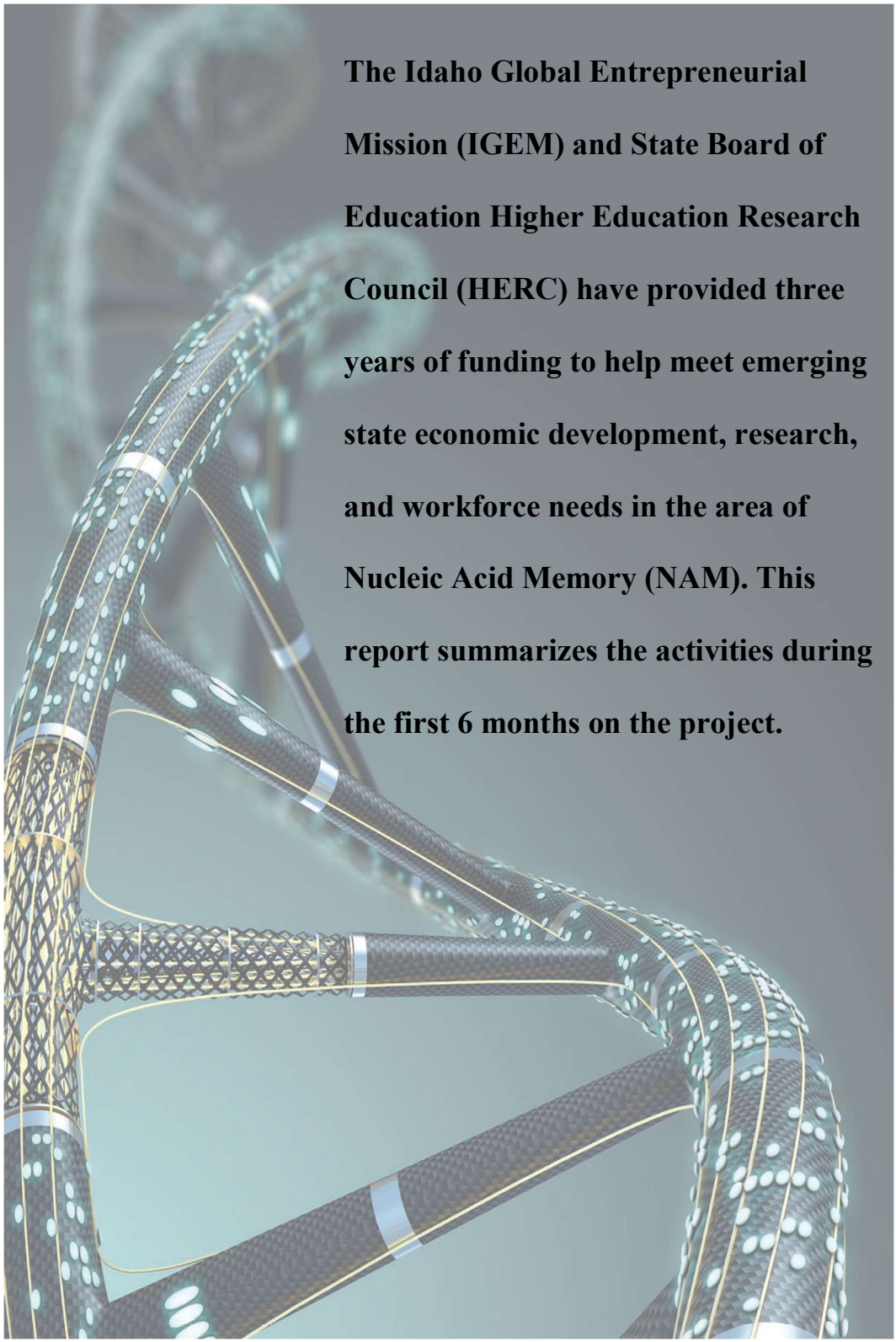
BOISE STATE UNIVERSITY

IGEM # 19-002: Nucleic Acid Memory

July 1, 2018 – January 1, 2019 Progress Report

Will Hughes
Chad Watson
Elton Graugnard
Tim Andersen
Reza Zadegan
Eric Hayden
Wan Kuang

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 activities during the first 6 months on the project.

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 Institute 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 will be met over the life of the IGEM-HERC: Task 1 – Create improved algorithms for coding information into data strands. Task 2 – Create a high-throughput, integrated analytical engine to design and select data strands using quantitative metrics based on an in-house, evolutionary algorithm. Task 3 – Create a cellular factory for manufacturing DNA scaffolds using a rapid design, build, and test cycle of genomes. Task 4 – Design and fabricate NAM storage nodes using the DNA scaffolds. Task 5 – Read and write arbitrary files into NAM storage nodes using super-resolution microscopy.

III. Summary of project accomplishments

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

- Kelsey Suyehira successfully defended her Master of Science in Computer Science in September 2018. Her thesis topic was entitled, *Using DNA for Data Storage: Encoding and Decoding Algorithm Development*. Briefly described, when encoding binary data into sequences of DNA, algorithms should account for biological constraints representing the idiosyncrasies of working with nucleic acids. In response, Kelsey created the REDNAM software package (a.k.a. Robust Encoding and Decoding of Nucleic Acid Memory). REDNAM includes a novel-mapping scheme that converts digital information into codons while accounting for important constraints when working with DNA. For example, it removes biologically active codes — such as start codons and some known promoter regions — avoids multiple repeats of unique nucleotides, and excludes repeating sequence strings. In doing so, Kelsey developed a schema mimicking how information has evolved to be efficiently encoded into natural DNA while also accounting for the errors that often arise when working with synthetic DNA. She also integrated her mapping scheme into a fountain code in an implementation that balanced information density with error correction. The result is that REDNAM recovers 100% of its data in spite of introducing random errors into the DNA. It also achieved a speed up of 2x for encoding and 435x for decoding digital information when compared to state-of-the-art fountain codes found in the literature. As shown below, Kelsey's thesis resulted in one publication and two conference proceedings that established a foundation for this award, with two more publications that are in preparation.

    - K. Suyehira, S. Llewellyn, R.M. Zadegan, W.L. Hughes, T. Anderson, "A Coding Scheme for Nucleic Acid Memory (NAM)," IEEE Workshop on Microelectronic and Electron Devices, pp 1-3, 2017.
    - R.M. Zadegan, K. Suyehira, S. Llewellyn, T. Andersen, W.L. Hughes, A Coding Scheme for Digital Data Storage in Nucleic Acid Memory (NAM), DNA 23, (September 2017), Austin, TX, USA.
    - R.M. Zadegan, K. Suyehira, S. Llewellyn, T. Andersen, W.L. Hughes, A Biologically Inspired Coding Scheme for Nucleic Acid Memory, FNANO, (April 2017), Snowbird, UT, USA.

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

- Michael Tobiason is completing his PhD in the Micron School of Materials Science & Engineering. With an expected graduation date in 2019, his dissertation topic is entitled, *Engineering Kinetically Uniform DNA Devices*. Briefly described, the relationship between DNA sequence and the rate of DNA reactions is not well understood. In response, Mike has hypothesized that observed kinetic variations in the literature arise due to unintentional base pairing in DNA. He has found that ranking model DNA devices based first on the size (in base-pairs) of the largest unintentional structure and then the count (number of structures of this size) reliably identifies sequences with improved kinetic reproducibility. To engineer DNA devices based on this principle, Mike created an evolutionary algorithm and software package called Sequence-Evolver. By engineering DNA devices with favorable interference profiles using Sequence-Evolver, Mike experimentally demonstrated that DNA kinetics vary by a factor of two or less when his sequences satisfy four conditions: (1) no intramolecular interferences longer than 2 base-pairs, (2) no intermolecular interferences longer than 4 base-pairs, (3) no stretches of consecutive cytosines or guanines longer than 3 base-pairs, and (4) no stretches of consecutive adenines or thymines longer than 6 base-pairs. Taken together, his findings support the hypothesis that kinetic variation arise due to interfering events and that kinetic reproducibility is possible through sequence optimization.

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

- Steven Burden is completing his PhD in Biomolecular Sciences and is expected to graduate in 2020. His dissertation topic is the development of nucleic acid biosensors with allosteric fluorescence signals. Supporting the Vertically Integrated Project (see section VI), Steven is organizing the training of six undergraduate students to produce, purify, and ensure the quality control of single-stranded DNA scaffolds from a manufacturing perspective. During the Fall 2018 semester, the undergraduate students built basic synthetic biology research skills including DNA primer design and validation in polymerase chain reaction, digital design and sharing of

DNA sequences, bacterial transformation and cloning, gel electrophoresis, and DNA quantification and quality control using ultra-violet absorbance.

- A simple, cost-efficient, and time-saving method for the generation of modified and unmodified long linear ssDNA molecules up to 40 kilobases is under development using a method called asymmetric PCR (aPCR). This method enables direct synthesis of the single-stranded DNA from the template DNA and does not require purification steps. The single-stranded DNA fidelity has been verified by gel electrophoresis and by sequencing.

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

- Sadao Takabayashi is completing his PhD in the Micron School of Materials Science & Engineering while working full time at Micron. He is expected to graduate in 2019 and his dissertation topic is *Patterning and Fabricating with DNA*. Foundational to this IGEM-HERC award, Sadao demonstrated high density and selective adsorption of DNA origami onto boron implanted silicon substrates made by Micron, which resulted in the below listed publication. He has since observed that surface adsorption is inversely proportional to the pattern feature size, and the smaller the pattern, the more pronounced the effect.
    - S. Takabayashi, S. Kotani, J. Flores-Estrada, E. Spears, J.E. Padilla, L. Godwin, E. Graugnard, W. Kuang, S. Sills, W.L. Hughes, "Boron-Implanted Silicon Substrates for Physical Adsorption of DNA Origami," International Journal of Molecular Sciences, vol 19, issue 9, number 2513, pp 1-12, 2018.
- Dr. Reza Zadegan, Assistant Research Professor on the project, has designed, built, and tested preliminary digital NAM (dNAM) structures. He has tested three iterations of rectangular NAM structures that have the capacity to contain 16-256 bits of binary information. Initial screening is underway to evaluate the resolution during super resolution microscopy, rate of structural errors, and direction/orientation of NAM. When completed, these parameters will inform future NAM prototypes.

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

- In support of this task, Drs. Wan Kuang and Elton Graugnard are pursuing a new super-resolution microscope to push the ultimate optical resolution for NAM. As a starting point, Nikon demoed a microscope at Boise State which had an ultimate demonstrated resolution of 20 nm, which is lower

than the resolution of our in-house microscope of ~14 nm point-to-point. In response, Kuang and Graugnard visited Leica labs in the Bay Area and UC Davis to evaluate two Leica super resolution microscopy systems. The research team is now in conversation with MadCityLabs about real-time camera-based drift correction and the demo is pending. This is in alignment with them pursuing options for building an advanced super-resolution microscope at Boise State.

- Using their current SRM system, the team demonstrated staple strand yield improvement due to PAGE filtration and docking-sequence dependent PAGE filtration.

IV. Demonstration of economic development and impact

Demonstration of Economic Development and Impact	Number
External Funding	$ 1,500,000
News Releases	3 articles
Private Sector Engagement	14 companies
University Engagement	11 universities
Federal Agency Engagement	5 agencies
Industry Involvement	2 companies
Patents	0
Copyrights	0
Plant Variety Protection Certificates	0
Technology Licenses Signed	0
Start-up Businesses Started	0
Jobs Created outside of Boise State University	0

Shortly after the IGEM-HERC award, 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*. Boise State was one of the few universities in the country to receive the prestigious award in the first round of competition. Other awardees included: MIT, Stanford University, University of Washington, and UT Austin. Prior to the release of this award mechanism, Drs. Will Hughes and Reza Zadegan coauthored the Semiconductor Synthetic Biology Roadmap in collaboration with the SRC, which helped steer the federal investments.

Because of the below listed consortium, industry involvement on the research project includes Gurtej Sandhu (Micron Technology Vice President) and Victor

Zhirnov (SRC Chief Scientist) who jointly serve as the co-chairs of the NAM Institute at Boise State University. According to Gurtej Sandhu, "*the leadership and innovation of this research team has brought them to the threshold of becoming a world class player in the research, development and education of nucleic acid memory*."

Industry Partners (14)	University Partners (11)	Federal Partners (5)
Autodesk	Boise State	- Army Research Office (ARO)
GenoCAD	Boston University	- Department of Defense (DoD)
Gingko Bioworks	Brigham Young U.	- Office of Naval Research (ONR)
Globalfoundries	Columbia University	- National Institute of Standards &
IBM	Dartmouth	Technology (NIST)
Intel	Georgia Tech	- National Science Foundation (NSF)
International Data Corp	NC State University	- Intelligence Advanced Research
Mentor Graphics	UCLA	Projects Activity (IARPA)
Micron	UIUC	
Microsoft	UNC Greensboro	
Mubadala Technology	U. of Washington	
Raytheon		
SynBioBeta		
Twist Biosciences		

For additional information, below are three news releases related to our work.

- **New NSF awards support the creation of bio-based semiconductors**, Sarah Bates, National Science Foundation, July 16, 2018. www.nsf.gov/news/news_summ.jsp?cntn_id=295968&org=NSF
- **How Micron's business could change dramatically from this research at Boise State**, David Staats, Idaho Statesman, September 27, 2018. www.idahostatesman.com/news/business/article218442875.html
- **Boise State University awarded $3.5 Million to research storing data on DNA**, Sherry Squires, August 28, 2018. https://news.boisestate.edu/update/2018/08/28/boise-state-university-awarded-3-5-million-to-research-storing-data-on-dna

In response to this momentum, Will Hughes has been invited to give a keynote talk on the Nucleic Acid Memory Institute at Boise State to VentureCapital.org (VCO) Investor's Choice Conference on February 20, 2019. The stage will be shared with senior leadership at Micron Technology. As a non-profit organization, VCOs mission is to improve the human condition by helping technology-based entrepreneurs "get started, find money, and change the world". Supported by the Wayne Brown Institute in Salt Lake City, VCO pulls together a powerful network of venture professionals who are actively engaged in advancing the impact of entrepreneurs in the United States.

V. Numbers of student, staff, and faculty participation

Classification	Number
Tenured or Tenure Track Faculty	5
Research Faculty	1
Project Manager	1
Graduate Students	6
Undergraduate Students	6

Critical to the success of any research initiative are the people that make up the project team. As part of the IGEM-HERC, we have six faculty (Will Hughes, Tim Andersen, Wan Kuang, Elton Graugnard, Eric Haden, and Reza Zadegan) and a project manager (Chad Watson). We have also transitioned three graduate students (Mike Tobiason – Task 2; Steven Burden – Task 3; Chris Green – Task 5) to this project and have recruited and hired three additional graduate students (Shoshi Llewellyn and Golam Md Mortuza – Task 1; Elijah Spears – Task 4 and 5). In addition, we hired Kelsey Suyehira, who is a recent graduate student from Computer Science that completed her Master of Science on the project. Ms. Suyehira helped transition graduate students focused on Task 1, while also working toward two project-related publications in which she is the lead author. In support of increasing the research productivity, the faculty are aggressively recruiting two postdoctoral fellows in the areas of nanofabrication and super-resolution microscopy to respectively work with Graugnard and Kuang. After an international search consisting of two advertisements, 42 applicants have applied, 8 finalists have been down-selected and 6 have been interviewed.

In support of this multidisciplinary team, a Vertically Integrated Project (VIP) was also launched in Fall 2018. As active participants in and co-owners of the Vertically Integrated Project courses called VIP 200, 400, 500 Bio-Innovations, undergraduate and graduate students enroll into a multi-year and multi-disciplinary research team that provides ongoing course and teaching credit. As of Spring 2019, six students have enrolled and are engaging in research activities aimed toward the production, purification, and quality control of new single-stranded DNA origami scaffolds. These students are being mentored by two previous VIP students with over 2 years of experience in the Eric Hayden lab, a fourth year PhD student, as well as the faculty members involved in the project. The students include sophomore Biology Major Ben Balzer, Junior Biology Major Madison Edwards, Hailey Jorgenson a senior pre-med student with a visual art minor, a senior Biology major Isaiah Keylor, and a senior Health Studies major Tia Senger. We are excited by the representation of women in this group,

the diversity of the majors and minors, and different levels of progress towards degrees. All students reported positive experiences, and all have signed up for the VIP course this semester (Spring 2019).

VI. Description of future plans

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

- Experimentally validate REDNAM by encoding and decoding digital information using synthetic DNA. The information will be read using commercially available next-generation sequencing. Once validated, the synthetic DNA will be randomly degraded under various doses to test REDNAM's ability to tolerate defects. Based on the performance of the tool, we will improve its robustness by accounting for better error correction techniques for insertion/deletion errors.
- Building on the success of REDNAM, we will create an optimal encoding/decoding algorithm specific to the digital NAM (dNAM). Concurrently, we will develop and publish a website for DNA-based encoding/decoding of data.

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

- Once published, the Sequence-Evolver software package will become an open source collaborative initiative available in the GitHub repository (https://github.com). It will also be posted on the website outlined above in Task 1. Once published, we will optimize the Sequence-Evolver software package to generate sequences specific to NAM, including but not limited to the scaffold strands, staple strands, and the strands used during super-resolution microscopy. In anticipation of needing to use the software tool for more complex jobs, Sequence-Evolver will be expanded to work on a supercomputer platform for larger NAM applications.

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

- Building on the initial success of the VIP project, graduate and undergraduate students will continue to collaborate to design and build synthetic phagemids for single-stranded DNA synthesis, and to quantify single-stranded DNA harvested from bacterial culture.

- The design space for scaffolds is quite large, and students will be initially challenged to build larger scaffolds with high yield. Our initial goal will be to double the size of the scaffold with the same molar yield of single-stranded DNA as M13mp18, which is the gold standard. To maximize exploration, the students will be divided up into two teams, each organized by an experienced undergraduate researcher, and each with guidance and mentorship from PhD student Steven Burden and Dr. Reza Zadegan.

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

- DNA origami will be designed for dNAM using in-house scaffolds. In addition, they will be tested in-house using our super-resolution microscope. Synthesis yield and defect rates for the origami structures will be quantified using super-resolution microscopy. In addition, initial origami structures will be designed for sequence NAM (seqNAM).

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

- Incorporate upgrades to existing super-resolution microscope to both improve its resolution and extend its fluid handling.
- Toward higher resolution, the team will design and build a custom super resolution NAM reading platform capable of sub-10 nm resolution.

VII. Summary of Budget Expenditures

The below table summaries expenditures associated with the project. O&E has helped support the postdoctoral research scientist searches and the purchase of modified and unmodified DNA oligos. The oligos are used to assemble NAM blocks and to perform super-resolution microscopy studies. Funds were also allocated to purchase polymerase enzymes and primer oligos necessary for asymmetrical Polymerase Chain Reactions (PCR). The team is currently assessing super-resolution equipment to be purchased under capital.

Category	Expended
Faculty and Staff Salary	$32,344
Graduate Students Salary	$40,425
Fringe Benefits	$14,824
Graduate Student Tuition and Fees	$13,791
O&E	$17,839
Capital	0

VIII. Commercialization Revenue

Commercialization	Revenue
None.	$0

IX. Additional metrics established specific to individual project

Metrics	Number
External Funding	$1,500,000
Software Tools Created and Initially Validated	3
Master of Science Thesis Awarded	1
Peer-Reviewed Publications	1
Manuscripts in Preparation	4
VIP Program Enrollment (grad and undergrad)	7
National and International Postdoc Recruitment	42

Listed above are specific, objective, measurable, and realistic performance metrics to gauge project success and economic impact, many of which have been distributed throughout this report and are consolidated here for ease of review.