Title: Simulations of Cosmological Structure and Machine Learning

Speakers: Simeon Bird

Series: Cosmology & Gravitation

Date: October 19, 2021 - 11:00 AM

URL: https://pirsa.org/21100002

Abstract: The large scale distribution of matter in the Universe contains the answers to many mysteries, such as the nature of dark matter, the reionization of the Universe, and the growth of galaxies. Cosmological simulations are the only way to understand these questions. I will talk about how modern current simulation models, work, discuss some new models and improvements in our latest simulation runs, especially our implementations of reionization and cosmology. I will then talk about some new work to dramatically expand the region of applicability of these simulations using machine learning. This can both to expand their dynamic range and combine different simulations to infer the physical parameters of the Universe.

Simulations of Cosmological Structure and Machine Learning

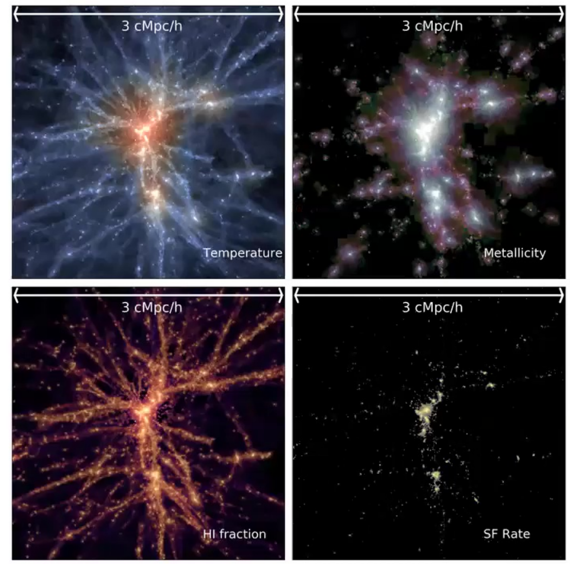
Simeon Bird, UCR

Cosmological Simulations: What?

Simulations of initially uniform 20-1000 Mpc box.

Fair sample of structures, box scale linear.

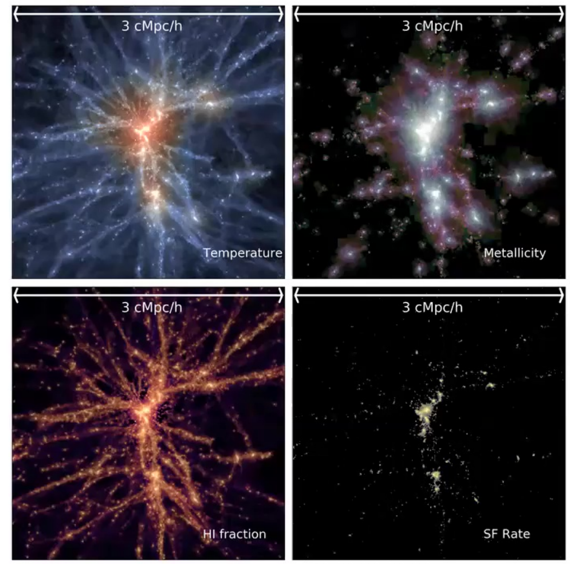
Full-physics



Cosmological Simulations: Why?

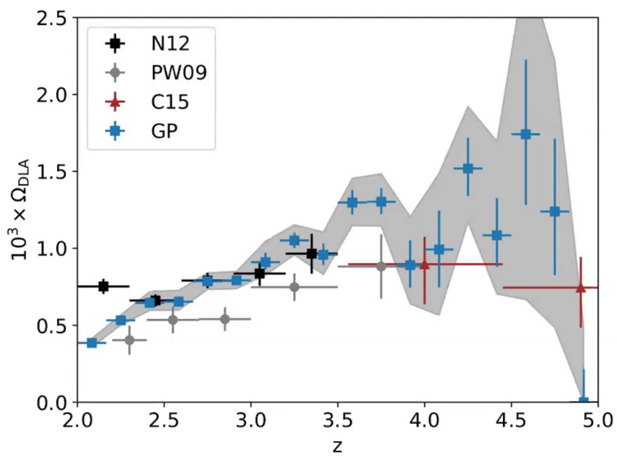
For:

- Galaxy formation
- Dark matter
- Cosmology

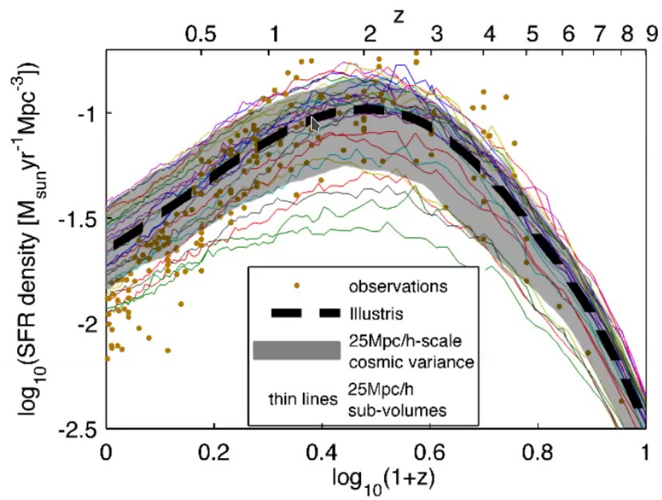


Galaxy Formation

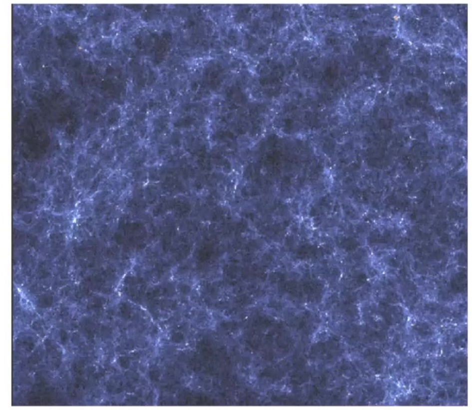
How does gas become stars?



Ho+ 2020



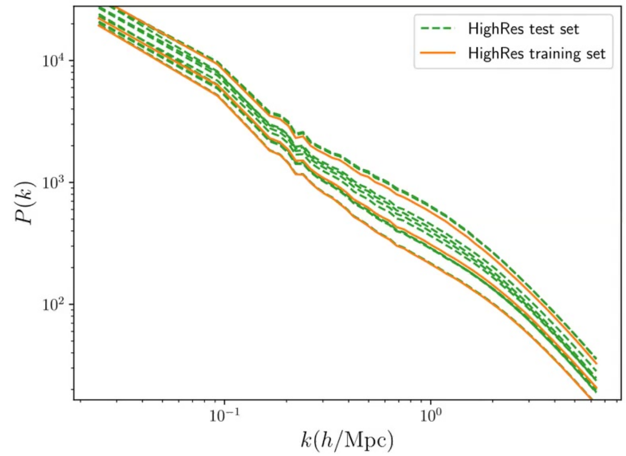
Dark Matter



- Cosmic structure proves that dark matter is 5/6 of the matter in the Universe.
- Non-minimal dark matter models can leave their imprint in cosmic structure
- … if we understand structure formation

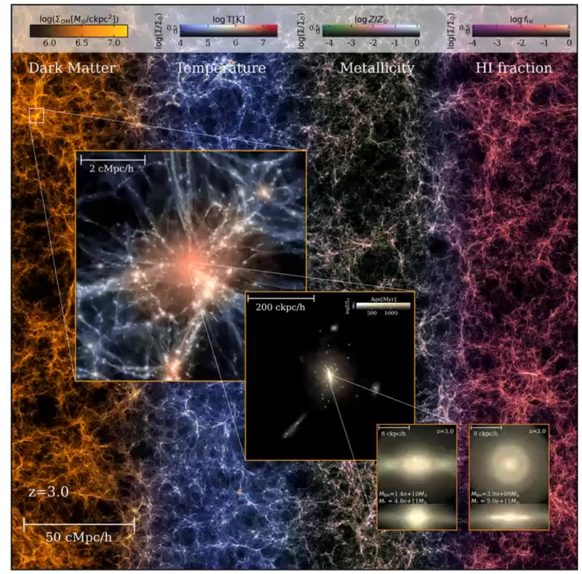
Cosmology

Parameter estimation:
simulated summary statistic
(eg: galaxy power spectrum)



How to Simulate

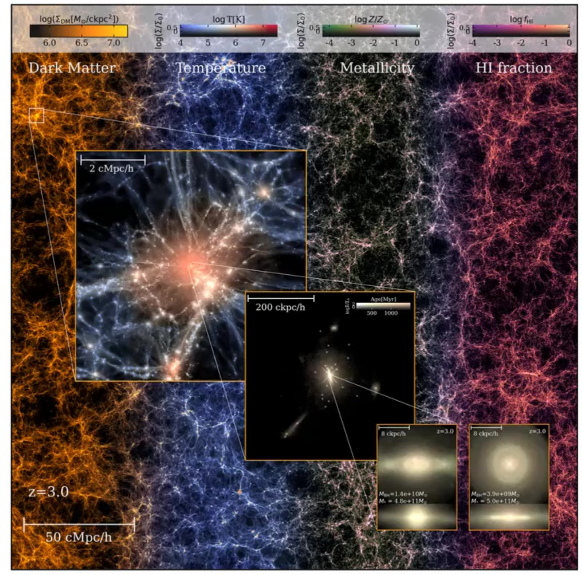
- We simulate initially ~uniform density gas & DM until it forms galaxies
- Need:
  - Dark matter, gravity
  - Gas pressure
  - Gas cooling and ionization
  - Supernova heating
  - Black hole heating and merging



ASTRID simulation

How to Simulate

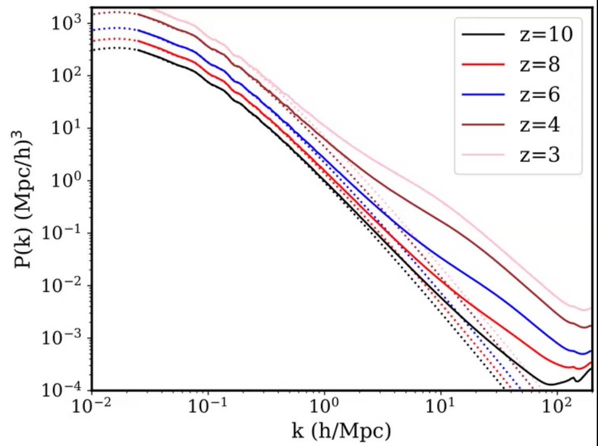
- Dark matter, gravity
  - Most well understood part
  - **Very important!**
  - Treat as many collisionless 'particles':
  - TNG100: 1820^3
  - ASTRID: 5500^3
  - DM-only: 10000^3



Black Hole Heating

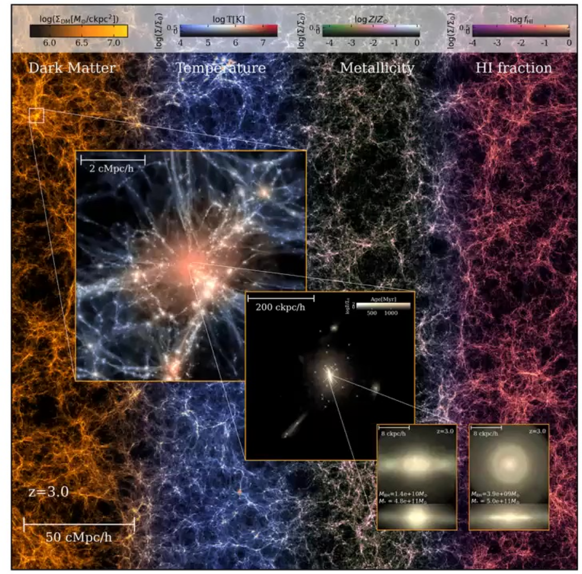
Black hole heating expels gas to ~1 Mpc distance.

- Lowers P(k) 20-30% at z=0
- $(\Omega_b / \Omega_m)^2$
- Subgrid: not well understood



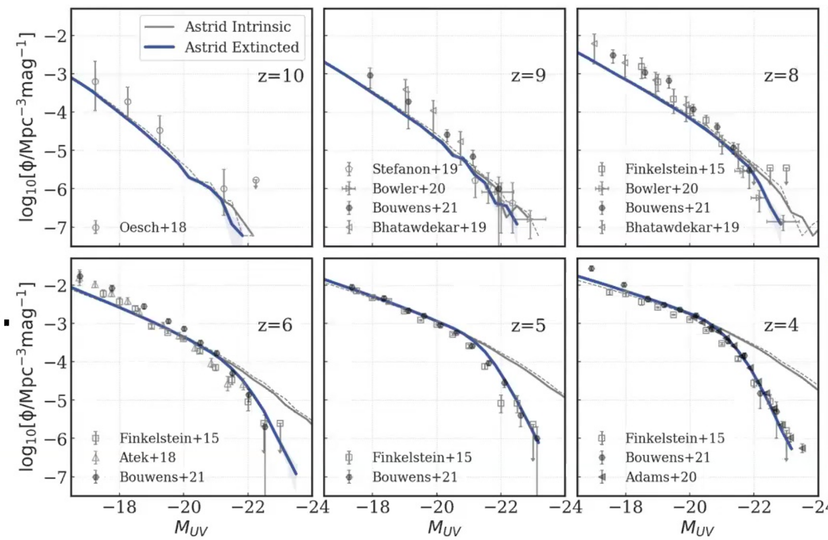
How to Simulate

- We simulate initially ~uniform density gas & DM until it forms galaxies
- Need:
  - Dark matter, gravity
  - Gas pressure
  - Gas cooling and ionization
  - Supernova heating
  - Black hole heating and merging



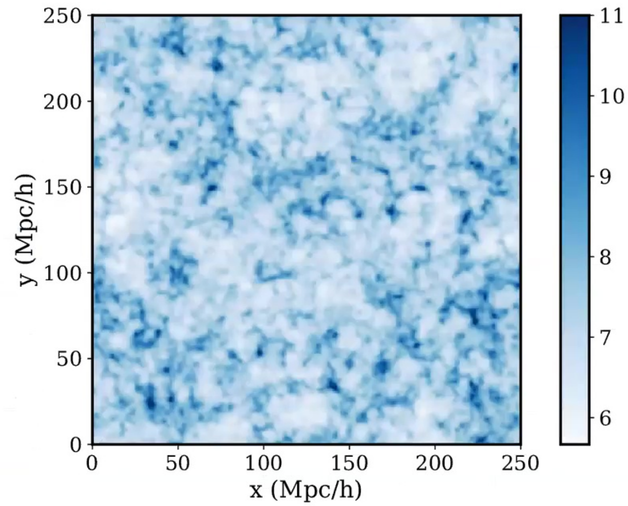
New simulation: ASTRID

- 250 Mpc/h
- 2x5500^3 particles
- Designed to have ~10 MW halos at z=5, ~200 MW halos at z=4.



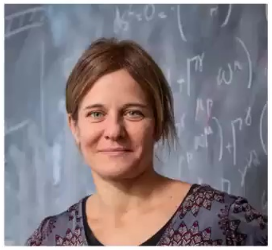
New simulation: ASTRID

- Hydrogen Reionization
- Helium Reionization
- Massive neutrinos
- Baryon relative velocities

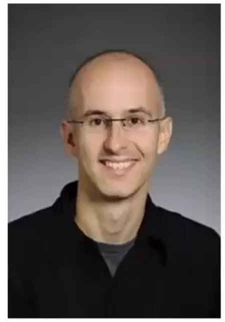


ASTRID

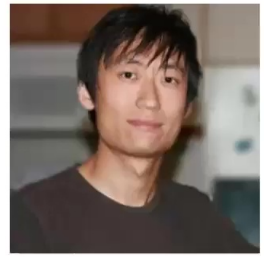
ASTRID Team



Tiziana Di Matteo (CMU)



Rupert Croft (CMU)

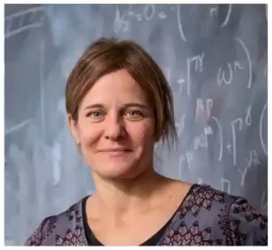


Yu Feng (Google)

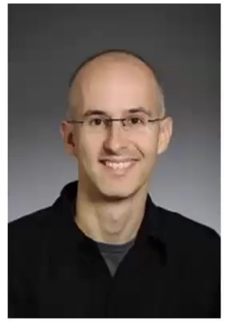
Yueying Ni (CMU)

(and me!)

ASTRID Team



Tiziana Di Matteo (CMU)



Rupert Croft (CMU)

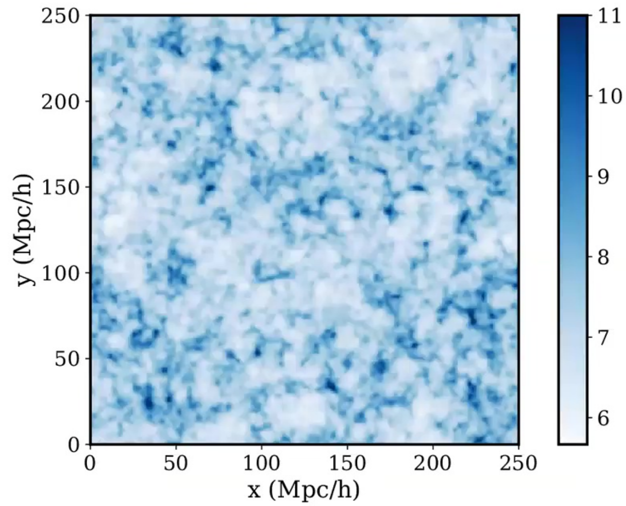
Yu Feng (Google)

Yueying Ni (CMU)

(and me!)

New simulation: ASTRID

- Hydrogen Reionization
- Helium Reionization
- Massive neutrinos
- Baryon relative velocities

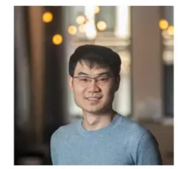


ASTRID

Machine Learning

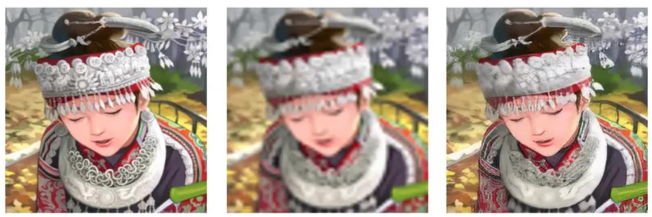
- Enhance dynamic range
- Super-resolution images applied to particle displacement

Yueying Ni (CMU)



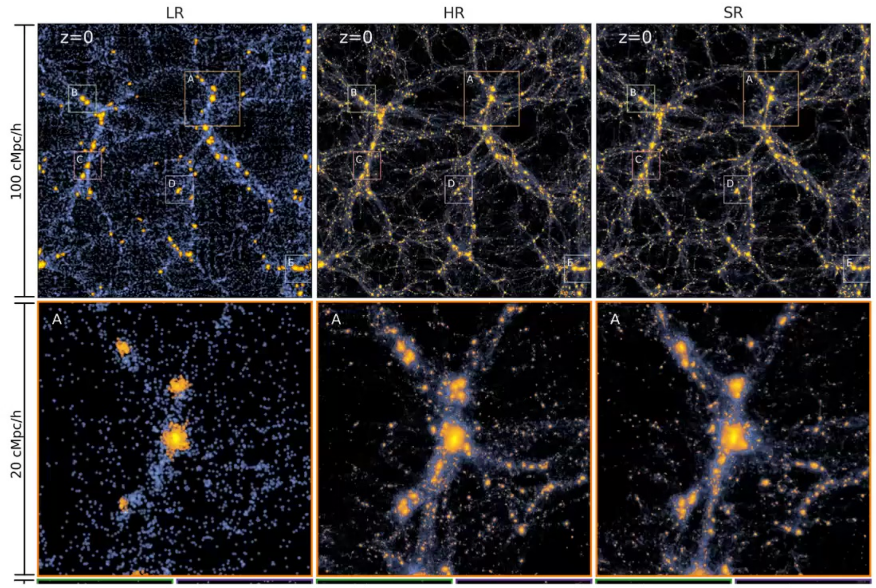
Yin Li (Flatiron)

High Low Enhanced



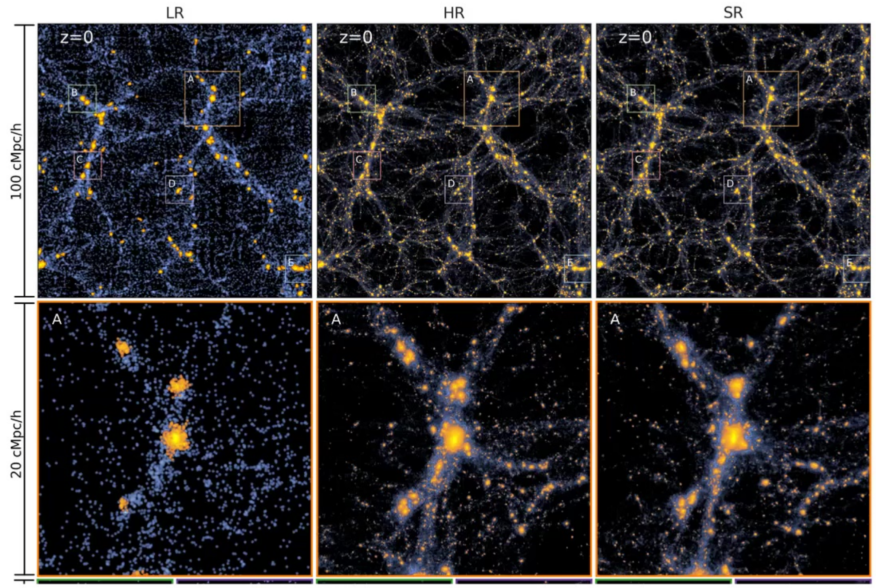
(SRSGan)

Super-resolution Simulations



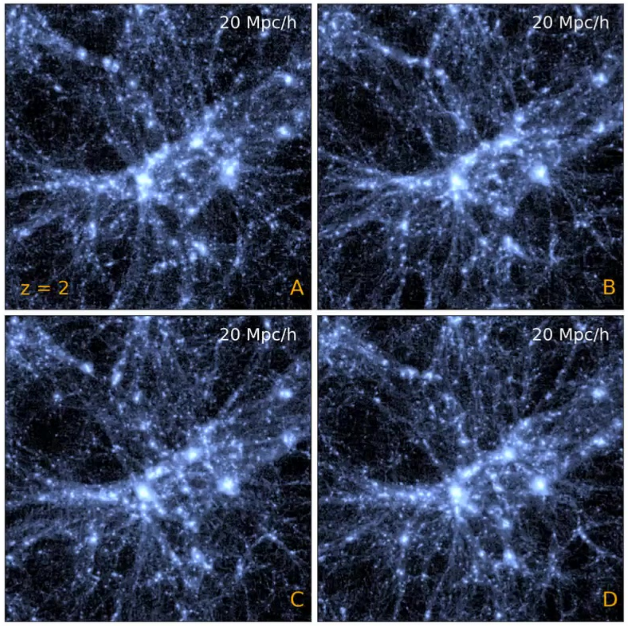
- Uses Generative Adversarial Networks - like AlphaGo – to learn particle displacement field
- Knows mapping from low-res displacement to high-res displacement
- Outputs full particle snapshot

Super-resolution Simulations



- Outputs full particle snapshot
- LR: 64^3 100 Mpc/h
- HR: 512^3 100 Mpc/h
- 16 pairs for training
- Same cosmology
- Different realisations

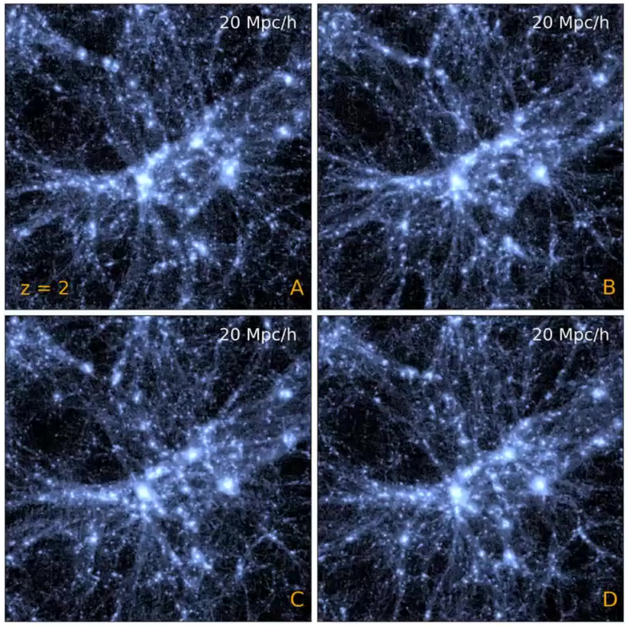
Super-resolution Simulations



- Multiple draws consistent with statistics
- Which is the true HR?

A

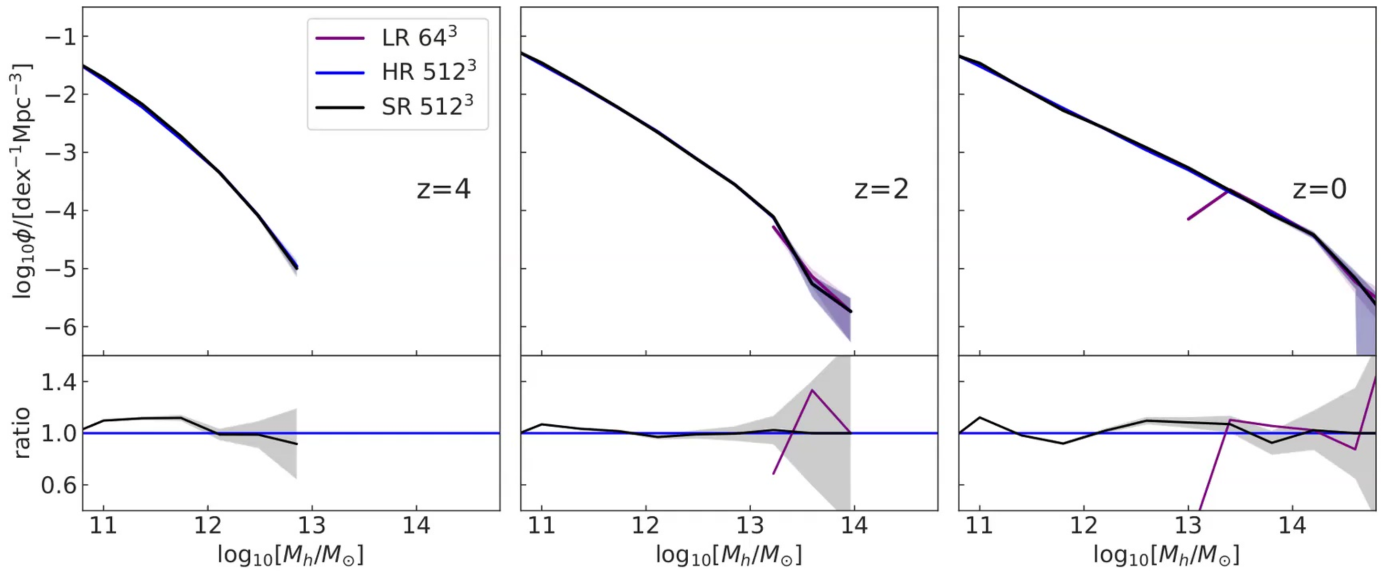
Super-resolution Simulations



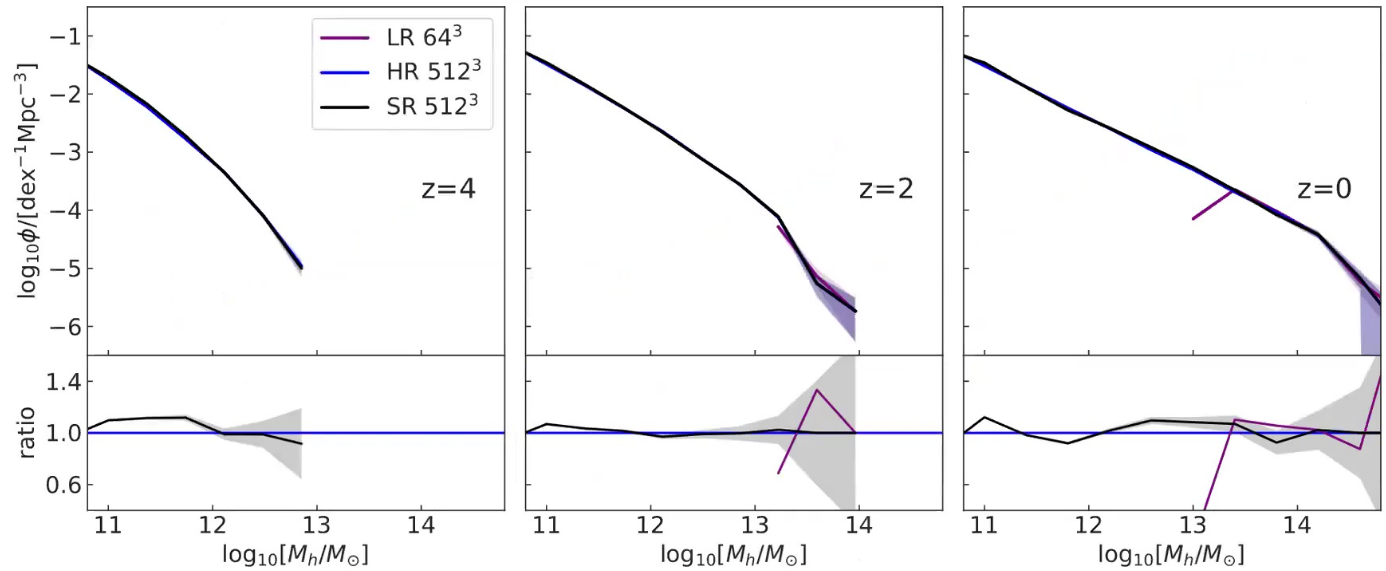
- Multiple draws consistent with statistics
- Which is the true HR?

A

Halo Mass Functions



Halo Mass Functions

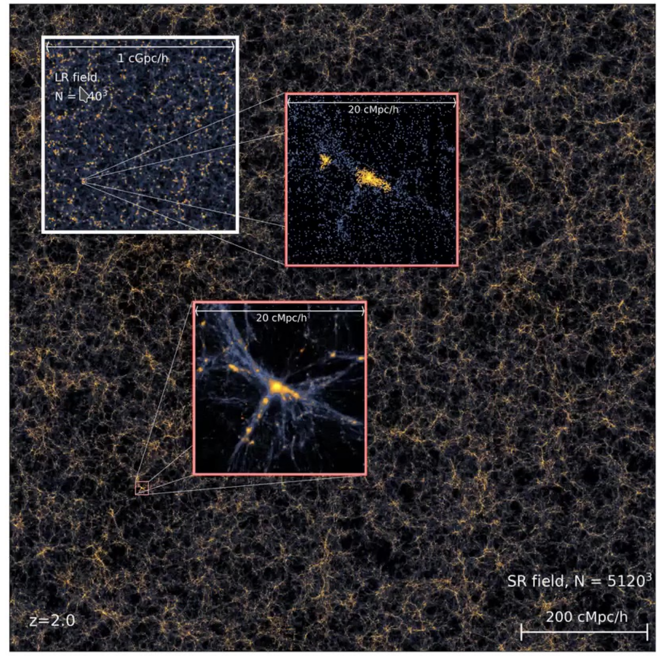


Large Mock Catalogues

Can combine large box with high resolution.

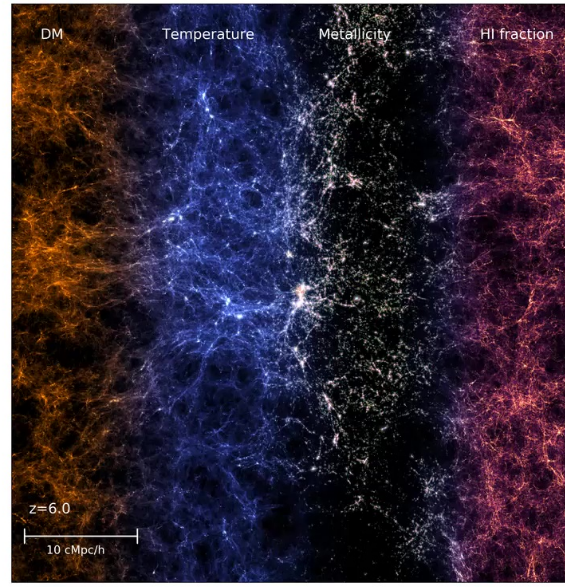
Simulate **one patch** of a large volume at high resolution.

Enhance the rest

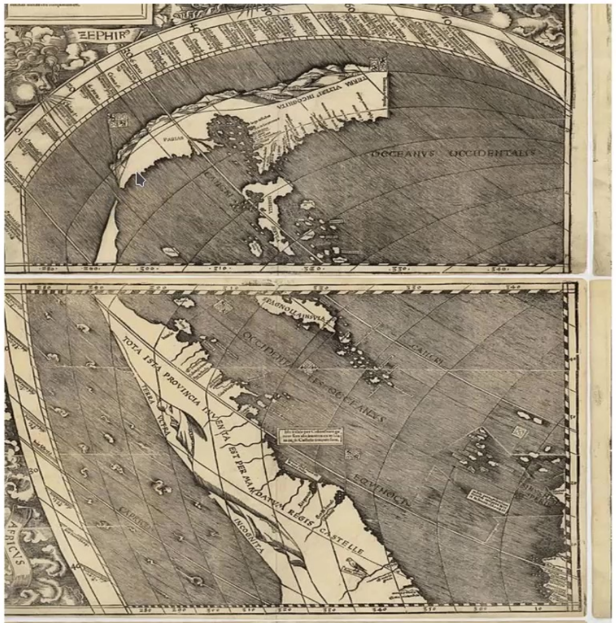


Simulation Interpolation

- We often want to have multiple simulations with different initial conditions.
- Need a map: cosmology → observables
- Interpolate; estimate output of simulations from existing set without running them.
- We need ~20-200.



Structure Simulations



- Early maps of the world
- had a similar problem

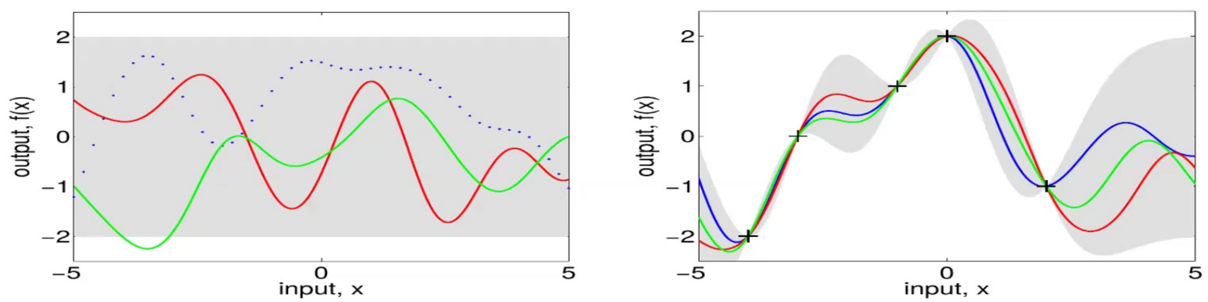
Measurements

- We are mapping Panama to 1% without ever visiting!

Universalis Cosmographia, the Waldseemüller map dated 1507 (wikipedia)

Traditional: use Gaussian Process on homogeneous simulations

- Bayesian function interpolation, which computes probability distribution of f(x) conditional on input set.
- Magic in **kernel function:** how correlation between function depends on parameter distance.

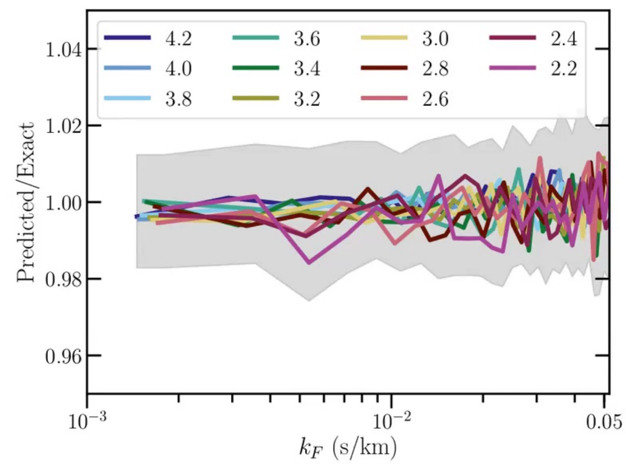


Rasmussen & Williams (GPML)

Traditional: use Gaussian Process on homogeneous simulations

- Works well at 1% level, but each simulation must be full resolution and expensive.

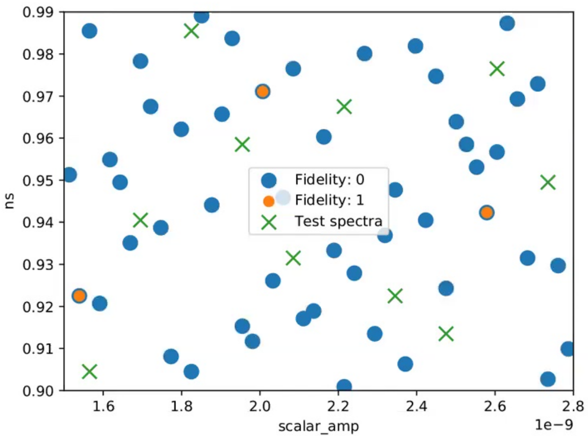
- Do we need all 50 high resolution simulations?
- Some information is duplicated



Multi-Fidelity Emulation: Ming-Feng Ho

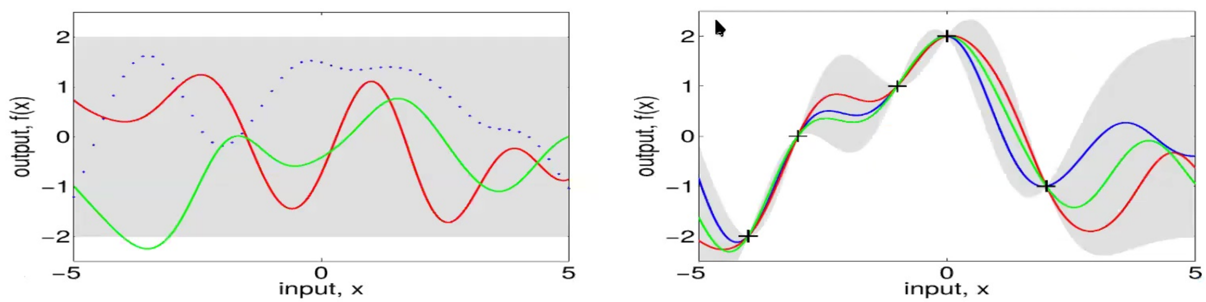
Combine simulations at **different resolutions**.

- Low resolution for cosmology
- GP on low resolution, correct with high resolution.
- Correction depends on cosmology.



Traditional: use Gaussian Process on homogeneous simulations

- Radial basis function kernels, separate length scale for each parameter



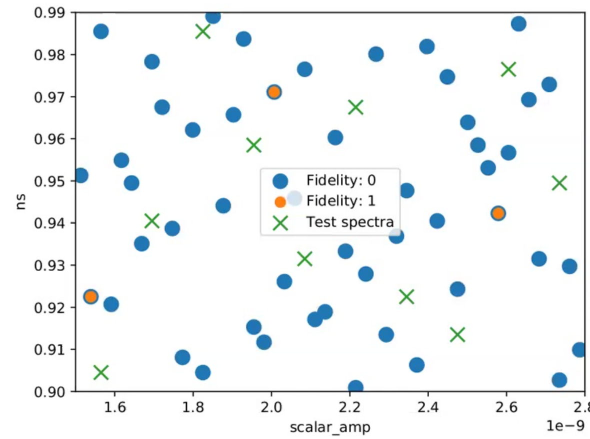
Rasmussen & Williams (GPML)

Multi-Fidelity Emulation: Ming-Feng Ho

Correction function:

f is the GP at fidelity t

$$f_t(k, x) = \rho_{t,j}\, f_{t-1}(x, k) + \delta_t(x, k),$$



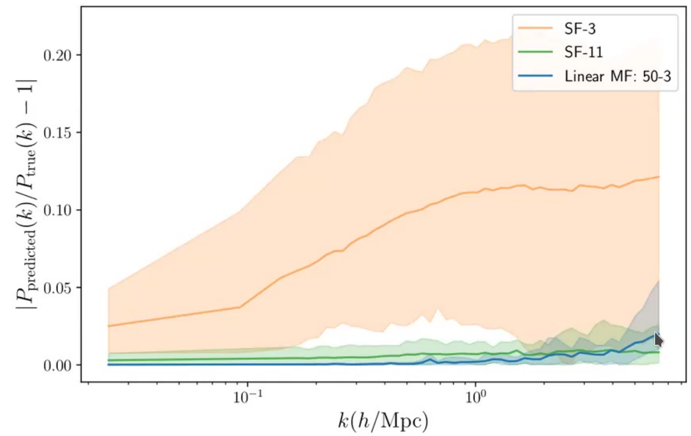
Optimize for rho and delta

1 HR GP, 1 LR GP, correction

Multi-Fidelity Emulation

Proof of concept emulator has 1% avg. accuracy

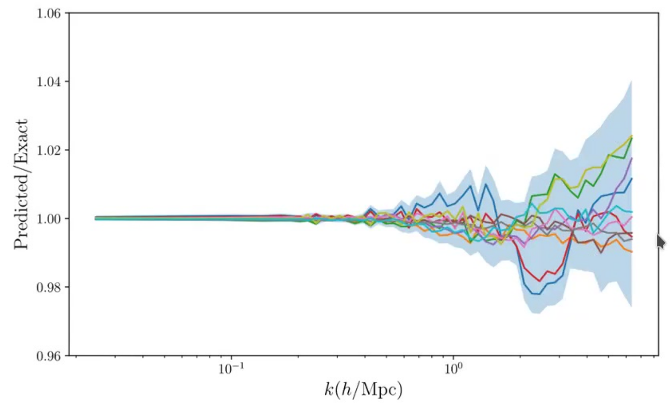
- 50 LR, 3 HR beats 11 HR
- Equivalent to 50 HR on large scales



Multi-Fidelity Emulation

Relative test error:

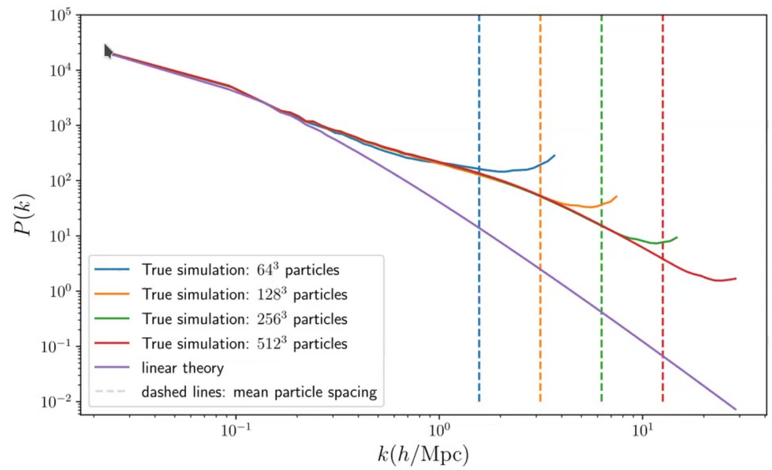
- Worst-case: 2%
- Probably dominated by
  our small HR sims



Multi-Fidelity Emulation

Works because of **halo model**

- 2-halo term learned by low fidelity simulations.
- Resolution correction is 1 halo term almost cosmology independent.



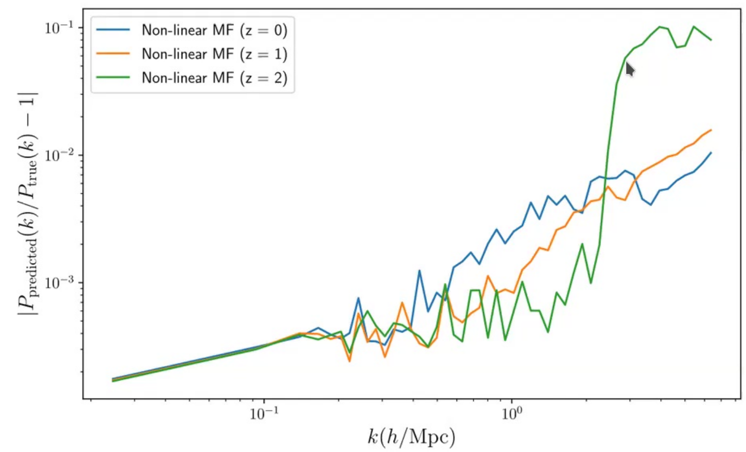
Multi-Fidelity Emulation

Good at z=0,1

Worse at z=2

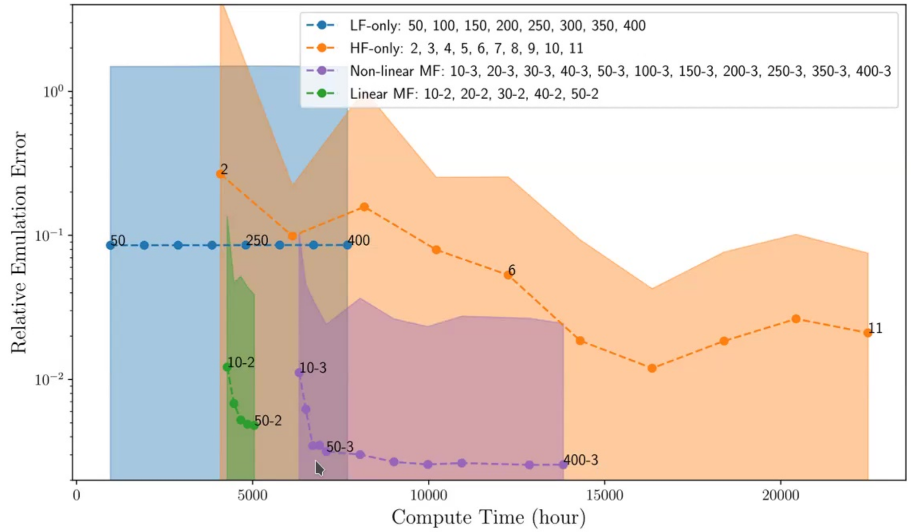
Why?

Low fidelity at z=2
is too linear



Multi-Fidelity Emulation

HR is 512, LR is 128: LR x4^3 cheaper



Conclusions

- An introduction to cosmological simulations and to ASTRID
- Machine Learning increases the effective scale of the simulation

- Parameter estimation trick: emulator from a few high resolution simulations with low resolution.

TODO: Do a hydro emulator.
TODO: Expand existing emulators with new parameters. eg: some non-trivial dark matter