BOISE STATE UNIVERSITY

Final Report

IGEM # 19-002: Nucleic Acid Memory

Will Hughes

Chad Watson

Tim Andersen

Eric Hayden

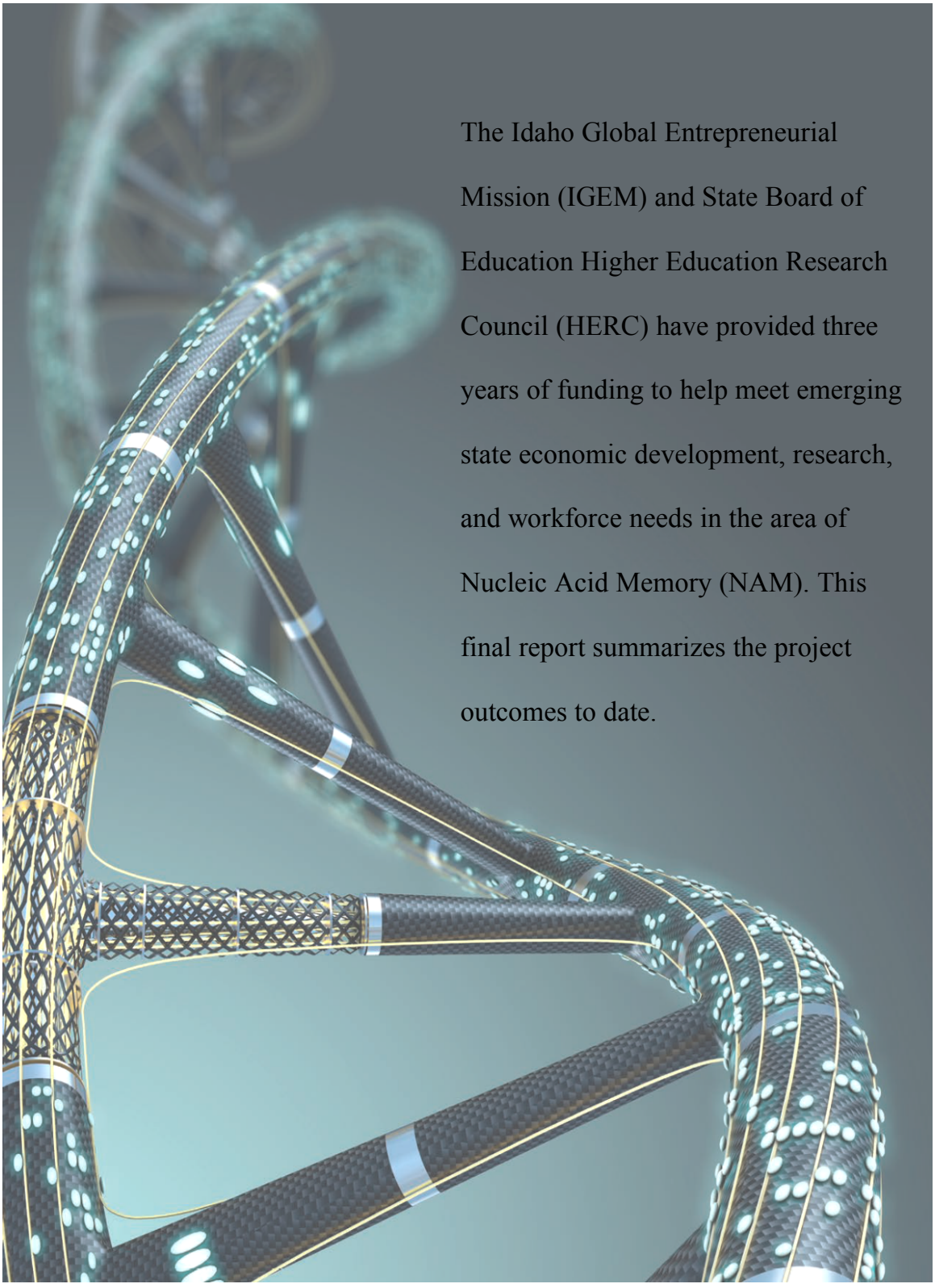
Wan Kuang

George Dickinson

Will Clay

Luca Piantanida

Mike Tobiason



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 final report summarizes the project outcomes to date.

I. Executive Summary

According to our theoretical study with Micron, Harvard, and the Semiconductor Research Corporation[1], 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. With support from IGEM/HERC, our vision was to prototype a digital data storage paradigm by designing, building, and testing non-volatile nucleic acid memory (NAM) technologies that are inspired by DNA circuits and made possible by innovations in DNA nanotechnology. The focal point for this research was to prototype digital nucleic acid memory (dNAM), a storage medium where data is encoded into the physical address of DNA strands within a DNA origami breadboard.

To achieve working implementations of the dNAM prototype, our team prioritized the following objectives: NAM coding (*objective 1*), NAM sequences (*objective 2*), NAM scaffolds (*objective 3*), NAM fabrication (*objective 4*), and NAM reading (*objective 5*). The following deliverables are aligned to the project objectives and have been coded to reflect progress on each. Green, yellow and red mean that the deliverable has respectively been accomplished, has been partially accomplished, and has not been accomplished. The red deliverables have not been accomplished for two important reasons: (1) the seqNAM prototype in *objective 5* was replaced with a new technique called $_{3D}$NAM – which is described below. In addition, selective area immobilization in *objective 4* was unnecessary once our computer algorithms compensated for the orientation of the DNA nanostructures.

Objective	Year	**Original Deliverable** (progress to date)
1	1	An information encoding/decoding algorithm for dNAM
	2	An information encoding/decoding algorithm for seqNAM
2	1	A high-throughput data pipeline for metrics-based optimization of data cells
	2	Sets of optimized codons and words for seqNAM data encoding
	3	Optimized insertion sequences for custom scaffold synthesis
3	1	Create viable phage modules from E.coli
	2	Validate phage genomes for use in DNA origami synthesis
4	1	Process for design and fabrication of DNA origami storage nodes
	2	Protocols for statistical correlated AFM/SRM defect metrology in DNA origami
	3	Protocols for selective area immobilization of DNA origami
5	1	Optical readout of dNAM
	2	Optical readout of seqNAM
	3	3 nm imaging resolution for SRM

Summarized here are products created during the award period, including patents, journals, commentaries, news briefs, software packages, select presentations, dissertations, and a company. Items highlighted in blue were accomplished since the last reporting period.

Product	Status	Details
Patent	filed	**NUCLEIC ACID MEMORY (NAM) / DIGITAL NUCLEIC ACID MEMORY (DNAM):** The present application claims priority to the earlier filed U.S. Provisional Application having Serial No. 62/705,995, and hereby incorporates subject matter of the provisional application in its entirety. The invention relates generally to nucleic acid memory (NAM). More specifically, the invention relates to digital Nucleic Acid Memory (dNAM) which use a nucleic acid architecture to create a physical address by providing docking sites for single stranded nucleic acid for information processing. The invention further relates to methods for enhanced data retention and retrieval and systems for use.
Journal	published	Dickinson, G.D., Mortuza, G.M., Clay, W., Piantanida, L., Green, C.M., Watson, C., Hayden, E.J., Andersen, T., Kuang, W., Graugnard, E., Zadegan, R. Hughes, W.L. An alternative approach to nucleic acid memory. *Nature Communications* 12, 2371 (2021).
Journal	published	Green, C.M., Hughes, W.L., Graugnard, E., Kuang, W. Correlative Super-Resolution and Atomic Force Microscopy of DNA Nanostructures and Characterization of Addressable Site Defects. ACS Nano Article ASAP
Journal	submitted	M. Tobiason, B. Yurke, W.L. Hughes, "Generation of DNA Oligonucleotides with Similar Hybridization Rates," **Nucleic Acid Research**, Submitted 2021.
Journal	in preparation	Llewellyn, S., Mortuza, G., Guerrero, J., Suyehira, K., Hughes, W.L., Andersen, T., Zadegan, R. **Algorithms for Digital Data Storage in Nucleic Acid Memory**, Summer 2021 Submission Goal.
Commentary	published	G. Dickinson, L. Piantanida, W.L. Hughes, "DNA 'Lite-Brite' is a promising way to archive data for decades or longer," The Conversation, May 10, 2021. https://theconversation.com/dna-lite-brite-is-a-promising-way-to-archive-data-for-decades-or-longer-157856
Commentary	accepted	L. Piantanida, W.L. Hughes, "A PCR-free approach to random access in DNA," **Nature Materials**, Accepted 2021 (NM21051571A).
News	published	NSF Research News – Researchers advance DNA as a storage material, May 13, 2021. www.nsf.gov/discoveries/disc_summ.jsp?cntn_id=302717&org=NSF&from=news
News	published	Chemical & Engineering News – Method offers new approach to DNA data storage, April 29, 2021. https://cen.acs.org/biological-chemistry/dna/Method-offers-new-approach-DNA/99/i16
Software	open-access	The DeviceProfiler (DevPro) program by Michael Tobiason calculates the fitness of an existing set of DNA oligonucleotides (oligos). https://github.com/MTobiason/Sequence-Analysis
Software	open-access	The SequenceEvolver (SeqEvo) program by Michael Tobiason generates fit sets of DNA oligos. https://github.com/MTobiason/Sequence-Analysis
Software	open-access	The NAM program by Golum Mortuza encodes, decodes, and performs error correction on nucleic acid memory. https://github.com/BoiseState/NAM
Presentation	poster	Amanda Wolf, Sarah E. Kobernat, Luca Piantanida, Eric J. Hayden. DNA origami FRET Ruler for Nucleic Acid Memory. Idaho Conference on Undergraduate Research, virtual, July 21-22, 2021.
Presentation	poster	Benjamin Balzer, Amanda Wolf, Sarah E. Kobernat, Luca Piantanida, Eric J. Hayden. Enhancement of Digital Nucleic Acid Memory by Customizing DNA Origami Scaffolds. Idaho Conference on Undergraduate Research, virtual, July 21-22, 2021.
Presentation	oral	Christopher M. Green, William L. Hughes, Elton Graugnard and Wan Kuang. Correlative DNA-PAINT/AFM Microscopy of DNA Nanostructures and Characterization of Addressable Sites, FNANO 2021: 18th Annual Conference Foundations of Nanoscience, April 12–14, 2021.
Presentation	oral	George D. Dickinson, Golam Md Mortuza, William Clay, Luca Piantanida, Christopher M. Green, Chad Watson, Eric J. Hayden, Tim Andersen, Wan Kuang, Elton Graugnard, Reza Zadegan and William L. Hughes. Digital Nucleic Acid Memory: A New Approach to DNA-based Data Storage, FNANO 2021: 18th Annual Conference Foundations of Nanoscience, April 12–14, 2021.

Presentation	oral	Luca Piantanida. Digital Nucleic Acid Memory, ASU Center of Molecular Design & Biomimetics Annual Symposium, June 6-9, 2021.
Presentation	oral	G.D. Dickinson, G.M. Mortuza, W. Clay, L. Piantanida, C.M. Green, C. Watson, E.J. Hayden, T. Andersen, W. Kuang, E. Graugnard, R. Zadegan, and W.L. Hughes, "Digital Nucleic Acid Memory" *TECHCON*, September 15-17, 2020.
Presentation	poster	Sarah Kobernat, George Dickinson, William Clay, Luca Piantanida, Chad Watson, Tim Andersen, Wan Kuang, William Hughes, and Eric Hayden. "Improving DNA origami through scaffold optimization" *26th International Conference on DNA Computing and Molecular Programing*, September 14–17, 2020.
Presentation	poster	Sarah Kobernat, George Dickinson, William Clay, Luca Piantanida, Chad Watson, Tim Andersen, Wan Kuang, William Hughes, and Eric Hayden. "Improving DNA origami through scaffold optimization" *26th International Conference on DNA Computing and Molecular Programing*, September 14–17, 2020.
Presentation	oral and poster	M. Tobiason, B. Yurke, and W.L. Hughes (2019). Engineering Kinetically Uniform DNA Devices, 2019 TECHCON conference presented by the SRC. Austin, Texas.
Presentation	poster	M. Tobiason, B. Yurke, and W.L. Hughes (2019). Engineering Kinetically Uniform DNA Sequences, 25th *International Conference on DNA Computing and Molecular Programing (DNA25)*. Seattle, Washington.
Presentation	poster	M. Tobiason, B. Yurke, and W.L. Hughes (2019). Engineering Kinetically Reproducible DNA Devices, 16th *Annual Conference on Foundations of Nanoscience: Self-*Assembled *Architectures and Devices (FNANO19).* Snowbird, UT.
Presentation	poster	M. Tobiason, B. Yurke, and W.L. Hughes (2019). Engineering Kinetically Reproducible DNA Devices, *Semiconductor Research Corporation: Semiconductor Synthetic Biology (SemiSynBio) Annual Review*. College Park, MD.
Presentation	oral and poster	C.M. Green, M. Tobiason, R.M. Zadegan, W.L. Hughes (2019). Nucleic Acid Memory, *Semiconductor Research Corporation: Semiconductor Synthetic Biology (SemiSynBio) Annual Review*, College Park, MD.
Presentation	poster	C.M. Green, G. Dickinson, R.M. Zadegan, W.L. Hughes, E. Graugnard, W. Kuang (2019). Correlative metrology and defect analysis of DNA origami, *Semiconductor Research Corporation: Semiconductor Synthetic Biology (SemiSynBio) Annual Review*, College Park, MD.
Presentation	oral	W. Clay, G. Dickinson, L. Piantanida, C. Watson, W.L. Hughes, W. Kuang, "Real-Time Drift Correction for Super-Resolution Microscopy using Multiple Tracking Markers" *TECHCON*, September 15-17, 2020.
Presentation	oral	Christopher M. Green, William L. Hughes, Elton Graugnard and Wan Kuang. Correlative DNA-PAINT/AFM Microscopy of DNA Nanostructures and Characterization of Addressable Sites, FNANO 2021: 18th Annual Conference Foundations of Nanoscience, April 12–14, 2021.
Dissertation	published	Green, Christopher Michael, Nanoscale Optical and Correlative Microscopies for Quantitative Characterization of DNA Nanostructures (2019). *Boise State University Theses and Dissertations*. 1639.
Dissertation	published	Tobiason, Michael D., In Silico Sequence Optimization for the Reproducible Generation of DNA Structures (2019). *Boise State University Theses and Dissertations*. 1614.
Dissertation	published	Burden, Steven J., The Development of Nucleic Acid Biosensors with Allosteric Fluorescence Signals (2019). *Boise State University Theses and Dissertations*. 1627.
Dissertation	published	Suyehira, Kelsey, "Using DNA For Data Storage: Encoding and Decoding Algorithm Development" (2018). *Boise State University MS Thesis*.10.18122/td/1500/boisestate
Company	launched	Facible is a purpose-driven biodiagnostic technology company focused on a new hospital-grade fast track test that offers speed, accessibility, and accuracy. Steven Burden is the founder and CEO of the company, which includes 25 employees.

Based on the above listed outcomes, the Nucleic Acid Memory (NAM) Institute at Boise State was invited to join the DNA Data Storage Alliance. The alliance is the first and most extensive bridge between industry and academic organizations that are pioneering DNA data storage. Its mission is to "*create and promote an interoperable storage ecosystem based on DNA as a data storage medium*". The alliance will recommend the creation of specifications and standards (e.g., encoding, reliability, retention, file systems) which enable end-users to add interoperable DNA-based storage solutions to their existing storage hierarchies. The founders include Illumina, Twist Biosciences, Western Digital, and Microsoft. Member organizations include but are not limited to: Ansa Biotechnologies, Battelle, Catalog, The Cloude Nobs Foundation, DNA Script, EPFL, ETH Zurich, Imagene, IMEC, Iridia, Kioxia, Molecular Assemblies, PFU, Quantitative Scientific Solutions, Quantum, Seagate, Semicondcutor Research Corporation, Spectra Logic, University of Arizona, University of Washington, Digital Preservation, Oligo Archive, Lost Alamos National Laboratory, Cinémathèque Suisse, 21e8, DNAli, and University of Marburg. This network is critical as Boise State attempts to license the NAM intellectual property and/or the research team spins-off companies in the memory/biotechnology arena.

Supported by this research project, Steven Burden (founder/CEO) and Clementine Gibard Bohachek (co-founder/CSO) spun-off Facible, a biodiognostics company in Boise that has 25 employees and is seeking FDA approval for a novel COVID-19 screening technology. In service to future generations of biotechnology start-ups in Idaho, Facible and the Nucleic Acid Memory Institute are actively exploring the creation of a biotech incubator in Boise.

[1] V. Zhrinov, R. Zadegan, G. Sandu G.M. Church, W.L. Hughes, "Nucleic Acid Memory," **Nature Materials**, 15, 366-370 (2016). doi.org/10.1038/nmat4594

[2] M. Tobiason, B. Yurke, W.L. Hughes, "Generation of DNA Oligonucleotides with Uniform Structure Formation," **Nucleic Acid Research**, Submitted 2021.

[3] C.M Green, W.L. Hughes, E. Graugnard, W. Kuang, "Correlative DNA-PAINT/AFM Microscopy for Characterization of Strand Defects in DNA Nanostructures," **ACS NANO**, Accepted 2021 (nn-2021-01976x).

[4] G.D. Dickinson, G.M. NatureMortuza, W. Clay, L. Piantanida, C.M. Green, C. Watson, E.J. Hayden, T. Andersen, W. Kuang, E. Graugnard, R. Zadegan, and W.L. Hughes; An Alternative Approach to Nucleic Acid Memory, **Nature Communications**, vol 12, number 2371, 2021 (doi.org/10.1038/s41467-021-22277-y).

II. Technical Summary

To realize our vision, we report the successful use of in-silico fitness score optimization to generate DNA sequences for dNAM with similar hybridization rates. Three optimization criteria were utilized: a network fitness score (N) where points are accumulated for inter-oligo secondary structures, an oligo fitness score (O) where points are accumulated for intra-oligo secondary structures, and a class of weighted fitness scores (W) which are linear combinations of N and O. Hybridization rates for both optimized and non-optimized oligo-sets were experimentally characterized and compared. A total of 144 hybridization rates were measured using fluorescent quenching and reported. For a duplex-formation reaction, W-fit oligo-sets were found to exhibit Arrhenius temperature dependence with consistent Arrhenius parameters. However, non-optimized oligo-sets exhibited an Arrhenius temperature dependence with variable Arrhenius parameters for the same duplex-formation rate. Optimization was observed to substantially decrease hybridization-rate variation, with three W-fit oligo-sets exhibiting typical hybridization-rate dispersions of ± 7.7% (duplex-formation) and ± 14% (strand-displacement). For the duplex-formation of both optimized and non-optimized oligo-sets, a very strong linear relationship between the two Arrhenius parameters was observed, indicating that this model may be over parameterized. For comparison, an alternative model describing the experimental data using a single variable parameter was derived. Further analysis of hybridization rates reported in the literature indicated a statistically significant ($p < 0.05$) correlation between decreasing O values and decreasing hybridization-rate dispersion in five separate datasets. This work has been submitted to Nucleic Acids Research[2]. The resulting computer programs created for this study (DevPro / SeqEvo) are freely available for academic use—and can design NAM prototypes with predictable kinetic performance.

To further realize our vision, we developed a metrology technique for analyzing defects in DNA-origami that combined super resolution microscopy (SRM) and atomic force microscopy (AFM); achieving strong correlations between structures visualized with both tools[3]. With the ability to detect single molecules, we resolved data sites with: (**1**) no observed defects in AFM and SRM (74 ± 2%), (**2**) defects observed in AFM only (8 ± 2%), (**3**) defects observed in SRM only (16 ± 1%), and (**4**) defects observed in AFM and SRM (2 ± 1%). In doing so, we observed that unresolved data sites in the SRM images are not strongly correlated with defects seen with AFM, revealing that most site defects do not arise from unicorporated strands but from strands that are present in the structure, and are most likely damaged due to photo-oxidation and UV damage. Our analysis indicates that there is significant room for progress in the design of data sites to overcome strand defects. We believe this method, in conjunction with the software tools above, will extract defect mechanisms and inform new design principles for increasing the yield and fidelity of NAM prototypes.

The culmination of the above listed research was published in Nature Communications[4], where we encoded '*Data is in our DNA\n*' into dNAM. In dNAM, data is encoded by selecting combinations of single-stranded DNA with (1) or without (0) docking-site domains. When self-assembled with scaffold DNA, staple strands form DNA origami breadboards. Information encoded into the breadboards is read by monitoring the binding of fluorescent imager probes using SRM. To enhance data retention, a custom multi-layer error correction scheme that combined fountain and bi-level parity codes was used. Each origami encoded unique data-droplet, index, orientation, and error-correction data. The error-correction algorithms fully recovered the message when individual docking sites, or entire origami, were missing. Our prototype achieved an areal density of 1000 Gbit/cm^2. After accounting for using 2/3 of the bits for indexing and error correction, this resulted in an areal data density of 330 Gbit/cm^2. Although relevant only for reading throughput, for comparison, recent advancements in tape report an areal information density of 31 Gbit/cm^2. Unlike other approaches to DNA data storage, reading dNAM did not require sequencing. As such, the technology offers a novel approach to explore NAM viability.

To improve our data density, we designed, built, and started to optimize a custom SRM system that is capable of 3 nm resolution. Our design is highly rigid with no moving parts and is compatible with the enclosures and vibration isolation tables traditionally used to stabilize scanning probe microscopes. We also replaced seqNAM with 3DNAM. Briefly described, 3DNAM integrates time-correlated SRM and DNA self-assembly to read non-volatile information with sub 5 nm lateral and 1 nm axial resolution. To enable time-correlated imaging measurements, we developed a TCI array in a 180nm semiconductor process provided by a commercial foundry (X-FAB). We also used an industry standard toolset (Cadence) to simulate and verify the design of our imager. X-FAB provided a comprehensive model of the SPAD to enable co-simulation with our design. The photon detection efficiency of our SPAD is around 25%, near the minimum required for our application. We anticipate improving this to 50% in our next revision due to refinements in the fabrication process by X-FAB. We submitted our design for fabrication in January 2021 and received the bare imager die in June 2021. We are in the process of packaging the die for benchtop characterization and then microscope integration. Back of the envelope calculations indicate that 3DNAM could have information densities above 10 Tbit/cm^2 and read speeds over 56 Tbit/day. It also has the potential to be used as a new sequencing technology. Regardless if this can be achieved, the time correlated imager we are creating has potential for commercial development. Single photon detection and precision timing capabilities are only available as bulky and low-throughput devices. Thus, by developing TCI, we are not only providing an enabling technology for 3DNAM, but are also addressing an unmet commercial need for this class of scientific instrumentation.

What follows are select project accomplishments that reinforce the technical summary.

III. Project Accomplishments

Objective 1: NAM Coding

1. Introduction

Objective 1 addresses the development and testing of algorithms for encoding and decoding information stored in NAM prototypes, which are robust to high levels of insertion, deletion, and substitution errors, and, in the case of seqNAM, which avoid biologically deleterious sub-sequences. Both objectives have been met, with encoding/decoding algorithms developed for both seqNAM and dNAM, and validated with in-silico simulations, as well as wet-lab experiments.

2. Deliverables

Objective	Deliverables
1A	An information encoding/decoding algorithm for dNAM (year 1)
1B	An information encoding/decoding algorithm for seqNAM (year 2)

3. Most significant results, outcomes and deliverables.

Deliverable 1A — An information encoding and decoding algorithm for dNAM. Robust dNAM-specific information encoding and decoding algorithms were developed and validated. With dNAM, the presence and absence of individual DNA molecules (staple strands) on a DNA-origami scaffold is used to store bit values. These bit values, *i.e.* the presence/absence of the DNA staple strands at various locations on the origami surface, are read using SRM, which is subject to relatively high read errors due to incomplete staple strand incorporation, defective imager strands, fluorophore bleaching, and background fluorescence. This high error rate necessitated the development of robust information encoding and decoding algorithms, combining multiple levels and strategies of error correction and error handling. Information is encoded using a fountain code, combined with a custom, bi-level, parity-based, and orientation-invariant error detection scheme (***Fig 1***). Fountain codes are an optimal mechanism for transmission of data over extremely noisy and unreliable channels, and work by dividing a data file into small segments, combining these segments via XOR into what are called droplets, and then sending the droplets at random to a receiver. Our algorithm encodes each droplet onto a single origami and adds additional bits of information for error correction to help ensure that individual droplets can be recovered. Together, the error correction and fountain codes increase the probability that the message can be fully recovered while minimizing the number of DNA origami that must be observed.

Through wet lab experiments, as well as through exhaustive simulation we have validated that the combination of the multi-level error correction strategy and fountain codes provide extremely robust storage and recovery of file information for dNAM. This approach took dNAM from an idea to practice.

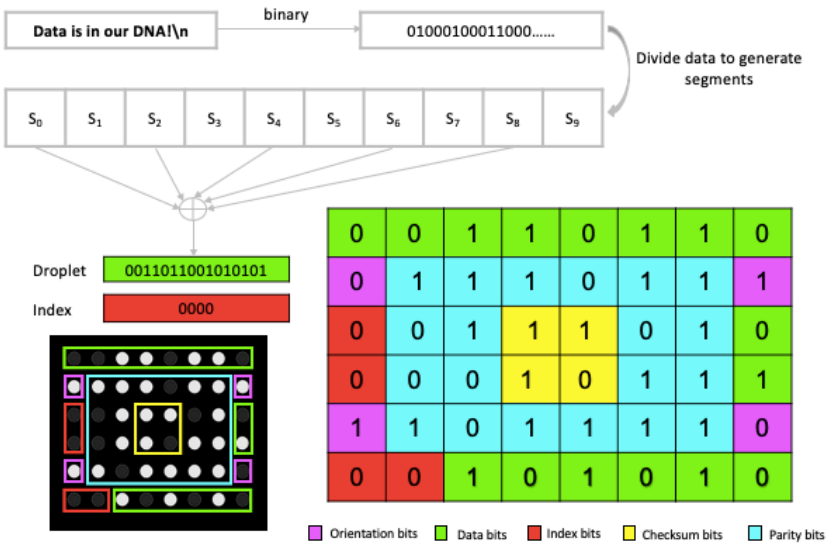


Figure 1. Example of Fountain Code implementation of dNAM digital encoding. The figure illustrates some of the main steps involved in encoding a digital message into dNAM. First a character string is divided into non-overlapping segments. These segments are combined in various patterns via an XOR operation to generate data droplets. Each droplet is assigned an index, error-correcting (checksum and parity) and orientation information and positioned within a grid to form the design used to synthesize a dNAM origami. Origami index 0 is depicted from the prototype.

Deliverable 1B — An information encoding and decoding algorithm for seqNAM. When encoding binary data into sequences representing DNA strands, the algorithms should account for biological constraints representing the idiosyncrasies of working with a molecular substance. In response, we developed REDNAM (Robust Encoding and Decoding of Nucleic Acid Memory). REDNAM includes a novel mapping scheme and translation stage which converts hexadecimal data to codons while accounting for four constraints: removing start codons, avoiding repeating nucleotides, excluding longer repeating sequences, and maintaining close to 50% GC content. We have integrated this mapping scheme into the fountain code algorithm to balance information density with error correction and parity data.

The primary innovation of REDNAM is the mapping approach, which is inspired by the codon to protein mapping scheme used in living cells. Uniquely, our codon-base mapping removes biologically active sequences—such as start codons and some known promoter regions—avoids multiple repeats of unique nucleotides, and excludes repeating sequence strings. This promotes more robust encoding and decoding of the information stored in the DNA, as it avoids structural problems that lead to synthesis and sequencing errors, and is also safer from a biological perspective, as our algorithm avoids the occurrence of start and other codons involved in transcription and translation.

As with dNAM, our implementation of REDNAM is used in combination with a fountain code to increase robustness. ***Figure 2*** shows the basic steps in the information encoding/decoding process. The fountain code algorithm is ideal because the mapping

scheme is entirely separate from the rest of the encoding and decoding processes. This allows us to easily include our mapping and translation stage, which takes more biological constraints into account.

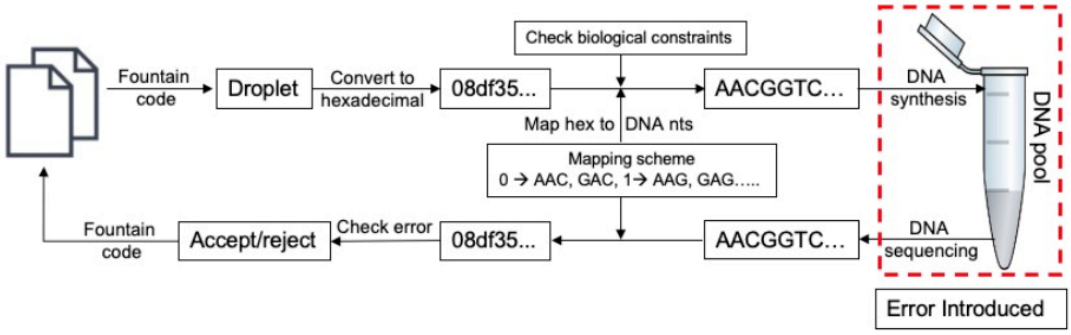


***Figure 2**. Overview of the REDNAM DNA based storage system.*

To validate the REDNAM algorithm, we encoded a JPEG file of size 13,170 bytes, resulting in 604 DNA sequences, each 250 nucleotides long. The synthesized sequences were sequenced using the Illumina Miseq platform, producing 78 million reads of sequences, where 5M reads were unique to the pool (***Fig 3***). To check the robustness of our algorithm, we sub-sampled portions of the 78 million reads and tested the decoders ability to recover the original message. For sub-samples greater than 9000 reads, the decoder successfully recovered the file 100% of the time. While this seems like a large number of required samples, it is primarily due to the repetition of the reads in the sub-samples, as some sequences occur much more frequently than others.

To further test the robustness and efficiency of our algorithm, we performed simulations on randomly generated files, testing the encoding and decoding of files ranging from 1 to 49 MB at 1 MB intervals, where each of the encoded files was subjected to varying levels of simulated errors—including insertion, deletion, or mutation of any random nucleotide or even total deletion of any random sequence. In all cases, the decoding algorithm was able to recover missing/corrupted data.

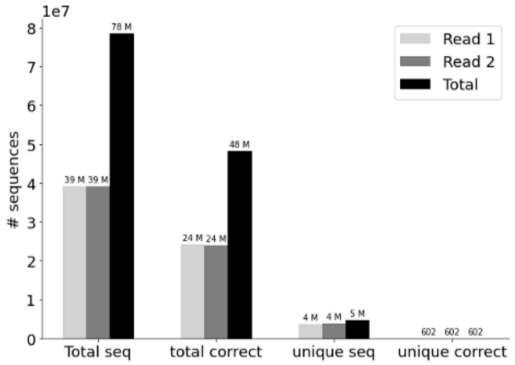


***Figure 3.** Sequence frequency distribution among two reads.* In total there were 78 million sequences read where 5 million reads were unique. Out of all the reads, 62% of the reads were correct.

Objective 2: NAM Sequences

1. Introduction

For NAM applications, it is ideal to have uniform hybridization rates. For these purposes, hybridization-rate variation results in inconsistent synthesis or inconsistent kinetics during DNA-PAINT. *Objective 2* focused on improving sequence-performance relationships and improving our ability to generate new sequences for NAM devices. Important outcomes include the creation of: (**1**) three high-performing metrics for *in-silico* sequence optimization, (**2**) one computer program (SeqEvo, which optimizes these metrics), (**3**) one kinetic model describing the temperature-dependence of DNA duplex-formation rates, and (**4**) new sequences for future 2D and 3D-NAM devices. Together these outcomes increase our predictive capacity for engineering NAM systems.

2. Deliverables

Objective	Deliverables
2A	A high-throughput data pipeline for metrics-based optimization of data cell anchor strands (year 1)
2B	Sets of optimized codons and words for seqNAM data encoding (year 2)
2C	Optimized insertion sequences for custom scaffold synthesis (year 3)

3. Most significant results, outcomes and deliverables.

Three fitness scores were developed for quantifying inadvertent hybridization reactions a given oligo-set may undergo: (**1**) a network fitness score (N) – which quantifies inadvertent inter-oligo simple secondary structures; (**2**) an oligo fitness score (O) – which quantifies inadvertent intra-oligo simple secondary structures; and (**3**) a weighted fitness score (W) – which is a linear combination of N and O.

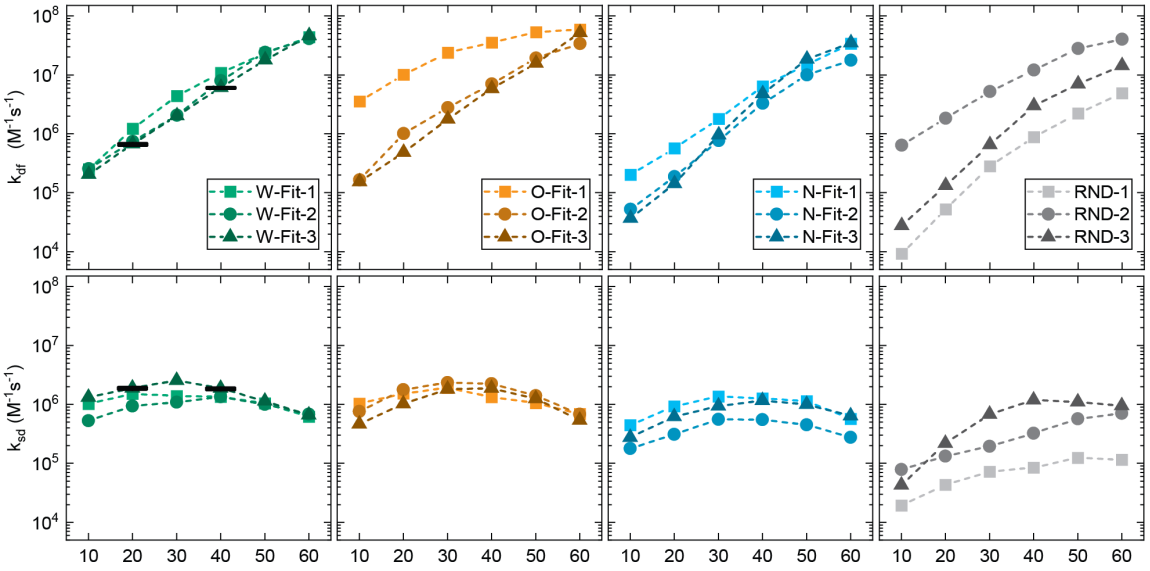


Figure 4. *Hybridization rates were experimentally measured for 12 unique oligo-sets (aka "DNA sequences").* These sets were named according to the method used to generate them (i.e. the W-Fit-1 oligo-set was the first generated using optimization of the W fitness score). Rates were measured for six experimental temperatures, for both a duplex-formation (k_{df}, above) and a strand-displacement reaction (k_{sd}, below).

A model DNA system was used to study the relationship between DNA sequence and hybridization-rates. *In-silico* optimization of the three fitness scores were used to generate novel sets of oligos. In total, twelve oligo-sets (*i.e.*, DNA sequences) were generated and experimentally characterized. 144 hybridization-rate measurements (***Fig 4***) were collected, which enabled the study of hybridization-rate dispersion as a function optimization criteria (***Fig 5***). Optimization of the W-fitness score was observed to result in low hybridization rate dispersions. Typical dispersions of ±7.7% for the duplex-formation reaction and ±14% for the strand-displacement reaction were observed for the oligo-sets.

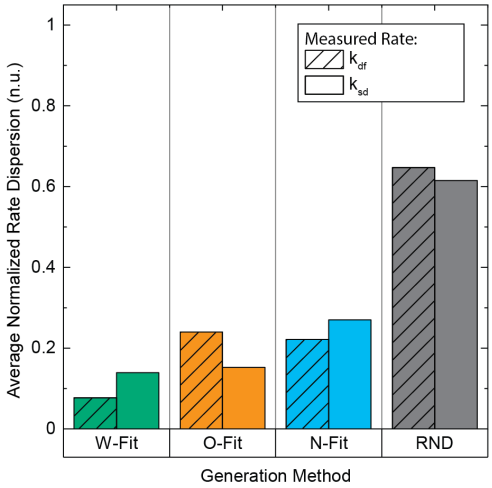


Figure 5. *In-silico optimization of the W fitness score was observed to yield oligo-sets with the most favorable hybridization-rate dispersions.*

Duplex-formation rates were observed to exhibit an Arrhenius temperature dependence (***Fig 6***) with strongly correlated activation energy (E_a) and pre-exponential factor (A). This enabled the derivation of a new kinetic model for the duplex-formation reaction, which reduced the number of sequence-dependent parameters necessary to predict duplex-formation rates from 2 to 1. This new model simplifies the task of predicting *in-silico* hybridization-rates. In addition, the hybridization rates observed for the sampled sequences (***Fig 4***) suggest that this parameter results from inadvertent intra-oligo structures, and that the reaction rates can be accurately predicted *in-silico*.

The SeqEvo code was designed to generate sequences for relatively small networks of interacting DNA oligos. In order to generate oligo-sets large enough for NAM, refactoring of the code was necessary. This process reduced the time to calculate the N, O, or W fitness scores, enabling larger oligo-sets to be generated. By the end of this objective, run-time was sufficiently low to enable the design of an oligo-set containing 8,000 total bases and 517 total oligos. The refactored SeqEvo code was used to generate four new oligo-sets for novel designs (***Table 1***). This included new sequences for a 10x10x10 DNA molecular canvas,[a] which is an attractive substrate for future 2D-NAM and 3D-NAM devices. It is speculated that this structure will have hybridization-rate dispersions similar to those observed in ***Fig 5*** (i.e. conservatively estimated at ± 14%).

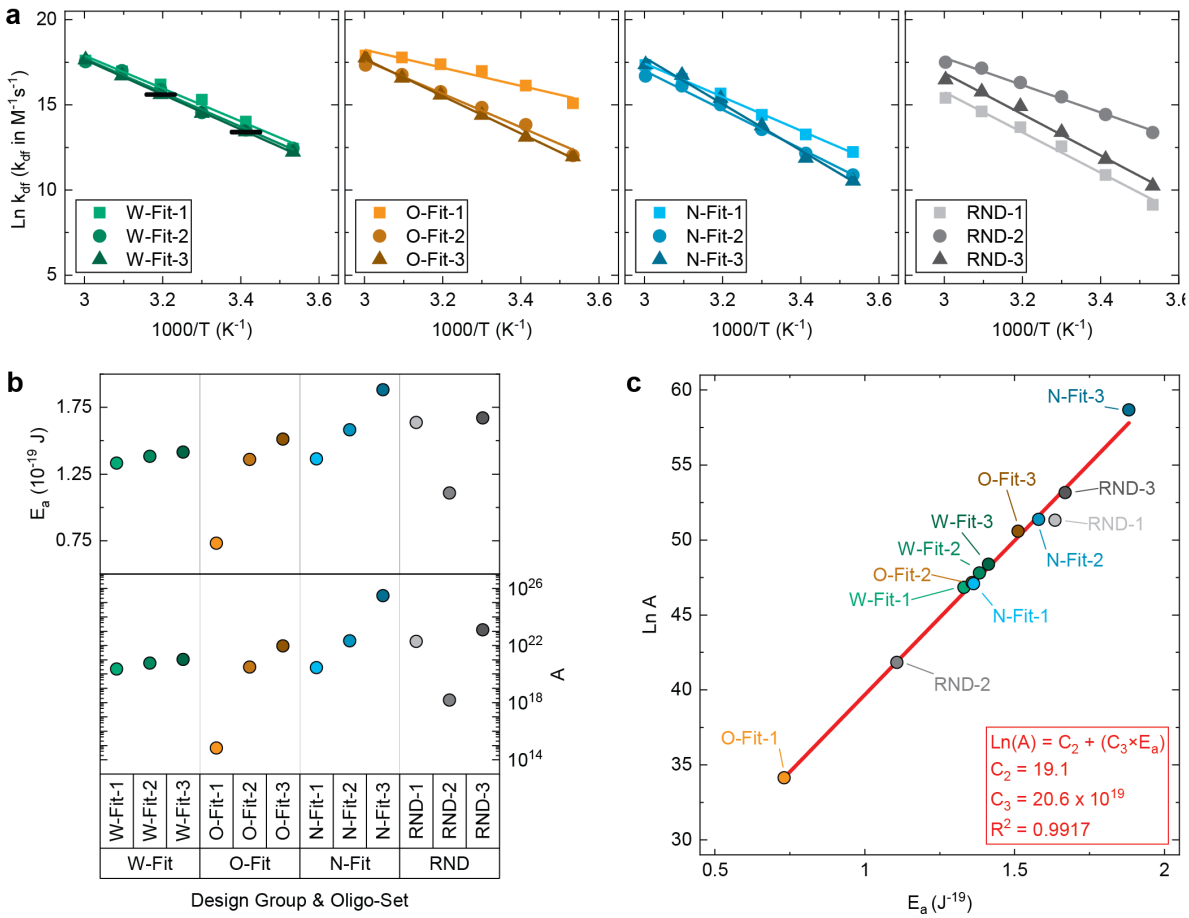


Figure 6. *Duplex formation rates for the twelve oligo-sets were found to exhibit an Arrhenius temperature dependence with strongly correlated values of the activation energy (E_a) and pre-exponential (A).* This enabled the derivation of a new kinetic model for this reaction which reduced the number of variable parameters necessary to predict duplex-formation rates from 2 to 1.

Table 1. *Oligo-sets generated for four designs using SeqEvo.* The source of these designs included: [a] Ke, Y.G., Ong, L.L., Shih, W.M. and Yin, P. (2012) Three-dimensional structures self-assembled from DNA bricks. Science, 338, 1177-1183; [b] Qian, L. and Winfree, E. (2011) Scaling up digital circuit computation with DNA strand displacement cascades. Science, 332, 1196-1201; [c] Kotani, S. and Hughes, W.L. (2017) Multi-Arm Junctions for Dynamic DNA Nanotechnology. J Am Chem Soc, 139, 6363-6368; and [d] Zhang, D.Y., Turberfield, A.J., Yurke, B. and Winfree, E. (2007) Engineering entropy-driven reactions and networks catalyzed by DNA. Science, 318, 1121-1125.

Design	№ Oligos	Oligo-Set	*N*	*O*
10x10x10 Canvas [a]	517	As published	2.22 x 10^{25}	2.15 x 10^{8}
		W[10^{8}] Optimized	1.96 x 10^{11}	1.94 x 10^{6}
Four-Input OR Network [b]	45	As published	2.71 x 10^{22}	8.08 x 10^{4}
		W[10^{4}] Optimized	3.13 x 10^{10}	5.18 x 10^{4}
Autocatalytic Network [c]	10	As published	1.11 x 10^{45}	6.35 x 10^{5}
		W[10^{6}] Optimized	3.97 x 10^{7}	9.78 x 10^{4}
Autocatalytic Network [d]	6	As published	1.32 x 10^{6}	7.50 x 10^{4}
		W[10^{2}] Optimized	4.10 x 10^{5}	1.51 x 10^{4}

Objective 3: NAM Scaffolds

1. Introduction

Objective 3 addresses the design, construction and production of single stranded DNA used as scaffolds for the DNA origami that dNAM is built upon. We have produced several ssDNA scaffolds of different sizes in E. coli. We have designed and are building a larger ssDNA scaffold by shuffling together the DNA from these different sized scaffolds. This larger scaffold will be used to synthesize a dNAM node with increased per node data.

2. Deliverables

Objective	Deliverables
3A	Create viable phage modules from E. coli
3B	Validate phage genomes for use in DNA origami synthesis

3. Most significant results, outcomes and deliverables.

Use viable phage modules to build larger scaffolds for improved data density. In our dNAM design, several data sites are used for orientation and error correction. Larger origami could include more data sites used for encoded information. Towards this goal, we have designed a scaffold that can produce a 78x120nm origami, which is 60% larger in area than our previously demonstrated structure. This larger origami can achieve an 8x10 data grid with two extra rows and columns of data compared to our previous 6x8 data grid. Synthesis of this scaffold is underway, as described next.

In our scaffold production approach, E. coli are transformed with phagemids, which are circular DNA plasmids that can produce ssDNA upon subsequent infection with "helper phage". We developed a strategy to build the DNA phagemid needed to produce the larger scaffold by shuffling together parts of smaller phagemids. Modular phagemid parts are extracted from smaller scaffolds by PCR, and can be combined back together in numerous combinations to meet design requirements. For our design goal of a larger scaffold, three unique 11,054 nt scaffolds were designed using a 10,080 bp phagemid combined with inserts from the smaller phagemids. Distinct 982 bp regions of the 3kb, 5kb, and 8kb phagemids were amplified with PCR primers that add KpnI and BglII cleavage sites. They were then cut with restriction enzymes and pasted into the matching restriction sites in a 10,080 bp phagemid using DNA ligase. These ligated phagemids were transformed into E. coli which will be grown and screened for the desired size. Scaffold production with helper phage when folded into an origami rectangle, the 11kb dNAM contains 8x10 data sites, compared to the original 6x8 dNAM (60x90nm); increasing the operable size by 60%.

Objective 4: NAM Fabrication

1. Introduction

Objective 4 addresses the design, build, and test of NAM prototypes. dNAM was created and while outside the scope of this project, 3D-NAM has been designed and is currently under test. Single-molecule defect analysis was performed by correlating SRM and AFM together; which was enabled by creating cross-compatible substrates.

2. Deliverables

Objective	Deliverables
4A	Process for design and fabrication of DNA origami storage nodes (year 1)
4B	Protocols for statistical correlated AFM/SRM defect metrology in DNA origami (year 2)
4C	Protocols for selective area immobilization of DNA origami (year 3)

3. Most significant results, outcomes and deliverables.

dNAM was successfully designed, built, and tested. Our first prototype included data domains spaced at 10 nm intervals to achieve an areal density of 1000 Gbit/cm^2. After accounting for using 2/3 of the bits for indexing and error correction, this resulted in an areal data density of 330 Gbit/cm^2. Although only comparable for reading throughput, not storage, this is significant because recent advancements in magnetic tape have reported a two-dimensional areal information density up to 31 Gbit/cm^2, though the current commercially available material typically has lower density. Our dNAM prototype is the first and only example of reading and writing DNA without the need for sequencing technology. It is also the only DNA-based memory system that does not require custom sequences to be synthesized to change the encoded and decoded message.

3D-NAM is a newly proposed modification to dNAM that has the potential to provide an order of magnitude higher information density. This technique relies on super resolution microscopy to perform spatial and temporal readout with sub 5 nm lateral and 1 nm axial resolution. While outside of the scope of this project, the DNA nanostructures for 3D-NAM have been designed and are currently being tested. This work, if successful, is significant because it would enable a new way to sequence DNA, at a single-molecule level, without signal amplification.

In Support of dNAM and 3D-NAM, an accessible strategy for high resolution, correlative DNA-based points accumulation for imaging in nanoscale topography (DNA-PAINT) super-resolution and atomic force microscopy (AFM) of DNA nanostructures was created, enabled by a simple and robust method to selectively bind DNA origami to cover glass. Using this correlative microscopy technique, addressable “docking” sites on DNA origami were examined to distinguish between two defect scenarios – structurally incorporated but inactive docking sites, and unincorporated docking sites. In addition to creating a new microscopy technique, the results are significant because over 75% of defective docking

sites were incorporated but inactive, suggesting unincorporated strands played a minor role in limiting the availability of addressable sites. The effects of strand purification, UV irradiation, and photooxidation on availability were also explored, providing insight on potential sources of defects and pathways towards improving the fidelity of DNA nanostructures for 2D and 3D-NAM.

To enable the above listed outcomes, cross-compatible SRM and AFM substrates that combined transparency, favorable DNA origami adsorption, low affinity for single-stranded DNA imager strands, and near atomic-level flatness were created via glow discharge. The results were validated by prior observations of DNA origami adsorption to piranha/HF-cleaned, thermally-grown silica, for which it was postulated that pH-dependent adsorption resulted from the deprotonation of silanol groups generated during cleaning. This is significant because DNA origami are typically bound to cover glass by biotin-avidin binding between biotinylated DNA present in the origami and surface-bound, biotinylated proteins (commonly biotinylated bovine serum albumin – BSA-biotin). While the surface proteins passivate the surface to diffusing imager strands during image acquisition, they are too rough to perform high-resolution AFM and/or SRM.

Objective 5: NAM Reading

1. Introduction

Objective 5 addresses NAM readout. The information stored in DNA is read by a super-resolution fluorescent microscopy technique. We achieved optical resolution as small as 5 nanometers to enable high areal density data storage. We also worked with Objectives 2-4 to create an optimized set of sequences that produces the least imaging defects. However, the imaging resolution is insufficient for seqNAM. A change in direction (see *III. Significant Changes in Direction*) was introduced in 2020 to address the resolution with a 3D super-resolution technique.

2. Deliverables

Objective	Deliverables
5A	Optical readout of dNAM (year 1)
5B	Optical readout of seqNAM (year 2)
5C	Sub-nanometer imaging resolution for SRM (year 3)

3. Most significant results, outcomes, and deliverables.

NAM readout with the use of super-resolution imaging of DNA. The result is published recently in Nature Communications[3].

Image drift is one of the limiting factors for single-nanometer resolution super-resolution imaging. Several improvements are made to the real-time active drift correction system. A temperature stabilized single-frequency laser source is utilized to increase the wavelength stability. The position of the slides is determined in all three dimensions by imaging the reference gold nanoparticles with a resolution of 3 nanometers. The resulting resolution as small as 4.5 nanometers is achieved which is important for NAM data density.

An additional pathway to improving resolution is through the elimination of mechanical and optical noise from the microscope system. Commercial optical microscopes have many moving parts that are susceptible to both vibration and drift coming from mechanical and thermal sources. They also have external light leaks, internal stray light, and internal optics that are not needed for DNA-PAINT microscopy. To improve performance, a custom SRM system has been designed. The design eliminates all macroscopic moving parts, is mechanically rigid, and uses a minimal number of components so to minimize mechanical sources of drift and noise. The design is also light-tight and designed to minimize stray light and optical loss internally. Additionally, the system's compact and lightweight design will make it possible to integrate it onto mechanical and thermal isolation systems used for AFM and other forms of high-sensitivity microscopy. A rendering of the mechanical design is shown in ***Figure 7A***.

The prototype design has been constructed and tested on our dNAM samples. An averaged image of a dNAM tile is shown in ***Figure 7B***. The system hasn't been integrated with mechanical or thermal isolation, but does utilize the active drift correction system described above. The system has not been fully optimized in terms of the illumination and experimental conditions but our initial experiment with the system has shown 4.5 nm FWHM resolution. This is superior to our best results of 7.2 nm on the commercial system, and we're optimistic more improvements can be made through optimization.

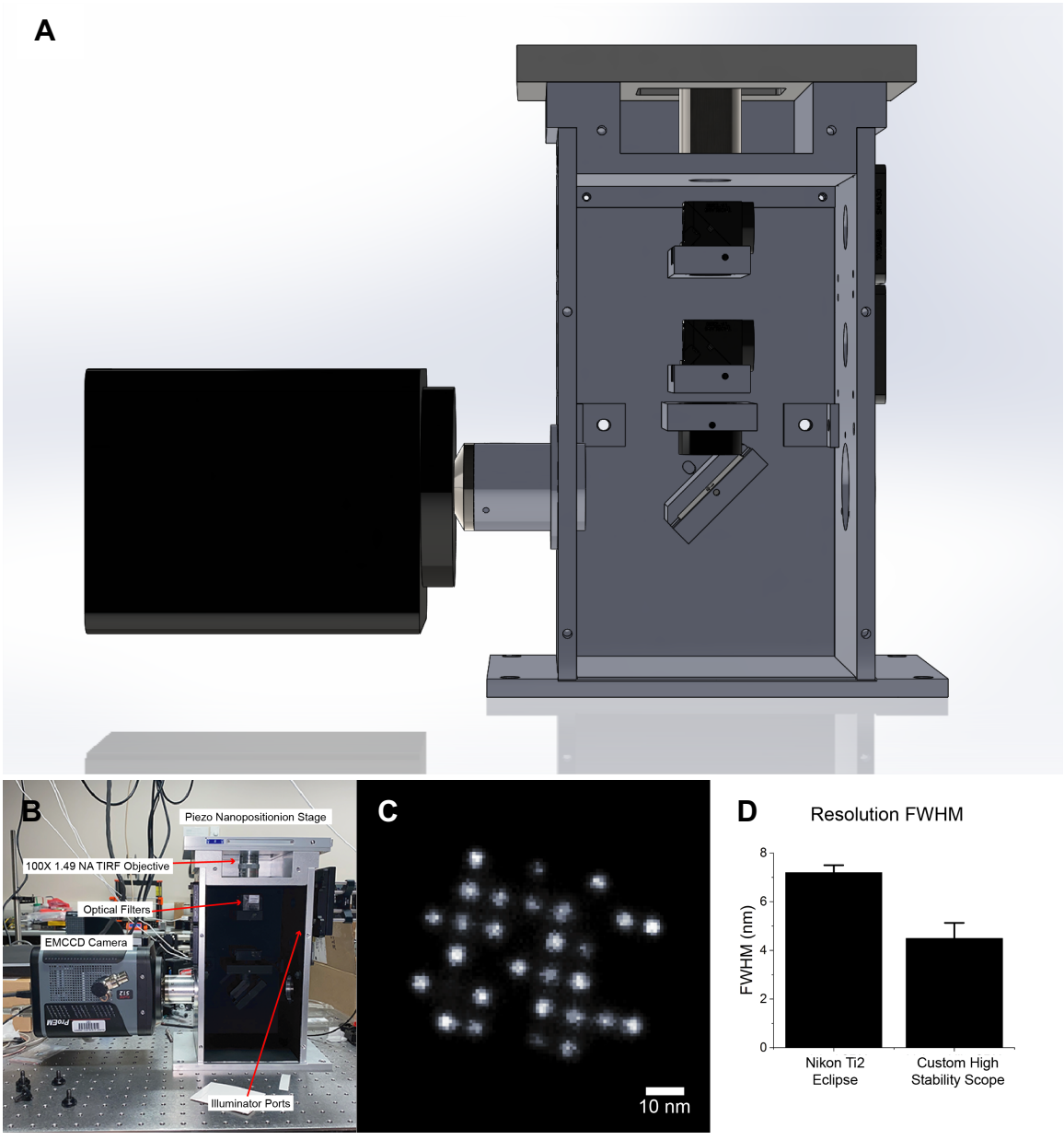


Figure 7. (**A**) CAD rendering of custom microscope design. showing third-party components attached to custom body. (**B**) Fabricated Microscope. (**C**) Super-resolution image of DNAM Tile recorded with custom scope. Showing 4.5 nm FWHM resolution. (**D**) Head-to-head resolution comparison between a highly optimized Nikon SRM system and the new custom SRM system.

IV. Significant Changes in Directions

Objective 1. None to date.

Objective 2. The initial strategy proposed for this objective was to build upon the previously-developed SeqEvo computer program to generate new NAM sequences. Experimental results collected while validating the SeqEvo software created a feedback-loop which changed the course of this research. As a result, further effort than anticipated was spent studying the kinetics of DNA oligo hybridization and developing new methods of in-silico optimization.

Objective 3. While we had originally planned to modify phage directly, the phagemid and helper phage approach was determined to be more modular. In addition, we are able to use numerous phagemids developed by others as modular sources of DNA that can be shuffled together without concern of reproduction effects.

Objective 4. While protocols for immobilization of DNA origami were achieved (described above), selective area protocols proved unnecessary for two algorithmic reasons: (**1**) custom pattern-recognition software was created that distinguished, rotated, and registered NAM structures so that they could be successfully read, and (**2**) custom encoding and decoding algorithms enabled reading NAM structures, with multiple messages, without needing to physically partition the data via typical approaches to random access.

Objective 5. In the project, an increased localization accuracy and enhanced drift correction are achieved. Theoretically, localization accuracy can be continuously improved by increasing photons collection since the imager strands are continuously replenished from solution. In practice, it is observed that the docking sites can become unavailable after 10s of minutes of imaging due possibly to photo damage. It limits how much further the resolution can be reduced. Sub-nanometer imaging resolution may not be achieved with resolution enhancement alone. Instead, $_{3D}$NAM is being explored to achieve an effective nanometer resolution. The data imager strands are designed in a way that imager strands will be attached at multiple distance from the donor fluorophore.

V. Future Directions

Objective 1.

dNAM. We will examine more advanced error correction codes in order to reduce the space devoted to error correction on the origami. We will determine how to utilize the information in the SRM image to guide the decoding algorithm as it corrects errors. i.e. look at how we can prioritize bits in the error correction search based on the information that supports that bit as found in the SRM image. Finally, we are working on developing deep NN based algorithms for reading the SRM image.

seqNAM. We are currently researching how we can improve the encoding algorithm by viewing the encoding process as the search for a shortest path in a weighted graph. The approach we are developing starts by constructing a weighted graph from the local constraint matrix and the information to be encoded, and then proceeds to search for the shortest paths through this weighted graph using a uniform-cost search. We are also interested in how we can incorporate non-local constraints in this process, such as the avoidance of long repeats or palindromic sequences, or other structural issues. To this end, we are exploring using a transformer deep learning model to speed structure problem prediction during the graph search, so that this can be run real-time during the search to provide guidance to the shortest path algorithm.

Objective 2. As a result of this research, it is now possible to generate high-quality DNA sequences for relatively large oligo-sets for NAM-based memory systems, as well as other DNA-based systems. However, it is not yet clear how much this will improve synthesis yields and/or performance of NAM devices. A quantitative study of the performance of the newly generated NAM sequences will be necessary to validate and further improve the sequence generation process.

Objective 3. We will continue to synthesize our 8x10 dNAM using the larger scaffold. In addition, we plan to design larger structures using multiple orthogonal smaller scaffolds stitched together with staple strands.

Objective 4. During this project, three super-resolution microscopes have come online. The first is a commercial-grade SRM system with vibrational and environmental control. The second is a modularly built SRM system that is used to experiment with novel approaches to SRM including but not limited to time-correlated SRM for $_{3D}$NAM. And the third is a state-of-the art custom designed and built SRM system from the frame-up. In the future, the first microscope will be used to test new NAM prototypes in a controlled environment, the second will pilot new SRM techniques, and the third will push the resolution limits of SRM as close to its theoretical limit as possible. Together, they will explore $_{3D}$NAM and eventually a new approach to sequencing DNA.

As an important step in this direction, ***Figure 8*** is a calibration standard created on a dNAM substrate to probe the physical limits of SRM. The top and left images are three- and two-dimensional representations of the substrate, where the red and blue protruding strands are respectively orientation makers and data sites with various lengths. The SRM image in the center and the corresponding photon intensity versus distance plot on the right indicate that we have imaged length 1 through 7 nts in increments of 1 nt. As shown in the intensity plot, when the imager probe was a mini-hairpin structure, the short data sites could be read. In comparison, when the linear probe was used as a control, the data sites could not be read. The ability to image 1 nt with SRM is below the resolution of AFM. Looking to the future, we will validate or debunk our findings through careful design and execution of control studies. If validated, will also explore the imaging mechanism.

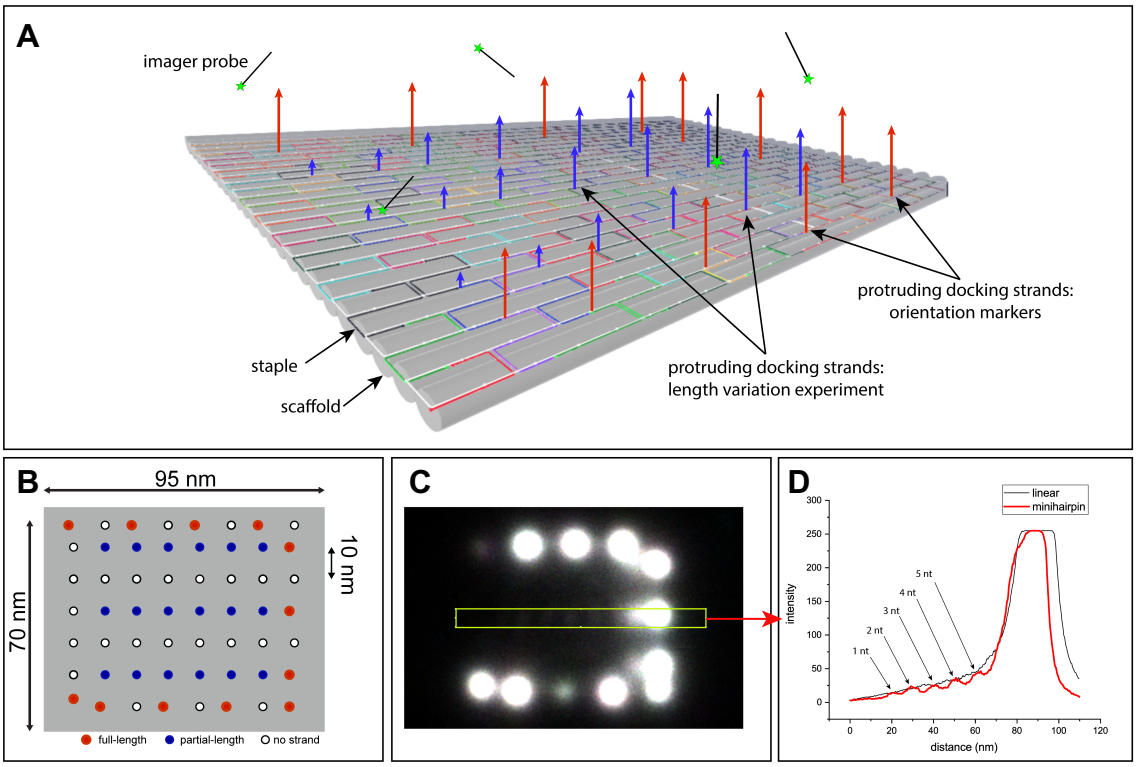


Figure 8. (**A**) Three and (**B**) two-dimensional representation of a dNAM substrate, where the red protruding strands are orientation makers and the blue protruding strands are data sites with various lengths. (**C**) Corresponding SRM image of a dNAm substrate with (**D**) the corresponding intensity versus distance profile.

Objective 5. To enable 3DNAM, we are developing a custom imaging array that combines high resolution, high light sensitivity, and high timing sensitivity. Conventional imagers use a lengthy exposure time to capture an image in a low-light environment, as we have in imaging the fluorescence of 3DNAM. Due to this exposure time, conventional imagers are incapable of extracting fluorescent lifetime. The time correlated imager (TCI) we developed uses single-photon avalanche diodes (SPADs) which have a binary response to a single photon, meaning we can extract the exact moment of a photon's arrival. While there are commercially available SPAD imagers, they are unsuitable for 3DNAM imaging

as they only have a single pixel or have poor photon detection efficiency. Fortunately, SPADs can be integrated directly into integrated circuit technology, meaning we can develop our own imager with supporting circuits to extract lifetime information. ***Figure 9A*** shows the cross section of the SPAD structure we used to implement our TCI. ***Figure 9B*** shows a simplified architecture of our first TCI prototype. It is comprised of a 16x16 SPAD array, column-level monostable circuits to stretch SPAD events and then reset the SPAD after photon detection, and shared time-to-digital converters (TDCs) that convert the photon's arrival time with respect to the laser into a digital code. Our TDC has a timing resolution of 62 ps and a selectable range to accommodate up to an 80MHz laser pulse repetition rate (***Figure 9C***). Our next prototype will include lifetime computation on-chip to compress data. We will use a center-of-mass method that directly extracts fluorescent lifetime with low computational overhead and is scalable to larger arrays (***Figure 9E***).

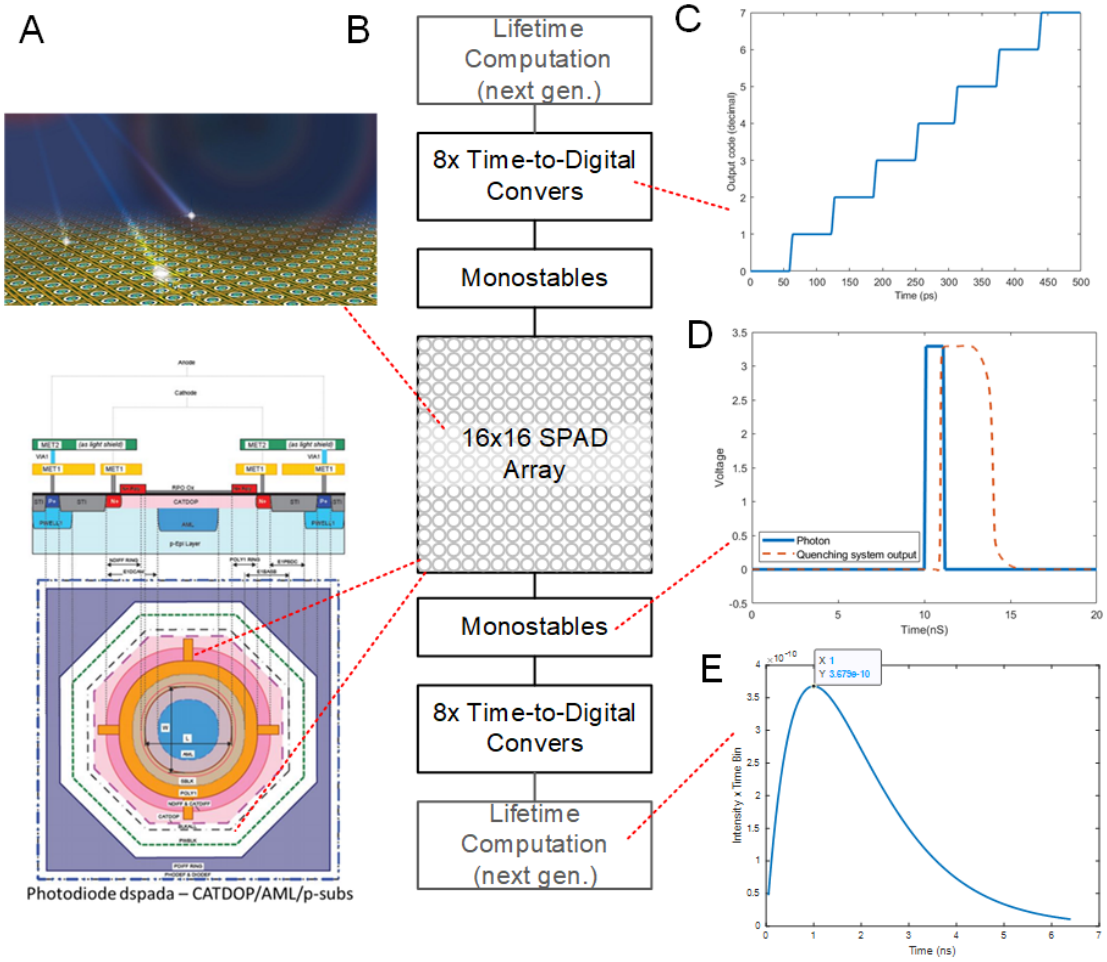


Figure 9*. Time correlated imager (TCI) overview.* (**A**) Cross section and layout view of the implemented SPAD. (**B**) System architecture of the TCI consisting of a 16x16 SPAD array, column-level monostable circuits, time-to-digital converters (TDCs), and on-chip fluorescent lifetime computation. (**C**) Plot of the simulated transfer function (photon arrival time vs. digital output) for the TDCs. (**D**) Plot of simulated photon arrival and corresponding monostable circuit output. (**E**) Time histogram of simulated fluorescent events with a fluorophore lifetime of 1ns. The center-of-mass closely matches the fluorescence lifetime.

VI. Demonstration of Economic Development and Impact

External Networks. As outlined in the Executive summary, the Nucleic Acid Memory (NAM) Institute at Boise State was invited to join the DNA Data Storage Alliance. The alliance is the first and most extensive bridge between industry and academic organizations that are pioneering DNA data storage. Its mission is to "*create and promote an interoperable storage ecosystem based on DNA as a data storage medium*". The alliance will recommend the creation of specifications and standards (e.g., encoding, reliability, retention, file systems) which enable end-users to add interoperable DNA-based storage solutions to their existing storage hierarchies. The founders include Illumina, Twist Biosciences, Western Digital, and Microsoft. Member organizations include but are not limited to: Ansa Biotechnologies, Battelle, Catalog, The Cloude Nobs Foundation, DNA Script, EPFL, ETH Zurich, Imagene, IMEC, Iridia, Kioxia, Molecular Assemblies, PFU, Quantitative Scientific Solutions, Quantum, Seagate, Semicondcutor Research Corporation, Spectra Logic, University of Arizona, University of Washington, Digital Preservation, Oligo Archive, Lost Alamos National Laboratory, Cinémathèque Suisse, 21e8, DNAli, and University of Marburg. This network is critical as Boise State attempts to license the NAM intellectual property and/or the research team spins-off companies in the memory/biotechnology arena.

External Impacts. As outlined in the Executive Summary, Steven Burden (founder/CEO) and Clementine Gibard Bohachek (co-founder/CSO) spun-off Facible, a biodiognostics company in Boise that has 25 employees and is seeking FDA approval for a novel COVID-19 screening technology. In service to future generations of biotechnology start-ups in Idaho, Facible and the Nucleic Acid Memory Institute are actively exploring the creation of a biotech incubator in Boise.

External Funding. Beyond the IGEM/HERC investment, the NAM Institute has secured $1,549,995 in grants from the National Science Foundation and the Semiconductor Research Corporation. According *The Implementation Group* (TIG) -- which is a research development firm specializing in strategic positioning, proposal development, and team science to increase Boise State's competitiveness for external funding – the awarded grants are among the most competitive and prestigious within the Boise State portfolio because we outcompeting MIT, Stanford and many other premier institutions that were positing for SemiSynBio funding. Building on our initial success, the NAM Institute is preparing to submit proposals to the NSF SemiSynBio III proposal opportunity this academic year, as well as the NSF Partnership for Innovation pathway to help the team evaluate if and how it should spin-off a company. To strengthen our proposal, we have designed, built, and started to test a custom SRM with 4.5 nm resolution. We have also designed and fabricated a custom 16×16 time-correlated imaging (TCI) array for super resolution and fluorescence

lifetime imaging microscopy (FLIM). Both the microscope and the TCI array are viable scientific instrumentation products that could be licensed or sold.

Future Funding. In support of future funding, the research team has provided thought partnership to IARPA on its proposed Biologically Templated Nanofabrication (IGATA) initiative; including but not limited to sharing technical ideas, suggesting performance metrics for the community to consider, introducing IARPA to leaders in the DNA nanotechnology community, reviewing drafts of their whitepaper (which will translate into an RFP), and offering to support their workshop once the RFP is approved.

In addition, the NSF Germination program aims to foster the development of frameworks, platforms, or environments to enable faculty to form research questions and ideas with potentially transformative outcomes. Based on the success of 2 NSF Germination Awards (# 1745944, 1629659), the PI has been invited to design, test, evaluate, and implement frameworks, platforms and/or environments that enable academics to formulate research questions and ideas that have the potential to address critical societal challenges.

VII. Demonstration of Economic Development and Impact

Demonstration (07/01/2018–07/27/2021)	**Amount**
External Funding	$ 1,549,995
Number of External Grants	3
Private Sector Engagement	~ 20 companies
University Engagement	~ 20 universities
Industrial Alliances Joined	1 (*DNA Data Storage Alliance*)
Federal Agency Engagement	5 agencies (*NSF, SRC, IARPA, DARPA, NIH*)
Industry Involvement	2 companies (*Micron, SRC*)
U.S. Patents Submitted	1
Publicly Available Software Packages	3
Plant Variety Protection Certificates	0
Technology Licenses Signed	0
News Releases	3 articles
Start-up Businesses Started	1 (Facible with 25 employees)
Jobs Created outside of BSU	~ 10

VIII. Numbers of Student, Staff, and Faculty participation

We were fortunate to have a diverse team of perspectives, experiences, and expertise. From the initial ideation phase leading to our proposal, to the research that led to our outcomes, we have embraced team science in addressing the future information storage needs outlined in the SemiSynBio Roadmap. In the following table, we recognize the people that have enabled the research during the project, from our students, staff, and principal investigators.

Contributor	Objective/Support	Experience	Professional Outcome
Steven Burden, PhD student	Objective 3	Biology	Earned his PhD, Co-founder and CEO of Facible
Chris Green, PhD student	Objective 4 and 5	Materials Science	Earned his PhD; NRC postdoctoral fellow
Mike Tobiason, PhD student, postdoctoral researcher	Objective 2 (PhD student), Objective 1, 2, and 5 as postdoctoral researcher)	Materials Science	Postdoctoral researcher for the NAM Institute
Golam Md Mortuza, PhD student	Objective 1	Computer Science	Passed his PhD proposal; intern at Facebook during summer 2021
Reza Zadegan, postdoc	Objective 1 and 4	Materials Science	Tenure-track faculty at NCA&T
Chad Watson, project manager	N/A	Project Management, Research Development	Boise State's Division of Research and Economic Development and the Center for Advanced Energy Studies
Kelsey Suyehira, MS student	Objective 1	Computer Science	Earned MS in Computer Science; Software Development Engineer at Cradlepoint
Elton Graugnard, co-PI	Objective 4	Materials Science, Physics	Transitioned off project to focus on developing atomically-thin semiconducting materials; awarded $126k by the Micron Foundation
Will Hughes, PI	Objectives 1-5	Materials Science	N/A
Wan Kuang, co-PI	Objective 5	Electrical Engineering	N/A
Tim Andersen, co-PI	Objective 1	Computer Science	N/A
Eric Hayden, co-PI	Objective 3, VIP	Biology	N/A
Shoshi Llewellyn, MS student	Objective 1	Computer Science	N/A
Will Clay, postdoctoral researcher	Objective 5	Optical Physics	N/A
Luca Piantinada, postdoctoral researcher	Objective 4, VIP	Bionanotechnology	N/A
George Dickinson, postdoctoral researcher	Objective 1, 4, 5	Biology, Computer Programing, Optical Physics	N/A
Clementine Gibard Bohachek, postdoc	Objective 3, VIP	Biology	Cofounder and CSO of Facible
Ben Johnson, collaborator	Objective 5	Electrical Engineering	N/A
Mehdi Bandali. PhD Student	Objective 5	Electrical Engineering	N/A
Natalya Hallstrom. Lab Man.	Objective 2–4, VIP	Biology	N/A
Sarah Kobernat, PhD student	Objective 3, VIP	Biology	N/A
Jacob Elmore, Julie Ramirez, Levi Orr, Amanda Wolf, Kaelee Ryner, Ben Balzer, Madia Bazso, Baylee Zanone, Ashlyn Trapp, Tia Senger, Hailey Jorgensen, Isaiah Keylor Aidan Poe, Katie Mateo Kelly Mazur, Hannah Hernandez, Gabe Frandsen, Lauren Grillo, Kayla Jonas, Olivia Paulsen, Brendan Yoshino, Hagen Shults, Madison Edwards, Tanner Pollock	A total of 24 students participated in our NAM VIP course over the last three years. These students range from freshman to seniors and span multiple majors: biology, pre-med, health sciences, chemistry, and psychology. The VIP students supported Objective 3.		Ben Balzer and Amanda Wolf selected to be summer undergraduate researchers for the NAM Institute; Ashlyn Trapp selected for an NSF REU on Data-Driven Security.

IX. Dissemination

See the summarized table of disseminated products on page 3 of this report.

X. Summary of Budget Expenditures

The investments from IGEM/HERC were largely infrastructure-centric; bringing the biological, computer, and materials sciences closer together by moving the NAM Institute into the Micron Center for Materials Research. Equipment that supported this project and its team integration included a:

- Custom super-resolution microscope – for pushing the physical limits of SRM
- Commercial super-resolution microscope – for routine NAM characterization
- SRM Environment Chamber – to minimize humidity and temperature effects
- SRM Vibrational Table – to minimize noise from vibrations during SRM
- Autoclave – for sterilizing solutions and equipment
- Shaker incubator – for microbial growth
- Sterile incubator – for growth and maintenance of E. coli strains.
- Gel imager – for validation, quantification and documentation of all Nucleic activities d materials including oligos, plasmids, scaffolds, and origami structures.
- Refrigerator – for storage of temperature sensitive biological material

With a desire to grow our computational capacity, and in anticipation of the long-term effects of the pandemic on our experimental research, our team also invested into a:

- Dell PowerEdge DSS 8440 RTX 8000 GPU node with 8 NVIDIA Quadro RTX 8000 48 GB GPU cards and 384 GB of system memory.

The above node continues to support our research via deep neural models to improve and speed up performance on such tasks as localization of fluorescent markers in SRM imagery, translation of SRM imagery to binary strings, and secondary structure prediction for our NAM encoding algorithms.

For this and more, the faculty, staff, and students would like to extend their greatest appreciation to IGEM/HERC. This project, and its resulting outcomes would not have been possible in the absence of your investment.