The Use of Coding and Protocols Within Molecular Communication Systems

Mark S. Leeson, Matthew D. Higgins, Chenyao Bai, Yi Lu, Xiayang Wang and Ruixiao Yu

Abstract This chapter focuses upon the use of coding and protocols within diffusion based molecular communication systems, laying the groundwork for future development in test bed implementations. The chapter starts with an introduction that briefly discusses coding and protocols used in traditional communication systems. Following this, details of the molecular channel are given, including the energy consumption constraints and a relevant mathematical framework. This discussion then leads onto potential encoding and decoding technologies. Next, original results on the use of Hamming codes in molecular communication systems are presented with a quantitative comparison against an uncoded molecular system. The impact of specific design parameters such as the number of molecules, energy, and transmission distance on the bit error rate (BER) is considered. Finally, a protocol, based upon the use of an acknowledgement (ACK) packet is presented as a further advancement to the field that the reader may wish to consider when designing future systems.

1 Introduction

Today, research on nano-communications can be broadly divided into the following domains [1]: nano-mechanical, acoustic, electromagnetic, and molecular. In the first of these, information exchange between transmitter and receiver is implemented through mechanical contact such as hard junctions between linked devices. In acoustic communication the transmitted information is encoded as ultrasonic waves. This method relies upon the ability to implement ultrasonic transducers which are capable of sensing the rapid variations of pressure and emitting acoustic signals [1] at the nano-scale. Electromagnetic communication transmits information

M.S. Leeson (✉) · C. Bai · X. Wang · R. Yu
School of Engineering, University of Warwick, Coventry CV4 7AL, UK
e-mail: mark.leeson@warwick.ac.uk

M.D. Higgins · Y. Lu
WMG, University of Warwick, Coventry CV4 7AL, UK

through modulated electromagnetic waves which can propagate with minimal losses either along wires or through an air or fluidic medium. However, it is known to be a challenge to integrate the transceivers at the nano-scale due to the relatively complexity and furthermore, assuming that integration was possible, the output power of the nano-transceiver would be insufficient to guarantee a bidirectional communication channel. As a result, it has been postulated that electromagnetic communication could be used to transmit information from a micro-device to a nano-device, but not from nano-machines to micro-machines, nor among nano-machines [1]. This is still an open research area. Finally, in molecular communications the transmitted information is encoded using molecules. This new and interdisciplinary field combines nano-, bio-, and traditional communication technologies and concepts [2]. Thus, the performance of molecular communication systems is affected by the physical laws governing these fields. In molecular communication, the molecular transmitters are able to release specific molecules in response to an internal command. Similarly, the receivers have the ability to react to specific molecules and thus the local concentration of the molecules at the transmitter and the receiver may be used to understand the molecular bit transmitter sent. Generally, it is common to use existing biological nano-techniques, with design traits found in traditional communications systems to construct the molecular communication system and this idea has typically been proven successful. It is this mind-set that this chapter capitalises upon.

Error detection and correction techniques that enable the reliable delivery of information over unreliable channels are essentially ubiquitous to all communications systems found in our daily lives. Such coding techniques are generally based upon the addition of redundant bits into the source coding representations, followed by the transfer of these bits into the channel where noise is present. Upon reception, an appropriate decoding method is used whereby the redundant bits can be used to check whether or not the information is corrupted [3]. Depending upon the type of source coding, the error may be correctable and the original information appropriately restored. After a thorough analysis of the molecular channel in Sects. 2 and 3 of this chapter, it will be shown in Sect. 4 how this traditional element of communication theory, but more specifically, the use of Hamming codes, can be applied to the molecular channel.

In the same way that error correction techniques are commonplace in all modern systems, so too are a form of protocol. The most well-known examples are the Transmission Control Protocol (TCP) and Internet Protocol (IP), with their v4 and v6 incarnations [4]. In Sect. 5 of this chapter, the implementation of a relatively simple protocol, namely Stop and Wait Automatic reQuest (SW-ARQ) [5, 6] is described, which is also compatible with the work on error correction in Sect. 4.

The use of coding and protocol techniques is still in its infancy at the nano-scale with many outstanding challenges. For example, all the functionalities, including encoding and decoding, or the retransmission mechanics, require energy capabilities which are likely to be one of the limits of any nanomachine that is also constrained mechanically. Moreover, transmission and device speeds are extremely slow in comparison with conventional data networks and devices. In addition,

whilst the conceptual design of the communication layer model can be inspired from existing models, the actual implementation of a full protocol stack is also dictated by the availability and capability of each synthetic components used to implement each layer of the stack.

This chapter aims to highlight to the reader how the use of both error correction and protocol can make a significant contribution to the design of any molecular communications system.

2　Analysis of the Diffusive Medium

Modelling molecular propagation in the channel medium is one of the key challenges in predicting molecular communication system performance [1]. The information molecules can be chosen to be proteins, protein complexes, peptides, DNA sequences or other molecular structures [7]. The motion of information molecules is governed by the forces produced by the constant random thermal motion of the molecules within the fluid. To simplify the communication system, the medium is considered to be of extremely large dimensions compared to the size of the information molecules. In addition, collisions between the information molecules are neglected. For this diffusion based molecular communication system, the transmitted information is represented by a sequence of symbols which are distributed over sequential and consecutive time slots with one symbol in each slot. The *intended symbol* refers to the symbol sent by the transmitter and the *received symbol* represents the symbol received by the receiver. The information is encoded by concentration with binary representation. Specifically, if the number of information molecules arriving at the receiver in a given time slot exceeds a threshold τ, the symbol is interpreted as a "1". Otherwise, it will be interpreted as a "0". However, errors may be caused by Intersymbol Interference (ISI), which is an unavoidable consequence of both wired and wireless communication systems and is known to have adverse effects in communication systems, particularly when the system is stochastic [8]. The ISI effect is related to the properties of the medium used, the distance of the symbol propagation and the selection of the threshold value. In the diffusion communication system here, some information molecules may arrive at the receiver after the current time slot according to the diffusion dynamics, which will lead to the possible incorrect decoding of the received symbol in the next time slot.

Here, a three dimensional (3D) diffusion based communication system is to be considered as shown in Fig. 1. The information molecule is at a distance r from the centre of the receiver which has a radius of $R = 5\,\mu m$ [7].

The diffusion coefficient is taken to be $= 79.4\,\mu m^2 s^{-1}$, which is a conservative value for insulin in water at human body temperature [7]. The escape probability $S(r, t)$ in the 3D diffusion medium can be described with the following backward difference equation at a given time t [9]:
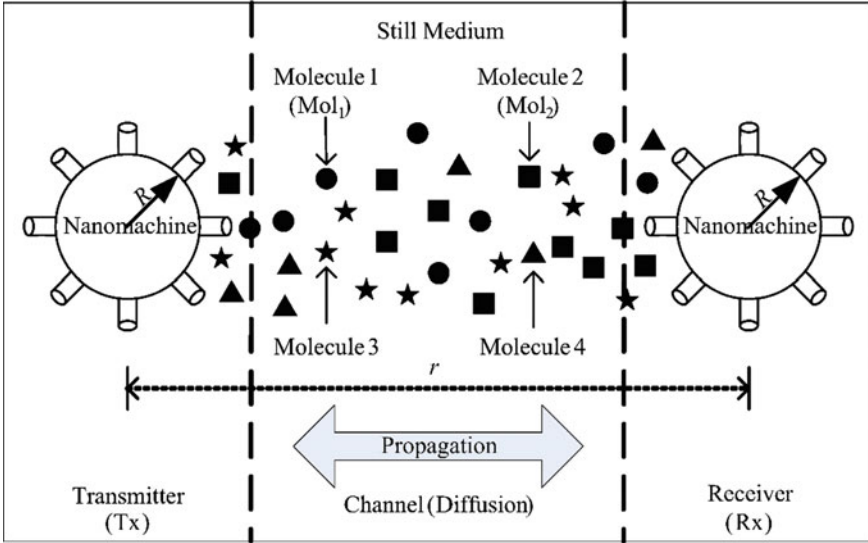
Fig. 1 Model of the 3D diffusion based communication used herein

$$\frac{\partial S(r,t)}{\partial t} = D\nabla^2 S(r,t) \tag{1}$$

For this communication model, the capture probability, rather than the escape probability, is more important. By solving Eq. (1), the capture probability $P_{hit}(r,t)$ can be calculated as:

$$P_{hit}(r,t) = 1 - S(r,t) = \frac{R}{r}\mathrm{erfc}\left\{\frac{d}{2\sqrt{Dt}}\right\} \tag{2}$$

where erfc($\cdot$) is the complementary error function, and $d = r - R$, the distance between the information molecule and the boundary of the receiver.

Figure 2 illustrates that for a given distance, the capture probability increases with the increasing of time, which means that as time increases, more and more molecules will arrive at the receiver. In addition, for a given period of time, the capture probability decreases when the distance between the information molecule and the receiver increases.

To extract the hit time probability, which refers to the probability that an information molecule arrives at the receiver at a particular time t, Eq. (2) is differentiated with respect to time, obtaining the hit time distribution:

$$h(t) = \frac{R}{r}\frac{d}{2\sqrt{\pi D}}\frac{1}{t^{3/2}}\exp\left(-\frac{d^2}{4Dt}\right) \tag{3}$$
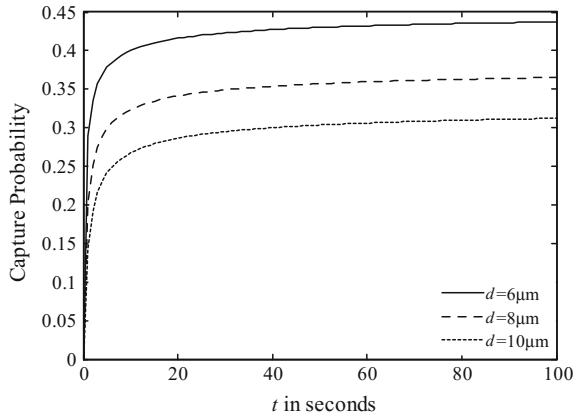
Fig. 2 Capture probabilities versus time for different transmission distances, $d = 6\,\mu m$, $d = 8\,\mu m$ and $d = 10\,\mu m$
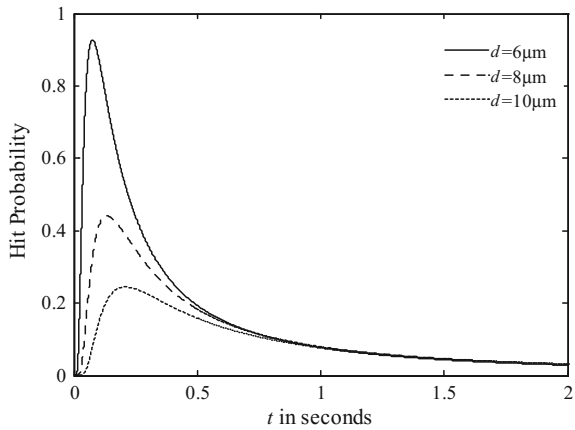
Fig. 3 Hit time distribution plot for different distances $d = 6\,\mu m$, $d = 8\,\mu m$ and $d = 10\,\mu m$

The capture probability shows the probability of successful transmission of the information symbols in a given period of time. To get a better understanding of the probability of success at a specific time, the hit time probability distribution, which encapsulates the response of the molecular channel to the release of an impulse of molecules, is investigated. In Fig. 3, for a given distance, it can be seen that this increases sharply in an extremely short period, it then decreases rapidly and approaches zero as time progresses. This is because that during the whole propagation process, most of the information molecules arrive in a short time whereas a few molecules arrive after a very long period of time, exhibiting a long diffusive tail leading to ISI. In addition, a longer time will be taken for most of the molecules arriving at the receiver when the distance between the molecule and the receiver is larger.

3   Communication Channel Model

Before moving further into the channel model, a brief review the modulation techniques used in molecular communication systems is presented.

Three main modulation techniques have been proposed for use in molecular communication: Concentration Shift Keying (CSK) [10], Molecule Shift Keying (MoSK) [10] and Ratio Shift Keying (RSK) [11]. Here, MoSK and RSK will be outlined before focusing on CSK in more detail. MoSK represents the information using different kinds of molecules, and then the receiver decodes them based on the concentration of each type of molecules during a time slot. In [12], an example is shown using hydrofluorocarbons as the information molecules where the messenger molecules are composed of three parts: header, trail and the chemical bit element, and the information bits is dependent on the form of the chemical element. RSK encodes the information based on the ratios of messenger molecules. Theoretically, it has an infinite modulation order, and needs just two types of molecules in the simplest system; it can also be developed for use in more complex systems. In addition, the authors in [11] proposed a technique which uses isomers, also known as hexoses, as the messenger molecules. The simulation results show that this technique can obtain a higher SNR gain than the techniques in [10, 13].

CSK uses the concentration of the received molecules in a time slot as the criterion. When the number of molecules exceeds a threshold, the receiver gives "1", otherwise it gives "0". It can be viewed as analogous to amplitude shift keying (ASK) in conventional communication systems. Similarly, its naming convention is based on the number of bits per symbol. For example, when the number of bits per symbol is 1, it is denoted as Binary CSK (BCSK) and when the number of bits per symbol is 2, it is denoted as Quadruple CSK (QCSK). These CSK techniques are illustrated in Fig. 4.

Fig. 4 CSK techniques for 1 and 2 bits per symbol
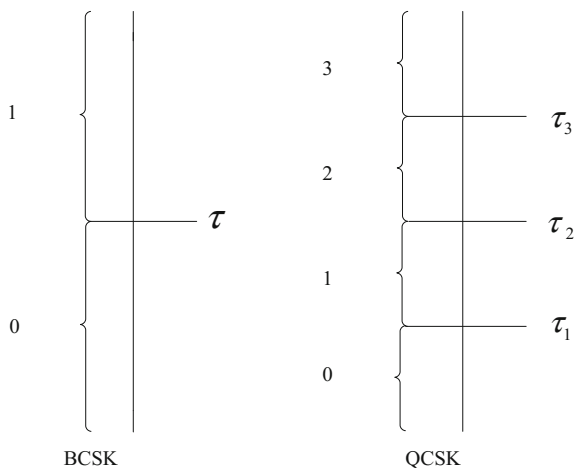
CSK is susceptible to ISI, and the severity of the effect is dependent on the threshold values and also on the number of bits per symbol. As the number of bits per symbol increases then the impact of ISI errors will be on an increasing number of bits simultaneously.

An outline of the modelling of the communication channel used to analyse the performance of the communication method will now be presented. To effectively represent the transmitted symbols, the propagation time is divided into time slots, also called symbol durations, which have the equal length. Only one symbol propagates in a single time slot the length of which is denoted by t_s. As shown in Sect. 2, there is a probability that the molecule will hit the receiver in a given time slot that depends upon the distance between the transmitter and receiver, the symbol duration, the environment and the chosen type of molecule [7].

The communication channel is a Binomial one, where each molecule arrives at the receiver or does not. It has been stated above that the previous bits can have an influence on the current bit due to ISI. Here, only the previous one time slot will be taken into consideration since this has been shown to be a reasonable approximation [7]. Thus the number of molecules received in a time slot which is denoted by N_{hit} is made up of the molecules sent at the start of the current time slot N_c and the start of the previous symbol duration N_{pre}. N_c and N_{pre} are both Binomial distributions which can be denoted as:

$$N_c = Binomial(N, P_{hit}(d, t_s)) \tag{4}$$

$$N_{pre} = Binomial(N, P_{hit}(d, 2t_s)) - Binomial(N, P_{hit}(d, t_s)). \tag{5}$$

Since no closed form is known for the probability mass function of the difference between two dissimilar Binomial distributions, a Gaussian approximation [7] is employed. Given that the one-bit information of the current intended symbol is S_C and that of the previous time slot is S_{Pre}, the current received symbol depends on both of them. For this binary channel model, both S_{Pre} and S_C can be taken as zero or one as shown in Fig. 5.

As may be seen in Fig. 5, there are four different cases for the binary channel model which are denoted by bit pairs {00, 01, 10, 11} for received symbol decoding, according to the different values of S_C and S_{Pre}. The probability represents the probability of success to receive the current intended symbol in the current

Fig. 5 The binary channel model
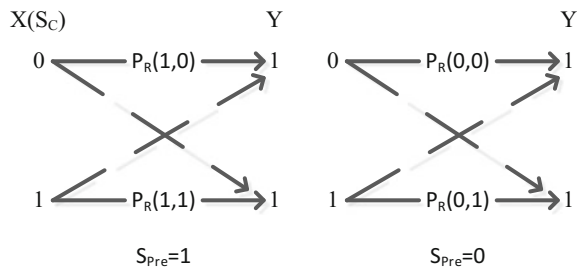
time slot, where S_{Pre} is the one-bit information represented by the previous intended symbol and c is that of the current one. The different four cases are displayed below.

Case {11}: Both the one-bit information represented by the previous intended symbol and current one are "1". In this case, the probability of successfully receiving the current symbol "1" increases because some of the molecules sent at the start of previous time slot will also receive at the current symbol duration. Thus, this is often considered as a favourable case as the ISI adds to the "1". Assuming that $P_1 = P_{hit}(d, t_s)$ and $P_2 = P_{hit}(d, 2t_s)$, the probability of success of this case can be described as:

$$P_{R(1,1)} = Q\left(\frac{\tau - nP_2}{\sqrt{n[P_2(1-P_2) + 2P_1(1-P_1)]}}\right) \tag{6}$$

where $Q(\cdot)$ is the usual tail probability Q-function.

Case {10}: No molecules are sent at the start of the current time slot which means that the received symbol should be "0" if it is correctly decoded. In this case, the molecules overflowing from the previous time slot have a negative influence on the successful decoding of the current intended symbol. It is considered to be an unfavorable case due to ISI. For successful decoding, the received molecules should not exceed the threshold such that:

$$P_{R(1,0)} = 1 - Q\left(\frac{\tau - n(P_2 - P_1)}{\sqrt{n[P_2(1-P_2) + P_1(1-P_1)]}}\right) \tag{7}$$

Case {01}: It means that the number of molecules received in the current time slot is only dependent on the molecules sent at the start of the current time slot, without overflowing molecules from the previous symbol duration. The information will be decoded successfully when the number of received molecules exceeds the threshold. The probability of success can be written as:

$$P_{R(0,1)} = \sum_{k=\tau}^{n} \binom{n}{k} P_1^k (1-P_1)^{n-k} = I_{p_1}(\tau, n - \tau + 1) \tag{8}$$

where $I_{p_1}(\tau, n - \tau + 1)$ is the regularized incomplete beta function.

Case {00}: There are no overflowing molecules from the previous time slot and no molecules are sent at the start of the current time slot. Thus, to decode the information successfully, the number of received molecules should smaller than the threshold. Because the threshold is always greater than zero, the received symbol is always zero in this case, leading to a 100% probability of success.

BER is considered as a key parameter which is often employed to assess the performance of systems that transmit information from one position to another. Various kinds of noise, interference and phase jitter may cause degradation of the transmitted signal. Even a single bit error can render the program inoperable when transmitting a computer program. The total average BER can be obtained by

Table 1 Values for 60% of molecules to arrive

Distance (μm)	t_s (s) in [7]	t_s (s) based on (2)
1	0.03	0.023
2	0.11	0.092
4	0.4	0.37
8	1.54	1.47
16	5.9	5.86
32	22.01	23.45

Fig. 6 BER versus molecules per bit
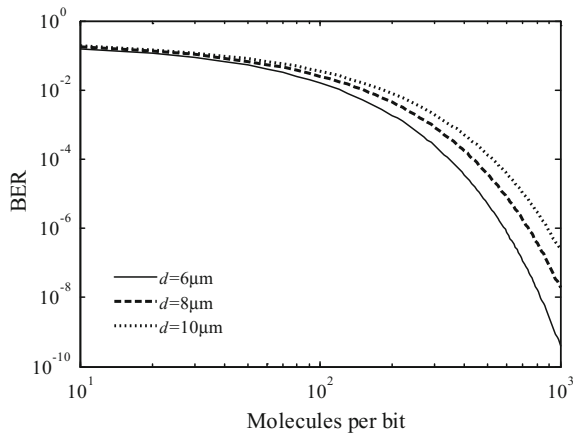
calculating the average BER of all the four states stated above. In this chapter, the symbol duration t_s is chosen as the time before 60% of the molecules arrive at the receiver. A comparison with the results from the literature based on Monte Carle simulation for several values of t_s is given in Table 1. It may be observed that the agreement is excellent and the gain in calculation time is considerable.

From Table 1, it is known that the value of t_s is set according to the transmission distance. Thus, for a fixed transmission distance d, the BER is a function of the threshold τ and the number of molecules per bit N, which is set in the range of 10–1000 here. Also, for each value of N, the threshold must be a specific value such that $\tau \in (1, N)$. Thus, the minimum bit error rate can be obtained for each N, which is the number of molecules per bit.

The optimized BER versus molecules per bit for different transmitted distance when the receiver radius is 5 μm is shown in Fig. 6, which shows that for a fixed distance d, the BER decreases with the increasing of number of molecules per bit N.

This demonstrates that as one might expect, the communication system has a better performance with a superior BER if large numbers of molecules are sent during one time slot. In addition, for a chosen value of N molecules sent in the time slot, the BER increases as the distance between the transmitter and receiver increases as a consequence of by the random property of the diffusion process. Thus, to achieve a better performance of the communication system, it is better to have a smaller propagation distance d and a larger number of transmitted molecules per bit.

4   Error Correction Coding

Coding is considered as a trade-off between reduced throughput and a reduction in the BER because of the redundant bits. Error correction coding covers different families of codes. Some of the traditional codes are shown in Fig. 7, which is far from being exhaustive. At the highest level, codes can be divided into two classes: those that can only detect errors and those can correct errors as well.

Digital communication systems transmit random quantities rather than reproducing the waveform of the transmitted signal [14]. A typical Shannon communication system is shown in Fig. 8.

For a molecular communication system, noise which results from ISI, may cause a distortion of transmitted information, making it necessary to use coding theory for controlling errors and achieving reliable data transmission. The BER can often be reduced and improved by choosing a slow and robust modulation scheme or coding scheme, and by using channel error correction coding schemes which can be employed to detect and possibly correct a certain number of errors that occur when information symbols are transmitted in a communication system. Here Hamming coding is considered since although these relatively simple block codes are not powerful by modern standards, they are still employed, and are very efficient in terms of the energy budget for nanoscale communications. Hamming codes have also typically proven to be a simple and efficient code over a channel where the errors are burst-free widely dispersed. Hamming codes are denoted in the form (n, k), where $n = 2^m - 1$, is the coded length output for a the number of parity check bits m, and $k = 2^m - m - 1$ is the number of date bits per block. The minimum distance d_m of this kind of block code is 3, which means that only one error can be corrected in each block. To compare the coded and uncoded transmission performance, the BER for the Hamming coded operation should be decided which is not an easy task for coding systems in general. However, for binary linear block codes in this case, an approximate BER expression is [3]:
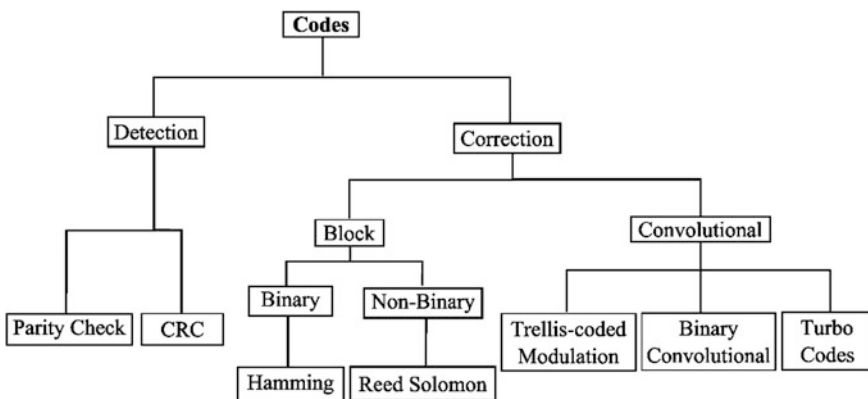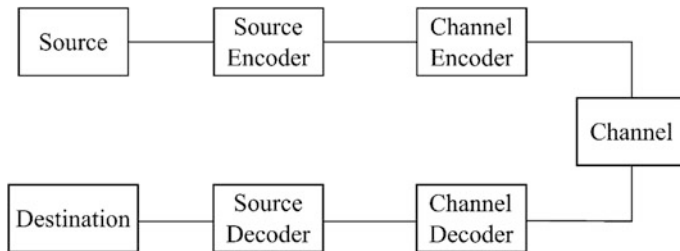

Fig. 7 Error detection and correction code classes

Fig. 8 Typical blocks of the Shannon communication system

$$P_{ec} = \frac{1}{n} \sum_{j=t+1}^{n} j \binom{n}{j} p^j (1-p)^{n-j} \tag{9}$$

where n is the length of code word, $t = (d_m - 1)/2$ is the maximum number of errors that the code can correct, and p is the probability of one bit error. The bit error probability p can be set in different ways [15]. In the case of molecular communications, there are four different probabilities of errors depending on the data pattern because of ISI, which has been stated in the previous section. With the calculation of the optimized BER in Sect. 3, the channel can be considered as a binary symmetric channel (BSC), which means that only bit flip is taken into consideration. Here, the bit flip probability is set as the values of the optimized BER.

Figures 9a through (c) show the BER results as a function of molecules per bit for uncoded, (7, 4) and (15, 11) Hamming-coded transmission by using (9) over distances of 1 μm, 4 μm and 8 μm respectively.

It is clearly seen that the system performance improves when the information sequence is Hamming coded. Generally speaking, to evaluate the performance of the system, the coding gain in dB is often assessed at a specific value of BER for a specific code. This coding gain can be directly obtained by finding out the ratio of molecules needed because there is approximately a linear relationship between the transmission energy and the number of molecules per bit. It can be derived that for a chosen BER level, here 10^{-9}, the coding gain for a (15, 11) Hamming code is 1.71 dB while for a (7, 4) Hamming code, it is 0.89 dB. In addition, an extra distance of 2 μm can be obtained by employing a (15, 11) code and 1 μm for a (7, 4) code. This indicates that the Hamming code is useful to extend the transmission range and achieve a coding gain, which can improve the system performance.

A coded system can give a better performance than an uncoded one but this improvement comes at the extra energy cost associated with the coding and decoding method. The energy saving (or loss) ΔE is defined as:

$$\Delta E = 2450(N_{uncoded} - N_{coded}) - E_{encode} - E_{decode} \tag{10}$$

In (10), $N_{uncoded}$ and N_{coded} are the number of molecules needed to achieve a given BER for un-coding and coding system, E_{encode} and E_{decode} are represent the

Fig. 9 BER versus molecules per bit for Hamming code with R = 5 µm and **a** $d = 6$ µm, **b** $d = 8$ µm, **c** $d = 10$ µm
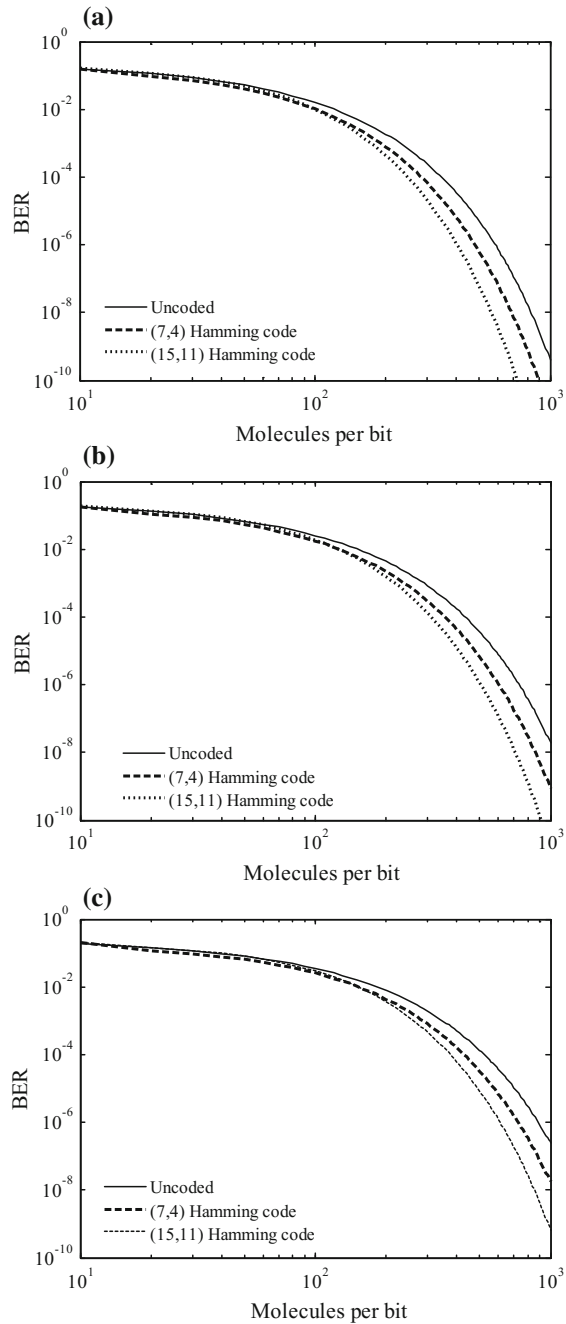
energy required to encode and decode, $2450\,K_BT$ is the coat of synthesizing a molecule, where K_B is the Boltzmanm constant. It is also assumed that the system is operating at an absolute temperate of $T = 300$ K. Here K_BT is used as an unit when measuring the energy. It is easy to see that when $\Delta E \geq 0$, it is worth to apply hamming code into the system.

Adenosine triphosphate (ATP) is a nucleoside triphosphate used in cells, it is often called the "molecular unit of currency" of energy transfer between cells in living organisms [16]. Encoding and decoding methods can be developed by using some logic circuits, an sound argument for every transmission through a NAND gate to cost 1 ATP is given by Sauro and Kholodenko [17, 18] and NANDs are universal in that one can construct all other gates from them via the principles of Boolean algebra. So here the NOT gates, AND gates, XOR gates and shift register units cost one, two, five and four ATPs respectively. This provides a basis for obtaining the total energy cost for the coding and decoding systems.

Figure 10a, b show the encoder and decoder for Hamming codes [19]:

The energy consumption for Hamming codes with different orders, m are given by:
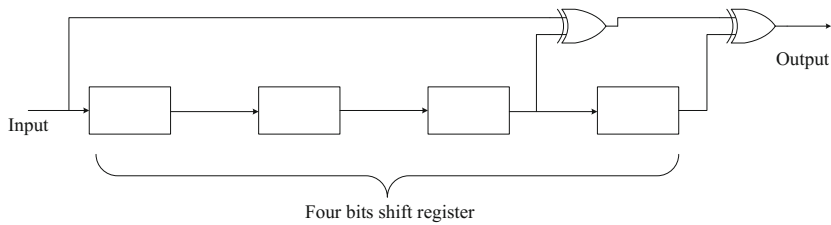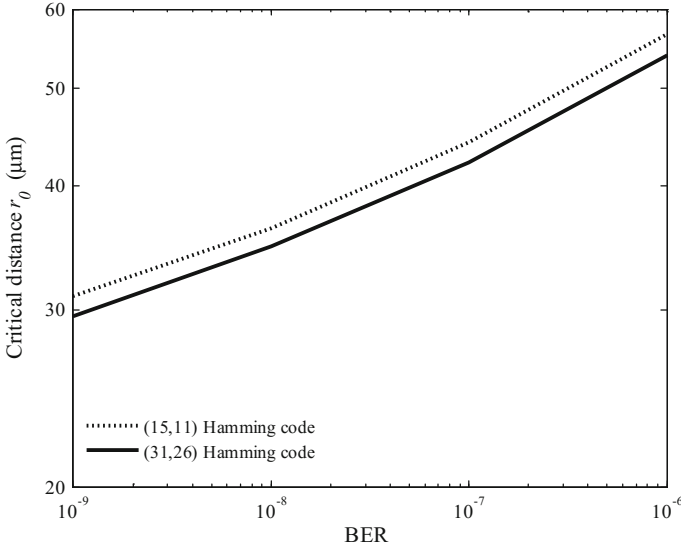

(a) Non-systematic encoder.
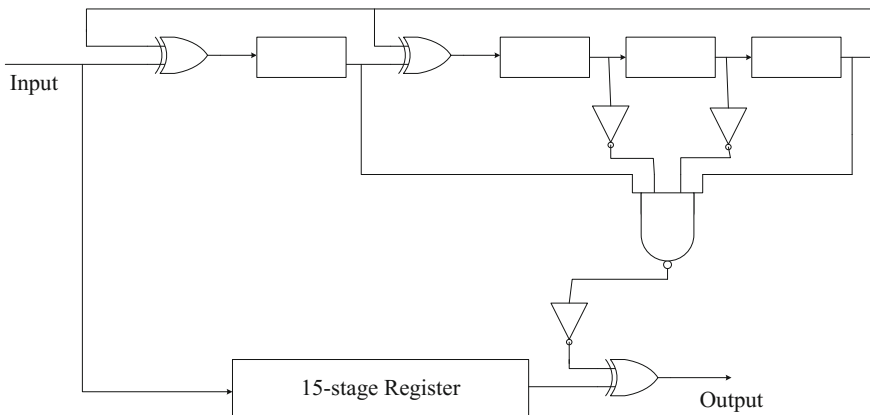
(b) Meggit decoder.

Fig. 10 Encoder and decoder for (15, 11) Hamming codes

Fig. 11 Critical distance r_0 with BER for m = 4 and m = 5 Hamming codes

$$E_{encode} = 20N_{tx}(4m + 10) + 2450N_{tx} \tag{11}$$

$$E_{decode} = 20N_{rx}(4m + 4n + 19) + 2450N_{rx} \tag{12}$$

N_{tx} and N_{rx} are the numbers of code generation molecules required to encode and decode the data. Assuming that $N_{tx} = N_{rx} = 300$ and one ATP reaction is about $20\,K_BT$ at 300 K. The critical distance [20] is defined as the distance at which $\Delta E = 0$ so that greater distances are more energy efficient with coding than without it.

Figure 11 shows the critical distances for m = 4 and m = 5 Hamming codes for BERs ranging from 10^{-9} to 10^{-6}. It shows that with the increases of the BER, the critical distance r_0 also increases, and the (31, 26) Hamming code needs a lower critical distance values to meet ΔE than (15, 11) Hamming code. The introduction of coding is beneficial for energy usage when $r_0 \geq 30$ μm.

5 SW-ARQ Schemes Within the Nano-Scale

Recently, the principle of *assured state transfer* was proposed [21], which is similar to the Stop-and-Wait Automatic Repeat reQuest (SW-ARQ) scheme in macro-communication applications [5, 6]. Assured state transfer is designed to enable the reliable transmission of signalling molecules from the source to the destination. The procedure of protocol abstraction is as follows, and shown in Fig. 1.

Firstly, the message molecules (represented as Mol_1) are emitted by the transmitting (Tx) nano-machine and propagate to the receiving (Rx) nano-machine. Upon reception, the Mol_1 molecules stimulate Rx to release another kind of molecules, known as an ACKnowledgement or ACK (represented as Mol_2). When Mol_2 is received by Tx, they stimulate the nano-machine to stop releasing Mol_1. In this way, this protocol abstraction guarantees that no more Mol_1 is transmitted to Rx, thus attempting to make each state more reliable. Further work on this protocol abstraction has been carried [22], where simulations on a molecular based reliable communication protocol with nano-logic computation were discussed. In this section, the abstraction is enhanced, proposing two schemes, and combine it with a realistic physical end-to-end model.

The system considered in this chapter, shown in Fig. 1, consists of two nanomachines, the transmitter, Tx, and the receiver, Rx, which communicate via transmission molecules. It is assumed that the nanomachines behave in a way that is analogous to biological entities and that they communicate via a natural ligand-receptor binding mechanism [23]. It is further assumed that, as per the work in [24], that each nano-machine can emit several specific types of molecules which can be recognised by the other nano-machine. For Tx to communicate a message, the total number of bits are broken up into $(n - i)$ bits per packet and then ordered into the packed structure shown in Fig. 12a with i index bits and that the receiver transmits an ACK with structure shown in Fig. 12b. For a message with an index, Ix, ranging from 0 to $2^i - 1$, two different kinds of molecules are required: one for the packet with the index Ix and one for the corresponding ACK with index Ix such that each molecule does not interfere with each other. This is otherwise known as pheromone diversity [13]. It is finally assumed the molecules are chemically inactive so they will not react with other molecules in the medium and that they are easily eliminated after decoding or after a fixed period of time. The medium is assumed to be a diffusing channel [18], where the molecules have a capture probability given by (2).

Two schemes are considered, one of which is very similar to macro-scale SW-ARQ and one that employs multiple acknowledgements.

(1) Scheme 1

As shown in Fig. 13a, the Tx transmits a packet and waits for the ACK from the Rx. If the waiting time is longer than a pre-defined limit, t_{out}, the Tx retransmits the packet. When receiving the packet, the Rx sends back an ACK and waits for the next packet from the Tx. If the waiting time is longer than the same pre-defined limit, the Rx retransmits the ACK.

Fig. 12 The structure of: **a** the transmission packet; **b** the ACK
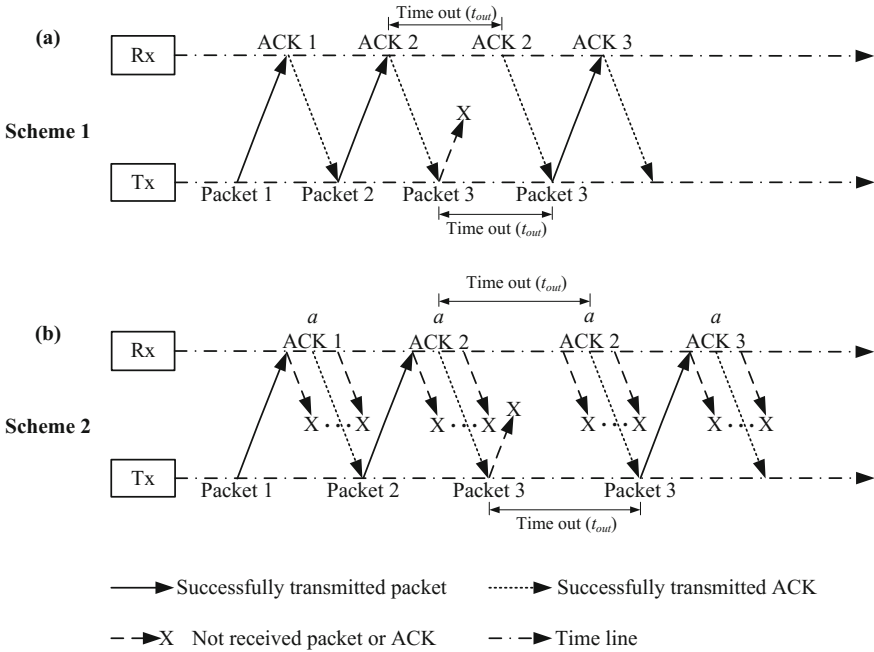

(a)

Index	Message
i bits	n-i bits

(b)

Index
i bits
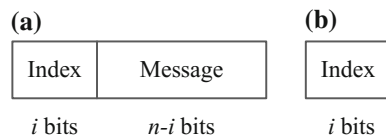
Fig. 13 The two proposed SW-ARQ Schemes

(2) Scheme 2

This scheme, as shown in Fig. 13b is based on the traditional SW-ARQ scheme. But when the Rx receives a packet, it sends back *a* copies of the corresponding ACK simultaneously. As soon as at least one of the *a* ACKs is received by the Tx, the next packet will be transmitted. Based on the structures shown in Fig. 12, a length ratio, *Ratio,* can be defined as:

$$Ratio = \frac{\text{The length of the packets}}{\text{The length of the ACKs}} = \frac{n}{i}, \tag{13}$$

which implies that when *Ratio* is large (that the length of the ACKs is significantly smaller than those of packets), when a packet does arrive at Rx, but the corresponding ACK is lost, this scheme will use less energy compared to the Scheme 1.

The first figure of merit in evaluating these transmission schemes is energy consumption required for a complete transmission. Setting $n = 17$ and $i = 2$, the length of each packet and ACK is $l_{pac} = 17$ and $l_{ack} = 2$ (in bits) respectively. Thus the total amount of energy required for a successful complete transmission can be calculated by counting the overall bits transmitted by both Tx and Rx as:

$$\text{Energy} = l_{pac} \times N_{pac} + l_{ack} \times N_{ack} = 17 \times N_{pac} + 2 \times N_{ack} \qquad (14)$$

where N_{pac} is the number of packets required and N_{ack} is the number of ACKs required assuming that each bit requires the same amount of energy. It can be noted that the unit of energy in this case is thus normalised to bits, thus allowing the reader to substitute their own energy model given by their own Tx-Rx nano-machine. The second figure of merit is the time for completion of a successful transmission under a given scheme and is calculated though simulation. It should also be noted, that due to this work being simulation based, each of the subsequent results is the average of 5000 retrials.

Considering the Scheme 2, varying a in the interval between 1 and 10, and setting $t_{out} = 15$ s, it can be seen from Fig. 14 that the scheme 2 greatly reduces the time cost compared to the scheme 1 ($a = 1$). It can also be seen that this scheme agrees with the known relationship of a 'longer distance requires a longer transmission time', but what is interesting is that for a given r, for an increasing a, the improvement differential (the gain between a and ($a - 1$)) decreases, implying an optimum value for a given application under this scheme. Furthermore, as can be seen in Fig. 14, using the Scheme 2, more energy is required to perform a successful transmission for any fixed a due to the capture probability decreasing with larger r values. However, what is not so intuitive is that, by increasing a, the probability of a successful transmission is increased, lowering the number of re-transmitted packets. Increasing a though does require more energy, so if a is too high, the cost of transmitting the excess ACKs (or the gain is having more ACKs) may outweigh the energy gain in requiring a lower number of transmission packets. This leads to an interesting trade-off where, as per Fig. 14, it can be seen that the energy cost is actually lowest when $a = 3$ (Fig. 15).

Referring to Figs. 16 and 17, t_{out} is varied to confirm that as it is reduced, the time to complete the transmission reduces (as the wait is on average shorter), and
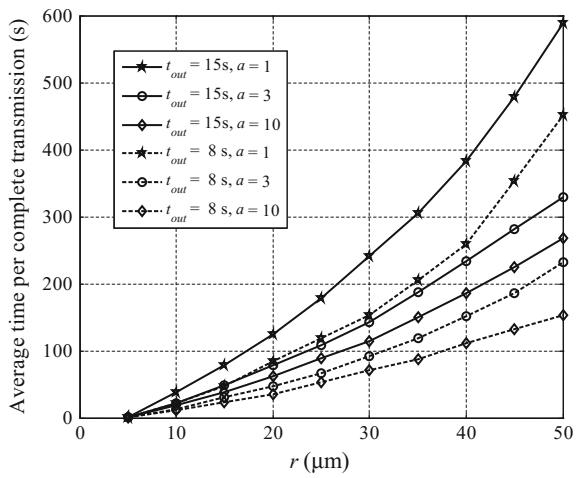

Fig. 14 The time required for a complete transmission using the Scheme 2 under varying r and with $t_{out} = 15$ s. *Note $a = 1$ is also Scheme 1*
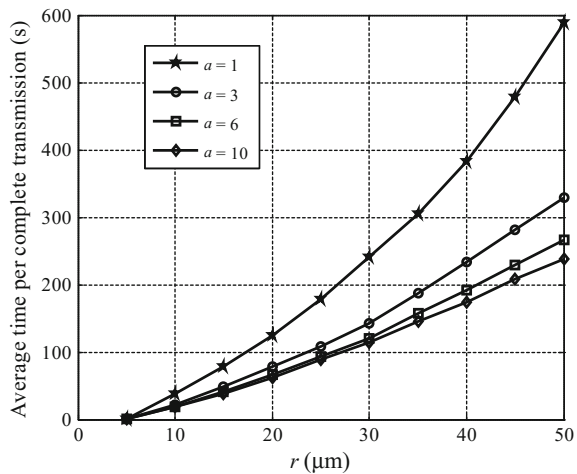
Fig. 15 The energy required for a complete transmission using Scheme 2 under varying *r* and with $t_{out} = 15$ s. *Note a = 1 is also the Scheme 1*
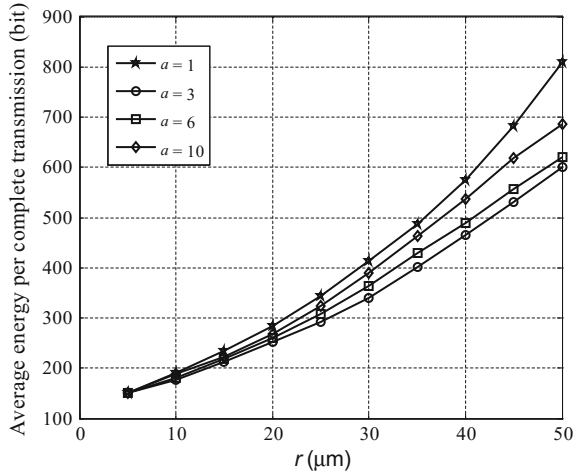
Fig. 16 The time required for a complete transmission using the Scheme 2 under varying *r* and *a* at $t_{out} = 8$ s and 15 s. *Note a = 1 is also the Scheme 1*

Fig. 17 The energy required for a complete transmission using the Scheme 2 under varying *r* and *a* at $t_{out} = 8$ s and 15 s. *Note a = 1 is also the Scheme 1*
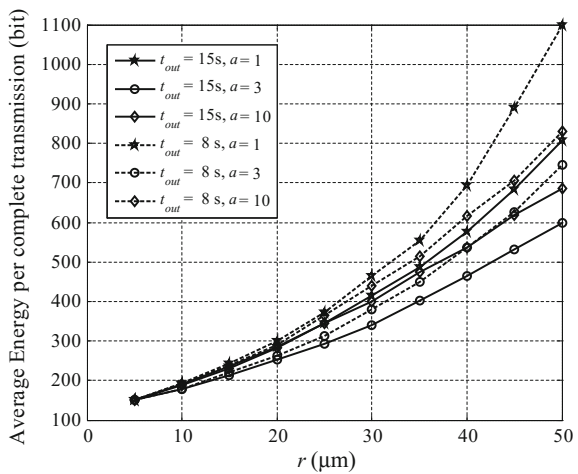
uses more energy (as more re-transmissions are made). However, this is again somewhat counter intuitive, it can also be seen that there may be a benefit in both time and energy to by carefully selecting t_{out} and a. This is an interesting result as for example, with $r = 50$ µm, and for $t_{out} = 8$ s and $a = 3$, the energy requirement is 747 bits and the time cost is 232 s. For $t_{out} = 15$ s and $a = 3$, the energy consumption is 600 bits, and the time cost is 329 s. That is, with the t_{out} increasing from 8 to 15 s, the energy consumption reduces 20% and the time cost increases 41%. Depending upon the scenario and system requirements, these parameters are available for the system designer to further refine.

Thus, it can stated, based upon the simulation results for the SW-ARQ schemes that Scheme 1 is the best choice for adjacent communications. Access to a lightly higher energy budget will mean that Scheme 2 will provide better performance than scheme 1 for longer communication distances.

6　Summary

This chapter has presented an overview of the essential elements needed to describe analysis and design molecular communication systems employing coding and protocols. After a brief introduction to codes and protocols, and the relevant state of the art, the key differences between macro-scale and nano-scale communications have been outlined. Following this, the diffusive medium and its impact on molecular communications were described leading to quantitative performance illustrations. The modulation techniques used in diffusion based molecular communication system were then introduced. One of them (CSK) formed the basis for more detailed analysis and performance predictions. Next, it was shown that Hamming codes are useful to extend the transmission range and achieve a coding gain, which can improve the system performance. In addition, the results also take energy budget into consideration, thus determining that the critical distance at which coding become beneficial in energy terms is approximately 30 µm. The final section presents the implementation of two SW-ARQ schemes for nano-communication networks, each of which has its own superiority and drawbacks. For short distance, near-adjacent communications ($r \leq 10$ µm) standard SW-ARQ performs well. However, use of multiple acknowledgments is more effective as the distance increases consuming only 50–80% of the energy needed for standard SW-ARQ by the time 50 µm is reached. System parameters need to be carefully selected to achieve the optimal performance and designers should choose the appropriate scheme dependent on the application. Overall, the fundamental aspects germane to the utilization of coding a protocol for molecular communications have been presented with novel insights into these areas. Thus, coding and protocols within molecular communications open up new opportunities and as nano-communications research develops, and it can be expected that new ways of coding and protocols within this exciting paradigm will be investigated.

References

1. Akyildiz IF, Brunetti F, Blázquez C (2008) Nanonetworks: a new communication paradigm. Comput Netw 52:2260–2279
2. Moore M, Enomoto A, Nakano T, Egashira R, Suda T, Kayasuga A et al (2006) A design of a molecular communication system for nanomachines using molecular motors. In: IEEE international conference on pervasive computing and communications workshops, pp 554–559
3. Bernard S (1988) Digital communications: fundamentals and applications. Prentice-Hall
4. Kurose JF (2005) Computer networking: a top-down approach featuring the Internet. Pearson
5. Moeneclaey M, Bruneel H (1984) Efficient ARQ scheme for high error rate channels. Electron Lett 20:986–987
6. De Munnynck M, Lootens A, Wittevrongel S, Bruneel H (2002) Transmitter buffer behaviour of stop-and-wait ARQ schemes with repeated transmissions. In: IEE proceedings in communications, pp 13–17
7. Kuran MŞ, Yilmaz HB, Tugcu T, Özerman B (2010) Energy model for communication via diffusion in nanonetworks. Nano Commun Netw 1:86–95
8. Leeson MS (2000) Performance analysis of direct detection spectrally sliced receivers using Fabry-Perot filters. J Lightwave Technol 18:13–25
9. Ziff RM, Majumdar SN, Comtet A (2009) Capture of particles undergoing discrete random walks. J Chem Phys 130. 27 Mar 2009
10. Kuran MS, Yilmaz HB, Tugcu T, Akyildiz IF (2011) Modulation techniques for communication via diffusion in nanonetworks. In: IEEE International Conference on Communications (ICC), pp 1–5
11. Kim N-R, Chae C-B (2012) Novel modulation techniques using isomers as messenger molecules for molecular communication via diffusion. In: IEEE International Conference on Communications (ICC), pp 6146–6150
12. Freitas RA (1999) Nanomedicine, volume I: basic capabilities. Landes Bioscience
13. Giné LP, Akyildiz IF (2009) Molecular communication options for long range nanonetworks. Comput Netw 53:2753–2766
14. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423
15. Yan H, Choe HS, Nam S, Hu Y, Das S, Klemic JF et al (2011) Programmable nanowire circuits for nanoprocessors. Nature 470:240–244
16. Knowles JR (1980) Enzyme-catalyzed phosphoryl transfer reactions. Annu Rev Biochem 49:877–919
17. Sauro HM, Kholodenko BN (2004) Quantitative analysis of signaling networks. Prog Biophys Mol Biol 86:5–43
18. Leeson MS, Higgins MD (2012) Forward error correction for molecular communications. Nano Commun Networks 3:161–167
19. Blahut RE (2003) Algebraic codes for data transmission. Cambridge University Press
20. Howard SL, Schlegel C, Iniewski K (2006) Error control coding in low-power wireless sensor networks: when is ECC energy-efficient? EURASIP J Wireless Commun Networking 2006
21. Akyildiz IF, Fekri F, Sivakumar R, Forest CR, Hammer BK (2012) Monaco: fundamentals of molecular nano-communication networks. IEEE Trans Wireless Commun 19:12–18
22. Walsh F, Balasubramaniam S, Botvich D, Donnelly W, Sergeyev S (2007) Development of molecular based communication protocols for nanomachines. Presented at the proceedings of the 2nd international conference on nano-networks, Catania, Italy
23. Krivan V, Lánský P, Rospars JP (2002) Coding of periodic pulse stimulation in chemoreceptors. BioSystems 67:121–128
24. Atakan B, Akan O (2007) An information theoretical approach for molecular communication. In: Bio-inspired models of network, information and computing systems, pp 33–40