Meet the SciDAC Visualization and Analytics Center for Enabling Technologies

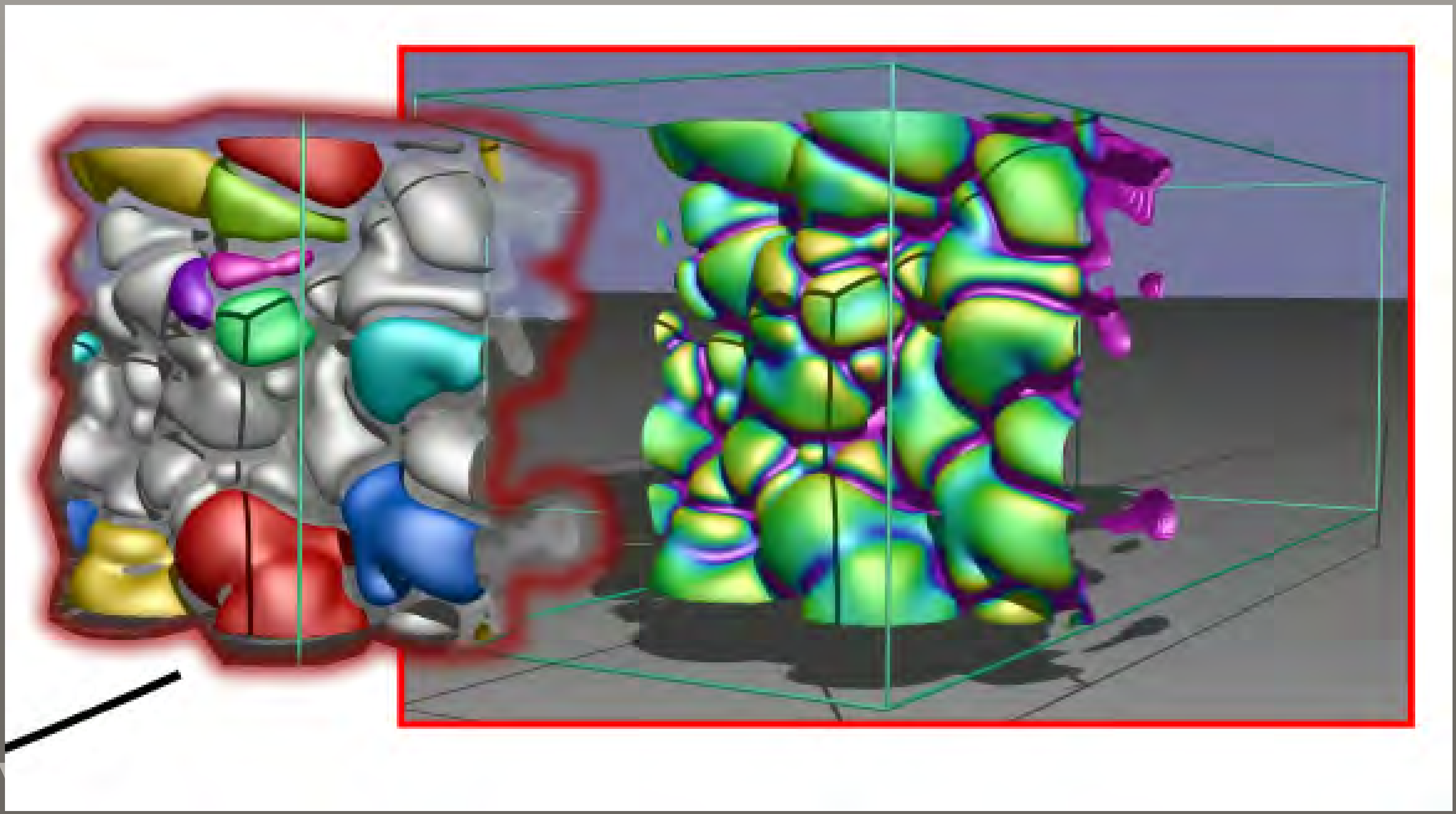
E. Wes Bethel (Coordinating PI)

LBNL

23 October 2008

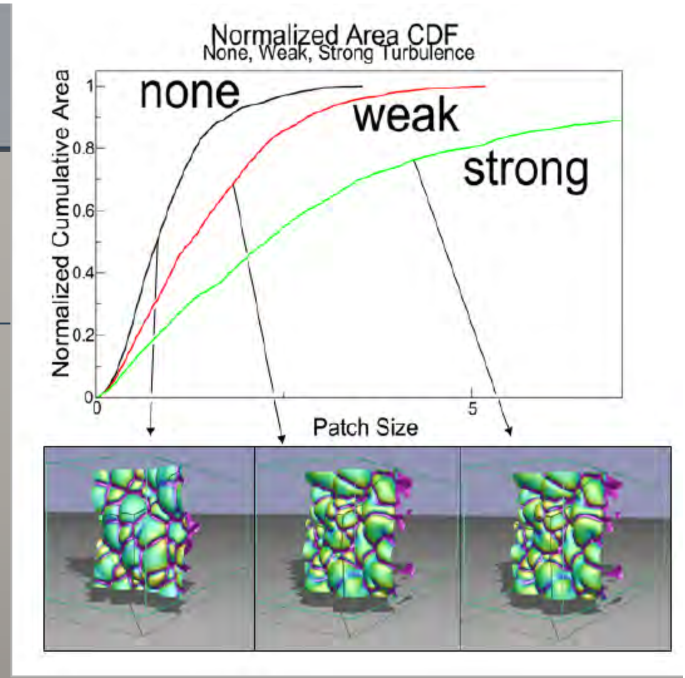
Case Study #1

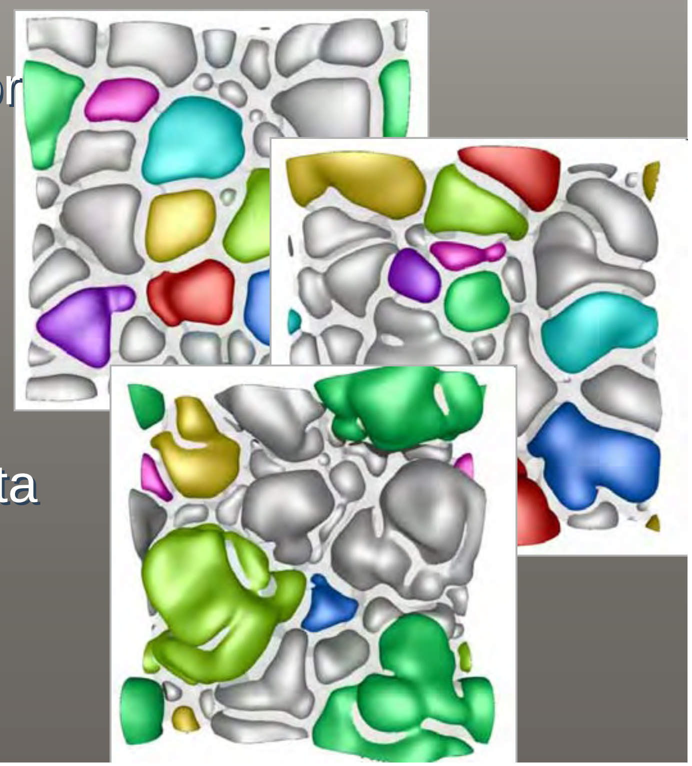
- Topological Analysis of Lean, Pre-mixed Hydrogen Flame Simulation



Combustion, Part 1

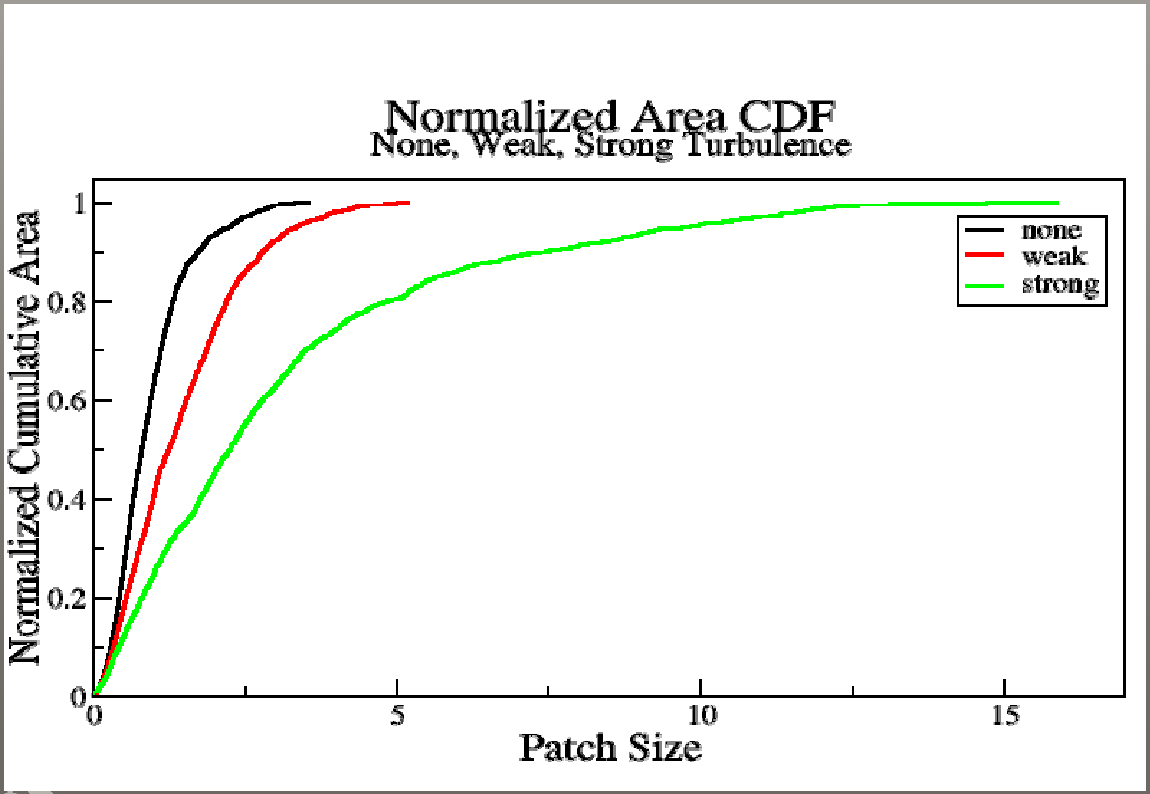
- PI: John Bell (LBNL), SciDAC Community Astrophysics Consortium Partnership, Incite Awardee.
- Accomplishments:
  - New topological analysis techniques for studying relationship between parameters and their effect.
  - Joint publications with stakeholder.
- Science Impact:
  - First-ever quantitative analysis large, time-varying combustion simulation data to study influence of turbulence on size/shape of combustion regions in lean, premixed hydrogen flames.





Dispelling Myths

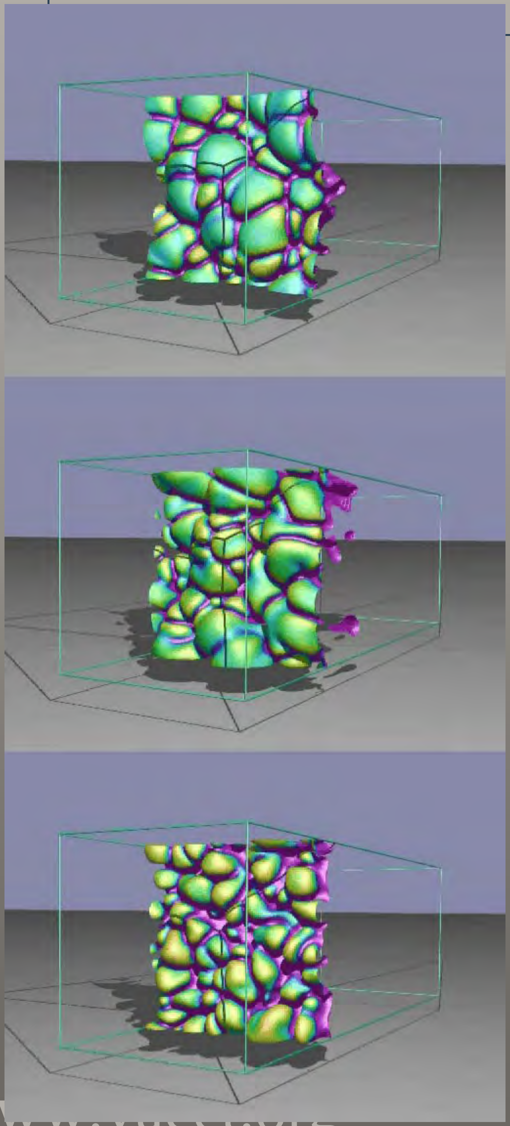
- You don't need sophisticated tools to do a simple x/y chart.



Target Application: Understanding Combustion Processes in Lean Premixed Flames

- Understanding combustion processes is important, for example, in engine and power plant design.
- Lean (fuel poor) flames are of interest since they reduce emissions.
- As the amount of fuel decreases creating stable flames becomes challenging.
- One major influence on the combustion process is the amount of turbulence imposed on the fuel air mixture

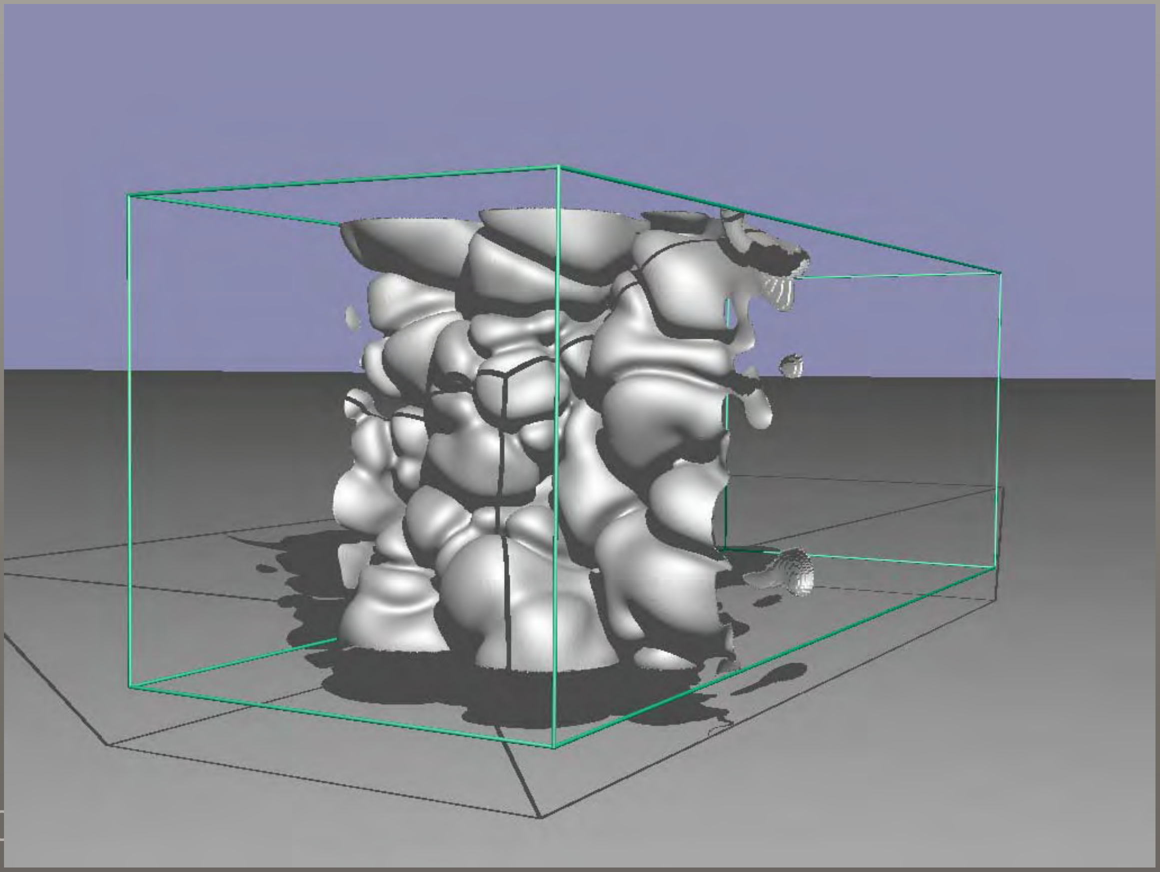
Analyzing Varying Levels of Turbulence



- **Input Data**: 621, 540, and 427 time steps of a 256x256x768 grid and 102, 82, and 91 time steps of a 512x512X1536 grid each storing temperature and fuel consumption rate ≈ 400GB compressed floating point data.
- **Objective**:
  - Analyze the cellular burning structures of the flame front as defined by the local fuel consumption rate.
  - Track individual burning cells to understand the temporal dynamics.
- **Challenges**: Extensive parameter studies are required to determine appropriate values but the large number of time steps make repeated analysis infeasible.

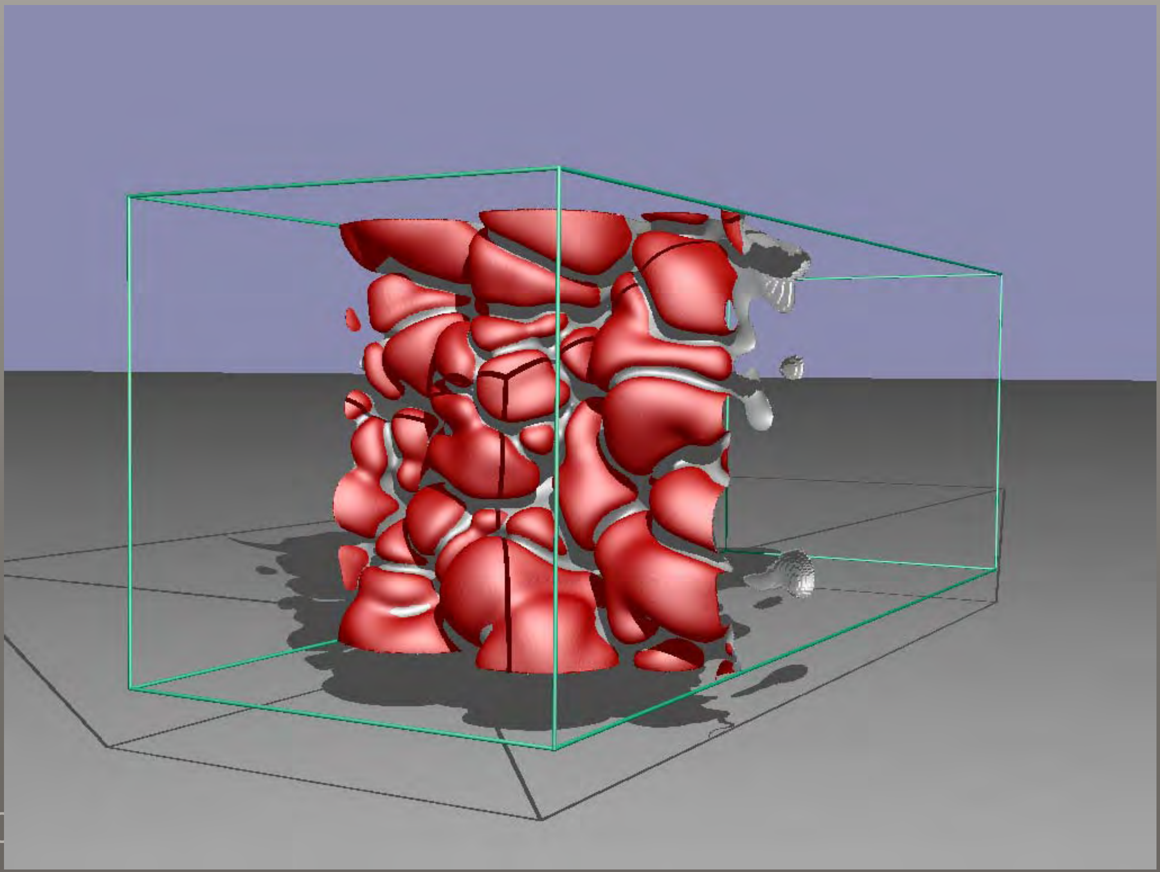
Feature Identification

- Extract the flame front as temperature isosurface.



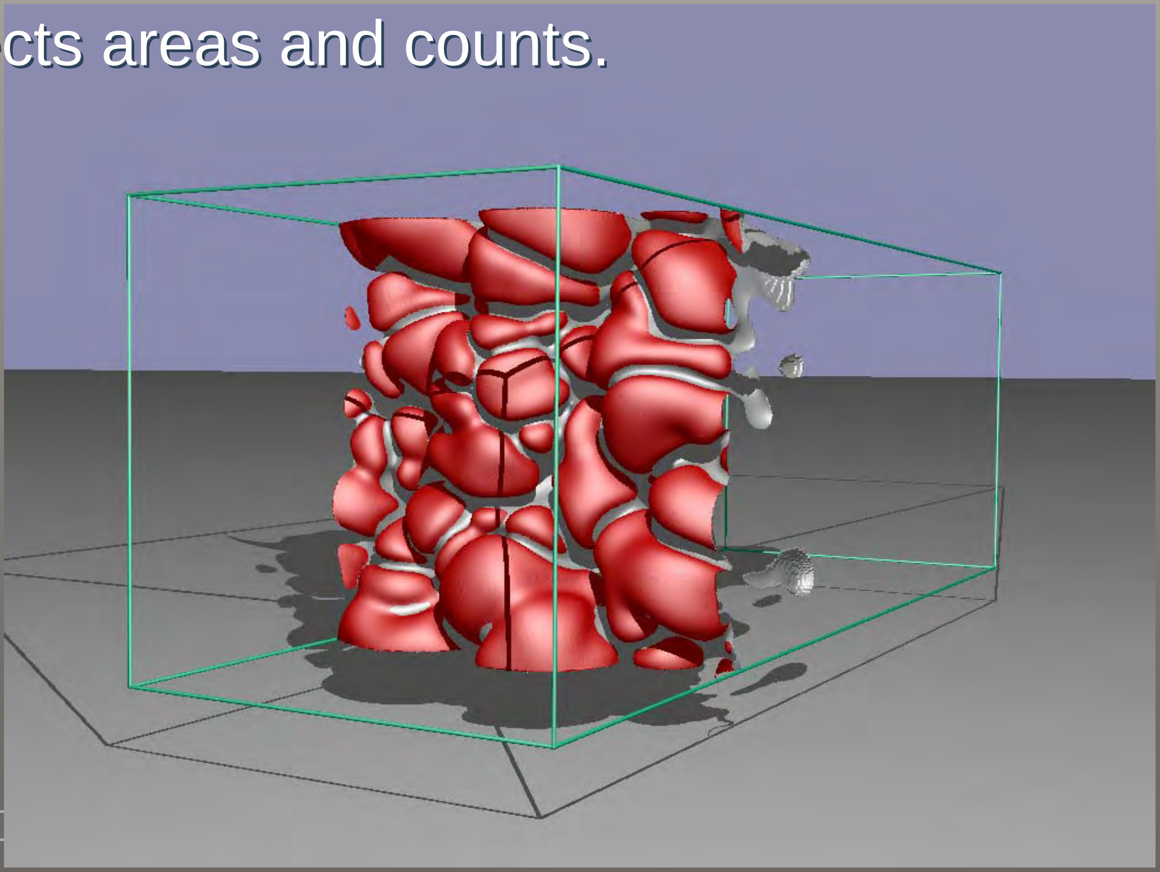
Feature Identification

- Extract the flame front as temperature isosurface.
- Threshold surface vertices.



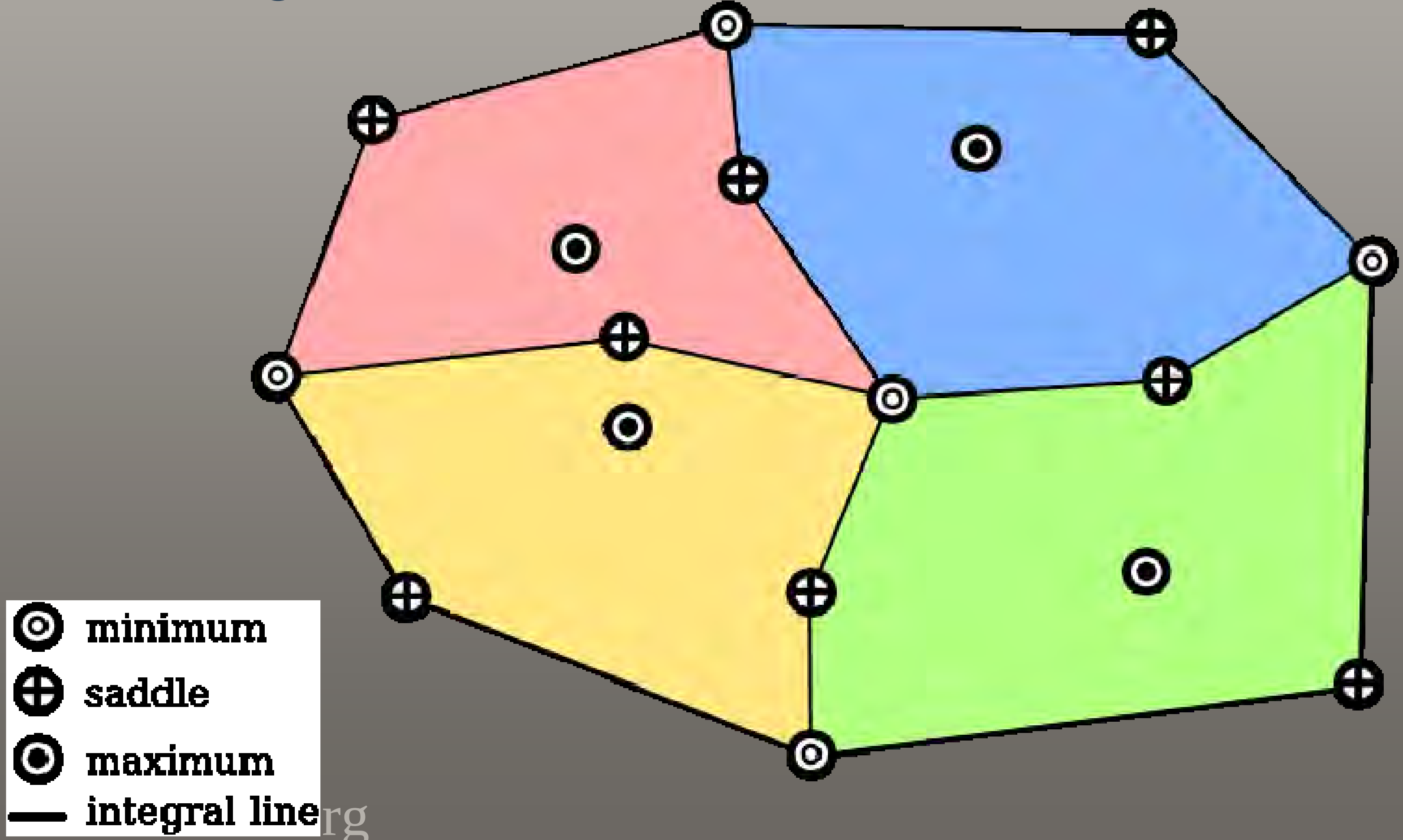
Feature Segmentation

- Extract the flame front as temperature isosurface.
- Threshold surface vertices.
- Collects areas and counts.



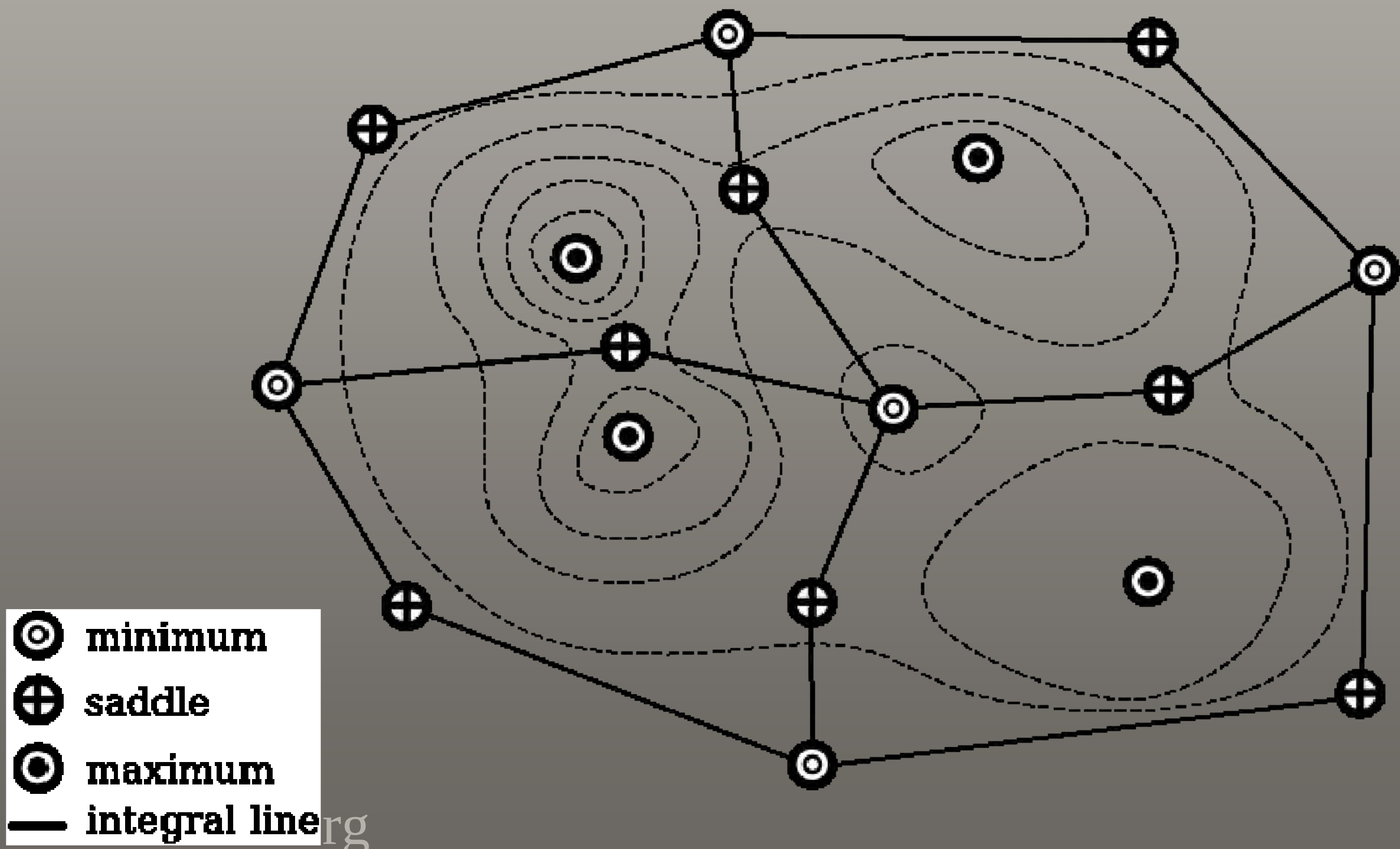
Express Segmentations Using the Morse Complex of the Fuel Consumption Rate

- Segment the surface into stable manifolds.



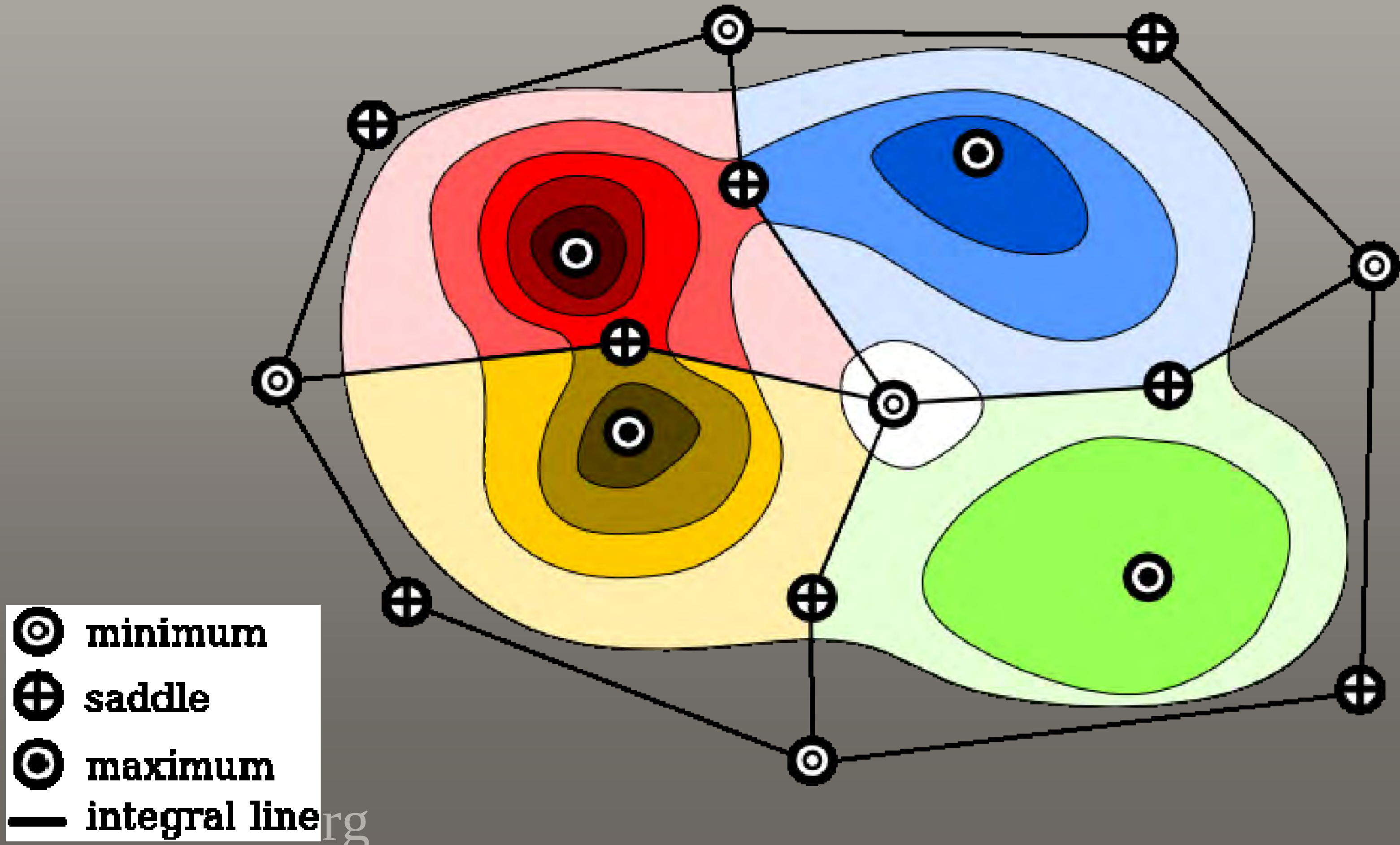
Express Segmentations Using the Morse Complex of the Fuel Consumption Rate

- Subdivide the function range into regular intervals.



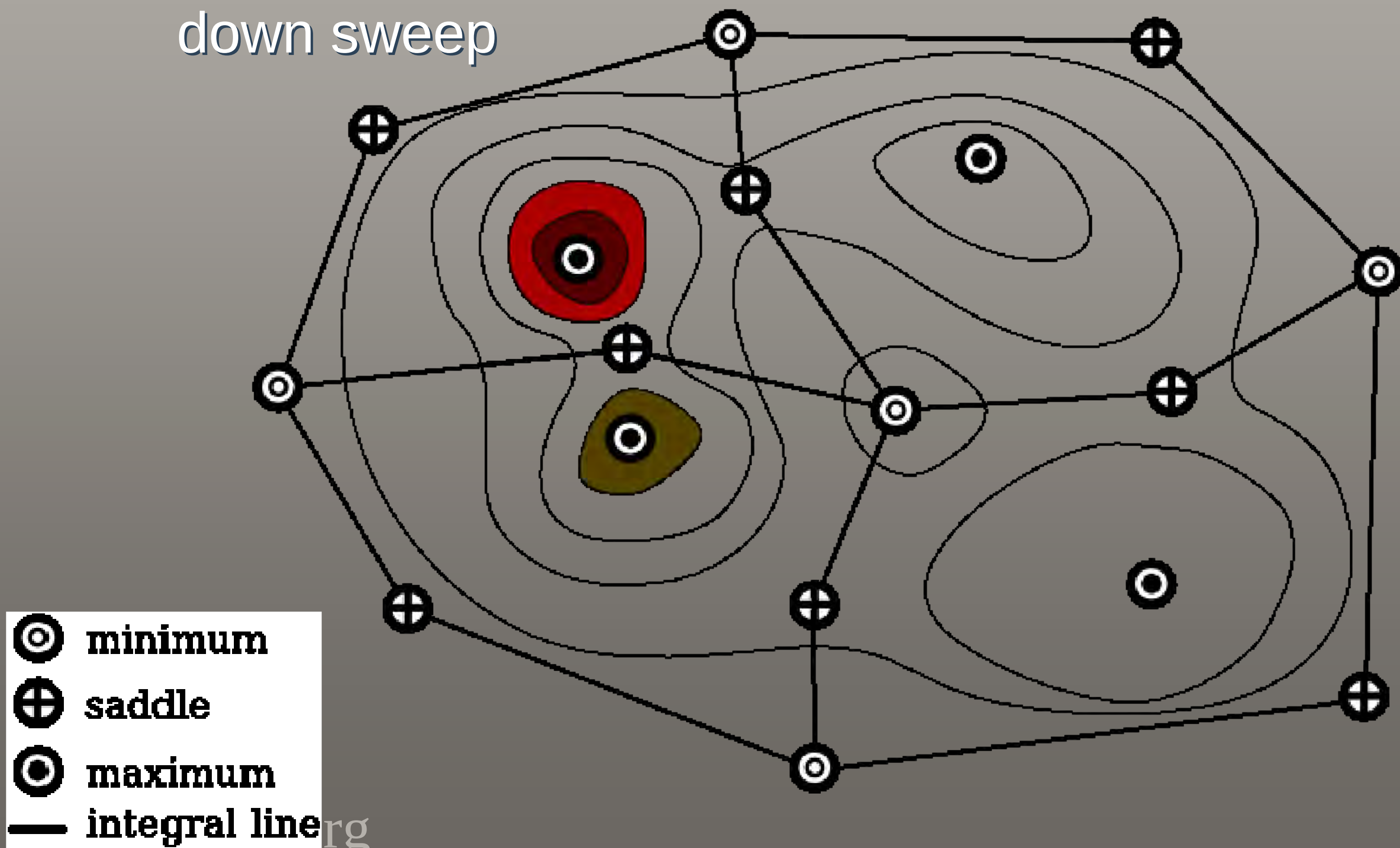
Express Segmentations Using the Morse Complex of the Fuel Consumption Rate

- Compute the surface area of each subrange.



Express Segmentations Using the Morse Complex of the Fuel Consumption Rate

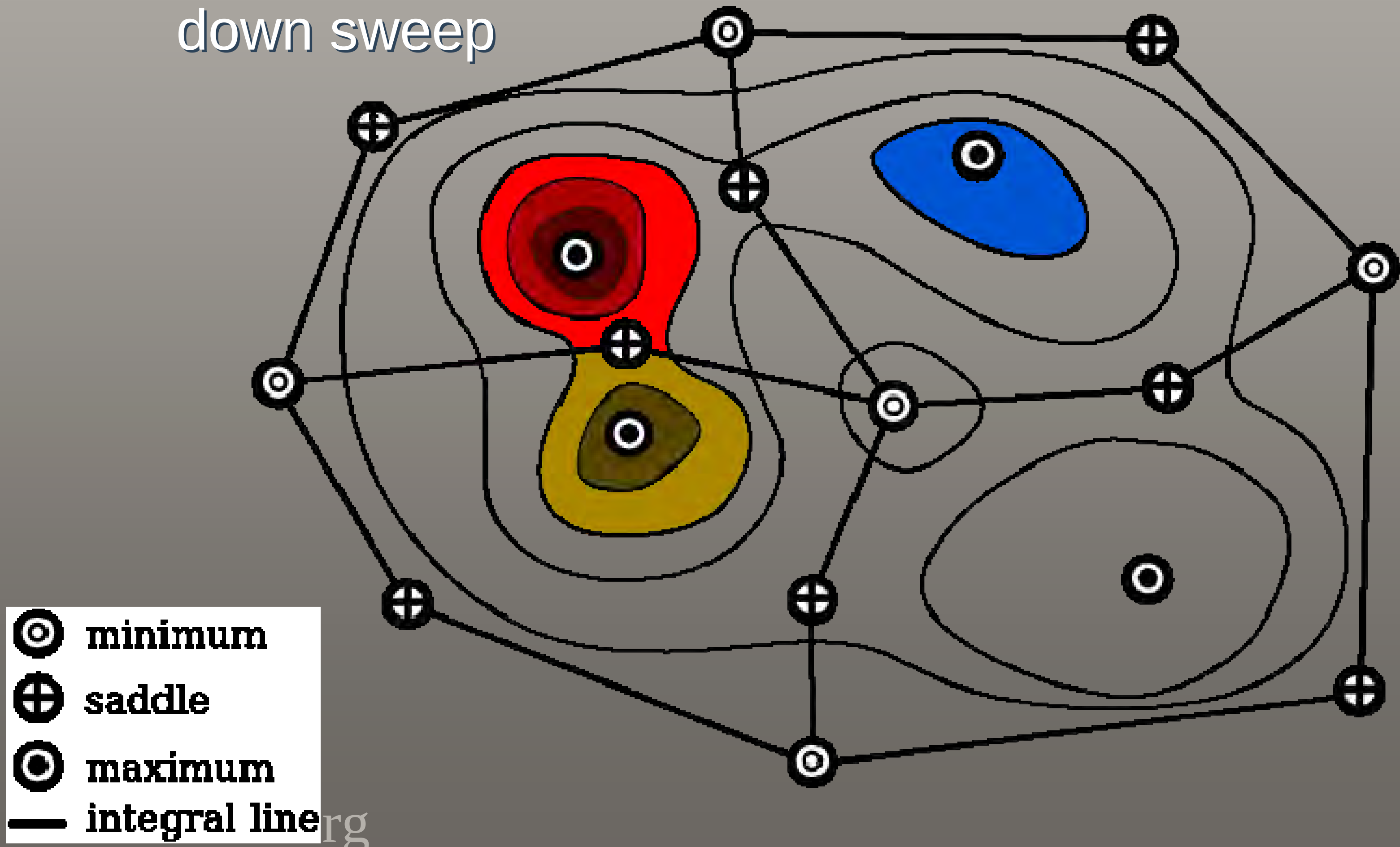
- Collect the areas/counts for all thresholds in a top-down sweep



rg

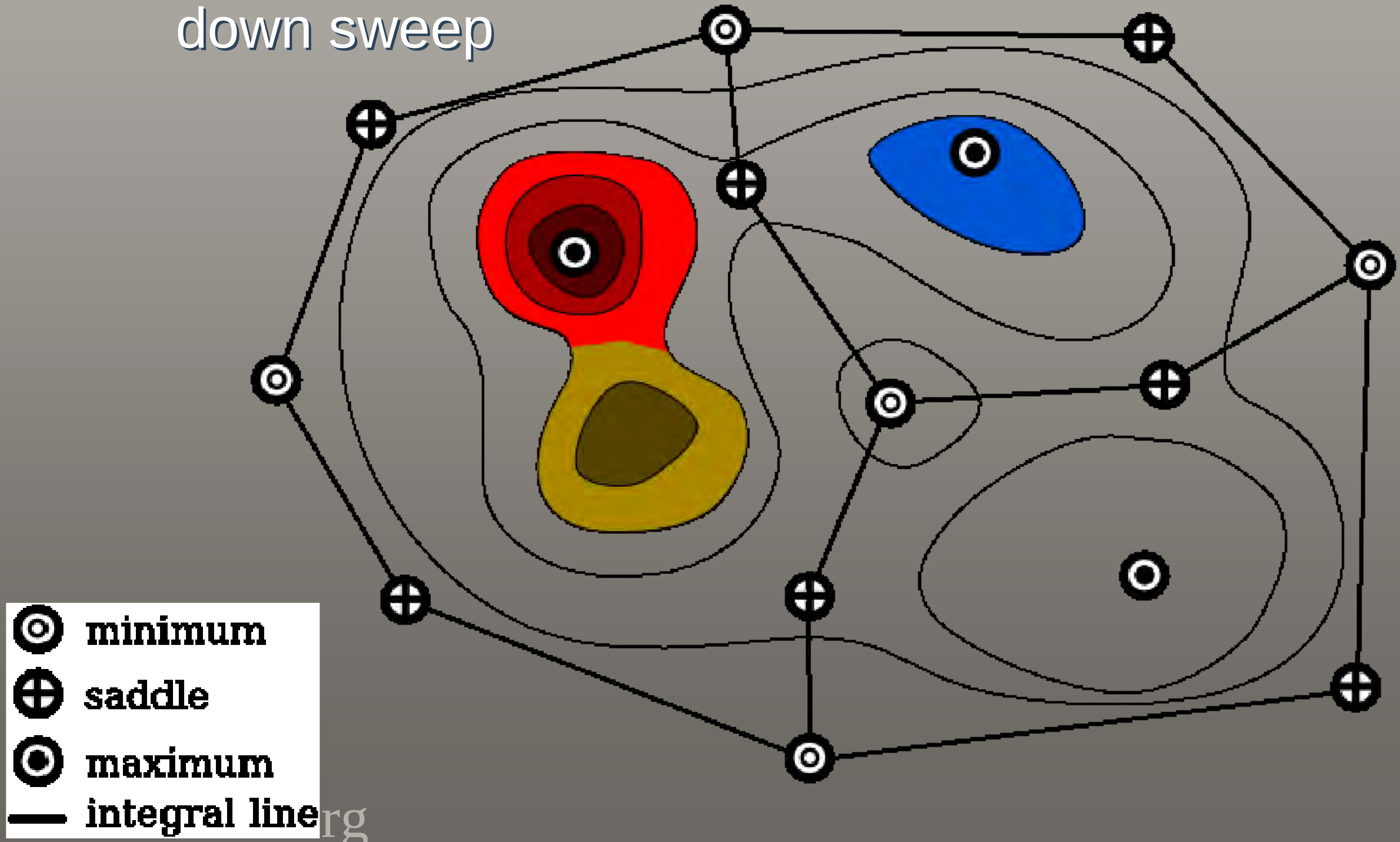
Express Segmentations Using the Morse Complex of the Fuel Consumption Rate

- Collect the areas/counts for all thresholds in a top-down sweep



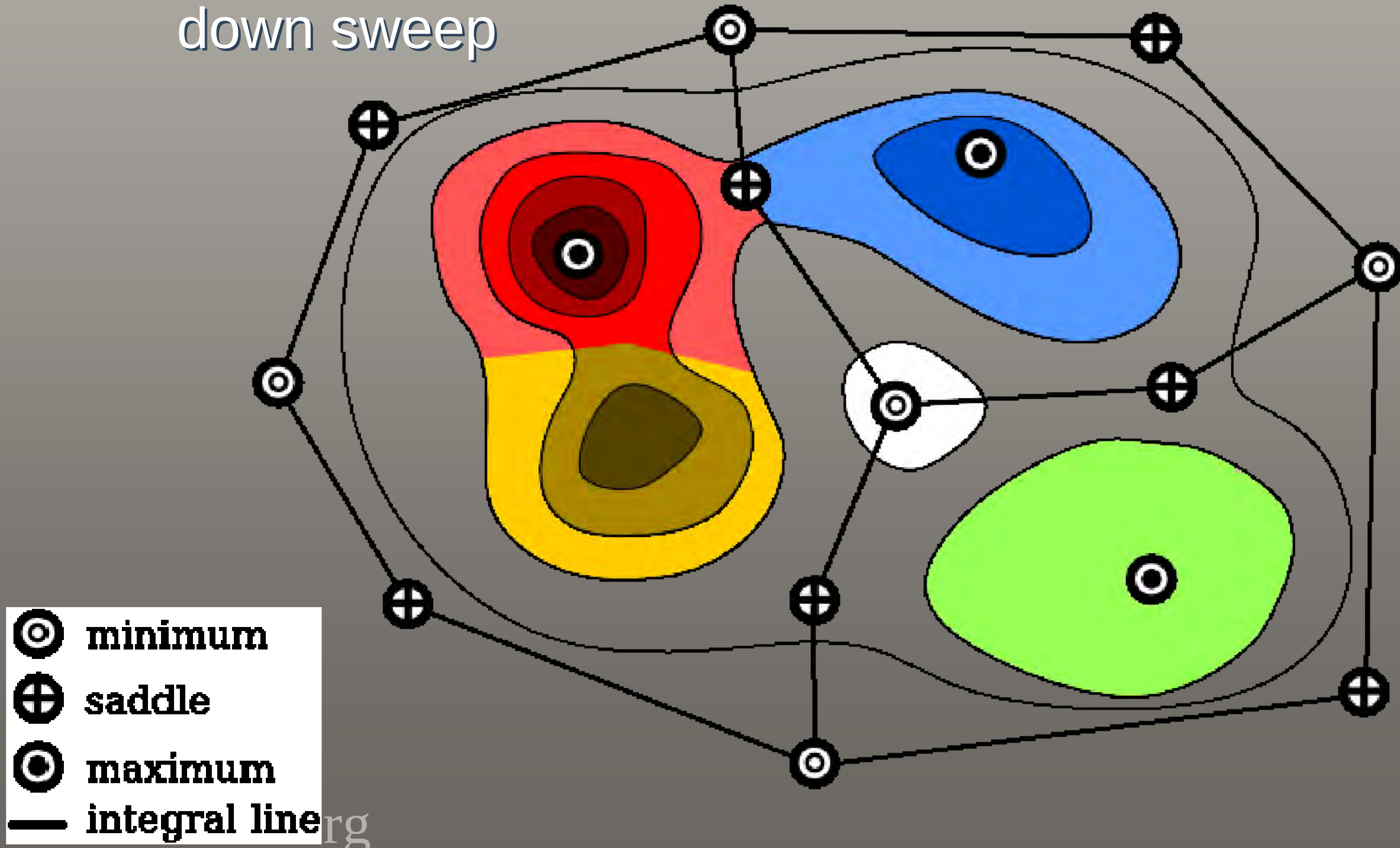
Express Segmentations Using the Morse Complex of the Fuel Consumption Rate

- Collect the areas/counts for all thresholds in a top-down sweep



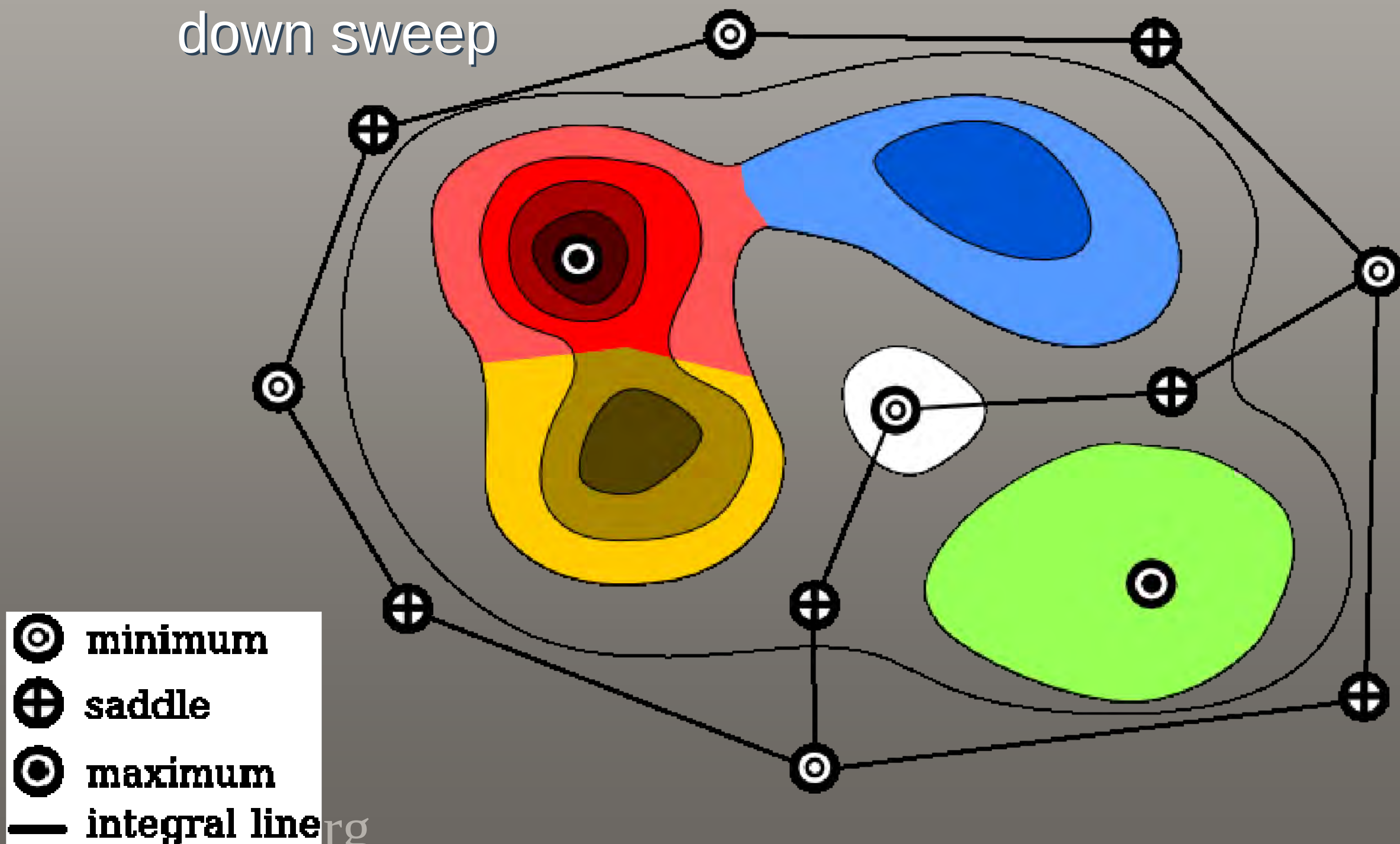
Express Segmentations Using the Morse Complex of the Fuel Consumption Rate

- Collect the areas/counts for all thresholds in a top-down sweep



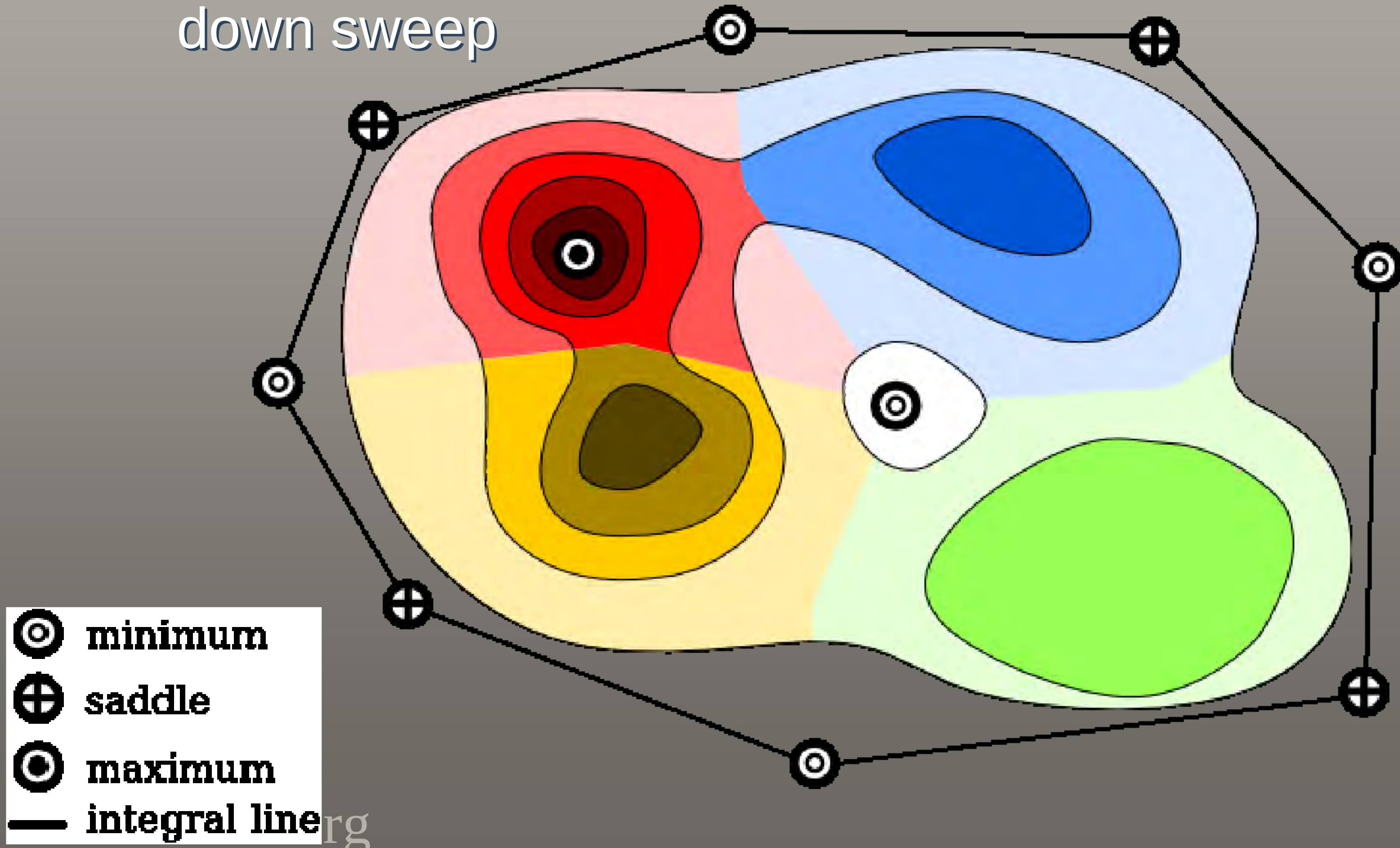
Express Segmentations Using the Morse Complex of the Fuel Consumption Rate

- Collect the areas/counts for all thresholds in a top-down sweep

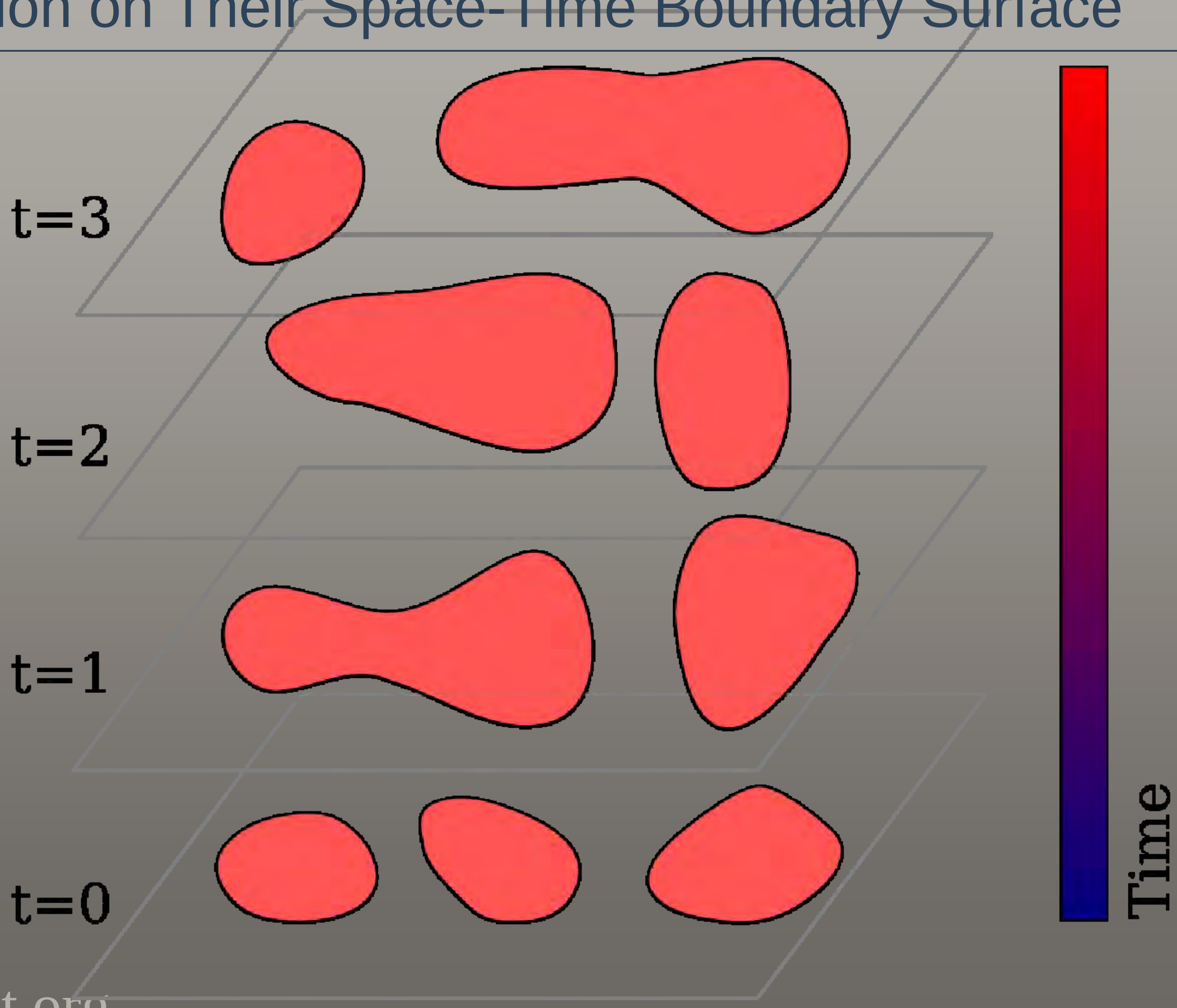


Express Segmentations Using the Morse Complex of the Fuel Consumption Rate

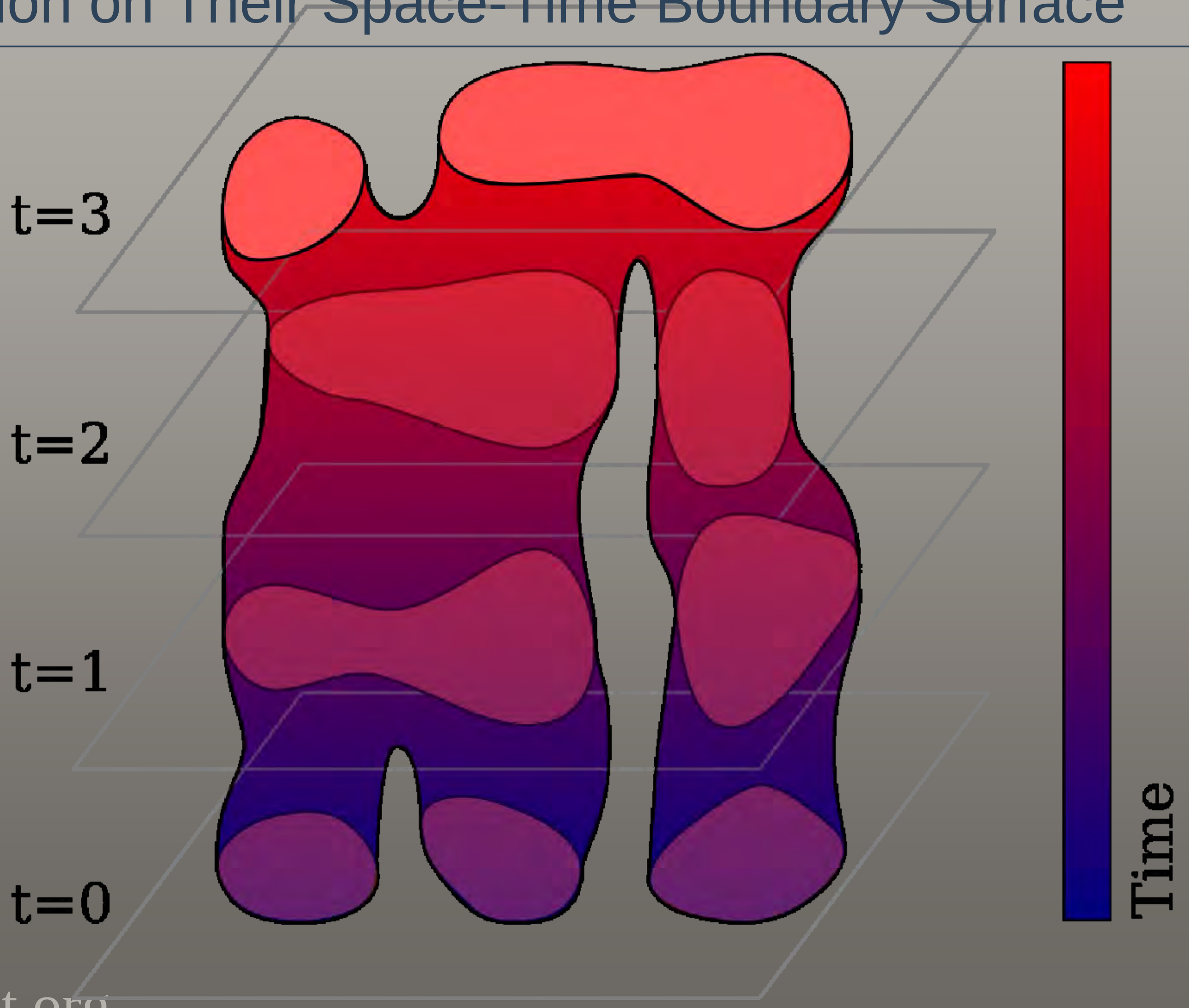
- Collect the areas/counts for all thresholds in a top-down sweep



Track Cells by Computing the Reeb Graph of Time as Function on Their Space-Time Boundary Surface

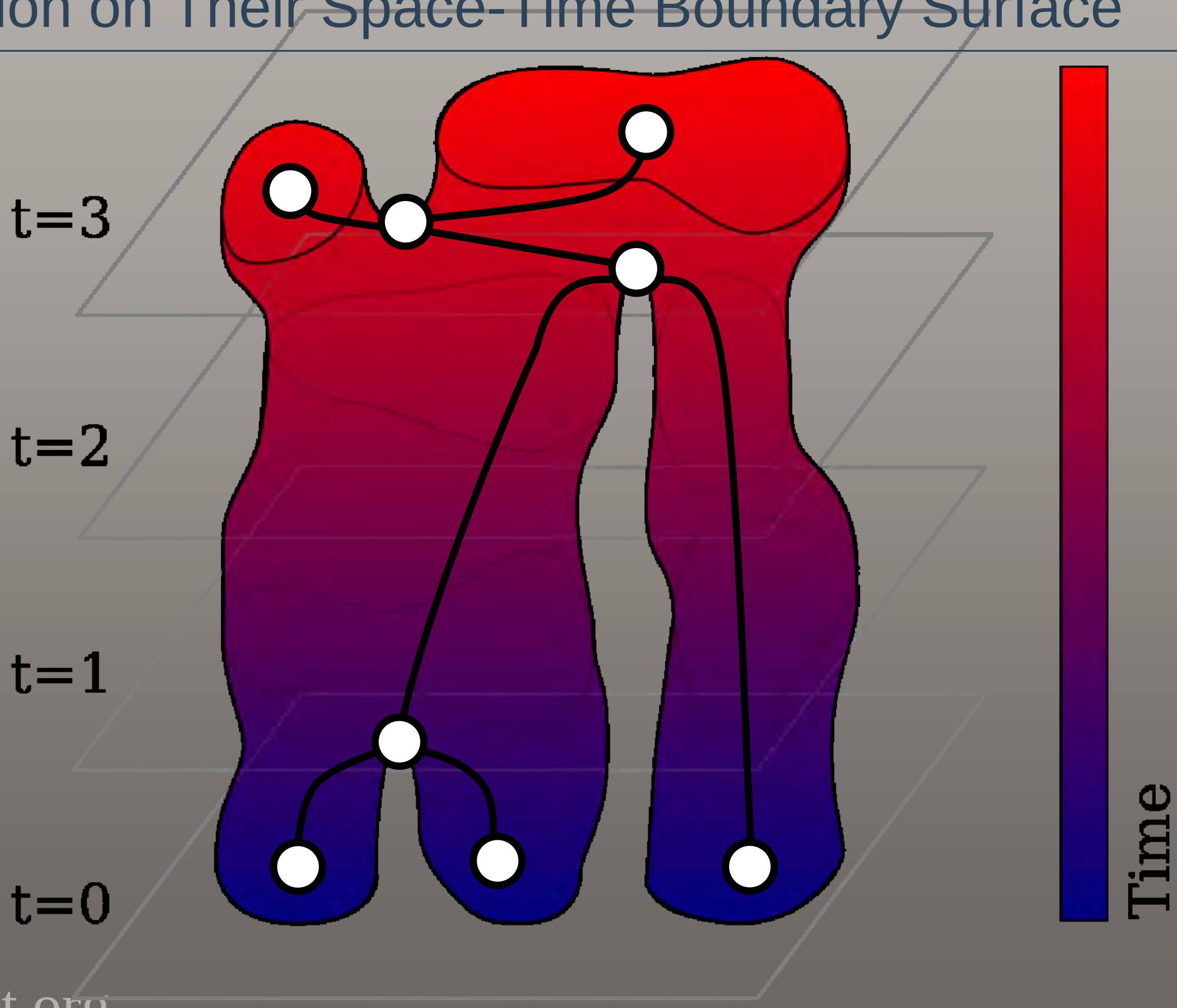


Track Cells by Computing the Reeb Graph of Time as Function on Their Space-Time Boundary Surface

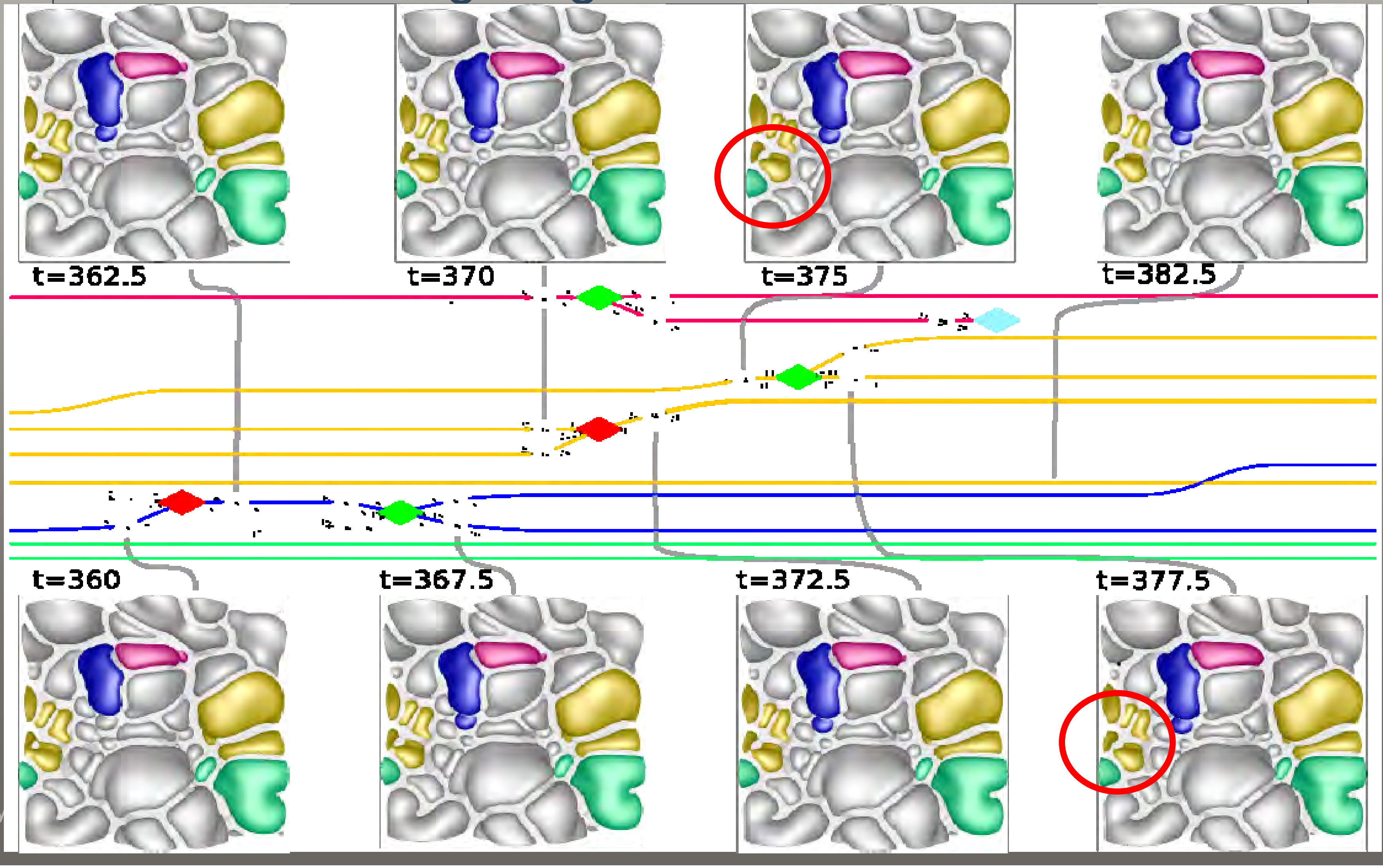


VACET

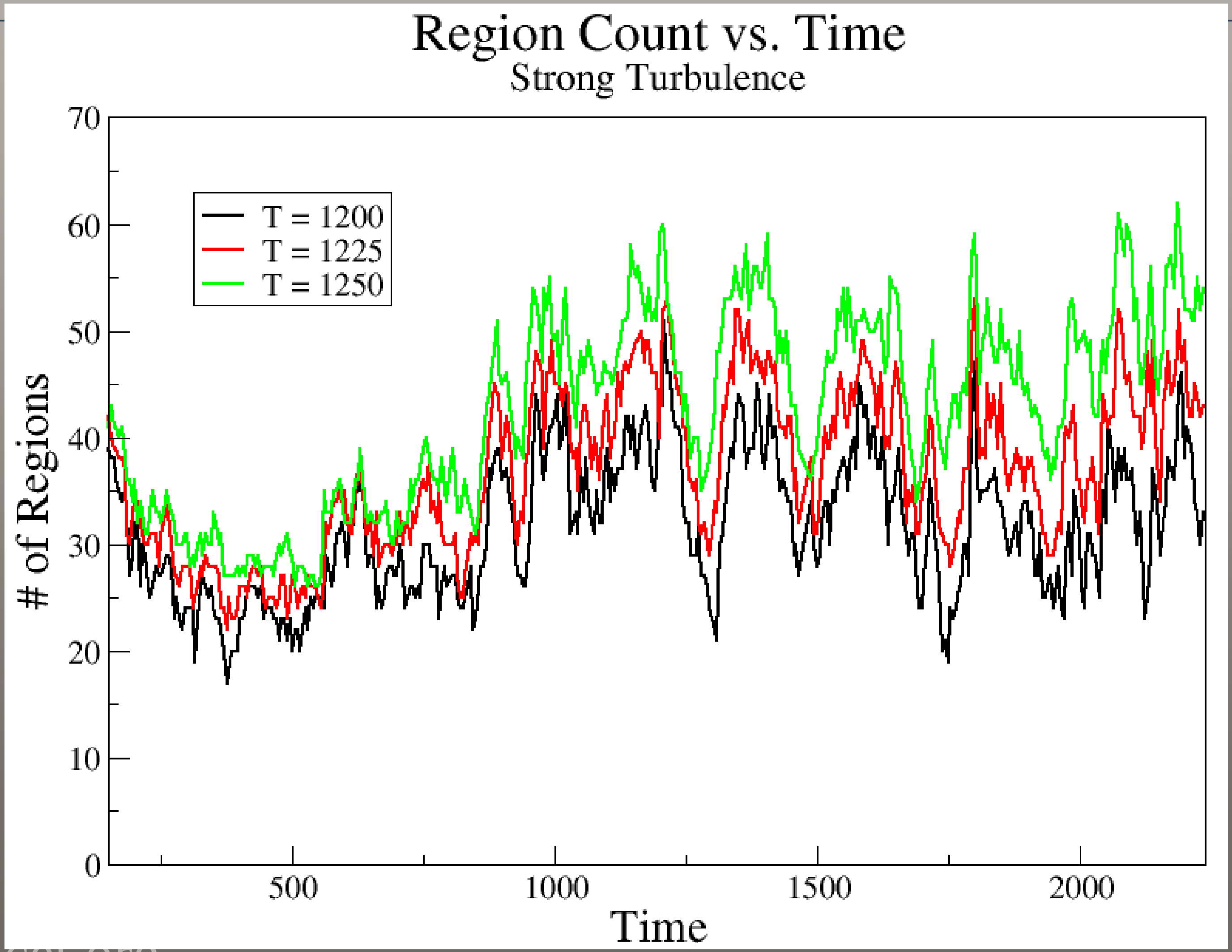
Track Cells by Computing the Reeb Graph of Time as Function on Their Space-Time Boundary Surface



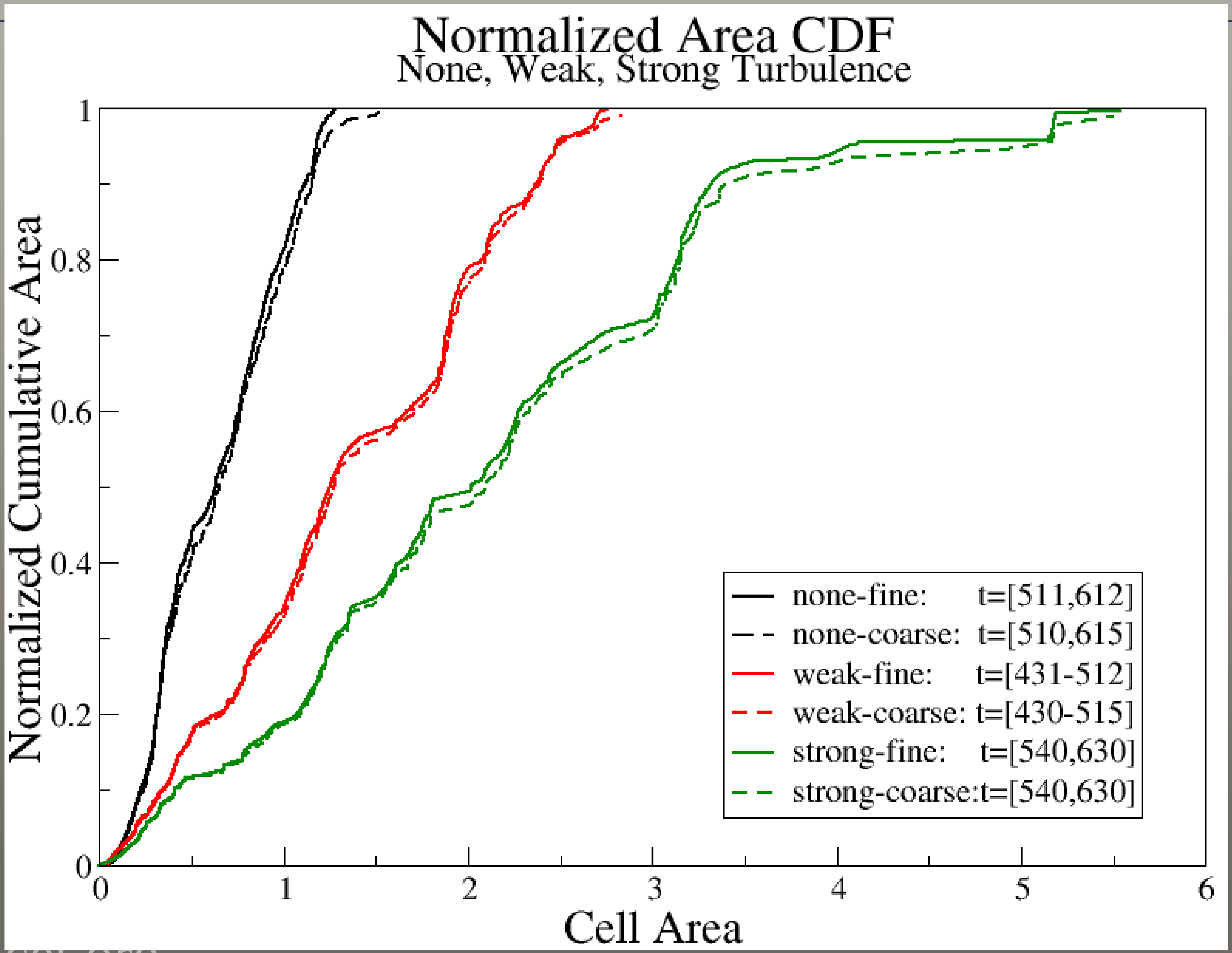
A Detailed Tracking Graph Shows the Evolution of Each Burning Region



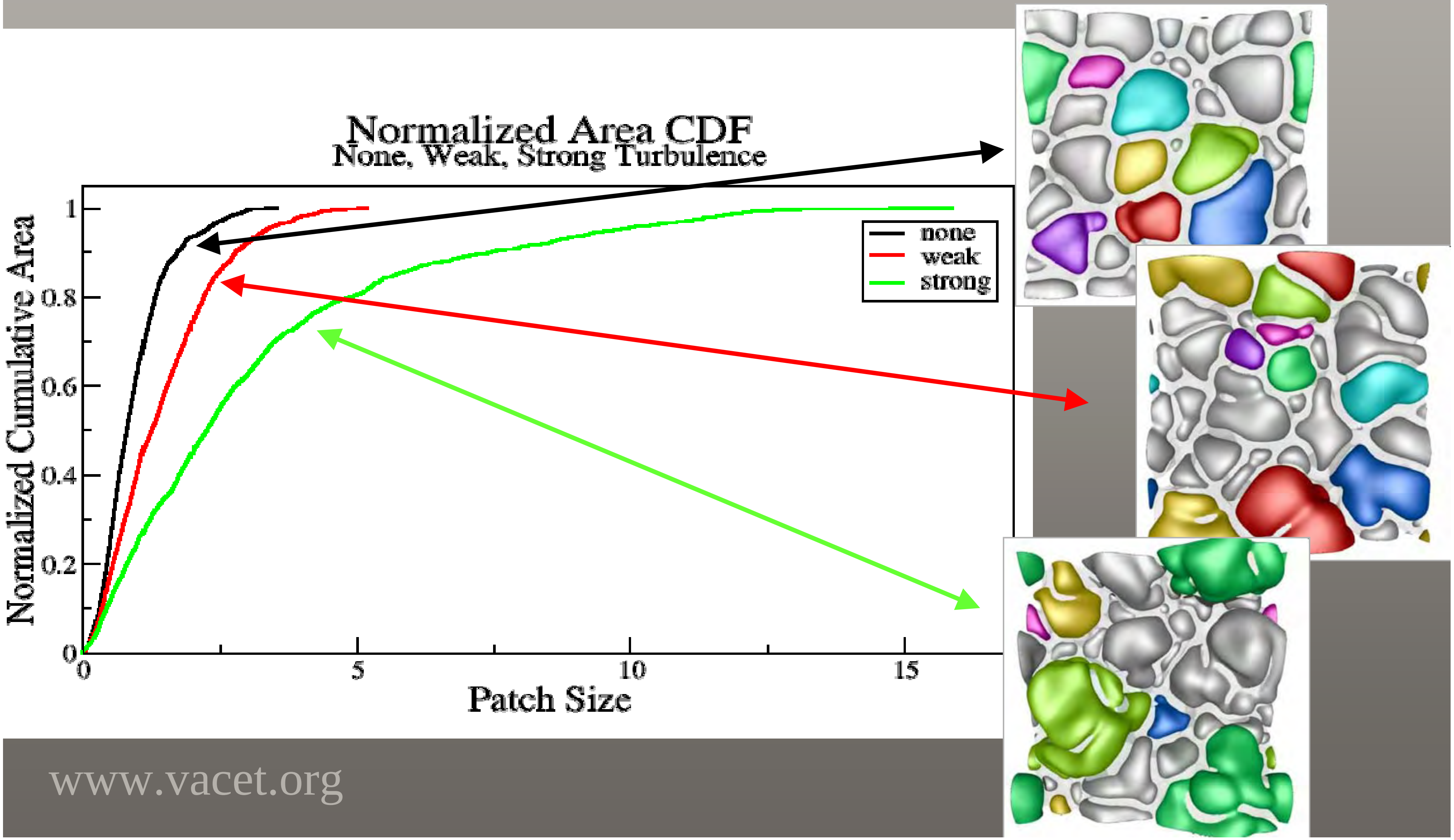
Parameter Studies: Flame Temperature



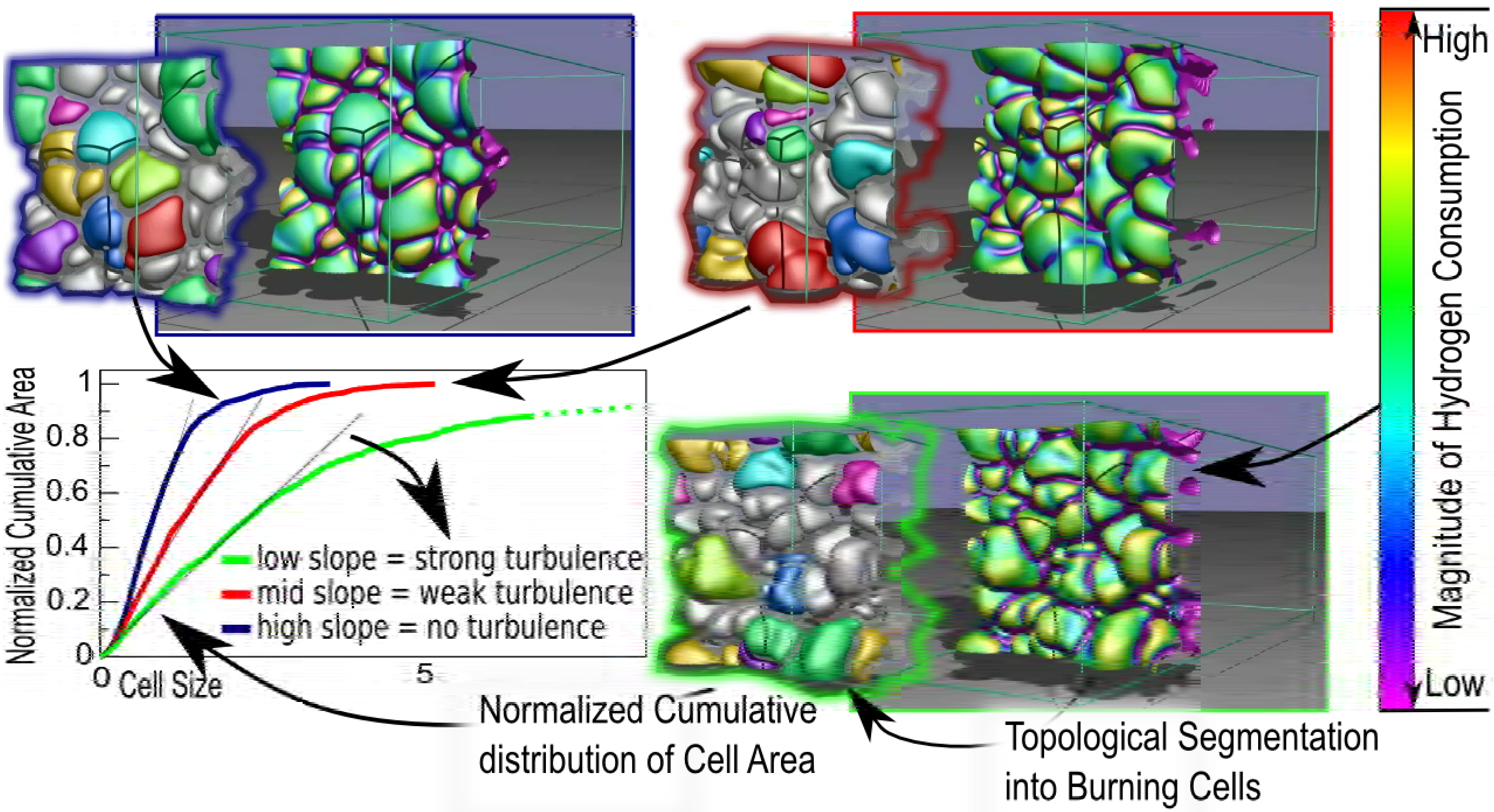
Convergence Study



Stronger Turbulence Leads to Larger Patches Burning More Intensely Than Expected



Topological Segmentation Allows Quantifying Turbulence From the Slope of Normalized Cumulative Distribution of Burning Cell Area



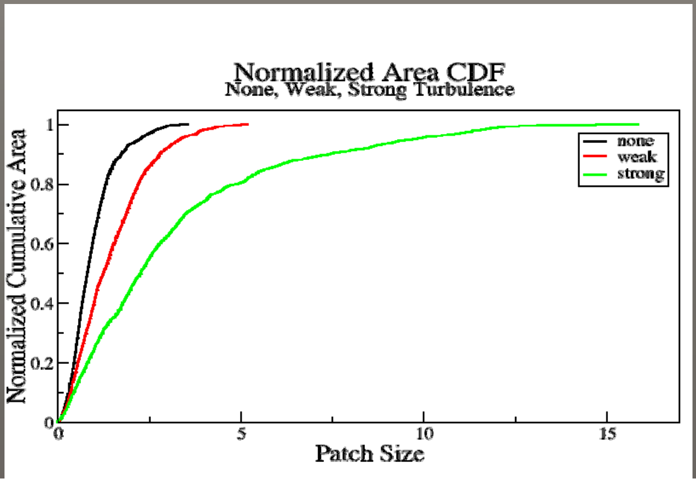
VACET

Recent Publications

- (Science Journal Paper) Turbulence Effects on Cellular Burning Structures in Lean Premixed Hydrogen Flames. J. Bell, M. Day, V. Pascucci, P-T. Bremer, G. Weber. In *Combustion and Flame.* (Accepted, to appear).
  - Note: *Combustion and Flame* is the top journal in the field of combustion (impact factor 1.4).
- (Book Chapter)Scientific Data Management Challenges in High Performance Visual Data Analysis. W. Bethel, H. Childs, V. Pascucci, Prabhat, A. Mascarhenas. In *Scientific Data Management: Challenges, Existing Technology, and Deployment* (to appear).

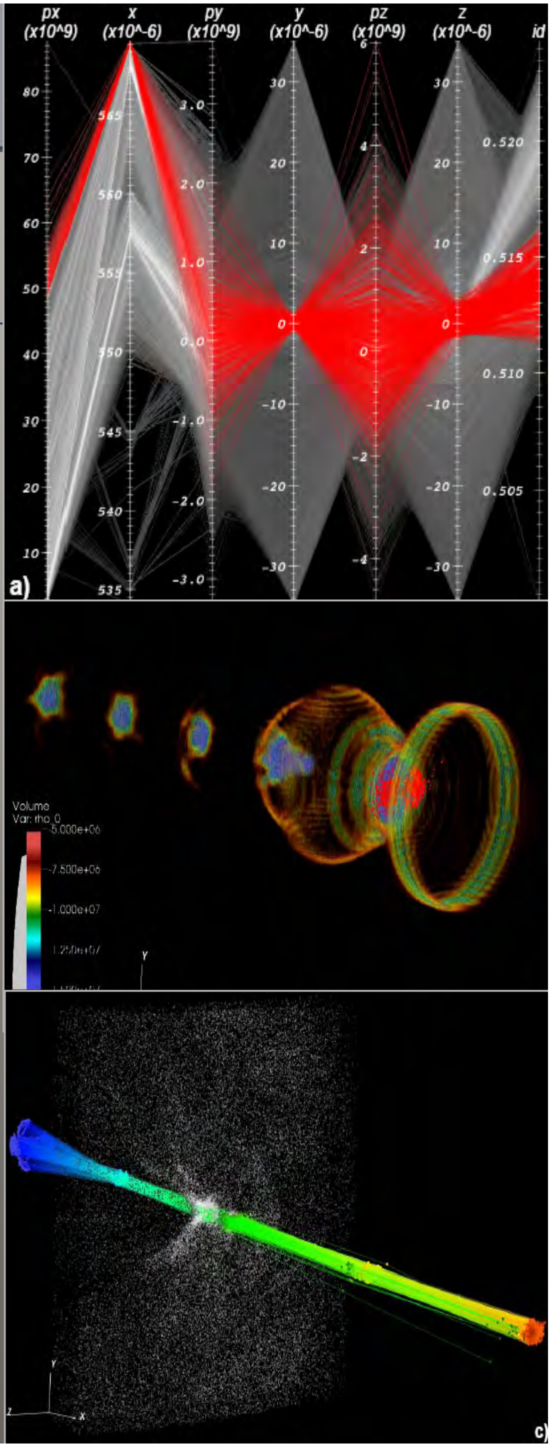
Dispelling Myths

- You don't need sophisticated tools to do a simple x/y chart.
- To go from simulation data to this x/y chart:
  - 10s of K of CPU hours performing feature detection, tracking, and analysis.
  - Many person-month's of effort conceiving, implementing algorithms, running algorithms on simulation data.

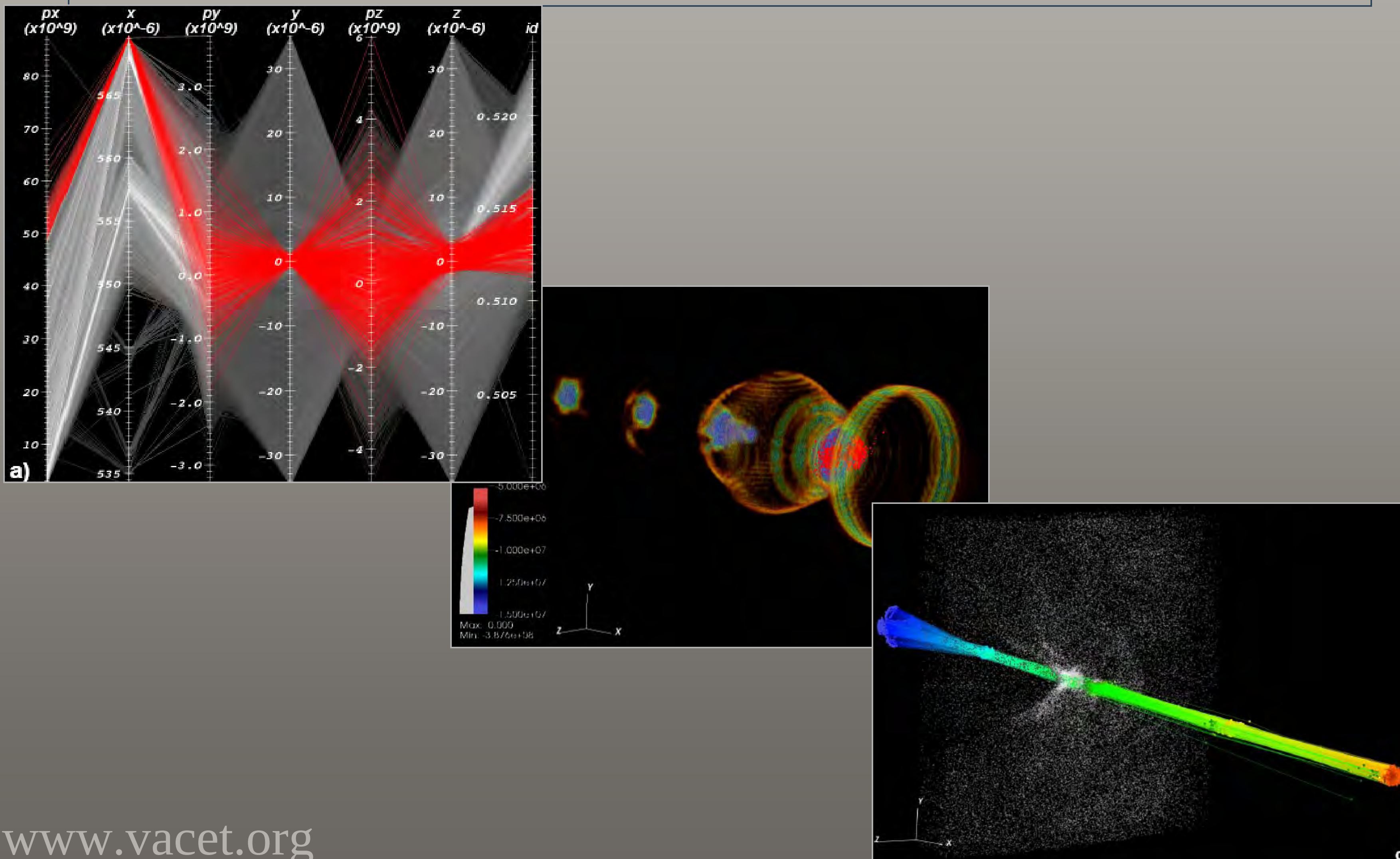


Accelerator

- PI: C. Geddes (LBNL), part of SciDAC COMPASS project, Incite awardee.
- Accomplishment:
  - Algorithms and production-quality s/w infrastructure to perform interactive visual data analysis (identify, track, analyze beam particles) in multi-TB simulation data.
- Science Impact:
  - Replace serial process that took hours with one that takes seconds.
  - New capability: rapid data exploration and analysis.
- Collaborators:
  - SciDAC SDM Center (FastBit)
  - Tech-X (Accelerator scientists)



Case Study – Accelerator Modeling



Laser Wakefield Particle Acceleration

Advantages:

• Can achieve electric fields thousands of times stronger than in conventional accelerators ➔ Can achieve high acceleration in very short distance.

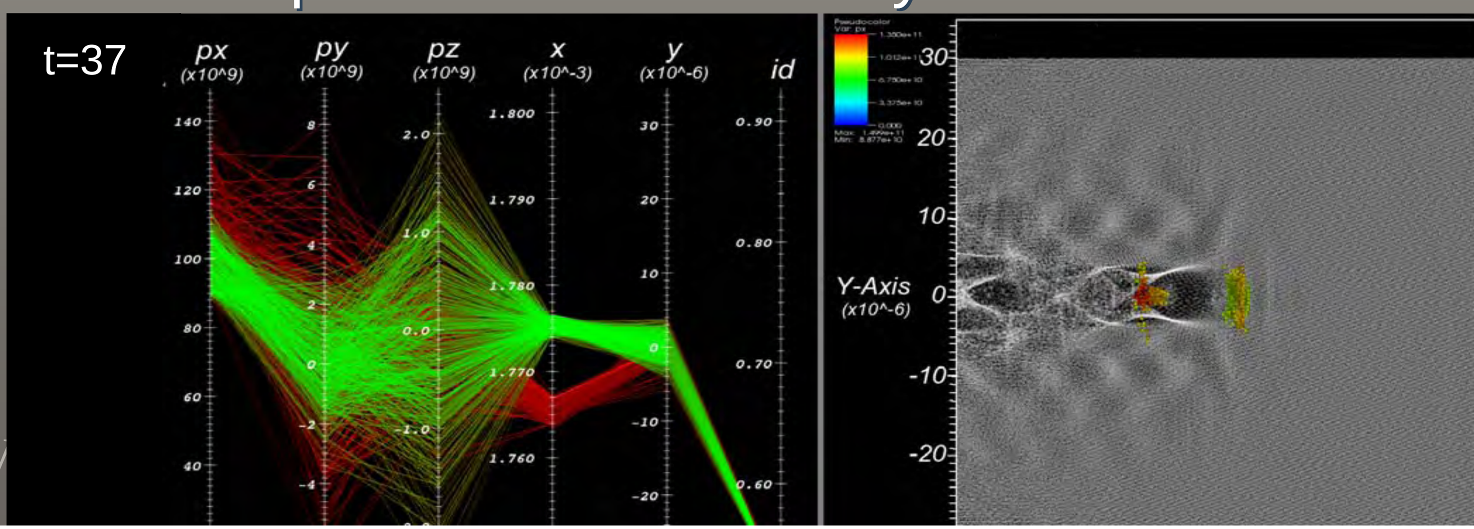
References:
• C.G.R. Geddes, C. Toth, J. van Tilborg, E. Esarey, C. Schroeder, D. Bruhwiler, C. Nieter, J. Cary, and W. Leemans, "High-Quality Electron Beams from a Laser Wakefield Accelerator using Plasma-Channel Guiding," *Nature,* vol. 438, pp. 538-541, 2004

Data Overview

- Simulation: VORPAL, 2D and 3D.
- Particle data:
  - X,y,z (location), px,py,pz (momentum), id.
  - No. of particles per timestep: ~ $0.4*10^6$ – $30*10^6$ (in 2D) and ~$80*10^6$– 200 $*10^6$ (in 3D)
  - Total size: ~1.5GB – >30GB (in 2D) and ~100GB – >1TB (in 3D)
- Field data:
  - Electric, magnetic fields, RhoJ
  - Resolution: Typically ~0.02-0.03µm longitudinally, and ~ 0.1-0.2µm transversely
  - Total size: ~3.5GB - >70GB (in 2D) and ~200GB - >2TB (in 3D)

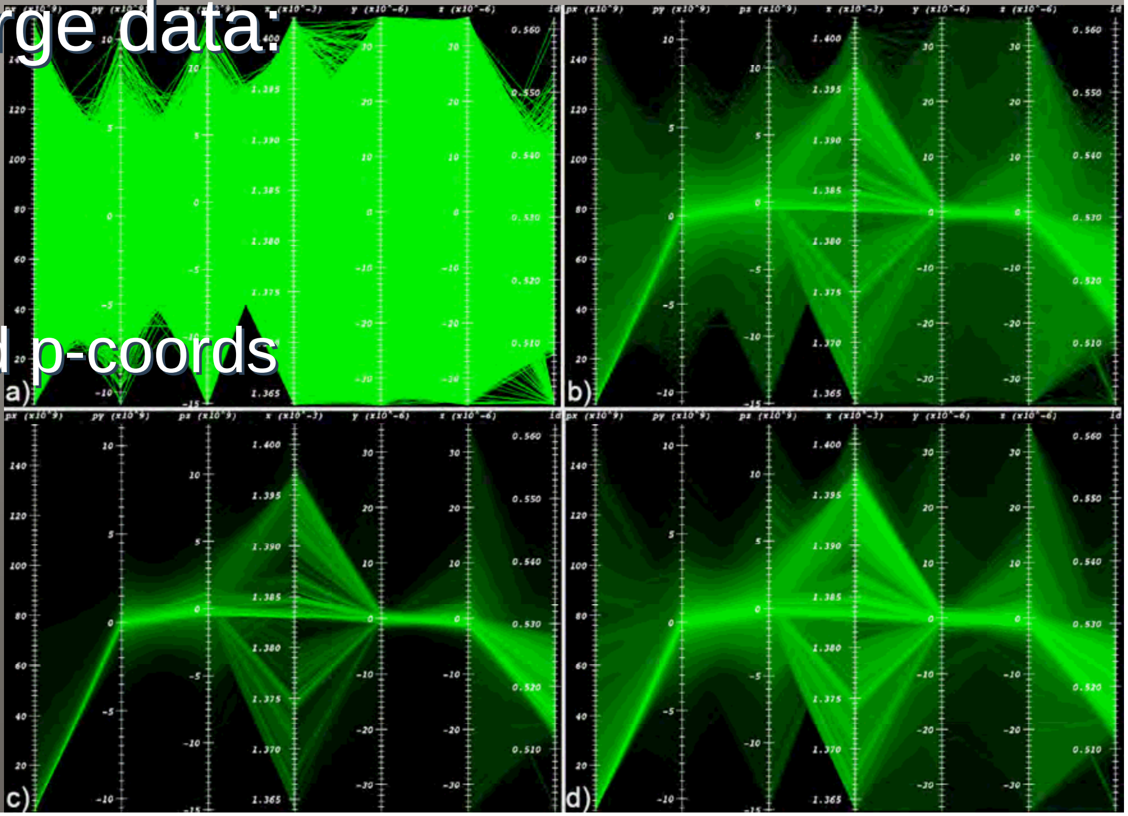
Analysis Task(s)

- Identify particles that form a beam
  - Interactive visual data exploration
  - Data subsetting
- Track them over time
  - Given particle ID's from a given time step,
  - Find all those particles in all time steps
  - Subsequent visual data analysis.

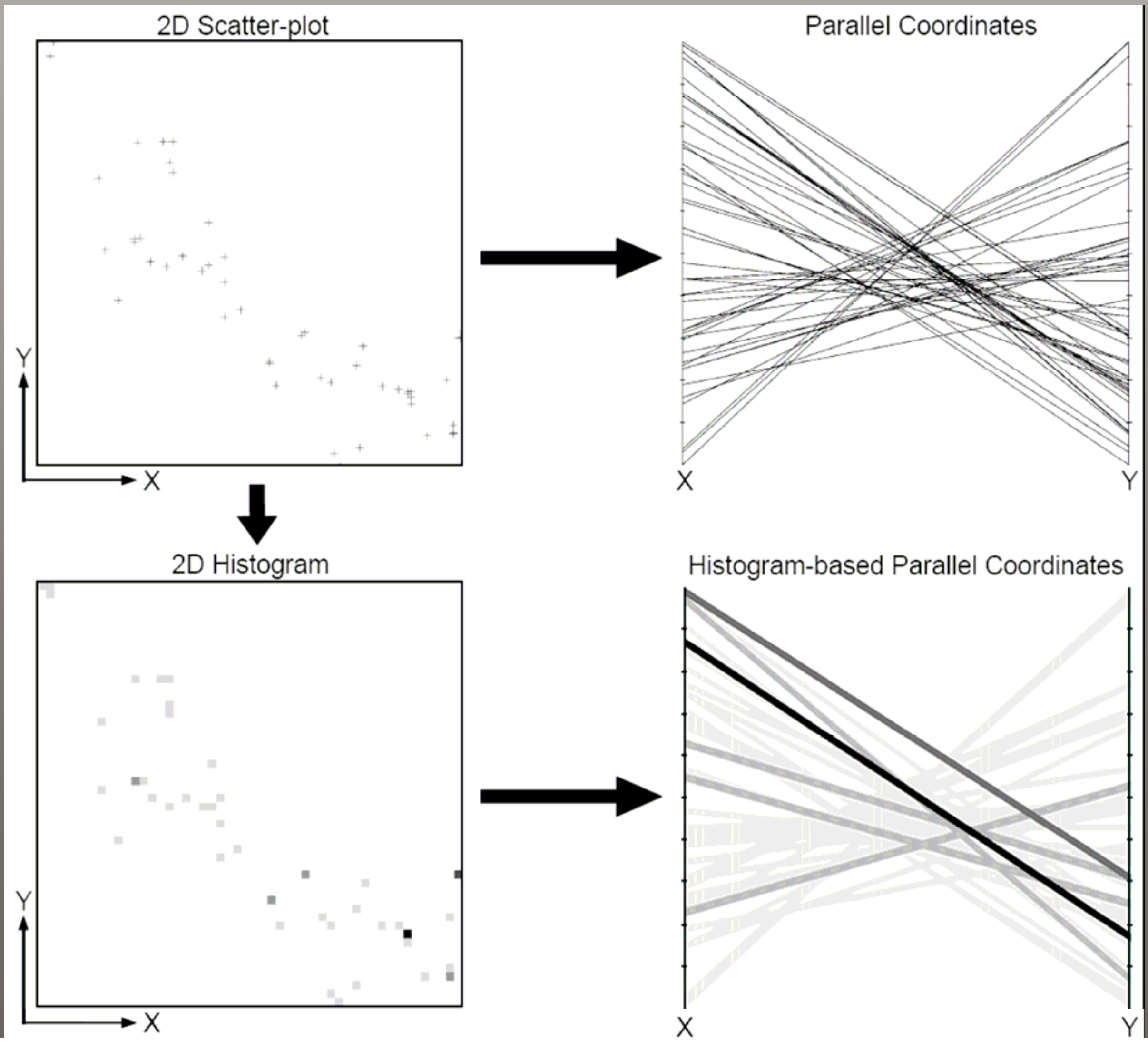


Fundamental Problem #1 - Interface

- Parallel coordinates
  - An interface for subset selection.
  - A mechanism for displaying multivariate data.
- Problems with large data:
  - Visual clutter
  - O(n) complexity
- Solution
  - Histogram-based p-coords



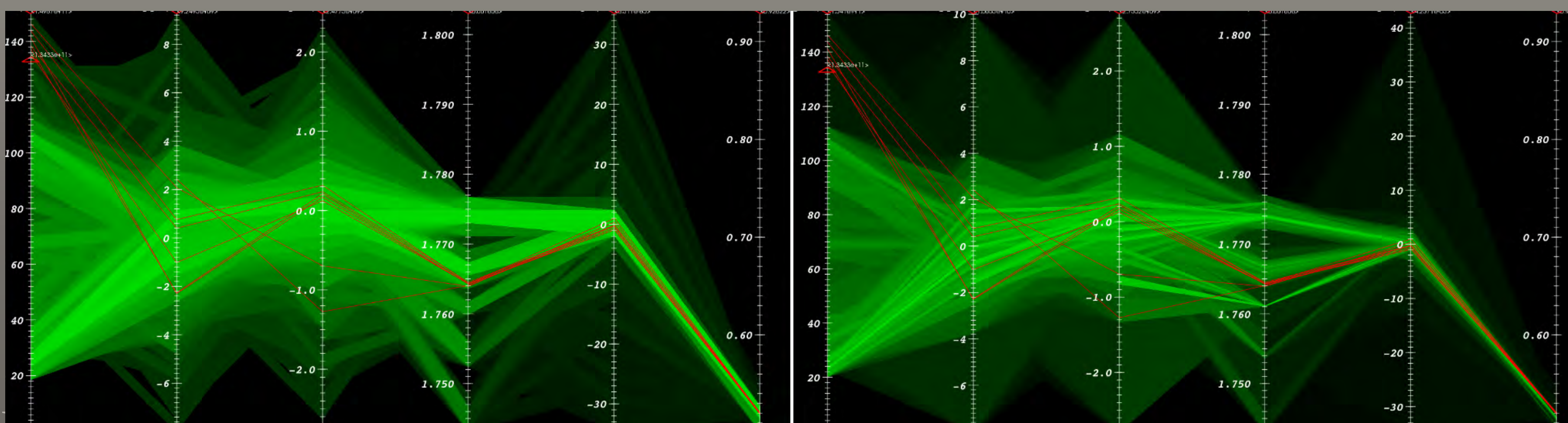
Histogram-Based Parallel Coordinates



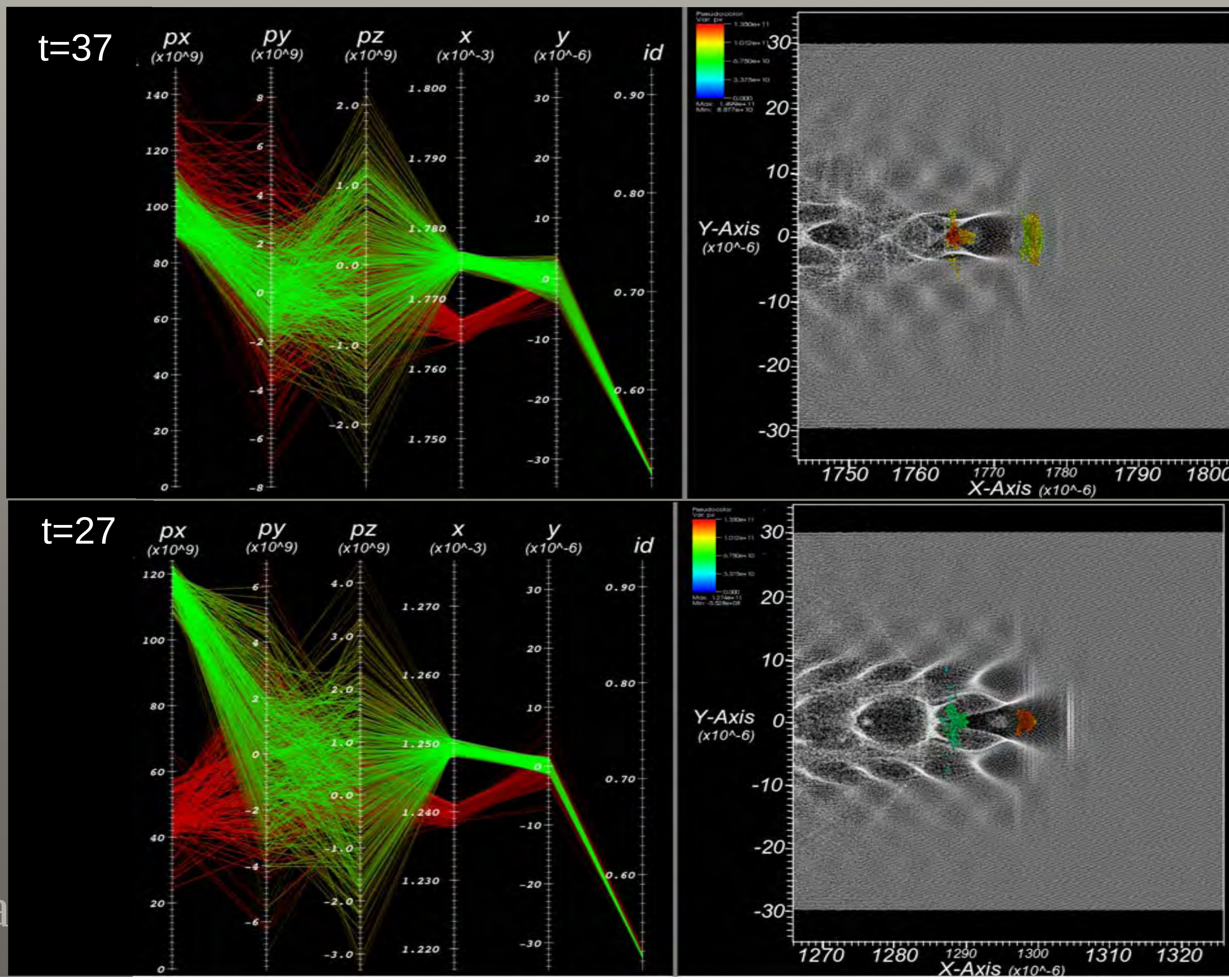
Histogram-based Parallel Coordinates

Histograms are computed on request:
- Enable rendering also of data subsets using histogram-based parallel coordinates
- Enable close zoom-ins and smooth drill-downs into the data
- Enable rendering with arbitrary number of bins

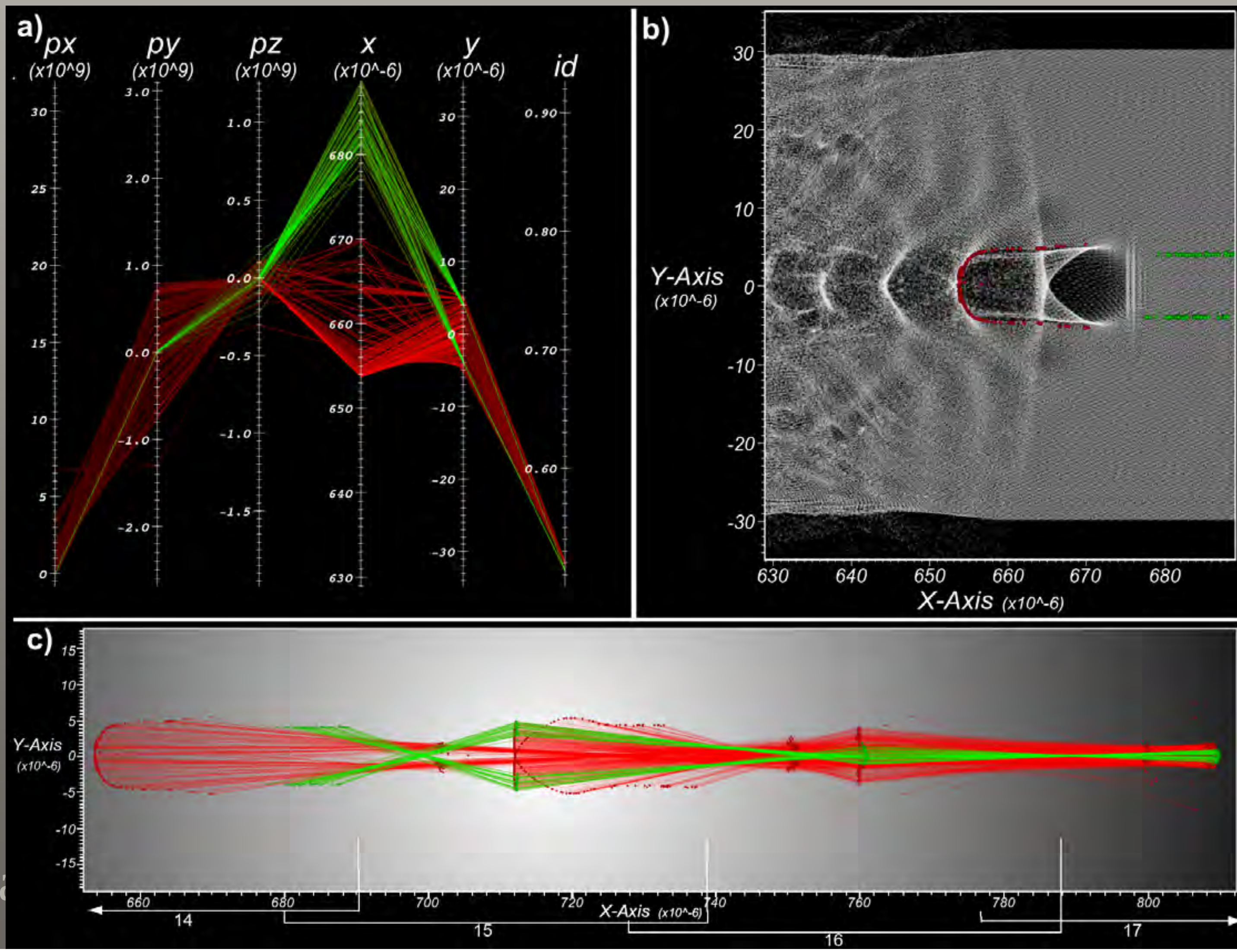
Allow use of adaptively binned histograms:
- Enable more accurate representation of the data in lower-level-of-.detail views



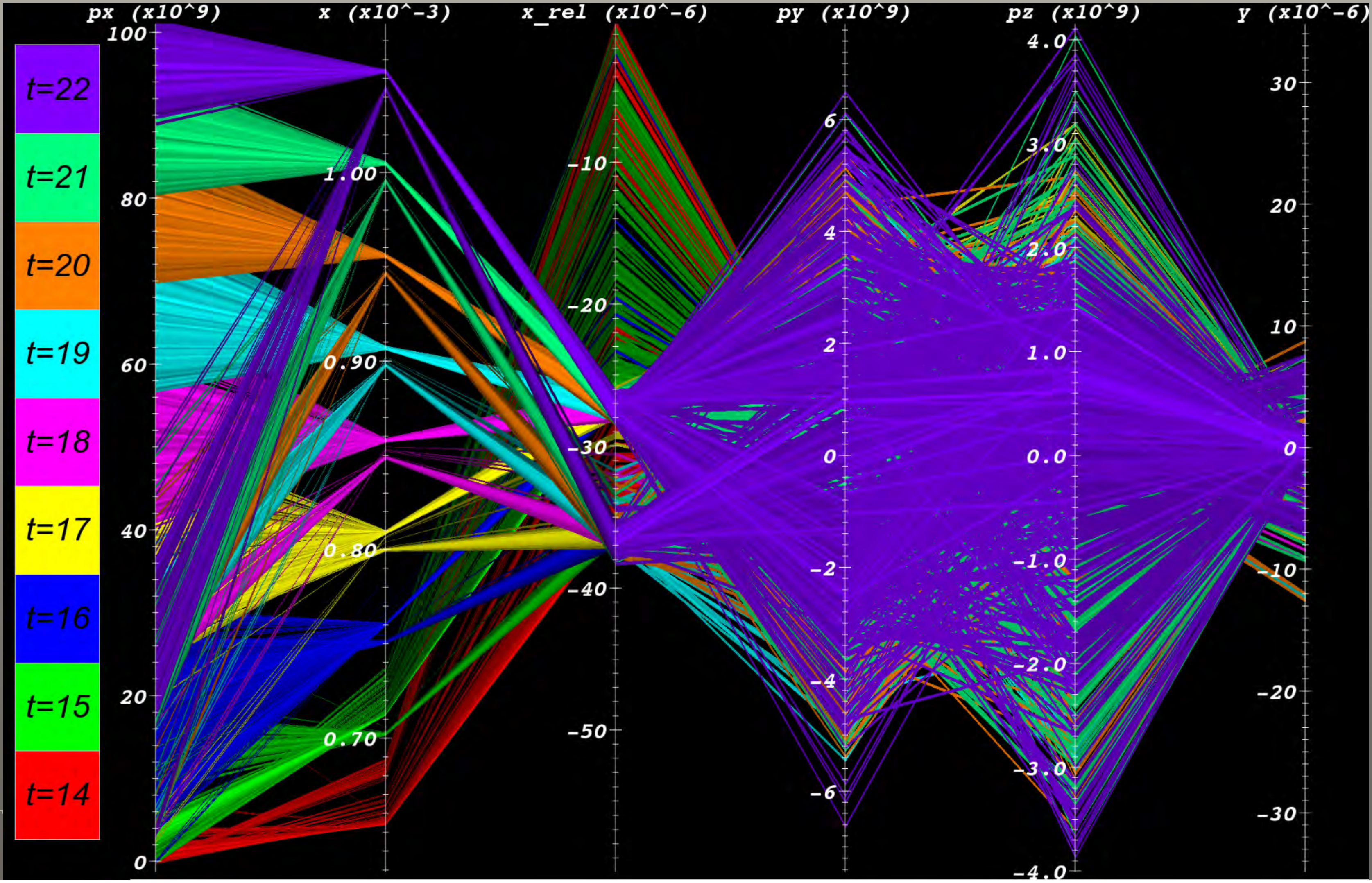
Beam Selection



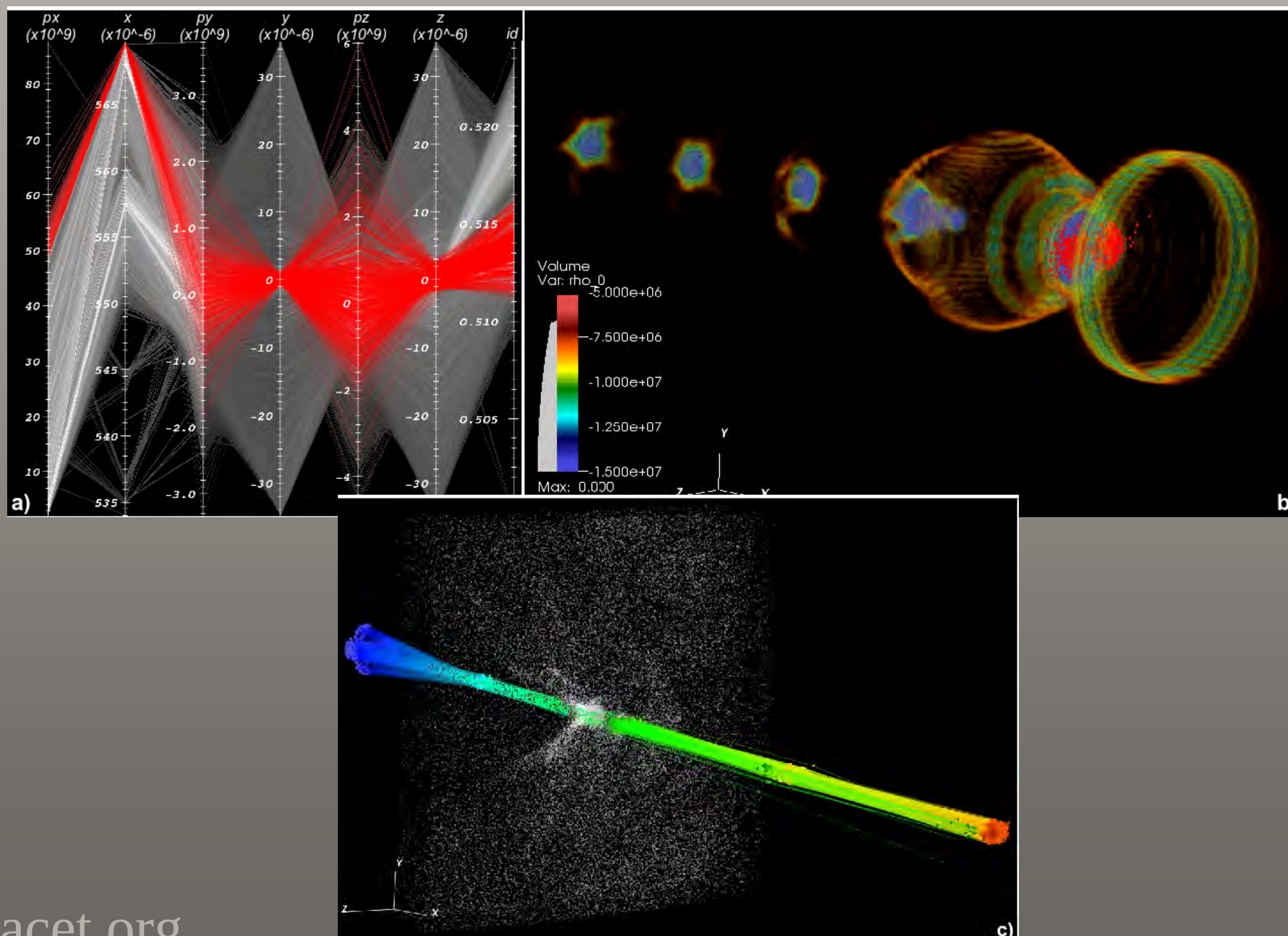
Beam Refinement



Beam Evolution



3D Example


px (x10^9)
x (x10^-6)
py (x10^9)
y (x10^-6)
pz (x10^9)
z (x10^-6)
id
a)
Volume
Var: rho_0
-5.000e+06
-7.500e+06
-1.000e+07
-1.250e+07
-1.500e+07
Max: 0.000
b)
c)

Fundamental Problem #2 – Performance

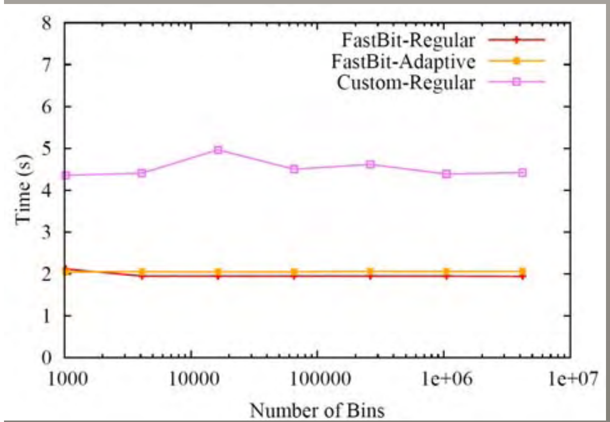
- How to efficiently construct a histogram?
  - Naïve approach: O(n)
  - Better approach: “cheat” (use FastBit)
- How to efficiently do particle tracking?
  - Naïve approach: $O(n^2)$
  - Better approach: O(H*t) (use FastBit)

Serial Performance

Dataset:
- 3D dataset consisting of 30 timesteps
- ~90 million particles per timestep
- ~7GBper timestep (including ~2GB for the index)
- ~210GB total size

Test platform:
- Workstation
- CPU: 2.2GHz AMD Opteron
- Memory: 4GB RAM
- OS: SuSE Linux

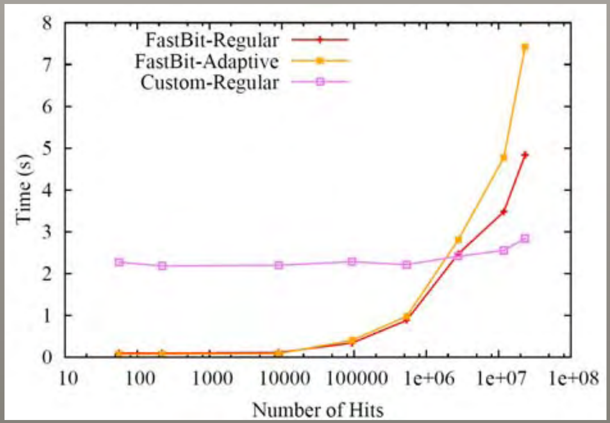
Unconditional Histograms:



Setup:
- Test performance with increasing bin counts: 32x32 to 2048x2048

Custom:
- Perform sequential scan

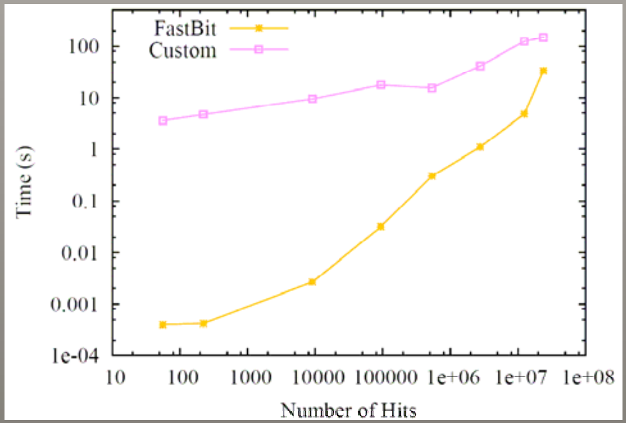
Conditional Histograms:



Setup:
- Compute 1024x1024 histogram with varying condition (px>...)
- By increasing the threshold the number of hits decreases

Custom:
- Perform sequential scan

Particle Selection:



Setup:
- Perform ID query at a single timestep and vary the size of the search set S

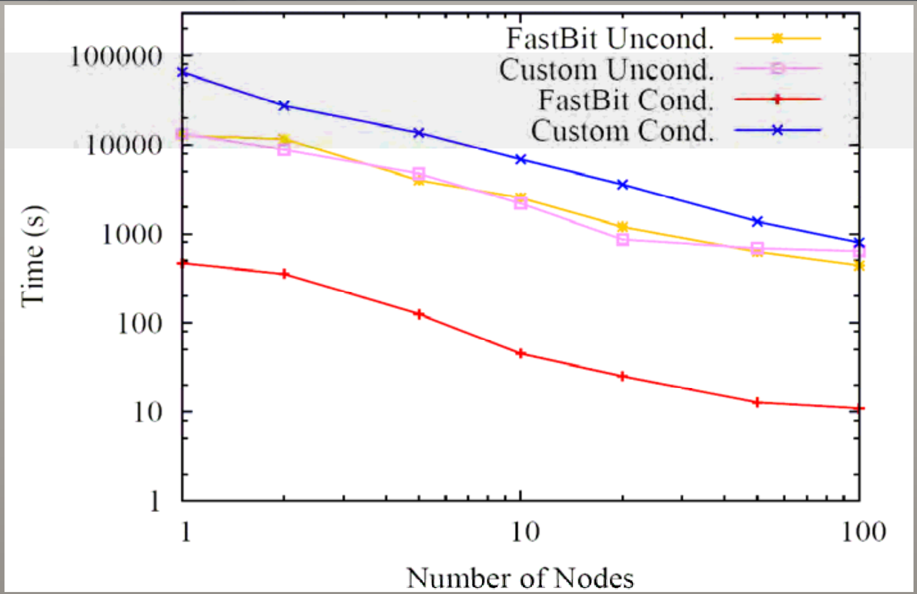
Custom:
- Compare particle ID of each data record to the search set
- Use efficient search algorithm with O(log(S)) complexity

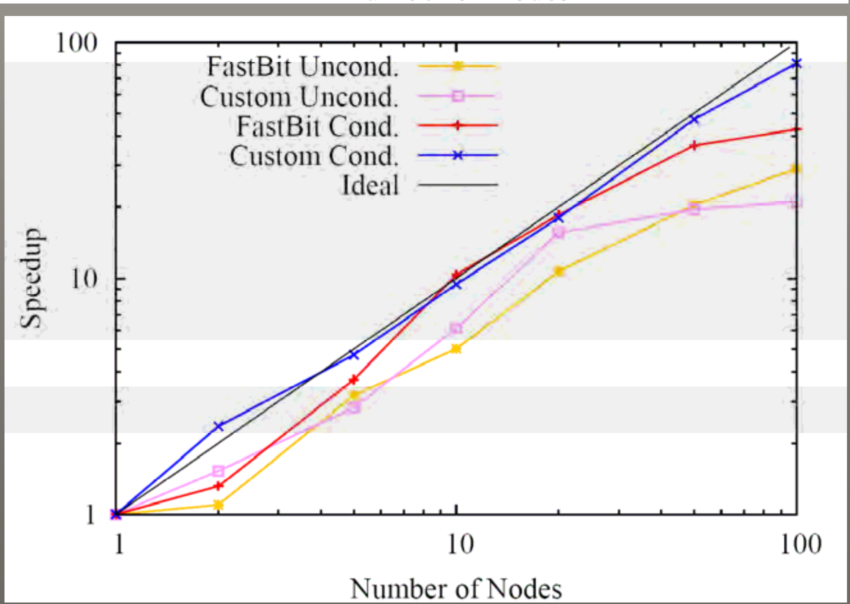
Parallel Performance I: Histograms

Dataset:
- 3D dataset consisting of 100 timesteps
- ~177 million particles per timestep
- ~10 GB per timestep
- ~1TB total size

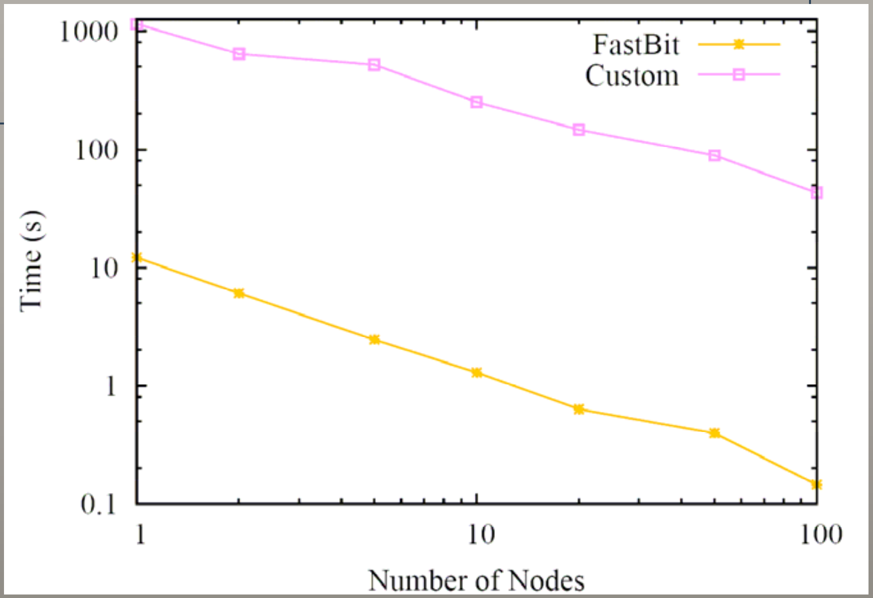
Test platform: (as of July.2008)
- franklin.nersc.gov
- 9,660 nodes, 19K cores Cray XT4 system
- Filesystem: Lustre Parallel Filesystem
- Each node consists of:
    - CPU: 2.6 GHz, dual-core AMD Opteron
    - Memory: 4GB
    - OS: Compute Node Linux

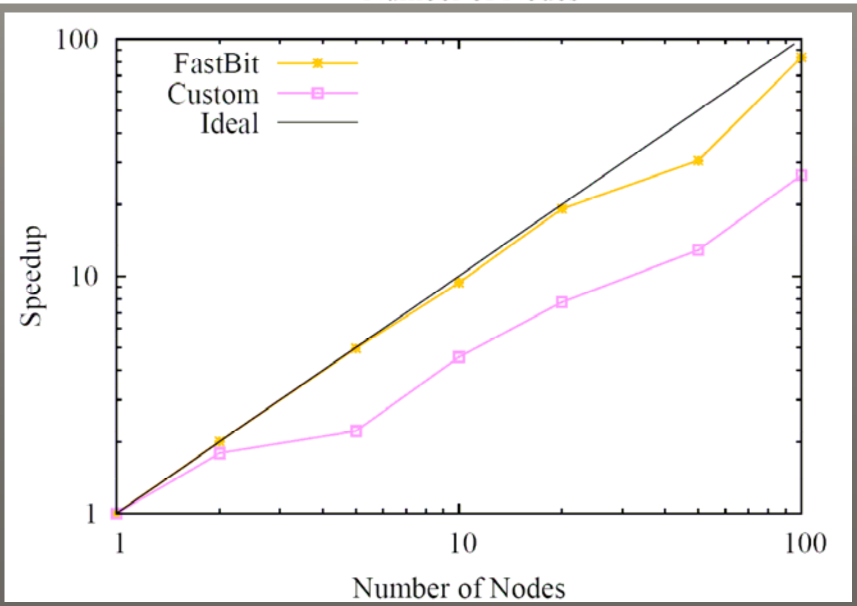
Test setup:
- Restrict operations to a single core of each node to maximize I/O bandwidth available to each process
- Assign data subsets corresponding to individual timesteps to individual nodes for processing
- Generate five 1024x1024 histograms for position and momentum fields at each timestep
- Conditon: $px>7*10^{10}$
- Levels of parallelism: 1, 2, 5, 10, 20, 50, 100





Parallel Performance II: Particle Tracking





Test setup:
- Same as for histogram computation
- Track 500 particles (Condition: $px>10^{11}$) over 100 timesteps

Results:
- FastBit is able to track 500 particles over 1.5TB of data in 0.15 seconds

Performance of original IDL scripts:
- ~2.5 hours to track 250 particles in small 5GB dataset

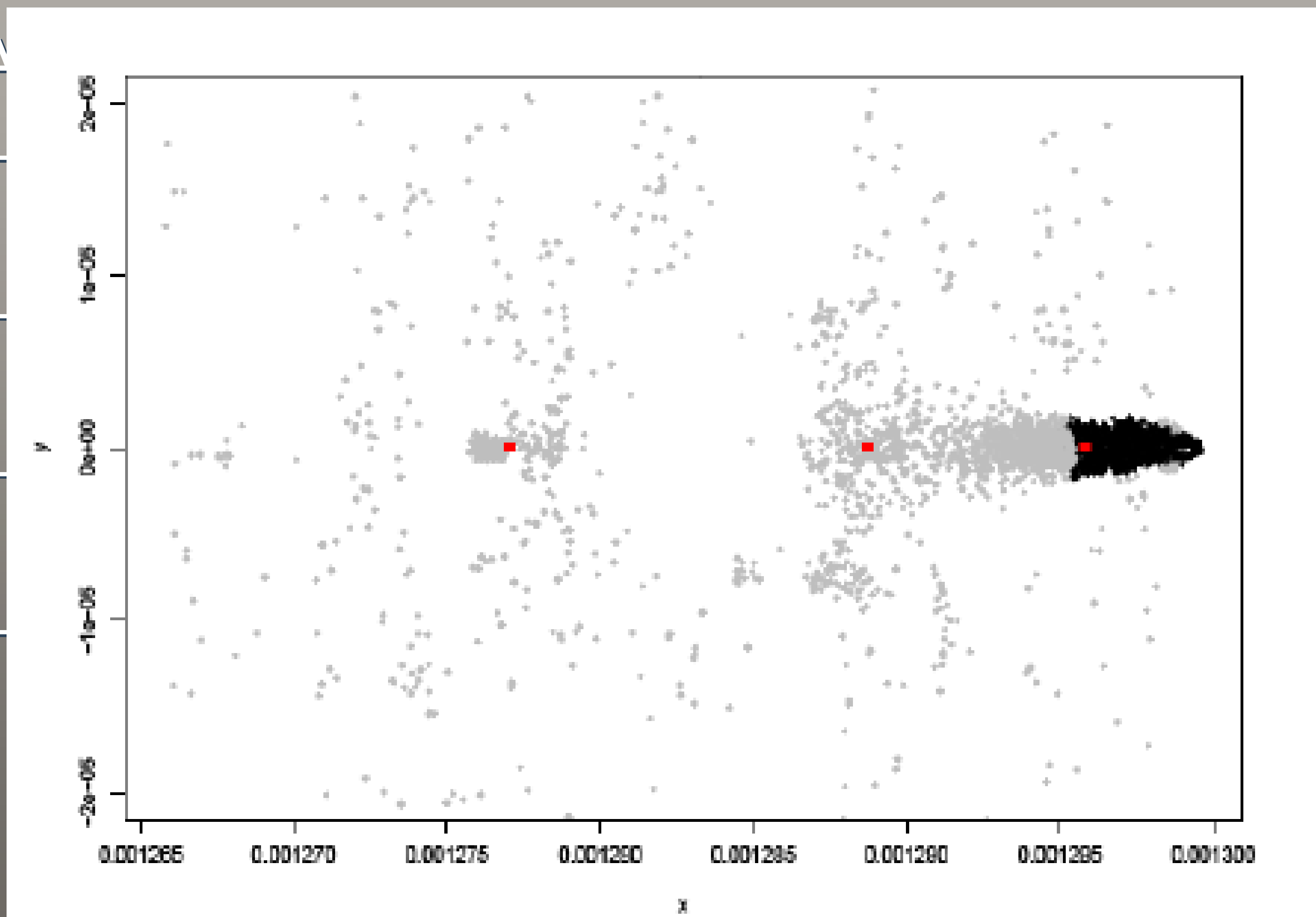
"Traditional Analysis" Applied to LWFA

- Approach
  - Identify particle "bunches" that have high momentum and spatial coherency.
  - For each bunch, use a graph algorithm to track bunch movement and evolution across timesteps.
  - Separately, use "fuzzy clustering" to compute probability a particle is "beam" or "non-beam".
  - Compare fuzzy clustering and space/momentum classification results. Where high agreement, have beam particles.

“Traditional Analysis” Applied to LWFA

- A

s.



Recent Publications

- SC08 Technical Paper: High-Performance Multivariate Visual Data Exploration for Extremely Large Data. O. Rubel, et al.
- 2008 International Conference on Machine Learning: Automated Analysis for Detecting Beams in Simulations. D. Ushizima, et al.

Observations about These Case Studies

- New science results from multidisciplinary team working on a challenging data understanding problem.
  - Such collaborative efforts require a substantial investment of time – thanks to SciDAC program!
- Work spans:
  - Data I/O, data models, veneer data I/O APIs
    - Encapsulating complexity, scalability.
  - Visualization algorithm architectures
  - Computational topology
  - Scalability, tuning, debugging.

Take Home Message

- VACET mission: deliver production-quality visual data analysis s/w infrastructure.
  - Target: difficult scientific data understanding problems
- VACET as a CET:
  - Delivering the goods.
  - Helping SciDAC as a whole: quantifiably enabling scientific knowledge discovery.
  - Strong science community support.
  - Business model addresses software lifecycle issues AND a healthy science-driven research effort.

The End

- Thanks for your time.
- More information: www.vacet.org