博 士 論 文

Study on Diffusion-Based Molecular Communication

（拡散型分子通信に関する研究）

by

Yao Sun

孫　堯

A dissertation submitted in partial fulfillment

for the Doctor of Philosophy degree

in the

Department of Information and Communication Engineering

Graduate School of Information Science and Technology

The University of Tokyo

Supervisor Prof. Kaoru Sezaki

指導教員　瀬崎　薫　教授

December 2018

東京大学
THE UNIVERSITY OF TOKYO

一歩ずつ前へ！

December 2018

Abstract

Molecular communication (MC) is a recently proposed paradigm which aims at enabling the communication among nano-machines. Dissimilar to the conventional communication techniques using electromagnetic wave as the information carrier, the molecular communication applies molecules to modulate signals to solve the diffraction problem of currently in use electromagnetic waves at nano-scale. In MC, the physical or chemical characters of molecules can be used to modulate signals. The modulated signals propagate through the medium via the physical displacement of the information molecules. Therefore, the channel has different features from the conventional communication techniques. Among all schemes that implement molecular communication, the diffusion-based channel is considered as the most promising way because of its simplicity and low extra energy cost. This dissertation focuses on the diffusion-based molecular communication (DBMC) and is composed of three components addressing three prominent problems of diffusion-based molecular communication: (1) communication distance measurement; (2) inter-symbol interference; (3) low channel capacity.

Communication distance measurement is always a crucial problem in wireless communications. It also plays an essential role in DBMC. A distance measurement approach based on the arrival time difference (ATD) of two molecular signals is

proposed. The ATD approach utilizes two types of molecules with different diffusion coefficients. Thus, when the node to be measured transmits the signals with the molecules simultaneously, there is an arrival time difference. Applying the arrival time difference, the distance between the two nodes can be estimated based on the channel model. The ATD approach avoids the complicated synchronization of the nodes, and at the same time overcomes the error accumulation and inefficiency of using a round trip signal. Besides, the ATD is a time-based approach so that it barely relies on accurate measurement of signal amplitude, namely the concentration in diffusion-based molecular communication.

Inter-symbol interference (ISI) is one of the most crucial problems of diffusion-based molecular communication. The ISI is the interference of the information molecules remaining in the transmission medium of the former bits on the later ones. It is an unavoidable phenomenon if diffusion-based molecular communication because of the diffusion features. It has a severe impact on the bit error rate, especially for a system consists of mobile nodes. We propose two protocols to mitigate the influence of ISI. The first one is a retransmission protocol. When the receiver sensed the signal quality becomes weaker due to the node movement, it transmits a simple acknowledgment to the transmitter. Based on the acknowledgment, the transmitter calculates the current channel condition, estimates the bits transmitted during the weak connection period and retransmits them. Another one is an adaptive code width (ACW) protocol. This protocol is based on the fact that the communication distance influences the ISI. In the ACW protocol, after the connection is established, the receiver transmits distance feedback to the transmitter periodically. The transmitter estimates the communication distance and adjusts

the code width accordingly. At the same time, the receiver knows the code width of the next bits using the estimated communication distance from the currently received signal. Consequently, the communication quality is guaranteed. Moreover, the ACW can reduce the redundant code width when the nodes are close to each other, to increase the transmission rate.

Due to the uncertainty and the long-delay character of the diffusion process, DBMC is considered with an extremely low channel capacity compared with the conventional communication techniques. We propose a novel MC utilizing DNA as the information packets. A simple modulation scheme is proposed in which the two types of nucleobase pairs are used as different bits. Therefore, a series of binary bits can be encoded onto a single DNA molecule. Additionally, the diffusion coefficient of DNA molecule is negatively correlated to the base pair number. By applying an observing window and focus slot scheme, we design a diffusive DNA based molecular communication model. Furthermore, the channel capacity of the proposed system is also analyzed. However, the diffusion leads to the disordered arrival of the DNA packets. To this problem, we propose a method using multiple duplicated packets and a buffer in the receiver to compare the number of received packet DNA molecules. The proposed method does not only reduce the probability of disordered arrival, but also increases the effective transmission rate of the diffusive DNA based MC.

Overall, this dissertation makes contributions to the theoretical research of DBMC. Three crucial problems are addressed. In particular, we propose an ATD approach to measure the communication distance to increase the accuracy and efficiency. Two protocols are proposed to mitigate the ISI in DBMC. DNA is adopted to design a novel DBMC as packets to increase the possible transmission rate.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Kaoru Sezaki for his excellent guidance, insightful suggestions on my research and great support on my study life in Japan. He has taught me to be open-minded and be ready to discover problems in a new filed; to be positive and optimistic facing with difficulties. With his guidance, I have learned how to become a researcher who can study individually. I could never have completed this work without his kind encouragement and support.

I want to thank all the defense committee members including Prof. Kanta Matsuura, Prof. Yoshihiro Kawahara, Prof. Hideya Ochiai, Prof. Kazuyuki Aihara, Prof. Shinya Sugiura for their dedication of time to discuss my research and precious advice.

I also would like to thank Prof. Tohru Asami, Prof. Yoichi Sato. I appreciate their valuable time having meetings with me as advisors. They have given many many helpful suggestions for my research from a different point of view.

I would like to thank Prof. Masaki Ito, Prof. Hill Hiroki Kobayashi, Yuya Matsuno, and other lab members including Shunsuke Aoki, Hongyang Chen, Hao Niu, Dunstan Matekenya who have also encouraged and helped me with my research. Thanks very much to Ms. Kaho Matsumoto and Ms. June Naito, who helped me

deal with a lot of paperwork.

I also appreciate the Japanese government for the MEXT scholarship that supported me for the years. Many thanks to the staffs of Komaba Lodge, Tokyo International Exchange Center, Kouraku-ryou, Ota-kinenkan for their help with my daily life in Japan. Their kindness made me feel like not being away from home.

Finally, I would like to convey special thanks to my family. My parents for their unconditional love, understanding, and perpetual support. My wife Yue Jiang, who suffered from long distance relationship with me for five years, and finally married me during my Ph.D. Her patient love encouraged me to overcome all the difficulties through all the years and finally complete this work.

December 2018

Contents

List of Figures

List of Tables

Chapter 1

Background

Nano-technology has been rapidly developing and inspiring numerous novel applications in the last two decades. Among all, information communication technology (ICT) as a discipline using devices processing data, seems not one of the fields that can be influenced by the nano-technology. However, the fact is, as the basic installations such as the computers that we are currently using have started reaching their limits of getting faster and lower energy consumption, progressively more demands for the miniaturization of the components that constitute the current ICT devices are drawing the significant attention of both academic circles and industrial community.

1.1 Challenges for Communication at Nano-Scale

It is believed that the Nobel laureate physicist Richard Feynman first pointed out the concept of nanotechnology in his famous speech entitled "There's Plenty of Room at the Bottom: An Invitation to Enter a New Field of Physics" [1] in December 1959. In this speech, he depicted the future that he believed humans

could make *nanoscale machines* which are sufficiently tiny and capable so that they can arrange the atoms the way we want, and even do chemical synthesis by mechanical manipulation. However, at first, limited activities exploring nano-technology got started, until the early 21st Century the advancements begun to accelerate significantly in this field. Some of the depicted application scenarios, such as the synthesis with a programmable nanoscale machine [2], have become true recently.

The nanoscale machines, also known as nano-machines or nano-robotics, are defined as a kind of artificial devices consist of nanoscale or molecular components. Due to the tiny components, a nano-machine can be sized ranging from 10 nanometers -10 micrometers [3]. Even though constrained by the size, the nano-machines are expected to have capabilities include but not limited to fulfilling the intended tasks by following a particular set of instructions, self-assembly, self-replication, locomotion, and communication. For these expected features, several methods are investigated to develop and manufacture nano-machines, including the top-down approach [4], bottom-up approach [5], and bio-hybrid approach[6]. Whichever approach to choose, the nano-machines, towards the expected features, should be consist of components such as processing and control unit, power supply unit, actuators, sensors, and transceiver [7]. The potential applications of nano-machines are limitless, ranging from industrial to environmental. The most promising application scenario lays in the biomedical field. For instance, nano-machines can be utilized for immune system support [8], bio-hybrid implants [9], targeted drug delivery systems [10], cellular level surgery [11], and so forth.

Because of the constrained size, a single nano-machine usually has limited re-

sources to accomplish complex tasks. Therefore, instead of a single nano-machine, a swarm of them cooperating for the same task is a more reasonable and practical way to utilize the nano-machines [12]. To this end, the communication among the nano-machines is necessary. Nevertheless, enabling the communicating among nano-machines faces a series of problems. First of all, the conventional communication techniques are not suitable for the communication at the nanoscale because of the size and power consumption constraints. Secondly, and more importantly, the currently in use wireless communication techniques count on the electromagnetic wave whose spectrum ended at the highest frequency of 300 GHz, i.e., the wavelength is as shortest as approximately 1 millimeter, according to the spectrum usage in Japan [13] and the USA [14]. Because the wavelength is longer than the size of nano-machines, the diffraction is unavoidable result in the current wireless techniques is not feasible for the communication at the nanoscale. Finally, since the in-body application is among the potential use of the nano-machines, the radio waves can easily attenuate inside the body, meanwhile the electronic components of the transceivers may be not sufficiently bio-friendly and cause health problems in the long term.

To solve the problems mentioned above and enable the nano-machines to communicate, two different approaches are proposed in recent years. The first one utilizes novel nano-materials, such as carbon nanotube [15] and graphene [16]. Thanks to the properties of the nano-materials, nano-scaled logical circuitry and antenna can be used for signal modulation based on amplitude or frequency. The potential frequency can even reach the terahertz region whose frequency is at the terahertz gap between radio waves and infrared light [17]. The second solution

makes use of molecules as the information carrier to enable the communication among nano-machines. Thus, this kind of approach is called *molecular communication (MC)*. Considering the relations with other research area, the MC locates at the place as shown in Fig. 1.1, is the interdisciplinary of the information theory, the biology, and the Nano-technology. In light of the signal propagation means, there are several ways to implement communication through molecules, including via diffusion, gap-junction, the molecular motor, bacterial motor among others. We focus more on the diffusion-based molecular communication (DBMC). The more detailed introduction of the MC and the reason why we choose the DBMC as our research focus will be given in the next chapter. Generally speaking, both of the paradigms mentioned above, namely terahertz band communication and the MC, are aiming at the communication at the nano-scale or using nano-scale materials, and thus they are also referred as *nano-networks* at the same time.

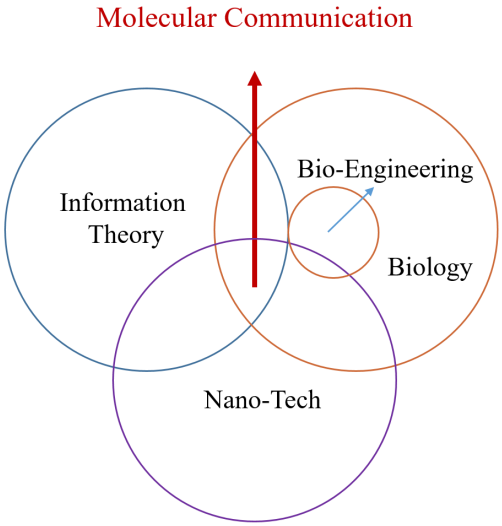

Figure 1.1: The research area of the molecular communication

1.2 Motivation and Contributions

The MC is capable to enable the communication among nano-machines. Since this paradigm was proposed a decade ago, many researchers including us have made efforts to the theoretical investigations, especially of the the DBMC. However, it is not easy to implement the MC to practical applications because there are several crucial problems at this moment.

First of all, the DBMC relies on the diffusion to propagate signals. And the diffusion is a stochastic process so that there are massive noises in the channel leading to the result that the information molecules arrive at the receiver occasionally. In addition, the diffusion process is relatively slow compared with the electromagnetic waves. As a result, there are consistently remaining molecules of former bits in the medium. Those remaining molecules impact the later bits and even raises the bit error rate. Therefore how to mitigate the influence of inter-symbol interference on the bit error rate becomes a key problem on the way to make the DBMC more practical.

Secondly, from simulation experiments, we know that the inter-symbol interference is directly correlated with the communication distance and the bit width. Therefore efficient and accurate distance measurement plays a key role in the mitigation of inter-symbol interference. Moreover, similar to the conventional wireless communication techniques, the communication distance measurement is also important for various application scenarios, for instance even deployment of the nano-machines for monitoring health condition and improving the connection reliability with the communication distance information.

Finally until this research, the studies on the DBMC mainly focus on small

molecules or ions that only transmit a bit with one or more particles at once. However, on the other hand, DNA has the capability to store massive information onto a single molecule. Meanwhile, as more in-depth research of DNA has been carried, more and more applications using DNA technology, such as genetic engineering, molecular computing, and drug discovery based on the DNA-encoded chemical library, have been developed. Combining the DBMC and nano-machines, we believe that the DNA-based application can make more contributions to the academic circles and industrial community.

The promising prospects of the DBMC and the problems mentioned above motivate us to devote ourselves to this study. The main contributions of this dissertation can be summarized as follows.

- *Efficient distance measurement.* We propose a distance measurement protocol based on arrival time difference. The protocol uses two types of information molecules with different diffusivities. Thus when the information molecules are released at the same time, there will be an arrival time difference of the two signals. It solves the efficient problem of round-trip signal based protocol. Meanwhile, the error accumulated in the round trip can be reduced.

- *Mitigation of bit error rate.* Aiming at the inter-symbol interference problem of concentration shift keying, we present two methods to reduce the error bits. The first one is a passive way. It uses an acknowledgement information from the receiver to the transmitter. The transmitter knows the connection quality is getting poorer from the acknowledgement, and then retransmits the data bits were not received clearly. The second method solves the problem in

an active way. It introduces a real-time distance feedback from the receiver to the transmitter. The transmitter can adjust the bit width corresponding to the communication distance to guarantee the connection quality during the whole transmission process.

- *Scope-expansion with DNA molecules.* We design a novel DBMC with the DNA molecules. Thanks to the massive information can be stored on the DNA molecules, the effective transmission rate of the DBMC can be significantly increased. In addition, this novel paradigm makes the DBMC has a more close relation with other molecule-based IT applications such as molecular computing [18] and integrated gene circuits [19].

1.3 Dissertation Outline

The remainder of this dissertation is organized as follows.

In Chapter 2, we present some preliminary knowledge of the MC. First of all, how the MC is implemented is introduced. In particular, the encoding/decoding schemes and signal propagation means using molecules instead of the well-known electromagnetic waves are introduced with examples. Secondly, the DBMC, which is our research focus is modeled in detail. From the micro- and macro-scope point of view, the mathematical models of diffusion channel are derived focusing on one single molecule and the collective behavior as a swarm of molecules.

Chapter 3 addresses the communication distance measurement issue of the DBMC. Two related works, a signal attenuation protocol and a round-trip time protocol, are first introduced. Then aiming at improving the efficiency and estimation accuracy, we propose a method using arrival time difference. In the proposed

method, signals base on two types of molecules with different diffusion coefficient are used to generate an arrival time difference. First of all it is a time-based approach so that the requirement of accurate concentration measurement is avoidable. In addition applying this method, only one way signal rather than the round trip signal is used to estimate the communication distance. Consequently, the efficiency is improved and the error can be accumulated in the round trip is avoided. Finally, the simulation results validate the effectiveness and accuracy of the proposed method. Moreover, due to two types of molecules are used, the performance is optimizable by rationally choose the value of the diffusion coefficients.

In Chapter 4, we study the reduction of bit error rate caused by the inter-symbol interference, which is one of the most crucial problems of the DBMC. Based on the results from analysis and simulation experiments that the inter-symbol interference has a strong correlation with the communication distance and the bit width/code width, two protocols are proposed. The first one applies a retransmission scheme which retransmits the potential error bits to reduce the influence of error bits caused by the inter-symbol interference. The transmitter calculates the potential error bits with the channel situation estimated from the acknowledgement signal. The second approach adopts a distance feedback based on which the suitable bit width is calculated and utilized to guarantee the reliable connection. This approach also reduces the redundant bit width when the connection quality is acceptable in order to secure the transmission efficiency. Both of the proposed methods are validated through simulation experiments.

In Chapter 5, we propose a novel DBMC by applying DNA molecules. which are capable to store massive information on a single molecule are used as data

packets, in order to increase the effective transmission rate of the DBMC. To this end, we apply a simple encoding/decoding scheme relies on the base pair types. We also analyze the features of DNA diffusion that is used in our proposed system. Lastly, the channel capacity is discussed with simulation results.

Finally, Chapter 6 makes conclusions of this dissertation, and discusses the potential future research directions.

Chapter 2

Overview of Molecular Communication

The molecular communication (MC) and the terahertz radio communication are the currently possible solutions to enable the communication of the nano-machines. The MC has advantages such as that it is more suitable for the bio-related applications of nano-machines over the terahertz communication. In this chapter, we introduce some preliminary knowledge of the MC. In particular, we start from the introduction of how the MC works without the conventional electromagnetic waves. The feasible ways to modulate signals onto information molecules and the signal propagation schemes of MC are presented. The mathematical models of our research focus, i.e., the diffusion-based molecular communication (DBMC) are given at the end of this chapter.

2.1 Flow of Molecular Communication

The purpose of a general communication system is to transmit some pieces of information from one geographical point to another through time via a succession process. Generally speaking, there are three fundamental elements in a simple

point-to-point communication system, including a transmitter, the channel, and a receiver. For an easier, more efficient, more secure transmission, the transmitter and the receiver can be divided into more specific modules [20]. Fig. 2.1 presents the block diagram showing each module included in the transmitter, the channel, and the receiver with the information flow between the modules.

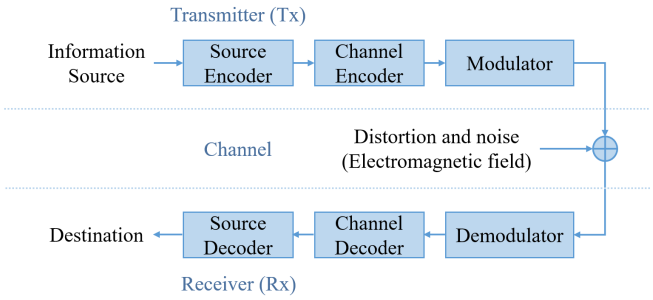

Figure 2.1: Block diagram of digital communication systems

Similarly, in an MC system, the message of interest is sent by the transmitter, then goes through the channel, and finally arrives at the receiver. The fundamental difference between the conventional communication and the MC is the form of signal. The conventional communication uses electrical signals. Correspondingly, the MC is based on information molecules. This difference has a consequence that the components of the transceivers and the noise in the channels are different. Fig. 2.2 shows the process of the information transmission from the information source to the desired destination in a general MC system.

In particular, the information of interest is modulated onto the information molecules by the processing unit, namely the information molecule generator of the transmitter. By this process, a specific amount of the molecules carrying the information are generated. The second step is that the transmitter releases the in-

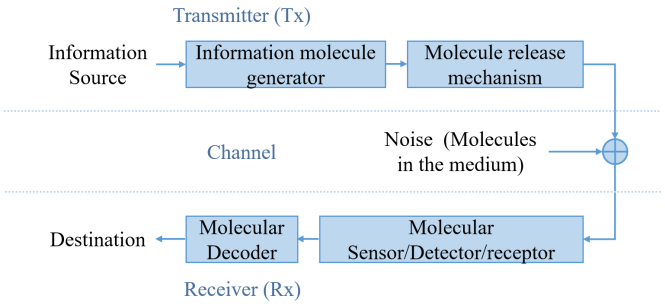

Figure 2.2: Block diagram of molecular communication systems

formation molecules directly into the medium [21] or embeds them onto a higher level carrier such as a bacterium [22] acting as a mobile node so that the information can propagate with the molecules. The information molecules then travel through the medium via a propagation mechanism from the transmitter to the receiver. The cause of the undesired fluctuation on the signals, namely the distortion and noise are different in the traditional communication techniques and the MC. In the conventional communication systems, the noises are caused by the electromagnetic fields in the environment. Meanwhile, the primary noise in the MC is a result of the molecules in the medium [23]. When the information molecules arrive at or near the receiver, they can be captured or sensed by the receptors on the receiver [24]. Then the information encoded onto the molecules can be decoded by the processing unit in the receiver. Correspondingly to the received information, a series of reactions or relaying the message are initiated in the processing unit as well.

2.2 Encoding and Decoding Schemes

Several schemes can be applied for the encoding and decoding process with molecules. Generally, all of the chemical and physical features of molecules can be used to modulate information. For example in nature, from where the idea of molecular communication is originally inspired, the neurons in the human brain use neurotransmitters and Ca^{2+} to transmit signals [25]. Similarly, in the MC, artificial nano-machine systems can use concentrations, types, released time, structure and so forth of the information molecules as the applicable candidates for the modulation of MC.

A. Concentration Shift Keying

Information can be encoded in the number of the molecules released by the transmitter [26]. For most of the cases, the molecules are difficult to count individually. Instead, the molecule concentration, namely the approximate number of the molecules per unit volume, is much easier to distinguish for the sensors of both macro-scale devices and nano-machines. As the concentration is used, this scheme is called concentration shift keying (CSK). Let us consider a binary system with only two states, bit 0 and bit 1. To transmit a binary sequence using CSK-based MC, we can define a particular amount of the molecules released into the medium represents the bit 1 while releasing no molecules represents the bit 0. It is similar to the idea of the difference between 0V and +5V in the integrated circuits. This particular binary scheme is also known as the on-off keying [27]. It is expandable to more than binary states with higher resolution sensors. The basic idea is similar to the amplitude shift keying techniques of conventional communication.

B. Molecule Shift Keying

The second scheme applies different types of molecules to encode different symbols [26], known as molecular shift keying (MoSK). For instance, still considering a binary bit scenario, two types of molecules are necessary for the system. Molecule A and B can be used to represent bit 0 and 1 respectively. This scheme requires some circumstances such as the two types molecules do not react with each other, and the receiver nano-machines need to be installed with different types of receptors. However, it has an advantage over the CSK that is can achieve a higher signal quality by reducing the accumulated interference of the former bits of the same type of information molecules.

C. Timing-Based Modulation

It is also possible to use the time of releasing and receiving the information molecules to distinguish the different patterns of the states [28]. This scheme is known as the timing channel or time-based channel. It borrows the idea of phase shift keying in the conventional radio communication. Let us consider a protocol in which the time slot width is pre-defined as T_s. The binary bit 0 and 1 can be defined as the released time at the beginning of the slot and the middle time in the slot. Consequently, there is a pattern difference in the probability of the time when the information molecules arrive at the receiver. Focusing on the encoding phase, the mechanism can be expressed as the bit in the current slot i is 0, when the time of releasing the information molecules $t_r = iT_s$, otherwise when the releasing time $t_r = (i + 1/2)T_s$, the bit is 1.

D. Structure-Based Modulation

The structure is also a candidate for the encoding and decoding scheme. The different types of molecules used in the MoSK also have different structures. However, in the structure-based modulation scheme, we may consider a more direct way of using DNA as the information molecules. The information, as a form of a bit sequence, can be directly encoded onto the DNA molecules. As well known, DNA is used to store, transcript, and translate genomic information by the Eukaryota in nature. It initially has the feature to store massive of information [29]. The genomic information is encoded onto DNA molecules using only four types of nitrogenous bases with different combination in the structure. Take a simple case as an example of a binary bit storage system, we can artificially define the base pair of adenine and thymine as a bit 1, meanwhile the other kind of base pair of cytosine and guanine as a bit 0 [30]. Thus, a bit sequence can be encoded on to a DNA chain for the data transmission purpose of MC.

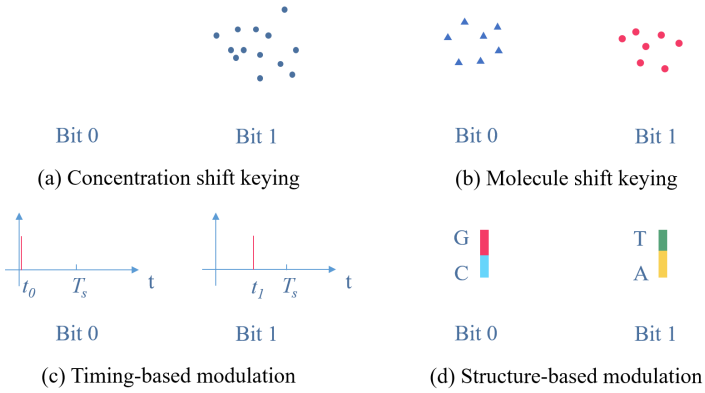

Figure 2.3: Examples of modulation schemes used in molecular communication

The modulation schemes described above are demonstrated respectively in Fig. 2.3. Moreover, they can be implemented not only individually, but also in a multiplexing way. For example, the CSK and MoSK can be applied simultaneously to obtain more efficient transmission with lower error rate [31]. Two types of molecules A_1 and A_2 are used in the odd and even time slots respectively. Also, in each time slot, different concentration of the molecules are used. Thus, this scheme can reduce the interferences between consecutive time slots, and then improve the transmission reliability.

2.3 Signal Propagation

After being encoded onto the molecules, the information needs to propagate with them through space to the desired destination, namely the receiver. In the conventional wireless communication systems, the information signal can spontaneously propagate at the speed of light with the carrier signal that is electromagnetic waves. Dissimilarly, the signals in the MC propagate via the physical displacement of the particles. By learning from nature, the ways that molecules carrying information propagate through the time and space are applicable to the MC for nano-machines. We list four typical ways as described as follows [32].

A. Signal Propagation through Diffusion

Molecules have an attribute that they randomly move from a position to another, specifically in the liquid or gaseous medium. This phenomenon is called Brownian motion, also known as diffusion. It enables us to use this character of molecules to propagate desired information through the liquid or gaseous medium.

Such MC is known as diffusion-based molecular communication (DBMC). In this signal propagation scheme, the molecules with the encoded information are first released by the transmitter. They can travel through the medium via the Brownian motion spontaneously. Because all the molecules are close to the transmitter at the moment when they are released, there is a concentration gradient from the transmitter to the receiver. This concentration gradient can help the information molecules propagate. Finally, some of the information molecules can arrive the receiver through time to finish the information transfer, as shown in Fig. 2.4-(a).

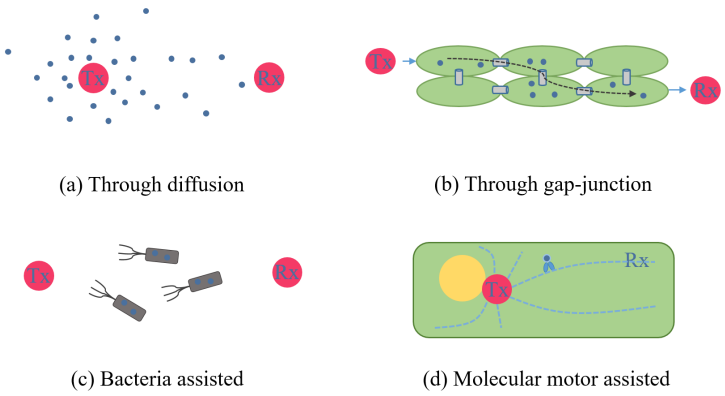

Figure 2.4: Examples of signal propagation in molecular communication

Compared with conventional communication using electromagnetic or acoustic waves, DBMC is relatively slow because the information molecules have to be physically transported from the transmitter to the receiver to finish the signal propagation. What is more, the diffusion process is stochastic so that whether the molecules can arrive at the receiver is indeterminate. Therefore, a directional flow in the medium, known as a drift in the diffusion process, can be applied to assist the diffusion process. It can increase the probability and accelerate the speed of the information molecules to arrive at the receiver.

B. Signal Propagation through Gap-Junction

The second method to implement the signal propagation in MC is through a gap-junction structure [33]. The gap-junctions are widely existing on the cells of vertebrate animals. They can allow not only electrical impulses but also chemical signals based on small particles go through, for example, calcium ions Ca^{2+} as secondary messengers in responding the inositol trisphosphate (IP3). By the mechanism, it can realize the communication between adjacent cell-based nano-machines. This mechanism is as illustrated in Fig. 2.4-(b). When the nano-machines locate close to each other, the gap-junctions can form tunnels from one to another. By controlling the opening and closing states of the gap-junction, similar to a switch used in a conventional network, we can define the routing path. During this process, the signals can also be amplified by multiplying the information molecules in an intermediate node nano-machine. Compared with diffusion, the gap-junction is more suitable for the scenario requires more directivity since we have more control on the routing path.

C. Bacteria Assisted Signal Propagation

Bacteria can also be used to assist the information transportation as mobile nodes [34]. In this scheme, a type of circled DNA segment, called plasmids are used as the information molecules. The plasmids are not the DNA of the bacteria, and play little role in the bacteria functioning. They naturally appear in the bacteria but also can be injected artificially. This bacteria assisted signal propagation is as shown in Fig. 2.4-(c). The bacteria, thanks to the flagellum of such as Escherichia

coli (E. coli), can move a short distance. Therefore, the bacteria are capable of carrying the information molecules, i.e., the plasmids injected into the bacteria, to propagate through the medium. Besides, the movement can be guided to the desired direction towards the receiver by applying the chemotaxis to increase the efficiency [35].

D. Molecular Motor Assisted Signal Propagation

The last possible method has been investigated is using the molecular motor to carry out the signal propagation for MC. This mechanism is originally existing in a living cell. Thus, compared with the other schemes, it is more feasible at a tinier level. The process is as demonstrated in Fig. 2.4-(d). There are microtubules and some motor proteins, for instance, kinesin and dynein, in the Eukaryota cells. The motor proteins can convert the energy to mechanical movement, and walk along the microtubules. By using the microtubules and the motor proteins, information molecules can be transmitted from positions to positions within a small scope. Moreover, it can be also applied in the inversely. Namely, the fixed molecular motors can be used to transmit the information molecules [36], similar to the cargo on a conveyor. This mechanism is suitable for lab-on-a-chip devices.

2.4 Diffusion-Based Molecular Communication

We focus on the DBMC because it has several advantages over the others in realizing the nano-networks currently. First of all, the DBMC requires simple components to carry out the communication. Therefore it is a friendly solution for the communication among nano-machines which are still under research and

difficult to assemble with complex modules. Secondly, the diffusion process can happen in the liquid or gaseous medium spontaneously. This spontaneity means that it is not necessary to consider controlling the switches or higher level carrier such as the gap-junction, the bacteria or the motor proteins used in other schemes. Finally, after being released into the medium, the free-diffusion consumes no extra energy so that it can save the limited power of the nano-machines.

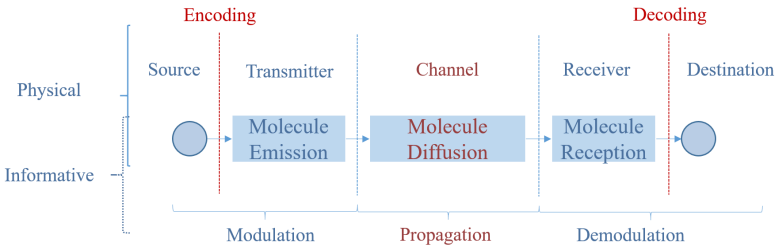

Figure 2.5: Physical and informative flow in the diffusion-based molecular communication

The process of the DBMC from the perspective of physical and informative flow is as shown in Fig. 2.5. The diffusion is the random motion of a particle such as an information molecule, which is a Wiener process. Considering a discrete time, n dimensional system, the random motion of each step during the period of Δt, can be described as

$$\vec{r}_{t+\Delta t} = \vec{r}_t + \Delta\vec{r} \tag{2.1}$$

where, $\vec{r}_t$ denotes the molecule position at time t, and $\vec{r}_t = (x_{1,t}, ..., x_{n,t})$, $\Delta\vec{r}_t$ is the displacement of the molecule during the time period Δt, $\Delta\vec{r}_t = (\Delta x_1, ..., \Delta x_n)$. The displacement of each dimension, nemely, $\forall i \in \{1, ..., n\}$, Δx_i independently

follows a Gaussian distribution

$$\Delta x_i \sim \mathcal{N}(\mu_i, {\sigma_i}^2) \tag{2.2}$$

in which, the mean $\mu_i = 0$, and variance ${\sigma_i}^2 = 2D\Delta t$, D is the diffusion coefficient, used to describe the diffusion speed for a particular type of molecules and fluid environment.

At the micro level, the diffusion of a molecule is partially caused by the collision with other particles in its vicinity. Consequently, the state of the medium, such as temperature, viscosity, etc., also has an impact on the diffusion process. For instance, in a liquid medium with low Reynolds number, the diffusion coefficient of a small (compared with the size of medium molecules) spherical particle is given by the Stokes-Einstein equation as

$$D = \frac{K_b T}{6\pi\mu r} \tag{2.3}$$

where K_b is the Boltzmann constant, T denotes medium temperature in Kelvin, μ is the viscosity of the fluid, and r is the radius of the diffusing particle.

Focusing on the location variation through time of a molecule and applying the probability model of displacement of each dimension, we can derive the probability of a molecule appearing at a position $(x_i, ..., x_n)$ when the time is t as [37]

$$f(x_i, t) = \frac{1}{\sqrt{(4\pi D t)^n}} e^{\frac{-\Sigma {x_i}^2}{4Dt}} \tag{2.4}$$

Because the displacement probability of each dimension is independent and identically distributed, the probability density function (PDF) is the product of the

Gaussian probability density function of each dimension.

From another point of view, for the scenario of a large number of diffusing molecules, the diffusion behavior is governed by the Fick's laws, which take into account the molecule concentration. The Fick's second law describes how the concentration varies through the continuous time as

$$\frac{\partial C(\vec{r},t)}{\partial t} = D\nabla^2 C(\vec{r},t) \tag{2.5}$$

in which $C(\vec{r},t)$ is the concentration at point $\vec{r}$ when t, and ∇ is the Laplacian, whose expansion is $\Sigma(\partial^2/\partial {x_i}^2), i = 1, ..., n$, in n dimensions. From (2.5), we can see that the variation of concentration through time is proportional to its Laplacian, and the proportionality is the diffusion coefficient D.

Given an initial condition that an amount of molecules Q, are released at $t = 0$ by the transmitter located at the origin, namely, the concentration in the whole unbounded-region at $t = 0$ is zero, $C(\forall\vec{r}, 0) = 0, \vec{r} \in \mathbb{R}^3$, the differential equation (2.5) has the solution as

$$C(\vec{r},t,Q) = \frac{Q}{\sqrt{(4\pi Dt)^n}} e^{-\frac{\|\vec{r}\|}{4Dt}} \tag{2.6}$$

The results of (2.4) and (2.6) show that under the same condition, the arrival probability of a single molecule at a particular point in the space, has the same form with the concentration at the location r when time is t, despite the number of released molecules Q.

A molecular impulse, i.e., one molecule or a unite amount of molecules are released by the transmitter when $t = 0$, is considered as an input to the diffusion-based channel, as shown in Fig. 2.6. According to the model above in (2.4) and

(2.6), the output signal of the channel can be generated. It varies through space and time in the medium. For example, observing a position located $20\mu m$ apart from the transmitter, and the information molecule has the diffusion coefficient of $1000\mu m^2/s$, the impulse response of the diffusion-based channel is shown in Fig. 2.7.

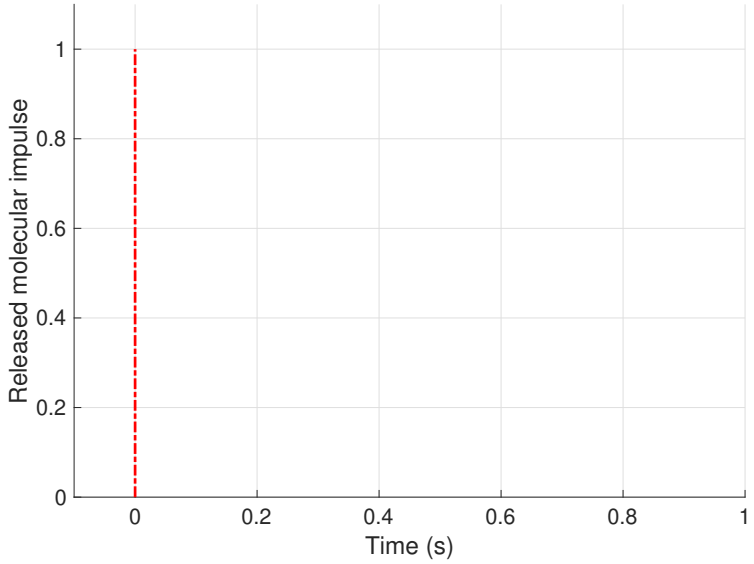

Figure 2.6: Channel input of a molecular impulse (a molecule/a unit amount of molecules)

From the figures, we can see that the maximum concentration of the information molecules at a specific point is five orders of magnitude lower than the channel input, even the measuring point is only $20\mu m$ apart from where the molecules are released. It verifies that the DBMC is for the communication on a tiny scale. Furthermore, the considerable uncertainty of whether the information molecules can arrive the receiver makes a significant influence on the DBMC as a type of noise. Consequently, the DBMC opens a new field for nano-scale communication but also faces many problems to solve.

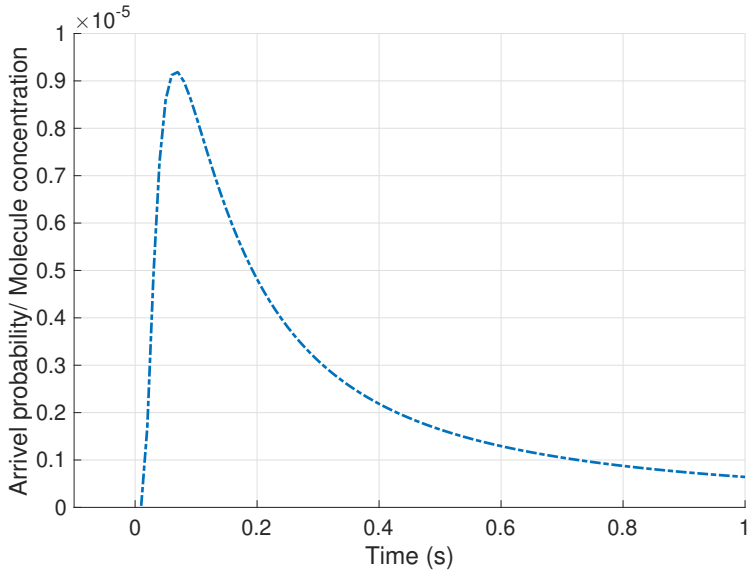

Figure 2.7: Channel output (arrival probability/sensed molecule concentration)

2.5 Summary

MC as a novel paradigm for the communication among nano-machines has distinct features comparing with the conventional electromagnetic wave based communication techniques. The information carrier in the MC is molecules rather than the radio waves. Thus, the information is encoded onto and decoded from the information molecules using their physical or chemical characters, for instance, the concentrations, types, released times, composition structures, along with others. The physical channel of MC, namely the signal propagation schemes rely on the physical replacement of the information molecules. It includes the ways such as the diffusion, the gap-junction, with the assistance of bacteria and molecular motors. Among all, the DBMC draws our attention because of its advantages such as energy efficiency. Its channel model can be derived through probability and physical approaches. The impulse response of a diffusion-based channel can be obtained with (2.4) and (2.6) and is as shown in Fig. 2.6 and 2.7. Our study on DBMC is based on this channel model.

Chapter 3

Communication Distance Measurement in Diffusion-based Molecular Communication

In conventional wireless communication techniques, communication distance plays an important role in many aspects and helps various applications in different scenarios. Similarly, in the DBMC, the communication distance is also a crucial factor for the protocol designing. In this chapter, we propose a communication distance measurement method based on the arrival time difference. We first propose three noise models in the DBMC channel. In the proposed method, we apply two types of information molecules with different diffusion coefficient. Thus, when the signals with the two type information molecules are transmitted at the same time, there is a time difference between the received signals. Based on the time difference and the channel character, we can estimate the communication distance between the transceivers. This method uses time-based protocol and one-way signal. Consequently, it is efficient and has a simple requirement for the nano-machine nodes.

3.1 Motivation

Distance information plays an important role in numerous wireless communication associated applications. Many localization systems have been designed and developed based on distance measurement or estimation, including the well-known Global Positioning System (GPS) [38]. Furthermore, some real-time target tracking systems using wireless sensor networks [39] are also based on the distance estimation. Not only used for practical applications, but the communication distance information can also be directly used for the optimization of wireless communication. For example, the wireless sensor networks are required to sense and collect data from a field of interest for some time with the limited power source. To this end, the communication distance can be used to design an energy-efficient routing protocol [40].

Similarly, distance measurement is also a crucial problem of the DBMC. First of all, the connection quality of even end-to-end DBMC relies significantly on the communication distance. Because of the uncertainty and the long-delay character, a tiny variation on the communication distance can cause severe fluctuations in the balance between the transmission rate and the bit error rate due to the interference [41]. Secondly, distance information between the mobile nano-machine nodes can contribute to the even deployment to maximize the coverage rate of a bio-sensing system [42]. Thirdly, within a particular distance, the nano-machines can be used for target tracking tasks using the repellents and attractants [43], as a type of DBMC. Therefore the distance is also essential information to increase the efficiency of the usage of the information molecules for target tracking. The applications of nano-machine and the MC that are based on the communication

distance information are not limited to the examples listed here.

3.2 System Model

For the target of communication distance estimation, we consider a system consisting of two nano-machines. They apply the DBMC to communicate with each other. For simplicity, we assume the nodes are deployed in a one-dimensional aqueous environment to simulate a microfluid-based lab-on-a-chip system [44]. The similar analysis and numerical results can be easily extended to two- or three-dimensional cases. The communication distance between the two nodes is fixed and denoted as d_c. The size of both nodes is negligible. In addition, both nodes are capable to transmit and receive signals based on the information molecules. However, when applying one type of information molecules, the two nodes can not transmit signals at the same time, because the information molecules can interfere in the medium. To solve this problem, half-duplex or multiple types of information molecules should be applied. Namely, the nodes transmit messages one by one according to a specific protocol.

Because our interest is the distance estimation, we do not take the encoding and decoding schemes into account. Moreover, we focus on a single time distance estimation assuming that the distance information is for the transmitter to adjust parameters to improve the transmission quality and efficiency. Thus, we define one of the nano-machine nodes as the transmitter (Tx), which is the node requires the distance information. Another node is assumed as the receiver (Rx), which receives the messages when the connection is established. We map the system to a Cartesian coordinate system. The transmitter is set at the origin(0), and the

receiver is deployed d_c apart from the transmitter on the axis, i.e., (d_c). The system is as demonstrated in Fig. 3.1

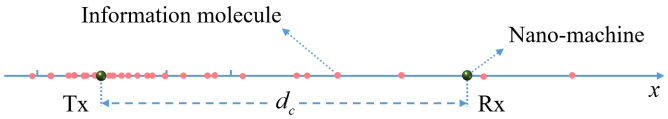

Figure 3.1: An example of a diffusion-based molecular communication system

Our objective is to estimate the distance between the transceivers based on the channel response to an input signal as simple as possible. The distance information can be obtained from either molecule concentration, i.e. amplitude, or arrival time of a particular amplitude of the received signal. However, there are several types of noises in the channel that can have influences on the signal and then further affect the distance estimation. In the following section, we model the noises that exist in the diffusion-based channel in three categories..

3.3 Noise Model

We estimate the communication distance from a received signal which is responding to a simple input signal according to the diffusion-channel character. There are three steps of the molecule-based signal propagation in the diffusion-based channel: a) Tx releases molecules into the medium as input signal; b) information molecules propagate via diffusion according to channel response model; c) Rx measures the peak concentration or its arrival time to calculate communication distance.. In each step, there is a kind of factor that can influence the signal propagation, which can be seen as a type of noise. The noise of each step of the signal

transmission in the channel is unavoidable and have the effects on the distance estimation as well, we categorize the noises into three types based on the appearance phases, as the input noise, the convection noise, and the reception noise. The signal transmission flow of each type of noises is as demonstrated in Fig. 3.2. We detail the noise models in the following subsections.

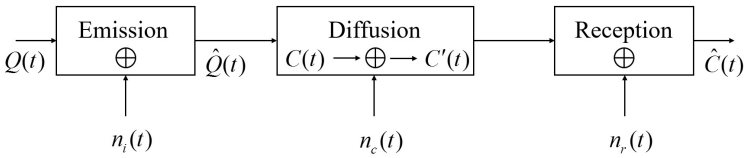

Figure 3.2: The signal transmission flow and noises

A. Input Noise

The imperfectness of the nodes in the system is unavoidable, especially when it comes to the nano-machine nodes that have limited capabilities. Namely, there are perturbations on the number of released molecules, because of the imperfect transmitter [45], as channel input. Such perturbation can be seen as a noise at the input signal. We consider this kind of noise as the first type of noise in the diffusion-based MC system. This noise is called input noise and denoted with $n_i(t)$, and empirically it is modeled as a simple additive white Gaussian noise (AWGN), follows the normal distribution with variance σ_i^2, i.e., $n_i(t) \sim \mathcal{N}(0, \sigma_i^2)$. As a result, we can derive the input signal for distance measurement with noise as

$$\hat{Q}(t) = Q(t) + n_i(t) \tag{3.1}$$

where $Q(t)$ is the perfect channel input. And we take two forms of the input signal into account. One is a Dirac delta impulse input. With this input signal,

the Tx node releases a particular number Q, of information molecules at once for instance, when $t = 0$

$$Q_i(t) = \begin{cases} Q, & \text{if } t = 0 \\ 0, & \text{else} \end{cases} \tag{3.2}$$

Another one is a single square wave. In other words, information molecules are released to the medium at a constant rate of Q, from $t = 0$, for a period T_i. We denote the square wave-based channel input $Q_s(t)$ as

$$Q_s(t) = \begin{cases} Q, & \text{if } 0 \leq t < T_i \\ 0, & \text{else} \end{cases} \tag{3.3}$$

B. Convection Noise

The DBMC is typically used for scenarios of in-body usage such as the targeted drug delivery. The environment of organs or tissues can be simplified as a viscous liquid for the diffusion-based channel. In such an environment, the unexpected tiny convection or advection in the medium is unavoidable. In the diffusion process, especially for the tiny scale for the DBMC, even a small fluctuation in the medium can have severe impacts on the concentration variation and then influence the received signal. We approximately consider the convection or advection as an uncertain drift and a type of noise in the diffusion-based channel. The uncertain drift of the medium various at different locations (d) through time t, with velocity $\overrightarrow{v}(d, t)$, which is a set of random vectors. Empirically, we let the velocity $\overrightarrow{v}(d, t)$ follows a normal distribution $\mathcal{N}(0, \sigma_v^2|(d, t))$, the positive and negative represents

the direction on the axis. We can have the noise caused by convection or advection, denoted as $n_c(d,t)$, and $n_c(d,t) \sim \mathcal{N}(0, \sigma_c^2)$. When considering the convection in the medium, the diffusion equation is transformed as

$$\frac{\partial C(d,t)}{\partial t} = D\nabla^2 C(\vec{r},t) - \overrightarrow{v}(t)\nabla C(d,t) \tag{3.4}$$

For simplicity, we only consider the impulse input in (3.2), we can derive the channel response from the convection-diffusion equation, as

$$C'(d,t,n_c(t),Q_i) = \frac{Q_i}{\sqrt{4\pi Dt}} e^{-\frac{(d,-n_c(d,t))^2}{4Dt}} \tag{3.5}$$

Similarly, channel response to the rectangular wave input can be done as

$$C''(d,t,n_c(t),Q_s) = \int_0^{T_i} \frac{Q_s(\tau)}{\sqrt{4\pi Dt}} e^{\frac{-d^2}{4Dt}} d\tau \tag{3.6}$$

Since the drift affects the communication distance in the channel response equation, the convection noise $n_c(d,t)$ does not only affect signal amplitude but also influence the arrival time of the signal.

C. Reception Noise

The last type of noise in the diffusion-based MC system is2 the reception noise. It is due to the randomness of the diffusion process as well as the molecule counting precision. According to the analysis in [23], the number of molecules received by Rx, denoted as $\hat{C}(t)$, conforms to a Poisson distribution with the molecule number counted at Rx $C(t)$, , i.e. $\hat{C}(t) \sim \mathcal{P}oiss(C(t))$. Considering the number of molecules should be a large number, $C(t) \gg 1000$, it is reasonable to use

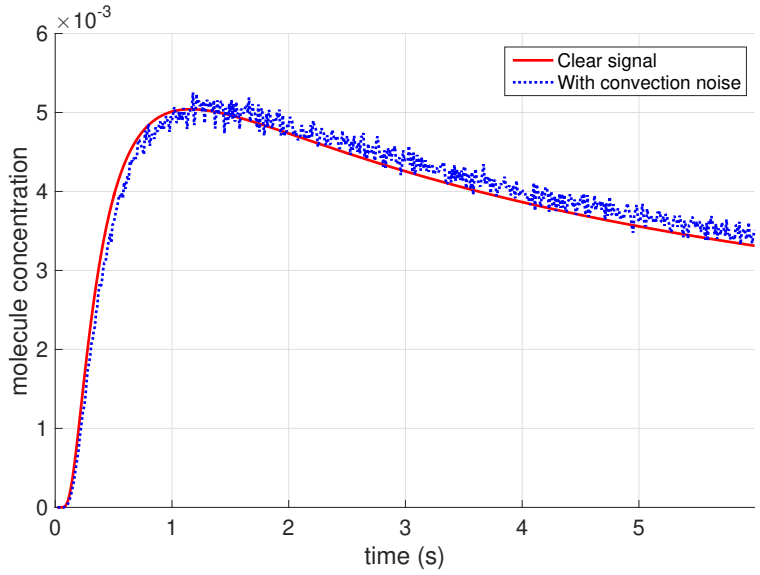

Figure 3.3: Demonstration of convection noise

Gaussian distribution to approximates the Poisson distribution. Thus, we model the reception noise of diffusion-based channel as another Gaussian noise, in other words, $\hat{C}(t) \sim \mathcal{N}(C(t), C(t))$. The reception noise $n_r(t)$ is the difference between measured concentration $\hat{C}(t)$ and expected concentration $C(t)$. Consequently, $n_r(t) \sim \mathcal{N}(0, \sigma_r^2)$, in which $\sigma_r^2 = C(t)$. Finally, the reception noise directly affects the amplitude of the received signal. The channel out amplitude with reception noise can be expressed as

$$\hat{C}(t) = C(t) + n_r(t) \tag{3.7}$$

Conclusively, based on the above analysis, considering the flow of the information molecules from the transmitter to the receiver, we categorize the noises into three types including the input noise, the convection noise, and the reception noise, appearing at different phases of the transmission. All the three types of noises are

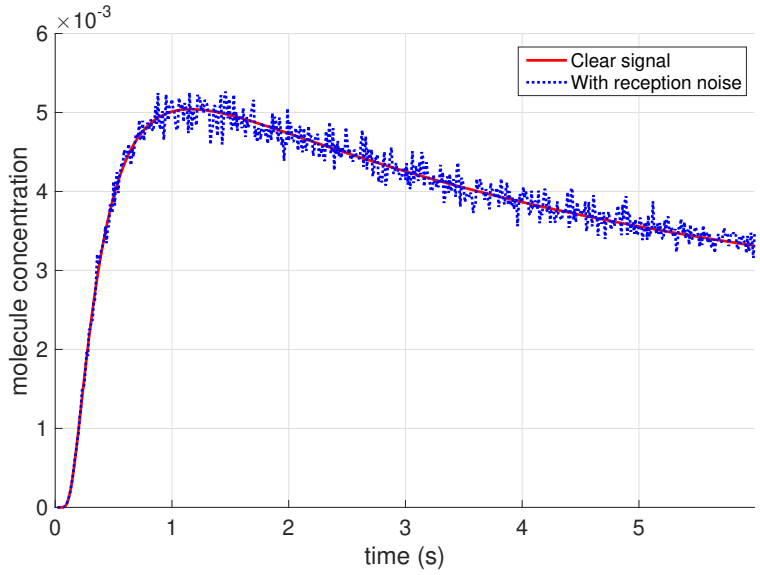

Figure 3.4: Demonstration of reception noise

modeled as AWGN. For a fixed communication distance d_c, considering all of the noises, the channel response to an impulse input Q_i can be derived as

$$\hat{C}(t, n_i(t), n_c(t), n_r(t)) = \frac{Q + n_i(t)}{\sqrt{4\pi D t}} e^{-\frac{(d_c - n_c(t))^2}{4Dt}} + n_r(t) \tag{3.8}$$

where, $n_i(t) \sim \mathcal{N}(0, \sigma_i^2)$, $n_c(t) \sim \mathcal{N}(0, \sigma_c^2)$, $n_r(t) \sim \mathcal{N}(0, \sigma_r^2)$, $\sigma_r^2 = C(t)$. In practical scenarios, all the noises should exist simultaneously. However, for the sake of distance measurement performance analysis, we classify them and test the impacts individually. The performance analysis is given in 3.5.2. In the following section, we describe two methods also focusing on the distance estimation in DBMC in related works, and then present the proposed method.

3.4 Related Works

Even though distance measurement is crucial to implement DBMC to practical applications, to the best of our knowledge, very limited works have been carried out in this field. Inspired by the received signal strength index (RSSI) and time of arrival (TOA) approaches, similar techniques for distance measurement in DBMC are investigated in [46]. The mathematical derivation is fundamental for our proposed method as well. Thus we briefly introduce the two related works in the following subsections.

3.4.1 Round Trip Time Protocol

At the beginning of this approach, Tx transmits an impulse input signal and starts a timer. When Rx senses the peak value of the received signal, it sends back an impulse reply. Finally, Tx waits for the responding peak value, and measures the time elapsed, which is the round trip time, denoted as t_{RTT-P}. When the peak concentration of the forwarding signal reaches the target node, the slope of channel output is zero

$$\left.\frac{\partial C(t)}{\partial t}\right|_{t=t_{RT},C(t)=Max(C)} = 0 \tag{3.9}$$

where $C(t)$ is (3.2) or (3.3), depends on the input formation, and it yields the trip time, i.e., delay as

$$t_{RT} = \frac{d^2}{2D} \tag{3.10}$$

If the forward signal and feedback signal use the same information molecule, namely they have the same diffusivity D, the round trip time t_{RTT-P} is twice of the trip time t_{RT-P}, i.e., $t_{RTT-P} = 2t_{RT-P}$. Thus, we can derive the communication based on the round trip time as

$$d_{RTT-P} = \sqrt{Dt_{RTT-P}} \tag{3.11}$$

3.4.2 Signal Attenuation Protocol

In this protocol, the peak molecule concentration, namely the maximum amplitude of the received signal, is measured at one node to calculate the distance from the other one. Based on the above analysis (3.9), we can simply substitute (3.10) into $C(t)$ (2.6) to yield the relation between communication distance and peak concentration of received signal.

$$d_{SA-P} = \frac{Q}{C_{peak}\sqrt{2\pi e}} \tag{3.12}$$

in which d_{SA-P} is the communication distance calculation with SA-P, C_{peak} denotes the peak concentration value.

3.5 Distance Measurement based on Arrival Time Difference

The method introduced above can estimate the communication distance of the nano-machine nodes using the DBMC. The attenuation of the signals in the diffusion-based channel is very harsh. However, the SA-P protocol relies on precise measurement of the signal amplitude, namely the molecule concentration. Therefore it requires a high-resolution nano-scale concentration sensor for

the nano-machine nodes. On the other hand, the RTT-P protocol is a time-based method, so that it has more simple requirement of the nano-machines. Nevertheless, in order to avoid the synchronization problem of the Tx and Rx, the RTT-P uses a round-trip signal. It has the problem that not only the error can be accumulated in the round-trip, but also lack of efficiency. Aiming at more efficient time-based distance estimation, we propose the arrival time difference (ATD) based method.

3.5.1 Methodology

Because in the same medium, different molecules usually have different diffusion coefficient D, according to the Einstein relation (2.3). In the same MC system, where medium temperature and fluid viscosity are same, D is determined by the size of the information molecule. As long as two types of particles used for signal propagation are different and have different radiuses, they have different diffusion speed in the medium, e.g. sodium chloride (NaCl) and sucrose in an aqueous environment. As a result, there is a difference between the concentration peaks of the signals carried by the two molecules. Based on this arrival time difference (ATD), the communication distance can be calculated.

We assume the two types of information molecule are type A and B. Parameters for each are subscripted with A and B. In the ATD approach, Rx transmits signal A and B at the same time $t = 0$, both with the same input form. Then when Tx will receive two different signals. Assuming molecule A has a higher diffusivity, $D_A > D_B$, physically signal A arrives Tx earlier. Since Tx senses the peak of signal A at time t_A, it starts counting the time until the peak of signal B

arrives when $t = t_B$. We denote the elapsed time as t_{BA}, then same as the analysis in 3.4.1, the relation between communication distance and delay yields

$$d_{ATD} = \sqrt{\frac{2D_A D_B(t_B - t_A)}{D_A - D_B}} = \sqrt{\frac{2D_A D_B t_{BA}}{D_A - D_B}} \tag{3.13}$$

where $t_{BA} = t_B - t_A$.

Similar to RTT-P, ATD also calculates communication distance by measuring signal arrival time. However, by adopting two types of information molecules, ATD overcomes the efficiency shortcoming of RTT-P, because Tx only needs to count the arrival time difference rather than to wait for a round trip of the signal.

3.5.2 Performance Analysis

In this section, we present the performance evaluations of the proposed ATD approach. The performance comparisons with the other two approaches in [46] are also carried out with simulation experiments. We carry out 2000 times Monte Carlo simulation experiments to evaluate the root mean square error (RMSE) of distance measurements versus several parameters such as the variances of noises, for each method. RMSE for n time simulations is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(d - \hat{d})^2}{n}} \tag{3.14}$$

in which, d is the pre-set communication distance, $\hat{d}$ is the estimated value. In each simulation, we use $0.001s$ as the step width and $500s$ as diffusion duration (t_{max}). The communication distance (d_c) in our experiments is set as $48\mu m$. The channel inputs an impulse signal as modeled in (3.2), the value of Q is 1. Diffusion coeffi-

Table 3.1: Simulation parameters used for distance measurement

Symbol	Description	Value
t_{step}	Simmulation step width	$0.001s$
t_{max}	Diffusion time	$500s$
d_c	Communication distance	$48\mu m$
Q	Impulse input	1
D_A	Diffusion coefficient of molecule A	$1000\mu m^2/s$
D_B	Diffusion coefficient of molecule B	$370\mu m^2/s$

cient (D or D_A) is set to be $1000\mu m^2/s$, which is the diffusivity of NaCl in water when the temperature is $300K$. In ATP, the second molecule type is assumed as sucrose, whose diffusivity denoted as D_B is $370\mu m^2/s$ in the same environment as above. The parameters that are used for the simulation are summarized in TABLE 3.1.

First, we test the impacts of noises to the distance measurement accuracy Figures 3.5 to 3.7. Each types of noises are added to the simulation individually. We calculate signal to noise ratio (SNR) in dB as the metric to indicate strength of noises, which is calculated according to

$$SNR_{dB} = 10\log_{10} \int_0^{t_{max}} \frac{C^2(t)}{[C(t) - \hat{C}(t)]^2} dt \tag{3.15}$$

where $C(t)$ denotes the amplitude of a clean signal without noise at time t, meanwhile we use $\hat{C}(t)$ to denote amplitude of distorted signal.

Fig. 3.5 indicates the relation between RMSE and the input noise. We test the input noise with variance σ_i^2 form 0 to 0.1, which leads to the SNR varies from $12dB$ to $21dB$. From the curves, we can see that the input noise barely influence the measurement accuracy with arrival time-based methods, namely ATD and RTT-P. On the contrary, SA-P is significantly affected by input noise. However,

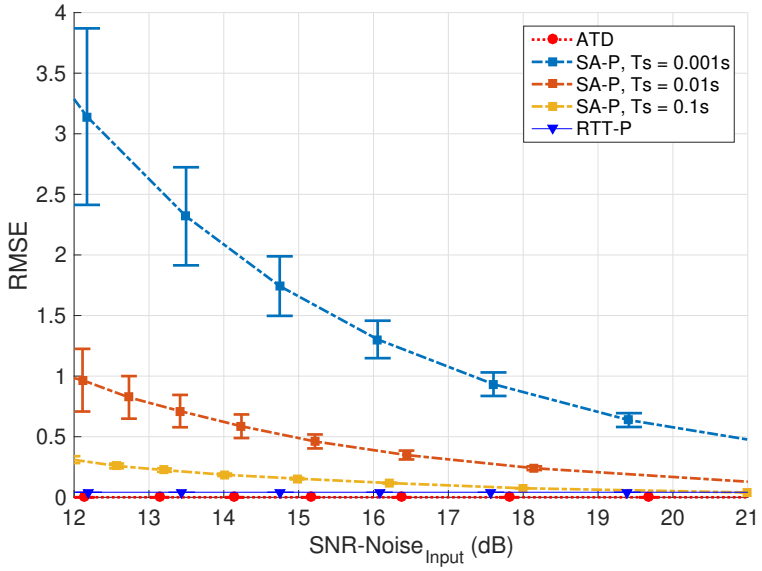

Figure 3.5: Estimation error vs. input noise

by adopting the square wave input (3.3), the RMSE of SA-P can be reduced. In addition, as the input signal width (T_s) increases, the SA-P can be more robust to the input noise. The impacts of convection noise and reception noise are as shown in Fig.4 and 5. The variance of these noises, σ_c^2 and σ_r^2 are both set to be 0 to 0.1. Fig. 3.6 and Fig. 3.7 both indicate that among the three approaches, SA-P has outperformance over the other two methods. Meanwhile, the results prove the analysis that ATD can reduce the measurement error comparing with RTT-P.

The tendency of RMSE versus communication distance is shown in Fig. 3.8. With the given parameters, as d_c increases from $50\mu m$ to $1950\mu m$, RMSE of SA-P increases slightly. Its value keeps steady below $1\mu m$. As to the other two approaches, it shows that ATD has less RMSE than RTT-P for any d_c, which proves that ATD has a more accurate measurement to variable communication distance. Dislike SA-P, the RMSEs of arrival time-based methods ATD and RTT-P, various along with d_c. When d_c is shorter than $200\mu m$, RMSEs of bot ATD and RTT-P in-

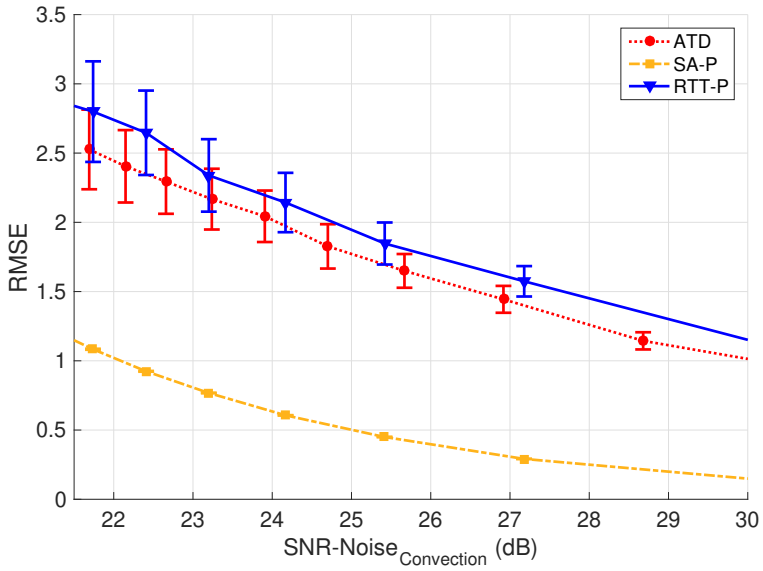

Figure 3.6: Estimation error vs. convection noise

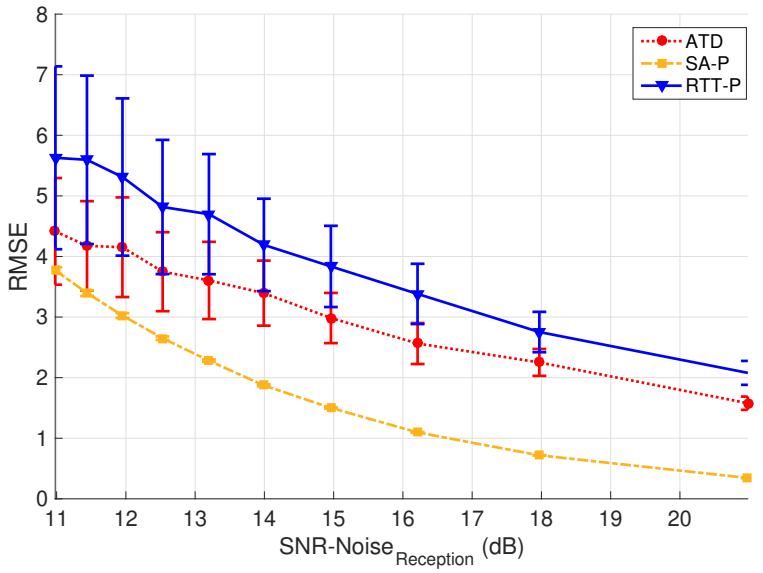

Figure 3.7: Estimation error vs. reception noise

creases with d_c. After then, they tend to decrease as d_c becomes longer. When d_c is longer than nearly $750\mu m$ and $1050\mu m$ respectively, ATD and RTT-P turn to have better measurement accuracy performance than SA-P. This tendency is because when d is short, the signal arrival time is small, so that it is easily affected by the uncertainty of diffusive signal, leading to an increase of RMSE. However, when d

is sufficiently long, the signal arrival time is long as well. Then the diffusion can only have limited impacts on the accuracy of distance measurement.

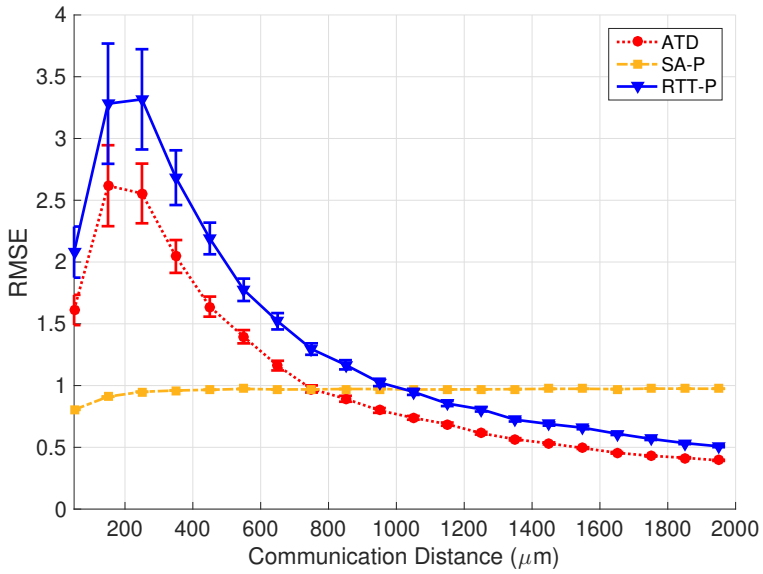

Figure 3.8: Estimation error vs. communication distance

The last parameter used for comparison is the diffusivity D. Results are illustrated in Figures 3.9 and 3.10 . Fig. 3.9 compares RMSEs of the three methods when the diffusion coefficient D or D_A varies from 800 to 1200$\mu m^2/s$. Notably, in Fig. 3.9, the difference between D_A and D_B is fixed to 600$\mu m^2/s$. The figure shows that accuracy performance of SA-P and RTT-P has no relation to the variation of D. The only accuracy of ATD is deteriorated as D increases. From (3.13), we can see that the error in measuring signal time determines the accuracy of ATD. The ATD method needs to count signal time two times, introducing more uncertainty to the algorithm. However, in an actual MC system, the diffusivity of information molecules are determined. Using the selected molecules in our system, which are sodium chloride and sucrose, as emphasized in Fig. 3.9, the accuracy of ATD is still over RTT-P. Namely, the performance of ATD can be

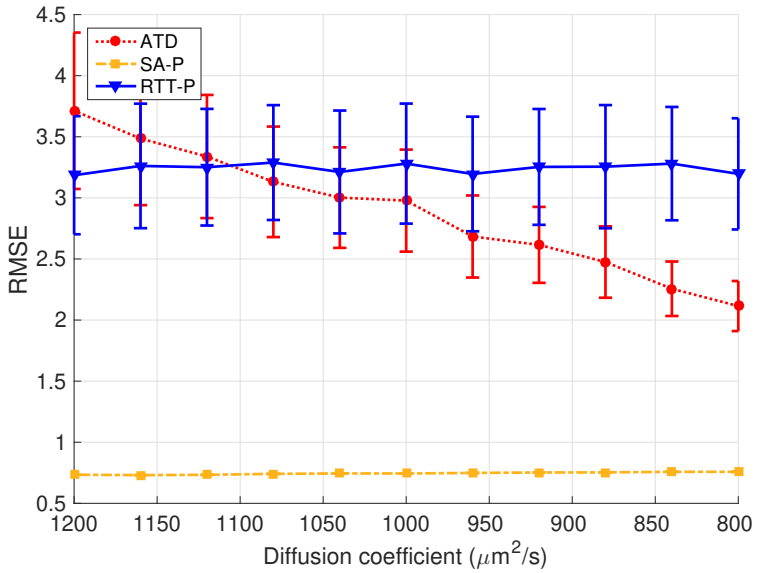

Figure 3.9: Estimation error vs. diffusion coefficient

affected by what types of information molecules are used. Therefore, we show the relation of RMSE and the difference of diffusivity of two types molecules used by ATD in Fig. 3.10. The horizontal label shows the diffusion coefficient of the type B information molecule when the diffusivity of type A is fixed to $1000\mu m^2/s$. From the figure, we can also get to the inference that the performance of ATD is relative to the chosen type of information molecules. When the diffusivities are small, and the difference between diffusivities of the two kinds of molecules is sufficiently large, for instance, NaCl and sucrose, the performance of ATD can be guaranteed.

3.6 Summary

This chapter addressed the communication distance estimation problem of the DBMC. Distance information can be helpful for various applications. Due to the limited capability of nano-machines, existing distance estimation methods apply

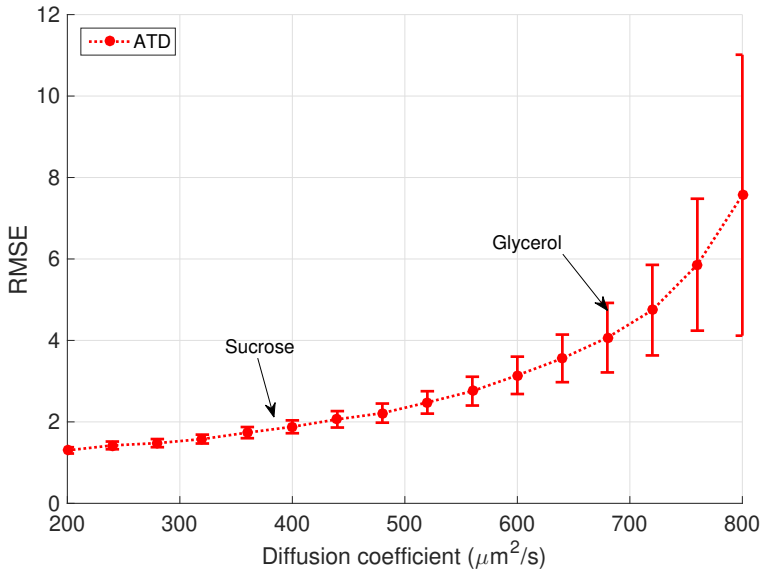

Figure 3.10: Estimation error vs. the difference of diffusion coefficients of ATD

either a signal attenuation protocol or a round-trip time protocol. The first one relies on accurate amplitude measurement, and the later one uses round-trip signal because the synchronization of nano-machine nodes is difficult. Aiming at these problems, we proposed a simple time-based protocol using two types of information molecules. Because of the different diffusion coefficient, there is an arrival time difference of the received signals transmitted at the same time. Therefore, the proposed method is simple since it is a time-based protocol, and more efficient and accurate because it uses one-way signals, and the error accumulated in the round trip is avoided. Simulation experiments have verified the efficiency and accuracy of the proposed method.

Chapter 4

Mitigation of Bit Error Rate caused by Inter-symbol Interference

In this chapter, we try to address the inter-symbol interference (ISI) problem. The ISI is one of the most crucial issues of DBMC. The ISI is the interference on the later transmitted bits of the remaining information molecules of former bits that are still diffusing in the medium. It is caused by the uncertainty and the long delay of the diffusion process. It may result in a high bit error rate (BER) in the transmission of using DBMC. Aiming at this problem, we propose two methods including a retransmission scheme and an adaptive code width (ACW) protocol in an end-to-end system with mobile nodes. Both ways apply an acknowledgment message from the receiver to the transmitter contains either the retransmission request or the periodical impulse signal based on which the transmitter can estimate the communication distance. According to the acknowledgment, the transmitter calculates the bits need to retransmit and adjust the code width for the following transmission. The proposed methods can reduce the error bit rate caused by the ISI.

4.1 Motivation

The CSK is currently the most widely studied modulation scheme for the DBMC. The advantage of the CSK over the others is mainly the simple requirement of the system. Thus it is more suitable for the DBMC, considering the limited capability of the nano-machine. Many types of molecules including the insulin, sucrose, ions such as K^{+}, Ca^{2+}, NO_3^{-}, etc. are adequately applicable for the CSK of DBMC.

However, it faces an ISI problem due to the uncertainty and the long delay of the diffusion process. Because of the features of diffusion, some information molecules may fail to arrive at the receiver in their corresponding slot. They may remain and accumulate in the medium for a while. Thus, when these information molecules finally arrive at the adjacent scope of the receiver and sensed by it, they will be counted as a part of the concentration of the transmitted bit in the current slot. Such influence of the former transmitted bits on the later bits is considered as the ISI in the DBMC using CSK modulation scheme.

The ISI is a crucial problem of DBMC because it considerably affects the transmission quality. In general, CSK uses a predefined threshold at the receiver to determine whether the received bit is a bit 0 or 1. It saves lots of the limited computational resources of the nano-machine node for other processing tasks. However, the false alarm, i.e. the receiver judges the received bit as 1 when the transmitted bit is a 0 by mistake, happens abundantly due to the ISI in the channel. The false alarm happens when the amplitude of the received signal is above the threshold, yet most of the ingredient is the ISI rather than the intended signal.

Thus, it is essential to mitigate the influence of ISI of CSK-based DBMC, to

reduce the BER and ensure the transmission quality.

4.2 Related Work

A number of studies of DBMC have been done trying to deal with the ISI problem.

In [47], the channel capacity of the DBMC is analyzed taking the ISI into account. The optimal threshold on the receiver side is mathematically derived using the Skellam distribution. By doing so, the channel capacity can be achieved with fewer information molecules by controlling the channel transmission probability.

The thresholds used for the decision of received signals can be calculated empirically [48]. The optimal threshold with which the error rate can be minimized, is obtained after receiving the information of interest. In another work [49], the threshold method is improved to be adaptive, during the information transmission. Through time, the optimal threshold can be found by the algorithm.

A reduced pulse width method is proposed in [50]. In this method, the effort focuses on the pulse width during the transmission. The idea is to decrease the time period of releasing the information molecules to reduce the molecule concentration which composes the ISI.

[51, 41] proposed two methods to mitigate the inter-symbol interference. One adopts MoSK additionally to the CSK. The two different types of molecules in odd and even slots can increase the slot width for each type of molecules so that the diffusion time for each of them is increased. As a result, the ISI is then mitigated. The second method actively adjusts the amplitude, namely the number of the molecules, of the transmission of bit 1s, according to the states of the former

bits. When there are many bit 1s in the former slots, the next bit 1 is transmitted with a smaller amplitude. It is a pre-coding technique applied at the transmitter side. Not only the ISI can be reduced, but also the efficiency of the information molecule usage is improved by this method.

However, these works have not considered the situation that the DBMC is used for mobile nano-machine nodes. The practical applications often require the nano-machines to move, such as flow within the circulation system to accomplish drug delivery. The channel status changes dramatically as the variation of the communication distance. Therefore the ISI is more crucial for the nano-networks using DBMC for mobile nodes.

4.3 System Model

In this chapter, we consider a system consists of two mobile nodes. They apply the unicast, half duplex communication through the DBMC. For simplicity we assume one node is used as the transmitter (Tx) to transmit messages encoded onto the information molecules. The other one acts as the receiver (Rx). Both nodes are capable of transmitting and receiving signals on the molecules.

The system is deployed in a two dimensional environment under the assumption that the system is first applied to a Petri dish for test. Both of the nodes have the mobility in the medium with a fixed velocity to simulate the case of mobile nodes in a nano-network. The system is as illustrated in Fig.4.1.

Since our objective is to address the ISI problem in this chapter, the CSK is applied in the system. We consider the most simple CSK scheme, namely the on-off keying (OOK) of the DBMC. In DBMC, the OOK can be modeled as follows.

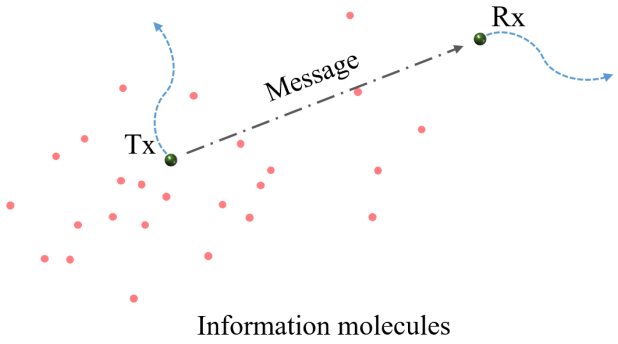

Figure 4.1: A DBMC system with two mobile nano-machine nodes

There are two states of a binary bit, '1' or 'on' and '0' or 'off'. The transmitter instantly releases a particular number Q_1 information molecules into the medium to represent a bit '1'. The released molecules can generate an impulse of the concentration at the time when they are released. On the other hand, transmitting a bit '0' with the OOK can be done through the action that the transmitter releases a fewer number Q_0 molecules. More generally, $Q_0 = 0$, namely in a slot of bit '0', the transmitter releases no molecules for the purpose of more distinct from the bit '1'. For instance, in the nth slot, the bit to be transmitted is defined as

$$b_{T,n} = \begin{cases} 1, Q_n = Q_1 \\ 0, Q_n = 0 \end{cases}, n = 1, 2, ... \tag{4.1}$$

According to the channel impulse response model (2.6), the amplitude, namely the molecule concentration of the received signal is a function of the communication distance and time. Apparently for a bit '0' when $Q_0 = 0$, the channel outputs no variation in the molecule concentration around the receiver. Thus, a pre-defined threshold Q_{Thr} can be applied at the receiver to distinguish the transmission of bit

'1's and '0's. When the concentration sensed by the receiver is higher than Q_{Thr}, it holds that the bit in the slot is a '1'. Otherwise, it is thought to be a '0'. The determination of the received bit in the nth slot is

$$b_{R,n} = \begin{cases} 1, Q_n > Q_{Thr} \\ 0, Q_n < Q_{Thr} \end{cases}, n = 1, 2, ... \tag{4.2}$$

4.4 Inter-symbol Interference in Diffusion-Based Molecular Communication

4.4.1 Mathematical model

The ISI is the interference of the remaining molecules of the former bit '1's on the bit in the current slot. To the end of modeling the ISI, we first calculate the received signal amplitude $A_{r,n}$of the bit in the nth slot with

$$A_{r,n}(d,t) = b_{T,n} * C(d, t - nTs, Q_1), t > nTs \tag{4.3}$$

where Ts is the time slot width, namely the time interval of two consecutive bits, therefore it is also known as code width, $C(d,t,Q)$ is the concentration-based channel response function (2.6). Accordingly, the ISI on the nth bit is the sum of remaining molecule concentration of the former bit '1's

$$ISI_n(d,t) = \sum_{i=1}^{n-1} A_{r,i}(d,t) \tag{4.4}$$

The definition of ISI suggests that the residual molecules in the medium of all the former bits can cause a higher molecular concentration of the following bits. Thus when the nth bit is '1', the amplitude is enlarged. The enlarged amplitude is

Table 4.1: Simulation parameters used in Chapter 4

Symbol	Description	Value
t_{step}	Simmulation step width	$0.01s$
Q_1	Channel input of bit '1'	1
Q_0	Channel input of bit '0'	0
D	Diffusion coefficient	$1500\mu m^2/s$
d	Communication distance	$[1-100]\mu m$
Ts	Time slot width	$[0.1, 0.25, 0.5, 1, 2.5, 5, 10]s$
fs	Signal frequency	$[10, 4, 2, 1, 0.4, 0.2, 0.1]Hz$

still higher than the threshold Q_{Thr}. However, it is a problem when the nth bit is a'0'. If the enlarged amplitude is higher than the threshold Q_{Thr}, this bit will be determined as a '1' erroneously. It is considered as the false alarm. Consequently, there will be an error at this bit. Moreover, the residual molecules will keep accumulating in the medium. As a result, there is a high probability that the following bits '0's will be false alarmed.

4.4.2 Impacts on Intersymbol Interference

According to the definition equation of the ISI, we can see that it is a function of communication distance d and time t. And the time slot width Ts can also influence the value of ISI. To evaluate the impact of d and Ts on the ISI, we carry out several transmission simulations of a bit sequence, with different parameter settings. The sequence consists of a series of randomly generated bit '0's and '1's. The simulation parameters are set as shown in TABLE 4.1. In addition, we use the signal to interference ratio (SIR) as the metric to evaluate the ISI level, which is defined as

$$SIR = \log\left[\frac{RMS(\bar{A}_s)}{RMS(ISI)}\right]^2 \tag{4.5}$$

in which, $\bar{A}$ is the amplitude of the clear signal, which can be calculated by (4.3), ISI is the amplitude of the ISI (4.4), and $RMS(\cdot)$ is the root mean square function.

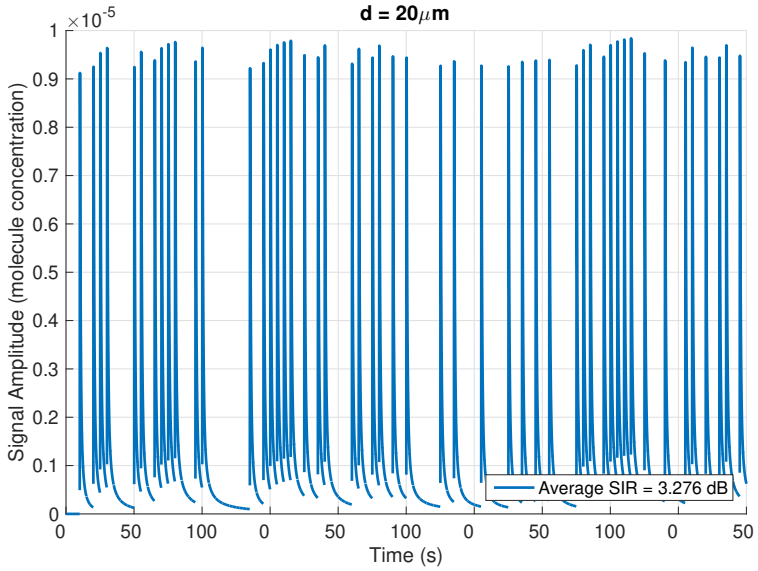

Figure 4.2: Received molecule concentration when distance $d = 20\mu m$

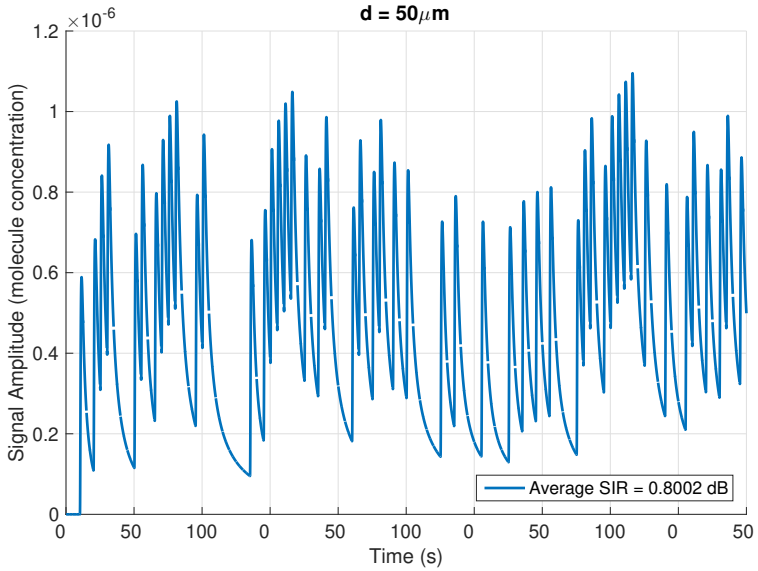

Figure 4.3: Received molecule concentration when distance $d = 50\mu m$

First, we simulate the various communication distance d. In the experiments, communication distance changes from 1 to 100 μm, while Ts is fixed at 1s. Some

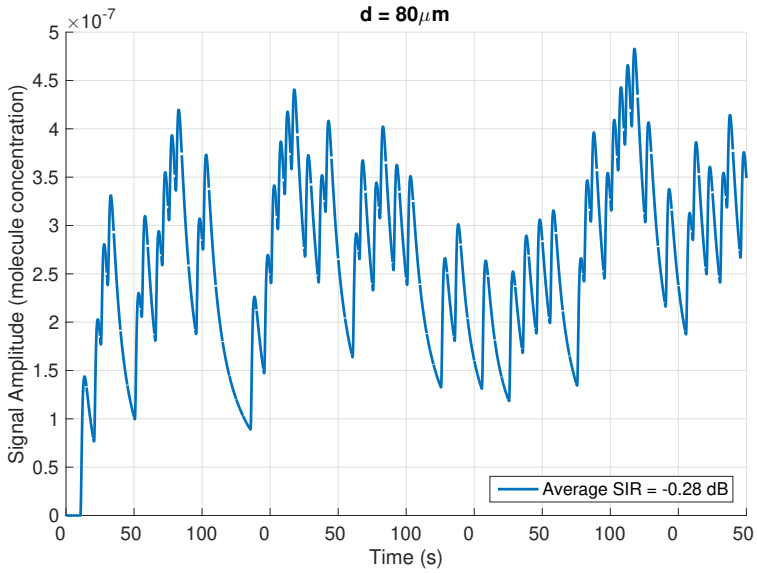

Figure 4.4: Received molecule concentration when distance $d = 80\mu m$

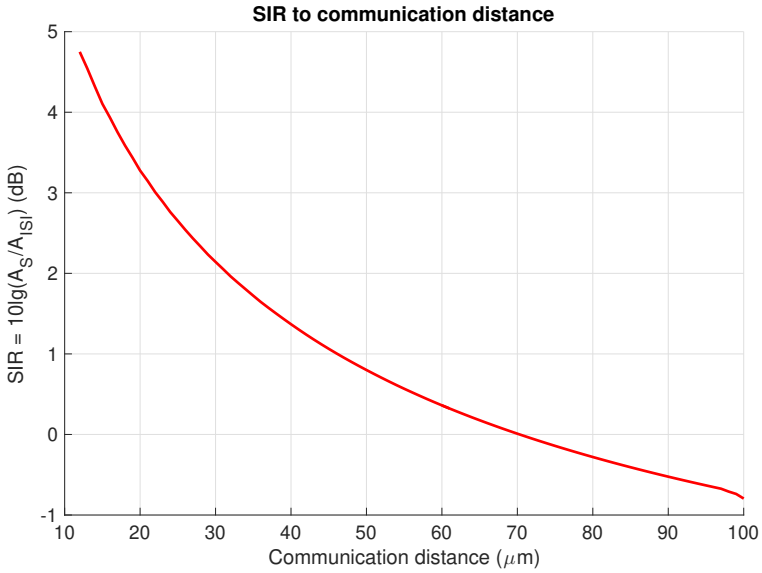

Figure 4.5: Signal-interference ratio vs. communication distance

of the received signals are intuitively shown in Fig. 4.2 to 4.4. From which we can see that as d increases, the average amplitude of the received signal becomes smaller. It means, at the same time, the SIR changes from 3.276 dB down to even -0.28 dB when $d = 80\mu m$. As d increases, there is more ISI in the received signal. The tendency of the SIR along with the increasing d can be seen from Fig. 4.5.

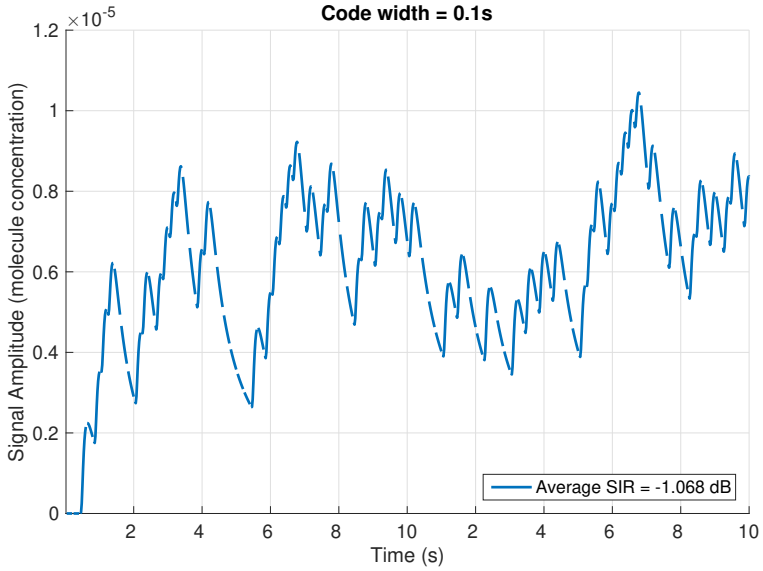

Figure 4.6: Received molecule concentration when code width $= 0.1s$

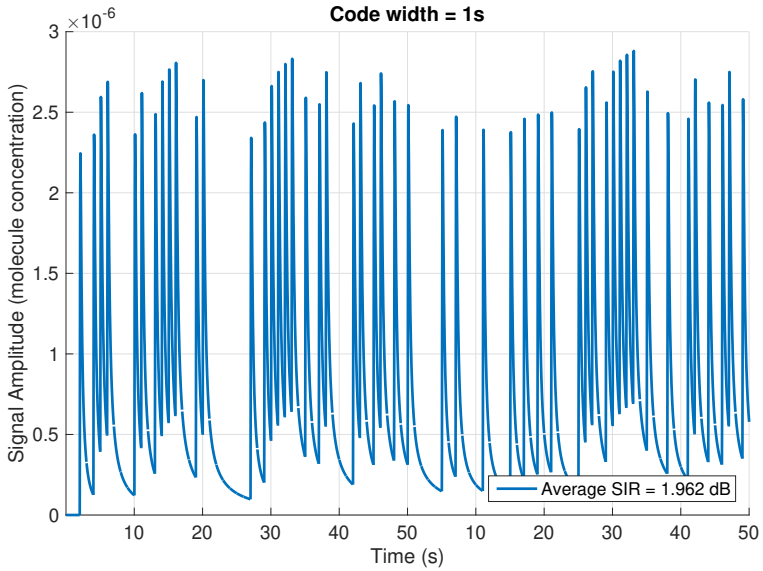

Figure 4.7: Received molecule concentration when code width $= 1s$

The time slot, namely the code width Ts is another factor that influences the magnitude of the ISI. As Ts increases, there is more time for the molecules of former bits to arrive the receiver. Thus the ISI can be smaller. The simulated received signal verifies this result, as shown in Fig. 4.6 to 4.8. When Ts is short, for example $Ts = 0.1$s, the average SIR is -1.068 dB. In other words, there is more in-

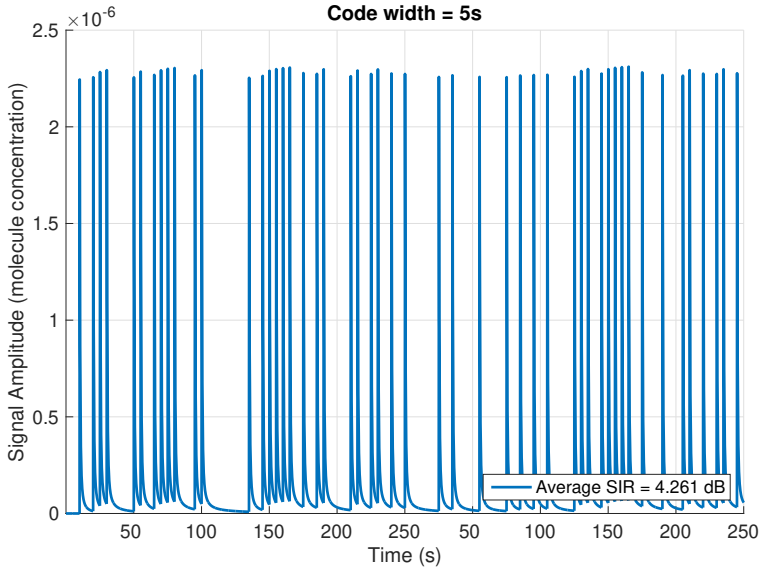

Figure 4.8: Received molecule concentration when code width $= 5s$

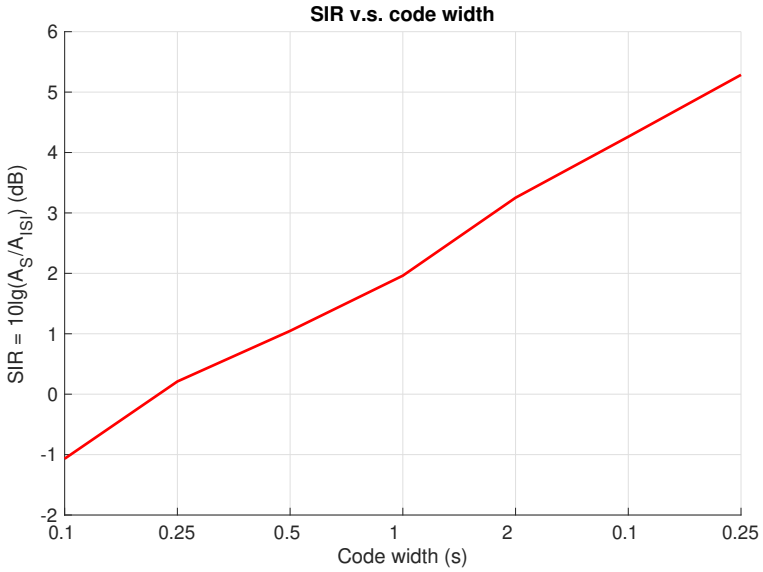

Figure 4.9: Signal-interference ratio vs. code width

terference than the intended signal in the received signal. Obviously, in such case, the BER is extremely high compared with it when $Ts = 5$s. However, the longer code width causes the lower transmission rate, namely decreases the transmission efficiency. Therefore, it is an optimization problem between the signal quality and the efficiency in the DBMC.

4.5 A Retransmission Scheme

From the above analysis, we know that the ISI is mainly influenced by the communication distance d and the time slot Ts. For the scenario of DBMC with mobile nodes that we consider, the variable d is the most crucial factor that may cause severe ISI. The BER can be significantly high due to the ISI. To reduce the effect of the error bit, we firstly propose a retransmission scheme for the DBMC with mobile nodes.

4.5.1 Retransmission Protocol Design

In the system as described in section , we consider a continuous transmission of messages. Before the transmission, the transceivers need to establish the connection. It can be easily done via transmitting a data packet contains the message of 'START' by the transmitter. The receiver replies a message of 'READY' after it receives the 'START'. Then the connection is established, and the transmitter starts the transmission of the message packet as soon as the 'READY' arrives it.

During the transmission, the signal quality varies because of the time-varying communication distance. In the proposed protocol, when the receiver discovered that the average received signal strength (RSS) becomes poorer, it transmits a 'retransmission request (RR)' acknowledgment (ACK). The 'RR' ACK can consist of only a bit '1', which is sufficient for the transmitter to notice the variation of the channel status and calculate the number of bits that need to retransmit. The time evolution diagram of the proposed retransmission protocol is as shown in Fig. 4.10.

According to the channel function (2.6), we know that for ACK of a bit '1', at

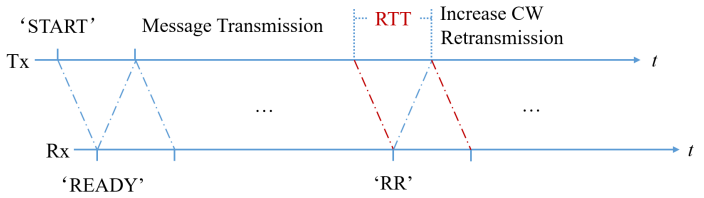

Figure 4.10: Time evolution diagram of the retransmission protocol

its RSS peak, the slope is zero. On this basis, we can calculate the trip time (TT) using (4.6), which is also the delay time of one-way signal transmission, as

$$t_{delay} = TT = \frac{d^2}{6D} \tag{4.6}$$

in which, D is the diffusion coefficient of the information molecules.

And the current communication distance can be calculated based on the results in Chapter 3 using the any protocol, for instance, the signal anttinuation method, as

$$d = \left(\frac{3}{2\pi e}\right)^{1/2} \left(\frac{Q_1}{RSS_{ACK}}\right)^{1/3} \tag{4.7}$$

or using the ATD protocol assuming the messages and two signals of ACK, are encoded on the molecules with diffusion coefficient of D_m, D_A, and D_B, as

$$d = \sqrt{\frac{6D_A D_B(t_B - t_A)}{D_A - D_B}} \tag{4.8}$$

Therefore, the round trip time (RTT), during which the transmitted bits need to retransmit, can be estimated by measuring the RSS of the ACK (RSS_{ACK}) as

$$RTT = 2t_{delay} = \frac{1}{2\pi De} \left(\frac{Q_1}{RSS_{ACK}}\right)^{2/3} \tag{4.9}$$

or the arrival time difference (t_{BA}) of the ACK signals as

$$RTT = 2t_{delay} = \frac{2D_A D_B t_{BA}}{D_m(D_B - D_A)} \tag{4.10}$$

Thus, the number of the bits to retransmit can be calculated by dividing RTT by the code width Ts as

$$n_{brt} = \frac{RTT}{Ts} = \frac{d^2}{3DTs} \tag{4.11}$$

After knowing the bits need to retransmit, the transmitter retransmits them with a doubled code width. This is because that increasing the code width can reduce the BER caused by the ISI.

4.5.2 Performance Analysis

Some simulation experiments are carried out to evaluate the influence of code width T_s on the BER to verify the feasibility of the proposed retransmission protocol. From Fig. 4.11, we can see that the BER maintains at 0 when d is relatively short. Longer than a particular distance, the BER starts to rise rapidly for all T_s. Besides, the increase of T_s can contribute to the improvement of connection quality. Therefore, in the proposed protocol, the retransmission with enlarged code width is proved feasible and practical.

4.6 Adaptive Code Width Protocol

As the results in subsection 4.4.2 indicate, the ISI is correlated with the communication distance d, and the code width Ts at the same time. Accordingly,

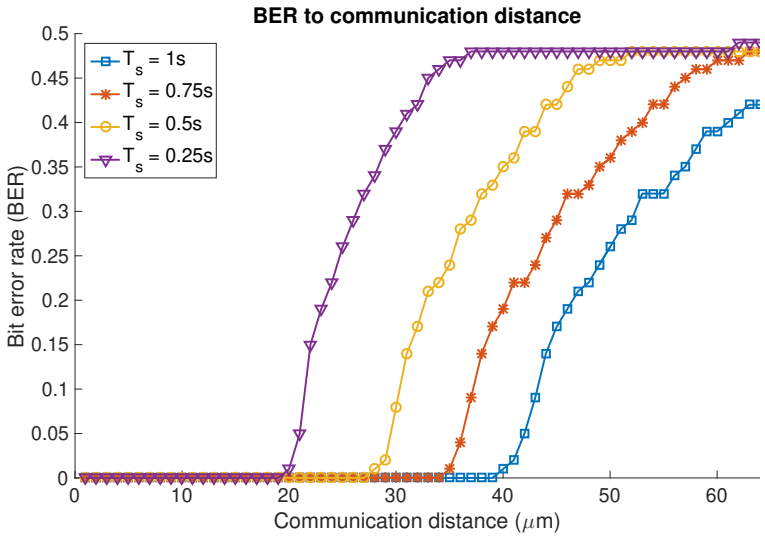

Figure 4.11: BER to communication distance d with various code with T_s

in a DBMC system with mobile nodes, the distance measurement can assist the transmission parameter settings. In other words, the time-varying communication distance information can be applied by the transmitter to optimize the transmission quality by trying to adapt the code width to mitigate the ISI. Based on this idea, we propose an adaptive code width protocol for the DBMC system with mobile nodes.

4.6.1 Protocol Design

In a DBMC system, the ISI is influenced by the communication distance d and the code width Ts. With the default parameters in TABLE 4.1, we plot the correlation of the SIR with both d and Ts, as shown in Fig. 4.12. The horizontal axis is communication distance d. The vertical axis represents the code width Ts. Different colors stand for different values of the SIR. Similar to the result in subsection 4.4.2, the SIR has a negative correlation with d and is positively influenced by Ts respectively. Thus, the SIR value decreases from left top to right

bottom in the heat map. However, there is a corresponding relation of d and Ts, with which the SIR can be maintained at a constant level. Namely, when knowing the instantaneous communication distance, the transceivers can adjust the code width used for the transmission according to a protocol to keep the SIR at a low level. Thus the BER can be controlled at a low level as well.

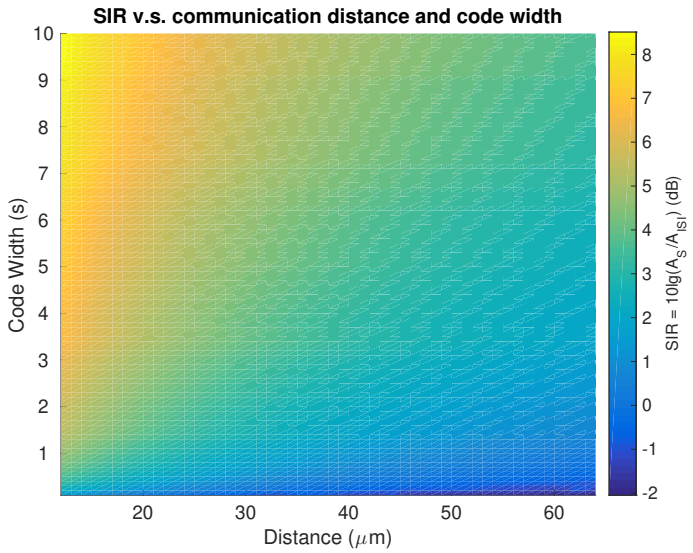

Figure 4.12: Heat map of SIR to communication distance d and code width

To this end, in the proposed adaptive code width (ACW) protocol, the Tx starts to establish the connection by sending a 'START' message to the Rx similar to section 4.5. The difference is that in ACW, the Rx transmits a bit '1' as distance feedback (DF). Based on the received signal of the DF at the Tx, the rational code width Ts_o is initialized to transmit the following packets. Meanwhile, the Rx can know the Ts_o based on the estimation of the current d with the RSS of the 'START' request. Thus, both the transceivers can start to communicate with DBMC using the Ts_o.

During the packet transmission, the communication distance d varies due to the

mobility of the transceivers. We assume the direction and velocity of the mobile nodes are both random. As a result, d can be seen as a random variable that follows a normal distribution.

After the Rx receives the 'START' request, it wakes up from the state of idle to transmitting the 'DF'. Then it starts to calculate the code width Ts_o that can optimize the transmission quality according to the analysis result in Fig. 4.12. It gets prepared to receive the packets after then. During the reception, the Rx monitors the RSS every time it receives messages. Either when the RSS is above a higher threshold $\mathrm{Thr}_{\mathrm{high}}$ or below a lower threshold $\mathrm{Thr}_{\mathrm{low}}$, the Rx transmits the 'DF' and turns to update the new Ts_o, $\mathrm{Thr}_{\mathrm{high}}$, and $\mathrm{Thr}_{\mathrm{low}}$. Firstly, it is obvious that when the RSS becomes poorer, the received signal is with higher ISI, and the BER will increase as well. The Ts used for the following transmission needs to be increased to guarantee the connection quality. On the other hand, when d becomes shorter, there will be redundant code width leading to inefficient transmission but bring no contribution to the received signal quality. The Rx then repeats the above process of changing the states among reception, transmitting 'DF', and updating parameters for the following transmission until the transmission stops. The state machine diagram of Rx is as shown in Fig. 4.13.

As for the Tx, it is similar to the Rx. After the transmission is started, it waits for the 'DF' continuously. When it receives the 'DF', which means that the channel condition has changed, the Tx calculates the new Ts_o based on the estimation of the current communication distance. The calculation of the new Ts_o is also according to the pre-defined corresponding relation of the Ts and d shown in Fig. 4.12. After then, the Tx starts to transmit the following messages with the

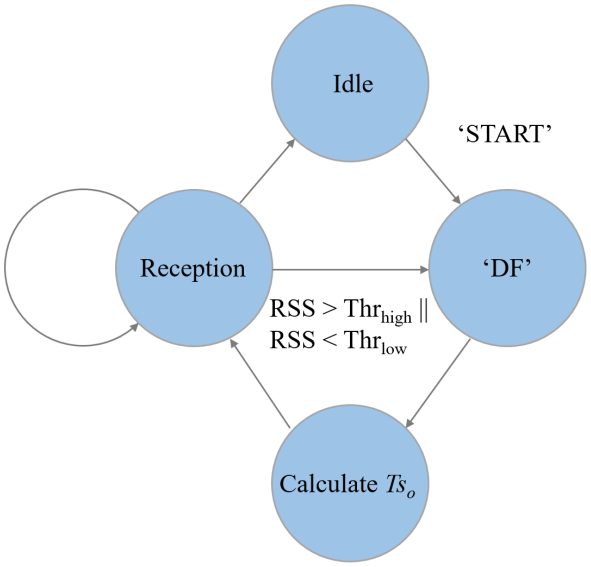

Figure 4.13: State machine diagram of Rx

updated Ts. And as the Rx also updates the Ts, the transmission is not influenced by the change of Ts.

Consequently, not only the connection quality but also the transmission efficiency is optimized by the proposed ACW protocol. The temporal evolution diagram of the whole transmission process is as shown in Fig. 4.15.

4.6.2 Performance Analysis

Some simulation experiments are carried out in this section to verify the feasibility of ISI mitigation with the proposed protocol. The parameters are set according to TABLE 4.1. We adopt two patterns of the communication distance variation. In one pattern, the distance tends to increase. The results are shown in the in Fig. 4.16 and 4.17. In the other pattern, the distance tends to decrease, as shown in Fig. 4.18 and 4.19. For both patterns, the SIR of proposed ACW protocol can be controlled higher than 3.5 dB. In another word, however the distance varies, the ISI is

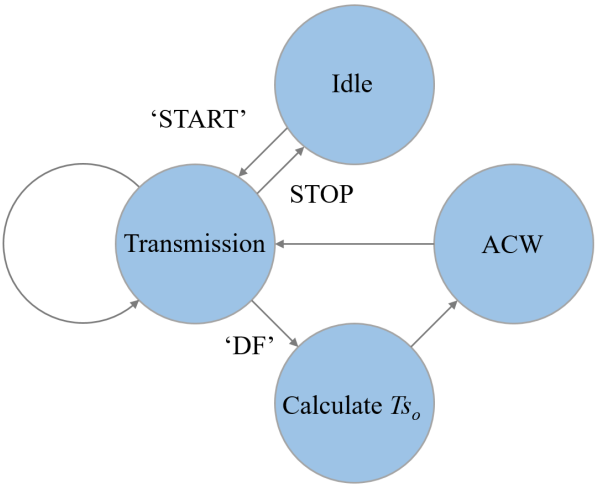

Figure 4.14: State machine diagram of Tx

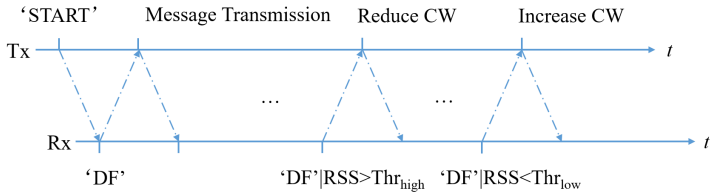

Figure 4.15: Time evolution diagram of adaptive code width protocol

less than 1/33 of the signal amplitude according to the SIR definition (4.5) in this paper. For comparison, we also do simulations of a fixed code width approach. In this approach, the code width is fixed to a value which is only determined based on the initial communication distance. The ISI of this method varies notably along with the communication distance. Moreover, the ISI decreases to 1.78 dB when the distance reaches 48 μm. On the other hand, when the distance decreases, the fixed code width approach has a better SIR performance than the ACW protocol. However, such performance does not contribute to the improvement of connection quality. And it is gained on the cost of the transmission efficiency.

Considering the transmission efficiency, the proposed ACW protocol does not

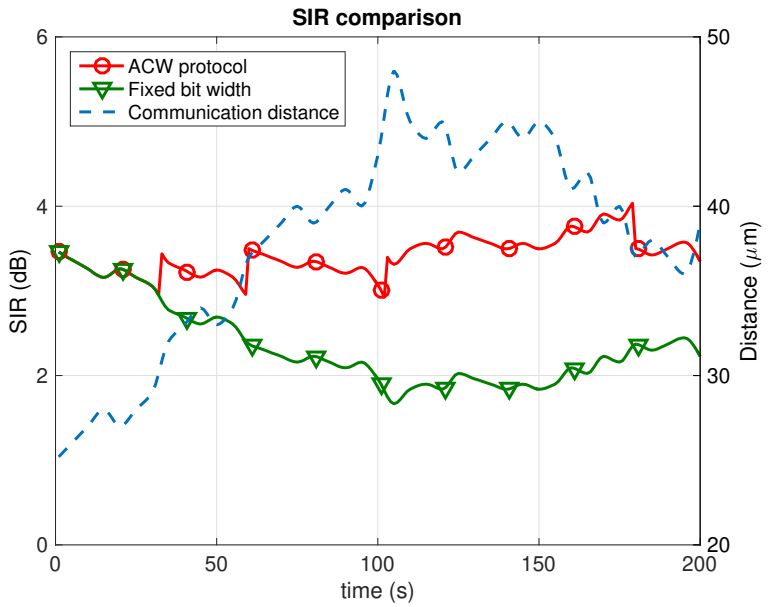

Figure 4.16: SIR comparison when distance increases

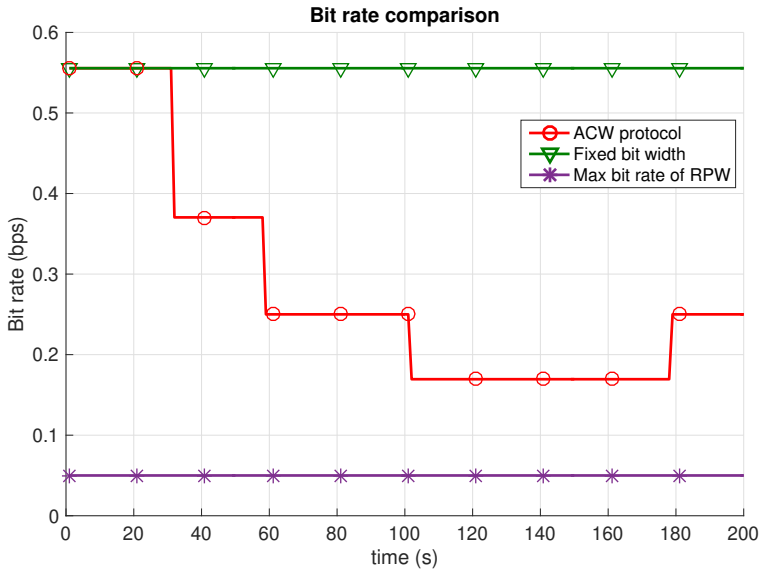

Figure 4.17: Bit rate comparison when distance increases

blindly increase code width for mitigating ISI. Especially, when the communication distance is short, the ISI is within an acceptable range regardless of the code width, according to Fig. 4.12. Hence, to optimize bit rate, the code width is set to a smaller value if the distance is measured to be short. From Fig. 4.18 and 4.19,

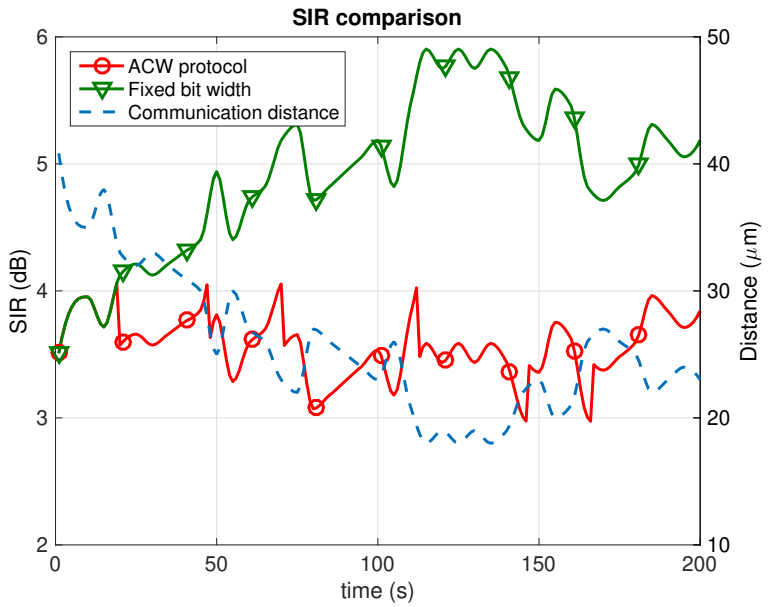

Figure 4.18: SIR comparison when distance decreases

we know that the ACW protocol controls the SIR higher than 2.91 dB. Moreover, the bit rate with ACW protocol is much higher than the fixed code width method because of using the optimal code widths. In addition, by using the proposed modulation model (4.1) and (4.2), the code width is much higher than the values presented in [50], which are 0.0025 Hz, 0.01 Hz and 0.05 Hz. Namely, bit rate with the ACW protocol is reasonably improved.

Consequently, the proposed ACW protocol can mitigate ISI in diffusion-based MC with mobile nodes and give attention to bit rate at the same time. When the communication distance is long, ISI is the main problem in communication performance. The ACW mitigate ISI based on the *'distance feedback'*. On the other hand, when the communication distance is short, the ISI becomes less serious. The bit rate is optimized by the proposed adaptive code width protocol.

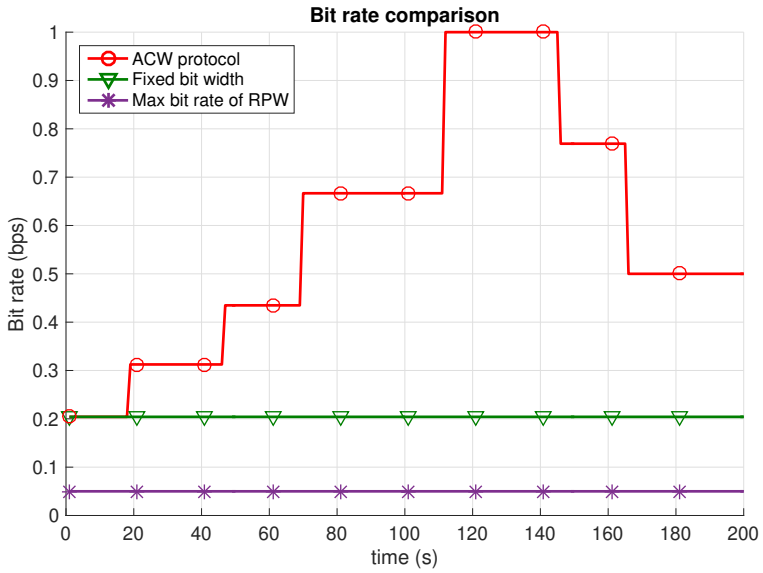

Figure 4.19: Bit rate comparison when distance decreases

4.7 Summary

In this chapter, we analyse the impacts on ISI in diffusion-based MC. We find out that ISI increases as communication distance increases or as code width decreases. Accordingly, we design an ACW protocol for diffusion-based MC with mobile nodes. In this protocol, a *'distance feedback'* is transmitted from Rx to Tx when the channel varies. The communication distance is measured using the signal attenuation model. Then the code width is adapted according to the measured distance. Through simulation, this ACW protocol is proved feasible to mitigate ISI regardless of the communication distance variation. It also guarantees the bit rate when communication distance is short and ISI is not serious.

Chapter 5

Diffusive DNA based Molecular Communication

In this chapter, we propose a novel molecular communication in which the DNA is used as the information molecule. In particular, we design a scheme utilizing the two types of nucleobase pairs to represent binary bits so that a bit sequence, i.e., a data packet, can be encoded onto a single DNA molecule. Because the stochastic character of the diffusion process and the variable diffusion coefficient of DNA molecules correlated to the size of the molecules, namely the number of base pairs composing the DNA fragment, the uncertainty of arrival order is a crucial factor that influences the reliable transmission. By modeling the observation window and focus slot, we concentrate on the arriving order of the DNA packets. Based on the probability of each state, we derive the channel capacity of the proposed diffusive DNA based molecular communication. We also testify the effect of some parameters including base pair number, communication distance, and time slot width, on the channel capacity. The analysis shows that due to the high-density information encoded onto the DNA molecule, the proposed diffusive DNA based molecular communication has a higher channel capacity than other diffusion-based molecular communication. Additionally, to solve the dis-

order arrival problem, we present a buffer aided method. The transmitter emits multiple duplicated DNA packets in each slot. The buffer in the receiver can temporarily store DNA molecules within a time slot. It compares the number of each packet and determines the one of the largest quantity as the packet that should be received in this slot. The simulation results show that this method can improve the successful transmission comparing with the order-based method. drrddfdffdfg

5.1 Motivation

The concentration shift keying and the time-based channel are currently the most widely studied modulation schemes of diffusion-based molecular communication because of their spontaneity, energy efficiency, and the low requirement of the nano-machine capability. Therefore, speaking of diffusion-based molecular communication, it generally refers to the above two situations. However, the studies also show us that the channel capacity of such systems is limited, for instance can be as low as 0.0008 bit according to [52], because the diffusion is a stochastic process with a long delay compared with electromagnetic wave based conventional wireless communication techniques. The effective transmission rate of diffusion-based molecular makes it inefficient to transmit extensive information. Besides the noises in the channel such as the remaining molecules of other slots can cause severe inter-symbol interference (ISI). Even though solutions to mitigate the ISI have been studied. Generally, it is still a crucial problem regarding the bit error rate and results in poor transmission quality.

On the other hand, DNA is capable of storing extensive information on one single molecule. There are four types of nucleobases including adenine (A), thymine

(T), cytosine (C), guanine (G) to constitute the DNA molecules. Each two of them are bounded with another. Namely, the four types of nucleobases appear in the DNA molecules in the form of base pairs. With the different combinations, which can be done through the artificial assembling of DNA molecules [53], a simple scheme using different base pairs to represent the binary bits can be applied to encode information onto the DNA molecules [54]. DNA can contain plenty of nucleobases in one single molecule. Thus data packets rather than simple bits can be encoded onto the DNA molecules. Besides, the DNA molecules are capable of diffusing within a small scope. Thus it is feasible to use DNA molecules to transmit the information via diffusion through the medium. Thanks to the high-density information that can be stored in DNA, it is possible to remarkably increase the potential transmission rate of diffusion-based MC with DNA molecules. Consequently, the diffusive DNA based molecular communication can be a promising solution for high channel capacity DBMC.

The possible high channel capacity and the possibility to combine with other DNA based applications such as the DNA computing [55], inspire us to investigate the diffusive DNA based molecular communication for bio-nano-machines from the information theory point of view.

5.2 Related Work

To design a technique using DNA molecules for the DBMC, we investigate several related works including the solutions to the information storage on DNA molecules and the DNA molecule diffusion in the potential medium of DBMC for bio-nano-machines.

First of all, the channel capacity is one of the most widely studied topics in the field of DBMC. The channel capacity, namely the maximum transmission rate is derived in [56] considering a ligand-receiver for the CSK-based DBMC channel. A 4-input-2-output discrete channel model rather than the binary symmetric channel is used to approximate the DBMC channel in [57]. It also shows that the theoretical channel capacity can be achieved by adjusting the frequency of binary zero bits higher than the binary ones. In [58], the diffusion process is decomposed into two separate processes, the Fick‘s diffusion, and the particle location displacement. The number of individual molecules emitted into the medium is novelly modeled as the bandwidth. And then, the channel capacity is comprehensively derived considering various parameters. [47] considers the inter-symbol interference for the channel capacity analysis. The optimal threshold value is derived from the maximum-a-posterior (MAP) by using the Skellam distribution. Higher capacity can be achieved with fewer information molecules with this model compared with other CSK-based DBMC.

The information storage in DNA molecules has been considered as a solution for massive information storage for a long time. [59] investigated the possibility to store data such as text with DNA molecules for a long term by encoding the English alphabet with DNA codons. The simple bit per base method is adopted, and an extremely high information storage density is achieved by the work presented in [60]. Massive Shannon information can be stored into and reconstructed from a synthesized DNA [61]. The reconstruction accuracy of this system can be 100%. [30] incorporated the DNA cellular storage into the MC, analyzed and compared the achievable transmission rate and the bit error rate with other modulation

schemes of MC such as CSK and MoSK.

The DNA molecule diffusion has been studied through practical experiments. The experiments in [62] show that the diffusion coefficient is approximately negative linear correlated with the size (the chain length in micrometer) of the DNA molecules. [63] derives the correlation of the diffusion coefficient of DNA molecules and the number of base pairs in water and cytoplasm. Another experimental study in [64] shows that the shape of the diffusion space also influences the diffusion coefficient of DNA molecules. [65] also verifies that the diffusion coefficient correlates negatively with the number of base pairs in mucus gel. Additionally, it takes the topology of the DNA molecules into account as well.

5.3 Diffusive DNA based Channel

In this section, we present the system of diffusive DNA based MC, and the analysis of the possible max transmission rate, namely the channel capacity.

5.3.1 System Model

We consider a simple system that consists of a transmitter (Tx) and a receiver (Rx). They are deployed in a one-dimensional environment using water as the medium simulating a part of a Lab-on-a-chip system neglecting the diameter of the tube channel. The Tx is set up at the origin point. It is capable of assembling the four types nitrogen-containing nucleobases of A, T, C, G into DNA fragments according to the given instructions. It can also emit the DNA molecules one by one with a particular time interval. Namely, in this section, we assume there are no duplicated packets, i.e., only one DNA molecule appears in a specific time

slot. The Rx locates at a point that is d apart from the transmitter and is capable of decoding the information from the DNA molecules. The system is as demonstrated in Fig.5.1.

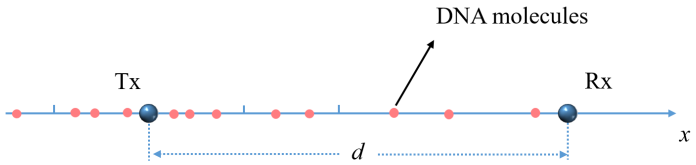

Figure 5.1: A demonstration of a diffusive DNA based MC system

A. DNA-based encoding scheme

There are four types of nucleobases. Each of them pairs with another one. Thus, regardless of the left-right direction, we can only find two kinds of nucleobase pairs in DNA molecules, represented as A-T base pairs and C-G base pairs. Take this into account, in this paper we utilize a simple encoding scheme similar to [30, 60], using different types of the base pairs to represent binary bits '0's and '1's. We assume the A-T pairs represent bits'1's, and the G-C pairs as '0's. Then by assembling nucleobase pairs connected one by one according to a binary bit sequence to be transmitted, we can encode this sequence to a DNA molecule-based packet. For simplicity, we only consider fixed-length bit sequences, namely the number of nucleobase pairs in each DNA packets is the same. Fig. 5.2 shows a demonstration of an encoded DNA packet with 11 nucleobase pairs, namely 11 binary bits.

Similar to all other data storage systems, the errors are unavoidable when using DNA as the data storage medium. The error bit can be caused by mistaken assembling and mutation. The mistaken assembling happens within the transmitter when

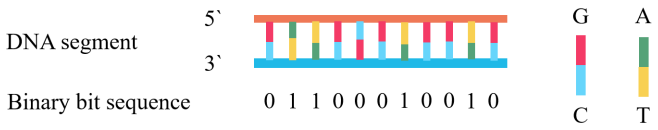

Figure 5.2: An example of using a DNA segment to represent a binary bit sequence

it encodes the binary bit sequence into a DNA-based packet. On the other hand, the mutation is a phenomenon that nucleobase changes to another type during its lifetime. It happens when there are mutagens in the environment such as oxidizing agent, strong alkali and high-energy electromagnetic radiation. Both the mistaken assembling and mutation are the cause of the error bits in the DNA-based packets. We only take into account the point mutation, namely the error occurs to a single bit independently, it is reasonable to consider the mistaken assembling and mutation as the same. The analysis of the transmission error caused by the mutation is given in the later subsection of 5.3.3.

B. Diffusion coefficient of DNA molecule

In the proposed diffusive DNA based MC system, DNA molecules carrying the information encoded with them to the receiver through free diffusion. According to the diffusion equation in (2.6), the arrival probability of a diffusing molecule through time depends on the traveling distance d and diffusion coefficient D. For a particular system with no mobile nodes, the communication distance is determined. Thus the arrival probability only relies on the diffusion coefficient D. In a liquid medium, the diffusion coefficient of DNA fragment is positively correlated to the size of the molecule, in other words, how many nucleobases in the DNA molecule. We apply the model in [63] to our system, which shows that the DNA diffusion coefficient in water has a relationship with the number of nucleobase

pairs in the DNA fragment as

$$D_n = 490\mu m^2 \cdot [n(bp)]^{-0.72} \tag{5.1}$$

in which, n is the number of nucleobase pairs in the DNA fragment molecule, and D_n is the diffusion coefficient of a DNA molecule with n base pairs in water. From this equation, we know that as the number of base pairs increases, the diffusion coefficient rises as well, and then the arrival probability to a particular point by a specific time decreases according to the diffusion equation in (2.6). Thus, there is a trade-off problem between the information that one DNA molecule can contain and the probability of successful arrival at the receiver.

5.3.2 Channel Capacity

In this section, we derive the mutual information of the proposed diffusive DNA based MC. To this end, we first decompose the whole information transmission from the source to the destination as three cascaded sub-systems according to the physical processes. The system is separately modeled into the source, released molecule, received molecule, and destination. The physical processes include encoding, diffusion, and transcription respectively. The noises that may occur during each of the physical processes are considered caused by the mutation error, diffusion uncertainty, and transcription error respectively. The information transmission process as a whole is as illustrated in Fig. 5.3.

In the model diagram, S, X, Y, D are used to denote the source, DNA molecules released and received by the transmitter and receiver, and the destination respectively. Based on this separation, the factors that can influence the channel capacity

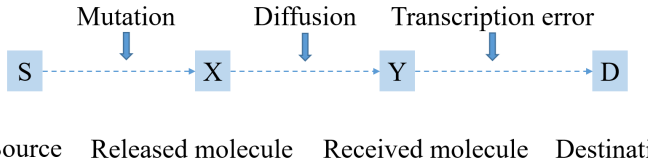

Figure 5.3: Diagram of transmission process in the diffusive DNA based MC system

during the information transmission are modeled as mutation during encoding, noise in the diffusion channel, and the transcription error. As the transmission is decomposed into cascaded sub-systems, we discuss the influence individually in the following sub-sections.

5.3.3 Encoding with mutation

The mutation in an unavoidable problem that causes error bits while applying the DNA-based packet using the encoding scheme proposed above, i.e., two types of nucleobase pairs are used to represent different binary bits. There are various types of mutation in the DNA molecules in nature including substitution, deletion, and insertion, etc. The mutation rate of different species varies substantially. Due to the potential limit capabilities of the Nano-machines, we tested the influence of the mutation to the mutual information of the sources and encoded information in the transmitted DNA molecules consists of 2, 4, 6, 8, 10 base pairs. For simplicity, we only consider the point mutation, namely, the substitution. The mutual information of the information source and the DNA molecule that is released by the transmitter is calculated with the basic formula as follows

$$I(S,X) = \sum_{s\in \mathrm{ALL}} \sum_{x\in \mathrm{ALL}} p_{(s,x)} \log \frac{p_{(s,x)}}{p_s p_x} \tag{5.2}$$

where s denotes the potential binary sequence contained in the source packets, meanwhile x is used to represent the sequence encoded on the released molecules, considering the mutation phenomenon.

The numerical results are as shown in Fig. 5.4. We can see that the mutual information is slightly affected by the mutation, even when the tested mutation rate increases to 0.25, which is much larger than the spontaneous mutation rate in nature. Moreover, the effect becomes weaker as the base pair number increases. Consequently, the mutation of DNA has a minimal influence on the mutual information between the encoding sources and assembled DNA segment packets.

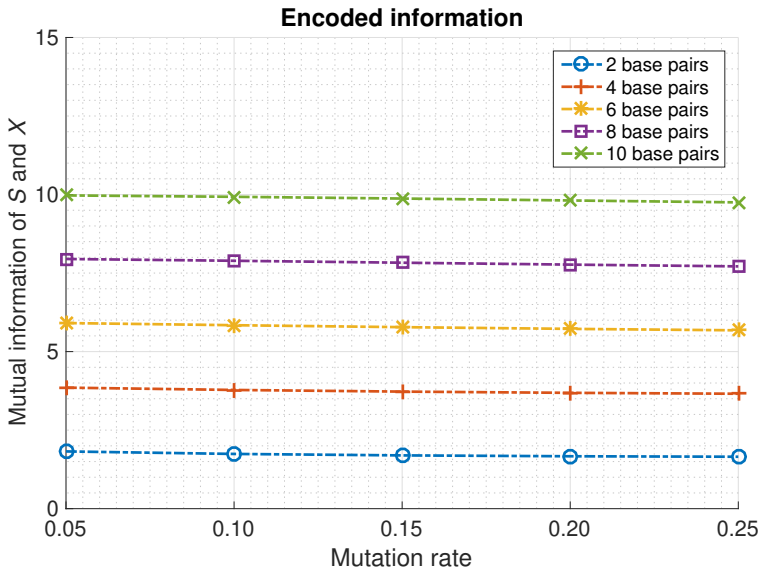

Figure 5.4: Mutual information of source and encoded information

5.3.4 Observing window and arrival order

The crucial factor that impacts the channel capacity in diffusive DNA based MC is whether the arrival order of the DNA molecules corresponding to the re-

leased order. Therefore to analyze the channel capacity, we first define the observing windows and focus slot, based on which if the arrival order is correct can be simply evaluated. From the diffusion equation (2.6) we know that the arrival probability becomes smaller and negligible as the time elapses. Thus we consider a limited number of observing windows, as illustrated in Fig. 5.5. Besides, the focus slot is used to measure if the molecule received by the receiver is in the correct slot. Taking 'Focus 2nd, 3 Windows (F=2, W=3') for example, we observe the order of the arrival molecules received by the receiver in three consecutive slots and mainly focus on checking which molecule is received in the second slot. Only if the secondly received molecule is the one released by the transmitter in the second slot, the packet is transmitted successfully. Otherwise, the disordered arrival DNA packets are considered as error packets. Because considering the order of a packet sequence, there are chances that the later transmitted molecule arrives before the former ones, we also take into account the situation that the focus slot is not the last one when designing the observing windows. Based on this assumption, the mutual information can be analyzed.

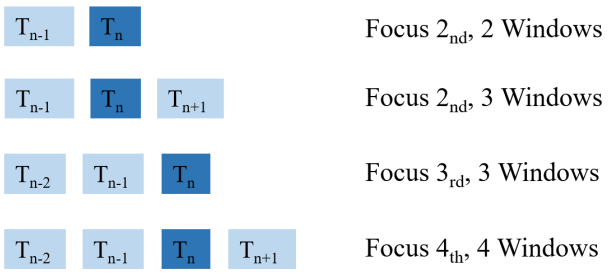

Figure 5.5: Demonstration of observing window and focus slot

Then we calculate the probability of each arrival states. The states here refer to the different orders of the received DNA packets in all observing windows.

Under the assumption that all DNA molecules will arrive the receiver through time, the number of states when considering n observation windows is $n!$. Taking $n = 2$ as an example, the received molecule sequence $Ms_a = \{M1, M2\}$ equals that the arrival time of the two molecules has the relation $ta_{M1} < ta_{M2}$. The random variable of arrival time t_a has a probability density function (2.6). Thus we can use the following equation to calculate the probability of the arrival sequence $Ms_a = \{M1, M2\}$.

$$p(t_{m1} < t_{m2}) = \int_0^{\infty} \int_0^{t_{m2}} f(t_{m1}) f(t_{m2}) dt_{m1} dt_{m2} \tag{5.3}$$

The probability of the four arrival states of two consecutive molecules is as shown in Fig. 5.6. $M1, M2$ are used to denote the molecules released in the first and second slot. The M with the overline means the state that this molecule has not arrived.

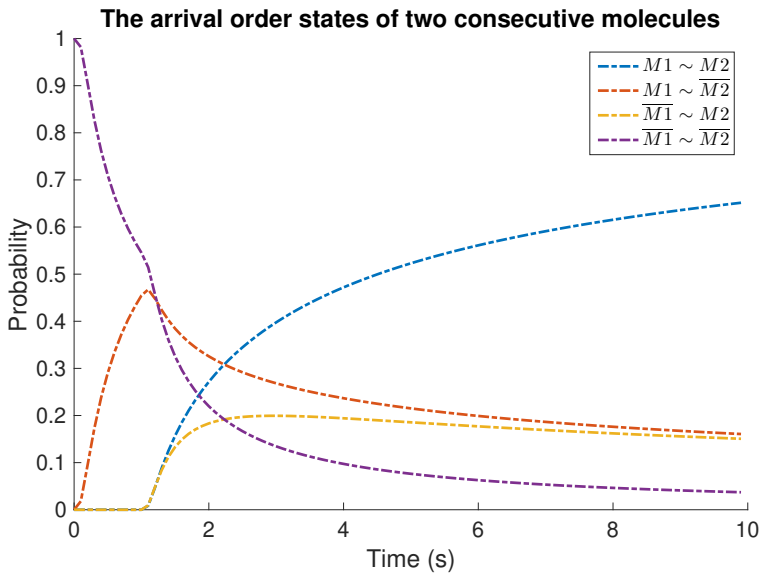

Figure 5.6: Arrival states of two consecutive DNA packet molecules

Similarly, we can readily obtain the probability of the molecule arrival orders

for three or four observation windows with multiple time integrals.

Based on the above derivation of state probability, we can calculate the channel capacity, which is the maximum value of the mutual information of the channel input and output. Here, the channel input is considered as a specific order of the released DNA packets. Meanwhile, the channel output is the order of the received packets. The mutual information can be calculated using the formula below.

$$\begin{aligned} C &= \max I(X;Y) \\ &= \max \sum_{x \in ALL} \sum_{y \in ALL} p(y\,|x)p(x) log \frac{p(y\,|x)}{p(y)} \end{aligned} \tag{5.4}$$

in which, the x and y are the transmitted packets and received packets, taking the arrival order into account.

5.3.5 Transcription error

Table 5.1: Transcription error in DNA molecules

	Read as			
	A	T	C	G
A	95%	2%	2%	1%
T	2%	92%	4%	2%
C	2%	4%	92%	2%
G	2%	4%	2%	92%

The final problem that needs to be noted of the diffusive DNA based MC is the transcription error, namely the variation between received DNA packets and the translated binary sequences. According to [66], the Nanopore-sequencer system has transcription errors for each type of the nucleobase. The transcription error rate of this system can be found in TABLE 5.1. It is reasonable to think that

the transcription performance of a Nano-machine is not better than the Nanopore-sequencer system, i.e., the error happens with at least similar probabilities. Consequently, the coefficient in the channel capacity of the last cascade sub-system, namely between the received packets and transcribed packets can be easily calculated with

$$\alpha = I(Y, D) = 1 - \sum_{i \in \text{ALL}} -p_{te\_i} \log_2 p_{te\_i} \tag{5.5}$$

where p_{te_i} refers to each transcription error probability in the TABLE 5.1.

5.3.6 Performance Analysis

Based on the above analysis, we carry out some simulation experiments in Matlab to obtain the numerical results of the channel capacity. The simulation is set up according to the models described in the former subsections. We evaluate the influence of the number of base pairs, communication distance, and time slot on the channel capacity respectively. The default parameters are set as follows, the maximum transmission time $t_{Max} = 1000$s, simulation step time $t_{Step} = 0.1$s, default communication distance $d = 10\mu$m, default base pair number $n = 100$, default time slot $t_{Slot} = 1$s.

Fig. 5.7 evaluates the influence of the number of based pairs. We can see that the channel capacity increases as the number of base pairs rise from 2^1 to 2^{13}. It is because that the diffusion coefficient of a DNA molecule becomes smaller as the number of base pairs increases. This phenomenon results in a slower diffusion and relatively lower arrival probability of a larger DNA molecule packet within the same period. However, the more massive information amount encoded on the

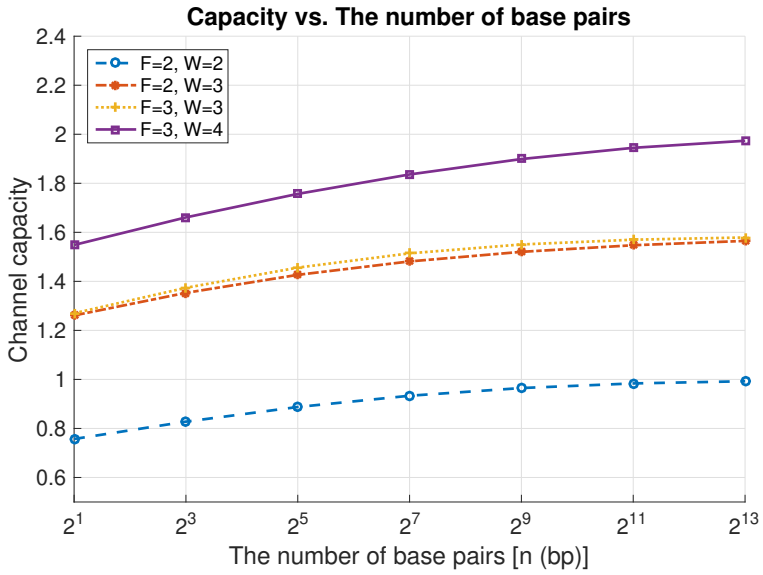

Figure 5.7: Channel capacity vs. the number of base pairs

molecule dominants when the number of base pair increases. Among the four types of observation schemes, as expected when observing with more windows, the channel has a higher capacity. Meanwhile, when the number of observation windows is the same, the different focus slots affect the channel capacity very slightly.

The simulation result of channel capacity against communication distance is shown in Fig. 5.8. Similar to the above analysis, the observation scheme with more windows has a higher channel capacity. Besides, the channel capacity decreases as the communication distance increases from 2μm to 32μm. For all observation schemes when the communication distance is shorter than 2μm, the channel capacity keeps at the same value, namely the upper bound. After that point, it starts to drop. This tendency is because when the time slot is fixed, and the communication distance is short, the DNA molecules are more likely to arrive at the receiver in the correct order so that there is not much uncertainty, i.e., the noise in the chan-

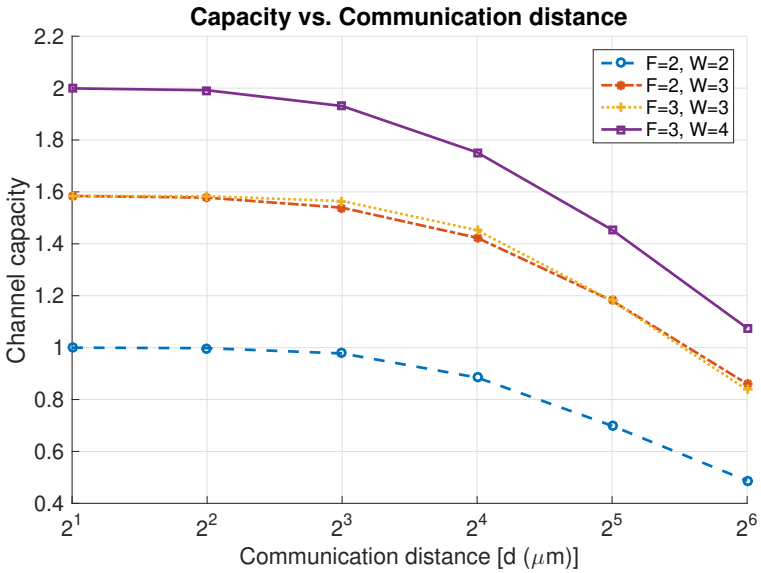

Figure 5.8: Channel capacity vs. communication distance

nel. However, when the communication distance reaches a particular value, the diffusion channel noises start to dominate the transmission quality. As a result, the achievable transmission rate without error decreases. This result also verifies that the conclusion that the communication distance is a crucial parameter of DBMC.

We finally evaluate the time slot width t_{Slot}, namely the time interval between two consecutive DNA molecules released by the transmitter. Fig. 5.9 shows that the channel capacity and time slot have a negative correlation. The channel capacity decreases as the time slot width is increased. Increasing the time slot can contribute to higher probability of correct arrival order of the DNA packets. However, on the other hand, the transmission efficiency is lower because the packets that can be transmitted per unit time are less. And the figure shows that in this case, the later result, namely the influence on the transmission rate, plays a more critical role.

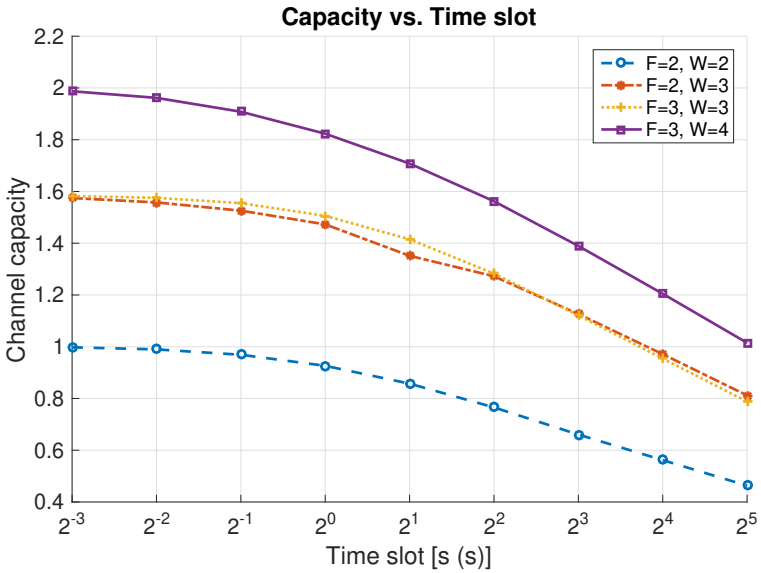

Figure 5.9: Channel capacity vs. time slot width

5.4 Buffer Aided Diffusive DNA based Molecular Communication

The usage of DNA molecules as data packets can increase the information density on a single information molecule in the DBMC. However, there is still an unavoidable problem, which is the disordered arrival packet in the diffusive DNA based MC. To solve this problem, we propose a multiple duplicated packets and buffer-aided (MDP&BA) method. In this method, several duplicated DNA packets are used to increase the correct arrival probability. A buffer can store the packets that the receiver receives in a slot, and compare the number of different packets to determine which is the one released in the corresponding slot.

5.4.1 Methodology

In a similar system as described in section 5.3.1, two nano-machine nodes located at a distance of d_c are considered in a 1-D environment. They communicate

with the diffusive DNA based MC. The main difference from the former section is that in this section, the transmitter releases multiple duplicated DNA molecules in a slot. The process is as follows. When the data packet of interest is fixed, the transmitter encodes the information onto the DNA molecule and replicates the DNA packet to a specific number. Then it releases these duplicated DNA packets into the medium simultaneously at the beginning of the time slot. For the following data packets, the transmitter repeats the above process. Consequently, because the motion of the molecules in the medium is independent of each other, for a particular slot, the arrival probability of each duplicated DNA molecules is independent and identically distributed. Because limited duplicated DNA molecules are used for the sake of reducing the base pair cost, the number of molecules that arrive at the receiver in the corresponding slot can be seen follows a Binomial distribution,

$$Pr(k; n, F(d, Ts)) \sim B(n, F(d, Ts)) \tag{5.6}$$

where k is the number of molecules arrive in the correct slot, n is the number of molecules released by the transmitter, i.e., the number of duplicated packets, Ts is the time slot width, and F denotes the cumulative distribution function (CDF) of the arrival probability of each single molecule in the diffusion-based channel. According to the diffusion function (2.6), the arrival probability of a molecule has the probability density function (PDF),

$$f(d, t) = \frac{1}{\sqrt{4\pi D t^3}} e^{\frac{-d^2}{4Dt}} \tag{5.7}$$

Its CDF denotes the accumulated probability of the arrival happens within the period of $(0, t]$, and can be derived as

$$F(d,t) = \text{erfc}(\frac{d}{2\sqrt{Dt}}) \tag{5.8}$$

where the erfc$(\cdot)$is the complementary error function.

By applying multiple duplicated DNA molecules in one slot, the received packet sequence is no longer relying on the arrival order of each molecule as it is in the system in section 5.3. However, using multiple duplicated packets increases not only the probability of the successful transmission but also the probability that the receiver receives several packets from different slots. It leads to the confusion of the receiver on the judge which of the received packet is the correct one in the current slot. To this problem, we apply a buffer in the receiver to temporarily store the received packets for the period of the current slot. Besides, it is also capable of selecting the packet of the largest amount as the one that it should receive in this slot. In other words, the DNA packet received in the ith slot, $Pa_i$is determined as

$$Pa_i = M_{i,j}, Nm_{i,j} = Max\{Nm \in ALL\} \tag{5.9}$$

where $Nm_{i,j}$ denotes the number of DNA molecules transmitted in the jth slot, and received in the ith slot.

The figure 5.10 shows the demonstration of the transmission of the multiple duplicated DNA packets and the packets received in a slot, in the proposed MDP&BA method.

The probability that the receiver receives the correct packet, namely the DNA molecules emitted by the transmitter in the corresponding slot, can be derived as

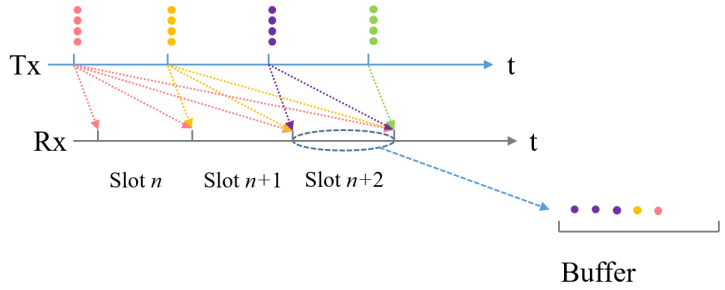

Figure 5.10: Demonstration of MDP&BA method

follows. The arrival probability of each molecule decreases dramatically after its slot, as shown in Fig. 2.7. Therefore the arrival probability of DNA molecules from several slots ago can be ignored. We take it as an example that the current slot is the ith slot, and only two former slots are considered. The probability of that the packet Pa_i is determined as the DNA molecules emitted in the ith slot by comparing the number of each received packets, namely a successful transmission in the ith slot is

$$\begin{aligned} &\Pr(Pa_i = M_{e,i}) \\ &= \sum_{j=1}^{BS-k-l} \sum_{k=0}^{j} \sum_{l=0}^{j} \Pr(N_{r,i} = j) \Pr(N_{r,i_{-1}} = k) \Pr(N_{r,i-2} = l) \end{aligned} \tag{5.10}$$

where $M_{e,i}$ is the DNA molecule emitted in the ith slot, $N_{r,i}$, $N_{r,i-1}$, $N_{r,i-2}$ are the numbers of received molecules that are emitted in each slots, BS is the buffer size. And according to (5.6), the probability of j DNA packets are received by the receiver in the ith slot can be calculated with

$$\Pr(N_{r,i} = j; n, F_i) = \binom{n}{j} F_i^{\,j} (1 - F_i)^{n-j} \tag{5.11}$$

where the F_i is the CDF of the ith slot.

5.4.2 Performance Analysis

Table 5.2: Default simulation parameters for buffer-aided diffusive-DNA-based MC

Symbol	Description	Value
NBP	Number of base pairs (packet length)	100
d	Communication distance	5μm
Ts	Time slot width	$1s$
BS	Buffer size	10
NP	Number of duplicated DNA molecules (packets)	10

Some simulation experiments in the MATLAB are carried out to evaluate the performance of the proposed MDP&BA method for diffusive DNA based MC. We take the practical packet error ratio (PER) as the criterion to evaluate the transmission quality and compare with the method without the buffer and only depending on the arrival order of the DNA packets (order-based method) presented in the section 5.3. The PER is the ratio of error packets to all the packets that are transmitted, is approximated to the probability of the failed transmission, can be calculated with

$$PER = \frac{P_E}{P_{ALL}} \doteq \Pr(P_E \,|\theta) = 1 - \Pr(Pa_i = M_{e,i}) \tag{5.12}$$

in which, $\theta = \{NBP, d, ts, BS, NP\}$ is the set of the parameters of the system including the number of base pairs (NBP), communication distance (d), time slot width (ts), buffer size (BS), and the number of duplicated packets (NP).

The default parameters that are used in the simulation are set as follows. The packet length NBP is assumed as fixed to 100 base pairs, communication distance $d = 5\mu$m, time slot width $ts = 1s$, buffer size and number of duplicated packets are both 10. The parameters are as summarized in TABLE 5.2.

We first evaluate the effect of the packet length on the PER. First of all, accord-

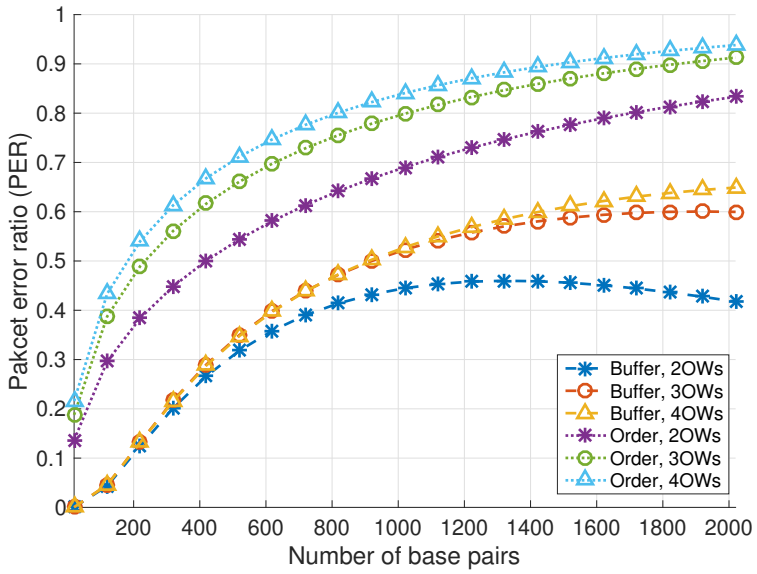

Figure 5.11: Packet error rate vs. number of base pairs

ing to (5.1), we know that the diffusion coefficient of DNA molecules is correlated with the number of base pairs in it. As shown in Fig. 5.11, the MDP&BA has smaller PER comparing with the order-based method for any observing schemes. We can also see that the PER increases as the number of base pairs become more extensive for the order-based method. Meanwhile for the MDP&BA, when there are not too much base pairs in the DNA molecule packets, for example, the number is smaller than 1200 for four observing windows, the PER also increases along with the number of base pairs. However, when it is sufficiently large, the PER turns to decrease slightly. But the similar tendency does not appear in the order-based method. This phenomenon is because when the number of base pairs is large, the diffusion coefficient is small so that the probability of each molecule to arrive successfully is small as well. It results in the uncertainty of the arrival order of the transmitted packets. On the other hand, the buffer helps to reduce the influence of error packets that arrive at the receiver earlier than the correct packets. Therefore,

the MDP&BA has a better performance than the order-based method.

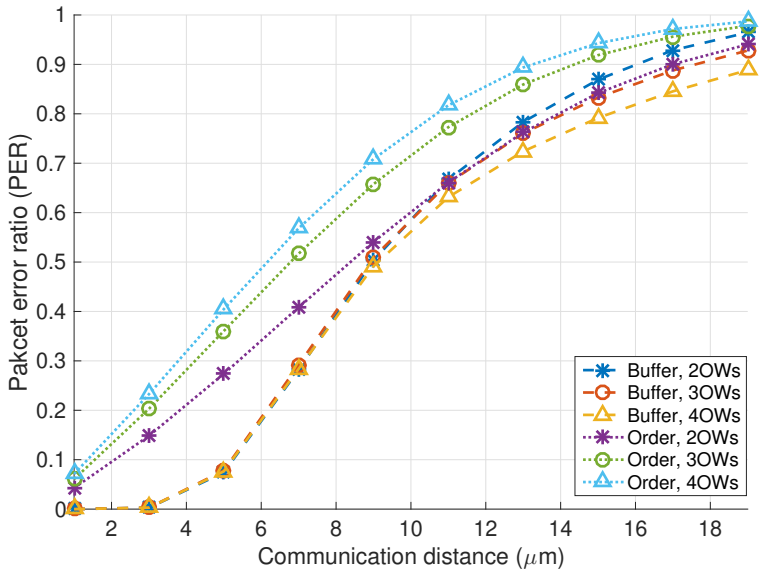

Figure 5.12: Packet error rate vs. communication distance

Fig. 5.12 shows the effect on the PER of the communication distance, which is one of the most crucial parameters for DBMC. For the order-based method, the PER increases from around 0.05 as we start the simulation of the communication distance from 0μm. At the same time, the PER is approximately 0 and hardly increases at the beginning when d is shorter than 6μm. In other words, when the communication distance is short, the MDP&BA can even mitigate the error packets. As d ranging from 6μm to 10μm, the PER of the MDP&BA increases significantly. Nevertheless, it is still lower than the order-based method. Conclusively, for the parameter d, we can see that the buffer-aided method outperforms the order-based method as well.

Fig. 5.13 shows that the PER of both of the methods tends to decrease when the time slot width increases. It is because the longer slot width can contribute to the higher accumulated arrival probability of the molecules in their corresponding

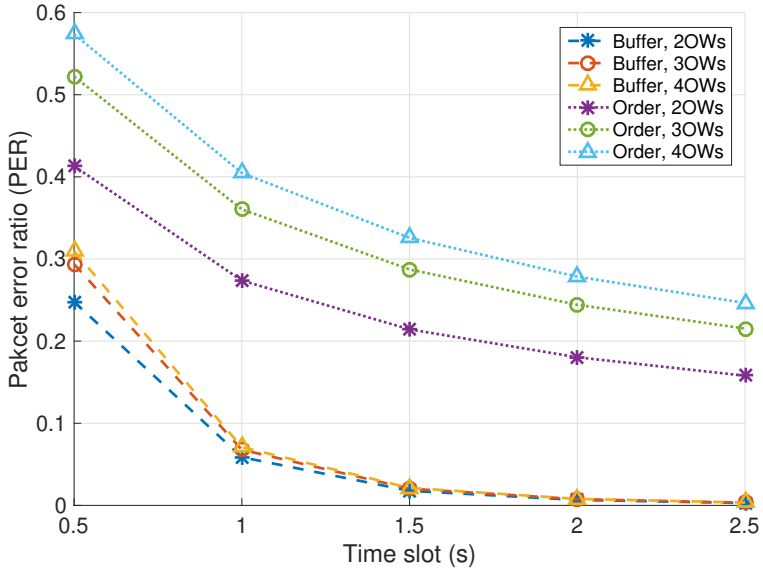

Figure 5.13: Packet error rate vs. time slot width

slot. We can also see that the average value of the PER and its lower bound of the order-based method is much higher than the MDP&BA method. When the time slot is longer than 1.5s, the PER of the MDP&BA method can even reach 0. It is similar to the result of the small d. The observing scheme does not affect the PER performance when evaluating the effect of ts. However on the other hand, for the order-based method, when $ts = 1.5$s the PER for any observing schemes is more extensive than 0.2. Even though as the time slot width increases, the PER turns to be smaller, the order-based method still has a lower bound higher than 0.1.

We plot the correlation of PER and the buffer size for the MDP&BA method in Fig. 5.14. When the buffer size is 1, namely the receiver only receives one packet that arrives the earliest in each slot, the result is the same as the order-based method. As the buffer size increases, the PER dramatically drops and reaches the lower bound as the buffer size is around 7 when the other parameters are set as default. The lower bound means that the buffer size is not necessary to be large to

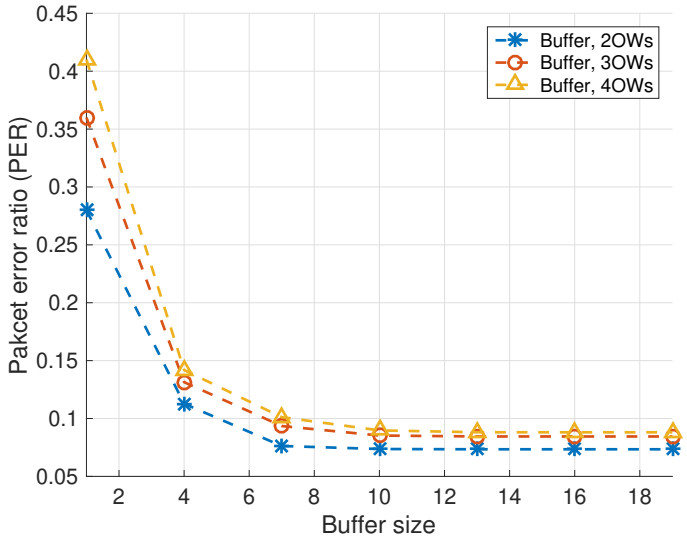

Figure 5.14: Packet error rate vs. buffer size

improve the transmission quality considering the number of DNA molecules are used in one slot. As the number of observing windows increases, the PER becomes slightly higher. However, it is still not a crucial factor to the PER, especially when the PER reaches its lower bound.

The curve demonstrating the variation of PER against the number of duplicated DNA packets is as shown in Fig. 5.15. The increase in the number of duplicated DNA molecules can contribute to the better transmission quality with lower PER. We can also see that for the default buffer size $BS = 10$ when the duplicated packet reaches 20, which is twice as the buffer size, the PER becomes nearly 0 regardless of the observing schemes. Consequently, it is reasonable to believe that with the assistance of the buffer in the receiver, the successful transmission ratio can be optimized by carefully choosing the buffer size and the number of duplicated packet molecules, meanwhile considering the cost of base pairs.

We also calculate the effective transmission rate (ETR). In the diffusive DNA

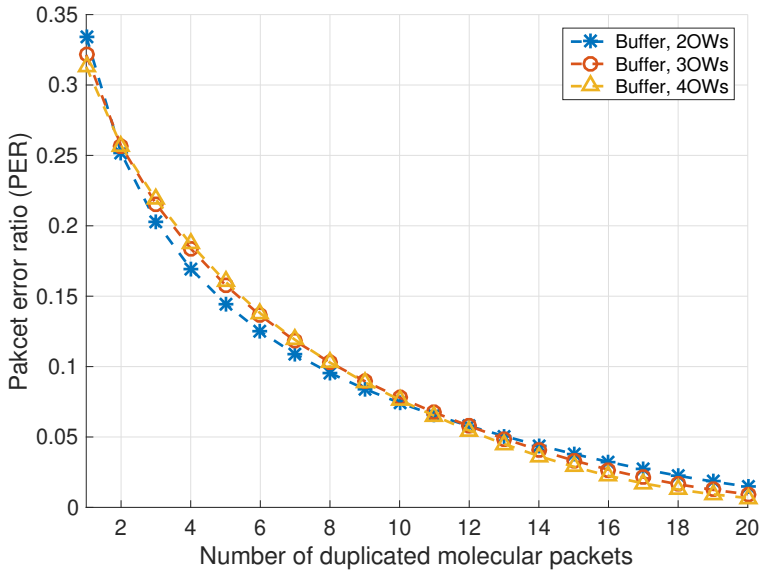

Figure 5.15: Packet error rate vs. number of duplicated DNA packets

based MC, we can believe that the ETR is the bits, namely the number of base pairs contained in the transmitted DNA molecule multiplied with the successful transmission ratio, as

$$ETR = NBP * (1 - PER) \tag{5.13}$$

Fig. 5.16 shows the tendency of ETR versus the various base pair number in one DNA packet. Thanks to the higher successful transmission ratio of the MDP&BA method, the ETR is significantly higher than the order-based method. We can also see that as the number of observing windows increases, the ETR tends to be lower. It is because with more observing windows, there will be more packets can be the error packet. It also shows that the ETR of the MDP&BA method monotonically increases as the number of base pairs increases. Thus we can conclude that for the MDP&BA method, the increasing number of base pairs does not influence the successful packet transmission. Therefore we can improve

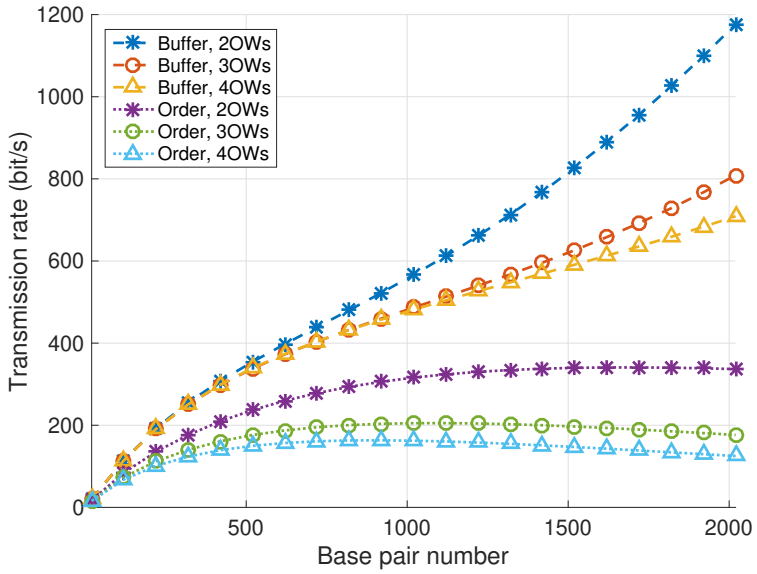

Figure 5.16: Effective transmission rate vs. number of base pairs

the transmission efficiency by increasing the number of base pairs in the DNA packet if irrespective of the base pair cost.

5.5 Summary

Aiming at increasing the channel capacity of DBMC, in this chapter we have presented a novel paradigm which uses DNA molecules as the data packets. In particular, we started from modeling a simple DNA-based encoding scheme utilizing the two types nucleobase pairs to represent binary bits. We considered the variation of DNA diffusion coefficient correlated with the number of base pairs. The mutation phenomenon and transcription error are taken into account as well. Moreover, based on the modeling of different observation schemes we derived the capacity of the proposed diffusive DNA based MC channel. The influences on the channel capacity of parameters including the number of base pairs, communication distance, and time slot width are evaluated in simulation experiments. Furthermore, the disordered arrival packet problem in the diffusive DNA based MC has been addressed. We proposed an MDP&BA method, in which multiple

duplicated DNA packets are applied to increase the probability of the successful arrival of the packets in the correct slot. A buffer used to store and compare the number of received DNA packets is also proposed. Simulation results show that the MDP&BA method improves the transmission success rate, compared with the order-based method.

Chapter 6

Conclusions and Future Work

The primary objective of this dissertation is to summarize our study on diffusion-based molecular communication (DBMC). In this chapter, we firstly generally conclude our work presented in this dissertation. Then some open opportunities for the possible directions of our research in the future are discussed in the later part of this chapter.

6.1 Conclusions

The MC is a novel paradigm to enable the communication among nano-machines. In a molecular communication system, the information of interest is encoded onto molecules rather than the electromagnetic waves in the conventional wireless communication networks. Among all the possible solutions to implement MC, the diffusion is the most straightforward scheme. However, there are still several problems in the most fundamental physical layer study of MC, such as the ISI, low channel capacity, among others. Therefore we have devoted our research focus on the DBMC.

Chapter 2 has introduced some fundamental preliminary knowledges of MC. Dissimilar to the conventional electromagnetic wave based communication techniques, the MC uses molecules to modulate signals of information. The modulation relies on the physical and chemical characters of molecules, including the concentration, the types, the released time, and the structures among others. The modulation schemes can be used not only individually, but also combinatorially to enable multiplexing. The signal propagation means are also different from the conventional techniques. In general, in the MC, information of interest propagates in the medium vis the physical displacement of the information molecules. Several methods, such as diffusion, gap-junction, bacteria, and molecular motors, can be used to implement the signal propagation of MC. Among all, the DBMC has the most simple requirement of the nano-machine nodes, and barely cost extra energy. Thus the DBMC becomes the currently most studied scheme among all. The channel model can be obtained through probability and physical ways. According to the distinctive channel characters, there are several problems need to solve in DBMC, such as the ISI and low channel capacity because of the extremely long delay due to the stochastic diffusion process.

Communication distance measurement

In Chapter 3, we have proposed a communication distance measurement approach based on the ATD. Two types of information molecules with different diffusion coefficient are used. Because of the different diffusion coefficient, the information molecules released in the medium have different diffusion speed. Thus when they are released by the node to be measured simultaneously, there is an ATD of the received signals at the node which measures the communication distance.

According to the channel model, the relation of the communication distance and the ATD can be derived. This method relies on measuring time. And as a result, it does not require the accurate concentration measurement, which is difficult for the nano-machines. On the other hand, the synchronization is a currently unsolved problem for the nano-machine nodes because of their limited capabilities. An RTT method estimates the communication distance by counting the travel time of a round trip signal has been proposed in the literature. The proposed ATD method also avoids the synchronization but without using the round trip signal. Therefore, comparing with RTT, the proposed method has a better efficiency performance and is more accurate because it overcomes the errors accumulated in the round trip signal. Additionally, the accuracy performance of ATD can be optimized by adjusting the diffusion coefficients of the chosen information molecules.

Mitigation of ISI for DBMC with mobile nodes

Chapter 4 has tried to address the ISI problem in DBMC for the communication among mobile Nano-nodes. In the CSK-based DBMC, the ISI is caused by the remaining information molecules of former transmitted bits in the medium. Because of the stochastic character of the diffusion process, such a phenomenon is unavoidable. Consequently, there are always molecules cannot arrive at the receiver in their corresponding slots. They are considered as interference when they finally reach at the receiver with the information molecules of the later transmitted bits.

Two main factors influence the ISI. The first one is the code width. The ISI and the code width have a negative correlation. Based on this, we have proposed a retransmission protocol. When the receiver sensed the received signal qual-

ity becomes poorer because of the channel variation due to the movement of the nodes, it transmits an ACK message to the transmitter. After the ACK arrives, the transmitter knows that the connection is no longer reliable and calculates the bits that need to retransmit. Then the bits are retransmitted with an increased code width. Another factor that influences the ISI is the communication distance. With other parameters fixed, the ISI becomes more severe when the communication distance increases. Therefore we have proposed an ACW protocol. It is based on the idea that there is a code width corresponding to a particular communication distance that can guarantee the connection quality as well as the transmission rate. In this protocol, the receiver transmits a distance-feedback periodically. Based on the feedback, the transmitter adjusts the code width according to the estimated communication distance. Simultaneously, the receiver calculates the code width of next bits based on the estimated communication distance as well. Thus the code width used to transmit the messages is always correspondingly adapted to the communication distance. Therefore, the ACW protocol can not only mitigate the impact of ISI but also optimize the possible transmission rate by reducing the redundant code width when the communication distance is short.

Both proposed protocols are aiming at mitigating the ISI of DBMC with mobile nodes. Comparing to the ACK protocol, the ACW is more real-time but on the cost of extra information molecules in the feedback signal and the computation resources of both the transmitter and the receiver.

Diffusive DNA based MC

The DBMC is considered with a low channel capacity because of the uncertainty and the long-delay character of the diffusion process. Aiming at this prob-

lem, we have proposed a novel DBMC using DNA molecules as data packets in Chapter 5. To this end, we adopt a simple modulation scheme using the two types of nucleobase pairs to represent different binary bits. As a result of the extensive information encoded onto one DNA molecule, a bit sequence (data packet) rather than a single bit can be transmitted by only one molecular transmission. The diffusion coefficient variation corresponding to the number of base pairs in the DNA molecules has been taken into account. Therefore, there is more uncertainty on the arrival probability of the DNA packets, leading to a disordered arrival problem. By modeling the observing windows and focus slot, we have derived the channel capacity of the proposed diffusive DNA based MC. Results have shown that the channel capacity is significantly increased and stable against different parameters.

Furthermore, to solve the disordered arrival problem of diffusive DNA based MC, we have proposed an MDP&BA protocol. In this method, the transmitter emits multiple duplicated DNA packets in each slot to increase the successful-arrival probability of the correct packet in the corresponding slot. Meanwhile, the buffer in the receiver stores the received packets temporarily and determines the received packet as the one with the most enormous amount. Simulation experiment results show that the PER is notability lower than the order-based method.

6.2 Future Work

In the future, we will mainly continue the research on unsolved problems and designing novel protocols for the DBMC. Multi-input Multi-output (MIMO) is one of the directions that can expand the current study of DBMC. To the end of designing MIMO for DBMC, the problems need to solve in the physical layer

include the interference of the signals of the MIMO channel. Network coding can be also applied to DBMC for more efficient and more secure transmission. Coding technology such as the redundant code can be used to solve the disordered arrival problem of diffusive DNA based MC. Studies on the higher layer are also necessary to implement MIMO, such as the medium access control (MAC). It is also essential to verify our theoretical and simulative research with practical experiments.

Appendix

A Molecular Communication-Based Targeted Drug Delivery

In the main body of this thesis, we have discussed several problems of the DBMC for the communication among Nano-machines. The potential applications of the Nano-machines and MC have been introduced in Chapter 1. Among all, the targeted drug delivery is considered as the most important one. In this appendix section, we present a simulation of an efficient TDD system using MC.

A1. System Framework

In our simulation system, the Nano-machines are considered as the artificially modified bacteria, which is capable of sensing, moving with the flagellum, and carrying and releasing specific molecules including drugs and information molecules for MC. These Nano-machines are deployed in the circulation system with drug and information molecules. They can flow in the vessels with the blood. However, by making use of the chemotaxis, we can guide the Nano-machines with the information molecules to a designed direction to deliver the drugs at the pathology more efficiently. The process of the proposed system is as shown in Fig. A.1.

For simplicity, we focus on a small part of the circulation system in a patient body. This part of is modeled as a tree-shaped structure, similar to [67]. The

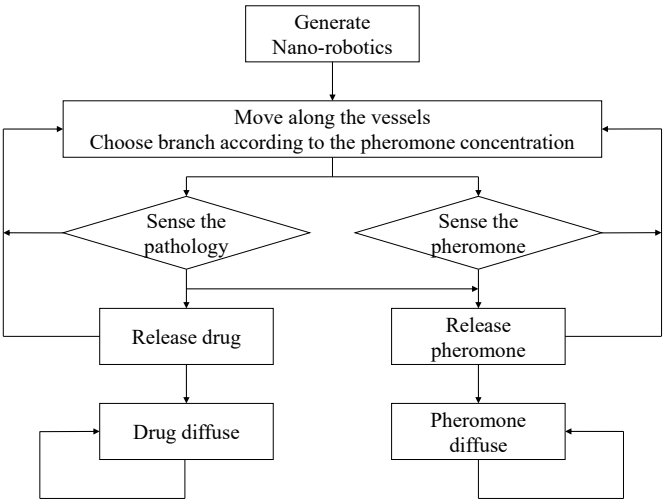

Figure A.1: Flow chart of the proposed system

pathology is assumed at one of the end branches, as shown in Fig. A.2 and A.3, noted with a star. Because it is a part of the circulation system, the drug carrier Nano-machines which are injected at some point of the circulation system. Through time they can randomly flow to this simulated part.

A. Diffusion Model

Because of the Brownian motion, the molecules in aqueous environment follow a diffusion principle described as Fick's laws as introduced in Chapter 2. In the Fick's second law, the concentration of molecules varies according to

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} \tag{A.1}$$

in which, C is the concentration of the molecule, x denotes the coordinate, D is the diffusion coefficient, which is determined by the temperature, the dynamic viscosity of the fluid and the size of the molecules.

Since in the system, the information molecules are the attractants [43] to the bacteria-based Nano-machines, we do not take the encoding into account. The molecules, both the information molecules and the drug, simply diffuse according to the Fick's second law model in (A.1).

B. Mobility Model of Nano-machines

Some of the bacteria has flagellate, which enables them to change direction and to move within a limited range. Meanwhile, the bacteria have chemotaxis. Namely, they can be attracted or propelled by some specific chemicals. For instance, Amino acids and low concentration salt are typical attractants for bacteria, while alcohols are normally considered as repellents. Moreover, the concentration of these molecules in different directions also affects the mobility of the bacteria [43].

Based on these characters, directing bacteria-based Nano-machines to an expected region becomes possible. The Nano-machines in the proposed system can be guided by the chemical attractant, which can be seen as a form of guiding information on molecules, also known as pheromone. When a Nano-machine reaches one bifurcation of the tree-structured vessel system, it will more possibly choose the branch with a higher concentration of attractant. The switching model can be described as

$$p_i(k) = \frac{C_i}{\sum C} \tag{A.2}$$

where, $p_i(k)$ is the probability that the Nano-machine k choose the branch i, C represents the attractant concentration at each branch.

C. Guiding Information Propagation Model

The attractant molecules normally do not exist in the vessels when the Nano-machines have not found the pathology. All the Nano-machines randomly flow with the blood. As long as one Nano-machine arrived the pathology, it releases the attractant and the drug at that region. Then, the attractant and drug start to diffuse through time. When the following Nano-machines sense that the attractant concentration is higher than a threshold, they will release more attractant to speed up the attractant spreading. Thus, the guiding information with the attractant can be amplified and propagates more efficiently. The propagation model of the guiding information is given as

$$I\left(x, t+1\right) = C\left(I\left(x, t\right) + C_T\left(x_n, t\right) \cdot C_R\right) \tag{A.3}$$

where I represents the guiding information, C is diffusion function, C_T is a judgment of whether the attractant concentration at x_n is higher than the threshold, C_R is the attractant can be released within a step.

A2. Simulation Results

The simulation experiments are carried out with Matlab. At the beginning point of the tree-structured vessels, we randomly generate the Nano-machines, then let them act based on the models illustrated above. The simulation result of how the guiding information propagate is as shown in Fig. A.2.

The two figures illustrate different phases in the propagation process. They show the propagation states at simulation steps of 600, 1200, respectively. From

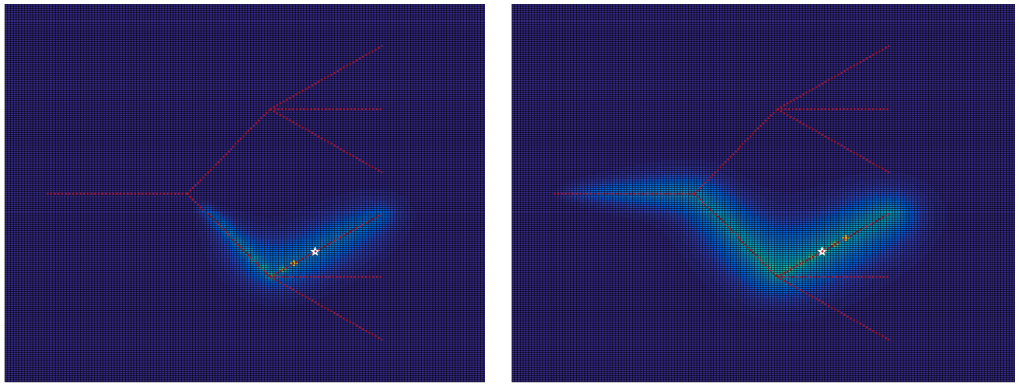

Figure A.2: Guiding information propagation process

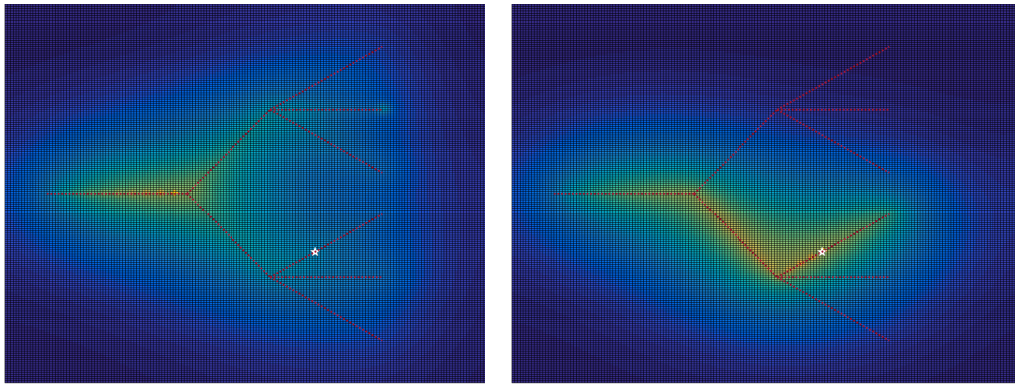

Figure A.3: Comparison of random delivery and the proposed system

the figures, it is obvious that the guiding information propagates along with the path in the vessels, which can lead the Nano-machines to the pathology. Namely, we can believe that with this system, the drug delivery efficiency can be improved compared with traditional drugs. To compare with our system, we also did the simulation of a random mobility-based system in the same environment, which can be seen as the conventional drug.

Fig. A.3 shows how the pheromone distributed in the simulation domain, namely how frequently the Nano-machines take each path, in which, the left figure is for the Nano-machines randomly choose branches, while the right one indicates the result of it with the proposed mechanism. From the figures, we can see that

when the drug is randomly delivered, they appear evenly at each branch. Meanwhile on the other hand, with the proposed system, more drugs carried by the Nano-machines can be delivered to the pathology. In other words, to obtain a particular amount of the drugs at the pathology, the MC-aided TDD system consumes fewer drugs.

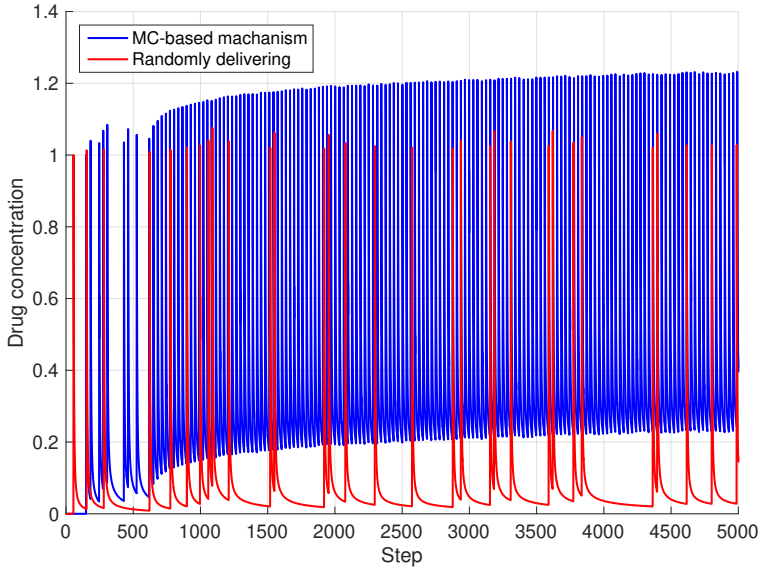

Figure A.4: Drug concentration comparison.

We also recorded the drug concentration variation at the pathology, as shown in Fig. A.4. The blue curve represents the drug concentration with the proposed method. We can see that the drug concentration can keep steady at a relatively high level after the guiding information propagates to upper reach. By adapting the number of the Nano-machines and the drug amount carried with them, it is reasonable to believe that the drug concentration can be controlled artificially. In contrast, the red curve shows drug concentration when the Nano-machines randomly choose the path, simulating conventional drugs, which is less efficient in the delivery rate. The accumulated drug concentration is as shown in Fig. A.5. It

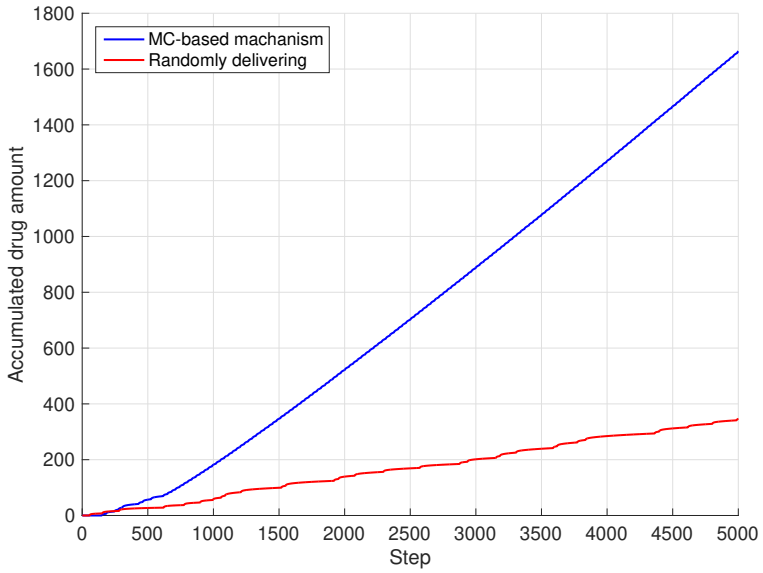

Figure A.5: Drug concentration comparison.

proves more straightforwardly that the proposed system is more efficient in delivering drugs to the pathology.

A3. Summary

This appendix proposed a targeted drug delivery system using molecular communication among modified bacteria-based Nano-machines. Attractant chemicals molecules are used for guiding the Nano-machines to the pathology to where the drug is intended to be delivered. The diffusion of both the drugs and the information molecules are taken into consideration in the simulation. The mobility of the Nano-machines and the propagation of the guiding information are modeled as well. Simulation results show that the proposed method has a higher drug delivery efficiency compared with conventional drugs.

List of Publication

Journal Articles

1. Yao Sun, Masaki Ito, Kaoru Sezaki, “Diffusion-based Molecular Communication using DNA Molecules as Data Packets”, Nano Communication Networks (Under preparation)

International Conference Papers

1. Yao Sun, Masaki Ito, Kaoru Sezaki, “Channel Capacity Analysis of a Novel Molecular Communication based on Diffusive DNA”, IEEE Wireless Communications and Networking Conference (IEEE WCNC), 2019 (Accepted)
2. Yao Sun, Masaki Ito, Kaoru Sezaki, ”An Efficient Distance Measurement Approach in Diffusion-based Molecular Communication based on Arrival Time Difference”, 4th ACM International Conference on Nanoscale Computing and Communication, Washington DC, USA, Sep. 2017
3. Yao Sun, Masaki Ito, Kaoru Sezaki, ”Adaptive Code Width Protocol for Mitigating Intersymbol Interference in Diffusion-based Molecular Communication with Mobile Nodes”, 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom), Munich, Germany, Sep. 2016

Domestic Conference Papers

1. Yao Sun, Masaki Ito, Kaoru Sezaki, “Buffer Aided Receiver for Diffusive DNA Based Molecular Communication”, 2019 IEICE General Conference, Tokyo, March, 2019 (Submitted)
2. Yao Sun, Masaki Ito, Kaoru Sezaki, “Channel Capacity of Diffusive DNA-based Molecular Communication”, 2018 IEICE General Conference, BS-2-30, Tokyo, March, 2018
3. Yao Sun, Masaki Ito, Kaoru Sezaki, “A Novel Molecular Communication through Diffusive DNA Molecules”, IEICE Ambient intelligence and Sensor Networks (IEICE ASN), Yogyakarta, Indonesia, Nov. 2017
4. Yao Sun, Masaki Ito, Kaoru Sezaki, “A Retransmission Scheme in Unicast Diffusion-based Molecular Communication”, 2016 IEICE Society Conference, Network and Service Design, Control and Management, BS-5-6, Sapporo, Sep. 21, 2016
5. Yao Sun, Masaki Ito, Kaoru Sezaki, ”A Targeted Drug Delivery Approach using Molecular Communication”, 2016 IEICE General Conference, vol.BS-3-19, Fukuoka, March 16, 2016

Bibliography

[1] Richard P Feynman. There ' s plenty of room at the bottom: An invitation to enter a new field of physics. In *Handbook of Nanoscience, Engineering, and Technology, Third Edition*, pages 26–35. CRC Press, 2012.

[2] Salma Kassem, Alan TL Lee, David A Leigh, Vanesa Marcos, Leoni I Palmer, and Simone Pisano. Stereodivergent synthesis with a programmable molecular machine. *Nature*, 549(7672):374, 2017.

[3] Alain Codourey, Wolfgang Zesch, Roland Buchi, and Roland Siegwart. A robot system for automated handling in micro-world. In *Intelligent Robots and Systems 95.'Human Robot Interaction and Cooperative Robots', Proceedings. 1995 IEEE/RSJ International Conference on*, volume 3, pages 185–190. IEEE, 1995.

[4] Hee Hyun Lee, Etienne Menard, JARNG Tassi, and G Blanchet. Large area microcontact printing presses for plastic electronics. *Materials Research Society Bulletin*, 846:731–736, 2005.

[5] K Eric Drexler. Nanosystems. *Molecular machinery, manufacturing and computation.*, 1992.

[6] Ricky K Soong, George D Bachand, Hercules P Neves, Anatoli G Olkhovets, Harold G Craighead, and Carlo D Montemagno. Powering an inorganic nanodevice with a biomolecular motor. *Science*, 290(5496):1555–1558, 2000.

[7] Tatsuya Suda, Michael Moore, Tadashi Nakano, Ryota Egashira, Akihiro Enomoto, Satoshi Hiyama, and Yuki Moritani. Exploratory research on molecular communication between

nanomachines. In *Genetic and Evolutionary Computation Conference (GECCO), Late Breaking Papers*, volume 25, page 29. sn, 2005.

[8] Baris Atakan, Ozgur B Akan, and Sasitharan Balasubramaniam. Body area nanonetworks with molecular communications in nanomedicine. *IEEE Communications Magazine*, 50(1), 2012.

[9] R Freitasjr. What is nanomedicine? nanomedicine: nanotechnology. *Biol. Med*, 1:2, 2005.

[10] Chun-Hua Lu, Bilha Willner, and Itamar Willner. Dna nanotechnology: from sensing and dna machines to drug-delivery systems. *ACS nano*, 7(10):8320–8332, 2013.

[11] Bo Song, Ruiguo Yang, Ning Xi, Kevin Charles Patterson, Chengeng Qu, and King Wai Chiu Lai. Cellular-level surgery using nano robots. *Journal of laboratory automation*, 17(6):425–434, 2012.

[12] Gianfranco Cerofolini, Paolo Amato, Massimo Masserini, and Giancarlo Mauri. A surveillance system for early-stage diagnosis of endogenous diseases by swarms of nanobots. *Advanced Science Letters*, 3(4):345–352, 2010.

[13] Ministry of Internal Affairs and Communications. Current situation of spectrum use in japan as of june 2018. http://www.tele.soumu.go.jp/e/adm/freq/search/myuse/0002/index.htm.

[14] National Telecommunications and Information Administration. United states frequency allocations: The radio spectrum chart as of 2016. https://www.ntia.doc.gov/page/2011/united-states-frequency-allocation-chart.

[15] Peter J Burke, Shengdong Li, and Zhen Yu. Quantitative theory of nanowire and nanotube antenna performance. *IEEE transactions on nanotechnology*, 5(4):314–334, 2006.

[16] Josep Miquel Jornet and Ian F Akyildiz. Graphene-based nano-antennas for electromagnetic nanocommunications in the terahertz band. In *Antennas and Propagation (EuCAP), 2010 Proceedings of the Fourth European Conference on*, pages 1–5. IEEE, 2010.

[17] Masayoshi Tonouchi. Cutting-edge terahertz technology. *Nature photonics*, 1(2):97, 2007.

[18] Leonard M Adleman. Molecular computation of solutions to combinatorial problems. *Science*, 266(5187):1021–1024, 1994.

[19] Nagarajan Nandagopal and Michael B Elowitz. Synthetic biology: integrated gene circuits. *science*, 333(6047):1244–1248, 2011.

[20] Bernard Sklar. *Digital communications*, volume 2. Prentice Hall Upper Saddle River, 2001.

[21] Massimiliano Pierobon and Ian F Akyildiz. A physical end-to-end model for molecular communication in nanonetworks. *IEEE Journal on Selected Areas in Communications*, 28(4), 2010.

[22] Arash Einolghozati, Mohsen Sardari, and Faramarz Fekri. Design and analysis of wireless communication systems using diffusion-based molecular communication among bacteria. *IEEE transactions on wireless communications*, 12(12):6096–6105, 2013.

[23] Massimiliano Pierobon and Ian F Akyildiz. Diffusion-based noise analysis for molecular communication in nanonetworks. *IEEE Transactions on Signal Processing*, 59(6):2532–2547, 2011.

[24] Deniz Kilinc and Ozgur B Akan. Receiver design for molecular communication. *IEEE Journal on Selected Areas in Communications*, 31(12):705–714, 2013.

[25] Ege T Kavalali. The mechanisms and functions of spontaneous neurotransmitter release. *Nature Reviews Neuroscience*, 16(1):5, 2015.

[26] Mehmet S Kuran, Huseyin Birkan Yilmaz, Tuna Tugcu, and Ian F Akyildiz. Modulation techniques for communication via diffusion in nanonetworks. In *Communications (ICC), 2011 IEEE International Conference on*, pages 1–5. IEEE, 2011.

[27] Mohammad Upal Mahfuz, Dimitrios Makrakis, and Hussein T Mouftah. On the characterization of binary concentration-encoded molecular communication in nanonetworks. *Nano Communication Networks*, 1(4):289–300, 2010.

[28] Yonathan Murin, Nariman Farsad, Mainak Chowdhury, and Andrea Goldsmith. Communication over diffusion-based molecular timing channels. In *Global Communications Conference (GLOBECOM), 2016 IEEE*, pages 1–6. IEEE, 2016.

[29] JA Tenreiro Machado. Shannon entropy analysis of the genome code. *Mathematical Problems in Engineering*, 2012, 2012.

[30] Sohil Shah, Ashwin Raghavachari, Chieh Lo, and Radu Marculescu. Molecular communication with dna cellular storage system. In *Proceedings of the 4th ACM International Conference on Nanoscale Computing and Communication*, page 24. ACM, 2017.

[31] Hamidreza Arjmandi, Amin Gohari, Masoumeh Nasiri Kenari, and Farshid Bateni. Diffusion-based nanonetworking: A new modulation technique and performance analysis. *IEEE Communications Letters*, 17(4):645–648, 2013.

[32] NAKANO Tadashi, Michael Moore, ENOMOTO Akihiro, and S Tatsuya. Part ii recent research and development in molecular communication technology molecular communication : Biological communications technology.

[33] Tadashi Nakano, Tatsuya Suda, Takako Koujin, Tokuko Haraguchi, and Yasushi 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*, pages 139–146. IEEE, 2007.

[34] Vitaly Petrov, Dmitri Moltchanov, Sasitharan Balasubramaniam, and Yevgeni Koucheryavy. Incorporating bacterial properties for plasmid delivery in nano sensor networks. *IEEE Transactions on Nanotechnology*, 14(4):751–760, 2015.

[35] Maria Gregori and Ian F Akyildiz. A new nanonetwork architecture using flagellated bacteria and catalytic nanomotors. *IEEE Journal on selected areas in communications*, 28(4), 2010.

[36] Satoshi Hiyama, Riho Gojo, Tomohiro Shima, Shoji Takeuchi, and Kazuo Sutoh. Biomolecular-motor-based nano-or microscale particle translocations on dna microarrays. *Nano Letters*, 9(6):2407–2413, 2009.

[37] Tadashi Nakano, Andrew W Eckford, and Tokuko Haraguchi. *Molecular communication*. Cambridge University Press, 2013.

[38] Adrian J Gray, David Jenkins, Mark H Andrews, Dennis R Taaffe, and Megan L Glover. Validity and reliability of gps for measuring distance travelled in field-based team sports. *Journal of Sports Sciences*, 28(12):1319–1325, 2010.

[39] Wooyoung Kim, Kirill Mechitov, J-Y Choi, and Soo Ham. On target tracking with binary proximity sensors. In *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*, pages 301–308. IEEE, 2005.

[40] Mustapha Reda Senouci, Abdelhamid Mellouk, Hadj Senouci, and Amar Aissani. Performance evaluation of network lifetime spatial-temporal distribution for wsn routing protocols. *Journal of Network and Computer Applications*, 35(4):1317–1328, 2012.

[41] Burcu Tepekule, Ali E Pusane, H Birkan Yilmaz, Chan-Byoung Chae, and Tuna Tugcu. Isi mitigation techniques in molecular communication. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, 1(2):202–216, 2015.

[42] Xiaolong Yang, Yanan Tang, Sean D Mason, Junbo Chen, and Feng Li. Enzyme-powered three-dimensional dna nanomachine for dna walking, payload release, and biosensing. *ACS nano*, 10(2):2324–2330, 2016.

[43] Yutaka Okaie, Tadashi Nakano, Takahiro Hara, Takuya Obuchi, Kazufumi Hosoda, Yasushi Hiraoka, and Shojiro Nishio. Cooperative target tracking by a mobile bionanosensor network. *IEEE transactions on nanobioscience*, 13(3):267–277, 2014.

[44] Howard A Stone, Abraham D Stroock, and Armand Ajdari. Engineering flows in small devices: microfluidics toward a lab-on-a-chip. *Annu. Rev. Fluid Mech.*, 36:381–411, 2004.

[45] Nariman Farsad, Na-Rae Kim, Andrew W Eckford, and Chan-Byoung Chae. Channel and noise models for nonlinear molecular communication systems. *IEEE Journal on Selected Areas in Communications*, 32(12):2392–2401, 2014.

[46] Michael J Moore, Tadashi Nakano, Akihiro Enomoto, and Tatsuya Suda. Measuring distance from single spike feedback signals in molecular communication. *IEEE Transactions on Signal Processing*, 60(7):3576–3587, 2012.

[47] Zhen Cheng, Yihua Zhu, Kaikai Chi, Yanjun Li, and Ming Xia. Capacity analysis for diffusive molecular communication with isi channel. *Nano Communication Networks*, 13:43–50, 2017.

[48] Mehmet Şükrü Kuran, H Birkan Yilmaz, Tuna Tugcu, and Bilge Özerman. Energy model for communication via diffusion in nanonetworks. *Nano Communication Networks*, 1(2):86–95, 2010.

[49] Peng He, Yuming Mao, Qiang Liu, and Kun Yang. Adaptive threshold variation algorithm for diffusion-based molecular communication. In *Proceedings of the Second Annual International Conference on Nanoscale Computing and Communication*, page 20. ACM, 2015.

[50] Mohammad Mahfuz, Dimitrios Makrakis, and Hussein T Mouftah. Characterization of intersymbol interference in concentration-encoded unicast molecular communication. In *Electrical and Computer Engineering (CCECE), 2011 24th Canadian Conference on*, pages 000164–000168. IEEE, 2011.

[51] Burcu Tepekule, Ali E Pusane, H Birkan Yilmaz, and Tuna Tugcu. Energy efficient isi mitigation for communication via diffusion. In *Communications and Networking (BlackSeaCom), 2014 IEEE International Black Sea Conference on*, pages 33–37. IEEE, 2014.

[52] Tadashi Nakano, Yutaka Okaie, and Jian-Qin Liu. Channel model and capacity analysis of molecular communication with brownian motion. *IEEE communications letters*, 16(6):797–800, 2012.

[53] Natalia V Ivanova, Mehrdad Hajibabaei, Paul DN Hebert, et al. Assembling dna barcodes. In *Environmental Genomics*, pages 275–294. Springer, 2008.

[54] Nathaniel G Portney, Yonghui Wu, Lauren K Quezada, Stefano Lonardi, and Mihrimah Ozkan. Length-based encoding of binary data in dna. *Langmuir*, 24(5):1613–1616, 2008.

[55] Gourab Chatterjee, Neil Dalchau, Richard A Muscat, Andrew Phillips, and Georg Seelig. A spatially localized architecture for fast and modular dna computing. *Nature nanotechnology*, 12(9):920, 2017.

[56] Arash Einolghozati, Mohsen Sardari, and Faramarz Fekri. Capacity of diffusion-based molecular communication with ligand receptors. In *Information Theory Workshop (ITW), 2011 IEEE*, pages 85–89. IEEE, 2011.

[57] Dogu Arifler. Capacity analysis of a diffusion-based short-range molecular nano-communication channel. *Computer Networks*, 55(6):1426–1434, 2011.

[58] Massimiliano Pierobon and Ian F Akyildiz. Capacity of a diffusion-based molecular communication system with channel memory and molecular noise. *IEEE Transactions on Information Theory*, 59(2):942–954, 2013.

[59] Carter Bancroft, Timothy Bowler, Brian Bloom, and Catherine Taylor Clelland. Long-term storage of information in dna. *Science*, 293(5536):1763–1765, 2001.

[60] George M Church, Yuan Gao, and Sriram Kosuri. Next-generation digital information storage in dna. *Science*, page 1226355, 2012.

[61] Nick Goldman, Paul Bertone, Siyuan Chen, Christophe Dessimoz, Emily M LeProust, Botond Sipos, and Ewan Birney. Towards practical, high-capacity, low-maintenance information storage in synthesized dna. *Nature*, 494(7435):77, 2013.

[62] Douglas E Smith, Thomas T Perkins, and Steven Chu. Dynamical scaling of dna diffusion coefficients. *Macromolecules*, 29(4):1372–1373, 1996.

[63] Gergely L Lukacs, Peter Haggie, Olivier Seksek, Delphine Lechardeur, Neal Freedman, and AS Verkman. Size-dependent dna mobility in cytoplasm and nucleus. *Journal of biological chemistry*, 275(3):1625–1629, 2000.

[64] Anthony Balducci, Pan Mao, Jongyoon Han, and Patrick S Doyle. Double-stranded dna diffusion in slitlike nanochannels. *Macromolecules*, 39(18):6273–6281, 2006.

[65] Hong Shen, Yueyue Hu, and WM Saltzman. Dna diffusion in mucus: effect of size, topology of dnas, and transfection reagents. *Biophysical journal*, 91(2):639–644, 2006.

[66] Olgica Milenkovic, Ryan Gabrys, Han Mao Kiah, and SM Hossein Tabatabaei Yazdi. Exabytes in a test tube. *IEEE Spectrum*, 55(5):40–45, 2018.

[67] Youssef Chahibi, Massimiliano Pierobon, and Ian F Akyildiz. Pharmacokinetic modeling and biodistribution estimation through the molecular communication paradigm. *IEEE Transactions on Biomedical Engineering*, 62(10):2410–2420, 2015.