75,000,000,000 Streaming Inserts/Second Using Hierarchical Hypersparse GraphBLAS Matrices

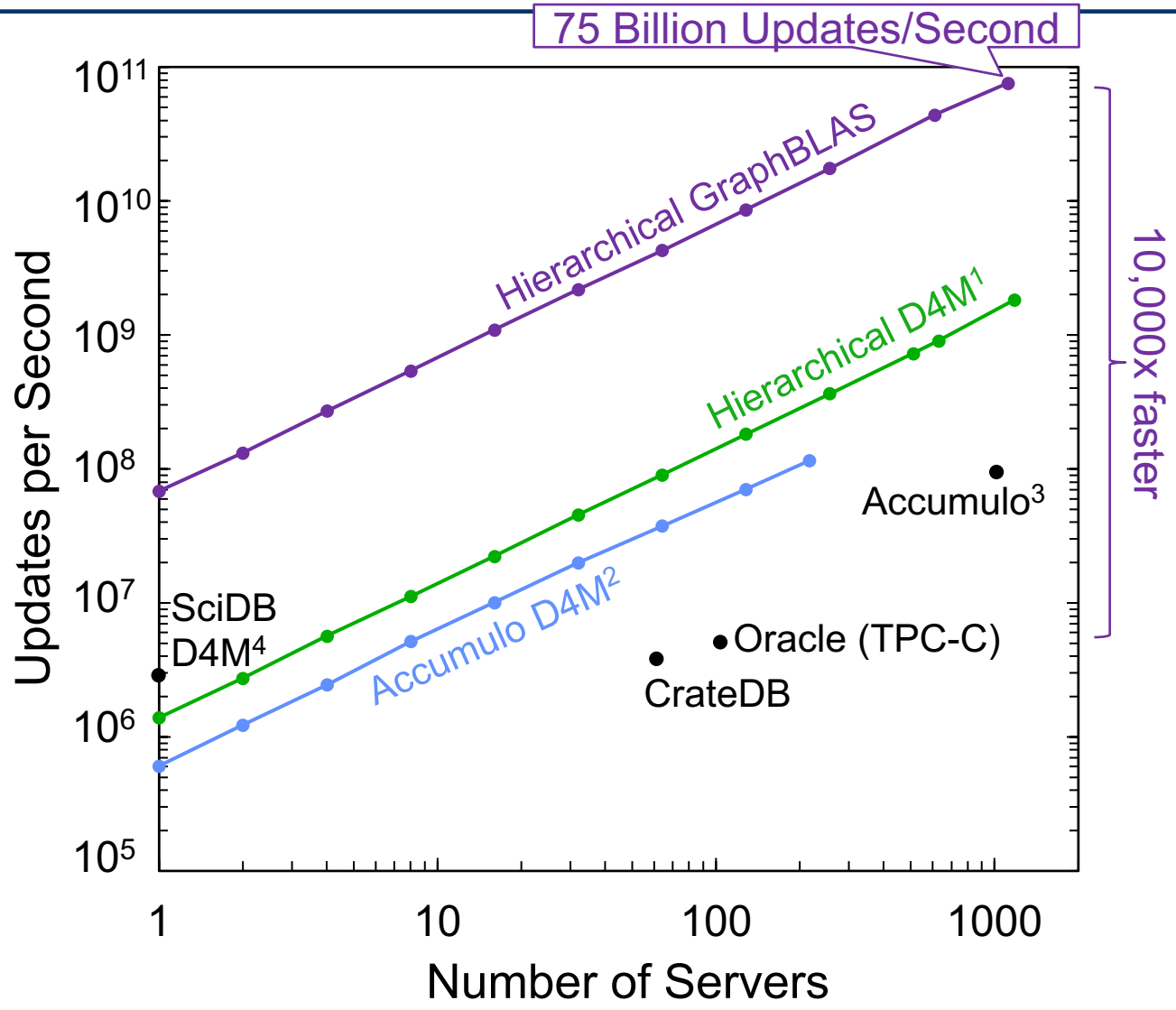
Jeremy Kepner, Tim Davis, Chansup Byun,
Siddharth Samsi, William Arcand, David Bestor, William Bergeron, Vijay Gadepally,
Matthew Hubbell, Micheal Houle, Micheal Jones, Anne Klein, Peter Michaleas, Lauren Milechin, Julie Mullen, Andrew Prout, Antonio Rosa, Charles Yee, Albert Reuther

May 2020

MIT LINCOLN LABORATORY
SUPERCOMPUTING CENTER

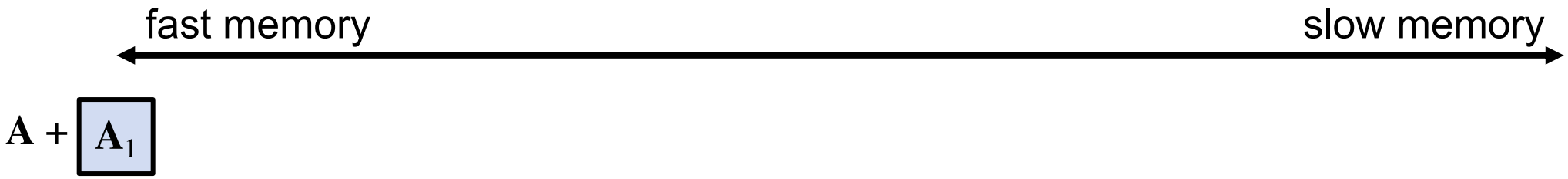
Hierarchical GraphBLAS In-Memory Database Performance

- **High performance streaming network analysis requires**
  - **High level programming environments**
  - **Rapid interaction and fast turnaround**
- **MIT SuperCloud optimizes every aspect of HPDA system to enable**
  - **34,000+ hierarchical GraphBLAS in-memory databases**
  - **1000+ servers**
  - **Launched in seconds**
- **Achieves 75 billion updates/seconds**
  - **Smashes prior Hierarchical D4M world record of 1.9 billion inserts/second[1]**
  - **10,000x faster than standard databases**
  - **20x the packet rate of projected 2022 North American non-Video Internet***



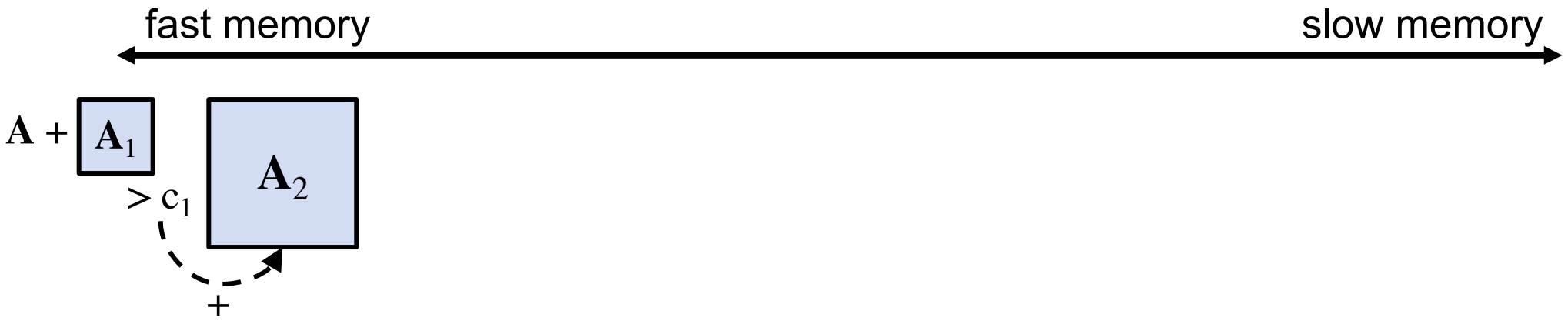
*Cisco Visual Networking Index: Forecast and Trends, 2017–2022; [1]*Streaming 1.9 Billion Hypersparse Network Updates per Second with D4M*, Kepner et al, IEEE HPEC 2019; [2]*Achieving 100,000,000 database inserts per second using Accumulo and D4M*, Kepner et al, IEEE HPEC 2014; [3]*Benchmarking Apache Accumulo BigData Distributed Table Store*, Sen et al, IEEE Big Data Congress 2013; [4]*Benchmarking SciDB Data Import on HPC Systems*, Samsi et al, IEEE HPEC 2016

Hierarchical GraphBLAS Arrays

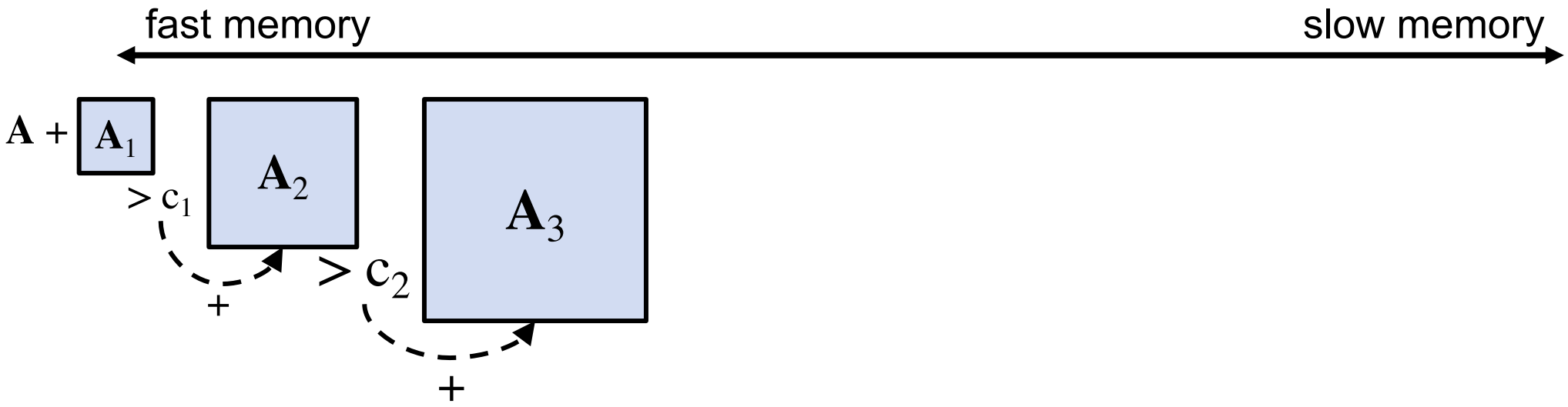


Hierarchical GraphBLAS Arrays

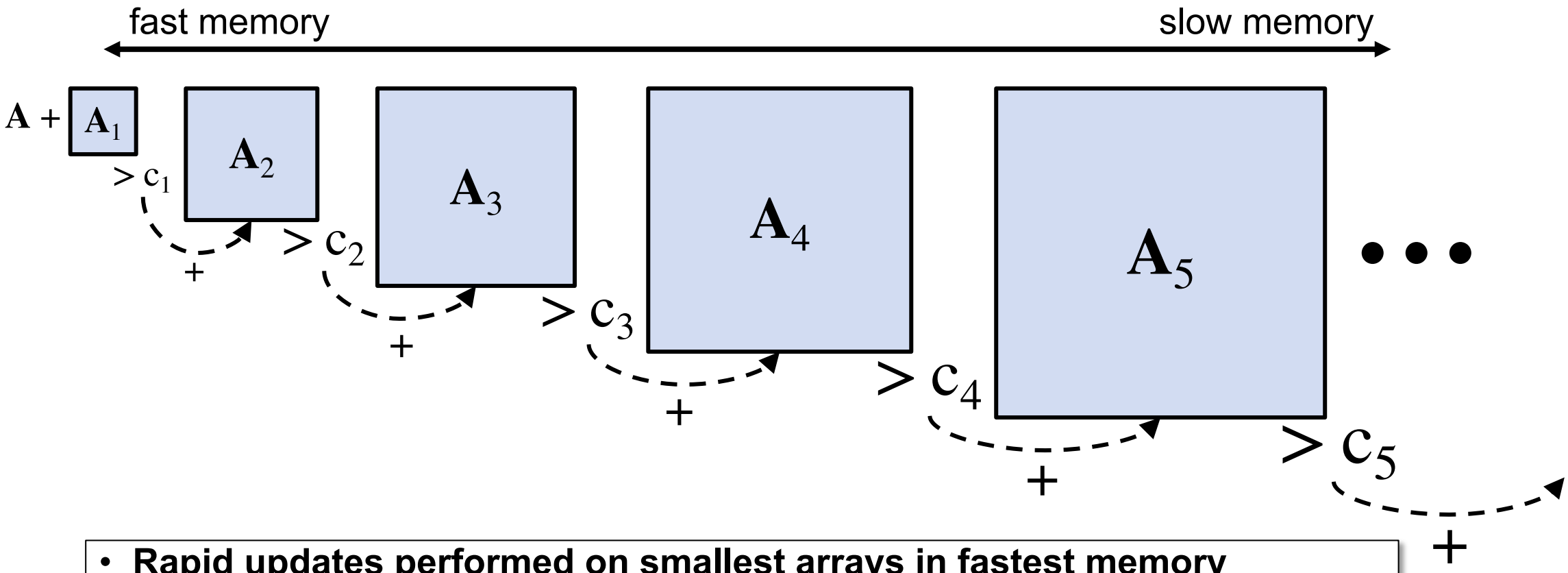
Hierarchical GraphBLAS Arrays



Hierarchical GraphBLAS Arrays



Hierarchical GraphBLAS Arrays



- **Rapid updates performed on smallest arrays in fastest memory**
- **If number of entries exceeds c_i, then A_i is added to A_{i+1} and A_i is cleared**
- **Dramatically reduces the number of updates to slow memory**
- **Upon query, all are summed into largest array**