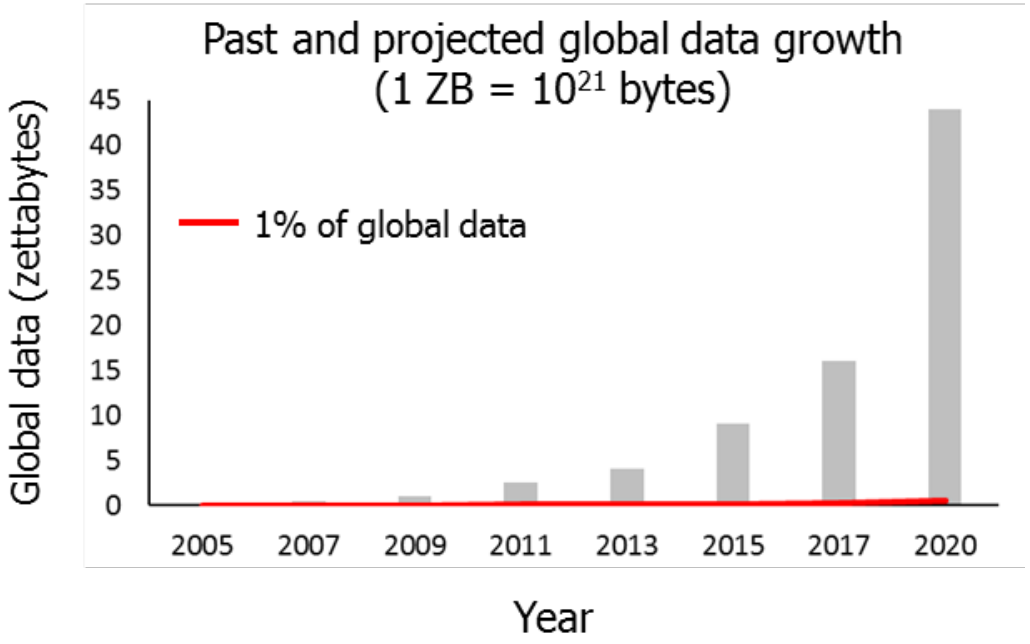
DARPA Molecular Informatics Program: Molecular Data Storage and Computing

Anne Fischer, Ph.D.
Program Manager
Defense Sciences Office

Library of Congress Designing Storage Architectures

September 10, 2019

Motivation: We're creating vast amounts of data, much of which is not stored or analyzed



Projected storage densities of 10^{18} bytes/mm^3

Projected processing speeds of ~10^{20} reactions/s

Ezziane et al. *Nanotech.*, 2006, *17*, R27-R39

For DNA: All data produced through 2015 could be stored in an SUV

Rapid progress in DNA-based data storage, now content addressable

2013: First practical demonstration – 739 kB

European Bioinformatics Institute

Text, Image and Audio

Goldman et al. *Nature*, 2013, *494*, 77-80

Write → 0101110011...

...0101110011...
...ACTGCGATC...

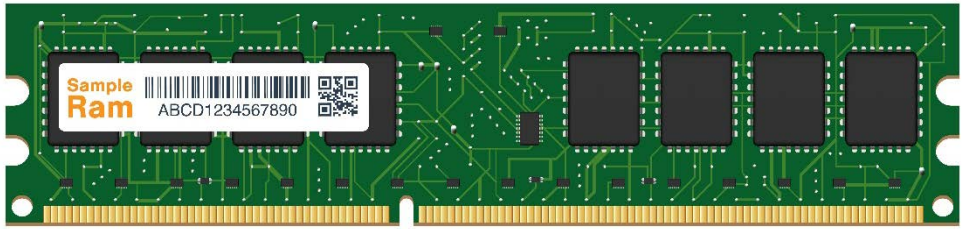
Read → 0101110011...

Dr. Martin Luther King's "I Have a Dream" speech

2016: 200 MB of data stored at ~10^{17} bytes/mm^3

2016: First random access demonstration from a set of images

University of Washington/Microsoft



Ceze et al. *Proc. of ACM Int. Conf. on Architectural Support for Programming Languages and Operating Systems*, 2016

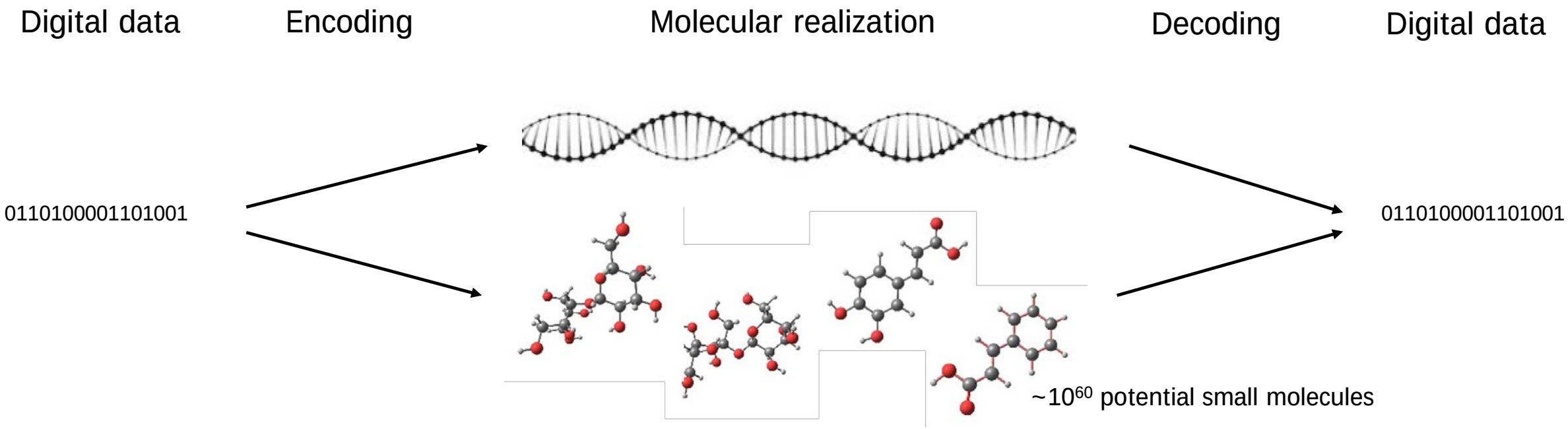
2019: Encoded features of 1.6M images*

DARPA Molecular Informatics
University of Washington/Microsoft

in-silico

DNA

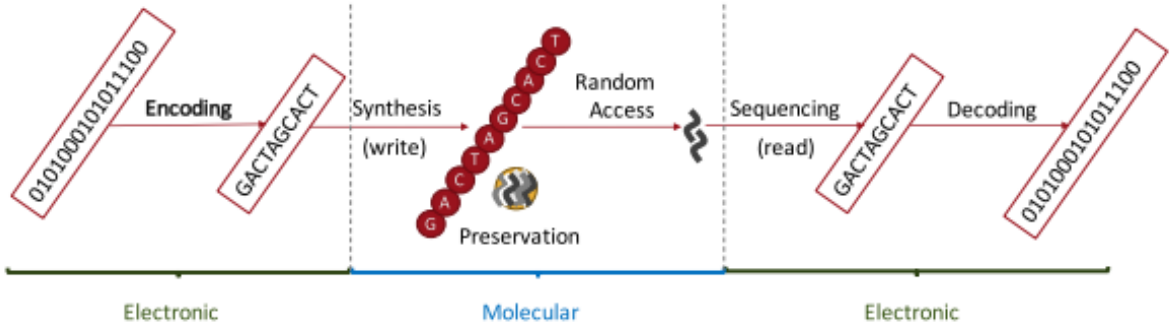
Molecular data storage is not all about DNA



Selection of molecular substrate depends on factors such as density, stability, scalability, accessibility, security and exploitability -- application dependent

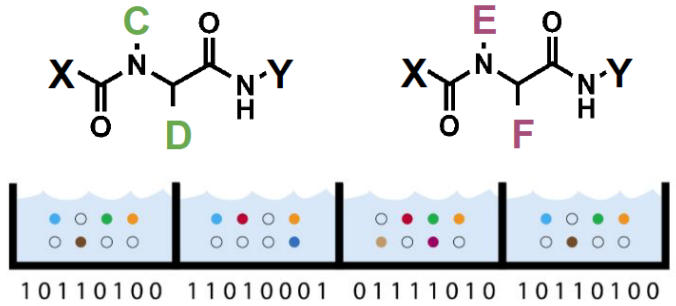
Diverse approaches to storage of digital data in molecules

DNA

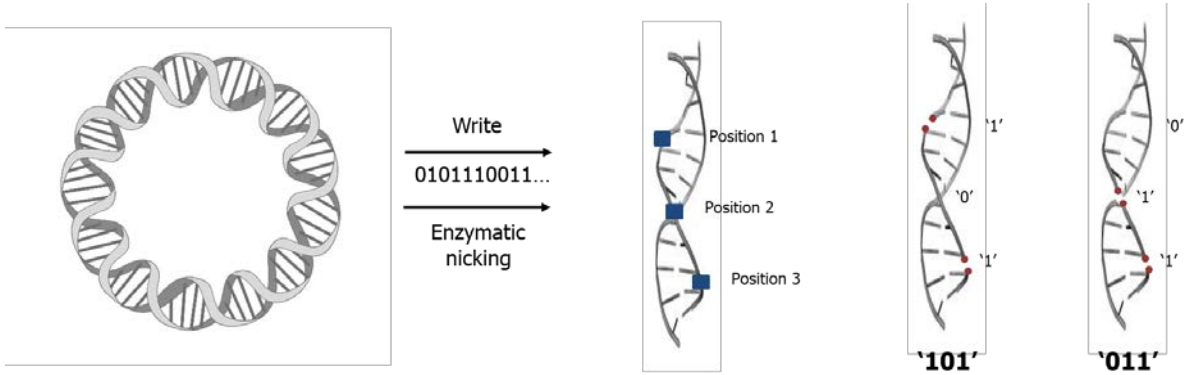


Washington: Synthetic DNA sequences

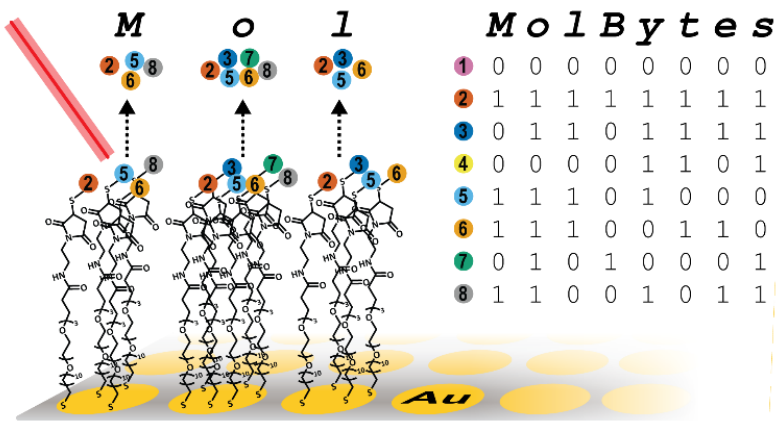
Molecular mixtures



Brown: Large synthetic library of small molecules



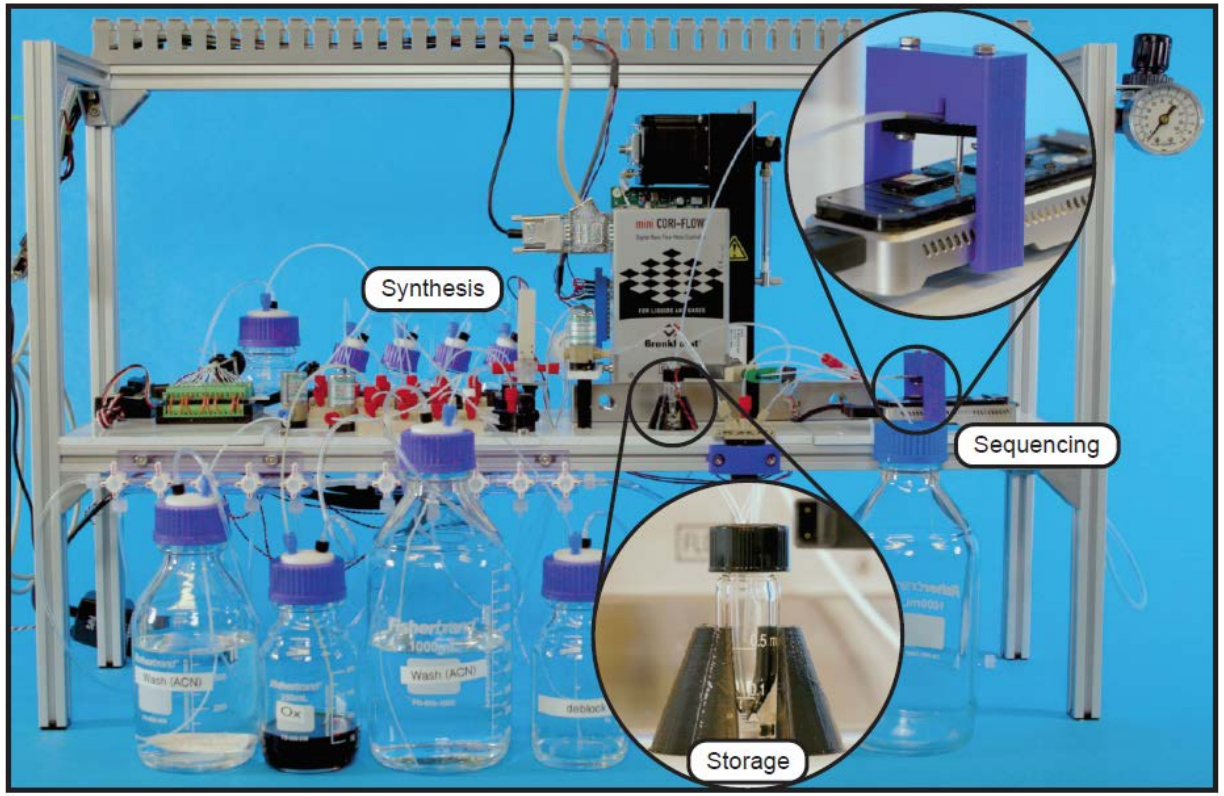
Illinois: Enzymatic nicking of natural DNA



Harvard: Small library of peptide mixtures

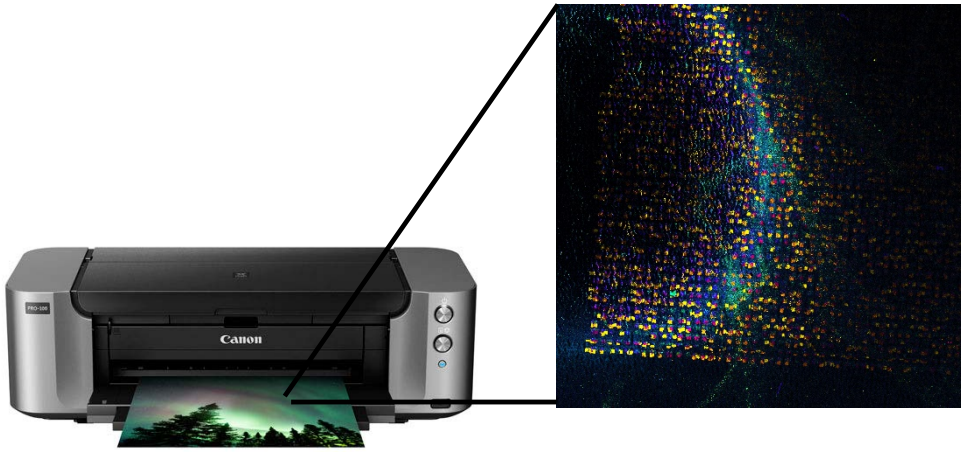
Moving to practical scales

DNA

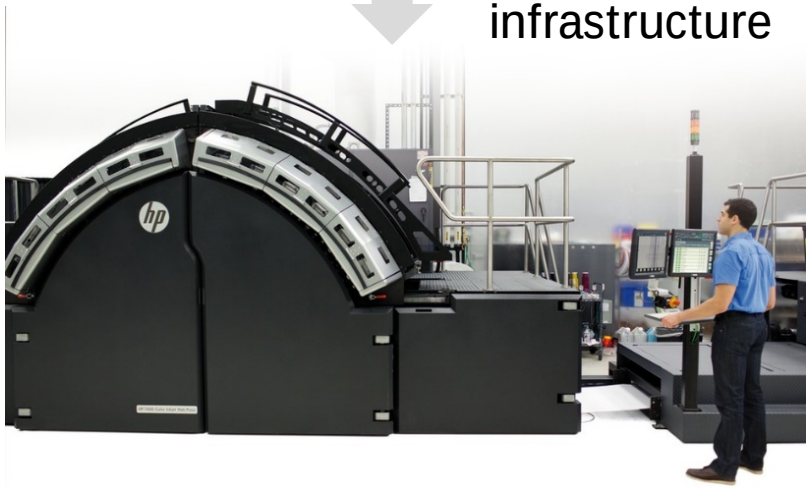


End-to-end automation

Molecular mixtures



Exploiting existing infrastructure



www.darpa.mil