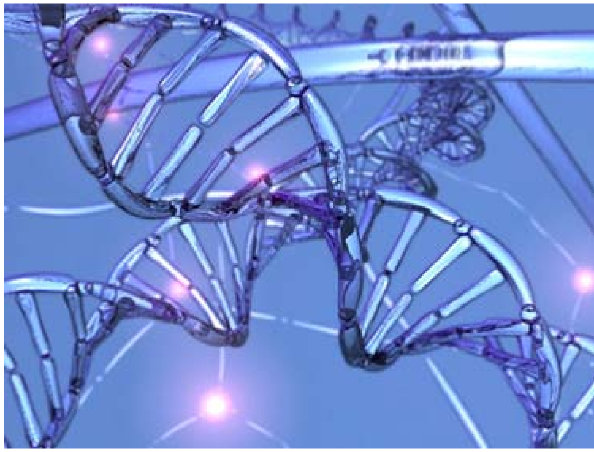
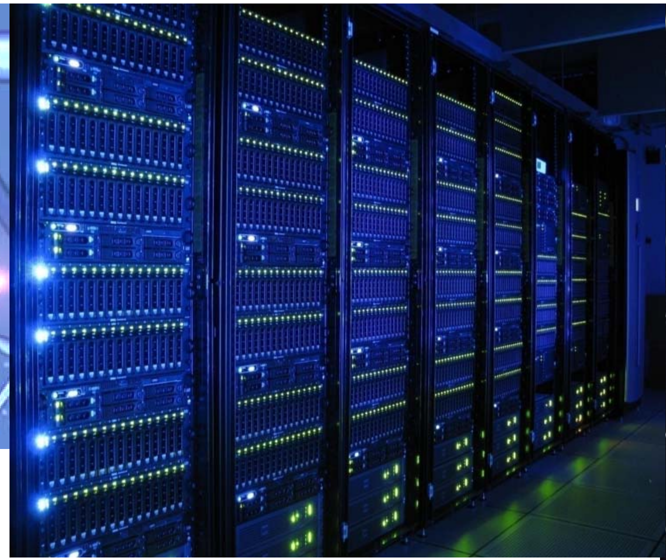




Exascale Numerical Laboratories

Alex Szalay
JHU

idies

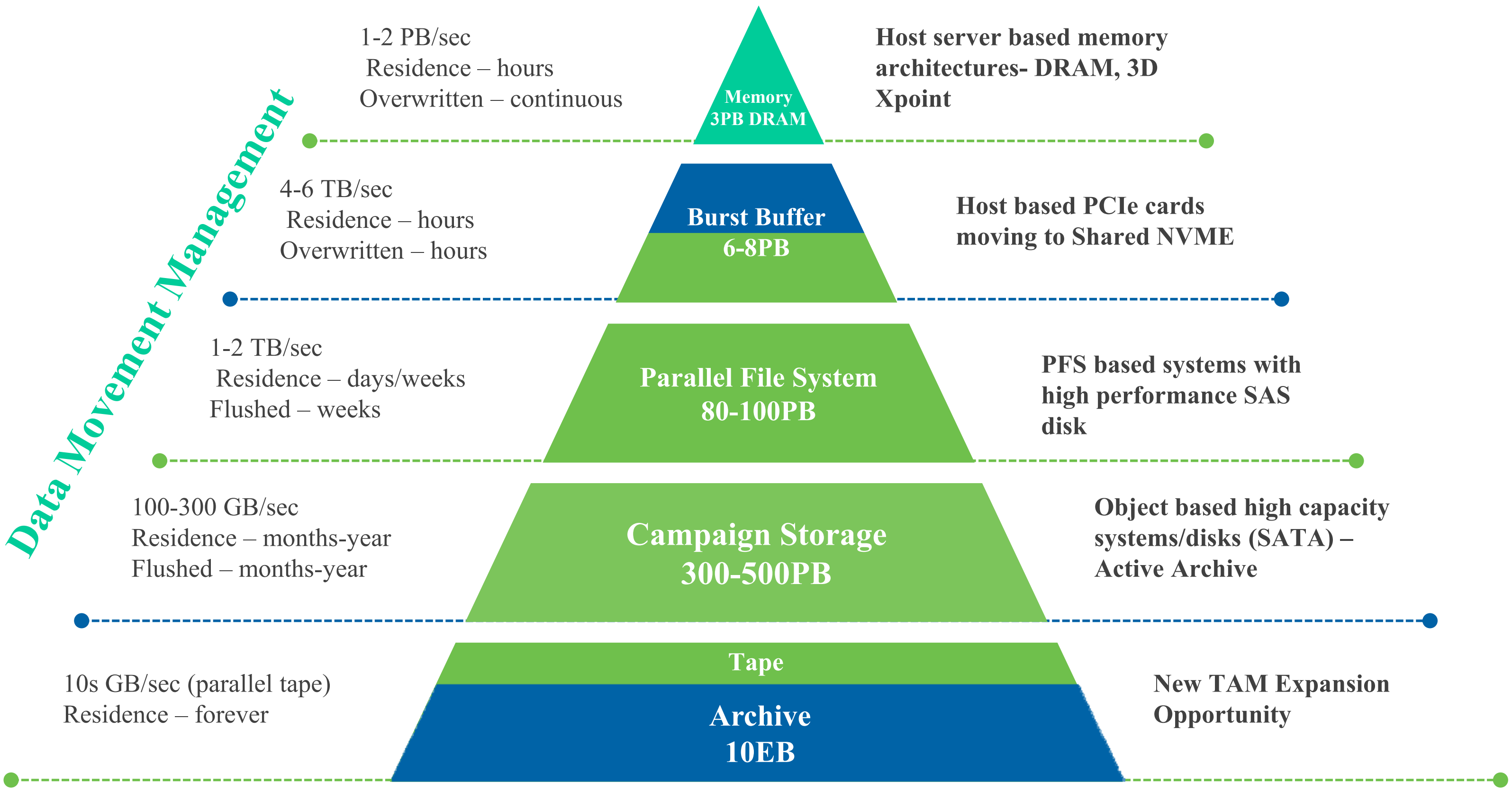
Data in HPC Simulations

- HPC is an instrument in its own right
- Largest simulations approach petabytes today
  - *from supernovae to turbulence, biology and brain modeling*
- Need public access to the best and latest through interactive **Numerical Laboratories**

- Examples in turbulence, N-body
- Streaming algorithms (annihilation, halo finders)
- Exascale coming

Towards Exascale

The 'Trinity' System at LANL is leading the way



Exascale Numerical Laboratories

- **Interactive analysis** of simulations becoming popular
  - *Comparing simulation and observational data crucial!*
- Similarities between Turbulence/CFD, N-body, ocean circulation and materials science
- Differences as well in the underlying data structures
  - *Particle clouds / Regular mesh / Irregular mesh*
- Innovative access patterns appearing
  - *Immersive virtual sensors/Lagrangian tracking*
  - *Posterior feature tagging and localized resimulations*
  - *Machine learning on HPC data*
  - *Joins with user derived subsets, even across snapshots*
  - *Data driven simulations/feedback loop/active control of sims*

Numerical Simulations

- HPC became an instrument in its own right
  - *Largest simulations exceed several petabytes*
  - *Directly compare to the experiments*
- Need public access to the best and latest
  - *Cannot just do in-situ analyses*
  - *Ensembles of simulations for UQ*
- Different access patterns
  - *What architectures can support these?*
- On Exascale everything will be a Big Data problem
  - *Memory footprint will be >2PB*
  - *With 5M timesteps => 10,000 Exabytes/simulation*
- Hard tradeoffs – cannot store it all
  - *We cannot keep all the snapshots*

How Do We Prioritize?

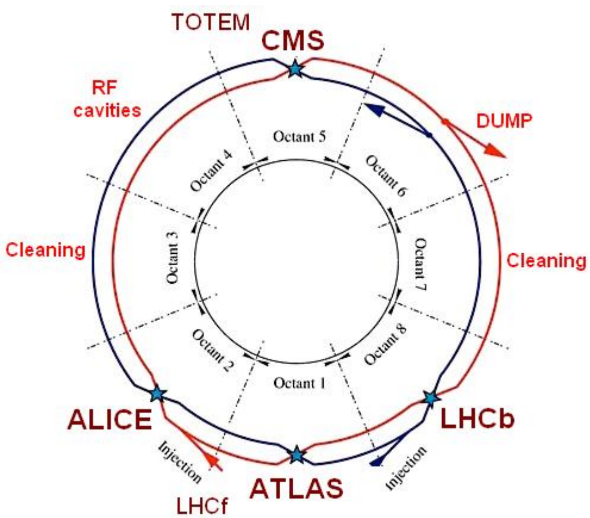
- Data Explosion: science is becoming data driven
- It is “too easy” to collect even more data
- Robotic telescopes, next generation sequencers, complex simulations

“Do you have enough data or
would you like to have more?”

- No scientist ever wanted less data….
- How can we decide how to collect data that is ***more relevant*** ?
- How to arrive at these tradeoffs?

LHC Lesson

- LHC has a single data source, $$$$$
- Multiple experiments tap into the beamlines
- They each use **in-situ** hardware triggers to filter data
  - *Only 1 in 10M events are stored*
  - *Not that the rest is garbage, just sparsely sampled*



- Resulting "small subset" analyzed many times **off-line**
  - *This is still 10-100 PBs*
- Keeps a whole community busy for a decade or more

Exascale Simulation Analogy

- Exascale computer running a community simulation
- Many groups plugging their own “triggers” (in-situ), the equivalents of “beamlines”
  - *Keep very small subsets of the data*
  - *Plus random samples from the field*
  - *Immersive sensors following world lines or light cones*
  - *Buffer of timesteps: save precursor of events*
- Sparse output analyzed offline by broader community
- Cover more parameter space and extract more realizations (UQ) using the saved resources

Architectural Implications

- In-situ: global analytics and “beamline” triggers, two stage, light-weight, and scheduler
- Simple API for community buy-in
- Very high selectivity to keep output on PB scales
- Burst buffers for near-line analyses
- Need to replace DB storage with smart object store with additional features (seek into objects)
- Build a fast DB-like index on top (SQL or key-value?) for localized access patterns
- Parallel high level scripting tools (iPython.parallel?)
- Simple immersive services and visualizations

Nature is Sparse

- Many natural phenomena are dominated by a few processes and described by a sparse set of parameters
- Compressed Sensing has emerged to find in high dimensional data the underlying sparse representation (Candes, Donoho, Tao, et al)
- This enables signal reconstruction with much less data!
- The resolution depends not on the pixel count but on the information content of an image…

Principal Component Pursuit

- Low rank approximation of data matrix: X
- Standard PCA:

$$\min \|X - E\|_2 \quad subject\ to\ \ rank(E) \le k$$

  - *works well if the noise distribution is Gaussian*
  - *outliers can cause bias*
- Principal component pursuit

$$\min \|A\|_0 \quad subject\ to\ \ X = N + A,\ \ rank(N) \le k$$

  - *"sparse" spiky noise/outliers: try to minimize the number of outliers while keeping the rank low*
  - *NP-hard problem*
- The L1 trick:

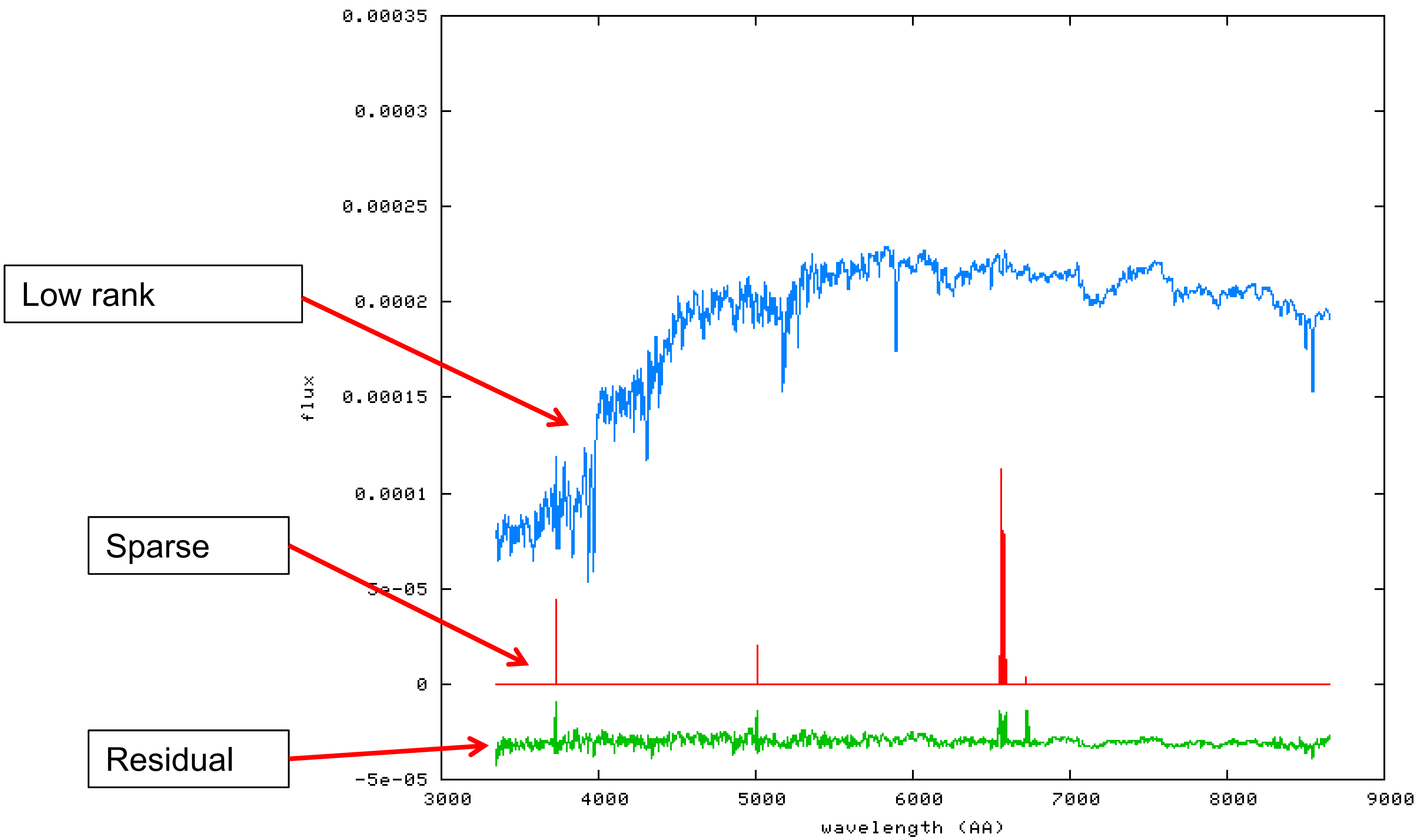
$$\min_{N,A} \left(\|N\|_* + \lambda \|A\|_1 \right) \ subject\ to\ \ X = N + A$$

  - *numerically feasible convex problem (Augmented Lagrange Multiplier)*

$$\min_{N,A} \left(\|N\|_* + \lambda \|A\|_1 \right) \quad subject\ to\ \ \|X - (N + A)\|_2 < \varepsilon$$

* E. Candes, et al. "Robust Principal Component Analysis". preprint, 2009.
Abdelkefi et al. ACM CoNEXT Workshop (traffic anomaly detection)

Principal component pursuit

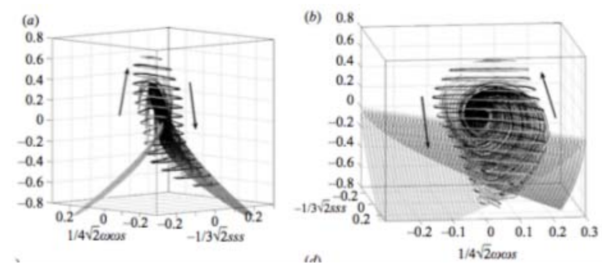


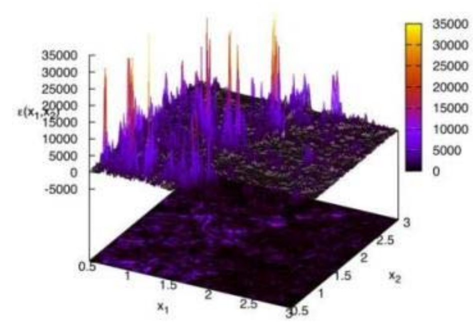
λ=0.6/sqrt(n), ε=0.03

Immersive Turbulence

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

- **Understand the nature of turbulence**
  - *Consecutive snapshots of a large simulation of turbulence: 30TB*
  - *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 step was 50TB MHD simulation*
  - *Channel flow 100TB, MHD2 256TB*





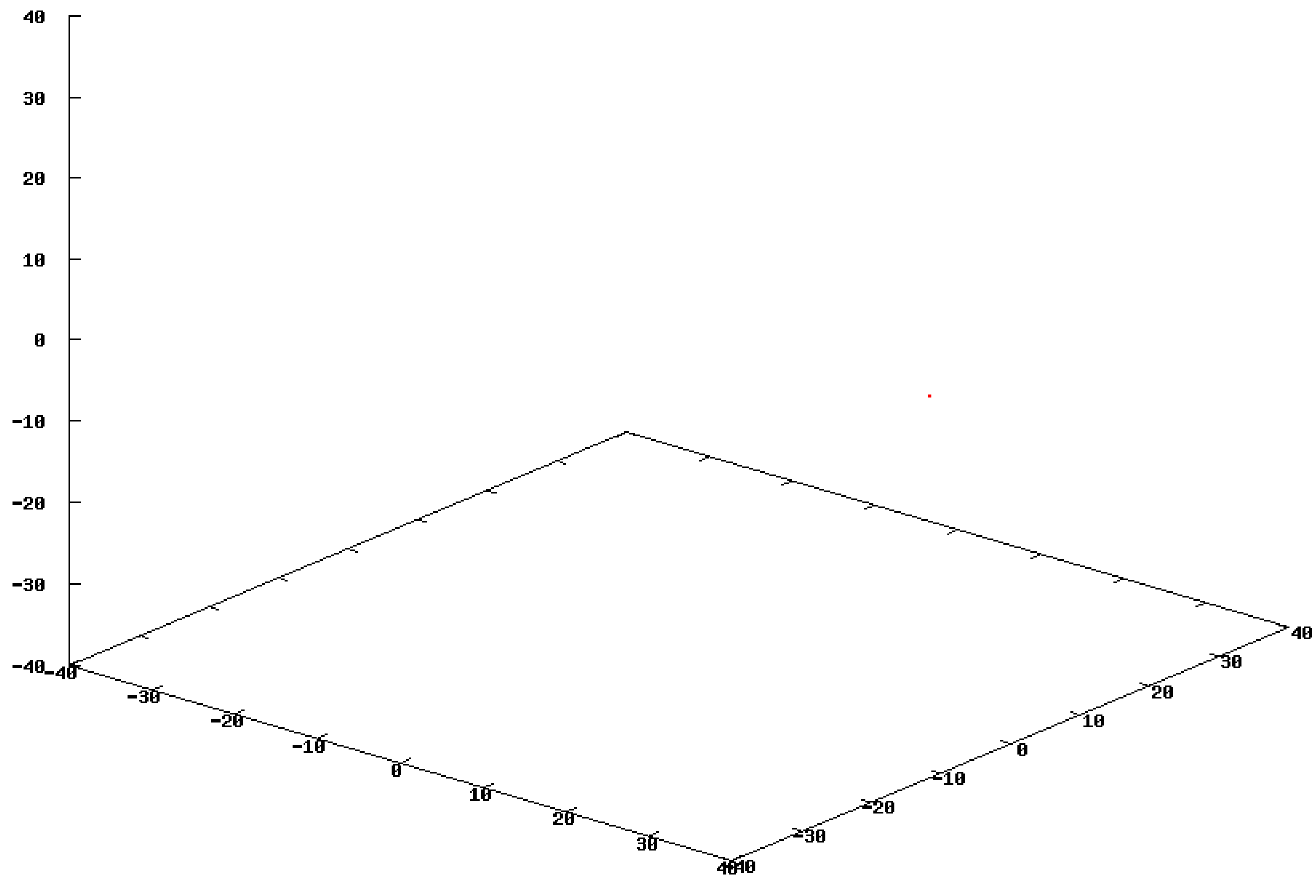
- **New paradigm** for analyzing simulations

20 trillion points queried to date!

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

Bring Your Own Dwarf (Galaxy)

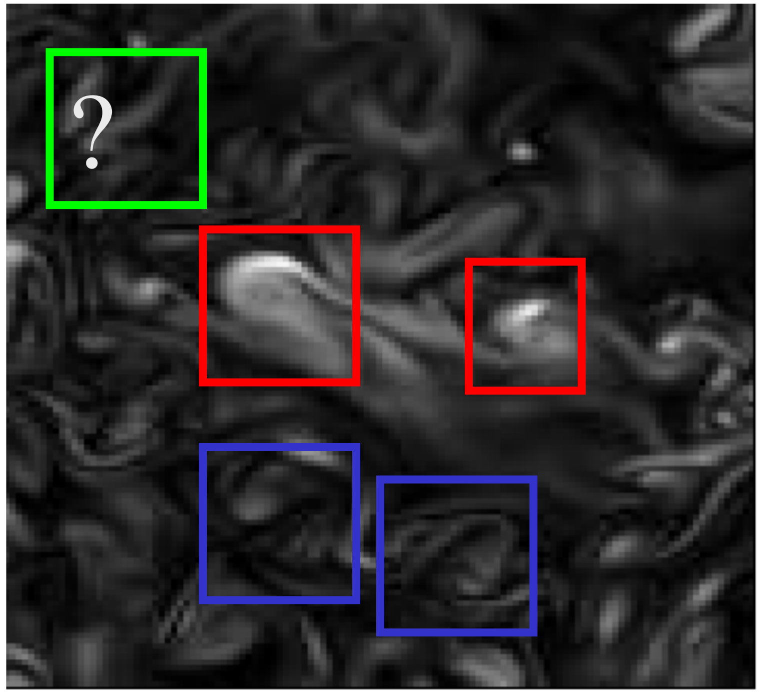
Wayne Ngan
Brandon Bozek
Ray Carlberg
Rosie Wyse
Alex Szalay
Piero Madau



Via Lactea-II
forces from halos

Machine Learning in Turbulence

Renyi divergence



Vorticity

Similarity between regions ⇒

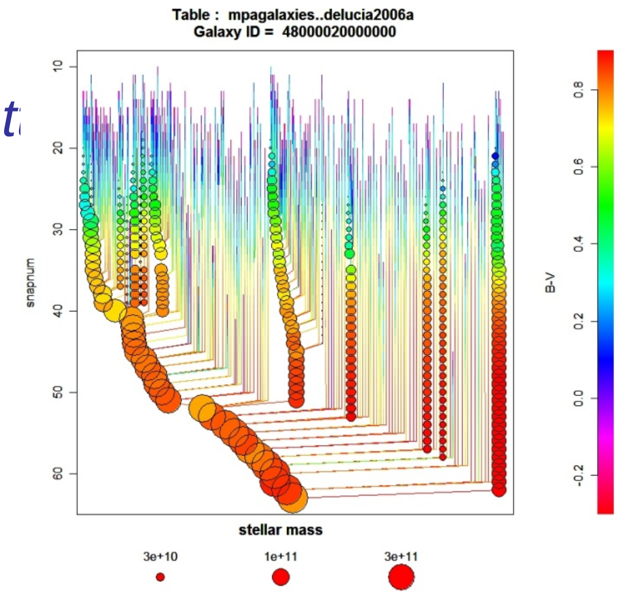
- ❑clustering,
- ❑classification,
- ❑anomaly detection

with J. Schneider, B. Poczos, CMU

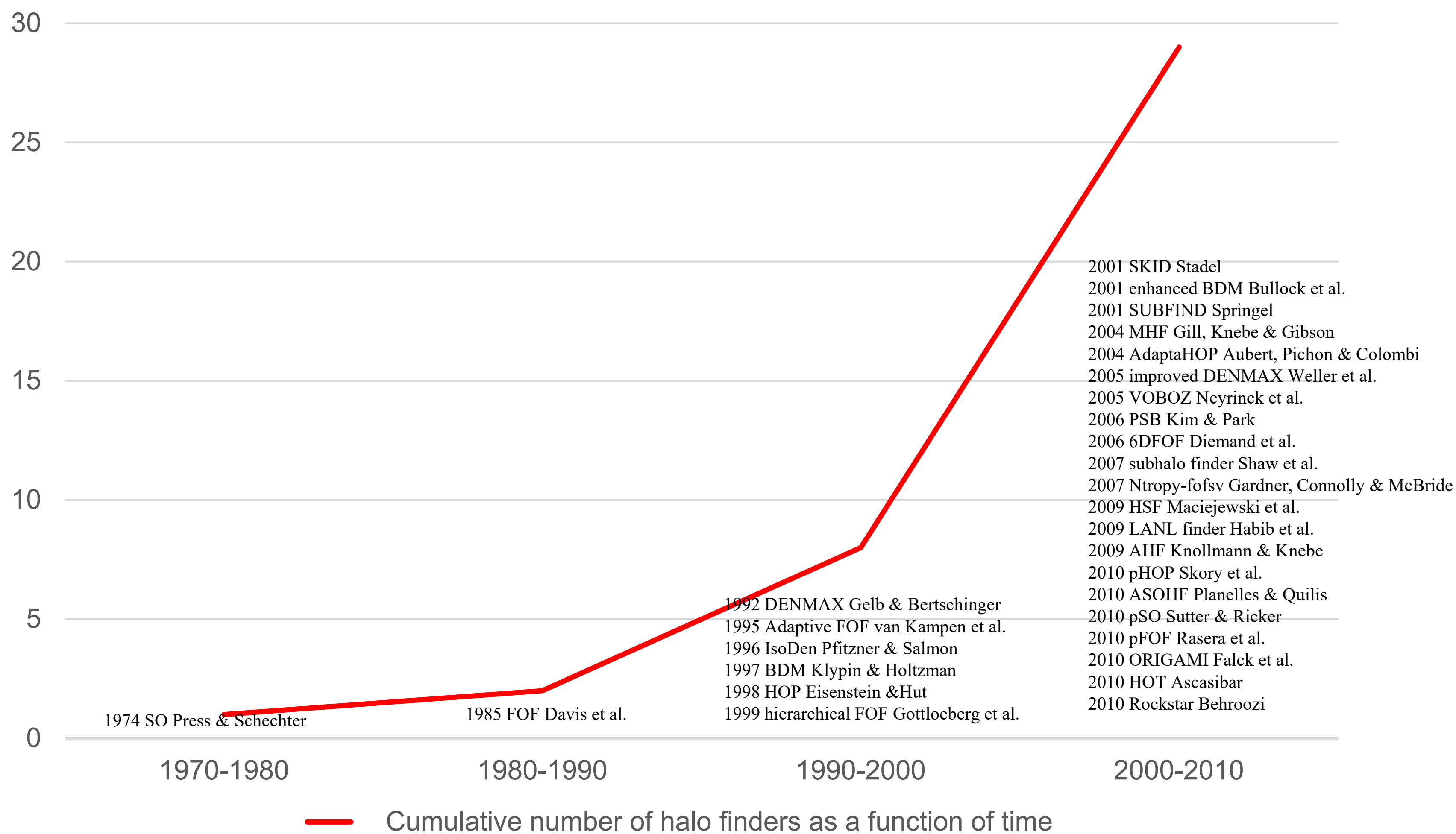
Cosmology Simulations

- Simulations are becoming an instrument on their own
- Millennium DB is the poster child/ success story
  - *Built by Gerard Lemson*
  - *600 registered users, 17.3M queries, 287B rows*

    http://gavo.mpa-garching.mpg.de/Millennium/
  - *Dec 2012 Workshop at MPA: 3 days, 50 people*
- Data size and scalability
  - *PB data sizes, trillion particles of dark mat*
- Value added services
  - *Localized*
  - *Rendering*
  - *Global analytics*



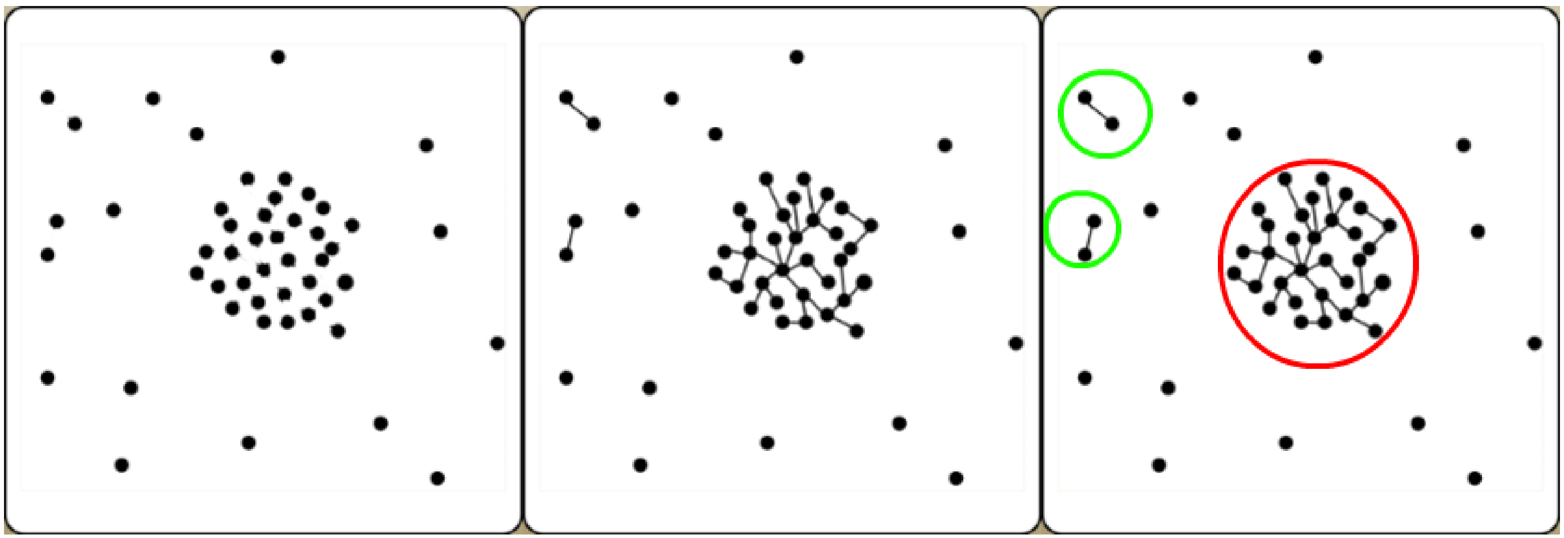
Halo finding algorithms



The Halo-Finder Comparison Project
[Knebe et al, 2011]

Friends-of-Friends Algorithm

- FOF is one of the very first halo finding algorithms [Davis et al, 1985]
- Simple conceptually, is the first step in many other algorithms
- Has a single free parameter called the linking length θ.
  - *Two particles are “friends” if the distance less than θ.*
  - *Two particles are in the same cluster if there exists a chain of links between them.*



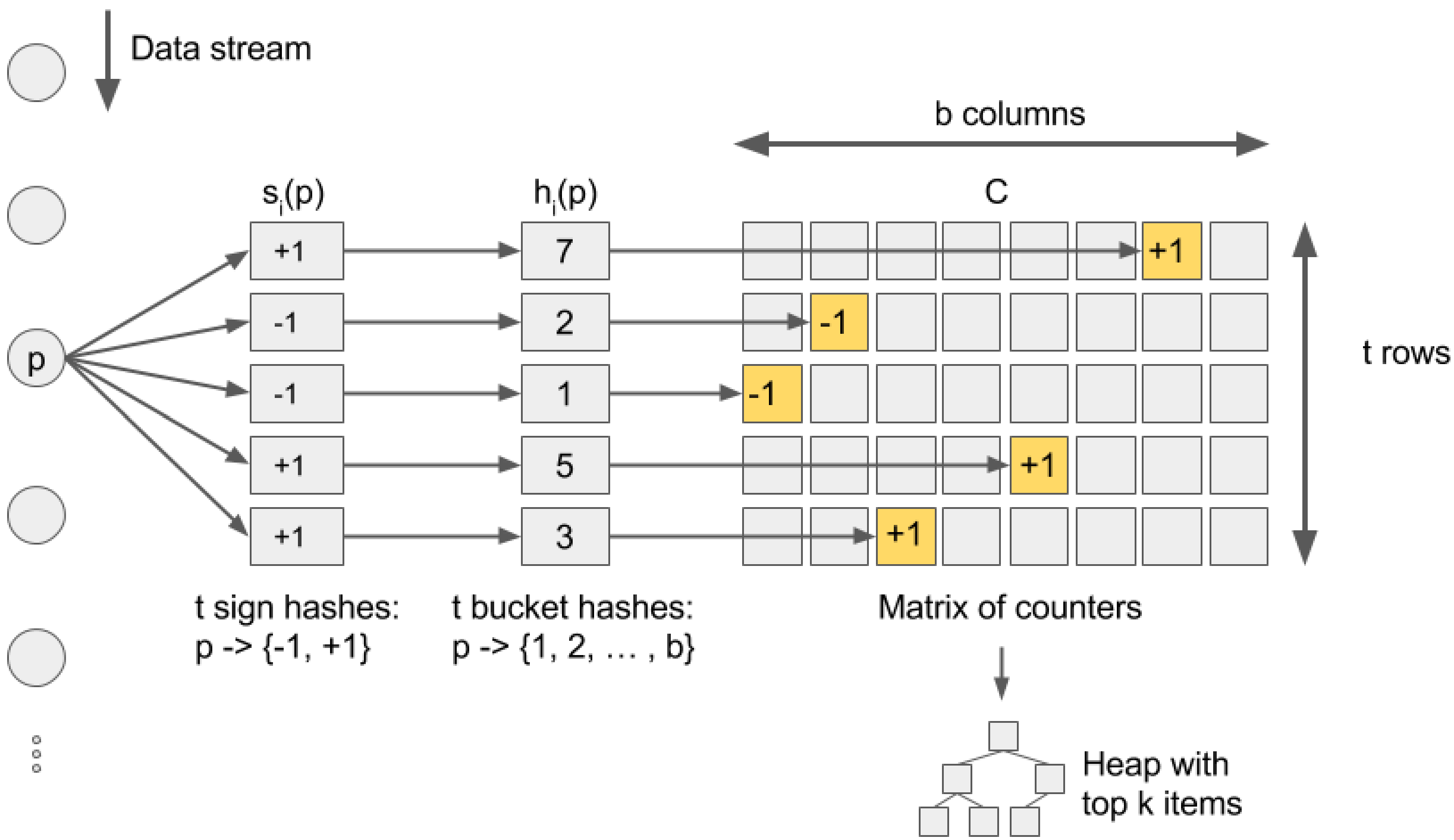
Approximate Aggregations

- Find the approximate top K cells in a simulation with 10B particles at a resolution of 0.1Mpc $(5K)^3$ cells, above a count threshold corresponding ~1M cells, and characterize the uncertainty
- Brute force would require a histogram of a size

$$5K^3 \text{ x } 4B = 500GB$$

- We can solve it in 0.5GB (fits on a low-end GPU)
- We use the Count-Sketch algorithm for finding heavy hitters

Count Sketch



Memory

- Memory is the most significant advantage of applying streaming algorithms.
- Dataset size: $\sim 10^{10}$ particles (Millennium DM)
  - *Any in-memory algorithm: 120 GB+*
  - *Count-Sketch: 640 MB*
- GPU acceleration
  - *One instance of Count-Sketch algorithm can be fully implemented by separate thread of GPU*
  - *Different parts of the volume (use PH index) can go to different streams/GPUs*
- Results are quite insensitive to subsampling by a factor of 2-8

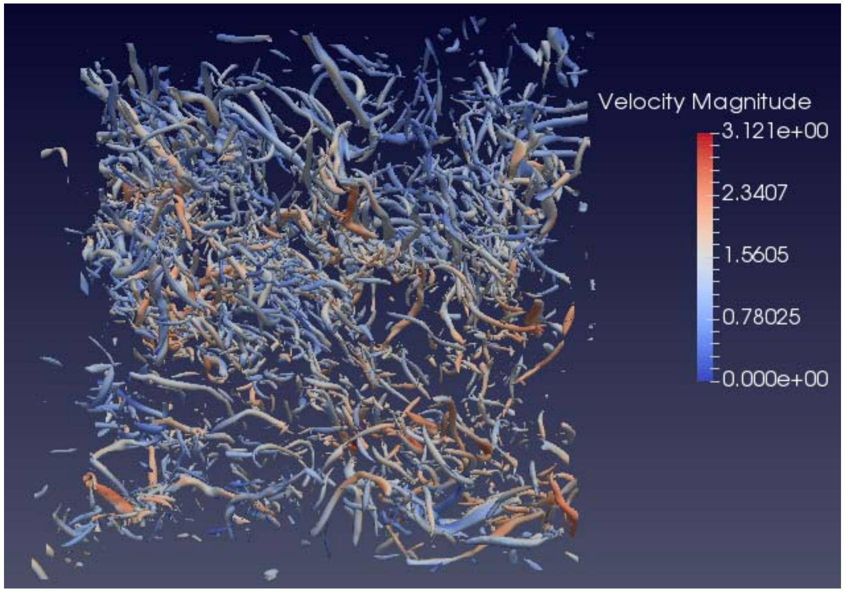
Testing Burst Buffer Triggers

- Use data in the Trinity Burst Buffers
- Allocate about 2%of CPU to compute triggers in-situ
- Store results in secondary storage for viz
- Extract high-vorticity regions from turbulence simulation
- Data compression/reduction not very high (5:1) for this use, but good illustration of concept
- Model also applies to light-cones in N-body, cracks in Material Science

Hamilton, Burns, Ahrens, Szalay et al (2016)

Data Extraction: Vorticity Mesh

- Extraction technique where high vorticity (Q-magnitude) regions are defined by setting a vorticity threshold
- Marching cubes are utilized to create a mesh structure around high vorticity regions
- Results in significant reduction of data since original velocity data is discarded and only mesh data is stored
- Provides good visualization of vorticity

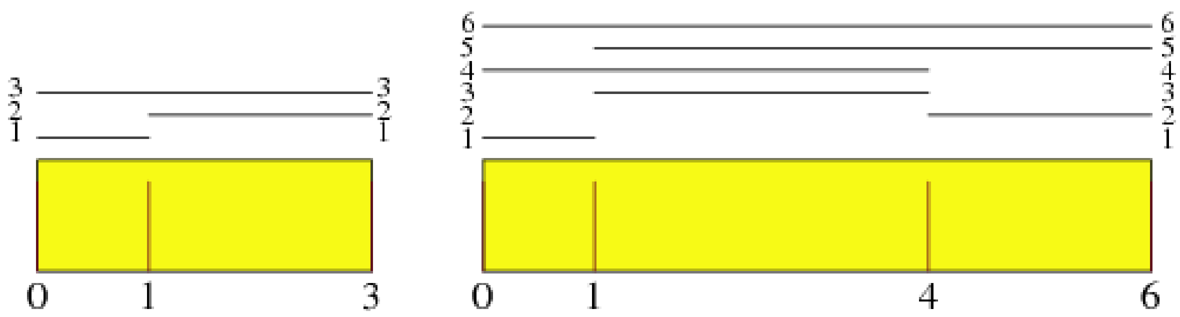


Temporal Sampling

Temporal correlations need uniform sampling of the differences

Golomb's Ruler

- In one dimension, set of marks at integer positions, so that their distances computed over all possible pairs are distinct
- If it measures all distances up to its length, it is "perfect"



- Sparse sample timesteps, with a Golomb Ruler one can optimally estimate temporal correlations

N=7, L=25: 0 1 4 10 18 23 25

N=48, L = 1887

Summary

- Simulations are becoming first-tier instruments
- Changing sociology – archival storage analyzed by individuals
- Need Numerical Laboratories for the simulations
  - *Provide impedance matching between the HPC experts and the many domain scientists*
- Razor-sharp balance of in-situ triggers and off-line
- Need computable **approximate** statistics
- Streaming, sampling, robust techniques
- Clever in-situ use of burst buffers promising
- **On Exascale everything is a Big Data problem**