

GenASM: A High Performance, Low-Power Approximate String Matching Acceleration Framework for Genome Sequence Analysis

Damla Senol Cali¹

Carnegie Mellon University
(dsenol@andrew.cmu.edu)

Gurpreet S. Kalsi², Zual Bingol³, Can Firtina⁴, Lavanya Subramanian²,
Jeremie Kim^{1,4}, Rachata Ausavarungnirun^{5,1}, Mohammed Alser⁴,
Juan Gomez-Luna⁴, Amirali Boroumand¹, Anant Nori², Allison Scibisz¹,
Sreenivas Subramoney², Can Alkan³, Saugata Ghose^{6,1}, and Onur Mutlu^{4,1,3}

¹ **Carnegie Mellon**

² **intel**



Bilkent University ⁴ **ETH zürich**



⁶ **I** UNIVERSITY OF
ILLINOIS
URBANA - CHAMPAIGN

^{1,4} **SAFARI**

Genome Sequence Analysis (GSA)

- ❑ *Read mapping*: First key step in genome sequence analysis (GSA)
 - Aligns **reads** to one or more possible locations within the **reference genome**, and
 - Finds the **matches** and **differences** between the read and the reference genome segment at that location
- ❑ Multiple steps of read mapping require *approximate string matching*
 - Approximate string matching (ASM) enables read mapping to account for **sequencing errors** and **genetic variations** in the reads
- ❑ Bottlenecked by the **computational power and memory bandwidth limitations of existing systems**

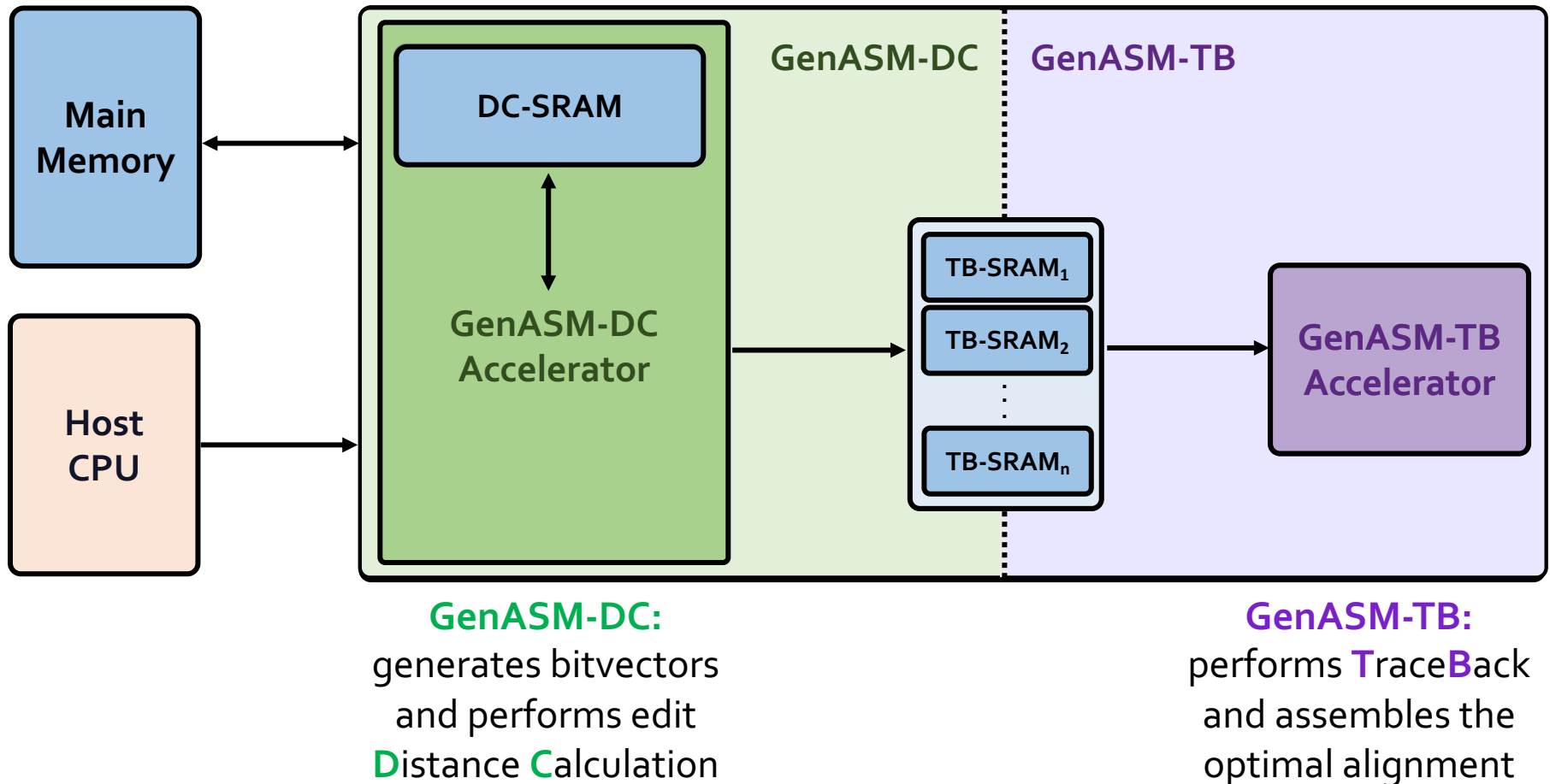
GenASM: ASM Framework for GSA

Our Goal:

Accelerate approximate string matching
by designing a fast and flexible framework,
which can accelerate *multiple steps* of genome sequence analysis

- **GenASM:** First ASM acceleration framework for GSA
 - Based upon the *Bitap* algorithm
 - Uses fast and simple bitwise operations to perform ASM
 - Modified and extended ASM algorithm
 - Highly-parallel Bitap with long read support
 - Bitvector-based novel algorithm to perform *traceback*
 - Co-design of our modified scalable and memory-efficient algorithms with low-power and area-efficient hardware accelerators

GenASM: Hardware Design



Use Cases of GenASM

(1) Read Alignment Step of Read Mapping

- Find the **optimal alignment** of how reads map to candidate reference regions

(2) Pre-Alignment Filtering for Short Reads

- Quickly identify and **filter out the unlikely** candidate reference regions for each read

(3) Edit Distance Calculation

- Measure the **similarity** or **distance** between two sequences
- We also discuss **other possible use cases of GenASM** in our paper:
 - Read-to-read overlap finding, hash-table based indexing, whole genome alignment, generic text search

Key Results

(1) Read Alignment

- ❑ **116×** speedup, **37×** less power than **Minimap2** (state-of-the-art SW)
- ❑ **111×** speedup, **33×** less power than **BWA-MEM** (state-of-the-art SW)
- ❑ **3.9×** better throughput, **2.7×** less power than **Darwin** (state-of-the-art HW)
- ❑ **1.9×** better throughput, **82%** less logic power than **GenAx** (state-of-the-art HW)

(2) Pre-Alignment Filtering

- ❑ **3.7×** speedup, **1.7×** less power than **Shouji** (state-of-the-art HW)

(3) Edit Distance Calculation

- ❑ **22–12501×** speedup, **548–582×** less power than **Edlib** (state-of-the-art SW)
- ❑ **9.3–400×** speedup, **67×** less power than **ASAP** (state-of-the-art HW)

GenASM: A High Performance, Low-Power Approximate String Matching Acceleration Framework for Genome Sequence Analysis

Damla Senol Cali¹

Carnegie Mellon University
(dsenol@andrew.cmu.edu)

Gurpreet S. Kalsi², Zual Bingol³, Can Firtina⁴, Lavanya Subramanian²,
Jeremie Kim^{1,4}, Rachata Ausavarungnirun^{5,1}, Mohammed Alser⁴,
Juan Gomez-Luna⁴, Amirali Boroumand¹, Anant Nori², Allison Scibisz¹,
Sreenivas Subramoney², Can Alkan³, Saugata Ghose^{6,1}, and Onur Mutlu^{4,1,3}

¹ **Carnegie Mellon**

² **intel**



Bilkent University ⁴ **ETH zürich**



⁶ **I** UNIVERSITY OF
ILLINOIS
URBANA - CHAMPAIGN

^{1,4} **SAFARI**