DNA Computing With an Associative Memory

CATALOG

David Turek
CTO
June 21, 2025

Based on “Demonstration of a Scalable DNA Computing Platform: Writing and Selection”, Bhatia et al, ACM Journal on Emerging Technologies in Computing Systems, June 2025

Why DNA Computing

“To Out Compete, one must Out Compute”

Council on Competitiveness, 2006

- Computing is a pathway to strategic insight
  - New products
  - New actions/behaviors
  - New insights and theories
- Computing infrastructure is a shared resource
  - You rarely have what you need

CATALOG

State of Computing-The Virtue of Parallelism (and the “Myth” of Big Data)

- Vocabulary
  - Strong Scaling
    - Addition of more resources reduces the time to solve a fixed problem
      - Ex: doubling processors cuts compute time in half
  - Weak Scaling
    - Increasing resources proportionately (data and compute) maintains solution time
  - Amdahls Law
    - The serial portion of an application is the limiting factor to the impact of parallelism

- Adding resources is complicated and expensive
  - Incremental CapEX
    - Adding one unit of incrementat computing to Summit was $65K for .00025 increase in compute
  - Disruption of integration

Our Core Belief: DNA Computing can add computing capacity cheaply and non-disruptively for strong scaling applications

CATALOG

Example Applications (the "Myth of Big Data")

Strong Scaling with Small Data

- Molecular Dynamics (drug design, medical research)
  - Few MB of data
  - Computationally intensive (N-Body Problem)
- Cryptographic Key Search
  - Key space is less than an MB
- Graph Optimization (scheduling and routing)
  - 1000 edges is a few MB
- ML Hyperparameter Tuning for Edge Devices
  - Few MB
- CFD (auto and aero)
  - 500MB
- Finite Element Analysis
  - 1GB

Near term DNA Computing opportunities

Strong Scaling with Moderate Data

- Protein Folding: 5-20GB
- Antenna Design 10-40 GB
- Portfolio Analysis 5-30GB
- Clustering Customer data 10-50GB
- Log Analysis for Cyber Forensics 5-40GB
- Supply Chain Optimization 10-30 GB

Intermediate term DNA Computing opportunities

CATALOG

DNA is one of the mediums that could store and compute data

You can store data in DNA…

DNA storage has been demonstrated by various research organizations and developed

[History of DNA storage]

- **'88** ▪ 5x7 matrix image was stored in a DNA sequence in E.coli
- **'16** ▪ 22 MB of a MPEG compressed movie sequence were stored and recovered from DNA
- **'18** ▪ Method known as DNA Fountain that stored data at a density of 215 petabytes per gram of DNA was published
- **'19** ▪ 16 GB of Wikipedia have been encoded into synthetic DNA by Catalog Technology

…And compute DNA

Concept of DNA computing has been researched and developed, laying the groundwork for commercial viability

[History of DNA computing]

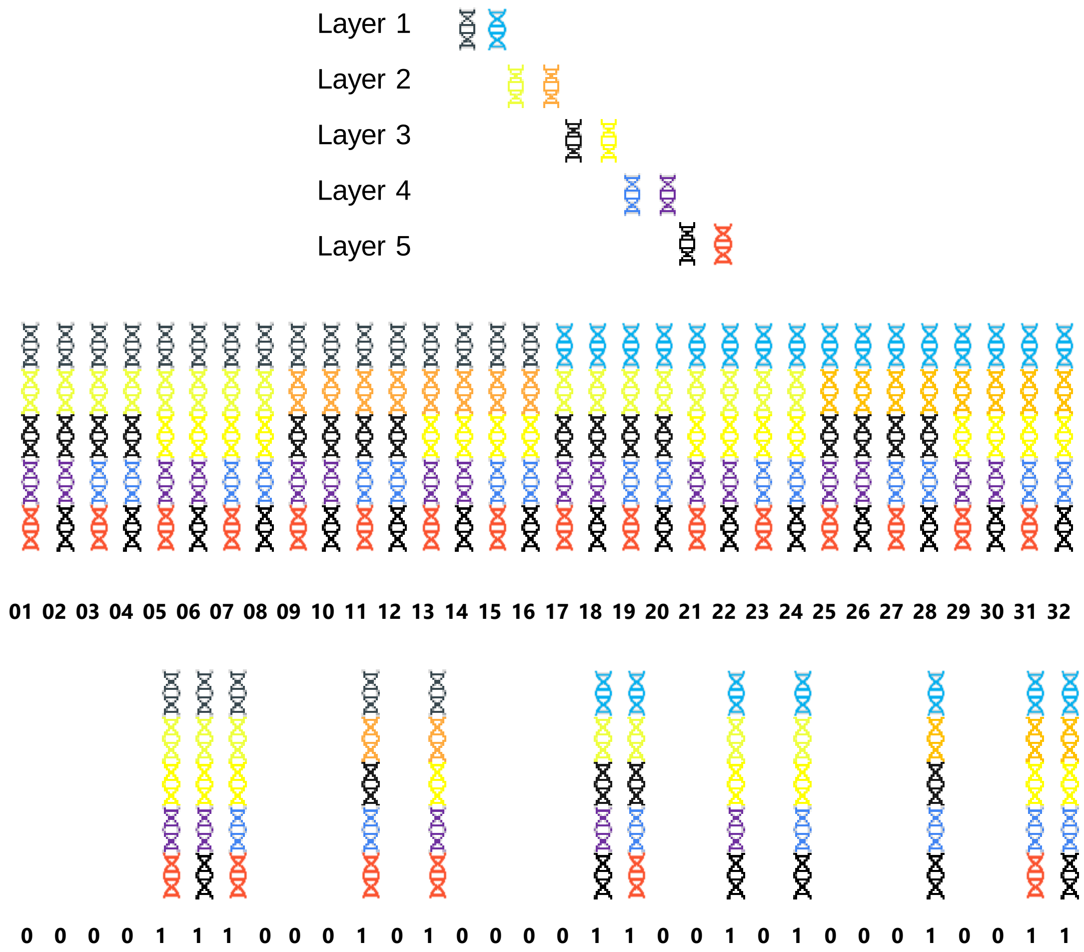
- **'94** ▪ Adelman presented the first prototype of a DNA computer, solving a seven-point Hamiltonian path problem
- **'02** ▪ Created a DNA computer able to play tic-tac-toe against a human player
- **'11** ▪ Developed a DNA-based artificial neural network that can recognize 100-bit hand-written digits.
- **'19** ▪ Produced renewable DNA logic circuit bringing the technology one step closer to the silicon-based computing

DNA Computing Platform

- High data ***density*** and massive ***parallelism*** (the power of "one step") will illuminate "dark data" that would otherwise be in accessible let alone analyzable
- Unlike quantum, DNA computing is ***broadly applicable*** in a variety of latency-tolerant applications

CATALOG

How CATALOG encodes information in DNA



Pre-made short pieces of DNA are combined into longer molecules

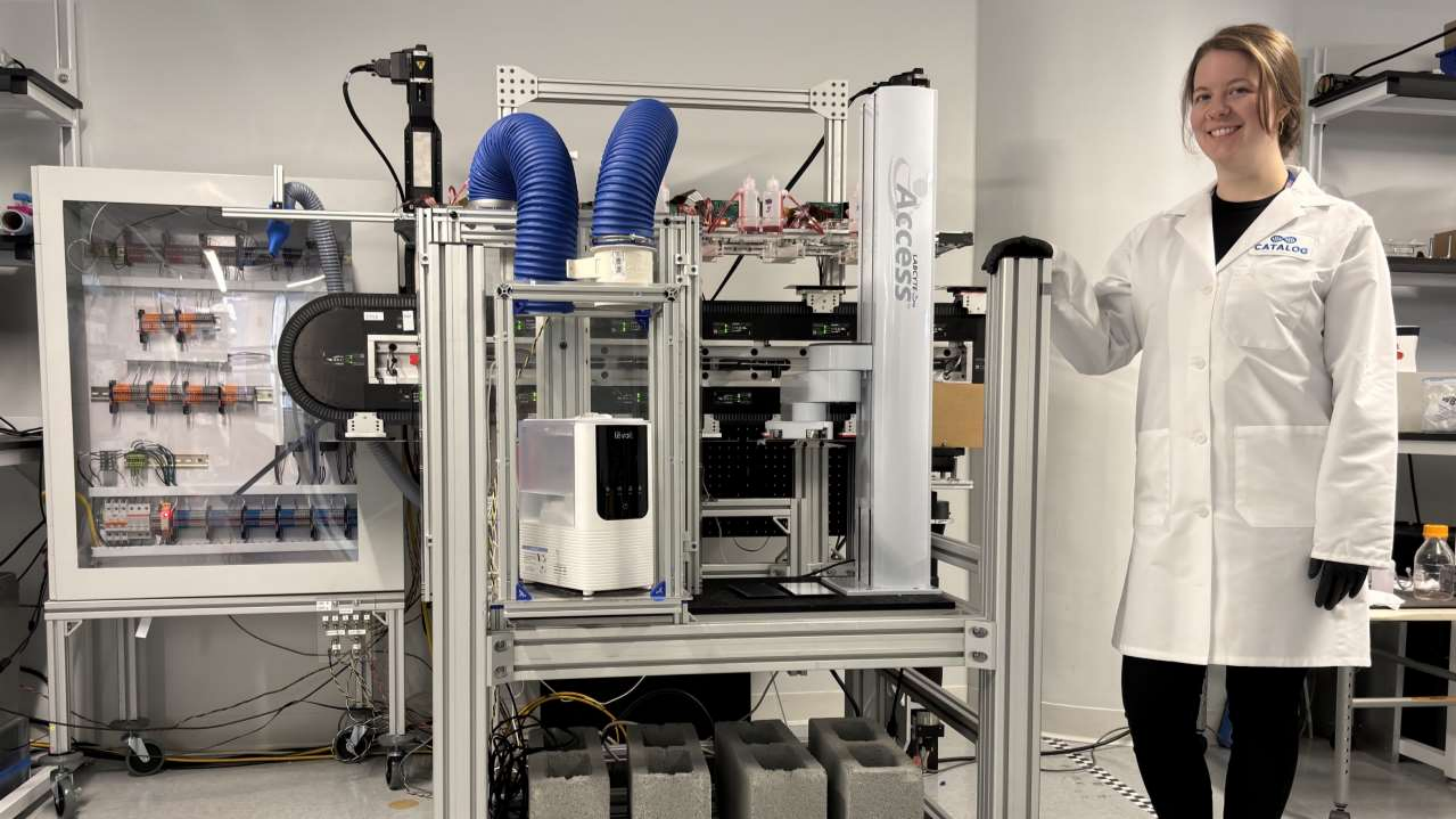
All possible assemblies of the DNA sequences create an ordered **combinatorial space**

Use DNA sequences analogously to characters in a typesetting machine

https://koreajoongangdaily.joins.com/2023/04/12/culture/koreanHeritage/jikji-bnf-gutenberg/20230412195850132.html

CATALOG

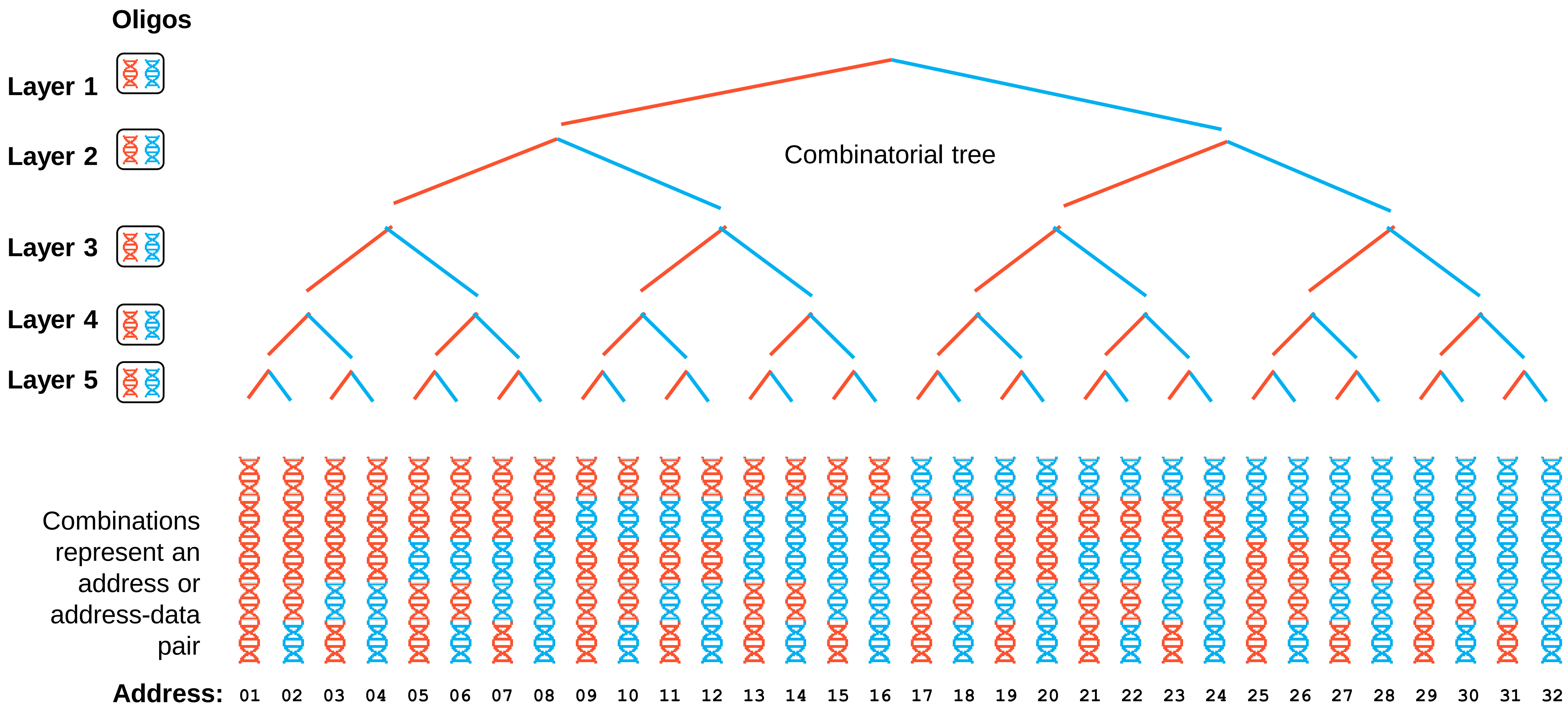

CAUTION


Access
LABCYTE
CATALOG

What is a Parallel Energy-Efficient In-Memory DNA Computer?

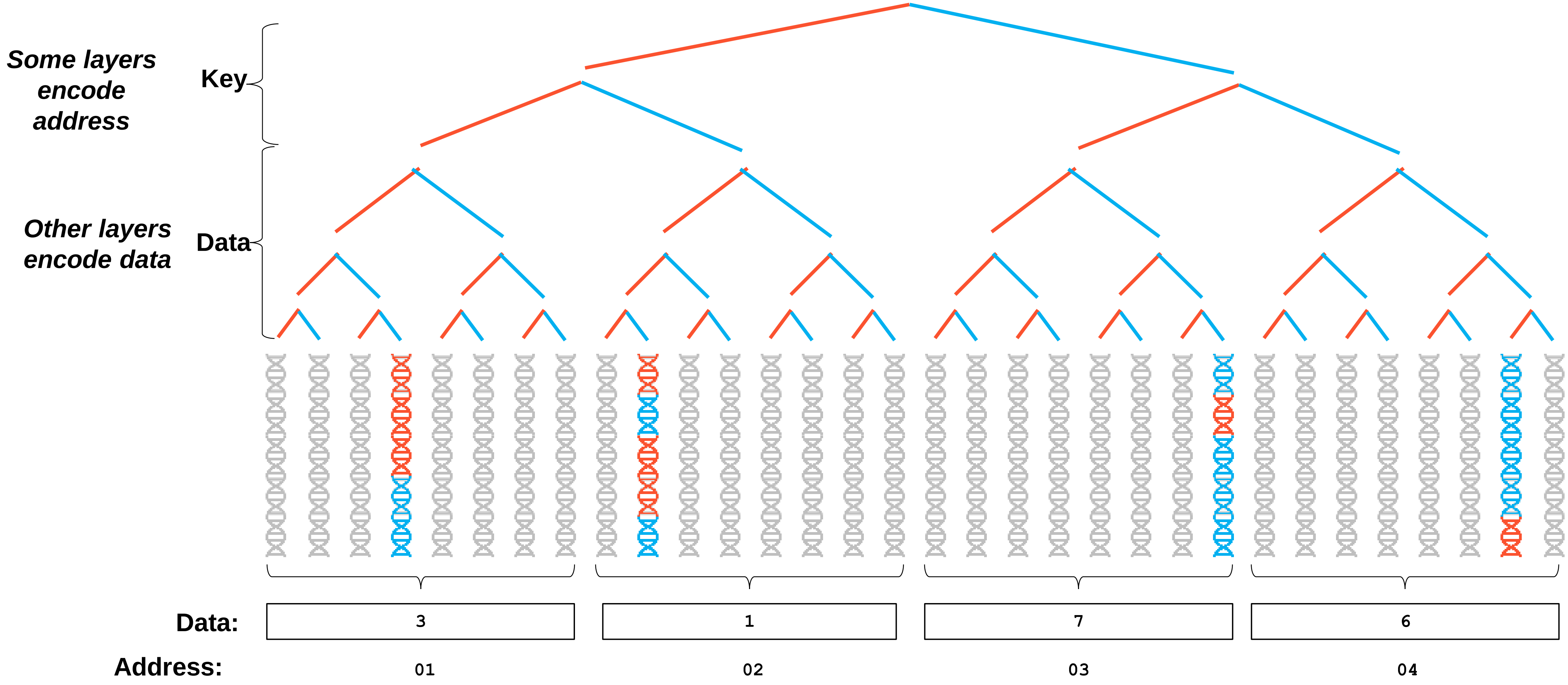
- Uses synthetic DNA to encode input, memory, and output
  - High density and longevity: inputs, memory, and output can be stored at micron scale for 1000s of years at ~zero energy cost without bit rot
- Uses DNA chemistry to transform inputs to outputs
  - Chemistry is executed on inputs encoded in DNA by automated robots or fluidic chips
- Computation happens within DNA memory without any fetch/store
- An instruction can be applied to all desired bits in parallel
  - Unlocks million- to billion-fold parallelism
- Instructions manipulate DNA, an energy-efficient molecule
  - Theoretically, a petabyte in DNA can be copied with only 55 J of energy (< J in 1 MLB pitch)
- Higher latency, of the order of minutes to hours
  - Great for latency-tolerant applications on high value data, and high throughput

CATALOG writes data using combinations of DNA oligos



Scales exponentially: ***L* layers of *N* components enables N^L bits of information with only LN oligos**

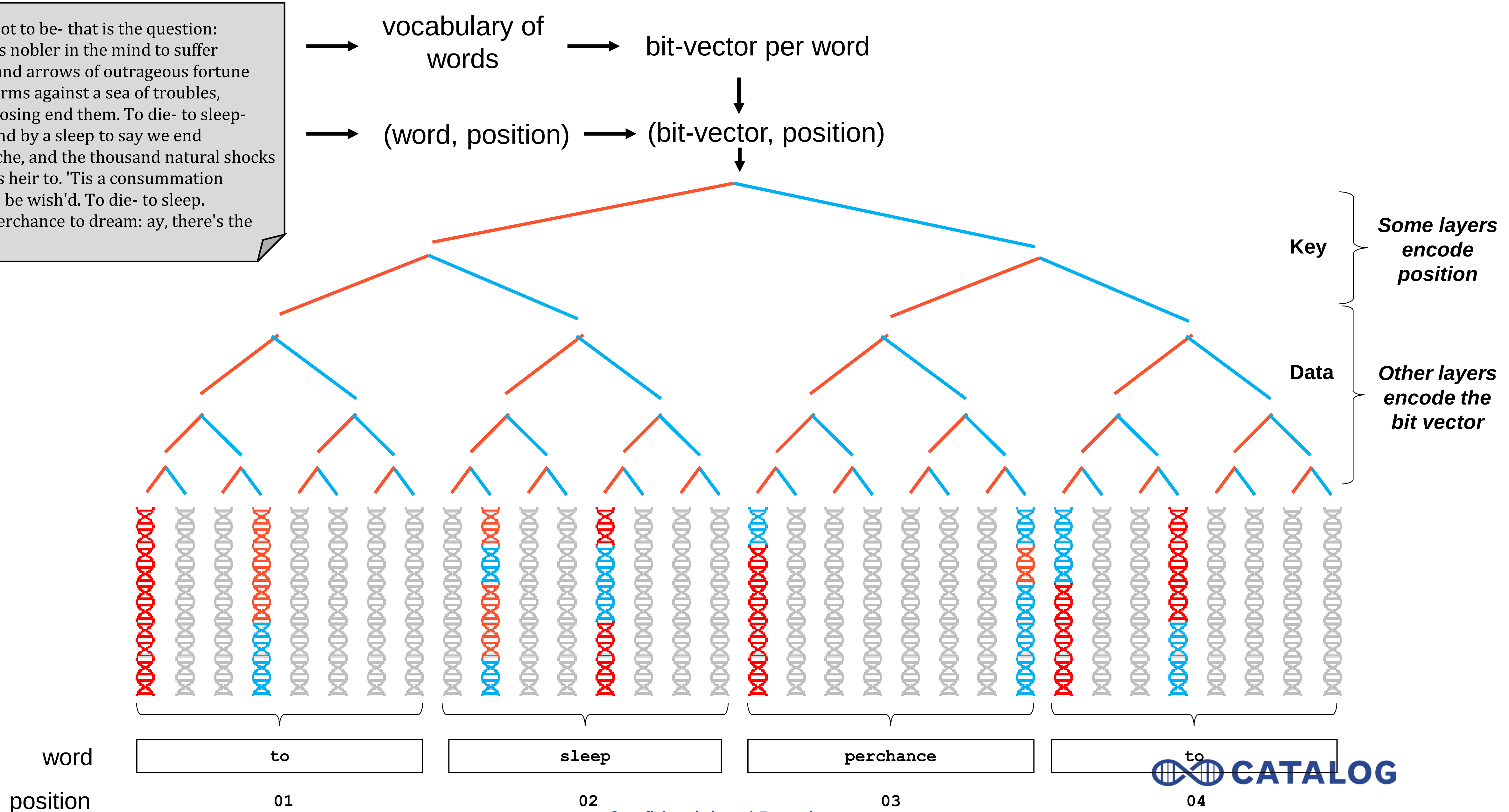
Each combination represents an address-data pair



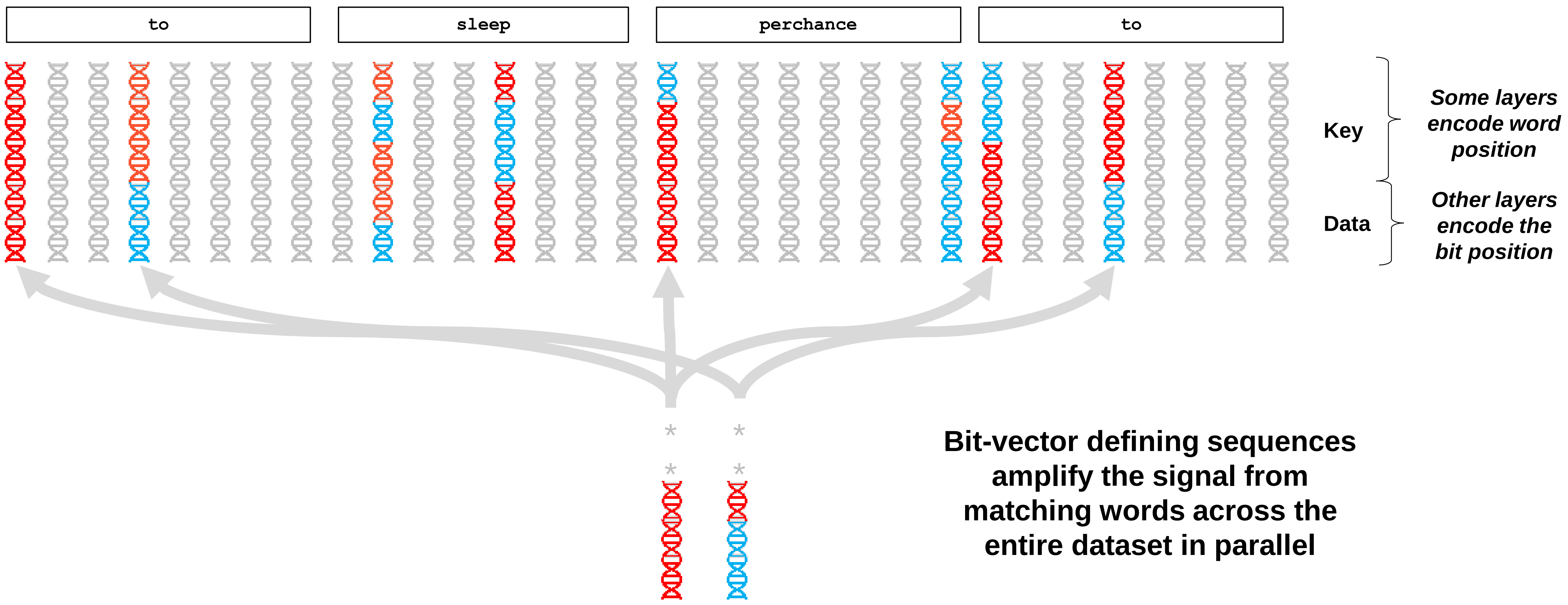
CATALOG

Demonstration of the *select* instruction in a parallel text search example

"To be, or not to be- that is the question:
Whether 'tis nobler in the mind to suffer
The slings and arrows of outrageous fortune
Or to take arms against a sea of troubles,
And by opposing end them. To die- to sleep-
No more; and by a sleep to say we end
The heartache, and the thousand natural shocks
That flesh is heir to. 'Tis a consummation
Devoutly to be wish'd. To die- to sleep.
To sleep- perchance to dream: ay, there's the
rub!"



Demonstration of the *select* instruction in a parallel text search example



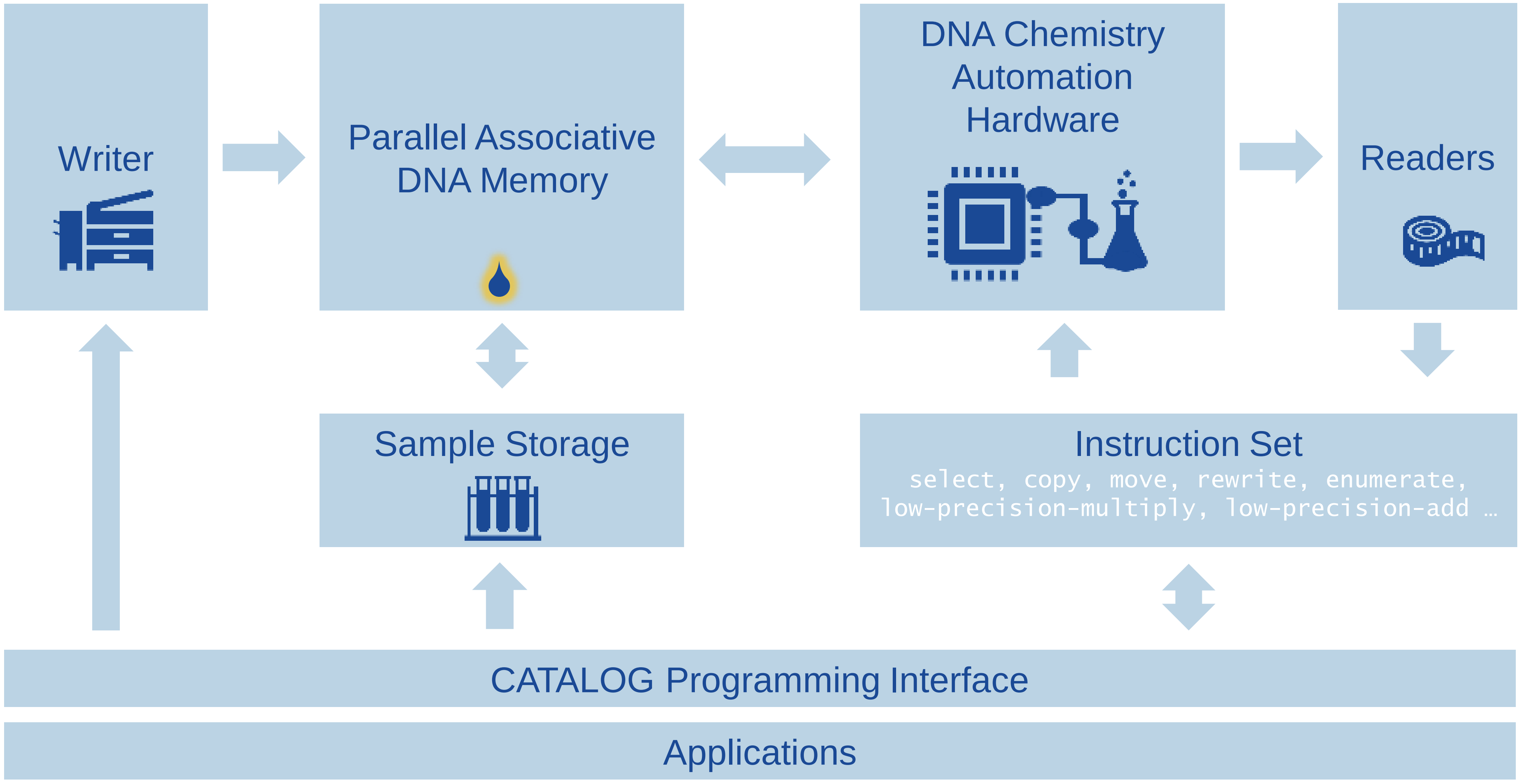
Bit-vector defining sequences amplify the signal from matching words across the entire dataset in parallel

query word → bit-vector → DNA sequences defining bit vector

"to" (10010000) (10010000)

CATALOG

A Data Parallel Energy Efficient In-Memory DNA Computer



Summary

- DNA Computing can beat conventional computers on *operations per joule* because DNA is amenable to more efficient transformation and preservation than electromagnetic symbols. DNA Storage beats conventional storage on *longevity, density, and energy consumption* because DNA is error-resistant, information-dense, and energy efficient
- Together DNA-based storage and computing can beat conventional computing on energy-efficiency, density, and longevity (*e.g.*, a 10-year-old petabyte can be searched with a few ~KJ in a small DNA computing lab)
- DNA Computing is broadly applicable to applications with
  - long-term valuable data
  - latency-tolerant
  - sufficiently large scale of data
  - addressable with CATALOG's growing instruction set
- Current applications include search, inference, and signal processing, intractable problems, and ML training

CATALOG