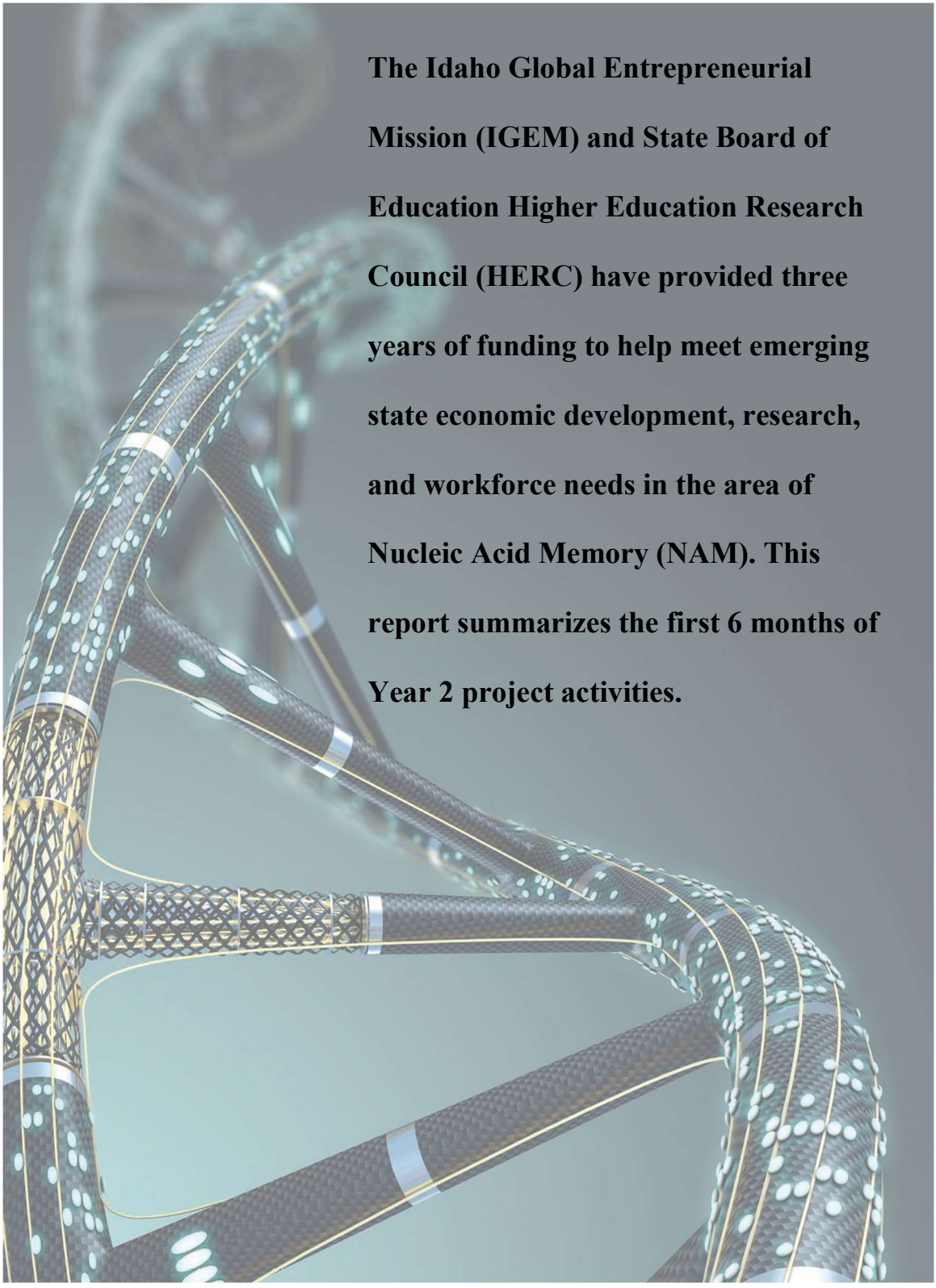
BOISE STATE UNIVERSITY

IGEM # 19-002: Nucleic Acid Memory

July 1, 2019 – January 1, 2020 Progress Report

Will Hughes
Tim Andersen
Eric Hayden
Wan Kuang
Will Clay
George Dickinson
Luca Piantanida
Chad Watson

I. Project Summary



The Idaho Global Entrepreneurial Mission (IGEM) and State Board of Education Higher Education Research Council (HERC) have provided three years of funding to help meet emerging state economic development, research, and workforce needs in the area of Nucleic Acid Memory (NAM). This report summarizes the first 6 months of Year 2 project activities.

II. Project Overview

In 2016, the digital universe produced 16 ZB (1 ZB = 1 trillion GB) of data. In 2025 it will create 163 ZB. These data, once generated, cascade through the information lifecycle — from primary storage media in the form of hard disks and solid-state drives to archival media such as tape. While the semiconductor industry maximizes the density, stability, and energy efficiency of electronic and magnetic memory, both are fast approaching their physical and economic finish lines. As envisioned by the new Semiconductor Synthetic Biology Roadmap, DNA-based massive information storage is a fresh start for memory manufacturing in the United States. According to our study with Micron, Harvard, and the Semiconductor Research Corporation (SRC), DNA has a retention time that ranges from thousands to millions of years, 1 kg of DNA can store the projected digital universe in 2040, and DNA's energy of operation is 100 million times less than current electronic memory. As a result, nucleic acid memory has become a global conversation, a national investment, an industrial opportunity, and a local strength in Idaho.

Our vision is to pioneer a digital data storage paradigm in Idaho by designing, building, and testing accessible, editable, and non-volatile nucleic acid memory (NAM) technologies that are inspired by DNA circuits and made possible by our innovations in DNA nanotechnology. With support from IGEM-HERC, we are creating a Nucleic Acid Memory Institute to meet critical innovation, economic, and workforce development needs in Idaho. To expedite our vision of Idaho becoming a global leader in NAM, five tasks are being addressed over the life of the IGEM-HERC: **Task 1** – Create efficient algorithms for coding information into data strands. Error correction strategies will account for DNA insertions, deletions, and substitutions, as well as screen for biological sequences to ensure that the data has no genetic function. **Task 2** – Create a high-throughput, integrated analytical engine to design and select data strands using quantitative metrics based on an in-house, algorithm. **Task 3** – Create synthetic biological factories for manufacturing DNA scaffolds using rapid design-build-test cycles of genomes. Genome size and structure will be engineered. **Task 4** – Design and fabricate NAM storage platforms using the DNA scaffolds, and validate the functionality of genome scaffolds using atomic force microscopy. **Task 5** – Read arbitrary data files into NAM storage nodes using super-resolution microscopy. Realize sub-nanometer imaging resolution to enable high areal density data storage.

This progress report spans July 1, 2019 to January 1, 2020. Listed below is a technical summary of our accomplishments during this time period.

III. Summary of project accomplishments

The support provided by IGEM-HERC during year 1 of this project provided the infrastructure and team to create the first digital Nucleic Acid Memory (dNAM) proof-of-concept. Building on this foundation, which was described in the July 1, 2018 to June 30, 2019 Annual Report, we conducted a series of experiments to validate dNAM, which provides key insights into the future of a DNA-based information storage roadmap. This work has led to a manuscript to be submitted next reporting period to either Nature Materials or Nature Communications. The content of the manuscript is highlighted below and represents the project tasks during this reporting period. The work is transdisciplinary by nature and highlights collaboration between materials science, computer science, biomolecular sciences, and electrical engineering at Boise State University.

Encoding and recovery of 'Data is in our DNA!\n' using dNAM

We designed an information encoding/decoding algorithm that combines fountain codes with a custom error detection scheme for dNAM. Fountain codes are intended for transmission of data over noisy channels and have great potential to generate a limitless sequence of encoding packets (called "droplets") from a single source message. Droplets can be received in any order and still be decoded to retrieve a message, regardless of the exact distribution of the droplets collected. With our error correction scheme, decoding of individual origami (each origami encodes a single droplet, plus indexing and error-correction information) is robustly achieved even in the presence of high noise by determining the minimum number of operations required to resolve the errors detected by the error detecting code (i.e. minimum edit distance). Combined with the redundancy provided by the fountain code, this leads to recovery of the entire data file with high reliability.

To test our dNAM data storage protocol, we synthesized fifteen unique 6 x 8 grid DNA-origami structures with the encoded random-access message '*Data is in our DNA!\n*'. Each of the memory blocks contained a 4-bit binary index (0000 - 1110), a 16-bit data droplet, 20 bits for parity checks, 4 bits for checksums, and 4 bits acting as orientation markers (Fig. 1A). To rapidly recover the message encoded in the blocks, we used DNA-PAINT to optically image a mixture of the dNAM blocks—below the diffraction limit of light—in a single recording (Fig. 1B, DNA-PAINT). Once DNA-PAINT had been used to identify the positions of all data strands in the 512 x 512 pixel field of view, a custom image processing algorithm was used to rotate and fit the data strands to a 6 x 8 grid and translate the signal detected at each grid location to a 48 bit binary string (Fig. 1B, Processing) for error correction, data recovery, and message reconstruction (Fig. 1B, Error Correction and Decode). We were able to identify all data strands in every origami (Fig. 2) from a single DNA-PAINT recording of 4 µl of a 5 nM mixture of DNA-origami (approximately 4,500 individual DNA-origami). By doing so, all fifteen origami were decoded to successfully retrieve the message '*Data is in our DNA!\n*'.

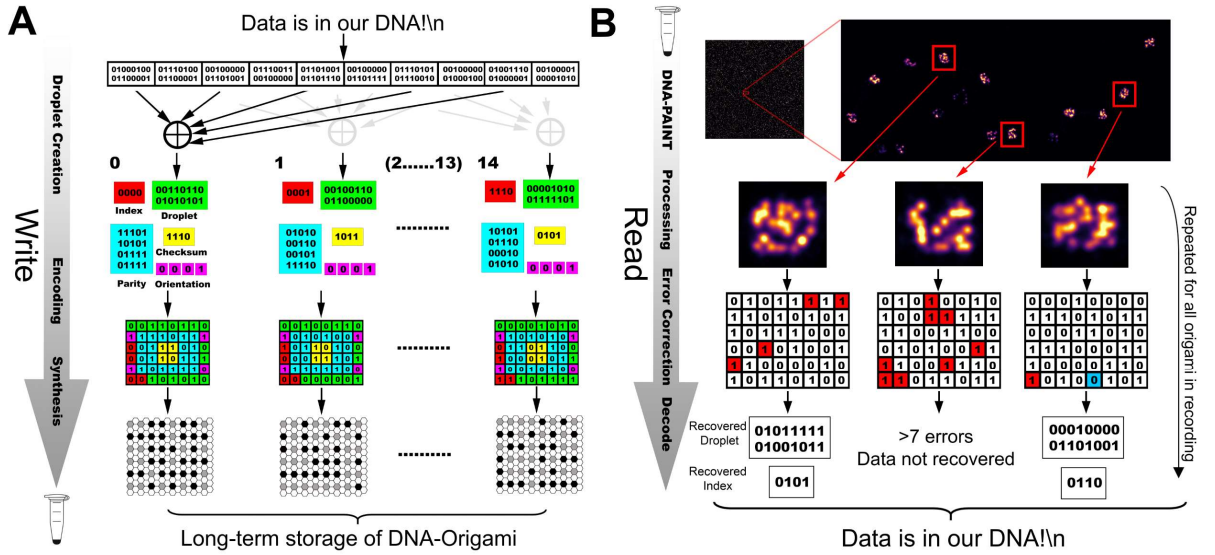


Figure 1. (A) Write: dNAM data packets can be written, stored, and read using DNA-origami and DNA-PAINT imaging. To 'Write' our test message ('Data is in our DNA!\n') fifteen DNA-origami were synthesized based on designs generated by the encoding algorithm in three steps depicted vertically in (**A**). For clarity the writing process is only shown for three of the fifteen structures (0, 1 and 14), however, the same process was carried out for all origami synthesized. First, the test message was partitioned into 16 bit non-overlapping segments and droplets formed by combining the segments via binary XOR (⊕) operations (**A. Droplet Creation**). The data droplets were assigned binary indexes (0000-1110), had parity and checksum bits calculated and were mapped to a 6 x 8 array, along with 4 orientation maker bits. The colors of the index (red), droplet (green), parity (blue), checksum (yellow) and orientation (magenta) bits correspond to their mapped positions in the array (**A. Encoding**). Finally, 2D DNA-origami with 46 addressable data-staples situated at ~10 nm intervals in a 95 x 70 nm hexagonal lattice were synthesized. Data-strands including binding domains for M1 imager probes represented binary '1' data points (black), while data-strands representing '0' (light grey) did not have a binding site extension (**A. Synthesis**). Following synthesis, the origami were gel purified, combined in equal amounts (~0.33 nM final concentration of each origami) and stored at 4°C. **(B) Read:** The four 'Read' steps used to recover the test message from the DNA-origami suspension are shown in (**B**). 4 ul of the DNA-origami mixture (containing all 15 origami) was diluted to 0.05 nM total origami DNA (0.5 TBE, 18 mM $MgCl_2$), deposited onto a glass coverslip using Glow discharge and imaged using DNA-PAINT with Cy3b-labelled imager probes complementary to the binding domains of the data-strands. A rendered image of the full-chip recording (~ 54 x 54 µm) is shown (**top left**) with a zoomed-in view of several individual DNA-origami (**top right**), one of which is a triangular structure used to aid drift correction (**D. DNA-PAINT**). The origami in the rendered image were identified using cluster analysis, individually fit to a 6 x 8 grid and thresholded to obtain binary outputs for each position of the array (**B. Processing**). Once the origami images were converted to 48 bit binary strings the decode algorithm used the parity, checksum and orientation bits to correct errors where possible. In the three representative origami structures, both false negative (red squares) and false positive (blue squares) are observable (**B. Error Correction**). When there were ≤7 total errors in a binary string with at least 13 data droplets and indexes recovered, the decode algorithm corrected errors and successfully retrieved the entire message (**B. Decode**). At the concentration of DNA deposited, all of the data droplets and indexes, and thus the entire message, were correctly decoded from a single 40,000 frame DNA-PAINT recording.

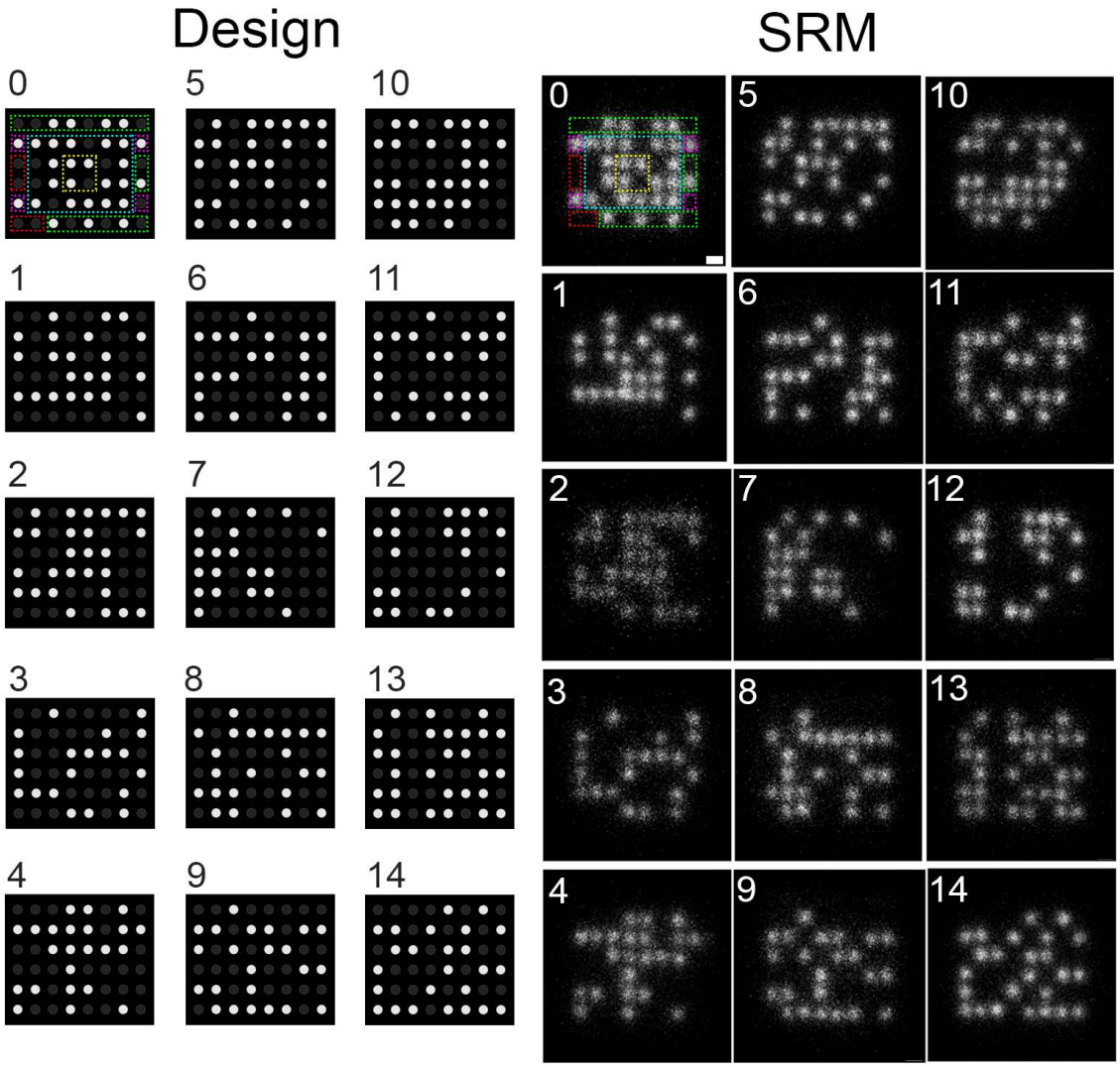


Figure 2. DNA-PAINT imaging of origami designs indicate all sites are recovered in a single recording. DNA-origami from a DNA-PAINT recording were identified and classified by aligning and template matching them with the 15 origami designs (**Design**) in which all potential docking sites are shown, filled circles indicate sites encoded '0' (black) or '1' (grey). Dashed boxes indicate the regions of the matrices used for indexing (red), droplet data (green), error correction (blue), checksums (yellow) and orientation (magenta) — to avoid repetition, only origami design 0 is highlighted. 'Averaged' images of 4560 randomly selected DNA origami, grouped by index, are depicted right (**SRM**). Scale bar, 10 nm.

Input / Output Variability

Comparing the decode algorithm output with each of the 15 input origami designs (each having a total of 223–338 structures within the 5 nM mixture mixture) indicate varying numbers of each origami were successfully recovered. For example, while on average only ~6 copies of origami design 2 were correctly decoded per experiment, ~147 copies of origami design 6 were successfully decoded (Fig. 3A, B). These differences in recovery can partially be explained by the variability in the numbers of errors seen in each structure (Fig. 3C). Specifically, the decode algorithm was only able to error-correct origami with 7 or fewer total errors, and only up to two false positive errors. The mean errors of the best recovered origami designs (1, 5, 6, 12 and 14) are all lower than these thresholds. A plot of the origami decoded against the mean error rate indicates that there is a strong relationship between both the total number of errors and the number of false negative errors and the ability of the algorithm to decode an individual origami's data (Fig. 3D). False positives errors, however, are randomly distributed around a mean of 2, but with a wide

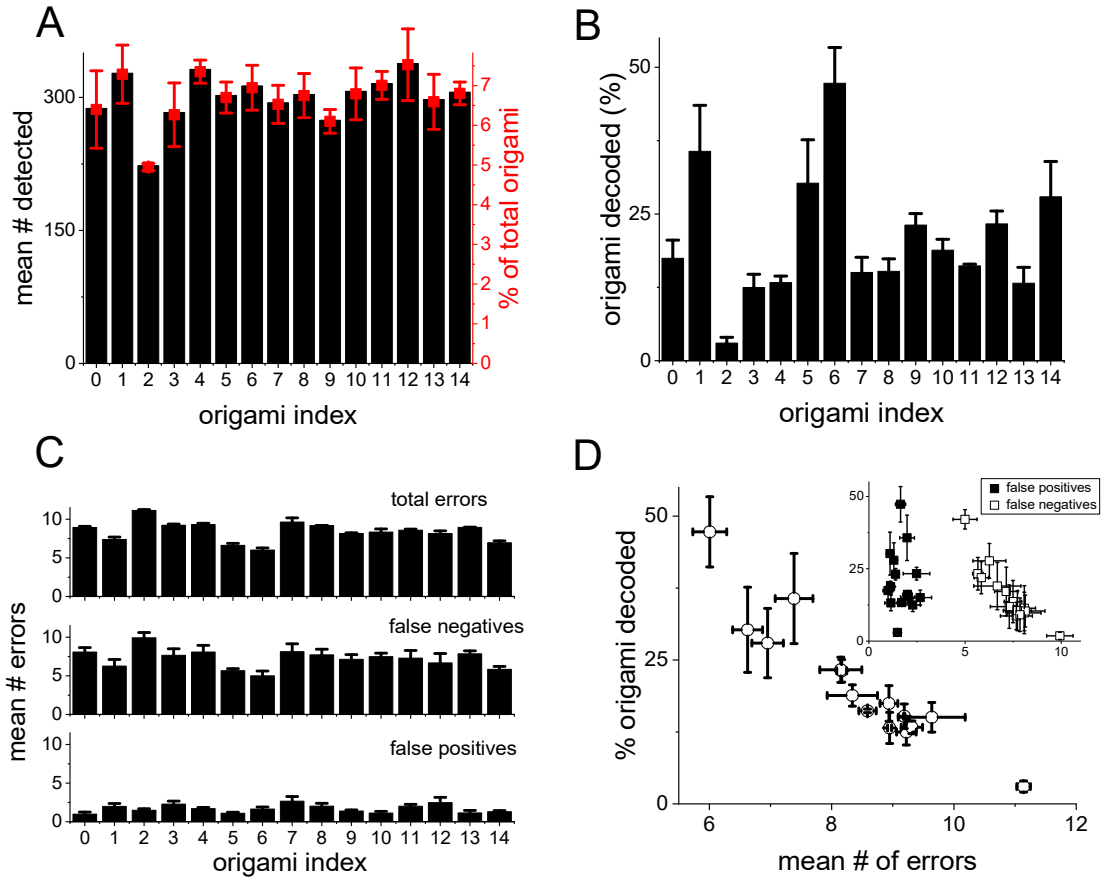


Figure 3. After image processing the origami were assigned an index based on their similarity to the origami designs. The left histogram (**A**) indicates the numbers of individual origami identified from a single full-chip DNA-PAINT recording (512 x 512 pixel sensor) based on similarity to the designs (mean counts shown as black bars, percentage of total origami in red). These origami were also decoded using the decode algorithm. (**B**) Depicts the percent of origami passed to the decode algorithm that had both their indexes and data strings correctly identified. (**C**) The variability in the mean number of false positive (top) and negative (below) error for each structure is shown. In graph (**D**) the percentage of each origami decoded are plotted against the mean number of errors for each structure. Mean values for three experiments are shown throughout, error bars indicate ±SD.

spread of decoded origami (~1–49%) — suggesting that the location of a false positive error within a structure plays an important role in determining whether the origami designs can be successfully decoded.

Source of Errors

By categorizing the errors of origami with 15 or less errors by their positions within an array (outer edge, mid and inner region) and normalizing (dividing the position error by the mean error of all of the sites), it is possible to pool all fifteen different origami designs to see if there are consistent differences between regions. Plots of the mean normalized errors indicate that the outer edges and interior of the array are differentially prone to false negative and false positive errors (Fig. 4). Our results indicate differences in successful strand incorporation between the inner and outer regions that could explain the differences in the numbers of errors observed. However, an alternative explanation could be that data-strands on the outer edge of an array are less likely to be affected by contaminating signals from neighboring sites (which would increase the chances of false positives) and vice-versa for interior sites.

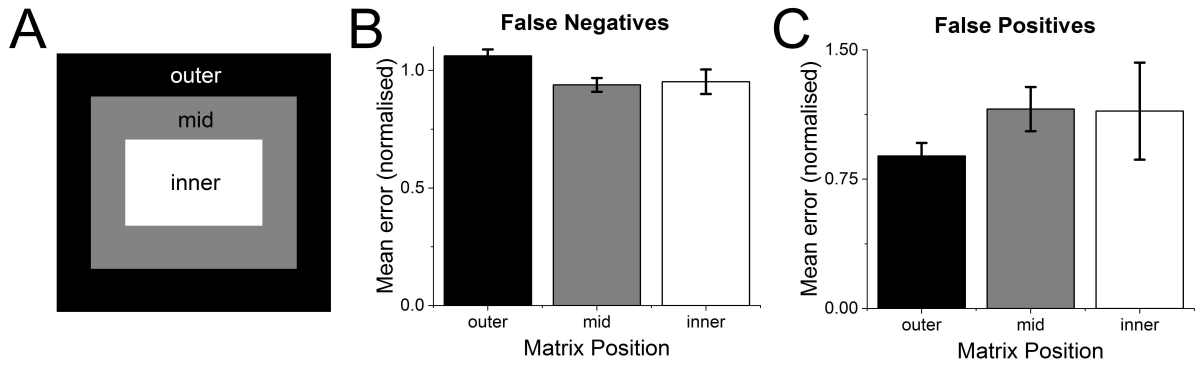


Figure 4. The outer edges and inner regions of origami are differentially error prone. The array positions of origami with 15 or less errors (as identified by pattern matching) were classified as either 'outer', 'mid' or 'inner' depending on their position in the array (**A**). The mean error for each classification was calculated and normalized by dividing by the overall mean error for that zone. Plots of the mean normalized false negative (**B**) and false positive values (**C**) for each zone are shown. Mean values for three experiments are shown, Error bars indicate SD.

To further investigate sources of potential sources of error in our array designs, we performed atomic force microscopy (AFM) imaging on individual DNA-origami deposited on mica. From the averaged SRM images in Fig. 1, it can be seen that every data strand was recorded at least once for all expected positions in all arrays. This suggests that there were no systematic failures in strand incorporation or data-strand binding domains. This is further substantiated by the AFM images, in which origami were typically both well formed (lacking holes and having the expected dimensions) and appeared to have incorporated the majority of their data-strands. Although it was possible to resolve the majority of data strands positions (Fig. 5), a strict analysis on missing data strands using AFM would not be completely reliable since tip-sample interactions could easily promote strand compression and displacement.

We also used AFM to examine DNA-origami deposited onto a glass coverslip immediately following SRM imaging. We were not able to resolve individual binding sites in these images (most likely due to the increased roughness of glass, as compared to mica), however it was possible to count the total number of origami in a field of view for comparison with structures visualized by SRM. The densities of origami estimated from the images were 2.4 and 1.4 origami/μm^2 for AFM and SRM respectively, suggesting that ~60% of the total origami deposited have their binding sites facing away from the coverslip and available for imager strand binding.

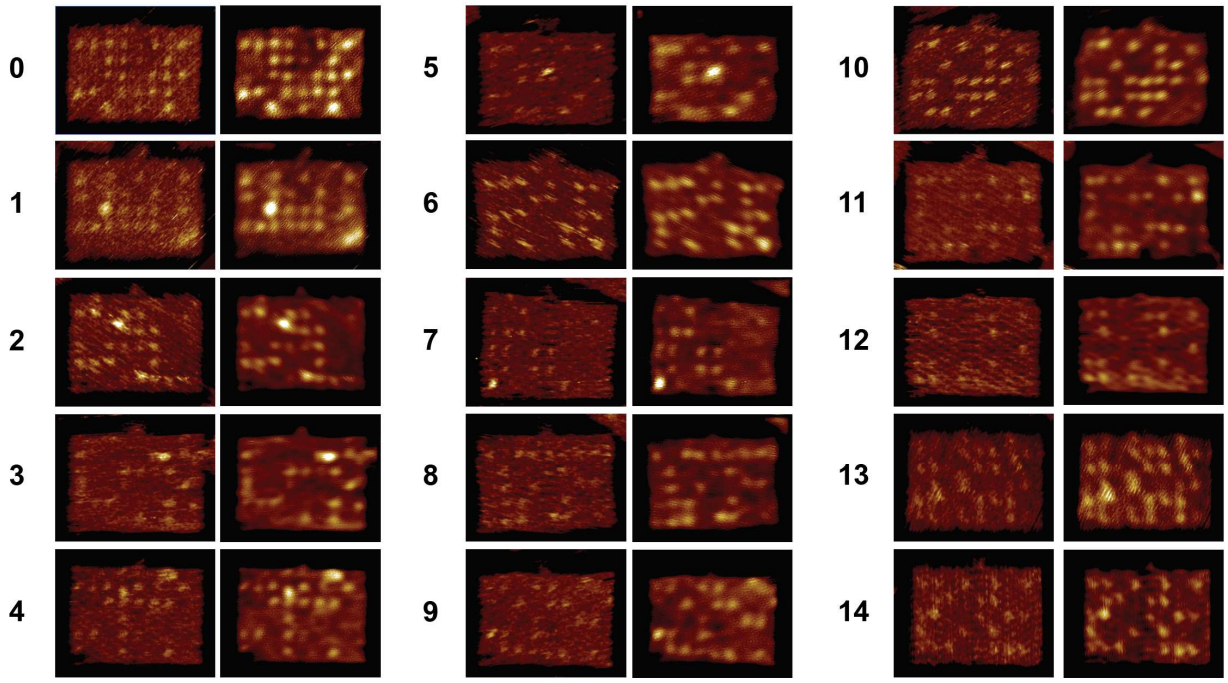


Figure 5. AFM images of DNA origami data nodes. Representative AFM images of all 15 dNAM "Data is in our DNA!/n" data node origami, where most of dockings sites are visible. (An inverse FFT analysis with a band rejected filter has been applied to highlight the dockings positions in right-hand panels). Every image is 90 x 110 nm and the color scale ranges over 250 pm.

Implications

We have created—through a systems-engineering approach—a new technology where digital data is encoded into multiple DNA origami structures and can be retrieved optically below the diffraction limit of light via super-resolution microscopy. By encoding our data using Fountain codes, combined with bi-level error detection/correction, the amount of redundancy required for successful data recovery is minimized (with 100% data retrieval ensured with sufficient node creation). While we have encoded the short message '*Data is in our DNA!/n*' as a proof-of-principal, dNAM platform is scalable and thus has potential for competing with current data storage technologies. For comparisons, our dNAM prototype currently allows for a data storage density 480 Gbit/cm^2, a more than ten-fold improvement relative to state-of-the-art magnetic tape capacity of 31 Gbit/cm^2.

IV. Demonstration of economic development and impact

Demonstration of Economic Development and Impact	Year 1 Reporting Period 07/01/2018–06/30/2019	Current Reporting Period 07/01/2019–01/01/2020
External Funding	$ 1,549,995	0, because we are doing the work
Number of External Grants	3	0, because we are doing the work
Private Sector Engagement	14 companies	2 companies, 1 VC group
University Engagement	11 universities	~20 universities
Federal Agency Engagement	5 agencies	4 (NSF, SRC, NRL, NIST)
Industry Involvement	2 companies	2 companies (Micron, SRC)
Patents	0	0
Copyrights	0	0
Plant Variety Protection Certificates	0	0
Technology Licenses Signed	0	0
News Releases	3 articles	0
Start-up Businesses Started	0	1
Jobs Created outside of BSU	0	6

External Funding

During this reporting period, we did not pursue external funding opportunities. Instead, we devoted ourselves to our current funding from IGEM/HERC, the National Science Foundation, and the Semiconductor Research Corporation. In doing so, we established dNAM as a new memory technology that is both information dense and has the promise of being stable for neo-archival applications. When published, this work will further position our team as a pioneer in DNA memory. Moving forward, the research team is on track to submit one proposal to an external funding agency before the end of the next reporting period.

Engagement

With the backing of the NSF Office of Emerging Frontiers and Multidisciplinary Activities, Hughes hosted the 2019 Germination Meeting at Boise State University on August 15-16, 2019. The meeting focused on new approaches in cultivating risk-taking and impact-driven research culture. As noted in the Year 1 Annual Report, the National Science Foundation (NSF) in collaboration with the Semiconductor Research Corporation (SRC) jointly awarded the research team $1,500,000 to address the scientific challenges facing NAM technologies. The funding mechanism was called *Semiconductor Synthetic Biology for Information Processing and Storage Technologies*. As part of this funding, the SRC holds an annual conference to showcase "*the quality and breadth of the SRC research portfolio, the excellence of SRC students and faculty, and the magnitude of the collaborative research investment made by industry through SRC*." Hughes and two PhD students (now graduated) on the NAM team, Chris Green and Mike Tobiason, attended the conference, which was held in Austin, Texas from September 8–10. "*The conference features student presentations and posters and gives SRC member companies multiple formal and*

informal occasions to network with SRC students. This is a great opportunity for students to meet with SRC member companies, including 7 of the top 10 semiconductor companies in the world. These networking occasions with SRC member companies give student opportunities to open the door to future full-time employment."

Hughes was also among a select group of scholars, industry stakeholders, and program managers to participate in a workshop on Nucleic Acid Nanotechnology. The workshop, held in Boston, Massachusetts on Dec. 7, 2019, was co-sponsored by the Materials Research Society and the prestigious Kavli Foundation. The goals of the workshop were to establish "priority research areas for next-generation applications of nucleic acid nanotechnology across diverse domains spanning computation, sensing, fabrication, therapeutics, and other areas." Through this process, Hughes reinforced relationships with Harvard University (George Church, William Shih), MIT (Mark Bathe), NIST (James Liddle), NRL (Igor Medintz), John's Hopkins (Rebecca Schulman), as well as established new relationships with the editors of Nature and Nature Materials. Based on ideas shared, George Church opened his lab to members of the NAM Institute at Boise State.

Business Development

Steven Burden, who successfully completed his PhD in Biomolecular Sciences, graduated December 2019 as a member of the NAM Institute. Burden's dissertation topic was on the development of nucleic acid biosensors with allosteric fluorescence signals. For the NAM team, Burden played a lead role in our Vertically Integrated Project (VIP), where he trained undergraduate students to produce, purify, and ensure the quality control of single-stranded DNA scaffolds. Prior to graduating from Boise State, Burden co-founded a biotechnology startup (FACible BioDiagnostics — https://www.facible.com/). Based in Boise, Idaho, FACible BioDiagnostics is focused on developing rapid, low-cost, diagnostics. Burden began full time employment as the company's CEO on January 1, 2020. In addition, one of the co-founders, Clementine Gibard Bohachek, was a postdoctoral research scientist at Boise State University and was part of the NAM team during the spring of 2019, where she developed VIP training materials and trained VIP and NAM graduate students on practical laboratory approaches to synthetic biology. In all, FACible BioDiagnostics employs 6 people — three full time and three part time. The financial, scientific and professional support that Burden received during his PhD was critical for his ability to secure venture capital needed to start his company. The success of Burden highlights the entrepreneurial environment that is being cultivated by the NAM Institute and team.

V. Numbers of student, staff, and faculty participation

Classification	Number
Tenured or Tenure Track Faculty	4 (*2 full professors, 2 associate professors*)
Research Faculty	1 (*started a tenure-track faculty position*)
Project Manager	1 (*also focused on business development*)
Senior Lab Research Associate	1 (*manages the laboratory & supports team*)
Postdoctoral Fellows	3 (*performing at a research faculty level*)
Graduate Students	5 (*3 of the 5 graduated in December*)
Undergraduate Students	10 (*5 female and 5 male*)

From a professional development perspective, the goal of the NAM Institute is to ensure the success of the people that make up the team, from students and postdoctoral research scientists to the faculty and staff that enable open innovation, ideation, and collaboration. And with any academic environment, matriculation to graduation is expected, supported, and applauded. Thus, we are happy to report that during this reporting period three PhD students on the NAM project successfully defended their PhD dissertation and graduated:

- **Steven Burden**, PhD in Biomolecular Sciences, Dissertation — *The Development of Nucleic Acid Biosensors with Allosteric Fluorescence Signals*
- **Chris Green**, PhD in Materials Science and Engineering, Dissertation — *Nanoscale Optical and Correlative Microscopies for Quantitative Characterization of DNA Nanostructures*.
- **Mike Tobiason**, PhD in Materials Science and Engineering, Dissertation — *Engineering Kinetically Uniform DNA Devices*

In addition, Reza Zadegan has started a tenure track faculty position at North Carolina A&T this past August. His professional development included but was not restricted to: grant writing support by Watson and Hughes; germination of research directions and intellectual risk management by Hughes; helping create his faculty package by Hughes, Andersen, and Hayden; mock interviews by Hughes; national and international networking opportunities by Hughes; technical training by Andersen, Hayden, Kuang, and Hughes; and professional training from Hughes and Watson. We also would like to acknowledge that one of the project principle investigators, Elton Graungnard, has transitioned from the team to focus his efforts on developing atomically-thin semiconducting materials for high performance, energy-efficient electronic devices. While Graungard's absence will be missed, we have hired a new postdoctoral research scientist, Luca Piantanida, who started on August 5, 2019. Piantandia's complementary expertise enables the team to move forward without needing to make course adjustments. Piantanida has a PhD in Nanotechnology from University of Trieste, where his dissertation was on developing DNA origami nanoactuators functionalized with gold nanoparticles for plasmon resonance tuning. Piantanida recently concluded a postdoctoral position at Durham University, UK under the supervision of Prof. Kislon Voïtchovsky, where he developed a novel atomic force microscopy approach for imaging

biological interfaces in fluid. His expertise with DNA origami and high-resolution imaging, coupled with the scientific expertise and productivity of postdoctoral research scientists Drs. Will Clay and George Dickinson, position the NAM Institute to accelerate our development rate.

Vertically Integrated Project

The Vertically Integrated Project (VIP) model integrates teaching and learning into one framework in support of work-force development of students that can work at the interface of semiconductor manufacturing and synthetic biology. These students are engaging in research activities aimed toward the production, purification, and quality control of new single-stranded DNA origami scaffolds. The students range from sophomore to seniors and span four different majors: biology, chemistry, health sciences, and psychology. Specifically, the VIP students synthesized and purified several large DNA scaffolds. They used *E. coli* cultures to express single stranded DNA ~8,000 and 10,000 bp in length. Currently, the bacteriophage M13mp18 is used to make the DNA scaffolds, but it limited to 7249 nucleotides. In addition to being longer than M13mp18, each of these scaffolds has a different sequence, potentially enabling orthogonal origami synthesis.

VI. Description of future plans

Team Management – Integration and graduation

- Manage the financial risk of the anticipated higher education budget cuts in Idaho that have the potential to impact the NAM Institute.
- Target the next round of grant opportunities and start working towards their submission. Leverage the grant writing process as an opportunity for professional development of the postdoctoral fellows.
- Help the postdoctoral fellows identify the intellectual space they want to lead in the future; periodically meeting with them to establish their professional development plans.
- Seek collaborations with key internationally recognized research groups; with an eye for cross-training our laboratories.
- Support our project manager to visit federally funded research centers to solicit effective practices in how to mobilize, pitch, and manage a large federally-funded effort.

Task 1 – Create improved algorithms for coding information into data strands.

- Explore the integration of Tasks 1 and 2 together to expand *DevPro* and *SeqEvo* to include machine learning so that our experimental results help inform future sequence design.
- Submit the initial manuscript describing and encoding/decoding algorithm into DNA to **DNA 26**, as well as a review article on encoding information into DNA to **IEEE** or equivalent.

Task 2 – Create a high-throughput, integrated analytical engine to design select data strands using quantitative metrics based on an in-house, algorithm.

- Resubmit the initial manuscript describing *SeqEvo* and *DevPro* to **Nucleic Acid Research**. Initial reviews from the journal reinforced the value of the research, as well as the need to communicate it more clearly because of our attempt/need to bridge multiple disciplines.
- The first version of our evolutionary sequence-generation tool (*SeqEvo*) and our sequence-analysis tool (*DevPro*) have been released to select research groups at Boise State University, the National Institute of Standards and Technology, and the Naval Research Laboratory to solicit their technical feedback in preparation for releasing both tools publicly. Based on their feedback, finish and then make publicly available *DevPro* and *SeqEvo* versions 2.0.
- Scale the adoption and adaption of *SeqEvo* and *DevPro* in support of Tasks 3 and 4.

Task 3 – Create a synthetic biological factory for manufacturing DNA scaffolds using a rapid design, build, and test cycle of genomes.

- With the successful development of software to optimize sequences (Task 2), we will next set out to design and synthesize large scaffolds with sequences optimized for our specific origami designs. Several designs will be synthesized and compared. The super resolution microscopy advancements will aid in this comparison. This will require Tasks 2, 3, and 4 to further integrate.
- Develop quality control metrics for scaffolds. Each scaffold synthesis will need external quality control metrics to ensure batch to batch consistency in order to enable comparison.

Task 4 – Design and fabricate NAM storage nodes using the DNA scaffolds.

- Submit the initial manuscript describing *dNAM* to **Nature Communications** or **Nature Materials**.
- Explore the application of short Locked Nucleic Acid (LNA) and other DNA analogues in dNAM to increase the resolution of the super-resolution microscope during DNA-PAINT, as well as explore if sequence Nucleic Acid Memory (seqNAM) is as viable as dNAM.

Task 5 – Read arbitrary files into NAM storage nodes using super-resolution microscopy.

- Test methods to improve resolution on existing microscope, including reducing drift, improving drift correction, and increasing the signal-to-noise ratio.
- Use knowledge gained from optimizing existing microscope to design and test components for custom built super-resolution microscopy system while working toward a full prototype.
- Use simulation to better understand optimal imaging and sample design parameters to maximize data reading rate.

VII. Summary of Budget Expenditures

The below table summarizes expenditures associated with the project from July 1, 2019 to December 20, 2019. In establishing dNAM as a new memory technology, *Salary* and *Fringe* supported five graduate students, three postdoctoral research scientists, an undergraduate student, an assistant research professor, and a project manager. *Other Expenses* were used to purchase modified and unmodified DNA oligos, supplies to process modified and unmodified DNA oligos into dNAM, super-resolution microscopy supplies, atomic force microscopy supplies, computers, and an upgrade to a liquid handling system (epMotion). The oligos are used to assemble NAM blocks and to perform super-resolution microscopy studies. The atomic force microscopy supplies complement the super-resolution studies by confirming the design and structural stability of the dNAM. The computers were purchased in support of our algorithm development and newest postdoctoral research scientist. The epMotion system enables us automate the mixing of solutions to synthesize DNA origami in both an accurate and efficient manner. The system was malfunctioning and was approaching its end-of-life. The upgrade ensures vendor support throughout the life of this project. The encumbered *Capital* is allocated to the purchase of a server to significantly improve our image processing efficiency and infrastructure. As part of analyzing dNAM, we compile over 60,000 high resolution images (~40 GB) per experiment. Post-processing of each series of experiments, and the 60,000+ images, are computationally intensive. When performed on a desktop computer, processing requires hours to days of processing time per experiment. The server will enhance productivity.

Category	Current Budget	Expenditures	Encumbered	Remaining Budget
Salary	$282,671.00	$118,649.66		$164,021.34
Fringe Benefits	$96,375.00	$29,017.86		$67,357.14
Other Expenses	$93,500.00	$52,004.98	$9,587.64	$31,907.38
Travel	$15,000.00	$480.00	$288.00	$14,232.00
Capital	$150,000.00		$23,689.56	$126,310.44
Student Costs	$28,954.00	$15,881.40		$13,072.60
Total	$666,500.00	$216,033.90	$33,565.20	$416,900.90

VIII. Commercialization Revenue

Commercialization	Revenue
None.	$0

IX. Additional metrics established specific to individual project

Metrics	Number
External Funding	$ 1,549,995
Graduate Degrees Awarded	4 (3 PhD, 1 MS)
Dissertations Published	4 (3 PhD, 1 MS)
Invited Technical Presentations	15 (5 oral, 10 poster)
Software Tools Created	3
Peer-Reviewed Publications	1
Manuscripts in Preparation	4
Number of Graduate Students on the Project	2
Number of VIP Students Enrolled (grad and undergrad)	10
Number of National and International Postdocs Hired	3
Number of Scientists that have become Tenure Track Faculty	1 (North Carolina A&T)
Number of PhD Students that have received Postdoc Fellowships	1 (NRC Fellowship)
Number of PhD Students that started a Company in Idaho	1 (6 employees)

Note: Listed above are specific, objective, measurable, and realistic performance metrics over the lifetime of the project. These metrics, many of which have been distributed throughout this report, are a reflection of project success and inform economic impact.