High-speed 3D DNA PAINT and unsupervised clustering for unlocking 3D DNA origami cryptography

Gde Bimananda Mahardika Wisna[a,e,1], **Daria Sukhareva**[b,e], **Jonathan Zhao**[c,e], **Deeksha Satyabola**[b,e], **Michael Matthies**[e], **Subhajit Roy**[a,e], **Chao Wang**[d,e], **Petr Šulc**[b,e], **Hao Yan**[b,e], **and Rizal F. Hariadi**[a,e,1]

[a]Department of Physics, Arizona State University, Tempe, Arizona, USA.; [b]School of Molecular Sciences, Arizona State University, Tempe, Arizona, USA.; [c]School of Computing and Augmented Intelligence, Arizona State University, Tempe, Arizona, USA.; [d]School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, Arizona, USA.; [e]Center for Molecular Design and Biomimetics at the Biodesign Institute, Arizona State University, Tempe, Arizona, USA.

[1]To whom correspondence should be addressed. E-mail: gwisna@asu.edu and rhariadi@asu.edu

This manuscript was compiled on July 4, 2024

Abstract

DNA origami information storage is a promising alternative to silicon-based data storage, offering a secure molecular cryptography technique that conceals information within arbitrarily folded DNA origami nanostructures. Routing, sliding, and interlacing staple strands lead to the creation of a large 700-bit key size. The realization of practical DNA data storage requires high information density, robust security, and accurate and rapid information retrieval. To meet these requirements, advanced readout techniques and large encryption key sizes are essential. In this study, we report an enhanced DNA origami cryptography protocol to encrypt information in 2D and 3D DNA origami structures, increasing the number of possible scaffold routings and increasing the encryption key size. We employed all-DNA-based steganography with fast readout through high-speed 2D and 3D DNA-PAINT super-resolution imaging, which enables higher information density. By combining 2D and 3D DNA-PAINT data with unsupervised clustering, we achieved accuracy up to 89% and high ratios of correct-to-wrong readout despite the significant flexibility in the 3D DNA origami structure shown by oxDNA simulation. Furthermore, we propose design criteria that ensure complete information retrieval for the DNA origami cryptography protocol. Our findings demonstrate that DNA-based cryptography is a highly secure and versatile solution for transmitting and storing information, making it an attractive choice for the post-silicon era.

Key words: 3D DNA origami cryptography, 3D DNA origami storage, High-speed DNA-PAINT, unsupervised clustering.

Introduction

The development of semiconductor-based transistors in the mid-20th century marked the beginning of the information revolution, enabling data storage and computing, while the advent of fiber optics and optical amplifiers necessitated secure communication systems.[1] Modern cryptography, such as the Advanced Encryption System (AES) with keys up to 256 bits, evolved to meet this need, relying heavily on semiconductor materials for processing power.[2–4] However, the limited availability of these materials, especially silicon[5], and the high energy demand for operating devices[6] present significant challenges. DNA,

with its stability, programmability, high information density, and low maintenance needs, has emerged as a promising substitute.[7–10] Pioneering work on DNA computing[11–17] and data storage[18–21] has demonstrated DNA's potential for these applications, making the development of molecular cryptography protocols crucial for secure DNA-based data storage management in the future.

Recent advancements in molecular cryptography for 2-way communication have been demonstrated through a multitude of chemical approaches, such as harnessing optical and physical properties of molecules[22–25] and exploiting Watson-Crick base pairing of DNA.[26–30] Zhang et al. developed DNA origami cryptography, which relies on the challenge of accurately predicting the arrangement and interlacing of DNA strands to achieve determined geometric shapes that function as templates for encrypting patterns.[31] This approach, similar to steganography, involves folding a long DNA strand from the M13mp18 bacteriophage with hundreds of short single-stranded staple strands, each <100 nt in length.[32–34] The central principle of modern cryptography, which employs complex mathematical problems to generate large possibilities of keys[2], can be seamlessly translated into the DNA-origami system.[28] Zhang et al. estimated that the key space can exceed 700 bits with a DNA scaffold of 7249 nucleotides, as found in the M13mp18, potentially surpassing the Advanced Encryption Standard (AES) by a factor of >2.[31]

Despite the theoretical large key space, the practicality of the current demonstration of DNA origami is limited by the inherently slow imaging of conventional atomic force microscopy (AFM)[35] coupled with the reliance on biotin-streptavidin conjugations, which frequently result in undesired aggregations.[36,37] High-speed DNA-PAINT,[38] a DNA-based super-resolution imaging technique, offers a solution to enhance readout speed and eliminate aggregation induced by protein conjugation. This method capitalizes on the stochastic binding of single-stranded DNA (ssDNA), referred to as docking strand, and short fluorophore-labeled ssDNA called imager strands to achieve resolutions as high as sub-nm[39] to 10 nm.[40] DNA-PAINT enables faster imaging and a larger imaging area of nearly 100 μm by 100 μm, allowing for the imaging of thousands of DNA origami structures per field of view.[38] Despite the application of DNA origami and DNA-PAINT techniques for alternative DNA storage, high error rates during experimental origami folding and low detection efficiency in DNA-PAINT imaging necessitate the development of error-correction post-processing algorithms.[21] Therefore, optimizing these strategies is crucial for improving information retrieval in DNA-based systems.

In this work, we report the application of high-speed DNA-PAINT and unsupervised clustering to achieve DNA origami cryptography in 2D and 3D DNA origami with fast readout while maintaining high detection efficiency of docking strands bearing pattern information. By extending the binding length to the scaffold, we achieve a detection efficiency of ∼90% using DNA-PAINT. The process allows quick analysis of thousands of DNA origami, each with unique patterns, within ∼24 min at a resolution of 10 nm. Combining high-speed DNA-PAINT readout with unsupervised K-means clustering enables fast and accurate assignment of centroids to 2D or 3D clusters and template alignment based on least squared distance minimization with pattern matching. Integration of bit redundancy improves the accuracy in the retrieval of data from 2D and 3D DNA origami encryption and decryption processes, achieving a precision of 70–89%, despite the inherent flexibility of 3D DNA origami architectures, as predicted by computational modeling of oxDNA. The 3D alignment of the 3D DNA-PAINT data to the mean structures as simulated by oxDNA gives better RMSD compared to fitting the 3D DNA PAINT data to the unrelaxed structures. Our method streamlines DNA origami cryptography, with the advantage of many possible folding paths, using fast DNA-PAINT and unsupervised clustering to achieve secure information transmission with high readout accuracy.

Results

The DNA origami cryptography protocols. The 2-way DNA origami cryptography protocol uses symmetric keys between a sender, Alice, and a receiver, Bob. Alice aims to securely transmit the message "ASU" (Fig. 1A). Using the principles of symmetric cryptography, Alice encrypts the message and generates keys and cipher text, which we refer to as a cipher mixture comprising multiple DNA strands. The encryption

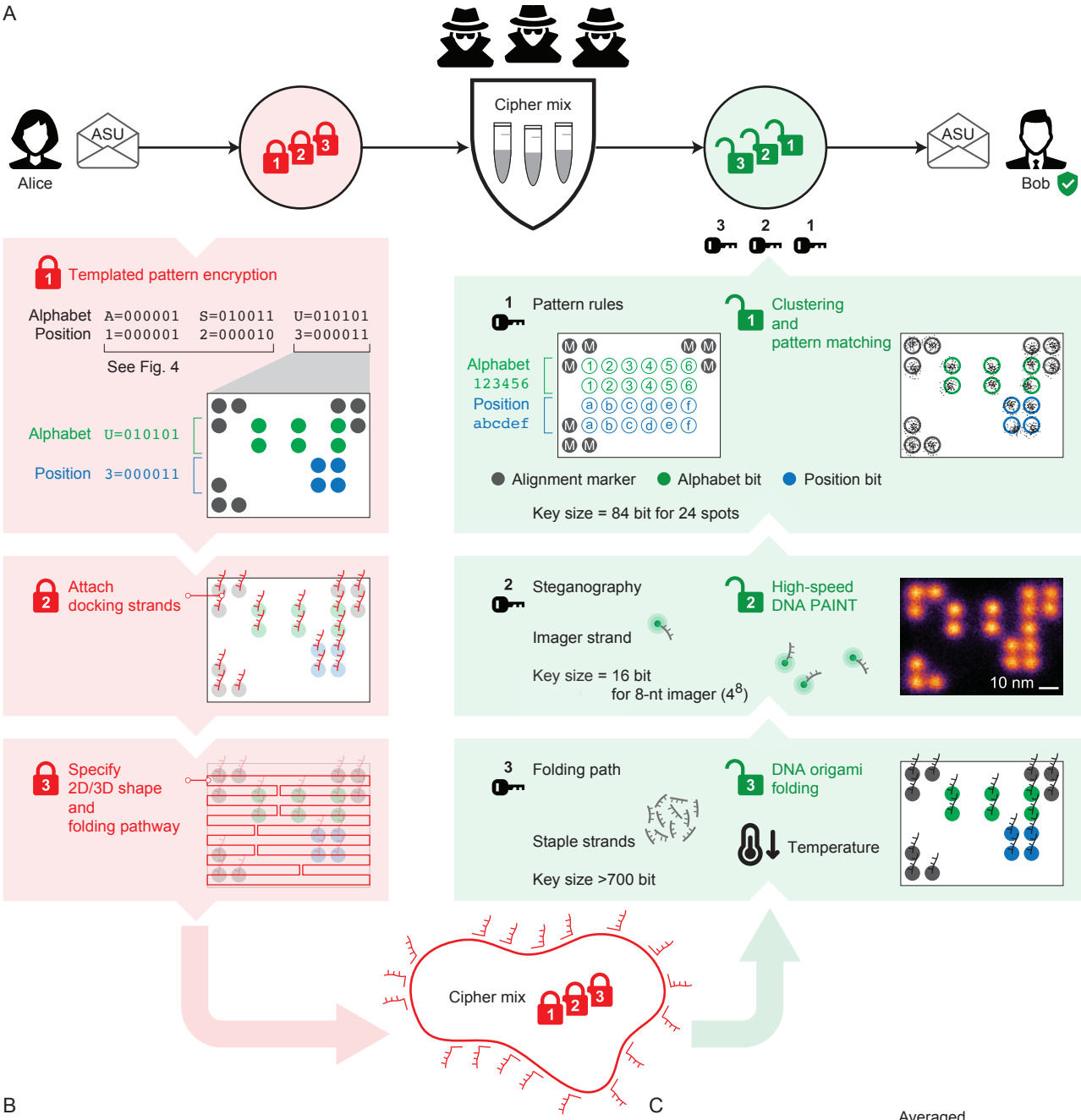
and the description processes involve 3 steps: encoding the letters and their corresponding positions as a binary pattern on DNA origami (🔒 1), docking sequence assignment (🔒 2), and specifying the shape and folding pathway of the scaffold strand (🔒 3). Alice and Bob can exchange the cipher mixtures, consisting of M13mp18 scaffolds and docking strands that encode the entire letters and positions of the message, through insecure communication channels. However, they must securely exchange the 3 keys required for decryption. To decrypt the message, Bob reverses the encryption process by first folding the DNA origami (🔓 3) using the staple strand mix (key 3), performing high-speed DNA-PAINT imaging to reveal the pattern (🔓 2) using the imager strand (key 2), and then applying clustering and template alignment to extract the letters and their corresponding positions (🔓 1) based on the pattern rules (key 1). In the absence of any of the required encryption keys, potential adversaries will be incapable of decrypting the encoded data, thereby ensuring the security of the information (Fig. 1C).

DNA origami encryption. The message encryption employs the DNA origami technique, which involves uniquely designing the routing of a long M13mp18 scaffold and staple strands for DNA origami geometries, such as a 2D Rothemund Rectangular Origami (RRO) or a 3D wireframe cuboctahedron origami (Fig. 1B; Tables S1 and S2). The encryption protocol (Fig. 1A) flows downward (red arrows) and results in the creation of 3 keys. The process begins by converting the text message (Fig. 1A) into binary codes (Table S3), which denote each letter and its position within the text (Fig. 1A). Next, a geometric shape of DNA origami is selected to serve as a template for the pattern, and a pattern rule that the template can accommodate is chosen. In our first demonstration, we select the 2D RRO with dimensions of approximately $90\times60\text{nm}^2$ as the template and devise a pattern encryption rule (Fig. 1A) that allows the arrangement of the binary code pattern, consisting of "0" and "1" bits, on the RRO template (Fig. 1C(iv)). The patterns on the DNA origami can achieve a resolution of sub-100 nm, with a separation of 10–20 nm between two bits in a pattern. The alignment markers (gray circles in Fig. 1A), are necessary to break the rotational symmetry within the plane of the Rothemund Rectangular Origami (RRO).

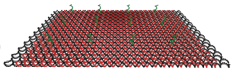
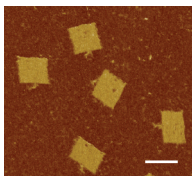
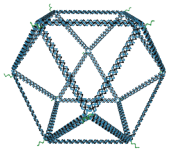
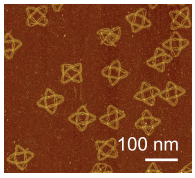
The second step of encryption is to design a unique docking strand sequence for high-speed DNA-PAINT imaging where we adopted the sequence used by Strauss et al.[38] (Fig. 1A); this step is called docking assignment. The third and final step is to generate the staple strands that will fold the M13mp18 scaffold strand into an RRO and to assign docking sequences to extend specific staple strands on the RRO DNA origami that are designated for the "1" bits in our binary codes (Fig. 1A). These extended strands are interchangeably called "information strands" or "docking strands" throughout the article. The strands on an RRO corresponding to "0s" will have no docking sequence extension and are called "staple strands key". The possibility space for the generation of the set of staples and the routing of the scaffold is estimated to be >700 bits for M13mp18, ensuring the security of information transfer.[31] Finally, the cipher mixture is generated by mixing the scaffold and the information strands. In this demonstration, we have 3 cipher mixtures since we have 3 letters: "A" at position 1, "S" at position 2, and "U" at position 3. Each cipher mixture consists of universal M13mp18 scaffold strands with all corresponding information strands for each letter and position (Fig. 1A; Table S4). The encryption process has 3 keys: the pattern encryption rules (key 1), the docking sequence (key 2), and the set of staple strands (key 3). The pattern and docking encryption add extra layers of security. In our first demonstration for 24 pattern spots, the key size for the pattern encryption is 84 bits[31]; a larger number of spots results in a larger key size. For the docking encryption, an 8-nt length docking sequence with 4 possible nucleotides at each position results in 4^8 possible combinations, equivalent to a 16-bit key. However, these additional layers of security are not the primary security measures, as the total key sizes for these two steps are significantly smaller than the key size of the DNA origami encryption.

The DNA origami decryption and readout. The decryption process involves the sequential application of 3 keys to the cipher mixtures (Fig. 1A, green boxes). First, the DNA staple strands key is applied to the mixtures, forming the RRO DNA origami through appropriate annealing (Fig. 1A). Next, DNA-PAINT super-resolution imaging is performed by applying the reverse complementary sequence of the docking strand,

A



B

Cypher mix	1	2	3		AFM image
+	+	+	2D mix (rectangle)		
+	+	+	3D mix (cuboctahedron)		

C

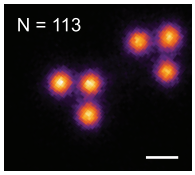
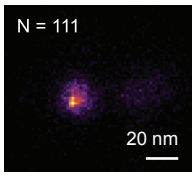
Cypher mix	1	2	3	Averaged High-speed DNA-PAINT image
+ (Fig. 3)	+	+	+	
+ (Fig. 3)	+	+	−	

Fig. 1 (Previous page) **DNA origami cryptography protocol.** A 3-step encryption/decryption process of information encoded in DNA strand cipher mix. Alice encrypts a message ("ASU") through templated pattern encryption, docking strand placement, and DNA origami folding. Bob decrypts using DNA origami formation through annealing, high-speed DNA-PAINT imaging, clustering and pattern matching. (**B**) 2D and 3D DNA structures for DNA origami cryptography and their corresponding AFM images. (**C**) Averaged DNA-PAINT images demonstrate successful decryption with all keys (top) and failed decryption with 1 missing key (bottom).

labeled with a Cy3B fluorophore, as the imager strands. High-speed DNA-PAINT docking enables the acquisition of a super-resolution image containing thousands of origami within a single 12-mins, 15,000-frame field of view. Averaged images of hundreds of DNA-PAINT images for each letter (Fig. 1C(top)) reveal the exact designed and encrypted pattern (compared with Fig. 3A). Without the staple strand key, DNA-PAINT yields random patterns, effectively concealing the true pattern (Fig. 1C(bottom)).

High-speed DNA-PAINT super-resolution imaging offers advantages over AFM due to its localization-based data, which seamlessly integrate with many clustering methods such as supervised machine learning,[41,42] Bayesian approaches,[43] and unsupervised machine learning for clustering[44] In our implementation, we employ K-means clustering to obtain the centroids of each cluster in a pattern and use them for template alignment. The encoded binary is then extracted by comparing the alignment with the pattern encryption rules to retrieve the "ASU" information. This process highlights the importance of the 3 keys in the decryption process. Without the staple strands key, the true pattern remains concealed (Fig. 1C(bottom)). In the absence of the docking sequence knowledge, DNA-PAINT cannot be performed to reveal the pattern, and attempting AFM imaging of the pattern formed by only 19-nt ssDNA docking sequence extensions is not practically feasible, as demonstrated by the invisibility of 12 docking extensions in the case of RRO and cuboctahedron (Figs. 1C(bottom) and S1). The pattern encryption rules key is crucial for translating the revealed pattern into binary codes to retrieve the encrypted information.

Detection and incorporation efficiency of information strands on 2D DNA origami via DNA-PAINT. Ensuring secure information transmission requires a high level of information retention and retrieval. Our encryption-decryption protocol relies on DNA origami architecture with integrated DNA-PAINT and unsupervised clustering readout. Previous work on DNA origami combined with DNA-PAINT readout for storage applications[21] showed a high error rate, thus requiring a mechanism and algorithm for robust error correction for each origami data droplet. Therefore, it is crucial to assess the incorporation and detection efficiency of the information strands on the RRO origami template to achieve higher information retention and retrieval. We investigate the detection efficiency of information strands on the 2D RRO DNA origami template via DNA-PAINT, where 1 information strand will be translated into 1 docking spot. To this end, we vary the number of spots per origami between 12, 24, and 48 as shown in Figs. 2A and S2). The increasing number of spots per origami leads to a reduction in the separation between spots. Specifically, the shortest distance between spots in origami with 12, 24, and 48 spots was systematically designed to be 20, 14, and 10 nm, respectively.

Extending the length of the information strand binder on DNA origami leads to higher incorporation and detection efficiency. We start with an RRO containing 12 spots/ origami, with a 32-nt information strand binder (Fig. 2A(i)), and estimate detection efficiency following the method developed by Strauss et al.[45] (Fig. S3). To improve the incorporation efficiency, we increase the information strand binder to 64-nt (Fig. 2A ($ii - iv$)), a strategy that has been shown to increase the melting temperature of the staple-scaffold complex during the annealing process.[46] We hypothesize that extending the information strand binders prolongs the stability of the information strand attachment on the DNA origami, notwithstanding physical perturbations arising from PEG purification, freeze-thaw cycles, pipetting, mixing, and local joule heating incurred during high-intensity DNA-PAINT imaging, thereby enhancing incorporation and detection efficiency.

Increasing the information strand binder length from 32-nt to 64-nt improves the probability of origami

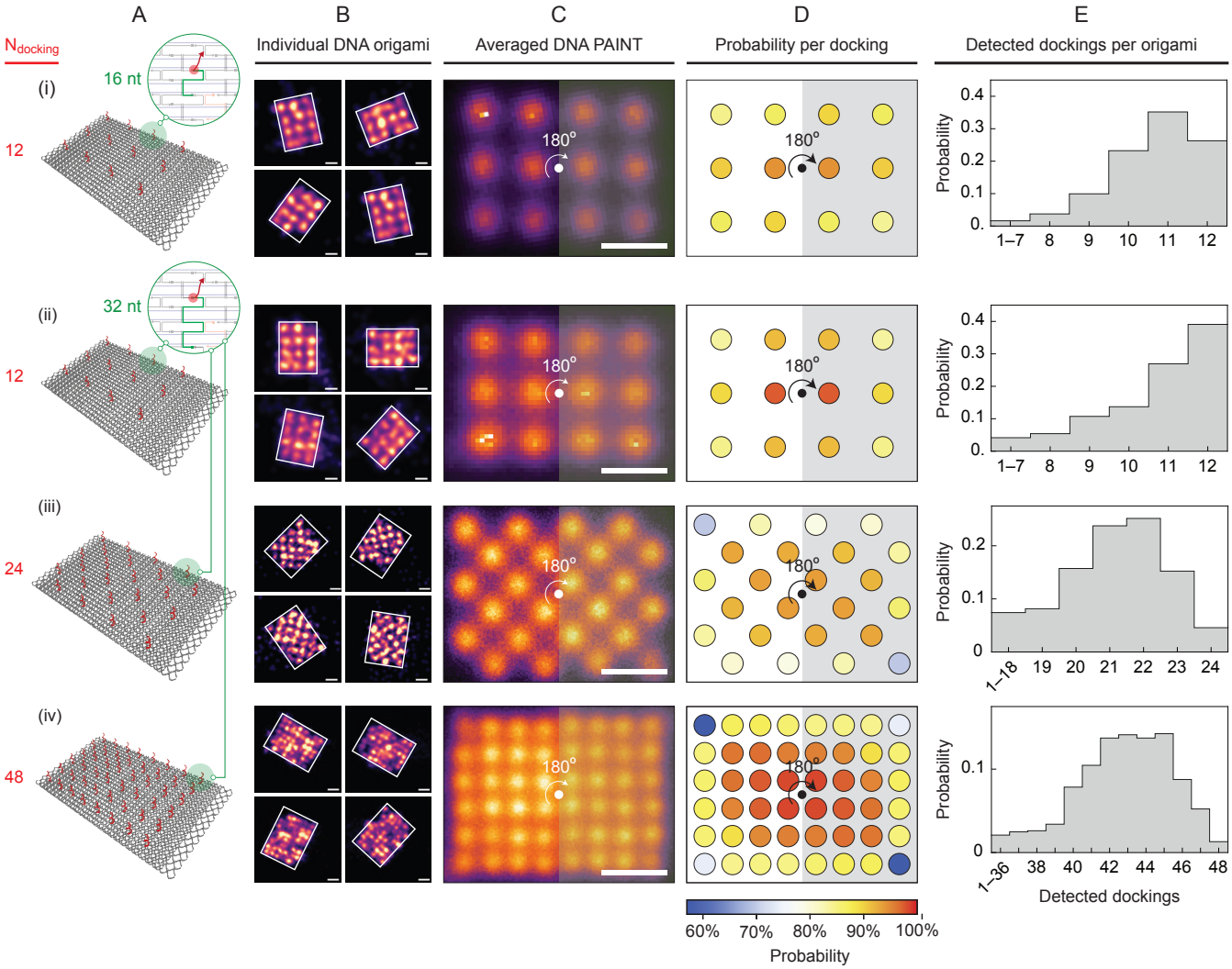


Fig. 2 **Detection efficiency of information strands on 2D DNA origami.** (**A**) Schematics of 2D RRO with red-colored docking strands: (i) 12 strands with 32-nt binders, (ii) 12 strands with 64-nt binders, (iii) 24 strands with 64-nt binders, and (iv) 48 strands with 64-nt binders. (**B**–**E**) Results for 2D RRO with varying numbers and lengths of symmetrically-placed information strands. (**B**) Examples of 4 individual DNA-PAINT origami of denoted system. Scale bars: 20 nm. (**C**) Averaged DNA-PAINT images ($N \geq$175 2D RROs). (**D**–**E**) Detection efficiency of each docking position (**D**) and the number of detected docking (**E**). Due to rotational symmetry, pairs of symmetric docking positions have identical averaged probabilities.

with all 12 dockings, while the detection efficiency varies depending on the docking location on 2D DNA origami. Fig. 2B–E presents the results for 12 spots per origami with varying binder lengths and numbers of information strands. Each row displays samples of DNA-PAINT from 4 individual origamis (Fig. 2B), an averaged image from N origamis or particles (Fig. 2C), the detection efficiency of each docking position (Fig. 2D), and the distribution of detected dockings (Fig. 2E). The probability of origami with all 12 dockings increased from 0.25 to 0.4 when the binder length was increased from 32-nt to 64-nt, despite a negligible increase in the mean detection efficiencies from (88.75%±9.99%) to (89.10±12.43%) (Fig. 2B and C). Using the offset of 7% established by Strauss et al.[45], the mean incorporation efficiencies were (95.75±9.99%) and (96.10±12.43%) for the 32-nt and 64-nt binders, respectively. These findings suggest that longer binders increase the likelihood of information strands remaining bound to the origami. The detection efficiency is highest (>94%) in the center areas of the origami and lowest at the corners and edges (Fig. 2B and C), consistent with the findings of Strauss et al.[45]

Increasing the density of docking spots per origami with a 64-nt binder affects the detection efficiency while necessitating the optimization of DNA-PAINT imaging parameters. We investigated the effect of higher docking density on detection efficiency by increasing the number of docking spots to 24 and 48 per origami (Fig. 2(*iii* and *iv*)). Consistent with the binomial distribution, the highest probability in the distribution did not correspond to full docking detection in both cases. The mean detection efficiencies were (87.92±7.1%) and (89.16±5.6%) for 24 and 48 docking spots, respectively, translating to incorporation efficiencies of (94.92±7.1%) and (96.16±5.6%). These results suggest that detection efficiency is influenced by both docking strand incorporation and the combined effects of docking density and DNA-PAINT imaging. Increasing docking density poses challenges for DNA-PAINT imaging, as closer spatial separation of docking spots (20 nm, 14 nm, and 10 nm for 12, 24, and 48 spots, respectively) necessitate adjustments in the imager strand concentration and acquisition frames to fully resolve the dockings (Table S5). Despite these adjustments, high-speed DNA-PAINT allows for reasonably fast imaging times of 12 mins and 75 mins for 12 and 48 docking spots, respectively, which is 4–15 times faster than AFM imaging for pattern readout on hundreds to thousands of origami per field of view.

The detection efficiency of docking spots via DNA-PAINT exhibits a systematic gradient across RRO, with its maximum value near the center and progressively lower values toward the edges and corners of RRO. However, increasing the docking density allows for more dockings with high detection values. The center locations of the origami consistently have detection efficiencies >90%, while the edge and corner dockings have lower values (Fig. 2D). A higher detection efficiency ensures high information retention and retrieval for our DNA origami encryption protocol. Despite the constraint imposed by our strategy of using longer binding to the scaffold for the information/docking strand limits the density to 48 docki

ngs/origami (corresponding to 10 nm resolution), this resolution allows for the theoretical encryption of $2^{28} \approx 268.4$ million combinations of numbers, letters, and punctuation marks that form texts and paragraphs, assuming near 100% incorporation and detection of information/docking strands (Fig. S4).

The DNA-PAINT and unsupervised clustering readout. We successfully executed the encryption and decryption the text "NSF" using DNA origami cryptography, showcasing the robustness and accuracy of our encryption and readout protocols (Fig. 3). The pattern encryption rules can accommodate $2^3 = 8$ combinations of letters, numbers, and punctuation marks; however, in this demonstration, we encode merely a 3-letter text of "NSF". Using a 32-nt binder for the docking/information strands, DNA-PAINT imaging is performed for 12 mins, yielding several hundred to a thousand super-resolution images of each pattern (Table S5). Around 100 particles of each pattern are selected for readout accuracy analysis (Figs. 3A and S5). K-means clustering is employed to obtain centroids corresponding to each docking (Fig. 3A), with the optimal number of clusters determined by the elbow method (see Method for details). Template alignment is then used to align the centroids to the pattern encryption rules, accommodating origami with shear angles α (Fig. 3A). Although the majority of DNA origami structures exhibit minimal shear angles nearing 0°, some correctly aligned origami experience significant shear deviations, presumably induced during the immobilization process on the glass coverslip. Bit extraction and readout are executed following

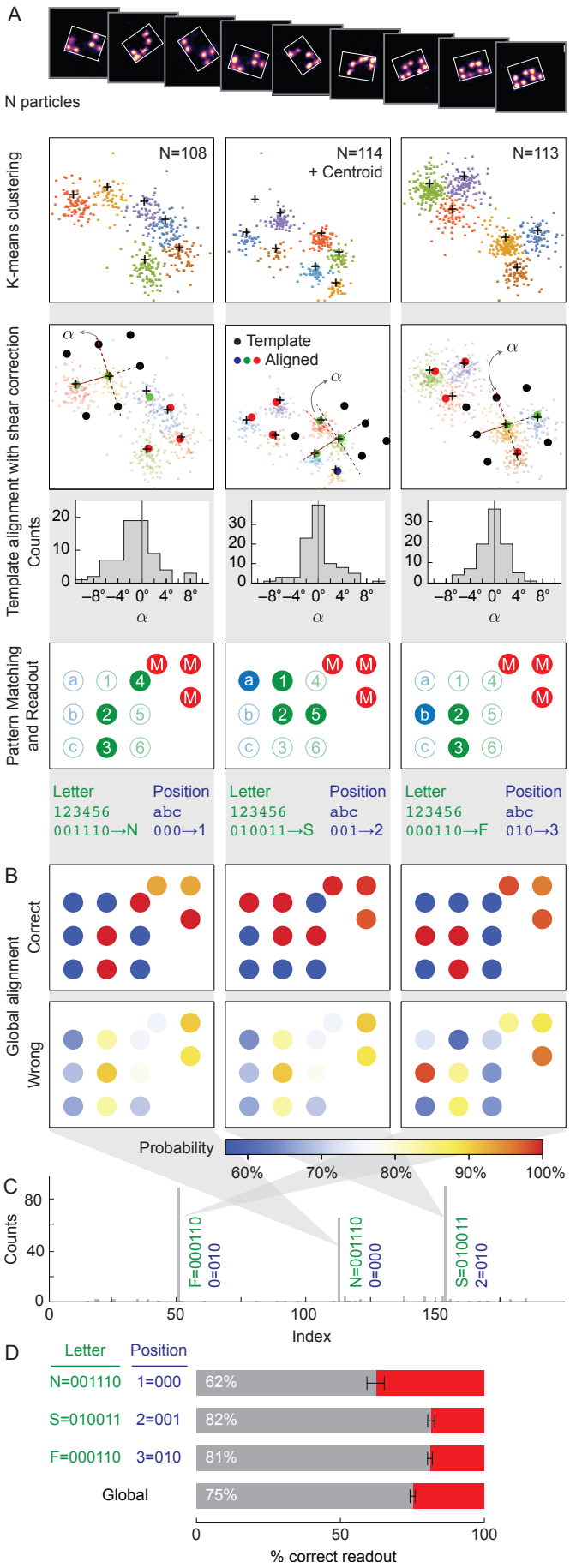

A
N particles
K-means clustering
N=108
N=114
+ Centroid
N=113
Template alignment with shear correction
Template
Aligned
Counts
Pattern Matching and Readout
Letter
Position
123456
abc
001110→N
000→1
010011→S
001→2
000110→F
010→3
B
Global alignment
Correct
Wrong
Probability
60%
70%
80%
90%
100%
C
Counts
Index
F=000110
0=010
N=001110
0=000
S=010011
2=010
D
Letter
Position
N=001110
1=000
62%
S=010011
2=001
82%
F=000110
3=010
81%
Global
75%
% correct readout

Fig. 3 (Previous page) **Unsupervised k-means clustering and template alignment for "NSF" data readout.** (**A**) Workflow of the process: N particles, K-means clustering, template alignment with shear correction and distribution of shear angle α in degrees for each correctly read letter, and pattern matching and readout. (**B**) Global bits alignment of correct (top) and wrong (bottom) readouts, displaying incorporation efficiency for each bit position using the method described in (A). (**C**) "NSF" data readout result presented as letter index vs. counts, showing 3 correct-readout major peaks and wrong-readout minor peaks. (**D**) Correct and wrong readout percentage for each letter. Error bars = $\pm$SD (N=3 runs of the process in panel A using the same dataset.

the template alignment process.

The analysis of individual letter accuracies and the global accuracy of the "NSF" text demonstrates the robustness and suitability of the DNA origami encryption and readout protocols for secure communication between two parties. Global bit alignment analysis (Fig. 3B) reveals 100% detection efficiency for non-alignment marker dockings in correct readouts. In contrast, incorrect readouts show undetected information-bearing dockings and non-zero detection in locations where no information dockings should be present, indicating misalignment of some origami. The combined readout results (Fig. 3C) show three prominent peaks corresponding to the "N,1", "S,2", and "F,3" letters, with wrong readouts occupying many small peaks. The count ratio between the lowest main peak and the highest false readout peak is 13.4, whereas the ratio between the highest main peak and the highest false readout peak is 18.4. Fig. 3D shows the individual letter accuracies with error bars denoting the standard deviation from three different runs of K-means clustering, template alignment by cost function minimization, and bit extraction. Despite the non-deterministic nature of the minimization, each letter has reasonable accuracy (>62%) with a relatively small error of $\leq$3%, indicating the method's robustness. The global accuracy of the "NSF" text is $\sim$75% with a relatively small error of 0.9%.

Incorporating bit redundancy in the encryption rules improves the readout accuracy of higher information density DNA origami with 48 docking spots per origami. Following the successful demonstration 3-letter DNA origami encryption and readout with 20 nm spaced docking spots, we design encryption rules to increase the information density using 48 docking spots per origami. Based on the detection efficiency results (Fig. 2(iv)), we assign 3 spots from each of the 4 corners as orientation markers, providing 3 redundancies to compensate for the lower detection and incorporation efficiency of corner dockings as opposed to other docking sites (Fig. 4A). We allocate 6 bits for encrypting letters, numbers, and punctuation marks through binary representation, alongside an additional 6 bits for position/order, accommodating up to $2^6 = 64$ characters to form a text. A total of 12 bits are mapped onto 12 docking spots on the origami, with 1 additional redundancy for each bit (highlighted in red in Fig. 4A). This scheme requires 24 docking spots in total (Fig. 4), with green circles representing the character bits and blue circles indicating the position bits.

The effect of redundancy on readout accuracy is investigated by encrypting the "ASU" text ("A,1", "S,2", and "U,3") (Fig. 4). We selected 200 particles of each pattern for readout accuracy analysis, with averaged images shown in Fig. 4B (see Fig. S6 for all analyzed picks). DNA-PAINT imaging takes over a duration of 24 mins (30,000 frames), generating data sets ranging from hundreds to thousands of super-resolution images per pattern (Table S5). K-means clustering and template alignment are performed. The global bit alignment maps for each letter (Fig. 4E) demonstrate that with one redundancy, the main bits do not need to be 100% present to retrieve the correct letter information, as the redundant bits support the main bits. In the correct readout map, the main bits are not always 100% present, but the redundant bits compensate for this. The incorrect readout map shows that some dockings without information strands have been falsely assigned a "1" bit. The combined readout of the 3 letter patterns is compared with and without bit redundancy (Fig. 4C and D). With redundancy, 3 clear main peaks corresponding to the correct letters and positions are observed, along with small peaks of wrong readout (Fig. 4C(top)). The ratio between the lowest main peak and the highest false readout peak is 13.2, and 19.3 between the highest main peak and the highest false readout peak. In contrast, without redundancy, the combined readout fails to retrieve the

encrypted information, showing comparable false readout peaks of "@,1" and "Q,2" that interfere with the final readout (Fig. 4C bottom panel).

Adding one-bit of redundancy significantly improves the accuracy of each letter and the global accuracy in higher docking density encryption. Fig. 4D shows the accuracy for each letter with and without bit redundancy. The global accuracy improves from ∼37% to ∼71% by adding one redundancy. The small standard deviation (<3%) from 3 different runs of K-means clustering, template alignment, and bit extraction demonstrates the robustness of our readout method. Two redundancies, utilizing all 48 docking spots, do not yield significantly better results than 1 redundancy (Figs. S7 and S8), indicating that the maximum global readout accuracy achievable by redundancy for these 3 letters ("ASU") is around 70–75%. False readouts can arise from factors such as imperfect imaging, causing the closest 2 spots to be poorly resolved, leading to failures in K-means centroid assignment and template alignment. We observe that the accuracy decreases as the number of bits occupied by a specific letter increases, as evident when comparing "A,1" (2 bits), "S,2" (4 bits), and "U,3" (5 bits) (Fig. 4D and S9).

The 3D wireframe DNA origami encryption, decryption, and fast readout. Implementing DNA origami cryptography in 3D nanostructures enhances security by concealing the encrypted pattern within complex 3D architectures, as opposed to the prior approach using 2D origami nanostructures.[31] Several groups have demonstrated various 3D DNA origami shapes, including the cuboctahedron (Figs. 1B and 5), which has a radius of ∼35 nm.[33,47,47–50] Revealing the true shape of and the encrypted pattern on 3D DNA origami requires 3D super-resolution imaging with sub-25 nm resolution. In comparison, AFM, which is restricted to 2D imaging, may misinterpret the 3D wireframe cuboctahedron as a 2D construct, leading to potential confusion about the encrypted pattern (Figs. 1B and S10). We identify the high-speed DNA-PAINT, capable of imaging 3D patterns on 3D DNA origami with ≤10 nm z-resolution,[39,51,52] as a superior readout technique for 3D DNA origami cryptography.

To demonstrate DNA origami cryptography in 3D, we encrypted the date "0407" (4th of July) on a 3D wireframe DNA origami cuboctahedron template. In our design, the docking/information strand binding length to the scaffold ranges from 52–54-nt (Table S2). The pattern encryption allocates 4 bits (ABCD) for number encryption, 3 bits (EFG) for position encryption, and 5 bits for alignment markers (3 filled and 2 empty) to break the in-plane (xy) rotational symmetry (Fig. 5A). The bits are located at the vertices of the cuboctahedron, with the possibility of placing more bits on the edges. The distance between 2 stacking bits along the z axis is 65–70 nm based on the ideal design (Fig. S11). We acquired 43,510 frames with an exposure time of 50 ms, resulting in a total acquisition time of ∼35 mins. We then selected 150–210 3D DNA origami structures for each pattern using visual inspection (Fig. S12).

The localization data for each selected DNA origami (Fig. 5B) reveals the challenges in resolving the vertical (z) separation of 60–70 nm with our experimental setup. Distinguishing 2 clusters in the localization data is difficult, as localizations from a docking spot occupy a 3D space with a z value range of approximately 50–60 nm (Fig. 5B and C). The flexibility of 3D wireframe origami, as revealed by AFM, further compromises the z resolution. Comparing our results to previous studies is challenging, as no groups have attempted 3D DNA-PAINT on single-scaffolded 3D wireframe DNA origami, and the z separation of <70 nm is beyond the state-of-the-art resolution achieved by many groups.[51–53] Despite the imperfect 3D DNA-PAINT localization data, the K-means clustering and elbow method can still assign centroids (Fig. 5B middle panel). However, the ∼30 nm separation between the 2 centroids from the K-means clustering does not conform to the designed docking position, which we attribute to the large spread of the z value of localization data for 2 stacking clusters compared to the x and y spread of 2 neighboring clusters.

The averaged 3D localization data images post-alignment indicate that the top and side perspectives of each pattern exhibit a high degree of match with the schematic in the top row of Fig. 5C. The side views demonstrate a clear height difference between a single cluster and 2 stacking clusters. Despite the challenge of obtaining clearly resolved 2 stacking dockings, we successfully aligned the centroids with the template (Fig. 5B). The combined readout shows 4 distinct peaks corresponding to the encrypted numbers "0,1", "4,2", "0,3", and "7,4", with minimal false readout peaks in the background (Fig. 5D). By employing the

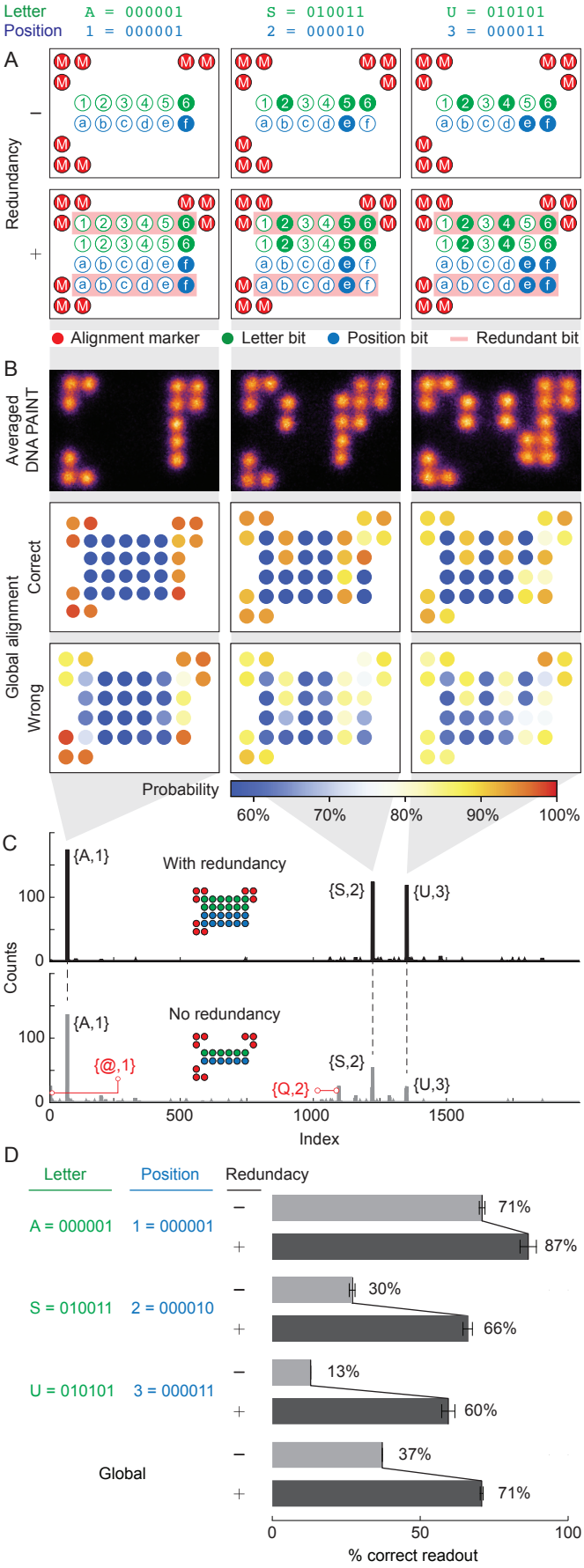

Letter A = 000001 S = 010011 U = 010101
Position 1 = 000001 2 = 000010 3 = 000011
A
Redundancy
Alignment marker
Letter bit
Position bit
Redundant bit
B
Averaged DNA PAINT
Global alignment
Correct
Wrong
Probability
60%
70%
80%
90%
100%
C
Counts
With redundancy
No redundancy
{A,1}
{S,2}
{U,3}
{@,1}
{Q,2}
Index
D
Letter
Position
Redundancy
A = 000001
1 = 000001
S = 010011
2 = 000010
U = 010101
3 = 000011
Global
71%
87%
30%
66%
13%
60%
37%
71%
% correct readout

Fig. 4 (Previous page) **Bit redundancy on higher density 2D RRO.** (**A**) Pattern encryption rules for "ASU" dataset showing alignment marker, letter bits, position bits, and redundancy (left); summed DNA-PAINT images of 3 letters of "ASU" (right), scale bars: 10 nm. (**B**) Examples of K-means clustering and template alignment for each letter (left); distribution of shear angle in degrees for each letter (right). (**C**) Readout of "ASU" dataset presented as letter index vs. counts, analyzed with redundancy (top) and without redundancy (bottom). (**D**) Readout percentage of correct and wrong readout for each letter and global; error bars represent standard deviation from 3 processing runs with the same dataset. (**E**) Global bits are aligned with the correct readout (top) and wrong readout (bottom), showing incorporation efficiency for each bit position.

method in Fig. 5B with the addition of alignment marker filtering, we extract the bits and retrieve the encrypted information with a global accuracy of 88% (Fig. 5D right panel). The alignment marker filtering is necessary due to the lack of redundancy in the alignment markers compared to the 2D origami pattern encryption rules. Our result demonstrates, for the first time, the successful application of 3D DNA-PAINT on a 3D wireframe cuboctahedron DNA origami as a method for 3D DNA origami encryption readout, achieving complete information retrieval.

Coarse-grained dynamic simulations using oxDNA reveal the flexibility inherent in the 3D DNA origami cuboctahedron, with the mean structure having root mean squared fluctuations (RMSF) ranging from 7.1 to 11.45 nm. The vertices become more rounded, and the edges bend, accounting for the observed compromise in the resolution along the x, y, and z axes resolution in the 3D DNA-PAINT data (Figs. 5C And 6A). Aligning the 3D DNA-PAINT localization data ($N = 173$ particles) bearing all vertices' docking with the docking center points in the mean structure, using 3D K-means clustering (Fig. S13A) and additional z-axis scaling, results in an alignment with a root mean squared distance (RMSD) of 100.88 nm (Fig. 6A(right)). However, alignment on the 2D projection of the 3D DNA-PAINT data and the mean structure produces better results, characterized by an RMSD of 26.07 nm (Figs. 6B and S13D), indicating that the oxDNA simulation agrees well with the experimental DNA-PAINT data in the xy plane without the need for scaling. The observed discrepancy along the z-axis could be attributed to potential inaccuracies in our 3D calibration. The intrinsic flexibility of the 3D wireframe DNA origami cuboctahedron becomes increasingly evident upon examination of the oxDNA trajectory (Fig. 6C). Comparing the 3D alignment of DNA-PAINT in the mean structure with the alignment in the unrelaxed structure (Fig. S13B, C, and E) corroborates the interpretation that the structure exhibits significant flexibility and agrees better with the mean structure obtained from the oxDNA simulations.

Discussion

In this study, we demonstrate a DNA origami cryptography protocol that significantly increases the practicality and potential for secure information storage and transmission using 2D and 3D DNA origami structures. Our protocol leverages 3D DNA origami structures and fast DNA-PAINT readout to achieve high-density and secure data storage for communication. Using high-speed DNA-PAINT, we achieve a >6-fold increase in data acquisition time compared to previously published DNA-PAINT readout of DNA origami information storage.[21] Building upon the foundational work by Dickinson et al.[21] and Zhang et al.[31], our approach offers several key advantages, including increased information density, improved design criteria, faster imaging and readout speeds, enhanced geometric complexity, and up to 89% accuracy in data retrieval. By exploiting the vast possibilities of scaffold routing, we generate a large key space for secure information encryption and retrieval in both 2D and 3D nanostructures. Furthermore, we demonstrate the crucial role of error correction codes in maximizing the chance of complete message retrieval, enabling the encryption of information using a one-redundancy code with 48 dockings per origami.

Using the results obtained in our study, design criteria that have never been discussed before in the DNA origami cryptography proof of concept work by Zhang et al.[31] can be laid out to achieve high information retention and retrieval with high-speed pattern readout utilizing the DNA-PAINT technique combined with unsupervised clustering and template alignment. We recommend having a docking strand binder to be

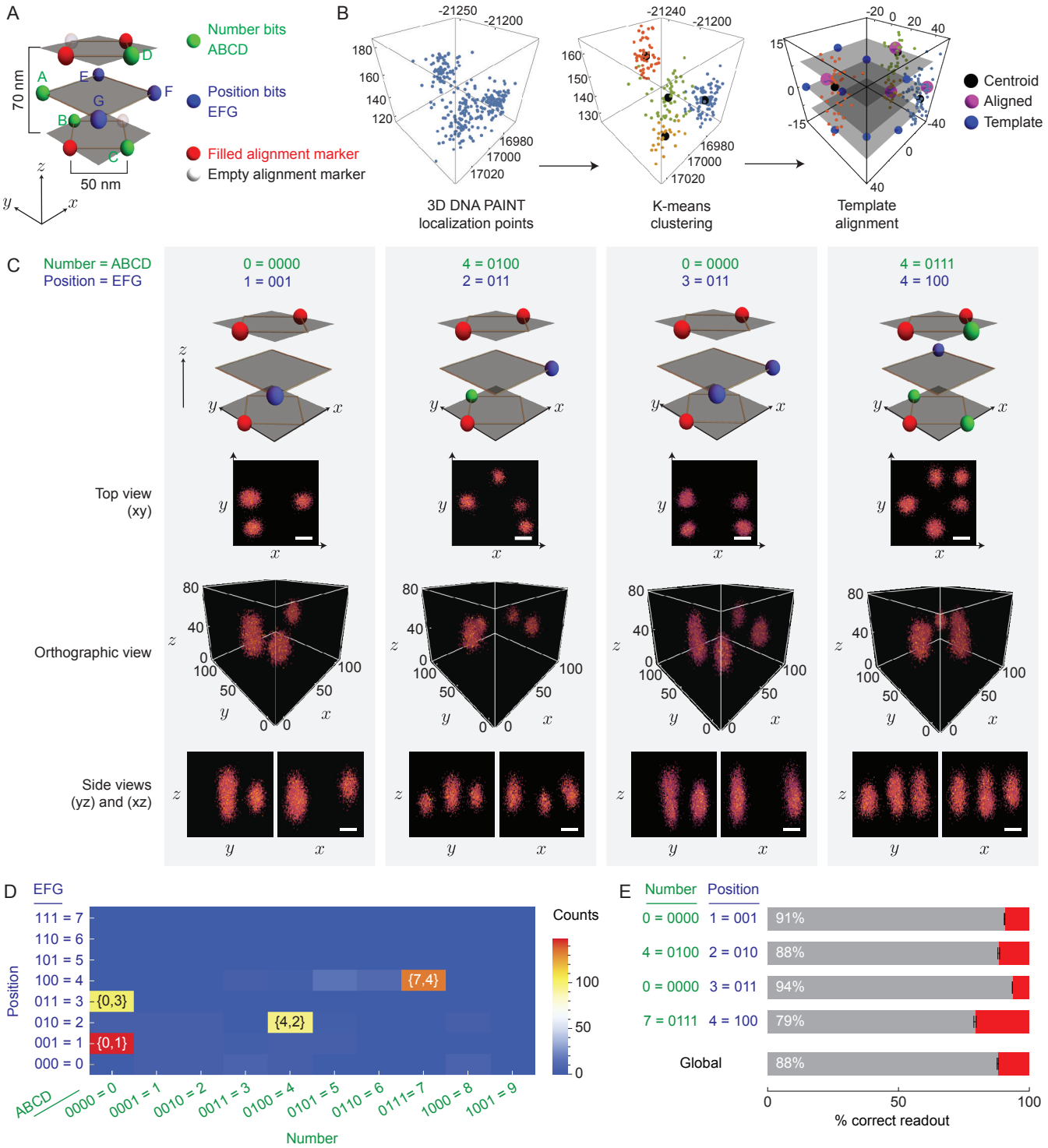


Fig. 5 **3D DNA PAINT encryption applied to information encoded in a 3D wireframe DNA origami cuboctahedron.** (**A**) Encryption rule for the pattern comprising number bits (green), position bits (blue), and alignment markers (red). Three z planes are colored transparent gray, occupied by the corner points. (**B**) Example of 3D DNA-PAINT localization data from a pattern on the 3D wireframe cuboctahedron DNA origami (left), K-means clustering result (middle), and 3D alignment with a cuboctahedron template (right). (**C**) Superposition images of correctly aligned and read 4 patterns from identified 3D DNA origami ($N \geq 95$). Scale bars: 20 nm. (**D**) One run of readout presented as positions vs. numbers vs. counts, showing 4 main peaks of letters. (**E**) Individual letter readout mean percentage of correct (gray) and wrong (red) readouts with uncertainties (mean±SD; N = 3 runs of the same dataset processed as in (B)).

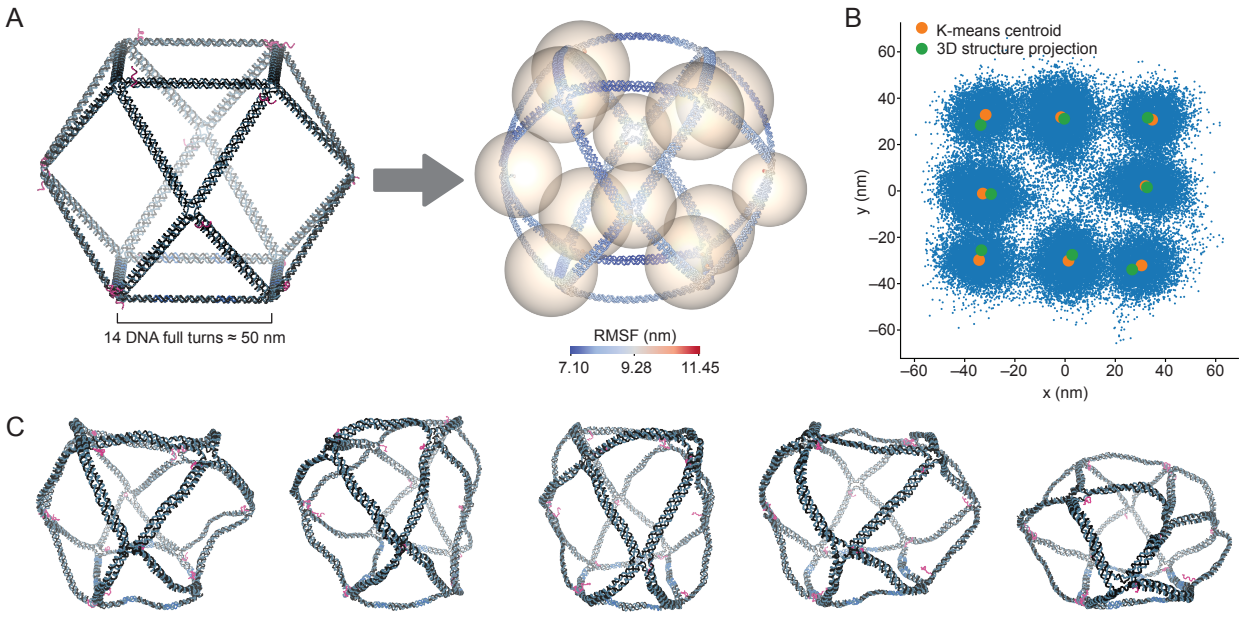


Fig. 6 3D wireframe cuboctahedron oxDNA and DNA-PAINT alignment with the mean structure. (**A**) Orthographic views of the original unrelaxed 3D cuboctahedron structure with red-colored docking handles (left) and the mean structure after oxDNA simulation, with nucleotide coloring based on RMSF (right). Spheres represent DNA-PAINT data, with centers based on centroids of clusters after 3D K-means clustering and radii based on the average distance of each localization to their centroids. Spheres fitting with the mean structure are obtained after scaling the z value of DNA-PAINT data by 3.3. (**B**) 2D projection of DNA-PAINT localizations from 3D cuboctahedrons (N = 173). The K-means centroids (averaged if there are 2 stacking centroids along the z-axis) are shown by orange disks, while th2D projections of the docking ssDNA centers post-oxDNA simulation (similarly averaged if there are 2 stacking projections along the z-axis) are denoted by green disks. (**C**) Time lapse from the oxDNA trajectory of the cuboctahedron reveals structural fluctuations. The dockings are colored in red and biotinylated strands are colored in light blue.

at least 32 nt, with the longer, the better (Fig. 1B–C) for high efficiency of information retrieval through DNA-PAINT and the readout protocols. The assignment of dockings is also important. We suggest utilizing the dockings in the center area of a 2D structure for higher incorporation and detection efficiency, thus giving higher information retention and retrieval (Fig. 1B–E). For 3D origami, we encourage the user to have a z dimension of at least 100 nm due to the resolution limit of 3D DNA-PAINT. The redundancy of bits is significant when dealing with high-density docking origami (Fig. 4). We recommend having an error correction code to maximize the chance of complete message retrieval. In our case of a one-redundancy code with 48 dockings per origami, by using 12 dockings at corners for alignment markers, 36 dockings are available for the information bits, which will then be split into 6 bits for letters plus 6 bits for the redundancy and 12 bits for positions plus 12 bits for position bits redundancy. The 12 bits correspond to a total of 4096 letters, which is equivalent to 2 pages of text information (assuming 1 page can have 2000 characters). The possibility of 2 and 3D geometries will increase the complexity of the scaffold routing and the possibility of staple strands interlacing; thus, we expect to have key sizes bigger than the estimated 700 bits for standard M13mp18 scaffold from Zhang et al.[31] Our protocol also improves the information density, design criteria, imaging and readout speed, and especially geometric complexity for the benefit of security. Since the key size depends on the length of the scaffold, another possible avenue to increase the key size is using several orthogonal scaffolds, thus enabling bigger DNA origami. Multi-orthogonal scaffolding of bigger DNA origami has been demonstrated.[54] As an implication, with bigger DNA origami and the same information density per origami, more information strands can be incorporated, thus giving higher total information that can be encrypted in a specific multi-orthogonal scaffold DNA origami.

While our protocol exhibits high accuracy in information retrieval, there is still room for improvement in resolution and density. First, state-of-the-art DNA-PAINT has achieved Ångström[39] and sub-5nm[40] resolution, surpassing what we demonstrated in this work. To achieve this resolution, shorter binding lengths between the docking strands and the scaffold would be required, potentially compromising the retention of the docking/information strands on the origami. To mitigate this limitation, we propose the use of modified information strands or scaffolds with covalent linkers,[54,55] thereby strengthening the binding affinity and ensuring robust information retrieval. Second, to further increase information density, multicolor DNA-PAINT can be used.[40,51] For instance, in the case of a 2D RRO template with 48 docking strands per origami, using 2 fluorophores such as ATTO 488 and ATTO 647 for 2 orthogonal high-speed docking sequences would allow for an additional 36 docking spots (assuming that 12 docking spots of the new color are also being dedicated for alignment markers). These docking spots can be used for 18 additional position bits and 18 redundancy bits, collectively resulting in 28 position bits, capable of encoding ∼268 million characters (Fig. S14).

For data classification, we conducted a comprehensive evaluation of multiple methodologies, ultimately adopting K-means clustering proposed by Huijben et al.[56]. This decision was predicated on its efficacy in 2D and 3D space, coupled with its direct implementation and fast computational performance. Alternative methodologies, including DBSCAN, mean-shift clustering, and Gaussian mixture models, require fine-tuning numerous parameters tailored to the dataset. Conversely, the K-means approach primarily hinges on determining the optimal number of clusters through the application of the elbow method. The method employed by Huijben et al. has a runtime that increases quadratically with the number of origami due to the requirement of pairwise registration to create the dissimilarity matrix, whereas our method has a linear runtime. We acknowledge that using their classification methodology enables us to achieve an enhanced accuracy of ∼90% for the "NSF" dataset, albeit at the expense of speed. However, we encounter limitations in accurately decoding the "ASU" dataset with one redundancy (Fig. S15). Furthermore, we have explored supervised learning classification frameworks such as MLP[41] and ResNet convolutional neural networks (CNN), as demonstrated in (Fig. S16). These models require a training dataset encompassing all possivle permutations of the pattern for each encryption rule, rendering the approach impractical. Specifically, for our "ASU" pattern encryption rules, which involve 24 information bits, the generation of approximately 2^{24}, or roughly 16.7 million, patterns would be required for the training dataset. Our readout method, which employs K-means clustering alongside template alignment via minimization, eliminates the need for training data and facilitates a fully automated readout process when integrated with auto-picking. Further automation and speed improvement are feasible through the application of DNA-PAINT augmented by deep neural networks.[57]

In conclusion, our DNA origami cryptography protocol significantly advances the field of molecular and DNA cryptography by offering new design considerations and demonstrating the versatility of our approach through the exploration of alternative machine learning methods for readout. The successful integration of 3D DNA origami structures, fast DNA-PAINT readout, and machine learning algorithms for data classification demonstrates the power of interdisciplinary collaboration in advancing the frontiers of biophysics and computer science. Our work showcases the potential for secure and high-density information storage and communication using DNA nanostructures, representing a crucial step forward in developing practical and secure methods for data storage and transmission at the molecular scale. As we continue to explore the vast potential of DNA as a medium for information storage and transmission, future research could focus on further increasing the information density and security of the protocol by integrating advanced techniques such as multicolor DNA-PAINT, the utilization of information strands or scaffolds composed of modified DNA with covalent linkers, and the application of deep neural networks for faster automation. The implications of our work extend beyond cryptography, with potential applications in various fields, including healthcare, data storage, and computing, highlighting the interdisciplinary nature of this research and its potential to inspire new developments in biophysics, computer science, and applied mathematics. Ultimately, our work lays the foundation for future research and development in the rapidly

evolving field of DNA-based information storage.

Funding

The research was funded by the National Science Foundation (NSF) through the SemiSynBio II (2027215) and SemiSynBio III (2227650) to H. Yan, C. Wang, and R.F. Hariadi. G.B.M. Wisna is supported by an American Heart Association (AHA) predoctoral fellowship (23PRE1029870). R.F. Hariadi acknowledges the funding from the National Institutes of Health (NIH; 1DP2AI144247-01) and NSF (2341002).

Acknowledgements

The authors thank Yinan Zhang, Pengkun Xia, Md Ashikur Rahman Laskar, Nirbhik Acharya, Ranjan Sasmal, Franky Djutanta, Ritvik Warrier, and Raymond Nucuta for their insightful discussions on our work. We also extend our thanks to Phillip R. Steen, Luciano A. Masullo, Hao Liu, and Raphael Jorand for the valuable discussions on 3D calibration and 3D DNA PAINT and to Teun A.P.M. Huijben for his helpful insights on unsupervised classification. Additionally, we thank Vaughn McGill-Adami for enabling ResNet.

Correspondence

All correspondence and requests for materials should be addressed to G.B.M. Wisna and R.F. Hariadi.

References

1. Riordan M, Hoddeson L (1997) *Crystal fire: The birth of the information age.* (WW Norton & Company).
2. Katz J, Lindell Y (2020) *Introduction to modern cryptography.* (CRC press).
3. Goldreich O (1998) *Modern cryptography, probabilistic proofs and pseudorandomness.* (Springer Science & Business Media) Vol. 17.
4. National Institute of Standards and Technology (2001) Advanced encryption standard (AES), (Gaithersburg, MD), Technical report.
5. Voas J, Kshetri N, DeFranco JF (2021) Scarcity and global insecurity: the semiconductor shortage. *IT Professional* 23(5):78–82.
6. Shehabi A, et al. (2016) United states data center energy usage report.
7. Matange K, Tuck JM, Keung AJ (2021) DNA stability: a central design consideration for DNA data storage systems. *Nat Commun* 12(1):1358.
8. Goldman N, et al. (2013) Towards practical, high-capacity, low-maintenance information storage in synthesized dna. *Nature* 494:77–80.
9. Extance A (2016) How DNA could store all the world's data. *Nature* 537.
10. Cox JP (2001) Long-term data storage in dna. *Trens Biotechnol* 19(7):247–250.
11. Liu Q, et al. (2000) DNA computing on surfaces. *Nature* 403(13):175–178.
12. Adleman LM (1998) Computing with dna. *Scientific american* 279(2):54–61.
13. Braich RS, Chelyapov N, Johnson C, Rothemund PW, Adleman L (2002) Solution of a 20-variable 3-sat problem on a DNA computer. *Science* 296(5567):499–502.
14. Păun G, Rozenberg G, Salomaa A (1998) *DNA computing: new computing paradigms.* (Springer).
15. Qian L, Winfree E, Bruck J (2011) Neural network computation with DNA strand displacement cascades. *Nature* 475(7356):368–372.
16. Qian L, Winfree E (2011) Scaling up digital circuit computation with DNA strand displacement cascades. *Science* 332(6034):1196–1201.
17. Seelig G, Soloveichik D, Zhang DY, Winfree E (2006) Enzyme-free nucleic acid logic circuits. *Science* 314(5805):1585–1588.
18. Church GM, Gao Y, Kosuri S (2012) Next-generation digital information storage in dna. *Science* 337(6102):1628–1628.
19. Ceze L, Nivala J, Strauss K (2019) Molecular digital data storage using dna. *Nature Reviews Genetics* 20(8):456–466.
20. Organick L, et al. (2018) Random access in large-scale DNA data storage. *Nature Biotechnology* 36(3):242–248.
21. Dickinson GD, et al. (2021) An alternative approach to nucleic acid memory. *Nat. Commun.* 12(1):2371.

22. Sarkar T, Selvakumar K, Motiei L, Margulies D (2016) Message in a molecule. *Nat. Commun.* 7(1):11374.
23. Boukis AC, Reiter K, Frölich M, Hofheinz D, Meier MA (2018) Multicomponent reactions provide key molecules for secret communication. *Nat. Commun.* 9(1):1439.
24. Lustgarten O, Motiei L, Margulies D (2017) User authorization at the molecular scale. *ChemPhysChem* 18(13):1678–1687.
25. Vrijsen JH, Rubens M, Junkers T (2020) Simple and secure data encryption via molecular weight distribution fingerprints. *Polymer Chemistry* 11(40):6463–6470.
26. Clelland CT, Risca V, Bancroft C (1999) Hiding messages in DNA microdots. *Nature* 399(6736):533–534.
27. Leier A, Richter C, Banzhaf W, Rauhe H (2000) Cryptography with DNA binary strands. *Biosystems* 57(1):13–22.
28. Gehani A, LaBean T, Reif J (2003) DNA-based cryptography in *Aspects of Molecular Computing*, Lecture notes in computer science. (Springer Berlin Heidelberg, Berlin, Heidelberg), pp. 167–188.
29. Cui G, Qin L, Wang Y, Zhang X (2008) An encryption scheme using DNA technology in *2008 3rd International Conference on Bio-Inspired Computing: Theories and Applications*. (IEEE), pp. 37–42.
30. Heider D, Barnekow A (2007) Dna-based watermarks using the dna-crypt algorithm. *BMC bioinformatics* 8(1):1–10.
31. Zhang Y, et al. (2019) DNA origami cryptography for secure communication. *Nat Commun* 10:5469.
32. Rothemund PW (2006) Folding DNA to create nanoscale shapes and patterns. *Nature* 440(7082):297–302.
33. Dietz H, Douglas SM, Shih WM (2009) Folding DNA into twisted and curved nanoscale shapes. *Science* 325(5941):725–730.
34. Provos N, Honeyman P (2003) Hide and seek: An introduction to steganography. *IEEE security & privacy* 1(3):32–44.
35. Hansma PK, Schitter G, Fantner GE, Prater C (2006) High-speed atomic force microscopy. *Science* 314(5799):601–602.
36. Gole A, Murphy CJ (2005) Biotin- streptavidin-induced aggregation of gold nanorods: tuning rod- rod orientation. *Langmuir* 21(23):10756–10762.
37. Aslan K, Luhrs CC, Pérez-Luna VH (2004) Controlled and reversible aggregation of biotinylated gold nanoparticles with streptavidin. *The Journal of Physical Chemistry B* 108(40):15631–15639.
38. Strauss S, Jungmann R (2020) Up to 100-fold speed-up and multiplexing in optimized DNA-paint. *Nat. Methods* 17(8):789–791.
39. Reinhardt SC, et al. (2023) Ångström-resolution fluorescence microscopy. *Nature* 617(7962):711–716.
40. Dai M, Jungmann R, Yin P (2016) Optical imaging of individual biomolecules in densely packed clusters. *Nature Nanotechnol* 11(9):798–807.
41. Auer A, Strauss MT, Strauss S, Jungmann R (2020) NanoTRON: a Picasso module for mlp-based classification of super-resolution data. *Bioinformatics* 36(11):3620–3622.
42. Williamson DJ, et al. (2020) Machine learning for cluster analysis of localization microscopy data. *Nat. Commun.* 11(1):1493.
43. Fazel M, et al. (2022) High-precision estimation of emitter positions using bayesian grouping of localizations. *Nat. Commun.* 13(1):7152.
44. Simoncelli S, et al. (2020) Multi-color molecular visualization of signaling proteins reveals how c-terminal src kinase nanoclusters regulate t cell receptor activation. *Cell Reports* 33(12):108523.
45. Strauss MT, Schueder F, Haas D, Nickels PC, Jungmann R (2018) Quantifying absolute addressability in DNA origami with molecular resolution. *Nat. Commun.* 9(1):1600.
46. Ramakrishnan S, et al. (2019) Enhancing the stability of DNA origami nanostructures: staple strand redesign versus enzymatic ligation. *Nanoscale* 11(35):16270–16276.
47. Zhang F, et al. (2015) Complex wireframe DNA origami nanostructures with multi-arm junction vertices. *Nature Nanotechnol* 10(9):779–784.
48. Douglas SM, et al. (2009) Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* 459(7245):414–418.
49. Han D, et al. (2011) DNA origami with complex curvatures in three-dimensional space. *Science* 332(6027):342–346.
50. Veneziano R, et al. (2016) Designer nanoscale DNA assemblies programmed from the top down. *Science* 352(6293):1534–1534.
51. Jungmann R, et al. (2014) Multiplexed 3d cellular super-resolution imaging with DNA-PAINT and exchange-PAINT. *Nat. Methods* 11(3):313–318.
52. Auer A, et al. (2018) Nanometer-scale multiplexed super-resolution imaging with an economic 3d-DNA-PAINT microscope. *ChemPhysChem* 19(22):3024–3034.
53. Iinuma R, et al. (2014) Polyhedra self-assembled from DNA tripods and characterized with 3d DNA-paint.

Science 344(6179):65–69.
54. Engelhardt FA, et al. (2019) Custom-size, functional, and durable DNA origami with design-specific scaffolds. *ACS Nano* 13(5):5015–5027.
55. Gerling T, Kube M, Kick B, Dietz H (2018) Sequence-programmable covalent bonding of designed DNA assemblies. *Sci. Adv.* 4(8):eaau1157.
56. Huijben TA, et al. (2021) Detecting structural heterogeneity in single-molecule localization microscopy data. *Nat. Commun.* 12(1):3791.
57. Narayanasamy KK, Rahm JV, Tourani S, Heilemann M (2022) Fast DNA-paint imaging using a deep neural network. *Nat. Commun.* 13(1):5047.
58. Schnitzbauer J, Strauss MT, Schlichthaerle T, Schueder F, Jungmann R (2017) Super-resolution microscopy with DNA-paint. *Nature Protocols* 12(6):1198–1228.
59. Stahl E, Martin TG, Praetorius F, Dietz H (2014) Facile and scalable preparation of pure and dense DNA origami solutions. *Angewandte Chemie* 126(47):12949–12954.
60. Huang B, Wang W, Bates M, Zhuang X (2008) Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* 319(5864):810–813.
61. Schindelin J, et al. (2012) Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9(7):676–682.
62. Virtanen P, et al. (2020) Scipy 1.0: fundamental algorithms for scientific computing in python. *Nat. Methods* 17(3):261–272.
63. He K, Zhang X, Ren S, Sun J (2015) Deep residual learning. *Image Recognition* 7.

Materials and methods

Materials. Unmodified DNA staple strands and biotinylated staple strands for 2D and 3D origami were obtained from Integrated DNA Technology (IDT). M13mp18 scaffold strands were ordered from Bayou Biolabs. Amine-modified DNA strands for imager strands were purchased from IDT DNA. Cy3B-NHS Ester fluorophores (PA63101) were ordered from General Electric Healthcare. Chemicals were purchased from Sigma-Aldrich. Glass coverslips (48466-205) and microscope slides (16004-430) were purchased from VWR. Kapton tape (PPTDE-2) for the flow chamber was purchased from Bertech.

Imager strands. Imager strands were prepared by conjugating amine-modified DNA strands to Cy3B using NHS Ester coupling. The conjugated imager strands were then purified using high-performance liquid chromatography (HPLC).

2D and 3D DNA origami folding and purification. To fold 2D and 3D DNA origami, M13mp18 scaffold strands were mixed with staple, biotinylated strands, and the corresponding docking strands for each pattern in 1X TAE 12.5 mM Mg^{2+} buffer. The scaffold strand concentration was 20 nM, while the staple strands were at a 10X concentration compared to the scaffold. To optimize the incorporation of docking and biotinylated strands, an excess concentration of 50–60X was used, as described by Strauss et al.[45] The mixture was annealed by ramping up to 80 °C for 5 minutes, followed by a slow ramp down to 4 °C at a rate of 3 minutes 12 seconds per degree Celsius, based on the protocol by Schnitzbauer et al.[58] After annealing, the DNA origami solution was purified using a PEG precipitation method with a PEG buffer (15% (m/v) PEG 8000, 12.5 mM $MgCl_2$, 505 mM NaCl, 5 mM Tris-HCl, 1 mM EDTA, pH=8.0) for three rounds, as described by Stahl et al.[59] Finally, the solution was redispersed in 1X TAE 12.5 mM Mg^{2+} buffer and stored at –20 °C. The entire process encompasses the decryption of DNA origami through origami folding.

AFM imaging of 2D and 3D DNA origami Bruker multimode AFM was used to image DNA origami after PEG purification. 3 μL of 1-2 nM DNA origami solution was deposited on a freshly cleaved mica surface on an AFM metal sample slab. Then, 60 μL of 1X TAE 12.5 mM $MgCl_2$ and 4 μL of 0.2 M $NiCl_2$ solution were added to immobilize the origami onto the mica surface. This sample was imaged in the AFM using fluid mode with SCANASYST-FLUID+ model tip by Bruker. The image was processed using NanoScope Analysis AFM software.

Pattern readout by 2D and 3D DNA-PAINT super-resolution imaging and image processing DNA-PAINT superresolution imaging was performed following the detailed protocols previously described by Schnitzbauer et al.[58] Briefly, DNA origamis were immobilized on a BSA-biotin-streptavidin coated coverslip that forms a flow chamber with microscope slide glass through double-sided Kapton tape. Buffer A+ (10 mM Tris-HCl, 100 mM NaCl, 0.05% (v/v) Tween 20, pH 8.0) was used to dilute BSA-Biotin and streptavidin to 1 mg/ml and 0.5 mg/ml concentration, respectively. Buffer B+ (5 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM EDTA, 0.05% (v/v) Tween 20, pH 8.0) was used to dilute DNA origami at experimental concentrations. Buffer B+ was also used to dilute the imager strands (AGGAGGA/3' Cy3B/) to experimental concentrations. Oxygen scavenger solutions PCA, PCD and Trolox with final concentrations of 1.25X PCA, 1X PCD, and 1X Trolox were mixed with the imager strands to make the final imaging solution. The experimental conditions for each figure are described in the Table S4.

We used an Oxford nanoimager (ONI) of a Benchtop Nanoimager S Mark II with a total internal reflection fluorescence (TIRF) setup. Cy3B was excited using a 532 nm laser with a power density ranging from 800 W/cm^2 to 1250 W/cm^2. A 549-623 nm Bandpass filter was installed on the emission path to select Cy3B emission. For the "NSF" dataset, an Olympus objective with 100X magnification and 1.4 NA with oil immersion was used. For the "ASU" and "0407" dataset, an Olympus objective with 100X magnification and 1.49 NA with immersion oil was used. A Hamamatsu ORCA-Flash4.0 V3 digital sCMOS camera was used to acquire the DNA-PAINT movies with a camera exposure time of 50 ms. The microscope was equipped with z-lock autofocus with a piezo stage. For 3D DNA-PAINT, an additional 3D lens was inserted into the optical path that modifies the point spread function (PSF) from circular to elliptical, which is dependent on the z distance of the emitter from the focal spot through astigmatism.[52,53,58,60] The z-calibration was done by using 2D RRO by scanning the z from -500 nm to $+500$ nm with an increment of 10 nm made possible by the piezo stage. A standard 3D calibration curve (Fig. S17) was generated through the Picasso Localize module, which was used as a calibration curve to process the 3D DNA-PAINT movies. The imaging conditions for each figure are described in Table S5.

The DNA-PAINT movies were processed using FIJI[61] to crop the image into 256 px by 256 px. Then, each movie was fed into the Picasso Localize module to fit the PSF into precise localization data (Table S6 for Picasso Localize module parameters). The Picasso Render module was used to perform multiple redundant cross-correlation (RCC) drift corrections to obtain the final super-resolution images. The Picasso Filter module was used to remove outlier localizations based on x and y localization precision and localization background. The whole processing was performed on an Alienware Desktop Computer with an Intel Core i7-6800K CPU, 32 GB RAM, and NVIDIA GeForce GTX 1080 graphics card.

2D DNA-PAINT incorporation efficiency analysis The incorporation efficiency analysis followed the method described by Strauss et al.[45]. Briefly, Picasso processed localization data of super-resolution images of the patterns using the Picasso pick automatic function. All picks were aligned using the Picasso Average module. The aligned picks were then unfolded in the Picasso Render module to get arrays of picks that had been aligned in the same orientation. Then, it was fed into a Matlab script that analyzes the incorporation efficiency (Fig. S3).

2D and 3D data clustering, template alignment, and analysis The protocol followed the pipeline as shown in Fig. 3A. The patterns on each origami were picked by visual inspection on the Picasso Render module. The picks were then used as input for the following processing of K-means clustering, which assigns centroids based on the optimal number cluster followed by cluster filtering. Based on the pattern encryption rules, the centroids were aligned with their corresponding templates. The processing was done in a System 76 Thelio Desktop with AMD Ryzen 9 5950X 16-Core Processor, 32GB RAM, and NVIDIA GeForce RTX 3070 GPU. The detailed explanations for each process are as follows.

K-means Clustering was performed using the Scipy Python package[62] for the 2D data and a Matlab function for the 3D data. In the 3D data, extra data filtering was performed using the DBSCAN function

in Matlab by empirically assigning the radius (ϵ) to be 7 and the minimum number of neighbor points to determine the core to be 4. In the implementation, K-means clustering was used to cluster localizations into N through M clusters, where N and M were chosen according to the pattern encryption template. K-means clustering takes one parameter K for the number of clusters. The optimal K value was determined using the elbow method by running K-means for $(M - N)$ times over each origami, initially with N clusters and increasing by 1 every iteration until M clusters, where M was determined by the maximum docking numbers for specific templates. The inertias, centroids, and cluster memberships resulting from the K-means clustering were recorded at each iteration. Centroids are defined as the center of each cluster found by K-means. Inertias are defined as the squared distances between each localization and its corresponding centroid. The differences between the inertias, at each iteration, were termed gradients, and they were calculated, then negated and normalized to have a maximum value of 1. The optimal number of clusters was chosen by comparing these differences and locating the point at which the difference negligibly improves. The optimal number of clusters was the point at which the gradient value starts to saturate to a value of 0.95 or more (maximum value is 1) in the case of 3D data or closest to some threshold T_I in the case of 2D data. See Table S7 for the parameters. The elbow method is a relatively standard practice.

Cluster filtering was done after K-means clustering over the 2D data. Outlying clusters were filtered out on the basis of size, defined as the number of localizations belonging to that cluster. To do this, the mean size of all clusters was calculated and the size threshold was calculated by multiplying the mean size by some value T_S. Clusters smaller than the threshold size were discarded and localizations belonging to those clusters were filtered out.

Template creation was performed based on the pattern encryption rules of 2D and 3D DNA origami. Four templates were created for the grids used in the "NSF" dataset, the "ASU" one redundancy dataset, the "ASU" two redundancy dataset, and the 3D cuboctahedron "0407" dataset. These templates use prior knowledge of the possible locations of the binding site and orientation marker positions. Each point in the template represents the approximate position (x, y) of the binding site in an idealized origami. Note that all binding sites are included, not just those used in any particular pattern.

2D Template Alignment. The centroids, template, cluster sizes, template weights, and orientation markers were used to calculate a transform of the template that minimizes the following cost function modified from the Euclidean squared distance formula:

$$C = \frac{\sqrt{D+1}}{n} \sum_{i=1}^{n} (|A_i - B_i| P_i S_i)^2, \qquad [1]$$

where A_i and B_i are the Euclidean coordinates of the ith centroid and the closest point in the transformed template to A_i by Euclidean distance, respectively. S_i is the number of localizations in the ith centroid's cluster. $|A_i\text{-}B_i|$ is the Euclidean distance between A_i and B_i. P_i is the weighting of B_i defined as $P_i = W_0$ if B_i is an orientation marker, otherwise $P_i = 1$. Finally, D is the number of orientation markers that are not the closest point to any centroid.

We assume there is a higher probability that binding sites lie closer to larger clusters; hence, the cost penalty is larger if the template is misaligned from larger clusters. In addition, the orientation markers are very important for readout, so these points in the template are weighted at W_0>1, and any missing orientation markers are penalized with the $D + 1$ term. Besides these additions, the cost function is a squared nearest-neighbor distance metric.

The transform parameters include rotation by an angle θ, X or horizontal scaling, Y or vertical scaling, X translation, and Y translation, which are applied in that order. The rotation prior to scaling allows the transformed templates to shear in response to distorted origami. At each iteration of optimization, the current transform is applied to the template, then the cost is calculated.

C is minimized according to any optimization method, and the final transform is applied to the template. For the "NSF" dataset, the template was first aligned to the origami by translating and rotating the template at fixed intervals around the origami until the minimal distance was achieved. The rough transform was then fine-tuned using the L-BFGS-B method of the minimize function from Scipy.[62] For the "ASU" 1-redundancy and 2-redundancy datasets, the transform was directly optimized using the differential evolution method of the minimize function from Scipy. For both datasets, at each iteration of the Scipy algorithm, all transformation parameters were simultaneously optimized and applied in the order specified above. Parameter selection was done by empirically finding good values for N and M. If origami with a larger number of binding sites is imaged, a higher value of M may be needed to account for false positives. T_I and T_S were selected through grid search. W_O was empirically determined. The best method for alignment for each dataset was chosen empirically. See Table S7 for the parameters.

Finally, the closest points in the template to each centroid were converted into a binary string using the knowledge of the template. In the case of repetition data sets, if any of the binding sites corresponding to a single bit were activated, then the bit was set to 1. The decimal value of this binary string could then be converted using the alphabet encoding.

3D Template Alignment. Using the same general idea in 2D template alignment, Euclidean coordinates of centroids and the 3D cuboctahedron template were used to minimize the Euclidean squared distances of each centroid to the closest point on the template using the formula:

$$C = \sum_{i=1}^{n} (|A_i - B_i|)^2 \qquad [2]$$

Initially, the 3D centroids and the 3D template were brought close by aligning the center of mass of the 3D centroids and the 3D template. The template was then z-scaled to fix the z-scale discrepancy due to centroid assignment by K-means clustering. Then, the minimization was done through a series of steps of x and y translation, then rotation about the z-axis with a total of 360°. Z translation could also be performed, but it would cause the running time to be longer. The z translation did not significantly affect the alignment result due to the small difference in the z value of the template and pattern center of mass. In all steps, the C was calculated and the minimum value was taken for the final transformation. Following the alignment process, origamis that lacked at least one orientation marker were discarded, and the remaining results were analyzed. The process of converting binary was carried out in a similar manner to the 2D method, utilizing aligned patterns and the understanding of binary translation employed in the rules for pattern encryption.

oxDNA simulation and 3D alignment of the unrelaxed and mean structure with experimental DNA-PAINT data. The simulation was carried out at 25°C in a periodic boundary cube of length around 168 nm using oxDNA2 forcefield and Langevin thermostat, for 1.52×10^{-5} s. An average of 6 such replicas were considered for the study, accounting for a total simulation time of 9×10^{-5} s.

The DNA PAINT data was projected onto the 2D $x-y$ plane and compared with simulation results to determine the best possible 2D configuration. The optimal 2D configuration conserves the z-axis and is subsequently scaled with a small factor to obtain a desirable structure. To obtain the best 2D configuration, the average configuration from the simulation was rotated such that the plane formed by the biotin strands was at the bottom of the frame and the normal of the plane was perpendicular to the xy plane, similar to the experimental setup. The center of mass of the 12 docking handles was projected to the $x-y$ plane and fitted with the projection of the PAINT data points. Due to the geometry of the origami, when the structure is projected onto an $x-y$ plane, the 4 points at the extreme top along the z-axis and 4 at the very bottom overlap, making them indistinguishable from the clustering algorithm, causing them to look only as 8 clusters instead of 12. Using the K-means algorithm, the projected data points from the DNA-PAINT were clustered into 8 points instead of 12. The two configurations were fitted using the SVD

superimposition technique (Fig. 6B for mean structure and Fig. S13D for unrelaxed structure). The 2D configuration generated in this step remained unchanged throughout the later procedures.

With 2D alignment being fixed, the z-axis of PAINT data was increased by a small factor and clustered into 12 clusters. The sum of the RMSD between the centroids and the mean position of the docking handles was minimized for an optimal z-axis scaling factor (Fig. S13E). To further confirm that the K-means produced meaningful clusters, the 12 centroids generated were again projected into the 2D plane, and the total RMSD between the current configuration and the old 2D configuration was noted. If the resultant RMSD exceeded the uncertainty, which was taken to be the RMSD from SVD 2D fitting from the previous step, the model was rejected. These results were visually confirmed as well to validate our findings. The K-means algorithm for the 3D fitting has an accuracy of 84%, which was predicted by executing the algorithm multiple times with random seeds and checking the projection 2D RMSD generated after 3D clustering (Fig. S13E). With the final check of 2D RMSD, the algorithm was successfully able to figure out the scaling factor within a deviation of 6%.

Code Availability The scripts and code for processing the data can be found in https://github.com/Jonathanzhao02/smlm_classification2d and https://github.com/gwisna