Cosmology @EOSC
the HPC Universe in the Cloud

Francisco Prada
IAA-CSIC

Our view of the distant Universe:

Hubble Ultra-deep Field

Is the visible material all there is?

Dark Cosmology

Much more gravity than we would expect:

➡ Dark Matter (27%)

Galaxies moving apart faster and faster:

➡ Dark Energy (68%)

New Physics, beyond the Standard Model

Observation and Theory interplay

Dark matter and dark energy have predictable effects on:

- Cosmic Microwave Background
- Patterns of galaxies
- Bending of light in Universe
- Brightness of supernovae

The Nobel Prize in Physics 2011

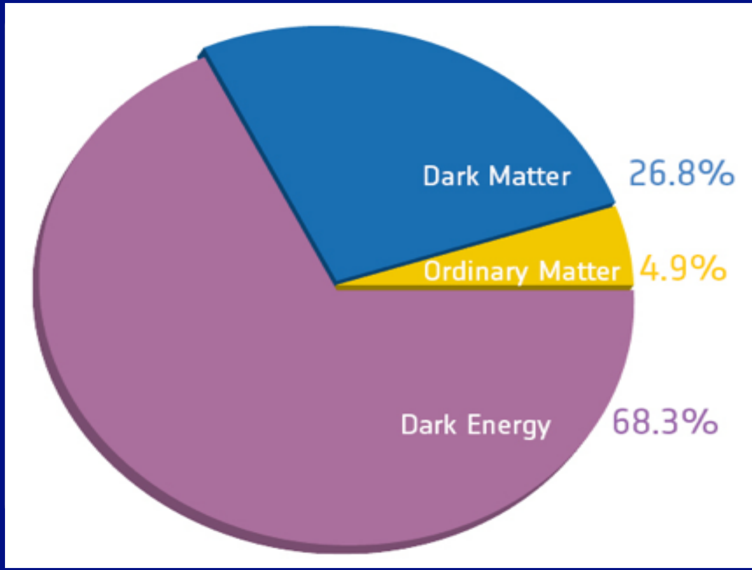
➡ Measure these effects in our large surveys

➡ Compare with theories and simulations containing dark Physics

Top Scientific Objectives

Physics of the Universe

Understanding Scientific Principles



The two highest level questions in the field are the following:

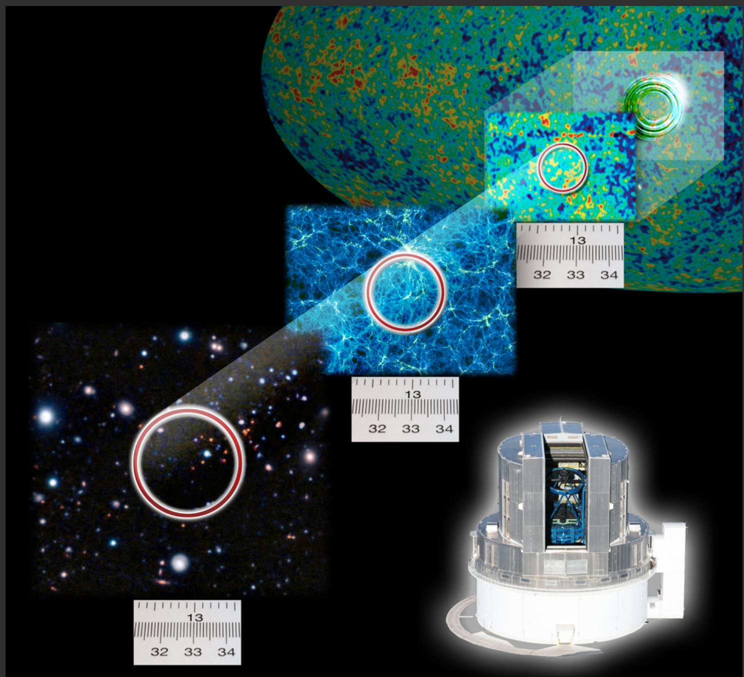
- Is cosmic acceleration caused by a breakdown of Einstein General Relativity on cosmological scales, or is it caused by a new energy component with negative pressure ("dark energy") within General Relativity?
- If the acceleration is caused by "dark energy," is its energy density constant in space and time and thus consistent with quantum vacuum energy or does its energy density evolve in time and/or vary in space?

From surveys to measurements of dark physics

- ➡ Measure statistics of galaxy properties in large surveys
- ➡ Compare with predictions for particular dark energy physical values

e.g. Baryonic Acoustic Oscillations (BAO):

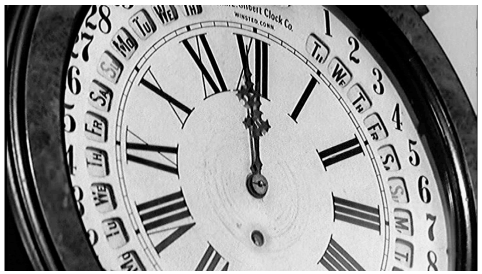
Measure correlation function of galaxy positions – BAO peak is found at a physical scale that depends on dark energy!



Large Survey Projects

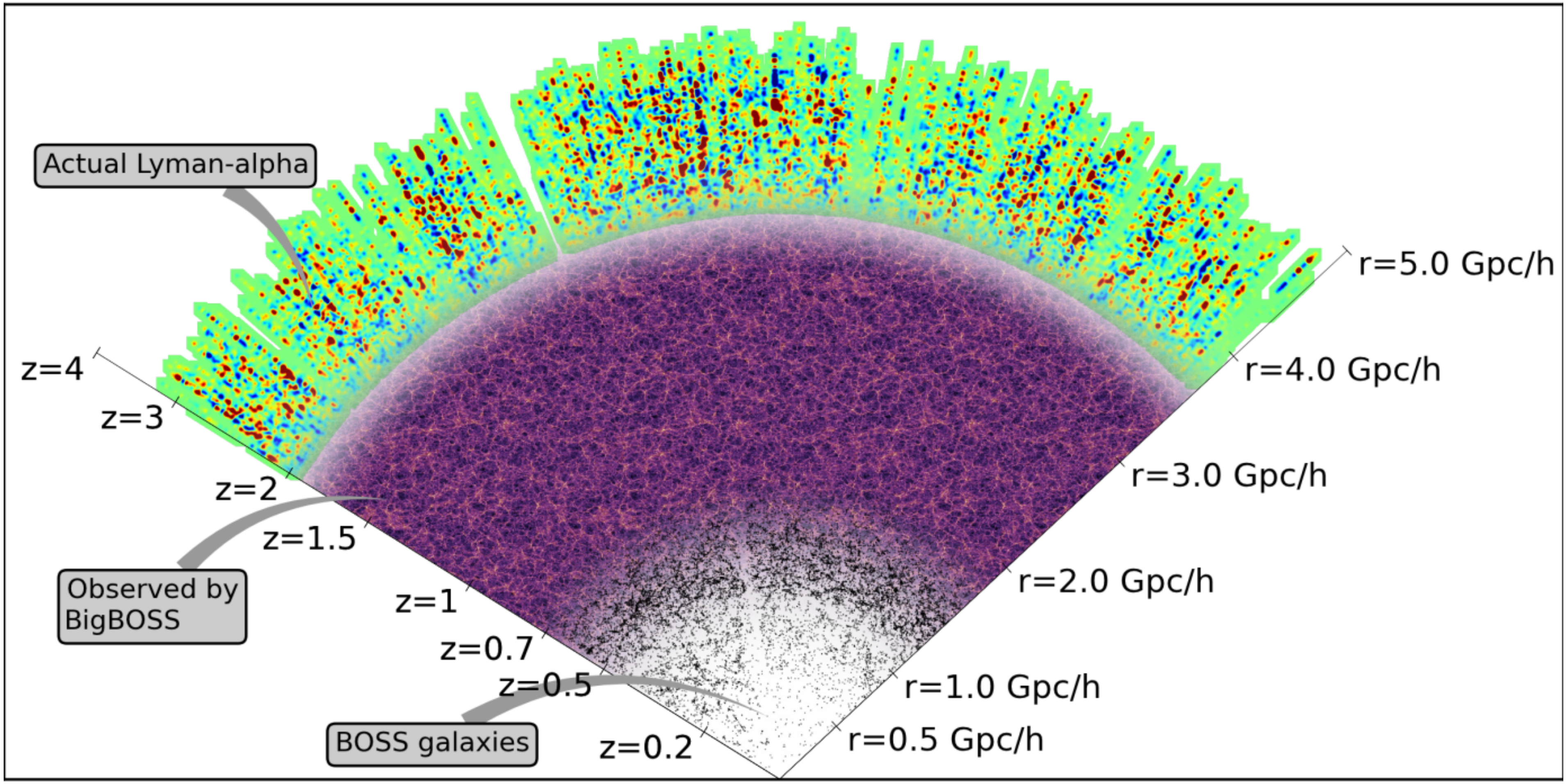
Table B2.1: **[S]** designates a spectroscopic redshift survey, **[I]** an imaging survey and **[R]** a radio survey.

Projects	Status	Ref.
[S] VIPERS	2009-2015	http://vipers.inaf.it/
[S] SDSS-III/BOSS	2009-2014	http://www.sdss3.org/surveys/boss.php
[I] DES	2012-2017	http://www.darkenergysurvey.org/
[I] VST/KIDS	2011-2016	http://kids.strw.leidenuniv.nl/
[I] eROSITA	2015-2020	http://www.mpe.mpg.de/erosita/
[S] HETDEX	2015-2017	http://hetdex.org/
[S] SDSS-IV/eBOSS	2014-2020	http://www.sdss3.org/future/eboss.php
[I+S] Euclid	2021-2027	http://sci.esa.int/euclid/
[S] DESI	2019-2022	http://desi.lbl.gov/
[I] J-PAS	2015-2020	http://j-pas.org/
[S] 4MOST	2019-2024	http://www.4most.eu/
[I] VISTA-VHS	2010-2017	http://www.vista-vhs.org/
[I] iPTF	2013-2015	http://ptf.caltech.edu/iptf/
[I] ZTF	2016-2020	-
[I] LSST	2023-onwards	http://www.lsst.org/
[R] LOFAR	2013-2018	http://www.lofar.org/
[R] Meerkat SKA-Pathfiner	2016-onwards	http://www.ska.ac.za/meerkat/
[R] SKA	2019-onwards	http://www.skatelescope.org/
[R] CMB (COrE/PRISM)	Proposal	http://www.prism-mission.org/
[R] PLANCK	2009-2014	http://sci.esa.int/planck/

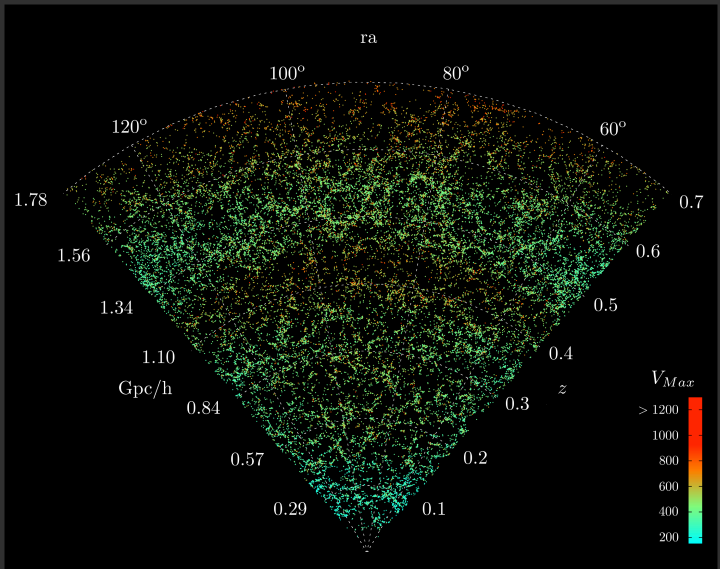


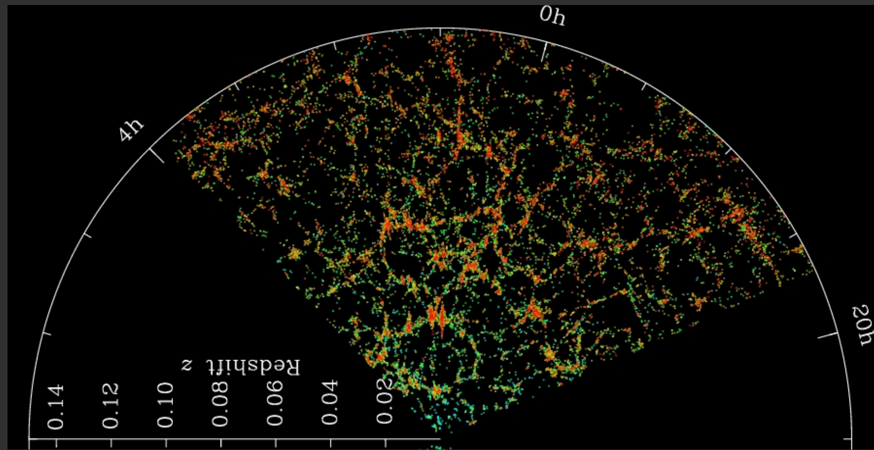
Dark Energy racing season has started!

Large Scale Structure Spectroscopic Surveys

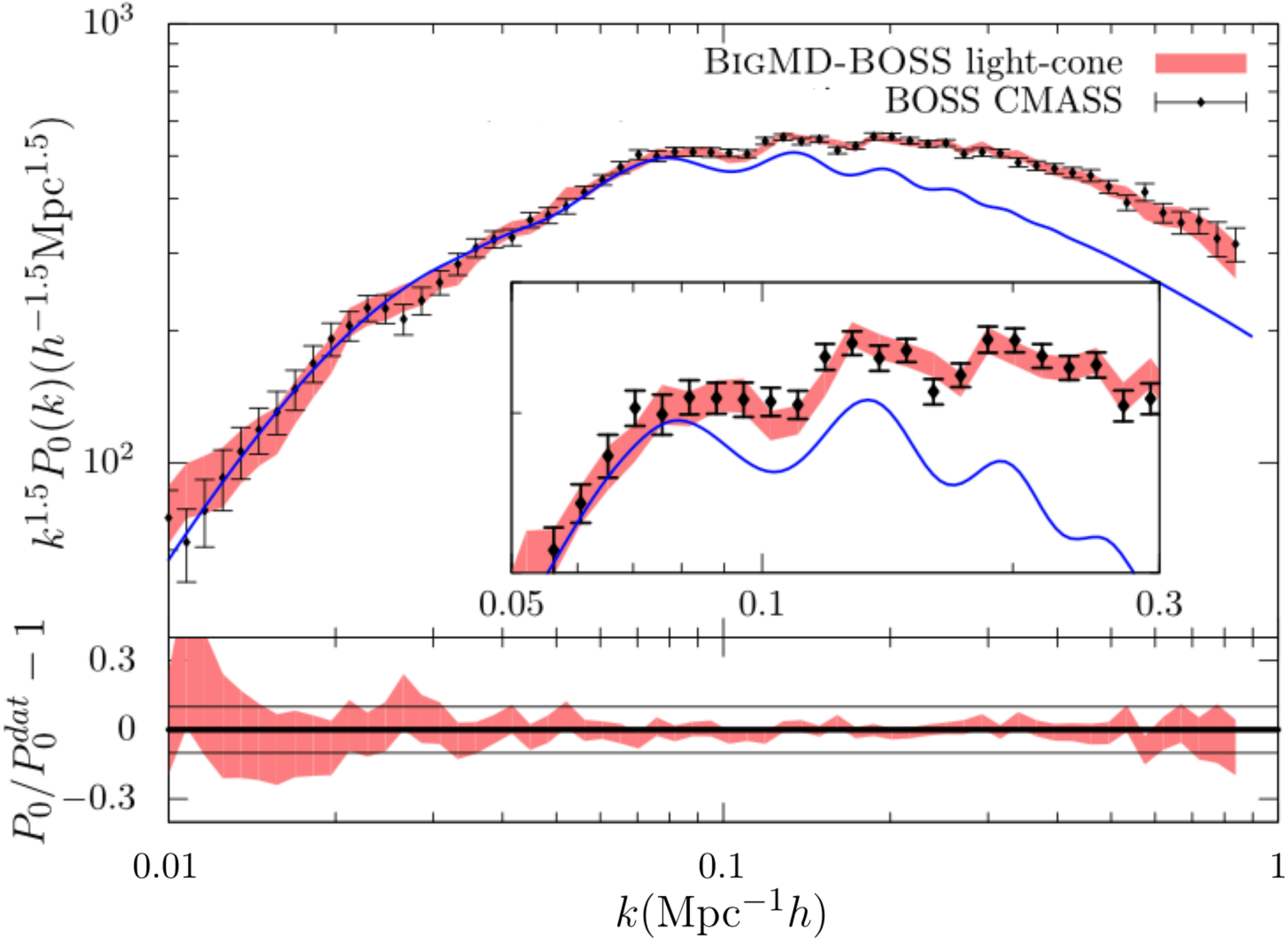


Simulations calculate consequences of dark physics





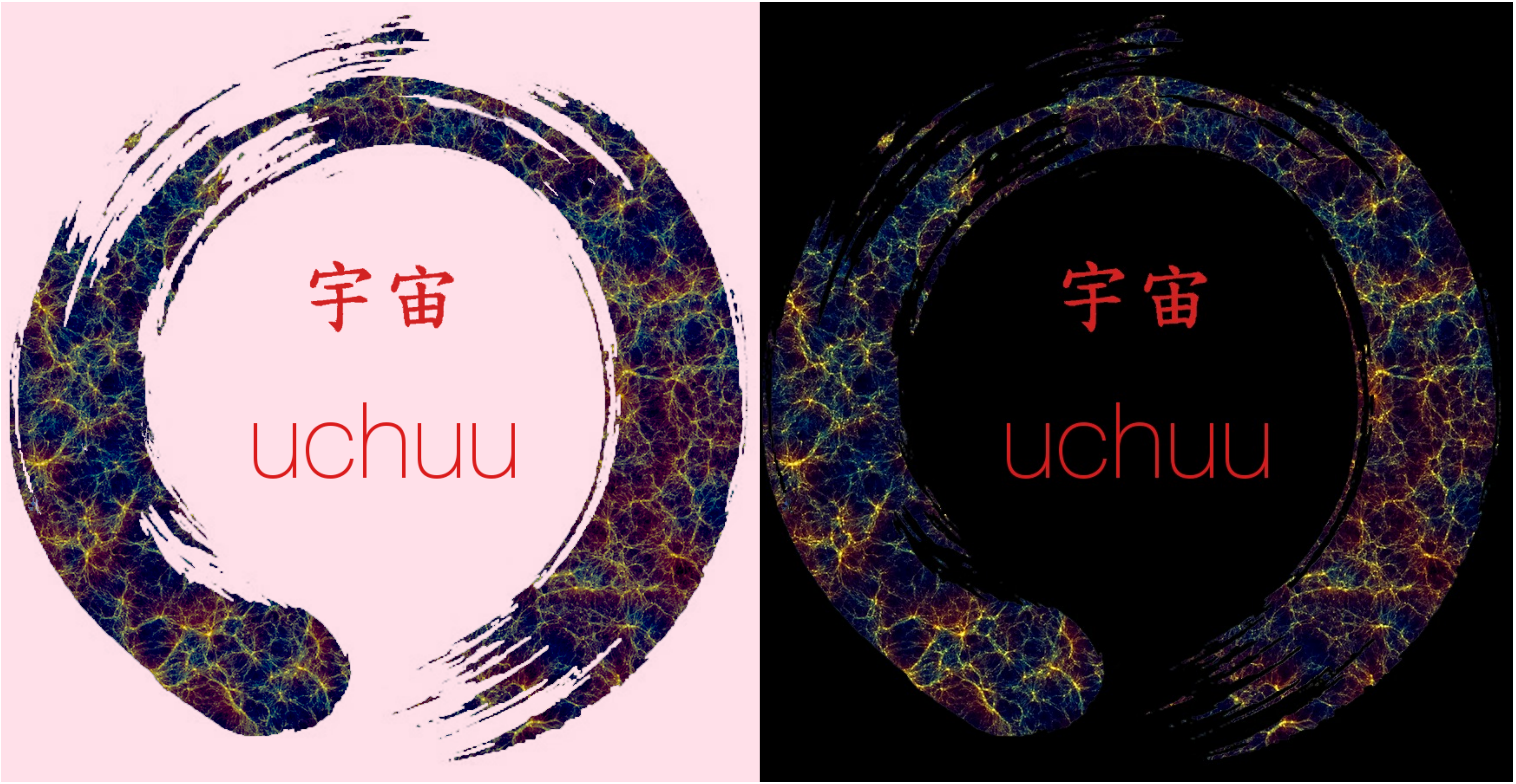
We provide cosmological simulations for large surveys
www.skiesanduniverses.org



Power spectrum from the BigMD-BOSS light cone and the BOSS CMASS sample at z=0.5

Rodríguez-Torres et al. 2016

The Uchuu Project



The **Uchuu** Collaboration

Collaboration Core Members:

Bruno Altieri, ESAC, **Spain** (*)
Sofia Cora, Institute of Astrophysics of La Plata, **Argentina** (*)
Darren Croton, Swinburne University, **Australia** (*)
Tomoaki Ishiyama, Chiba University, **Japan** (*)
Eric Jullo, LAM, **France** (*)
Anatoly Klypin, Virginia University, **USA** (*)
Ben Metcalf, Bologna University, **Italy**
David Millan, IAA-CSIC, Spain
Francisco Prada, IAA-CSIC, Spain (*)
Manodeep Sinha, Swinburne University, Australia
Taira Oogi, Kavli IPMU / University of Tokyo, Japan
Cristian Vega-Martínez, University of La Serena, **Chile**

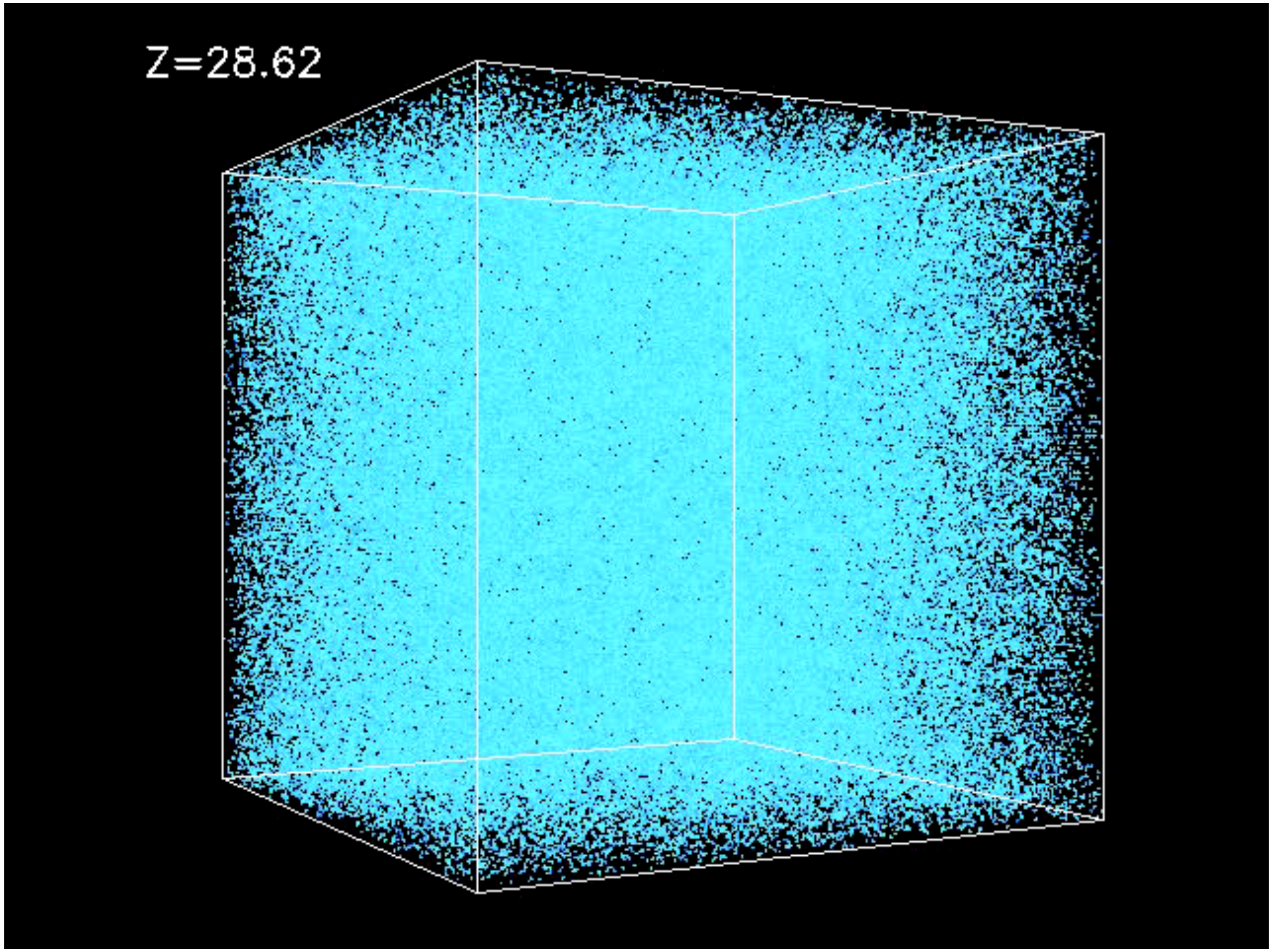
(*) members of the Collaboration Board (CB);
CB Chairs: Tomoaki Ishiyama and Francisco Prada

Our Team Expertise on Computational Cosmology

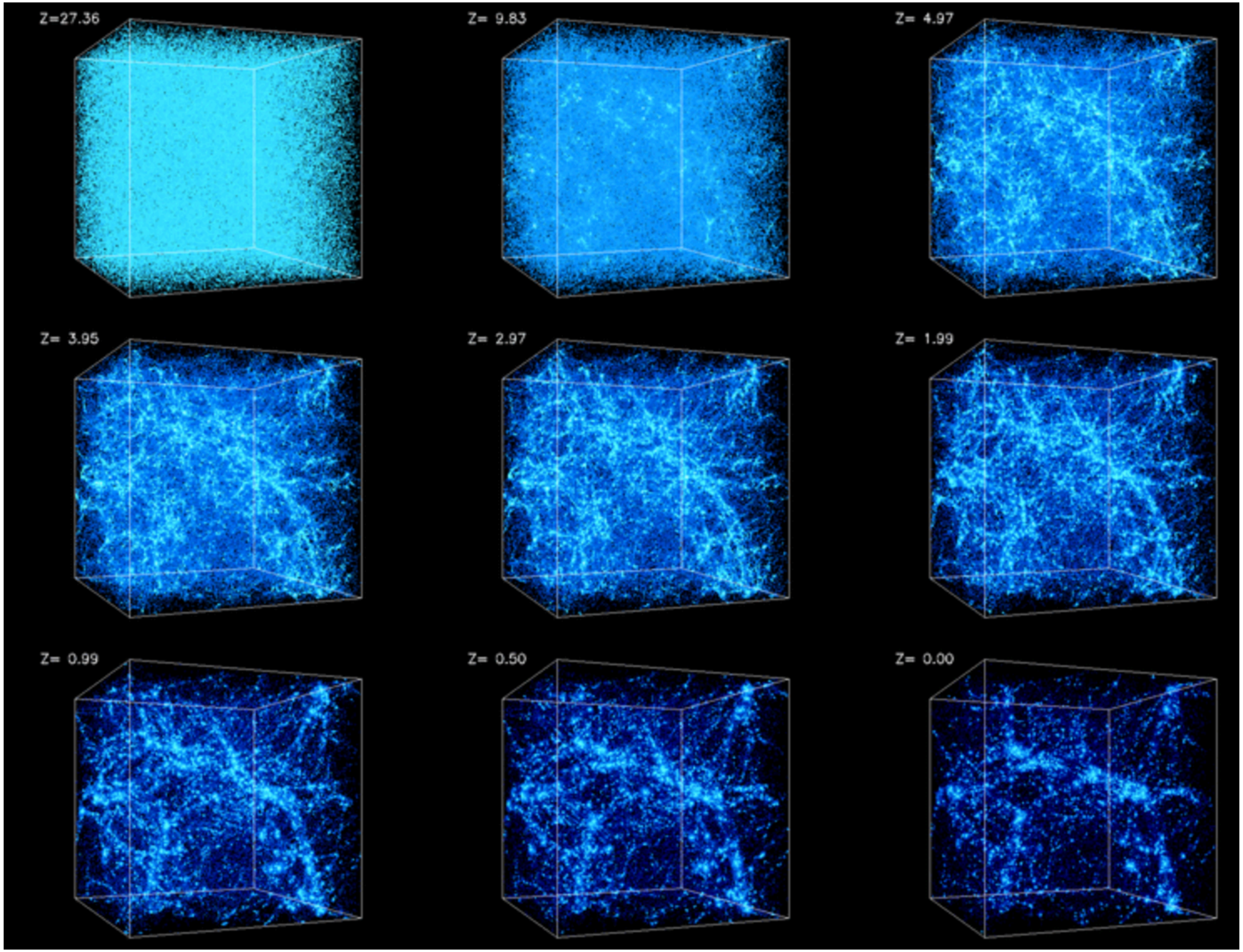
- 160 years of experience on cosmological simulations and virtual observations:
  - ART, GreeM3, GLAM N-body codes
  - MultiDark, GLAM, ν^2GC suite of simulations (data and products)
  - SAGE & SAG semi-analytical models of galaxy formation
  - Gravitational lensing maps generation

- 60 years of experience building publicly available databases and data platforms for cosmological simulation products and virtual laboratory that host mock observations for massive galaxy survey data (SDSS, DESI, Euclid)
  - MultiDark Simulations Database (*www.cosmosim.org*)
  - Skies & Universes (www.skiesanduniverses.org)
  - Theoretical Astrophysical Observatory (https://tao.asvo.org.au/tao/)

HPC Simulations of the Universe



Credit: A. Kravtsov



Credit: A. Kravtsov

Supercomputer Astronomy: The Next Generation

```
* Number of particles: 12,800^3
* Particle mass: 3.27018e+08 Msun/h
* Box size: 2000.0 Mpc/h
* Force resolution assumed: 0.00427 Mpc/h
* Total number of time steps: 6,000
```

Uchuu Parameters

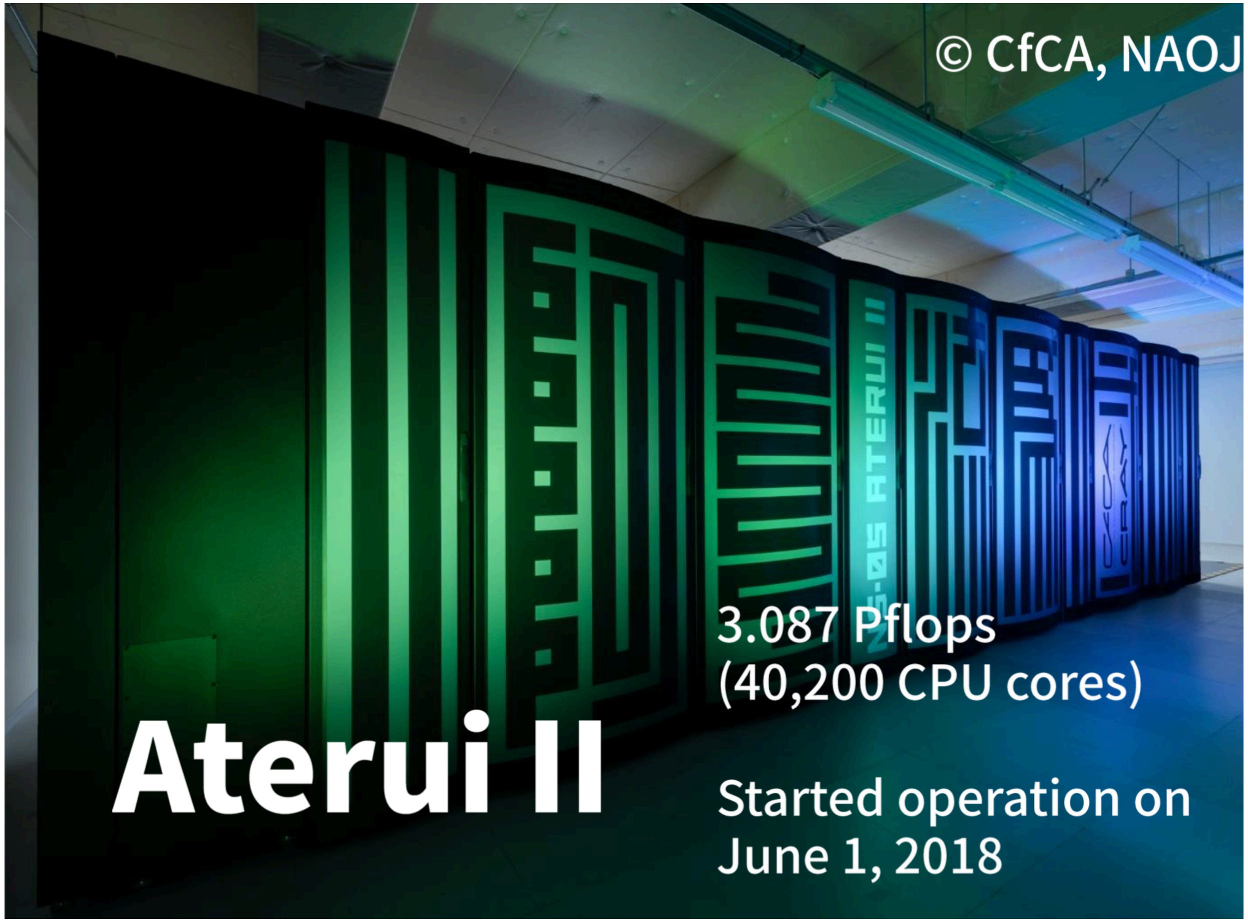
2,097,152,000,000

two trillion particles!

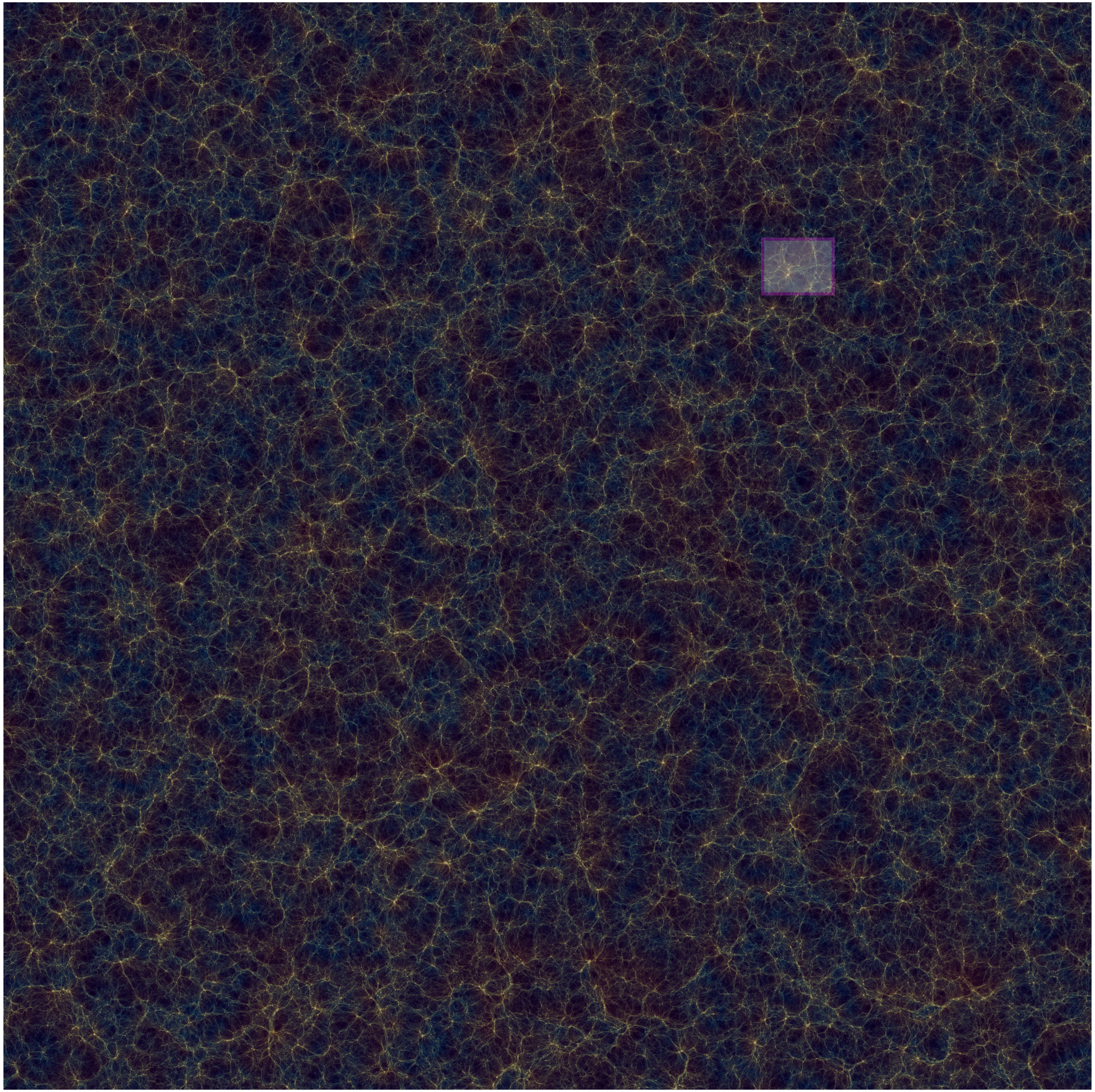
Run with GreeM3 N-body code
by Tomoaki Ishiyama

40,000 CPU hours
Exclusive used over 48 hrs / month
July 2018 - October 2019

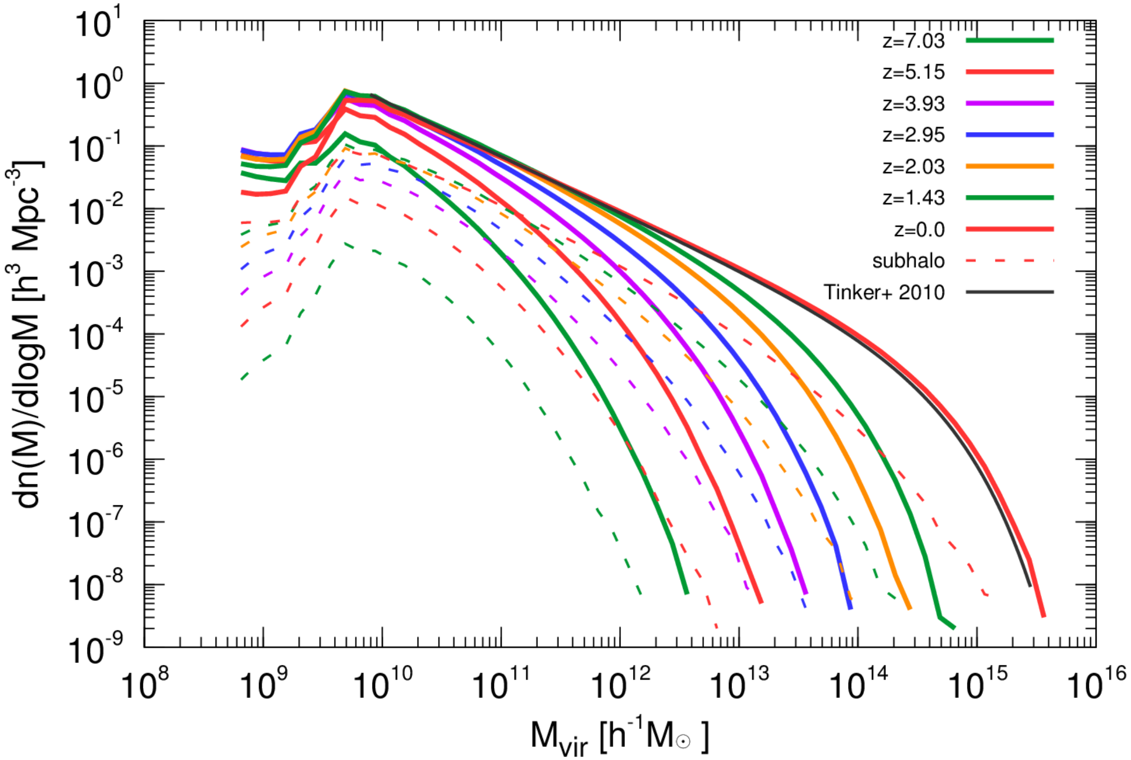
4 PB of raw data!



Uchuu is a simulation of the formation of the Large Scale Structure of the Universe based on the mutual gravitational interactions between the dark matter particles, and it can resolve smaller features to depict the Universe in greater detail.

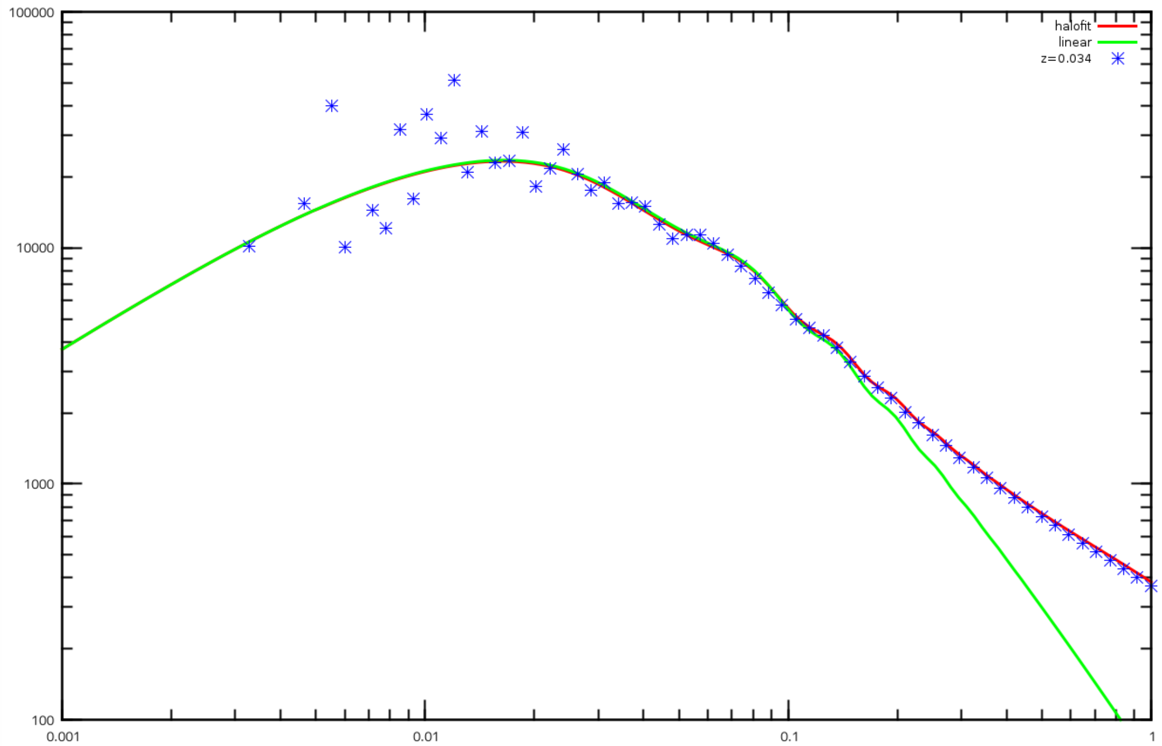


Uchuu Slice



Mass function of dark matter halos at different epochs

Power spectrum of matter density at present epoch



Uchuu Products on the Cloud

To store and allow analysis with about **300Tb of data:**

- 2 (out of 50) outputs of all particles (128Tb)
- 50 snapshots of halo catalogs (1Tb each)
- 50 snapshots of MergingTrees (1Tb each)
- 50 snapshots of semi-analytical models (1Tb each)
- Gravitational Lensing Maps and Catalogs (20Tb total)

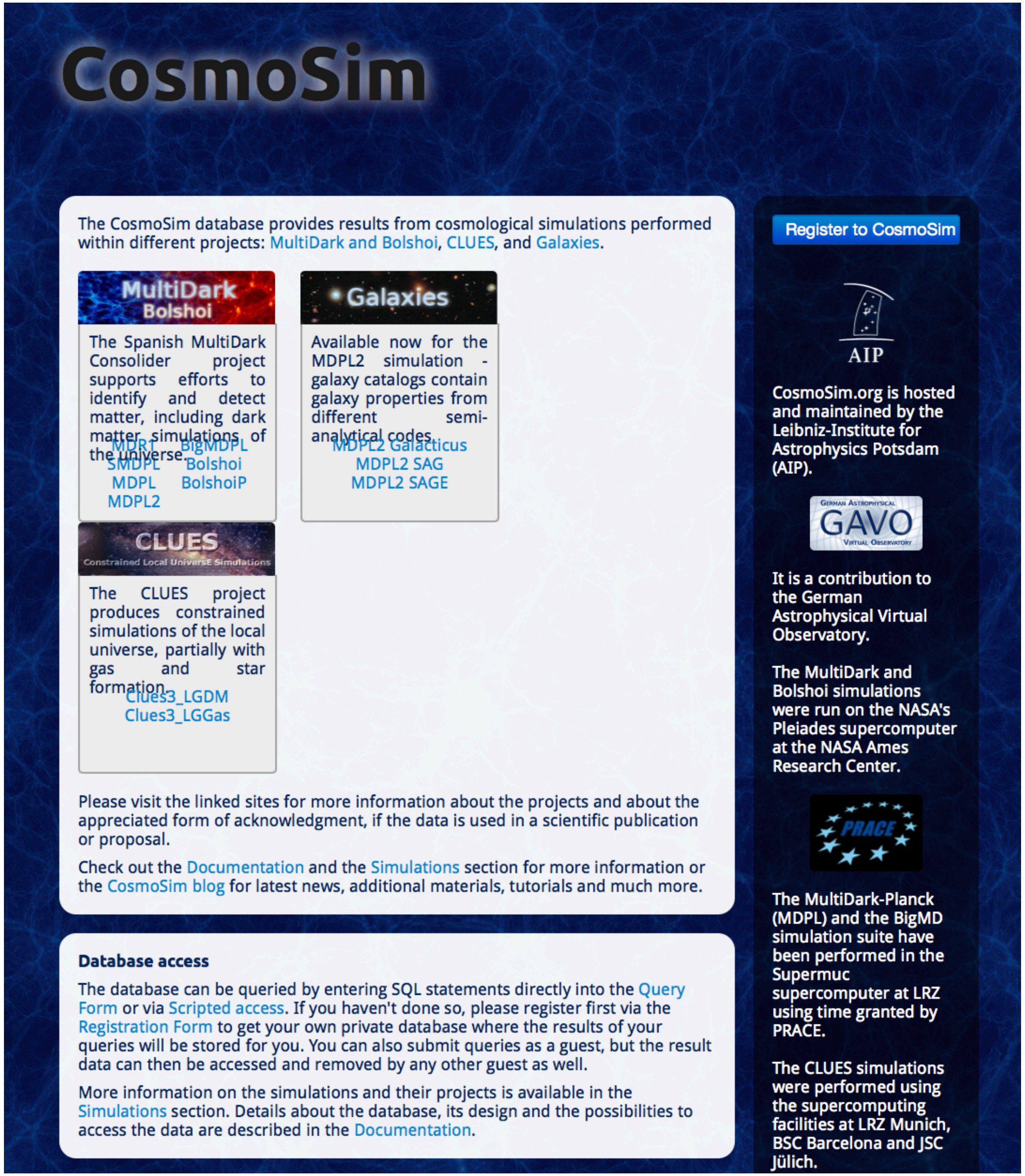
WHAT ARE WE GOING TO DO WITH ALL THIS FUTURE?

OR-
WHAT IS ALL THIS FUTURE GOING TO DO WITH US?

#cococapitanwritting

Dissemination of the data to the public: Our operational developments

MultiDark Project
&
AIP-Potsdam
(German VO)



SQL queries

www.cosmosim.org

IAA-CSIC & NMSU

Skies & Universes

Home | Simulations | Products | Surveys

The main goal of the Skies and Universes website is to provide access for the astronomical community with *results from cosmological simulations*. These results are useful on their own to understand processes related with the non-linear evolution of cosmic structures. Some other products are specifically focused on large survey observational projects. To facilitate the interaction between the theory and observations we also provide links to recent observational projects.There are other databases and websites that provide access to results of simulations. Their links can also be found here.

Specifics of our portal is that it provides raw data of simulations. For example, one can find here *coordinates and velocities* of simulations with billions of particles. We also provide *halo catalogs, light-cones for mock galaxy/qso samples*, *power spectra* for many thousands of N-body realizations and some observational products. While we provide some assistance in the form of codes to read the data, it is responsibility of users to write their own analysis software. Unlike the CosmoSim.org and Millennium databases, we do not support SQL queries. Users of *Skies & Universes* are expected to download data and write their own routines to analyze the data. At present we do not provide computing facilities to do the analysis.

Downloading files: You are welcome to use more than 30Tb of raw data on this site. However, special care should be given when downloading large files. Please, do not download all of the large files at once. You will not get anything, if you do. Get 1-2 files, wait until the downloading finishes and then get another bunch files.

More info at arXiv:1711.01453

Simulations:
* Main Page
* GLAM
* MultiDark
* Bolshoi
* Guest simulations:
  * Lomonosov
  * MassiveNuS
  * ν2GC

Products:
* Main page
* Mock Catalogs

Other databases and websites:
* CosmoSim
* Millennium
* Bolshoi
* DarkSky
* TAO
* ν2GC
* Abacus

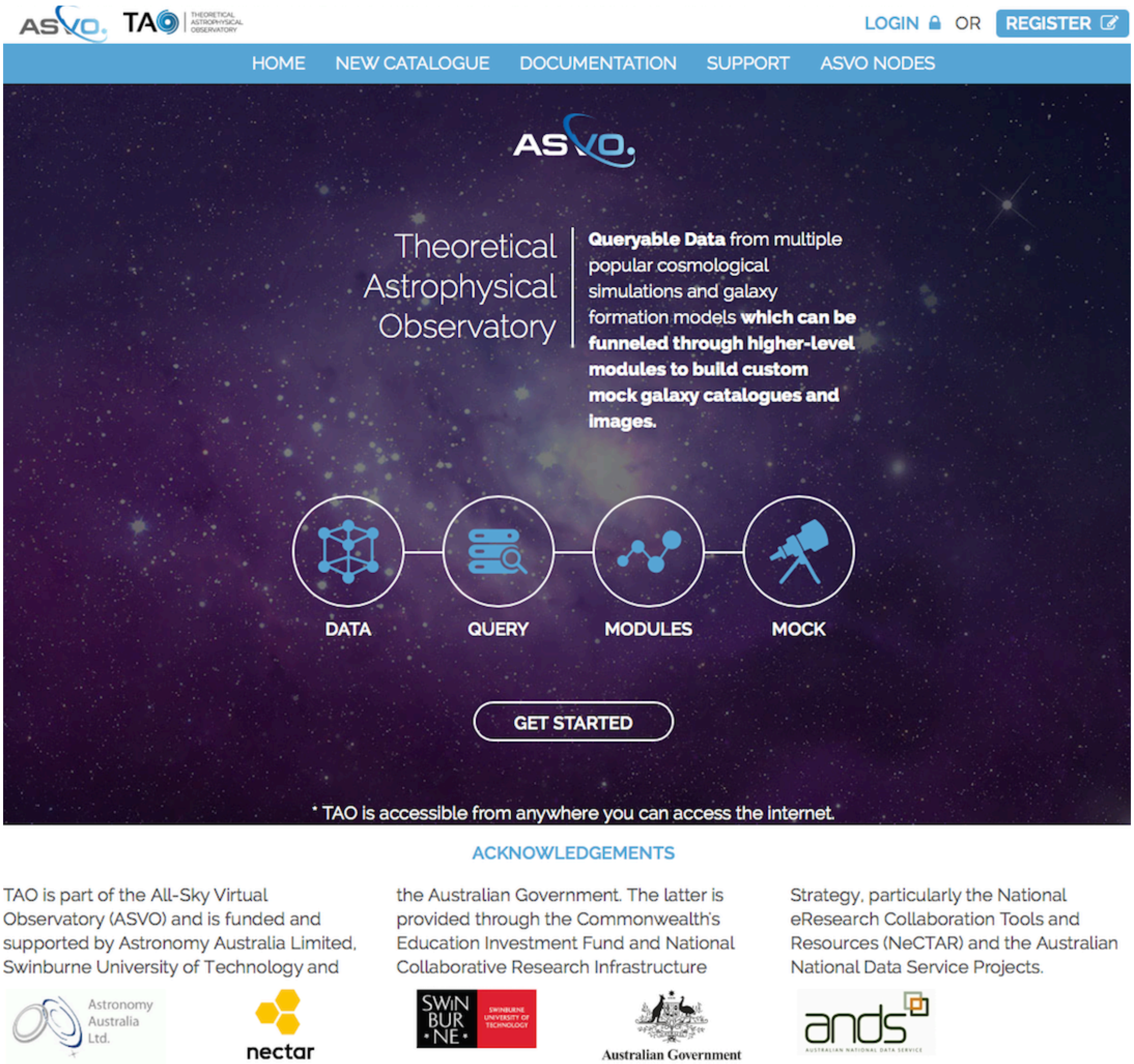
Credits and Registration:
* Credits * Registration

Alternative Skies & Universes sites:
Europe
USA

If you use our results, acknowledge www.skiesanduniverses.org Contact us

Files repository

www.skiesanduniverses.org



Swinburne

Cloud oriented

tao.asvo.org.au

Other simulation public platforms

ILLUSTRIS ABOUT PEOPLE RESULTS PRESS IMAGES/VIDEOS DATA ACCESS THE EXPLORER GALAXY OBSERVATORY

Public Data Access Overview / Web API Documentation

Welcome! You are currently not logged in. In order to access any Illustris[TNG] data, you need to first **Login**.

user@domain.com | enter password | Login

Don't have an account yet? **New User Registration**.

The web-based interface (API) can respond to a variety of user requests and queries, and can be used in addition to, or in place of, the download and local analysis of large data files. At a high level, the API allows a user to search, extract, visualize, and analyze. In each case, the goal is to reduce the data response size, either by extracting an unmodified subset, or by calculating a derivative quantity.

This page has three sections: getting started guide, cookbook, reference.

We provide examples of accessing the API in a few languages. Select one to show all the content on this page specifically for that language.

Python | IDL | Matlab | Julia

API Getting Started Guide

First, start up your interface of choice and define a helper function, whose purpose is to ma
and verify that the response is successful. If the response type is JSON, then automatically

```
>>> import requests
>>>
>>> baseUrl = 'http://www.illustris-project.org/api/'
>>> headers = {"api-key":"INSERT_API_KEY_HERE"}
>>>
>>> def get(path, params=None):
>>>     # make HTTP GET request to path
>>>     r = requests.get(path, params=params, headers=headers)
>>>
>>>     # raise exception if response code is not HTTP SUCCESS
>>>     r.raise_for_status()
>>>
>>>     if r.headers['content-type'] == 'application/json':
>>>         return r.json() # parse json responses automatical
>>>     return r
```

Issue a request to the API root.

```
>>> r = get(baseUrl)
```

The response is a dictionary object with one key, "simulations", which is a list of N (

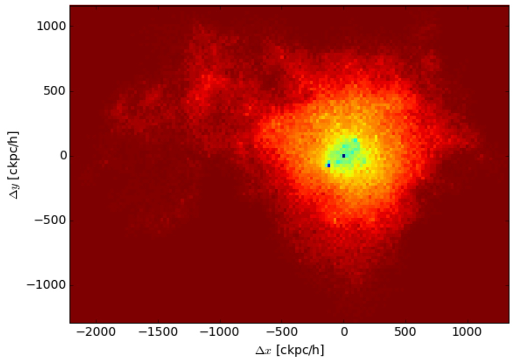
```
>>> r.keys()
['simulations']

>>> len(r['simulations'])
18
```

A web-based interface (API)

ILLUSTRIS ABOUT PEOPLE RESULTS PRESS IMAGES/VIDEOS DATA ACCESS THE EXPLORER GALAXY OBSERVATORY

Make a quick 2d histogram visualization of the distribution of gas bound to this subhalo, weighted by the log of gas cell mass, and with position relative to the subhalo center.

```
>>> with h5py.File(cutout,'r') as f:
>>>     x = f['PartType0']['Coordinates'][:,0] - sub_prog['pos_x']
>>>     y = f['PartType0']['Coordinates'][:,1] - sub_prog['pos_y']
>>>     dens = np.log10(f['PartType0']['Masses'][:])
>>>
>>> plt.hist2d(x,y,weights=dens,bins=[150,100])
>>> plt.xlabel('$\Delta x$ [ckpc/h]')
>>> plt.ylabel('$\Delta y$ [ckpc/h]');
```



That's enough! You can also explore the independent examples below.

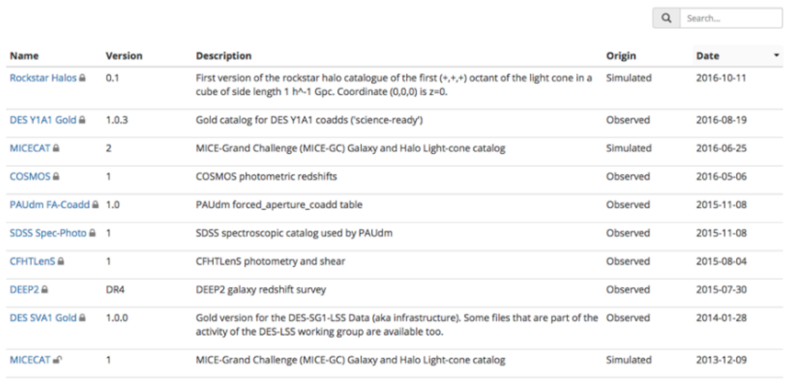
API Cookbook

Examples for how to accomplish specific tasks, covering some different API response formats.

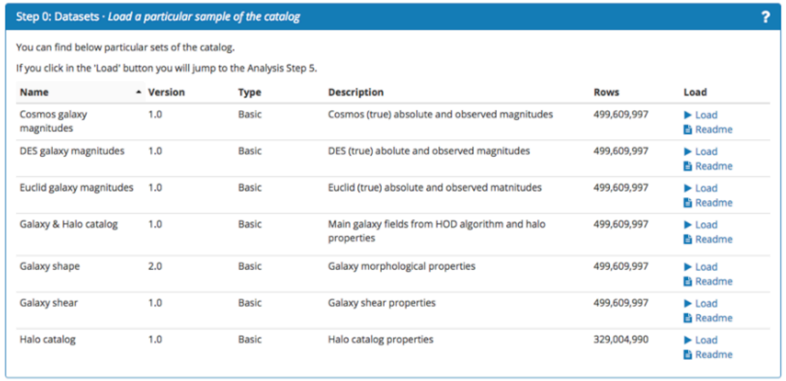
http://www.illustris-project.org

CosmoHub: See following Talk!

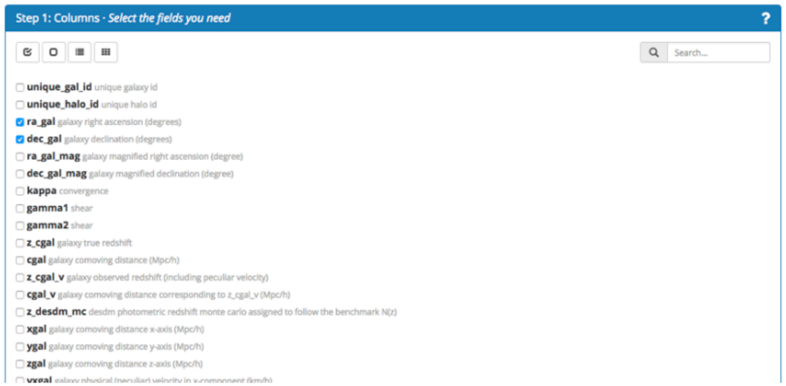
How does it work?



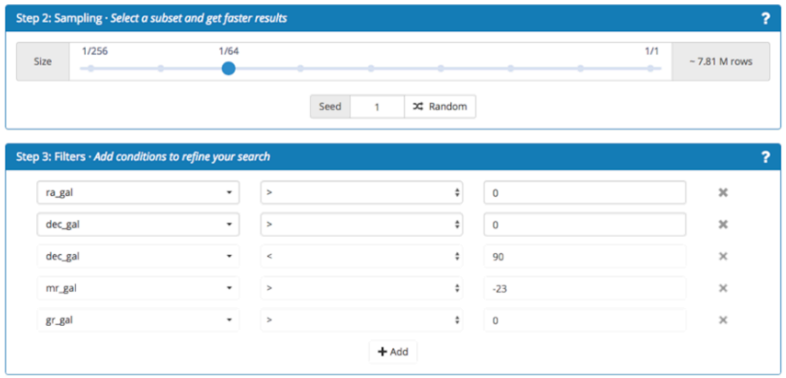
1. Select a catalog



2a. Select one of the prebuilt datasets...



2b. Or build your own custom set



3. Select the sampling and filters

CosmoHub portal uses the Apache Hive infrastructure, which provides a very fast data query and is built on top of Hadoop

Our vision for Uchuu in the Cloud @ EOSC

(NEED YOUR FEEDBACK & HELP!)

- Previous examples (TAO, CosmoHub) provide a convenient interface for “SQL” queries and allow preview the catalogs you are building with online plot generator

- Do we want the users to be able to carry out a certain set of pre-defined actions in the Cloud or do we want them to able to run arbitrary analysis snippets?

    - Uchuu @ EOSC requires the latter
    - Technologies? Spark, **Vaex** …
    - Cloud platform with necessary computing and storage resources (RAM is key 2TB, 128 cores, 500TB storage)
    - Legacy: Long live preservation (critical)

vaex.io

HOME VAEX.IO SERVICES TECHNOLOGY TEAM CONTACT US

About vaex.io

Power up your business with our data driven solutions. With our unique, state-of-the-art technology, we provide fast and scalable solutions that will make you more agile, while limiting unnecessary resources.

BE AGILE

Cut developement time by 80%. Your prototype is your solution. Create automatic pipelines for any model.

MAKE IMPACT

Improve your business right away. Deploy your models to Amazon Web Services or Google Cloud Platform with a single command.

BE RESOURCEFUL

Empower your data scientists. Turn any laptop into a big data powerhouse. No clusters, no engineers.

What is Vaex?

Vaex is a python library for lazy **Out-of-Core DataFrames** (similar to Pandas), to visualize and explore big tabular datasets. It can calculate *statistics* such as mean, sum, count, standard deviation etc, on an *N-dimensional grid* up to **a billion** (10^9) objects/rows **per second**. Visualization is done using **histograms**, **density plots** and **3d volume rendering**, allowing interactive exploration of big data. Vaex uses memory mapping, a zero memory copy policy, and lazy computations for best performance (no memory wasted).

https://vaex.io

http://vaex.astro.rug.nl/

https://vaex.readthedocs.io/

Why vaex

- **Performance:** works with huge tabular data, processes $> 10^9$ rows/second
- **Lazy / Virtual columns:** compute on the fly, without wasting ram
- **Memory efficient** no memory copies when doing filtering/selections/subsets.
- **Visualization:** directly supported, a one-liner is often enough.
- **User friendly API:** you will only need to deal with the DataFrame object, and tab completion + docstring will help you out: `ds.mean<tab>`, feels very similar to Pandas.
- **Lean:** separated into multiple packages
  - `vaex-core`: DataFrame and core algorithms, takes numpy arrays as input columns.
  - `vaex-hdf5`: Provides memory mapped numpy arrays to a DataFrame.
  - `vaex-arrow`: Arrow support for cross language data sharing.
  - `vaex-viz`: Visualization based on matplotlib.
  - `vaex-jupyter`: Interactive visualization based on Jupyter widgets / ipywidgets, bqplot, ipyvolume and ipyleaflet.
  - `vaex-astro`: Astronomy related transformations and FITS file support.
  - `vaex-server`: Provides a server to access a DataFrame remotely.
  - `vaex-distributed`: (Proof of concept) combined multiple servers / cluster into a single DataFrame for distributed computations.
  - `vaex-qt`: Program written using Qt GUI.
  - `vaex`: Meta package that installs all of the above.
  - `vaex-ml`: Machine learning
- **Jupyter integration**: vaex-jupyter will give you interactive visualization and selection in the Jupyter notebook and Jupyter lab.

宇宙
uchuu

Remarks

- **Uchuu** is coming public on Spring 2020
- Need desperately your HELP!

Grazas!
Obrigado!