Using MEMS-based Storage to Boosting Disk Performance

Feng Wang, Bo Hong
Scott A. Brandt, Darrell D. E. Long

Storage Systems Research Center
University of California, Santa Cruz

UC Santa Cruz

MEMS Storage Technology

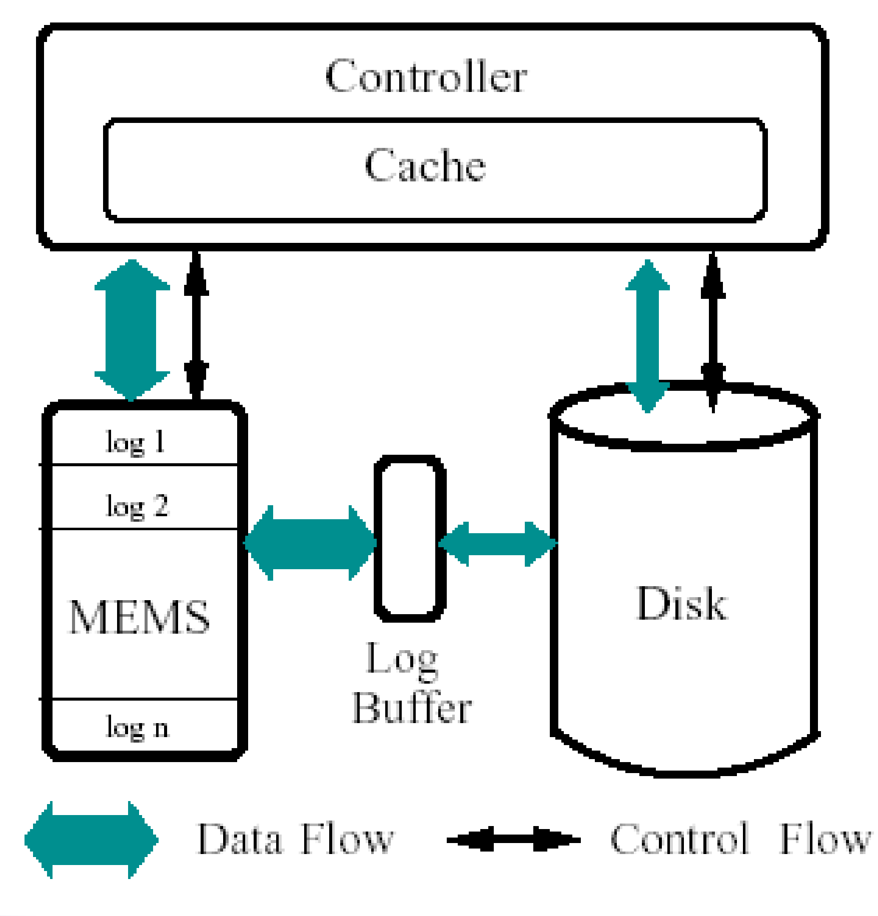
- **M**icro-**E**lectro-**M**echanical **S**ystems (MEMS) storage
  - A promising alternative secondary storage technology
  - Hardware Research: IBM, HP, **CMU**, Nanochip
- Radical differences between MEMS storage and magnetic disk technologies

	Disk	MEMS
Recoding media	Magnetic	**Magnetic** or physical (non-volatile)
Recoding technique	Longitudinal	Orthogonal (higher density)
R/W head	Single	Thousands – tip array (Higher bandwidth and parallelism)
Media movement	Rotation	Media sled moves in X and Y independently (no rotation delay)

Alternative Storage Hierarchies

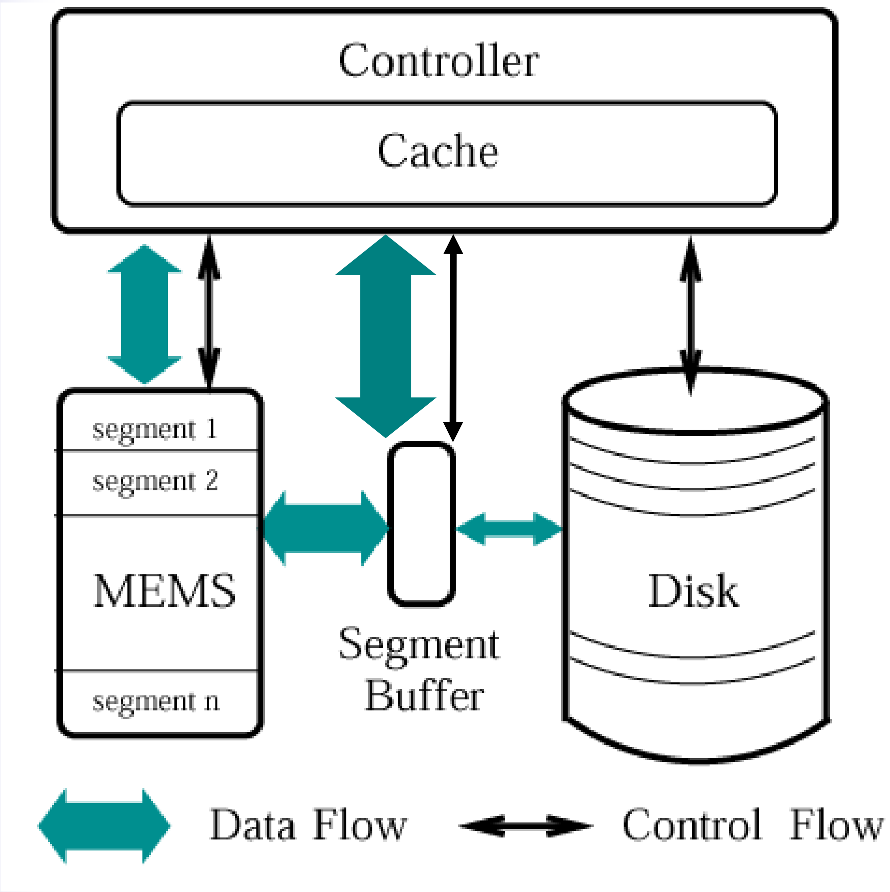
- Replacing disk with MEMS
  - Can be expensive
  - Limited by Capacity
- Hybrid MEMS/disk storage devices
  - MEMS is easy to integrate with disk
    - Non-volatile, block access
  - MEMS can mask relatively large disk access latencies
  - Can be as fast as MEMS and as large and cheap as disk
- Workload characteristics
  - Often write-dominant – use MEMS as a disk write buffer
  - Data reference localities – use MEMS as a disk cache

MEMS as Disk Write Buffer



- All writes appended to logs on MEMS
- Logs written to disk when disk is idle
- Mapping info duplicated in log headers

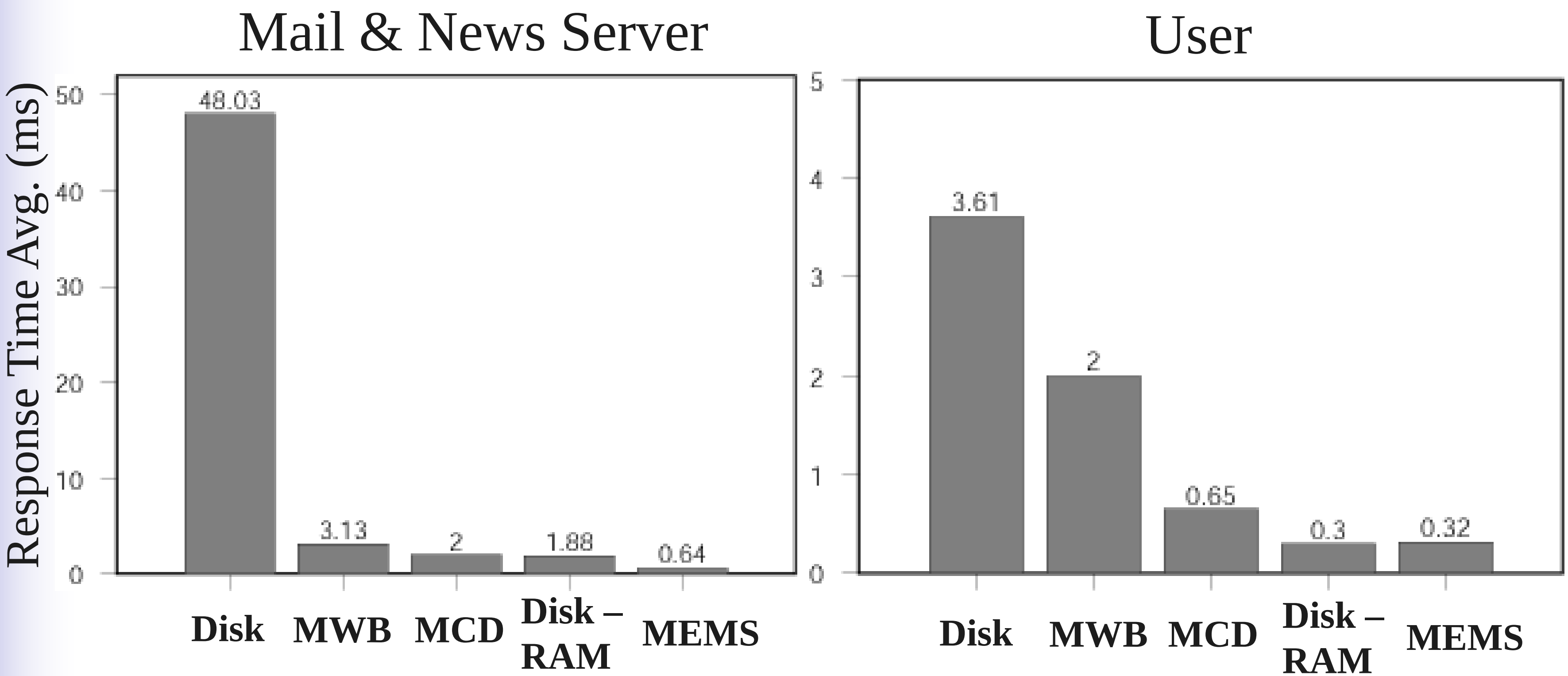
MEMS as Disk Cache



- MEMS as front-end data store of disk
- All requests serviced by MEMS
- Data exchanged between MEMS and disk in segments
- Mapping info duplicated in segment headers

- Optimizations

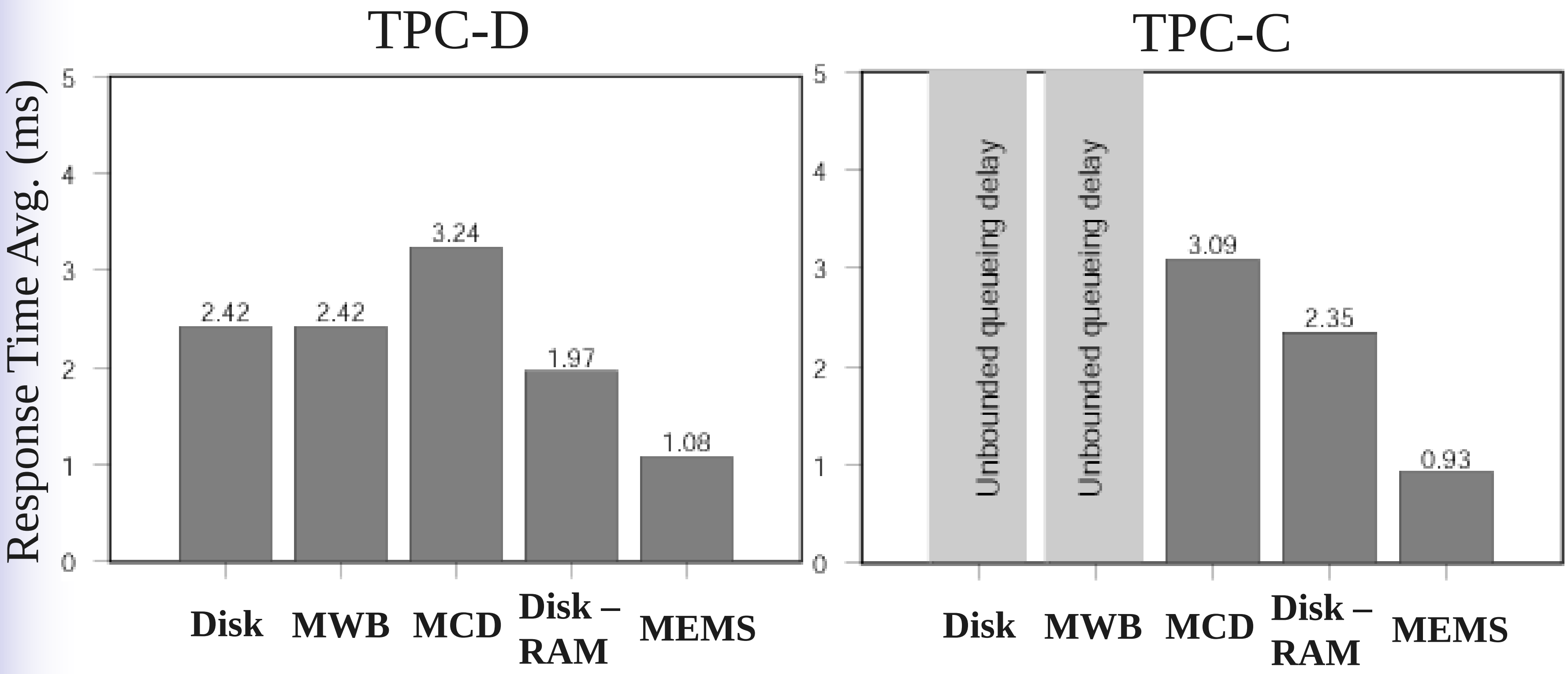
Experimental Results



Disk – disk only
MWB – MEMS Write Buffer
Disk-RAM – Replace MEMS in MCD with the same amount of RAM

MEMS – MEMS only
MCD – MEMS-Cached Disk

Experimental Results (cont.)



Conclusions

- The performance improvement by MEMS write buffer depends on workloads
- With only a small amount of MEMS, MEMS-cached disk can …
  - Provide significant performance improvement in the user, news server, and TPC-C workloads
  - Achieve 30-49% of the MEMS performance

Thank You!

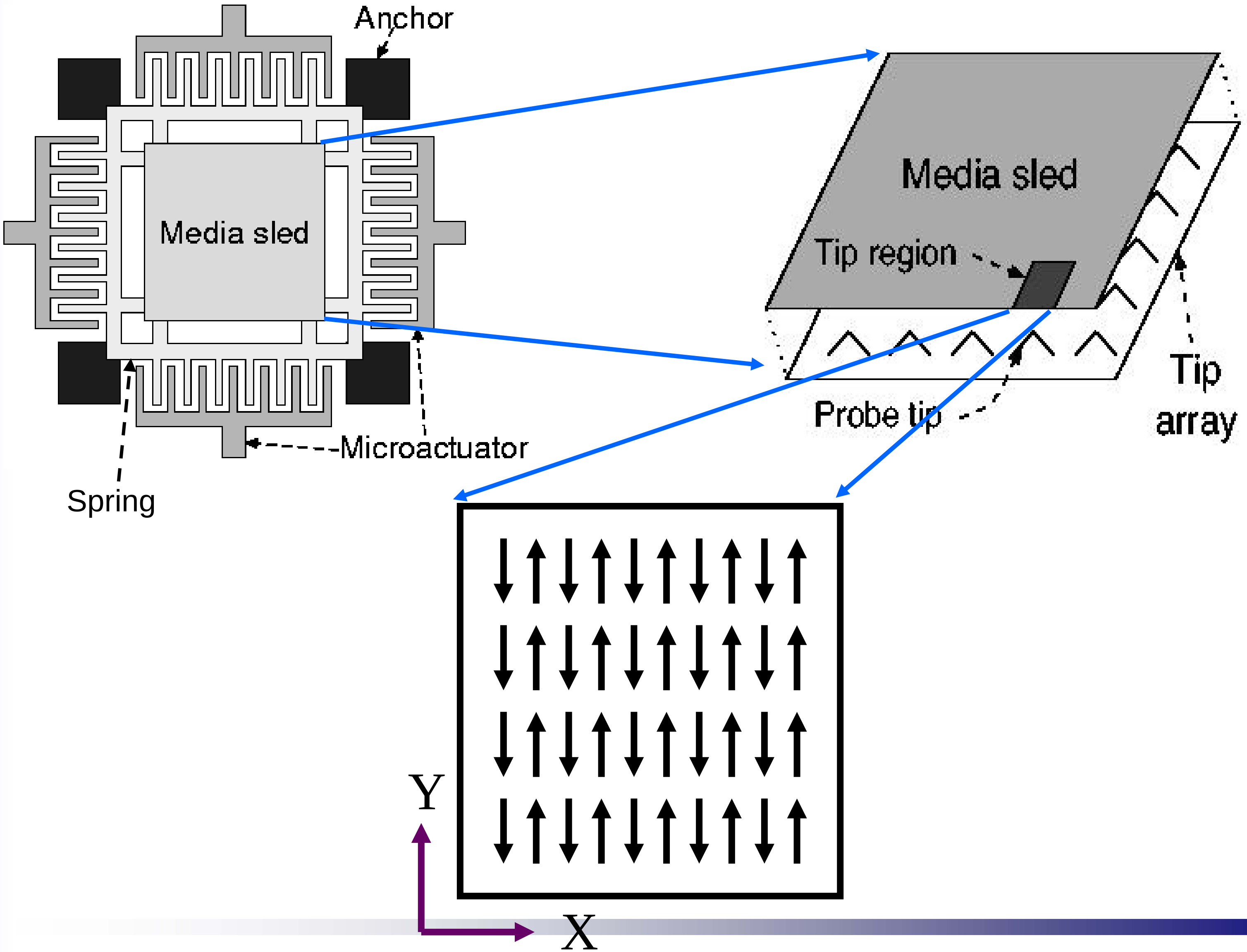
Experimental Methodology

- Disk model – Quantum Atlas 10K, 1999
  - 8.6 GB, 10,025 RPM
  - 5.7/6.19 ms avg. read/write seek times
- MEMS model – CMU G2
  - 0.55 ms avg. seek time
  - 89.6 MB/s streaming bandwidth
- MEMS size
  - 256 MB – 3% of the disk capacity
- Controller cache – 4 MB
- Speed matching buffer – 2 MB
- Segment replacement policy – LRU

A Brief Overview

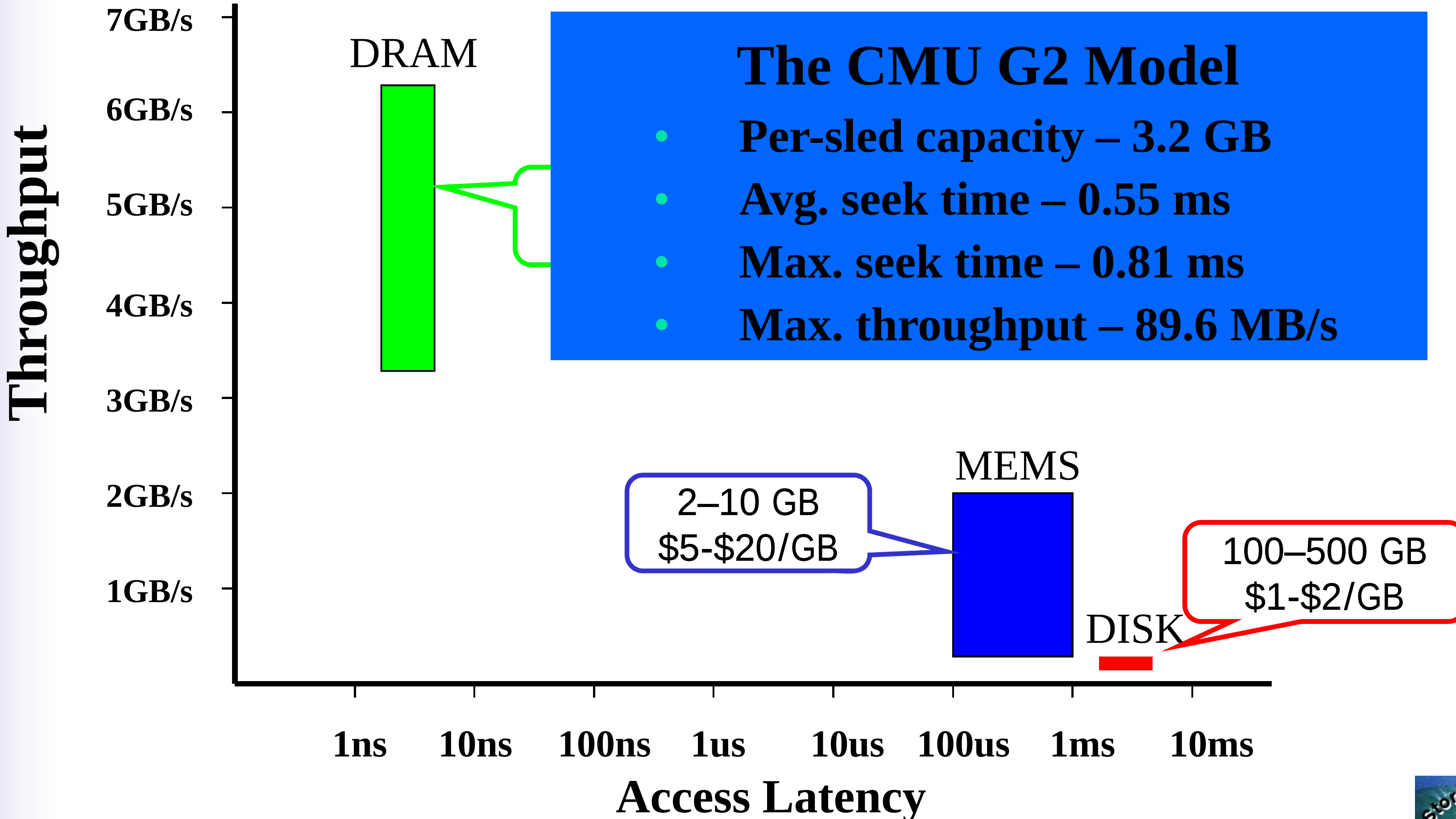
- File and storage systems for MEMS-based storage
  - MEMS-based storage
  - Device modeling and performance characterization
  - Request scheduling
  - Alternative storage hierarchies
  - MEMS storage enclosures – reliable storage bricks
- Other research activities
  - I/O workload modeling and synthesis
    - Irrelevance of long-range dependence in I/O traffic
    - Multifractal-based I/O trace synthesis
    - Cluster-based I/O trace synthesis
  - Modeling and managing flash crowds on the Internet
  - Duplicate data elimination in a SAN file system
- Future research directions

MEMS Storage Device



MEMS Storage Device Characteristics

- Physical size: 1 – 2 cm^2
- Recording density: 250 – 750 Gb/in^2

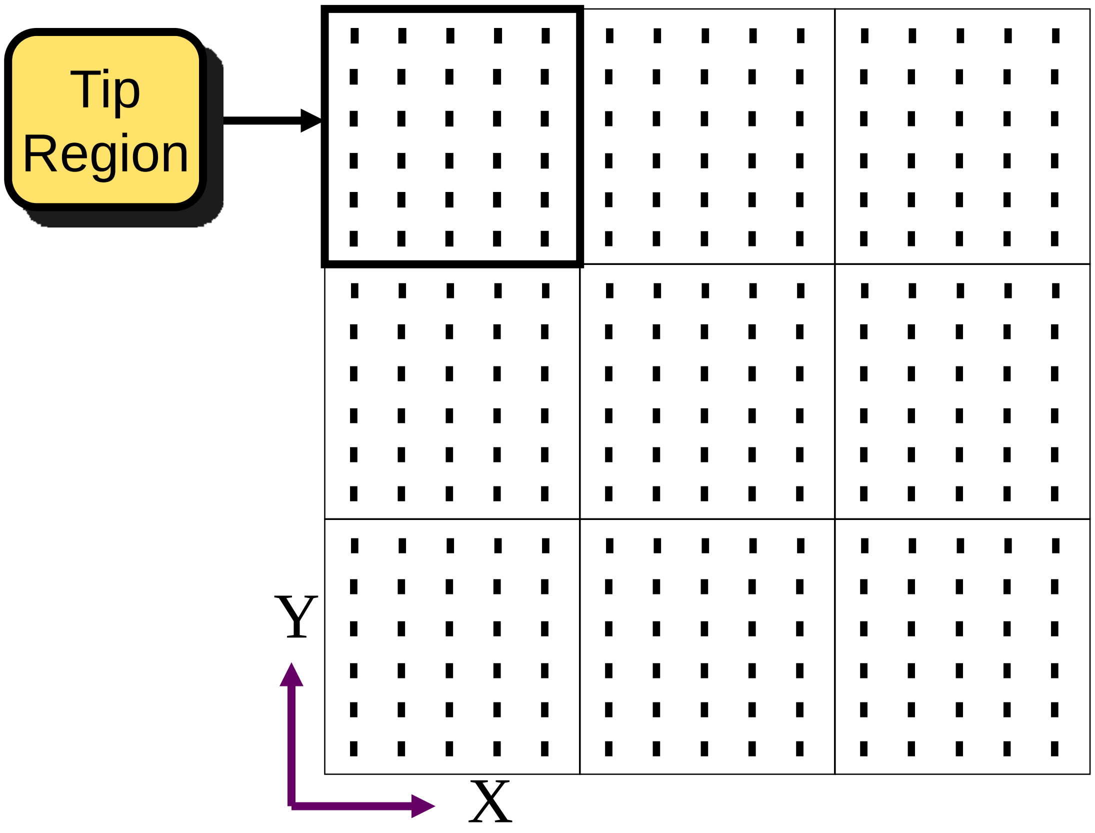


How We Use MEMS-Based Devices

Improve system performance, cost/performance, and reliability

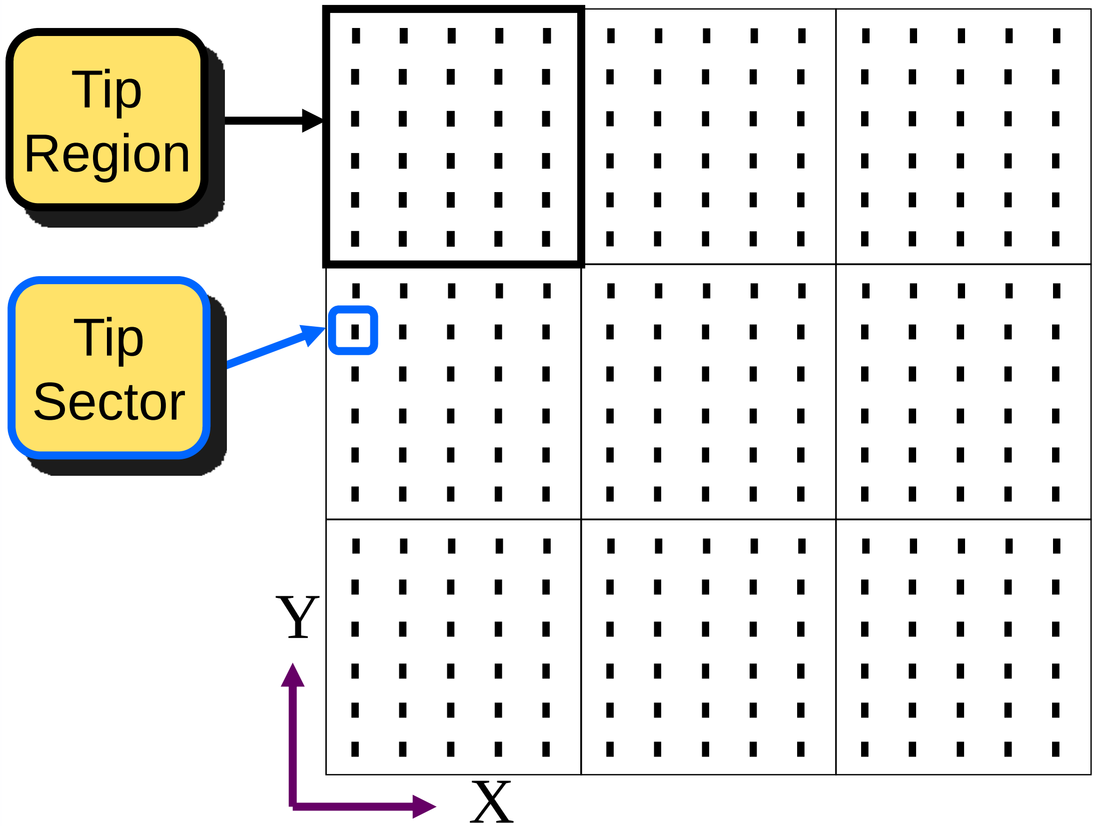
- Leverage low-level device-specific properties
  - Two-dimensional motions
    - Modeling and performance characterization
    - Request scheduling
- Leverage generally superior high-level properties
  - Non-volatility, block-access, easy to work with disks
    - Alternative storage architectures
  - Fast full device scan, little in size / power / entry cost
    - Reliable storage building bricks

Definitions: MEMS Disk Analogies

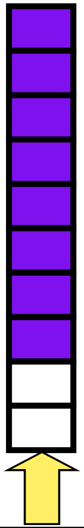


Tip Region:
The portion of the media sled accessible by a single tip

MEMS Disk Analogies

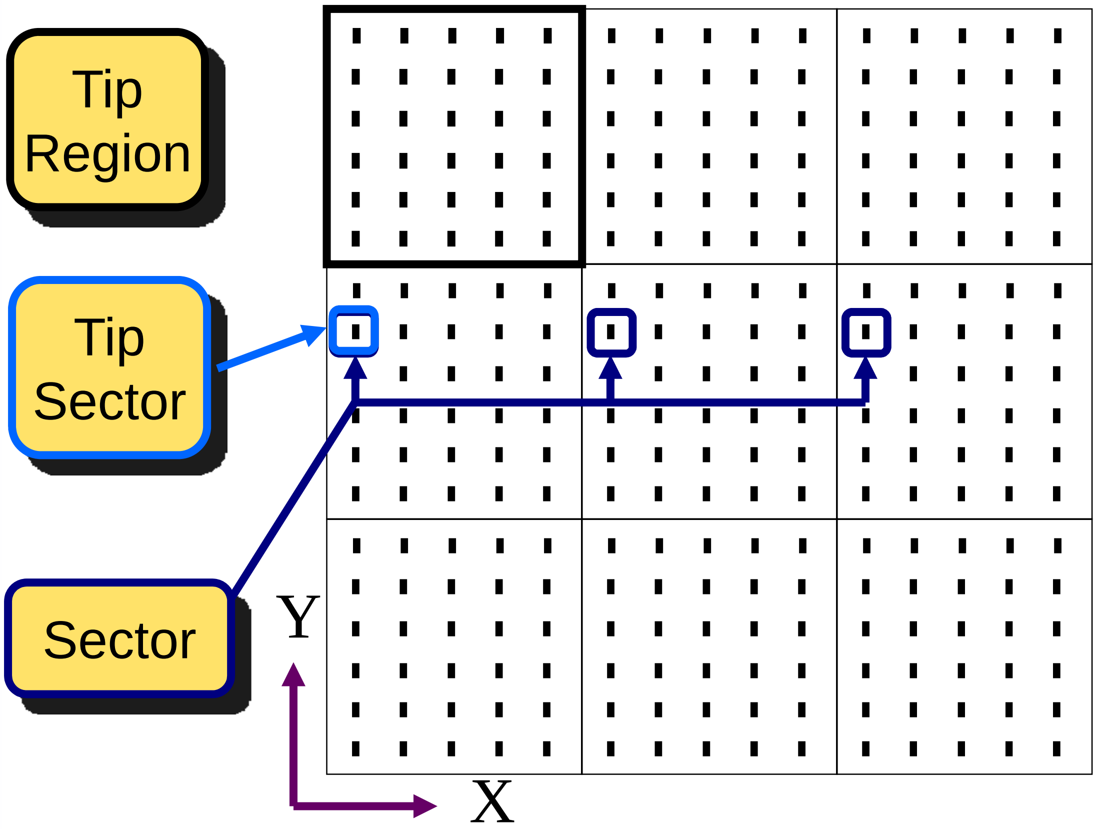


Tip Sector: The smallest unit of data accessible by a single tip



Tip Sector:
8 data bytes +
ECC +
Servo info

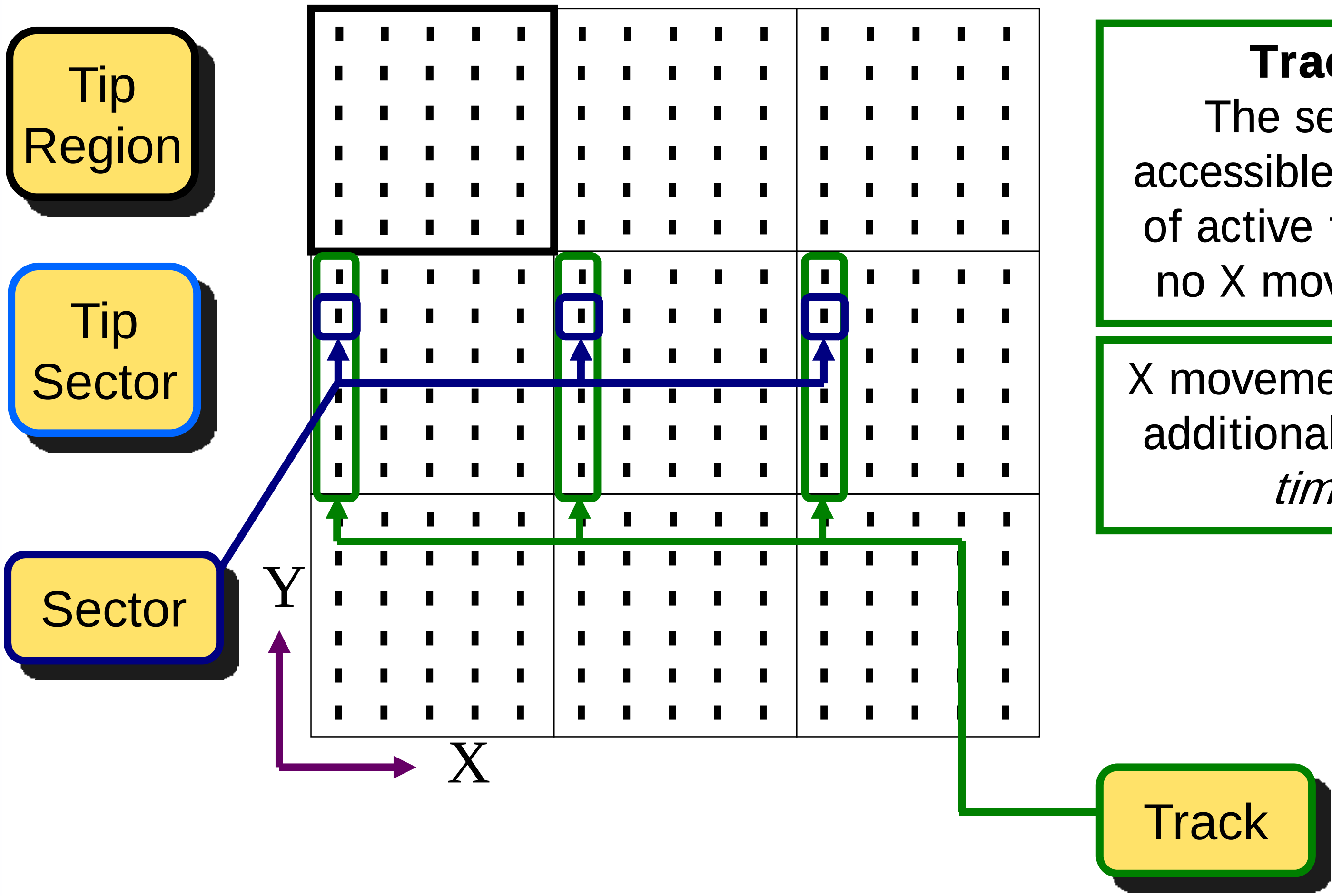
MEMS Disk Analogies



Sector:
The tip sectors accessed by *n* simultaneously active tips. The standard unit of data access.

Recall: all tips are over the same relative tip sector at the same time

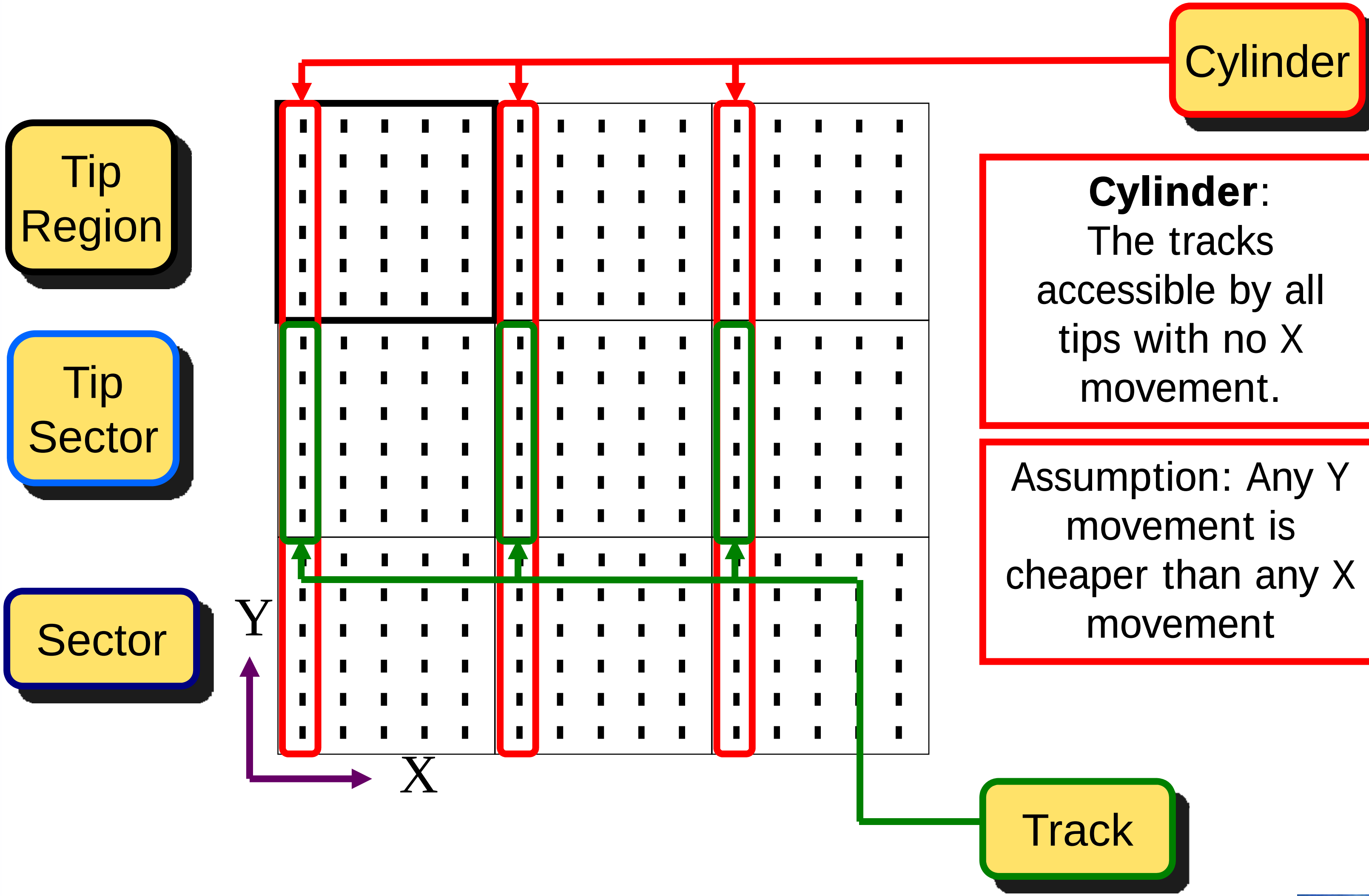
MEMS Disk Analogies



Track:
The sectors accessible by a set of active tips with no X movement.

X movement incurs additional *settling time*.

MEMS Disk Analogies



Cylinder:
The tracks accessible by all tips with no X movement.

Assumption: Any Y movement is cheaper than any X movement