NEWS | SEMICONDUCTORS

How Scientists Encoded "The Wizard of Oz" Into DNA › University of Texas researchers unveil protocol to shuffle large data stores into strands of genetic material

BY MARGO ANDERSON

27 JUL 2020 | 4 MIN READ

Margo Anderson is senior associate editor and telecommunications editor at IEEE Spectrum.

Synthetic DNA as a high-density data storage medium has fascinated digital futurists for years. The entire internet could be coded into DNA strands that fit inside a shoebox, while the DNA molecule is so stable it can last tens of thousands or even hundreds of thousands of years. In 2013, for instance, scientists sequenced the entire genome of a 700,000 year-old horse fossil.

The trick to date has involved shoehorning vast sums of bytes—a data standard tailor-made for linear and sequential stores like RAM and hard drives—into wet, squiggly forests of nano-sized deoxyribonucleic spaghetti noodles. Translating one data format to the other has been anything but straightforward.

Enter William Press's team at the University of Texas at Austin. They've pioneered a set of DNA data encoding and decoding algorithms that could jumpstart a new field of high-density, long-term data storage. Their work, reminiscent in its generative ambition of the landmark BB84 protocol that launched the field of quantum cryptography, could one day form the basis for a world of genomic data storage applications that come from reimagining information in terms of petabytes per gram.

Stephen Jones, a postdoc in Press's group and co-author of the *Proceedings of the National Academy of Sciences* paper that describes their research, says it's best to begin by understanding where data storage errors typically creep in. In traditional hard-drive and flash memory devices, bit-flips and erasures are the enemies of zeroes and ones.

"We have decades of beautiful work finding solutions to these two kinds of errors," Jones said. "But DNA is fundamentally different."

To make a workable DNA data storage standard, you need instead to worry about substitutions, insertions and deletions. The first is similar to a bit flip in which, say, an A nucleotide is substituted in the place where a T nucleotide used to be. (A, C, T and G and not 0 and 1 are the base language of DNA information.) The latter two classes of error represent cases, as the names suggest, where DNA base pairs are inserted or deleted from a strand.

Crucially, however, with DNA there is no reliable, inherent way of knowing that the strand you're reading off contains any substitution, insertion or deletion errors. There's no such thing as a countable and quantifiable DNA "memory register." Every base pair is just another nucleotide in a long sequence. And together they all form just another strand of DNA.

The relative nature of DNA data storage, in fact, is a key to Press, Jones and co-authors's HEDGES protocol (standing for Hash Encoded, Decoded by Greedy Exhaustive Search). No single isolated nucleotide in their protocol contains useable data. Rather, it's the accumulation of sequences of nucleotides that provides a robust storage system that they predict could achieve DNA's high-density potential while still enduring the eons.

The group used L. Frank Baum's *The Wizard of Oz,* translated into Esperanto no less, as their sample data set to be stored. Synthetic DNA these days, Jones said, typically “outer” coding layer and an “inner” coding layer. (Think of these steps as two separate algorithms in a complex cryptographic standard.)

The outer layer diagonalized the source data so that any given strand of DNA would contain shards of many portions of the message. The inner layer, HEDGES, then translates each bit into an A, C, T or G according to an algorithm that depends both on the zero or one value of that bit plus additional information about its place in the data stream as well as the data bits immediately preceding it.

Then, once *Oz* is translated into the language of nucleotides, it’s now ready to be written onto strands of synthetic DNA. Once encoded, the strands sat in storage where, Jones said, his job was to artificially age the genetic information—attempting to biochemically mutate the DNA strands and subjecting the sample to heat and cold damage.

---

out which bits are "address" bits, they can string the remaining information bits back together into a single, concatenated data file.

Fellow postdoc and fellow co-author John Hawkins said one of the most attractive features of their new protocol is how robust it is to technological and data format changes over the centuries to come.

"Reading DNA will never become obsolete," he said. "Data surviving into the future is only half the problem. You still need to be able to read it on the other end. [But] DNA is uniquely future-proof on this front because we are made of it. As long as humans are made of DNA, we will always want machines around that can read it."

MORE FROM SPECTRUM

L. Frank Baum's Cautionary Tale About the Gifts of Electricity

TAGS

MEMORY NANOTECHNOLOGY MATERIALS DNA-DATA-STORAGE

ABOUT THE AUTHOR

Margo Anderson is senior associate editor and telecommunications editor at *IEEE Spectrum*.
See full bio →

READ ALSO

NEWS | THE INSTITUTE

IEEE Honors Global Dream Team of Innovators

8 HOURS AGO | 3 MIN READ

NEWS | AEROSPACE

Why Artemis Faces Challenges in Lunar Ambitions

13 HOURS AGO | 6 MIN READ

NEWS | SEMICONDUCTORS

New Devices Might Scale the Memory Wall

14 HOURS AGO | 3 MIN READ

Related Stories

NEWS SEMICONDUCTORS

Hybrid Memory Designed to Cut AI Energy Use

FEATURE | AI

AI HUNTS FOR THE NEXT BIG THING IN PHYSICS

There's a crisis in particle physics. Researchers hope AI can help.

BY MATTHEW HUTSON

03 FEB 2026 | 18 MIN READ

In 1930, a young physicist named Carl D. Anderson was tasked by his mentor with measuring the energies of cosmic rays—particles arriving at high speed from outer space. Anderson built an improved version of a cloud chamber, a device that visually records the trajectories of particles. In 1932, he saw evidence that confusingly combined the properties of protons and electrons. “A situation began to develop that had its awkward aspects,” he wrote many years after winning a Nobel Prize at the age of 31. Anderson had accidentally discovered antimatter.

Four years after his first discovery, he codiscovered another elementary particle, the muon. This one prompted one physicist to ask, “Who ordered that?”

Keep Reading ↓

SPONSORED ARTICLE | ENERGY

Real-World Diagnostics and Prognostics for Grid-Connected Battery Energy Storage Systems

› The core challenge of clean energy

BY THE UNIVERSITY OF SHEFFIELD

12 DEC 2025 | 7 MIN READ

THE UNIVERSITY OF SHEFFIELD

electricity. Grid operators face increasingly steep ramp events, larger frequency excursions, faster transients, and prolonged periods where fossil generation is minimal or absent.

Keep Reading ↓

WHITEPAPER | TELECOMMUNICATIONS

Breaking Boundaries In Wireless Communication › Simulating Animated, On-Body RF Propagation

BY REMCOM

03 FEB 2026 | 1 MIN READ

This paper discusses how RF propagation simulations empower engineers to test numerous real-world use cases in far less time, and at lower costs, than in situ testing alone. Learn how simulations provide a powerful visual aid and offer valuable insights to improve the performance and design of body-worn wireless devices.

NEWS COMPUTING

New Tool Enhances 3D Modeling for Visually Impaired › A11yShape lets blind coders design and verify models on their own

BY SAMANTHA HURLEY

07 FEB 2026 | 3 MIN READ

for blind and low-vision users. As a result, a range of hardware design, robotics, coding, and engineering work is inaccessible to interested programmers. A visually-impaired programmer might write great code. But because of the lack of accessible modeling software, the coder can't model, design, and verify physical and virtual components of their system.

However, new 3D modeling tools are beginning to change this equation. A new prototype program called A11yShape aims to close the gap. There are already code-based tools that let users describe 3D models in text, such as the popular OpenSCAD software. Other recent large-language-model tools generate 3D code from natural-language prompts. But even with these, blind and low-vision programmers still depend on sighted feedback to bridge the gap between their code and its visual output.

Keep Reading ↓

TAGS

ASSISTIVE TECH | CODING | COGNITIVE LOAD | PROGRAMMING | SCREEN READERS

- Career Alert
- Robotics News
- The Future Lane
- University Spotlight
- Product Spotlight

Your email

I agree to the IEEE Privacy Policy

SUBSCRIBE

FEATURE | AEROSPACE

THE QUEST TO BUILD A RADIO TELESCOPE THAT CAN HEAR THE COSMIC DARK AGES

The catch: It will have to be on the moon

BY NED POTTER

20 JAN 2026 | 14 MIN READ

Isolation dictates where we go to see into the far reaches of the universe. The Atacama Desert of Chile, the summit of Mauna Kea in Hawaii, the vast expanse of the Australian Outback—these are where astronomers and engineers have built the great observatories and radio telescopes of modern times. The skies are usually clear, the air is arid, and the electronic din of civilization is far away.

It was to one of these places, in the high desert of New Mexico, that a young astronomer named Jack Burns went to study radio jets and quasars far beyond the Milky Way. It was 1979, he was just out of grad school, and the Very Large Array, a constellation of 28 giant dish antennas on an open plain, was a new mecca of radio astronomy.

Keep Reading ↓

SPONSORED ARTICLE | AI

From Bottleneck to Breakthrough: AI in Chip Verification › How AI is transforming chip design with smarter verification methods

BY PRIYANK JAIN

30 OCT 2025 | 8 MIN READ

Using advanced machine learning algorithms, Vision AI analyzes every error to find groups with common failure causes. This means designers can attack the root cause once, fixing problems for hundreds of checks at a time instead of faster, and packed with more features than ever before. But creating these chips is not just a question of sheer engineering talent or ambition. The design process itself has reached staggering levels of complexity, and with it, the challenge to keep productivity and quality moving forward.

Keep Reading ↓

WHITEPAPER | AI

How AI Accelerates PMUT Design For Biomedical Ultrasonic Applications › Neural networks trained on 10,000 FEM simulations deliver reliable performance forecasts.

BY QUANSCIENT

08 JAN 2026 | 1 MIN READ

This whitepaper provides MEMS engineers, biomedical device developers, and multiphysics simulation specialists with a practical AI-accelerated workflow for optimizing piezoelectric micromachined ultrasonic transducers (PMUTs), enabling you to explore complex design trade-offs between sensitivity and bandwidth while achieving validated performance improvements in minutes instead of days using standard cloud infrastructure.

What you will learn about:

Keep Reading ↓

NEWS | THE INSTITUTE

IEEE Mini-MBA Aims to Fill Leadership Skills Gaps in AI › Curriculum includes strategic decision-making and risk management

BY ANGELIQUE PARASHIS

06 FEB 2026 | 4 MIN READ

ISTOCK

redefinition of the skills required for business leadership.

Keep Reading ↓

TAGS

AI CAREERS EDUCATION IEEE PRODUCTS AND SERVICES RUTGERS