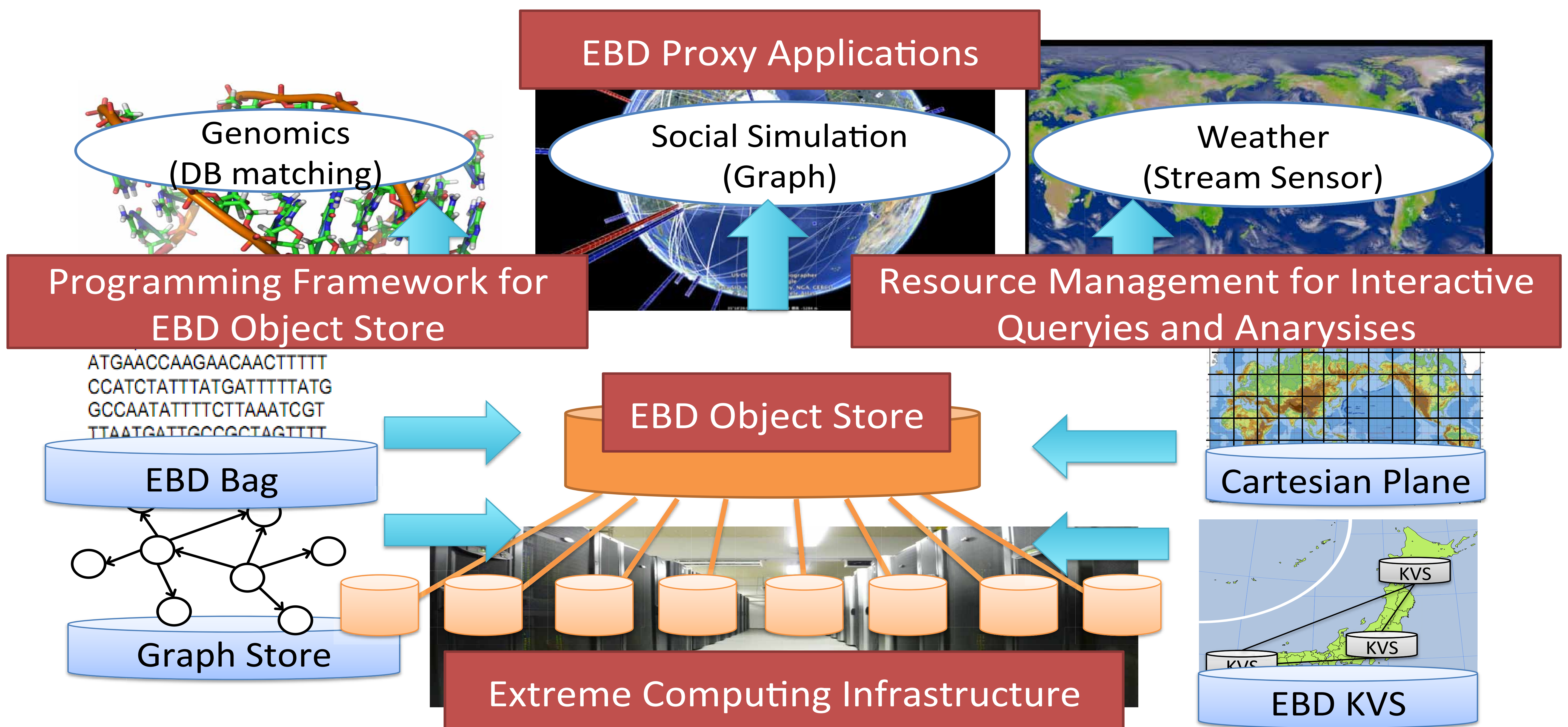




# Big Data and Extreme Computing

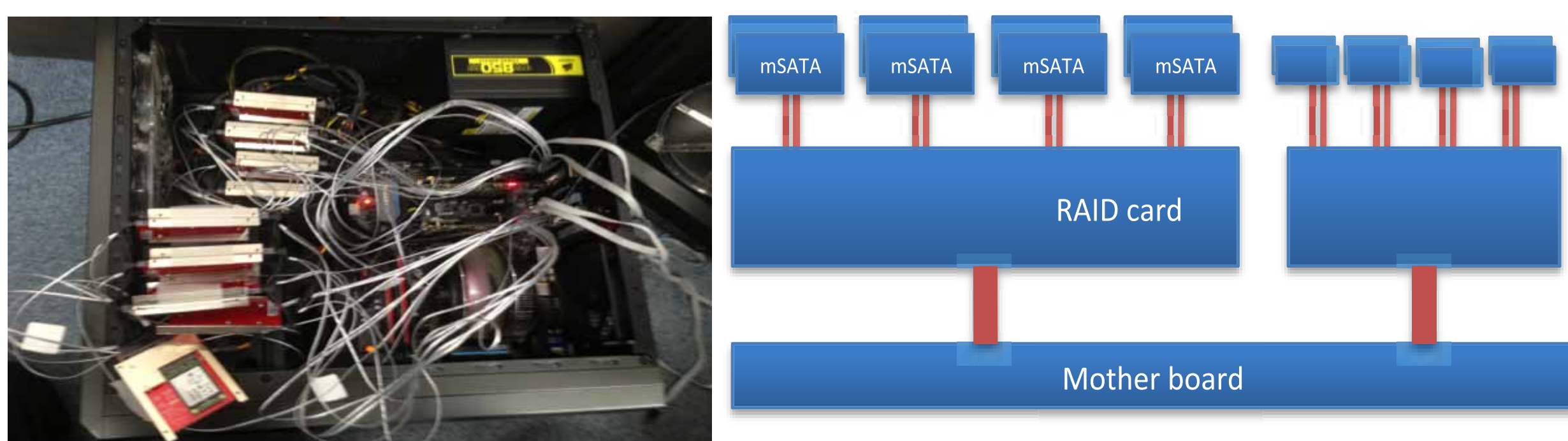
## Extreme Big Data (EBD) Infrastructure for Data Science



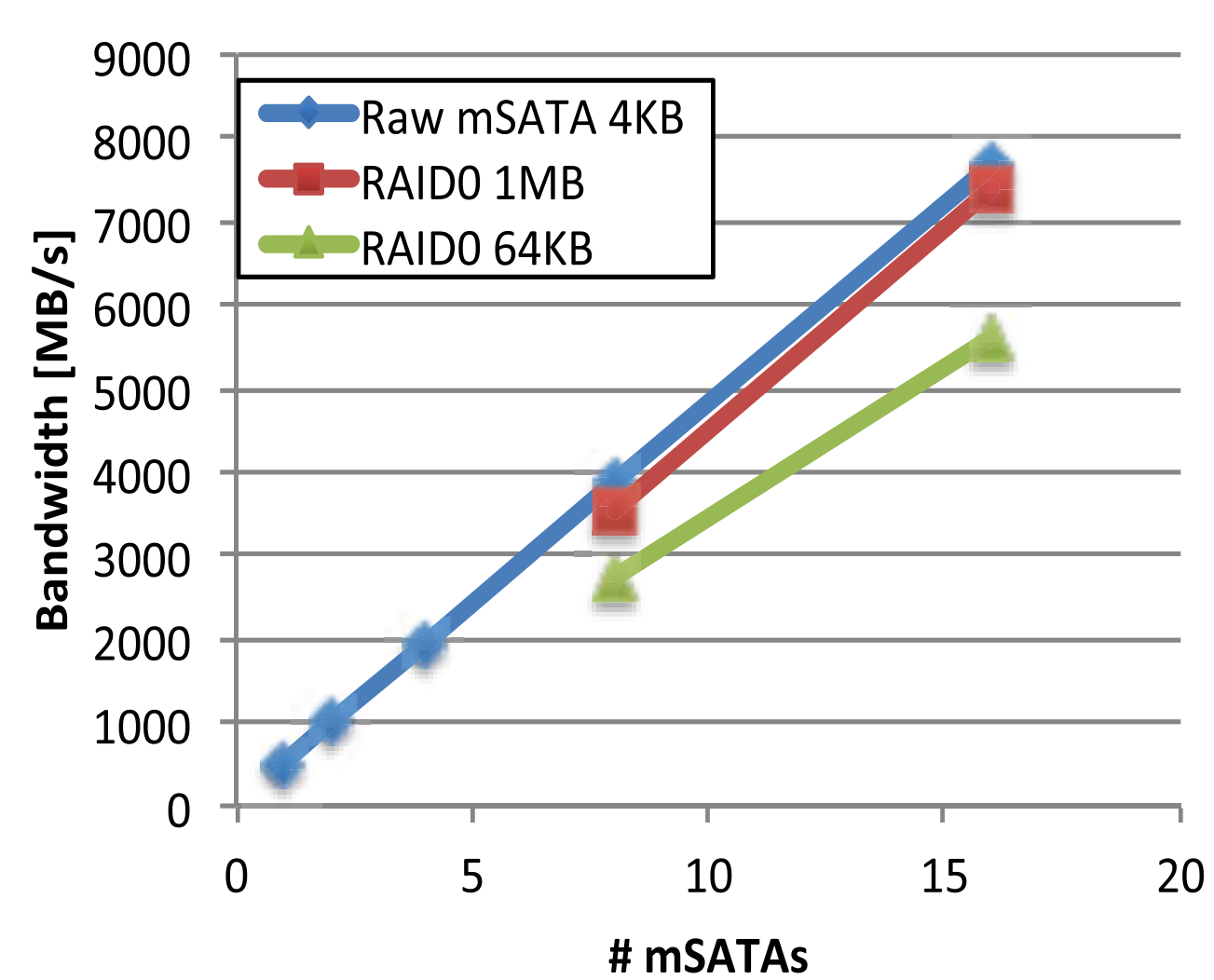
## EBD I/O

How to design local storage for next-gen supercomputers?

- EBD I/O prototype using local 16 mSATA SSDs
- Capacity: **4TB**
- Read Bandwidth: **8GB/s**

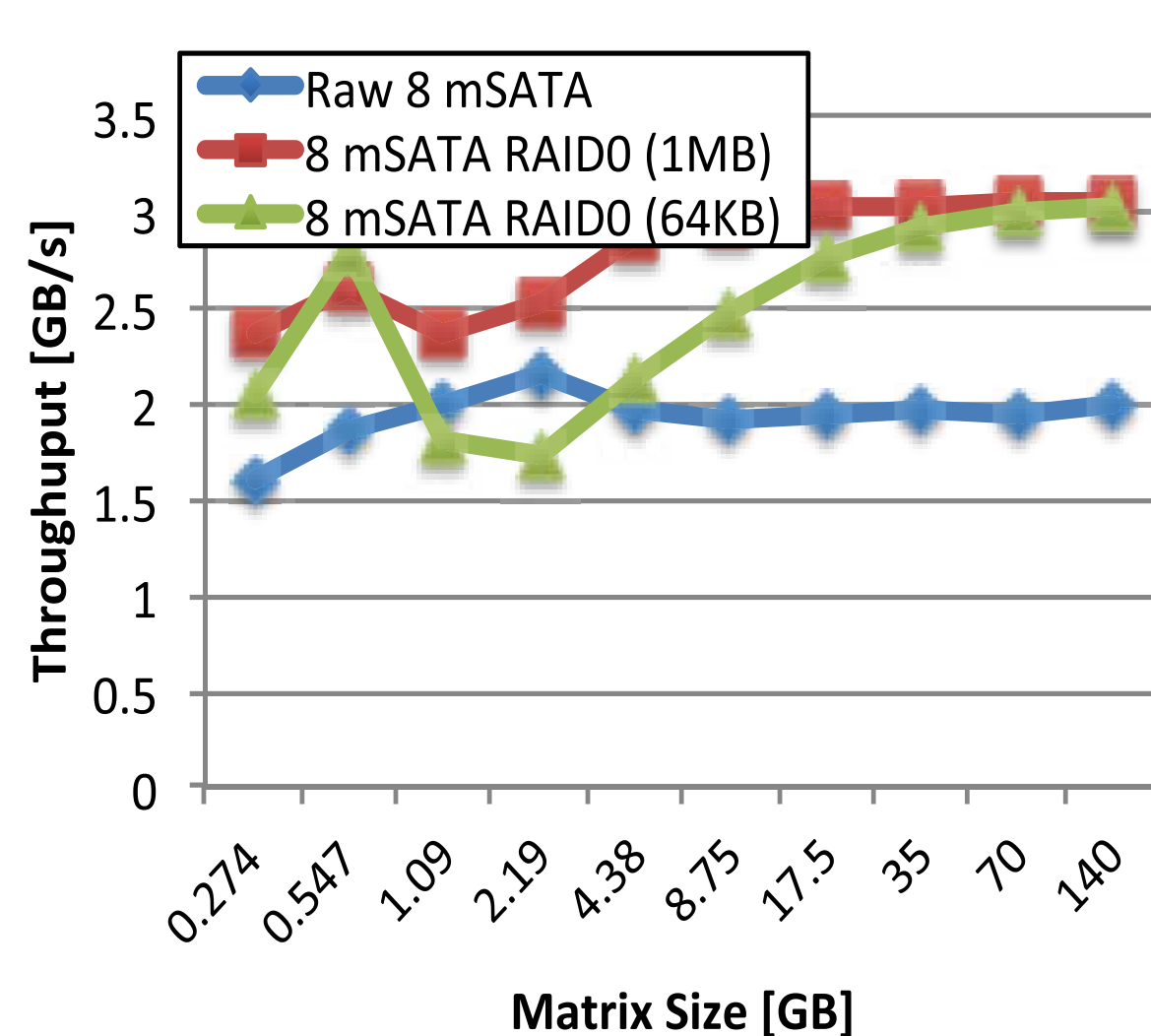


I/O performance of multiple mSATA SSD



~ 7.39 GB/s from 16 mSATA SSDs (Enabled RAID0)

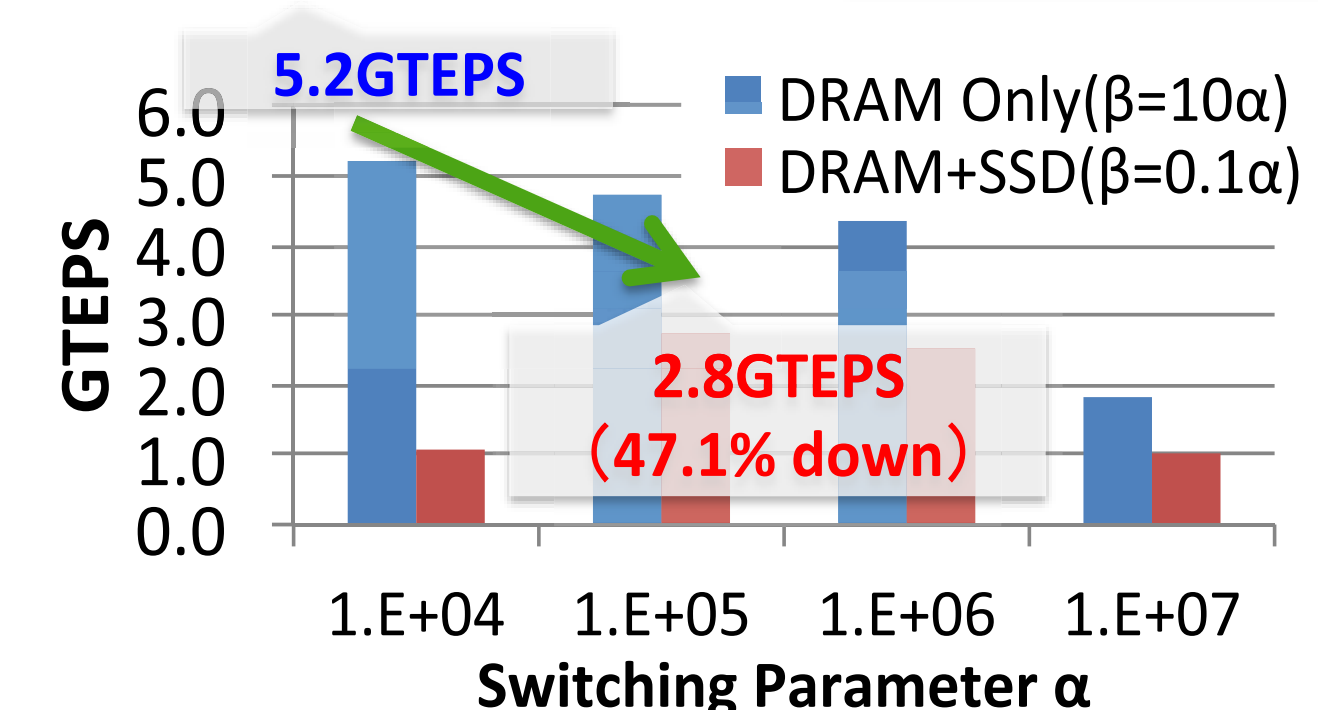
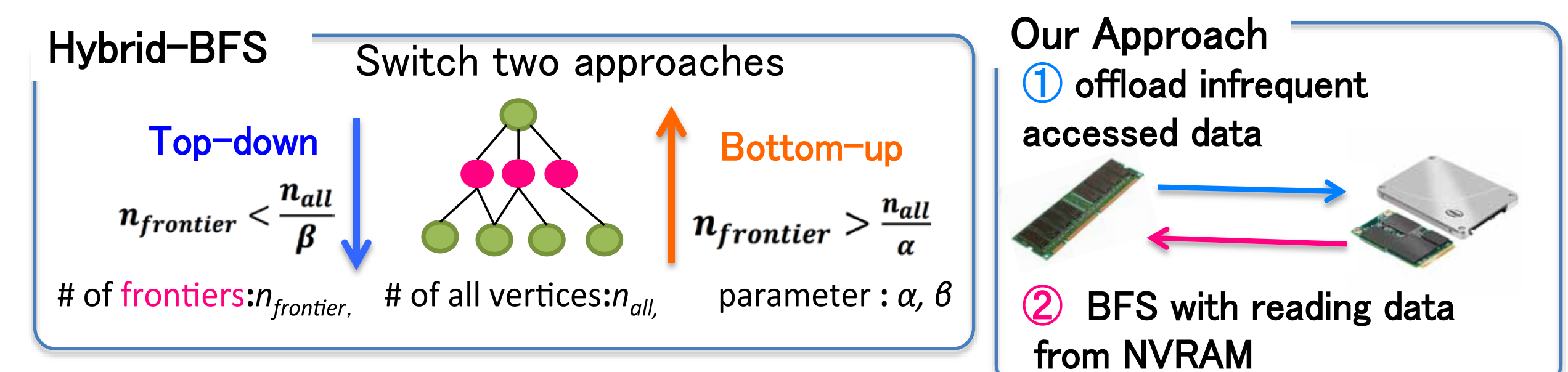
I/O performance from GPU to multiple mSATA SSDs



~ 3.06 GB/s from 8 mSATA SSDs to GPU

## Graph

Hybrid BFS Approach Using NVRAM for Extreme Large-scale Graphs



## Sort

High Performance Sorting Comparison-based (quicksort, mergesort, etc.)

Distribution Sorts: not cmp based -> linear/sublinear time

MSD radix sort variable length keys high efficiency on small alphabets long keys

LSD radix sort (THRUST) good for short length keys only fixed-length keys

integer sorts GPUs are good at counting numbers

Computational Genomics (A,C,G,T) don't have to examine all characters

Single node/device performance

