Talk at Google 30 May 2017

New Insights on Galaxy Formation from Comparing Simulations and Observations

Joel Primack
Distinguished Professor of Physics Emeritus, UCSC

Brief introduction to modern cosmology, based on ΛCDM: dark energy and dark matter

Cosmic large scale structure simulations and star formation in galaxies

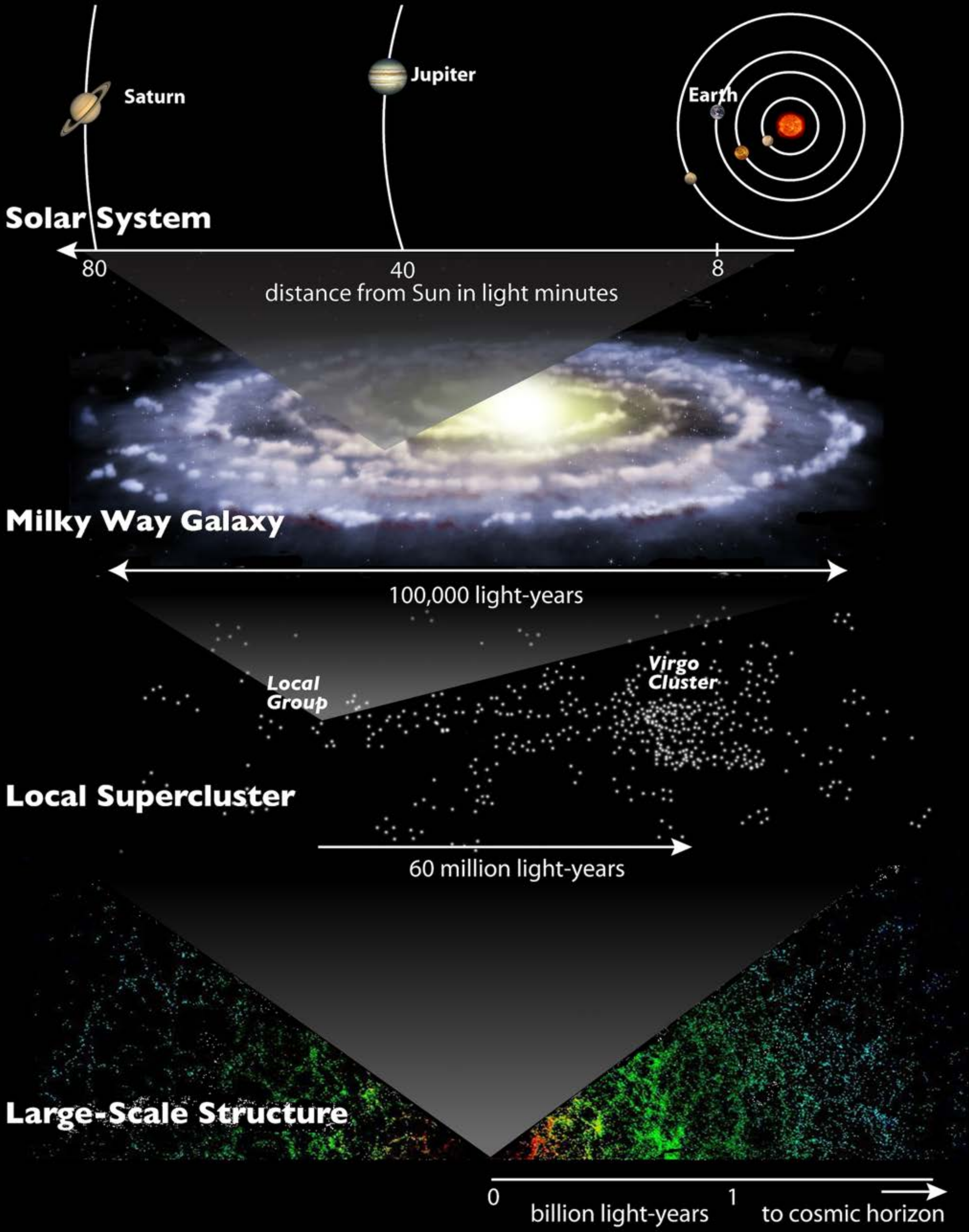
Comparing high-resolution hydrodynamic galaxy simulations with observations

Astronomers used to think that galaxies form as disks, that forming galaxies are pretty smooth, and that galaxies generally grow in radius as they grow in mass — but Hubble Space Telescope data show that all these statements are false, and our simulations may explain why.

We are using these simulations and deep learning to improve understanding of galaxy formation, with support from Google.

The Modern Scientific Cosmos

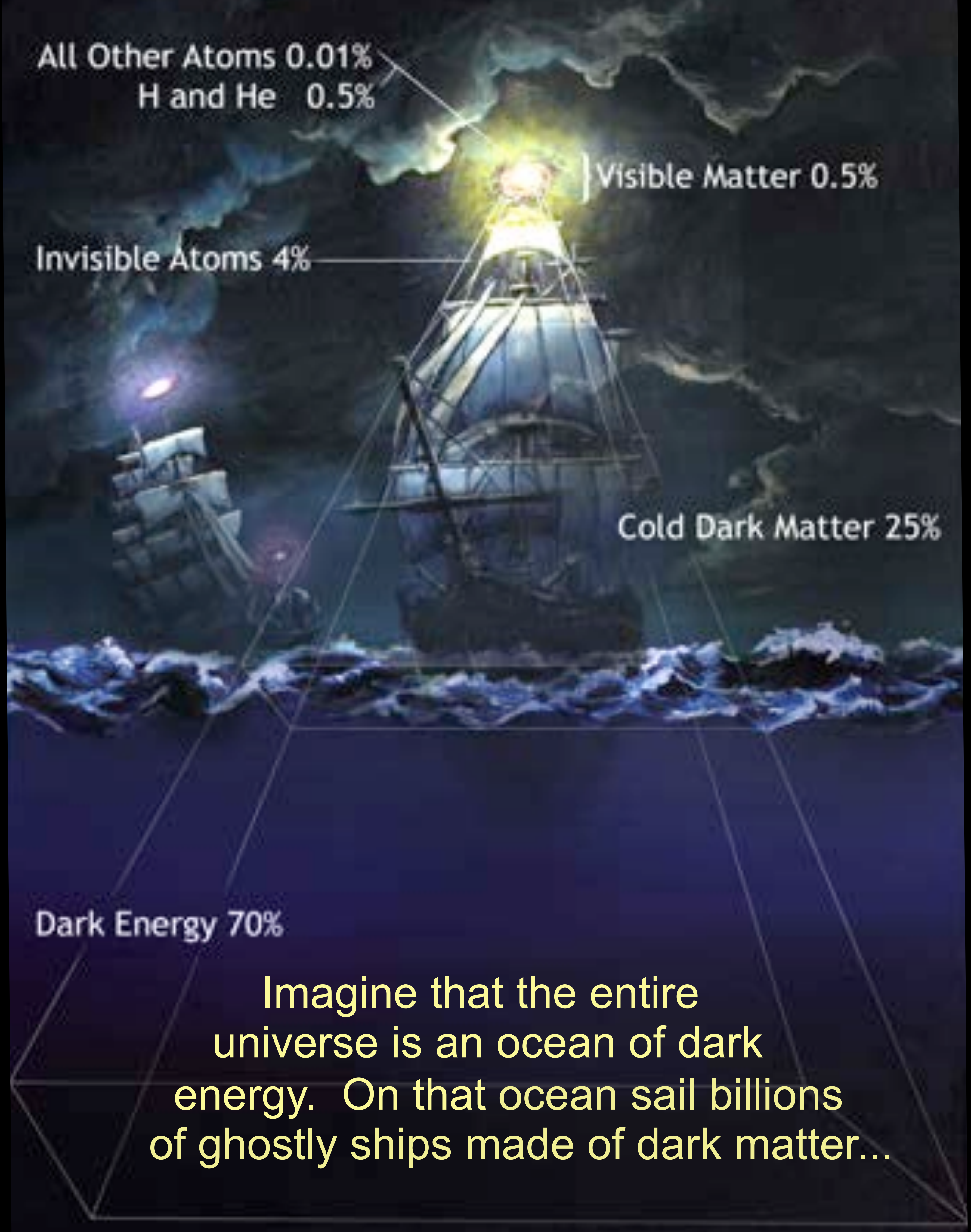
Our Cosmic Address



each dot is a big galaxy

Sloan Digital Sky Survey

Matter and Energy Content of the Universe



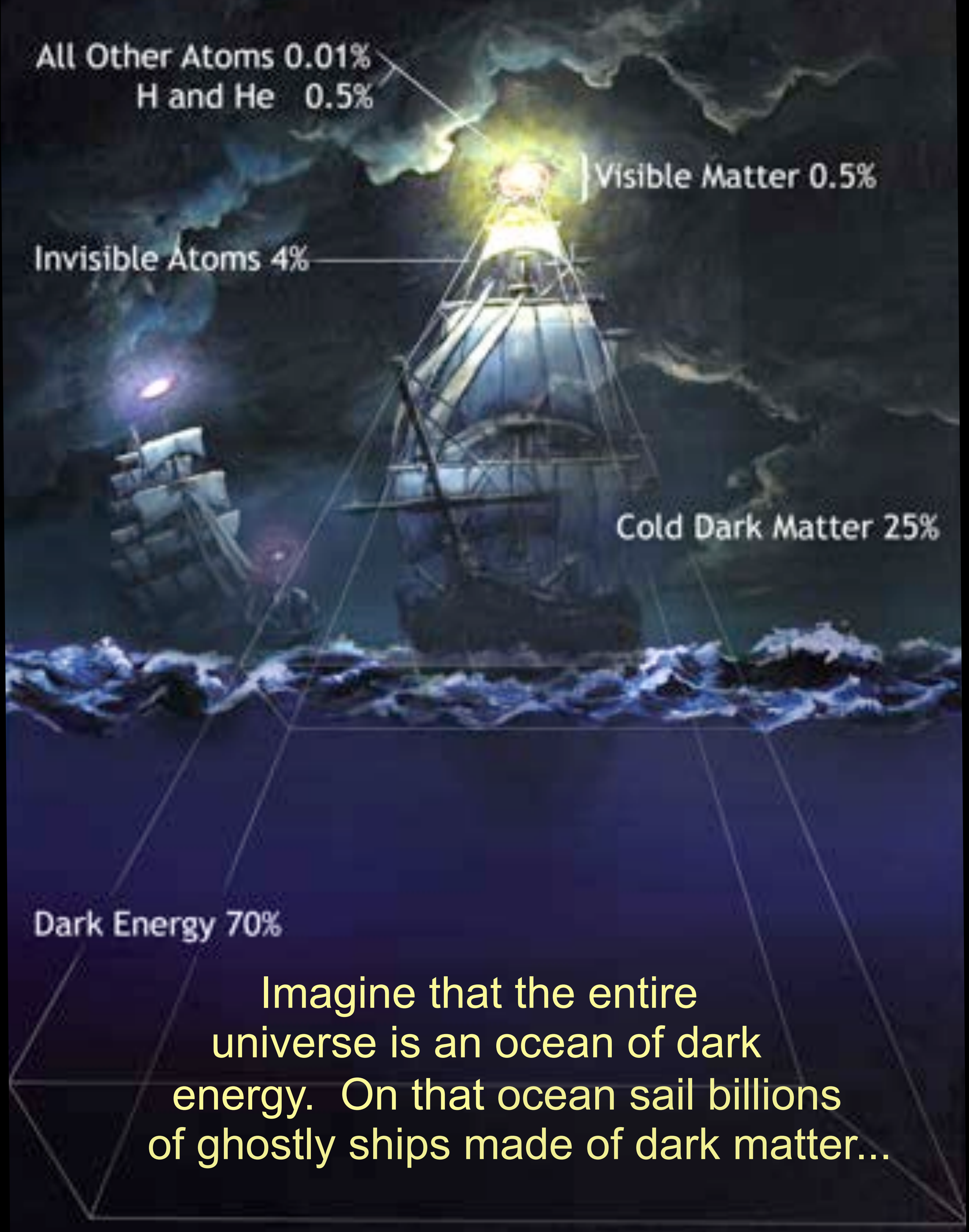
Imagine that the entire universe is an ocean of dark energy. On that ocean sail billions of ghostly ships made of dark matter...

Matter and Energy Content of the Universe

Dark Matter Ships

on a

Dark Energy Ocean



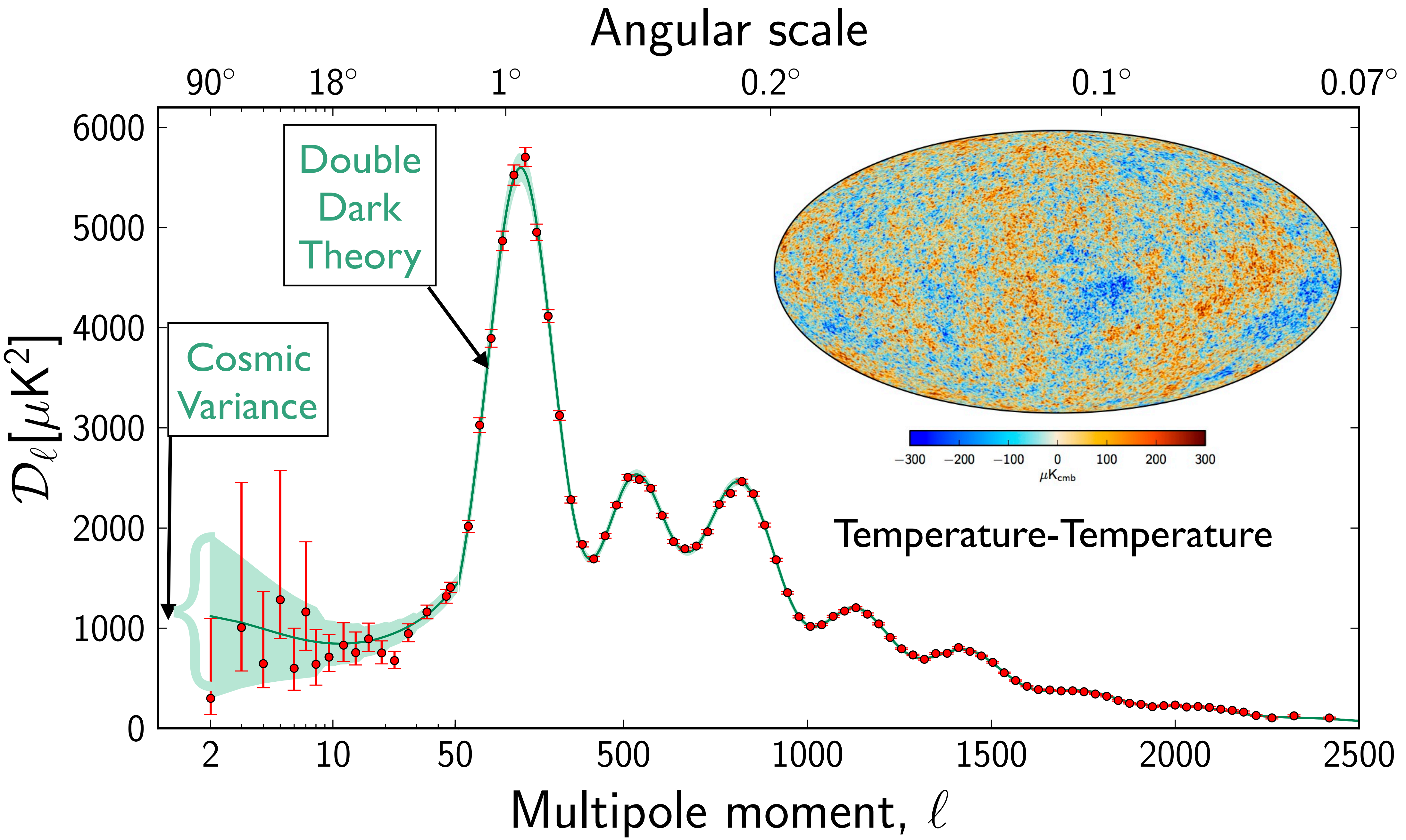
Imagine that the entire universe is an ocean of dark energy. On that ocean sail billions of ghostly ships made of dark matter...

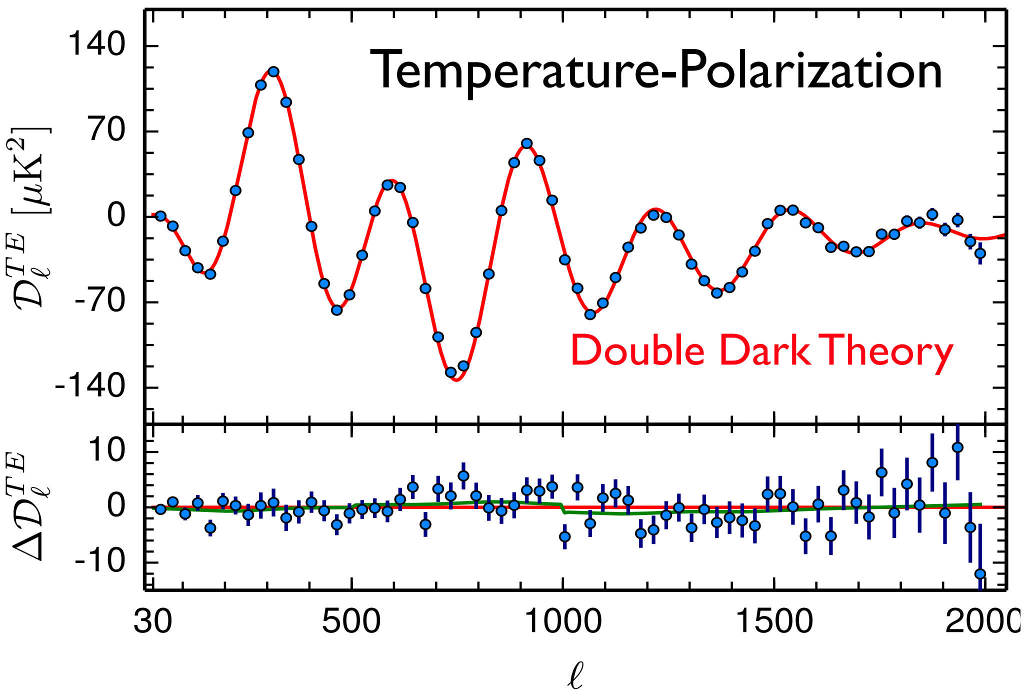
ΛCDM

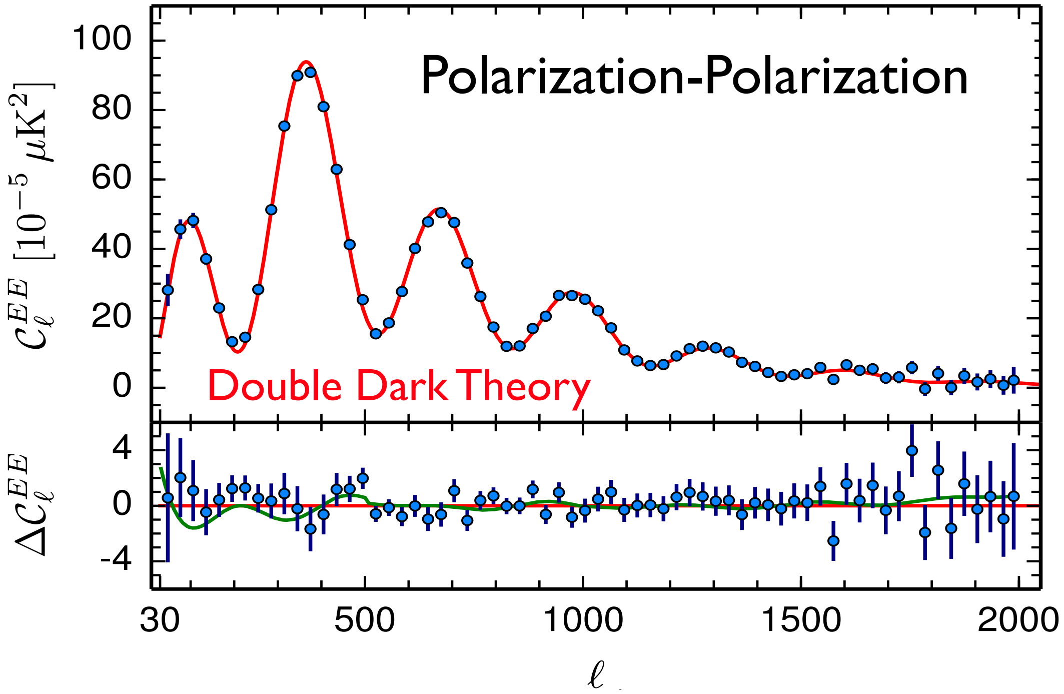
Double Dark Theory

European Space Agency PLANCK Satellite Data

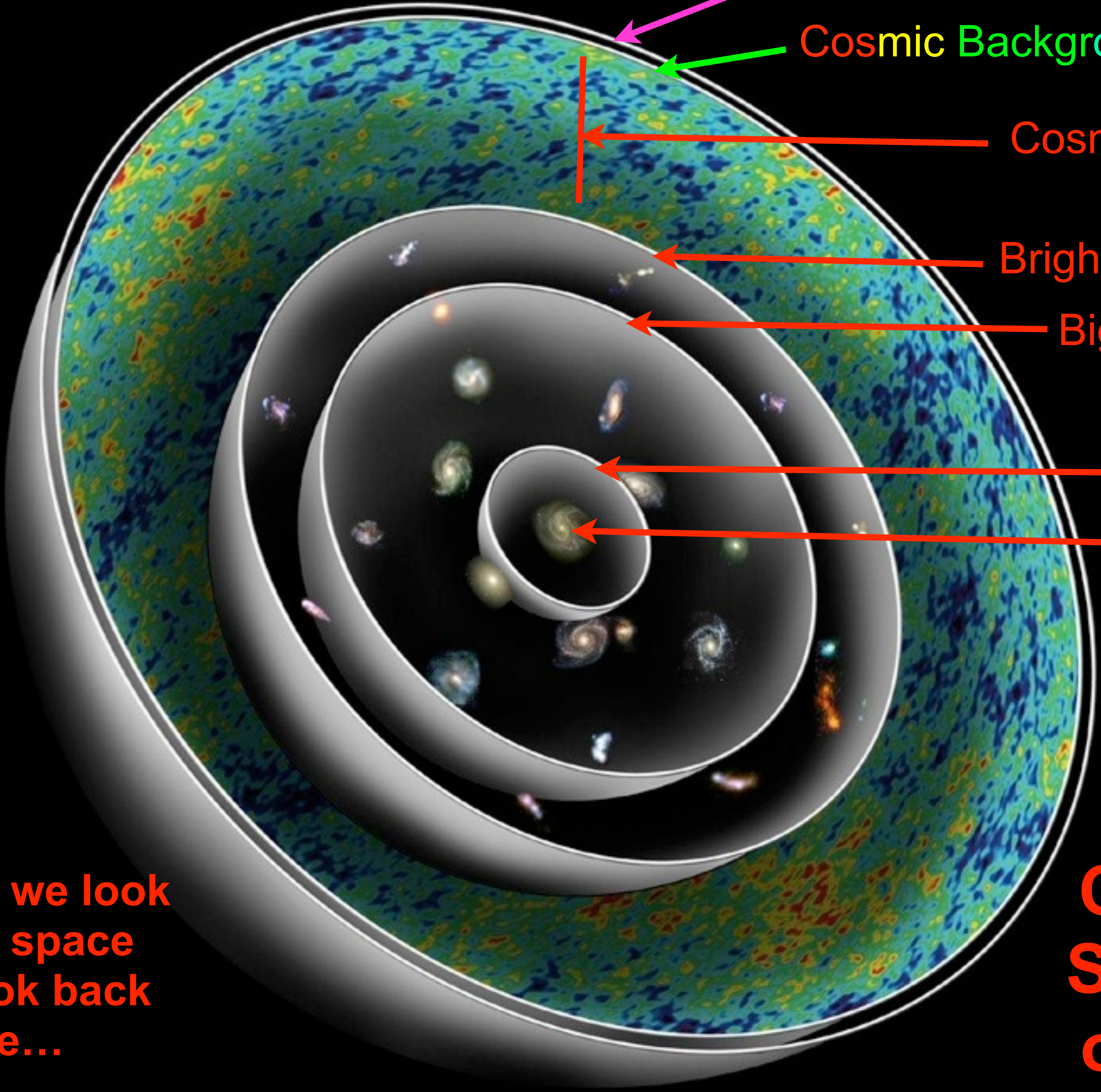
Released February 9, 2015







Cosmic Spheres of Time



Cosmic Horizon (The Big Bang)

Cosmic Background Radiation

Cosmic Dark Ages

Bright Galaxies Form

Big Galaxies Form

Earth Forms

Today

When we look out in space we look back in time…

Cosmological Simulations

Astronomical observations represent snapshots of moments in time. It is the role of astrophysical theory to produce movies -- both metaphorical and actual -- that link these snapshots together into a coherent physical theory.

Cosmological dark matter simulations show large scale structure and dark matter halo properties, basis for semi-analytic models

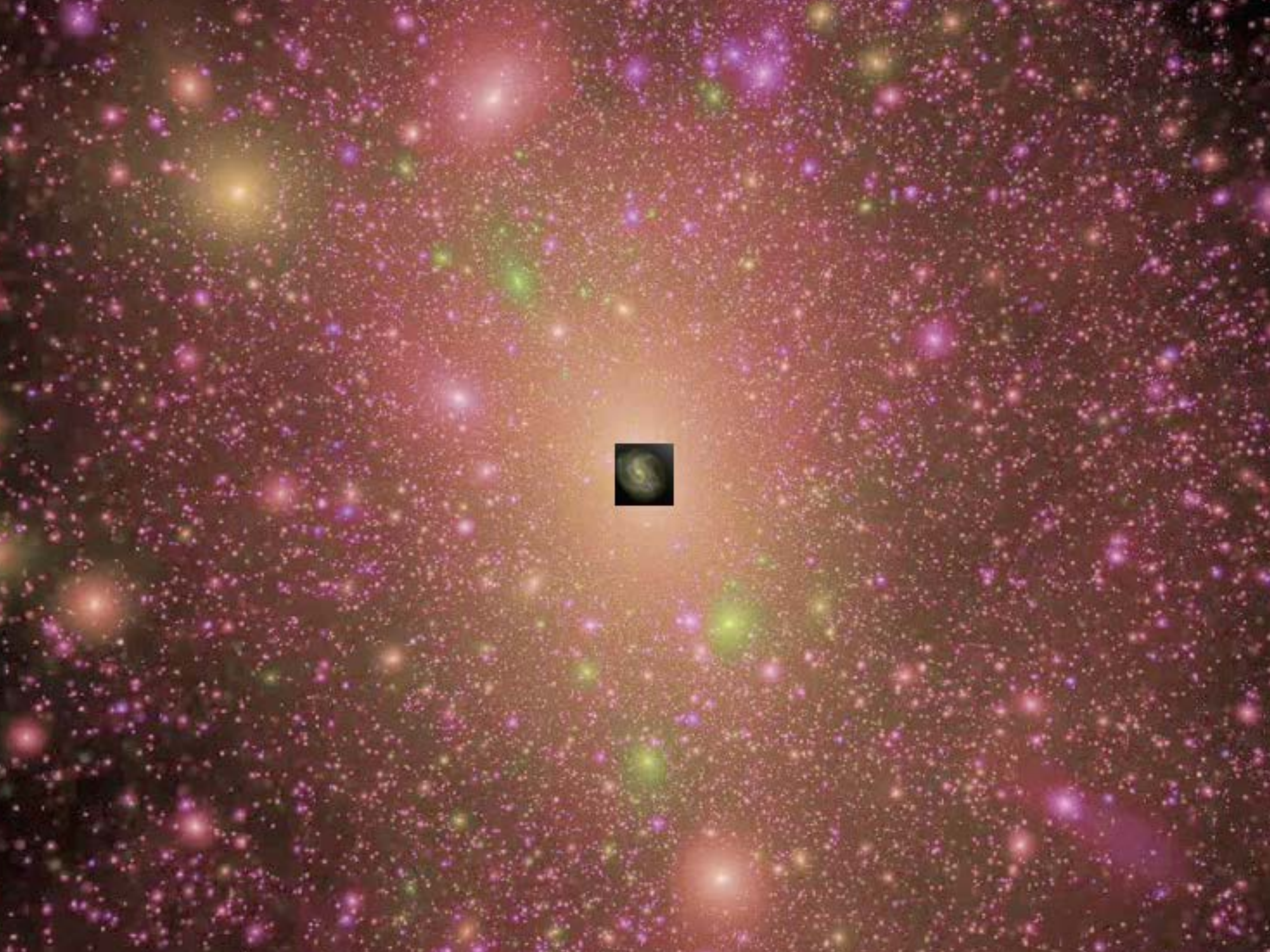
Hydrodynamic galaxy formation simulations: evolution of galaxies, formation of galactic spheroids, mock galaxy images and spectra including stellar evolution and dust effects

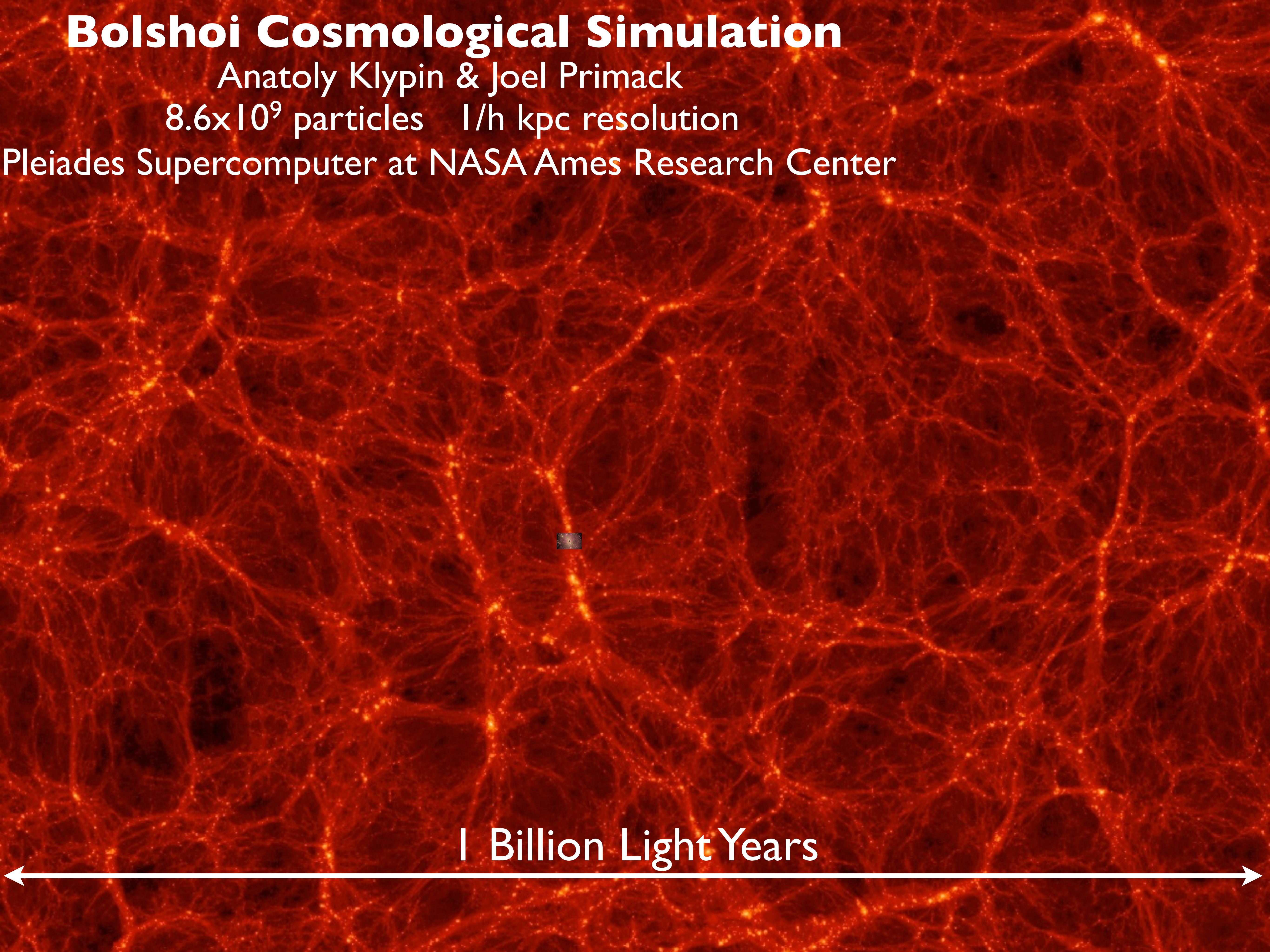
Aquarius Simulation

Volker Springel

Milky Way
100,000 light years

Milky Way Dark Matter Halo
1.5 million light years



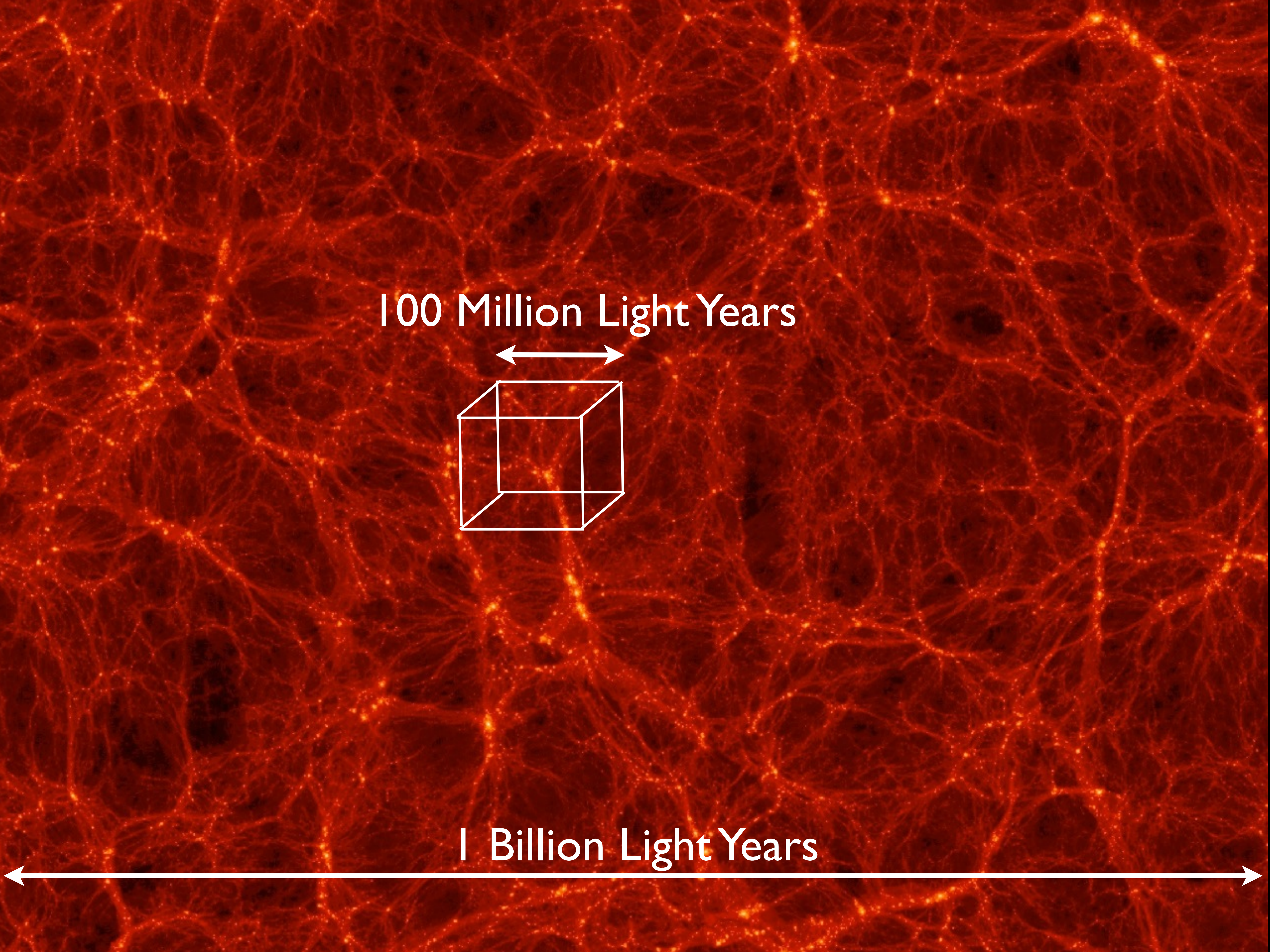


Bolshoi Cosmological Simulation

Anatoly Klypin & Joel Primack

8.6×10^9 particles   1/h kpc resolution

Pleiades Supercomputer at NASA Ames Research Center


100 Million Light Years
1 Billion Light Years

How the Halo of the Big Cluster Formed

100 Million Light Years

Bolshoi-Planck Cosmological Simulation

Merger Tree of a Large Halo

Peter Behroozi & Christoph Lee

We theorists make very complicated models of the star formation rate (SFR) in galaxies — but

Is Main Sequence SFR Controlled by Halo Mass Accretion?

by Aldo Rodríguez-Puebla, Joel Primack, Peter Behroozi, Sandra Faber MNRAS 2016

One can show that this must be true on average

Our radical SHARC (stellar halo accretion rate co-evolution) hypothesis is that this may be true halo-by-halo for many dark matter halos hosting star-forming galaxies

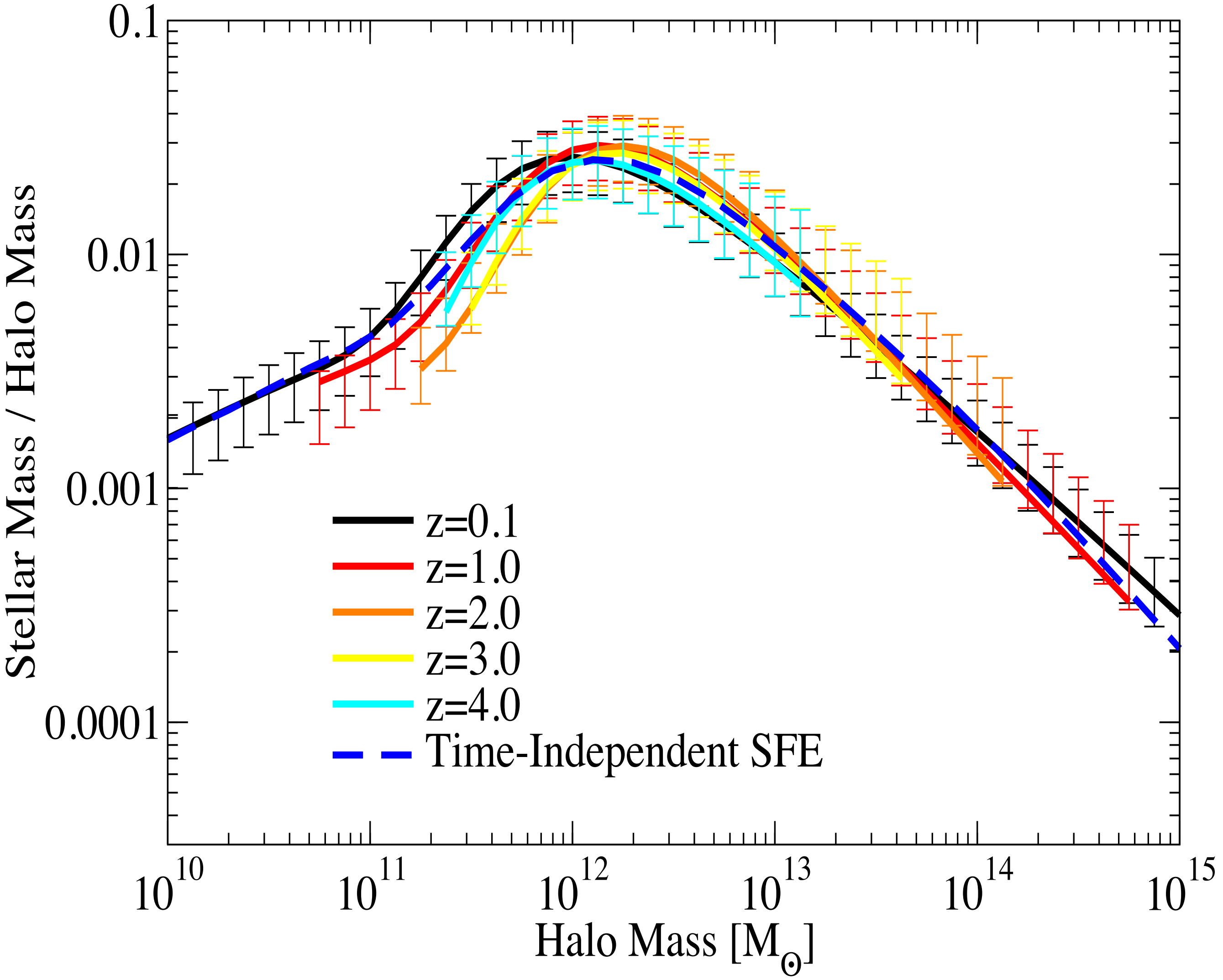
We then put SHARC in the bathtub, by combining the SHARC hypothesis with "bathtub" galaxy models

KEY BACKGROUND INFORMATION

- the stellar/halo mass relation
- the galaxy main sequence

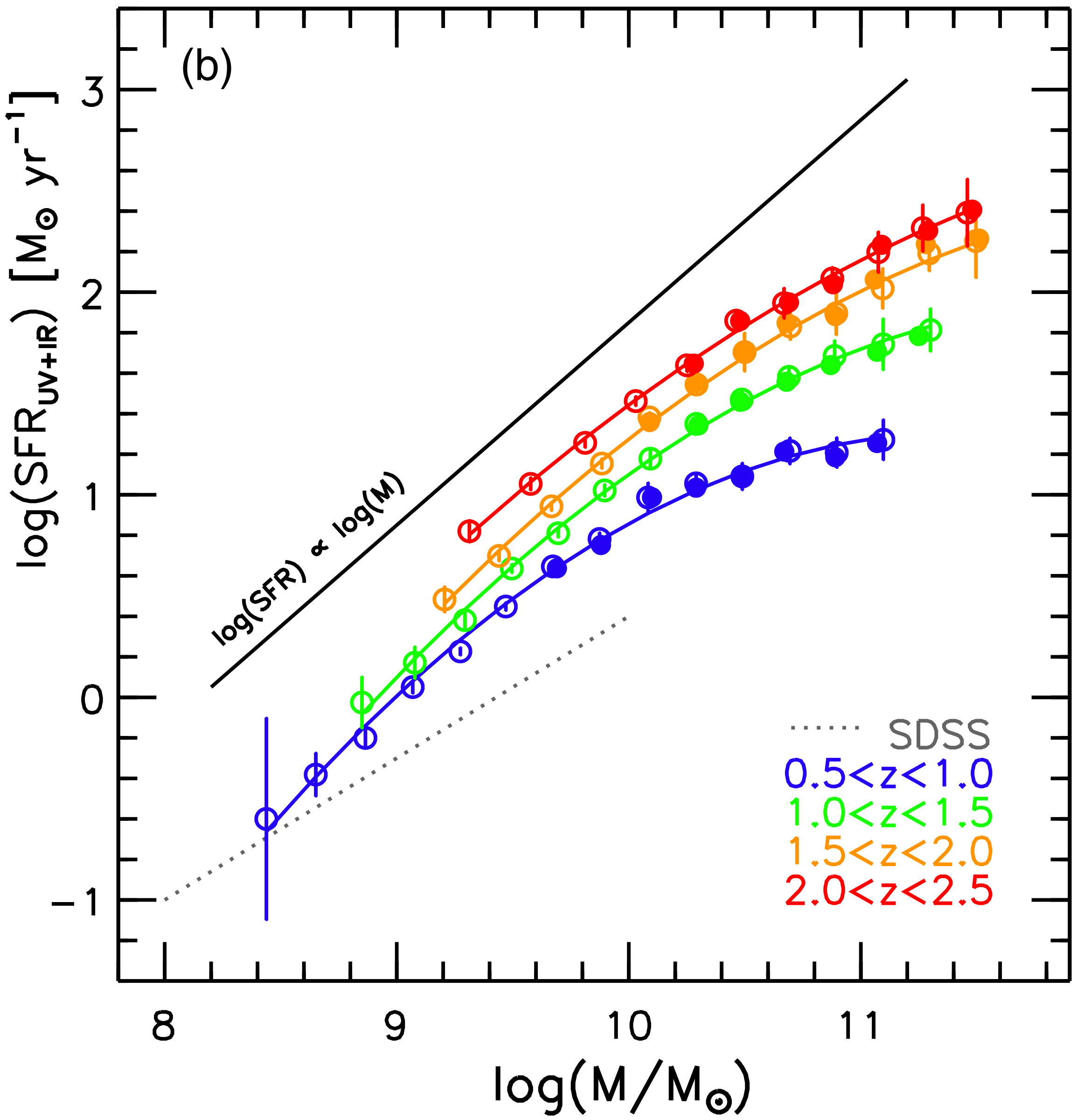
Two Key Discoveries About Galaxies

Relationship Between Galaxy Stellar Mass and Halo Mass



The stellar mass to halo mass ratio at multiple redshifts as derived from observations compared to the Bolshoi cosmological simulation. Error bars show 1σ uncertainties. A time-independent Star Formation Efficiency predicts a roughly **time-independent stellar mass to halo mass relationship**. (Behroozi, Wechsler, Conroy, ApJL 2013)

Star-forming Galaxies Lie on a "Main Sequence"

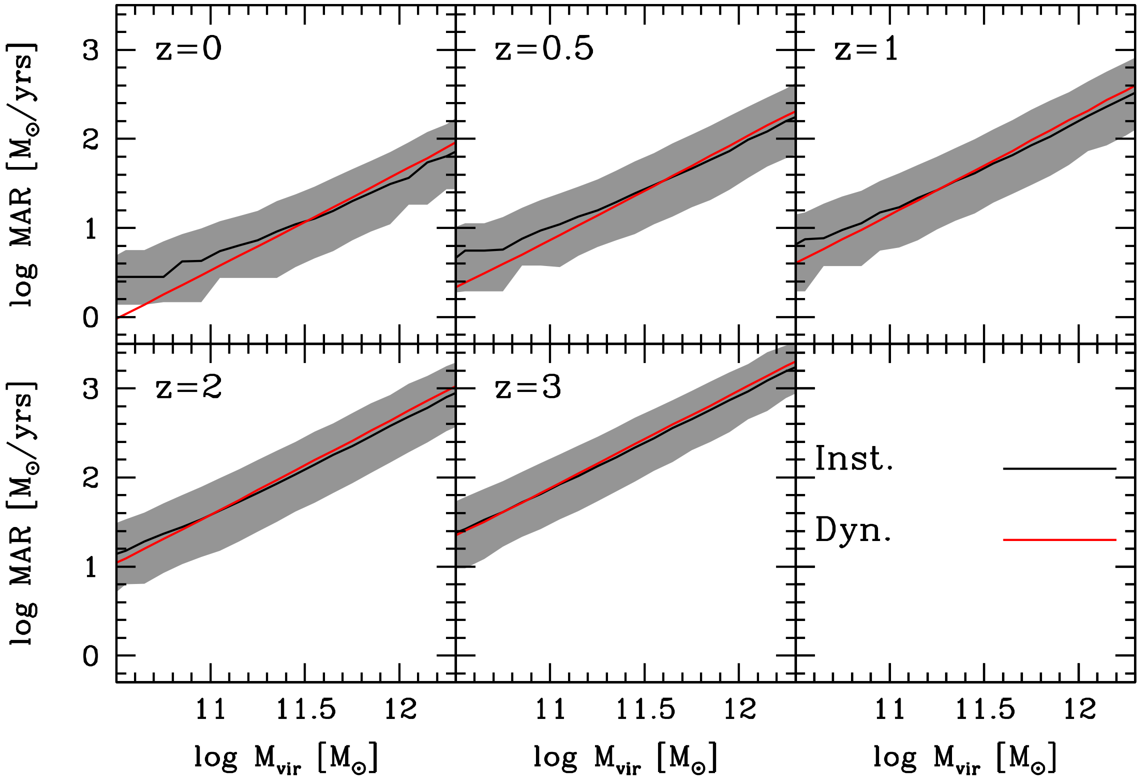


Just as the properties of hydrogen-burning stars are controlled by their mass, **the galaxy star formation rate (SFR) is approximately proportional to the stellar mass**, with the proportionality constant increasing with redshift up to about $z = 2.5$. (Whitaker et al. ApJ 2014)

Is Main Sequence SFR Controlled by Halo Mass Accretion?

by Aldo Rodríguez-Puebla, Joel Primack, Peter Behroozi, Sandra Faber MNRAS 2016

Halo mass accretion rates z=0 to 3



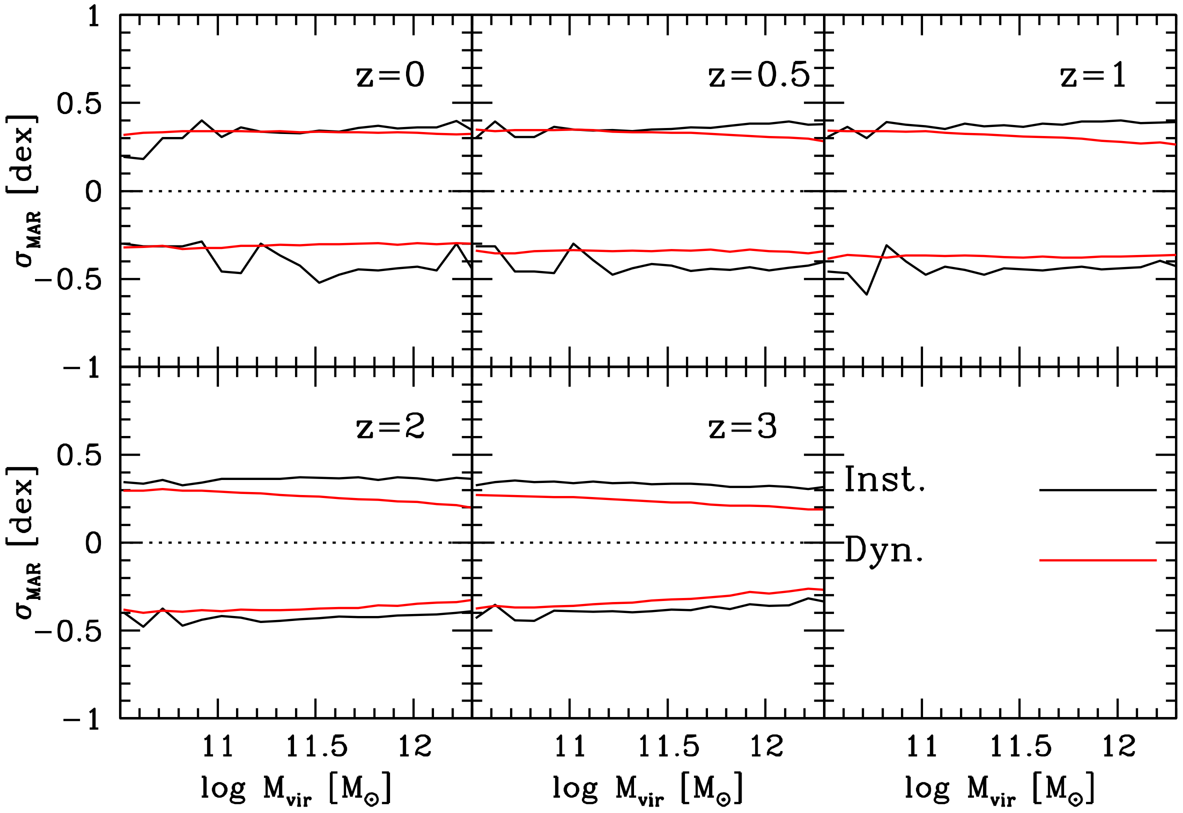
$$\frac{dM_*}{dt} = \frac{\partial M_*(M_{\rm vir}(t), z)}{\partial M_{\rm vir}} \frac{dM_{\rm vir}}{dt} + \frac{\partial M_*(M_{\rm vir}(t), z)}{\partial z} \frac{dz}{dt}$$

but if the M_*–$M_{\rm vir}$ relation is independent of redshift then the stellar mass of a central galaxy formed in a halo of mass $M_{\rm vir}(t)$ is $M_* = M_*(M_{\rm vir}(t))$. From this relation star formation rates are given simply by

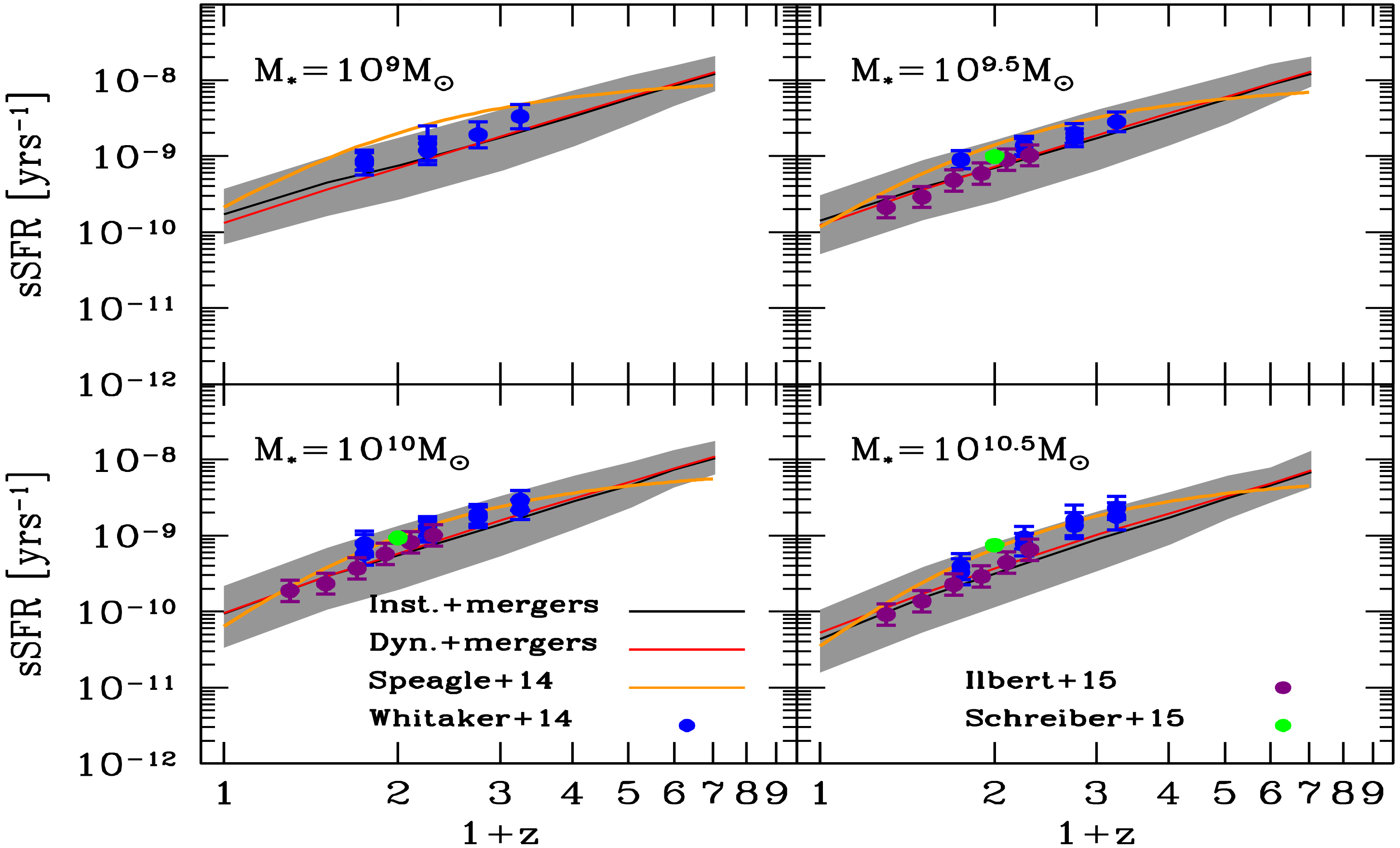
$$\frac{dM_*}{dt} = f_* \frac{d\log M_*}{d\log M_{\rm vir}} \frac{dM_{\rm vir}}{dt},$$

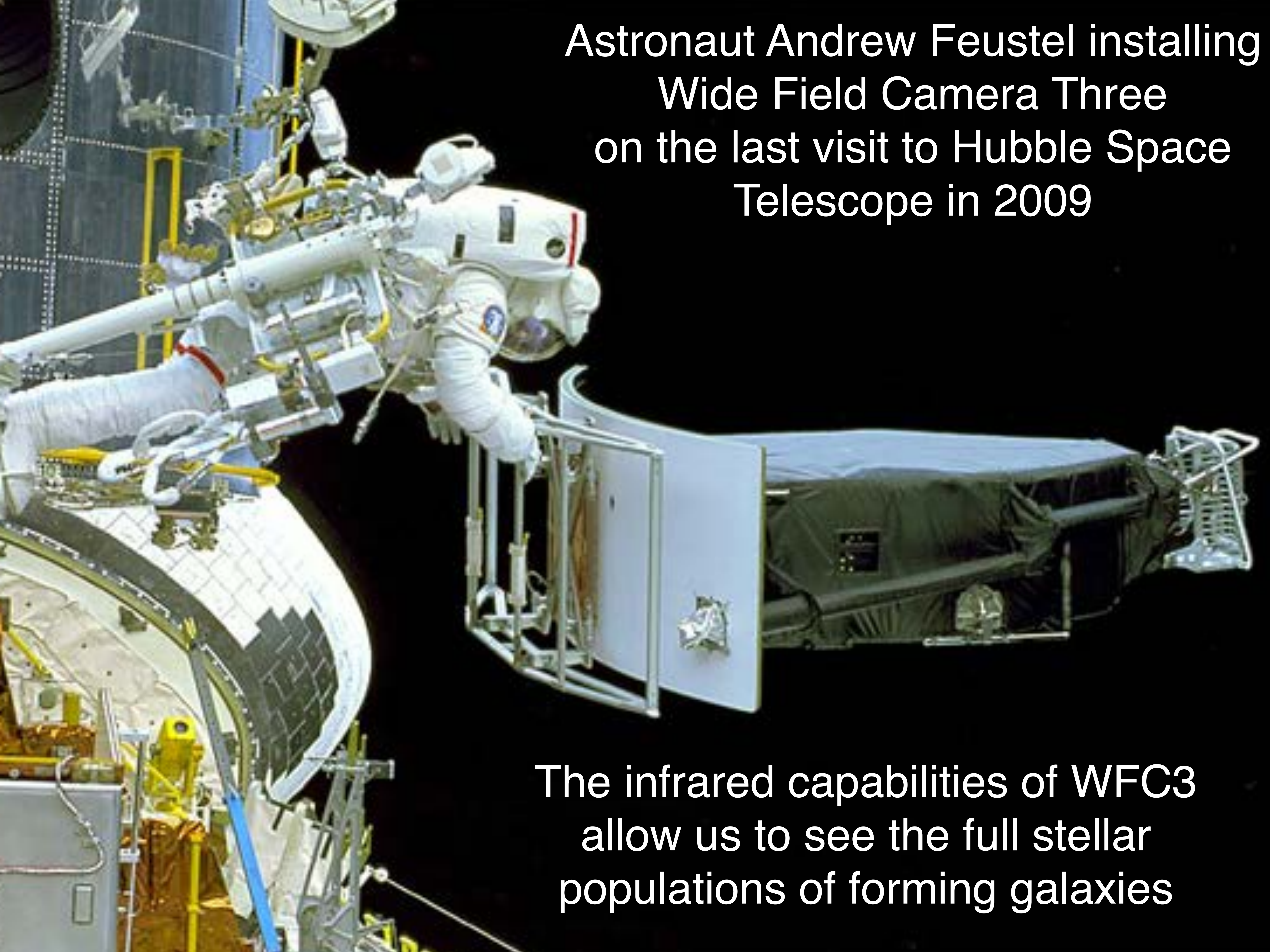
where $f_* = M_*/M_{\rm vir}$. We call this **Stellar-Halo Accretion Rate Coevolution (SHARC)** if true halo-by-halo.

Scatter of halo mass accretion rates



SHARC correctly predicts star formation rates to z ~ 4



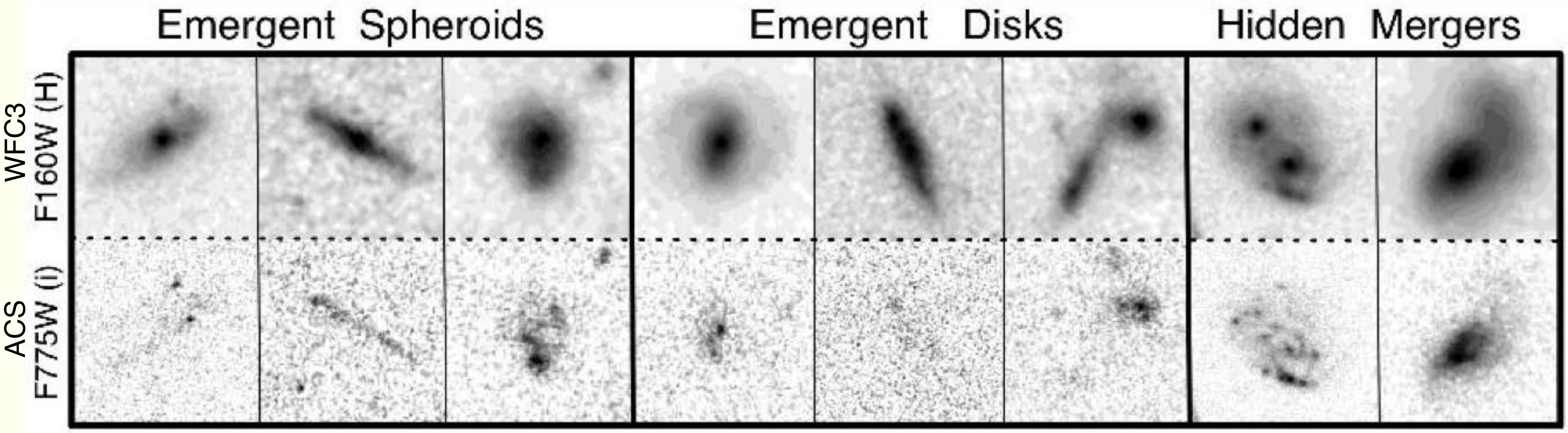


Astronaut Andrew Feustel installing
Wide Field Camera Three
on the last visit to Hubble Space
Telescope in 2009

The infrared capabilities of WFC3
allow us to see the full stellar
populations of forming galaxies

The CANDELS Survey

candels.ucolick.org



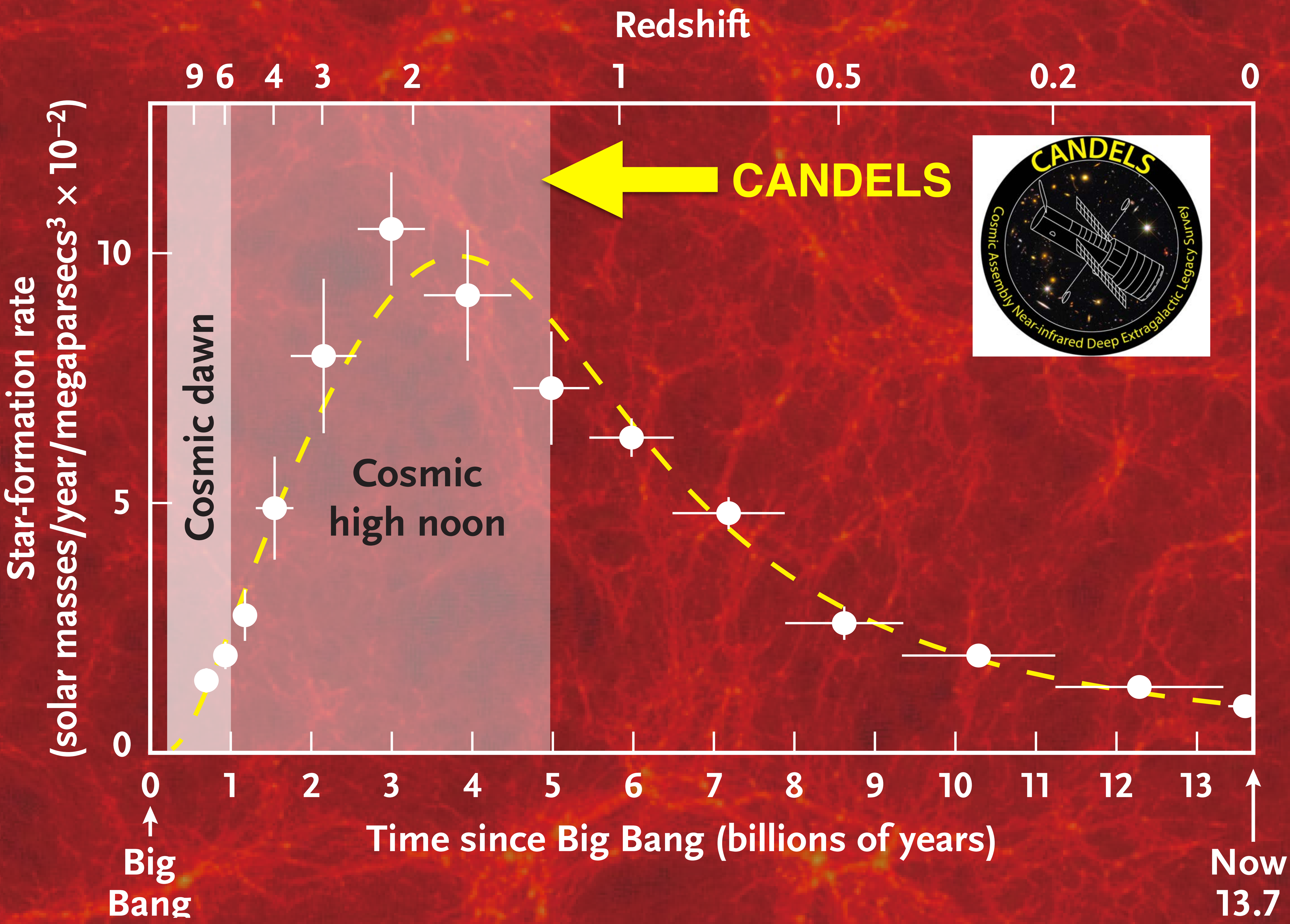
CANDELS: A Cosmic Odyssey

(blue 0.4 μm)(1+z) = 1.6 μm @ z = 3

(red 0.7 μm)(1+z) = 1.6 μm @ z = 2.3

CANDELS is a powerful imaging survey of the distant Universe being carried out with two cameras on board the Hubble Space Telescope.

- CANDELS is the largest project in the history of Hubble, with 902 assigned orbits of observing time. This is the equivalent of four months of Hubble time if executed consecutively, but in practice CANDELS will take three years to complete (2010-2013).
- The core of CANDELS is the revolutionary near-infrared WFC3 camera, installed on Hubble in May 2009. WFC3 is sensitive to longer, redder wavelengths, which permits it to follow the stretching of lightwaves caused by the expanding Universe. This enables CANDELS to detect and measure objects much farther out in space and nearer to the Big Bang than before. CANDELS also uses the visible-light ACS camera, and together the two cameras give unprecedented panchromatic coverage of galaxies from optical wavelengths to the near-IR.
- CANDELS will exploit this new lookback power to construct a "cosmic movie" of galaxy evolution that follows the life histories of galaxies from infancy to the present time. This work will cap Hubble's revolutionary series of discoveries on cosmic evolution and bequeath a legacy of precious data to future generations of astronomers.


Redshift
9
6
4
3
2
1
0.5
0.2
0
Star-formation rate
(solar masses/year/megaparsecs^3 × 10^{-2})
10
5
0
Cosmic dawn
Cosmic
high noon
CANDELS
CANDELS
Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey
0
1
2
3
4
5
6
7
8
9
10
11
12
13
Time since Big Bang (billions of years)
Big
Bang
Now
13.7

Cosmological Simulations

Astronomical observations represent snapshots of moments in time. It is the role of astrophysical theory to produce movies -- both metaphorical and actual -- that link these snapshots together into a coherent physical theory.

Cosmological dark matter simulations show large scale structure and dark matter halo properties, basis for semi-analytic models

Hydrodynamic galaxy formation simulations: evolution of galaxies, formation of galactic spheroids, mock galaxy images and spectra including stellar evolution and dust effects

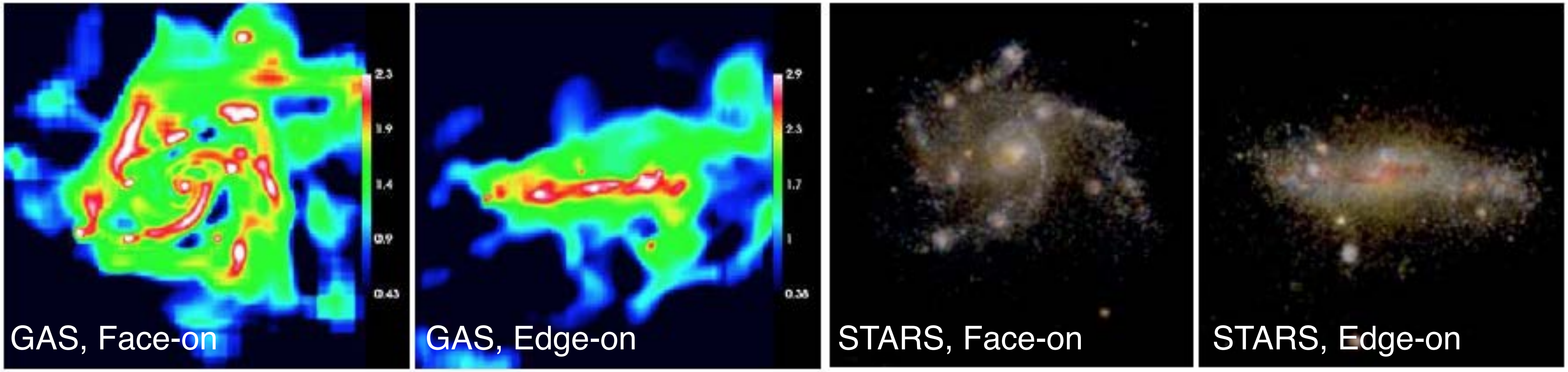
Galaxy Hydro Simulations: 2 Approaches

1. Low resolution (~ kpc)

Advantages: it's possible to simulate many galaxies and study galaxy populations and their interactions with CGM & IGM.
Disadvantages: we learn little about how galaxies themselves evolve, and cannot compare in detail with high-z galaxy images and spectra.
Examples: Overwhelmingly Large Simulations (OWLs, EAGLE), AREPO simulations in 100 Mpc box (Illustris)

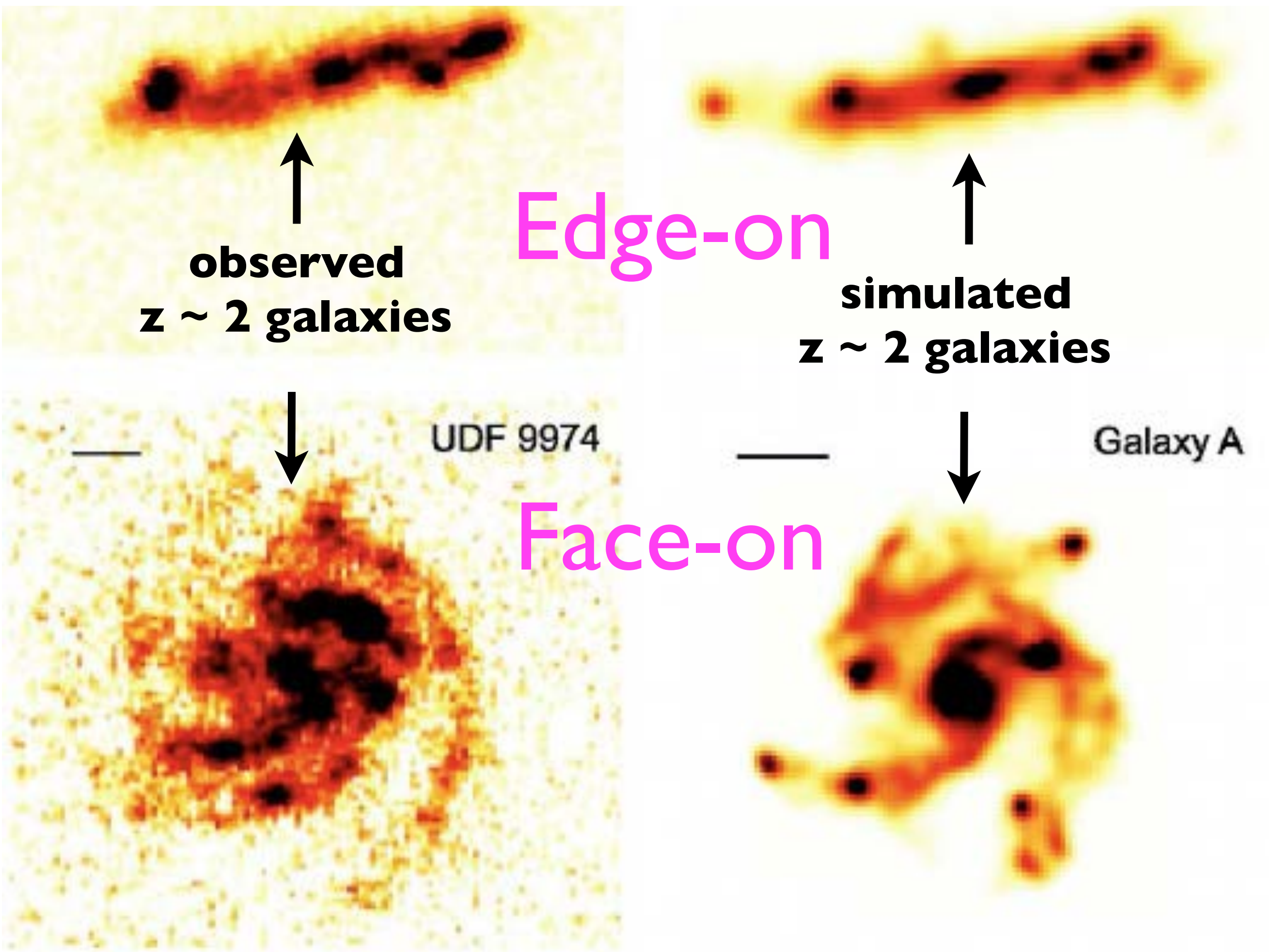
2. High resolution (~10s of pc)

Advantages: it's possible to compare in detail with high-z galaxy images and spectra, to discover how galaxies evolve, morphological drivers (e.g., galaxy shapes, clumps and other instabilities, origins of galactic spheroids, quenching). Radiative feedback essential?
Disadvantages: it's hard to run statistical galaxy samples, so the best approach puts simulation insights into SAMs.
Examples: ART and FIRE simulation suites, AGORA simulation comparison project

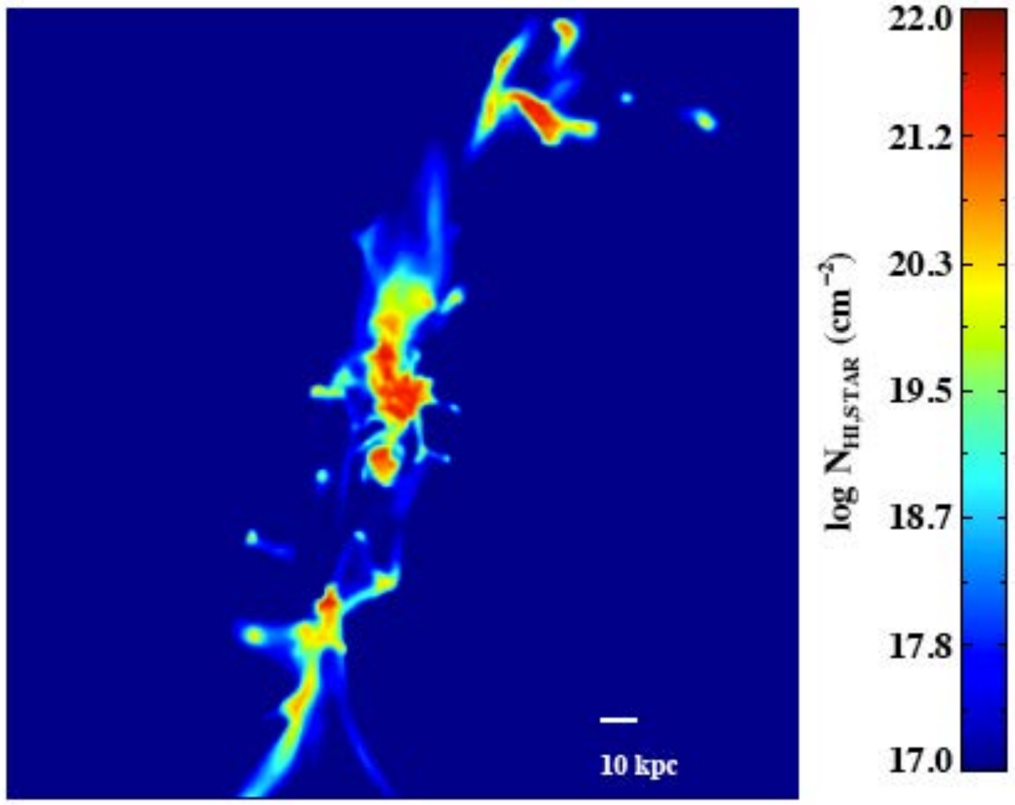


Clumpy Galaxies in hydroART Generation 1 Simulations

Figure 1: Violently unstable disks in $\sim 10^{11} M_\odot$ halos with $\sim 10^{9} M_\odot$ clumps at $z = 2.3$: (a) face-on, (b) edge-on (Ceverino et al. 2009, resolution 70 pc, images 10 kpc across). RGB color images of the same simulated galaxy through dust using *Sunrise*: (c) face-on, (d) edge-on, illustrating how the clumps can be reddened and obscured when viewed edge-on.



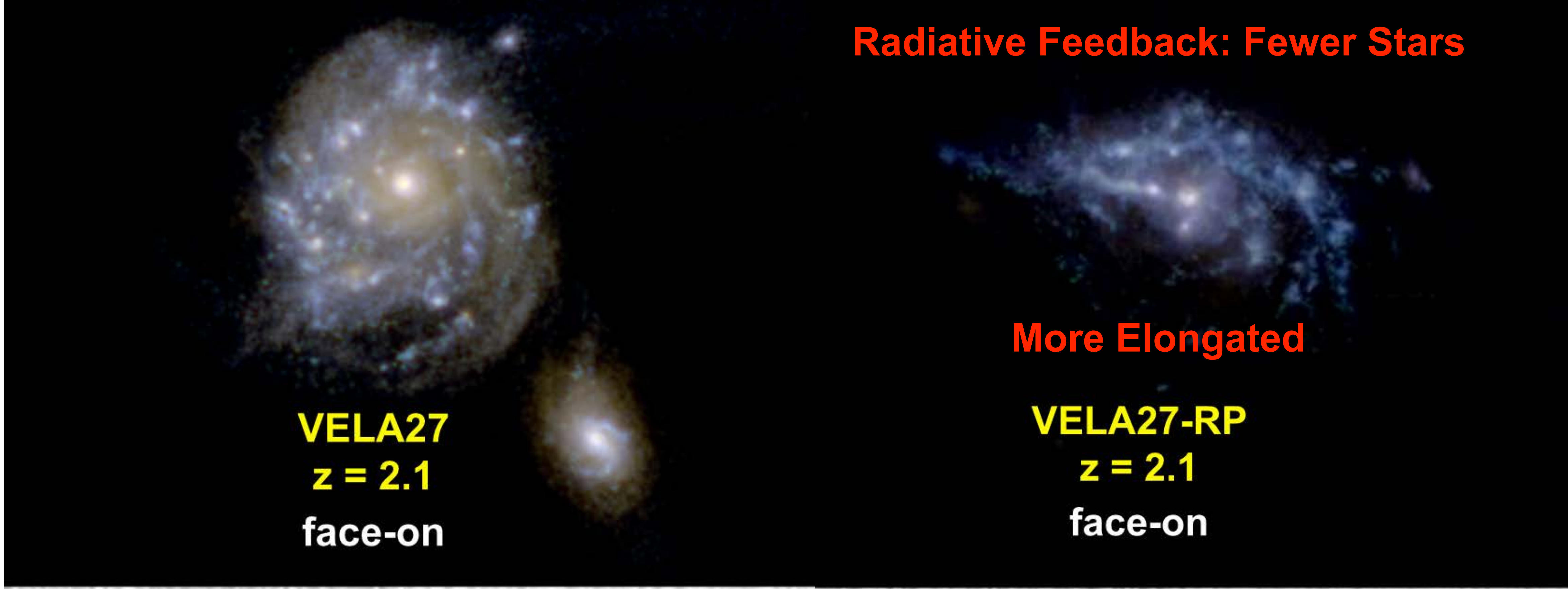
Ly alpha blobs from same simulation



Fumagalli, Prochaska, Kasen, Dekel, Ceverino, & Primack 2011

Simulated Galaxy 10 billion years ago

as it would appear nearby to our eyes



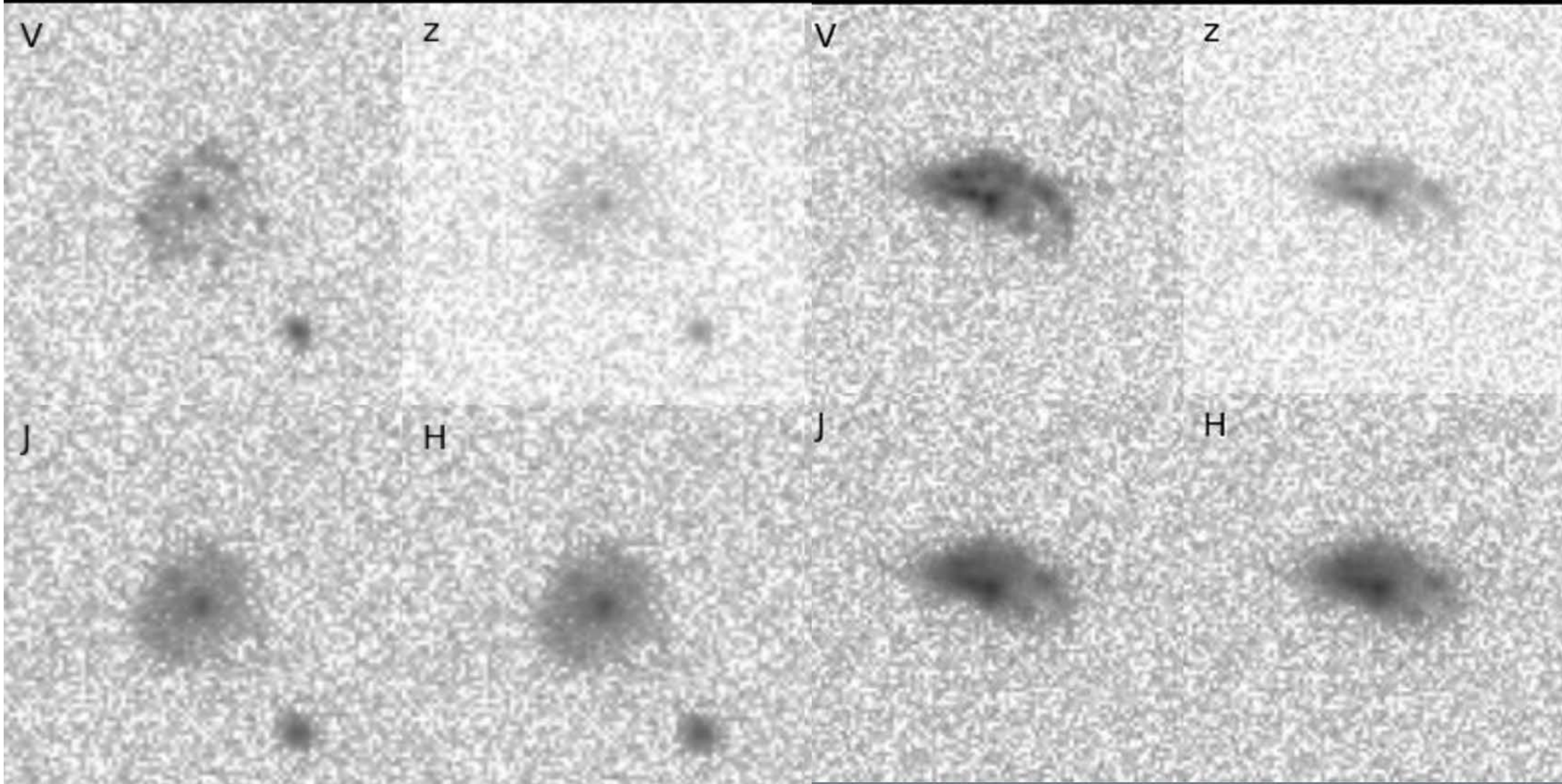
Radiative Feedback: Fewer Stars

More Elongated

CANDELized

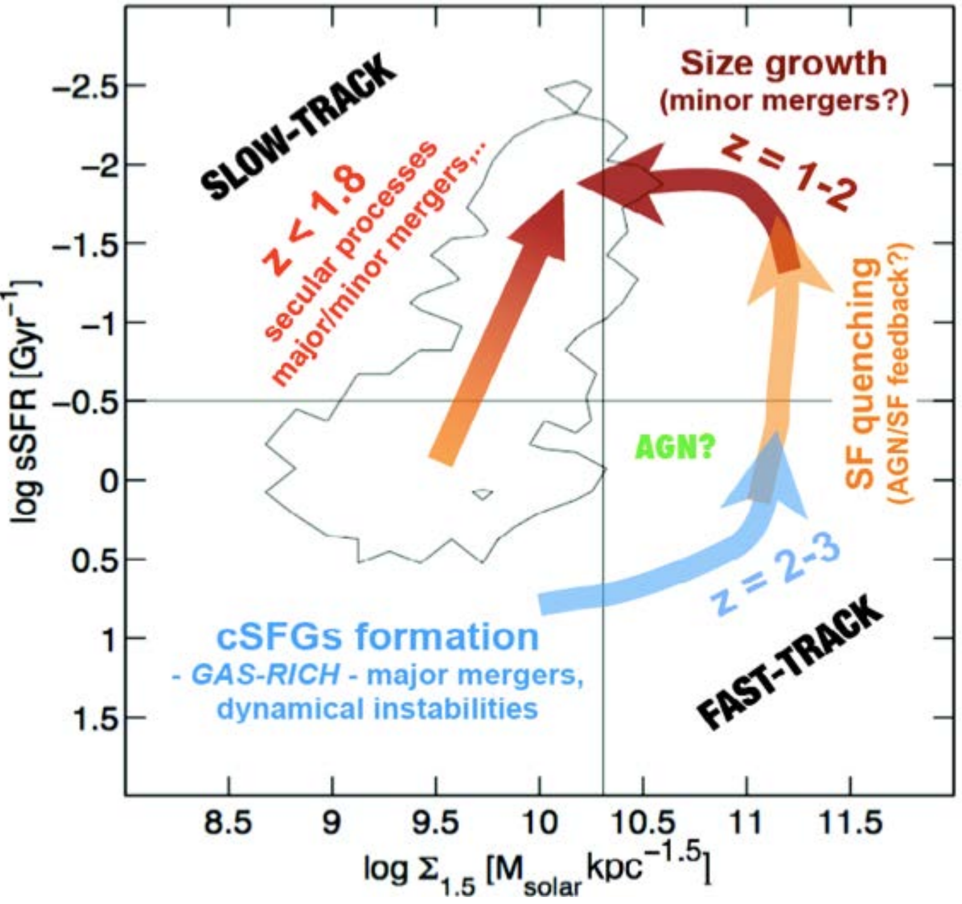
as it would appear to Hubble's ACS visual camera

as it would appear to Hubble's WFC3 infrared camera

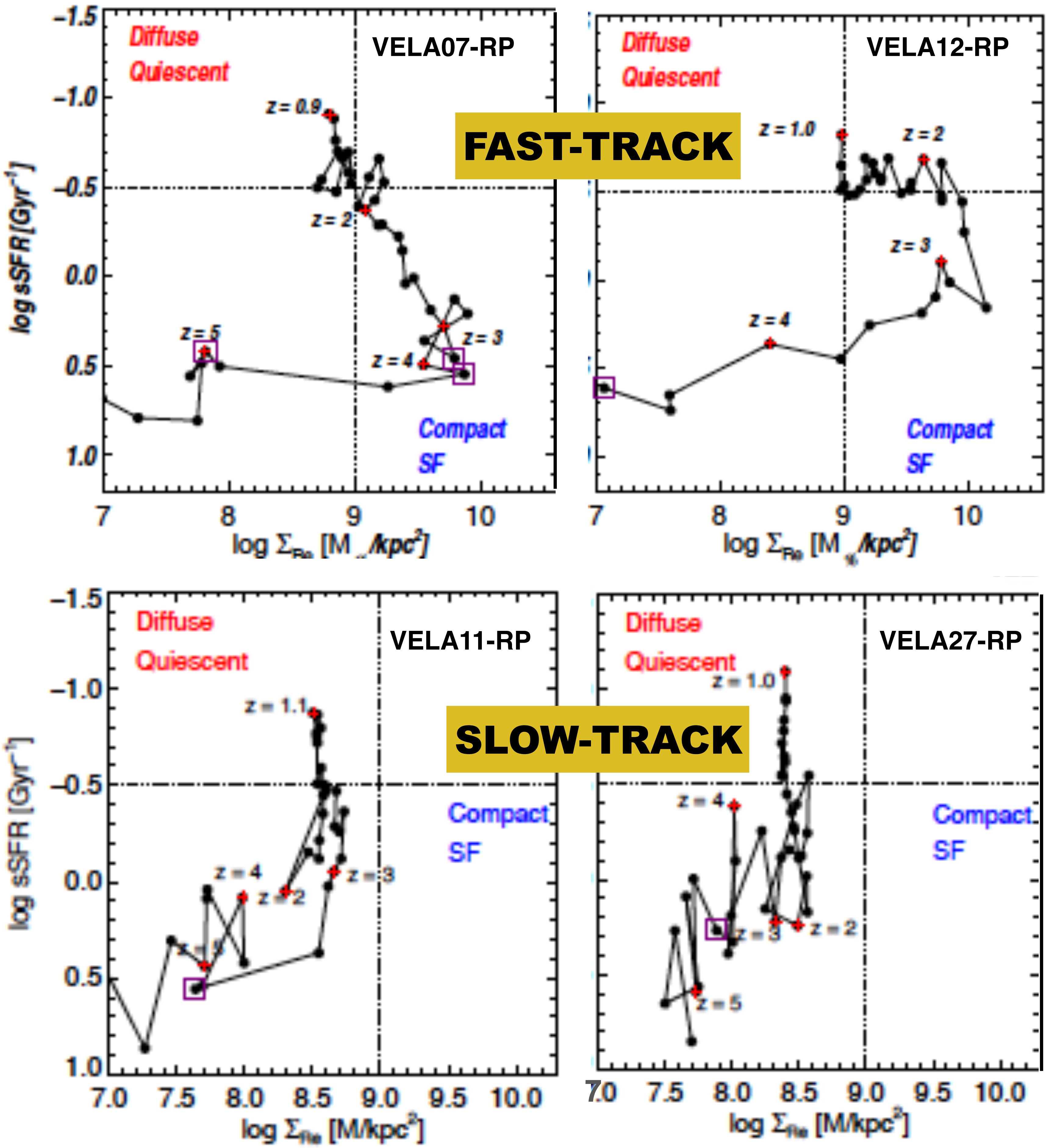


Ceverino+ RP simulations analyzed by Zolotov, Dekel, Tweed, Mandelker, Ceverino, & Primack MNRAS 2015

Barro+ (CANDELS) 2013



COMPACTION —>

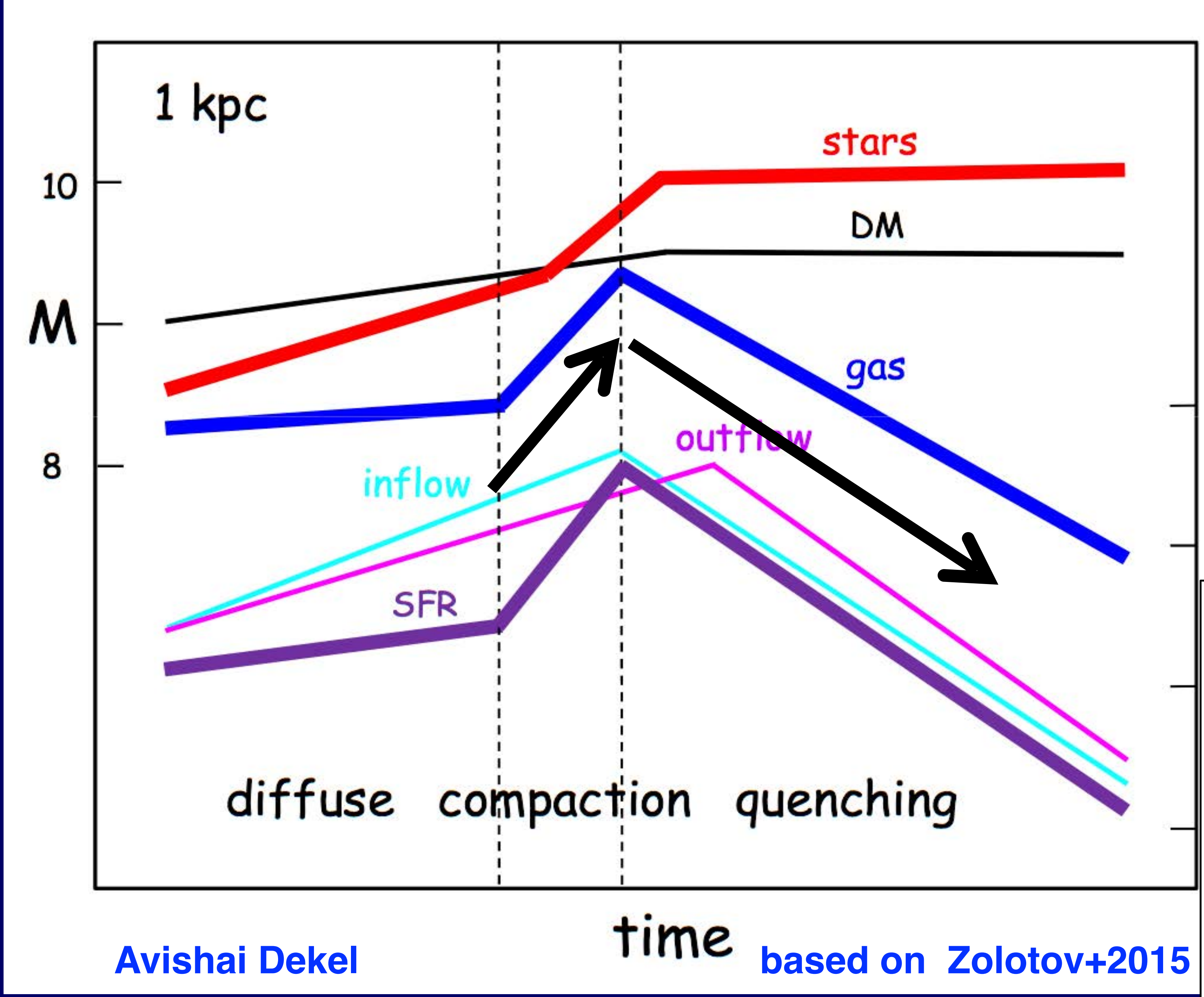


▣ minor merger

▣ major merger

Zolotov+2015

Compaction and Quenching in the Inner 1 kpc

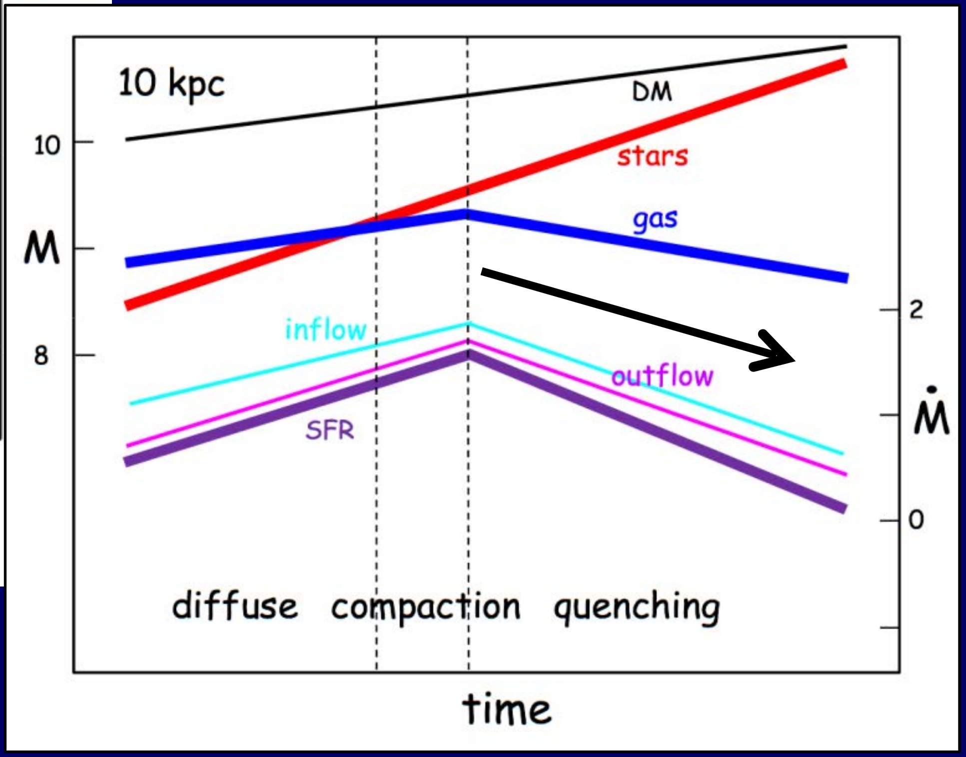


Avishai Dekel

based on Zolotov+2015

inner 1 kpc

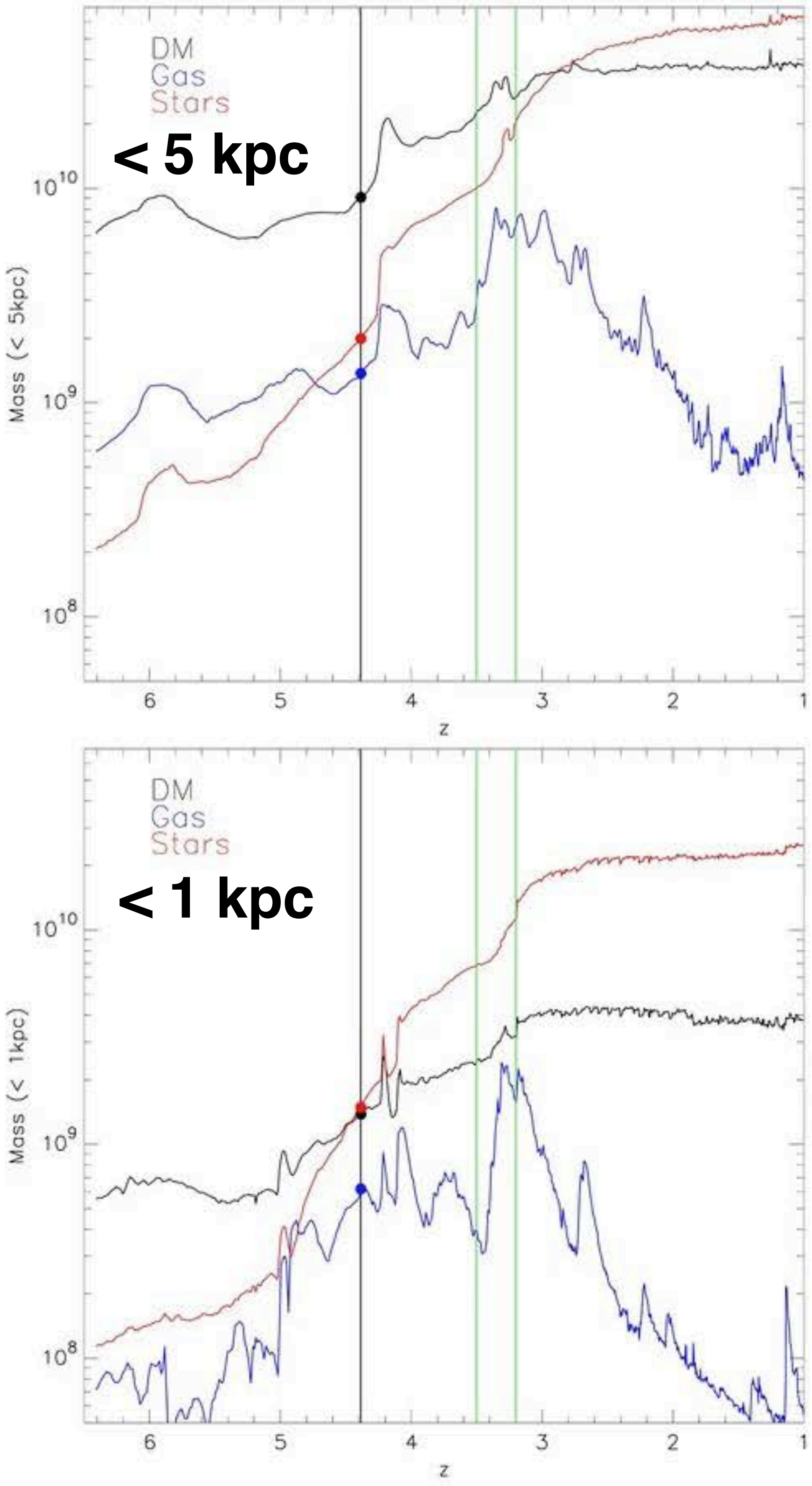
Inner 10 kpc



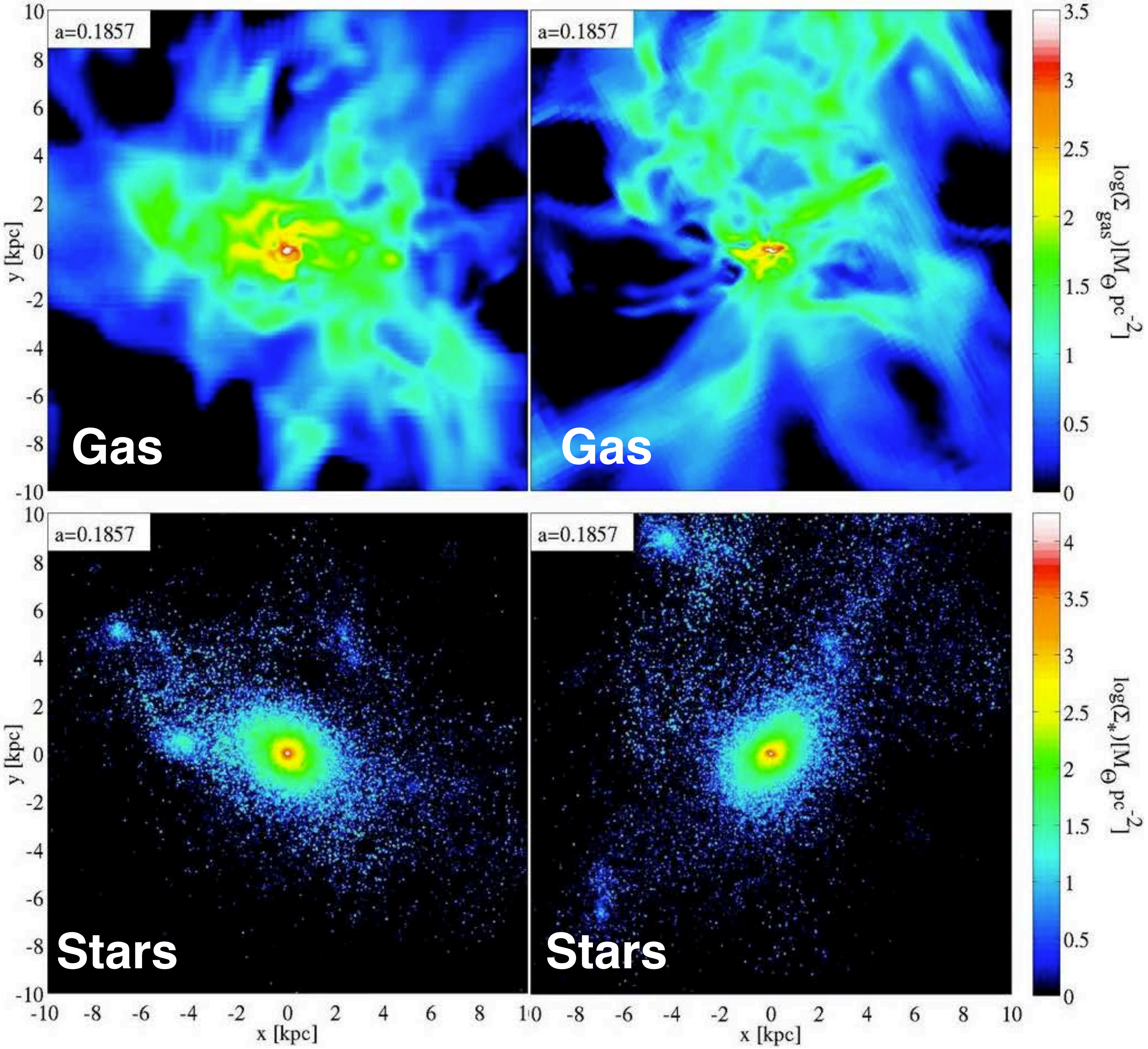
Gen 3 VELA07-RP Animations z = 4.4 to 2.3

DM
Gas
Stars

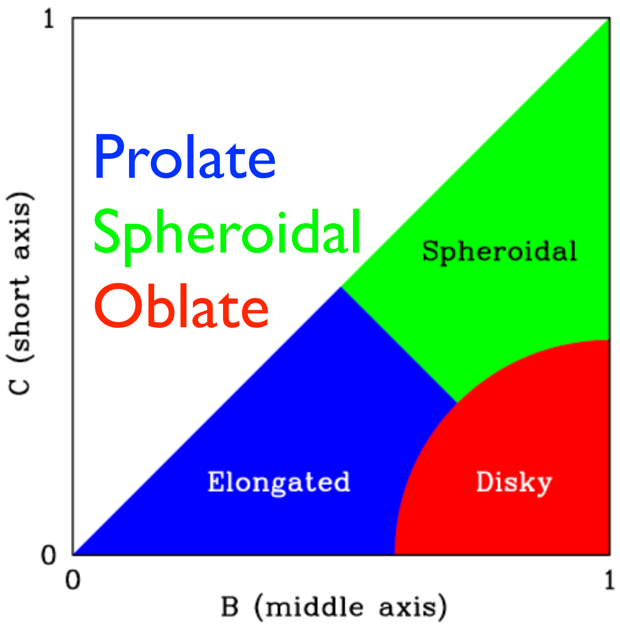
Compaction

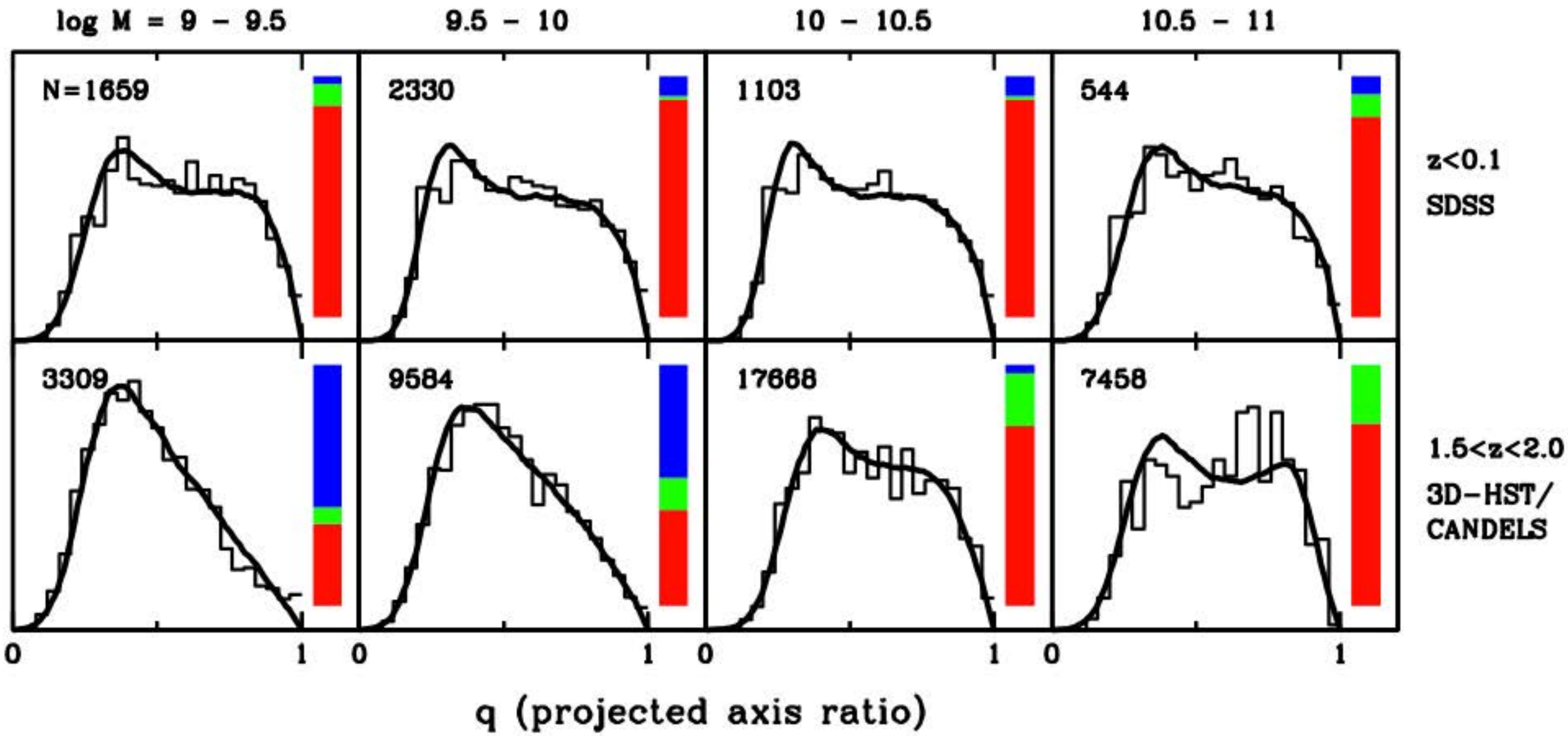


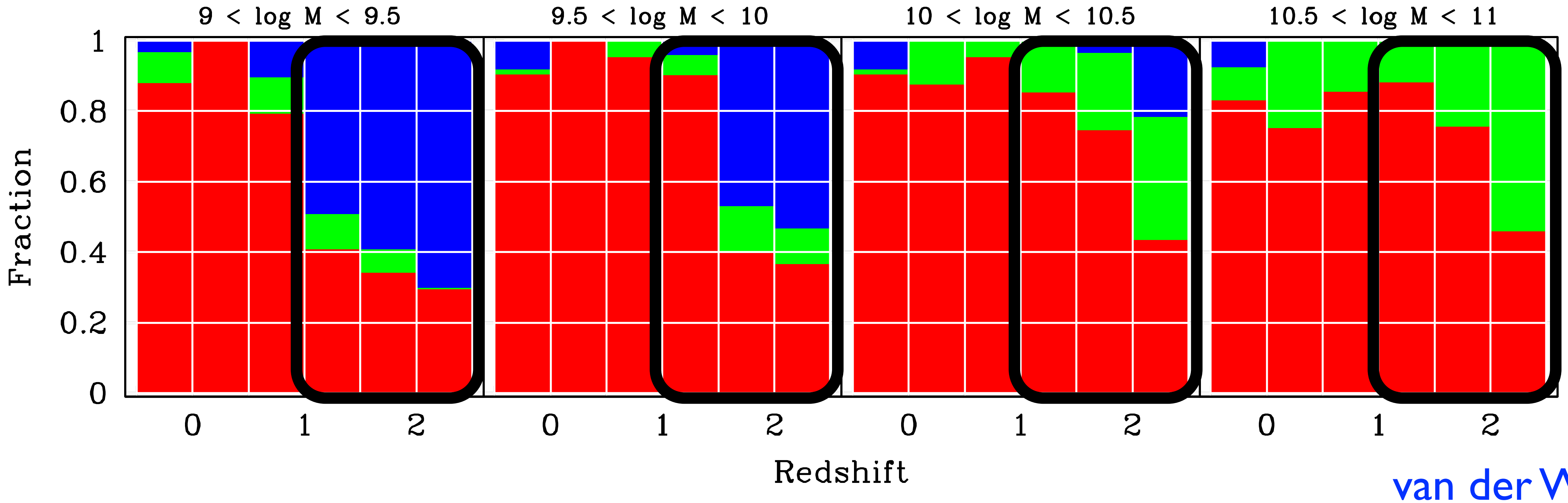
Face-on **Edge-on**



Prolate Galaxies Dominate at High Redshifts & Low Masses







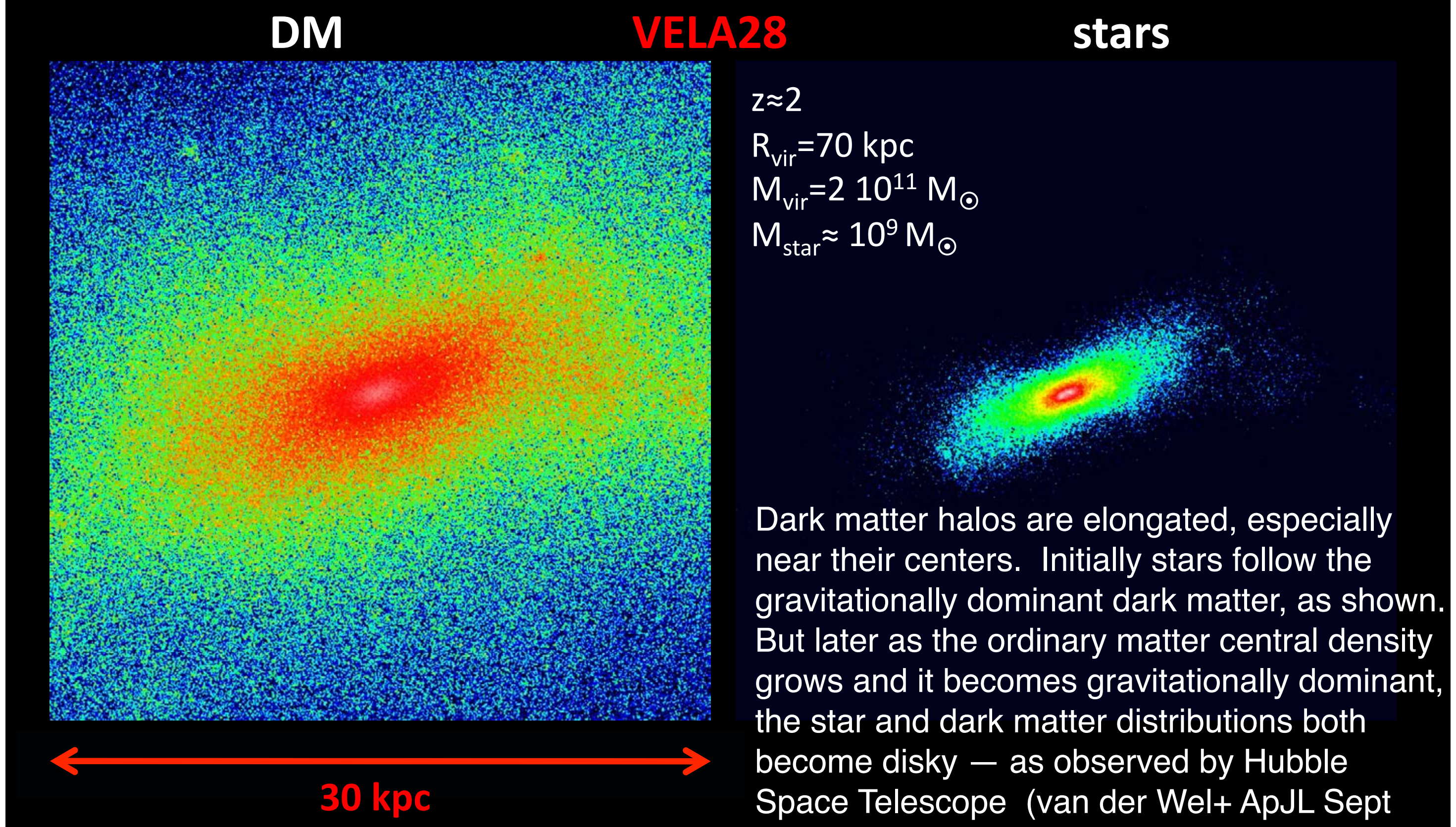
van der Wel+2014

See also Morphological Survey of Galaxies z=1.5-3.6 Law, Steidel+ ApJ 2012

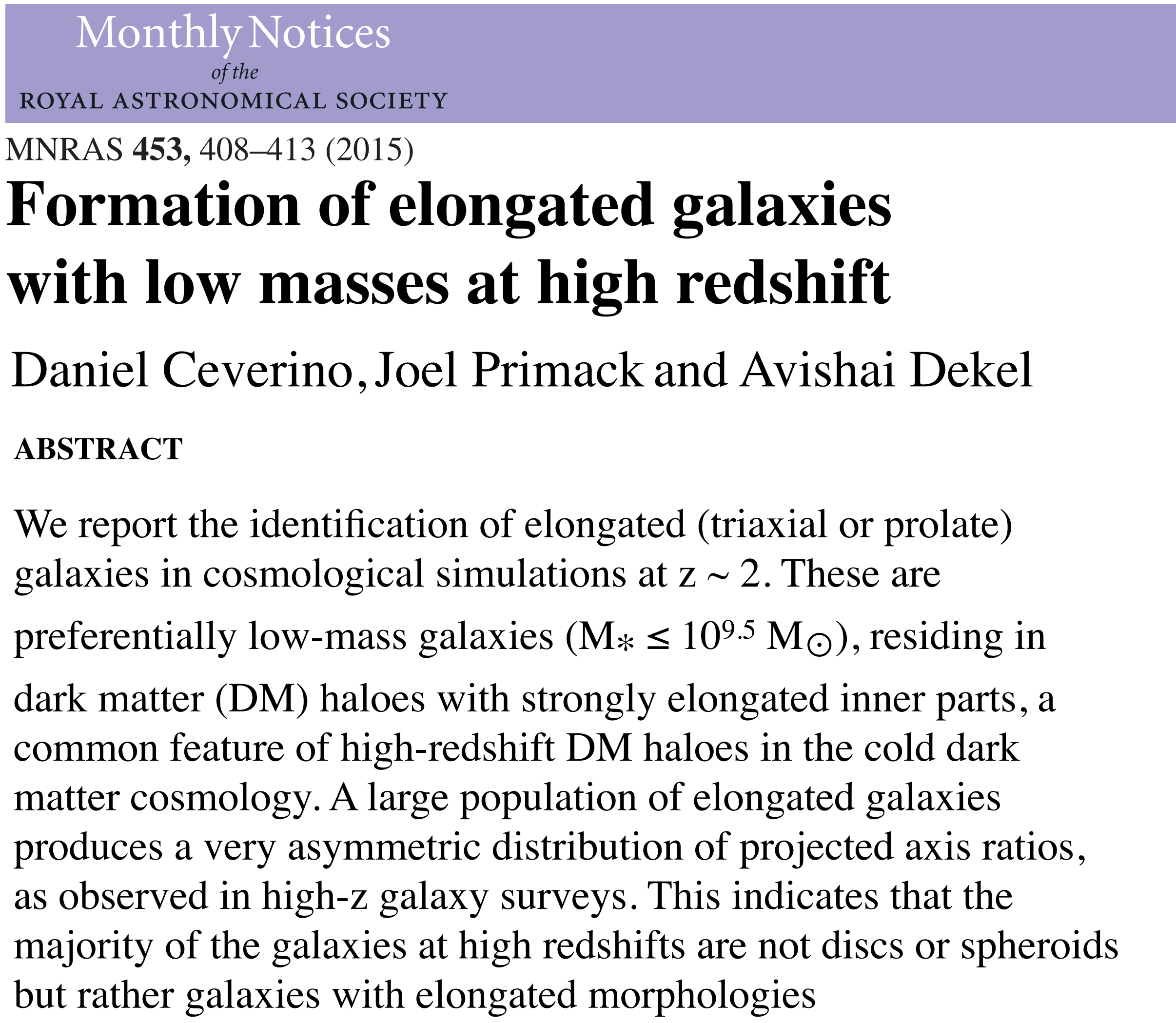
When Did Round Disk Galaxies Form? T. M. Takeuchi+ ApJ 2015

Our cosmological zoom-in simulations often produce elongated galaxies like observed ones. The elongated stellar distribution follows the elongated inner dark matter halo.

Prolate DM halo → elongated galaxy



Dark matter halos are elongated, especially near their centers. Initially stars follow the gravitationally dominant dark matter, as shown. But later as the ordinary matter central density grows and it becomes gravitationally dominant, the star and dark matter distributions both become disky — as observed by Hubble Space Telescope (van der Wel+ ApJL Sept 2014).



Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **453,** 408–413 (2015)

Formation of elongated galaxies with low masses at high redshift

Daniel Ceverino, Joel Primack and Avishai Dekel

ABSTRACT

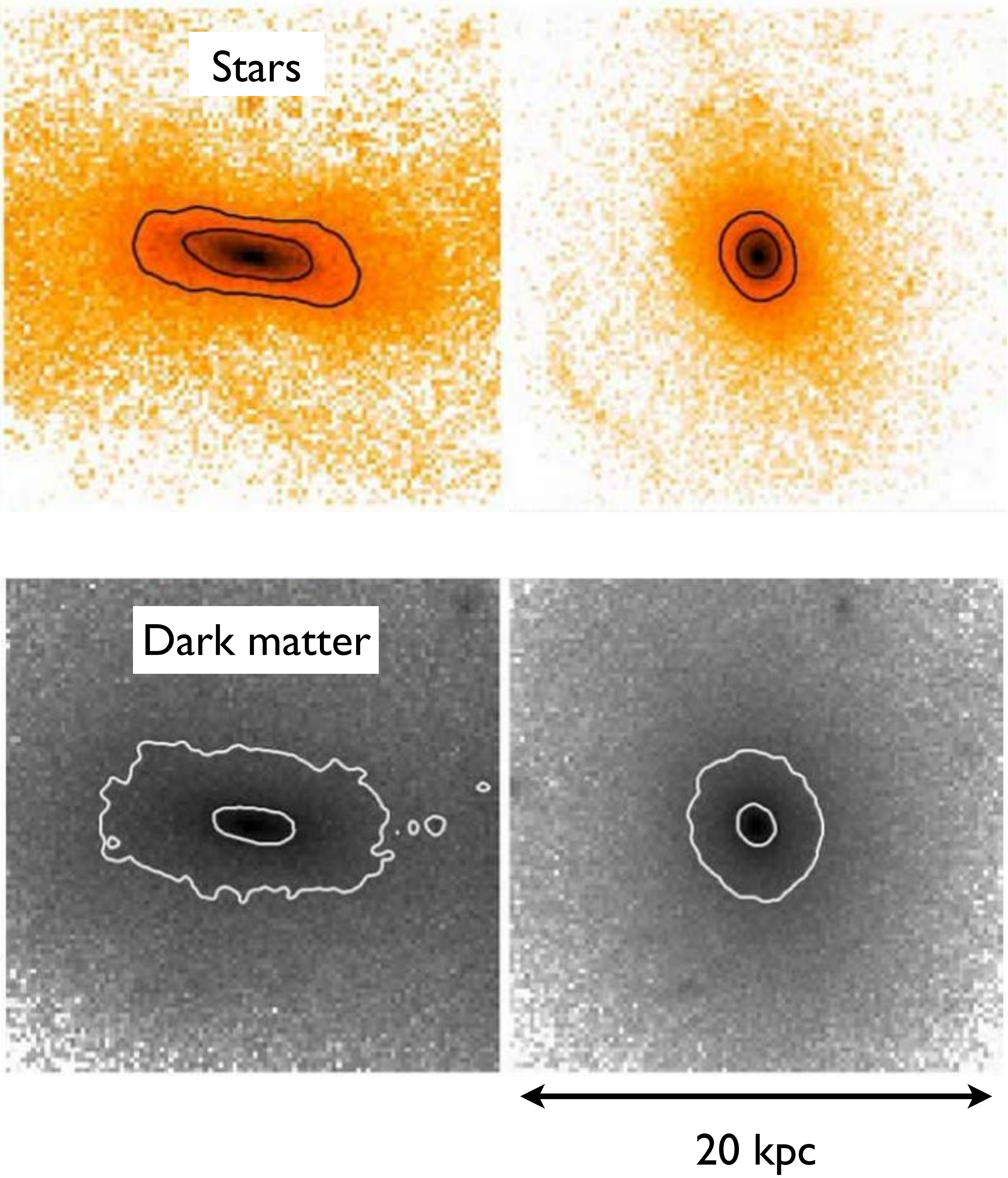
We report the identification of elongated (triaxial or prolate) galaxies in cosmological simulations at z ~ 2. These are preferentially low-mass galaxies ($M_* \leq 10^{9.5}$ $M_\odot$), residing in dark matter (DM) haloes with strongly elongated inner parts, a common feature of high-redshift DM haloes in the cold dark matter cosmology. A large population of elongated galaxies produces a very asymmetric distribution of projected axis ratios, as observed in high-z galaxy surveys. This indicates that the majority of the galaxies at high redshifts are not discs or spheroids but rather galaxies with elongated morphologies

Nearby large galaxies are mostly disks and spheroids — but they start out looking more like pickles.

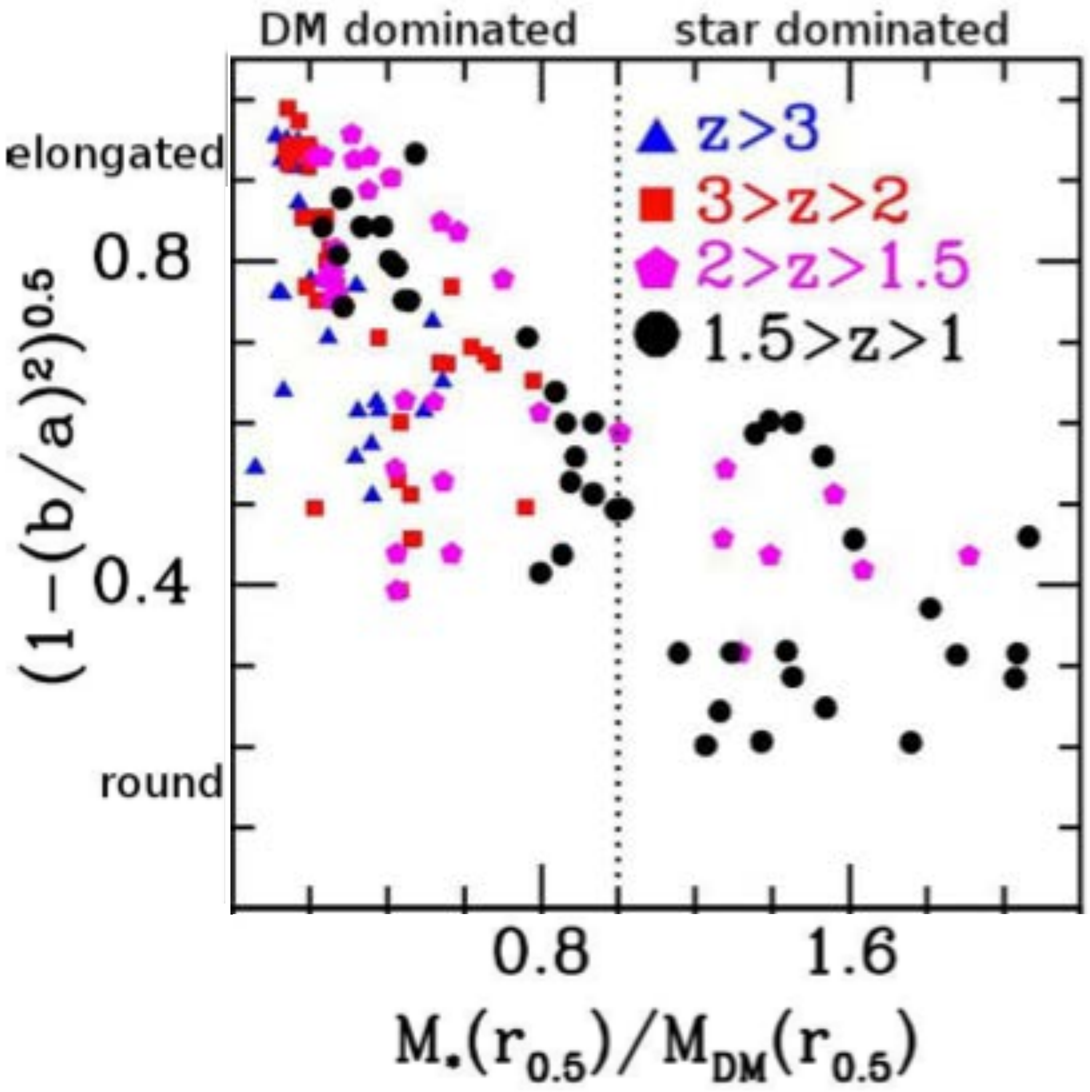
Formation of elongated galaxies with low masses at high redshift

Daniel Ceverino, Joel Primack and Avishai Dekel

MNRAS 2015



$M_* < 10^{10}\,M_\odot$ at z=2



Also Tomassetti et al. 2016 MNRAS

Simulated elongated galaxies are aligned with cosmic web filaments, become round after compaction (gas inflow to center)

How we are using galaxy simulations and deep learning to improve understanding of galaxy formation, with support from Google.

Sander Dieleman used a deep learning code to predict Galaxy Zoo nearby galaxy image classifications with 99% accuracy, winning 2014 Kaggle competition

Marc Huertas-Company used Dieleman's code to classify CANDELS galaxy images

H-C et al. 2015, Catalog of Visual-like Morphologies in 5 CANDELS Fields Using Deep Learning

H-C et al. 2016, Mass assembly and morphological transformations since z ~ 3 from CANDELS

Google supports Marc H-C's visits to UCSC Summer 2016 and 2017, and his grad student Fernando Caro's visit March-August 2017 using deep learning, CANDELS images, and Primack group's galaxy simulations to understand galaxy formation

UCSC group here today: Profs. David Koo, Joel Primack; grad students Fernando Caro, Christoph Lee, Viraj Pandya, astrophysics senior thesis student Sean Larkin

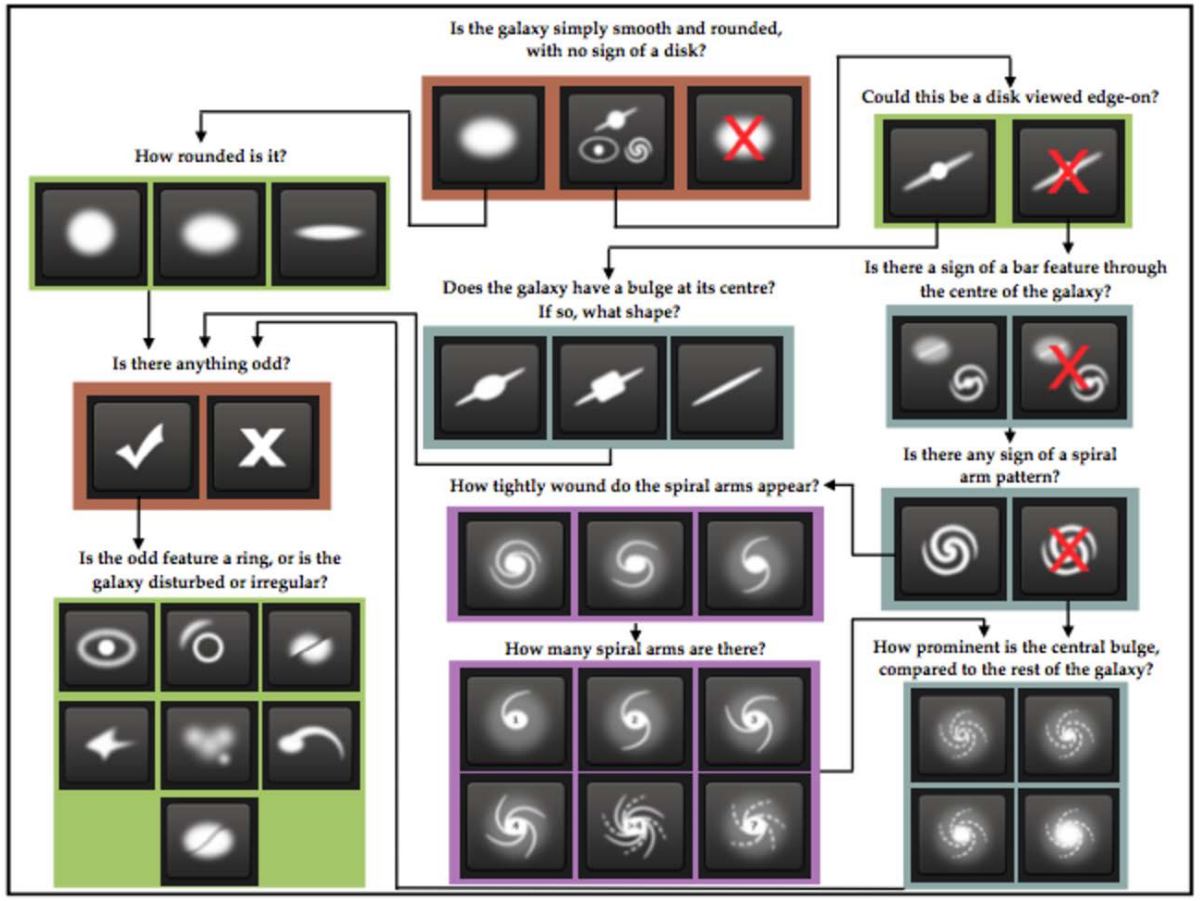
Related UCSC deep learning project: better galaxy environment estimates

Here today: Dr. Doug Hellinger, grad students James Kakos, Dominic Pasquali

Related UCSC deep learning project: damped Lyα systems in SDSS spectra

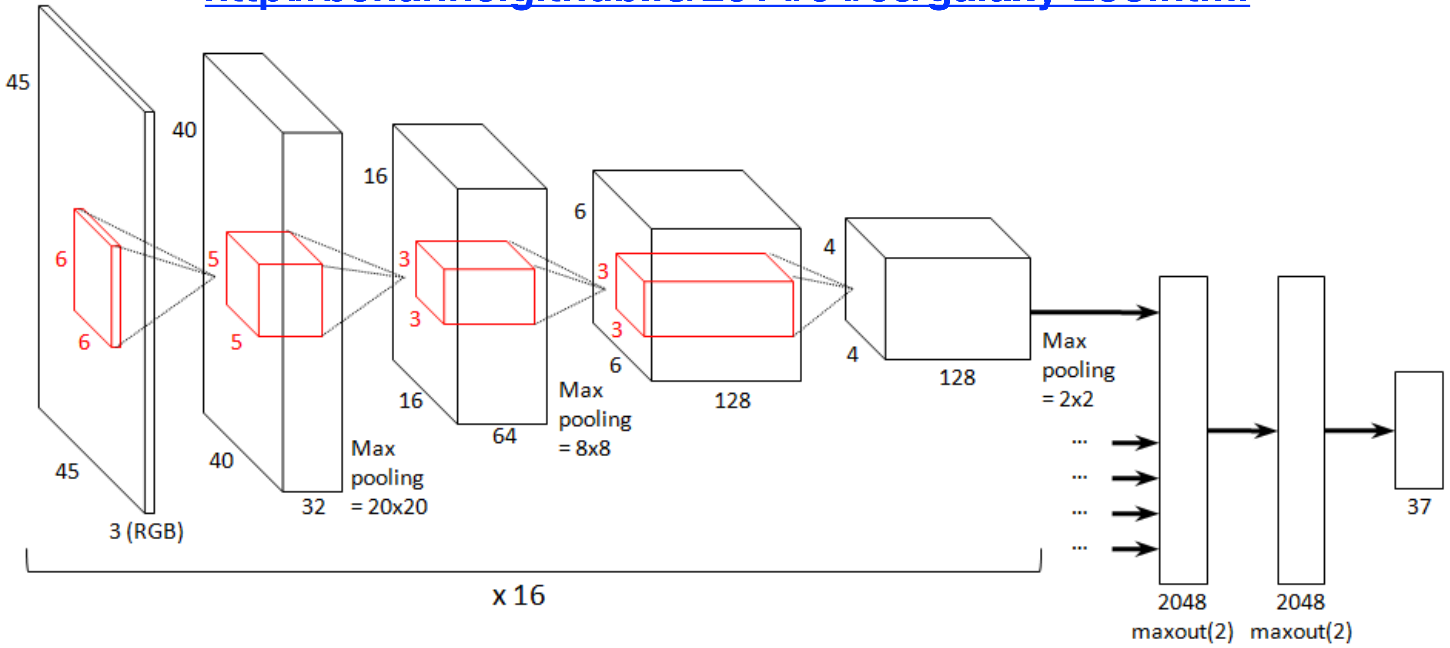
Here today: Dr. Shawfeng Dong

Sander Dieleman used a deep learning code to predict Galaxy Zoo nearby galaxy image classifications with 99% accuracy, winning 2014 Kaggle competition



The Galaxy Zoo 2 decision tree. Reproduced from fig.1 in Willett et al. (2013).

http://benanne.github.io/2014/04/05/galaxy-zoo.html



Krizhevsky-style diagram of the architecture of the best performing network.

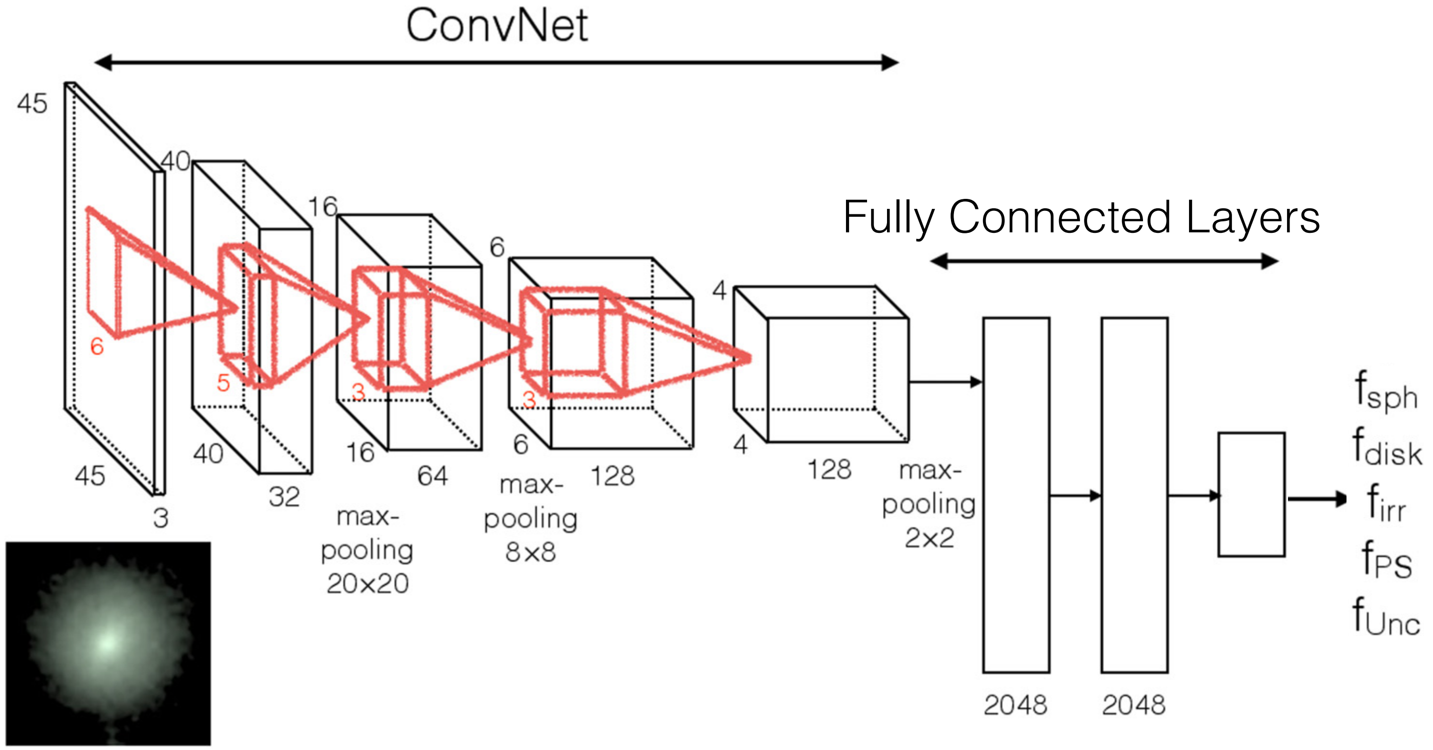
Dieleman, Willett, Dambre 2015, Rotation-invariant convolutional neural networks for galaxy morphology prediction, MNRAS

From the ABSTRACT: The Galaxy Zoo project successfully applied a crowdsourcing strategy, inviting online users to classify images by answering a series of questions. We present a deep neural network model for galaxy morphology classification which exploits translational and rotational symmetry. For images with high agreement among the Galaxy Zoo participants, our model is able to reproduce their consensus with near-perfect accuracy (>99 per cent) for most questions.

Marc Huertas-Company used Dieleman’s code to classify CANDELS galaxy images

H-C et al. 2015, Catalog of Visual-like Morphologies in 5 CANDELS Fields Using Deep Learning

In this work, we mimic human perception with deep learning using convolutional neural networks (ConvNets). The ConvNet is trained to reproduce the CANDELS visual morphological classification based on the efforts of 65 individual classifiers who contributed to the visual inspection of all of the galaxies in the GOODS-S field. It was then applied to the other four CANDELS fields. The galaxy classification data was then released to the astronomical community.



Configuration of the Convolutional Neural Network used in this paper, based on the one used by Dieleman et al. (2015) on SDSS galaxies. It is made of 5 convolutional layers followed by 2 fully connected perceptron layers.

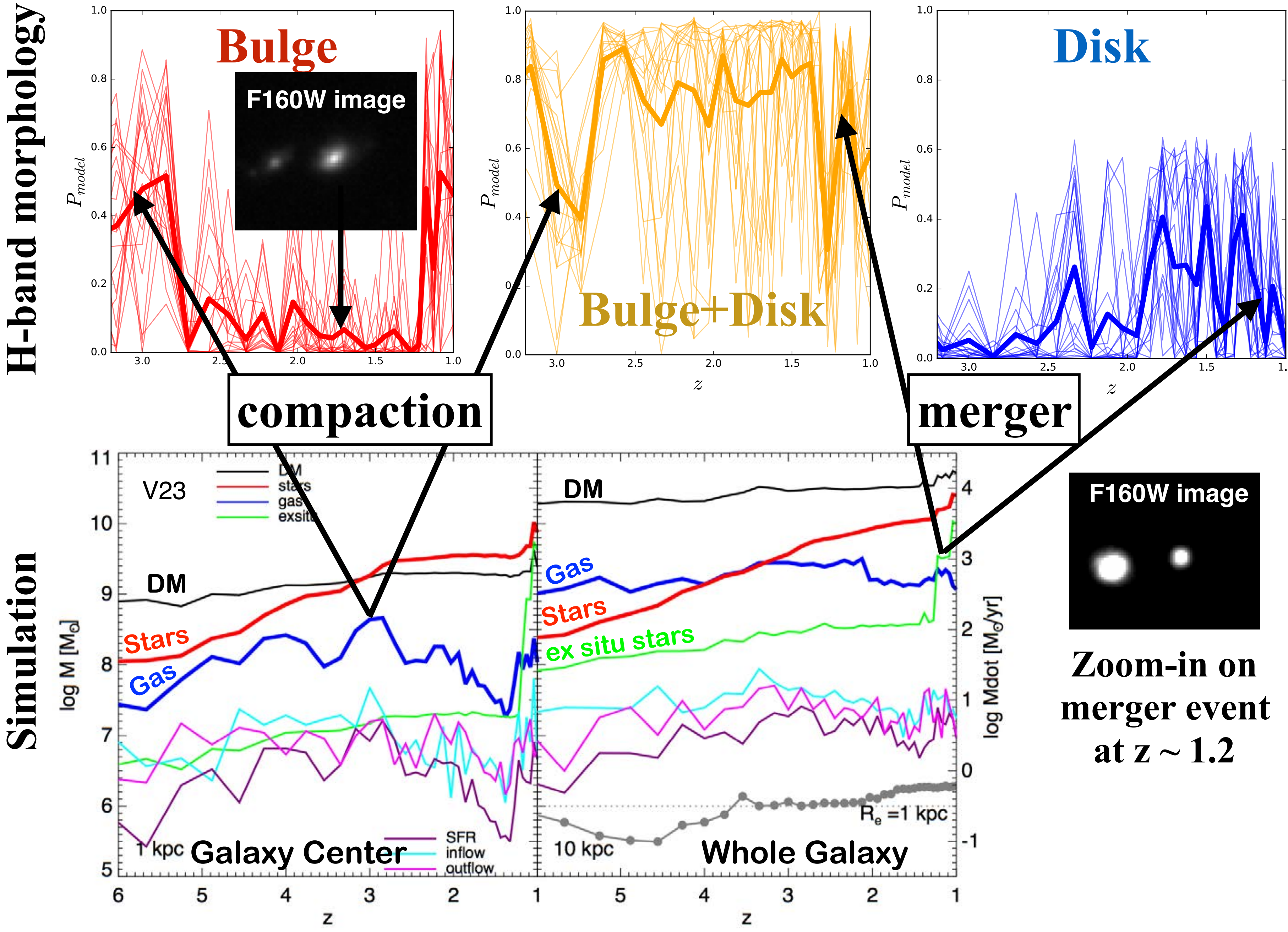
Following the approach in CANDELS, we associate five real numbers with each galaxy corresponding to the frequency at which expert classifiers flagged a galaxy as having a bulge, having a disk, presenting an irregularity, being compact or point-source, and being unclassifiable. Galaxy images are interpolated to a fixed size, rotated, and randomly perturbed before feeding the network to (i) avoid over-fitting and (ii) reach a comparable ratio of background versus galaxy pixels in all images. ConvNets are able to predict the votes of expert classifiers with a <10% bias and a ~10% scatter. This makes the classification almost equivalent to a visual-based classification. The training took 10 days on a GPU and the classification is performed at a rate of 1000 galaxies/hour.

H-C et al. 2016, Mass assembly and morphological transformations since $z \sim 3$ from CANDELS

From the ABSTRACT: We quantify the evolution of star-forming and quiescent galaxies as a function of morphology from $z \sim 3$ to the present. Our main results are: 1) At $z \sim 2$, 80% of the stellar mass density of star-forming galaxies is in irregular systems. However, by $z \sim 0.5$, irregular objects only dominate at stellar masses below $10^9 M_\odot$.

2) Quenching: We confirm that galaxies reaching a stellar mass $M_* \sim 10^{10.8} M_\odot$ tend to quench. Also, quenching implies the presence of a bulge: the abundance of massive red disks is negligible at all redshifts

Google supports Marc H-C's visits to UCSC Summer 2016 and 2017, and his grad student Fernando Caro's visit March-August 2017 using deep learning, CANDELS images, and Primack group's galaxy simulations to understand galaxy formation

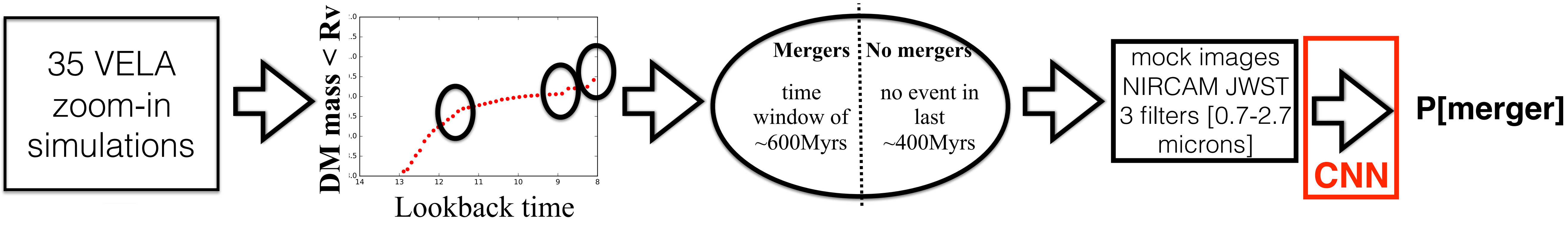


Zoom-in on
merger event
at z ~ 1.2

Evolution of zoom-in galaxy simulation VELA23-RP. The upper three panels show the probabilities that the galaxy is best fit by GALFIT as a single-Sersic Bulge or Disk, or instead as a double Sersic Bulge+Disk, based on classifications by a deep learning code trained using synthetic images. (Note that these probabilities do not need to sum to unity, since they are independent.) Classifications are plotted for 20 different orientations, with the medians plotted as heavy lines.

Google supports Marc H-C's visits to UCSC Summer 2016 and 2017, and his grad student Fernando Caro's visit March-August 2017 using deep learning, CANDELS images, and Primack group's galaxy simulations to understand galaxy formation

UCSC group here today: Profs. David Koo, Joel Primack; grad students Fernando Caro, Christoph Lee, Viraj Pandya, astrophysics senior thesis student Sean Larkin


35 VELA zoom-in simulations

Mergers | **No mergers**

time window of ~600Myrs | no event in last ~400Myrs

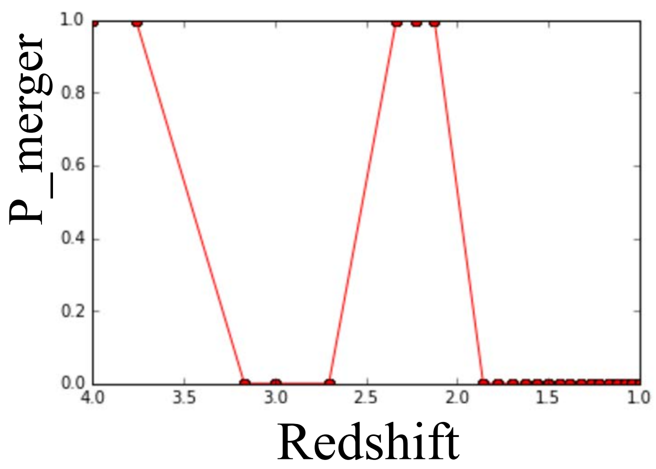
mock images NIRCAM JWST 3 filters [0.7-2.7 microns]

CNN

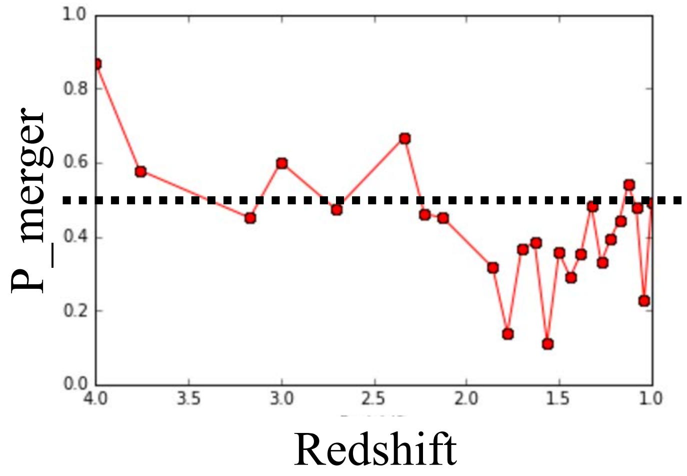
P[merger]

Example of 1 simulation

INPUT

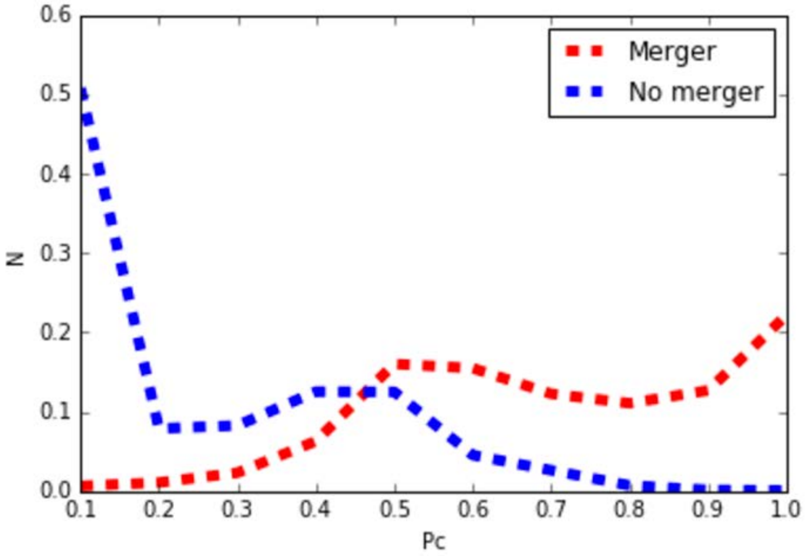


OUTPUT



This is an oversimplified example, where we just used the total dark matter mass within the halo radius R_V to estimate when a major merger occurred. We are now analyzing the entire satellite galaxy population to determine when major and minor mergers and satellite fly-bys occur.

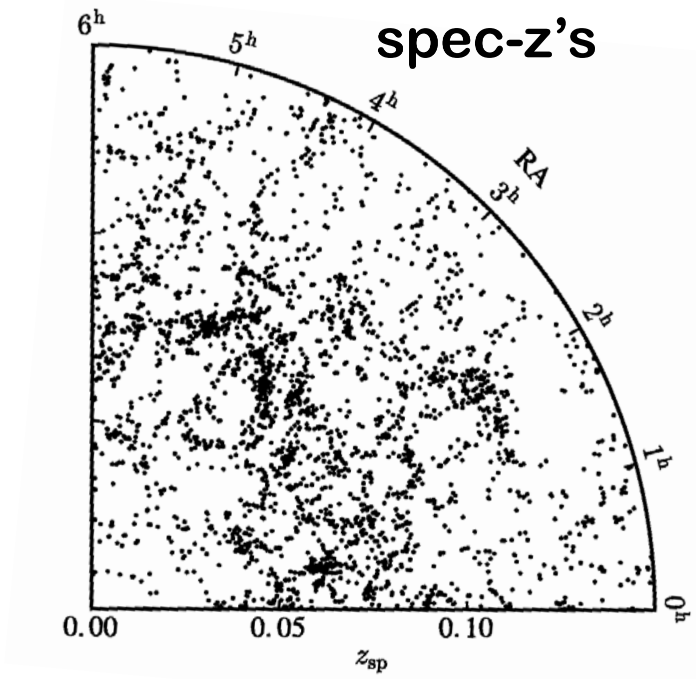
Applied to all 35 simulations



Related UCSC deep learning project: better galaxy environment estimates

Here today: Dr. Doug Hellinger, grad students James Kakos, Dominic Pasquali

Images at various wavelengths (=>photometric redshifts, photo-z's) are much more plentiful than spectroscopic redshifts. How can we best combine a few spectroscopic z's with many photo-z's to estimate the environment of each galaxy? A preprint by Nicholas Tejos, Aldo Rodriguez-Puebla, and Joel Primack introduces a method ("sort") to do this. Can deep learning do better?


spec-z's

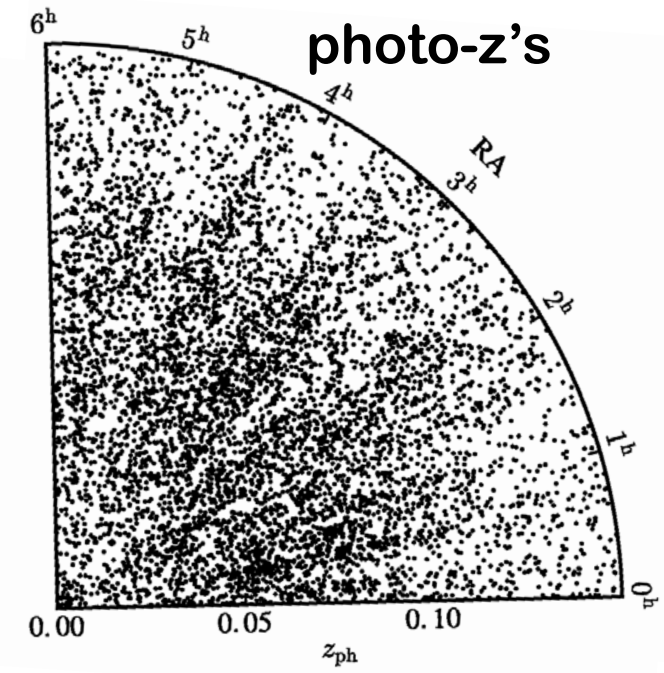
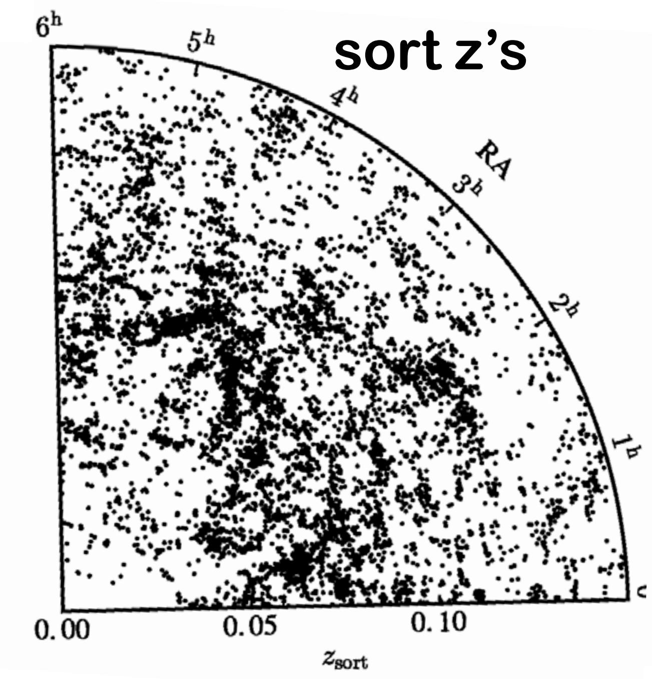
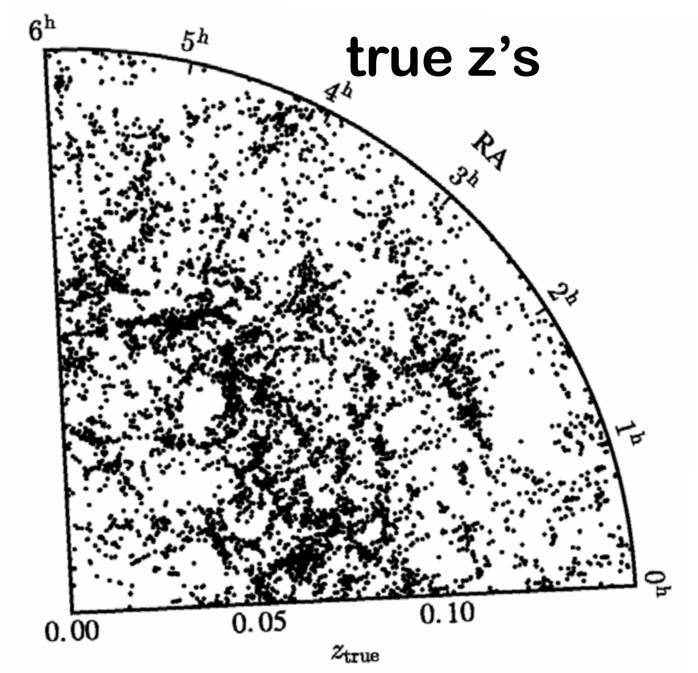
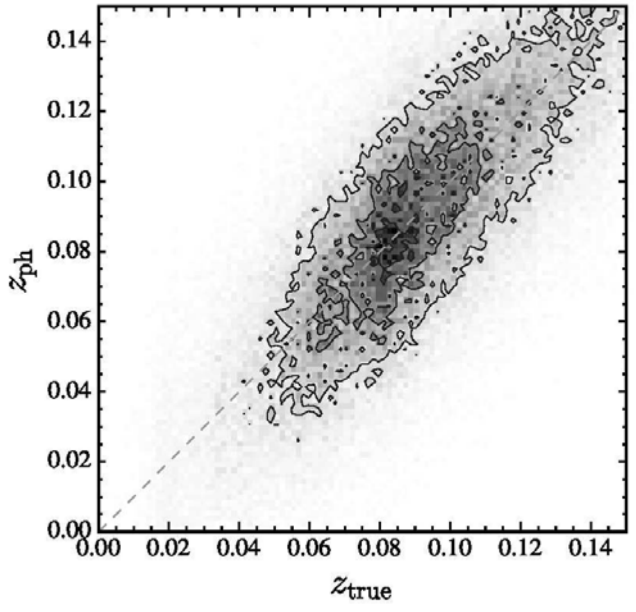
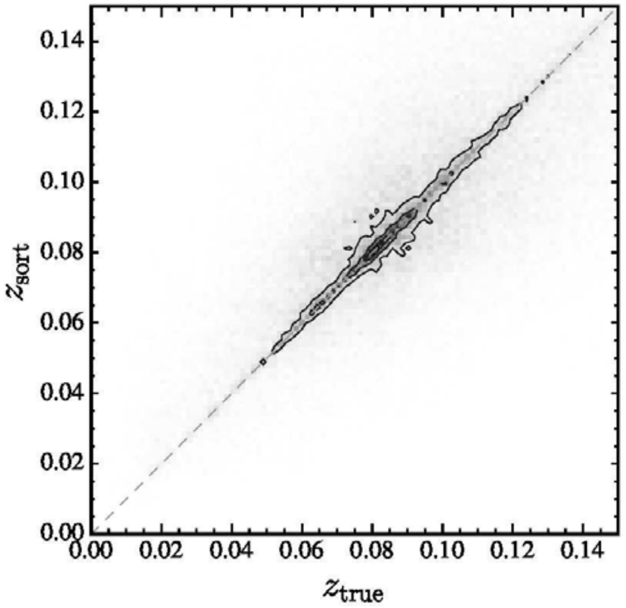


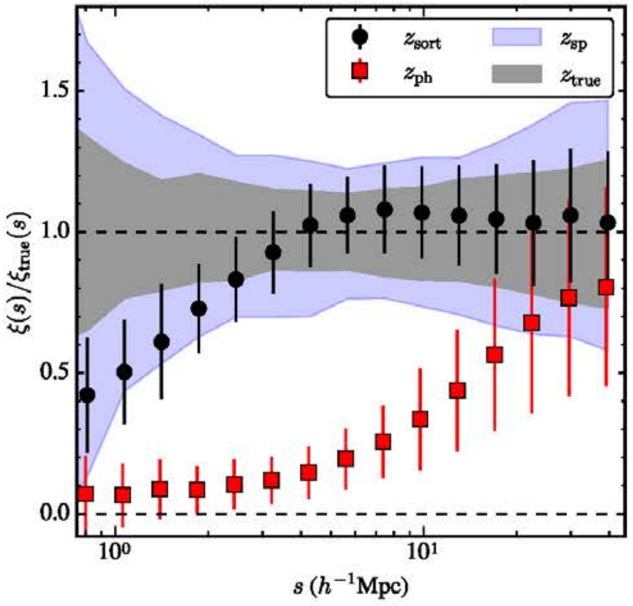
photo-z's


sort z's


true z's





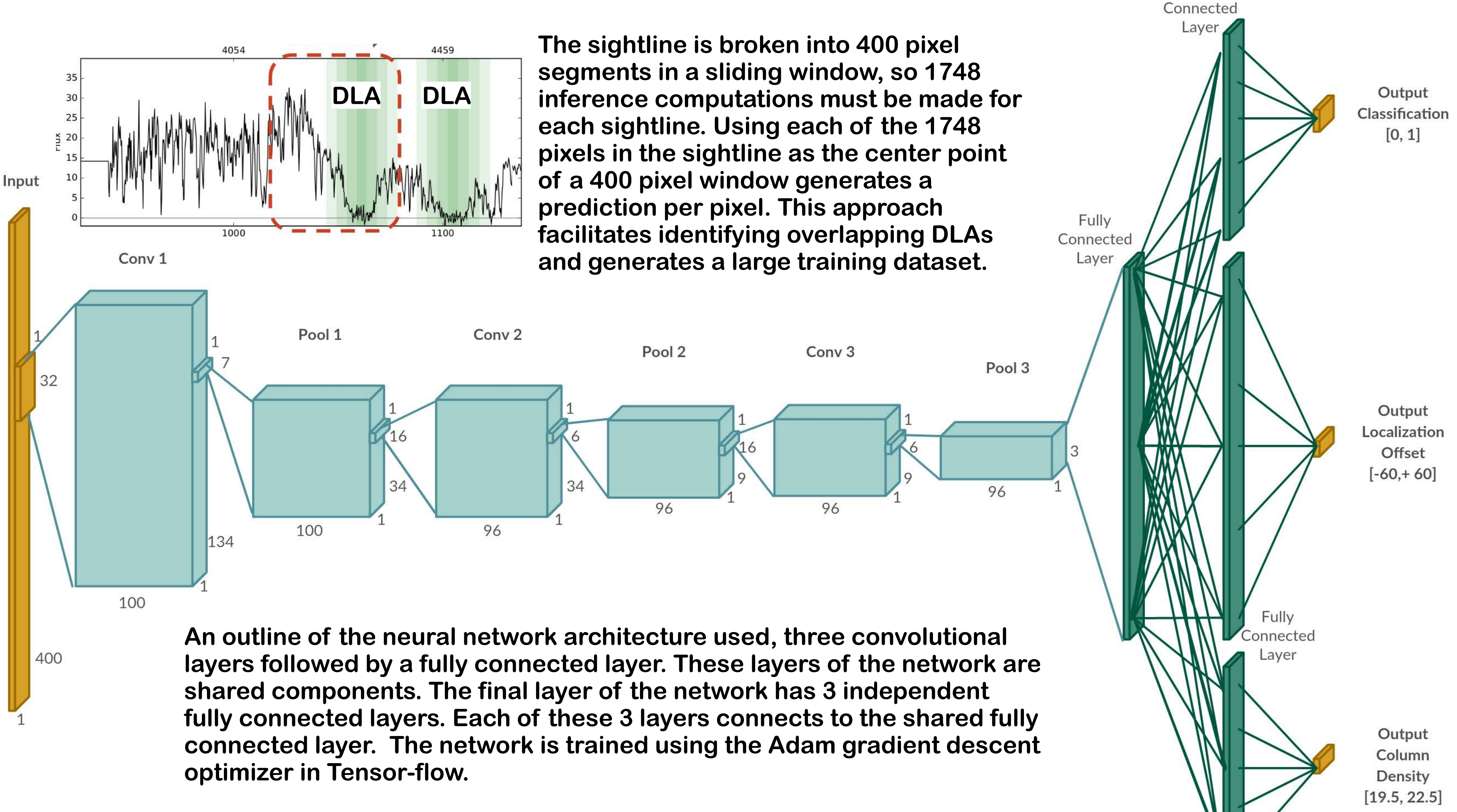


Ratio of measured and true 2-point correlation function as a function of redshift space distance s. Sort gets it right for s > 4 h^{-1} Mpc, while photo-z's fail even at s > 40 h^{-1} Mpc.

Related UCSC deep learning project: damped Lyα (DLA) systems in SDSS spectra

Here today: Dr. Shawfeng Dong; co-authors David Park, Prof. J. Xavier Prochaska, Dr. Zheng Cai

DLA systems seen in quasar spectra, corresponding to at least 2×10^{20} hydrogen atoms/cm^2, represent most of the neutral hydrogen in the universe at redshifts $z = 2$ to 4. About 7000 DLAs were identified by astronomers in about 100,000 quasar spectra. The additional 270,000 sightlines that recently became available from the Sloan Digital Sky Survey were scanned for DLAs by a deep learning code, and the resulting DLA catalog will be made publicly available.



The sightline is broken into 400 pixel segments in a sliding window, so 1748 inference computations must be made for each sightline. Using each of the 1748 pixels in the sightline as the center point of a 400 pixel window generates a prediction per pixel. This approach facilitates identifying overlapping DLAs and generates a large training dataset.

An outline of the neural network architecture used, three convolutional layers followed by a fully connected layer. These layers of the network are shared components. The final layer of the network has 3 independent fully connected layers. Each of these 3 layers connects to the shared fully connected layer. The network is trained using the Adam gradient descent optimizer in Tensor-flow.

Talk at Google 30 May 2017

New Insights on Galaxy Formation from Comparing Simulations and Observations

Joel Primack

Distinguished Professor of Physics Emeritus, UCSC

Brief introduction to modern cosmology, based on ΛCDM: dark energy & dark matter

Cosmic large scale structure simulations and star formation in galaxies

Comparing high-resolution hydrodynamic galaxy simulations with observations

Astronomers used to think that galaxies form as disks, that forming galaxies are pretty smooth, and that galaxies generally grow in radius as they grow in mass — but Hubble Space Telescope data show that all these statements are false, and our simulations may explain why.

We are using these simulations and deep learning to improve understanding of galaxy formation, with support from Google.