

---

---

# **Sapphire Analysis of DNS of Rayleigh-Taylor Instability**

---

---

***Chandrika Kamath, Abel Gezahegne, and  
Paul L. Miller  
Lawrence Livermore National Laboratory***

***Fall Creek Falls Conference  
September 26, 2007***

UCRL-PRES-234421 This work was performed under the auspices of the U.S.  
Department of Energy by University of California Lawrence Livermore National Laboratory  
under contract no. W-7405-Eng-48.



# Sapphire: using data mining techniques to address the data overload problem

---

---

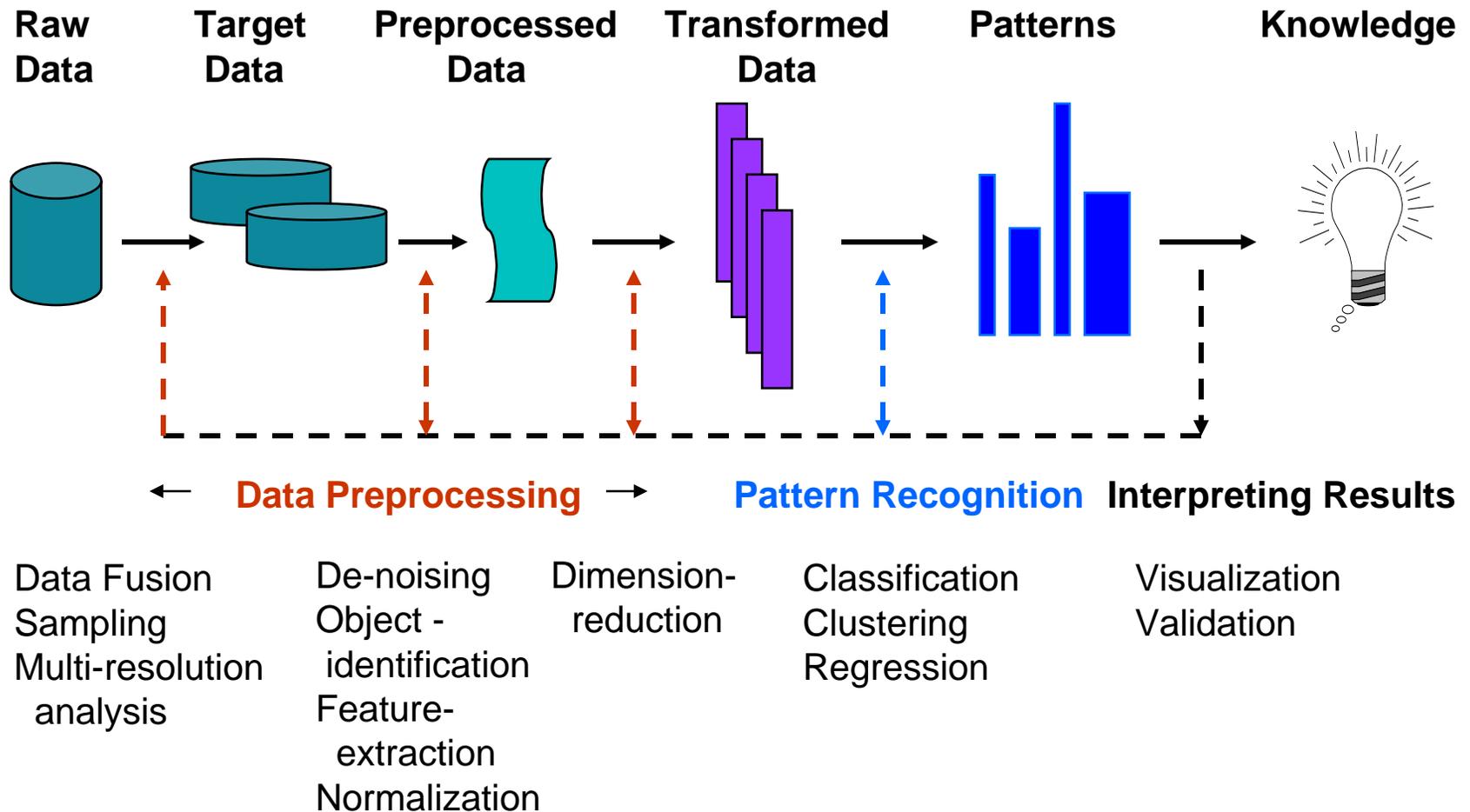
- Analyze science data from experiments, observations, and simulations

- Sapphire has a three-fold focus

- — **analysis** of data
- **software** development
- **research** in algorithms

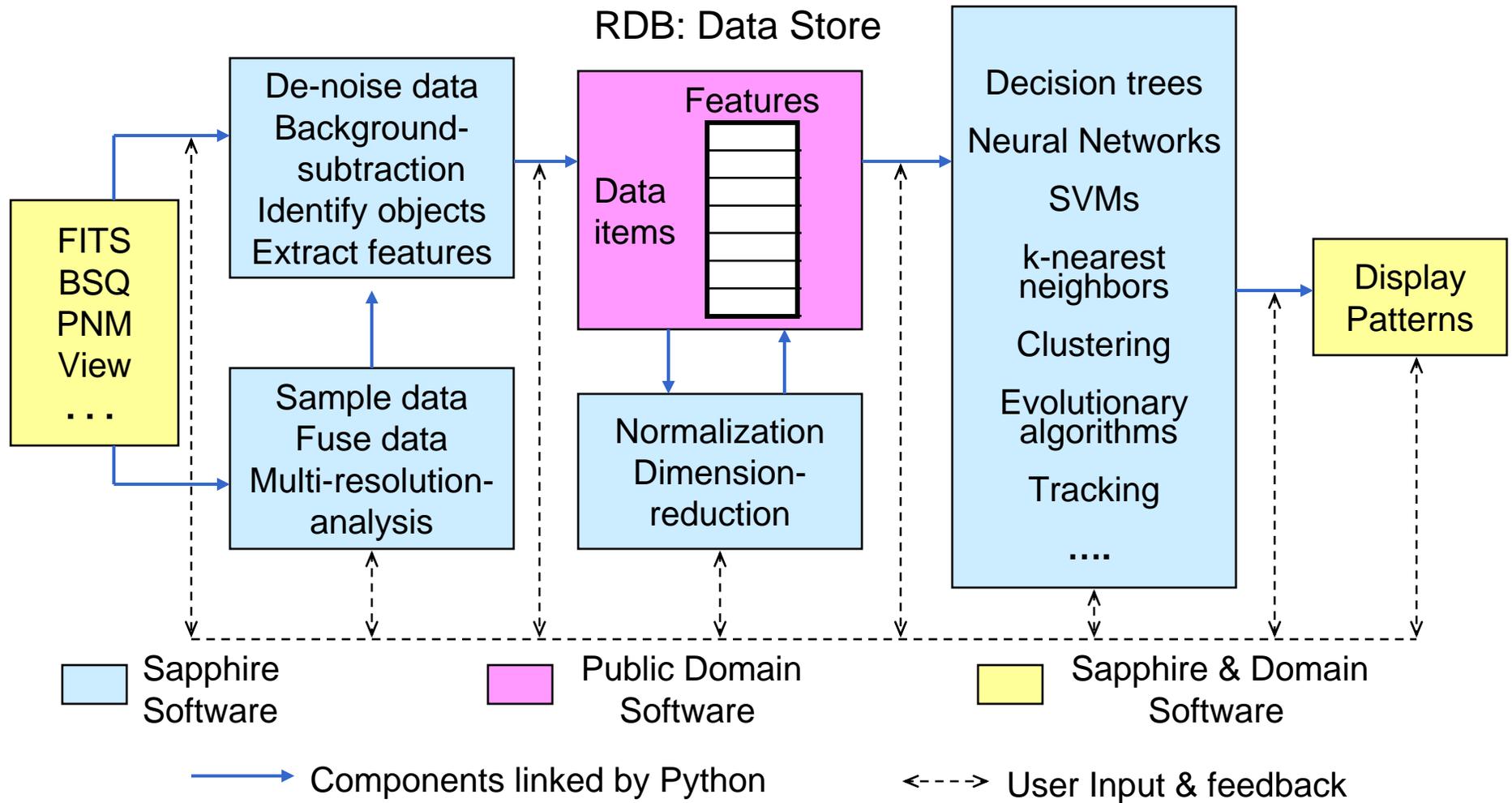
→ Visual data analysis for massive complex data sets is impractical given its subjective nature and human limitations in absorbing detail.

# Scientific data mining - from a Terabyte to a Megabyte



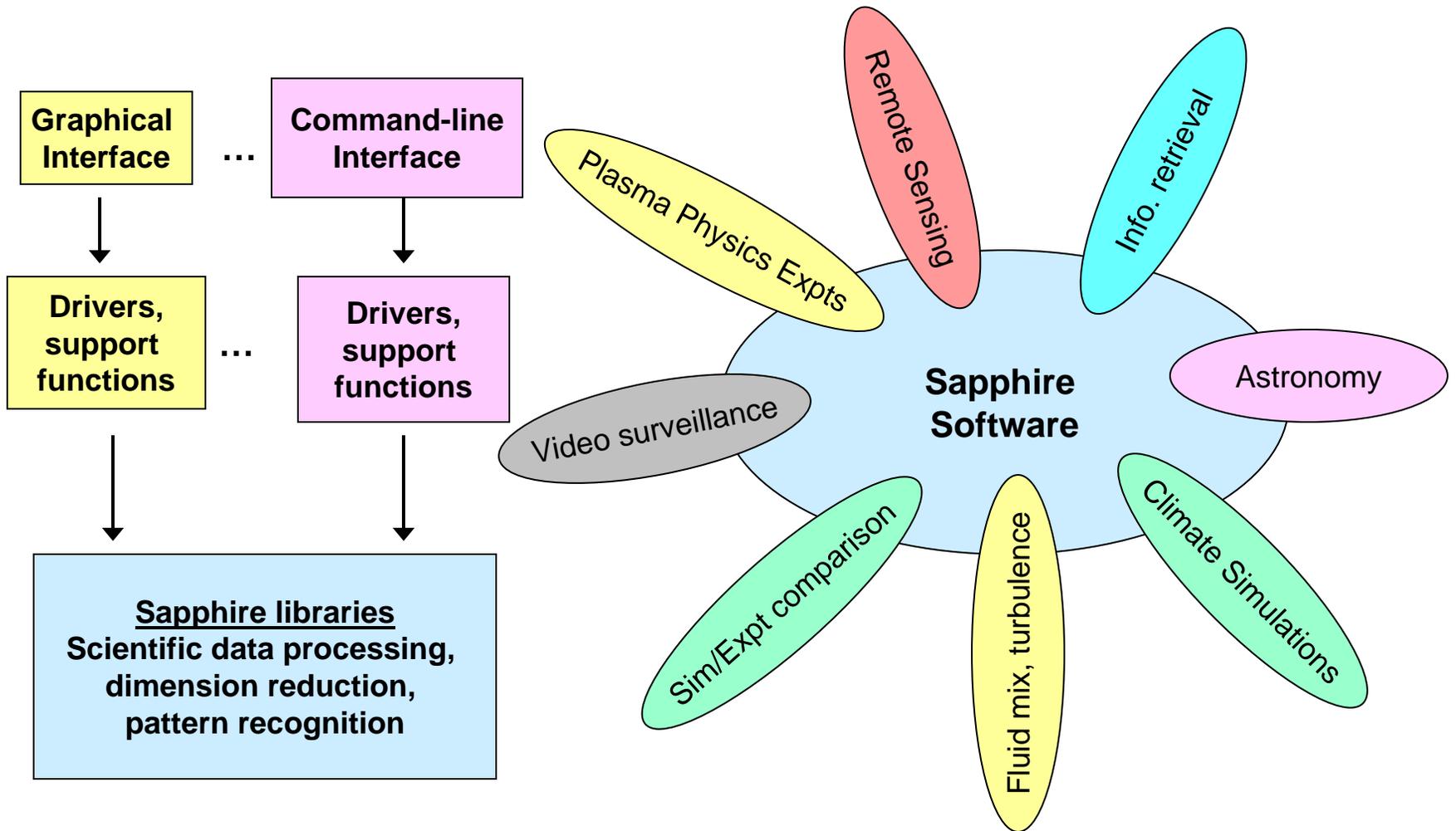
**An iterative and interactive process**

# The Sapphire system architecture: flexible, portable, scalable



US Patents 6675164 (1/04), 6859804 (2/05), 6879729 (4/05), 6938049 (8/05), 7007035 (2/06), 7062504 (6/06)

# The modular software allows us to meet the needs of different applications



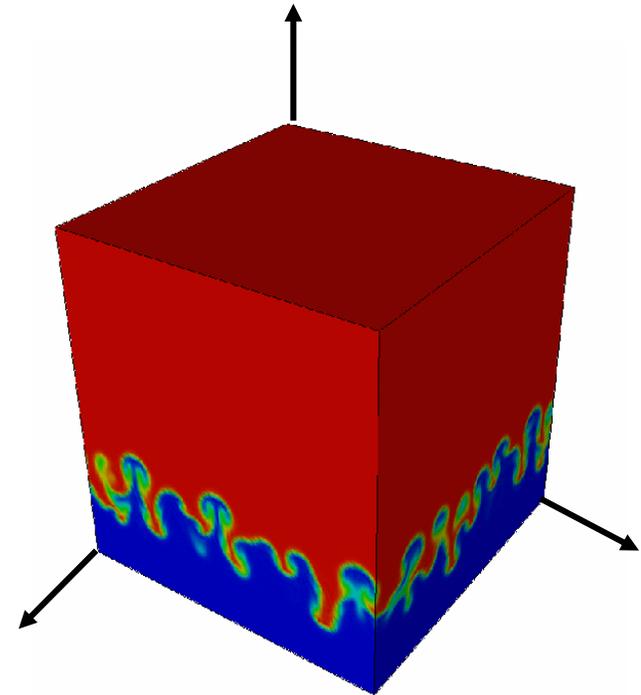


# **Analysis of Bubbles and Spikes in Rayleigh-Taylor Instability**

# Goal: use image analysis to characterize and track bubbles and spikes

---

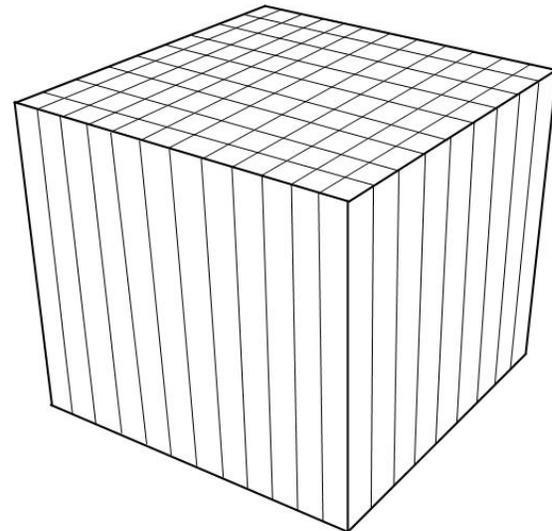
- Two high-fidelity simulations of the Rayleigh-Taylor instability
  - density ratio 3:1
- Goals of the analysis
  - **bubble counts**
  - bubble sizes
  - distances between bubbles
  - bubble dynamics



# The data is obtained from the Miranda code on a 3-D regular Cartesian grid

---

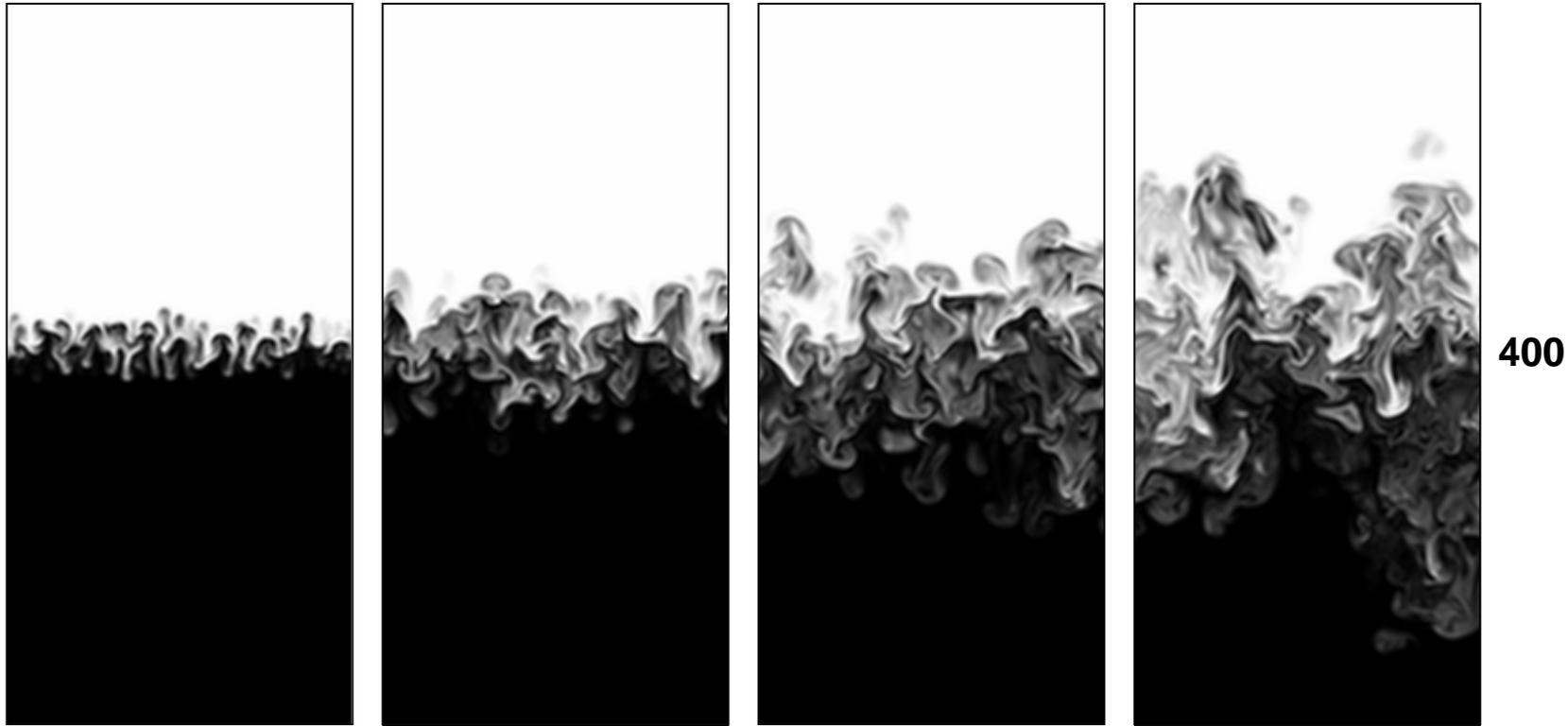
- LES simulation\*
  - 1152\*\*3 grid points
  - 7 variables per grid point
  - 758 time steps
  - **30TB analysis data**
- DNS simulation\*\*
  - 3072\*\*3 grid points
  - 5 variables per grid point
  - 249 time steps
  - **80TB analysis data**



\* Cook, Cabot, and Miller, *Journal of Fluid Mechanics*, 511, 2004.

\*\* Cabot and Cook, *Nature Physics*, 2, 2006.

# The first step is to define a bubble...

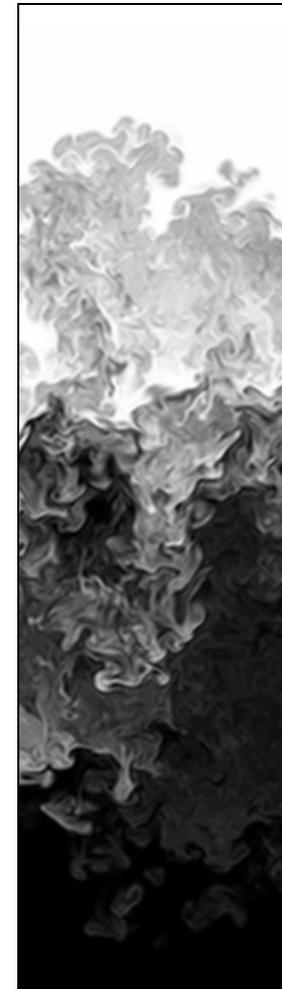
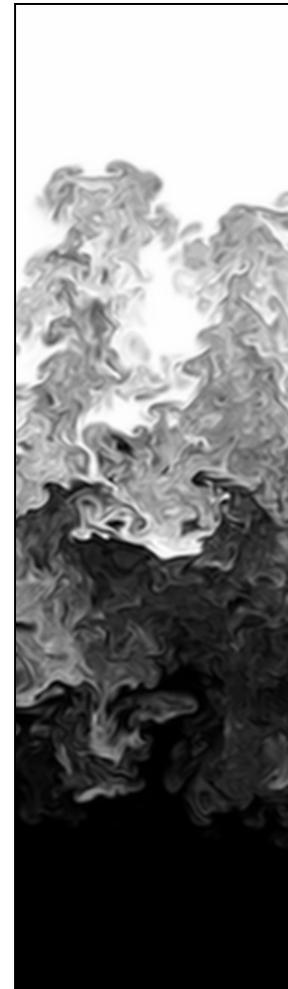
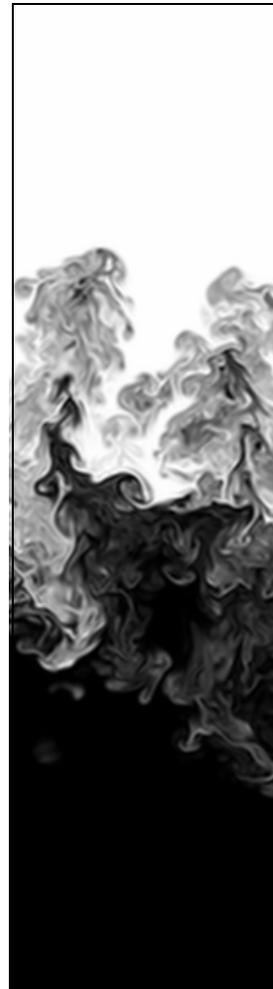


A slice through the density variable: LES data at  
time steps 100, 200, 300, 400

**Convention:** Smaller values are darker in image.

**... which can be a challenge, especially at the later time steps**

**Density variable, LES data  
time steps 500, 600, 700**



**700**

**192**

# There are several challenges to the analysis

---

---

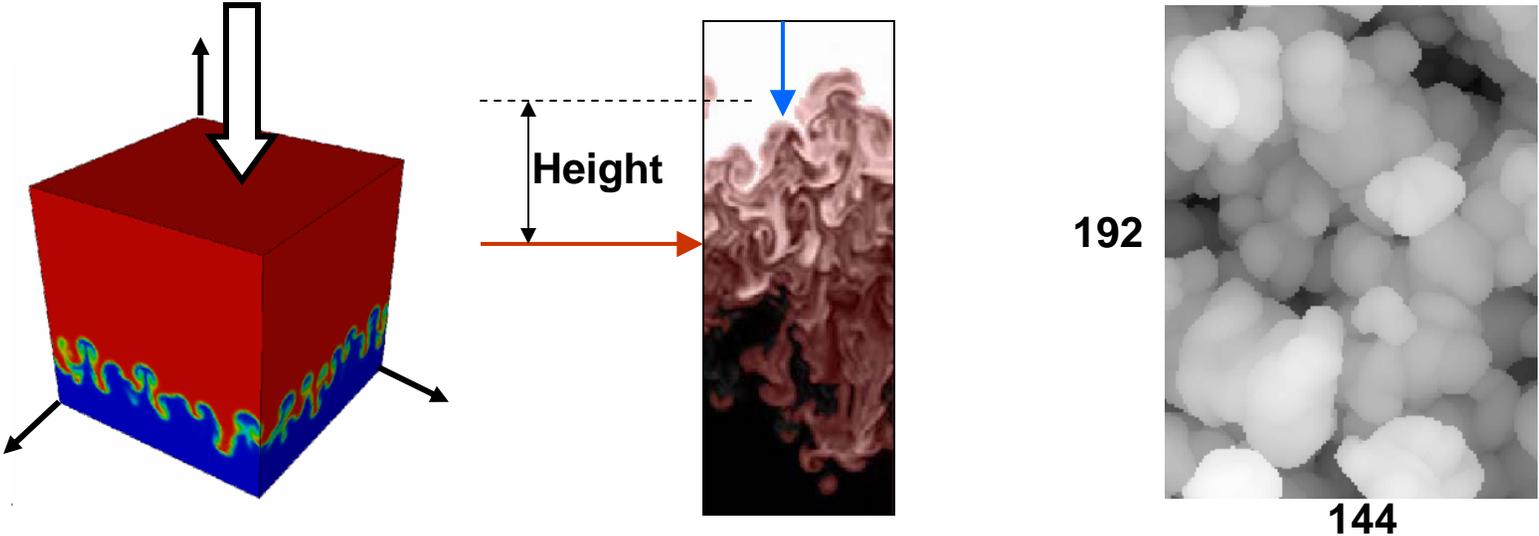
- **Lack of a precise definition of a bubble**
  - range of scales of the structures of interest
- **Massive size of the data**
  - distributed nature of output at each time step
- **Requirements of the analysis algorithms**
  - low computational cost
  - applicable to distributed data
  - few parameters
  - relatively insensitive to choice of parameters
  - a single algorithm and parameters for all time steps
  - multiple algorithms for verification

# Our solution approach: refine bubble definition on a small subset of the data

---

- Start with a small subset (0.04%) of the full data
  - LES data, 6x6 columns (= 144x192x1152)
  - every 50-th time step
- Consider all variables
- Attempt a definition of a bubble
- Iterate to determine parameter settings
  - evaluate on 6x6 columns, every 50-th time step
  - evaluate on full data, every 50-th time step
- Apply to full data – **only once**

# We use the bubble height to generate a 2-D image: the height-depth map (HDM)

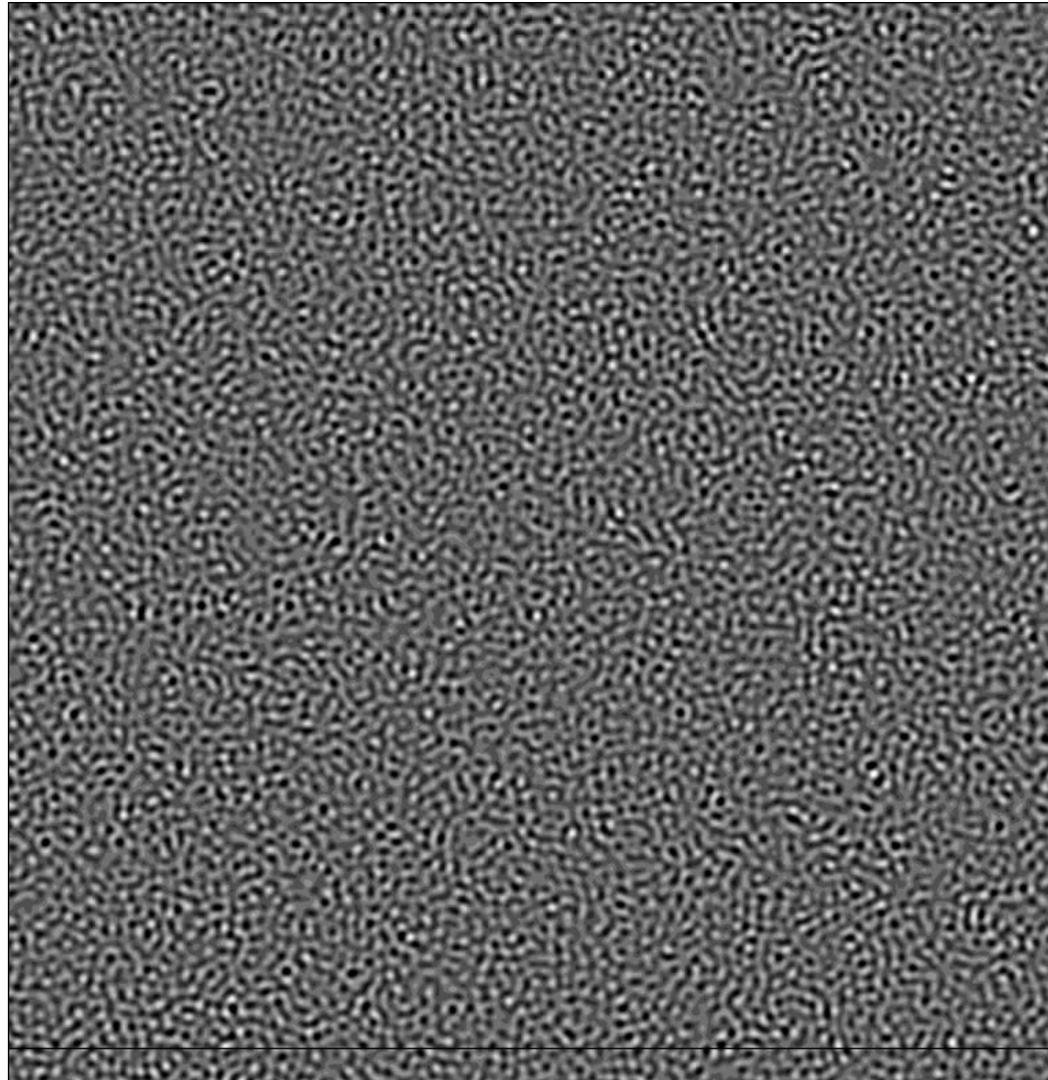


→ Original fluid interface

# Movie 1: Bubble height map for the DNS data (3Kx3K sub-sampled to 1Kx1K)

---

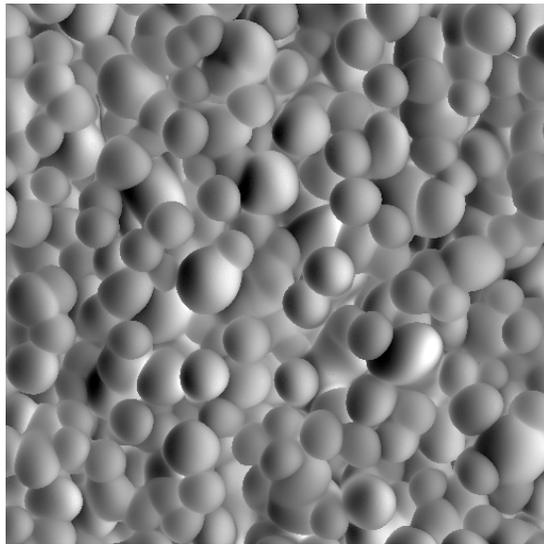
---



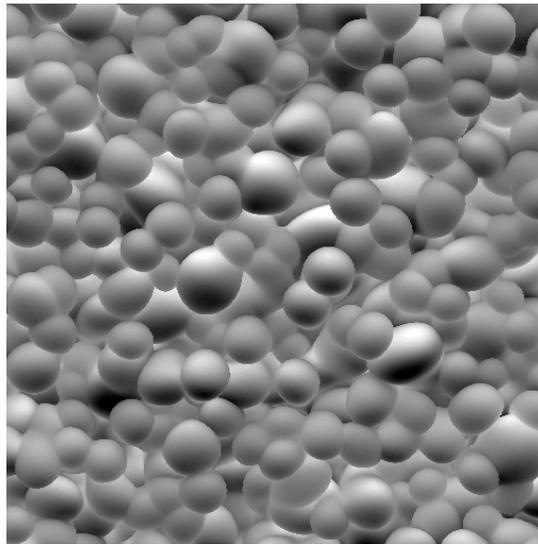
# Bubble counting – the mag-X-Y velocity (DNS, time step 50)

---

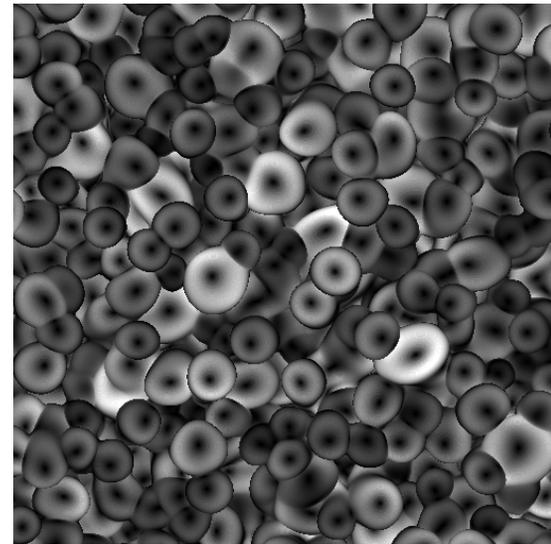
---



**X velocity**



**Y velocity**

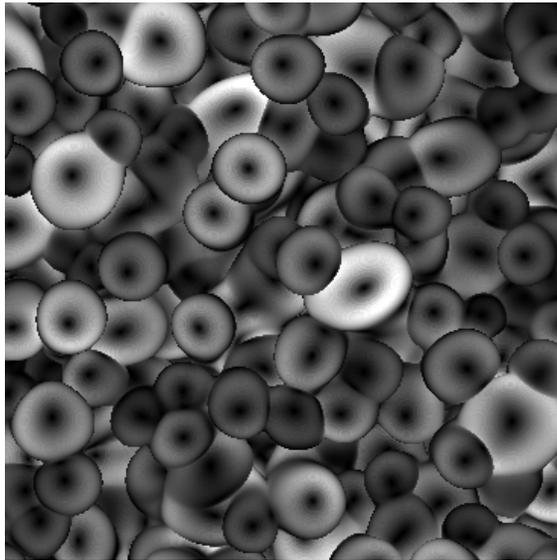


**Mag X-Y velocity**

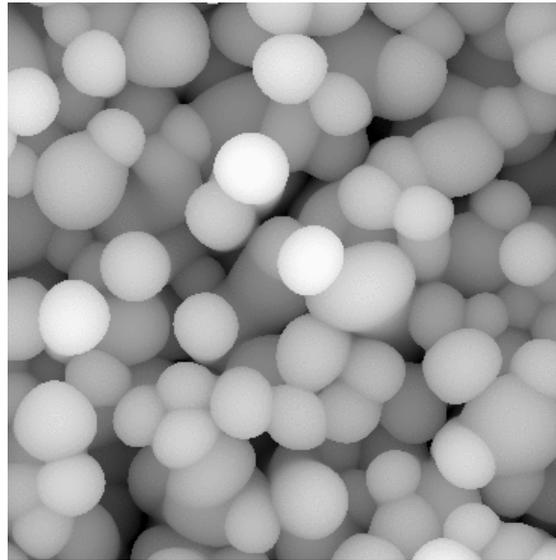
# Bubble counting – identifying the bubble tips

---

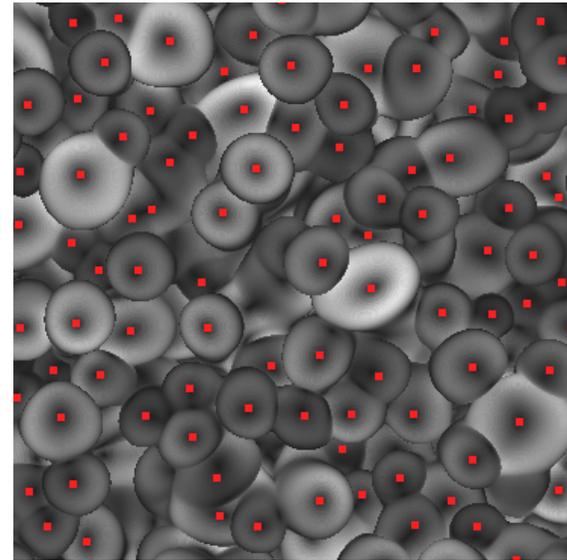
---



**Mag X-Y velocity**

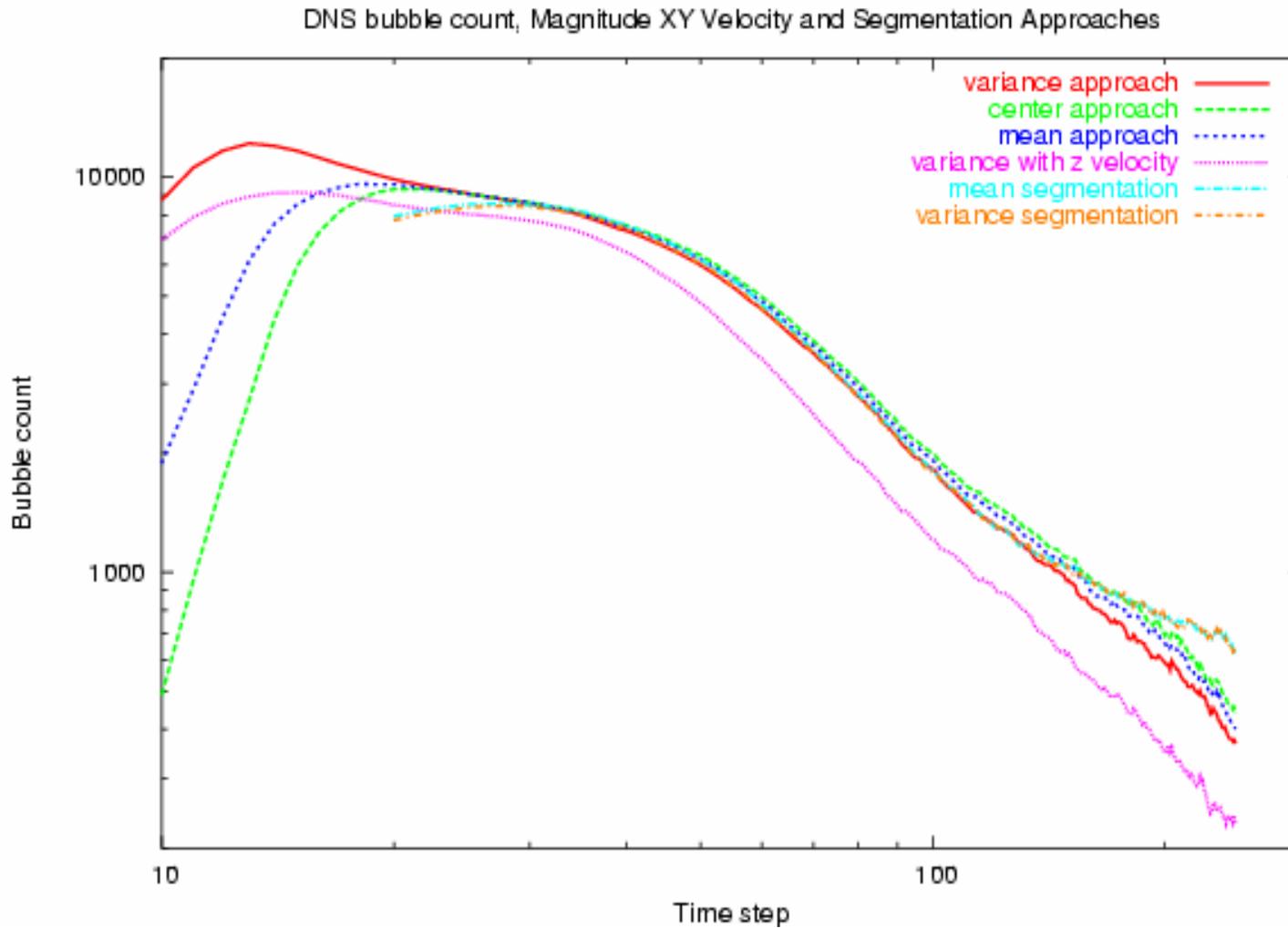


**Height-depth map**

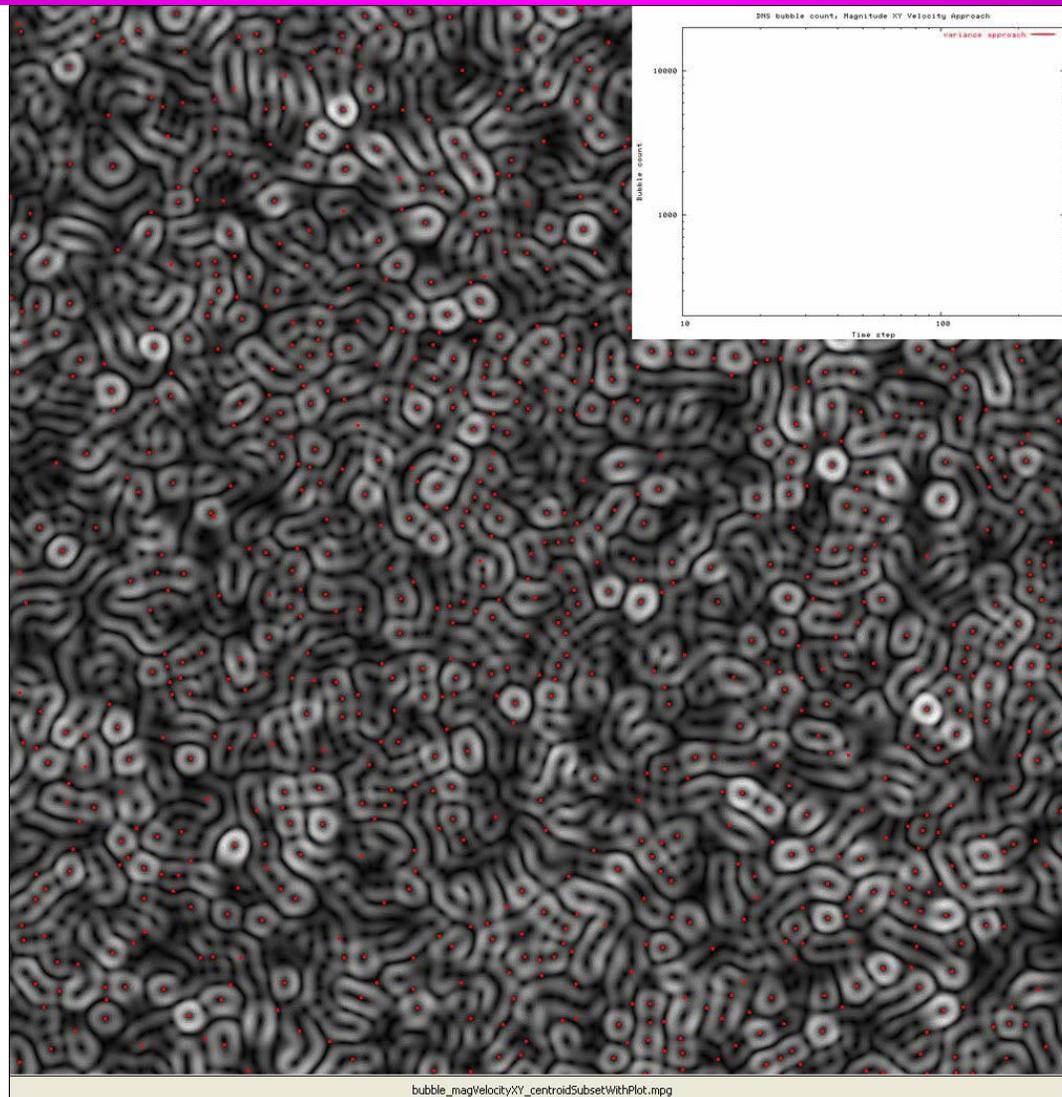


**Bubble tips**

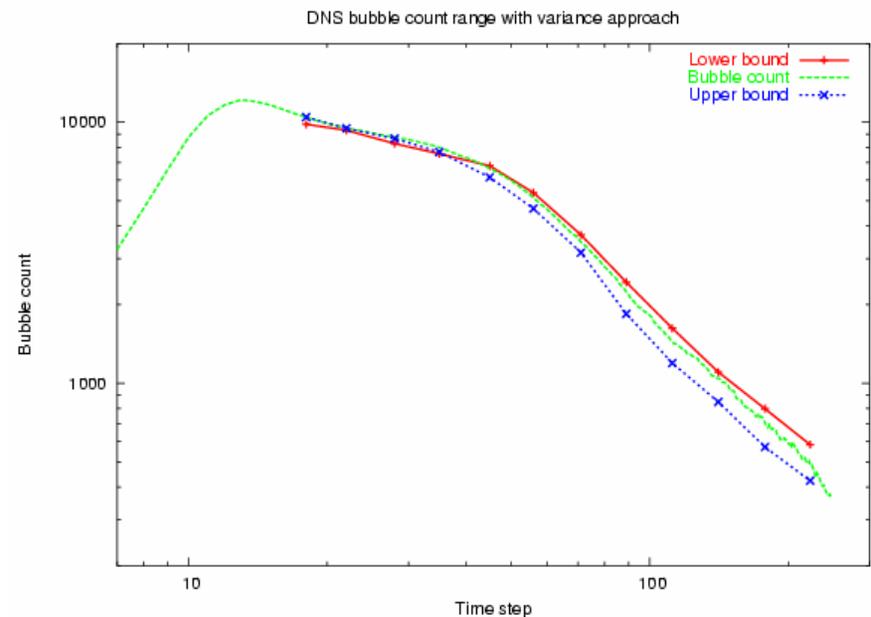
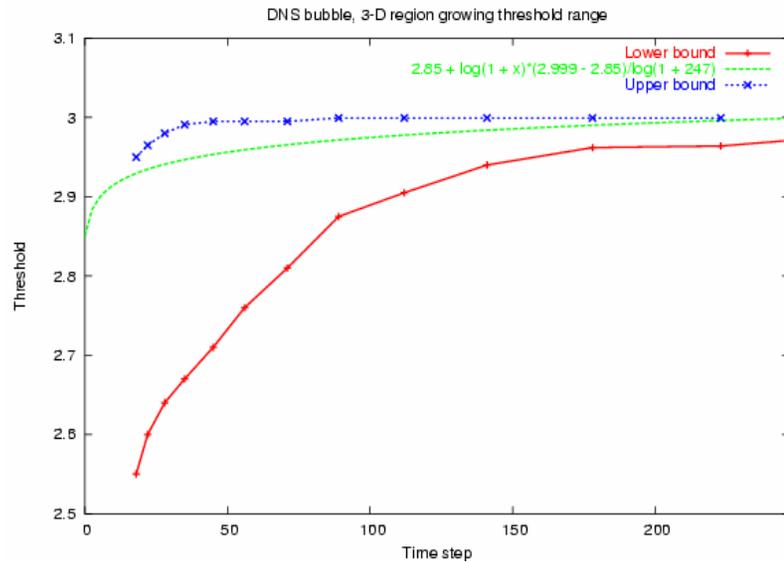
# Bubble counts over time: DNS data



# Movie 2: Bubble tips for DNS data (1Kx1K subset of full image)

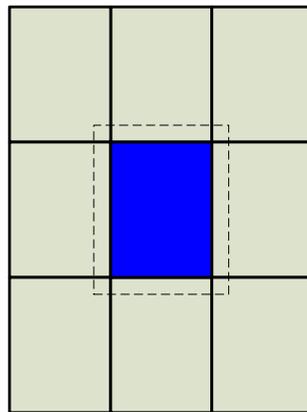
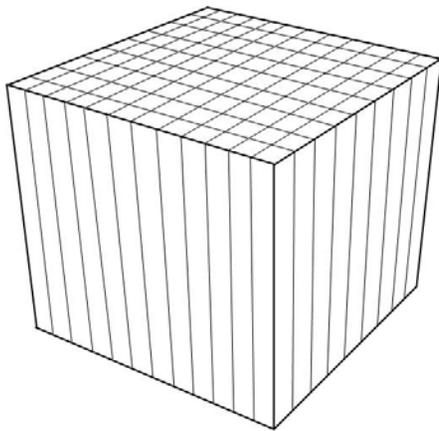


# Sensitivity of the DNS bubble count to changing 3-D region-growing threshold

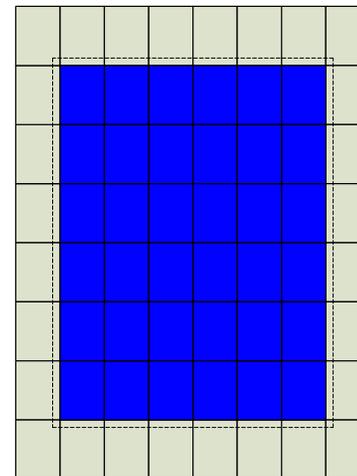


# We had to address several issues related to the large size of the data

- Various options for processing the 3D columns
- Converting the 3-D data to 2-D
  - pipeline the analysis: read, process, write
  - using several machines, with multiple jobs
- 2-D data analyzed mainly on local workstations



8 extra reads/column  
34.5K seconds/dump  
(LES)



~0.78 extra reads/column  
17.5K seconds/dump  
(LES)

# Computational times for analysis

- Reducing 3-D data to 2-D : limited by time taken to get data from storage

	Using 1 proc.	Using available proc.
LES	25 days	7 days
DNS	88 days	4 days (data obtained prior)

- Identification of bubble tips for DNS

Time step	Accessing data	3D to 2D	Mag X-Y	Segmentation
50	6960s	669s	8s	1457s
150	9960s	2459s	8s	2850s
247	9780s	3041s	8s	3596s

DNS data: took 2303 single-CPU years to generate; used 16K-64K processors

# Challenges in mining scientific data at the megabyte, terabyte, and beyond ....

---

- **Data-driven** analysis
  - need for objectivity
  - sensitivity of results to algorithms and parameters
  - physically-meaningful results
- **Complexity** of the data, regardless of the size
  - robust feature extraction
  - improving the quality of data: noise, outliers, missing values, ...
  - handling variations in data, scale of structures of interest
- **Massive size** can create additional challenges
  - data access
  - structures of interest distributed across processors/files

# Acknowledgements

---

- **Data: William Cabot and Andrew Cook (LLNL)**
- **Livermore Computing folks for support with the management of massive datasets**
- **Rebecca Springmeyer and Jeffery Long for their support of the work**
- **Sapphire team for the software used in the analysis**
- **Funding sources for Sapphire**
  - DOE NNSA ASC program
  - DOE SciDAC program
  - LLNL LDRD program

[www.llnl.gov/casc/sapphire](http://www.llnl.gov/casc/sapphire)