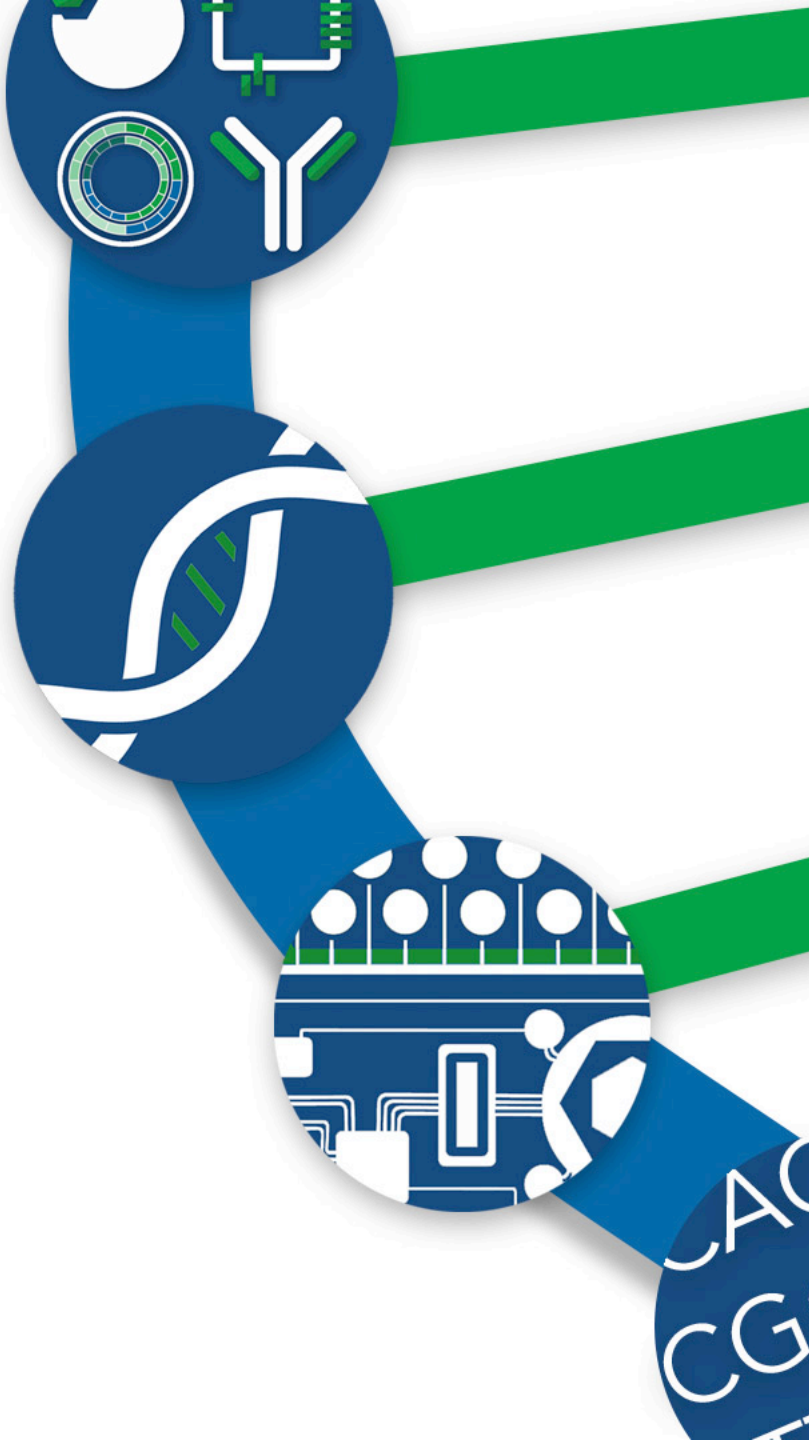
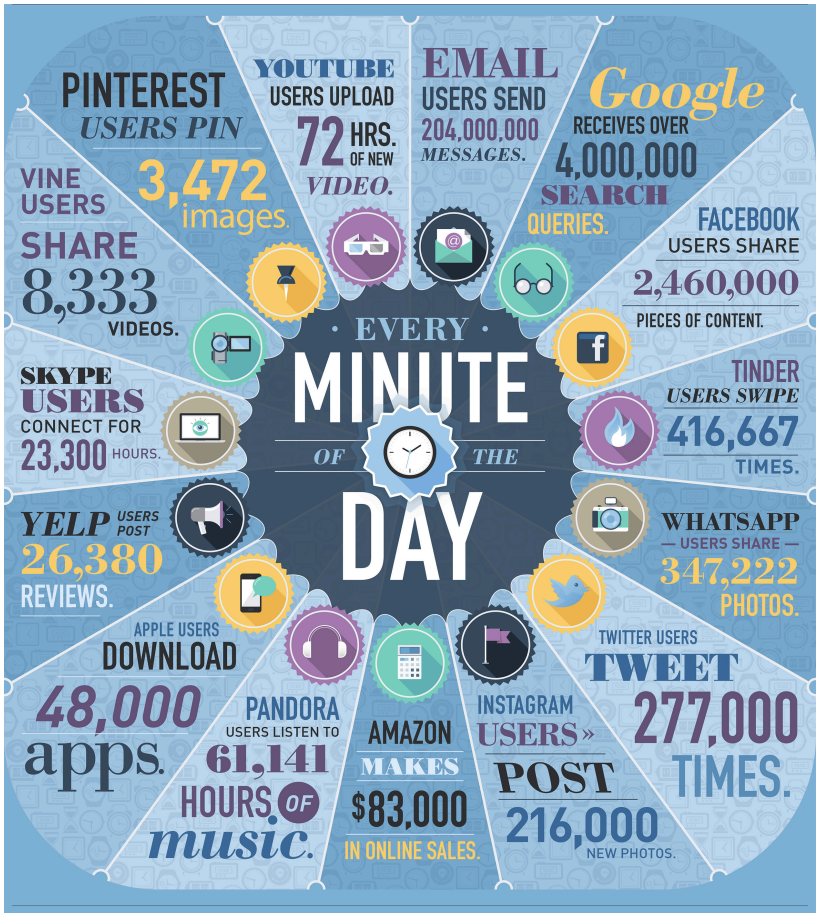
DNA-based Data Storage

Devin Leake
VP of R&D

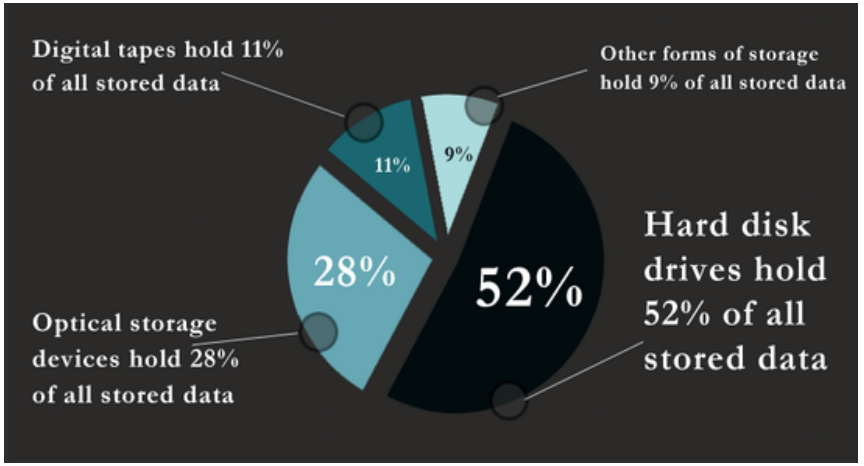
Gen9



There is a growing need for storage



Content created by DOMO, "Data Never Sleeps"



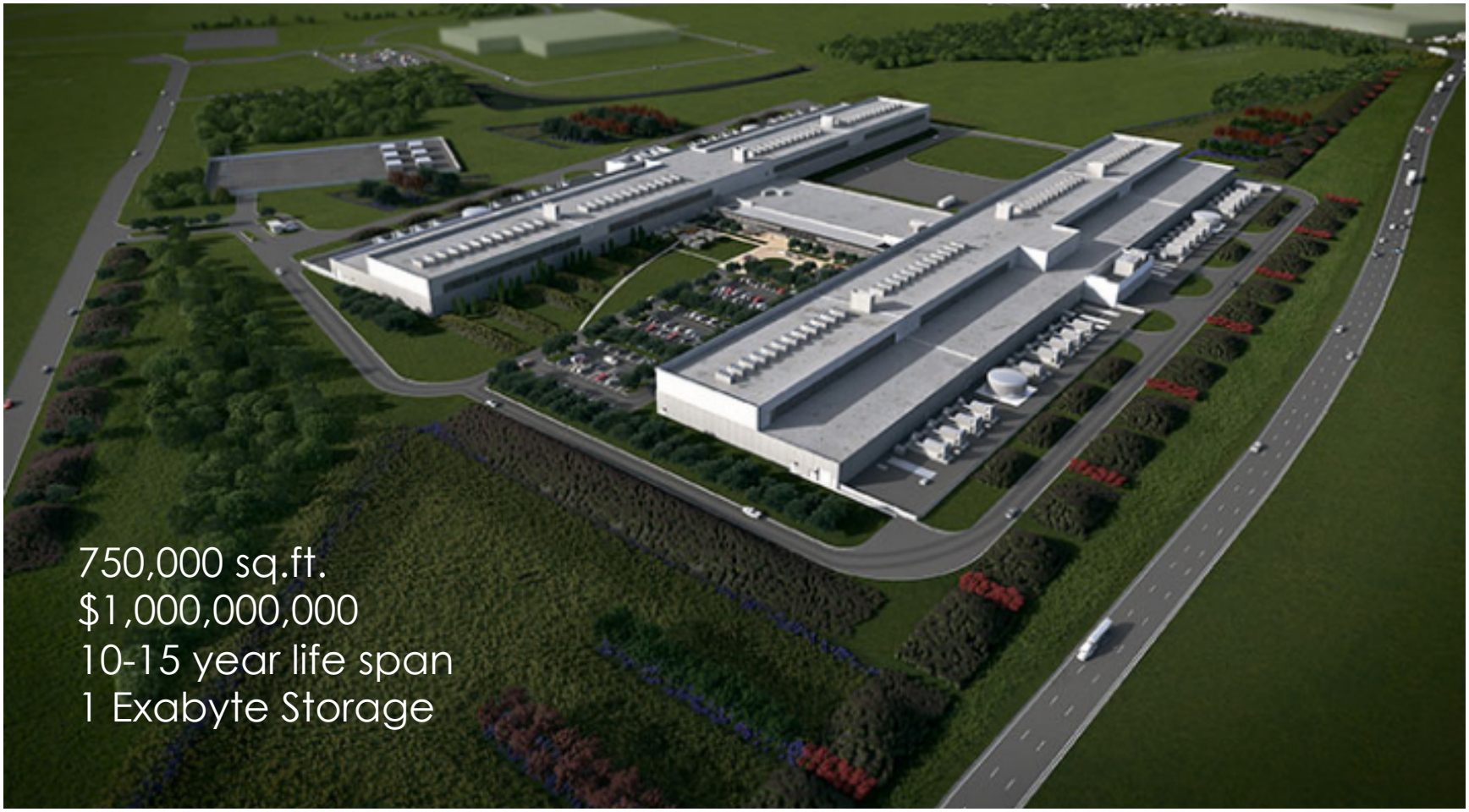
https://mozy.com/infographics/where-is-the-worlds-data-stored

- By 2020, 40 zettabytes will be generate
- Predominant form of storage is tape

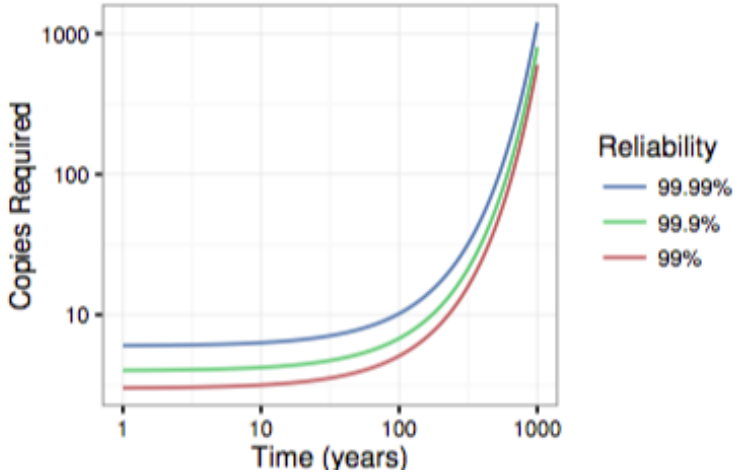
IDC (International Data Corporation)

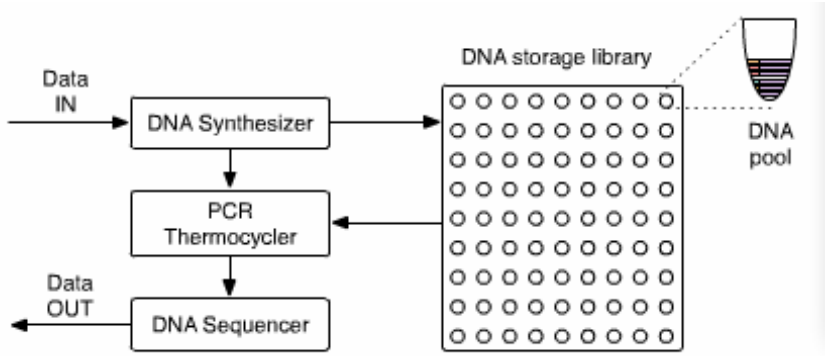
Gen9

Facebook Facility in Fort Worth



DNA-based Data Storage

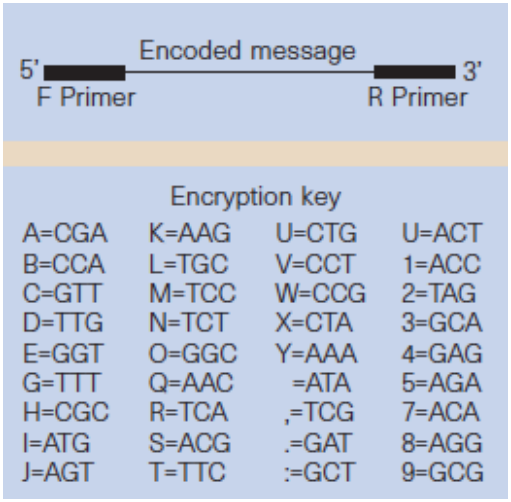




Bornholt, et.al. ASPLOS 2016

- DNA has been recovered and analyzed from:
  - Wooly mammoth (20,000 years ago)
  - Neanderthal (40,000 years ago)
  - Bison (60,000 years ago)
  - Pollen samples (~500,000 years ago)
- DNA Storage Process:
  - Data IN: Synthesize DNA that contain user information
  - Data OUT: Sequencing is performed via high-throughput sequencing technologies (e.g. Illumina, PacBio, Oxford Oxford Nanopore Technologies)

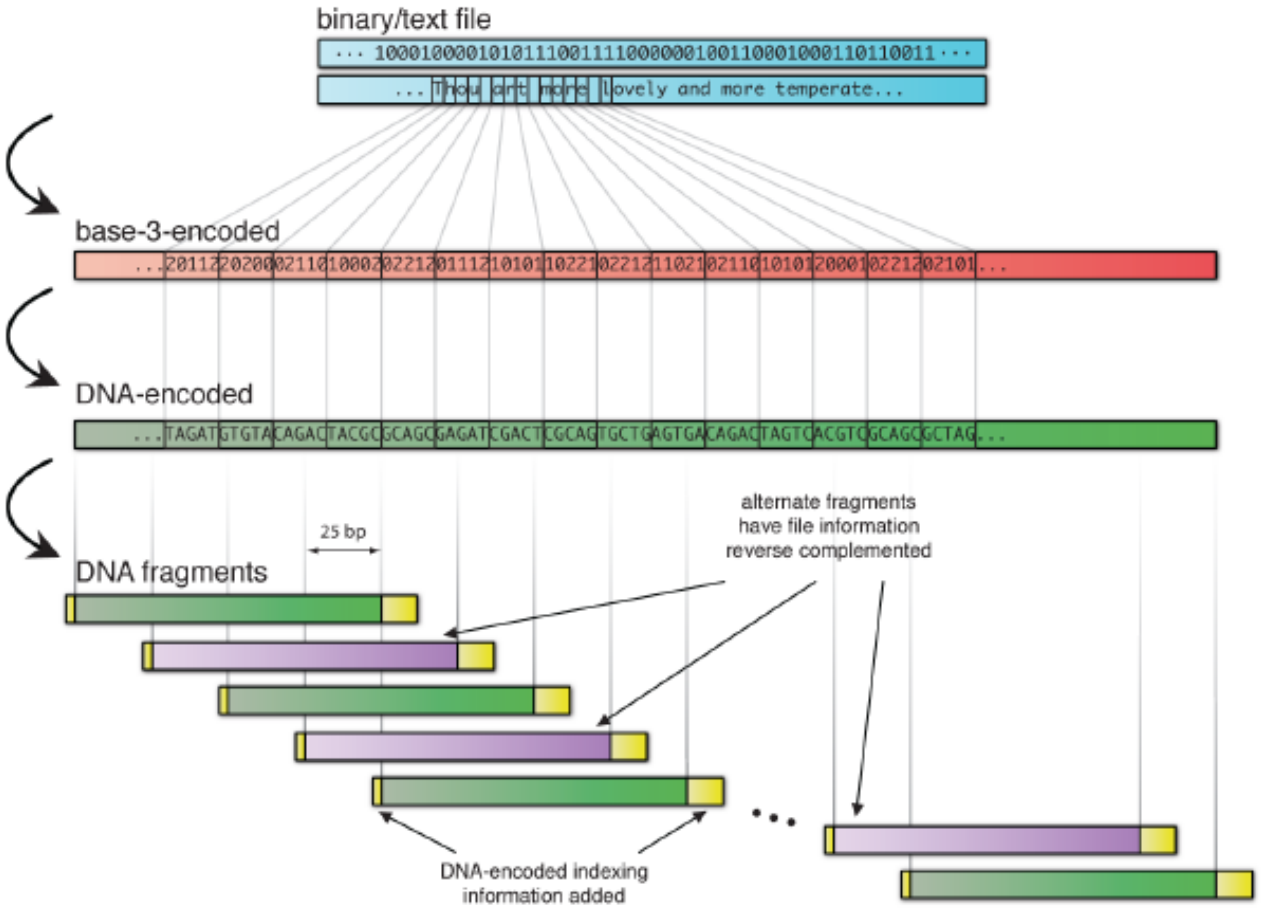
Proof-of-concept for DNA-based Storage



Clelland, Risca and Bancroft Nature 1999

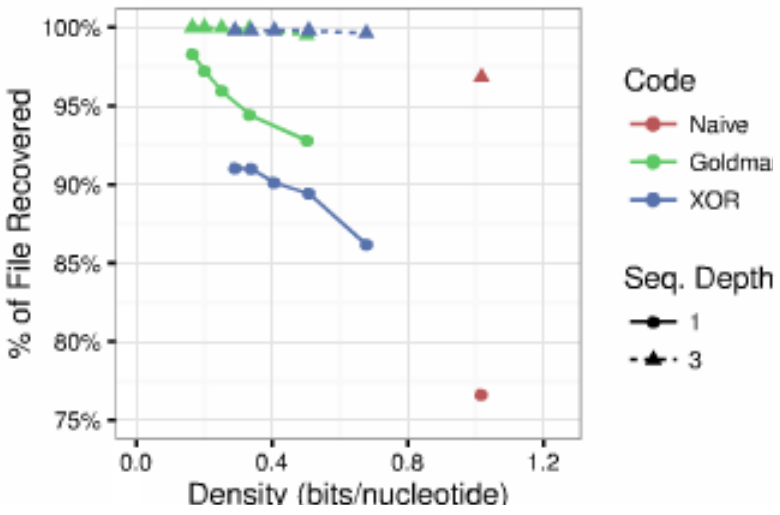
- Simple encoding scheme (substitution cipher) used for initial experiments
- Microdot storage similar to prior 'espionage' formats
- Indexing and access primers flank DNA message

Evolution of Encoding Architecture

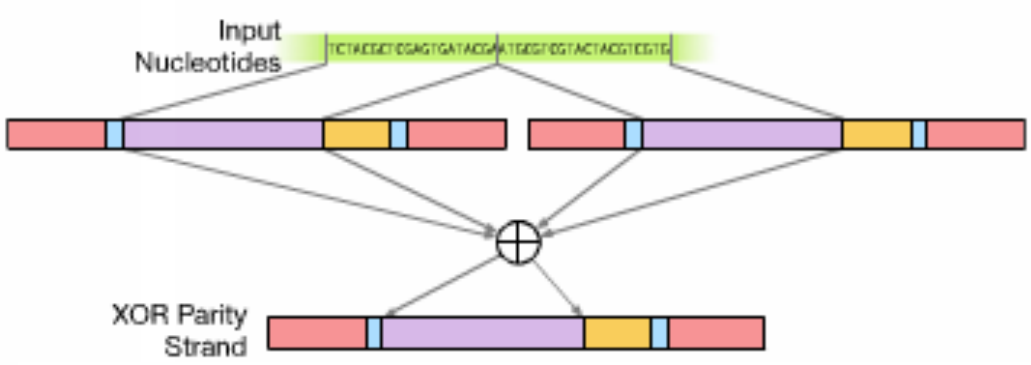


Goldman, et.al. Nature 2013

Evolution of Encoding Architecture



Bornholt, et.al. ASPLOS 2016



Encoding architecture maximizes content density and redundancy while avoiding biophysical features (e.g. homopolymers) that impair storage performance

Volumetric Density of DNA

	Volumetric Density (Bits/cm^3)
Hard Disk	**10^13**
Flash Memory	**10^16**
DRAM	**10^13**
Cellular DNA	**10^19**

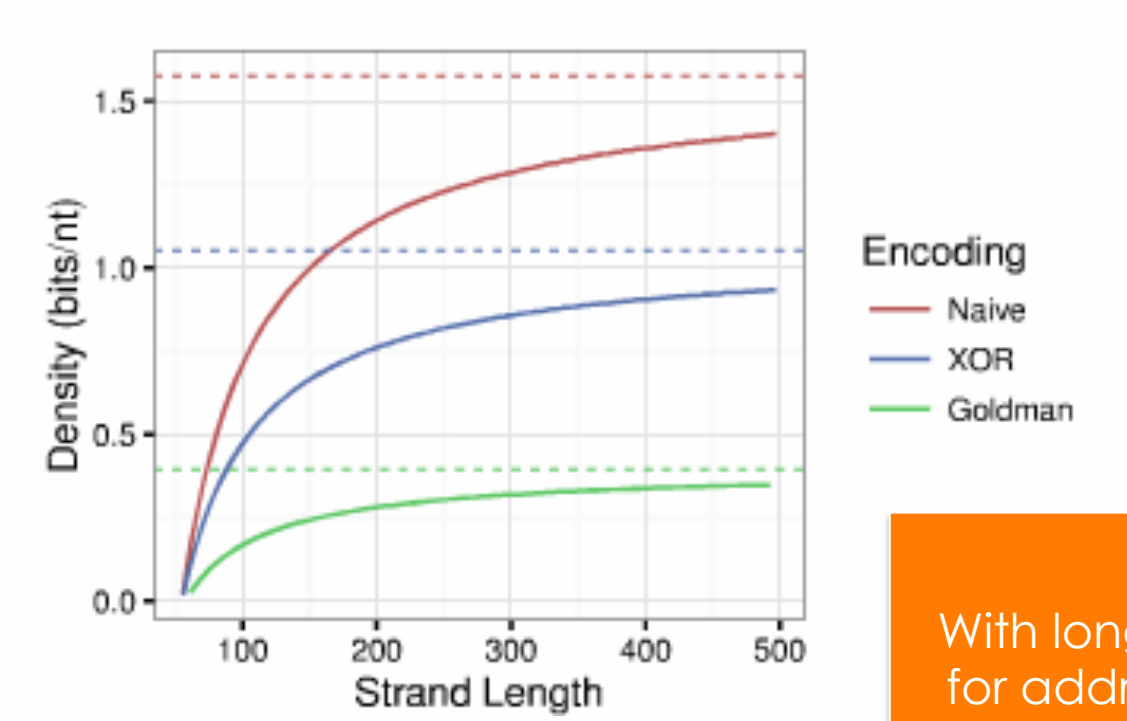
Zhirnov et.al. Nature Materials 2016

12.5 Zettabytes of Storage using DNA

DNA provides several orders of magnitude higher density than traditional storage media

Gen9

Determining Optimal Length of DNA



Bornholt, et.al. ASPLOS 2016

With longer DNA, sequences for addressing and indexing 'payload' are a smaller percentage of overall length

Types of Data Storage Based on Need

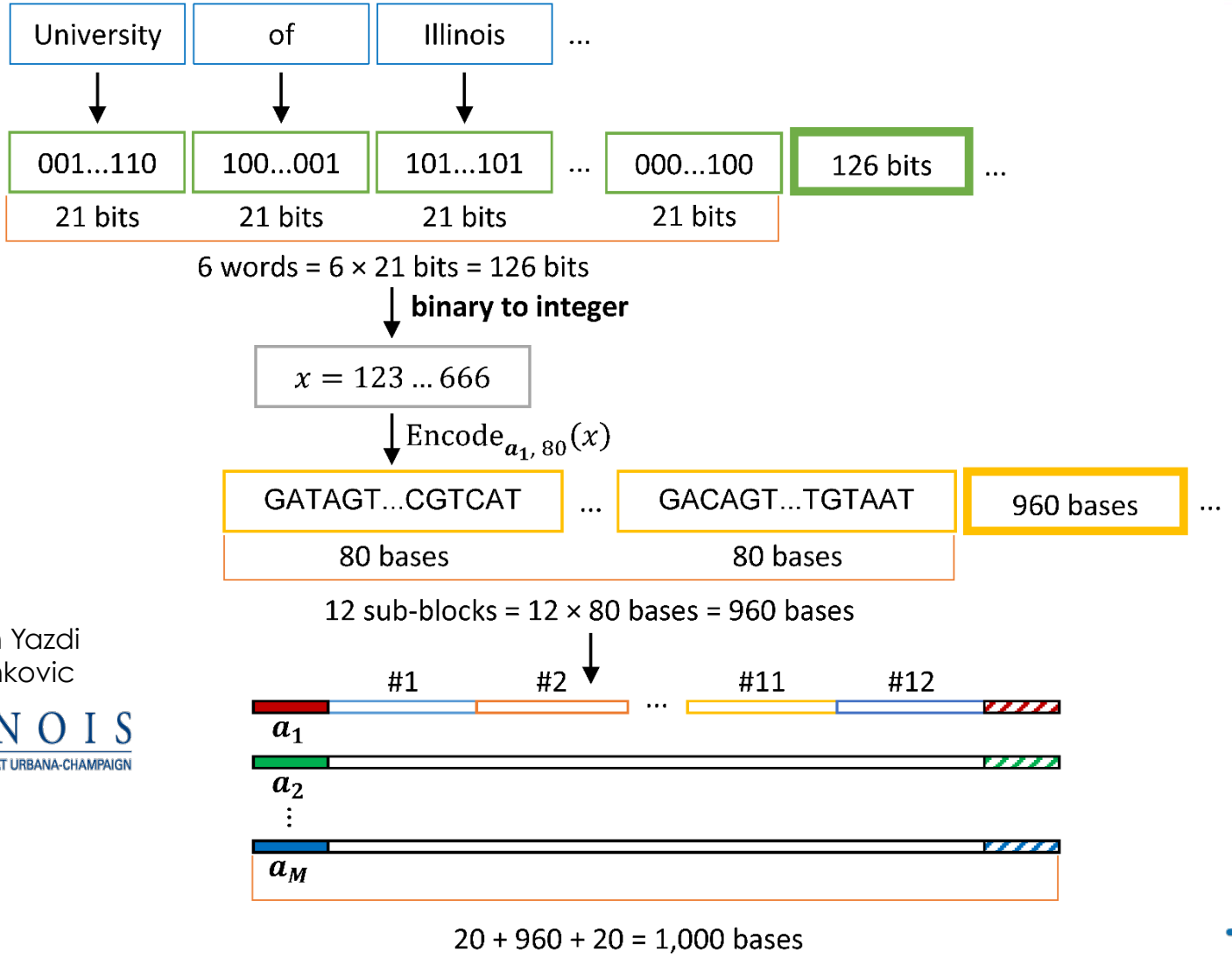
- Uninterrupted availability
- High performance read/ write access

- High media capacity
- High performance read/ write access
- Low storage cost per GB

- Data authenticity – data is protected from modification
- Extended media longevity/ durability
- **High performance random read access**
- Extremely low storage cost per GB

Gen9

Coding for Random Access Scheme

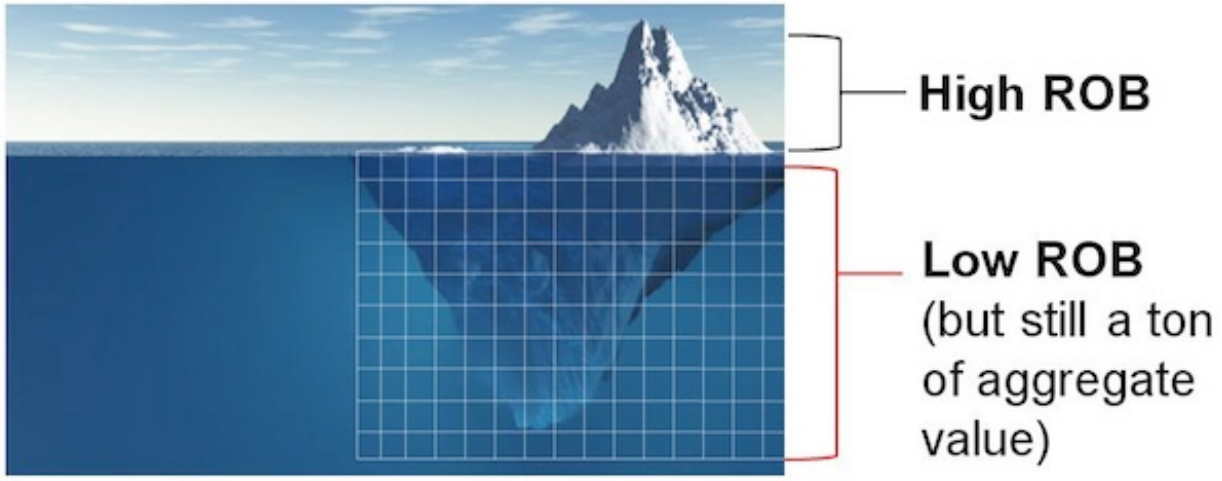


S. M. Hossein Yazdi
Olgica Milenkovic

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Gen9

Active Archive Model

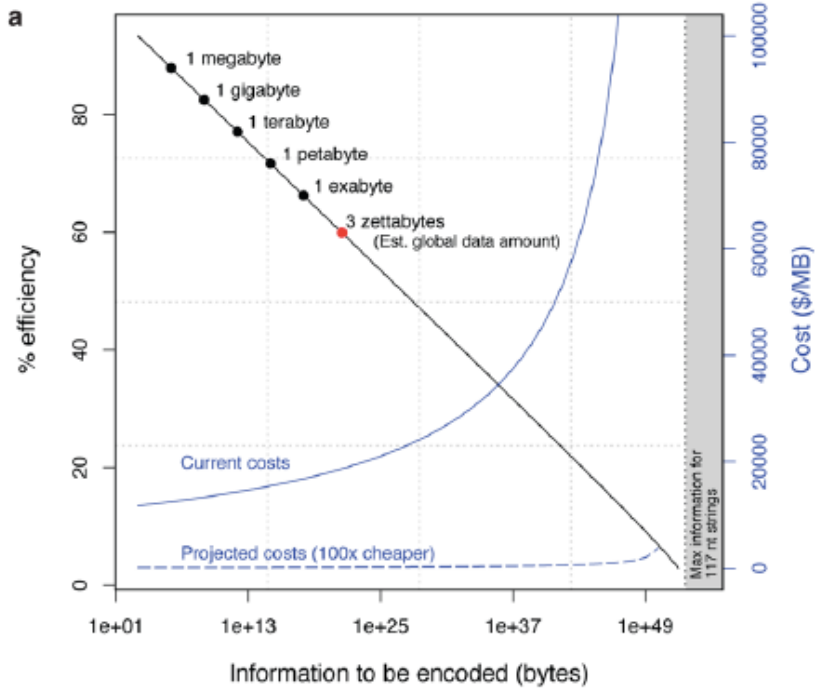


Big Data's New Use Cases: Transformation, Active Archive, and Exploration. Cloudera. February 21, 2013 by Amr Awadallah

- Method of tiered storage using multiple systems
  - Multiple media employed for optimal storage (Return on Byte)
  - Metadata used to organize stored data
- Model facilitates necessary deployment of support systems, such as organizational AI

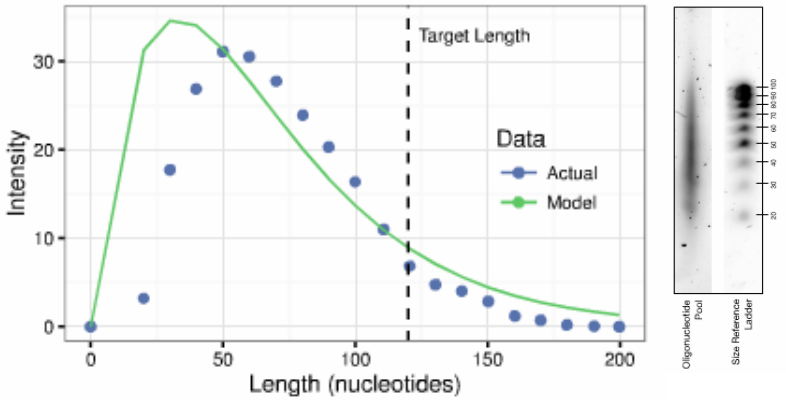
Gen9

Reducing Cost of DNA-based Storage



Goldman, et.al. Nature 2013

- Synthesis is currently the primary driver for price (2 orders of magnitude more than 'reading')
- Improving synthesis efficiency regain ~1 order of magnitude)



Bornholt, et.al. ASPLOS 2016

FOUNDERS

JOSEPH jacobson
MIT
FABRICATION

GEORGE church
Harvard
GENETICS

DREW endy
Stanford
BIO-ENGINEERING

INVESTORS

Agilent Technologies

THE KRAFT GROUP

DFJ

CAMBRIDGE
MASSACHUSETTS

2009
COMPANY FOUNDED

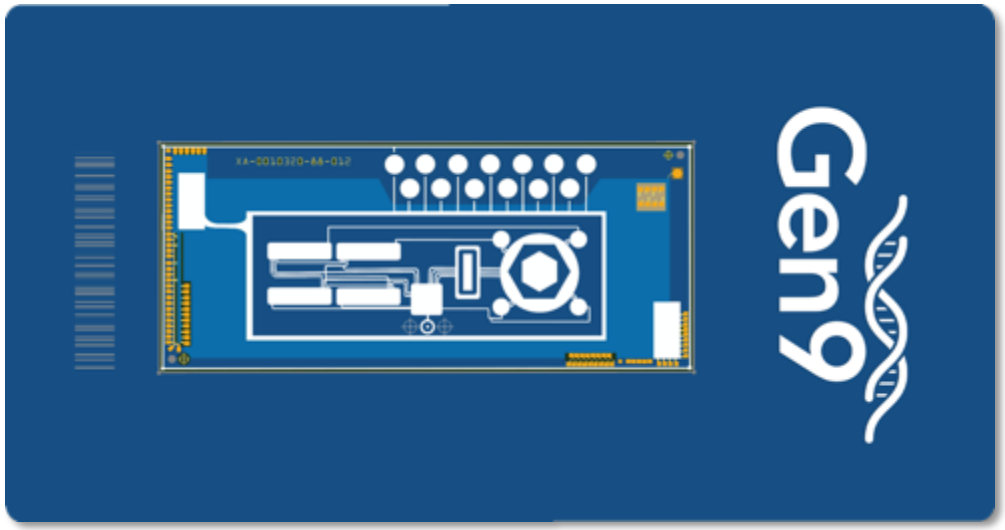
2012
FIRST COMMERCIAL SALE

2015
OVER 50 EMPLOYEES

Gen9

The Synthetic Biology Solution

Using **semiconductor** technology, Gen9 developed a high-throughput process to build **DNA** at a very high quality and low cost



Thank You!

Gen9