

Image Encoding Using Multi-Level DNA Barcodes with Nanopore Readout

Jinbo Zhu, Niklas Ermann, Kaikai Chen, and Ulrich F. Keyser*

Deoxyribonucleic acid (DNA) nanostructure-based data encoding is an emerging information storage mode, offering rewritable, editable, and secure data storage. Herein, a DNA nanostructure-based storage method established on a solid-state nanopore sensing platform to save and encrypt a 2D grayscale image is proposed. DNA multi-way junctions of different sizes are attached to a double strand of DNA carriers, resulting in distinct levels of current blockades when passing through a glass nanopore with diameters around 14 nm. The resulting quaternary encoding doubles the capacity relative to a classical binary system. Through toehold-mediated strand displacement reactions, the DNA nanostructures can be precisely added to and removed from the DNA carrier. By encoding the image into 16 DNA carriers using the quaternary barcodes and reading them in one simultaneous measurement, the image is successfully saved, encrypted, and recovered. Avoiding any proteins or enzymatic reactions, the authors thus realize a pure DNA storage system on a nanopore platform with increased capacity and programmability.

and readout speed,^[11–14] but it is almost helpless for DNA nanostructure-based storage. Even if the sequence is decoded, it is still difficult to reproduce the whole structure accurately, which indicates the higher security of this storage method from this perspective. Additionally, hardware encryption with physical keys can further improve the security of the information saved by this strategy.^[10,15]

Solid-state nanopore provides us with a versatile tool to investigate the biomolecules, such as DNA, ribonucleic acid (RNA), and proteins, at the single-molecule level.^[16,17] We recently introduced a solid-state nanopore sensing platform to read bits encoded by 56 DNA hairpins on a 7.2 kb DNA carrier.^[18] Readout of the structures required glass nanopores with diameters of around 5 nm. Such a small diameter complicates the fabrication process

and reduces the success rates compared to easily manufactured >10 nm glass nanopores. The smaller diameter also leads to more non-specific interaction and pore clogging.^[19] In further work, we used a streptavidin-labelled DNA scaffold to build a secure data storage method with binary codes.^[15] While protein labels can be used with larger diameter pores, they complicate the molecular assembly and set an upper bound on the data density through their defined size. Additionally, the number of usable protein labels is limited due to the high monovalent salt concentrations required for the readout. The presence of proteins also reduces the lifetime of solid-state nanopores and hence has a detrimental effect on the viability of the readout.

In this work, we replace protein barcodes with multi-way DNA junctions to address the shortcoming and limitations of our prior approaches. Inspired by the work of Wang and Seaman^[20] we designed a multi-level storage architecture. Building on our established DNA carrier-based rewritable storage system with 14 nm diameter nanopores,^[19] we use three DNA junction structures of different sizes (4-way junction, 6-way junction, and 12-way junction) to generate a quaternary encoding system (0–3) on the long linear DNA carrier, which increases the data density compared to classic binary encoding. Through toehold-mediated strand displacement reaction (SDR),^[21–23] the presence of nanostructures on the carrier can be precisely controlled, allowing data reading and writing. Based on this storage system, we successfully save a grayscale image into the DNA nanostructures on 16 different carriers and read out the information in the mixture. Furthermore, using SDR the image information can be easily encrypted and decrypted.

1. Introduction

Data storage in synthetic deoxyribonucleic acid (DNA) sequences has proved to be a promising means due to DNA high storage density, long-term durability, and low energy cost compared to traditional storage media,^[1–4] however it suffers from slow readout, high synthesis cost, and difficulty in rewriting. As a natural molecule, besides carrying genetic information, DNA is an ideal programmable molecule for the construction of nanostructures from 1D to 3D.^[5–7] Storage based on DNA nanostructures offers a solution to achieve easily rewritable and editable DNA storage.^[8–10] Data decoding of DNA nanostructure-based storage is often based on gel electrophoresis, which severely restricted the reading speed and data capacity. As an emerging technique, nanopore sequencing has been utilized to decode data saved in DNA sequences with advantages in terms of cost

Dr. J. Zhu, Dr. N. Ermann, Dr. K. Chen, Prof. U. F. Keyser
Cavendish Laboratory
University of Cambridge
JJ Thompson Avenue, Cambridge CB3 0HE, UK
E-mail: ufk20@cam.ac.uk

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/smll.202100711>.

© 2021 The Authors. Small published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/smll.202100711

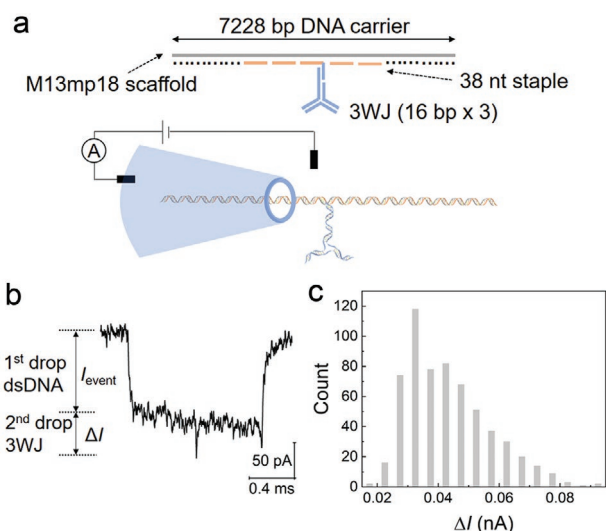


Figure 1. Detection of single DNA three-way junction through a conical quartz nanopore. a) Schematic of DNA carrier 1 consisting of a M13mp18 scaffold, 190 staples, and a three-way DNA junction bound in the middle. The translocation of the carrier through a pore is shown below. b) Current signal of carrier 1. c) Distribution of second current drop ΔI in the middle of 622 unfolded events. The average ΔI is 0.039 ± 0.001 nA and dsDNA current drop is 0.147 nA.

2. Results and Discussion

We initially investigated the design of DNA structures detectable on the ≈ 14 nm nanopore sensing platform.^[19] As shown in Figure 1a, the backbone of the DNA carrier is comprised of a long M13mp18 scaffold (7228 nt) and 190 short synthetic DNA

staples (38 nt).^[19] When the DNA carrier is driven through the nanopore under an electric force in an unfolded configuration, a first level current drop (I_{event}) is recorded (Figure 1b). In our previous work, we found that a single short ssDNA or dsDNA protruding out from the carrier is insufficient to induce an obvious second level signal with the ≈ 14 nm diameter nanopore,^[24] so we placed a DNA three-way junction (3WJ) in the center to create carrier 1 (Figure 1a, Figure S1a, Supporting Information).^[25] The 3WJ structure on carrier 1 is indeed detectable. About 84% of events showed a clear peak in the middle and gave a second current drop (ΔI) centered at 0.039 ± 0.001 nA upon a backbone dsDNA level of 0.147 nA (Figure 1c). The average ΔI and error are obtained by fitting a Gaussian function to the data in Figure 1c.

In order to obtain a stable signal and generate more current levels, we increased the arm number and length of the junction structures to build a multi-level system.^[20,26] We first prepared three different DNA multi-way junctions, 4WJ, 6WJ, and 12WJ (Figure S3, Supporting Information), then linked them onto carrier 2 (Figure 2a, Figure S1b, Supporting Information). A reference DNA structure composed of 11 DNA dumbbells was designed at one end to distinguish the translocation direction and determine the relative presence time (Δt). In the example event (Figure 2b), four downward peaks were clearly observed, in which the first one is from the dumbbells, and the following three with average ΔI at 0.041, 0.087, and 0.176 nA are from 4WJ, 6WJ, and 12WJ, respectively (dsDNA current drop I_{event} is 0.140 nA). The second current drop and presence time were normalized ($\Delta I/I_{\text{event}}$ and $\Delta t/t_{\text{event}}$) to eliminate the influence of different pore sizes. Based on the analysis of 84 unfolded translocation events (Table S2, Supporting Information), a plot showing the relationship between ΔI and Δt was

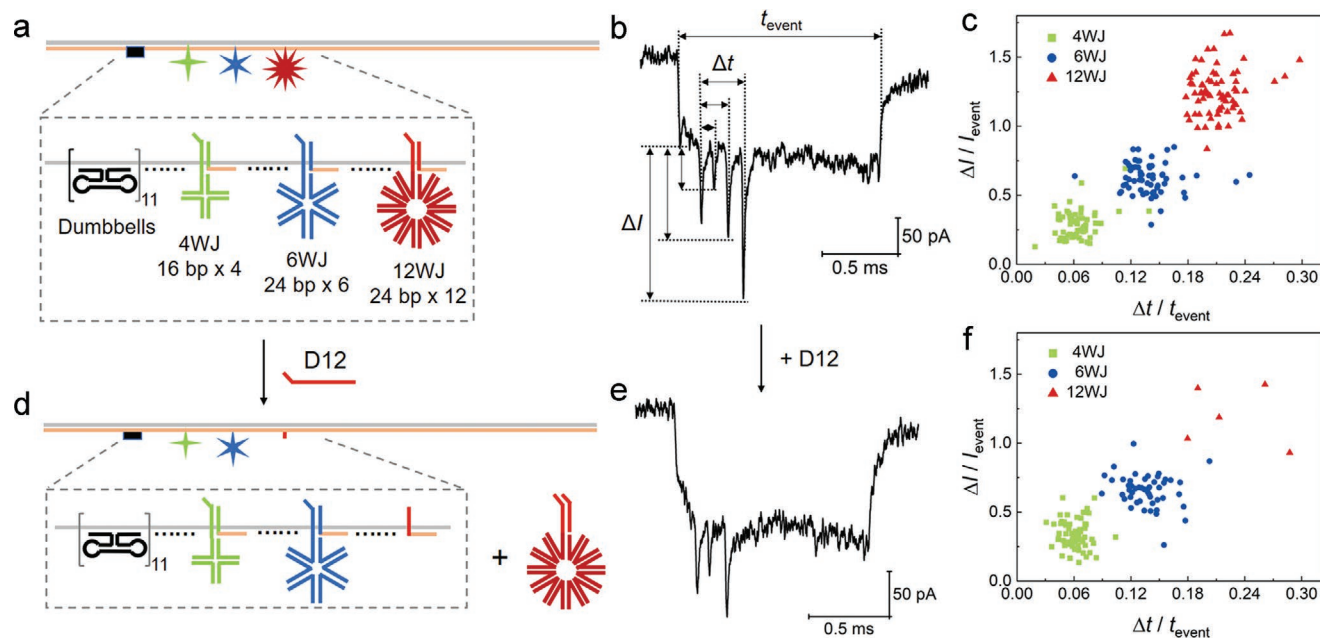


Figure 2. Detection of different DNA multiway junctions through conical quartz nanopores and modification by toehold-mediated strand displacement reactions. a,d) Schematic of carrier 2 with or without 12WJ-deletion strand D12. The detailed design and more sample events can be found in Figure S1, Supporting Information. b,e) Current signal of carrier 2 with or without D12. c,f) Scatter diagram of ΔI and Δt of 4WJ, 6WJ and 12WJ on carrier 2 with (c) or without (f) D12. The same number (84) of unfolded events were analyzed for both cases.

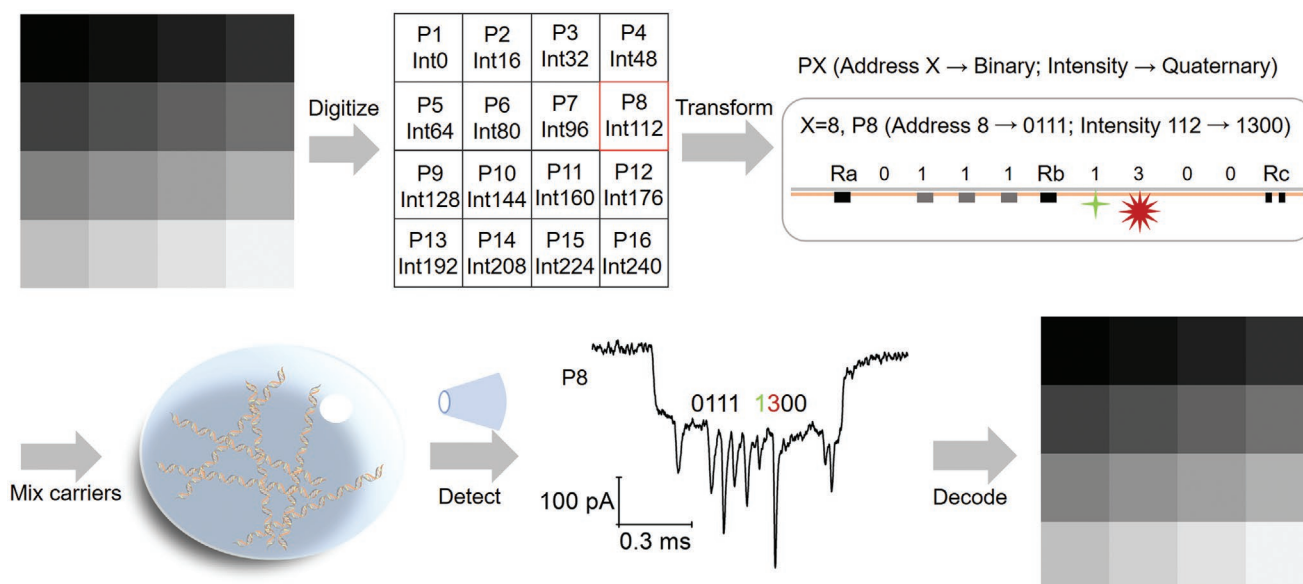


Figure 3. Storage of a 4×4 grayscale image using 16 carriers with different DNA nanostructures. Each pixel of the image is saved through address and intensity information, which are transformed into binary and quaternary codes and represented by DNA dumbbells and multiway junctions, respectively. The design for Pixel 8 (highlighted by the red box) is shown as an example here. The mixture of P1–P16 is analyzed with a glass nanopore and a typical event for P8 is shown. Measuring 20 unfolded events of P1–P16 allows for error-free decoding of the images. Details of the design and more example events can be found in Figures S5 and S6, Supporting Information, respectively.

obtained (Figure 2c). At the three designed areas, three levels of ΔI were observed as expected with occupied fractions above 70% for each binding site (Figure S2, Supporting Information). Therefore, we can use 4WJ, 6WJ, and 12 WJ to represent three distinct digits. Based on the normalized average $\Delta I/I_{\text{event}}$ of the three DNA structures (0.29 for 4WJ, 0.62 for 6WJ, and 1.26 for 12WJ), two thresholds at $\Delta I/I_{\text{event}} = 0.46$ and 0.93 were set to separate them.

The next challenge is how to individually control each DNA nanostructures independently, that is, to edit the information content of the carrier at each site. This can be achieved by SDR on the designed toehold end reserved on the DNA structures. In Figure 2d, taking the 12WJ as an example, the added invading strand D12 (Table S1, Supporting Information) will bind to the toehold end on 12WJ and remove the whole structure from carrier 2. A typical example event is shown in Figure 2e, and the statistics are given in Figure 2f and Table S2, Supporting Information. As expected, the occupied fraction of 12WJ at its designed binding site has fallen to less than 10% (Figure S2, Supporting Information), which indicates that the target DNA nanostructure can be accurately displaced by the relevant invading strand. This feature can be applied to erase or encrypt data, and then further addition of any specific DNA construct will rewrite or recover the data.

Based on the above DNA nanostructures, we designed a quaternary coding system to store information. A grayscale image was stored as a proof of concept (Figure 3). The basic composition unit of a digital image is a pixel, and intensity information is carried in each pixel. As a monochrome image, the intensity of grayscale describes the brightness in each individual pixel on a scale from black (zero) to white (full). For an 8-bit grayscale image, 256 (2^8) levels of intensity can be provided to each

pixel. In the example image in Figure 3, there are 16 pixels with different intensities composing the square graph. The first step for storage is the digitization of the address and intensity information of each pixel as shown in Figure 3.

The second step of the storage system is the transformation of digital information into our DNA nanostructures. We use 16 DNA carriers to save the information of 16 pixels. There are two kinds of information for each pixel, address and intensity, so the coding area on each carrier was split into two parts (Figures S4 and S5, Supporting Information). In the address information area, 4 binding sites for DNA dumbbells (11 for each) are designed to establish a binary coding system, so 16 (2^4) codes for pixel addresses can be stored here. In the intensity information area, another 4 binding sites are reserved for the three different DNA multi-way junctions to establish a quaternary encoding system (4WJ for bit “1”, 6WJ for bit “2”, and 12WJ for bit “3”), so 256 (4^4) codes for grayscale intensity can be saved here. To distinguish the signal, three reference structures, Ra, Rb and Rc, are designed on the carrier. Eleven DNA dumbbells are used for Ra and Rb, and two groups of five dumbbells are used for Rc. As an example, the design of carrier P8 for information storage of pixel 8 (P8) is shown in Figure 3. All designs are shown in the Supporting Information.

After preparing the 16 carriers separately and mixing them into one solution, the final step will be reading and decoding. One advantage of our method is that all the reading work can be done in a simultaneous measurement, because we can distinguish the current signals of 16 carriers by their unique address codes. A typical example event of P8 is given in Figure 3 and more example events of all carriers representing the pixels can be found in Figure S6, Supporting Information. After dividing the unfolded events into 16 groups based

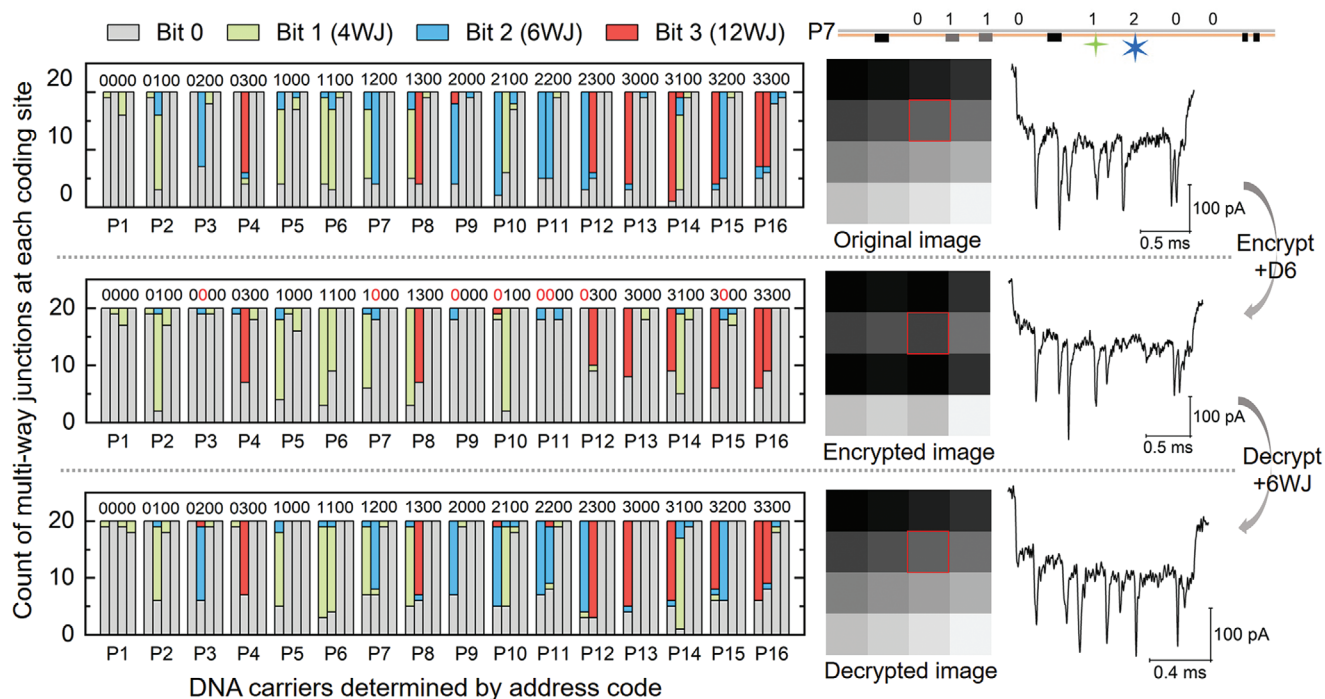


Figure 4. Nanopore results of the image storage (top row), encryption (middle row), and decryption (bottom row) using DNA nanostructures and strand displacement reactions. First, the mixture of the 16 carriers was measured by the nanopore (top row). Second, strand D6 was added into the mixture to remove the 6WJ for encryption (middle row). Finally, 6WJ was added to recover the information for decryption (bottom row). The bar graph shows the proportions of each DNA multi-way junction at each designed coding site on the 16 carriers. The unfolded events were classified by the address code into 16 groups, and then the intensity coding area of the first 20 events for each carrier was analyzed. Statistics can be found in Tables S3–S5, Supporting Information, and more example events can be found in Figure S6, Supporting Information. The decoded images are given next to the bar graph. Carrier P7, which carries the information of Pixel 7 highlighted by the red box in the image, is shown on the right as an example.

on the address codes, we subsequently use the two thresholds obtained from Figure 2c to analyze the signal and decode the intensity information. To reduce measurement time and simplify the analysis, we try to minimize the number of events needed for correct reading on each encoding site. Based on initial analysis on carrier P8 and P5, we find that a high confidence level for nanopore readout can be achieved when more than nine unfolded events are analyzed (Section S4, Supporting Information). Herein, to ensure correct readout, twenty unfolded events are analyzed for each carrier to decode the contained information. After successful identification of the secondary current drops in the translocation events, the statistics are given in the top row in **Figure 4** and Table S5, Supporting Information. The final digit depends on the DNA structure with the highest frequency of occurrence on the corresponding site. Based on the result reported from the nanopore sensing platform, we can restore the image as shown in Figure 3, which is the same as the original one.

The above storage strategy is straightforward and easy to read out, but this of course raises another question: how can the data be encrypted to ensure information security? This point can be addressed in the design of the system. On each DNA multi-way junction, a specific toehold end has been reserved for further modification. The information can be easily changed by adding invading strands as shown in Figure 2. In Figure 4, as a physical encryption key, invading strand D6 was added to the carrier

mixture to remove the 6WJ on the carriers and encrypt the data. Taking P7 as an example (right side in Figure 4), the downward peak of 6WJ is missing after the addition of D6 and the intensity code of pixel 7 changes from 1200 to 1000. As shown in the middle row in Figure 4 and statistics in Table S6, Supporting Information, the switch of “2 to 0” in the intensity code also happened for all the other pixels, hence the image decoded from the encrypted codes is distinct from the original one.

Decryption is an inverse process of the encryption in our system. 6WJ plays the role of a physical decryption key. As shown in the bottom row of Figure 4 and statistics in Table S7, Supporting Information, after the addition of 6WJ, the medium size peak of 6WJ reappears in the signal of carrier P7. This is because the single-stranded overhang for 6WJ binding is free on the carrier as in the case of 12WJ in Figure 2d, and thus adding excess 6WJ “consumes” the D6 strand in solution and reoccupies the binding site on the carrier. It is the same case for the other carriers, so the intensity codes read out after the decryption are restored to the original codes before encryption (Figure 4), recovering the image. Furthermore, we took carrier P8 as an example and conducted encryption and decryption on it using the D4/4WJ and D12/12WJ systems, respectively. The expected nanopore results for both systems are shown in Table S8, Supporting Information, which proves the encryption and decryption strategy works with different DNA invading strands and nanostructures.

3. Conclusion

In summary, we have established an information storage system based on the sensing of differently sized DNA nanostructures with solid-state nanopores. Taking a grayscale image as an example, we have shown how to design multi-way DNA junctions to save, encrypt, and decrypt complex information using quaternary barcodes on DNA carriers. Using toehold-mediated SDR, the target structure on the scaffold can be accurately removed and added. Thus, the encryption and decryption can be easily realized on a molecular level, not merely through an encryption algorithm. Even though only a simple image is securely stored here, it serves as a proof of concept for more complex encryption schemes based on more complex molecular designs in the near future. This secure data storage mode will be useful for the encoding and transmission of confidential information and due to storage in DNA shape being difficult to read by DNA sequencing. Benefiting from the single-molecule nanopore sensing platform, the speed for data reading is superior to sequencing, fluorescence, or gel electrophoresis-based readout approaches. Combined with deep learning techniques,^[27] real-time nanopore data analysis may be achievable in the near future.

Compared with our previous nanopore-based storage work,^[15,18] the method reported here is a pure DNA system using easily manufactured 14 nm nanopores. DNA multi-way junctions as signal-inducing structures offer the advantage of tunable size compared to protein tags such as streptavidin. The quaternary encoding scheme doubles the storage capacity compared to a binary system and given the signal-to-noise ratio hexadecimal and higher are attainable. Their programmability makes DNA multi-way junctions a powerful tool in the future development of nanopore storage and sensing.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors thank the support from an ERC Consolidator Grant (Designerpores No. 647144). Filip Bošković and Ran Tivony are acknowledged for experiment and suggestions on the manuscript.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Research data are not shared.

Keywords

DNA nanostructures, encryption, information storage, nanopores, strand displacement

Received: February 3, 2021

Revised: March 30, 2021

Published online:

- [1] V. Zhirnov, R. M. Zadegan, G. S. Sandhu, G. M. Church, W. L. Hughes, *Nat. Mater.* **2016**, *15*, 366.
- [2] L. Ceze, J. Nivala, K. Strauss, *Nat. Rev. Genet.* **2019**, *20*, 456.
- [3] J. Koch, S. Gantenbein, K. Masania, W. J. Stark, Y. Erlich, R. N. Grass, *Nat. Biotechnol.* **2020**, *38*, 39.
- [4] Y. Dong, F. Sun, Z. Ping, Q. Ouyang, L. Qian, *Nat. Sci. Rev.* **2020**, *7*, 1092.
- [5] N. C. Seeman, H. F. Sleiman, *Nat. Rev. Mater.* **2018**, *3*, 17068.
- [6] W. Bae, S. Kocabay, T. Liedl, *Nano Today* **2019**, *26*, 98.
- [7] Y. Zhang, J. Tu, D. Wang, H. Zhu, S. K. Maity, X. Qu, B. Bogaert, H. Pei, H. Zhang, *Adv. Mater.* **2018**, *30*, 1703658.
- [8] J. S. Shin, N. A. Pierce, *Nano Lett.* **2004**, *4*, 905.
- [9] K. Halvorsen, W. P. Wong, *PLoS One* **2012**, *7*, e44212.
- [10] A. R. Chandrasekaran, O. Levchenko, D. S. Patel, M. MacIsaac, K. Halvorsen, *Nucleic Acids Res.* **2017**, *45*, 11459.
- [11] L. Organick, S. D. Ang, Y. J. Chen, R. Lopez, S. Yekhanin, K. Makarychev, M. Z. Racz, G. Kamath, P. Gopalan, B. Nguyen, C. N. Takahashi, S. Newman, H. Y. Parker, C. Rashtchian, K. Stewart, G. Gupta, R. Carlson, J. Mulligan, D. Carmean, G. Seelig, L. Ceze, K. Strauss, *Nat. Biotechnol.* **2018**, *36*, 242.
- [12] K. Doroschak, K. Zhang, M. Queen, A. Mandyam, K. Strauss, L. Ceze, J. Nivala, *Nat. Commun.* **2020**, *11*, 5454.
- [13] S. M. H. T. Yazdi, R. Gabrys, O. Milenkovic, *Sci. Rep.* **2017**, *7*, 5011.
- [14] M. Wanunu, *Phys. Life Rev.* **2012**, *9*, 125.
- [15] K. Chen, J. Zhu, F. Bošković, U. F. Keyser, *Nano Lett.* **2020**, *20*, 3754.
- [16] C. Dekker, *Nat. Nanotechnol.* **2007**, *2*, 209.
- [17] M. Thakur, M. Macha, A. Chernev, M. Graf, M. Lihter, J. Deen, M. Tripathi, A. Kis, A. Radenovic, *Small Methods* **2020**, *4*, 2000072.
- [18] K. Chen, J. Kong, J. Zhu, N. Ermann, P. Predki, U. F. Keyser, *Nano Lett.* **2019**, *19*, 1210.
- [19] N. A. W. Bell, U. F. Keyser, *Nat. Nanotechnol.* **2016**, *11*, 645.
- [20] X. Wang, N. C. Seeman, *J. Am. Chem. Soc.* **2007**, *129*, 8169.
- [21] D. Y. Zhang, E. Winfree, *J. Am. Chem. Soc.* **2009**, *131*, 17303.
- [22] R. R. F. Machinek, T. E. Ouldrige, N. E. C. Haley, J. Bath, A. J. Turberfield, *Nat. Commun.* **2014**, *5*, 5324.
- [23] F. C. Simmel, B. Yurke, H. R. Singh, *Chem. Rev.* **2019**, *119*, 6326.
- [24] J. Kong, J. Zhu, U. F. Keyser, *Chem. Commun.* **2017**, *53*, 436.
- [25] Z. Zhu, X. Duan, Q. Li, R. Wu, Y. Wang, B. Li, *J. Am. Chem. Soc.* **2020**, *142*, 4481.
- [26] L. He, P. Karau, V. Tabard-Cossa, *Nanoscale* **2019**, *11*, 16342.
- [27] K. Miunas, N. Ermann, U. F. Keyser, *Nano Lett.* **2018**, *18*, 4040.