DNA-Based Molecular Computing, Storage, and Communications

Qiang Liu, *Member, IEEE*, Kun Yang, *Senior Member, IEEE*, Jialin Xie, and Yue Sun, *Member, IEEE*

***Abstract*—DNAs exist in nature and could provide solutions to computing, storage, and communications as existing ones approach their physical limits. Plenty of research work has been conducted on DNA-based computing, data storage and molecular communications (MCs), but largely in isolation. There is a lack of a unified place where the triples are put together to be discussed. This article aims to filling in this gap by providing an overview of each triplet. Starting from an overall description of DNA features and their reading and writing in practical terms, this article goes on to describe each of the three from three aspects: 1) requirements and differences from current electronic-dominating technologies; 2) their working principles; and 3) practical considerations. Recent advancement in each area is summarized and discussed. Furthermore, this article intends to call for researches that go beyond the boundary of each and encourages interconnection and joint research among the three. It proposes a molecular information and communication technology (ICT) system architecture with all its three components underpinned by DNAs. This article also identifies and discusses some new future directions, such as joint coding for storage and communications, directional DNA-based MCs, interfaces between molecular DNA systems and electronic systems. It is hoped that this article can spark more joint research across computing, storage, and communications in this exciting field of DNA-based molecular ICT systems.**

***Index Terms*—DNA, DNA-based communications, DNA-based storage, DNA computing, molecular communication (MC).**

I. INTRODUCTION

CURRENT information and communication technology (ICT) systems face serious challenges. For instance, Intel processors have gone to the scale of several nanometres making it almost impossible to be made denser and thus faster in processing due to physical limitations. The data generated by us are growing exponentially to a point where current storage technologies cannot cope in a sustainable way. 5G mobile systems are rolling out commercially and discussions on 6G have also started. However, their performance gain is at a high cost of energy consumption. A revolutionary thinking that goes beyond the current electronic domain and provides a much high information density and greener way for communications is imperative.While the Moore's Law that dominates the electronic domain is slowing down, our ability to manipulate DNA's has been growing in an unprecedented pace. The rapid advancement of biotechnologies nowadays has enabled us to manipulate DNA in a more cost-effective manner and at large scale. And this has come to a point where massive production of DNA products becomes commercially feasible. In the meanwhile, DNAs, like other biological components, exist in nature and are extremely energy efficient. DNAs also exhibit other advantageous features, such as longevity and enormous information density. These all make DNA an attractive medium for computing, data storage and communications.

This has given rise to researches into DNA-based molecular ICT. Here, this collective term ICT specifically refers to computing, storage, and communication these three most fundamental aspects of any ICT system. Current ICT systems are predominantly based on electronics, which have been pushed to their physical limits. This imperatively calls for alternative solutions to problems caused by ever increasing demands for processing, data storage and communication bandwidth. We believe going to the molecular level and in particular DNA is inevitable, as also observed by Carmean *et al.* [1]. DNA-based molecular ICT started with computing, dating 25 years back to Adleman's research on molecular computation of solutions to combinatorial problems that was published in the Science Journal [4]. Together with photonic computing and quantum computing, DNA computing is regarded as one of the three most-promising future computing modalities. DNA's twin-strand structure has been found particularly suitable for image matching [2], at a speed hundreds of times faster than nowadays' approach. Recently, researchers have developed techniques to enable DNA strand displacement reactions (SDRs), which can be utilized to fulfil general purpose computations, such as Boolean operations [5]. More complex circuits (and thus operations) can be created by using various enzymes for DNA processing, including cutting, joining, replication, and editing [1].

Manuscript received April 8, 2021; accepted May 11, 2021. Date of publication May 25, 2021; date of current version January 7, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62071101; in part by the Fundamental Research Funds for the Central Universities under Grant ZYGX2019J001; and in part by the Special Science Foundation of Quzhou. *(Corresponding author: Kun Yang.)*

Qiang Liu is with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China, and also with the Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, Quzhou 324000, China (e-mail: liuqiang@uestc.edu.cn).

Kun Yang is with the School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, U.K., and also with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: kunyang@essex.ac.uk).

Jialin Xie is with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: 201921010530@std.uestc.edu.cn).

Yue Sun is with the School of Information Science and Technology, Chengdu University of Technology, Chengdu 610059, China (e-mail: sunyuestc90@126.com).

Digital Object Identifier 10.1109/JIOT.2021.3083663

DNA-based data storage became materialized around the same period of time as DNA computing, though the concept of using DNA as storage could go back to around the mid-1960s, as represented by N. Wiener's suggestion of "genetic" memory for computers. A representative work was Baum's research published in the Science Journal [9]. From then on, DNA's potential as an alternative for data storage has been intensively researched into across the world. Significant advancement has been achieved in recent years. In a recent work, Carmean *et al.* [1] have implemented a real-life system that can archive digital information in the form of DNAs and then retrieve it from the DNAs. An image search example is presented in this work illustrating the real usefulness of the system. Meiser *et al.* [10] provided technical approaches of reading and writing digital data in DNA with particular focus on error correction.

Communications are needed between DNA-based storage and processors and also from one storage to another. And this has naturally given rise to DNA-based molecular communications (DNA-MCs) in recent five years. DNA-based communication is a type of bioinspired molecular communication (MC) method that uses DNA to encode, transmit and receive information. For instance, Bilgin *et al.* [2] proposed a novel DNA-based communication protocol in order to achieve high capacity communication between nano-machines. Shah *et al.* [11] have incorporated DNA-based data storage into an MC system via encoding mechanisms. A thorough review of the practicality of MC physical systems, with DNA included, is presented in [12], indicating increased degree of maturity of MC networks and indeed DNA-based communication systems.

Plenty of research work has been conducted on DNA-based computing, data storage and MCs, but largely in isolation. As a matter of fact they are very much related. For instance, coding is a common function for both storage and communication aiming to reduce errors. A coding mechanism that considers features from both storage and communication will be naturally and inherently beneficial to a DNA-based ICT system that includes both storage and communications. More discussions on this will be presented later in this article.

As is known to us all, the wireless sensor networks remain significant part of the IoT. Raut and Sarwade [6] proposed a nano-machines network which is based on MC, and that special structure has the obvious advantages of ubiquitous, energy saving, biocompatibility, and can operate in water medium, which makes up for the shortcomings of the traditional IoT fields. Moreover, the idea of seamless connection between DNA-based molecular nano networks and macro networks is to make the Internet of bionano things possible. Biological nano materials are expected to be applied in cutting-edge fields, especially in the medical domain. Murat's group designed a molecular antenna for biosensors based on field effect transistors, which can transmit signals and convert molecular information into electrical signals [7], and a micro/nanoscale biocyber gateway is needed in this structure. The gateway can decode molecular information and encode it into the concentration or the type information of molecules, and send the decoded information to the macro network node through the wireless link. The relationship between DNA-based ICT and IoT is clear and mutually reinforcing. During this process, DNA plays a key role because of its excellent ability of information storage and concealment. Bansal *et al.* [8] focused on the elliptic curve cryptography (ECC) calculation and Rivest–Shamir–Adleman (RSA) calculation which operated in the DNA computing related fields, and this investigation can sharply decrease the danger of leaking the delicate information. There are many related studies which reveal the strong correlation between DNA-based ICT and IoT fields.

In summary, the aim of this article are twofold. First, it serves to provide an overview of DNA-based molecular ICT covering computing, storage, and communications all in one place with cross references and discussions. Second, this article intends to call for researches that go beyond the boundary of each and encourages interconnection and joint research among the three. For this purpose, it proposes a molecular ICT system architecture with all its three components underpinned by DNAs. This article also identifies some new future directions, such as joint coding for storage and communications, directional DNA-based communications, interfaces between molecular DNA systems and electronic systems. It is hoped that this article can spark more joint research across computing, storage, and communications in this emerging, exciting and promising field of DNA-based molecular ICT systems.

The remainder of this article is organized as follows. Section II gives a brief overview of DNA structure and biological means of reading and writing them. DNA-based computing, storage, and communications are discussed in Sections III–V, respectively. Each of them is discussed from three aspects: 1) requirements and differences from current electronic-dominating technologies; 2) their working principles; and 3) practical considerations. Recent advancement in each area is also summarized. Before conclusions in Section VII, some future directions are identified and discussed in Section VI.

II. DNA PRINCIPLES

Nowadays, permanent data storage is very desirable to archive key records of our civilization and keep vital information of our time. Although there are various data storage technologies at present, the existing technologies may become obsolete due to the limited hardware life and high social storage cost for archiving data. Many scientists believe that DNA is a very promising medium of data storage, for its incredible longevity and extremely large information density. However, its research direction is not a thought in recent years, which is already proposed in the latter half of last century [3], while human exploration of it can be traced back to the end of last century [3], [4]. In this section, DNA structure is first introduced, while its composition, structure and characteristics elaborated. And DNA engineering from traditional to potential DNA reading and writing techniques is expounded.

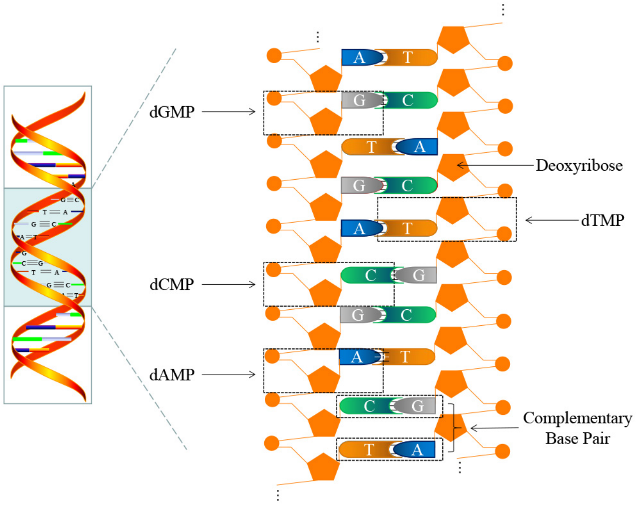


Fig. 1. Three-dimensional and plane structures of completely complementary DNA. (dAMP: DeoxyAdenosine MonoPhosphate, dCMP: DeoxyCytidine MonoPhosphate, dGMP: DeoxyGuanosine MonoPhosphate, dTMP: DeoxyThymidine MonoPhosphate).

A. DNA Structure

1) Composition of DNA: DNA, the full name of DeoxyriboNucleic Acid, is the main genetic material of organisms, whose basic molecular unit is deoxynucleotide. Each deoxynucleotide consists of three relatively simple compounds, namely phosphoric acid (P), deoxyribose and nitrogenous base. In fact, each nucleotide contains one of the following four bases: 1) adenine (A); 2) cytosine (C); 3) guanine (G); or 4) thymine (T), thus forming four different types of deoxynucleotides. Since a single DNA molecule is double chain composed of four kinds of deoxynucleotides, it can be simply represented by bases' initial string, e.g., AGTCGA.

2) Structure of DNA: Generally speaking, DNA molecule is a famous double helix spatial structure, as exemplified in Fig. 1.

1) DNA is made up of two long reverse parallel deoxynucleotides chains.
2) Deoxyribose and phosphoric acid in DNA molecules are alternately connected and arranged on the outside to form the basic skeleton. Specifically, the phosphoric acid of a single nucleotide (located at the carbon atom of deoxyribose 5') and the hydroxyl of another nucleotide (located at the carbon atom of deoxyribose 3') form a phosphate diester bond. Therefore, there will be a phosphoric acid or a hydroxyl group at both ends of the nucleotide chain. The end that connects phosphoric acid is named the five ("five prime") end, while the end with hydroxyl is named the three ("three prime") end.
3) The bases on the chain are linked to their complement by hydrogen bond, which conforms to the complementary pairing principle of bases: A and T, G and C. Therefore, as a DNA molecule with double helix structure, its sequence is completely complementary to its anticomplement.

However, sometimes the two sequences are not completely matched, which is called partially complementary structure. Since the structure determines the characteristics, whether bases are completely matched or not will directly affect some characteristics of DNA.

3) Characteristics of DNA: Diversity: The variety, quantity and sequence of base pairs in long chains are all changeable, which determines the diversity of DNA molecules. For example, for a short DNA molecule with 4000 base pairs, the possible sequencing methods can be up to 4^{4000}.

Stability: First, the alternating arrangement of phosphoric acid and deoxyribose sequence in DNA molecules is stable. Second, the hydrogen bond of base pair in DNA molecule keeps the double helix structure stable, which is because of the ineluctable energy consumption of hydrogen bond breaking. In addition, between G and C there are three hydrogen bonds while only two exist between A and T, which means that the more content of G and C, the more stable the DNA structure. However, the thermodynamic stability of some partially hybridized[1] structures is poor, which is difficult to exist at higher solution temperature. That is to say, the higher melting temperature[2] indicates that DNA molecules are more stable. In addition, the melting temperature is not always related to the amount of unpaired bases [9]. When the types of unpaired bases change, the melting temperature may change even if the amount of unpaired bases remains the same [13]. In practical engineering, the melting temperature can be accurately calculated by nupack and other thermodynamic simulation software.

Specificity: Every DNA molecule contains a specific base sequence, and each specific DNA molecule stores specific genetic information. This property makes the specificity of each DNA molecule. The structure of DNA molecule makes it unique and becomes the first condition in DNA-based molecular computing, storage, and communication.

B. DNA Engineering for DNA Writing and Reading

Since the information of DNA is contained in different arrangements, to master the writing and reading of DNA is the basis of DNA-based data computing, storage, and communication.

1) Writing—Synthesis of DNA: Gene synthesis is the technique of synthesizing double stranded DNA molecules *in vitro*, which is one of the means of gene acquisition. Compared with obtaining genes from organisms, gene synthesis does not need template since it is not limited by gene source [45]. This technology has been widely utilized due to its short synthesis cycle and no restriction of gene source.

At present, the most commonly adopted method is phosphoramite chemistry method, which is proposed by Caruthers [14]. It has the characteristics of high efficiency, high speed and relatively stable initial reactants. Specifically, the method is to fix the initial nucleotide by covalent binding with its 3' hydroxyl group on solid-phase carrier, such as controlled pore glass

---

[1] A biphasic process is defined as hybridization.

[2] The melting temperature is the melting point of DNA molecules. Specifically, when the temperature rises to a certain value, the double strand of DNA molecule sets out to rapidly disintegrate into random coil. Since the temperature range of denaturation is very narrow, the temperature required for the double stranded DNA molecule to untangle 50% is called the melting point of the DNA, which is abbreviated to T_m.

(CPG) to complete the synthesis reaction. And dimethoxytrityl (DMT) is connected to the 5' of the nucleotide. The specific steps are as follows.

1) *Deblocking:* DMT was removed from nucleotide linked to CPG by trichloroacetic acid to obtain free 5' hydroxy group.
2) *Activation:* Obtain the activated intermediate at the 3' end by mixing with activator and the 5' hydroxy group was still protected by DMT.
3) *Coupling:* Condensation reaction occurs between activated intermediate and free 5' hydroxy in solution.
4) *Capping:* Stop the continuous reaction of 5' hydroxy which was not involved in the condensation reaction.
5) *Oxidation:* Under the action of oxidant iodine, the unstable phosphite bond is transformed into more stable triphosphate.
6) *Loop:* Repeat the above steps until the nucleotide chain reaches the required length.

After synthesis and a series of treatment of cutting, deprotection, purification, quantification and storage, the stable storage time can be more than one year.

The above traditional method of DNA synthesis is limited to the direct production of oligonucleotide chains with about 200 bases. With the increase of the length, there will be inevitable errors in the synthesis process. Moreover, this process adopts toxic organic reagents, expensive and slow.

In recent years, a new technique for DNA synthesis has been proposed, due to Palluk *et al.* [15]. This method is totally different from the above-mentioned traditional technique. It depends on a DNA synthetase in cells from the immune system, which is naturally capable of adding nucleotides to DNA molecules in water. Owing to the stability of DNA in water, it is undoubtedly a landmark breakthrough. Compared with the former, the new technique is expected to improve the accuracy, so that the length of the synthesized DNA chain is able to be increased to ten times that of the traditional method, which contains thousands of bases.

2) Reading—Sequencing of DNA: DNA sequencing is the basis of further research and transformation of target genes in molecular biology. Generally, the traditional sequencing techniques are Sanger sequencing and chemical degradation of Maxam and Gilbert. These two methods differ greatly in principle, while they all initiate from a fixed nucleotide and terminate randomly at a specific base. The progress ultimately generates a series of nucleotides of different lengths, and then carry out electrophoresis detection to obtain DNA sequence.

At present, the DNA sequencing platform based on image processing and sequencing by synthesis is most commercial, following [13]. The steps are as follows.

1) *Base Pairing:* One by one, the complementary bases with fluorescent markers are connected to the single strand DNA sequence connected to the substrate.
2) *Image Generation:* Capture the fluorescence image and correlate the fluorescence points with the bases on the DNA chain one by one.
3) *Remove Markers:* Adopt chemical methods remove fluorescent markers and establish the next base for identification.

Besides, there are several technologies for DNA sequencing. The silver sequencetm DNA sequencing system of Promega company has faster speed, lower cost, and higher accuracy than traditional method, since it does not need expensive reagents and careful operation of isotopes in the radioactive method. It is a nonradioactive sequence analysis system that detects strips in gel by sensitive silver staining. The Taq DNA polymerase utilized in this system has a greatly strong thermal stability at 95s, which has a good effect on the double stranded DNA template. Besides, DNA sequencing via Illumina dye sequencing is adopted for data storage in DNA following [15].

In recent years, nanopore sequencing, which has a considerable development prospect, utilizes electrophoretic technique to drive individual molecules through nanopores one by one. Owing to the tiny diameter of nanopore that only a single nucleic acid polymer is allowed to pass through, a series of methods may be adopted for high-throughput detection on this basis. As a novel platform, nanopore detection technique has the superiority of low cost, high throughput and non labeling. Specifically, for single-stranded DNA (ssDNA) molecules with 1000 bases, nanopore sequencing is able to be adopted for detection without amplification or labeling, which makes it possible to sequence DNA inexpensively and rapidly. In addition, the nanoscale pore size ensures a stable continuity of detection, so that it has the high accuracy of sequencing. At present, Oxford Nanopore technology has been commercialized and nanopore DNA analyzer has been successfully developed.

III. DNA-BASED MOLECULAR COMPUTING

A. Classification of DNA-Based Molecular Computing

Process in classic semiconductor-based computing is facing challenges, while Moore's law in the electronic domain is slowing down. While the past two decades, the rapid development in bionanotechnology, including areas of materials, fabrication and self-assembly, has resulted in a tremendous improvement in the ability to manipulate molecules. This trend provides the possibility of computing directly with molecules. DNA-based molecular computing, as the major type, employs DNA SDRs to implement complex computing tasks, whereas classic computers implement silicon-based circuits fabricated on chips. The SDRs is an exchange reaction between double-stranded DNA (dsDNA) and its complementary ssDNA. Due to the Watson–Crick complementary pairing structure, DNA, with precisely predicted behavior, has proven to be an excellent candidate for logic operating systems design [16]. Since extraordinary information density on DNA strand, DNA computing provides vast parallelism computing ability. Thus, DNA-based molecular computing is regarded as promising next-generation computing modalities.

In 1994, Adleman proposed the first DNA-based molecular computer, which is designed to solve the seven-city Hamiltonian path problem [4]. It breaks the shackles of the classic computer architecture and found the first milestone to design DNA-based computing. Its basic idea is to encode the DNA strand sequence as an information carrier and perform SDRs under the control of catalyzing enzymes in a test tube as an implementation process booster. Inspired by Adleman, several DNA-based molecular computing systems have been designed and developed. These systems are classified by four types; followed by enzyme-free SDRs computing [17]–[19], [26], enzyme-catalyzed SDRs [20], [27], DNAzyme facilitated reactions [21], [22], and algorithmic self-assembly of DNA structure [24].

Among the four types of approaches, the most simple are enzyme-free SDRs computation systems [16]. The input and output signals are in the same form, and all represented by ssDNA. Thus, with the idea of mixing the input ssDNA and a set of programmed dsDNA for displacement, a series of DNA SDRs are directed by the toehold and a detectable ssDNA output was produced [25].

The majority of DNA computing designs are enzyme-catalyzed SDRs, which are aiming to boost the displacement reactions. Although the high information capacity of DNA sequence within the SDRs operation provides massively parallel processing, SDRs are extremely slow compared with classic semiconductor-based computing. For example, in 2011, it required several hours to perform one logic gate operation using seesaw gate strategy SDRs [26]. Therefore, before DNA computing is useful to large-scale implementation, SDRs must be accelerated. For example, in [27], with the cationic copolymer, SDRs operation times are markedly reduced from hours to minutes. Meanwhile, it shows the potential to further boosting the speed of DNA logic gate operations with the copolymer.

The third type of DNA computing employs DNAzyme facilitated reactions to perform a set of Boolean logic functions. Since any more complex arbitrary Boolean operation, such as XOR, NAND, ANDNOT can be expressed as a combination of basic operations, such as YES, NOT, AND, and OR functions. DNAzymes, known as a library of catalytic nucleic acids, are the basic elements combined to perform inhibit functions. According to [22], these inhibitory agents are widely used in various experimental environments of disease therapies, such as cancer, HIV and other viral infections. From a broader perspective, DNA computing based on Boolean operations has great potential and can be used to describe programmable functions in biomedical devices, which may be susceptible to cancer markers and external stimuli (such as light or magnetic fields) [23]. Thus, it will be necessary to develop a library of DNAzyme-based computing elements to implement more complex operation function. In [21], it reports the construction of a DNA-based computing system to perform basic logic gates that use a library of DNAzymes, and their substrates. To solve elimination of leakage, it usually designed a caged substrate to protect the output DNA strand until the DNAzyme is released [21]. Later, in 2013, [19] presents a new programmable DNAzyme library in a cell environment, which is composed of a variety of Boolean logic gates, and even a 3-input-AND gate. It also demonstrates the ability to operate logic gates in living mammalian cells [22].

The last types of DNA computing employ self-assembly structure of DNA. In general, the same starting material may form an infinite number of different structures. While the information in seed molecule can determine the direction and when and where to grow [24]. As demonstrated by the programable self-assembly of DNA tiles, the phenomena have the potential to be implement in growth of more complex molecular structures. In [24], it introduces a programmable DNA origami seed which can display 32 different binding sites. This origami self-assembly reactions involves up to 300 DNA strands containing more than 17 kb of sequence information [24]. However, the major challenge is the lack of the reliable and effective seeds in the algorithmic crystals. Thus, it demonstrates the potential route toward DNA-based self-assembly fabrication that the easy, reliable, and high-productive growth of algorithmic crystals enables with DNA seeds.

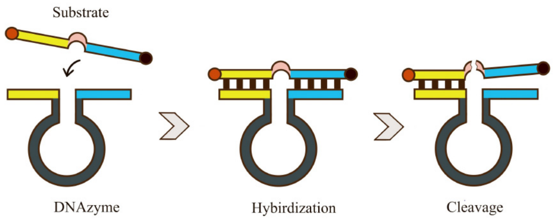


Fig. 2. Design of 10-23 DNAzyme Dz13.

B. Logic Gates Design of DNA Computing

The biological processes regarding medical disorders are extremely sophisticated, which requires complex nonlinear functions for precise mathematical description [23]. However, in practice we have to simplify the performance to the binary 1/0 behavior, due to significant differences with different patients. For instance, It is nature to employ the threshold to distinguish the two ranges cancer marker: 1) the safe level (0) and 2) the dangerous level (1), rather than the precisely concentration of cancer markers. Boolean logic gates operate the basic functions YES, NOT, AND, and OR functions to perform more complex computing tasks. Thus, logic gates design plays a dominating roll in DNA-based Molecular computing, which rely on DNA SDRs. Different SDRs provides a variety of design for a library of logic gates. In addition, a set of more complex intelligent toolboxes are developed based on logic gates design of DNA computing.

Here, we introduce the results of a set of logic gates design by DNAzyme library in [22]. The reaction operations are only performed by three basic mechanisms, as illustrated in Fig. 2.

1) The DNAzyme is split in the core catalytic region. So that only with the presence of appropriate input molecule, the subcomponents forms a complete DNAzyme complex.
2) Use a stem-loop structure to fix the DNAzyme arm. So that when the appropriate input exits stem-loop is uncaged.
3) Toe exchange reaction requires a long hybridization time, as it will change the conformation of the components when input molecule presents.

Fig. 3 illustrates the process to form the YES gate. When there is only one input molecule, a simple "YES" gate is activated. According to the previous design [21], DNAzyme is divided into two parts in the catalytic core region. Only when

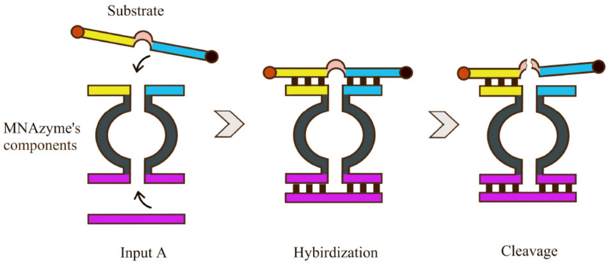


Fig. 3. YES gates demonstration.

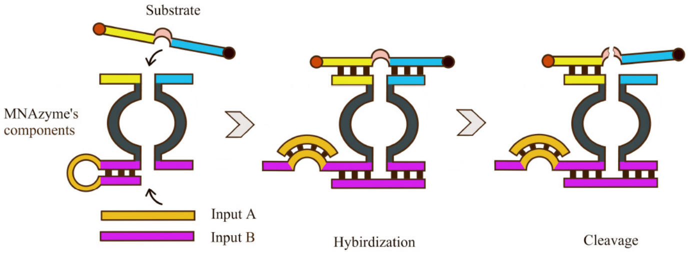


Fig. 4. AND gates demonstration.

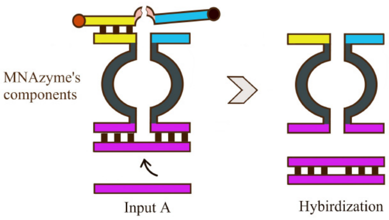


Fig. 5. NOT gates demonstration.

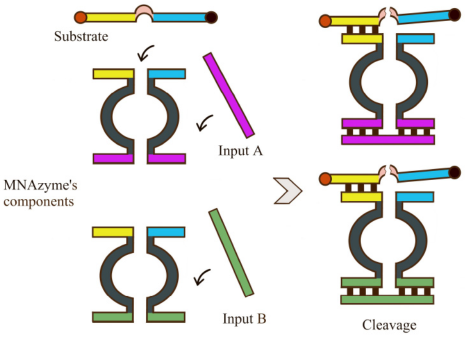


Fig. 6. OR gates demonstration.

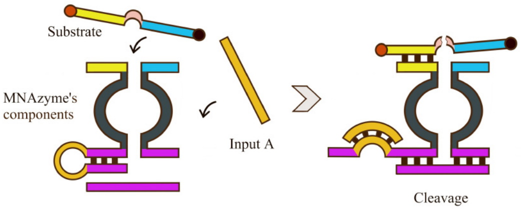


Fig. 7. ANDNOT gates demonstration.

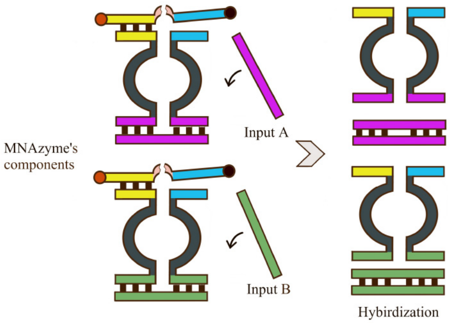


Fig. 8. NAND gates demonstration.

the appropriate input molecule is present, these two parts are combined to form a complete DNAzyme complex, and RNA cleavage occurs, which is resulting in a “True” output.

Fig. 4 presents the demonstration of AND gates. To form an AND gate, it consists of an additional binding loop and a cage sequence in a stem-loop structure. The binding loop is the complementary sequence of input B (the purple strand in the Fig. 4) and followed by a cage sequence (the yellow part) which is complementary to the DNAzyme. When the input B occurs, the arm is in the state of uncaged and accept the first input A which connects the two subcomponents, since it is conducive to an open construction. Only when the presence of both input A and B, a complete DNAzyme complex is formed to operate AND gates.

As the basic logic gates in this library, the demonstration of NOT gate is shown in Fig. 5. In order to realize the NOT gate, it implements MNAzyme’s components to add an additional strand (the second purple one), which contains the input complementary sequence. When the input is present, it bonds to the complementary sequence, as the reverse of input, the MNAzyme’s components are separated. And this process is resulting in a “False” output.

AS shown in Fig. 6, To form a OR gate, two YES gates are implemented in parallel based on two different inputs.

With the design of basic Boolean gates, including YES, AND, NOT and OR gates, more complex gates, which combines two or more basic gates, as the extension of logic gates library. As shown in Fig. 7, the ANDNOT gate combines the AND and NOT gates. When input A and B are present together, the input B bonds to the complementary sequence, resulting in the presence of input A and the reverse of input B, which means a “False” output.

As shown in Fig. 8, the NAND gate is combines two NOT gates with two different inputs. As a consequence, the two gates are inactive only when input A and B are present together.

As the NOR gate is the inverse of AND gate, we can find the difference between Figs. 4 and 9. When either one input is present, resulting in disassembling the complete DNAzyme complex, which means a “False” output.

As shown in Fig. 10, the XOR gate combines two ANDNOT gates. When there is any input, it will activate the subgates, but also break the orthogonal subgates by combining with its complement sequence of inputs.

Finally, as shown in Fig. 11, it extends its library to support 3 inputs. The 3-input-AND gate is operating AND gates with the MNAzymes, in which both arms are caged. Only when all 3 inputs are present, resulting in forming a complete

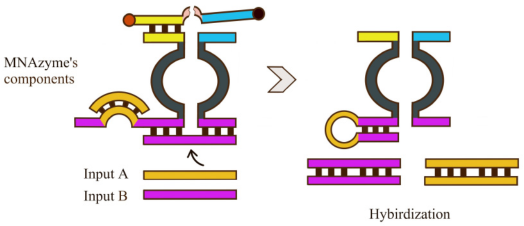


Fig. 9. NOR gates demonstration.

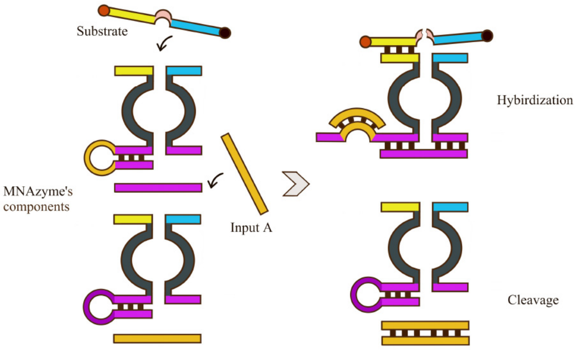


Fig. 10. XOR gates demonstration.

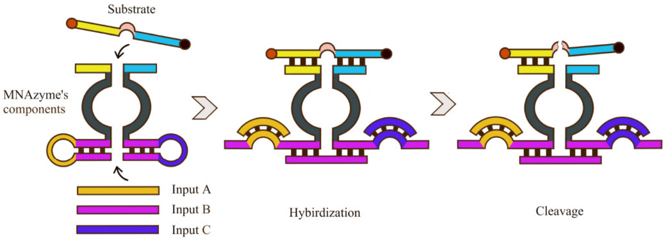


Fig. 11. 3-input-AND gate demonstration.

DNAzyme complex by breaking the cages on the left and right side, which means a "True" output.

Besides this design of logic gates library, a new three-input logic gate and multiple-input logic circuit were introduced in [19]. A few more complicated computing circuit including binary square root and neural memory mimicry have been demonstrated available using SDRs. The most encouraging breakthrough in [28], 7 years later in 2020, an SDR-based DNA digital switching circuits (DSCs) are developed for implementing digital computing. This work demonstrates that the arbitrary Boolean functions can be represented by DSCs and shows the potential in designing molecular digital switches with high speed. This contribution minimizes the gap between high-bandwidth digital communication and biomolecules devices. With the encouraging development of DNA-based Boolean logic gates design over the last decades, our expectation of next-generation DNA computing is gradually clear.

C. Practical Considerations and Application

The major advantages of DNA computing consists of four aspect.

1) Massive parallelism of DNA strand.
2) Watson–Crick Complementary pairing.
3) Exceptional energy efficiency.
4) Environment friendly in biologic system.

However, due to the SDRs of DNA computing, the disadvantages is obvious compared with classic computing.

1) Reaction time is relatively long, ranging from minutes to hours.
2) Relatively expensive.

Thus, it is not suitable for frequent serial computing. In the parlance of today's IT community, cold data is data that is not frequently accessed or actively used. It is perfect to get collected and stored for a long time in DNA storage without being retrieved, analyzed or transferred to another part of the system in low frequency. In [29], it cryptanalyzes the image encryption algorithm based on DNA computing by combining the chaotic map and recognizes its weaknesses. This proposed scheme has proven a enhanced strength in modifying color maps. In the past few years, researchers have proposed many encryption schemes based on obfuscation and diffusion [30]. In the diffusion part, the scheme uses DNA computation as it is reversible. These works demonstrate the potential of DNA computing in cryptography and imaging processes.

Another potential application scenario is the advanced theranostic intergraded with DNA or molecular computing [23]. In the last decades, a lot of research work, regarding biomedicine, has focused on the design of the properties and behavior of nanomaterials can be programable. The development of DNA computing in drug delivery systems is represented by the concept of nanomedicine that can achieve logical gated targeting, sensing, imaging and therapy [23]. In [23], it summaries three types of the potential applications based on logic-gates, including selective targeting, nanoparticles design for sensing, imaging and detection tasks and last type design for therapeutic tasks.

In 2016, [32] has demonstrated the ability to perform complex synthetic genetic programs task on continuous environmental signals in living cells, which involve analog and digital information processing. Inspired by this work, hybrid biosemiconductor systems has shown the great potential in key applications with significant scientific and economic impact [31]. DNA computing with its unique features will play an important role in the development of hybrid biosemiconductor systems design.

IV. DNA-BASED MOLECULAR STORAGE

The past decade has witnessed the rapid development of storage technologies for data information, thanks to the gradually digitalization of the whole world. The storage media has been updated for several generations, dating from the original punched tape, magnetic tape, to the modern hard disk and flash memory. An ideal storage system should have the characteristics of low read/write latency, high throughput as well as high reliability, which is efficient to be in line with the requirement of other services, for the instance, computing and communication. Recently, the rapidly increasing global information volumn is posing a challenge on the current storage technologies of portable hard disk, USB flash memory or integrated circuit, which are exposing the deficiencies of short

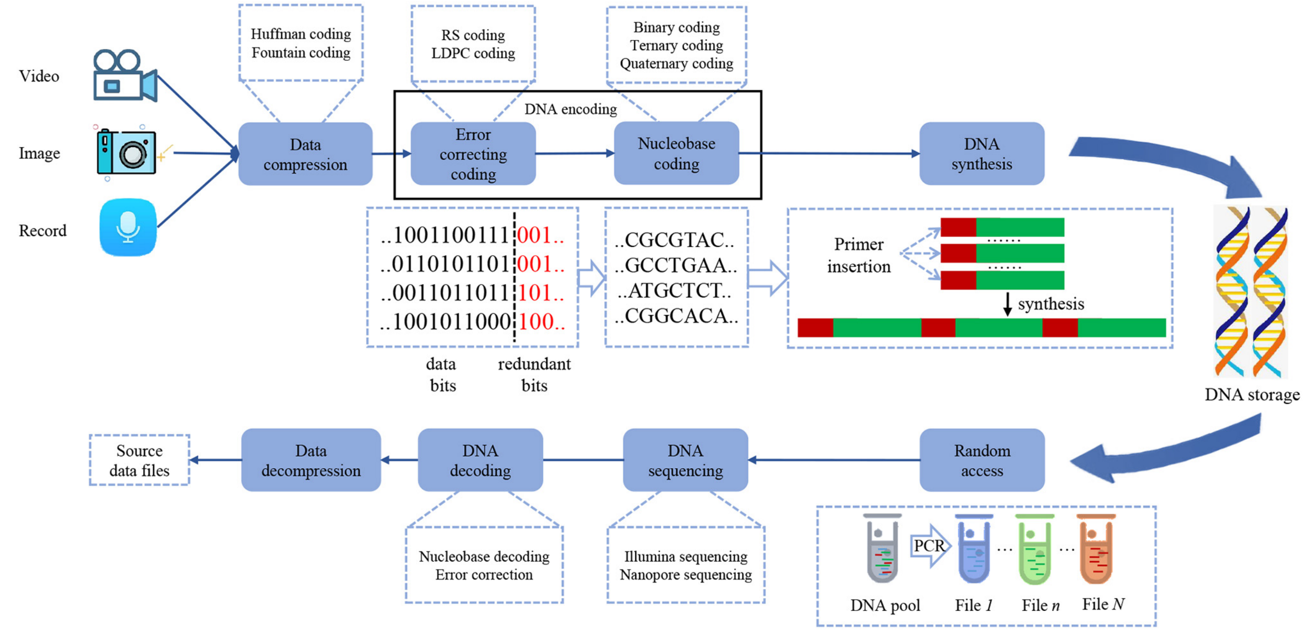


Fig. 12. Working principle of DNA-based storage.

storage lifetime, low stability and environmental pollution. Therefore, DNA-based molecular storage has been thought of as a key technology in the future to cope with these intractable problems.

A. *Characteristics of DNA-Based Molecular Storage*

DNA is responsible for storing genetic information of various kinds of organisms in the world, by synthetising specific nucleobase sequences in the double strand. Inspired by the biological technology, we are attempting to store our data information, e.g., an image, a record or a video, into DNA by exploiting the tremendous genetic diversity of DNA sequences. Compared with the traditional storage technologies, DNA-based molecular storage has the following characteristics.

Long Storage Lifetime: The lifetime of this article-based storage is hundreds of years, while that of stone-based storage could be thousands of years. As for the DNA-based storage, if the storage medium were kept in the dark, dry and cold enviroment, the lifetime could be as long as thousands of years [33]. Moreover, in [34], the experimental results show that the dehydrated DNA can maintain its primary and secondary structure for a long time which beyonds the most actual requirements of conservation, validating that DNA is an ideal medium for the long-term storage in room-temperature.

High Stability: DNA is composed of four kinds of nucleobase (A, G, C, T), each of which has stable chemical properties and is prevented from chemical reaction with surroundings. Meanwhile, the double strands are also built in DNA by the nucleotide pairing to guarantee the correctness in replication and transcription. The unique structure of DNA results in a lower loss ratio of data information and a higher stability.

High Storage Capacity: DNA is considered as an ideal storage media, especially due to its high capacity, since the whole genetic information of organisms all over the world is stored by this approach. A unit of DNA may have the storage capacity of 200PB, which is the same capacity as 0.1 million portable hard disks. According to the statistics, only three boxes of DNA are required to store the whole data information all over the world.

High Replicability: The DNA sequence, which carries the data information of an image, a record or a video, is able to be inserted into the gene of a bacteria, without causing adverse physiological impact. By the aid of the reproductive characteristics of microorganisms, enormous amount of repetitions are produced without any artificial manipulation, which saves lots of labor cost.

Environmental Friendly: As DNA is ubiquitous in nature environment, the zero-pollution feature is a huge advantage compared with other existing storage technologies.

B. *Working Principle*

The working principle of DNA-based storage is illustrated in Fig. 12. First, the data compression is responsible for compressing the source files (a record, an image, a video or others) into less bits by Huffman coding or fountain coding. Then, the redundant error correction bits are inserted, in order to alleviate the impact caused by possible errors during DNA-based storage. Nucleobase coding is indispensible for the DNA storing process, since it provides an appropriate mapping rule between binary files and nucleobase sequences. After inserting some primers into the head of each short DNA strand, the data information is stored into DNA by splicing these short strands into a longer one. The primer is responsible for information retrieving in a vast data archive. Using the technology of polymerase chain reaction (PCR), random access is realized by amplifying the strands in the DNA pool, who have the same primers as the target ones. Then, the information in nucleobase sequence is obtained by DNA sequencing, either using the Illumina sequencing technology or the nanopore

TABLE I
MAPPING RULE OF TERNARY NUCLEOBASE CODING

Previous Nucleobase	**Ternary digital encodes**		
	0	1	2
A	C	G	T
C	G	T	A
G	T	A	C
T	A	C	G

sequencing technology. Finally, the source files are recovered after the DNA decoding and data decompression. In the next part, we select some key technologies in the working principle for detailed explanation.

1) DNA Encoding: DNA encoding consists of two parts, namely error correction coding and nucleobase coding. Error correction coding has the same function as that in traditional wireless communications, which inserts redundant bits after original data information for the sake of detecting and correcting potential errors. Low-density parity-check (LDPC) code, along with reed-solomon (RS) code, are two candidates to efficiently cope with the inevitable errors caused by DNA storage process.

In order to store data information in DNA strands, a mapping rule between binary files and nucleobase sequences is essential, which is the function of nucleobase coding. Three different kinds of approaches for nucleobase coding are proposed, namely, binary coding, ternary coding, and quaternary coding.

1) In the binary nucleobase coding, each binary bit corresponds to two kinds of nucleobases. For an instance, the nucleobases “A” and “G” represent the bit “0,” while the nucleobase “C” and “T” represent the bit “1.” By following these mapping rule, the input binary file “100110011” is encoded as “CAGTCAGTC.”
2) In the ternary nucleobase coding, the current coded nucleobase is also correlated with the previous one in the DNA strands. An example of ternary coding table is examplified in Table I.
3) Quaternary nucleobase coding is the most efficient approach, since each kind of nucleobase carries specific information. For an instance, the data bits “00,” “01,” “10,” “11” could be encoded as “A,” “G,” “C,” “T,” respectively.

Moreover, since DNA homopolymer and a high-content of “C” and “G” nucleobases may cause adverse impacts to DNA synthesis, an approximate DNA encoding method is also indispensible for DNA-based storage.

2) DNA Synthesis: DNA synthesis is the process to link the encoded nucleobase together into a double strand. Due to the exclusivity and susceptibility, DNA synthesis is always operated manually outside the organisms. Meanwhile, short DNA strands are considered as the carrier of data information, since they have better performance on error rate compared with longer DNA strands. For the sake of information retrieving, primers are inserted as the head of each short DNA strands, before they are spliced into a longer one.

3) Random Access: Random access is indispensible in DNA-based storage, since it’s impractical for us to retrieve target data information in a vast data archive. After the DNA synthesis, the preassigned primers are allocated for every DNA short strand carrying specific data information. Then, PCR technology is adopted to amplify the DNA strands in the DNA pool who have the same primer as the target one. A new DNA pool is built after PCR, where exist many duplicates of target DNA strands as well as a few of irrelated DNA strands. Since efficient error correction coding is responsible to cope with this problem, this may not degrade the efficiency of DNA-based storage. Although PCR is a promising approach for random access, it’s impractical to store the whole data information into a single DNA pool, since it may be more time-wasting. An ideal approach is to store the data information into a DNA library, which contains of multi pools. The whole data can be divided into several data segments, and the data segments can be classified into different DNA pools according to the segment’s address in the whole information chain. Then, the double-layered indexing is designed, where a target DNA pool is first selected accroding to the segment’s address before amplifying the target DNA strands using PCR.

C. Recent Advance and Practical Consideration

A lot of experiments have been done during the past decade. In 2013, Goldman *et al.* [33] successfully encoded 739 kB message into DNA molecules by using a new encoding method, while data of 200Mb was written into DNA molecules by University of Washington [35] in 2016. However, since DNA synthesis and sequencing have a considerable latency, the effectiveness could not reach the fairly high level as flash memory during the process of writing or reading. Some researchers are working on the random-access techniques to improve its efficiency of DNA-based storage [36], and thus new constrained coding techniques as well as DNA editing techniques were proposed [39], [40].

In order to reduce the cost of DNA storage, researchers are focusing on new techniques that help compress encoded data and thus control the length of DNA strands. In 2017, Erlich and Zielinski [38] introduced fountain codes to DNA storage architecture, which effectively alleviated the redundancy to 20.71% and greatly reduced the cost of DNA storage. In 2019, Dimopoulou *et al.* [39] first introduced image compression techniques into long-term DNA storage to reduce the length of DNA strands.

As DNA strands need to be duplicated many times and transferred via different channels, additional noise and DNA mutation can occur while synthesizing and sequencing. For example, in an experiment of Goldman, some data were lost after recovered from DNA-based storage [33]. Blawat *et al.* proposed a technique of forward error correction for DNA storage in 2016 [40] to help correct erroneous DNA sequences. Later, researchers from Israel Institute of Technology proposed a method which could recreate the correct DNA strand out of a set of erroneous sequences [41].

Inspite of the above-mentioned advantages compared with tradition storage technology, some practical considerations still pose the challenges on the development of DNA-based storage.

High Cost of DNA-Based Storage: Different with the traditional storage media, DNA-based storage requires more precise and advanced technologies. The great cost lies in the very high purity of reagents needed for synthesis. Therefore, synthesizing high fidelity, long strand DNA, and then sequencing that back to recover the information involve quite a bit of laboratory expertise and labor. However, the cost is coming down recently, as the technology matures and progresses, which indicated that it's more close from the large scale of application of DNA-based storage.

High Read/Write Latency: DNA synthesis and sequencing costs a lot of time, especially when the strand length of DNA is very large. For an instance, retrieving data information from DNA strands using a DNA sequencer is a long process, which takes about 10 h. The resultant high write/read latency degrades the efficiency of DNA-based storage, when compared with other traditional storage technologies. However, high read/write throughput is competitive to make up for the performance degradation caused by the latency. Meanwhile, DNA-based storage is more suitable for some latency-insensitive offline applications.

Reliability Requirement: Unique characteristics of DNA results in the inevitable errors in synthesis and sequencing. The error ratio could be up to 1% per position, which is much higher than other technologies, for an instance, the BER in wireless communications. The maximum strand length should be constrained based on the current synthesis methods and technologies, which is approximately 150 nucleotides. Moreover, effective and robust data compression method and error correction techniques are also indispensible to make DNA-based storage more stable.

Security Requirement: secuirity is the an urgent demand for the iformation storage system. The data stored in the DNA strands can be displayed in the form of A-T or C-G base pairs. To make data more secure from potential attack and embezzlement, DNA steganography and cryptography has applied in the flied of cloud storage. For instance, Murugan and Thilagavathy have proposed an encryption model adding Zigzag pattern to imprve the security of cloud storage [44]. To sum up, it is of great significance to propose a reliable and efficient encryption method based on the DNA characteristics.

V. DNA-BASED MOLECULAR COMMUNICATIONS

DNA-MC is an important means to complete point-to-point communication in a specific environment. The transmitted information is encoded into the DNA molecules using their specific physical or biological characteristics, and the communication between transmitter (Tx) and receiver (Rx) is completed by the diffusion of the molecules. DNA has a special double-strand structure that can store binary bit information via base pairs sequence encoding. Moreover, the length of DNA strands can also be applied to represent the specific binary information in data transmission. Due to the application of multiple characteristics (e.g., base pairs sequence, length, motifs/labels) of DNA strands, using DNA as information carrier between nano-machines can improve the channel capacity to a certain extent. Besides, the simulation results show that using DNA as information carrier in MC can partly reduce the intersymbol interference (ISI) and interlink interference (ILI) [2].

The DNA molecules are released into the body fluid environment, and there are mainly two different scenarios, one is relatively static humoral environment, and other is cardiocirculatory system. Specific molecular diffusion process stands for the information transmission process actually. The movement in former scenario can be characterized by the Brownian diffusion, and the Brownian diffusion is referred to the movement of small particles in a fluid medium, because of the collisions between the small particles, the direction and velocity of molecular motion at any time are random. It has clear advantages that the movement almost cost no external energy, and the diffusive situation can take place spontaneously in the gases and liquid environment [59]. The later scenario is usually refer to the molecular transmission in the blood vessels. In addition to the basic diffusion, the complex molecular movement in the blood vessels includes a positive drift caused by the bloodstream, and the propagation delay can be described as an inverse Gaussian distribution with two determining parameters, i.e., λ (the shape parameter), μ (the mean of the propagation time). The channel parameter estimation and analysis for the MC in the blood vessels are proposed by Chen's group [60]. In this article, the movement of each molecule is regarded as the Brownian diffusion.

Based on the properties of this communication scheme, the application of DNA as inforamtion carrier in the fields of drug delivery, directional communications, water environment detection, body fluid environment detection etc., has aroused widely concern. Moreover, there are multiple application scenarios and the requirements of transceiver varies greatly in different scenarios. Take body fluid environment analysis as an example, in this part, the requirements of the design of Tx and Rx are discussed as follows, and the properties of diffusive channel in DNA-MC are briefly introduced as well.

A. Requirement and Its Difference From EM Communications

1) Transmitter Requirements: In DNA-MC, the Tx end needs to encode the binary information into DNA molecules, and the information can be represented as the form of a nucleotide bases sequence, length, release time, etc. Tx releases the corresponding DNA molecule in predetermined slot to reduce the interference between each kind of molecules. Therefore, the Tx needs to have functional modules for editing, replacing, and cutting the DNA strands. In addition, the container of information molecules and a continuous energy source are also functional units that the Tx should have. Currently, there are mainly two kinds of design ideas for the Tx: 1) the physical controlled structure using nanomaterials and 2) the bacterial structure transformed by biotechnology. The first scheme is suitable for the coordinated control of different nano-machines, and it is necessary to consider how to maintain the sustainability of energy and information molecules. The second scheme has good biocompatibility and has solved the problems of energy and molecule sustainability. However, the bacteria construction that meet the requirements is hard to materialize by biotechnology. The description of each functional requirement on the Tx side is as follows.

Nanoscale Volume: DNA-MC is operating in body fluidic environment, and to realize the properties of communicating with cells and capturing the Bioinformation more accurately [49]. The Tx architecture should minimize the design volume to nanoscale/microscale as far as possible [12].

Energy Consumption: Embedding battery into nanomachines seems to be infeasible, and the solution to enable nano-machines processing in body fluid environment continuously needs to be solved, and energy harvesting (EH) module are currently proposed [12]. Reforming bacteria by biotechnology satisfy the requirement of energy [50], but allowing bacteria to respond regularly to communication tasks is another point of consideration.

DNA Recognition and Editing: In DNA-MC, the Tx and Rx transform information via the diffusion of DNA molecules, so Tx should be capable of distinguishing the DNA base pairs. As introduced above, the information can be encoded into the multiple characteristics of DNA, such as bases sequence, length, releasing slot, molecular concentration etc. An ideal coding mechanism can utilize the multiple information of DNA to encode information. Thus, Tx needs to be capable of distinguishing DNA molecules and synchronous with Rx. Moreover, to transfer instant messages, the Tx can equip the DNA editing function module to add temporary information.

Molecule Sustainability: The nanoscale volume restriction has introduced another important concern, that is the limited storage of DNA molecules. In DNA-MC system, if the information molecules are exhausted, the system will remain useless. Therefore, the replenishment rate of DNA molecules has a huge impact on the system communication performance [55], [56]. The application of genetically engineered bacteria which is capable of producing desired proteins and DNA is a perfect solution of the sustainability problem, and the molecule harvesting method is another ideal scheme [57].

Biocompatibility: The nanomachines are operating in the body fluidic environment, and it's well known that the potential toxicity of materials at nanoscale is nonnegligible. Therefore, the biocompatibility of the whole system has been regarded as an important and challenging issue in this field [12]. In addition to the biological toxicity, it has been proved that the body immune reactions will reduce the lifetime of nanomachines, and these risks can be removed by applying reforming bacteria to the DNA-MC system [58].

Transmission Performance: Actual performance of Tx is a crucial factor that should be considered when it comes to design the structure of Tx. The main performance metrics include unnecessary leakage, release rate, device diameter, transmission range & delay, response time, etc. [12]. All these metrics should be considered according to features of MC system. An ideal Tx should minimize the effects of these metrics to make communication flexible and efficient.

2) Receiver Requirements: Rx is designed for the information collection and processing in the target. Because the DNA molecules move randomly in the body fluidic environment, the Rx should be capable of computing, detecting and information storage individually. The detailed description of functional requirements of Rx are introduced as follows.

Nanoscale Volume: Due to the actual requirements of DNA-based communication, the Rx is supposed to receive the information transmitted from the Tx or gene-engineered bacteria. In body fluidic environment, Rx in nanoscale/microscale gains more flexibility.

DNA Recognition: DNA molecules is used as information carrier in the communication process, so the Rx should be capable of decoding the information embodied in the DNA molecules. Currently, the emergence and development of "biochip" makes it possible for this requirement to become a reality. Lund *et al.* have proposed an amorphous silicon nanostructure for the detection of pH and DNA hybridization [42], and this structure can convert the DNA recognition and binding events to electronic signals. Moreover, the recognition process can be also accomplished by detecting the length of the DNA chain, which can be realized through nanopores.

EH: Because of the limitation of Rx volume, it is not practical to embed the battery in the Rx, and the energy usage of Rx must be considered and optimized. To accomplish the long-term activities, the EH module is also an ideal method in Rx end.

Continuous Operation: The body fluidic environment can be described as a free diffusive model, and the DNA molecules do Brownian motion in this environment. Transmission time and distance do not satisfy the linear relationship, so the Rx is ought to observe the diffusive channel continuously. In this way, the system will get an optimal result of molecules reception ratio [47].

3) Differences From EM Communications: In conventional wireless communication system, the Tx encodes information into the frequency, amplitude and phase of electromagnetic signals. However, the traditional kind of system is not suitable in the body fluid environment. Not only because the body fluid environment may greatly shortens the transmission distance of electromagnetic wave, but also the ideal wavelengths are longer than the diameter of transmission ends, and this phenomenon can lead to the severe diffraction. As a result, DNA molecules are chosen as the information carrier in MC scheme in nano-scale. Compared with the conventional communication system, the diffusion process means relatively low speed, uncertain location and direction, and these features lead to a low arrival probability in Rx side. Moreover, the persistence of different molecules in body fluidic environment can causes signal noise remaining in the channel. The detailed feature of diffusive channel will be introduced in the following part.

B. Working Principle

As is shown in Fig. 13, the DNA-MC can be divided into 5 detailed steps: 1) encoding; 2) emission; 3) diffusion; 4) reception; and 5) decoding [59], and each subprocess has fully combined the characteristics of DNA strands and MC systems. This part is mainly based on the general cognition of DNA-MC system, and some related variables and ICT assessment indicators are also considered.

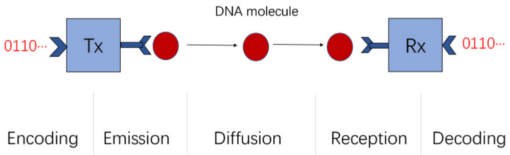


Fig. 13. Five subprocess of DNA-based molecules communication.

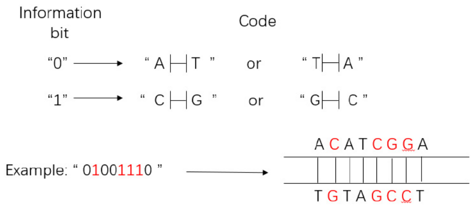


Fig. 14. Base pairs sequence encoding scheme.

1) Encoding: DNA strands are consist of four kinds of nucleobases (adenine [A], thymine [T], cytosine [C], and guanine [G]), and it also has multiple characteristic information, such as base pair sequence, length, proportion of base pairs, etc. Because each base has four choices and the strands have 4^N (N represents the number of base pairs) combinations, this encoding scheme is qualified for large capacity information transfer scenarios. Moreover, the different encoding schemes will influence transmission performance to a great extent, transmission capacity and computational complexity are often two tradeoff options. Choosing a suitable encoding mode based on the system requirements is very important. Following are two widely used encoding methods in DNA-MC.

1) *Base Pairs Sequencing Encoding:* Tx is capable of assembling the DNA fragments according to the predetermined method. Even though there are 4 types of nucleobases, each of them always pair with another one, that is A-T base pairs and C-G base pairs. Because the double helix structure has two opposite directions, take one strand encoding scheme into consideration, there are four types of nucleobases, each nucleobase can represent two bits of information. Such as "A" represents bits "00," "C" represents bits "01," "G" represents bits "10," and "T" represents bits "11." This kind of modulation greatly increased the encoding efficiency, but the Rx must figure out the right coded single strand. Otherwise, it will get the completely wrong decoding information. For simplification, the proposed system has adopted "base pair encoding scheme," that is "A-T" and "T-A" pairs represent the bits "1" samely, while "C-G" and "G-C" stands for the bits "0." The encoding method is described in Fig. 14.
2) *Length Encoding:* In base pair sequence encoding scheme, the transceiver should be capable of recognizing and editing the nucleobases, and it has increased the design difficulty of Tx and Rx due to the limitations of nano-volume and processing time. Another desirable method is length encoding, and the detailed encoding process is described in Fig. 15. On the condition that the Tx transmits only one bit of information per slot. For instance, 4 slots are required to transmit the information sequence "0110," and the length of DNA molecules sent at each time slot is different. As shown in Fig. 15, the first bit "0" means that the Tx should not the DNA molecules of 3 kilo base pairs (kbp), and the second bit "1" means the Tx should release the DNA molecules of 4.5 kbp at the second slot, etc. The Rx determines the information of each bit by detecting the length of the DNA molecules arriving at the receiving end. Moreover, the number of different length molecules, the length increment and the duration of each slot have a significant influence on the system performance [2]. Thus, these parameters need to be optimized according to the actual demands.

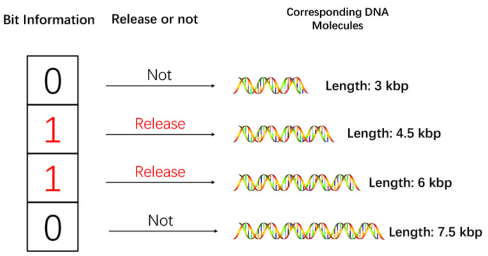


Fig. 15. Length encoding scheme.

2) Emission: Assuming that the transmission is deployed in 1-D environment, and molecules can be treated as individual points in the diffusion process. The diffusive rate of molecules is depending on the molecules size and environment concentration, and the diffusive movement may cause wrong reception in Rx side. Therefore, the emission of molecules should be fixed in specific time slots, and the Tx only release one kind of DNA molecules during one slot. If the Tx and Rx is synchronized, Rx will receive the DNA packets one by one in correct order. The transmission process become more clear in this situation, and the wrong reception ratio surely will reduce to some extent. Another point of consideration is the Tx is supposed to avoid the leakage of molecules. Due to the stable biological characteristics of DNA strands, the information molecules will exist in the fluidic environment for quite a long time, and this phenomenon will introduce undesirable noise and wrong reception in Rx side.

3) Diffusion: Diffusion is the main transport mode of information molecules in the body fluid environment. After DNA molecules are encoded and released in Tx side. Molecules of nano-size move randomly in the environment of liquid, which can be described as Brownian motion. Therefore, the probability that the molecule will reach the point with a distance of d after a time t from the origin is [48], [51]

$$f(d,t) = \frac{d}{\sqrt{4\pi \mathrm{Dt}^3}} e^{-\frac{d^2}{4\mathrm{Dt}}}, \quad t > 0. \tag{1}$$

In the formula, D is the diffusion coefficient, and it is related to the number of base pairs of DNA strands [43]

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

Assume that the molecules are released at intervals of time slot T_s , the arrival probability can be further described as

$$f(d,t) = \frac{d}{\sqrt{4\pi D[t-(i-1)T_s]^3}} e^{-\frac{d^2}{4D[t-(i-1)T_s]}}. \tag{3}$$

Here, T_s refers to the length of each slot, and the probability of the *ith* packet arrives at the Rx within time t can be derived by calculating the integral of the formula. Because of the randomness of Brownian motion, the diffusive channel will disrupt the arrival sequence of information packages, and it can greatly decrease receiving efficiency [46]. The Rx should set up a suitable mechanism to ensure correct reception.

Moreover, mRNA can also be used for the communication part which can be an interesting direction to investigate the communication 'chain reaction' consist of 'DNA-mRNA-protein (enzyme)-DNA'.

mRNA, the single strand message RNA, which is the product of DNA transcription process associated with RNA polymerase, is also an intriguing and fundamental component of DNA-based computing and communication paradigm. With similarity to DNA, mRNA, regarded as the input signal, is regulated by co-factor enzymes. For instance, benefiting from hybridization chain reaction (HCR) of DNA amplification [52], mRNA is amplified massively deriving from corresponding DNA domains, rusting an amplified signal. Thus, mRNA is implemented as an activator or repressor in DNA computing, which enhances or suppresses the rate of gene transcription, respectively. According to [53], a synthesis genetic current is established to mimic the dynamics of an electrical circuit. The methodology of generating synthetic gene networks is introduced to regulated the behavior of living organisms [54].

Furthermore, A chain of reactions, involving three processes, including transcription process from DNA to mRNA, translation process from mRNA to enzymes proteins, catalysis of enzymes back to DNA. Therefore, mRNA plays a critical role in modulation process of enzyme reaction-based MC. However, the research to date has tended to focus on DNA-MC individually rather than the integrated mRNA-DNA-MC.

4) Reception: Reception is the process of capturing the diffusive molecules in Rx, and the reception ratio is a determined factor of communication efficiency. Some special designs of reception are trying to realize the error free reception. The nanopore structure of Rx is utilized to improve the reception efficiency [1]. Nanopores are designed based on the physical characteristics of DNA molecules. The surface of nanopores is charged, so the molecules which are negative charged can pass through, and diameter of nanopores are predetermined that each pore can only pass through one molecule. Due to the diffusion characteristics of body fluidic environment. The arrival time and probabilities are stochastic at the Rx side, and there is always the situation that some nanopores are capturing the molecules while others are not. Therefore, the number of nanopores and releasing molecules can be optimized for the communication performance.

5) Decoding: For the base pairs sequence encoding in Tx, there is the corresponding decoding method in the Rx. The decoding process in transferring biological information into binary information by base pairs recognition technology. One of the advantages of DNA as information carrier is the stable double helix structure, and no specific error correction measures are introduced in the communication process, the decoding accuracy at the Rx can nearly achieve 100%. However, the phenomenon of nucleotides change into another type also exist. The electromagnetic radiation, oxidizing agent, etc. can play the role of mutagens which lead to mistaken assembling and mutations in DNA-based packets.

For the DNA length encoding scheme, when the molecules are pass the nanopores, the surface current will be blocked, and the different strand length corresponds to the different blocking time. The Rx can figure out each bit information according to the current state. Due to the randomness of arrival time, the Rx has set a decision threshold, if the number of detected molecules of a certain length is greater than the threshold, the corresponding bit will turn into "1." Otherwise, it will remain "0."

Because of the limitations of DNA strands length, the large capacity information will be divided into dozens of DNA molecules, and the molecules may not arrive in the right order after the diffusion. Assuming that the Tx and Rx are time synchronized, increasing the time interval of releasing different molecules may eliminate the phenomenon of wrong reception. Moreover, adding locating bits into the DNA packets to connect different DNA molecules is another solution, but that is hard to materialize due to the bottleneck of DNA editing technology. To sum up, more complex and innovative system design will emerge if both the system design and the biotechnology get developed in the long run.

C. Recent Advances

DNA-MC is a new interdisciplinary research field. Compared with traditional MC, DNA as an information carrier can effectively improve the information transmission rate and reduce the ISI. In the design of communication system, how to apply the multiple features of DNA molecules to MC system is the focus of consideration now. According to the research status in this field, several innovative ideas are found.

In [2], the length encoding scheme is introduced in the MC system to realize high capacity communication between nanomachines. In traditional MC system design, the processing time in Rx is often neglected, but this encoding scheme considers the effect of processing times which are related to the DNA lengths. Furthermore, the processing time can be used to distinguish different molecules in the Rx side, then arrival probability of specific DNA molecules can be caculated. As is shown in the stimulation of ICT metrics, the length encoding MC system can achieve capacity levels up to 6 bps.

In [59], the communication system has applied the base pairs sequence encoding, the matching degree between arrival order of DNA molecules and the released order has crucial impacts on the channel capacity. Therefore, to analyse the ICT metrics, the article define the observation window and focus slot for evaluating whether the arrival packets are consistent with the sending sequences. The window is the observation time, and the focus slot is used for checking the order. Only if the Rx receive the molecule which is released by the Tx in focus slot, the DNA packet is transmitted successfully. Therefore, the mutual information can be derived based on this assumption. In this model, Sun *et al.* analyze the influence of arrival order on channel capacity by mathematical derivation.

D. Practical Considerations

The DNA-MC system design is on the theoretical level, when we start to think about the implementation of the system: 1) the technical bottlenecks; 2) transmission performance; and 3) realization cost should be considered. This field is an interdisciplinary research direction, and the achievements in MC greatly depend on the development of related biotechnology. The existing biological and nano-material technology gives us the basic design of the transceiver, but in order to complete more complex communication tasks, there is still a large space for improvement in the structure design. As is known, DNA editing technology has become more and more mature, but how to improve the processing speed of DNA strand in the instant communication scenario is another point of consideration.

In order to realize the large capacity transport properties of DNA, some system indexes need to be optimized, such as DNA strand length, length increment, number of nanopores, transport distance, slot length, etc. These indicators often have a significant impact on communication performance. However, it is a tradeoff problem to improve the performance and the difficulty of implementation. It is an ideal solution to set the system parameters according to the actual needs of the system.

When considering the large-scale implementation of DNA-MC systems, the actual cost evaluation is unavoidable. According to existing DNA editing techniques, the cost of synthesizing DNA is currently about $ 12400/MB [33]. Given the high cost of synthesis and high latency of DNA writing and reading, using DNA as information carrier is currently not feasible. It takes a long time to accomplish the encoding and decoding process via biotechnology. How to apply this technology to satisfy the instantaneity of communication is also a heated point.

VI. FUTURE DIRECTIONS

A. Directional Communications

MC has numerous advantages, compared to electromagnetic communication systems, in zero energy propagation, robustness in challenging environments, information storage capacity, and feasibility for intrabody system. However, it has some defects, such as low data rates, high transmission delay, and instability to the surrounding. When the propagation occurs between Tx and Rx with a long distance, the robustness of transmission would decline severely. Normally, it takes a significant amount of time if more than few μm exists between Tx and Rx. To improve the efficiency of propagation process, the directional communication is necessary in MC. In some special MC applications, such as targeting therapy of tumor, it is very important to keep all drug molecules moving along a fixed direction to the location of tumor cell. In the latest research of MC, cancer cell detection or targeted drug delivery becomes a hot area as their great application future in the directional MC.

A typical directional MC is based on Chemotaxis, as shown in Fig. 16, the mobile node (i.e., nanomachine) could swim toward to the direction of higher concentrations of preferred chemical molecules (i.e., attractant), which are released by the target, such as tumor cells. Taking advantage of this feature, the artificial nanomachine can carry drugs to find the tumor cells directionally, and release drugs around them. Many bacteria have the characteristic of Chemotaxis. *E.coli* is one of bacteria with this characteristic, that can recognize more than 12 diverse attractants and the attractant concentration gradient could be remembered though its short-term memory. *E.coli* can select a favorite attractant to move to with a precise navigation. *E.coli* gets the propulsion through its flagella, which is a long tail-like appendage, to propel and rotate during the process of the movement. Bacteria could move toward the direction of higher attractant concentration due to the ability to sense preferred attractants' concentration gradient. There are some literatures investigating the directional communication based on chemotaxis. In [61], nanomachines who detect the position of target are able to release some attractants around the target, helping other nanomachines to find the area of target rapidly. In [62], two types of attractants are used to extend the scope of concentration gradient, so that deep targets can be detected using a decentralized rule. Much deeper works on this field will be kept on studying in the future.

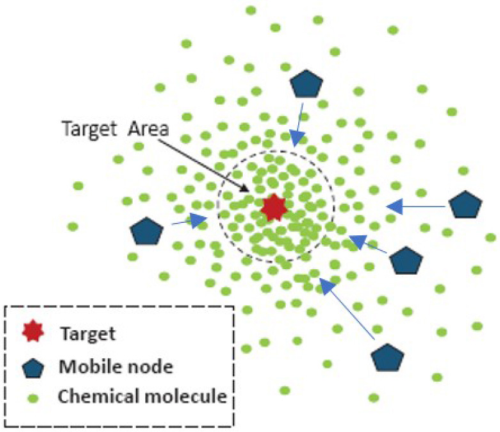


Fig. 16. Mobile MC system for target detection.

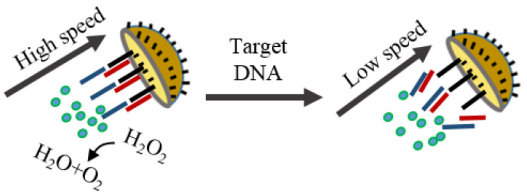


Fig. 17. Schematic of micromotor-based DNA sensing.

Due to some special characteristics, DNA could be employed in directional communication. For instance, the characteristic of active motion of nanomotors could be utilized to detect the target DNA. We can use the signal of speed change of nanomotor interacting with the target to estimate the position of target. In [63], a jellyfish-like micromotor is employed to detect the target DNA, as shown in Fig. 17. On the dented surface of the micromotor, the catalase (a substance carried in the hybrid DNA) is placed at first. When the target DNA is found, micromotors release the catalase to result in slowing down on the motion speed of micromotor. Thus, we can detect the target DNA through recording the signal of micromotor speed.

miRNAs detection is another significant application in clinical cancer diagnosis. The change of miRNAs is related to cancer initiation, so it is very important to real-time monitor the miRNAs for cancer diagnosis and prognosis. Nanomotors has the ability to recognize the target miRNA once its surface is assembled with different biological molecules or receptors. In [64], an interesting method is proposed to detect the target miRNA-21 (a kind of endogenous content of the microRNA-21) with intracellular “off-on” fluorescence. The target miRNA can course displacement of the dyed ssDNA probe, and fluorescence switches swiftly the state from “off” to “on.” Thus, this nanomotor-based detecting method provides new ways to detect target in directional MC.

As vast information can be encoded into the DNA strands, it is foreseeable that DNA-based communication is high efficiency interaction between nanomachines. This feature can be used in cooperation among nanomachines to perform more complicated tasks, such as group directional motion. MCs via diffusion is delay-sensitive communications. Existing modulation schemes are not suitable to communicate between nanomachines because of the serious delay. DNA-MC can transmit a large amount of data each time by means of bacteria carrying DNA with coded information. This function makes it possible to cooperate between nanomachines in a complicated nanonetwork topology and it could expend the directional communication function from single molecule to a group of molecules. Recently, more applications of bionano-machines and the DNA-MC can be used in Internet of Nanothings (IoNT) or Internet of Bio-Nanothings (IoBNT), especially in the industrial and biomedical. One of them is that bionano-machines are suggested as the communication scheme between biological entities and their biocyber interfaces in IoBNT for body health monitoring.

B. Joint Designs of DNA Coding and Addressing

Though much research work has been conducted on DNA-based computing, data storage and MCs, they are often carried out separately. Shah *et al.* have realized the necessity and significance of combining DNA-based data storage into an MC system [11] but did not provide a joint design. Rather it presented a DNA-based modulation scheme, which only focuses on the performances on communications but does not take storage into account. It is obvious that there are many common properties between storage and communication. For instance, both data storage and communications require the same process: coding. Storage coding is usually for error correction as the information stored in DNA may get eroded and corrupted due to the influence of physical and chemical factors. Channel coding is compulsory for any communication system to cope with unavoidable errors during transmission as introduced by various interferences, such as interchannel interference (ICI), ISI, etc. They are correlated and thus a joint design has a great potential to lead to optimized solutions.

Another example is addressing, which is largely neglected in the literature. Addressing is a common mechanism in communication networks to identify a node (e.g., the famous IP address) or a packet (e.g., TCP sequence number). Meanwhile, this mechanism is also applicable to DNA-MC networks. Addressing is equally needed in DNA-based data storage systems to enable accurate random access of a desired data in large sized DNA storage because it is infeasible to always start from the beginning to retrieve a piece of information. It is worthwhile to investigate correlations between these two types of addressing as this may result in a streamlined addressing scheme across both storage and communications. Furthermore, experimental work is encouraged to make sure the proposed scheme is easy to operate in bioengineering sense.

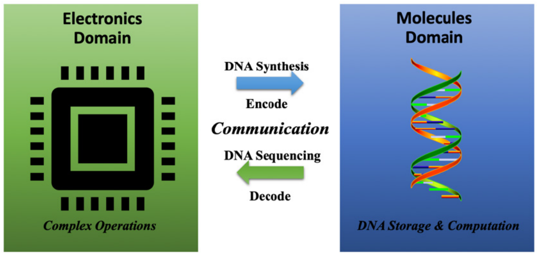


Fig. 18. Communication between electronics domain and molecules domain.

C. Hybrid Molecular-Electronic Systems

Since the second industrial revolution, human has stepped into the age of electricity and almost everything today is directly or indirectly related to electronic things. While exploring and exploiting molecular world, it can be very beneficial if we can leverage both strengths of molecules and electronics by hybridizing molecular domain with electronic domain. For example, [65] envisions that DNA based computer could plausibly be 103 times faster than the fastest current supercomputer, because of the utilization of advantageous properties (faster speed, better energy efficiency and larger storage) of DNA computing.

In 2019, University of Washington built a hybrid molecular-electronic system model [1]. The key to the system is to use DNA as storage medium and computing substrates. As data and computing units both exist and float in solution, computation can be done simultaneously and parallelly and thus theoretically achieves infinite bandwidth. Similar conclusion was obtained in [65], that such DNA-based computer system could potentially utilize more processors than all electronic computers in the world combined.

As with most heterogeneous system, the communication between different parts is of vital importance. The communication within hybrid molecular-electronic system relies on DNA synthesis(write) and sequencing(read). Fig. 18 Shows a possible process of transferring data from electronic domain to molecular domain and vice versa. Actually, the device that connects the two domains is nano-machine. Many nanothing components have already been prototyped and tested. However, there are still several challenges from the device perspective that need to be addressed in order to realize a nanomachine. On the other hand, from existing literature, the feasible communication technology is terahertz. Akyildiz *et al.* [66]

described the application scenario of bionanothings network and [67] points the challenges to use terahertz band in Internet of Nano-Things. Although molecule domain is not as precise as electronic domain, it can offer orders of magnitude improvements of performance with its parallel operations and extremely large throughput. However, the latency of sequencing DNA is another issue. It is proportional to the length of DNA strand being sequenced, as discussed in [2], which used nanopore to sequence DNA strand with latency taken into account. In addition, the interfaces between two domains are also limited. It is better to send a small amount of information and make most work be done in parallel. This way, it just acts like GPUs which compute parallelly and thus improve the performance. In addition, by leveraging DNA SDRs, some researchers have developed a novel DNA-based switching circuit that could work in molecular domains and do arbitrary digital computing [28].

In reality, there are many existing constraints and there is a long way to go to fully realize such hybrid systems. Both the cost and delay of DNA reading/writing need to be reduced to be practical and DNA computing is still being studied and could not provide desired performance. In addition, the necessary operations like DNA synthesis and sequencing are done by human in a lab, but a real system need to be capable of doing so automatically. Our current microfluidic technology could not yet support such automation.

D. Cross-Function Molecular ICT System Architecture

Section III focuses on how to use DNA to design logic gates in order to carry out calculation and processing. This is analogous to a processor, or precisely the arithmetic and logic unit (ALU) within a processor/central processing unit (CPU), in the classical von-Neumann's computer architecture. However, as in a digital electronic computer, a DNA computer needs an integration of DNA processor, DNA memory (or storage in general) and DNA-enabled communications. This gives rise to the need to discuss the interconnection of these three functions that have been presented in the previous three sections, computation, storage and communications.

We will start with a DNA computer (i.e., a cross-function DNA ICT system in its own right) that follows the fundamental concepts of electronic computers, i.e., Turing machine and von-Neumann architecture, as one major purpose of building biological computers is to assist or even replace their counterpart electronic computers. This will also give us a better understanding of DNA computers. Then our discussions will go beyond the boundary of current computer's von-Neumann architecture and explore the specialities of DNAs and its impact on biocomputers and bio-ICT systems.

1) Following Turing Machine and von Neumann Architecture: The current digital computation is fundamentally based on the computing model proposed by Alan Turing in the 1930s [68], named as the Turing machine, which defines the key concept of computability. This set up the foundation for modern computation. Its basic procedure is for a processor to take inputs, to carry out processing and then to produce outputs. It does not mention stored programs and assumes unlimited memory. Von Neumann later proposed the famous von-Neumann computer architecture [69] to fulfil the type of computations described by the Turing Machine. In addition to inputs, outputs and processing, this architecture also introduces memory/storage to store processing programs, data and execution status, as illustrated in Fig. 19(a).

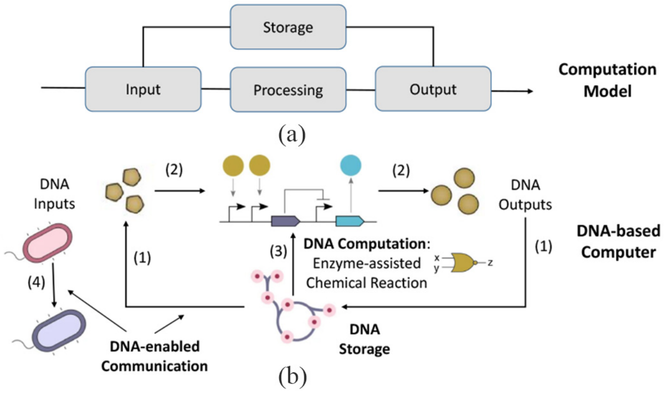


Fig. 19. DNA-based computer architecture.

In a direct comparison with these two key concepts in electronic computers, we propose an equivalent DNA-based computer as shown in Fig. 19(b). Not like its electronic counterpart which can use only electronic signals, DNA-containing biological units (such as cells) are able to input/output a wide range of physical, chemical and biological signals and information. For instance, blood pressure can be encoded as a string of DNA. DNA computation, enabled by DNA processors, is typically carried out in terms of enzyme-assisted chemical reactions. These chemical reactions will generate various combinations, out of which the wrong combinations will be eliminated to get the right coding sequences (the right answers). This process needs to involve synthetic biological engineering, which applies engineering methods to living things, such as bacteria, to provide new substrates for computation [70], rather than the traditional silicon used by electronic devices. Depending on computation capability, DNA computation may involve from simple gene regulation, to metabolism and to biological evolution [71].

DNA storage is responsible for storing programs and their execution status in genetic information forms by synthetizing specific nucleobase sequences in the DNA's double strand. Massive amount of information can be stored thanks to the tremendous genetic diversity of DNA sequences. Metabolism presents a volatile memory, while DNA sequences are able to store information in a more stable fashion [71].

As illustrated and numbered in Fig. 19(b), DNA communications serve the following four purposes.

1) Data transfer between the input/output devices and the DNA storage: input devices can take data from the storage in addition to their own data and output devices can store the generated data back into the storage.
2) Data transfer between input/output devices and the DNA processor as they may not be connected directly.
3) Data transfer between the DNA processor and the DNA storage: the processor may retrieve data from the storage

to assist the ongoing processing and may (temporarily or permanently) store its intermediary or final results back into the storage.

4) Communications between different cells/biocomputers.

2) Going Beyond von Neumann Architecture: Fig. 19 provides a conceptual mapping from electronic computer to DNA computer based on von Neumann architecture. Much research on DNA-based computational models has been focused on molecular emulation of logic gates and circuit models, as presented in Section III of this article. However, this may fundamentally limit the types of functions that may be engineered inside a cell as it does not exploit the inherent and diverse features of DNAs and living things that are not available in electronic machines. As a matter of fact, molecular computers cannot compete again their electronic counterparts in terms of speed and cost, at least given the foreseeable technologies. We would argue, like the authors in [71], the real power of DNA-based ICT systems, including computation, storage and communications, lies in DNA's inherent specialities, such as analogue, parallel processing, low energy consumption, and energy self-supply, e.g., cells can absorb ambient nutrition to get energy), and stochastic nature. Exploring these features will lead to a more powerful and promising DNA-enabled ICT era.

VII. CONCLUSION

Much research work has been conducted on DNA-based computing, data storage and MCs in recent years, partially ignited by the rapid advancement of DNA engineering technologies. This article has provided an overview of each from the following aspects: requirements and differences from current electronic-dominating technologies, their working principles, recent advancement and practical considerations. However, these three aspects of researches are largely carried out in isolation. This article encourages a joint design of them. For this purpose, this article has proposed a molecular ICT system architecture with all its three components underpinned by DNAs. The following future directions of DNA-enabled molecular ICT have also been discussed and their recent development being summarized: joint coding for storage and communications, directional DNA communications, interfaces between molecular DNA systems and electronic systems. It is the authors' belief that DNA-based molecular ICT systems will develop even more rapidly and in a cross-function manner in the coming years.

REFERENCES

[1] D. Carmean, L. Ceze, G. Seelig, K. Stewart, K. Strauss, and M. Willsey, "DNA data storage and hybrid molecular–electronic computing," *Proc. IEEE*, vol. 107, no. 1, pp. 63–72, Jan. 2019.

[2] B. A. Bilgin, E. Dinc, and O. B. Akan, "DNA-based molecular communications," *IEEE Access*, vol. 6, pp. 73119–73129, 2018.

[3] M. S. Neiman, "Some fundamental issues of microminiaturization," *Radiotekhnika*, vol. 2, no. 1, pp. 3–12, 1964.

[4] L. M. Adleman, "Molecular computation of solutions to combinatorial problems," *Science*, vol. 266, no. 5187, pp. 1021–1024, Nov. 1994.

[5] D. Y. Zhang and G. Seelig, "Dynamic DNA nanotechnology using strand-displacement reactions," *Nat. Chem.*, vol. 3, pp. 103–113, Jan. 2011. [Online]. Available: https://doi.org/10.1038/nchem.957

[6] P. Raut and N. Sarwade, "Coverage modelling of molecular communication based nano machines network for the IoT," in *Proc. IEEE Region 10 Symp.*, Ahmedabad, India, 2015, pp. 21–24, doi: 10.1109/TENSYMP.2015.7.

[7] M. Kuscu and O. B. Akan, "Modeling and analysis of SiNW BioFET as molecular antenna for Bio-cyber interfaces towards the Internet of bio-NanoThings," in *Proc. IEEE 2nd World Forum Internet Things (WF-IoT)*, Milan, Italy, 2015, pp. 669–674, doi: 10.1109/WF-IoT.2015.7389134.

[8] M. Bansal, S. Gupta, and S. Mathur, "Comparison of ECC and RSA algorithm with DNA encoding for IoT security," in *Proc. 6th Int. Conf. Inventive Comput. Technol. (ICICT)*, Coimbatore, India, 2021, pp. 1340–1343, doi: 10.1109/ICICT50816.2021.9358591.

[9] E. B. Baum, "Building an associative memory vastly larger than the brain," *Science*, vol. 268, no. 5210, pp. 583–585, Apr. 1995.

[10] L. C. Meiser *et al.*, "Reading and writing digital data in DNA," *Nat. Protocol*, vol. 15, pp. 86–101, Jan. 2020. [Online]. Available: https://doi.org/10.1038/s41596-019-0244-5

[11] S. Shah, A. Raghavachari, C. Lo, and R. Marculescu, "Molecular communication with DNA cellular storage system," in *Proc. 4th ACM Int. Conf. Nanoscale Comput. Commun. (NanoCom)*, 2017, p. 24.

[12] M. Kuscu, 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," *Proc. IEEE*, vol. 107, no. 7, pp. 1302–1341, Jul. 2019.

[13] S. A. Tsaftaris, A. K. Katsaggelos, T. N. Pappas, and T. E. Papoutsakis, "DNA-based matching of digital signals," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Montreal, QC, Canada, May 2004, pp. 581–584.

[14] M. H. Caruthers, "The chemical synthesis of DNA/RNA: Our gift to science," *J. Biol. Chem.*, vol. 288, no. 2, pp. 1420–1427, 2013.

[15] S. Palluk *et al.*, "De novo DNA synthesis using polymerase-nucleotide conjugates," *Nat. Biotechnol.*, vol. 36, pp. 645–650, Jun. 2018.

[16] W. Li, F. Zhang, H. Yan, and Y. Liu, "DNA based arithmetic function: A half adder based on DNA strand displacement," *Nanoscale*, vol. 8, no. 6, pp. 3775–3784, 2016.

[17] B. Yurke, A. J. Turberfield, A. P. Mills Jr., F. C. Simmel, and J. L. Neumann, "A DNA-fuelled molecular machine made of DNA," *Nature*, vol. 406, pp. 605–608, Aug. 2000.

[18] A. J. Turberfield, J. C. Mitchell, B. Yurke, A. P. Mills Jr., M. I. Blakey, and F. C. Simmel, "DNA fuel for free-running nanomachines," *Phys. Rev. Lett.*, vol. 90, no. 11, p. 4, 2003.

[19] W. Li, Y. Yang, H. Yan, and Y. Liu, "Three-input majority logic gate and multiple input logic circuit based on DNA strand displacement," *Nano Lett.*, vol. 13, no. 6, pp. 2980–2988, 2013.

[20] Y. Benenson, B. Gil, U. Ben-Dor, R. Adar, and E. Shapiro, "An autonomous molecular computer for logical control of gene expression," *Nature*, vol. 429, no. 6990, pp. 423–429, 2004.

[21] J. Elbaz, O. Lioubashevski, F. Wang, F. Remacle, R. D. Levine, and I. Willner, "DNA computing circuits using libraries of DNAzyme subunits," *Nat. Nanotechnol.*, vol. 5, no. 6, pp. 417–422, 2010.

[22] M. Kahan-Hanum, Y. Douek, R. Adar, and E. Shapiro, "A library of programmable DNAzymes that operate in a cellular environment," *Sci. Rep.*, vol. 3, p. 1535, Mar. 2013.

[23] A. A. Tregubov, P. I. Nikitin, and M. P. Nikitin, "Advanced smart nanomaterials with integrated logic-gating and biocomputing: Dawn of theranostic nanorobots," *Chem. Rev.*, vol. 118, no. 20, pp. 10294–10348, 2018.

[24] R. D. Barish, R. Schulman, P. W. K. Rothemund, and E. Winfree, "An information-bearing seed for nucleating algorithmic self-assembly," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 15, pp. 6054–6059, 2009.

[25] W. Li, Y. Yang, S. Jiang, H. Yan, and Y. Liu, "Controlled nucleation and growth of DNA tile arrays within prescribed DNA origami frames and their dynamics," *J. Am. Chem. Soc.*, vol. 136, no. 10, pp. 3724–3727, 2014.

[26] L. Qian and E. Winfree, "Scaling up digital circuit computation with DNA strand displacement cascades," *Science*, vol. 332, no. 6034, pp. 1196–1201, 2011.

[27] N. Shimada, K. Saito, T. Miyata, H. Sato, S. Kobayashi, and A. Maruyama, "DNA computing boosted by a cationic copolymer," *Adv. Funct. Mater.*, vol. 28, no. 17, pp. 1–6, 2018.

[28] F. Wang *et al.*, "Implementing digital computing with DNA-based switching circuits," *Nat. Commun.*, vol. 11, no. 1, pp. 1–8, 2020.

[29] H. Hermassi, R. Rhouma, and S. Belghith, "Improvement of an image encryption algorithm based on hyper-chaos," *Telecommun. Syst.*, vol. 52, pp. 539–549, May 2011.

[30] B. Mondal and T. Mandal, "A light weight secure image encryption scheme based on chaos & DNA computing," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 29, no. 4, pp. 499–504, 2017.

[31] V. Zhirnov, *2018 Semiconductor Synthetic Biology Roadmap*, NIST, Gaithersburg, MD, USA, Oct. 2018, p. 36.

[32] J. R. Rubens, G. Selvaggio, and T. K. Lu, "Synthetic mixed-signal computation in living cells," *Nat. Commun.*, vol. 7, Jun. 2016, Art. no. 11658.

[33] N. Goldman *et al.*, "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA," *Nature*, vol. 494, no. 7435, pp. 77–80, 2013.

[34] J. Bonnet *et al.*, "Chain and conformation stability of solid-state DNA: Implications for room temperature storage," *Nucleic Acids Res.*, vol. 38, no. 5, pp. 1531–1546, 2010.

[35] J. Bornholt, R. Lopez, D. M. Carmean, L. Ceze, G. Seelig, and K. Strauss, "Toward a DNA-based archival storage system," *IEEE Micro*, vol. 37, no. 3, pp. 98–104, Jun. 2017, doi: 10.1109/MM.2017.70.

[36] S. M. H. T. Yazdi, Y. Yuan, J. Ma, H. Zhao, and O. Milenkovic, "A rewritable, random-access DNA-based storage system," May 2015. [Online]. Available: https://arxiv.org/abs/1505.02199

[37] L. Organick *et al.*, "Random access in large-scale DNA data storage," *Nat. Biotechnol.*, vol. 36, no. 3, pp. 242–248, 2018.

[38] Y. Erlich and D. Zielinski, "DNA fountain enables a robust and efficient storage architecture," *Science*, vol. 355, no. 6328, pp. 950–954, 2017. [Online]. Available: https://science.sciencemag.org/content/355/6328/950

[39] M. Dimopoulou, M. Antonini, P. Barbry, and R. Appuswamy, "A biologically constrained encoding solution for long-term storage of images onto synthetic DNA," 2019. [Online]. Available: arXiv:1904.03024.

[40] M. Blawat, K. Gaedke, I. Huetter, X. M. Chen, and G. M. Church, "Forward error correction for DNA data storage," *Procedia Comput. Sci.*, vol. 80, pp. 1011–1022, 2016, doi: 10.1016/j.procs.2016.05.398.

[41] M. A. Sini and E. Yaakobi, "Reconstruction of sequences in DNA storage," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 290–294, doi: 10.1109/ISIT.2019.8849740.

[42] J. Lund, C.-J. Wang, and B. Parviz, "Electronic detection of DNA with amorphous silicon nanostructures," in *Proc. 19th IEEE Int. Conf. Micro Electro Mech. Syst.*, Istanbul, Turkey, 2006, pp. 474–477, doi: 10.1109/MEMSYS.2006.1627839.

[43] L. Lukacs, P. Haggie, O. Seksek, D. Lechardeur, N. Freedman, and A. S. Verkman, "Size-dependent DNA mobility in cytoplasm and nucleus," *J. Biol. Chem.*, vol. 275, no. 3, pp. 1625–1629, 2000.

[44] A. Murugan and R. Thilagavathy, "Cloud storage security scheme using DNA computing with morse code and zigzag pattern," in *Proc. IEEE Int. Conf. Power Control Signals Instrum. Eng. (ICPCSI)*, Chennai, India, 2017, pp. 2263–2268, doi: 10.1109/ICPCSI.2017.8392120.

[45] Y. Mori, Y. Kuroe, and T. Mori, "A synthesis method of gene networks based on gene expression by network learning," in *Proc. SICE-ICASE Int. Joint Conf.*, Busan, South Korea, 2006, pp. 4545–4550, doi: 10.1109/SICE.2006.315086.

[46] M. Pierobon and I. F. Akyildiz, "A statistical–physical model of interference in diffusion-based molecular nanonetworks," *IEEE Trans. Commun.*, vol. 62, no. 6, pp. 2085–2095, Jun. 2014, doi: 10.1109/TCOMM.2014.2314650.

[47] A. Noel, K. C. Cheung, and R. Schober, "Optimal receiver design for diffusive molecular communication with flow and additive noise," *IEEE Trans. NanoBiosci.*, vol. 13, no. 3, pp. 350–362, Sep. 2014, doi: 10.1109/TNB.2014.2337239.

[48] A. Noel, K. C. Cheung, and R. Schober, "Joint channel parameter estimation via diffusive molecular communication," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 1, no. 1, pp. 4–17, Mar. 2015, doi: 10.1109/TMBMC.2015.2465511.

[49] Q. H. Abbasi *et al.*, "Nano-communication for biomedical applications: A review on the state-of-the-art from physical layers to novel networking concepts," *IEEE Access*, vol. 4, pp. 3920–3935, 2016, doi: 10.1109/ACCESS.2016.2593582.

[50] B. D. Unluturk, A. O. Bicen, and I. F. Akyildiz, "Genetically engineered bacteria-based biotransceivers for molecular communication," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1271–1281, Apr. 2015, doi: 10.1109/TCOMM.2015.2398857.

[51] T. Nakano, Y. Okaie, and J.-Q. Liu, "Channel model and capacity analysis of molecular communication with Brownian motion," *IEEE Commun. Lett.*, vol. 16, no. 6, pp. 797–800, Jun. 2012, doi: 10.1109/LCOMM.2012.042312.120359.

[52] R. M. Dirks and N. A. Pierce, "Triggered amplification by hybridization chain reaction," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 43, pp. 15275–15278, 2004, doi: 10.1073/pnas.0407024101.

[53] R. Sarpeshkar, "Analog synthetic biology," *Philos. Trans. A, Math. Phys. Eng. Sci.*, vol. 372, Feb. 2014, Art. no. 20130110.

[54] R. Weiss *et al.*, "Genetic circuit building blocks for cellular computation, communications, and signal processing," *Nat. Comput.*, vol. 2, pp. 47–84, Mar. 2003.

[55] J.-T. Huang and C.-H. Lee, "On capacity bounds of two-way diffusion channel with molecule harvesting," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.

[56] H. G. Bafghi, A. Gohari, M. Mirmohseni, and M. Nasiri-Kenari, "Diffusion based molecular communication with limited molecule production rate," 2018. [Online]. Available: https://arxiv.org/abs/1802.08965

[57] A. B. Bilgin and O. B. Akan, "A fast algorithm for analysis of molecular communication in artificial synapse," *IEEE Trans. Nanobiosci.*, vol. 16, no. 6, pp. 408–417, Sep. 2017.

[58] F. Asghari, M. Samiei, K. Adibkia, A. Akbarzadeh, and S. Davaran, "Biodegradable and biocompatible polymers for tissue engineering application: A review," *Artif. Cells Nanomed. Biotechnol.*, vol. 45, no. 2, pp. 185–192, 2017.

[59] Y. Sun, M. Ito, and K. Sezaki, "Channel capacity analysis of diffusive DNA based molecular communication," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Marrakesh, Morocco, 2019, pp. 1–6.

[60] Y. Chen, Y. Li, L. Lin, and H. Yan, "Parameter estimation of diffusive molecular communication with drift," *IEEE Access*, vol. 8, pp. 142704–142713, 2020, doi: 10.1109/ACCESS.2020.3014094.

[61] T. Nakano *et al.*, "Performance evaluation of leader–follower-based mobile molecular communication networks for target detection applications," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 663–676, Feb. 2017.

[62] W.-K. Hsu, X. Lin, and M. R. Bell, "Deep-target delivery of nanosensors with bacteria-inspired coordination," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017, pp. 1–6.

[63] X. Zhang, C. Chen, J. Wu, and H. Ju, "Bubble-propelled jellyfish-like micromotors for DNA sensing," *ACS Appl. Mater. Interfaces*, vol. 11, pp. 13581–13588, Mar. 2019. [Online]. Available: https://doi.org/10.1021/acsami.9b00605

[64] D. Esteban-Fernandez *et al.*, "Single cell real-time miRNAs sensing based on nanomotors," *ACS Nano*, vol. 9, pp. 6756–6764, Jun. 2015. [Online]. Available: https://doi.org/10.1021/acsna no.5b028 07

[65] A. Currin *et al.*, "Computing exponentially faster: Implementing a non-deterministic universal Turing machine using DNA," *J. Roy. Soc. Interface*, vol. 14, no. 128, 2017, Art. no. 20160990.

[66] I. F. Akyildiz, M. Pierobon, S. Balasubramaniam, and Y. Koucheryavy, "The Internet of bio-nano things," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 32–40, Mar. 2015, doi: 10.1109/MCOM.2015.7060516.

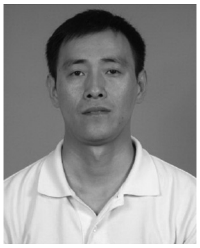
[67] J. M. Jornet and I. F. Akyildiz, "The Internet of multimedia nano-things in the Terahertz band," in *Proc. Eur. Wireless 18th Eur. Wireless Conf.*, Poznan, Poland, 2012, pp. 1–8.

[68] A. M. Turing, "On computable numbers, with an application to the Entscheidungsproblem," *Proc. London Math. Soc.*, vol. s2-42, no. 1, pp. 230–265, 1937.

[69] J. von Neumann, "First draft of a report on the EDVAC," *IEEE Ann. Hist. Comput.*, vol. 15, no. 4, pp. 27–75, 1993, doi: 10.1109/85.238389.

[70] M. Amos and A. Goñi-Moreno, "Cellular Computing and Synthetic Biology," in *Computational Matter* (Natural Computing Series), S. Stepney, S. Rasmussen, and M. Amos, Eds. Cham, Switzerland: Springer, 2018, ch. 7, pp. 93–110, doi: 10.1007/978-3-319-65826-1_7

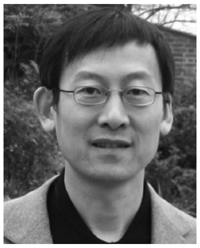
[71] L. Grozinger *et al.*, "Pathways to cellular supremacy in biocomputing," *Nat. Commun.*, vol. 10, p. 5250, Nov. 2019.



Qiang Liu (Member, IEEE) received the B.S., M.S., and Ph.D. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, 1996, 2000, and 2012, respectively.

He has worked with the School of Communication and Information Engineering, UESTC, where he is an Associate Professor. He had worked with the University of Essex, Colchester, U.K., from December 2012 to December 2013 and the University of California at Davis, Davis, CA, USA, from August 2017 to August 2018, as a visitor scholar. His researches focus on wireless sensor networks, Internet of Things, broadband wireless networks, and molecular communication.

Dr. Liu is the member of IEEE ComSoc. He is the Reviewer of IEEE TRANSACTIONS ON NANOBIOSCIENCE and *International Journal of Communication Systems*.



Kun Yang (Senior Member, IEEE) received the B.Eng. and M.Sc. degrees from Jilin University, Changchun, China, in 1994 and 1991, respectively, and the Ph.D. degree from the Department of Electronic and Electrical Engineering, University College London (UCL), London, U.K., in 2006.

He worked with UCL on several European Union research projects for several years. In 2003, he joined the University of Essex, Colchester, U.K., where he is currently a Chair Professor with the School of Computer Science and Electronic Engineering, leading the Network Convergence Laboratory. He is also an Affiliated Professor with UESTC, Chengdu, China. He has published over 150 journal papers and filed 10 patents. 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 U.K. EPSRC, EU FP7/H2020, and industries.

Dr. Yang serves on the editorial boards of both IEEE and non-IEEE journals. He has been a Fellow of IET since 2009.

Jialin Xie received the B.S. degree (Highest Hons.) in communication engineering from Sichuan University, Chengdu, 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, Chengdu.

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, Oxford, U.K., in 2019.

Mr. Xie graduated with Honor Awards from Sichuan Province and Sichuan University in 2019.

Yue Sun (Member, IEEE) received the B.Sc. degree in telecommunication and information system from the University of Electronic Science and Technology of China, Chengdu, China, in 2012, and the M.Sc. and Ph.D. degrees from the University of Essex, Colchester, U.K., in 2013 and 2018, respectively.

He is currently a Lecturer with the School of Information and Technology (Cyber Security), Chengdu University of Technology, Chengdu. His current research interests include molecular communication, DNA computing, and drug delivery systems.