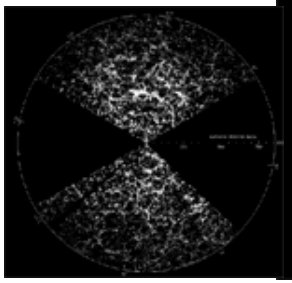
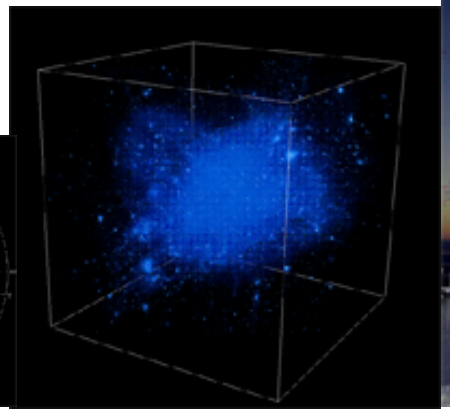
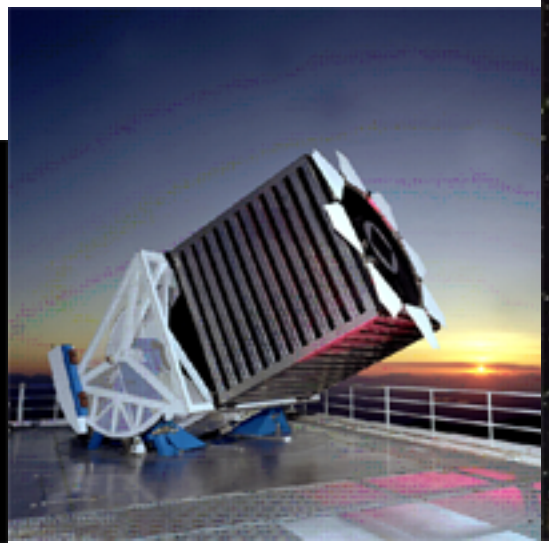
Institute for Data Intensive Engineering and Science
idies

Extreme Data-Intensive Scientific Computing: The Fourth Paradigm

Alex Szalay
JHU


Big Data in Science

- Data growing exponentially, in all science
- All science is becoming data-driven
- This is happening very rapidly
- Data becoming increasingly open/public
- Non-incremental!
- Convergence of physical and life sciences through Big Data (statistics and computing)
- The “long tail” is important
- A scientific revolution in how discovery takes place

  => a rare and unique opportunity

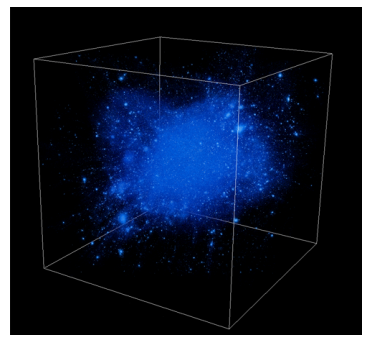
Science is Changing

THOUSAND YEARS AGO
science was empirical
describing natural phenomena

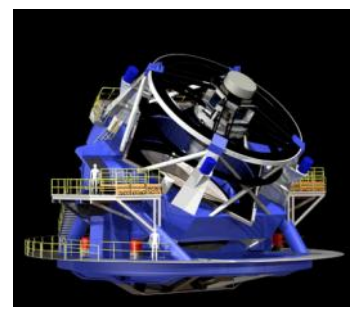
LAST FEW HUNDRED YEARS
theoretical branch using models, generalizations

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - \mathrm{K}\frac{c^2}{a^2}$$

LAST FEW DECADES
a computational branch simulating complex phenomena



TODAY
data intensive science, synthesizing theory, experiment and computation with statistics
►new way of thinking required!



Scientific Data Analysis Today

- Scientific data is doubling every year, reaching PBs
  - *CERN is at 22PB today, 10K genomes ~5PB*
- Data will never will be at a single location
- Architectures increasingly CPU-heavy, IO-poor
- Scientists need special features (arrays, GPUs)
- Most data analysis done on midsize BeoWulf clusters
- Universities hitting the “power wall”
- Soon we cannot even store the incoming data stream
- **Not scalable, not maintainable…**

Non-Incremental Changes

- Multi-faceted challenges
- New computational tools and strategies

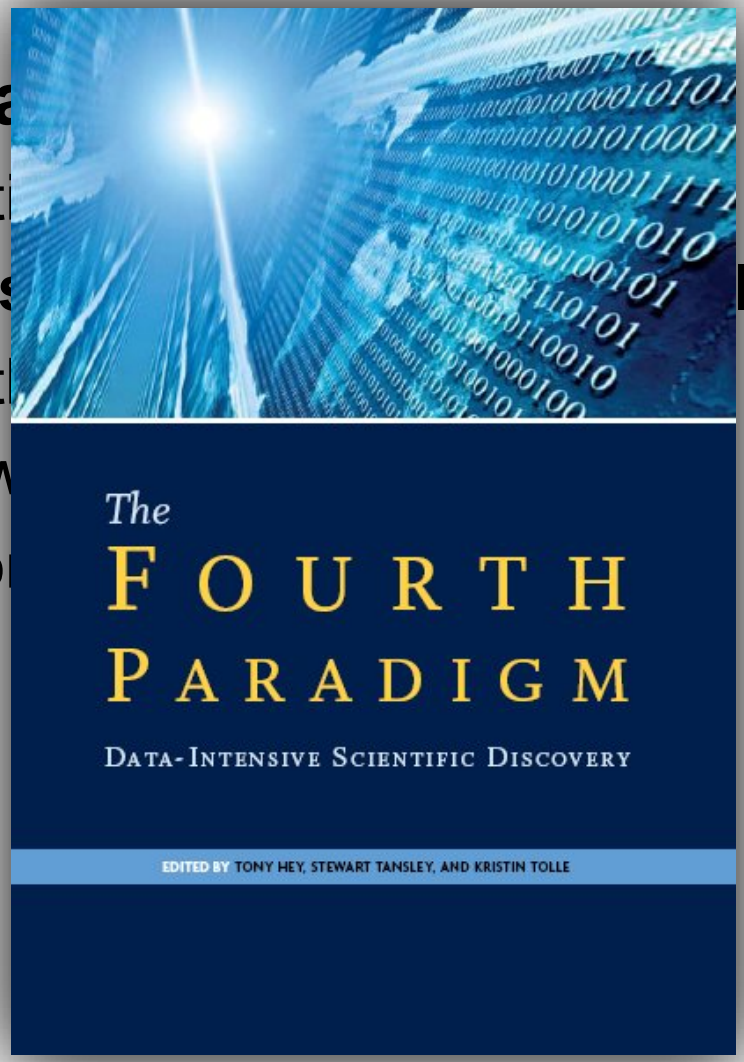
  … not just statistics, not just computer science, not just astronomy, not just genomics…
- Need new data intensive scalable architectures
- Science is moving increasingly from hypothesis-driven to data-driven discoveries

Astronomy has always been data-driven…. now this is becoming more accepted in other areas as well

Gray's Laws of Data Engineering

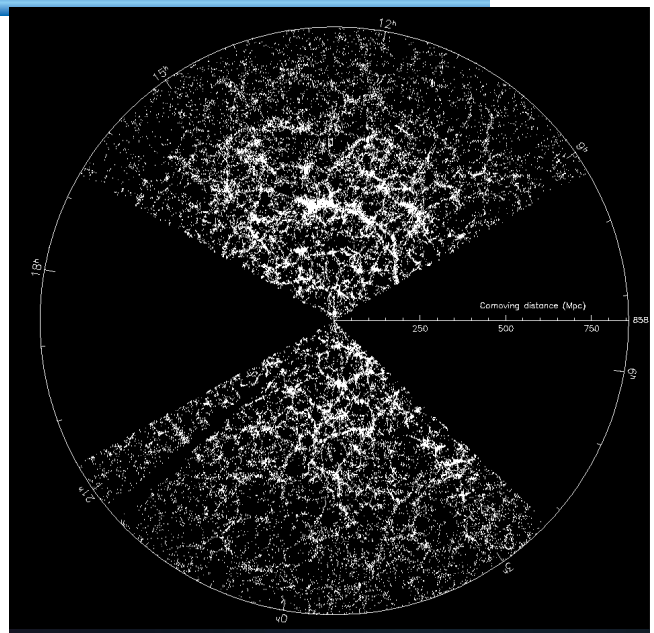
Jim Gra

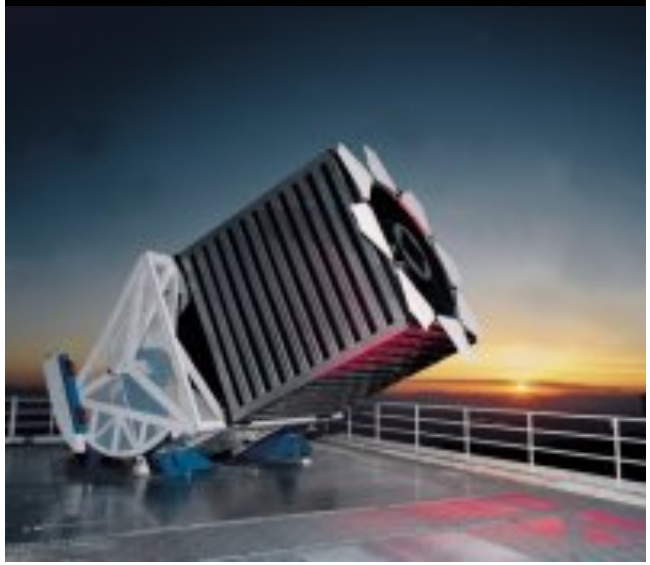
- Scienti around **data**
- Need s lysis
- Take t
- Start w
- Go fro



Sloan Digital Sky Survey

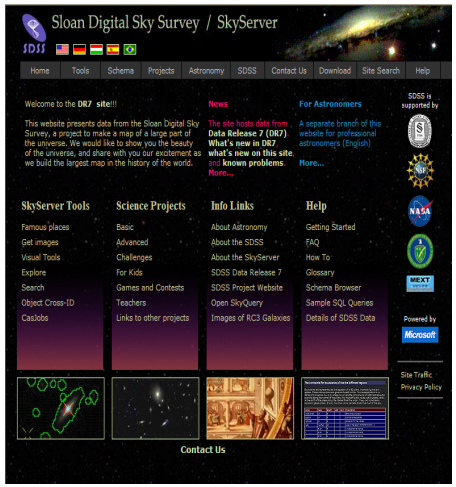
- **"The Cosmic Genome Project"**
- Two surveys in one
  - Photometric survey in 5 bands
  - Spectroscopic redshift survey
- Data is public
  - 2.5 Terapixels of images => 5 Tpx
  - 10 TB of raw data => 120TB processed
  - 0.5 TB catalogs => 35TB in the end
- Started in 1992, finished in 2008
- Database and spectrograph built at JHU (SkyServer)





Skyserver

- Prototype in 21st Century data access
  - *993 million web hits in 10 years*
  - *4,000,000 distinct users vs. 15,000 astronomers*
  - *The emergence of the "Internet scientist"*
  - *The world's most used astronomy facility today*
  - *Collaborative server-side analysis done by 5K astronomers (30%)*



- GalaxyZoo (Lintott et al)
  - *40 million visual galaxy classifications by the public*
  - *Enormous publicity (CNN, Times, Washington Post, BBC)*
  - *300,000 people participating, blogs, poems…*
  - *Original discoveries by the public (Voorwerp, Green Peas)*

Impact of Sky Surveys

Astronomy

Sloan Digital Sky Survey tops astronomy citation list

NASA's Sloan Digital Sky Survey (SDSS) is the most significant astronomical facility, according to an analysis of the 200 most cited papers in astronomy published in 2006. The survey, carried out by Juan Madrid from McMaster University in Canada and Duccio Macchetto from the Space Telescope Science Institute in Baltimore, puts NASA's Swift satellite in second place, with the Hubble Space Telescope in third (arXiv:0901.4552).

Madrid and Macchetto carried out their analysis by looking at the top 200 papers using NASA's Astrophysics Data System (ADS), which charts how many times each paper has been cited by other research papers. If a paper contains data taken only from one observatory or satellite, then that facility is awarded all the citations given to that article. However, if a paper is judged to contain data from different facilities – say half from SDSS and half from Swift – then both facilities are given 50% of the citations that paper received.

Top 10 telescopes

Rank	Telescope	Citations	Ranking in 2004
1	Sloan Digital Sky Survey	1892	1
2	Swift	1523	N/A
3	Hubble Space Telescope	1078	3
4	European Southern Observatory	813	2
5	Keck	572	5
6	Canada–France–Hawaii Telescope	521	N/A
7	Spitzer	469	N/A
8	Chandra	381	7
9	Boomerang	376	N/A
10	High Energy Stereoscopic System	297	N/A

The researchers then totted up all the citations and produced a top 10 ranking (see table). Way out in front with 1892 citations is the SDSS, which has been running since 2000 and uses the 2.5 m telescope at Apache Point in New Mexico to obtain images of more than a quarter of the sky. NASA's Swift satellite, which studies gamma-ray bursts, is second with 1523 citations, while the Hubble Space Telescope (1078 citations) is third.

Although the 200 most cited papers make up only 0.2% of the references indexed by the ADS for papers published in 2006, those 200 papers account for 9.5% of the citations. Madrid and Macchetto also ignored theory papers on the basis that they do not directly use any telescope data. A similar study of papers published in 2004 also puts SDSS top with 1843 citations. This time, though, the European Southern Observatory, which has telescopes in Chile, comes second with 1365 citations and the Hubble Space Telescope takes third spot with 1124 citations.

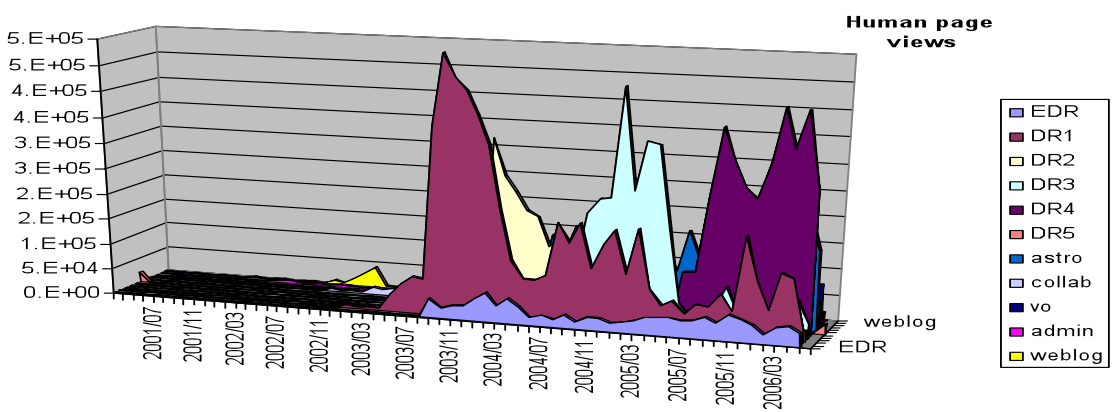
Michael Banks

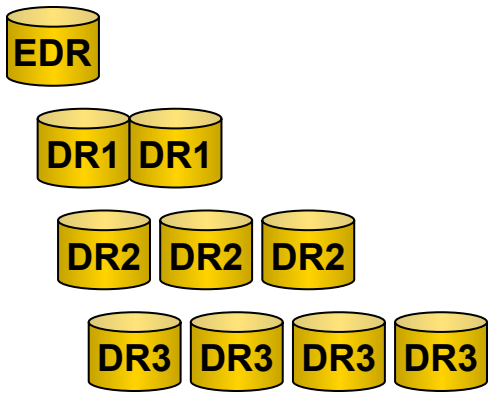
Database Challenges

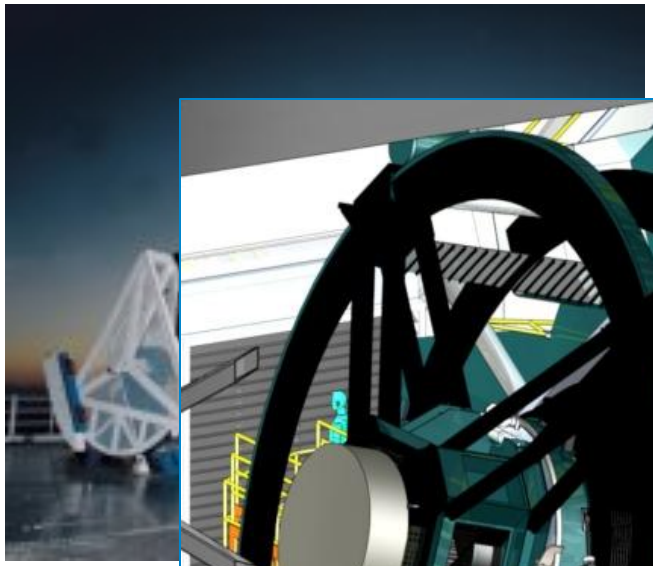
- Loading (and scrubbing) the data
- Organizing the data (20 queries, self-documenting)
- Accessing the data (small and large queries, visual)
- Delivering the data (workbench)
- Analyzing the data (spatial, scaling…)

Data Versions

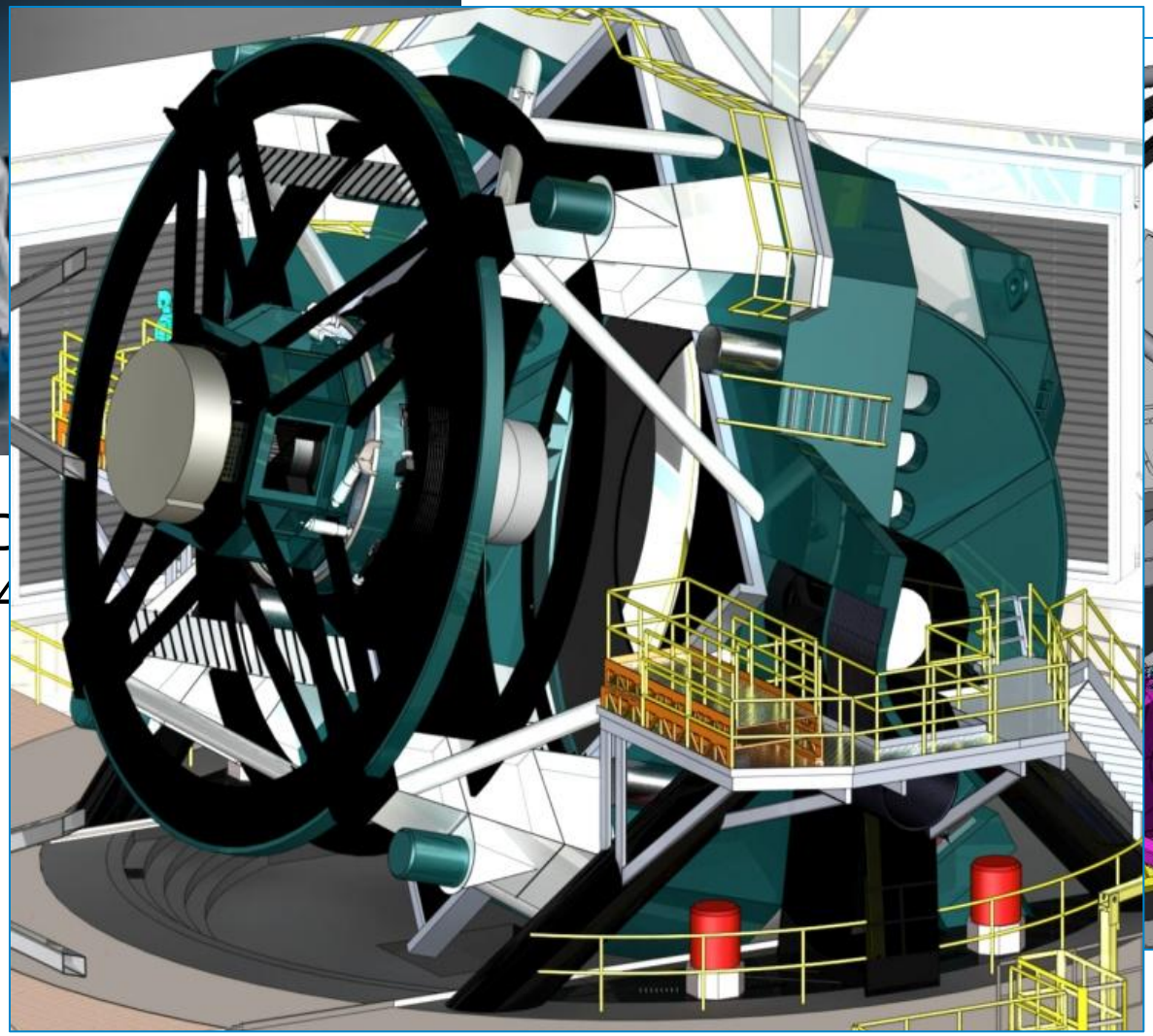
- June 2001: EDR
- Now at DR5, with 2.4TB
- 3 versions of the data
  - *Target, Best, Runs*
  - *Total catalog volume 5TB*
- Data publishing: once published, must stay
- SDSS: DR1 is still used

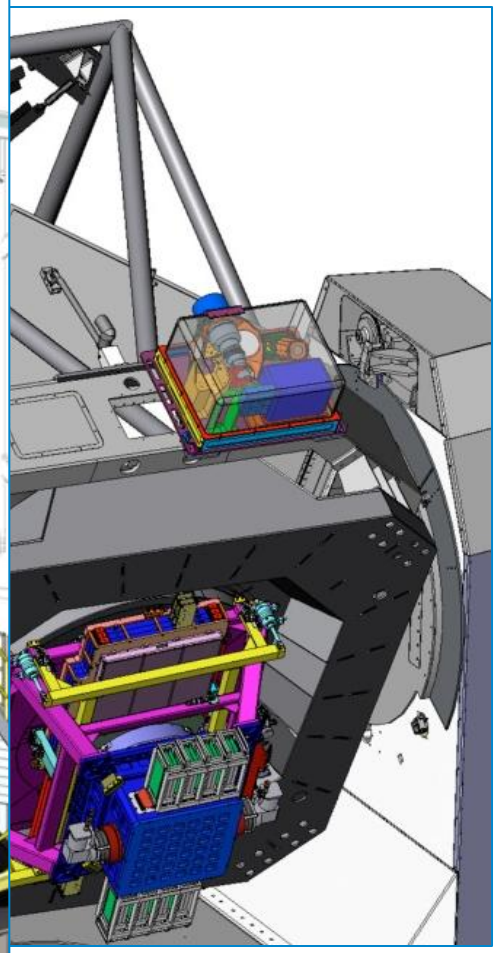






SD
2.4

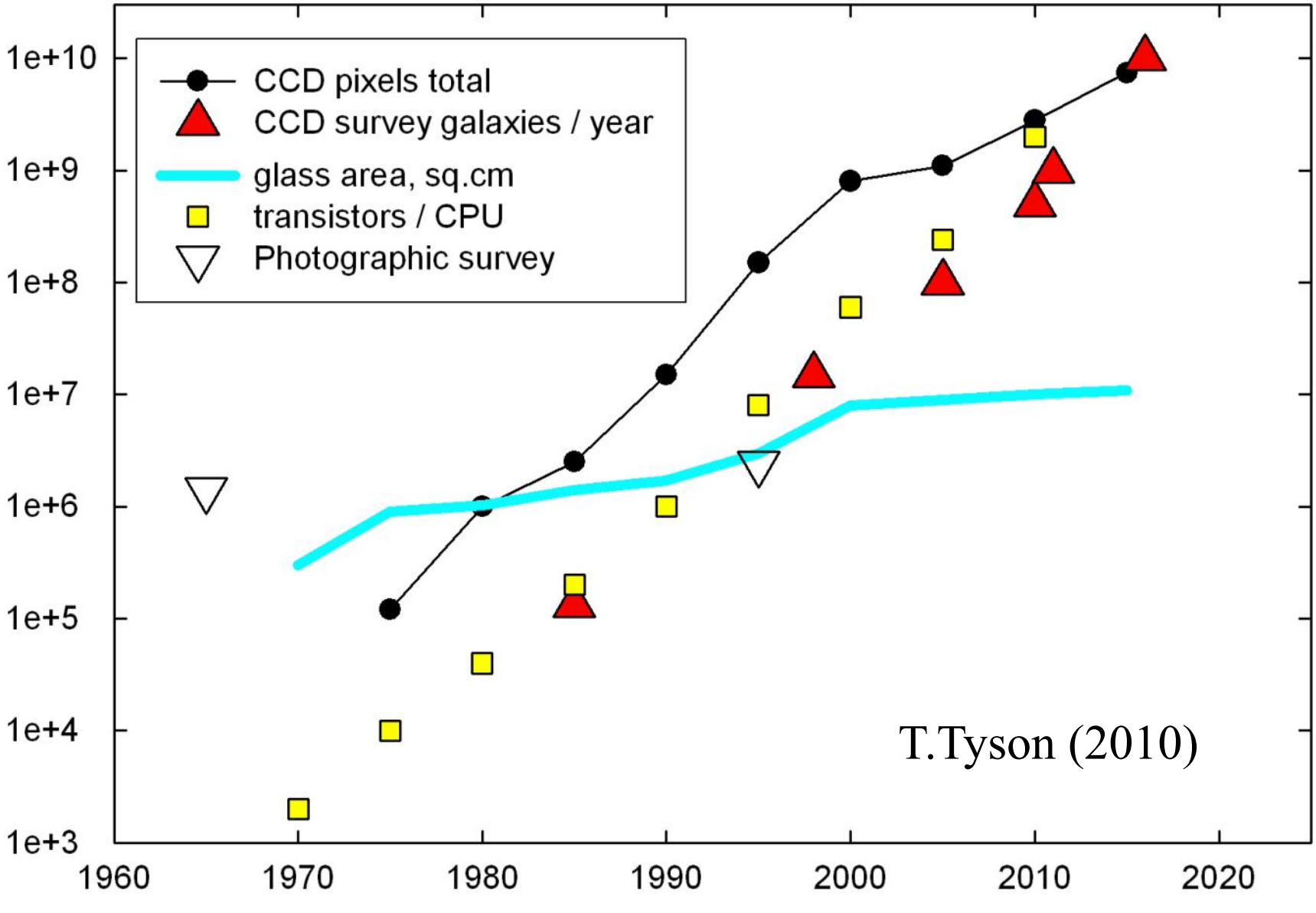




LSST
8.4m 3.2Gpixel

PanSTARRS
1.8m 1.4Gpixel

Survey Trends



Why Is Astronomy Interesting?

- Approach inherently and traditionally data-driven
  - *Cannot do experiments…*
- Important spatio-temporal features
- Very large density contrasts in populations
- Real errors and covariances
- Many signals very subtle, buried in systematics
- Data sets large, pushing scalability
  - *LSST will be 100PB*

*"Exciting, since it is **worthless**!"*

— Jim Gray

Virtual Observatory Challenges

- Most challenges are sociological, not technical
- Trust: scientists want trustworthy, calibrated data with occasional access to low-level raw data
- Career rewards for young people still not there
- Threshold for publishing data is still too high
- Robust applications are hard to build (factor of 3…)
- Archives (and data) on all scales, all over the world

- Astronomy has successfully passed the first hurdles… but it is a long journey… no instant gratification

Data in HPC Simulations

- HPC is an instrument in its own right
- Largest simulations approach petabytes
  - *from supernovae to turbulence, biology and brain modeling*
- Need public access to the best and latest through interactive numerical laboratories
- Creates new challenges in
  - *how to move the petabytes of data (high speed networking)*
  - *How to look at it (render on top of the data, drive remotely)*
  - *How to interface (virtual sensors, immersive analysis)*
  - *How to analyze (algorithms, scalable analytics)*

Data Access is Hitting a Wall

FTP and GREP are not adequate

- You can GREP 1 MB in a second
- You can GREP 1 GB in a minute
- You can GREP 1 TB in 2 days
- You can GREP 1 PB in 3 years

- You can FTP 1 MB in 1 sec
- You can FTP 1 GB / min (= 1 $/GB)
- … 2 days and 1K$
- … 3 years and 1M$

- Oh!, and 1PB ~4,000 disks

- At some point you need
  indices to limit search
  parallel data search and analysis
- This is where **databases** can help

Slide from Jim Gray (2005)

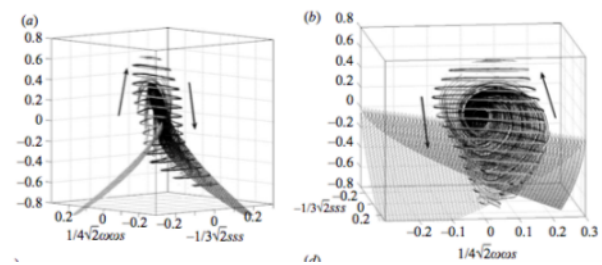
Silver River Transfer

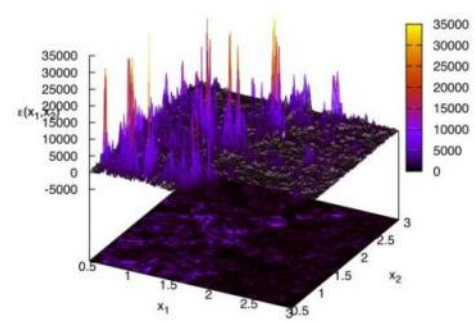
- 150TB in less than 10 days from Oak Ridge to JHU using a dedicated 10G connection

Immersive Turbulence

"… the last unsolved problem of classical physics…" Feynman

- **Understand the nature of turbulence**
  - *Consecutive snapshots of a large simulation of turbulence: now 30 Terabytes*
  - *Treat it as an experiment, **play** with the database!*
  - ***Shoot test particles** (sensors) from your laptop into the simulation, like in the movie Twister*
  - *Next: 70TB MHD simulation*

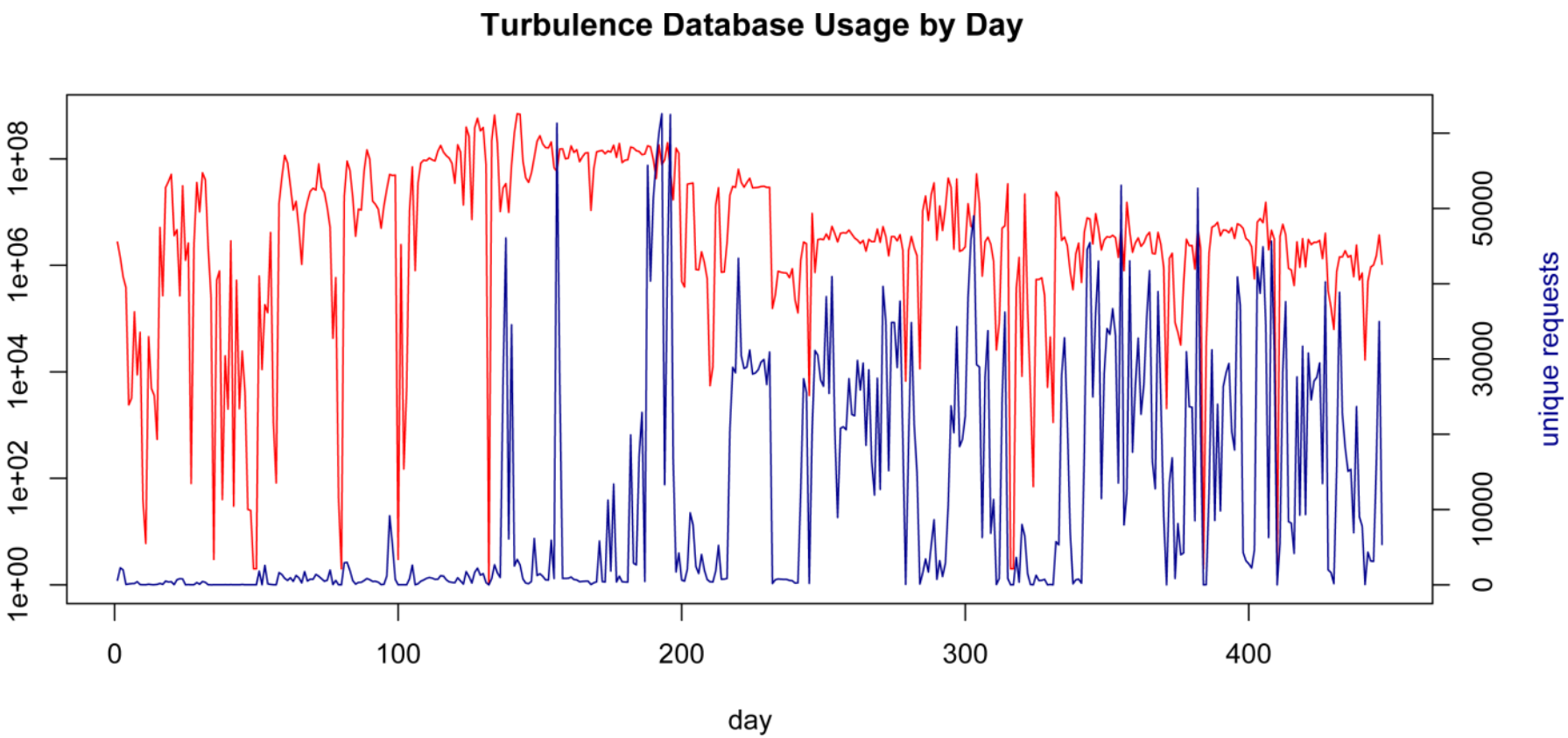




- **New paradigm** for analyzing simulations!

with C. Meneveau, S. Chen (Mech. E), G. Eyink (Applied Math), R. Burns (CS)

Daily Usage

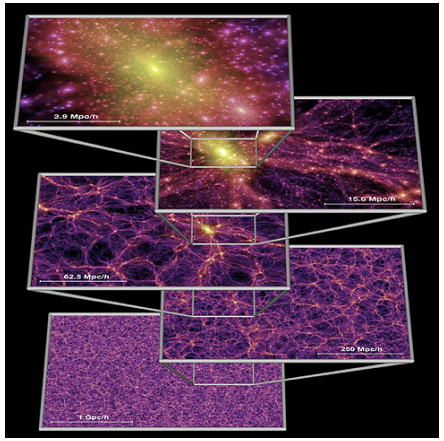


2011: exceeded 100B points, delivered publicly

Cosmological Simulations

In 2000 cosmological simulations had 10^{10} particles and produced over 30TB of data (Millennium)

- Build up dark matter halos
- Track merging history of halos
- Realistic distribution of galaxy types
- Reference for the whole community

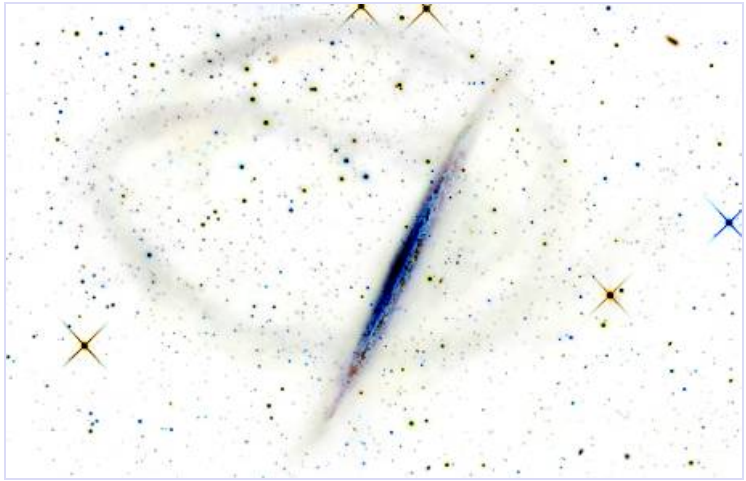


- Today: simulations with 10^{12} particles and PBs of output are under way (MillenniumXXL, Silver River, etc)
- Hard to analyze the data afterwards
- What is the best way to compare to real data?

The Milky Way Laboratory

- Use cosmology simulations as an immersive laboratory for general users
- Via Lactea-II (20TB) as prototype, then Silver River (50B particles) as production (15M CPU hours)
- 800+ hi-rez snapshots (2.6PB) => 800TB in DB
- Users can insert test particles (dwarf galaxies) into system and follow trajectories in pre-computed simulation
- Users interact remotely with a PB in 'real time'

Madau, Rockosi, Szalay, Wyse, Silk, Kuhlen, Lemson, Westermann, Blakeley



Data-Intensive Research at JHU

- Sloan Digital Sky Survey
- Virtual Observatory
- Pan-STARRS
- LSST
- Earth circulation modeling
- Turbulence
- LHC

- Computational Biology
- High Throughput Genomics
- Biophysics
- Neuroscience/ fMRI
- OncoSpace
- BIRN
- Life Under Your Feet

- IDIES
- Data Conservancy
- I4M
- Discovery grants

- GrayWulf
- Amdahl-Blades
- Data-Scope
- CDI/ITR/MRI…

Institute for Data Intensive Engineering and Science

idies

Common Analysis Patterns

- Huge amounts of data, aggregates needed
  - *But also need to keep raw data*
  - *Need for parallelism*
  - *Heavy use of structured data, multi-D arrays*
- Requests enormously benefit from indexing
- Computations must be close to the data!
- Very few predefined query patterns
  - *Everything goes….*
  - *Rapidly extract small subsets of large data sets*
  - *Geospatial/locality based searches everywhere*
- Data will never be in one place
  - *Remote joins will not go away*
- No need for transactions
- Data scrubbing is crucial

DB!

Increased Diversification

One shoe does not fit all!

- Diversity grows naturally, no matter what
- Evolutionary pressures help
- Individual groups want specializations

- Large floating point calculations move to GPUs
- Big data moves into the cloud (private or public)
- RandomIO moves to Solid State Disks
- High-Speed stream processing emerging
- noSQL vs databases vs column store vs SciDB …

At the same time

- What remains in the middle?
  - Common denominator is Big Data
- Data management
  - Everybody needs it, nobody enjoys doing it
- We are still building our own… over and over

Amdahl's Laws

Gene Amdahl (1965): Laws for a balanced system

i. Parallelism: max speedup is S/(S+P)

ii. One bit of IO/sec per instruction/sec (BW)

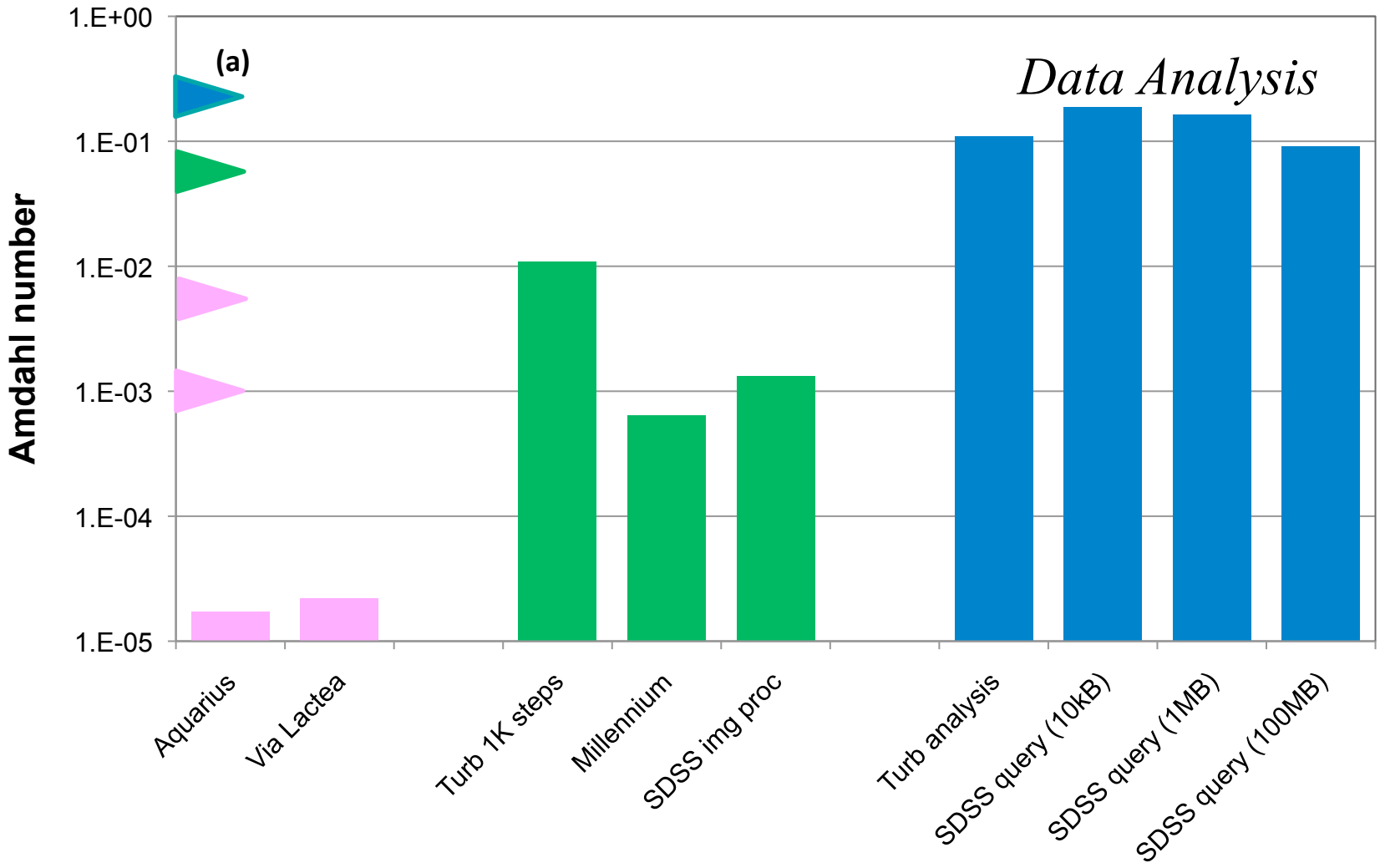
iii. One byte of memory per one instruction/sec (MEM)

Modern multi-core systems move farther away from Amdahl's Laws (Bell, Gray and Szalay 2006)

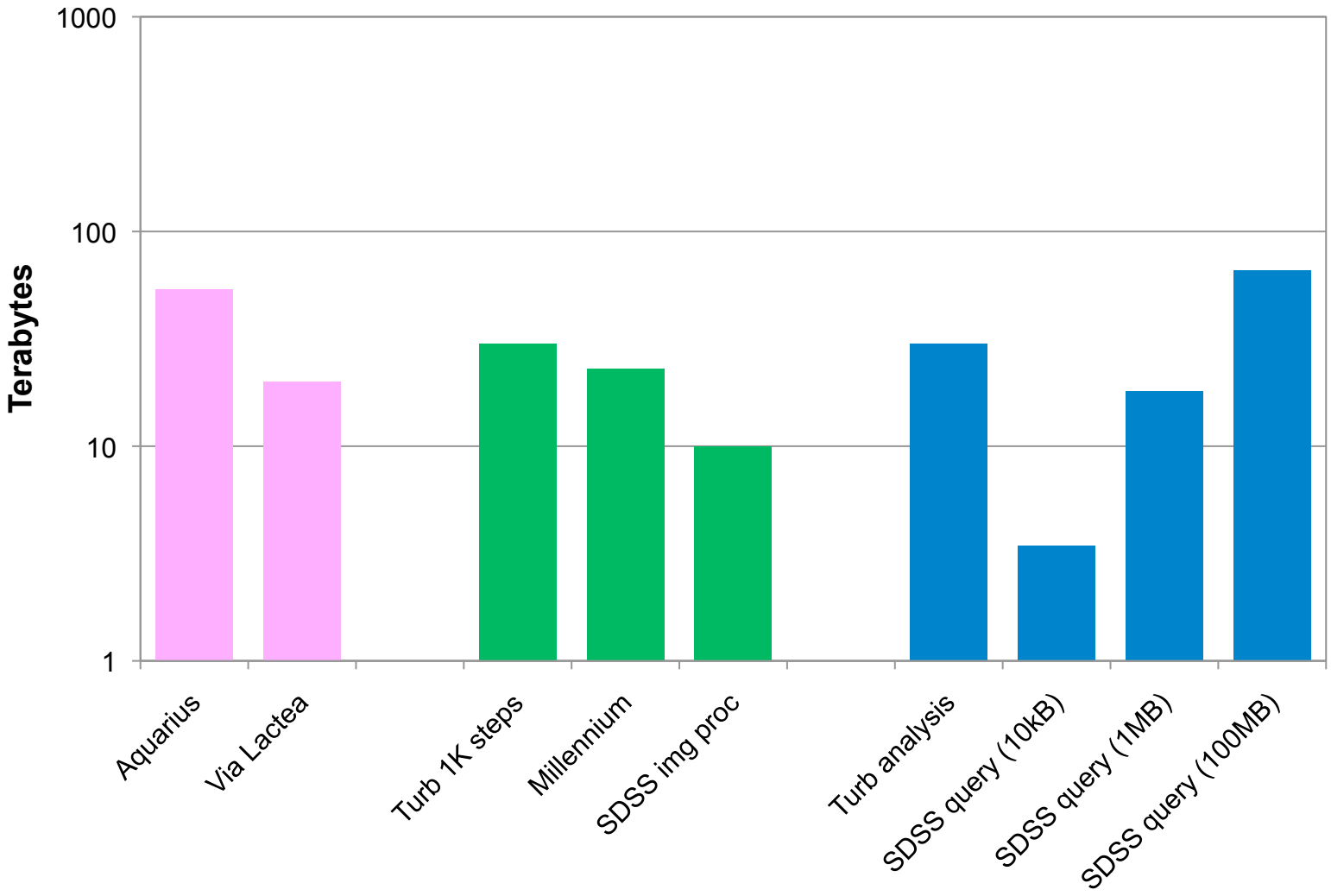
Typical Amdahl Numbers

System	CPU count	GIPS [GHz]	RAM [GB]	diskIO [MB/s]	Amdahl	
					RAM	IO
BeoWulf	100	300	200	3000	0.67	0.08
Desktop	2	6	4	150	0.67	0.2
Cloud VM	1	3	4	30	1.33	0.08
SC1	212992	150000	18600	16900	0.12	0.001
SC2	2090	5000	8260	4700	1.65	0.008
GrayWulf	416	1107	1152	70000	1.04	0.506

Amdahl Numbers for Data Sets



The Data Sizes Involved



DISC Needs Today

- Disk space, disk space, disk space!!!!
- Current problems not on Google scale yet:
  - *10-30TB easy, 100TB doable, 300TB hard*
  - *For detailed analysis we need to park data for several months*
- Sequential IO bandwidth
  - *If analysis is not sequential for large data set, we cannot do it*
- How do can move 100TB within a University?
  - *1Gbps* *10 days*
  - *10 Gbps* *1 day (but need to share backbone)*
  - *100 lbs box* *few hours*
- From outside?
  - *Dedicated 10Gbps or FedEx*

Tradeoffs Today

Stu Feldman: Extreme computing is about tradeoffs

Ordered priorities for data-intensive scientific computing

1. *Total storage* *(-> low redundancy)*
2. *Cost* *(-> total cost vs price of raw disks)*
3. *Sequential IO* *(-> locally attached disks, fast ctrl)*
4. *Fast streams* *(->GPUs inside server)*
5. *Low power* *(-> slow normal CPUs, lots of disks/mobo)*

The order will be different every year…

JHU Data-Scope

- Funded by NSF MRI to build a new ‘instrument’ to look at data
- Goal: 102 servers for $1M + about $200K switches+racks
- Two-tier: performance (P) and storage (S)
- Large (6.5PB) + cheap + fast (500+GBps), but …
  . ..a special purpose instrument

	Revised					
	1P	*1S*	*All P*	*All S*	*Full*	
servers	1	1	90	6	102	
rack units	4	34	360	204	564	
capacity	24	720	2160	4320	6480	TB
price	8.8	57	8.8	57	792	$K
power	1.4	10	126	60	186	kW
GPU*	1.35	0	121.5	0	122	TF
seq IO	5.3	3.8	477	23	500	GBps
random IO	240	54	21600	324	21924	kIOPS
netwk bw	10	20	900	240	1140	Gbps

Everything is a Fractal

- The Data-Scope created a lot of excitement but also a lot of fear at JHU…
  - *Pro: Solve problems that exceed group scale, collaborate*
  - *Con: Are we back to centralized research computing?*
- Clear impedance mismatch between monolithic large systems and individual users
  - *Multi-tier architecture needed, like the LHC model*
- eScience needs different tradeoffs from eCommerce
- Larger systems are more efficient
- Smaller systems have more agility
- How to make it all play nicely together?

Cloud vs Cloud

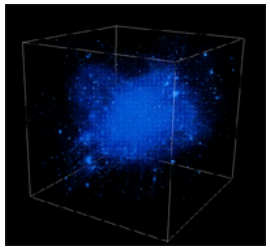
- Economy of scale is clear
- However:
  - *Commercial clouds are too expensive for Big Data*
  - *Smaller private clouds with special features are emerging*
  - *May become regional gateways to larger-scale centers*
  - *Trust!!!!*
- The “Long Tail” of a huge number of small data sets
  - *The integral of the “long tail” is big!*
- Facebook: bring many small, seemingly unrelated data to a single cloud and new value emerges
  - *What is the science equivalent?*

Technology+Sociology+Economics

- Neither of them is enough
  - *We have technology changing very rapidly*
  - *Sensors, Moore's Law*
  - *Trend driven by changing generations of technologies*
- Sociology is changing in unpredictable ways
  - *YouTube, tagging,…*
  - *In general, people will use a new technology if it is*
    - Offers something entirely new
    - Or substantially cheaper
    - Or substantially simpler
- It is not granted that if we build it they will come…
- Funding is essentially level

Changing Sociology

- Broad sociological changes
  - *Convergence of Physical and Life Sciences*
  - *Data collection in ever larger collaborations*
  - *Virtual Observatories: CERN, VAO, NCBI, NEON, OOI,…*
  - *Analysis decoupled, off archived data by smaller groups*
  - *Emergence of the citizen/internet scientist*
  - *Impact of demographic changes in science*
- Need to start training the next generations
  - *Π-shaped vs I-shaped people*
  - *Early involvement in "Computational thinking"*



Summary

- Science is increasingly driven by data (large and small)
- Large data sets are here, COTS solutions are not
- Changing sociology
- From hypothesis-driven to data-driven science
- We need new instruments: “microscopes” and “telescopes” for data
- There is also a problem on the “long tail”
- Similar problems present in business and society
- Data changes not only science, but society
- A new, Fourth Paradigm of Science is emerging…

A convergence of statistics, computer science, physical and life sciences…..

"If I had asked people what they wanted, they would have said faster horses…"

Henry Ford

From a recent book by Eric Haseltine:
"Long Fuse and Big Bang"