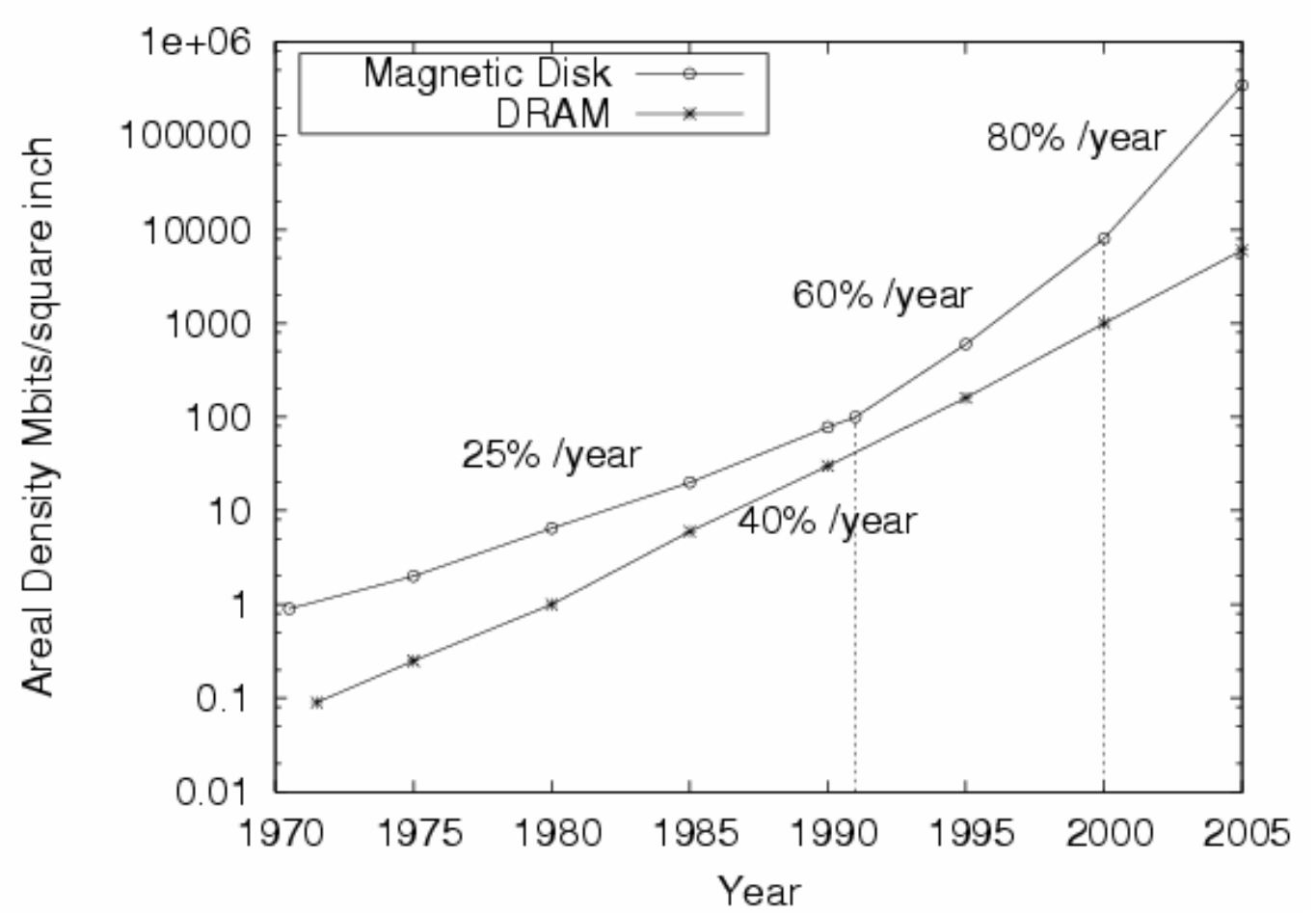
High Performance Embedded Computing
September 2005

Embedding Applications within a Storage Appliance

Roger D. Chamberlain

Exegy Inc.
and
Washington University in St. Louis

Technology Trend #1: Storage Density

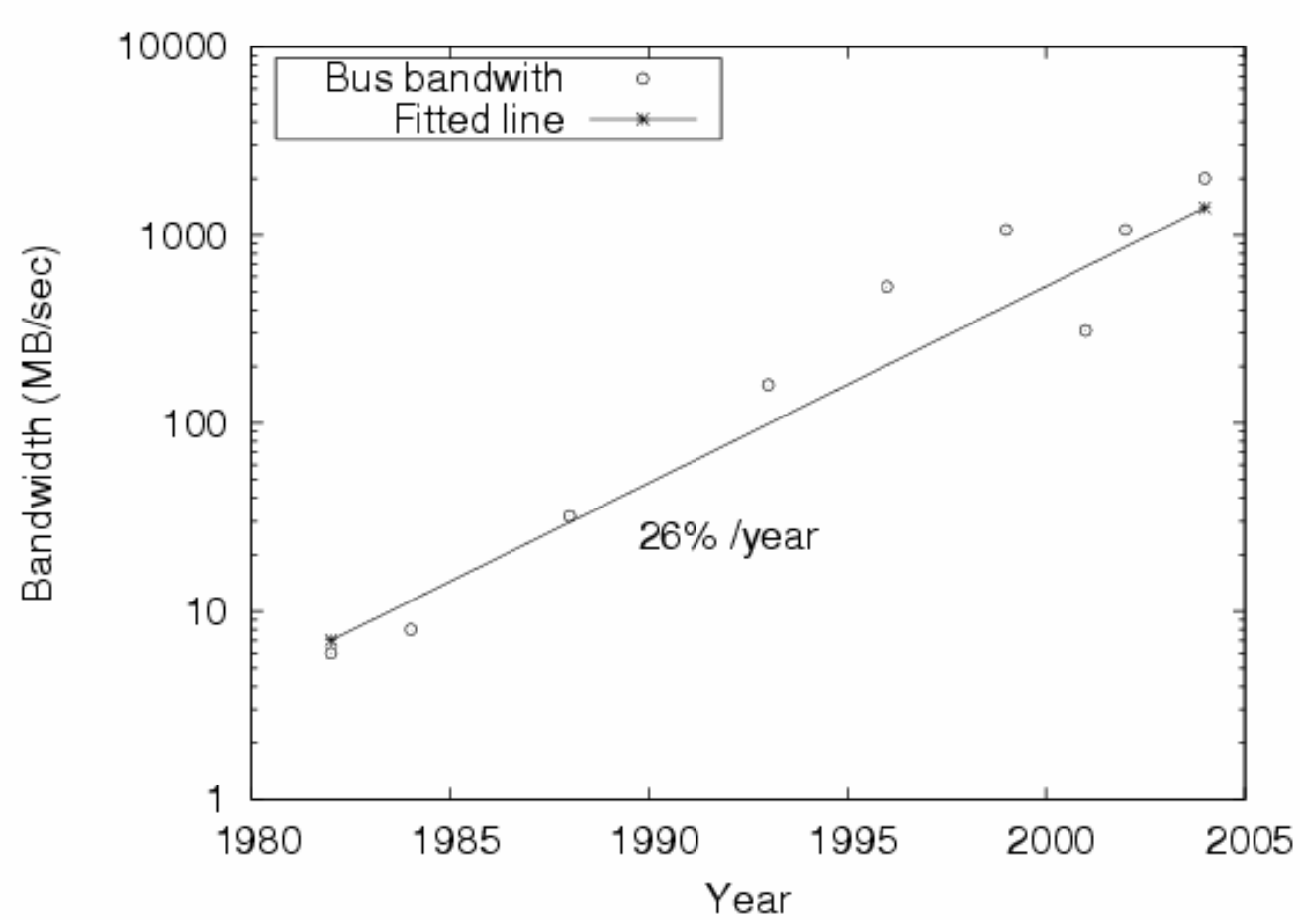


Disk density increasing faster than Moore's Law

EXEGY™

Washington University in St. Louis

Technology Trend #2: Interconnect Bandwidth



Increase is much slower than storage density improvements

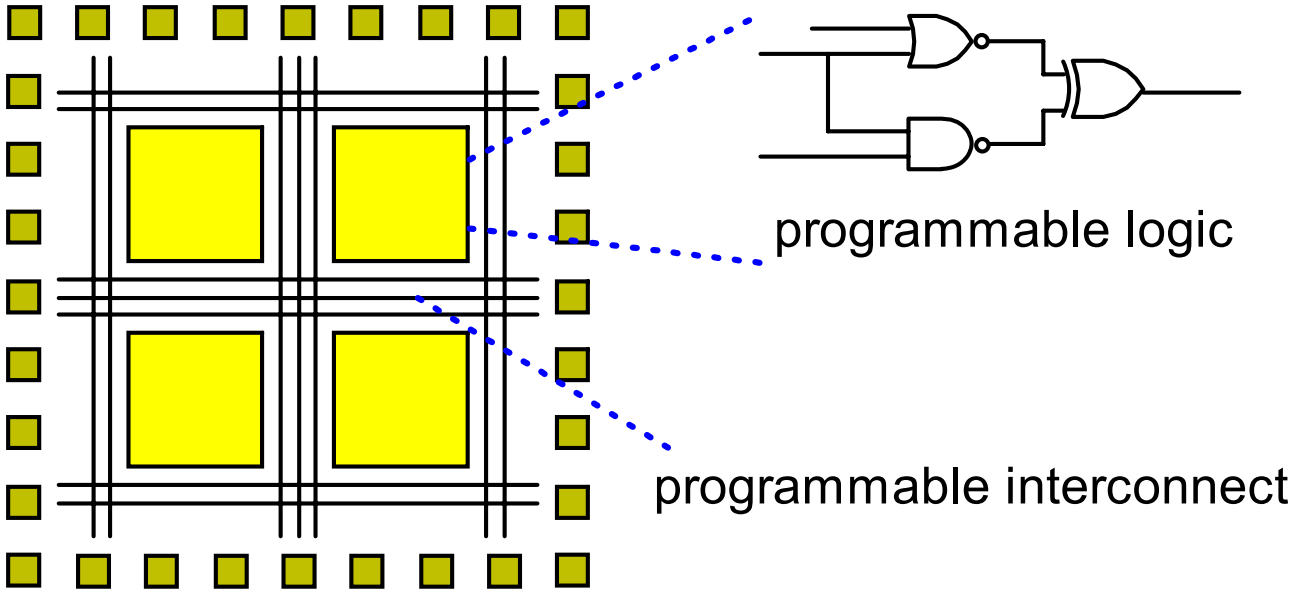
EXEGY™

Washington University in St. Louis

Massive Data

- Storage industry will ship 18,000,000,000,000,000,000 bytes this year
- "The size of the databases we now deal with is no longer measured in terabytes, but in exabytes," John Reynders, CIO Celera Genomics, at 2002 HPEC
- VERITAS gamma-ray telescope will soon generate 24 terabytes/year of event data

Enabling Technology: Reconfigurable Hardware



- Field Programmable Gate Arrays (FPGAs) provide custom logic function capability
- Operate at hardware speeds
- Can be altered (reconfigured) in the field to meet specific application needs

What are we doing?

We are combining the capabilities of these two enabling technologies to build extremely fast data processing engines.

We do this by moving the processing closer to the data, and performing it in hardware rather than software.

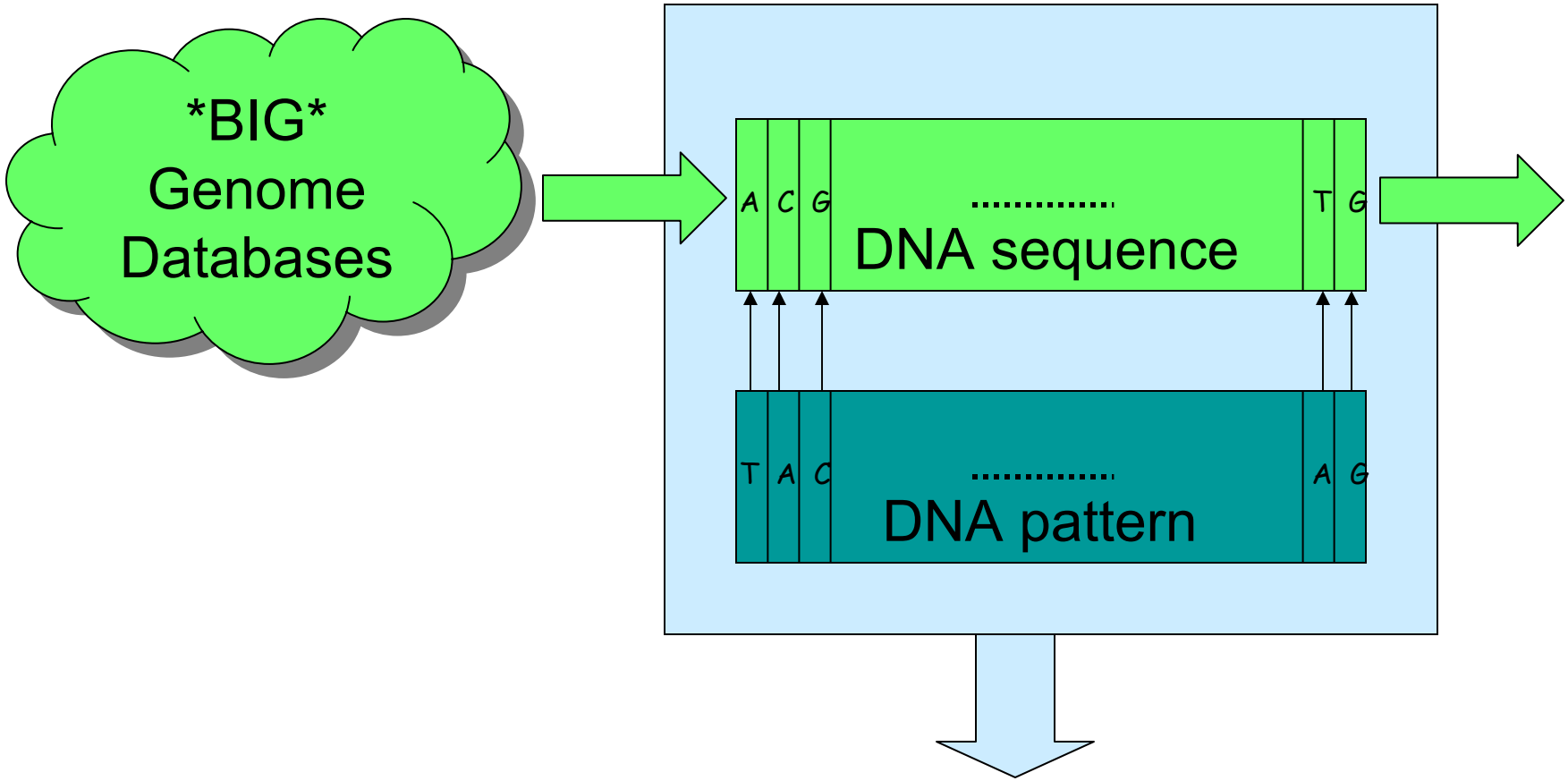
EXEGY™

Washington University in St. Louis

Applications

- Biosequence search
- Science data mining
- Text search (exact and approximate)
- Compression/decompression
- Encryption/decryption
- Structured record search
- Signature hashing

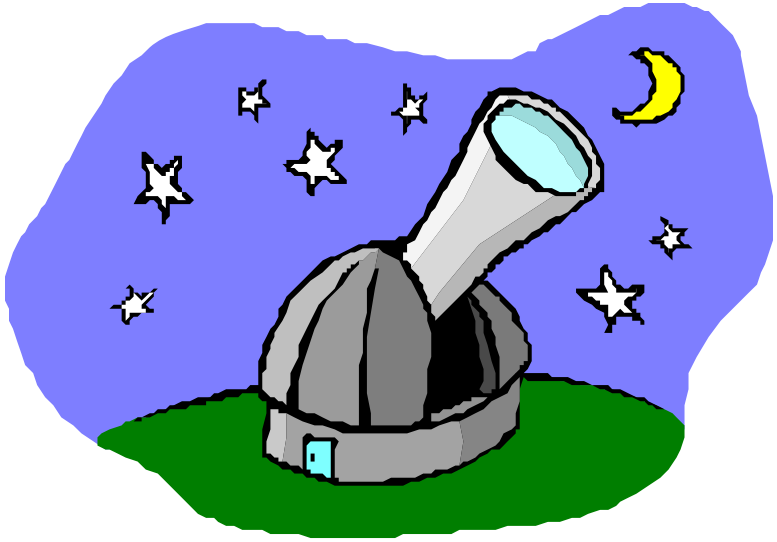
Biosequence Similarity Search



EXEGY™

Washington University in St. Louis

VERITAS Telescope



- Array of 12 m telescopes being constructed in Arizona.
- Looking for Cherenkov radiation from 50 GeV to 50 TeV gamma-ray interactions with upper atmosphere.
- Early indicator of supernovae, so timely data analysis is central to scientific mission.

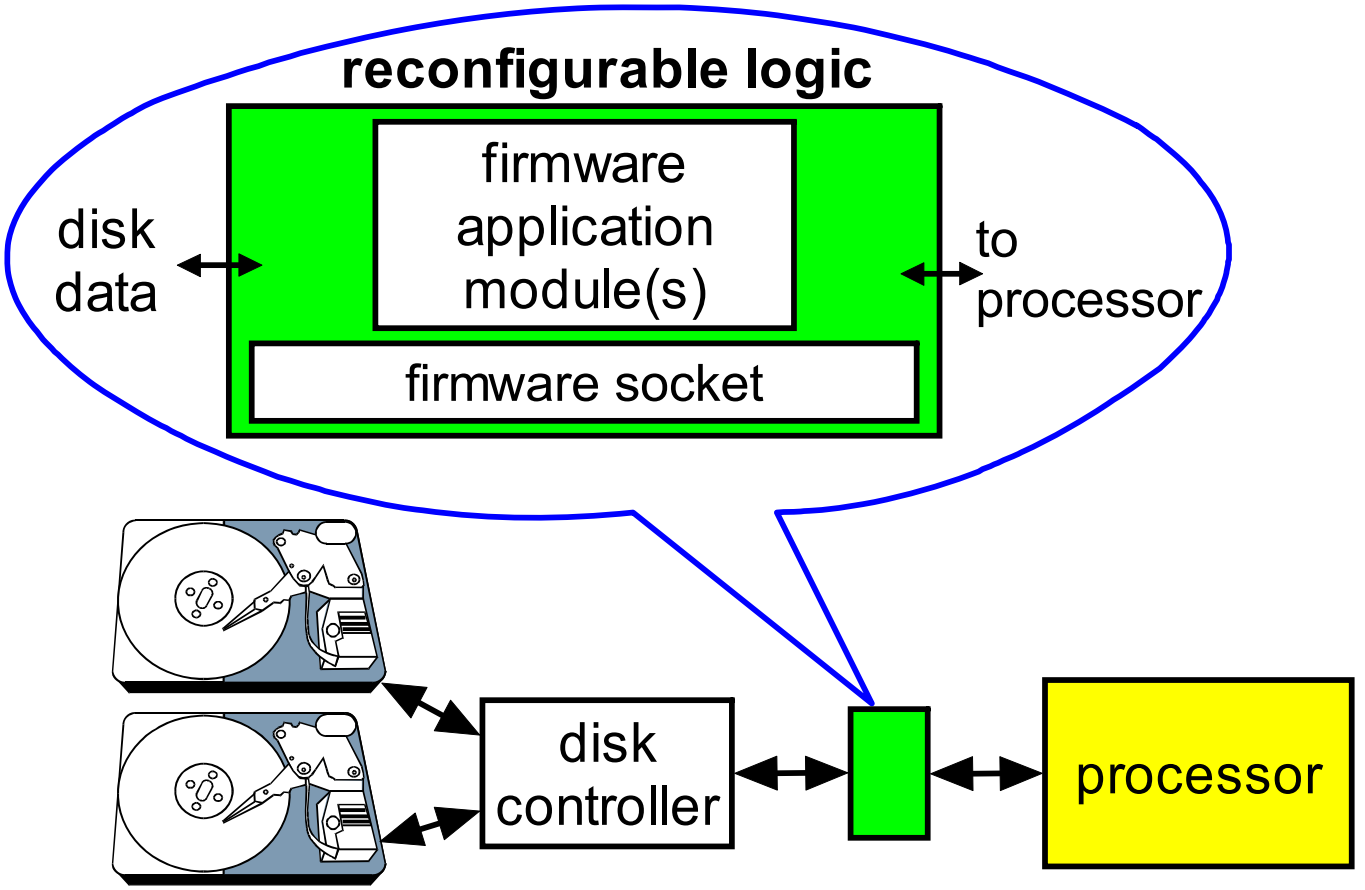
EXEGY™

Washington University in St. Louis

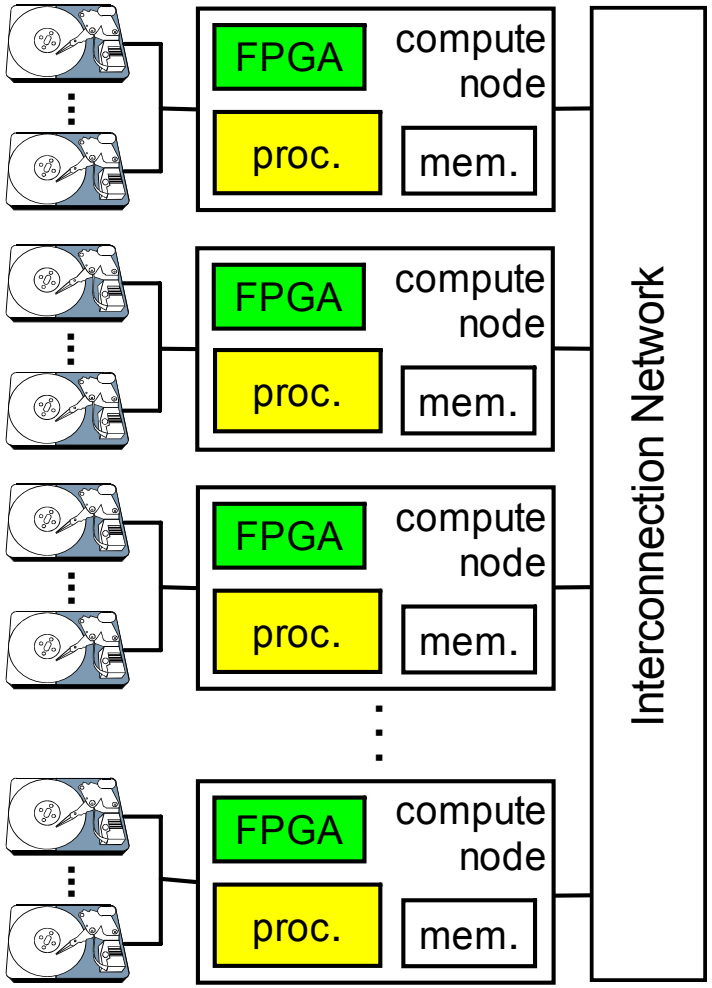
Intelligence Data

- Lots of data
  - Public (e.g., web pages)
  - Clandestine (e.g., via national technical means)
- Growing constantly
- Many perturbations of individual words
  - Tzar, Tsar, Czar, …
- Query and field types aren't known *a priori*

Individual Compute Node



Scaled System



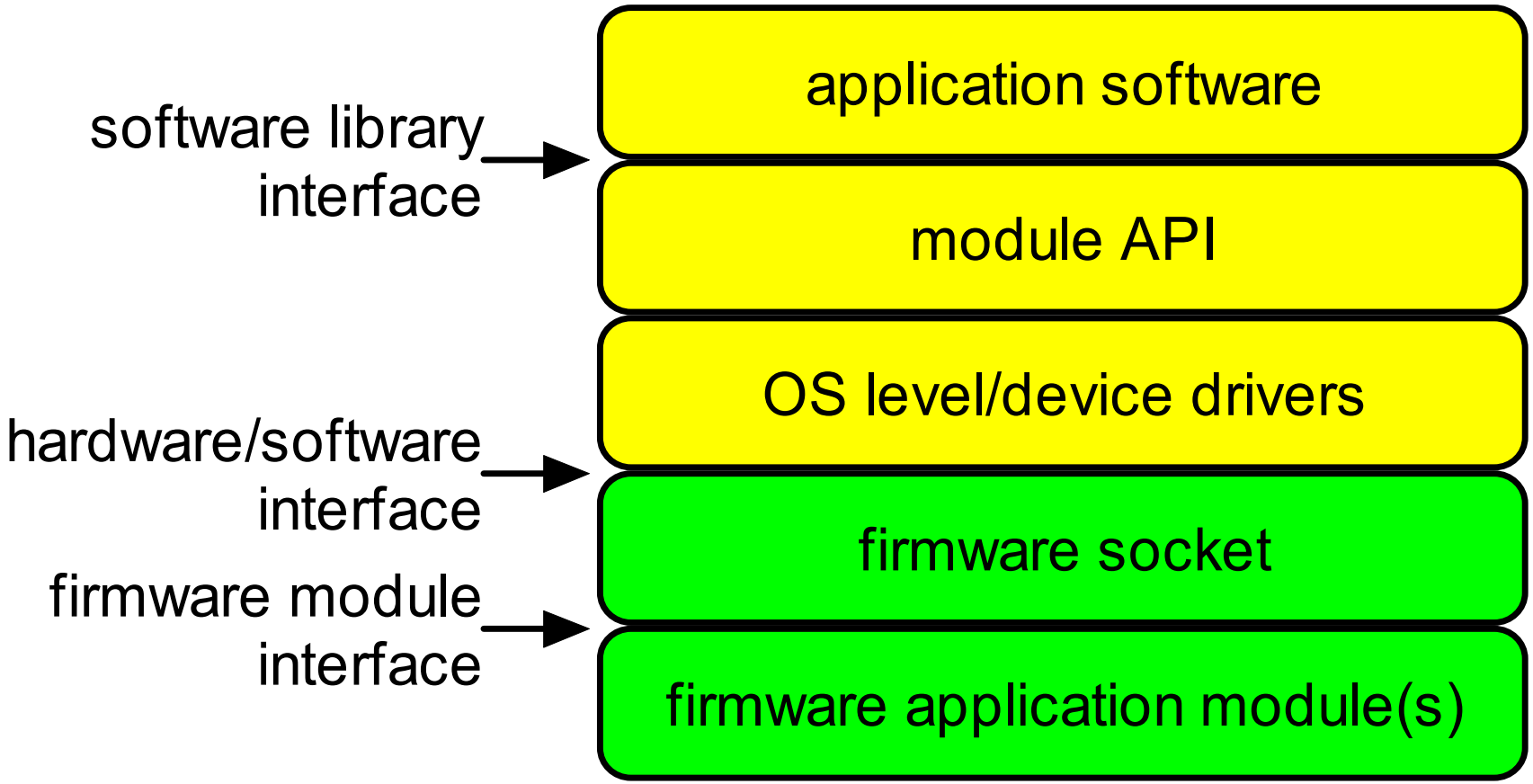
EXEGY™

Washington University in St. Louis

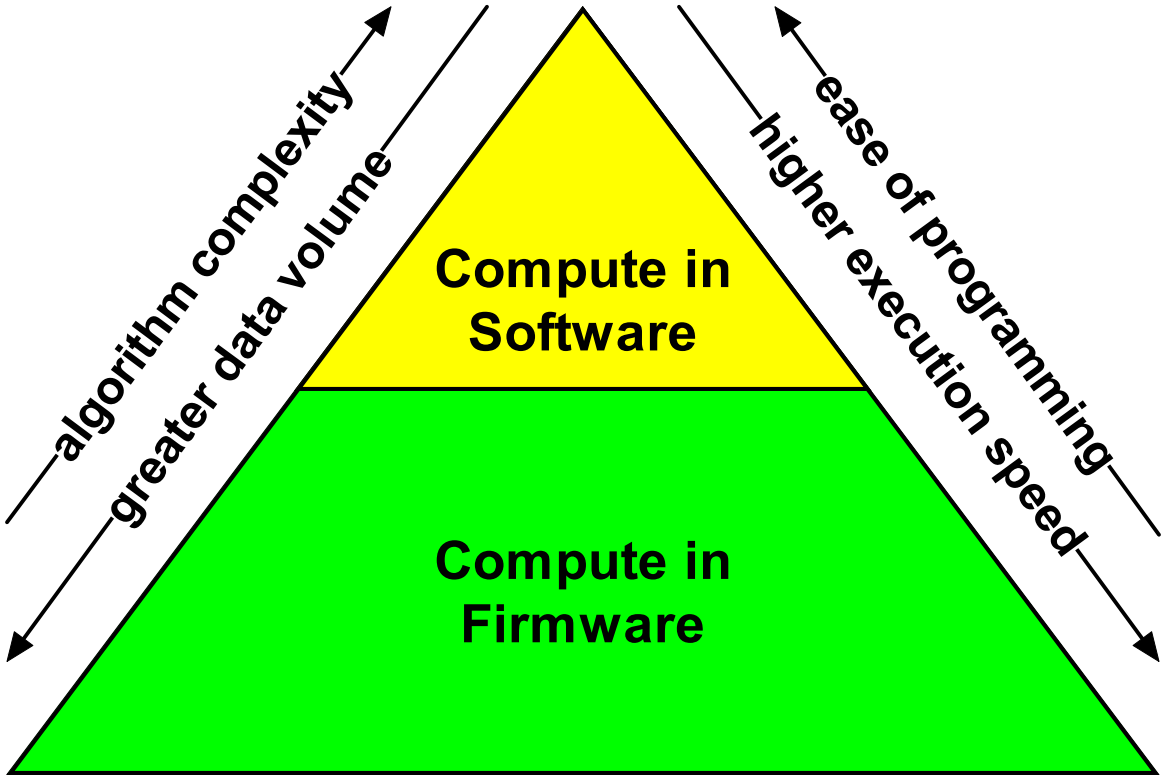
Challenges in Delivering Performance to Applications

- Application development
  - Hardware/software codesign problem
  - Partitioning of functions
- Delivering sustained data throughput
  - Any bottleneck causes throughput to drop

Application Development Framework

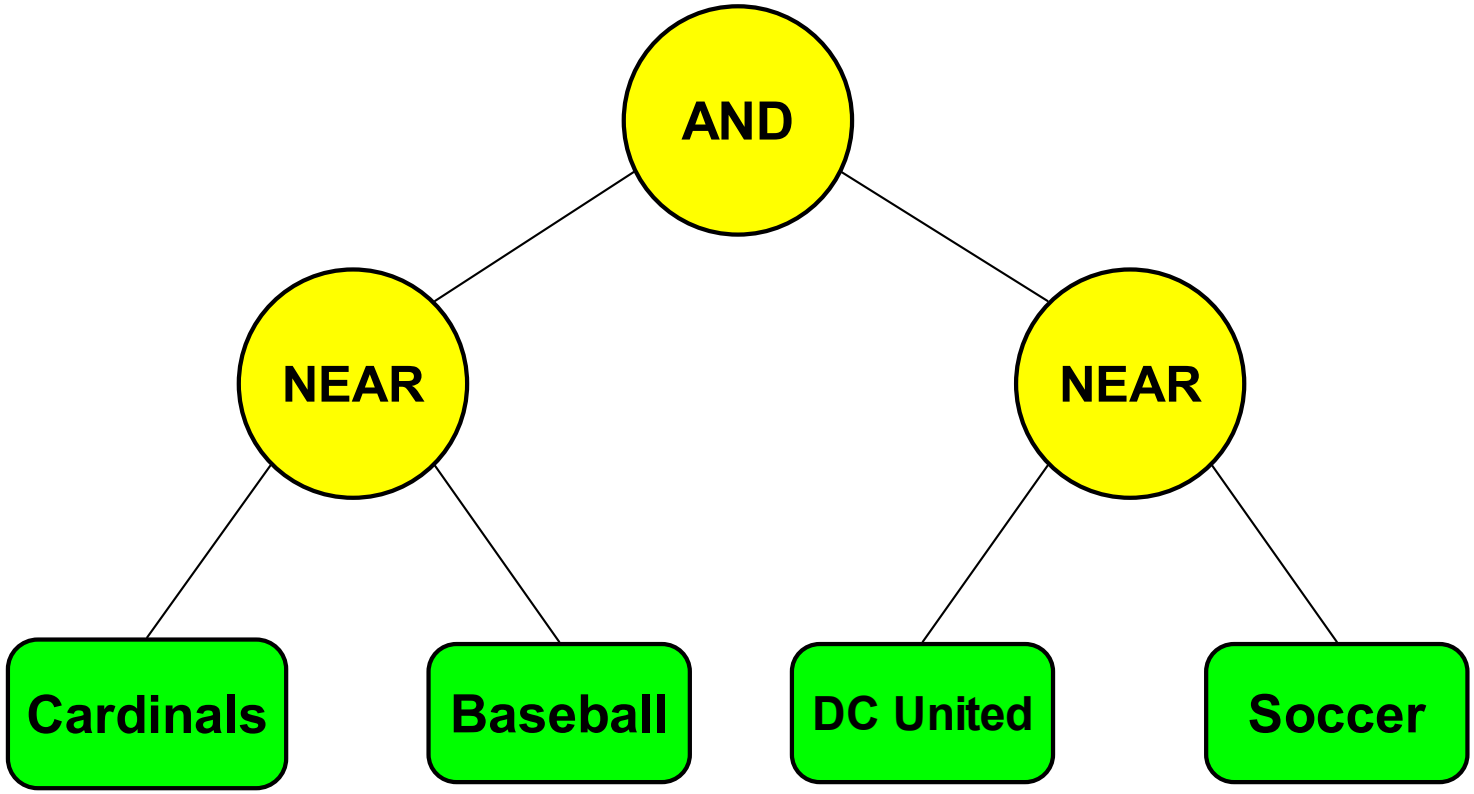


Partitioning an Application



Example Text Search Query

(Cardinals NEAR Baseball) AND (DC United NEAR Soccer)



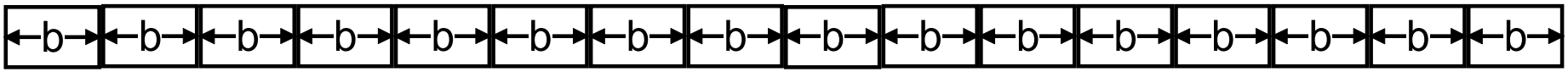
EXEGY™

Washington University in St. Louis

Delivering Sustained Data Throughput

- Layout on data store
  - Incrementally defragmenting file system
  - Data placement cognizant of application
- Extreme diligence in tuning, e.g.:
  - OS parameter settings impact performance in ways that are often surprising

Mostly Contiguous File System

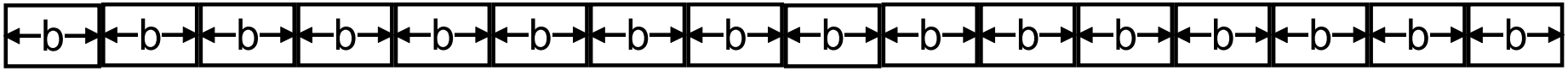


Traditional file system uses fixed-size blocks

EXEGY™

Washington University in St. Louis

Mostly Contiguous File System

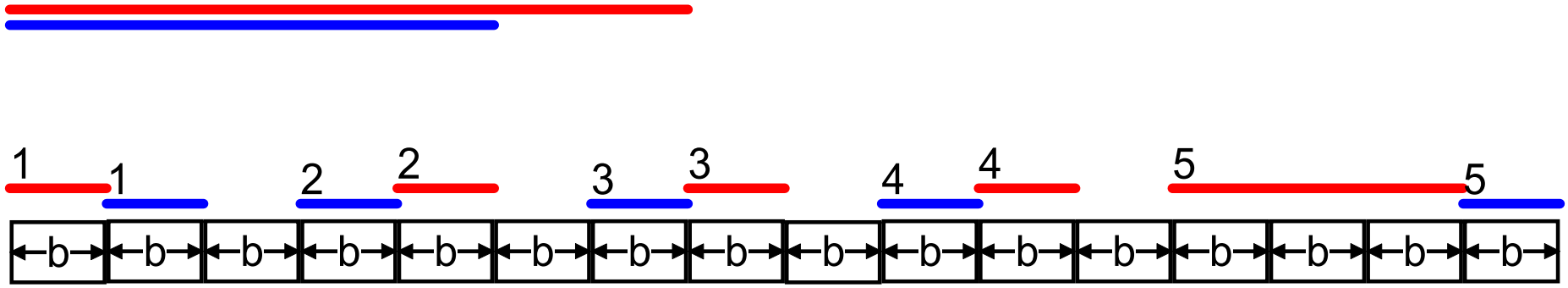


With two files to store, with file size *nb*, ...

EXEGY™

Washington University in St. Louis

Mostly Contiguous File System

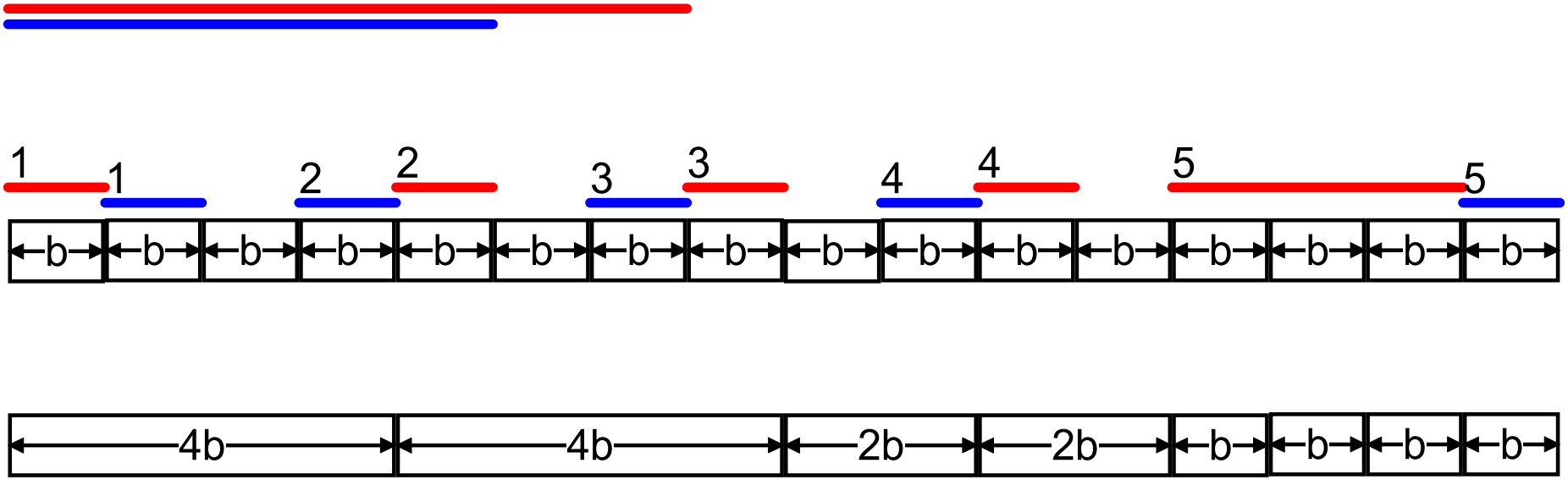


... the number of contiguous components is O(*n*)

EXEGY™

Washington University in St. Louis

Mostly Contiguous File System

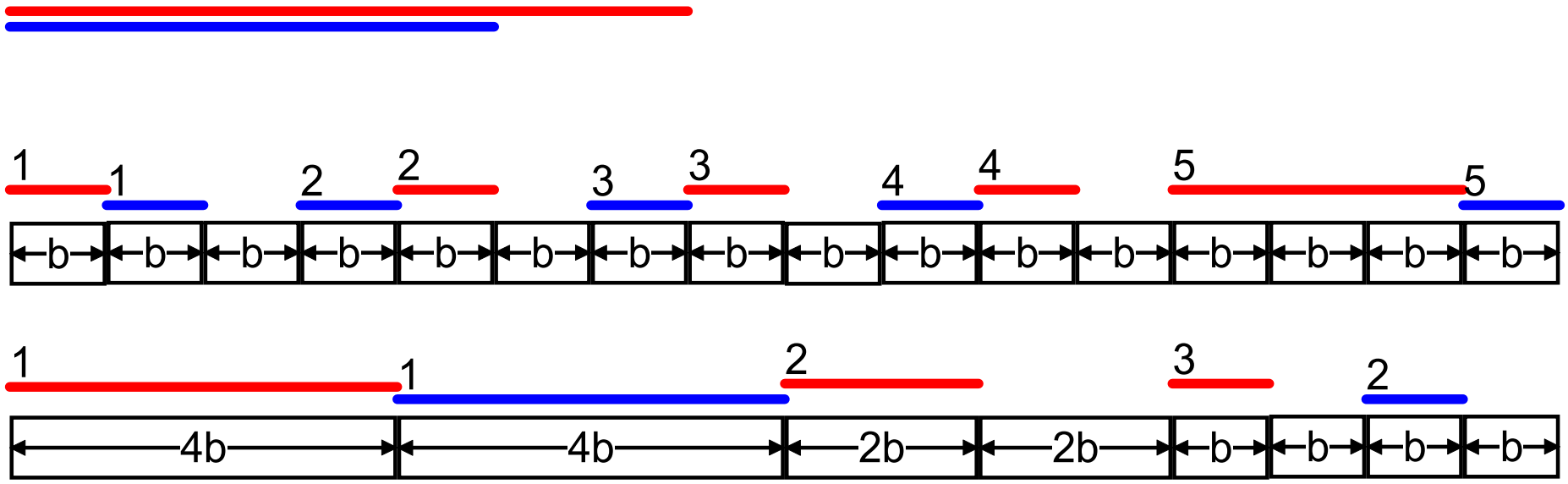


Our file system uses variable-size blocks ...

EXEGY™

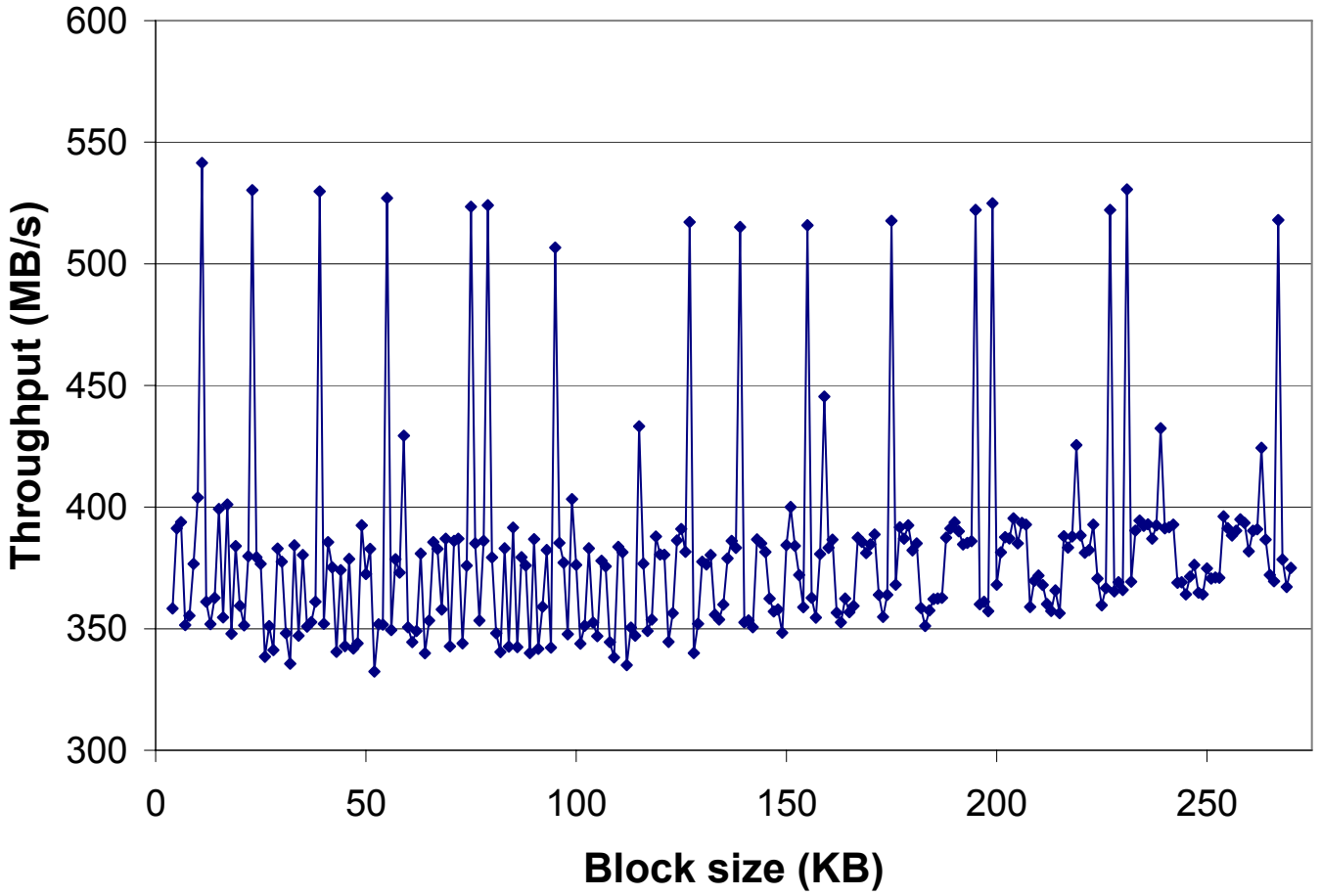
Washington University in St. Louis

Mostly Contiguous File System



... resulting in O(log *n*) contiguous components

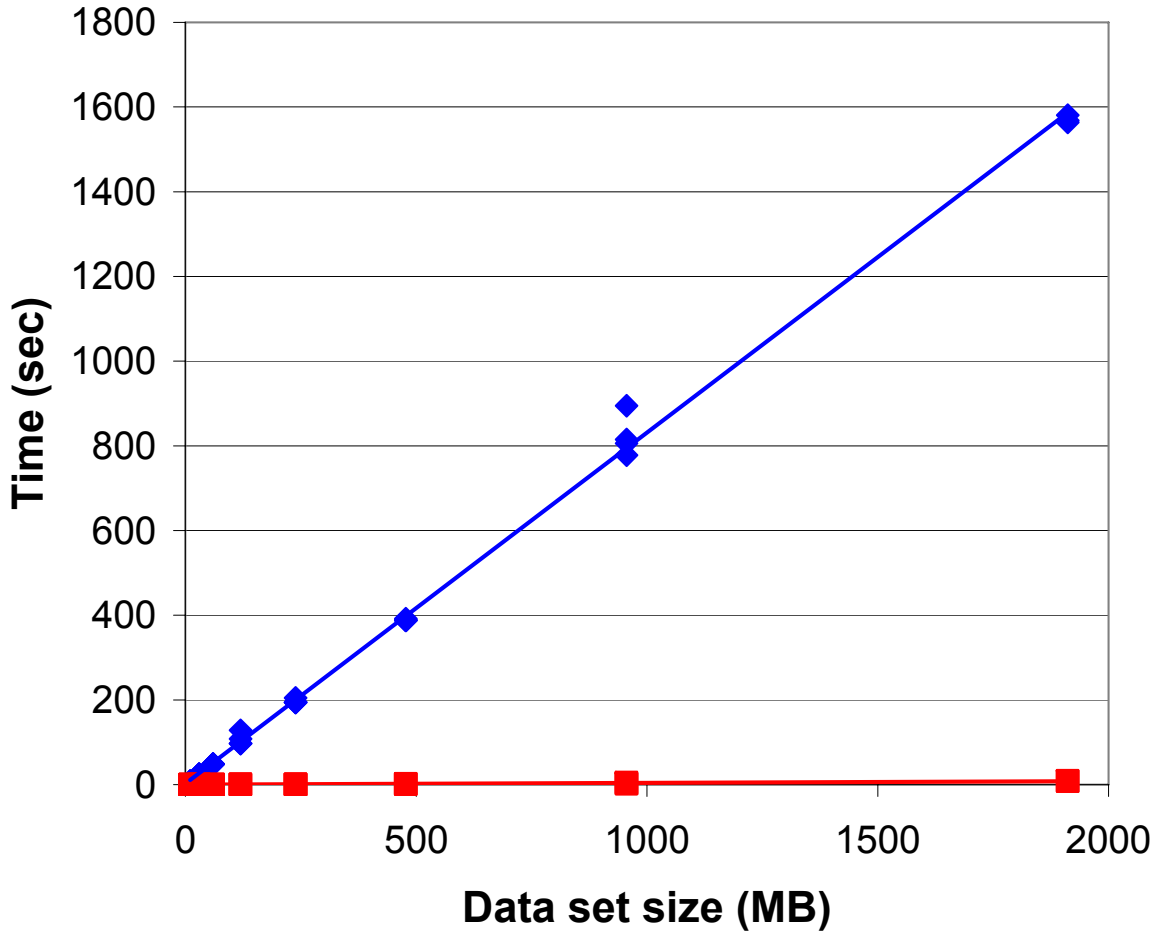
Example of Sensitivity to OS Parameters



EXEGY™

Washington University in St. Louis

Performance on Approximate Text Search Application



Firmware Speedup: 200

EXEGY™

Washington University in St. Louis

Summary

- We are exploiting both inexpensive, high-volume storage and reconfigurable hardware technology
- 750 MB/s sustained throughput achieved on single node
- Performance is scalable and uses conventional disk drives
- Framework for application development
- Data layout on disks is important
- Result is applications running at hardware speeds with the flexibility of software

Current Activities

- Operational applications
  - Approximate keyword search
  - Structured record search
  - Encryption/decryption
  - Signature hashing
- Applications under development
  - Biosequence similarity search
  - Science data mining
  - High-speed compression/decompression
- Investigating SATA RAIDs
  - Order of magnitude price drop vs. SCSI
  - RAID5 support within/across controllers