This article has been accepted for publication in IEEE Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2023.3309312

Novel Interleaved Code for High-throughput Parallel DNA-based Molecular Communications

Qingwen Wang, Yue Sun*, *Member, IEEE,* Wanli Cheng, Yifan Chen, *Senior Member, IEEE,* and Kun Yang, *Fellow, IEEE*

***Abstract*—DNA-based molecular communication (DNA-MC) is a biological communication mechanism that uses DNA strands as information carriers. The longevity, stability, high information density, massive parallelism, and biological compatibility of DNA offer a dramatic potential for DNA-based storage, computing, and communication. This paper extends our previous work, which used the directional and controllable molecular hopper along the track to replace the slow and random diffusion mechanisms. This paper proposes a multiple-track-hopper parallel communication mechanism to achieve high throughput by parallel transmission and sequencing. We recommend utilizing interleaved coding to mitigate the bit error rate (BER) caused by the back-stepping motion, resulting in successive symmetric errors. Additionally, we have explored the proper interleaving depth necessary to preserve the diminished DNA information density that results from the redundancy for error correction. Simulations show that interleaved coding efficiently reduces BER in parallel DNA-MC while requiring less redundancy. This paper demonstrates the feasibility and potential of high-throughput and low-error DNA-MC, which could enable novel interdisciplinary advances between DNA communication, nanotechnology, and synthetic biology.**

***Index Terms*—DNA-based Molecular Communication, High-throughput Parallel Communication, Interleaved code, Deduplication algorithm, Molecular hopper.**

I. Introduction

MOlecular communication (MC) has gained popularity recently due to its biocompatibility, energy efficiency, and bio-environmental friendliness [1]. DNA, composed of four bases (adenine, cytosine, guanine, and thymine) that follow the Watson-Crick complementary pairing law [2], offers extra-high information density [3], $\sim 4.5 \times 10^7 GB/g$, and longevity in potential information carriers [4]. DNA synthesis and sequencing techniques have advanced with synthetic biology and organic chemistry, which has led to DNA writing with encoded oligonucleotide chains and DNA reading with nanopore sequencing [5]. We believe that DNA-based MC (DNA-MC) will become a promising possibility.

*Corresponding author: Yue Sun

Qingwen Wang and Wanli Cheng are with the School of Mechanical and Electrical Engineering, Chengdu University of Technology, Chengdu 610059, China (email: wangqw818@126.com; cwl1716@163.com).

Yue Sun is with the School of Mechanical and Electrical Engineering, Chengdu University of Technology, Chengdu 610059, China, and also with the School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610000, China (email: sunyuestc90@126.com).

Yifan Chen is with the School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610000, China (email:yifan.chen@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 (email: kunyang@essex.ac.uk).

The past works of DNA-MC primarily focus on diffusion. Bilgin [6] proposed the initial diffusive DNA-MC using the nanopore as the receiver and DNA strands of varying lengths as information carriers, minimizing inter-symbol interference by adjusting the threshold level of DNA length. Yao [7] investigated the order of arrival of DNA molecules in a diffusive MC, demonstrating that diffusive DNA-MC has a higher channel capacity over the rest of MC in diffusion due to the high information density of DNA. However, diffusive DNA-MC still suffers from slow transmission speed and high latency due to the slow stochastic nature of diffusion. However, it should be noted that while molecular motor-based MC powered by chemical fuels may increase transmission rates, it is prone to off-track behaviour [8]. Molecular track-hopper is a new approach that overcomes the limitations of the above methods. It has the desired characteristics of a moving molecule, including processivity, no chemical fuel requirement, directional motion, and external control by reversing the applied potential with step-wise motion [9]. Furthermore, [10] has shown the potential of precise and directional DNA sequencing through a nanopore using the track-hopper mechanism, with a step size of about 0.70 nm, corresponding to two DNA bases. Motivated by molecular motion control mechanisms advances, we present a robust DNA-MC system with a single track-hopper, which offers higher capacity and lower latency than diffusive DNA-MC.

A major challenge of DNA-MC is scaling the throughput and the cost of DNA reading/writing by several orders of magnitude beyond currently required by the life sciences industry [11], [12]. For example, the requirement transmission rate for real-time sequence reaches 3GB/s [13]; however, it is far beyond the single-channel diffusive DNA-MC, 6b/s [6]. Thus, it is important to investigate the high-throughput DNA-MC in parallel. In addition, the DNA reading/writing processes are relatively slow, with state-of-art strand displacement reactions in minutes [14]; however, the DNA-MC system still provides the potential of massively parallel communication [11], requiring parallel ratcheting of multiple DNA strands at once [9]. Nguyen et al. [15] parallelized over millions of nanoelectrode wells within one μm^2, successfully coding and decoding a message in DNA, which could reach the maximum data rate of megabytes per second. Adam [16] reported that hybrid nanopores could create a platform for wafer-scale nanopore arrays for the reliable controlling of the passage of multiple DNA strands [17], namely the parallel DNA sequencing. Therefore, we exploit the possibility of high-throughput DNA-MC using parallel tracks.

This paper presents a novel DNA-MC system with eight parallel tracks that enable high-rate transmission. The system utilizes hybrid nanopores managed by hoppers carrying DNA

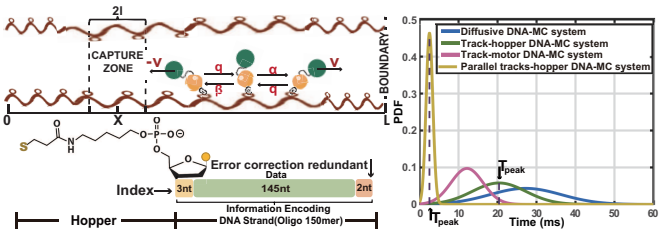
This article has been accepted for publication in IEEE Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2023.3309312

strands on cysteine footholds. An external electric field drives the DNA strands through the nanopores, and the residual current reveals the hopper's position and the DNA length. A sophisticated algorithm translates the current signals into DNA bases in real-time. Additional inter-channel interference (ICI) was observed due to the high density of parallel tracks, which resulted in the molecular hopper being captured by nearby tracks. In addition, we employ interleaved coding and deduplication techniques to reduce error rates caused by back-stepping motion and eliminate symmetric regions in the DNA sequences.

The rest of this paper is organized as follows: Section II presents a DNA molecular communication model based on parallel multiple protein tracks and identifies the potential challenges in the parallel track mechanisms. In Section III, we compare different interleaving methods and suggest using an interleaved coding approach to decrease the bit error rate (BER) and enhance the efficiency of the deduplication algorithm. Section IV evaluates the performances of interleaved coding by numerical simulations. Section V concludes and suggests future directions.

II. Illustration of Parallel DNA-MC

A. DNA-MC with single track-hopper



(a) (b)

Fig. 1. (a) Chemical structure of molecular track-hopper and designed coding scheme of DNA strands (b) Distribution of translocation times for DNA strands with 1200nt.

We have improved our understanding of how a molecular hopper moves along the protein track by refining our previous propagation mechanism. The kinetics of the molecular hopper can be described as a Markov process with three states: forward state with velocity v, stationery state, and backward state with velocity $-v$. As indicated in Fig. 1(a), the molecular hopper could switch at each foothold randomly with transition rates α, β, and q. From the Markov process model of molecular hoppers, we can derive the master equations of the hoppers and develop these to a first-order Taylor approximation concerning time-step dt [8]. The resulting following master equations govern the evolution of the probability densities $h_s(x,t)$ of the molecular hopper for the states $s \in \{-,0,+\}$:

$$\frac{\partial h_-(x,t)}{\partial t} = v\frac{\partial h_-(x,t)}{\partial x} - qh_-(x,t) + \beta h_0(x,t), \quad (1)$$

$$\frac{\partial h_0(x,t)}{\partial t} = qh_+(x,t) + qh_-(x,t) - \{\alpha+\beta+K(x)\}h_0(x,t), \quad (2)$$

$$\frac{\partial h_+(x,t)}{\partial t} = -v\frac{\partial h_+(x,t)}{\partial x} - qh_+(x,t) + \alpha h_0(x,t). \quad (3)$$

Also, we assume that the molecular hopper cannot propagate beyond the boundary at L, and propose a capture zone of length $2l$ at X, in which each molecular hopper has a one-time probability of being captured by foothold on the nearby track with capture speed k. To indicate this, we employ the indicator function $K(x)$ [18], which is defined as

$$K(x) = k\mathcal{X}(x-X) = \begin{cases} k, & |x-X| < l \\ 0, & otherwise \end{cases}. \quad (4)$$

To reduce the high dimension of this model and sustain acceptable computational accuracy. Along the lines of the approaches proposed in [8], [18], the one-dimensional Fokker-Planck equation governing the probability density function $h(x,t)$ for the presence of the molecular motor at position x and time t can be found from the aforementioned master equation (1-3).

$$\frac{\partial h(x,t)}{\partial t} = -\Lambda h(x,t) - V\frac{\partial h(x,t)}{\partial x} + D\frac{\partial^2 h(x,t)}{\partial x^2}. \quad (5)$$

Where the reaction term Λ is the random force from the thiol-disulfide interchange, which is described as

$$\Lambda = \frac{qK(x)}{\gamma\alpha\beta} + \frac{q^2K(x)}{\gamma^2\alpha^2\beta^2}\left(\frac{1}{\alpha}+\frac{1}{\beta}\right), \quad (6)$$

with the normalization factor $\gamma = 1/\beta + 1/\alpha + q/\alpha\beta$.

The drift term V is the propulsion derived from the electrical potential, which is expressed as

$$V = V_0 + \frac{K(x)}{\gamma^2\alpha\beta}\left[\frac{2-(1+\gamma)V_0}{\beta} - \frac{2+(1+\gamma)V_0}{\alpha}\right], \quad (7)$$

with $V_0 = 1/\gamma \cdot (1/\beta - 1/\alpha)$.

Here, we set an average hopper velocity v of $0.02nm/\mu s$ [9] to calculate diffusion coefficient D, given as

$$D = \frac{v}{l}\left[\frac{(1-V_0)^2}{\gamma\beta q} + \frac{(1+V_0)^2}{\gamma\alpha q}\right]. \quad (8)$$

The probability density function (PDF) distribution over transmission times of DNA strands with 1200 nucleotide (nt) for the same distance is shown in Fig. 1(b), compared with the diffusive, track-hopper, and track-motor DNA-MC. Although the transmission rate in the track-hopper DNA-MC system is slower than that in the DNA-MC system with a motor, the characterization of manually setting the step size of the molecular hopper and externally controllable direction might improve sequencing accuracy. The ability to reverse the chemical ratcheting process and obtain many-fold coverage of an individual DNA strand [9] is much more critical.

In the field of high-throughput molecular communication (MC), the use of parallel track-hopper systems has proven to be a potential solution to overcome certain limitations. Specifically, when compared to single-track-hopper systems, the peak translocation time for a parallel track-hopper is approximately $2.55ms$, as opposed to $20.4ms$ for a single-track-hopper. This decrease in translocation time and increase in transmission rate compensates for the relatively slow speed of a molecular hopper. However, the synthesis of strands of DNA longer than a few hundred nucleotides remains a challenge in practice [19]. In the following subsection, we will delve deeper into the parallel track mechanism and its potential for high-throughput track-hopper MC.

This article has been accepted for publication in IEEE Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2023.3309312

B. DNA-MC with parallel tracks

Due to the challenge of the de novo synthesis of long sequences of DNA, long-strand sequences are suggested to be divided into smaller fragments of $150 \sim 200$ bases [5]. One base represents two bits of information. The DNA-MC system with a single track-hopper can only handle a maximum of 400 bits during one transmission. Furthermore, the reading and writing processes are relatively slow, which is insufficient for the high throughput requirement of DNA sequencing in the biomedical field. To address this challenge and facilitate the advancement in the fabrication of multiple hybrid nanopores in synthesis biology, this paper proposed eight parallel tracks with nanopores, which increase the transmission rates by increasing the number of parallel channels, greatly enhancing the capacity limits of the MC system. As the hopper progresses through nanopores with each step, the receiver sequences and obtains sixteen bits of information, equivalent to two bytes.

Parallel MC transmission can be viewed as the combination of eight independent tracks, whereas sequential order matters on different tracks and determines the output. The receiver should store the DNA sequences according to the correct sequential order. Then, the decoder in the receiver could output the decoded information. The illustration of the structure and the eight cascaded processes of the parallel DNA-MC system associated with interleaved coding are shown in Fig. 2.

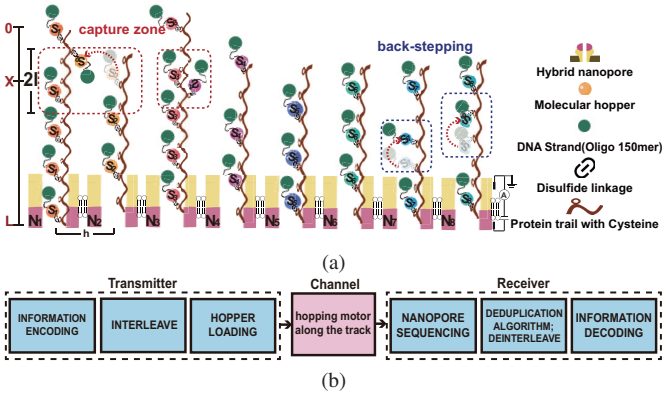


Fig. 2. (a) Molecular hoppers move along protein tracks through continuous disulfide exchange reactions, carrying the DNA strands through hybrid nanopores with external electrical potential added at both ends, meanwhile, the back-stepping of motions causes continuous symmetric errors and inter-channel interference in the system. (b) The Parallel DNA-MC system with Interleaved coding.

C. The errors in parallel DNA-MC

High-throughput MC systems are more likely to become a reality due to the characteristics of external controllable direction, higher transmission speed, and processive in track-hopper systems. There are interferences and sequence errors in parallel DNA-MC. As shown in Fig. 2(a), the interference caused by high-density protein tracks capturing other molecular hoppers at speed k, resulting in ICI, leading the receiver to decode DNA sequences from the other tracks incorrectly. The sequence errors include the back-stepping motion causing undesired symmetric errors and an additional 2% error in nanopore sequence [20]. This paper assumes that the molecular hopper's back-stepping error occurs within continuous times N. Both sequencing and back-stepping motion determine the combined error rates of DNA-MC. The sequencing error rate can be calculated by multiplying the probability of each bit being sent with the probability of each bit being misidentified as other bits [21]. The back-stepping error rate can be expressed as the ratio of the current number of error bits after deinterleaving and the 2nt error correction to the total number of bits. The sets of BER generated by the parallel tracks are averaged to obtain the average BER.

In order to improve the transmission rate and reduce BER, [22] and [23] have implemented new modulation methods or incorporated error-correcting mechanisms that use additional bits. However, these approaches reduce the information density of DNA sequences [24]. Due to the high cost of synthesizing DNA [2], this paper introduces an interleaved coding scheme to eliminate burst redundancy induced by sequencing errors and back-stepping of the molecular hopper. While reducing the BER, interleaved coding maintains the information density. In dealing with ICI, a 3nt index structure is designed to store corresponding nanopore numbers at the receiver in the first three bits and transfer times of DNA sequences in the last three. The purpose of the 3nt index is to accurately organize the original DNA sequences at the receiver, which aligns with the concept of synchronicity. As a result, a separate synchronization mechanism is not required for a parallel DNA-MC system.

III. Interleaved Coding

A. The selection of interleaver

In DNA-MC, the back-stepping motion of the molecular hopper typically introduces $2N$ burst errors above the error correction capability of 2nt we set. As a solution, interleaved coding is introduced to convert burst errors into random-like errors. The interleaving operation $\pi(i)$ and the deinterleaving operation $\pi^{-1}(i)$ of row interleaver can be generally expressed as

$$\pi(i) = \lfloor \frac{i}{d} \rfloor + (i \ mod \ d) \cdot \left\lceil \frac{mn}{d} \right\rceil, \tag{9}$$

$$\pi^{-1}(i) = \left(i \ mod \ \left\lceil \frac{mn}{d} \right\rceil \right) \cdot d + \lfloor \frac{i}{\lceil \frac{mn}{d} \rceil} \rfloor, \tag{10}$$

where $i \in [0, mn-1]$ is the index of the input data, mn is the length of input data, where m is the number of parallel tracks, n is the length of a DNA sequence, and d denotes the interleaved depth. The interleaving operations of the $m \times n$ block interleaver can be described as column-wise reading $\pi_c(i,j)$:

$$\pi_c(i,j) = \left((\pi(i_c)+1) \ mod \ n, \left\lceil \frac{\pi(i_c)+1}{m} \right\rceil \right), \tag{11}$$

here, $i \in [1,m]$, and $j \in [1,n]$ are the index of the input data block. With $i_c = m(j-1)+i-1$, and $\pi(\cdot)$ follows the mapping rules defined in Eq. (9). For no-interleaver, $\pi(i) = i$.

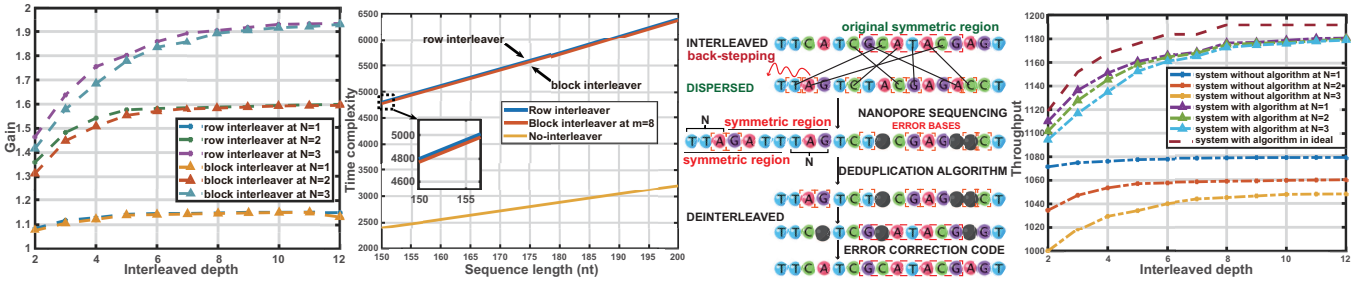
The interleaved gain G is defined as [25]

$$G = 10 \times log \frac{N_c \ with \ interleaver}{N_c \ without \ interleaver}, \tag{12}$$

where N_c are the number of correct sequencing bases.

Fig. 3(a) presents a similar increase in the gain G of row and block interleaver at different N with successive increases of the interleaved depth d. However, as d exceeds eight, the increase in gain is insignificant. Fig. 3(b) compares the

This article has been accepted for publication in IEEE Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2023.3309312



(a) (b) (c) (d)

Fig. 3. (a) The gain of parallel DNA-MC system with different interleavers. (b) The comparison of time complexity. (c) The illustration of interleaved coding. (d) The throughput of parallel DNA-MC system.

time complexity of three types of interleaving methods. These are row interleaving, block interleaving, and no-interleaving, expressed as $O(4mn)$, $O(4mn-4m+4)$ [26], and $O(2mn)$, respectively. Although the no-interleaved has the lowest time complexity, it is less effective at correcting burst errors caused by the back-stepping motion of molecular hoppers. The row interleaver, compared to the 8×150 block interleaver with row-wise reading, has less significant drawbacks in terms of time complexity. Further, the row interleaver has more suitable for complex channel environments in parallel DNA-MC systems due to its flexible interleaving rules for specific channel error correction coding. Therefore, we adopted the row interleaver and proposed an optimized deduplication algorithm based on the analysis of interleaved coding to correct burst and sequencing errors in the following subsection.

B. The principle and performance of interleaved coding

Taking 18.5kB data as an example, the coded information is equally divided into the length of 290 bits of information, which will be added with the 3nt corresponding address and the 2nt error correction redundant. The interleaving operation arranges the coded information row-wise using interleaving depth d and outputs the result column-wise, generating a new 150nt DNA strand. During nanopore sequencing, the molecular hoppers have a 12% probability of backward motion [9], leading to yield erroneous bases with two partially overlapping symmetric bases of the length $2N+1$ feature. The interleaved encoding will disperse the original continuous symmetric region to avoid incorrect deduplication, followed by error removal and sequence sorting. The contiguous error bits will be dispersed and decoded by the error correction code to complete the information recovery as shown in Fig. 3(c).

The pseudo-code for the proposed interleaved encoding algorithm is shown on the right. The ***Reshape***(X_n, m, n) function is the key statement used to implement interleaved coding: input in m column by column and output in n row by row.

C. The analysis of throughput

The derivation of the throughput T defined as $V_{code} \cdot (1 - \textit{average BER})$, in which V_{code} is the transmission rate of DNA strands. As shown in Fig. 3(d), the difference between the throughput of the system with and without algorithm groups is significant, which demonstrates the higher throughput performance with the interleaved coding. Compared with

Algorithm 1 The Interleaved Coding Scheme

Input: DNA sequence $X_n\,(m)$; maximum continuous back-stepping times N; interleaving depth d.
Output: Possible sequences without redundancy Y.

$S_{1+3N} = \{x_i, x_{i+1}, \cdots, x_{i+2N}, x_{i+2N+1}, \cdots, x_{i+3N}\}$
while $M \leftarrow N$ **do**
  for $S_{1+3N}\,(i)$ in X_n **do**
    if $S_{1+3N}\,(i) \equiv S_{1+3N}\,(i+2N)$ **then**
      if $S_{1+3N}\,(i+2N) \equiv S_{1+3N}\,(i+3N)$ **then**
        Remove bases $\{x_{i+N+1}, ., x_{i+3N}\}$
      end if
    end if
  end for
  $M \leftarrow M-1$
end while
Add $X_n(m)$ to Y; Update X_n, S_{1+3N}.
$Y = \{X_n\,(1), X_n\,(2), \cdots, X_n\,(m)\}$
Step 2: Deinterleaved coding
Reshape ($(Y$, the length of a DNA sequence n, $d)$,1, $8 \times d)$

the throughput in the ideal case, the difference could be argued that the algorithm misjudges and removes original symmetric bases, leading to an increase in the symbol error rate. It is worth noting that the greater the interleaving depth improves resistance to burst errors, resulting in longer processing times and transmission delays, potentially overprotecting the system. Thus, we analyze the interleaved coding with the depth from 3 to 8.

IV. Simulation and Analysis

This paper simulates the average BER value over 100 times to verify the superior performance of the interleaved coding system to the benchmarks of previous works. The simulation parameters are given in Table 1.

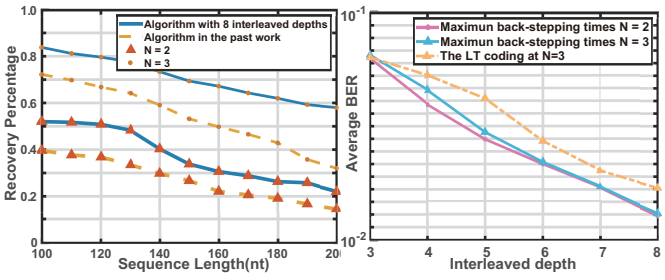
TABLE I
Parameters Used For Simulation

Parameters	Values
Link Length L	$1.0\,\mu$m
Interfoothold Distances $\boldsymbol{\Delta d}$	0.34 nm
Hopper Speed v	$35.0\,\mu s^{-1}$
Forward State $\boldsymbol{\alpha}$	$30.8\,\mu s^{-1}$
Backward State $\boldsymbol{\beta}$	$3.2\,\mu s^{-1}$
Stationery State q	$1.0\,\mu s^{-1}$
Sequencing Error Rate	2.0 %
Continuous Back − stepping Times N	3.0

This article has been accepted for publication in IEEE Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2023.3309312

Fig. 4(a) compares the performance of the deduplication algorithm at eight interleaved depths with benchmarks. The most prominent finding is that interleaved successful recovery curve is higher than without interleaved coding. In addition, success recovery refers to the probability of restoring the received sequence to the original sequence. The curves are affected by maximum continuous back-stepping times N and decline significantly as sequence length increases. The additional sequencing errors that appear in the symmetric region resulted in the deduplication algorithm being inapplicable and damaging specific statistical characteristics. It is likely that the appearance probability for symmetric regions with five bases after interleaving will be higher than those with seven bases after three back-stepping times BER.

Subsequently, the average BER performance, affected by the interleaved depths comparison of the proposed algorithm and LT coding scheme in different maximum continuous back-stepping times, is shown in Fig. 4(b). In BER simulation, continuous maximum back-stepping times are assumed to be three. The averaged BER decreases as the interleaved depth increases, revealing the fact that high d causes sequence errors to be widely dispersed. The curve of LT-coded BER is higher than curve of proposed coding BER, where deviation can be attributed to randomness in LT coding and decoding processing. It is worth noting that the increased interleaving depth, however, requires a higher computing capacity and longer processing times. Future research should investigate resource allocation balancing with computing complexity and processing delay.



(a) (b)

Fig. 4. Scheme of the parallel DNA-MC system with interleaved coding.

V. Conclusion

We have proposed an interleaved DNA-MC system with multiple parallel tracks. Under ideal conditions, the data transmission rate of the system is eight times faster than that of a track-hopper MC system. Moreover, this paper has demonstrated that interleaved coding efficiently reduces the BER of parallel nanopore sequencing while requiring fewer redundancy bits, and optimized the deduplication algorithm by dispersing the original symmetric regions. It is worth noting that the model based on interleaving coding has a time delay, which is not considered here, so it is necessary further to balance the cost of DNA synthesis and time delay. Limited by the length of letter, the packet loss rate and the impact of inter-channel interference on the data rate, the addressing and sorting algorithm, mature interleavers, and state-of-art coding benchmarks, such as a fountain, RS, Cyclone code, and LT coding, will be considered in future work.

References

[1] T. Nakano, M. J. Moore, F. Wei, A. V. Vasilakos, and J. Shuai, "Molecular communication and networking: Opportunities and challenges," *IEEE Transactions on Nanobioscience*, vol. 11, no. 2, pp. 135-148, Jun. 2012.

[2] Q. Liu, K. Yang, J. Xie, and Y. Sun, "DNA-based molecular computing, storage, and communications," *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 897-915, Jan. 2022.

[3] Y. Sun, Y. Xiao, Z. Gong, X. Zhong, and Y. Chen, "Novel sequential DNA circuits using DNAzymes," *IEEE Transactions on Nanotechnology*, vol. 21, pp. 613–619, Sep. 2022.

[4] L. Organick *et al.*, "Probing the physical limits of reliable DNA data retrieval," *Nature Communications*, vol. 11, no. 1, pp. 1–7, Jan. 2020.

[5] A. Doricchi *et al.*, "Emerging approaches to DNA data storage: Challenges and prospects," *ACS Nano*, vol. 16, no. 11, pp. 17552–17571, Oct. 2022.

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

[7] Y. Sun, M. Ito, and K. Sezaki, "Channel capacity analysis of diffusive DNA based molecular communication," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, Marrakesh, Morocco, Apr. 2019, pp. 1-6.

[8] Y. Chahibi, I. F. Akyildiz, and I. Balasingham, "Propagation modeling and analysis of molecular motors in molecular communication," *IEEE Transactions on Nanobioscience*, vol. 15, no. 8, pp. 917-927, Dec. 2016.

[9] Y. Qing, S. A. Ionescu, G. S. Pulcu, and H. Bayley, "Directional control of a processive molecular hopper," *Science*, vol. 361, no. 6405, pp. 908–912, Aug. 2018.

[10] Y. Qing, H. Bayley, "Enzymeless DNA base identification by chemical stepping in a nanopore," *Journal of the American Chemical Society*, vol. 143, no. 43, pp. 18181-18187, Nov. 2021.

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

[12] Suyama, Yoshihisa *et al.*, "Complementary combination of multiplex high-throughput DNA sequencing for molecular phylogeny," *Ecological Research*, vol. 37, no. 1, pp. 171-181, Oct. 2021.

[13] S. Vargas-Pérez and F. Saeed, "A hybrid MPI-OpenMP strategy to speedup the compression of big next-generation sequencing datasets," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 10, pp. 2760-2769, Oct. 2017.

[14] F. Wang *et al.*, "Implementing digital computing with DNA-based switching circuits," *Nature Communications*, vol. 11, no. 1, pp. 1–8, Jan. 2020.

[15] B. H. Nguyen *et al.*, "Scaling DNA data storage with nanoscale electrode wells," Science Advances, vol. 7, no. 48, pp. 1-7 Nov. 2021.

[16] A. R. Hall, A. Scott, D. Rotem, K. K. Mehta, H. Bayley, and C. Dekker, "Hybrid pore formation by directed insertion of α-haemolysin into solid-state nanopores," *Nature Nanotechnology*, vol. 5, no. 12, pp. 874-877, Dec. 2010.

[17] Y. L. Ying *et al.*, "Nanopore-based technologies beyond DNA sequencing," *Nature Nanotechnology*, vol. 17, no. 11, pp. 1136–1146, Nov. 2022.

[18] J. M. Newby and P. C. Bressloff, "Directed intermittent search for a hidden target on a dendritic tree," *Physical review. E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 80, no. 2, pp. 24-29, Aug. 2009.

[19] R. Heckel, G. Mikutis, and R. N. Grass, "A characterization of the DNA data storage channel," *Scientific Reports*, vol. 9, no. 1, pp. 1-12, Jul. 2019.

[20] C. Delahaye and J. Nicolas, "Sequencing DNA with nanopores: Troubles and biases," *PLoS ONE*, vol. 16, no. 10, pp. e0257521, Oct. 2021.

[21] L. Deng *et al.*, "Optimized code design for constrained DNA data storage with asymmetric errors," *IEEE Access*, vol. 7, pp. 84107-84121, June. 2019.

[22] Y. Huang *et al.*, "Frequency domain analysis and equalization for molecular communication," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1952-1967, Mar. 2021.

[23] S. Bhattacharjee, M. Damrath, L. Stratmann, P. A. Hoeher and F. Dressler, "Digital communication techniques in macroscopic air-based molecular communication," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 8, no. 4, pp. 276-291, Dec. 2022.

[24] W. H. Press, J. A. Hawkins, J. M. Schaub, J. M. Schaub, and I. J. Finkelstein, "HEDGES error-correcting code for DNA storage corrects indels and allows sequence constraints," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 31, pp. 18489–18496, July. 2020.

[25] Y. Lu, M. D. Higgins, and M. S. Leeson, "Diffusion based molecular communications system enhancement using high order hamming codes," in *2014 9th International Symposium on Communication Systems, Networks & Digital Sign (CSNDSP)*, Manchester, UK, 2014, pp. 438-442.

[26] N. C. Shekhar, B. Amulya, G. Praveen and K. S. Reddy, "Bit error rate performance of hybrid interleaver in turbo codes," in *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, Dec. 2021, pp. 539-546.