

# Lost in Translation: Harnessing the Power of Nanopore Electrical Signals in Genomics

CDSL Seminar Talk, MD, USA  
February 11, 2026

**Can Firtina** | [canfirtina@gmail.com](mailto:canfirtina@gmail.com) | [cs.umd.edu/~firtina](https://cs.umd.edu/~firtina)

Assistant Professor of Computer Science at UMD



DEPARTMENT OF  
COMPUTER SCIENCE

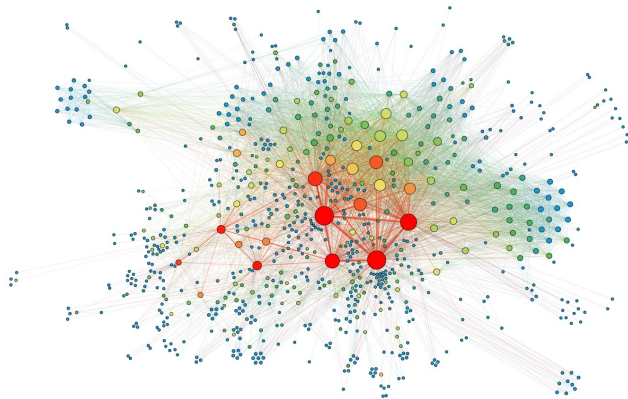
# Big Data is Everywhere



**Large recommender systems**



**Large Language Models**



**Graph/Tree Processing**

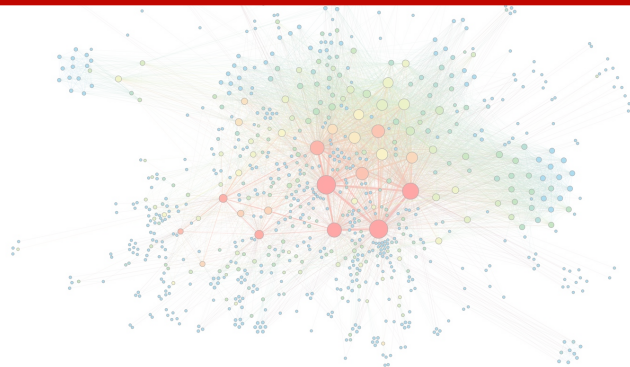


**Genomics**

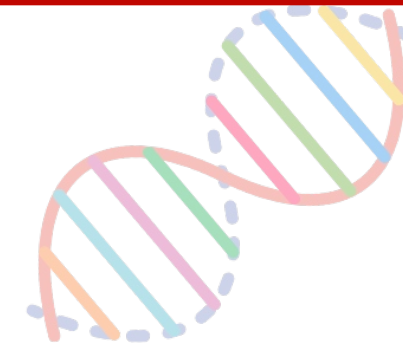
# Computing is Bottlenecked by Data



**Data movement →  
Performance & Energy Bottleneck**

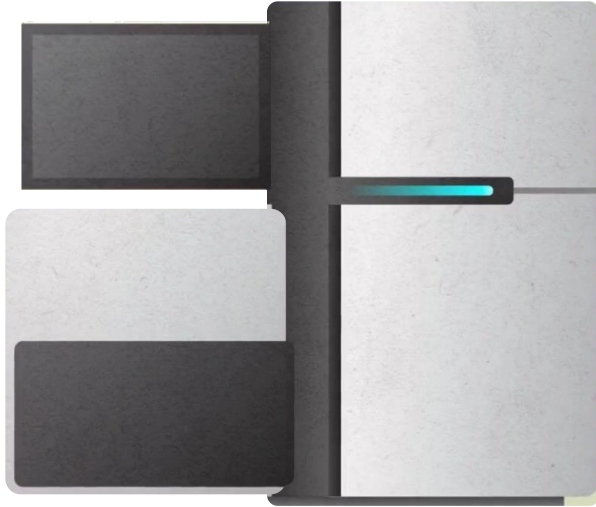


**Graph/Tree Processing**



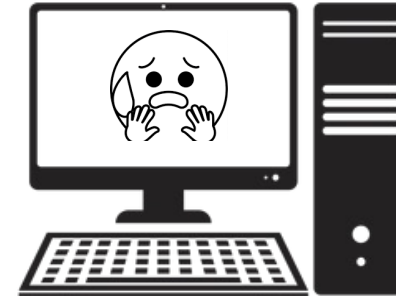
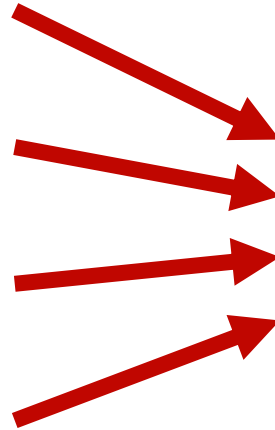
**Genomics**

# Problems with Data Analysis Today



**Special-Purpose** Machine  
for **Data Generation**

**FAST**



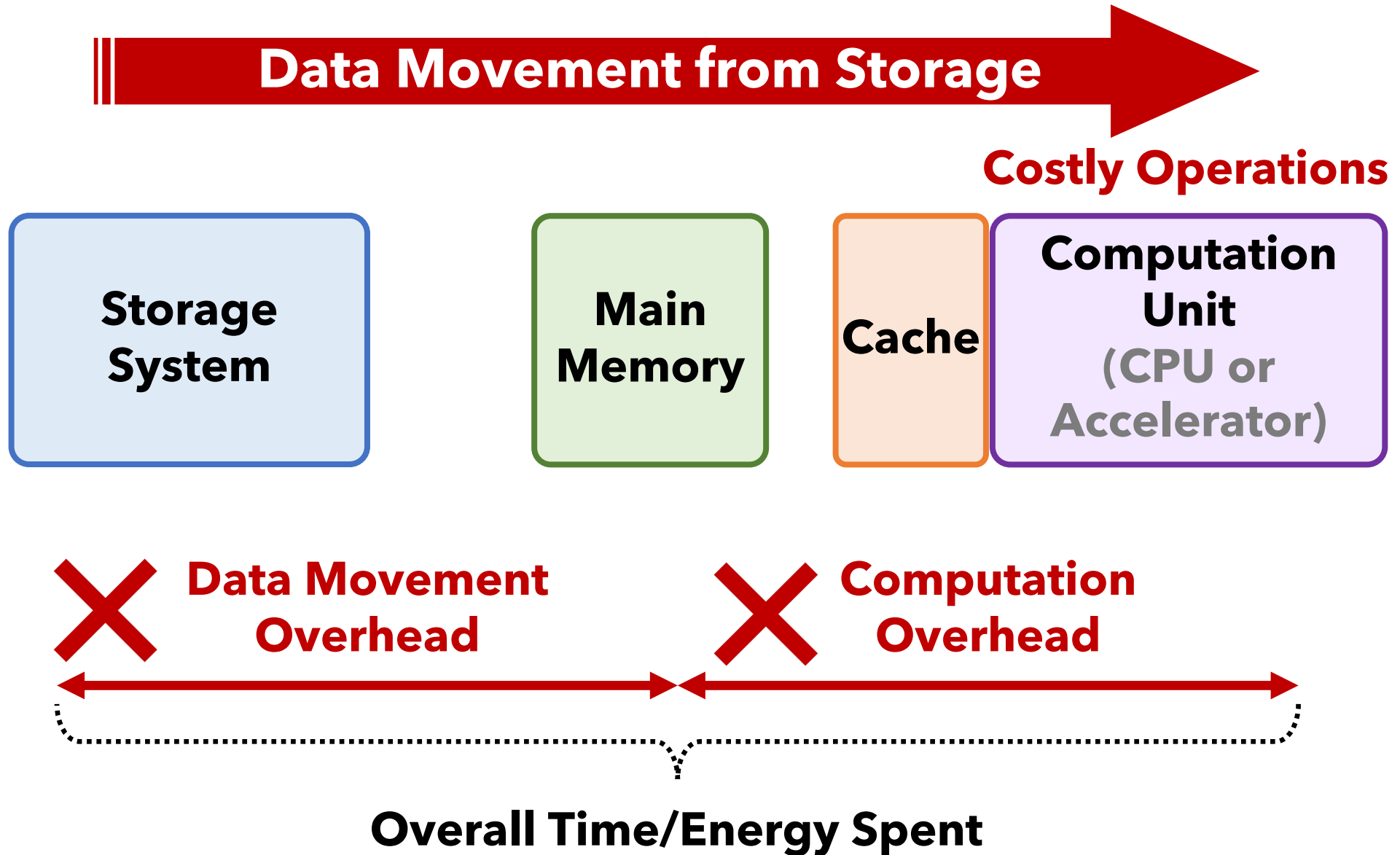
**General-Purpose** Machine  
for **Data Analysis**

**SLOW**

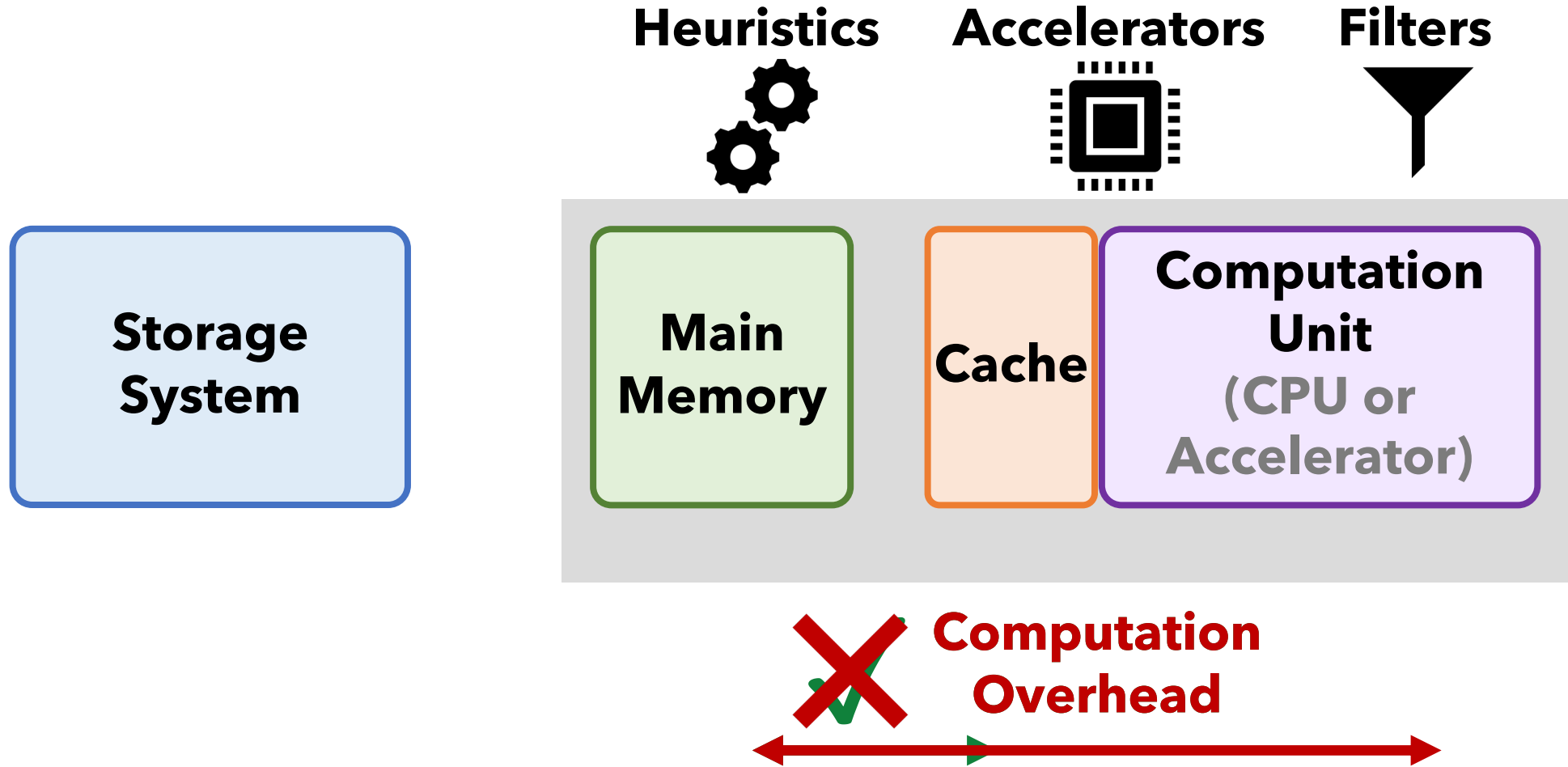
**Processing capabilities oblivious to data analysis requirements**

**Large amounts of data movement**

# A Typical (Biological) Data Analysis



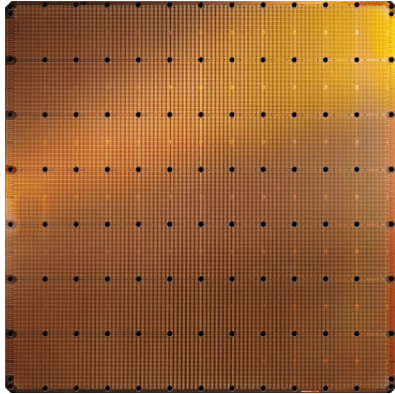
# A Typical Accelerated (Biological) Data Analysis



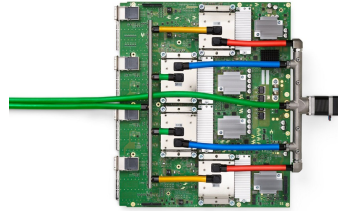
# Acceleration Efforts for AI and Genomics

## Cerebras WSE-3

Wafer-scale ML accelerator chip



## Google's TPUs:



## NVIDIA H100



### Scale your studies with ease

Process high-throughput data quickly with hardware acceleration

Process high-throughput data quickly with hardware acceleration. With four field-programmable gate arrays (FPGAs) onboard you have the most powerful DRAGEN analysis ever, enabling you to process NovaSeq X System data easily. Perform up to four simultaneous applications per flow cell in a single run.

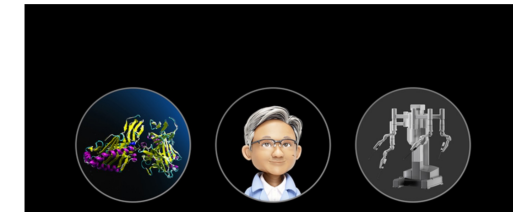
Reduce data footprint, manage and store data easily with lower costs and lower energy consumption, with built-in compression that reduces FASTQ file sizes by up to 80%.2

Stream data directly to Illumina Connected Analytics or BaseSpace Sequence Hub on the cloud for scalable data management, analysis, and aggregation.

### NVIDIA Partners With Industry Leaders to Advance Genomics, Drug Discovery and Healthcare

IQVIA, Illumina, Mayo Clinic and Arc Institute Harness NVIDIA AI and Accelerated Computing to Transform \$10 Trillion Healthcare and Life Sciences Industry

January 13, 2025

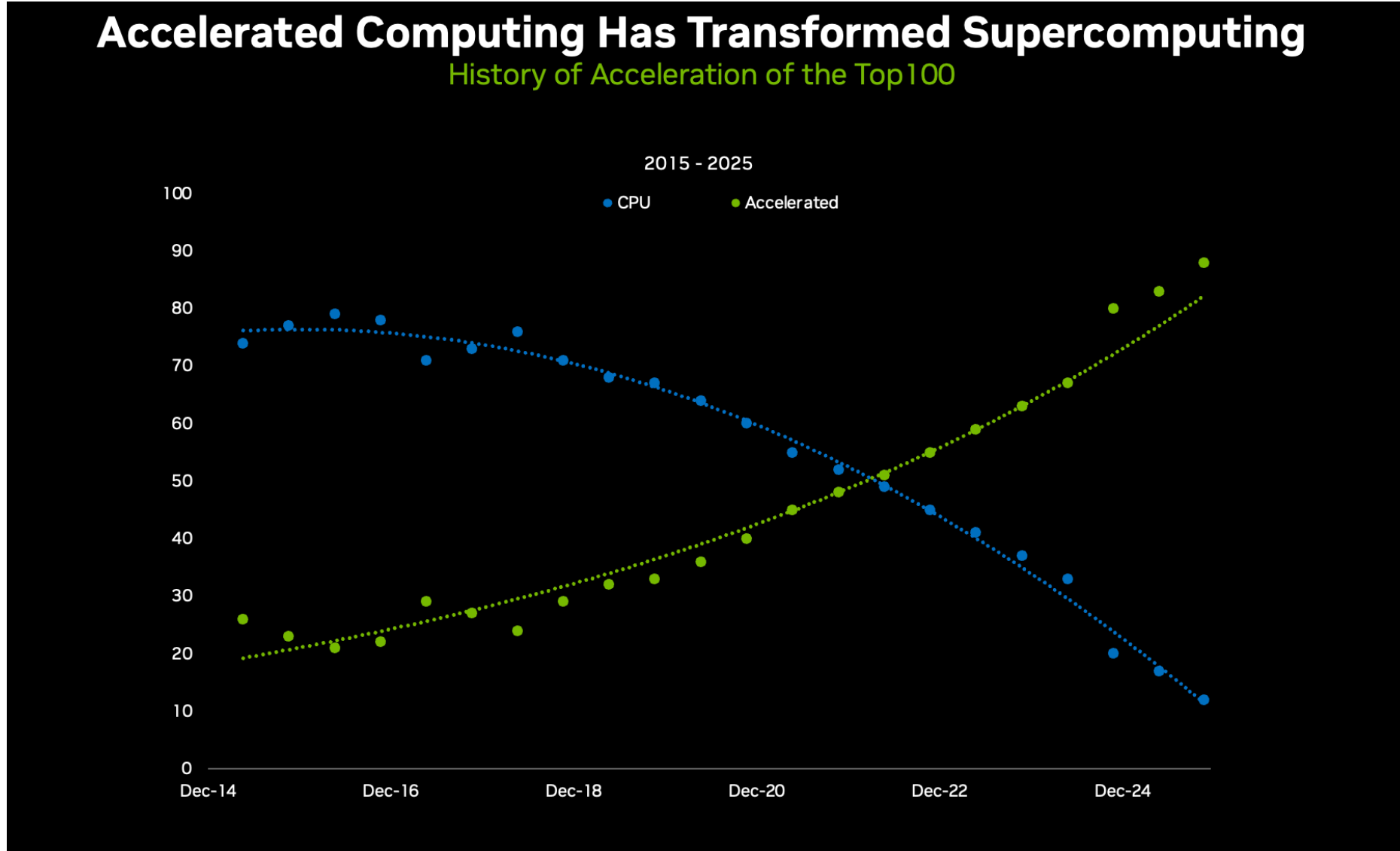


## Absci and AMD Announce Collaboration and Strategic Investment to Accelerate the Future of AI Drug Discovery

01/08/2025

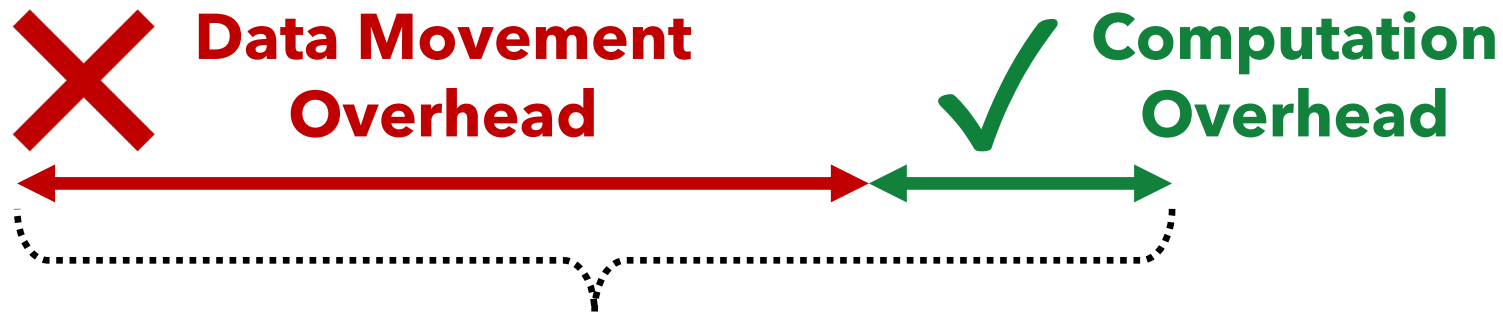
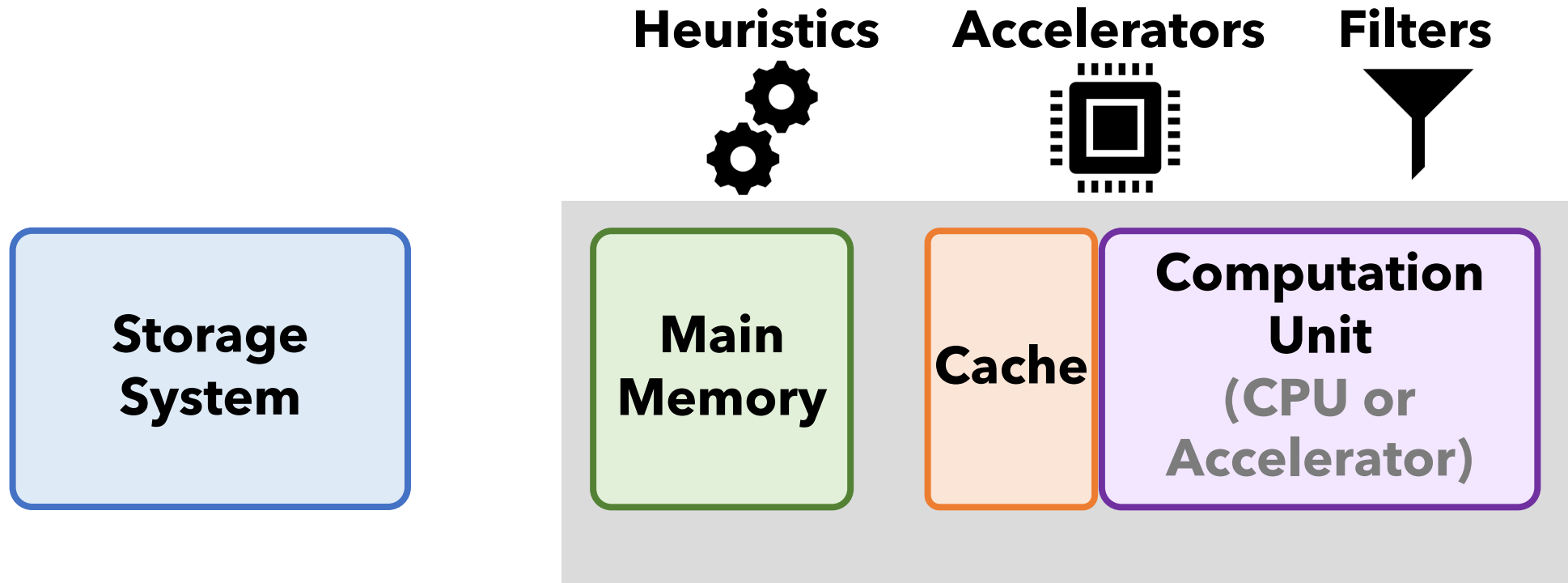
PDF Version

# Accelerated Computing is Mainstream Now



<https://blogs.nvidia.com/blog/accelerated-scientific-systems/?linkId=100000392637737>

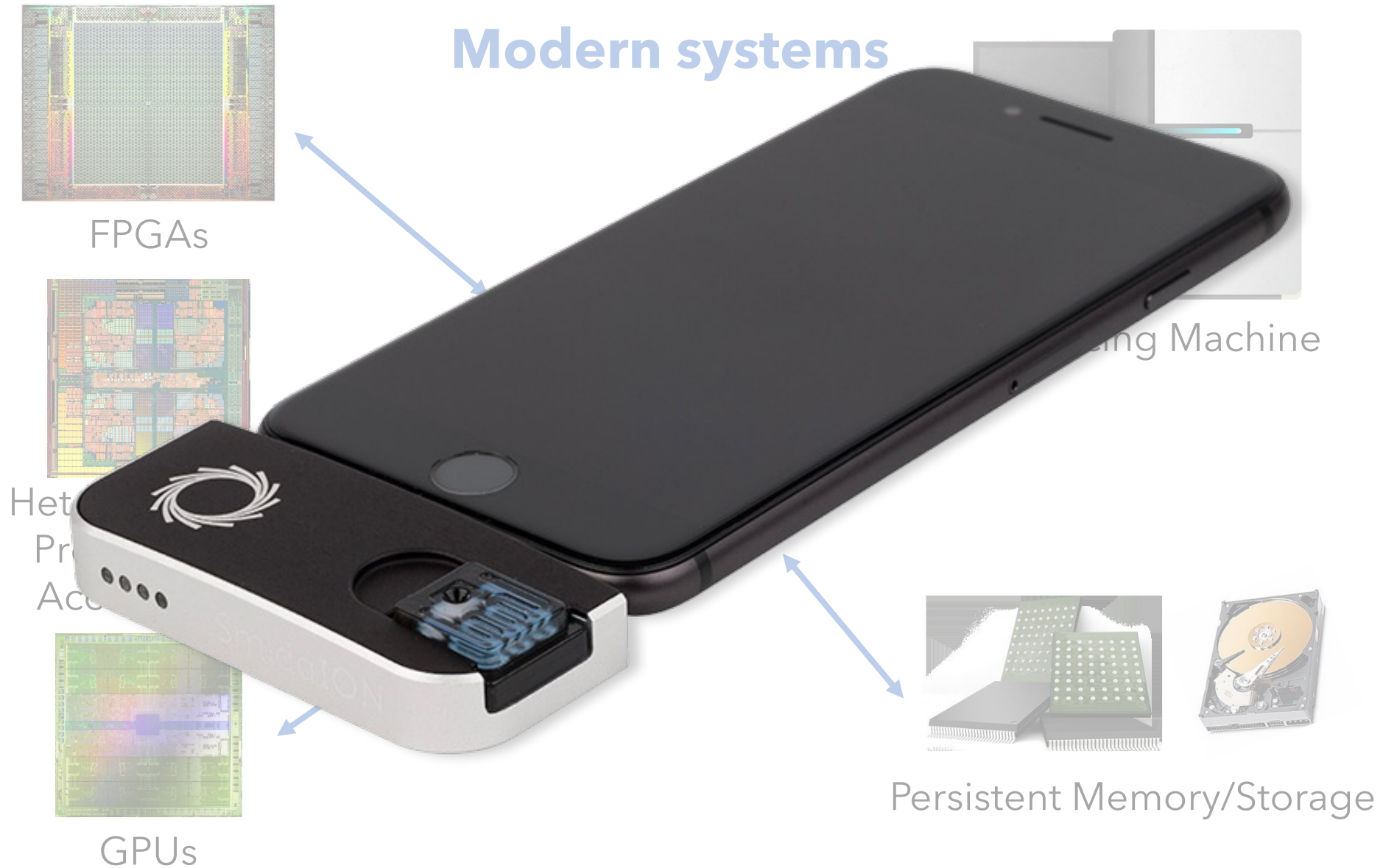
# Data Movement Dominates Performance



**Overall Time/Energy Spent**

**We need to co-design  
algorithms and architecture  
to handle data well**

# Pushing Towards New Architectures



# Recommended: Accelerating Genome Analysis [DAC '23]

- Onur Mutlu and [Can Firtina](#),

**"Invited: Accelerating Genome Analysis via Algorithm-Architecture Co-Design"**

Proceedings of the [60th Design Automation Conference \(DAC 2023\)](#), San Francisco, CA, USA, July 2023.

Paper [\[online\]](#) [\[pdf\]](#)

## Accelerating Genome Analysis via Algorithm-Architecture Co-Design

Onur Mutlu    Can Firtina

# Analyzing Genomes Reveals Key Insights



Personalized Medicine



Genome Editing

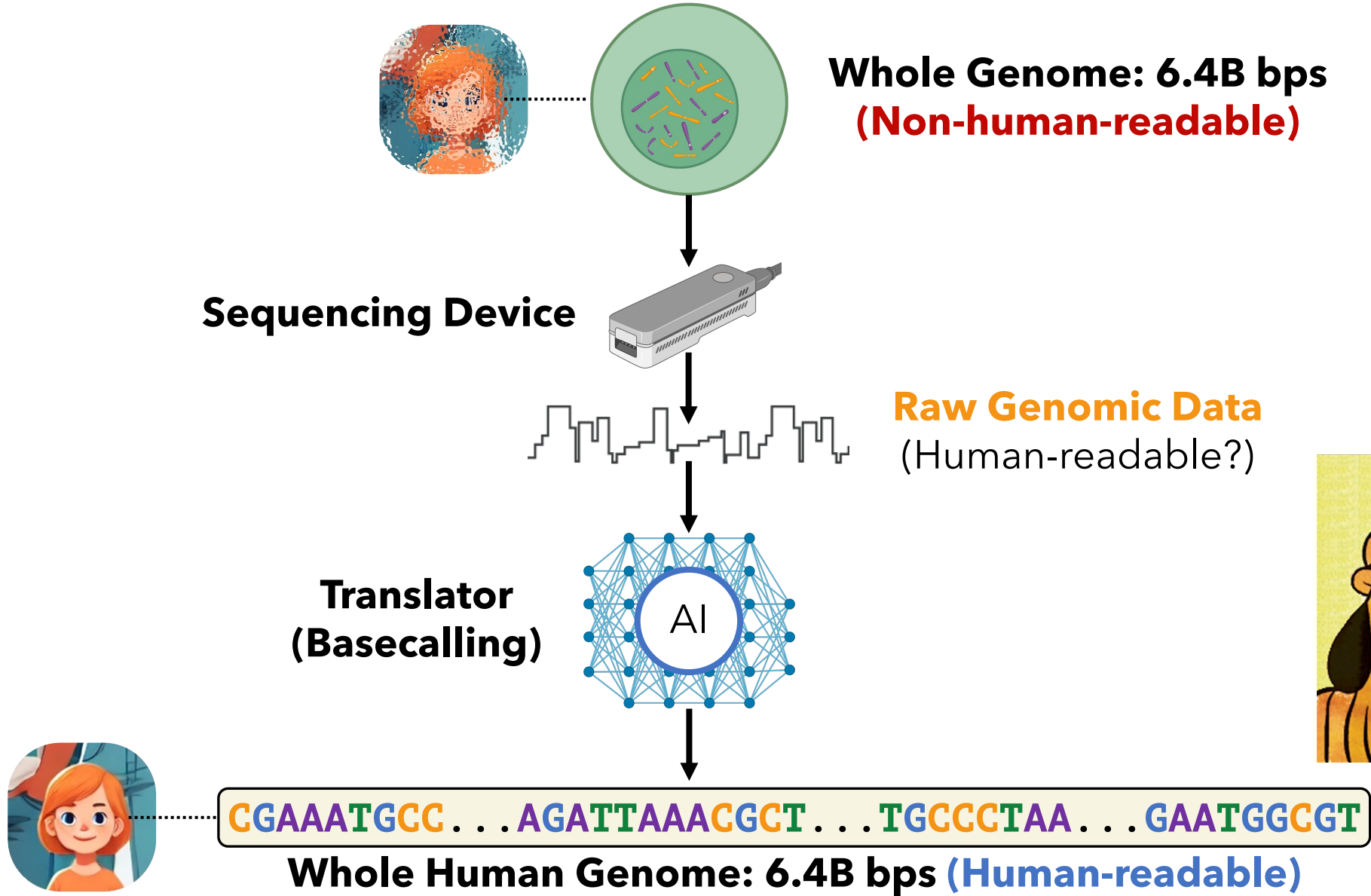


Detecting **pathogens**  
in the environment

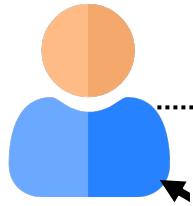


Rapid surveillance of  
**disease outbreaks**

# Generating the Entire Genomic Sequence



# Genomic Data: Giant Jigsaw Puzzle to Solve







**Human (Reference) Genome:** 3.2 Billion bps (haploid)

CGAAATGCC...AGATTAAACGGCT...TGCCCTAA...GAATGGCGT



CGAAATGCC...AGATTAAATGGT...TGCCCTAA...GAATGGCGT

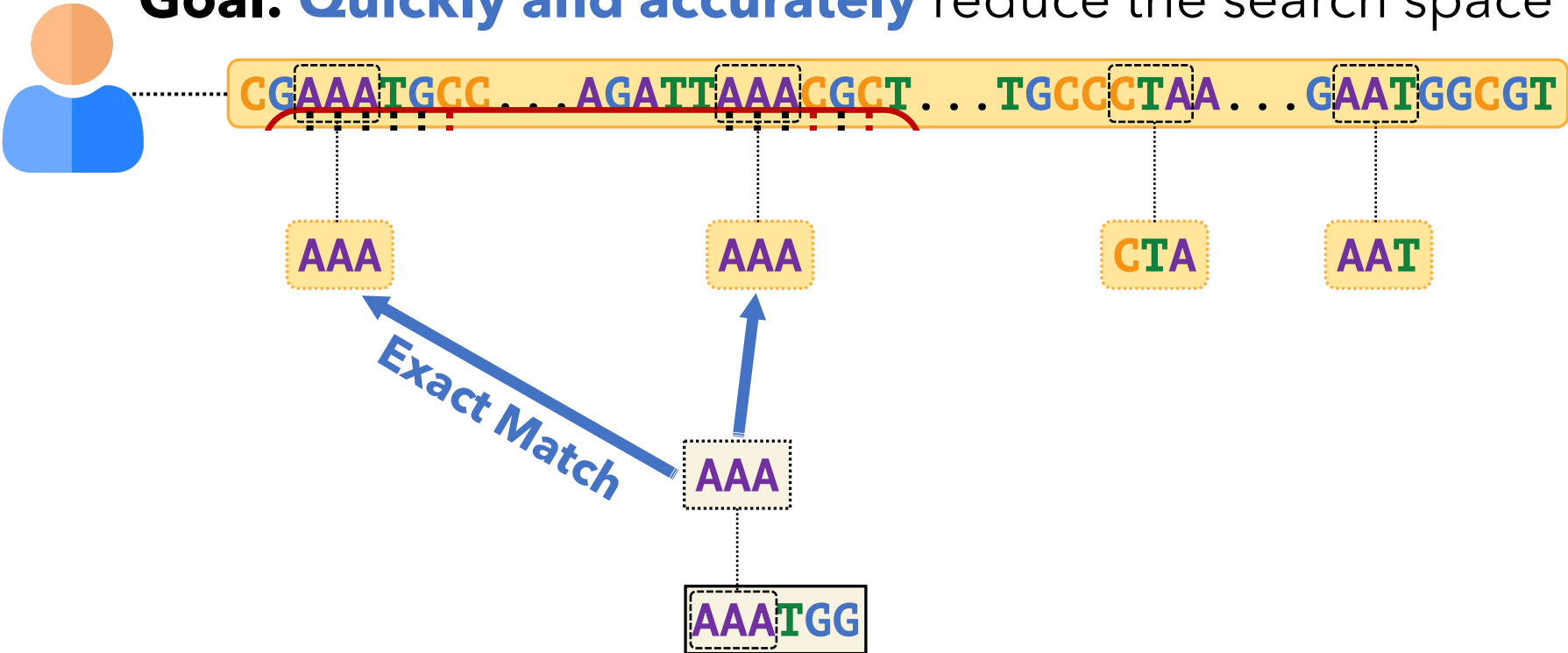
**Entire Genomic Sequence of an Organism**

-  **Small pieces (reads)**
-  **Unknown origins**
-  **Non-identical reference genome**
-  **Large volume of data to analyze**

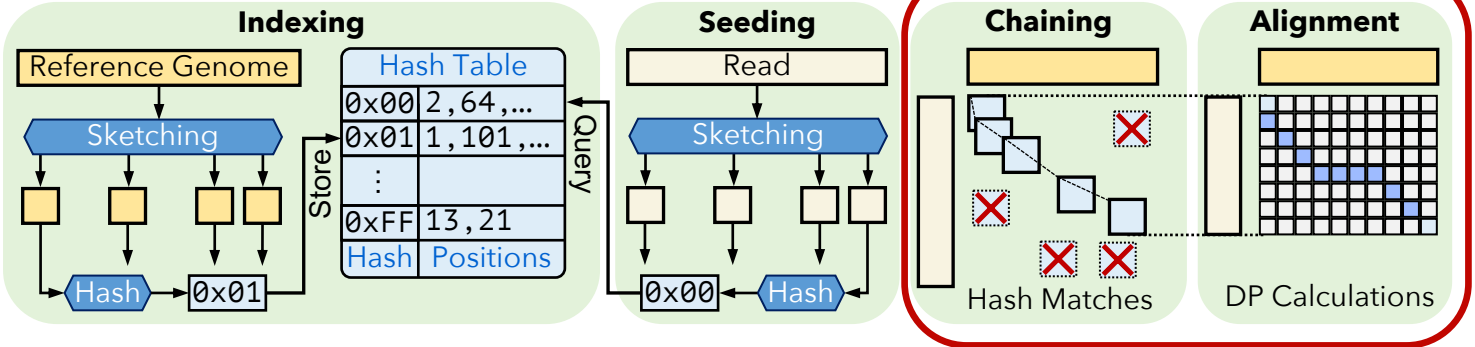


# Solving the Puzzle: The Practical Way

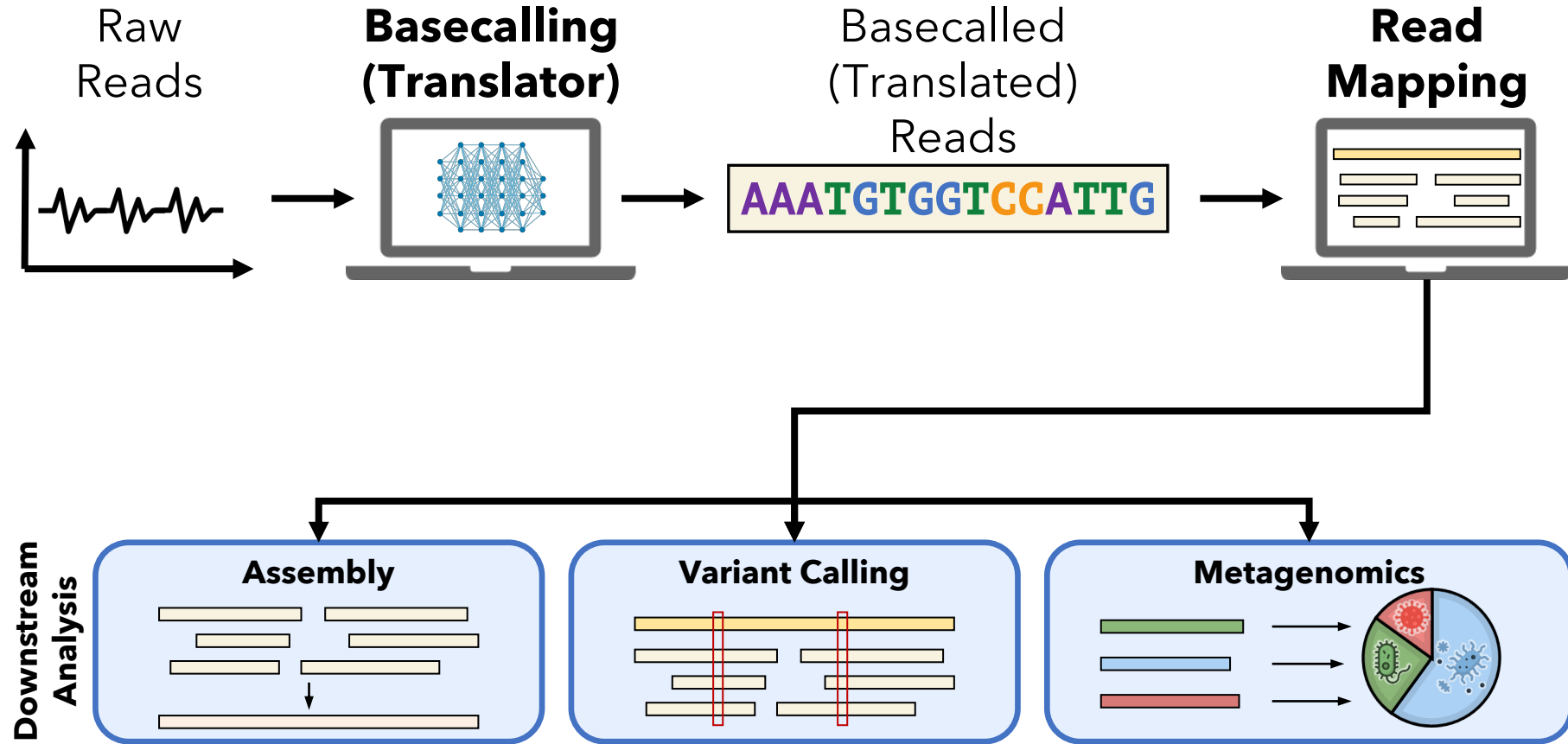
**Goal: Quickly and accurately** reduce the search space



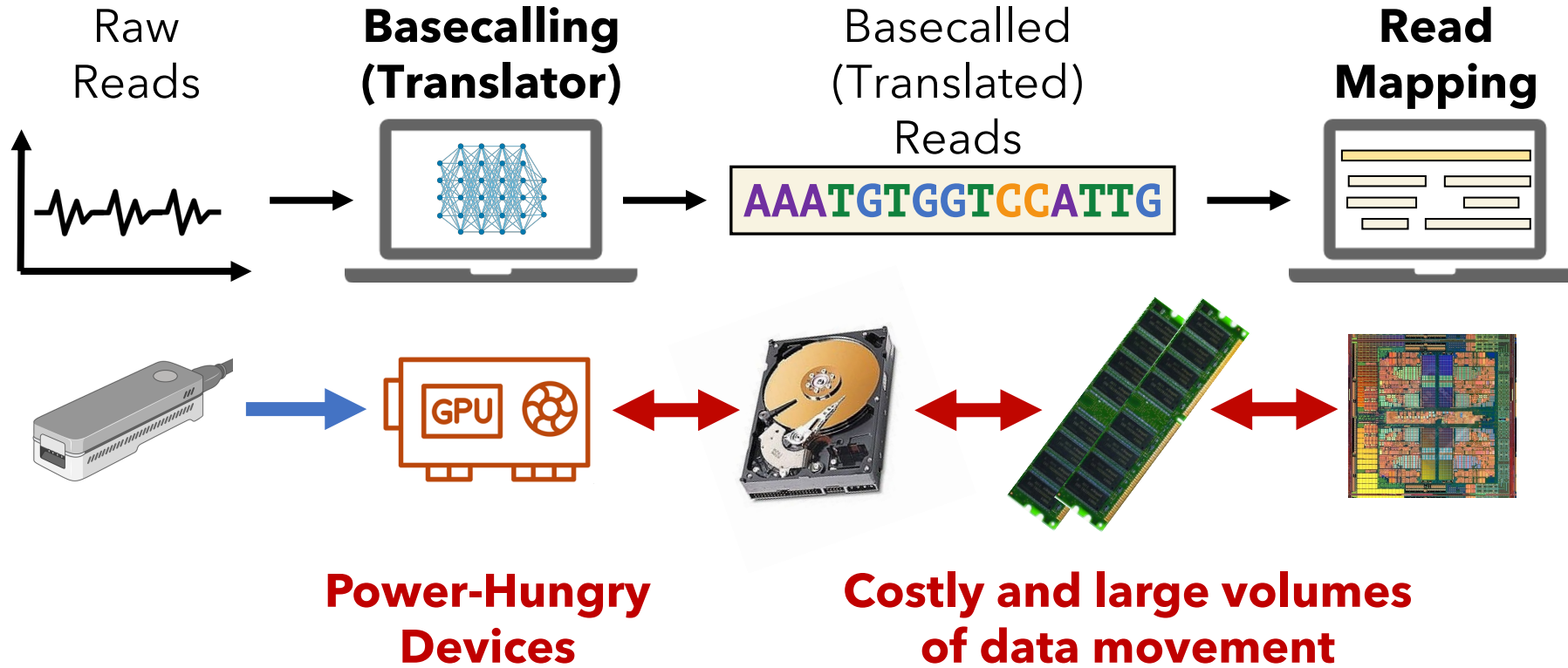
## Read Mapping



# A Common Genome Analysis Pipeline



# Genome Analysis is Energy Inefficient



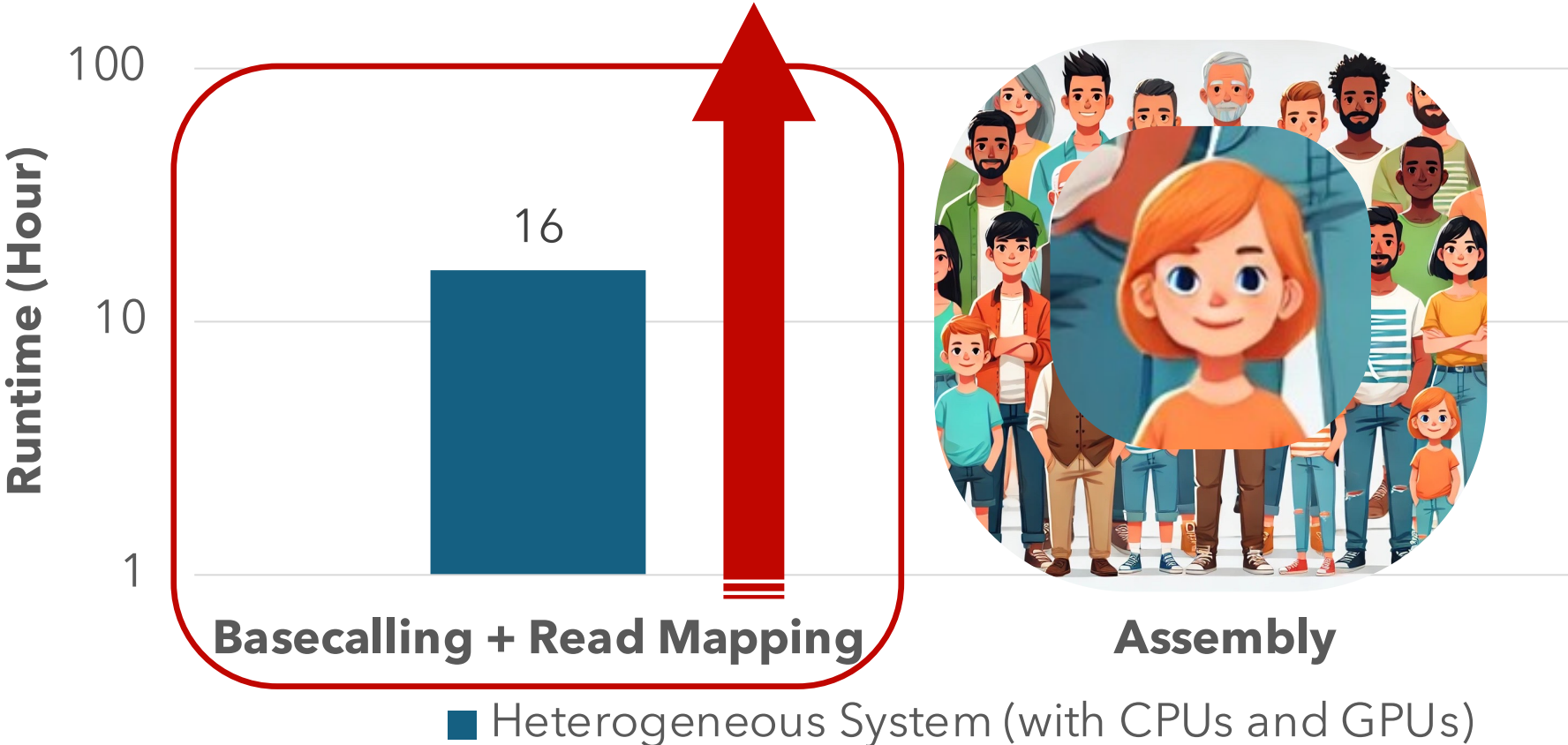
Single **memory** request consumes **>160× - 800× more energy** compared to performing an **addition operation**


- Boroumand et al., "Google Workloads for Consumer Devices: Mitigating Data Movement Bottlenecks," ASPLOS 2018

- Kestor et al., "Quantifying the Energy Cost of Data Movement in Scientific Applications," IISWC 2013

- Pandiyan and Wu, "Quantifying the energy cost of data movement for emerging smart phone workloads on mobile platforms," IISWC 2014

# Genome Analysis is Computationally Costly



 **Significant computational overhead**

# Genome Analysis Pipeline is Unbalanced

## Short Reads (Illumina)

<b>1 Sequencing</b> Library preparation: 6.5 hours Sequencing: 68.2 Gb/hour	<b>2 Basecalling</b> 104.4 Gb/hour	<b>3 Quality Control</b> 1339.2 Gb/hour	<b>4 Read Mapping</b> 0.2 Gb/hour	<b>5 Variant Calling</b> 1.2 Gb/hour
---	---------------------------------------	--	--------------------------------------	---

## Ultra-long Reads (ONT)

<b>1 Sequencing</b> Library preparation: 24 hours Sequencing: 4.1 Gb/hour	<b>2 Basecalling</b> 0.833 Gb/hour	<b>3 Quality Control</b> 3420 Gb/hour	<b>4 Read Mapping</b> 1.7 Gb/hour	<b>5 Variant Calling</b> 0.044 Gb/hour
---	---------------------------------------	--	--------------------------------------	---

## Accurate Long Reads (PacBio)

<b>1 Sequencing</b> Library preparation: 24 hours Sequencing: 5.3 Gb/hour	<b>2 Basecalling</b> 8.3 Gb/hour	<b>3 Quality Control</b> 1081 Gb/hour	<b>4 Read Mapping</b> 1.4 Gb/hour	<b>5 Variant Calling</b> 1.1 Gb/hour
---	-------------------------------------	--	--------------------------------------	---

# Limited Application Scope and Accuracy



**Energy-efficient analysis for  
Resource-constrained devices**

**Urgent analysis  
within minutes**



**Large-scale analysis  
without sacrificing accuracy  
and performance**



# STORM Research Group

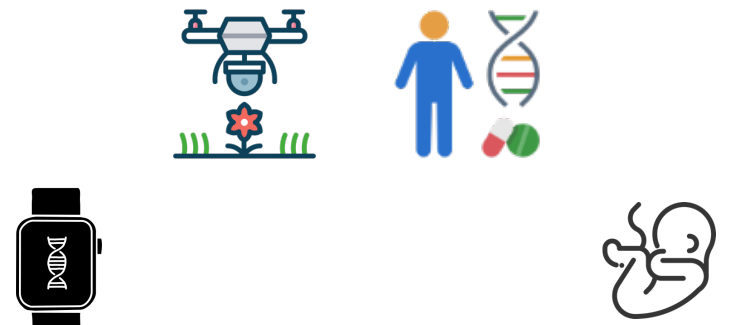
## Research Directions

 **Real-Time and Portable**  
Genomic Signal Analysis

 **Data-centric and AI-Driven** Solutions to Analyze Genomic Data

 Accurate and Fast Algorithms Scalable to **Large Volumes of Genomic Data**

### Today's Talk



### Goal

Enable fundamentally better integration of biological data analysis in our **everyday lives**



# New Frontiers: Raw Signal Analysis [ISMB '23]

- [Can Firtina](#), Nika Mansouri Ghiasi, Joel Lindegger, Gagandeep Singh, Meryem Banu Cavlak, Haiyu Mao, and Onur Mutlu,

## "RawHash: enabling fast and accurate real-time analysis of raw nanopore signals for large genomes"

Proceedings of the [31st Annual Conference on Intelligent Systems for Molecular Biology \(ISMB 2023\)](#) and the [22nd European Conference on Computational Biology \(ECCB 2023\)](#), Lyon, France, July 2023.

Paper [\[online\]](#) [\[pdf\]](#) [\[code\]](#)

Conference Talk [\[video\]](#) [\[pptx\]](#) [\[pdf\]](#)

Preprint (bioRxiv) [\[online\]](#) [\[pdf\]](#)

Social Media Thread [\[Twitter \(X\)\]](#)

*Bioinformatics*, 2023, **39**, i297–i307  
<https://doi.org/10.1093/bioinformatics/btad272>  
ISMB/ECCB 2023



---

## RawHash: enabling fast and accurate real-time analysis of raw nanopore signals for large genomes

Can Firtina <sup>1,\*</sup>, Nika Mansouri Ghiasi <sup>1</sup>, Joel Lindegger <sup>1</sup>, Gagandeep Singh <sup>1</sup>,  
Meryem Banu Cavlak <sup>1</sup>, Haiyu Mao <sup>1</sup>, Onur Mutlu <sup>1,\*</sup>

<sup>1</sup>Department of Information Technology and Electrical Engineering, ETH Zurich, 8092 Zurich, Switzerland

\*Corresponding author. Department of Information Technology and Electrical Engineering, ETH Zurich, Gloriastrasse 35, 8092 Zurich, Switzerland.  
E-mail: [firtinac@ethz.ch](mailto:firtinac@ethz.ch) (C.F.), [omutlu@ethz.ch](mailto:omutlu@ethz.ch) (O.M.)

# Real-Time Raw Signal Analysis [Bioinformatics '24]

- [Can Firtina](#), Melina Soysal, Joël Lindegger, and Onur Mutlu, "**RawHash2: Mapping Raw Nanopore Signals Using Hash-Based Seeding and Adaptive Quantization**" [Bioinformatics](#), July 2024.  
Paper [\[online\]](#) [\[pdf\]](#) [\[code\]](#)

*Bioinformatics*, 2024, **40(8)**, btae478  
<https://doi.org/10.1093/bioinformatics/btae478>  
Advance Access Publication Date: 30 July 2024  
**Applications Note**



---

Sequence analysis

## RawHash2: mapping raw nanopore signals using hash-based seeding and adaptive quantization

Can Firtina <sup>1,\*</sup>, Melina Soysal <sup>1</sup>, Joël Lindegger <sup>1</sup>, Onur Mutlu <sup>1,\*</sup>

<sup>1</sup>Department of Information Technology and Electrical Engineering, ETH Zurich, Zurich 8092, Switzerland

\*Corresponding authors. Department of Information Technology and Electrical Engineering, ETH Zurich, Zurich 8092, Switzerland.  
E-mail: [firtinac@ethz.ch](mailto:firtinac@ethz.ch) (C.F.); [omutlu@ethz.ch](mailto:omutlu@ethz.ch) (O.M.)

# New Directions: Assembling Raw Signals

- [Can Firtina](#), Maximilian Mordig, Harun Mustafa, Sayan Goswami, Nika Mansouri Ghiasi, Stefano Mercogliano, Furkan Eris, Joel Lindegger, Andre Kahles, and Onur Mutlu, **"Rawsamble: Overlapping and Assembling Raw Nanopore Signals using a Hash-based Seeding Mechanism"**

[Accepted to Bioinformatics] [arXiv](#), October 2024.

Paper [\[online\]](#) [\[pdf\]](#) [\[code\]](#)

[ISMB 2024 Talk](#) [\[video\]](#) [\[pptx\]](#) [\[pdf\]](#)

[CSHL 2024 \(Biological Data Science\) Talk](#) [\[video\]](#) [\[pptx\]](#) [\[pdf\]](#)

Social Media Thread [\[Twitter \(X\)\]](#) [\[LinkedIn\]](#)

## **Rawsamble: Overlapping and Assembling Raw Nanopore Signals using a Hash-based Seeding Mechanism**

Can Firtina<sup>1</sup> Maximilian Mordig<sup>1,2</sup> Harun Mustafa<sup>1,3,4</sup> Sayan Goswami<sup>1</sup> Nika Mansouri Ghiasi<sup>1</sup>

Stefano Mercogliano<sup>1</sup> Furkan Eris<sup>1</sup> Joël Lindegger<sup>1</sup> Andre Kahles<sup>1,3,4</sup> Onur Mutlu<sup>1</sup>

<sup>1</sup>*ETH Zurich*      <sup>2</sup>*Max Planck Institute for Intelligent Systems*

<sup>3</sup>*University Hospital Zurich*      <sup>4</sup>*Swiss Institute of Bioinformatics*

# Quickly Aligning Raw Signals [IEEE Access '24]

- Joël Lindegger, Can Firtina, Nika Mansouri Ghiasi, Mohammad Sadrosadati, Mohammed Alser, and Onur Mutlu,

## **"RawAlign: Accurate, Fast, and Scalable Raw Nanopore Signal Mapping via Combining Seeding and Alignment"**

[IEEE Access](#), December 2024.

Paper [\[online\]](#) [\[pdf\]](#) [\[code\]](#)

## **RawAlign: Accurate, Fast, and Scalable Raw Nanopore Signal Mapping via Combining Seeding and Alignment**

**JOËL LINDEGGER<sup>ID</sup>, CAN FIRTINA<sup>ID</sup>, NIKA MANSOURI GHIASI, MOHAMMAD SADROSADATI<sup>ID</sup>,  
MOHAMMED ALSER<sup>ID</sup>, AND ONUR MUTLU<sup>ID</sup>, (Fellow, IEEE)**

Department of Information Technology and Electrical Engineering, ETH Zürich, 8092 Zürich, Switzerland

Corresponding authors: Joël Lindegger (jmlindegger@gmail.com), Can Firtina (firtinac@ethz.ch), and Onur Mutlu (omutlu@ethz.ch)

This work was supported in part by the Semiconductor Research Corporation (SRC), in part by the ETH Future Computing Laboratory, in part by European Union's Horizon Programme for Research and Innovation under Grant 101047160-BioPIM, and in part by the Swiss National Science Foundation (SNSF) under Grant 200021\_213084.

# Benchmarking Raw Signal Analysis [ACM-BCB '25]

- Furkan Eris, Ulysse McConnell, Can Firtina, and Onur Mutlu,  
**"RawBench: A Comprehensive Benchmarking Framework  
for Raw Nanopore Signal Analysis Techniques"**

Proceedings of the [16th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics \(ACM-BCB 2025\)](#), Philadelphia, PA, USA, October 2025.

Paper [\[online\]](#) [\[pdf\]](#)

Conference Talk [\[pptx\]](#) [\[pdf\]](#)

## RawBench: A Comprehensive Benchmarking Framework for Raw Nanopore Signal Analysis Techniques

Furkan Eris  
ETH Zurich  
Zurich, Switzerland  
furkan6olm@gmail.com

Ulysse McConnell  
ETH Zurich  
Zurich, Switzerland  
umcconnell@student.ethz.ch

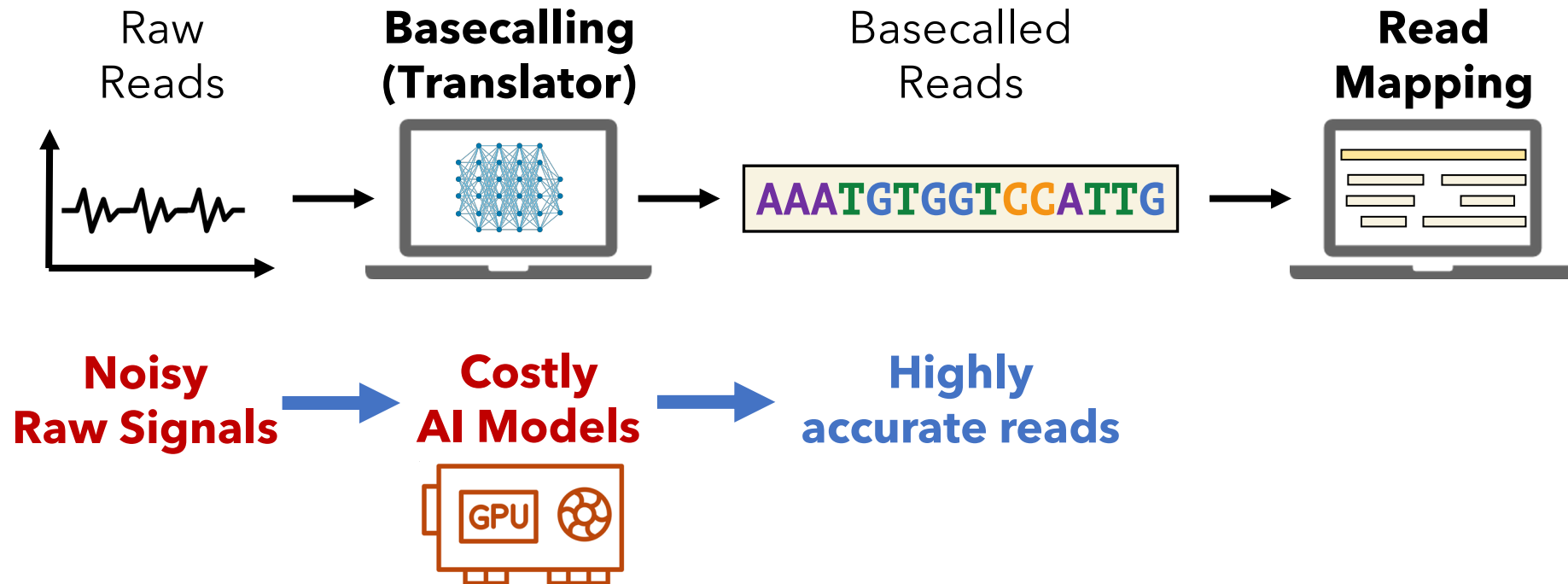
Can Firtina\*  
University of Maryland, College Park  
MD, USA  
firtina@umd.edu

Onur Mutlu\*  
ETH Zurich  
Zurich, Switzerland  
omutlu@gmail.com

# Basecalling is Accurate yet Costly

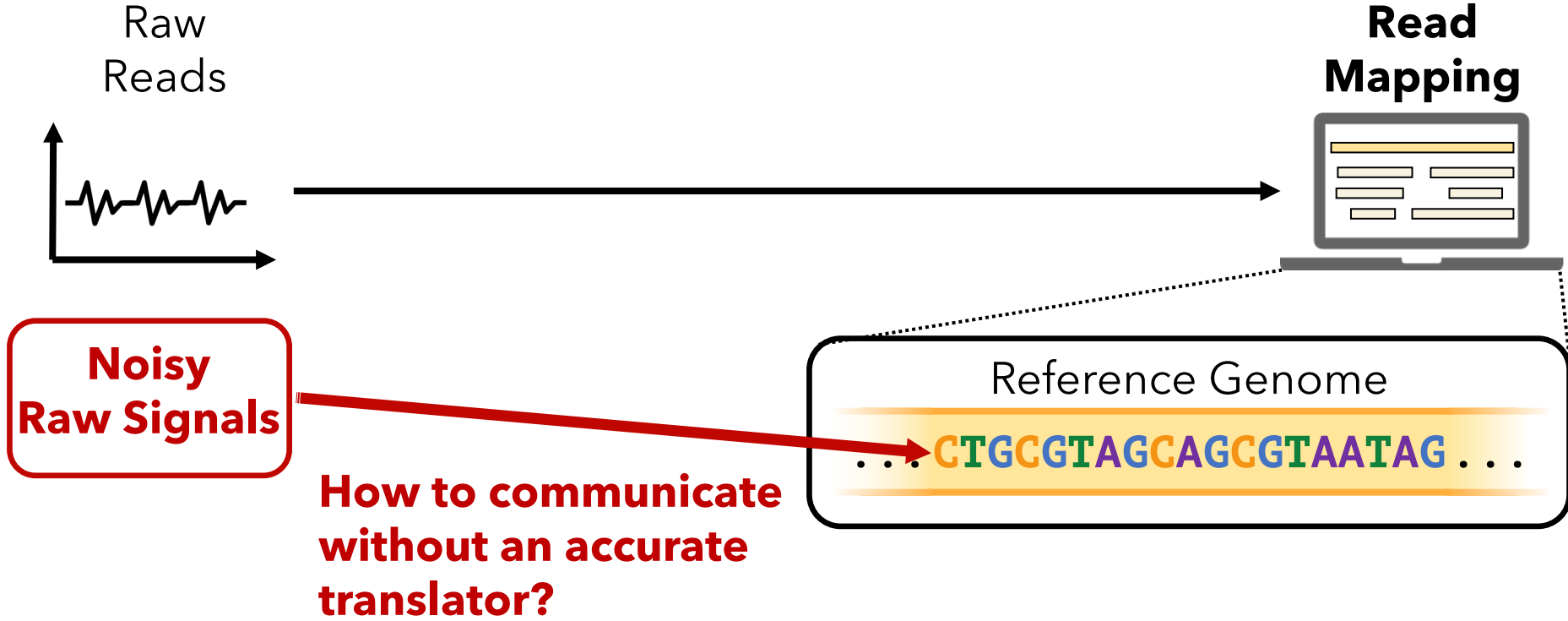


Significant computation and energy overhead

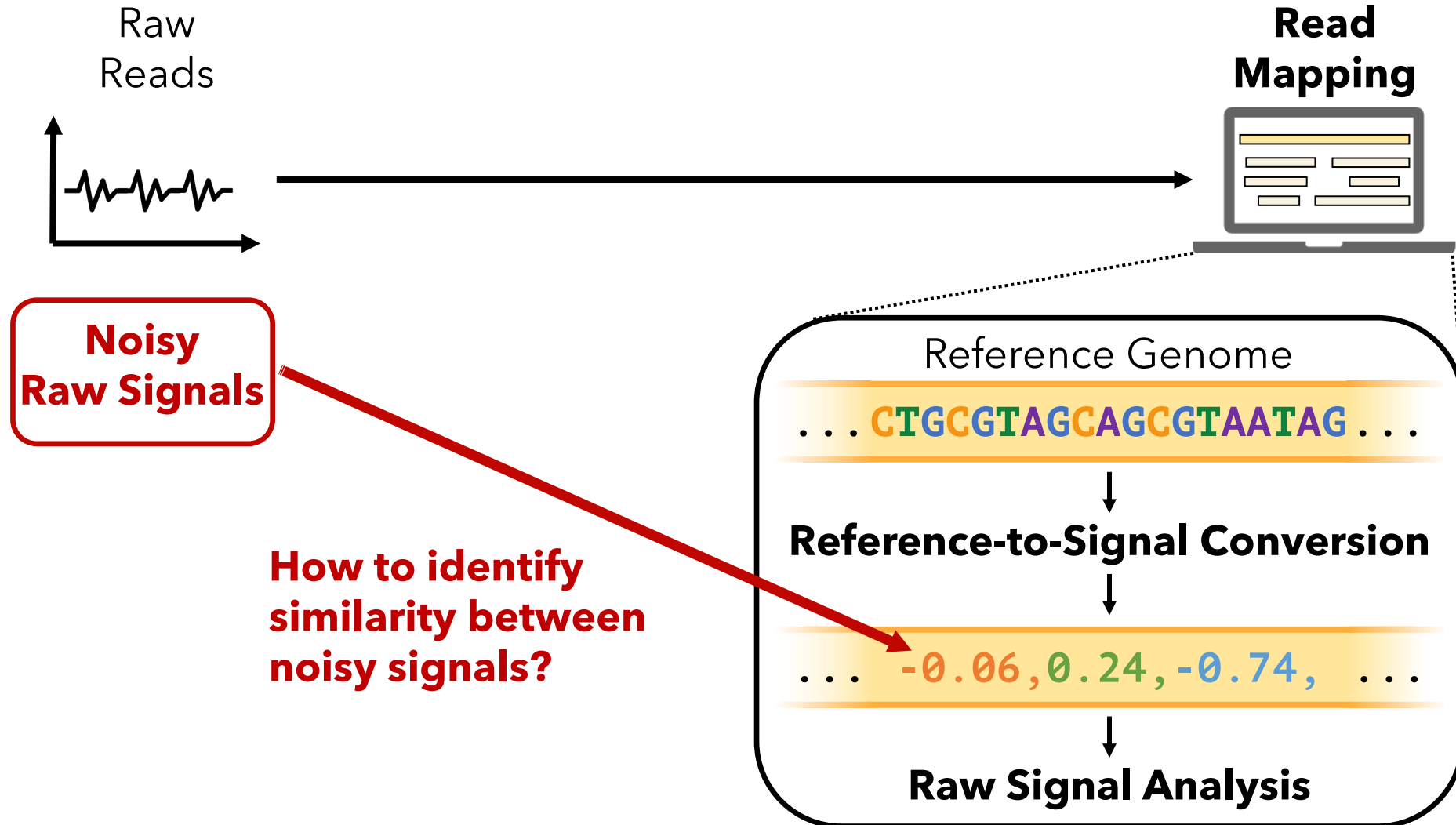


How to make this pipeline cheaper?

# Can We Survive Without Translation?



# Can We Survive With Noisy Signal Analysis?



# Noise in Raw Electrical Signals

**Observation:** Identical molecules generate similar raw signals

Sequencing **CTGCGTAGCA**  
Raw Signals: [-0.06, -0.22, -0.75, 0.96, -0.49]

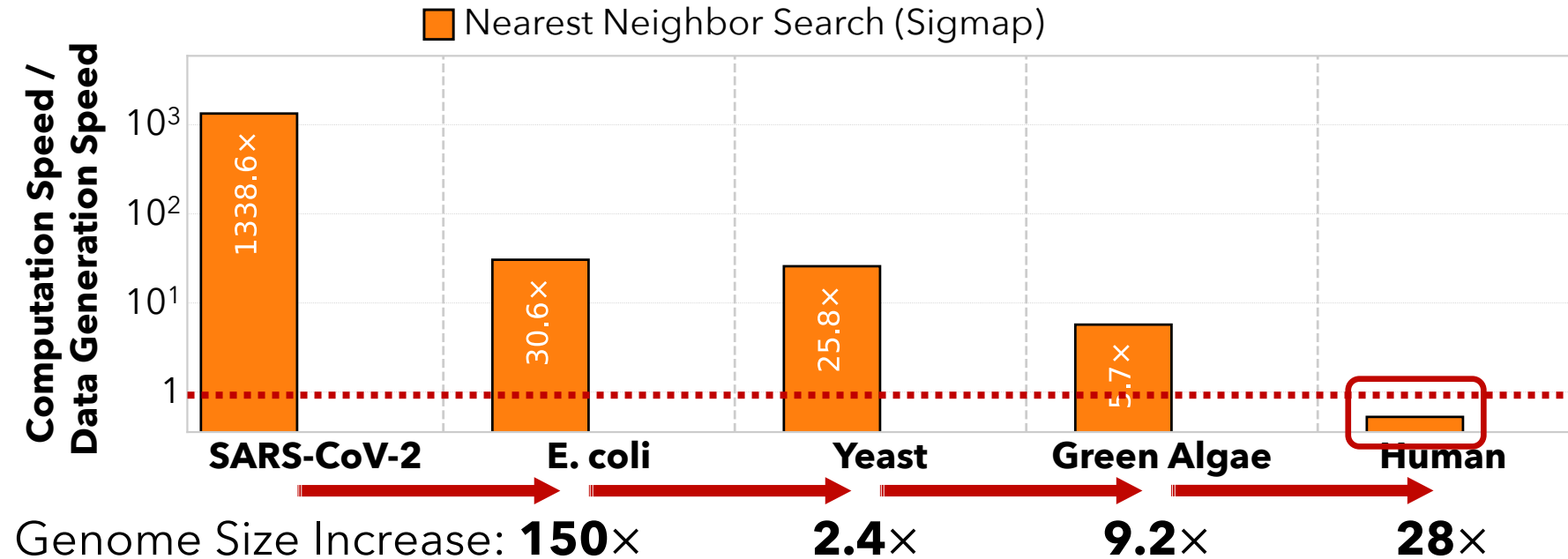
**Key Idea:** Perform similarity estimation based on **vector similarity search**  
**Nearest Neighbor Search**

Raw Signals: [-0.05, -0.21, -0.76, 1.01, -0.51]  
Sequencing **CTGCGTAGCA**



**Problem: Too Costly for Large Genomes**

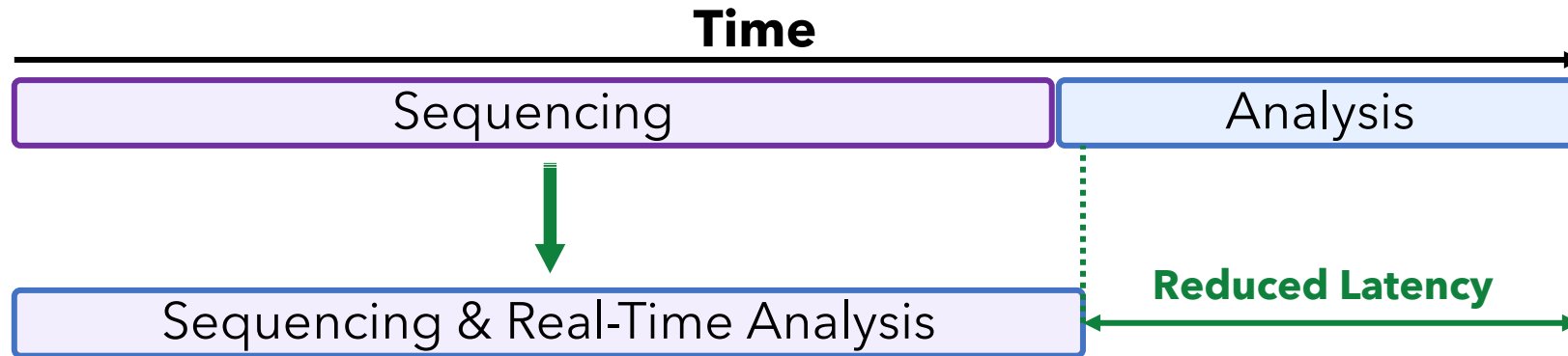
# Scalability Issues Limit Real-Time Analysis



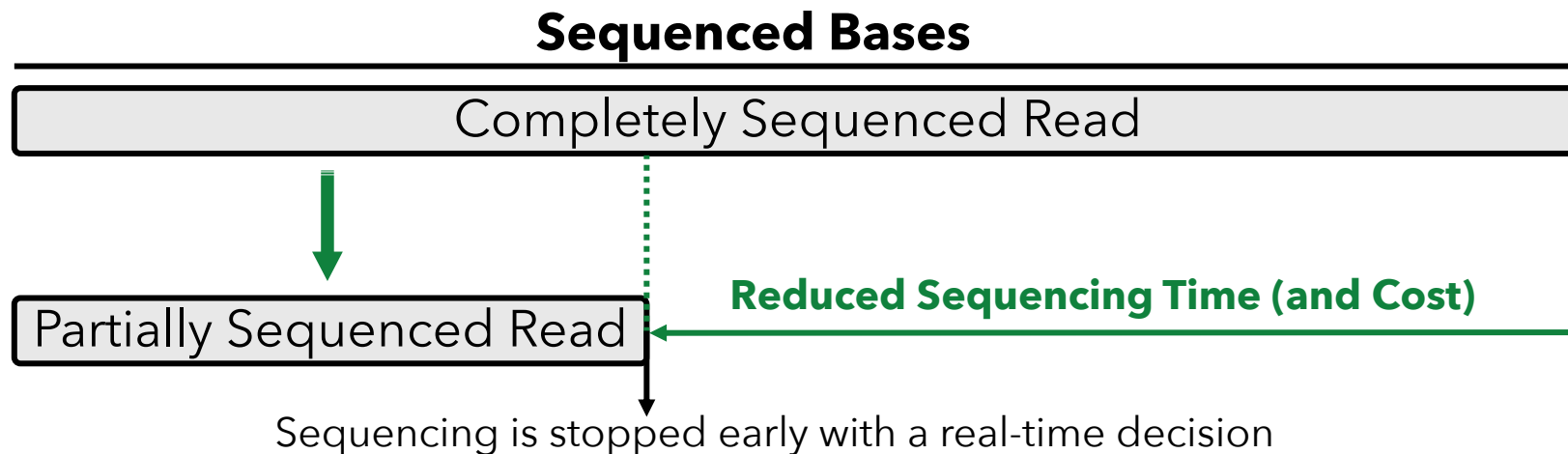
**No Real-Time Analysis for Larger Genomes:**  
Searching mechanism does not scale well

# Benefits of Real-Time Analysis

- ✓ **Reducing latency** by overlapping analysis with sequencing



- ✓ **Reducing sequencing time and cost** by stopping sequencing early

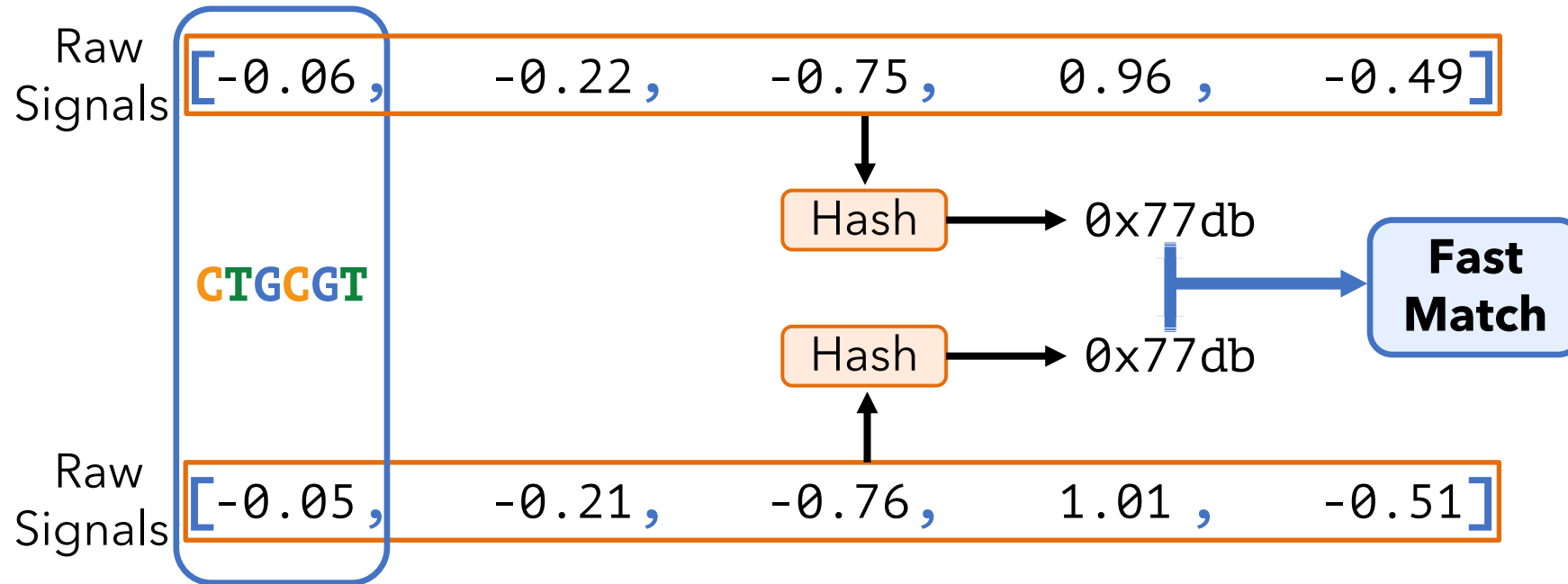


# RawHash's Goal

Enable **Fast, Accurate, and Scalable Real-Time Analysis** for Larger Genomes

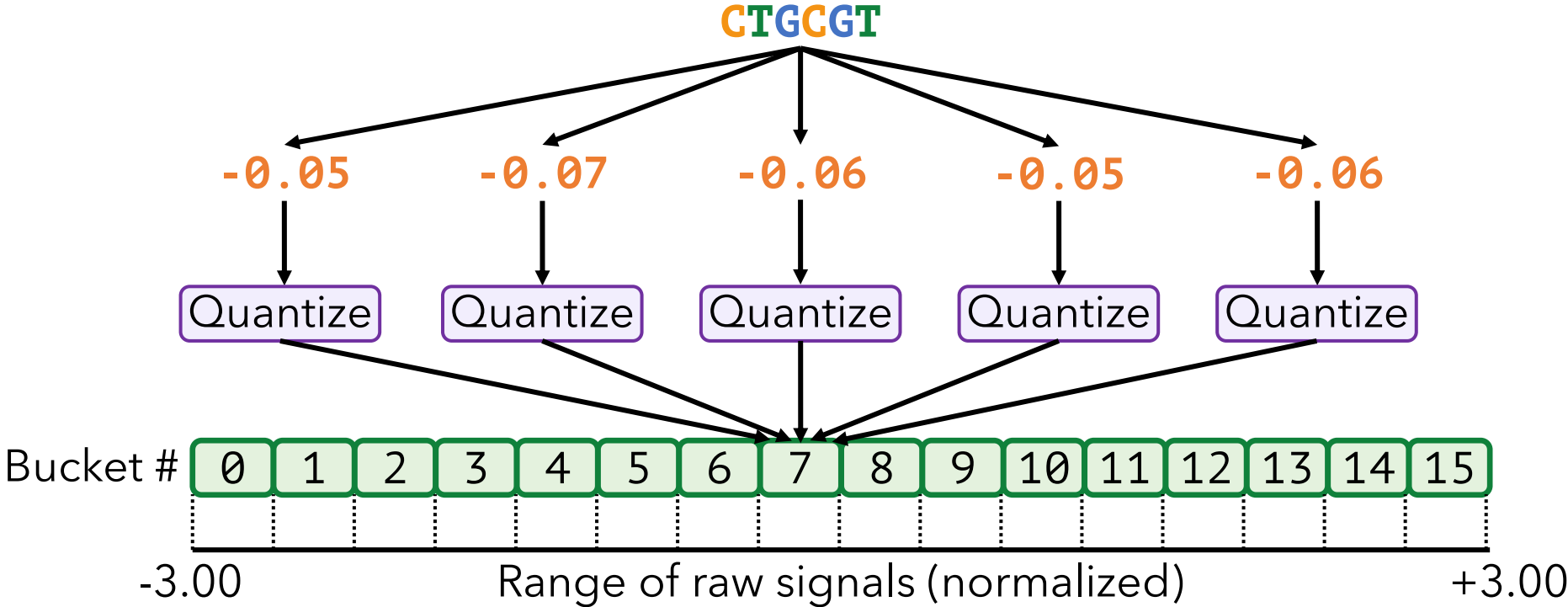
# RawHash's Key Idea

**Key Idea:** Perform similarity estimation based on **quick exact matches**



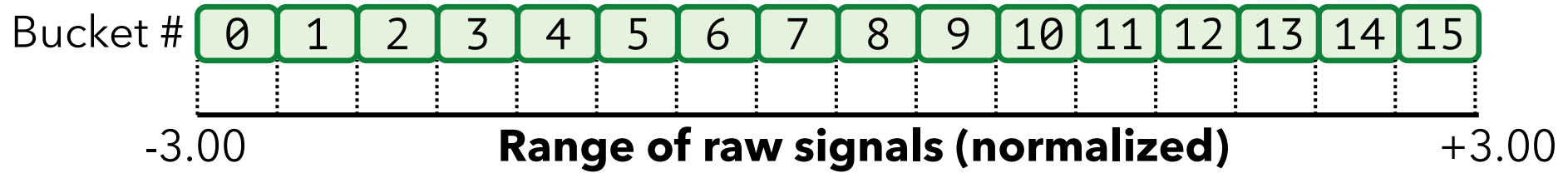
**Challenge:** How to generate **the same hash value** for **similar enough signals with various noise types?**

# RawHash's Key Idea – Quantization

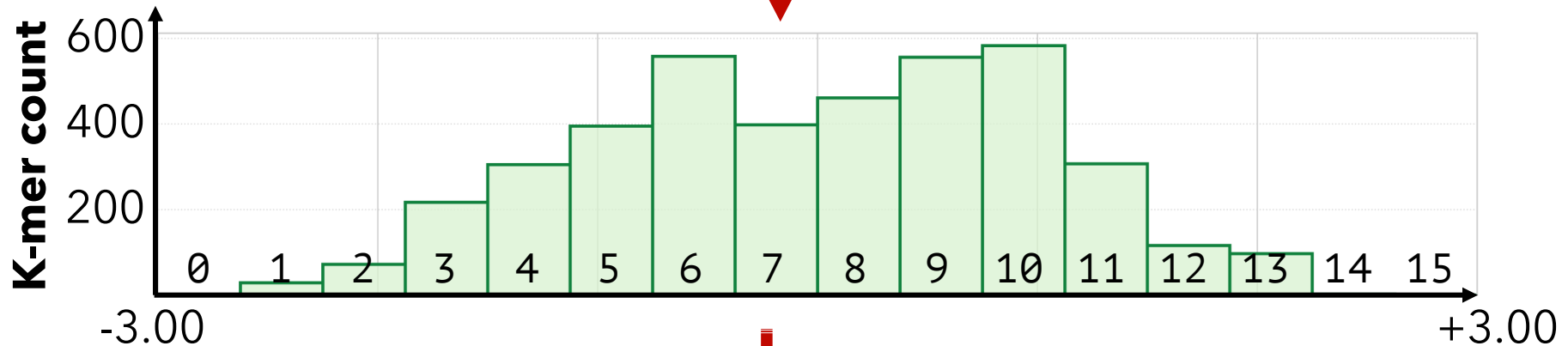


✔ **Enables matching raw signals** by eliminating slight differences

# How to Better Quantize Signals



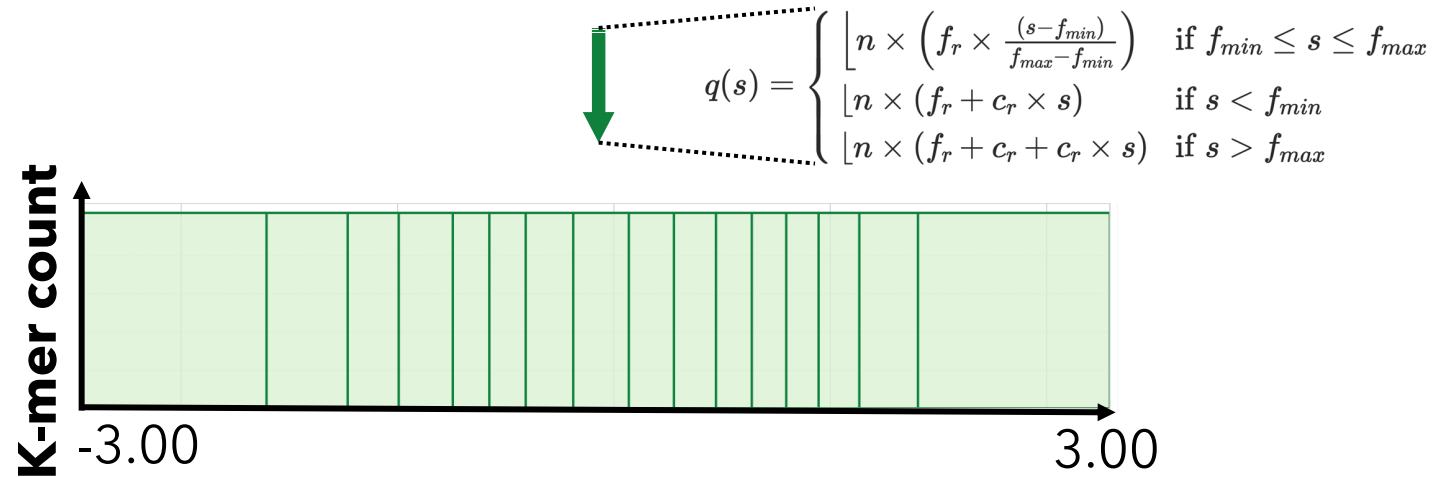
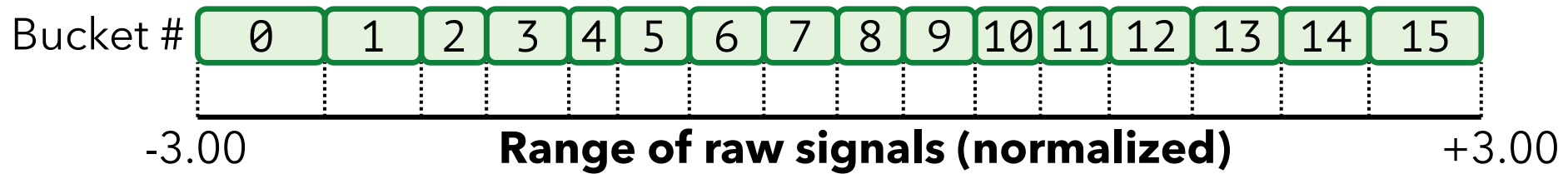
✗ Equal-width buckets leads to **unbalanced bucket loading**



✗ Reduced uniqueness and poor resource utilization

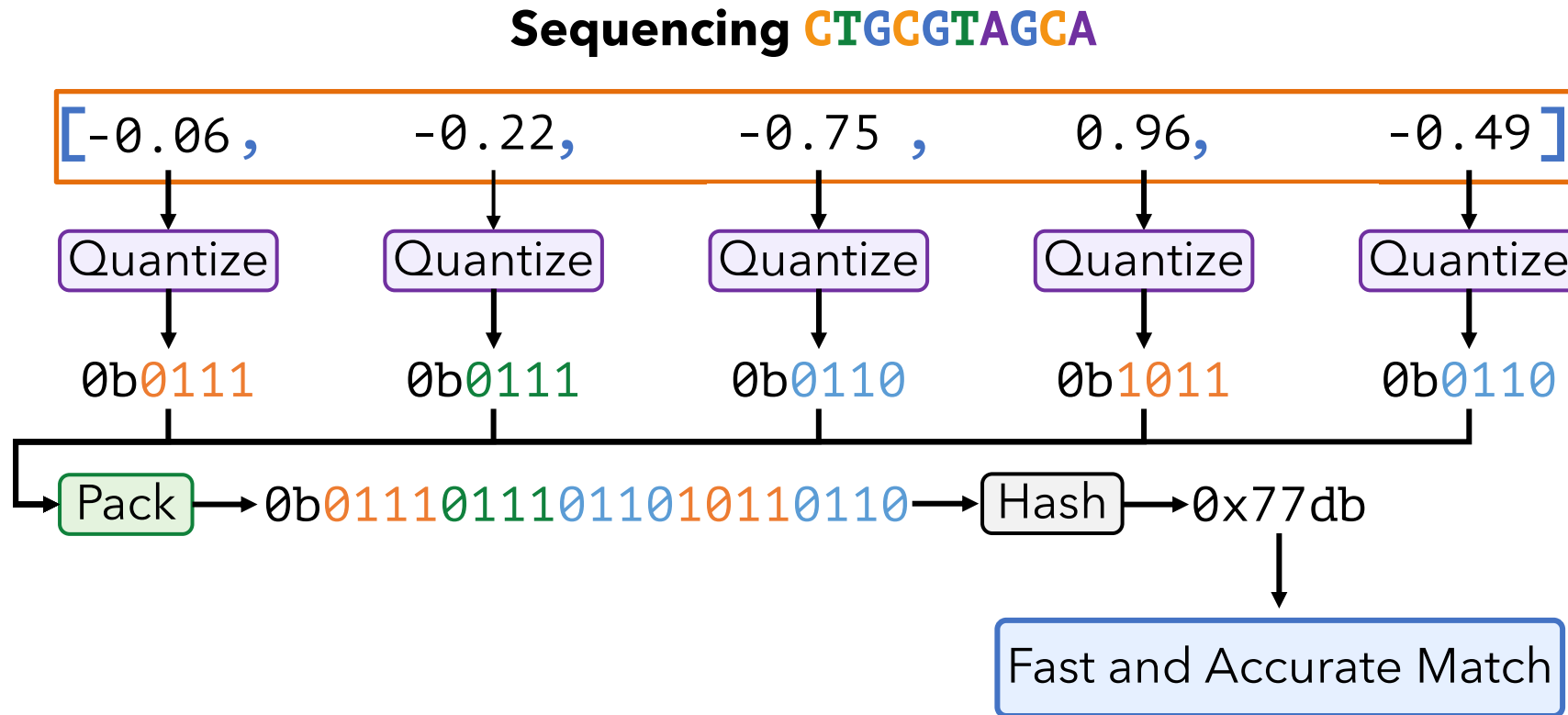
# Adaptive Quantization

- **Key Idea:** Quantizing raw signals with **non-equal bucket widths** by leveraging raw signal distribution



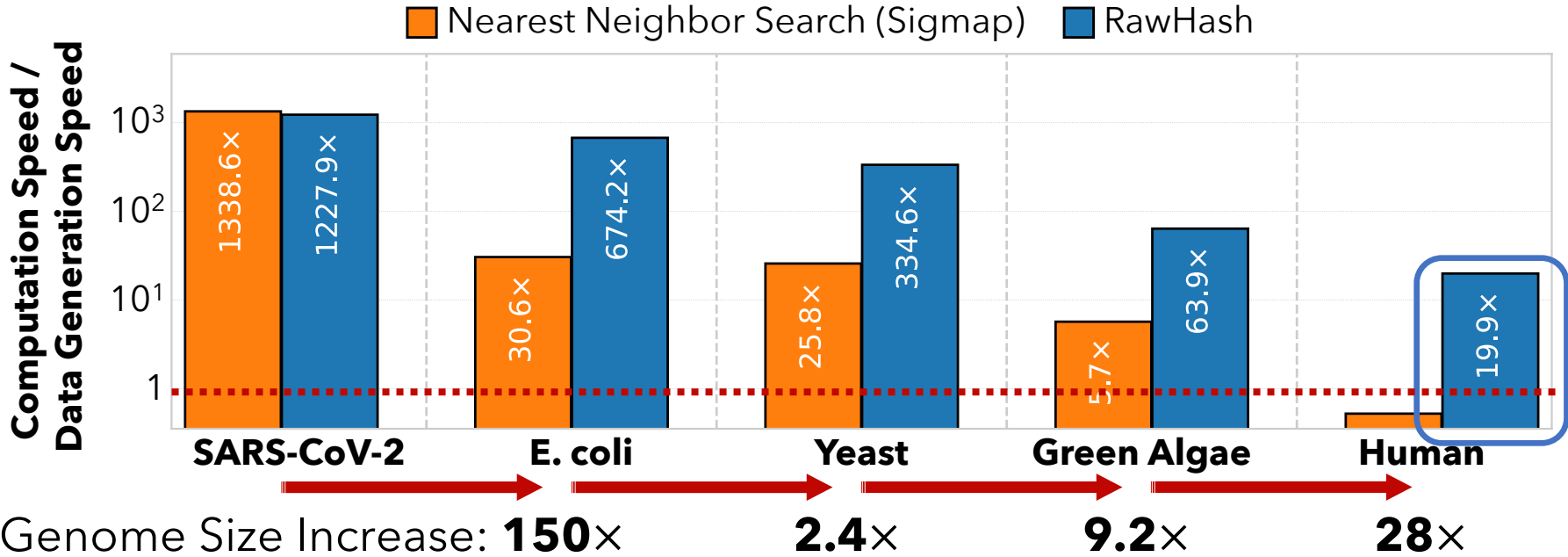
**Better** bucket utilization

# RawHash Key Idea – Hash-based Matching



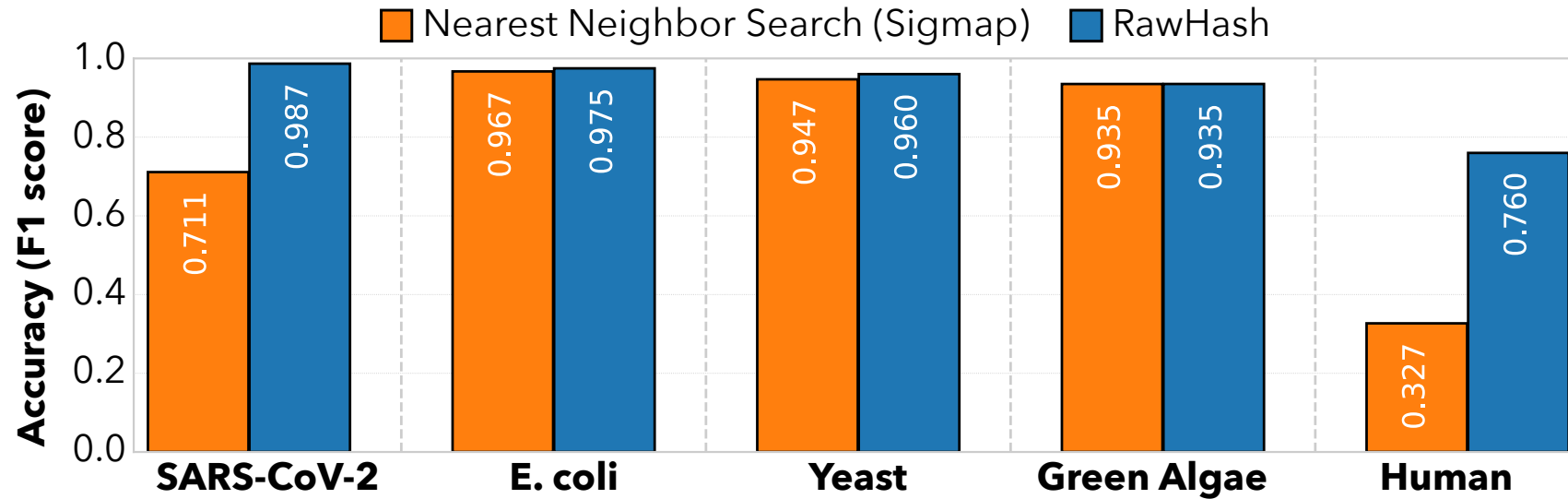
**Key Contribution:** **The first** hash-based search (and indexing) mechanism for raw signals

# Real-Time Analysis Benefits from Fast Search



**Real-time Analysis for Human Genomes:**  
Fast search instead of costly neighbor search

# Fast Search Leads to Accurate Search



## Best accuracy for all genomes:

Hash values span longer sequences than vectors ->  
More unique and informative matches

## (Almost) Identical accuracy to basecalled analysis:

Is the small difference because RawHash is even better?

# RawHash is Open Source

SAFARI RawHash Public

Edit Pins Unwatch 7 Fork 10 Starred 64


main 9 Branches 3 Tags

Go to file Add file Code

canfirtina Rawsamble evaluation scripts cc9f2b9 · 29 days ago 49 Commits

extern	new R10 parameters	6 months ago
gitfigures	rawsamble info in README	2 years ago
src	new R10 parameters	6 months ago
test	Rawsamble evaluation scripts	29 days ago
.gitignore	Rawsamble evaluation scripts	29 days ago
.gitmodules	ZSTD submodule for POD5	3 years ago
LICENSE	R10 k-mer models can be parsed now as well.	3 years ago
Makefile	A minor fix in Makefile	3 years ago
README.md	new R10 parameters	6 months ago
code_of_conduct.md	Moving to multiple headers than a single one to improve ...	3 years ago

README Code of conduct GPL-3.0 license



# RawHash

<https://github.com/CMU-SAFARI/RawHash>

About

RawHash can accurately and efficiently map raw nanopore signals to reference genomes of varying sizes (e.g., from viral to a human genomes) in real-time without basecalling. Described by Firtina et al. (published at [https://academic.oup.com/bioinformatics/article/39/Supplement\\_1/i297/7210440](https://academic.oup.com/bioinformatics/article/39/Supplement_1/i297/7210440)).

[academic.oup.com/bioinformatics/art...](https://academic.oup.com/bioinformatics/art...)

- bioinformatics
- nanopore
- seeding
- segmentation
- event-detection
- genome-analysis
- hash-tables
- contamination
- read-mapping
- relative-abundances
- nanopore-sequencing
- nanopore-analysis-pipeline
- nanopore-reads
- nanopore-data
- nanopore-minion
- raw-signal
- rawhash
- raw-nanopore-signal-analysis

Readme GPL-3.0 license Code of conduct Activity Custom properties 64 stars 7 watching

So What?

# New Directions: Assembling Raw Signals

- [Can Firtina](#), Maximilian Mordig, Harun Mustafa, Sayan Goswami, Nika Mansouri Ghiasi, Stefano Mercogliano, Furkan Eris, Joel Lindegger, Andre Kahles, and Onur Mutlu, **"Rawsamble: Overlapping and Assembling Raw Nanopore Signals using a Hash-based Seeding Mechanism"**

[Accepted to Bioinformatics] [arXiv](#), October 2024.

Paper [\[online\]](#) [\[pdf\]](#) [\[code\]](#)

[ISMB 2024](#) Talk [\[video\]](#) [\[pptx\]](#) [\[pdf\]](#)

[CSHL 2024 \(Biological Data Science\)](#) Talk [\[video\]](#) [\[pptx\]](#) [\[pdf\]](#)

Social Media Thread [\[Twitter \(X\)\]](#) [\[LinkedIn\]](#)

## **Rawsamble: Overlapping and Assembling Raw Nanopore Signals using a Hash-based Seeding Mechanism**

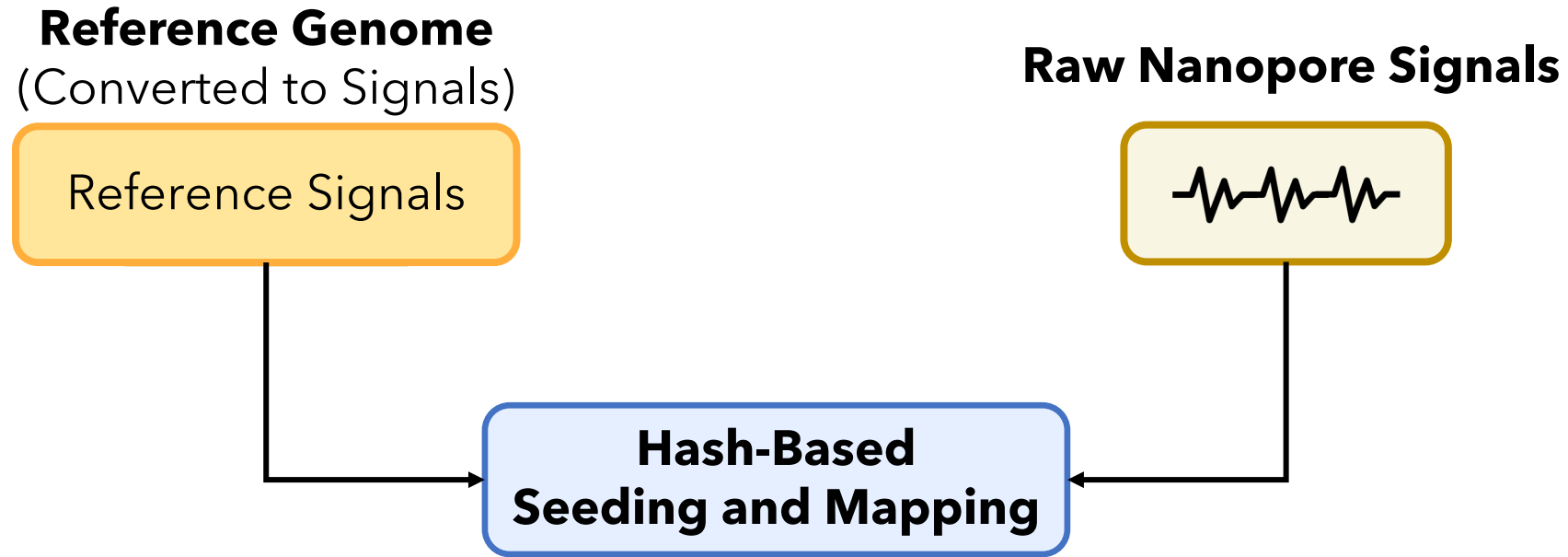
Can Firtina<sup>1</sup> Maximilian Mordig<sup>1,2</sup> Harun Mustafa<sup>1,3,4</sup> Sayan Goswami<sup>1</sup> Nika Mansouri Ghiasi<sup>1</sup>

Stefano Mercogliano<sup>1</sup> Furkan Eris<sup>1</sup> Joël Lindegger<sup>1</sup> Andre Kahles<sup>1,3,4</sup> Onur Mutlu<sup>1</sup>

<sup>1</sup>*ETH Zurich*      <sup>2</sup>*Max Planck Institute for Intelligent Systems*

<sup>3</sup>*University Hospital Zurich*      <sup>4</sup>*Swiss Institute of Bioinformatics*

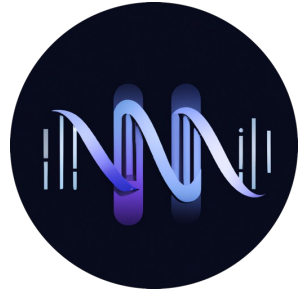
# Beyond Reference Mapping: Overlapping



 **Challenge:** Reference genome is not always available

 **Assembly:** Constructing genome from **overlapping reads**

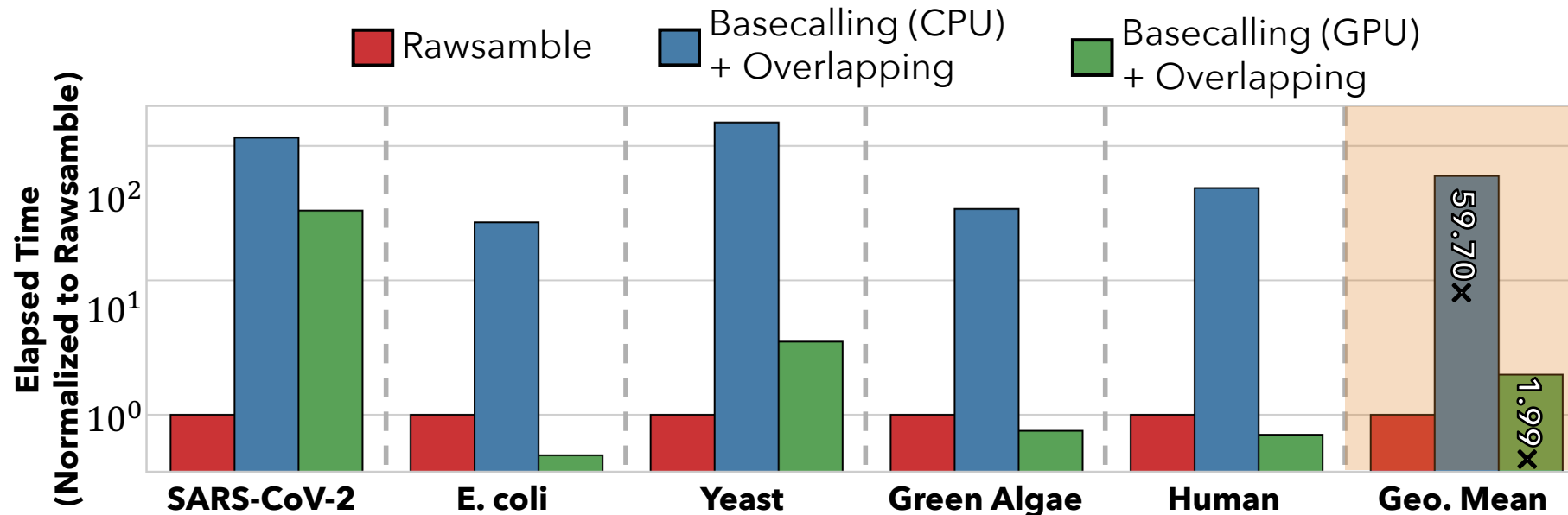
 Existing solutions cannot find overlapping reads **without basecalling**



# Rawsamble

**The first mechanism that can perform all-vs-all overlapping from raw signals**

# Eliminating Basecalling from the Pipeline

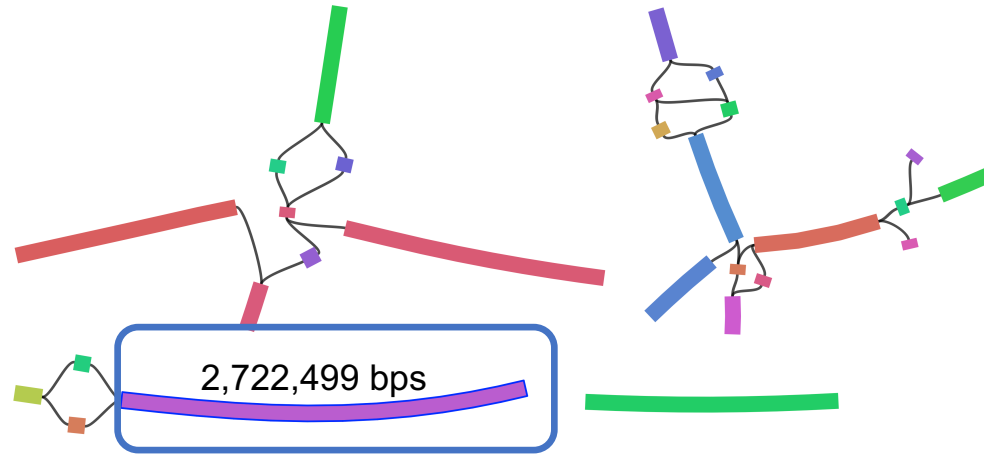


**Average speedup of 60x** than CPU-based basecalling:  
**Eliminating basecalling** leaves a significant room  
for performance improvements

**Average speedup of 2x** than the **GPU-based basecalling**:  
GPU vs. GPU comparison is likely to provide even better results

# Key Results – First Ever Assemblies

**E. coli Assembly  
(From the Rawsample Overlaps)**



**First assemblies  
ever constructed**  
without basecalling

**~400× longer assembled pieces**  
than the average read length:  
Informative overlaps lead to  
longer paths

**It is time to rethink if we want to translate  
individual reads separately**

# Quickly Aligning Raw Signals [IEEE Access '24]

- Joël Lindegger, Can Firtina, Nika Mansouri Ghiasi, Mohammad Sadrosadati, Mohammed Alser, and Onur Mutlu,

## "RawAlign: Accurate, Fast, and Scalable Raw Nanopore Signal Mapping via Combining Seeding and Alignment"

[IEEE Access](#), December 2024.

Paper [\[online\]](#) [\[pdf\]](#) [\[code\]](#)

## RawAlign: Accurate, Fast, and Scalable Raw Nanopore Signal Mapping via Combining Seeding and Alignment

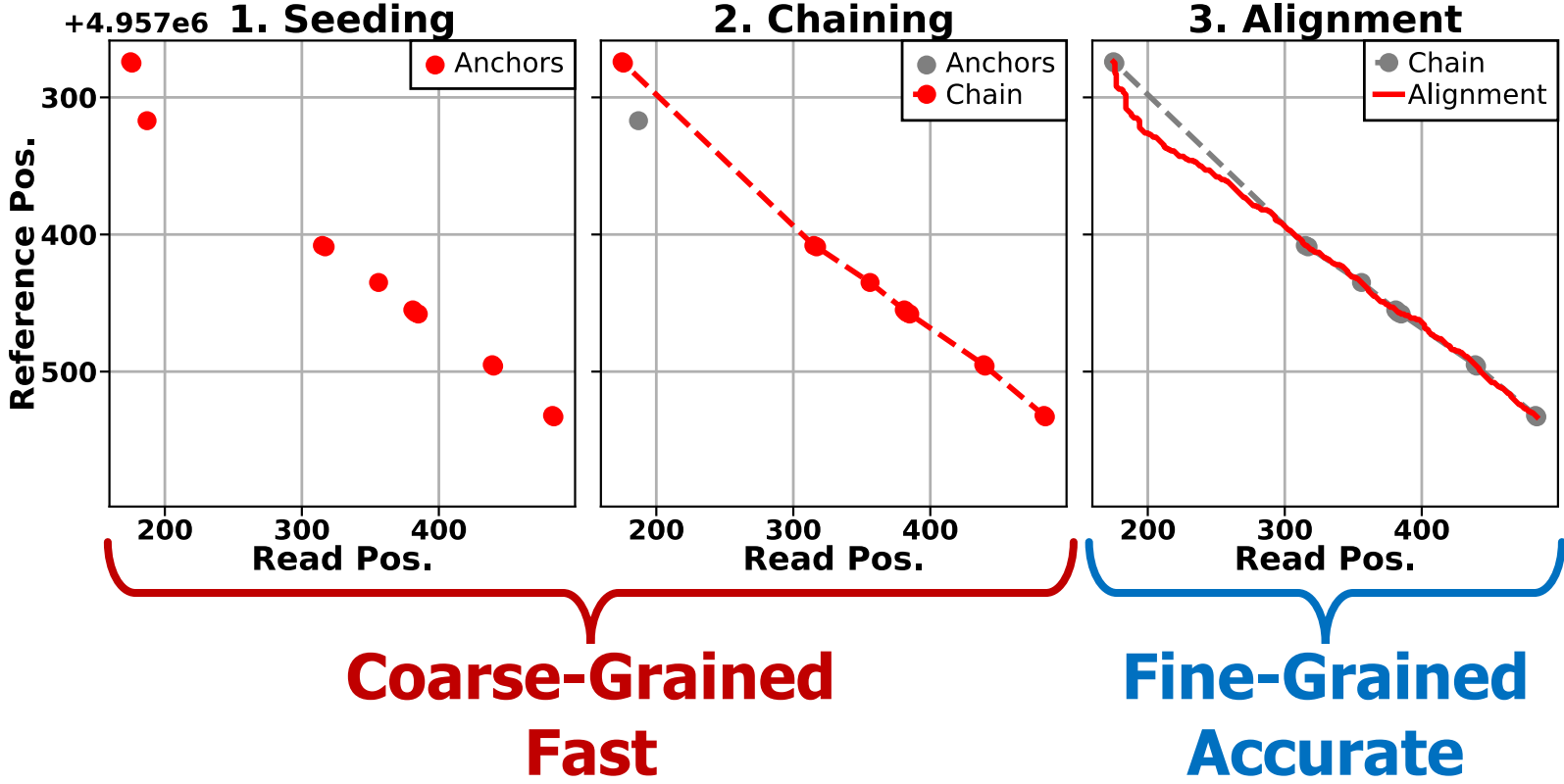
**JOËL LINDEGGER<sup>ID</sup>, CAN FIRTINA<sup>ID</sup>, NIKA MANSOURI GHIASI, MOHAMMAD SADROSADATI<sup>ID</sup>,  
MOHAMMED ALSER<sup>ID</sup>, AND ONUR MUTLU<sup>ID</sup>, (Fellow, IEEE)**

Department of Information Technology and Electrical Engineering, ETH Zürich, 8092 Zürich, Switzerland

Corresponding authors: Joël Lindegger (jmlindegger@gmail.com), Can Firtina (firtinac@ethz.ch), and Onur Mutlu (omutlu@ethz.ch)

This work was supported in part by the Semiconductor Research Corporation (SRC), in part by the ETH Future Computing Laboratory, in part by European Union's Horizon Programme for Research and Innovation under Grant 101047160-BioPIM, and in part by the Swiss National Science Foundation (SNSF) under Grant 200021\_213084.

# RawAlign Overview



# Benchmarking Raw Signal Analysis [ACM-BCB '25]

- Furkan Eris, Ulysse McConnell, Can Firtina, and Onur Mutlu,  
**"RawBench: A Comprehensive Benchmarking Framework  
for Raw Nanopore Signal Analysis Techniques"**

Proceedings of the [16th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics \(ACM-BCB 2025\)](#), Philadelphia, PA, USA, October 2025.

Paper [\[online\]](#) [\[pdf\]](#)

Conference Talk [\[pptx\]](#) [\[pdf\]](#)

## RawBench: A Comprehensive Benchmarking Framework for Raw Nanopore Signal Analysis Techniques

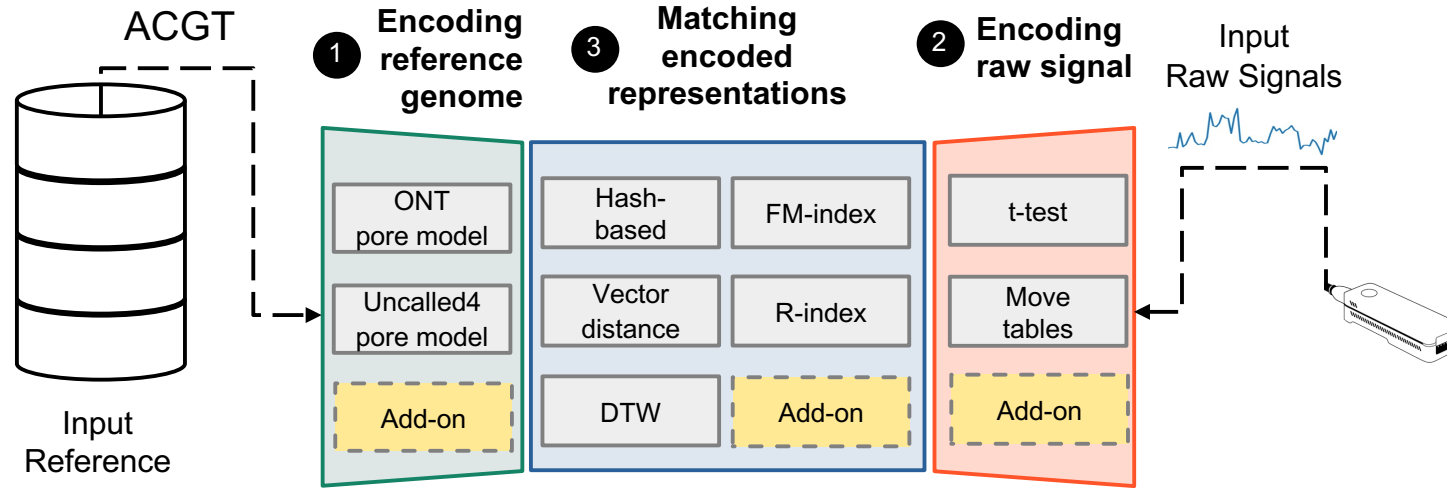
Furkan Eris  
ETH Zurich  
Zurich, Switzerland  
furkan6olm@gmail.com

Ulysse McConnell  
ETH Zurich  
Zurich, Switzerland  
umcconnell@student.ethz.ch

Can Firtina\*  
University of Maryland, College Park  
MD, USA  
firtina@umd.edu

Onur Mutlu\*  
ETH Zurich  
Zurich, Switzerland  
omutlu@gmail.com

# RawBench Benchmarking Framework



## 1 Encoding Reference Genome

- Uses learned pore models by ONT and community

## 2 Encoding Raw Signal

- Utilizes statistical and ML-based encoding techniques

## 3 Matching Encoded Representations

- Compares representations that are now encoded into the same space

# In-Storage Raw Signal Analysis [ICS '25]

- Melina Soysal, Konstantina Koliogeorgi, Can Firtina, Nika Mansouri Ghiasi, Rakesh Nadig, Haiyu Mao, Geraldo Francisco Oliveira Junior, Yu Liang, Klea Zambaku, Mohammad Sadrosadati, and Onur Mutlu,

## **"MARS: Processing-In-Memory Acceleration of Raw Signal Genome Analysis Inside the Storage Subsystem"**

Proceedings of the [39th International Conference on Supercomputing \(ICS 2025\)](#), Salt Lake City, Utah, USA, June 2025.

Paper [\[online\]](#) [\[pdf\]](#)

Conference Talk [\[video\]](#) [\[pptx\]](#) [\[pdf\]](#)

## **MARS: Processing-In-Memory Acceleration of Raw Signal Genome Analysis Inside the Storage Subsystem**

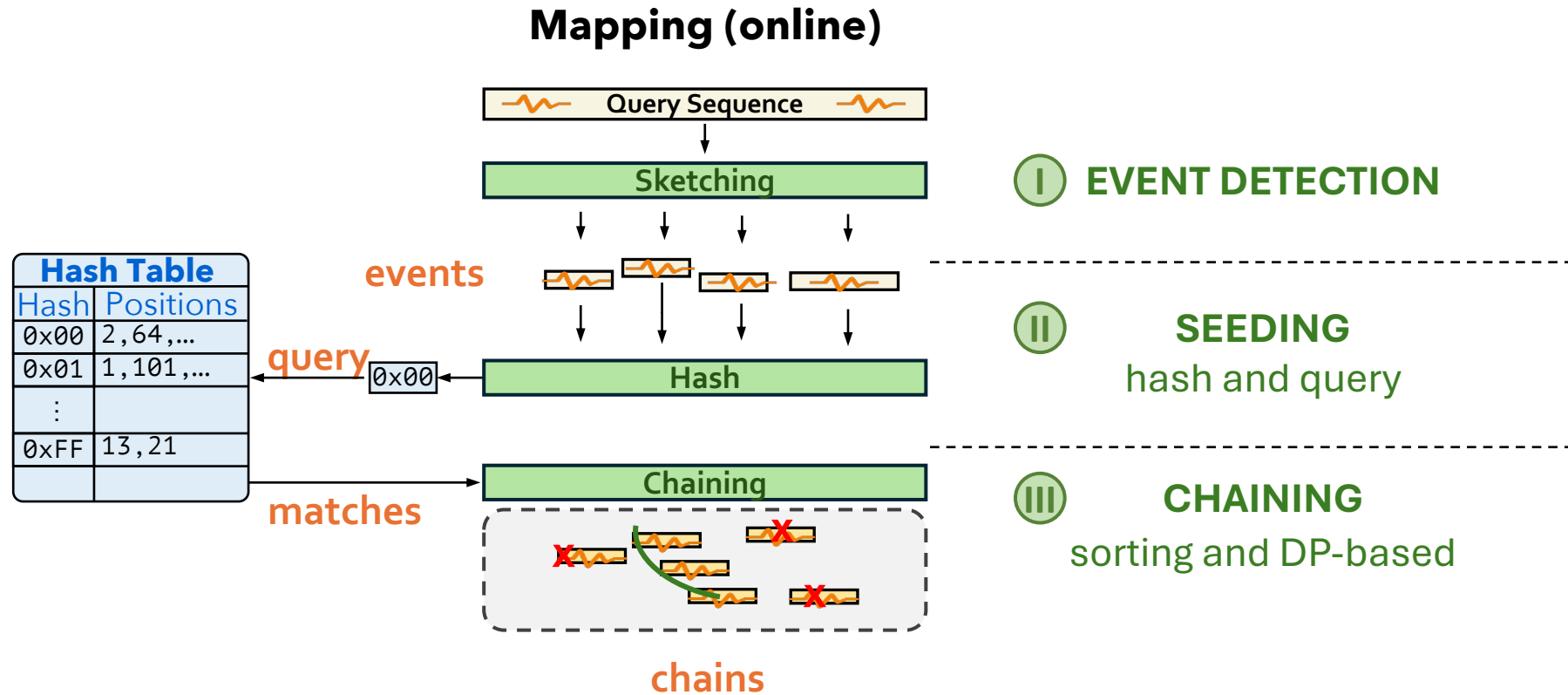
Melina Soysal<sup>†</sup>   Konstantina Koliogeorgi<sup>†</sup>   Can Firtina<sup>†</sup>   Nika Mansouri Ghiasi<sup>†</sup>  
Rakesh Nadig<sup>†</sup>   Haiyu Mayo<sup>\*</sup>   Geraldo F. Oliveira<sup>†</sup>  
Yu Liang<sup>†</sup>   Klea Zambaku<sup>†</sup>   Mohammad Sadrosadati<sup>†</sup>   Onur Mutlu<sup>†</sup>

<sup>†</sup> *ETH Zürich*

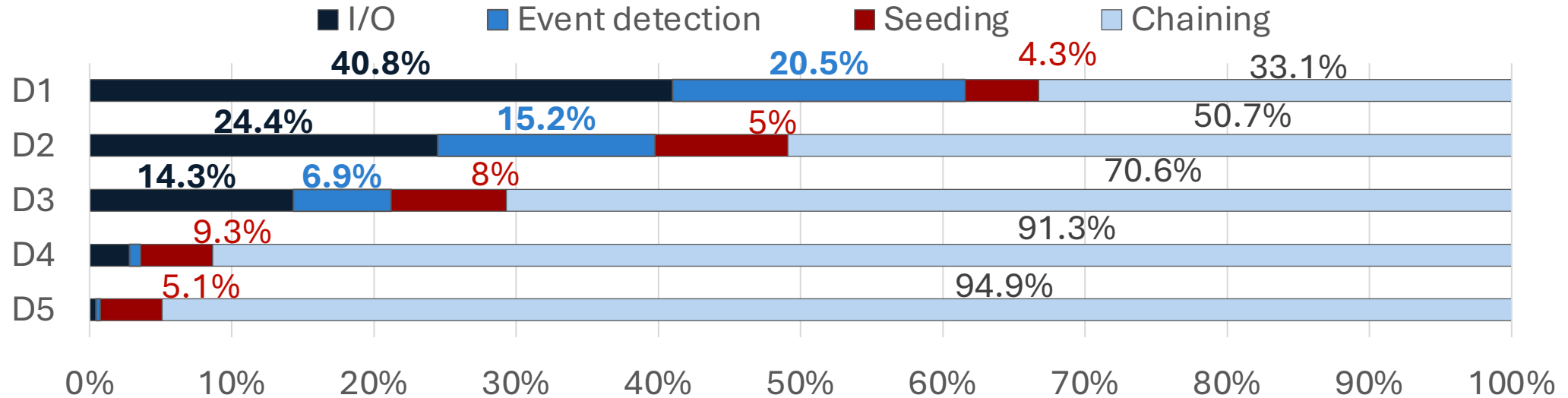
<sup>\*</sup> *King's College London*

# Read Mapping with Raw Signals

## High Level Overview of Read Mapping in Raw Signal Genome Analysis (RSGA)



# Motivation

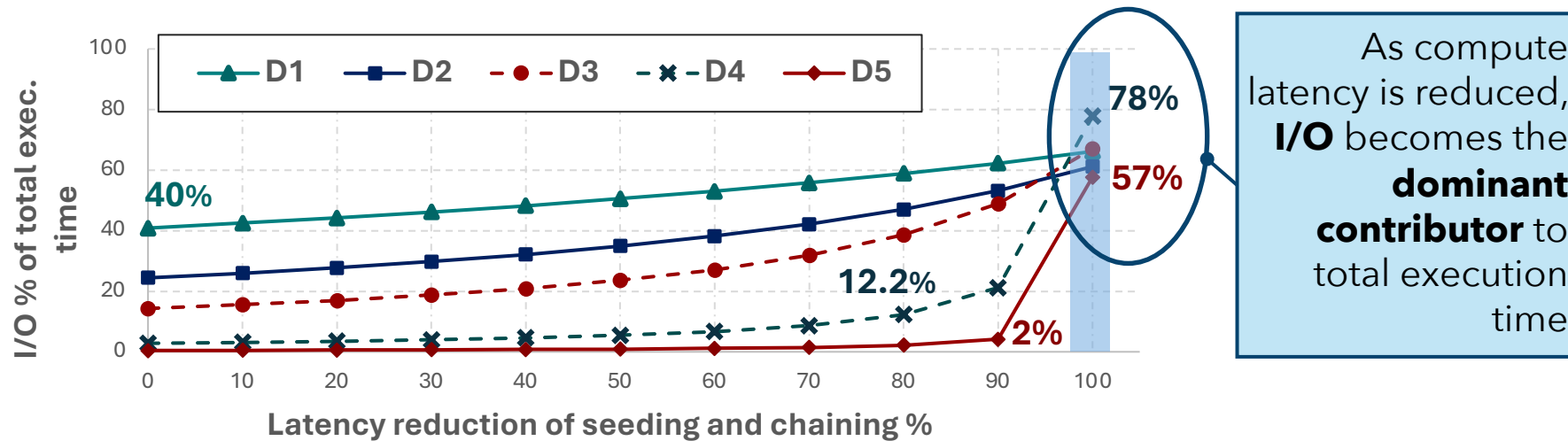


**Chaining** & **seeding** consistently dominate runtime

**Event detection** & **I/O overhead** are non-negligible for some datasets

# Impact of I/O data movement

We model **I/O data movement overhead** when accelerating commonly targeted steps in the pipeline: **seeding** and **chaining**

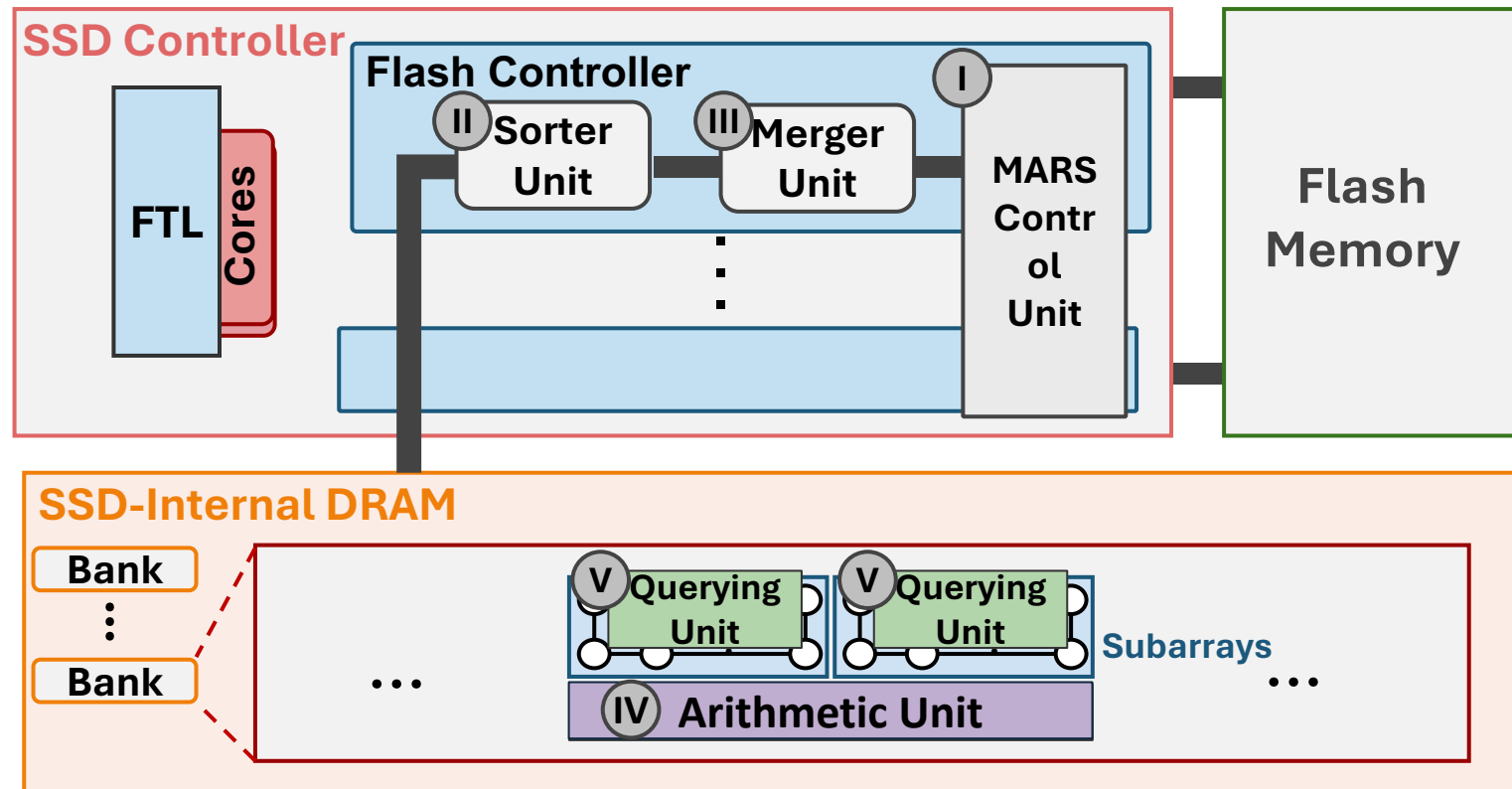


**I/O data movement from SSD becomes dominant overhead as compute steps are accelerated in RSGA**

# MARS Architecture & System - Overview

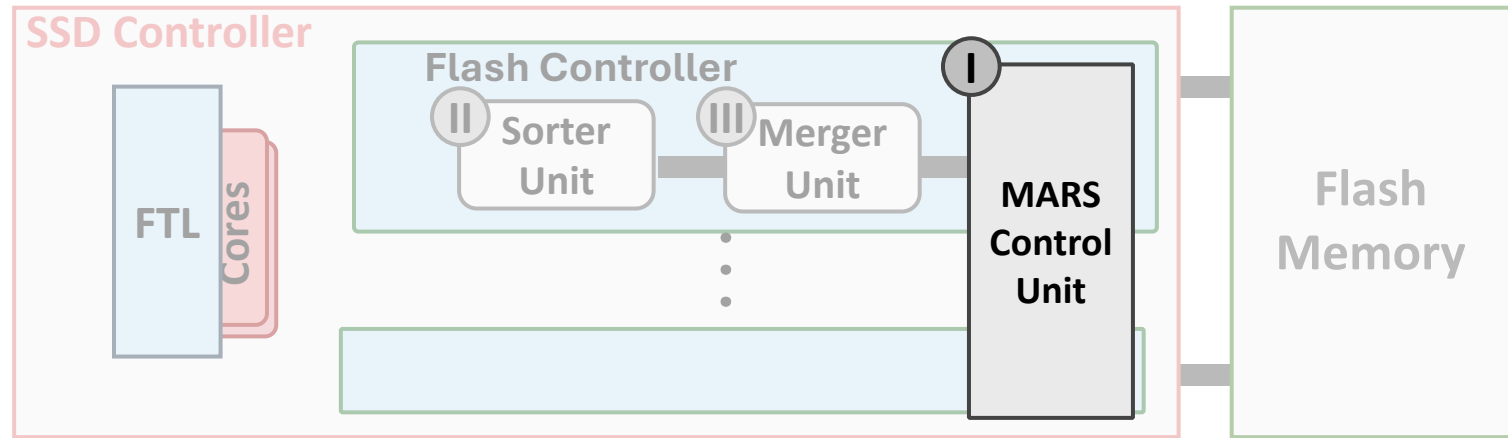
## In-Storage-Processing system for RSGA

- **Conventional Mode:** SSD performs standard SSD storage operations
- **Accelerator Mode:** SSD executes the full RSGA pipeline without host intervention



# MARS Architecture & System (I/IV)

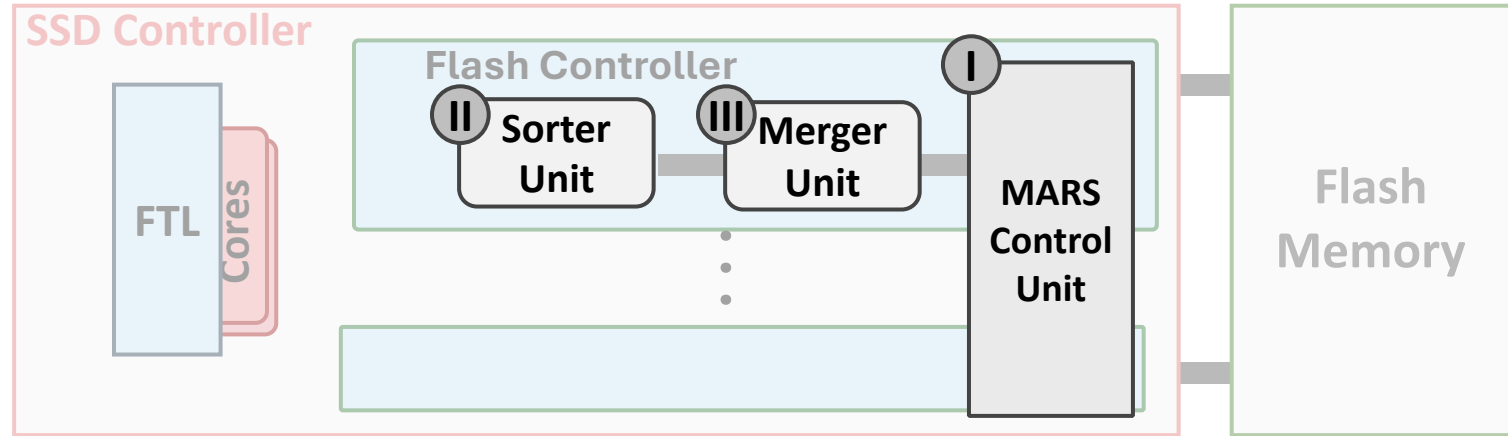
## SSD Controller Components



**I** **MARS Control Unit**  
Finite State Machine  
that orchestrates  
pipeline execution,  
control and data flow

# MARS Architecture & System (II/IV)

## SSD Controller Components



**I** **MARS Control Unit**  
Finite State Machine that orchestrates pipeline execution, control and data flow

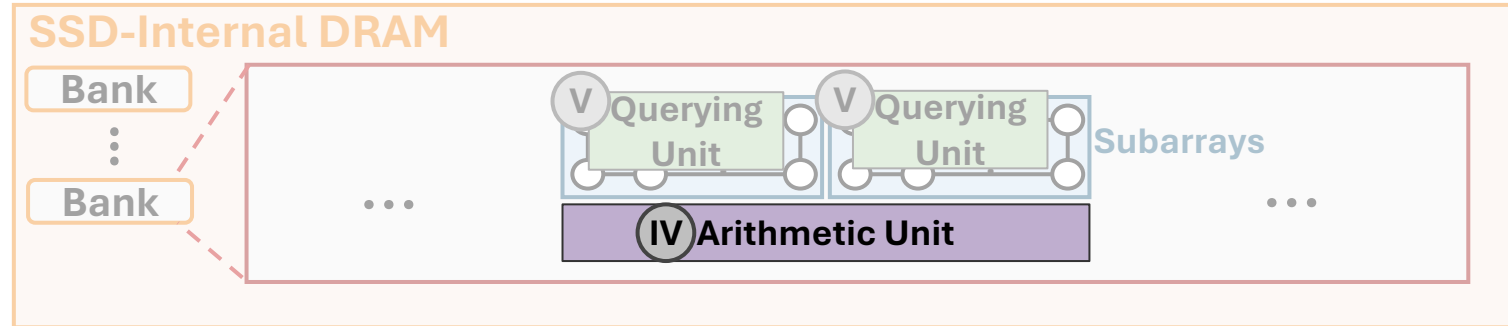
**II** **Sorter Unit**  
Performs parallel sorting of sequences using dedicated logic

**III** **Merger Unit**  
Combines sorted subsequences into longer sorted ones

**Paradigm:** Implements **Processing Near DRAM**  
*1 Sorter-Merger pair per flash controller*

# MARS Architecture & System (III/IV)

## SSD-Internal DRAM Components



**IV**

### Arithmetic Unit

*Performs arithmetic and logical operations required for RSGA*

**Paradigm:** Implements Processing Near DRAM

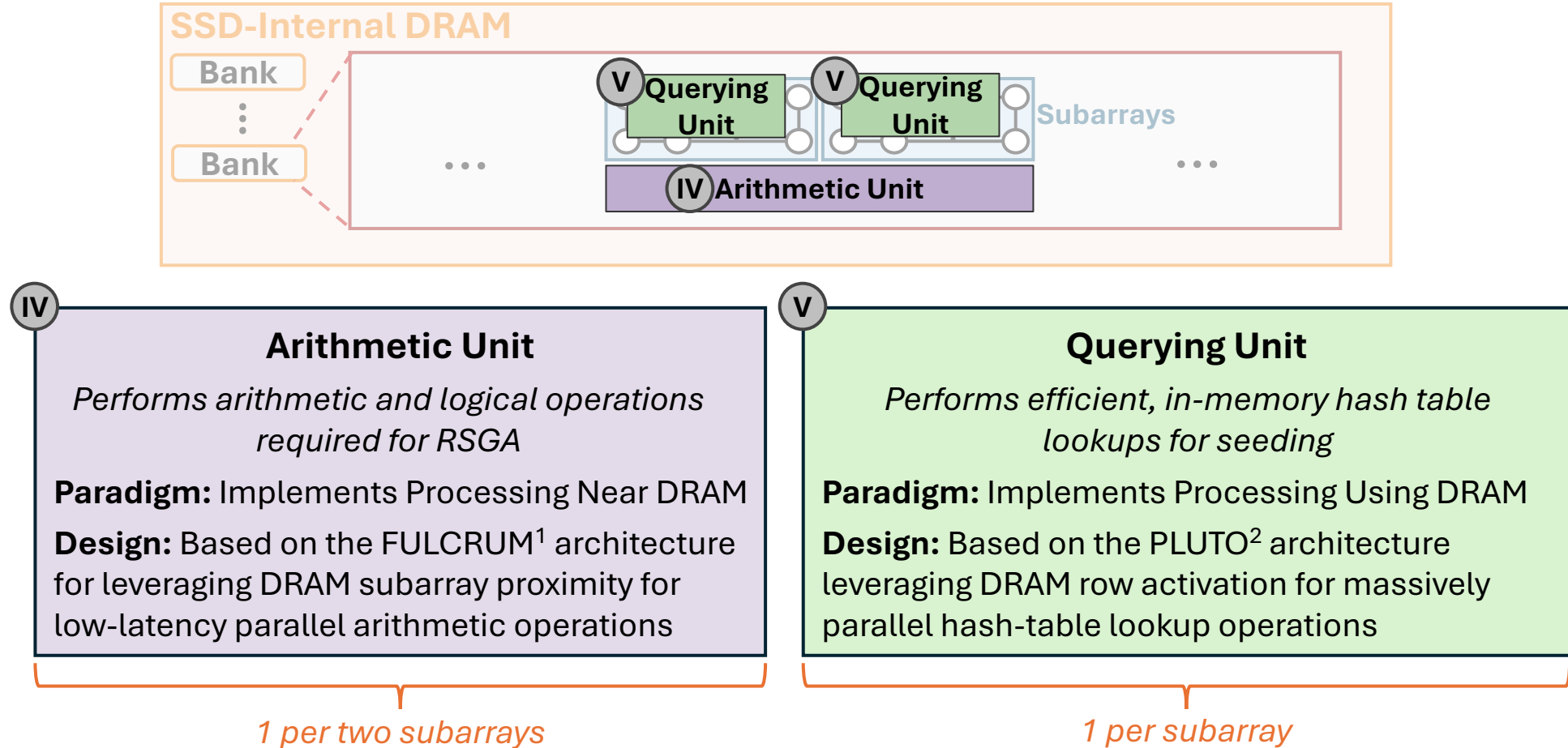
**Design:** Based on the FULCRUM<sup>1</sup> architecture for leveraging DRAM subarray proximity for low-latency parallel arithmetic operations

*1 per two subarrays*

1: [Lenjani, HPCA'20];

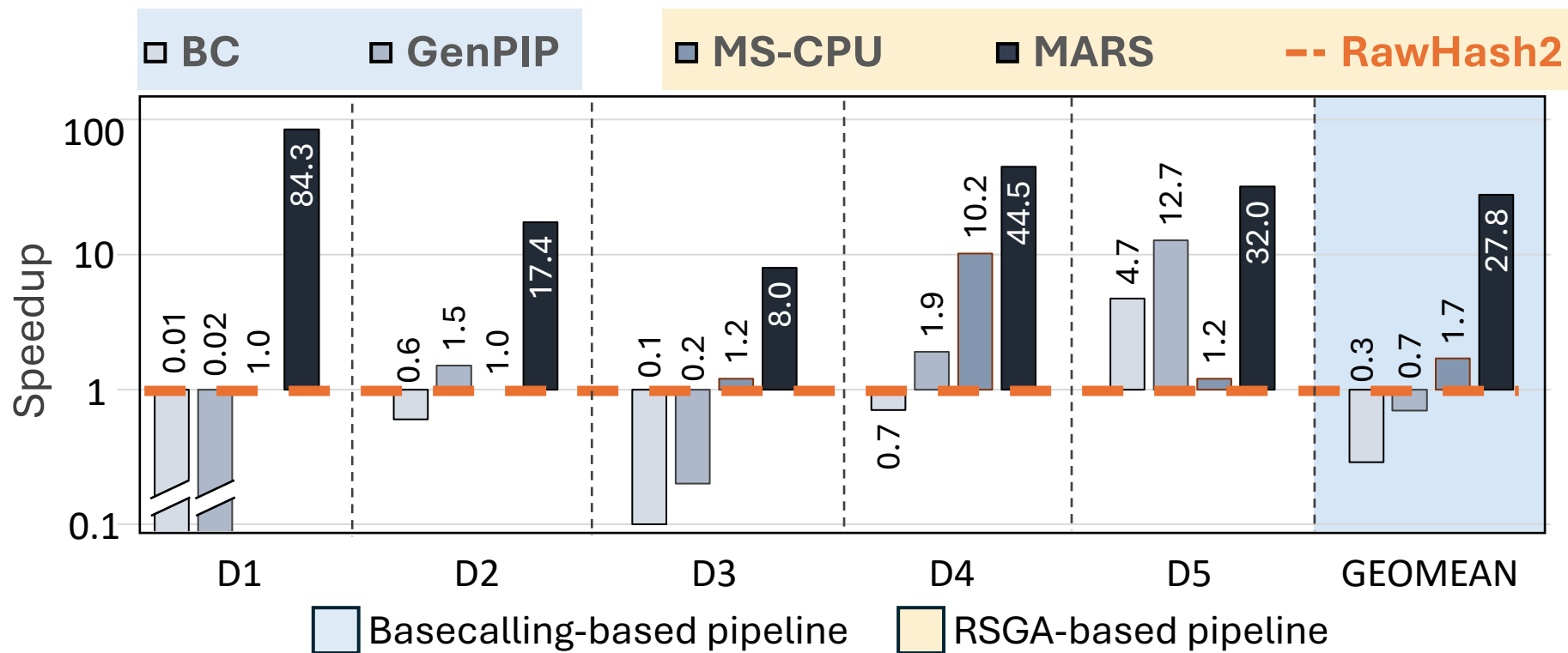
# MARS Architecture & System (IV/IV)

## SSD-Internal DRAM Components



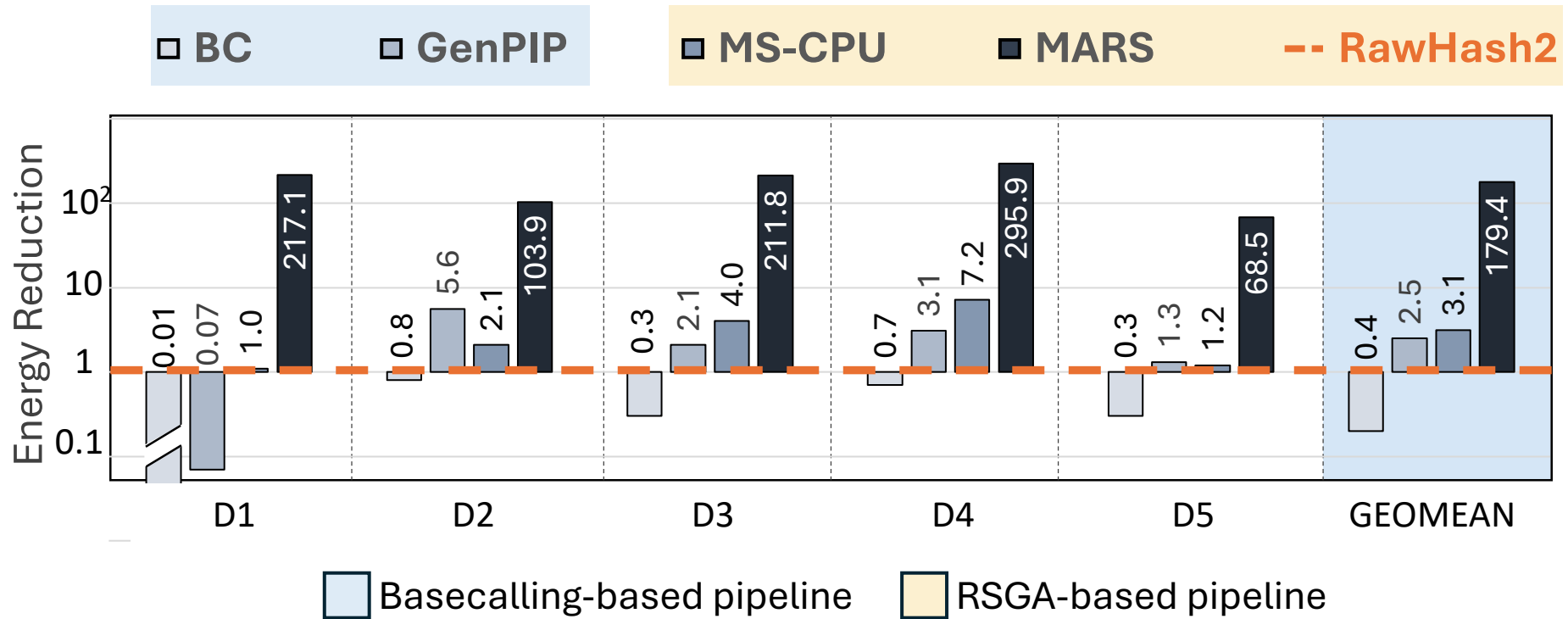
1: [Lenjani, HPCA'20]; 2: [Ferreira+, MICRO'22]

# Performance Evaluation



- Key Observations**
- MARS outperforms all other baselines across all datasets
  - MARS outperforms GenPIP adopting PIM outside storage  
*We fundamentally solve the I/O data movement overhead problem*
  - MARS-CPU provide speedup of 1.7x over RH2  
*Our SW modifications are effective in reducing the computational load*

# Energy Evaluation



## Key Observation

- MARS reduces energy consumption over all other baselines across all datasets
  - Improvement stems from*
  - (1) *reduced execution time*
  - (2) *fundamentally addressing the I/O data movement overhead*

Processing Data Where It Makes Sense  
Provides Significant Performance  
and Energy Benefits

# Some Interesting Challenges in Signal Analysis

- Identifying novel base modifications in various organisms
- Better integrating the chemical structure information in genomics
  - Distinguishing noise from useful information
- Redesigning effective solutions for storage
- Enabling end-to-end and energy-efficient analysis with (portable) sequencing devices

# STORM Research Group

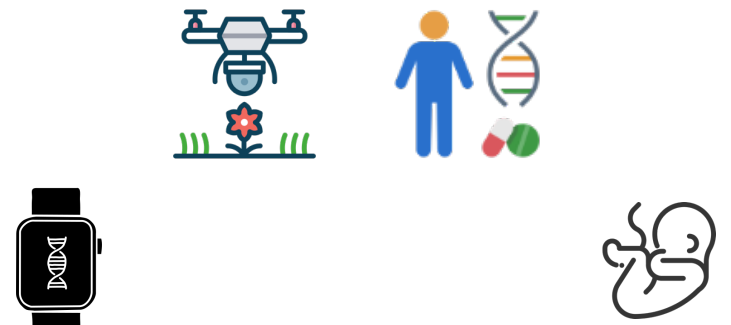
## Research Directions

 **Real-Time and Portable**  
Genomic Signal Analysis

 **Data-centric and AI-Driven** Solutions to Analyze Genomic Data

 Accurate and Fast Algorithms Scalable to **Large Volumes of Genomic Data**

### Today's Talk



### Goal

Enable fundamentally better integration of biological data analysis in our **everyday lives**



# Lost in Translation: Harnessing the Power of Nanopore Electrical Signals in Genomics

CDSL Seminar Talk, MD, USA  
February 11, 2026

**Can Firtina** | [canfirtina@gmail.com](mailto:canfirtina@gmail.com) | [cs.umd.edu/~firtina](https://cs.umd.edu/~firtina)

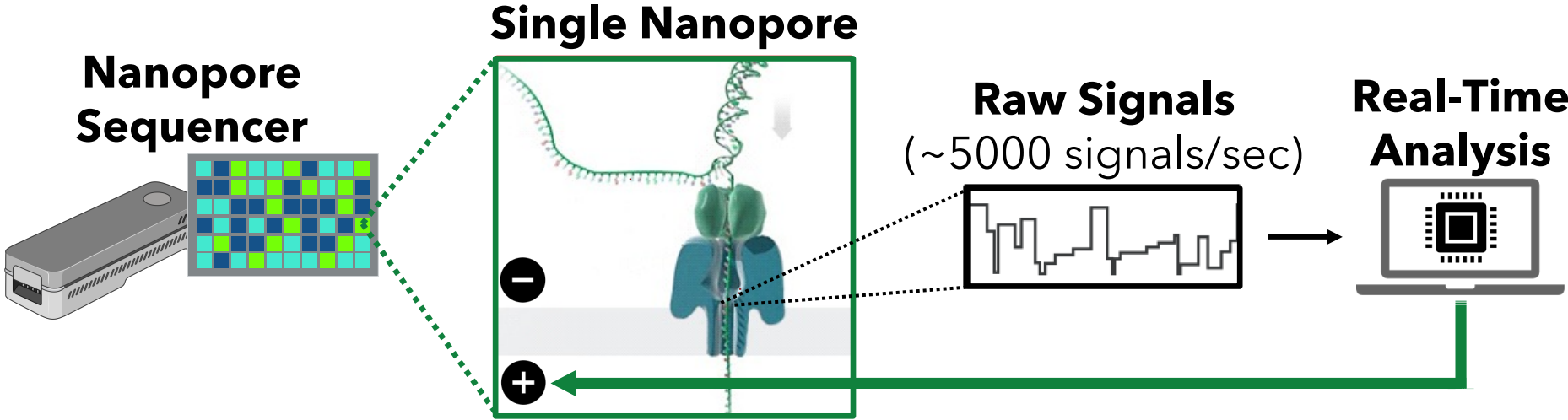
Assistant Professor of Computer Science at UMD



DEPARTMENT OF  
COMPUTER SCIENCE

# Backup Slides

# Generating Electrical Signals from DNA



# Basecalling using PIM [MICRO '23]

- Taha Shahroodi, Gagandeep Singh, Mahdi Zahedi, Haiyu Mao, Joel Lindegger, Can Firtina, Stephan Wong, Onur Mutlu, and Said Hamdioui,  
**"Swordfish: A Framework for Evaluating Deep Neural Network-based Basecalling using Computation-In-Memory with Non-Ideal Memristors"**  
*Proceedings of the 56th International Symposium on Microarchitecture (MICRO), Toronto, ON, Canada, November 2023.*  
[\[Slides \(pptx\) \(pdf\)\]](#)  
[\[arXiv version\]](#)

## **Swordfish: A Framework for Evaluating Deep Neural Network-based Basecalling using Computation-In-Memory with Non-Ideal Memristors**

Taha Shahroodi<sup>1</sup> Gagandeep Singh<sup>2,3</sup> Mahdi Zahedi<sup>1</sup> Haiyu Mao<sup>3</sup> Joel Lindegger<sup>3</sup> Can Firtina<sup>3</sup>  
Stephan Wong<sup>1</sup> Onur Mutlu<sup>3</sup> Said Hamdioui<sup>1</sup>

<sup>1</sup>TU Delft   <sup>2</sup>AMD Research   <sup>3</sup>ETH Zürich

# DNN Hardware Acceleration

DNN execution is dominated by:

Vector-Matrix  
Multiplication (VMM)

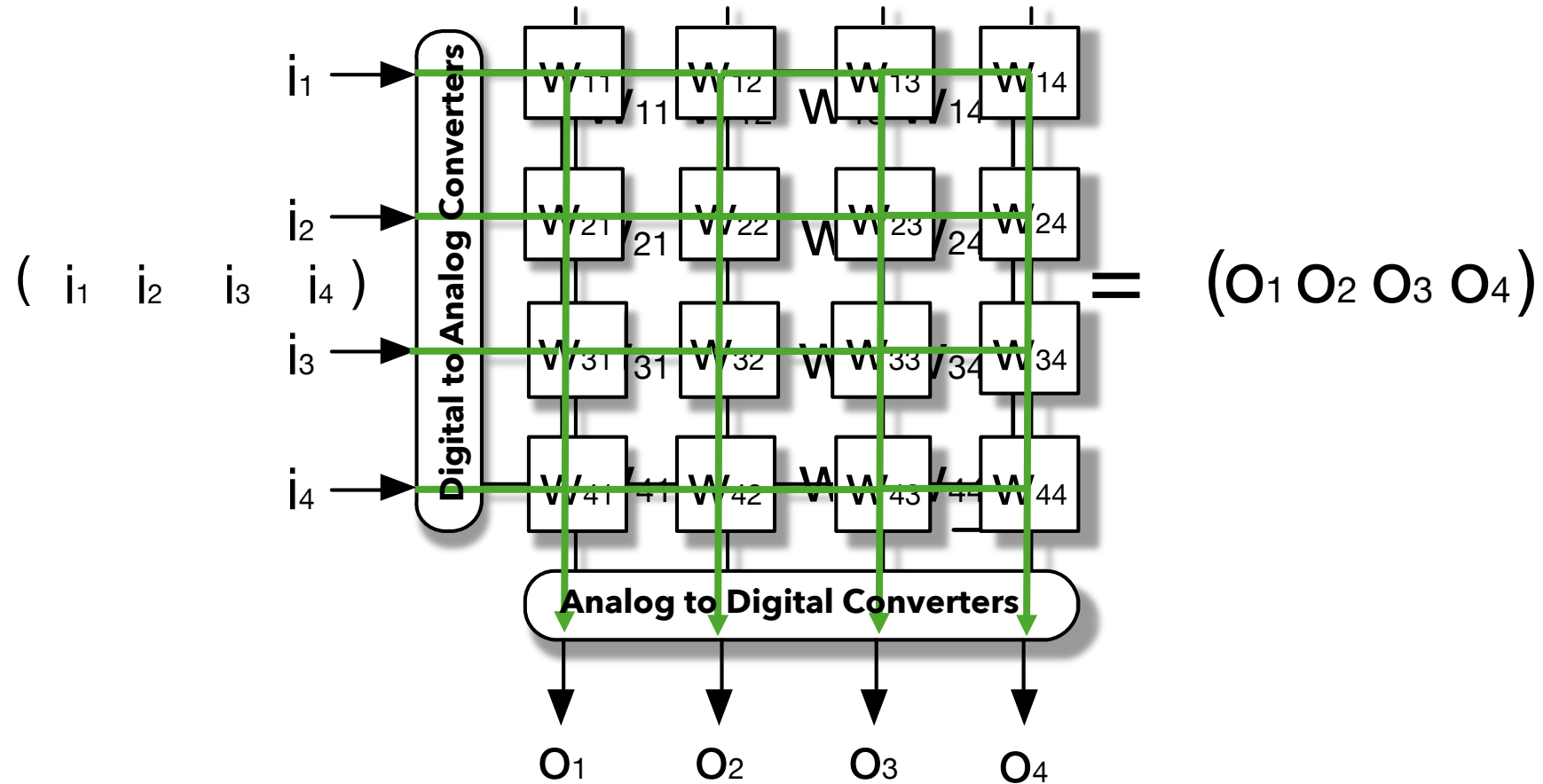
Data movement  
between memory and  
accelerator

## Memristor-based PIM for DNN Acceleration

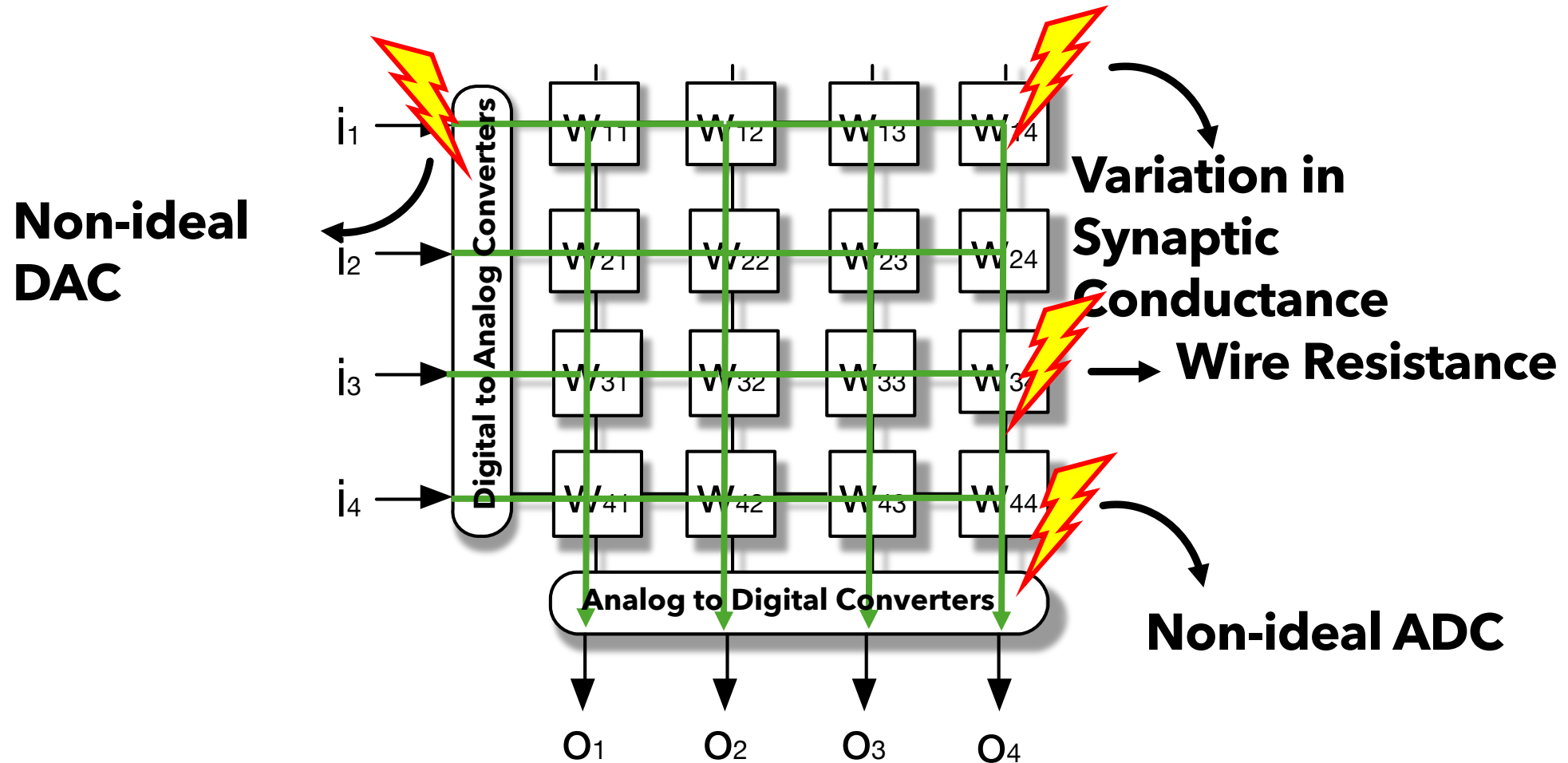
Memristor-based  
crossbars support  
VMM

Processing in Memory  
(PIM) minimizes data  
movement

# VMM in Memristor-based Crossbars

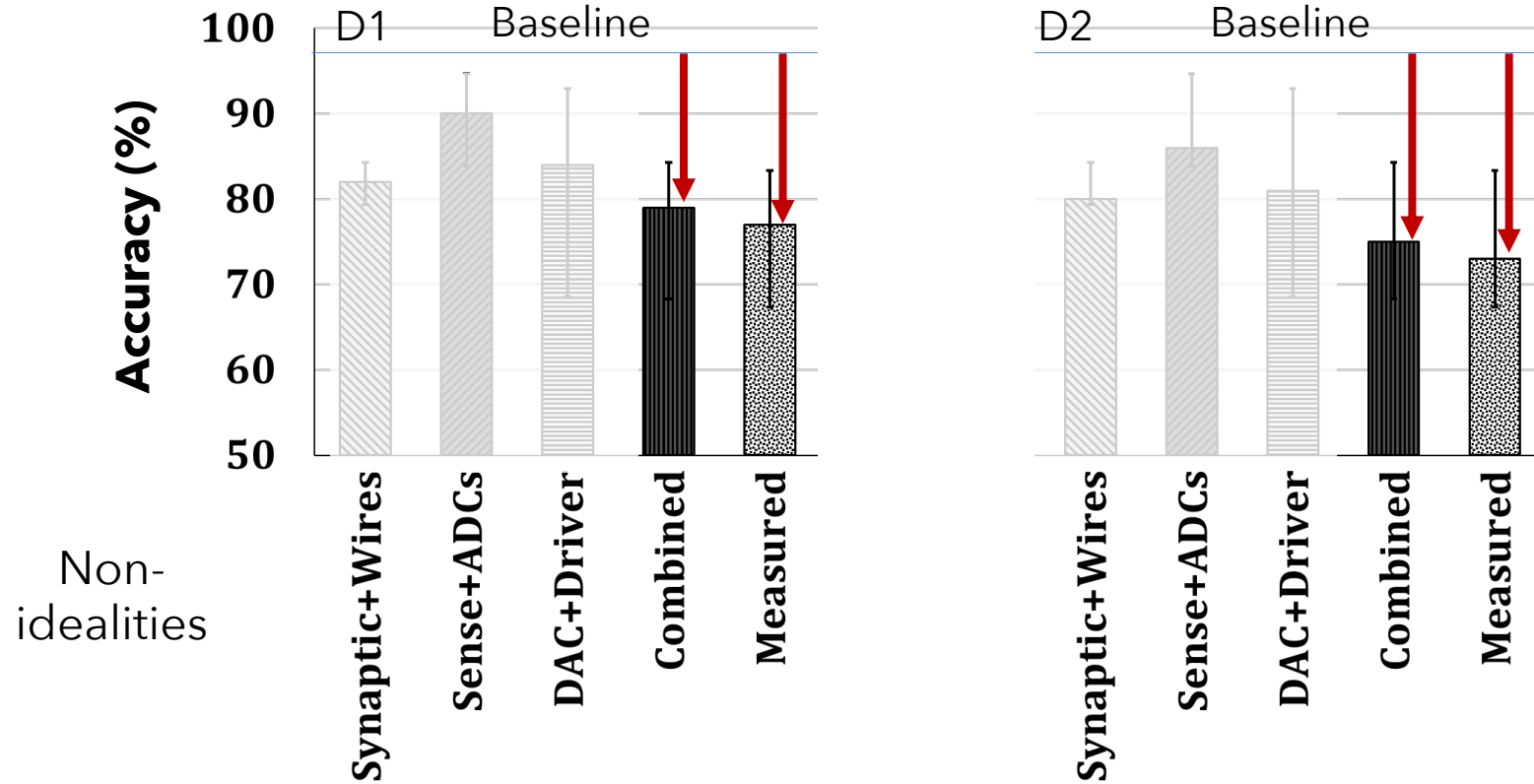


# Non-idealities in Memristor-based Crossbars



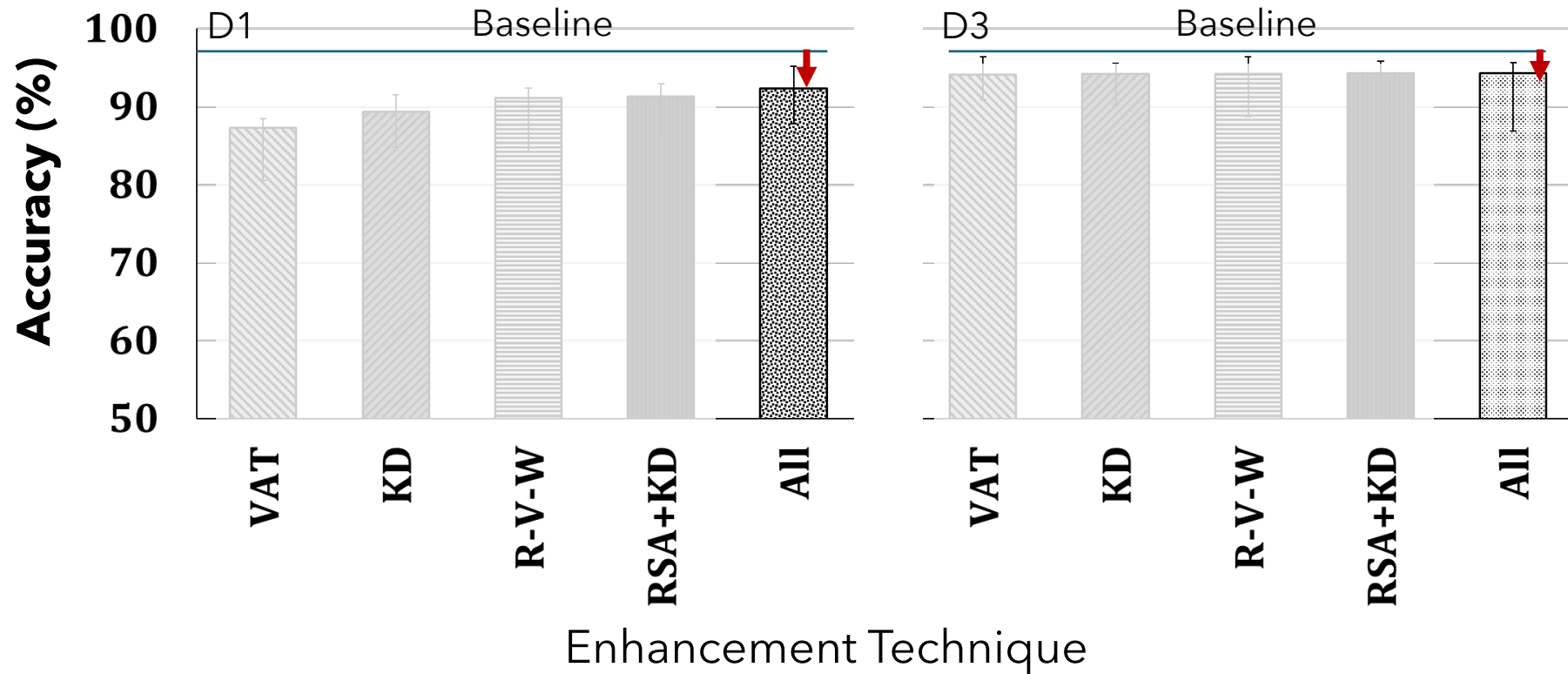
Subpar utilization of trained models  
**due to non-idealities**

# Accuracy: All Non-idealities without Mitigation



Combined non-idealities leads to significant accuracy loss (>18%)

# Accuracy: Enhancement Techniques on All Non-idealities



**Accuracy loss drops to 6% from 18%** after handling all non-idealities

# ML-based Genome Analysis via PIM [MICRO '22]

- Haiyu Mao, Mohammed Alser, Mohammad Sadrosadati, Can Firtina, Akanksha Baranwal, Damla Senol Cali, Aditya Manglik, Nour Almadhoun Alserr, and Onur Mutlu, **"GenPIP: In-Memory Acceleration of Genome Analysis via Tight Integration of Basecalling and Read Mapping"**

Proceedings of the 55th International Symposium on Microarchitecture (**MICRO**), Chicago, IL, USA, October 2022.

[[Slides \(pptx\)](#) ([pdf](#))]

[[Longer Lecture Slides \(pptx\)](#) ([pdf](#))]

[[Lecture Video](#) (25 minutes)]

[[arXiv version](#)]

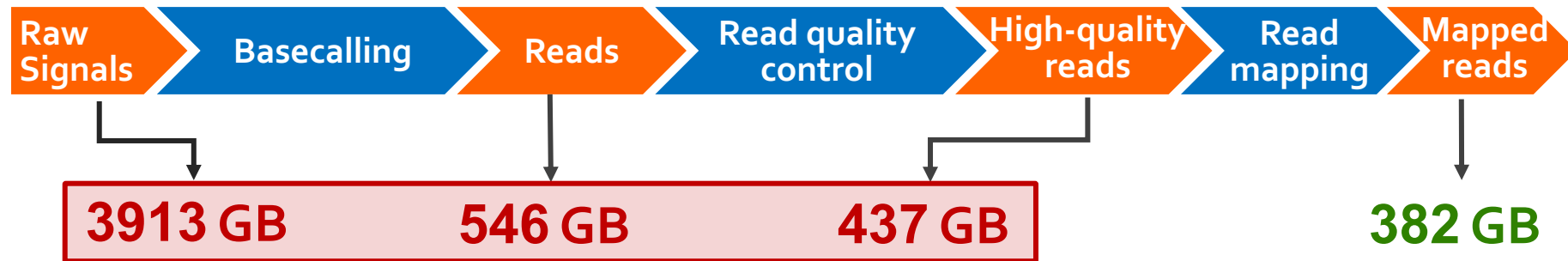
## **GenPIP: In-Memory Acceleration of Genome Analysis via Tight Integration of Basecalling and Read Mapping**

Haiyu Mao<sup>1</sup> Mohammed Alser<sup>1</sup> Mohammad Sadrosadati<sup>1</sup> Can Firtina<sup>1</sup> Akanksha Baranwal<sup>1</sup>  
Damla Senol Cali<sup>2</sup> Aditya Manglik<sup>1</sup> Nour Almadhoun Alserr<sup>1</sup> Onur Mutlu<sup>1</sup>

<sup>1</sup>*ETH Zürich*

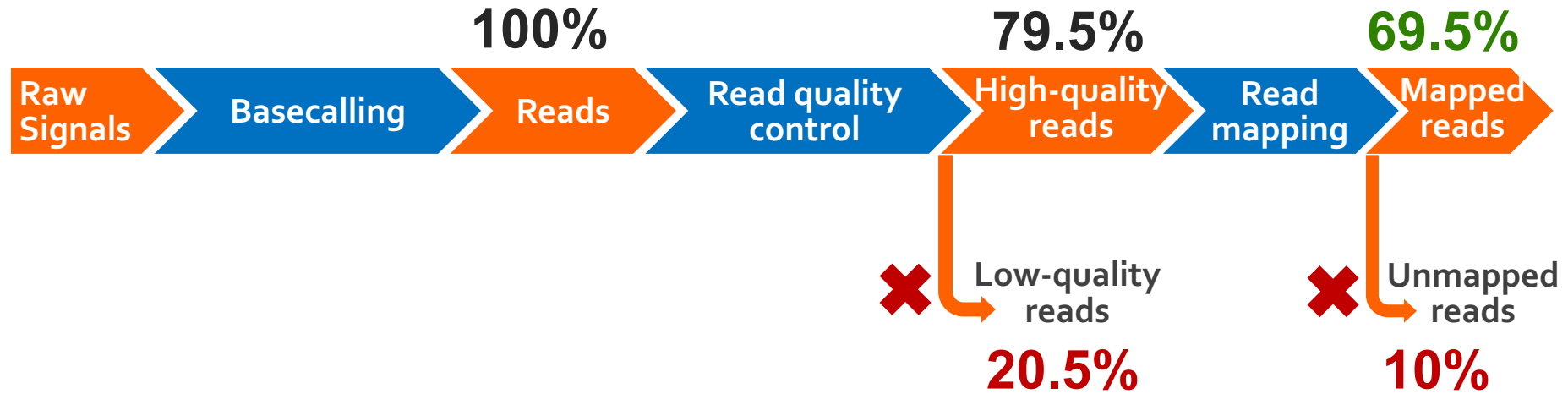
<sup>2</sup>*Bionano Genomics*

# Problem 1: Large Data Movement



**Large data movement** between genome analysis steps

# Problem 2: Wasted Computation



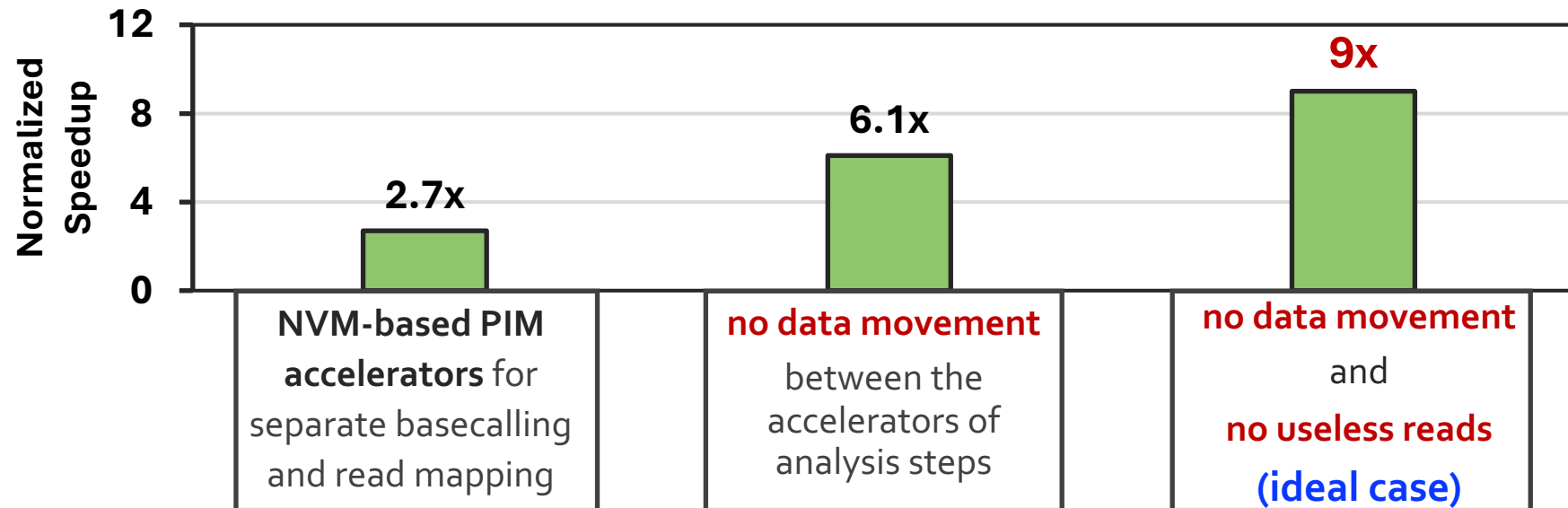
A considerable amount of computation on **useless data** due to

- Low-quality reads
- Unmapped reads

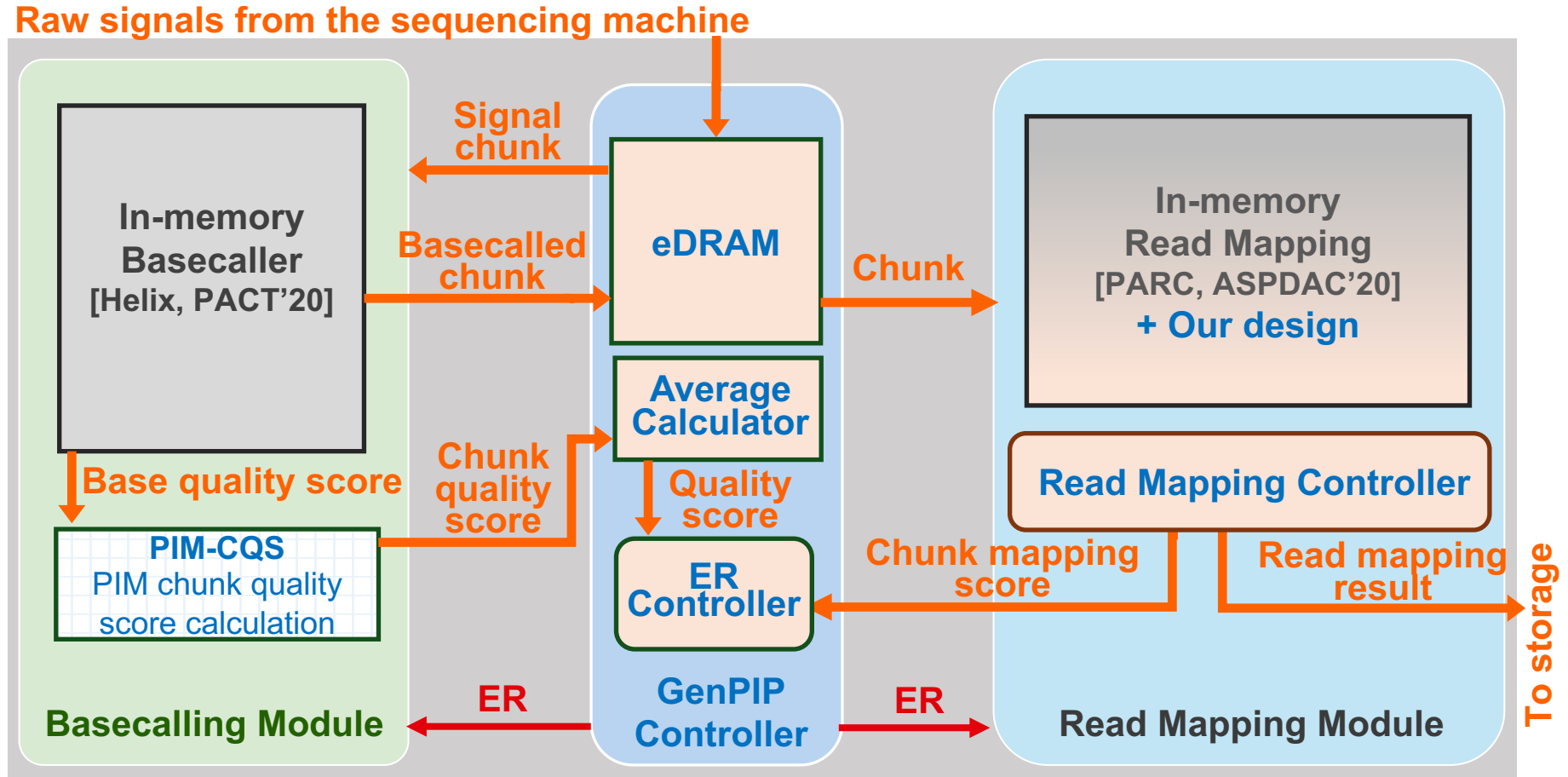
# Goal and Opportunities

**Goal:** Efficiently accelerate the entire genome analysis pipeline while **minimizing data movement and useless computation**

We perform a study to quantify potential performance benefits



# GenPIP Overview

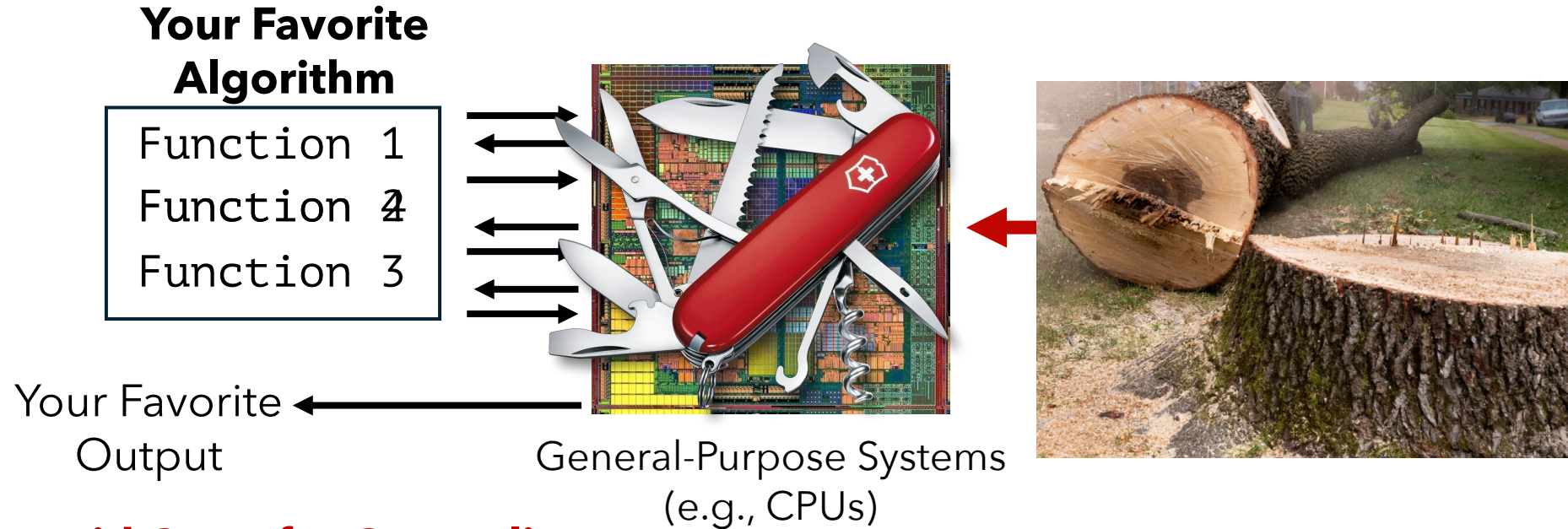


<https://arxiv.org/pdf/2209.08600.pdf>

# General-Purpose "Things"



# Algorithm Design Demands General-Purpose Computing



Your Favorite  
Output

General-Purpose Systems  
(e.g., CPUs)

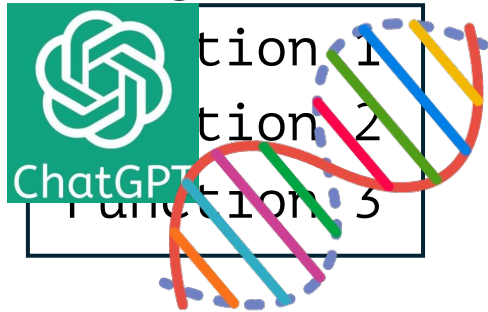
## **Paid Costs for Generality:**

Supporting unnecessary operations  
Inefficient data and control flow  
Limited parallelism & bandwidth

**High latency &  
Substantial energy**

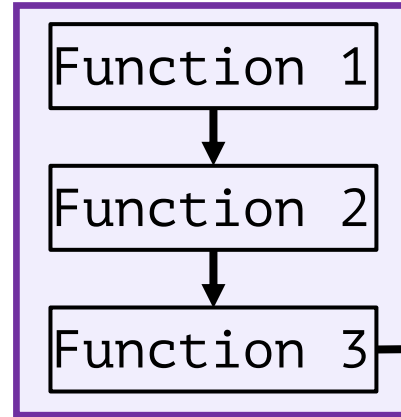
# Great Algorithms Demand Specialization

## Everyone's Favorite Algorithm



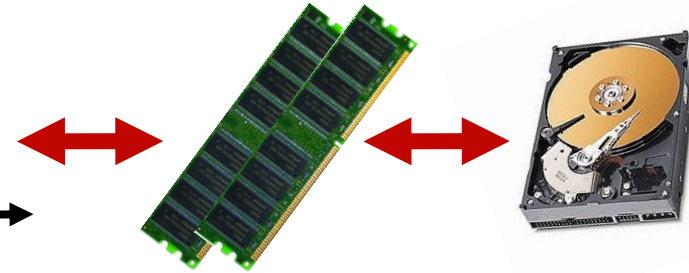
**Goal: Avoid wasting computation & energy**

## Application-Specific Architecture



## Opportunities:

Better data flow with minimal latency & power  
Efficient processors customized for the task  
Better parallelism



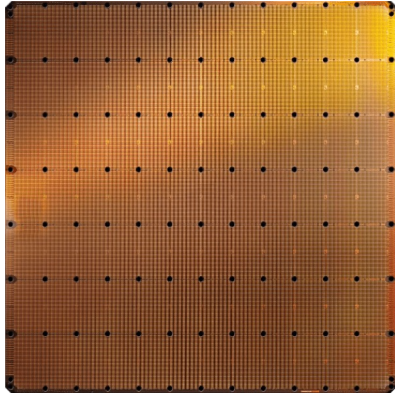
## Paid Costs for Specialization:

Poor programmability  
High manufacturing and engineering costs

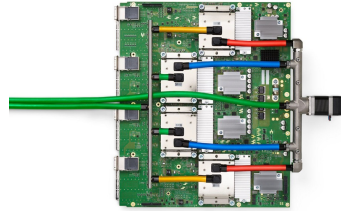
# Acceleration Efforts for AI and Genomics

## Cerebras WSE-3

The largest ML accelerator chip



## Google's TPUs:



## NVIDIA H100



### Scale your studies with ease

Process high-throughput data quickly with hardware acceleration

Process high-throughput data quickly with hardware acceleration. With four field-programmable gate arrays (FPGAs) onboard you have the most powerful DRAGEN analysis ever, enabling you to process NovaSeq X System data easily. Perform up to four simultaneous applications per flow cell in a single run.

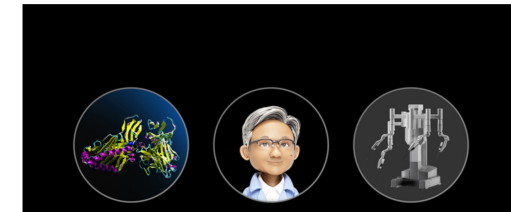
Reduce data footprint, manage and store data easily with lower costs and lower energy consumption, with built-in compression that reduces FASTQ file sizes by up to 80%.2

Stream data directly to Illumina Connected Analytics or BaseSpace Sequence Hub on the cloud for scalable data management, analysis, and aggregation.

### NVIDIA Partners With Industry Leaders to Advance Genomics, Drug Discovery and Healthcare

IQVIA, Illumina, Mayo Clinic and Arc Institute Harness NVIDIA AI and Accelerated Computing to Transform \$10 Trillion Healthcare and Life Sciences Industry

January 13, 2025



## Absci and AMD Announce Collaboration and Strategic Investment to Accelerate the Future of AI Drug Discovery

01/08/2025

PDF Version

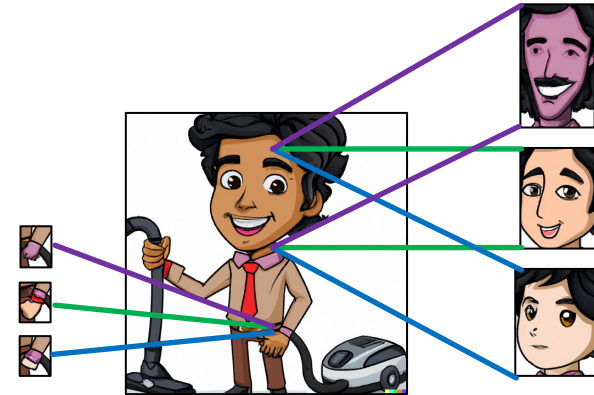
# Important and Commonly Used Applications Demand Specialization

# Probabilistic Graphs in Genomics

**Probabilistic graphs** are commonly used **to identify variations** between sequences

To avoid **redundant** comparisons and storage

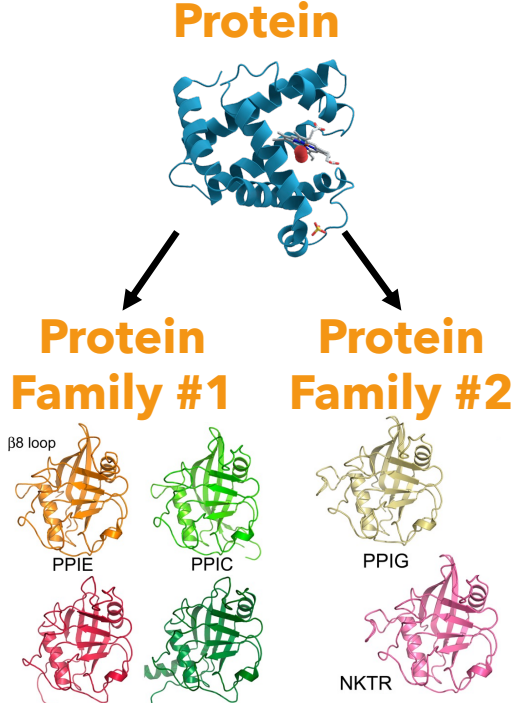
To provide **rich information** on expected variations



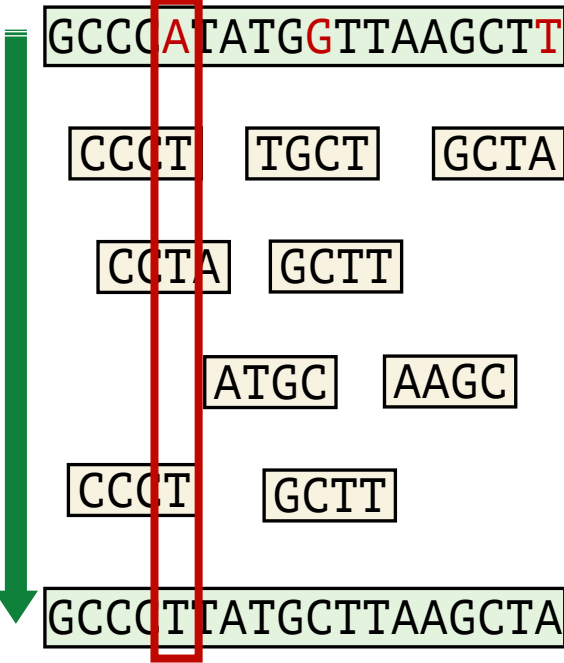
**Reference Genome  
Graph**

# Graph-Based Applications in Genomics

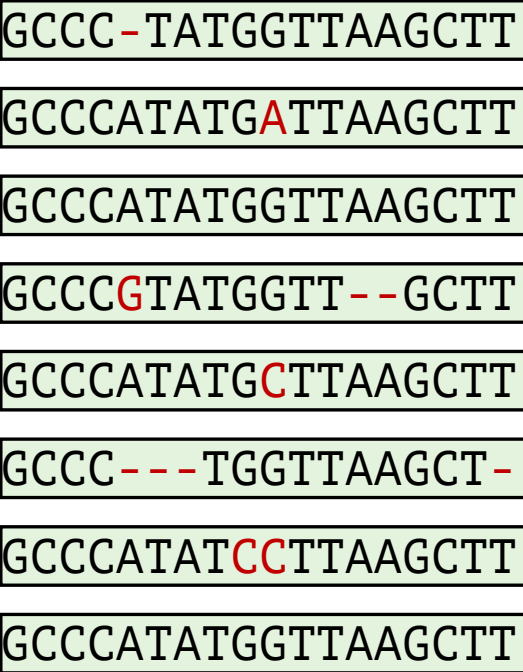
## Protein Family Search with Graph Alignment



## Error Correction with Probabilistic Graphs

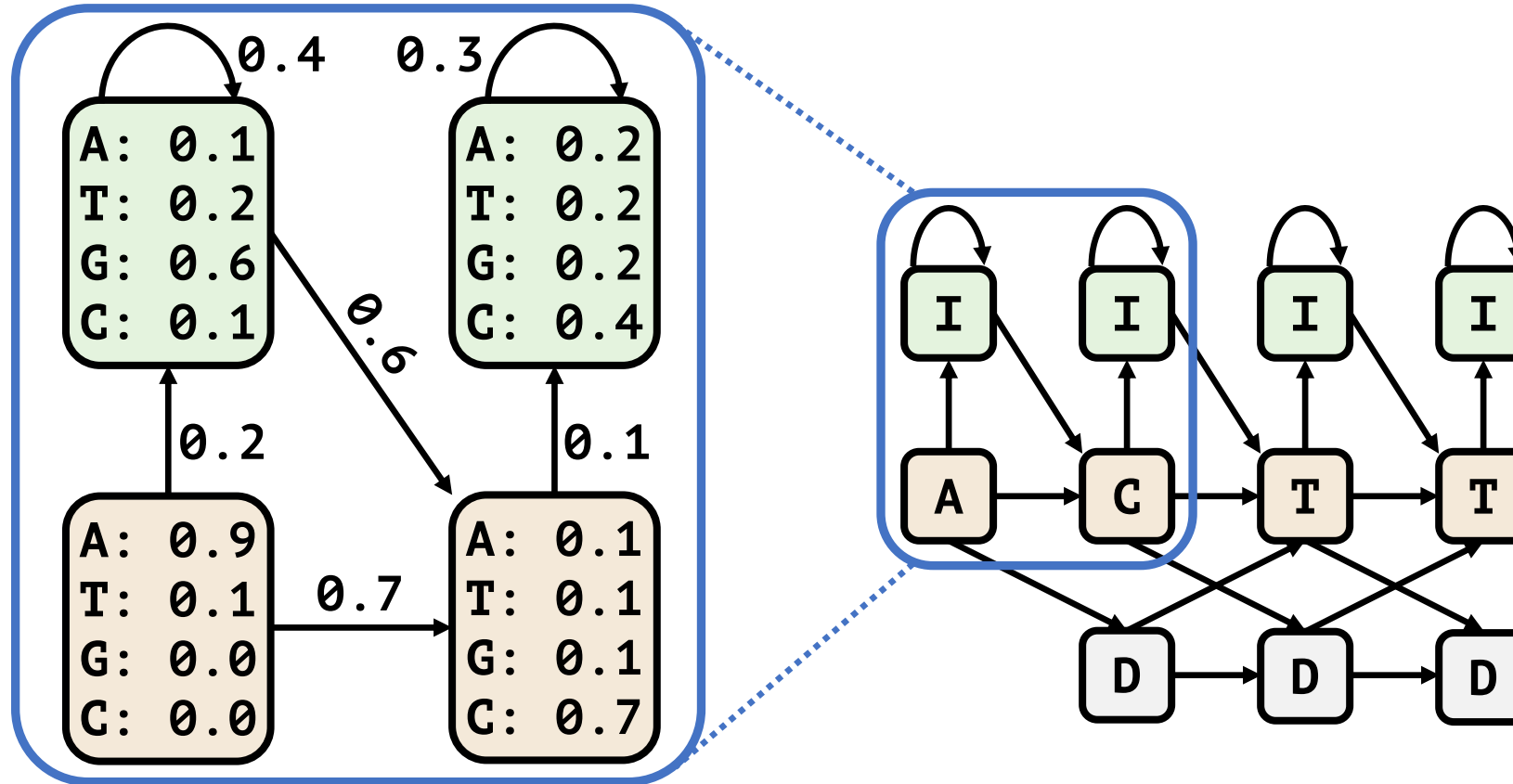


## Multiple Sequence Alignment

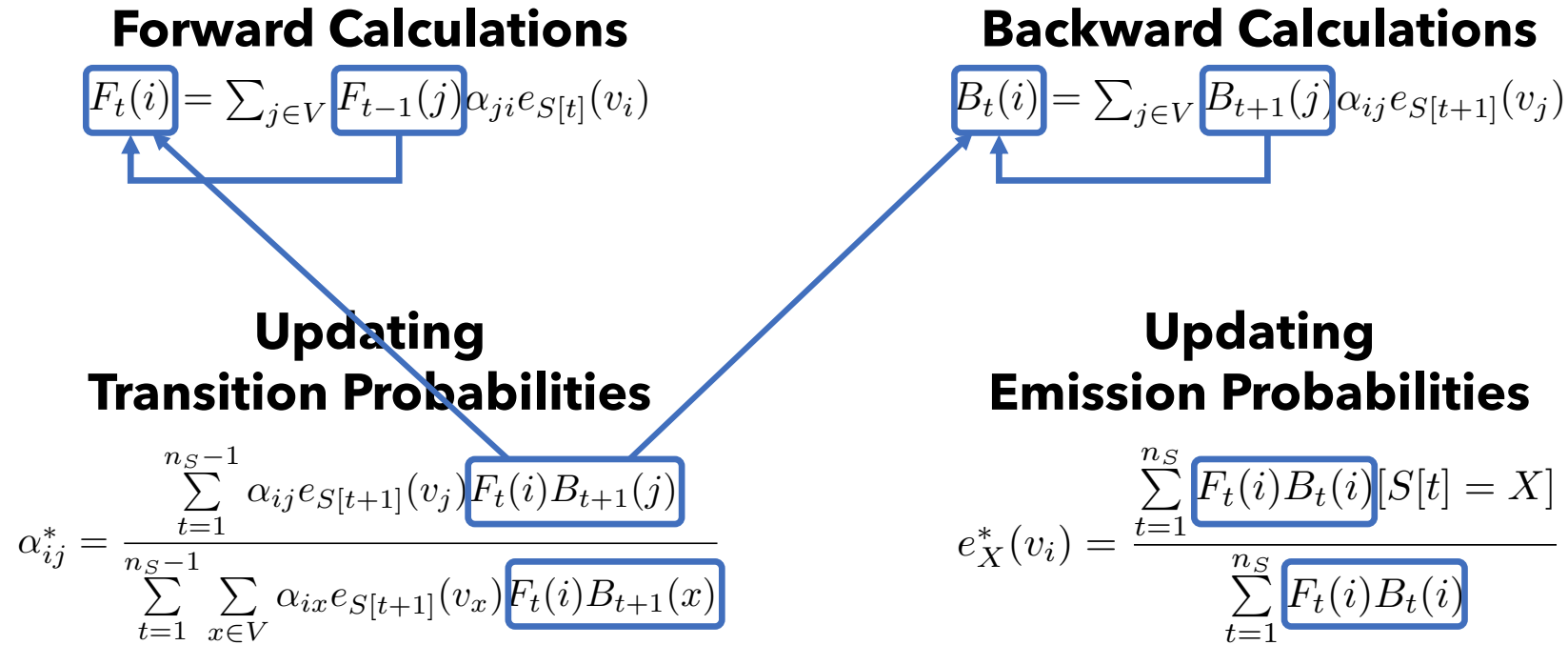


# A Common Graph Structure in Genomics

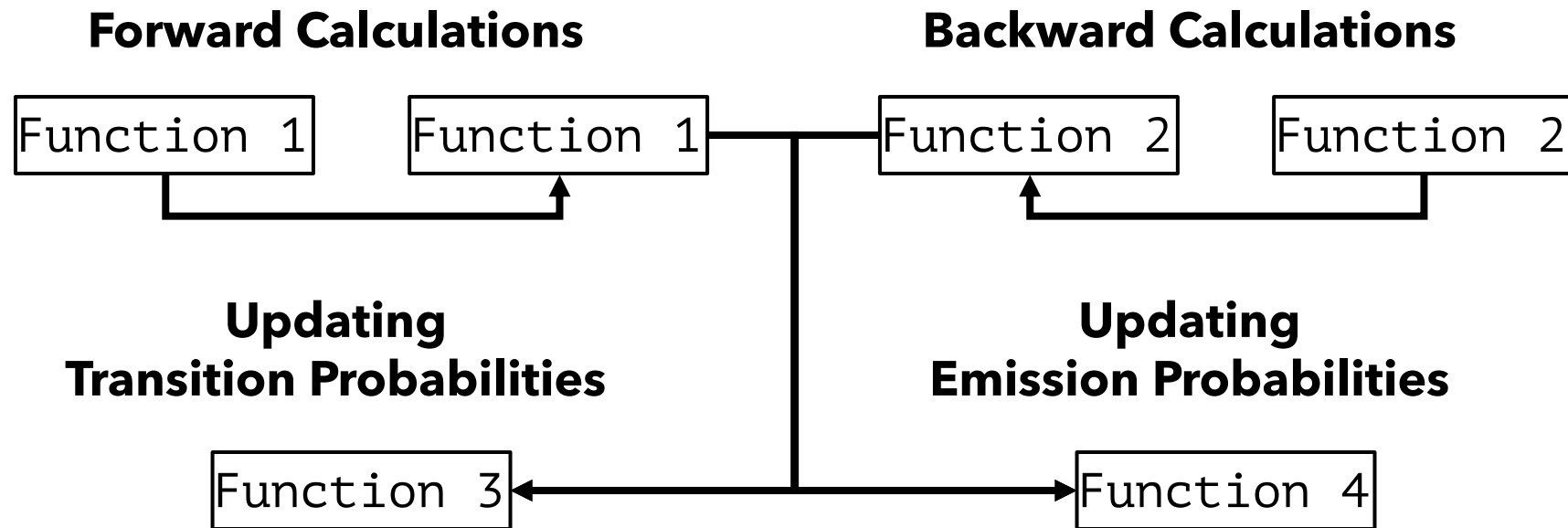
Profile Hidden Markov Models (PHMMs):  
A probabilistic graph structure



# Underlying ML Technique in Genomic Graphs

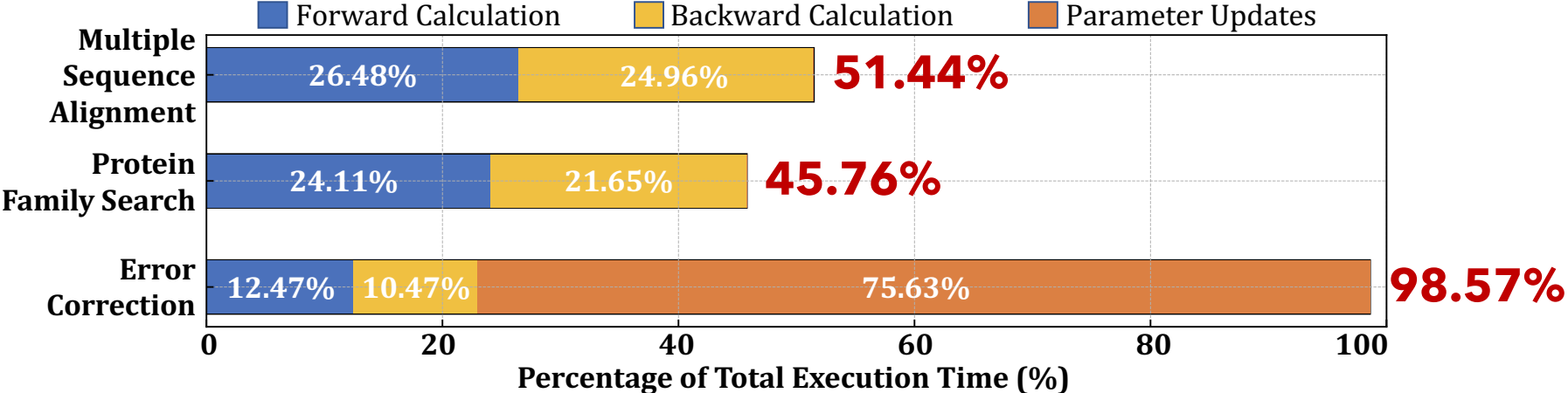


# Underlying ML Technique in Genomic Graphs



# Costly Use of ML in Genomic Graphs

## Cost of Using Machine Learning (ML) in pHMMs:



# Goal

Enable **rapid and power-efficient** use of the underlying ML technique in genome graphs

# Key Approach: HW-SW Co-Design

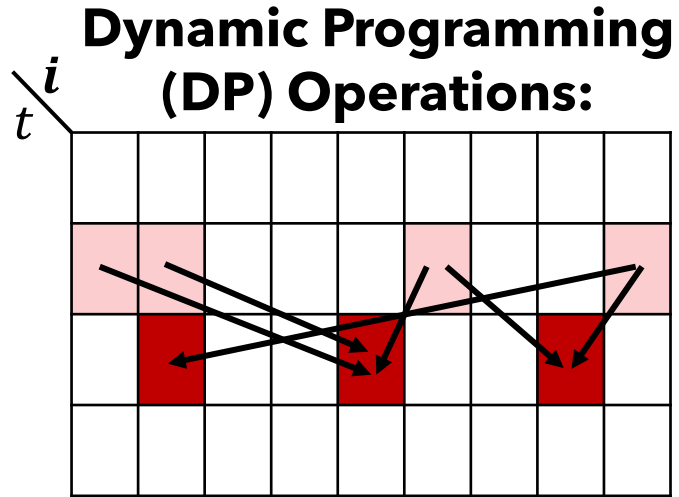
- **Reduce redundant data storage** by utilizing the fixed data pattern
- **Reduce unnecessary computations** with quick **SW** filtering

- **Avoid repeated operations** by utilizing lookup tables

**HW**

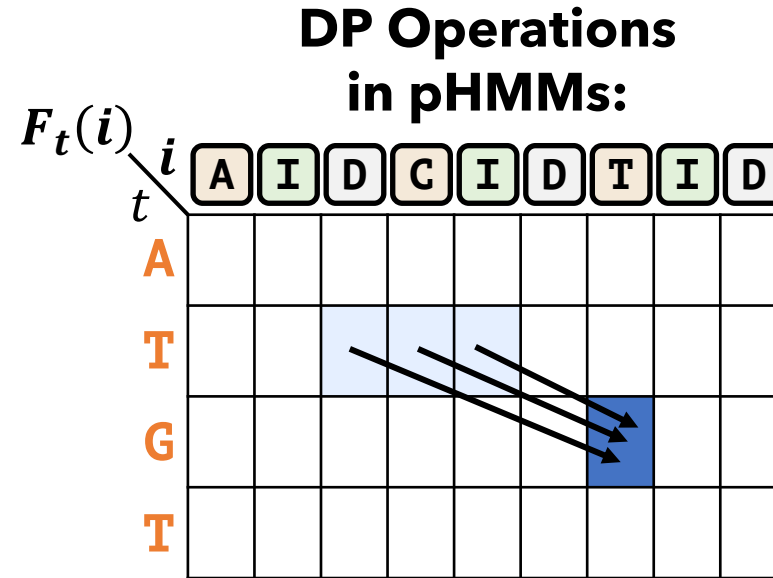
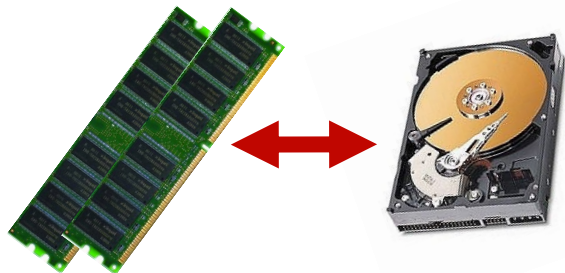
- **Reduce data movement** by processing data where it makes sense

# HW: Simple Data Pattern in DP Calculations

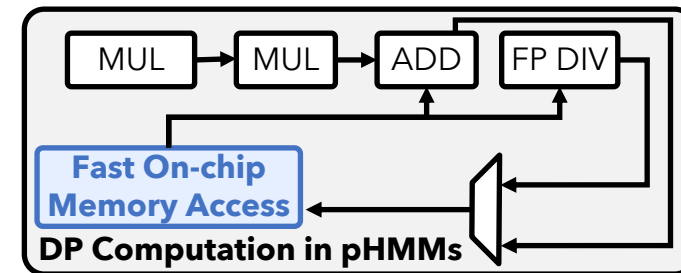


**Many Different Patterns**

**Frequent off-chip access**



**Simple Pattern**



**Store and access data**  
where it makes sense

# SW: Reducing Unnecessary Computations

**Observation:**  
Calculating the entire DP  
is not needed

$F_t(i)$

$t \backslash i$	A	I	D	G	I	D	T	I	D
A									
T									
G									
T									

**Dense DP Calculations**

**Filter by  
sorting**



$F_t(i)$

$t \backslash i$	A	I	D	G	I	D	T	I	D
A									
T									
G									
T									

**Sparse (Filtered) DP Calculations**

**Challenge:**  
Sorting is costly in hardware

# SW: Reducing Unnecessary Computations

## Software co-design:

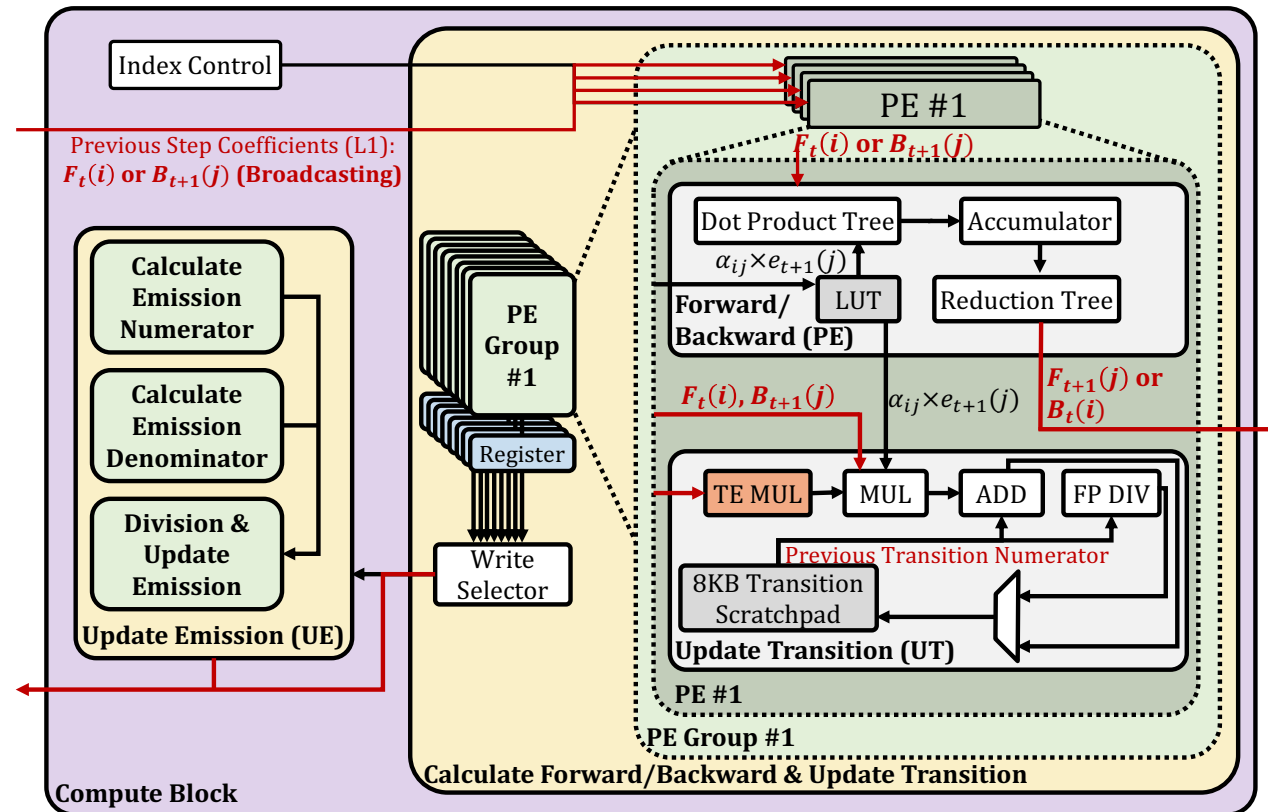
Replace sorting with a cheaper filtering

State IDs	Range
8, 9	1.00 - 0.94
10, 14	0.94 - 0.88
15, 16, 18	0.88 - 0.82
11, 20, 21, ...	0.82 - 0.76
13, 17, 19, ...	0.76 - 0.70
...	...
...	...
...	...
...	0.05 - 0.00

**Histogram Filter**

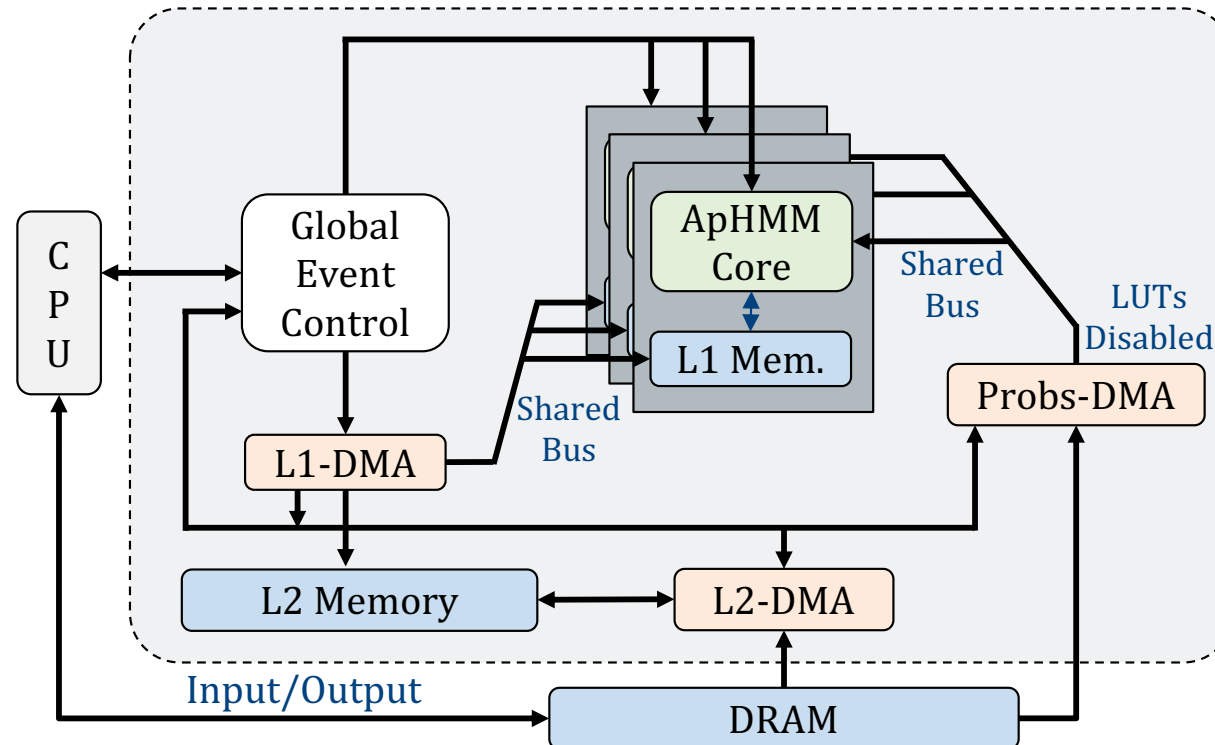
# Putting the Optimizations Together

Effectively implementing the hardware and software optimizations  
**to improve performance and reduce power requirements**

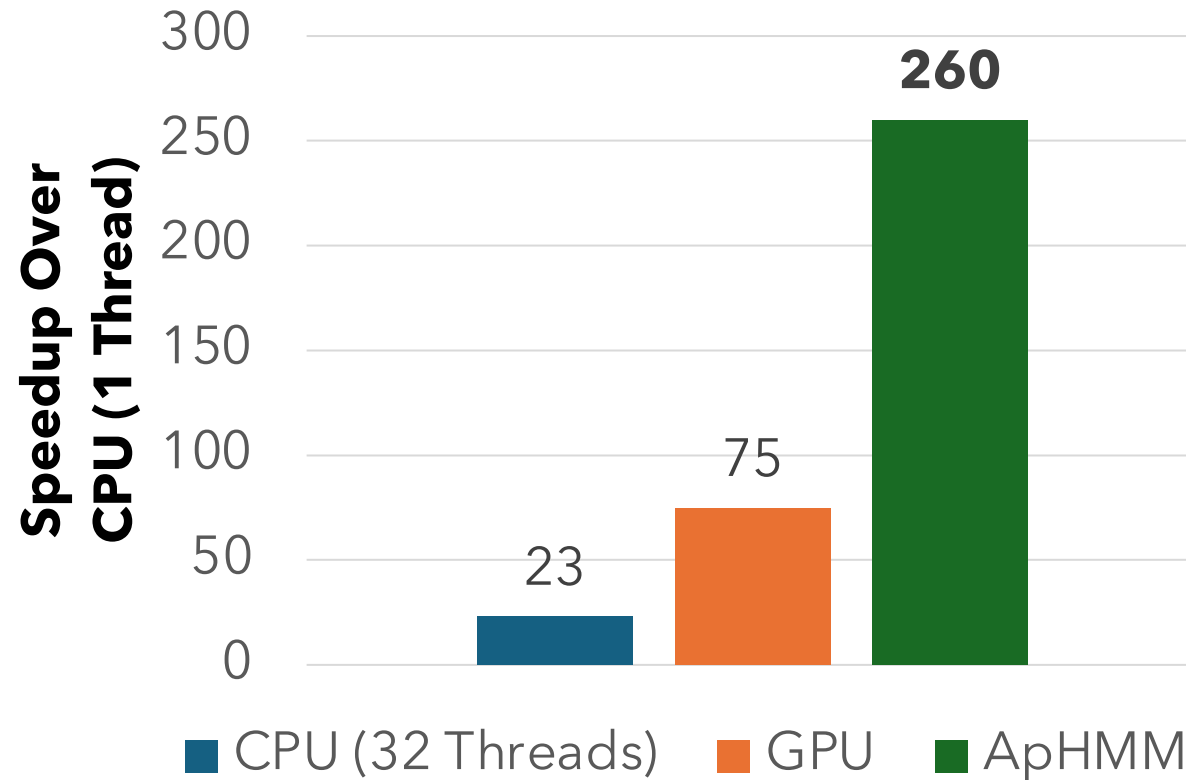


# Integrating ApHMM in a Complete System

The specialized task can be **offloaded**  
to an accelerator **whenever needed**

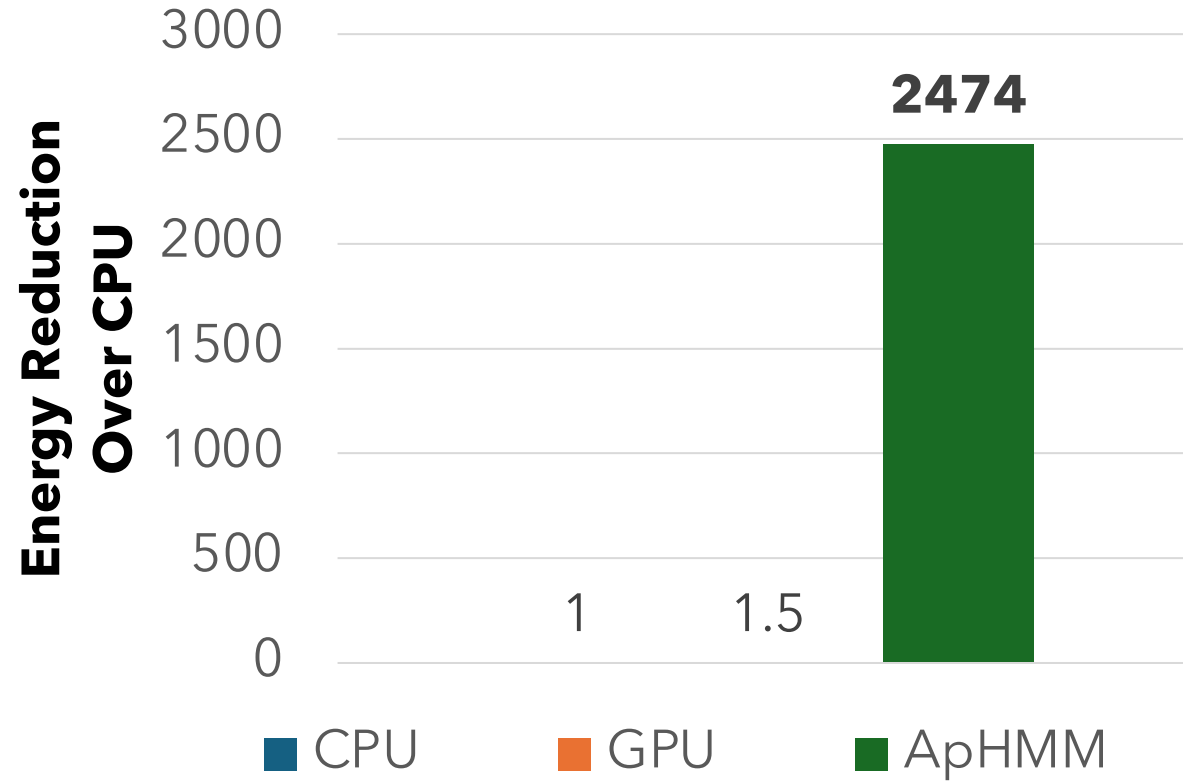


# Key Results: Performance



Specialized design enables computation  
**up to two orders of magnitude faster:**  
Better data flow with reduced latency and improved parallelism

# Key Results: Energy Consumption



...and a significant reduction in energy consumption  
**by up to three orders of magnitude:**  
simpler hardware with a minimized off-chip memory accesses  
and data movement