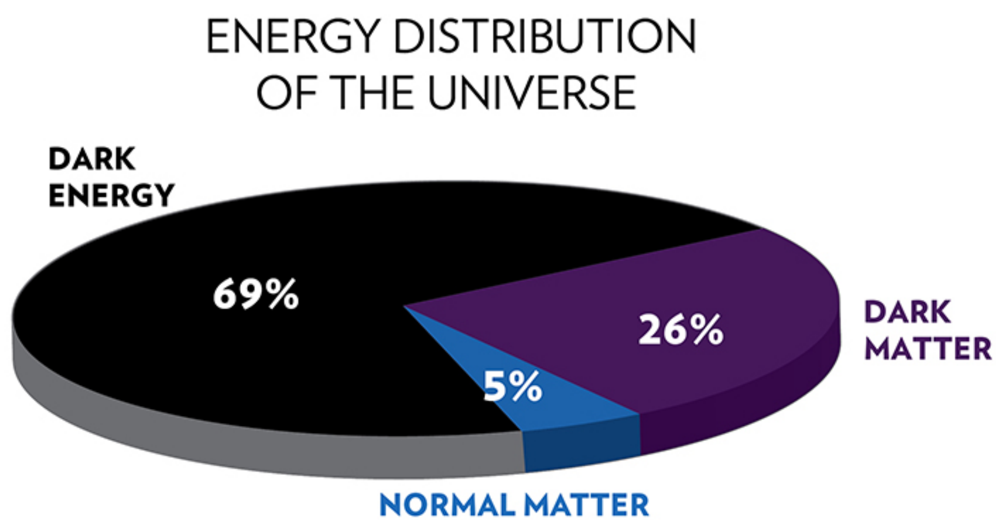
Learning the Initial Conditions of the Universe

A Simons Collaboration

Carnegie Mellon University,
Columbia University,
Center for Computational Astrophysics,
Harvard University,
Institut d'Astrophysique de Paris,
LBL/Berkeley,
Max-Planck Institute,
Princeton University,
Université de Montréal,
Stockholm University

MEASURING THE UNIVERSE: THE ΛCDM MODEL

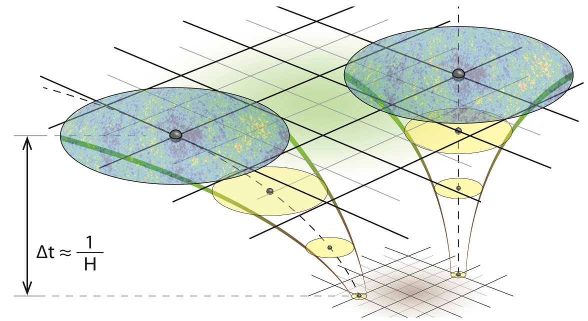
- Current observations remarkably well fit by a model with:
  - Cosmological constant ("Dark Energy") which drives current acceleration
  - Cold Dark Matter plus some baryonic matter (gas and stars)
  - Initial conditions from an early period of "inflation"

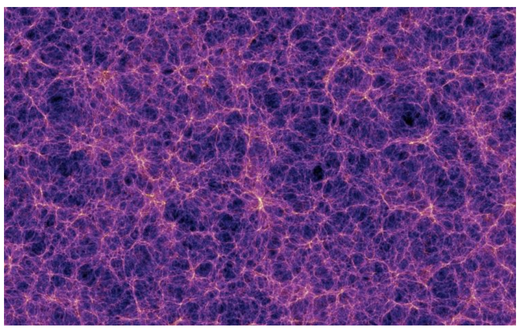


BUT WE DO NOT UNDERSTAND THE MODEL COMPONENTS

Inflation (initial conditions):

The nature, properties, or origin of the field causing inflation are completely unknown



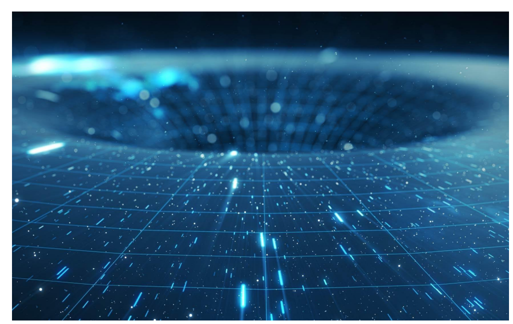


Dark Matter:

What is known:
only that it exists and gravitates

Dark Energy:

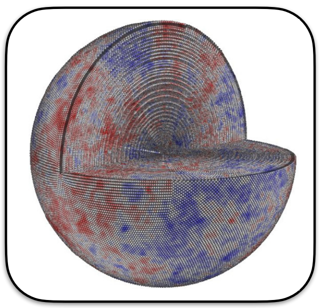
Is not even remotely understood.



Collaboration Goals

Use cosmological observations to infer:

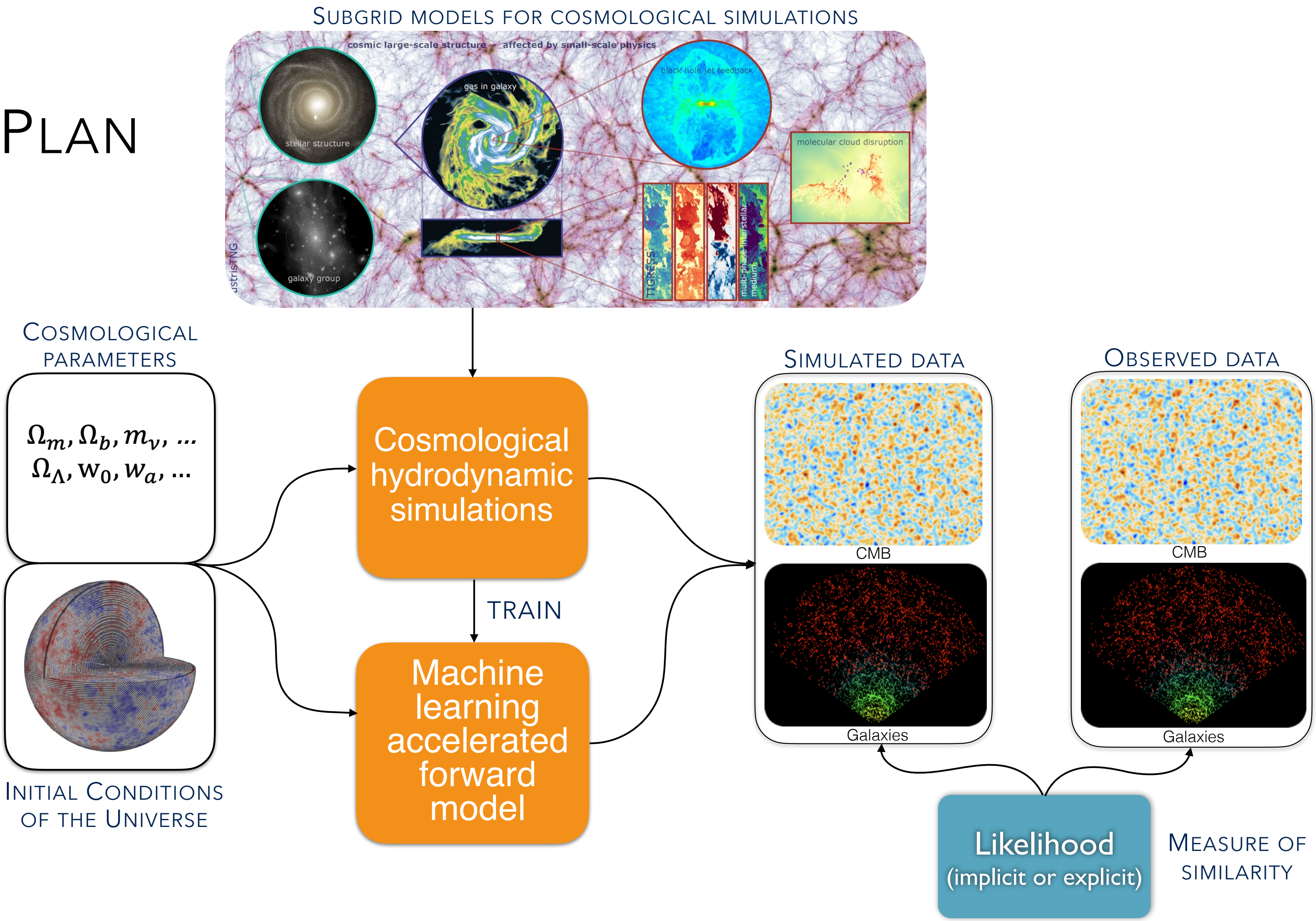
Initial conditions of the universe:
Phases and amplitudes



Cosmological parameters

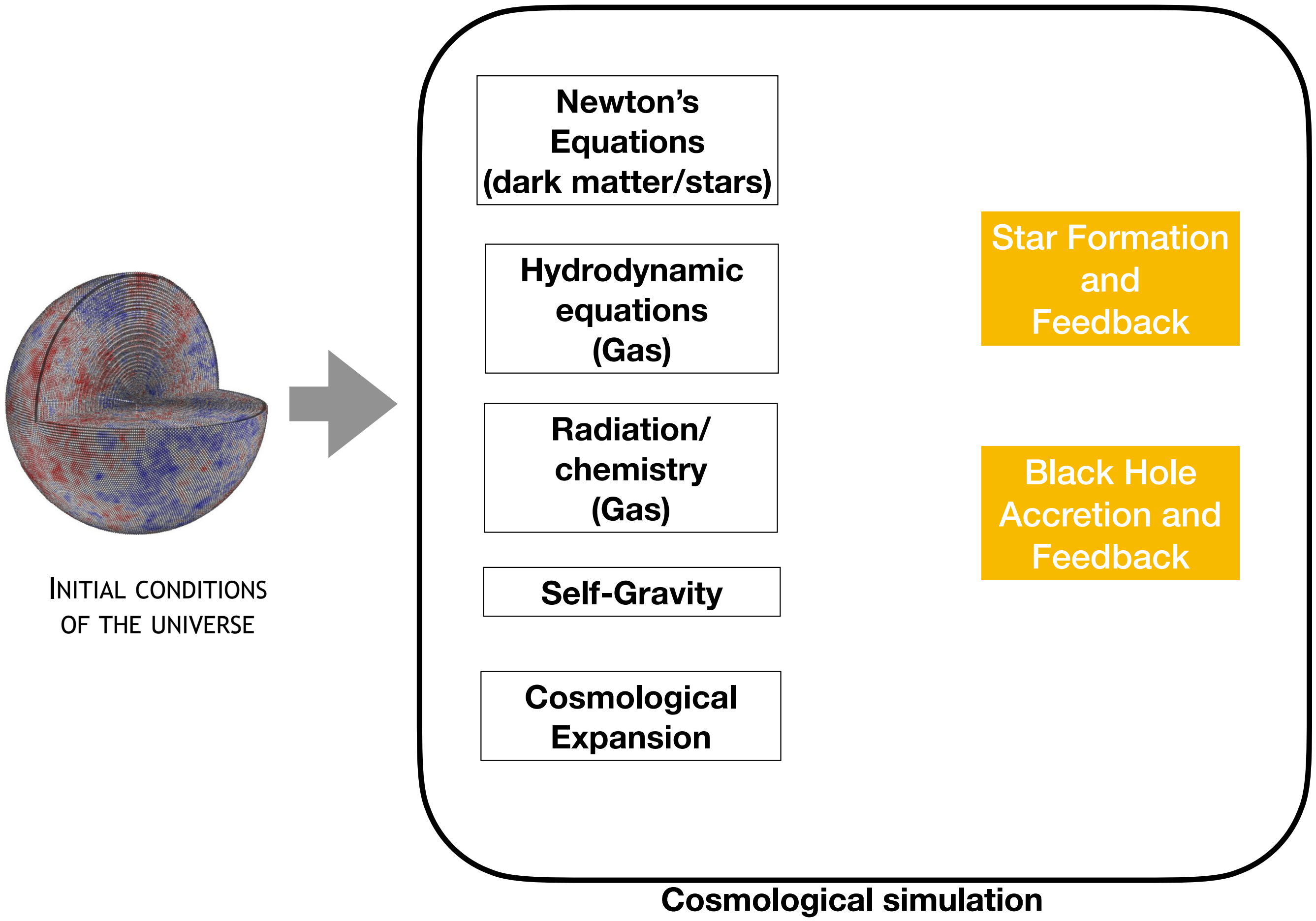
$\Omega_m, \Omega_b, m_\nu, \ldots$
$\Omega_\Lambda, w_0, w_a, \ldots$

The Plan

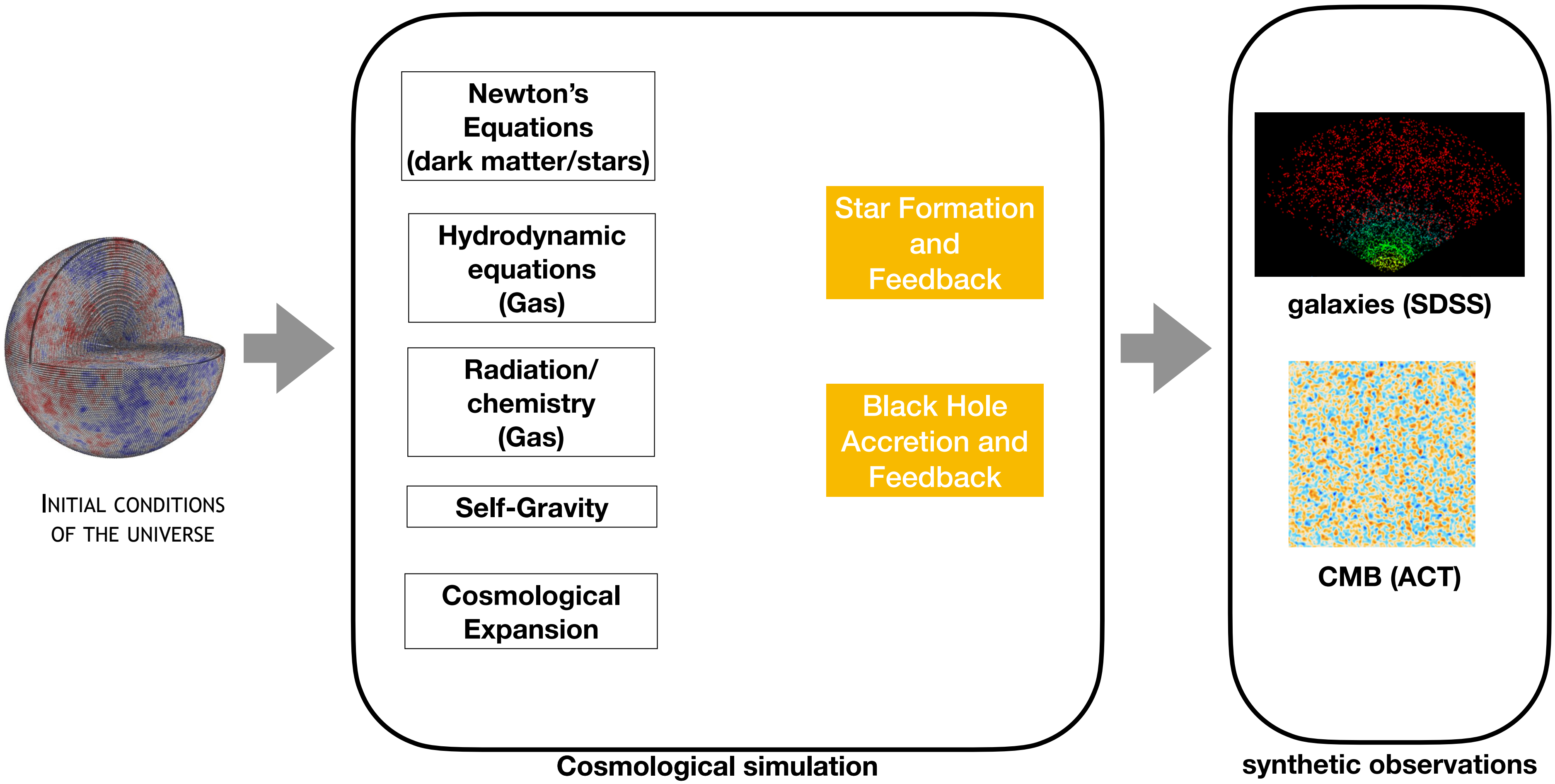


Subgrid models for cosmological simulations
cosmic large-scale structure – affected by small-scale physics
stellar structure
gas in galaxy
black hole jet feedback
molecular cloud disruption
galaxy group
Cosmological parameters
$\Omega_m, \Omega_b, m_\nu, \ldots$
$\Omega_\Lambda, w_0, w_a, \ldots$
Initial Conditions of the Universe
Cosmological hydrodynamic simulations
TRAIN
Machine learning accelerated forward model
Simulated data
CMB
Galaxies
Observed data
CMB
Galaxies
Likelihood
(implicit or explicit)
Measure of similarity

1. (A) Develop New Stellar & BH Feedback Models

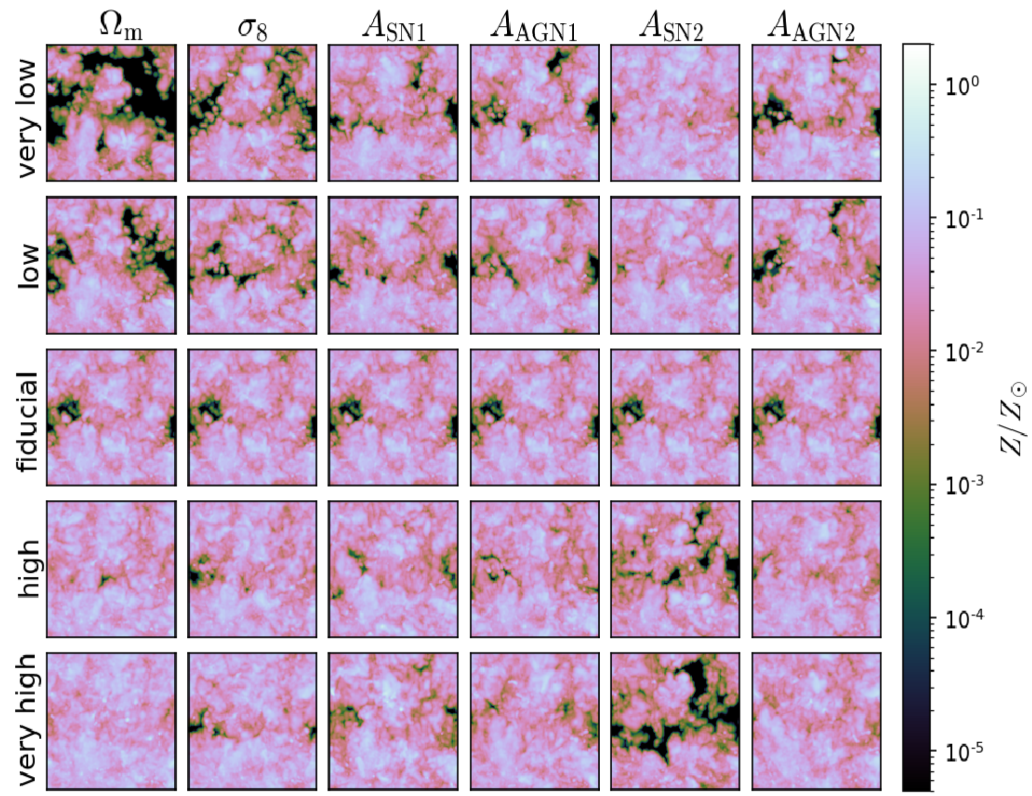


1. (B) Forward Model to Observational Space



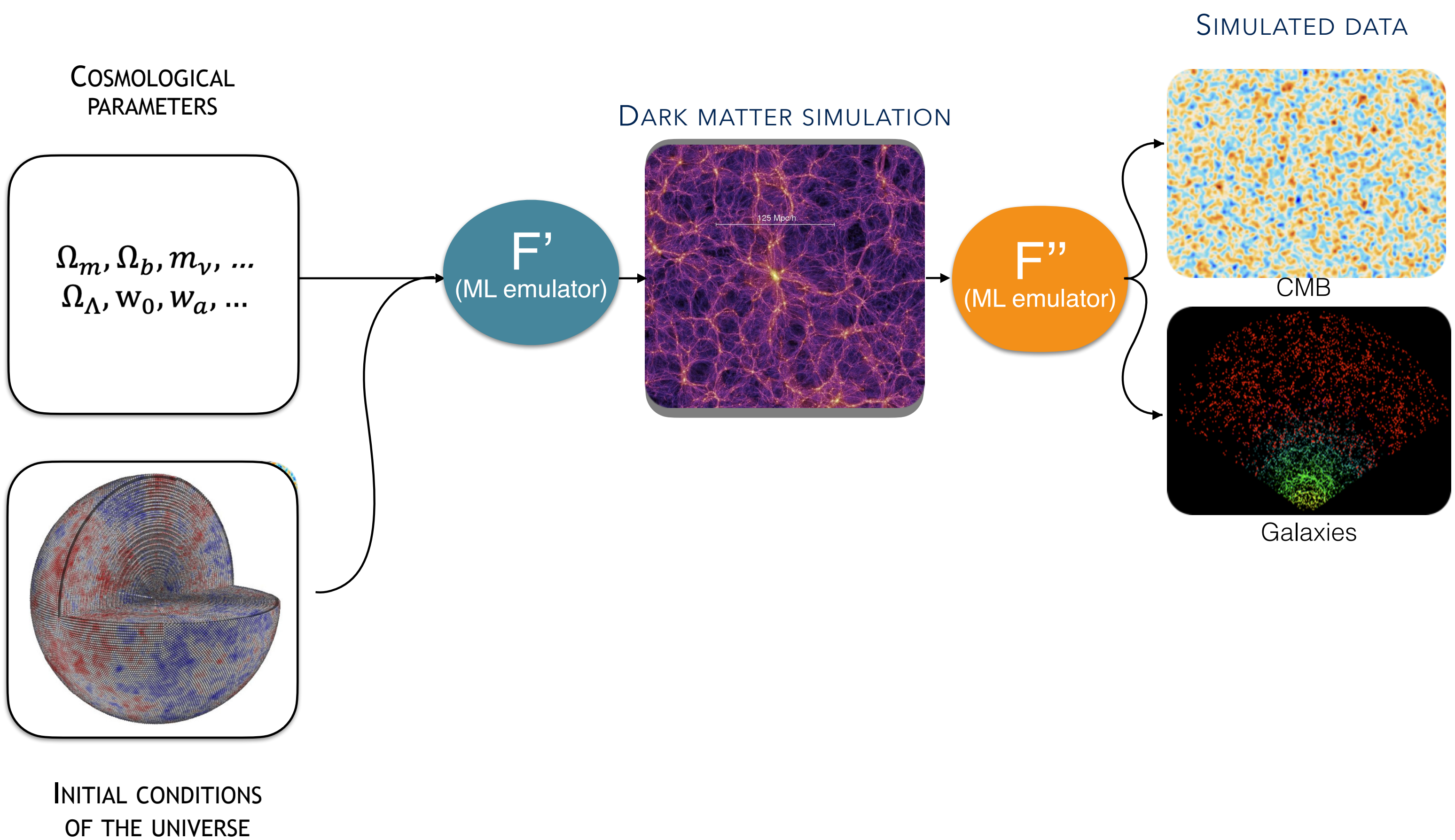
2. (A) GENERATE LARGE SUITES OF TRAINING SIMULATIONS

- Star Formation and Black Hole models will be implemented in the AREPO cosmological hydrodynamics code.
  - Well-tested, widely used, scales well
- Large suite of cosmological simulations required for emulator training
  - Vary cosmological parameters
  - Vary initial conditions
  - Vary astrophysics parameters
- Training sets:
  - "small boxes" (~25-50 Mpc) for galaxy properties
  - "zoom simulations" (small regions in large boxes) for rare clusters of galaxies



Navarro-Villaescusa et al. (2021)

2. (B) Accelerate forward modeling with machine learning



3. INFERENCE WITH FULL PHYSICAL FORWARD MODELS

- Infer the posterior pdf of cosmological parameters and the initial curvature perturbations based on galaxy surveys and CMB maps enabled by high-fidelity, fast forward modeling of cosmic structure, and galaxy formation and evolution

- Principled approach: compute the posterior pdf $p(\theta \mid d)$ using *two techniques:*

  A. *An explicit likelihood-based (**EL**) inference approach: BORG. Will serve as validation benchmark.*

  B. *Multiple simulation-based inference (**SBI**) approaches: more flexible, but more heavily reliant on machine learning.*

Our approach

1. Develop next-generation galaxy formation simulations

- Calibrated sub-grid model for star formation and galactic winds from resolved (small-scale) simulations
- Comprehensive set of sub-grid models for black hole accretion/feedback from resolved simulations
- Carry out a large suite of cosmological simulations, varying parameters
- Create synthetic observations from this suite

2. Develop Machine-learning techniques to accelerate forward models

- Use large suite of N-body simulations to train neural network to predict dark matter distribution
- Use cosmological simulation suite to train machine to predict galaxy properties from dark matter
- Use active learning to minimize number of galaxy simulations required to train

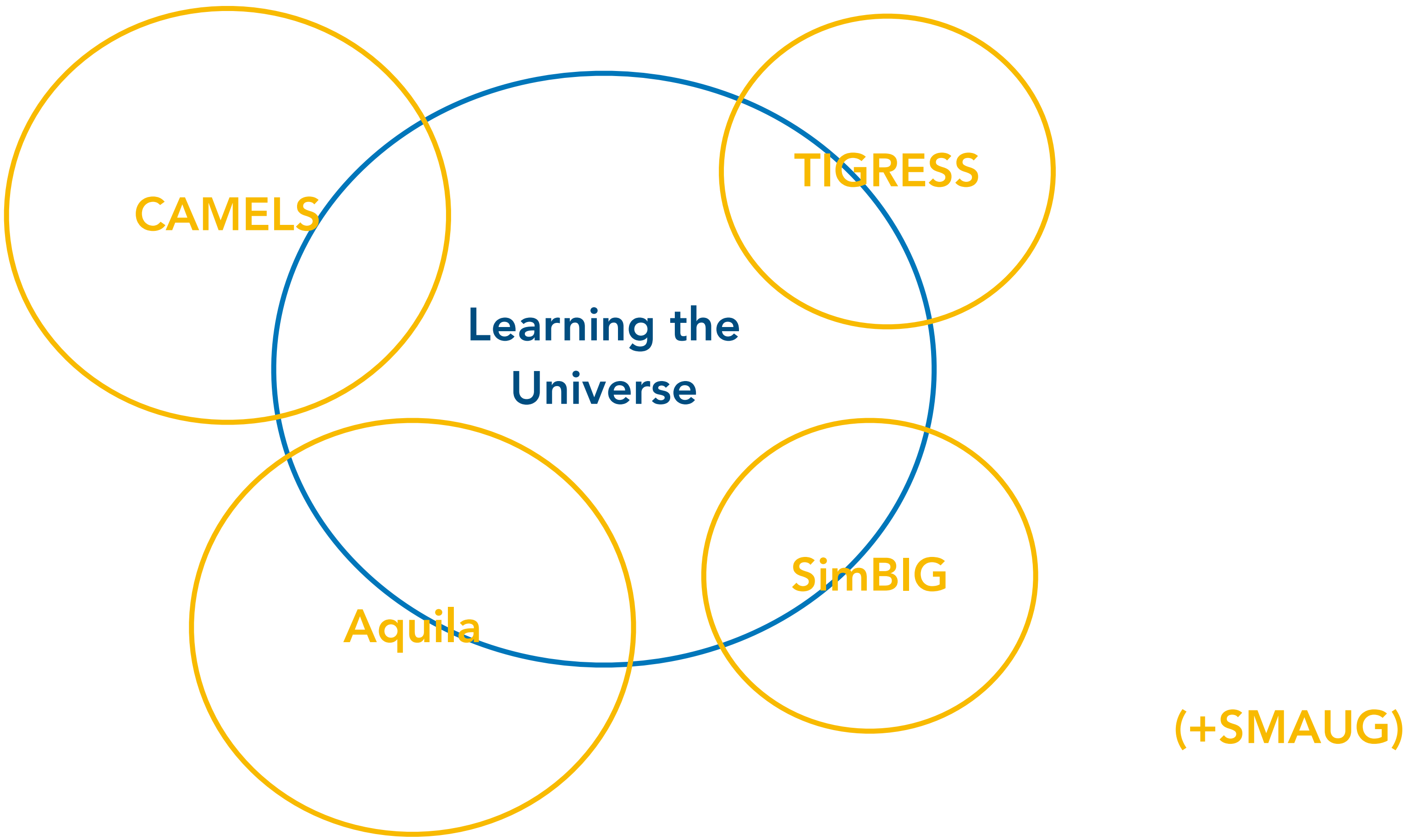
3. Infer "initial conditions" (parameters and phases)

- Use the accelerated forward models to constrain parameters using two approaches:
    1. A forward model built on a physical map from initial conditions to survey likelihood (BORG)
    2. Simulation-based inference (SBI): use machine to learn the posterior

LtU Working groups

Star formation/wind subgrid model (Ostriker, Kim, …)
Black hole accretion and feedback (Hernquist, …)
Cosmological Modeling (Springel, Burger, …)
Synthetic observations (Somerville/Ferraro, …)
Training set generation (Genel, Navarro, Angles-Alcazar, …)
Accelerated forward modeling (S. Ho, Lavasseur, Lemos, …)
Inference with physical model (BORG) (Jasche/Lavaux, …)
Simulation-based (IL) inference (Wandelt, M. Ho, …)
Understanding models/Inference robustness (Singh, …)

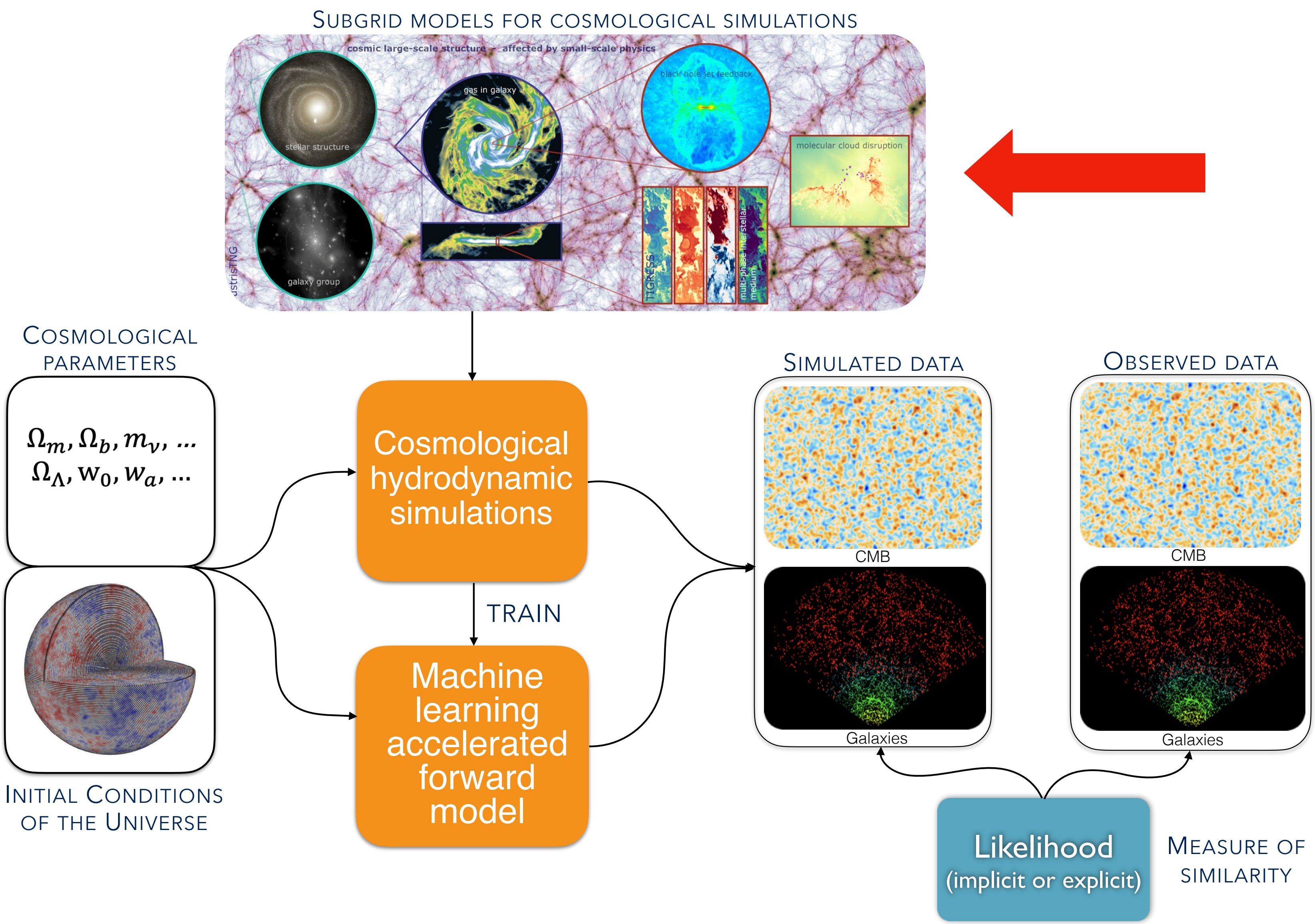
A COLLABORATION OF COLLABORATIONS?



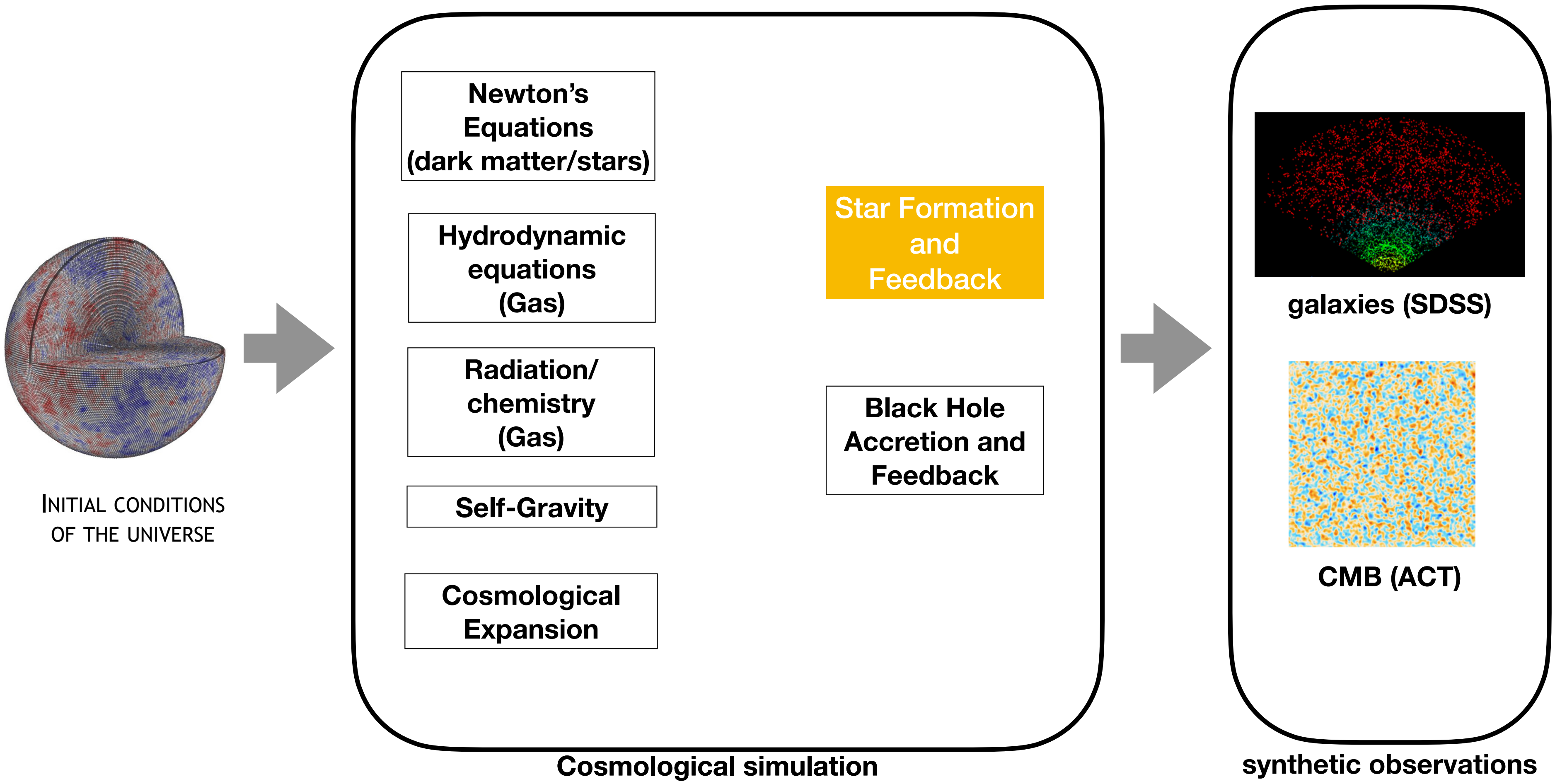
A COLLABORATION OF COLLABORATIONS?

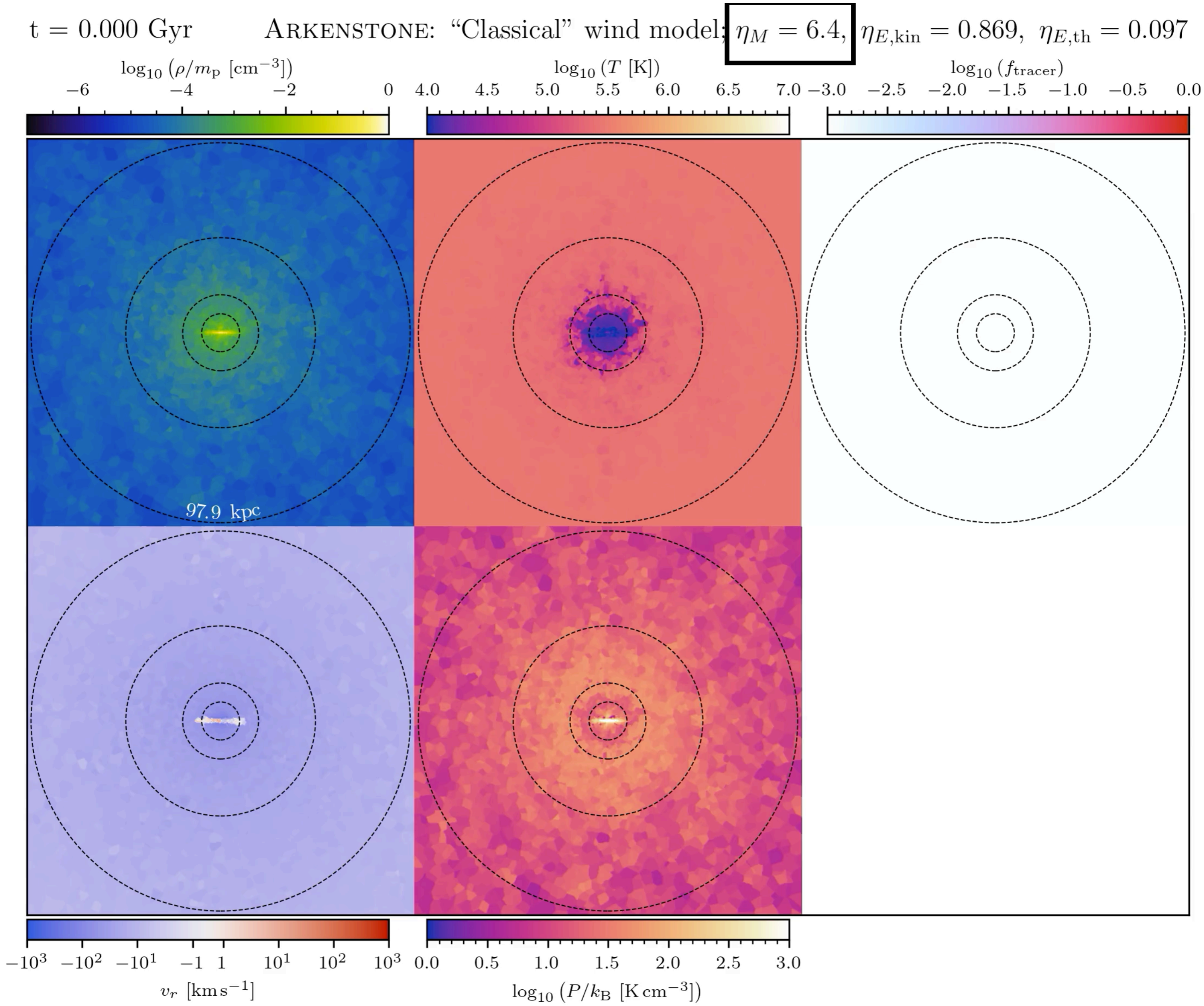
See talks by:

Ana Maria Delgado
Shivam Pandey
Ben Wandelt
Boon Kiat Oh
Francisco Villaescusa-Navarro
Yongseok Jo
Tjitske Starkenburg
Daniel Angles-Alcazar
Ulirch Steinwandel
Yueying Ni
Lucia Perez
Christopher Lovell


Subgrid models for cosmological simulations
cosmic large-scale structure
affected by small-scale physics
stellar structure
gas in galaxy
black hole jet feedback
molecular cloud disruption
galaxy group
TIGRESS
multi-phase interstellar medium
Cosmological parameters
$\Omega_m, \Omega_b, m_\nu, \ldots$
$\Omega_\Lambda, w_0, w_a, \ldots$
Initial conditions of the Universe
Cosmological hydrodynamic simulations
TRAIN
Machine learning accelerated forward model
Simulated data
CMB
Galaxies
Observed data
CMB
Galaxies
Likelihood
(implicit or explicit)
Measure of similarity

Galaxy Formation Simulations: Modeling Star Formation and Feedback





$v = 380\ \mathrm{km\,s^{-1}}$

$T = 3.8 \times 10^5\ \mathrm{K}$

"CLASSIC" WINDS

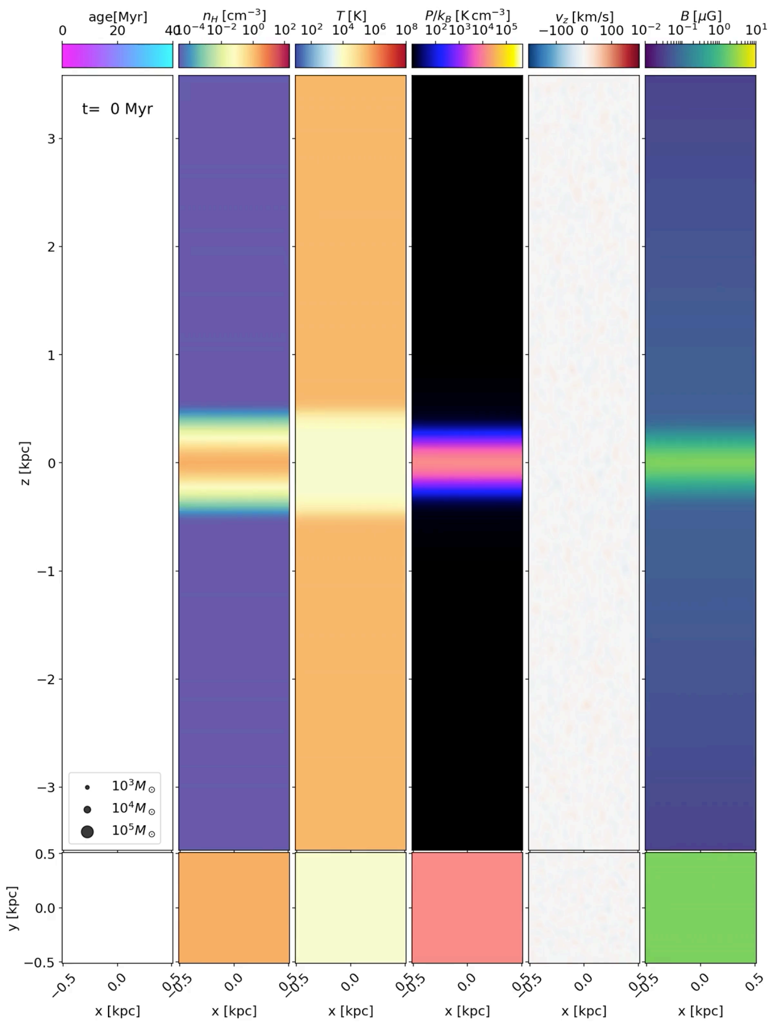
Matthew Smith (MPIA)
w/Drummond Fielding

MULTIPHASE OUTFLOWS

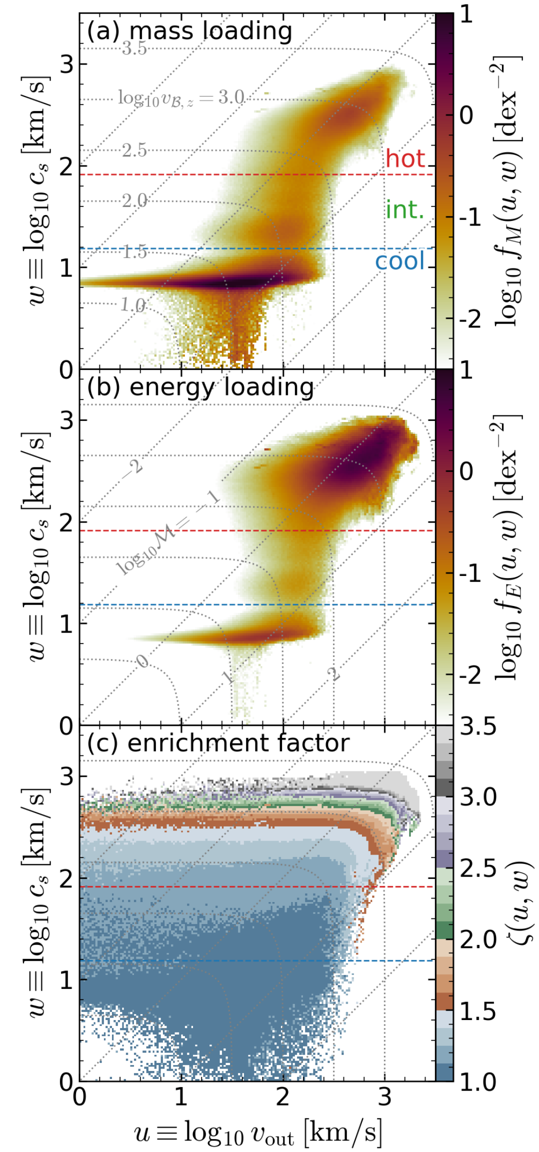
TIGRESS simulations produce highly multiphase outflows

Mass loading dominated by cold/warm slow moving material

Energy loading dominated by hot, fast, metal enriched outflow



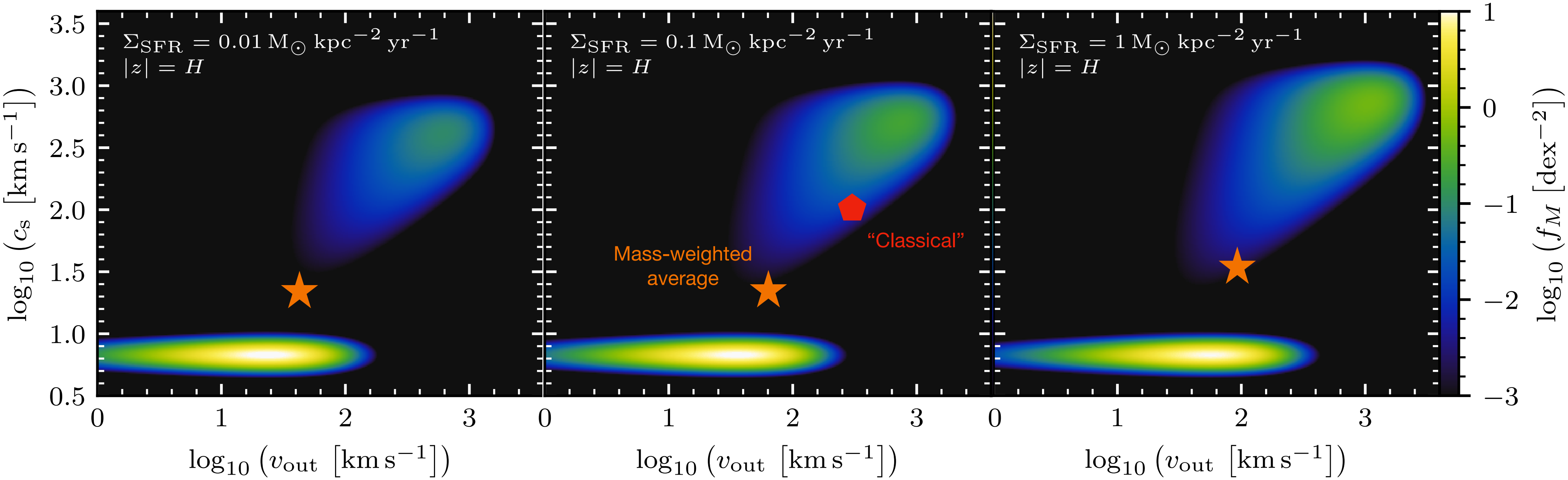
Kim & Ostriker 2017
Kim, Ostriker & SMAUG 2020



MULTIPHASE LAUNCHING

When launching a wind particle, draw velocity and temperature from a distribution, instead of using single value.

Kim, Ostriker & SMAUG 2020



Hot, fast winds

High specific energy outflows are hard to model:

- η_E/η_M is large, so coarse mass resolution means poor time sampling.
- Low density means poor spatial resolution.

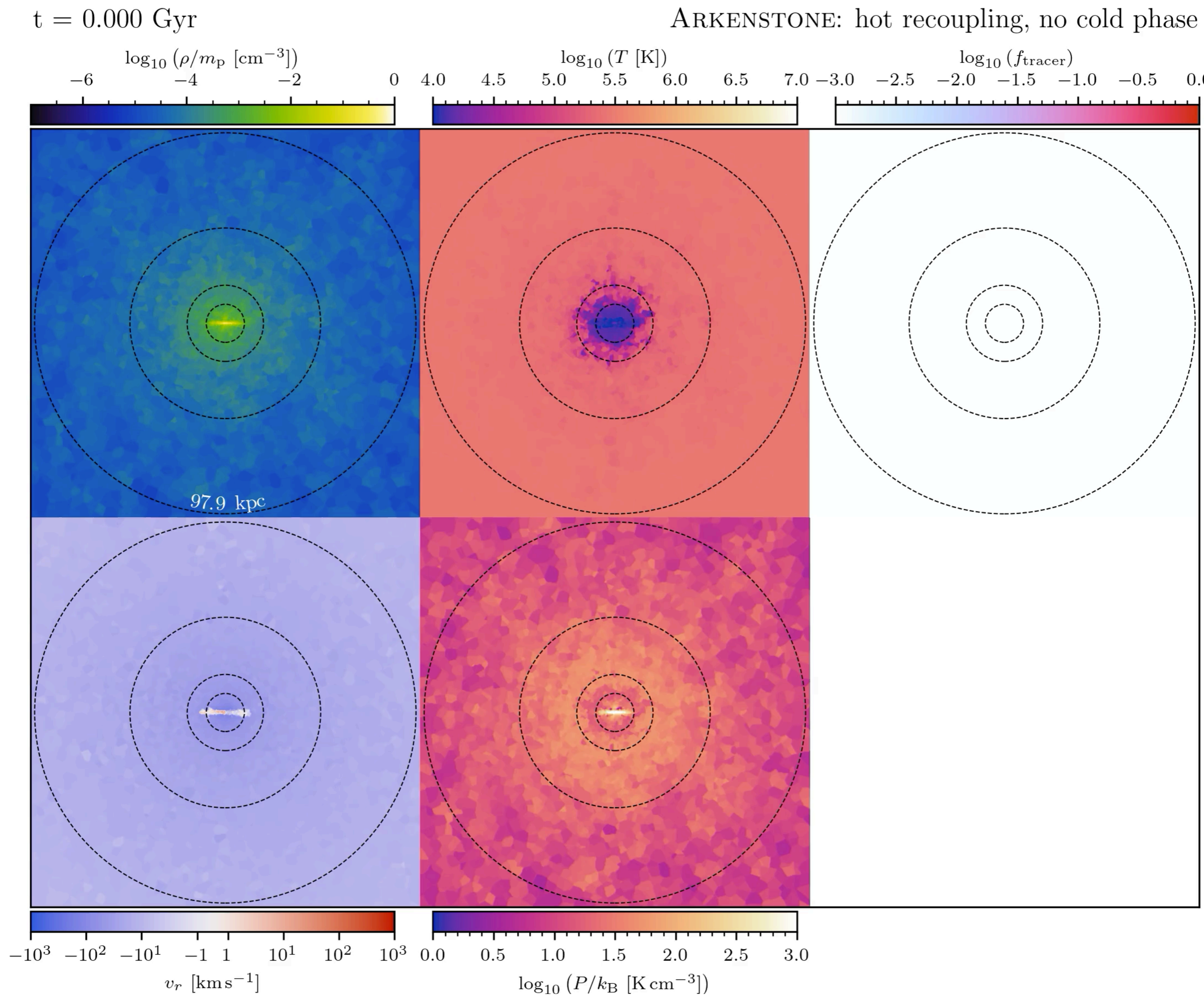
Arkenstone "hot recoupling" model:

1. Throw low mass wind particles (e.g. 100 times lower mass than gas cells)
2. Refine when recoupling

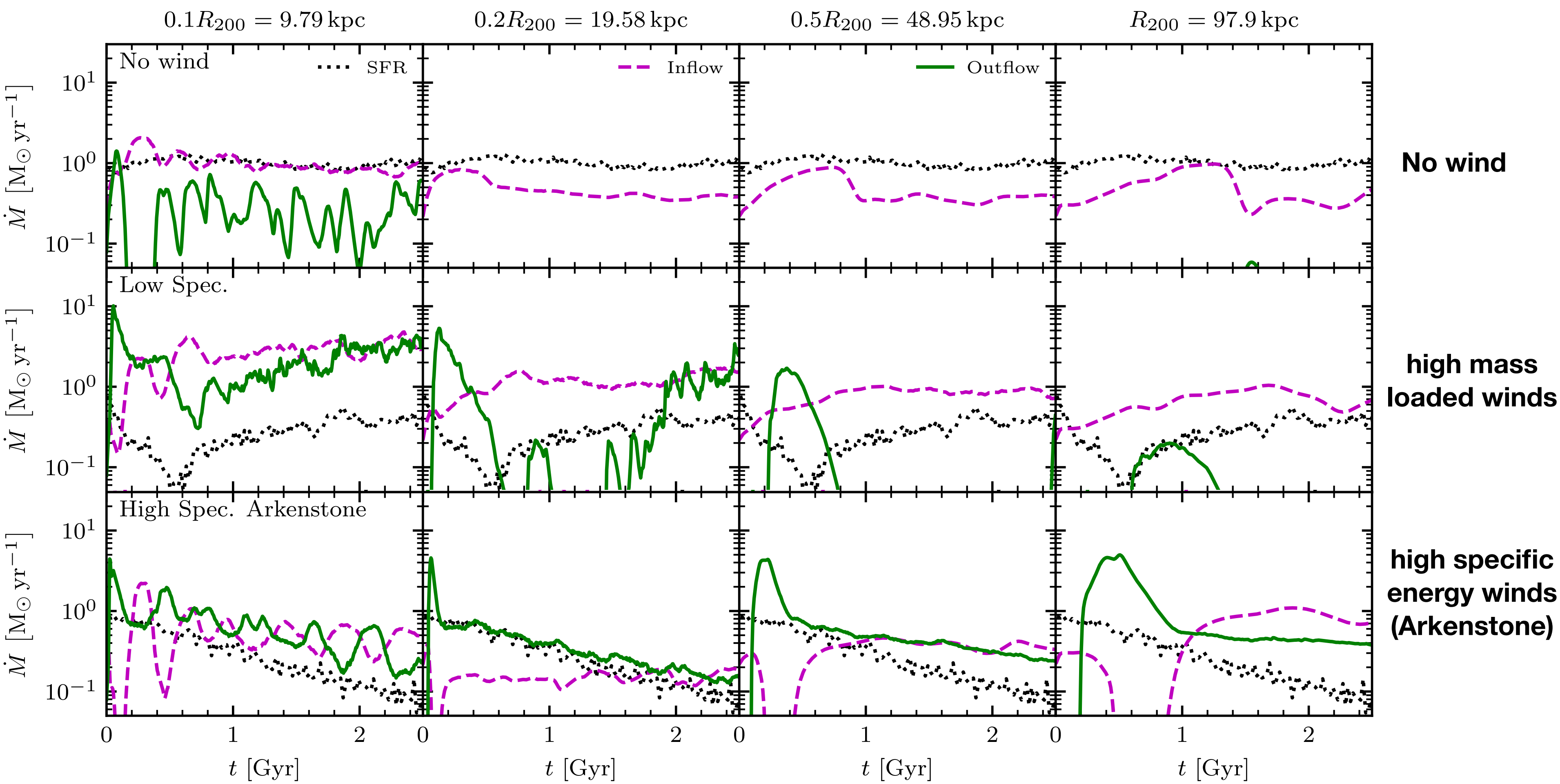
Gives smooth energy injection, high temperatures and refined wind.

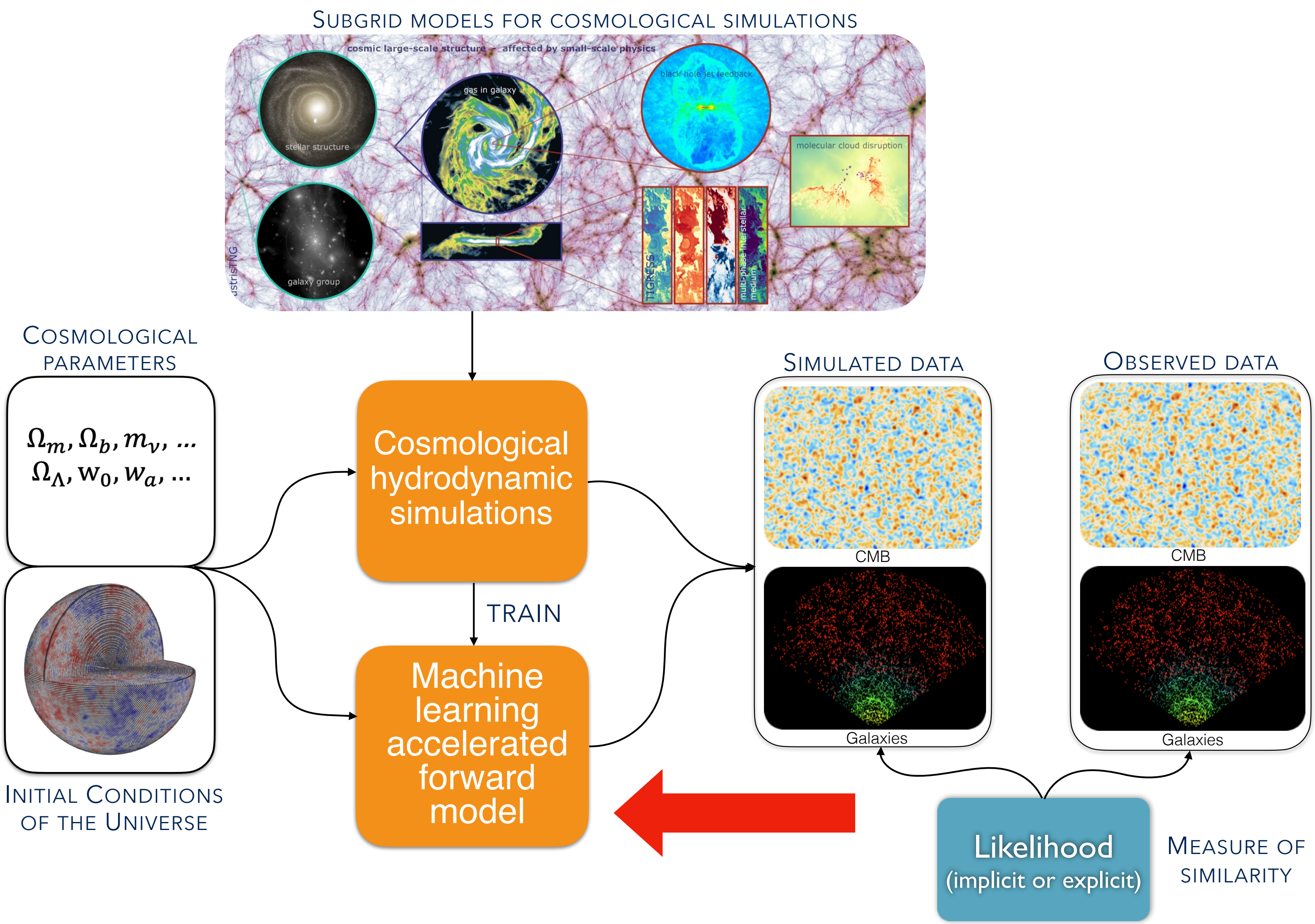
ARKENSTONE WINDS

(HOT RECOUPING)



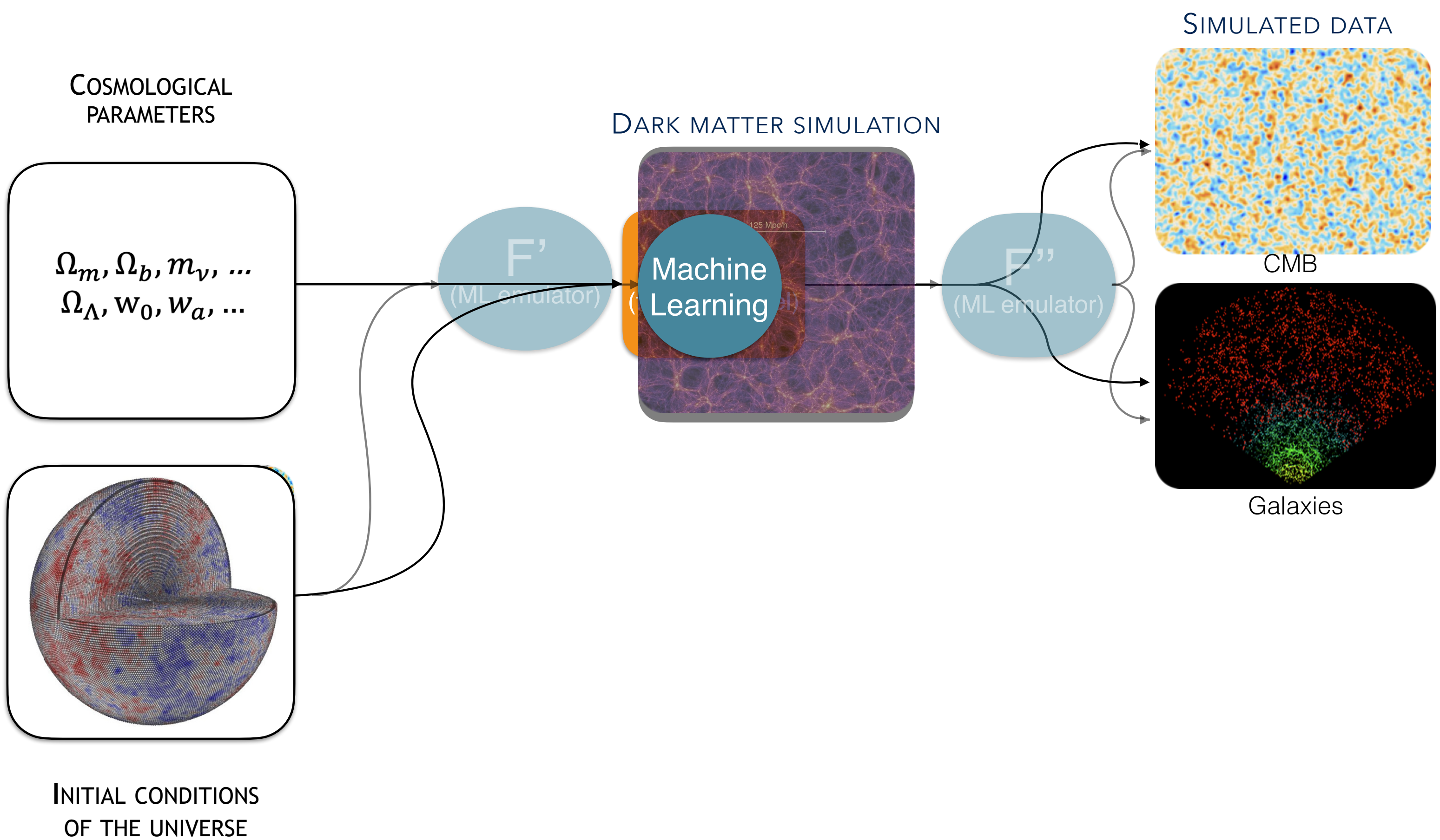
ARKENSTONE WINDS REGULATE STAR FORMATION WITH LOW MASS-LOADING





SUBGRID MODELS FOR COSMOLOGICAL SIMULATIONS
cosmic large-scale structure → affected by small-scale physics
stellar structure
gas in galaxy
black hole jet feedback
molecular cloud disruption
galaxy group
TIGRESS
multi-phase interstellar medium
COSMOLOGICAL PARAMETERS
$\Omega_m, \Omega_b, m_\nu, \ldots$
$\Omega_\Lambda, w_0, w_a, \ldots$
INITIAL CONDITIONS OF THE UNIVERSE
Cosmological hydrodynamic simulations
TRAIN
Machine learning accelerated forward model
SIMULATED DATA
CMB
Galaxies
OBSERVED DATA
CMB
Galaxies
Likelihood
(implicit or explicit)
MEASURE OF SIMILARITY

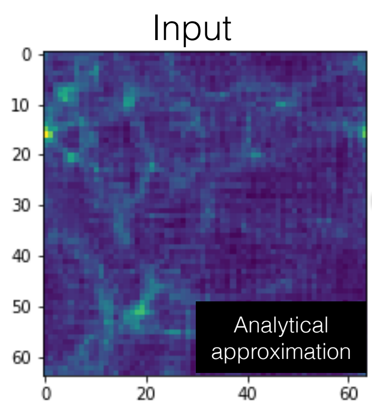
Accelerating forward modeling with machine learning

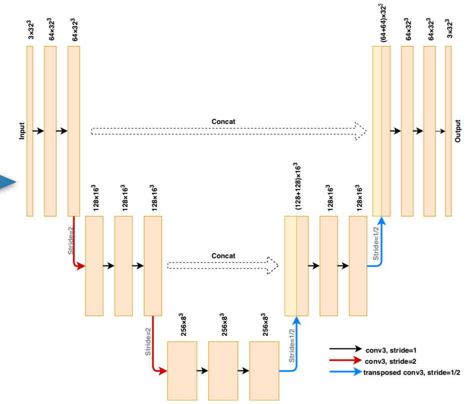


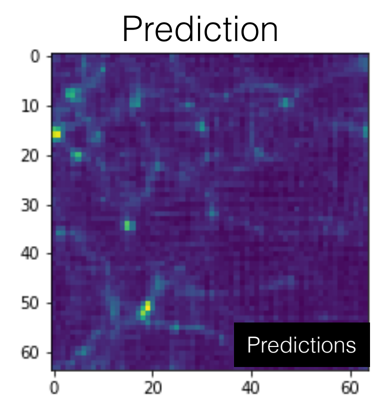
ACCELERATING N-BODY (DARK MATTER ONLY) SIMULATIONS WITH MACHINE LEARNING (F')

Instead of using **numerical simulations** as an approximation of newton's laws for DM particles, we make use of the universal approximation theorem to approximate them with a deep model.
Goal: using machine learning to
"learn"/ interpolate from a large number of pre-run simulations.

First step:
For a **fixed set of cosmological parameters**, over a **small volume** (512 (h^{-1}Mpc)^3), at **low resolution** (mean separation of particles 1 h^{-1} Mpc)







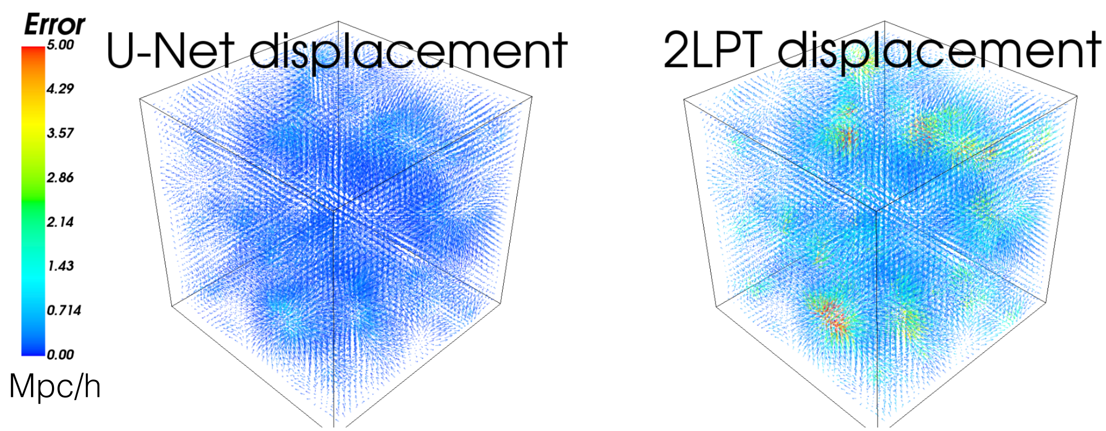
S. He, Y. Li, Y. Feng, S.Ho., S. Ravanbaksh, B. Poczos, **PNAS 2019**

Assessing Performances

Errors in **displacement field** (difference between current position to the initial position of the particles), predicted by the benchmark model (2LPT), and the ML model

Machine Learning Model
prediction errors

Benchmark (2LPT)
prediction errors



S. He, Y. Li, Y. Feng, S.Ho., S. Ravanbaksh, B. Poczos, **PNAS 2019**

CNN WITH "STYLE" - GENERALIZE TO DIFFERENT COSMOLOGIES

Comparing the following:

1) The average **power spectrum** of 1000 sims,

$$\hat{P}_{A\times B}(k) = \frac{1}{V}\int \frac{\mathrm{d}\Omega_{\boldsymbol{k}}}{4\pi}\delta_A(\boldsymbol{k})\delta_B^*(\boldsymbol{k})$$

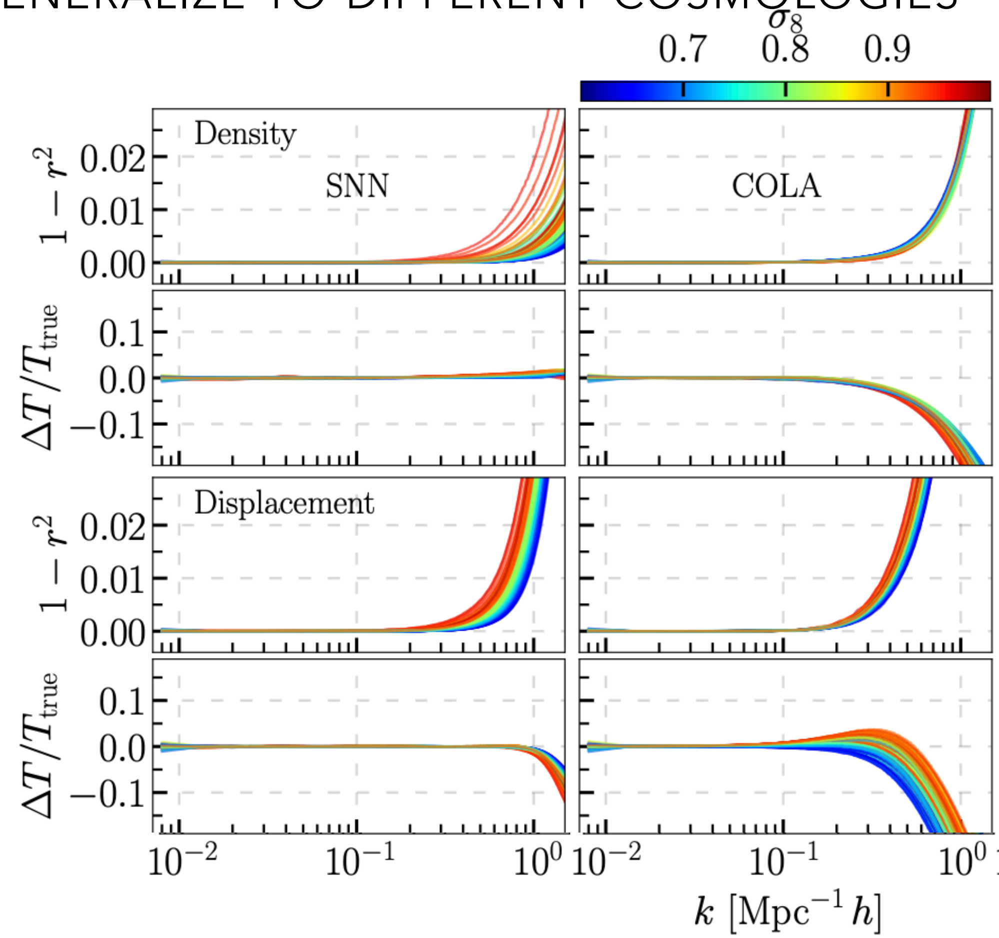
2) The **ratios** to the true power-spectrum (T(k)),

$$T(k) = \frac{P_{\mathrm{pred}}(k)}{P_{\mathrm{true}}(k)}$$

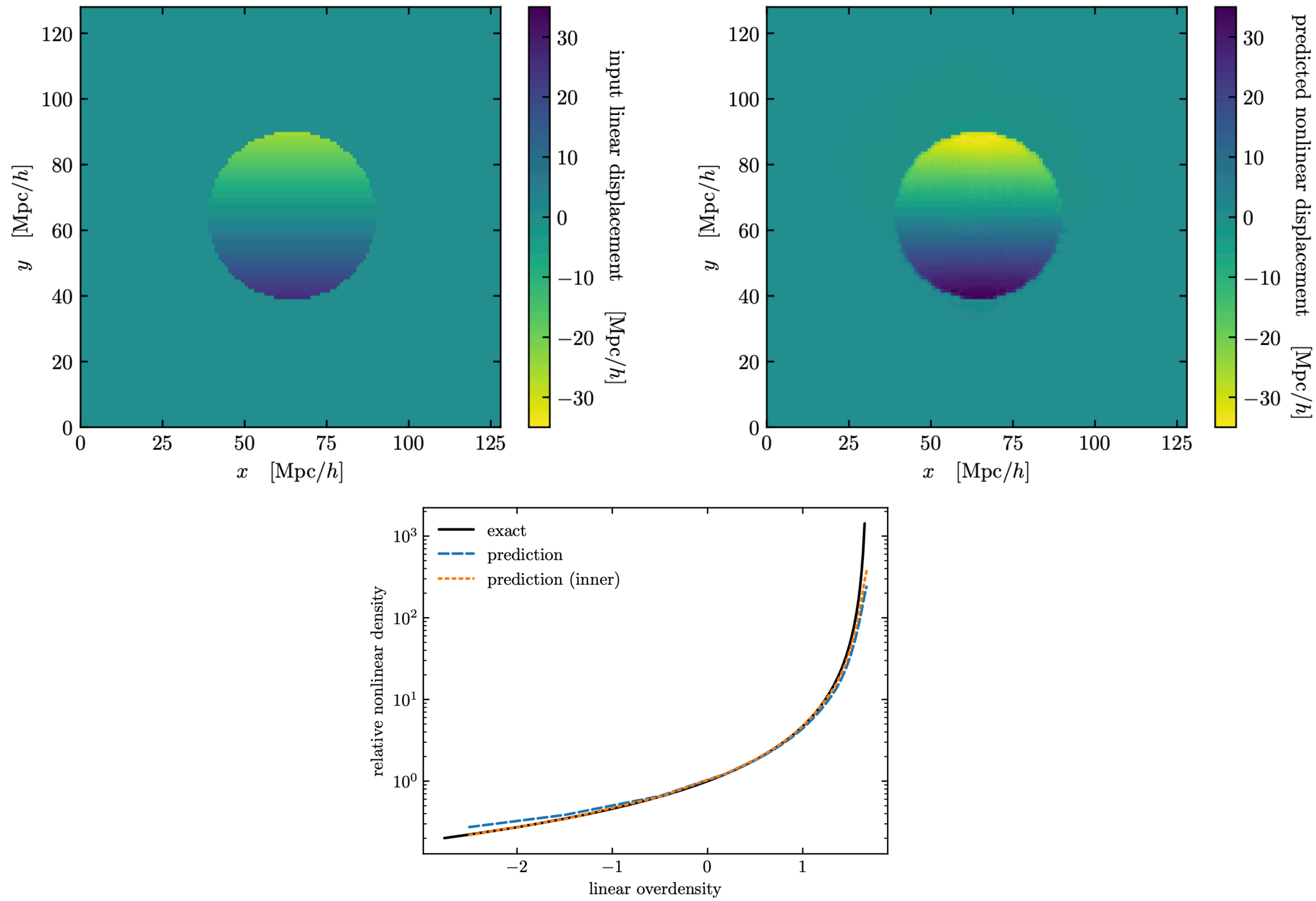
3) The **cross-correlation** coefficients.

$$r(k) = \frac{P_{\mathrm{pred}\times\mathrm{true}}(k)}{\sqrt{P_{\mathrm{pred}}(k)P_{\mathrm{true}}(k)}}$$

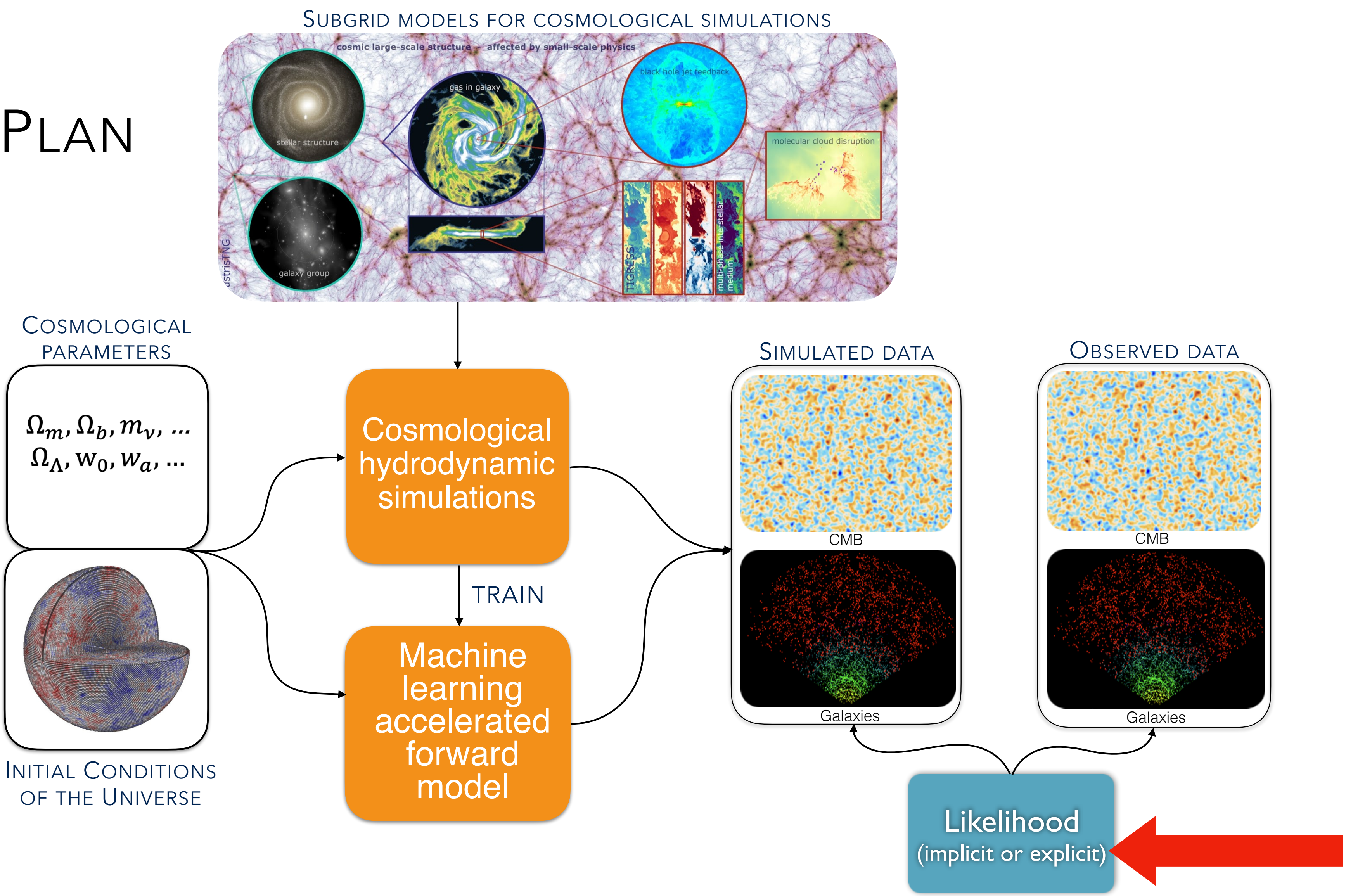
Jamieson+ 2022



What is the network learning?

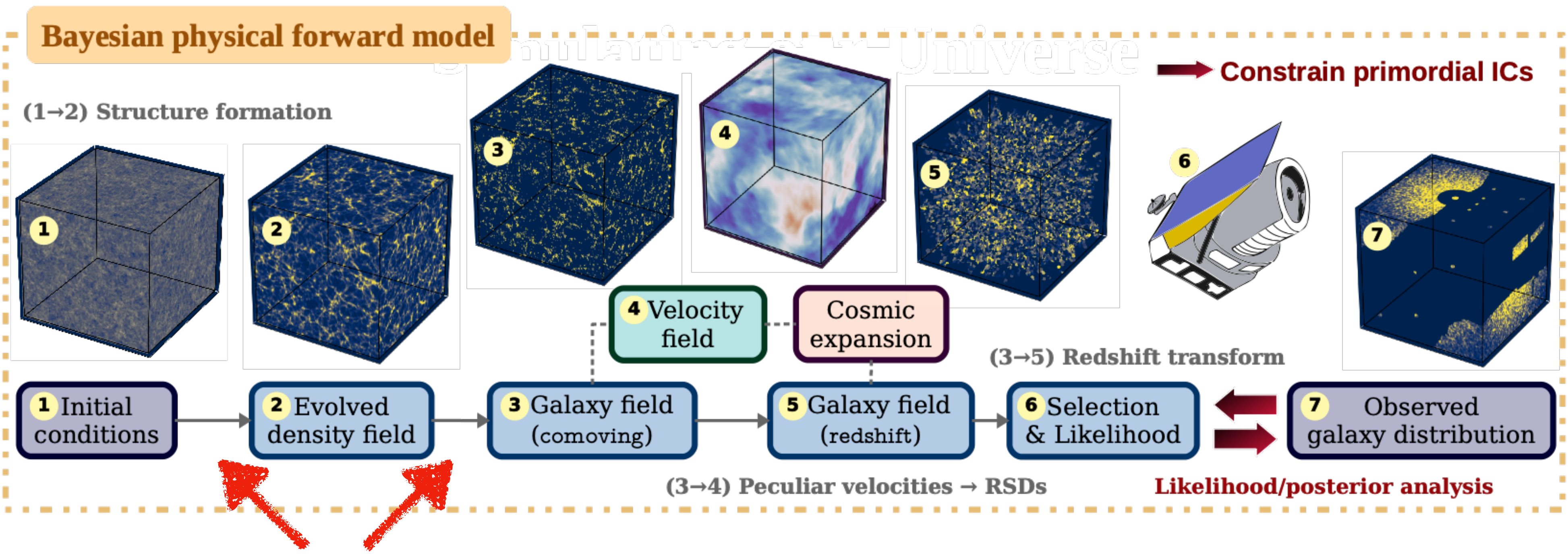


The Plan



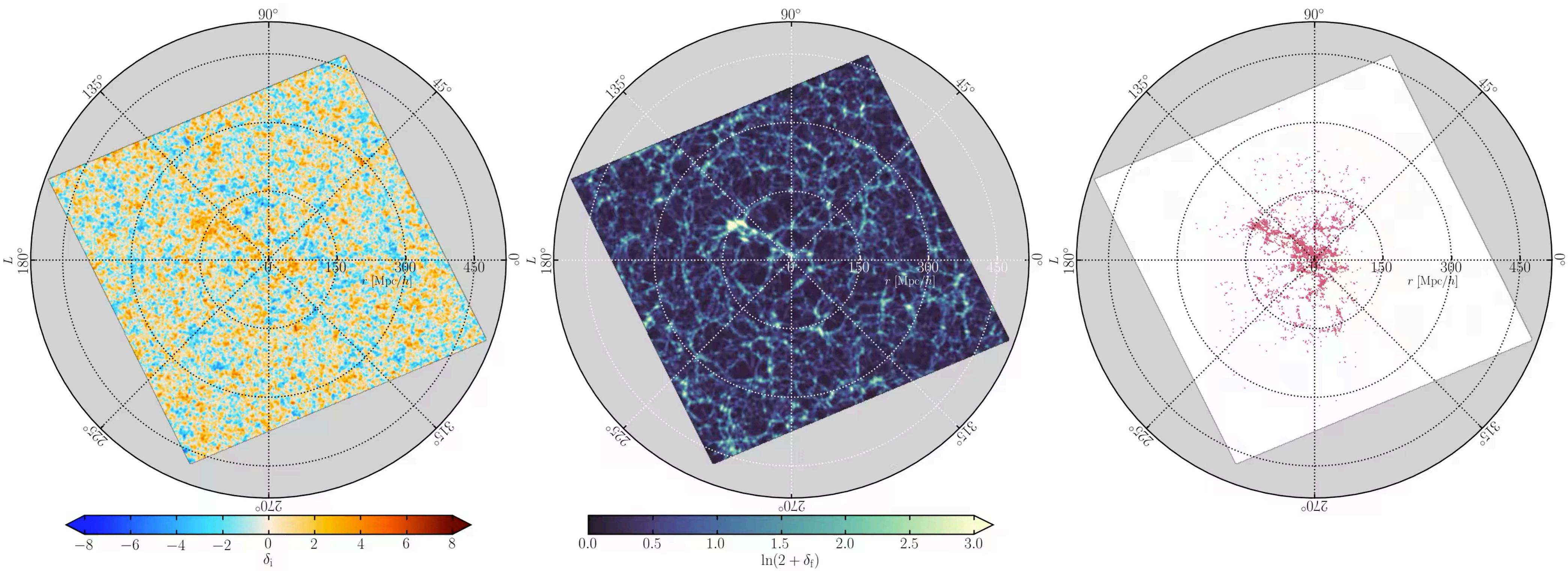
Subgrid models for cosmological simulations
cosmic large-scale structure
affected by small-scale physics
stellar structure
galaxy group
gas in galaxy
black hole jet feedback
molecular cloud disruption
Cosmological parameters
$\Omega_m, \Omega_b, m_\nu, \ldots$
$\Omega_\Lambda, w_0, w_a, \ldots$
Initial conditions of the Universe
Cosmological hydrodynamic simulations
TRAIN
Machine learning accelerated forward model
Simulated data
CMB
Galaxies
Observed data
CMB
Galaxies
Likelihood
(implicit or explicit)

BENCHMARK EXPLICIT LIKELIHOOD APPROACH WITH BORG (BAYESIAN ORIGIN RECONSTRUCTION FROM GALAXIES)



ML-accelerated gravity and hydro model for 3-5

MCMC EXPLORATION OF THE INITIAL CONDITIONS WITH BORG



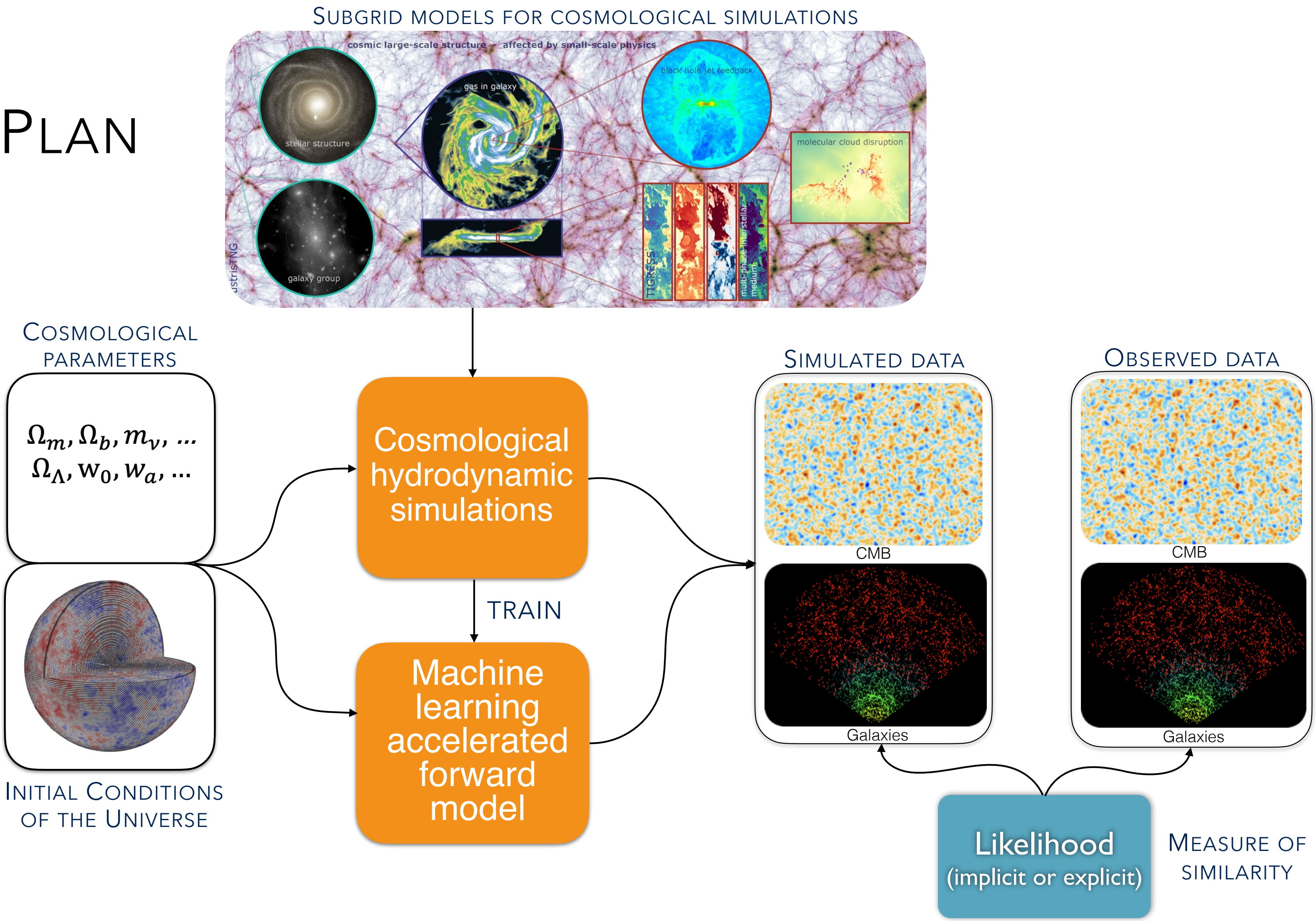
Initial Conditions

Evolved DM

2M++ Galaxy survey
(centered on Milky Way)

Jasche & Lavaux 2019

The Plan



Subgrid models for cosmological simulations
cosmic large-scale structure
affected by small-scale physics
stellar structure
gas in galaxy
black hole jet feedback
molecular cloud disruption
galaxy group
TIGRESS
multi-phase interstellar medium
Cosmological parameters
$\Omega_m, \Omega_b, m_\nu, \ldots$
$\Omega_\Lambda, w_0, w_a, \ldots$
Initial conditions of the Universe
Cosmological hydrodynamic simulations
TRAIN
Machine learning accelerated forward model
Simulated data
CMB
Galaxies
Observed data
CMB
Galaxies
Likelihood
(implicit or explicit)
Measure of similarity