Fast, Efficient, and Low-Cost Computing Systems for Health and Life Sciences

Nika Mansouri Ghiasi

SAFARI

ETH zürich

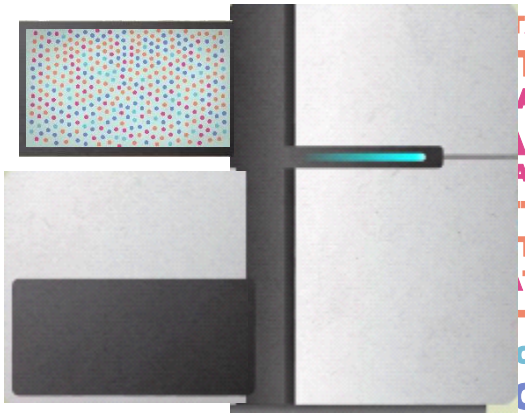
Generating and analyzing

large-scale biological data

provides transformative capabilities

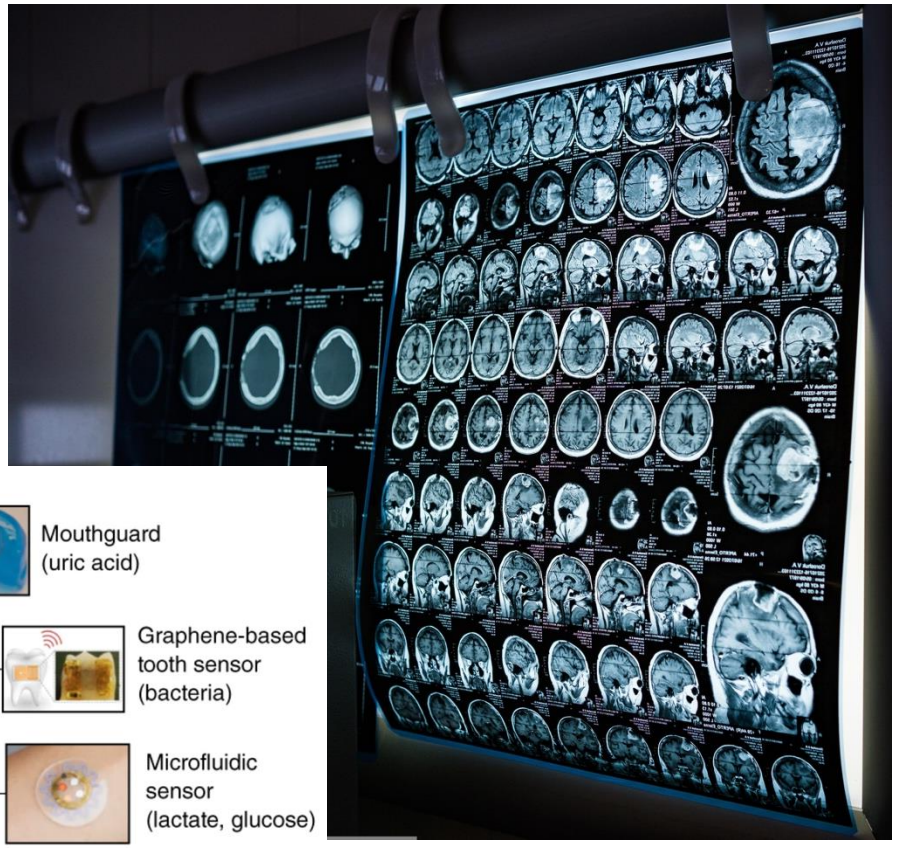
in healthcare and life sciences

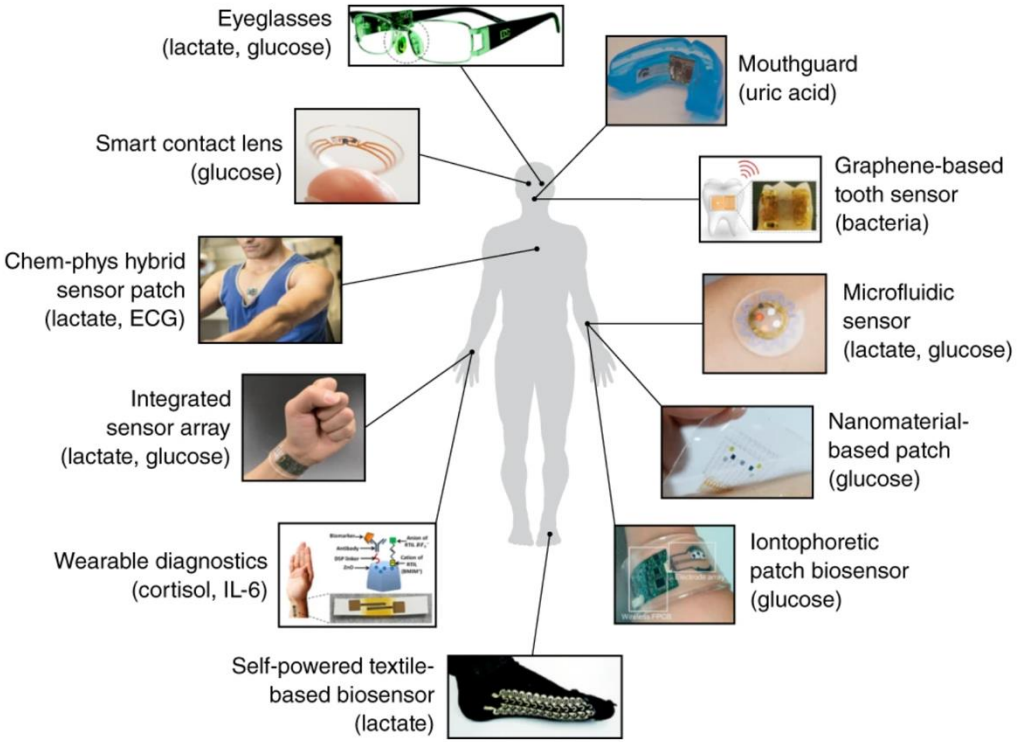
Biological Data



Billions of Short Reads

TATATATACGTACTAGTACGT
TTTAGTACGTACGT
ATACGTACTAGTACGT
CG CCCCTACGTA
ACGTACTAGTACGT
TTAGTACGTACGT
TACGTACTAAAGTACGT
TACGTACTAGTACGT
TTTAAAACGTA
CGTACTAGTACGT
GGGAGTACGTACGT





Kim et al, [Nature Biotech'19]

Transformative Capabilities:
Diagnosis in Complex and Urgent Settings

For critically ill infants:

- Reduces **morbidity**
- Reduces **hospital stay length** by 6%-69%
- Reduces **inpatient cost** by $800,000-$2,000,000

Farnaes+, "Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization", NPJ Genomic Medicine, 2018

Transformative Capabilities:
Personalized Treatments

For cancer:



Diagnosis **months (years) earlier**



Treatment **before the cancer spreads** and symptoms appear

Chen +, "Non-invasive early detection of cancer four years before conventional diagnosis using a blood test," *Nature Communications*

Transformative Capabilities:
Tracking Disease Outbreaks

For disease outbreaks

- Detecting new virus/bacteria variants in wastewater
- Containment before the outbreak spreads

Karthikeyan+, “Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission”, Nature, 2022

Transformative Capabilities

Article | Open Access | Published: 05 May 2020

Clinical utility of 24-h rapid trio-exome sequencing for critically ill infants

Huijun Wang, Yanyan Qian, Yulan Lu
Ping Zhang, Lin Yang, Bingbing Wu

npj Genomic Medicine 5, Article nu

Article | Open Access | Published: 04 April 2018

Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization

Lauge Farnaes, Amber Hildreth, Nathaly M. Sweeney, Michelle M. Clark, Shimul
H. Kaplan,
J. Gold, Yan
smore
le

nature biotechnology

ARTICLES

https://doi.org/10.1038/s41587-019-0156-5

Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection

Themoula Charalampous[1,8], Gemma L. Kay[1,2,8], Hollian Richardson[1,8], Alp Aydin[2], Rossella Baldan[1,3], Christopher Jeanes[4], Duncan Rae[4], Sara Grundy[4], Daniel J. Turner[5], John Wain[1,2],

The NEW ENGLAND JOURNAL of MEDICINE

CURRENT ISSUE SPECIALTIES TOPICS

ORIGINAL ARTICLE

100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care — Preliminary Report

RTICLES
://doi.org/10.1038/s41591-020-1072-4

nature medicine

Check for updates

Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer

Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study

Luke W Meredith*†, William L Hamilton*†, Ben Warne, Charlotte J Houldcroft, Myra Hosmillo, Aminu S Jahun, Martin D Curran, Surendra Parmar, Laura G Caller, Sarah L Caddy, Fahad A Khokhar, Anna Yakovleva, Grant Hall, Theresa Feltwell, Sally Forrest, Sushmita Sridhar, Michael P Weekes, Gordon Dougan,

Direct detection of early-stage cancers using circulating tumor DNA

Jillian Phallen,[1] Mark Sausen,[2] Vilmos Adleff,[1] Alessandro Leal,[1] Carolyn Hruban,[1] James White,[1] Valsamo Anagnostou,[1] Jacob Fiksel,[1] Stephen Cristiano,[1] Eniko Papp,[1]* Savannah Speir,[1] Thomas Reinert,[3] Mai-Britt Worm Orntoft,[3] Brian D. Woodward,[4] Derek Murphy,[2] Sonya Parpart-Li,[2] David Riley,[2] Monica Nesselbush,[2] Naomi Sengamalay,[2] Andrew Georgiadis,[2] Qing Kay Li,[1] Mogens Rørbæk Madsen,[5] Frank Viborg Mortensen,[6] Joost Huiskens,[7,8]
Meijer,[10] Hatim Husain,[4]
Torben Ørntoft,[3]
lescu[1‡]

nature communications

Article

https://doi.org/10.1038/s41467-0

SARS-CoV-2 genomic surveillance in wast water as a model for monitoring evolutio endemic viruses

Received: 16 February 2023

Accepted: 30 August 2023

Published online: 10 October 2023

Check for updates

Mukhlid Yousif[1,2], Said Rachida[1], Setshaba Taukobong[1], Nkosenhle
Chinwe Iwu-Jaja[1], Wayne Howard[1], Shelina Moonsamy[1], Nompilo
Sipho Gwala[1], Joshua I. Levy[3], Kristian G. Andersen[3], Cathrine Sch
Anne von Gottberg[5,6], Nicole Wolter[5,6], Jinal N. Bhiman[5,6],
Daniel Gyamfi Amoako[5], Arshad Ismail[7,8], Melinda Suchard[9] &
Kerrigan McCarthy[1,2]

Real-time, portable genome sequencing for Ebola surveillance

Joshua Quick[1]*, Nicholas J. Loman[1]*, Sophie Duraffour[2,3]*, Jared T. Simpson[4,5]*, Ettore Severi[6]*, Lauren Cowley[7]*, Joseph Akoi Bore[2], Raymond Koundouno[2], Gytis Dudas[8], Amy Mikhail[7], Nobila Ouédraogo[9], Babak Afrough[2,10], Amadou Bah[2,11], Jonathan H. J. Baum[2,3], Beate Becker-Ziaja[2,3], Jan Peter Boettcher[2,12], Mar Cabeza-Cabrerizo[2,3], Álvaro Camino-Sánchez[2], Lisa L. Carter[2,13], Juliane Doerrbecker[2,3], Theresa Enkirch[2,14], Isabel García-Dorival[2,15], Nicole Hetzelt[2,12], Julia Hinzmann[2,12], Tobias Holm[2,3], Liana Eleni Kafetzopoulou[2,16], Michel Koropogui[2,17], Abigael Kosgey[2,18], Eeva Kuisma[2,10], Christopher H. Logue[2,10], Antonio Mazzarelli[2,19], Sarah Meisel[2,3], Marc Mertens[2,20], Janine Michel[2,12], Didier Ngabo[2,10], Katja Nitzsche[2,3], Elisa Pallasch[2,3], Livia Victoria Patrono[2,3], Jasmine Portmann[2,21], Johanna Gabriella Repits[2,22], Natasha Y. Rickett[2,15,23], Andreas Sachse[2,12], Katrin Singethan[2,24], Inês Vitoriano[2,10], Rahel L. Yemanaberhan[2,3], Elsa G. Zekeng[2,15,23], Trina Racine[25], Alexander Bello[25], Amadou Alpha Sall[26], Ousmane Faye[26], Oumar Faye[26], N'Faly Magassouba[27], Cecelia V. Williams[28,29], Victoria Amburgey[28,29], Linda Winona[28,29], Emily Davis[29,30], Jon Gerlach[29,30], Frank Washington[29,30], Vanessa Monteil[31], Marine Jourdain[31], Marion Bererd[31], Alimou Camara[31], Hermann Somlare[31], Abdoulaye Camara[31], Marianne Gerard[31], Guillaume Bado[31], Bernard Baillet[31], Déborah Delaune[32,33], Koumpingnin Yacouba Nebie[34], Abdoulaye Diarra[34], Yacouba Savane[34], Raymond Bernard Pallawo[34], Giovanna Jaramillo Gutierrez[35], Natacha Milhano[6,36], Isabelle Roger[34], Christopher J. Williams[6,37], Facinet Yattara[17], Kuiama Lewandowski[10], James Taylor[38], Phillip Rachwal[38], Daniel J. Turner[39], Georgios Pollakis[15,23], Julian A. Hiscox[15,23], David A. Matthews[40], Matthew K. O'Shea[41], Andrew McD. Johnston[41], Duncan Wilson[41], Emma Hutley[42], Erasmus Smit[43], Antonino Di Caro[2,19], Roman Wölfel[2,44], Kilian Stoecker[2,44], Erna Fleischmann[2,44], Martin Gabriel[2,3], Simon A. Weller[38], Lamine Koivogui[45], Boubacar Diallo[34], Sakoba Keïta[17], Andrew Rambaut[8,46,47], Pierre Formenty[34], Stephan Günther[2,3] & Miles W. Carroll[2,10,48,49]

But in reality, can we provide these capabilities to…

- All patients in critical care units?
- Everyone, to be screened for cancer before symptoms arise?
- The whole population, to prevent a new pandemic?


GOODS CARRIER

Are we there yet?

No

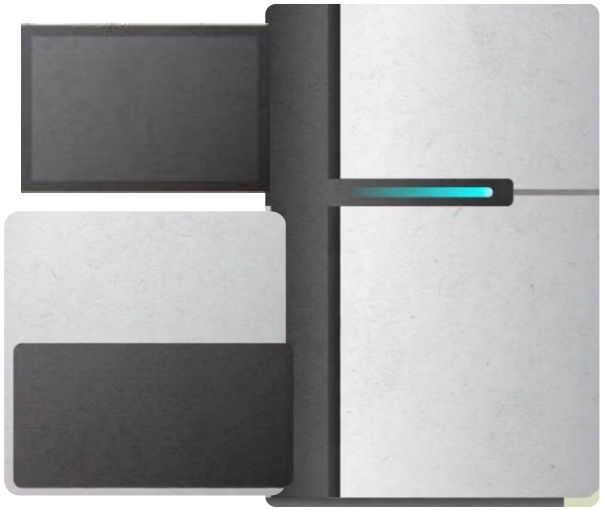
Generating and **analyzing**

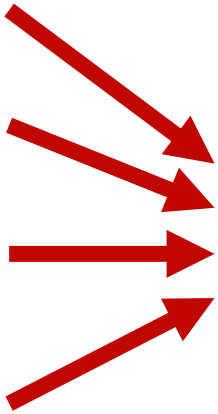
large-scale biological data

provides transformative capabilities

in healthcare and life sciences

Generating and analyzing







Special-Purpose

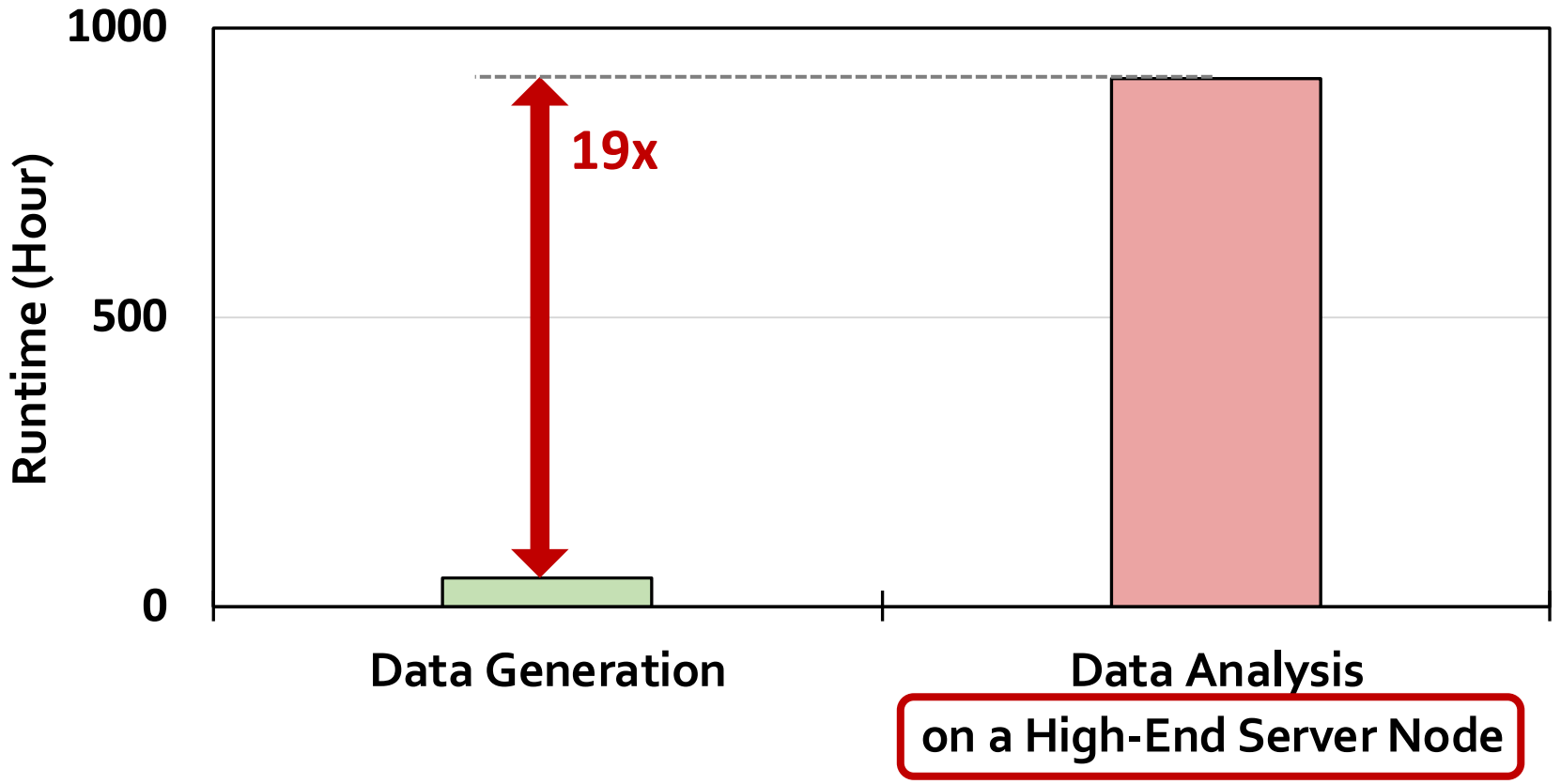
Fast & Efficient

General-Purpose

Slow & Inefficient

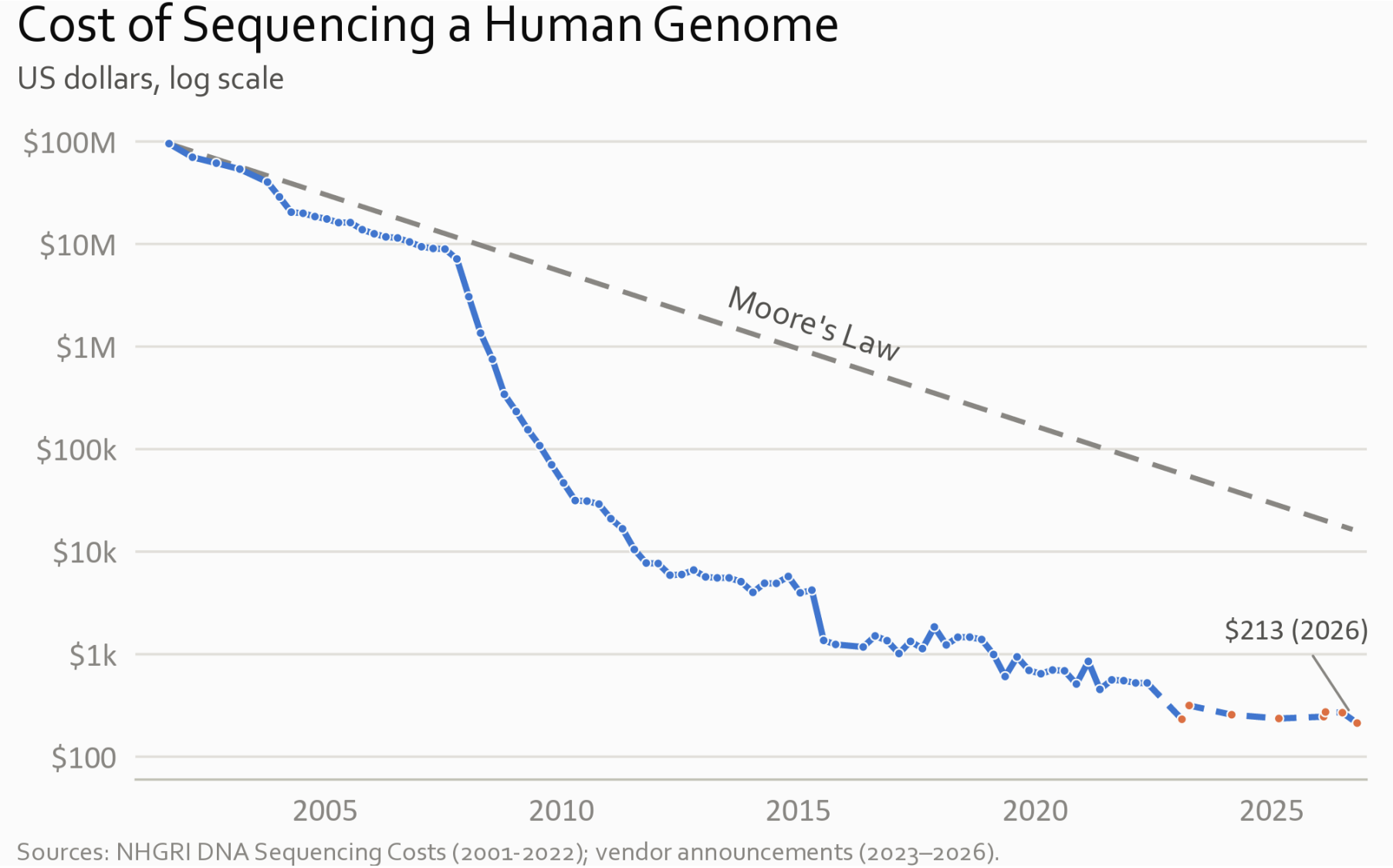
Example

Time to generate vs. time to analyze a metagenomic sample



Merely scaling conventional computing systems is cost- and energy-inefficient

While Data Generation Cost Keeps Dropping…



Goal

Provide transformative capabilities

in healthcare and life sciences

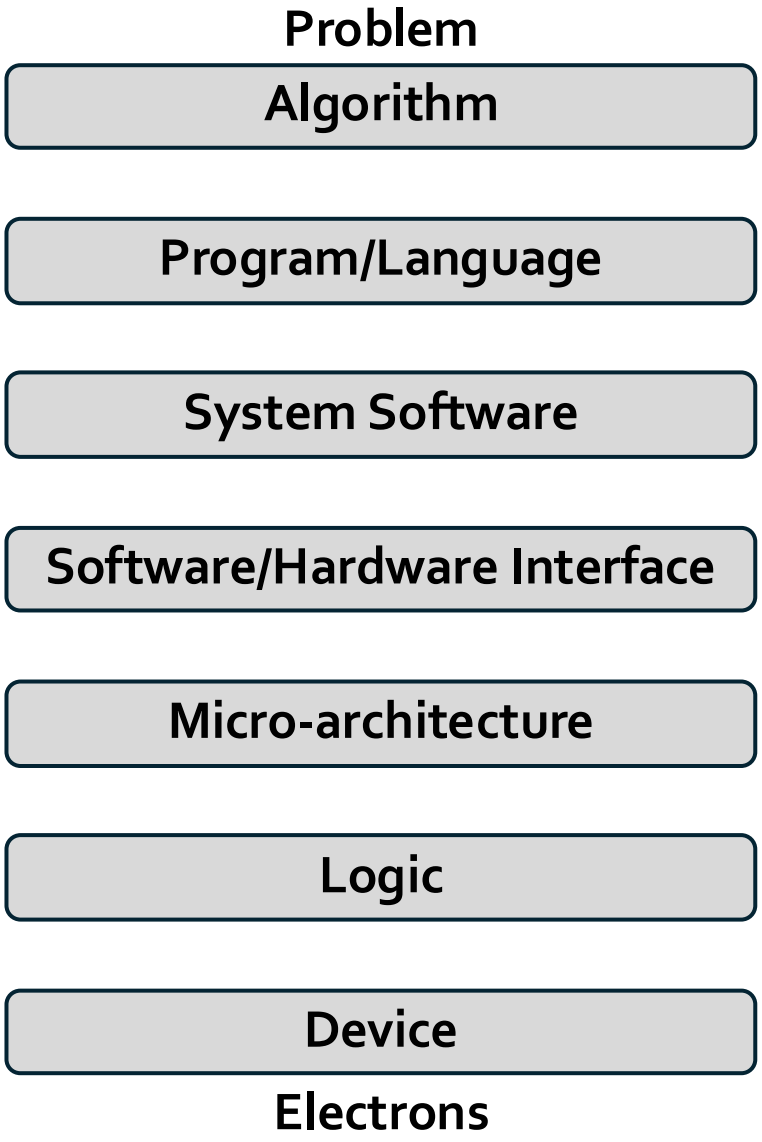
across the population

by designing

fast, efficient, and low-cost computing systems

for analyzing large-scale biological data

Computing Stack



Computing Stack

Problem

Algorithm

Program/Language

System Software

Software/Hardware Interface

Micro-architecture

Logic

Device

Electrons

Computing Stack

Problem

Algorithm

Program/Language

System Software

Software/Hardware Interface

Micro-architecture

Logic

Device

Electrons

Computing Stack

Problem

Algorithm

Program/Language

System Software

Software/Hardware Interface

Micro-architecture

Logic

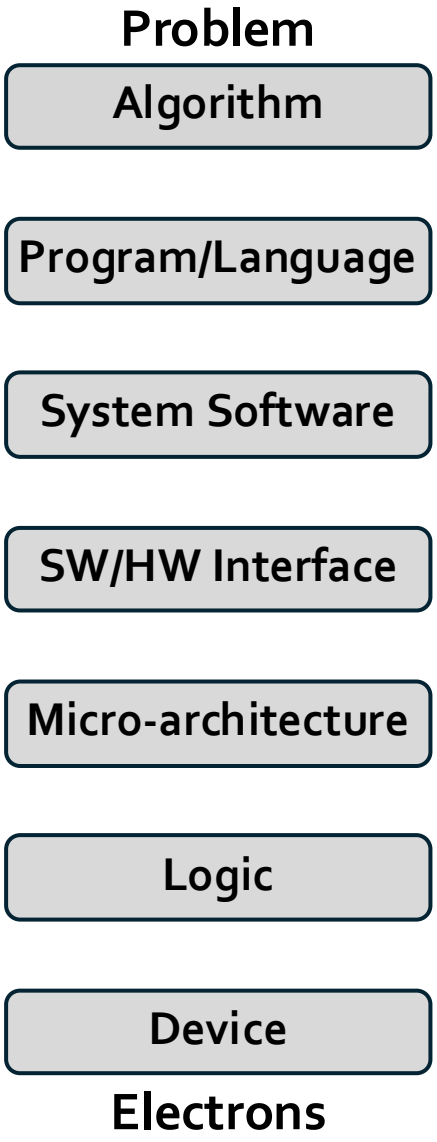
Device

Electrons



My Journey: Across the Computing Stack for Health, Life Sciences, and Beyond

https://nikamgh.github.io



My Journey: Across the Computing Stack for Health, Life Sciences, and Beyond

https://nikamgh.github.io

Problem

Algorithm

Program/Language

System Software

SW/HW Interface

Micro-architecture

Logic

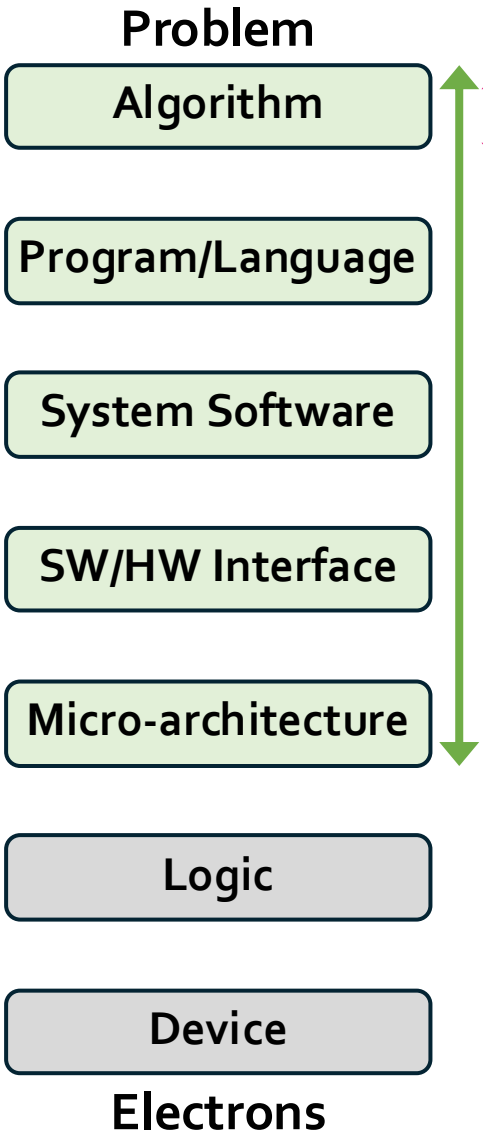
Device

Electrons

❶ ***High-Performance Algorithms*** *[Nature Genetics'26], [Bioinformatics'26], [Nature Reviews'25], [ISMB'24], [IEEE Access'24], [APBC'23], [NAR'23]*

My Journey: Across the Computing Stack for Health, Life Sciences, and Beyond

https://nikamgh.github.io



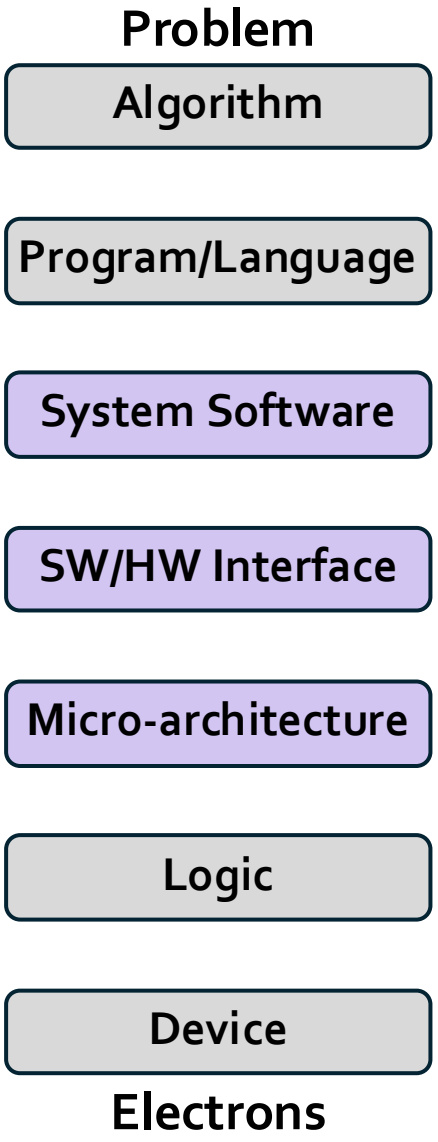
1 ***High-Performance Algorithms*** *[Nature Genetics'26], [Bioinformatics'26], [Nature Reviews'25], [ISMB'24], [IEEE Access'24], [APBC'23], [NAR'23]*

2 ***Data-Centric Computing for Performance, Energy-Efficiency, Cost-Effectiveness, and Security*** *[ISCA'26], [HPCA'26], [ICS Workshops'26], [ISCA'25], [DSN'25], [ICS'25], [ISCA'24], [IEEE Access'24], [IEEE TETC'22], [ASPLOS'22], [ASPLOS'21], [Under Submission]*

3 ***Other Algorithm-Architecture Co-Designs*** *[HPCA'26], [ISCA'25], [HPCA'25], [Bioinformatics'23], [ISCA'22], [MICRO'19], [Under Submission]*

My Journey: Across the Computing Stack for Health, Life Sciences, and Beyond

https://nikamgh.github.io



1 ***High-Performance Algorithms*** *[Nature Genetics'26], [Bioinformatics'26], [Nature Reviews'25], [ISMB'24], [IEEE Access'24], [APBC'23], [NAR'23]*

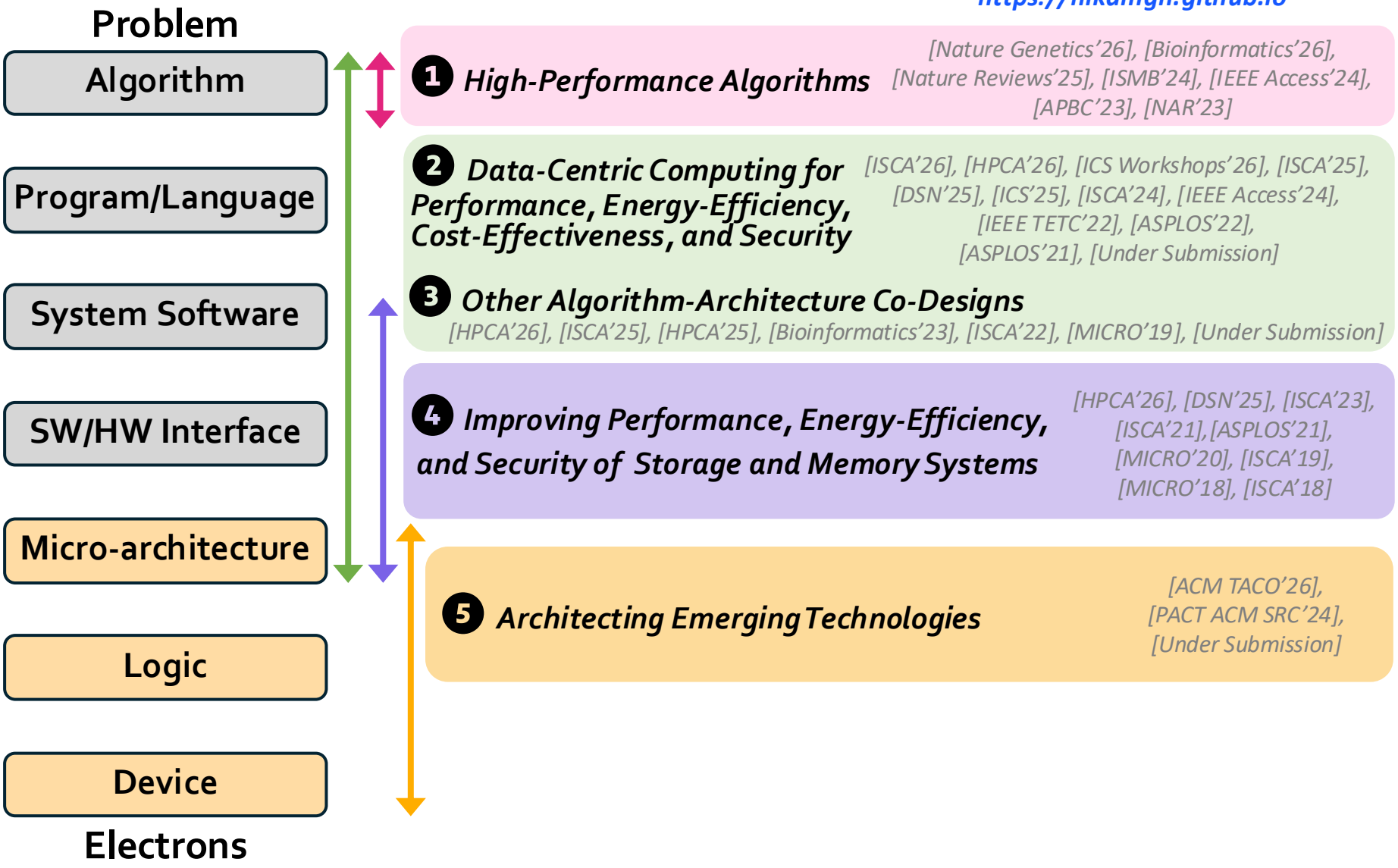
2 ***Data-Centric Computing for Performance, Energy-Efficiency, Cost-Effectiveness, and Security*** *[ISCA'26], [HPCA'26], [ICS Workshops'26], [ISCA'25], [DSN'25], [ICS'25], [ISCA'24], [IEEE Access'24], [IEEE TETC'22], [ASPLOS'22], [ASPLOS'21], [Under Submission]*

3 ***Other Algorithm-Architecture Co-Designs*** *[HPCA'26], [ISCA'25], [HPCA'25], [Bioinformatics'23], [ISCA'22], [MICRO'19], [Under Submission]*

4 ***Improving Performance, Energy-Efficiency, and Security of Storage and Memory Systems*** *[HPCA'26], [DSN'25], [ISCA'23], [ISCA'21], [ASPLOS'21], [MICRO'20], [ISCA'19], [MICRO'18], [ISCA'18]*

My Journey: Across the Computing Stack for Health, Life Sciences, and Beyond

https://nikamgh.github.io



My Journey: Across the Computing Stack for Health, Life Sciences, and Beyond

https://nikamgh.github.io

Problem

- Algorithm
- Program/Language
- System Software
- SW/HW Interface
- Micro-architecture
- Logic
- Device

Electrons

1 ***High-Performance Algorithms*** — *[Nature Genetics'26], [Bioinformatics'26], [Nature Reviews'25], [ISMB'24], [IEEE Access'24], [APBC'23], [NAR'23]*

2 ***Data-Centric Computing for Performance, Energy-Efficiency, Cost-Effectiveness, and Security*** — *[ISCA'26], [HPCA'26], [ICS Workshops'26], [ISCA'25], [DSN'25], [ICS'25], [ISCA'24], [IEEE Access'24], [IEEE TETC'22], [ASPLOS'22], [ASPLOS'21], [Under Submission]*

3 ***Other Algorithm-Architecture Co-Designs*** — *[HPCA'26], [ISCA'25], [HPCA'25], [Bioinformatics'23], [ISCA'22], [MICRO'19], [Under Submission]*

4 ***Improving Performance, Energy-Efficiency, and Security of Storage and Memory Systems*** — *[HPCA'26], [DSN'25], [ISCA'23], [ISCA'21], [ASPLOS'21], [MICRO'20], [ISCA'19], [MICRO'18], [ISCA'18]*

5 ***Architecting Emerging Technologies*** — *[ACM TACO'26], [PACT ACM SRC'24], [Under Submission]*



6 ***Architecture for Health (Arch4Health) Initiative*** — ***Workshops in*** *[MICRO'26], [SOSP'26], [ICS'26], [HPCA'26], [MICRO'26]*

My Journey: Across the Computing Stack for Health, Life Sciences, and Beyond

https://nikamgh.github.io

Problem

- Algorithm
- Program/Language
- System Software
- SW/HW Interface
- Micro-architecture
- Logic
- Device

Electrons

1 ***High-Performance Algorithms*** — *[Nature Genetics'26], [Bioinformatics'26], [Nature Reviews'25], [ISMB'24], [IEEE Access'24], [APBC'23], [NAR'23]*

2 ***Data-Centric Computing for Performance, Energy-Efficiency, Cost-Effectiveness, and Security*** — *[ISCA'26], [HPCA'26], [ICS Workshops'26], [ISCA'25], [DSN'25], [ICS'25], [ISCA'24], [IEEE Access'24], [IEEE TETC'22], [ASPLOS'22], [ASPLOS'21], [Under Submission]*

3 ***Other Algorithm-Architecture Co-Designs*** — *[HPCA'26], [ISCA'25], [HPCA'25], [Bioinformatics'23], [ISCA'22], [MICRO'19], [Under Submission]*

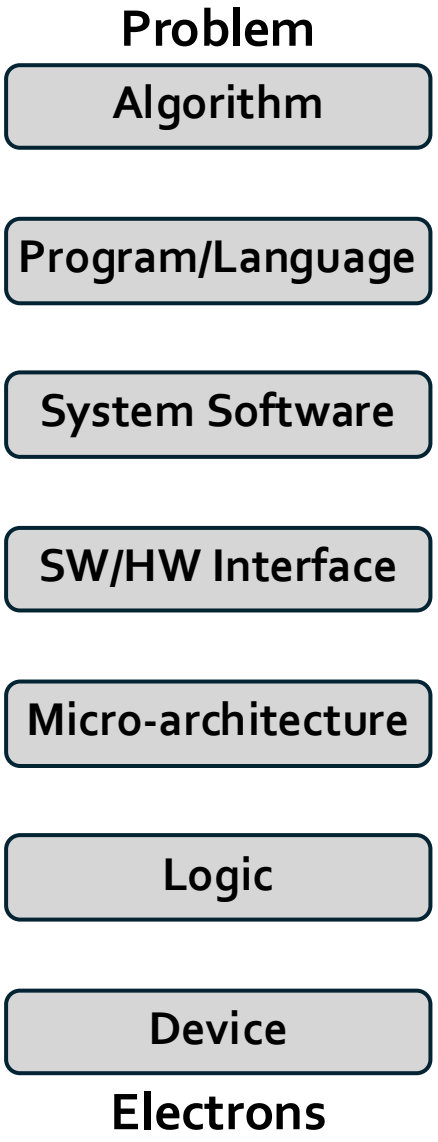
4 ***Improving Performance, Energy-Efficiency, and Security of Storage and Memory Systems*** — *[HPCA'26], [DSN'25], [ISCA'23], [ISCA'21], [ASPLOS'21], [MICRO'20], [ISCA'19], [MICRO'18], [ISCA'18]*

5 ***Architecting Emerging Technologies*** — *[ACM TACO'26], [PACT ACM SRC'24], [Under Submission]*

6 ***Architecture for Health (Arch4Health) Initiative*** — ***Workshops in*** *[MICRO'26], [SOSP'26], [ICS'26], [HPCA'26], [MICRO'26]*

My Journey: Across the Computing Stack for Health, Life Sciences, and Beyond

https://nikamgh.github.io



1 *High-Performance Algorithms* [Nature Genetics'26], [Bioinformatics'26], [Nature Reviews'25], [ISMB'24], [IEEE Access'24], [APBC'23], [NAR'23]

2 ***Data-Centric Computing for Performance, Energy-Efficiency, Cost-Effectiveness, and Security*** [ISCA'26], [HPCA'26], [ICS Workshops'26], [ISCA'25], [DSN'25], [ICS'25], [ISCA'24], [IEEE Access'24], [IEEE TETC'22], [ASPLOS'22], [ASPLOS'21], [Under Submission]

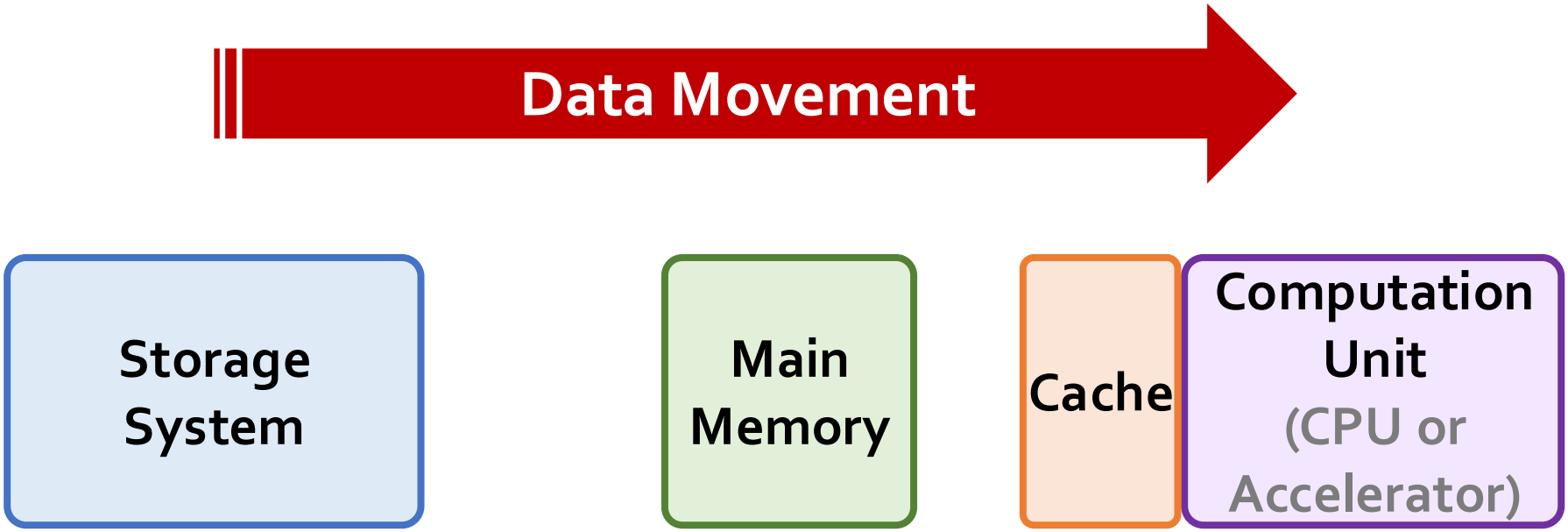
3 ***Other Algorithm-Architecture Co-Designs*** [HPCA'26], [ISCA'25], [HPCA'25], [Bioinformatics'23], [ISCA'22], [MICRO'19], [Under Submission]

4 ***Improving Performance, Energy-Efficiency, and Security of Storage and Memory Systems*** [HPCA'26], [DSN'25], [ISCA'23], [ISCA'21], [ASPLOS'21], [MICRO'20], [ISCA'19], [MICRO'18], [ISCA'18]

5 *Architecting Emerging Technologies* [ACM TACO'26], [PACT ACM SRC'24], [Under Submission]

6 *Architecture for Health (Arch4Health) Initiative* Workshops in [MICRO'26], [SOSP'26], [ICS'26], [HPCA'26], [MICRO'26]

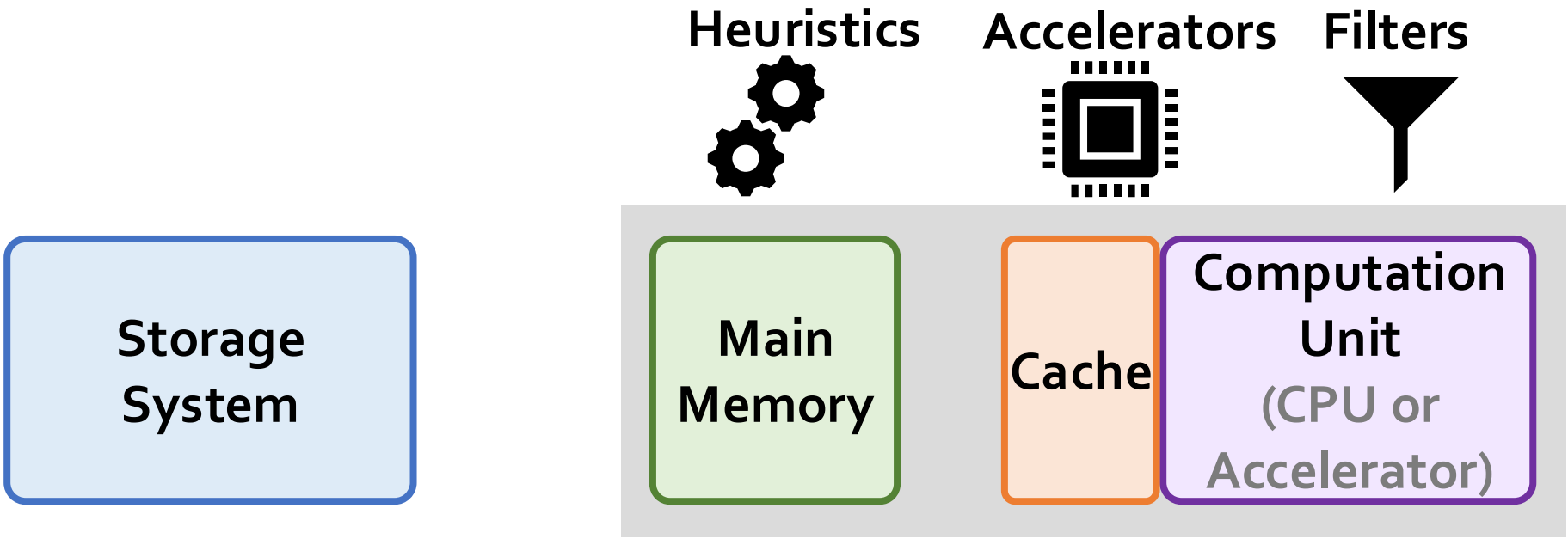
Overheads of (Meta)Genomic Analyses



✕ Computation overhead

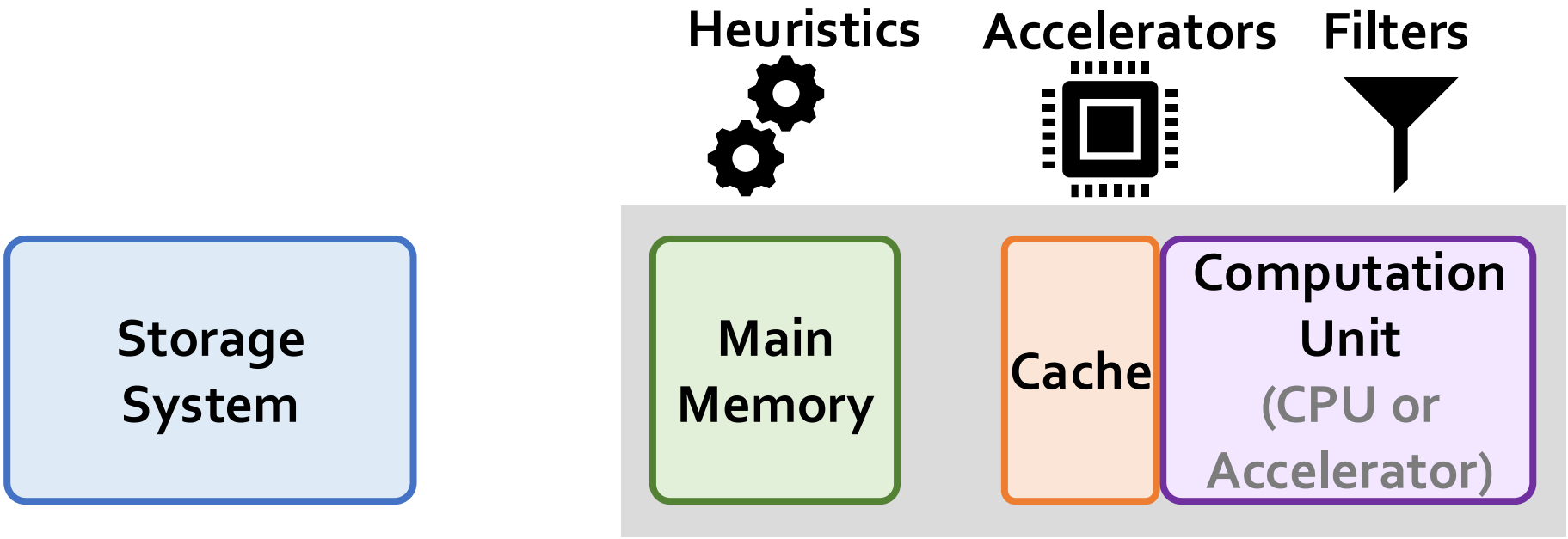
✕ Data movement and preparation overhead

Accelerating (Meta)Genome Sequence Analysis



✓ Computation overhead

Accelerating (Meta)Genome Sequence Analysis



✓ Computation overhead

✗ Data movement and preparation overhead

Our Approach

*Designing customized **storage-centric systems** that efficiently*

① ***analyze** genomic and metagenomic data **inside the storage system***

② *enable **highly-compressed storage** and **high-performance access** of large-scale sequence data*

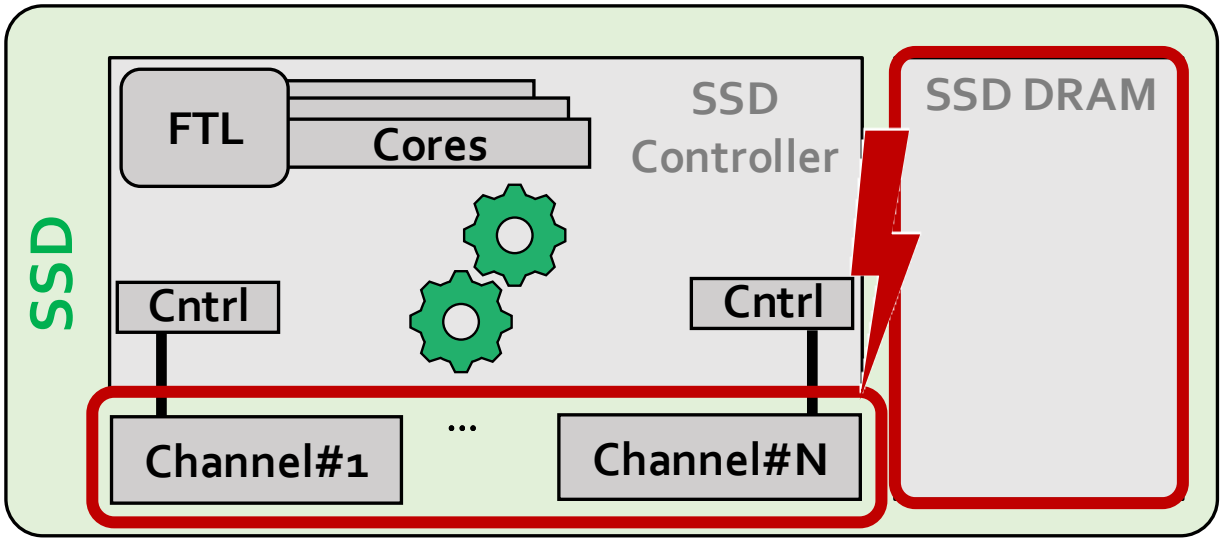
we can alleviate the overheads of

data movement *computation* *data preparation*

Challenges of Storage-Centric Computing

(Meta)genomic analysis tools cannot run in-storage due to SSD limits

- Long **latency of NAND flash** chips
- Limited **DRAM capacity** inside the SSD
- Limited **DRAM bandwidth** inside the SSD



Addressing Challenges

New storage-aware HW/SW co-designs

inspired by

the key properties of genomic data
(i.e., impacted by genetic variations & sequencing devices)

Our Approach

*Designing customized **storage-centric systems** that efficiently*

① ***analyze** genomic and metagenomic data **inside the storage system***

1 GenStore [ASPLOS'22]
In-storage filtering
for genomic analysis

2 MegIS [ISCA'24]
Cooperative in-storage processing
for metagenomic analysis

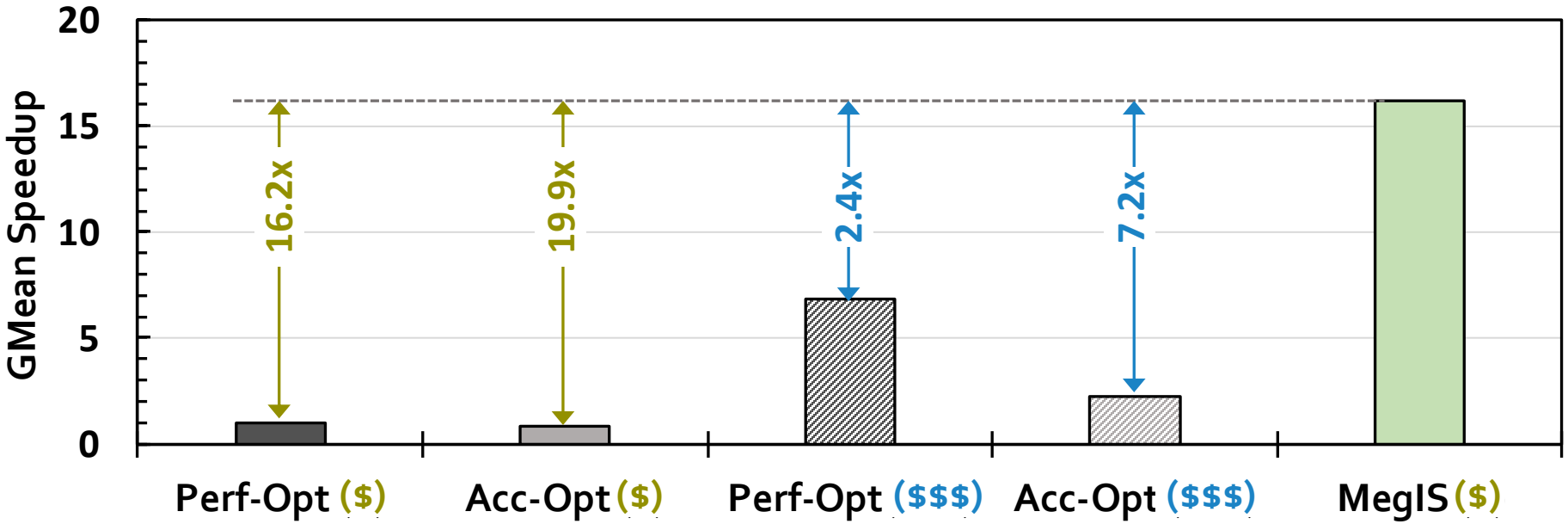
3 GRAINS [ISCA'26]
Storage-aware algorithm-architecture co-design
for graph-based (meta)genomic analyses

② *enable **highly-compressed storage** and **high-performance access** of large-scale sequence data*

4 SAGe [HPCA'26]
Algorithm-architecture co-design
for storing and accessing sequence data

System Cost-Efficiency

- **Cost-optimized system ($):** With SSD-C and 64-GB host DRAM
- **Performance-optimized system ($$$):** With SSD-P and 1-TB host DRAM



MegIS outperforms the baselines

***even* when running on a much less costly system**

My Journey: Across the Computing Stack for Health, Life Sciences, and Beyond

https://nikamgh.github.io

Problem

- Algorithm
- Program/Language
- System Software
- SW/HW Interface
- Micro-architecture
- Logic
- Device

Electrons

1 ***High-Performance Algorithms*** — *[Nature Genetics'26], [Bioinformatics'26], [Nature Reviews'25], [ISMB'24], [IEEE Access'24], [APBC'23], [NAR'23]*

2 ***Data-Centric Computing for Performance, Energy-Efficiency, Cost-Effectiveness, and Security*** — *[ISCA'26], [HPCA'26], [ICS Workshops'26], [ISCA'25], [DSN'25], [ICS'25], [ISCA'24], [IEEE Access'24], [IEEE TETC'22], [ASPLOS'22], [ASPLOS'21], [Under Submission]*

3 ***Other Algorithm-Architecture Co-Designs*** — *[HPCA'26], [ISCA'25], [HPCA'25], [Bioinformatics'23], [ISCA'22], [MICRO'19], [Under Submission]*

4 ***Improving Performance, Energy-Efficiency, and Security of Storage and Memory Systems*** — *[HPCA'26], [DSN'25], [ISCA'23], [ISCA'21], [ASPLOS'21], [MICRO'20], [ISCA'19], [MICRO'18], [ISCA'18]*

5 ***Architecting Emerging Technologies*** — *[ACM TACO'26], [PACT ACM SRC'24], [Under Submission]*



6 ***Architecture for Health (Arch4Health) Initiative*** — ***Workshops in*** *[MICRO'26], [SOSP'26], [ICS'26], [HPCA'26], [MICRO'26]*

Are we there yet?

No...

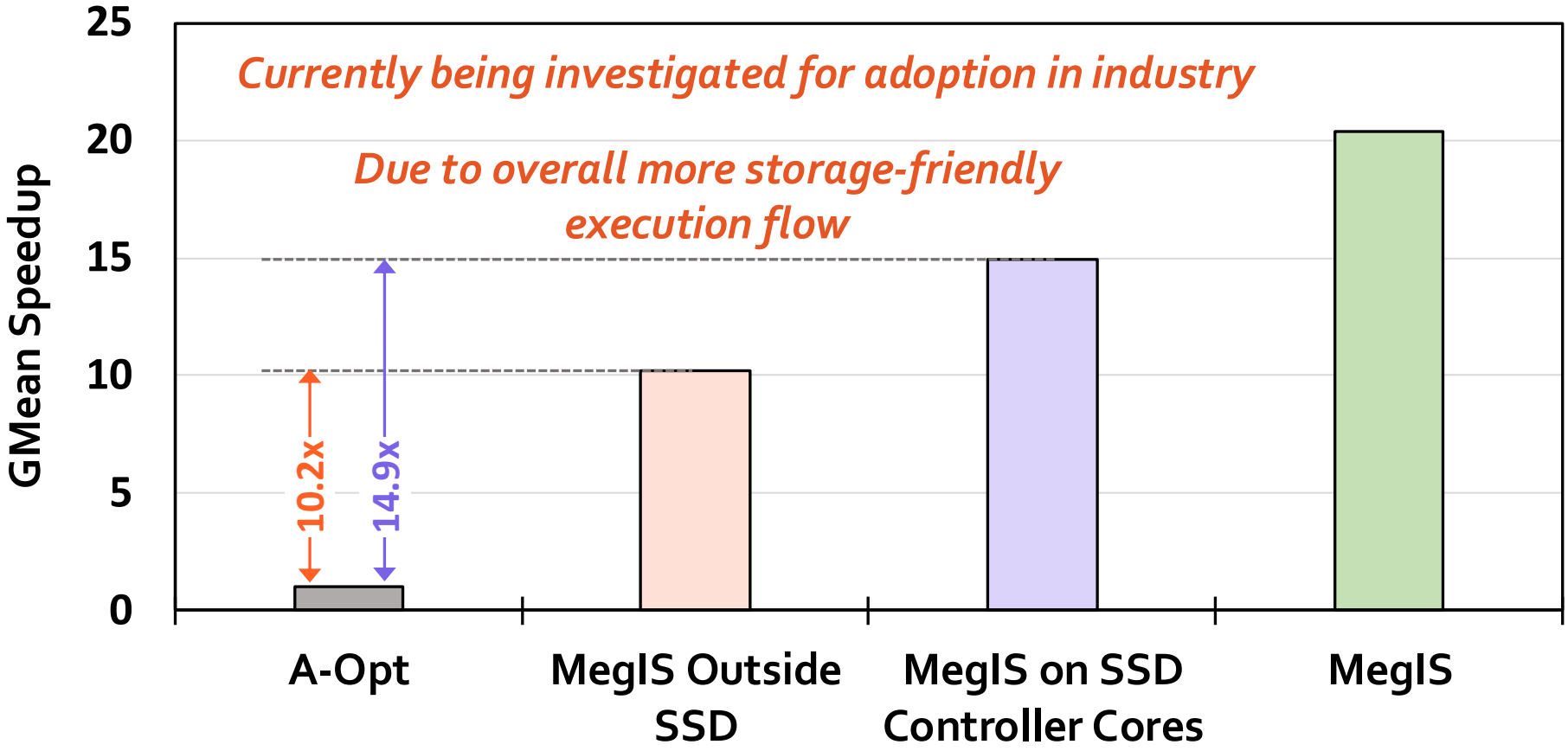
But getting closer!

How can we get there?

Next Steps

1. Facilitating wider adoption

Example: Adopting MegIS



Execution on general-purpose cores facilitates near-term adoption

Facilitating Wider Adoption

Data-centric computing
(particularly storage-centric computing)

provides large benefits in

analyzing **large-scale biological data**

(& other data in other data-intensive domains such as AI/ML)

We need to facilitate wider adoption

Facilitating Wider Adoption

① Security and privacy

The devices holding the data must now be trusted to compute on it

② End-to-end system support

Compute-capable devices deliver their full benefits only when the entire computing stack exposes their capabilities

③ Programmability

Models, abstractions, and benchmarks that enable both human programmers and AI coding agents to target data-centric systems

Next Steps

1. Facilitating wider adoption

2. Increase reach to other critical problems and settings

Increase Reach

Other critical problems



Mental health support



Single-cell analysis

Other critical settings



Portable, in-field analysis

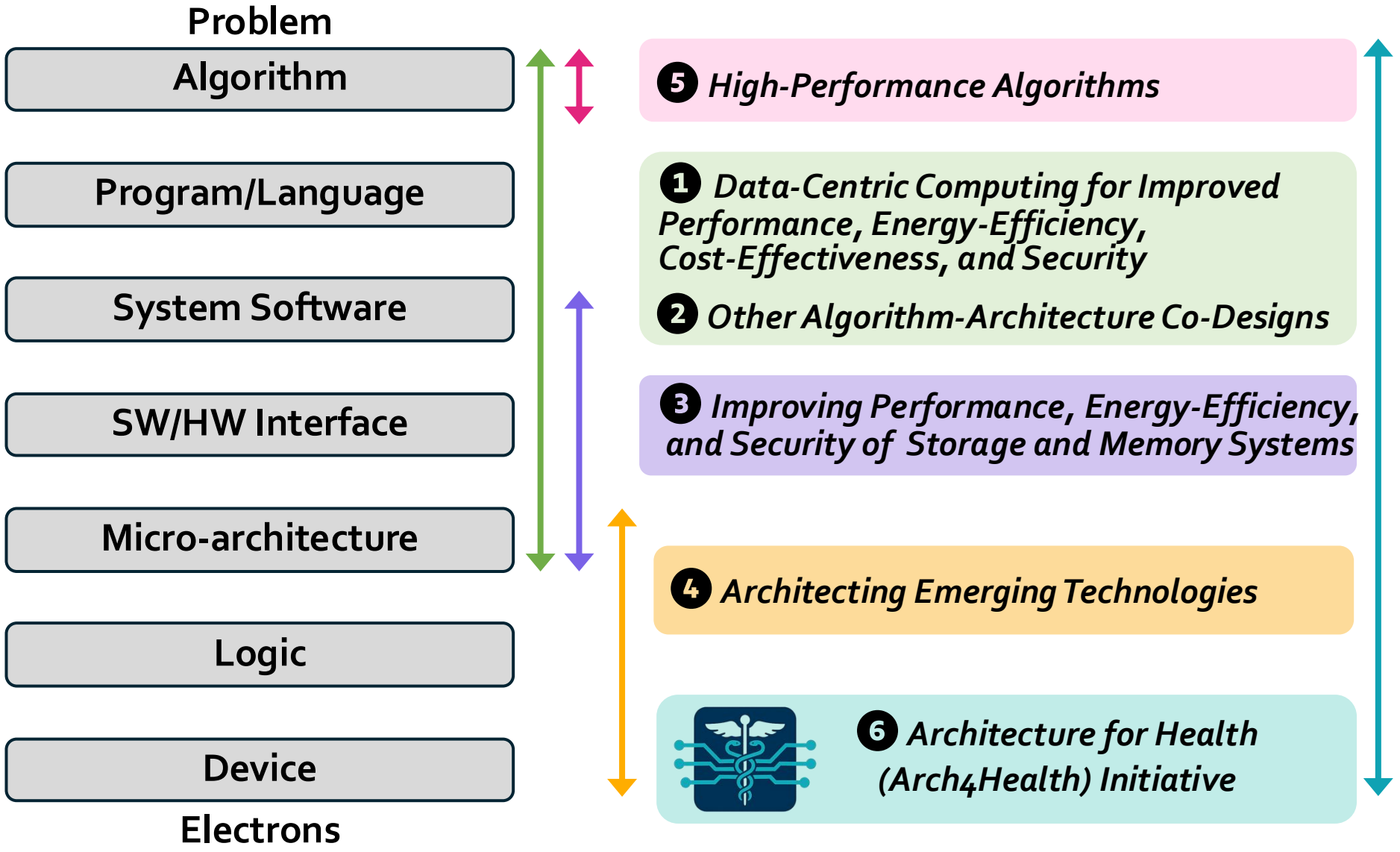
Next Steps

1. Facilitating wider adoption

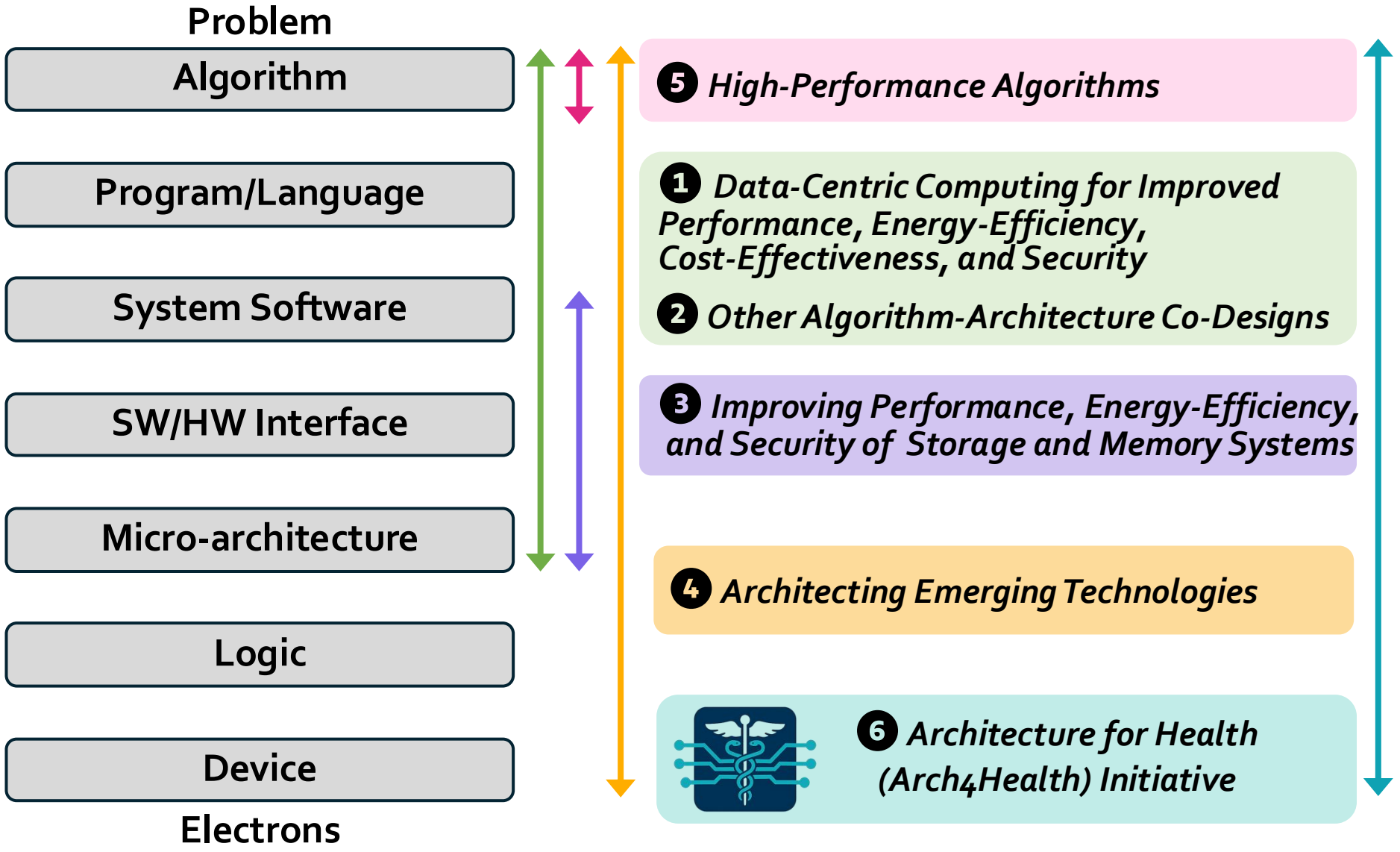
2. Increase reach to other critical problems and settings

3. From emerging technologies to emerging capabilities

Across the Computing Stack for Health, Life Sciences, and Beyond



Across the Computing Stack for Health, Life Sciences, and Beyond



How can we get there?

How can we get there?

The path will cross

Different layers of the computing system stack

Different expertise in different fields

Different perspectives

Fast, Efficient, and Low-Cost Computing Systems for Health and Life Sciences

Nika Mansouri Ghiasi
mnika@ethz.ch

SAFARI

ETH zürich