

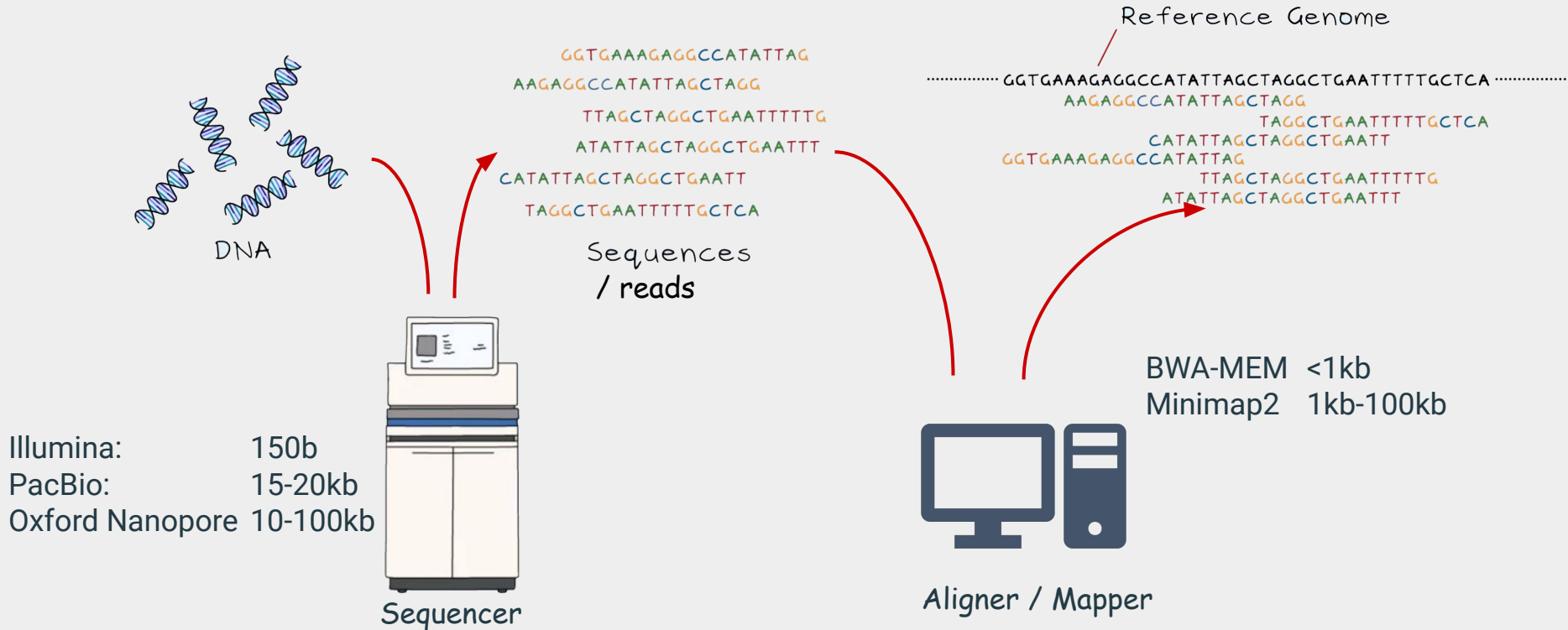
@BioSys Workshop'24

mm2-gb: GPU Accelerated Minimap2 for Long Read DNA Mapping

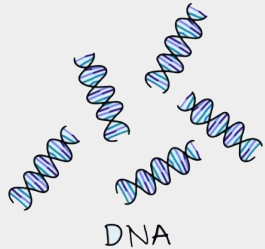
Juechu Dong*¹, Xueshen Liu*¹, Harisankar Sadasivan², Sriranjani
Sitaraman², Satish Narayanasamy¹

1. University of Michigan 2. AMD Inc.

Long Read Mapping is slow



Long Read Mapping is slow



GGTGAAAGAGGCCATATTAG
AAGAGGCCATATTAGCTAGG
TTAGCTAGGCTGAATTTTTG
ATATTAGCTAGGCTGAATTT
CATATTAGCTAGGCTGAATT
TAGGCTGAATTTTTGCTCA

Sequences / reads

**Challenging to map:
Irregular, noisy, long Reads.**

Reference Genome
..... GGTGAAAGAGGCCATATTAGCTAGGCTGAATTTTTGCTCA
AAGAGGCCATATTAGCTAGG
TAGGCTGAATTTTTGCTCA
CATATTAGCTAGGCTGAATT
GGTGAAAGAGGCCATATTAG
TTAGCTAGGCTGAATTTTTG
ATATTAGCTAGGCTGAATTT

Illumina: 150b
PacBio: 15-20kb
Oxford Nanopore 10-100kb

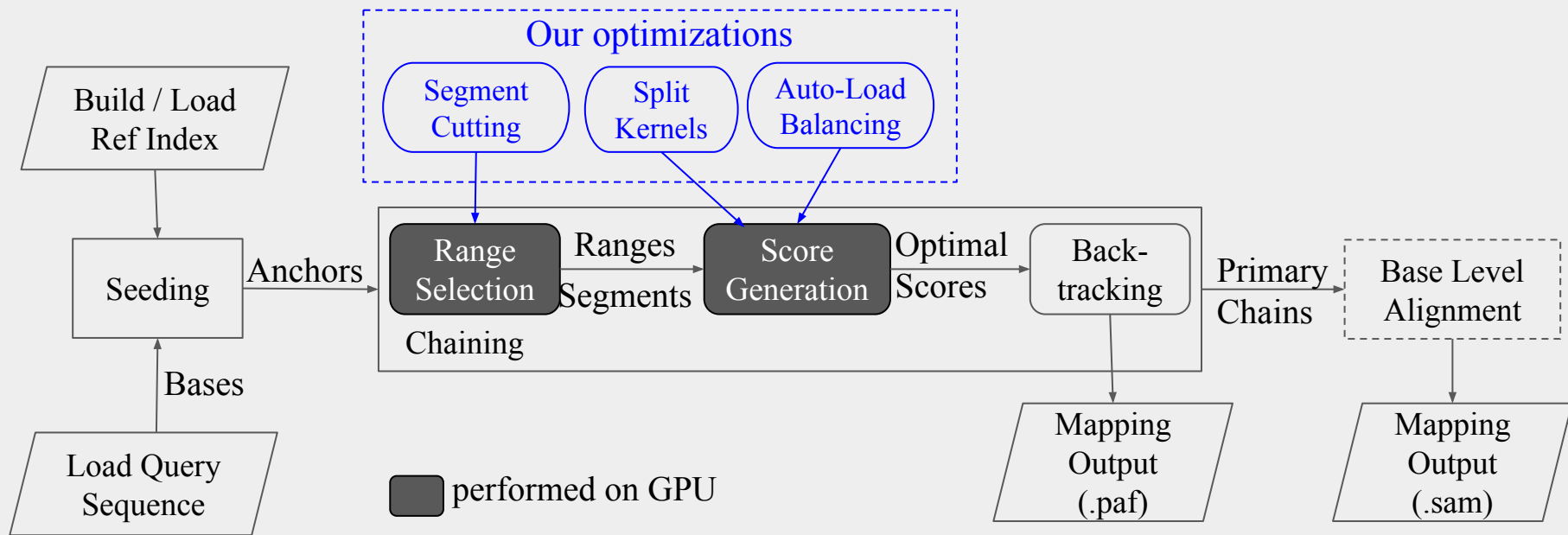


Aligner / Mapper

BWA-MEM <1kb
Minimap2 1kb-100kb

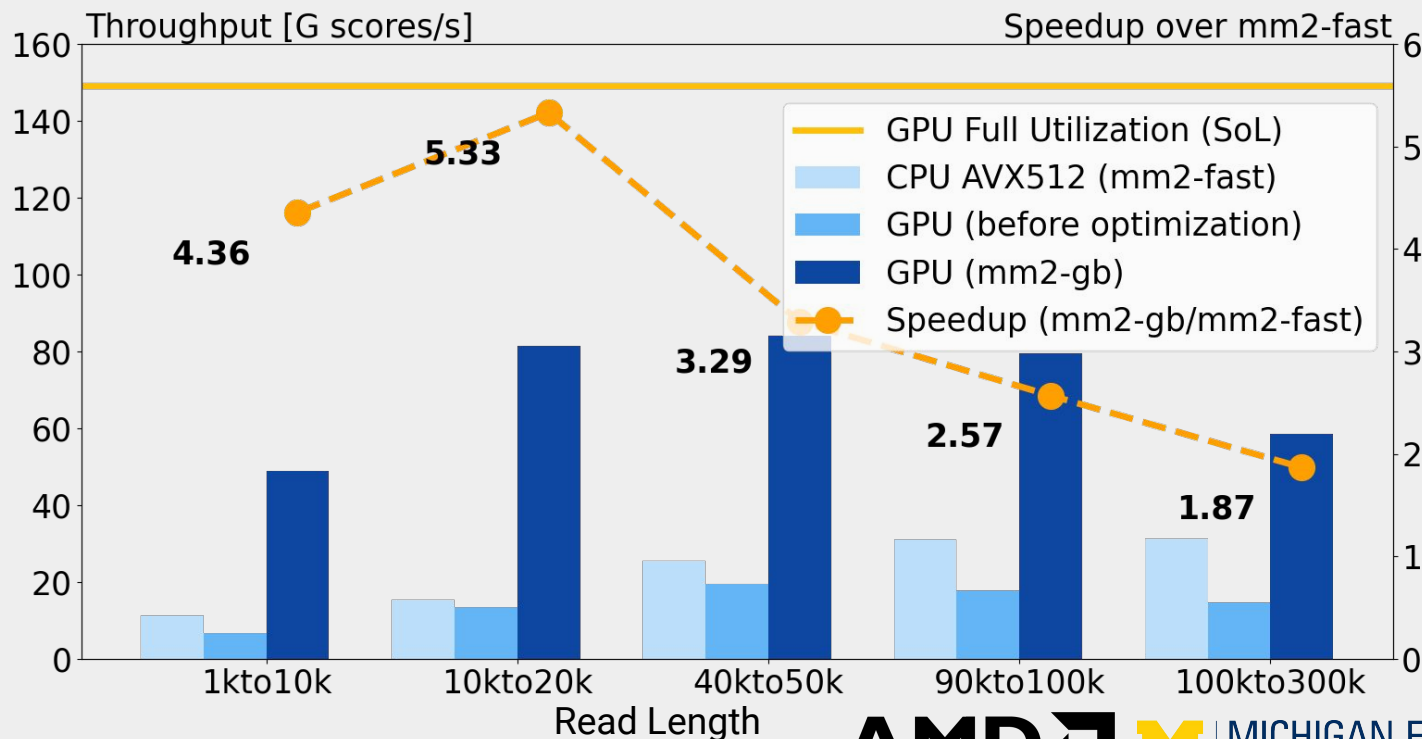
**SOA, runs on CPU,
slow.**

Accelerating *minimap2* on GPU



mm2-gb offers 5.33x faster chaining

No accuracy loss
Open sourced



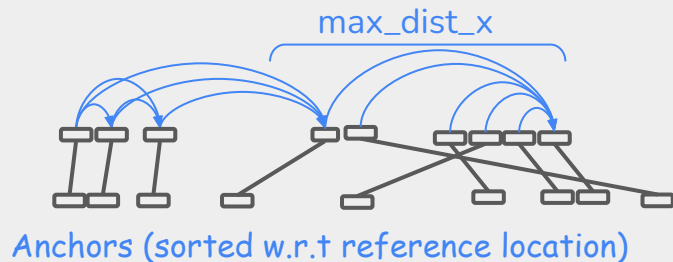
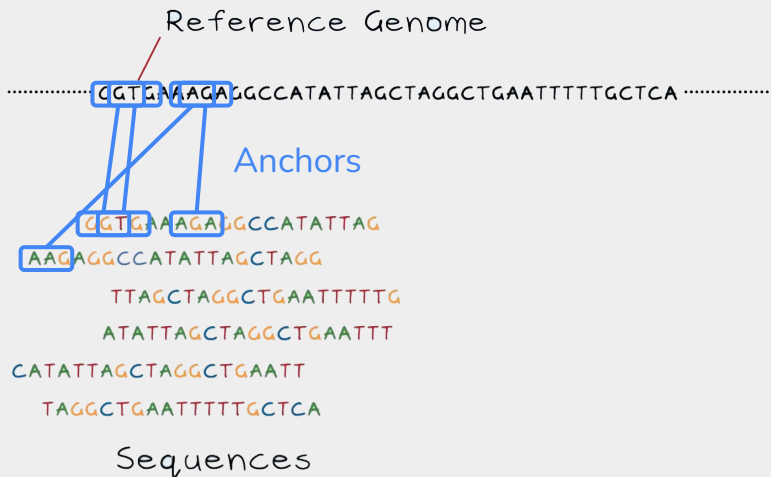
M Agenda

- What is minimap2 & why it is challenging to parallelize
- How we accelerate minimap2
- Hardware Setup and Results

minimap2 does chaining sequentially

Chaining

1D dynamic programming

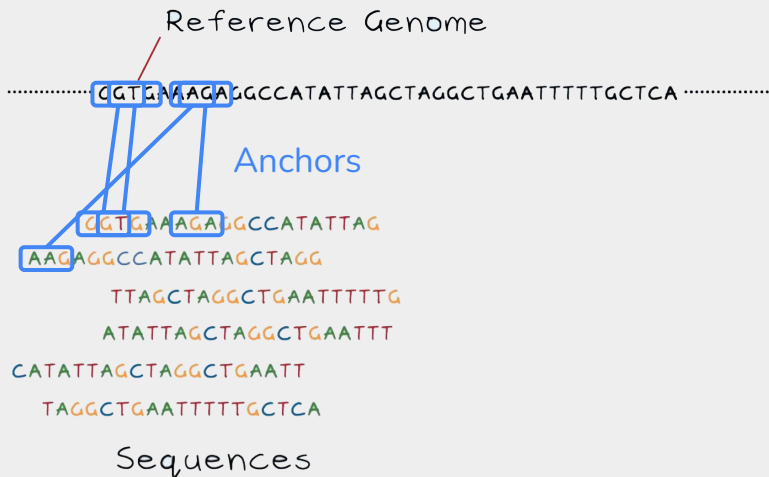


Seeding

minimap2 does chaining sequentially

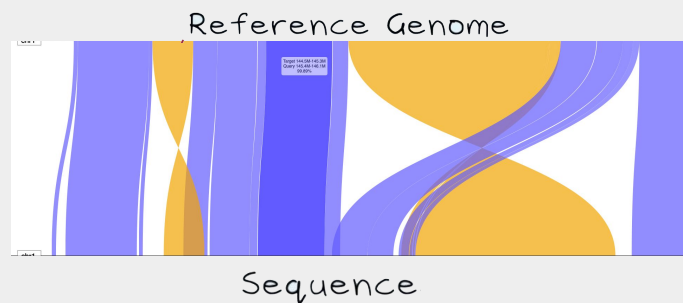
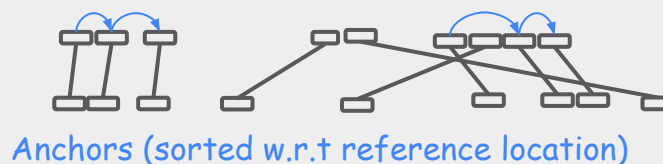
Chaining

1D dynamic programming

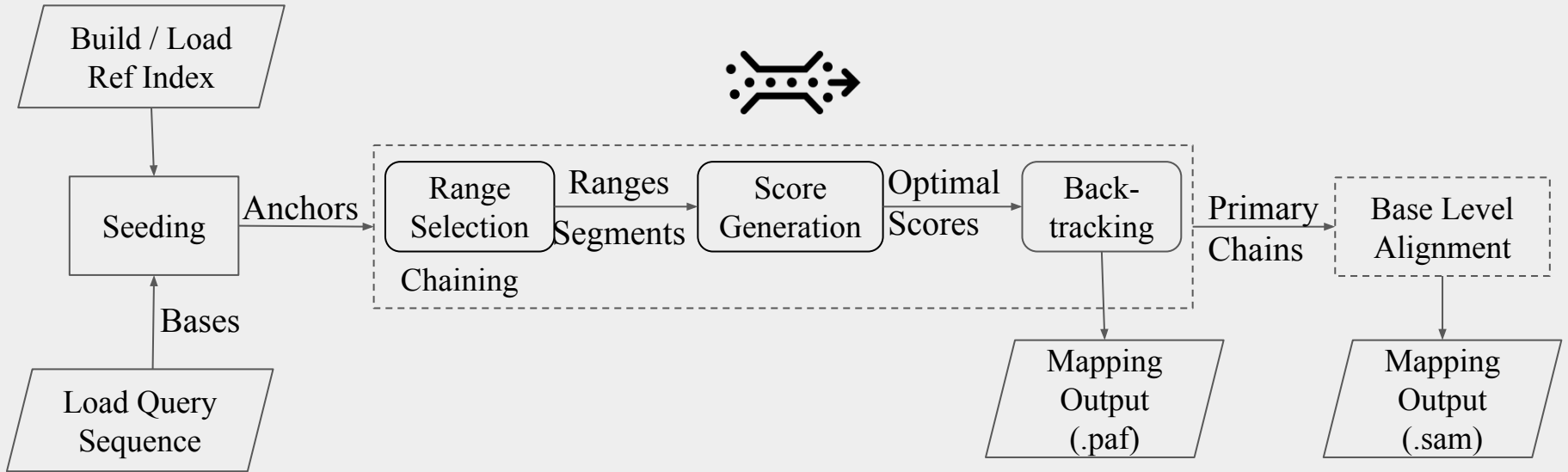


Seeding

max_dist_x



Chaining is the bottleneck

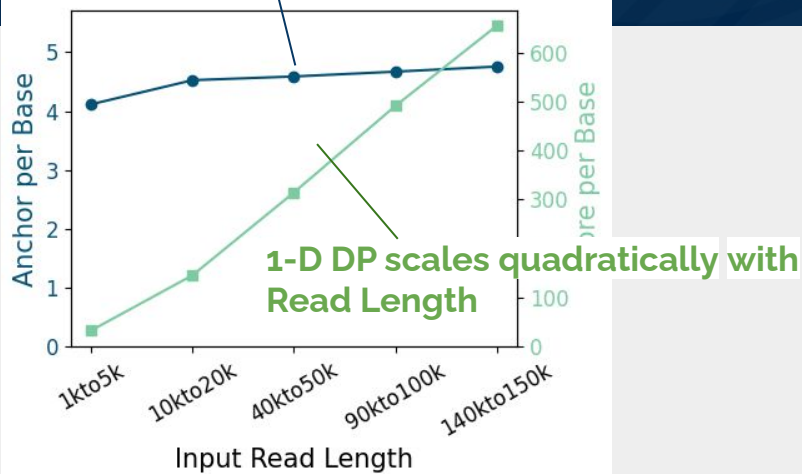


67%

of CPU time in minimap2 is spent on chaining



Seeding scales linearly



Chaining is the scaling bottleneck

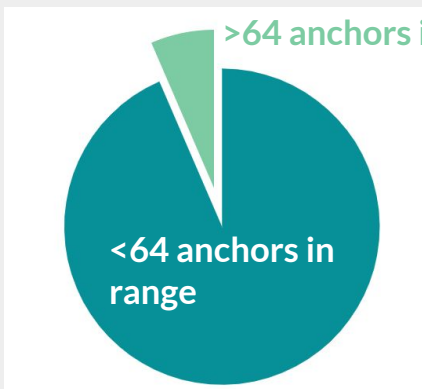
Irregular Compute Intensity

Reference Genome

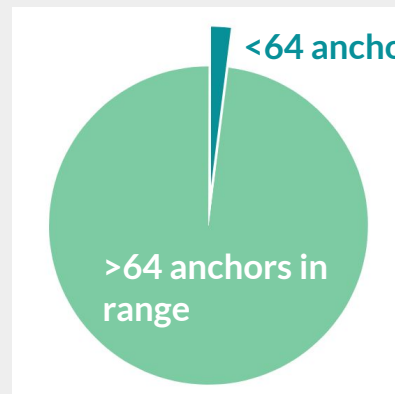


Potential map regions:
A lot of anchors in `max_dist_x`

Sparse Anchors:
Only a few anchors in `max_dist_x`

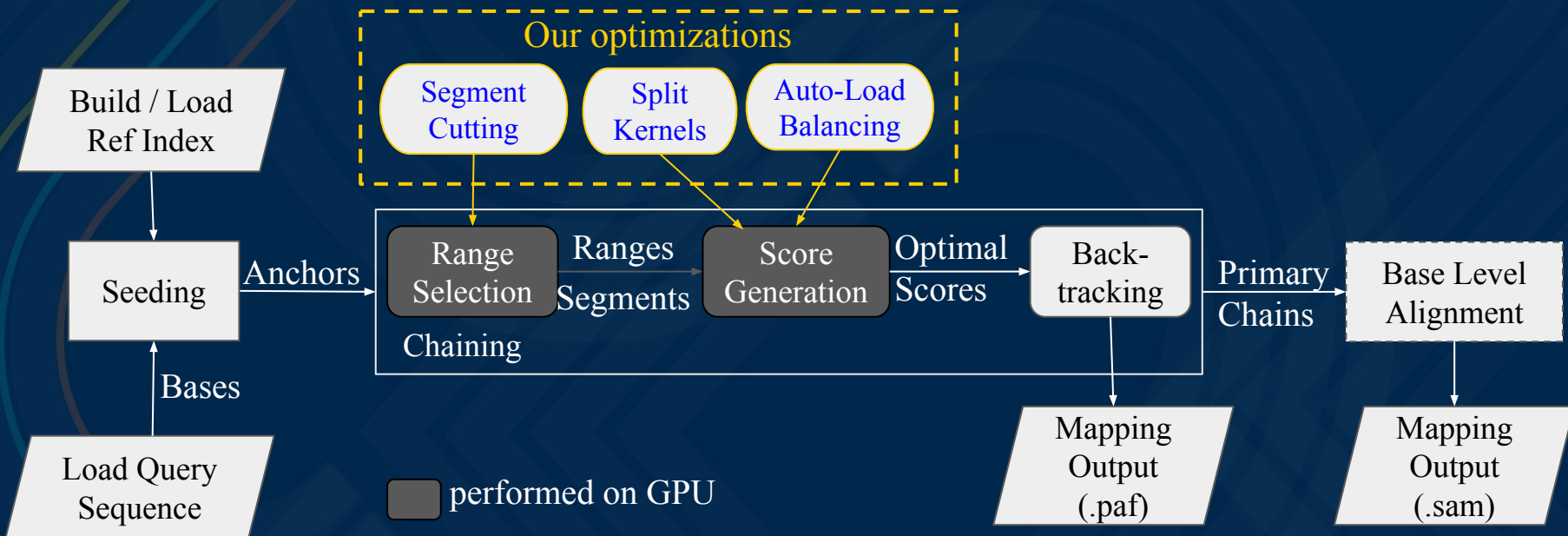


Number of anchors



Number of scores to compute

mm2-gb Design Overview



Parallelize Read Processing by Sequence Cutting

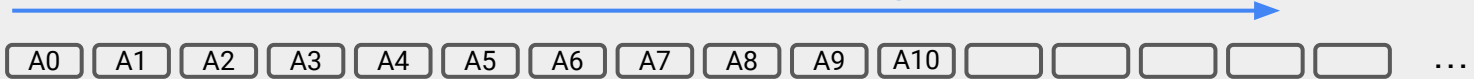
Read#0:



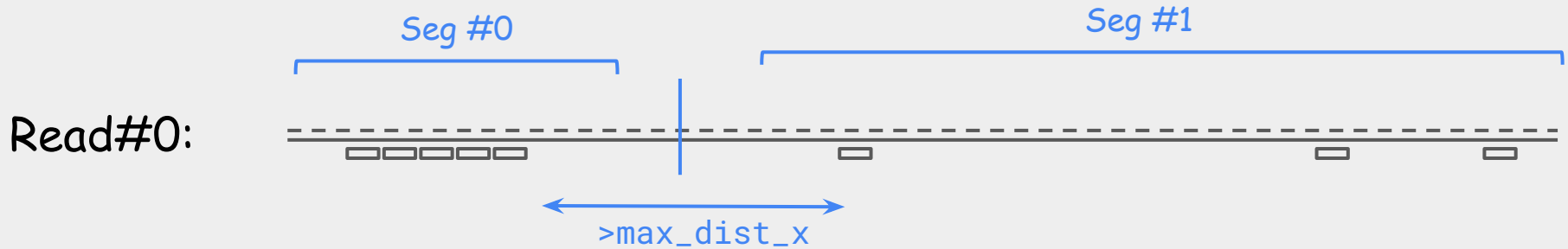
Anchor score is dependent on its predecessors

Sequential Processing

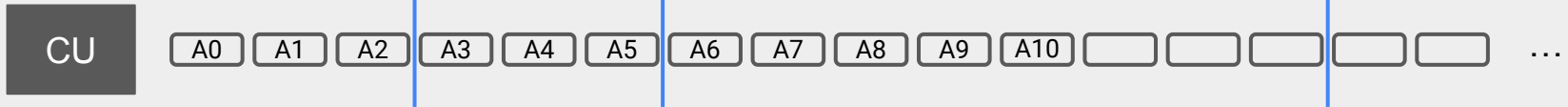
CU



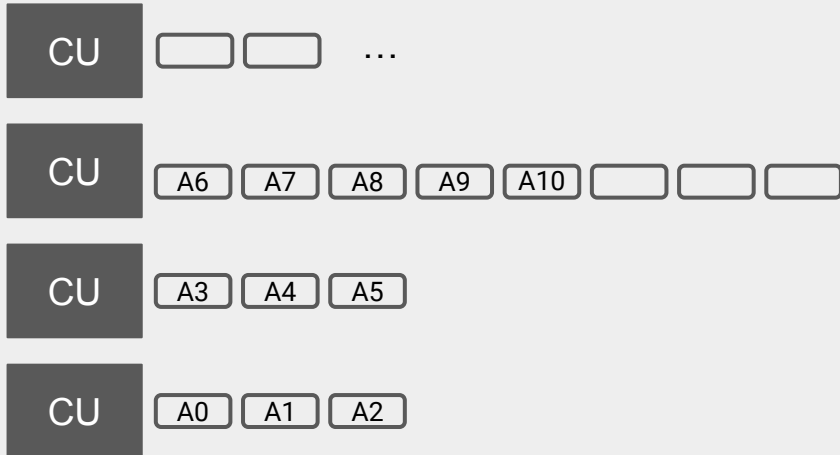
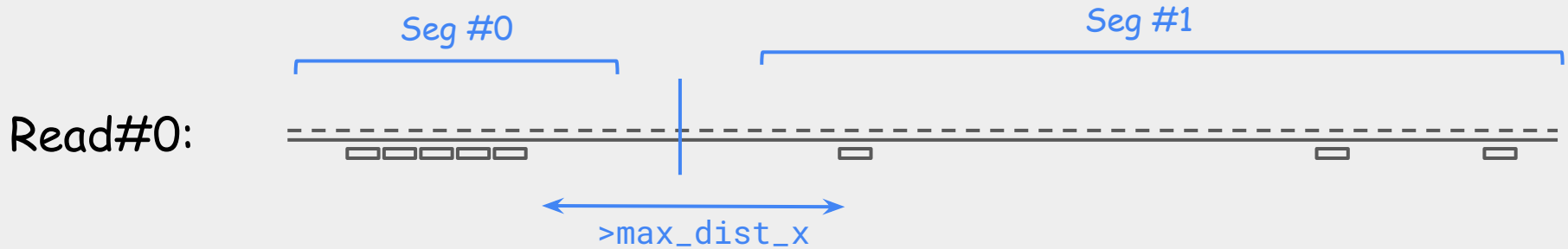
Parallelize Read Processing by Sequence Cutting



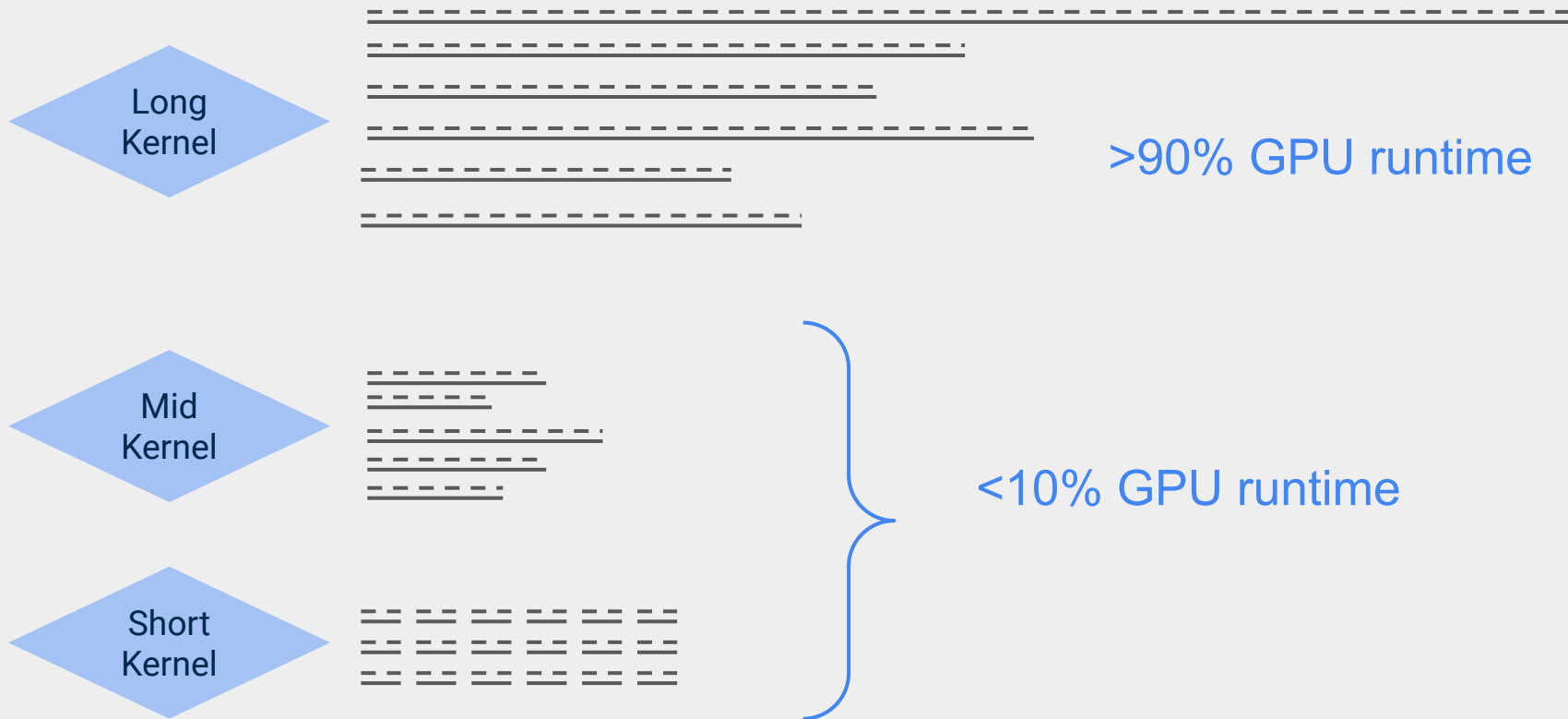
Anchor score is dependent on its predecessors
– Unless there is no anchor within max_dist_x .



Parallelize Read Processing by Sequence Cutting



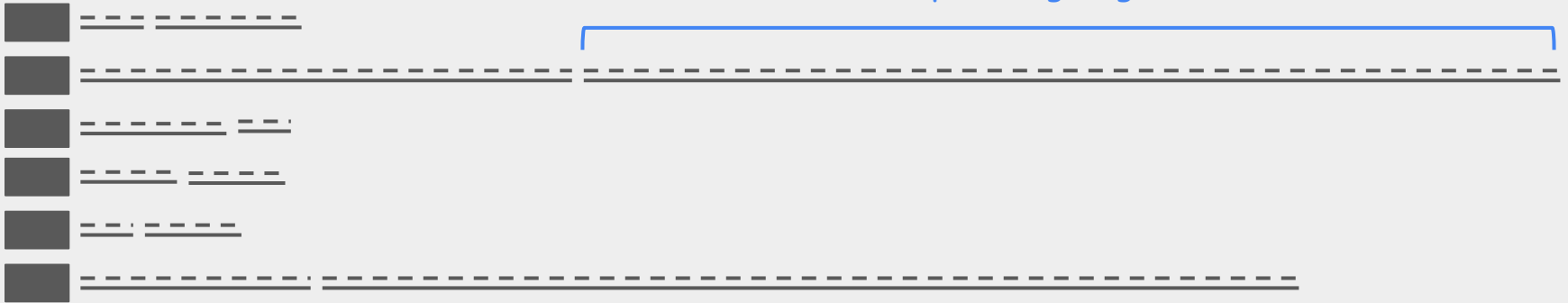
Custom Kernels Tailored to Compute Intensities



Prioritized Auto Workload balancing

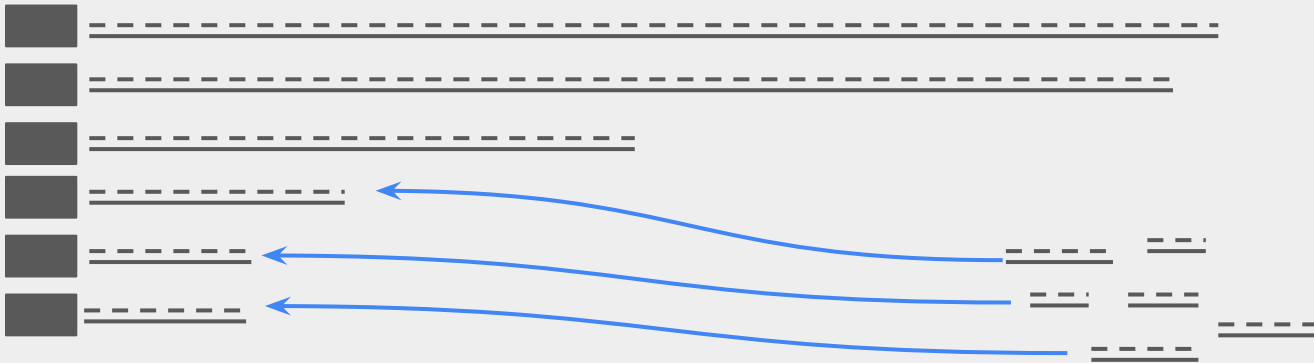
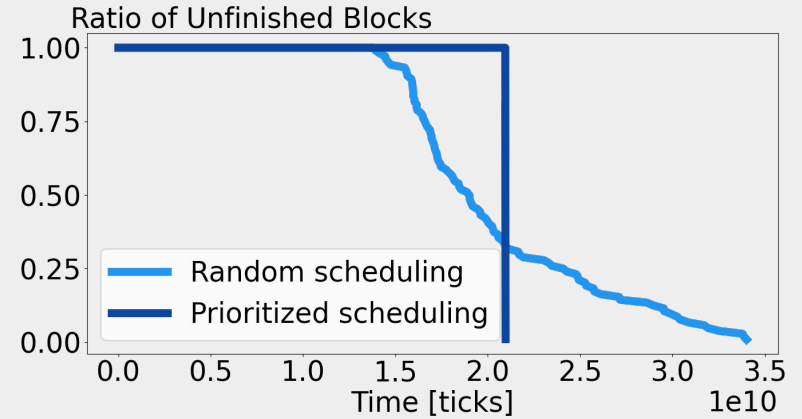


Super Long Segments



Prioritized Auto Workload balancing

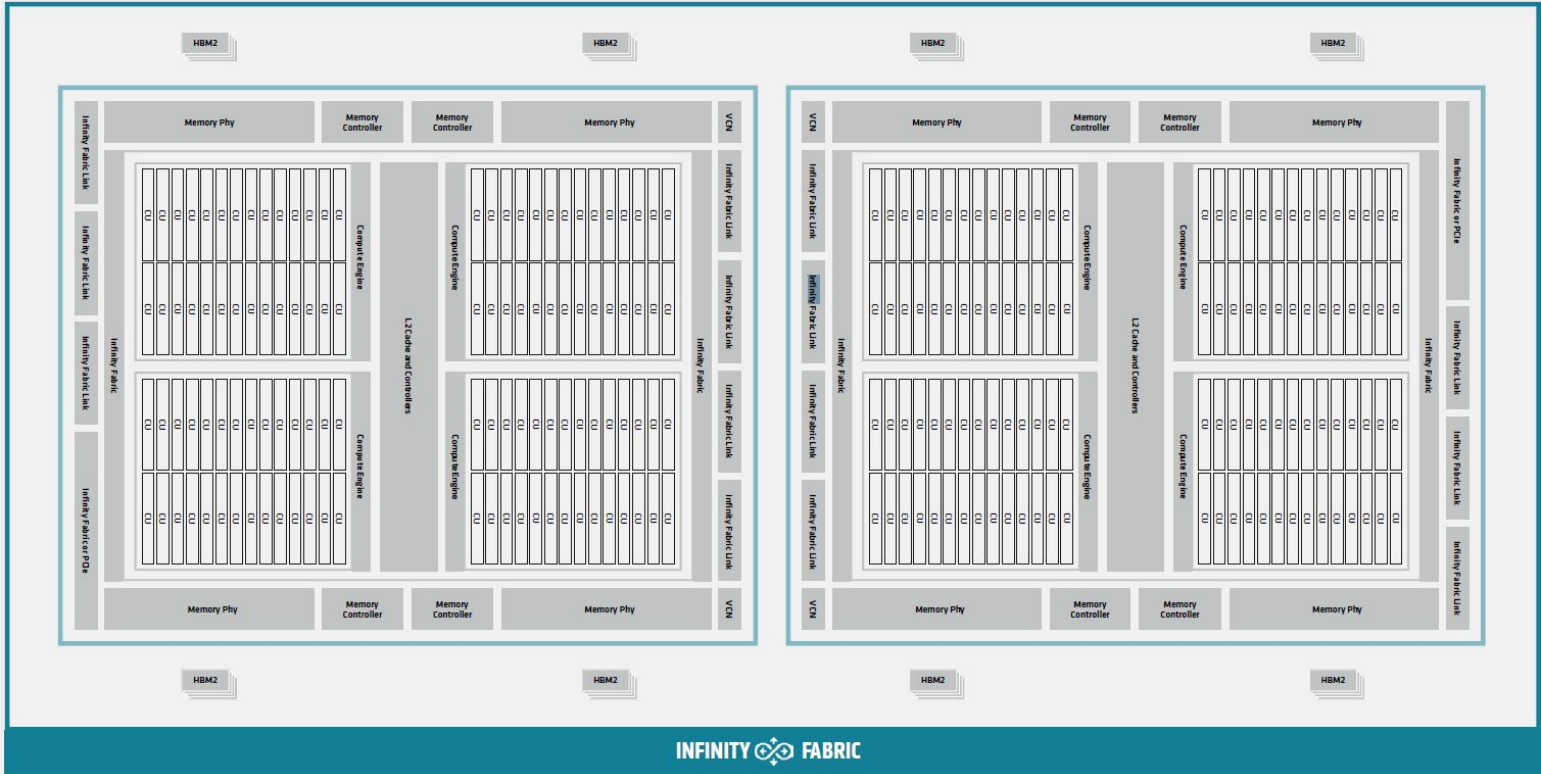
Long Kernel



Dynamically Fetch from TODO pool

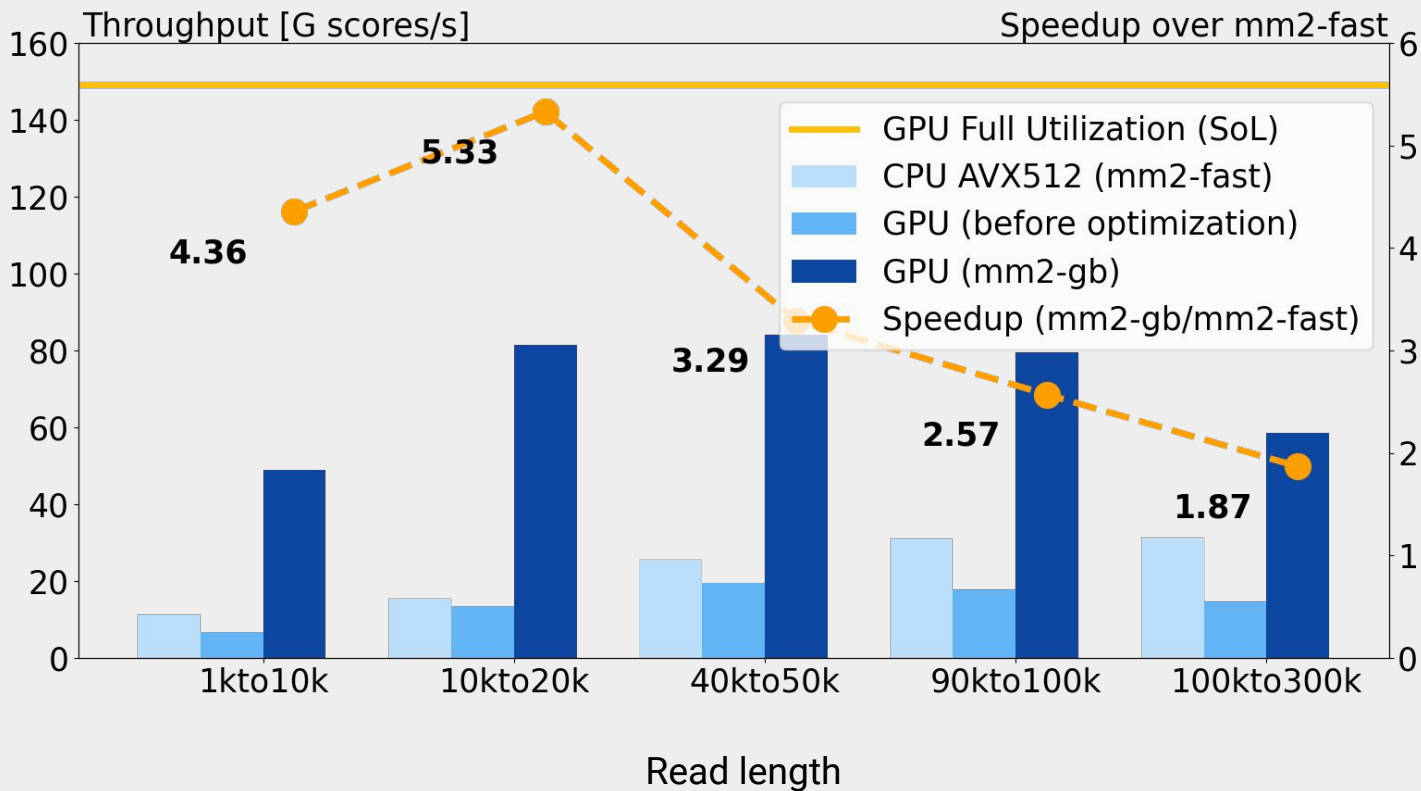
Results

AMD Instinct™ MI210 Accelerators



INFINITY FABRIC

Speedup Compared to mm2-fast* using avx-512 on Intel® Icelake



The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. The information contained herein is subject to change and may be rendered inaccurate for many reasons, including but not limited to product and roadmap changes, component and motherboard version changes, new model and/or product releases, product differences between differing manufacturers, software changes, BIOS flashes, firmware upgrades, or the like. Any computer system has risks of security vulnerabilities that cannot be completely prevented or mitigated. AMD assumes no obligation to update or otherwise correct or revise this information. However, AMD reserves the right to revise this information and to make changes from time to time to the content hereof without obligation of AMD to notify any person of such revisions or changes.

THIS INFORMATION IS PROVIDED ‘AS IS.’ AMD MAKES NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE CONTENTS HEREOF AND ASSUMES NO RESPONSIBILITY FOR ANY INACCURACIES, ERRORS, OR OMISSIONS THAT MAY APPEAR IN THIS INFORMATION. AMD SPECIFICALLY DISCLAIMS ANY IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY, OR FITNESS FOR ANY PARTICULAR PURPOSE. IN NO EVENT WILL AMD BE LIABLE TO ANY PERSON FOR ANY RELIANCE, DIRECT, INDIRECT, SPECIAL, OR OTHER CONSEQUENTIAL DAMAGES ARISING FROM THE USE OF ANY INFORMATION CONTAINED HEREIN, EVEN IF AMD IS EXPRESSLY ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

© 2024 Advanced Micro Devices, Inc. All rights reserved. AMD, the AMD Arrow logo, ROCm, Radeon, CDNA, Instinct, and combinations thereof are trademarks of Advanced Micro Devices, Inc. Other product names used in this publication are for identification purposes only and may be trademarks of their respective companies.