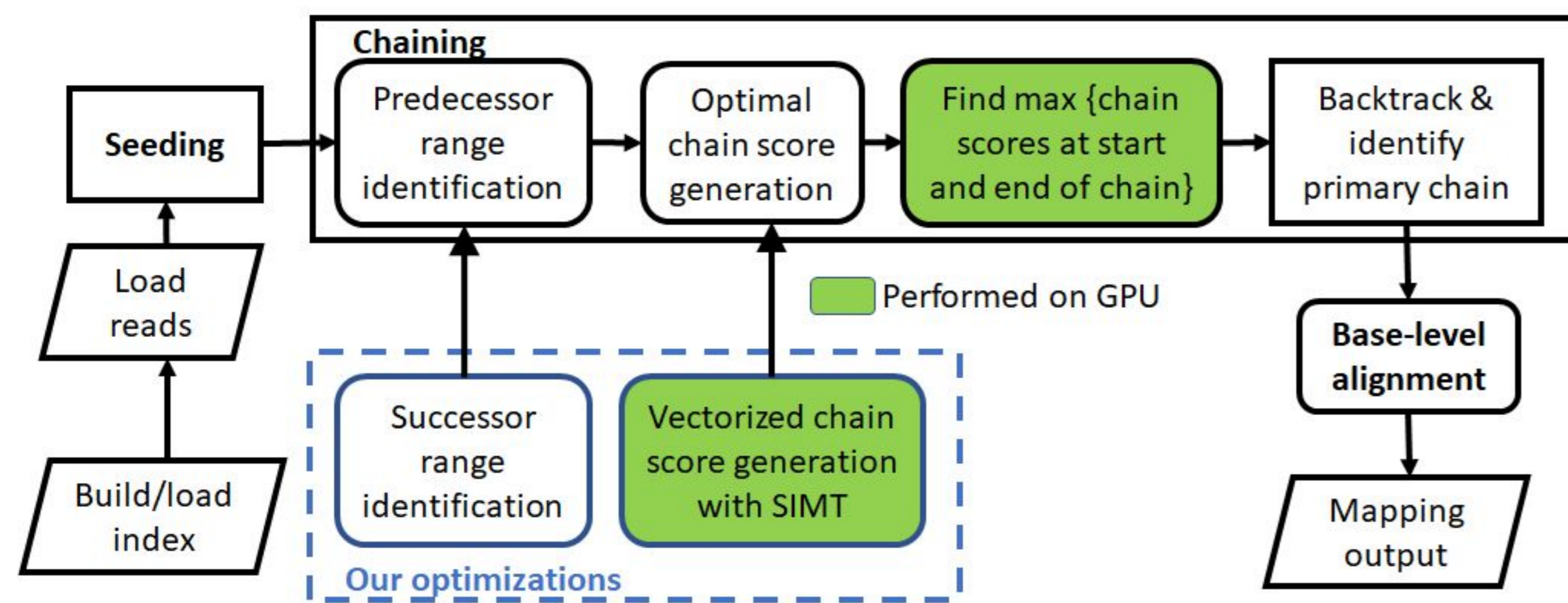


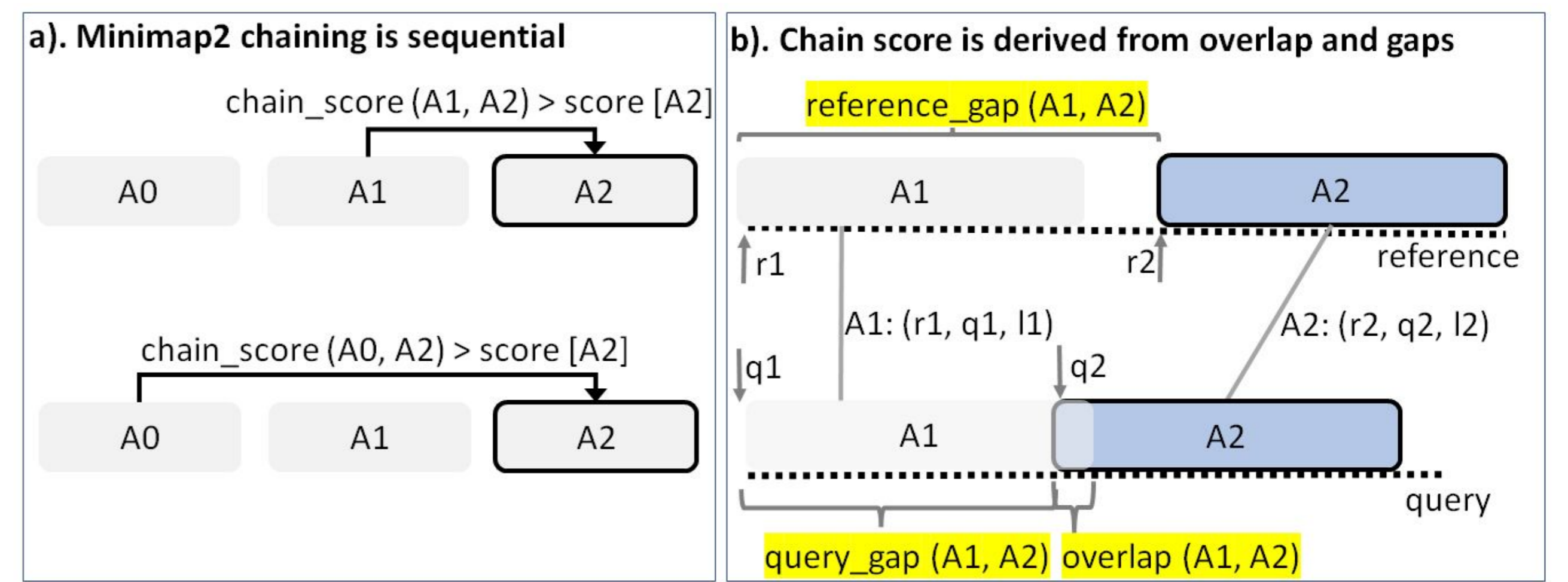
# Accelerating Minimap2 for accurate long read alignment on GPUs

ALIGNMENT

Minimap2 aligns sequences to identify regions of similarity

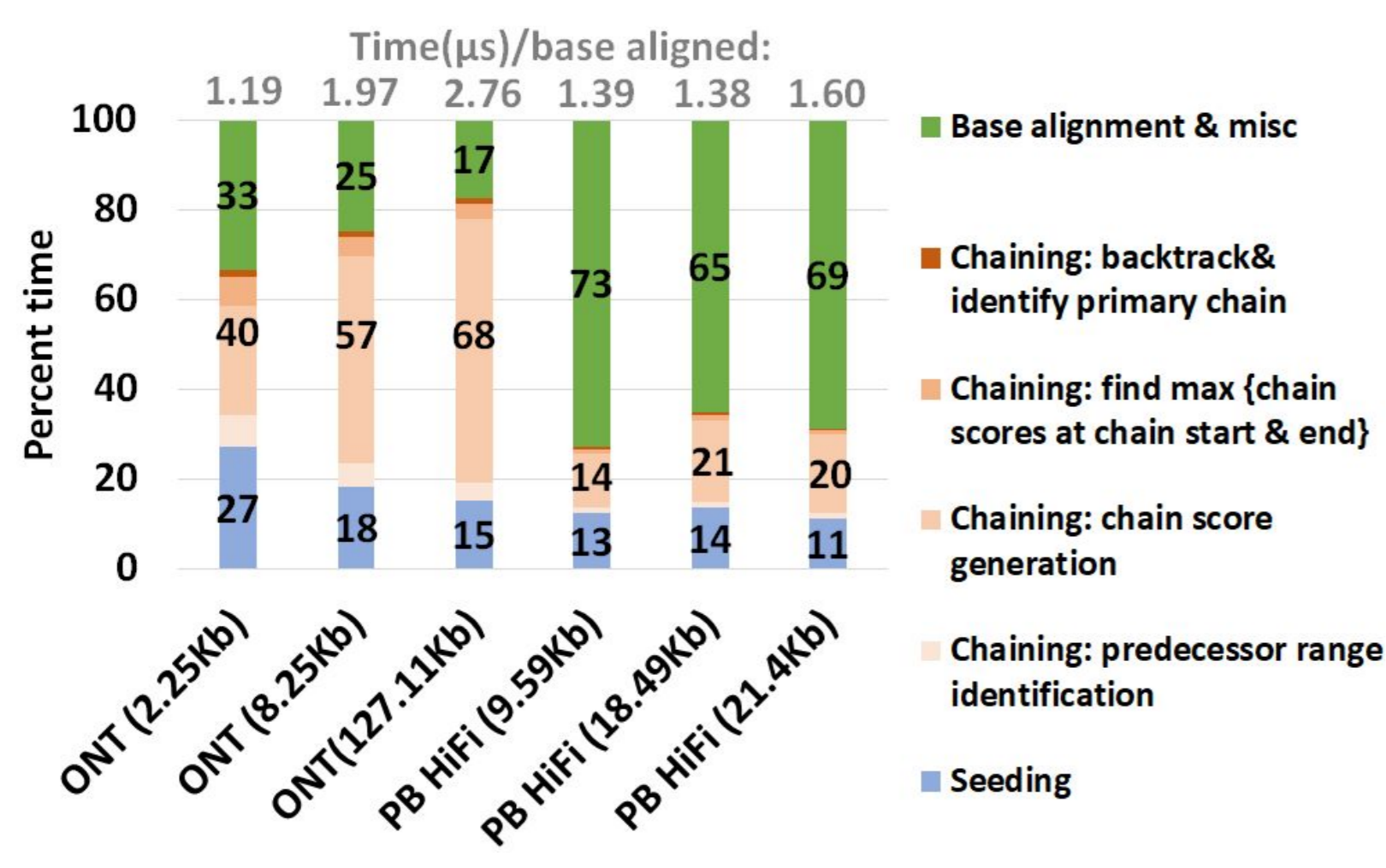


Sequential chaining step is the bottleneck for ONT

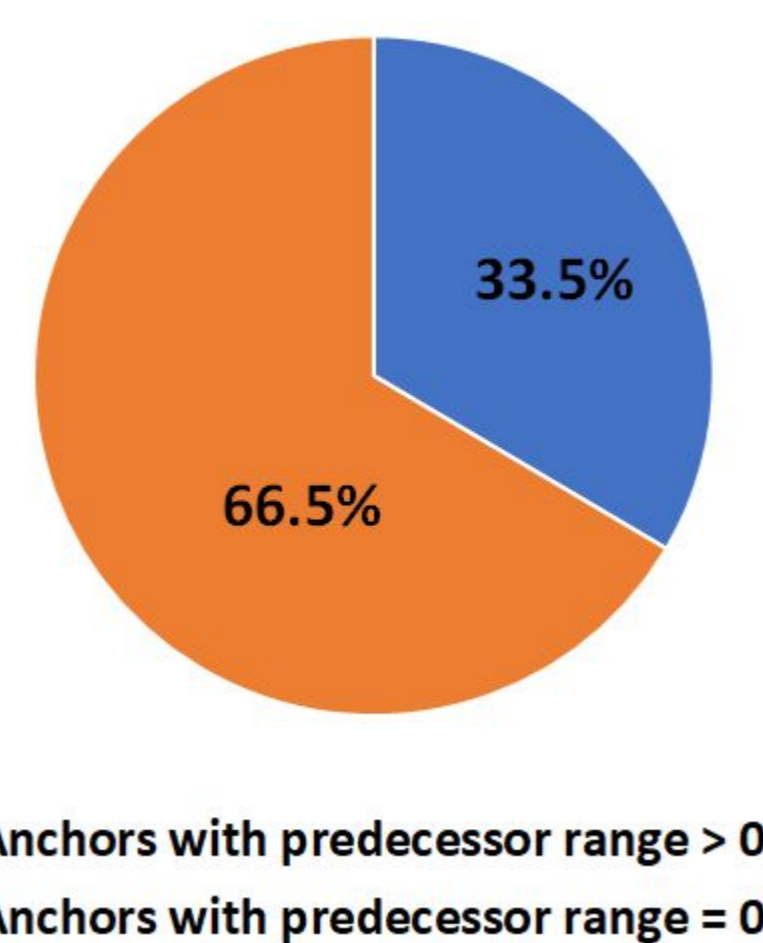


MOTIVATION

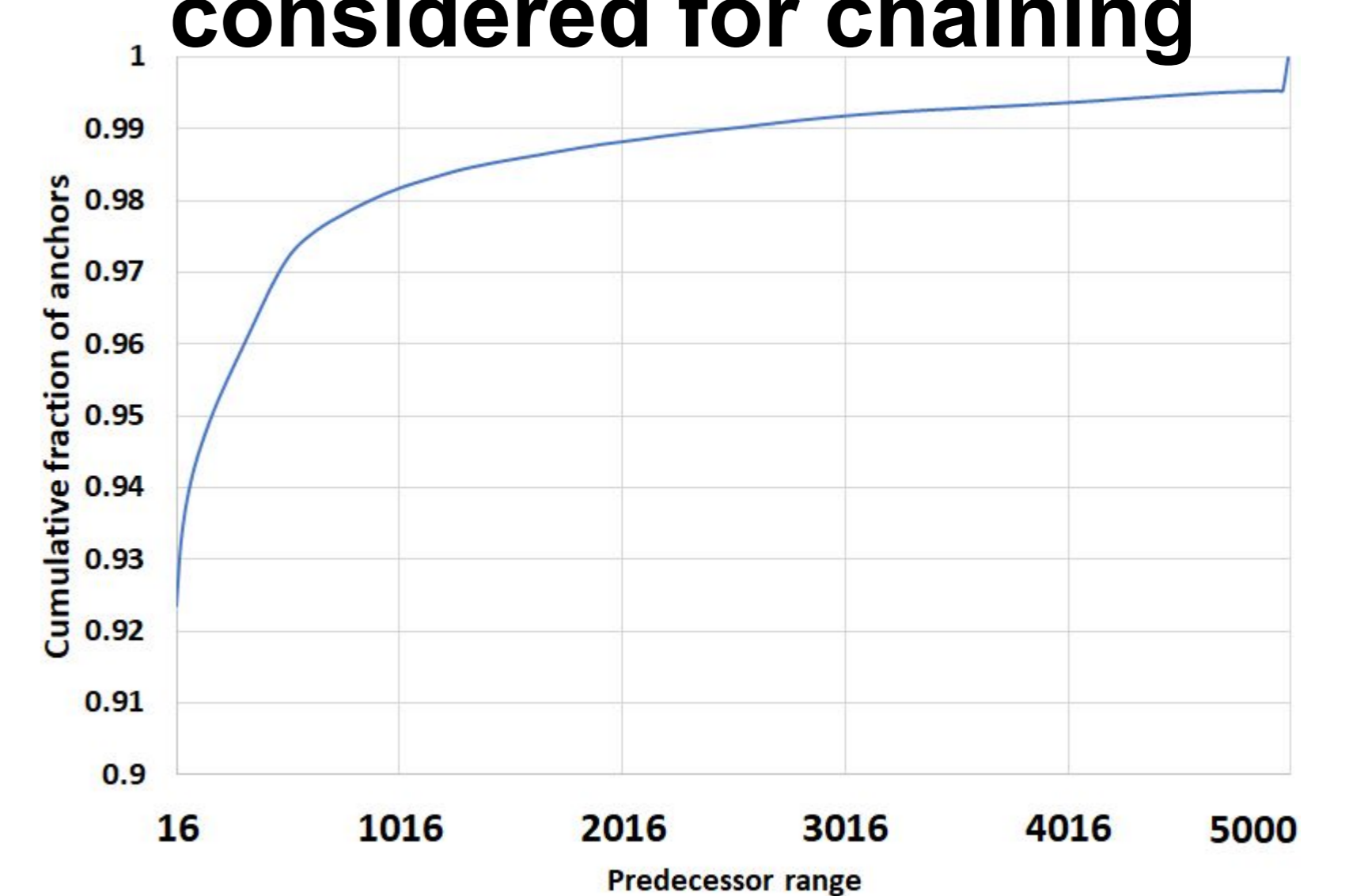
Sequential: Chaining is ~60-70% of run-time



Sparse: ~67% seeds don't start a chain

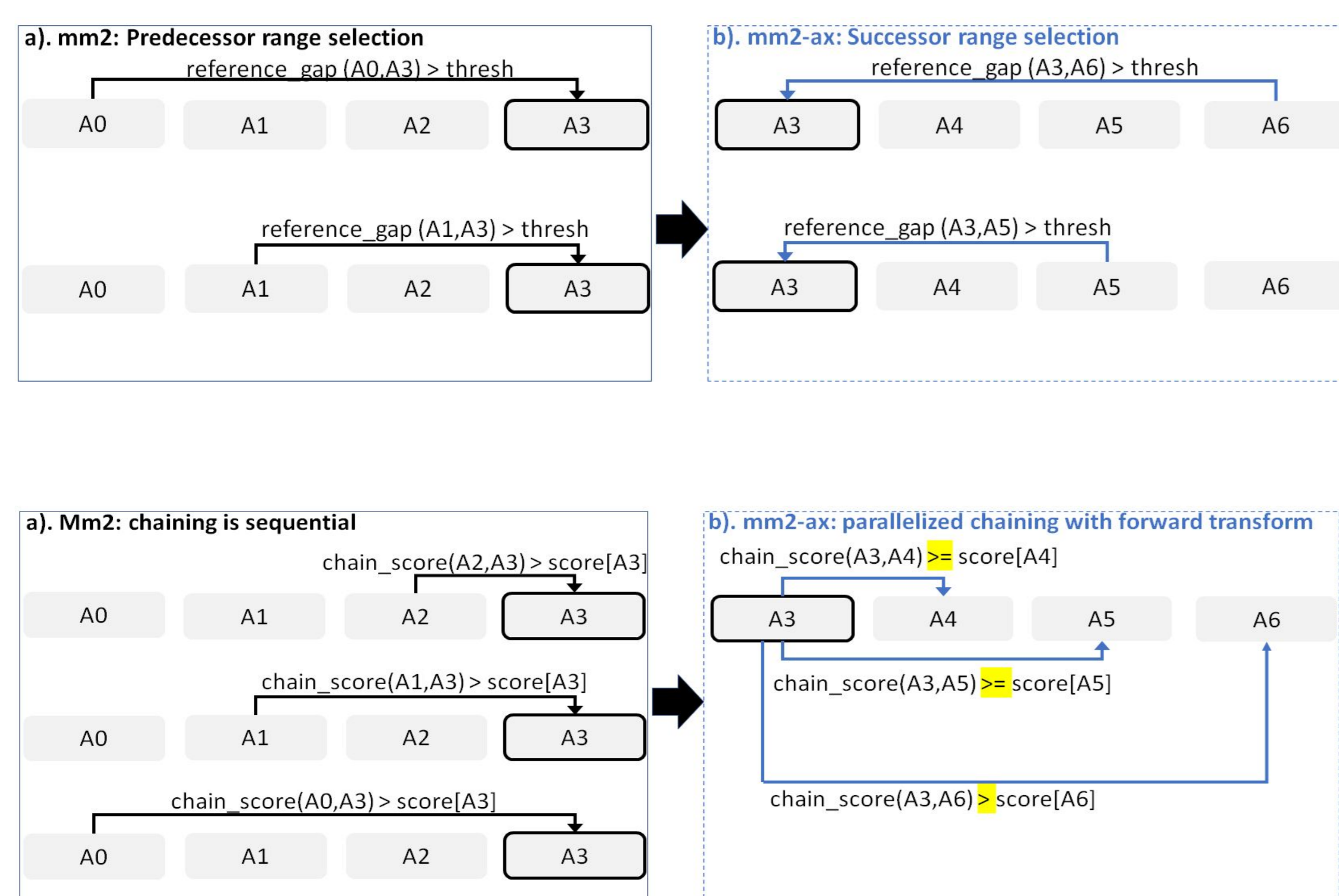


Irregular: 0-5000 neighbor seeds considered for chaining



METHODS

Forward chaining transformation enables better parallelization without losing mapping accuracy



Dynamic successor range selection:

- Finds successor range to attempt chaining
- Retains mapping accuracy

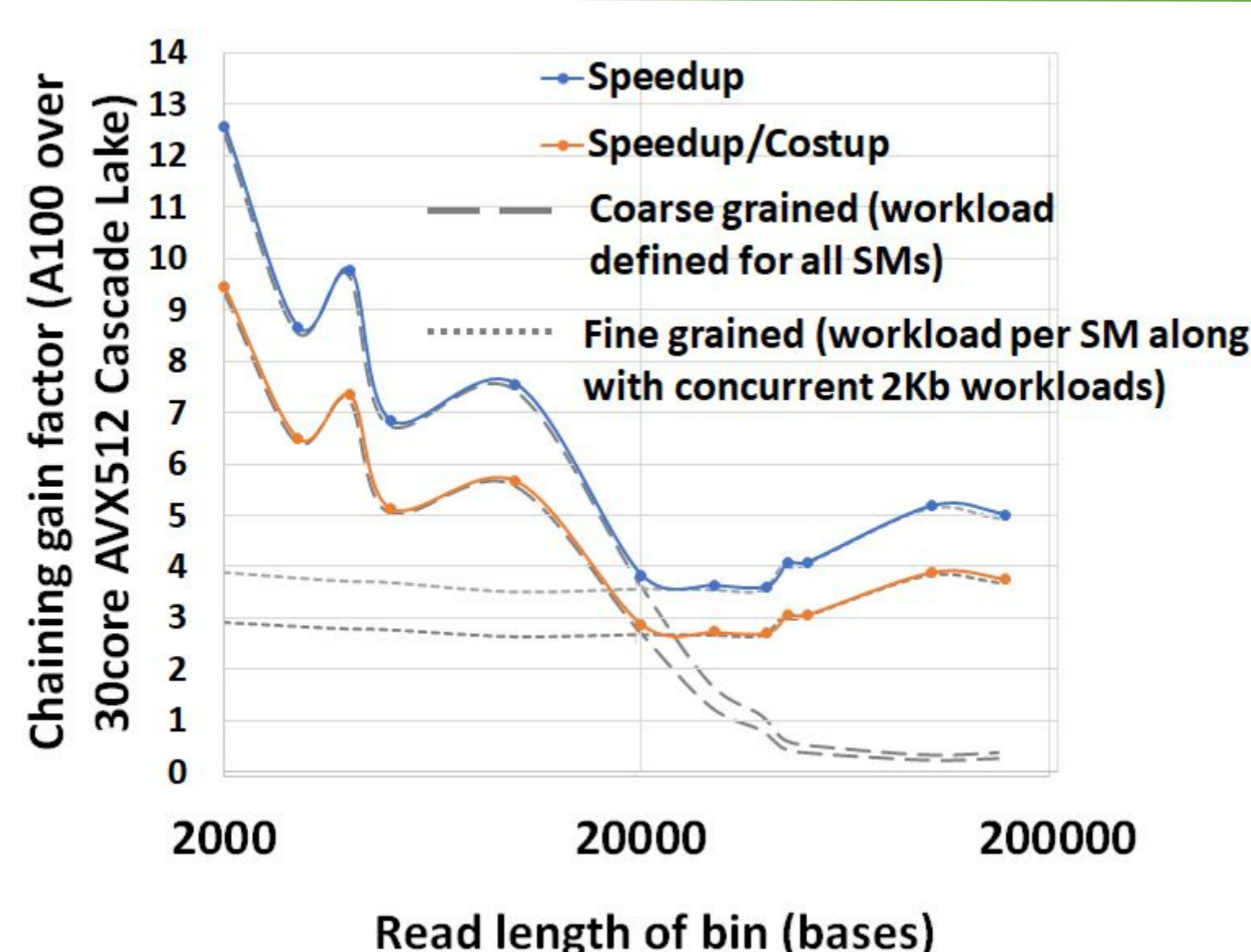
Forward chain score generation :

- Better intra-range parallelism
- Minimal branch divergence
- Exploit spatial data locality

Better GPU occupancy:

- Reads of similar length batched together
- Multiple concurrent reads per CUDA block
- Condensed chaining workload

RESULTS



- ~12.6 - 5X Speedup
- ~9.44 - 3.77X Speedup: Costup
- 99.998% primary alignments match



<sup>1</sup>Compute Science and Engineering, University of Michigan, Ann Arbor, MI 48109, USA  
<sup>2</sup>NVIDIA Corporation, Santa Clara, CA 95051, USA

Harisankar Sadasivan<sup>1</sup>, Milos Maric<sup>2</sup>, Eric Dawson<sup>2</sup>, Vishanth Iyer<sup>2</sup>, Johnny Israeli<sup>2</sup>, and Satish Narayanasamy<sup>1</sup>