

# GenStore: A High-Performance In-Storage Processing System for Genome Sequence Analysis

Session 6A: Thursday 3 March, 3:00 PM CEST

Nika Mansouri Ghiasi, Jisung Park, Harun Mustafa, Jeremie Kim, Ataberk Olgun, Arvid Gollwitzer, Damla Senol Cali, Can Firtina, Haiyu Mao, Nour Almadhoun Alserr, Rachata Ausavarungnirun, Nandita Vijaykumar, Mohammed Alser, and Onur Mutlu

**SAFARI**

**ETH** zürich

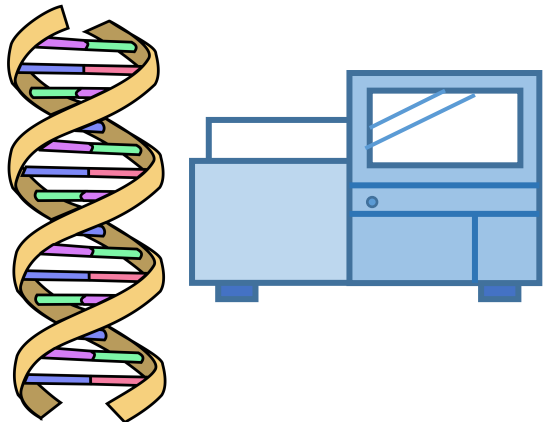
**bionano**  
GENOMICS



UNIVERSITY OF  
**TORONTO**

# Genome Sequence Analysis

- **Genome sequence analysis (GSA)** is critical for many applications.
  - Personalized medicine
  - Outbreak tracing
  - Evolutionary studies
- Genome sequencing machines extract smaller fragments of the original DNA sequence, known as **reads**.



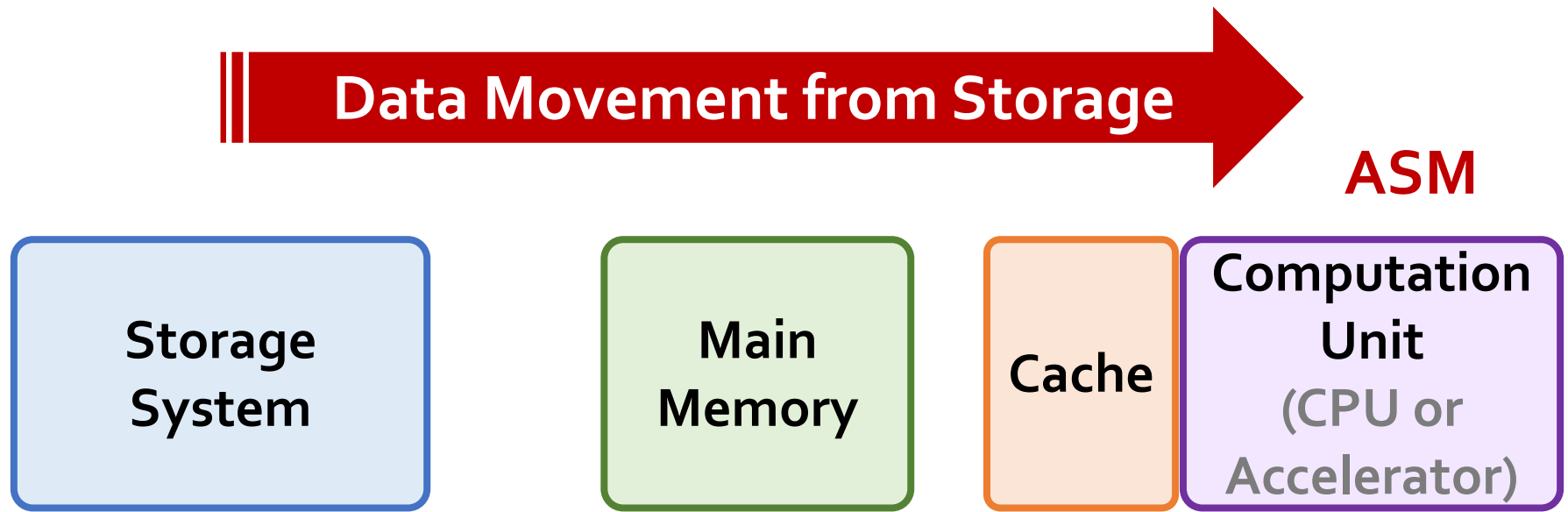
# Genome Sequence Analysis

- **Read mapping:** First key step in genome sequence analysis
  - Aligns reads to potential matching locations in the reference genome



- Read mapping requires **computationally-expensive approximate string matching (ASM)** to account for differences between reads and the reference genome due to:
  - Sequencing errors
  - Genetic Variation

# Genome Sequence Analysis

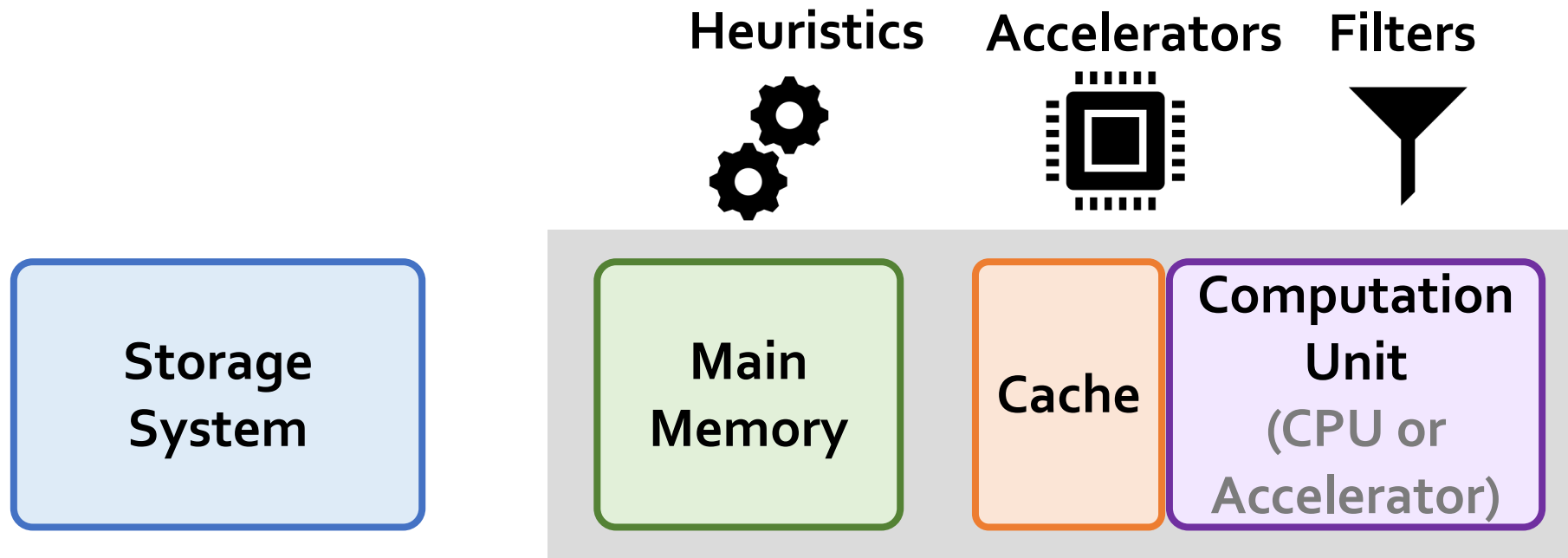


Computation overhead



Data movement overhead

# Improving the Performance of GSA



Computation overhead

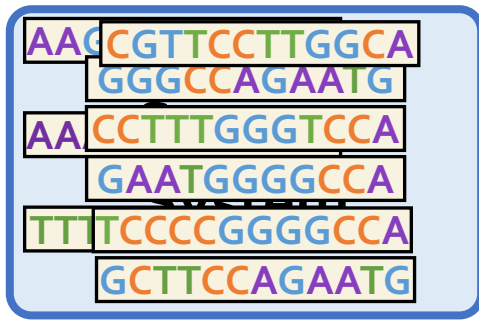


Data movement overhead  
(The effect becomes even larger)

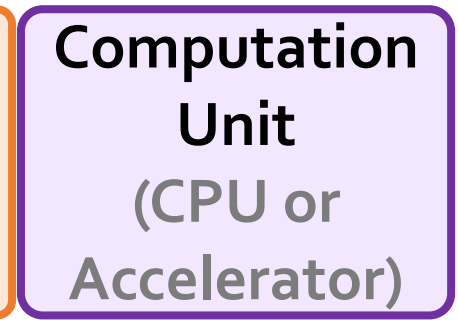
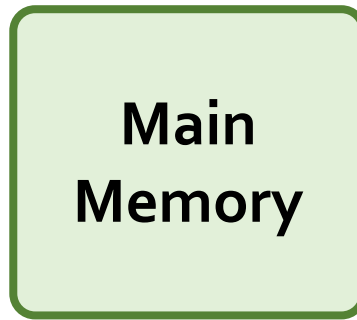
# Key Idea



*Filter reads that do not require ASM inside the storage system*



Filtered Reads



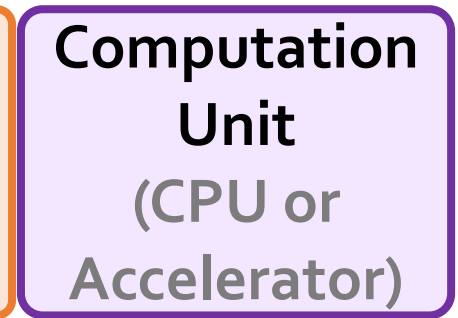
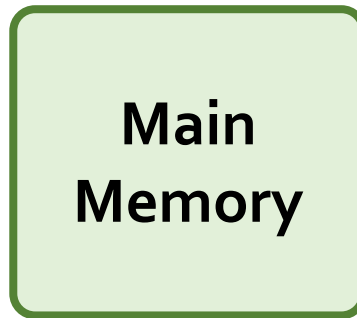
Read mapping workloads can exhibit different behavior

There are **limited available hardware resources** in the storage system

# GenStore



*Filter reads that do not require ASM inside the storage system*



Computation overhead



Data movement overhead

GenStore provides significant speedup (1.4x - 33.6x) and energy reduction (3.9x - 29.2x) at low cost

# GenStore: A High-Performance In-Storage Processing System for Genome Sequence Analysis

**Session 6A: Thursday 3 March, 3:00 PM CEST**

**Nika Mansouri Ghiasi** ([mnika@ethz.ch](mailto:mnika@ethz.ch))

Jisung Park, Harun Mustafa, Jeremie Kim, Ataberk Olgun,  
Arvid Gollwitzer, Damla Senol Cali, Can Firtina, Haiyu Mao, Nour Almadhoun Alserr,  
Rachata Ausavarungnirun, Nandita Vijaykumar, Mohammed Alser, and Onur Mutlu

**SAFARI**

**ETH** zürich

**bionano**  
GENOMICS



UNIVERSITY OF  
**TORONTO**