

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/80028>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

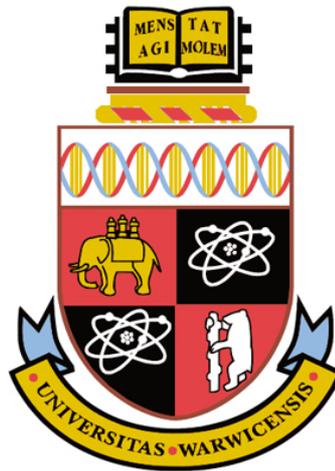
Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

Error Control in Bacterial Quorum Communications



Chenyao Bai

A thesis Submitted to the University of Warwick
in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Engineering

©May 2016



Table of Contents

List of Tables	vi
List of Figures	vii
Acknowledgements	xii
Declaration	xiv
List of Publications	xv
Abstract	xvi
Abbreviations	xvii
CHAPTER 1. Introduction	1
1.1 Background	1
1.1.1 Nanotechnology.....	1
1.1.2 Nanomachines	2
1.1.3 Interaction between nanomachines.....	4
1.1.4 Nanoscale communication requirements.....	5
1.1.5 Open challenges	6
1.2 Motivation of this work.....	8
1.3 Research objectives.....	10
1.4 Thesis contributions.....	13
1.5 Thesis outline	15
CHAPTER 2. Molecular communications	17
2.1 Introduction	17
2.2 MC architectures.....	18
2.2.1 MC components	19
2.2.2 MC process.....	20
2.3 MC types	21
2.3.1 Wired MC	21
2.3.2 Wireless MC.....	23
2.4 MC characteristics	27
2.5 Research aspects of MC.....	30

2.5.1	Modulation techniques.....	31
2.5.2	Channel modelling and noise analysis.....	33
2.5.3	Coding techniques	35
2.5.4	Network architectures and protocols	36
2.5.5	Simulation tools.....	39
2.6	Applications of MC	40
2.6.1	Biomedical applications.....	40
2.6.2	Industrial applications.....	41
2.6.3	Environmental applications.....	42
2.6.4	Information technology applications	42
2.7	Summary	43
CHAPTER 3. Bacterial communications		45
3.1	Introduction	45
3.2	Capabilities of bacteria	48
3.3	Quorum sensing.....	52
3.3.1	QS: an overview	52
3.3.2	QS in gram-negative bacteria.....	54
3.3.3	QS in gram-positive bacteria.....	62
3.4	QS applications.....	64
3.4.1	Engineered QS system.....	65
3.4.2	Biosensors	66
3.4.3	Biofouling control	66
3.5	Summary	67
CHAPTER 4. Communication model.....		70
4.1	Introduction	70
4.2	Existing models of QS	71
4.3	Two QS systems in <i>V. fischeri</i> bacteria	74
4.4	Bacterial communication scheme	75
4.4.1	Bacteria functionality model.....	75
4.4.2	Environmental illustration.....	79
4.4.3	Transmitter and receiver model.....	81
4.5	Channel model.....	85
4.5.1	Overview	85

4.5.2	The diffusion medium.....	86
4.5.3	The communication channel model.....	90
4.5.4	Communication discussions.....	99
4.6	Channel capacity analysis	100
4.7	Summary	105
CHAPTER 5.	FEC in bacterial communications.....	107
5.1	Introduction	107
5.2	Existing error correction coding for MC.....	110
5.3	FEC mechanisms	112
5.3.1	Digital communication system.....	112
5.3.2	Linear block codes.....	117
5.3.3	Cyclic codes	119
5.3.4	Convolutional codes	120
5.3.5	Turbo codes.....	122
5.3.6	Discussion	123
5.4	Hamming coding	124
5.4.1	Introduction of Hamming codes.....	124
5.4.2	Analytical results	128
5.5	Minimum Energy Coding	132
5.5.1	Introduction of MECs.....	132
5.5.2	Analytical results	135
5.6	Luby Transform (LT) coding	137
5.6.1	Introduction of LT codes	137
5.6.2	Analytical results.....	141
5.7	Summary	143
CHAPTER 6.	ARQ in bacterial communications.....	145
6.1	Introduction	145
6.2	Existing transmission protocols for MC	147
6.3	ARQ mechanisms	151
6.3.1	Overview	151
6.3.2	Cyclic redundancy check (CRC) codes	153
6.3.3	ARQ implementation.....	156
6.3.4	Discussions.....	164

6.4	Simulation results for SW-ARQ.....	167
6.4.1	Parameter setup.....	167
6.4.2	Simulation process.....	169
6.4.3	Results and discussions.....	171
6.5	Simulation results for GBN-ARQ.....	184
6.5.1	Parameter setup.....	184
6.5.2	Simulation process.....	185
6.5.3	Results and discussions.....	187
6.6	Simulation results for SR-ARQ.....	197
6.6.1	Parameter setup.....	197
6.6.2	Simulation process.....	198
6.6.3	Results and discussions.....	199
6.7	Comparisons of different ARQ schemes.....	209
6.8	Summary.....	211

CHAPTER 7. Throughput and energy efficiency based packet size

optimization of ARQ.....	216	
7.1	Introduction.....	216
7.2	Engineered molecular communication.....	219
7.2.1	Synthetic biology hierarchy.....	219
7.2.2	Technical advances.....	223
7.2.3	Synthetic biological examples.....	225
7.2.4	Discussions.....	230
7.3	Energy model.....	231
7.4	Mathematical analysis of throughput and energy efficiency.....	238
7.4.1	SW-ARQ.....	238
7.4.2	GBN-ARQ.....	242
7.4.3	SR-ARQ.....	244
7.5	Numerical results.....	245
7.5.1	Parameter setup.....	246
7.5.2	SW-ARQ.....	246
7.5.3	GBN-ARQ.....	249
7.5.4	SR-ARQ.....	251
7.6	Summary.....	254

CHAPTER 8. Conclusions and future works.....	256
8.1 Concluding remarks.....	257
8.2 Future work.....	270
Appendix A	273
Appendix B.....	278
References	280

List of Tables

Table 2.1 Comparison of communication characteristics.....	28
Table 3.1 Summary of characteristics of typical bacterial cell structures.	50
Table 4.1 Parameter settings [186].	84
Table 4.2 Binary Channel Model (Possibility Representation).....	92
Table 5.1 Parameters of Hamming coding path.	125
Table 6.1 Standard generator polynomials [240].	154
Table 7.1 Promoters and proteins lengths	242

List of Figures

Figure 2.1 Generic communication architecture.	18
Figure 2.2 Microtubule motor proteins [37].....	23
Figure 2.3 Calcium signalling MC based on (a) gap junction and (b) diffusion [5].	25
Figure 2.4 Flow of information through a MC network.	37
Figure 2.5 A set of functions provided by each layer [79].....	39
Figure 3.1 The structural components of a typical bacterial cell.....	49
Figure 3.2 The symbiotic relationship between bacteria and the host.....	58
Figure 3.3 The LuxI/LuxR QS system.	59
Figure 3.4 QS system in <i>V. harveyi</i>	61
Figure 3.5 A general model for peptide-mediated QS in gram-positive bacteria.	63
Figure 3.6 Overview of current QS applications.....	65
Figure 4.1 Bacterial communication: (a) bacterium structure; (b) GFP production..	77
Figure 4.2 Bacterial communication setup consisting of the transmitter node, the diffusion channel and the receiver node.....	79
Figure 4.3 AHL/LuxR binding probability for different concentration of signalling molecules surrounding the bacterium.	83
Figure 4.4 Luminescence over time for different AHL concentrations.....	85
Figure 4.5 Channel BER performance.....	95
Figure 4.6 BER considering more than one previous time slot.....	99
Figure 4.7 Binary asymmetric channel (BAC).....	102
Figure 4.8 Effect of threshold on the mutual information.	104
Figure 4.9 Channel capacity.....	105
Figure 4.10 Selected threshold when calculating capacity.	105
Figure 5.1 A general framework for digital communications [202].....	115

Figure 5.2 (3,1,3) convolutional encoder with constraint length 3.....	122
Figure 5.3 Hamming coding path.	125
Figure 5.4 BER comparison between (7, 4) and (15, 11) Hamming codes for transmission distance of $4\mu m$	128
Figure 5.5 Average energy per bit comparison between (7, 4) and (15, 11) Hamming codes for transmission distance of $4\mu m$	130
Figure 5.6 BER comparison between (7, 4) and (15, 11) Hamming codes for the number of molecules per bit of 500.	131
Figure 5.7 BER comparison between MECs and Hamming codes.	136
Figure 5.8 Average energy per bit comparison between MECs and Hamming codes.	136
Figure 5.9 BER comparison between LT codes and Hamming codes.	142
Figure 5.10 Average energy per bit comparison between LT codes and Hamming codes.	142
Figure 6.1 Relationship between packets and frames in ARQ.	152
Figure 6.2 Block diagram of the basic ARQ system.	156
Figure 6.3 Basic elements of ARQ [225].	158
Figure 6.4 Example of SW-ARQ.	160
Figure 6.5 Examples of GBN-ARQ and SW-ARQ schemes.	163
Figure 6.6 Sliding window example.	165
Figure 6.7 Channel performance of SW-ARQ for different transmission distances: (a) Transmission delay; (b) Efficiency; (c) BER; (d) channel throughput.	172
Figure 6.8 Channel performance of SW-ARQ for different number of bacteria: (a) Transmission delay; (b) Efficiency; (c) BER; (d) channel throughput.	174
Figure 6.9 Channel performance of SW-ARQ for different CRC polynomials: (a) Transmission delay; (b) Efficiency; (c) BER; (d) channel throughput.	175
Figure 6.10 Channel performance of SW-ARQ for different frame length: (a) Transmission delay; (b) Efficiency; (c) BER; (d) channel throughput.	177

Figure 6.11 Transmission time of SW-ARQ for different combinations of the main factors discussed in the text.....	179
Figure 6.12 Transmission efficiency of SW-ARQ for different combinations of main factors discussed in the text.....	181
Figure 6.13 Error performance of SW-ARQ for different combinations of the main factors discussed in the text.....	182
Figure 6.14 Channel throughput of SW-ARQ for different combinations of the main factors discussed in the text.....	184
Figure 6.15 Channel performance of GBN-ARQ for different transmission distances: (a) Transmission delay; (b) Efficiency; (c) BER; (d) channel throughput.....	188
Figure 6.16 Channel performance of GBN-ARQ for different window sizes: (a) Transmission delay; (b) Efficiency; (c) BER; (d) channel throughput.....	189
Figure 6.17 Channel performance of GBN-ARQ for different number of bacteria: (a) transmission delay; (b) efficiency; (c) BER; (d) channel throughput.....	191
Figure 6.18 Channel performance of GBN-ARQ for different frame lengths: (a) transmission delay; (b) efficiency; (c) BER; (d) channel throughput.....	192
Figure 6.19 Transmission time of GBN-ARQ for different combinations of the main factors discussed in the text.....	194
Figure 6.20 Transmission efficiency of GBN-ARQ for different combinations of the main factors discussed in the text.....	195
Figure 6.21 Error performance of GBN-ARQ for different combinations of the main factors discussed in the text.....	196
Figure 6.22 Channel throughput of GBN-ARQ for different combinations of the main factors discussed in the text.....	197
Figure 6.23 Channel performance of SR-ARQ for different transmission distances: (a) Transmission delay; (b) Efficiency; (c) BER; (d) channel throughput.....	200
Figure 6.24 Channel performance of SR-ARQ for different window sizes: (a) transmission delay; (b) channel throughput.....	202
Figure 6.25 Channel performance of SR-ARQ for different number of bacteria: (a) transmission delay; (b) efficiency; (c) BER; (d) channel throughput.....	202

Figure 6.26 Channel performance of SR-ARQ for different frame lengths: (a) transmission delay; (b) efficiency; (c) BER; (d) channel throughput.....	204
Figure 6.27 Transmission time of SR-ARQ for different combinations of the main factors discussed in the text.....	206
Figure 6.28 Transmission efficiency of SR-ARQ for different combinations of the main factors discussed in the text.....	207
Figure 6.29 Error performance of SR-ARQ for different combinations of the main factors discussed in the text.....	208
Figure 6.30 Channel throughput of SR-ARQ for different combinations of the main factors discussed in the text.....	209
Figure 6.31 Comparisons of channel performance for different ARQ schemes: (a) transmission delay; (b) efficiency; (c) BER; (d) channel throughput.....	210
Figure 7.1 A hierarchal analogy between computer engineering and synthetic biology [179].	222
Figure 7.2 Toggle switch design.....	227
Figure 7.3 Oscillator design.	228
Figure 7.4 Synthetic encoder and decoder for CRC-8.....	233
Figure 7.5 Four synthetic logic gates employed to establish the energy model.....	234
Figure 7.6 Throughput efficiency versus frame length for SW-ARQ	247
Figure 7.7 Energy efficiency versus frame length for SW-ARQ	248
Figure 7.8 SW-ARQ optimal frame length versus molecules per bit to maximize the throughput and energy efficiency, for different CRC polynomials.....	249
Figure 7.9 Throughput efficiency versus frame length for GBN-ARQ.....	250
Figure 7.10 Energy efficiency versus frame length for GBN-ARQ.....	250
Figure 7.11 GBN-ARQ optimal frame length versus molecules per bit to maximize the throughput and energy efficiency, for different window sizes.	251
Figure 7.12 Throughput and energy efficiency performances for SR-ARQ.....	252
Figure 7.13 SR-ARQ optimal frame length versus molecules per bit to maximize the throughput and energy efficiency, for different window sizes.....	253

Figure B.1 Comparison of BER performances between the proposed model in this thesis and the new model.279

Acknowledgements

I would never have been able to accomplish my thesis without the guidance of my supervisor, my friends, and support from my family.

First and foremost, I would like to express my deepest appreciation to my supervisor Dr Mark S. Leeson for the continuous support of my Ph.D. study and research, for his patience, motivation, enthusiasm, and immense knowledge. He has been supportive since the days I began working on nano-communications as an M.Sc. student. Since then, Mark has supported me not only by providing research guidance over four years, but also academically and emotionally through the rough road to finish this thesis. And during the most difficult times in my Ph.D. life, he gave me the moral support and the freedom I needed to move on.

I would also like to thank my second supervisor, Dr Matthew D. Higgins, for his encouragement, productive discussions and helping me to revise and submit my papers for the past few years.

I acknowledge the financial support of School of Engineering, University of Warwick. I am grateful to students in the Communication Networks (ComNet) Laboratory, especially Xiayang Wang, Yi Lu, Enzo and Roy, for providing fun and stimulating research environment. Also, I would thank my friend Yipei Ye and Hu Yuan. My experience at Warwick has been enjoyable thanks to the members of different nationalities and research interests.

Special thanks to my boyfriend Ji Hu, who was always there cheering me up and stood by me through the good times and bad.

Finally, and most importantly, I would like to express my thanks to my parents and my brother for their smile, support and motivation that helped me to get through many ups and downs. They always give me the strength and wisdom during these four years abroad and it is to them that I dedicate this thesis.

Declaration

This thesis is submitted in partial fulfilment for the degree of Doctor of Philosophy under the regulations formulated by the School of Engineering of the University of Warwick. I herewith declare that this thesis contains my own research performed under the supervision of Dr. Mark S. Leeson and Dr. Matthew D. Higgins, without assistance of third parties, unless stated otherwise. The research materials have not been submitted in any previous application for a degree at any other university. All sources of information are specifically acknowledged in the content.

List of Publications

Published Journals:

- [1]. **Chenyao. Bai**, M. S. Leeson, and M. D. Higgins, "Minimum energy channel codes for molecular communications," *Electronics Letters*, vol. 50, no. 23, pp. 1669-1671, 2014. DOI: 10.1049/el.2014.3345.
- [2]. **Chenyao. Bai**, M. S. Leeson, and Matthew D. Higgins, "Performance of SW-ARQ in bacterial quorum communications," *Nano Communication Networks*, vol. 6, no. 1, pp. 3–14, 2015. DOI: 10.1016/j.nancom.2014.11.001.
- [3]. **Chenyao Bai**, M. S. Leeson, and M. D. Higgins, "Analysis of ARQ protocols for bacterial quorum communications," *Nano Communication Networks*, vol. 7, no. 1, pp. 65–79, 2016. DOI: 10.1016/j.nancom.2015.12.001.
- [4]. **Chenyao Bai**, M. S. Leeson, M. D. Higgins and Yi Lu, "Throughput and Energy Efficiency Based Packet Size Optimization of ARQ Protocols in Bacterial Quorum Communications," *Transactions on Emerging Telecommunications Technologies*, accepted for publication, 2016.

Published Book Chapter:

- [5]. M. S. Leeson, M. D. Higgins, **Chenyao Bai**, Y. Lu, X. Wang and R. Yu, "The Use of Coding and Protocols within Molecular Communication Systems", for Modeling, Methodologies and Tools for Molecular and Nano-scale Communications, J. Suzuki, T. Nakano and M. J. Moore (eds.), Springer, due summer 2016.

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 (MC), bacterial populations can be programmed as bio-transceivers to establish information transmission using molecules. In this work, to investigate the key features for MC, 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 and energy efficiency, different error control techniques, including forward error correction (FEC) and Automatic Repeat reQuest (ARQ) are taken into consideration. For FEC, this work presents a comparison of the performance of traditional Hamming codes, Minimum Energy Codes (MEC) and Luby Transform (LT) codes over the channel. In addition, it applied several ARQ protocols, namely Stop-N-Wait (SW-ARQ), Go-Back-N (GBN-ARQ), and Selective-Repeat (SR-ARQ) combined with error detection codes to achieve better reliability. Results show that both the FEC and ARQ techniques can enhance the channel reliability, and that ARQ can resolve the issue of out-of-sequence and duplicate packet delivery. Moreover, this work further addresses the question of optimal frame size for data communication in this channel capacity and energy constrained bacterial quorum communication system. A novel energy model which is constructed using the experimental validated synthetic logic gates has been proposed to help with the optimization process. The optimal fixed frame length is determined for a set of channel parameters by maximizing the throughput and energy efficiency matrix.

Abbreviations

3-oxo-C12-HSL:	<i>N</i> -(3-oxododecanoyl)-L-homoserine lactone
3-oxo-C6-HSL:	<i>N</i> -(3-oxo-hexanoyl)-L-homoserine lactone
ABC:	ATP-binding cassette
ACK:	positive acknowledgement frames
ADP:	adenosine diphosphate
AHL:	<i>N</i> -acyl homoserine lactones
AI-1:	<i>N</i> -(3-hydroxybutanoyl) homoserine lactone
AIGN:	additive inverse Gaussian noise
ARP:	address resolution protocol
ARQ:	Automatic Repeat reQuest
ASK:	amplitude-shift keying
ATP:	adenosine triphosphate
BAC:	binary asymmetric channel
BAN:	body area networks
BER:	bit error rate
BSC:	binary symmetric channel
C4-HSL:	<i>N</i> -butanoyl-L-homoserine lactone
C8-HSL:	<i>N</i> -3-oxo-octanoyl-L-Homoserine lactone
CA:	concentration aware
CAI-1:	(<i>S</i>)-3-hydroxytridecan-4-one
CPEC:	Circular Polymerase Extension Cloning
CRC:	cyclic redundancy check
CSK:	Concentration Shift Keying
CtS:	Communication through Silence

FBD:	furanosyl borate diester
FEC:	Forward Error Correction
FSK:	frequency-shift keying
GBN-ARQ:	Go-Back-N Automatic Repeat reQuest
GFP	Green fluorescent protein
HARQ	hybrid automatic repeat request
HLDC:	High-Level Data Link Control
HSL:	homoserine lactone
ICMP:	Internet Control Message Protocol
ICT:	Information and Communication Technologies
IP:	Internet Protocol
IRSK:	isomer-based ratio shift keying
ISI:	Inter-symbol interference
LDPC:	Low-Density Parity-Check
LT:	Luby Transform
MAC:	Medium Access Control
MAP:	Maximum A Posteriori
MARCO:	Molecular ARray-based Communication
MC:	molecular communication
MEC:	Minimum Energy Codes
ML:	maximum likelihood
MoSK:	Molecule Shift Keying
MTSK:	Molecular Transition Shift Keying
NAK:	negative acknowledgement frames
NAODV:	Nano Ad hoc On-Demand Distance Vector
NEMS/MEMS:	nano- and micro-electromechanical systems
NMR:	nuclear magnetic resonance

OOK:	On-Off Keying
OSI:	Open System Interconnection
PAM:	Pulse Amplitude Modulation
PCR:	polymerase chain reaction
PMF:	Poisson-Binomial probability mass function
PMID:	PubMed Identifier
PPM:	pulse position modulation
PSK:	phase-shift keying
QAM:	quadrature amplitude modulation
QS:	quorum sensing
RM:	Reed Muller
RS:	Reed Solomon
SLIC:	ligation-independent cloning
SOCC:	self-orthogonal convolutional codes
SOVA:	Soft Output Viterbi Algorithm
SR-ARQ:	Selective-Repeat Automatic Repeat reQuest
ssDNA:	single-stranded DNAs
SW-ARQ:	Stop-N-Wait Automatic Repeat reQuest
TCP:	Transmission Control Protocol
UAS:	upstream activating sequences
UDP:	User Datagram Protocol

CHAPTER 1.

INTRODUCTION

1.1 Background

1.1.1 Nanotechnology

Nanotechnology is a new interdisciplinary field, based on the knowledge of diverse scientific areas such as chemistry, physics, molecular biology, materials science, computer science, and engineering [1]. It focuses on the design, characterization, production and application of functional materials, devices, and systems through the control of matter on the nanoscale (1-100 nanometres) and the exploitation of novel phenomena and properties (physical, chemical, and biological) at that length scale [2], where the Newtonian approximations are no longer applicable, resulting in quantum physics. One of the top concerns regarding the future of nanotechnology is molecular manufacturing, which involves the use of nanoscale tools and non-biological processes to build structures, devices and systems at the molecular level, with virtually no defects. Also, it is expected that more wonders occur in the medical field. Due to the fact that the research of nanotechnology is not mature, it would be almost impossible to predict the future developments over the next ten or twenty years in this field. However, it has been proved that nanotechnology could be a huge contributor in shaping the world from atom up, using atomic level manipulation to transform and construct a broad range of new materials, devices, living organisms and technological systems [3]. In the past few years, nanotechnology has been

utilized in many fields such as biomedical engineering, information and communication, environmental research and industrial technology. The detailed applications will be discussed in the following chapter.

1.1.2 Nanomachines

One of the main objectives of nanotechnology is to successfully arrange nanomachines, which can be considered as the most basic functional unit [4]. Generally, nanomachines are devices consisting of nanoscale components and being able to perform specific tasks at nano-level, such as communicating, computing, data storing, sensing and/or actuation [5]. The actual molecular biological systems existing in the human body are nanomachines themselves. However, nowadays only some primitive nanomachines have been tested. Most of them are still in the theoretical research and development stage. In addition, nanomachines can be employed as building blocks for more complex systems, such as nano-robots [6]. A single nanomachine is fairly sensitive to its environment, especially when perturbed by thermal collisions with the surrounding molecules. There are three different approaches for the development of nanomachines, namely, the top-down, bottom-up and bio-hybrid approaches [5]. Among them, the top-down approach is the most commonly used technique currently, especially in electronic fields such as microelectromechanical systems (MEMS) technologies. Top-down nanomachines are developed by scaling down existing microelectronic components, without atomic level manipulation [7]. Bottom-up technology, which is also called molecular nanotechnology or molecular manufacturing, is a promising method which creates nanomachines from molecular components either manually or through self-assembly, operated with the chemical, biological and physical forces between molecules [5]. The bio-hybrid approach is based on the use of existing biological systems and

mechanisms as components in synthetically engineered bio-nanomachines [8]. It is beneficial particularly to biomedical applications due to their inherent compatibility with natural biological processes and reduced energy requirements [9]. At the present stage, complex nanomachines cannot be established following the top-down approach. However, the development of nanomachines following the bottom-up and bio-hybrid approaches can offer promising solutions in the near term.

Generally, a nanomachine may consist of one or more components, with different levels of complexity. For a fairly comprehensive nanomachine, the components include the control unit, the communication unit, the reproduction unit, the power unit and sensor and actuators [5]. The control unit is the central processing unit, dealing with instructions of particular tasks and controlling all the other components of the nanomachine. The communication unit (e.g., gap junctions and ligand receptors in living cells) is used to transmit and receive the transmitted information. The reproduction unit intends to fabricate each component of a nanomachine using external sources and then assemble them to duplicate the nanomachine. The power unit provides energy to the components of the nanomachine when it is in operation, by scavenging energy from external sources (e.g., light, temperature) and reserving this power for later consumption. Sensors and actuators are used to sense the fluctuation of the environment and react accordingly, acting as the interface between the environment and the nanomachine [5]. However, such nanomachines have not been constructed artificially because of the extremely high complexity and the lack of fabrication technologies to handle and assemble molecular structures precisely. Considering the components of the typical nanomachine stated above, nanomachines have some significant expected features, including self-assembly, self-replication, locomotion and communication and being self-contained [5]. Self-assembly refers to

the ability of a nanomachine to form an organized structure from disordered elements, without external forces. Self-replication allows a nanomachine to make copies of itself using external elements, in order to implement complex and macroscopic tasks. Locomotion is the ability to move from one place to another. Communication between nanomachines, just as in living cells, is required, enabling them to complete more complex tasks in a cooperative manner and achieving decentralized and distributed intelligence [5]. In addition, to become self-contained, each nanomachine contains a set of instructions to realize the intended task, which can be embedded in their molecular structure [5].

1.1.3 Interaction between nanomachines

As stated above, nanomachines cannot execute complex tasks individually. Nanonetworks, the interconnection of different nanomachines, will expand the capabilities of single nanomachines by providing them with a method to cooperate and exchange information and commands, allowing the fulfilment of more complex tasks [10]. Also, instead of the extremely limited workspace of a single nanomachine, nanonetworks will enable the interaction with remote nanomachines by means of broadcasting and multi-hop communication mechanisms, expanding the application range from metres to kilometres [11]. Communication between nanomachines can be realized through nano-mechanical, acoustic, electromagnetic, chemical or molecular communications (MC) [11]. Among them, MC, inspired by the communication mechanisms in living organisms and defined formally as the use of molecules as messages between transmitters and receivers, is considered as the most promising approach for nanonetworks. The process of MC consists of the encoding of information inside the molecules, the release of these molecules as a response to a certain command by the transmitter, and the reaction to specific particles by

particular signal-transducing mechanisms at the receiver. The reasons that MC outperforms the other communication schemes are as follows. On one hand, unlike the other communication techniques, the integration process of molecular transceivers, which are intrinsically at the nanoscale, is more feasible due to the size and natural domain, making MC a novel and promising way to transmit information between nanomachines [5]. On the other hand, in nanomechanical communication, transceivers need to be in direct contact; this is not a restriction for MC over large areas, where the transmitter can be located remotely as long as the transmitted molecules are able to reach the intended receiver [5]. There are several MC modes, some of which will be presented in Chapter 2. A promising method of MC is bacterial quorum sensing (QS), exchanging chemical signals to estimate the bacterial population [12]. A brief introduction of QS will be given in the following sections. It should be noted that this thesis focuses on the diffusion-based MC through QS, where information symbols are encoded into the number of signalling molecules and propagate within the medium undergoing diffusion. The receiver observes the received concentration of the molecules at its location and thus decodes the information accordingly.

1.1.4 Nanoscale communication requirements

In order to establish MC networks, bio-nanomachines must have the ability to encode messages with molecules using either concentration encoding or molecular encoding. The former involves modulating the concentration of molecules in order to transmit information. The latter uses the physical shape or the structure of a molecule to encode information [13]. Both encoding techniques must be decodable at the intended receiver. This indicates that the bio-nanomachines should have the capabilities of encoding, sending and receiving information via MC, which means

that the implementation of MC protocols is required. For instance, the IEEE P1906.1 Standards Working Group for Nanonetworking was established in 2011 to promote the standardization of MC [14]. In addition, synthetic biology is helpful to the implementation of MC, providing methods of reusing and reengineering existing genetic, enzymatic and cellular systems to design and create synthetic molecular or genetic computational parts (e.g., promoter, transcription factors, degradation tags and transcriptional terminators), devices and circuits such as oscillators, toggle switches and biologic gates [15].

1.1.5 Open challenges

The progress and the potential of nanonetworks have been presented in the previous sections. Although it is very promising that nanomachines will possibly be built in the near future, it is still not clear how nanomachines will actually communicate [7]. At current stage, communication between nanomachines is experiencing several key challenges that are required to be overcome before the development of the system design.

The rapid increase of the propagation delay as a function of the transmission distance will become a significant limitation in diffusion-based MC networks, making it expensive for the application of some handshake-based protocols and retransmission protocols in some situations where the transmission delay is fairly large [16]. Also, due to the propagation delay, synchronization among nanomachines will be another challenging task.

Channel attenuation is another restriction, which will limit the transmission range of nanomachines in an MC network [16]. One possible solution is to use a cooperative approach for the transmission of information. Moreover, MC is a very slow (nm to μm per second) communication scheme compared to conventional radio-based

communications [17]. For example, for messenger molecules with a radius of 5 nm diffusing in human blood, it takes 60 s for the maximum number of messenger molecules to reach the receiver which is located 100 μm away from the transmitter [18]. Thus, improving the communication speed is another big challenge in MC. In addition, most of the currently investigated MC systems are based on molecular diffusion. In these situations, because of the stochastic nature of Brownian motion, the molecular channel distorts the transmitted signals, which leads to a further challenge to the successful detection of the signals and represents the main limitation in the achievable bandwidth in diffusion-based MC. Apart from these effects, the signal distortion will cause intersymbol interference (ISI), resulting in a highly unreliable communication system [19]. Therefore, novel techniques to combat the effect of ISI need to be investigated.

Furthermore, it has been stated above that synthetic biology will help with the development of MC but it remains an appreciable challenge to combine the protocols of both domains. For example, engineering an interface for a chemical enzyme-based logic circuit to control an existing and entirely separate calcium signalling process in a biological cell is difficult. Taking a bio-hybrid approach does have the advantage of bio-compatibility but significant constraints still need to be addressed in the integration of distinct biological processes [20]. In addition, when using existing biological mechanisms to design and create MC networks, the characteristics of these biological components and processes may become significant challenges. For example, some biological components (e.g., viruses or DNA) have a degradation rate or half-life, which will have an impact on the effective communication range. Moreover, biological signalling mechanisms, such as calcium signalling, have a significant refractory time after the signals are released, during which they cannot

generate or forward signals [20]. This is considered as a key constraint in multiple access channels which contain more than one transmitting nanomachine.

1.2 Motivation of this work

Nanotechnology in general, and nanonetworks in particular, are research areas that have attracted a great deal of attention over the last couple of decades, and this interest will continue to increase in the future. The first significant motivation of this study is to discover new schemes for the communication between nanomachines. For MC, synchronization between elements of a nanonetwork is a commonly required feature to build any network architecture, which, however, is not easy to accomplish due to the inherent characteristics of the nanomachines, including issues involving operation complexity and energy requirements. It would be possible to maintain nanomachines working continuously using individual clocks but this is not energy efficient. Also, a global clock is not suitable for diffusion-based MC networks because of the low transmission speed of information [21]. Thus, these factors inspire this work to apply natural existing biological mechanisms, in particular bacterial QS [12], to achieve synchronization between two bacterial colony nodes in a nanonetwork. QS is the mechanism for bacteria to communicate with each other using signalling molecules; and 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, and therefore the population as a whole can make a coordinated response [12]. 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 [12], which will in turn increase the density of signalling molecules over time producing a positive feedback process. The output of the QS mechanism can be in

various forms and one example is the production of luminescence [22]. Also, because of the fact that many biological mechanisms found in biological cells are able to provide interfacing and switching functions that could be re-purposed for MC [20], in this thesis, it is believed that QS can provide nanomachines with communication capabilities to enable synchronization or coordination with the nodes within the environment.

In diffusion-based MC, the molecules which propagate undergoing Brownian motion arrive at the receiver in a probabilistic manner. In other words, not all of the molecules transmitted can reach the receiver located at a given distance from the transmitter at a given time, resulting in concentration attenuation. Although the diffusion mechanism has been known for more than a hundred years [23], the concept of using diffusion as a communication technique is relatively new [5]. Thus, inspired by the phenomenon of bacterial QS which depends on the diffusion of signalling molecules, an MC system based on diffusion is presented in this thesis.

Across all branches of communications, a primary concern has always been on the reliability of the channel, and being able to control or correct any errors that have been introduced into the channel. The majority of diffusion-based MC channels are characterized by a poor quality physical link, caused by the random diffusion process. As a result, the channel error performance is often not good enough. Generally, the MC channel experiences time-varying signal fading due to random walk motion of the molecules, which will contribute to the ISI effect that this thesis has focused on. In particular, ISI degrades the performance of a diffusion-based MC system, especially at high transmission rates, possibly resulting in data corruption. Thus, to establish reliable communication over such a channel, error correction techniques, including forward error correction (FEC) and Automatic Repeat reQuest (ARQ), are

required by which erroneously received data packets are corrected or retransmitted [24]. In addition, due to their extremely small size and limited energy storage capabilities, nanomachines can only utilise limited energy, which makes it essential to develop energy efficient communication techniques. Thus, any error control schemes for MC should consider energy dissipation as an essential metric. Generally, ARQ protocols are always implemented together with error detection techniques. Compared with FEC techniques, ARQ protocols can ensure in-sequence delivery and are considered to be energy efficient because error detection requires much simpler decoding operations than error correction does [7]. Furthermore, since the bio-entities are constrained by information transmission rate, channel capacity and energy limitations [9], the optimal frame lengths for different ARQ schemes should be investigated to obtain better throughput and energy performances. Although there have been several studies on packet size optimisation in traditional wireless and wired networks [10], none of them are directly applicable to the bacterial scenario. In addition, inspired by the advancement of MC and other relevant fields, the implementation of coding or protocol components can be achieved by using biological computation techniques and components from synthetic and molecular biology, for example, using existing synthetic biological logic gates [25] to design and develop cell-based protocol components that can be integrated with the physical channel.

1.3 Research objectives

The focus of this Ph.D. thesis is on bacterial QS based, diffusive MC, where the propagation of information molecules between two populations of bacteria is realized through free diffusion in the medium. It has been stated in the previous section that this choice is motivated by two aspects: (a) as will be discussed in detail

in Chapter 2, the diffusion-based approach is identified as the most fundamental type of MC among different communication modes discussed in the literature; (b) QS is supposed to be an efficient method for the synchronization of transceivers. Also, as a consequence of the differences between MC and classical electromagnetic communication paradigms that will be described in detail in Section 2.4, the classical communication engineering models and techniques are not directly applicable to the design of QS based diffusive MC systems. Therefore, an entirely new MC paradigm is required to be built from scratch. The research objectives addressed in this thesis have been identified to meet this requirement, and are summarized in the following.

The first research objective is to develop a bacterial QS based diffusive MC model, involving the mathematical characterization of the main communication processes, including the transmission, propagation and reception of molecules for the exchange of transmitted information between the transmitter and receiver nodes. In this design, both the transmitter and receiver nodes are composed of a population of bacteria. In addition, this design models the three physical processes, including the transmitter model involved in the emission of molecules at the transmitter, the channel model involved in the diffusion of molecules within the medium, and the receiver model involved in the chemical reaction. The probability distribution function has been provided for each of the three processes.

Besides the mathematical modelling of the transceivers and the communication channel, the second research objective is to identify and model the noise that affects the system performance. For a communication system, it is fundamental to model the noise source to help the design of system components, achieving a better reliability of exchanging information between the transmitter and receiver. In this thesis, the major noise source for diffusion-based MC, caused by ISI effect, is taken into

account. A mathematical analysis of the physical processes which generate the ISI effect has been presented in Chapter 4. The diffusion-based MC system has two main features, including a channel and a signal-dependent noise. Based on some simplifying assumptions, a pure theoretical approach has been made to achieve analytical closed-form expressions that relate the channel performances, including bit error rate (BER) and channel capacity, to physical parameters, such as the diffusion coefficient, the bacterial population in the transceivers, the transmission distance, and the number of molecules released at the start of each time slot.

The third research objective is to apply classical channel coding techniques in MC systems. Specifically, the scope is to show that conventional error correction codes are sufficient to provide acceptable error performance when transmitting information messages over noisy MC channels. The objective is intended to be achieved by performing numerical simulations, in order to observe the performance of various kinds of channel codes in the proposed MC system. Also, under some simplifying assumptions, the energy consumption of implementing channel codes will also be investigated.

The fourth research objective is the mapping of conventional protocols, in particular, ARQ protocols, to MC systems, to enhance system reliability. Specifically, an investigation into how the existing communication protocol concepts can be applied to bio-nanomachines to successfully use MC paradigms for communication among nanomachines has been made. The aim of an ARQ protocol is to organize retransmission of erroneously received frames or packets. Various system parameters need to be investigated to help design and develop an ARQ protocol which is as efficient as possible (with better error performance), but simple to implement.

Based on the fourth objective, the fifth objective is to utilise synthetic biological components for protocol enhancement, in particular, optimising the throughput and energy efficiency by controlling the packet size. The central idea of this optimization is to estimate the throughput efficiency or energy efficiency and adjust the packet size accordingly.

It should be noted that all the simulation work throughout this thesis is implemented using the industry standard tool MATLAB.

1.4 Thesis contributions

It is believed that this thesis makes significant contributions to the research of MC and bio-nanotechnology. The main achievements contained in the thesis are summarised in this section.

In this thesis, the first contribution is the establishment of a bacterial QS based MC system, transmitting information over ranges up to a few hundred micrometres via the diffusion of identical signalling molecules within the medium containing the transmitter node and the receiver node. Both the transmitter and receiver nodes are composed of a population of the well-studied bacteria species, *Vibrio fischeri*, which is most famous for bioluminescence and has been used to study the toxicity of aquatic environments. The transmitter releases carrier molecules into the medium in which the nanomachines are suspended and then the molecules diffuse by means of Brownian motion which causes noisy reception in terms of arrival time at the receiver. 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 signalling molecules, presents a major challenge for the robust function and performance of natural and engineered QS networks [6]. For the bacterial communication system proposed in the thesis, the effect of ISI caused by molecular diffusion is considered as a primary source of impairment, which makes it necessary to apply error correction techniques for reliable transmission.

Thus, the second major contribution is to enhance the channel reliability using channel codes, including minimum energy codes (MEC), and Luby Transform (LT) codes, which have been applied for the first time in MC and compared with the more traditional Hamming codes, both in respect to BER and energy consumption.

Moreover, although ARQ techniques, including Stop-N-Wait (SW-ARQ), Go-Back-N (GBN-ARQ) and Selective-Repeat (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, which is considered as the third major contribution of this thesis. 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 communications. 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.

Finally, another contribution is that this is the first time that the application of known experimentally validated genetic logic gates has been proposed for mapping existing ARQ protocols to QS based MC. Instead of applying the previously presented energy models in [15, 21], the thesis proposes a new energy model, which utilises

the recently invented synthetic biological logic gates in [20]. Also, it addresses the question of optimal frame size for data transmission in channel capacity and energy constrained bacterial quorum communication system to enhance throughput and energy efficiency. Specifically, this research shows how different protocols and frame lengths can be chosen to optimize various 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.

1.5 Thesis outline

This Ph.D. thesis is organized as follows:

In Chapter 1, the background, the motivation, the research objectives and the contributions of this thesis are presented.

In Chapter 2, a general discussion of MC is presented. The communication components, communication processes and communication types of MC are introduced and explained at first, followed by the research aspects and potential applications that could benefit from it.

In Chapter 3, the characteristics of bacterial communications are outlined. The capabilities of bacteria are discussed, as well as the QS mechanisms in both the gram-negative and gram-positive bacteria. Also, the potential applications of QS are presented.

In Chapter 4, the proposed bacterial communication model, which includes the transmitter and receiver model and the channel model, is discussed. Before the description of the characteristics of the transceivers and the channel, existing QS

models are reviewed. Also, the QS systems of the bacteria used in the proposed model, specifically *V. fischeri*, are analysed. Besides, channel capacity analysis is also been presented.

In Chapter 5, the existing error correction coding schemes for MC are reviewed, followed by a brief description of FEC mechanisms. In addition, two kinds of channel codes, specifically MEC and LT codes, are applied to the diffusion-based MC channel and compared with traditional Hamming codes. Both error performance and energy consumption are considered to analyse the channel performances.

In Chapter 6, a brief review of the existing transmission protocols for MC is given, followed by the description of ARQ mechanisms. In addition, the ARQ protocols, including SW-ARQ, GBN-ARQ, SR-ARQ, and error detection codes, are applied to enhance the system performance, and the simulation results and discussions in terms of transmission delay, transmission efficiency, error performance and channel throughput are presented. Moreover, this chapter makes a comparison of the three different ARQ schemes.

In Chapter 7, a brief discussion of synthetic biology, including the technical advances and some synthetic biological examples, is presented. Also, with the introduction of a new energy model, applying existing synthetic logic gates to achieve the implementation of ARQ protocols as well as error detection codes, the mathematical analysis of the throughput and energy efficiency is introduced. This is followed by the numerical results and discussions, where the optimal packet size is determined in order to maximize the throughput and energy efficiency.

Finally, Chapter 8 gives the concluding remarks of the thesis and possible future work.

CHAPTER 2.

MOLECULAR COMMUNICATIONS

2.1 Introduction

MC is a novel and promising approach for the nanoscale communication paradigm, enhancing capabilities for nano-devices, and at the same time opening new opportunities for numerous applications. It has been used for the intracellular, inter-cellular, inter-organ and inter-species communications [26]. Also, MC has been used extensively as one of the key technologies in the synthetic biology and bio-manufacturing fields of research [27]. As a bio-inspired communication paradigm, it uses molecules to transport information between bio-nanomachines over a short (nanoscale or microscale) range, which can be realized based on molecular diffusion or molecular motor-based microtubules [14]. The nanomachine is defined as the most basic functional unit, which is able to perform various tasks, such as computing, data storage, sensing and actuation [5]. In MC, information is encoded onto and decoded from molecules, for example protein and DNA molecules, rather than utilising electrons or electromagnetic waves as in conventional telecommunication schemes. Information molecules are transmitted by the communication sender, propagated in the environment, and received by the communication receptor, which in turn decodes the information (i.e., some chemical reacts to the information molecules). Currently, research on MC demands novel solutions, including the identification of existing MC mechanisms, the establishment of the foundations of

molecular information theory, and the development of architectures and networking protocols for nanomachines. This chapter addresses the architectures and features of MC, as well as the current research topics of this new communication paradigm.

This chapter is organized as follows: in Section 2.2, MC architectures, including the communication components and communication process, are introduced, followed by the description of MC types in Section 2.3. In Section 2.4, the key features and characteristics of MC are explained. After that, a brief review of the current research aspects of MC is given in Section 2.5, followed by the illustration of application fields in Section 2.5.5. Finally, Section 2.7 presents a brief summary of the chapter.

2.2 MC architectures

From Information and Communication Technologies (ICT) theory, the components of a typical MC system are similar to those of the traditional communication system. The five ordinary primary components are the transmitter, the receiver, the information message, the carrier and the propagation medium. All of these components are indispensable. Figure 2.1 shows a schematic diagram of the generic communication architecture.

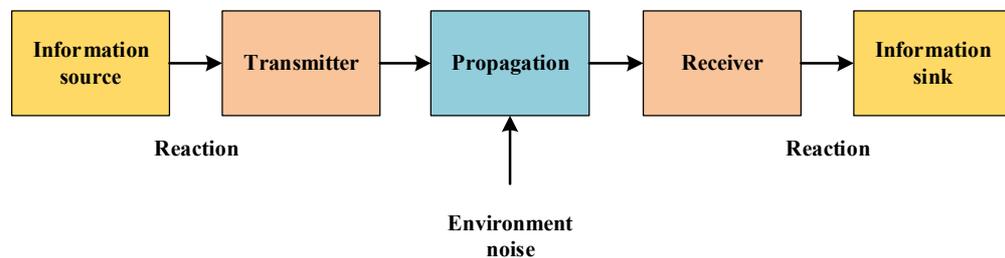


Figure 2.1 Generic communication architecture.

2.2.1 MC components

As shown in Figure 2.1, the generic MC system consists of information molecules, such as proteins or ions, which represent the information to be transmitted, transmitter nanomachines that emit the information molecules, receiver nanomachines that receive and react to information molecules and the environment in which the information molecules propagate from the transmitter to the receiver. The communication system may also include transport carrier molecules, such as myosin and dynein molecular motors [14], which can transport information molecules through the environment using chemical energy.

In Figure 2.1, the information source, such as some events and chemical reactions, has information to send through a transmitter to a receiver. The information sink is the destination where the desired biological reaction is performed. For a MC system, the transmitter, which is always a modified living cell or a biomedical implant, is used to encode a sensed chemical reaction onto information molecules and releases the information molecules into the environment. Some molecules reach the destination by free diffusion, while others need a carrier (e.g., the aforementioned molecular motors which will be discussed later) to bind and facilitate transport of information molecules. Upon arriving, the information molecules are detected and decoded by the receiver, resulting in a desired biological reaction at the information sink. The receiver should have the ability to extract the information molecules from the medium. Here, the information molecule is supposed to have a specific structure which can be easily recognized by the receiver and not to react with other molecules in the medium. In MC, information can be described as a biological reaction at a receiver; and the information molecules are used to store such reaction information. In some situations, to transmit the information successfully and efficiently, the

molecular carriers are used to protect molecules from external noise sources from the environment, making it possible for the creation of multiple independent channels within the same medium [5]. The most widely used information molecules in biological systems are molecular motors, including some specific proteins such as kinesin, dynein and myosin, and calcium ions. For MC, the medium, which represents the space in which the information molecules propagate, can be either in-body or environmental monitoring medium. Generally, the medium should be aqueous with necessary molecules at appropriate biological conditions (e.g., human body temperature) [8].

2.2.2 MC process

The communication processes of MC include encoding, sending, propagation, receiving and decoding. Encoding is the process by which a transmitter nanomachine senses some information (e.g., biological reaction or stimulus) inside or outside the nanomachine and translates the information into information molecules that the receiver can capture or detect. The information can be encoded by the polarity, motion, magnetization and structures of molecules according to [28]. Also, information can be encoded in the environment, for example, by having the transmitter release molecules that modify the environment and by having a receiver detect the changes in the environment [29]. Sending is the process by which the transmitter releases the information molecules into the environment. For example, a transmitter may emit ligands towards corresponding ligand receptors of nearby cells, leading to the generation of calcium waves which can propagate between cells [29]. Propagation is the process by which the information molecules travel through the medium (e.g., undergoing Brownian motion or along microtubules). Receiving is the process of detecting and capturing the carrier molecules or information molecules

propagating in the environment by the receiver. The receiver may contain a selective receptor, such as a receptor that is sensitive to calcium ions or specific peptides, to capture molecules, or contain gap junctions that allow molecules to flow into the cell. Decoding refers to the process of extracting the useful information from the carrier molecules by means such as some specific reaction with the cell membrane, data storage and actuation commands.

2.3 MC types

Generally, MC can be divided into wired and wireless branches [30]. In wired MC, information is transmitted via some specific types of physical link, including the neuron, blood vessel and microtubule [31]. However, in wireless MC, information is transmitted through the diffusive propagation of chemical molecules, and can be further classified depending on whether there is energy consumption during the propagation process.

2.3.1 Wired MC

In wired MC, a separate physical mechanism is used to guide the transport of information molecules. Several such mechanisms exist in nature providing wired transport directionality for molecule payloads that can be applied to create nanonetworks. For example, membrane nanotubes, which have important functions in various processes throughout cell biology, are considered as a significant type of molecular channel for both inter- and intracellular communications [32]. This approach represents a novel mechanism that could enable the formation of supra cellular structures where nutrients, genetic material and signalling molecules could be efficiently shared between interconnected cells [32]. Besides, molecular motor transport is also a promising method for MC. Generally, in intracellular transport, communication within a biological cell is performed using molecules carried by

molecular motors, which are always proteins or protein complexes that can transform chemical energy into mechanical work at the molecular scale [33]. Instead of free diffusion in the intermediate environment, the molecular motors walk internally along cytoskeletal tracks. This means that molecules move in the direction corresponding to the cytoskeletal track, which is made up of microtubules or actin filaments and is organized by a cell for various functionalities [34]. Microtubules are responsible for a great amount of cell movements, for example, the intracellular transport, the positioning of membrane vesicles and organelles, the chromosome segregation during the process of cell mitosis, and the beating of cilia and flagella [35]. Microtubules can be considered as tracks within the cell, on which biological materials such as vesicles or organelles can be transported. Movement along microtubules relies on the action of motor proteins (e.g., kinesin and dynein) which utilize energy from ATP hydrolysis to produce force and movement [35]. Generally, kinesin and dynein move along microtubules in opposite directions, specifically, kinesin moves towards the positive end while dynein moves towards the negative end. The movements of both motors are shown in Figure 2.2 [15]. Thus, molecular motors can control the directionality of transport by using different kinesin and dynein ratios. For example, more active kinesin motors will result in walking towards the positive end of a microtubule, while more dynein will lead to the reverse effect. In addition, actin filaments can also function as tracks for intracellular traffic in a similar fashion to microtubules. A molecular motor, named myosin, moves along actin filaments towards their positive extremity [36].

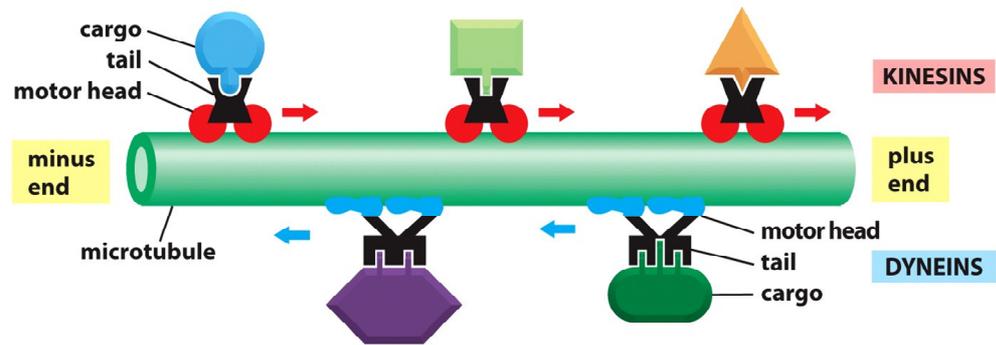


Figure 2.2 Microtubule motor proteins [37].

Kinesin and dynein move along microtubules in opposite directions, towards the plus (i.e., away from the centre of the cell) and minus (i.e., towards the centre of the cell) ends, respectively. Both groups of motor proteins have some features in common. For example, both of them have a globular ATP-binding heads that function as the motor domain and interact with the microtubules. Also, both of them have a tail domain that is involved in binding the cargo.

The wired approaches of MC provide a direct physical connection between the transmitter and receiver, which can operate in a unicast mode. Compared with the wireless approaches discussed later, the wired solutions will not suffer from attenuation or dilution due to transmission distance and noise. Thus they do not need to transmit as many message molecules as in the case of wireless approaches, leading to a more energy efficient transmission. In the next section, the wireless approaches will be discussed to support these features.

2.3.2 Wireless MC

Compared with conventional wireless communications which typically use electromagnetic waves as information carriers to propagate through space, MC perspective utilises molecules for communication among nanomachines. In wireless MC, the transport of signalling molecules can be classified as passive or active [36].

In passive transport, the signalling molecules propagate through the medium by diffusion undergoing Brownian motion, while in active transport, the signalling molecules propagate in a particular direction, ideally towards the intended receiver, with the involvement of some form of propulsion using chemical energy.

A. *Passive transport based MC*

Passive transport provides a simple method of transmitting signalling molecules for both intra- and intercellular communications [33]. In passive transport, signalling molecules randomly diffuse in the intermediate environment, uniformly distributed in all available directions, making the method particularly suited to environments that are highly dynamic, unfeasible and unpredictable [33]. Generally, passive transport requires a large number of molecules to be transmitted to reach a remote destination.

Passive transport can be classified into two mechanisms: (a) calcium signalling MC which is either free diffusion-based or gap junction mediated diffusion-based; (b) reaction-diffusion-based MC [33].

In MC based on calcium signalling, the information is represented by the concentration level of molecules, modulated in frequency or in amplitude. There are two different communication schemes for calcium signalling based communication. One is through indirect contact, termed free diffusion-based MC, where nanomachines are located in separated places without any physical contact. Most of the research in wireless MC is based on the molecular diffusion process, whereby the transmitter emits a quantity of molecules which diffuse passively to the receiver. The communication steps are as follows. In the initial step, the information molecules, which may be of the order of hundreds or even thousands, are released into the medium, which will cause an abruptly increasing concentration of molecules around

the transmitter. Then, the molecules will propagate through the medium randomly undergoing Brownian motion. At the receiver, the molecules may or may not bind to the receivers, in accordance with the type of molecule received and different stimuli [6]. The diffusion may cause noisy reception, especially in time-slotted MC systems [38]. For passive transport based MC, no external energy is consumed during the transmission process. The other kind of calcium signalling based communication is through direct contact, gap junctions, where nanomachines are located one next to each other. A gap junction acts as a gate of a cell membrane for specialized intercellular connection, connecting the cytoplasm of two cells directly and allowing various molecules and ions to pass freely between cells [39]. One cell can be in contact with several different cells, enabling different gap junctions open to exchange molecules. The permeability and selectivity of gap junctions are dependent on various internal and external factors, including cytosolic calcium ions, membrane potential, connexin phosphorylation, electromagnetic fields, temperature and pH in the environment [40]. The schemes of free diffusion and gap junction based MC are shown in Figure 2.3 [5].

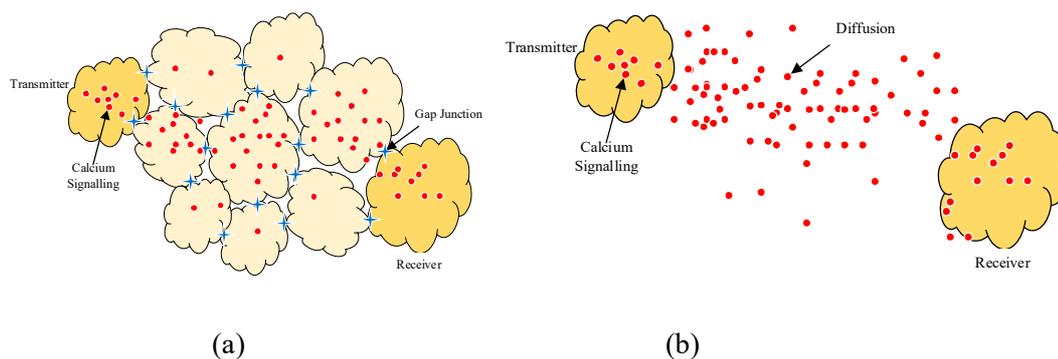


Figure 2.3 Calcium signalling MC based on (a) gap junction and (b) diffusion [5].

Besides the communication mode of calcium signalling, which uses the concentration of signalling molecules to represent the transmitted information,

diffusion of signalling molecules can involve biochemical reactions to achieve a different mode of communication, which uses the propagation of impulses to represent information [33]. The rapid increase and decrease of the signalling molecular concentrations would result in the generation of an impulse that propagates in the environment. For instance, neurons can produce ion impulses to propagate over the length of the neuron [33]. A nerve impulse refers to the transmission of a coded signal from a given stimulus along the membrane of the neuron. Two types of phenomenon are involved in processing the nerve impulse: electrical and chemical events, which propagate a signal within a neuron, and from one neuron to another or to a muscle cell, respectively [41]. The chemical interaction process between neurons and between neurons and target cells happen at the end of the axon, the structure of which is called a synapse. When a neuron is activated by a stimulus from the environment or from other nerve cells, it sends electrical impulses, which will rapidly propagate along its axon. Upon reaching the synapse, the impulse triggers secretion of a chemical signal, which will then be delivered specifically to receptors of the target cell. It should be noted here that the chemical impulse diffuses quickly in the synapse and reaches the target cell since the length of the synaptic gap is very small [36].

B. *Active transport based MC*

Active transport enables a communication system to transport signalling molecules directionally to specific destinations. Compared with passive transport, active transport can transmit information over longer distances (e.g., up to metres), and consumes chemical energy to generate the force for directionally transporting information molecules. In addition, active transport can achieve a high reliability even when the number of transmitted signalling molecules is small. For wireless MC,

an example of active transport is bacterial chemotaxis based MC [33]. Chemotaxis describes the ability of bacteria to direct their movements in reaction to particular sensed chemicals in their environment [42]. Bacteria swim by means of thin helical flagella on their surface, driven by molecular motor machines embedded in the bacterial membrane [43]. The motors can change the rotation direction, steering the bacteria according to the environment. For example, *E. Coli* have several flagella that rotate to produce a biased motion at several $\mu m \cdot s^{-1}$ [14]. In chemotaxis based MC, flagellated bacteria which are released by the transmitter, containing the information encoded in DNA molecules, are guided to swim towards the receiver, which continuously releases an attractant substance [44]. The characteristic of chemotaxis enables various bacterial based MC [45, 46]. For example, authors in [43] proposes a multi-hop routing mechanism which uses bacteria as carriers of information molecules (in this case DNA).

2.4 MC characteristics

MC exhibits some unique features that are not commonly found in conventional communication technology. The comparisons between two communication paradigms have been given in Table 2.1.

Table 2.1 Comparison of communication characteristics.

Type of communication	Traditional	Molecular
Devices	Electronic devices	Bio-nanomachines
Carrier	Electromagnetic field	Molecule
Signal type	Electronic or optical	Chemical
Range	Long distance ($\sim km$)	Short distance ($nm - m$)
Propagation speed	Speed of light	Extremely slow
Energy consumed	Electrical/high	Chemical/low
Receiver behaviour	Receiver interprets encoded data	Chemical reactions
Medium conditions	Air/cables	Aqueous

In addition to the features listed in Table 2.1, biocompatibility is another unique feature of MC, which is defined as the capacity of a device to operate in biological environments without affecting them negatively. Since MC uses communication mechanisms derived from biological systems, it can directly interact with cells, tissues and organs. Functions including receiving, interpreting, and releasing molecules would enable biological nanomachines to interface directly with natural processes. For example, a bio-nanomachine engineered from a biological cell (e.g., a red blood cell) can be safely injected into a human body, avoiding unintended reactions with existing components in the human body (e.g., immune cells) [14]. Also, to avoid procedures for clean-up of devices, bio-nanomachines may be programmed to be broken down after use [8].

Furthermore, functional complexity is a particular feature of MC, especially using cells as system components. A cell is a highly functional and integrated component with information processing capabilities [47]. There are a number of sensors (e.g., receptors to sense the environment), logic circuits (e.g., complex signal transduction pathways), memory for information storage, and actuators for motion generation in a

cell [48]. For example, in *E. Coli*, a cell stores a 4.6 million base pair chromosome in a $2\mu\text{m}^2$ cross-sectional area, which has the equivalent of a 9.2-Mbit memory that encodes a number of functional polypeptides [49]. The functional complexity property can help with designing high compacted engineered systems, such as NEMS/MEMS (nano- and micro-electromechanical systems) and labs-on-a-chip [50]. The information representation is also a key feature for MC. Transmitted information can be encoded using the physical or chemical properties of molecules, such as chemical structures (e.g., protein structure), sequence information (e.g., DNA sequence), relative positioning, and concentration that the receiver is able to react to. Thus, MC can provide different methods for manipulating and interaction with the information.

MC exhibits stochastic characteristics since thermal noise and other environmental factors may affect the system performance in the aqueous operating environment [48]. The stochastic features could be caused by factors such as the unpredictable movement of molecules in the environment, bio-nanomachines probabilistically reacting to information molecules, and information molecule degradation over time [14]. The probabilistic property inherently influences the design of MC systems. For example, in some situations, such probabilistic aspects can be overcome by relying on a large number of molecules for communication.

Moreover, MC may achieve a high degree of energy efficiency of chemical processes by reusing materials and processes from biological systems. MC systems may be energy efficient with low heat dissipation as cellular components are energy efficient [48]. For example, biomolecular motors, such as kinesin and myosin, can convert chemical energy into mechanical work by the hydrolysis of ATP, with high

energy efficiency [51]. Also, bio-nanomachines are able to harvest energy from the environment, requiring no external energy sources [14].

Self-assembly is also a significant property of MC systems [48]. Molecular affinities between elements of a nanomachine can enable the disordered elements to construct an organized structure, without external forces. For example, cells can divide and grow to assemble into a structure like organs. In addition, self-assembly MC systems exhibit a high degree of fault tolerance, since damaged parts of a system may be recovered through the division and growth of nearby cells [52].

Finally, robustness and fragility are also features that are required to be taken into consideration. Cells or cellular mechanisms exhibit some degree of robustness, which enables the system to maintain its functionalities against external and internal perturbations [53]. Meanwhile, cells can be extremely fragile to various factors such as temperature and pH changes that can destroy system behaviour easily.

2.5 Research aspects of MC

Compared to conventional radio based telecommunication systems, the MC research is still in its infancy. In this section, a brief review of some of the recent work on MC will be given, from a communication engineering perspective. Generally, the MC research can be categorized based on five partially overlapping topics:

- Modulation techniques: This describes the way in which the information is encoded onto the information molecules or chemical signals.
- Channel modelling and noise analysis: Due to the fact that the communication mechanism and noise sources are quite different from that of radio based communication channels, new channel models as well as noise models are required for MC.

- Coding techniques: Since the channel and noise sources are unique, new error correcting codes for MC systems are required.
- Network architectures and protocols: Since there are no standard protocols for MC, new protocols which satisfy the characteristics of MC are required.
- Simulation tools: Because laboratory experimentation for MC can be time consuming and expensive, various simulation tools have been developed for MC.

In the next few subsections, the main concepts and existing research of each category will be presented.

2.5.1 Modulation techniques

Modulation is the process of adding information to the carrier signal, varying one or more properties of the carrier signal. In conventional radio-based wireless communication systems, depending on whether the amplitude, frequency or phase of the carrier signal is modified, there are three basic modulation techniques, namely amplitude-shift keying (ASK), frequency-shift keying (FSK) and phase-shift keying (PSK) [54]. In essence, the transmission symbols are encoded using the variation of the carrier signal characteristics. In MC, in which the information carriers are small molecules, information can be encoded in the number of information molecules released. Also, it is possible to encode information in the type or structure of information molecules released. For example, one bit of information can be encoded by two different types (or sequences) of molecules [55]. Moreover, information can be encoded in the time of release of information molecules [56]. Another timing based approach is pulse position modulation (PPM), which has been proposed in [56]. In the PPM scenario, the information can only be encoded in the time of release of the molecule.

One of the first works that considered modulation in MC was [57]. In this study, two modulation schemes, including ON-OFF amplitude modulation and frequency modulation for the transmission rate of molecules, have been considered in a MC system. In ON-OFF amplitude modulation, the presence of bit ‘1’ or ‘0’ is encoded in the ‘presence’ or ‘absence’ of a particular kind of molecules, respectively. This modulation scheme is similar to the on-off keying (OOK) scheme in conventional communications. In the second scheme, the information is encoded via the changes in frequency value of the transmission rate, in sinusoidal transmission. In [58], novel modulation techniques including Concentration Shift Keying (CSK) and Molecule Shift Keying (MoSK) are considered for coding and decoding information of the so-called messenger molecule concentration waves in MC networks. Specifically, CSK modulates the information via the variation in the concentration of information molecules, while MoSK utilizes different types of messenger molecules to represent the information. In [59], besides CSK and MoSK, a novel modulation technique, named isomer-based ratio shift keying (IRSK) is presented, using isomers as information molecules for in-body nanonetworks via diffusion. It should be noted that the information-carrying isomers must be selected carefully such that they would not be harmful to the body. In [60], four pulse-based modulation techniques have been presented: Pulse Amplitude Modulation (PAM), PPM, Communication through Silence (CtS) and Rate Modulation. In PAM, at the start of each bit interval, the transmission of a pulse represents a binary ‘1’, and no pulse represents a binary ‘0’. In PPM, a symbol time is divided into two equal durations, with the first half representing a binary ‘1’ and the second half representing a binary ‘0’. CtS uses time as a new dimension to encode the information, specifically, the symbol is encoded into the time between two pulses. In Rate Modulation, the symbol time is fixed and

information is encoded into the rate of transmission of pulses. In [61], an ISI mitigation technique, titled Molecular Transition Shift Keying (MTSK), for diffusion-based MC channels has been proposed. The authors in [62] proposed another ISI-reducing modulation technique, named the Molecular ARray-based COmmunication (MARCO) scheme, encoding and decoding molecular information using the transmission order of different molecules.

2.5.2 Channel modelling and noise analysis

In MC, channel modelling and noise analysis are also key directions of research. Information theory is the mathematical foundation of any communication system [63]. Similar to any digital communication system, the design of a MC system needs to be based on a well-justified channel model including models for the signal propagation and corrupting noise sources. Physically, the channel is the medium over which the transmission signal propagates. During the transmission process, noise may occur, which will cause the received signal to be different from the transmitted signal. Traditional communication and information theory are based on a set of mathematically precise channel models, such as the additive white Gaussian noise channel. However, information theory based solid mathematical foundations for analysis of MC have not yet been built. In particular, no widely accepted general channel and noise model exists for MC. Instead, depending on the scenario, it is likely that several different channel models are required [14]. In MC, a major source of noise is the random propagation of information particles. As mentioned above, different propagation schemes exist for MC channels. Thus, for each propagation scheme and modulation technique, there exist different channel models.

Most of the current channel modelling research focuses on diffusion-based MC. Within this, the channel models can be divided into three categories: discrete models,

continuous diffusion-based models, and diffusion models with flow. In [64], the time slotted, OOK modulation based channel has been represented as a binary symmetric channel (BSC), where ligand receptors are applied for detection at the receiver. Extended works on broadcast channels, relay channels and multiple access channels are taken into consideration in [65, 66], where the impact factors, including the temperature of environment, concentration of emitted molecules, transmission distance and duration of molecule emission have been considered. In [67, 68], the communication channel, which has ISI between consecutive transmission symbols, has been modelled. In [69], a z-channel has been applied to represent the diffusion-based channel. Also, a simplified ISI model, where only the previous one time slot is taken into account, has been proposed. In [70], an additive noise channel is proposed, where the noise has a Poisson-Binomial probability mass function (PMF). In addition, for continuous diffusion-based models, the transmission signal is an analogue signal, and the received signal is convolved with the impulse response of the channel, plus the channel noise. In [71], a physical end-to-end model for continuous diffusion channels has been proposed; the transfer functions for transmission, diffusion and reception processes have been derived. Based on this work, further research has been presented in [72], where the most relevant diffusion-based noise sources, including particle sampling noise and particle counting noise, are considered. Similarly, the authors in [73] provide a mathematical study of the noise at the reception of molecular information when ligand-binding reception is employed. The reception noise is modelled by two different approaches, namely, through the ligand-receptor kinetics and through stochastic chemical kinetics. Finally, regarding the diffusion models with flow, authors in [74] consider MC where information is conveyed in the time of release of molecules, over a fluid medium,

propelled by a positive drift velocity and Brownian motion. The communication is represented by an additive inverse Gaussian noise (AIGN) channel model, and the channel noise follows an IG distribution.

Besides diffusion, active transport based MC is also a promising scheme. The work in [75] considers both schemes including diffusion and directional transport along protein filaments. In the latter scheme, kinesin motor proteins carry the information particles on the microtubule tracks. In [76], a mathematical model for microtubule filaments moving over a kinesin covered substrate has been presented. Moreover, the gap junction channel is also a propagation scheme that has been considered in multiple works. For example, the authors in [26] proposed a calcium ion relay channel, where the propagation of molecules depends on gap junction. Also, the information transfer capacity has been investigated and optimized by modifying channel properties. It should be stated that the information-theoretic capacity of MC, or the maximum rate at which data can be reliably transmitted, is also a significant research topic.

2.5.3 Coding techniques

In conventional communication systems, channel coding is applied to mitigate the effects of noise and fading that is introduced into the system. The basic idea of channel coding is that the transmitter encodes the message in a redundant way by using a proper error correction code. In this thesis, several channel codes will be applied to the proposed channel model. Thus the detailed descriptions of channel codes and the existing research of coding techniques in MC will be later given in Chapter 5.

2.5.4 Network architectures and protocols

For communication engineering, the approach to deal with the system complexity is to define the network architecture [77]. In computer networks, system functionality is divided into a stack of layers. Each layer implements a service for the layer above by using services provided by the layer below and adding the necessary functionality. The Open System Interconnection (OSI) model, which is a conceptual model that characterizes and standardizes the internal functions of a communication system, defines a networking framework by partitioning it into abstraction layers [78]. Also, the TCP/IP model divides the communication system into four layers: link layer, Internet layer, transport layer and application layer.

Because the development of MC is still in its early stage, most of the research has still been focusing on channel modelling, as well as the modulation and coding techniques for setting up a reliable communication link. However, once the communication link is established, the network architecture and protocols need to be considered. The architecture of MC may be discussed from a computer networks perspective, based on the conventional OSI or TCP/IP reference models. Establishing a layered architecture of MC may help design and develop functionalities and applications of MC. Here, the layered architecture proposed in [79] is considered, which has been shown in Figure 2.4. It indicates that the communication system can be divided into five layers, including the physical layer, the molecular link layer, the molecular network layer, the molecular transport layer and the application layer. Figure 2.4 illustrates the flow of information from a source to the destination.

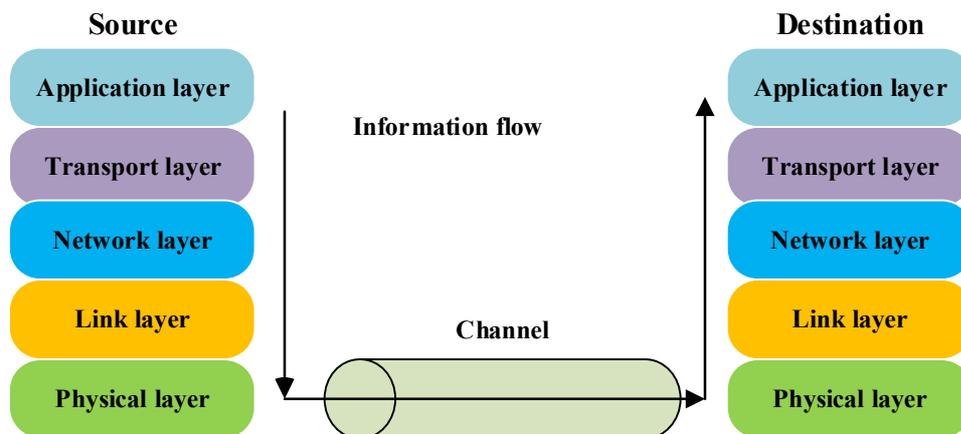


Figure 2.4 Flow of information through a MC network.

The physical layer provides the biophysical basis to address issues related to transmission, propagation, and reception of information molecules [14]. It is made up of two sublayers. Specifically, the bio-nanomachine sublayer abstracts physical details of bio-nanomachines and the signalling sublayer provides functionalities for signalling and modulating information onto molecules [79]. In particular, the physical layer determines the hardware and interfaces of bio-nanomachines. For example, the hardware includes the transmitter and receiver nanomachines which are capable of encoding and decoding a message onto and from information molecules. Also, the physical layer provides the modulation techniques for MC. Moreover, the mechanisms to propagate information molecules in the environment, such as calcium signalling, and the mechanisms for signal amplification and relay are provided by the physical layer. For instance, positive feedback loops can be used by biological systems to amplify the signal strength and to relay the chemical message over a long transmission distance [14].

The molecular link layer provides a set of mechanisms to enhance reliability for a group of nanomachines to communicate within a communication range [14]. In this

layer, information is represented by molecular frames. Framing is the functionality of creating molecular frames from physical layer signal molecules. Two kinds of framing method exist, including vesicle-based framing and DNA-based framing [79]. Also, the link layer provides mechanisms for error control, transmission rate control (i.e., flow control), medium sharing (i.e., media access control), addressing, synchronizing clocks in bio-nanomachines, and distance measurement among bio-nanomachines. Error control is an important function of link layer. In biological systems, a transmitter bio-nanomachine may transmit a redundant number of information molecules to increase the signal-to-noise ratio and reduce the impact of noise or fluctuation. Also, frame loss can be handled by detecting loss and correcting loss through retransmission [79]. In this thesis, a link layer retransmission protocol has been applied to the proposed channel model to enhance channel reliability in Chapter 6. In addition, the link layer provides addressing mechanisms to specify a receiver that receives molecular frames or a location to which molecular frames are delivered; this is another big issue in MC, because of its probabilistic nature. Moreover, the mechanisms for synchronization of clocks in bio-nanomachines are supported by link layer. Synthetic biology demonstrates that biochemical clocks can be introduced artificially into bacterial cells [80]. For example, in Chapter 6, biological clocks have been used by both the transmitter and receiver when the retransmission protocols are applied, to maintain flow of frames. Synchronization of clocks may be used when a group of bio-nanomachines perform some function at the same time or to avoid interference through time-slotted communication [81].

The network layer provides functionalities and mechanisms for communication over larger distances than at the molecular link layer. In this layer, information is represented by molecular packets. The network layer provides mechanisms which

include network formation, network routing, network processing and congestion control [14]. In MC, most of the current research is based on the physical layer and link layer. Thus the network layer and upper layers will not be discussed in detail here. Instead, Figure 2.5 gives some typical functionalities of each layer. Readers can refer to [14] for further understanding. The existing research works considering communication protocols in each layer will be given in Chapter 6.

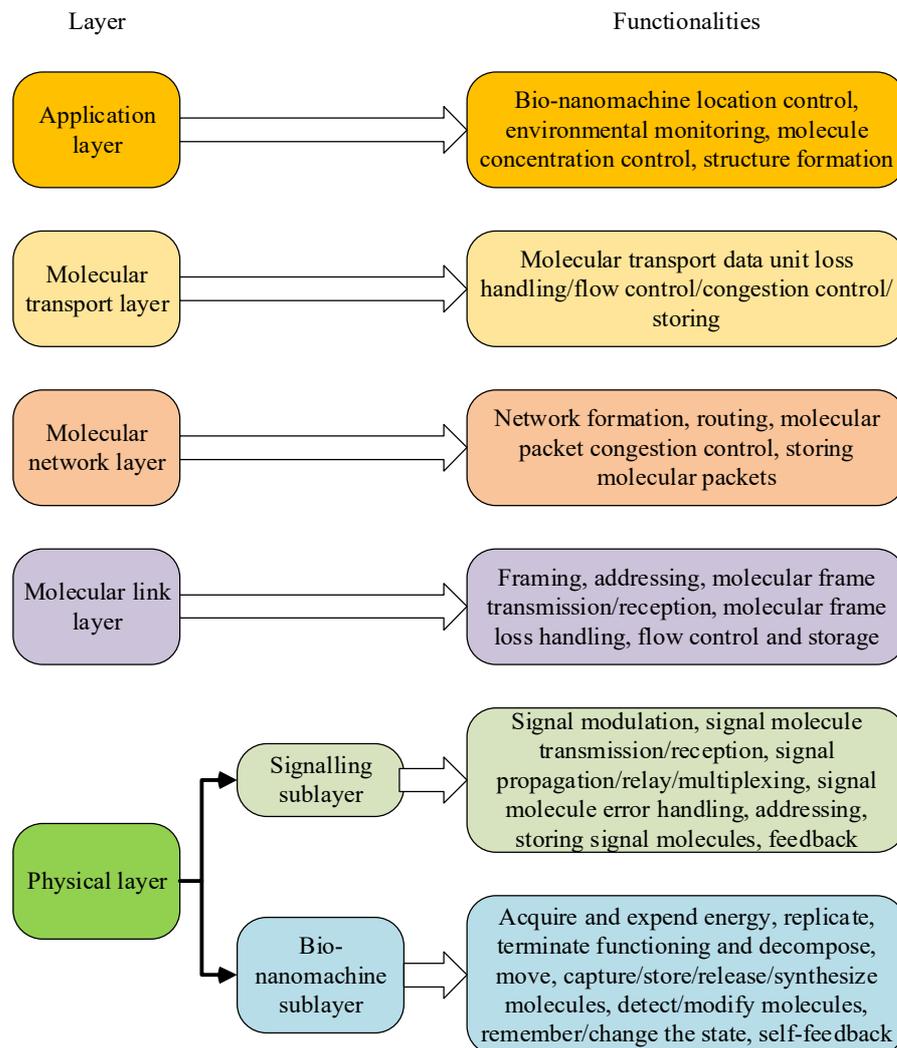


Figure 2.5 A set of functions provided by each layer [79].

2.5.5 Simulation tools

For MC, laboratory experiments are significantly more expensive and difficult to perform than simulations. At the current stage, simulations are a key tool for

determining the performance of MC systems. Some simulation tools, including dMCS [82], N3Sim [83], MUCIN [84], NanoNS [85], BINS [86], and BNSim [87], have been developed for MC. The development languages of dMCS, N3Sim, BINS and BNSim are JAVA, and the language for MUCIN is MATLAB. Also, simulation tools including dMCS, MUCIN, NanoNS, BINS and BNSim can be used for 3-D environments. In this thesis, all the simulation work is based on MATLAB.

2.6 Applications of MC

MC is a completely new paradigm and would potentially enable many new applications in various fields. Examples of future applications that MC enables are described below.

2.6.1 Biomedical applications

Biomedical applications are the most promising and major application group that would benefit from MC. For example, with the help of nanotechnology, it is possible to interact with cell components including cells, tissues and organs, to manipulate the cell proliferation and differentiation, and to implement the production and organization of extracellular matrices [88]. The principal applications in the biomedical field would be in developing enhanced immune systems, bio-hybrid implant systems, targeted drug delivery systems, health monitoring systems, and tissue engineering systems [5, 89].

Targeted drug delivery may be performed by encapsulating drug molecules in drug delivery carriers, sending the carriers to the target site (e.g., disease cells or tumours), and releasing the drug molecules from the carriers at the target site [79]. These nanoscale capsules can bind to specific receptors of the targeted cells. Therefore, targeted drug delivery can reduce the potential adverse effects of drug molecules on

non-targeted organs [90]. Also, MC may provide alternative techniques to improve the targeting accuracy and the therapeutic effect through the coordinate of bio-nanomachines. For example, when a nanomachine can detect some special molecules released from a sick cell, it can share information with the nanomachines nearby, which would become alerted and can behave cooperatively against the sick cell.

Tissue engineering, which aims at assembling functional constructs that restore, maintain, or improve damaged tissues or whole organs [91], is also a promising application of MC. Tissue engineering refers to the process of extracting the stem cells from patients, culturing them in vitro, and eventually returning the cultured stem cells to the lost tissue part of the patient. MC can provide additional mechanisms to help control the growth and differentiation of the autologous cells into specific structures [27]. The products of tissue engineering, such as bio-artificial skins with viable and nonviable cells and autologous cultured chondrocytes, have been put into the market [92].

The research object of enhanced immune systems is to enhance the immune responses of the human body [93]. Introducing artificial immune systems into the human body may be helpful to protect from infectious agents such as parasites, virus, bacteria and fungi [94]. In enhanced immune systems, MC may enable the communication and coordination of a group of bio-nanomachines, for example, the tracking of moving targets (e.g., pathogens) and the notification to external devices of the target location for further treatment [95].

2.6.2 Industrial applications

MC has promising applications in industrial fields, such as the monitoring and control of microbial formations. For example, bacterial biofilms have been used to clean residual waters coming from different manufacturing processes or to treat

organic waste in [96, 97], using MC networks. Also, MC can be used to improve the quality of food, including the production, processing, safety and packaging. For example, a number of studies have indicated that QS, which is a promising method of MC, plays a major role in food spoilage, biofilm formation, and food-related pathogenesis [98]. Besides, the agricultural industry can benefit from food materials which contain multiple bio-nanomachines through which the growth process can be controlled [14]. In manufacturing industry, molecular manufacturing can solve a number of current problems, such as water shortage [99]. Also, advanced molecular manufacturing technology will enable the fabrication of clean, efficient, and low cost complex products, for example, materials much stronger than steel, and superior military systems [99].

2.6.3 Environmental applications

MC may also be applied in monitoring an environment that may be contaminated with toxic or radioactive agents. Specifically, the integration of bio-nanomachines can be performed into large or small scale devices, which are deployed in the environment to form a large scale biosensor network [14]. Identification of the location of foreign contamination or toxins can be achieved through the utilization of bio-nanomachines to monitor molecules in the environment, which can be helpful to identify and amplify molecular signals with the cooperation of a group of nanomachines, which will be helpful to remove toxins or contaminations [100, 101].

2.6.4 Information technology applications

The area of information technology may be advanced by integrating bio-nanomachines into currently available silicon-based electrical systems using MC. For example, future mobile phones may be equipped with a biochip, which may be possible to diagnose diseases or stress by directly analysing biomolecules, such as a

drop of sweat, tear, saliva and blood [102]. The molecular transport system would be packaged in the biochip and the acquired results would be transmitted to a medical specialist via traditional cellular networks. A dermal screen, which is made from as many as 3 billion bio-nanomachines and embedded below the skin surface of a human body, is another example of the application of MC in advancing information technology [103]. In addition, MC may apply to non-silicon-based computing paradigms, which has promising features including extremely high functional complexity and large scale parallelism that cannot be achieved with conventional silicon-based electronic circuits [14].

2.7 Summary

MC, which is considered as one of the most promising communication paradigm at nanoscale, has developed rapidly in recent years. The transmitted information is encoded in chemical molecules, transmitted via specific channels and exchanged between bio-nanomachines [10]. Bio-nanomachines, which are made of biological materials, can perform various tasks, for example, moving in the environment, sensing a specific type of molecules in the environment, or catalysing specific biochemical reactions. A typical MC system is made up of an information source, a transmitter, a propagation channel, a receiver and information sinks. Generally, MC can be divided into wired and wireless communications. One example of the former is based on microtubule motor proteins whereas in the latter, the molecules can be transported through calcium signalling or bacterial chemotaxis. The main features of MC include the use of molecules as communication carriers, the limited communication range, the slow-speed propagation, the low energy consumption, the chemical reactions at the receiver and the fluid transmission medium. Other features of MC include its stochastic nature of communication, high compatible with

biological systems, functional complexity, self-assembly capability and so on. These capabilities would potentially enable many applications in various fields, including biomedical, information technology, industrial and environmental applications. Until recently, most of the research about MC has focused on modulation techniques, channel modelling and noise analysis, coding techniques, transmission protocols and simulation tools. In this thesis, coding techniques and transmission protocols will be taken into consideration for the proposed communication channel in the following chapters.

CHAPTER 3.

BACTERIAL COMMUNICATIONS

3.1 Introduction

It has been stated in the previous chapter that MC enables communication between nanomachines, increasing their functionalities and developing new possible applications. Bacterial communication, as a promising approach of MC due to some of the biological characteristics, including the ability to swim and migrate between locations, to carry DNA contents that could be utilized for information storage, and to interact and transfer plasmids to other bacteria, will be described in detail in this chapter. Bacteria could be used to establish nano communication networks, termed as bacterial nano-networks, that operate in microfluidic devices, body area networks (BAN) or other circumstances [43]. Current research for bacterial nano-networks have not considered the internal structures of the nanomachines that can facilitate the use of bacteria as transceivers or information carriers. There are mainly two different communication mechanisms for different bacterial species: molecular-based communication and plasmid-based communication. In the former, bacteria emit specific molecules, which can diffuse randomly through the cell membrane into the local environment, and then eventually be picked up by the bacterial population that are in close vicinity. The communication process is carried out using essential entities including signalling bacteria cells, target cells, signalling molecules, and receiver proteins. The signalling cells have the responsibility of diffusing the

molecules to target cells, where the information is encoded onto the signalling molecules. The receiver proteins are in charge of decoding the transmitted information. In the latter, the bacteria are able to transfer plasmids, which are small DNA molecules within bacterial cells that are physically separated from chromosomal DNA and can replicate independently, between each other. The plasmid can be transferred among cells through conjugation, nanotubes, and bacteriophages. Bacteria are capable of forming a physical connection using the pilus, which is a type of tubular protein that stems from membranes. This process, called conjugation, facilitates the transfer of plasmids between bacterial cells by direct cell-to-cell contact. Nanotubes are also applied by bacteria to establish physical connections on a solid surface, transferring not only plasmids, but also including ions and proteins. Additionally, bacteriophages, which is a type of virus that infects and replicates within the bacteria, also provide an approach for transferring genetic information [104]. Experiments show that *E. coli* emit the phages with encapsulated plasmids, where the phages could diffuse into the medium and be received by the surrounding bacterial population [104].

In the thesis, the focus is on the molecular-based bacterial communication. Instead of language, bacteria communicate with each other through a process called QS, using chemical molecules as sensing signals, called autoinducers, which are released into the immediate environment [12]. 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) [12]. Thus, the concentration of external autoinducers is correlated with the bacterial cell population density. 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 [105]. The discovery of this phenomenon changed the general perception of many individual, simple organisms existing in the environment. 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, attacking suitable hosts, and rapid adaptation to environmental changes, based on the local density of the bacterial population [106]. 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 [107]. 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 each class of autoinducers in a different way [108]. In addition, it has also been proved that interspecies communication via QS exists, which is referred to as cross talk [108]. Nowadays, it is well acknowledged that most bacteria strains communicate using secreted signalling molecules to regulate their group behaviours. Also, many different classes of molecules can be used as signals, and individual species of bacteria can synthesise, detect and respond to multiple classes of chemical signals.

In this chapter, the essential properties and capabilities of bacteria which are necessary for bacterial cells acting as transceivers are discussed in Section 3.2. Also, it provides an overview of the different types of QS mechanisms of various species of bacteria in Section 3.3. The applications and possible challenges that bacterial communications can bring to reality are also presented in Section 3.4. Finally Section 3.5 gives a brief summary of this chapter.

3.2 Capabilities of bacteria

For bacterial communication networks, bacteria can be utilized as nanomachines and also information carriers. Bacteria provide a large amount of attractive properties that make them ideal and appropriate for MC. Although previous studies have investigated the utilization of bacteria to enable MC, a majority of these research has not considered how nanomachines for bacterial nano-networks could be constructed and designed to support the communication process. In this section, the bacterial structure is first explained, followed by the description of some essential properties, capabilities and functionalities, which are required for individual bacterium or population of bacteria acting as nanomachines.

Generally, a bacterial cell has eight essential structural components (shown in Figure 3.1), including the cell wall, cytoplasmic membrane, chromosome, plasmid, ribosome, flagella, inclusion body and pili. The interface between the bacterium cell and its external environment is the cell surface, which contains the cell membrane and cell wall plus an outer membrane (if one is present), protecting the cell interior from external hazards and maintaining the cell integrity as a discrete bio-entity. The interface enables the transport of large molecules, including carbohydrates, vitamins, amino acids, nucleosides as well as proteins, into and out of the cell. The cell membrane is selectively permeable, which means that only a select group of molecules can pass through freely; the movement involves both active and passive processes. Passive transport describes the movement of biochemical or molecular substances across cell membranes, driven by the growth of the entropy of the system, without energy consumption. In contrast, in active transport, the movement of molecules across a cell membrane is in the opposite direction of concentration gradient, using cellular energy. The chemical composition of bacteria includes

protein (55%), total RNA (20.5%), DNA (3.1%), phospholipid (9.1%), lipopolysaccharide (3.4%), murein (2.5%) and inorganic ions (1.0%) in general. The detailed functions and chemical compositions of each structural component are given in Table 3.1 [109]. In 1884, the Danish investigator Christian Gram conducted an experiment, namely the Gram staining method, leading to the discovery of two fundamentally different types of bacterial cells, specifically gram-negative and gram-positive bacteria [110]. Generally, the surface of gram-negative bacterial cells is much more complicated than that of gram-positive cells. The cytoplasmic membrane of bacterial cells is a lipid bilayer composed of phospholipids, glycolipids, and a variety of proteins, which are able to provide either structural support to the membrane, or the transport of sugars and metabolic intermediates. For gram-positive and gram-negative bacteria, the transport systems for transiting materials across the cell membrane are different. Specifically, the cytoplasmic membrane of gram-positive bacteria has direct access to ingredients of medium. While for gram-negative bacterial cells, there exist pores formed by protein triplets in their outer membrane which enable the transit of macromolecules into the periplasmic space.

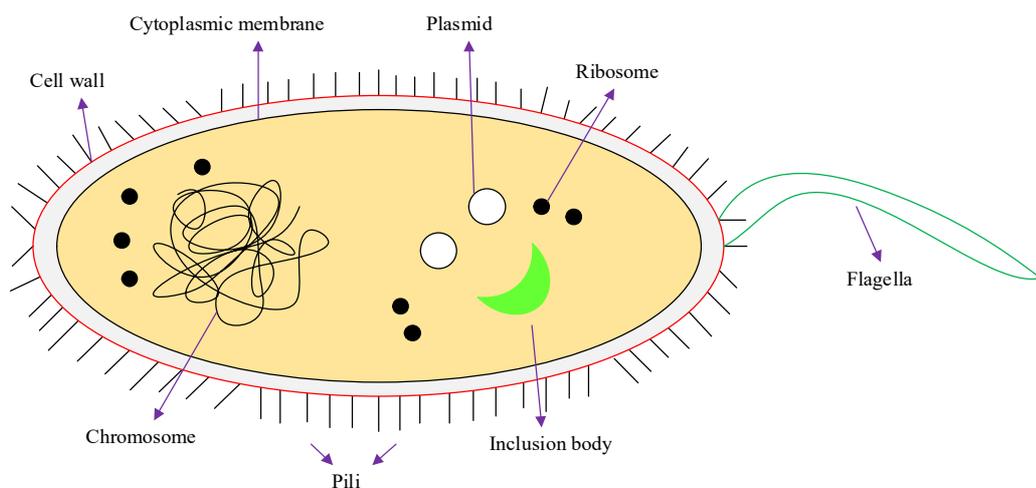


Figure 3.1 The structural components of a typical bacterial cell.

Table 3.1 Summary of characteristics of typical bacterial cell structures.

Cell structures	Relative functions	Chemical composition
Cell wall	Protects cells against osmotic shock (most important) and physical damage	peptidoglycan
Cytoplasmic membrane	Regulation of substance transport into and out of cells	Phospholipid and protein
Chromosome	Contains genome	DNA
Plasmid	Contains supplemental genetic information such as resistance to antibiotics, production of toxins and tolerance to toxic environment	DNA
Ribosome	Takes part in protein synthesis	RNA and protein
Flagella	Movement of cells	Protein
Inclusion body	Mineral storage of cells	Highly variable; carbohydrate, lipid, protein or inorganic
Pili	Attachment to host, bacterial mating	Protein

Generally, in bacterial communications, biological nanomachines are molecular scale complexes, such as multiple protein sub-units, bacterial cells and vesicles, which could act as energy transduction devices and molecular storage devices that respond to specific biological stimuli to perform cooperative behaviours. These bacterial nanomachines are able to display diverse functionalities, including cell motility, chromosome and plasmid segregation, cytokinesis, DNA replication, energy generation, and protein synthesis and secretion [111]. In addition, even the simplest bacterial nanomachine is formed by highly coordinated action of numerous proteins assembled into multiproteins. Their molecular architectures, nanoscale conformational changes, binding and unbinding forces, as well as the dynamic stoichiometry are all significant characteristics that require investigation. Until recently, the development of nuclear magnetic resonance (NMR) spectroscopy, x-ray crystallography, mass spectrometry and cryo-electron microscopy has provided

plenty of biochemical, structural and functional characterization of bacterial nanomachines [111].

Through the process of QS, bacteria can simultaneously and cooperatively respond to environmental changes, such as temperature, pH, concentrations of ions, nutrients and enemies, enabling reproduction or migration to a new favourable location. From the experimental results to date, it has been observed that the bacterial colony performs collective sensing, distributed information processing, division of labour, and regulation of gene expressions of the individual bacterium in the population [104]. In addition, bacteria are able to collaborate to maintain the bacterial population and to ensure the maximum profit for the colony. For example, in undernutrition conditions, bacteria in the colony would cooperate to slow down the reproduction process in order to decrease the consumption of nutrition. Under other circumstances, bacteria in the colony are able to synthesise and secrete specific bacteriocins that are toxic to the strains of the same bacterial species (e.g., colicines secreted by *E. coli*), inhibiting the growth rate of other bacteria [112]. Besides, each bacterium is a biotic autonomous system with its own internal cellular gel that possesses informatics capabilities, which means that bacteria are also able to store information, perform decision making and learn from past experiences collectively (e.g. exposure to antibiotics) when solving newly encountered problems, generating an erasable, collective inheritable memory [104]. A population of bacteria is also able to perform distributed information processing, where each bacterium in the colony is capable of storing, processing, and interpreting information [112]. Through biochemical communication, such as QS, the bacteria in the colony can sense environmental changes and perform internal information processing as well as coordinate the information. The information processing can occur either among

individual bacteria as well as between the colony and the environment or by intracellular communication which obtains and analyses the information extracted from the environment. Such capabilities greatly improve the bacterial cooperation level during the communication process.

In addition, the individual bacterial cell is a highly specialized energy transformer. Chemical energy generated by substrate oxidations is conserved by the synthesis of high-energy compounds such as adenosine diphosphate (ADP) and ATP [113]. ATP, which is a universal energy currency used in many biological processes that require energy, is one of the most significant small molecules in bacteria. In the presence of proper enzyme systems, these high-energy compounds can be used as energy sources to synthesize the new complex organic compounds needed by the cell, thus carrying out cellular functions that are necessary for their survival, growth and replication.

3.3 Quorum sensing

3.3.1 QS: an overview

QS is a mechanism that allows bacteria to “talk” to each other. In general, a quorum is defined as the minimal number of members of a committee required to validate a decision. This means that many bacterial behaviours controlled by QS, including symbiosis, virulence, antibiotic production, and biofilm formation [114], are induced or repressed only when a critical cell population is reached. The benefit of QS is twofold: it is a collaborative communication process that enables bacteria to coordinate behaviours (e.g., formation of biofilms) that would be inefficient for an individual bacterium and it also helps the bacteria to sense the population density. In this section, a similarity-difference analysis of some well-characterized QS systems for particular bacterial species are represented, in terms of the type of autoinducers, stimulant, receptors, signal transduction mechanisms and QS system output.

Bacteria communicate with each other through the release, detection and response to the accumulation of specific chemical autoinducers in QS systems. The term *autoinduction* was introduced to describe the process by which bacteria sense signals that were released by themselves [105]. With the increase of the bacterial population towards the stationary phase, autoinducers accumulate in the culture medium and, above a certain threshold level, regulate the gene expressions involved in diverse functions. This QS process enables a population of bacteria to control the gene expression of the entire community, cooperatively and synchronously. The detailed procedure of intercellular communication in bacteria was first described in the bioluminescent marine bacterium *V. fischeri* [115], which is frequently found in symbiotic relationships with marine animals like the bobtail squid [116]. It shows that two regulatory components, LuxI and LuxR, are essential for QS process, which will be presented in detail in subsection 3.3.2. Beyond regulation of gene expression on an intracellular level, QS allows bacteria to communicate within and between species. A deeper understanding of QS systems will offer a broader perspective on the complicated interaction mechanisms that exist widely among bacteria.

Basically, a cell-cell bacterial signalling system could be classified into two main categories, in accordance with the type of signalling molecules and the detection mechanisms:

- Gram-negative bacteria: LuxI/LuxR type QS system which uses *N*-acyl homoserine lactones (AHL) as signalling molecules in *V. fischeri*-like bacteria [117], or two component signalling circuits which can identify three different autoinducers, including AHLs, furanosyl borate diester (FBD) and an uncharacterized *V. cholera* autoinducer (CAI-1) molecule in *V. harveyi*-like bacteria [118, 119].

- Gram-positive bacteria: two-component QS systems which use modified oligopeptides as signalling molecules [120].

3.3.2 QS in gram-negative bacteria

In microbiology, the visualization of bacteria at the microscopic level is facilitated by the use of stains, which react with components present in some cells but not others. This Gram staining technique is a rapid diagnostic tool to group bacterial species as either a gram-positive or a gram-negative bacteria, depending on their colour following a specific staining procedure [110]. The reason bacteria are either gram-positive or gram-negative is due to the structure of their cell surface. Gram-negative bacteria are defined as the types of bacteria which do not retain the crystal violet dye in the Gram staining method, appearing red or pink due to the effects of the counterstain.

A. *LuxI/LuxR-type QS system*

The vast majority of the well-studied gram-negative QS systems utilize AHLs as signalling molecules. AHLs have been isolated and characterized from different bacteria species, all of which share a common homoserine lactone ring moiety, while important distinctions that give specificity are found in the acyl side chains with variations in length, in the substitution at the β position and in the degree of saturation of the acyl chain bonds [121]. AHLs are soluble in the liquid condition, and are able to cross phospholipids layers of cell membranes, which enables them to freely diffuse through the environment. These two properties are necessary for signalling molecules. AHL-producing bacteria have been identified in more than 37 genera within the *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria* [122-124] and the *Cyanobacteria* [125], in which AHL-based QS mechanisms are proved to be crucial to regulate gene expressions and bacterial behaviours. More

than 50 LuxI/LuxR pairs have been experimentally identified and studied in more than 70 different species in the phylum *Proteobacteria* [126]. Therefore, LuxI/LuxR-type model acts as a prototype of the QS control in gram-negative bacteria. Using this QS mechanism, gram-negative bacteria can efficiently couple gene expression to fluctuations in cell-population density.

The most intensely studied LuxI/LuxR-type QS system in gram-negative bacteria is that of the bioluminescent marine bacterium *V. fischeri*, which exists naturally either in a free-living planktonic state or as a symbiotic of certain luminescent fish or squid [127, 128], and is found in a specialised light organ where it can reproduce to quite a high cell density (10^{11} cells/ml) [129]. As a typical example, *Euprymna scolopes* (the Hawaiian bobtail squid) and *V. fischeri*, which are shown in Figure 3.2, have an obligate mutualism with one another. The squid provides the *V. fischeri* bacteria with abundant nutrition and allows proliferation in numbers unachievable in seawater, while the *V. fischeri* provide the squid with luminescence which provides its camouflage and protection. The regulatory circuit controlling QS in *V. fischeri* was discovered in [130, 131], which showed that bioluminescence in this bacterium is controlled by the QS system. This is composed of two regulatory genes, luxI and luxR, coding for proteins LuxI and LuxR, respectively. The signal N-(3-oxo-hexanoyl)-L-homoserine lactone (3-oxo-C6-HSL or AHL) is synthesised by the protein LuxI, which is absolutely required and sufficient for AHL synthesis with no additional cofactors and energy source needed. LuxR is responsible for binding the signalling molecule and activating the transcription of the bioluminescence operon, which is made up of the luxCDABE genes, also called the *lux* box, at high cell density. Thus the *lux* operon, which is a 20bp (base pairs) conserved inverted repeat region located in the promoter region of luxR genes, is linked to the luxI gene

encoding for the AHL synthase. LuxAB genes encode a heterodimeric enzyme *luciferase* which catalyses the light emission and luxCDE encode proteins required for biosynthesis of the aldehyde substrate [132]. When the concentration of the freely diffusible signal molecules reaches a threshold or a micro-molar range, the LuxR protein binds the signal molecule and interacts with the luxCDABE genes, activating the transcription of them. The process of bioluminescence is illustrated in Figure 3.3. It shows that at low bacterial cell densities, the LuxI protein secretes AHL at a low basal level. With the increase of the cell population, the autoinducer concentration increases as a function of increasing bacterial population density. When the autoinducer concentration reaches a certain threshold, it can interact with the LuxR protein, and the AHL-activated LuxR dimerises and binds the luciferase promoter to activate transcription of luminescence genes, inducing therefore luminescence production. Importantly, the LuxR-AHL complex also induces the expression of luxI because it is also encoded in the *luciferase* operon, which floods the environment with the signal. This phenomenon creates a positive feedback loop for AHL production and the spatial distribution of cells, causing the entire population to switch into QS mode and generate light. In other words, this QS system enables bioluminescence to be tightly correlated with the bacterial population. Also, bioluminescence is only observed at high cell density when a high concentration of autoinducers are presented. In addition, *V. fischeri* emits light only inside the specialized light organ of the symbiotic host which is free-living in the ocean [114]. This is because that bacteria can grow to high density population only when there is abundant nutrition in the light organ, and that trapping the diffusible autoinducers enables them to accumulate to a featured threshold cellular concentration which is sufficient for LuxR protein to bind with.

For many bacteria species, more than one set of LuxI/LuxR QS systems act synergistically, and are often connected and form hierarchical signalling cascades. Each QS system synthesises various AHL molecules, providing individual differences and peculiarity to each system. For example, the opportunistic human bacterial pathogen *Pseudomonas aeruginosa* owns two LuxI/LuxR QS systems, one is LasI/LasR type and the other one is RhII/RhIR type, both of which are arranged in a hierarchical gene regulatory networks. Protein LasI synthesis signal *N*-(3-oxododecanoyl)-L-homoserine lactone (3-oxo-C12-HSL), and protein RhIR catalyse the formation of signal *N*-butanoyl-L-homoserine lactone (C4-HSL). The two QS systems can be activated sequentially and in the right order [133]. The richness of fluorescence, including small organic fluorescent dyes, nanocrystals (quantum dots), autofluorescent proteins, and small genetic encoded tags complexed with fluorochromes, and advances in labelling or marking methods, such as immunolabelling and genetic tagging techniques, enable the resolution and the simultaneous observations of the distributions of multiple specific signalling molecules in bacterial communication systems [111, 134]. In addition, some bacteria contain multiple LuxR homologous series controlling the expressions of different genes, but responding to the same autoinducer [133].

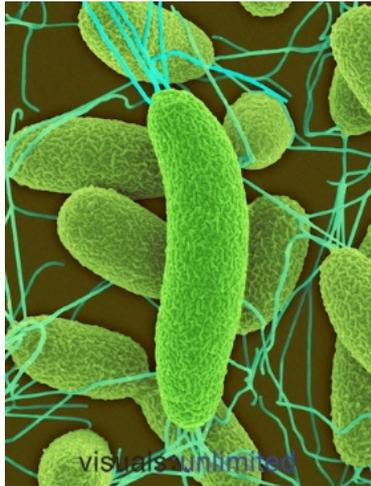


Figure 3.2 The symbiotic relationship between bacteria and the host.

The symbiotic relationship between (a) *V. fischeri* (300 × 248 pixels, photographed by Dr. Dennis Kunkel) [135] and (b) *Euprymna scolopes* (302 × 254 pixels, photographed by Prof. Margaret McFall-Ngai) [136], a small Hawaiian squid, and *V. fischeri* provides an example of specific cooperatively during the development and growth of both organisms.

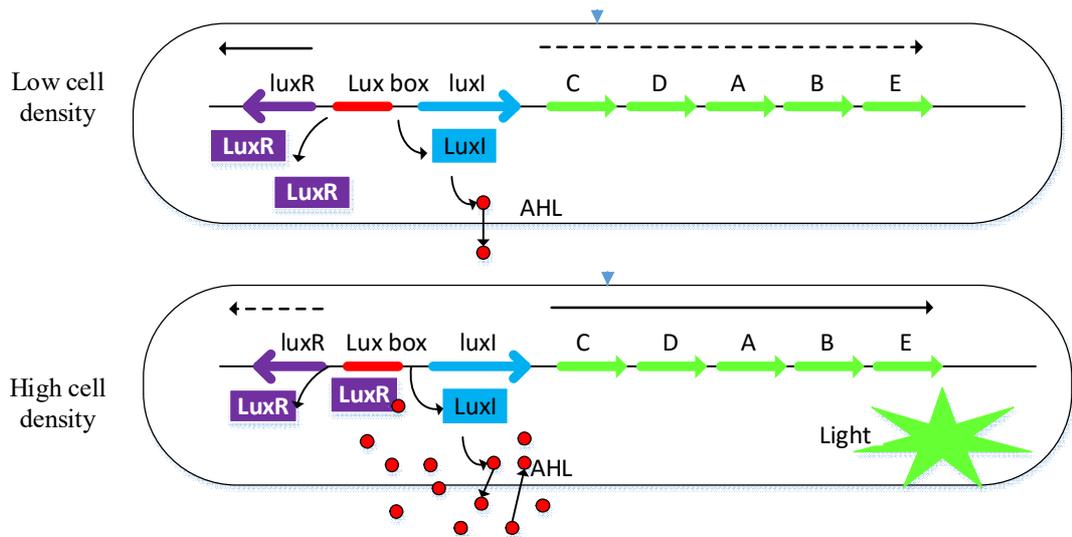


Figure 3.3 The LuxI/LuxR QS system.

(a) The system is not active and there is basal transcription of *luxR*, *luxI* and *luxCDABE*. The concentration of the autoinducer increases only when the bacterial population increases in cell number. With the bacterial population increasing, the autoinducers accumulate until the concentration exceeds a threshold which allows the binding between AHL and LuxR. (b) The LuxR-AHL complex binds the *lux* box and induces both luminescence and LuxI. This process forms a positive feedback autoinduction circuit, inducing the full expression of the system in an exponential way.

B. *V. harveyi*-type QS system

In Gram-negative bacteria species, besides LuxI/LuxR-type QS system, there is *V. harveyi*-type QS system, in which three separate QS systems have been identified and organised in a hierarchical regulatory circuit [137]. *V. harveyi* is also a bioluminescence marine bacteria, which can exist free swimming in seawater, adhered to abiotic surfaces, as a constituent of biofilm consortia in marine animals, and in pathogenic associations with marine hosts [137]. The QS system consists of three autoinducers and three cognate receptors functioning in parallel to channel information into a shared regulatory pathway. *V. harveyi* produces the system 1

autoinducer termed AI-1, *N*-(3-hydroxybutanoyl) homoserine lactone (HSL), which is synthesised by LuxM, which shares no homology to the LuxI family, although the biosynthetic pathway is probably identical [118]. Then it binds to its ligand receptor, a membrane bound sensor histidine kinase (LuxN). The system 2 autoinducer is a furanosyl borate diester termed AI-2, whose synthesis requires the LuxS enzyme. AI-2 and the gene required for its synthesis have recently been proved to be present in a variety of gram-negative and gram-positive bacteria; thus it can be the link between different bacterial species [138]. AI-2 is bound in the periplasm by the protein LuxP, which is similar to periplasmic ribose binding proteins, and the LuxP-AI-2 complex interacts with another membrane bound sensor histidine kinase, LuxQ. The system 3 autoinducer, termed CAI-1, is (S)-3-hydroxytridecan-4-one and is produced by the CqsA enzyme. Again, this signal interacts with a membrane bound sensor histidine kinase, CqsS.

At low cell densities, when the autoinducer levels are not high enough to be detected by their cognate sensors, LuxN, LuxQ and CqsS act as protein kinases and autophosphorylate their respective response regulator domains. Phosphate groups are then transferred to a shared phosphotransferase protein LuxU, which in turn transmits the signal to the transcriptional activator LuxO. The phosphorylated LuxO, together with σ^{54} (a protein required for the initiation of RNA synthesis), repress the transcriptional activator LuxR by activating the expression of at least five small regulatory RNAs (sRNAs). Repression occurs via incomplete base-pairing in coordination with the chaperone Hfq, which is a protein essential for binding sRNA. At high cell densities, LuxN, LuxQ and CqsS act as protein phosphatases upon binding of their cognate autoinducers. The phosphatase activity of the three sensors dephosphorylates LuxO, relieving the repression of LuxR expression, which can then

inhibits the transcription of the downstream target genes involved in diverse functions including bioluminescence and the production of exopolysaccharides [139].

The mechanism of QS in *V. harveyi* is displayed in Figure 3.4.

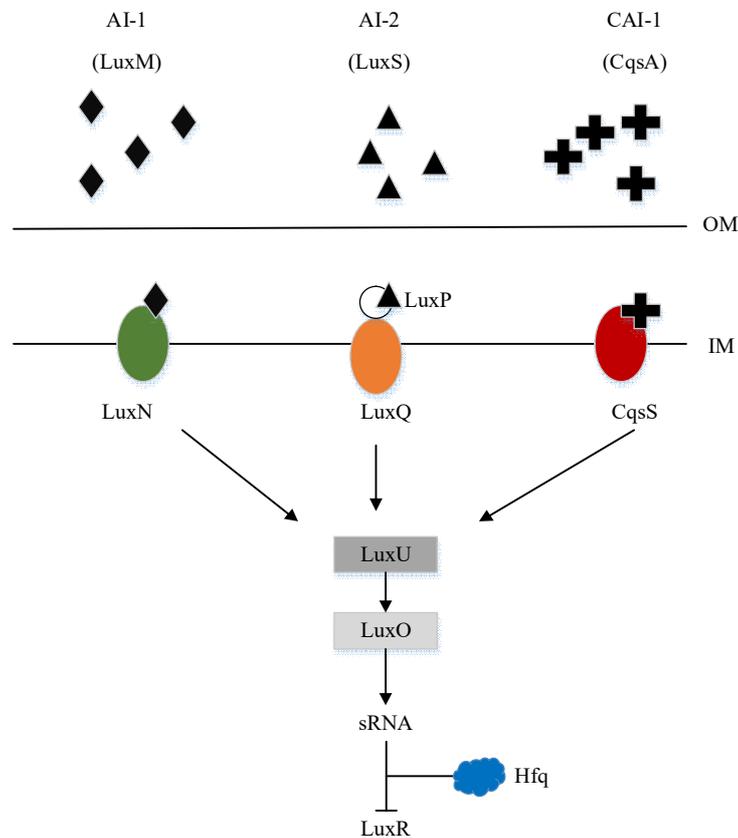


Figure 3.4 QS system in *V. harveyi*.

There are three parallel QS systems, including AI-1/LuxM, AI-2/ LuxS, and CAI-1/CqsA, that converge to a common transcriptional activator via a phosphotransferase mechanism. At low cell densities, no autoinducers are present and detected. At high cell densities, LuxN, LuxQ and CqsS act as protein phosphatases upon binding of the signalling molecules, AI-1, AI-2 and CAI-1, respectively. The phosphatase activity of LuxN, LuxQ and CqsS dephosphorylates LuxO, inactivating it and relieving the repressing LuxR expression. All the three signal transfer pathways are achieved by a series of hierarchical phosphorylation and dephosphorylation by the same components [139].

In this thesis, the work below will be mainly based on the QS system in *V. fischeri*. Subsequent research has identified that *V. fischeri* possesses three QS signalling systems organised in a hierarchical circuit, two *V. harveyi*-type systems, AinS-AinR and LuxS-LuxPQ, and one LuxI/LuxR-type system given above. The detailed description will be given in Chapter 4.

3.3.3 QS in gram-positive bacteria

Gram-positive bacteria are defined as the group of bacteria that retain the colour of the crystal violet stain in the Gram staining method, appearing to be purple through a microscope [110]. Such bacteria regulate different kinds of phenotypes in response to increasing cell population density. Different bacterial species employ QS to regulate different phenotypes. For example, *Bacillus subtilis* and *Streptococcus pneumoniae* use QS to develop bacterial competence. *Staphylococcus aureus* employs QS to regulate virulence while *Enterococcus faecalis* use it to regulate conjugation [133]. The signalling molecules used in gram-positive bacteria QS systems are known to be different from those of gram-negative organisms [140]. Instead of using AHLs as the signalling molecule, gram-positive QS systems make use of peptides, which can interact with the sensor element of a histidine kinase two-component signal transduction system [140]. These extracellular peptides, often called pheromones, are generally synthesized first as precursors inside the bacterial cell and are then processed and secreted via a dedicated ATP-binding cassette (ABC) transporter [138]. Instead of using LuxR protein to detect autoinducers in gram-negative bacteria QS system, gram-positive bacteria use two-component adaptive response proteins for detection [138]. A general model for peptides mediated QS system in gram-positive bacteria is shown in Figure 3.5, where a peptide signal precursor locus is translated into a precursor protein (red and green diamonds) that is

cleaved (arrows) to produce the processed peptide autoinducer signal (red diamonds). The peptide signal is then processed and transported out of the cell membrane via an ABC transporter (grey protein complex). When the concentration of the peptide signal surrounding the cell exceeds a certain threshold, a histidine sensor kinase protein of a two-component signalling system detects it. The binding of the peptide signal to the sensor domain triggers the transmission of the signal via auto-phosphorylation or dephosphorylation of the sensor histidine kinase domain on a conserved histidine residue (H). The phosphate group is then transferred to the aspartate domain of the cognate response regulator, which is phosphorylated on a conserved aspartate residue (D), activating it and allowing it to regulate transcription of various targeted genes [138].

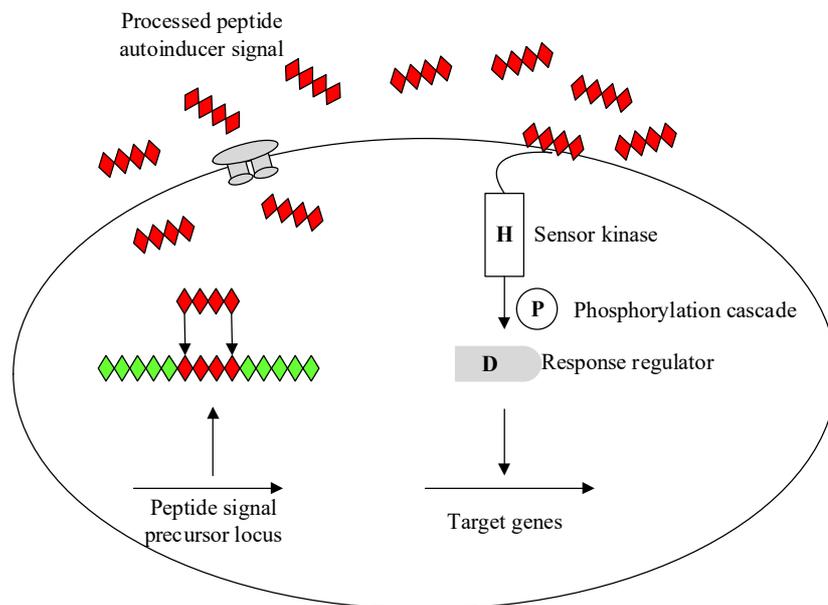


Figure 3.5 A general model for peptide-mediated QS in gram-positive bacteria.

3.4 QS applications

The development of synthetic biology enables the formation of many functional artificial genetic circuits. These can be applied in innovative gene regulation systems, gene therapy, drug development, cellular computation and bioremediation [141], using the components of bacterial QS systems. QS is considered as a predominant method of cell-cell communication, which is the basis for numerous communication modules utilized in the field of synthetic biology. In previous work in the synthetic biological field, devices such as switches, oscillators, timers, and logic gates were mostly designed and constructed to operate at the single cellular level [142]. Until recently, the research emphasis has been concentrated on the creation of cell-cell communication modules to coordinate population-level behaviours in synthetic biological systems [143, 144]. The reason of changing the research topic from unicellular to multicellular programming is that with the extensive ranging importance of cell-cell communication in multiple types of natural biological systems, synthetic biologists could use engineered, communication-based systems to explore basic communication equations. Also, programming cellular behaviour on a multicellular level will be more robust and efficient from an engineering perspective [142]. Considering these factors, the use of multicellular communication may be critical in generating robust bacterial communication systems which are capable of biocomputing and bioprocessing. In this section, the applications of engineered QS systems in various areas including industry, environment, and biomedical field, which are shown in Figure 3.6, are detailed represented.

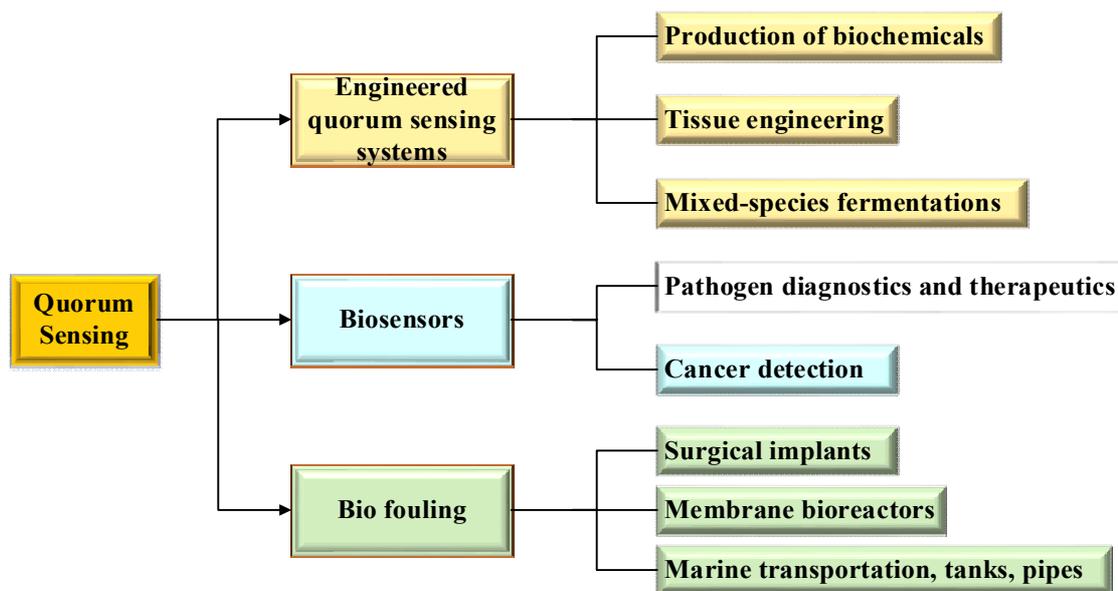


Figure 3.6 Overview of current QS applications.

3.4.1 Engineered QS system

The emerging field of synthetic biology enables researchers to establish and optimize complicated genetically engineered devices, such as genetic circuits which can be applied to modify cellular functions and create cellular responses to environmental conditions, using biological parts inside a cell [145]. Several genetic modules can be integrated into complex gene networks, which can be implanted into a stable host cell to supply the necessary materials and functional support, through a process called plug-and-play [146]. In engineered QS systems, cells synthesize autoinducers by themselves, which enables the cells to control their cell density and modulate their behaviours accordingly, thus decreasing the need for external management [147]. Also, the foundation of the BioBricks database [148] provides freedom in the design of synthetic QS systems. The authors in [146] achieve the incorporation of QS components in complicated genetic circuits, combining various input, processor and output modules to construct novel functional systems, which is expected to be

significant in the industrial production of toxic gene products and in designing environmental biosensors [149].

3.4.2 Biosensors

QS can be applied in whole cell microbial biosensors engineering to identify pathogenic microbes present in the surroundings, mount a concerted response against them and infected host organisms [149]. It has been stated in [150] that QS signals alone can be used as markers for the presence of pathogenic bacteria in clinical and environmental samples. Another exciting application of QS is to construct engineered bacteria which are capable of detecting and invading cancer cells. For example, the signalling molecule 3-oxo-C12-HSL inhibits cancer cell proliferation and promotes cell death [151]. Also, QS can be applied in the synthesis of genetically modified plant which are able to protect themselves against bacterial pathogens [149].

3.4.3 Biofouling control

Biofouling, or biological fouling is the undesirable accumulation of bacteria, plants, algae or animals on wetted surfaces such as pipes, tanks, membrane bioreactors and medical/dental implants [152]. It leads to, *inter alia*, contamination, colonization and corrosion of equipment components exposed to water and decreasing machinery efficiency [149]. Incorporation of QS inhibitors on the equipment surface is a possible way to reduce *P. aeruginosa* biofouling of surgical implants, which suggests that QS inhibition may be employed to give defence against many pathogens which start biofilm formation depending on QS [153]. Also, QS inhibition may be very valuable in improving the membrane bioreactors efficiency and preventing marine biofouling during marine transportation [149].

Considering Chapter 2 and Chapter 3, in particular, with specific reference to the use of genetically engineered bacteria [10], and their interconnection in MC networks, some application examples are given as follows. Genetically engineered bacteria can be deployed in the human gastrointestinal tract, serving as sensors of particular biomarkers generated by inflammations or ulcers [154]. When an inflammation is detected, the bacteria colony would communicate with each other, perform a consensus decision-making process and cooperate to generate drug molecules, resulting in a localized and timely treatment. Also, MC based nanonetworks of genetically engineered bacteria could be used to achieve high-throughput microfluidic devices for chemical analysis by enabling on-chip information exchange, where the MC between bacteria could enhance procedures for chemical analysis applications such as enzymatic assays, drug discovery assays and protein crystallization techniques [155]. Also, the dynamic molecular exchanges within these nanonetworks could possibly be used to combine multi-stage reactions on a single microfluidic device, which will lead to not only higher throughput and accuracy, but also the automation of chemical analysis, instead of using the intervention of an external human operator [156]. Another application is the environmental monitoring which has been displayed in Chapter 2, so is not further illustrated here.

3.5 Summary

In this chapter, it is clear that the QS mechanism is a pattern that bacteria extensively use to control fundamental functions involved in every important stage of their different life cycles, including survival and host colonization. It belongs to a set of signalling systems that interact in response to diversified autoinducers to integrate a variety of inputs into an appropriate cell response. It allows a population of

individuals to coordinate global behaviour and thus act as a multicellular unit. Also, this process is fundamental to microbiology and synthetic biology, including industrial and clinical microbiology. Recently, it has become apparent that QS promotes both intraspecies and interspecies bacterial communications, enabling populations of bacteria to sense their cell numbers and regulate their corresponding gene expressions. The processes controlled by QS are diverse and reflect the specific needs of particular species inhabiting unique niches. In general, gram-negative bacteria use AHLs as autoinducers, and gram-positive bacteria use processed oligopeptides to enable cell-cell communication.

Furthermore, the essential capabilities of bacterial nanomachines which are vital for the establishment of bacterial communication networks are presented in the chapter. A bacteria based hybrid model can be constructed where individual bacterial cells or bacterial colonies are represented as transceivers, endowed with individual decision making and information storing capabilities. These capabilities undertake the commitments to provide satisfactory and reasonable explanations for the bacterial threat to human health. For example, an increasing number of bacterial strains of pathogenic bacteria can today resist multiple antibiotic drugs. Bacteria are clearly capable of developing antimicrobial resistance at a higher rate than the development and invention of new drugs, and humans seem to be losing the crucial battle for health. One promising efficient solutions is to perform gene-expression research during the colonial reproduction stage, when bacteria are exposed to antibiotic stress and most challenging during bacterial learning from experience. It is expected that future experiments could change the prevailing incorrect attitudes to bacteria, for example, simple creatures that are merely solitary, of little threat, and possessing limited capabilities.

In addition, the chapter has discussed recent studies of synthetic biological systems that utilize QS communication systems and the corresponding applications. Previous research concerning bacterial communication has focused mainly on the development of signalling molecules, which have the potential for use as antimicrobial drugs aimed at bacteria that utilize QS to control virulence. Also, the biosynthetic enzymes which are related to autoinducer synthesis and detection are considered as potential targets for novel antibacterial drug design. Furthermore, research into QS could be used to develop industrial production of products like antibiotics, and to give us insights into novel intra- and intercellular transmission mechanisms and intra- and interspecies communication. Although research into QS is only in its initial stage, sufficient insight has been gained to offer fundamental insights into how bacteria establish communication and information networks. There are still nevertheless many subjects in need of further research such as how bacteria employ multiple types of signals, the nature of the signal biosynthesis pathways and the mechanism by which the information encoded in these chemical signals is processed and transduced to regulate gene expression. The challenges of this field will include the biological, such as the unintended consequence of host-circuit interactions [157], the influence of stochastic gene expressions [158], and the limitations of metabolic capacity of bacterial cells [159]. In addition, there will be substantial technical challenges, including the real-time monitoring of genetic circuit components [160], the development of high-throughput functional analysis to inspect novel biological devices [161], and the limits of high-fidelity DNA synthesis technologies [162].

CHAPTER 4.

COMMUNICATION MODEL

4.1 Introduction

In the previous chapters, nanotechnology, MC, and more in-depth biological background of bacterial communications, especially QS, have been presented. Starting from this chapter, the objective is to focus on the mathematical modelling of QS system, thus enabling mathematical physical analysis from the natural biological phenomenon. This is to make the study of bacterial communications more approachable from an engineering perspective.

QS is a particular communication process, which allows bacteria to activate genetic programs cooperatively, providing an instructive and tractable example illuminating the causal relationships between the molecular organization of gene networks and the complex phenotypes they control. A purely experimental approach to investigate such networks becomes increasingly difficult due to the additional components revealed. The following methodology is to model the bacterial communication from the traditional transmitter and receiver perspective, which are also connected conversely. The proposed model is made up of two populations of bacteria, the *transmitter* node and the *receiver* node, respectively, and the communication channel, using the well-studied bacteria species, *V. fischeri*, which is most famous for its previously discussed 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'. This approach has been chosen mainly because it provides a useful tool to enable the understanding of the biological processes that occur in nature and its dynamics. 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 QS networks [163]. For the bacterial communication system proposed in this work, the effect of ISI caused by the stochastic nature of molecular diffusion is considered as a major source of impairment.

The rest of the chapter is organized as follows. First, a brief review of the existing modelling approaches of QS is presented in Section 4.2, followed by the description of two QS systems in *V. fischeri* in Section 4.3, which will be used in the following research. After that, the communication model, including the communication environment, the transmitter and receiver model, and the channel model, is developed in Section 4.4 and 4.5, describing the steps followed to achieve the natural communication system. Then the channel capacity analysis is given in Section 4.6, followed by the summary in Section 4.7.

4.2 Existing models of QS

QS was first modelled mathematically in gram-negative bacteria [164-166], and a large proportion of mathematical models, also referred to as stochastic models, still focus on the bioluminescence system and its homologues which are quite common in this kind of bacteria. The dynamic behaviour of QS networks can be analysed with a wide range of theoretical approaches, including Boolean networks, Kauffman networks, dynamic Bayesian networks, Petri Nets, as well as differential equation

models [167-169]. The majority of the models are based on the use of differential equations that describe the chemical kinetics of the reactions among bacteria, ranging from subcellular approaches, wherein the relevant gene regulation network is captured explicitly, to the cellular and population levels where the focus is on the proportion of cells in each state of QS induced regulation and on the concentration of the autoinducers. Specifically, the researchers introduce parameters for each mRNA, protein and enzyme concentration, and indicate the parametric variation in terms of protein production and degradation. This series of reactions result in autoinducer production in accordance with the intracellular and extracellular concentration, leading to bioluminescence. Thus, the properties of the system depend upon both the parameter choice and the particular scenario which is being modelled. For instance, research in [170, 171] illustrates how the variable values and the number of feedback loops within the QS system could influence the number of feasible steady states. Besides, there exist some other scenarios in which QS is modelled, for instance, the use of the signalling system in a burn wound [172] or in a biofilm [173]; both studies focus on the bacterium *P. aeruginosa*. In addition, even for the same type of bacterial species, either gram-negative bacteria or gram-positive bacteria, the enzyme reaction chain that induces gene expression varies among them. Thus, different species are supposed to have unique parts in their communication scheme apart from the universal structure. Hence, specific models for specific bacteria have also been established, for example, the case of *P. aeruginosa* [173], and the *agrobacterium* family [174]. Unlike the relatively well developed QS models for gram-negative bacteria, the mathematical models specific to gram-positive bacteria are significantly fewer in number. Some examples of QS modelling in gram-positive bacteria include the studies of the linear peptide-based QS system in *B. subtilis* and

its influence on the initiation of sporulation [175], rather than focusing on gene regulation. However, the drawback of most of the existing mathematical models is the lack of connection between bacteria, their enzyme catalysed reactions and the environment. For example, variables, including the colony population, the interaction between bacteria, and the bacterial multiplication, are essential to achieve QS but few of them have been taken into consideration.

Besides mathematical models, computational models have also been developed, simplifying the chemical reactions within the bacteria, and being more focusing on establishing a macroscopic framework in which bacteria are basic elements in the environment. For example, the authors in [21] model QS bacteria as automata, which could be used as the control unit of a nanomachine.

A wide range of software is now available for the numerical solution of the types of differential equations mentioned above, some of which could be used with more general purposes, such as MATLAB, NAG routines or XPPAUT. In addition, there are also some software packages specifically designed for biological applications, such as Berkeley Madonna, CellDesigner, COPASI or the Systems Biology Toolbox for MATLAB. However, many biological aspects regarding QS are still relatively unknown and thus requires detailed study to fully understand the phenomenon. Therefore, the creation of a 3D simulator, which could take all these detailed aspects into consideration, would be an attractive approach to help obtaining more reliable results. In this thesis, MATLAB is utilized to model and analyse the system performance.

4.3 Two QS systems in *V. fischeri* bacteria

It has been stated in previous chapters that the typical gram-negative bacterium *V. fischeri* is utilized to establish the QS model. Thus it is essential to understand the QS systems for *V. fischeri*.

V. fischeri possesses two major QS systems, *ain* and *lux*, using AHLs as signalling molecules. QS based on the *lux* system, as the principal process directly regulating luminescence production, has been demonstrated in Section 3.3.2. The protein LuxI synthesizes the autoinducer 3-oxo-C6-HSL, denoted as Type-I autoinducer, which binds and activates the transcription factor LuxR at a threshold concentration (100 ~ 200nM). The LuxR/Type-I complex binds as a dimer upstream of the *lux* operon and utilizes RNA polymerase to initiate the transcription of luminescence genes *luxCDABE*. The *V. fischeri lux* system is essential for both luminescence production in vivo and persistence of the bacterial symbiont in the host light organ. Besides the *lux* system, *V. fischeri* has a second QS system based on *AinS*. This synthesizes N-octanoyl-homoserine lactone (C8-HSL), denoted as a Type-II autoinducer, which is an autoinducer detected by the histidine kinase *AinR*, a hybrid sensor kinase. In addition, the *ain* locus exhibits no obvious sequence similarity to *luxI*, suggesting that it arose independently [176]. The Type-II autoinducer regulates a number of activities through the transcriptional regulator *LitR*, such as *rpoS* expression and normal motility [177]. Also, it is required for normal colonization of the squid and is necessary for a normal growth yield. As a matter of fact, there exist a third autoinducer, LuxS dependent signal AI-2, which was proved to be involved in the regulation of both luminescence production and colonization competence [178]. However, compared to that of the *AinS-AinR* and *LuxI-LuxR* QS, the impact of AI-

2 was quantitatively small [178]. Thus in this thesis, only Type-I and Type-II autoinducers will be taken into consideration.

4.4 Bacterial communication scheme

4.4.1 Bacteria functionality model

Because of the high degree of randomness and limited capabilities of a particular bacterium, communication between two individual bacteria can be unreliable. In addition, the delay in the communication process can be fairly large due to biological actions such as transcription and translation. Also, the effects of gene expression noise, mutation, cell death, undefined and changing extracellular environments, and interactions with cellular context currently make it a challenge to engineer single cells. Most applications or tasks in present synthetic biological systems are generally completed by a population of cells, not any single cell [179]. Hence, to achieve reliability of the communication system, here, the communication model between two populations of bacteria which is proposed in [180] is taken into consideration. In this model, a cluster of bacteria trapped in a chamber is considered as a node. These clusters of biological nodes, which are able to transport information from one point to another, are considered to be the basic building blocks of the communication system. In short, an individual bacterium is very primitive and unreliable and hence incapable of transferring information by itself. Instead, a cluster of bacteria has been applied, which are collectively capable for reliable transmission and reception of molecular information, to form a biological node [180].

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, generally by means of readily isolating plasmids from bacterial cells and altering in

vitro by inserting or deleting specific sequences of DNA [181] [182]. Molecular communication between two bio-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 [180].

In this proposed model, both the transmitter and receiver nodes contain m instances of the bacterium, *V. fischeri*, which was introduced in Chapter 3 and is the most commonly studied QS system in gram-negative bacteria. These bacteria are motile, gram-negative rods, 0.8-1.3 μm in diameter and 1.8-2.4 μm in length [183]. As a marine bacterium, *V. fischeri* exists at low cell densities when free living and at high cell densities when colonising the light organ. The luminescence is governed by the expression of certain genes, called the *lux* operon, in the cell, which in turn is controlled by the density of cells in a population. The regulation of the luminescence genes, named *luxCDABE*, depends on the production and detection of the signal (Type-I autoinducer), which is synthesised by the protein LuxI and sensed by the protein LuxR. Figure 4.1 illustrates the process with the structure of a bacterium used in such a node shown in Figure 4.1 (a). The features of the production of Green fluorescent protein (GFP), which is the protein responsible for detection or

production of molecules, is stored in the plasmid which can be added to a bacterium who does not naturally emit GFP through genetic engineering techniques. The receiver node senses the surrounding concentration of Type-I autoinducers, which will trigger the production of GFP and the process is shown in Figure 4.1 (b).

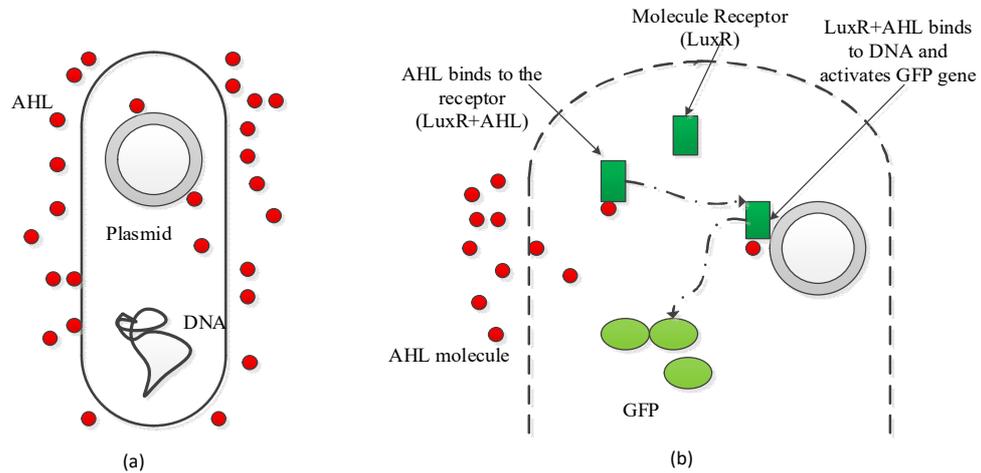


Figure 4.1 Bacterial communication: (a) bacterium structure; (b) GFP production.

In *V. fischeri*, bioluminescence 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 AHL, which is synthesized 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 [180].

Figure 4.2 shows the schematic for the communication between two populations of bacteria. The specifics of the nodes will be discussed in the next section. 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 [10]. It is assumed that each bacterium can sense and produce two different types of AHL molecules, specifically Type-I (3-oxo-C6-HSL) and Type-II (C8-HSL) autoinducers mentioned above [184]. 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. The transmitted information which will be transmitted at the transmitter is encoded into the concentration of signals, denoted by A_0 . The bacteria inside the transmitter node can produce various concentrations of Type-I molecules to be transmitted through the channel by the stimulation of different levels of concentration of specific stimulants. The emitted signalling molecules would then diffuse through the channel to the receiver which is at the distance d from the transmitter. At the receiver, each bacterium senses the concentration of Type-I molecules through the corresponding Type-I receptors (LuxR), followed by the production of GFP by bacteria, which is used to decode the input signal concentration A_0 . However, it should be noted that 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-II receptors (AinR) are activated and the gene expression of AinS is repressed, while for bacteria in the receiver node, only Type-I receptors (LuxR) are enabled and the gene expression of luxI is repressed, a process which can be controlled by proper enzymes. The corresponding reasons will be discussed later.

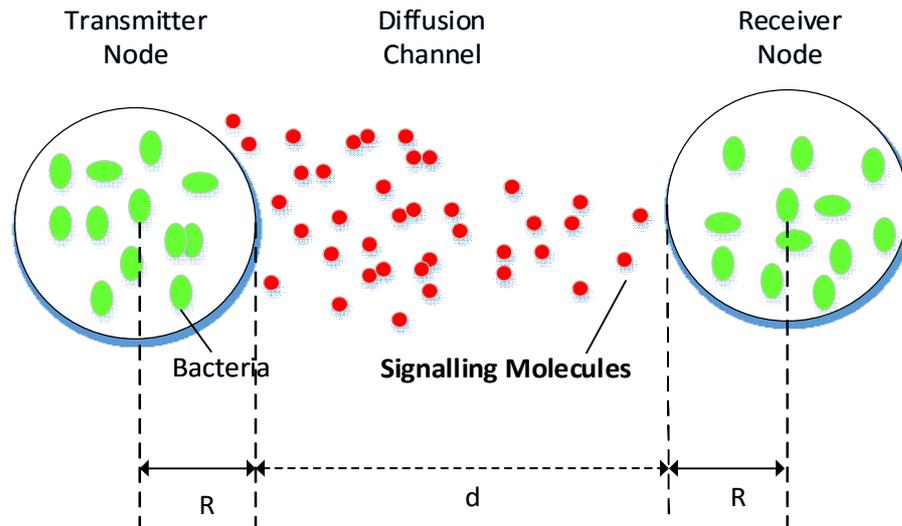


Figure 4.2 Bacterial communication setup consisting of the transmitter node, the diffusion channel and the receiver node.

4.4.2 Environmental illustration

QS relies on the generation, release and detection of autoinducers, which will propagate through the environment which the bacteria are colonising. It is essential to understand the existing circumstance so as to be capable of implementing its properties in the following research.

It has been stated in previous chapters that different bacterial species possesses various QS systems, which means that the living circumstances of these species should be also diverse. However, in order to simplify the modelling and simulation processes for the research, some hypotheses are going to be made with respect to the sceneries.

For *V. fischeri*, the bacterial colony may reach populations of approximately 10^5 individual bacteria [177, 178]. Also, the size of the bacterium is of the order of a few micrometres, which has been stated in section 4.4.1. Therefore, the required space for the culture medium of bacteria involved in QS system would not be larger than a

few millimetres. Generally, the term *small culture medium* refers to cultures not exceeding a diameter of 2mm , while large cultures are supposed to those larger than 5mm in diameter. Therefore, it is obvious that the volume of the culture medium in which the bacteria are involved in QS are small and thus considered to be finite. In other words, it should be reasonable to consider a cluster of bacteria trapped in a chamber as a node, which has been displayed in Figure 4.2. The chamber could be a chemostat, in which a bacterial population could be maintained at a constant density; this situation has some similarity with bacterial growth in natural environments [109]. Besides, to ensure that the calculation of the diffusion of particles is done correctly, the medium in which the autoinducers diffuse is considered to be infinite, which means it is considered to be much larger than both the size of the bacterial nodes and the distance between the transmitter and receiver. The concept and mechanism of diffusion will be described in the next section.

Moreover, the communication environment is considered to be homogeneous, which means that its properties are identical throughout all its extension. Generally, in real environments, the environmental properties will vary both spatially and temporally, by default. Thus, if the environment is considered to be heterogeneous, which means that its characters vary as a function of space and time, numerous variables are required for the modelling and simulation process to describe the heterogeneity, which would bring a significant challenge to this research. Moreover, it has been stated above that the size of the chambers colonising the bacteria populations are small, and since transmission distance is set at the micrometre level (which will be discussed later), the variations of the environmental properties are limited. As a result, the bacterial communication system is assumed to be in a homogeneous environment, with identical properties throughout all its extension.

In addition, in terms of the diffusion medium of autoinducers, the properties, especially the diffusion coefficient of signalling molecules, is set based on its composition and the size of the entities involved. Due to the fact that the autoinducers are nanoscale molecules in size, together with the viscosity of the medium, the signalling molecules propagating in the medium will be mainly affected by viscous forces, instead of inertial forces. This mechanism may be explained using the Reynolds number, which describes a ratio between inertial and viscous forces [185]. The nutrients can be added to the chamber from a reservoir of sterile medium. After certain initial adjustments, the overflow rate may be regulated and the inflow rate of nutrients would then be capable of regulating and maintaining a constant population in the growth chamber [109]. In this work, the fluids in which the autoinducers propagate is assumed to be a low Reynolds number environment, specifically water, in which the diffusion coefficient of autoinducers has been well studied; the value will be given later.

As a conclusion, this section defines the environment of both bacterial culture medium and autoinducers diffusion medium. The autoinducers released by the bacterial nodes will move in a homogeneous low Reynolds number, infinite space, following a process called molecular diffusion.

4.4.3 Transmitter and receiver model

For both the transmitter and receiver nodes, the radius R is related to the number of bacteria m in the node. As discussed in section 4.4.1, each bacterium must be equipped in general with two distinct receptor types (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 receptors is activated. For the bacteria in the transmitter node, only Type-II receptors are

enabled, while for bacteria in the receiver node, only Type-I receptors are enabled. It is assumed that for each type of molecules, there are enough ligand receptors, which should be much larger than the number of molecules released at the transmitter node. Both the transmitter and receiver nodes have a probabilistic nature. For the transmitter and receiver nodes, the process of reception is governed by a set of equations, but with possibly different coefficients [180]. The response of different kinds of bacteria to different levels of concentration of signalling molecules has been well studied in the literature. In this work, the model in [186] is adopted, which uses a series of linear differential equations to describe the average dynamic behaviour of bacteria and also their steady state, and to explain the process of luminescence in response to AHL concentration difference. These equations take the three main phases in the luminescence process as shown in Figure 4.1 (b) into account:

- Binding of signalling molecules to the LuxR receptors.
- Generation of the AHL/LuxR complex and transcription of the luminescence genes which are responsible for the production of GFP.
- GFP production.

According to [186], the probability p_{bind} of binding the signalling molecule AHL to the ligand receptor LuxR is given by:

$$\frac{dp_{bind}(t)}{dt} = -\kappa p_{bind}(t) + A\gamma(1 - p_{bind}(t)) \quad (4.1)$$

where A is the concentration of signalling molecules surrounding the bacterium, γ is the input gain and κ is the dissociation rate of captured molecules in the cell receptors; $\frac{dp_{bind}(t)}{dt}$ is the derivative of p_{bind} with respect to time. Solving for binding probability that:

$$p_{bind}(t) = \frac{A\gamma}{A\gamma + \kappa} [1 - e^{-(A\gamma + \kappa)t}] \quad (4.2)$$

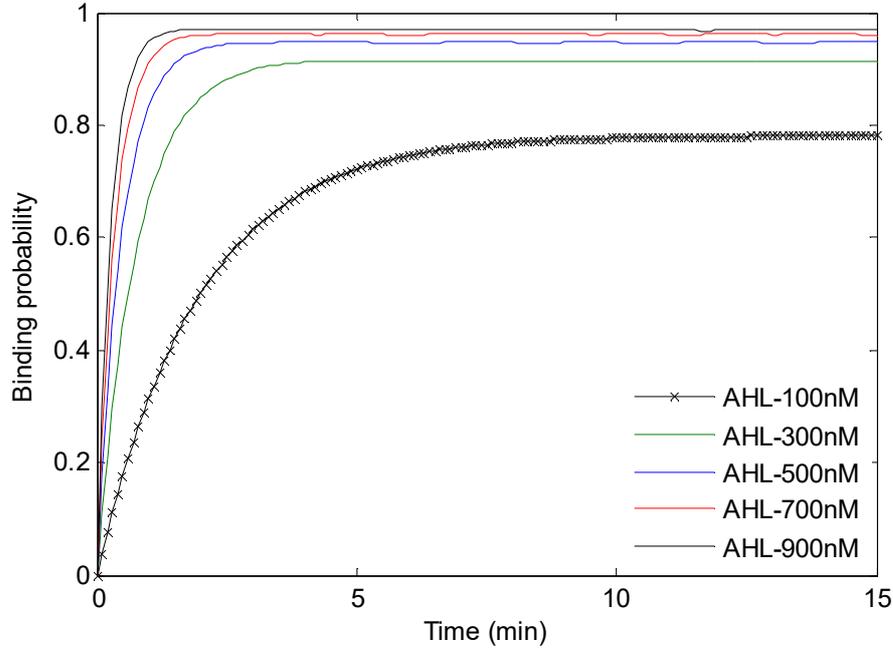


Figure 4.3 AHL/LuxR binding probability for different concentration of signalling molecules surrounding the bacterium. (*nM* denotes nanomole)

Figure 4.3 shows that each cell receptor is activated by capturing one signalling molecule with probability p_{bind} , which depends on the molecular concentration around it. Also, it can be inferred that the steady state value is denoted by:

$$p_{bind}^* = \frac{A\gamma}{A\gamma + \kappa} \quad (4.3)$$

This shows that with a higher concentration of molecules, the capture probability is higher, moreover as one would expect; it approaches unity for very high concentrations.

In addition, it can be concluded that for a given AHL concentration, the time until the binding probability $p_{bind}(t)$ reaches a certain threshold p_0 is given by:

$$t^* = \frac{1}{A\gamma + \kappa} \log\left(\frac{A\gamma}{(1 - p_0)\gamma A - \kappa p_0}\right) \quad (4.4)$$

The production of the LuxR/Type-I autoinducer complex, the transcription of genes and the production of luminescence are described by the equations [186]:

$$\begin{cases} \frac{dS_1(t)}{dt} = (bp_{bind}(t) + a) - b_1S_1(t) \\ \frac{dS_2(t)}{dt} = S_1(t) - b_2S_2(t) \end{cases} \quad (4.5)$$

where S_1 is a post-transcription messenger. a_i and b_i are some constants [186]. S_2 is the intensity of luminescence. The values of these parameters are given in Table 4.1.

Table 4.1 Parameter settings (nM denotes nanomole) [186].

Parameter	Value	Unit		Parameter	Value	Unit
$p_{bind}(0)$	0	1		γ	0.004	$min^{-1}nM^{-1}$
$S_1(0)$	0.44	nM		a, b	0.0085	nM/min
$S_2(0)$	5.7	nM		b_1	0.0189	min^{-1}
κ	0.112	min^{-1}		b_2	0.0901	min^{-1}

Figure 4.4 gives the intensity of luminescence for different signalling molecules concentrations. It shows that there is an initial decline for any concentration of AHLs surrounding the bacteria, which is not well understood [186].

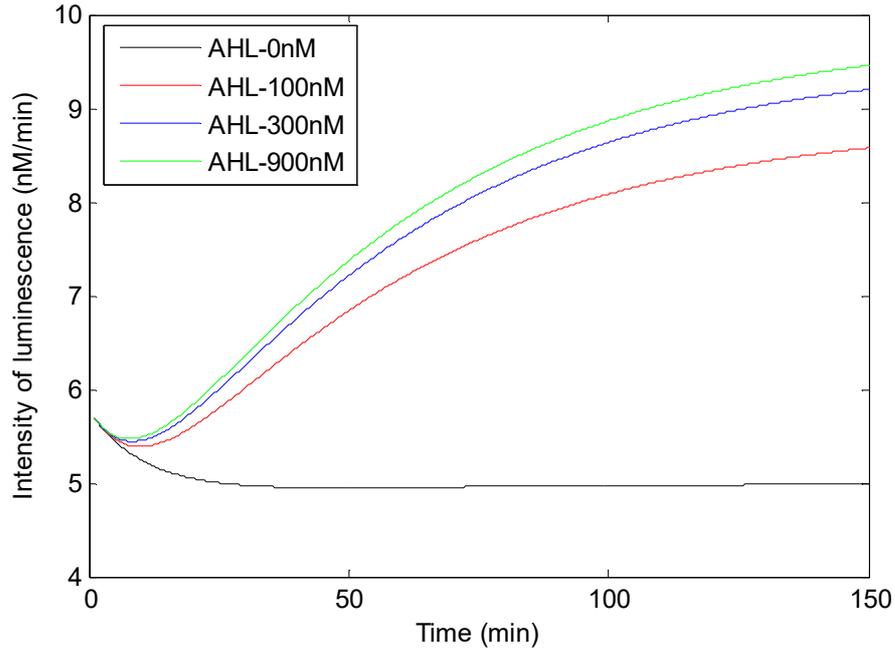


Figure 4.4 Luminescence over time for different AHL concentrations (nM denotes nanomole).

4.5 Channel model

4.5.1 Overview

A communication channel based on diffusion, inspired by bacterial communication systems will be described in this section. Information molecules propagate through the medium via diffusion. Since the diffusion process is a random propagation procedure, there is a probability that the information molecules do not arrive at the receiver or are delayed which will lead to arrival at the receiver in an incorrect time slot [24]. In this section, the focus is on the transmission of information over ranges up to a few micrometres through a diffusion process of identical molecules within the medium which includes the transmitter, the receiver and the propagation channel. The information molecules are released into the medium where the nanomachines are suspended. The molecules propagate through Brownian motion whose random

transmission process will cause noisy reception in terms of the receiver arrival time [24].

4.5.2 The diffusion medium

In diffusion-based MC, the information molecules propagate through the fluidic transmission medium between the transmitter and receiver via diffusion. Generally, the information molecules can be chosen as proteins, protein complexes, peptides, DNA sequences or other molecular structures [38]. The motion of information molecules is inspired by the forces produced by the constant random thermal motion of the molecules within the fluid medium. To be available to communicate through the diffusion channel, the information molecules must be able to be easily fabricated by the transmitter. Also, there must be enough building blocks for the messenger molecules to reside in the environment. In addition, the information molecules must be no harm to the components of the communication system [38].

Molecular diffusion is the thermal motion of all molecules at temperatures above absolute zero. Following this principle, when an uneven distribution of particles exist in a certain environment, they tend to diffuse away in order to reach a uniform concentration throughout all the space. Molecular diffusion could also be considered as a special situation of a random walk or Brownian motion, which is used to model the random motion of particles suspended in a fluid or gas, and also some other phenomena in diverse fields. The emission and propagation of the autoinducers are subject to this physical law. When a certain amount of autoinducers are released by the transmitter node, a peak of concentration appears adjacent to the node. Then the autoinducers would diffuse away as explained before, following the gradient of the concentration, thus going away from the transmitter. The phenomenon of molecular diffusion is typically described mathematically using Fick's laws of diffusion [187].

To simplify the communication system, the three dimensional medium 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 bacteria inside the transmitter node can produce various concentrations of Type-I molecules to be transmitted through the channel. 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-I molecules through the corresponding Type-I receptors (LuxR), followed by the production of GFP by the bacteria, which would be used to decode the transmitted information.

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-II receptors in the node are not activated, and gene *luxI* is repressed, which means that no extra Type-I autoinducers could be generated. Thus the channel can be modelled as a diffusion-based MC channel as follows.

The proposed channel is a binary asymmetric channel (BAC) with binary input and binary output and an average probability of error. To effectively represent the transmitted symbols, the propagation time is divided into time slots, also called symbol durations, of equal length, denoted by t_s , in each of which only one symbol propagates. The “intended symbol” and the “received symbol” refer to the symbol sent by the transmitter and received by the receiver, in the current time slot,

respectively. The 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, 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 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 have adverse effects in communication systems, particularly when the system is stochastic [188]. It should be noticed that the received signals tend to spread to adjacent symbols and smear into each other when a sequence of symbols are transmitted [189]. 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 diffusion 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 incorrect decoding of the received symbol of the next time slot.

As shown in Figure 4.2, which shows the communication setup of the QS system, the transmitter is at a distance d away from the receiver which has a radius of R , the value of which is related to the number of bacteria m in the receiver node.

In essence, the information molecules propagate through the fluid medium undergoing Brownian motion which is a random procedure and a probabilistic behaviour, which means that the molecules are not ensured to arrive at the receiver. In other words, there is a probability that the molecule will hit the receiver at a time slot. According to Fick's second law of diffusion [187], the escape probability

$\mathcal{S}(d, t)$ in a three dimensional environment can be described with the following backward difference equation at a given time t :

$$\frac{\partial \mathcal{S}(d, t)}{\partial t} = D \nabla^2 \mathcal{S}(d, t) \quad (4.6)$$

The diffusion coefficient D is $4.9 \times 10^{-6} \text{cm}^2 \text{s}^{-1}$, which is settled as a conservative value for AHLs in water at 25°C [190]. Considering the medium homogeneous, the coefficient will be a constant for all the points in space. It shows that the time rate of change in escape probability is proportional to the curvature of escape probability and to the diffusion coefficient.

However, for this communication model, the capture probability, rather than the escape probability, is more important. Thus by solving equation (4.6), the capture probability $P(d, t)$ can be calculated by:

$$P(d, t) = \frac{R}{R + d} \text{erfc} \left\{ \frac{d}{2\sqrt{Dt}} \right\} \quad (4.7)$$

where $\text{erfc}\{\cdot\}$ is the complementary error function [191]. Equation (4.7) shows the probability that a molecule arrives at the receiver at a time slot from mathematical approach.

To achieve the hit time probability, which refers to the probability that an information molecule arrives at the receiver at a certain time t , equation (4.7) is differentiated with respect to time, obtaining the hit time distribution as:

$$h(d, t) = \frac{R}{R + d} \frac{d}{2\sqrt{\pi D}} \frac{1}{t^{3/2}} \exp \left(-\frac{d^2}{4Dt} \right) \quad (4.8)$$

The capture probability and the hit time probability are both affected primarily by the diffusion coefficient D , the radius of the receiver R which is affected by the

number of bacteria in the receiver node m , and the distance between the transmitter and receiver, d .

4.5.3 The communication channel model

The basic operation principles of the diffusion-based MC channel are explained above, as well as the behaviour of the molecules in the medium. In this part, the communication channel is modelled to analyse the performance of the communication method. As stated above, only one symbol propagates in single time slot which is denoted by t_s .

In this work, the communication channel is a binary one, where each molecule either arrives at the receiver or does not. When n molecules are sent at the start of the time slot, the number of molecules received within the current time duration N_C is a random variable and follows a binomial representation, which can be described as:

$$N_C \sim \text{Binomial}(n, P_1) \quad (4.9)$$

where P_1 represents the capture probability with receiver radius R , transmission distance d and symbol duration t_s , which can be calculated by:

$$P_1 = \frac{R}{R+d} \operatorname{erfc} \left\{ \frac{d}{2\sqrt{Dt_s}} \right\} \quad (4.10)$$

It has been stated above that the previous bits can have an influence on the current bit due to ISI. Thus 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 start of the previous symbol durations. Molecules which are received from the previous time slots during the current one are called residual molecules denoted by N_P , which is also a random variable. Also, it is assumed that the transmitted information includes k bits. Here, two cases are considered to analyse the channel:

- Only the immediately previous time slot is taken into consideration.
- All the previous time slots are taken into consideration.

A. *Only the immediately previous time slot is considered.*

When only one previous time slot is taken into consideration, 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 and the start of the previous symbol duration. The number of residual molecules N_p can be obtained by calculating the difference between two binomial distributions, the number of molecules received during two time slots and the number of molecules received by the current time slot, which can be described as:

$$N_p \sim \text{Binomial}(n, P_2) - \text{Binomial}(n, P_1) \quad (4.11)$$

where P_2 is the capture probability for two consecutive time slots, which can be obtained by:

$$P_2 = \frac{R}{R+d} \operatorname{erfc} \left\{ \frac{d}{2\sqrt{D * 2t_s}} \right\} \quad (4.12)$$

The binomial distribution can be approximately replaced by a normal or Gaussian distribution which has the same mean and standard deviation. In other words, if a binomial distribution $\mathbb{P} \sim B(n_t, p_{suc})$ has mean $n_t p_{suc}$ and standard deviation $\sqrt{n_t p_{suc} (1 - p_{suc})}$, it can be approximated by a normal distribution $\mathbb{P}^* \sim \mathcal{N}(n_t p_{suc}, n_t p_{suc} (1 - p_{suc}))$ when p_{suc} is far from 0.5 and $n_t p_{suc}$ is large enough. Thus, by replacing the binomial distribution with normal distribution, the number of current molecules N_C and residual molecules N_p can be described as:

$$N_C \sim \mathcal{N}(nP_1, nP_1(1 - P_1)) \quad (4.13)$$

$$N_P \sim \mathcal{N}(nP_2, nP_2(1 - P_2)) - \mathcal{N}(nP_1, nP_1(1 - P_1)) \quad (4.14)$$

Given that the one-bit information of the current intended symbol is s_{cu} and that of the previous time slot is s_{pr} , the current received symbol depends on both of them. For this binary channel model, both s_{pr} and s_{cu} can be taken as zero or one. The possibilities of outputs with respect to different inputs, which are represented by the combinations of s_{pr} and s_{cu} , for this binary channel model are shown in Table 4.2.

Table 4.2 Binary Channel Model (Possibility Representation).

$s_{pr}s_{cu}$ (input) Y(output)	00	01	10	11
0	$P_{R(0,0)}$	$1 - P_{R(0,1)}$	$P_{R(1,0)}$	$1 - P_{R(1,1)}$
1	$1 - P_{R(0,0)}$	$P_{R(0,1)}$	$1 - P_{R(1,0)}$	$P_{R(1,1)}$

As shown in Table 4.2, there are four different cases for the binary channel model which are denoted by bit pairs “00”, “01”, “10”, “11” for received symbol decoding, according to the different values of s_{cu} and s_{pr} . Y represents the received symbol in the current time slot. The probability $P_{R(s_{pr},s_{cu})}$ represents the probability of successfully receiving the current intended symbol in the current time slot, where s_{pr} is the one-bit information represented by the previous intended symbol and s_{cu} is that of the current one. The different four cases are displayed below.

Case “11”: Both the one-bit information represented by the previous intended symbol and the current one are ‘1’. Thus the current received symbol is affected by both the previous and current time slot. The number of molecules received at the current symbol duration can be described as:

$$\begin{aligned}
N_{hit} = N_P + N_C &\sim \mathcal{N}(nP_2, nP_2(1 - P_2)) - \mathcal{N}(nP_1, nP_1(1 - P_1)) \\
&+ \mathcal{N}(nP_1, nP_1(1 - P_1)) \sim \mathcal{N}(nP_2, n[P_2(1 - P_2) + 2P_1(1 - P_1)])
\end{aligned} \tag{4.15}$$

The probability of success is the case when the number of received molecules exceeds the chosen threshold which is denoted by τ . Thus it can be calculated by:

$$P_{R(1,1)} = P(N_{hit} \geq \tau) \approx Q\left(\frac{\tau - nP_2}{\sqrt{n[P_2(1 - P_2) + 2P_1(1 - P_1)]}}\right) \tag{4.16}$$

where $Q(\cdot)$ is the tail probability of the standard normal distribution with zero mean and unity variance, often referred to as Q -function, which can be calculated by the tail integration of normal distribution.

Case “10”: The one-bit information represented by the previous intended symbol is ‘1’ while for the current intended symbol it is ‘0’. No molecules are sent at the start of the current time slot which means that the received symbol should be ‘0’ if it is correctly decoded. In this case, the molecules overflowing from the previous time slot have a negative influence on the successful decoding of the current intended symbol. It is considered to be an unfavourable case due to ISI. For successful decoding, the received molecules should not exceed the threshold. The number of molecules received can be written as:

$$\begin{aligned}
N_{hit} = N_P &\sim \mathcal{N}(nP_2, nP_2(1 - P_2)) - \mathcal{N}(nP_1, nP_1(1 - P_1)) \\
&\sim \mathcal{N}(n(P_2 - P_1), n[P_2(1 - P_2) + P_1(1 - P_1)])
\end{aligned} \tag{4.17}$$

The probability of success can be calculated by:

$$P_{R(1,0)} = P(N_{hit} < \tau) \approx 1 - Q\left(\frac{\tau - n(P_2 - P_1)}{\sqrt{n[P_2(1 - P_2) + P_1(1 - P_1)]}}\right) \tag{4.18}$$

Similarly, the expressions of the other two cases are given by [38]:

$$P_{R(0,1)} = P(N_{hit} \geq \tau) = I_{P_1}(\tau, n - \tau + 1) \quad (4.19)$$

$$P_{R(0,0)} = 1 \quad (4.20)$$

where $I_{P_1}(\tau, n - \tau + 1)$ is the regularized incomplete beta function.

The 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. Various kinds of noise, interference and phase jitter may cause degradation of the transmitted signal. Here, the BER refers to the probability of one bit error when information symbols are transmitted in the diffusion-based communication channel. The total average error probability can be obtained by calculating the average BER of all the four states stated above. According to [38], 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 proposed model, the symbol duration t_s is chosen as the time at which 60% of the molecules have arrived at the receiver, i.e. $P(d, t)$ in equation (4.7) is taken as 0.6. It has been stated that the error probability is a function of parameters n, τ, m, d, t_s . Thus in this work, for given values of n, m, d, t_s , the BER is taken as the value when the average error probability reaches its minimum for $\tau \in [1, n]$. The optimized BER versus molecules per bit for different transmission distance and different number of bacteria in the receiver node is shown in Figure 4.5.

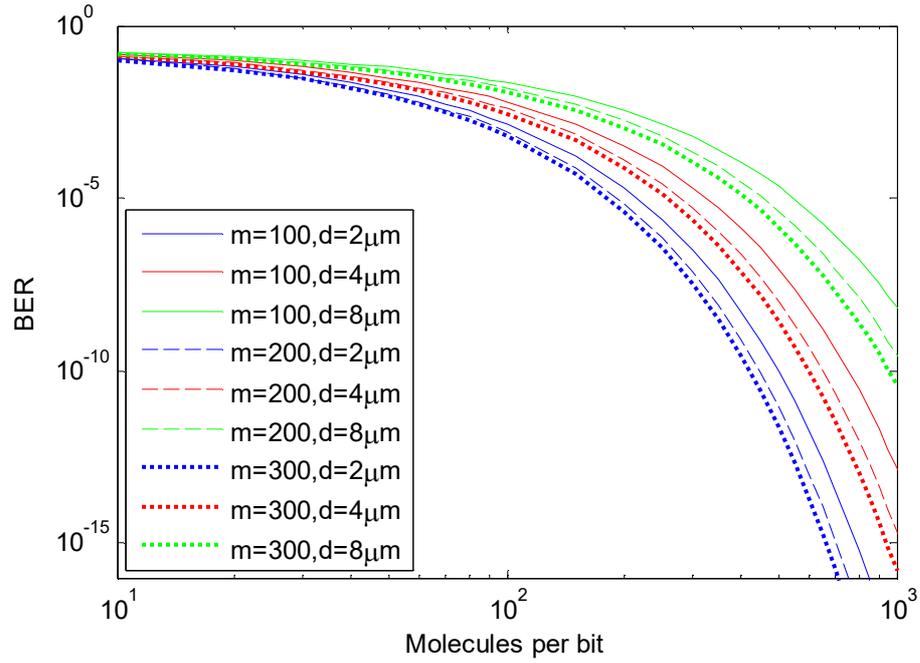


Figure 4.5 Channel BER performance.

BER vs. molecules per bit for different number of bacteria in the receiver node ($m = 100, 200, 300$) and different transmission distance ($d = 2\mu\text{m}, 4\mu\text{m}, 8\mu\text{m}$).

Figure 4.5 shows the bit error probabilities for this communication channel, with respect to different numbers of bacteria in the receiver node and different transmission distances, which can be considered as a noise for the channel. It demonstrates that the communication system has a better performance with a smaller BER if large numbers of molecules are sent at the start of one time slot. In addition, the channel performs better with a smaller transmission distance and a larger population of bacteria in the receiver node.

B. *All the previous time slots are considered.*

As stated in the previous section, it is assumed that n information molecules are sent at the start of each symbol duration. Also, the transmitted information is assumed to be 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. Being consistent with the case when only the previous one time slot is taken into consideration, here, t_s is set to be the time when 60% of the signaling molecules reach the receiver node. Also, α_j ($j \in [2, i]$) is used to represent the transmitted information symbol from the $(j - 1)^{\text{th}}$ previous time slot; and α_1 represents the transmitted symbol in the current time slot. According to equation (4.13), 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 [192]:

$$N_c \sim \alpha_1 B(n, P_1) \sim \alpha_1 \mathcal{N}(nP_1, nP_1(1 - P_1)) \quad (4.21)$$

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^{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 [192]:

$$\begin{aligned} N_{p(q)} &\sim \sum_{j=2}^i \alpha_j (B(n, P_j) - B(n, P_{j-1})) \sim \\ &\sum_{j=2}^i \alpha_j (\mathcal{N}(nP_j, nP_j(1 - P_j)) - \mathcal{N}(nP_{j-1}, nP_{j-1}(1 - P_{j-1}))) \quad (4.22) \\ &\sim \mathcal{N}(\mu_{p(q)}, \sigma_{p(q)}^2) \end{aligned}$$

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 (4.10). $\mu_{p(q)}$, $\sigma_{p(q)}^2$ 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 $N_{hit(q)}$ 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' ($\alpha_1 = 1$), 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:

$$N_{hit(q)} = N_{p(q)} + N_c \sim \mathcal{N}(\mu_{hit(q)}, \sigma_{hit(q)}^2) \quad (4.23)$$

where $\mu_{hit(q)}$ and $\sigma_{hit(q)}^2$ are the expectation and variance of the distribution of $N_{hit(q)}$.

Thus, the error probability for this case is:

$$p_{e1(q)} = \frac{1}{2^i} P(N_{hit(q)} < \tau) = \frac{1}{2^i} \left(1 - Q \left(\frac{\tau - \mu_{hit(q)}}{\sqrt{\sigma_{hit(q)}^2}} \right) \right) \quad (4.24)$$

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:

$$N_{hit(q)} = N_{p(q)} \sim \mathcal{N}(\mu_{p(q)}, \sigma_{p(q)}^2) \quad (4.25)$$

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

$$p_{e0(q)} = \frac{1}{2^i} P(N_{hit(q)} > \tau) = \frac{1}{2^i} Q\left(\frac{\tau - \mu_{p(q)}}{\sqrt{\sigma_{p(q)}^2}}\right) \quad (4.26)$$

It is assumed that the transmitter sends ‘0’s and ‘1’s with equal frequency. Due to the fact that there exist 2^{i-1} different error patterns for a fixed transmitted symbol in the current time slot, 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 using:

$$p_{er} = \sum_{q=1}^{2^{i-1}} (p_{e1(q)} + p_{e0(q)}) \quad (4.27)$$

The BER is the value when the error probability reaches the minimum for $\tau \in [1, n]$, denoted by p in the subsequent chapters.

Figure 4.6 shows the BER performances when different numbers of previous time slots are taken into consideration. It is clear that the error probability is smaller when fewer previous time slots are considered. In addition, for the current time slot, the first previous time slot has the most significant effect on the error performance in the current time slot, and the second previous time slot has a smaller influence, and so forth. It means that in some complex conditions, it is reasonable to consider only the previous one or two time slots.

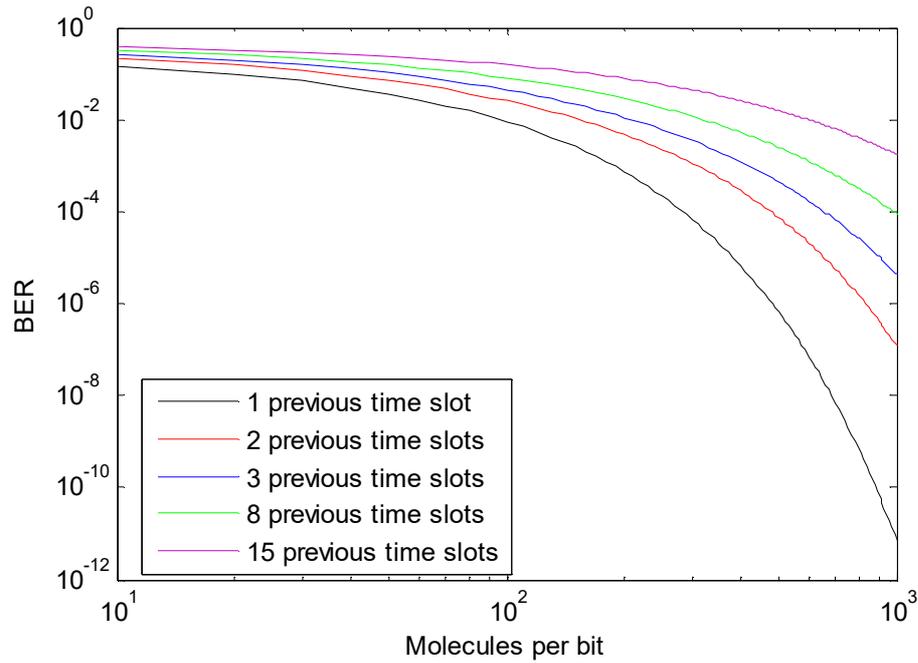


Figure 4.6 BER considering more than one previous time slot.

4.5.4 Communication discussions

In consideration of the discussions made in the previous section, the communication process in QS is slightly different from that of the classical approach. The transmitter encodes the information message into a certain number of autoinducers, instead of a single autoinducer. Thus it does not make any sense to decode from a single particle of the environment. Therefore, a unicast or point-to-point communication scheme could not be considered.

For a given transmission distance and a constant emission rate of autoinducers, the only message to be decoded and interpreted is the bacterial population in the colony, which makes the classical approaches not applicable in this case. The transmitted message is inherently constructed by the colony of bacteria, using the concentration of autoinducers in a sort of OOK modulation. From the communication perspective, it could be considered as a collective communication, which is defined as

communication that involves a group of processes, because that in a broad sense, all the bacteria are emitters and receivers of the global message in the meantime.

The main feature of the proposed QS communication scheme is that each bacterium is able to realize bidirectional communication and make a corresponding response to achieve a collective message that is understood by all bacteria. Specifically, the bidirectional property is supported by the fact that there exist different types of molecules acted as autoinducers, which can bind to distinct receptor proteins. Therefore, the message is only intended to be accepted by a certain group of bacterial colonies. From the perspective of the information theory, the addressing of the transmitted information is basically encoded in the chemical structure of the autoinducer. Also, it should be noted that in this chapter, the feedback channel which is used for retransmission is not taken into consideration; the retransmission issue will be discussed in detail in Chapter 6. Besides, it should also be noted that some sources of noise, attenuation or interference for this communication system are neglected. For example, the secretion of enzymes that destroy the autoinducers, and the production of autoinducer antagonists could be considered a source of attenuation, since they diminish the concentration of autoinducers that the receiver node would detect.

4.6 Channel capacity analysis

In this section, the fundamental concept of the capacity of the QS communication channel will be discussed, which is a measure of how much of information per channel usage could be achieved through the channel. According to the information theory, the capacity of the channel is given by the maximum of the mutual information between the input and the output of the channel, where the maximization is with respect to the input distribution.

Let X and Y be the random variables representing the channel input and output, respectively. Also, it is assumed that $P_{Y|X}(y|x)$ is the conditional distribution function of Y given X , which is considered an inherent fixed property of the communication channel. Then the joint distribution $P_{X,Y}(x,y)$ could be obtained by:

$$P_{X,Y}(x,y) = P_{Y|X}(y|x)P_X(x) \quad (4.28)$$

where $P_X(x)$ is the marginal distribution of X .

Thus the mutual information, which is a measure of mutual dependence between variables X and Y , is defined as:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} P_{X,Y}(x,y) \log \left(\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)} \right) \quad (4.29)$$

The information channel capacity, which is defined as the maximum mutual information, could be described as $C = \max I(X,Y)$.

It has been stated in previous sections that the quorum communication channel is modelled as a binary channel, which means that for both the input X and output Y , the value would be either binary '1' or '0'. Also, in Figure 4.6, it is clear that the majority of the noise results from the previous three time slots, while the effect of other time slots is slight when compared. Thus, to simplify the investigation of the channel capacity, only the previous three time slots are taken into consideration.

Similar with the case when only the previous one time slot is considered in Section 4.5.3 (A), for the case of considering three previous time slots, there are 16 different cases for the binary channel model, according to the different values of the symbol in the current time slot and the three symbols in the previous time slots. The possibilities of unsuccessfully decoding the current intended symbol in the current time slot are denoted by $P_{0000}, P_{0010}, P_{0100}, P_{0110}, P_{1000}, P_{1010}, P_{1100}, P_{1110}, P_{0001},$

$P_{0011}, P_{0101}, P_{0111}, P_{1001}, P_{1011}, P_{1101}, P_{1111}$, where the four bits of the subscripts represent the symbol in the third, second, first, and current time slot, in sequence. Following the derivation method mentioned in Section 4.5.3 (A), the expressions of the 16 probabilities could be achieved, which will be given in the Appendix A.

Figure 4.7 shows the binary channel which takes three previous time slots into consideration. Here, ϵ_0 and ϵ_1 represent the probability of unsuccessfully decode the information when the symbol in the current time slot is binary ‘0’ and ‘1’, respectively, which could be expressed as:

$$\epsilon_0 = [P_{0000} + P_{0100} + P_{0100} + P_{0110} + P_{1000} + P_{1010} + P_{1100} + P_{1110}]/8 \quad (4.30)$$

$$\epsilon_1 = [P_{0001} + P_{0011} + P_{0101} + P_{0111} + P_{1001} + P_{1011} + P_{1101} + P_{1111}]/8 \quad (4.31)$$

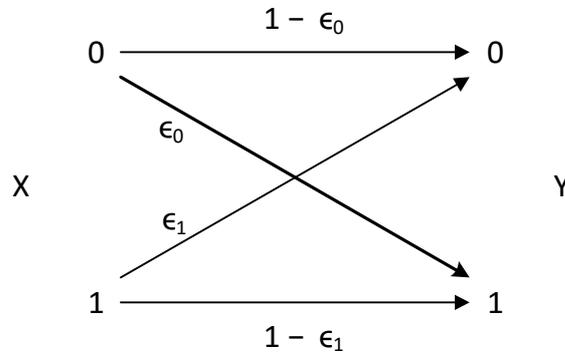


Figure 4.7 Binary asymmetric channel (BAC).

It is assumed that for the input variable X , which represents the symbol in the current time slot, the probabilities of occurrence for binary ‘1’ and ‘0’ are identical, which means that

$$P_X(x = 0) = P_X(x = 1) = \frac{1}{2} \quad (4.32)$$

Also, here the output variable Y represents the received symbol in the current intended time slot. As shown in Figure 4.7, it is clear that the conditional probabilities could be represented by:

$$P_{Y|X}(y = 0|x = 0) = 1 - \epsilon_0; \quad (4.33)$$

$$P_{Y|X}(y = 1|x = 0) = \epsilon_0; \quad (4.34)$$

$$P_{Y|X}(y = 0|x = 1) = \epsilon_1; \quad (4.35)$$

$$P_{Y|X}(y = 1|x = 1) = 1 - \epsilon_1; \quad (4.36)$$

Thus for the output variable Y , the probability when the received symbol is binary '0' can be calculated by:

$$\begin{aligned} P_Y(y = 0) &= P_{X,Y}(x = 1, y = 0) + P_{X,Y}(x = 0, y = 0) \\ &= P_{Y|X}(y = 0|x = 1)P_X(x = 1) + P_{Y|X}(y = 0|x = 0)P_X(x = 0) \\ &= \frac{1}{2}(\epsilon_1 + 1 - \epsilon_0) \end{aligned} \quad (4.37)$$

Similarly, it can be derived that

$$P_Y(y = 1) = \frac{1}{2}(1 - \epsilon_1 + \epsilon_0) \quad (4.38)$$

Therefore, with equations (4.32) to (4.38), the mutual information, which has been defined by equation (4.29), can be obtained. Using the t_s values selected, the mutual information is calculated, using different values of the number of molecules emitted at the start of each time slot n and the threshold τ , over different transmission distances ranging from 5 to 30 μm . The results are shown in Figure 4.8, where n is fixed as 100 and 500 molecules, respectively. Also, the number of bacterial molecules in the transceiver node is set to be 100. It shows that for each distance and n couple, there exists an ideal threshold value (τ^*) which maximizes the mutual information rate. In addition, when the value of threshold is in a lower range, the

receiver node would decide that the current symbol represents binary ‘1’ after the arrival of a few molecules. However, at high threshold values, a large number of autoinducers are required to arrive at the receiver to decode the current symbol as ‘1’. Thus, for the case when the transmitted symbol in the current time slot is ‘0’, if the threshold value is selected smaller than τ^* , ϵ_0 attains a very high value and results in a low mutual information since the residential molecules from the previous time slots makes it easy to exceed the threshold. On the other hand, for the case when the current transmitted symbol is ‘1’, if $\tau > \tau^*$, as threshold values increases, the value of ϵ_1 would increase. After a certain threshold value, it would be almost impossible for an adequate number of autoinducers to exceed the threshold value, even with the help of the residential molecules from previous time durations. This would lead to a sharp decrease of the mutual information, which matches the results in Figure 4.8.

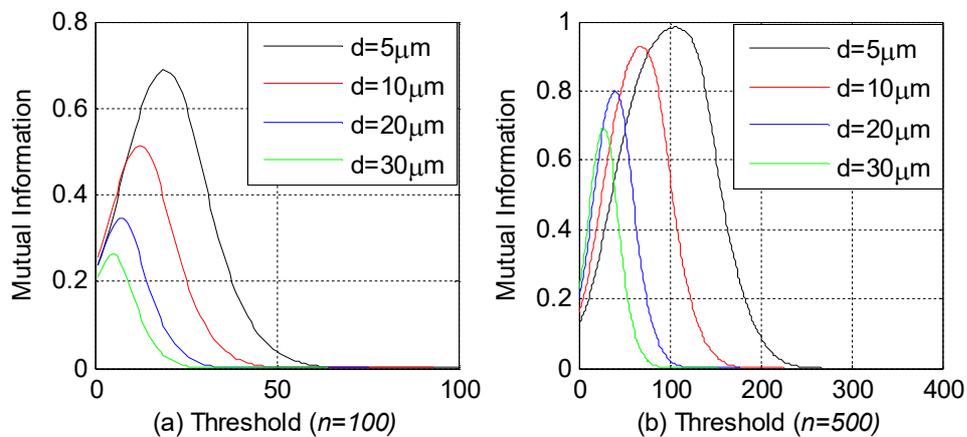


Figure 4.8 Effect of threshold on the mutual information.

Therefore, by calculating the maximum values of the mutual information, the channel capacity for OOK may be achieved, which has been shown in Figure 4.9. The corresponding values to get the maximum mutual information are shown in Figure 4.10. It is clear from these figures that the values of optimal threshold τ^* and channel capacity C decrease as the distance increases for a fixed value of n . In

addition, the channel capacity could attain very high values at each distance after a certain n value. However, with the increase of the transmission distance, the smallest value of n that achieves high capacity values also increase. Moreover, since the capture probabilities for longer distances should be lower than that for shorter distances, the τ^* values decrease with distance. Also, there is an almost linear relationship between the values of τ^* and n for fixed values of distance.

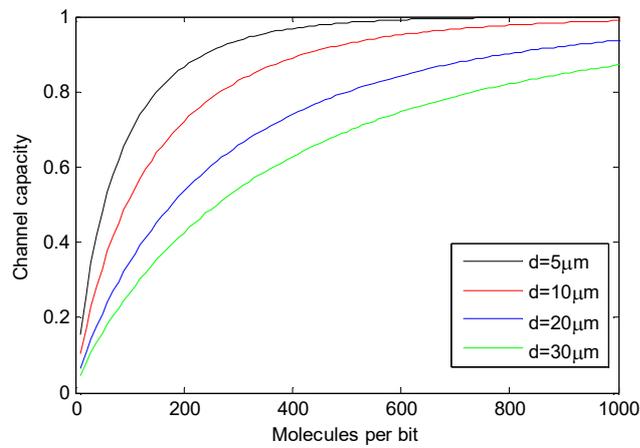


Figure 4.9 Channel capacity.

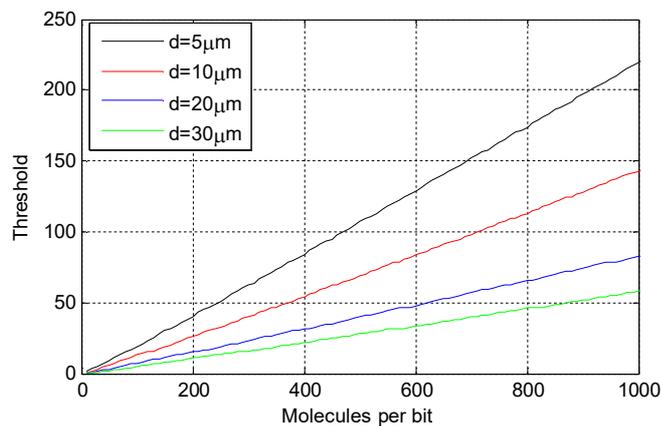


Figure 4.10 Selected threshold when calculating capacity.

4.7 Summary

In recent years, bacteria have been considered as one approach for MC using QS. In this chapter, a review of existing models of QS has been given at first. After that, a

bacterial communication network model through a diffusion channel between transmitter and receiver bacterial populations is proposed. The communication depends on free molecular diffusion to transport the information symbols from the transmitter to the receiver. A time slotted structure is used in the communication channel model. In order to investigate the channel performance, two optimization problems have been defined, including getting the minimum error probability and maximum channel capacity. Type-I autoinducer, 3-oxo-C6-HSL, is applied as the messenger molecule. It should be noted that Type-II autoinducer which is going to be used in the feedback channel in Chapter 6 is not applied in this chapter.

Based on the investigations in this chapter, it has been shown that the symbol duration of the system is mainly influenced by the distance between the transmitter and receiver nodes. Also, the symbols belonging to the previous time slots would have a significant ISI effect on the current intended symbol, especially the symbols from the previous three symbol durations. Based on these discussions, the successful reception probabilities have been derived. In addition, it has been indicated that the selection of the appropriate threshold value is critical for the BER, the mutual information and channel capacity calculation.

However, there exists a major challenge for the proposed QS communication model. Specifically, in this model, the bacteria are organized into clusters. Operations such as the emission of autoinducers are considered to be performed by the bacterial colony, instead of individual bacterium. This becomes the major shortcoming of this option, as bacteria sense the environment, release autoinducers, and reproduce themselves as individuals. It is difficult to ensure that all these actions are able to be performed by all the bacteria at the same time and with the same approach.

CHAPTER 5.

FEC IN BACTERIAL COMMUNICATIONS

5.1 Introduction

It has been stated in previous chapters that bacterial communications, which operates in aqueous environments and utilises molecules to encode and transmit information between bacterial nanomachines, represents a new communication paradigm of MC. Bacteria communicate with each other through the process of QS, using signalling molecules (autoinducers), which are released into the surrounding environment. Also, bacteria in a colony are able to sense their cell density through the accumulation of autoinducers, which means that the concentration of external autoinducers is correlated with the bacterial cell population density. In Chapter 4, a diffusive bacterial communication network between two populations of bacteria has been modelled and analysed, using the well-studied bioluminescent bacteria species, *V. fischeri*, which has been used in environmental applications. With OOK modulation, 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 transmission and performance of natural and engineered QS networks [163]. For the bacterial communication system proposed in this thesis, 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 control techniques, which refer to mechanisms to detect and correct errors that occur in transmitting packets or frames, for reliable transmission. According to information and coding theory, error detection is the detection of errors caused by noise or other impairments during transmission from transmitter to the receiver, while error correction refers to the detection of errors and reconstruction of the original, error free data. Generally, to achieve error detection and correction, redundant bits are required to be appended to the transmitted message. These can then be used by receivers to check consistency of the delivered message, and to recover data determined to be corrupted. Good error control performance requires the selected error control scheme to be in harmony with the features of the communication channel. In this thesis, it is aimed that the transmission errors are corrected to maximum extent. Generally, error correction can be achieved via two different methods, ARQ and FEC [193]. Specifically, ARQ, which also refers to backward error correction, combines an error detection scheme with requests for the retransmission of erroneous data. When a data frame is received, it is checked using error detection techniques; if the check displays errors, retransmission of the frame is requested. The process is operated repeatedly, until the data frame is verified. For FEC, there is no retransmission but check bits are appended to the original information bits prior to transmission. After the transmitted encoded information is

received, the receiver reconstructs the information using the redundancy bits according to specific error correction coding schemes. It should also be noted that FEC can operate at either the bit, the byte or the packet level; for each case, appropriate redundant bits are added to the data by the sender to recover either bits, bytes or whole packets efficiently. In short, ARQ attempts to correct errors via retransmission, whereas FEC attempts to correct errors using parity bits. In some complex situations, ARQ and FEC are combined for error correction, which refers to the hybrid automatic repeat request (HARQ) technique [194]. For example, in some HARQ schemes, minor errors are corrected through relatively simple error correction codes, while major errors would be corrected via retransmissions.

Although some traditional FEC techniques have been developed for MC [195-198], there is still no single internationally accepted standard for the channel. This stems mainly from a number of technical challenges encountered while trying to establish efficient bacterial or molecular communications, the most important of which is the poor propagation of molecular signals. Specifically, there is significant influence from diffusion and ISI effects, and the attenuation suffered as a result of the former is severe except for small transmission distances.

Design of a proper error control mechanism for bacterial communications requires consideration of two main issues: reliability and energy consumption. Both of these issues are related to each other for the reason that the achievement of better reliability requires error correction techniques, which in turn requires extra energy consumption in coding and decoding the messages. This extra energy consumption required for reliability has detrimental effects on those applications that require transceiver nodes to be operated in the human body or other similar environments where energy supply and storage is difficult to implement. For a MC system, the

energy budget is a considerable limit because the nodes will consume energy from the environment. However, for most of the energy harvesting devices the output is only in the pW range which makes it too energy intensive to apply state of the art recursive coding schemes [199]. Thus, due to the severe energy constraints on bacterial or molecular communications, some appropriate schemes have to be designed, which not only achieves reliable information transmission but also ensures lower energy consumption at the same time. To summarise the above, there exist two well-established error correction techniques for reliable data transmission: FEC and ARQ.

In this chapter, FEC mechanisms will be applied to improve the proposed channel performance. The insertion of message redundancy using error correction coding is applied to increase transmission reliability and achieve lower power or energy consumption but at the cost of additional power consumption in coding and decoding process. This chapter is organized as follows: in Section 5.2, an overview of existing error correction coding for MC is given, followed by the brief introduction of FEC techniques in Section 5.3. Afterwards, in Section 5.4, Hamming codes are applied to the channel, and the analysis of the BER and energy consumption performance is given. Subsequently, MEC and LT codes are used in the channel model and explained in Section 5.5 and Section 5.6, respectively. The simulation results and discussions of each code have been given in the corresponding sections. Finally, Section 5.7 concludes the chapter.

5.2 Existing error correction coding for MC

Until recently, only a small amount of research has been carried out concerning the usage of the error correction coding schemes in bacterial or diffusion-based MC systems to increase their robustness. In [198], the authors utilize a simple block code

as a means to enhance the communication performance between two nanomachines in a diffusion-based MC system, which is considered as the first attempt to apply error correction coding techniques in MC. Results have shown that by employing Hamming codes, a coding gain of 1.7 dB over $1\mu\text{m}$ transmission distance can be achieved, which means that the use of an error correcting code is considered as a method of enhancing the performance of future MC networks. In addition, the paper also considered the energy budget of the implementation of Hamming coding to determine the critical distance, which refers to the distance at which the energy gain of the code overcomes its operational energy cost. Based on this paper, the authors in [200] extend the research using high order Hamming codes, specifically up to an order of 10. Results show that Hamming (31, 26) provides the shortest critical distance whilst Hamming (63, 57) provides the largest coding gain. Moreover, in [196], more complex codes, including Low-Density Parity-Check (LDPC) and C-RM codes, have been introduced into a diffusion-based MC system, and the performance enhancements they bring have been compared against Hamming codes in accordance with both coding gain and energy requirements. Also, in the same communication scenario, self-orthogonal convolutional codes (SOCCs) have been applied in [197]. Furthermore, in [19], a new channel code, ISI-free coding scheme, is proposed to increase the communication reliability while keeping the encoding and decoding complexity reasonably low in a diffusion-based MC system. Also, the BER performance of conventional repetition codes and convolutional codes has also been investigated. Results show that compared with convolutional codes, the proposed ISI-free code and conventional repetition codes offer similar performance with much lower complexity. Furthermore, baseband modulation techniques, specifically Return-to-zero (RZ) and Non-return-to-zero (NRZ), have been used in

combination with Reed Solomon (RS) channel coding to improve the performance of a calcium signalling-based MC system in [201]. These preliminary studies have demonstrated the feasibility of applying conventional modulation and coding techniques to calcium based MC systems.

5.3 FEC mechanisms

5.3.1 Digital communication system

This section is a brief introduction to the theory of FEC. The understanding of coding and its limitations requires some awareness of information theory and how it would help with a digital communication system. Here, information refers to a physical quantity which can be measured, transformed, stored, and moved from place to place [202]. Figure 5.1 illustrates a fairly general framework for a sink-to-sink digital communication link, which contains key features to perform physical actions on information. It is clear that the data message from the source is encoded and modulated for communication over the channel. After the channel, the message is demodulated, decoded and sent to the destination or receiver. According to information theory, all the elements in this link have mathematical descriptions and principles which govern their performance. The source of information can be analogue (e.g. audio or video signal) which is stored in the form of a wave signal, or digital (e.g. a data file) which is stored in the form of bits. The signal produced by the source is converted if necessary into a digital signal which is made up of ‘1’s and ‘0’s, by the source encoder. The number of bits to represent the data may exceed the number of bits of actual information content. Shannon’s Source Coding Theorem [63] ensures that a particular source of data can be compressed without any loss of information (lossless compression). When compressing a data stream, a source encoder thus removes redundancy present in the data. Generally, the source encoder

employs specific types of codes to achieve compression, called collectively source codes or data compression codes [202], including Huffman coding, run-length coding, arithmetic coding and Lempel-Ziv coding. In short, source encoding is the process of efficiently converting the output of either an analogue or a digital source into a sequence of binary digits. The information sequence is then passed through the channel encoder, with the purpose of introducing controlled redundancy in the binary information sequence via channel codes that can be used at the receiver to overcome the effects of noise and interference encountered in the transmission through the channel. The redundancy removed by the source encoder, which typically depends on the source in an unstructured way, is unstructured. In contrast, the appended redundancy for the channel encoder is structured and able to provide both uniform protection to all the information in the stream and indications of the error occurrences plus the way to correct them. In other words, the appended redundancy bits help to increase the channel reliability and improve the fidelity of the received signal. The channel encoder is the first step in the error correction or error detection process. Often, this operates by accepting a block of k input symbols and producing a block of $\varpi (> k)$ symbols at the output. The input to the channel encoder is referred to as the message symbols or information symbols (or bits, for binary codes). The binary sequence at the output of the channel encoder is passed to a digital modulator, serving as the interface to the communication channel. The modulator converts the channel encoded symbol sequences into signals appropriate for transmission over the channel, since specific channel-conforming representations are required in many channels. For example, in some cases, the channel requires the signal be sent as a continuous-time voltage, or an electromagnetic waveform in a specified frequency band. The most fundamental digital modulation techniques are

based on keying, including PSK (phase-shift keying), FSK (frequency-shift keying), ASK (amplitude-shift keying), and QAM (quadrature amplitude modulation). In the case of this thesis, it has been stated in Chapter 4 that the signal is represented by the concentration of molecules. Thus, simple ASK or *OOK* modulation is applied, which means that the emission of molecules represents a binary '1', whilst no release represents a binary '0'. Generally, the communication channel is the physical medium over which the information is conveyed between two distinct places. Common examples of channels include telephone lines, fibre-optic lines, microwave radio channels, internet cables and underwater acoustic channels. Whatever the physical medium used for transmission of the information, the essential feature is that the transmitted signal can be corrupted in a random manner by a variety of possible mechanisms. For example, some noise may be added to the signal or it may suffer from time delay, time jitter, attenuation due to the transmission distance or carrier offset, or inadvertent interference from other channels. Moreover, interference among symbols may occur if the signal is filtered by the channel response [202]. For the purpose of analysis, channels are always characterized by specific mathematical models, which are sufficiently accurate to represent the attributes of the communication scenario. For example, the bacterial communication channel proposed in this work is represented by a BAC channel, the detailed properties of which have been stated in Chapter 4. After the transmission across the channel, the demodulator or equalizer receives the signal and converts it into a sequence of symbols that represents estimates of the transmitted data symbols. This sequence of symbols then passes through the channel decoder which attempts to reconstruct the original information sequence from the channel encoding algorithm used by the channel encoder and the redundancy contained in the received data. Then

the source decoder provides an uncompressed representation of the data according to the source encoding algorithm, which is finally received by the destination. It should be noted that in this chapter, the source encoder and decoder, modulation and demodulation are not considered. Instead, the focus will be on channel encoding and decoding methods.

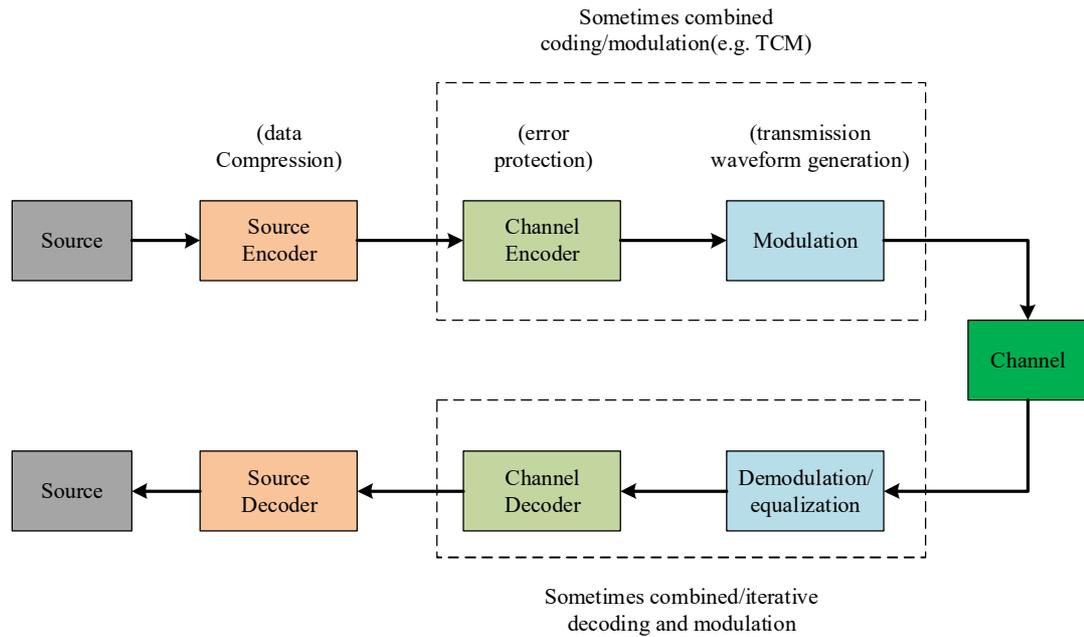


Figure 5.1 A general framework for digital communications [202].

To summarise, FEC attempts to control data transmission errors over unreliable or noisy communication channels by transmitting sufficient redundant data to allow the receiver to recover from errors unaided, i.e. without sender retransmissions. Redundant bits are formed using functions of the original message bits, which may be present in an unmodified form in the encoded output (systematic codes) or not (non-systematic codes) [202]. Therefore, FEC allows the receiver to correct errors without the need for a reverse channel to request data retransmission but at the expense of a fixed and a larger forward channel bandwidth. It is therefore applied in cases where retransmissions are either very expensive or almost impossible, such as

one-way communication links and when broadcasting to different receivers in multicast scenarios [203]. The FEC information is sparsely used in modems and generally stored in receiver mass storage devices so as to be able to recover the corrupted data. Different FEC codes are chosen for different conditions. Specifically, the code rate and the type of coding used determine the number of errors that can be corrected using FEC. A simple example is to send '000' ('111' correspondingly) instead of sending only one '0' ('1' correspondingly) to the channel. Due to noise in the channel, the received bits may become '001'. However, either '000' or '111' could have been sent but a maximum likelihood (ML) decoding scheme will decode the message correctly as '000' and therefore '0'. The error correction capability of the code is dependent upon many factors, but is usually improved by increasing the amount of redundancy added to the message. The drawback to adding a lot of redundancy is that either the transmission rate is decreased as the link must be shared among the significant data information as well as the redundant bits or the bit rate must increase on the link. The benefit, however, is that the receiver has a better chance of correcting the errors without having to request a retransmission of the message. The major established categories of FEC codes are:

- Block codes (such as linear block codes, cyclic codes)
- Convolutional codes
- Modern codes (such as Turbo codes)

Block codes operate on fixed-size blocks (packets) and their decoding is generally performed in polynomial time. There are many different types of block codes, such as linear block codes (e.g. Hamming codes), and cyclic codes (e.g. RS, Golay and BCH codes). Convolutional codes operate on bit or symbol streams of arbitrary length and are usually decoded using the Viterbi algorithm but their decoding may

become extremely complex [204]. In recent years, codes such as Turbo codes have been discovered that offer performance close to the Shannon theoretical limit [205].

5.3.2 Linear block codes

A block code is a code having all its codewords of the same length. For block codes, generally at the channel encoder, the input message bits will be divided into blocks of length k message bits and replaces each k message bits block with a ϖ -bit codeword by introducing r parity-check bits to each message block to improve reliability, where $r = \varpi - k$. Among the family of block codes, a block code \mathbb{C} of length ϖ with 2^k codewords is called a linear (ϖ, k) block code if any linear combination of codewords is also a codeword. Here, the elements of \mathbb{C} are its codewords, and the collection of all codewords is referred to as a codebook. Block codes are not necessarily linear, but in general all block codes used in practice are linear. Some typical linear block codes include Hamming codes, RS codes, Hadamard codes, Expander codes, Golay codes, and Reed-Muller (RM) codes. A (ϖ, k) block code has a code rate of k/ϖ , which refers to the level of redundancy that is applied to a block code. If z is an input sequence, then the output codeword is calculated by:

$$\varsigma = zG \quad (5.1)$$

where G is a $k \times \varpi$ generator matrix. The generator matrix for a linear block code is one of the bases of the vector space of valid codewords. The generator matrix defines the length ϖ and number of information bits k in each codeword, and the type of redundancy that is added; the code is completely defined by its generator matrix. The minimum distance of the code \mathbb{C} is the minimum weight of its nonzero codewords, or equivalently, the minimum distance between distinct codewords.

Associated with every linear block code generator G is a matrix H called the parity check matrix whose rows span the null space of G . If ζ is a codeword, the parity check matrix is used to detect errors in the received code by using the fact that $\zeta * H^T = 0$. For a received sequence $v = \zeta + e$, where ζ is the codeword and e is the error sequence, the syndrome is calculated by:

$$S = vH^T = (\zeta + e)H^T = eH^T \quad (5.2)$$

From equation (5.2), it can be determined that if a received vector is a codeword. Specifically, if S is 0, then the received message is correct. Otherwise, if S is not 0 and S is the j th row of H^T , then it implies an error occurs in the j th position of v . It should be noted that the syndrome depends only on the error pattern and is independent of the transmitted codeword.

When designing a block code to be used, there are several key considerations. Two concepts including Hamming distance and Hamming weight are required to be considered. Here, the (Hamming) distance between two codewords refers to the number of coordinates in which the two codewords differ. The (Hamming) weight of a codeword is the number of its nonzero coordinates. Generally, for an (ϖ, k, d_m) code, where d_m is the distance of the code, the detection capability (that is, the numbers of errors that the code can detect) is $(d_m - 1)$, while the correction capability (that is, the numbers of errors that the code can correct) is calculated by $\lfloor (d_m - 1)/2 \rfloor$. The values of ϖ and k must be chosen so that the minimum distance of a codeword is as large as possible. The distance of a codeword refers to the number of bits that are different in a codeword after transmission. Therefore, the minimum distance of a codeword signifies the number of single bit errors that can be detected by an error correction code. In addition, the number of redundancy bits

should be appropriate to balance the bandwidth of a communication channel and the error rate.

5.3.3 Cyclic codes

Cyclic codes are error correction codes that became popular around the time of block codes because of their use of finite-field algebraic structures [202]. Many widely used linear codes, including Hamming and Golay codes, have an equivalent cyclic representation. Cyclic codes are considered to be a subclass of linear block codes due to the additional property of invariance under a cyclic shift of ϖ -tuple codewords. In other words, if a codeword $\mathbb{C} = (\zeta_1, \zeta_2, \dots, \zeta_{\varpi-1}, \zeta_{\varpi})$ exists, then $(\zeta_2, \zeta_3, \dots, \zeta_{\varpi}, \zeta_1)$, $(\zeta_3, \zeta_4, \dots, \zeta_1, \zeta_2)$, etc. are also codewords, where each ζ_i is either binary ‘1’ or ‘0’. Therefore, each (ϖ, k) codeword can be represented by a code polynomial:

$$\zeta(\mathfrak{x}) = \zeta_1 \mathfrak{x}^{\varpi-1} + \zeta_2 \mathfrak{x}^{\varpi-2} + \dots + \zeta_{\varpi} \quad (5.3)$$

Also, each code polynomial can be represented by the product of the code generator polynomial $g(\mathfrak{x})$ and the data polynomial $z(\mathfrak{x})$:

$$\zeta(\mathfrak{x}) = z(\mathfrak{x})g(\mathfrak{x}) = (z_1 \mathfrak{x}^{k-1} + z_2 \mathfrak{x}^{k-2} + \dots + z_k)g(\mathfrak{x}) \quad (5.4)$$

Actually, this equation is the polynomial representation of equation (5.1). Even if this way to encode is the simpler, another procedure is used to obtain a systematic encoding, which again exploits some properties of the polynomial ring. The message $z(\mathfrak{x})$ is multiplied by $\mathfrak{x}^{\varpi-k}$, and the result is then divided by $g(\mathfrak{x})$, obtaining the remainder $\vartheta(\mathfrak{x})$. Then the codeword can be calculated by joining the message $z(\mathfrak{x})$ with the remainder $\vartheta(\mathfrak{x})$, which is displayed as:

$$\zeta(\mathfrak{x}) = z(\mathfrak{x})\mathfrak{x}^{\varpi-k} + \vartheta(\mathfrak{x}) \quad (5.5)$$

To check if received signal is error free, the remainder from dividing $\zeta(\mathfrak{x}) + e(\mathfrak{x})$ by $g(\mathfrak{x})$ is obtained (syndrome). If the result is zero, then the received signal is

considered error free, otherwise error pattern is detected from known error syndromes. Among the class of cyclic codes, the RS code is a typical cyclic block code that can detect and correct multiple errors, which is particularly good at dealing with bursts of errors and erasure errors. Generally, an (ϖ, k) RS code can recover the original k source symbols if any k of the ϖ transmitted symbols are received. For example, if a two-dimensional barcode, such as a QR code, is damaged, RS code treats the damage as an erasure and attempts to read the correct, undamaged portion of the barcode and infer the rest of the barcode.

5.3.4 Convolutional codes

Convolutional codes are similar to the block codes discussed in the previous section in that they involve the transmission of parity bits that are computed from message bits. Unlike block codes in systematic form, however, the transmitter does not send the message bits followed by (or interspersed with) the parity bits; in a convolutional code, the sender sends only the parity bits instead. Convolutional codes are often preferred in practice over block codes, because they provide excellent performance when compared with block codes of comparable encoding and decoding complexities. Whereas block codes take discrete blocks of k symbols and produce blocks of ϖ symbols that depend only on the k input symbols, convolutional codes are frequently considered as stream codes, by operating on continuous streams of symbols not partitioned into discrete message blocks. However, they can still be considered as rate k/ϖ codes, accepting k new symbols and producing ϖ new symbols. In addition, convolutional codes differ from block codes in that the encoder contains memory and the ϖ encoder outputs at any time unit depend not only on the k inputs but also on K previous input blocks. A convolutional code is specified by three parameters (ϖ, k, K) , where K is called the constraint length of the code. In

general, the longer the constraint length, the larger the number of parity bits that are influenced by any given message bit. Because the parity bits are the only bits sent over the channel, a larger constraint length generally implies a greater resilience to bit errors. The trade-off, though, is that it will take considerably more complicated operations to decode codes of long constraint length, so it is not appropriate to increase the constraint length arbitrarily and expect fast decoding. The convolutional encoding starts with K memory registers, each holding one input bit. Generally, all memory registers start with a value of 0. Also, the encoder has ϖ modulo-2 adders, which can be implemented with single Boolean XOR gates, and ϖ generator polynomials, one for each adder. An input bit z_1 is fed into the leftmost register. Using the generator polynomials and the existing values in the remaining registers, the encoder outputs ϖ symbols, which will be transmitted or punctured depending on the designed code rate. Then the bits in all the registers shift to the right (e.g. the bit in register 1 moves to register 2, the bit in register 2 moves to register 3, as shown in Figure 5.2), waiting for the next input bit. The encoder continues shifting in this way until all registers have returned to zero state. Figure 5.2 shows a (3,1,3) convolutional encoder with constraint length K of 3. The input bits are represented by $Z = (z_3, z_2, z_1)$, where z_1 is the first fed into the leftmost register (e.g. register 1), followed by the input z_2 and z_3 . Generator polynomials are $G_1 = (1,1,1)$, $G_2 = (0,1,1)$, $G_3 = (1,0,1)$. Therefore, the output bits can be calculated by:

$$\begin{cases} \zeta_1 = z_3 + z_2 + z_1 \\ \zeta_2 = z_2 + z_1 \\ \zeta_3 = z_3 + z_1 \end{cases} \quad (5.6)$$

It is clear that each encoded bit is a function of the present input bits and their past ones. Generally, decoding of convolutional codes is mostly performed by the Viterbi Algorithm [202].

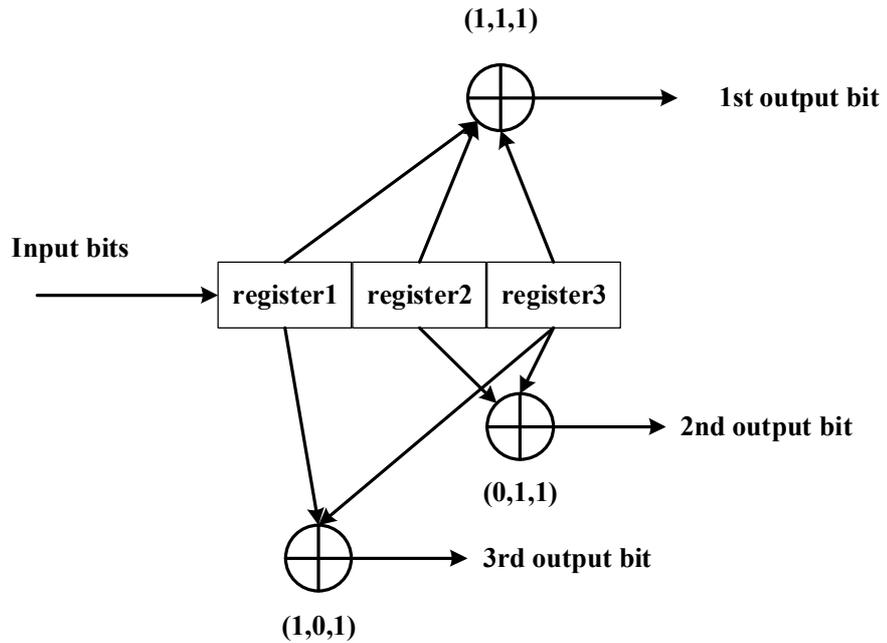


Figure 5.2 (3,1,3) convolutional encoder with constraint length 3.

5.3.5 Turbo codes

The first turbo code, based on convolutional encoding, was introduced in 1993 by Berrou, Glavieux, and Thitimajshima [206] as a class of high performance FEC codes. It was the first practical code to closely approach the channel capacity, a theoretical maximum for the code rate at which reliable communication is still possible given a specific noise level. Turbo codes have also been termed parallel concatenated codes, since their decoding complexity is relatively small for the dimension of the code, very long codes are possible, so that the bounds of Shannon's channel coding theorem become, for all practical purposes, achievable [207]. The turbo code encoder consists of two or more systematic block codes which share messages via interleavers. A key element of turbo codes is the iterative decoding algorithm, in which decoders for each constituent encoder operate on the received data in sequence. Each decoder is a soft output decoder, producing an estimate of the

probabilities of the transmitted symbols. Probabilities of the symbols from one encoder, known as extrinsic probabilities, are passed to the other decoder (in the symbol order appropriate for the encoder), where they are used as prior probabilities for the other decoder. The decoder thus passes probabilities back and forth between the decoders, with each decoder combining the evidence it receives from the incoming prior probabilities with the parity information provided by the code. After a number of iterations, the decoder converges to an estimate of the transmitted codeword. Since the output of one decoder is fed to the input of the next decoder, the decoding algorithm is called a turbo decoder. The two main types of decoder are Maximum A Posteriori (MAP) and the Soft Output Viterbi Algorithm (SOVA), looking for the most likely symbol received and the most likely sequence, respectively [208]. MAP was selected by Berrou as the optimal decoder for turbo codes, which looks for the most probable value for each received bit by calculating the conditional probability of the transition from the previous bit, given the probability of the received bit. Due to the relatively complex encoding and decoding details of turbo codes, readers referred to Sklar [209] for more details of them.

5.3.6 Discussion

It has been stated above that for a bacterial/molecular communication system, there is a considerably restricted energy budget due to the energy storage capabilities of the bio-nanomachines, although the bacterial nodes are able to harvest energy from the nutrient solution. Here, energy harvesting is a developing technology that seeks to exploit naturally-occurring energy to power systems, rather than relying on external sources such as batteries. However, the output from most energy harvesting devices is usually small and intermittent, making it quite energy intensive to apply the advanced recursive coding schemes. In a communication system, the receiver

side BER can often be reduced and improved by increased signal strength, by choosing a slow and robust modulation scheme, by line coding schemes or by using channel coding schemes which can be employed to detect and possibly correct certain number of errors that occurring in the transmitted information symbols.

Four broad methods of channel coding, including linear block coding, cyclic coding, convolution coding and turbo coding, have been outlined above. In the following section, Hamming coding, MEC and LT coding, which map a certain number of message symbols to a certain number of code symbols in three different methods, are applied for error correction coding of the channel. Although these relatively simple codes are not particularly powerful by the current standards, they are still employed, and are proved to be very efficient in terms of the channel reliability and energy budget for MC. According to Chapter 4, the proposed channel is a BAC, which has an average error probability of p . Here, it should be noted that even if there are two different probabilities for ones and zeros for the BAC channel, the average error probability can still be used to analyse the system performance. Theoretically speaking, the application of redundancy for error correction will decrease the channel BER.

5.4 Hamming coding

5.4.1 Introduction of Hamming codes

Hamming codes, which are the earliest and simplest example of linear block codes, can detect up to two-bit errors or correct one-bit error without detection of uncorrected errors. Also, Hamming codes are perfect codes, which means that they achieve the highest possible rate for codes with their block length and minimum distance of three [210]. Hamming codes, which are described as $(2^r - 1, 2^r - r - 1)$ codes, are used to form coded output blocks of length $\varpi = 2^r - 1$, from original

input blocks of length $k = 2^r - r - 1$, where r is the number of parity check bits. The minimum distance, d_m , of this type of block code is 3, which means that only one error can be corrected in each block. The structure of a transmission with a Hamming code is illustrated in Figure 5.3.

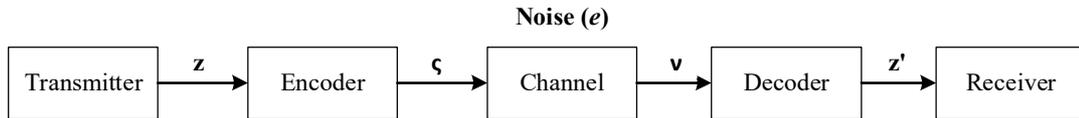


Figure 5.3 Hamming coding path.

The parameters in Figure 5.3 are explained in Table 5.1.

Table 5.1 Parameters of Hamming coding path.

Parameters	Definition
z	Original message
ζ	Codeword
e	Noise sequence
v	Received sequence
z'	Decoded received sequence

A good tutorial example of Hamming codes is the (7, 4) version, which transmits $\varpi = 7$ bits for every $k = 4$ source bits, so the number of parity check bits is 3. Generally, the number of parity check bits needed is given by the Hamming rule, a function of the number of bits of information transmitted. Each of the three parity bits are parity for three of the four data bits, and no two parity bits are for the same three data bits. All of the parity bits are even parity. Given data bits z_1, z_2, z_3, z_4 , the rules for obtaining the bits that compose the parity matrix are:

$$\begin{cases} \rho_1 = z_2 + z_3 + z_4 \\ \rho_2 = z_1 + z_3 + z_4 \\ \rho_3 = z_1 + z_2 + z_4 \end{cases} \quad (5.7)$$

where ρ_1, ρ_2, ρ_3 are the parity bits. There is a fourth equation for a parity bit that may be used in Hamming codes: $\rho_4 = z_1 + z_2 + z_3$. Here, one method to create a generator matrix G will be explained. It is denoted that Z is a 1×4 vector, which is $[z_1, z_2, z_3, z_4]$; each data bit is represented with a column vector as follows:

$$\begin{cases} z_1 = [1\ 0\ 0\ 0]^T \\ z_2 = [0\ 1\ 0\ 0]^T \\ z_3 = [0\ 0\ 1\ 0]^T \\ z_4 = [0\ 0\ 0\ 1]^T \end{cases} \quad (5.8)$$

Then each parity bit can be represented with a column vector using the parity bit definitions in equation (5.7):

$$\begin{cases} \rho_1 = [0\ 1\ 1\ 1]^T \\ \rho_2 = [1\ 0\ 1\ 1]^T \\ \rho_3 = [1\ 1\ 0\ 1]^T \end{cases} \quad (5.9)$$

Thus the generator matrix can be created by arranging the column vectors from the previous steps into a 4×7 matrix such that the columns are ordered to match their corresponding bits in a codeword. The result is represented by $G = [\rho_1\ \rho_2\ \rho_3\ z_1\ z_2\ z_3\ z_4]$ and the parity matrix is represented by $\mathbb{P} = [\rho_1\ \rho_2\ \rho_3]$.

To decode the Hamming codes, a 3×7 parity check matrix H can be constructed using $H = [I_3 | \mathbb{P}^T]$, where I_3 is a 3×3 identity matrix. The decoding task can be re-expressed as syndrome decoding as explained in equation (5.2). If the syndrome is zero, it indicates that all three parity checks agree with the corresponding received bits, then the received vector is a codeword, and the most probable decoding is given by reading out its first four bits. In this case z' is supposed to be the same as z . Whereas if the syndrome is non-zero, it is required to correct the bit in the position which matches the column of the parity check matrix that matches the syndrome, which has been explained in Section 5.3.2.

The BER for the Hamming coded operation can be approximated by [211]:

$$P_{ec} = \frac{1}{\bar{\omega}} \sum_{j=t_m+1}^{\bar{\omega}} j \binom{\bar{\omega}}{j} p^j (1-p)^{\bar{\omega}-j} \quad (5.10)$$

where $\bar{\omega}$ is the length of codeword, $t_m = \lfloor (d_m - 1)/2 \rfloor$ is the maximum number of errors that the code can correct, and p is the probability of one bit error. In this case, p is set as the value of the optimised probability of error appropriate to the code rate according to Chapter 4. Encoding the transmission information with (7, 4) and (15, 11) Hamming codes, produces coding rates of 4/7 and 11/15 respectively.

Power consumption for codeword i is $\varphi_i = w_i \varphi_s$, where w_i is the codeword weight and φ_s is the symbol power, which is here normalised to be unity. For different source distributions, information per codeword will be different from an information theoretic point of view. However, for simplicity, it is assumed that each codeword carries $\log(M)$ bits of information, where M is the number of codewords, so M transmitted codewords contain $M \log(M) \varepsilon_d$ bits of information. Here ε_d is the probability that transmitted codeword is correctly decoded and can be given by:

$$\varepsilon_d = \sum_{i=0}^{\lfloor \frac{d_m-1}{2} \rfloor} \binom{\bar{\omega}}{i} p^i (1-p)^{\bar{\omega}-i} \quad (5.11)$$

The average energy per information bit is given by:

$$\eta = \frac{E(\varphi) \times t_s}{\log(M) \varepsilon_d} \quad \text{Joules (bit)}^{-1} \quad (5.12)$$

where $E(\varphi)$ is the expected value of power consumption per codeword and t_s is the symbol duration.

5.4.2 Analytical results

In this section, a (7, 4) Hamming code with three parity check bits and a (15, 11) Hamming code with four parity check bits are both applied to the proposed bacterial communication channel, compared with the uncoded situation. Here, the transmission distance is set as $4\mu\text{m}$. Results are shown in Figure 5.4 and Figure 5.5.

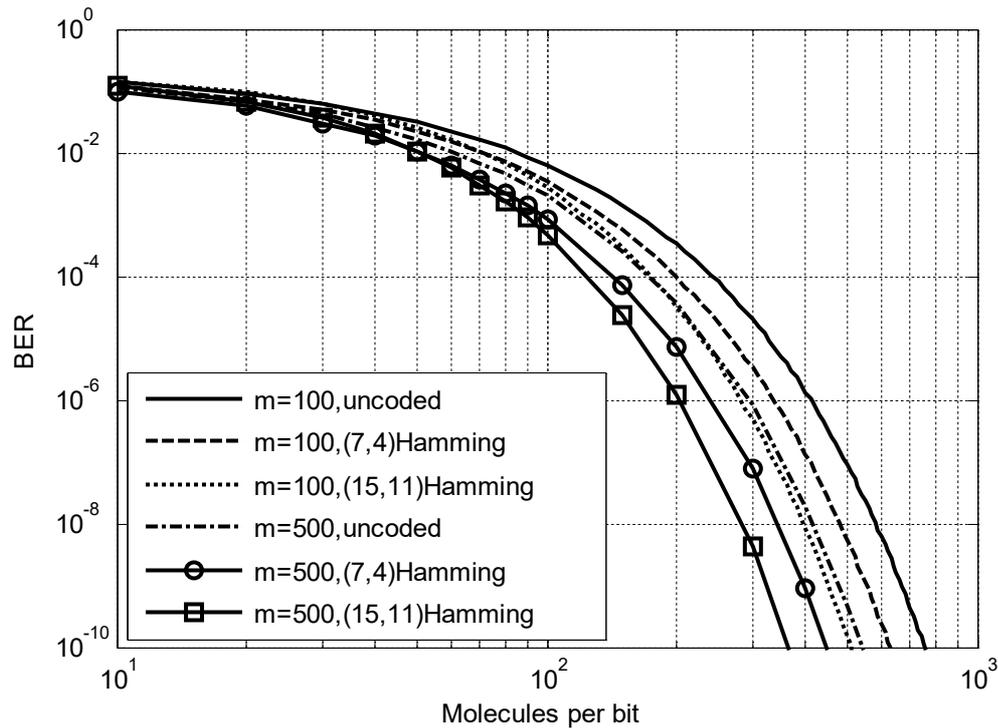


Figure 5.4 BER comparison between (7, 4) and (15, 11) Hamming codes for transmission distance of $4\mu\text{m}$.

Figure 5.4 indicates that at a given value of the number of bacteria in the receiver node m , the performance of the communication system is improved when the information sequence is coded with Hamming (7, 4) code and (15, 11) code. Generally speaking, to evaluate the performance of the system, the coding gain in dB is often assessed at a certain value of BER for a specific code. However, in MC scenario, the coding gain can also be directly obtained by finding out the ratio of molecules needed because there is approximately a linear relationship between the

transmission energy and the number of molecules per bit. When the receiver node contains $m = 100$ bacteria, the number of molecules per bit is 280 for (15, 11) Hamming code, 330 for (7, 4) Hamming code and 410 for uncoded information symbols at a BER level of 10^{-6} which gives a relatively effective error free operation. Thus, it can be derived that the coding gain for (15, 11) Hamming code is:

$$10\log_{10}\left(\frac{410}{280}\right) = 1.65dB \quad (5.13)$$

Also, in this case the coding gain for the (7, 4) Hamming code can be calculated by:

$$10\log_{10}\left(\frac{410}{330}\right) = 0.94dB \quad (5.14)$$

Similarly, when the receiver node contains $m = 500$ bacteria, the number of molecules per bit is 200 for (15, 11) Hamming code, 250 for (7, 4) Hamming code and 300 for uncoded information symbols for a BER of 10^{-6} . Thus, it can be calculated that the coding gain for (15, 11) and (7, 4) Hamming codes are 1.76 dB and 0.79 dB, respectively. Also, it is clear that when the same Hamming codes are applied for the communication channel, the channel which has a receiver node with a larger bacterial population has a better channel performance, with smaller BER values, confirming the results from the channel analysis in Chapter 4.

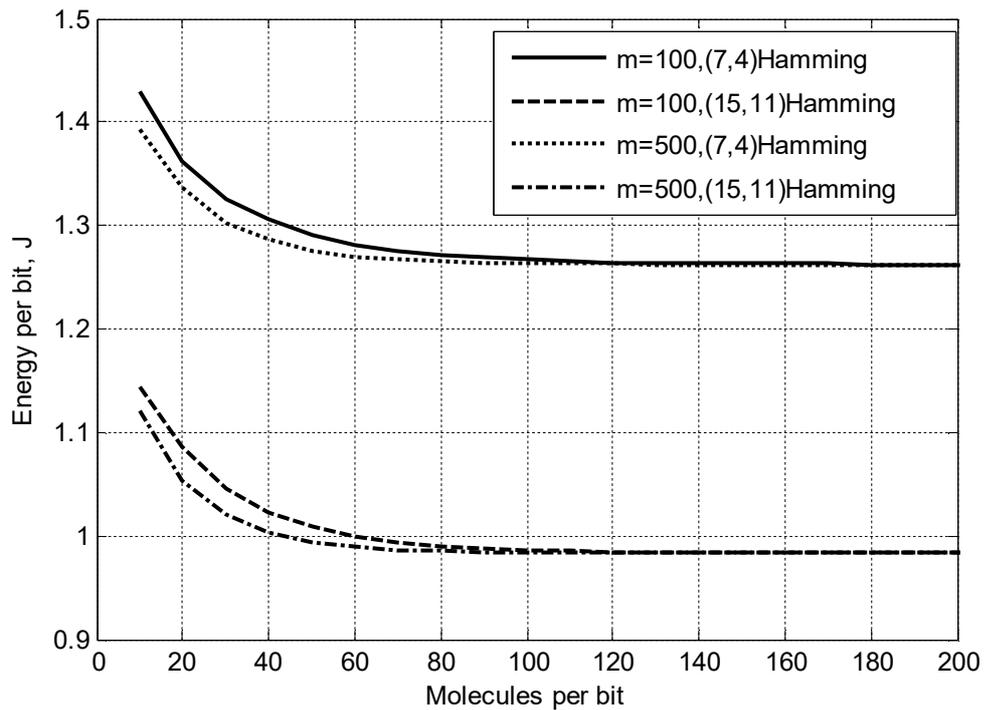


Figure 5.5 Average energy per bit comparison between (7, 4) and (15, 11) Hamming codes for transmission distance of $4\mu m$.

Figure 5.5 shows the average energy per bit comparison between (7, 4) and (15, 11) Hamming codes for a transmission distance of $4\mu m$. It is clear that the energy consumption performance is better with longer codewords. Specifically, the (15, 11) Hamming code exhibits superior average energy per bit values, compared with the (7, 4) Hamming code. For small numbers of molecules per bit (approximately less than 80), extra energy consumption is required to deal with unreliable decoding but this effect levels out as the number of molecules emitted as the start of each time slot increases. In addition, it indicates that when the same Hamming codes are applied, the communication system in which the receiver has a larger bacterial population consumes less average energy per bit. However, when the number of molecules per bit is larger than approximately 100, the consumption approaches the same value.

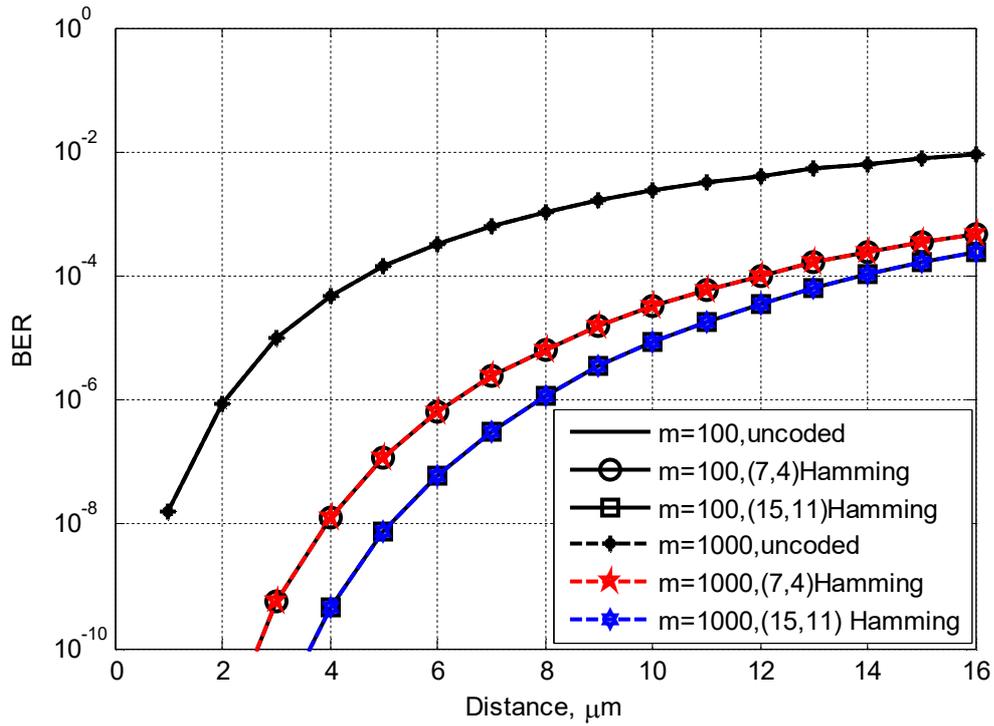


Figure 5.6 BER comparison between (7, 4) and (15, 11) Hamming codes for the number of molecules per bit of 500.

The relationship between error performance and the transmission distance for fixed values of the number of molecules per bit n and the number of bacteria in the receiver node m , is shown in Figure 5.6. This shows that the BER increases with increasing distance between the transmitter and receiver because of the random property of the diffusion process. Also, the figure indicates that when the same coding scheme is applied, the performance curve for the smaller value of m is almost coincidental with that for larger values of m . In addition, at the BER level of 10^{-6} , the transmission distance is $2\mu\text{m}$, $6\mu\text{m}$ and $8\mu\text{m}$ for uncoded, (7, 4) Hamming coded and (15, 11) Hamming coded systems, respectively. In other words, the (15, 11) Hamming code achieves a $6\mu\text{m}$ transmission distance extension, while the (7, 4) Hamming code obtains an additional distance extension of $4\mu\text{m}$. Thus, compared

with a (7, 4) Hamming code, a (15, 11) Hamming code can lead to a reduced BER and an extended transmission distance.

5.5 Minimum Energy Coding

5.5.1 Introduction of MECs

As stated at the start of this chapter, it is essential to develop energy efficient communication techniques for MC. Choosing codewords for each source outcome such that the mean codeword energy is less than any other choice of codeword mappings is called minimum energy coding. A novel, minimum energy coding scheme, which is proved to be energy efficient, is provided in [212] for a THz wireless nanosensor network. Unlike conventional studies, MEC maintains the desired code distance to provide reliability while minimizing energy. In theory, by using OOK modulation, MEC with Hamming distance constraints can reduce energy consumption by minimising the average weight of codewords [212]. In this section, the MEC proposed in [212], which is considered reliable and suitable for nanoscale communications, is used as the channel code to improve the system performance in the proposed diffusion-based MC system, especially the energy performance. With OOK modulation, the average codeword energy is the symbol energy times average codeword weight, which means that average energy can be minimized by minimizing the average code weight. In other words, codewords with a lower weight result in reduced energy consumption since the transmission of a '0' symbol requires less energy than the transmission of a '1' symbol. Therefore, codeword weights and original message-codeword mappings are chosen such that the expected code weight is minimized at the cost of increased codeword length, hence increased delay. The idea of using low weight channel codes together with OOK modulation to reduce energy consumption is first introduced in [213] for sensor networks, in which

codewords are sorted in increasing code weight order and original symbols are assigned in decreasing probability order, such that the most probable source symbol is mapped to the codeword with the smallest weight, resulting to the optimum average code weight. Development of reliable MECs has been an open issue, however it is the first attempt in MC, which has met with some success [214].

The source message, which is of length k , can be encoded into a codeword which is of length ϖ in the following way. For a given set of source symbols, which have a specific source distribution, and a given set of codewords, codewords are sorted in increasing code weight order and source symbols are assigned in decreasing probability, which has been explained above. For example, the least probable source symbol is mapped to the largest weight codeword. In the work of this thesis, it is assumed that each codeword has the same probability of occurrence as the source outcomes to which they are mapped, since no source coding mechanism is applied. For OOK modulation, transmitting bit ‘0’ requires no energy. Thus, as has been stated earlier, the minimisation of energy consumption is equivalent to the minimisation of the average codeword weight. The weight enumerator of a code \mathbb{C} is the polynomial $W_{\mathbb{C}}(\mathfrak{x}) = \sum_i \mathring{A}_i \mathfrak{x}^i$, where \mathring{A}_i is the number of codewords with weight i and \mathfrak{x} is a symbol which is called an indeterminate that does not represent any value. Assuming that M is the number of codewords, d_m is the minimum Hamming distance and p_{max} is the maximum probability in the source probability distribution, the weight enumerators of MEC codewords are given by [212]:

$$W_{\mathbb{C}}(\mathfrak{x}) = \begin{cases} \mathfrak{x}^0 + (M - 1)\mathfrak{x}^{d_m} & p_{max} > 0.5 \\ \mathfrak{x}^{\lfloor \frac{d_m}{2} \rfloor} + (M - 1)\mathfrak{x}^{\lceil \frac{d_m}{2} \rceil} & p_{max} \leq 0.5 \end{cases} \quad (5.15)$$

Also, for a desired code distance d_m , the minimum expected code weight is given by:

$$\min(E(w)) = \begin{cases} (1 - p_{max})d_m, & p_{max} > 0.5 \\ \frac{d_m}{2}, & p_{max} < 0.5, \text{ if } d_m \text{ even} \\ \left\lceil \frac{d_m}{2} \right\rceil - p_{max}, & p_{max} < 0.5, \text{ if } d_m \text{ odd} \end{cases} \quad (5.16)$$

The MEC only provides the limitation of the length of codeword and the minimum weight, rather than the actual codeword. Thus, different codebooks can exist for a single Hamming distance. For MECs, the decoding method is minimum distance decoding which means that the received ϖ -tuple is mapped to the closest codeword in terms of Hamming distance. More errors can be corrected when the minimum Hamming distance increases with the codeword length but this leads to a larger number of error patterns, which will decrease the reliability of the MEC [212]. The minimum codeword length is given by [212]:

$$\varpi_{min} = d_m + (M - 2) \left\lceil \frac{d_m}{2} \right\rceil \quad (5.17)$$

Here, ϖ_{min} is the minimum codeword length required to satisfy the MEC weight enumerator for given values of M and d_m . It is important since it indicates the minimum delay due to transmission of codewords.

Using equations (5.15) and (5.17), sample codebooks for $p_{max} < 0.5$ and $p_{max} > 0.5$ with $d_m = 3$ and $M = 3$ can be generated as:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}; \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \quad (5.18)$$

Generally, with the increase of the minimum Hamming distance between the codewords, more errors can be corrected. However, the length of the codewords will also increase with the code distance, leading to a larger number of error patterns. Thus, increasing code distance does not necessarily increase the channel reliability. Hence, it is required to analyse the error correction capability of MEC. Codes with

distance d_m can correct $\lfloor \frac{d_m-1}{2} \rfloor$ errors, and the channel reliability increases with code distance, since more error patterns can be corrected. Thus after minimum distance decoding of a MEC, the probability that transmitted codeword is correctly decoded is given by:

$$\varepsilon_{MEC} = \sum_{i=0}^{\lfloor \frac{d_m-1}{2} \rfloor} \binom{\varpi_{min}}{i} p^i (1-p)^{\varpi_{min}-i} \quad (5.19)$$

where p is the average bit error probability.

To calculate the energy consumption of MEC, method given in Section 5.4 can be applied with results given in the subsection below.

5.5.2 Analytical results

In this section, the error correction capability and energy efficiency of MECs are investigated via numerical evaluations of analytical parameters in MATLAB. Specifically, MECs are compared with Hamming (7, 4) and (15, 11) codes in terms of BER and energy consumption. MECs satisfy the minimum Hamming distance required by Hamming codes so here this is set to three. An (ϖ, k) code maps 2^k source words into length ϖ channel codewords. Thus, for better comparison, the corresponding MECs are thus $M = 2^4$ and $M = 2^{11}$. Also, the bacterial population in the receiver node is 100. The error correction and energy performances of MECs and Hamming codes over a $4\mu\text{m}$ transmission distance are illustrated in Figure 5.7 and Figure 5.8.

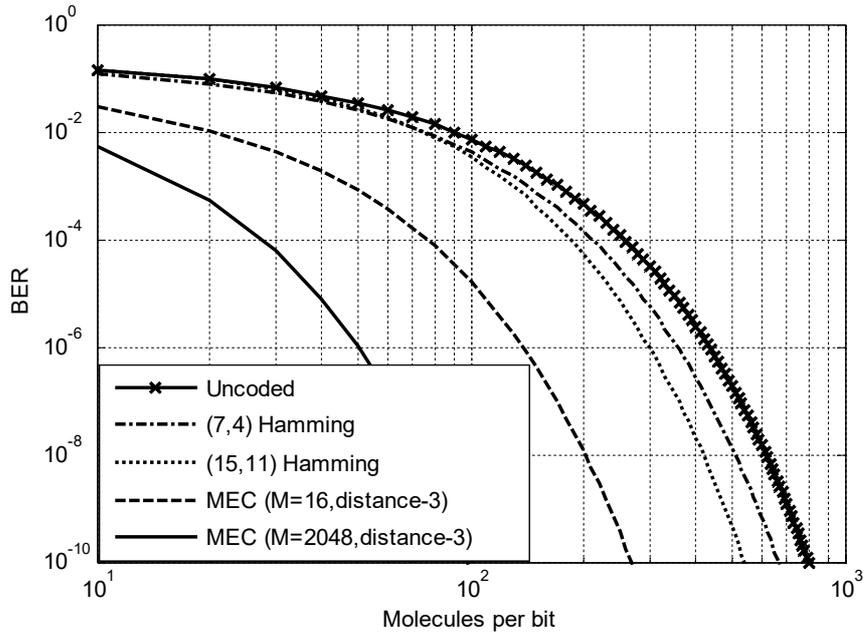


Figure 5.7 BER comparison between MECs and Hamming codes.

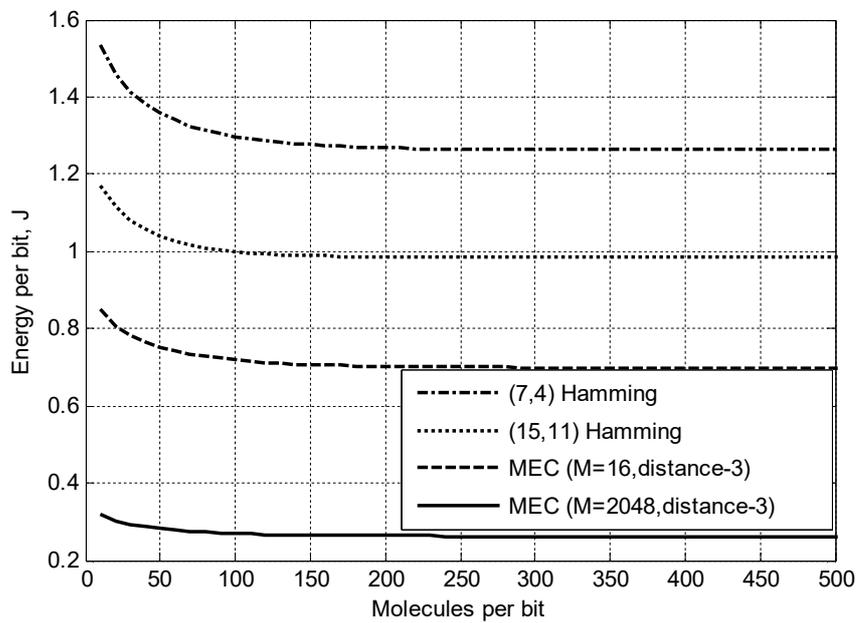


Figure 5.8 Average energy per bit comparison between MECs and Hamming codes.

Figure 5.7 shows that both Hamming Codes and MECs improve system performance for a transmission distance of $4\mu m$. The coding gain can be determined as above by

taking the ratio of the number of molecules for a given BER in the uncoded and coded cases. When the receiver node contains $m = 100$ bacteria, the number of molecules per bit is 280 for (15, 11) Hamming code, 330 for (7, 4) Hamming code, 140 for $M = 2^4$ MEC code, 50 for $M = 2^{11}$ MEC code and 410 for uncoded information symbols for a BER of 10^{-6} which gives a relatively effective error free operation. Thus, it can be derived that at 10^{-6} BER level, the coding gains for the Hamming codes are 0.94 dB and 1.65 dB for the (7, 4) and (15, 11) codes respectively, and for the MECs, the figures are 4.67 dB and 9.14 dB for $M = 2^4$ and $M = 2^{11}$ respectively. In general, MECs have a better performance than Hamming codes with a larger coding gain. Also, the system performance is better with a lengthy codeword. However, for MECs, since increasing the number of codewords means increasing the amount of information to be transmitted, more reliable channels are required to transmit the codewords, which is intuitively expected. Figure 5.8 shows that MECs exhibit superior average energy per bit values. The extra energy needed to deal with unreliable decoding for small numbers of molecules per bit becomes negligible as the number of molecules per bit increases.

5.6 Luby Transform (LT) coding

5.6.1 Introduction of LT codes

In the proposed channel model in this thesis, the symbols from the previous time durations have a relatively strong impact on the received symbol. In previous sections, block codes, including Hamming codes and MECs have been applied to improve the channel performance. For such codes, the code rate should be determined in compliance with the error probability, before transmission. However, in the proposed molecular channel in this work, the error probability is difficult to calculate precisely when the transmission distance is very small and the number of

molecules emitted at the start of each time slot is large, since it is affected by many previous time slots. Thus, in these cases when the error probability is less or more than that expected, it will either cause problems at the decode side or achieve a rate less than the transmission rate achievable. Also, another disadvantage of powerful block codes is the high computational cost of the overall encoding and decoding processes, which is a severe limit for bacterial nanomachines. Taking the above discussions into consideration, fountain codes [215], which are a new class of codes designed and ideally suited for reliable transmission of information over a communication channel with unknown error probability, are utilized as channel codes in the proposed channel model. Fountain codes are rateless codes, which means that the encoder can produce a potentially infinite number of output symbols. In this section, LT codes, as a type of fountain codes, are applied in the channel.

The concept of LT codes was presented and put into practice by Luby [216], to reduce the encoding and decoding complexity. For LT codes, each encoded symbol is obtainable using $O(\sqrt{k} \ln^2\{k/\delta\})$ symbol operations, where k is the number of input symbols and $1-\delta$ is the probability that k input symbols can be recovered [216]. Since as many or as few encoding symbols can be generated as needed and any data copy can be recovered from any set of encoding symbols that is slightly longer than the transmitted symbols in length, encoding symbols can be generated and sent over the communication channel until a certain number of symbols have arrived at the decoder to recover the data, regardless of the channel model [216].

Assuming that the message bits are denoted by $Z = (z_1, z_2, \dots, z_k)$ and the codeword is denoted by $\mathbb{C} = (\zeta_1, \zeta_2, \dots, \zeta_{\varpi})$ where ϖ is the codeword length, the LT encoding process can be expressed as follows:

- Randomly choose a degree ℓ of the encoding symbol from a degree distribution $\Phi(\ell)$. Suitable choice of the degree distribution depends on the message length.
- Uniformly randomly choose ℓ distinct message bits as neighbours of the encoding symbol, bitwise sum these bits modulo 2 and set the result as the value of the encoding symbol.

The encoding scheme mentioned above for LT codes is identical for any LT code design and can be associated with many kinds of decoding schemes, such as Gaussian Elimination, ML Detection, or Belief Propagation. For each kind of degree distribution, a specific decoding algorithm is chosen in terms of fast and accurate decoding. The encoded bit is then transmitted over the noisy channel, and the decoder receives a corrupted version of this bit. For a diffusion-based MC system, all the nanomachines can be synchronized through bio-inspired approaches [217]. Thus they share a common random number generator, which means that the decoder knows which ℓ bits are used to generate any given encoded bit, but not their values. Using the Belief Propagation algorithm [218], which is a simple and efficient method of decoding LT coded messages, the decoding process can be described as follows:

- Find a codeword $\zeta_i \in \mathbb{C}$ connected to only one original $z_j \in Z$. If there is no such a codeword, suspend the decoding process and report the decoding failure.
- Set the value of ζ_i to z_j : $z_j = \zeta_i$.
- Find all the codewords ζ_Λ ($\Lambda \in [1, \varpi]$) that are connected to z_j , update the codeword by: $\zeta'_\Lambda = \zeta_\Lambda \oplus z_j$.

- Remove the connections from the generator matrix.
- Repeat the procedure above until no more codewords have the degree of 1.

The key factor in the LT coding process is the LT degree distribution $\Phi(\ell)$, which refers to the probability that an encoding symbol has degree ℓ . Designing proper degree distributions is of great importance since it both affects the encoding and decoding costs and the overhead. The random procedure of the LT process is completely determined by $\Phi(\ell)$, the number of encoded symbols ϖ and the number of information symbols k . Here, the Robust soliton distribution, which has been proved to have a good performance by Luby [216], is applied as the degree distribution. Before going through the Robust Soliton distribution, the Ideal soliton distribution is introduced at first. It is defined as:

$$\xi(\ell) = \begin{cases} \frac{1}{\ell} & \ell = 1 \\ \frac{1}{\ell(\ell-1)} & \ell = 2, \dots, k \end{cases} \quad (5.20)$$

To obtain the Robust Soliton distribution $\mu(\ell)$, an auxiliary function $o(\ell)$ is defined:

$$o(\ell) = \begin{cases} \mathcal{R}/\ell k & \text{for } \ell = 1, \dots, k/\mathcal{R} - 1 \\ \mathcal{R} \ln(\mathcal{R}/\delta) & \text{for } \ell = k/\mathcal{R} \\ 0 & \text{for } \ell = k/\mathcal{R} + 1, \dots, k \end{cases} \quad (5.21)$$

Then:

$$\mu(\ell) = \frac{\xi(\ell) + o(\ell)}{\sum_{\ell=1}^k [\xi(\ell) + o(\ell)]} \quad (5.22)$$

where δ is the allowable failure probability of the decoder to recover the data for a given number of encoding symbols and $\mathcal{R} = c \ln(k/\delta)$ for some suitable constant $c > 0$.

In addition to the BER, the energy efficiency of the code is also of interest. The power consumption for codeword i is $\varphi_i = w_i \varphi_s$, where w_i is the codeword weight and φ_s is the symbol power, which is here normalised to be unity as previous sections. However, it is assumed that each codeword carries $\log(\varpi)$ bits of information, which is different from previous sections.

5.6.2 Analytical results

In a similar way to the analysis of MECs, LT codes are compared with Hamming (7, 4) and (15, 11) codes in terms of BER and energy consumption as well. LT codes satisfy the minimum Hamming distance required by Hamming codes so here the respective LT information bit lengths are set to be 4 and 11. Furthermore, δ is 0.2 as chosen in [186] and the number of encoding symbols for LT codes is $\varpi = k\{\sum_{i=1}^k \xi(i) + o(i)\}$. The error correction and energy performances of LT codes and Hamming codes over a 4 μ m transmission distance are illustrated in Figure 5.9 and Figure 5.10.

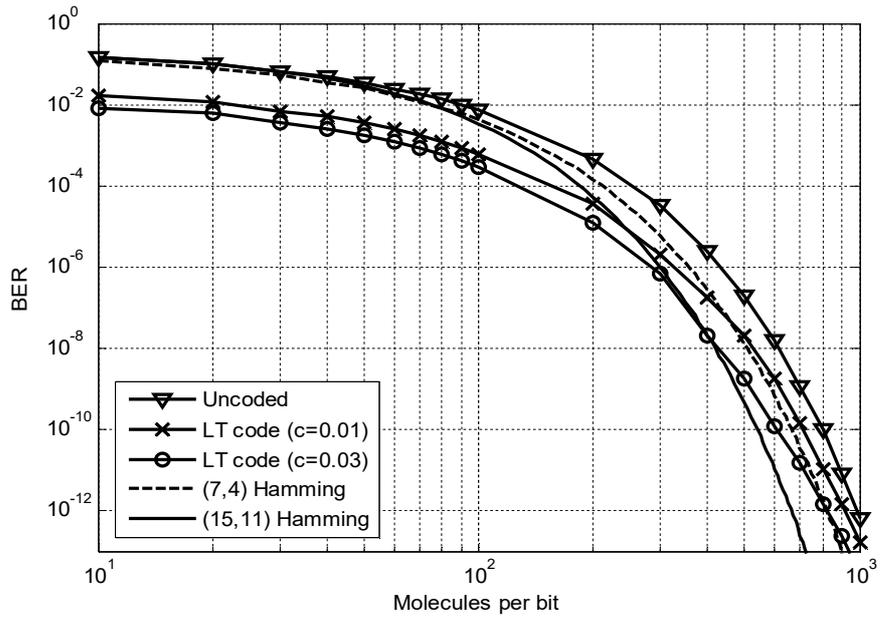


Figure 5.9 BER comparison between LT codes and Hamming codes.

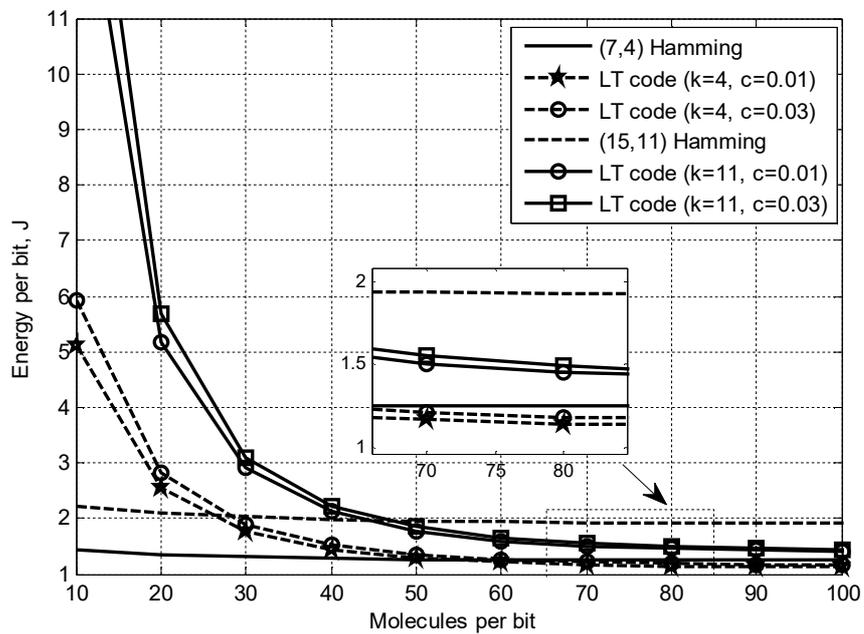


Figure 5.10 Average energy per bit comparison between LT codes and Hamming codes.

Figure 5.9 shows that although system reliability is improved by using both Hamming Codes and LT codes, LT codes have a better performance when the

number of molecules per bit is smaller than approximately 400, with a smaller BER. Also, for LT codes, larger values of the constant c achieve a smaller BER. It should be noted that for LT codes, the BER performances are almost the same for different values of k when the same value of c is applied, thus are not plotted in Figure 5.9. The coding gain can again be determined by taking the ratio of the number of molecules for a given BER in the uncoded and coded cases. Thus at the 10^{-6} BER level, the coding gains for the Hamming codes are 0.94 dB and 1.65 dB for the (7, 4) and (15, 11) codes respectively, and for the LT codes, the figures are 1.37 dB and 2.04 dB for $c = 0.01$ and $c = 0.03$ respectively. In general, the system performance is better with a lengthy codeword. Figure 5.10 shows that LT codes consume more energy per bit values on average when the number of molecules per bit is smaller than approximately 60. The overall shape of the curves conforms to those seen with the other codes. For small numbers of molecules per bit extra energy is needed to deal with unreliable decoding but this effect levels out as the number of molecules per bit increases.

5.7 Summary

Based on the bacterial communication model proposed in Chapter 4, it has been indicated that the diffusion-based MC channel depends on free molecular diffusion to transport the information symbols from the transmitter to the receiver. To obtain the channel reliability, the BER is required and this is strongly affected by channel noise caused by ISI. Moreover, due to the limited storage capabilities and output power range of the bio-nanomachines, energy consumption is considered as a key factor as well. An error correction coding mechanism is considered to be an effective way to enhance the channel reliability and reduce the energy consumption. In this chapter, an overview of the existing related work has been presented, followed by

brief summaries of the essential features of FEC techniques. Three different categories of error correction codes, Hamming code, MEC and LT code, with OOK modulation, have been developed and applied as channel codes. The proposed MEC coding achieves the minimum codeword energy and guarantees a minimum Hamming distance at the price of lengthy codewords. Also, MECs are developed by keeping the codeword distance unconstrained but with a minimum average code weight. The LT codes are also considered as rateless codes, providing an endless stream of codewords for decoding. The parameters chosen conform to the methods used in the existing literature. Thus the results are believed to be representative of the performance of the coding schemes in the field. Two key factors, viz. BER and average energy consumption per bit, have been investigated. The results show that all the three kinds of codes offer coding gains which can be several dBs. Specifically, when the number of bacteria in the receiver node is 100 and the transmission distance is $4\mu\text{m}$, at the 10^{-6} BER level the coding gains for the Hamming codes are 0.94 dB and 1.65 dB for the (7, 4) and (15, 11) codes respectively; for the MECs, the figures are 4.67 dB and 9.14 dB for $M = 2^4$ and $M = 2^{11}$ respectively, indicating that MEC outperforms Hamming codes and that (15, 11) Hamming codes perform better than (7, 4) Hamming codes, with a larger coding gain. For the LT codes, the coding gains are 1.37 dB and 2.04 dB for $c = 0.01$ and $c = 0.03$ respectively. In addition, among the three proposed coding schemes, LT codes consume more energy than Hamming codes and MECs, especially when the number of molecules per bit is smaller than approximately 60, whilst MECs exhibit the lowest mean energy consumption per bit. However, it should be noted that the energy analysis is approximate, since it must rely on estimates of the energy dissipation in the nano processing units as no complete bio-node architecture is yet available.

CHAPTER 6.

ARQ IN BACTERIAL COMMUNICATIONS

6.1 Introduction

According to the previous chapters, it is obvious that the QS process enables cell-to-cell communication between bacteria, which is mediated by small diffusible signalling molecules that trigger changes in gene expression in response to fluctuations in population density. It controls diverse phenotypes in numerous gram-positive and gram-negative bacteria. It has been considered as one of the most promising paradigms of MC, which uses the presence or absence of selected types of molecules to digitally encode information messages. In Chapter 4, a diffusion-based MC system between two populations of bacteria has been established and analysed, using a well-studied luminescence marine bacteria species, *V. fischeri*, in which the phenomenon of QS was first discovered. The transmitted information is represented by the concentration of signalling molecules, which will then be encoded into data frames using OOK modulation techniques, with the release of molecules represented by a binary '1' and no release represented by a binary '0'.

Generally, physical channels are never completely reliable due to many kinds of imperfections, such as queue overflows at switches because of congestion, repeated

collisions over shared media and routing failures. Moreover, packets can arrive out-of-sequence at the destination due to the fact that different packets sent in sequence take different paths or that some switches or routers rearrange packets by some chance [219]. For example, they may experience variable delays, in particular whenever they encounter a queue. In some cases, duplicate packets may even be delivered over an unreliable underlying network [220]. To achieve a reliable communication system, it is necessary to have a reliable data transfer protocol to ensure delivery of all packets in the right order and to enable the receiver to receive each packet only once. For many applications, including web page downloads, file transfers, and interactive terminal sessions, both the transmitter and receiver would like to receive a data stream that conforms to particular protocols and data types, exactly once and in-order [221]. To counter problems including packet losses, reordered packets and duplicate packets in transit through the network, it is desirable to use reliable transport protocols, which can provide a reliable packet stream. For the proposed diffusion-based QS communication network in this thesis, errors are mainly caused by ISI, which results from molecular diffusion and possibly information transmission rate, leading to data packet corruption and out-of-sequence delivery, making it necessary to apply error detection rules and ARQ mechanisms for reliable transmission. As stated in the previous chapter, error control techniques of communication can be divided into two basic categories: ARQ and FEC. In Chapter 5, FEC techniques, including Hamming coding, MEC and LT coding, have been applied to improve the channel performance and enhance reliability. However, FEC cannot solve the problem of out-of-sequence delivery. Thus, in this chapter, ARQ protocols are used rather than error correction schemes because that they can ensure in-order transmission, moreover, error detection requires much simpler

decoding operations than does error correction [222]. The term ARQ was first introduced by Chang [223], after which three widely used ARQ schemes, SW-ARQ, GBN-ARQ and SR-ARQ, have been presented and developed [224]. This work is the first attempt to apply ARQ protocols in a bacterial communication system. It can be helpful to develop more sensitive biosensors, drug delivery systems and water toxicity detection mechanisms. In this chapter, the performance of the three kinds of ARQ schemes will be considered separately and then compared for bacterial communications. In theory, SW-ARQ will solve the problem in the simplest way, but will do so somewhat inefficiently. The GBN-ARQ and SR-ARQ protocols, which are also called sliding window protocols [225], are believed to significantly improve the performance.

This chapter is organized as follows: in Section 6.2, a brief review of the existing transmission protocols for MC is given, followed by the detailed description of the ARQ mechanisms, including the illustration of both error detection codes and the three basic ARQ techniques, in Section 6.3. Then the simulation results for SW-ARQ, GBN-ARQ and SR-ARQ schemes are given in Section 6.4, 6.5 and 6.6, respectively, followed by the comparisons of the three ARQ schemes in Section 6.7. Finally, a summary of this chapter is given in Section 6.8.

6.2 Existing transmission protocols for MC

Generally, a communication protocol defines the rules and standards for sending data packets from one node to another in a network, which is normally defined in a layered manner, exhibiting specific properties, including access mechanisms to physical communication interfaces, encoding and addressing mechanisms, error detection/correction techniques, and routing of packets between connected nodes [226]. Various protocols have been used extensively in communication networks and

have generally been described using a layered architecture known as the OSI reference model, which is a conceptual framework that characterizes and standardizes the communication functions of a communication system, irrespective of their internal structure and mechanisms [227]. In the OSI model, the functions and standardization of the communication system are divided into seven abstraction layers: application layer, presentation layer, session layer, transport layer, network layer, data link layer and physical layer [228]. Conventional protocols include HDLC (High-Level Data Link Control), MAC (Medium Access Control) and ARP (address resolution protocol) in data link layer; IP (Internet Protocol) and ICMP (Internet Control Message Protocol) in the network layer; UDP (User Datagram Protocol) and TCP (Transmission Control Protocol) in the transport layer [226]. Later in this chapter, the ARQ protocol, which is a data link layer protocol in conventional networks, will be applied in the diffusive communication channel. Here, it should be noted that this approach assumes that MC networks will employ a stacked network architecture, which has been displayed in Chapter 2, to support the heterogeneity which is inherent in MC physical layers. The protocol can be implemented using biological computation techniques and components from synthetic and molecular biology research, and existing models and components from this domain which will be discussed in Chapter 7. According to Section 2.5.4, the molecular link layer, which is the second layer in the proposed molecular network architecture model, is implemented as a method of establishing and maintaining reliable links between different nodes or devices on a network using existing physical connections, performing the functions, including data sequencing, data routing, flow control and error control, for the efficient and error free transmission of

data. It provides a well-defined service interface to the network layer and also determines the way in which the bits are grouped into frames for transmission.

Most of the recent research on molecular or bacterial quorum communications has been focused on modelling the communication channel as well as the modulation and coding schemes which are necessary for reliable transmission. Based on these established communication links, it is considered possible to use layered architecture similar with the conventional one over a bacterial communication physical link. The IEEE Standards Association has started the IEEE P1906.1 project to provide recommended standards for nanoscale and molecular communication frameworks [229], to assist research in related fields.

It has been stated that the data link layer protocol can enhance the reliability of the communication channel. Some relevant research exists which concentrates on designing protocols to enhance the reliability of diffusion-based MC in aqueous environments [230-233]. In [230] and [231], the authors present an investigation on feedback based rate control schemes for molecular propagation in diffusive transports. The proposed schemes effectively prevent the transmitter from transmitting the molecules faster than the receiver can react when delivering a number of information molecules to the receiver. However, the issue of in-sequence delivery is not taken into consideration. Instead, in-sequence molecular transmission is studied with SW-ARQ schemes for diffusion-based MC in [232] and bacterial communication in [233], which is useful to further protocol design in related fields. In [234], a concentration aware (CA) routing protocol in MC networks is proposed, which utilizes the concentration gradient of diffused molecules to establish the transmission delay path in multi-hop nanonetworks. In addition, in communication scenarios with multiple transmitters and multiple receivers, a key property is an

addressing mechanism to provide the ability to send information to a specific bio-nanomachine connected to the network. For routing protocols using a specific molecule for each address in a multiple access network, the transmitter must have the capability to produce the distinct molecules for each address. The authors in [235] proposed a routing protocol NAODV (Nano Ad hoc On-Demand Distance Vector), which integrates the nanotechnology method in routing and shows improved network performance for a collision based molecular network system. Also, in nanoscale communications, the time duration is important for the channel reliability. In [236], an adaptive transmission protocol, termed the Adaptive Time-slot Protocol, has been proposed, which adapts the bit transmission duration for a MC system based on calcium signalling, dependent on the type and level of tissue deformation caused by the application of force. The protocol includes two stages: the first uses information metrics to infer the type of compression on the tissue, and the second which adapts the time-slot duration for the bit transmission. Results show that there is an average increase in performance of 15% for the proposed protocol compared to a static time-slot protocol. Furthermore, the physical link distance is a common constraint for reliable communication in nanonetworks and could be important for routing issues in multi-path networks. Several protocols for distance measurement have been proposed, analytically modelled and compared. For example, in [237], a relative coordinate system is established which uses distance measurements to guide molecular carriers capable of active transport to a target location. In addition, since genetically modified bacteria can play an important role in MC, there is a considerable body of work focused on network architectures using such bacteria. For example, in [43], an opportunistic routing process for multi-hop bacteria communication networks is proposed based on two inherent properties of bacteria,

including chemotaxis and conjugation. A series of simulations have been performed to validate the solution for three different types of topologies, termed grid, hexagon and T-shape [43]. Also, nucleic acid molecules can also provide a means of biological addressing. For example, in [238], single-stranded DNAs (ssDNAs) are used, which specify the receiver address, attached to message molecules. The receiver exposes a complimentary DNA molecule designed to attach to the respective ssDNA on the information molecule.

6.3 ARQ mechanisms

6.3.1 Overview

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 [239]. In terms of the OSI reference model for layered network architectures [240], an ARQ protocol is usually located at the data link layer, which takes the packets it gets from the network layer and encapsulates them into frames for transmission [240]. Each frame contains a frame header, a payload field for holding the packet, and a frame trailer which is shown in Figure 6.1. ARQ is an error control technique to ensure that a data stream is delivered accurately to the user despite errors in transmission; it also 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 positive acknowledgement frame (ACK) to indicate correct reception of a packet, the negative acknowledgement (NAK) to indicate error in a packet, and the timeout mechanism. The ACK/NAK signifies the receipt of a given frame. The timeout mechanism is required to maintain the flow of frames.

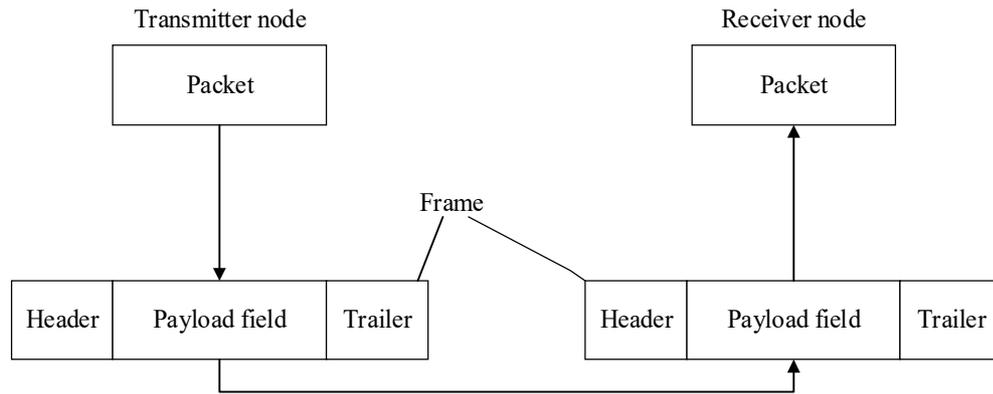


Figure 6.1 Relationship between packets and frames in ARQ.

According to Chapter 3, it is clear that a population of bacteria is able to perform distributed information processing, where each bacterium in the colony is capable of information storage, and the processing and interpreting of information [112]. Also, the advance of synthetic biology, particularly the foundation of the BioBricks database [148], enables many types of capabilities based on genetically engineered bacteria, including timing, counting, clocking, logic gates, pattern detection and intercellular communication [241]. Moreover, because ARQ mechanisms and error control operations can be implemented through circuits [222], it is also possible to implement both techniques in bacterial communication systems. In addition, in some specific communication scenarios, genetically engineered bio entities can harvest energy from biological systems and require no external energy sources, which is therefore expected to be energy efficient [242]. Such capabilities greatly improve the bacterial cooperation level during the communication process, making it possible to meet the requirements of high processing and memory intensive ARQ protocols. Generally speaking, the SW-ARQ protocol, in which the transmitter node sends a frame to the receiver and waits for its acknowledgement, has been investigated [243] and shown to suffer from inefficiency because 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 [244]. In this chapter, the three basic ARQ schemes will be analysed separately and then compared with each other.

Due to the noise in the communication channel, error detection techniques will be used in the model here to improve the error-rate performance. There are three kinds of error detection codes, parity check codes, the Internet checksum and polynomial codes which are also known as cyclic redundancy check (CRC) codes. The choice of method is an open question for nanoscale communication systems but should be energy efficient given the limited energy storage capabilities of nanoscale devices, since they can harvest energy from their surroundings [245]. In this work, CRC is adopted for error checking, which is widely used in data communication systems. In the next few subsections, the mechanisms of CRC, as well as ARQ protocols, will be described.

6.3.2 Cyclic redundancy check (CRC) codes

CRC is a technique for detecting errors in digital communication networks but not for making corrections when errors are detected. Before data transmission, a certain number of check bits, also called a checksum, are appended to the message being transmitted. CRC is often used cooperatively with ARQ. Specifically, when the CRC encoded data packets arrive at the destination, the receiver can determine whether or not an error has occurred in transmission, by checking if the check bits agree with the data. If an error occurs, the receiver sends a NAK back to the transmitter, requesting that the message be retransmitted. CRC has good error sensing performance, fast encoding and decoding capabilities and applicability to varying message lengths [246]. In CRC, the information symbols, the codewords and the error vectors are represented by polynomials with binary coefficients [247]. CRC is

specified by its generator polynomial $g(x)$, which plays the role of the generator matrix for linear block codes, which has been described in Section 5.3.2. Table 6.1 gives some generator polynomials that have been endorsed in a number of standards.

Table 6.1 Standard generator polynomials [240].

Name	Polynomial	Used in
CRC-8	$x^8 + x^2 + x + 1$	ATM header error check
CRC-10	$x^{10} + x^9 + x^5 + x^4 + x + 1$	ATM AAL CRC
CRC-12	$x^{12} + x^{11} + x^3 + x^2 + x + 1$	Bisync
CRC-16	$x^{16} + x^{15} + x^2 + x + 1$	Bisync
CCITT-32	$x^{32} + x^{26} + x^{23} + x^{22} + x^{16} + x^{12} + x^{11} + x^{10} + x^8 + x^7 + x^5 + x^4 + x^2 + x + 1$	IEEE802, DoD, V.42, AAL5

CRC is a typical kind of cyclic code, which has been briefly introduced in Section 5.3.3. It is based on polynomial arithmetic, in particular, on computing the remainder of dividing one polynomial by another [248]. The calculation processes, including addition and subtraction, are performed modulo 2, which is the same as the exclusive-or operator. The coding process of CRC is as follows:

- Generate the message polynomial $z(x)$ using the message bit stream, which is about to transmit over the medium, as the increasing order of the power of x . For example, the bit sequence 10101010 can be represented by $x^7 + x^5 + x^3 + x^1$.
- Take the predefined CRC polynomial $g(x)$ as a polynomial of x . Here, the simple CRC-3 polynomial $x^3 + x^2 + 1$ will be taken as an example.
- Take the highest degree term of x from the CRC polynomial $g(x)$ as x^i . Here, according to the example in the above steps, it should be x^3 .
- Multiply the message polynomial $z(x)$ by the x^i . Here, it should be $\lambda(x) = x^{10} + x^8 + x^6 + x^4$.

- Divide the product of $x^i z(x)$ by $g(x)$, and take the remainder data as a polynomial of x as $\vartheta(x)$. The quotient polynomial is discarded. Here, the remainder is calculated as $\vartheta(x) = x^2 + x^1$.
- Convert the remainder polynomial to the corresponding binary bits by placing a one where the respective power coefficient of x occurs and zero otherwise. Here, the result of CRC is 110 according to the polynomial representation of $\vartheta(x)$.
- Append the CRC bits calculated in the previous step to the message data to obtain the encoded data, which will be transmitted over the channel. Here, the encoder output message is (10101010 110).
- At the receiver, the same CRC polynomial is used to decode the message in the same process as encoding. Specifically, upon reception of the transmitted codeword, the remainder can be obtained by dividing the polynomial for the received codeword $v(x)$ by $g(x)$.
- Check the remainder from the previous step. If it is a polynomial of x as zero coefficient of each x term, it means that there is no error occurs during transmission. Otherwise, it indicates that an error has occurred in transmission and the received word is not a valid codeword.

However, it should be noted that not all error patterns can be detected by CRC codes. In this chapter, the investigations are mainly based on simulations, instead of mathematical derivations. So it is unnecessary to consider which kind of errors can be detected here. Thus the error detection capabilities will be explained in detail in Chapter 7.

6.3.3 ARQ implementation

ARQ is a protocol for error control in data transmission, which has been used extensively in communication networks. The general concept of ARQ is to detect frames with errors at the receiver and then to request the transmitter to repeat the information in those erroneous frames. Typically, the protocol calls for the receiver to send back special control frames bearing positive or negative acknowledgements about the incoming frames. If the sender receives an ACK about a frame, it indicates that the frame has arrived correctly and safely. On the other hand, a NAK means that something has gone wrong and the frame must be retransmitted. The block diagram of a basic ARQ system is shown in Figure 6.2.

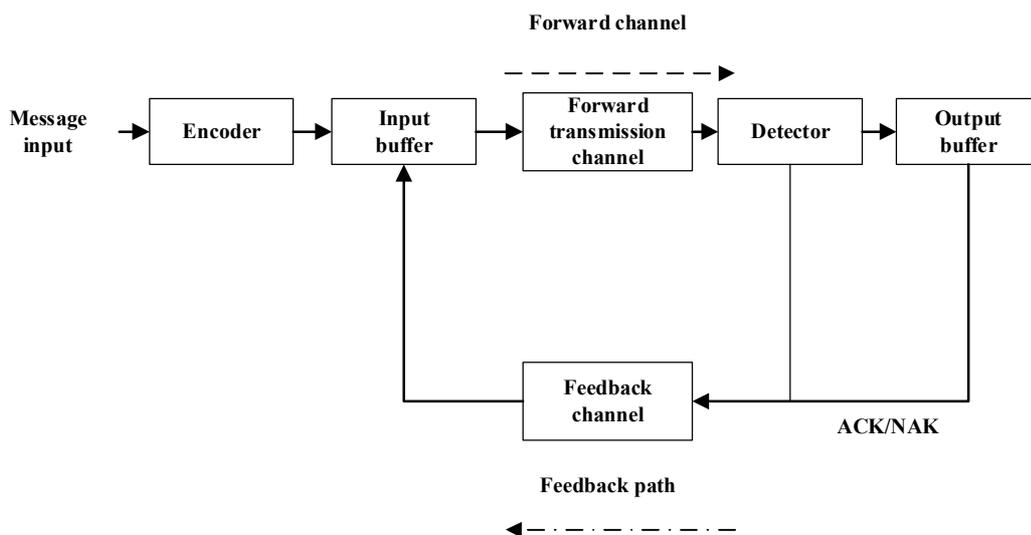


Figure 6.2 Block diagram of the basic ARQ system.

The operation process of ARQ mechanisms is described as follows. The encoder produces codewords for each information message sequence as its input. At the output of the encoder, each codeword is stored temporarily at the input buffer, and then transmitted over the forward transmission channel. After arriving at the destination, the decoder will check if errors occur in the received codewords. The decoder will output an ACK frame if no errors are detected, otherwise, the output

will be a NAK frame. When the controller at the encoder side in Figure 6.2 receives a NAK signal via the feedback channel, it will retransmit the appropriate codeword from the codewords stored in the input buffer. It should be noted here that a particular codeword may be transmitted more than once. Also, the output buffer and controller at the receiver side are capable of assembling the output data stream from the codewords accepted by the decoder.

For a typical transmission channel, the transmitter node generates a sequence of information frames for transmission. The ARQ mechanism requires the frame to contain a header with control information that is essential for proper operations. In addition, CRC check bits will also be appended in the frame to determine if an error occurs during transmission. Figure 6.3 shows the basic elements of ARQ protocols [225]. It contains the information frames that transfer the information packets, the acknowledgement frames, which signify the receipt of a given frame, and the timeout mechanisms. In this model, 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. This process is particularly needed since time varying channel conditions with high BER cause frequent packet corruptions in communication networks [239]. To date, three versions of ARQ have been standardized, including SW-ARQ, GBN-ARQ and SR-ARQ [240], which will be introduced in detail in the following subsections.

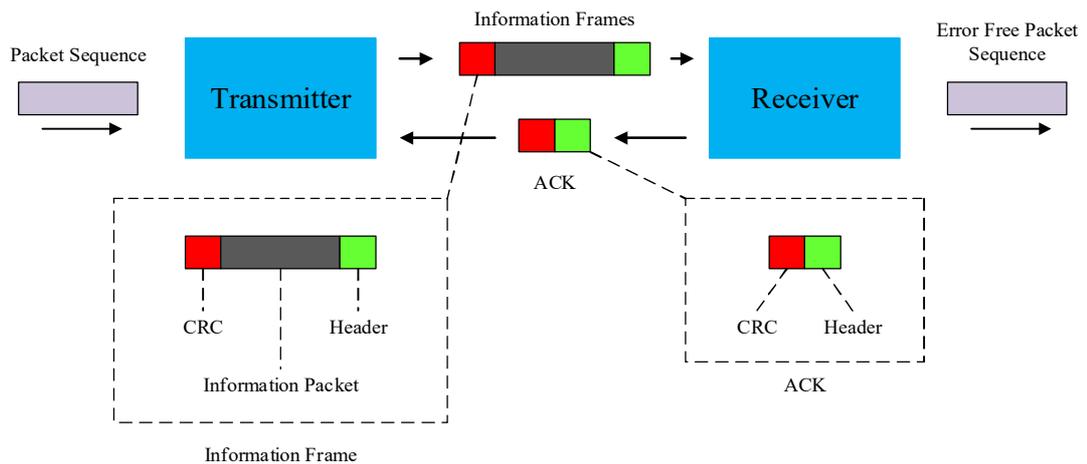


Figure 6.3 Basic elements of ARQ [225].

The data link layer encapsulates the packet, which is passed by the network layer, into an information frame by adding a header containing the sequence number of the frame and a trailer containing the CRC code. Upon arriving the data link layer at the receiver side, the data link layer extracts the sequence number from the header and the CRC code from the trailer. If no error is detected, the information frame is passed to the network layer and an acknowledgement frame is transmitted to the sender.

A. *SW-ARQ*

The simplest type of retransmission protocol is called SW-ARQ. In the SW-ARQ technique, after transmitting a frame, the transmitter waits for an ACK/NAK from the receiver before transmitting the next frame. No other frames can be sent until the acknowledgement frame arrives at the transmitter. There exist two sorts of errors which could occur during transmission process. First, the frames that arrive at the destination could be corrupted or damaged. In this case, 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. Otherwise, a NAK is transmitted back to the transmitter and the received erroneous frame is discarded. For the transmitter, if an ACK is received, it transmits the next frame. Otherwise, it

will retransmit the previously sent frame. It has been stated earlier that the transmitter should be equipped with a timer to maintain the flow of frames. Specifically, after a frame is transmitted, the transmitter will wait for a corresponding acknowledgement. If none is received by the time that the timer expires, the same frame is released again. Generally, the timer is designed to expire after an interval which is long enough for the frame to reach the destination, be processed there and have the acknowledgement propagate back to the transmitter. Here it should be noted that the timeout mechanism requires that the transmitter maintain a copy of the transmitted frame until an ACK is received for that frame. The second kind of error is a damaged acknowledgement frame. For example, if the ACK/NAK is lost, the transmitter waits for a timeout period until the next transmission. In this case, duplicate frames may be received at the receiver if the frames are not labelled [240]. Thus, to avoid ambiguities, it is generally necessary to assign sequence numbers to outstanding frames, helping the receiver to distinguish retransmissions from originals. Specifically, when a valid frame arrives at the receiver, its sequence number is checked to decide if it is a duplicate. If not, it is accepted, passed to the network layer, and an acknowledgement is generated. Duplicates and damaged frames are discarded. 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. The protocol continues in this manner until all the frames are transmitted successfully. Also, since errors can occur in the feedback channel as well as in forward transmission channel, the ACK/NAK is protected with error detection codes as well. Figure 6.4 shows an example of the process of transmitting a series of frames with ACKs and the timeout mechanism, in terms of lost data frames, damaged frames and lost acknowledgements.

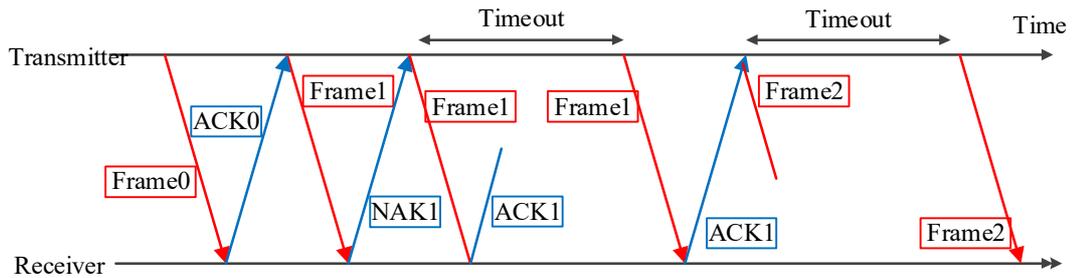


Figure 6.4 Example of SW-ARQ.

Each time the transmitter sends a frame, it starts a timer. At the initial point, frame 0 is transmitted and the corresponding ACK is received correctly. When frame 1 arrives at the receiver, the data link layer there makes error detection and notices that there is an error, so it sends back a NAK for frame 1. Upon receiving NAK1, the transmitter will resend frame 1 immediately and starts a timer. However, the acknowledgement frame ACK1 is lost when propagation in the feedback channel. In this situation, the transmitter will resend frame 1 until the timeout period expires. Eventually, when the transmitter gets the ACK1, frame 2 will be transmitted. If it is lost, it will be retransmitted until the timeout expires.

As discussed above, the ARQ protocol specifies that the transmitted frames are required to be numbered using sequence numbers, which are stored in the header displayed in Figure 6.3. One important consideration is the range of the sequence numbers. Generally, a header that allows λ bits for the sequence number produces sequence numbers in the range $[0, 2^\lambda - 1]$, for example zero to 7 for $\lambda = 3$ with sequence numbers $(0, 1, 2, 3, 4, 5, 6, 7, 0, 1, 2, 3, 4, 5 \dots)$; in other words, the sequence numbers are modulo- 2^λ . Similarly, to distinguish acknowledgements for different frames, the receiver returns the sequence number of the next packet awaited, instead of returning ACK or NAK on the reverse link [240]. This helps to provide all the information of the acknowledgements whilst avoiding ambiguities about which

frame is being acknowledged. An equivalent method would be to return the number of the packet just accepted.

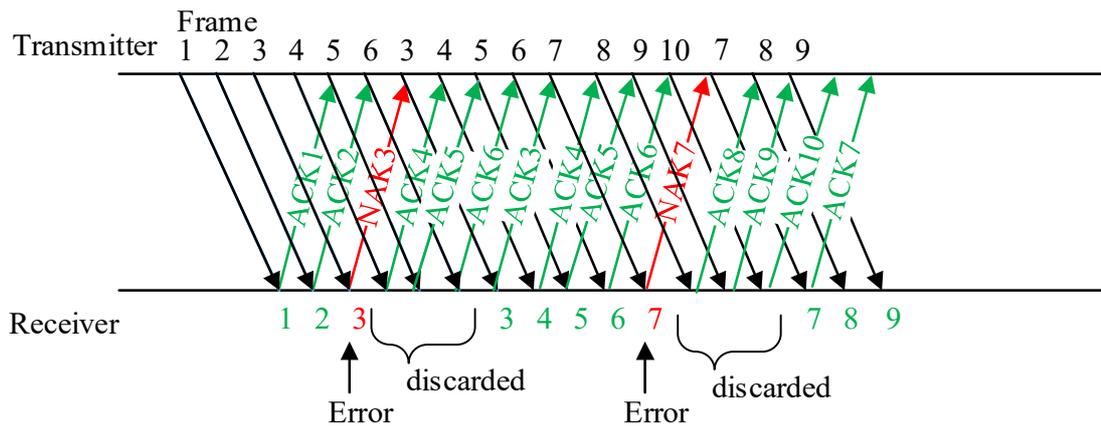
B. *GBN-ARQ*

In GBN-ARQ techniques, the transmitter continuously transmits a block of W (often known as the window size) frames without waiting for the individual frame acknowledgements. Each frame must be buffered (stored) until a valid ACK arrives, in case retransmission is needed. The window size 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 the 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 the W th frame has been transmitted, the transmitter will then discard the first three frames from its buffer and retransmit all the frames from the fourth to the W th. The best situation is when none of the frames in one block are negatively acknowledged and successful transmission of W packets

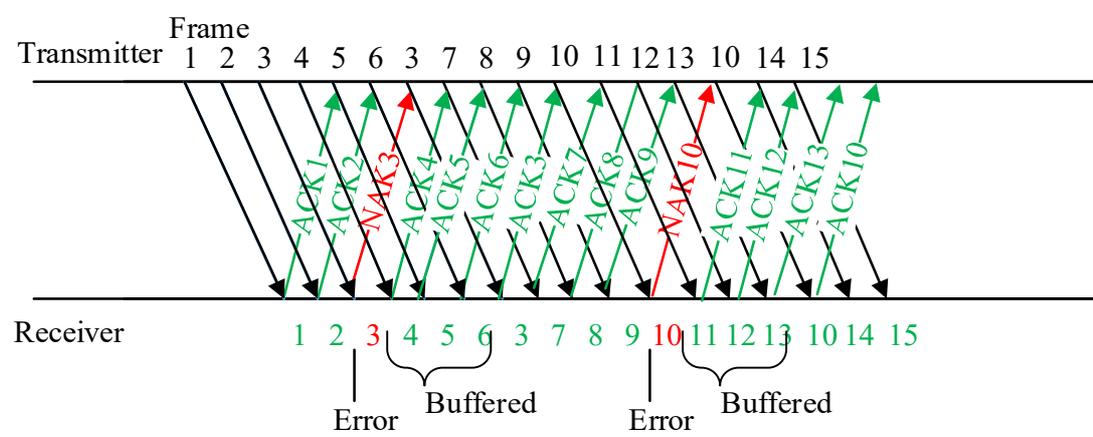
involves a minimum of two propagation delays. Compared with the best situation in SW-ARQ where one frame is involved with two-propagation delays, GBN-ARQ provides, in theory, a throughput improvement.

C. *SR-ARQ*

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 a more complex logic for implementation as the receiver will receive frames out-of-sequence and must have the capability to arrange frames in order. When the transmitter times out, only the earliest unacknowledged frame is retransmitted. If that frame arrives correctly, the receiver will pass to the network layer, in sequence, all the frames it has buffered. SR-ARQ is also often used combined with having the receiver send a NAK when it detects an error, including a parity check error or a frame out-of-sequence. NAKs can stimulate retransmission before the corresponding timer expires and thus can improve the transmission efficiency. Figure 6.5 shows examples of the process of transmitting a series of frames with ACKs and NAKs of GBN-ARQ and SR-ARQ, respectively. Here, it should be noted that if the NAK3 in Figure 6.5 gets lost, the transmitter will time out for frame 3 and send it again, which will be a time consuming process.



(a) Go-Back-6 ARQ



(b) Selective Repeat ARQ

Figure 6.5 Examples of GBN-ARQ and SW-ARQ schemes.

(a) GBN-ARQ with window size of 6: If the 3rd packet comes with an error, then the receiver will accept the 4th packet but sends NAK of 3 to the transmitter and buffer the 4th packet. The receiver repeats this process for the 5th and 6th packet. When the transmitter receives the NAK of 3rd packet, it immediately sends the 3rd packet again. After receiving the 3rd packet, the receiver sends the ACK of 3 and discards the buffered 4th to 6th packets. The transmitter will send the discarded packets again, even though they have been positively acknowledged before the retransmission of 3rd packet arrives. For the 7th packet which comes with an error as well, the operation process is the same. (b) SR-ARQ: The process is similar with that of GBN-ARQ in (a). However, after receiving the retransmitted 3rd packet, the buffered 4th, 5th, and 6th packets will be kept in the receiver buffer, instead of being discarded.

6.3.4 Discussions

The three kinds of ARQ protocols, SW-ARQ, GBN-ARQ and SR-ARQ, are bidirectional protocols that belong to the sliding window protocol family [240]. They are different from each other in terms of efficiency, operation complexity and buffer requirements. As has been discussed in the previous subsections, each outstanding frame contains a sequence number, ranging from 0 to $(2^\lambda - 1)$, fitting the sequence number exactly in an λ -bit field. Generally, the SW-ARQ protocol uses $\lambda = 1$, restricting the sequence number to 0 and 1. But more sophisticated schemes allow an arbitrary λ .

For all sliding window protocols, the essence is that both the transmitter and receiver use a window for each connection at any instant of time. Specifically, the transmitter maintains a set of sequence numbers corresponding to frames that have been sent or can be sent but are as yet not acknowledged [240]. These frames are said to fall within the sending window. When a new packet arrives from the network layer, it is given the next highest sequence number, and the upper edge of the window will slide forward by one. When an ACK comes in, the lower edge is advanced by one. In this way, the sending window continuously maintains a sequence of unacknowledged frames. Also, it should be noted that if the maximum window size is W , the transmitter needs W buffers to hold the unacknowledged frames. Similarly, the receiver also maintains a receiving window corresponding to the set of frames it is permitted to accept. When a frame whose sequence number is equal to the lower edge of the receiving window arrives, it is accepted and passed to the network layer and the window is sliding by 1. The upper edge moves by 1 when an ACK is sent. Any frame whose sequence number falls out of the receiving window is discarded. Both the sending window and the receiving window can slide over the buffer as the

protocol operates. The lower and upper limits and the size of the sending and receiving windows will be discussed later. Figure 6.6 shows an example of sliding window with a maximum window size of 7.



Figure 6.6 Sliding window example.

Both the sending window and receiving window have a maximum window size of 7. At the initially point, 7 frames are outstanding, so the lower and upper edges of both windows are equal. However, as time goes on, the situation progresses as shown. For the sending window, when frame 0 and frame 1 have been sent, the left edge shrinks by 1 for each frame, while when an ACK (ACK2) is received, the right edge moves until the window expands to 7. For the receiving window, when frame 0 and frame 1 have been received, the left edge moves to the right by 1 for each frame. Also, when ACK2 is sent, the right edge moves until the window expands to its maximum size 7. The process continues in this way until all the frames are transmitted and acknowledged.

In the proposed channel model, 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 the CRC codes discussed previously in Section 6.3.2; 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 Chapter 4, it is clear that for the transmitter node, only Type-II receptors are activated, while for the receiver node, only Type-I receptors are enabled. Also, it has been explained that the transmitted information is represented by the concentration of Type-I molecules. Thus the information frame is made up of certain number of information bits, which are encoded by Type-I molecules. Similarly, the acknowledgement frames are composed of fixed number of bits encoded by Type-II molecules. In addition, because the signalling molecules used to encode the acknowledgement frames in the reverse channel (Type-II molecules) are different from autoinducers used to encode the information frames in the forward channel (Type-I 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-I molecules, providing a more accurate decoding of the transmitted information at the receiver; this process can be achieved by using proper enzymes as stated in previous chapters. Therefore, both the transmitter and receiver are able to generate and accept 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 [249].

6.4 Simulation results for SW-ARQ

In this section the simulation results for SW-ARQ protocol are discussed in terms of transmission delay, transmission efficiency, BER and channel throughput. The results show that the major aspects that affect the system performance are transmission distance, number of bacteria in the receiver node, frame length and CRC polynomial.

6.4.1 Parameter setup

Through the simulation of the information transmission, in this section the channel performance results for the SW-ARQ protocol 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. All information frames are supposed to be of the same length. It has been explained in Chapter 4 that the transmitter node implements a 1 bit memory, which is either binary '0' or '1', based on the concentration of emitted autoinducers. Generally, specific biological units could be designed to act as molecule storage, which may simply either be the MC 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 [79]. 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 . For SW-ARQ, the window size is 1. In this channel, only one bit is allowed to transmit in one time slot. Thus, for a sliding window protocol, the basic delay, t_0 , in the absence of errors, which transpires from the moment a frame is transmitted into the channel to the moment when the ACK is confirmed is calculated by:

$$t_0 = 2t_p + 2t_{ps} + t_f + t_{ack} = 2t_p + 2t_{ps} + \frac{n_{cw}}{R_b} + \frac{n_a}{R_b} \quad (6.1)$$

In this equation, the first bit that is input into the channel appears at the output of the channel after a propagation time of t_p , which can be calculated according to Chapter 4; the end of the frame is received at the receiver after t_f additional seconds. 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 [202], which makes it possible to operate CRC coding using genetic circuits. Here, t_{ps} is the processing time for CRC implementation and is ignored for simplification in this chapter, n_{cw} is the total number of bits in the information frame including the information bits, sequence numbers and error checking bits and n_a is the number of bits for the acknowledgement frame which has the same value as the number of sequence bits; R_b is the bit rate of the transmission channel. For sliding window protocols, W frames are allowed to be transmitted continuously without waiting for ACK/NAK for individual frames. Here, for SW-ARQ, W is taken as 1. In this model, there is one bit transmitting in each time slot. The boundary condition to choose R_b is represented by:

$$(n_{cw} \cdot t_p + t_{ps}) \cdot R_b \leq n_{cw} \quad (6.2)$$

This indicates that in the time period of transmitting one codeword which contains n_{cw} bits of information, no more than n_{cw} bits are sent by the transmitter. Here, R_b is chosen as the maximum value according to equation (6.2). 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:

$$R_{\text{eff}}^0 = \frac{\text{bits delivered}}{\text{total time}} = \frac{n_{cw} - n_0}{t_0} \quad (6.3)$$

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:

$$R_{\text{eff}} = \frac{\text{information bits delivered}}{\text{average total time per frame}} = \frac{n_{cw} - n_0}{t_{\text{ave}}} \quad (6.4)$$

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 transmitted (here 2000 bits), and the total time for transmission. It should be noted that during the simulation process, the error probabilities of changing binary '1' to binary '0' and changing binary '0' to binary '1' are different and can be calculated separately according to Chapter 4.

6.4.2 Simulation process

This section describes the simulation process of the SW-ARQ protocol. The simulation begins with the generation of 2000 bits to be transmitted, in which there is an equiprobable distribution of '1's and '0's. The bit stream is then divided into

frames of equal length, which are then appended with sequence numbers and CRC check bits. For the transmitter, there exist two modes: Ready mode and Wait mode. Before data transmission, the transmitter is initialized to Ready mode. Here, two parameters are introduced: S_{las} – the sequence number of the frame which is sent by the transmitter; R_{next} – the sequence number of the frame that is expected by the receiver; both are initialized to zero.

The two modes of the transmitter are described as follows:

- Ready mode: The transmitter prepares the requested information frame with the sequence number S_{las} in the header. Then the transmitter enters the Wait mode, which will be described below, and waits for acknowledgement with sequence number $R_{next} = (S_{las} + 1) \bmod 2$.
- Wait mode: During this stage, the transmitter does not accept packets from the network layer. If an acknowledgement is not received, or is received incorrectly, the transmitter stays in the Wait mode until the timer expires. If the timeout period expires, the information frame with sequence number S_{las} is retransmitted. If a correct acknowledgement frame with sequence number R_{next} is received, then S_{las} is set to R_{next} and the transmitter returns to Ready mode and repeat the process of step 1.

Differently from the transmitter, the receiver is always in the Ready mode. Whenever an information frame arrives, it is checked for errors at the receiver.

- If no error is detected and the received frame number is the expected number, which is $S_{las} = R_{next}$, then the information frame is passed on to the network layer at the receiver side. Also, R_{next} is updated to $(R_{next} + 1) \bmod 2$, and an ACK with sequence number R_{next} is sent.

- If the received information frame has errors, the frame is discarded without other actions.
- If the received information frame has no errors but the sequence number is incorrect, an ACK with sequence number R_{next} is sent.

6.4.3 Results and discussions

The transmission distance is first taken into consideration in Figure 6.7. Here the number of bacteria in the receiver node is 100, the frame length is set as 100 bits, and CRC-8 is employed due to its relatively simple operation compared with higher order CRC polynomials. Compared with the system performance when neither ARQ nor error detection mechanisms is applied, the channel which uses ARQ protocol performs better with a lower BER, achieving a better system reliability. However, when the number of molecules per bit is in a lower range, approximately below 100, it takes longer to transmit the data stream, thus obtaining a lower channel throughput, which perfectly matches the results in Figure 6.7 (d). Also, the time for transmission, transmission efficiency, error probability and channel throughput all improve with smaller values of d . This is because over a larger distance, the bit error probability is higher according to Figure 4.5, which will result to more transmission times per frame on average. Here, it should be noted that although the proposed channel is a BAC, meaning that the crossover probabilities of changing ‘1’ to ‘0’ and ‘0’ to ‘1’ are different but the average one bit error probability can still be used for analysis as is conventional in established communication systems. Thus lower transmission efficiency should be observed with larger distances, which has been shown in Figure 6.7 (d). For different transmission distances, with the increase of the number of molecules per bit n , the transmission efficiency increases until it levels off, with a maximum value of 97%. Moreover, the channel throughput increases with the rising

of the value of n . However, for different transmission distances, the equilibrium values are different, specifically, 64.17bps , 16.04bps and 4.01bps for conditions of $2\mu\text{m}$, $4\mu\text{m}$ and $8\mu\text{m}$ respectively. In the subsequent research in this section, the distance is taken as $4\mu\text{m}$.

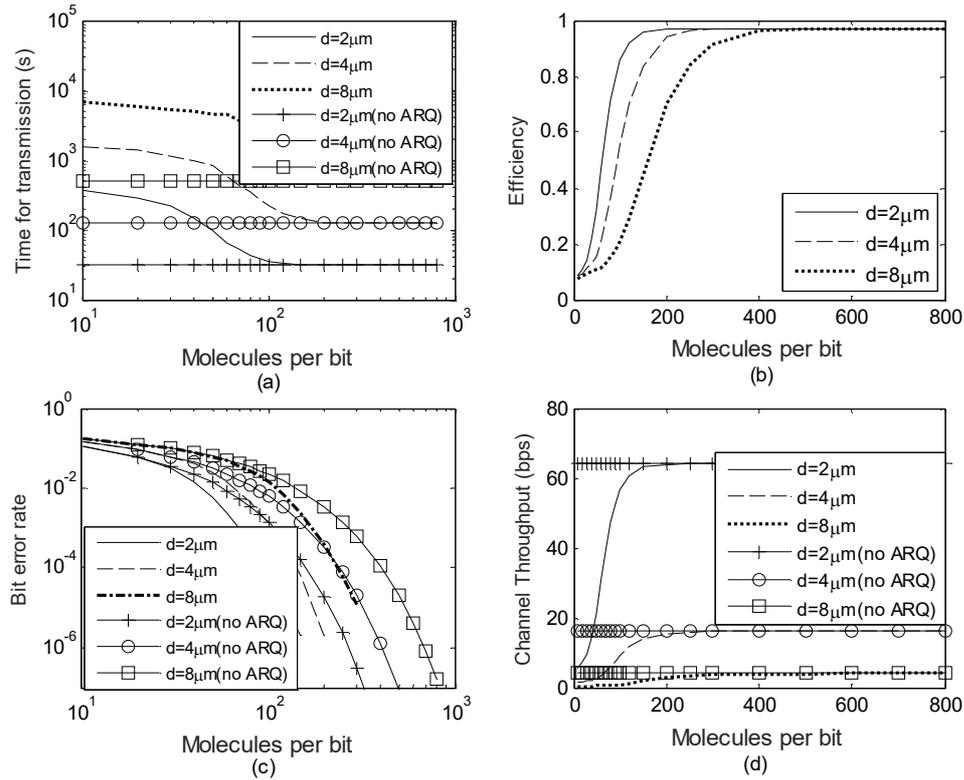


Figure 6.7 Channel performance of SW-ARQ for different transmission distances: (a) Transmission delay; (b) Efficiency; (c) BER; (d) channel throughput.

In Figure 6.8, the number of bacteria in the receiver node is taken into consideration, ranging from 100 to 1000, keeping the frame length at 100 bits with CRC-8 applied. 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, thus achieving a higher channel throughput. Moreover, fewer packet corruptions

happen during the transmission process and the transmission efficiency is higher in this situation. Also, when the number of molecules per bit is above approximately 250, the average error probability is at the level of 10^{-5} , which is an estimation according to Figure 4.5. In this situation, for the transmitted frames which have equal lengths of 100, there is little difference in retransmission times per frame. Thus with the increase in the number of molecules per bit, the transmission time, efficiency and channel throughput all reach equilibrium. In addition, the transmission time converges to identical stable states for different values of m , when the number of molecules per bit is approximately above 250. This is because that the intensity of impact of the bacterial population m on the receiver radius R is slight, leading to a little influence on the capture probability according to equation (4.7). It is obvious that similar situations occur for both the transmission efficiency and channel throughput performances.

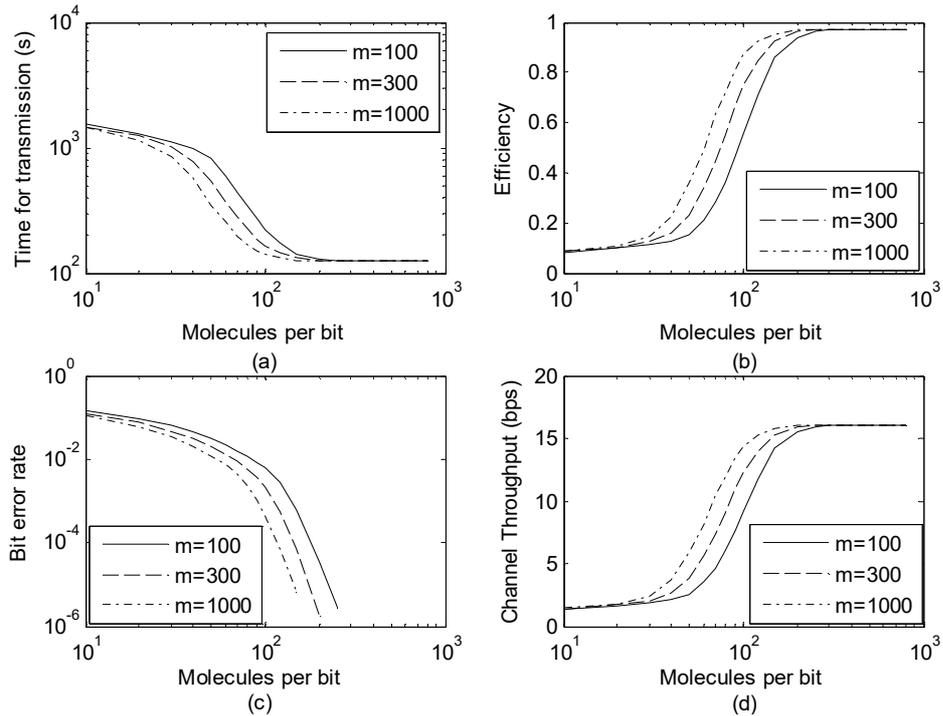


Figure 6.8 Channel performance of SW-ARQ for different number of bacteria: (a) Transmission delay; (b) Efficiency; (c) BER; (d) channel throughput.

Figure 6.9 displays the results for different CRC codes. Here the frame length is still 100 bits, the transmission distance is $4\mu m$, and the number of bacteria in the receiver is 100. It is clear that much more time is needed when more check bits are appended, especially when the number of molecules per bit is below 100. Also it should be noted that CRC-8 shows better transmission efficiency because not too many redundancy bits are added to the frames to occupy more time. In addition, the error probability is smaller with higher order CRC polynomials. Hence, to balance the channel throughput and error detection capability, CRC-8 will be adopted in the following investigations due to its relatively simpler operation, lower time consumption and higher transmission efficiency and channel throughput.

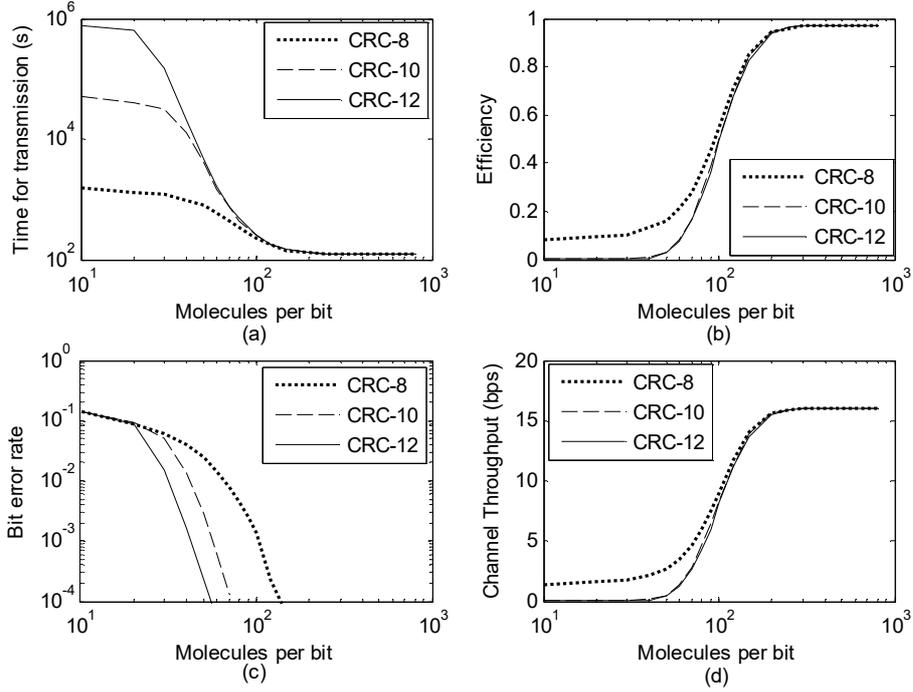


Figure 6.9 Channel performance of SW-ARQ for different CRC polynomials: (a) Transmission delay; (b) Efficiency; (c) BER; (d) channel throughput.

Figure 6.10 shows the system performances when the frame length is different with other parameters taking on the values stated above. Figure 6.10 (c) shows that the channel reliability is better with a smaller frame length. This is because the probability of an error frame is calculated by

$$p_f = 1 - (1 - p)^{n_{cw}} \quad (6.5)$$

where p is the average probability of one bit error. It should be noted here that even though the proposed channel is a BAC, equation (6.5) still works for rough estimations because binary ‘1’s and ‘0’s are equally distributed in the transmitted data stream. This one bit error probability will be also used for the analysis of the following communication scenarios. It is clear that when the frame length is larger, there is a larger probability that transmission error occurs in the frame. In addition, from Figure 6.10 (a), (b) and (d), it is obvious that there exist two intersection points,

specifically at the points when the numbers of molecules per bit are approximately 20 and 250, respectively. When the value of n is in the range between 20 and 250, the channel with smaller frame length performs better, with smaller transmission time and larger transmission efficiency and channel throughput. The possible reason is that the one bit error probability is relatively high in this range, leading to the increasing retransmission times for average per frame when the frame length is large. Thus the transmission time will be larger for large values of N . When the value of n is larger than 250, the one bit error probability is small enough for successful transmission of an information frame (at the level of 10^{-5}). In this case, the retransmission times per frame will decrease sharply and level off. Also, larger frame length means fewer frames to be transmitted, which would lead to slightly smaller transmission time of the data stream for large values of N . Similar situations happen for transmission efficiency and channel throughput performance.

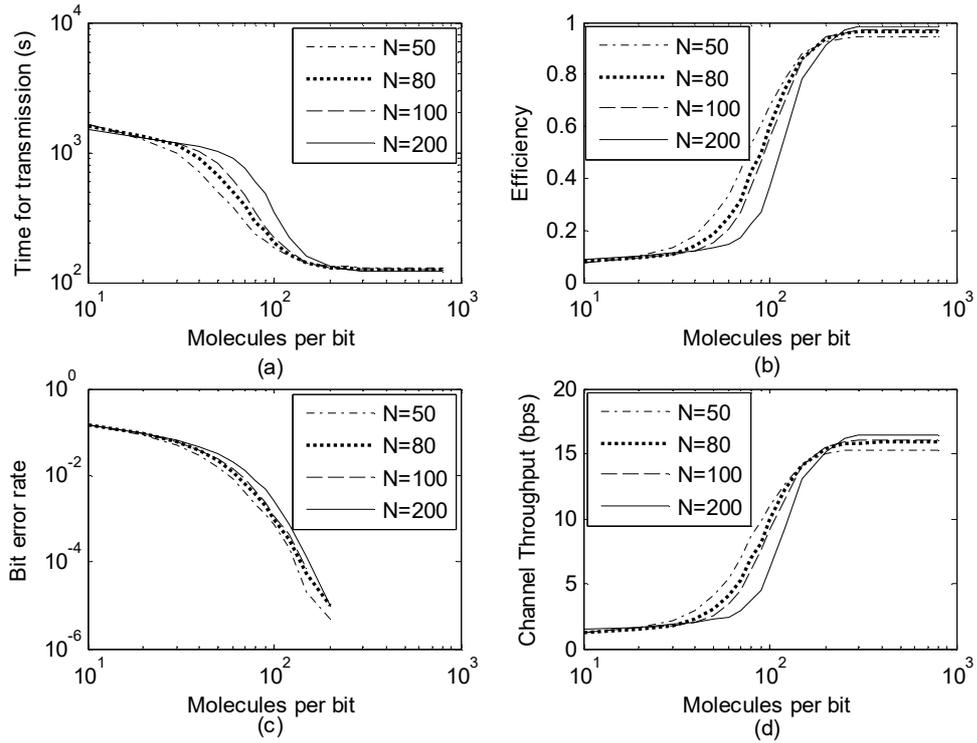


Figure 6.10 Channel performance of SW-ARQ for different frame length: (a) Transmission delay; (b) Efficiency; (c) BER; (d) channel throughput.

According to the investigations above, it is manifest that the four key parameters that have a big influence on the system performance are the number of bacteria in the receiver node m , the CRC polynomial, the transmission distance d and the frame length N . Hence, as shown in Figure 6.11 to Figure 6.14, the performances of different combinations of these four factors are taken into consideration.

In Figure 6.11, results show that for the 16 proposed conditions, the transmission delay varies mainly because of the CRC polynomial, especially when the number of molecules emitted at the start of each time slot is less than approximately 150. A communication channel with a lower order CRC polynomial achieves a smaller transmission time, regardless of the other three factors. Also, when the same CRC polynomial is applied, the communication channel with a smaller transmission

distance achieves a smaller transmission time, regardless of the values of m and N . Moreover, for conditions when the same transmission distances and error detection codes are applied, less time will be consumed for transit the data stream when there is a smaller population in the receiver node. Finally, with fixed CRC polynomials and fixed vales of d and m , the data stream transmits faster if it is broken into frames with smaller length. In addition, when the number of molecules per bit is larger than approximately 150, all the curves converge to their steady states, which are mainly decided by the transmission distance. Specifically, when the curves level off, less time is taken to transmit the data stream over a smaller transmission distance. In addition, when the distance between the transmitter and receiver is fixed, the transmission time for the channel with a larger frame size will be smaller. Also, when the values of d and N are fixed, the channel with a larger receiver bacterial population consumes less time for transmission. In short, under the condition $n < 150$, the CRC polynomial has the largest impact on the transmission time, followed by the transmission distance, the bacterial population in the receiver and the frame length, in sequence. However, when $n > 150$, the significance to the time for transmission is in the decreasing order of d , N , m and CRC polynomial.

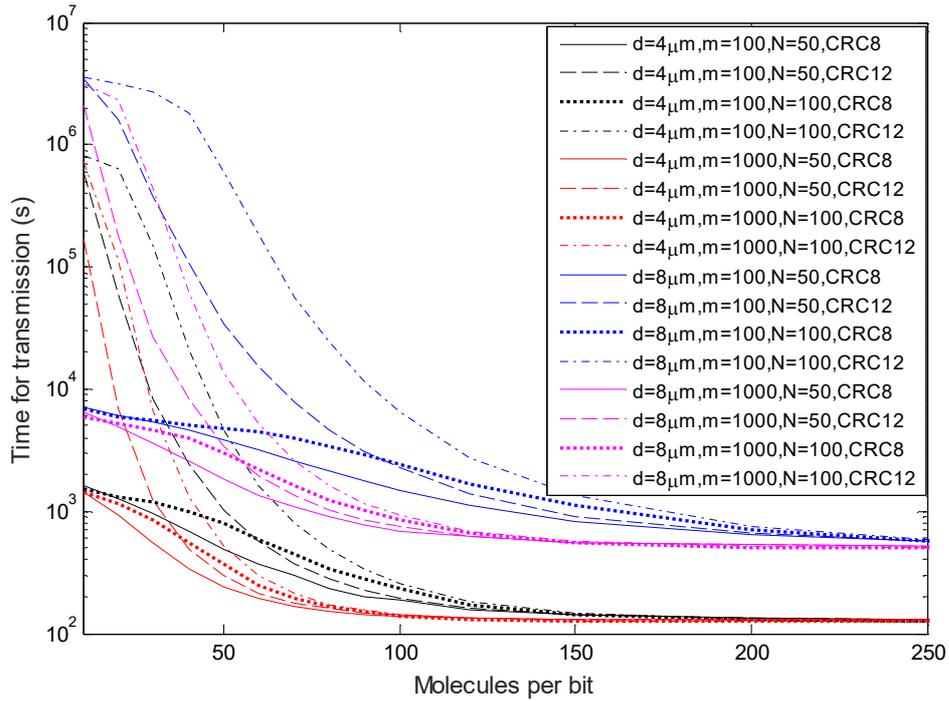


Figure 6.11 Transmission time of SW-ARQ for different combinations of the main factors discussed in the text.

Figure 6.12 shows the transmission efficiency of SW-ARQ for different parameter combinations. It is obvious that the curves can be roughly divided into three stages: stage 1 when the number of molecules per bit is below approximately 100, stage 2 when the curves have reached the steady states and stage 3, which is the period between stage 1 and stage 2. In the first stage, the transmission efficiency increases slowly with the increase of the number of molecules per bit, with the maximum transmission efficiency of approximately 15%. In this stage, the transmission efficiency is mainly decided by the CRC polynomial used, which means that the communication channel with CRC-8 performs better than that using CRC-12, regardless of the other three parameters. In stage 2, where the transmission efficiency is more than 90%, the same steady state values can be achieved with the same frame lengths. In other words, if the value of N is fixed, the maximum value is also determined, no matter what the values of the other three parameters are. Specifically,

for $N = 50$, the equilibrium value of transmission efficiency is 97%, while for $N = 100$, it is 94%. For the conditions which have the same frame length, the system can achieve a larger maximum transmission efficiency with a larger bacterial population in the receiver, regardless of the value of d and CRC polynomial. Also, with the same values of m and N , a higher steady state transmission efficiency can be achieved with a smaller transmission distance. Thus, in stage 2 the significance to the transmission efficiency of the four factors will be in the decreasing order of frame length, number of bacteria in the receiver, transmission distance and CRC polynomial. In state 3, there is a rapid increase in the transmission efficiency between 15% and 90%. It is obvious that in this stage, to achieve a particular transmission efficiency, the number of molecules required to be sent at the start of each time slot is the most under the condition $d = 8\mu m, m = 100$, while the least molecules are required to be sent at the start of each time slot under the condition $d = 4\mu m, m = 1000$.

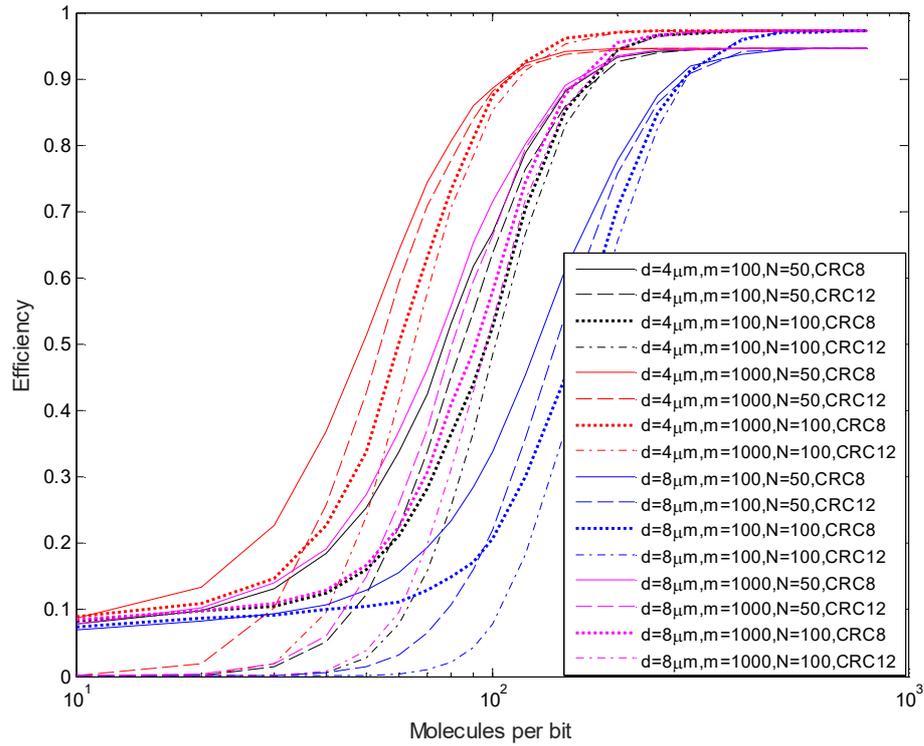


Figure 6.12 Transmission efficiency of SW-ARQ for different combinations of main factors discussed in the text.

Figure 6.13 shows the error performance of SW-ARQ for different parameter combinations. Overall, the error probability is smaller during the transmission process when a higher order CRC polynomial is applied. Also, with the same error detection code, the channel is more reliable with a smaller transmission distance, regardless of the values of N and m . Moreover, for the condition when the distance between the transmitter and receiver and the CRC polynomial are fixed, the communication channel which has a larger bacterial population in the receiver node is more reliable, regardless of the value of frame length. Thus, the importance to the channel reliability of the four parameters will be in the decreasing order of CRC polynomial, transmission distance, bacterial population in the receiver node and the frame length.

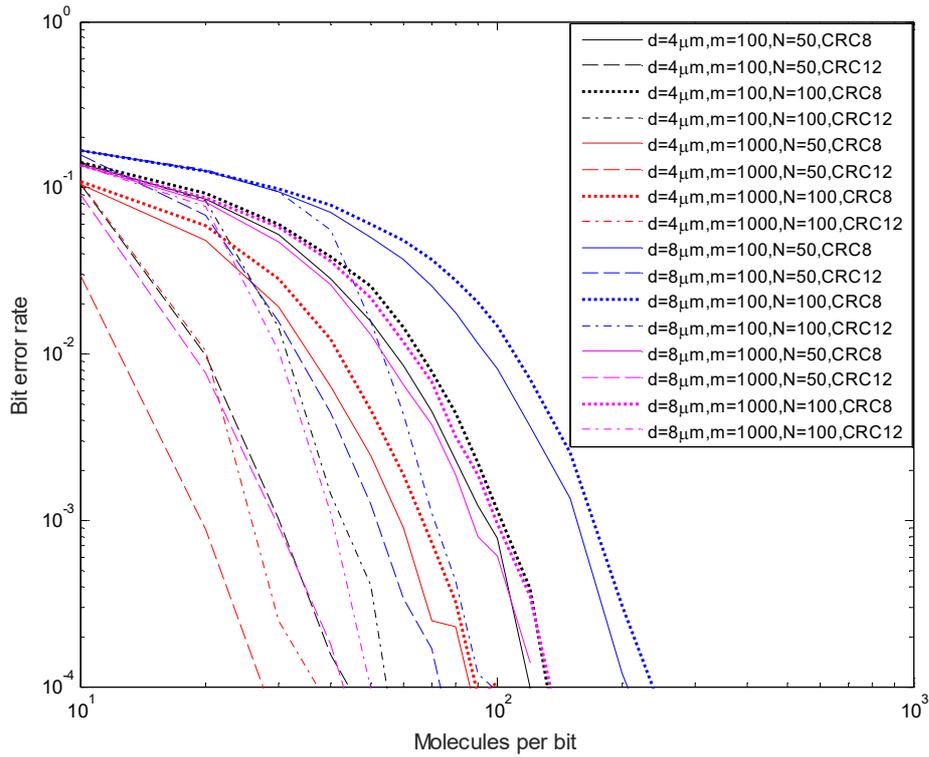


Figure 6.13 Error performance of SW-ARQ for different combinations of the main factors discussed in the text.

Figure 6.14 shows the channel throughput performance of SW-ARQ for different parameter combinations. It indicates that the performance of the channel throughput can be divided into two stages. In stage 1, the channel throughput is rapidly increasing with the increase of the number of molecules per bit. It shows that at the starting point where the number of molecules per bit is 10, the corresponding channel throughput is higher with a lower order CRC polynomial. Also, when the same CRC is applied, the channel which has a smaller transmission distance achieves a larger throughput. Specifically, for conditions when $d = 4\mu\text{m}$ and CRC-8 is applied, the initial channel throughput is 1.48bps ; for situations when $d = 8\mu\text{m}$ and CRC-8 is applied, the initial channel throughput is 0.35bps ; while for conditions when CRC-12 is applied, it shows quite a small initial channel throughput of circa $7 \times 10^{-3}\text{bps}$. Thus the significance to the initial point of transmission throughput of

the four factors will be in the decreasing order of CRC, d , m , N . After the starting point, the curves increase rapidly to reach their steady states. The results indicate that better channel throughput can be achieved with a smaller transmission distance, regardless of the other three factors. Moreover, according to the conditions which have the same transmission distance, the increasing of the number of bacteria in the receiver node leads to the growth of channel throughput. Also, for the conditions which have the same propagation distance and population size of the receiver node, the system which has a smaller frame length has a better performance, with a higher channel throughput, regardless of the employed CRC polynomial. Hence, in stage 1, the significance to the channel throughput of the four factors will be in the decreasing order of transmission distance, number of bacteria in the receiver node, frame length and CRC polynomial. In stage 2, the channel throughput has reached the steady state. For the conditions when the same propagation distance and the same frame length are applied, the system will reach identical steady states. Specifically, the equivalent values of channel throughput are $16.04bps$, $15.31bps$, $4.01bps$ and $3.5bps$ for conditions $d = 4\mu m, N = 100$; $d = 4\mu m, N = 50$; $d = 8\mu m, N = 100$ and $d = 8\mu m, N = 50$, respectively, regardless of the other two parameters.

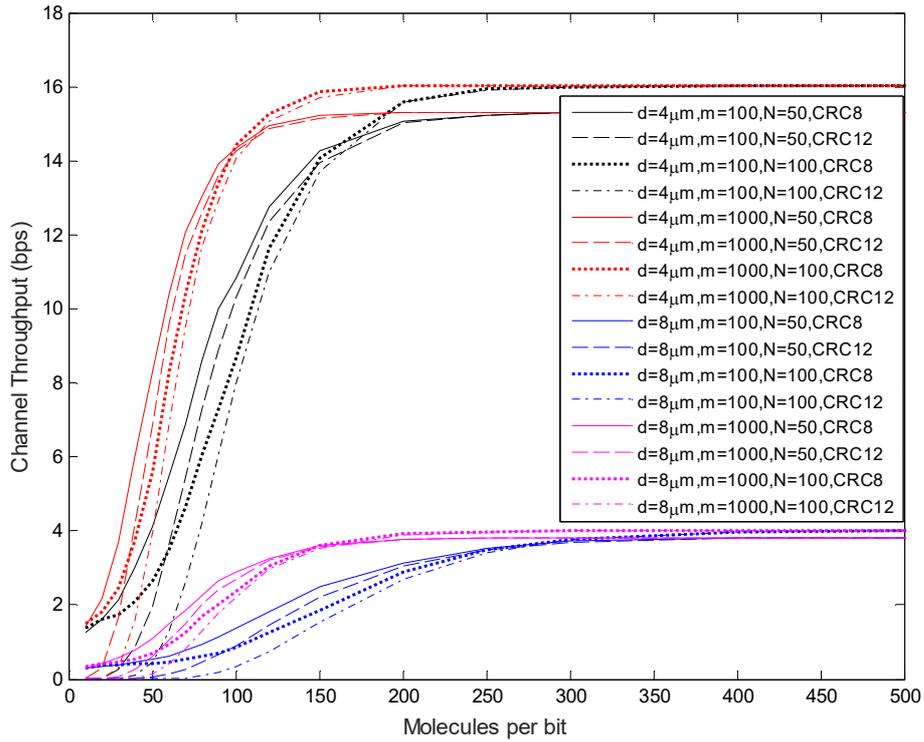


Figure 6.14 Channel throughput of SW-ARQ for different combinations of the main factors discussed in the text.

6.5 Simulation results for GBN-ARQ

In this section, when the influences of the parameters, including the transmission distance, the number of bacteria in the receiver node, the frame length and the window size, have been comprehensively taken into account, the performances of the channel, including transmission delay, transmission efficiency, BER and channel throughput, for GBN-ARQ protocol, are compared by numerical analysis based on the simulation results.

6.5.1 Parameter setup

Considering the proposed bacterial communication channel, simulations of the information transmission process have been made for GBN-ARQ in MATLAB. The total amount of transmitted information bits remains 2000 as per the simulations of

SW-ARQ. Also, CRC check bits and a 3-bit sequence number are appended to frames of equal length, in a controlled manner. The parameters discussed above for SW-ARQ, as well as the window size W , will be varied to investigate the system performance. According to Figure 6.9, although systems with the application of high order CRC polynomials are more reliable, the transmission efficiency and channel throughput are substantially decreased due to the complex operations for substantially increased numbers of CRC check bits [250]. Thus CRC-8, with polynomial representation $x^8 + x^2 + x + 1$, is applied here due to its relatively lower operation time and energy consumption. Each time slot only allows the transmission of 1 bit. The transmission delay t_0 can be calculated according to equation (6.1). The definitions of the parameters remain those given in Section 6.4.1, and the processing time for CRC implementation is ignored again for simplification. Also, the bit rate of the channel is taken as the maximum value according to equation (6.2). In addition, the timeout period is set to be exactly equal to the sum of round trip propagation delay, which can be calculated according to Chapter 4, and the CRC processing time. Moreover, the computing method for obtaining the transmission efficiency and channel throughput are the same as illustrated in Section 6.4.1.

6.5.2 Simulation process

This section describes the simulation process of the GBN-ARQ protocol, which again begins with the generation of the 2000 bits to be transmitted as previously, uniformly distributed with '1's and '0's. The data stream is then encapsulated into frames for transmission. Here, the transmitter has two modes: Ready mode and Wait mode; whilst the receiver is always in Ready mode. It has been stated that there is a sending window for the transmitter. The window size is denoted by W . Here, three symbols, including S_{last} , S_{recent} , and R_{next} are introduced to describe the simulation

process. The sending window buffers all the frames that can be transmitted simultaneously, the sequence numbers of which are within the range between S_{last} and $(S_{\text{last}} + W - 1)$. S_{last} represents the sequence number of the oldest unacknowledged frame, S_{recent} is the sequence number of the most recent transmitted frame, and $(S_{\text{last}} + W - 1)$ is the maximum sequence number allowed to transmit. R_{next} represents the sequence number of the frame that is expected by the receiver.

Before data transmission, the transmitter is initialized to Ready mode, which will be described later. The parameters including S_{last} , S_{recent} , and R_{next} are initialized to zero.

During the data transmission process, the transmitter mode details are:

- Ready mode: The transmitter prepares an information frame with the sequence number S_{recent} , in the header and starts a timer. Then S_{recent} is updated to $S_{\text{recent}} = (S_{\text{recent}} + 1) \text{ modulo } W$. This process is denoted as step 1.

If $S_{\text{recent}} = (S_{\text{last}} + W - 1) \text{ modulo } W$, the transmitter enters the Wait mode; otherwise, the transmitter goes to step 1.

- Wait mode: If a correct ACK/NAK with sequence number R_{next} is received before the timer expires, and R_{next} is between S_{last} and S_{recent} , the value of S_{last} is set to R_{next} , and the transmitter returns to step 1. However, if the timer of S_{last} expires, the value of S_{recent} is reset to S_{last} , and the transmitter returns to the Ready mode (step 1).

The receiver is always in the Ready mode. Whenever an information frame arrives, it is checked for errors at the receiver.

- If no error is detected and the received frame number S_{last} equals to R_{next} , the information frame is passed on to the network layer at the receiver side. Also, R_{next} is updated to $(R_{\text{next}} + 1) \bmod W$, and an ACK with sequence number R_{next} is sent.
- If the received information frame has errors, the frame is discarded and an NAK with sequence number $R_{\text{next}} = S_{\text{last}}$ is sent.
- If no error is detected but the received frame sequence number S_{last} does not equal to R_{next} , the frame will be discarded, without any other actions.

6.5.3 Results and discussions

For a communication system with GBN-ARQ applied, the transmitter has a limit on the number of frames that can be outstanding. As stated above, with an λ -bit sequence number appended to each information frame, the sequence of frames carry the sequence numbers with decimal representation $[0, (2^\lambda - 1)]$. In general, the window size W for GBN-ARQ needs to be $(2^\lambda - 1)$ or less to avoid ambiguities [251]. Thus, in this model with a 3-bit sequence number, 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 Figure 6.15. 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 mechanisms are used at all, the channel which uses ARQ performs better with a lower BER, achieving a higher channel reliability. Moreover, the transmission delay, BER, efficiency and channel throughput all

improve with smaller transmission distances because over a larger distance, the increased one bit error probability results, on average, in longer transmission times per frame. Thus, a lower transmission efficiency and channel throughput should be observed with larger distances confirmed by the results in Figure 6.15 (b) and Figure 6.15 (d). In what follows, a representative transmission distance is taken as $4\mu\text{m}$.

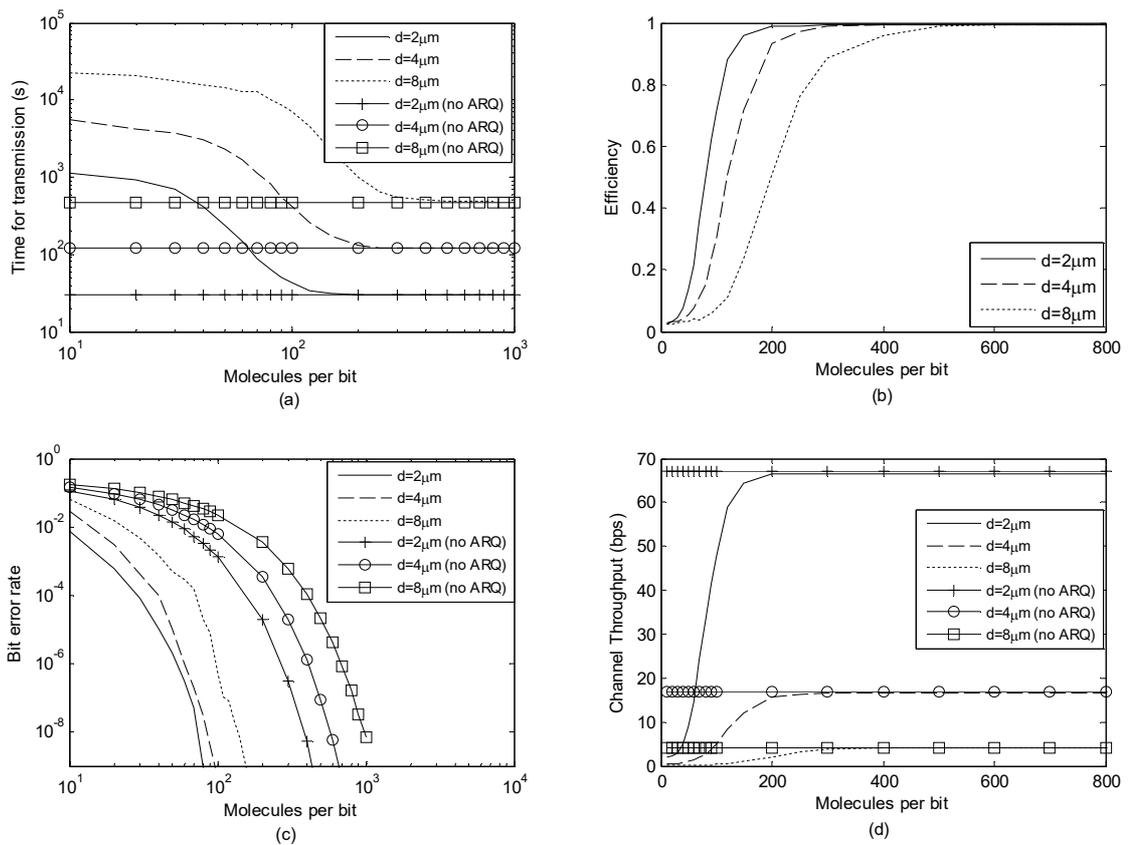


Figure 6.15 Channel performance of GBN-ARQ for different transmission distances: (a) Transmission delay; (b) Efficiency; (c) BER; (d) Channel throughput.

Figure 6.16 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, error performance does not vary significantly. Figure 6.16 (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 the lower end of what is practical, the channel probability for one-bit error 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 causes many more retransmissions, resulting in increased transmission time and lower efficiency and channel throughput, which perfectly matches the result in Figure 6.16 (b) and (d).

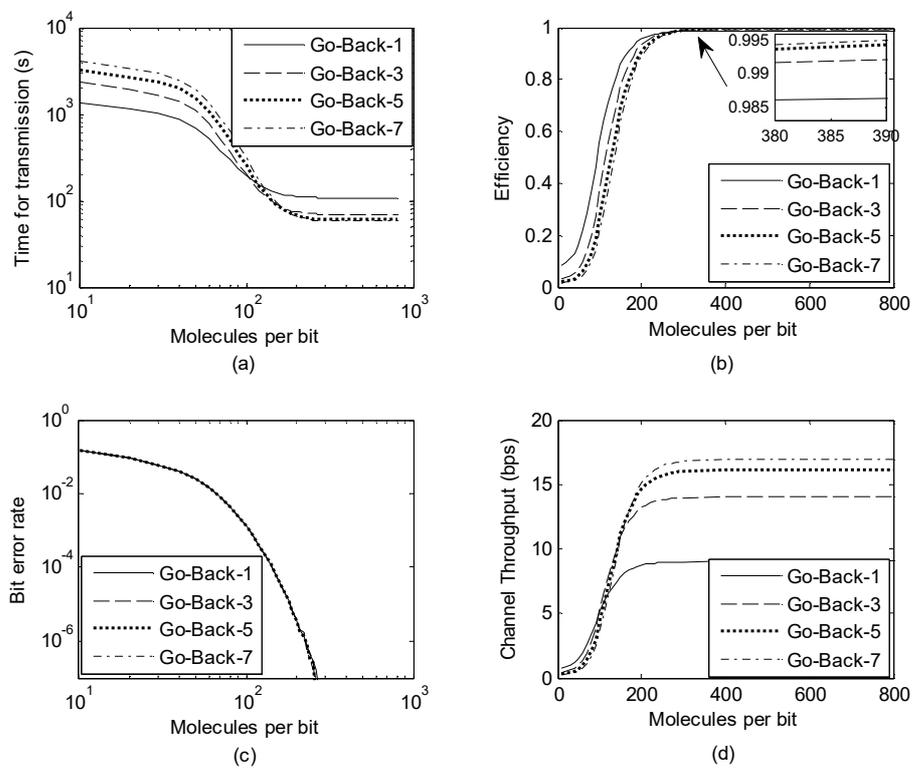


Figure 6.16 Channel performance of GBN-ARQ for different window sizes: (a) Transmission delay; (b) Efficiency; (c) BER; (d) channel throughput.

However, as the number of molecules per bit increases, the probability of one bit error 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 Figure 6.17, 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 (4.7). Also, with the increase of 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.

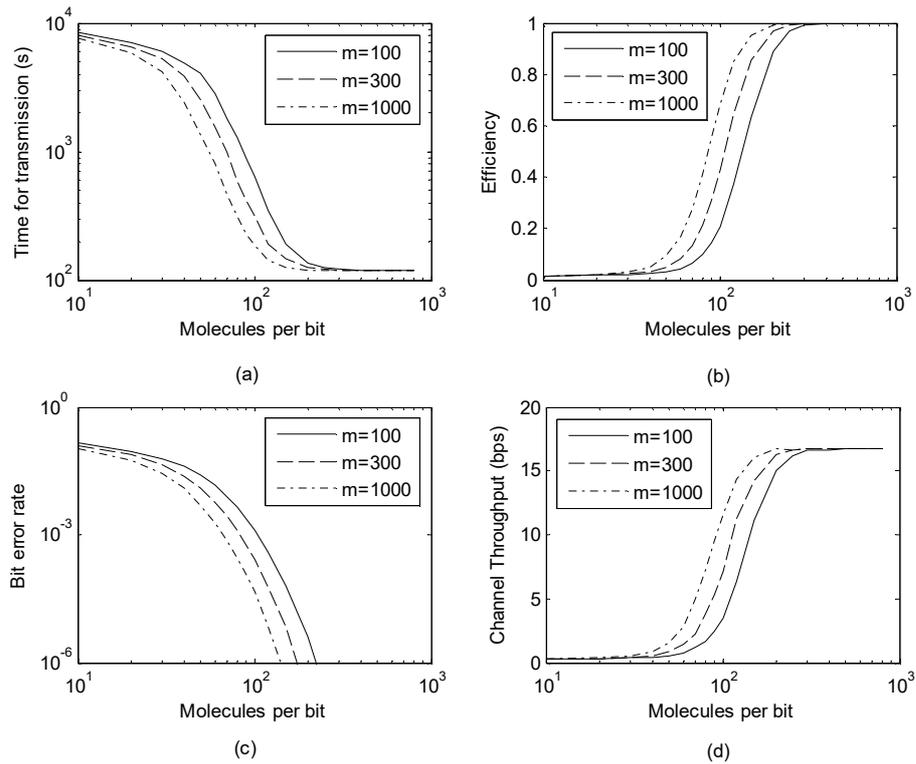


Figure 6.17 Channel performance of GBN-ARQ for different number of bacteria: (a) transmission delay; (b) efficiency; (c) BER; (d) channel throughput.

Figure 6.18 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 that according to equation (6.5), 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, smaller frame lengths perform better. In addition, the channel throughput with larger frame length is larger when the number of molecules emitted at the start of each time slot is greater than 200. These 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 the following chapter in this thesis.

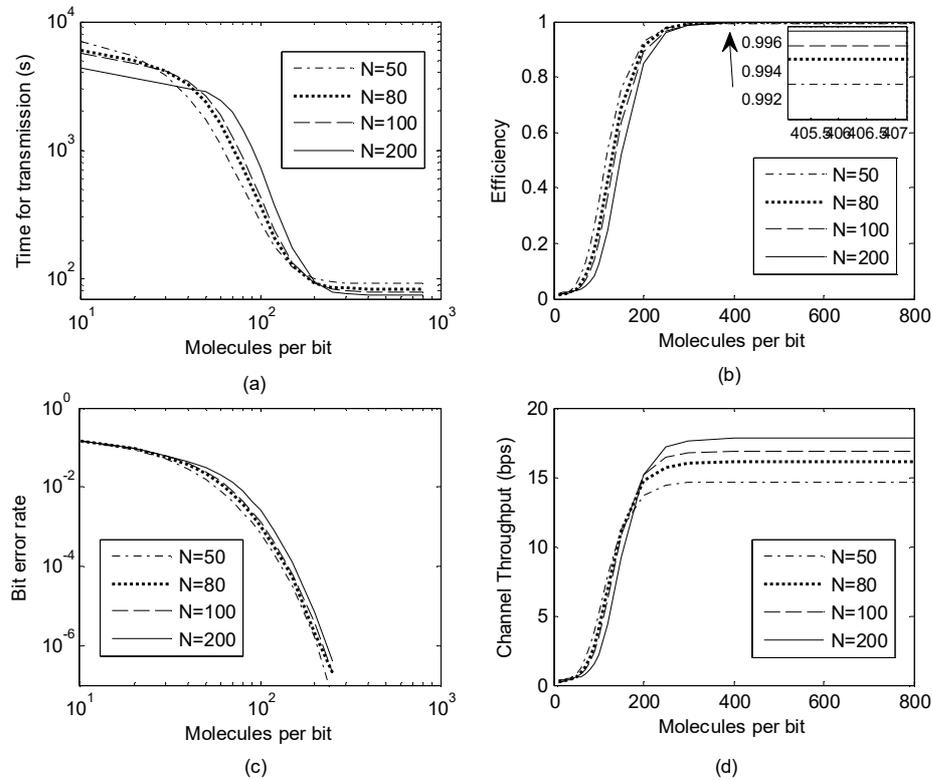


Figure 6.18 Channel performance of GBN-ARQ for different frame lengths: (a) transmission delay; (b) efficiency; (c) BER; (d) channel throughput.

According to the investigations above, results indicate that the system parameters, including the transmission distance d , the bacterial population m , the frame length N and the window size W , have an influence on the performance of the channel for GBN-ARQ scheme. In order to design a better GBN-ARQ scheme with appropriate system parameters, the influence degree of each of the four parameters is required to be taken into account. Also, in a particular communication scenario, the distance between the transmitter and the receiver should be a fixed value. Thus, as shown in Figure 6.19–Figure 6.22, the channel performances of different combinations of parameters m, N, W are considered.

Figure 6.19 shows the transmission time of the data stream when different combinations of parameters are applied for GBN-ARQ scheme. It is obvious that the transmission time for different communication scenarios can be divided into two stages. In stage 1 when the number of molecules per bit is less than approximately 250, the transmission time drops down rapidly with the increase of the value of n for all parameter combinations. Also, the communication channels which have the same window size and the same bacterial population in the receiver node have the same initial growth point. Moreover, in this stage, the transmission time is mainly decided by the window size, which means that the communication channel with smaller window size consumes less time for transmission regardless of the other two parameters. Also, when the value of W is fixed, the channel with larger value of m can achieve smaller transmission time. Thus in stage 1, the window size W has the biggest influence on the transmission time, followed by the receiver bacterial population and the frame size. In stage 2 when the number of molecules per bit is larger than 250, the channel which has a larger window size achieves a slightly smaller transmission time. However, there is little difference among these 12 communication scenarios.

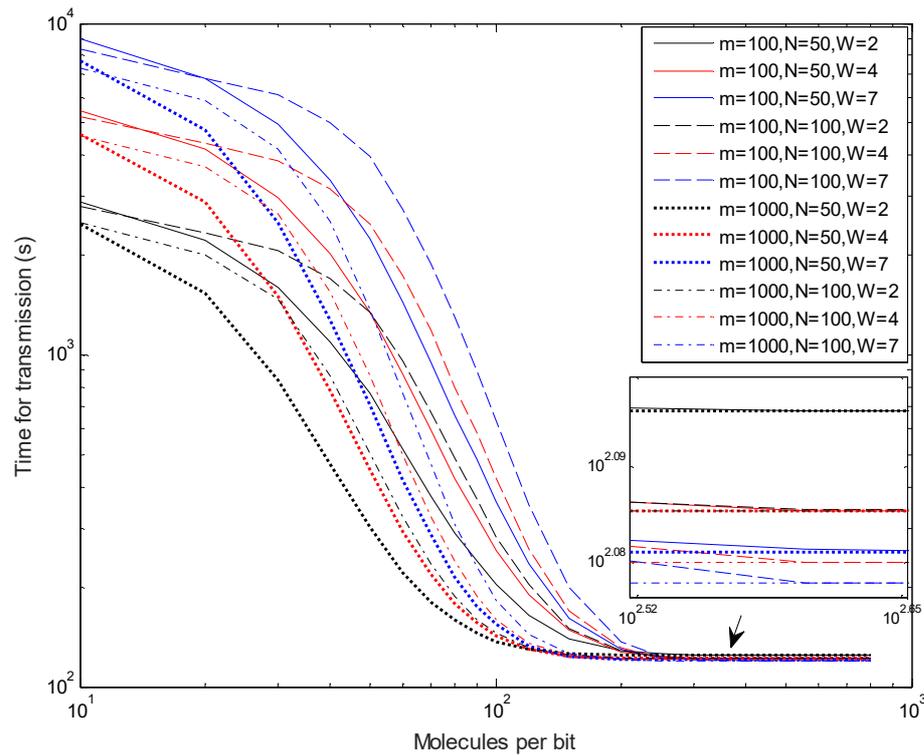


Figure 6.19 Transmission time of GBN-ARQ for different combinations of the main factors discussed in the text.

Figure 6.20 shows the transmission efficiency of the channel when different combinations of parameters are applied for GBN-ARQ scheme. Figure 6.22 shows the channel throughput of the communication system under different communication scenarios. The two figures are discussed together because of the similar tendency of the curves. Both the transmission efficiency and channel throughput can be divided into two stages, specifically stage 1 when $n < 250$ and stage 2 when $n > 250$, which is similar to the transmission time behaviour in Figure 6.19. In stage 1, the transmission efficiency and channel throughput increase rapidly with the increase in the number of molecules per bit. Also, the communication scenarios which have the same window sizes and receiver bacterial populations have the same initial points of growth. During the rapid increasing period, the transmission efficiency and channel throughput of the system which has a larger a larger window size grow faster,

regardless of the other two parameters. In addition, to achieve a particular value of transmission efficiency or channel throughput, more molecules are required to be released at the start of each time slot when the communication system has a receiver with larger bacterial population, regardless of the other two parameters. Moreover, when m is a fixed value, it requires a larger number of molecules emitted at the start of each time slot for the channel with a larger window size to achieve a specific level of transmission efficiency or channel throughput. Thus in stage 1, the significance to both the efficiency and channel throughput is in the decreasing order of m, N, W . In stage 2, when all the curves have reached their steady states, the channel which has a larger window size can achieve larger transmission efficiency and channel throughput, however, the difference is quite slight.

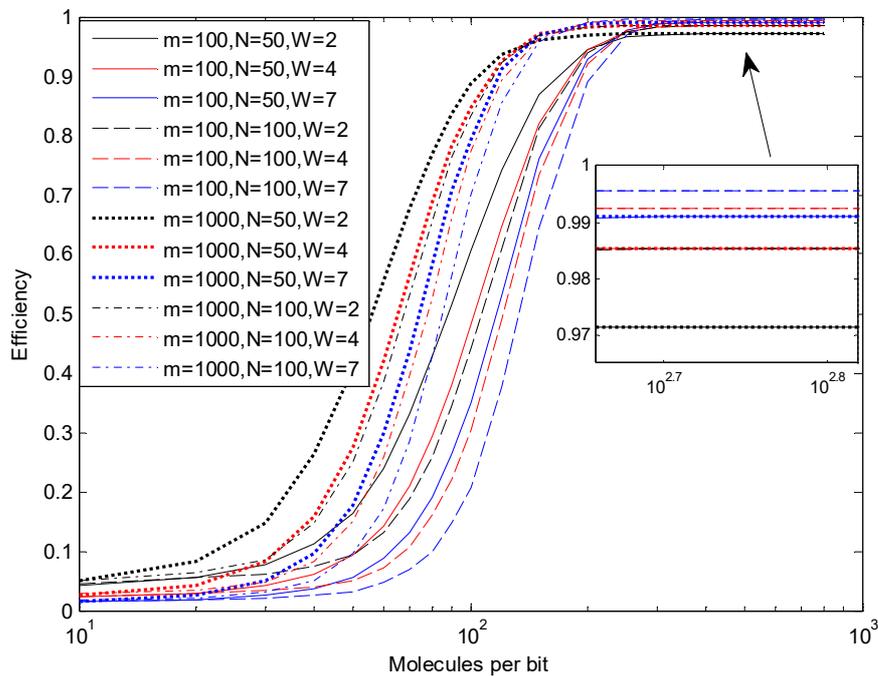


Figure 6.20 Transmission efficiency of GBN-ARQ for different combinations of the main factors discussed in the text.

Figure 6.21 shows the error performance of the channel for 12 different communication scenarios. It is clear that when the same values of m and N are

applied in the GBN-ARQ protocol, there is little variation in the error probabilities, regardless of the window size. Also, the BER of the channel will be smaller with a larger receiver bacterial population, regardless of the values of N and W , which means that the significance of the three parameters to the error performance is in the decreasing order of m, N, W .

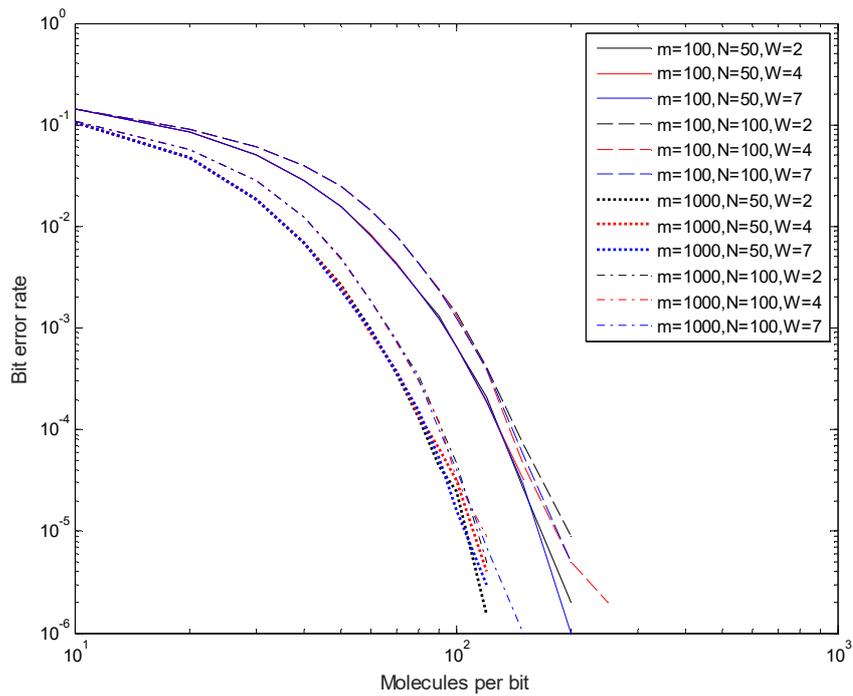


Figure 6.21 Error performance of GBN-ARQ for different combinations of the main factors discussed in the text.

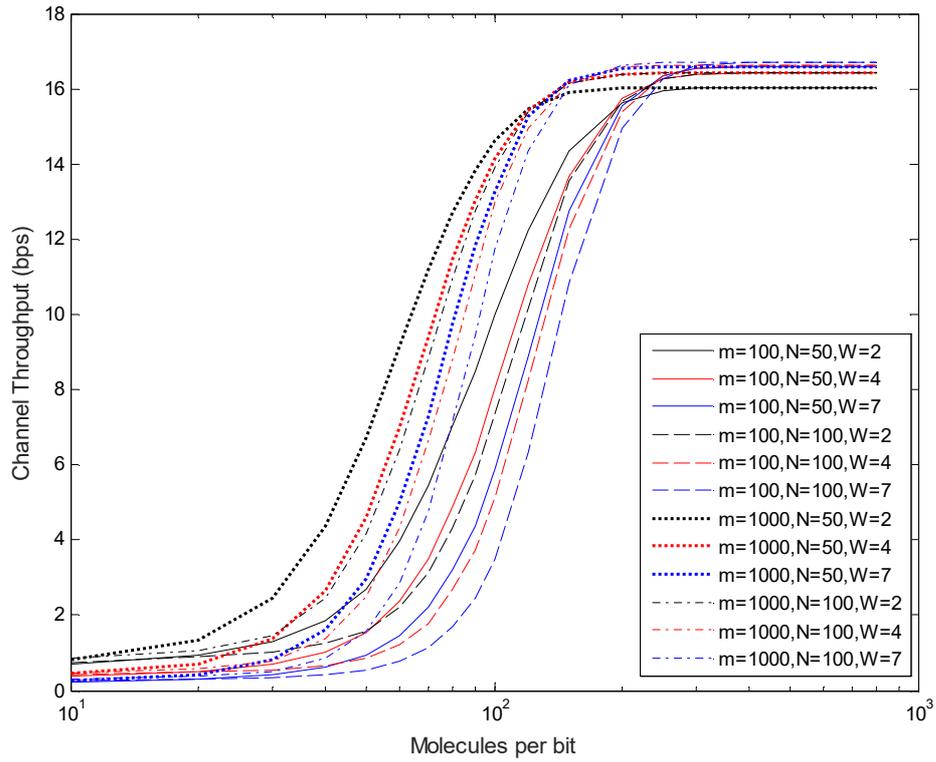


Figure 6.22 Channel throughput of GBN-ARQ for different combinations of the main factors discussed in the text.

6.6 Simulation results for SR-ARQ

In this section, when considering the application of SR-ARQ scheme in the proposed diffusion-based MC channel, the transmission time of the data stream, the transmission efficiency, the error probability and the channel throughput of the channel have been investigated by varying different channel parameters.

6.6.1 Parameter setup

For SR-ARQ, the parameters considered are similar with that for GBN-ARQ, including the transmission distance d , the number of bacteria in the receiver node m , the frame length N and the window size W . Here, CRC-8 will again be applied because of its relatively simple computation operation. In addition, for better

comparisons, the parameter settings, including the transmitted data length, the length of sequence bits, the transmission data rate, the time out period, are the same as that of the GBN-ARQ scheme described in Section 6.5.1.

6.6.2 Simulation process

This section describes the simulation process of SR-ARQ protocol, which again begins with the generation of 2000 bits to be transmitted. The transmitter has two modes, specifically Ready mode and Wait mode, whilst the receiver node is always ready for operations. The window sizes of both the transmitter and receiver are denoted by W . Here, three symbols, including S_{last} , S_{recent} , and R_{next} are introduced to describe the simulation process, the representations of which have been explained in detail in Section 6.5.2. Here, it should be noted that the receiving window buffers more than one frame, the sequence numbers of which could be within the range between R_{next} and $(R_{\text{next}} + W - 1)$ inclusive; the latter represents the maximum sequence number that can be accepted by the receiver.

Before data transmission, the transmitter is initialized to Ready mode. The parameters including S_{last} , S_{recent} , and R_{next} are initialized to zero. The behaviour of the transmitter is the same as that of GBN-ARQ described in Section 6.5.2, thus is left out here. For the receiver, it is always in the Ready mode. Upon arriving of an information frame, it is checked for errors at the receiver.

- If no error is detected and the received frame number S_{last} equals to R_{next} , the information frame is passed on to the network layer at the receiver side. Also, R_{next} is updated to $(R_{\text{next}} + 1) \bmod W$, and an ACK with sequence number R_{next} is sent.

- If the received information frame has errors, the frame is discarded and an NAK with sequence number $R_{\text{next}} = S_{\text{last}}$ is sent.
- If no error is detected but the received frame sequence number S_{last} does not equal R_{next} , the frame will be stored in the buffer, without any other actions.

6.6.3 Results and discussions

In a similar way to the GBN-ARQ scheme, there is also a limit on the maximum send window size for the SR-ARQ scheme. Also, due to the fact that the receiver needs to store the out-of-order positively acknowledged frame it receives, the receiving window is required to be larger than 1. Usually, the window sizes of the transmitter and the receiver are identical. In general, when an λ -bit sequence number is appended to each information frame, the window size W for SR-ARQ needs to be $2^{\lambda-1}$ or less to avoid duplicate transmission [251]. 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 Figure 6.23. Here the frame length is set to be 100 bits, the number of bacteria in the receiver is 100, 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 one bit error probability results in more transmission times per frame. In the following investigations, the transmission distance is thus taken as 4 μm to make better comparisons with the performance of GBN-ARQ.

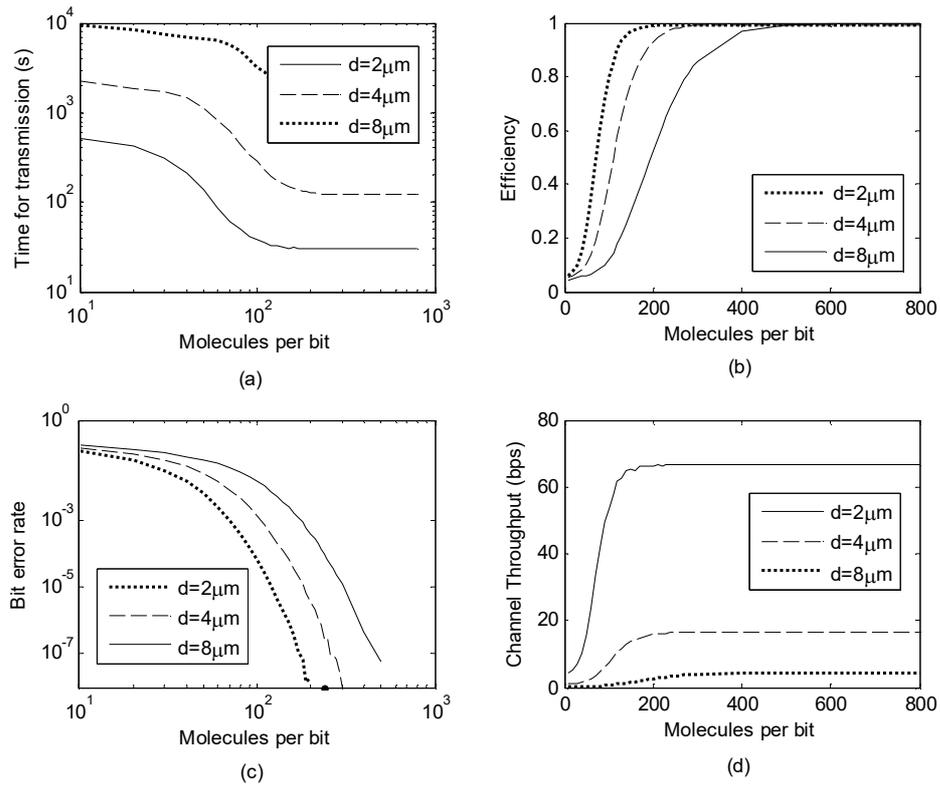


Figure 6.23 Channel performance of SR-ARQ for different transmission distances: (a) Transmission delay; (b) Efficiency; (c) BER; (d) channel throughput.

Figure 6.24 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 error performances for different window sizes are almost the same. Figure 6.24 (a) shows that when the number of molecules per bit is low (approximately within the range 10 to 200), the transmission delay is slightly larger with a larger window size. It should be noted that the condition of $W = 1$ reduces to SW-ARQ. In other words, when the number of molecules per bit is in a smaller range, the time consumption for SW-ARQ is smaller than that for SR-ARQ; and for SR-ARQ, larger window size can achieve larger transmission time. This is because that when the number of molecules per bit is small, the probability of one bit error is relatively large, leading to higher

mean transmission times per frame. Compared with the channel performance of GBN-ARQ in Figure 6.16, the difference is quite small between the performances of channels with different window sizes because only the negatively acknowledged frames need to be retransmitted in SR-ARQ. Thus both the transmission efficiency and channel throughput will be smaller with a larger window size when the number of molecules per bit is less than 200, which perfectly matches the results in Figure 6.24 (b) and (d). In addition, when the value of n is greater than 200, the transmission efficiency and channel throughput is slightly increased by the use of larger window sizes, due to the small probability of one bit error of the channel. In other words, when the number of molecules per bit is relatively small (<200), SW-ARQ ($W = 1$) is more efficient than SR-ARQ ($W > 1$), achieving smaller transmission time and larger transmission efficiency and channel throughput. When the number of molecules per bit is larger than 200, SR-ARQ with larger window size leads to better performance, with slightly improvement. Also, it should be noted that SR-ARQ requires a larger storage capability for the receiver.

In Figure 6.25, the number of bacteria in the receiver 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 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.

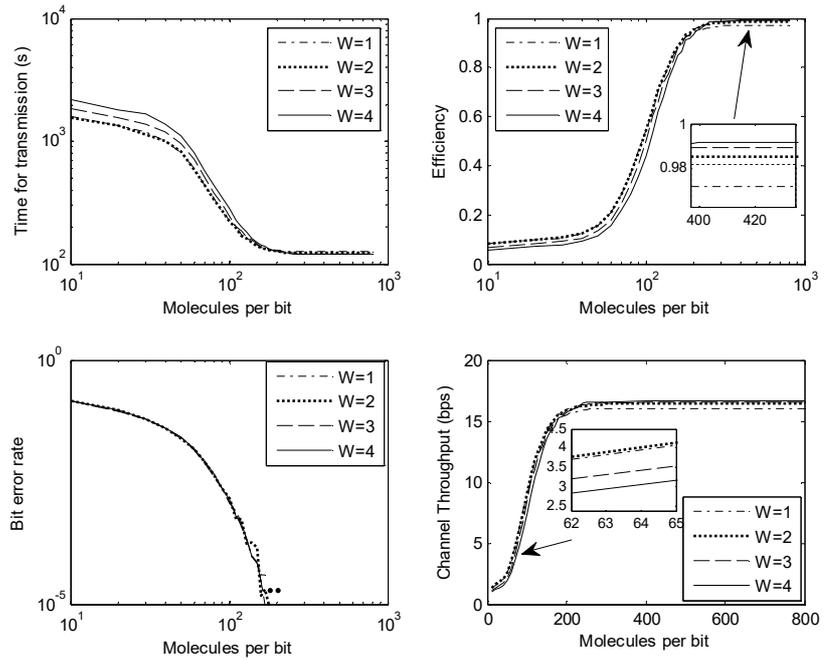


Figure 6.24 Channel performance of SR-ARQ for different window sizes: (a) transmission delay; (b) channel throughput.

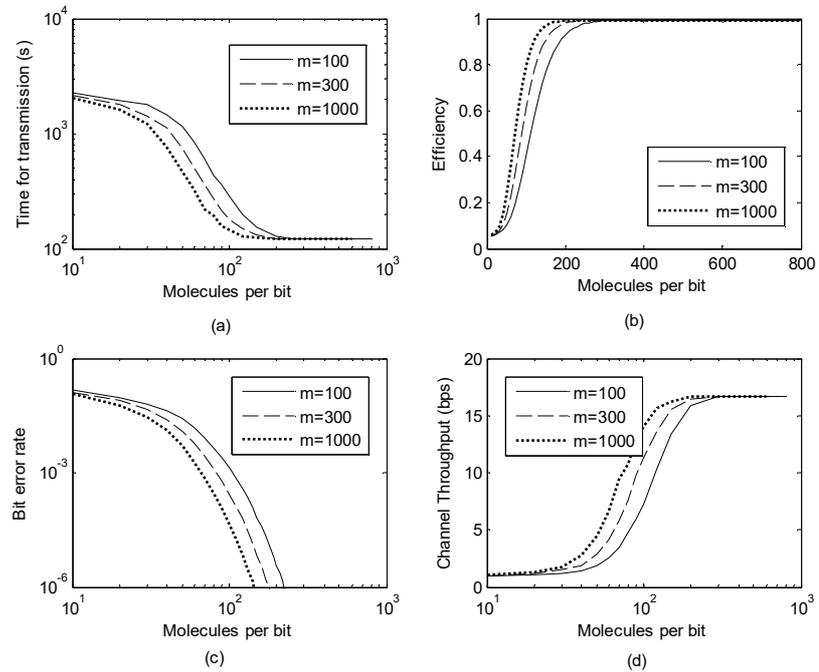


Figure 6.25 Channel performance of SR-ARQ for different number of bacteria: (a) transmission delay; (b) efficiency; (c) BER; (d) channel throughput.

Figure 6.26 shows the system performance for different frame lengths, 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 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 larger error probability for transmission. However, there are two crossover points for the transmission time and channel throughput, indicating that when the number of molecules per bit is between 40 and 250, smaller frame lengths perform better. In addition, the transmission 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 40 to 250. 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 be future investigated in Chapter 7.

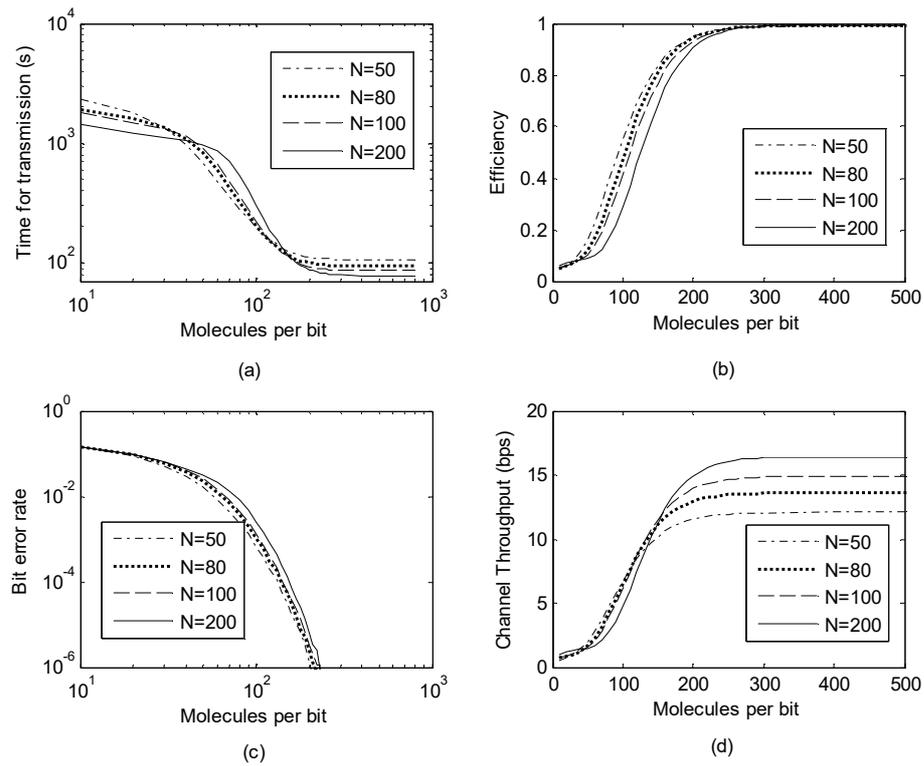


Figure 6.26 Channel performance of SR-ARQ for different frame lengths: (a) transmission delay; (b) efficiency; (c) BER; (d) channel throughput.

According to the investigations above, the system parameters, including the transmission distance d , the bacterial population m , the frame length N and the window size W , can affect the channel performance when SR-ARQ scheme is applied. In order to select appropriate system parameters to design a better SR-ARQ scheme, the impact degree of the four parameters is required to be taken into account. Also, the distance between the transmitter and the receiver is a fixed value, the reason of which has been given in the previous section. Hence, as shown in Figure 6.27–Figure 6.30, the channel performances of different combinations of parameters m, N, W are considered. It should be noted that for SR-ARQ, the window size can only be selected between 2 and 4.

Figure 6.27 shows the total transmission time for the 2000 transmitted bits used in the simulation process, when different parameter combinations are selected. It is clear that when the number of molecules per bit is small (approximately smaller than 200), the transmission time increases with the increased value of selected window size W , no matter how the values of m and N are chosen. Also, when the value of W is fixed, larger receiver bacterial population can achieve smaller transmission time, regardless of the frame length. In addition, for conditions when same values of W and m are applied, the transmission time of the system which has a larger frame length falls faster with the increase of the value of n . Here, it should be noted that the initial decreasing points are almost the same with same parameter combinations of W and m , regardless of the value of window size. When the number of molecules emitted at the start of each time slot is larger than 200, the transmission times for the 8 simulated scenarios have approximately identical steady state values. In short, under the condition of $n < 200$, the transmission time of the data stream mainly depends on the window size, followed by the receiver bacterial population and the frame length.

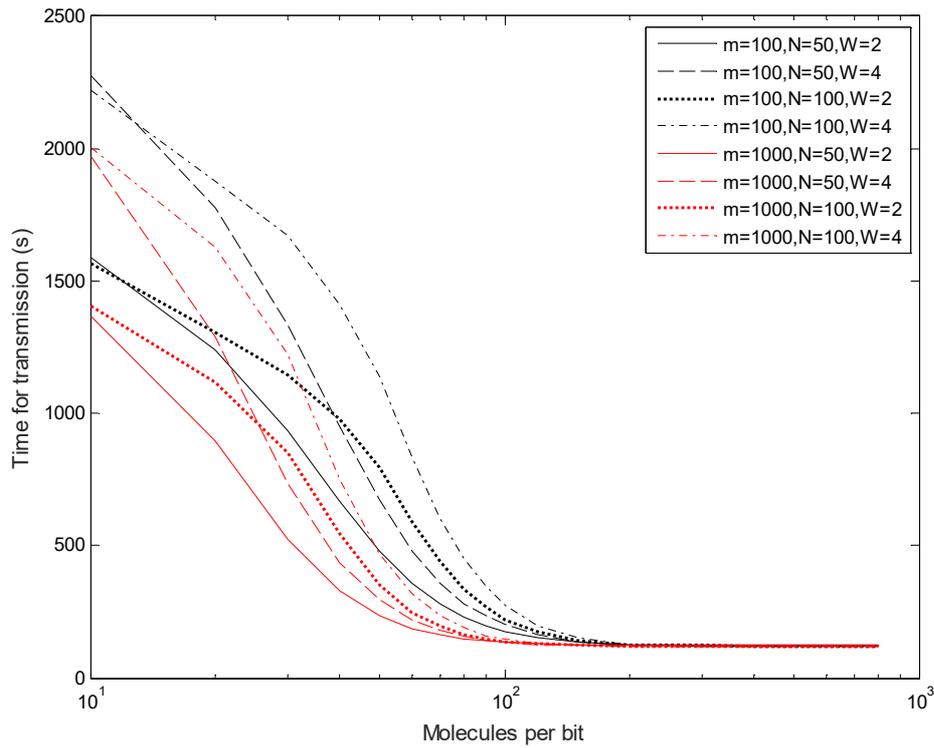


Figure 6.27 Transmission time of SR-ARQ for different combinations of the main factors discussed in the text.

Figure 6.28 and Figure 6.30 show the transmission efficiency and channel throughput of the communication system for SR-ARQ scheme, respectively, with different parameter combinations used. It should be noted that the efficiency and channel throughput are analysed together due to the uniform tendency of the curves. It may be seen that before the transmission efficiency or channel throughput achieves stability ($n < 200$), the channel performs better with a larger bacterial population in the receiver, regardless of the other two parameters, achieving a larger transmission efficiency and channel throughput. Also, when m is a fixed value, both the transmission efficiency and the channel throughput are larger with a smaller frame size, no matter what the value of W is. Moreover, for the condition when the bacterial population in the receiver is a fixed value, the transmission efficiency and channel throughput of the channel with larger frame size increases faster; and will be

even faster when the window size is larger. It should also be noted that the initial increasing points are almost the same for different channels when the corresponding values of W and m are identical. On the other hand, if the number of molecules per bit is larger than 200, the channel with parameter combination $N = 100, W = 4$ achieves the highest steady state values, which is 99% for transmission efficiency and 16.64 *bps* for channel throughput; and the channel with parameter combination $N = 50, W = 2$ achieves the lowest steady state values, which is 96% for transmission efficiency and 16.04 *bps* for channel throughput, regardless of the receiver bacterial population.

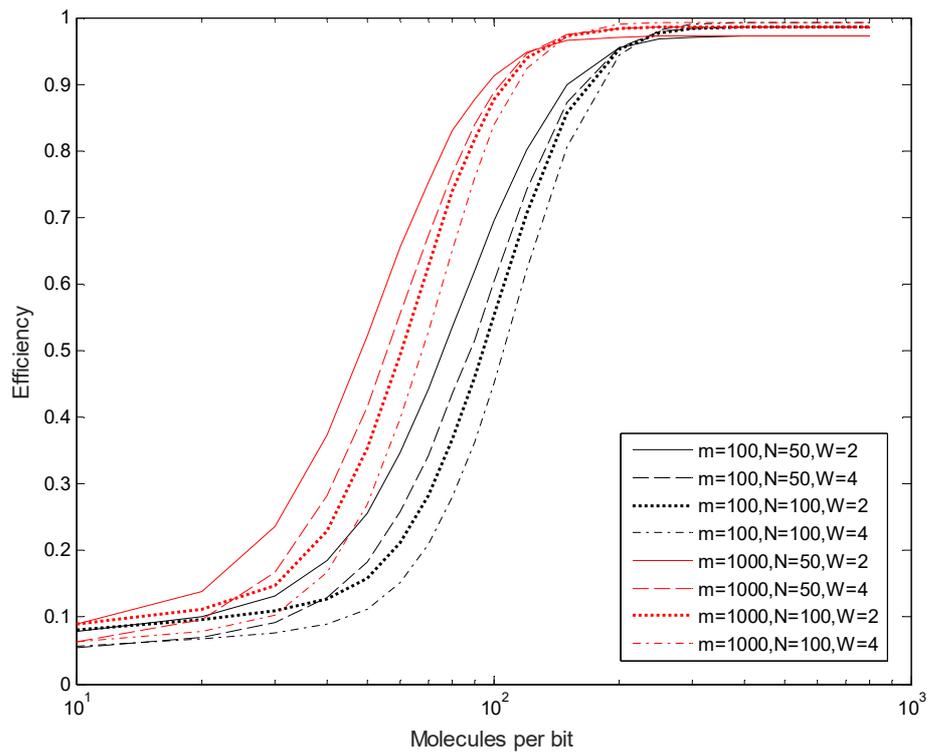


Figure 6.28 Transmission efficiency of SR-ARQ for different combinations of the main factors discussed in the text.

Figure 6.29 shows the error performance of different combinations of three factors for SR-ARQ scheme. Results show that the error probability varies mainly because

of the bacterial population in the receiver and the frame size. Specifically, with a larger value of m and smaller value of N , the error probability will be smaller, regardless of the window size. The significance of the three parameters to the BER of the channel is in the decreasing order of m, N, W .

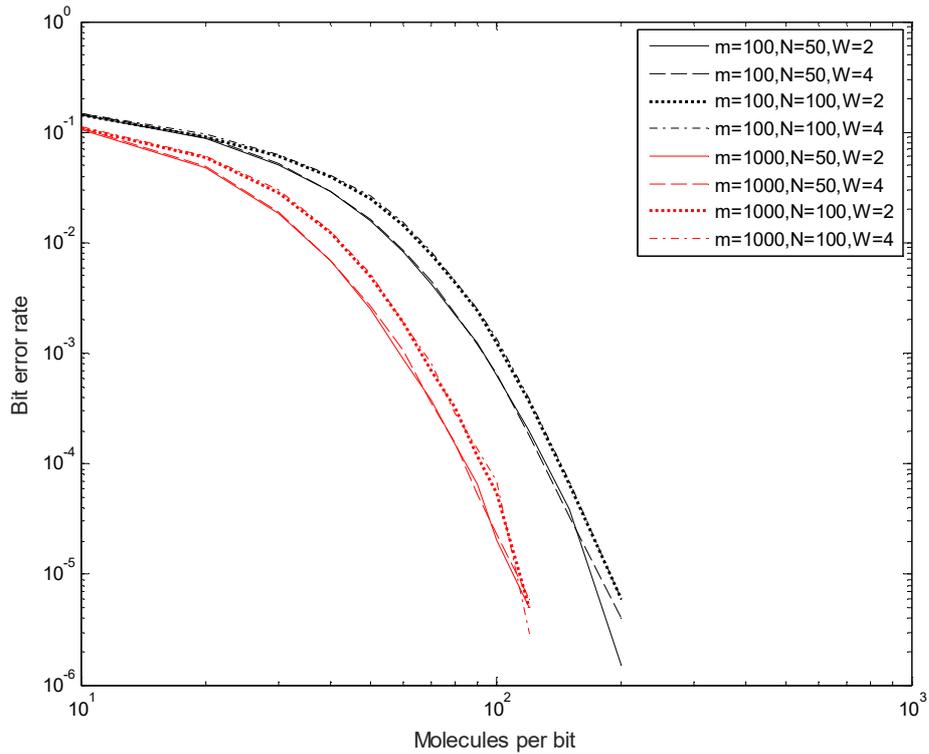


Figure 6.29 Error performance of SR-ARQ for different combinations of the main factors discussed in the text.

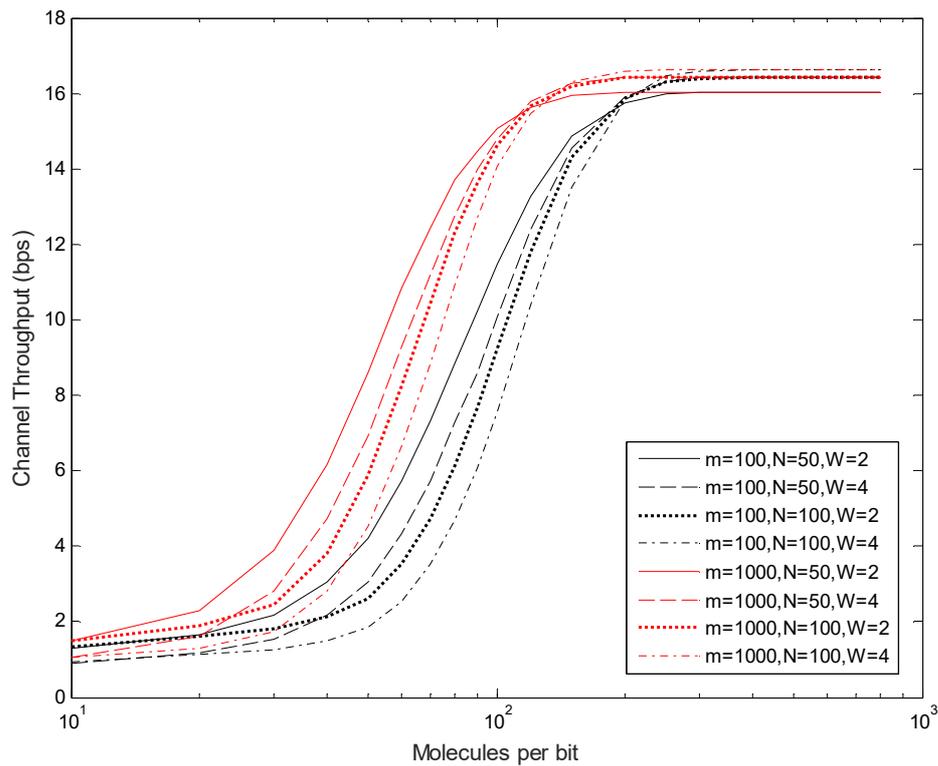


Figure 6.30 Channel throughput of SR-ARQ for different combinations of the main factors discussed in the text.

6.7 Comparisons of different ARQ schemes

In the previous sections, the simulation results of the three basic ARQ schemes including SW-ARQ, GBN-ARQ and SR-ARQ have been given separately, in terms of different parameters, which include the transmission distance, the bacterial population in the receiver, the frame length, the window size and the CRC polynomial. In this section, to make comparisons between the three ARQ schemes, the channel parameters are set to be the same, including a transmission distance of $4\mu m$, 100 bacteria in the receiver, a 100 bit frame length, a window size of 4 and CRC-8 ; the results are shown in Figure 6.31.

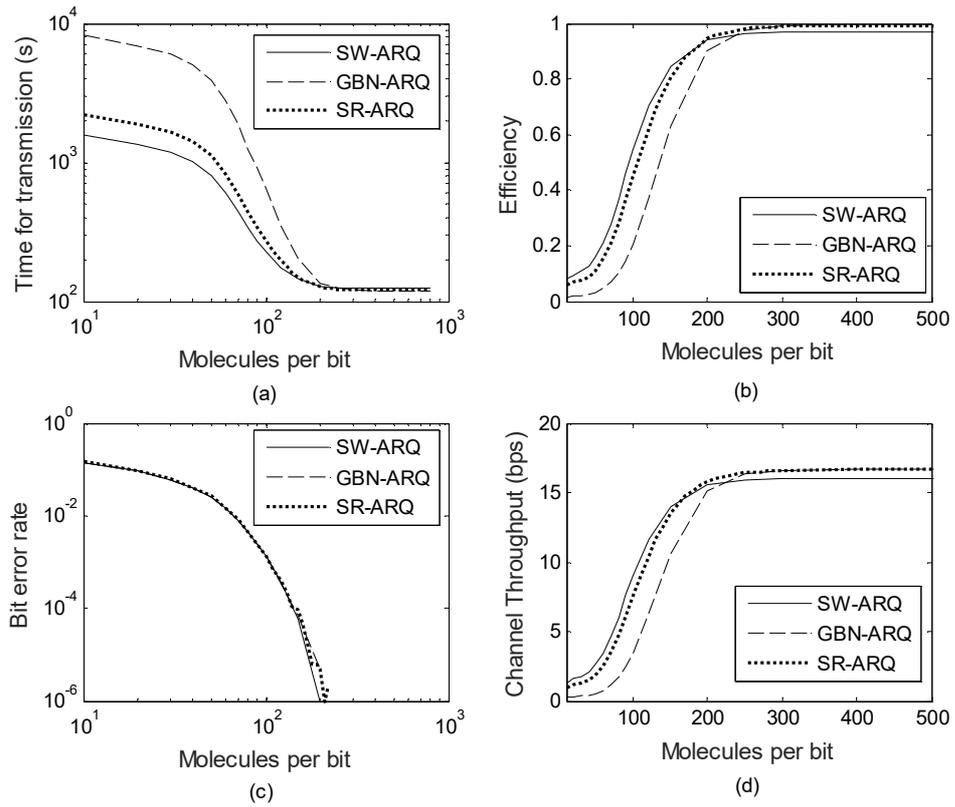


Figure 6.31 Comparisons of channel performance for different ARQ schemes: (a) transmission delay; (b) efficiency; (c) BER; (d) channel throughput.

Figure 6.31 shows that when the number of molecules per bit is in the range between 10 to approximately 250, SW-ARQ and SR-ARQ take less time and have larger transmission efficiencies and channel throughputs. Also, SW-ARQ has a slightly better performance than SR-ARQ. According to the non-ARQ BER performances shown in Figure 4.5, when the number of molecules per bit is in a smaller range, the one bit error 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 (>250), the one bit error 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, particularly 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. 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, the ARQ performances varies according to the range of the number of emitted molecules at the start of each time slot. Specifically, when the number of molecules per bit is in a smaller range, SW-ARQ performs better than SR-ARQ, and GBN-ARQ performs worst. However, when the number of molecules emitted at the start of each time slot is relatively large, GBN-ARQ and SR-ARQ have similar channel performances, while SW-ARQ performs worst.

Here, it should be noted that the conditions when the largest frame size of each ARQ protocols are selected, specifically $W = 1$ for SW-ARQ, $W = 7$ for GBN-ARQ and $W = 4$ for SR-ARQ, are simulated as well. But the results are similar to those shown in Figure 6.31, so are not plotted.

6.8 Summary

In recent years, bacteria have been considered as a promising approach for MC, through the process of QS. In the previous chapters, a diffusion-based bacterial

communication network model between two populations of bacteria has been established; and the error correction techniques have been applied to enhance the system reliability. In this chapter, to ensure in-sequence delivery of transmitted frames, which is not achievable using channel coding techniques, for the first time, the widely used ARQ protocols have been employed, specifically with the utilization of CRC coding techniques, as well as SW-ARQ, GBN-ARQ and SR-ARQ protocols. For SW-ARQ, results show that the diffusion channel has better transmission efficiency and channel throughput with smaller transmission distance and larger bacterial population in the receiver. In addition, less complex error detection codes have better performance, leading to smaller transmission time and error probability, as well as larger transmission efficiency and channel throughput. However, all the channel performances, including the transmission time, efficiency and channel throughput, have crossover points when varying the frame length, which means that there should exist an optimal frame size to minimise the transmission time, or to maximise the efficiency and channel throughput. Also, after having investigations of the four parameters individually and separately, the combinations of them are taken into consideration to decide the degree of their influences on the channel performance. Results show that when the number of molecules per bit is in a smaller range (fewer than approximately 150), the transmission time mainly depends on the CRC polynomials, which means that much time will be consumed with a higher order CRC polynomial. However, when the number of molecules per bit is in a larger range ($n > 150$), the transmission time will be mainly decided by the transmission distance and the frame size. For the transmission efficiency, there exist three stages. In stage 1, when the number of molecules per bit is below approximately 100, the transmission efficiency is mainly decided by the CRC

polynomial. In stage 2, which is the steady state, the efficiency will be depend on the frame size, which means that same steady state values can be achieved with the same frame lengths. The significance to the transmission efficiency of the four factors will be in the decreasing order of frame length, number of bacteria in the receiver node, transmission distance, and CRC polynomial. In stage 3, which refers to the period between stage 1 and stage 2, to achieve the same transmission efficiency, the required number of molecules emitted at the start of each time slot is the most for the condition of $d = 8\mu m, m = 100$; while the required value of n is the smallest when $d = 4\mu m, m = 1000$. For the channel error probability, it is mainly up to the CRC polynomial used. According to the conditions which have the same error detection codes, the dependency is in the decreasing order of transmission distance, bacterial population and frame length. Finally, for the channel throughput, it can be divided into two stages. In stage 1, the significance to the channel throughput of the four factors will be in the decreasing order of transmission distance, number of bacteria in the receiver node, frame length and CRC polynomial. Also, results show that the initial growth point mainly depends on the CRC polynomial. In stage 2, the channel throughput levels off. The system can achieve identical steady states with same parameter combinations of transmission distance and frame length. For a particular communication scenario, the transmission distance is fixed, leading to the choice of $4\mu m$ as a feasible and representative value for GBN-ARQ and SR-ARQ simulations. Given the increasing complexity and delay from high order CRC polynomial operations, which has been proved in SW-ARQ scheme, CRC-8 is selected when GBN-ARQ and SR-ARQ are considered. 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, upon varying the frame length, 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 60 to 250 for SR-ARQ, smaller frame lengths perform better, with reduced time consumption and increased channel throughput. In addition, the transmission efficiency with larger frame length 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 250 for SR-ARQ.

By applying different combinations of three parameters, including the bacterial population m , the frame length N and the window size W , in both GBN-ARQ and SR-ARQ schemes, further results have been given. Results show that for GBN-ARQ, the window size W has the biggest influence on the transmission time, followed by the receiver bacterial population and the frame size when $n < 250$. In addition, when $n < 250$, the significance to both the efficiency and channel throughput is in the decreasing order of m, N, W ; while for the condition when $n > 250$, the channel with a larger window size can achieve a higher steady state value of efficiency and channel throughput. For SR-ARQ, when the number of molecules per bit is smaller than approximately 200, the transmission time mainly depend on the window size. Also, in this range of $n < 200$, the significance of the three parameters to the transmission efficiency and channel throughput is in the decreasing order of m, N, W . In addition, for the condition when $n > 200$, the transmission time has little change when varying the channel parameters. However, in this range, the steady state values of both the transmission efficiency and channel throughput reach the maximum when $N = 100, W = 4$, while the values reach the minimum when $N = 50, W = 2$. Moreover, for both GBN-ARQ and SR-ARQ schemes, the error performance is

influenced by the bacterial population in the receiver node at the most extent, followed by the frame length and window size.

Finally, with identical parameter settings, the performance of the three ARQ schemes has been investigated and compared, indicating that when the number of molecules per bit is in a smaller range, SW-ARQ performs better than SR-ARQ, and GBN-ARQ performs worst. However, when the number of molecules emitted at the start of each time slot is relatively large, GBN-ARQ and SR-ARQ have similar channel performances, while SW-ARQ performs worst; 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. It should be recognized 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 [252] to create complex coding, windowing and sequencing operations. Nevertheless, it is considered 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, the optimization of the frame length to achieve the best channel throughput and the investigation of the energy consumption will be studied in the next chapter.

CHAPTER 7.

THROUGHPUT AND ENERGY EFFICIENCY BASED PACKET SIZE OPTIMIZATION OF ARQ

7.1 Introduction

In Chapter 6, ARQ protocols in molecular link layer and error detection mechanisms, in particular, CRC codes, have been applied to the proposed bacterial quorum communication system. The performances of SW-ARQ, GBN-ARQ and SR-ARQ have been presented and analysed. Due to the fact that CRC codes can be implemented through mathematical logic operations [202], this chapter intends to map synthetic biological components in the emerging discipline of synthetic biology to the implementations of ARQ protocols proposed in Chapter 6. In addition, results in Chapter 6 indicate that there exist optimal frame lengths to maximize the throughput and energy efficiencies. Thus, there are two foci in this chapter: (a) establishing a novel energy model that combines synthetic logical functions with the operation of coding and protocols; (b) obtaining the optimal frame lengths for different ARQ schemes to achieve better throughput and energy performances and to resolve the issue that the bio-entities are constrained by information transmission

rate, channel capacity and energy limitations [38]. Although there have been several studies on packet size optimisation in traditional wireless and wired networks [253], none of them are directly applicable to the bacterial scenario.

Synthetic biology, which is generally considered as the engineering of biology, involves the construction of sophisticated and biological inspired components, exhibiting functions that do not exist in nature. The improvements of existing biological systems and the creation of artificial ones are critical objectives of the emerging discipline of synthetic biology [254]. Considering this, synthetic biology is characterised by two aspects, specifically on one hand to improve and re-design parts, genomes and natural biological systems of existing microbes which can perform specific functions, and on the other to design and engineer new biologically based parts (e.g., promoters, transcriptional regulators/terminators), devices (e.g., inverters, toggle switch, logic gates and oscillators) and systems (e.g., programmed pattern formation, biofilm formation, and controlled invasion of cancer cells), through the approaches of abstraction, decoupling and standardisation [254]. Re-design of existing systems includes transcriptional, translational and post-translational parameters, and standard optimisation and control engineering methods to find the optimum operational parameters to achieve the desired objective [254]. Significant progress has been made in inventing well-characterised biological components based on basic science, from which standard molecular parts (e.g., nucleic acids and proteins) can be constructed and then assembled in a controlled manner to predict their behaviour. Synthetic biological components design depends on tools from multiple disciplines including genetic engineering, bioengineering, systems biology and many other engineering fields, with methods and criteria such as standardization, abstraction, modularity, predictability, reliability, and uniformity [255]. The majority

of the foundational research in this field was carried out within microbial species *E. coli*, which has been widely used in various relevant areas such as biological circuit design and metabolic engineering due to the deep mechanistic understanding of its biology and its ease of genetic manipulation [256]. A great deal of high-profile scientific success in DNA synthesis techniques, high-throughput DNA sequencing, large-scale metabolic and signalling networks modelling in which the forward-engineering principles of synthetic biology converged with basic research [256], enables the rapid and efficient development of synthetic biology [257]. The main challenge in this field is the selection of well-matched synthetic components that when coupled, reliably and robustly produce the desired behaviour. Difficulties in this area lead to difficulties in achieving the predictability and rapid iteration of design [256]. For example, synthetic biological components are not independent objects built in the absence of a biological milieu; instead, they typically function within a cellular environment or environmental context in which they were originally characterized [179], leading to the probable functional failure when placed into circuits with other parts [256]. When engineering synthetic biological devices, the components, resources and machinery of host cells will be used, resulting in the modification of not only the intended resources, but also the cells themselves. However, at the present stage, due to the lack of knowledge and established methods about inherent cellular characteristics and functionalities, it is almost impossible to completely predict the functions of even simple synthetic components in engineered cells [179].

The rest of this chapter is organized as follows: Section 7.2 gives a brief introduction to engineered MC systems, including the synthetic biology hierarchy in Section 7.2.1, the technical advances in relevant fields in Section 7.2.2, the typical synthetic

biological components examples in Section 7.2.3 and a brief discussion in Section 7.2.4. Also, a novel bacterial communication energy model for ARQ protocols has been established in Section 7.3, followed by the mathematical analysis of both throughput and energy efficiency of ARQ techniques in Section 7.4. The numerical results and analysis are given in Section 7.5. Finally, Section 7.6 provides a summary for this chapter.

7.2 Engineered molecular communication

The discovery of mathematical logic in gene regulation in the 1960s and advances in genetic engineering in the 1970s, for instance, the DNA recombinant and DNA sequencing technology [179], laid the foundation for the emerging discipline of synthetic biology. It is a new area enabling the design and manufacturing of new and unconventional biological systems [179]. In contrast to conventional genetic engineering approaches that resolve problems mainly focusing on modifying a few genes, synthetic biology deals with these issues from a novel engineering perspective [258]. Synthetic biology has mostly been concerned with the invention and perfection of genetic devices and small modules which are constructed by combining them. The design and construction of complex synthetic biological components or systems provide new methods for investigating and modifying natural phenomena or processes and has various applications [179]. In this section, an overview of synthetic biology is presented, covering the catalogue of emerging techniques and some examples of synthetic biological components.

7.2.1 Synthetic biology hierarchy

The synthetic biological systems are inspired by natural existing biological processes or engineered in vitro (DNA strands) or in vivo (bacteria) laboratory practices simulating different computational functions. For example, Boolean logic gates, as a

particular class of models in computer science, have been studied and designed unconventionally as genetic devices [252, 259]. Synthetic biology is expected to have a significant potential to change the way to interact with the environment and to approach human health, by manufacturing practical organisms which can be used to clean hazardous waste in unreachable places [260], to sense chemicals and respond accordingly [261], to produce fresh fuel in an efficient and sustainable fashion [262] or to recognize and destroy tumours [262]. According to [179], a useful analogy between the computer engineering hierarchy and the hierarchy of synthetic biology can be formed as shown in Figure 7.1 to conceptualize both the goal and methods of synthetic biology in a layered architecture. Within the hierarchy, each constituent part is embedded in a more complex system that provides its context. Synthetic biology can be considered as a bottom-up approach, which has been discussed in Chapter 1, to study biological systems by assembling lower layer components to form upper layer devices or systems.

As shown in Figure 7.1, at the bottom of the hierarchy are the biological building blocks such as DNA, RNA, proteins, and metabolites (including lipids and carbohydrates, amino acids, and nucleotides). These biological components are analogous to the physical layer which includes transistors, capacitors, and resistors in the computer engineering hierarchy. The upper layer, the device layer, comprises biochemical reactions that regulate the flow of information and manipulate physiological processes, equivalent to engineered logic gates that perform computations on a computer. Then, at the upper module layer, the assembling of complex pathways (or modules) that function as integrated circuits in a computer is implemented using a diverse library of biological devices. In the upper layer, the interconnection of pathways and their integration into host cells allow the synthetic

biologists to extend or modify the behaviour of cells in a predetermined and programmatic fashion. The concept of programmable cells inspires researchers to achieve innovative solutions to currently unresolved problems. For example, bacteria can be used as a living computational therapeutic tool to break tumour immune tolerance [262]. Upon the simultaneous detection of two conditions with the utilisation of a two-input logical AND gate, engineered bacteria will invade and kill tumorous cells. Correspondingly, in the computer engineering hierarchy, assembling integrated circuits can be interpreted as a programmable computer. Finally, similar to the case of computer networks, although different tasks of varying complexity can be performed with the implementation of independently engineered cells, more sophisticated coordinated tasks are possible with populations of programmable cells in the form of tissues or cultures. Furthermore, in a synthetic biological system, intercellular MC is also employed not only to make each cell more predictable and reliable [36], but also to enable the coordination of tasks across heterogeneous cell populations to elicit highly sophisticated behaviours [179]. Therefore, it is reasonable to design multicellular systems for the achievement of overall stability and reliability in performing complex tasks. Similarly, in computer engineering, interconnections among computers create a computer network.

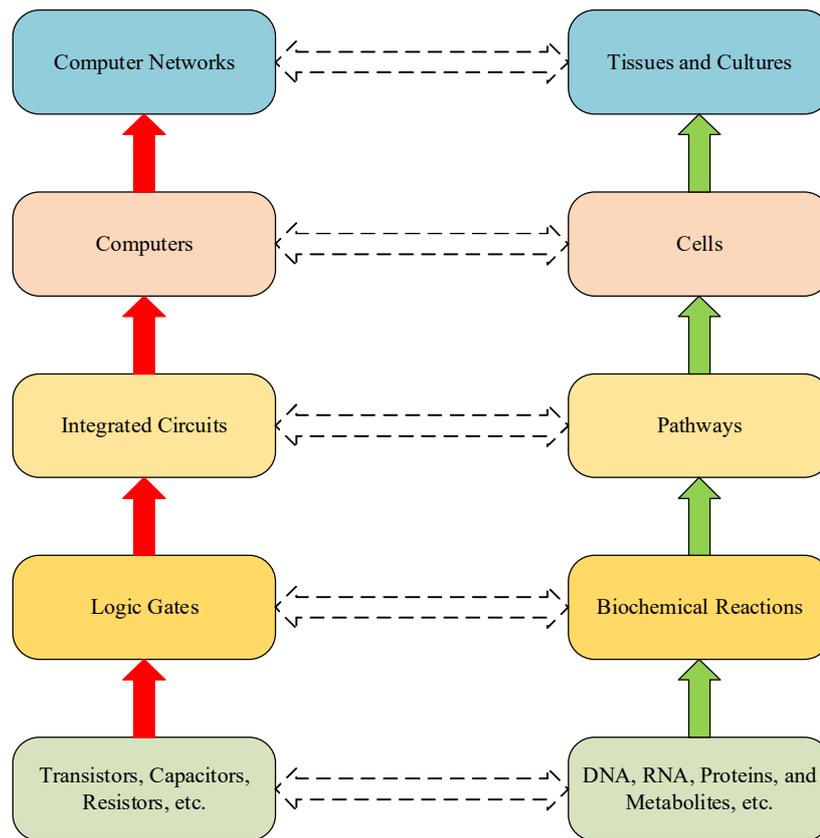


Figure 7.1 A hierarchal analogy between computer engineering and synthetic biology [179].

However, it should be noted that there exist some design problems when building biological systems. For example, a number of impact factors, including gene transcription and translation noise, gene mutation, cell death, undefined and constant changeable extracellular environments, and interactions with cellular context, may result in the limited capability of adequately predicting the functions of even simple devices in engineered cells [179]. Also, many biological parts have not been well defined, i.e., not standardised and quantitatively characterised, resulting in that their performance varies with distinct kinds of cells or under different environmental conditions or configurations. This makes the biological components lack modularity and reusability. In addition, many synthetic elements are incompatible, in other

words, once constructed and integrated into cells, they can have unintended effects on their hosts or interfere with existing elements in the engineered cell.

7.2.2 Technical advances

As stated before, synthetic biology can be described as the design and construction of novel biological organisms by integrating well-defined and reproducible essential biological ‘parts’, which is in correspondence with the bottom layer in Figure 7.1, (e.g., promoters, coding sequences, terminators, etc.) into genetic ‘devices’ (e.g., transcriptional units) in the second layer from bottom in Figure 7.1, which will then be assembled into basic genetic ‘modules’ (e.g., biochemical pathways, genetic circuits, etc.) [263]. Here, the upper three layers in Figure 7.1 can be considered as modules. Therefore, parts perform as the basic unit for synthetic biological systems.

Genetic engineering is one of the great advancements for synthetic biology, allowing the use of specific DNA fragments to modify existing cellular functions through techniques such as DNA recombination, DNA sequencing and synthesis and DNA amplification [264]. Among them, DNA synthesis technologies allow the creation of entire genomes, with high efficiency and low cost [254]. In synthetic biology, custom-made DNA can be used to build larger DNA segments, which can be further assembled into larger fragments until the desired DNA product is obtained [265]. Two of the most commonly used methods are BioBricks™ (or standard assembly) [148], which is a standard for interchangeable parts, and Gibson Assembly® [266], which allows for successful assembly of plentiful DNA fragments in spite of the fragment length or end compatibility [266], complemented by other systems such as GoldenBraid [263], which makes use of GBparts (fragments of DNA with four-nucleotide overhangs) as minimal standard building blocks and has been shown to serve as a modular assembly system in plant synthetic biology [254]. Other methods

include CPEC (Circular Polymerase Extension Cloning), Golden Gate, SLIC (ligation-independent cloning) and so on [254].

In addition, standardisation is necessary for precise reproduction of synthetic biological devices and modules, demanding fully characterization of parts. The advance of standardized biological components allows for the rapid assembly of sequences. However, this has not reached a sufficient state of maturity to date. Specifically, a lack of standardization in both gene assembly methods and the modules themselves is a current constraint for the efficient advancement of synthetic biology. Generally, parts need to be characterised in a specific genetic or environmental circumstance, and do not function in a predictable manner when separated from this environment. Standard assembly parts should provide a paradigm that addresses two fundamental aspects: (a) the mechanism of designing sequences that impart a particular function; (b) the way to construct such functions using DNA encoding that can be readily introduced into the cell, by recognizing that functional units of DNA sequence (e.g., promoters, ribosome binding sites) are frequently reused as non-reducible elements of genetic composition, or ‘parts’, in a variety of projects [267]. Each primary genetically engineered part has been designed not only associated with one particular assembly standard, but also able to comply with any of several publicly-available norms [267]. Currently, one significant standard assembly is the BioBricks™ standard, which was the first attempt at defining synthesis rules, allowing the assembly of standard biological parts using a single assembly chemistry [148]. The BioBricks Foundation [268] was formed in 2006 to standardize biological parts across the field, and to make sure that the engineering of biology is implemented in an open and ethical manner. A BioBricks™ standard biological part is a nucleic acid encoded molecular biological function, along with the associated

information defining and describing it [269]. Also, for a given part, there exists a set of restriction sites, allowing standard cloning and construction of a range of parts to form a synthetic gene unit such as a hybrid promoter [270], which makes it possible that all the parts can be shared and combined without compatibility issues.

To make use of parts efficiently, the requirement for a professional registry of parts was identified [271]. BioBricks™ parts are made available to synthetic biologists through the Registry of Standard Biological Parts, hosted at MIT (<http://parts.mit.edu>). The registry provides various types of biological parts, including promoters, ribosome binding sites, protein domains, protein coding sequences, translation units, terminators, DNA, plasmid backbones, plasmids, prime and composite parts, as well as devices such as protein generators, reporters, inverters, receivers and senders, and measurement devices [272]. The central object of the Registry is to catalogue a library of fully characterized biological modules, from individual genes or parts of genes to complete gene circuits, developing and providing educational and scientific materials to be accessed by the public and to improve existing standard parts and contribute new BioBricks™ standard biological parts [273].

7.2.3 Synthetic biological examples

Genetic devices can be established by combining parts that implement a defined function. Various electronic circuitry inspired synthetic biological devices, such as genetic toggle switches, timers, oscillators and logic evaluators [258], have been produced and their functioning is based on the control of transcription, translation or post-translational processing. Sophisticated devices can be assembled from well-documented parts, but it remains a challenge to predict their behaviour entirely. Also, the assembled devices may be less robust than natural systems, and endogenous

regulatory systems may interfere with the function of synthetic biology devices [274]. In this section, some recent advances in the field of synthetic biology will be explored, with particular emphasis on the development of biological modules for the regulation of artificial devices.

A. *Transcriptional control devices*

Generally, the functionalities of biological systems can be optimised through transcriptional control of gene expression, for example, by modifying the spacer region between DNA sequence of native promoters, or by utilising polymerase chain reaction (PCR) to introduce mutations into the entire promoter region [254]. A synthetic hybrid promoter, which combines the core promoter elements with enhancer elements consisting of tandem repeats or combinations of upstream activating sequences (UAS), can be applied to the creation of powerful promoter libraries [275]. Promoter strength can differ depending on flanking sequences upstream and downstream of the consensus boxes and promoter copy number. A genetic toggle switch was constructed in 2000 [276], where a one-bit memory was implemented with two genes inserted into a bacterium. Its bistability was achieved by the two genes under the control of two promoters that are mutually repressed by the product of the other gene, which is displayed in Figure 7.2. One promoter also transcribed a reporter gene, such as the green fluorescent protein (GFP), the expression of which could be switched on by the transient addition of the first inducer and switched off by the transient addition of the second inducer. The system was designed to inhibit the activity of one repressor over the other through an external stimulus and to push the system to one stable state, where one gene was repressed and the other fully transcribed. Generally, the establishment of these

synthetic components includes the processes of quantitative design, physical construction, experimental measurement and final debugging [256].

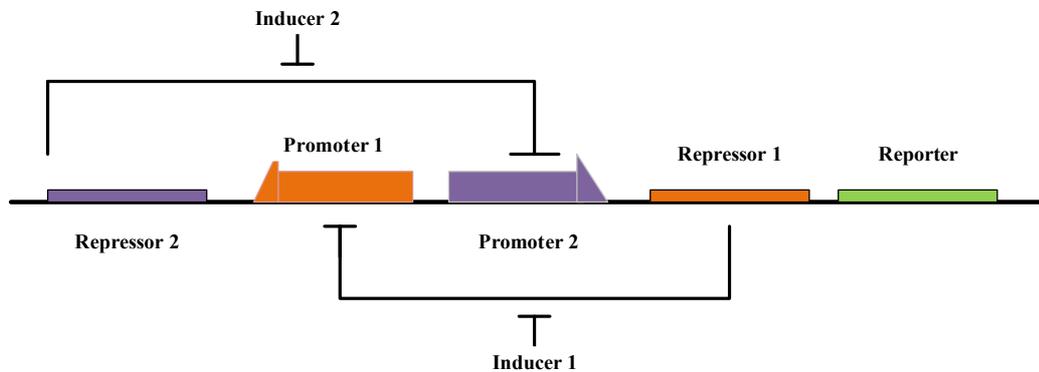


Figure 7.2 Toggle switch design.

Each promoter can be induced by the corresponding inducer, and repressed by the relevant promoter. In the absence of inducers, two stable states can be achieved: (a) Promoter 1 transcribes Repressor 2; (b) Promoter 2 transcribes Repressor 1. Transiently introducing an inducer of the currently active repressor can result in switching [276].

In addition, more complex assemblies of promoters and repressors exist, for example, transcriptional oscillators. The authors in [80] employ three transcriptional repressor systems in a ring topology that do not belong to any natural biological clock, to construct an oscillating network, known as the repressilator, in *E. coli*, exhibiting damped oscillatory behaviour in terms of luminescence output in individual cells. Specifically, a bacterium possesses inserted DNA sequences that cause the oscillation of the concentration of three protein products (TetR, cI, LacI) to oscillate within it over time. The DNA sequences to be inserted into the bacterium were selected so that TetR blocked the promoter sequence of cI, which blocked the promoter sequence of LacI, and this blocked the promoter sequence of TetR. As shown in Figure 7.3, the genetic circuit involved a triple negative feedback loop including three repressors, each repressing the transcription of another in the cycle. It

is obvious that both the toggle switch and repressilator were constructed using a similar range of synthetic parts, in particular inducible promoter systems, and employed luminescence production as the output to monitor the circuit.

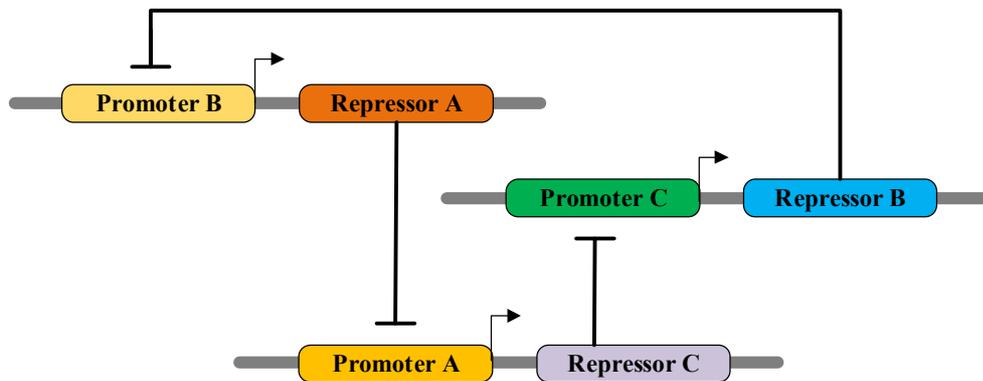


Figure 7.3 Oscillator design.

A circular network of three transcriptional repressors, each repressing the expression of the next in the series, acts as the basis of the genetic ring oscillator, or repressilator [80].

Another significant function of transcriptional control is signal integration, i.e. processing multiple inputs and producing well-defined outputs. An outstanding operational example that performs integrations of signal elements is represented as logic gates. The authors in [277] have engineered some computational building blocks based on DNA transcription and translation processes. An example building block is a biochemical inverter, where the input mRNA generates a repressor protein that prevents DNA transcription processes, producing no output mRNA. Similarly, signal transduction pathways in eukaryotic cells have been modified by synthetic signalling proteins to demonstrate logic behaviours such as AND and OR gates. Synthetic logic gates can be connected to execute complex biological tasks as in electrical circuits [278]. Genetic logical devices can communicate with each other through variations in gene expression and reaction. For example, upon the

stimulation of a sensor, a promoter will be activated, which can be considered as the input to a circuit [279]. Genetic logic gates can be regarded as the core of the mechanism of synthetic biology and a necessity for generating genetic logic systems, combining electronic engineering with cell biology.

Since the genetic logic gates will be applied to optimise the performance of the proposed channel in this thesis, further description is warranted here. SynBioLGDB is a resource for experimentally efficient browsing and visualization of genetic logic gates in synthetic biology [25]. It may enable more comprehensive understanding of the connection of genetic logic gates to execute complex cellular functions in living cells. The current version of SynBioLGDB contains more than 189 genetic logic gates with experimental evidence involving 80 AND gates, 16 NOR gates, 8 Buffer gates, 10 NAND gates, etc. in three species (Human, *E. coli* and *Bacillus clausii*). Each entry contains detailed information of a logic gate, including gate category, input symbols (e.g., gene, protein, and promoter), output symbols (e.g., gene, protein, and promoter), species, validated method, PubMed Identifier (PMID) and detailed descriptions [25].

Despite the existence of the individual synthetic biological devices, little works has been completed to investigate cell-cell communication using synthetic intercellular circuits, achieving coordinated behaviours among a cluster of cells or different species. Most of these synthetic intercellular communication systems employed QS mechanisms, for example a pattern formation system in *E. coli* involving LuxI, LuxR, an enzyme which controls the synthesis of AHL, a transcriptional activator dependent on AHL, LuxI promoter and GFP [280]; these studies make it possible to practically implement the model proposed in this work. Other synthetic intercellular systems include the transient pulse generating circuit [281], the population control

circuit that regulates the *E. coli* population density autonomously [282], the engineered cells that respond to biological signals in a programmable fashion [146] and so on.

B. *Posttranscriptional control devices*

As stated in the previous subsection, inducible promoters have been used to generate complex genetic circuits with diverse operation modes such as toggle switches, Boolean logic gates and oscillators. Also, the transcriptional level of inducible promoters can be adjusted by using compounds that are chemically related to the inducer molecules [283]. Although transcriptional control biological systems have been numerous and well-studied, few systems exist for posttranscriptional control, which mainly control target protein synthesis from transcribed mRNAs [284]. Typical elements of posttranscriptional control are synthetic ribozymes [285]. In these RNA devices, RNA-based gene regulation can repress gene expression in a programmable manner through either targeted mRNA degradation or inhibition of translation of the transcript. Within the last few years, a new class of synthetic devices, engineered riboswitches, came into focus, which refers to RNA devices that respond to various cell-permeable small molecular metabolites (e.g., to control chemotactic gene expression and guide the bacterial moving direction) [286]. Also, RNA devices that respond to small interfering RNAs (siRNAs) are called riboregulators [285].

7.2.4 Discussions

It has been stated previously that advanced synthetic biological techniques and improved understanding in cell and molecular biology have made it possible to design and engineering MC systems. A standard approach for engineering MC systems is to extend or modify existing biological systems, which can be

implemented with the utilisation of synthetic biology. Specifically, transceiver nanomachines, which are required to have the communication capabilities to synthesize, store and release molecules, to encode molecules into frames, to window and sequence frames, and to capture and react to specific signalling molecules, can be developed by synthetic logical components. For example, in [280], transmitter nanomachines are constructed and engineered to synthesize and release AHL molecules using specific metabolic pathways. Receiver nanomachines are also engineered which can respond to the signalling molecules by synthesizing specific reporter proteins depending on concentration. The AHL molecules are membrane permeable and can freely diffuse between transceivers [280]. The synthetic components presented in Section 7.2.3 can be used to build MC transceivers, as well as to increase their complexity according to their functions. For example, for the proposed model in this thesis, functions including error detection coding, frame buffer and receiver responses to received information molecules can be implemented using synthetic logical functions.

7.3 Energy model

In a conventional wireless communication systems, the energy budgets of the transmitters and receivers result in a constrained performance of the system. The energy, usually provided by limited battery power, should be used efficiently by the communication system to maximize the system performance while attaining acceptable operational device lifetime values [38]. 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 work, 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 and constructed to supply all the essential nutrients in solution for bacterial growth. During the communication process, protein and mRNA molecules have to be continuously synthesised to counter their degradation, which consumes power [287]. ATP provides the energy for the synthesis since the its hydrolysis releases energy, and it is the universal currency of energy transfer between cells in living organisms [287]. 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 approximately $20k_B T$ of energy, where k_B is Boltzmann's constant and T is the absolute temperature [288]. 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 [287]. Similarly, the energy cost of synthesizing a protein molecule in ATP is $E_{\text{prot}} \approx 50L_{\text{prot}}$, where L_{prot} represents the length of the protein molecule in amino acids. It must be stated that at this point that these estimates are speculative and the energy consumption in bacterial communications has yet to be fully investigated. It has been stated above that CRC codes are used as error detection codes in the ARQ implementations. In traditional communication areas, CRC codes can be implemented through logical operations [202]. For instance, the logical representation of the systematic encoder and decoder of CRC-8 are given in Figure 7.4.

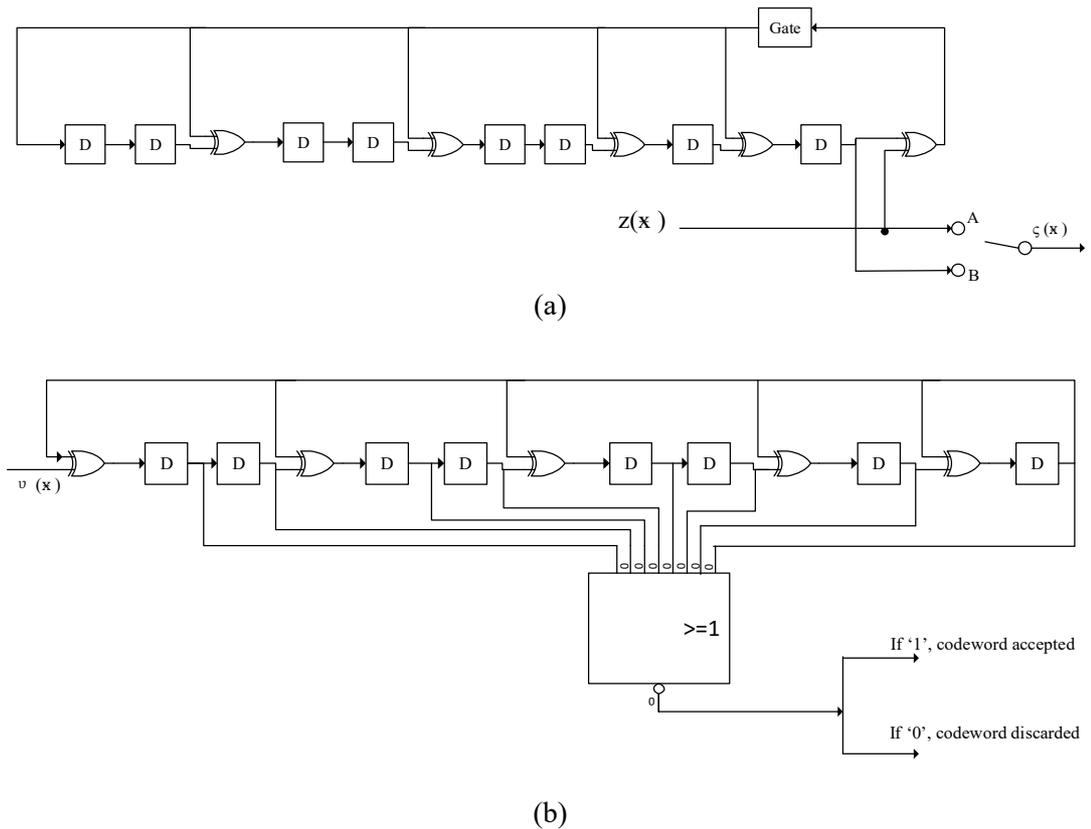


Figure 7.4 Synthetic encoder and decoder for CRC-8.

(a) Systematic encoder for CRC-8, where $z(x)$ is the information bits and $\zeta(x)$ is the code word; (b) Systematic decoder for CRC-8, where $v(x)$ is the received code word.

It is apparent in Figure 7.4 that the logical operations may be reduced to an arrangement of XOR gates, NOR gates and shift registers. Thus, the first stage in producing the energy estimate is to get an insight into the ATP cost of implementing the logical operations required for the encoder and decoder. In this paper, the synthetic genetic logic gates constructed in [252] are considered to design the energy model. The four logic gates which will be used to develop the energy model are displayed in Figure 7.5.

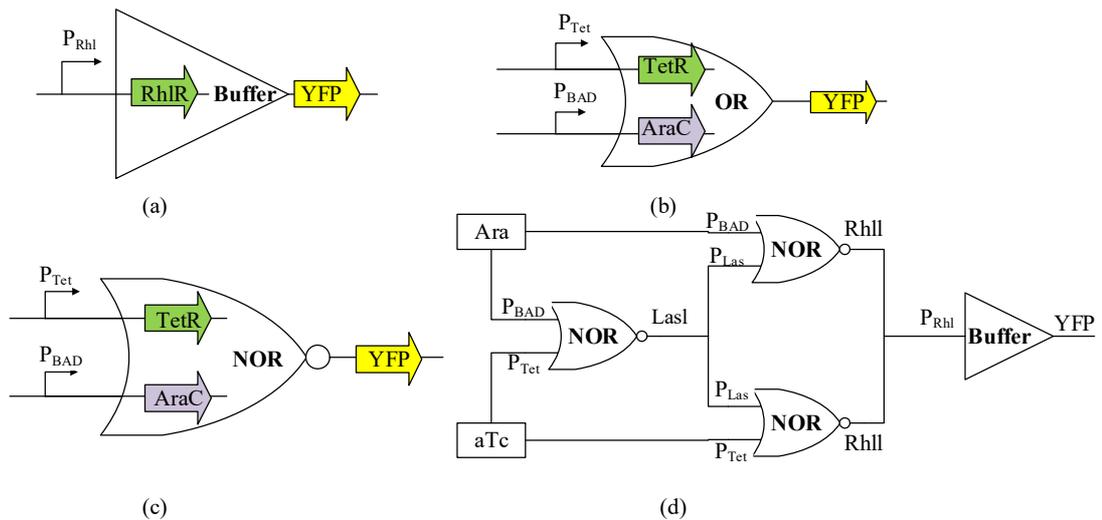


Figure 7.5 Four synthetic logic gates employed to establish the energy model.

(a) Buffer gate with P_{RhI} (downstream gene *rhIR* and corresponding protein RhIR) 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 the quorum signal LasI [252].

NOR gates are functionally complete, which means that any computational operation can be implemented by layering them. In Figure 7.5(d), the XOR gate is made up of three NOR gates and one buffer gate, which are connected in series. The first gate output is the expression of the synthase for particular AHL molecules (e.g., LasI or RhII), which will then diffuse through the cell membrane and bind to the 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 [252]. 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 will behave on the basis of the functions of the individuals [289].

Since not all the concentrations of the elements (e.g., promoters, mRNAs and proteins) involved in the logic gates have been given in [252], the energy consumption computations of logic gates are simplified by taking the number of all the inputs and outputs as 1, which means that the actual required concentration of the inputs and outputs of the logic gate implementations is not considered. Although this simplification may not be entirely accurate, 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. Nomenclature is introduced 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, equally probable zeros and ones are assumed 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:

$$E_{\text{buffer}1} = (E_{\text{rhlR}} + E_{\text{RhIR}}) + E_{\text{YFP}} = 30L_{\text{rhlR}} + 50L_{\text{RhIR}} + 50L_{\text{YFP}} \quad (7.1)$$

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

$$E_{\text{buffer}} = E_{\text{buffer}1}/2 = 15L_{\text{rhlR}} + 25L_{\text{RhIR}} + 25L_{\text{YFP}} \quad (7.2)$$

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:

$$E_{or_{01}} = E_{araC} + E_{AraC} + E_{YFP} = 30L_{araC} + 50L_{AraC} + 50L_{YFP} \quad (7.3)$$

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

$$E_{or_{10}} = E_{tetR} + E_{TetR} + E_{YFP} = 30L_{tetR} + 50L_{TetR} + 50L_{YFP} \quad (7.4)$$

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

$$\begin{aligned} E_{or_{11}} &= E_{araC} + E_{AraC} + E_{tetR} + E_{TetR} + E_{YFP} \\ &= 30L_{araC} + 50L_{AraC} + 30L_{tetR} + 50L_{TetR} + 50L_{YFP} \end{aligned} \quad (7.5)$$

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

$$E_{or} = \frac{1}{4}(E_{or_{00}} + E_{or_{01}} + E_{or_{10}} + E_{or_{11}}) \quad (7.6)$$

For the NOR gate shown in Figure 7.5(c), the output 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:

$$E_{nor_{00}} = E_{YFP} = 50L_{YFP} \quad (7.7)$$

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:

$$E_{nor_{01}} = E_{araC} + E_{AraC} = 30L_{araC} + 50L_{AraC} \quad (7.8)$$

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

$$E_{nor_{10}} = E_{tetR} + E_{TetR} = 30L_{tetR} + 50L_{TetR} \quad (7.9)$$

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

$$\begin{aligned} E_{nor_{11}} &= E_{araC} + E_{AraC} + E_{tetR} + E_{TetR} \\ &= 30(L_{araC} + L_{tetR}) + 50(L_{AraC} + L_{TetR}) \end{aligned} \quad (7.10)$$

Then, the NOR gate energy consumption is found from:

$$E_{nor} = \frac{1}{4}(E_{nor_{00}} + E_{nor_{01}} + E_{nor_{10}} + E_{nor_{11}}) \quad (7.11)$$

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:

$$E_{xor_{00}} = E_{LasI} + 2(E_{lasR} + E_{LasR}) = 50L_{LasI} + 60L_{lasR} + 100E_{LasR} \quad (7.12)$$

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 RhII are synthesised. Moreover, in this case, the promoter P_{RhI} is expressed, with the expression of gene $rhIR$ and synthesis of protein RhIR. Thus the energy cost can be given as:

$$\begin{aligned} E_{xor_{01}} &= 2(E_{tetR} + E_{TetR}) + E_{YFP} + E_{RhII} + (E_{rhIR} + E_{RhIR}) \\ &= 60L_{tetR} + 100L_{TetR} + 50L_{YFP} + 50L_{RhII} + 30L_{rhIR} + 50L_{RhIR} \end{aligned} \quad (7.13)$$

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 RhII are synthesised.

Moreover, in this case, the promoter P_{RhI} is expressed, with the expression of gene $rhIR$ and synthesis of protein $RhIR$. Thus the energy cost can be given as:

$$\begin{aligned}
E_{xor_{10}} &= 2(E_{araC} + E_{AraC}) + E_{YFP} + E_{RhII} + (E_{rhIR} + E_{RhIR}) \\
&= 60L_{araC} + 100L_{AraC} + 50L_{YFP} + 50L_{RhII} + 30L_{rhIR} \\
&\quad + 50L_{RhIR}
\end{aligned} \tag{7.14}$$

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:

$$\begin{aligned}
E_{xor_{11}} &= 2(E_{tetR} + E_{TetR}) + 2(E_{araC} + E_{AraC}) \\
&= 60L_{tetR} + 100L_{TetR} + 60L_{araC} + 100L_{AraC}
\end{aligned} \tag{7.15}$$

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

$$E_{xor} = \frac{1}{4}(E_{nor_{00}} + E_{nor_{01}} + E_{nor_{10}} + E_{nor_{11}}) \tag{7.16}$$

7.4 Mathematical analysis of throughput and energy efficiency

Section 6.3.3 has presented the detailed mechanisms for SW-ARQ, GBN-ARQ and SR-ARQ protocols. In this section, the mathematical analysis of both the throughput and energy efficiency of the three ARQ protocols will be presented.

7.4.1 SW-ARQ

In the SW-ARQ technique, the transmitter waits for a receiver ACK/NAK after transmitting a frame before transmitting the next frame. The retransmission continues until a frame is received correctly and is positively acknowledged, or the number of retransmissions reaches a certain threshold.

Firstly, the throughput efficiency is taken into consideration as a measurement of the performance of the SW-ARQ technique. It 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_{cw}, k) code is used in the protocol, where k is the number of information bits and n_{cw} is the total number of bits in the information frame including the information bits, sequence numbers and error checking bits. The probability p_f that a code word is received incorrectly at the receiver has been given in equation (6.5). It should be noted here that this work only considers the independent bit errors, other than burst error conditions with no interleaving. One successful transmission of a code word requires a total time t_0 , which has been given in equation (6.1). Some parameters are introduced in this equation: t_p is the propagation time between the transmitter and the receiver for one bit, which can be calculated according to Chapter 4; t_f or t_{ack} is the time from when the first bit of a frame arrives at the receiver to when the last bit arrives for an information frame or an ACK/NAK frame, respectively; the variable n_a 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. CRC codes can be implemented by logic gate operations [202], making it possible to operate CRC coding using genetic circuits. Also, t_{ps} is the processing time for CRC implementation, which is set as four hours in this chapter according to [252]. In this time interval, an equivalent logical operation behaviour can be produced [252]. Here, R_b is chosen as the maximum value according to equation (6.2). 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 [290]. 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 $(1 - p_f + P_u)$. Generally, CRC can detect all single, double and odd bits in error [291].

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:

$$N_{r_sw} = \sum_{i=1}^{\infty} i p_f^{i-1} (1 - p_f + P_u) = \frac{1 - p_f + P_u}{(1 - p_f)^2} \quad (7.17)$$

Therefore, the average time required to successfully transmit the code is: $T_{r_sw} = N_{r_sw} t_0$. 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:

$$\eta(\text{sw_th}) = \frac{T_{tu}}{T_{r_sw}} = \frac{k/R_b}{(2t_p + 2t_{ps} + \frac{n_{cw}}{R_b} + \frac{n_d}{R_b}) \left(\frac{1 - p_f + P_u}{(1 - p_f)^2} \right)} \quad (7.18)$$

It has been stated above that CRC implementations can be represented by combinations of logic gates. To calculate the energy efficiency of the ARQ protocols, the energy cost of CRC implementations needs to be investigated first. As shown in Figure 7.4, for CRC-8, the systematic encoder 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 gates and 1 NOR gate due to the 2-input limitation of the proposed logic gates in Figure 7.5. Hence, the energy cost of CRC-8 encoder can be described by:

$$E_{\text{encode8}} = 8E_{\text{buffer}} + 5E_{\text{xor}} \quad (7.19)$$

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

$$E_{\text{decode8}} = 8E_{\text{buffer}} + 5E_{\text{xor}} + 6E_{\text{or}} + E_{\text{nor}} \quad (7.20)$$

Similarly, according to [202], the energy cost for CRC-12 encoder and decoder can be represented by:

$$E_{\text{encode12}} = 12E_{\text{buffer}} + 5E_{\text{xor}} \quad (7.21)$$

$$E_{\text{decode12}} = 12E_{\text{buffer}} + 5E_{\text{xor}} + 10E_{\text{or}} + E_{\text{nor}} \quad (7.22)$$

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

$$E_{\text{encode16}} = 16E_{\text{buffer}} + 3E_{\text{xor}} \quad (7.23)$$

$$E_{\text{decode16}} = 16E_{\text{buffer}} + 3E_{\text{xor}} + 14E_{\text{or}} + E_{\text{nor}} \quad (7.24)$$

As stated in Chapter 4, the number of molecules emitted at the start of each time slot is denoted by n . Hence for an (n_{cw}, k) code, the energy required to construct an information frame is denoted by $E_{\text{info}} = 50n \cdot n_{cw} \cdot L_{\text{AI1}}$, where L_{AI1} is the length of the Type-I signalling molecule in amino acids. At the receiver, the energy required to construct the LuxR receptors for Type-I is given by $E_{\text{re}} = 50n \cdot n_{cw} \cdot L_{\text{LuxR}}$, where L_{LuxR} is the length of LuxR molecule. In the feedback channel, the energy required to construct an ACK/NAK frame is represented by $E_{\text{ack}} = 50n \cdot n_a \cdot L_{\text{AI2}}$, where L_{AI2} is the length of the Type-II 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 equations (7.19)–(7.24).

Hence, the average energy consumption for transmitting a code word is:

$$E_{sw} = E \cdot N_{r_{sw}} \quad (7.25)$$

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: $E_0 = 50n \cdot k \cdot L_{AI1} + 50n \cdot k \cdot L_{LuxR}$.

Here, 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 the SW-ARQ protocol, the energy efficiency can be described as: $\eta(sw_en) = E_0/E_{sw}$.

Using the GenBank database [292], the length of promoters and proteins mentioned above are given in Table 7.1.

Table 7.1 Promoters and proteins lengths

Promoters	length in nucleotides (bp)	Proteins	Length in amino acids (aa)
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

7.4.2 GBN-ARQ

For GBN-ARQ, the transmitter continuously transmits a window size of W frames without waiting for the acknowledgement for the individual frame. Before being positively acknowledged, each frame must be buffered in the transmitter in case

retransmission is required. The receiver only accepts correct and in-order frames, discarding all the erroneous or out-of-sequence frames. 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 the next block. For the continuous GBN-ARQ scheme, there are W packets in each block. The boundary condition is represented by $W > t_0/t_f$, where t_0 and t_f have the same meanings as in equation (6.1). It can be derived that:

$$W > \frac{2(t_p + t_{ps})R_b + n_{cw} + n_a}{n_{cw}} \quad (7.26)$$

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

$$T_{r_con_gbn} = t_f + \sum i p_f^i (1 - p_f + P_u) t_0 = \frac{n_{cw}}{R_b} + t_0 \frac{p_f (1 - p_f + P_u)}{(1 - p_f)^2} \quad (7.27)$$

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: $\eta(\text{con_gbn_th}) = T_{tu}/T_{r_con_gbn}$. For this continuous GBN-ARQ scheme, the average number of attempts required per successful transmission of a frame can be calculated by: $N_{r(\text{con-gbn})} = T_{r_con_gbn}/t_0$. 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 the continuous mode which is described in equation (7.26), 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 [290]:

$$\begin{aligned}
N_{r(\text{unc-gbn})} &= \frac{1}{W} \sum (iW + 1)p_f^i(1 - p_f + P_u) \\
&= \frac{(1 - p_f + P_u)(1 - p_f + Wp_f)}{W(1 - p_f)^2}
\end{aligned} \tag{7.28}$$

where $i = 0, 1, 2, \dots$. Therefore, the average time required to successfully transmit the code is: $T_{r_unc_gbn} = N_{r(\text{unc-gbn})}t_0$. Hence, it can be concluded that the throughput efficiency of non-continuous GBN-ARQ scheme is calculated by: $\eta(\text{unc_gbn_th}) = T_{tu}/T_{r_unc_gbn}$. Also, the energy efficiency for this case is given by: $\eta(\text{unc_gbn_en}) = E_0/(E \cdot N_{r(\text{unc-gbn})})$.

7.4.3 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 out-of-sequence frames and buffers them. This requires more buffer space at the receiver to store correct but out-of-order frames.

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

Case I: continuous scheme when $W > t_0/t_f$:

Here the average transmission time for successful transmission of a frame is straightforward as below [290]:

$$T_{r_con_sr} = \frac{T_{t_con_sr}}{1 - p_f + P_u} \tag{7.29}$$

where $T_{t_con_sr}$ 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 [290]:

$$T_{t_con_sr} = \frac{Wt_f + k_0 \cdot 2t_p}{W} \quad (7.30)$$

where k_0 is the number of errors out of W , which can be represented by [290]:

$$k_0 = \sum_{i=1}^W ip_f^i(1 - p_f + P_u)^{W-i} \quad (7.31)$$

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: $\eta(\text{con_sr_th}) = T_{tu}/T_{r_con_sr}$. For this continuous SR-ARQ scheme, the average number of attempts required per successful transmission of a frame can be calculated by: $N_{r(\text{con-sr})} = T_{r_con_sr}/T_{t_con_sr}$. 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 $W < t_0/t_f$:

In this case, the average number of transmissions required for successful transmission of a frame is given by [290]:

$$N_{r(\text{unc-sr})} = \frac{1}{W} \sum ip_f^{i-1}(1 - p_f + P_u) = \frac{1 - p_f + P_u}{W(1 - p_f)^2} \quad (7.32)$$

where $i = 0,1,2, \dots$. Therefore, the average time required to successfully transmit the code is: $T_{r_unc_sr} = N_{r(\text{unc-sr})}t_0$. It can be concluded that the throughput efficiency of the non-continuous SR-ARQ scheme is calculated by: $\eta(\text{unc_sr_th}) = T_{tu}/T_{r_unc_sr}$. Also, energy efficiency for this case is given by: $\eta(\text{unc_sr_en}) = E_0/(E \cdot N_{r(\text{unc-sr})})$.

7.5 Numerical results

This section presents investigations of 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 analysed.

7.5.1 Parameter setup

According to the analysis in Section 7.4, here the channel throughput and energy efficiency performance for SW-ARQ, GBN-ARQ and SR-ARQ protocols are discussed in terms of different channel parameters, with particular emphasis on the information frame length. In addition, optimal frame lengths for various channel conditions are being investigated, achieving the minimum channel throughput and energy efficiency, respectively. The information bits are made up from “0”s and “1”s with equal frequency. The total amount of information to be transmitted is set to be in the range of 1 to 1500 bits, which will be broken into frames for transmission. CRC check bits and a 3-bit sequence number are appended to each information frame. Previous chapters and previous work [243] indicate that the parameters including 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 have been varied to assess the system performance. In this chapter, the main focus is the optimisation of the frame length. 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. The transmission distance is set to be $4\mu\text{m}$ and there are 100 bacteria in the receiver node. The lengths of the promoters and proteins are those given in Table 7.1.

7.5.2 SW-ARQ

The throughput and energy efficiency in terms of frame length of the system are being simulated according to Section 7.4.1, with different CRC polynomials and different numbers of molecules per bit. The results are given in Figure 7.6 and Figure 7.7, which depict the throughput and energy efficiency for various values of frame

size, CRC polynomials and number of molecules emitted at the start of each time slot, respectively. Both of them indicate that when a particular CRC polynomial is used, a larger number of molecules per bit results in better throughput and energy performances. Also, with the same number of molecules per bit, CRC-8 performs better than 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 0.3 and 0.45, respectively, for the condition $n = 170$, CRC-8. In addition, it is clear that in each case, there exists a unique maximum value for either throughput or energy efficiency, with a corresponding optimal frame size.

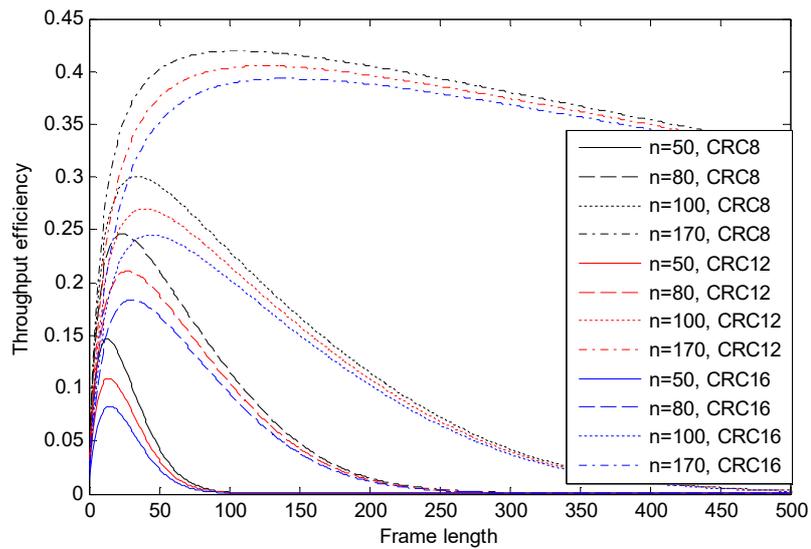


Figure 7.6 Throughput efficiency versus frame length for SW-ARQ

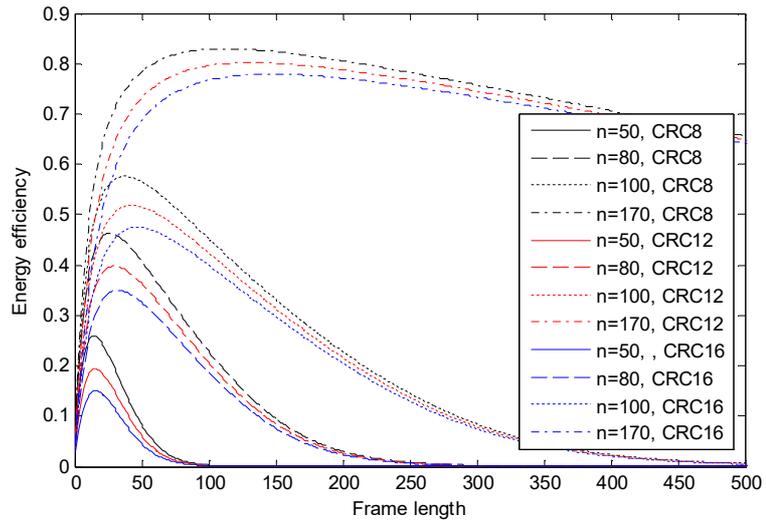


Figure 7.7 Energy efficiency versus frame length for SW-ARQ

Figure 7.8 shows the optimal frame lengths for maximizing the throughput and energy efficiency of SW-ARQ scheme. It is clear that with increasing numbers of molecules per bit, the optimal frame length will grow. Also, when a certain CRC is applied, the optimal frame length required to reach the maximum energy efficiency is slightly larger than that to reach the largest throughput efficiency.

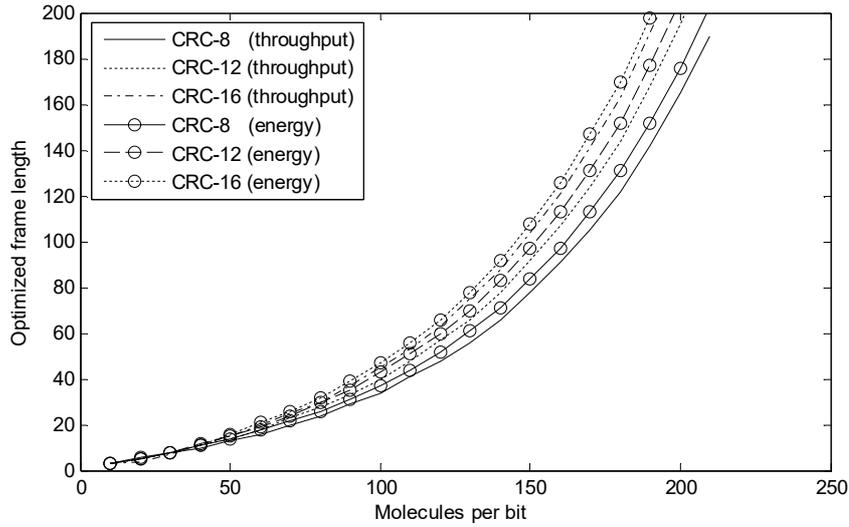


Figure 7.8 SW-ARQ optimal frame length versus molecules per bit to maximize the throughput and energy efficiency, for different CRC polynomials.

7.5.3 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 λ -bit sequence number to each information frame means that the sequence of frames carries sequence numbers with decimal representation $[0, (2^\lambda - 1)]$. In general, the window size W for GBN-ARQ needs to be $(2^\lambda - 1)$ or less to avoid ambiguities [251]. 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 [202], CRC-8 is applied in the following investigations due to its relatively lower operation time and energy consumption.

Following the method in Section 7.4.2, the throughput and energy efficiency of GBN-ARQ scheme versus different frame length for different numbers of molecules per bit and window sizes are displayed in Figure 7.9 and Figure 7.10.

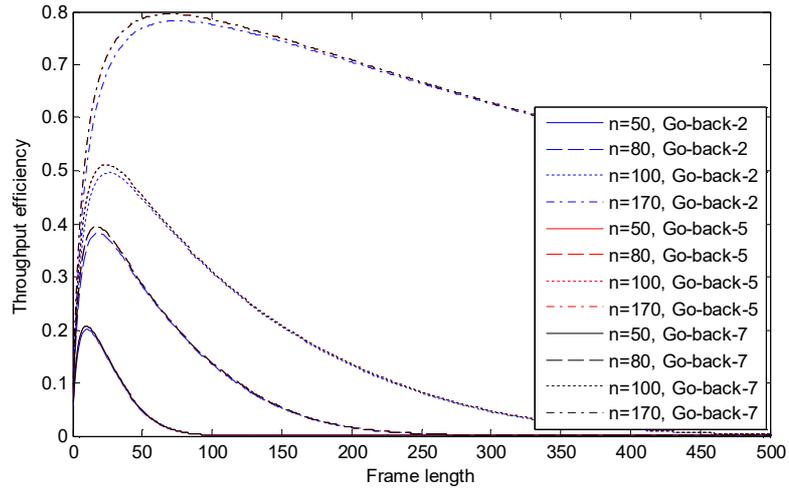


Figure 7.9 Throughput efficiency versus frame length for GBN-ARQ

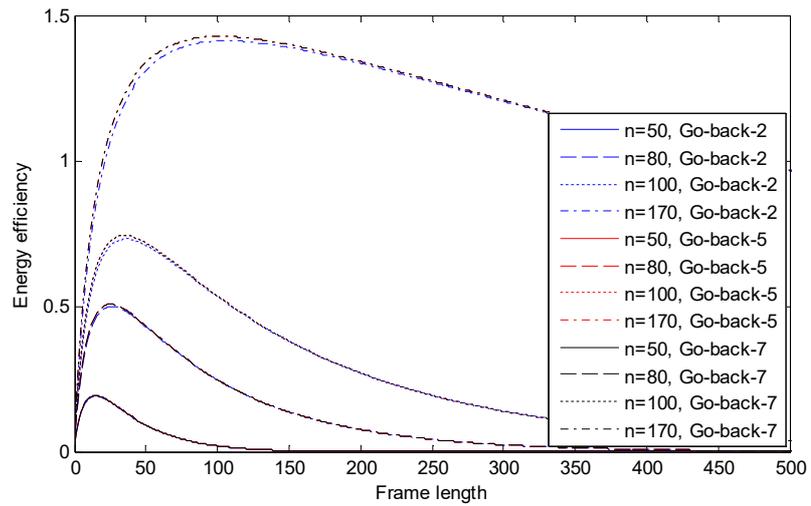


Figure 7.10 Energy efficiency versus frame length for GBN-ARQ

Figure 7.9 and Figure 7.10 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 a larger number of molecules per bit. Hence, under reliable 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.

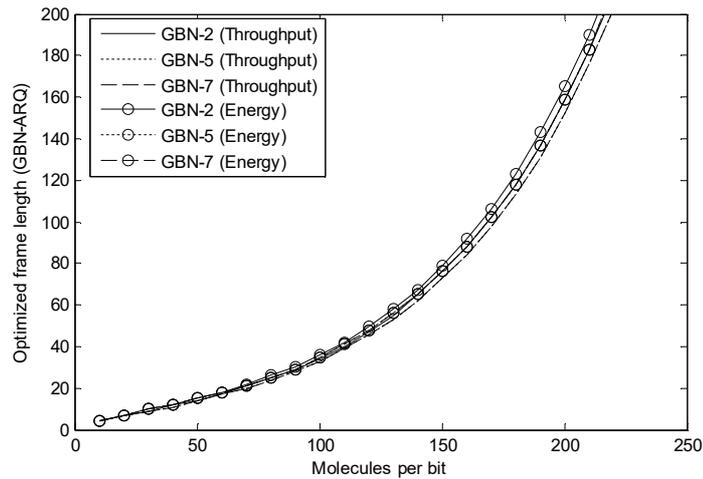


Figure 7.11 GBN-ARQ optimal frame length versus molecules per bit to maximize the throughput and energy efficiency, for different window sizes.

Figure 7.11 shows the optimal frame length of GBN-ARQ for maximizing the throughput and energy efficiency, respectively. It is obvious that as the number of molecules per bit increases, the optimal frame length grows. Also, when the number of molecules per bit is fixed, the corresponding optimal frame lengths are almost the same as SW-ARQ to achieve the maximum throughput and energy efficiency for different window sizes being slightly smaller than those of SW-ARQ in Figure 7.8.

7.5.4 SR-ARQ

As in 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 λ -bit sequence number is appended to each information frame, the window size W for SR-ARQ needs to be $2^{\lambda-1}$ or less to avoid duplicate transmission [251]. 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.

The throughput and energy efficiency of the SR-ARQ scheme are investigated in accordance with Section 7.4.3, for various values of frame lengths, number of molecules per bit and the window size.

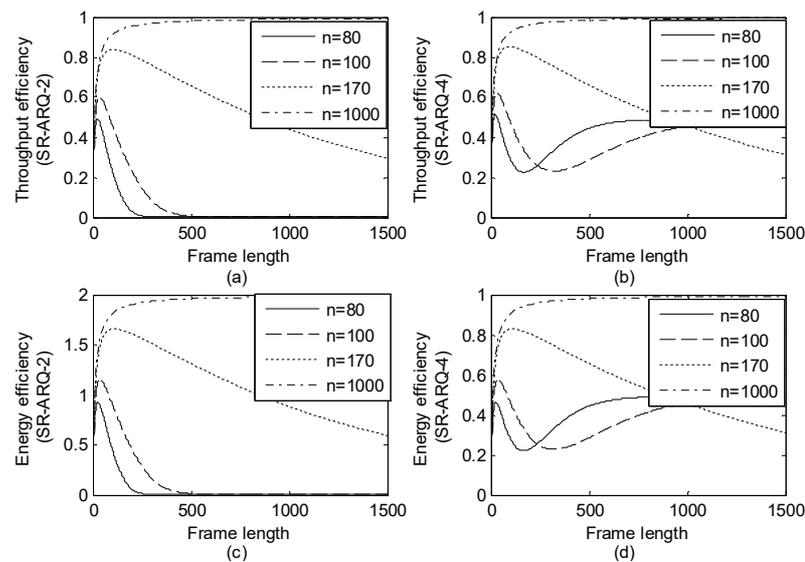


Figure 7.12 Throughput and energy efficiency performances for SR-ARQ.

(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.

Figure 7.12 shows results obtained for the throughput and energy efficiency of the SR-ARQ scheme using different frame lengths, numbers of molecules per bit and

window sizes. It should be 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 80 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 100 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.

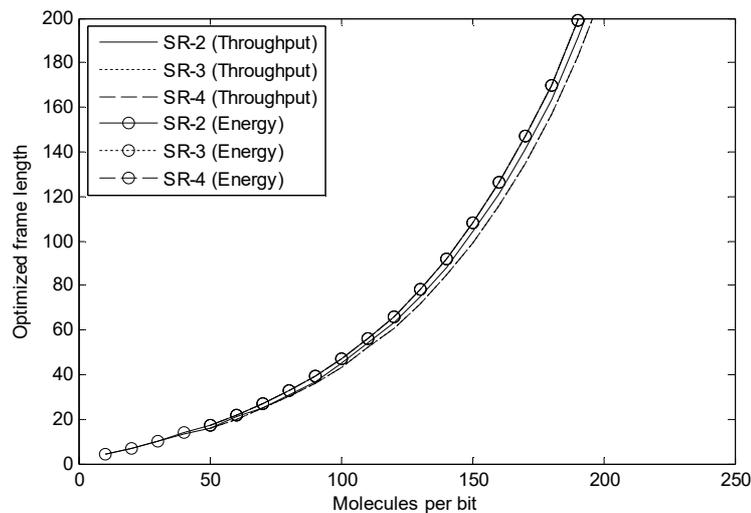


Figure 7.13 SR-ARQ optimal frame length versus molecules per bit to maximize the throughput and energy efficiency, for different window sizes.

Figure 7.13 shows the optimal frame lengths for maximizing the throughput and energy efficiency for the SR-ARQ scheme. It is clear that with the number of molecules per bit increasing, the optimal frame length will grow. Also, there is not

much difference between the optimal frame length to maximise the throughput efficiency and that to maximise the energy efficiency for different window sizes.

7.6 Summary

In Chapter 6, the widely used ARQ protocols have been employed, specifically utilising CRC coding, SW-ARQ, GBN-ARQ and SR-ARQ, in the proposed bacterial communication network model in this thesis. The simulation results indicate that there should be an optimal frame length to optimize the system performances. Existing frame size optimization techniques are not applicable in the case of bacterial communications due to its limited capabilities and that the materials used to be engineered in bacterial communication circuits are fundamentally different from their engineering counterparts. Instead, some synthetic biological components are employed to help implement the ARQ protocols and channel performance analysis. 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 the SR-ARQ scheme is slightly larger than the SW-ARQ scheme, followed by the GBN-ARQ scheme. The simulation results 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. It is recognised 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 [252] to create complex coding, windowing and sequencing operations. Nevertheless, it is considered that the broad conclusions with respect to the performance of the established ARQ schemes over the bacterial diffusion channel to be valid.

CHAPTER 8.

CONCLUSIONS AND FUTURE

WORK

Communication among devices at the nanoscale is required for expanding the capabilities of single nanomachines, increasing the complexity and range of operation of the system. With the enormous potential applications for nanoscale networks in various fields, MC is considered as the most appropriate option for this issue.

In the past few years, bacteria have been considered as one promising approach for MC using QS, which is a natural phenomenon that uses signalling molecules to coordinate the action of a group of bacteria, depending on the bacterial population density. This work has presented an in-depth study of the characteristics and performance of a QS based bacterial communication network model through a diffusion channel between transmitter and receiver bacterial populations. Intensive investigations of particular error control schemes, impacts of ISI, and the channel characteristics have been presented. Channel reliability has been improved by the first use of (a) widely used FEC techniques, including Hamming codes, MEC and LT codes; (b) ARQ protocols, specifically utilising CRC coding, SW-ARQ, GBN-ARQ and SR-ARQ. Further investigations in adjusting the frame size to optimise the channel throughput and energy efficiencies of ARQ schemes have also been

provided. In this work, the simulation processes are within the transmission distance range of $2\text{-}8\mu\text{m}$. When considering MC, instead of trying to connect electronic devices, we may be interconnecting biological entities, where each device has different limitations and requirements. Therefore, it is not perfectly clear if using standard communication theory for MC would be appropriate. It may take a long time for the realization of this theory in practice. Also, the results obtained in this work are estimations, which, however, are considered to be useful in future studies in this area, providing a method for combining standard communication theory with MC. In this chapter, Section 8.1 gives the concluding remarks of this work and Section 8.2 contains some suggestions for future projects.

8.1 Concluding remarks

In Chapter 1, the background to nanoscale communications, especially the communication mechanisms between nanomachines, and the possible challenges have been presented. In addition, this chapter gives the motivations, research objectives, major contributions and outline of the work.

In Chapter 2, a detailed description of the emerging discipline of MC has been presented, including the communication architectures, communication types, characteristics, research aspects and potential applications. The generic MC system is composed by information molecules (e.g., proteins or ions), the transmitter and receiver nanomachines and the environment through which the signals travel. Also, the MC processes include encoding, transmission, propagation, receiving and decoding. Based on the information transfer mechanism between nanomachines, MC can be divided into wired and wireless communications. In the former, the transport of information molecules needs to be guided by separate physical mechanisms, such as microtubules, providing a physical contact between the transmitter and receiver.

The latter can be further divided into passive transport (e.g., diffusion-based) in which signalling molecules are uniformly distributed in all directions, and active transport (e.g., based on bacterial chemotaxis) in which molecules propagate in a particular direction. The emphasis of this thesis is the free diffusion-based MC system, whereby the transmitter encodes information with a quantity of molecules which diffuse passively to the receiver. In addition, MC exhibits some unique features that are not commonly found in conventional communications, for example, molecular information carriers, limited transmission distance and speed, biocompatibility, functional complexity, stochasticity, low energy cost, self-assembly, robustness and fragility and so on. At the present stage, most of the MC research has been focusing on modulation techniques, channel modelling and noise analysis, coding techniques, protocol design and the development of simulation tools. As a new communication paradigm, MC would potentially enable many new applications in biomedical, industrial, environmental and information technological fields.

In Chapter 3, bacterial communications, which use bacteria to establish nano communication networks, is introduced as a promising MC approach due to some of the biological characteristics. The communication mechanisms based on bacteria can be divided into two categories: molecular-based communication where bacteria can emit specific molecules as information carriers between bacterial populations, and plasmid-based communication where bacteria are able to transfer plasmids to carry information. In this thesis, QS, as a typical molecular-based communication, is focused on. QS enables the bacteria in a population to control fundamental functions involved in important stages of their life cycles and respond to environmental changes, (e.g., temperature, pH and concentration of ions) simultaneously and

cooperatively. This chapter also gives the structures, essential properties and capabilities of bacteria which are necessary for bacterial cells acting as transceivers for the establishment of bacterial communication networks in subsequent chapters. Concerning the bacteria structural components, the cell membrane calls for special attention because it is selectively permeable, allowing specific groups of molecules pass through freely, which enables the autoinducers to diffuse in and out of the bacteria cells in the proposed model in Chapter 4. Also, it should be noted that ATP acts as the universal energy currency in most of the biological processes that require energy. Moreover, the QS mechanisms of two fundamentally different bacteria types, gram-negative and gram-positive bacteria, have been presented in this chapter. In addition, the recent studies of synthetic biological systems that utilize QS systems and the potential applications are also discussed. Most of the research concerning bacterial communications has focused on the development of signalling molecules, which have the potential use as antimicrobial drugs. QS has significant applications in industry, environment, and biomedical fields.

In Chapter 4, a brief review of existing models of QS has first been presented. Most current studies have concentrated on the mathematical modelling of bioluminescence systems and their homologues. Moreover, this chapter proposes a diffusion-based bacterial quorum communication model, consisting of the transmitter node, the receiver node, and the communication channel. Both the transmitter and receiver nodes are composed of a cluster of genetically modified bacteria (*V. fischeri*), which can sense specific types of signals and respond accordingly. A population of bacteria, rather than an individual bacterium, is used as a node because a single bacterium is very primitive and unreliable, and is incapable of transferring information by itself. *V. fischeri* possesses two QS systems, based on *ain* and *lux*, with the utilisation of

Type-I autoinducer (3-oxo-C6-HSL) and Type-II autoinducer (C8-HSL) as signalling molecules, respectively. In this chapter, only the former QS system is taken into account. The transmitted information is encoded via the concentration of signalling molecules, by OOK modulation, with the release of molecules representing a binary '1' and no release representing a binary '0'. The communication process is as follows: by specific stimulation, the transmitter node generates the signalling molecules (Type-I autoinducers) which then propagate through the medium undergoing Brownian motion and finally the receiver node senses the concentration of the local autoinducers through the corresponding receptors (LuxR), followed by the production of luminescence, which is used to decode the transmitted information. Depending on the functionalities, for the bacteria in the transmitter node, only Type-II receptors (AinR) are activated and the gene expression of *ains* is repressed, whilst for the bacteria in the receiver node, only Type-I receptors (LuxR) are enabled and the gene expression of *luxI* is repressed. The 3-D communication medium in which the signalling molecules diffuse is considered to be infinite and homogeneous, with identical properties throughout all its extension. Each of the three processes, including transmission, propagation and reception, has a probabilistic nature and the corresponding mathematical model has been presented in this chapter. It should be noted that throughout this thesis, the main investigation focus is on the channel or propagation model, rather than the transmitting and receiving models. The proposed time-slotted channel is modelled as a BAC, with binary inputs and outputs and a probability of error. Also, only one symbol is allowed to be transmitted in each time slot. Investigations show that the symbol duration of the channel is mainly influenced by the distance between the transmitter and receiver nodes. The major noise in the proposed model results from

ISI, which means that the molecules sent from the previous time slots can have an effect on the received symbol in the current time slot. Two optimization problems have been defined, including achieving the minimum error probability and maximum mutual information to analyze the channel performance. It has been indicated that the selection of the appropriate threshold value is critical for the BER, the mutual information and the channel capacity computations. Also, investigations on BER performances of the channel when different numbers of previous time slots are taken into account indicate that for the current time slot, the first previous time slot has the most significant effect on the error performance, the second previous time slot has a smaller influence and so forth. In addition, results indicate that a better channel capacity can be achieved with a smaller transmission distance. Also, the channel capacity attains the maximum value of 1 bit per channel use at each distance when using a few hundred molecules per bit at a distance of 5 μm but needs a few thousand molecules per bit at 30 μm .

Due to the fact that ISI caused by molecular diffusion may result in data packet corruption and out-of-sequence delivery of transmitted information, error control techniques, including FEC and ARQ, are required to detect and correct errors that occur in transmission frames for robust transmission. Both techniques can be implemented through circuits in conventional communication concepts. Also, due to the fact that a population of bacteria is able to perform distributed information processing, and that the development of synthetic biology enables many types of capabilities based on genetically engineered bacteria (e.g., logical functions), FEC and ARQ mechanisms are supposed to be able to implement in bacterial communication systems. Thus, Chapters 5 and 6 introduce FEC and ARQ techniques to the proposed quorum communication model, respectively.

Chapter 5 starts with the review of the existing FEC techniques for MC. At the present stage, there is only a small amount of research considering their application in nanoscale networks to enhance reliability, by transmitting sufficient redundant data to allow the receiver to recover from errors without sender retransmissions. Some traditional channel codes, such as Hamming codes, LDPC codes, C-RM codes and SOCCs have been applied previously. Also, a few new codes, for example the ISI-free code, have been specifically designed for the scenario of MC. After this literature review, a brief introduction to the mechanisms of error correction codes has been presented in this chapter. For a general communication system, the message from the source is encoded and modulated for transmission over the channel; and then demodulated, decoded and sent to the receiver. Channel codes can be broken into three fundamental categories: block codes, convolutional codes and modern codes. Redundant bits allow the receiver to correct error bits but at the expense of a larger forward channel bandwidth and extra energy consumption in coding and decoding the messages. Due to the restricted energy storage capabilities of bacterial nodes and the limited output power range of bio-nanomachines, the design of channel codes should not only achieve reliable information transmission, which is assessed using BER, but also ensure low energy consumption at the same time. In this work, MEC and LT codes, which map a certain number of message symbols to a certain number of code symbols in different methods, have been introduced and extensively investigated in the proposed bacterial communication system for the first time and the performance enhancements they bring have been compared against Hamming codes with respect to both coding gain and energy requirements. MECs are implemented by utilizing an unconstrained code distance but with a minimum codeword length. LT codes, which are a category of rateless codes, provide an

endless stream of codewords for decoding. The parameters chosen conform to the methods used in the existing literature, so that the results are believed to be representative of the performance of the coding schemes in the field. Although from the perspective of current coding standards, these relatively simple coding schemes are not particularly powerful, they are still applied and proved to be efficient in terms of the channel reliability and the energy budget of MC. Results show that the channel performance of the communication system can be improved with Hamming codes, MEC codes and LT codes, offering coding gains which can be several dBs. Specifically, at the BER level of 10^{-6} , when the receiver node contains 100 bacteria and the transmission distance is $4\mu m$, the coding gain for (7, 4) and (15, 11) Hamming codes is 0.94 and 1.65dB, respectively, indicating as expected that the higher order Hamming code exhibits superior error performance and can lead to an extended transmission distance. Moreover, for the same distance, number of bacteria and BER level, the MEC coding gains for $M = 2^4$ and $M = 2^{11}$ are 4.67 dB and 9.14 dB, respectively; for LT codes the figures are 1.37 dB and 2.04 dB for $c = 0.01$ and $c = 0.03$ respectively. Results indicate that MECs have a better performance than Hamming codes with a larger coding gain and that the system performance is better with a lengthy codeword. In addition, among the three proposed coding schemes, LT codes have higher energy consumption than Hamming codes and MECs, especially when the number of molecules per bit is smaller than approximately 60, whilst MECs exhibit the lowest energy cost. According to the results, the code designer can determine which code is appropriate by the energy cost for different operating BERs. For example, considering the case when a communication system operates at a BER larger than 10^{-8} , the $c = 0.03$ LT code is a better choice than (15, 11) hamming code, with a smaller BER. However, it should

be noted that in this chapter, the energy consumption analysis is an approximate prediction because that complete bio-nanomachines have not yet developed.

Although FEC can enhance channel reliability without retransmission, it cannot resolve the issue of out-of-sequence or duplicate delivery of information packets in transit through the network. Thus in Chapter 6, the reliable transport protocols, in particular ARQ protocols, are taken into consideration because that they can enable the receiver to receive each packet in-order and only once and that error detection requires much simpler decoding operations than does error correction. This work is the first attempt to apply ARQ protocols in a bacterial communication system. The chapter starts with a literature review of the existing transmission protocols for MC. The IEEE Standards Association has started the IEEE P1906.1 project to provide recommended standards for nanoscale and MC frameworks, and to assist research in related fields. Some relevant research, which concentrates on designing protocols aiming at rate control, routing, addressing, bit transmission duration adjustment, distance measurement and so on have been carried out. Also, this chapter presents a detailed description of ARQ mechanisms as well as CRC codes, which have been used as the error detection codes in this chapter due to their good error sensing performance and fast encoding/decoding capabilities. An ARQ protocol is commonly located at the link layer in layered network architectures, and takes the packets it gets from the upper layer and encapsulates them into frames, which contain sequence numbers, information sequence and error detection bits, for transmission. The ARQ process is generally as follows: when the CRC encoded data frame arrives at the destination, the receiver will check for errors using the CRC check bits. If an error occurs, the receiver sends a NAK back to the transmitter, requesting the transmitter to repeat the information in the erroneous frame, otherwise

an ACK will be sent back indicating that the frame has arrived correctly and safely. The timeout mechanism is usually required to maintain the flow of frames. It should be noted that the ARQ protocols are bidirectional protocols, with the flow of information only in the forward channel and the flow of ACKs/NAKs only in the reverse channel. Considering the proposed channel in Chapter 4, the information frame is made up of a certain number of information bits, which are encoded by Type-I molecules. Similarly, the acknowledgement frames are composed of a fixed number of bits encoded by Type-II molecules. The forward and reverse channels will not interfere with each other because that both the transmitter and receiver are able to generate and accept different types of signalling molecules. In this chapter, three categories of ARQ protocols, SW-ARQ, GBN-ARQ and SR-ARQ, have been applied to investigate the channel performances in terms of transmission time, transmission efficiency, BER and channel throughput. Results show that the major aspects that affect the system performances are transmission distance d , number of bacteria in the receiver node m , frame length N , CRC polynomial and window size W . The simulations start with the generation of a fixed length bit stream of 2000 bits, which is then be divided into equal length frames. The frames are appended by the CRC check bits and a 3-bit sequence number, for transmission. Also, the timeout period is set to be exactly equal to the sum of round trip propagation delay and the CRC processing time. The simulations first concentrate on the impact of individual variables on the channel performances with the other parameters fixed. Results show that compared with the system performance when neither ARQ nor error detection mechanisms is applied, the channel which uses an ARQ protocol performs better with a lower BER, achieving a better system reliability. For each of these three ARQ protocols, BER and delay performance worsen as the transmission distance increases,

leading to the choice of 4 μm as a feasible and representative value. Given the increasing complexity and delay from high order CRC polynomial operation, CRC-8 is selected. Moreover, all the three kinds of ARQ protocols exhibit a better channel performance with a larger number of bacteria in the receiver node. For SW-ARQ, the window size is 1. For GBN-ARQ and SR-ARQ, when the number of molecules per bit is low (approximately 10-100), the channel performs better using a smaller window size; while as n increases, a window size of 7 is found to be a highly satisfactory compromise for GBN-ARQ, and a value of 4 is the better for SR-ARQ. Moreover, concerning the impact of the frame size, there are two crossover points, indicating that when the number of molecules per bit is in a range of approximately 20 to 250 for SW-ARQ, 40 to 200 for GBN-ARQ and 10 to 250 for SR-ARQ, larger frame lengths perform better, with reduced time consumption and increased efficiency and throughput. After investigation into the impact of individual variables, the performances of different combinations of these factors are taken into consideration. For SW-ARQ, the transmission delay varies mainly because of the CRC polynomial when the number of molecules per bit is smaller than 150; whilst when $n > 150$, it mainly depends on the transmission distance. The performance of transmission efficiency can be divided into three stages, stage 1 when $n < 100$ (CRC polynomial dependent), stage 2 (frame length dependent) when reaching the steady states, and stage 3 (both transmission distance and population of the receiver dependent) which is the period between stage 1 and stage 2. Also, BER is mainly decided by CRC polynomials. The channel throughput performance can be divided into two stages. In stage 1, where the channel throughput is rapidly increasing with the increase of the number of molecules per bit, the significance to the channel throughput of the four factors will be in the decreasing order of d, m, N and CRC

polynomial. In stage 2, the system will reach identical steady states when the same combinations of d and n are applied, regardless of the other parameters. For GBN-ARQ, when the number of molecules per bit is less than approximately 250, the window size W has the biggest influence on the transmission time; while if $n > 250$, there is little difference among different communication scenarios. Also, the significance to both the efficiency and channel throughput is in the decreasing order of m, N, W for the case when $n < 250$. Once $n > 250$, the channel with a larger window size can achieve larger transmission efficiency and channel throughput. In addition, the significance of the three parameters to the error performance is in the decreasing order of m, N, W . For SR-ARQ, the performances of transmission time and BER are similar to those of GBN-ARQ. Also, before achieving stability ($n < 200$), the transmission efficiency or channel throughput mainly depends on the bacterial population in the receiver. When $n > 200$, the channel with parameter combinations $N = 100, W = 4$ and $N = 50, W = 2$ achieves the highest and lowest steady state values, respectively. Finally, with identical parameter settings, the performance of the three ARQ schemes has been investigated, indicating that when the number of molecules per bit is in a smaller range, SW-ARQ performs better than SR-ARQ, and GBN-ARQ performs worst. However, when the number of molecules emitted at the start of each time slot is relatively large, GBN-ARQ and SR-ARQ have similar channel performances, while SW-ARQ performs worst; this is quite different from the performance of ARQ schemes in traditional networking areas.

Based on the simulation results in Chapter 6 which present that optimal frame lengths exist for maximizing the channel throughput and energy efficiency performances, Chapter 7 mainly concentrate on two issues (a) establishing a novel energy model using a series of synthetic biological components; (b) finding the

optimum frame lengths of different ARQ protocols to optimize the throughput and energy efficiency. The chapter starts with the introduction of engineered MC based on synthetic biology, which represents a newly interdisciplinary field at the interface between engineering and biology and is defined as the design and engineer of novel biological components and re-design of natural biological systems for specific purposes. Also, this chapter describes the synthetic biology hierarchy, relevant technical advances and some typical examples of the start-of-art synthetic biological components, specifically toggle switches, oscillators and logic gates. However, it should be noted that a large amount of engineering principles may not be applied directly to the entire engineering process of gene circuits due to various challenges explained in this chapter. Also, a novel energy model, which has been established using four kinds of start-of-art experimentally validated synthetic transcriptional logic gates, is proposed and employed to implement the ARQ protocols and to analysis channel performances of the bacterial communication system. To optimize the throughput and the energy efficiency, the optimal frame length has been determined for different combinations of channel parameters. During the whole numerical computation process, the transmission distance is set as $4\mu m$ and there are 100 bacteria in the receiver node. Results show that lower order CRC is a better choice for ARQ protocols. Also, both the optimization matrix including the throughput and energy efficiency exhibit better performances with a larger number of molecules per bit and a larger window size, for the three ARQ schemes. Moreover, for a fixed frame size, both the optimization matrix for GBN-ARQ and SR-ARQ are higher than that of SW-ARQ. In addition, for fixed value of a smaller n , to achieve the maximum throughput or energy efficiency, the required frame length is in the

decreasing order of SW-ARQ, SR-ARQ, and GBN-ARQ; this phenomenon comprehensively matches the simulation results in Chapter 6.

It is important to note that the simulations and numerical computations are conceptual and contain approximations because they are intended to lay a foundation for future studies. Given the current state of the art, the actual concentrations of the synthetic gene components used to build the energy model are not considered and it is still not clear how to connect different logic gates for CRC operations. To take specific examples or mechanisms that are not clear are present: the transmitter control mechanism to ensure an appropriate molecule rate; the mechanism to implement encoding/decoding, windowing and sequencing using synthetic components; the interconnections between these components. However, the simulation results with respect to the performance of the established FEC and ARQ schemes provide methods of performing conventional codes and protocols in nanoscale communication networks and determining parameters to achieve a better channel performance, which are considered to be valid and useful for further investigations.

To conclude, the aim of this thesis was to design and develop a framework for MC among nanomachines based on QS, which is a valid technique that enables the coordination or synchronization of a group of entities at the nanoscale by means of MC. Moreover, conventional error correction techniques, including FEC and ARQ, can be mapped into bacterial quorum systems for channel performance enhancement, by adjusting system parameters.

8.2 Future work

This work provides a basic framework for bacterial quorum communication systems, which has a high potential to be used as one of the promising techniques of communication at nanoscale. However, the field is very new and there are still a great amount of unaddressed issues and challenges in this area. The achievements of this work inspire some potential research issues described subsequently.

This work mainly concentrates on the investigations into the channel model, without regard for the probabilistic nature of the transmitter and receiver nodes described in Chapter 4. Also, the degradation of signalling molecules and ligand receptors has not been considered. Hence, in future studies, the characteristics of transceivers, especially the binding probability and luminescence output, will be taken into account, establishing a relatively complete bacterial communication model. Concerning the channel model, recently a new channel model has been proposed in [192], which takes into account the dependency of the molecules received in the successive time slots due to one transmission at the start. This new model has been explained briefly in Appendix B. Thus a possible future work may make a comparison between the proposed mode of this thesis and the new model.

Also, since the detection mechanisms proposed in this thesis are based on temporal sampling of the signal concentration, the QS based MC system in this thesis has been mainly investigated in the time domain. Similar research in the frequency domain would also be considered as a future work.

Moreover, instead of using logic gates, the authors in [293] employ the original design of a toggle switch [276] as a memory unit, which may inspire the application of toggle switches as the transceiver buffers in ARQ implementations in future work.

Moreover, it may be possible to use biological oscillators as timers of transmitter and receiver nanomachines to build the timeout mechanism in further studies. In addition, based on the results presented in this thesis, a hybrid energy efficient error control scheme for bacterial communication networks could be designed and implemented in the near future. This hybrid protocol must adapt to the varying channel conditions, including the number of bacteria in the receiver node and the diffusion coefficient (if the system is in a different environment), and to the varying distances between the bacterial nodes. Also, based on the error probability of the channel and the encoding technique applied, an adaptation algorithm can be developed. Also, when designing error control schemes for bacterial communications, some issues must be taken into consideration. For example, due to the extremely small size, bacterial nanomachines are expected to have strict limitations in terms of processing power, memory storage, and the rate of emitting the signalling molecules. Also, another major constraint is the difficulty in linking molecular based logic components to create the complex logic required in communication protocols. In other words, although it is possible to conceptually create complex logic, practical implementation of such circuits is difficult due to the lack of linkable promoter or enzyme reactions. These constraints make designing computationally complex modulations and coding schemes full of challenges. Thus concerning these aspects, based on the new energy model proposed in Chapter 7, the complete design of the encoder and decoder with biochemical reactions can be considered as a future work of this research.

In addition, as shown in this thesis, communication in MC networks is characterised by a long propagation delay and poor reliability. In spite of the FEC and ARQ techniques employed here, network coding approaches, which are used in some wireless network applications, could be used to enhance the system performance.

Furthermore, these network coding approaches may be compatible with nucleic acid molecular computation approaches [20].

Moreover, the use of FEC and ARQ protocols described in this thesis will facilitate the creation of BAN, which would consist of an interconnected network of bio-nanomachines acting as sensors and actuators.

Security is also a key issue for MC nanonetworks. One possible future work is to investigate the mapping of existing security paradigms (e.g., intrusion detection, integrity protection and encryption mechanisms), particularly those used in wireless sensor networks, to MC.

Experimental verification constitutes another significant avenue of future research. In this regard, biological experiments can be implemented by using bacterial cells as transmitters and receivers and the interconnections between them can be studied and verified with the simulation results. However, this requires a considerable amount of research from interdisciplinary perspectives because the experimental setup for MC is now in its infancy.

Finally, it is strongly believed that the investigations presented in this thesis will help MC system researchers to develop more sensitive biosensors, drug delivery systems and water toxicity detection mechanisms.

Appendix A

In this section, some possibilities which were not given in detail in Section 4.6 have been presented. Specifically, when considering three previous time slots, there are 16 different cases for the binary channel model, according to the different values of the symbol in the current time slot and the three symbols in the previous time slots. The possibilities of unsuccessfully decoding the current intended symbol in the current time slot are denoted by $P_{0000}, P_{0010}, P_{0100}, P_{0110}, P_{1000}, P_{1010}, P_{1100}, P_{1110}, P_{0001}, P_{0011}, P_{0101}, P_{0111}, P_{1001}, P_{1011}, P_{1101}, P_{1111}$, where the four bits of the subscripts represent the symbol in the third, second, first previous time slot, and current time slot, in sequence. When n molecules are sent at the start of the time slot, the number of molecules received within the current time duration N_C is a random variable and follows a binomial representation which is approximated by a normal distribution described as:

$$N_C \sim \mathcal{N}(nP_1, nP_1(1 - P_1)) \quad (\text{A. 1})$$

where P_1 represents the capture probability with receiver radius R , transmission distance d and symbol duration t_s . Here, three parameters P_2, P_3, P_4 are introduced to represent the capture probability within time duration $2t_s, 3t_s$ and $4t_s$, respectively, which can be calculated by equation (4.7). Also, to simplify the derivation process, it is defined that:

$$\begin{cases} M_1 = P_1(1 - P_1) \\ M_2 = P_2(1 - P_2) \\ M_3 = P_3(1 - P_3) \\ M_4 = P_4(1 - P_4) \end{cases} \quad (\text{A. 2})$$

The number of left over molecules from the first previous time slot is described as:

$$N_{P_{-1}} \sim \mathcal{N}(nP_2, nP_2(1 - P_2)) - \mathcal{N}(nP_1, nP_1(1 - P_1)) \quad (\text{A. 3})$$

$$\sim \mathcal{N}(n(P_2 - P_1), n(M_2 + M_1))$$

The number of left over molecules from the second previous time slot is described as:

$$N_{P_{-2}} \sim \mathcal{N}(nP_3, nP_3(1 - P_3)) - \mathcal{N}(nP_2, nP_2(1 - P_2)) \quad (\text{A. 4})$$

$$\sim \mathcal{N}(n(P_3 - P_2), n(M_3 + M_2))$$

The number of left over molecules from the third previous time slot is described as:

$$N_{P_{-3}} \sim \mathcal{N}(nP_4, nP_4(1 - P_4)) - \mathcal{N}(nP_3, nP_3(1 - P_3)) \quad (\text{A. 5})$$

$$\sim \mathcal{N}(n(P_4 - P_3), n(M_4 + M_3))$$

For case “0000”, the number of molecules received at the current symbol duration can be described as:

$$N_{hit_{-1}} = 0 \quad (\text{A. 6})$$

Thus the error probability of decoding the current intended symbol in the current time slot for this case can be described as:

$$P_{0000} = P(N_{hit_{-1}} > \tau) = 0 \quad (\text{A. 7})$$

For case “0001”, the number of molecules received at the current symbol duration can be described as:

$$N_{hit_{-2}} = N_C \quad (\text{A. 8})$$

Thus the error probability of decoding the current intended symbol in the current time slot for this case can be described as:

$$P_{0001} = P(N_{hit_{-2}} < \tau) = 1 - I_{P_1}(\tau, n - \tau + 1) \quad (\text{A. 9})$$

For case “0010”, the number of molecules received at the current symbol duration can be described as:

$$N_{hit_{-3}} = N_{P_{-1}} \quad (\text{A. 10})$$

Thus the error probability of decoding the current intended symbol in the current time slot for this case can be described as:

$$P_{0010} = P(N_{hit_3} > \tau) = Q\left(\frac{\tau - n(P_2 - P_1)}{\sqrt{n(M_2 + M_1)}}\right) \quad (\text{A. 11})$$

For case “0011”, the number of molecules received at the current symbol duration can be described as:

$$N_{hit_4} = N_{P_1} + N_C \sim \mathcal{N}(nP_2, n(2M_1 + M_2)) \quad (\text{A. 12})$$

Thus the error probability of decoding the current intended symbol in the current time slot for this case can be described as:

$$P_{0011} = P(N_{hit_4} < \tau) = 1 - Q\left(\frac{\tau - nP_2}{\sqrt{n(2M_1 + M_2)}}\right) \quad (\text{A. 13})$$

For case “0100”, the number of molecules received at the current symbol duration can be described as:

$$N_{hit_5} = N_{P_2} \quad (\text{A. 14})$$

Thus the error probability of decoding the current intended symbol in the current time slot for this case can be described as:

$$P_{0100} = P(N_{hit_5} > \tau) = Q\left(\frac{\tau - n(P_3 - P_2)}{\sqrt{n(M_3 + M_2)}}\right) \quad (\text{A. 15})$$

For case “0101”, the number of molecules received at the current symbol duration can be described as:

$$N_{hit_6} = N_{P_2} + N_C \sim \mathcal{N}(n(P_3 - P_2 + P_1), n(M_1 + M_2 + M_3)) \quad (\text{A. 16})$$

Thus the error probability of decoding the current intended symbol in the current time slot for this case can be described as:

$$P_{0101} = P(N_{hit_6} < \tau) = 1 - Q\left(\frac{\tau - n(P_3 - P_2 + P_1)}{\sqrt{n(M_1 + M_2 + M_3)}}\right) \quad (\text{A. 17})$$

For case “0110”, the number of molecules received at the current symbol duration can be described as:

$$N_{hit_7} = N_{P_2} + N_{P_1} \sim \mathcal{N}(n(P_3 - P_1), n(M_1 + 2M_2 + M_3)) \quad (\text{A. 18})$$

Thus the error probability of decoding the current intended symbol in the current time slot for this case can be described as:

$$P_{0110} = P(N_{hit_7} > \tau) = Q\left(\frac{\tau - n(P_3 - P_1)}{\sqrt{n(M_1 + 2M_2 + M_3)}}\right) \quad (\text{A. 19})$$

For case “0111”, the number of molecules received at the current symbol duration can be described as:

$$N_{hit_8} = N_C + N_{P_2} + N_{P_1} \sim \mathcal{N}(nP_3, n(2M_1 + 2M_2 + M_3)) \quad (\text{A. 20})$$

Thus the error probability of decoding the current intended symbol in the current time slot for this case can be described as:

$$P_{0111} = P(N_{hit_8} < \tau) = 1 - Q\left(\frac{\tau - nP_3}{\sqrt{n(2M_1 + 2M_2 + M_3)}}\right) \quad (\text{A. 21})$$

The error probabilities of the remaining 8 cases can be achieved using the same methods described above, the results of which have been given as follows.

For case “1000”, the error probability can be described as:

$$P_{1000} = Q\left(\frac{\tau - n(P_4 - P_3)}{\sqrt{n(M_4 + M_3)}}\right) \quad (\text{A. 22})$$

For case “1001”, the error probability can be described as:

$$P_{1001} = 1 - Q\left(\frac{\tau - n(P_4 - P_3 + P_1)}{\sqrt{n(M_4 + M_3 + M_1)}}\right) \quad (\text{A. 23})$$

For case “1010”, the error probability can be described as:

$$P_{1010} = Q\left(\frac{\tau - n(P_4 - P_3 + P_2 - P_1)}{\sqrt{n(M_4 + M_3 + M_2 + M_1)}}\right) \quad (\text{A. 24})$$

For case “1011”, the error probability can be described as:

$$P_{1011} = 1 - Q\left(\frac{\tau - n(P_4 - P_3 + P_2)}{\sqrt{n(M_4 + M_3 + M_2 + 2M_1)}}\right) \quad (\text{A. 25})$$

For case “1100”, the error probability can be described as:

$$P_{1100} = Q\left(\frac{\tau - n(P_4 - P_2)}{\sqrt{n(M_4 + 2M_3 + M_2)}}\right) \quad (\text{A. 26})$$

For case “1101”, the error probability can be described as:

$$P_{1101} = 1 - Q\left(\frac{\tau - n(P_1 - P_2 + P_4)}{\sqrt{n(M_4 + 2M_3 + M_2 + M_1)}}\right) \quad (\text{A. 27})$$

For case “1110”, the error probability can be described as:

$$P_{1110} = Q\left(\frac{\tau - n(P_4 - P_1)}{\sqrt{n(M_4 + 2M_3 + 2M_2 + M_1)}}\right) \quad (\text{A. 28})$$

For case “1111”, the error probability can be described as:

$$P_{1111} = 1 - Q\left(\frac{\tau - nP_4}{\sqrt{n(M_4 + 2M_3 + 2M_2 + 2M_1)}}\right) \quad (\text{A. 29})$$

Appendix B

In this section, a new channel model encompassing ISI, which has been proposed in [192], is taken into account. The new model takes into account the dependency of the molecules received in the successive time slots due to one transmission at the start. Specifically, in this model, for the i^{th} symbol, there exist 2^{i-1} different error patterns for a fixed transmitted symbol in the current time slot. Also, as stated in Section 4.5.3, α_j ($j \in [2, i]$) is used to represent the transmitted information symbol from the $(j-1)^{\text{th}}$ previous time slot. Different from equation (4.22), for the q^{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 given as [192]:

$$N_{p'(q)} \sim \sum_{j=2}^i a_j \left(B(n, p_{j-1}^j) \right) \sim \tag{B. 1}$$

$$\sum_{j=2}^i a_j \left(\mathcal{N}(np_{j-1}^j, np_{j-1}^j(1-p_{j-1}^j)) \right) \sim \mathcal{N}(\mu_{p'(q)}, \sigma_{p'(q)}^2)$$

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 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 $N_{hit'(q)}$ received in the i^{th} symbol duration is the sum of N_c and $N_{p'(q)}$, where N_c has been given in equation (4.21). The average bit error probability can be calculated similar with equations (4.23-4.27) in Section 4.5.3. Figure B.1 gives comparisons of the channel BER performances between the proposed model in Chapter 4 and the new model in this section. It is obvious that the probability of one bit error is higher in the

proposed model in this thesis. In other words, when considering the incomplete independence between $B(n, P_j)$ and $B(n, P_{j-1})$, the error probability of the channel will be lower.

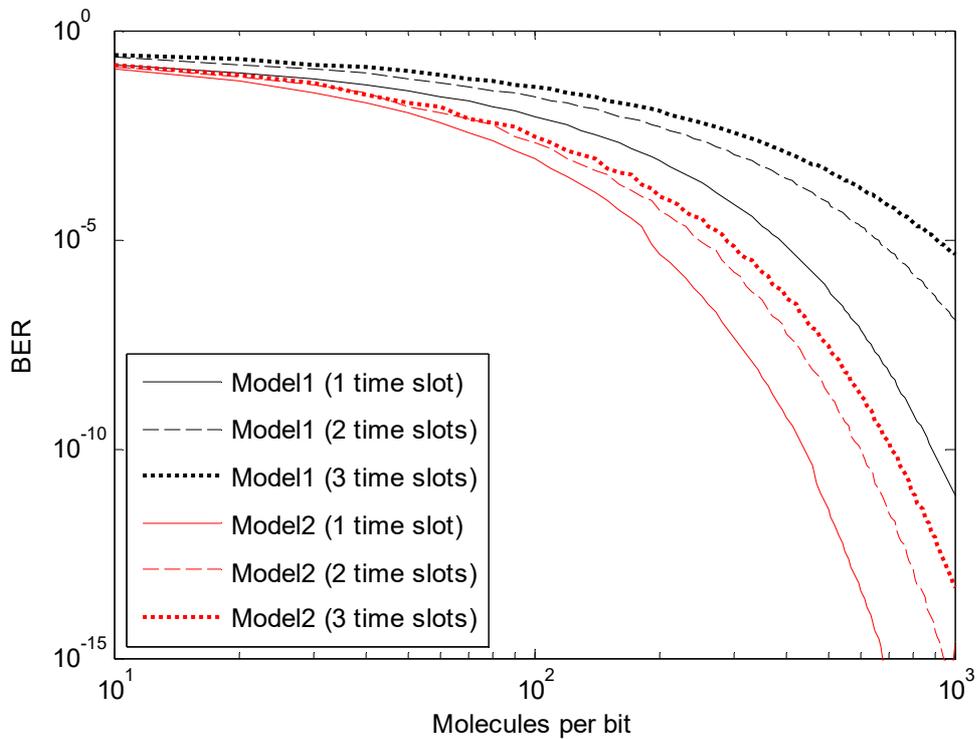


Figure B.1 Comparison of BER performances between the proposed model in this thesis and the new model.

Model1 represents the proposed model in Chapter 4 and Model2 represents the novel model proposed in this section. The relevant parameters include transmission distance $d = 8\mu m$ and the number of bacteria in the receiver $m = 100$.

References

- [1] A. L. Porter and J. Youtie, "How interdisciplinary is nanotechnology?," *Journal of nanoparticle research*, vol. 11, pp. 1023-1041, 2009.
- [2] M. C. Roco, R. S. Williams, and P. Alivisatos, *Nanotechnology Research Directions: IWGN Workshop Report: Vision for Nanotechnology in the Next Decade*: Springer Science & Business Media, 2000.
- [3] G. Hunt and M. Mehta, *Nanotechnology: "Risk, Ethics and Law"*: Routledge, 2013.
- [4] J. M. Jornet and I. F. Akyildiz, "Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band," *IEEE Transactions on Wireless Communications*, vol. 10, pp. 3211-3221, 2011.
- [5] I. F. Akyildiz, F. Brunetti, and C. Blázquez, "Nanonetworks: A new communication paradigm," *Computer Networks: The International Journal of Computer and Telecommunications Networking*, vol. 52, pp. 2260-2279, 2008.
- [6] N. Roca Lacasa, "Modeling the molecular communication nanonetworks," Universitat Politècnica de Catalunya, 2009.
- [7] I. F. Akyildiz, J. M. Jornet, and M. Pierobon, "Nanonetworks: a new frontier in communications," *Communications of the ACM*, vol. 54, pp. 84-89, 2011.
- [8] M. Moore, A. Enomoto, T. Nakano, R. Egashira, T. Suda, A. Kayasuga, *et al.*, "A design of a molecular communication system for nanomachines using molecular motors," in *Pervasive Computing and Communications Workshops, 2006. PerCom Workshops 2006. Fourth Annual IEEE International Conference on*, Pisa, 2006, pp. 554-559.
- [9] C. Mohanty, G. Arya, R. S. Verma, and S. K. Sahoo, "Nanobiotechnology: application of nanotechnology in therapeutics and diagnosis," *International Journal of Green Nanotechnology: Biomedicine*, vol. 1, pp. B24-B38, 2009.
- [10] I. F. Akyildiz, F. Fekri, R. Sivakumar, C. R. Forest, and B. K. Hammer, "Monaco: fundamentals of molecular nano-communication networks," *Wireless Communications, IEEE*, vol. 19, pp. 12-18, 2012.
- [11] S. G. Glisic, *Advanced Wireless Communications and Internet: Future Evolving Technologies*, 3rd ed.: John Wiley & Sons Ltd, 2011.
- [12] B. L. Bassler, "How bacteria talk to each other: regulation of gene expression by quorum sensing," *Current Opinion in Microbiology*, vol. 2, pp. 582-587, 1999.
- [13] S. F. Bush, *Nanoscale communication networks*. Boston: Artech House, 2010.
- [14] T. Nakano, A. W. Eckford, and T. Haraguchi, *Molecular Communication*: Cambridge University Press, 2013.
- [15] V. Singh, "Recent advancements in synthetic biology: current status and challenges," *Gene*, vol. 535, pp. 1-11, 2014.
- [16] I. Llatser, A. Cabellos-Aparicio, and E. Alarcon, "Networking challenges and principles in diffusion-based molecular communication," *Wireless Communications, IEEE*, vol. 19, pp. 36-41, 2012.
- [17] T. Nakano, T. Suda, T. Koujin, T. Haraguchi, and Y. Hiraoka, "Molecular communication through gap junction channels: System design, experiments

- and modeling," in *Bio-Inspired Models of Network, Information and Computing Systems, 2007. Bionetics 2007. 2nd*, Budapest, 2007, pp. 139-146.
- [18] H. Shahmohammadian, G. G. Messier, and S. Magierowski, "Optimum receiver for molecule shift keying modulation in diffusion-based molecular communication channels," *Nano Communication Networks*, vol. 3, pp. 183-195, 2012.
- [19] P. J. Shih, C. H. Lee, and P. C. Yeh, "Channel codes for mitigating intersymbol interference in diffusion-based molecular communications," in *2012 IEEE Global Communications Conference (GLOBECOM)*, , 2012, pp. 4228-4232.
- [20] F. Walsh, "Protocols for Molecular Communication Nanonetworks," Department of Computing, Mathematics, and Physics, Waterford Institute of Technology, Waterford, Ireland, 2013.
- [21] S. Abadal and I. F. Akyildiz, "Automata modeling of quorum sensing for nanocommunication networks," *Nano Communication Networks*, vol. 2, pp. 74-83, 2011.
- [22] A. Einolghozati, M. Sardari, and F. Fekri, "Design and Analysis of Wireless Communication Systems Using Diffusion-Based Molecular Communication Among Bacteria," *Wireless Communications, IEEE Transactions on*, vol. 12, pp. 6096-6105, 2013.
- [23] J. Philibert, "One and a half century of diffusion: Fick, Einstein, before and beyond," *Diffusion Fundamentals*, vol. 2, pp. 1-10, 2005.
- [24] M. S. Leeson and M. D. Higgins, "Forward error correction for molecular communications," *Nano Communication Networks*, vol. 3, pp. 161-167, 2012.
- [25] L. Wang, K. Qian, Y. Huang, N. Jin, H. Lai, T. Zhang, *et al.*, "SynBioLGDB: a resource for experimentally validated logic gates in synthetic biology," *Scientific reports*, vol. 5, 2015.
- [26] T. Nakano and J. Q. Liu, "Design and Analysis of Molecular Relay Channels: An Information Theoretic Approach," *NanoBioscience, IEEE Transactions on*, vol. 9, pp. 213-221, 2010.
- [27] T. Nakano, M. J. Moore, F. Wei, A. V. Vasilakos, and J. Shuai, "Molecular Communication and Networking: Opportunities and Challenges," *NanoBioscience, IEEE Transactions on*, vol. 11, pp. 135-148, 2012.
- [28] S. Hiyama, Y. Moritani, T. Suda, R. Egashira, A. Enomoto, M. Moore, *et al.*, "Molecular communication," *Journal-Institute of Electronics Information and Communication Engineers*, vol. 89, p. 162, 2006.
- [29] P. Liò, E. Yoneki, J. Crowcroft, and E. C. Verma, *Bio-Inspired Computing and Communication* vol. 5151: Springer-Verlag Berlin Heidelberg, 2008.
- [30] D. Malak and O. B. Akan, "Molecular communication nanonetworks inside human body," *Nano Communication Networks*, vol. 3, pp. 19-35, 2012.
- [31] P. He, Y. Mao, Q. Liu, and K. Yang, "A Diffusion-Neuron Hybrid System for Molecular Communication," *arXiv preprint arXiv:1507.01060*, 2015.
- [32] O. Hayden and K. Nielsch, *Molecular-and Nano-tubes*: Springer Science & Business Media, 2011.
- [33] T. Nakano, M. Moore, A. Enomoto, and T. Suda, "Molecular communication technology as a biological ICT," in *Biological functions for information and communication technologies*, ed: Springer, 2011, pp. 49-86.
- [34] C. Appert-Rolland, M. Ebbinghaus, and L. Santen, "Intracellular transport driven by cytoskeletal motors: General mechanisms and defects," *Physics Reports*, vol. 593, pp. 1-59, 2015.

- [35] G. M. Cooper, *The Cell*: Sinauer Associates, 2000.
- [36] B. Atakan, *Molecular Communications and Nanonetworks*: Springer, 2014.
- [37] B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, *et al.*, *Essential cell biology*: Garland Science, 2013.
- [38] M. Ş. Kuran, H. B. Yilmaz, T. Tugcu, and B. Özerman, "Energy model for communication via diffusion in nanonetworks," *Nano Communication Networks*, vol. 1, pp. 86–95, 2010.
- [39] G. Söhl, S. Maxeiner, and K. Willecke, "Expression and functions of neuronal gap junctions," *Nature Reviews Neuroscience*, vol. 6, pp. 191-200, 2005.
- [40] C. Elfgang, R. Eckert, H. Lichtenberg-Fraté, A. Butterweck, O. Traub, R. A. Klein, *et al.*, "Specific permeability and selective formation of gap junction channels in connexin-transfected HeLa cells," *The Journal of Cell Biology*, vol. 129, pp. 805-817, 1995.
- [41] R. J. Koscijew, *The Designing Theory of Transference* vol. 2: AuthorHouse, 2012.
- [42] V. Shah, *Emerging environmental technologies*: Springer, 2008.
- [43] S. Balasubramaniam, "Opportunistic routing through conjugation in bacteria communication nanonetwork," *Nano Communication Networks*, vol. 3, pp. 36-45, 2012.
- [44] M. Pierobon, "Fundamentals of diffusion-based molecular communication in nanonetworks," Georgia Institute of Technology, 2013.
- [45] Y. Gerchman and R. Weiss, "Teaching bacteria a new language," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, pp. 2221-2222, 2004.
- [46] M. Gregori and I. F. Akyildiz, "A new nanonetwork architecture using flagellated bacteria and catalytic nanomotors," *IEEE Journal on Selected Areas in Communications*, vol. 28, pp. 612-619, 2010.
- [47] R. Paton and M. Holcombe, *Information Processing in Cells and Tissues*: World Scientific, 1998.
- [48] C. Priami, F. Dressler, O. B. Akan, and A. Ngom, *Transactions on Computational Systems Biology X*: Springer, 2008.
- [49] B. Hanawalt, *Cellular Computing*: Oxford University Press, USA, 2004.
- [50] M. Caldorera-Moore and N. A. Peppas, "Micro-and nanotechnologies for intelligent and responsive biomaterial-based medical systems," *Advanced drug delivery reviews*, vol. 61, pp. 1391-1401, 2009.
- [51] H. Hess and G. D. Bachand, "Biomolecular motors," *MaterialsToday*, vol. 8, pp. 22-29, 2005.
- [52] R. Frei, R. McWilliam, B. Derrick, A. Purvis, A. Tiwari, and G. D. M. Serugendo, "Self-healing and self-repairing technologies," *The International Journal of Advanced Manufacturing Technology*, vol. 69, pp. 1033-1061, 2013.
- [53] H. Kitano, "Biological robustness," *Nature Reviews Genetics*, vol. 5, pp. 826-837, 2004.
- [54] D. K. Sharma, A. Mishra, and R. Saxena, "Analog & Digital Modulation Techniques: An overview," *International Journal of Computing Science and Communication Technologies*, vol. 3, p. 2007, 2010.
- [55] W. Chang and A. V. Vasilakos, *Molecular Computing*: Springer International Publishing, 2014.

- [56] S. Kadloor, R. S. Adve, and A. W. Eckford, "Molecular communication using brownian motion with drift," *IEEE Transactions on NanoBioscience*, vol. 11, pp. 89-99, 2012.
- [57] M. U. Mahfuz, D. Makrakis, and H. T. Mouftah, "On the characterization of binary concentration-encoded molecular communication in nanonetworks," *Nano Communication Networks*, vol. 1, pp. 289-300, 2010.
- [58] M. S. Kuran, H. B. Yilmaz, T. Tugcu, and I. F. Akyildiz, "Modulation techniques for communication via diffusion in nanonetworks," in *Communications (ICC), 2011 IEEE International Conference on*, Kyoto, 2011, pp. 1-5.
- [59] N. R. Kim and C. B. Chae, "Novel modulation techniques using isomers as messenger molecules for nano communication networks via diffusion," *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 847-856, 2013.
- [60] N. Garralda, I. Llatser, A. Cabellos-Aparicio, E. Alarcón, and M. Pierobon, "Diffusion-based physical channel identification in molecular nanonetworks," *Nano Communication Networks*, vol. 2, pp. 196-204, 2011.
- [61] B. Tepekule, A. E. Pusane, H. B. Yilmaz, and T. Tugcu, "Energy efficient ISI mitigation for communication via diffusion," in *Communications and Networking (BlackSeaCom), 2014 IEEE International Black Sea Conference on*, Odessa, 2014, pp. 33-37.
- [62] B. Atakan, S. Galmés, and O. B. Akan, "Nanoscale communication with molecular arrays in nanonetworks," *IEEE Transactions on NanoBioscience*, vol. 11, pp. 149-160, 2012.
- [63] T. M. Cover and J. A. Thomas, "Elements of information theory 2nd edition," 2006.
- [64] B. Atakan and O. Akan, "An information theoretical approach for molecular communication," in *Bio-Inspired Models of Network, Information and Computing Systems, 2007. Bionetics 2007. 2nd*, 2007, pp. 33-40.
- [65] B. Atakan and O. B. Akan, "Single and multiple-access channel capacity in molecular nanonetworks," in *Nano-Net*, ed: Springer Berlin Heidelberg, 2009, pp. 14-23.
- [66] B. Atakan and O. B. Akan, "Deterministic capacity of information flow in molecular nanonetworks," *Nano Communication Networks*, vol. 1, pp. 31-42, 2010.
- [67] A. Einolghozati, M. Sardari, and F. Fekri, "Capacity of diffusion-based molecular communication with ligand receptors," in *Information Theory Workshop (ITW), 2011 IEEE*, Paraty, 2011, pp. 85-89.
- [68] A. Einolghozati, M. Sardari, A. Beirami, and F. Fekri, "Capacity of discrete molecular diffusion channels," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, St. Petersburg, 2011, pp. 723-727.
- [69] D. Arifler, "Capacity analysis of a diffusion-based short-range molecular nano-communication channel," *Computer Networks*, vol. 55, pp. 1426-1434, 2011.
- [70] B. Atakan, "Optimal transmission probability in binary molecular communication," *Communications Letters, IEEE*, vol. 17, pp. 1152-1155, 2013.
- [71] M. Pierobon and I. F. Akyildiz, "A physical end-to-end model for molecular communication in nanonetworks," *Selected Areas in Communications, IEEE Journal on*, vol. 28, pp. 602-611, 2010.

- [72] M. Pierobon and I. F. Akyildiz, "Diffusion-Based Noise Analysis for Molecular Communication in Nanonetworks," *IEEE Transactions on Signal Processing*, vol. 59, pp. 2532-2547, 2011.
- [73] M. Pierobon and I. F. Akyildiz, "Noise Analysis in Ligand-Binding Reception for Molecular Communication in Nanonetworks," *IEEE Transactions on Signal Processing*, vol. 59, pp. 4168-4182, 2011.
- [74] K. V. Srinivas, A. W. Eckford, and R. S. Adve, "Molecular Communication in Fluid Media: The Additive Inverse Gaussian Noise Channel," *IEEE Transactions on Information Theory*, vol. 58, pp. 4678-4692, 2012.
- [75] M. J. Moore, T. Suda, and K. Oiwa, "Molecular Communication: Modeling Noise Effects on Information Rate," *IEEE Transactions on NanoBioscience*, vol. 8, pp. 169-180, 2009.
- [76] N. Farsad, A. W. Eckford, S. Hiyama, and Y. Moritani, "A simple mathematical model for information rate of active transport molecular communication," in *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*, Shanghai, 2011, pp. 473-478.
- [77] J. Kurose and K. Ross, *Computer Networking: A Top-Down Approach*, 6th ed.: Addison Wesley, 2012.
- [78] D. M. Piscitello and A. L. Chapin, *Open systems networking: TCP/IP and OSI*: Addison Wesley, 1993.
- [79] T. Nakano, T. Suda, Y. Okaie, M. Moore, and A. Vasilakos, "Molecular Communication Among Biological Nanomachines: A Layered Architecture and Research Issues," *NanoBioscience, IEEE Transactions on*, 2014.
- [80] M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators," *Nature*, vol. 403, pp. 335-338, 2000.
- [81] M. J. Moore and T. Nakano, "Synchronization of inhibitory molecular spike oscillators," in *Bio-Inspired Models of Networks, Information, and Computing Systems*, ed: Springer Berlin Heidelberg, 2012, pp. 183-195.
- [82] A. Akkaya, G. Genc, and T. Tugcu, "HLA based architecture for molecular communication simulation," *Simulation Modelling Practice and Theory*, vol. 42, pp. 163-177, 2014.
- [83] I. Llatser, D. Demiray, A. Cabellos-Aparicio, D. T. Altılar, and E. Alarcón, "N3Sim: Simulation framework for diffusion-based molecular communication nanonetworks," *Simulation Modelling Practice and Theory*, vol. 42, pp. 210-222, 2014.
- [84] H. B. Yilmaz and C. B. Chae, "Simulation study of molecular communication systems with an absorbing receiver: Modulation and ISI mitigation techniques," *Simulation Modelling Practice and Theory*, vol. 49, pp. 136-150, 2014.
- [85] E. Gul, B. Atakan, and O. B. Akan, "NanoNS: A nanoscale network simulator framework for molecular communications," *Nano Communication Networks*, vol. 1, pp. 138-156, 2010.
- [86] L. Felicetti, M. Femminella, and G. Reali, "A simulation tool for nanoscale biological networks," *Nano Communication Networks*, vol. 3, pp. 2-18, 2012.
- [87] G. Wei, P. Bogdan, and R. Marculescu, "Efficient modeling and simulation of bacteria-based nanonetworks with bnsim," *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 868-878, 2013.
- [88] M. Patil, D. S. Mehta, and S. Guvva, "Future impact of nanotechnology on medicine and dentistry," *Journal of Indian Society of Periodontology*, vol. 12, pp. 34-40, 2008.

- [89] J. Shi, A. R. Votruba, O. C. Farokhzad, and R. Langer, "Nanotechnology in Drug Delivery and Tissue Engineering: From Discovery to Applications," *Nano letters*, vol. 10, pp. 3223-3230, 2010.
- [90] T. M. Allen and P. R. Cullis, "Drug delivery systems: entering the mainstream," *Science*, vol. 303, pp. 1818-1822, 2004.
- [91] L. G. Griffith and G. Naughton, "Tissue engineering--current challenges and expanding opportunities," *Science*, vol. 295, pp. 1009-1014, 2002.
- [92] M. S. Chapekar, "Tissue engineering: challenges and opportunities," *Journal of biomedical materials research*, vol. 53, pp. 617-620, 2000.
- [93] E. De Leo, L. Galluccio, A. Lombardo, and G. Morabito, "Networked Labs-on-a-Chip (NLoC): Introducing networking technologies in microfluidic systems," *Nano Communication Networks*, vol. 3, pp. 217-228, 2012.
- [94] J. Timmis, T. Knight, L. N. de Castro, and E. Hart, "An overview of artificial immune systems," in *Computation in Cells and Tissues*, ed: Springer, 2004, pp. 51-91.
- [95] Y. Okaie, T. Nakano, T. Hara, K. Hosoda, Y. Hiraoka, and S. Nishio, "Cooperative target tracking by a bacterium-based mobile sensor network," *IEEE Transactions on Nanobioscience*, vol. 13, pp. 267-277, 2014.
- [96] J. W. Costerton, P. S. Stewart, and E. P. Greenberg, "Bacterial biofilms: a common cause of persistent infections," *Science*, vol. 284, pp. 1318-1322, 1999.
- [97] C. J. Kerr, K. S. Osborn, G. D. Robson, and P. S. Handley, "The relationship between pipe material and biofilm formation in a laboratory model system," *Journal of applied microbiology*, vol. 85, pp. 29S-38S, 1998.
- [98] A. J. Bai and V. R. Rai, "Bacterial quorum sensing and food industry," *Comprehensive Reviews in Food Science and Food Safety*, vol. 10, pp. 183-193, 2011.
- [99] P. P. Simeonova, N. Opopol, and M. I. Luster, *Nanotechnology--toxicological Issues and Environmental Safety*: Springer, 2007.
- [100] L. Marques and A. T. De Almeida, "Electronic nose-based odour source localization," in *Advanced Motion Control, 2000. Proceedings. 6th International Workshop on*, Nagoya, Japan, 2000, pp. 36-40.
- [101] F. Farhan and M. Mushfiq, "BIOLOGICAL CELL AND MOLECULAR COMMUNICATION TECHNOLOGY: OVERVIEW AND CHALLENGES," *European Scientific Journal*, vol. 11, 2015.
- [102] S. Hiyama, Y. Moritani, and T. Suda, "Molecular transport system in molecular communication," *NTT DOCOMO Technical Journal*, vol. 10, pp. 49-53, 2008.
- [103] R. A. Freitas, *Nanomedicine, volume I: basic capabilities: Georgetown, TX, USA: Landes Bioscience*, 1999.
- [104] E. Ben-Jacob, "Social behavior of bacteria: from physics to complex organization," *The European Physical Journal B*, vol. 65, pp. 315-322, 2008.
- [105] K. H. Nealson, T. Platt, and J. W. Hastings, "Cellular control of the synthesis and activity of the bacterial luminescent system," *Journal of Bacteriology*, vol. 104, pp. 313-322, 1970.
- [106] H. B. Kaplan and E. P. Greenberg, "Diffusion of autoinducer is involved in regulation of the *Vibrio fischeri* luminescence system," *Journal of Bacteriology*, vol. 163, pp. 1210-1214, 1985.

- [107] S. T. Rutherford and B. L. Bassler, "Bacterial quorum sensing: its role in virulence and possibilities for its control," *Cold Spring Harbor Perspectives in Medicine*, vol. 2, 2012.
- [108] W. B. Miller Jr, *The Microcosm Within: Evolution and Extinction in the Hologenome*: Universal-Publishers, 2013.
- [109] N. G. Ravichandra, *Fundamentals of Plant Pathology*: Prentice-Hall of India Pvt.Ltd, 2013.
- [110] H. C. J. Gram and C. Friedlaender, *Ueber die isolirte Färbung der Schizomyceten: in Schnitt-und Trockenpräparaten*: Theodor Fischer's medicinischer Buchhandlung, 1884.
- [111] S. W. Chiu and M. C. Leake, "Functioning nanomachines seen in real-time in living bacteria using single-molecule and super-resolution fluorescence imaging," *International journal of molecular sciences*, vol. 12, pp. 2518-2542, 2011.
- [112] E. B. Jacob, Y. Shapira, and A. I. Tauber, "Seeking the foundations of cognition in bacteria: From Schrödinger's negative entropy to latent information," *Physica A: Statistical Mechanics and its Applications*, vol. 359, pp. 495-524, 2006.
- [113] S. Baron, *Medical Microbiology*: University of Texas Medical Branch at Galveston.
- [114] S. Schauder and B. L. Bassler, "The languages of bacteria," *Genes & Development*, vol. 15, pp. 1468-1480, 2001.
- [115] J. W. Hastings and K. H. Nealson, "Bacterial bioluminescence," *Annual Reviews in Microbiology*, vol. 31, pp. 549-595, 1977.
- [116] M. McFall-Ngai, "Divining the essence of symbiosis: Insights from the squid-vibrio model," *PLoS Biology*, vol. 12, p. e1001783, 2014.
- [117] W. C. Fuqua, S. C. Winans, and E. P. Greenberg, "Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators," *Journal of bacteriology*, vol. 176, p. 269, 1994.
- [118] B. L. Bassler, M. Wright, R. E. Showalter, and M. R. Silverman, "Intercellular signalling in *Vibrio harveyi*: sequence and function of genes regulating expression of luminescence," *Molecular microbiology*, vol. 9, pp. 773-786, 1993.
- [119] B. L. Bassler, M. Wright, and M. R. Silverman, "Multiple signalling systems controlling expression of luminescence in *Vibrio harveyi*: sequence and function of genes encoding a second sensory pathway," *Molecular microbiology*, vol. 13, pp. 273-286, 1994.
- [120] M. Kleerebezem, L. E. Quadri, O. P. Kuipers, and W. M. De Vos, "Quorum sensing by peptide pheromones and two - component signal - transduction systems in Gram - positive bacteria," *Molecular microbiology*, vol. 24, pp. 895-904, 1997.
- [121] C. Fuqua and A. Eberhard, "Signal generation in autoinduction systems: synthesis of acylated homoserine lactones by LuxI-type proteins," *Cell-cell signaling in bacteria*. ASM Press, Washington, DC, pp. 211-230, 1999.
- [122] E. O. Burton, H. W. Read, M. C. Pellitteri, and W. J. Hickey, "Identification of acyl-homoserine lactone signal molecules produced by *Nitrosomonas europaea* strain Schmidt," *Applied and environmental microbiology*, vol. 71, pp. 4906-4909, 2005.
- [123] C. D' Angelo - Picard, D. Faure, I. Penot, and Y. Dessaux, "Diversity of N - acyl homoserine lactone - producing and - degrading bacteria in soil and

- tobacco rhizosphere," *Environmental microbiology*, vol. 7, pp. 1796-1808, 2005.
- [124] G. Dunphy, C. Miyamoto, and E. Meighen, "A homoserine lactone autoinducer regulates virulence of an insect-pathogenic bacterium, *Xenorhabdus nematophilus* (Enterobacteriaceae)," *Journal of bacteriology*, vol. 179, pp. 5288-5291, 1997.
- [125] U. M. Pinto, E. de Souza Viana, M. L. Martins, and M. C. D. Vanetti, "Detection of acylated homoserine lactones in gram-negative proteolytic psychrotrophic bacteria isolated from cooled raw milk," *Food Control*, vol. 18, pp. 1322-1327, 2007.
- [126] M. Boyer and F. Wisniewski-Dyé, "Cell-cell signalling in bacteria: not simply a matter of quorum," *FEMS microbiology ecology*, vol. 70, pp. 1-19, 2009.
- [127] E. G. Ruby and K. H. Nealson, "Symbiotic association of *Photobacterium fischeri* with the marine luminous fish *Monocentris japonica*: a model of symbiosis based on bacterial studies," *The Biological bulletin*, vol. 151, pp. 574-586, 1976.
- [128] E. G. Ruby and M. J. McFall-Ngai, "Oxygen-utilizing reactions and symbiotic colonization of the squid light organ by *Vibrio fischeri*," *Trends in microbiology*, vol. 7, pp. 414-420, 1999.
- [129] S. V. Nyholm and M. J. McFall-Ngai, "Sampling the light-organ microenvironment of *Euprymna scolopes*: description of a population of host cells in association with the bacterial symbiont *Vibrio fischeri*," *The Biological Bulletin*, vol. 195, pp. 89-97, 1998.
- [130] J. Engebrecht, K. Nealson, and M. Silverman, "Bacterial bioluminescence: isolation and genetic analysis of functions from *Vibrio fischeri*," *Cell*, vol. 32, pp. 773-781, 1983.
- [131] J. Engebrecht and M. Silverman, "Identification of genes and gene products necessary for bacterial bioluminescence," *Proceedings of the National Academy of Sciences*, vol. 81, pp. 4154-4158, 1984.
- [132] T. Charrier, G. Thouand, H. Gezekel, M. Affi, M. J. Durand, and S. Jouanneau, *Bacterial bioluminescent biosensor characterisation for on-line monitoring of heavy metals pollutions in waste water treatment plant effluents*, *Biosensors*, Pier Andrea Serra (Ed.): INTECH Open Access Publisher, 2010.
- [133] K. M. Gray and J. R. Garey, "The evolution of bacterial LuxI and LuxR quorum sensing regulators," *Microbiology*, vol. 147, pp. 2379-2387, 2001.
- [134] Ben N. G. Giepmans, S. R. Adams, M. H. Ellisman, and R. Y. Tsien, "The fluorescent toolbox for assessing protein location and function," *Science*, vol. 312, pp. 217-224, 2006.
- [135] D. Kunkel. (2015). *Vibrio fischeri* _ bioluminescent, symbiotic bacterium. Available: <http://www.denniskunkel.com/detail/14209.html>
- [136] McFall-Ngai. (2010). *Sustaining symbiosis – new clues*. Available: <http://whyfiles.org/2010/sustaining-symbiosis-new-clues/>
- [137] K. H. Nealson and J. W. Hastings, "Bacterial bioluminescence: its control and ecological significance," *Microbiological reviews*, vol. 43, p. 496, 1979.
- [138] M. B. Miller and B. L. Bassler, "Quorum sensing in bacteria," *Annual Reviews in Microbiology*, vol. 55, pp. 165-199, 2001.

- [139] H. Liu, S. Srinivas, H. He, G. Gong, C. Dai, Y. Feng, *et al.*, "Quorum sensing in *Vibrio* and its relevance to bacterial virulence," *Journal of Bacteriology & Parasitology*, vol. 4, p. 2, 2013.
- [140] T. R. De Kievit and B. H. Iglewski, "Bacterial quorum sensing in pathogenic relationships," *Infection and Immunity*, vol. 68, pp. 4839-4849, 2000.
- [141] T. Vo-Dinh, *Nanotechnology in biology and medicine: methods, devices, and applications*: CRC Press, 2007.
- [142] S. Payne and L. You, "Engineered cell–cell communication and its applications," in *Productive Biofilms*, ed: Springer, 2014, pp. 97-121.
- [143] J. S. Chuang, "Engineering multicellular traits in synthetic microbial populations," *Current opinion in chemical biology*, vol. 16, pp. 370-378, 2012.
- [144] A. Pai, Y. Tanouchi, C. H. Collins, and L. You, "Engineering multicellular systems by cell–cell communication," *Current opinion in biotechnology*, vol. 20, pp. 461-470, 2009.
- [145] R. A. Meyers, *Synthetic Biology*: Wiley-Blackwell, 2015.
- [146] H. Kobayashi, M. Kærn, M. Araki, K. Chung, T. S. Gardner, C. R. Cantor, *et al.*, "Programmable cells: interfacing natural and engineered gene networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, pp. 8414-8419, 2004.
- [147] K. Brenner, D. K. Karig, R. Weiss, and F. H. Arnold, "Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium," *Proceedings of the National Academy of Sciences*, vol. 104, pp. 17300-17304, 2007.
- [148] R. P. Shetty, D. Endy, and T. F. K. Jr., "Engineering BioBrick vectors from BioBrick parts," *Journal of Biological Engineering*, vol. 2, pp. 1-12, 2008.
- [149] S. Choudhary and C. Schmidt-Dannert, "Applications of quorum sensing in biotechnology," *Applied microbiology and biotechnology*, vol. 86, pp. 1267-1279, 2010.
- [150] T. Köhler, A. Buckling, and C. van Delden, "Cooperation and virulence of clinical *Pseudomonas aeruginosa* populations," *Proceedings of the National Academy of Sciences*, vol. 106, pp. 6339-6344, 2009.
- [151] L. Li, D. Hooi, S. R. Chhabra, D. Pritchard, and P. E. Shaw, "Bacterial N-acylhomoserine lactone-induced apoptosis in breast carcinoma cells correlated with down-modulation of STAT3," *Oncogene*, vol. 23, pp. 4894-4902, 2004.
- [152] D. M. Yebra, S. Kiil, and K. Dam-Johansen, "Antifouling technology—past, present and future steps towards efficient and environmentally friendly antifouling coatings," *Progress in organic coatings*, vol. 50, pp. 75-104, 2004.
- [153] M. Hentzer, K. Riedel, T. B. Rasmussen, A. Heydorn, J. B. Andersen, M. R. Parsek, *et al.*, "Inhibition of quorum sensing in *Pseudomonas aeruginosa* biofilm bacteria by a halogenated furanone compound," *Microbiology*, vol. 148, pp. 87-102, 2002.
- [154] E. M. Comelli, S. Lariani, M. C. Zwahlen, G. Fotopoulos, J. A. Holzwarth, C. Cherbut, *et al.*, "Biomarkers of human gastrointestinal tract regions," *Mammalian Genome*, vol. 20, pp. 516-527, 2009.
- [155] E. Livak-Dahl, I. Sinn, and M. Burns, "Microfluidic chemical analysis systems," vol. 2, pp. 325-353, 2011.
- [156] D. Skoog, D. West, F. L. Holler, and S. Crouch, *Fundamentals of analytical chemistry*: Cengage Learning, 2013.

- [157] P. Marguet, Y. Tanouchi, E. Spitz, C. Smith, and L. You, "Oscillations by minimal bacterial suicide circuits reveal hidden facets of host-circuit physiology," *PloS one*, vol. 5, p. e11909, 2010.
- [158] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, pp. 1183-1186, 2002.
- [159] A. Varma and B. O. Palsson, "Metabolic capabilities of Escherichia coli: I. Synthesis of biosynthetic precursors and cofactors," *Journal of theoretical biology*, vol. 165, pp. 477-502, 1993.
- [160] J. W. Young, J. C. W. Locke, A. Altinok, N. Rosenfeld, T. Bacarian, P. S. Swain, *et al.*, "Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy," *nature protocols*, vol. 7, pp. 80-88, 2012.
- [161] A. S. Khalil and J. J. Collins, "Synthetic biology: applications come of age," *Nature Reviews Genetics*, vol. 11, pp. 367-379, 2010.
- [162] S. Ma, N. Tang, and J. Tian, "DNA synthesis, assembly and applications in synthetic biology," *Current opinion in chemical biology*, vol. 16, pp. 260-267, 2012.
- [163] Y. Tanouchi, D. Tu, J. Kim, and L. You, "Noise reduction by diffusional dissipation in a minimal quorum sensing motif," *PLOS Computational Biology*, vol. 4, p. e1000167, 2008.
- [164] S. James, P. Nilsson, G. James, S. Kjelleberg, and T. Fagerström, "Luminescence control in the marine bacterium *Vibrio fischeri*: an analysis of the dynamics of lux regulation," *Journal of molecular biology*, vol. 296, pp. 1127-1137, 2000.
- [165] J. D. Dockery and J. P. Keener, "A mathematical model for quorum sensing in *Pseudomonas aeruginosa*," *Bulletin of mathematical biology*, vol. 63, pp. 95-116, 2001.
- [166] J. P. Ward, J. R. King, A. J. Koerber, P. Williams, J. M. Croft, and R. E. Sockett, "Mathematical modelling of quorum sensing in bacteria," *Mathematical Medicine and Biology*, vol. 18, pp. 263-292, 2001.
- [167] H. H. Atlan, F. Fogelman-Soulie, J. Salomon, and G. Weisbuch, "Random boolean networks," *Cybernetics and System*, vol. 12, pp. 103-121, 1981.
- [168] H. Lähdesmäki, S. Hautaniemi, I. Shmulevich, and O. Yli-Harja, "Relationships between probabilistic Boolean networks and dynamic Bayesian networks as models of gene regulatory networks," *Signal Processing*, vol. 86, pp. 814-834, 2006.
- [169] L. J. Steggles, R. Banks, and A. Wipat, "Modelling and analysing genetic networks: From Boolean networks to Petri nets," pp. 127-141.
- [170] K. Anguige, J. R. King, J. P. Ward, and P. Williams, "Mathematical modelling of therapies targeted at bacterial quorum sensing," *Mathematical biosciences*, vol. 192, pp. 39-83, 2004.
- [171] E. L. Haseltine and F. H. Arnold, "Implications of rewiring bacterial quorum sensing," *Applied and environmental microbiology*, vol. 74, pp. 437-445, 2008.
- [172] A. J. Koerber, J. R. King, J. P. Ward, P. Williams, J. M. Croft, and R. E. Sockett, "A mathematical model of partial-thickness burn-wound infection by *Pseudomonas aeruginosa*: quorum sensing and the build-up to invasion," *Bulletin of mathematical biology*, vol. 64, pp. 239-259, 2002.

- [173] D. L. Chopp, M. J. Kirisits, B. Moran, and M. R. Parsek, "The dependence of quorum sensing on the depth of a growing biofilm," *Bulletin of mathematical biology*, vol. 65, pp. 1053-1079, 2003.
- [174] A. B. Goryachev, D. J. Toh, K. B. Wee, T. Lee, H. B. Zhang, and L. H. Zhang, "Transition to quorum sensing in an *Agrobacterium* population: A stochastic model," *PLoS Comput Biol*, vol. 1, p. e37, 2005.
- [175] I. B. Bischofs, J. A. Hug, A. W. Liu, D. M. Wolf, and A. P. Arkin, "Complexity in bacterial cell-cell communication: Quorum signal integration and subpopulation signaling in the *Bacillus subtilis* phosphorelay," *Proceedings of the National Academy of Sciences*, vol. 106, pp. 6459-6464, 2009.
- [176] A. Kuo, N. V. Blough, and P. V. Dunlap, "Multiple N-acyl-L-homoserine lactone autoinducers of luminescence in the marine symbiotic bacterium *Vibrio fischeri*," *Journal of bacteriology*, vol. 176, pp. 7558-7565, 1994.
- [177] S. V. Studer, M. J. Mandel, and E. G. Ruby, "AinS quorum sensing regulates the *Vibrio fischeri* acetate switch," *Journal of bacteriology*, vol. 190, pp. 5915-5923, 2008.
- [178] C. Lupp and E. G. Ruby, "*Vibrio fischeri* uses two quorum-sensing systems for the regulation of early and late colonization factors," *Journal of bacteriology*, vol. 187, pp. 3620-3629, 2005.
- [179] E. Andrianantoandro, S. S. Basu, D. K. Karig, and R. Weiss, "Synthetic biology: new engineering rules for an emerging discipline," *Molecular systems biology*, vol. 2, 2006.
- [180] A. Einolghozati, M. Sardari, and F. Fekri, "Design and Analysis of Wireless Communication Systems Using Diffusion-Based Molecular Communication Among Bacteria," *Wireless Communications, IEEE Transactions on*, vol. 12, pp. 6096-6105, 2013.
- [181] T. Danino, O. Mondragon-Palomino, L. Tsimring, and J. Hasty, "A synchronized quorum of genetic clocks," *Nature*, vol. 463, pp. 326-330, 2010.
- [182] M. A. Palacios, E. Benito-Peña, M. Manesse, A. D. Mazzeo, C. N. LaFratta, G. M. Whitesides, *et al.*, "InfoBiology by printed arrays of microorganism colonies for timed and on-demand release of messages," *Proceedings of the National Academy of Sciences*, vol. 108, pp. 16510-16514, 2011.
- [183] F. Cousteau, *Ocean: the world's last wilderness revealed*: Dorling Kindersley Ltd, 2011.
- [184] J. G. Cao and E. A. Meighen, "Purification and structural identification of an autoinducer for the luminescence system of *Vibrio harveyi*," *Journal of Biological Chemistry*, vol. 264, pp. 21670-21676, 1989.
- [185] A. S. K. Agrawal, *Fluid Mechanics and Machinery*: Tata McGraw-Hill Education, 2001.
- [186] J. Muller, C. Kuttler, and B. A. Hense, "Sensitivity of the quorum sensing system is achieved by low pass filtering," *Biosystems*, vol. 92, pp. 76-81, 2008.
- [187] A. Fick, "V. On liquid diffusion," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 10, pp. 30-39, 1855.
- [188] M. S. Leeson, "Performance analysis of direct detection spectrally sliced receivers using Fabry-Perot filters," *Lightwave Technology, Journal of*, vol. 18, pp. 13-25, 2000.
- [189] L. Koschel, A. Kortke, and F. Undi, "Multi-gigabit communication with hybrid pulse shaping and fractional oversampling," in *Personal Indoor and*

- Mobile Radio Communications (PIMRC), 2012 IEEE 23rd International Symposium on*, 2012, pp. 2437-2442.
- [190] P. S. Stewart, "Diffusion in Biofilms," 2003-03-01 2003.
- [191] P. S. Stewart, "Diffusion in Biofilms," *Journal of Bacteriology*, vol. 185, pp. 1485-1491, 2003.
- [192] H. B. Yilmaz and C. B. Chae, "Arrival modelling for molecular communication via diffusion," *Electronics Letters*, vol. 50, pp. 1667-1669, 2014.
- [193] C. S. R. Prabhu and A. P. Reddi, *Bluetooth Technology: And Its Applications With Java And J2Me*: PHI Learning Pvt. Ltd., 2004.
- [194] A. Ghosh, J. W. Lee, and H. S. Cho, "Throughput and energy efficiency of a cooperative hybrid ARQ protocol for underwater acoustic sensor networks," *Sensors*, vol. 13, pp. 15385-15408, 2013.
- [195] P. J. Shih, C. H. Lee, P. C. Yeh, and K. C. Chen, "Channel Codes for Reliability Enhancement in Molecular Communication," *IEEE Journal on Selected Areas in Communications*, , vol. 31, pp. 857-867, 2013.
- [196] Y. Lu, M. D. Higgins, and M. S. Leeson, "Comparison of Channel Coding Schemes for Molecular Communications Systems," *IEEE Transactions on Communications*, , vol. 63, pp. 3991-4001, 2015.
- [197] Y. Lu, M. D. Higgins, and M. S. Leeson, "Self-orthogonal convolutional codes (SOCCs) for diffusion-based molecular communication systems," in *2015 IEEE International Conference on Communications (ICC)*,, 2015, pp. 1049-1053.
- [198] M. S. Leeson and M. D. Higgins, "Error correction coding for molecular communications," in *2012 IEEE International Conference on Communications (ICC)*, 2012, pp. 6172-6176.
- [199] X. Wang, J. Song, J. Liu, and Z. L. Wang, "Direct-current nanogenerator driven by ultrasonic waves," *Science*, vol. 316, pp. 102-105, 2007.
- [200] Y. Lu, M. D. Higgins, and M. S. Leeson, "Diffusion based molecular communications system enhancement using high order hamming codes," in *Communication Systems, Networks & Digital Signal Processing (CSNDSP), 2014 9th International Symposium on*, 2014, pp. 438-442.
- [201] M. T. Barros, S. Balasubramaniam, and B. Jennings, "Error control for calcium signaling based molecular communication," in *2013 Asilomar Conference on Signals, Systems and Computers*,, 2013, pp. 1056-1060.
- [202] T. K. Moon, *Error correction coding: mathematical methods and algorithms*.: Wiley-Blackwell, 2005.
- [203] J. Park, S. Mackay, and E. Wright, "8 - Introduction to protocols," in *Practical Data Communications for Instrumentation and Control*, J. P. M. Wright, Ed., ed Oxford: Newnes, 2003, pp. 186-198.
- [204] R. H. Morelos-Zaragoza, *The art of error correcting coding*: John Wiley & Sons, 2006.
- [205] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell System Technical Journal*, vol. 38, pp. 611-656, 1959.
- [206] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Transactions on Communications*, , vol. 44, pp. 1261-1271, 1996.
- [207] C. Heegard and S. B. Wicker, *Turbo coding* vol. 476: Springer Science & Business Media, 2013.

- [208] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "A soft-input soft-output maximum a posteriori (MAP) module to decode parallel and serial concatenated codes," *TDA progress report*, vol. 42, pp. 1-20, 1996.
- [209] S. Bernard, *Digital communications: fundamentals and applications*, 2001.
- [210] V. Castano and I. Schagaev, *Resilient computer system design*: Springer, 2015.
- [211] M. S. Leeson and M. D. Higgins, "Forward error correction for molecular communications," *Nano Communication Networks*, vol. 3, pp. 161-167, 2012.
- [212] M. Kocaoglu and O. B. Akan, "Minimum energy channel codes for nanoscale wireless communications," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 1492-1500, 2013.
- [213] C. Erin and H. H. Asada, "Energy optimal codes for wireless communications," in *Decision and Control, 1999. Proceedings of the 38th IEEE Conference on*, 1999, pp. 4446-4453.
- [214] C. Bai, M. S. Leeson, and M. D. Higgins, "Minimum energy channel codes for molecular communications," *Electronics Letters*, vol. 50, pp. 1669-1671, 2014.
- [215] D. J. C. MacKay, "Fountain codes," *IEE Proceedings-Communications*, vol. 152, pp. 1062-1068, 2005.
- [216] M. Luby, "LT codes," in *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*, 2002, pp. 271-280.
- [217] S. Abadal and I. F. Akyildiz, "Bio-Inspired Synchronization for Nanocommunication Networks," in *Global Telecommunications Conference 2011, IEEE*, 2011, pp. 1-5.
- [218] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, pp. 498-519, 2001.
- [219] J. C. Martinez, J. Flich, A. Robles, P. López, J. Duato, and M. Koibuchi, "In-order packet delivery in interconnection networks using adaptive routing," in *Parallel and Distributed Processing Symposium, 2005. Proceedings. 19th IEEE International*, 2005, pp. 101-101.
- [220] D. E. McDysan and D. L. Spohn, *ATM theory and application*, 1st ed.: McGraw-Hill Professional, 1998.
- [221] M. D. Yarvis, P. Reiher, and G. J. Popek, *Conductor: Distributed Adaptation for Heterogeneous Networks: Distributed Adaptation for Heterogeneous Networks*: Springer Science & Business Media, 2002.
- [222] G. Serban, C. Anton, L. Ionescu, I. Tutanescu, and A. Mazare., "Implementation of a 64-bit hybrid SR-ARQ algorithm on FPGA," in *2011 International Conference on Applied Electronics (AE), IEEE*, 2011.
- [223] S. Chang, "Theory of information feedback systems," *Information Theory, IRE Transactions on*, vol. 2, pp. 29-40, 1956.
- [224] J. Li and Y. Q. Zhao, "Resequencing Analysis of Stop-and-Wait ARQ for Parallel Multichannel Communications," *Networking, IEEE/ACM Transactions on*, vol. 17, pp. 817-830, 2009.
- [225] I. A. Bagad and V. S. Dhotre, *Data Communication*: Technical Publications, 2007.
- [226] F. Walsh, S. Balasubramaniam, D. Botvich, T. Suda, T. Nakano, S. Bush, *et al.*, "Hybrid DNA and Enzyme Based Computing for Address Encoding, Link Switching and Error Correction in Molecular Communication," in

- Nano-Net*. vol. 3, M. Cheng, Ed., ed: Springer Berlin Heidelberg, 2009, pp. 28-38.
- [227] S. Banzal, *Data and Computer Network Communication*: Laxmi Publications Pvt Limited, 2007.
- [228] N. K. Vijayakumaran and C. S. S. Vindo, *Informatics*: PHI Learning Pvt. Ltd., 2014.
- [229] "IEEE Draft Recommended Practice for Nanoscale and Molecular Communication Framework," *IEEE P1906.1/D2.0*, October 2014, pp. 1-63, 2015.
- [230] T. Nakano, Y. Okaie, and A. V. Vasilakos, "Transmission Rate Control for Molecular Communication among Biological Nanomachines," *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 835-846, 2013.
- [231] L. Felicetti, M. Femminella, G. Reali, T. Nakano, and A. V. Vasilakos, "TCP-like molecular communications," *IEEE Journal on Selected Areas in Communications*, vol. 32, pp. 2354-2367, 2014.
- [232] X. Wang, M. D. Higgins, and M. S. Leeson, "Simulating the performance of SW-ARQ schemes within molecular communications," *Simulation Modelling Practice and Theory*, vol. 42, pp. 178-188, 2014.
- [233] C. Bai, M. S. Leeson, and Matthew D. Higgins, "Performance of SW-ARQ in bacterial quorum communications," vol. 6, pp. 3-14, March 2015 2015.
- [234] Z. P. Li, J. Zhang, and T. C. Zhang, "Concentration Aware Routing Protocol in Molecular Communication Nanonetworks," *Applied Mechanics and Materials*, vol. 556, pp. 5024-5027, 2014.
- [235] T. Ganesan and N. Rajkumar, "Nanotechnology Integrated Routing Protocol Design for MANET Based on collision-Molecular communication," *International Journal on Information Technology and Computer Science (IJITCS)*, vol. 9, 2013.
- [236] M. T. Barros, S. Balasubramaniam, B. Jennings, and Y. Koucheryavy, "Adaptive transmission protocol for molecular communications in cellular tissues," in *Communications (ICC), 2014 IEEE International Conference on*, Sydney, NSW, 2014, pp. 3981-3986.
- [237] M. J. Moore and T. Nakano, "Addressing by beacon distances using molecular communication," *Nano Communication Networks*, vol. 2, pp. 161-173, 2011.
- [238] S. Hiyama, Y. Isogawa, T. Suda, Y. Moritani, and K. Sutoh, "A design of an autonomous molecule loading/transporting/unloading system using DNA hybridization and biomolecular linear motors," in *Proc. European Nano Systems*, Grenoble, France, 2005.
- [239] E. Ustunel, I. Hokelek, and O. Ileri, "A cross-layer goodput enhancement considering CRC coding and ARQ dynamics," in *Computers and Communications (ISCC), 2012 IEEE Symposium on*, 2012, pp. 000023-000028.
- [240] A. S. Tanenbaum, *Computer networks*, 5th ed. Boston, Mass. ; London: Pearson, 2011.
- [241] W. Guopeng, P. Bogdan, and R. Marculescu, "Efficient Modeling and Simulation of Bacteria-Based Nanonetworks with BNSim," *Selected Areas in Communications, IEEE Journal on*, vol. 31, pp. 868-878, 2013.
- [242] T. Nakano, M. J. Moore, W. Fang, A. V. Vasilakos, and S. Jianwei, "Molecular Communication and Networking: Opportunities and Challenges," *NanoBioscience, IEEE Transactions on*, vol. 11, pp. 135-148, 2012.

- [243] Chenyao Bai, Mark S. Leeson, and Matthew D. Higgins, "Performance of SW-ARQ in bacterial quorum communications," vol. 6, pp. 3–14, March 2015 2015.
- [244] H. Jianhua, K.R. Subramanian, D. Donghua, and W. Wei, "Novel methods for the performance analysis of adaptive hybrid selective repeat ARQ," *Computer Communications*, vol. 23, pp. 1548–1557, 2000.
- [245] V. Srikanth, "Energy Efficient, Scalable and Reliable MAC Protocol for Electromagnetic Communication among Nano Devices," *International Journal of Distributed and Parallel systems*, vol. 3, pp. 249-256, 2012.
- [246] T. S. Baicheva, "Determination of the Best CRC Codes with up to 10-Bit Redundancy," *Communications, IEEE Transactions on*, vol. 56, pp. 1214-1220, 2008.
- [247] M. Hassner, R. Kotter, and T. Tamura, "Method and means for computationally efficient error and erasure correction in linear cyclic codes," *US Patent No. 5,942,005*, 24 Aug. 1999.
- [248] C. H. J. Wu and J. D. Irwin, *Introduction to computer networks and cybersecurity*: CRC Press, 2013.
- [249] H. Sahoo, "Fluorescent labeling techniques in biomolecules: a flashback," *RSC Advances*, vol. 2, pp. 7017-7029, 2012.
- [250] Todd K. Moon *Error correction coding: mathematical methods and algorithms.*: ISBN 0-471-64800-0, 2005.
- [251] I. Widjaja and A. Leon-Garcia, *Communication Networks Fundamental Concepts and Key Architectures*: Mc Graw Hill, 2004.
- [252] A. Tamsir, J. J. Tabor, and C. A. Voigt, "Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'," *Nature*, vol. 469, pp. 212-215, 2011.
- [253] I. F. A. Y. Sankarasubramaniam, 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.
- [254] A. Ribarits, W. Stepanek, M. Wögerbauer, V. Peterseil, M. Kuffner, C. Topitschnig, *et al.*, "Synthetic Biology," Federal Ministry of Health, Vienna2014.
- [255] P. Fu and S. Panke, *Systems biology and synthetic biology*: John Wiley & Sons, 2009.
- [256] D. E. Cameron, C. J. Bashor, and J. J. Collins, "A brief history of synthetic biology," *Nature Reviews Microbiology*, vol. 12, pp. 381-390, 2014.
- [257] K. M. Esvelt and H. H. Wang, "Genome - scale engineering for systems and synthetic biology," *Molecular systems biology*, vol. 9, p. 641, 2013.
- [258] P. E. Purnick and R. Weiss, "The second wave of synthetic biology: from modules to systems," *Nature reviews Molecular cell biology*, vol. 10, pp. 410-422, 2009.
- [259] J. Beal, A. Phillips, D. Densmore, and Y. Cai, "High-level programming languages for biomolecular systems," in *Design and analysis of biomolecular circuits*, ed: Springer, 2011, pp. 225-252.
- [260] N. M .Kumar and N. B. Gilula, "The gap junction communication channel," *Cell*, vol. 84, pp. 381-388, 1996.
- [261] M. S. Kuran, T. Tugcu, and B. Edis, "Calcium signaling: Overview and research directions of a molecular communication paradigm," *Wireless Communications, IEEE*, vol. 19, pp. 20-27, 2012.

- [262] R. J. Lipton and E. B. Baum, *DNA based computers* vol. 27: American Mathematical Soc., 1995.
- [263] A. Sarrion-Perdigones, E. E. Falconi, S. I. Zandalinas, P. Juárez, A. Fernández-del-Carmen, A. Granell, *et al.*, "GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules," *PloS one*, vol. 6, p. e21622, 2011.
- [264] Baojun Wang, "Design and functional assembly of synthetic biological parts and devices," Doctor of Philosophy, Department of Bioengineering, Imperial College London, 2010.
- [265] M. G. Montague, C. Lartigue, and S. Vashee, "Synthetic genomics: potential and limitations," *Current opinion in biotechnology*, vol. 23, pp. 659-665, 2012.
- [266] D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, and H. O. Smith, "Enzymatic assembly of DNA molecules up to several hundred kilobases," *Nature methods*, vol. 6, pp. 343-345, 2009.
- [267] J. Anderson, J. E. Dueber, M. Leguia, G. C. Wu, J. A. Goler, A. P. Arkin, *et al.*, "BglBricks: A flexible standard for biological part assembly," *Journal of biological engineering*, vol. 4, pp. 1-12, 2010.
- [268] C. D. Smolke, "Building outside of the box: iGEM and the BioBricks Foundation," *Nature biotechnology*, vol. 27, pp. 1099-1102, 2009.
- [269] M. Rimmer and A. McLennan, *Intellectual property and emerging technologies: the new biology*: Edward Elgar Publishing, 2012.
- [270] S. Ayukawa, A. Kobayashi, Y. Nakashima, H. Takagi, S. Hamada, M. Uchiyama, *et al.*, "Construction of a genetic AND gate under a new standard for assembly of genetic parts," *BMC genomics*, vol. 11, p. S16, 2010.
- [271] R. Kitney and P. Freemont, "Synthetic biology—the state of play," *FEBS letters*, vol. 586, pp. 2029-2036, 2012.
- [272] M. Galdzicki, C. Rodriguez, D. Chandran, H. M. Sauro, and J. H. Gennari, "Standard biological parts knowledgebase," 2011.
- [273] D. A. Drubin, J. C. Way, and P. A. Silver, "Designing biological systems," *Genes & development*, vol. 21, pp. 242-254, 2007.
- [274] A. P. Arkin and D. A. Fletcher, "Fast, cheap and somewhat in control," *Genome Biol*, vol. 7, p. 114, 2006.
- [275] J. Blazeck, R. Garg, B. Reed, and H. S. Alper, "Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters," *Biotechnology and bioengineering*, vol. 109, pp. 2884-2895, 2012.
- [276] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in *Escherichia coli*," *Nature*, vol. 403, pp. 339-342, 2000.
- [277] R. Weiss, S. Basu, S. Hooshangi, A. Kalmbach, D. Karig, R. Mehreja, *et al.*, "Genetic circuit building blocks for cellular computation, communications, and signal processing," *Natural Computing*, vol. 2, pp. 47-84, 2003.
- [278] J. E. Dueber, E. A. Mirsky, and W. A. Lim, "Engineering synthetic signaling proteins with ultrasensitive input/output control," *Nature biotechnology*, vol. 25, pp. 660-662, 2007.
- [279] J. C. Anderson, C. A. Voigt, and A. P. Arkin, "Environmental signal integration by a modular AND gate," *Molecular systems biology*, vol. 3, p. 133, 2007.

- [280] S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, and R. Weiss, "A synthetic multicellular system for programmed pattern formation," *Nature*, vol. 434, pp. 1130-1134, 2005.
- [281] S. Basu, R. Mehreja, S. Thiberge, M. T. Chen, and R. Weiss, "Spatiotemporal control of gene expression with pulse-generating networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, pp. 6355-6360, 2004.
- [282] L. You, R. S. Cox, R. Weiss, and F. H. Arnold, "Programmed population control by cell-cell communication and regulated killing," *Nature*, vol. 428, pp. 868-871, 2004.
- [283] C. Krueger, C. Danke, K. Pfleiderer, W. Schuh, H. M. Jäck, S. Lochner, *et al.*, "A gene regulation system with four distinct expression levels," *The journal of gene medicine*, vol. 8, pp. 1037-1047, 2006.
- [284] F. J. Isaacs, D. J. Dwyer, and J. J. Collins, "RNA synthetic biology," *Nature biotechnology*, vol. 24, pp. 545-554, 2006.
- [285] J. A. J. Arpino, E. J. Hancock, J. Anderson, M. Barahona, G. B. V. Stan, A. Papachristodoulou, *et al.*, "Tuning the dials of synthetic biology," *Microbiology*, vol. 159, pp. 1236-1253, 2013.
- [286] A. Wittmann and B. Suess, "Engineered riboswitches: expanding researchers' toolbox with synthetic RNA regulators," *FEBS letters*, vol. 586, pp. 2076-2083, 2012.
- [287] R. Sarpeshkar, "Analog synthetic biology," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 372, p. 20130110, 2014.
- [288] J. M. Berg, J. L. Tymoczko, G. J. Gatto, and L. Stryer, *Biochemistry*. New York, NY: W.H. Freeman and Company, 2015.
- [289] C.C.Guet, M. B. Elowitz, W. Hsing, and S. Leibler, "Combinatorial synthesis of genetic networks," *Science*, vol. 296, pp. 1466-1470, 2002.
- [290] C. T. Bhunia, *Information Technology Network and Internet*: New Age International, 2006.
- [291] S. P. Pramod, A. Rajagopal, and S. K. Akshay, "FPGA Implementation of Single Bit Error Correction using CRC," *International Journal of Computer Applications*, vol. 52, pp. 15-19, 2012.
- [292] D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and D. L. Wheeler, "GenBank," *Nucleic Acids Research*, vol. 33, pp. D33-D38, 2005.
- [293] P. Hillenbrand, G. Fritz, and U. Gerland, "Biological signal processing with a genetic toggle switch," *PloS one*, vol. 8, p. e68345, 2013.