STORAGE DEVELOPER CONFERENCE
SDC 22
Fremont, CA
September 12-15, 2022
BY Developers FOR Developers

A SNIA Event

Nucleic Acid Memory

Super Resolution Microscopy Enhances Novel Approach to DNA Data Storage

Presented by : Luca Piantanida


Will Hughes


Chad Watson


William Clay


George Dickinson


Golam Mortuza


Ben Johnson


Mike Tobiason


Medhi Bandali


Eric Hayden


Wan Kuang


Tim Andersen


Elton Graugnard


Natalya Hallstrom


Chris Green


Sarah Kobernat

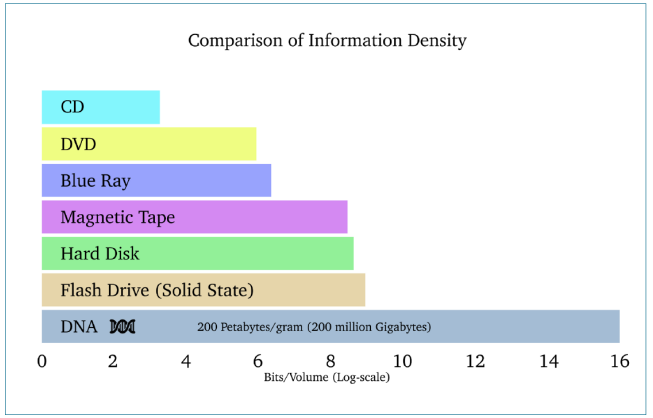

Reza Zadegan

Why store digital information using DNA?

Longevity

Hundreds to millions of years of information retention

Data density

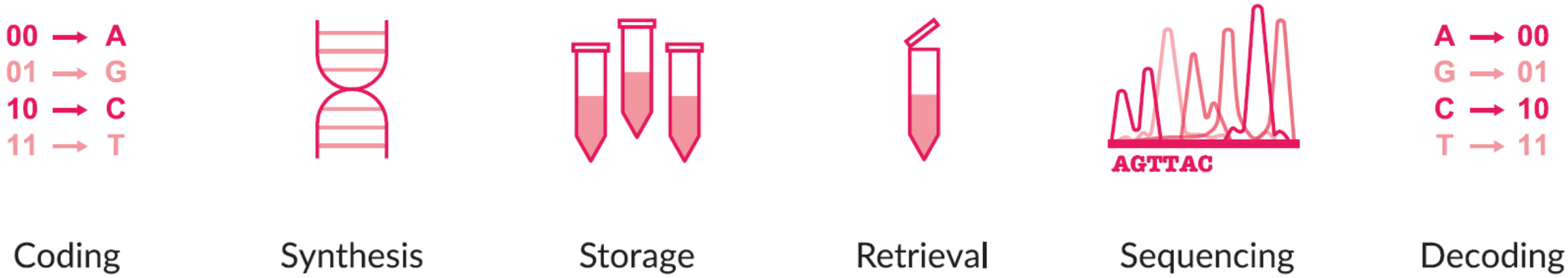


All the worldwide data stored in one room (predicted)

Promising material for storing large quantity of information for a long period of time

The future of DNA storage, **Potomac Institute for Policy Studies**, 2018.
Zhirnov, V. et al., **Nature Materials**, 2016.
Archiving big data beyond 2040: DNA as a candidate, **National Academy of France**, 2020.

DNA data storage fundamental steps



State of the art of DNA memory process

An introduction to DNA data storage. DNA data storage alliance (2021)

DNA data storage involved technologies

Encoding

Message

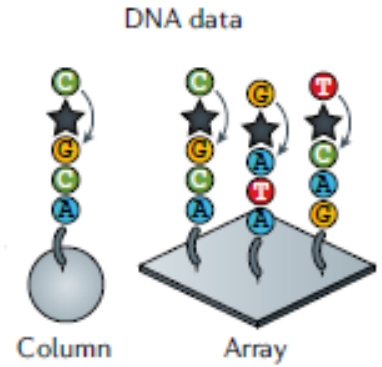
01001110
01000001
01001101

Digital DATA

Home developed algorithms

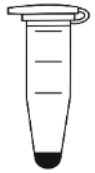
Writing

ATACGTT…



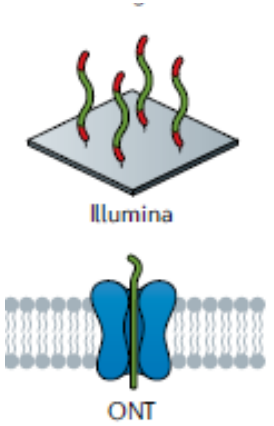
Synthetic DNA synthesis

Storage



In fluid or dried

Reading

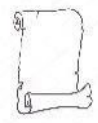


DNA sequencing:
NGS
Nanopore

Decoding

Digital DATA

01001110
01000001
01001101



Message

Ceze, L., et al. **Molecular digital data storage using DNA.** Nat Rev Genetics (2019)

digital Nucleic Acid Memory (dNAM)

DNA data storage system with an optical readout with no DNA sequencing required

Dickinson, G.D., et al. **An alternative approach to nucleic acid memory.** Nature Communications (2021)

Hughes, W. et al., US Patent, 17/443,312, “NUCLEIC ACID MEMORY (NAM) / DIGITAL NUCLEIC ACID MEMORY (DNAM)”, 2021.

How dNAM differs from other technologies

Encoding

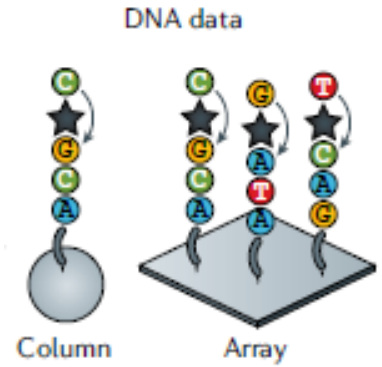
Message

01001110
01000001
01001101

Digital DATA

Writing

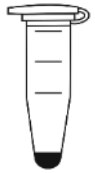
ATACGTT…



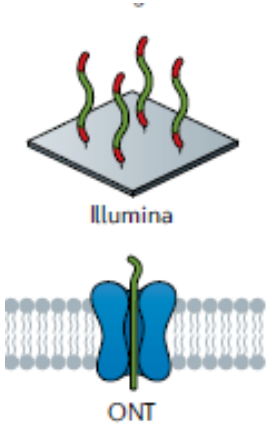
dNAM

DNA assembly as structural material

Storage



Reading

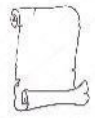


dNAM

DNA-PAINT TIRF microscopy

Digital DATA

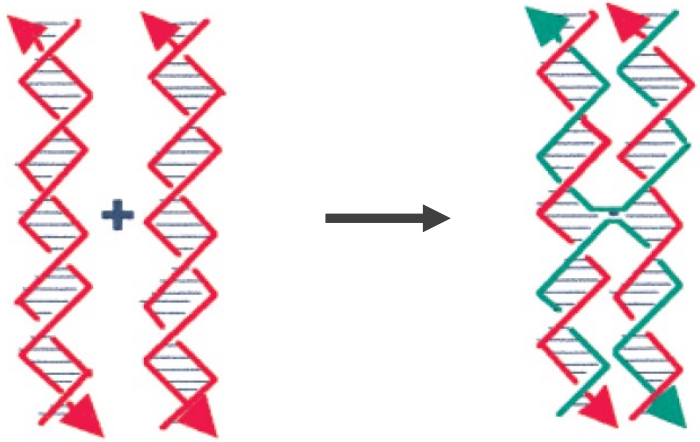
01001110
01000001
01001101



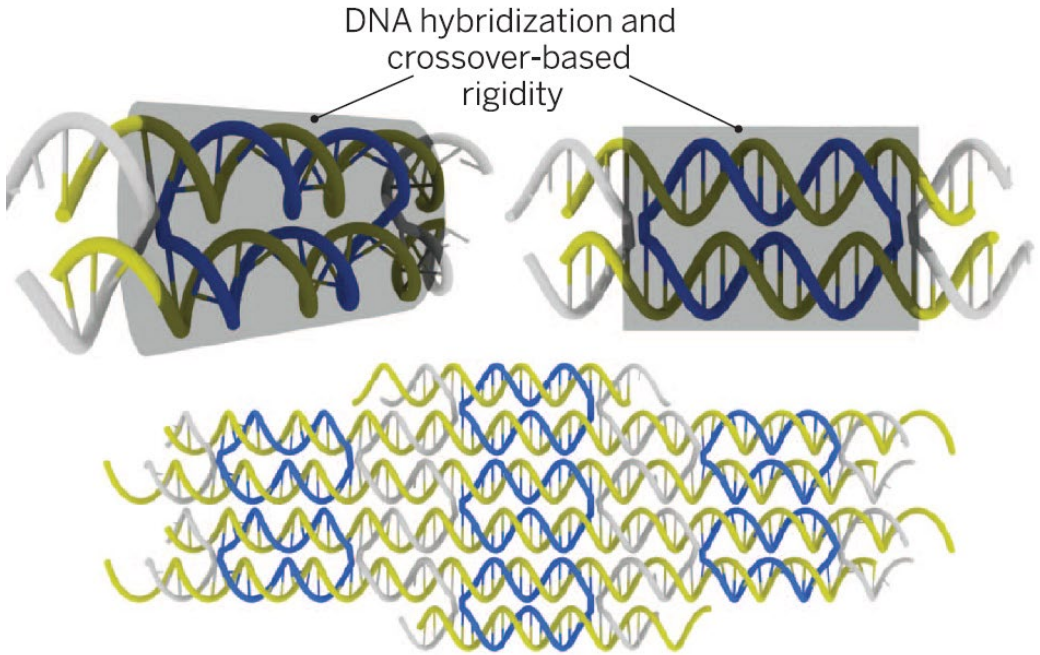
Message

Ceze, L., et al. **Molecular digital data storage using DNA.** Nat Rev Genetics (2019)

DNA as a programmable material



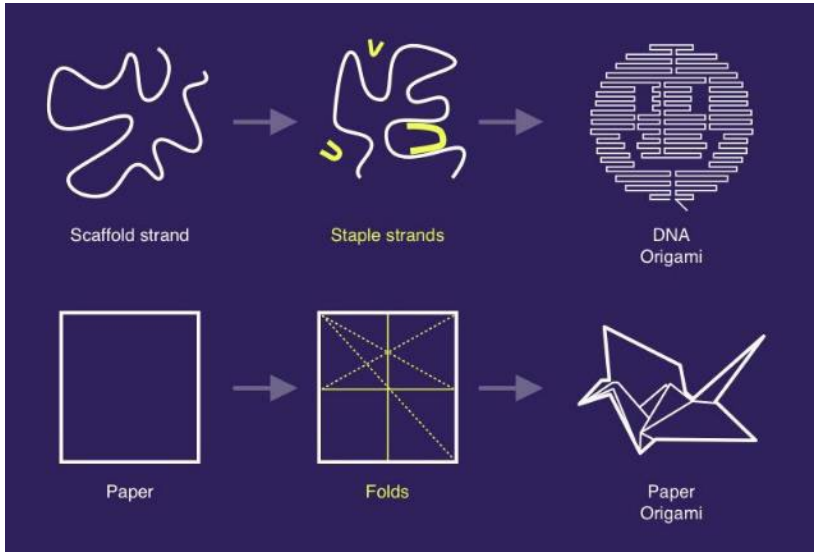
DNA bonds can be programmed at the nanoscale

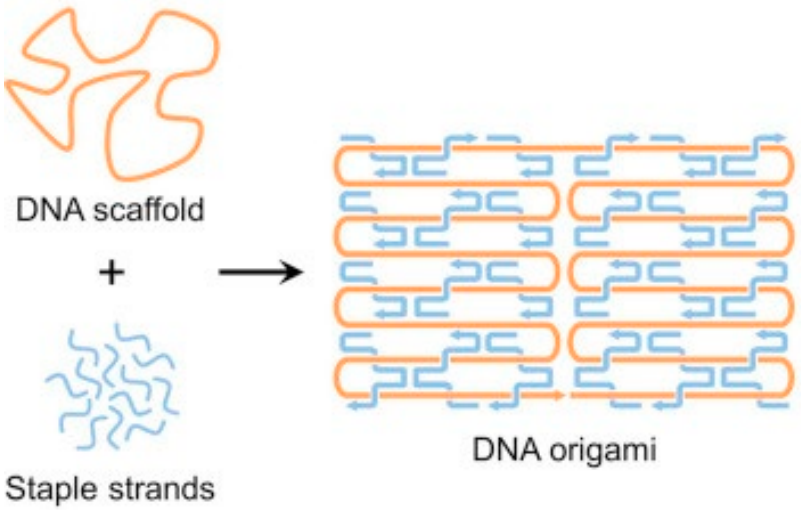


…allowing formation of fairly rigid and nanostructures

Li X., et al., JACS (1996)
Seeman N. C., Nano Letters (2001)
Jones M. R., et al., Science (2015)

DNA origami technology (writing)

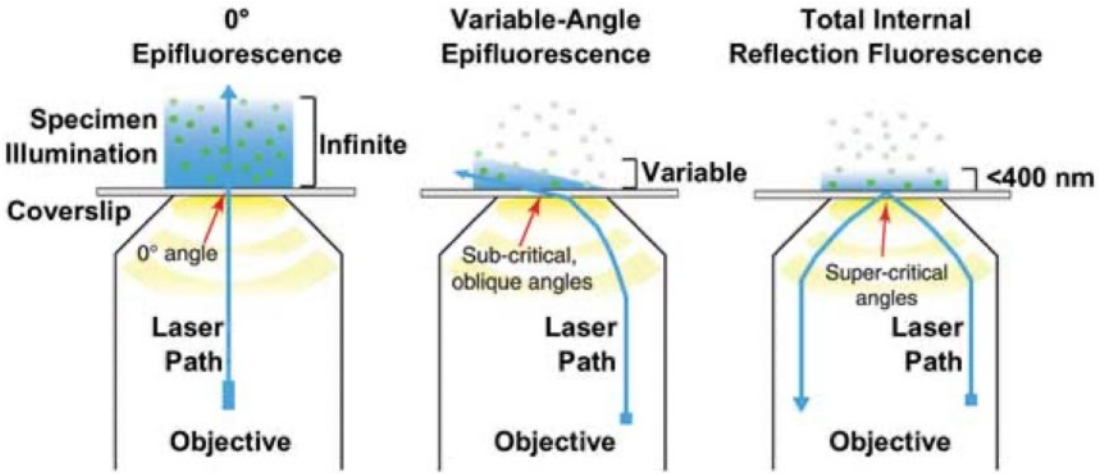


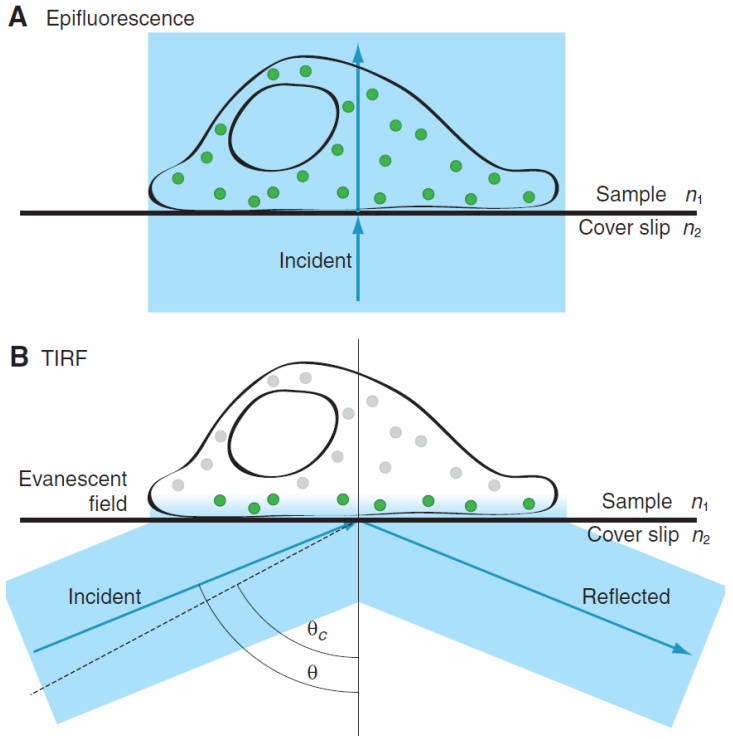


DNA as a structural and scaffolding material
highly programmable
with nanometer precision

Rothemund, P., Nature (2006).

TIRF (Total Internal Reflection Fluorescence) Microscopy

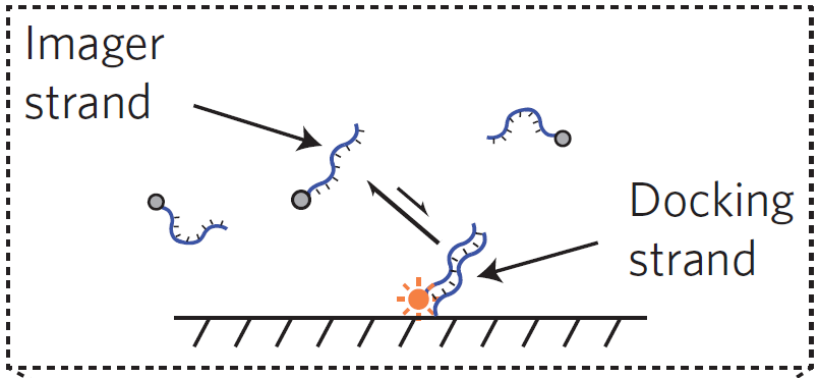




TIRF is a upgraded version of Fluorescence Microscopy that is able to cancel all the background from the bulk solution

Mattheyses. A. L., et al., J. Cell Science (2010)

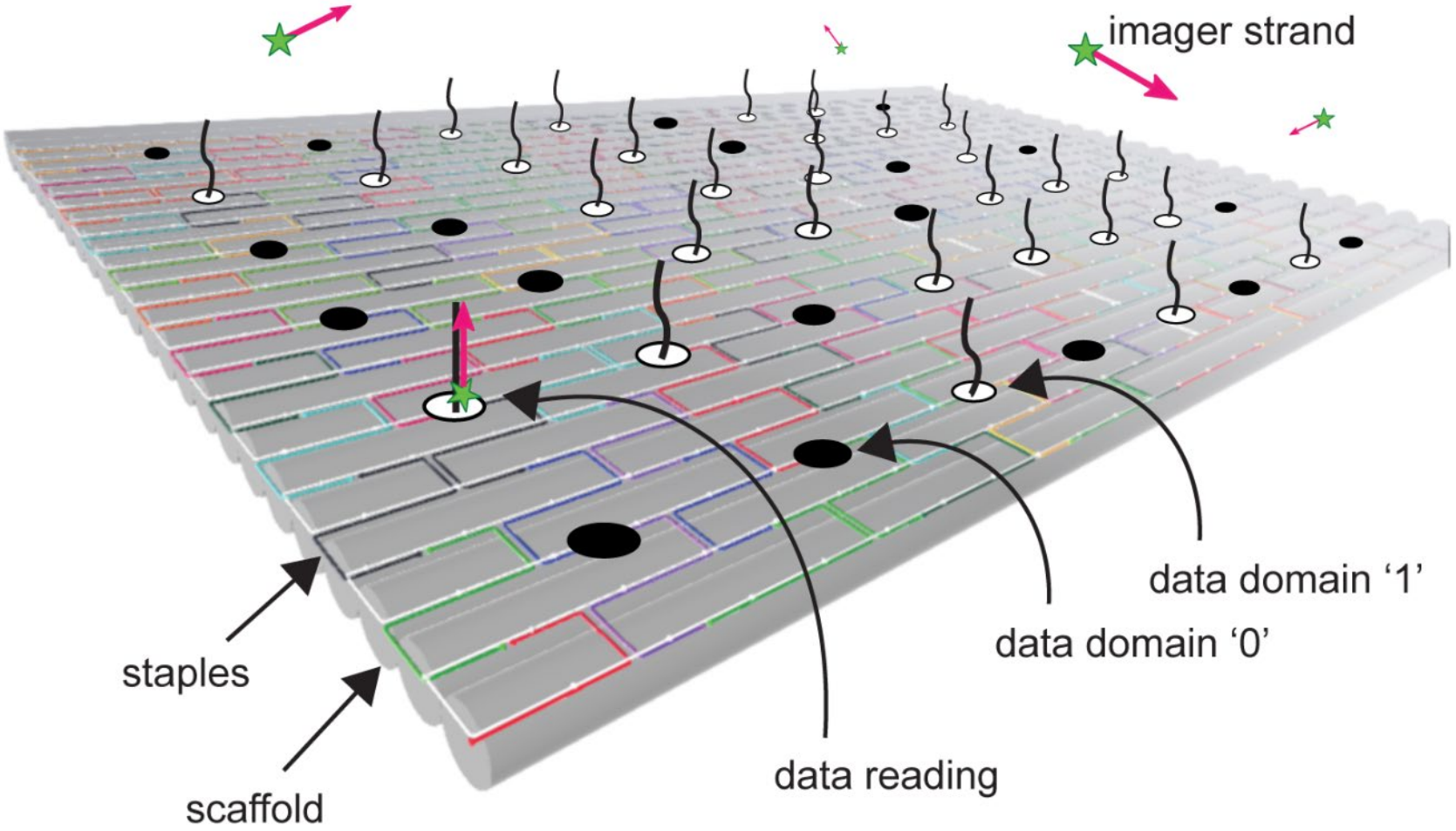
DNA-PAINT technology (reading)

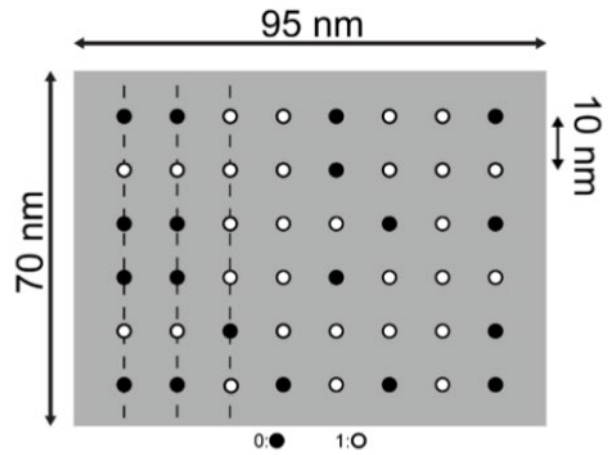


Transient hybridization binding between a dye-labelled DNA strand to its complementary attached to the surface

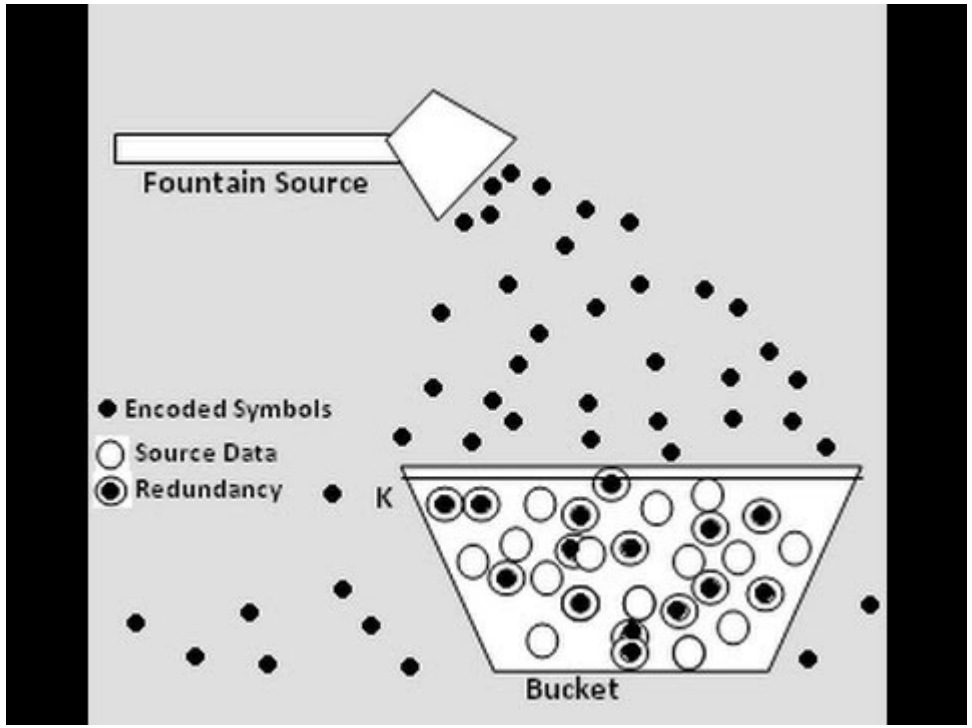
Schnitzbauer, J., et al., Nat Protocols (2017).

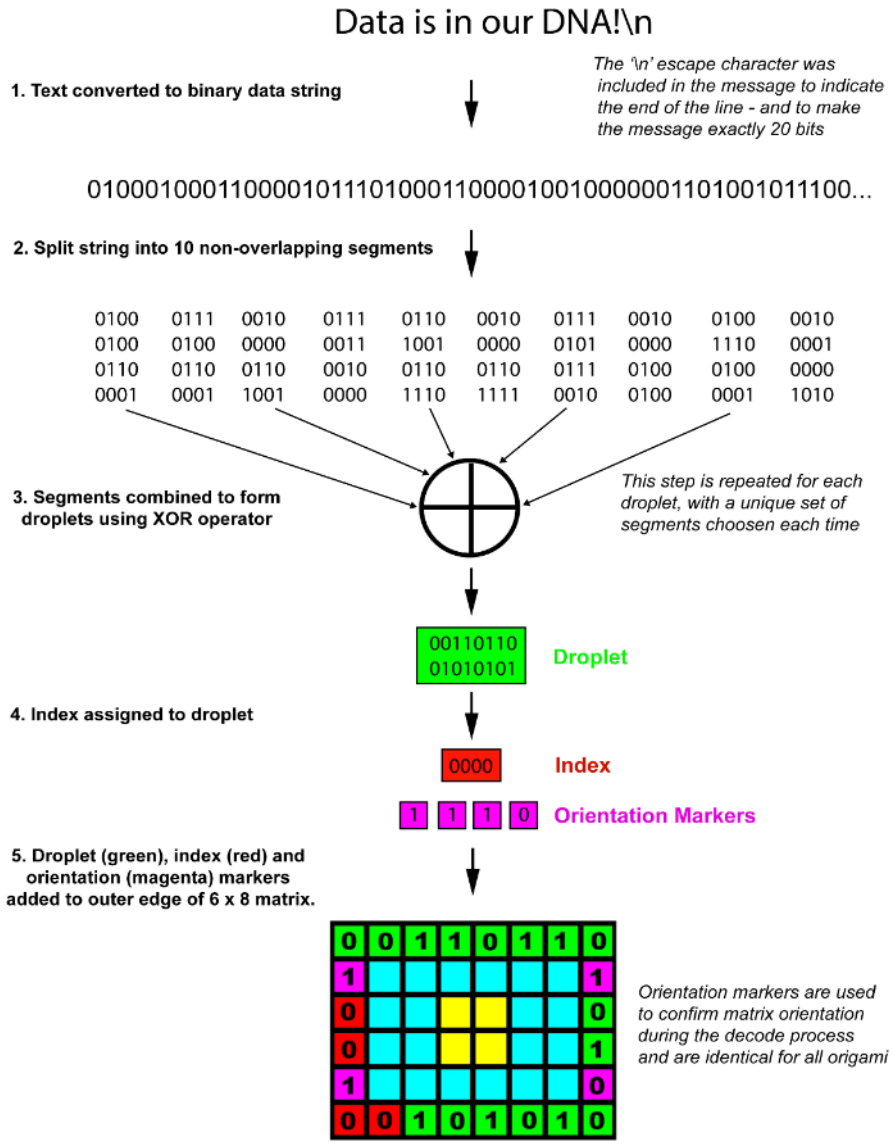
dNAM platform



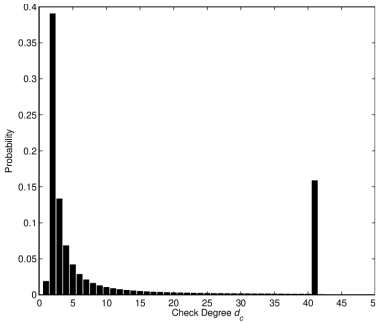


Encoding and storing information using a Fountain Code





Message converted into a binary string (20 bytes)



Divided in 10 non-overlapping segments

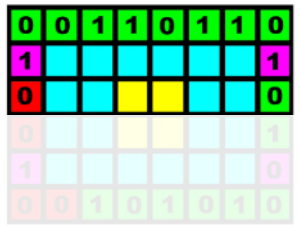
-Soliton distribution (1 to 10 XORed segments)

-Uniform distribution (distribute segments in droplets)

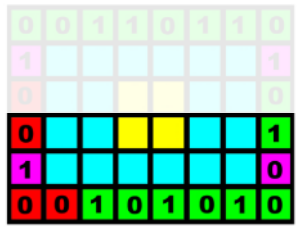
Trough XOR operator 15 “droplets”, with one or more data segments in it, are generated (5 droplets more than the # segments to have enough redundancy)

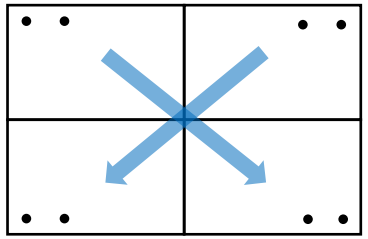
Checksum and parity bits are calculated with XOR operator following a 180º rotational symmetry:

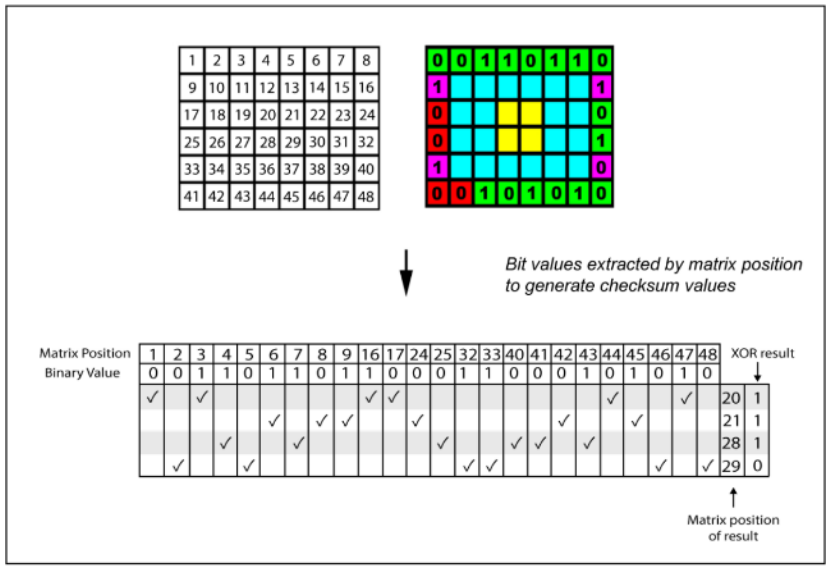
-matrix orientation is retrievable without having orientation markers in the correct order



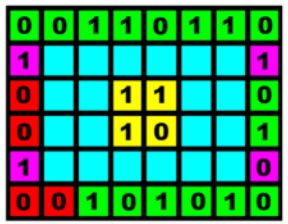
180º







7. Checksum bits added to center of matrix (yellow)



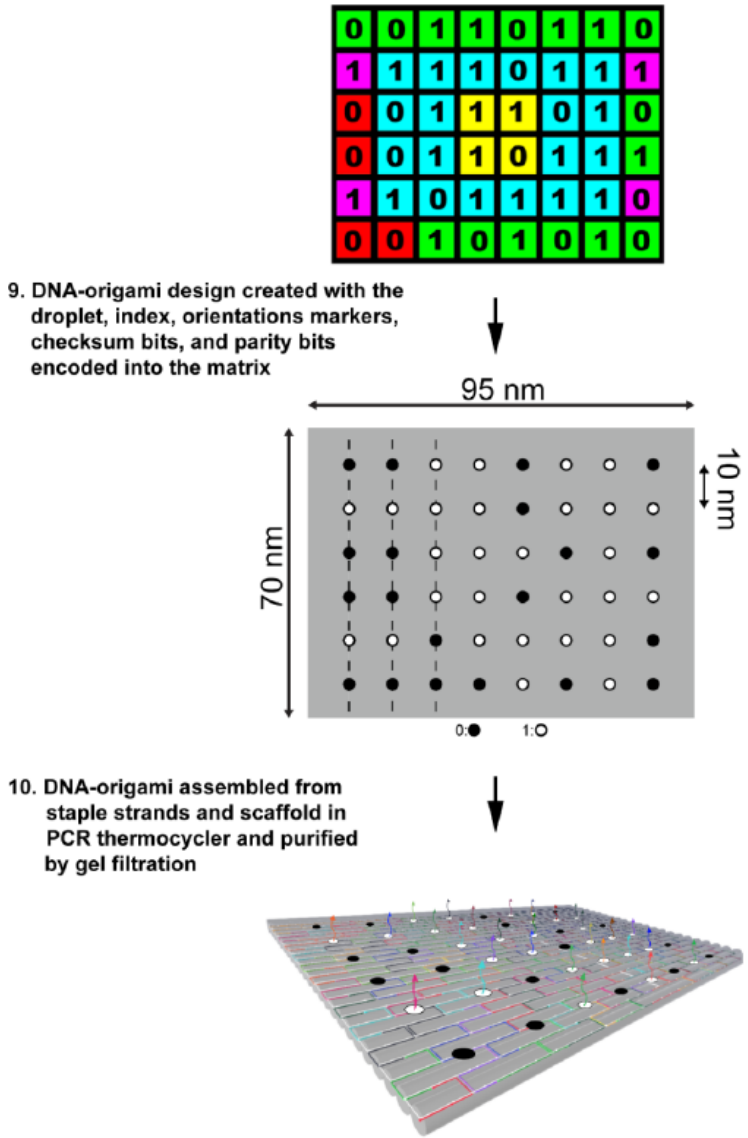
9. Parity bits added to remaining matrix positions (blue)

Checksum bits calculated from symmetrically positioned matrix edge values

And added to center of matrix (yellow)

Parity bits calculated from symmetrical positions and checksum and added to the matrix (blue)

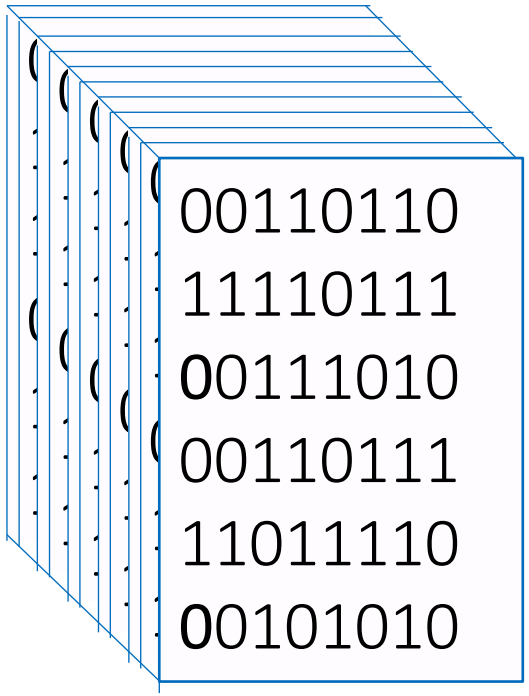
Encoding → Writing → Storage → Reading → Decoding



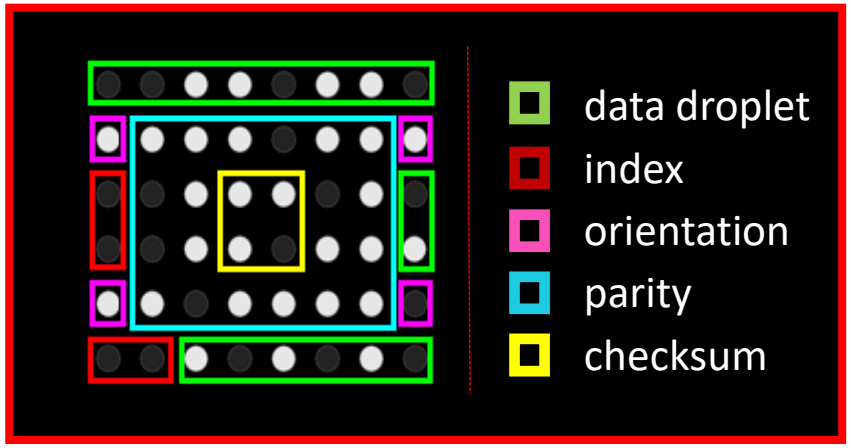
0 bit = no signal (dark spot)

1 bit = signal (bright spot)

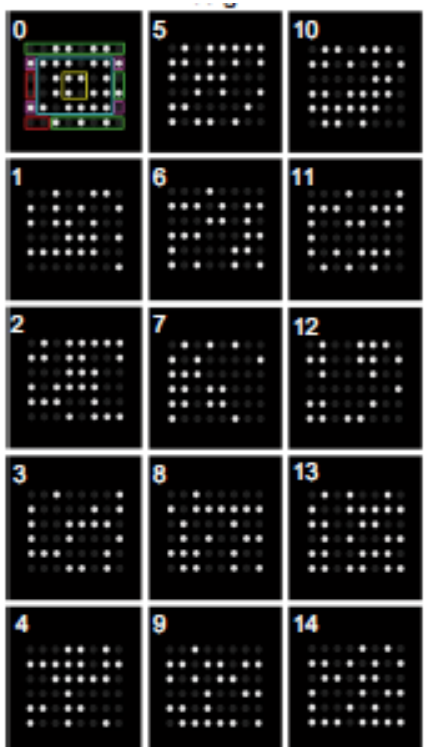
Encoding → Writing → Storage → Reading → Decoding



Every data droplet contains markers for:



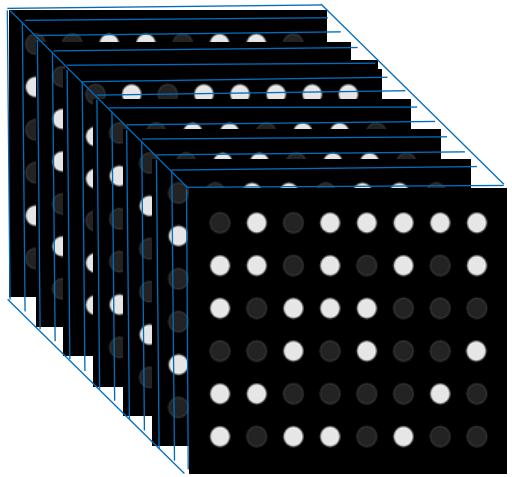
Matrices designs

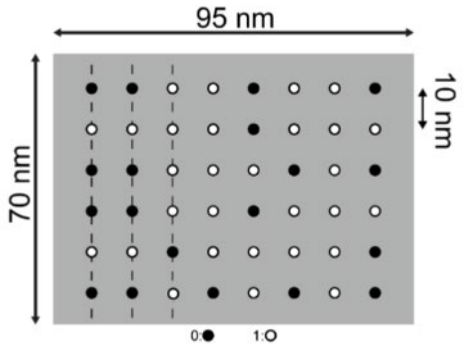


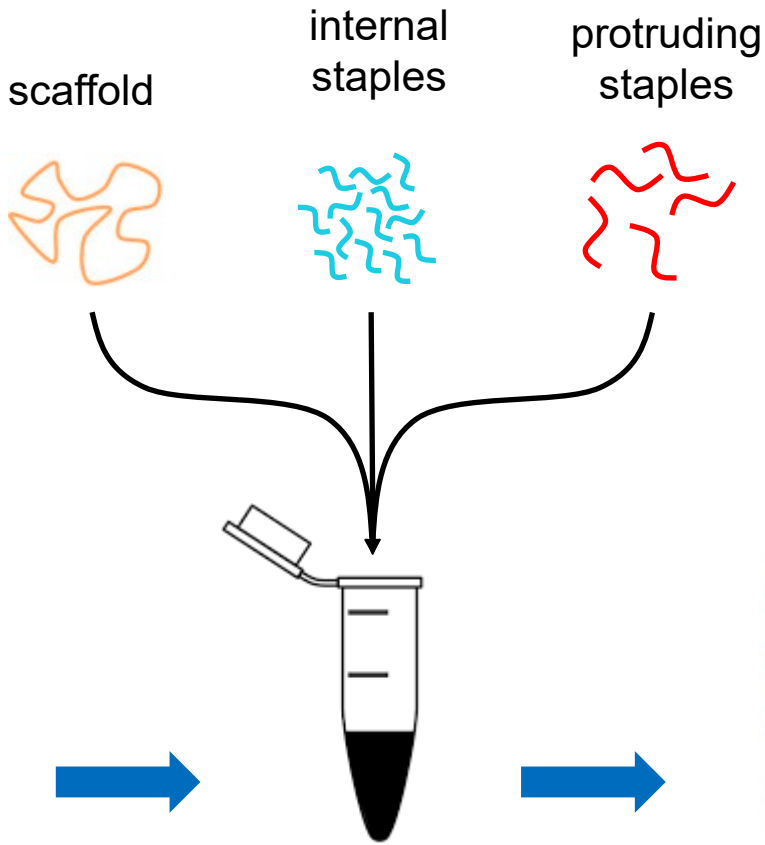
Data is in our DNA!/n → The message is converted in 15 different matrices designs

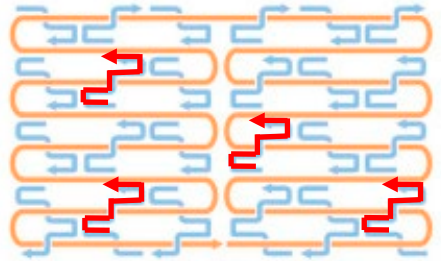
Encoding → **Writing** → Storage → Reading → Decoding

15 matrices designs



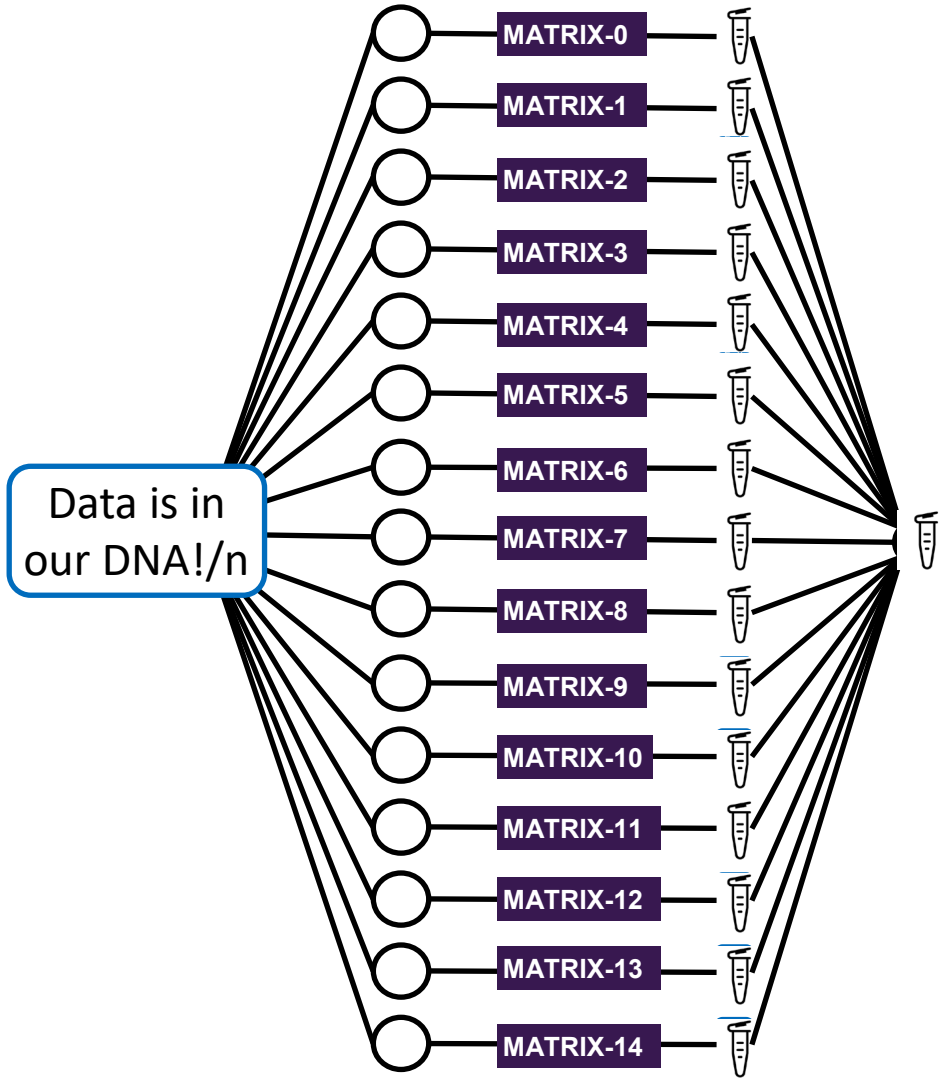






15 origami platforms with different matrices pattern

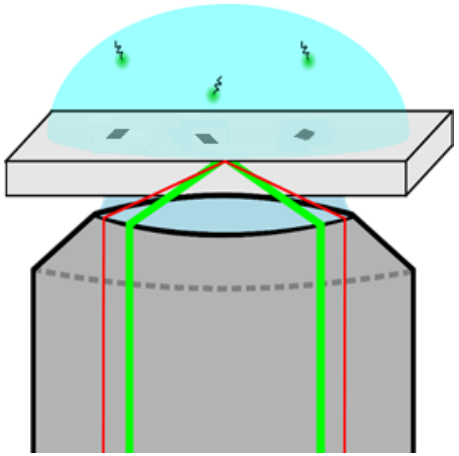
Encoding → **Writing** → **Storage** → Reading → Decoding

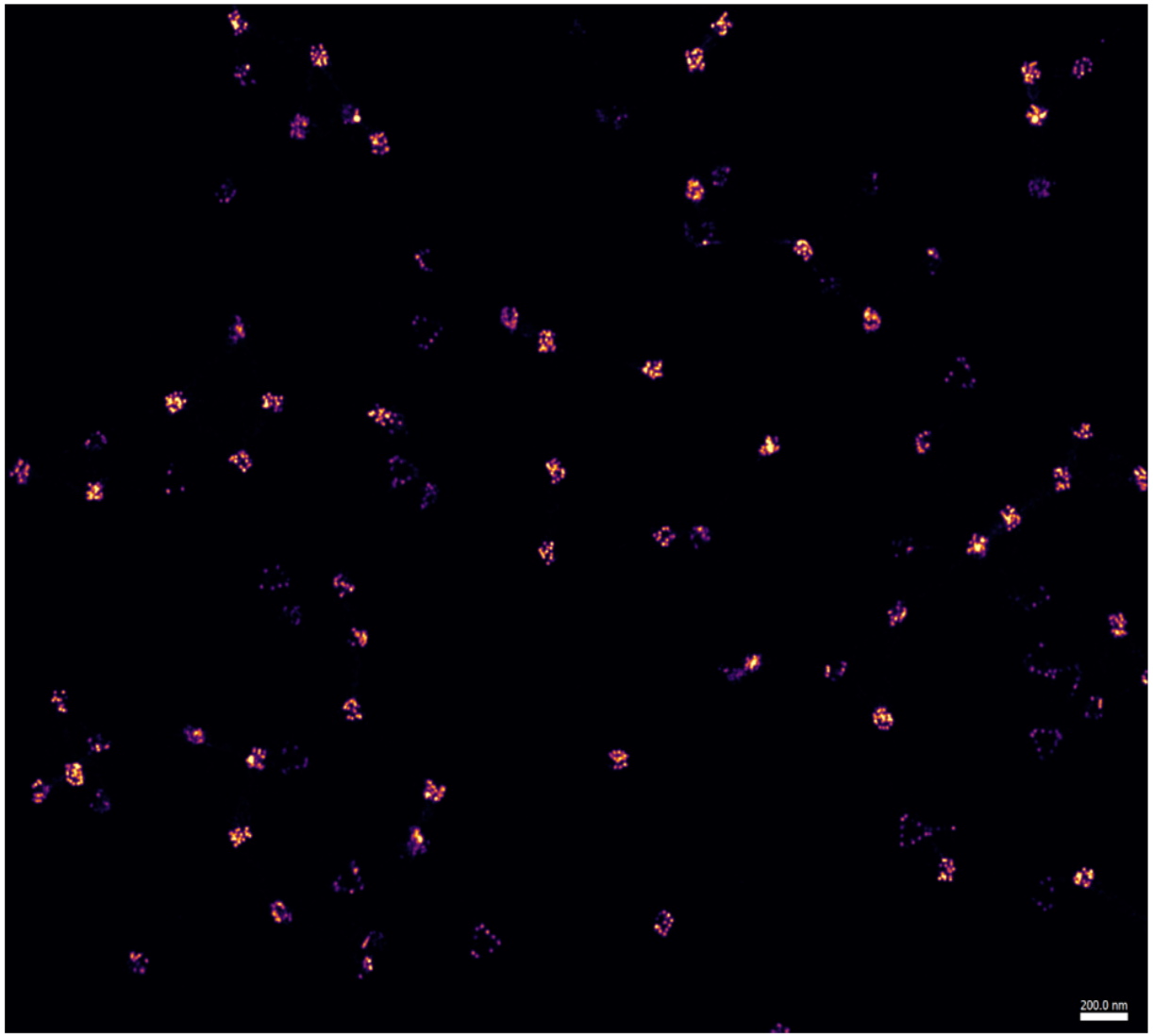


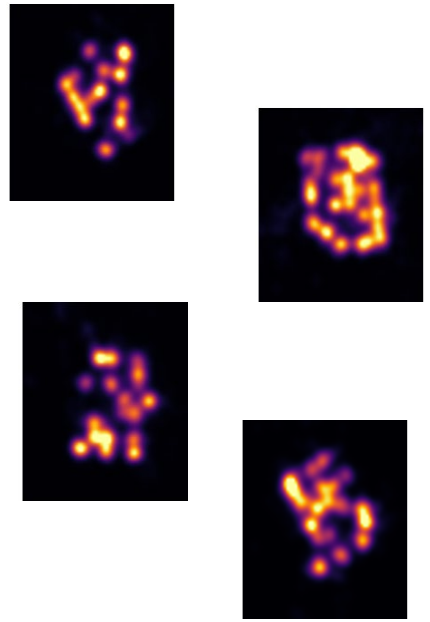
The message can now be stored as whole or in separated matrices designs

Encoding → **Writing** → **Storage** → **Reading** → Decoding

DNA PAINT TIRF Microscopy

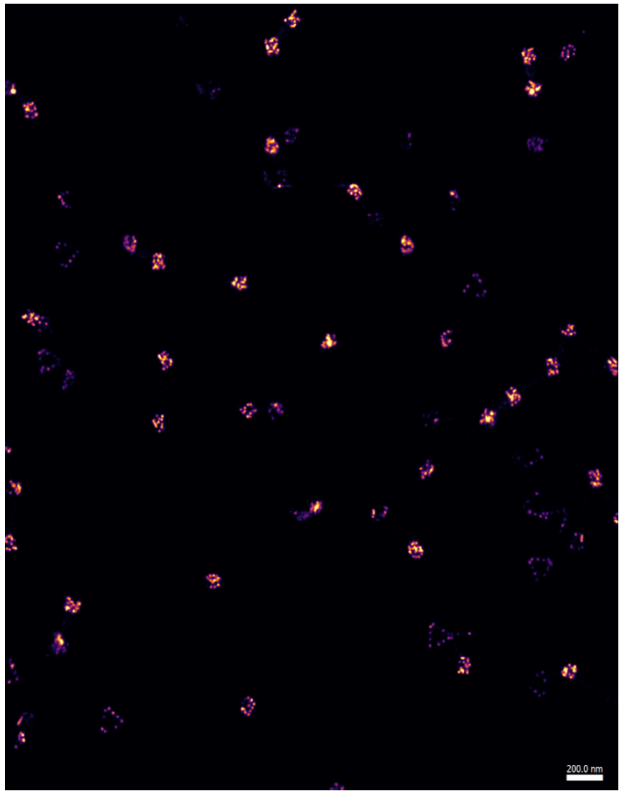






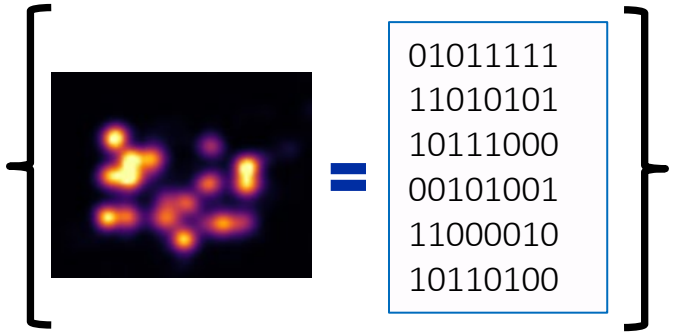
All origami are picked and super resolved

Encoding → **Writing** → **Storage** → **Reading** → **Decoding**

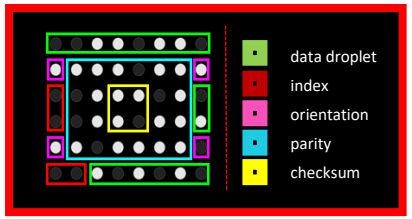


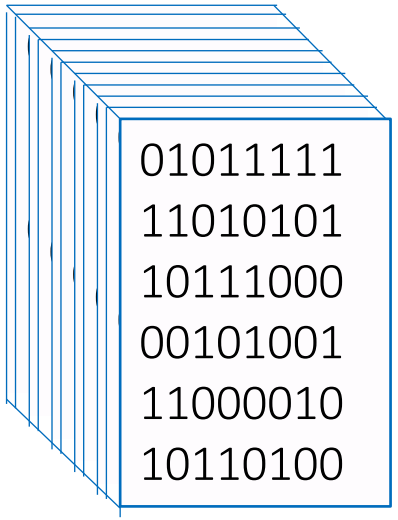
One matrix

One binary string



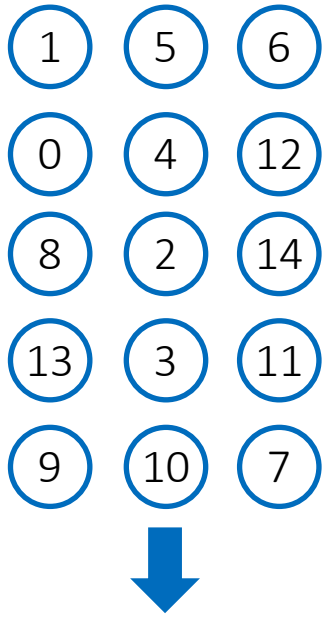
Decoding





Multiple binary strings decoded

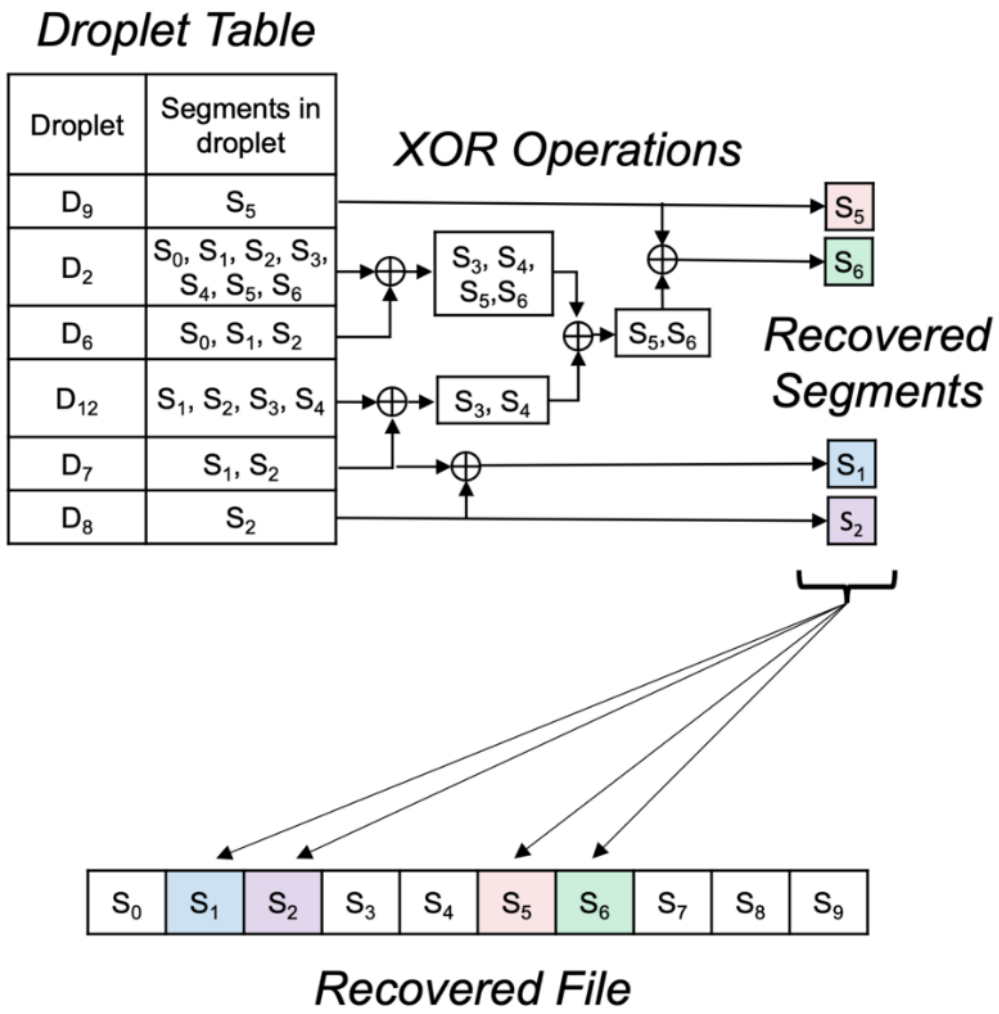
Every retrieved string is a code's droplet



Data is in our DNA!/n

The message is completely decoded

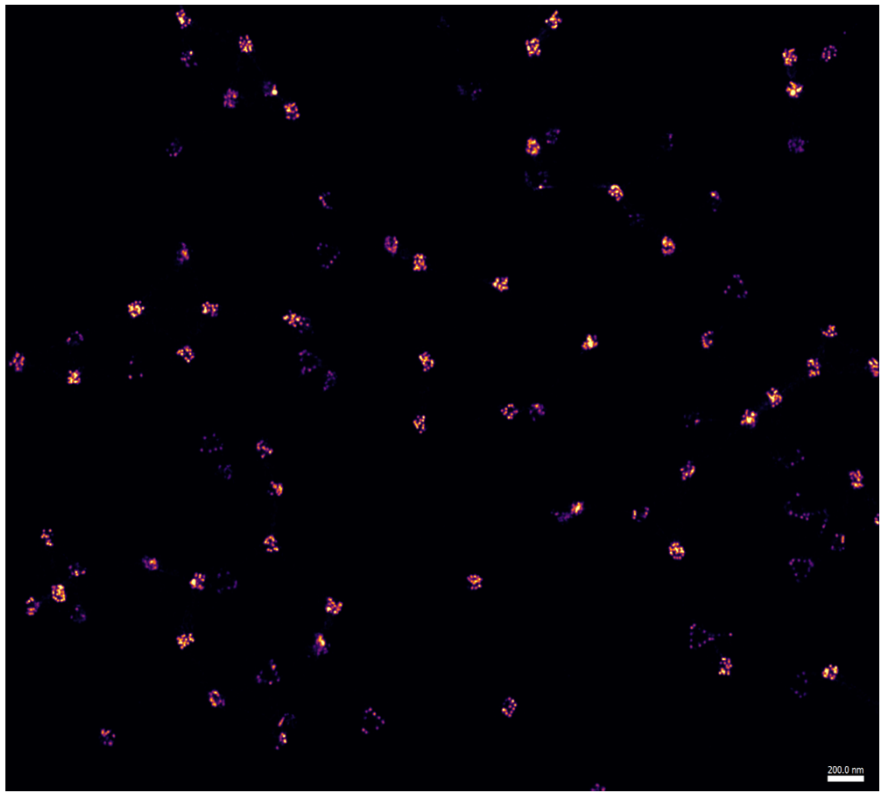
Decoding

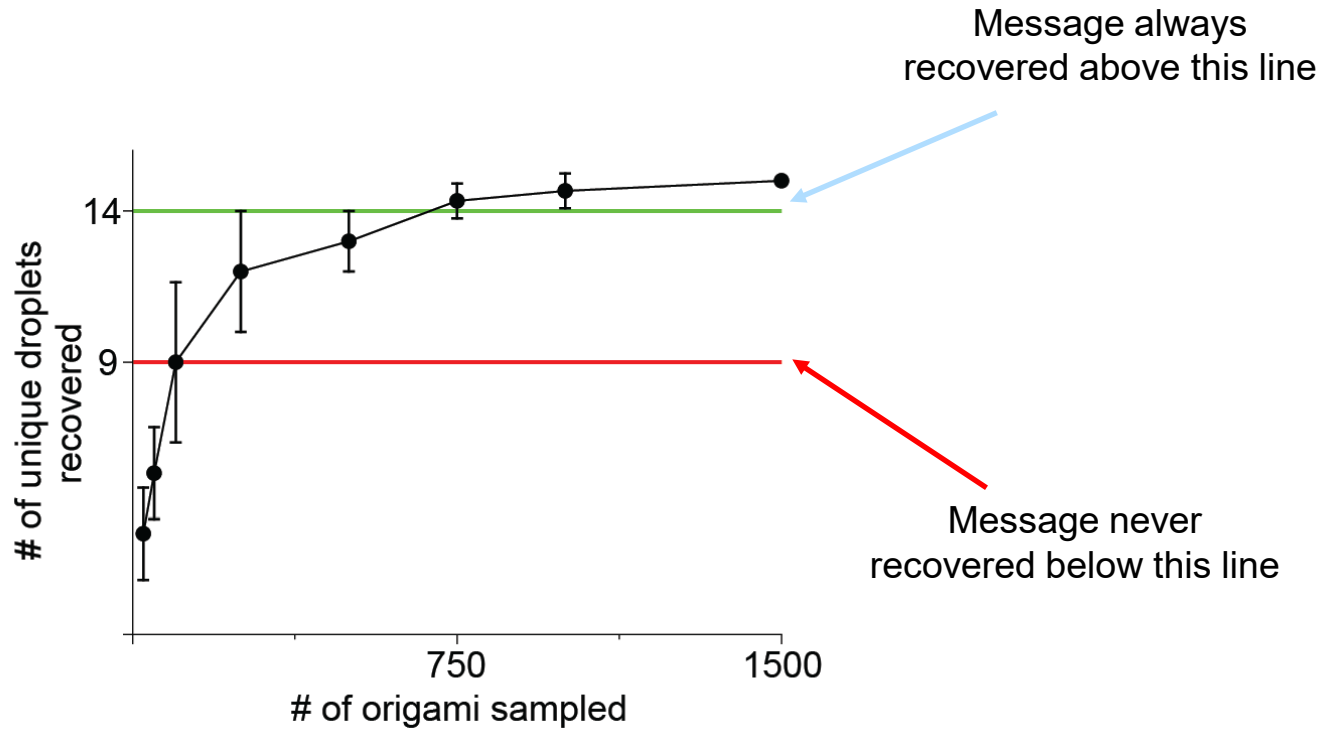


High redundancy of data segments in each droplet ensure the recovering of the file

Sampling

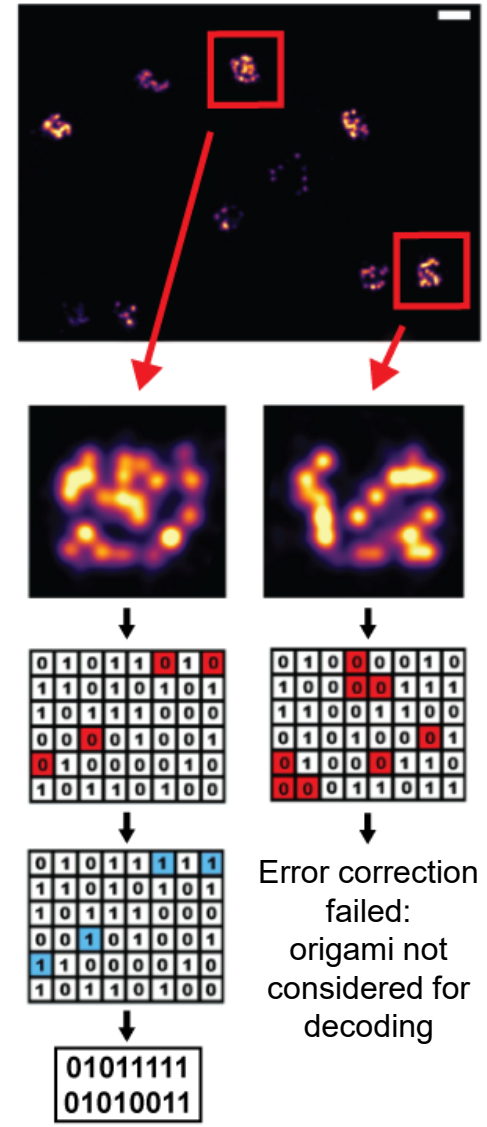
How much sampling I need to recover the complete message?

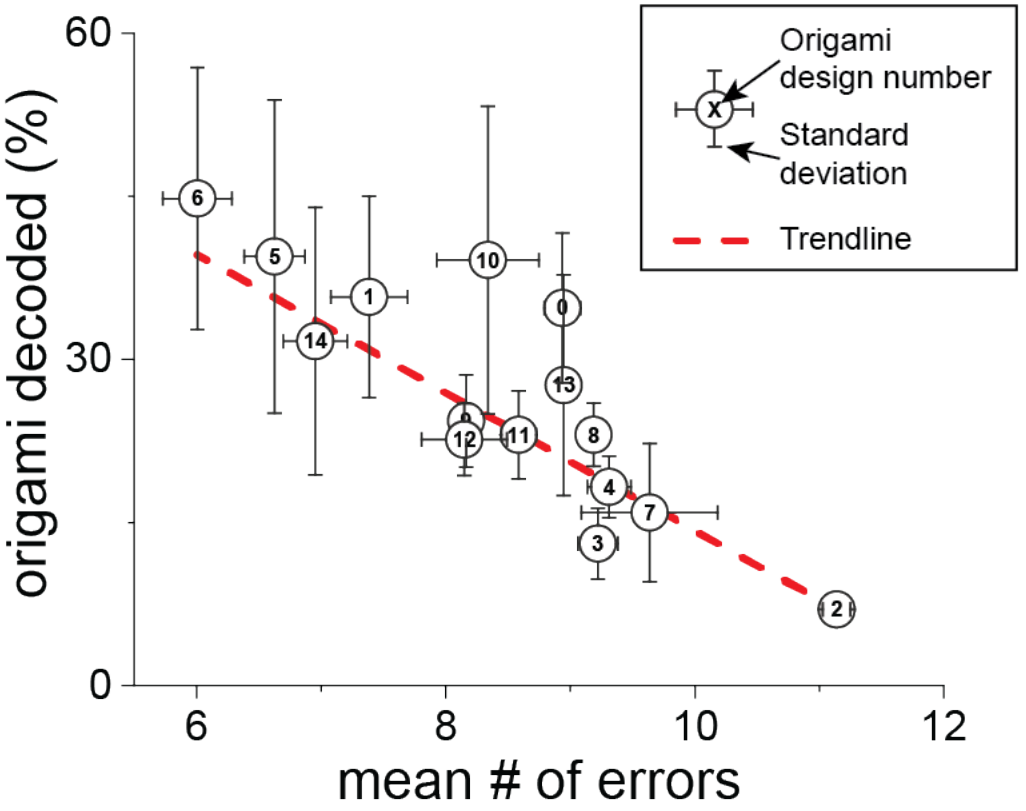




40% redundancy is needed to always recover the message but increasing file size this will drastically reduced

Errors correction





Designs are differentially error prone

Our decoding algorithm performs, as expected, less well as errors increase

dNAM as a novel prototype for storing digital information in DNA

Strengths :

- The technique does not require the synthesis or sequencing of custom DNA strands.
- A discrete library of sequences can encode arbitrary messages.
- Fountain codes and error corrections algorithms ensure 100% reading accuracy.

Weaknesses :

- The read times for SRM are inherently slow.
- Data density is low compared to other DNA-based data storage methods

dNAM is more suitable for archival applications than real-time access and due to its data redundancy and high copy number is a promising system for bar-coding, encryption and long-term storage applications.

Data density calculations

DNA Storage capacity (Zhirnov et al., 2016):

~$1x10^{19}$ bits/ cm^3 → [$1.25x10^{18}$ bytes/cm^3] [1250 PB/cm^3]

dNAM prototype:

[160 bits (20 bytes) / aliquot] [1 aliquot = 15 origami] [$6.25x10^{14}$ origami/cm^3]

then

$6.25x10^{14}$: 15 = $4.17x10^{13}$ aliquots

$4.17x10^{13}$ x 160 = $6.67x10^{15}$ bits/cm^3

$6.67x10^{15}$: 8 = $8.33x10^{14}$ bytes/cm^3 [833.3 TB/cm^3]

Aerial density:

Hard drive:

~1.1 TB / in^2 → [170.5 GB/cm^2]

Magnetic Tape:

~224 Gbit / in^2 → [34.7 GB/cm^2]

https://blocksandfiles.com/2020/06/29/fujifilm-400tb-magnetic-tape-cartridge-future/

Reading speed

NovaSeq 6000 sequencer (illumina):

~ 5 TB/day

This is for reads of long sequences, the pools of short oligonucleotides used in dNAM would be sequenced at a considerably slower rate as they require additional amplification or ligation steps.

dNAM prototype:

20 bytes / 3.3 h → 145 bytes/day [now doubled]

A combination of concentrated origami deposition, larger origami with tightly packed data domains, increased bit-depth, probe multiplexing, optimized binding kinetics and larger camera sensor could feasibly bring data collection to **Gigabytes/day** per microscope with current technology

Hard drive:

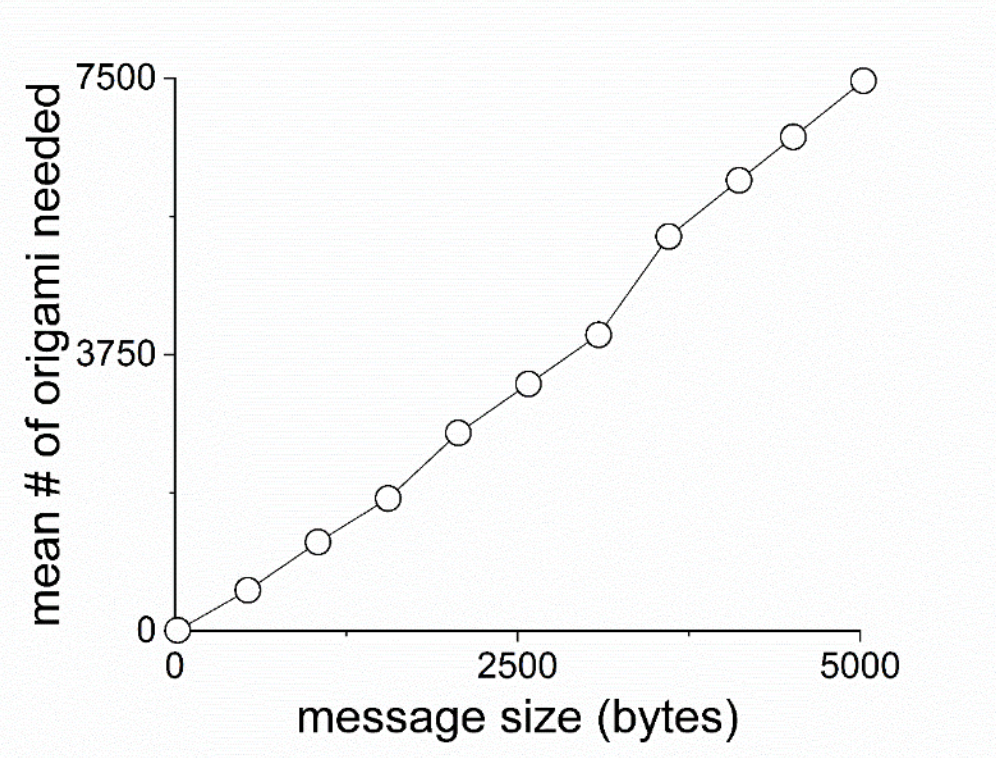
~80-160 MB/s

Magnetic Tape:

~300-800 MB/s

https://www.illumina.com/science/technology/next-generation-sequencing
https://www.ibm.com/docs/en/ts4500-tape-library?topic=performance-lto-specifications

Scalability



The simulation demonstrates a roughly linear increase in the number of matrices required up to 5000 bytes.

Given the scaling challenges, we are actively investigating different methods to increase the data density of dNAM!

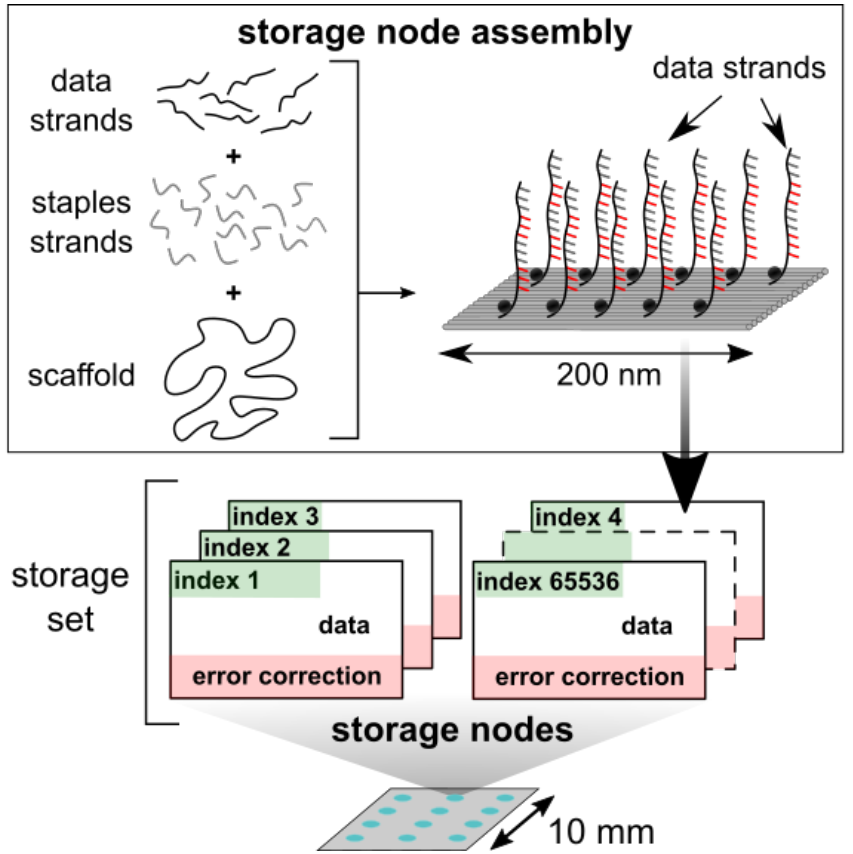
Increase data capacity

- Higher data density (closer data strands, multiplexing, etc.)
- Higher deposition density (more origami in a single field of view)
- Improved resolution (validation of a custom SRM system)
- Larger origami (more data and less algorithmic overhead)

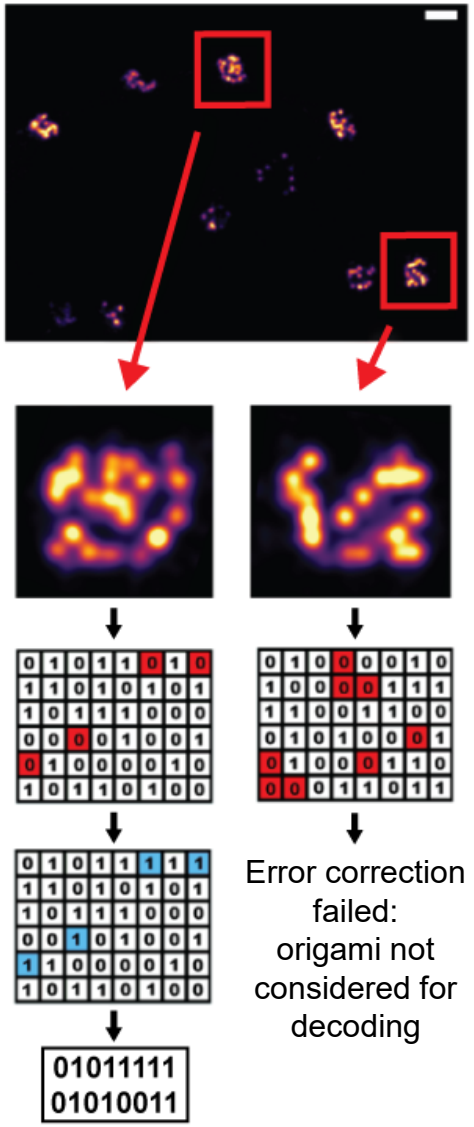
Future perspectives

We are working 3 main aspects of the device functionality:

- Increasing data density
- Decreasing errors probability
- Better reading automation/decoding

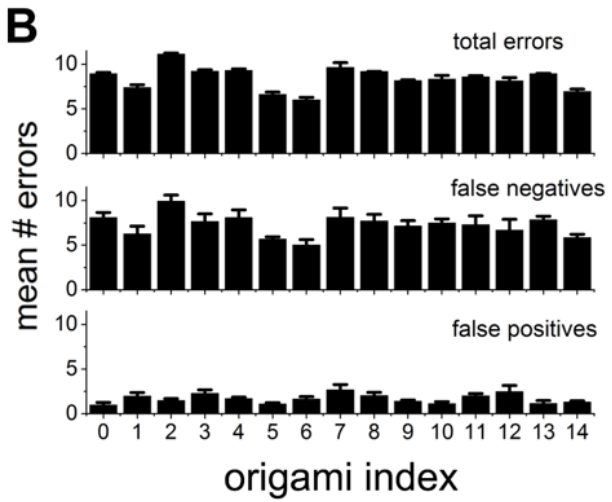


Errors correction



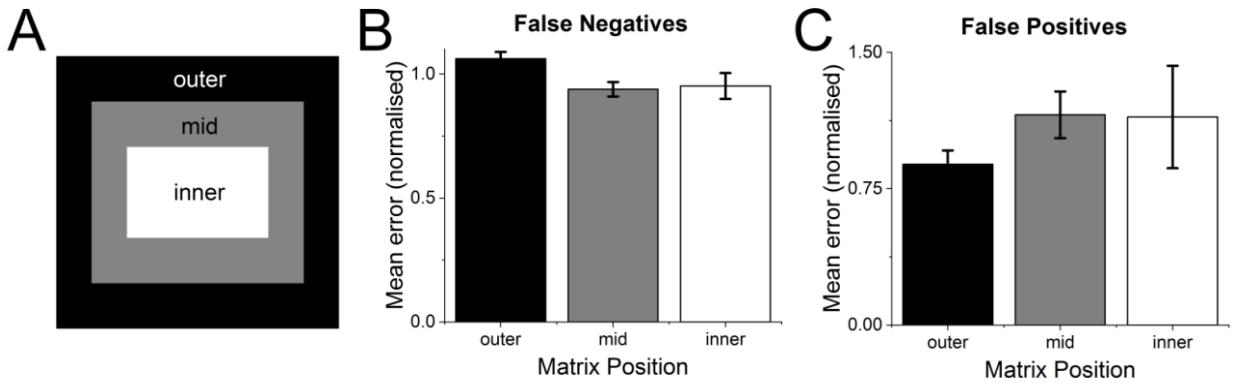
Droplet correctly decoded

Error correction failed: origami not considered for decoding

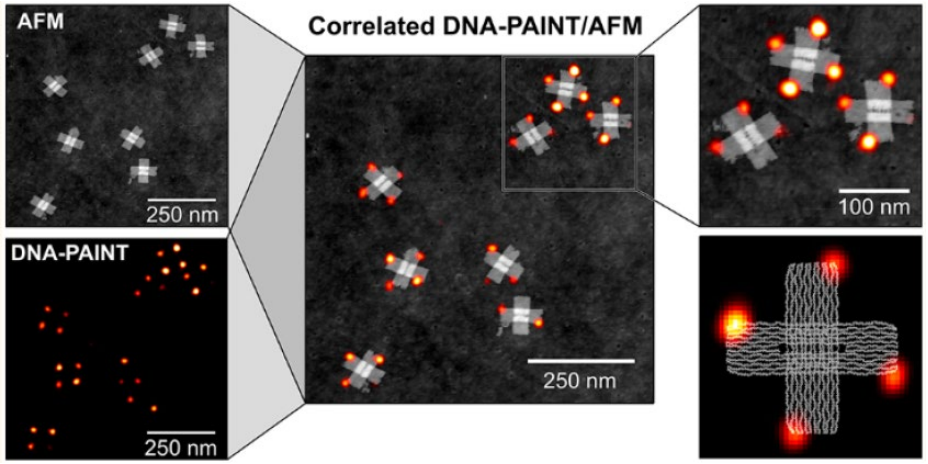


false negatives are occurring definitely more than false positives

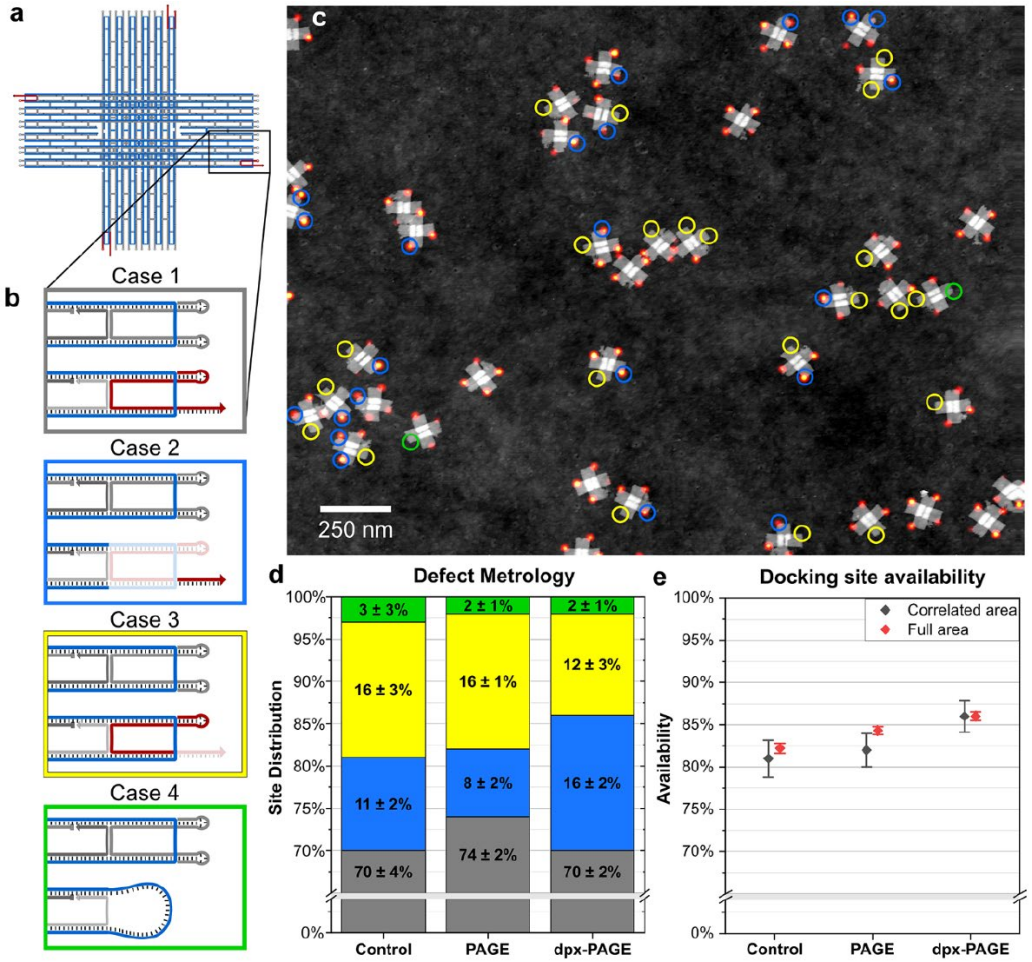
high variability along the matrices but no significant trend



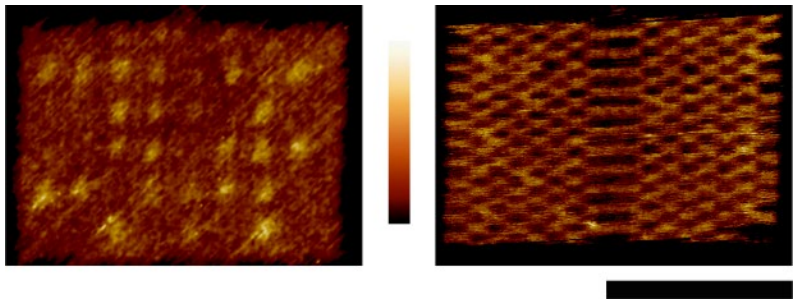
Metrology



Inspired by previous experience on correlating SRM errors to structural defects in the origami assembly



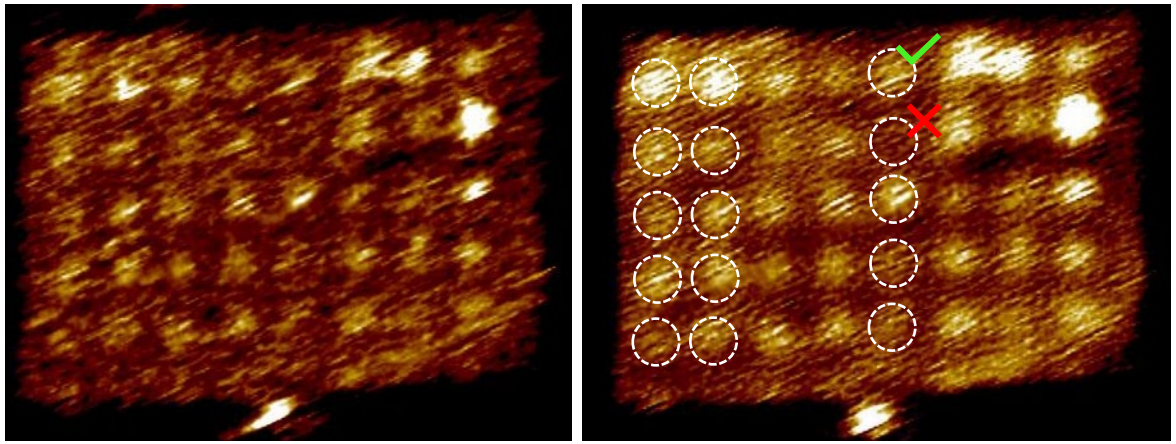
AFM analysis



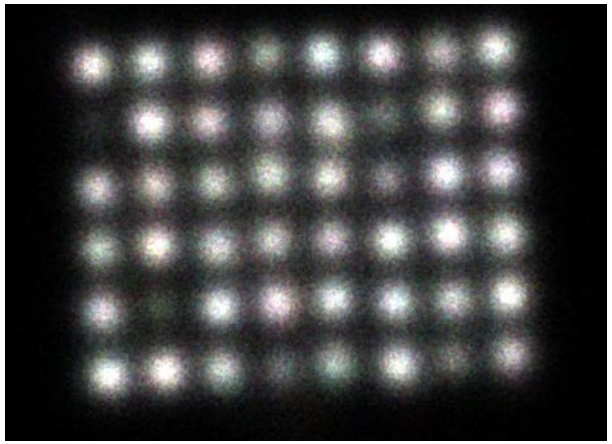
Soft-AFM to check protruding staples

HR-AFM to check integrity

The high copy number of the same DNA origami in one DNA PAINT acquisition allows us to always recover the full pattern



dNAM full matrix



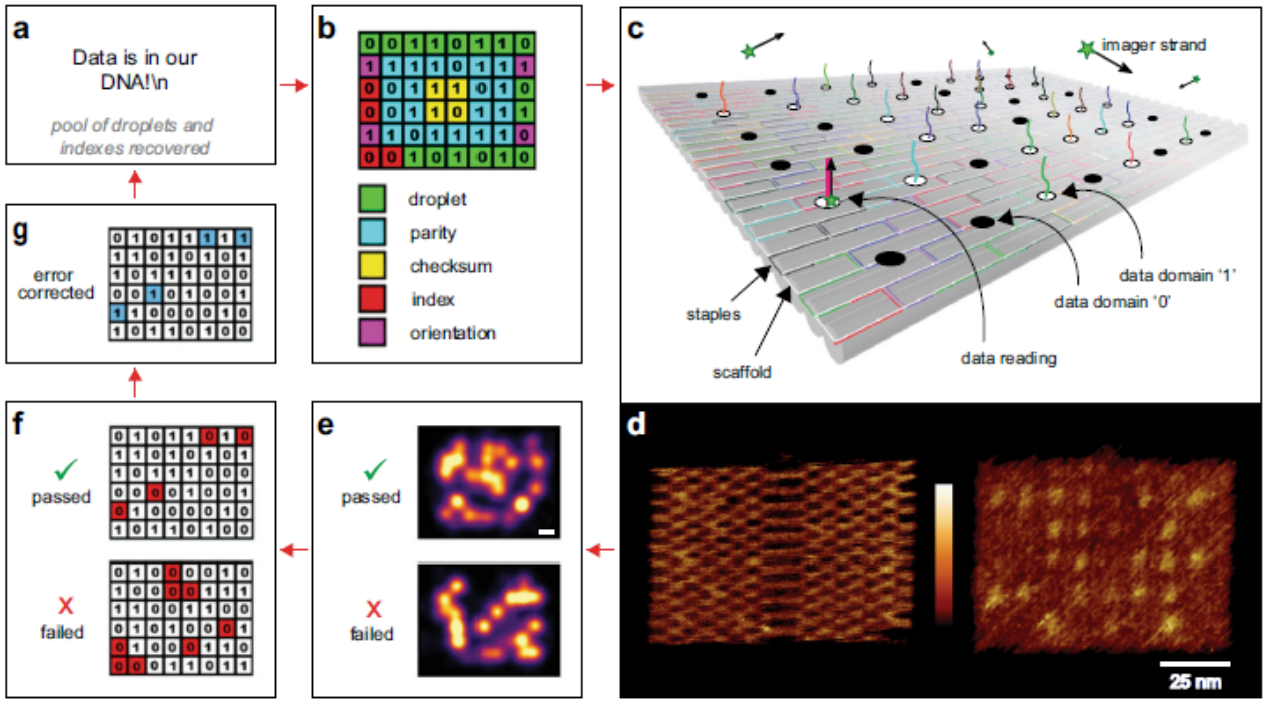
Averaged DNA PAINT

ARTICLE

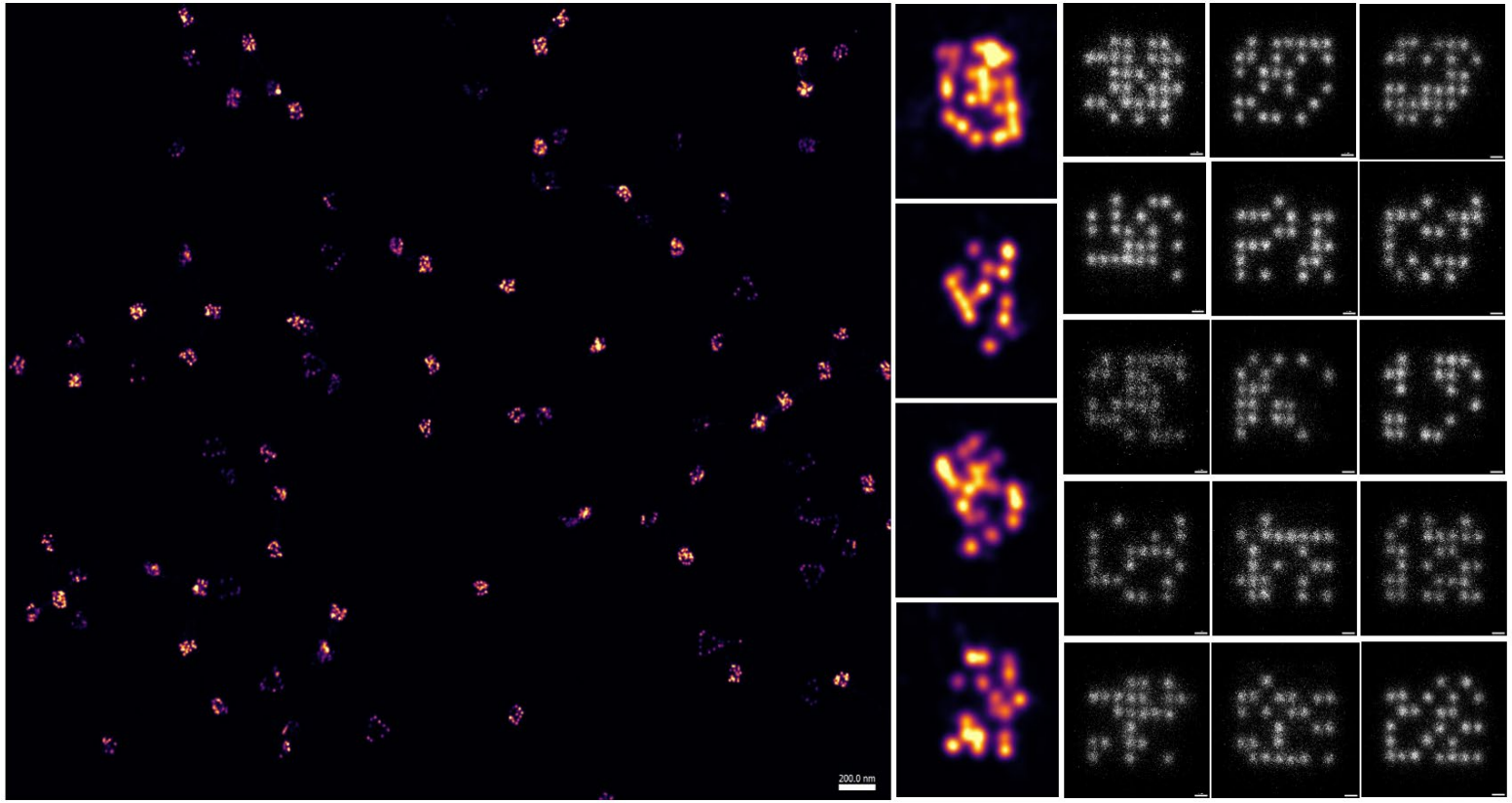
https://doi.org/10.1038/s41467-021-22277-y OPEN

An alternative approach to nucleic acid memory

George D. Dickinson[1,7], Golam Md Mortuza[2,7], William Clay[1,7], Luca Piantanida[1,7], Christopher M. Green[1,5], Chad Watson[1], Eric J. Hayden[3], Tim Andersen[2], Wan Kuang[4], Elton Graugnard[1], Reza Zadegan[1,6] & William L. Hughes[1✉]

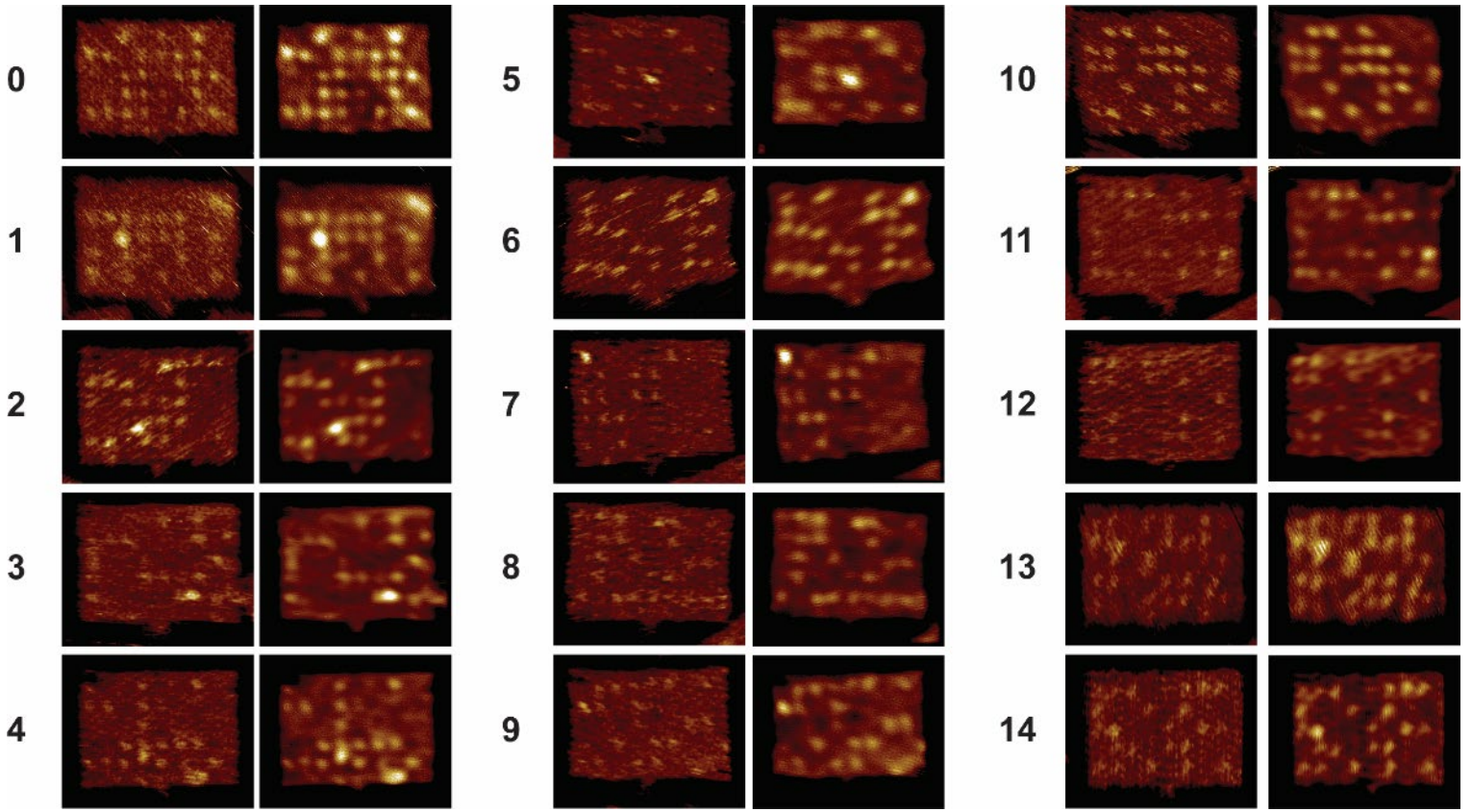


Super Resolution Microscopy

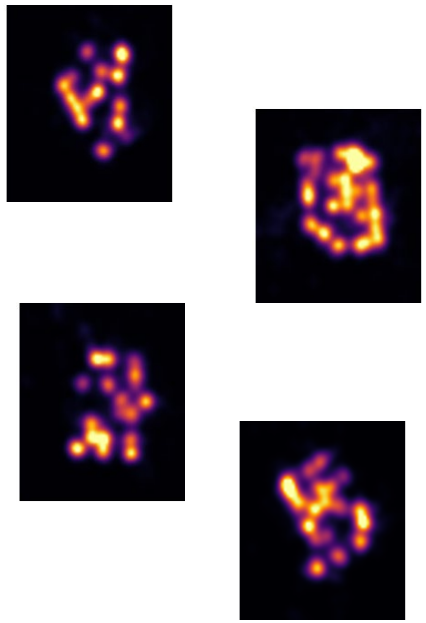


Binary data is recovered from fluorescence reading of DNA platforms

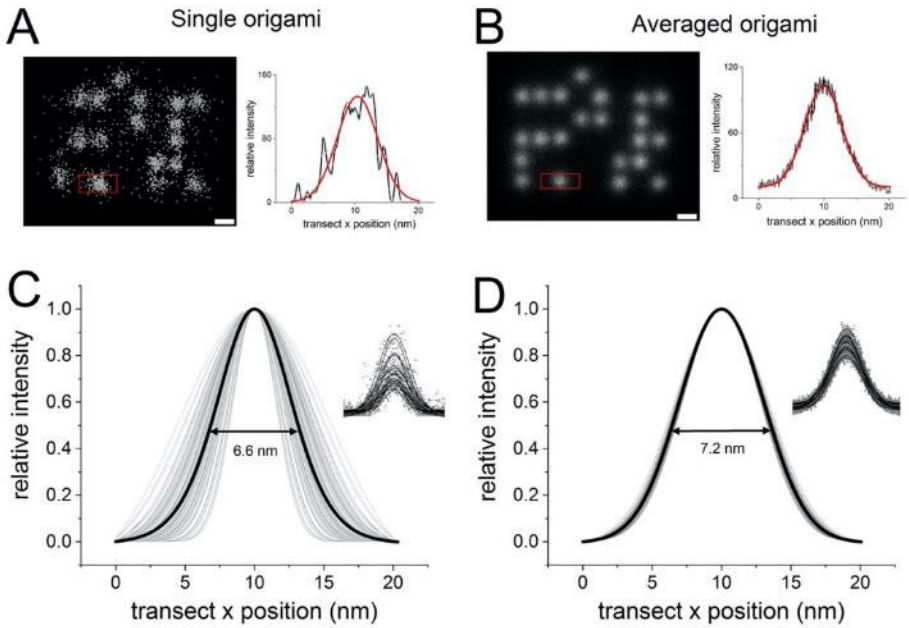
Atomic Force Microscopy

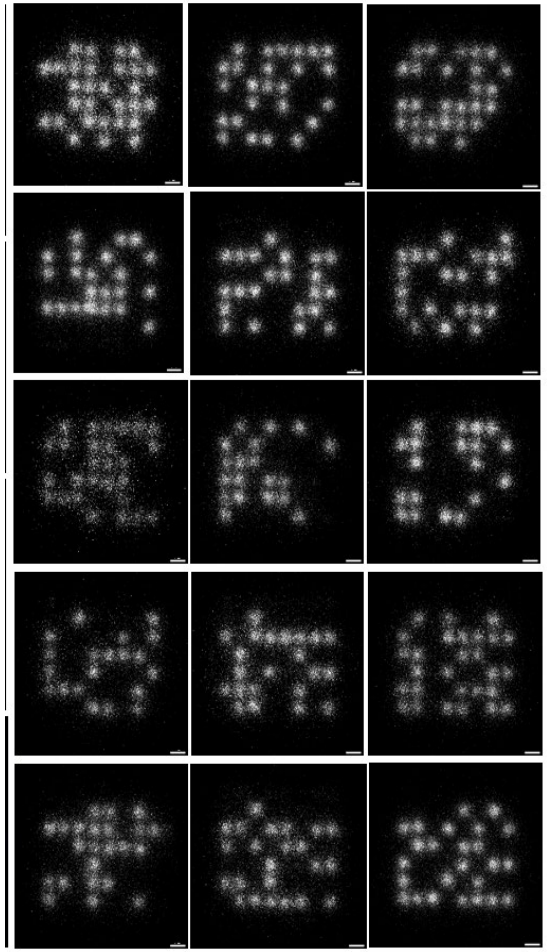


Microscopy technique used to check efficiency of DNA platforms formation



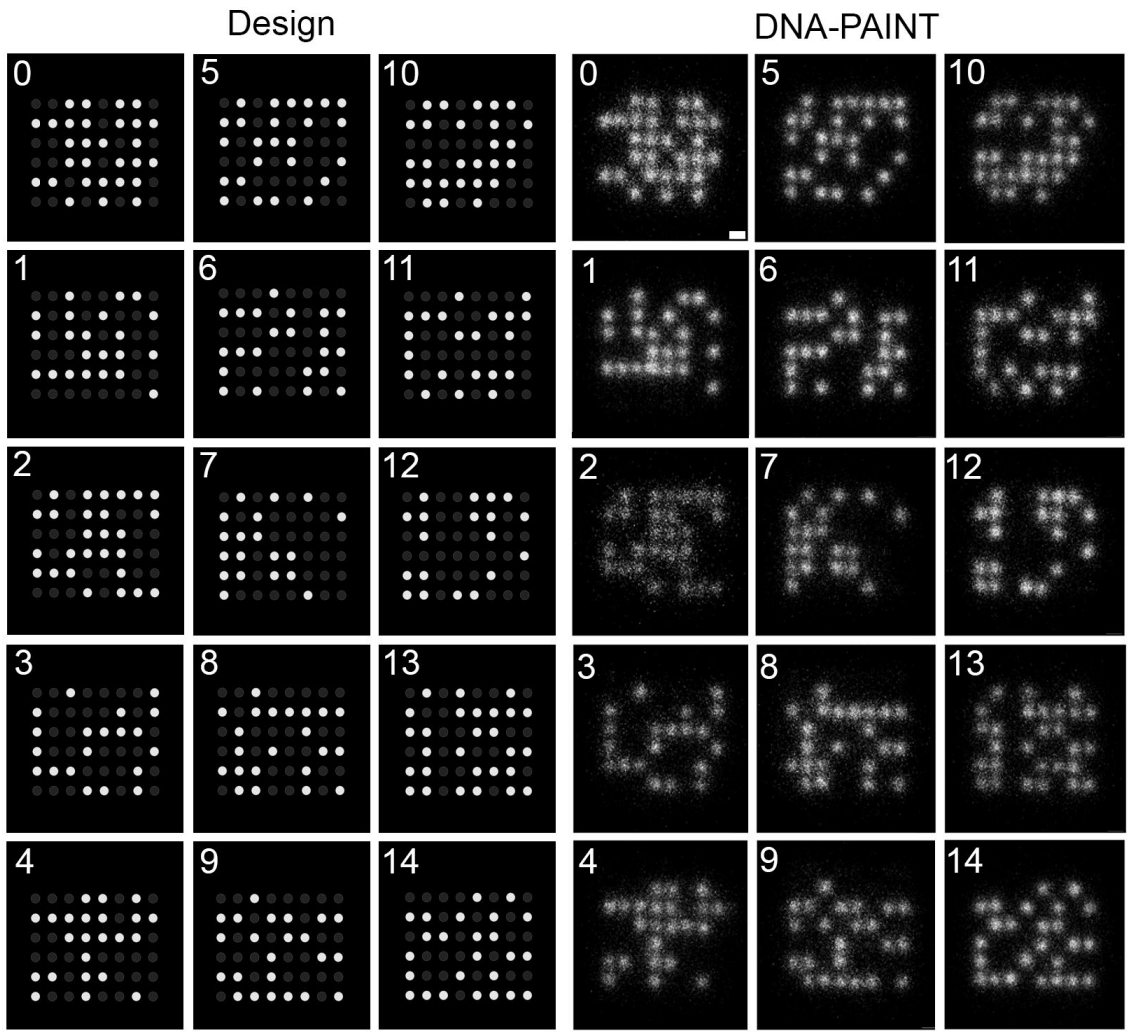
All origami are picked and super resolved





After being averaged all 15 matrices are retrieved from one single DNA PAINT acquisition

Encoding → Writing → Storage → Reading → Decoding



Averaging procedure is used to better present the 15 different droplets matrices but it's not needed in order to retrieve the digital data in dNAM