

MADALGO Core Research Areas

Massive Data

Society

- Increasingly “data driven”
- Sensors in buildings, cars, phones, goods, animals, cloth, name tags,...
- More networked devices that both acquire and process data
- Acquire, store and process massive data anywhere and any time

Massive data examples

- Phone:** AT&T +20TB phone call database, wireless tracking
- Consumer:** WalMart +70TB database, buying patterns
- Bank:** Danske Bank +250TB database
- WEB:** Google index +8 billion web pages
- Geography:** NASA satellites generate Terrabytes each day

Nature 2006

Highlights trends in sciences: “2020 – Future of computing”

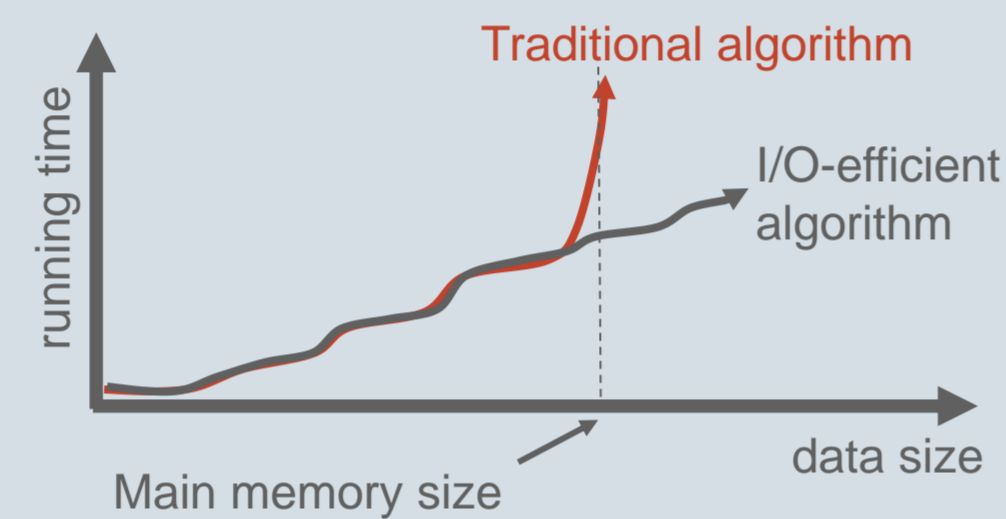


- Exponential growth of scientific data
- Due to large experiments, sensor networks, etc
- Paradigm shift: *Science is increasingly about mining massive data*

- Computer science paramount in all sciences
- Increased data availability: “nano-technology-like” opportunity

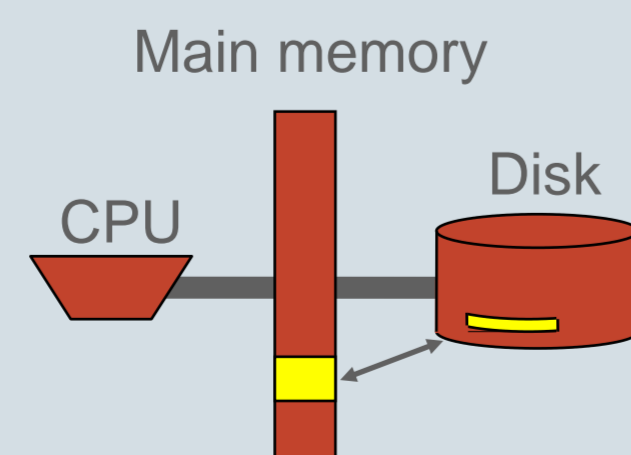
I/O-Efficient Algorithms

- Problems involving massive data on disk
- Disk access is 10^6 times slower than main memory access
- Large access time amortized by transferring large blocks of data
- Inadequacy of traditional algorithmics



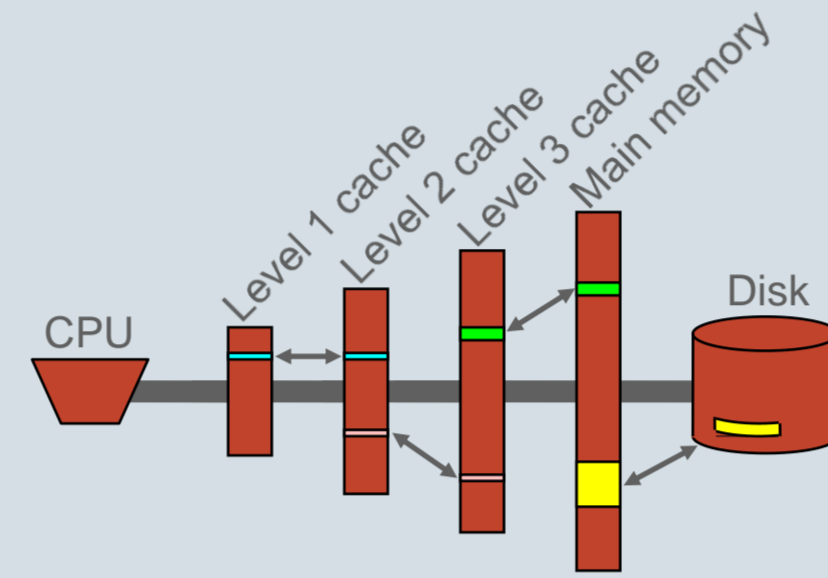
I/O-efficient algorithms

- Important to store/access data to take advantage of blocks
- Move as few disk blocks as possible to solve given problem

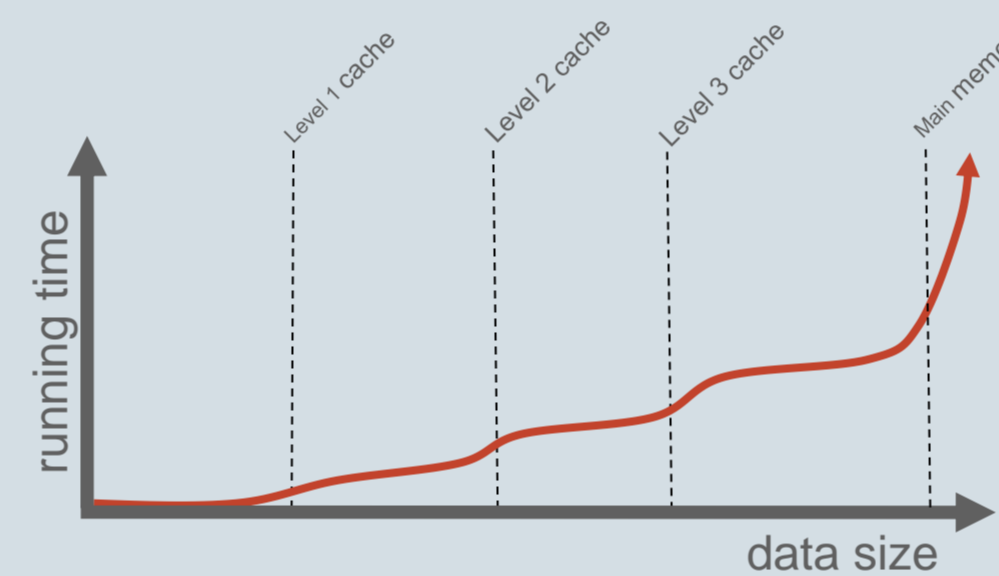


Cache-Oblivious Algorithms

- Block access important on all levels of memory hierarchy



- Memory hierarchies are very diverse

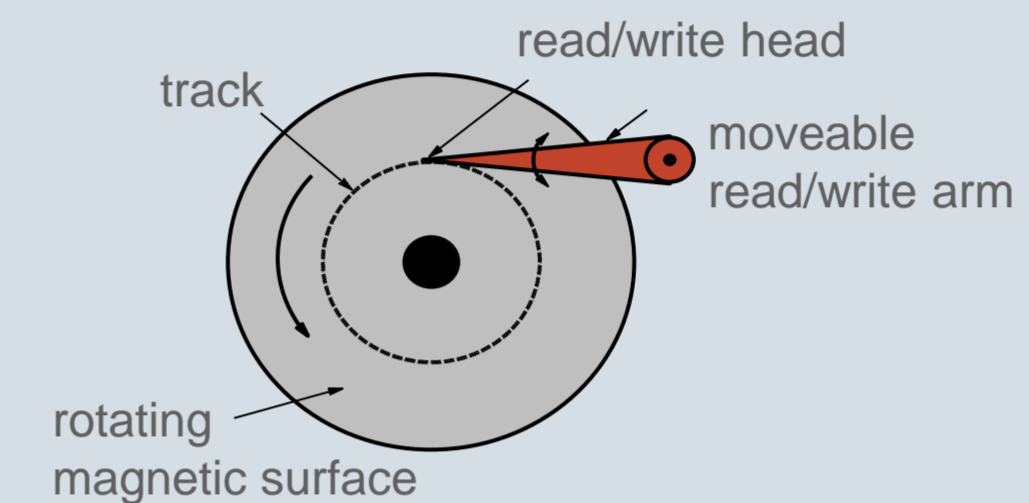


Cache-oblivious algorithms

- Use blocks efficiently on *all* levels of *any* hierarchy
- Algorithms do not know the parameters of the hierarchy
- Block transfers between memory levels is done automatically

Streaming Algorithms

- Problems involving truly massive data
- Sequential read of disk blocks much faster than random read

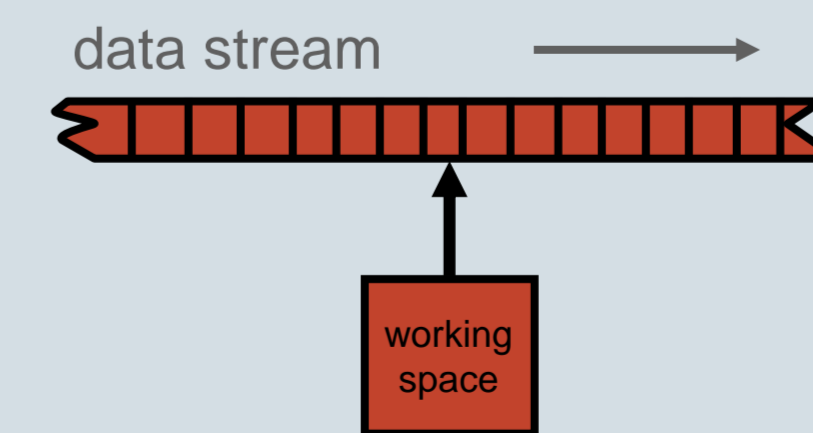


- In many modern (sensor) applications data arrive continually
- (Massive) problems often have to be solved in one sequential scan



Streaming algorithms

- Use single scan, handle each element fast, using small space



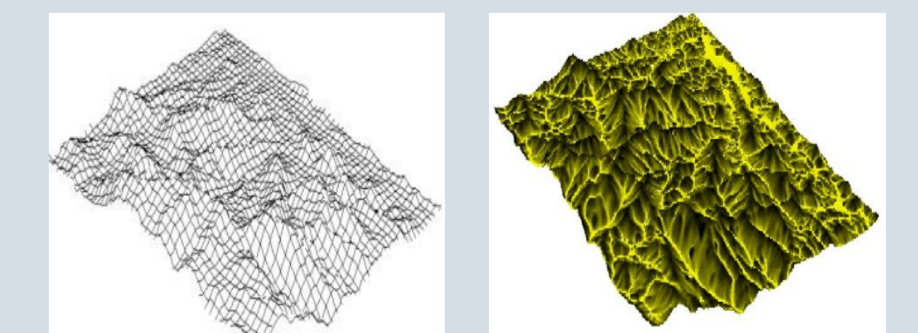
Algorithm Engineering

- Design and implementation of practical algorithms
- Experimentation
- Algorithm engineering work can lead to practical breakthroughs



Example

- Flow simulation on massive terrain models
- 18 billion points at 1 meter (>>1TB)



Implementation of I/O-efficient
→ two weeks to three hours !!!