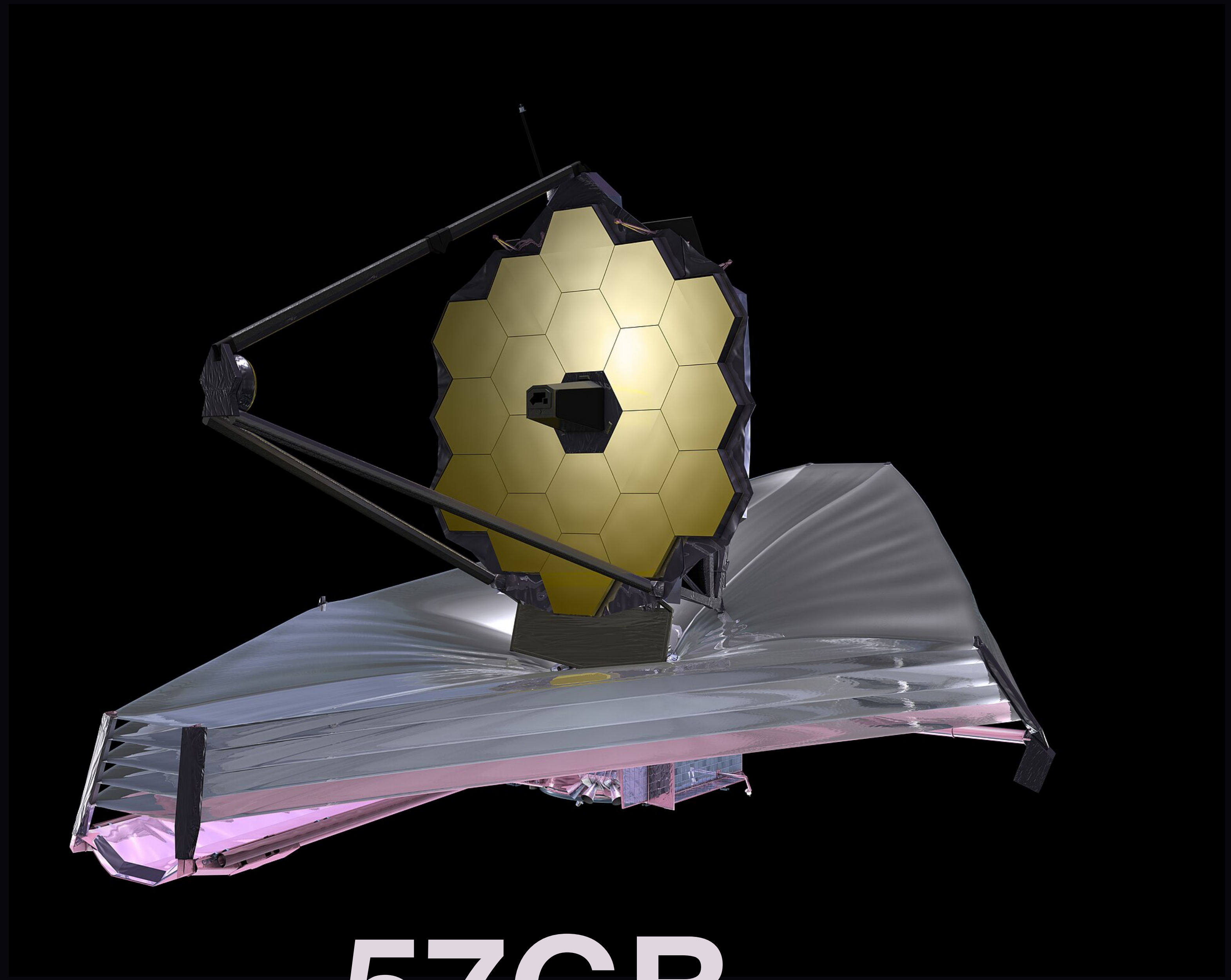
Research Examples — using AI for cosmology and galaxy formation studies

(how **not** to be used by **AI**)

Kentaro Nagamine

(U. Osaka / K-IPMU / UNLV)

`Big Data' Era in Cosmology



57GB

Daily data download from JWST

~500TB over lifetime

~20TB

/ day

3G pixel image / 30sec

~**100PB** over lifetime

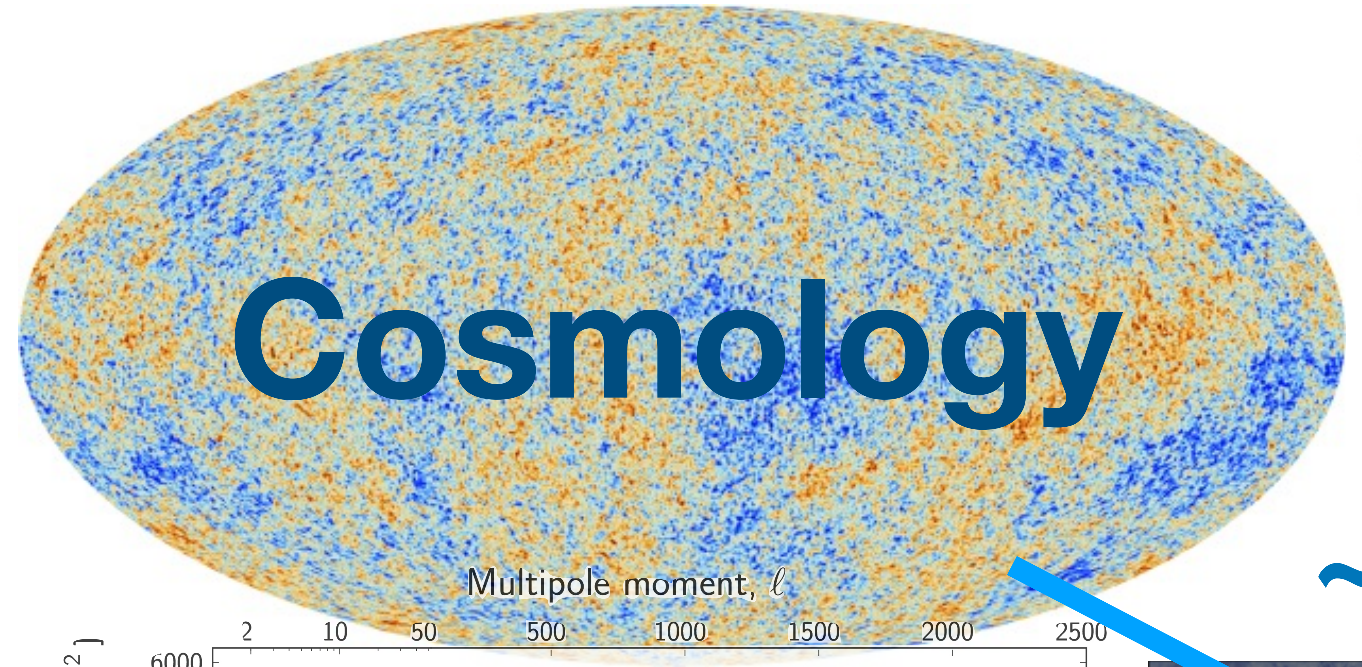
$\gtrsim 10^9$

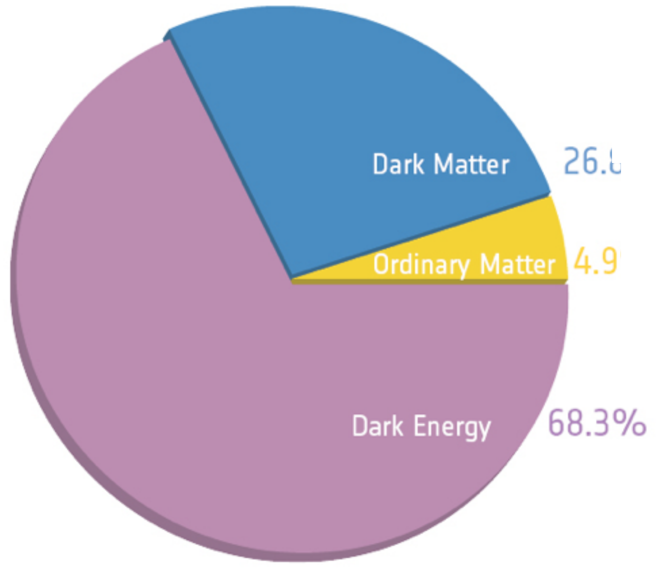
Galaxies

Observable in Euclid, Roman, ALMA, JWST, Subaru-PFS, SKA, etc.

- Massive datasets; difficult to handle.
- Sophisticated computational techniques required for analysis.
- Traditional methods cannot efficiently extract cosmological params.

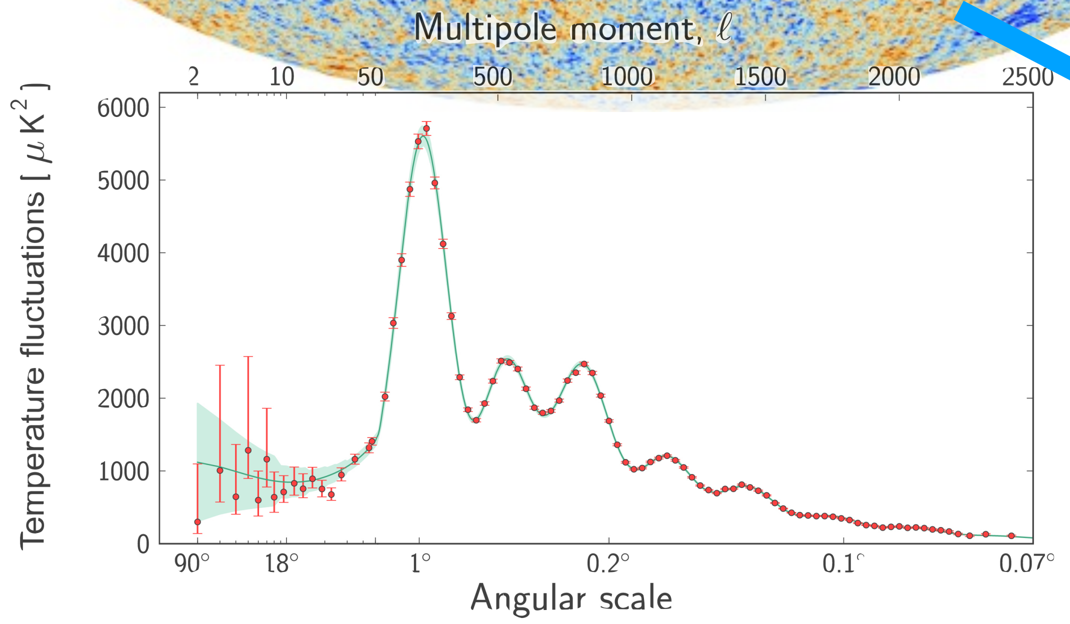
Cosmology





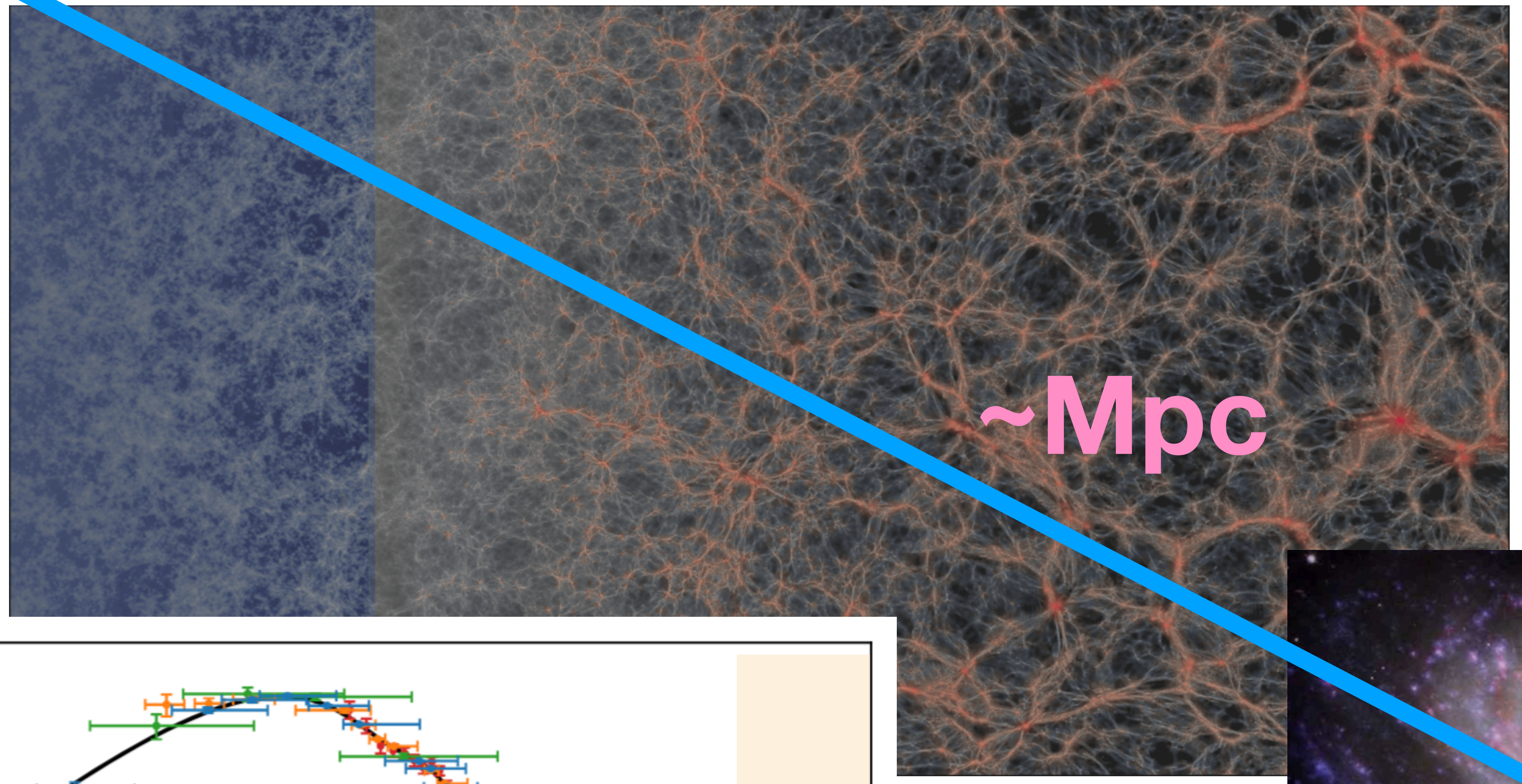
$t_H = H_0^{-1} = 9.78e9\, h^{-1}$ yr ~ 14 Gyr

$d_H = cH_0^{-1}$ ~ 3 Gpc



~Gpc

Large-scale Structure — Filaments, Nodes



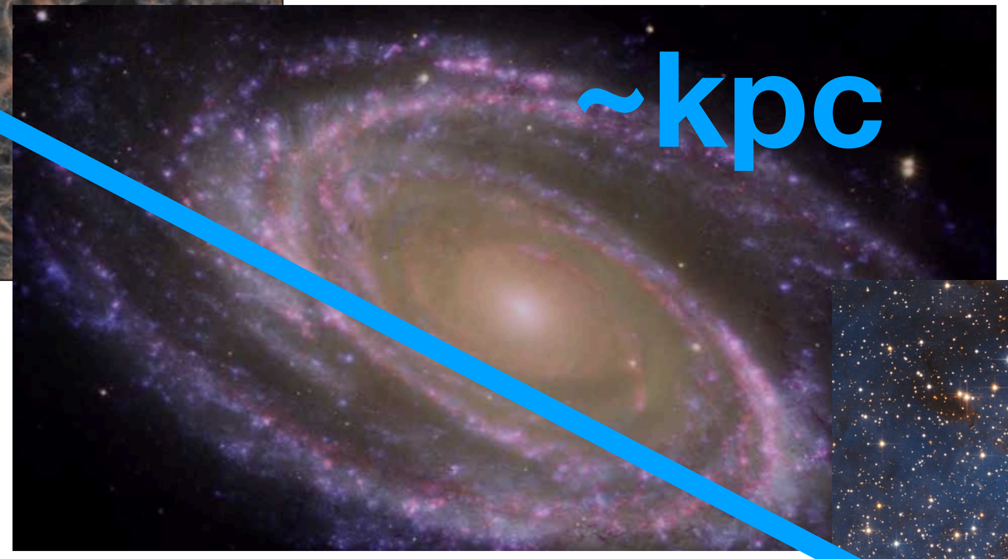
~Mpc

Galaxy Clusters

$$1000\,\mathrm{km/s} \times 10\,\mathrm{Gyr} = 10\,\mathrm{Mpc}$$

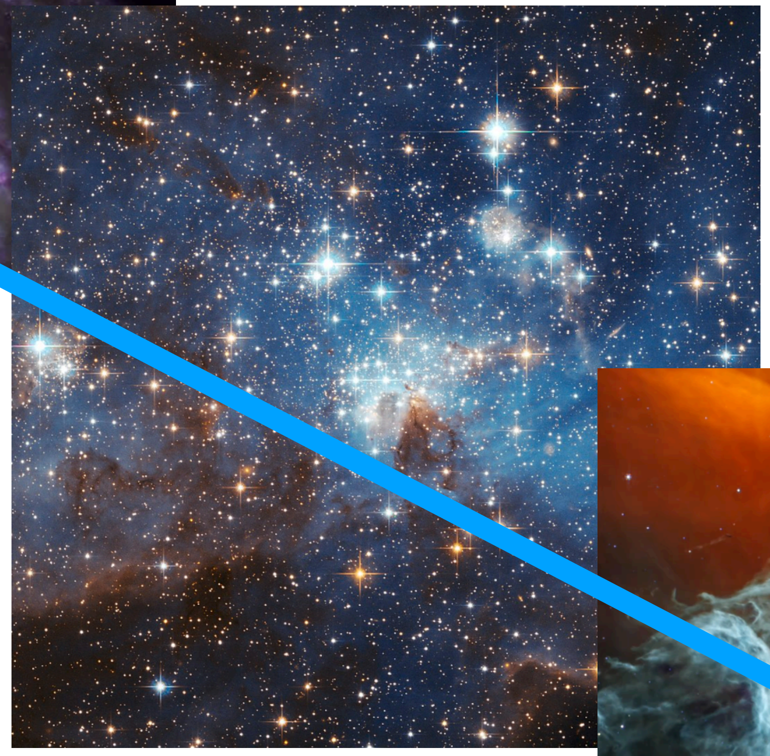
Galaxies

$$100\,\mathrm{km/s} \times 1\,\mathrm{Gyr} = 100\,\mathrm{kpc}$$



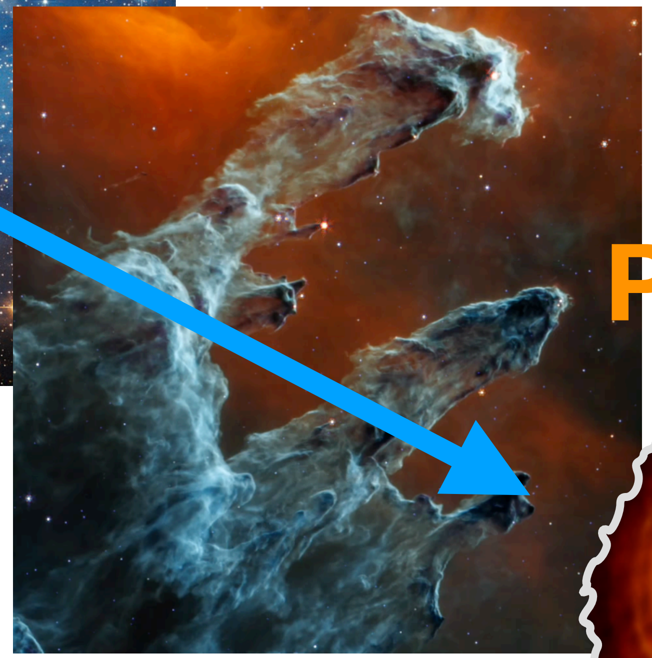
~kpc

Molecular clouds



~pc

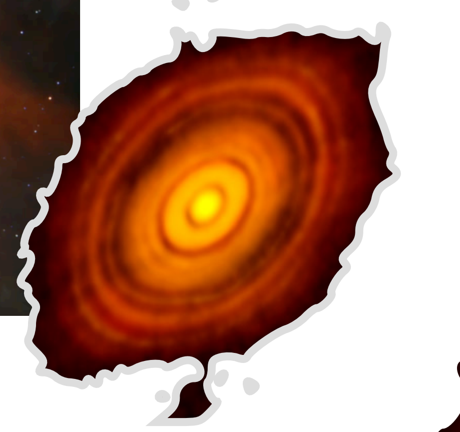
$$1\,\mathrm{km/s} \times 1\,\mathrm{Myr} = 1\,\mathrm{pc}$$



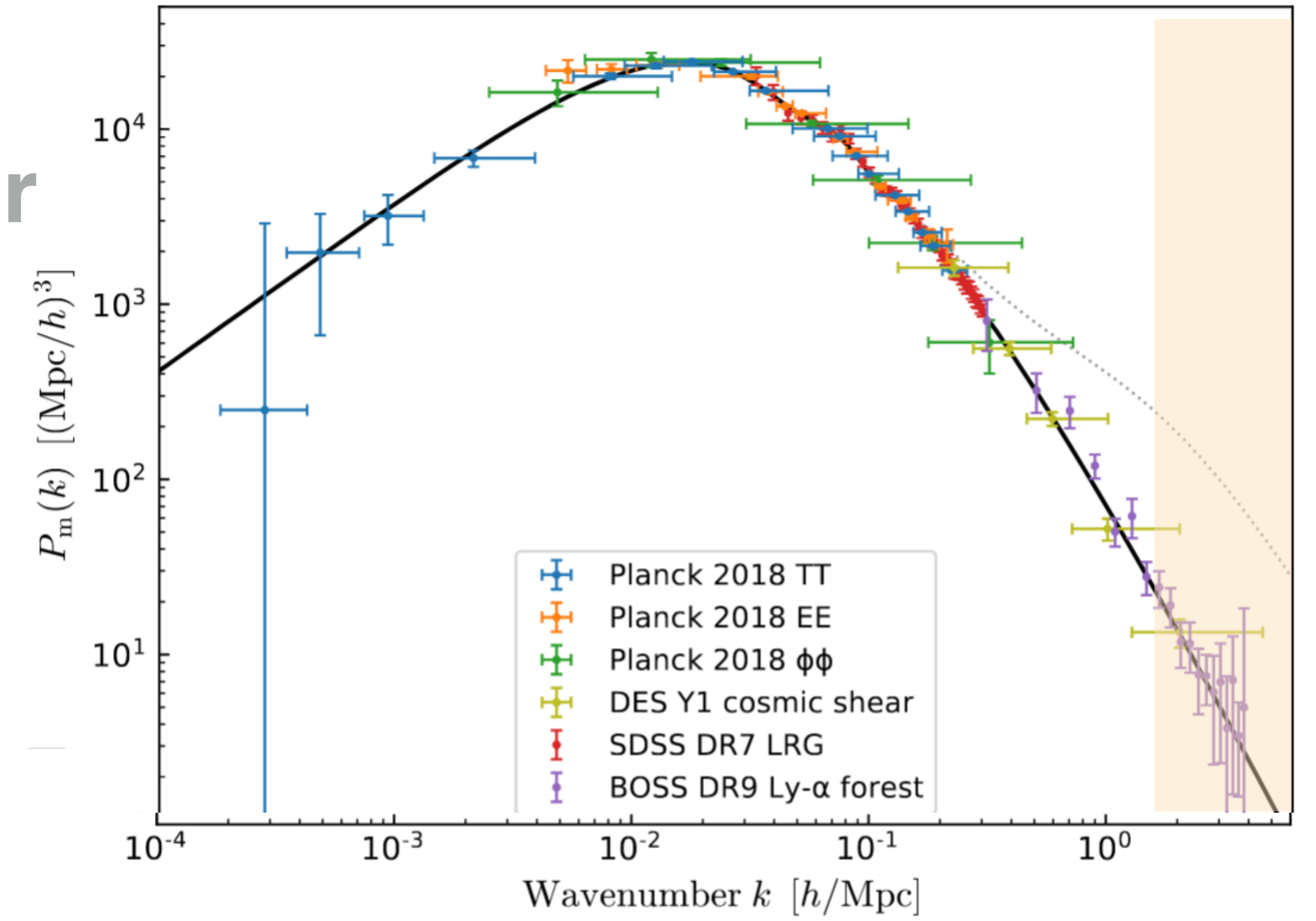
Star-forming region

AU

Planets

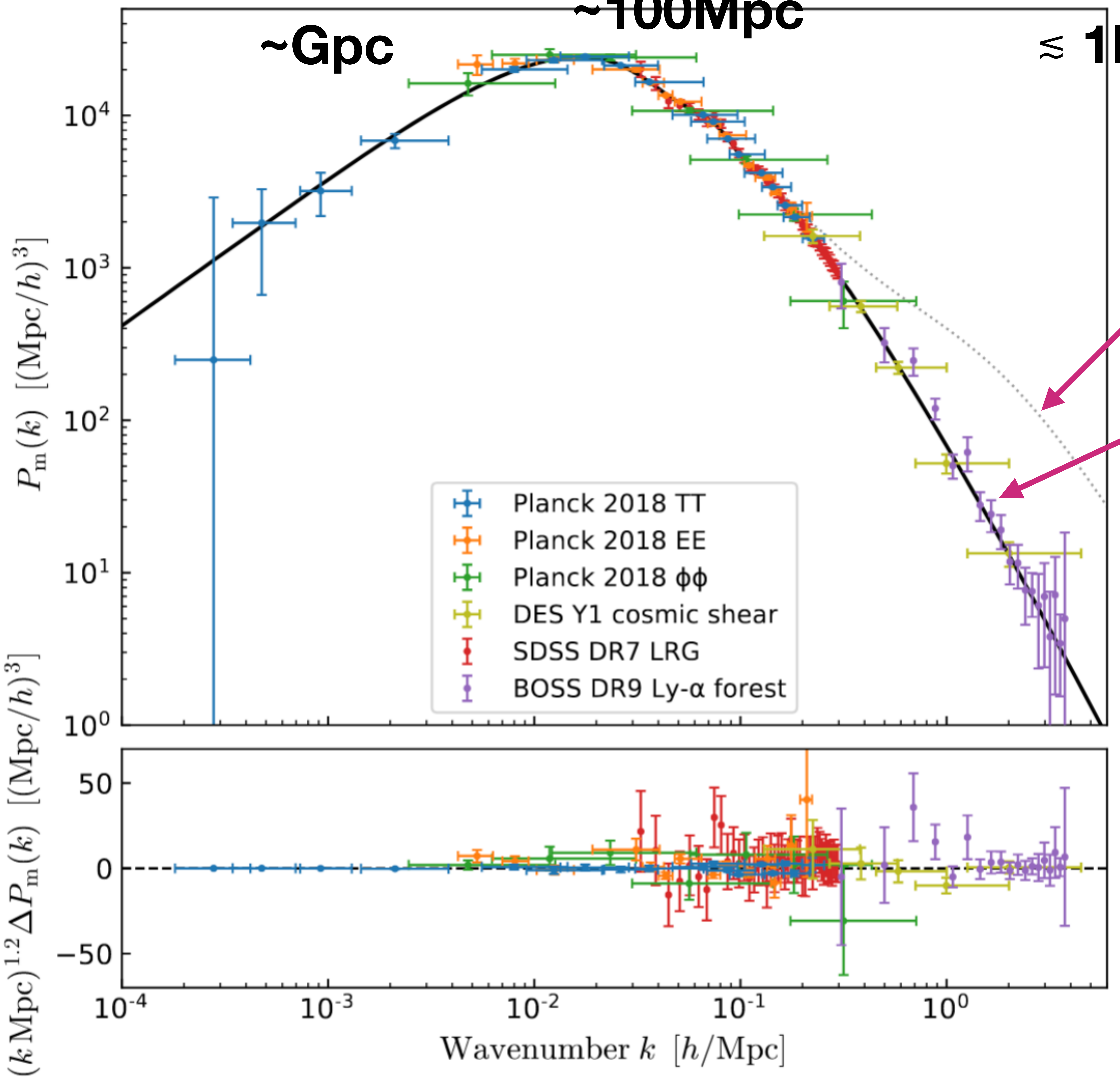


Matter power spectrum



wavenumber

~Gpc ~100Mpc ≲ 1Mpc



nonlinear P(k)

linear P(k)

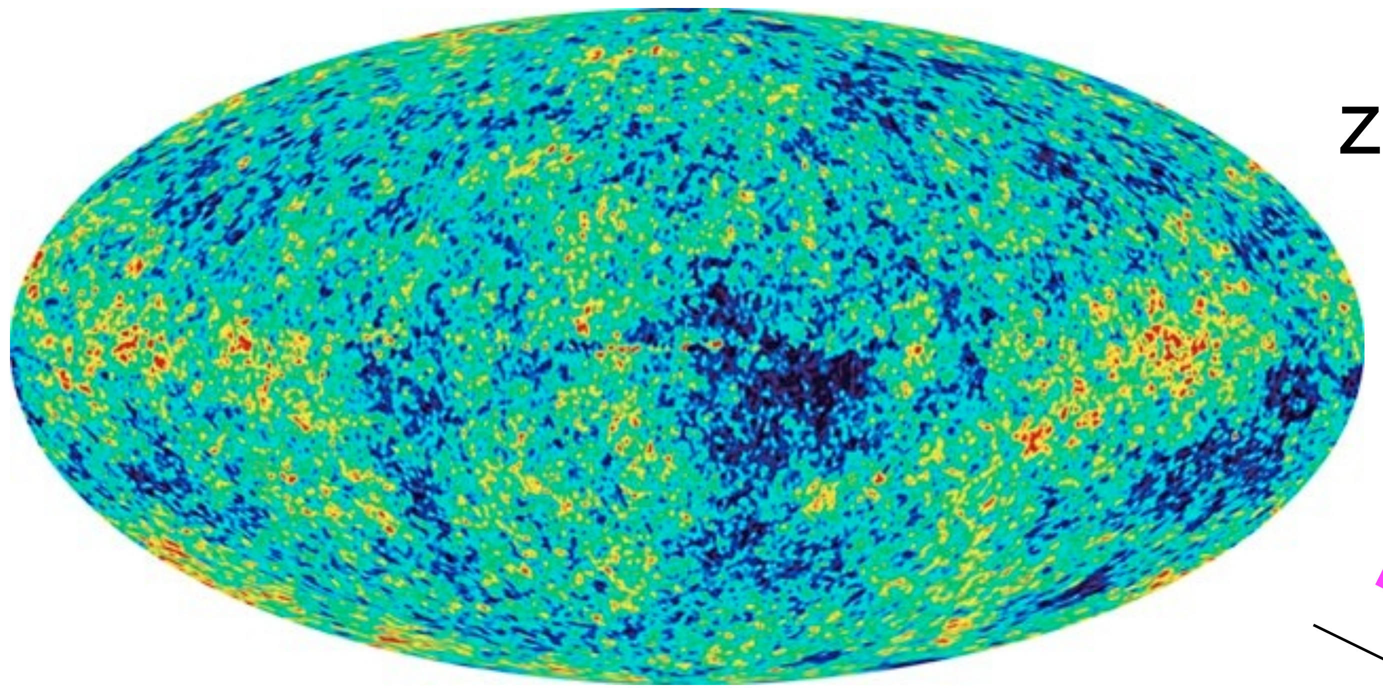
Continued support for ΛCDM

Chabanier+'19

Forward Modeling by “Computational Cosmology”

— started in the ‘90s.

Self-consistent galaxy formation scenario from first principles



z~1100

Initial conditions

Cosmological params,
Dark Energy, Dark Matter,
Baryons
(+ expanding universe)

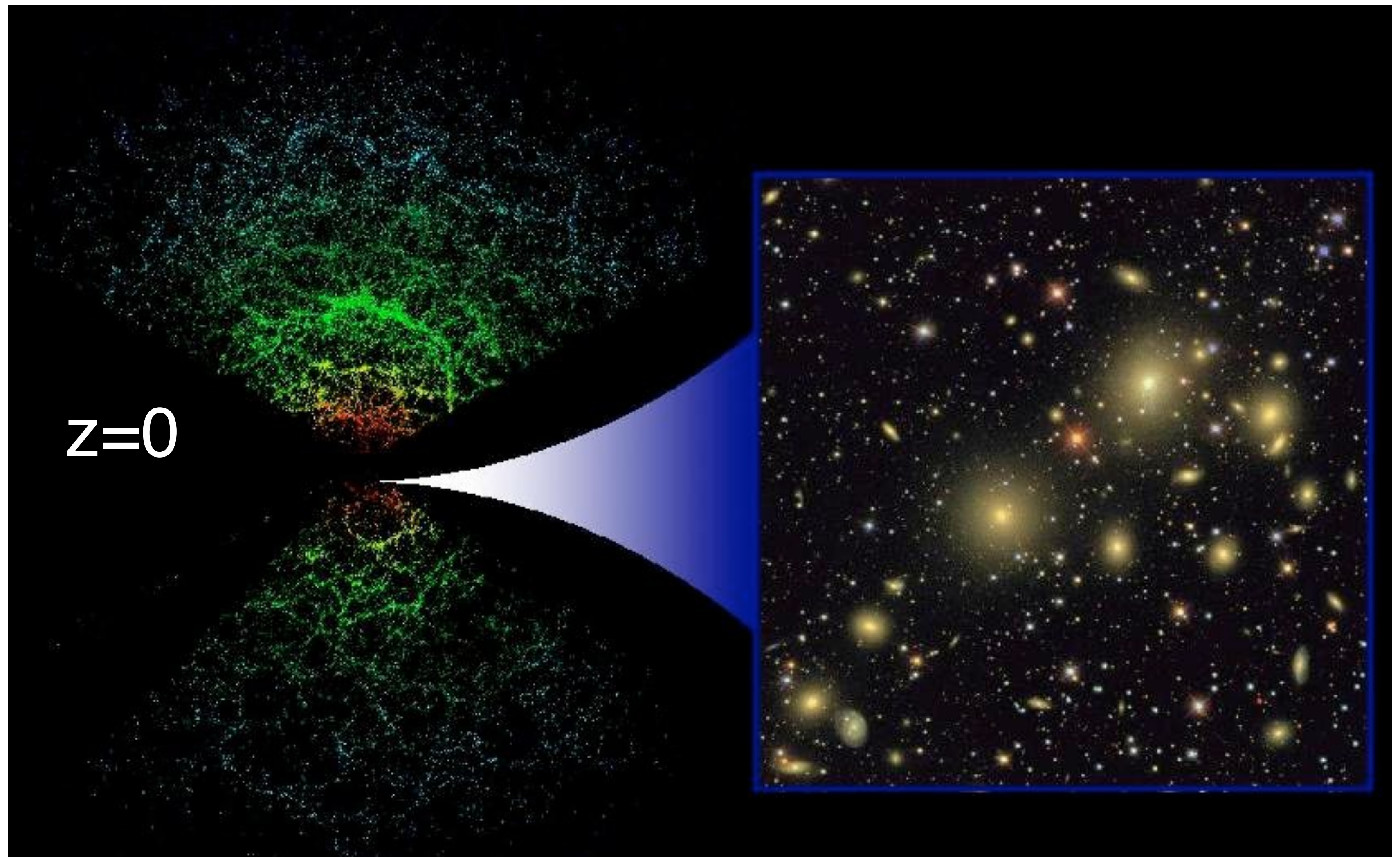
z=100

z=10

z=3

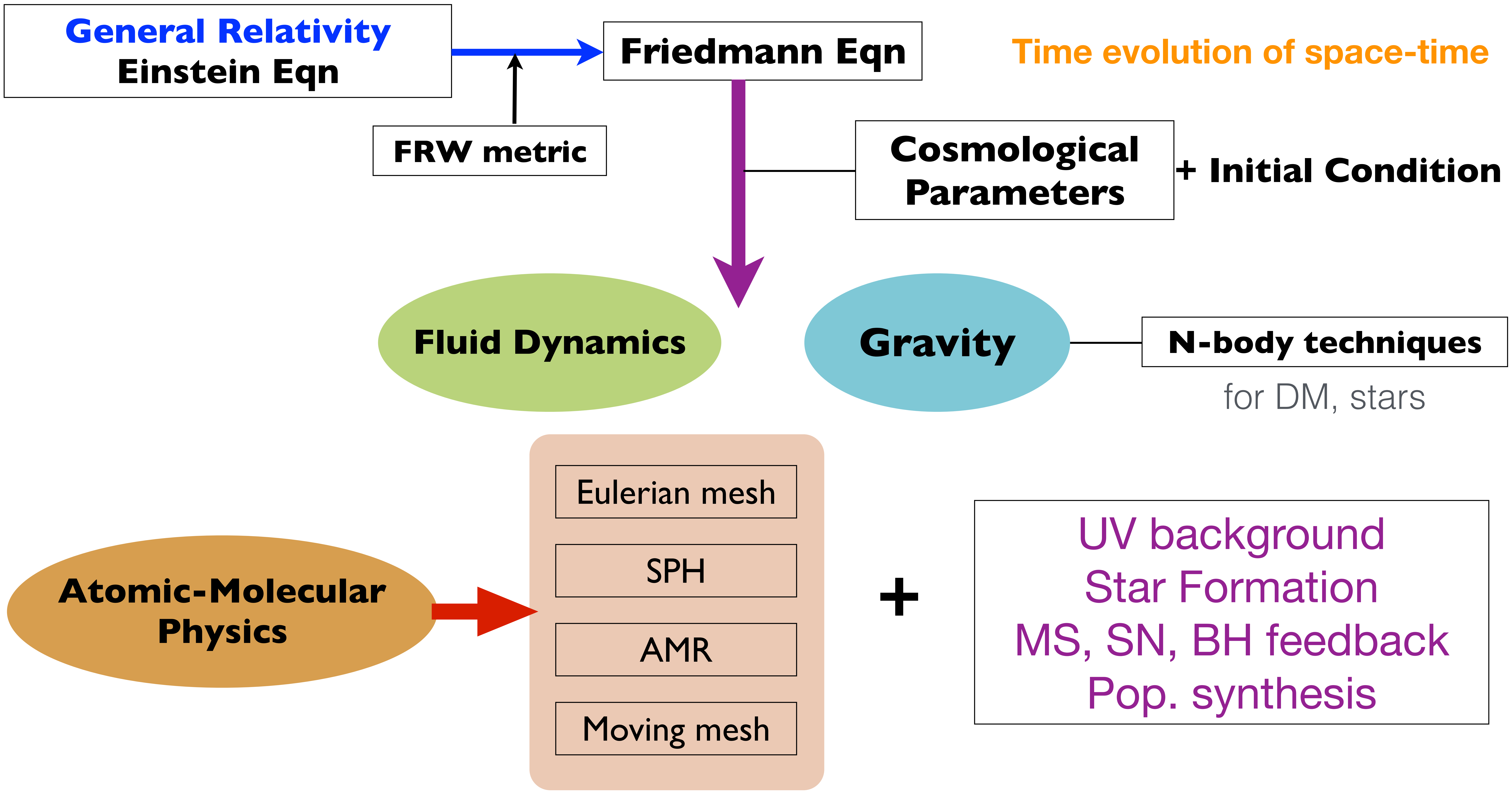
Baryonic physics

Radiative heating/cooling,
Star formation,
& Feedback



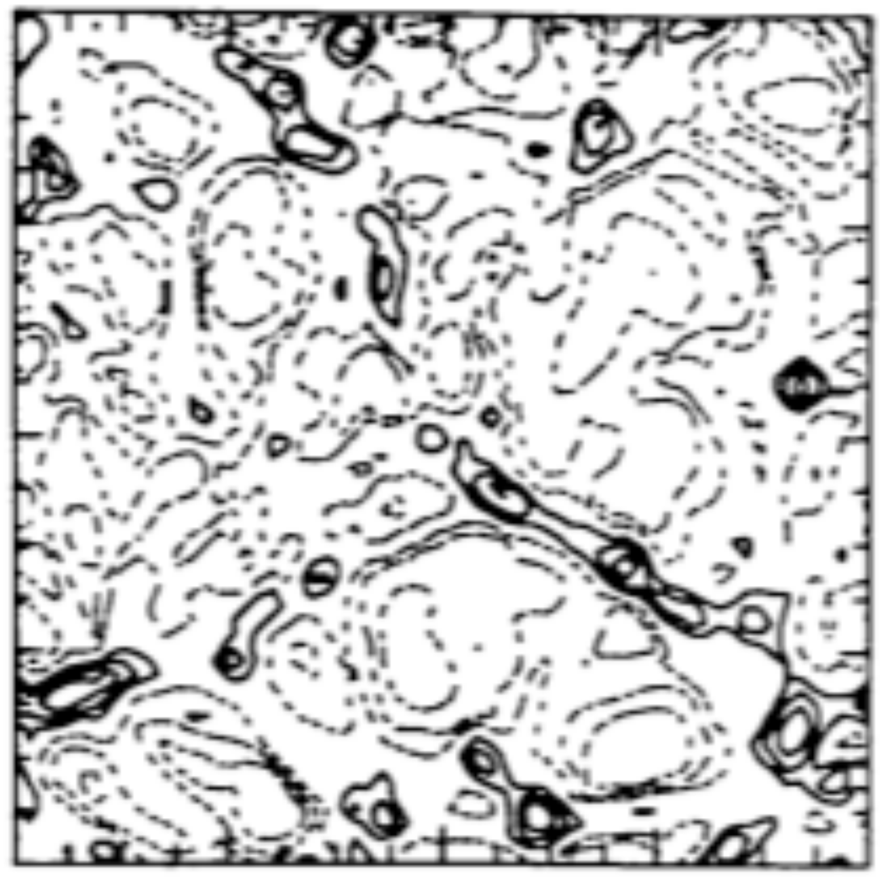
Gravity + Hydrodynamics

Framework of Computational Cosmology



Three Revolutions in Cosmo. Hydro Sims

1990': 1st Revolution

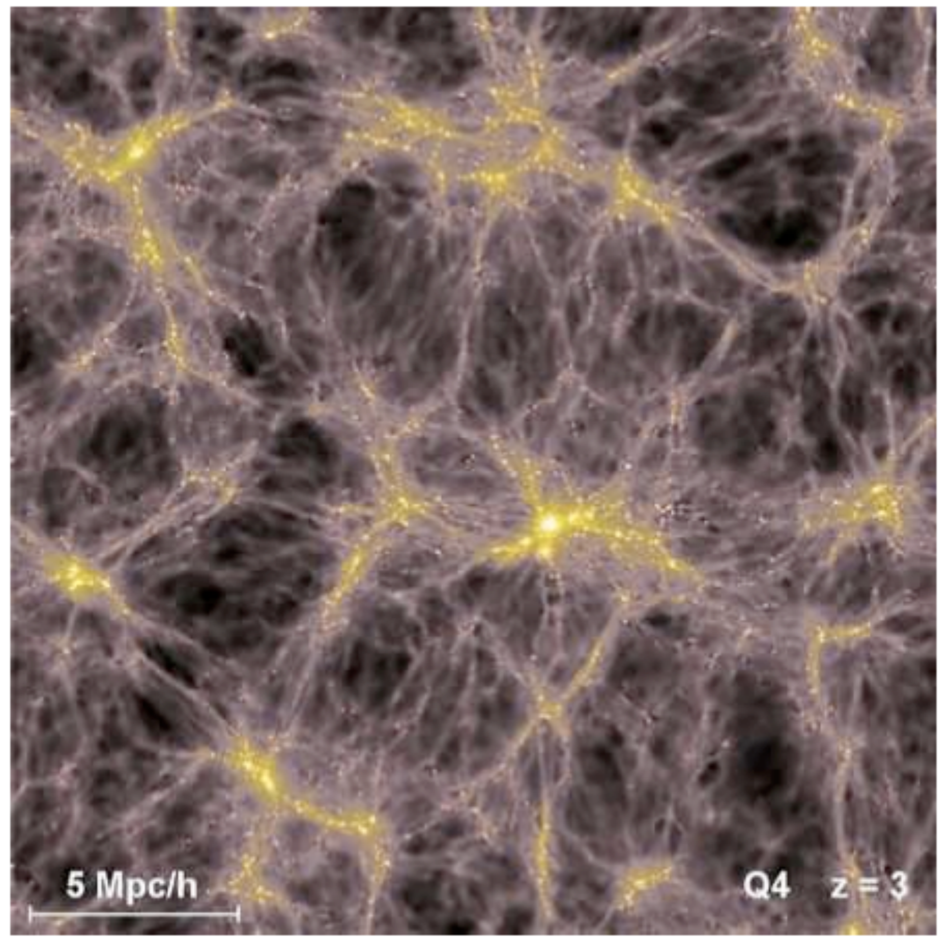


First cosmological, but coarse calculation

Resolution ~100 kpc

e.g. Cen, Ostriker '92-'93
Katz+ '96

2001-2011 2nd Rev.

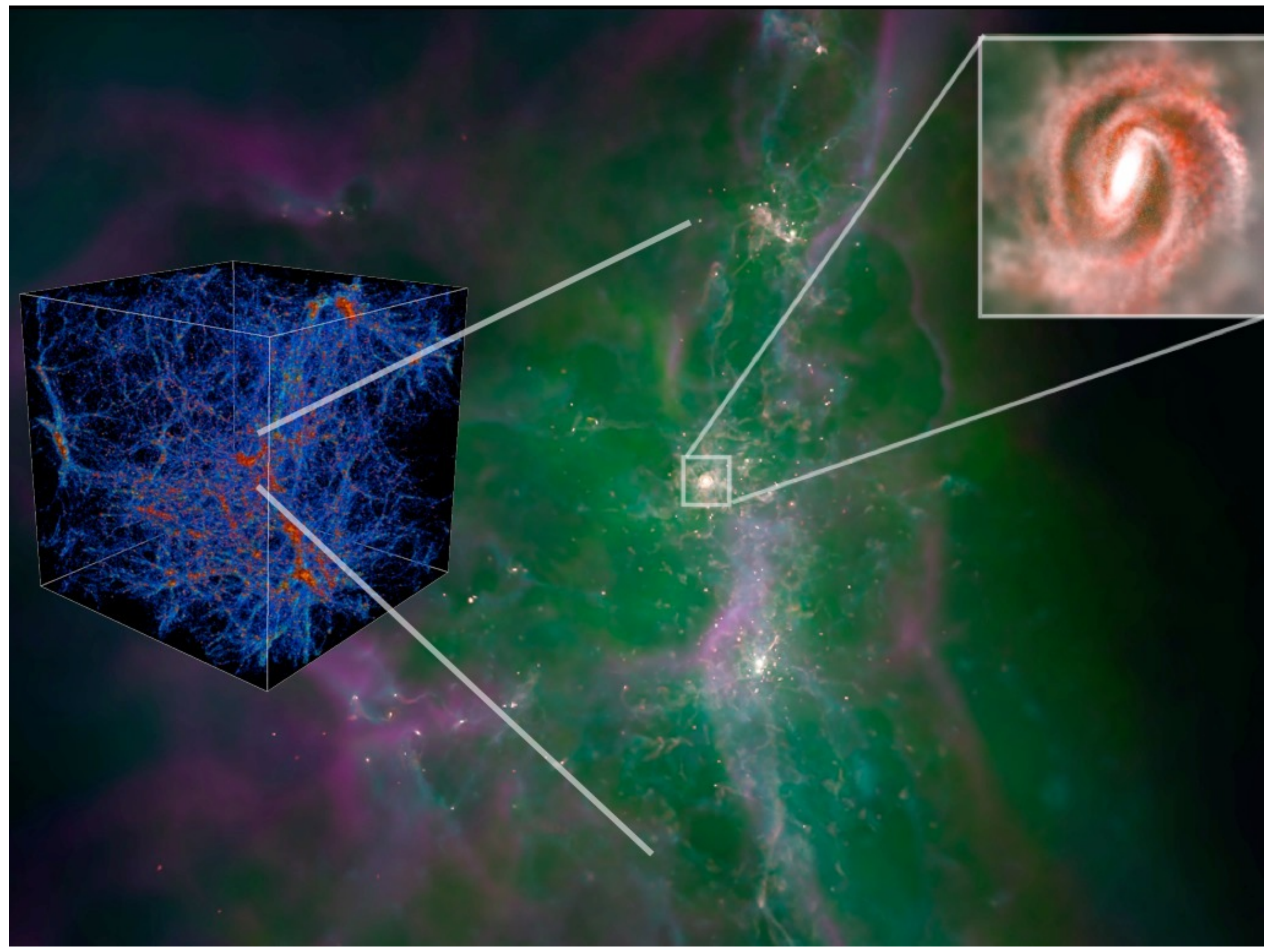


Larger scale, medium resolution
w. subgrid models

Resolution ~ kpc

e.g. Springel & Hernquist '03
KN+ '01, '04, '06

2012~ 3rd Rev.



Zoom-in method allows much higher res.

Resolution~ 10-100pc → **< 3pc**

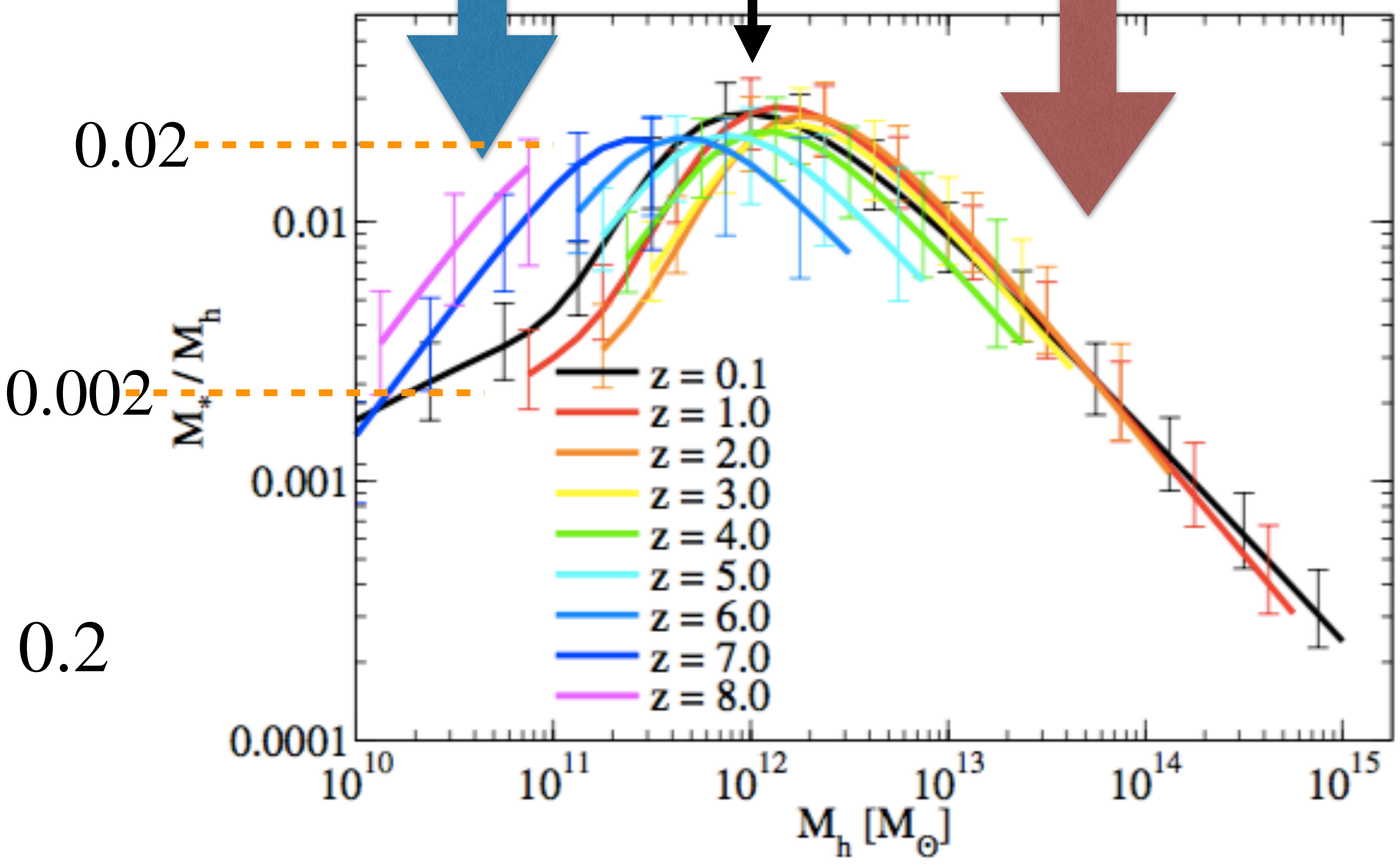
IC code: GRAFIC (Bertschinger)
MUSIC (Hahn & Abel '11)

Stellar-to-Halo Mass Ratio (SHMR)

SN Feedback

~$10^{12}M_{\odot}$

AGN Feedback?



Abundance Matching result

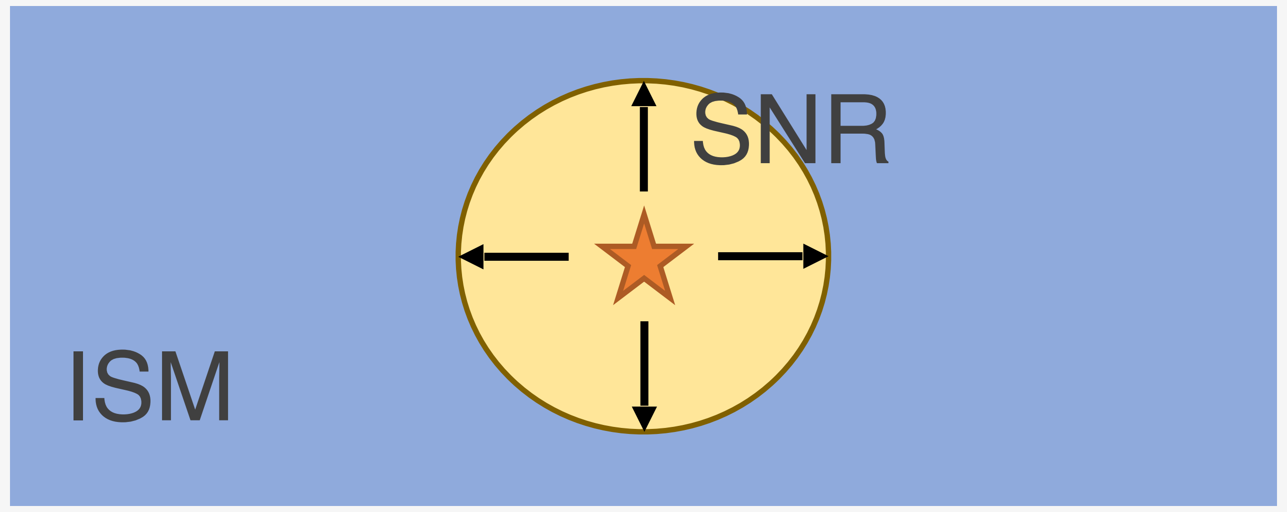
cf.

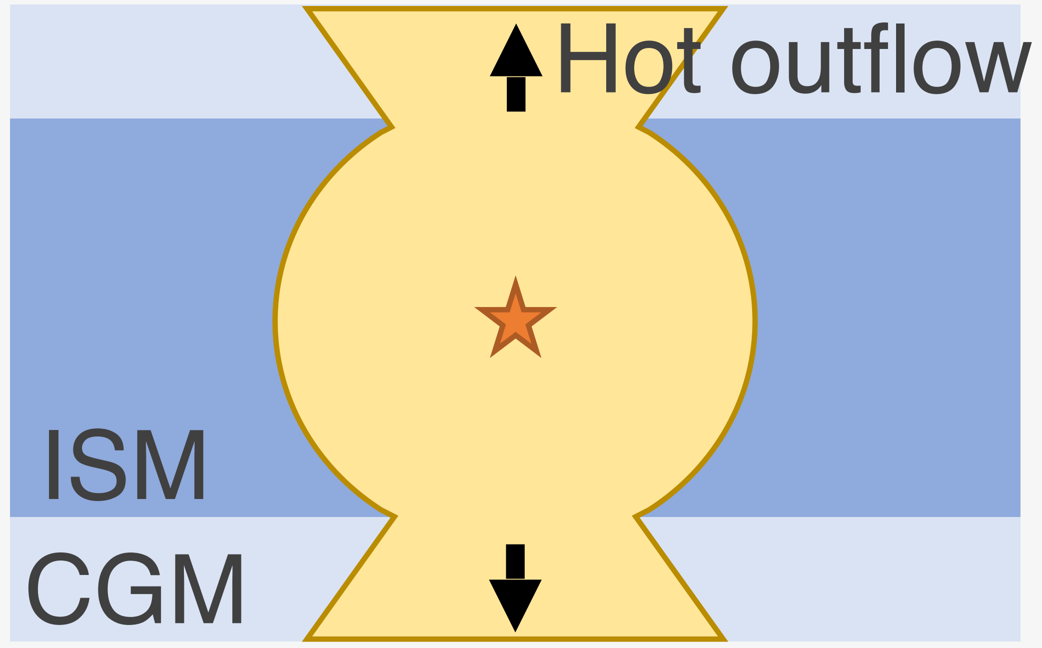
$$\frac{\Omega_b}{\Omega_{DM}} \sim \frac{0.045}{0.22} \sim 0.2$$

Behroozi+'10, '13, '18

(cf. Ilbert+'10; George+'11; Leauthaud+'12)

Two modes of SN Feedback: Kinetic & Thermal





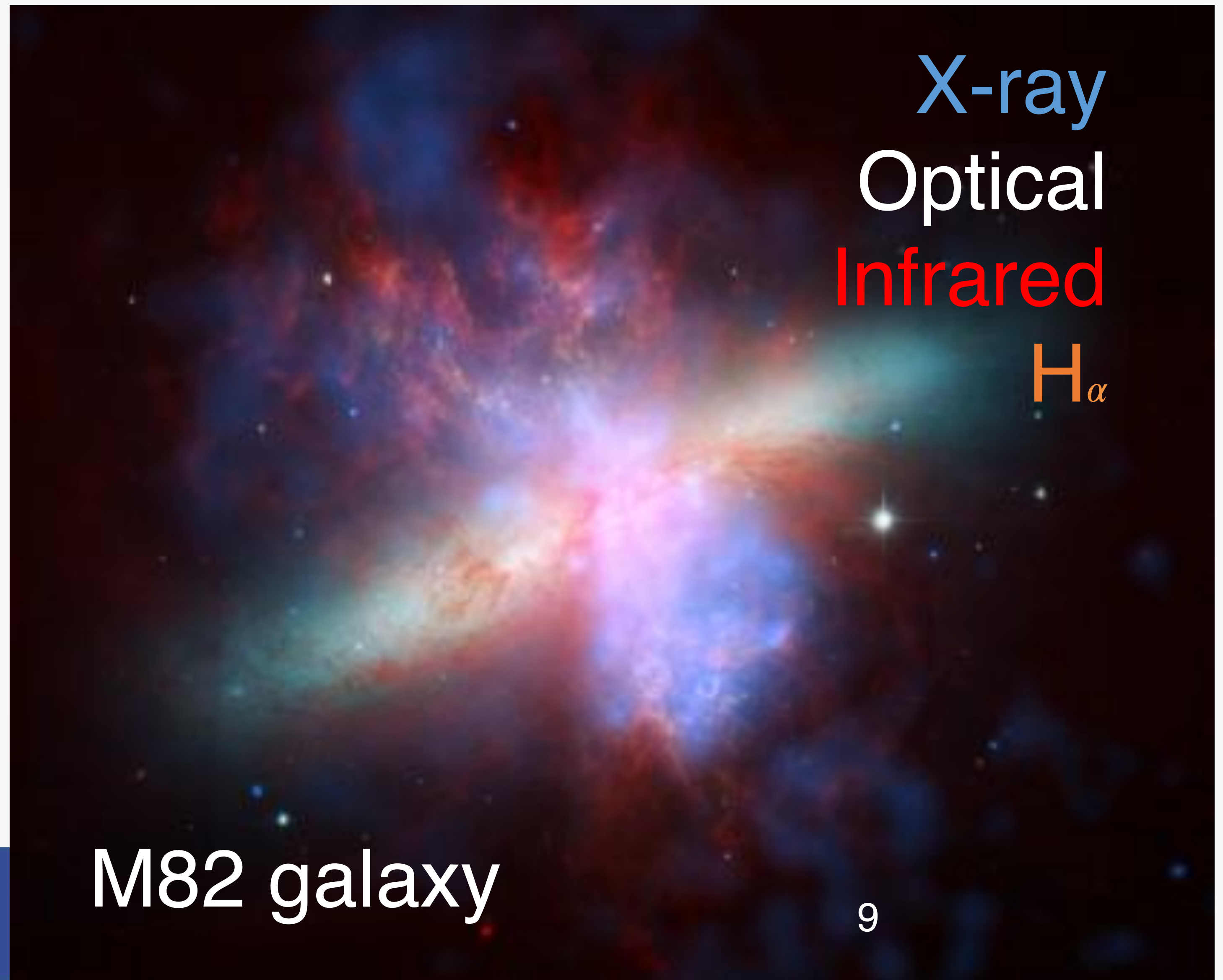
Kinetic FB

- SNR gives momentum to ISM
- Drive **turbulence**

Thermal FB

- Shock heating by SN ejecta
- Produce **hot outflow**
- **Transport metals** out from the galaxy

- The Adiabatic phase (Sedov-Taylor phase) is unresolved in large-scale cosmological simulation
- Explicit modeling of galactic wind is necessary to produce hot galactic wind



Historical Flow Chart of SN Feedback Treatment

1st phase '90s

Simple thermal feedback

↓ Overcooling problem

2nd phase
'00-'10

Phenomenological subgrid model (thermal + kinetic)
based on galactic properties (SFR, M_{star}, M_{halo})

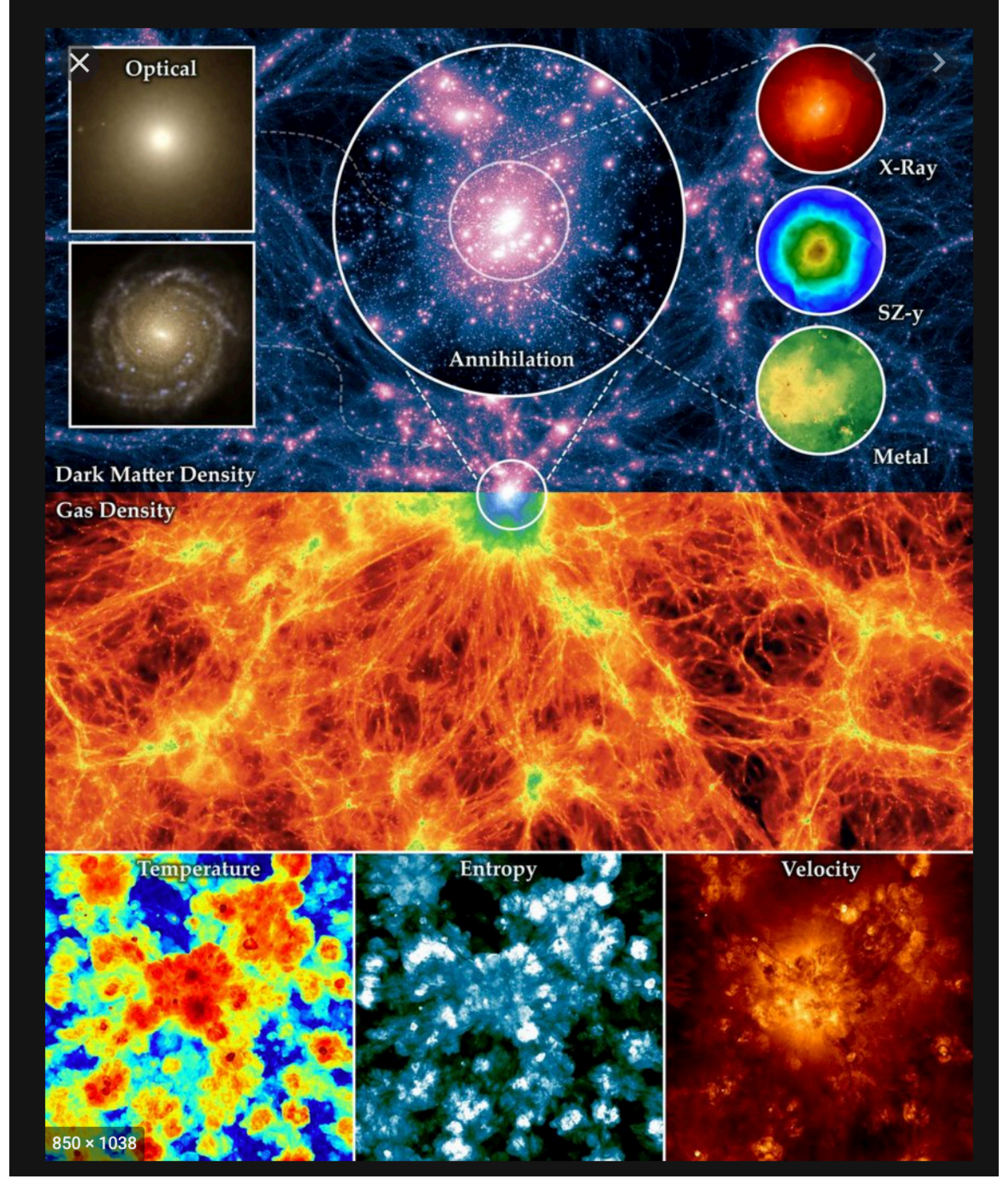
Multi-phase ISM model

↓

3rd phase
'10~
(w/ zoom sim)

More direct, local,
thermal + kinetic + radiative feedback

AGN feedback models



IllustrisTNG

AREPO

- **Two-mode AGN feedback model**

Eddington-limited accretion: $\dot{M} = \min\left(\dot{M}_{\rm Bondi}, \dot{M}_{\rm Edd}\right)$,

$$\dot{M}_{\rm Bondi} = \frac{4\pi G^2 M_{\rm BH}^2 \rho}{c_s^3}, \qquad \dot{M}_{\rm Edd} = \frac{4\pi G M_{\rm BH} m_p}{\epsilon_r \sigma_{\rm T}} c,$$

- **Eddington ratio threshold**

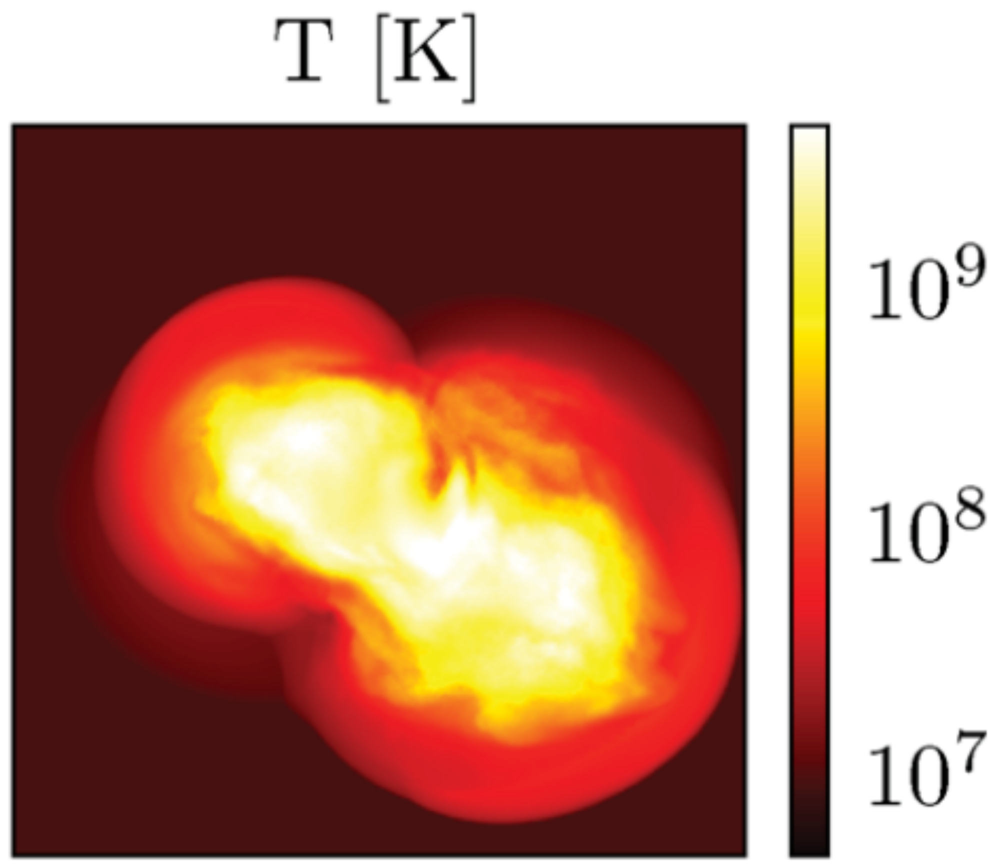
$$\chi = \min\left[0.002\left(\frac{M_{\rm BH}}{10^8\,{\rm M}_\odot}\right)^2, 0.1\right],$$

high

$\dot{E}_{\rm therm} = 0.02\,\dot{M}\,c^2,$ **thermal (quasar) mode**

low

$\dot{E}_{\rm kin} = \epsilon_{f,\rm kin}\,\dot{M}\,c^2,$ **kinetic (jet) mode** (maintenance mode)



$$\epsilon_{f,\rm kin} = \min\left(\frac{\rho}{0.05\,\rho_{\rm SFthresh}}, 0.2\right),$$

weaker coupling in low-ρ environment

Weinberger+'18

Angular Momentum / Torque model

- **EAGLE** (GADGET-3)

$$\dot{M}_{\rm BH} = \min(\dot{M}_{\rm Bondi} \times \underline{\min\left((c_{\rm s}/V_\Phi)^3/C_{\rm visc}, 1\right)}, \dot{M}_{\rm Edd}),$$

Rosas-Guevara+15,16

V_Φ : average circular speed of gas around BH

C_{visc} : free param for viscosity of subgrid accretion disk

$$\epsilon_r \epsilon_f = 0.1 \times 0.15 = 0.015.$$

Stochastic thermal heating only when sufficient to raise to $\Delta T = 10^{8.5}\,K$

- **SIMBA** (Gizmo)

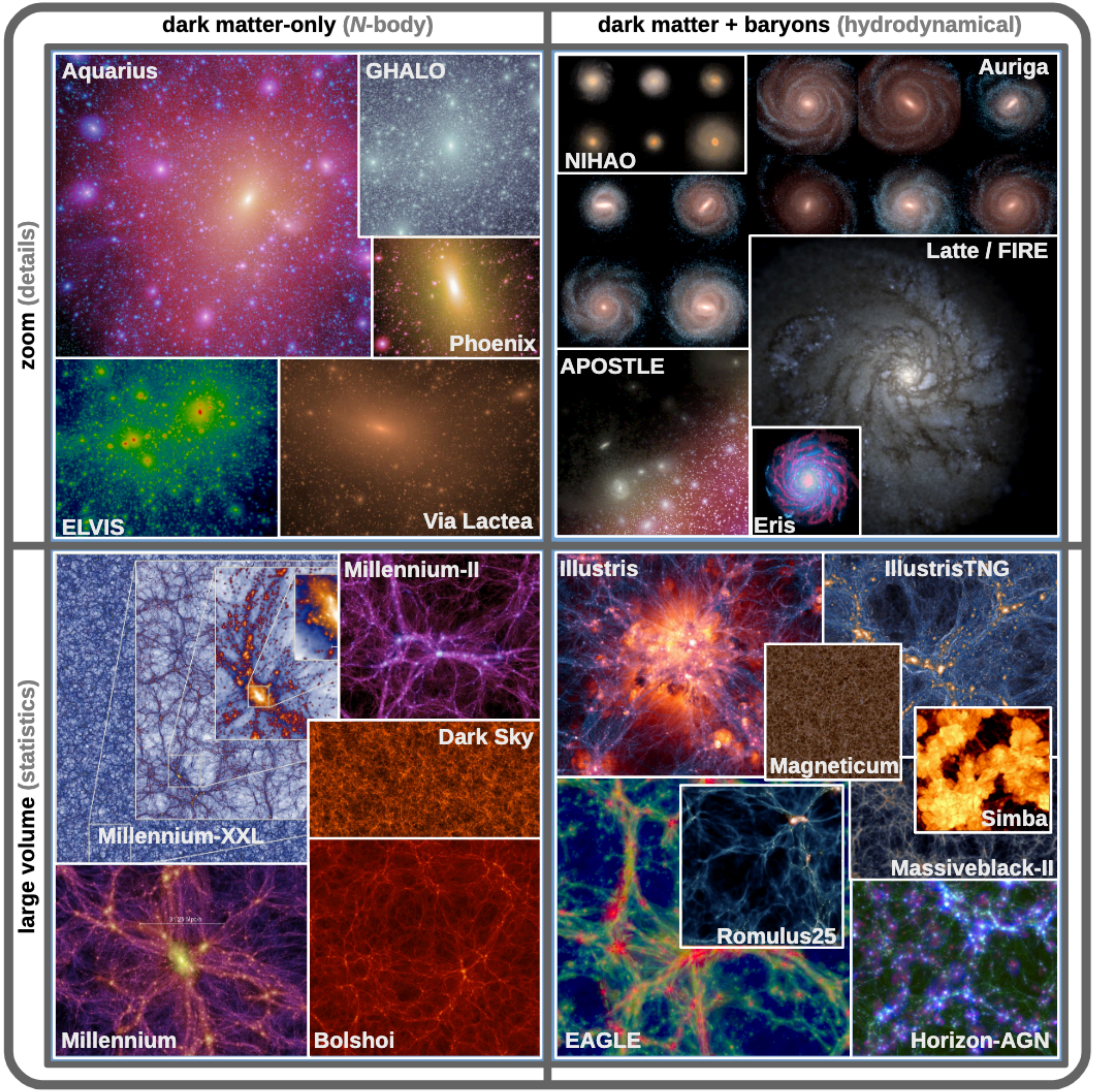
$$\dot{M}_{\rm BH} = (1-\epsilon_{\rm r}) \left[\min(\dot{M}_{\rm Bondi}, \dot{M}_{\rm Edd}) + \underline{\min(\dot{M}_{\rm Torque}, 3\,\dot{M}_{\rm Edd})}\right],$$

$$\dot{M}_{\rm Torque} = \epsilon_{\rm T}\, f_{\rm d}^{5/2} \times \left(\frac{M_{\rm BH}}{10^8\,{\rm M}_\odot}\right)^{1/6} \times \left(\frac{M_{\rm enc}(R_0)}{10^9\,{\rm M}_\odot}\right) \times \left(\frac{R_0}{100\,{\rm pc}}\right)^{-3/2} \times \left(1+\frac{f_0}{f_{\rm gas}}\right)^{-1} {\rm M}_\odot/{\rm yr},$$

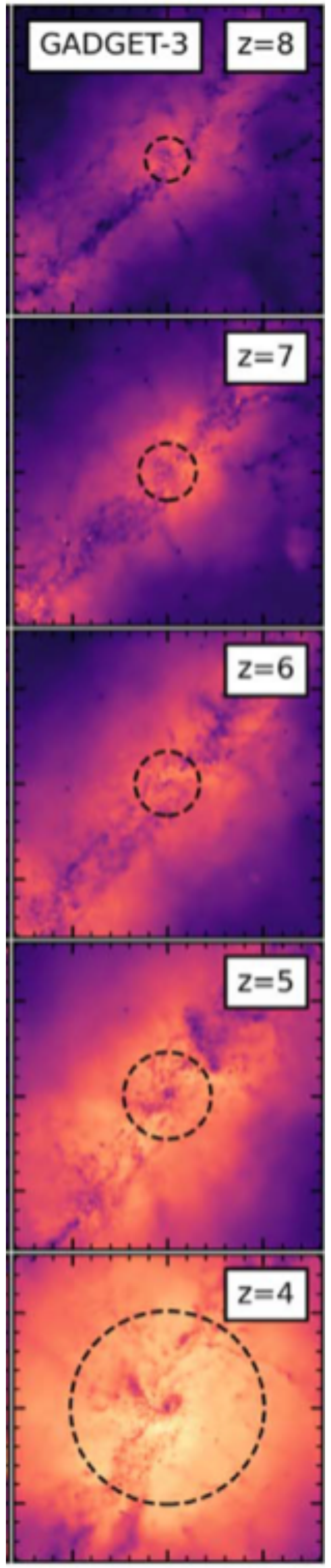
gas inflow rate driven by grav. instabilities from galactic to the accretion disk scale, within $R_0 = 2\,h^{-1}$ kpc (Hopkins & Quataert '10; Angl´es-Alc´azar+ '15,'17).

Table 1. List of common parameters for subgrid physics

Parameter	Adopted value	Description
$n_{\rm thres}$	$0.1\,{\rm cm}^{-3}$	Lower density threshold to allow star formation
$T_{\rm thres}$	$10^4\,{\rm K}$	Upper temperature threshold to allow star formation
ϵ_*	0.01	Star formation efficiency
$n_{\rm spawn}$	2	Maximum number of stars spawned from one gas particle
$N_{\rm ngb,fb}$	$8{\pm}2$	Number of gas particles subject to feedback
$n_{\rm event,snii}$	2	SNII feedback event number
$n_{\rm event,snia}$	8	SNIa feedback event number
$n_{\rm event,agb}$	8	AGB feedback event number
$M_{\rm seeding,FoF}$	$10^{10}\,h^{-1}\,{\rm M}_\odot$	FoF mass threshold to seed BH
$M_{\rm seeding,star}$	$10^{8}\,h^{-1}\,{\rm M}_\odot$	FoF stellar mass threshold to seed BH
ϵ_r	0.1	Radiative efficiency of BH
$\epsilon_{\rm FB}$	0.15	AGN feedback efficiency
$\Delta T_{\rm AGN}$	$10^{8.5}\,{\rm K}$	AGN feedback temperature
$C_{\rm visc}$	~ 10	Torque parameter of BH accretion

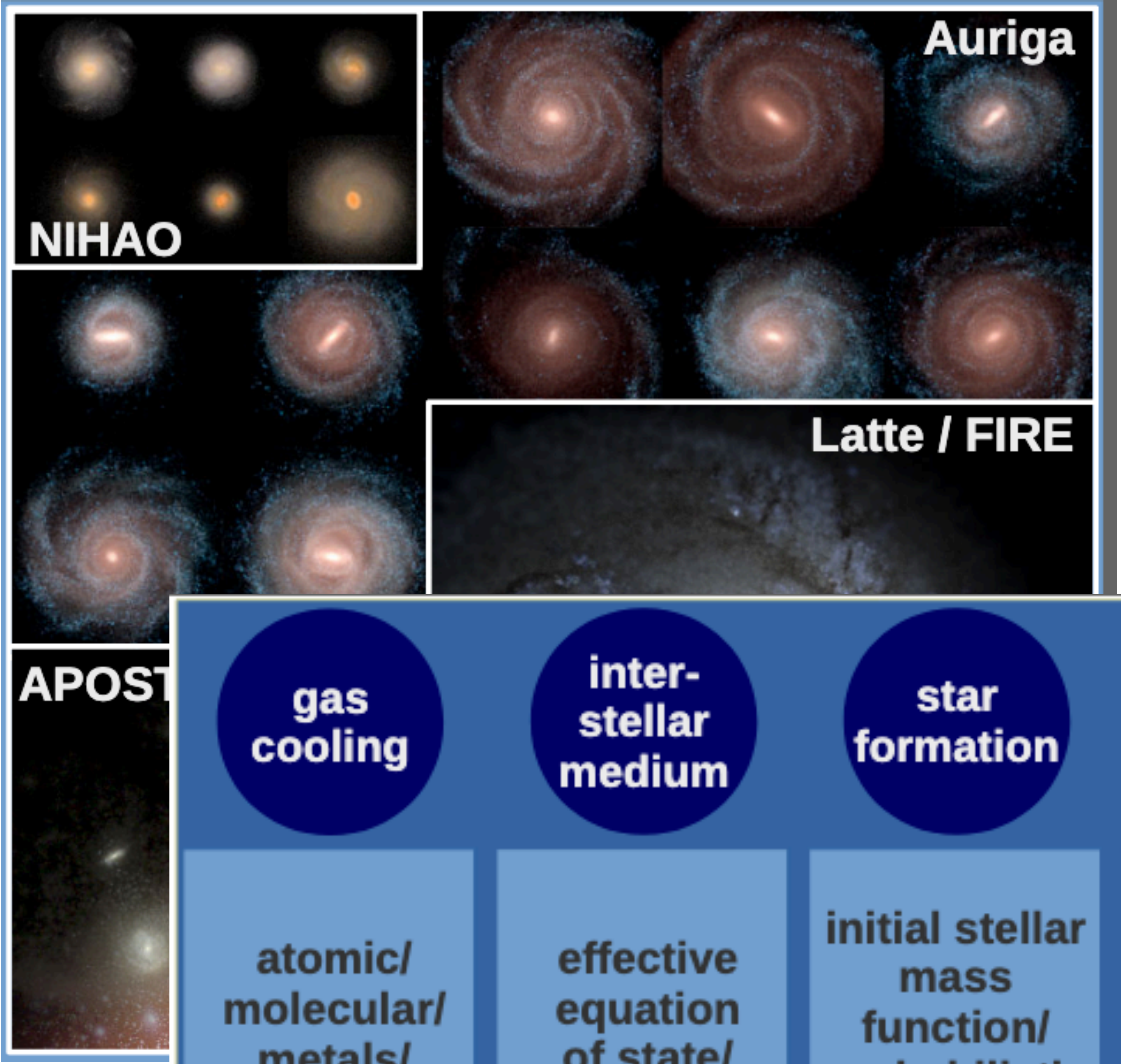


Vogelsberger+'19



AGORA
code comparison project

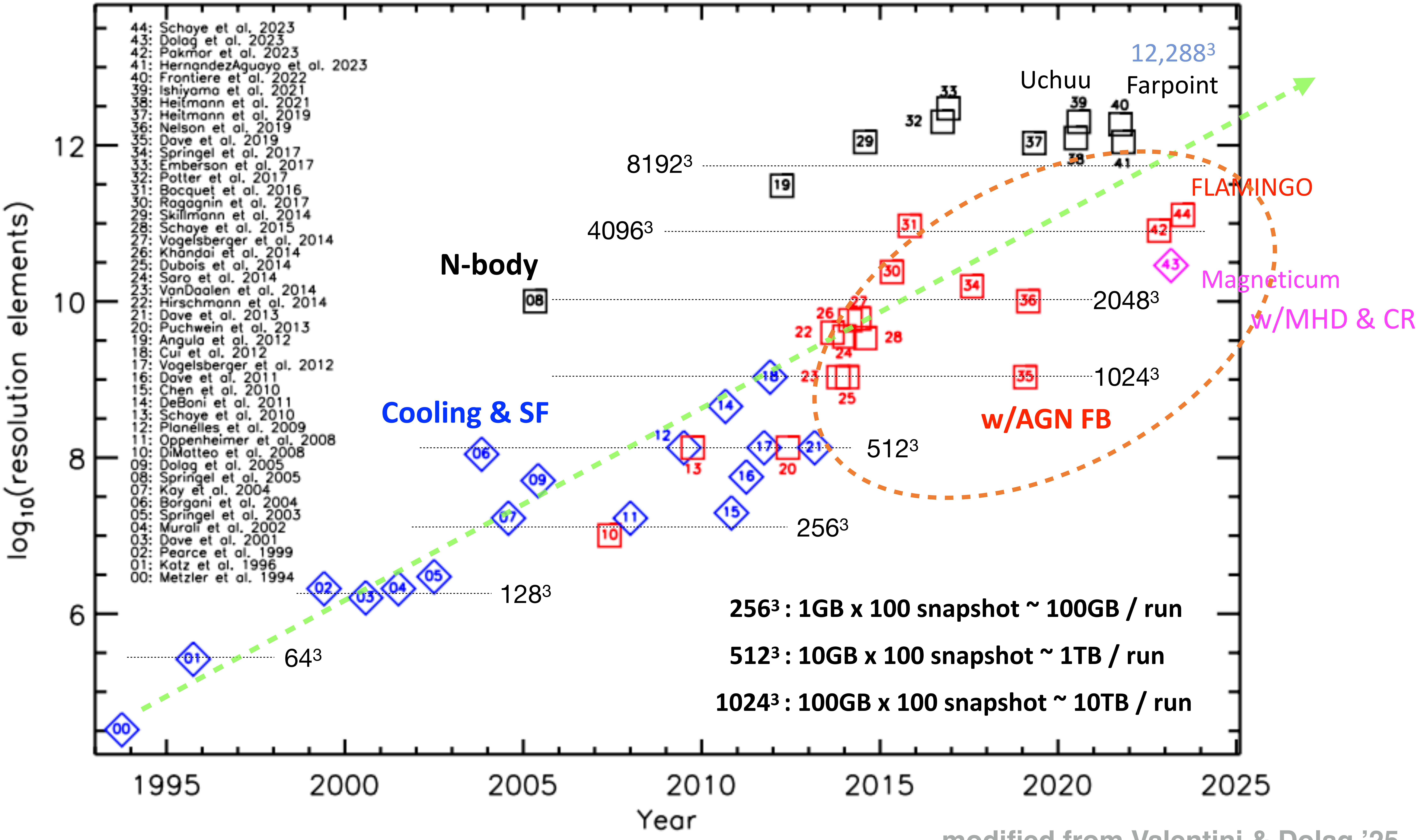
Roca-Fabrega+'21 (incl. KN)



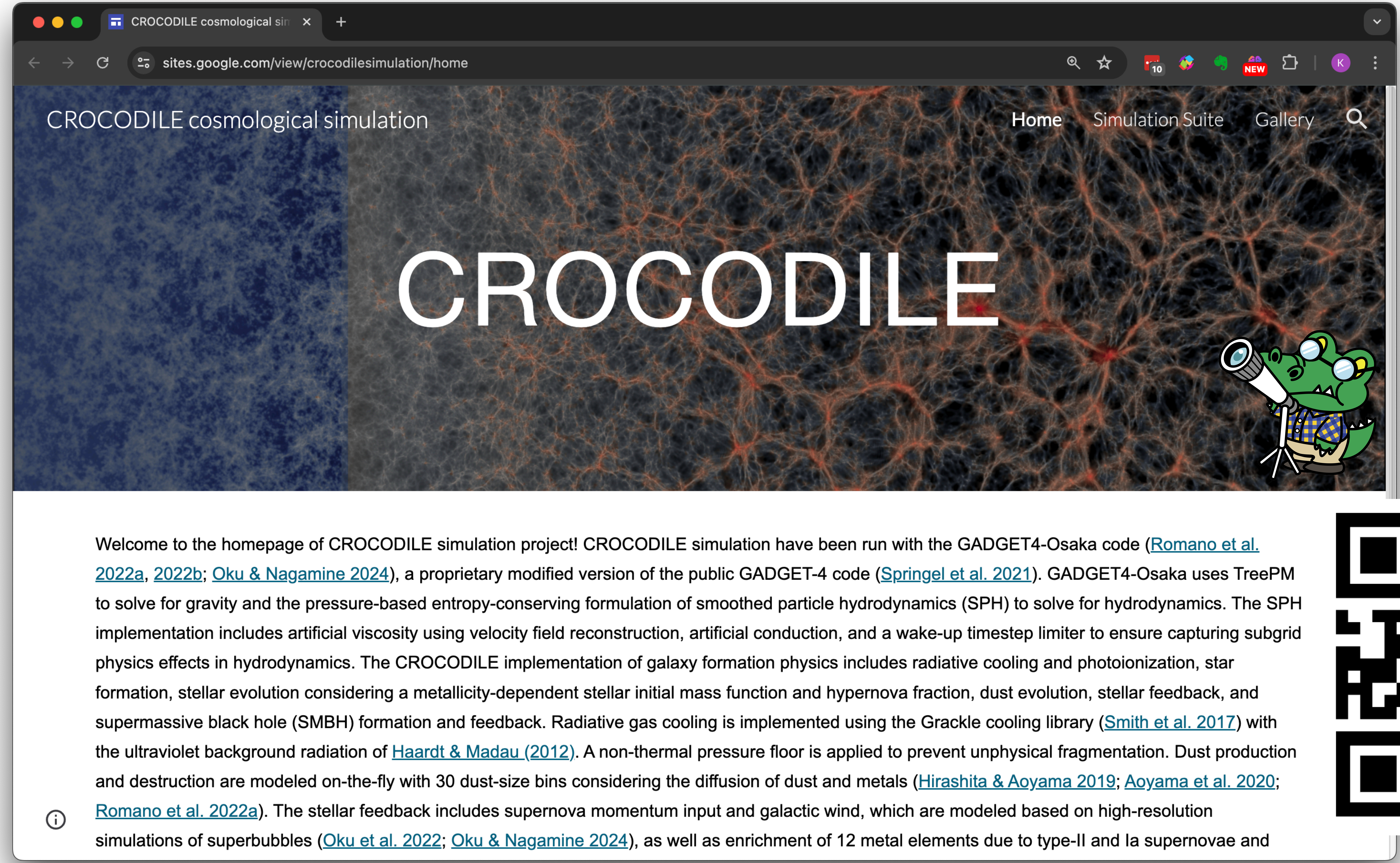
gas cooling	inter-stellar medium	star formation	stellar feedback	super-massive black holes	active galactic nuclei	magnetic fields	radiation fields	cosmic rays
atomic/ molecular/ metals/ tabulated/ network	effective equation of state/ multi-phase	initial stellar mass function/ probabilistic sampling/ enrichment	kinetic/ thermal/ variety of sources from stars, supernovae	numerical seeding/ growth by accretion prescription/ merging	kinetic/ thermal/ radiative/ quasar mode/ radio mode	ideal MHD/ cleaning schemes/ constrained transport	ray tracing/ Monte Carlo/ moment-based	production/ heating/ anisotropic diffusion/ streaming

most important astrophysical processes

Vogelsberger+'19



modified from Valentini & Dolag '25



Welcome to the homepage of CROCODILE simulation project! CROCODILE simulation have been run with the GADGET4-Osaka code (Romano et al. 2022a, 2022b; Oku & Nagamine 2024), a proprietary modified version of the public GADGET-4 code (Springel et al. 2021). GADGET4-Osaka uses TreePM to solve for gravity and the pressure-based entropy-conserving formulation of smoothed particle hydrodynamics (SPH) to solve for hydrodynamics. The SPH implementation includes artificial viscosity using velocity field reconstruction, artificial conduction, and a wake-up timestep limiter to ensure capturing subgrid physics effects in hydrodynamics. The CROCODILE implementation of galaxy formation physics includes radiative cooling and photoionization, star formation, stellar evolution considering a metallicity-dependent stellar initial mass function and hypernova fraction, dust evolution, stellar feedback, and supermassive black hole (SMBH) formation and feedback. Radiative gas cooling is implemented using the Grackle cooling library (Smith et al. 2017) with the ultraviolet background radiation of Haardt & Madau (2012). A non-thermal pressure floor is applied to prevent unphysical fragmentation. Dust production and destruction are modeled on-the-fly with 30 dust-size bins considering the diffusion of dust and metals (Hirashita & Aoyama 2019; Aoyama et al. 2020; Romano et al. 2022a). The stellar feedback includes supernova momentum input and galactic wind, which are modeled based on high-resolution simulations of superbubbles (Oku et al. 2022; Oku & Nagamine 2024), as well as enrichment of 12 metal elements due to type-II and Ia supernovae and

https://sites.google.com/view/crocodilesimulation/home

CROCODILE Suite

L200

L100

L50

L25

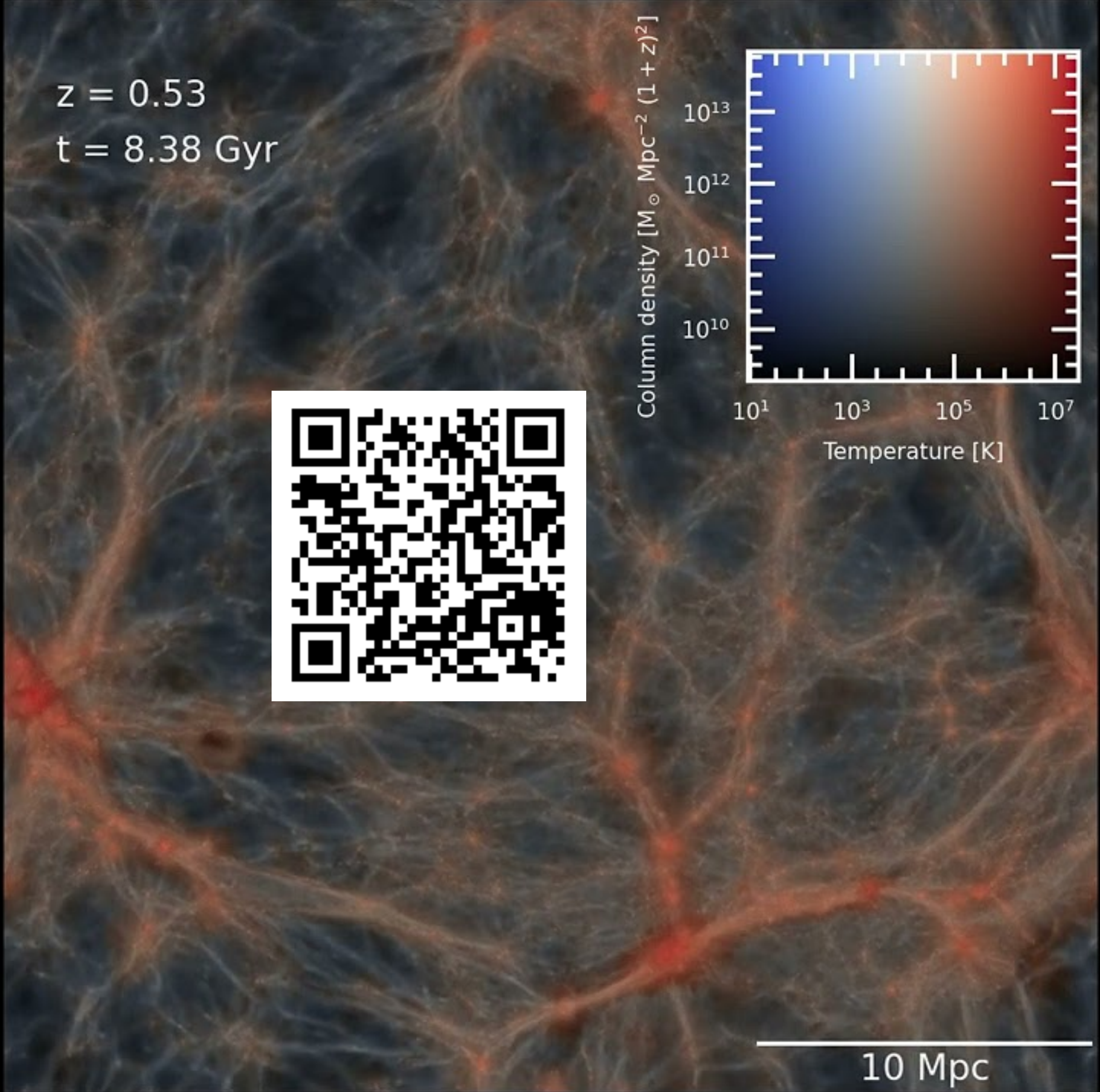
Osaka FB model III
(Oku & KN+ ’24)

							Feedback Type					
Name	**L_{box} [h^{-1} Mpc]**	**$N_{particles}$**	**m_{DM} [h^{-1} M_{sun}]**	**m_{gas} [h^{-1} M_{sun}]**	**C_{visc}**	**m_{BHseed} [h^{-1} M_{sun}]**	**SN Mechanical**	**SN GalWind**	**AGN**	**Stellar IMF**	**HN fraction**	**Note**
L200N1024_Fiducial	200	$2{\times}1024^3$	$5.39{\times}10^8$	$1.01{\times}10^8$	200π	$1{\times}10^5$	✔	✔	✔	variable[1]	Z-dependent[2]	
L200N1024_NoBH	200	$2{\times}1024^3$	$5.39{\times}10^8$	$1.01{\times}10^8$	-	-	✔	✔		variable	Z-dependent	No black holes
L200N1024_BF25	200	$2{\times}1024^3$	$5.39{\times}10^8$	$1.01{\times}10^8$	200π	$1{\times}10^5$	✔	✔	✔	variable	Z-dependent	25x boost to the Bondi accretion rate.
L200N1024_BF50NEL	200	$2{\times}1024^3$	$5.39{\times}10^8$	$1.01{\times}10^8$	200π	$1{\times}10^5$	✔	✔	✔	variable	Z-dependent	50x boost to the Bondi accretion rate. No Eddington limit.
L100N1024_Fiducial	100	$2{\times}1024^3$	$6.74{\times}10^7$	$1.26{\times}10^7$	200π	$1{\times}10^5$	✔	✔	✔	variable	Z-dependent	
L100N1024_NoBH	100	$2{\times}1024^3$	$6.74{\times}10^7$	$1.26{\times}10^7$	-	-	✔	✔		variable	Z-dependent	No black holes
L100N1024_NoFBNoBH	100	$2{\times}1024^3$	$6.74{\times}10^7$	$1.26{\times}10^7$	-	-				variable	Z-dependent	No black holes
L100N1024_SNonly	100	$2{\times}1024^3$	$6.74{\times}10^7$	$1.26{\times}10^7$	200π	$1{\times}10^5$	✔	✔		variable	Z-dependent	
L100N1024_Cvisc100pi	100	$2{\times}1024^3$	$6.74{\times}10^7$	$1.26{\times}10^7$	**100π**	$1{\times}10^5$	✔	✔	✔	variable	Z-dependent	
L50N512_Fiducial	50	$2{\times}512^3$	$6.74{\times}10^7$	$1.26{\times}10^7$	200π	$1{\times}10^5$	✔	✔	✔	variable	Z-dependent	
L50N512_NoSNGalWind	50	$2{\times}512^3$	$6.74{\times}10^7$	$1.26{\times}10^7$	200π	$1{\times}10^5$	✔		✔	variable	Z-dependent	
L50N512_AGNonly	50	$2{\times}512^3$	$6.74{\times}10^7$	$1.26{\times}10^7$	200π	$1{\times}10^5$			✔	variable	Z-dependent	
L50N512_SNonly	50	$2{\times}512^3$	$6.74{\times}10^7$	$1.26{\times}10^7$	200π	$1{\times}10^5$	✔	✔		variable	Z-dependent	
L50N512_NoZdepSN	50	$2{\times}512^3$	$6.74{\times}10^7$	$1.26{\times}10^7$	200π	$1{\times}10^5$	✔	✔	✔	Chabrier	0.01	
L50N512_NoBH	50	$2{\times}512^3$	$6.74{\times}10^7$	$1.26{\times}10^7$	-	-	✔	✔		variable	Z-dependent	No black holes
L50N512_NoFB	50	$2{\times}512^3$	$6.74{\times}10^7$	$1.26{\times}10^7$	200π	$1{\times}10^5$				variable	Z-dependent	
L50N512_NoFBNoBH	50	$2{\times}512^3$	$6.74{\times}10^7$	$1.26{\times}10^7$	-	-				variable	Z-dependent	No black holes
L50N512_LowCvisc	50	$2{\times}512^3$	$6.74{\times}10^7$	$1.26{\times}10^7$	**2π**	$1{\times}10^5$	✔	✔	✔	variable	Z-dependent	
L50N512_LowCviscLowMseed	50	$2{\times}512^3$	$6.74{\times}10^7$	$1.26{\times}10^7$	**2π**	**$1{\times}10^4$**	✔	✔	✔	variable	Z-dependent	

CROCODILE
simulation

GADGET4-OSAKA code

Dr. Wani
@Osaka U.



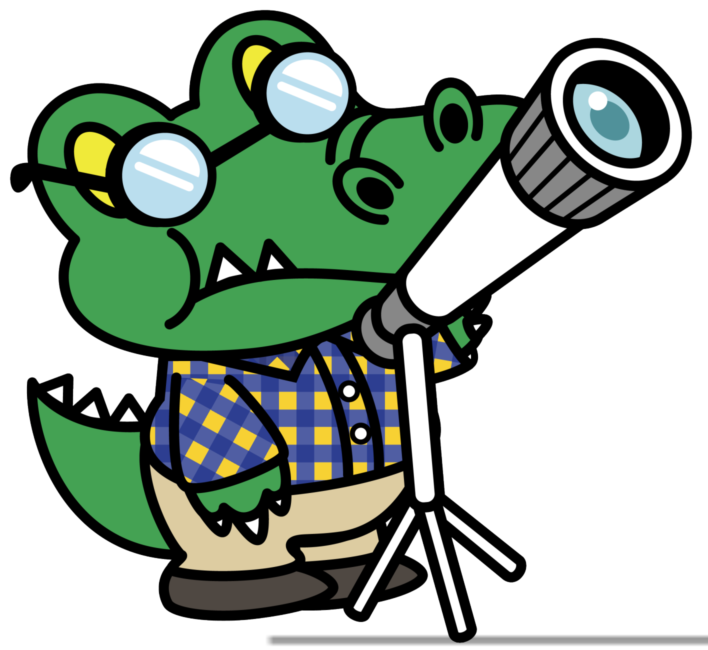
Understanding the matter distribution:
DM, gas, stars
HI, metals, dust, …

25 & 50 Mpc/h
Star formation
SN & AGN fb
UVB

+ 100, 200, 500 Mpc/h

Oku & KN '24 ApJ

CROCODILE

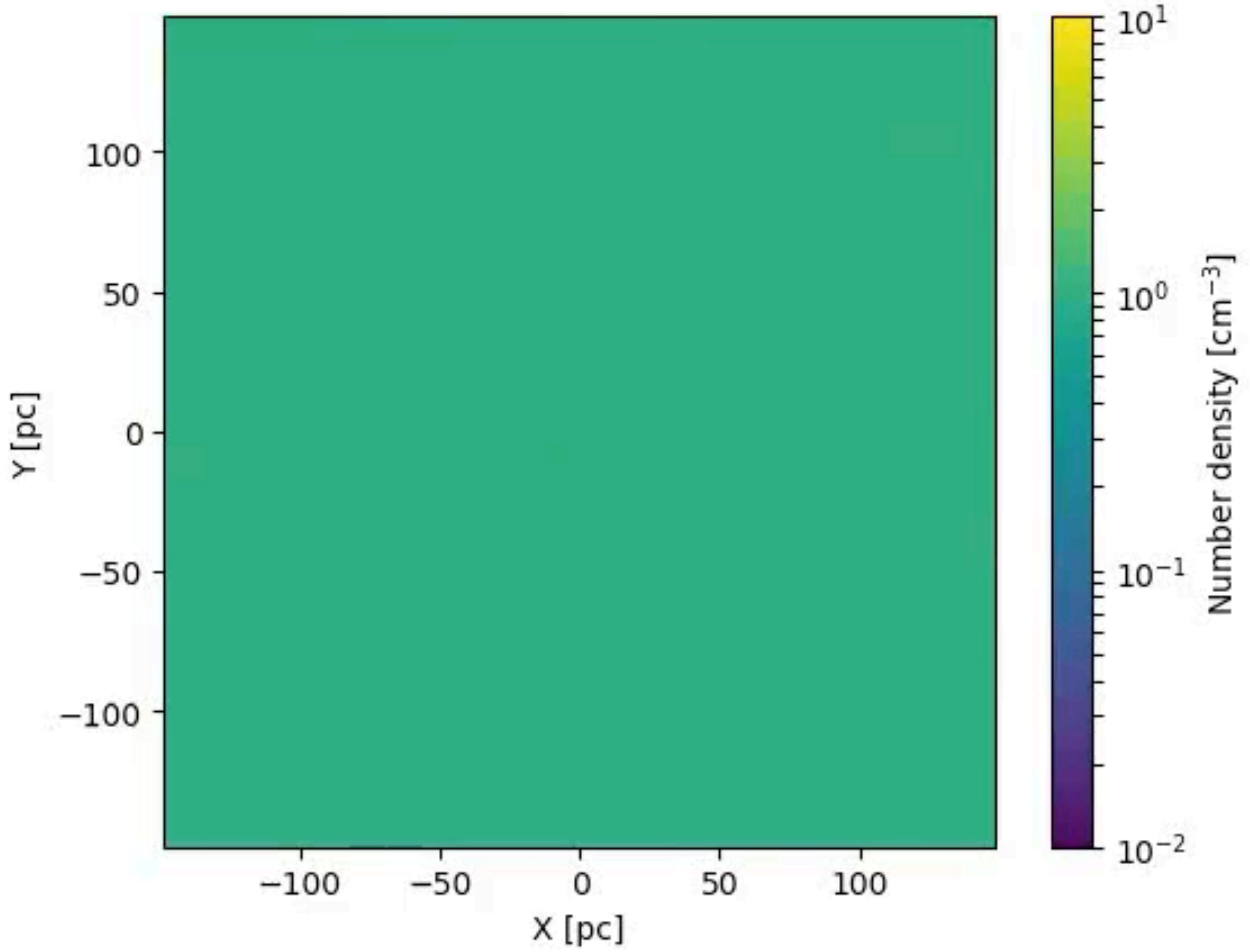


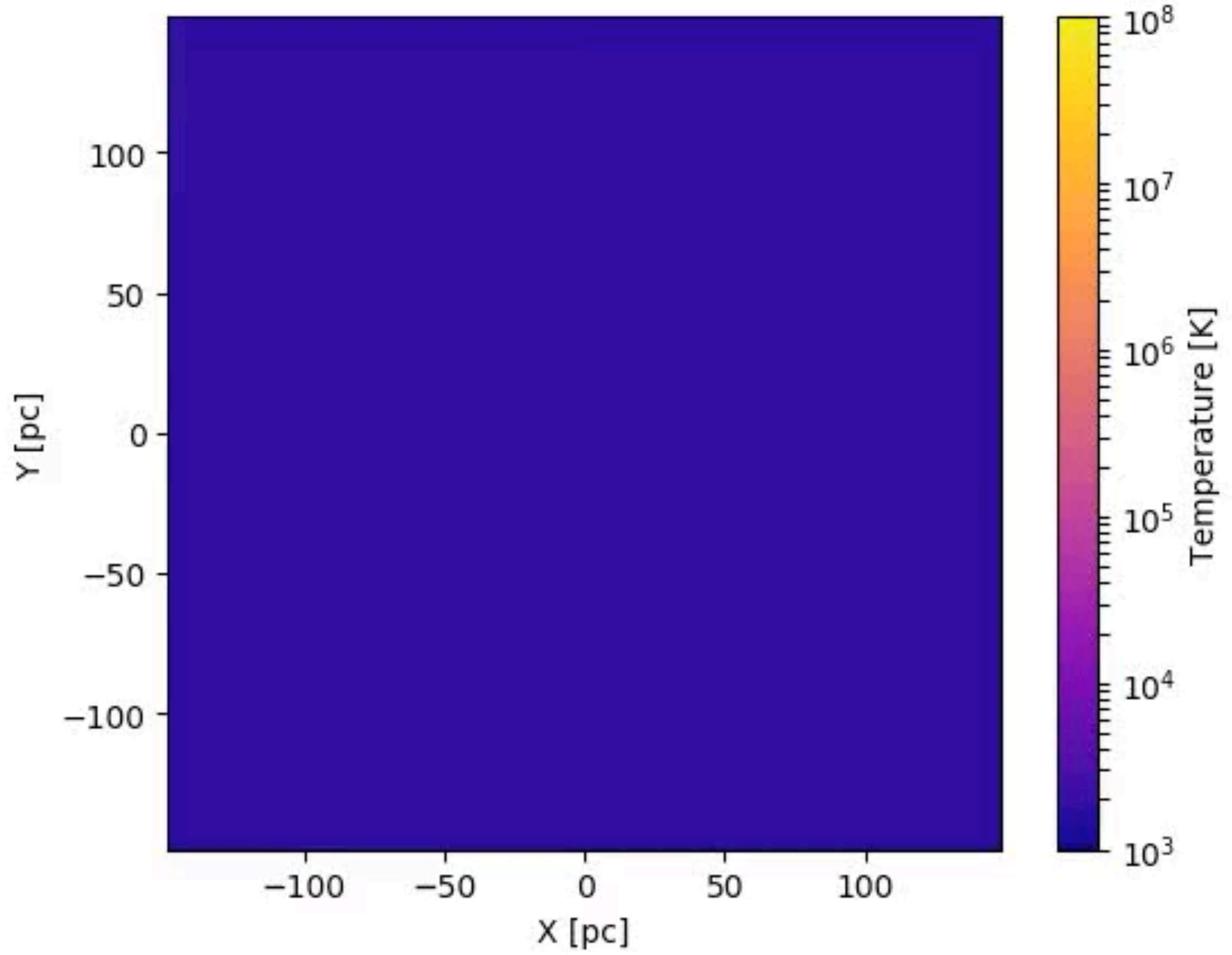
- **Cosmological SPH GADGET4-Osaka code** — Oku & KN '24
- **Star formation, SN / AGN feedback**, HM12/FG09 UVB
- **Chabrier / Top-heavy IMF at low-metallicities**
- 13 species chemistry & cooling module — CELib (Saitoh '17)
- variety of feedback model combinations in 25-500 Mpc/h
- dust formation/destruction w/ grain size evol. — Aoyama+'18, '19, '21

Name	$L_{\rm box}$	$N_{\rm particle}$	$C_{\rm visc}$	$m_{\rm seed}$	Feedback Type: SN Mechanical	Feedback Type: SN GalWind	Feedback Type: AGN	Stellar IMF	HN fraction
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Fiducial	50	2×512^3	200π	1×10^5	✓	✓	✓	variable[a]	Z-dependent[b]
NoSNGalWind	50	2×512^3	200π	1×10^5	✓		✓	variable	Z-dependent
AGNonly	50	2×512^3	200π	1×10^5			✓	variable	Z-dependent
SNonly	50	2×512^3	200π	1×10^5	✓	✓		variable	Z-dependent
NoFB	50	2×512^3	200π	1×10^5				variable	Z-dependent
LowCvisc	50	2×512^3	2π	1×10^5	✓	✓	✓	variable	Z-dependent
LowCviscLowMseed	50	2×512^3	2π	1×10^4	✓	✓	✓	variable	Z-dependent
L25N512	25	2×512^3	200π	1×10^5	✓	✓	✓	variable	Z-dependent
L25N256	25	2×256^3	200π	1×10^5	✓	✓	✓	variable	Z-dependent
L25N128	25	2×128^3	200π	1×10^5	✓	✓	✓	variable	Z-dependent
L25N256NoZdepSN	25	2×256^3	200π	1×10^5	✓	✓	✓	Chabrier	0.01

$m_{\rm DM} = 6.7 \times 10^7 M_\odot$ $\quad m_{\rm gas} = 1.3 \times 10^7 M_\odot$ $\quad \epsilon_g = 3.4 h^{-1}$ ckpc (≥ 500 phys pc)

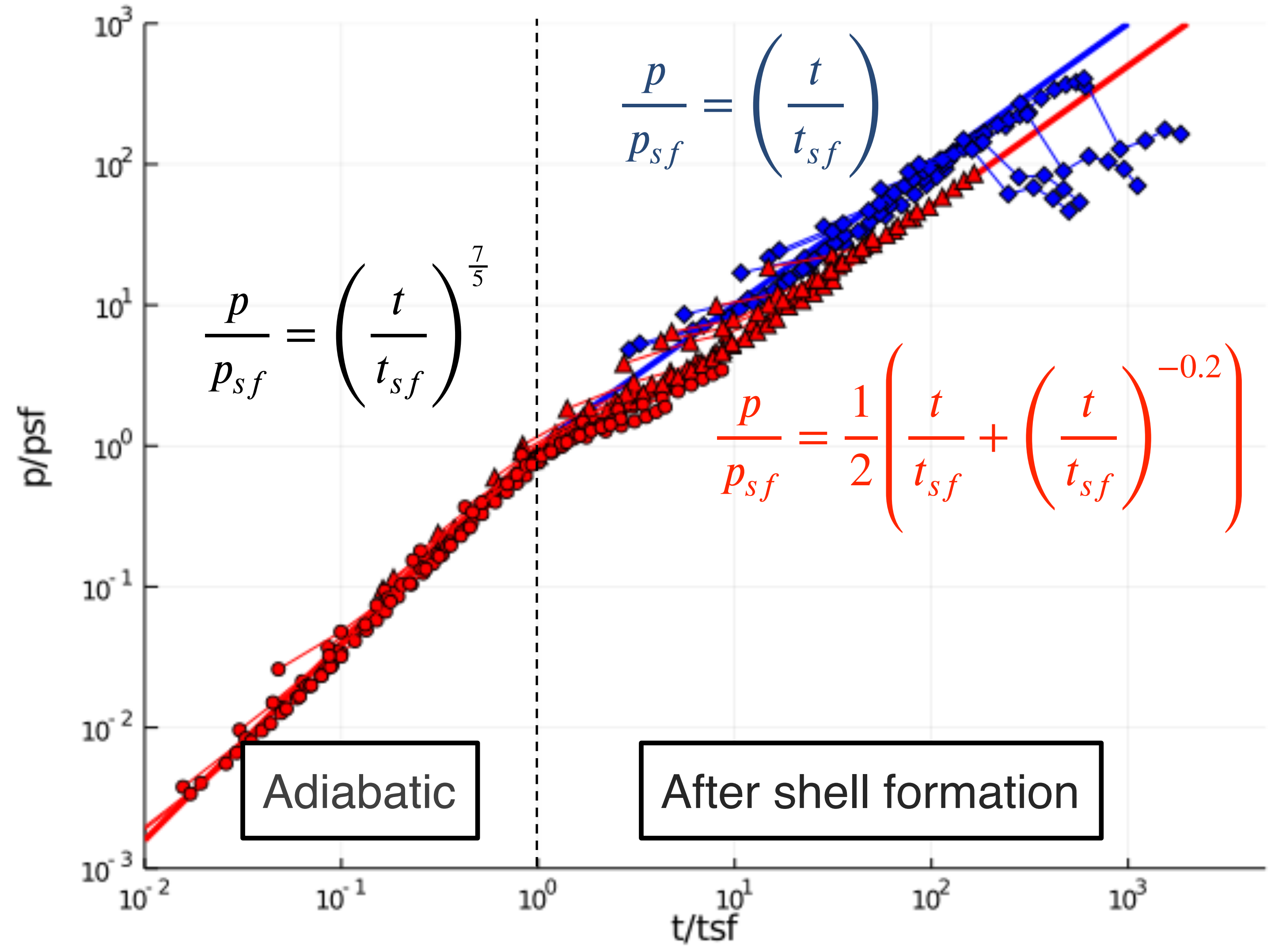
Athena++ Superbubble Simulation





A unified time evolution model of superbubble momentum

Oku+ ’22



Assuming initial stellar cluster mass function $\frac{dN}{dM_c} \propto M_c^{-2}$ (Krumholz+19), translate to momentum per SN

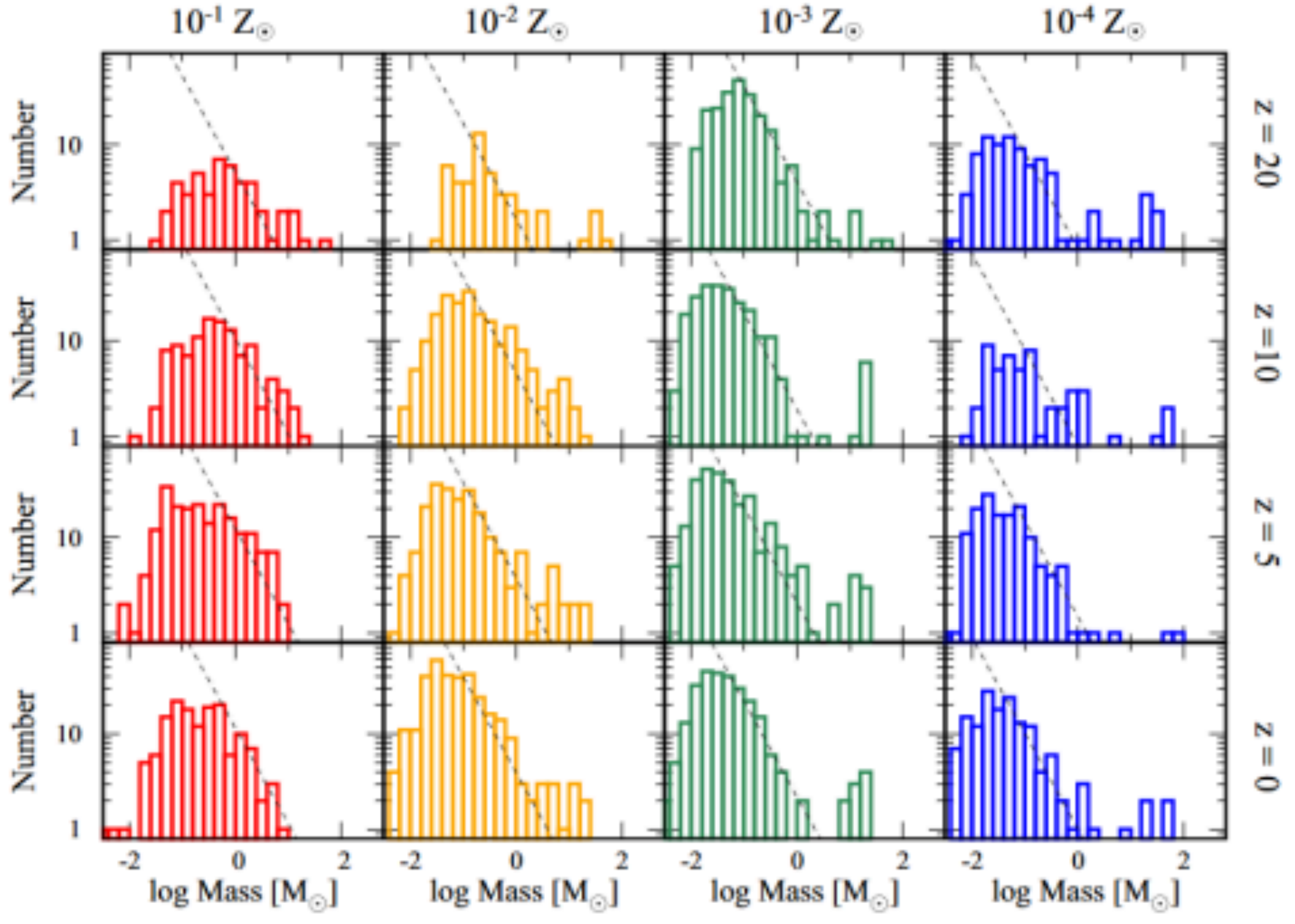
$$\hat{p} = 1.75 \times 10^5 \left(\frac{n}{1\ \mathrm{cm}^{-3}}\right)^{-0.05} \left(\frac{\Lambda(10^6\mathrm{K})}{10^{-22}\ \mathrm{erg\ cm^3\ s^{-1}}}\right)^{-0.17} \ [M_\odot \mathrm{km\ s^{-1}}]$$

x3 enhancement of momentum for Z=0 than Z⊙ ISM

cf. Kim & Ostriker ’18, …

Top-heavy IMF & Enhanced specific SN energy

Metallicity- & redshift-dependent **IMF**



(Chon+22)

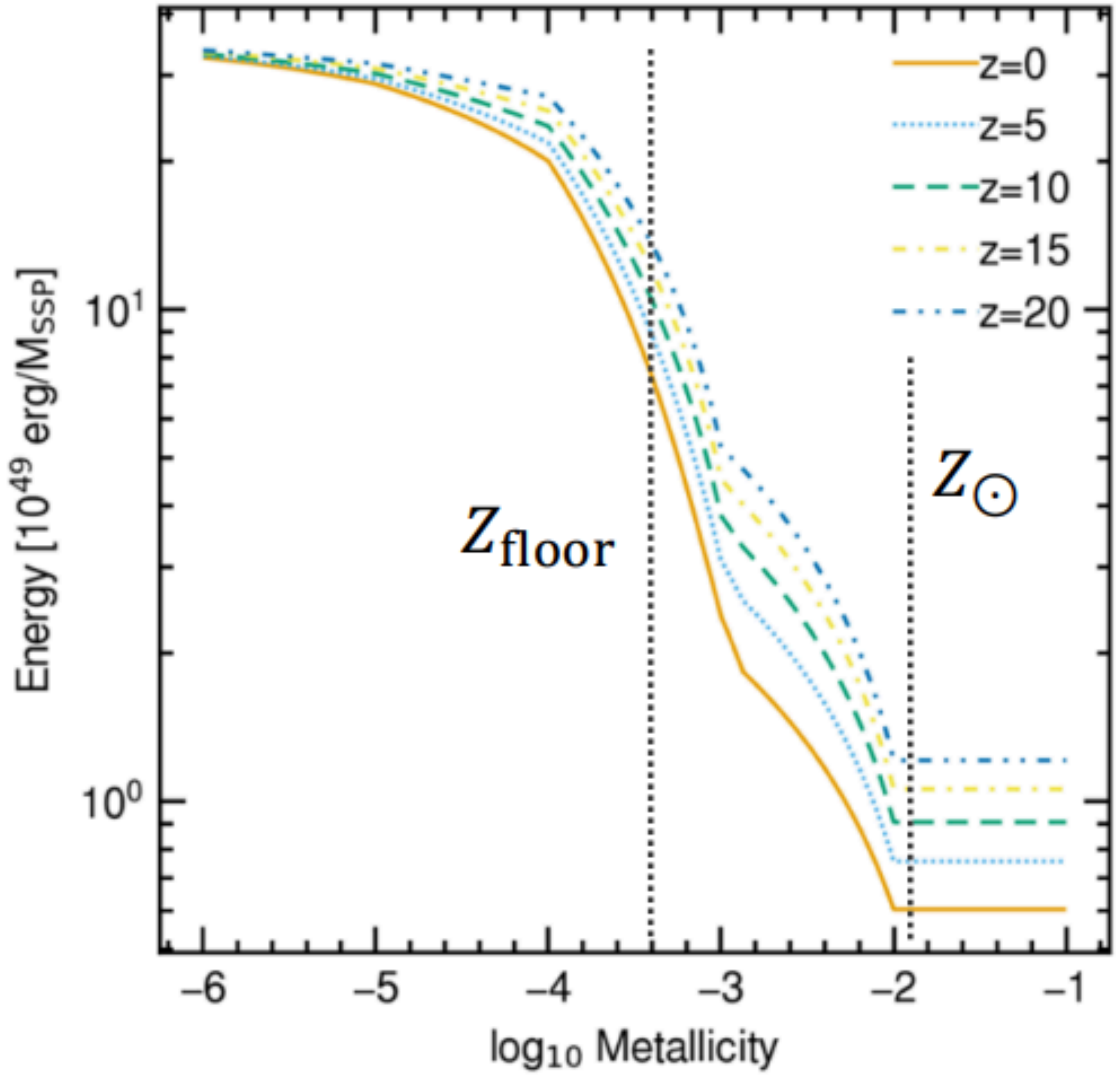
Metallicity-dependent

HN Fraction $(E_{\rm HN} = 10^{52}\ {\rm erg})$

$$f_{\rm HN} = \begin{cases} 0.5 & (Z < 0.001) \\ 0.01 & (Z > 0.001) \end{cases}$$

(Kobayashi+06)

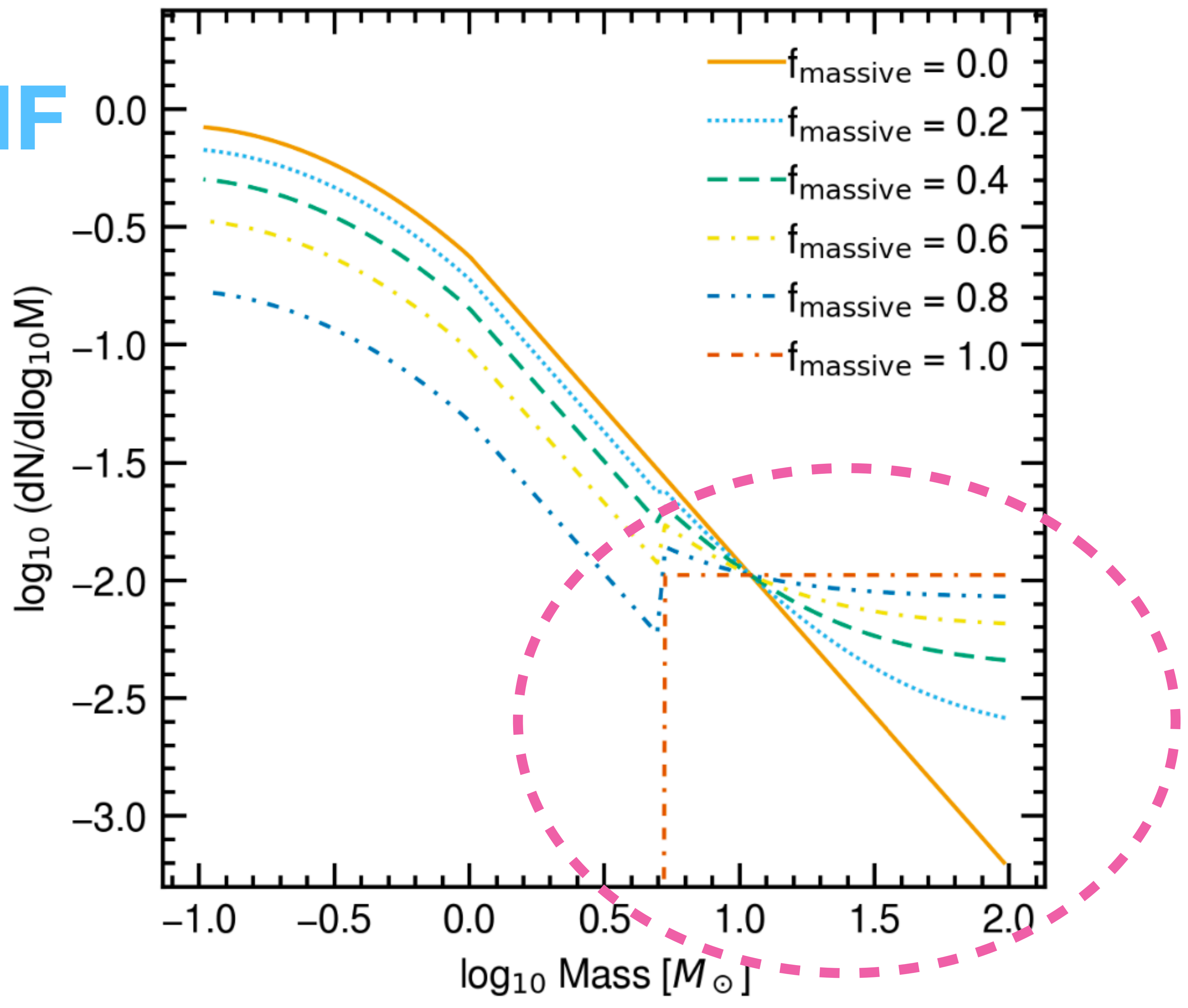
Metallicity- & redshift-dependent **SN Energy Yield**



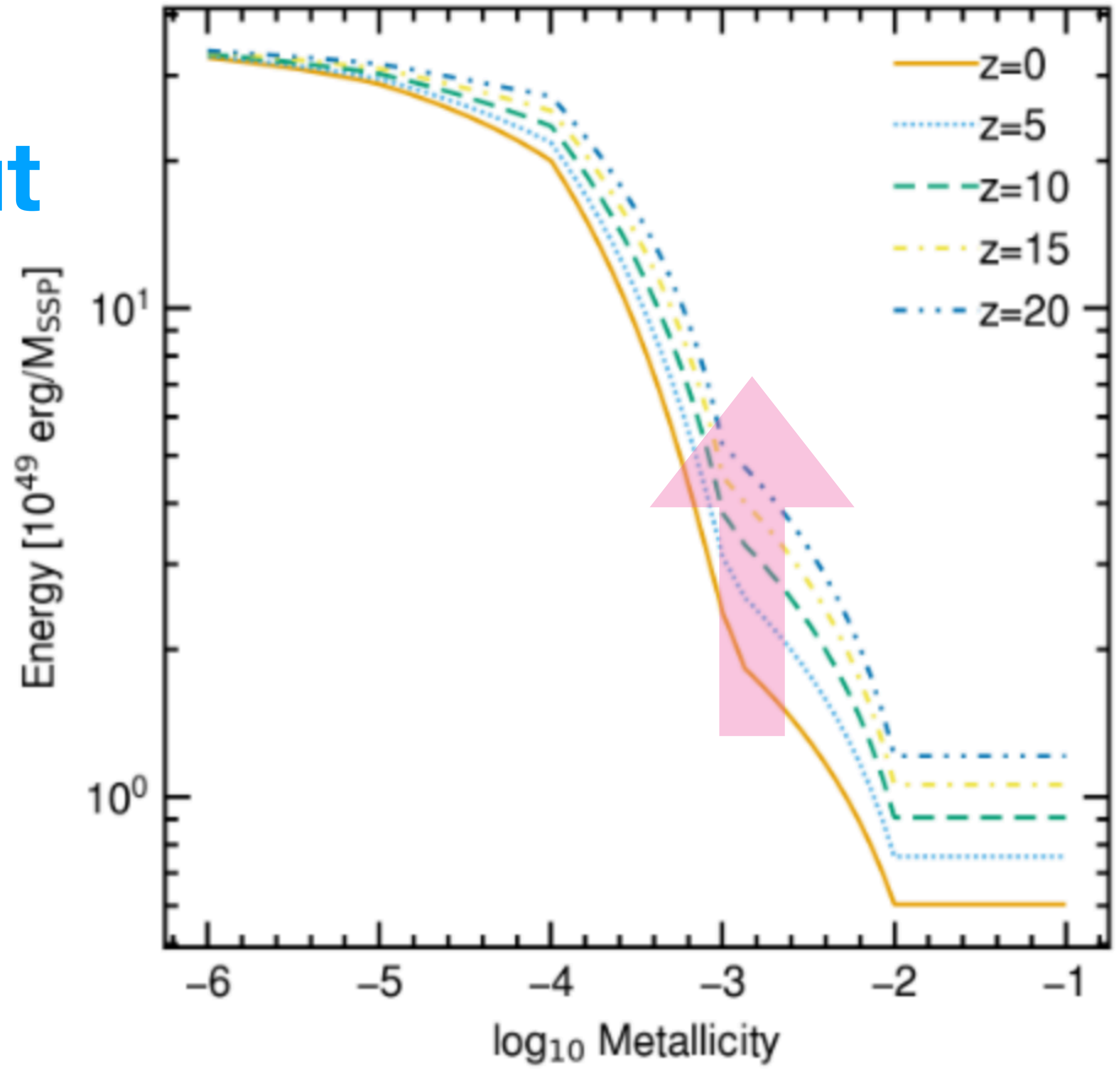
(Oku&Nagamine24)

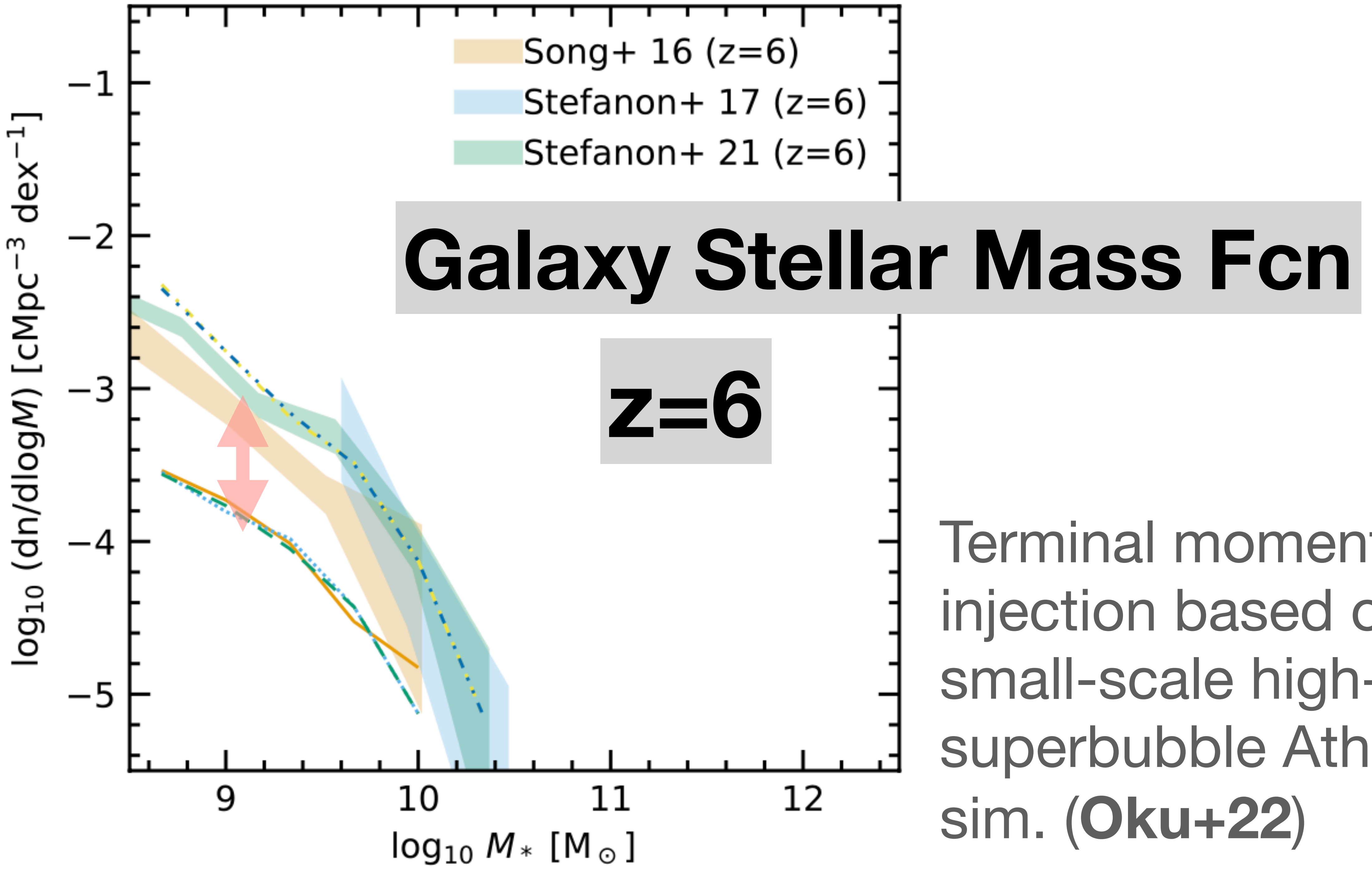
SN Energy Boost by >10

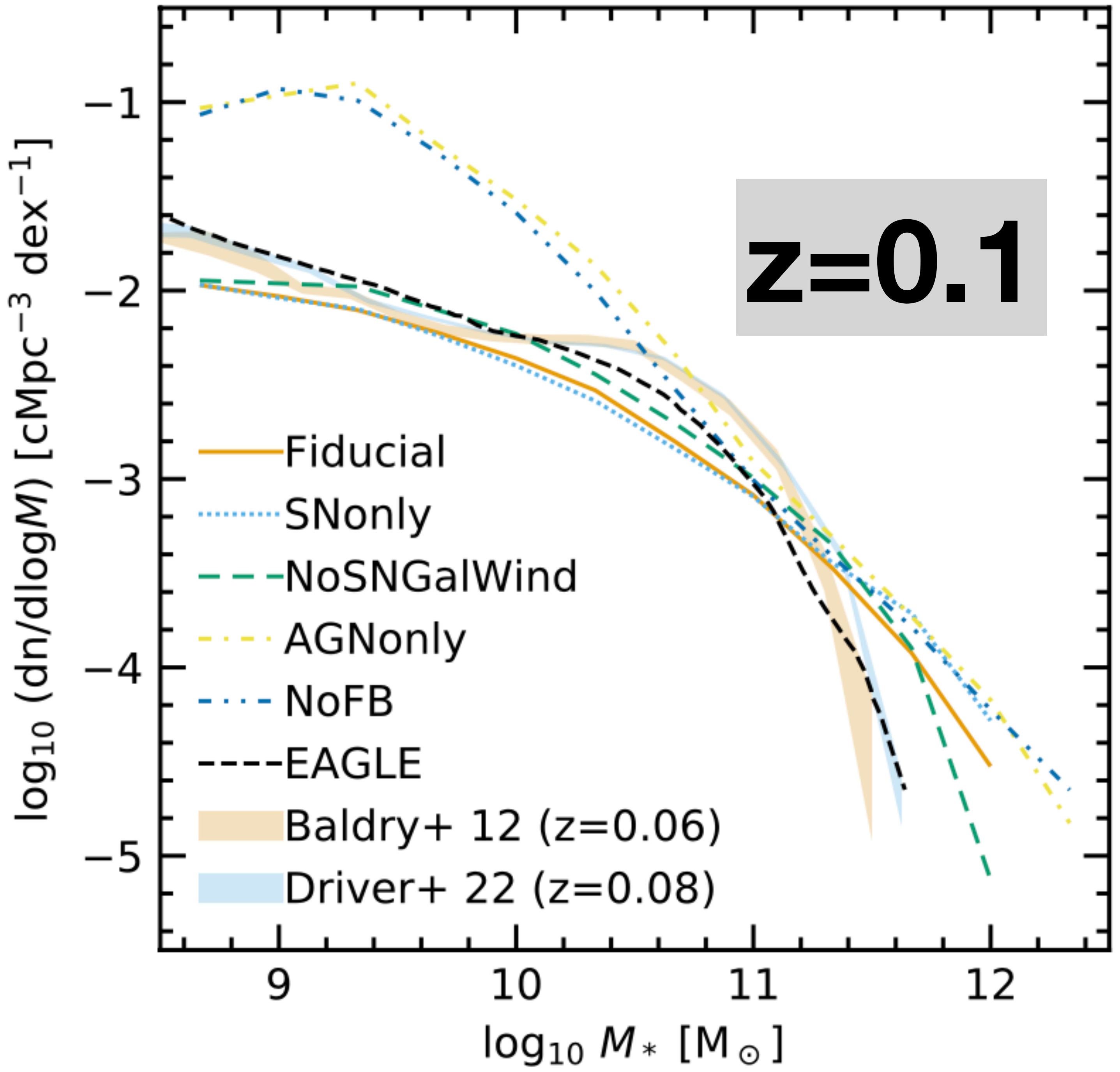
Top-heavy IMF model (Chon+22)



Specific SN energy input



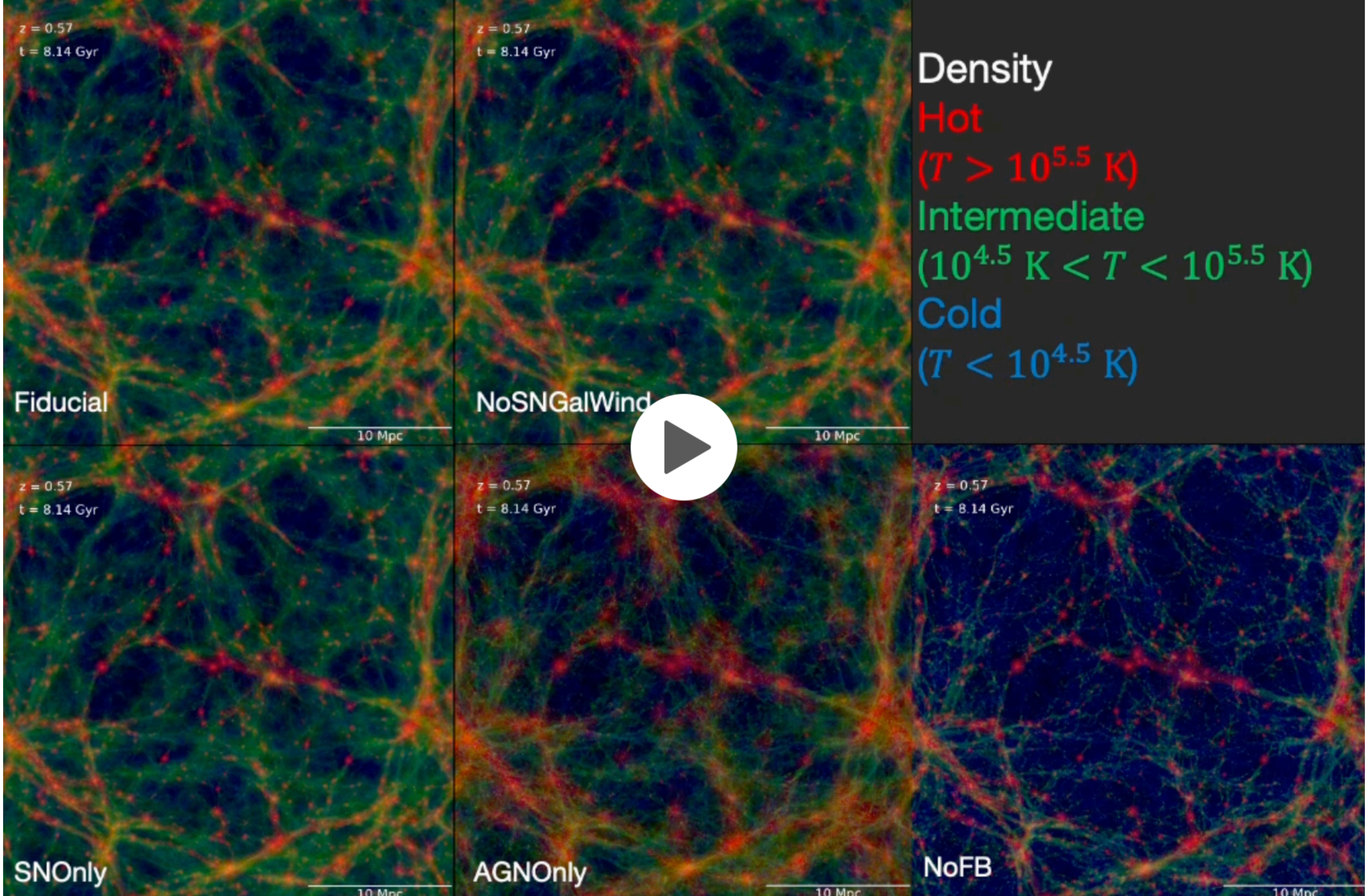




Terminal momentum injection based on small-scale high-res superbubble Athena++ sim. (**Oku+22**)

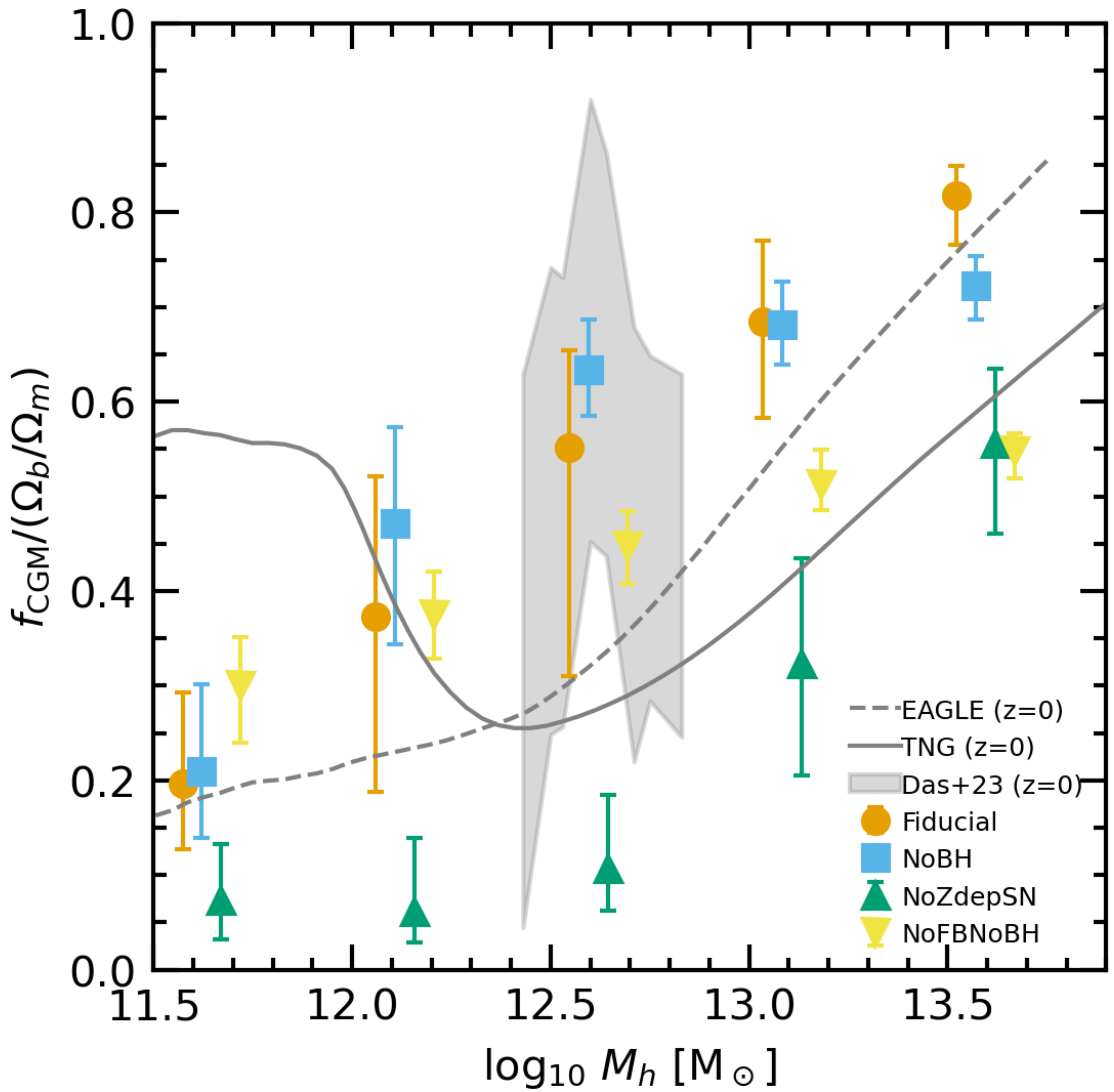
+

TIGRESS mass-loading factor (**Kim & Ostriker ’18; Oku & KN ’24**)



z = 0.57
t = 8.14 Gyr
Fiducial
10 Mpc
z = 0.57
t = 8.14 Gyr
NoSNGalWind
10 Mpc
Density
Hot
($T > 10^{5.5}$ K)
Intermediate
($10^{4.5}$ K $< T < 10^{5.5}$ K)
Cold
($T < 10^{4.5}$ K)
z = 0.57
t = 8.14 Gyr
SNOnly
z = 0.57
t = 8.14 Gyr
AGNOnly
z = 0.57
t = 8.14 Gyr
NoFB

Baryon fraction in DM halos



cf. Davies+'20

Zhang, KN+'25, submitted


IllustrisTNG
SIMBA
Ramses
CROCODILE
$z = 1.00$
https://camels.readthedocs.io/en/latest/codes.html
00:31
-00:11

Learning Feedback Physics with AI

AGN Feedback

Supermassive black holes regulate galaxy growth through energetic jets.

Stellar Feedback

Supernovae drive galactic winds, enriching the intergalactic medium.

Model Refinement

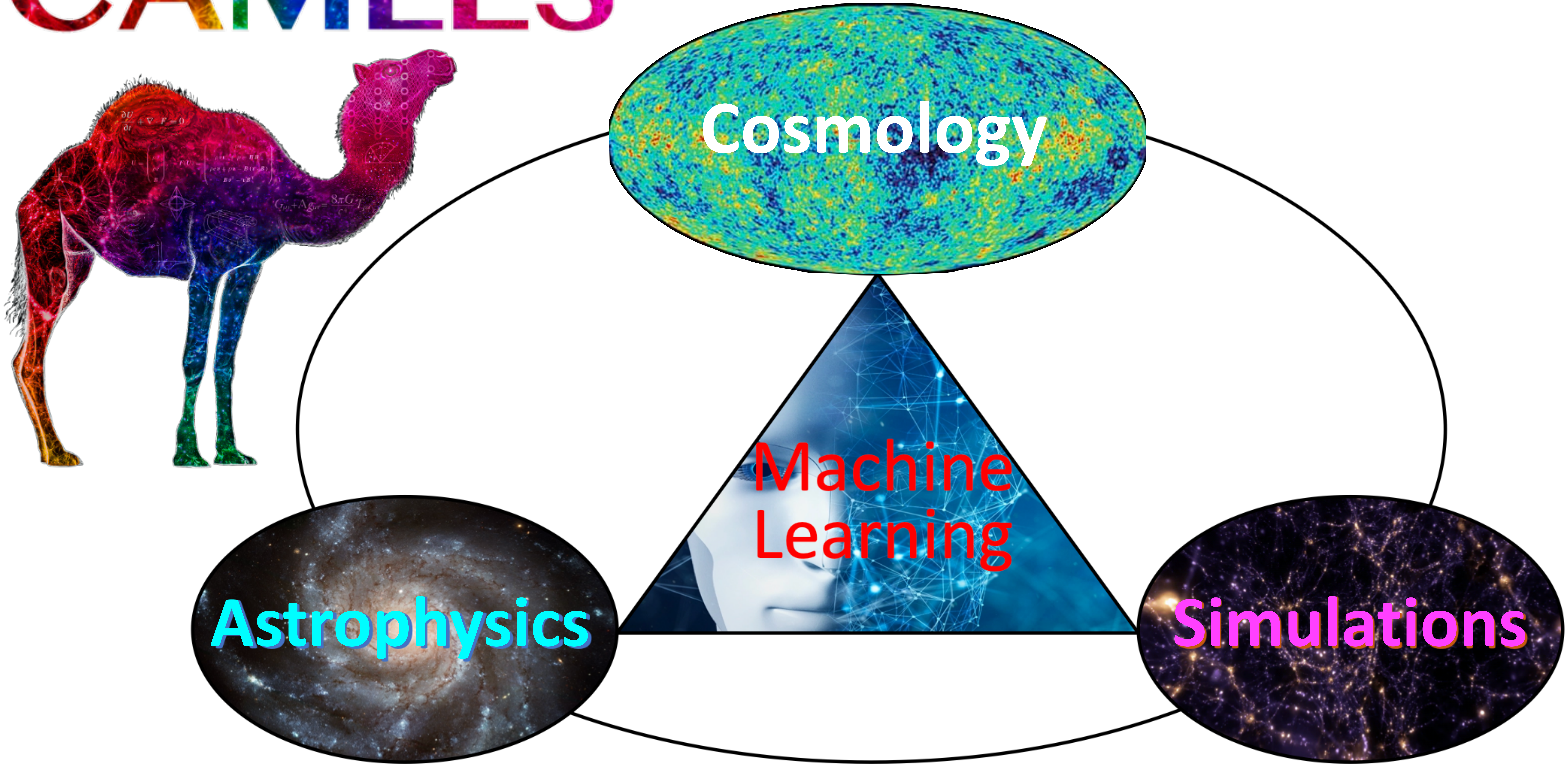
Improved physics implementations based on AI insights.

AI Analysis

ML identifies feedback signatures across cosmic scales.

CAMELS

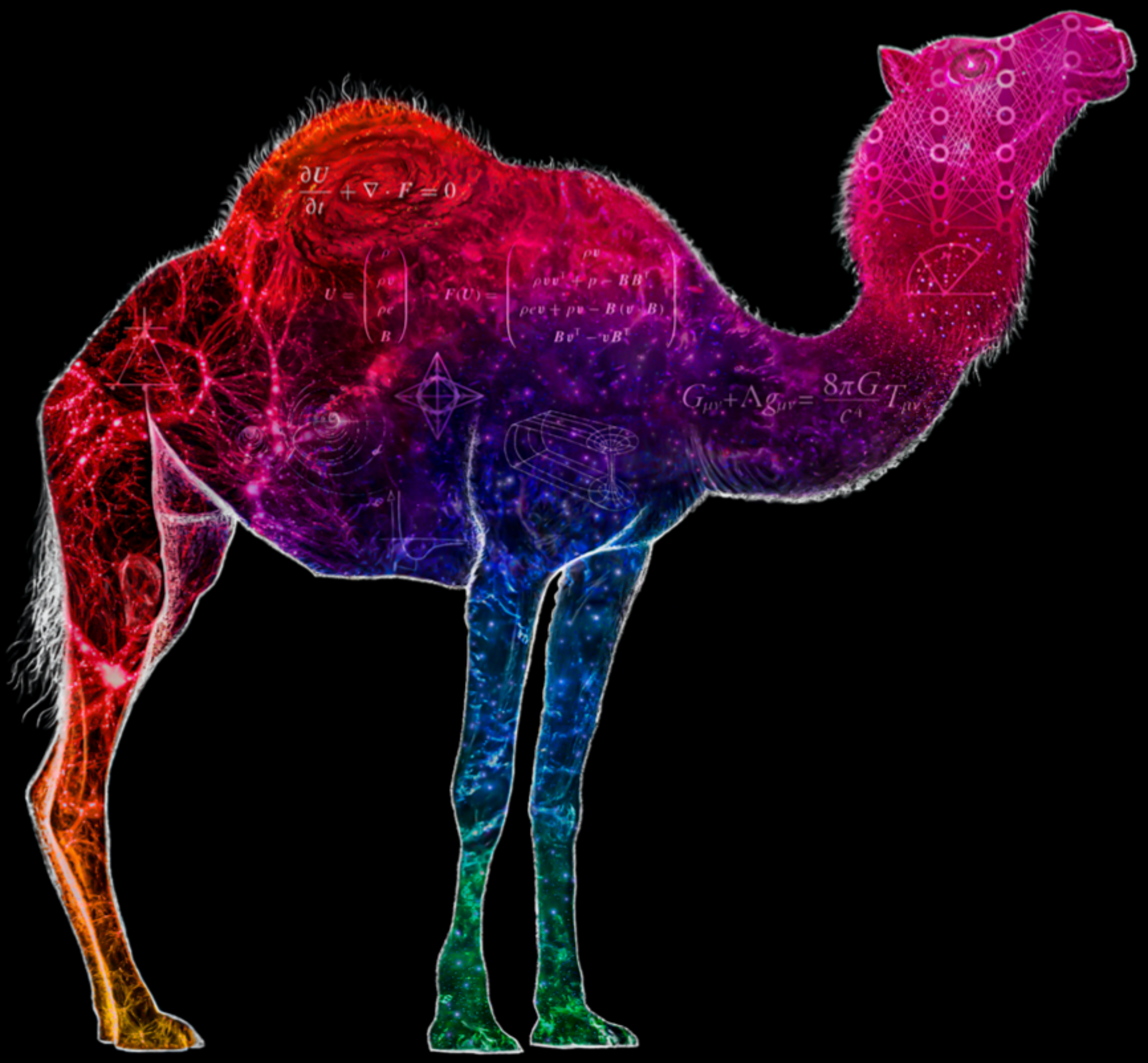
https://www.camel-simulations.org/



CAMELS

Cosmology and Astrophysics with MachinE Learning Simulations

https://www.camel-simulations.org



A set of 14,000+ simulations:

- 6,000+ N-body sims
- 8,000+ state-of-the art hydrodynamic sims

Hydrodynamic sims w/ 9 different codes/subgrid models:

1) IllustrisTNG, 2) SIMBA, 3) Astrid, 4) Magneticum, 5) SWIFT-EAGLE, 6) Ramses, 7) Enzo, 8) ***CROCODILE***, 9) Obsidian

- Variations in cosmological & astrophysical parameters

2-4 cosmological & 4-30 astrophysical parameters

- Hundreds of thousands of snapshots & galaxy catalogs

- Designed for ***Machine Learning Applications***

CAMELS

https://www.camel-simulations.org/

$\gtrsim$4PB and increasing.

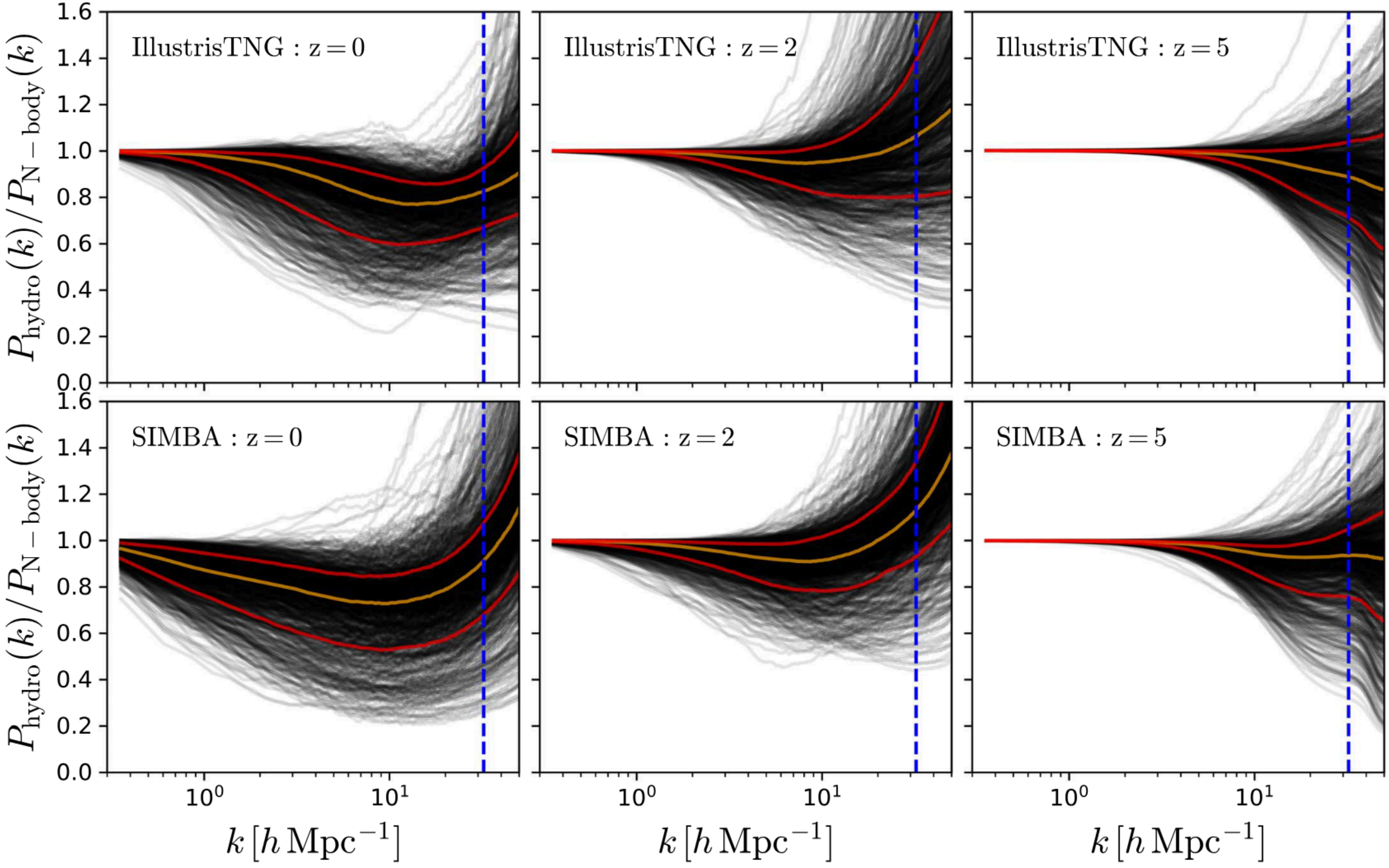
Now <u>L25</u> & <u>L50</u>

~1.6PB ~1.8PB

Suite

- IllustrisTNG
- SIMBA
- Astrid
- Magneticum
- CROCODILE
- SWIFT-EAGLE
- Ramses
- Enzo
- Obsidian

SB	LH	1P	CV	EX
5 cosmological parameters	Ω_m, σ_8	Ω_m, σ_8	Same cosmology & astrophysics	Same cosmology Extreme astrophysics
Most astrophysical parameters	$A_{\rm SN1}, A_{\rm SN2}$ $A_{\rm AGN1}, A_{\rm AGN2}$	$A_{\rm SN1}, A_{\rm SN2}$ $A_{\rm AGN1}, A_{\rm AGN2}$		
Different random seed	Different random seed	Same random seed	Different random seed	Same random seed
<u>1,024 simulations</u>	<u>1,000 simulations</u>	~80 <u>simulations</u>	<u>27 simulations</u>	<u>3 simulations</u>

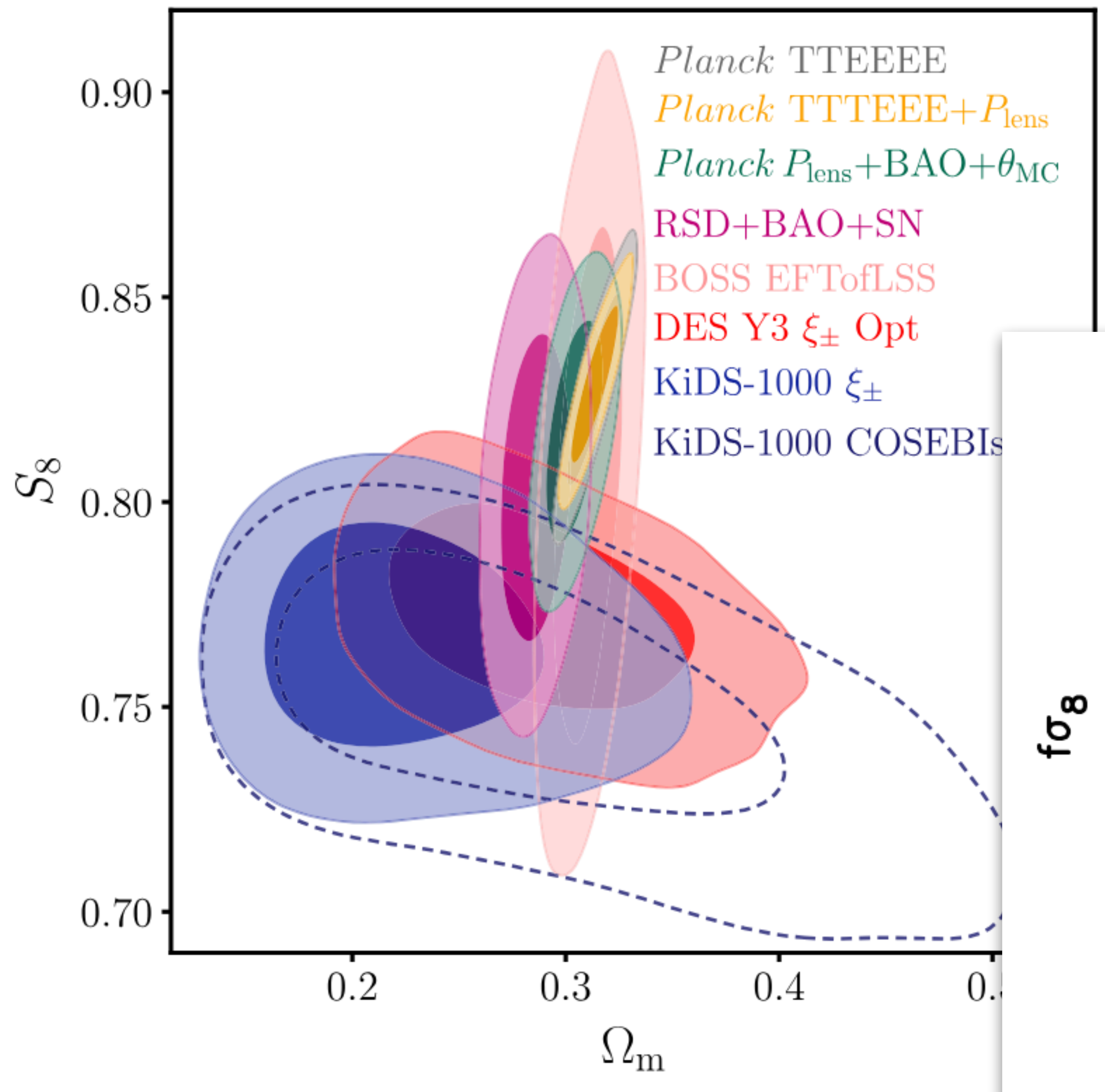


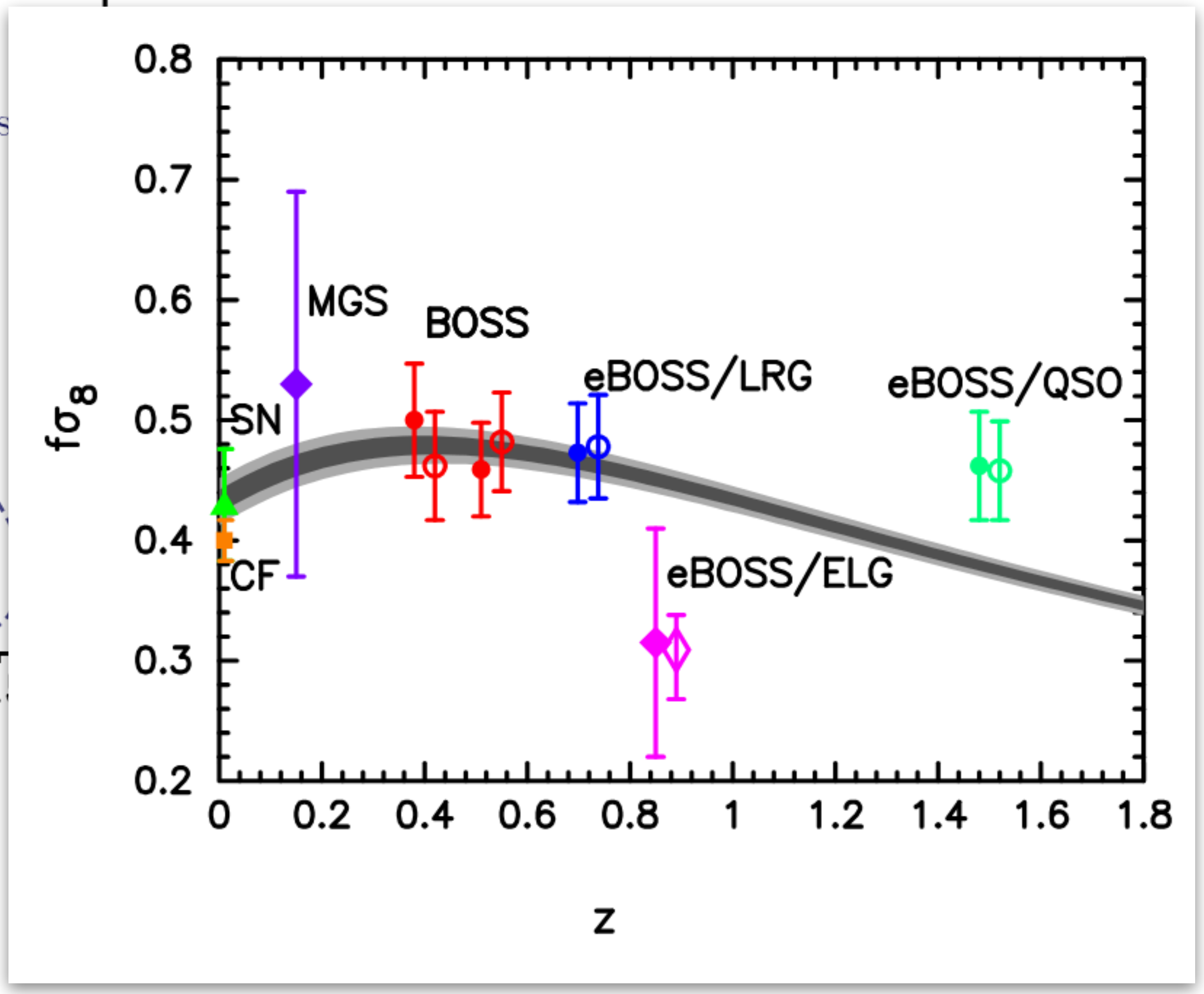
1P set variations

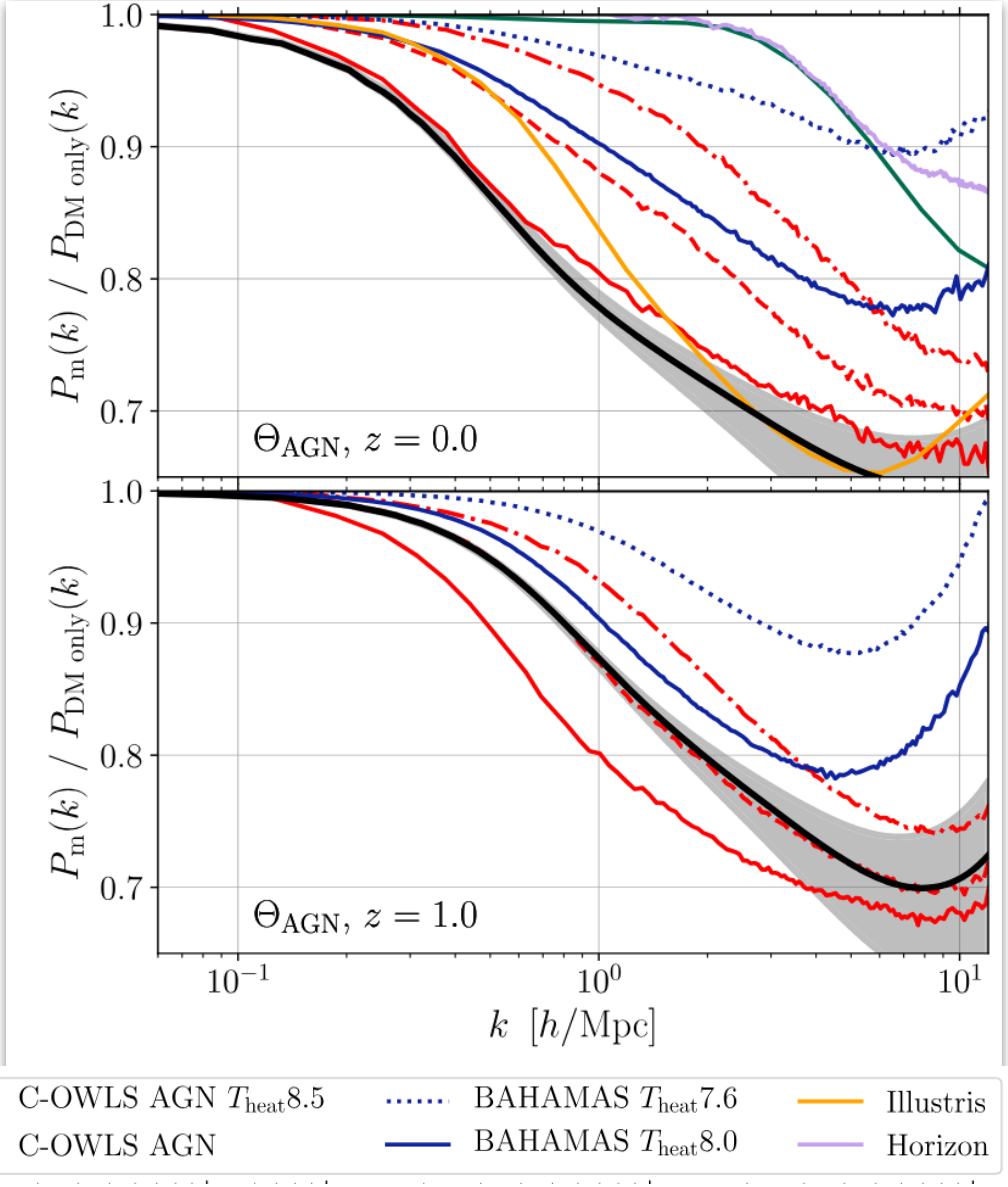
Villaescusa-Navarro+ ’21 (CAMELS flagship paper)

S_8 tension & P_k suppression

Amon & Efstathiou '22

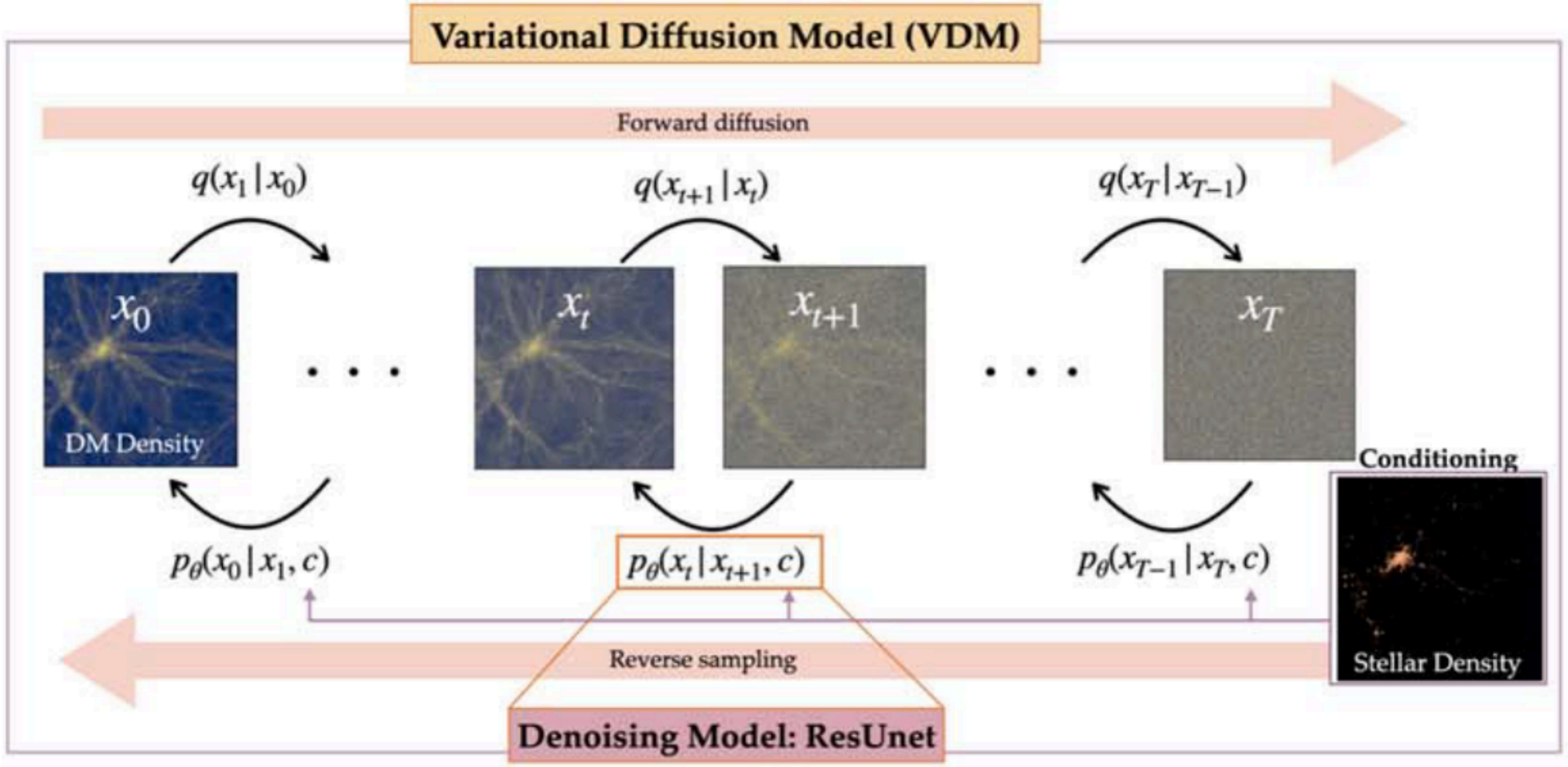


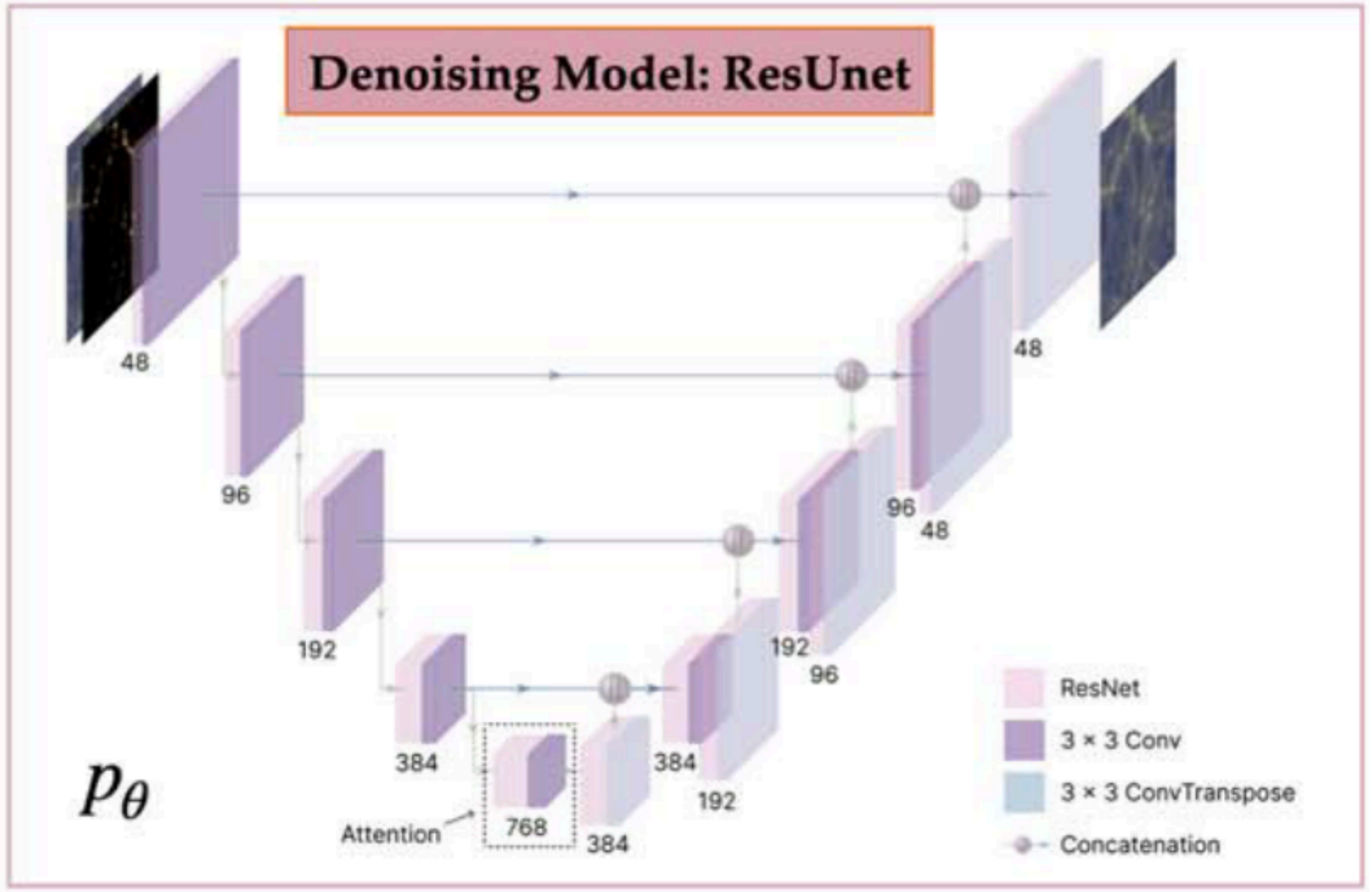


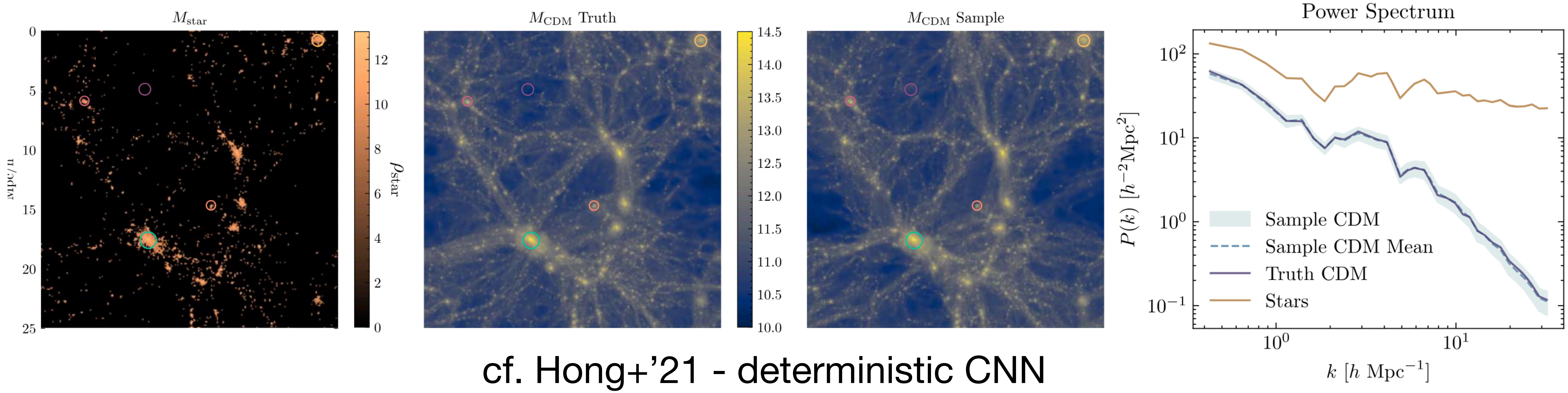


“**Debiasing with Diffusion: Probabilistic Reconstruction of Dark Matter Fields from Galaxies with CAMELS**”

Ono+ ’24, ApJ







cf. Hong+’21 - deterministic CNN

X-raying CAMELS: Constraining Baryonic Feedback in the Circum-Galactic Medium with the CAMELS simulations and eRASS X-ray Observations

Lau+ ’25, ApJ (arXiv:2412.04559)

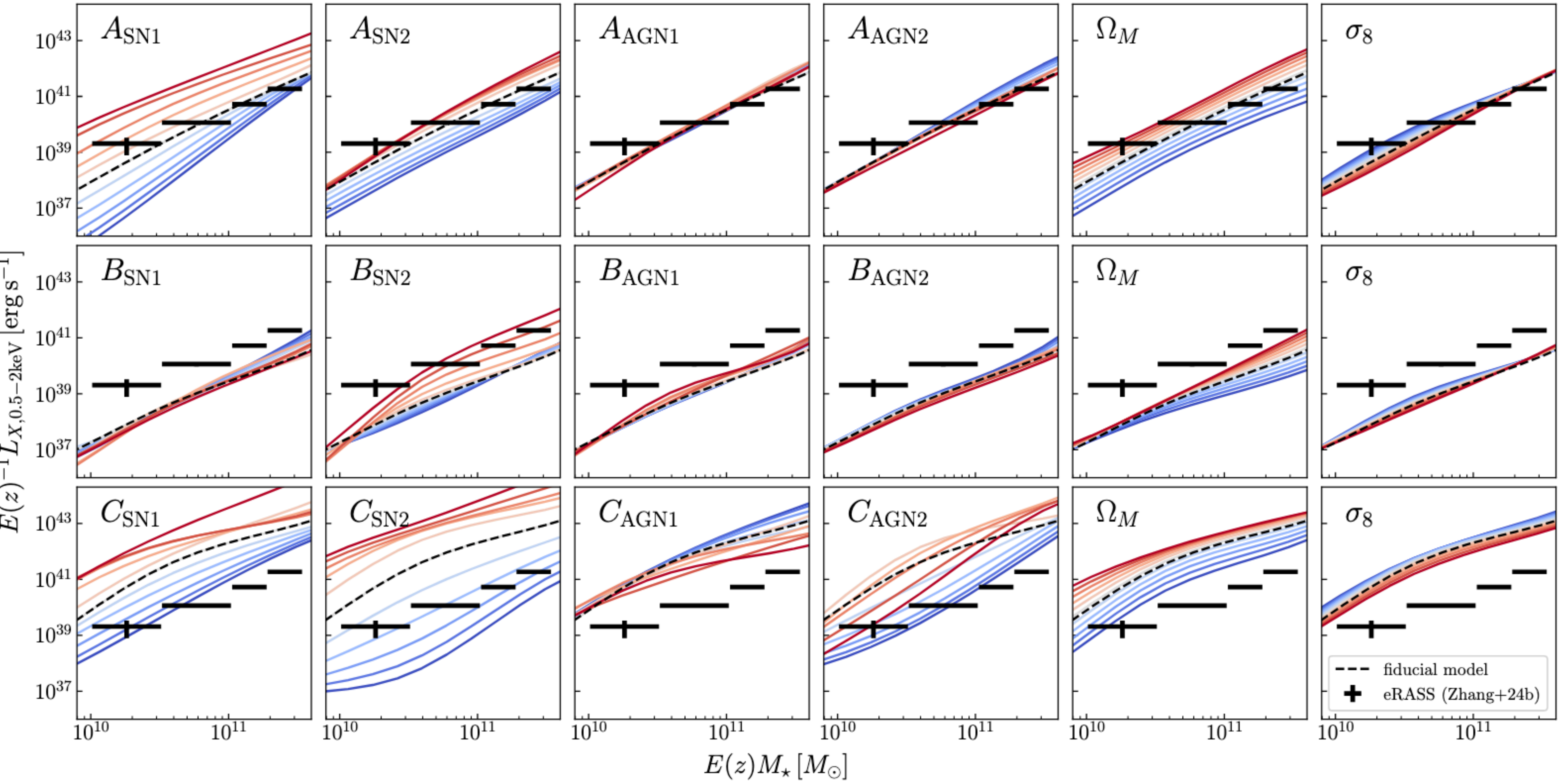
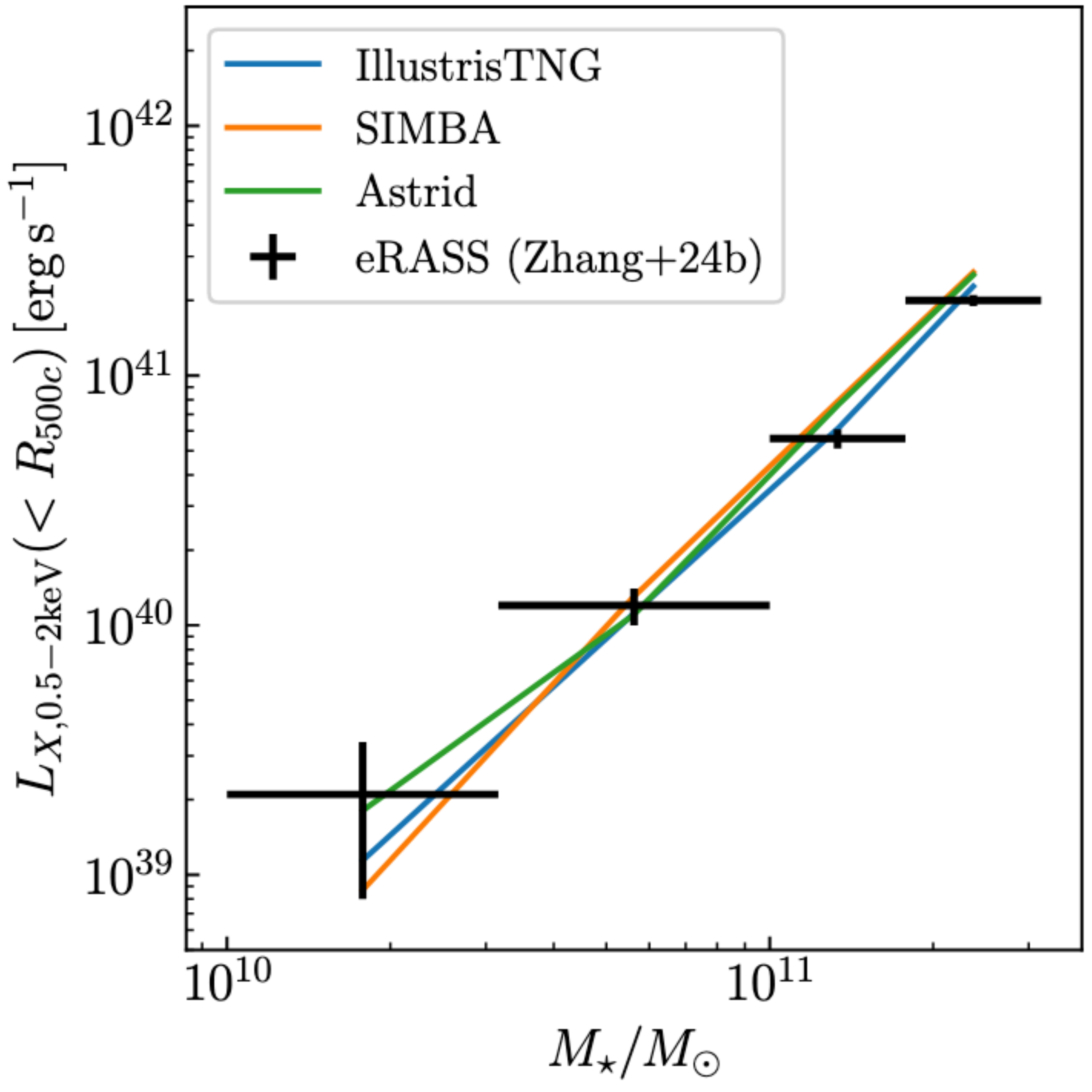


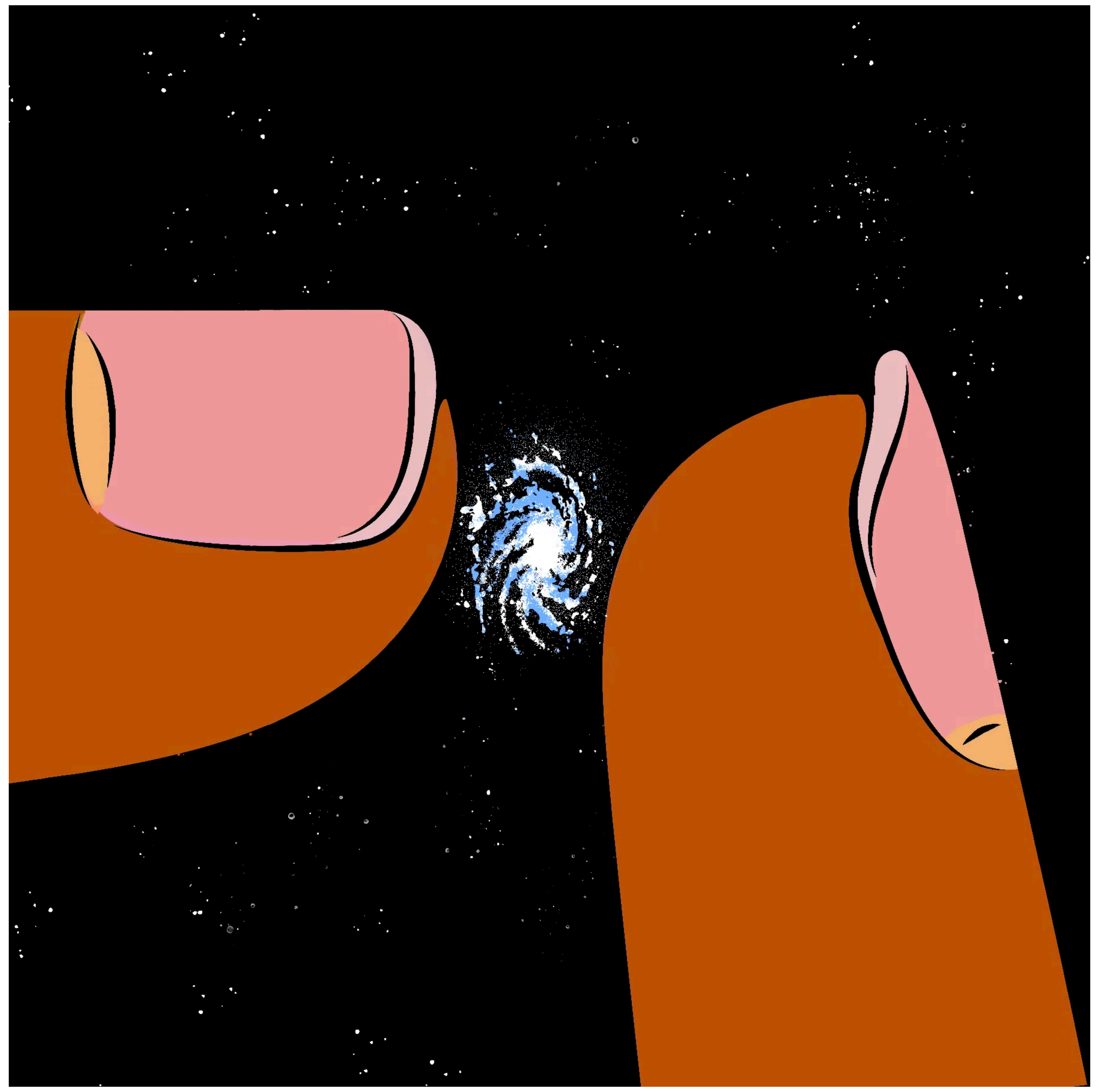
Table 1. Best-fit Feedback Parameters

Parameter	TNG	SIMBA	Astrid
SN1	$2.07^{+0.35}_{-0.34}$	$3.45^{+0.35}_{-0.38}$	$2.07^{+0.82}_{-1.00}$
SN2	$0.6^{+0.13}_{-0.07}$	$1.63^{+0.20}_{-0.05}$	$0.78^{+0.13}_{-0.13}$
AGN1	$2.18^{+1.56}_{-1.14}$	$1.79^{+0.55}_{-0.80}$	$2.42^{+0.84}_{-1.39}$
AGN2	$1.15^{+0.53}_{-0.54}$	$1.62^{+0.27}_{-0.53}$	$0.67^{0.65}_{-0.30}$



Cosmology with one galaxy?

Villaescusa-Navarro+ ‘22b



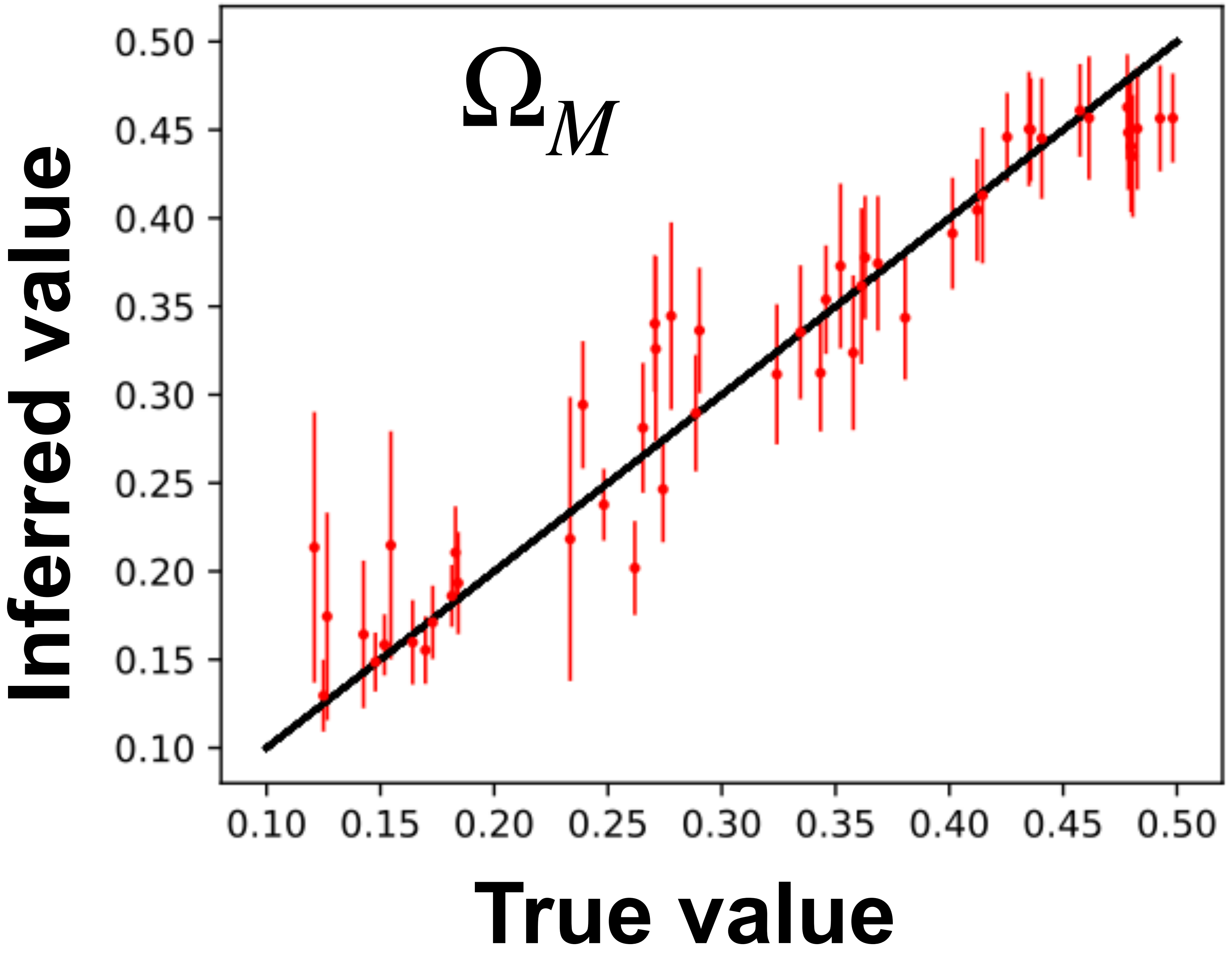
- Train NN w/ 2000 CAMELS LH set; Likelihood-free inference on params.
- Possible to infer Ω_M with ~10% precision from just one galaxy
- Link btw cosmology and astrophysics via ***low-dim manifold*** of gal properties
- New possibilities for connecting cosmology and galaxy formation with AI

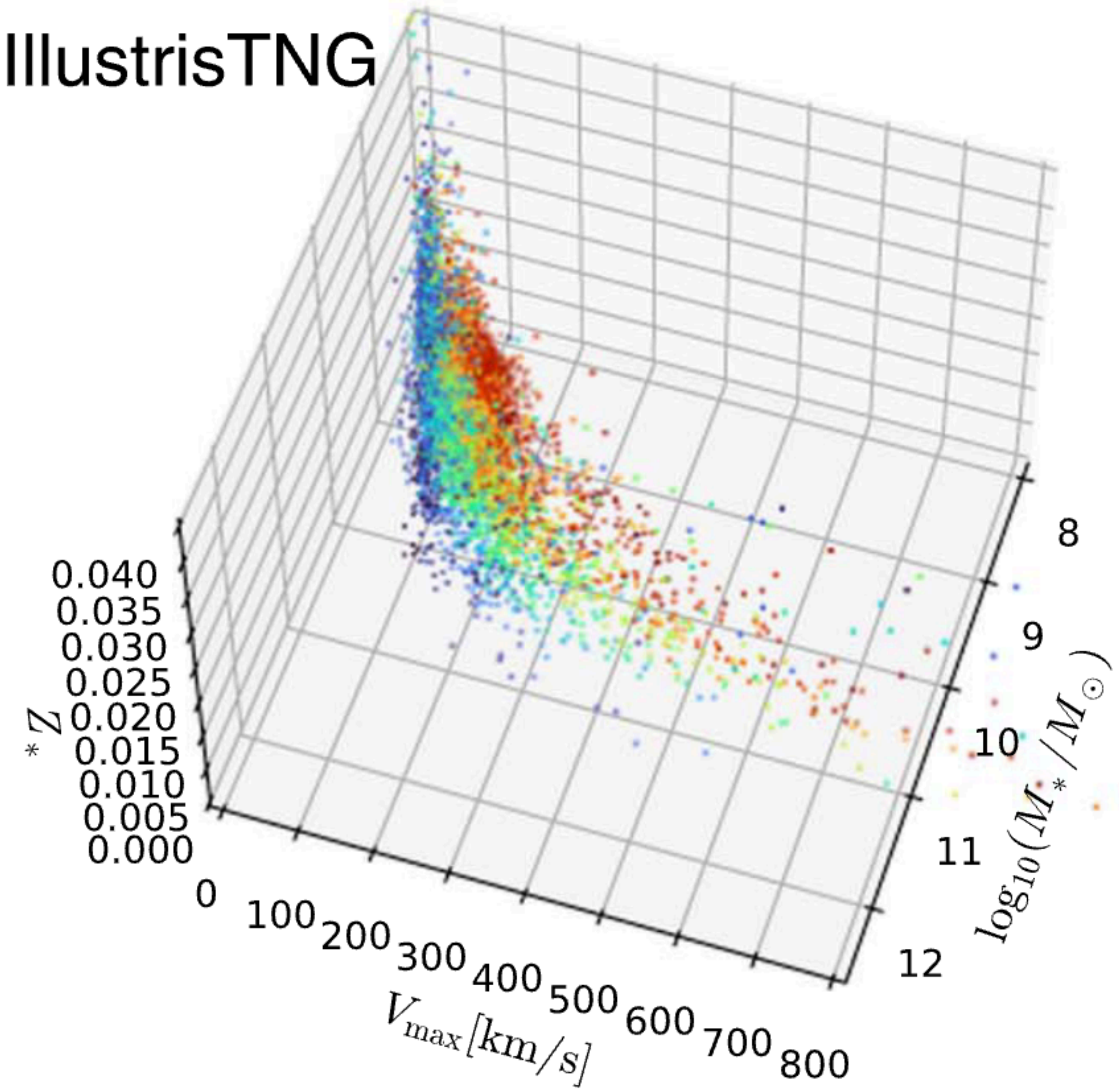
$$\Omega_m \in [0.1, 0.5]$$

$$\sigma_8 \in [0.6, 1.0]$$

$$A_{SN1}, A_{AGN1} \in [0.25, 4.0]$$

$$A_{SN2}, A_{AGN2} \in [0.5, 2.0]\ ,$$





Concerns & Issues of using AI models

1. Interpretability and Explainability （解釈可能性、説明可能性）

- Black-box problem: Complex AI models lack transparency in their decision-making processes.
- Difficulty in verifying scientific consistency or physical interpretability of results.
- Uncertainty whether the inferred cosmo. params or feedback mechanisms are physically meaningful.

2. Overfitting and Generalization Issues （過学習 & 汎化）

- AI models, particularly those with many parameters, risk overfitting training data.
- Training data set isn't representative for *generalization*.
- Incorrect cosmo. inferences due to poor generalization from simulated data to actual obs data.

3. Data *Bias* and *Quality;* Extrapolation limit

- AI models heavily depend on training data quality and completeness.
- Bias in training data (selection bias, obs. bias) may propagate and amplify errors.
- Biased predictions if simulations or training datasets are not fully representative.

4. Loss of Physical Insight （因果関係 & 相関関係の混同）

- Excessive reliance on AI can shift the focus from understanding physical processes to mere predictive accuracy.
- Risk of substituting physical understanding with model performance.
- Potential stagnation in theoretical advancements if AI predictions are not complemented by deeper physical insights.

5. Significant Computational Cost and Scalability （計算コスト & スケーラビリティ）

- Scaling AI-based analysis to petabyte-scale datasets (LSST, SKA) poses significant practical challenges.
- Increased cost and complexity in infrastructure needed to support large-scale ML applications.

6. Reproducibility and Robustness （検証の難しさ）

- Difficulty reproducing AI results due to stochastic training procedures or proprietary algorithms.
- Vulnerability of AI models to subtle changes in input data or initial conditions.
- Challenges in scientific reproducibility, validation, and robustness testing of cosmological results.

7. Ethical and Sociological Concerns （倫理学的・社会学的課題）

- Potential for unintentional ethical issues — open access, data privacy, fairness, and model biases.
- Unequal access to resources needed for AI research can exacerbate disparities.
- Barriers to collaboration or biases introduced by unequal access to AI methods and data resources.

8. Misuse or Over-Reliance on AI Results

- Scientists or policymakers might accept AI-driven results without sufficient critical scrutiny.
- Overconfidence in AI predictions without adequate error analysis or understanding of limitations.
- Misguided scientific or policy decisions based on inadequately vetted AI-driven cosmological findings.