

Channel Capacity Analysis of DNA-based Molecular Communication with Length Encoding Mechanism

Jialin Xie^{1,2}, Qiang Liu^{2*}, Kun Yang³ and Lin Lin⁴

¹ Yangtze Delta Region Institute (Quzhou),
University of Electronic Science and Technology of China,
Quzhou, Zhejiang 324000, China

² School of Information and Communication Engineering,
University of Electronic Science and Technology of China
Chengdu, Sichuan 611731, China
[e-mail: liuqiang@uestc.edu.cn]

³ School of Computer Science & Electronic Engineering, University of Essex
Wivenhoe Park, Colchester, Essex, UK
[e-mail: kunyang@essex.ac.uk]

⁴ Department of Information and Communication Engineering, Tongji University
Shanghai 201804, China

*Corresponding author: Qiang Liu

*Received January 20, 2021; revised June 22, 2021; accepted July 21, 2021;
published August 31, 2021*

Abstract

The double helix structure of DNA makes it diverse, stable and can store information with high density, and these characteristics are consistent with the requirements of molecular communication for transport carriers. In this paper, a specific structure of molecular communication system based on DNA length coding is proposed. Transmitter (Tx) adopts the multi-layer golden foil design to control the release of DNA molecules of different lengths accurately, and receiver (Rx) adopts an effective and sensitive design of nanopore, and the biological information can be converted to the electric signal at Rx. The effect of some key factors, e.g., the length of time slot, transmission distance, the number of releasing molecules, the priori probability, on channel capacity is demonstrated exhaustively. Moreover, we also compare the transmission capacity of DNA-based molecular communication (DNA-MC) system and concentration-based molecular communication (MC) system under the same parameter setting, and the peak value of capacity of DNA-MC system can achieve 0.08 bps, while the capacity of MC system remains 0.025 bps. The simulation results show that DNA-MC system has obvious advantages over MC system in saving molecular resources and improving transmission stability.

Keywords: channel capacity, DNA-based molecular communication, length encoding, multi-layer foil, nanopore

1. Introduction

Currently, DNA-based molecular communication (DNA-MC) becomes an emerging research direction. Just as its name implies, DNA-MC represents a communication system with DNA as its information carrier, and many research groups believe that DNA is a potential storage media for its structural stability. In the water-rich environment, the conventional communication model based on electromagnetic waves is infeasible [1]. This special information carrier enables DNA-MC system to complete the task of information transmission in special scenarios, such as: body fluid environment monitoring [2][3], drug delivery, molecular directional communication, and environmental monitoring, etc. According to the Fick's law, the characteristics of diffusive molecules arriving the destination are exhaustively studied, Tao et al. have proposed a synthetic molecular communication modulator which is realized by multi-layer assembly technique to immobilize DNA molecules on the golden foil, and this nanodevice can control the releasing order of DNA molecules through electrical signals [4]. Yao et al. have established a channel capacity model with DNA as the data packets in one-dimensional environment, and Yao's group has proposed the mechanism of observation window to ensure the reception of packets was in order [5], otherwise, it may lead to severe inter symbol interference (ISI) [6]. Bilgin et al. have proposed the structure of nanopores which can convert the length of the arriving DNA strands into the electrical signal in Rx, and the DNA length coding mechanism is applied to this system [7]. This paper has proposed a novel design of DNA-MC system based on these researches.

DNA, the full name of Deoxyribonucleic Acid, has the double helix spatial structure, and abundant research work have been accomplished studying the characteristics of DNA [8]. Owing the double helix structure, the phosphoric acid and deoxyribose sequence in DNA molecules is stable, and extra energy is needed to break hydrogen bond. Information is stored in the DNA strand in the form of base sequence firmly. Moreover, the base pairs are all independent of each other, and each nucleotide has four options: adenine (A), cytosine (C), guanine (G), or thymine (T), so the DNA strands can store information at an extremely high density [9]. Due to the property that different base sequences determine different DNA molecules, there are different expression patterns in a DNA strand with only base pairs. Recently, the technology of manipulating DNA is developing in an unprecedented pace, and this technology enable us to synthesize DNA molecules as needed in a quicker and more cost-effective way [10][11]. Especially, in the nanoscale communication system, the transceiver needs to be within the nano size, and there is limited space for the transceiver to store the information molecules. However, the DNA manipulating technology can convert massive of information into the base sequence [11][12] or the length information [13] of the DNA strands, hence satisfying the actual requirements of information transmission. All these features make DNA an ideal information carrier in diffusion environment characterized by low arrival probability at receiver (Rx).

The theoretical research of each part of the system is very exhaustive [14][15][16], and the multi-layer golden foil is assembled by soaking the golden foil with DNA molecules and ionic solution in turn. Electrical signals control the release of specific DNA molecules by dissolving the ionic layer of the multi-layer golden foil, so as to complete the conversion from electrical signal to biological signal. Moreover, the foil plays the role of interface between macro world and micro system [17][18][19]. In the general system, a container is embedded in the Tx to store the releasing molecules [20]. If this structure is adopted, the channel capacity of the system will be limited due to the volume of the container [21], and multiple kinds of molecules cannot be stored at the same time. Moreover, the replenishment of molecules is also a

significant issue which affects the system performance [22][23]. Compare with this structure, the multi-layer foil compensates these defects, and it can even fix the number of molecules released in each slot to a constant. Therefore, the application of multi-layer foil makes the release of molecules easier to control.

Brownian motion is an irregular motion caused by the collision between solute molecules and solvent molecules, and the motion is suitable for the case that solvent molecules are much smaller and lighter than the solute molecules, and the collisions should not lead to the change of two molecules [24]. The diffusion of molecules in body-fluid environment is often regarded as the simple Brownian motion, assuming that the body-fluid environment remains stable at all times, and according to Fick' law, the probability of one molecule diffuses to the Receiver (Rx) can be calculated in one-dimensional. Although the diffusion motion of a molecule in the body-fluid environment is stochastic, in the beginning of one time slot, the Tx releases a certain number of information molecules, and the probability that the Rx detects the corresponding molecule before the end of slot can be regarded as the channel response in this slot.

The structure of nanopore is often used in DNA analysis, and the paper [7] introduces the system which adopts the nanopore structure in Rx, and taking the processing time of Rx which is related to the length of DNA strands into account. Thanks to this mechanism, the length of DNA molecules can be applied to encode information. Moreover, the size of each nanopore is designed elaborately, and only one DNA molecule which is negatively charged can pass through the nanopore each time. Based on this structure, the Rx can convert the length of DNA molecules into the electrical signal that can be easily recognized. In DNA-MC system, the Rx consists of several nanopores, and the employ of multiple nanopores makes the Rx more efficient and more sensitive to the changes of molecules around. Thus, the nanopores is of great significance in the diffusive environment with low probability of arrival, i.e., the body-fluid environment.

This article has proposed a joint design of DNA-based molecular system, and the transmitter adopts the multi-layer foil structure, and the receiver uses the nanopores for reception. The researches on the two structures are exhaustive, but this paper is the first to combine the two structures together and apply them to the DNA-MC system at the same time. Besides, the influence of the key parameters of the transceiver on the transmission capacity is analyzed comprehensively, and it provides suggestions for the parameter setting of actual DNA-MC system.

As an information carrier, the stability and diversity of the double helix structure of DNA make is possible to encode information into the strand lengths and the base pair sequences. This unique way of information encoding can make up for the defect of low probability of arrival between transceiver in diffusive environment. Nevertheless, in the field of DNA-based molecular communication, the research on channel capacity is not comprehensive. Applying the DNA length encoding mechanism, the influences of key factors, e.g., the number of released molecules, the number of nanopores (N_n), distance (x), the length of time slot (T) and the probability of transmitting "1" (p) on channel capacity is exhaustively analyzed.

Moreover, this paper compares the performance differences between DNA-based molecular communication and concentration-based molecular communication, namely, using the molecular concentration at Rx to encode information. Although the reception process of the two systems is different, supposing two systems release the same number of DNA molecules N_c or have the same transmission distance x and the results show that DNA-MC is better than MC in terms of transmission capacity. The simulation results indicate the peak value of channel capacity of DNA-MC system can achieve 0.08 bps when the $x = 8\mu m$ and

$N_c = 400$ while the capacity of MC system remains 0.025 bps.

The remainder of the manuscript is organized as follows. Section II introduces the structure of DNA-MC system, and the factors which affect the successful reception of DNA molecules is analyzed as well. The channel capacity is deduced in Section III, and simulation results of the system performance and capacity comparison between DNA-MC system and MC system are presented in Section IV. Finally, some conclusions and the plans for future work are discussed in Section V.

2. DNA-based Molecular Communication System Analysis

In this section, we consider the DNA-based molecular communication (DNA-MC) scheme is in the one-dimensional situation, that is, the DNA molecules are transmitted between one pair of transmitters and receiver. In this system, the structures of Tx and Rx are delicately designed to meet the needs of release or distinguish different DNA strands in specific time slot, and the ICT metrics are calculated based on this model. The structure of DNA-MC is introduced as followed.

2.1 Basic system structure

The structure of DNA-Based molecular communication is divided into three parts: transmitter (Tx), diffusive channel and receiver (Rx). In order to increase the realizability and performance of the system, the Tx and Rx have applied practical structures, and the whole system is operating on the diffusive body fluid environment. The system model is described as Fig. 1.

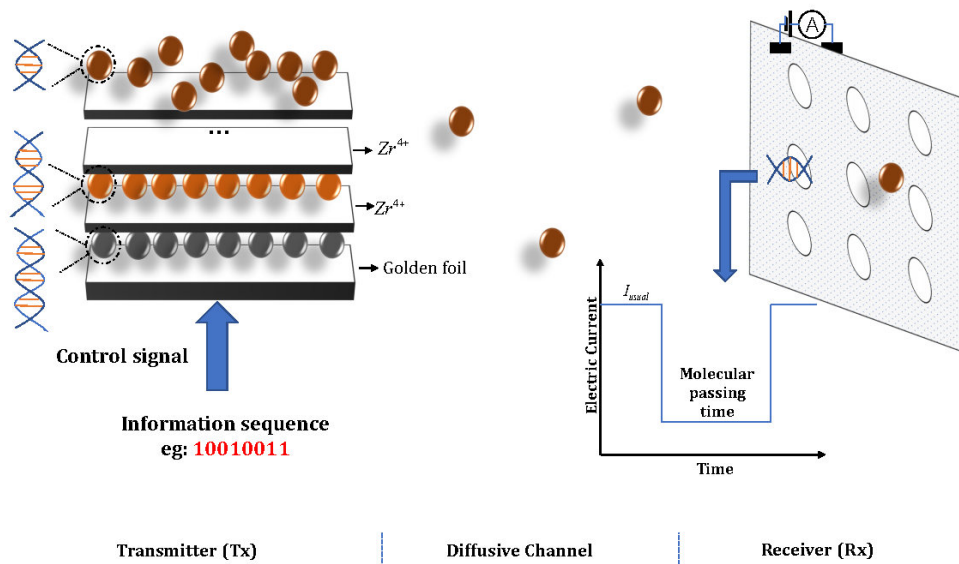


Fig. 1. The structure of DNA based molecular communication system

2.1.1 Transmitter

The Tx has introduced in the structure of multilayer film, which is fabricated by layer-by-layer assembly technique. As shown in Fig. 1, the Tx has contained various kinds of DNA molecules which are distinguished by strands length, and the molecules are attached to a gold thin film. Therefore, the gold thin film has immersed DNA molecules and Zr^{4+} alternately to separate different kinds of DNA molecules [4], and DNA molecules are immersed to gold thin film in descending order of strand length. To convert the information sequences into the release order of specific DNA molecules, the Tx generates the electric potential signal based on the information sequences, and the electric pulse acts on the multilayer and dissolves the Zr^{4+} film layer on the surface, so the shortest DNA strands were released if the current signal level is high. To sum up, OOK modulation is used at the transmitter for simplicity. Based on this mechanism, the Tx can release certain number N_c of corresponding DNA strands per slot according to the information sequence to be transmitted.

Furthermore, it is assumed that the Tx transmits N bits of information sequence by controlling the release of N kinds of DNA molecules with different lengths l measured in kilo base pairs (kbp). The initial bit releases the DNA strands which length is l_1 , then the length of DNA molecules representing each bit is fixed to increase σ in turn. According to the arrival probability curve of diffusive channel, the molecules released at the previous 8-th slot has little effect on the current slot, so the length of DNA molecules in each time slot circulates once every 8 slots, and the length of n-th DNA molecules is shown as (1) (\backslash represents the integer modular operation). Tx and Rx are synchronized, and Tx transmits each kind of molecules with the interval of T , that is, Tx releases the n-th DNA molecules at the beginning of the n-th slot $t = nT$.

$$l_n = l_1 + ((n \backslash 8) - 1) * \sigma \quad (1)$$

From the representation above, the main difference in encoding mechanism between DNA-MC system and MC system is the transmitted information is encoded into the length of DNA strands, and if the Rx receives one molecule released in current slot, the Rx will determine the current bit information is "1". While in MC system, the Rx needs to detect the molecular concentration continuously, and if the concentration reaches the threshold, the current bit information will be judged as symbol "1". Therefore, the Tx is required to release more molecules to ensure the concentration achieve the threshold, and with the increase of releasing molecules, the probabilities of nanopores are occupied by the interfering molecules become larger, and DNA-MC system is more resistant to the inter-symbol interference (ISI) caused by the molecules released in previous slots.

2.1.2 Diffusive channel

The DNA-MC mechanism is operating on the body fluid environment, and the movement of molecules can be described as simple Brownian motion. During the diffusion, no extra energy is required, and the molecules can hardly lead to toxic reaction of body organs which is consistent with the actual needs of the whole system. Obviously, the randomness of diffusion motion has a great influence on the information transmission efficiency between the receiver and the transmitter, and the diffusion motion of different information molecules is not the same. Therefore, it is necessary to find a universal model to represent the successful arrival of information molecules at the receiving end.

In the diffusive environment, there is a general mathematical model to describe the relation between the solute density $\rho(r,t)$ and the solute molecular flux $J(r,t)$, that is the average net increase of the solute molecules inside an infinitesimal volume element equals to the average net influx of the molecules through the element's six sides during the time interval.

$$\begin{aligned} & \rho(r,t+dt) \cdot dx dy dz - \rho(r,t) \cdot dx dy dz \\ &= (J_x(r,t) \cdot dy dz \cdot dt - J_x(r+\hat{x} dx,t) \cdot dy dz \cdot dt) \\ &+ (J_y(r,t) \cdot dx dz \cdot dt - J_y(r+\hat{y} dy,t) \cdot dx dz \cdot dt) \\ &+ (J_z(r,t) \cdot dx dy \cdot dt - J_z(r+\hat{z} dz,t) \cdot dx dy \cdot dt). \end{aligned} \quad (2)$$

In (2), dt, dx, dy, dz are all infinitesimals, let these variables approaches zero, we can conclude the continuity equation:

$$\frac{\partial \rho(r,t)}{\partial t} = - \left(\frac{\partial J_x(r,t)}{\partial x} + \frac{\partial J_y(r,t)}{\partial y} + \frac{\partial J_z(r,t)}{\partial z} \right) \quad (3)$$

According to the Fick's Law, the diffusion equation can be derived from the continuity equation, therefore, the Fick's Law asserts that:

$$J_u(r,t) = -D \frac{\partial \rho(r,t)}{\partial u} \quad (u = x, y, z) \quad (4)$$

D is the diffusive coefficient, then, substituting (4) into the right side of (3), we can obtain the diffusion equation:

$$\frac{\partial \rho(r,t)}{\partial t} = D \left(\frac{\partial^2 \rho(r,t)}{\partial x^2} + \frac{\partial^2 \rho(r,t)}{\partial y^2} + \frac{\partial^2 \rho(r,t)}{\partial z^2} \right) \quad (5)$$

the average density of solute molecules is only related to the one-dimensional distance x and time t , and the density satisfy the following equation [25]:

$$\frac{\partial \rho(x,t)}{\partial t} = D \frac{\partial^2 \rho(x,t)}{\partial x^2} \quad (6)$$

The solution for one-dimensional diffusion equation (6) must satisfy three conditions, those are (7)(8)(9),

$$\rho(x,0) = N \delta(x) \quad (7)$$

$$\int_a^b \rho(x,0) dx = N \quad (8)$$

$$\rho(\pm\infty,t) = 0 \quad (9)$$

(7) is the initial condition, N represents the total number of solute molecules close together at $x=0$, and δ is the Dirac delta function, (8) represents the total solute molecules at time 0 inside any interval (a,b) equals to N . The two boundary conditions for this problem are (9). Hence, the probability of molecules arriving at a certain point x after time t can be calculated as [24]:

$$g_d(x,t) = \frac{1}{\sqrt{(4\pi Dt)^3}} \exp\left(-\frac{x^2}{4Dt}\right) \quad (10)$$

D is the diffusive coefficient which represent the diffusive velocity and direction of molecules, and it is positively correlated with the length of the DNA strands [26]. Hence, this formula can calculate the probability of the current released DNA molecules reaching the Rx

at the end of the time slot. In this model, the arrival probability of DNA molecules has a great influence on the system performance, because it can be treated as the successful transmission probability of encoded information. Due to the randomness of diffusion, DNA molecules released from previous time slots often arrive in this time slot, and the formula can be used to evaluate the impact of this situation on the transmission of information molecules.

For instance, given the transmission distance $x = 8\mu m$, according to the formula, although the arrival probability of DNA molecules in the previous slots decreases with time, there is still a relatively high probability that the DNA molecules in the previous three slots will arrive at the Rx in the current slot. It should not be ignored the situation that these unwanted DNA molecules will occupy the receiving resources at the Rx. In our model, three significant factors, namely the length of the time slot interval (T), the number of molecules released (N_c) and the transmission distance (x) should be taken into account, and these factors have a direct impact on the occupancy of receiving resources at the Rx. In the simulation part, the range of the three variables can be obtained when the channel capacity is maximized.

2.1.3 Receiver

As shown in the Fig. 1, the transmitted information is encoded as the length of the DNA molecules. In order to efficiently decode the transmitted DNA molecules arriving at the Rx, the nanopore structure is introduced into this system, and the Rx is composed of several individual nanopores. The nanopore is negatively charged, assuming that, the molecules which arriving at the Rx side can pass through the Rx in a specific direction one by one. During the translocation time, the nanopores are blocked by DNA strands, so the current in Rx side will drop to a lower level. Therefore, the Rx can detect the duration of the current of lower level to decide which kind of DNA molecules are passing through, and this process completes the transition from length information to electrical information.

Because the length of DNA strand released in each time slot is fixed, and the length of DNA strands is increased with the number of time slots, the duration of blocked time should be increased. Based on this mechanism, the Rx can recognize the DNA molecules corresponding to the current slot regardless of the influence of the released molecules in the previous slot.

2.2 Reception index

As illustrated in Fig. 1, the receiving resource of Rx is limited. Too few molecules of current slot and too many molecules of previous slot arriving may cause great influence on the channel capacity. As discussed above, the length of the time slot interval (T), the transmission distance (x) and the number of molecules released (N_c) are the main factors that causes the change of molecular arrival probability. In this part, we will discuss how these three variables affect the arrival probability and the trade-off relationship between them.

Length of time slot interval: In DNA-MC system, if the one DNA molecule of the current slot has arrived the Rx, and been detected correctly, Rx will automatically determine that the current bit information is "1". This is quite different from the traditional molecular communication, cause the information is encoded into the concentration of Rx side in traditional situation, that means the Rx needs to detect new arrival information molecules constantly [27]. Obviously, the residual information molecules of the previous slots can affect the detection process. Under this mechanism, the value of arrival probability is the maximum value during the whole slot time.

According to the arrival probability formula, the probability of the arrival situation of

previous three time slots in current time slot should not be ignored. If the setting of the length of time slot interval T is too small, the probability of the previous slot reaching the Rx will increase, which will occupy the receiving resources of the current slot. On the contrary, if T is set too large, it surely will reduce the influence of previous slot molecules, but at the cost of time resource, which will hinder the improvement of channel capacity.

Transmission distance: When we discuss the transmission distance between the transmitter and the receiver, we should consider the actual scenario of DNA molecular communication, such as the relative size between the transmission distance and the designed transceivers. When the order of magnitude of the transmission distance is determined, there is still a trade-off relationship about the distance. Both the information and interfering molecules' arrival probability at Rx will increase when the transmission distance is too small, and the increase of interfering molecules may degrade the reception performance. Moreover, with the same whole transmission distance, the demand for the system's transceiver resources will increase. Otherwise, when the transmission distance increases, the influence of the reduced arrival probability of information molecules on the transmission efficiency should also be considered.

Number of releasing molecules: When the time slot length and transmission distance are fixed, the probability of a single molecule arriving at the receiver is about 10^{-5} . In order to ensure that each bit of information transmission is successful, it is assumed that the Tx released a fixed number of molecules in each time slot. When the number of transmitting molecules is too small, the information of the current bit may be missed because the molecules do not arrive. When the number of transmitting molecules is too large, the previously transmitted interfering molecules will occupy the receiving resources. Thus, it is of great significance to select the appropriate number of transmitting molecules. In the simulation part, the numerical results will give the number of molecules sent by the Tx and the channel capacity when the information transmission efficiency is maximum.

3. Channel capacity analysis

Due to the system model, in diffusive body fluid environment, the probability of one molecule arrives at the certain point x in a fixed time t can be described as (10). Moreover, the DNA molecules need to be decoded successfully, namely, the translocation time of various DNA strands should be recorded correctly. Assuming that, the length increment of DNA molecules released from adjacent time slots is 2 kbp, and the initial shortest DNA strand length is set to be 1 kbp [7], so the length of DNA molecules represented by the m -th bit can be expressed as $N_{pair} = (1 + 2m) \text{ kbp}$. Moreover, the length of each base pair is $l_{per} = 0.34 \mu\text{m}$, and the translocation speed of DNA molecules in Rx side is set to be $v_{trans} = 0.015 \text{ nm} / \mu\text{s}$ in an ideal situation [28]. Thus, the time for the m -th bit DNA strand to pass through the nanopore can be expressed as (11).

$$t_p = \frac{N_{pair} * l_{per}}{v_{trans}} \mu\text{s} \quad (11)$$

Considering the limitation of receiver operation accuracy, the molecular translocation time will not remain constant t_p . The actual translocation time which is less than t_p conforms to Gaussian probability distribution, as shown in (12). Based on the assumption of increasing translocation time before, the wrong reception probability of the current slot is the integral of $f(t)$ from 0 to t_p [29][30].

$$f(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-t_p)^2}{2\sigma^2}} \quad t \leq t_p \quad (12)$$

Therefore, the process of information transmission consists of two parts: molecular diffusion and decoding. The probability of successful information transmission of one molecule is: $q(t) = g(x, t) \cdot f(t)$. The multiple nanopores structure is applied in the Rx side, suppose the Rx contains N_n nanopores. If one nanopore receives and decodes the current molecules successfully, it is determined that the Rx has successfully received the information of the current slot. The probability of successful transmission of Rx is:

$$q_m = 1 - (1 - q(t))^{N_n} \quad (13)$$

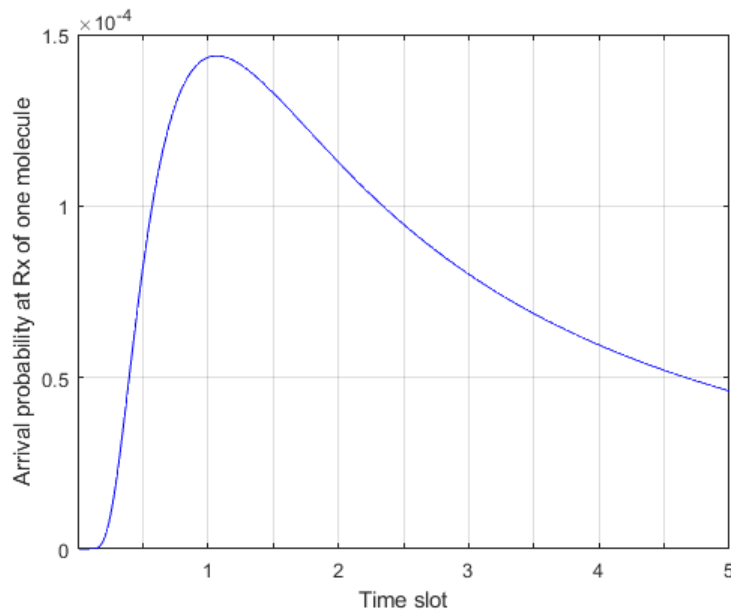


Fig. 2. The probability of one molecule arriving at Rx in different time slots

According to [3], **Fig. 2** depicts the probability of one molecule arriving at the Rx in different time slots when the transmission distance is fixed at $8 \mu m$, and obviously, the arrival probability reaches the peak at the end of the first slot. The magnitude of the peak value P_0 is only 10^{-4} yet, and the probability is too small to satisfy the demands of reliable communication between the transceivers. To increase the probability of information molecules arriving at the Rx, the Tx is required to release N_c number of DNA molecules at the beginning of each slot.

Nevertheless, due to the randomness of Brownian motion, the diffusive time per molecule is not fixed, and the molecules released in previous slot may reach the Rx in current slot. As illustrated in **Fig. 2**, although the arrival probability begins to decline after the first slot, the probability of the molecule arriving at the Rx after the first slot is not small enough to be ignored. Hence, we set a threshold P_{th} , when the arrival probability of the interfering molecules reaches this threshold, we need to consider the occupation of the receiving resources caused by such interfering molecules. P_{th} is an optimizable variable, and the parameter setting of P_{th} is supposed to satisfy the system demand of sensitivity towards the occupancy of nanopores. After the simulation data test, here we take $P_{th} = 0.4 \cdot P_0$.

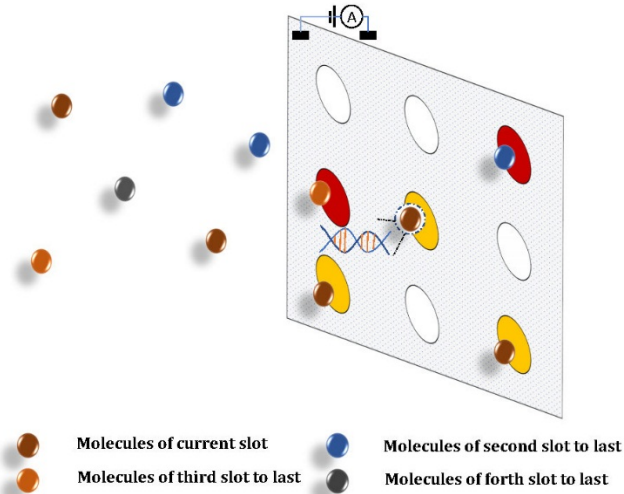


Fig. 3. The occupancy of nanopores caused by interfering molecules at Rx

As depicted in Fig. 3, these brown dots represent the information molecules released by current time slot, and the other colored dots represent the interfering molecules. The Rx uses several nanopores to detect the information molecules, the yellow colored nanopores indicate the successful detection of information molecules, and the red colored nanopores indicate the occupancy of receiving resources caused by interfering molecules. According to the threshold P_{th} , there are three main types of interfering molecules, thus, the exponent of the formula of successful reception can be modified to:

$$q_m = 1 - (1 - N_c \cdot q(t))^{(N_n - \sum_{P_i \geq P_{th}} n_i)} \tag{14}$$

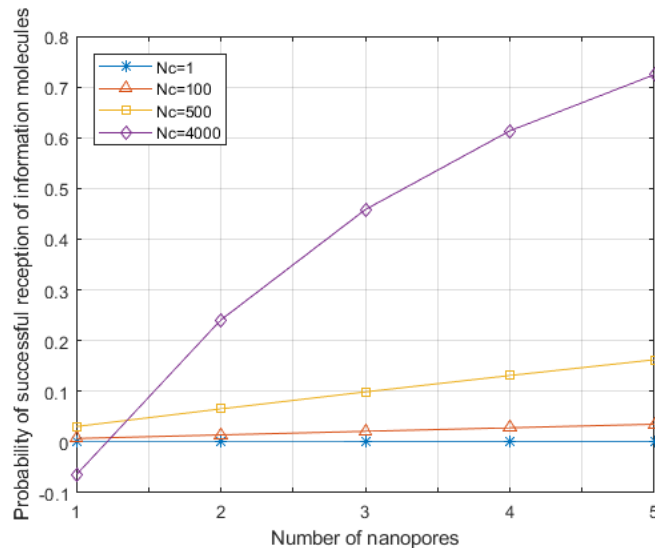


Fig. 4. The probability of successful reception considered the influences of occupancy

Here, n_i represents the number of molecules of the previous three slots arriving at the Rx in the current slot, and the value of $N_n - \sum_{P_i \geq P_{th}} n_i$ represents the number of nanopores that can be

used to detect information molecules in current slot. Fig. 4 depicts the probability curve of successful reception, when the number of nanopores is set to "1" and the number of releasing molecules equals to 4000, the value of probability of successful reception is negative. It represents the situation that numerous interfering molecules occupy the nanopores at Rx, and there is no remaining nanopores to detect information molecules.

As introduced in the system model, the OOK modulation is adopted in the Tx. For the transmitted information of each bit, there is a priori probability p to send symbol "1" and $1 - p$ to send symbol "0". Assuming that the Tx and Rx are time synchronized and only one bit information is transmitted in each time slot. Moreover, the probability for the Rx receipts and decodes properly is P and Q , that is:

$$P[Y_n = 0 | X_n = 0] = P_n \quad (15)$$

$$P[Y_n = 1 | X_n = 1] = Q_n \quad (16)$$

Thus, for an information sequence of length n , the probability of successful bit transmission can be expressed as [5]:

$$Q_n = 1 - \prod_{i=1}^n (1 - pq_i) \quad (17)$$

$$P_n = \prod_{i=1}^{n-1} (1 - pq_i) \quad (18)$$

Here, q_i represents the probability of successful transmission q_m which is calculated in (14). Based on the information theory, the information transmission capacity C is the maximum of mutual information.

$$C = \max\{I(X;Y)\} = \max\{H(Y) - H_Y(X)\} \quad (19)$$

Considering the characteristics of DNA-based molecular communication system, the mutual information for each slot can be expressed as [31]:

$$I(X;Y) = H(Y) - H(Y|X) = \chi((1-p)P_n + p(1-Q_n)) - p\chi(Q_n) + (1-p)\chi(P_n) \quad (20)$$

Moreover, if the length of information sequences $n > 2$, we can conclude the results of $I(X_n;Y_n)$:

$$I(X_n;Y_n) = \chi\left((1-p)\prod_{i=1}^{n-1}(1-pq_i) + p\left(1 - (1 - (1 - q_0)\prod_{i=1}^{n-1}(1 - pq_i))\right)\right) - \left\{p\chi(1 - (1 - q_0)\prod_{i=1}^{n-1}(1 - pq_i)) + (1-p)\chi\left(\prod_{i=1}^{n-1}(1 - pq_i)\right)\right\} \quad (21)$$

where $\chi(\xi) = -\xi \log_2 \xi - (1 - \xi) \log_2 (1 - \xi)$. Moreover, to analyze the influence of time slot length T on transmission efficiency, the channel capacity can be further deduced as follows:

$$C = \max \sum_{i=1}^n \frac{I(X_i;Y_i)}{n} \text{ bits / slot} \quad (22)$$

$$C = \frac{1}{n \times T} \sum_{i=1}^n \max I(X_i;Y_i) \text{ bps} \quad (23)$$

The simulation results are based on this capacity derivation.

4. Numerical Results

In this paper, the simulation is mainly focused on the main factors which affect the system performance of DNA-based molecular communication, such as: the priori probability of sending 1, transmission distance, the number of released molecules and nanopores, etc. The numeral results show the positive and negative effects on channel capacity of each factor, and the peak value is obtained assuming that other factors are constant. Moreover, the simulation also reveals the relationship between the DNA-MC system and the traditional molecular communication system, and the numeral results show that the DNA-MC system has better performance than the traditional molecular communication system from the perspective of transmission distance and transmission resource saving.

4.1 Performance analysis of DNA-based molecular communication

In Fig. 5, the channel capacity curve is computed based on a set of parameters, and the values are set as follows. The transmission distance λ is set to $8 \mu\text{m}$, the length of each time slot T is set to 10 s, the diffusion coefficient D which is related to the length of DNA strands is set to $1 \mu\text{m}^2/\text{s}$ for simplicity, the probability of sending symbol “1” (p) is set to 0.5, and the number of time slot is set to 10, and number of nanopores N_n at Rx is set to 10. The curve is obviously convex, and the capacity of n-th slot is obtained when p is set around 0.5 which is consistent with the common channel capacity conclusion. When the Tx releases an increasing number N_c of DNA molecules, the Rx has a greater probability of receiving information molecules, so the channel capacity increases when p is fixed. Moreover, the channel capacity is easily affected by the number of transmission molecules. If Tx releases a large number of molecules each time, the molecules released in the previous slot will occupy the receiving resources of current slot, accordingly, the curve peak moves slightly to the left which means a smaller probability of releasing molecules.

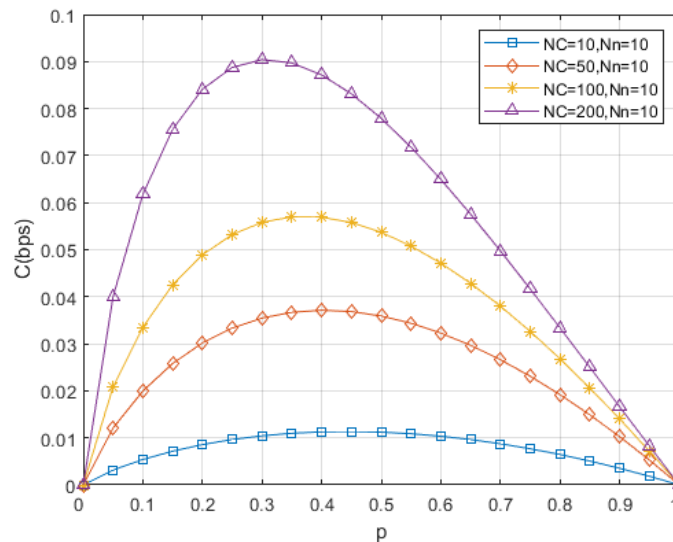


Fig. 5. The channel capacity curve when the probability of transmitting “1” (p) changes (with different N_c)

To get the results of Fig. 6, the N_c and N_n are set to 400 and 10 respectively, and other parameters are set as before. Although both Fig. 5 and Fig. 6 depict the relationship between transmission capacity and prior probability, the different descent speed of the curve indicates the extent to which transmission distance λ affects capacity. Take the peak point of the curve for comparison, it's obvious that the upper bound of transmission capacity that the system can reach gradually decreases with the increase of transmission distance. Moreover, the parameter of N_c is set slightly redundant, when p approaches 1, which means that the transmitter is more likely to release information molecules, that is, the receiver is more likely to accumulate redundant molecules. According to the results, when the transmission distance is set $8 \mu m$ and $10 \mu m$, the capacity declines more gently. When the transmission distance is set to $4 \mu m$ and $6 \mu m$, the peak point of the curve is obviously shifted to the direction of $p = 0$, and the transmission capacity decreases sharply along the direction of $p = 1$, which indicates that a large number of redundant molecules remain at the receiving end, causing interference to information molecules of the current slot. Declining the transmission distance will increase the peak capacity, but it will decrease more sharply when p is larger, and this is a trade-off. Thus, it is necessary to specify the transmission distance between the Tx and the Rx according to the characteristics of the transmission information.

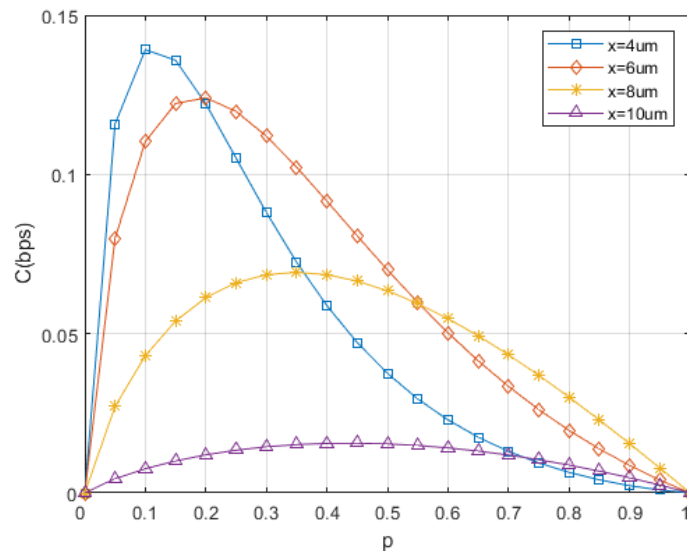


Fig. 6. The channel capacity curve when the priori probability (p) changes (with different λ)

In Fig. 7, λ is set to $5 \mu m$, N_c is set to 200, and other parameters remain the same. At Rx, the information molecules are detected by several nanopores, and this picture has revealed the relationship between N_n and transmission capacity. It is obvious that the peak value of transmission capacity increases about three times when $N_n = 13$ compared with that of $N_n = 8$, and by adding several nanopores, the transmission capacity can be promoted by several times. However, in (14), N_n is the exponential part of the equation of successful reception, so it can multiply the occupation impact of interference molecules at Rx. Thus, the transmission capacity decreases faster along the direction of p approaching 1 when $N_n = 13$, and this is a trade-off. Similarly, the number of nanopores at Rx should be determined based on the characteristics of the transmitted information to achieve the optimal transmission efficiency.

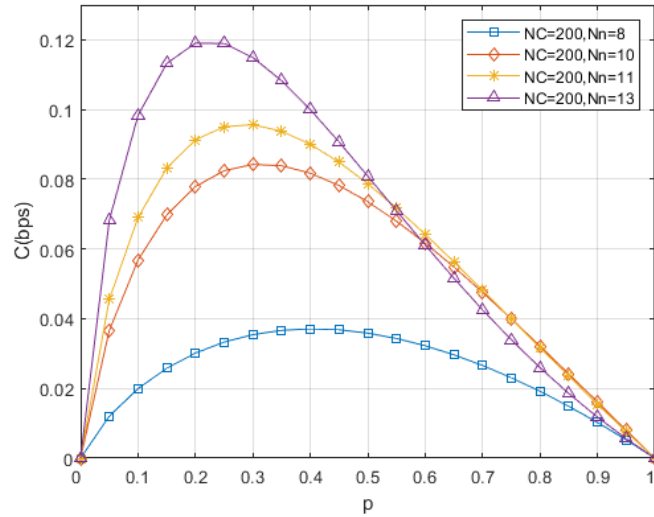


Fig. 7. The channel capacity curve when the priori probability (p) changes (with different N_n)

The results of **Fig. 8** are obtained under the set conditions of the N_c and N_n are set to 100 and 10 respectively, and other parameters are set as before. When the length of time slot \mathcal{T} is set to constant, the transmission capacity decreases as the distance increases. As the transmission distance increases, according to (10), the molecules have less chance of reaching the Rx, and the capacity declines due to the uncertainty of molecular arrival. When the transmission distance is longer than $15 \mu\text{m}$, the value of capacity remains very low and has little change with \mathcal{T} , this is owing to the limitation of the accessibility of diffusion motion, which indicates that the suitable transmission distance of DNA-based molecular communication must be controlled within $15 \mu\text{m}$. Moreover, when the transmission distance is longer than $12 \mu\text{m}$, the increase of the \mathcal{T} can rise the arrival probability slightly, then further enhances the capacity. However, the increment is very small at the expensive of time effectiveness. To sum up, it is not feasible to rise the transmission capacity by increasing the length of time slot in DNA-MC system.

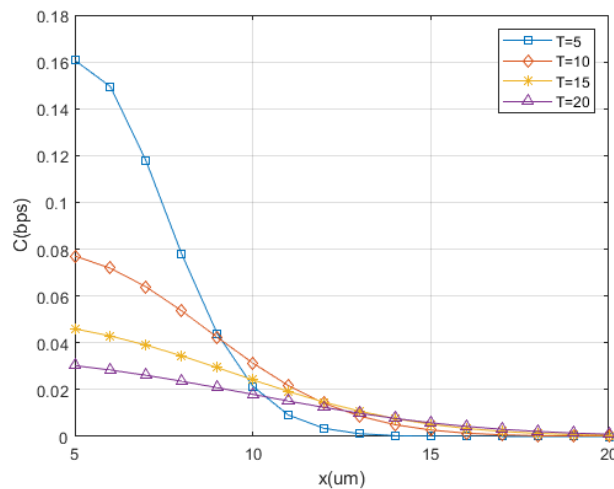


Fig. 8. The channel capacity curve when the transmitting distance changes

In Fig. 9, λ equals to $8 \mu m$, other parameters are set as before. From this picture, regardless of the number of nanopores N_n , the capacity first increases to the peak and then decreases to a stable value, and the decline can be explained by the nanopores occupancy caused by the interfering molecules reduces the probability of successful detection of current slot. With the increase of N_n , the capacity gradually reaches and stabilizes at about 0.08 bps, and the larger the N_n is, the lower number of molecules the peak shifts to. This is because the increasing of N_n can promote the sensitivity of the capacity to the change of the number of molecules.

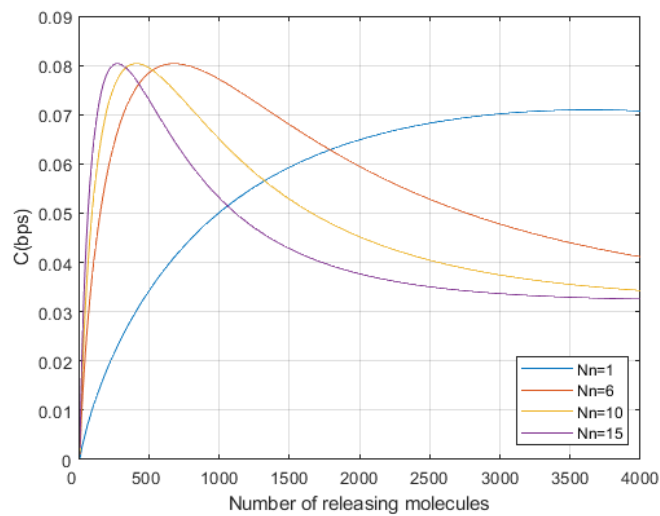


Fig. 9. The transmission capacity curve when the number of releasing molecules changes

4.2 Comparison between DNA-MC system and MC system

DNA-MC and MC mechanism are easy to distinguish in system structure, In DNA-MC system, the transmitted information is encoded into the length of DNA strands, while the MC system uses the molecular concentration at Rx to transmit information, and there are many improvement schemes in this field, such as Zebra-CSK coding mechanism, to significantly reduce the impact of ISI [32]. However, in order to maintain the same molecular utilization, this paper only compares the performance with the original CSK coding system. The length-based system adopts nanopore structure in Rx to decode information, and the other adopts multiple ligands to bind the molecules. Because both systems operate in the body fluid environment, and it's of great significance to compare the performance of the two from the perspective of transmission efficiency and resource utilization. Then, we will compare the transmission capacity of the two systems when transmitting the same number of information molecules, because the two systems adopt different structure of the transceivers, this part does not consider the influence of the system energy consumption on the channel capacity, and the enhancement degree of the capacity when increasing the number of released information molecules.

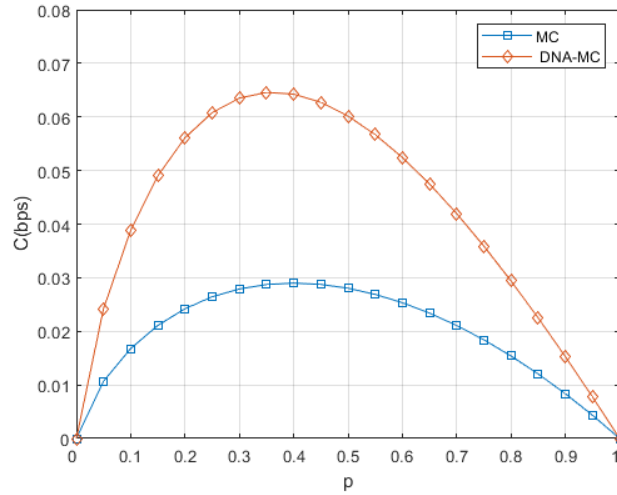


Fig. 10. The transmission capacity of MC and DNA-MC when priori probability p changes

In **Fig. 10**, the MC system modeling is different from DNA-MC, and some parameters are only applied in the MC system, based on [31], those are: the releasing rate of information molecules at Rx $k_{-1}=0.08$, the concentration of receptors at Rx $N=1 \mu\text{mol}/l$, the radius of Rx $r_0=0.8 \mu\text{m}$. Moreover, the common parameters are set as follows: $D=1 \mu\text{m}^2/s$, $x=8 \mu\text{m}$, $N_c=100$, and the number of time slot $M=10$. The capacity curves of the two system are both upwards convex, and in DNA-MC system the peak value appears at $p=0.35$ while the peak value is obtained at $p=0.4$ in MC system. The peak value of both curves shifted slightly to the left, indicating that the two systems were sensitive to the interferences caused by molecules sent in previous slots, and reducing the probability of sending molecules or increasing the interval of releasing different information molecules will promote the transmission capacity. The capacity of DNA-MC is always greater than MC, and the peak value can reach 0.065 bps. From this picture, we observe that the DNA-MC can achieve higher transmission efficiency with less information molecules.

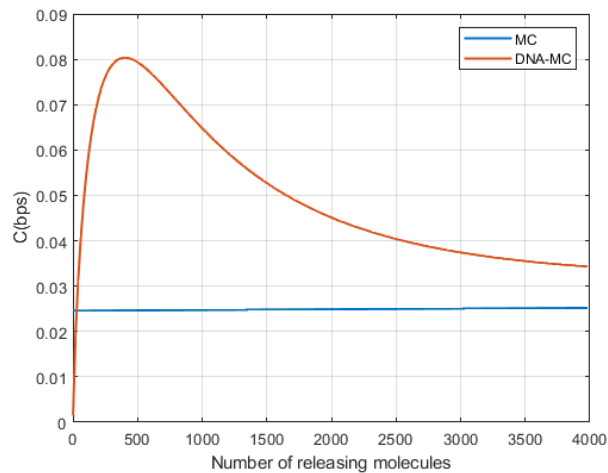


Fig. 11. The transmission capacity of MC and DNA-MC when the number of releasing molecules changes

In Fig. 11, the priori probability P is set to 0.5, and other parameters are set as the same as Fig. 10, and this picture describes the relationship between the transmission capacity and the number of releasing information molecules in both systems. Take DNA-MC system as example, when the number of released information molecules increases from 0 to 400, the transmission capacity just rises to the peak value. If the number of molecules continues to increase, the capacity gradually decreases instead. This is because when Tx releases few number of molecules, increasing the number can promote the probability of molecules arriving at the Rx. However, if the number continue to rise, the number of interference molecules reaching the Rx will also increase. These redundant molecules will occupy the nanopores at Rx, and it will degrade the performance on transmission capacity. On the other hand, in MC system, as the number of releasing molecules increases from 0 to 4000, the transmission capacity increases by 0.0006 bps. This shows that the effect of increasing the number of sending molecules on the capacity is not obvious at this order of magnitude. Although the performance of DNA-MC system is not stable under this condition, it still has advantages over MC system.

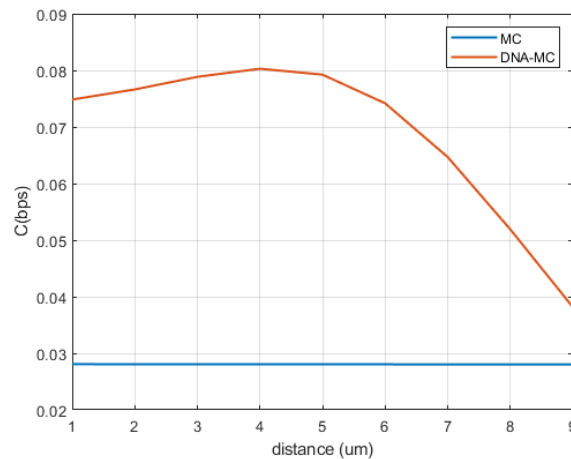


Fig. 12. The transmission capacity of MC and DNA-MC when the transmission distance changes

In Fig. 12, the number of releasing molecules N_c is set to 100, and the other parameters setting is the same as before. The simulation results show the relationship between the capacity and transmission distance. In DNA-MC system, when the transmission distance is less than $4 \mu m$, the number of molecules sent by the Tx is redundant for the Rx. Thus, with the increase of the distance, the capacity will also increase. When the transmission distance is longer than $4 \mu m$, increasing the distance will significantly degrade the probability of information molecules arriving at the receiving end, then degrading the transmission capacity. In addition, through observation, there is a fixed match between the number of molecules released and the transmission distance, so the optimal release number can be set according to the transmission distance. In MC system, when the distance is increased from $1 \mu m$ to $9 \mu m$, the capacity is reduced by 0.0001 bps. This phenomenon is mainly because of the too few molecules, and the capacity is not sensitive to the change of transmission distance. Obviously, no matter how long the transmission distance is, the capacity of DNA-MC system is always better than that of MC system.

5. Conclusion

This article has proposed a novel model for DNA-based molecular communication. The system is composed by three parts, e.g., transmitter (Tx), diffusive channel, receiver (Rx). The Tx is a multi-layer golden foil, and the release of DNA molecules in each layer is controlled by the information transmitted per bit, then the molecules reaches the Rx through the diffusive channel, and the Rx decodes each bit of information by detecting the length of the arriving DNA molecules per slot through the nanopores. There are some main factors e.g., the number of releasing molecules (N_c) and nanopores (N_n), the transmission distance (x), the length of time slot (T), which have a significant influence on the probability of molecules arriving the Rx and being detected correctly. According to the equation of q_m , the transmission capacity can be derived, and the simulation results reveal the trade-off relationship between each factor and the transmission capacity. Moreover, we have finished the comparison of transmission capacity between DNA-MC system and MC system. Since the concentration of molecules at the Rx is not used as the information carrier in the DNA-MC system, the equivalent messages can be transmitted through fewer molecules. The numerical results show the transmission capacity of DNA-MC system has an obvious advantage over MC system, and the capacity can reach 0.08 bps when the number of releasing molecules is set to 100.

This article provides reasonable suggestions for the parameters setting of DNA-MC system in body fluid environment, and the future work is mainly focused on the channel modeling of DNA-MC system which applies the base sequence coding, and the capacity analysis of a multi-transceiver system.

Acknowledgement

This paper was supported in part by the National Natural Science Foundation of China (Grant No.62071101), the Fundamental Research Funds for the Central Universities (Grant No. ZYGX2019J001) and the Special Science Foundation of Quzhou (No. 2020D002).

References

- [1] W. Guo, C. Mias, N. Farsad and J. Wu, "Molecular Versus Electromagnetic Wave Propagation Loss in Macro-Scale Environments," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 1, no. 1, pp. 18-25, March 2015. [Article\(CrossRef Link\)](#)
- [2] Li J, Esteban-Fernández de ávila, Berta, Gao W, et al., "Micro/nanorobots for biomedicine: Delivery, surgery, sensing, and detoxification," *Science Robotics*, 2(4), eaam6431, 2017. [Article\(CrossRef Link\)](#)
- [3] T. Nakano, T. Suda, Y. Okaie, M. J. Moore and A. V. Vasilakos, "Molecular Communication Among Biological Nanomachines: A Layered Architecture and Research Issues," *IEEE Transactions on NanoBioscience*, vol. 13, no. 3, pp. 169-197, Sept. 2014. [Article\(CrossRef Link\)](#)
- [4] T. Luo, R. Zheng, J. Song, L. Lin and H. Yan, "A Small-scale Modulator of Electric-to-biological Signal Conversion for Synthetic Molecular Communications," in *Proc. of ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, Dublin, Ireland, pp. 1-7, 2020. [Article\(CrossRef Link\)](#)
- [5] T. Nakano, Y. Okaie and J. Liu, "Channel Model and Capacity Analysis of Molecular Communication with Brownian Motion," *IEEE Communications Letters*, vol. 16, no. 6, pp. 797-800, June 2012. [Article\(CrossRef Link\)](#)

- [6] B. Tepekule, A. Pusane, et al., "ISI Mitigation Techniques in Molecular Communication," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 1, no. 2, pp. 202-216, June 2015. [Article\(CrossRef Link\)](#)
- [7] B. Bilgin, E. Dinc and O. Akan, "DNA-Based Molecular Communications," *IEEE Access*, vol. 6, pp. 73119-73129, 2018. [Article\(CrossRef Link\)](#)
- [8] G. Church, Y. Gao, S. Kosuri, "Next-Generation Digital Information Storage in DNA," *ence*, 337(6102), 1628, 2012. [Article\(CrossRef Link\)](#)
- [9] N. Goldman, P. Bertone, C. Chen, et al., "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA," *Nature*, 494(7435), 77-80, 2013. [Article\(CrossRef Link\)](#)
- [10] N. Bell, U. Keyser, "Digitally encoded DNA nanostructures for multiplexed, single-molecule protein sensing with nanopores," *Nature Nanotechnology*, 11, 645-651, 2016. [Article\(CrossRef Link\)](#)
- [11] K. Chen, M. Juhasz, F. Gularek et al., "Ionic Current-Based Mapping of Short Sequence Motifs in Single DNA Molecules Using Solid-State Nanopores," *Nano Letters*, 17(9), 5199-5205, 2017. [Article\(CrossRef Link\)](#)
- [12] N. Bell, U. Keyser, "Direct measurements reveal non-Markovian fluctuations of DNA threading through a solid-state nanopore," 2016.
- [13] N. Bell, M. Muthukumar, U. Keyser, "Translocation frequency of double-stranded DNA through a solid-state nanopore," *Physical Review E*, 93, 2016. [Article\(CrossRef Link\)](#)
- [14] F. Wang, X. Liu, G. Li, et al., "Selective electrodisolution of inorganic ions/DNA multilayer film for tunable DNA release," *Journal of Materials Chemistry*, 19(2), 286-291, 2009. [Article\(CrossRef Link\)](#)
- [15] F. Wang, D. Li, G. Li, et al., "Electrodisolution of Inorganic Ions/DNA Multilayer Film for Tunable DNA Release," *Biomacromolecules*, 9(10), 2645-2652, 2008. [Article\(CrossRef Link\)](#)
- [16] T. Drummond, M. Hill, J. Barton, "Electrochemical DNA sensors," *Nature Biotechnology*, 21(10), 1192-1199, 2003. [Article\(CrossRef Link\)](#)
- [17] W. Guo et al., "Molecular communications: channel model and physical layer techniques," *IEEE Wireless Communications*, vol. 23, no. 4, pp. 120-127, August 2016. [Article\(CrossRef Link\)](#)
- [18] N. Farsad, H. B. Yilmaz, A. Eckford, C. Chae and W. Guo, "A Comprehensive Survey of Recent Advancements in Molecular Communication," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1887-1919, thirdquarter 2016. [Article\(CrossRef Link\)](#)
- [19] B. Koo, C. Lee, et al., "Molecular MIMO: From Theory to Prototype," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 600-614, March 2016. [Article\(CrossRef Link\)](#)
- [20] M. Kuscü, E. Dinc, B. A. Bilgin, H. Ramezani and O. B. Akan, "Transmitter and Receiver Architectures for Molecular Communications: A Survey on Physical Design With Modulation, Coding, and Detection Techniques," in *Proc. of the IEEE*, vol. 107, no. 7, pp. 1302-1341, July 2019. [Article\(CrossRef Link\)](#)
- [21] I. F. Akyildiz, et al., "Nanonetworks: A new frontier in communications," *Communications of the ACM*, 54(11), 84-89, 2011. [Article\(CrossRef Link\)](#)
- [22] G. Bafghi, A. Gohari, et al., "Diffusion Based Molecular Communication with Limited Molecule Production Rate," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, 4(2), 61-72, 2018. [Article\(CrossRef Link\)](#)
- [23] J. Huang and C. Lee, "On capacity bounds of two-way diffusion channel with molecule harvesting," in *Proc. of 2017 IEEE International Conference on Communications (ICC)*, Paris, pp. 1-6, 2017. [Article\(CrossRef Link\)](#)
- [24] E. Seitaridou, "Simple Brownian Diffusion: An Introduction to the Standard Theoretical Models," *Physics Today*, 66(9), 50-52, 2013. [Article\(CrossRef Link\)](#)
- [25] X. Bao, Y. Zhu and W. Zhang, "Channel Characteristics for Molecular Communication via Diffusion With a Spherical Boundary," *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 957-960, June 2019. [Article\(CrossRef Link\)](#)
- [26] G. Lukacs, P. Haggie, O. Seksek, et al., "Size-dependent DNA mobility in cytoplasm and nucleus," *The Journal of biological chemistry*, 275(3), 1625-1629, 2000. [Article\(CrossRef Link\)](#)

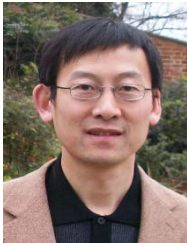
- [27] A. Noel, K. C. Cheung and R. Schober, "Optimal Receiver Design for Diffusive Molecular Communication With Flow and Additive Noise," *IEEE Transactions on NanoBioscience*, vol. 13, no. 3, pp. 350-362, Sept. 2014. [Article\(CrossRef Link\)](#)
- [28] A. Meller, D. Branton, "Single molecule measurements of DNA transport through a nanopore," *Electrophoresis*, 23(16), 2583-2591, 2002. [Article\(CrossRef Link\)](#)
- [29] Q. Liu, H. Wu, L. Wu, et al., "Voltage-Driven Translocation of DNA through a High Throughput Conical Solid-State Nanopore," *PLoS ONE*, 2012. [Article\(CrossRef Link\)](#)
- [30] C. Kong, M. Muthukumar, "Modeling of polynucleotide translocation through protein pores and nanotubes," *Electrophoresis*, 23(16), 2697-2703, 2002. [Article\(CrossRef Link\)](#)
- [31] Q. Liu, K. Yang and P. He, "Channel capacity analysis for molecular communication with continuous molecule emission," in *Proc. of 2013 International Conference on Wireless Communications and Signal Processing*, Hangzhou, pp. 1-6, 2013. [Article\(CrossRef Link\)](#)
- [32] P. Subodh & S. Shin & K. Kwak, "Robust Modulation Technique for Diffusion-based Molecular Communications in Nanonetworks," 1-4, 2015. [Article\(CrossRef Link\)](#)



Jialin Xie has received the B. S.(Highest Hons) degree in communication engineering from Sichuan University, Sichuan, China, in 2019. He is currently pursuing the M. S. degree with the School of Information and Communication Engineering, University of Electronic Science and Technology of China (UESTC), Sichuan, China. He is currently focusing on the topic of the coding method and the channel capacity of DNA based molecular communication. He has attended the Connected and Autonomous Vehicles Design program in the Lady Margret Hall (LMH), Oxford, UK, in 2019. He graduated with Honor Awards from Sichuan Province and Sichuan University in 2019.



Qiang Liu received his B. S. (1996), M. S. (2000) and Ph.D. (2012) degrees from University of Electronic Science and Technology of China (UESTC) respectively. After graduating from M.S. study in 2000, he has worked with School of Communication and Information Engineering, UESTC, China, and is an associate professor now. He had worked in the University of Essex in U.K. (2012.12-2013.12) and the University of California, Davis in U.S.A (2017.8-2018.8) as a visitor scholar. His researches focus on wireless sensor networks, Internet of Things, broadband wireless networks and molecular communication. He is the member of IEEE/IEEE ComSoc. He is the reviewer of IEEE Transactions on NanoBioscience, International Journal of Communication Systems (IJCS).



Kun Yang received his PhD from the Department of Electronic & Electrical Engineering of University College London (UCL), UK. He is currently a Chair Professor in the School of Computer Science & Electronic Engineering, University of Essex, leading the Network Convergence Laboratory (NCL), UK. He is also an affiliated professor at UESTC, China. Before joining in the University of Essex at 2003, he worked at UCL on several European Union (EU) research projects for several years. His main research interests include wireless networks and communications, IoT networking, data and energy integrated networks, and molecular communications. He manages research projects funded by various sources such as UK EPSRC, EU FP7/H2020 and industries. He has published 150+ journal papers and filed 10 patents. He serves on the editorial boards of both IEEE and non-IEEE journals. He is a Senior Member of IEEE (since 2008) and a Fellow of IET (since 2009).



Lin Lin graduated from Nanyang Technological University in Singapore with a doctor's degree, and he graduated from Tianjin University with a bachelor's degree and a master's degree. He completed the Curie scholar research program at Warwick University in the UK, and he is the EU Curie scholar. He chaired one belt, one road project, the European Union H2020 Marie Curie scholar program. He has published more than 70 academic papers in domestic and foreign academic journals (e.g. IEEE Transactions on NanoBioscience) and international academic conferences (e.g. IEEE Globecom). He has served as TPC chair of molecular communication BICT2020 conference, chairman of China Germany molecular communication seminar in 2019. Winner of ACM Shanghai New Star Award. Currently, he is the vice chairman of IEEE Committee on molecular, biological and multiscale communication technology.