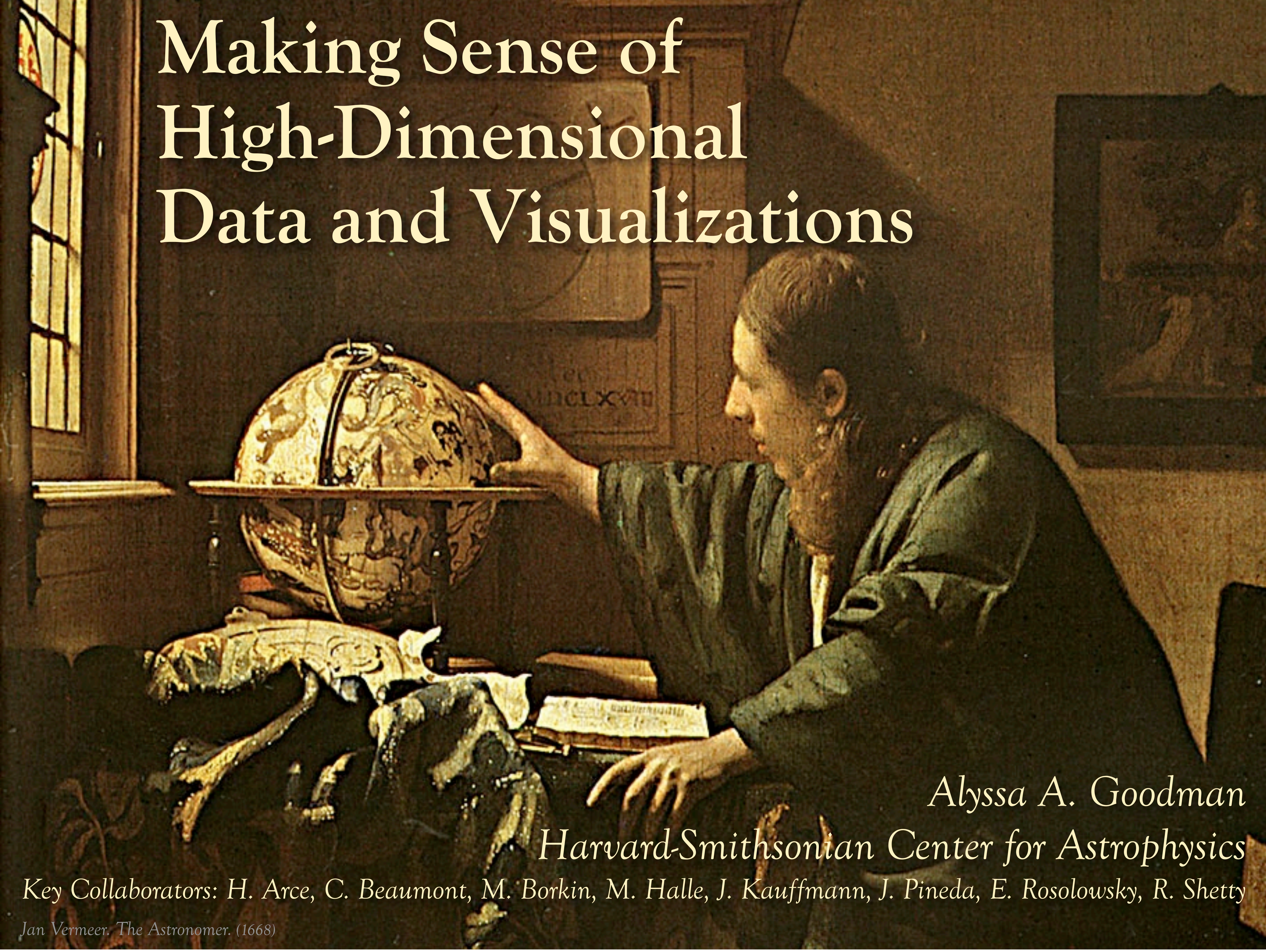


The "data deluge" in science is old news. Now, it's pouring, and we need working tools to collect, sort out, understand, and keep what is falling down on us. In astronomy, the greatest insights very often come from studies where more than one "band" of data (e.g. optical, infrared, radio, X-ray) is combined. And, data sets aren't just large--they are often also high-dimensional, in that they contain information about flux as functions not just of position on the sky, but also as functions of a third dimension (e.g. frequency, velocity), and/or of time. Life science, geophysical, and geospatial data all present similar challenges.

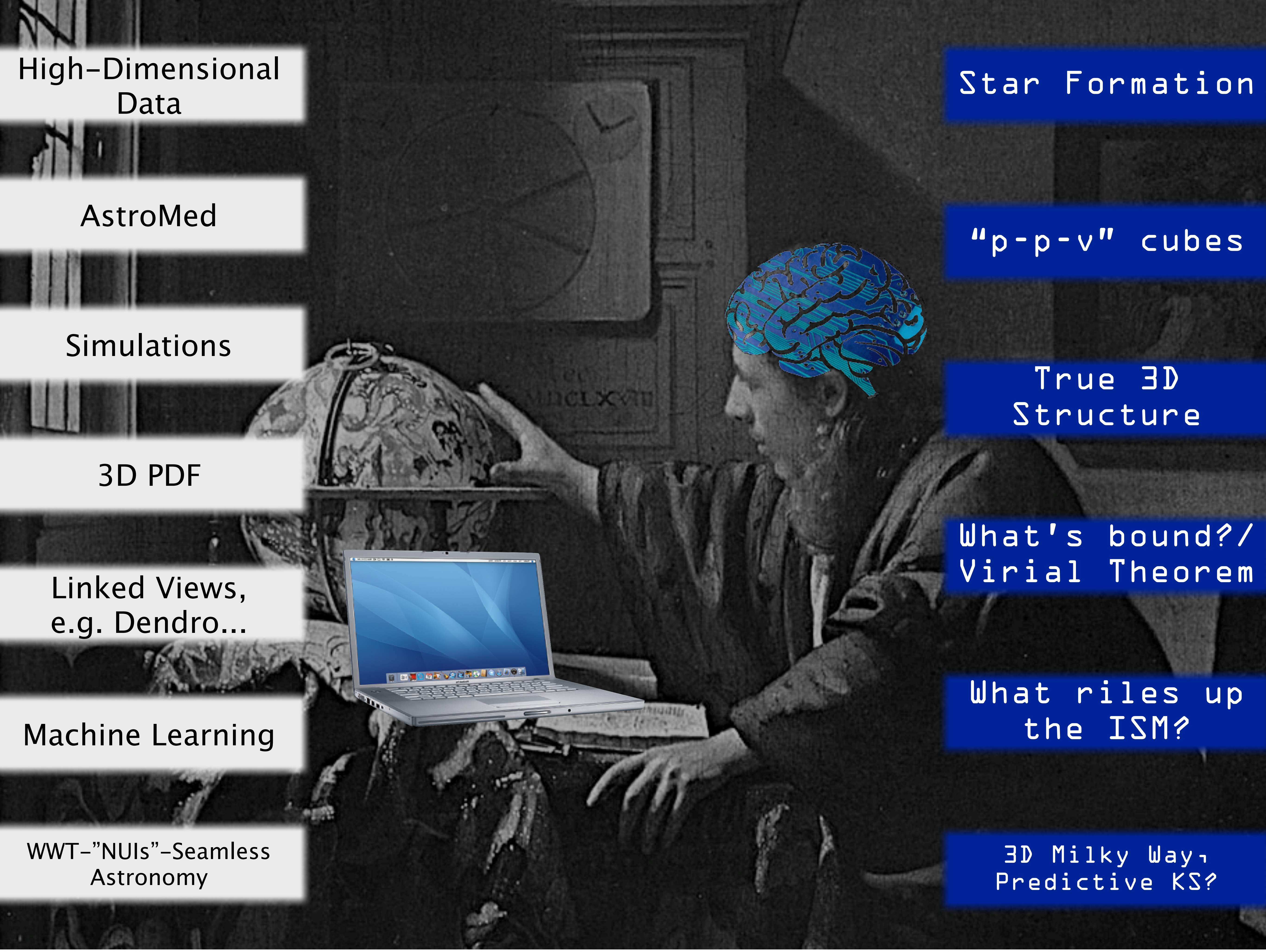
In this talk, I will focus on examples drawn from my group's research on star formation in molecular clouds. In particular, I will show how new visualization and statistical analysis techniques relying on interactive high-dimensional views of data and on automated algorithms for "segmenting" data give new insight. "Segmentation" in imaging terms refers to extracting the meaningful structures from data, and I will show results from both dendrogram (tree-hierarchy) and machine-learning approaches. I will emphasize how the visualization of segmentation results is critical for understanding. The highlighted science results will show how we can now--for the first time--quantitatively but intuitively understand the connections between the "real" (position-position-position) space where simulations (e.g. of star formation) can be made and the "observational" (e.g. position-position-velocity) space available to earthbound astronomers. As a result of this newfound understanding, we can place important limits on the validity of virial-theorem-based calculations of the properties of gas--allowing, for example, for better estimates of which gas in star-forming regions is most likely to stay bound long enough to form stars.

Even though this abstract may sound technical to non-star-formation or non-computational researchers, my goal will be to keep the talk accessible to non-experts, so people from other fields faced with high-dimensional data and visualization challenges should feel free to join in--and to ask questions

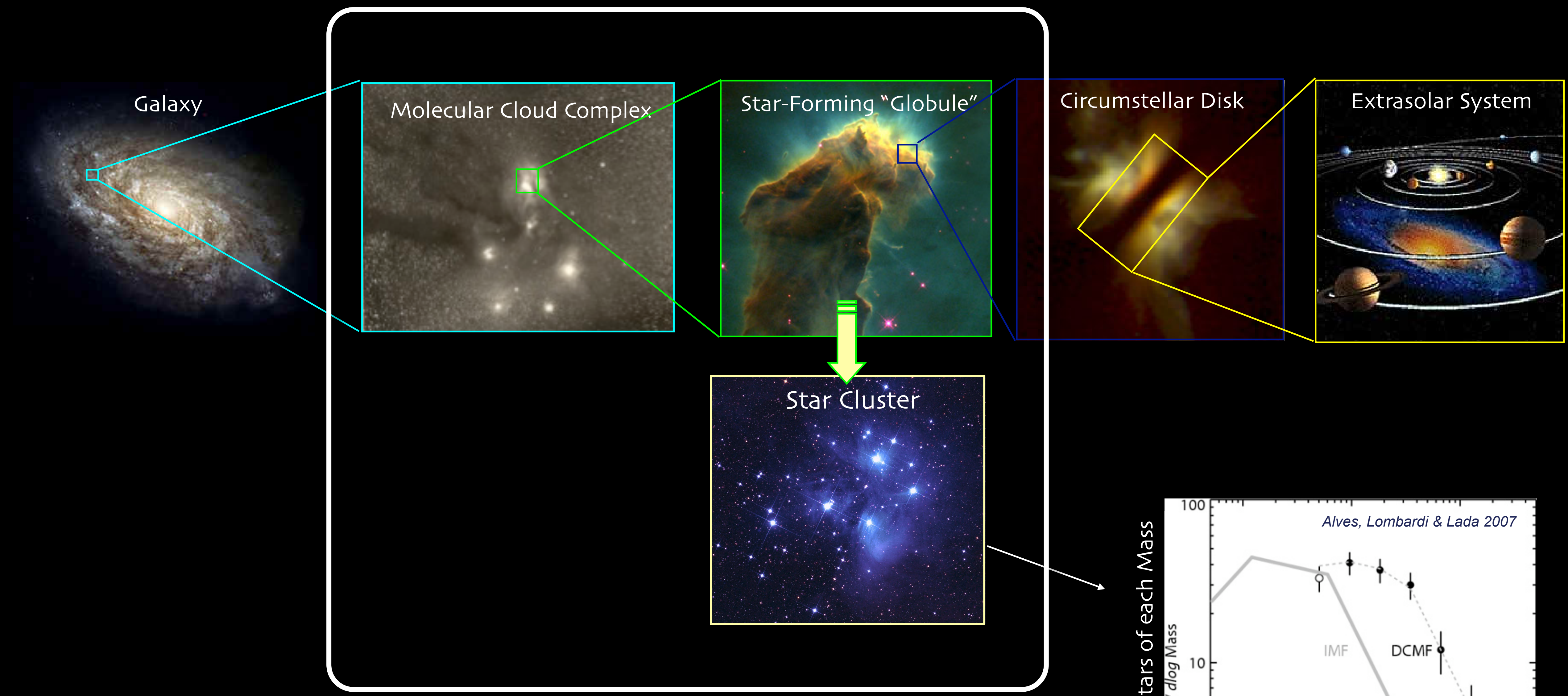
Relative Strengths

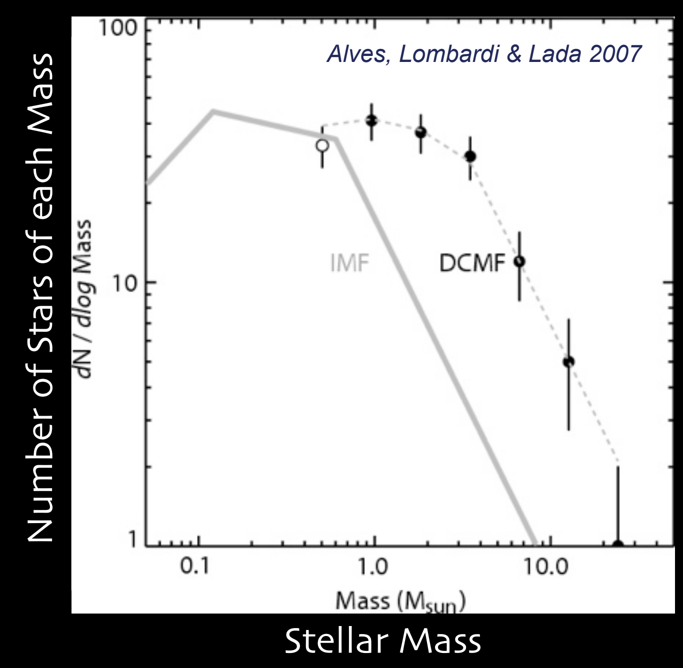
Calculations

Pattern Recognition
Creativity

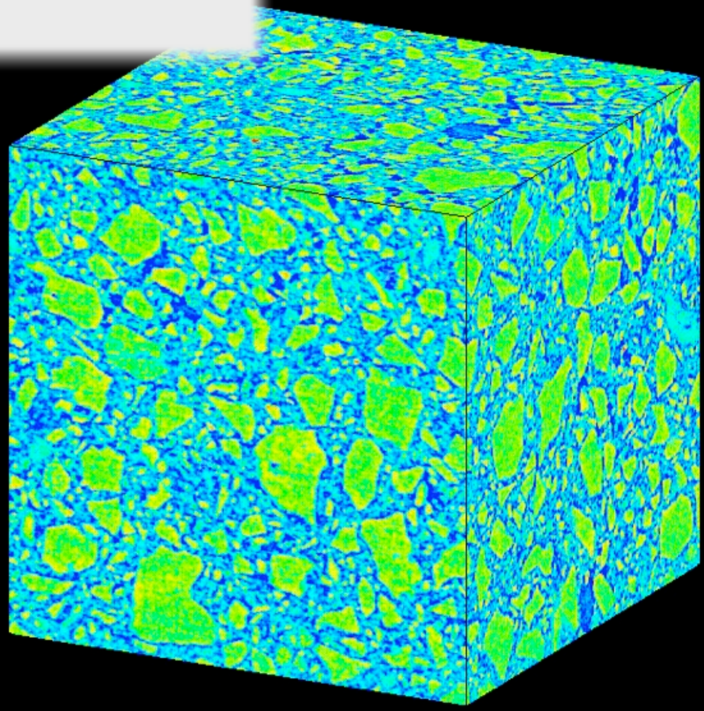

High-Dimensional Data
AstroMed
Simulations
3D PDF
Linked Views, e.g. Dendro...
Machine Learning
WWT-"NUIs"-Seamless Astronomy
Star Formation
"p-p-v" cubes
True 3D Structure
What's bound?/ Virial Theorem
What riles up the ISM?
3D Milky Way, Predictive KS?

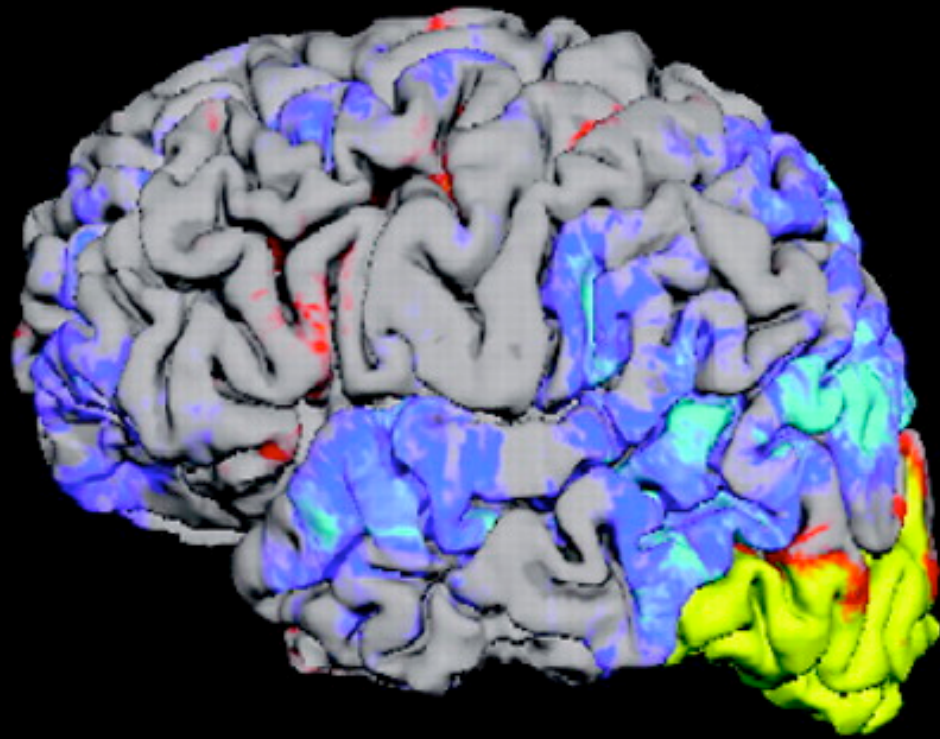
Star (and Planet, and Moon) Form

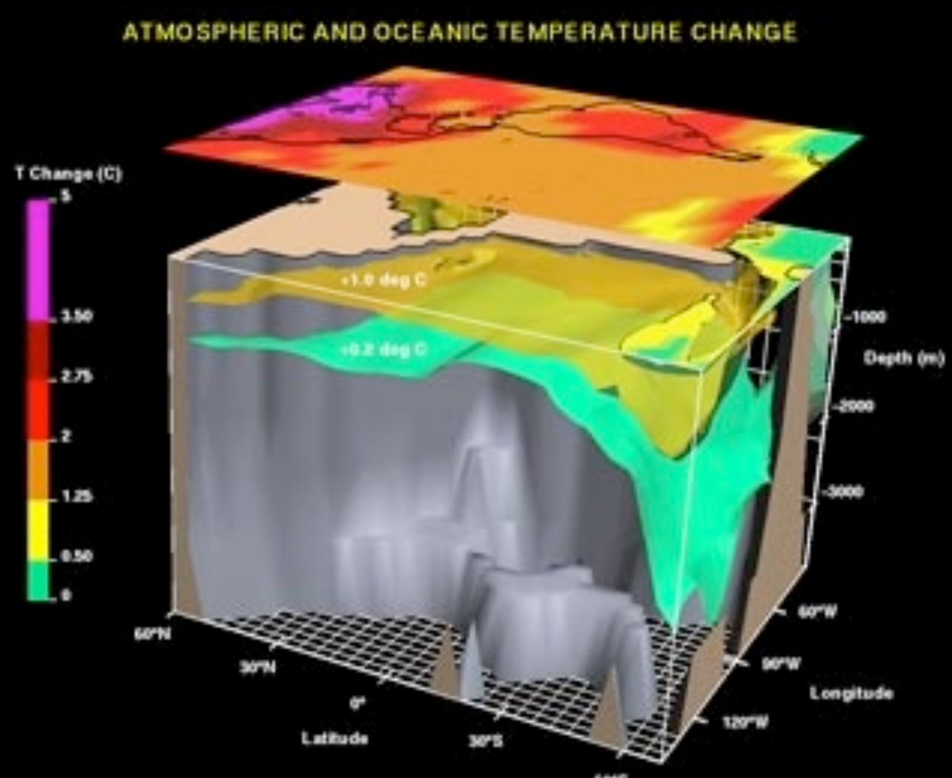


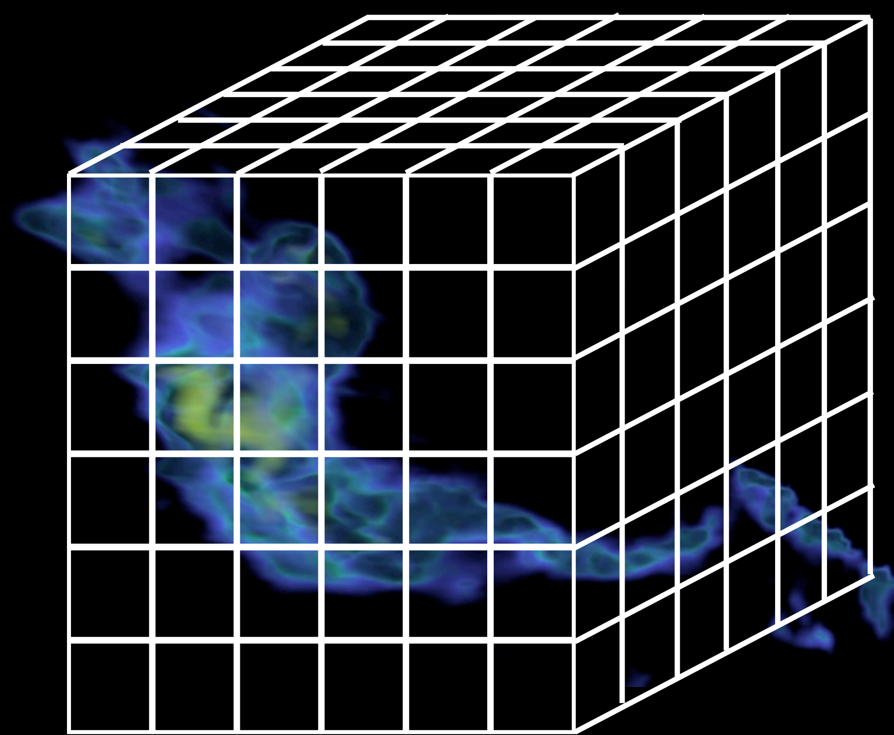


High-Dimensional Data

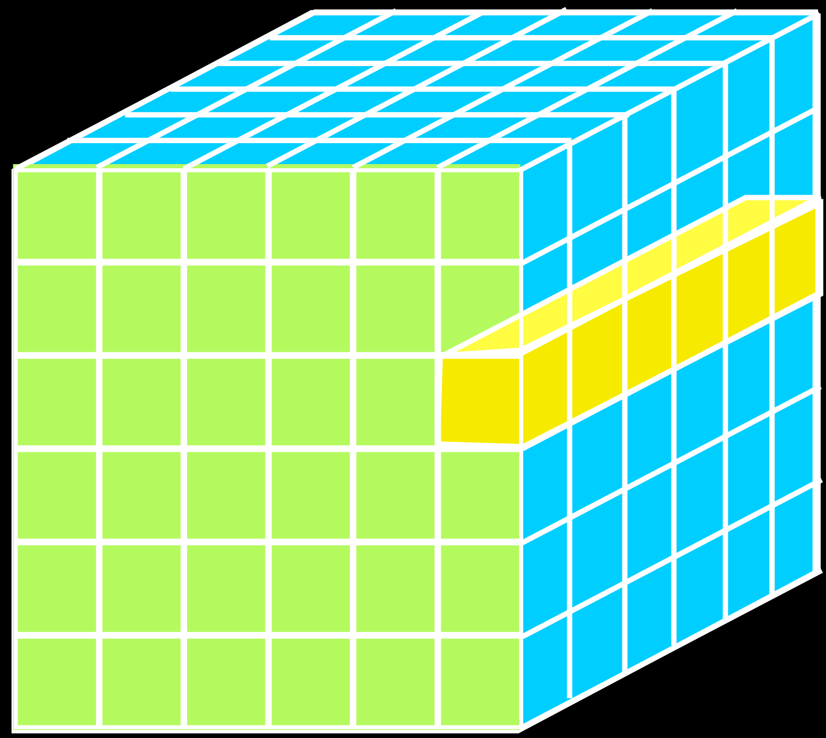








High-Dimensional Data



GENERALLY

1D: Columns = "Spectra", "SEDs" or "Time Series"

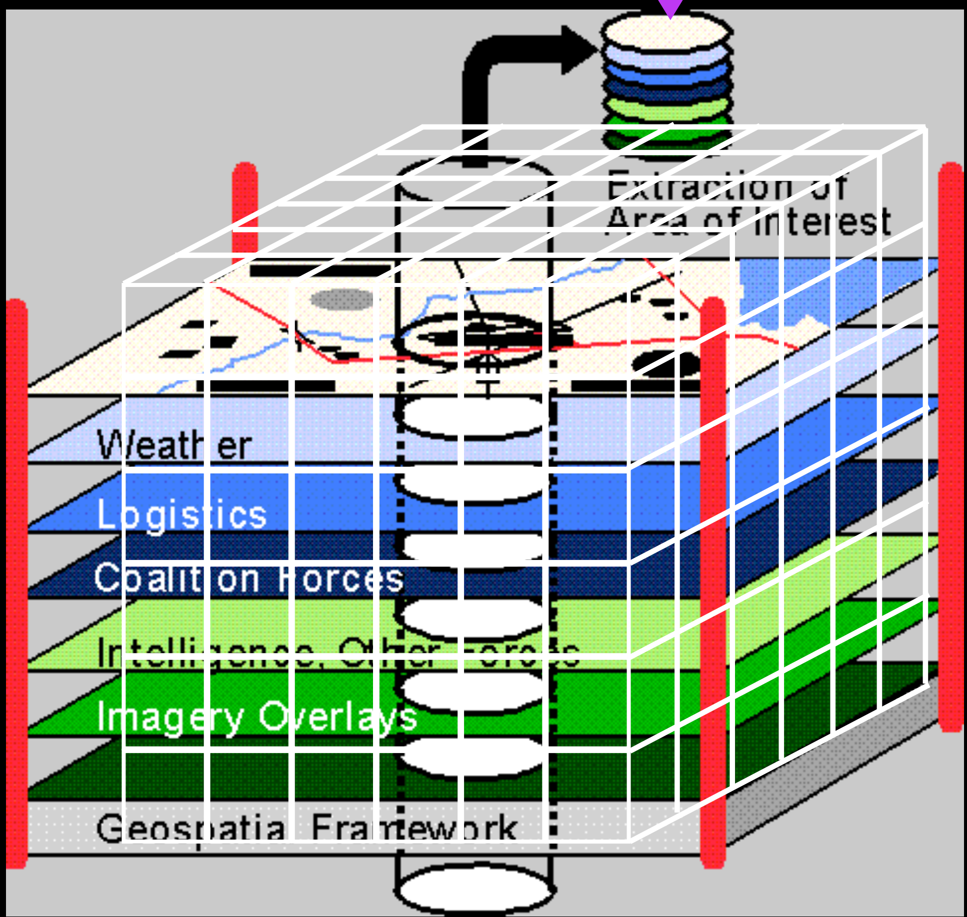
2D: Faces or Slices = "Images"

3D: Volumes = "3D Renderings"

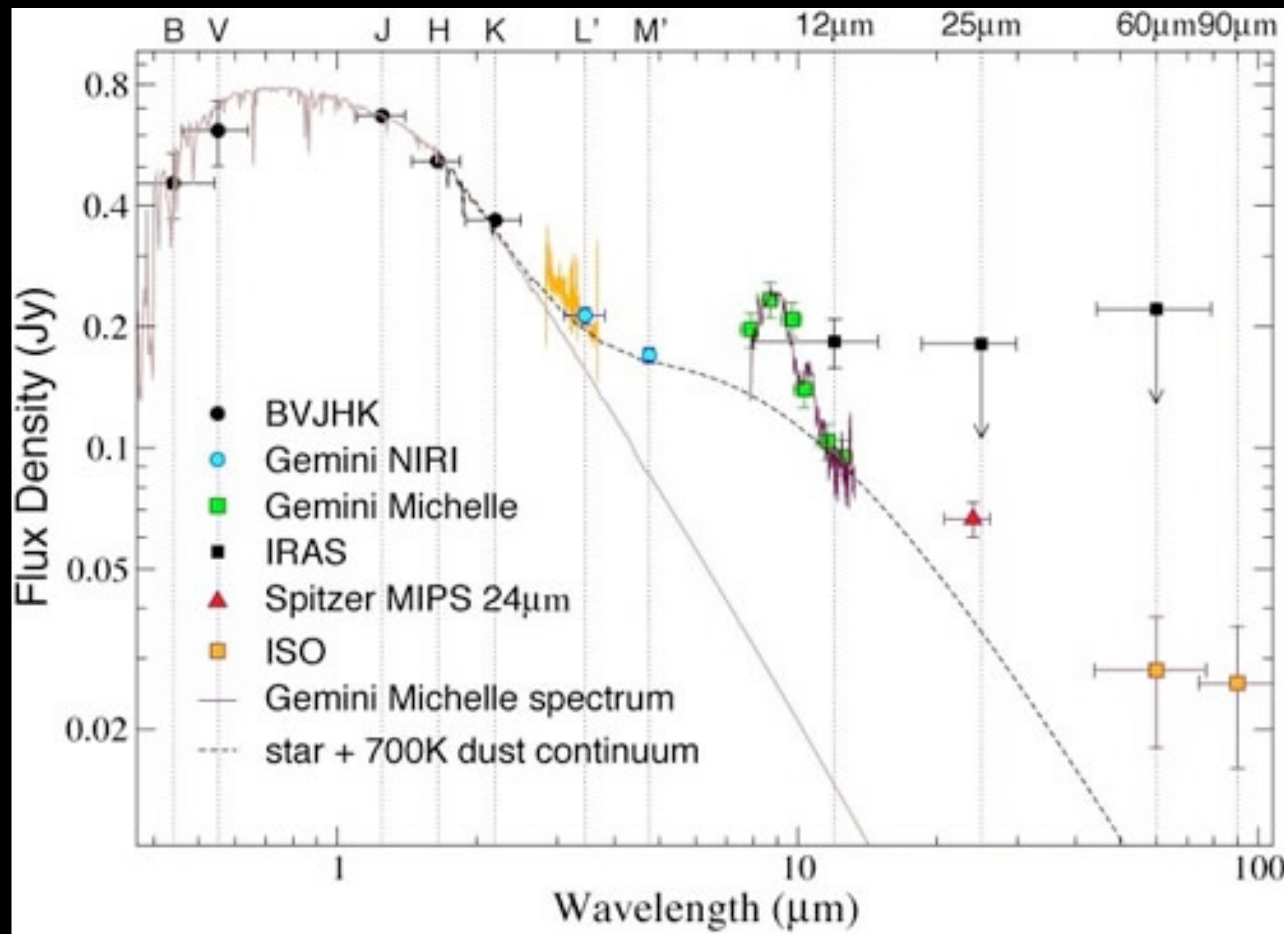
4D: Time Series of Volumes = "3D Movies"

High-Dimensional Data

This



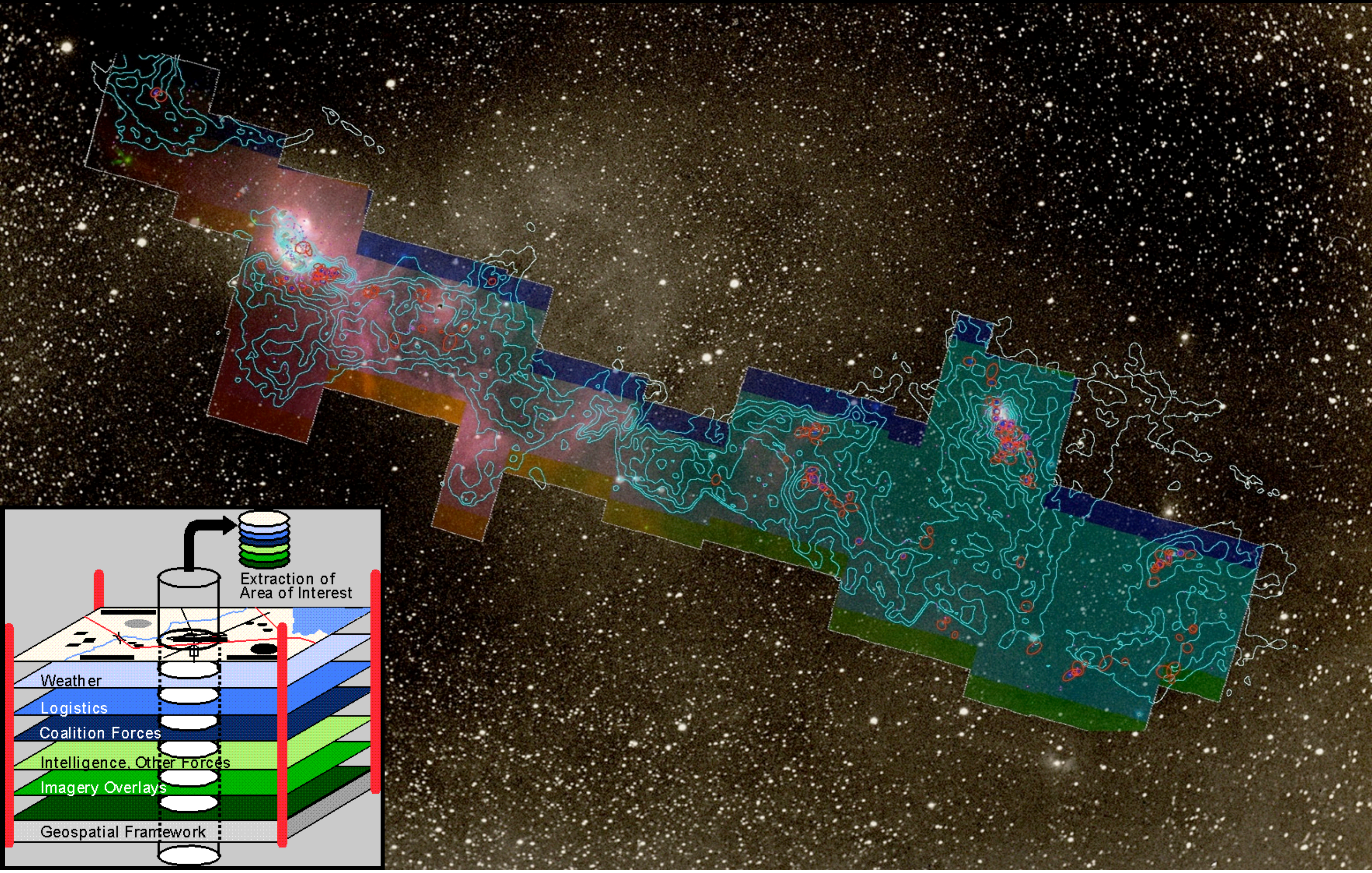
is a "spectral energy distribution"



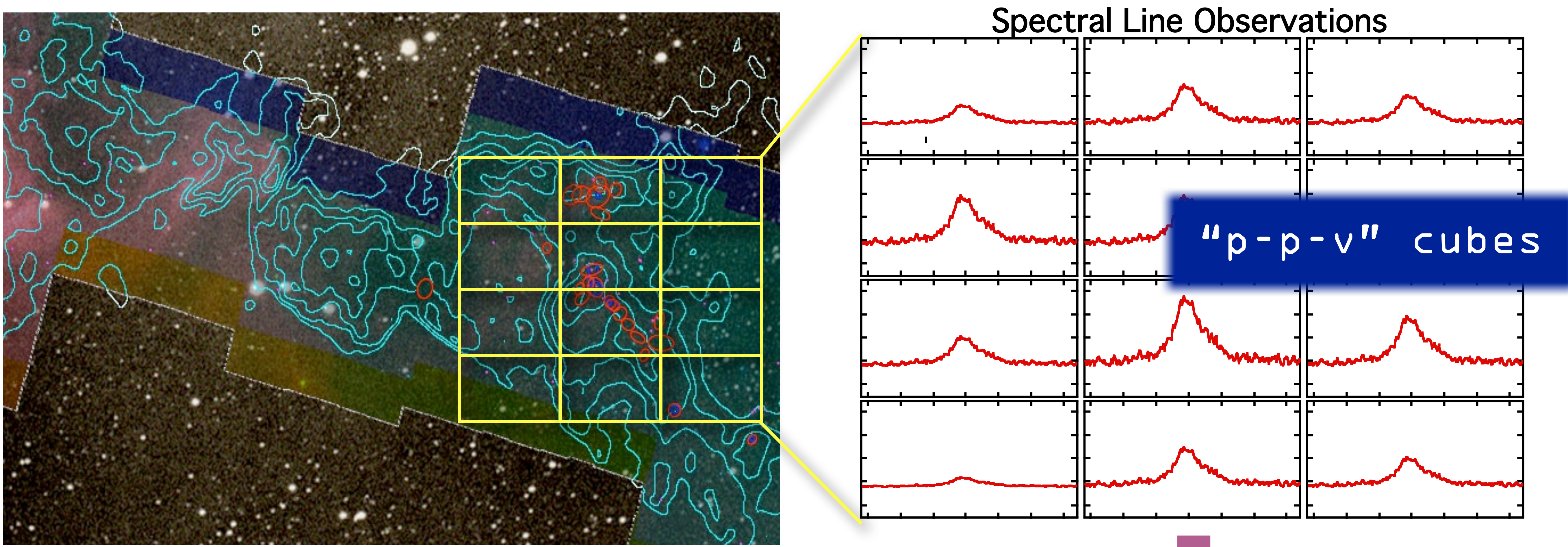
High-Dimensional Data

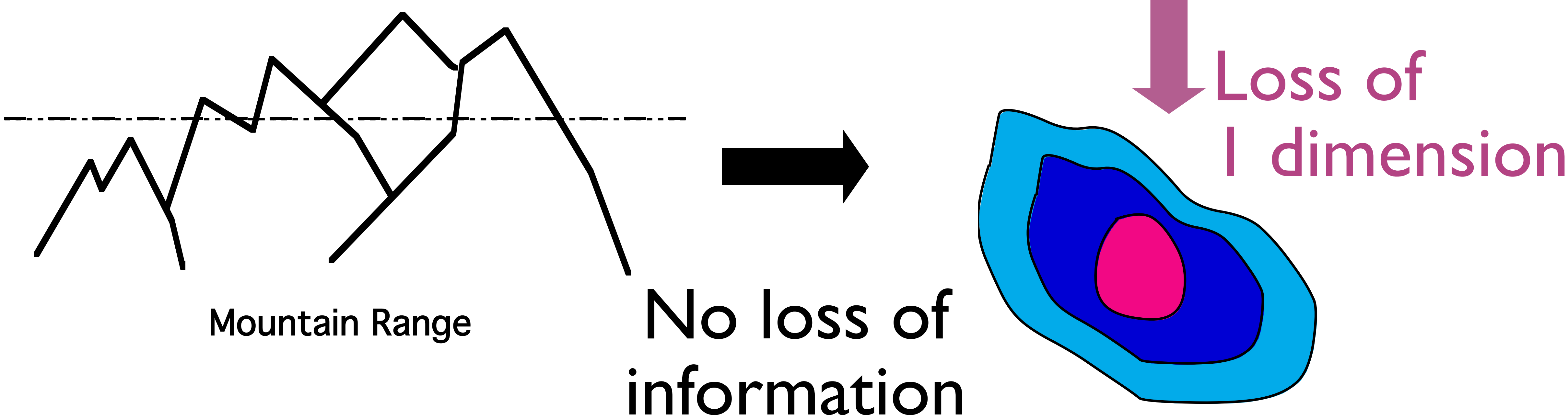
COMPLETE Perseus

Star Formation



Spectral-Line Mapping



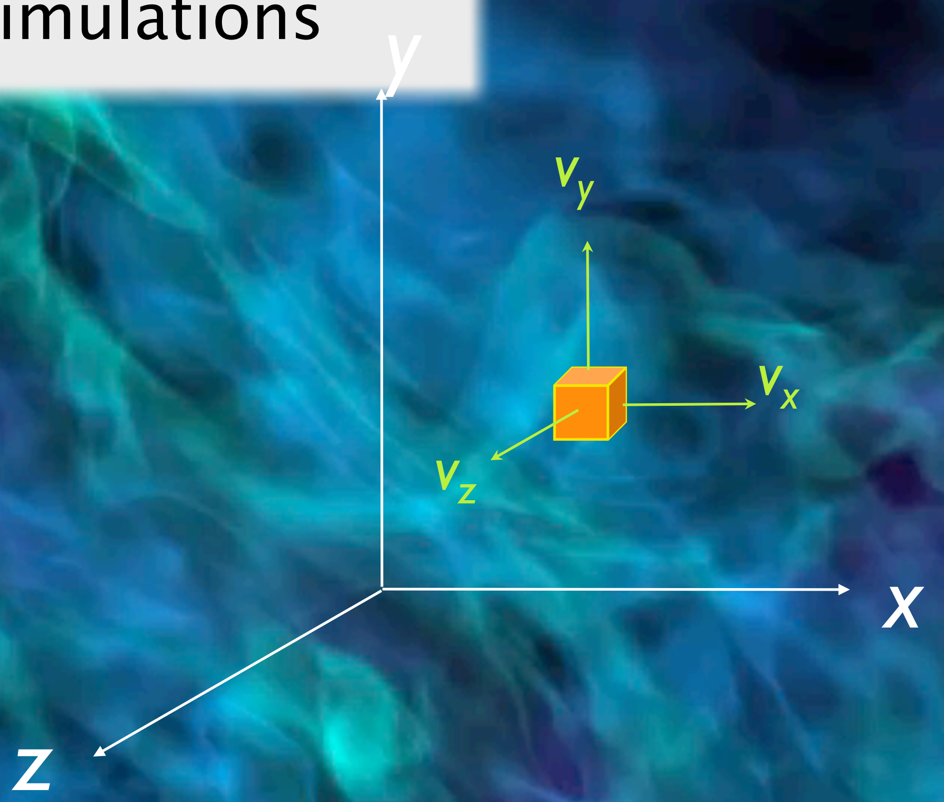


"Three" Dimensions: Spectral-Line Mapping

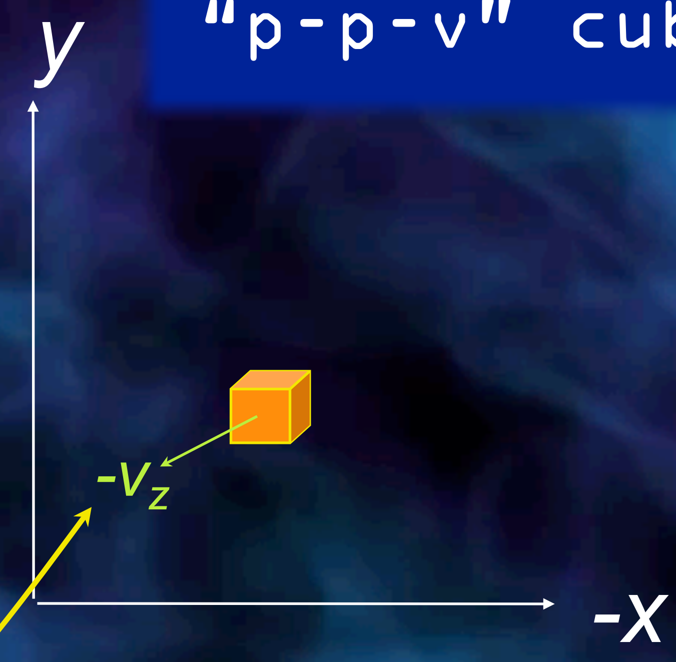
We wish we could measure…

But we can measure…

Simulations

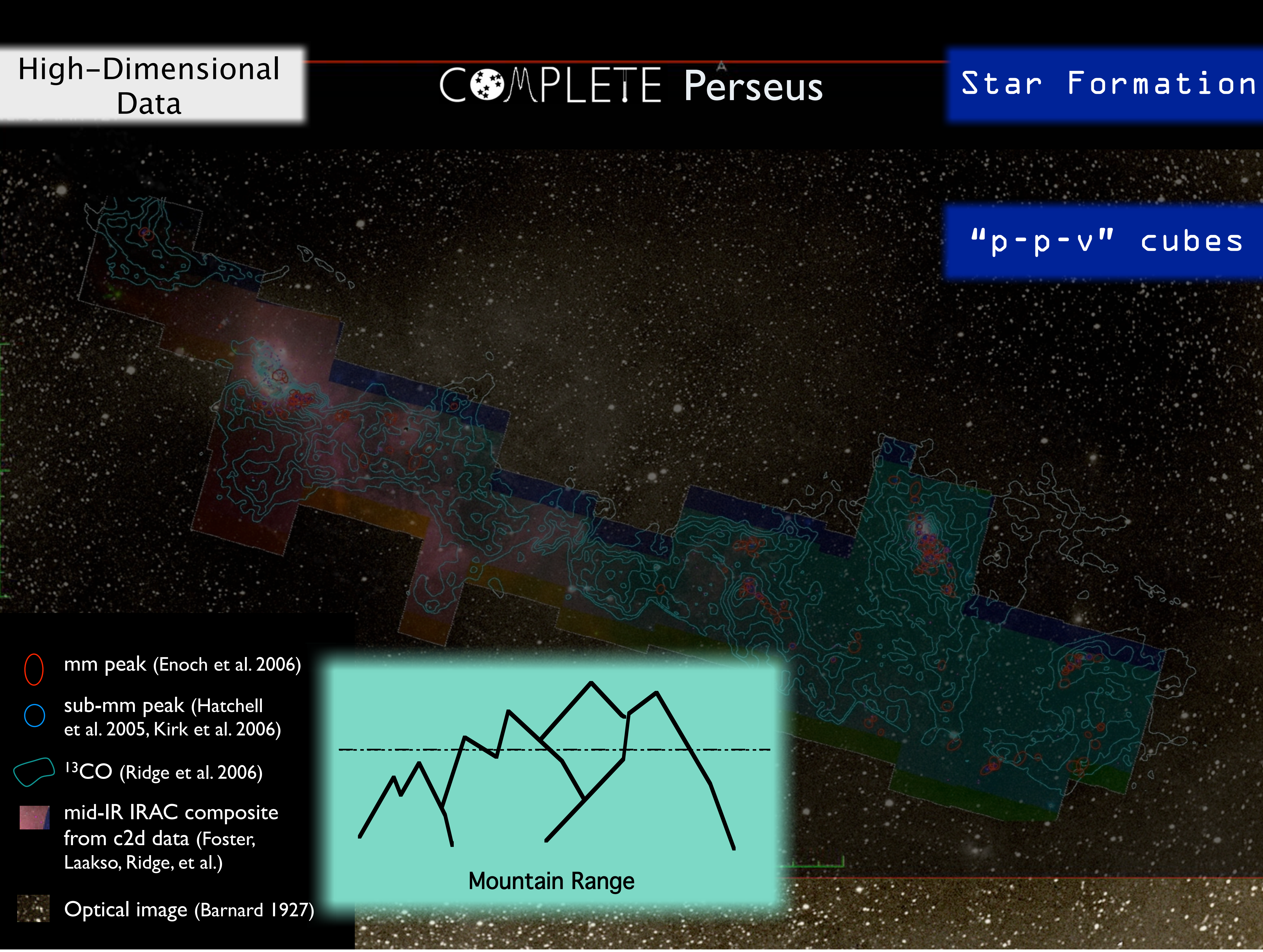


"p-p-v" cubes



v_z *only* from "spectral-line maps"

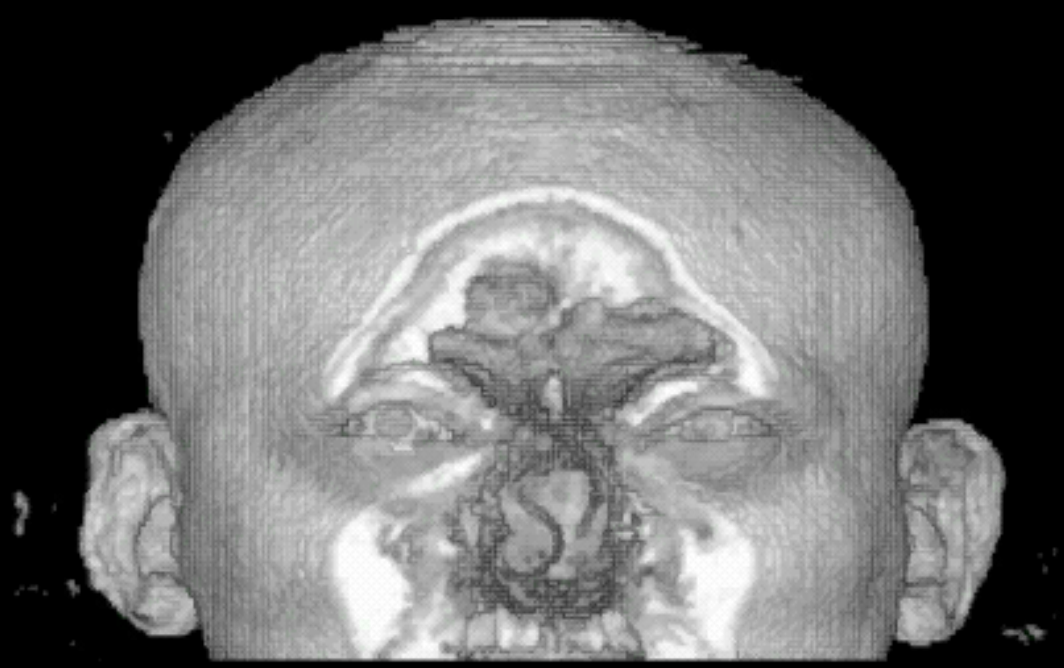
Hydrodynamic AMR Simulation, courtesy Stella Offner


High-Dimensional Data
COMPLETE Perseus
Star Formation
"p-p-v" cubes
mm peak (Enoch et al. 2006)
sub-mm peak (Hatchell et al. 2005, Kirk et al. 2006)
13CO (Ridge et al. 2006)
mid-IR IRAC composite from c2d data (Foster, Laakso, Ridge, et al.)
Optical image (Barnard 1927)
Mountain Range

“Astronomical Medicine”

AstroMed

“KEITH”



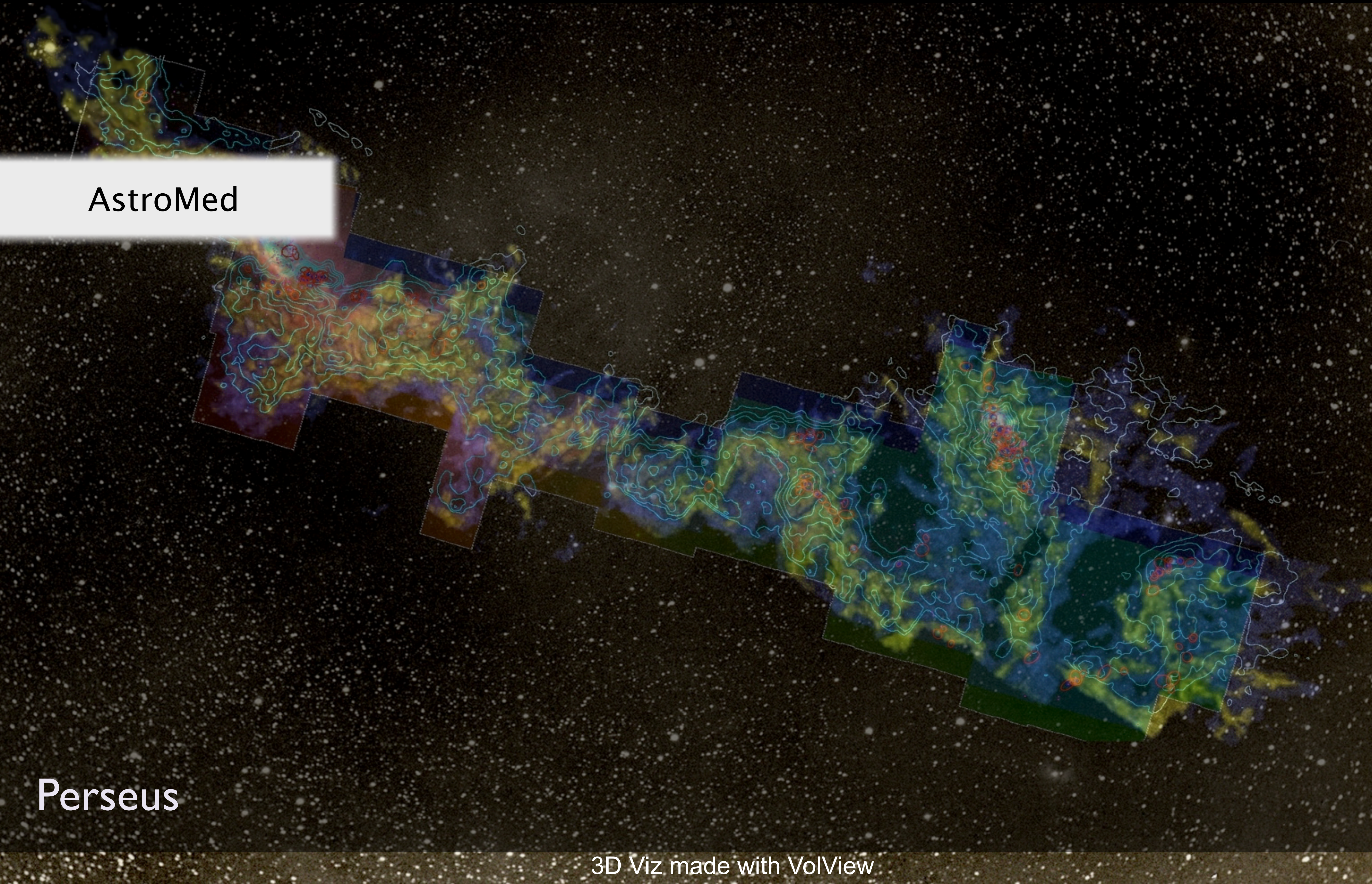
“z” is depth into head

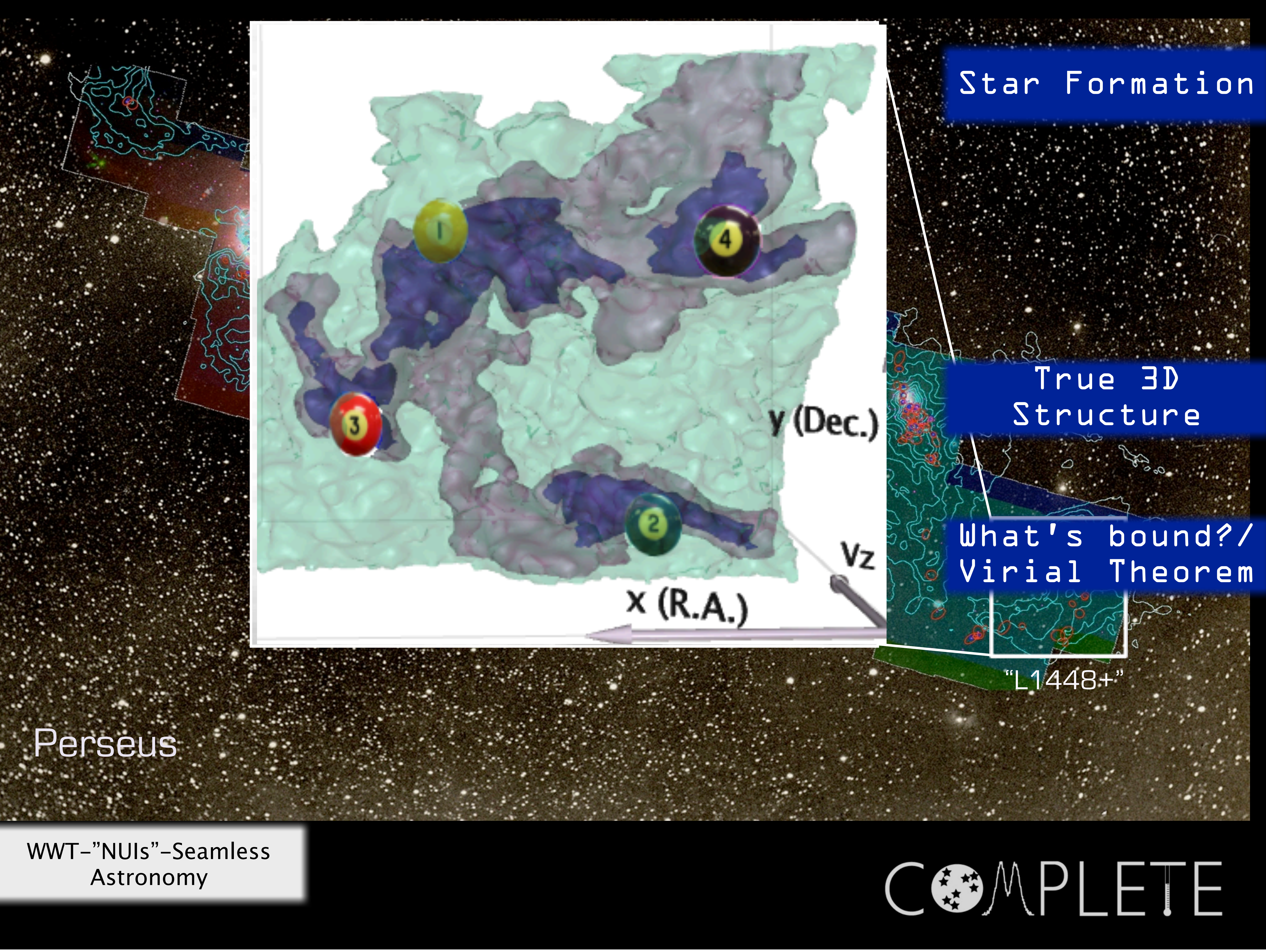
“PERSEUS”

“z” is line-of-sight velocity

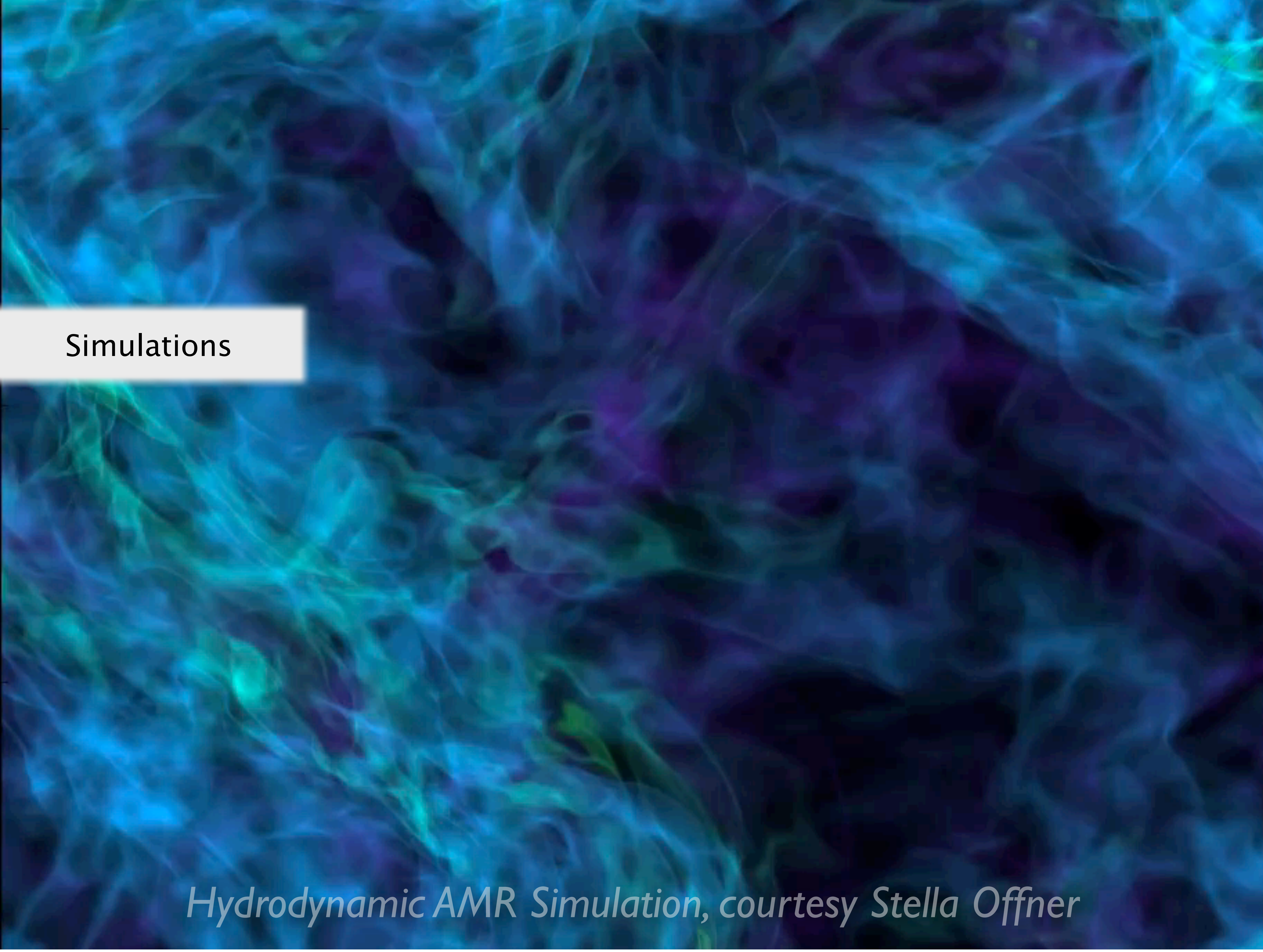
http://am.iic.harvard.edu/

AstroMed

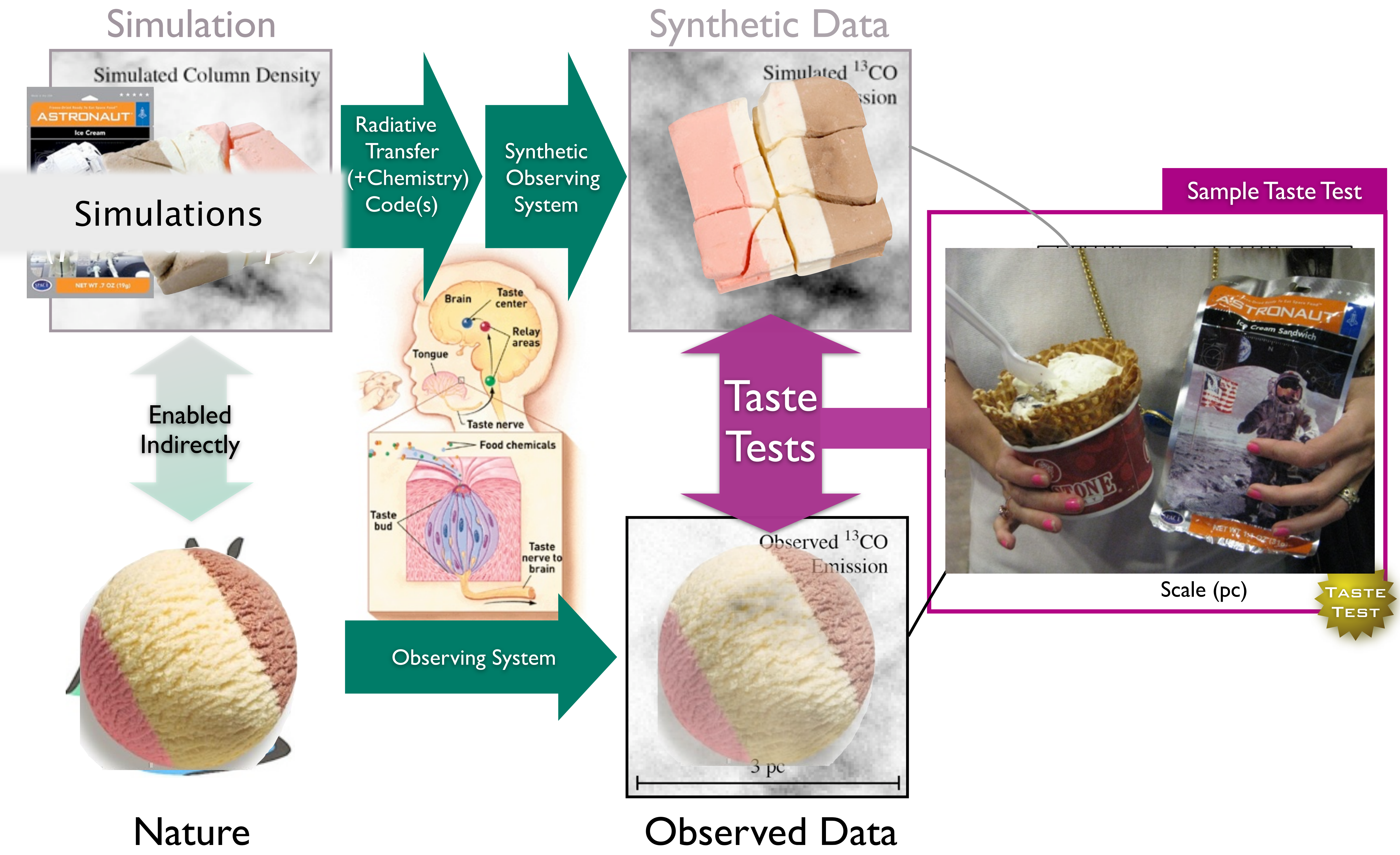




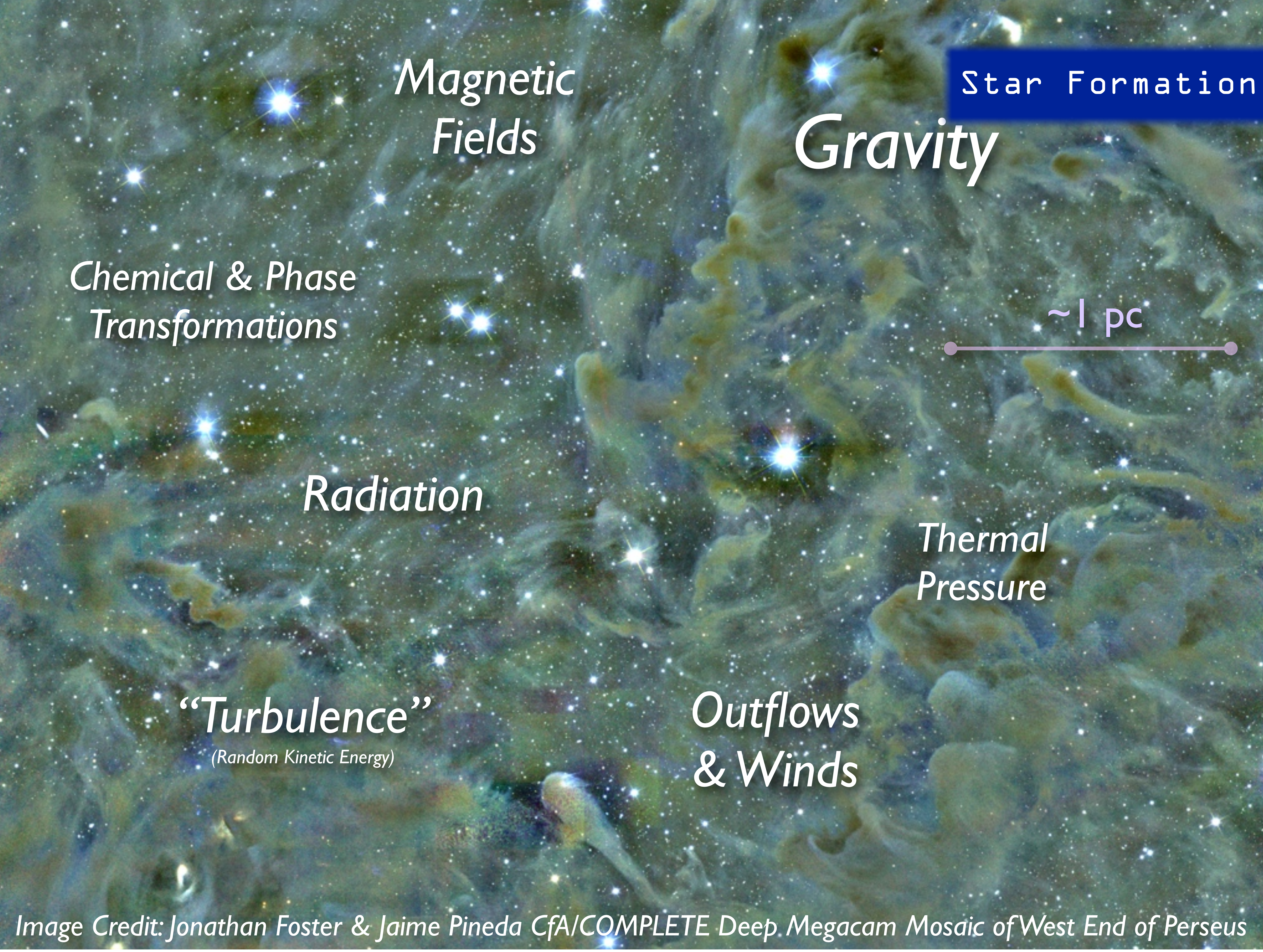
Star Formation
True 3D Structure
What's bound?/ Virial Theorem
y (Dec.)
Vz
x (R.A.)
"L1448+"
Perseus
WWT-"NUIs"-Seamless Astronomy
COMPLETE


Simulations
Hydrodynamic AMR Simulation, courtesy Stella Offner

"Taste-Testing" Simulations



Tuesday, March 22, 2011


Star Formation
Magnetic Fields
Gravity
Chemical & Phase Transformations
~1 pc
Radiation
Thermal Pressure
"Turbulence"
(Random Kinetic Energy)
Outflows & Winds
Image Credit: Jonathan Foster & Jaime Pineda CfA/COMPLETE Deep Megacam Mosaic of West End of Perseus

Taste-Testing "Gravity"

High-Dimensional Data

AstroMed

Simulations

3D PDF

Star Formation

"p-p-v" cubes

True 3D Structure

What's bound?/ Virial Theorem

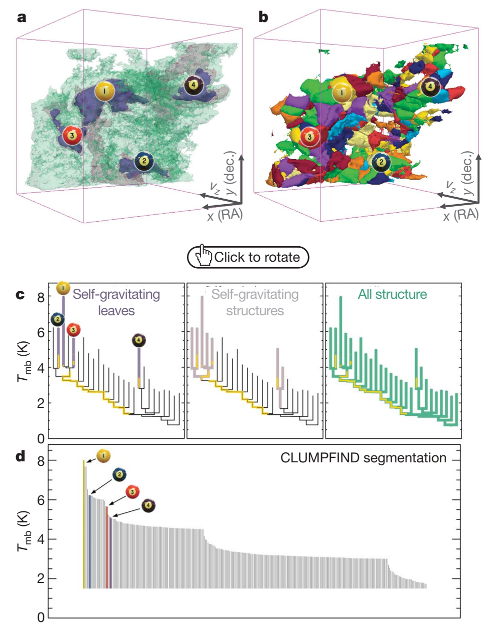


Figure 2 | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to ^{13}CO emission from the L1448 region of Perseus. a, 3D visualization of the surfaces indicated by colours in the dendrogram shown in **c**. Purple illustrates the smallest scale self-gravitating structures in the region corresponding to the leaves of the dendrogram; pink shows the smallest surfaces that contain distinct self-gravitating leaves within them; and green corresponds to the surface in the data cube containing all the significant emission. Dendrogram branches corresponding to self-gravitating objects have been highlighted in yellow over the range of T_{mb} (main-beam temperature) test-level values for which the virial parameter is less than 2. The x–y locations of the four 'self-gravitating' leaves labelled with billiard balls are the same as those shown in Fig. 1. The 3D visualizations show position–position–velocity (p–p–v) space. RA, right ascension; dec., declination. For comparison with the ability of dendrograms (**c**) to track hierarchical structure, **d** shows a pseudo-dendrogram of the CLUMPFIND segmentation (**b**), with the same four labels used in Fig. 1 and in **a**. As 'clumps' are not allowed to belong to …

"Dendrogram"

… interactive online version) corresponds exactly to the patch of sky shown in Fig. 1, and velocity with respect to the Local Standard of Rest increases from front ($-0.5\ \mathrm{km\,s^{-1}}$) to back ($8\ \mathrm{km\,s^{-1}}$).

data, CLUMPFIND typically finds features on a limited range of scales, above but close to the physical resolution of the data, and its results can be overly dependent on input parameters. By tuning CLUMPFIND's two free parameters, the same molecular-line data set[8] can be used to show either that the frequency distribution of clump mass is the same as the initial mass function of stars or that it follows the much shallower mass function associated with large-scale molecular clouds (Supplementary Fig. 1).

Four years before the advent of CLUMPFIND, 'structure trees'[9] were proposed as a way to characterize clouds' hierarchical structure using 2D maps of column density. With this early 2D work as inspiration, we have developed a structure-identification algorithm that abstracts the hierarchical structure of a 3D (p–p–v) data cube into an easily visualized representation called a 'dendrogram'[10]. Although well developed in other data-intensive fields[11,12], it is curious that the application of tree methodologies so far in astrophysics has been rare, and almost exclusively within the area of galaxy evolution, where 'merger trees' are being used with increasing frequency[13].

Figure 3 and its legend explain the construction of dendrograms schematically. The dendrogram quantifies how and where local maxima of emission merge with each other, and its implementation is explained in Supplementary Methods. Critically, the dendrogram is determined almost entirely by the data itself, and it has negligible sensitivity to algorithm parameters. To make graphical presentation possible on paper and 2D screens, we 'flatten' the dendrograms of 3D data (see Fig. 3 and its legend), by sorting their 'branches' to not cross, which eliminates dimensional information on the x axis while preserving all information about connectivity and hierarchy. Numbered 'billiard ball' labels in the figures let the reader match features between a 2D map (Fig. 1), an interactive 3D map (Fig. 2a online) and a sorted dendrogram (Fig. 2c).

A dendrogram of a spectral-line data cube allows for the estimation of key physical properties associated with volumes bounded by isosurfaces, such as radius (R), velocity dispersion (σ_v) and luminosity (L). The volumes can have any shape, and in other work[14] we focus on the significance of the especially elongated features seen in L1448 (Fig. 2a). The luminosity is an approximate proxy for mass, such that $M_{\mathrm{lum}} = X_{13\mathrm{CO}} L_{13\mathrm{CO}}$, where $X_{13\mathrm{CO}} = 8.0 \times 10^{20}\ \mathrm{cm^2\,K^{-1}\,km^{-1}\,s}$ (ref. 15; see Supplementary Methods and Supplementary Fig. 2). The derived values for size, mass and velocity dispersion can then be used to estimate the role of self-gravity at each point in the hierarchy, via calculation of an 'observed' virial parameter, $\alpha_{\mathrm{obs}} = 5\sigma_v^2 R / G M_{\mathrm{lum}}$. In principle, extended portions of the tree (Fig. 2, yellow highlighting) where $\alpha_{\mathrm{obs}} < 2$ (where gravitational energy is comparable to or larger than kinetic energy) correspond to regions of p–p–v space where self-gravity is significant. As α_{obs} only represents the ratio of kinetic energy to gravitational energy at one point in time, and does not explicitly capture external over-pressure and/or magnetic fields[16], its measured value should only be used as a guide to the longevity (boundedness) of any particular feature.

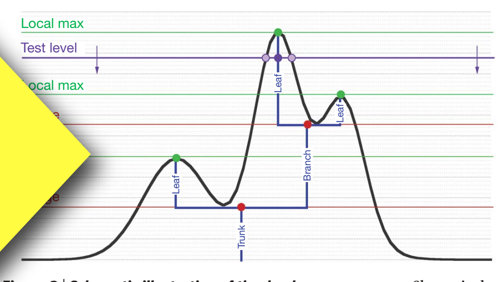
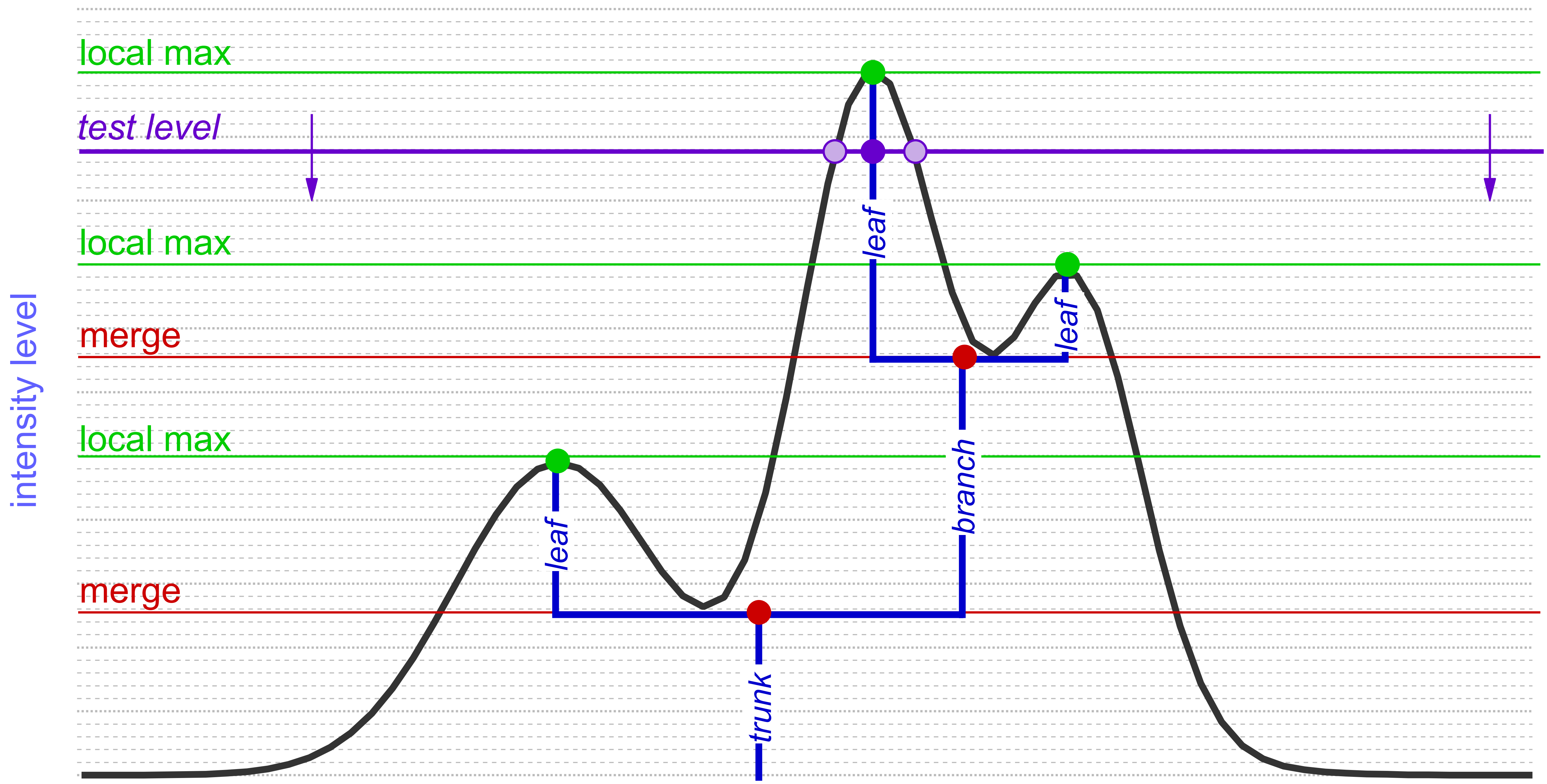


Figure 3 | Schematic illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (black). The dendrogram (blue) can be constructed by 'dropping' a test constant emission level (purple) from above in tiny steps (exaggerated in size here, light lines) until all the local maxima and mergers are found, and connected as shown. The intersection of a test level with the emission is a set of points (for example the light purple dots) in one dimension, a planar curve in two dimensions, and an isosurface in three dimensions. The dendrogram of 3D data shown in Fig. 2c is the direct analogue of the tree shown here, only constructed from 'isosurface' rather than 'point' intersections. It has been sorted and flattened for representation on a flat page, as fully representing dendrograms for 3D data cubes would require four dimensions.

Goodman et al. *Nature*, 2009

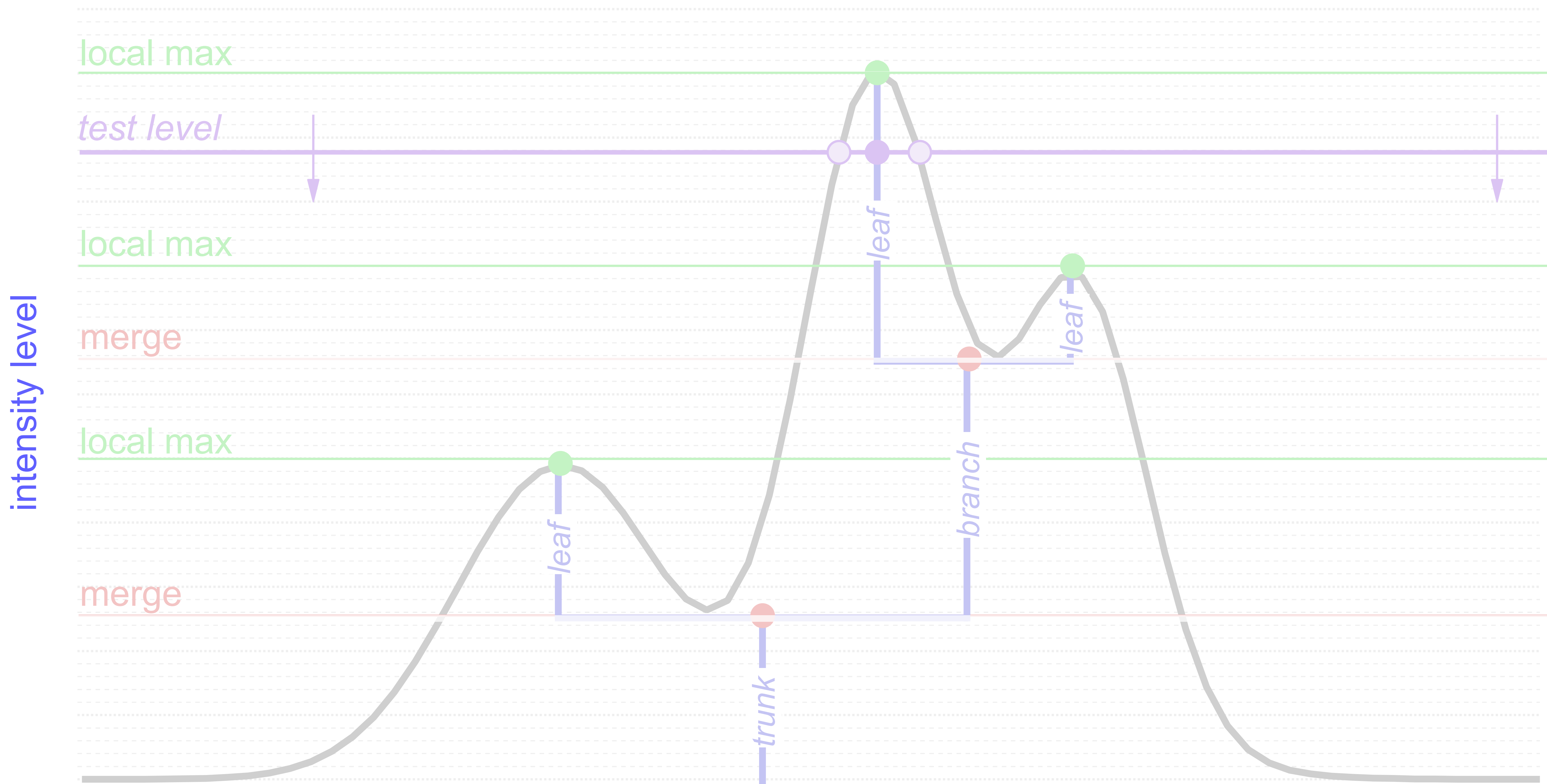
Dendrograms



Hierarchical "Segmentation"

Rosolowsky, Pineda, Kauffmann & Goodman 2008

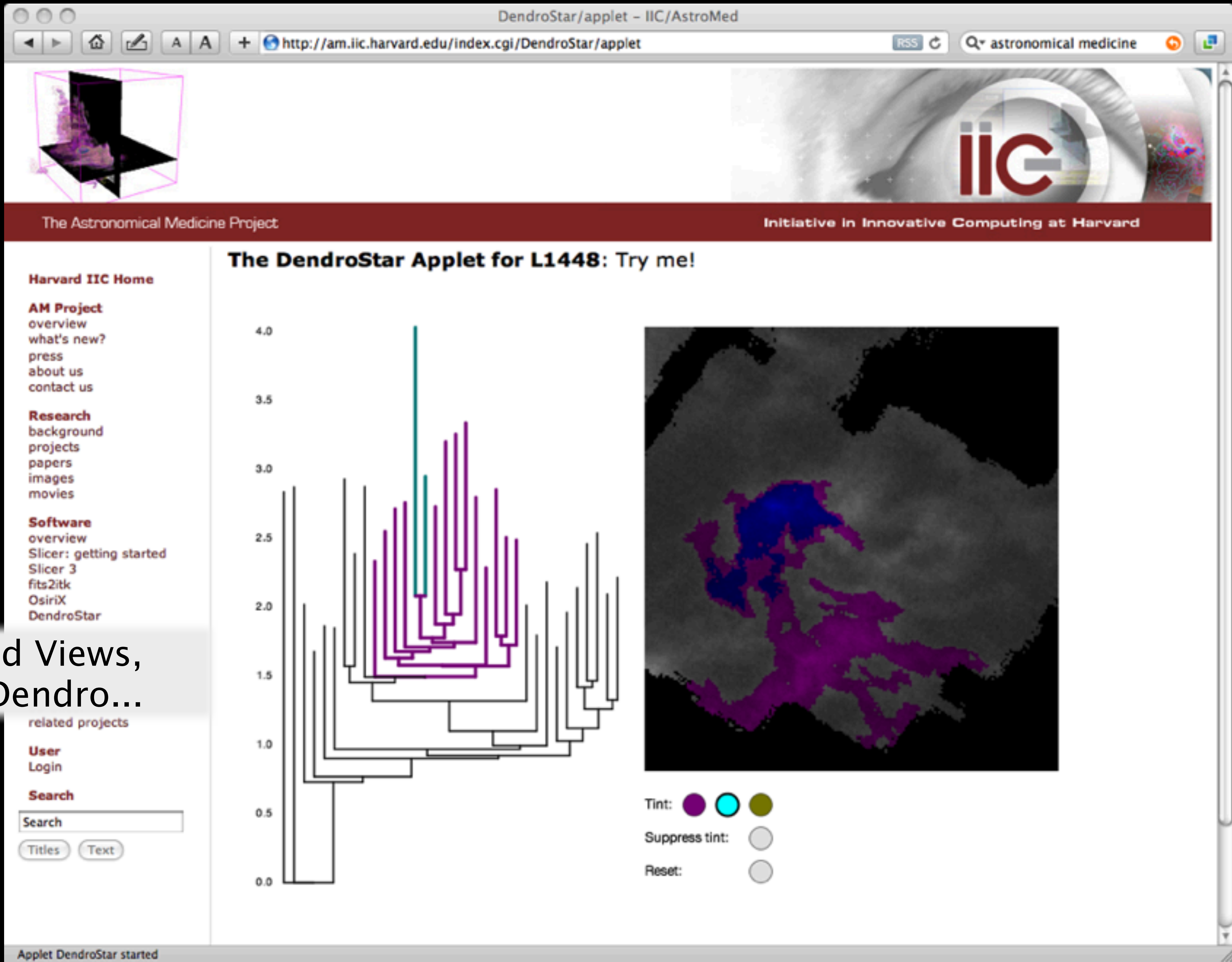
Dendrograms



1-D: points; 2-D closed curves (contours); 3-D surfaces enclosing volumes

see 2D demo at *http://am.iic.harvard.edu/index.cgi/DendroStar/applet*

Linked Views, e.g. Dendro...



http://am.iic.harvard.edu/index.cgi/DendroStar/applet

Dendrogram Algorithm by Erik Rosolwosky; Applet by Douglas Alan

Taste-Testing "Gravity"

3D PDF

True 3D Structure

What's bound?/ Virial Theorem

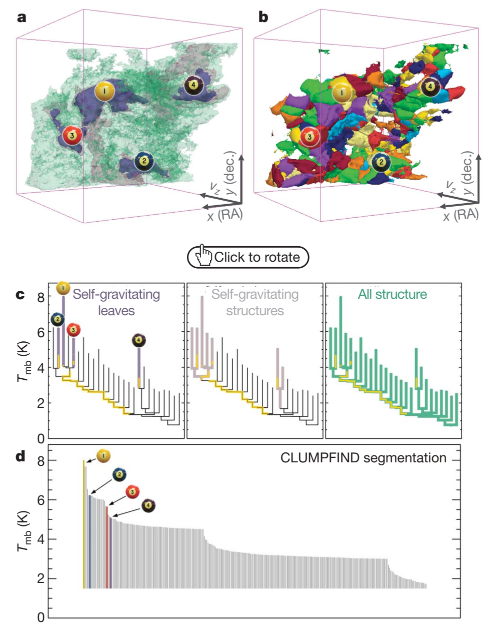


Figure 2 | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to ^{13}CO emission from the L1448 region of Perseus. a, 3D visualization of the surfaces indicated by colours in the dendrogram shown in **c**. Purple illustrates the smallest scale self-gravitating structures in the region corresponding to the leaves of the dendrogram; pink shows the smallest surfaces that contain distinct self-gravitating leaves within them; and green corresponds to the surface in the data cube containing all the significant emission. Dendrogram branches corresponding to self-gravitating objects have been highlighted in yellow over the range of T_{mb} (main-beam temperature) test-level values for which the virial parameter is less than 2. The *x–y* locations of the four 'self-gravitating' leaves labelled with billiard balls are the same as those shown in Fig. 1. The 3D visualizations show position–position–velocity (*p–p–v*) space. RA, right ascension; dec., declination. For comparison with the ability of dendrograms (**c**) to track hierarchical structure, **d** shows a pseudo-dendrogram of the CLUMPFIND segmentation (**b**), with the same four labels used in Fig. 1 and in **a**. As 'clumps' are not allowed to belong to larger structures, each pseudo-branch in **d** is simply a series of lines connecting the maximum emission value in each clump to the threshold value. A very large number of clumps appears in **b** because of the sensitivity of CLUMPFIND to noise and small-scale structure in the data. In the online PDF version, the 3D cubes (**a** and **b**) can be rotated to any orientation, and surfaces can be turned on and off (interaction requires Adobe Acrobat version 7.0.8 or higher). In the printed version, the front face of each 3D cube (the 'home' view in the interactive online version) corresponds exactly to the patch of sky shown in Fig. 1, and velocity with respect to the Local Standard of Rest increases from front ($-0.5\ km\ s^{-1}$) to back ($8\ km\ s^{-1}$).

data, CLUMPFIND typically finds features on a limited range of scales, above but close to the physical resolution of the data, and its results can be overly dependent on input parameters. By tuning CLUMPFIND's two free parameters, the same molecular-line data set[8] can be used to show either that the frequency distribution of clump mass is the same as the initial mass function of stars or that it follows the much shallower mass function associated with large-scale molecular clouds (Supplementary Fig. 1).

Four years before the advent of CLUMPFIND, 'structure trees'[9] were proposed as a way to characterize clouds' hierarchical structure using 2D maps of column density. With this early 2D work as inspiration, we have developed a structure-identification algorithm that abstracts the hierarchical structure of a 3D (*p–p–v*) data cube into an easily visualized representation called a 'dendrogram'[10]. Although well developed in other data-intensive fields[11,12], it is curious that the application of tree methodologies so far in astrophysics has been rare, and almost exclusively within the area of galaxy evolution, where 'merger trees' are being used with increasing frequency[13].

Figure 3 and its legend explain the construction of dendrograms schematically. The dendrogram quantifies how and where local maxima of emission merge with each other, and its implementation is explained in Supplementary Methods. Critically, the dendrogram is determined almost entirely by the data itself, and it has negligible sensitivity to algorithm parameters. To make graphical presentation possible on paper and 2D screens, we 'flatten' the dendrograms of 3D data (see Fig. 3 and its legend), by sorting their 'branches' to not cross, which eliminates dimensional information on the *x* axis while preserving all information about connectivity and hierarchy. Numbered 'billiard ball' labels in the figures let the reader match features between a 2D map (Fig. 1), an interactive 3D map (Fig. 2a online) and a sorted dendrogram (Fig. 2c).

A dendrogram of a spectral-line data cube allows for the estimation of key physical properties associated with volumes bounded by isosurfaces, such as radius (R), velocity dispersion (σ_v) and luminosity (L). The volumes can have any shape, and in other work[14] we focus on the significance of the especially elongated features seen in L1448 (Fig. 2a). The luminosity is an approximate proxy for mass, such that $M_{lum} = X_{13CO} L_{13CO}$, where $X_{13CO} = 8.0 \times 10^{20}\ cm^2\ K^{-1}\ km^{-1}\ s$ (ref. 15; see Supplementary Methods and Supplementary Fig. 2). The derived values for size, mass and velocity dispersion can then be used to estimate the role of self-gravity at each point in the hierarchy, via calculation of an 'observed' virial parameter, $\alpha_{obs} = 5\sigma_v^2 R/GM_{lum}$. In principle, extended portions of the tree (Fig. 2, yellow highlighting) where $\alpha_{obs} < 2$ (where gravitational energy is comparable to or larger than kinetic energy) correspond to regions of *p–p–v* space where self-gravity is significant. As α_{obs} only represents the ratio of kinetic energy to gravitational energy at one point in time, and does not explicitly capture external over-pressure and/or magnetic fields[16], its measured value should only be used as a guide to the longevity (boundedness) of any particular feature.

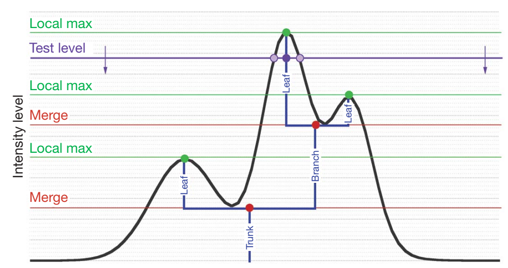
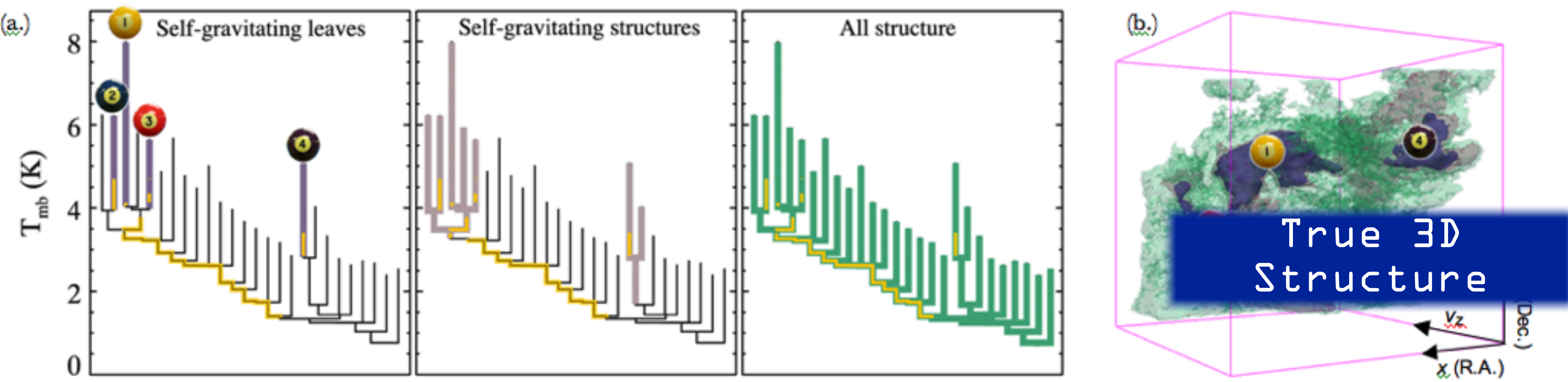


Figure 3 | Schematic illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (black). The dendrogram (blue) can be constructed by 'dropping' a test constant emission level (purple) from above in tiny steps (exaggerated in size here, light lines) until all the local maxima and mergers are found, and connected as shown. The intersection of a test level with the emission is a set of points (for example the light purple dots) in one dimension, a planar curve in two dimensions, and an isosurface in three dimensions. The dendrogram of 3D data shown in Fig. 2c is the direct analogue of the tree shown here, only constructed from 'isosurface' rather than 'point' intersections. It has been sorted and flattened for representation on a flat page, as fully representing dendrograms for 3D data cubes would require four dimensions.

Goodman et al. *Nature*, 2009

Tuesday, March 22, 2011

What's Bound? Can we Know?



What's bound?/ Virial Theorem

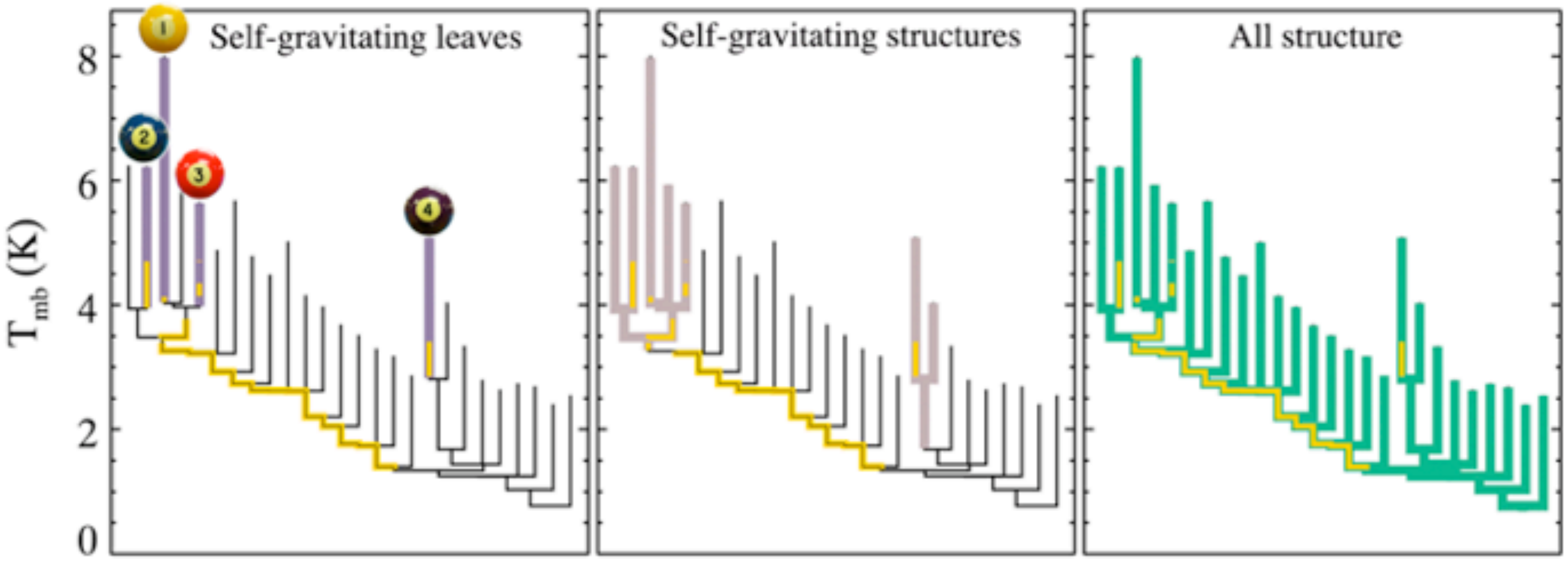
Yellow highlighting= "self-gravitating"

"Self-gravitating" here just means α_{vir} $(=5s_v^2 R/GM_{lum}) < 2$
(à la Bertoldi & McKee 1992–*BUT*–see Shetty et al. 2010)

Rosolowsky et al. 2008 (ApJ) &
Goodman et al. 2009 (Nature)

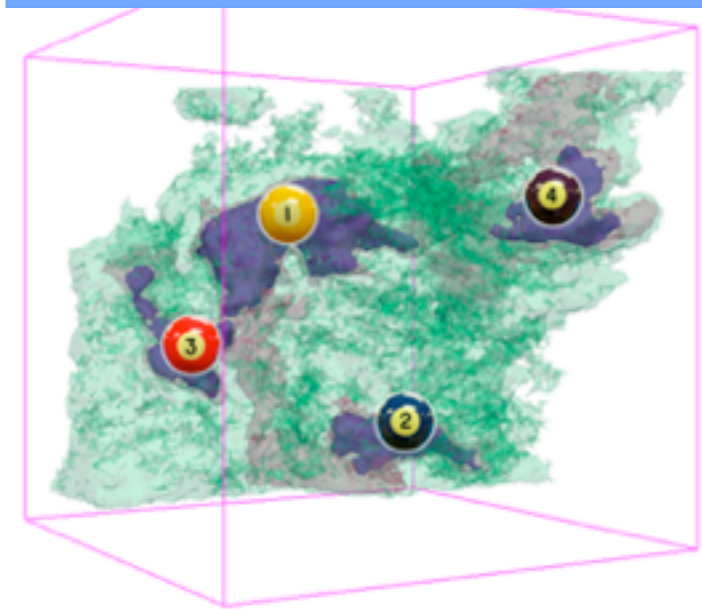
see PDF...

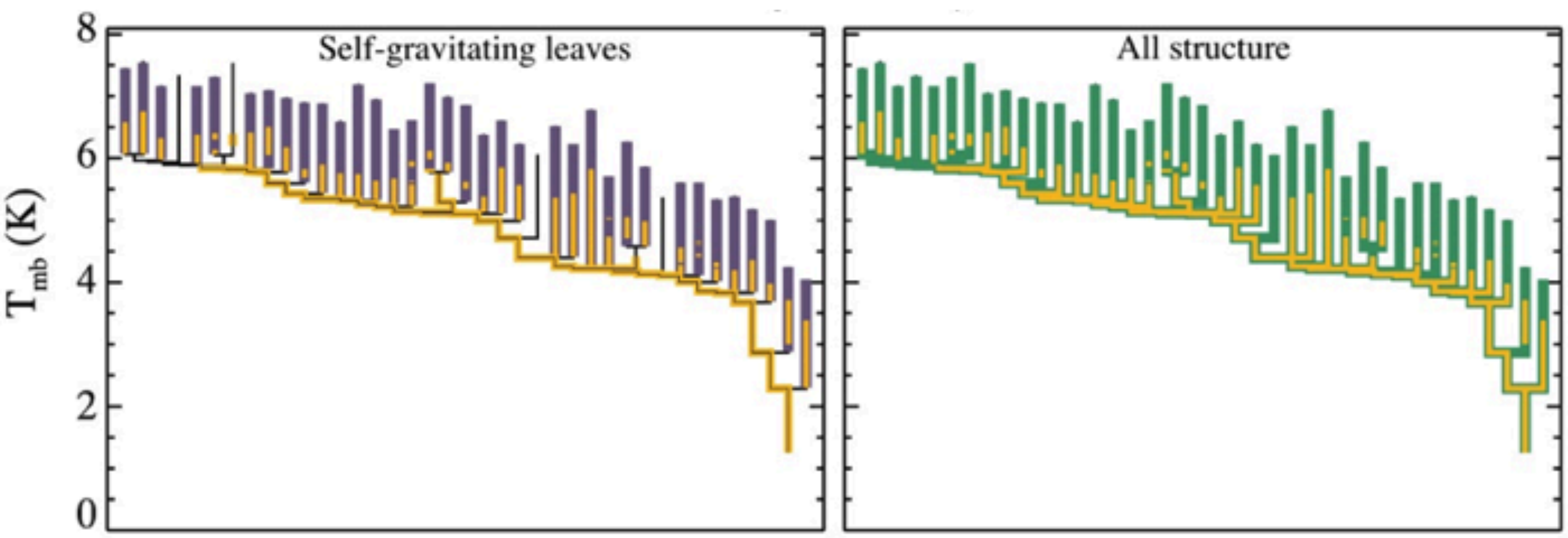
Real and Simulated ^{13}CO



(Yellow = self-gravitating components)

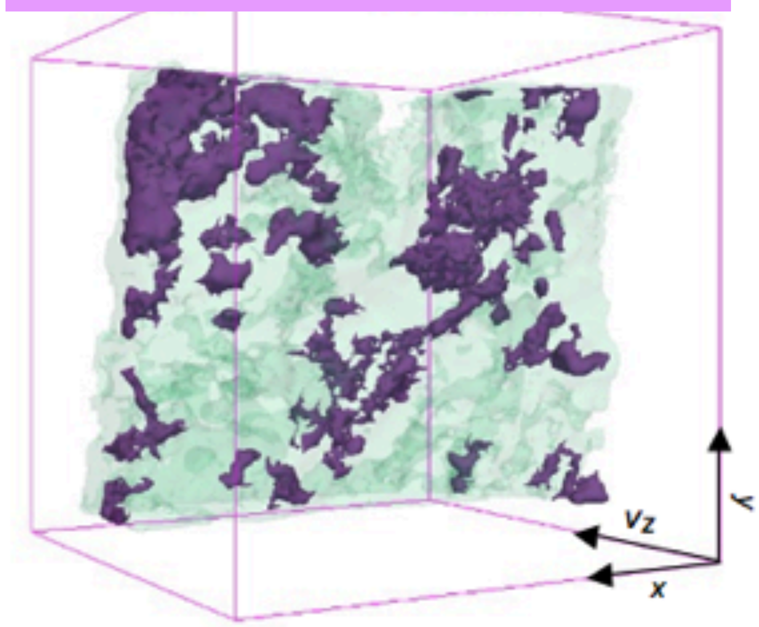
Real



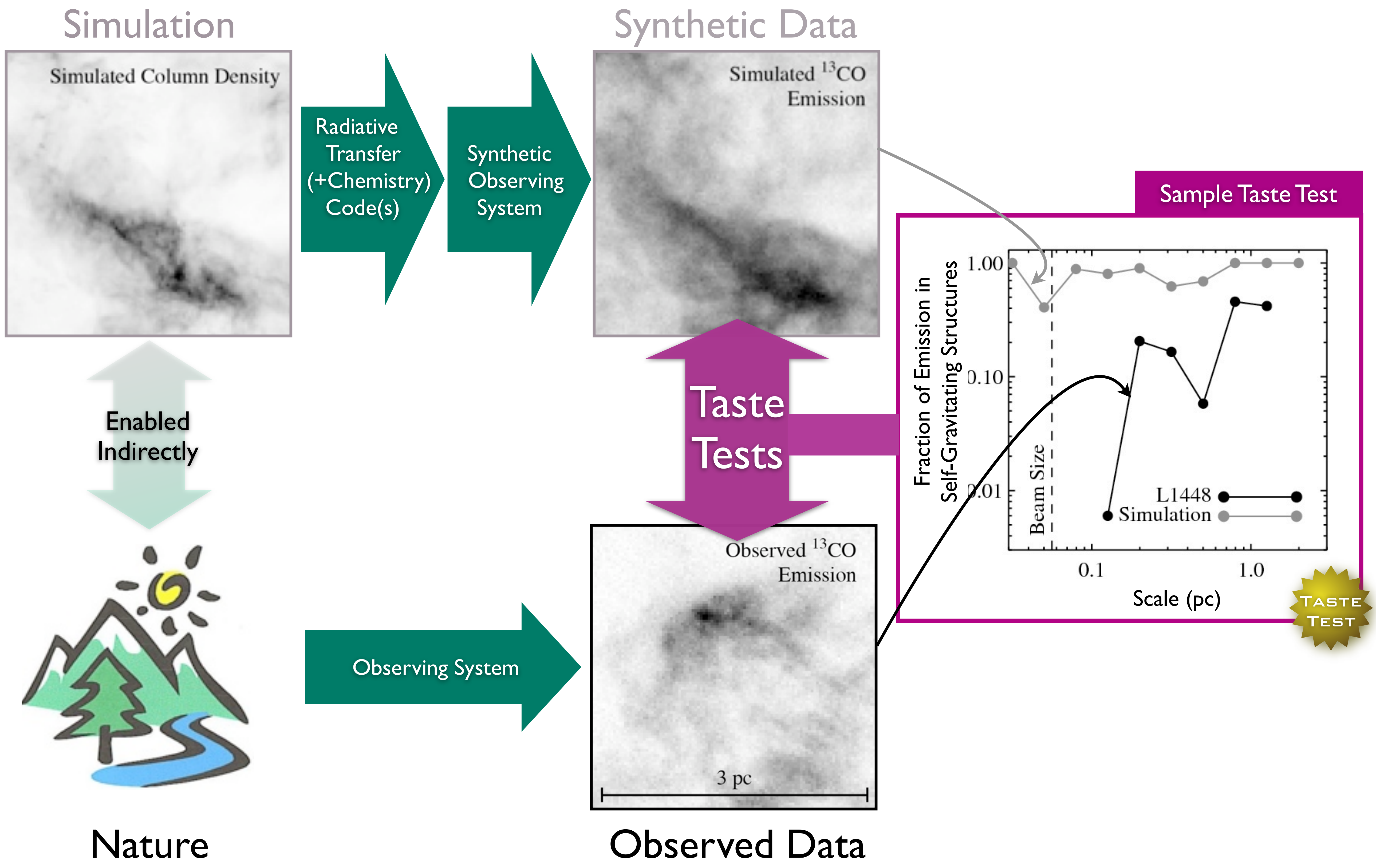


(Yellow = self-gravitating components)

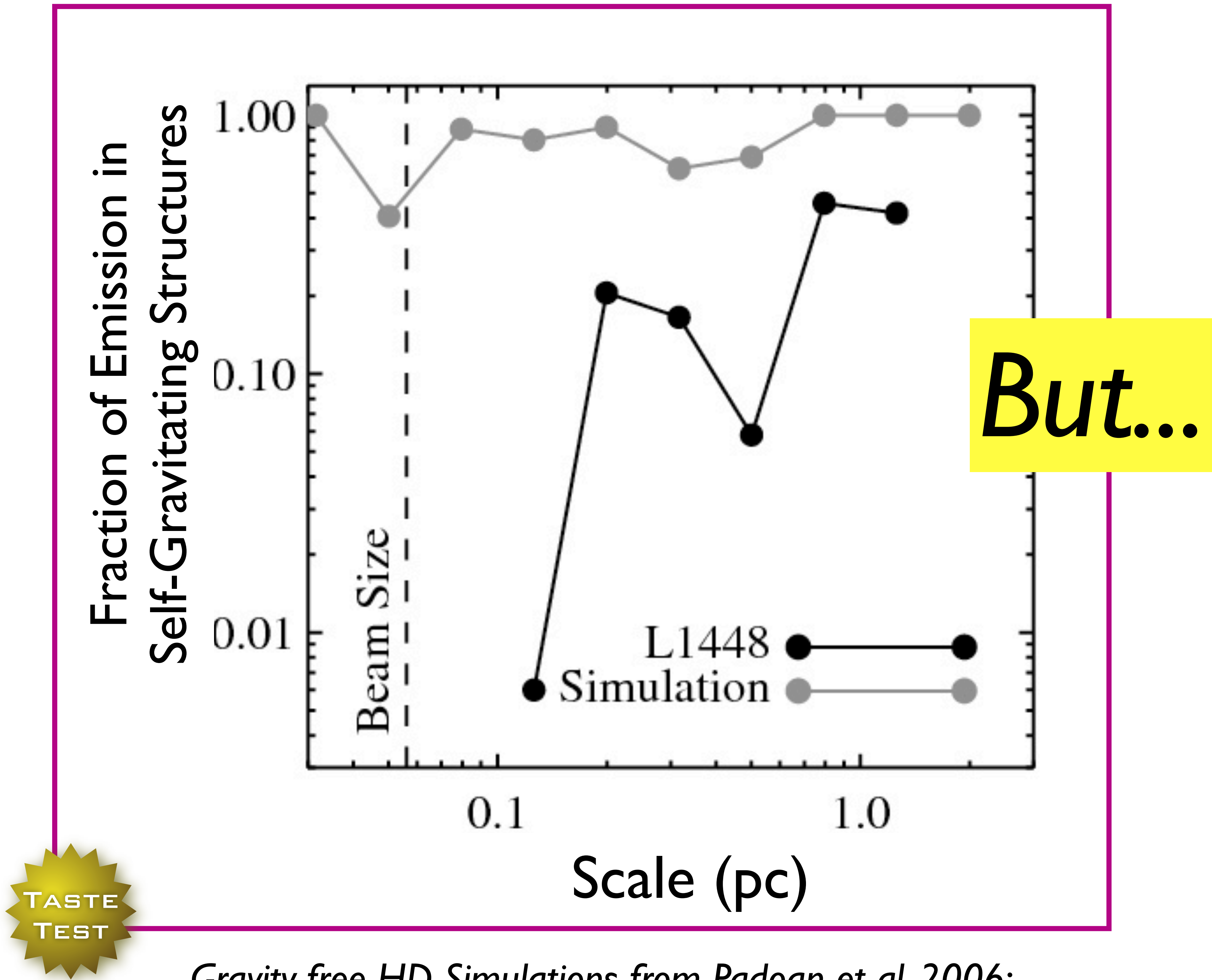
Simulated



The Taste-Testing Process

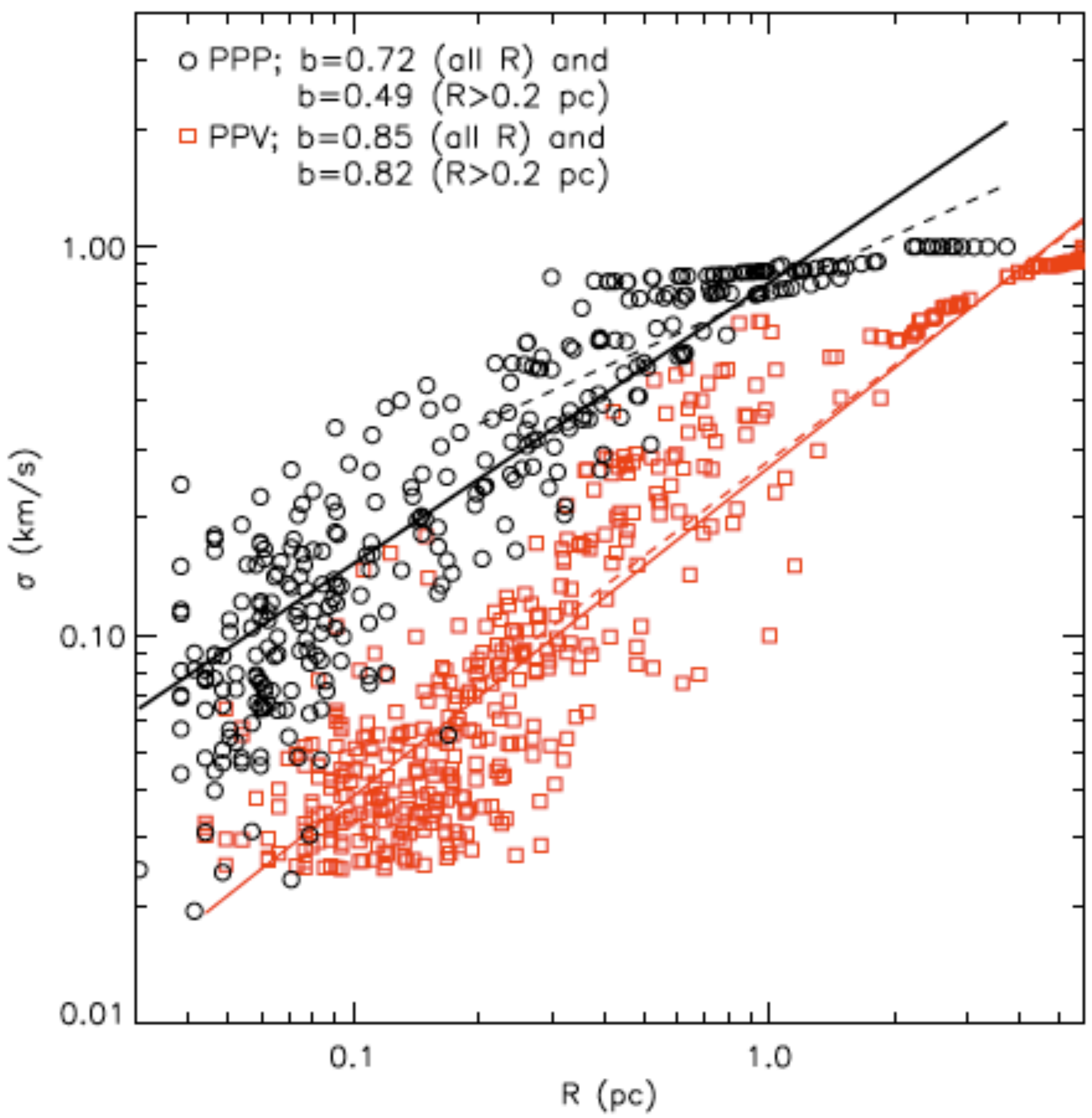


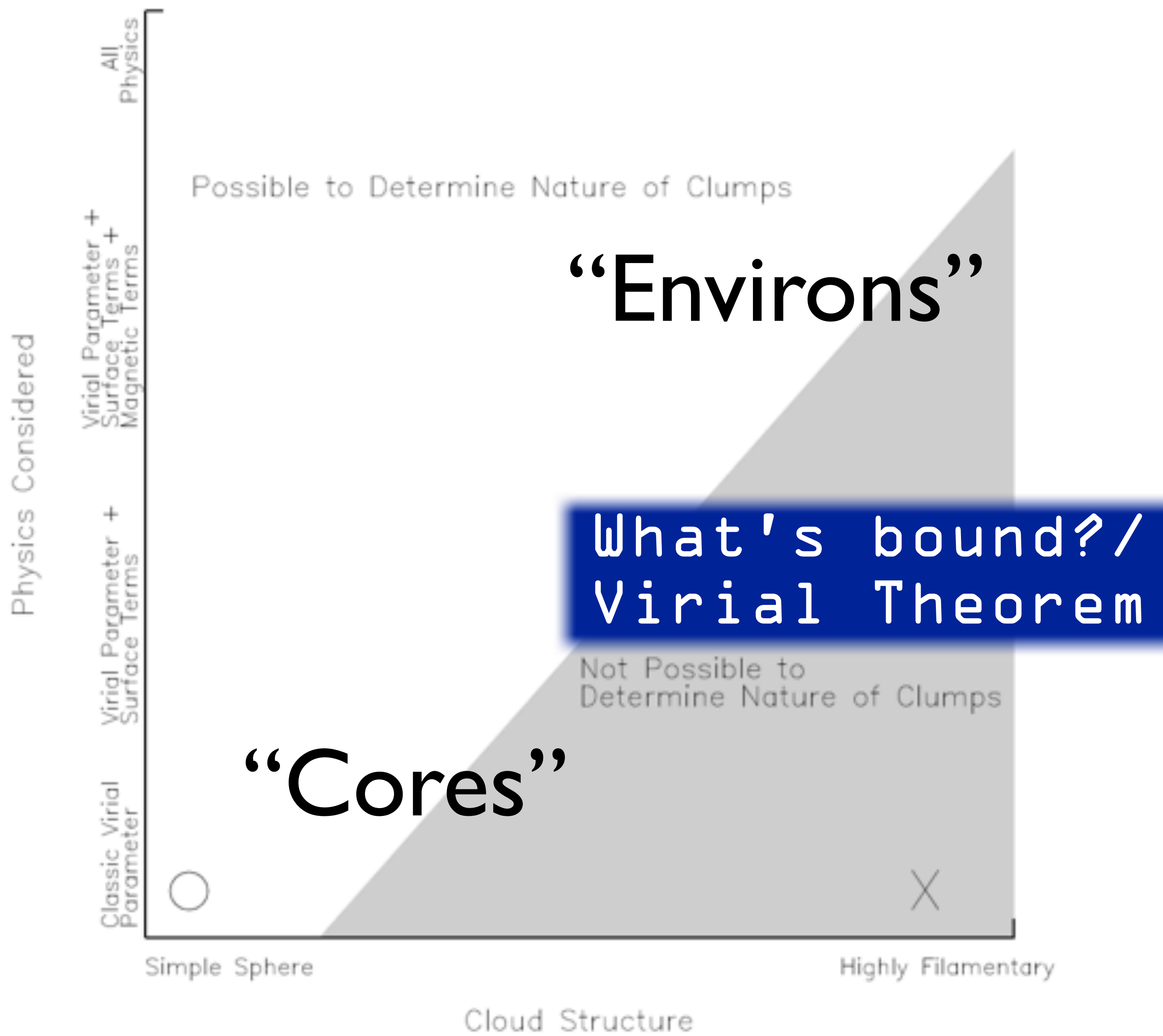
Taste-Testing "Gravity"



Gravity-free HD Simulations from Padoan et al. 2006;
L1448 analysis from Rosolowsky et al. 2008
both lines derived from ^{13}CO "observations"

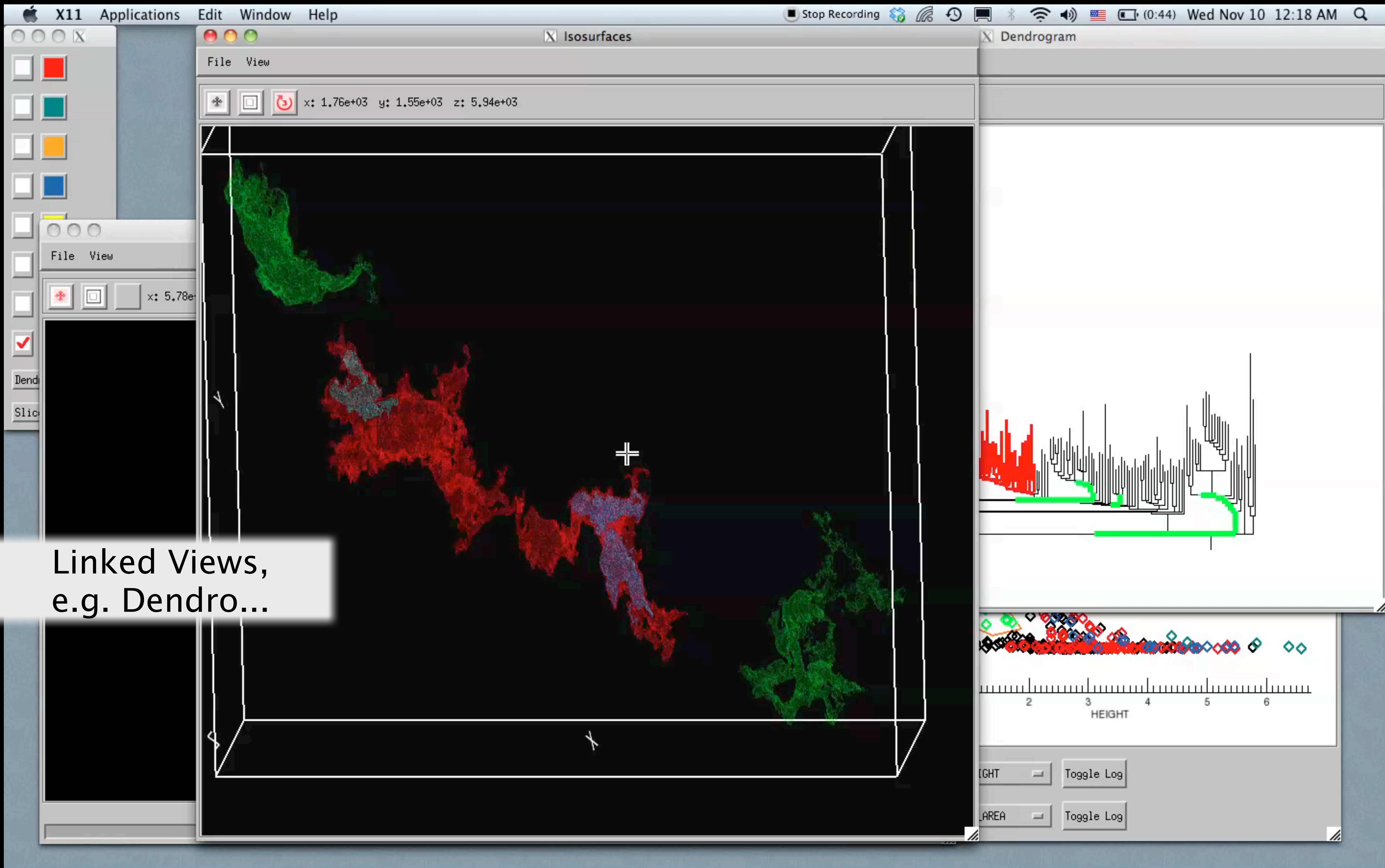
But... *Caveats*/Worries about *p-p-v* (bijection) ... and the virial parameter





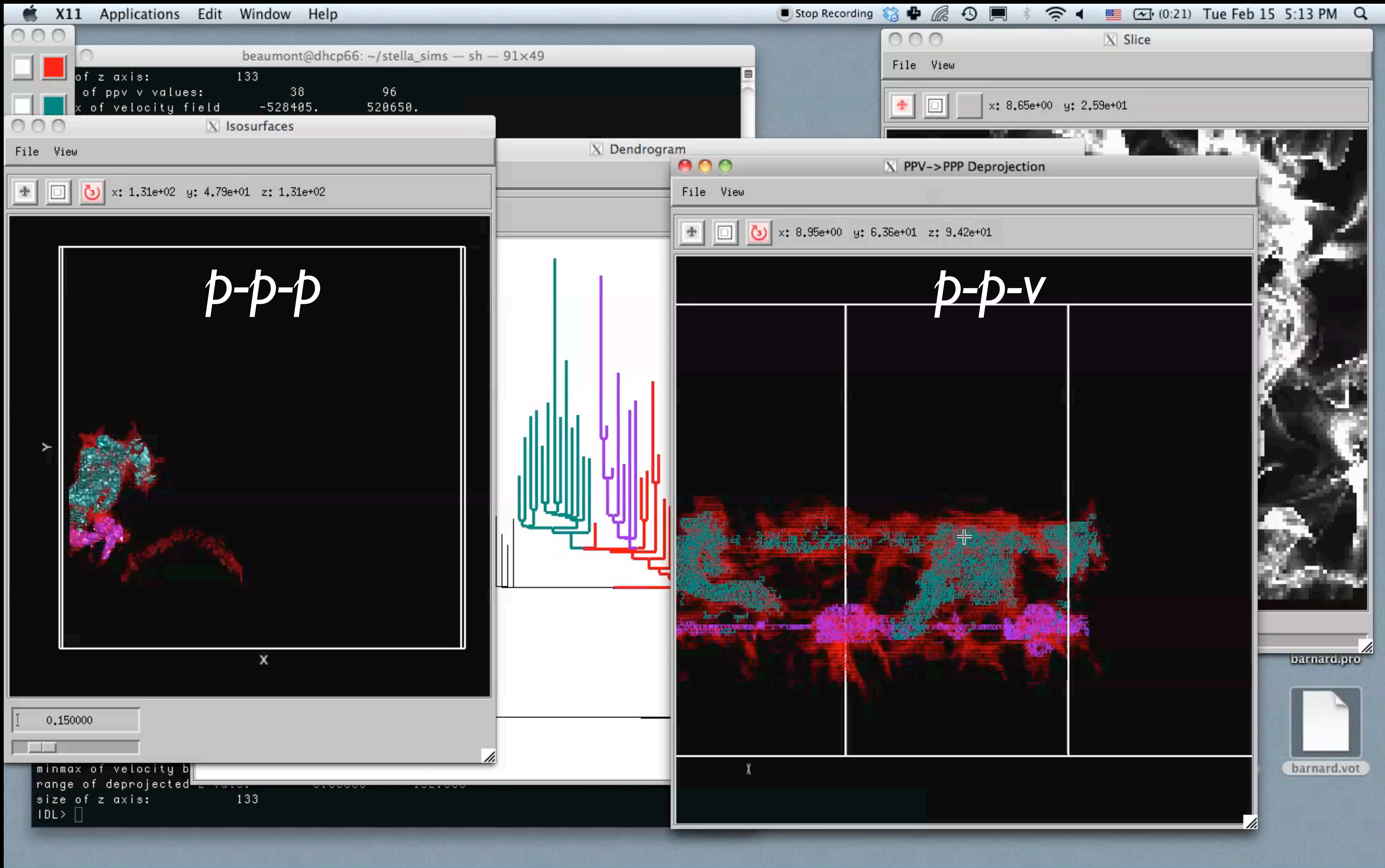
*from **Shetty**, Collins, Kauffmann, Goodman, Rosolowsky & M. Norman 2010;*
see also recent work of Dib et al., Ostriker et al., Ballesteros-Paredes et al., Myers, and Smith, Clark & Bonnell

Linked Dendrogram Views in IDL (I)

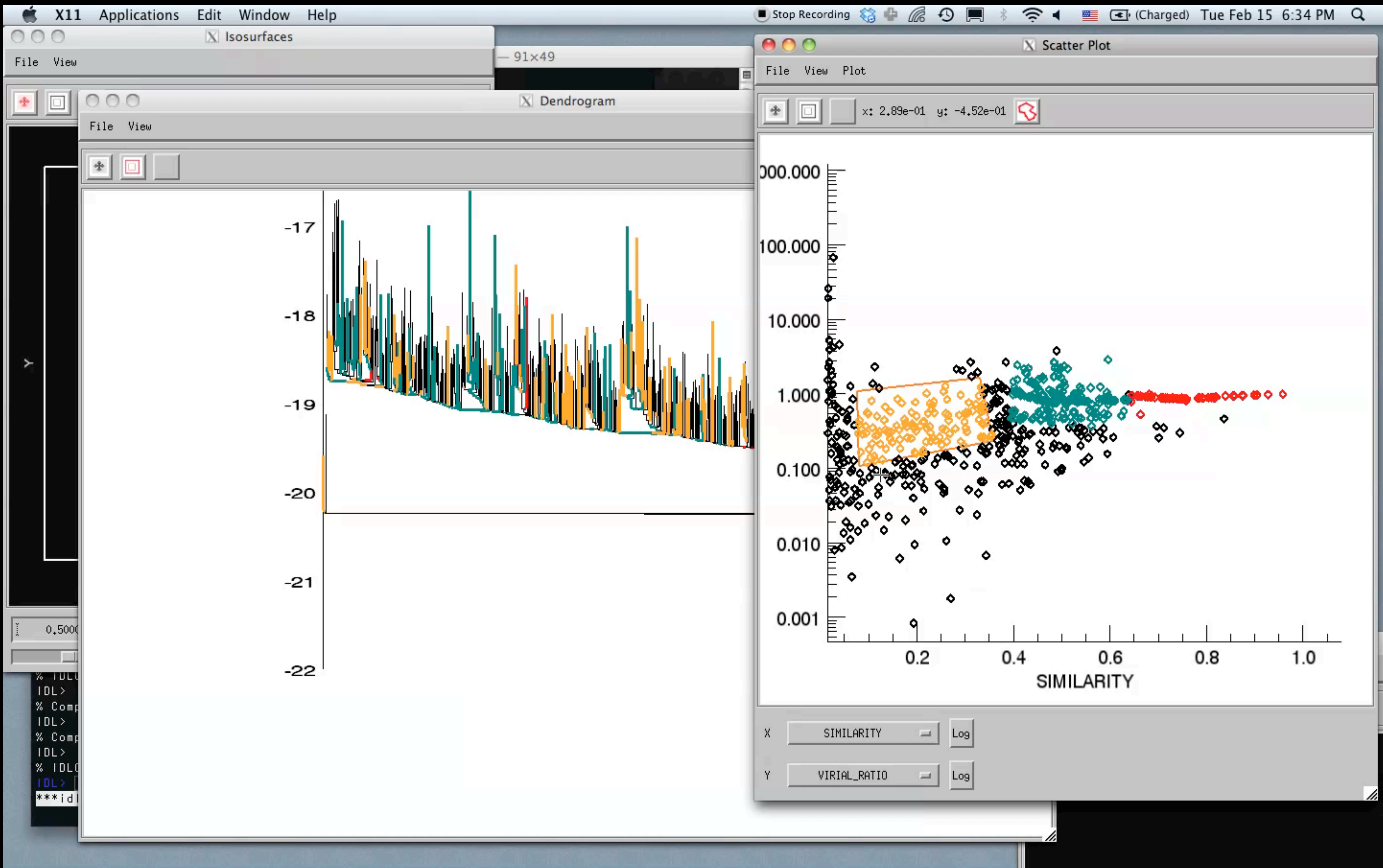


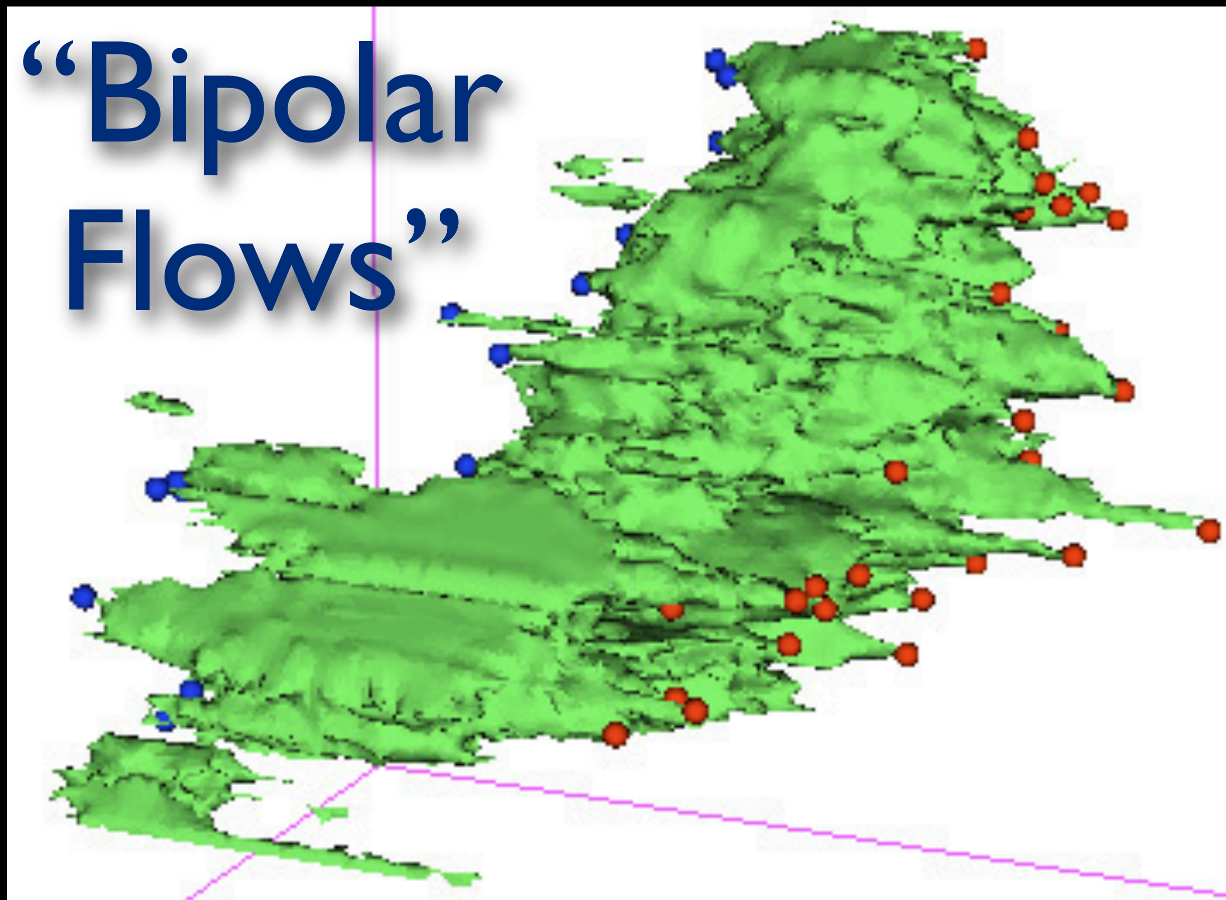
Video & implementation: Christopher Beaumont, CfA/UHawaii;
inspired by AstroMed work of Douglas Alan, Michelle Borkin, AG, Michael Halle, Erik Rosolowsky

Linked Dendrogram Views in IDL (2)



Linked Dendrogram Views in IDL (3)





The Astrophysical Journal, 715:1170–1190, 2010 June 1 doi:10.1088/0004-637X/715/2/1170

THE COMPLETE SURVEY OF OUTFLOWS IN PERSEUS

Héctor G. Arce[1], Michelle A. Borkin[2], Alyssa A. Goodman[3], Jaime E. Pineda[3], and Michael W. Halle[4,5]
[1] Department of Astronomy, Yale University, P.O. Box 208101, New Haven, CT 06520, USA; hector.arce@yale.edu
[2] School of Engineering and Applied Sciences, Harvard University, 29 Oxford Street, Cambridge, MA 02138, USA
[3] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
[4] Surgical Planning Laboratory, Department of Radiology, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, USA
[5] Initiative in Innovative Computing, Harvard University, 60 Oxford Street, Cambridge, MA 02138, USA
Received 2009 October 7; accepted 2010 April 9; published 2010 May 7

ABSTRACT

We present a study on the impact of molecular outflows in the Perseus molecular cloud complex using the COMPLETE Survey large-scale $^{12}CO(1–0)$ and $^{13}CO(1–0)$ maps. We used three-dimensional isosurface models generated in right ascension–declination–velocity space to visualize the maps. This rendering of the molecular line data allowed for a rapid and efficient way to search for molecular outflows over a large ($\sim$16 deg^2) area. Our outflow-searching technique detected previously known molecular outflows as well as new candidate outflows. Most of these new outflow-related high-velocity features lie in regions that have been poorly studied before. These new outflow candidates more than double the amount of outflow mass, momentum, and kinetic energy in the Perseus cloud complex. Our results indicate that outflows have significant impact on the environment immediately surrounding localized regions of active star formation, but lack the energy needed to feed the observed turbulence in the *entire* Perseus complex. This implies that other energy sources, in addition to protostellar outflows, are responsible for turbulence on a global cloud scale in Perseus. We studied the impact of outflows in six regions with active star formation within Perseus of sizes in the range of 1–4 pc. We find that outflows have enough power to maintain the turbulence in these regions and enough momentum to disperse and unbind some mass from them. We found no correlation between outflow strength and star formation efficiency (SFE) for the six different regions we studied, contrary to results of recent numerical simulations. The low fraction of gas that potentially could be ejected due to outflows suggests that additional mechanisms other than cloud dispersal by outflows are needed to explain low SFEs in clusters.

Key words: ISM: clouds – ISM: individual objects (Perseus) – ISM: jets and outflows – ISM: kinematics and dynamics – stars: formation – turbulence

Online-only material: color figures

Arce et al. 2010, 2011

COMPLETE Shells in Perseus

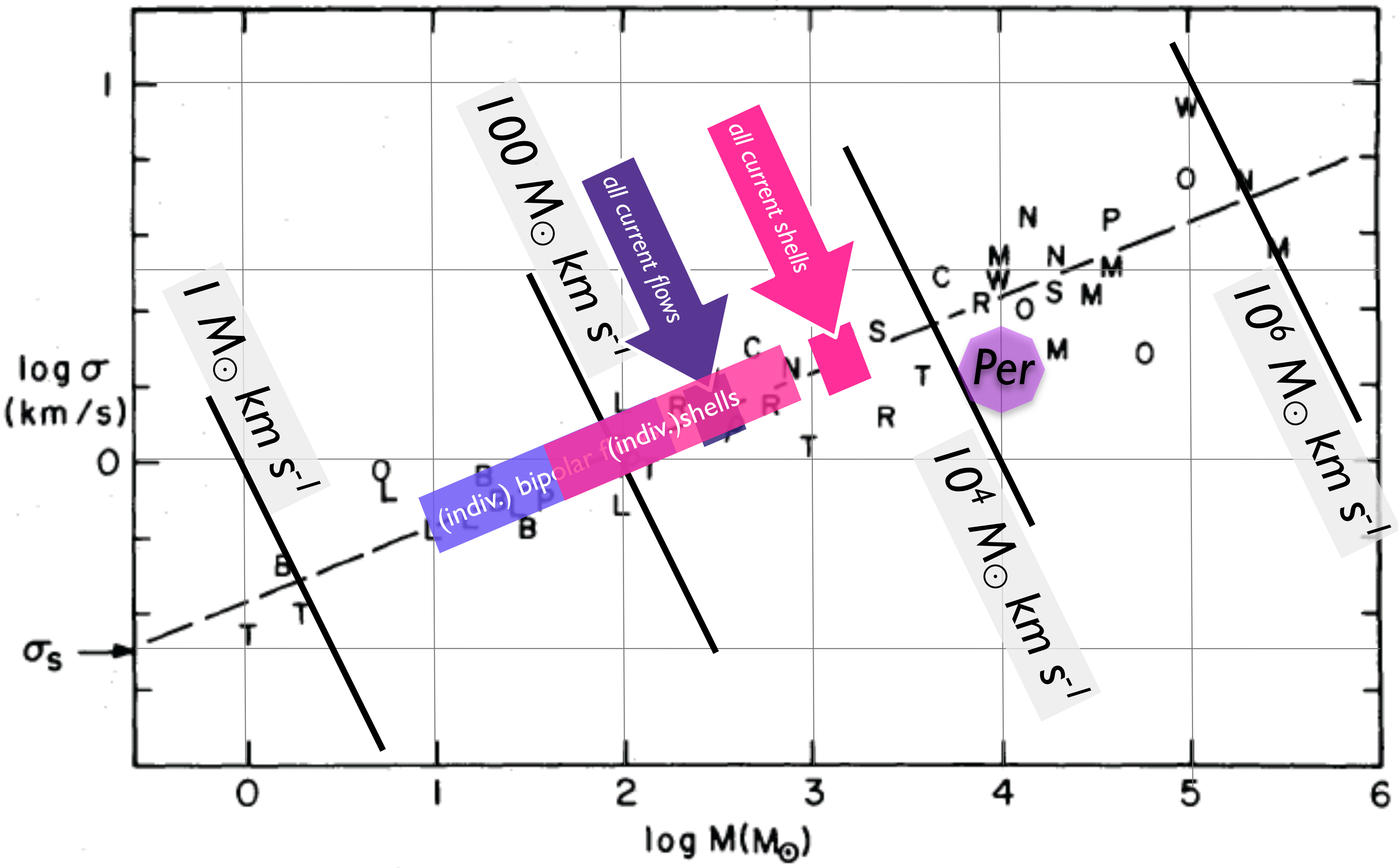
ABSTRACT

We present a study on the shells in the Perseus molecular cloud using the COMPLETE survey large-scale $^{12}CO(1-0)$ and $^{13}CO(1-0)$ maps. The shells are spread throughout most of the Perseus cloud and have circular or arc-like morphologies with a range in radius of about 0.2 to 3 pc. Most of the CO shells are coincident with near-infrared nebulosity of similar shape and have a candidate powering source near the center. We suggest they are formed by the interaction of spherical or very wide-angle winds powered by young stars inside or near the Perseus molecular cloud complex —a cloud that is commonly considered a low-mass star forming region. It is clear that two of the twelve shells are powered by high-mass stars near the cloud, while the others appear to be powered by low or interm
winds with a mass loss rate of about 10
observed shells, which are clearly impac
age of the pre-main sequence stars in t
mass loss rate. Our estimates indicate t
that create the shells is similar to the tu
energy input from both collimated protostellar outflows and powerful spherical winds from young stars is sufficient to maintain the turbulence in the molecular cloud. Most of the shells had not been detected before, most likely as maps of the region lacked the coverage *and* resolution needed to distinguish the shells. Large scale molecular line and IR continuum maps of a sample of other clouds will help investigate the frequency of powerful shells from low-mass stars and the impact from stellar winds from nearby massive stars on low-mass star forming regions.

Subject headings: star: formation — ISM: jets and outflows — ISM: clouds — ISM: individual (Perseus) — ISM: kinematics and dynamics — turbulence

What riles up the ISM?

Arce, Beaumont, Borkin, Pineda, Goodman



Properties of Molecular Clouds as "Equivalent Momentum" (using Larson 1981)

grey boxes mark lines of constant "momentum," as labeled

What riles up the ISM?

What "upshifts" are justified?....

IOTW, how do we go from a "snapshot" to cumulative effects?

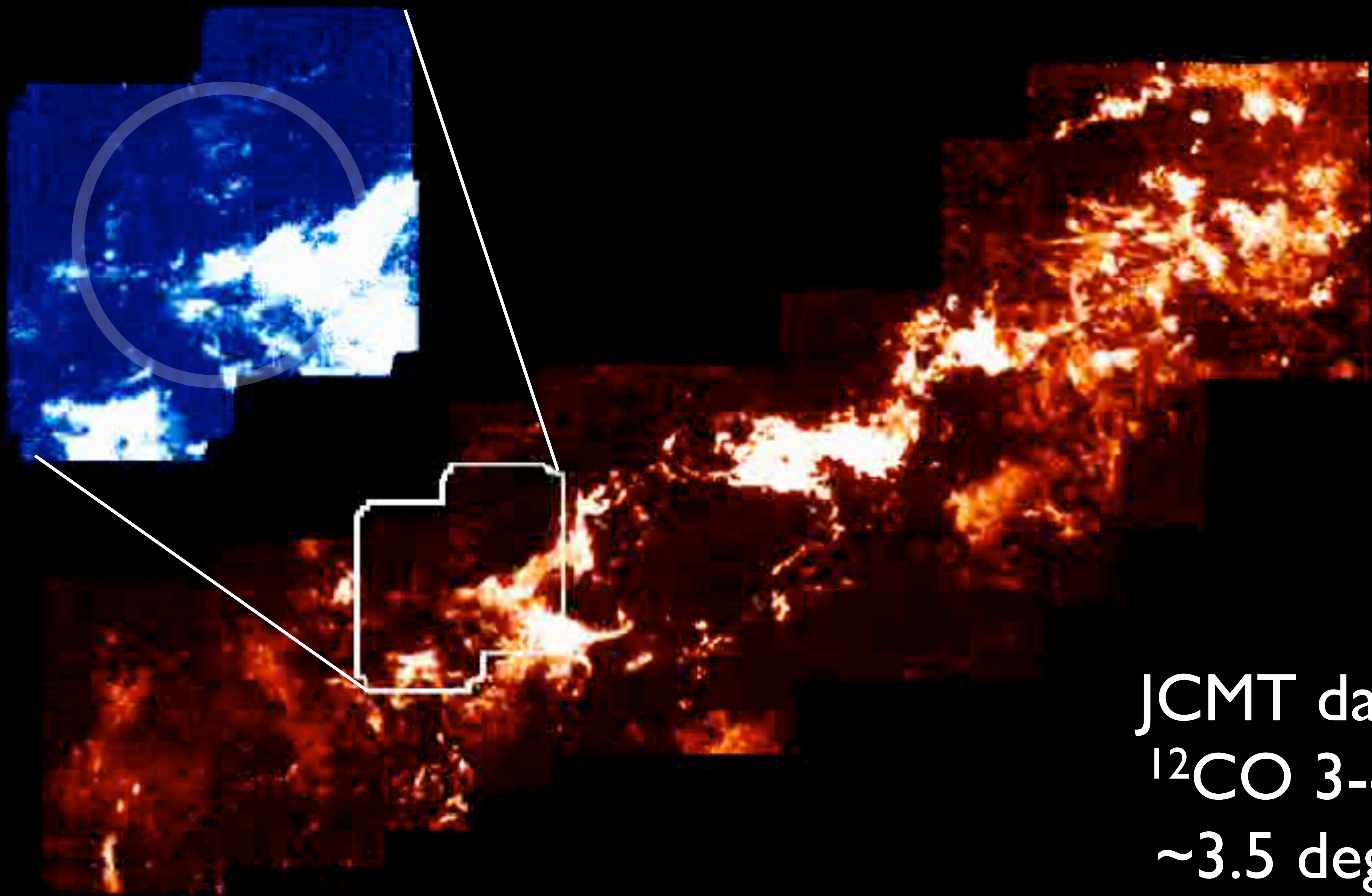
Note theory gives ~10 to 1000 $M_\odot$ km s^{-1} per B-star wind.


EN · The Milky Way Project is part of the ZOONIVERSE
...just like SOLAR STORMWATCH
THE MILKY WAY PROJECT
FOLLOW US ON TWITTER
VISIT THE BLOG
MILKY WAY TALK
HOME
TAKE PART
ABOUT
TUTORIAL
LOG IN
GALACTOMETER™
DRAW BUBBLES
WELCOME
The Milky Way Project aims to sort and measure our galaxy, the Milky Way. Initially we're asking you to help us find and draw bubbles in beautiful infrared data from the Spitzer Space Telescope.
Understanding the cold, dusty material that we see in these images, helps scientists to learn how stars form and how our galaxy anges and evolves with time.
ick here to see the full tutorial or browse e site to find out more about the science hind the Milky Way Project.
YOU CAN NOW SEE HOW CLOSE WE ARE TO 1,
/PROJECT.ORG/G... 12 DAYS AGO
194,943 IMAGES SERVED · 252,562 BUBBLES DRAWN ·
IATE GALAXIES · 597,054 OTHER OBJECTS
"Shells"
Machine Learning
What riles up the ISM?

(p-p-v) Case Study (Beaumont)

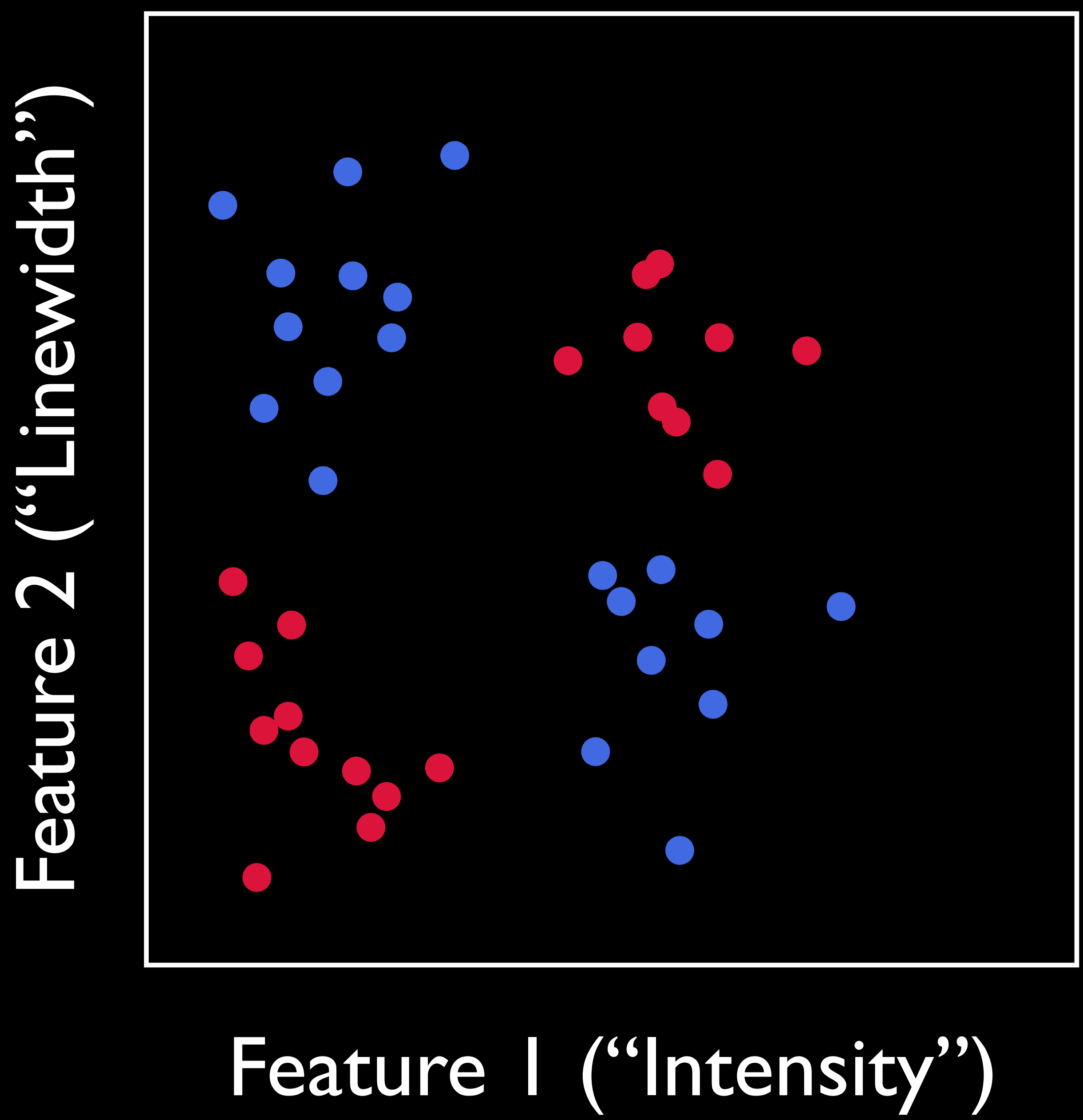
"Buried" SNR G16.05-0.57

All of **M17**

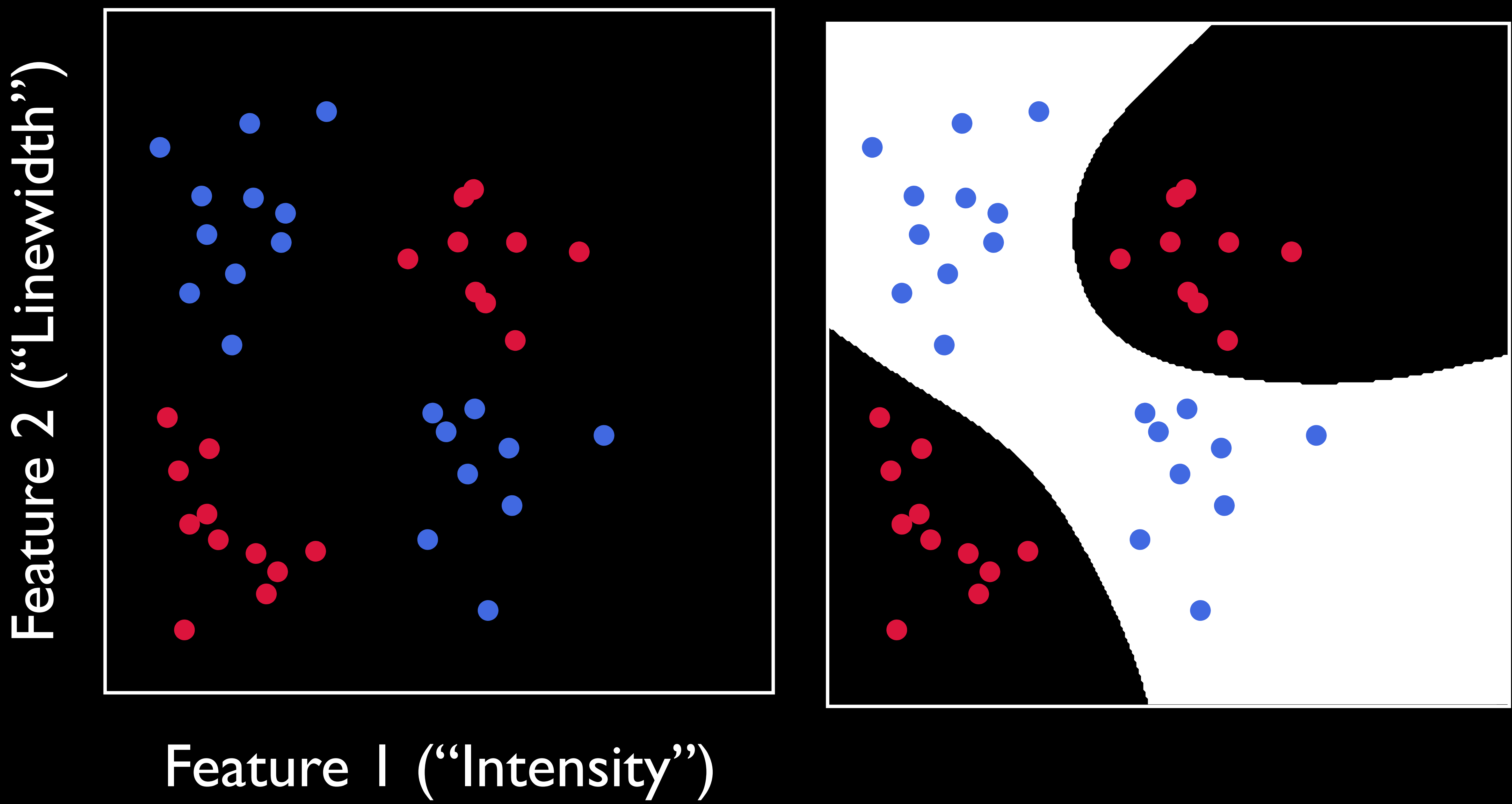


JCMT data
^{12}CO 3--2
~3.5 deg^2

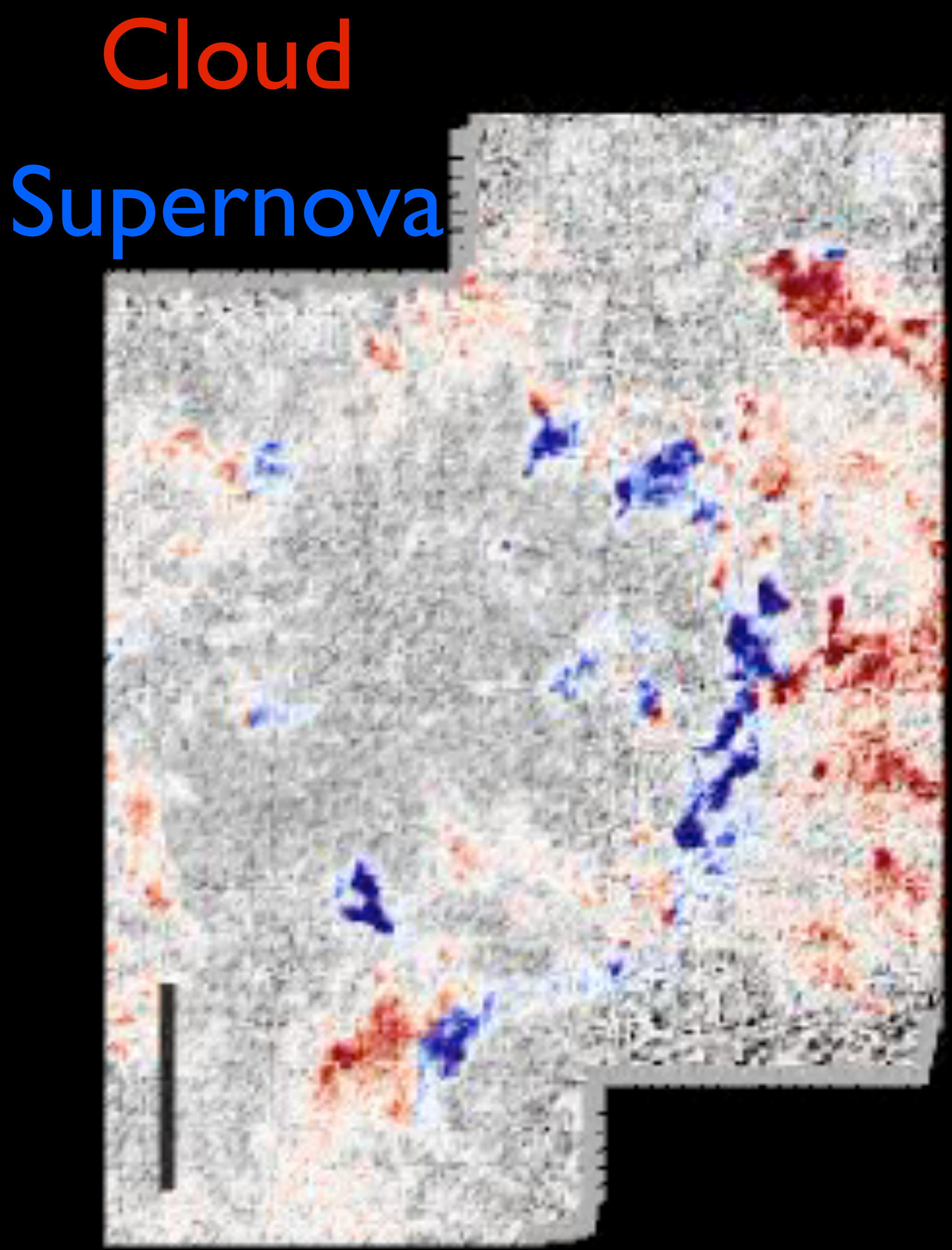
Support Vector Machines in One Minute
(SVM is a kind of "Machine Learning")

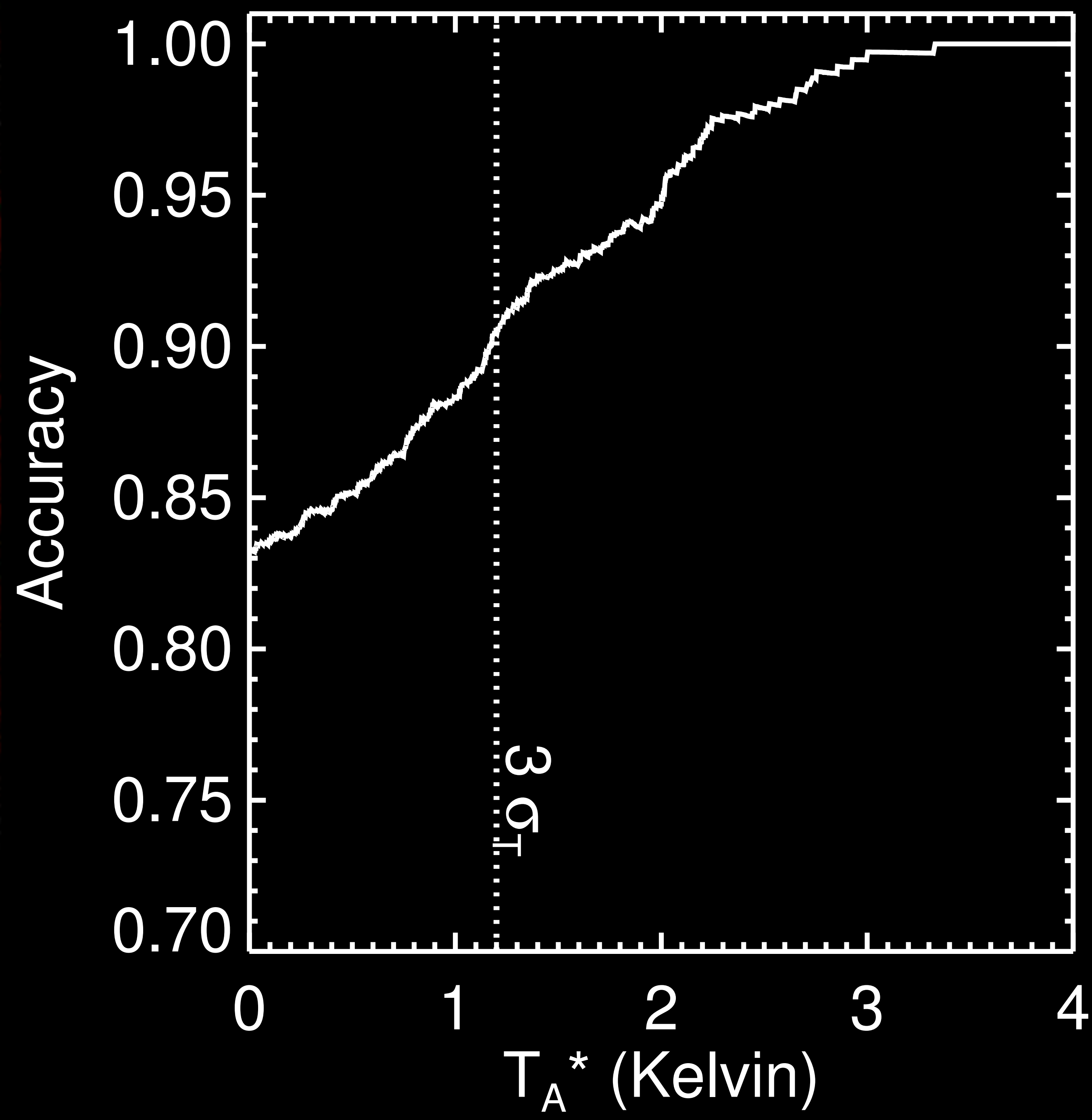


Support Vector Machines in One Minute



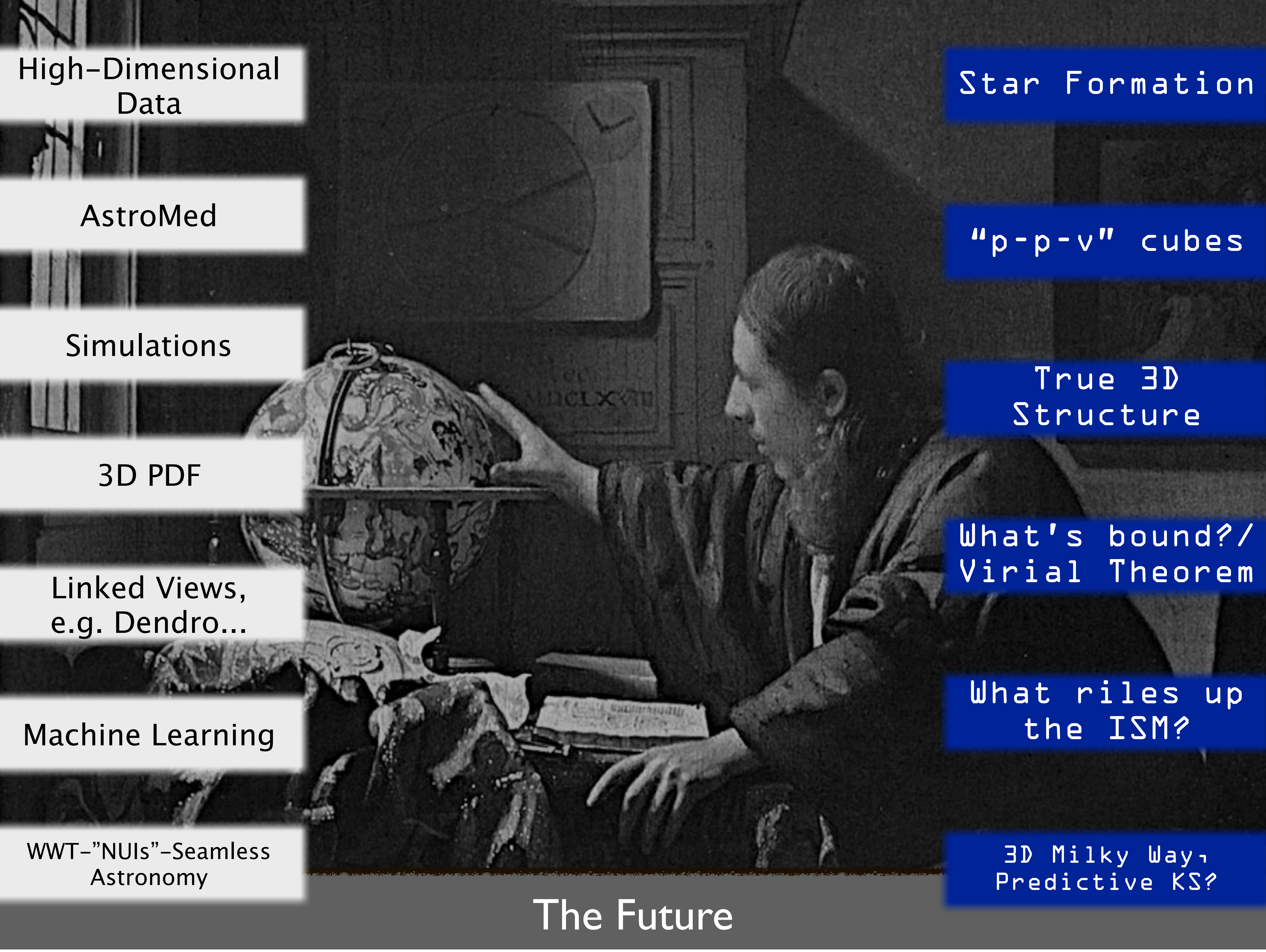
Results


Cloud
Supernova

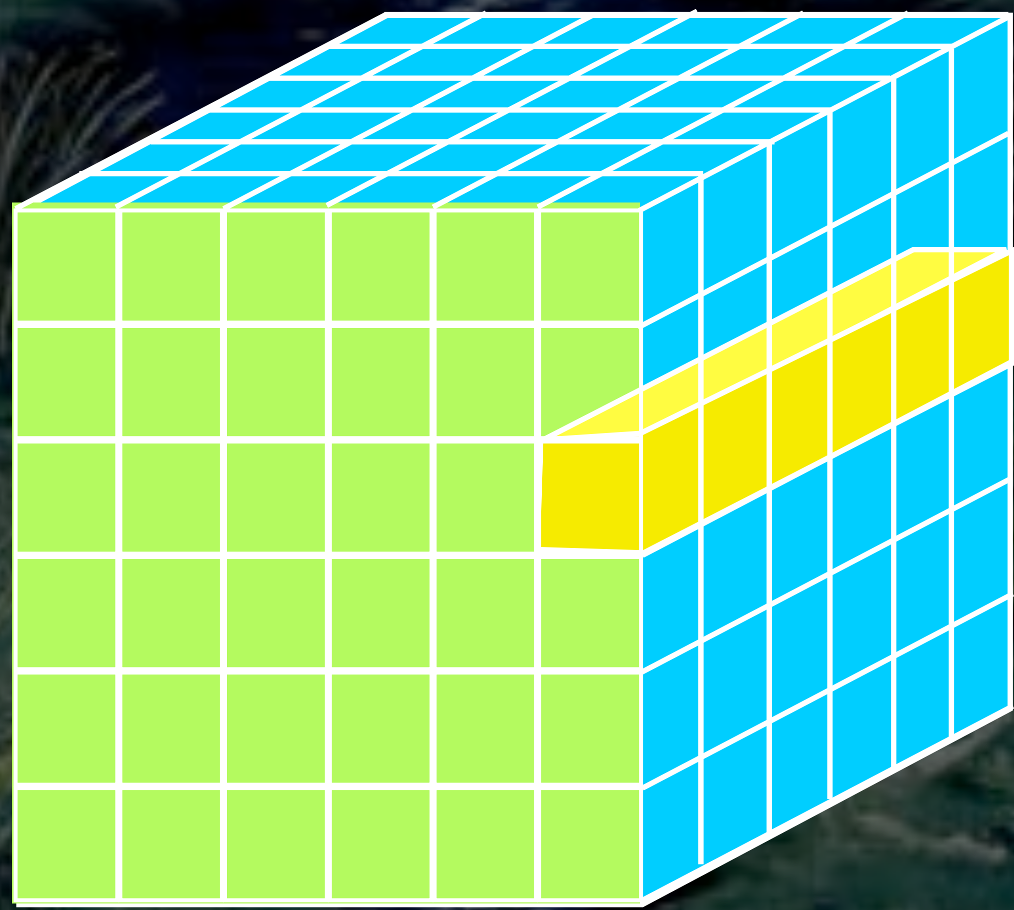

1.00
0.95
0.90
0.85
0.80
0.75
0.70
Accuracy
3 σ_T
0
1
2
3
4
T_A* (Kelvin)

Machine Learning

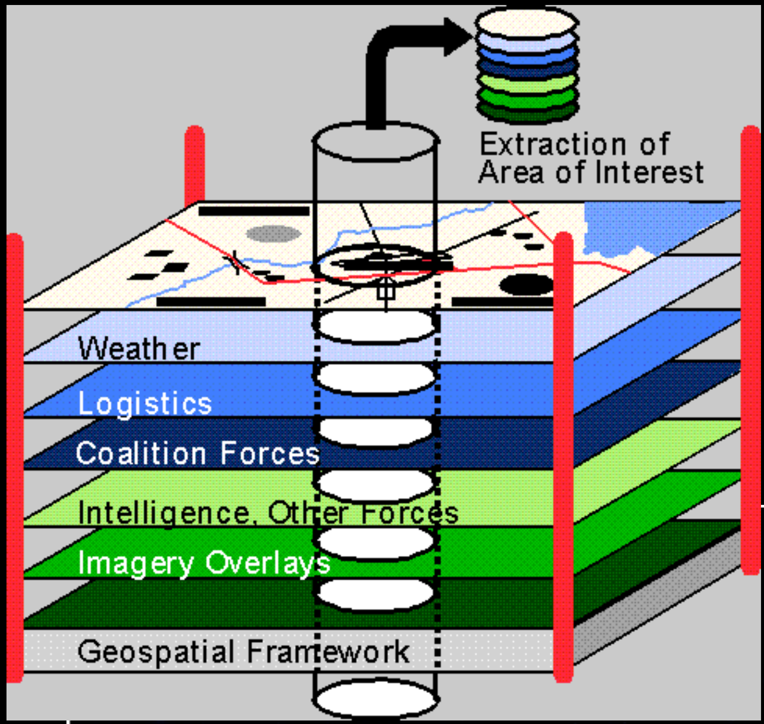
What riles up the ISM?


High-Dimensional Data
AstroMed
Simulations
3D PDF
Linked Views, e.g. Dendro...
Machine Learning
WWT-”NUIs”-Seamless Astronomy
Star Formation
“p-p-v” cubes
True 3D Structure
What's bound?/ Virial Theorem
What riles up the ISM?
3D Milky Way, Predictive KS?
The Future

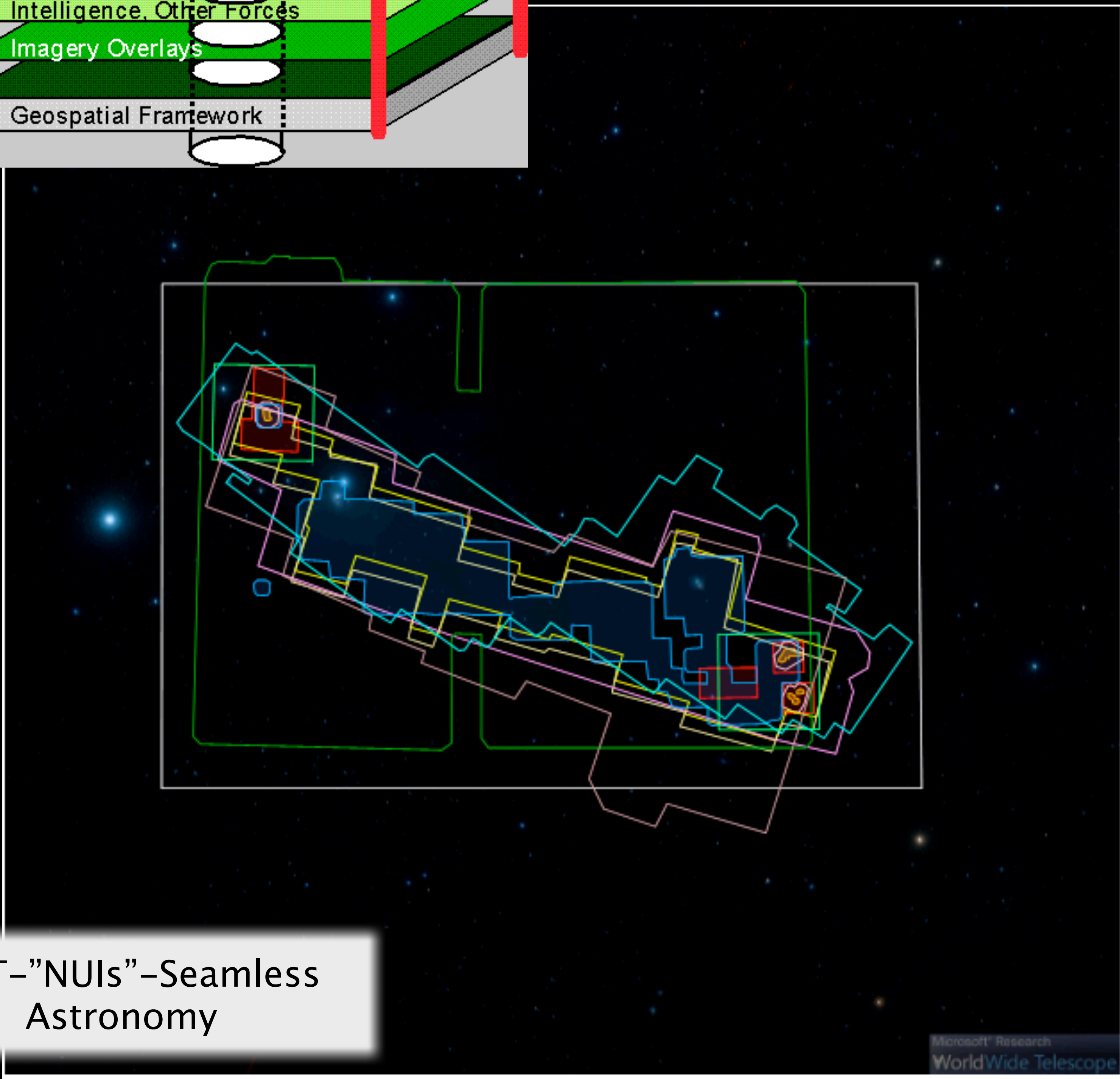
The dream scenario...



WWT-"NUIs"-Seamless Astronomy



COMPLETE



WWT-"NUIs"-Seamless Astronomy

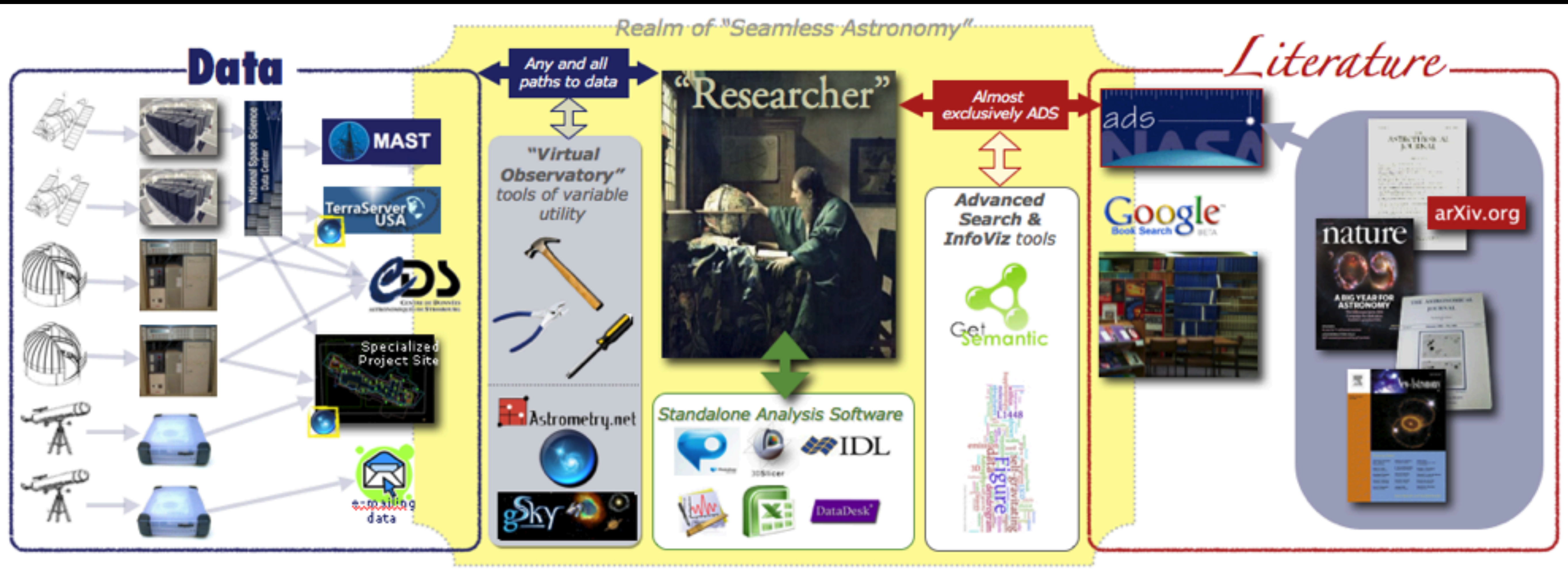
COMPLETE Data Available

Center on Perseus | Center on Ophiuchus | Center on Serpens

Dataset	Show	Perseus	Ophiuchus	Serpens	Link
Full-Cloud Data (Phase I, All Data Available)					
GBT: HI Data Cube	☑	√	√	∅	Data
IRAS: Av/Temp Maps	☑	√	√	√	Data
FCRAO: 12CO	☑	√	√	√	Data
FCRAO: 13CO		√	√	√	Data
JCMT: 850 microns	☑	√	√	∅	Data
Spitzer c2d: IRAC 1,3 (3.6,5.8 µm)	☑	√	√	√	Data
Spitzer c2d: IRAC 2,4 (4.5,8 µm)	☑	√	√	√	Data
CSO/Bolocam: 1.2-mm	☑	√	∅	∅	Data
Spitzer MIPS: Derived Dust Map	☑	√	∅	∅	Data
Targeted Regions (Phase II, Some Data Not Yet Available)					
CTIO/Calar Alto: NIR (J,H,Ks)	☑	√	√	∅	Data
IRAM 30-m: N2H+ and C18O	☑	√	∅	∅	Data
IRAM 30-m: 1.1-mm continuum	☑	√	∅	∅	Data
Megacam/MMT: r,i,z images	☑	√	∅	∅	Data
Catalogs & Pointed Surveys					
NH3 Pointed Survey	☐	√	∅	∅	Data
YSO Candidate list (c2d)	☐	√	√	√	Data

Seamless Astronomy

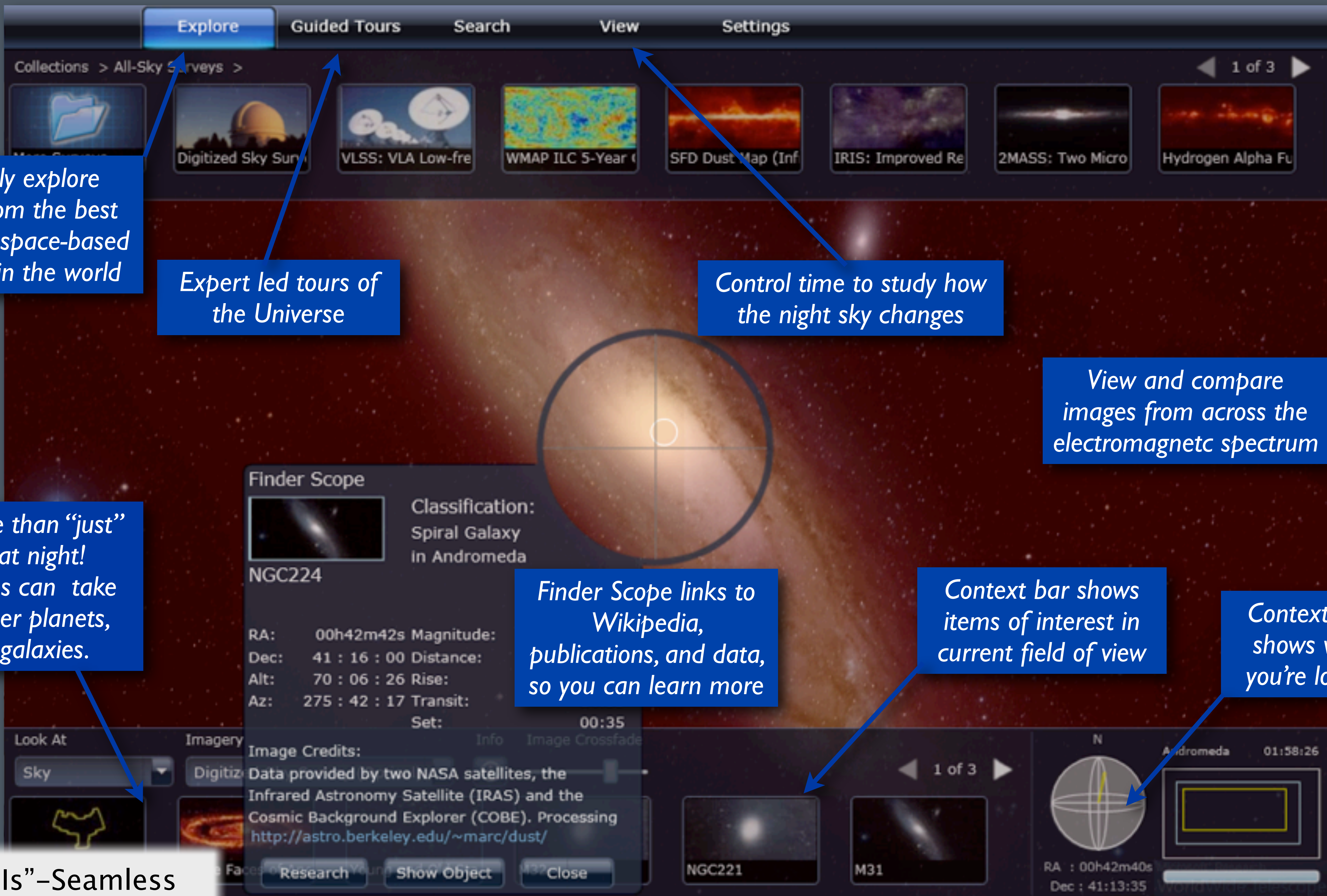
Alberto Accomazzi, Doug Burke, Alberto Conti, Carol Christian, Mercé Crosas, Raffaele D'Abrusco, Rahul Davé, Christopher Erdmann, Jonathan Fay, Jay Luker, Alyssa Goodman, Michael Kurtz, Gus Muench, Alberto Pepe, Curtis Wong



WWT-"NUIs"-Seamless Astronomy

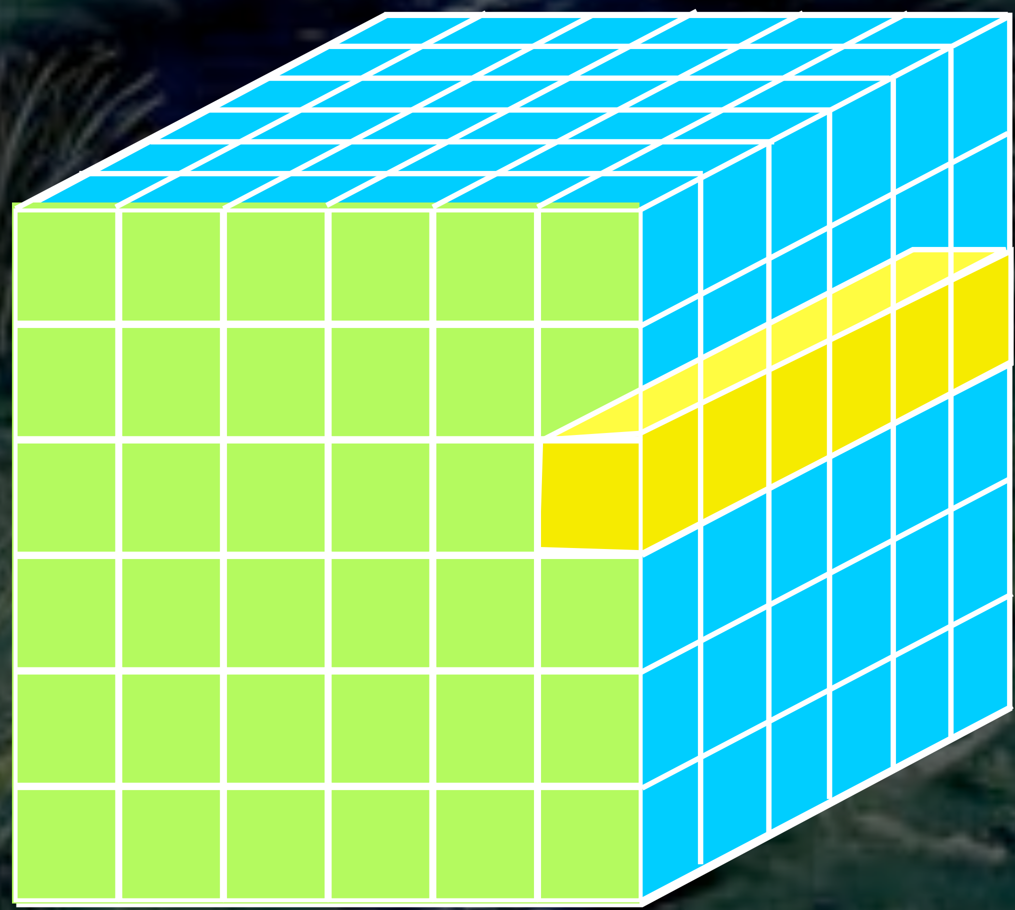
Microsoft® Research
WorldWide Telescope

Experience WWT at **worldwidetelescope.org**



WWT–"NUIs"–Seamless Astronomy

The dream scenario...



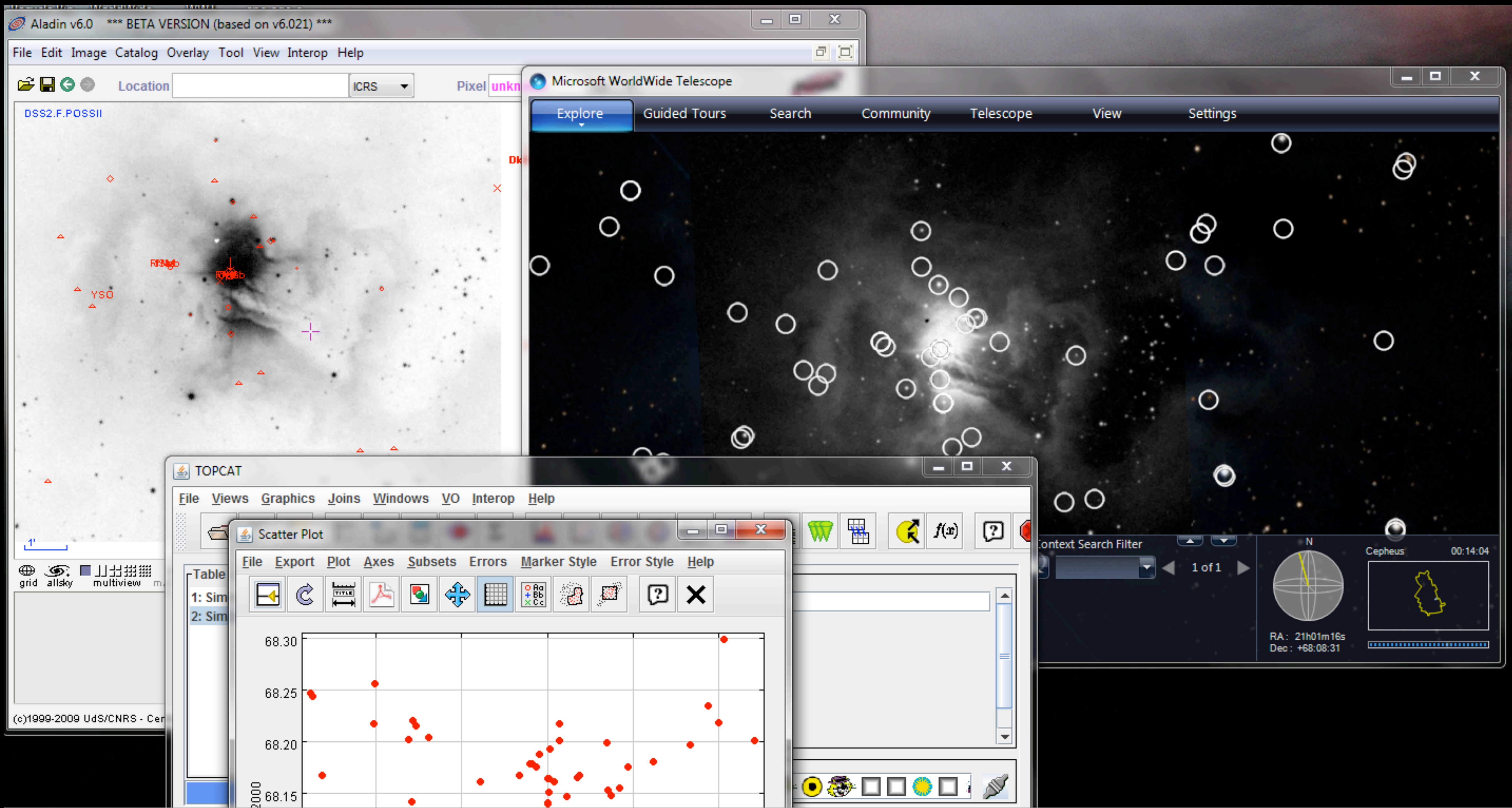
WWT-"NUIs"-Seamless Astronomy

Challenge #1: 3D Selection

- Wh
- How
- How … on?

visit our new “Viz-e-Lab”!

Challenge #2: Too many windows...



Tuesday, March 22, 2011

Challenge #3:

What does "Publication-Quality" Graphics Mean in an Interactive 3D World?

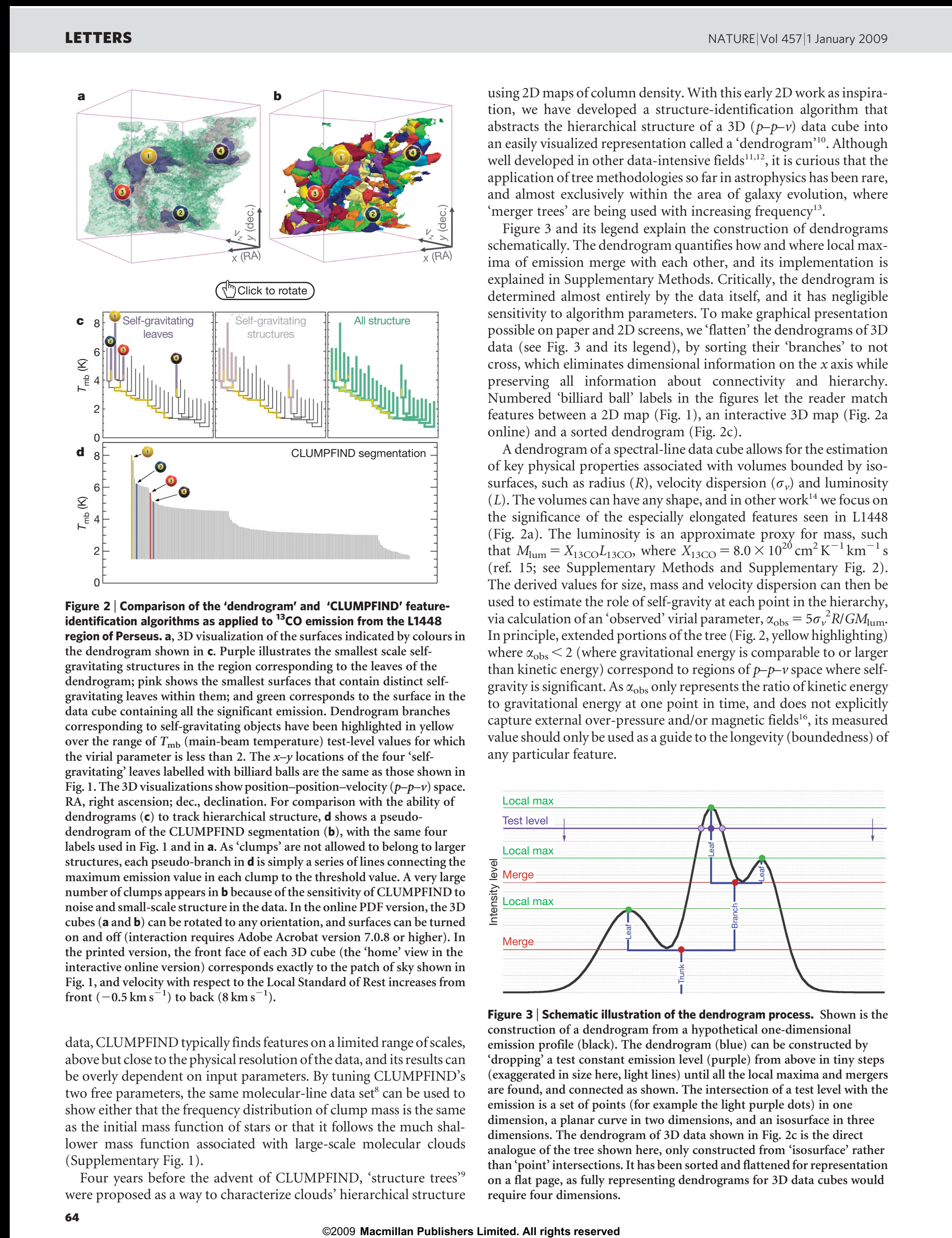


Figure 2 | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to ^{13}CO emission from the L1448 region of Perseus. a, 3D visualization of the surfaces indicated by colours in the dendrogram shown in c. Purple illustrates the smallest scale self-gravitating structures in the region corresponding to the leaves of the dendrogram; pink shows the smallest surfaces that contain distinct self-gravitating leaves within them; and green corresponds to the surface in the data cube containing all the significant emission. Dendrogram branches corresponding to self-gravitating objects have been highlighted in yellow over the range of T_{mb} (main-beam temperature) test-level values for which the virial parameter is less than 2. The x–y locations of the four 'self-gravitating' leaves labelled with billiard balls are the same as those shown in Fig. 1. The 3D visualizations show position–position–velocity (p–p–v) space. RA, right ascension; dec., declination. For comparison with the ability of dendrograms (c) to track hierarchical structure, d shows a pseudo-dendrogram of the CLUMPFIND segmentation (b), with the same four labels used in Fig. 1 and in a. As 'clumps' are not allowed to belong to larger structures, each pseudo-branch in d is simply a series of lines connecting the maximum emission value in each clump to the threshold value. A very large number of clumps appears in b because of the sensitivity of CLUMPFIND to noise and small-scale structure in the data. In the online PDF version, the 3D cubes (a and b) can be rotated to any orientation, and surfaces can be turned on and off (interaction requires Adobe Acrobat version 7.0.8 or higher). In the printed version, the front face of each 3D cube (the 'home' view in the interactive online version) corresponds exactly to the patch of sky shown in Fig. 1, and velocity with respect to the Local Standard of Rest increases from front ($-0.5\,km\,s^{-1}$) to back ($8\,km\,s^{-1}$).

data, CLUMPFIND typically finds features on a limited range of scales, above but close to the physical resolution of the data, and its results can be overly dependent on input parameters. By tuning CLUMPFIND's two free parameters, the same molecular-line data set[8] can be used to show either that the frequency distribution of clump mass is the same as the initial mass function of stars or that it follows the much shallower mass function associated with large-scale molecular clouds (Supplementary Fig. 1).

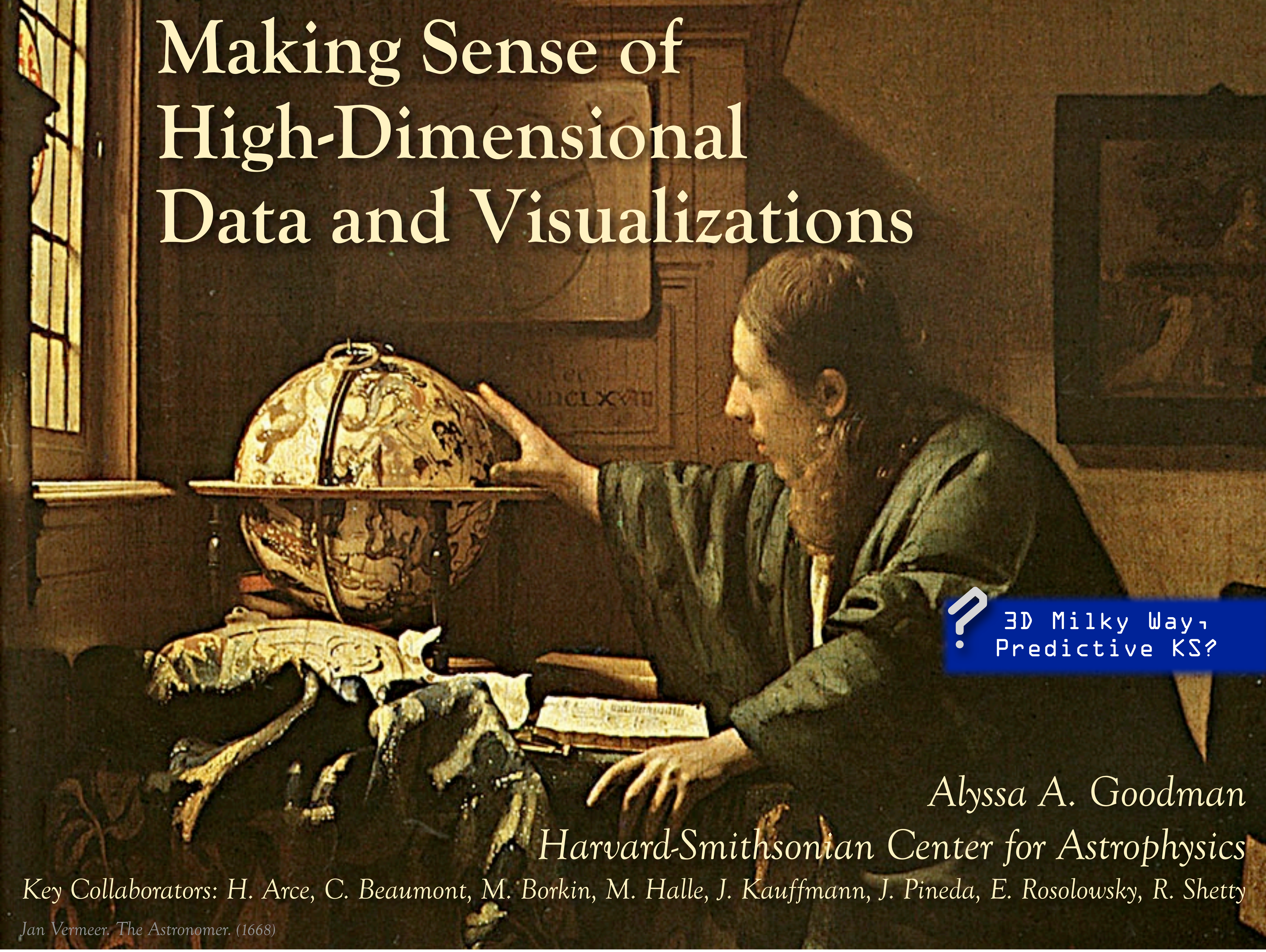
Four years before the advent of CLUMPFIND, 'structure trees'[9] were proposed as a way to characterize clouds' hierarchical structure using 2D maps of column density. With this early 2D work as inspiration, we have developed a structure-identification algorithm that abstracts the hierarchical structure of a 3D (p–p–v) data cube into an easily visualized representation called a 'dendrogram'[10]. Although well developed in other data-intensive fields[11,12], it is curious that the application of tree methodologies so far in astrophysics has been rare, and almost exclusively within the area of galaxy evolution, where 'merger trees' are being used with increasing frequency[13].

Figure 3 and its legend explain the construction of dendrograms schematically. The dendrogram quantifies how and where local maxima of emission merge with each other, and its implementation is explained in Supplementary Methods. Critically, the dendrogram is determined almost entirely by the data itself, and it has negligible sensitivity to algorithm parameters. To make graphical presentation possible on paper and 2D screens, we 'flatten' the dendrograms of 3D data (see Fig. 3 and its legend), by sorting their 'branches' to not cross, which eliminates dimensional information on the x axis while preserving all information about connectivity and hierarchy. Numbered 'billiard ball' labels in the figures let the reader match features between a 2D map (Fig. 1), an interactive 3D map (Fig. 2a online) and a sorted dendrogram (Fig. 2c).

A dendrogram of a spectral-line data cube allows for the estimation of key physical properties associated with volumes bounded by isosurfaces, such as radius (R), velocity dispersion (σ_v) and luminosity (L). The volumes can have any shape, and in other work[14] we focus on the significance of the especially elongated features seen in L1448 (Fig. 2a). The luminosity is an approximate proxy for mass, such that $M_{lum} = X_{13CO}L_{13CO}$, where $X_{13CO} = 8.0 \times 10^{20}\,cm^2\,K^{-1}\,km^{-1}\,s$ (ref. 15; see Supplementary Methods and Supplementary Fig. 2). The derived values for size, mass and velocity dispersion can then be used to estimate the role of self-gravity at each point in the hierarchy, via calculation of an 'observed' virial parameter, $\alpha_{obs} = 5\sigma_v^2 R/GM_{lum}$. In principle, extended portions of the tree (Fig. 2, yellow highlighting) where $\alpha_{obs} < 2$ (where gravitational energy is comparable to or larger than kinetic energy) correspond to regions of p–p–v space where self-gravity is significant. As α_{obs} only represents the ratio of kinetic energy to gravitational energy at one point in time, and does not explicitly capture external over-pressure and/or magnetic fields[16], its measured value should only be used as a guide to the longevity (boundedness) of any particular feature.

Figure 3 | Schematic illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (black). The dendrogram (blue) can be constructed by 'dropping' a test constant emission level (purple) from above in tiny steps (exaggerated in size here, light lines) until all the local maxima and mergers are found, and connected as shown. The intersection of a test level with the emission is a set of points (for example the light purple dots) in one dimension, a planar curve in two dimensions, and an isosurface in three dimensions. The dendrogram of 3D data shown in Fig. 2c is the direct analogue of the tree shown here, only constructed from 'isosurface' rather than 'point' intersections. It has been sorted and flattened for representation on a flat page, as fully representing dendrograms for 3D data cubes would require four dimensions.

*Goodman, Rosolowsky, Borkin, Foster, Halle, Kauffmann & Pineda, **Nature**, 2009*



Tuesday, March 22, 2011

The "data deluge" in science is old news. Now, it's pouring, and we need working tools to collect, sort out, understand, and keep what is falling down on us.
In astronomy, the greatest insights very often come from studies where more than one "band" of data (e.g. optical, infrared, radio, X-ray) is combined. And, data sets aren't just large--they are often also high-dimensional, in that they contain information about flux as functions not just of position on the sky, but also as functions of a third dimension (e.g. frequency, velocity), and/or of time. Life science, geophysical, and geospatial data all present similar challenges.

In this talk, I will focus on examples drawn from my group's research on star formation in molecular clouds. In particular, I will show how new visualization and statistical analysis techniques relying on interactive high-dimensional views of data and on automated algorithms for "segmenting" data give new insight. "Segmentation" in imaging terms refers to extracting the meaningful structures from data, and I will show results from both dendrogram (tree-hierarchy) and machine-learning approaches. I will emphasize how the visualization of segmentation results is critical for understanding. The highlighted science results will show how we can now--for the first time--quantitatively but intuitively understand the connections between the "real" (position-position-position) space where simulations (e.g. of star formation) can be made and the "observational" (e.g. position-position-velocity) space available to earthbound astronomers. As a result of this newfound understanding, we can place important limits on the validity of virial-theorem-based calculations of the properties of gas--allowing, for example, for better estimates of which gas in star-forming regions is most likely to stay bound long enough to form stars.

Even though this abstract may sound technical to non-star-formation or non-computational researchers, my goal will be to keep the talk accessible to non-experts, so people from other fields faced with high-dimensional data and visualization challenges should feel free to join in--and to ask questions