The CAMELS project

Francisco Villaescusa-Navarro

SIMONS FOUNDATION

FLATIRON INSTITUTE

PRINCETON UNIVERSITY

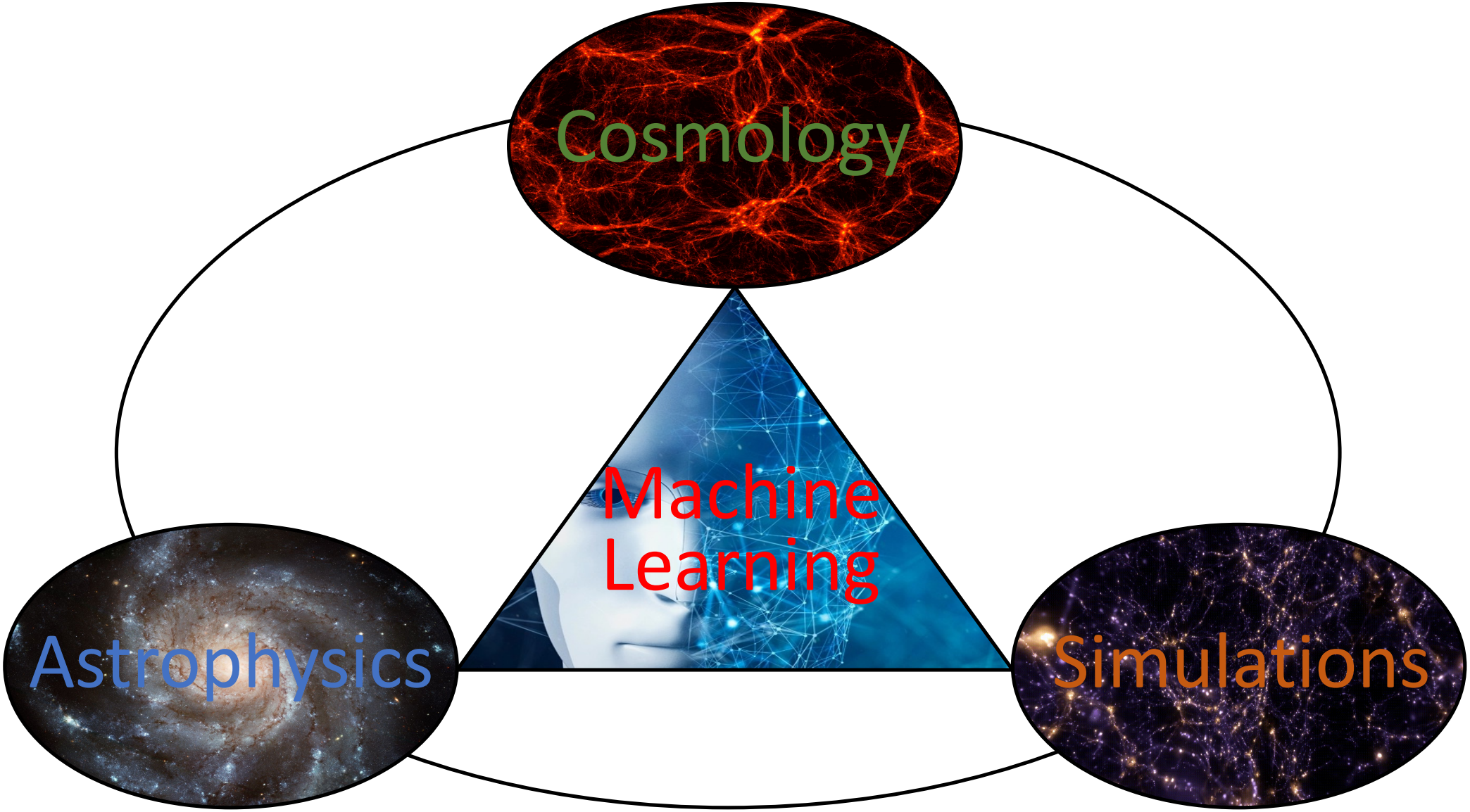
November 3rd 2023

The 2nd Shanghai Assembly on Cosmology and Structure Formation

Simulations

(https://quijote-simulations.readthedocs.io)

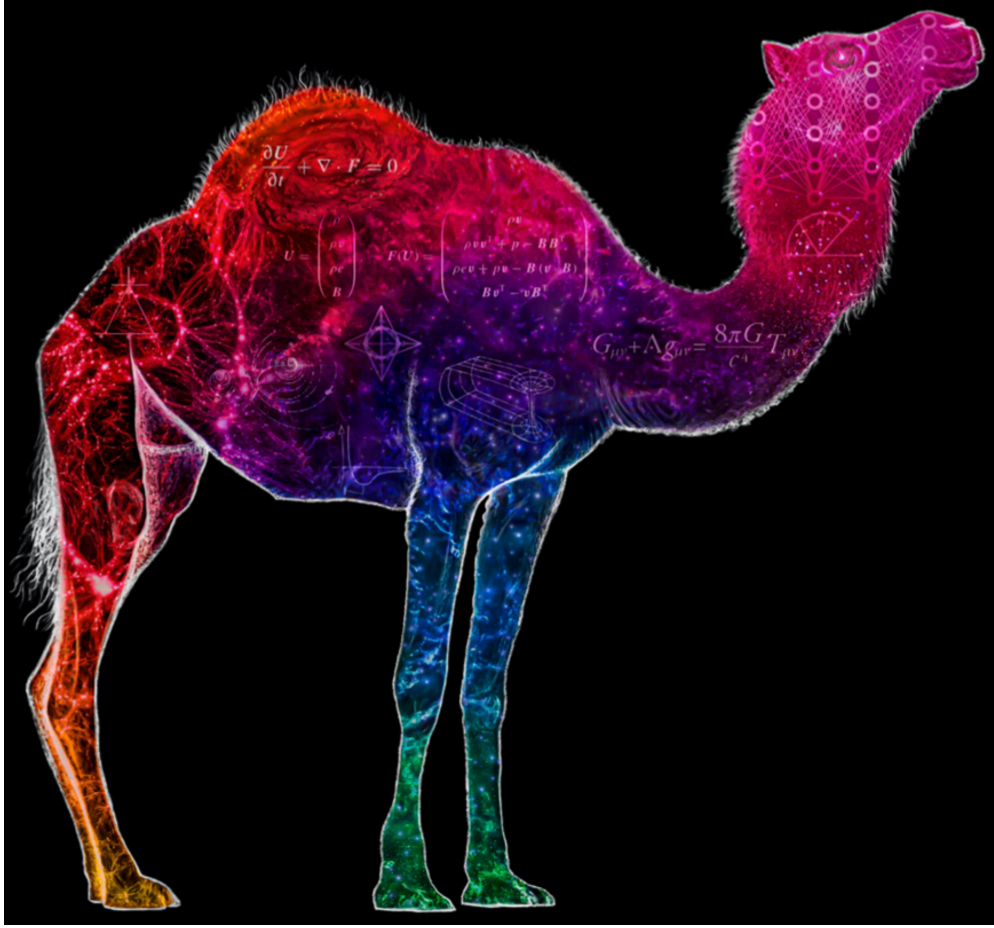
- A set of 44,100+**32768** full N-body simulations
- Around 40,000 cosmologies in
  $\{\Omega_m, \Omega_b, h, n_s, \sigma_8, M_\nu, w_0, \delta_b, f_{\rm NL}, g_{\rm NL}, f(R)\}$
- 12+ trillion particles over a volume larger than entire observable Universe
- Catalogues with billions of halos, voids (Gigantes), galaxies (Molino & Sancho). WL maps (Ulagam)
- 50+ Million CPU hours; 1+ Petabyte of data
- 140+ papers written using this data
- All data publicly available (binder & globus)


Cosmology
Machine Learning
Astrophysics
Simulations

CAMELS

Cosmology and Astrophysics with MachinE Learning Simulations

https://www.camel-simulations.org



<u>A set of 10,000+ simulations:</u>

- 5,000+ N-body sims
- 5,000+ state-of-the art hydrodynamic sims

<u>Hydrodynamic simulations run with 7 different codes/subgrid models:</u>

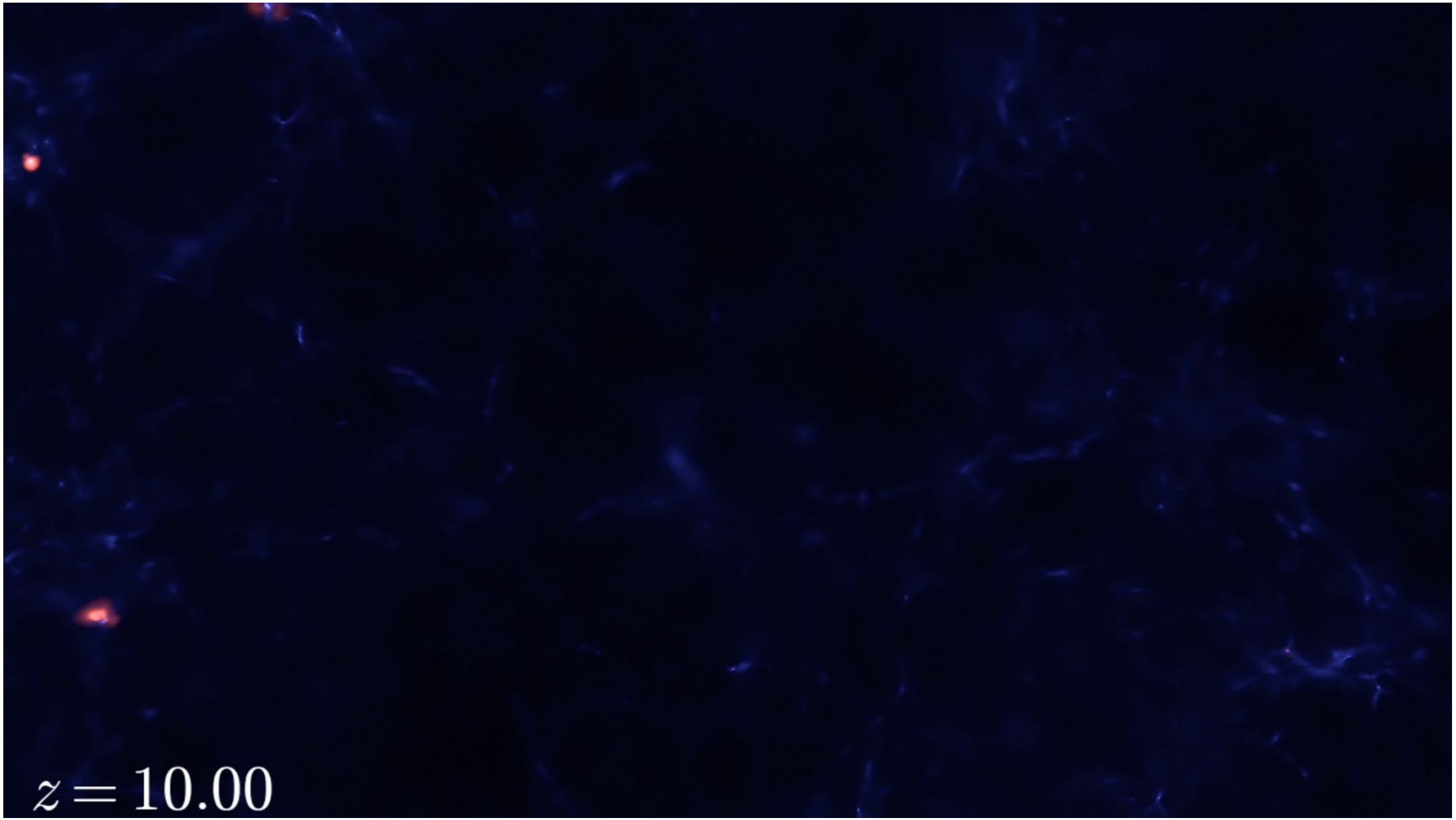
1) IllustrisTNG, 2) SIMBA, 3) Astrid, 4) Magneticum, 5) SWIFT-Eagle, 6) Ramses, 7) Enzo

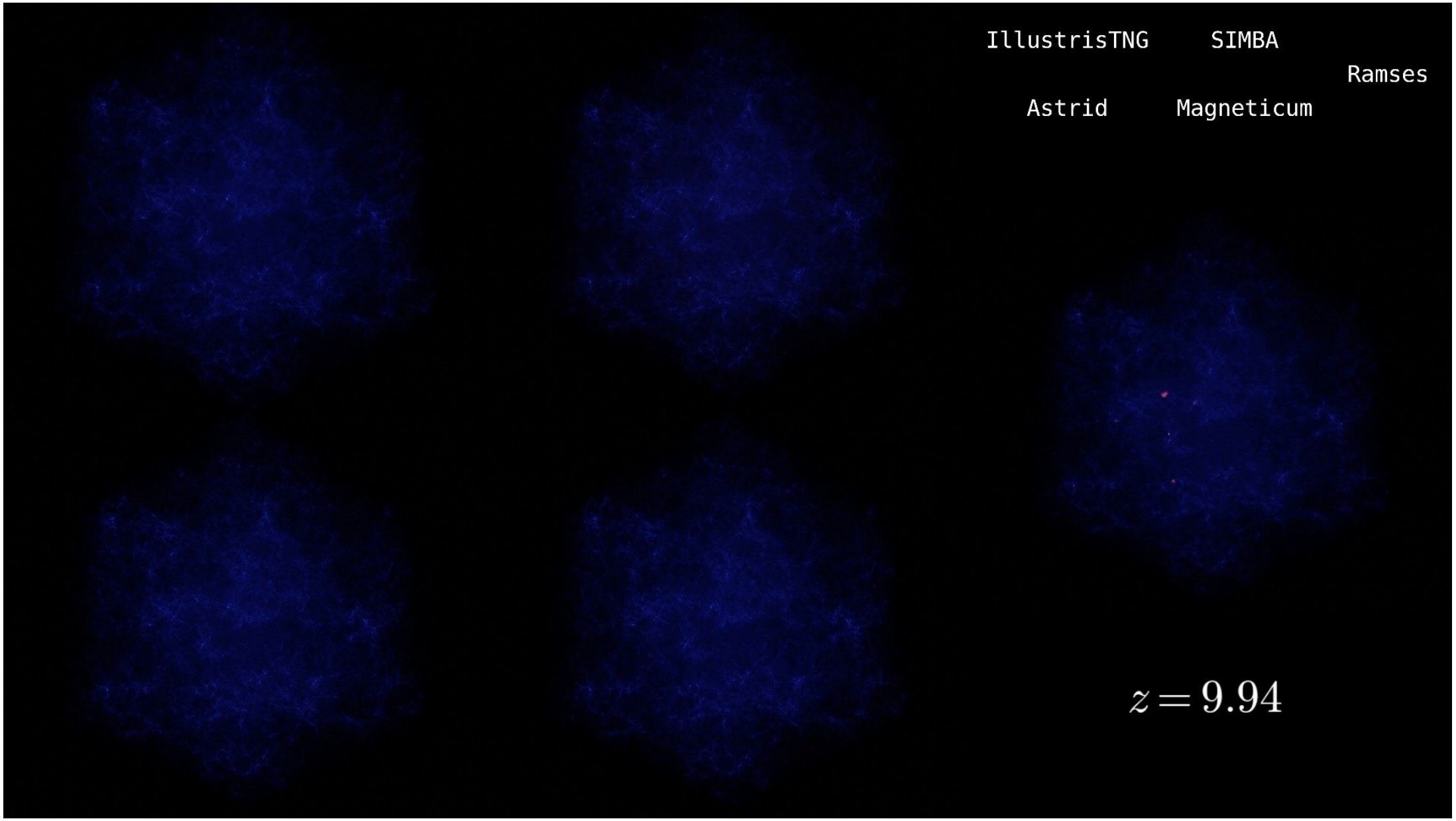
<u>Variations in cosmological & astrophysical parameters</u>

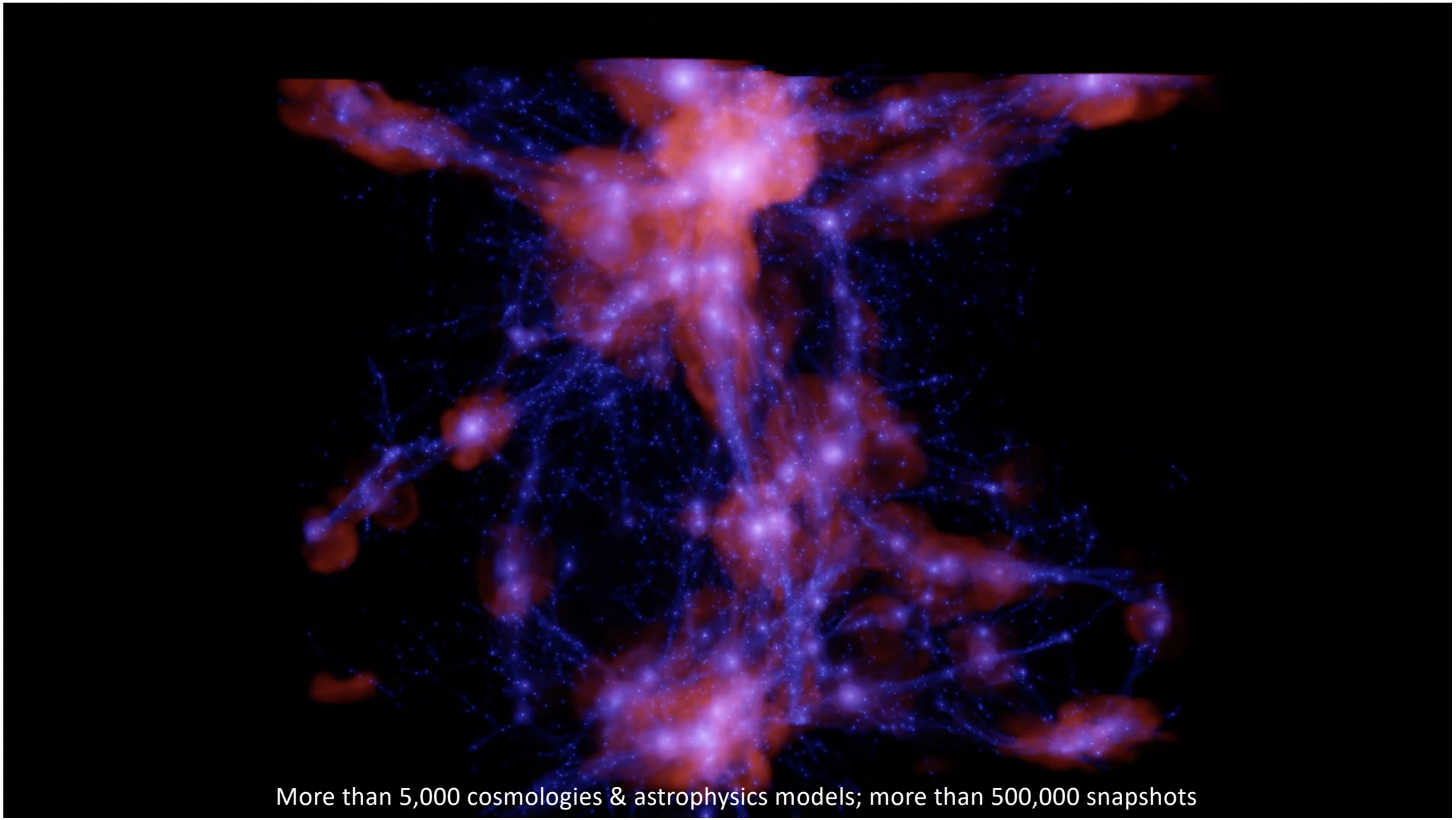
2 & 4 (LH set) 5 & 23 (SB set)

<u>Hundreds of thousands of snapshots</u>

<u>Designed for machine learning applications</u>


$z = 10.00$


IllustrisTNG
SIMBA
Ramses
Astrid
Magneticum
$z = 9.94$

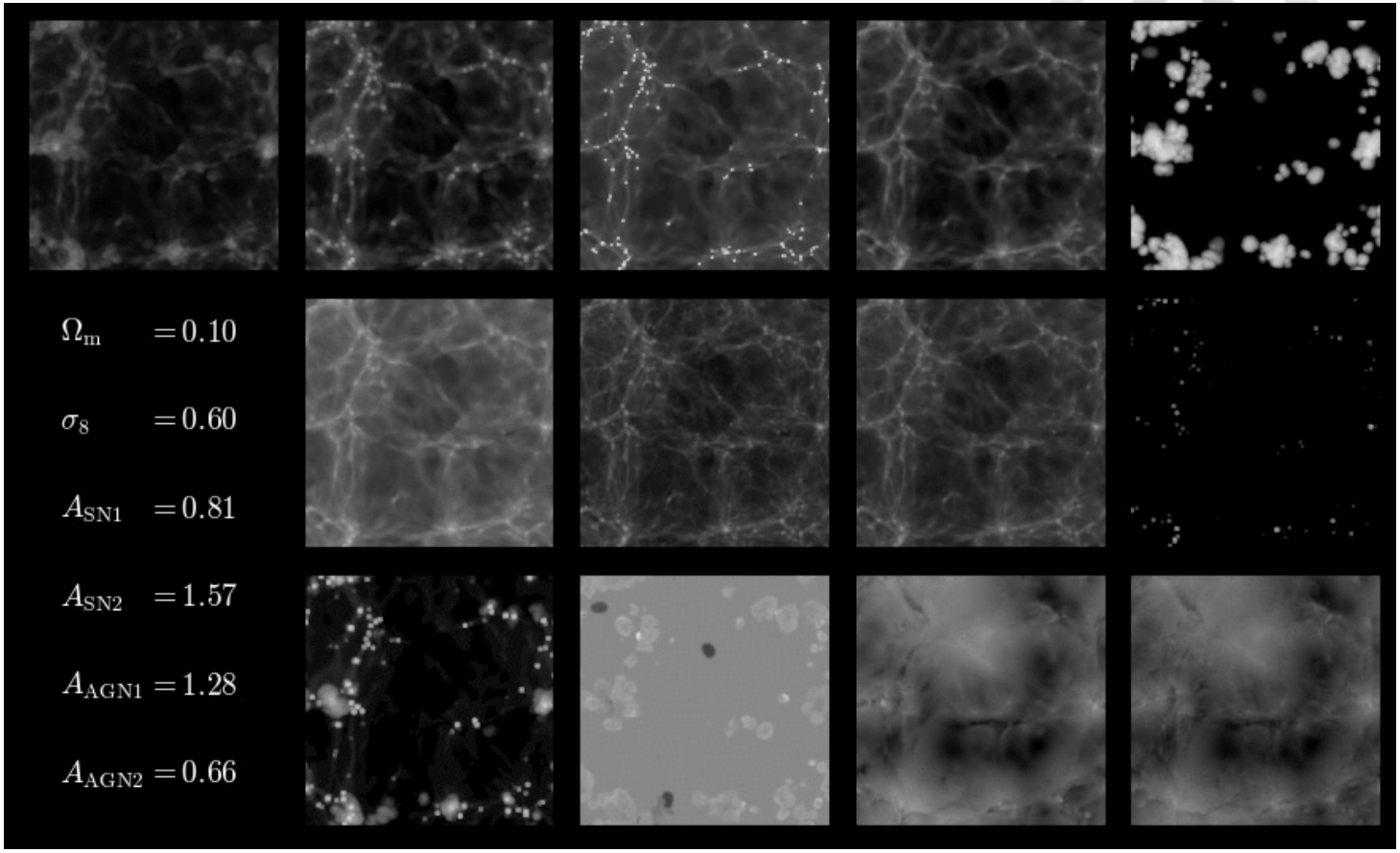

More than 5,000 cosmologies & astrophysics models; more than 500,000 snapshots

CAMELS Multifield Dataset

FVN et al. 2021c

https://camels-multifield-dataset.readthedocs.io

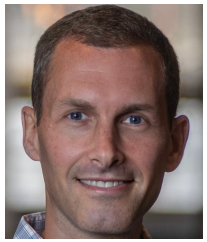
- Hundreds of thousands of labeled 2D maps and 3D grids
- Several redshifts: 0, 0.5, 1, 1.5, 2
- Three different resolutions
- 13 different fields:
  1. Gas density
  2. Gas temperature
  3. Gas metallicity
  4. Gas pressure
  5. Neutral hydrogen density
  6. Electron number density
  7. Dark matter density
  8. Total matter density
  9. Stellar mass density
  10. Gas velocity
  11. Dark matter velocity
  12. Magnetic fields
  13. Mg/Fe
- 70 Tb of data; Publicly available
- The MNIST of cosmology

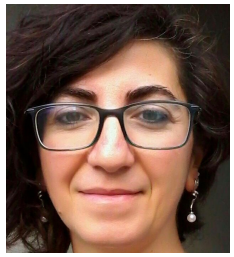


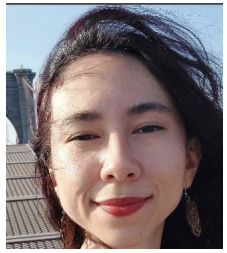
CAMELS-SAM


Lucia Perez
(Princeton/Flatiron)


Rachel Somerville
(Flatiron)


Shy Genel
(Flatiron)


Gabriella de Lucia
(Trieste Observatory)


Natali de Santi
(Flatiron/Sao Paolo)


Pablo Araya
(Sao Paolo)


Fabio Fontanot
(Trieste Observatory)


Robert Yates
(Hertfordshire)

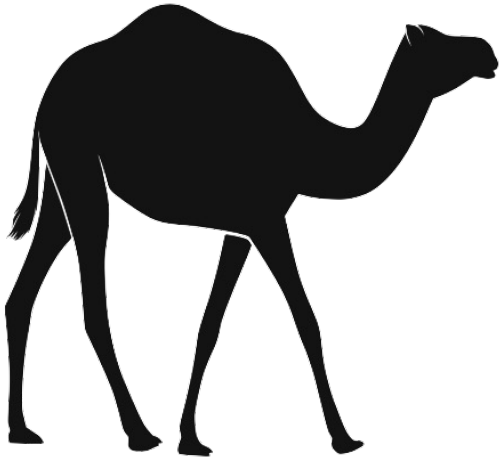
- Thousands of galaxy catalogs created from semi-analytic models:
  - 100 Mpc/h boxes
  - 25 Mpc/h boxes
- Different models:
  - SC-SAM
  - GAEA
  - L-Galaxies
  - Galform
  - …
- Varying cosmological and astrophysical parameters

Warning!!
Highly incomplete list

CAMELS

Francisco Villaescusa-Navarro, Daniel Angles-Alcazar, Shy Genel

Adrian Bayer
Alex Barreira
Ana Maria Delgado
Andrina Nicola
Alice Pisani
Benjamin Oppenheimer
Benjamin Wandelt
Blakesley Burkhart
ChangHoon Hahn
Colin Hill
Core Francisco Park
Daisuke Nagai
Desika Narayanan
David Spergel
Emily Moser
Erwin T. Lau
Faizan Mohammad
Gabrielle Parimbelli
Greg Bryan
Gabriella Contardo
Helen Shao
Jay Wadekar
Jingjing Shi
Joyce Caliendo
Lucia Perez
Lars Hernquist
Leander Thiele
Luis F. Machado Poletti
Matteo Viel
Matthew Gebhardt
Megan Tillman
Michael Eickenberg
Neerav Kaushal
Nicholas Battaglia
Oliver Philcox
Pablo Villanueva-Domingo
Rachel Somerville
Romeel Dave
Stephanie Tonnensen
Sultan Hassan
Romain Teyssier
Ulrich Steinwandel
Valentina La Torre
Vid Irsic
William Coulton
Yin Li
Yongseok Jo
Yueying Ni

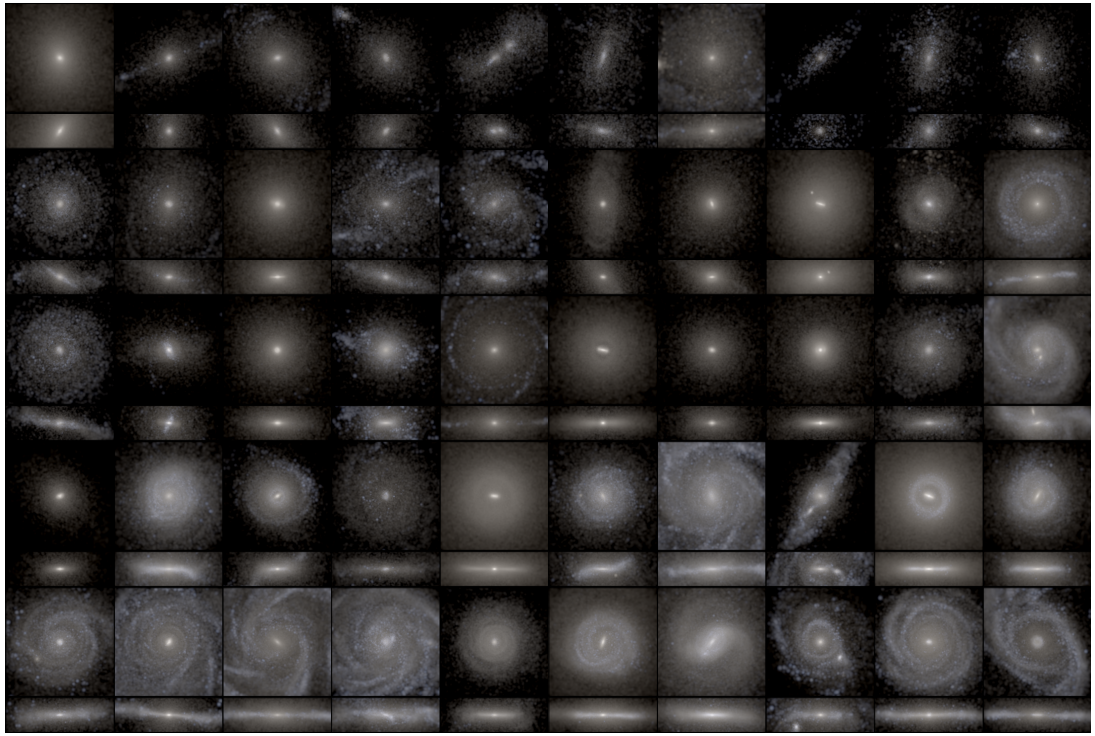


Dark CAMELS

Jonah Rose
(Florida/CCA)

Thousands of hydrodynamic simulations varying

- Cosmology
- Astrophysics
- Dark matter properties

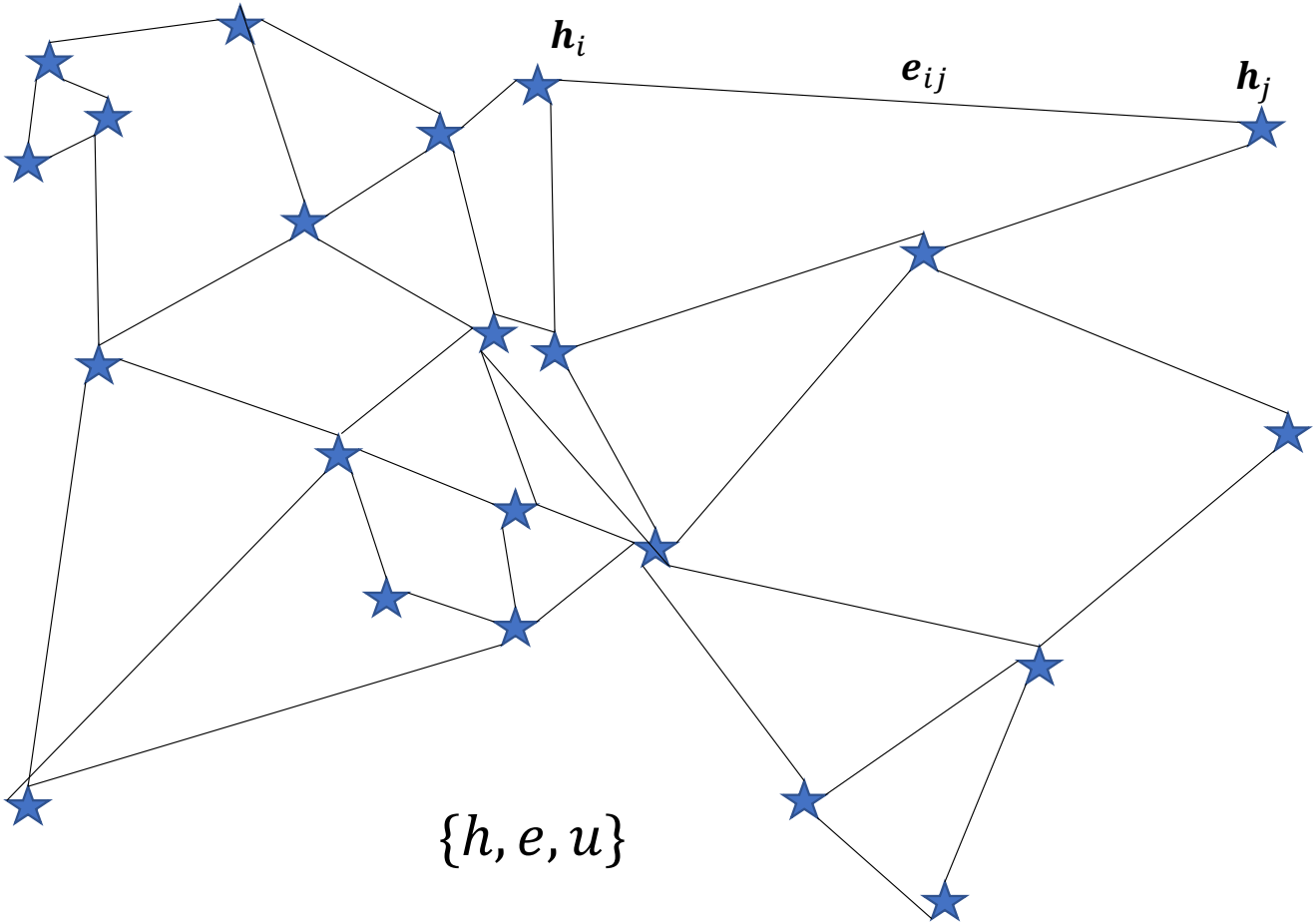


CDM WDM ETHOS SIDM ADM …

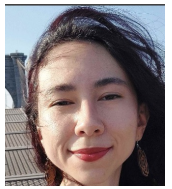
IllustrisTNG SIMBA Astrid Ramses …

25 Mpc/h boxes Milky-way zoom-ins Dwarf zoom-ins

Graphs and Graph neural networks



Pablo Villanueva-Domingo
(Barcelona)



Natali de Santi
(Flatiron/Sao Paolo)

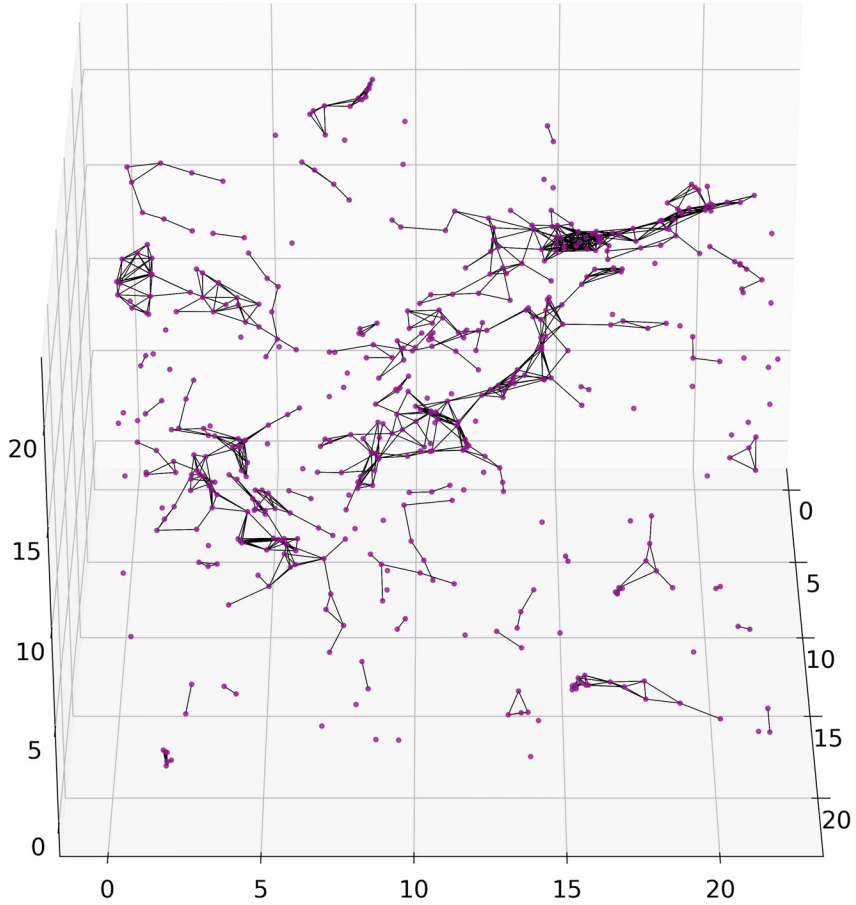
Helen Shao
(Princeton)

$$\mathbf{e}_{ij}^{(l+1)} = \phi_{l+1}([\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, \mathbf{e}_{ij}^{(l)}])$$

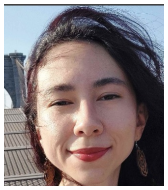
$$\mathbf{h}_i^{(l+1)} = \psi_{l+1}([\mathbf{h}_i^{(l)}, \bigoplus_{j \in \mathcal{N}_i} \mathbf{e}_{ij}^{(l+1)}, \mathbf{u}])$$

$$\mathbf{y} = \xi([\bigoplus_{i \in \mathcal{G}} \mathbf{h}_i^{(L)}, \mathbf{u}])$$

CAMELS graphs



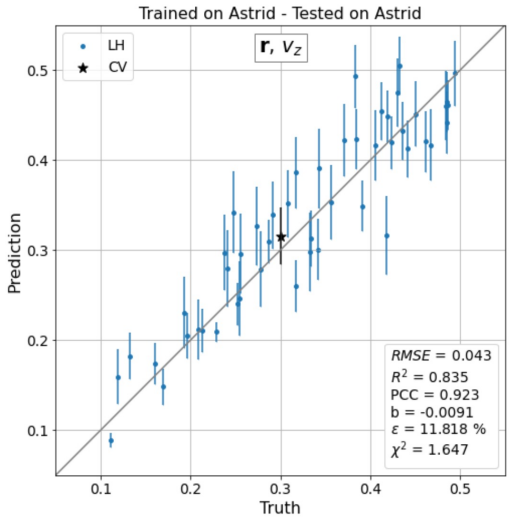
Results from 2302.14101

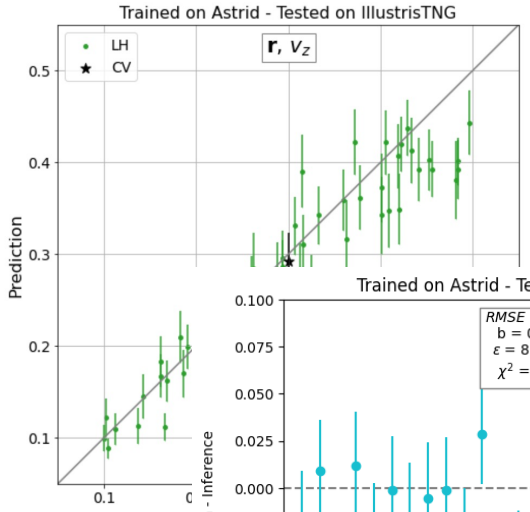


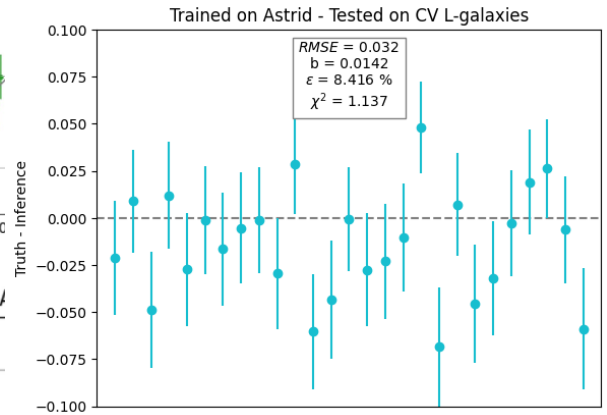
Natali de Santi
(Flatiron/Sao Paolo)

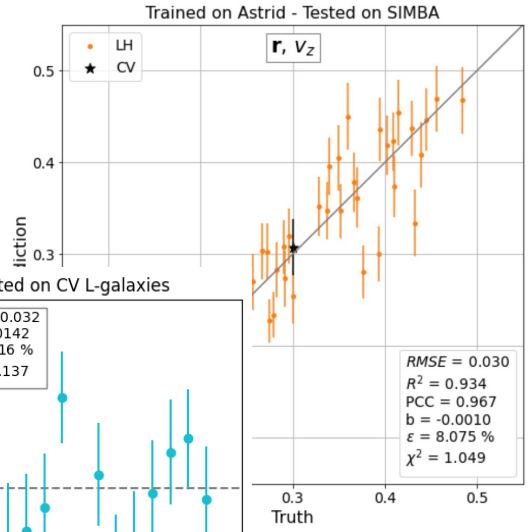
- 1,000 graphs
  - 1 graph per simulation (Astrid)
  - 1000 galaxies/graph
  - 25 Mpc/h boxes at z=0
  - Each graph is characterized by a value of Ω_m and σ_8
  - Catalogs only have galaxy positions and z-velocities
- 2 galaxies are linked if their distance is smaller than r_{link}
- Field-level and no cut on scale: $k_{\max} \sim 100\ \mathrm{h/Mpc}$
- By construction, rotational and translational invariant

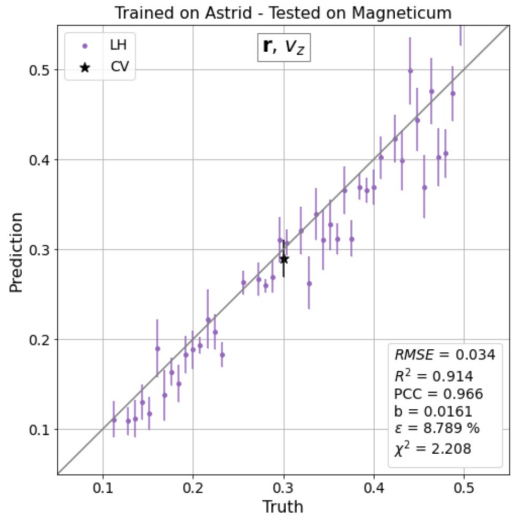
Robust field-level inference with GNNs

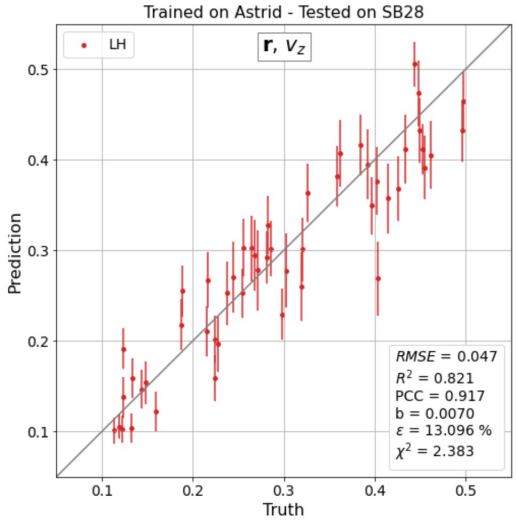


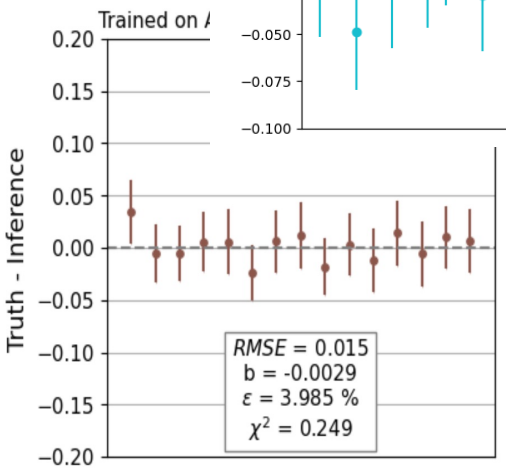


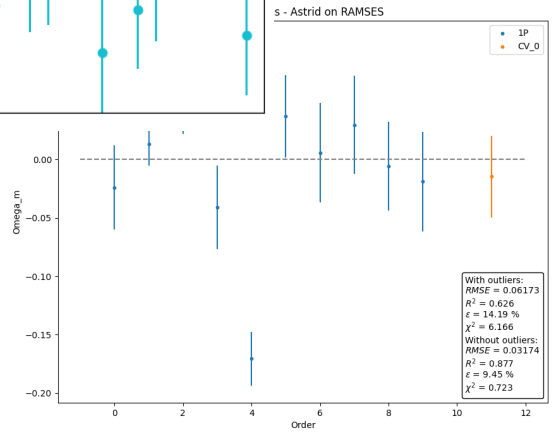


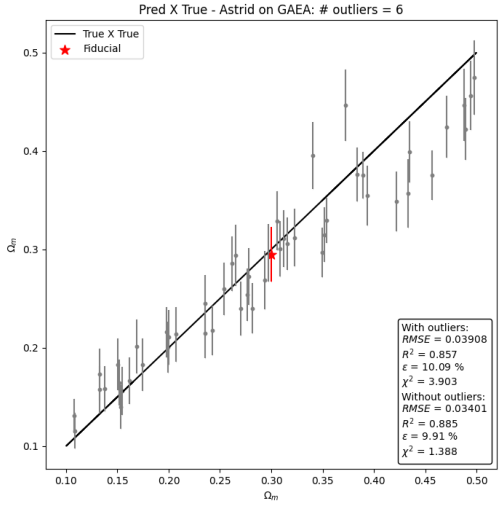












Robust field-level inference with GNNs: Interpretability

Helen Shao
(Princeton)

$$\mathbf{e}_{ij}^{(l+1)} = \phi_{l+1}([\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, \mathbf{e}_{ij}^{(l)}])$$

$$\mathbf{h}_i^{(l+1)} = \psi_{l+1}([\mathbf{h}_i^{(l)}, \bigoplus_{j\in\mathcal{N}_i} \mathbf{e}_{ij}^{(l+1)}, \mathbf{u}])$$

$$\mathbf{y} = \xi([\bigoplus_{i\in\mathcal{G}} \mathbf{h}_i^{(L)}, \mathbf{u}])$$

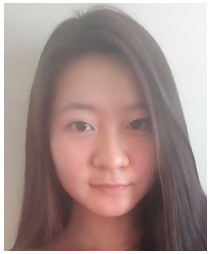
GNN Component	Formula
Edge Model: $e_1^{(1)}$	$1.32\lvert v_i - v_j + 0.21\rvert + 0.12(v_i - v_j) - 0.12(\gamma_{ij} + \beta_{ij} - 1.73)$
Edge Model: $e_2^{(1)}$	$\lvert 1.62(v_i - v_j) + 0.45\rvert + 1.98(v_i - v_j) + 0.55$
Node Model: $v_1^{(1)}$	$1.21^{v_i}(0.77^{3.29\sum_{j\in\mathcal{N}_j} e_1^{(1)} + \sum_{j\in\mathcal{N}_j} e_2^{(1)}}) + 0.12$
Node Model: $v_1^{(1)} + v_2^{(1)}$	$0.78 - \sqrt{\log(0.16^{\sum_{j\in\mathcal{N}_j} e_2 + \sum_{j\in\mathcal{N}_j} e_1 - 0.41 v_i - 1.05})} + 1.45$
Final MLP: μ_{Ω_m}	$4\times10^{-4}\cdot(-5.5\sum_{i\in\mathcal{G}} v_2^{(1)} + 2.21\sum_{i\in\mathcal{G}} v_1^{(1)} + \lvert 0.96\sum_{i\in\mathcal{G}} v_2^{(1)} + 0.82\sum_{i\in\mathcal{G}} v_1^{(1)}\rvert) - 0.103$

https://arxiv.org/abs/2302.14591

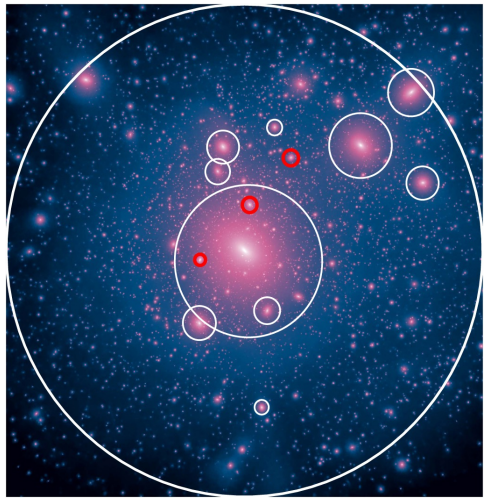
Work on thousands of halo and galaxy catalogs from 6 different N-body codes and 6 different hydrodynamic codes. Fundamental physics behind it; perhaps continuity equation.

Using AI to learn physics

Shao, FVN et al. 2021
(2109.04484)



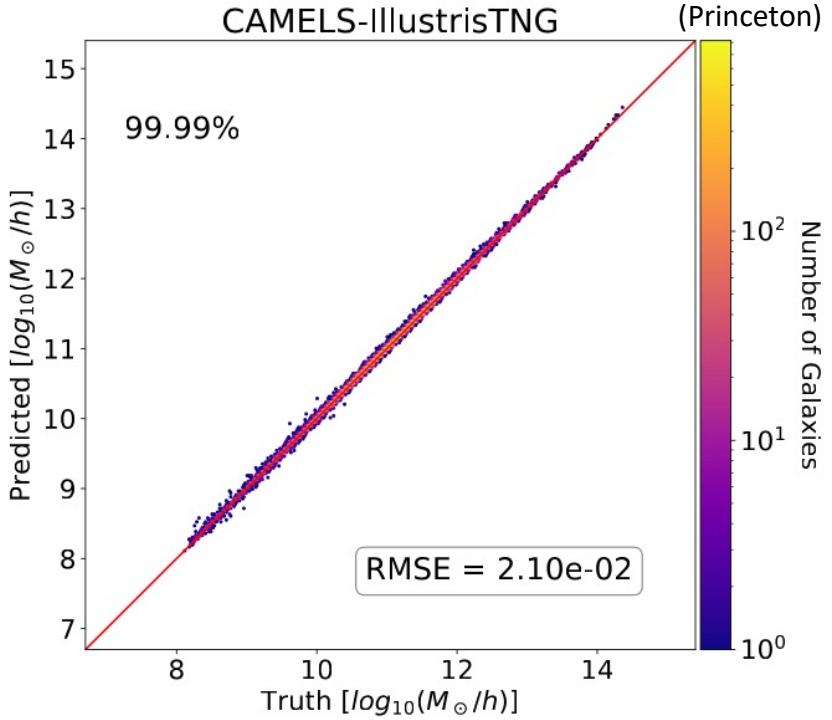
Helen Shao
(Princeton)



- Gas mass
- Stellar mass
- Black hole mass
- Gas metallicity
- Stellar metallicity
- Radius
- V_{max}
- Velocity dispersion
- Star-formation rate
- Spin

Total subhalo mass

$$M_{\rm tot} = A\sigma^{(\alpha_0+\alpha_1 \log \sigma)} R^{(\beta_0+\beta_1 \log R)} V_{\rm max}^{(\gamma_0+\gamma_1 \log V_{\rm max})}$$

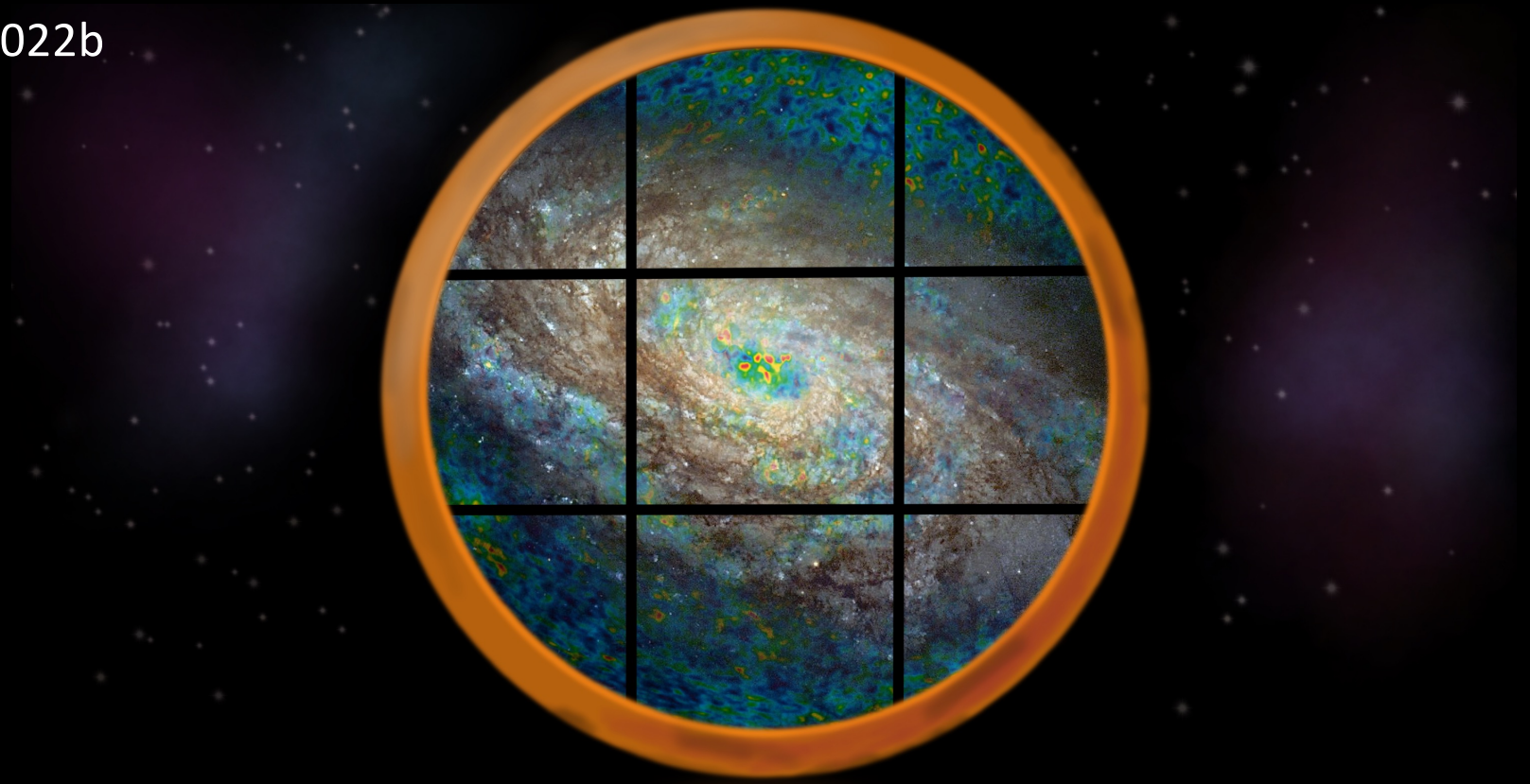


It works for any subhalo (central or satellite) containing any type of galaxy at any redshift from simulations with different cosmologies, different astrophysics, different subgrid physics, different resolutions, and different volumes.

Some version of the virial theorem

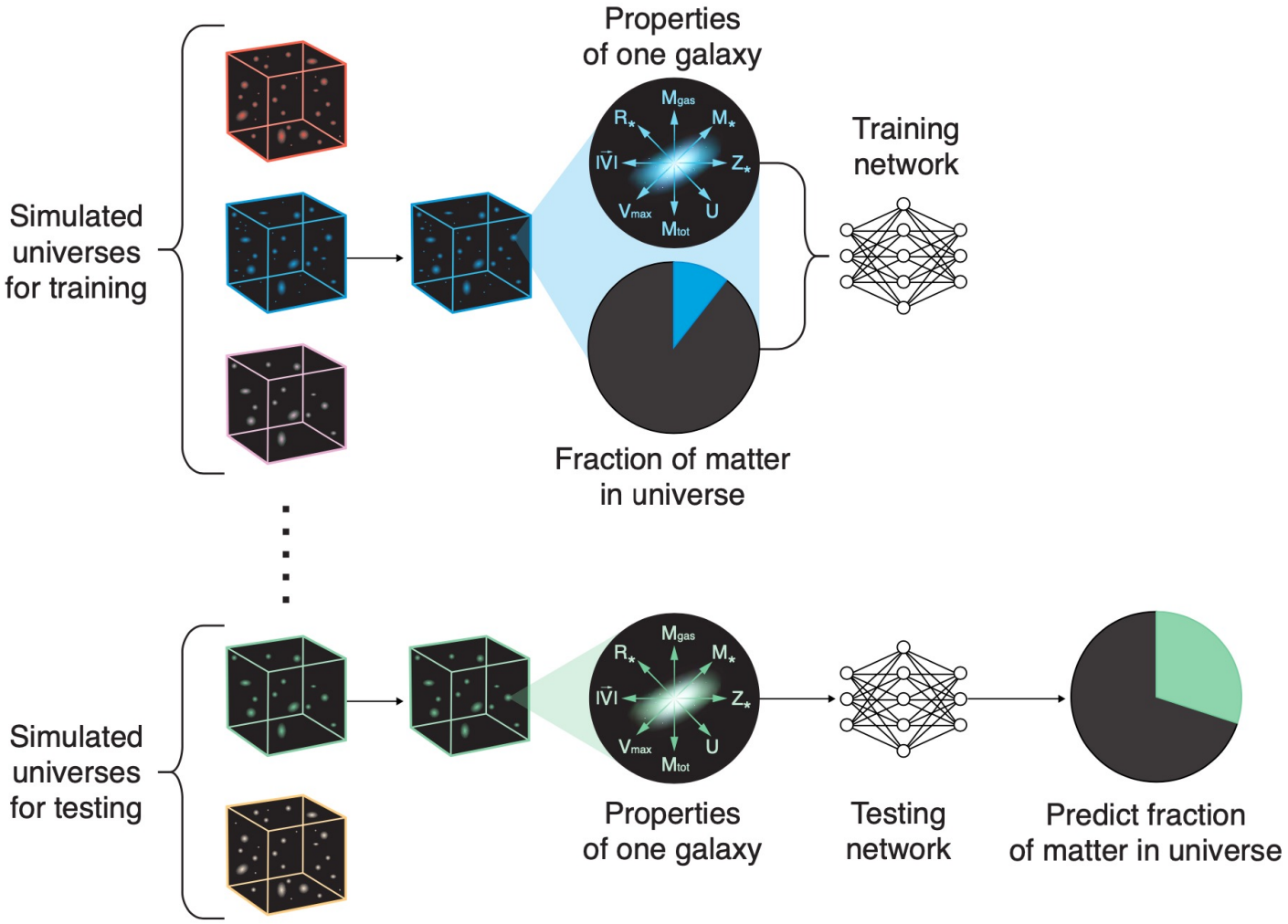
Cosmology with 1 galaxy?

FVN et al. 2022b

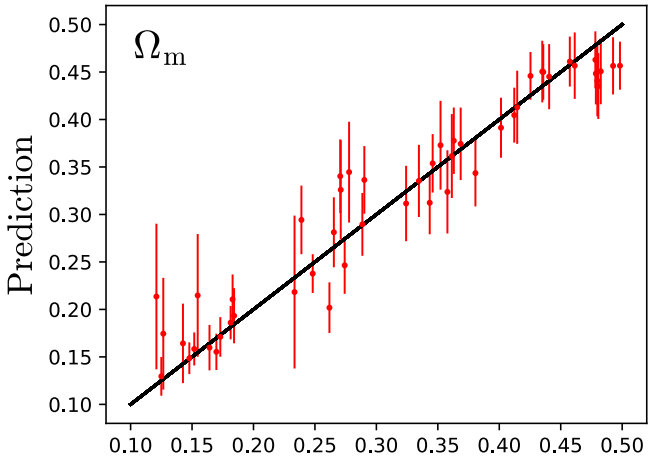


In collaboration with Jupiter Ding, Shy Genel, Stephanie Tonnesen, Valentina La Torre, David Spergel, Romain Teyssier, Yin Li, Caroline Heneka, Pablo Lesmos, Daniel Angles-Alcazar, Daisuke Nagai

Cosmology with one galaxy: setup



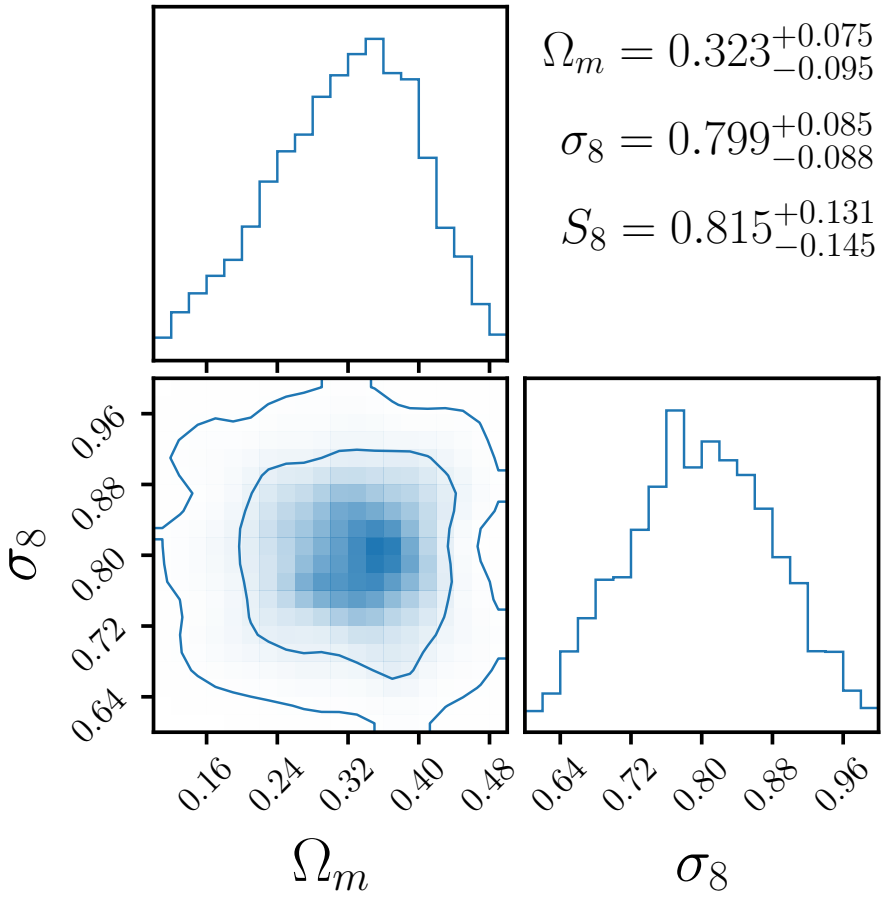
Cosmology with 1 galaxy



cosmological constraints from *only* the observed photometry of *22,338* NASA-Sloan Atlas galaxies

ChangHoon Hahn
(Princeton)

2310.08634



Caveats:

- Galaxy formation
- SED model

We can be picky with galaxies.
E.g. only take low dust, well constrained IMF…etc

Or marginalize over these parameters

Conclusions

- CAMELS
  - Largest and most diverse set of cosmological hydrodynamic simulations ever run
  - Designed to learn cosmology and astrophysics with artificial intelligence
  - Most of the data publicly available via globus and binder: 1+ Petabyte

- Dark CAMELS
  - Largest and most diverse set of simulations beyond CDM
  - CDM, WDM, ETHOS, SIDM, ADM…etc
  - Cosmological volumes, Milky-way zoom-ins, dwarf zoom-ins

- Quijote
  - Largest set of N-body simulations ever run: 77,000+ simulations
  - $\Lambda\mathrm{CDM}$, neutrinos, dark energy, primordial non-Gaussianities, parity violation, tails, modified gravity
  - Most of the data publicly available via globus and binder: 1+ Petabyte