



Scientists Race Toward DNA-Based Data Storage

New technology is needed to cope with the information boom.

If the pursuit of DNA-based data storage is a race, it is probably more of a long, arduous, challenge-laden Tough Mudder than a quick, straightforward 50-yard dash. Or it may be a tortoise and hare situation with data growing at an extraordinary pace while science moves steadily along in hopes of gaining the lead.

BY GEORGE I.
SEFFERS

Soon, conventional silicon-based computing technology may no longer adequately store the world's data, which is growing at a mind-boggling pace, so researchers across multiple governments, industry and academia are racing to explore DNA as an alternative. The National Science Foundation, for example, recently kicked off eight projects designed to tackle

some of the toughest challenges associated with storing data in DNA.

Late last year, International Data Corporation, a market research firm, predicted the global datasphere will grow from 33 zettabytes in 2018 to 175 zettabytes in 2025. A zettabyte is a measure of storage capacity and is 10^{21} (1,000,000,000,000,000,000,000 bytes) or 1 sextillion bytes. Cisco attempts to put that into terms non-math wizards can understand by explaining that if each terrabyte in a zettabyte were a kilometer, it would equal 1,300 trips to the moon and back, and if each gigabyte in a zettabyte were a brick, it would be enough to build 258 Great Walls of China, which is made up of 3,873,000,000 bricks.

"Globally, there is a data explosion ... a massive volume of data being generated continuously," notes Usha

Varshney, program director for Electronics, Photonics and Magnetic Devices, at the National Science Foundation (NSF).

And that data affects virtually every aspect of society, including healthcare, manufacturing, the electric grid and transportation. Officials with the military and the intelligence agencies also report being unable to effectively access, analyze and use all of the data being generated.

“All of these are going to be heavily impacted by data, big data, data analytics, and so having new technologies that allow you to manipulate data in ways that had not been

is not in silicon,” says Mitra Basu, NSF program director in the Directorate for Computer and Information Science and Engineering.

Just last year, the NSF kicked off eight new projects under the Semiconductor Synthetic Biology (SemiSynBio) program. New information technologies based on biological principles could enable stored data to be retained for more than 100 years and storage capacity to be 1,000 times greater than current capabilities, according to NSF officials.

While most people may not realize it, DNA actually generates information as part of its natural, biological functioning. The average human writes 40 exabytes each day while consuming comparatively little energy. “What we see in biology is that every day we are writing new information in our DNA. That happens every day, a large number of bytes,” Basu points out.

DNA continually changes based on a number of factors, including environment, diet and lifestyle. “Your DNA keeps changing. It adds things to it, and sometimes we have disease for that reason,” Basu explains.

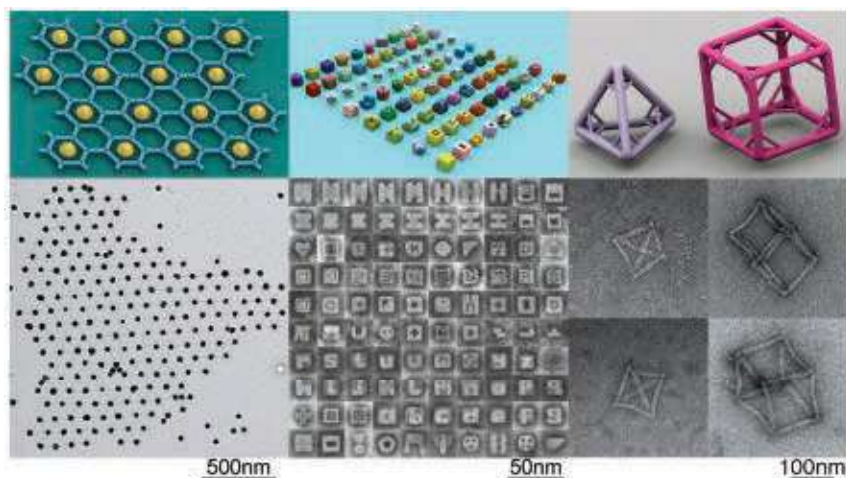
The new projects under the SemiSynBio program address a range of potential applications, such as automating the design of genetic circuits, creating bioelectronics, and exploring methods for molecular communication. Essentially, the idea is to integrate DNA and silicon, taking advantage of the best qualities of both.

“The semiconductor technology is approaching its limits of scaling. Currently, there is no technology that can replace it. The goal here is to seek synergy between the capabilities of the semiconductor technology with the properties of a biomolecule and create a DNA storage system,” Varshney offers. “Recently, there has been a real breakthrough in synthetic biology, and these bio molecules have the capability of carrying stored digital data for memory architectures. On the other hand, the semiconductor industry has developed tools to integrate biomolecules and other complex molecules at the system level.”

The NSF’s basic research efforts aim to resolve issues required to effectively unite DNA and silicon. Essential questions include how to input data into DNA and then extract it when needed, what kind of interface is required, and how to design and sequence DNA more rapidly.

Inputting data is known as “writing” and extracting it is referred to as “reading.” “The question is how we will be able to write and read in these DNA molecules. For that, the semiconductor comes in because we need an electronic interface to input data and also output data,” Basu explains. “We cannot do this at the moment because the cost of reading and writing is very expensive.”

She notes that Microsoft Corporation has invested heavily in synthetic DNA strands as part of its own research into DNA data storage solutions. Unfortunately, synthetic DNA remains too expensive for practical product development.



Researchers use DNA, the essential molecule for coding genetic information in biology, as a programmable building block to create sophisticated materials with custom-designed functions. New research seeks to integrate DNA and silicon-based computing.

possible before could have a lot of impact on an immensely growing capability in society,” says Filbert (Fil) Bartoli, who directs NSF’s Division of Electrical, Communications and Cyber Systems.

Bartoli adds that the quest for DNA-based data storage is really part of a much broader research challenge: the interface between biological systems and semiconductor microelectronics. “It could be part of prosthetics. It could be part of any number of systems that involve health. The way that the living system interfaces with electronics is a critical problem,” he declares.

Sheer longevity is one reason DNA seems a plausible alternative for data storage. DNA has been extracted from the bones of a horse that died approximately 700,000 years ago. “Data in DNA in the right conditions could keep for thousands of years. This is the basis for a real race toward DNA storage,” Varshney says.

In addition, DNA can hold massive amounts of information. Scientists estimate that 1 gram of DNA can hold up to 455 exabytes—1 quintillion bytes—of data and that all the data that existed in the world in 2015 could fit on a DNA hard drive the size of a teaspoon. Furthermore, DNA may enable virtually unbreakable encryption to keep data secure.

“We are looking at how to address this data explosion that is happening and how we can store the data that we have in a more efficient way. ... The future of data storage

“If you want to go and buy DNA to store 200 megabytes of something, it will be \$800,000, so this is not feasible. This is not ready for the market.”

One of the new projects, which is titled An On-Chip Nanoscale Storage System Using Chimeric DNA, aims to “reduce the cost-integration barrier between classical recorders and DNA-based data storage devices,” according to an NSF website. The goal is to develop a new system centered around “chimeric DNA,” using chemically modified nucleotides, the basic building blocks for DNA, to extend the coding alphabet from four basic symbols found in DNA to more than 20.

DNA is primarily made up of adenine, thymine, guanine and cytosine, which are abbreviated A, T, G and C. Adding another 16 nucleotides would dramatically increase storage density. “That seems to be a pretty innovative idea that you can change the nucleotides and control them in order to use them as memory,” Varshney says.

The reading process also is currently too slow. “We can read from DNA 400 bytes per second, as opposed to your computer, which can read 100 megabytes per second. If you are watching a movie, you cannot do it with 400 bytes per second,” Basu offers.

The interface between DNA and silicon technology also is a challenge. “You cannot just take a DNA molecule and put it on a chip. It has to be properly integrated together so that the DNA molecule does not die, and so that you can still write data into it or extract data from it,” Basu adds.

The NSF is not the only organization working on DNA-based data storage. Scientists at Sandia National Laboratories developed a novel method for encrypting text within synthetic DNA. The encryption is stronger than conventional technology and practically impossible to break, researchers say. (SIGNAL Magazine, February 2017, “The Infinite Promise of DNA-Based Data Encryption”)

In addition, the University of Washington and Microsoft recently announced that they have demonstrated the first fully automated system to store and retrieve data in manufactured DNA, which they describe as “a key step in moving the technology out of the research lab and into commercial data centers.” Also, a startup known as Catalog revealed in June that it has packed the entire contents of Wikipedia’s English-language version—about 16 gigabytes—into DNA molecules. And scientists from Brown University reported in June they have developed a method for storing data on metabolic molecules by manipulating those molecules to imitate the 1s and 0s used in binary code.

As research progresses, DNA data storage might first be used for archiving information or storing medical records, cases in which the information has to be stored for long periods but does not need to be retrieved very often. The technology also may find a role in medicine, such as the fight against cancer. Basu says it may be possible, relatively soon, to input data into DNA that would kill cancer cells.

“Part of the challenge going forward is to identify those applications that can be successful within the realm of performance that has been demonstrated or will be demonstrated in the near future,” Bartoli explains.

The NSF’s eight new projects will be funded for three years, but because it is basic, fundamental research that could take years to come to fruition, they may require additional funding. “Maybe after 25 years we will have a bio-flash drive so you and I can use it,” Varshney quips.

• • • — • •

To share or comment on this article
go to <http://url.afcea.org/September19>



contact: George I. Seffers, gseffers@afcea.org

In cooperation with


FUTURE FORCES FORUM International Platform for Trends & Technologies in Defence & Security
www.future-forces-forum.org

International Security Conference

SCADA SECURITY

Organized as a part of FUTURE FORCES FORUM

4 – 5 November 2019

Hotel DAP, Prague, Czech Republic

Patronage and professional sponsorship










Topics

- Current and future cyber threats and their solution
- New trends in technology in the ICS security
- Connected world, Mobility and the Internet of things
- Human factor and related topics
- Public safety in smart cities
- GDPR in environment of the smart cars

www.future-forces-forum.org