

April 7, 2016

UW team stores digital images in DNA – and retrieves them perfectly

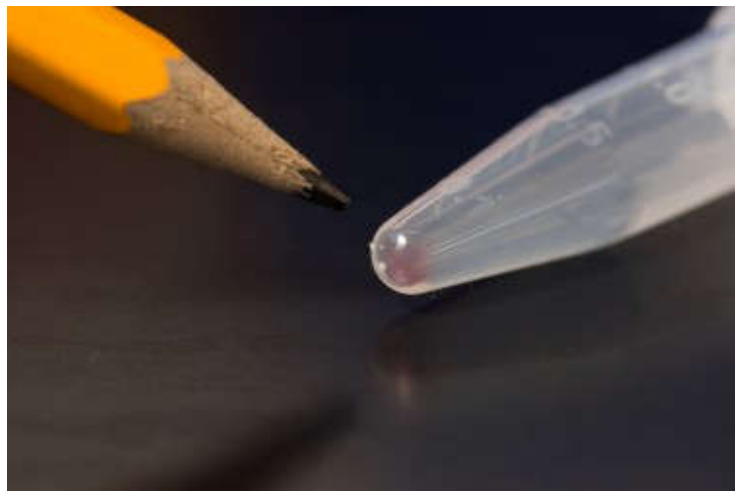
[Jennifer Langston](#)

UW News

Technology companies routinely build [sprawling data centers](#) to store all the baby pictures, financial transactions, funny cat videos and email messages its users hoard.

But a new technique developed by University of Washington and Microsoft researchers could shrink the space needed to store digital data that today would fill a Walmart supercenter down to the size of a sugar cube.

In a [paper](#) presented in April at the [ACM International Conference on Architectural Support for Programming](#)



All the movies, images, emails and other digital data from more than 600 basic smartphones (10,000 gigabytes) can be stored in the faint pink smear of DNA at the end of this test tube. *Tara Brown Photography/University of Washington*

For more information, visit the [Molecular Information Systems Lab](#).

[Languages and Operating Systems](#), the team of computer scientists and electrical engineers has detailed one of the first complete systems to encode, store and retrieve digital data using DNA molecules, which can store information millions of times more compactly than current archival technologies.

Authors of the paper are UW computer science and engineering doctoral student [James Bornholt](#), UW bioengineering doctoral student [Randolph Lopez](#), UW associate professor of computer science and

engineering [Luis Ceze](#), UW associate professor of electrical engineering and of computer science and engineering [Georg Seelig](#), and Microsoft researchers and UW CSE affiliate faculty Doug Carmean and Karin Strauss.

In one experiment outlined in the paper, the team successfully encoded digital data from four image files into the nucleotide sequences of synthetic DNA snippets. More significantly, they were also able to reverse that process — retrieving the correct sequences from a larger pool of DNA and reconstructing the images without losing a single byte of information.

The team has also encoded and retrieved data that authenticates archival video files from the UW's [Voices from the Rwanda Tribunal](#) project that contain interviews with judges, lawyers and other personnel from the Rwandan war crime tribunal.

“Life has produced this fantastic molecule called DNA that efficiently stores all kinds of information about your genes and how a living system works — it’s very, very compact and very durable,” said co-author [Luis Ceze](#), UW associate professor of computer science and engineering.

“We’re essentially repurposing it to store digital data — pictures, videos, documents — in a manageable way for hundreds or thousands of years.”



The digital universe — all the data contained in our computer files, historic archives, movies, photo collections and the exploding volume of digital information collected by businesses and devices worldwide — is expected to hit 44 trillion gigabytes by 2020.

That’s a tenfold increase compared to 2013, and will represent enough data to fill more than six stacks of computer tablets stretching to the moon. While not all of that information needs to be saved, the world is producing data faster than the capacity to store it.

DNA molecules can store information many millions of times more densely than existing technologies for digital storage — flash drives, hard drives, magnetic and optical media. Those systems also degrade after a few years or decades, while DNA can reliably preserve information for centuries. DNA is

Lee Organick, a UW computer science and engineering research scientist, mixes DNA samples for storage. Each tube contains a digital file, which might be a picture of a cat or a Tchaikovsky symphony. *Tara Brown Photography/ University of Washington*

best suited for archival applications, rather than instances where files need to be accessed immediately.

The team from the [Molecular Information Systems Lab](#) housed in the UW Electrical Engineering Building, in close collaboration with [Microsoft Research](#), is developing a DNA-based storage system that it expects could address the world's needs for archival storage.

The team from the [Molecular Information Systems Lab](#) housed in the UW Electrical Engineering Building, in close collaboration

with [Microsoft Research](#), is developing a DNA-based storage system that it expects could address the world's needs for archival storage.

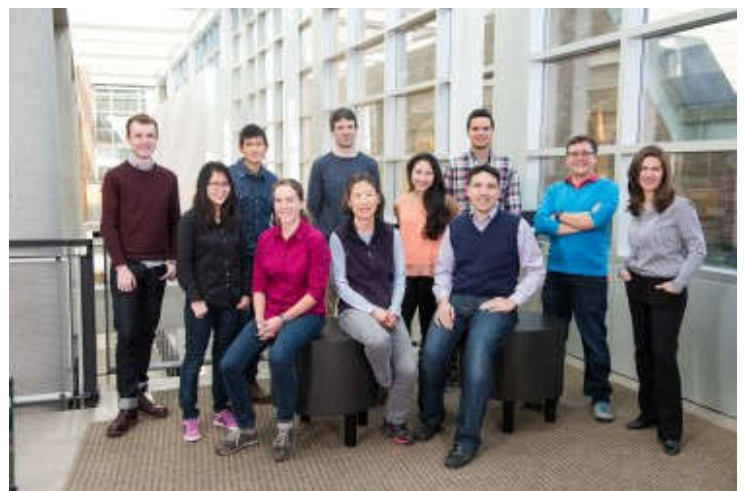
First, the researchers developed a novel approach to convert the long strings of ones and zeroes in digital data into the four basic building blocks of DNA sequences — adenine, guanine, cytosine and thymine.

“How you go from ones and zeroes to As, Gs, Cs and Ts really matters because if you use a smart approach, you can make it very dense and you don't get a lot of errors,” said co-author [Georg Seelig](#), a UW associate professor of electrical engineering and of computer science and engineering. “If you do it wrong, you get a lot of mistakes.”

The digital data is chopped into pieces and stored by synthesizing a massive number of tiny DNA molecules, which can be dehydrated or otherwise preserved for long-term storage.

The UW and Microsoft researchers are one of two teams nationwide that have also demonstrated the ability to perform “random access” — to identify and retrieve the correct sequences from this large pool of random DNA molecules, which is a task similar to reassembling one chapter of a story from a library of torn books.

To access the stored data later, the researchers also encode the equivalent of zip codes and street addresses into the DNA sequences. Using Polymerase Chain Reaction (PCR) techniques — commonly used in molecular biology — helps them more easily identify the zip codes they are looking for. Using DNA sequencing techniques, the researchers can then “read” the data and convert them back to a video, image or document file by using the street addresses to reorder the data.



The Molecular Information Systems Lab research team: Front (left to right): Bichlien Nguyen, Lee Organick, Hsing-Yeh Parker, Siena Dumas Ang, Chris Takahashi. Back (left to right): James Bornholt, Yuan-Jyue Chen, Georg Seelig, Randolph Lopez, Luis Ceze, Karin Strauss. Not pictured: Doug Carmean, Rob Carlson, Krittika

Currently, the largest barrier to viable DNA storage is the cost and efficiency with which DNA can be synthesized (or manufactured) and sequenced (or read) on a large scale. But researchers say there's no technical barrier to achieving those gains if the right incentives are in place.

d'Silva. Credit: Tara Brown Photography/University of Washington *Tara Brown Photography/University of Washington*

Advances in DNA storage rely on techniques pioneered by the biotechnology industry, but also incorporate new expertise. The team's encoding approach, for instance, borrows from error correction schemes commonly used in computer memory — which hadn't been applied to DNA.

"This is an example where we're borrowing something from nature — DNA — to store information. But we're using something we know from computers — how to correct memory errors — and applying that back to nature," said Ceze.

"This multidisciplinary approach is what makes this project exciting. We are drawing from a diverse set of disciplines to push the boundaries of what can be done with DNA. And, as a result, creating a storage system with unprecedented density and durability," said [Karin Strauss](#), a researcher at Microsoft and UW affiliate associate professor of computer science and engineering.

The research was funded by Microsoft Research, the National Science Foundation, and the David Notkin Endowed Graduate Fellowship.

For more information, contact Ceze at luisceze@cs.washington.edu or Seelig at gseelig@u.washington.edu. To reach Strauss, please contact TNRPR@we-worldwide.com.

[Facebook](#)

[Twitter](#)

[Email](#)

Tag(s): [College of Engineering](#) • [Georg Seelig](#) • [Luis Ceze](#) • [Paul G. Allen School of Computer Science & Engineering](#)

SEARCH UW NEWS

Search for:

CATEGORIES

[Browse](#)

UW EXPERTS

- [Hurricanes and weather](#)
- [Flooding and landslides](#)
- [Wildfires](#)
- [Public policy and politics](#)

[More topics](#)

Latest news



[**Common allergen, ragweed, will shift northward under climate change**](#)

[2 hours ago](#)



[**UW recognized for 27-year partnership with Northwest Mountain Minority Supplier Development Council**](#)

[22 hours ago](#)



[**University of Washington's Fall 2018 entering class is largest ever**](#)

[24 hours ago](#)

[More](#)

NEWSLETTER SIGNUP

UW Today Daily

UW Today Week in Review

