A Forward Modeling Approach to Understanding the Light from Galaxies

2023/10/10

BCCP/Cosmology Seminar

Suchetha Cooray

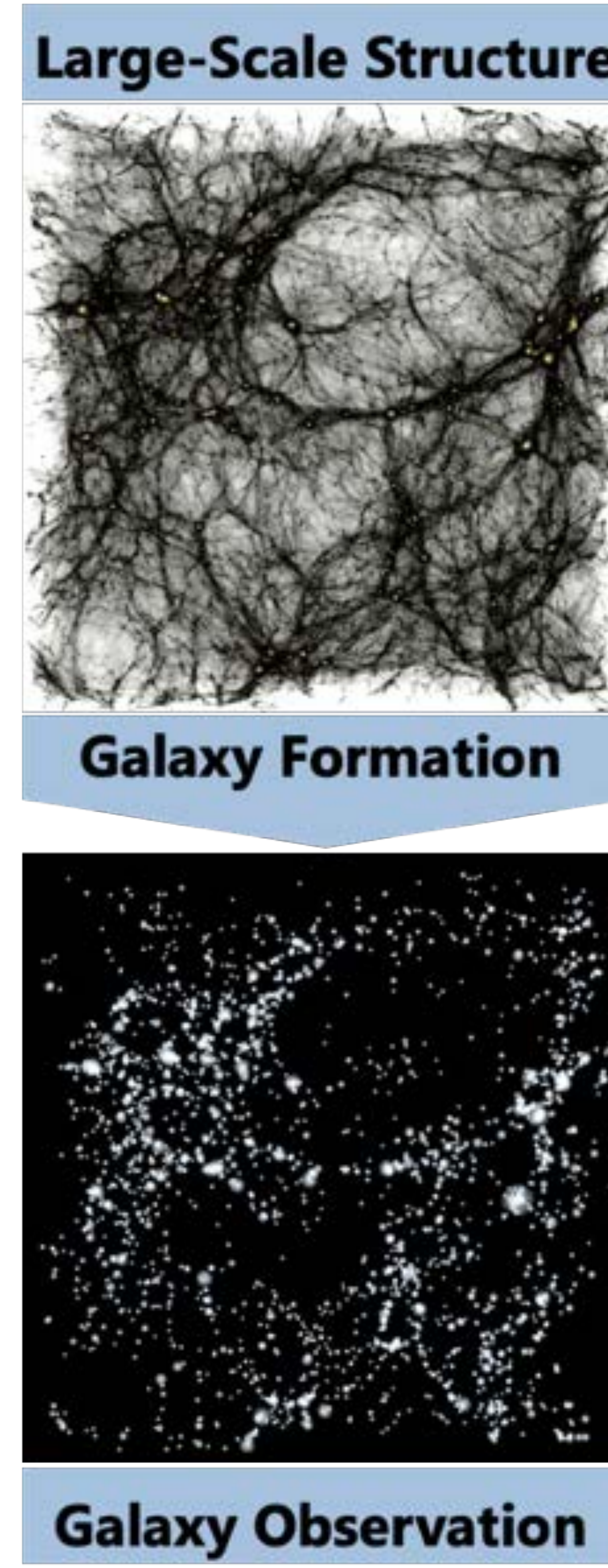
Suchetha Cooray

Nagoya U. Doctor of Science 2023
JSPS Postdoctoral Fellow at NAOJ

National Astronomical Observatory of Japan

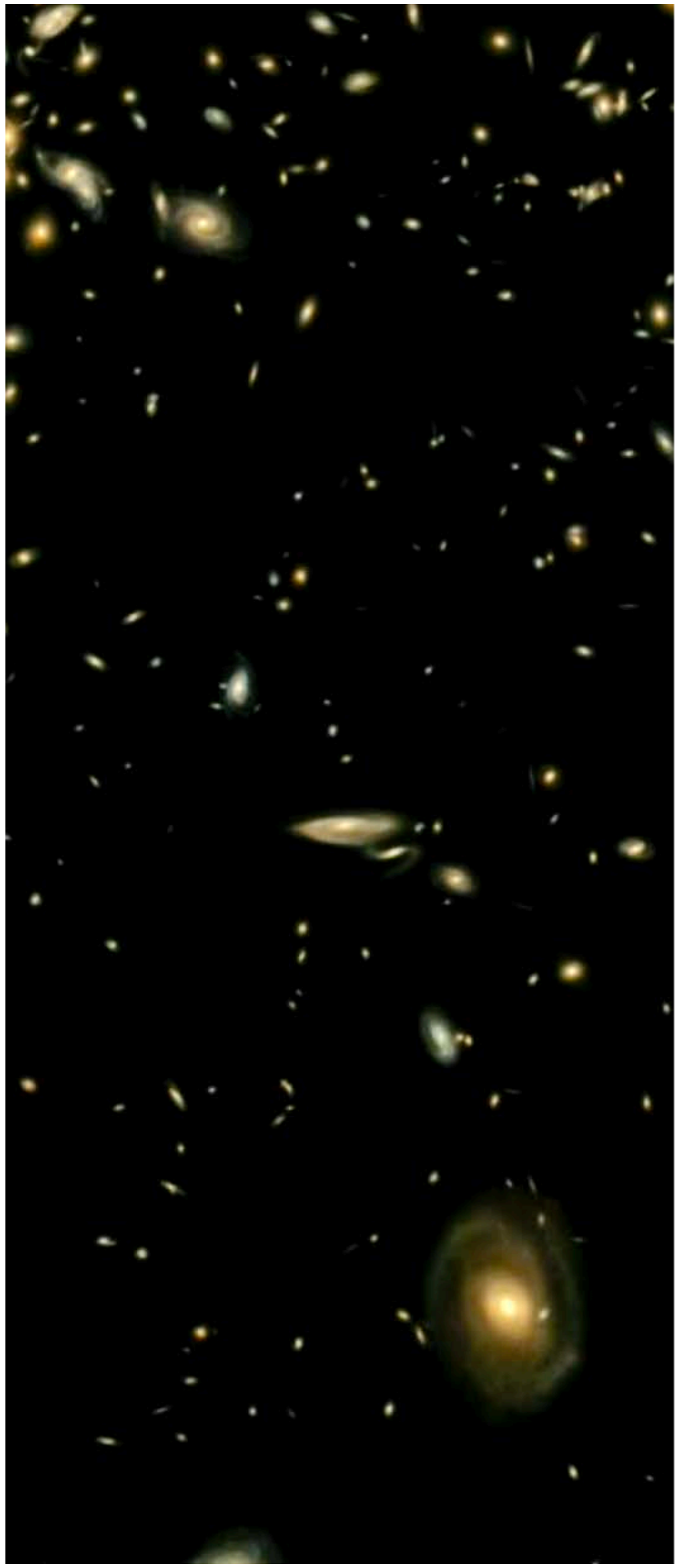
- *Understanding galaxy formation using ML with simulations and observations*

- Some interests are;
  - Forward modeling all galaxy observables within cosmological structure formation
  - Learning latent representations of galaxies
- Soccer, music, and socializing over drinks…
- In the job market & open for collaborations!



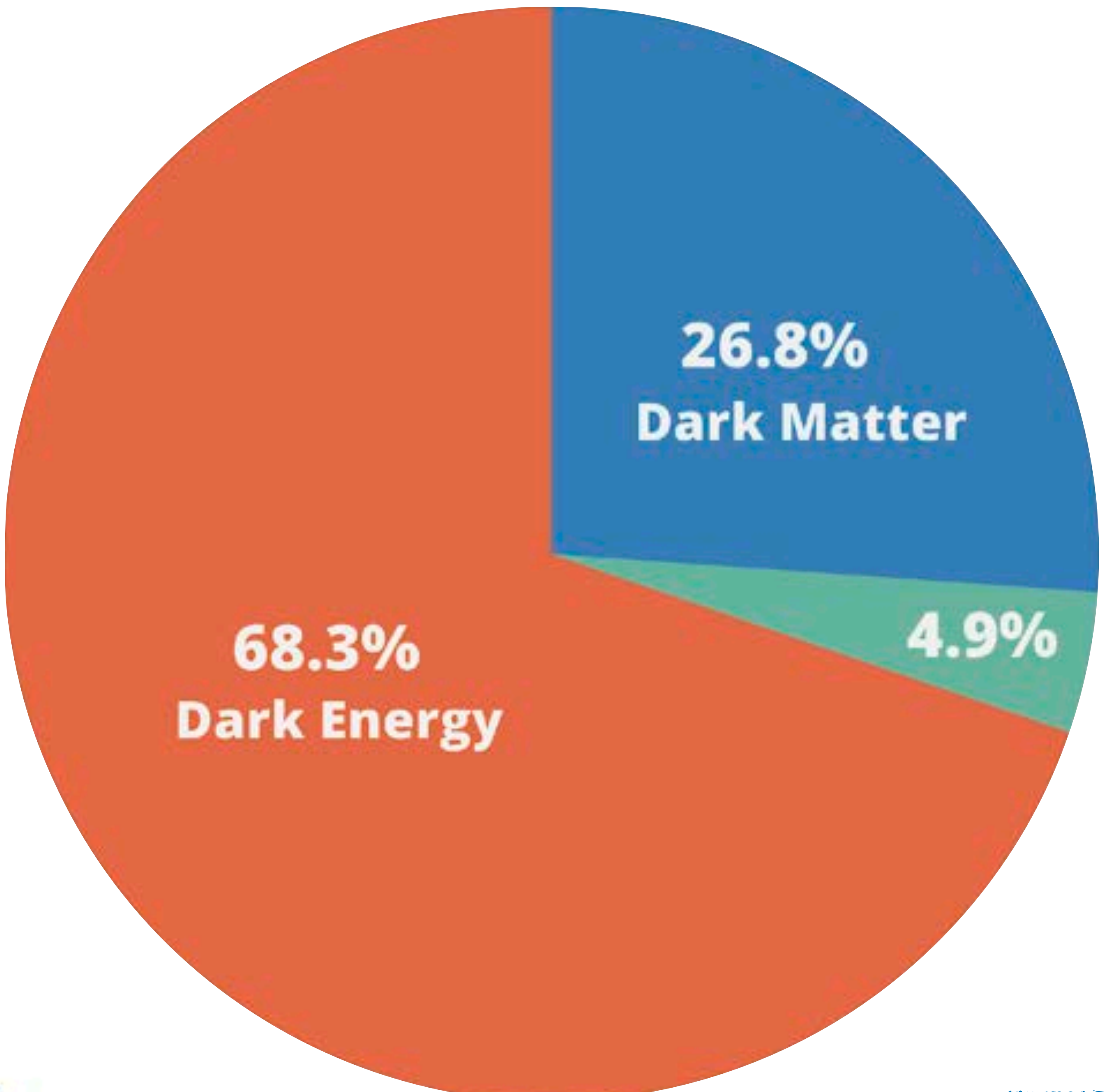
Talk Outline

- ➤ Why study galaxies?
- ➤ What is empirical forward modeling for galaxy formation?
- ➤ Challenges with current approaches to studying galaxies
- ➤ Proposed solution (field-level inference)
- ➤ Future of empirical forward modeling for galaxy formation and cosmology



Mock universe with UM

Why study galaxies in the Universe?



Baryons

<1% in galaxies!

Why study them?

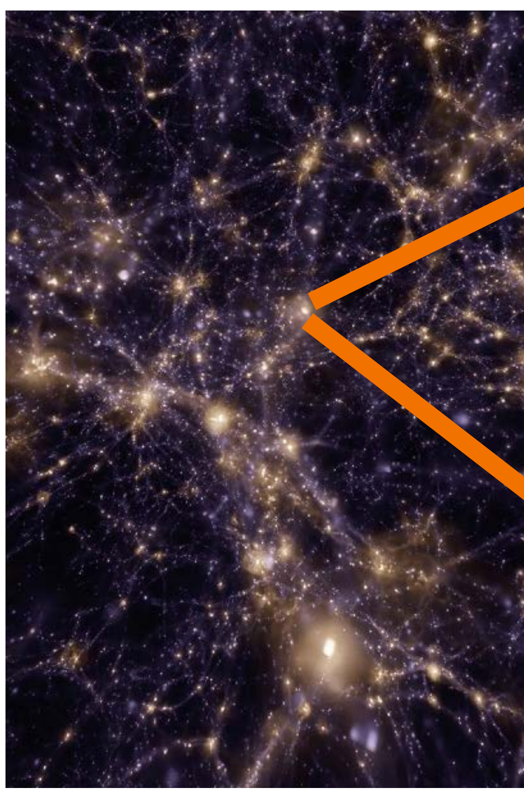
Because galaxies are what we see!

Difficulty in Understanding Galaxies

- Gas flows/IGM/CGM
- Dust/ISM
- Blackholes/AGNs
- Supernovae/Stellar evolution
- Chemical evolution
- Magnetic fields
- Cosmic rays
- Reionization
- And so on…

Difficulty in Understanding Galaxies

The key driver of galaxy formation, i.e., star formation,
is a multi-scale process



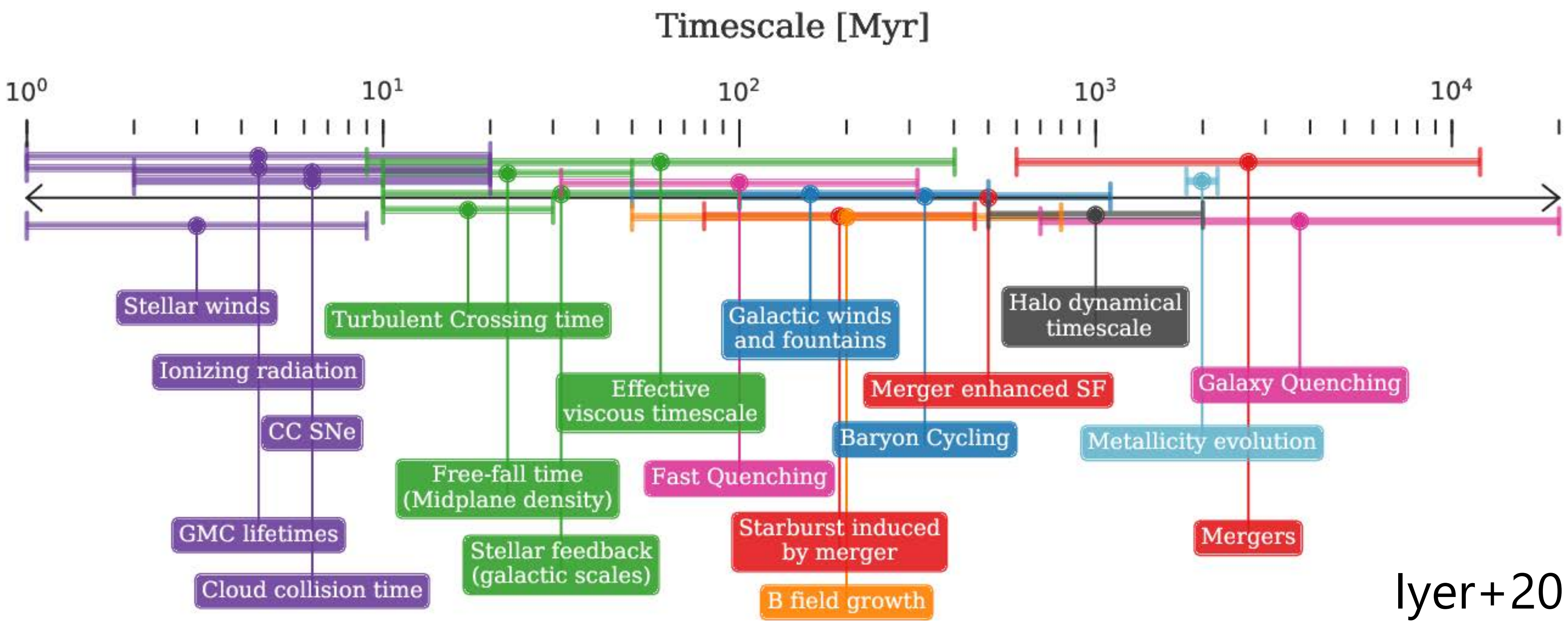
Large-scale physics

Galaxy

Small-scale physics

Difficulty in Understanding Galaxies

Physics and their times scales regulating star formation histories



Iyer+20

GMC Physics

Interaction between ISM and CGM

Mergers

Chemistry

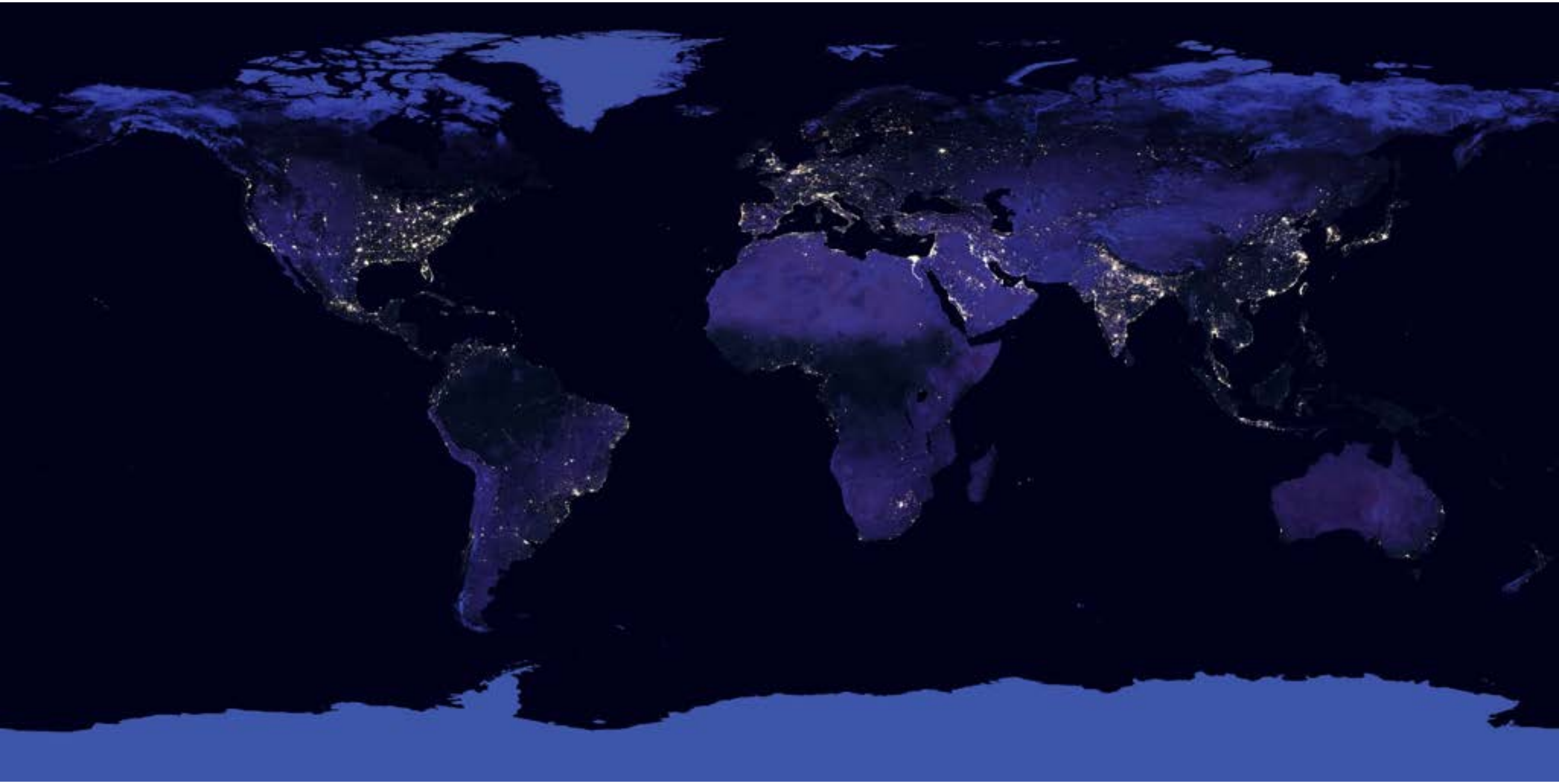
Quenching

Dynamics within galaxies

Magnetic Fields

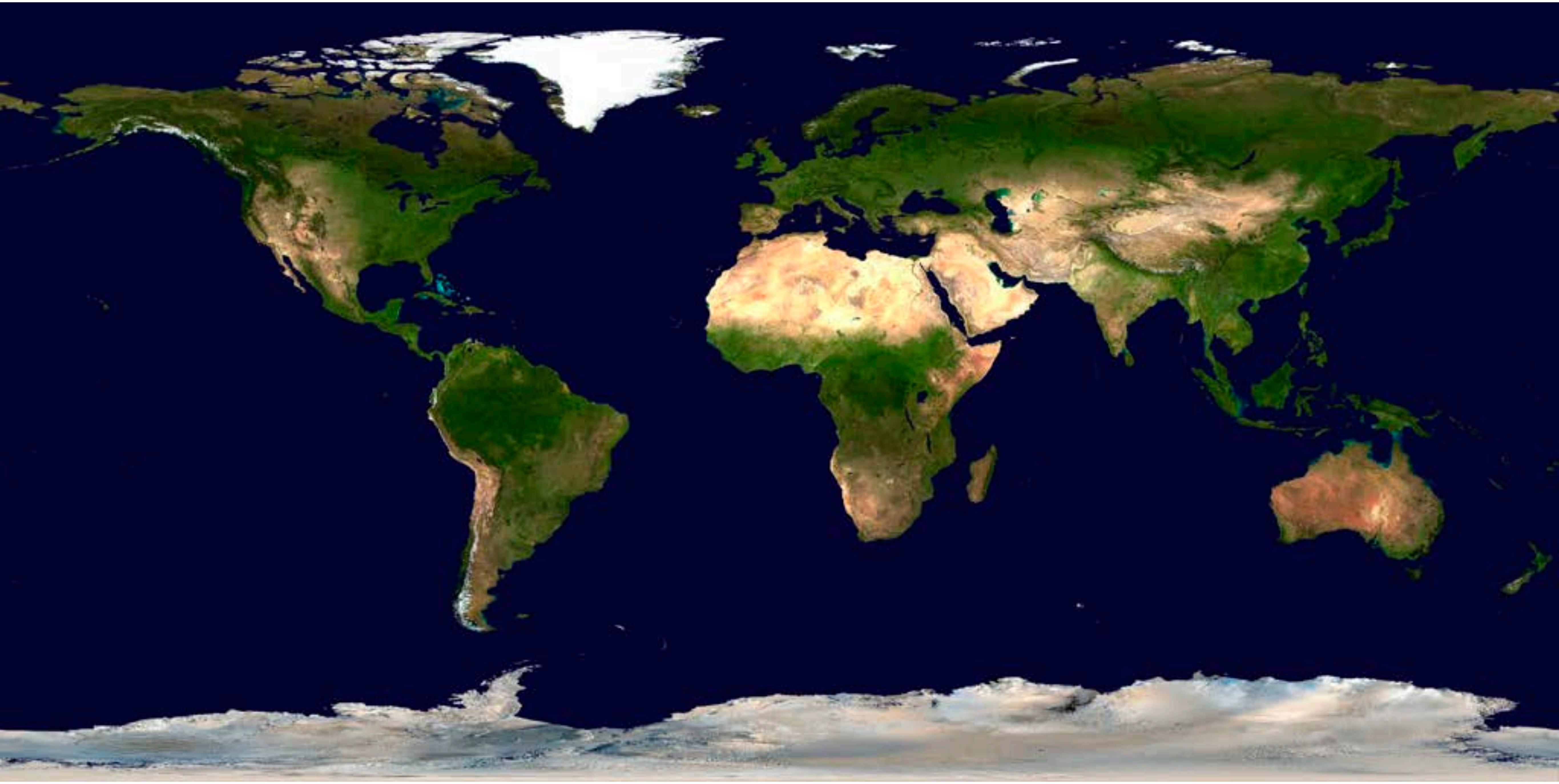
Galaxy environment & DM

Galaxies in the Cosmology Context



NASA

Galaxies in the Cosmology Context

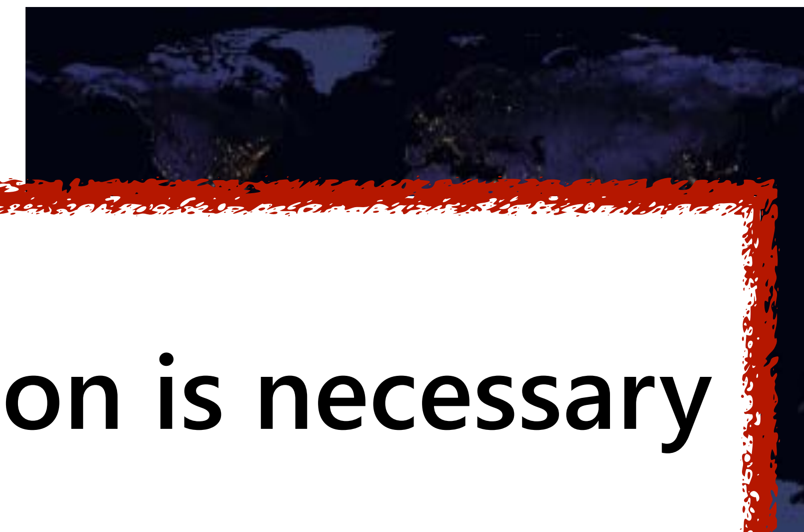


NASA

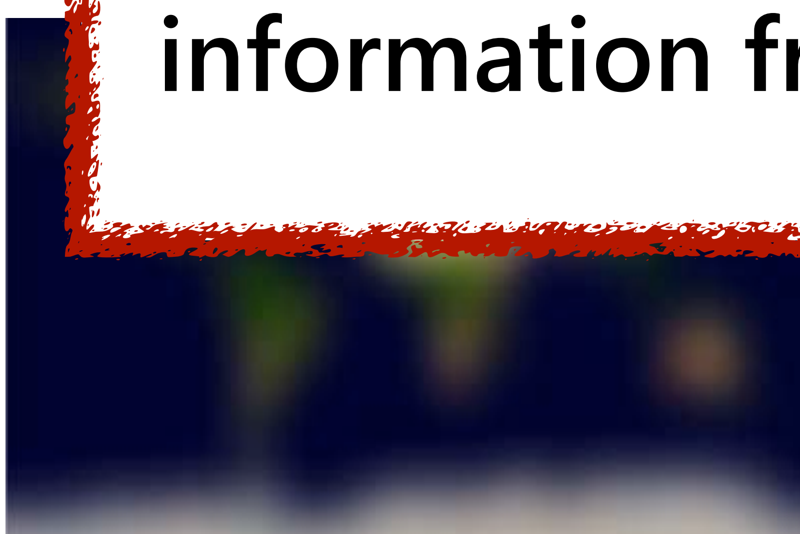
Galaxies in the Cosmology Context

Dark Matter (DM) Simulations

Wide-Field Galaxy Observations



Understanding galaxy formation is necessary for extracting the maximal cosmological information from upcoming galaxy surveys!

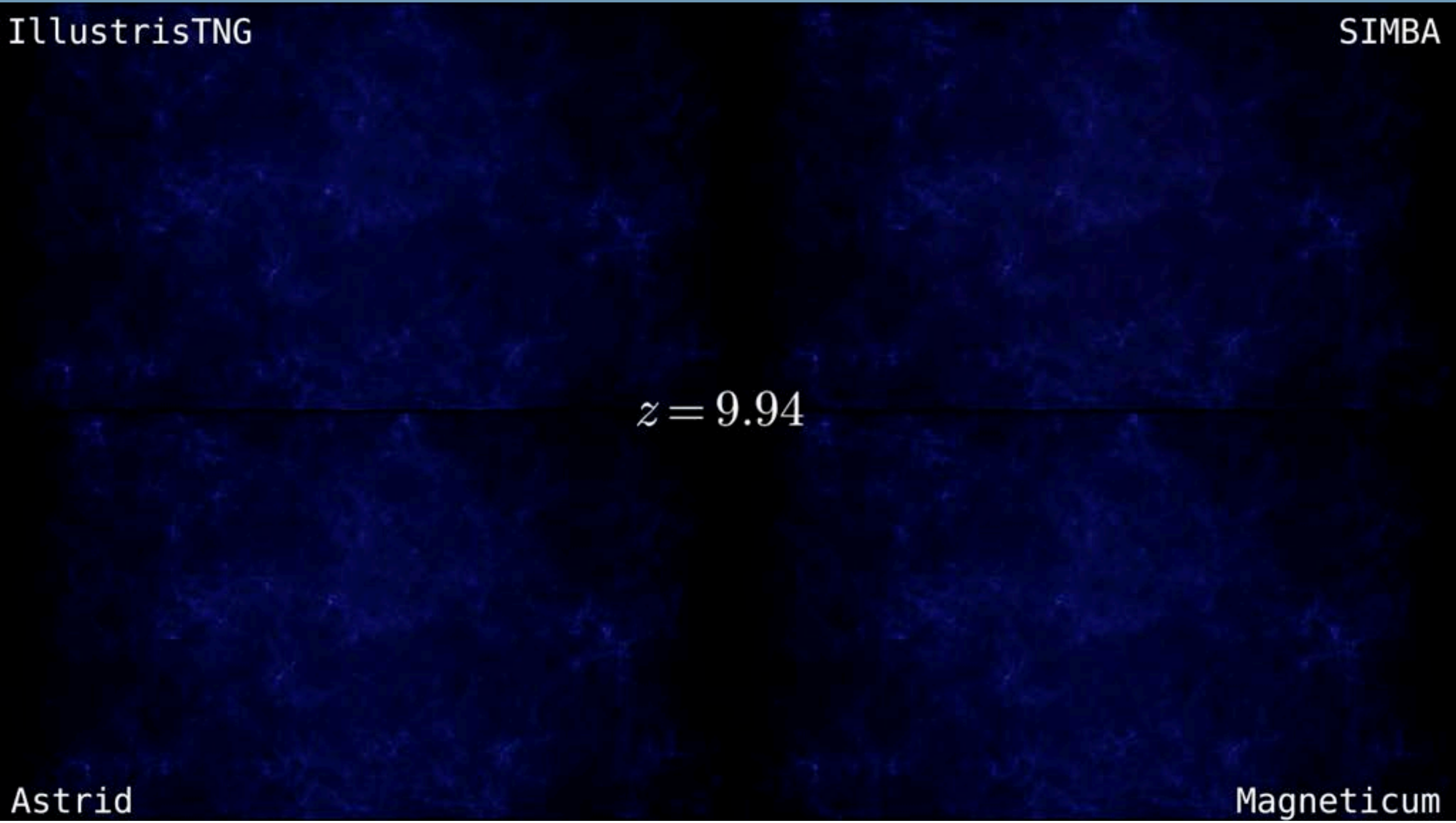


Linear theory of structure formation, halo model, and halo occupation distributions

Smoothed DM Distribution

Smoothed Galaxy Distribution

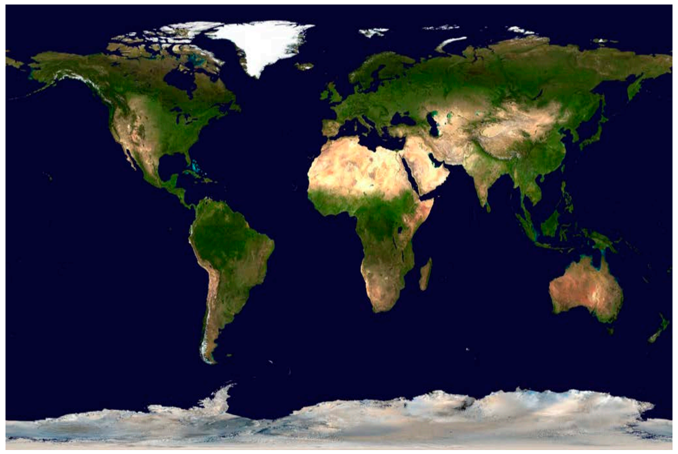
Significant Progress in Simulations!



CAMELS

We just don't know if they are right!

Can we Learn Galaxy Formation from the Universe?



High Res. DM Distribution

Galaxy Formation

Challenging!

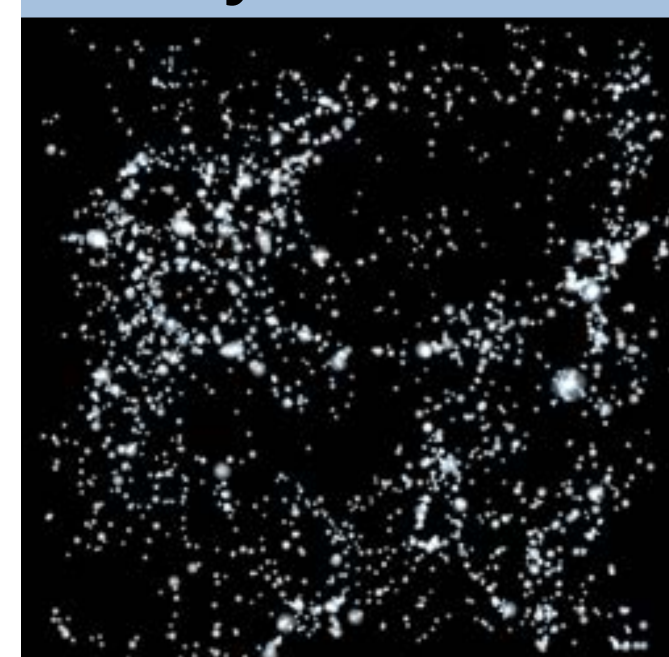
Galaxy Distribution

Cosmology + N-body

Construct a model that is flexible to be able to capture the real galaxy formation physics while being physically consistent (e.g, conservation)

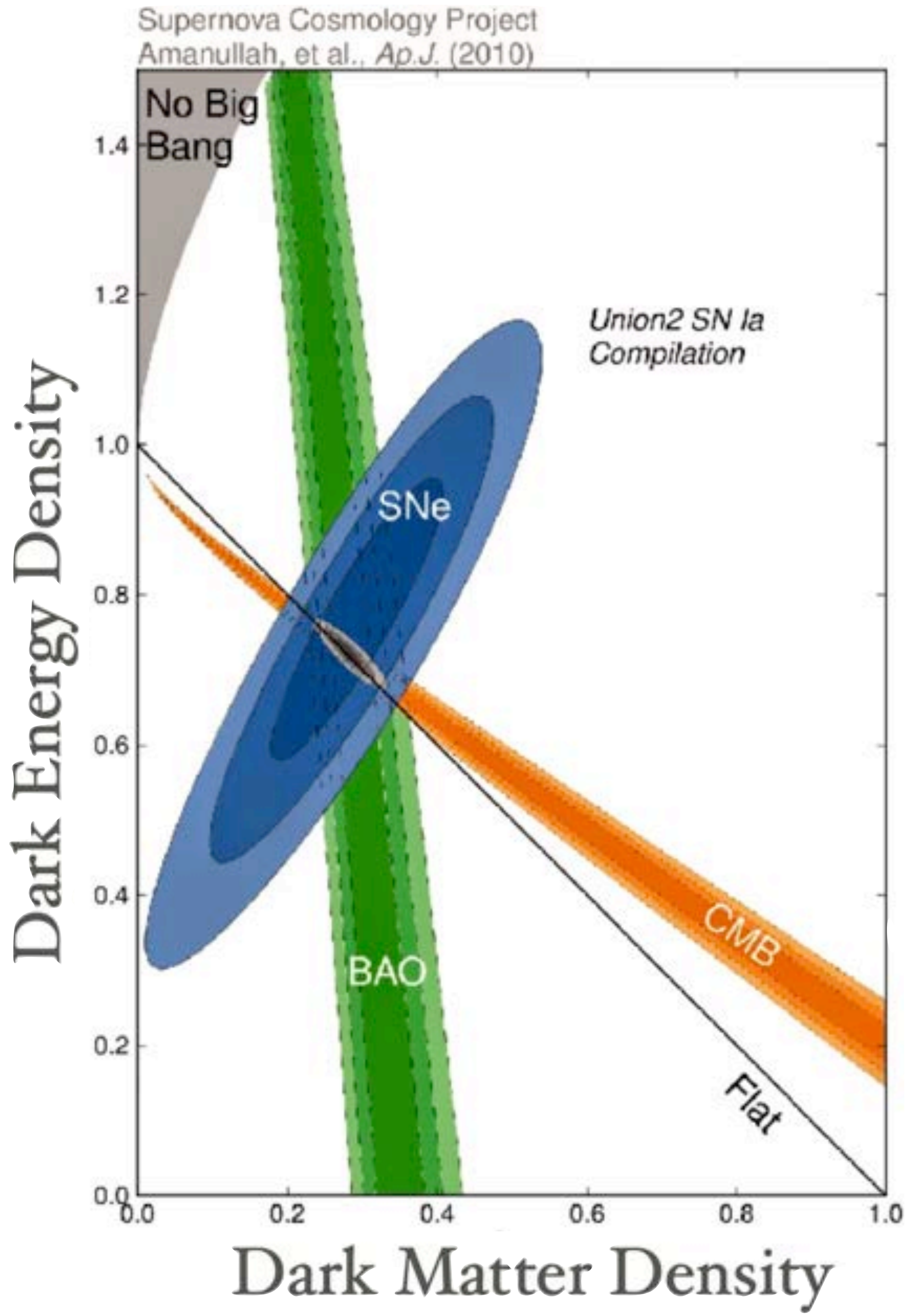
=> Let the machine learn!

Galaxy Observation

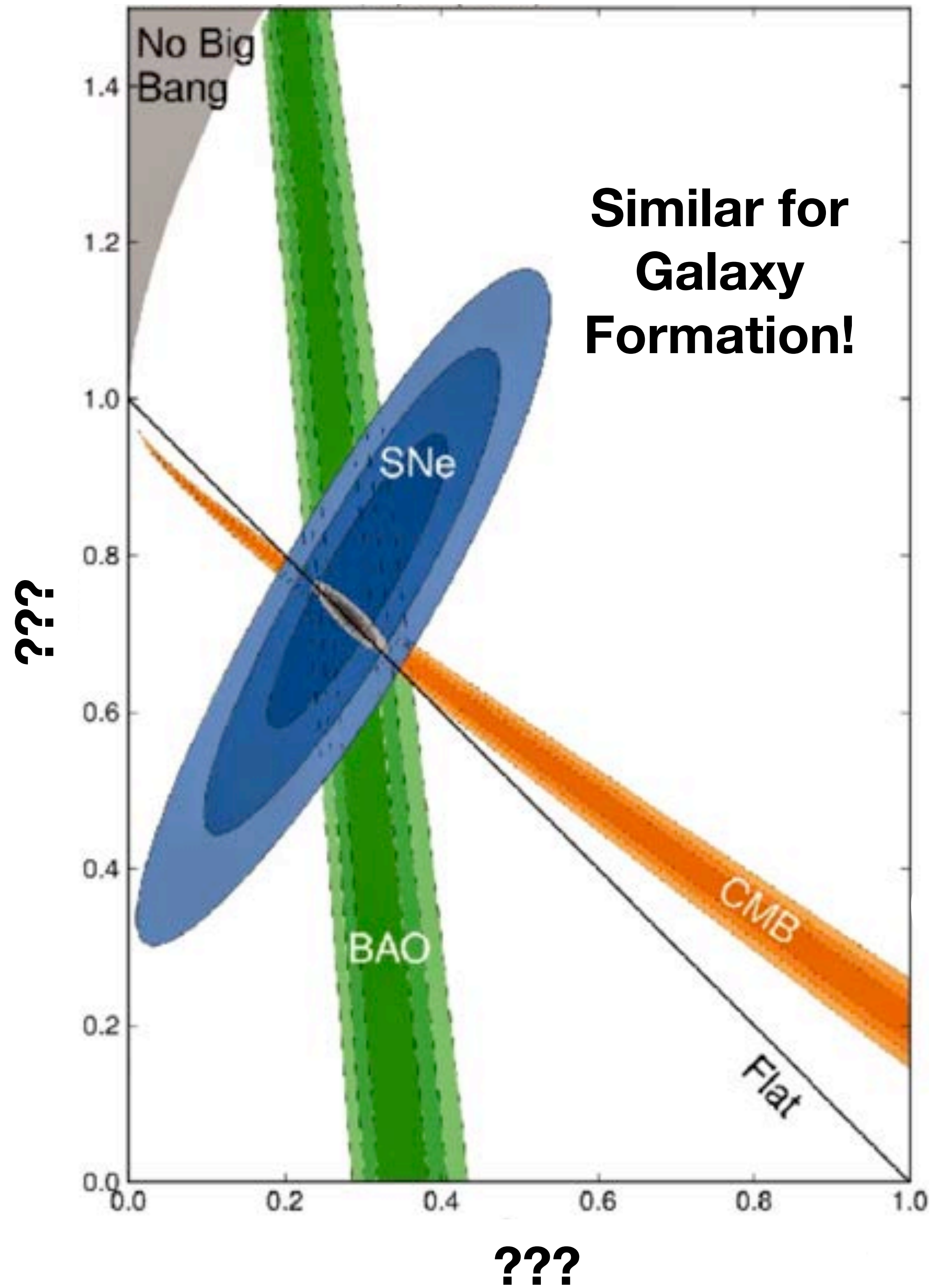


Wechsler & Tinker 2018

Combining Various Observations

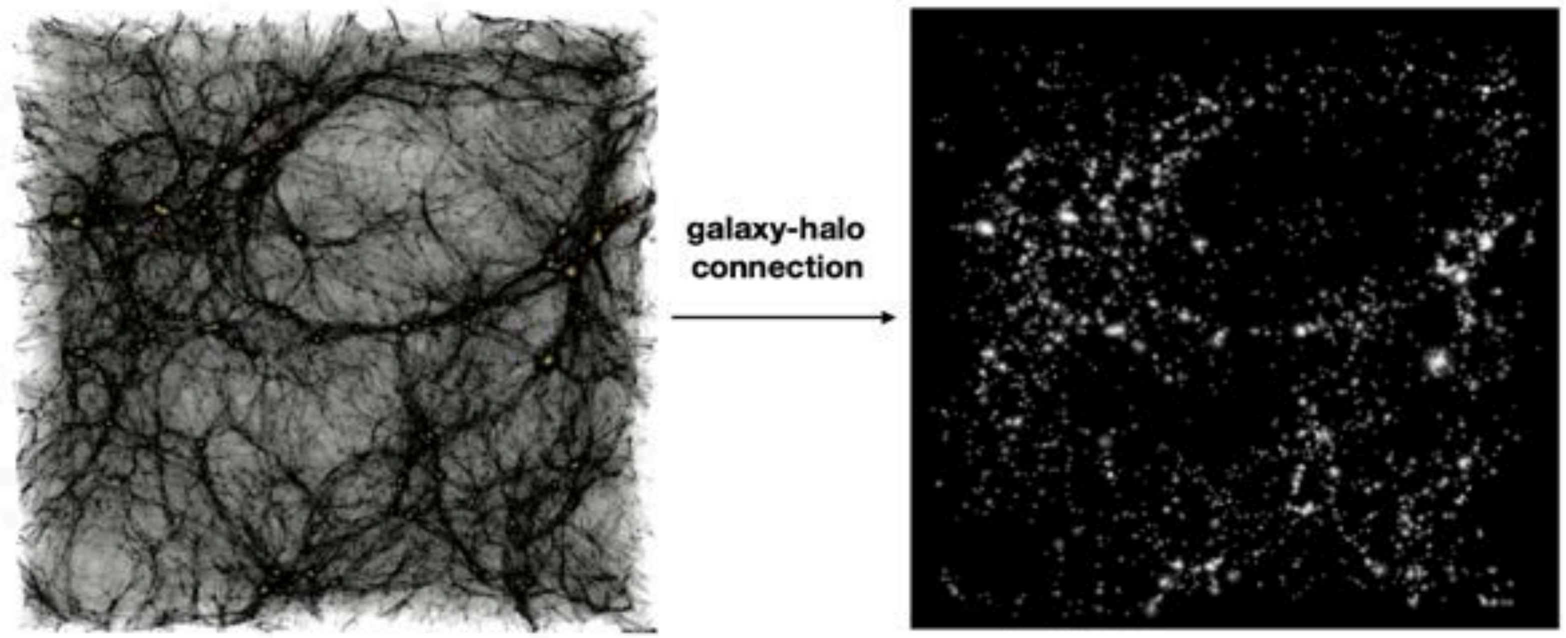


Combining Various Observations



Our Approach

Wechsler & Tinker 2018



Approaches to modeling the galaxy-halo connection

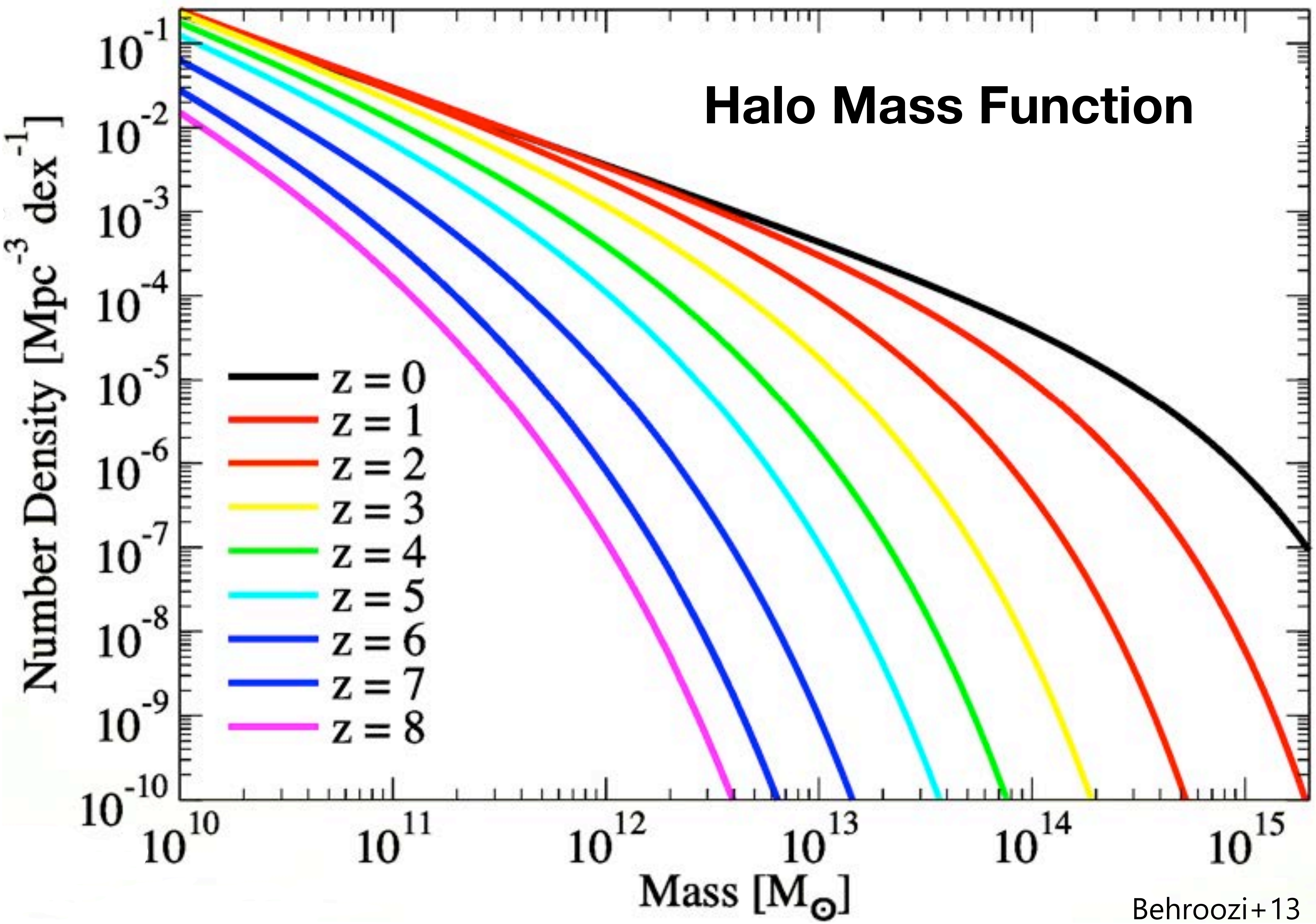
← physical models | empirical models →

Hydrodynamical Simulations	Semi-analytic Models	Empirical Forward Modeling	Subhalo Abundance Modeling	Halo Occupation Models
Simulate halos & gas; Star formation & feedback recipes	Evolution of density peaks plus recipes for gas cooling, star formation, feedback	Evolution of density peaks plus parameterized star formation rates	Density peaks (halos & subhalos) plus assumptions about galaxy—(sub)halo connection	Collapsed objects (halos) plus model for distribution of galaxy number given host halo properties

- **UniverseMachine (Behroozi+19) is an empirical galaxy formation model that has been constrained on the latest observations and produces consistent galaxy growth histories from z=10 to z=0**

UniverseMachine (Behroozi+19)

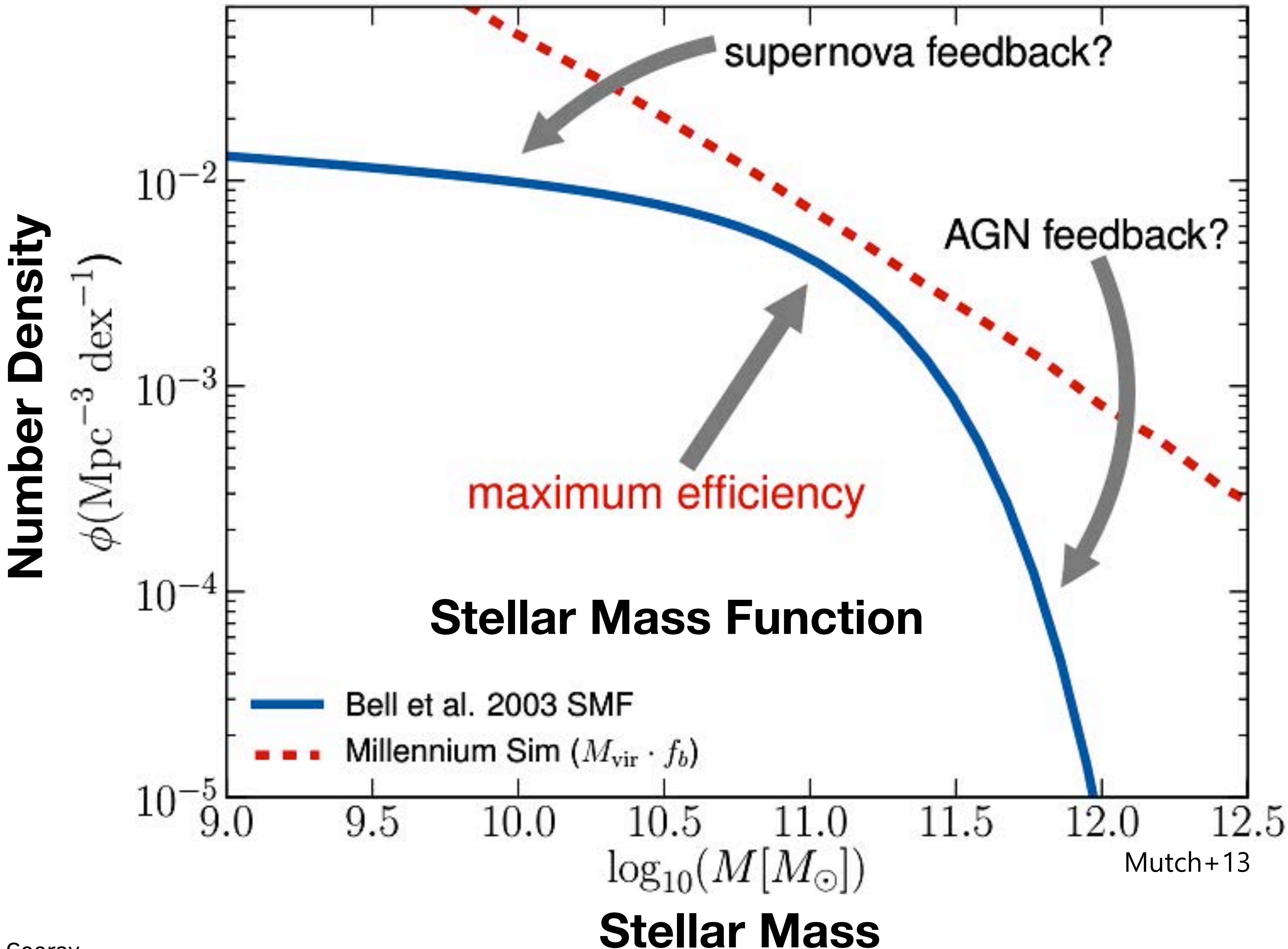
Key Concepts - HMF



Behroozi+13

DM Halo Mass

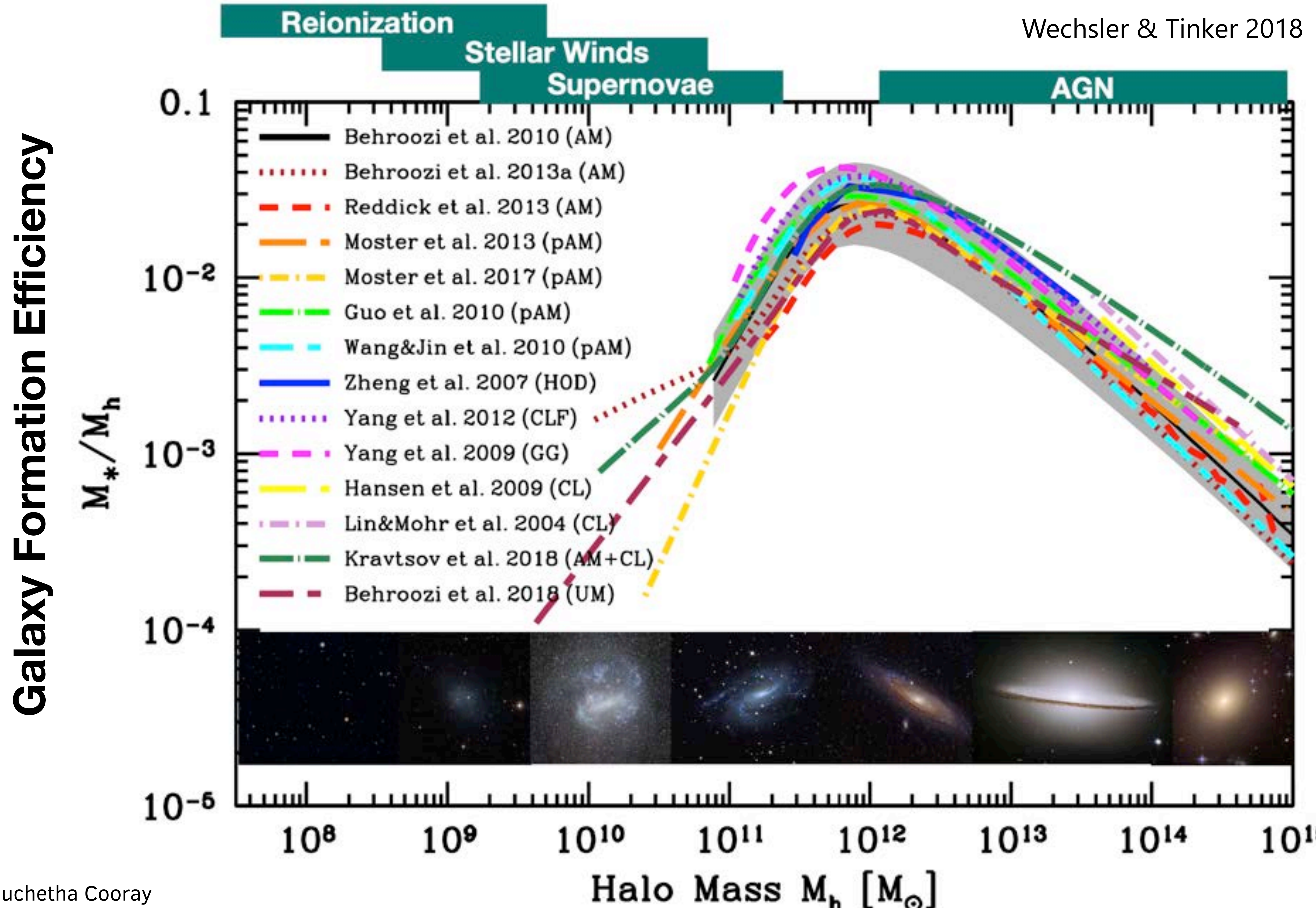
Key Concepts - SMF



Mutch+13

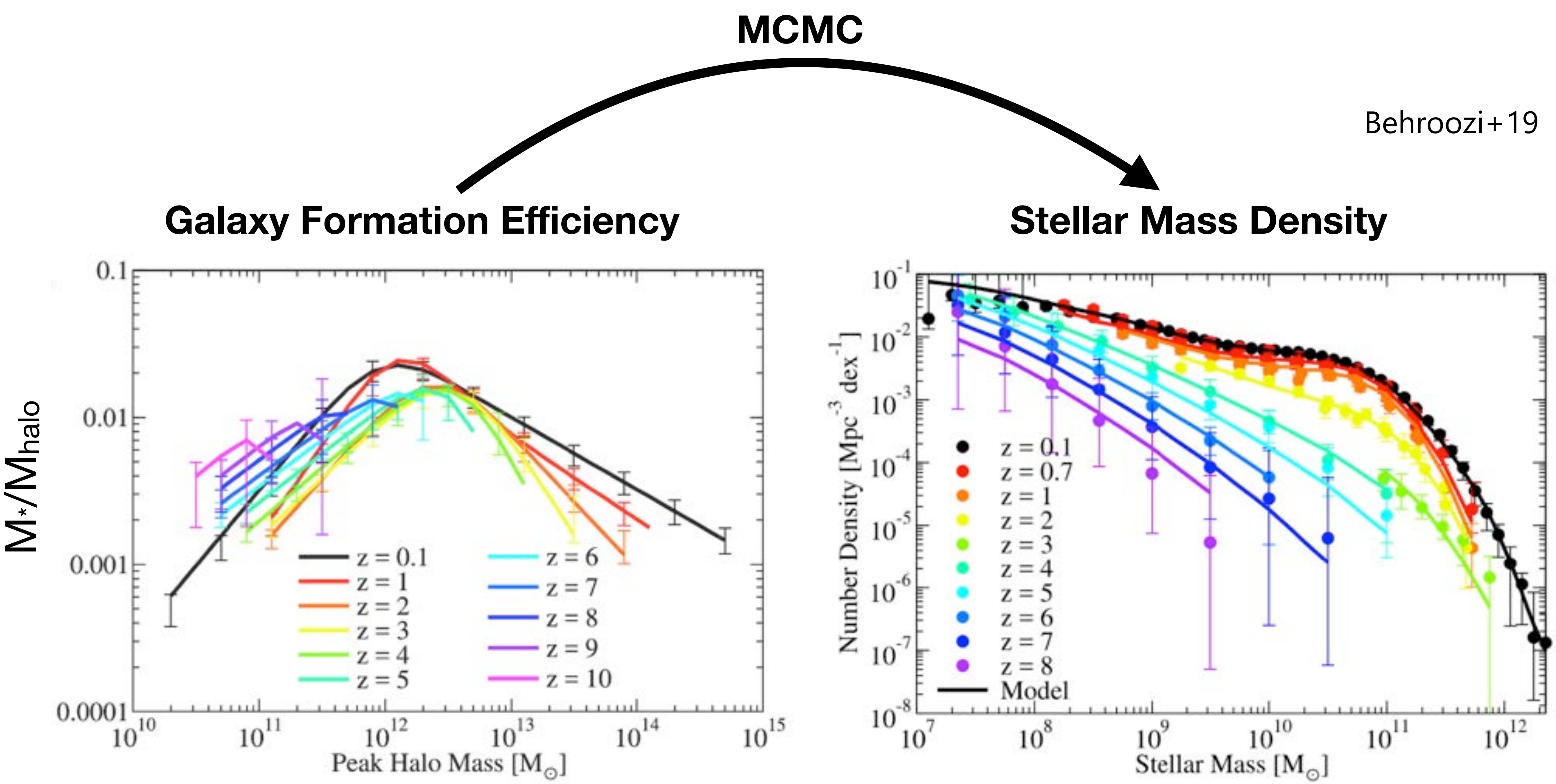
Key Concepts - SMHM Relation

Wechsler & Tinker 2018



Empirical Galaxy Formation Models

Behroozi+19



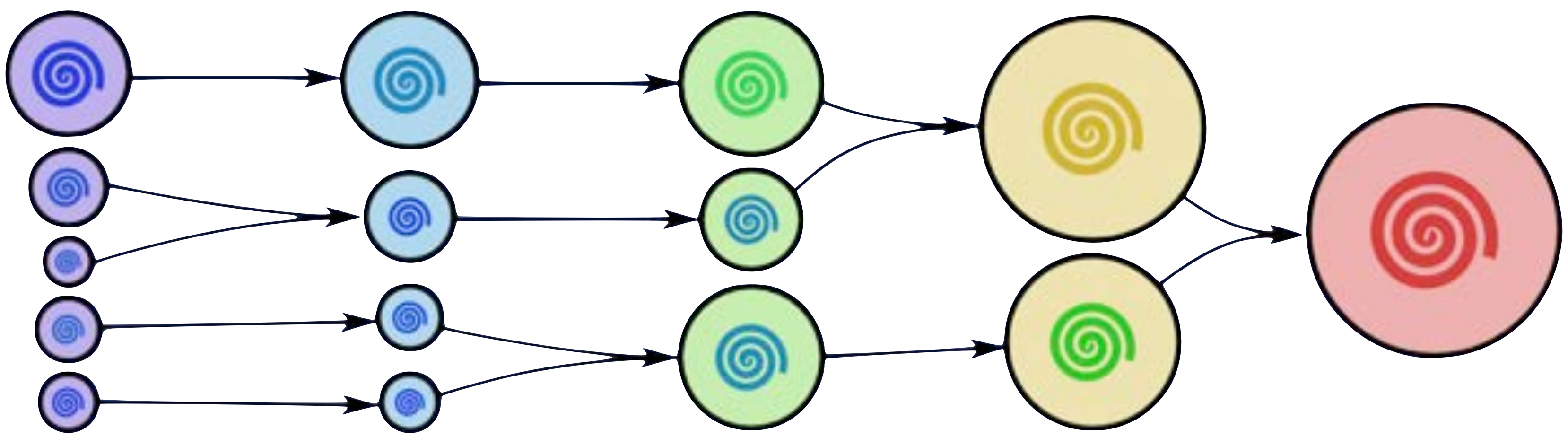
UniverseMachine Procedure

Cosmology + N-body

Minimal Galaxy
Formation Model

$$SFR = f(M_h, \dot{M}_h, z)$$

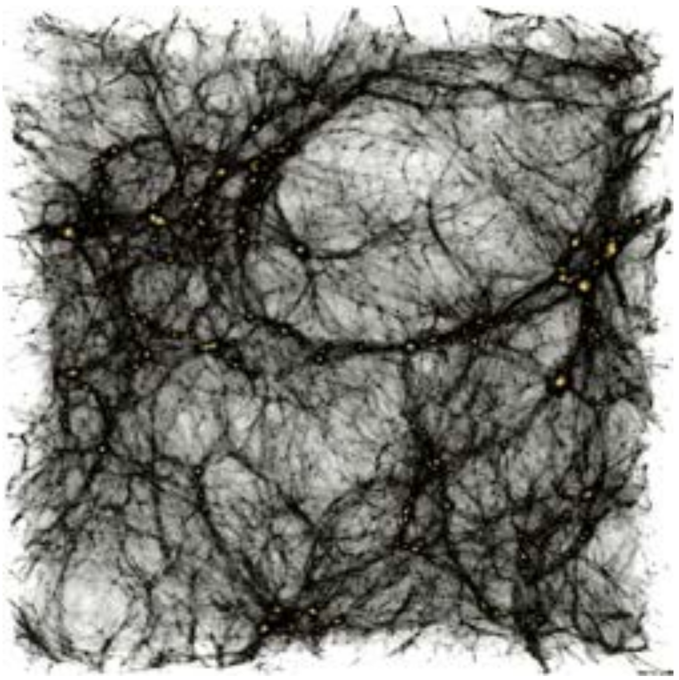
Galaxy Distribution



Populated Dark Matter Merger Trees

UniverseMachine Procedure

Cosmology + N-body

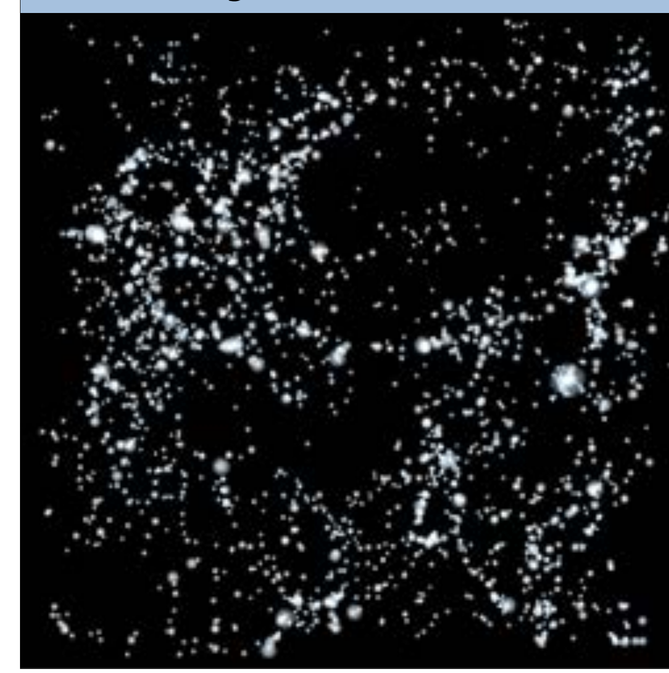


Minimal Galaxy Formation Model

$$SFR = f(M_h, \dot{M}_h, z)$$

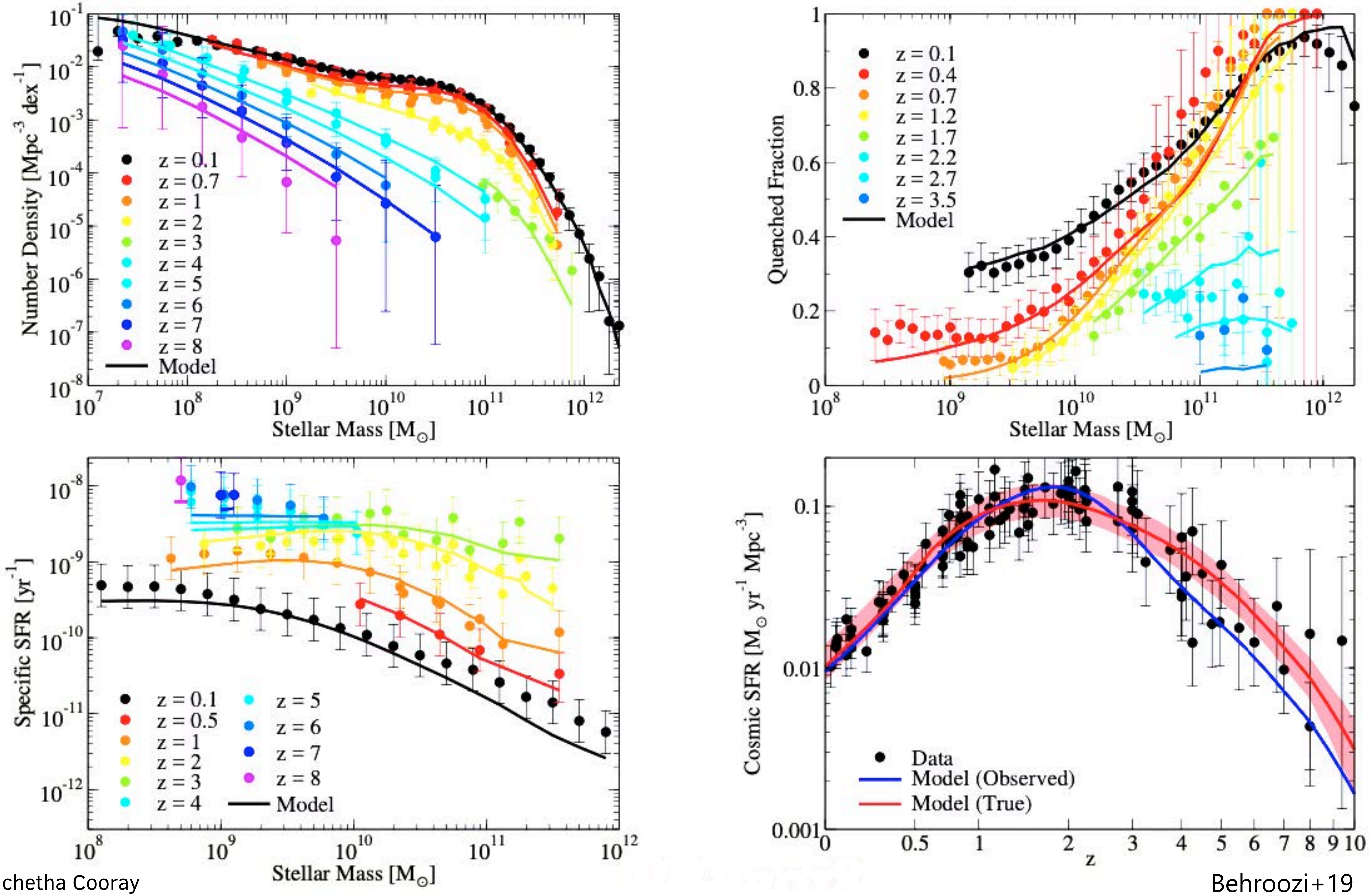
Constrain w/ Observations

Galaxy Distribution



Constrain based on observed stellar mass functions, SFRs (specific and cosmic), quenched fractions, ultraviolet (UV) luminosity functions, UV–stellar mass relations, IRX–UV relations, auto- and cross-correlation functions (including quenched and star-forming subsamples), and quenching dependence on environment

UniverseMachine Results - Observational Constraints



Suchetha Cooray

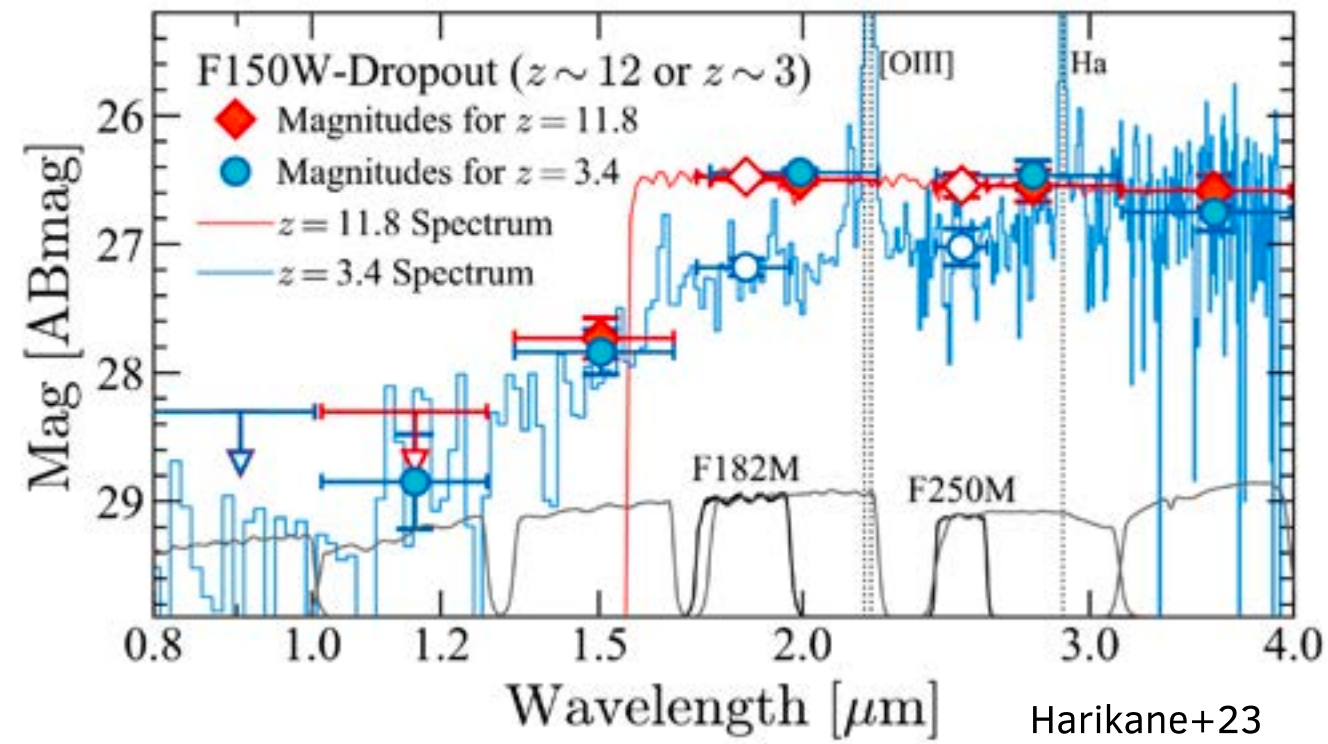
Behroozi+19

Uncertainties in the Galaxy-Halo Connection

Challenge of Extracting Galaxy Physical Properties

- Main source of uncertainty is in galaxy property measurements!
- Degeneracy between effects of **redshift, star formation history, dust, and metallicities** on the SED (e.g., Calzetti et al. 2001)
- Recovered galaxy physical properties are significantly and systematically affected by modeling assumptions

Redshift-SED Degeneracy



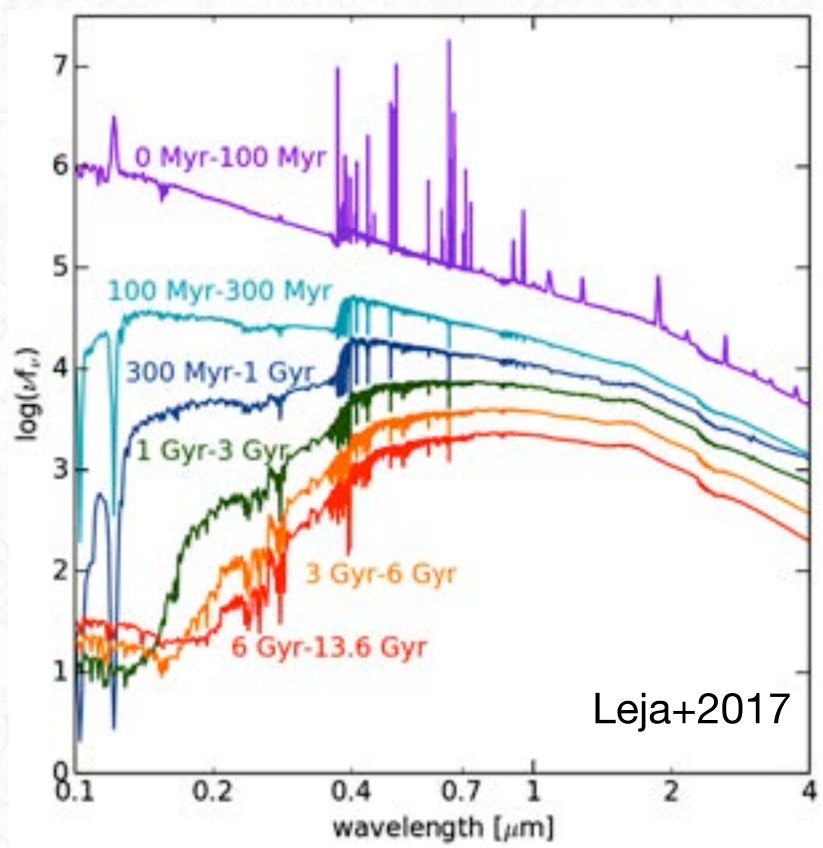
Harikane+23

SED Modeling Assumption	Stellar Mass Error*
Star Formation Histories	~0.2 dex
Dust	~0.2 dex
Metallicity	~0.1 dex
Photometric Redshifts	~0.1 dex
Stellar Population Synthesis	~0.1 dex
Total (in quadrature)	**~0.35 dex**

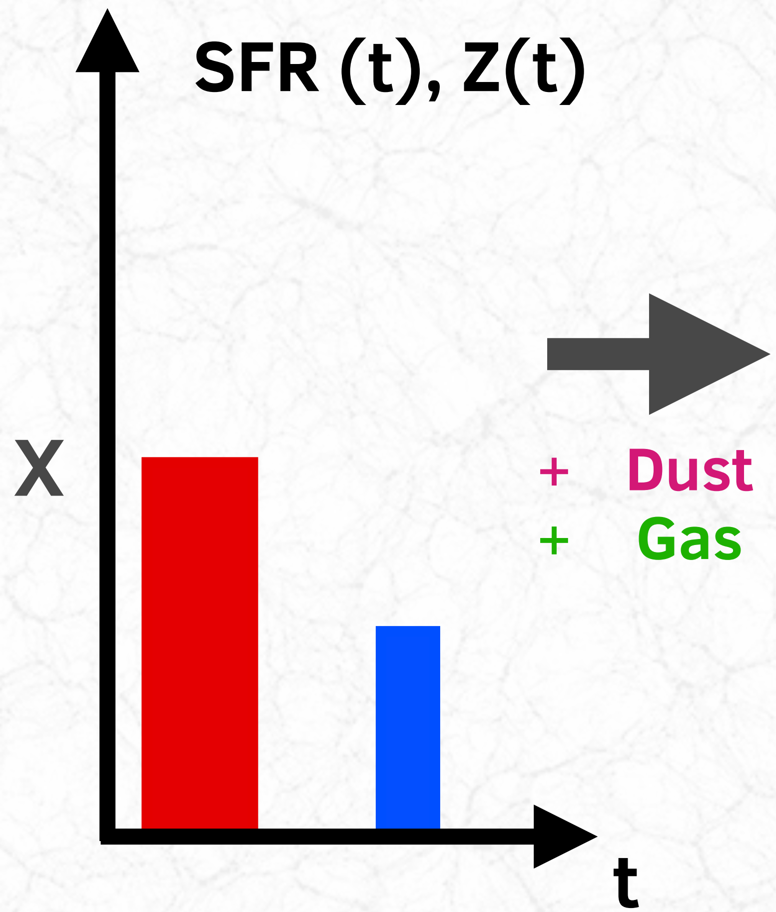
Conroy+13

Physical Properties from Observations

Most galaxy properties should be determined from observed fluxes of galaxies!

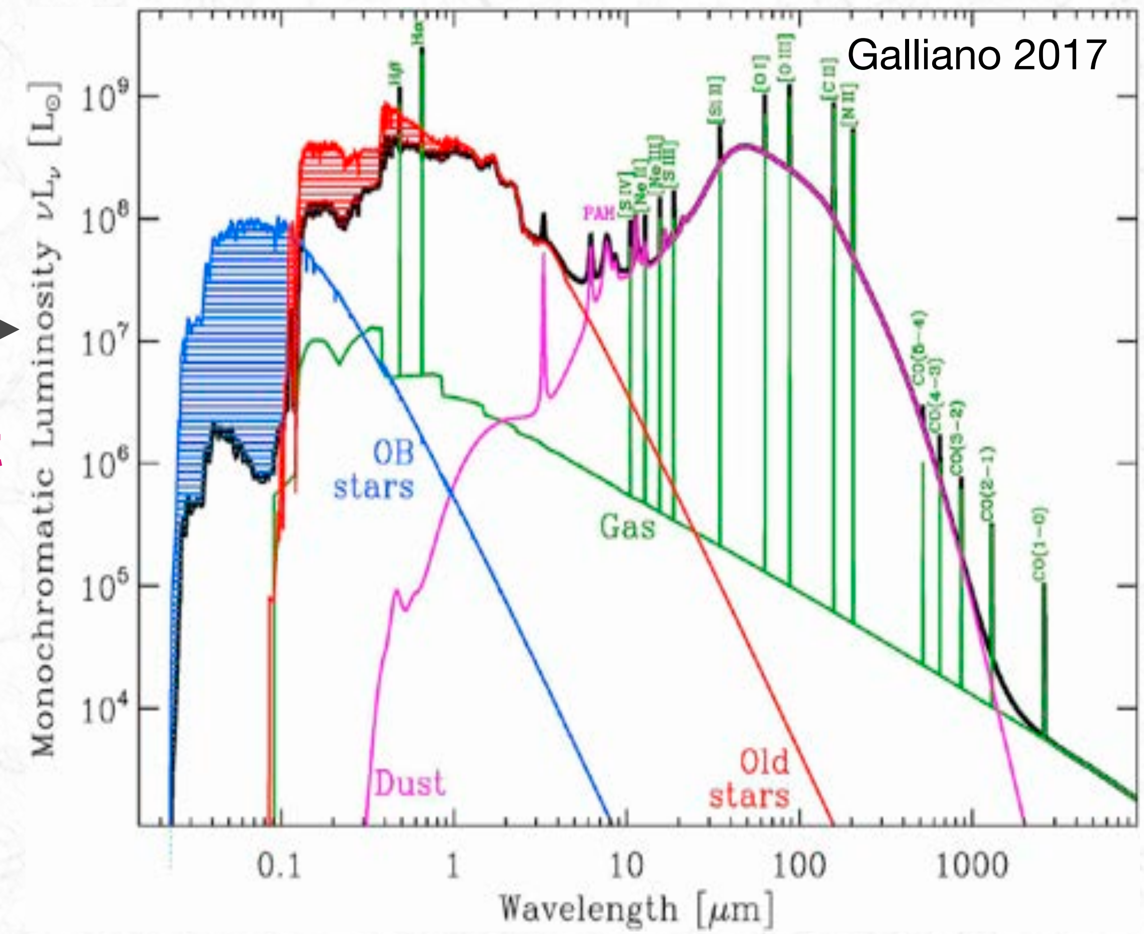


Spectra for simple stellar populations



Star Formation History,

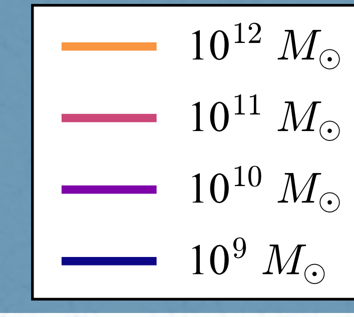
Metallicity History

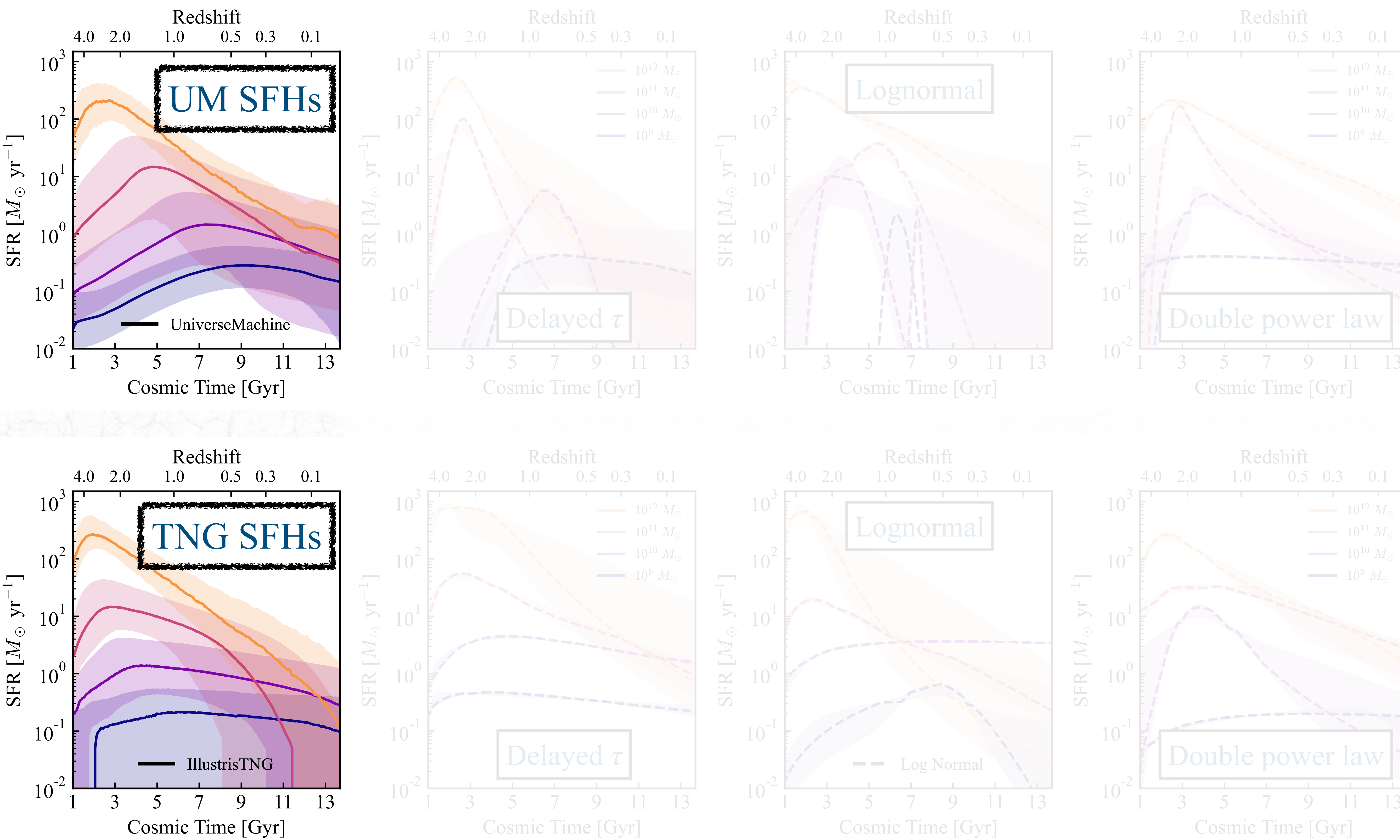


Spectral Energy Distribution (SED) of Galaxies

Inverse Problem!

Recovery of Simulated SFHs

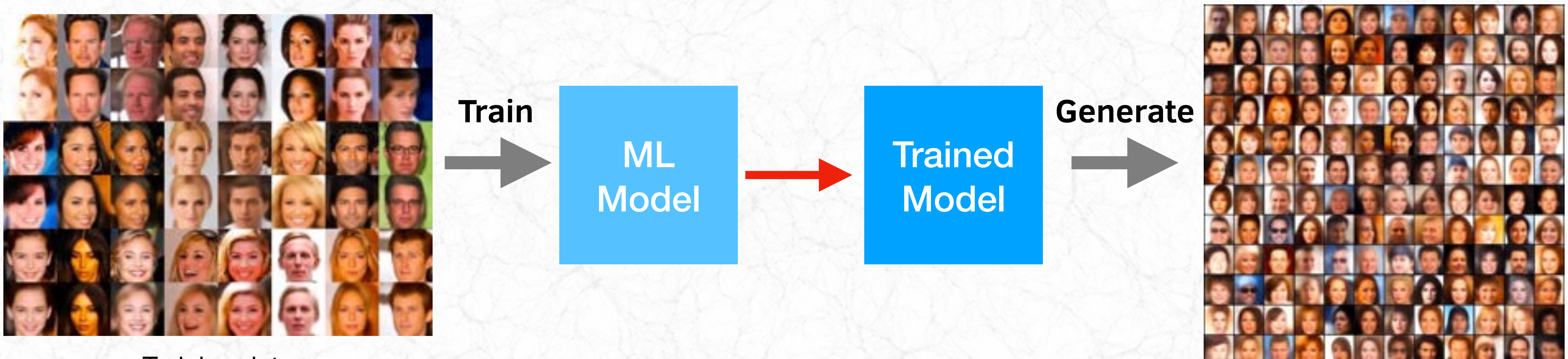




Cooray+ in prep.

Generative Models as a New SFH Model

- Generative models are machine learning (ML) algorithms that can generate new data instances from a learnt data distribution



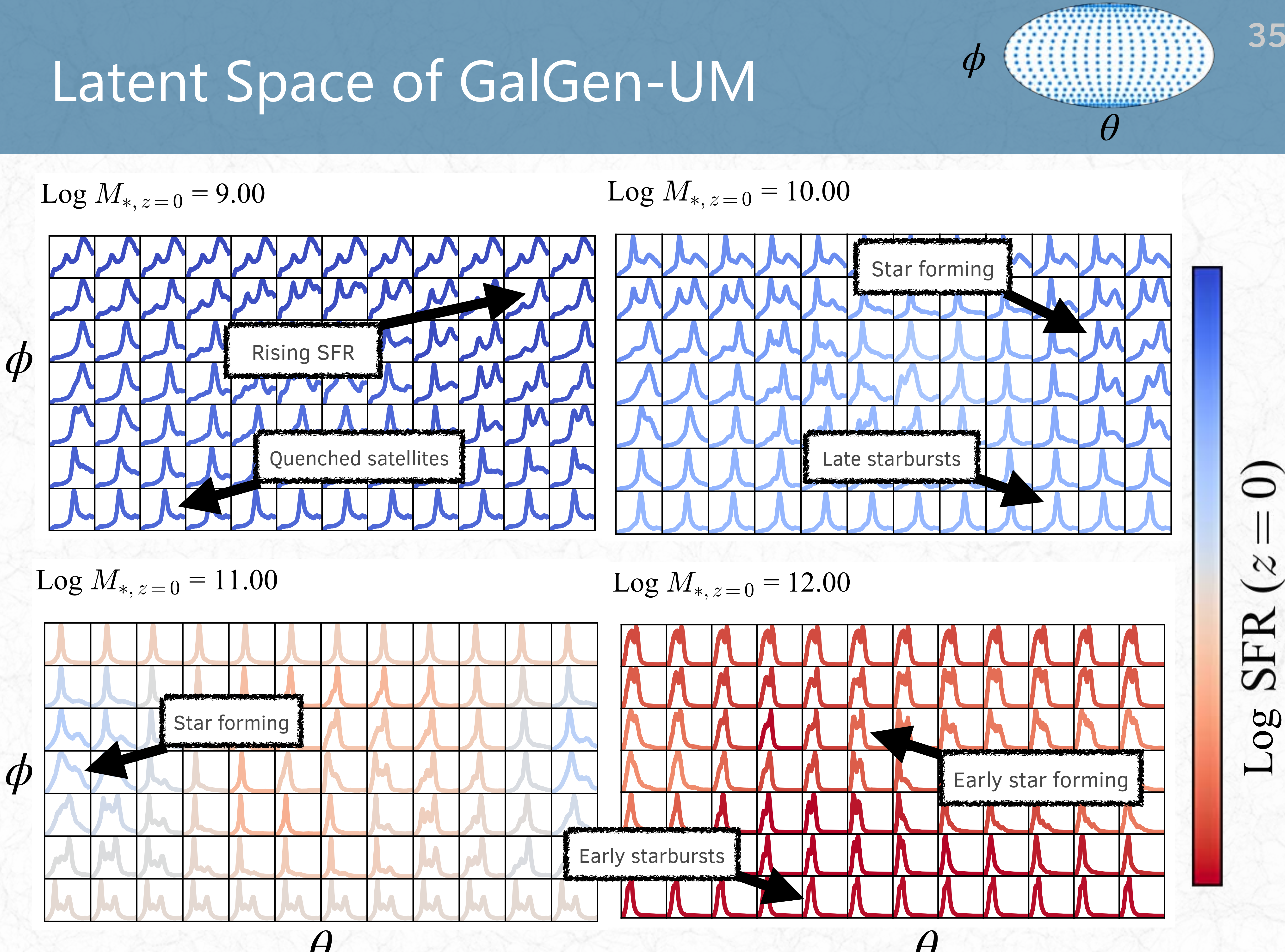
Training data

Generated data samples

- We try to train a model with simulated galaxy SFHs and **the model can learn characteristics and the diversity of simulated SFHs**

Generated SFHs follows the original data (simulated SFH) distributions and thus produce physically-motivated SFHs to fit real galaxies

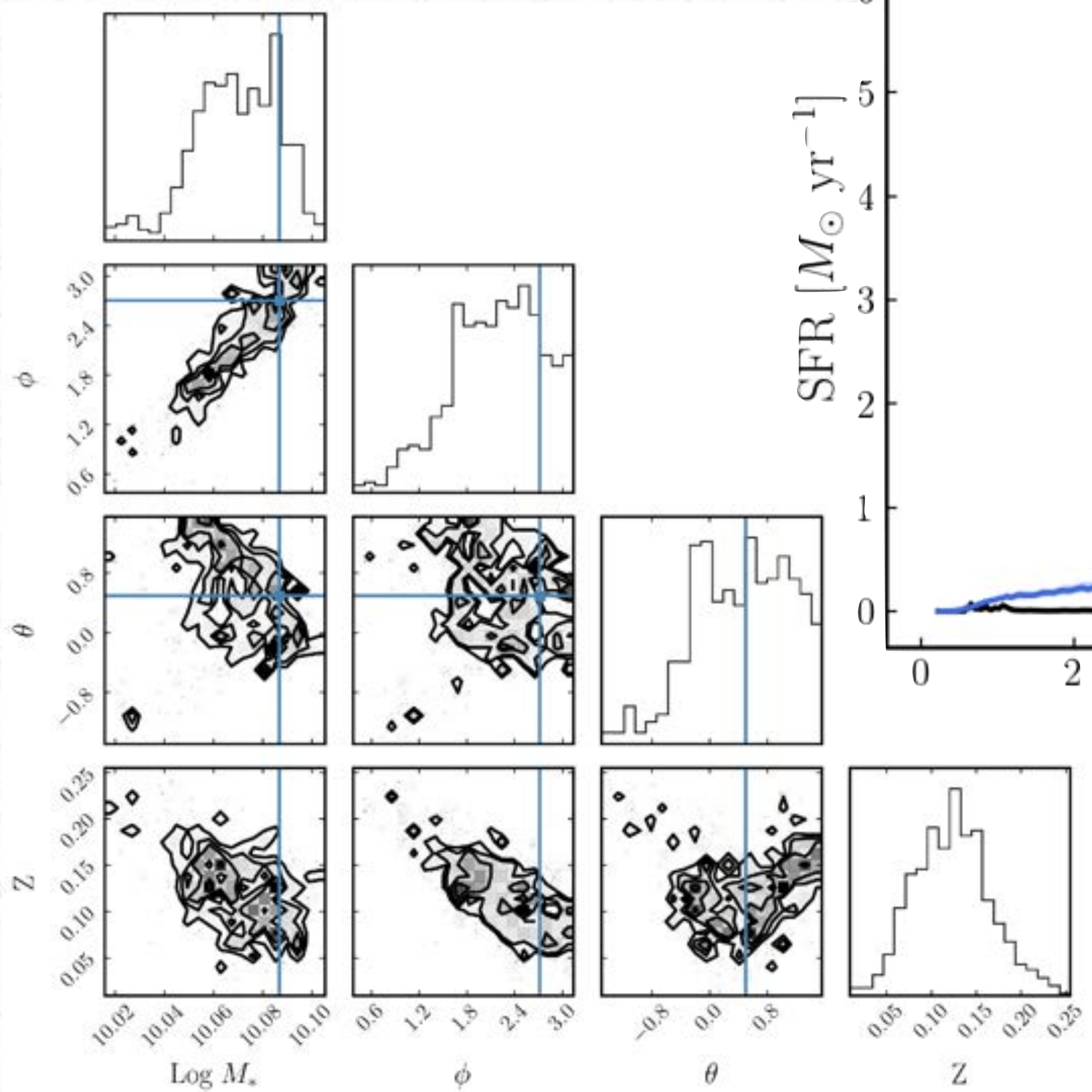
Latent Space of GalGen-UM

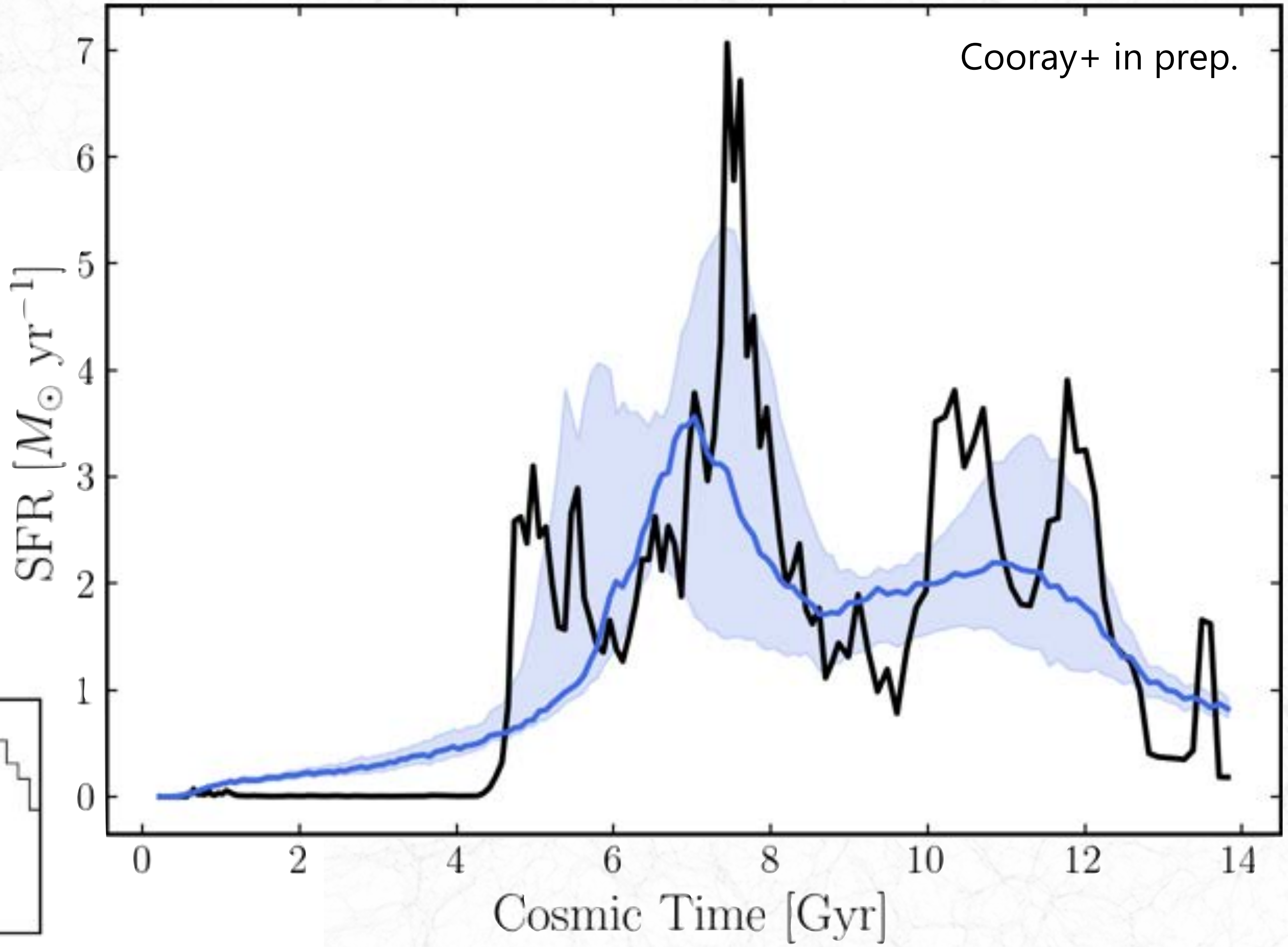


Cooray+ in prep.

Example SED fit of the GalGen model **Preliminary!**

➤Fitting SDSS-like spectra





➤Using the information from simulations for observations as priors

Suchetha Cooray

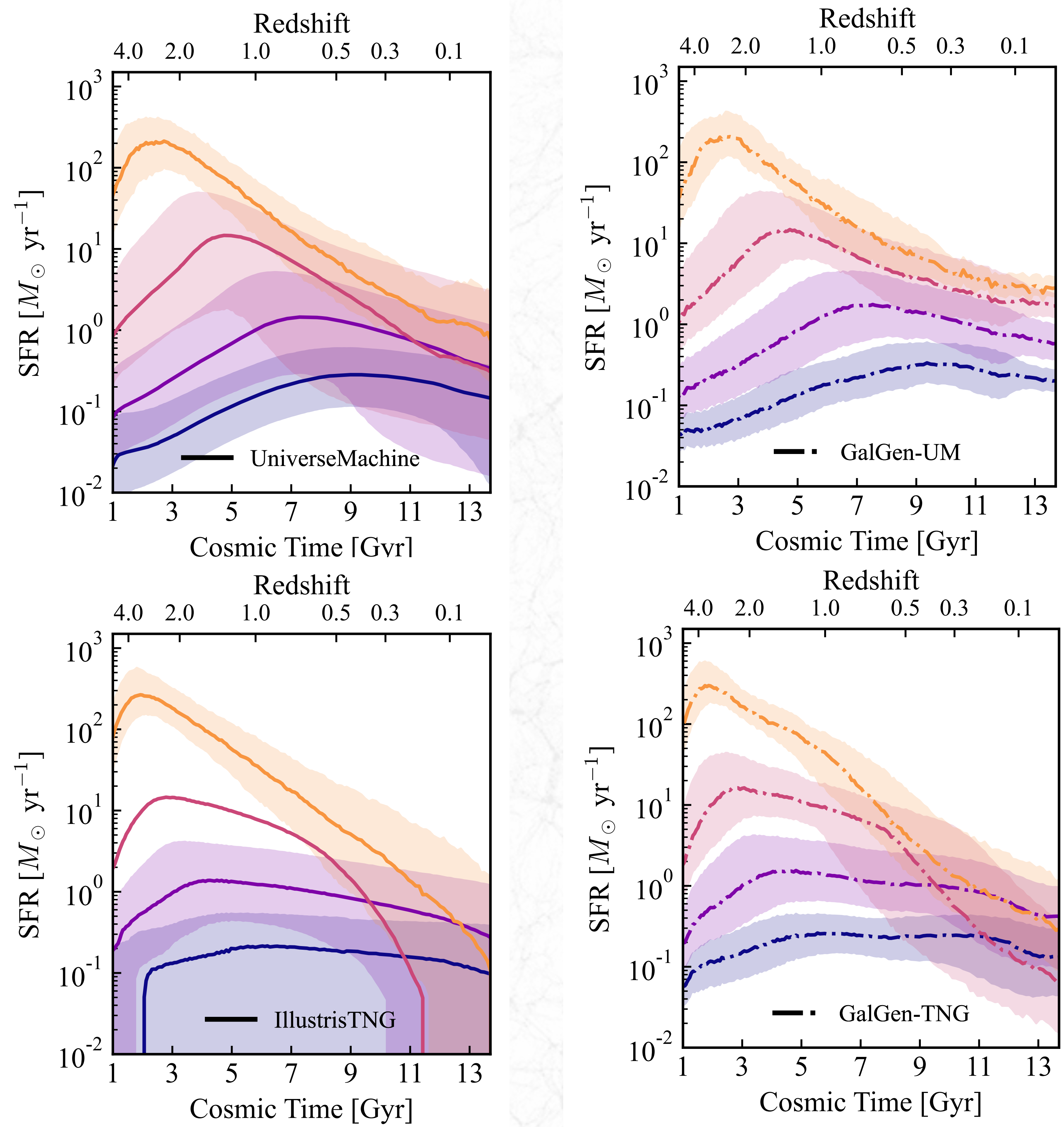
Recovery of Simulated SFHs

UM SFHs

TNG SFHs

GalGen-UM

GalGen-TNG



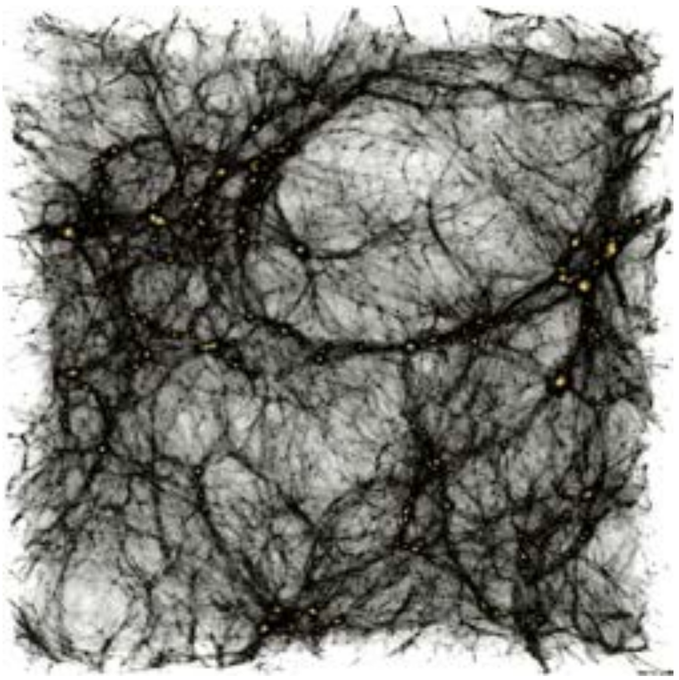
Cooray+ in prep.

Better way to do it?

Forward model galaxy formation up to SEDs and match all SEDs simultaneously!

New Approach to Forward Model SEDs

Cosmology + N-body



Galaxy Formation Model

$$SFR = f(z, M_h, \dot{M}_h, \ldots)$$
$$Z = f(z, M_h, \dot{M}_h, \ldots)$$
$$A_\lambda = f(z, M_h, \dot{M}_h, \ldots)$$

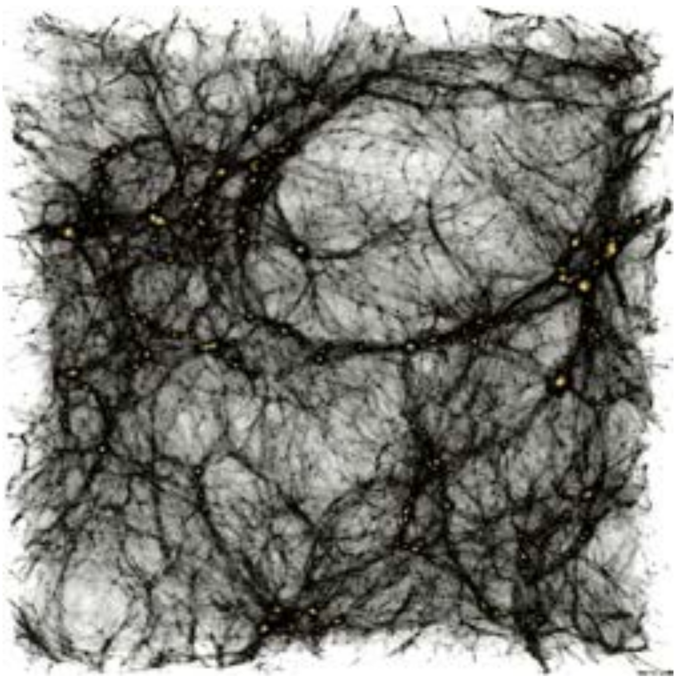
Constrain w/ existing observations

Galaxy distribution with consistent SEDs

Constrain based on observed multi wavelength color distributions, ultraviolet (UV) luminosity functions, auto- and cross-correlation functions

New Approach to Forward Model SEDs

Cosmology + N-body



Galaxy Formation Model

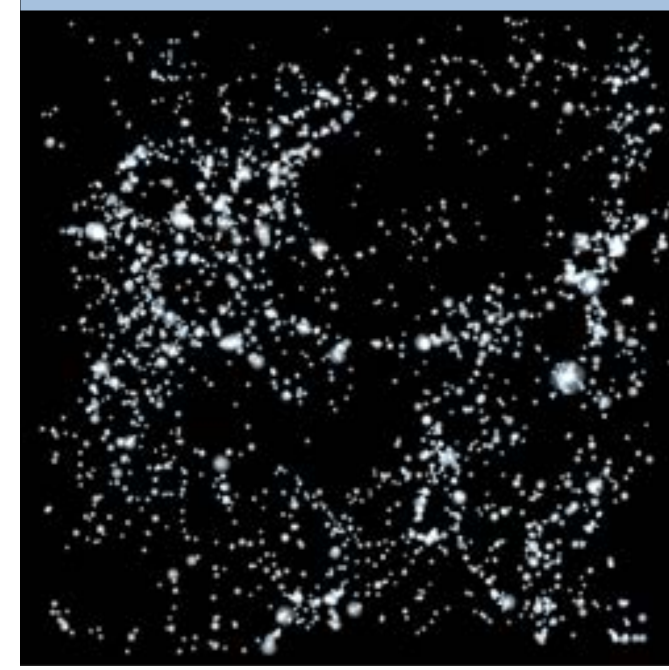
$$SFR = f(z, M_h, \dot{M}_h, \ldots)$$

$$Z = f(z, M_h, \dot{M}_h, \ldots)$$

$$A_\lambda = f(z, M_h, \dot{M}_h, \ldots)$$

Constrain w/ existing observations

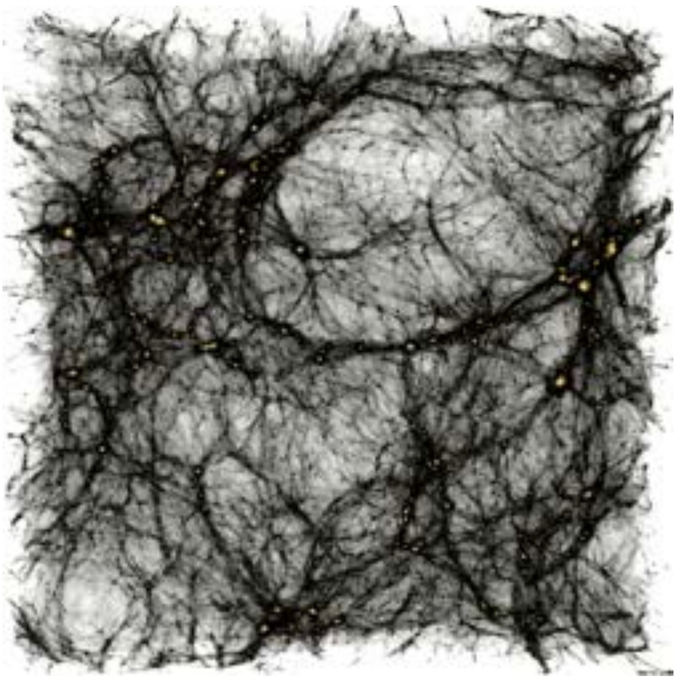
Galaxy distribution with consistent SEDs



Constrain based on observed multi wavelength color distributions, ultraviolet (UV) luminosity functions, auto- and cross-correlation functions

New Approach to Forward Model SEDs

Cosmology + N-body

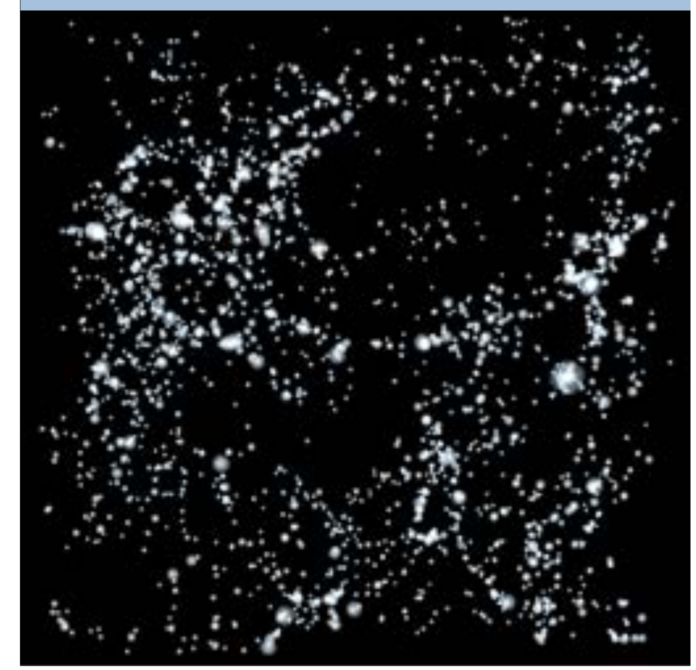


New SED Model

$$SED = f(z, M_h, \dot{M}_h, \ldots)$$

Constrain w/ existing observations

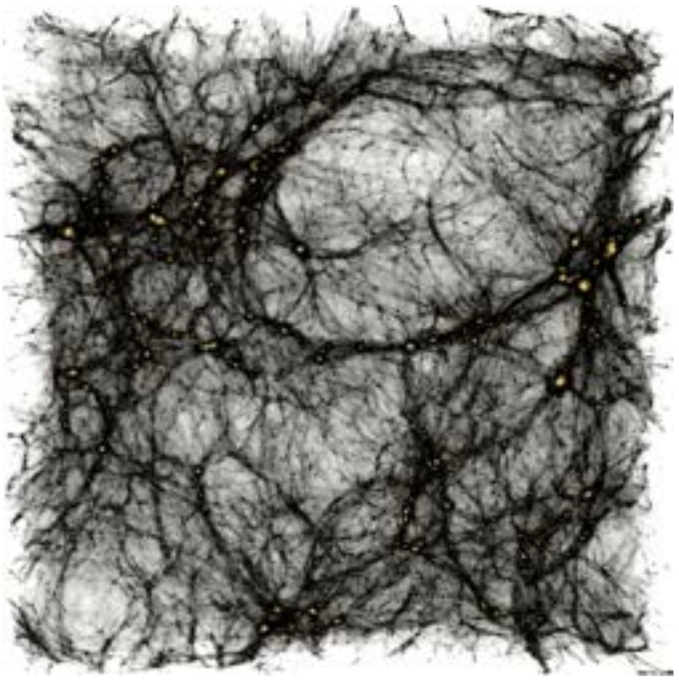
Galaxy distribution with consistent SEDs



Constrain based on observed multi wavelength color distributions, ultraviolet (UV) luminosity functions, auto- and cross-correlation functions

New Approach to Forward Model SEDs

Cosmology + N-body



New SED Model

$$SED = f(z, M_h, \dot{M}_h, \ldots)$$

Constrain w/ existing observations

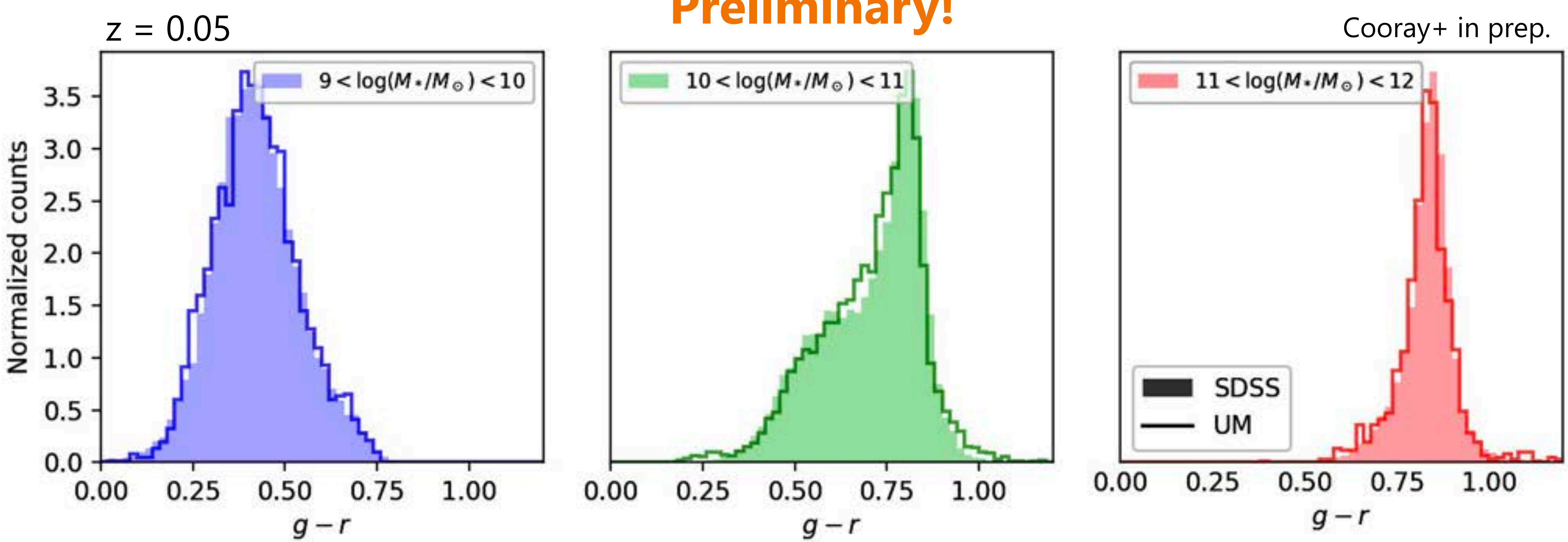
Galaxy distribution with consistent SEDs

Benefits:

- A maximum likelihood solution of the population instead of individual galaxies
- Self-consistent evolution solves much of the degeneracy in SED
- Encodes the environmental information
- Population-level SED fitting code that is both consistent with the population as well as individual!

Matching Observed Color Distributions

Preliminary!

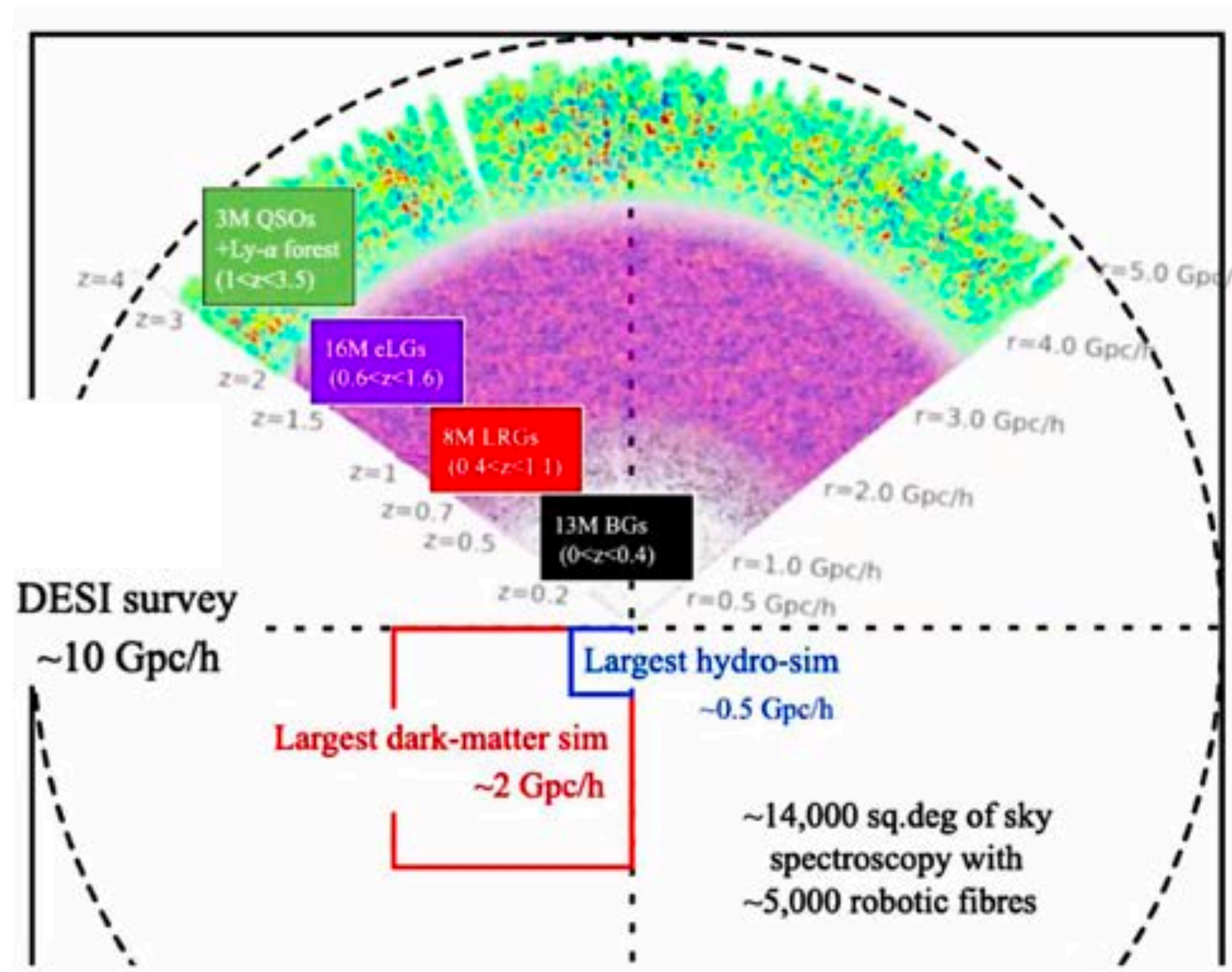


- ➤ Matching SDSS optical photometry, more coming soon
- ➤ Requires careful modeling of selection function, metallicity (w/ scatter), dust attenuation (w/scatter), and photometric errors, etc.
- ➤ Arbitrary bands can be implemented

Stay tuned!

Expected Outcomes

- **A fully physical, self-consistent picture of galaxy stellar masses, star formation histories, metallicity histories, and dust from z = 0 to 15**
- Significantly reduced systematic uncertainties in the extracted galaxy physical properties.
- Mock catalogs for arbitrary current and future surveys that simultaneously match currently observed galaxy number densities, colors, and clustering.
- Better photometric redshifts



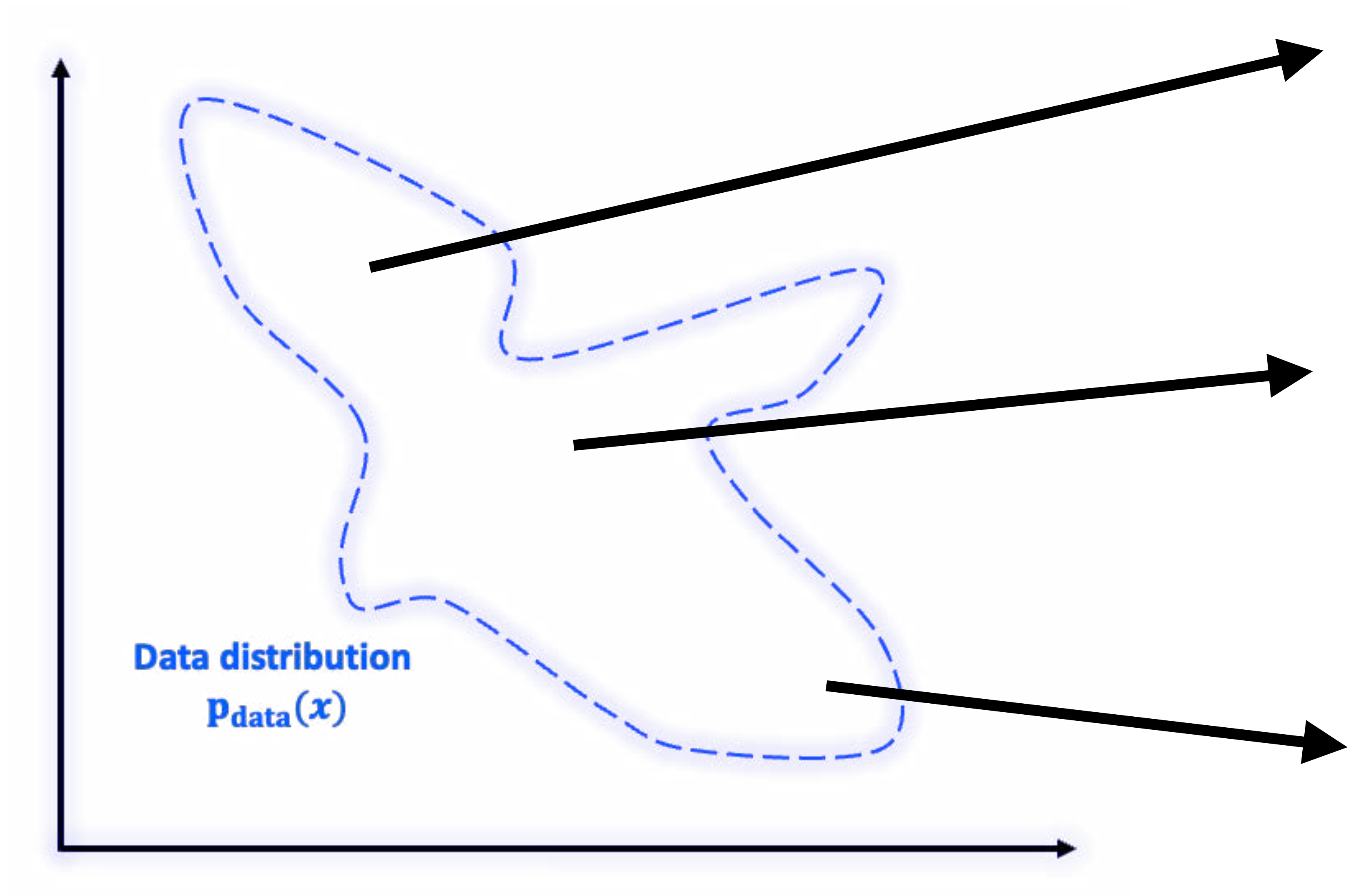
Francisco-Shu Kitaura

Should we stop here?

Forward model up to the image level!

Generative Models for Mock Observations

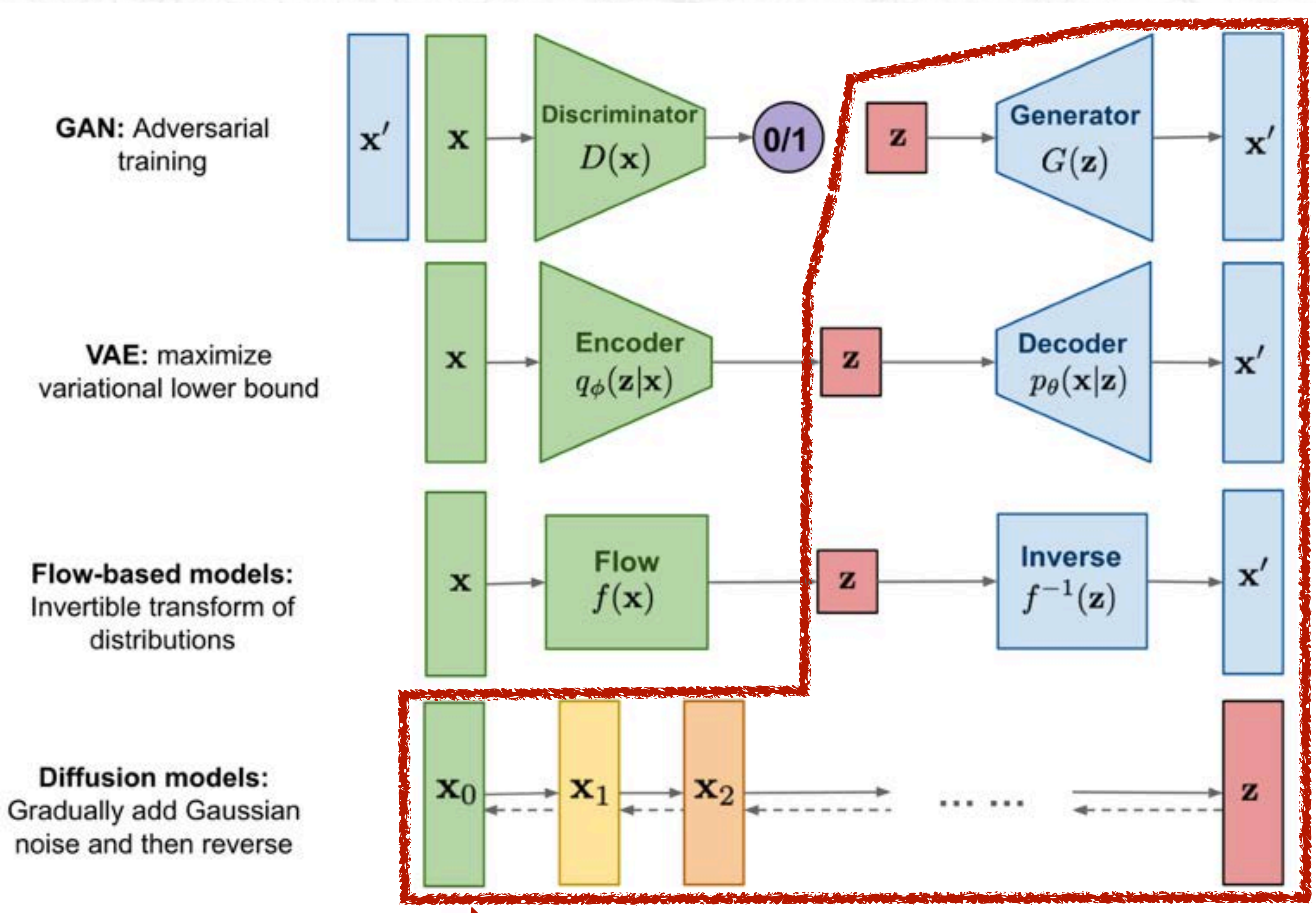
- We should be able to simulate observations!
- Can we simulate the image domain?



High dimensional space

From SEDs to Galaxy Images with Generators!

Generator

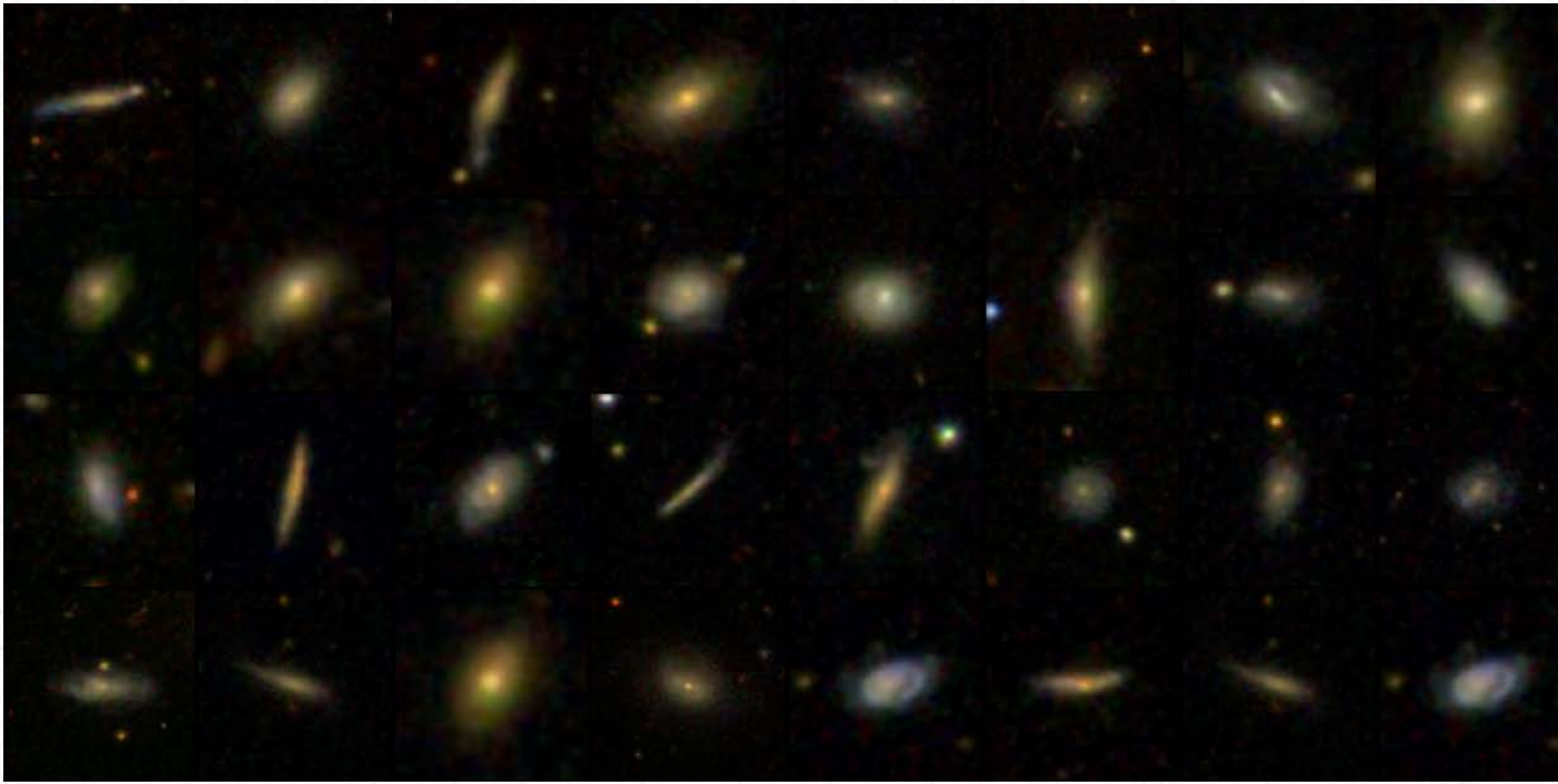


Conditional information from UM

From SEDs to Galaxy Images

Preliminary!

SDSS-like images generated with a denoising diffusion model



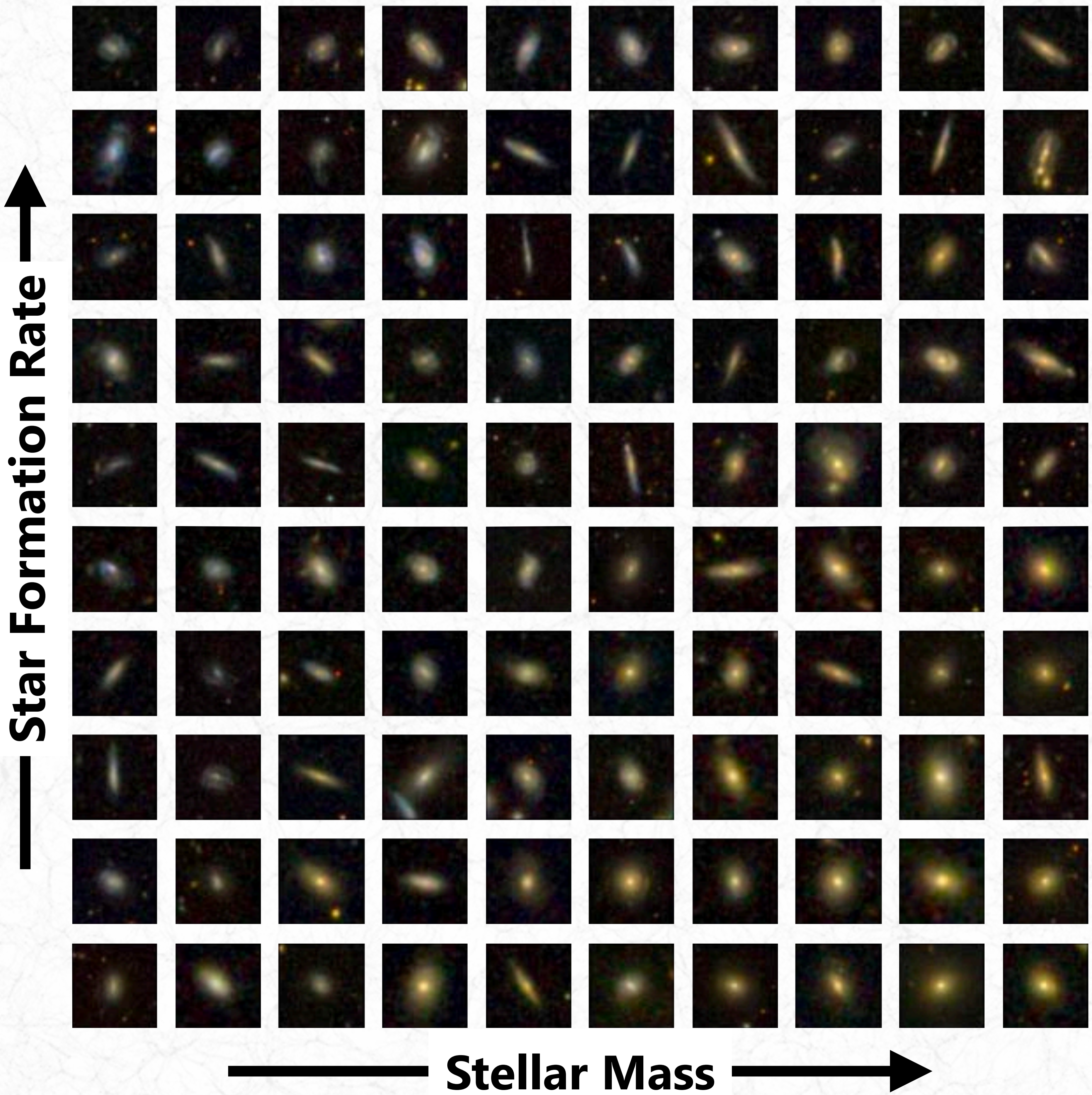
Cooray+ in prep.

Code: https://arxiv.org/abs/2006.11239

Galaxy Properties to Galaxy Images

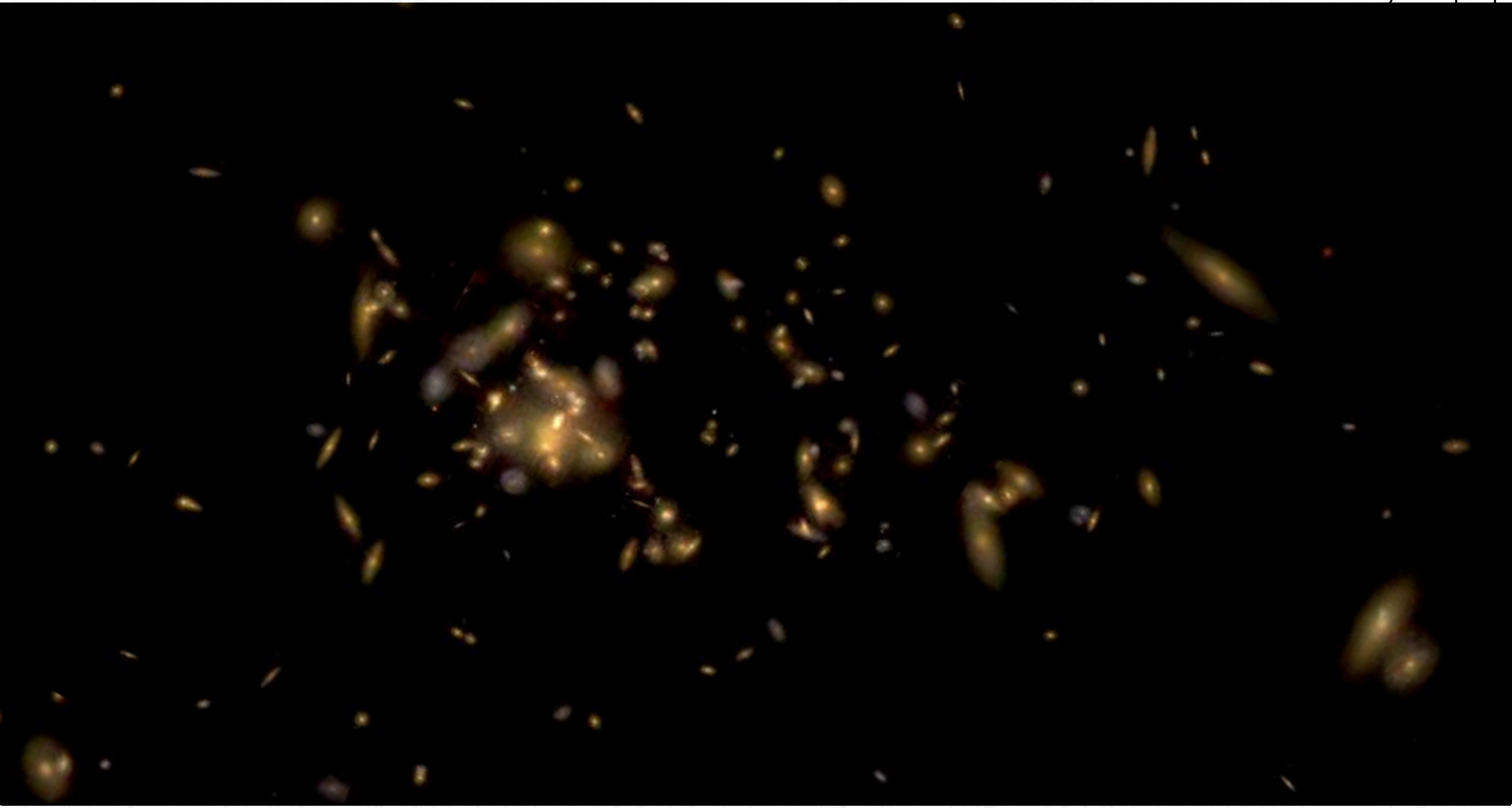
Preliminary!

SDSS-like images generated with a conditional diffusion model



Mock universes as telescope-level images **Preliminary!**

Cooray+ in prep.

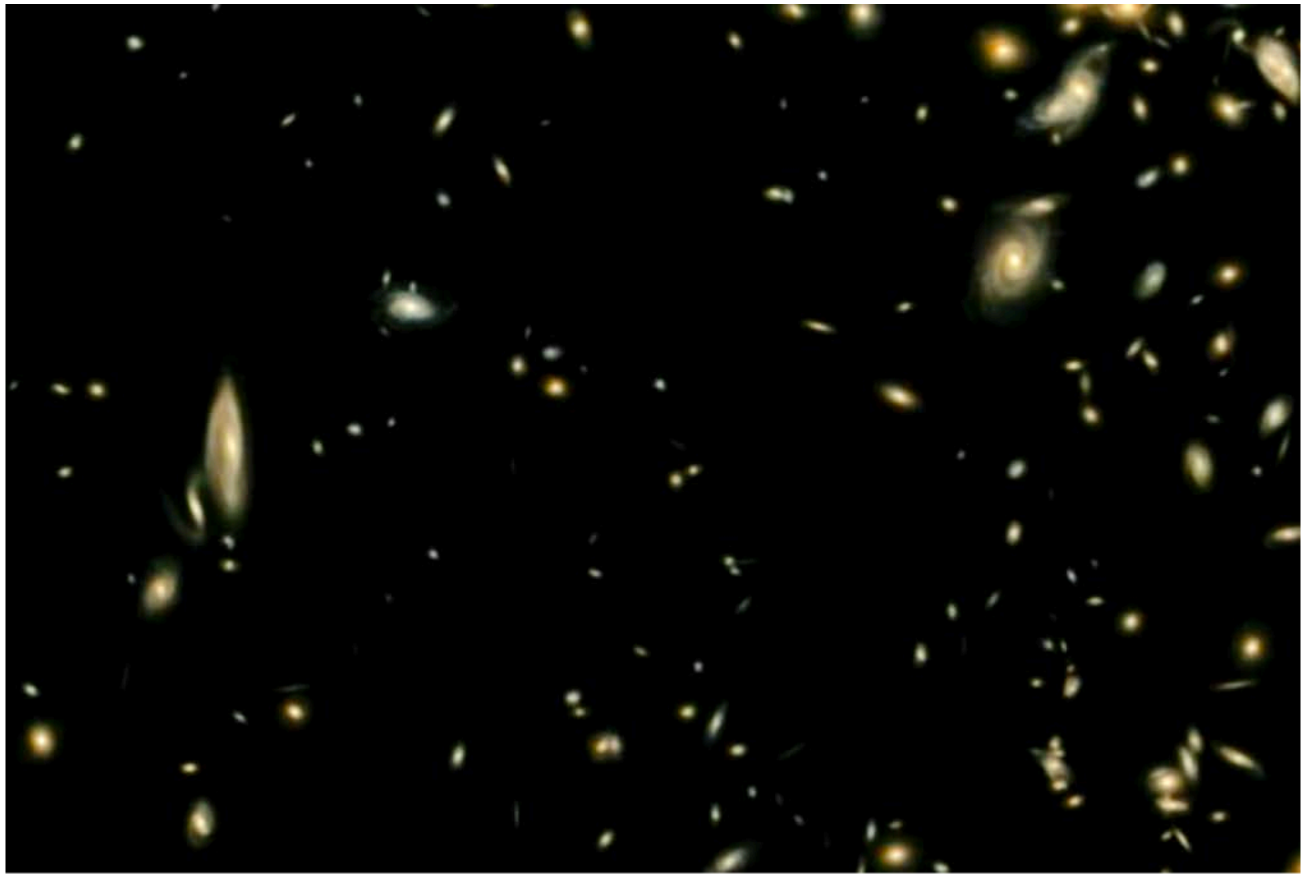


SDSS-like survey field. The colors and the color-dependent spatial information come from the empirical forward model. A conditional diffusion model trained on SDSS images and photometry generated individual galaxy image patches. Relative sizes come from the size-mass relation (Mawla_2019).

Mock Universes with Empirical Modeling

Parameters of cosmology and galaxy formation

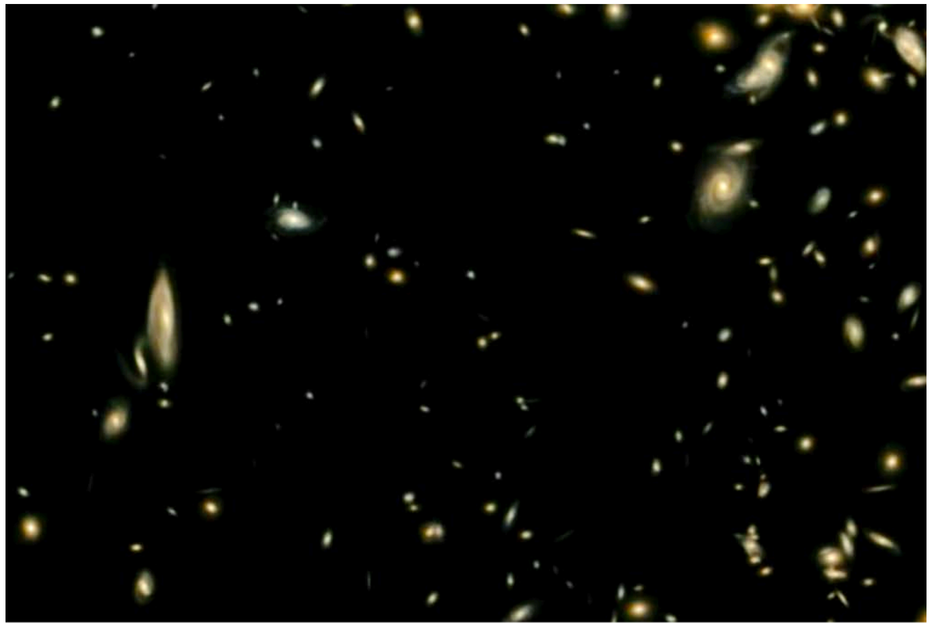
Forward modeling



Mock Universes with Empirical Modeling

Forward modeling

Parameters of cosmology and galaxy formation



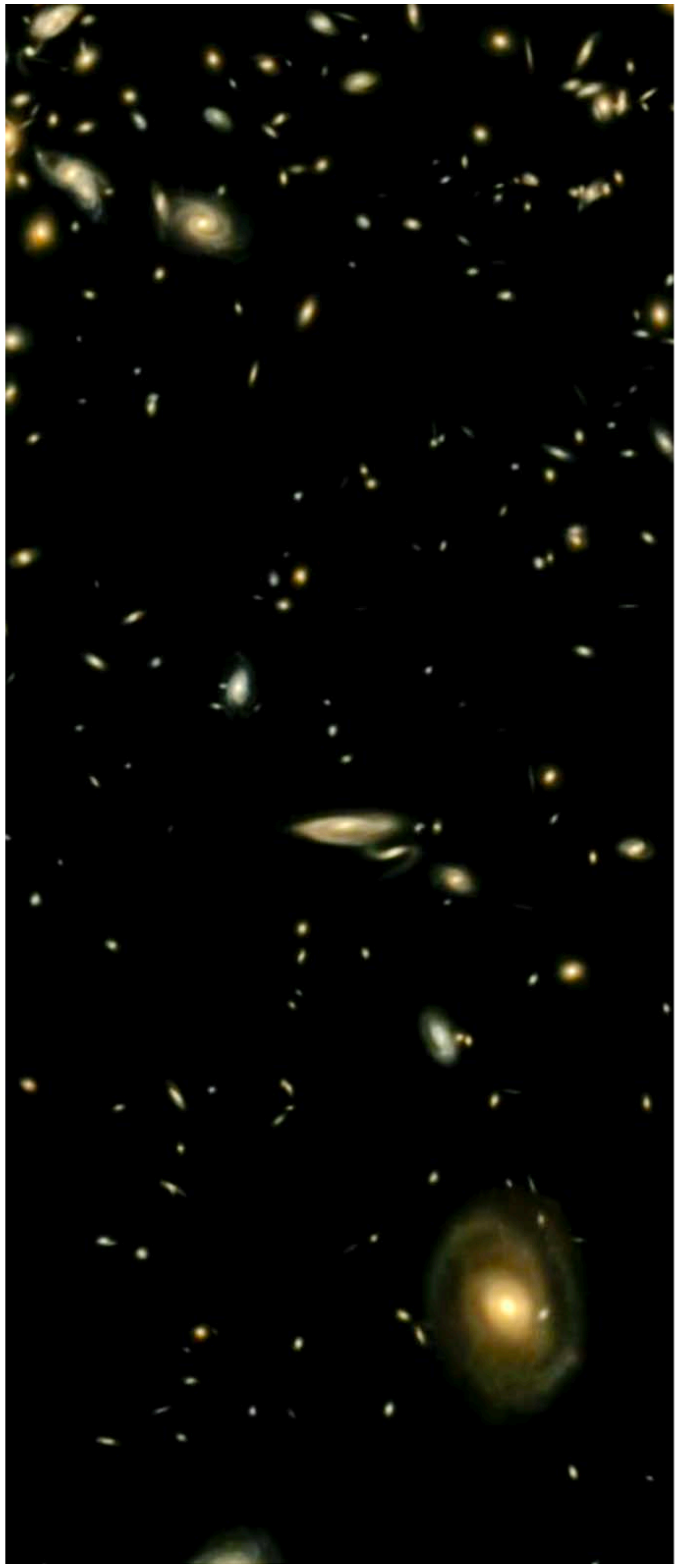
⋮

Expected Outcomes

- **Field-level (pixel-level) inference, maximizing the information from galaxy surveys**
- Data-driven model of galaxy morphologies with data from multiple telescopes
- Constraints on galaxy size evolution
- Observationally indistinguishable mock images

The way forward for galaxy formation

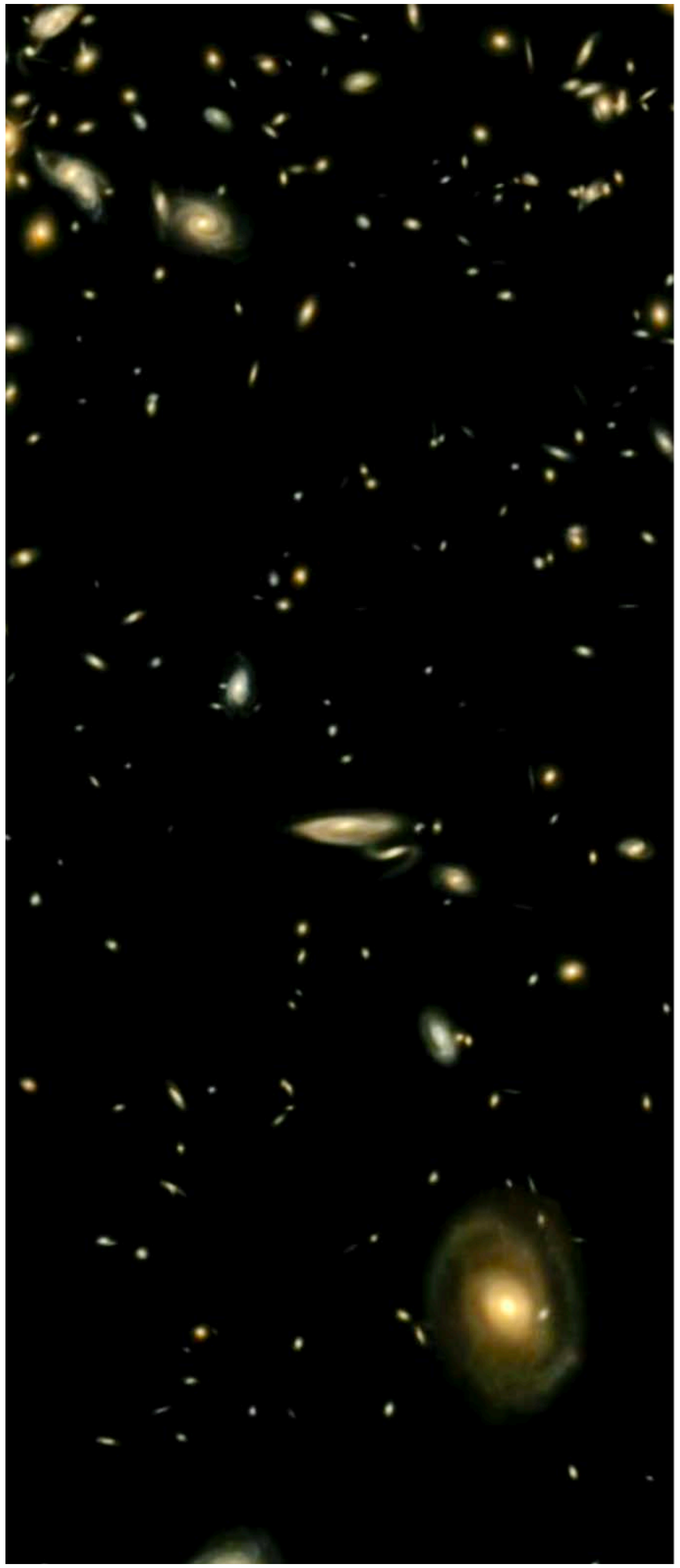
- ➤ Increase model complexity
  - ➡ Add other physical processes: blackholes (Zhang et al. 2023), gas (Hong et al. 2023), Metals (Ishigaki et al. in prep.)
- ➤ Distribution of light with distribution of DM
  - ➡ Model light and mass profiles with DM profiles
- ➤ Latent representations
  - ➡ Fundamental parameters
- ➤ Causality
  - ➡ Go beyond correlations
- ➤ Bayesian optimization approach for designing new surveys
  - ➡ Maximize science outcome while balancing costs



Mock universe with UM

The way forward for cosmology

- ➤ Mock catalogs
  - ➡ End-to-end tests of pipelines
- ➤ Cosmology + galaxy formation joint analysis
  - ➡ Benefits more for galaxy formation?
- ➤ Field-level inference for initial condition reconstructions
  - ➡ Differentiable codes



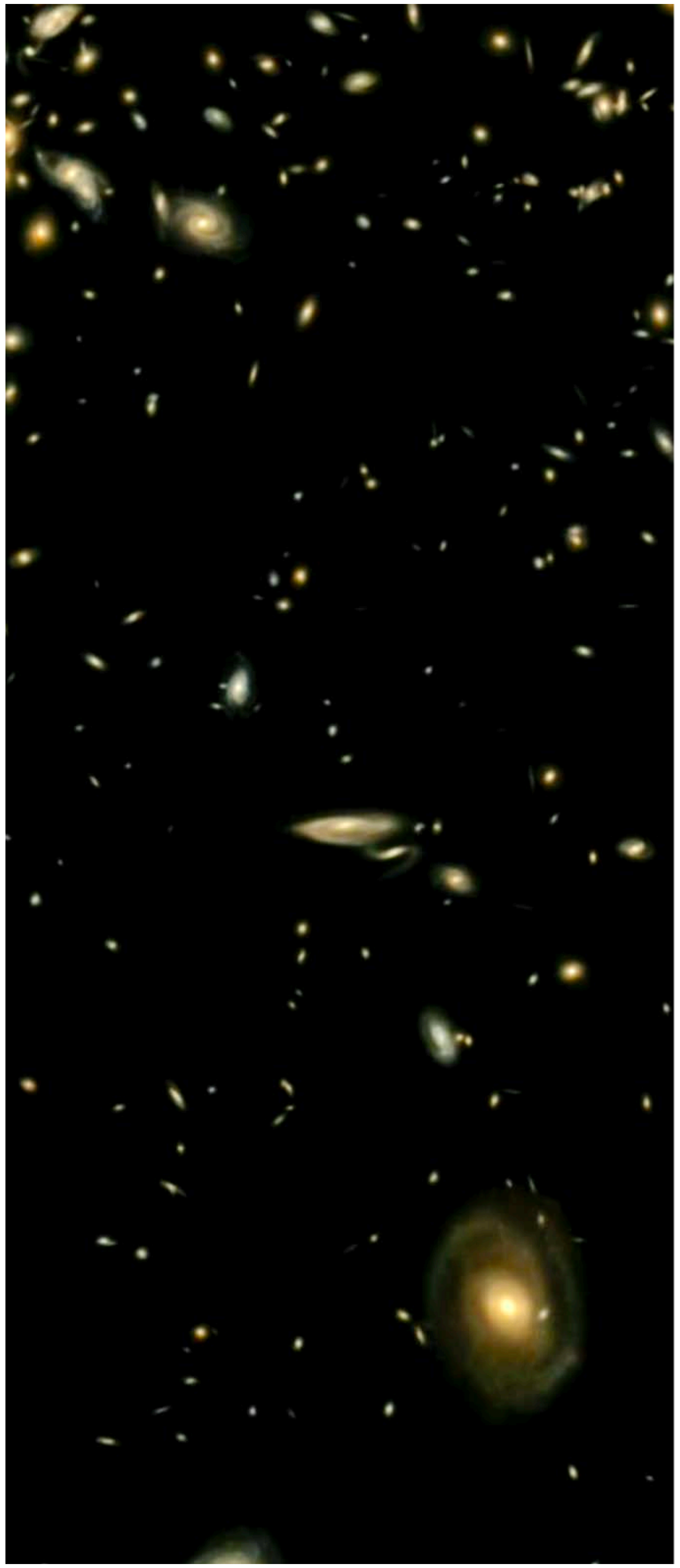
Mock universe with UM

Summary

- I am building a field-level (pixel-level) inference machine for galaxy formation

- Will combine all observations to simultaneously derive constraints on SFR, SM, dust, metallicity evolution and its connection to halos from z=0-15

- Some topics I am interested in include;
  - Learning physically meaningful representations and dimensionality reduction
  - Signal processing
  - Faraday tomography
  - Anything cosmology, galaxy formation, ML, etc.



Mock universe with UM

Thank you!

And to Peter Behroozi, Risa Wechsler, Andrew Hearin, Tsutomu Takeuchi, Daichi Kashino, Kartheik Iyer, Sultan Hassan, and many more!

Web: https://suchethac.github.io/