How can we learn about the physics of galaxy formation from Line Intensity Mapping?

rachel somerville

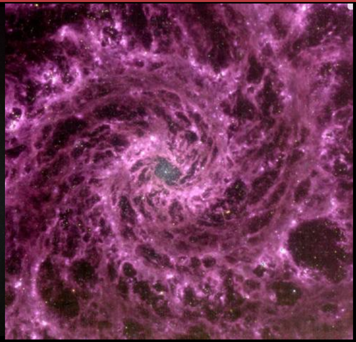
Center for Computational Astrophysics

FLATIRON INSTITUTE

SIMONS

Learning the Universe

SMAUG


CCA-NYU
SAM-LIM group

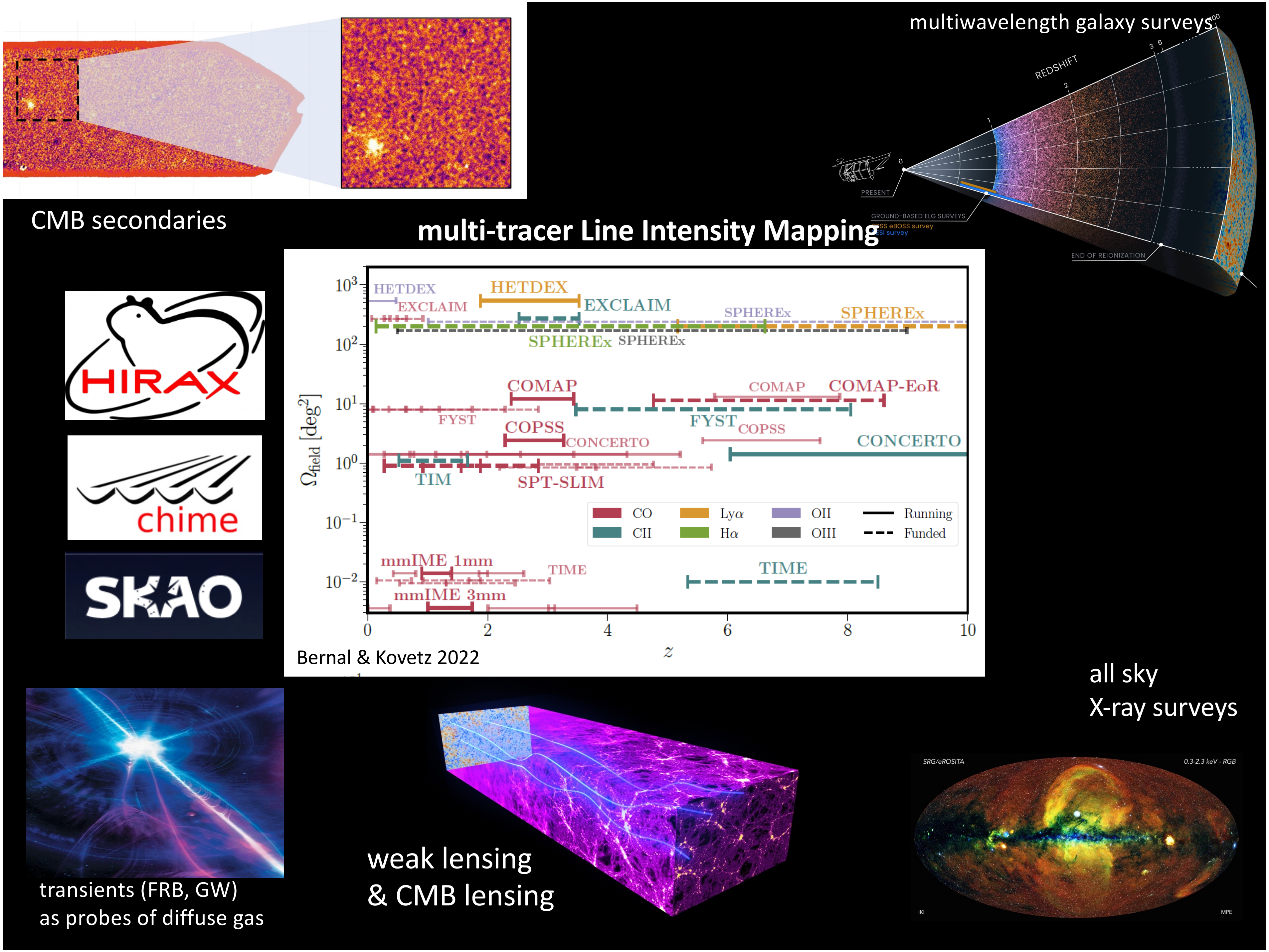
Anthony Pullen

Patrick Breysse

Anirban Roy

Shengqi Yang

John Forbes, Manush Manju, Gergö Popping, rachel somerville,
Laura Sommovigo, Aidan Subrahimovic, Rakshitha Thaman,
Aaron Yung, Yucheng Zhang

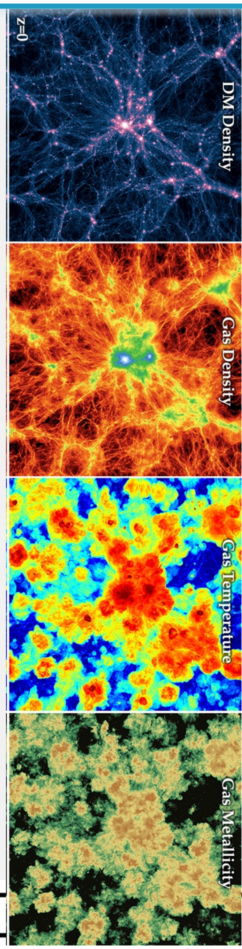


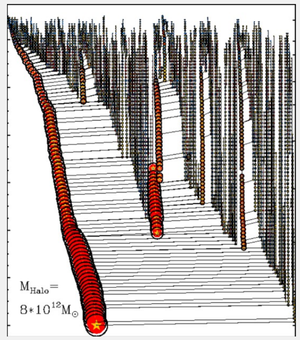
CMB secondaries
multiwavelength galaxy surveys
REDSHIFT
PRESENT
GROUND-BASED ELG SURVEYS
END OF REIONIZATION
multi-tracer Line Intensity Mapping
HIRAX
chime
SKAO
HETDEX
EXCLAIM
SPHEREx
COMAP
COMAP-EoR
FYST
COPSS
CONCERTO
TIM
SPT-SLIM
mmIME 1mm
mmIME 3mm
TIME
CO
CII
Lyα
Hα
OII
OIII
Running
Funded
$\Omega_{\rm field}$ [deg^2]
z
Bernal & Kovetz 2022
all sky
X-ray surveys
SRG/eROSITA
0.3-2.3 keV - RGB
IKI
MPE
transients (FRB, GW)
as probes of diffuse gas
weak lensing
& CMB lensing

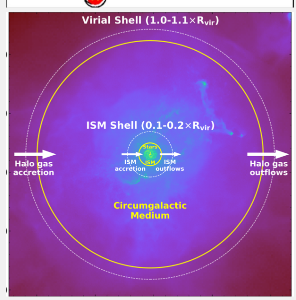
outline for this talk

- physics based galaxy formation models: status & open questions
- predictions for LIM (biased view)
- way forward

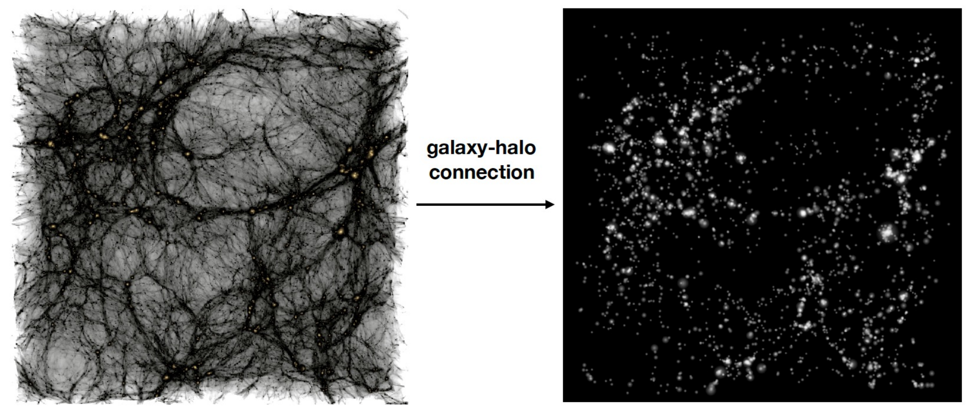
Somerville & Davé 2015
Naab & Ostriker 2017
Crain & van de Voort 2023







Wechsler & Tinker 2018



← physical models

empirical models →

Hydrodynamical Simulations	Semi-analytic Models	Empirical Forward Modeling	Subhalo Abundance Modeling	Halo Occupation Models
solve PDEs for DM, stars, gas sub-grid models for SF, feedback, BH, etc	solve ODEs for gas flows between global reservoirs; recipes for SF, BH growth, feedback, etc	assume gas inflows track DM; empirical recipes for SF, etc	mapping from DM (sub)-halos to galaxy properties	model for n_{gal} as function of halo mass (or other halo properties)

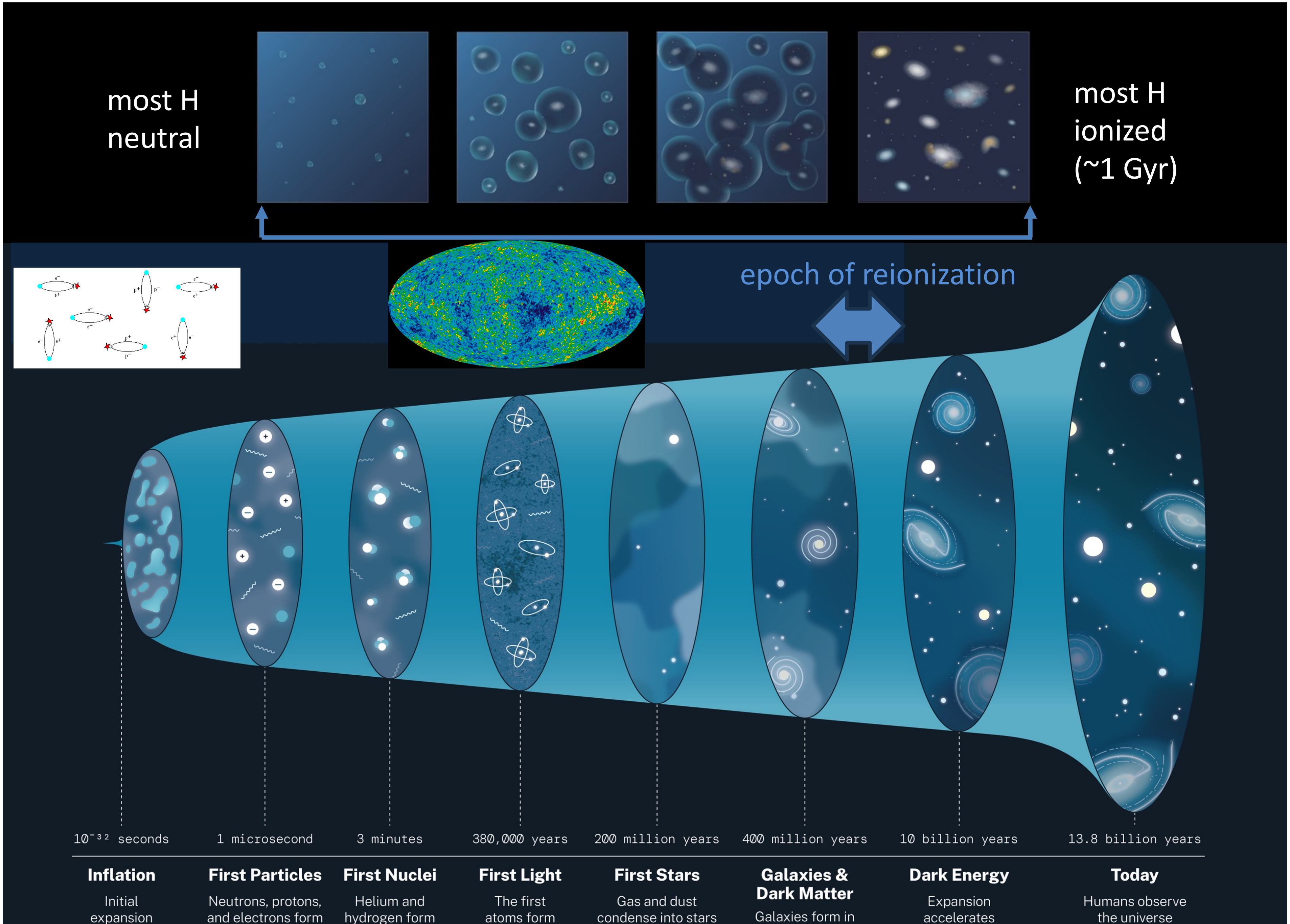
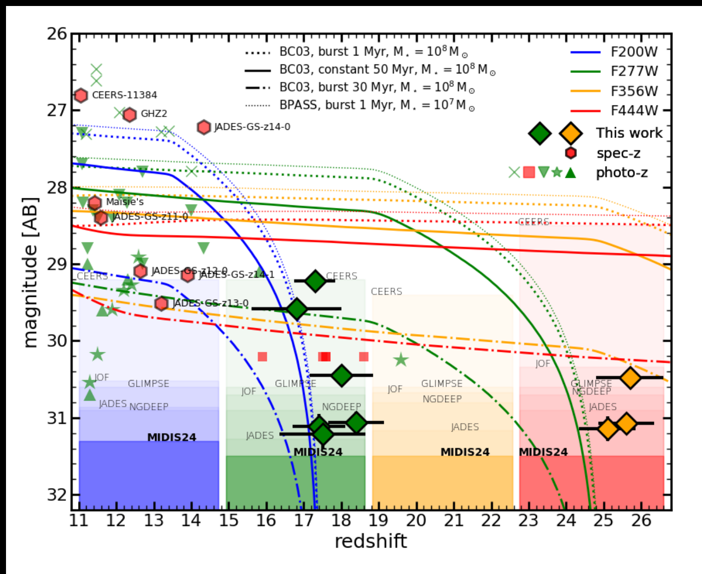


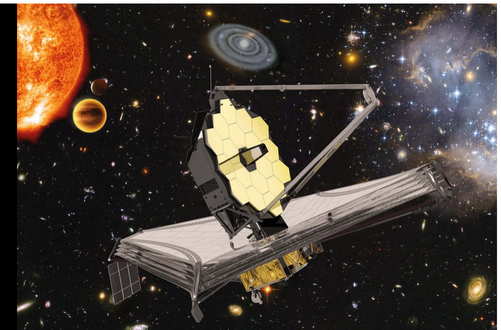
image credit: NASA

Perez-Gonzalez et al. 2025



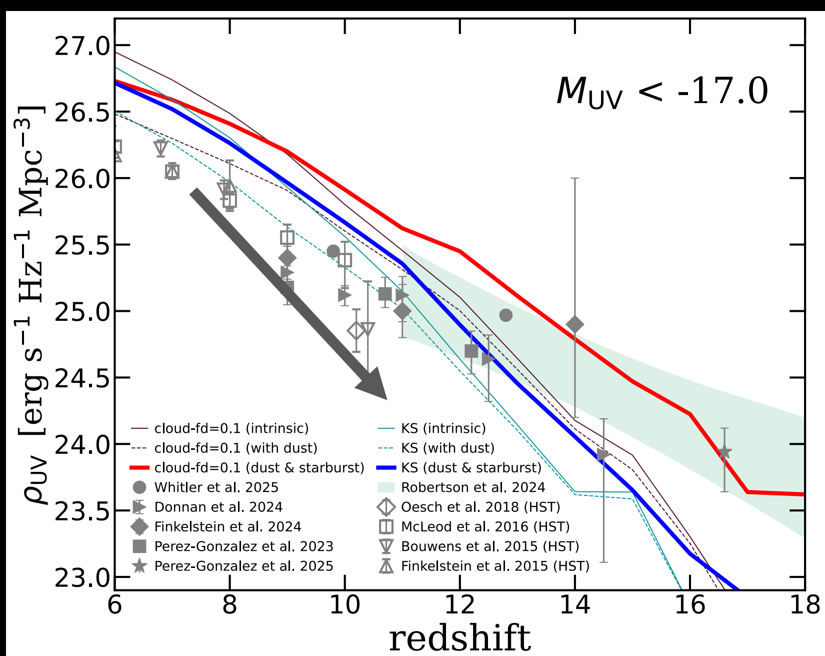
JWST has discovered a more abundant population of UV-luminous galaxies at z>8 than expected from pre-JWST observations or pre-launch models

--density modulated star formation efficiency, bursty star formation, evolving IMF, dust ejection?



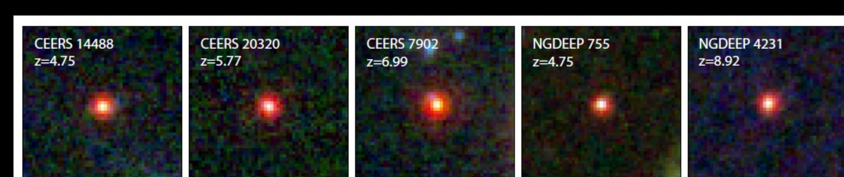
ionizing photon 'overproduction' crisis (Munoz+2024) or 'all bark and no bite' (Papovich+2025)?

cosmic UV luminosity density



rss et al. 2025

discovery of abundant low-luminosity high-z AGN population (many X-ray weak) – role of AGN in H reionization (Madau et al. 2024)?

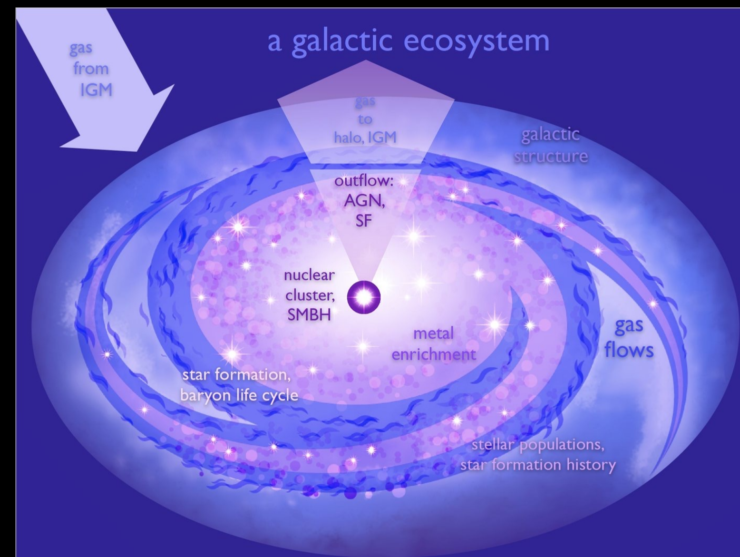


Kocevski et al. 2024

cosmic baryon cycle: unveiling the drivers of galaxy growth*

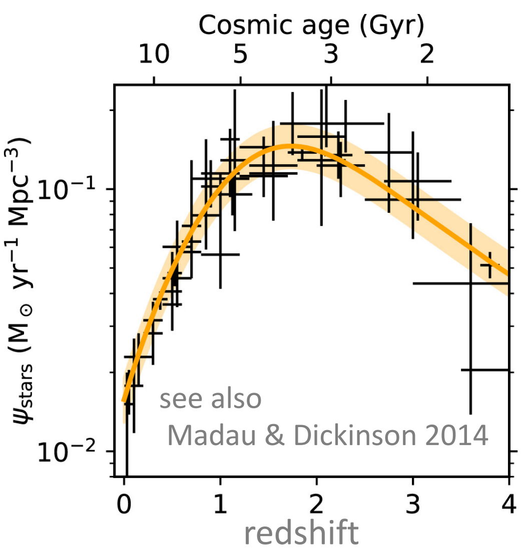
-how do mass, metals, and energy cycle in and out of the ISM/CGM?

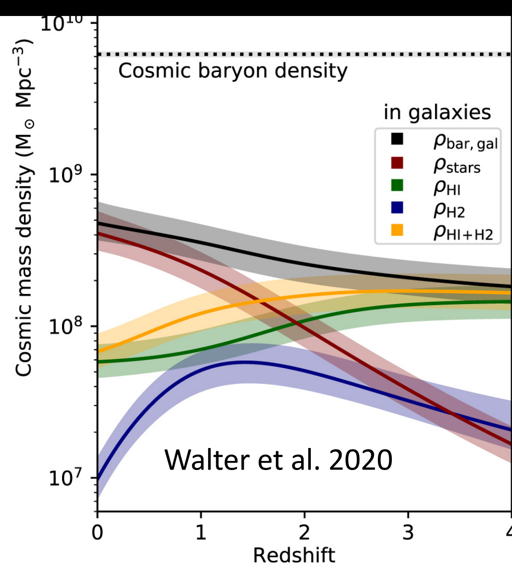
-*which physical processes drive this evolution?*

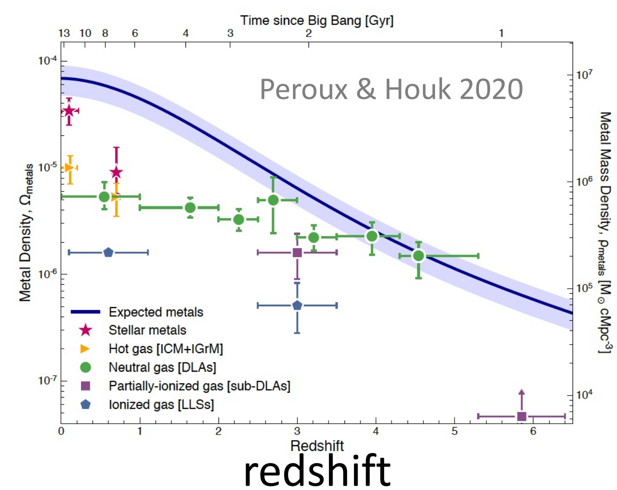


PHANGS collaboration, Design: Daniela Leitner

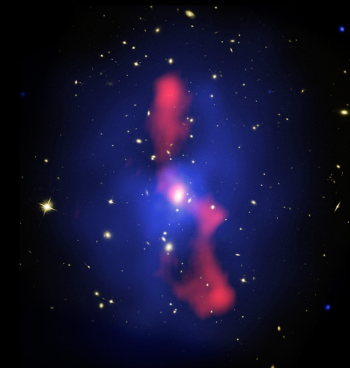
*priority area identified by Astro2020 Decadal Survey

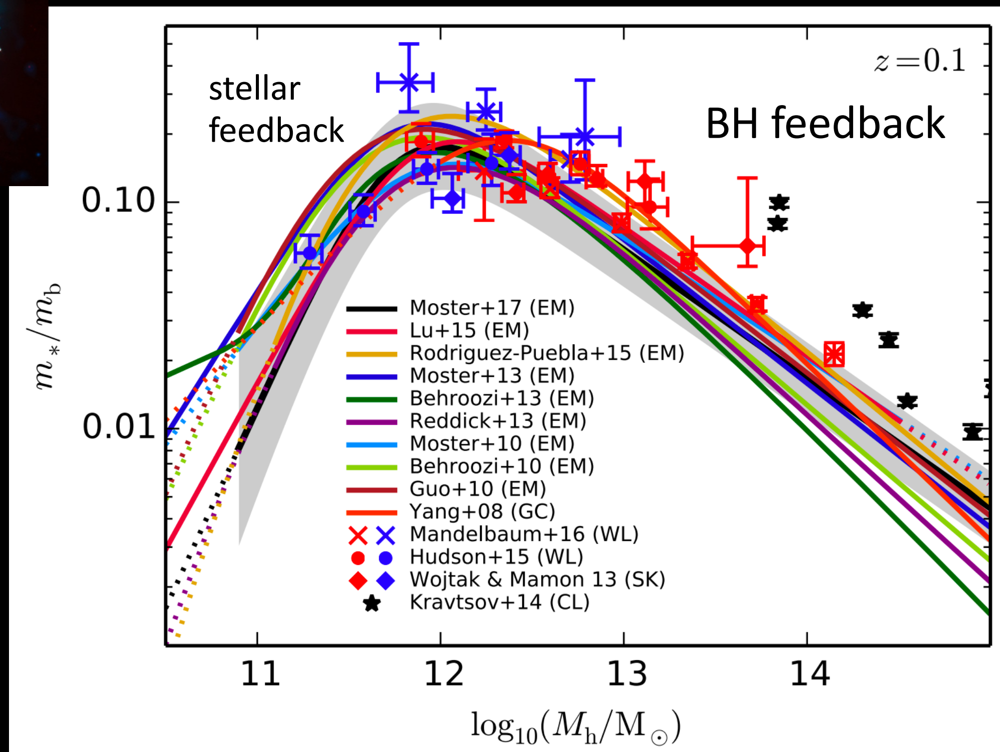






how do stellar and black hole feedback work?



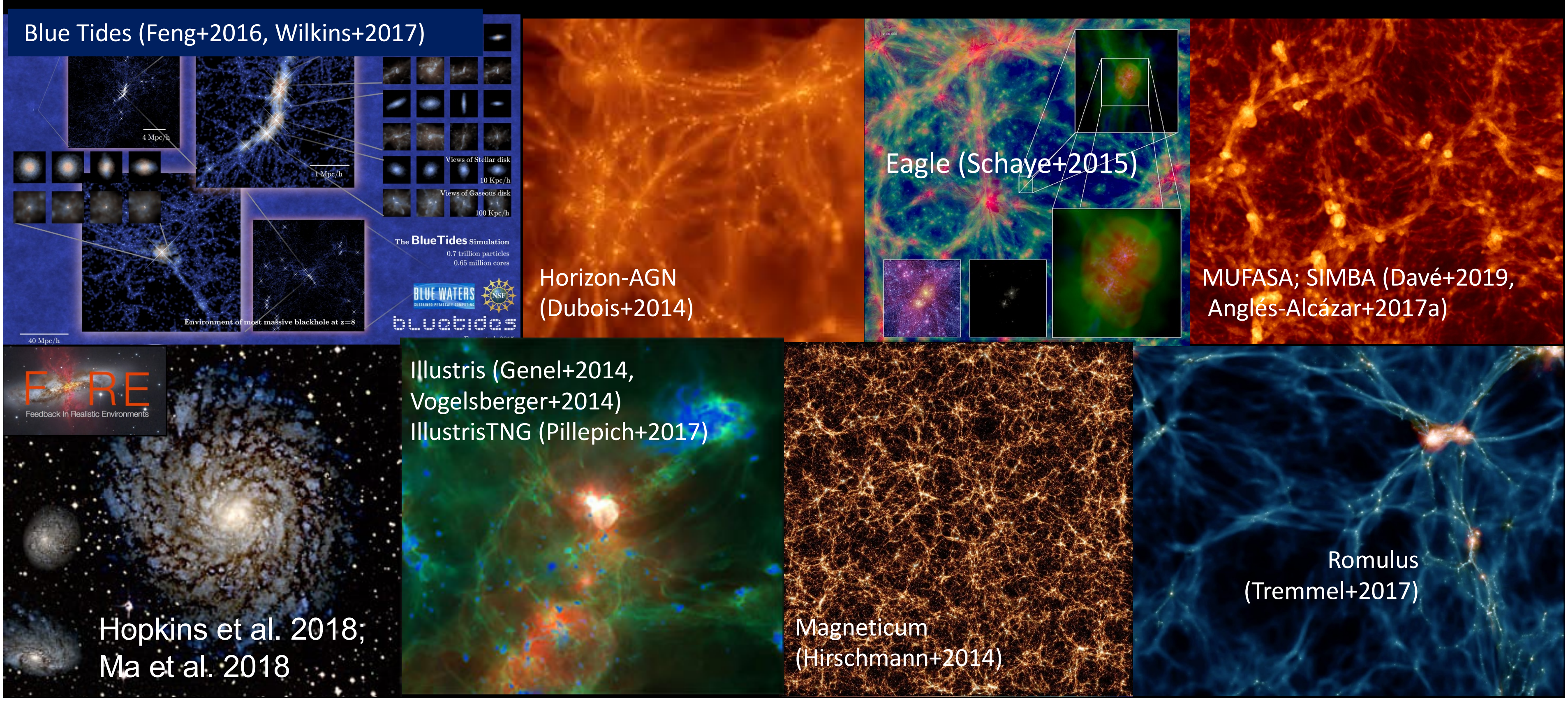


fraction of total baryonic mass in stars

mass of the dark matter halo

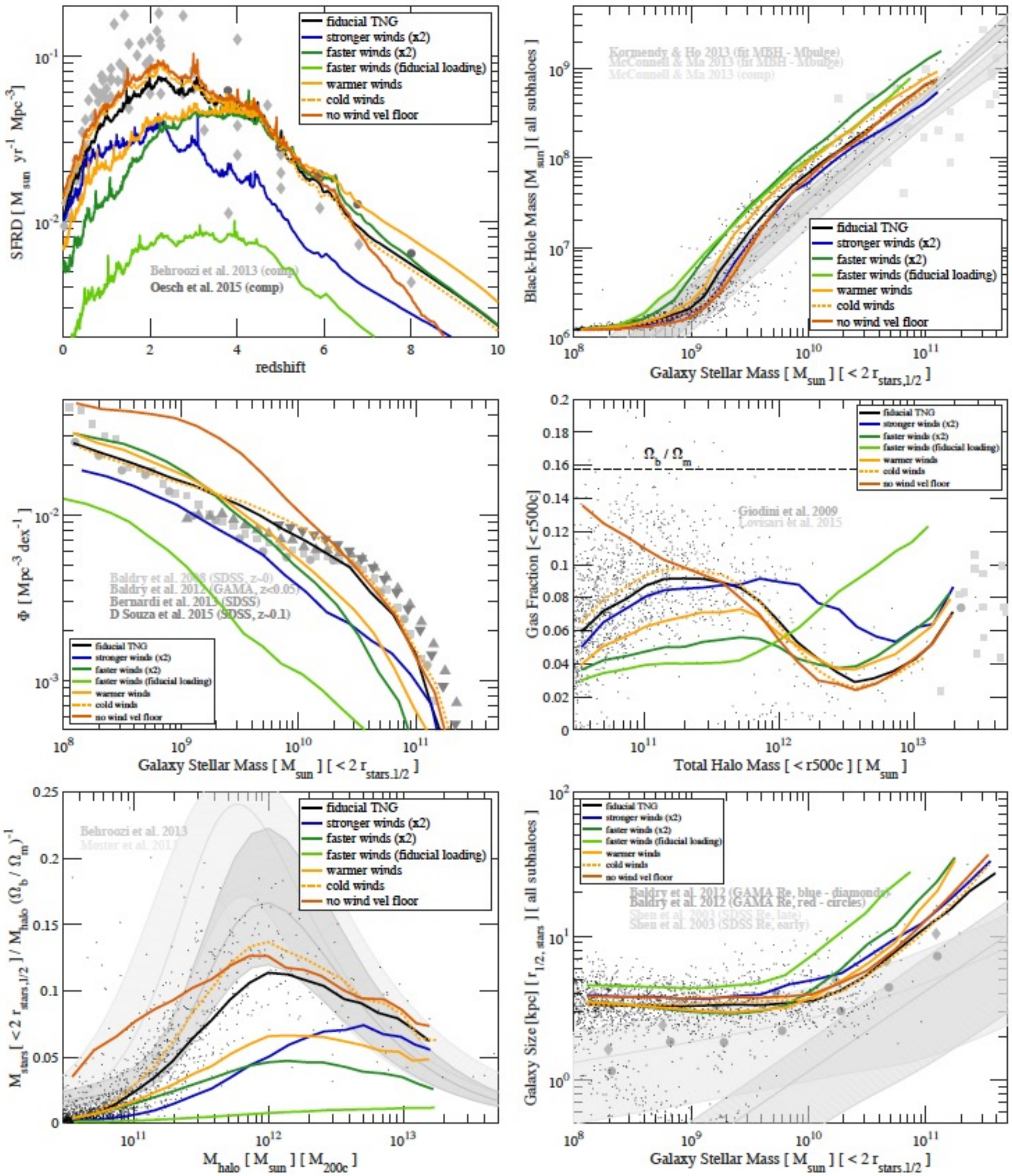
cosmological hydrodynamic simulations

equations of gravity, hydrodynamics, thermodynamics solved numerically for particles or grid cells representing dark matter, gas, stars, and black holes



an incomplete list of processes that currently require sub-grid treatment in large-volume cosmological simulations

- the multiphase interstellar medium (ISM)
- star formation (conditions for its onset, and its efficiency)
- stellar feedback (stellar winds, radiation, SNae)
- chemical evolution and metal diffusion
- black hole seeding
- black hole accretion
- black hole feedback (kinetic, thermal, radiation)

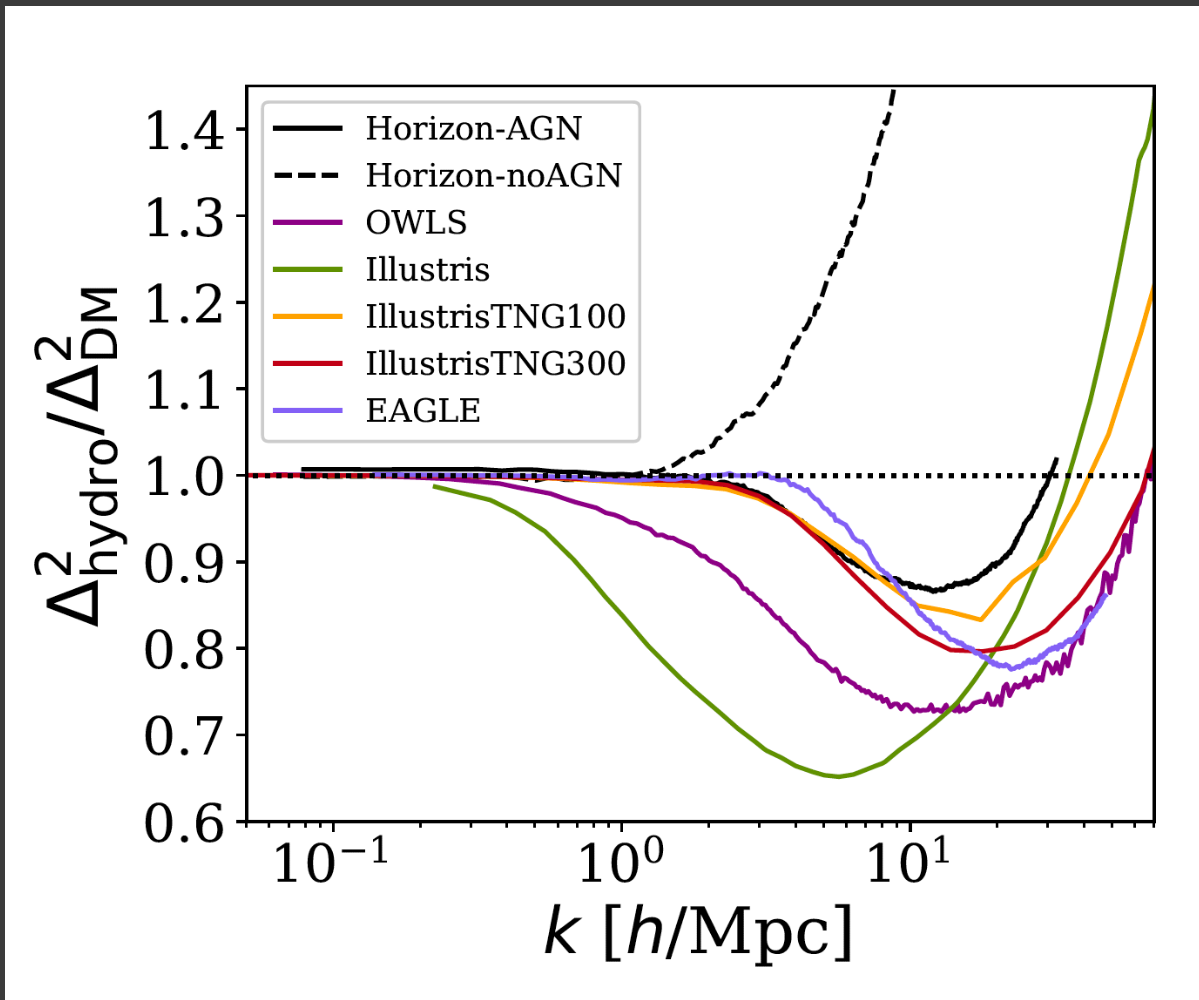


parameters are tuned to achieve a "best" match (by eye) to a set of calibration observations

other calibrated parameters:
- stellar wind mass loading and velocity
- wind metal loading
- SF timescale (efficiency)
- BH accretion efficiency
- parameters controlling mode & effect of BH FB

IllustrisTNG
Pillepich et al. 2018

the distribution of matter on Mpc-10 Mpc scales is altered by baryonic processes



differences due to different implementations of baryonic feedback are >> precision needed to constrain 'fundamental physics'

Chisari et al. 2018

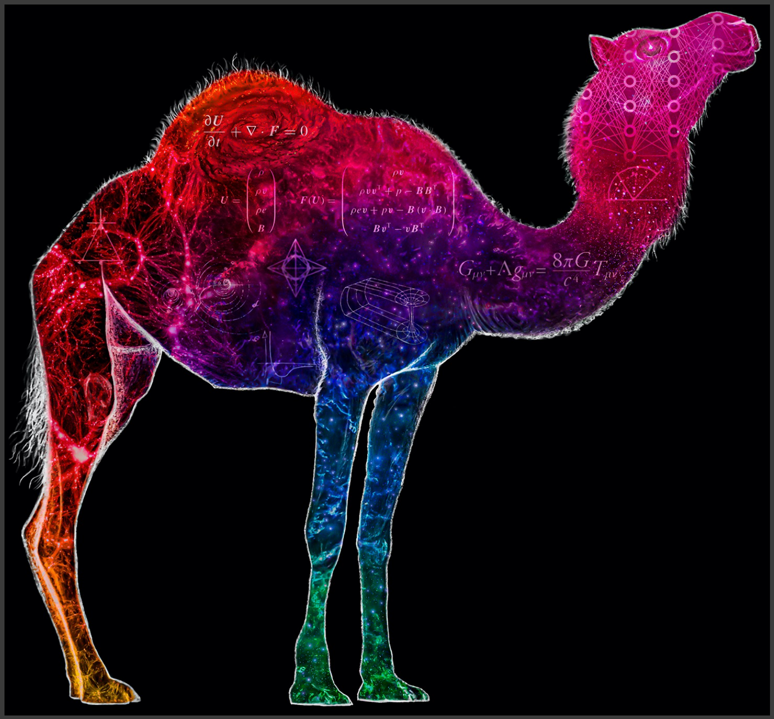
CAMELS

F. Villaescusa-Navarro

https://www.camel-simulations.org

Shy Genel

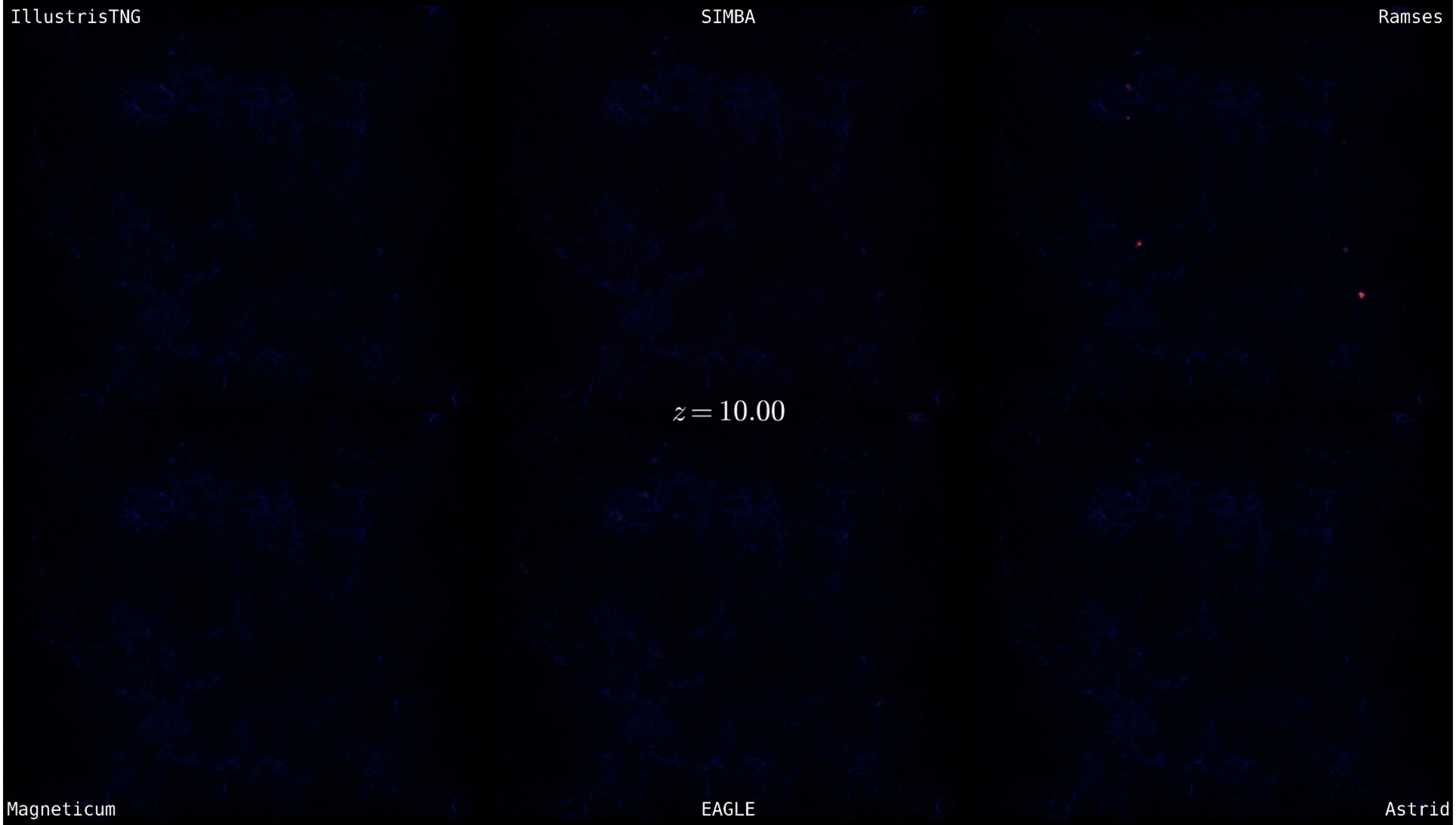
Daniel Angles-Alcazar



Cosmology and Astrophysics with MachinE Learning Simulations

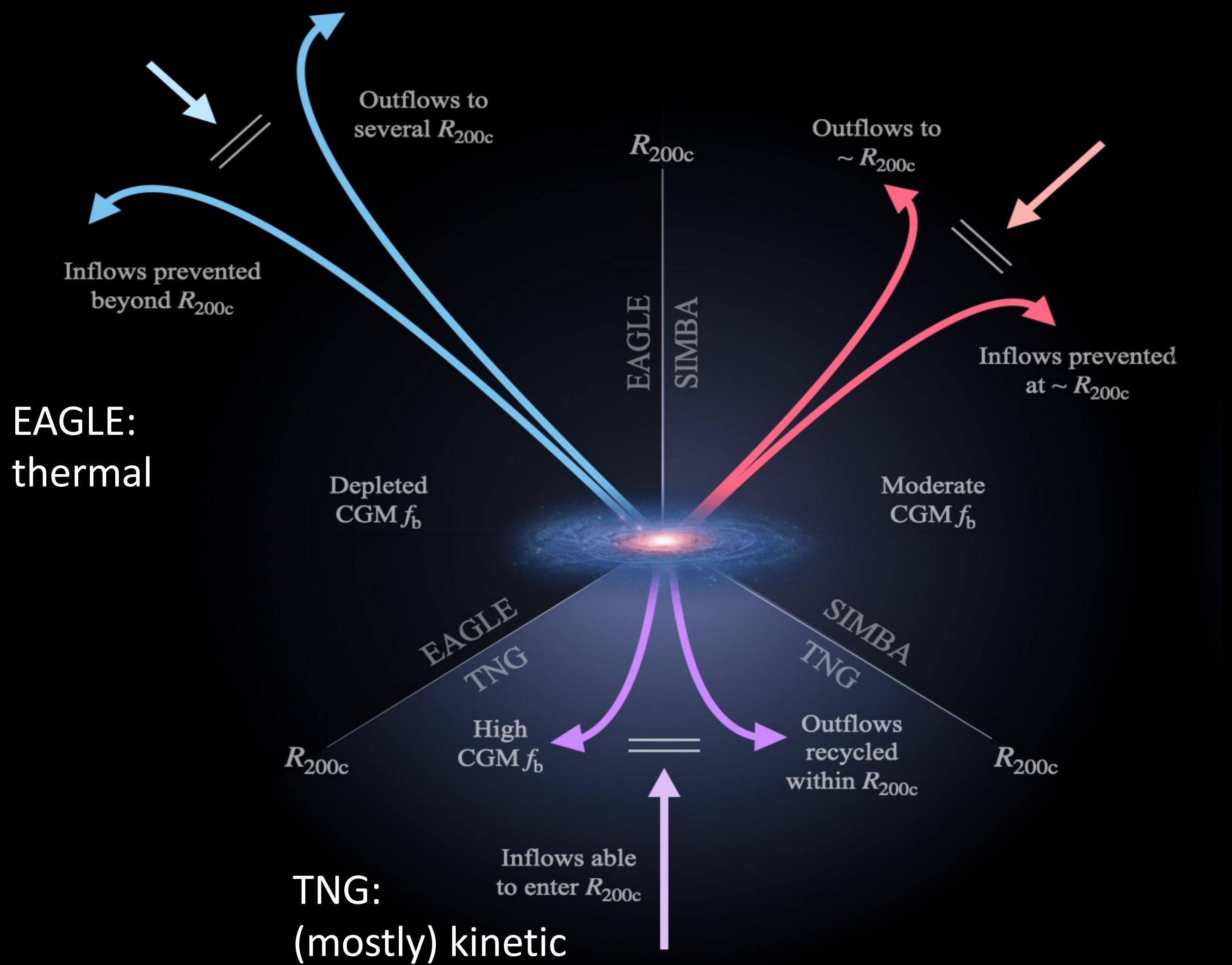
Villaescusa-Navarro et al. 2021

- suite of thousands of state-of-the-art hydrodynamic sims
- IllustrisTNG, SIMBA, Astrid, Magneticum, Swift-EAGLE, Ramses, Enzo, Crocodile, Obsidian
- 6 parameters: $\{\Omega_m, \sigma_8, A_{\mathrm{SN1}}, A_{\mathrm{SN2}}, A_{\mathrm{AGN1}}, A_{\mathrm{AGN2}}\}$
- extended parameter set for some models
- $(25/h\ \mathrm{Mpc})^3$ boxes; group/cluster zooms; $(50/h\ \mathrm{Mpc})^3$ boxes in progress
- Designed for machine learning applications
- all data public!



video from CAMELS project; F. Villaescusa-Navarro

kinetic wind sub-grid: weak preventative feedback on all scales
thermal wind sub-grid: stronger preventative feedback out to beyond R_{vir}



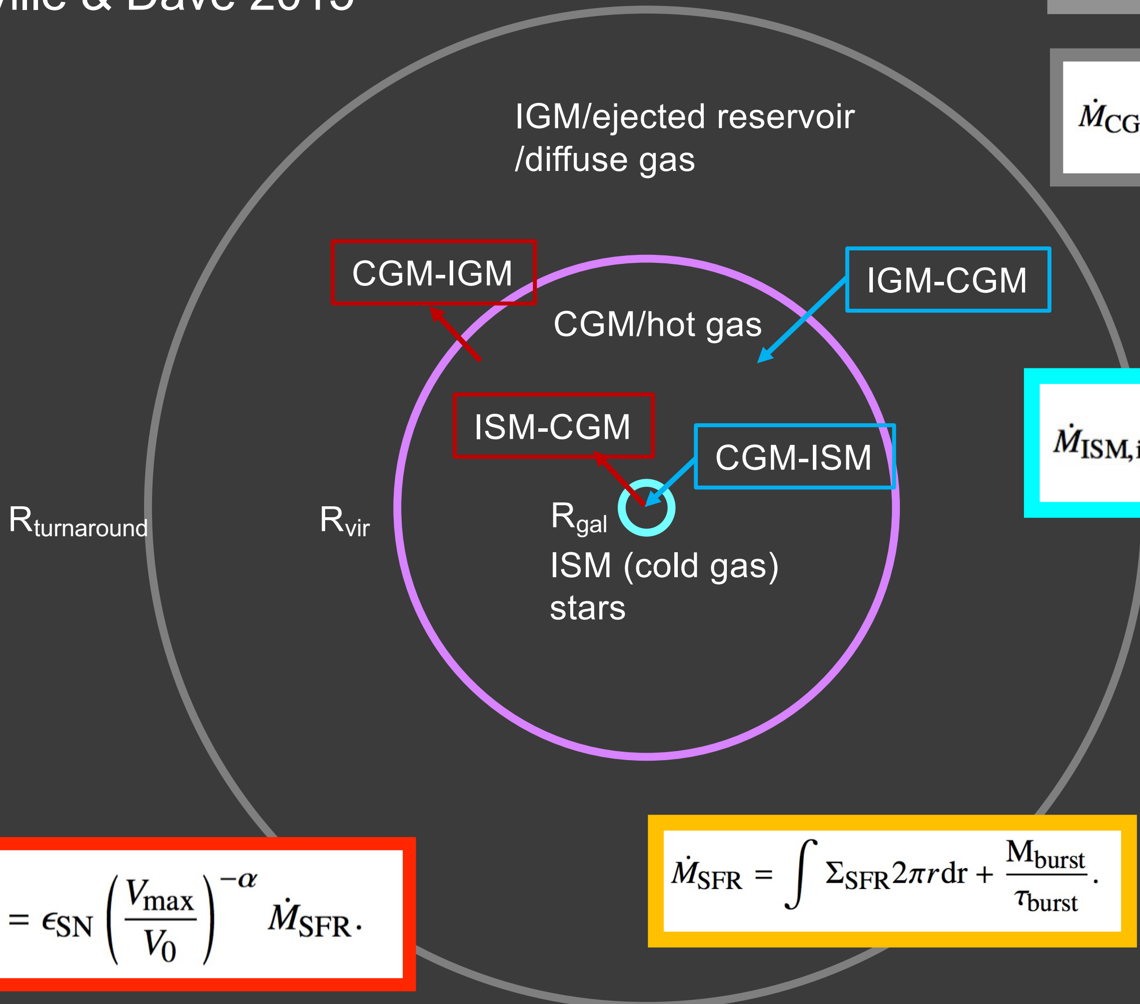
Wright, rss et al. 2024

traditional semi-analytic/flow/regulator models

Somerville & Davé 2015

$$\dot{M}_{\rm CGM,in,pristine} = f_{\rm UV} f_b \dot{M}_{\rm vir}.$$

$$\dot{M}_{\rm CGM,in,recycled} = \chi_{\rm re-infall} \left(\frac{M_{\rm ejected}}{t_{\rm dyn}} \right).$$

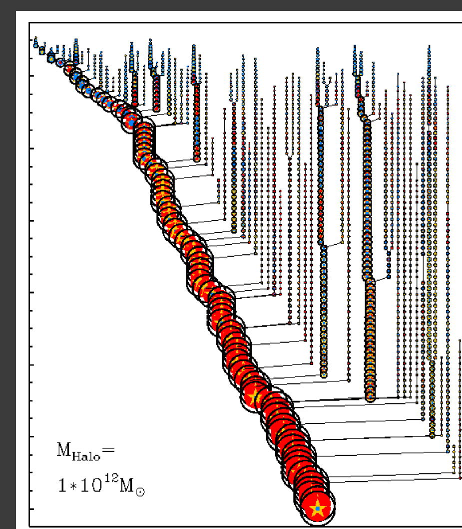


$$\dot{M}_{\rm ISM,in} = \frac{1}{2} M_{\rm CGM} \frac{R_{\rm cool}}{R_{\rm vir}} \frac{1}{t_{\rm dyn}}$$

$$\dot{M}_{\rm SFR} = \int \Sigma_{\rm SFR} 2\pi r {\rm d}r + \frac{\rm M_{burst}}{\tau_{\rm burst}}.$$

$$\dot{M}_{\rm ISM,out} = \epsilon_{\rm SN} \left(\frac{V_{\rm max}}{V_0} \right)^{-\alpha} \dot{M}_{\rm SFR}.$$

$$\dot{M}_{\rm CGM,out} = [1.0 + (V_{\rm vir}/V_{\rm eject})^{\beta}]^{-1} \times \dot{M}_{\rm ISM,out}.$$

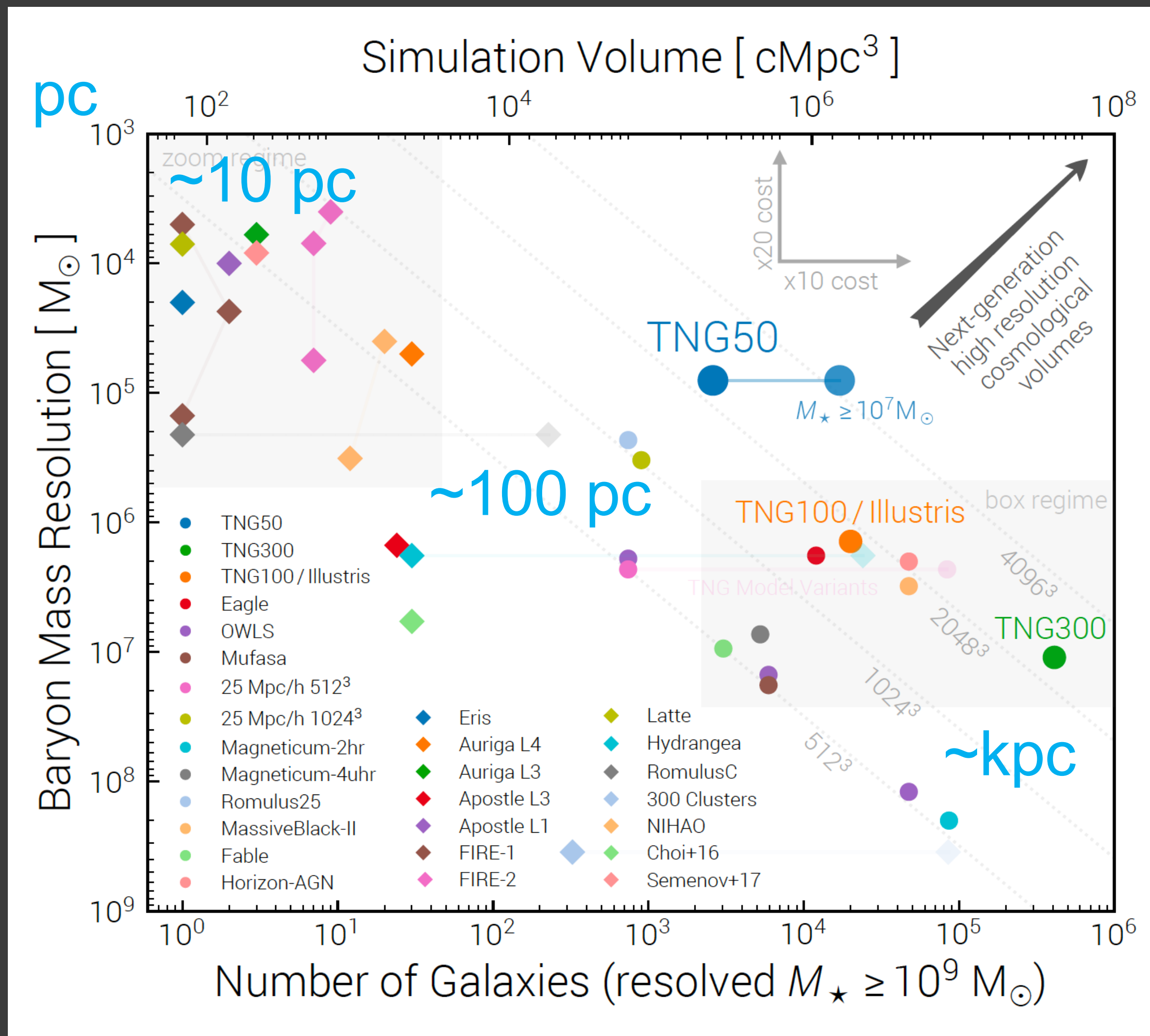


slide adapted from V. Pandya

“resolved”/ explicit physics

semi-resolved, mixed explicit+ sub-grid

sub-grid



10^9

SAMs

MTNG

FLAMINGO

Nelson et al. 2019

ΛCDM galaxy simulations report card

passing grades:

- stellar mass functions/UV/optical LF z~0-10
- reionization history
- clustering of optically selected galaxies, including dependence on color/mag/size
- qualitative correlations of galaxy properties (color, SFR, morphology) with larger scale over-density
- cold gas fractions/cold gas MF z=0
- galaxy (optical/stellar) size vs. stellar mass z=0-2ish
- galaxy color/SFR bimodality z~0-1
- CGM scale baryon (hot gas) fractions z~0

Somerville & Davé 2015; Naab & Ostriker 2017; Crain & van de Voort 2023

ΛCDM galaxy simulations report card

mixed/uncertain:

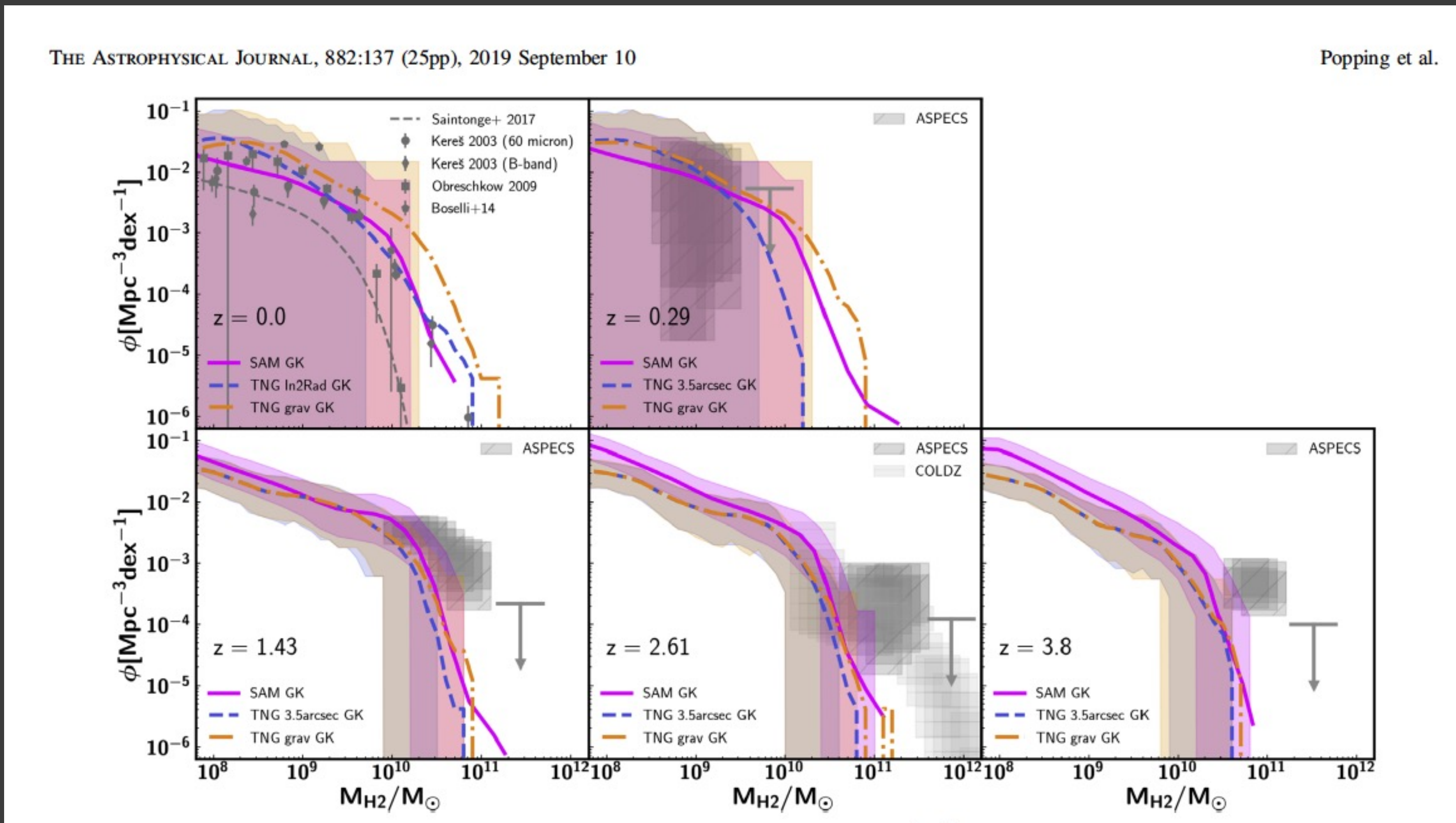
- cold gas fractions/cold gas mass functions ~cosmic noon (Popping et al. 2019; Davé+2020)
- numbers of ULIRGy galaxies (e.g. sub-mm counts, Herschel, etc) ~cosmic noon (e.g. Wang et al. 2019; Hayward et al. 2021)
- gas phase stellar mass-metallicity relation?

ΛCDM galaxy simulations: report card

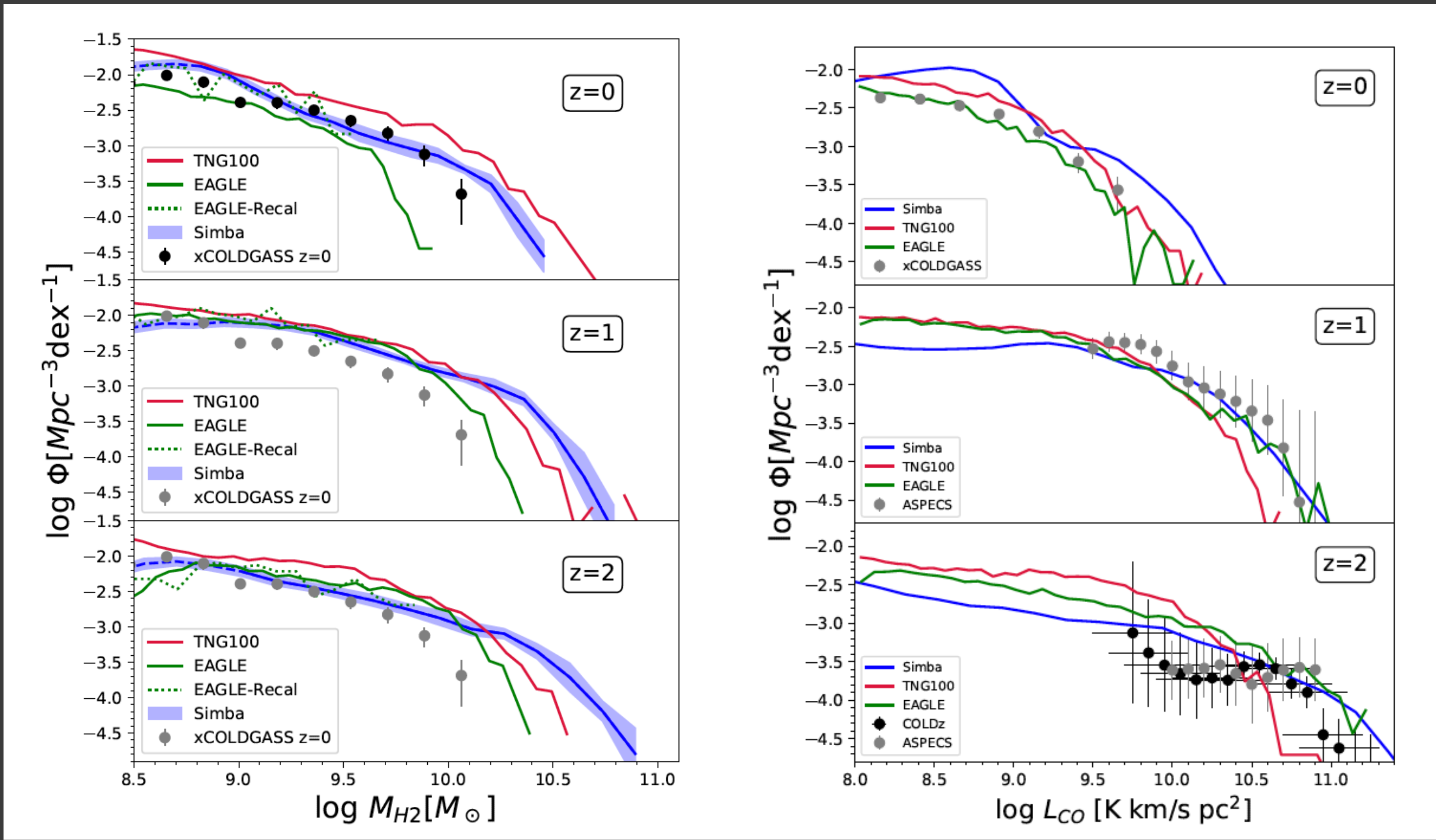
failing grades:

- galaxy UVLF z>10 (e.g. Finkelstein+22)
- diffuse IGM z<1 as probed by Lyman-alpha forest (Tillman et al. 2023)
- extended CGM as probed by Sunyaev-Zeldovich (Amodeo et al. 2021)
- numbers of massive quenched galaxies z>3-4 (Valentino+23, Lagos+24)
- AGN luminosity functions at all z (Habouzit+22)

some models have difficulty reproducing enough galaxies with large cold gas reservoirs at cosmic noon



Popping et al. 2019

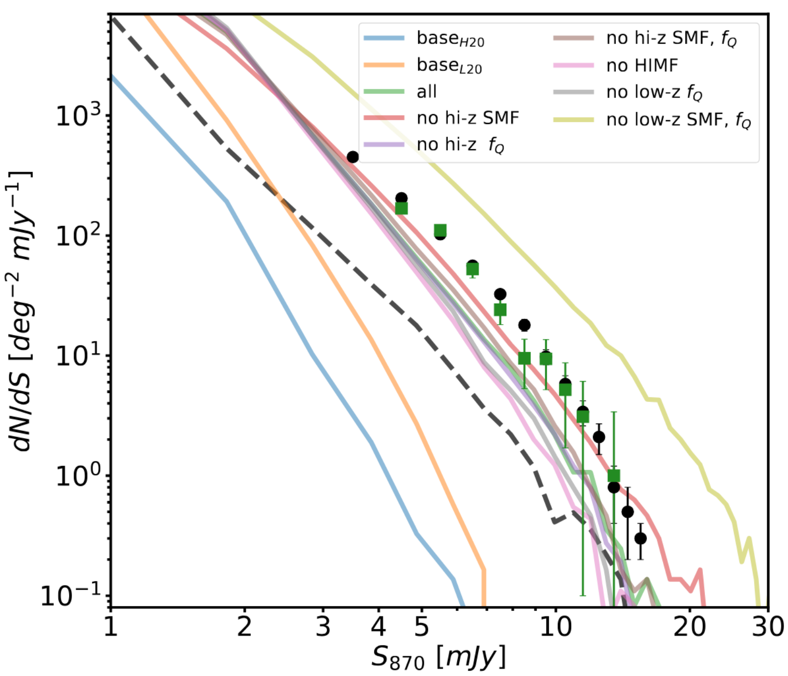


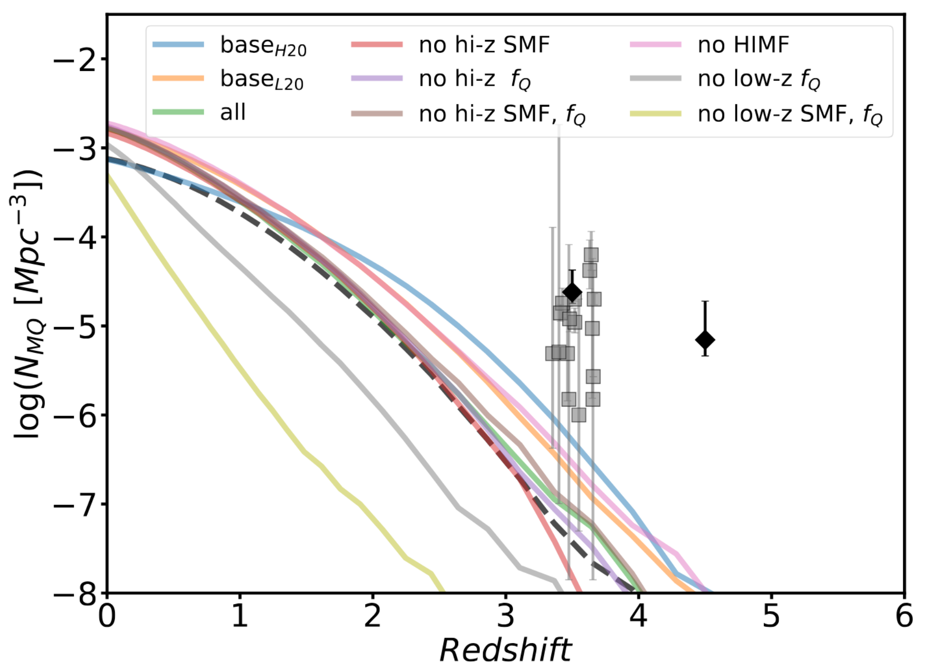
Davé et al. 2020

models that do better at reproducing highly star forming galaxies at z~2-4 have more difficulty matching the number of massive quenched galaxies

highly SF galaxies

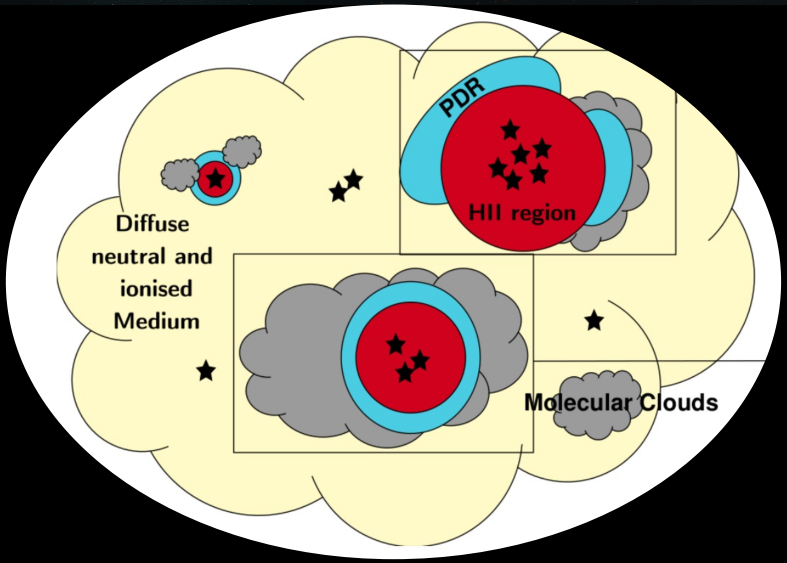
massive quenched galaxies



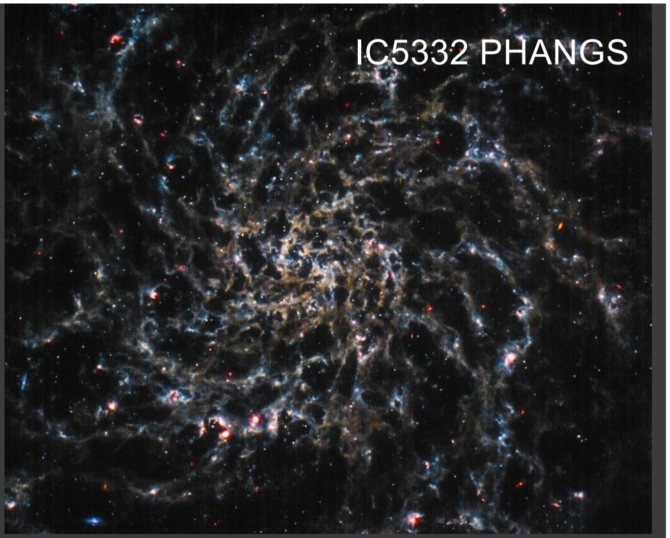


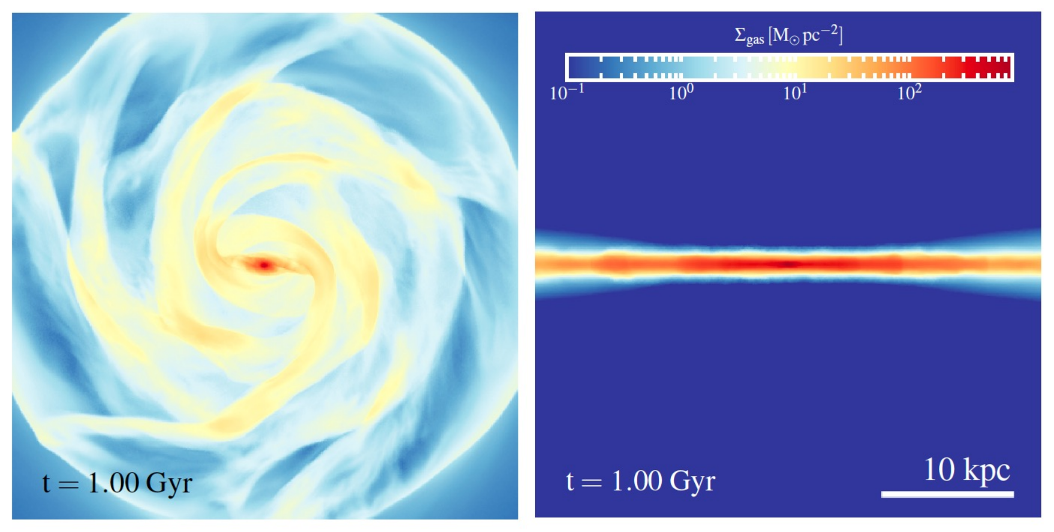
Hayward et al. 2021; Araya-Araya et al. 2025

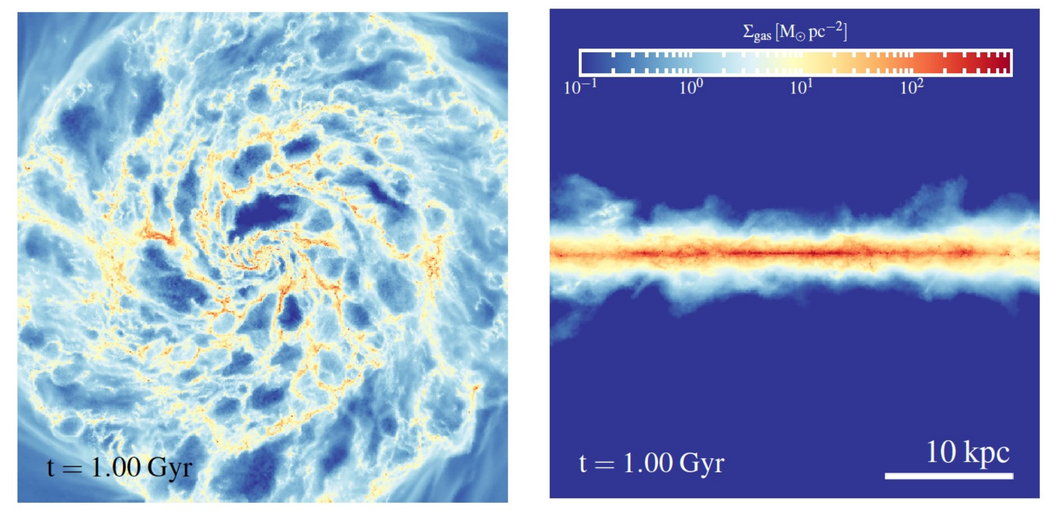
emission from the ISM

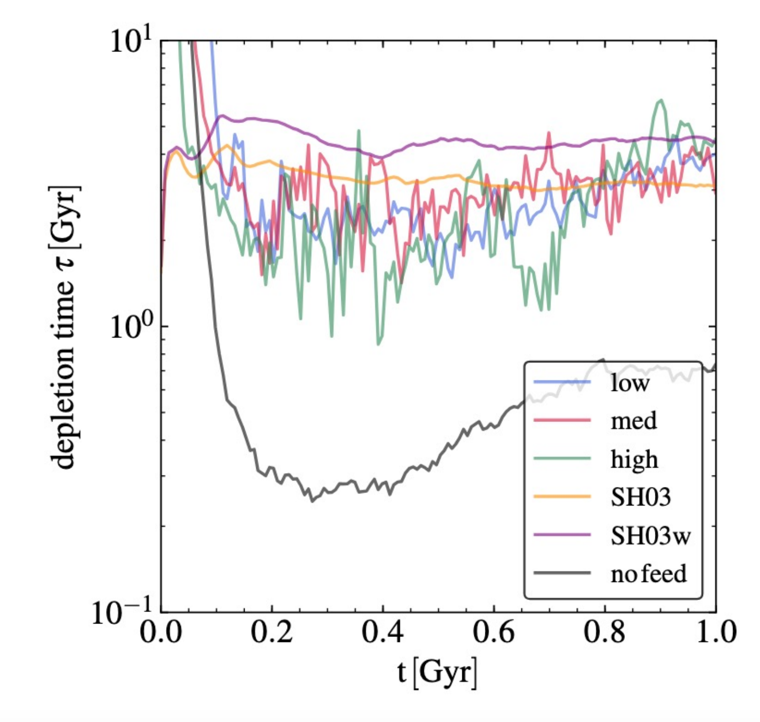


most large volume cosmo sims adopt an ‘effective equation of state’; artificially pressurizes and ‘smooths’ ISM



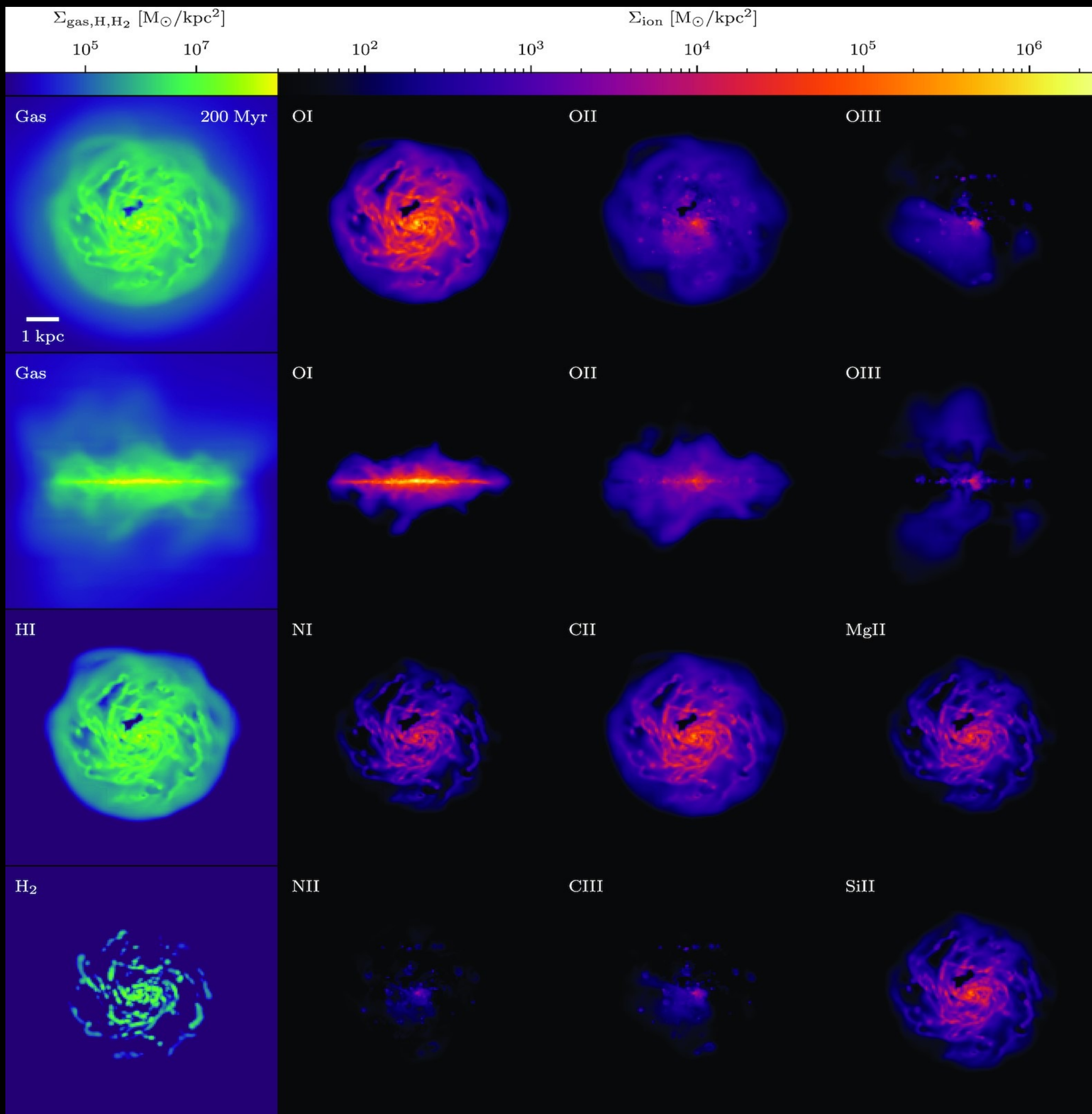






Marinacci et al. 2019

simulations with non-equilibrium thermo-chemistry + on the fly radiative transfer



RAMSES-RTZ → MEGATRON

line ratios are very sensitive to conditions in the ISM and hence to details of star formation & feedback physics implementations in sims

e.g. [O III] λ5007 / [O II] λλ3727
C IV λλ1550/[C III] λλ1908
[OIII]4363/5007

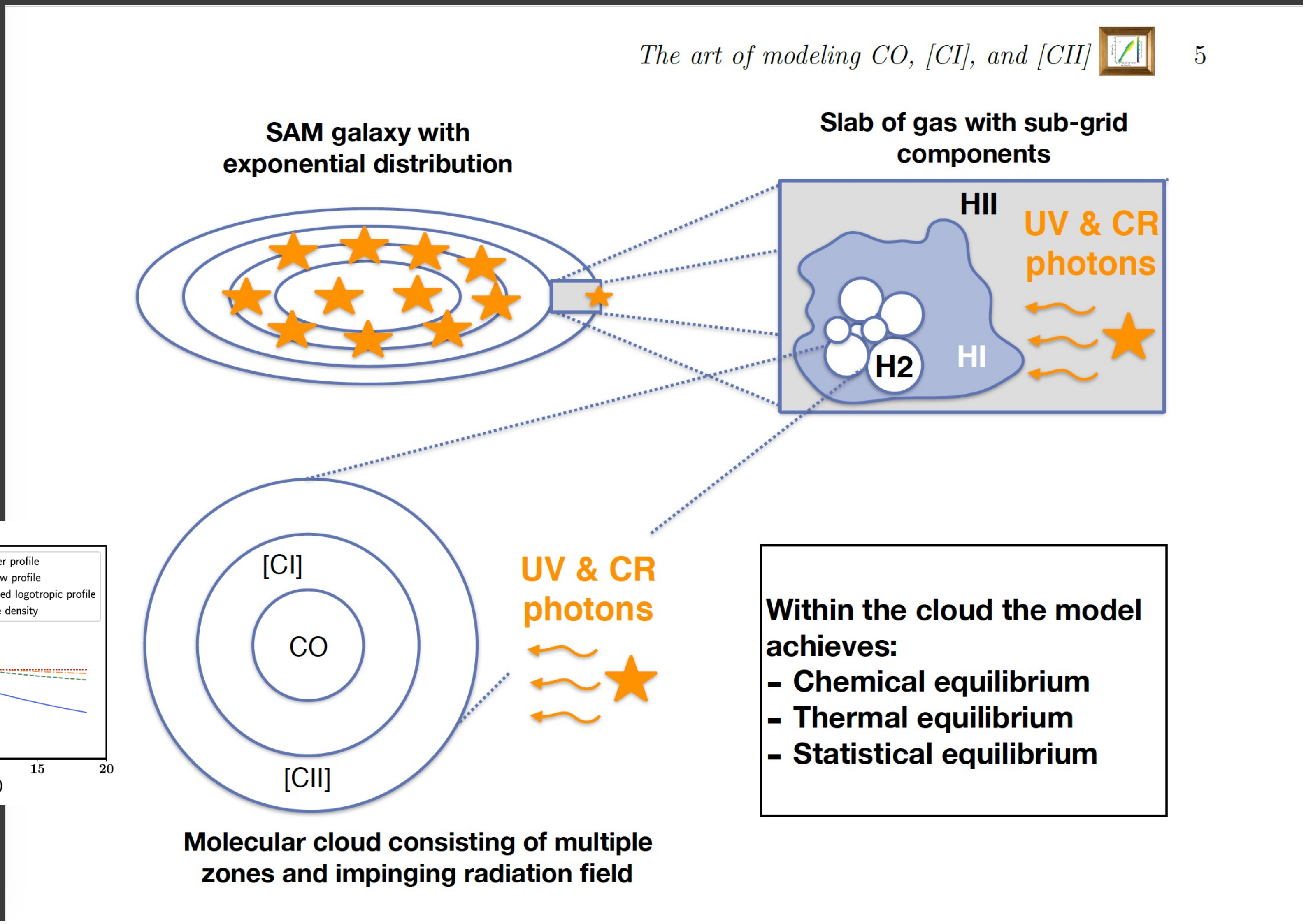
(Katz et al. 2024)

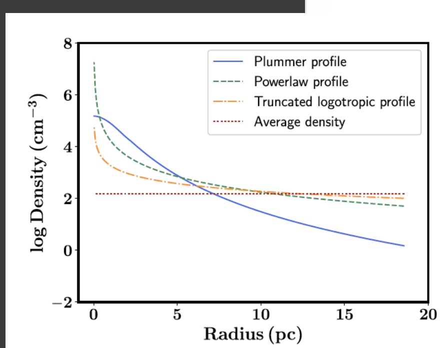
Katz 2022

The art of modeling line emission from the ISM

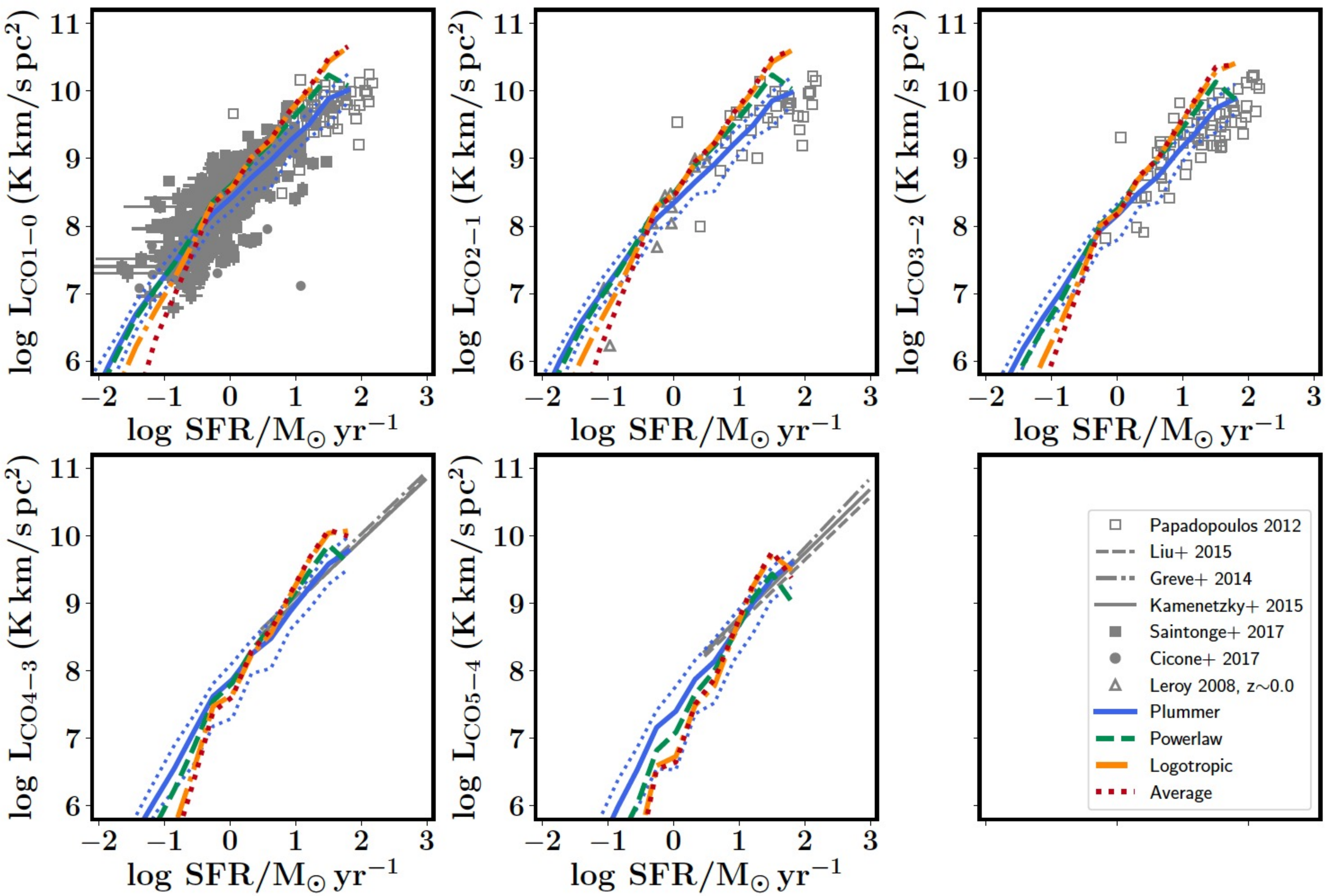
Popping et al. 2019; see also Garcia et al. 2024

Gergö Popping

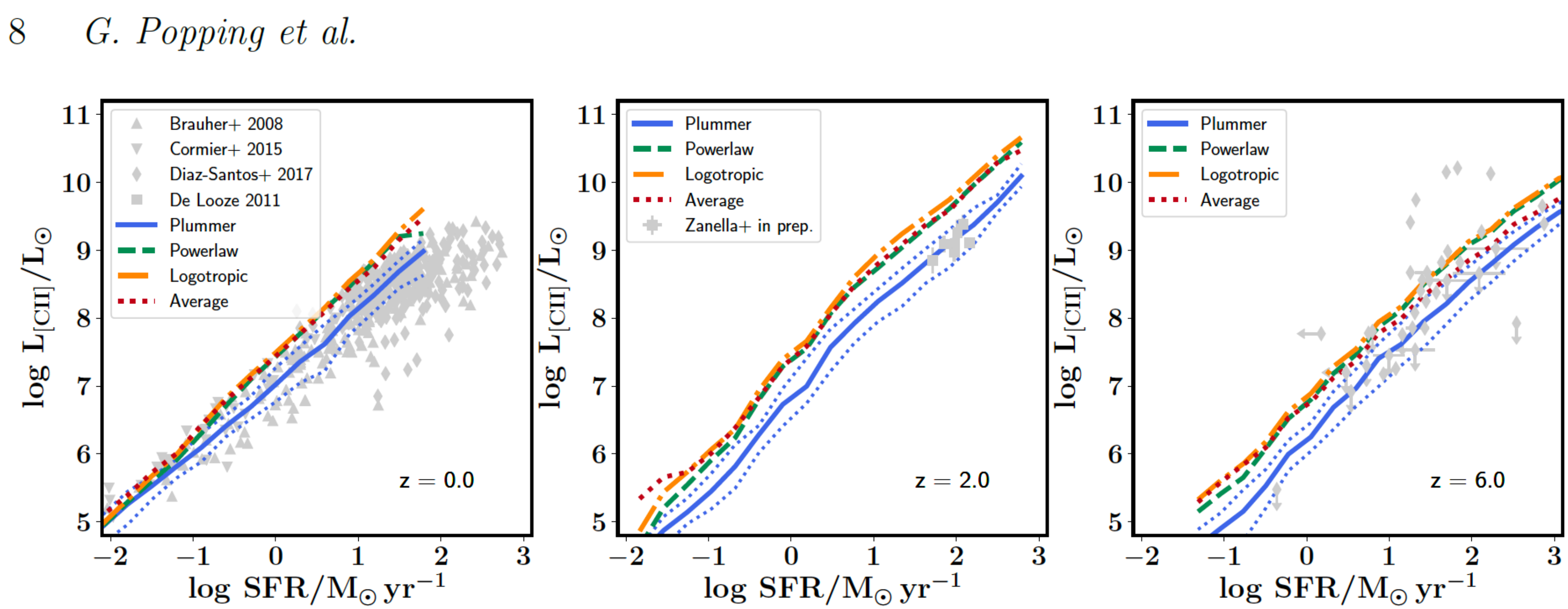




Popping et al. 2019; Popping et al. 2016; Popping et al. 2014



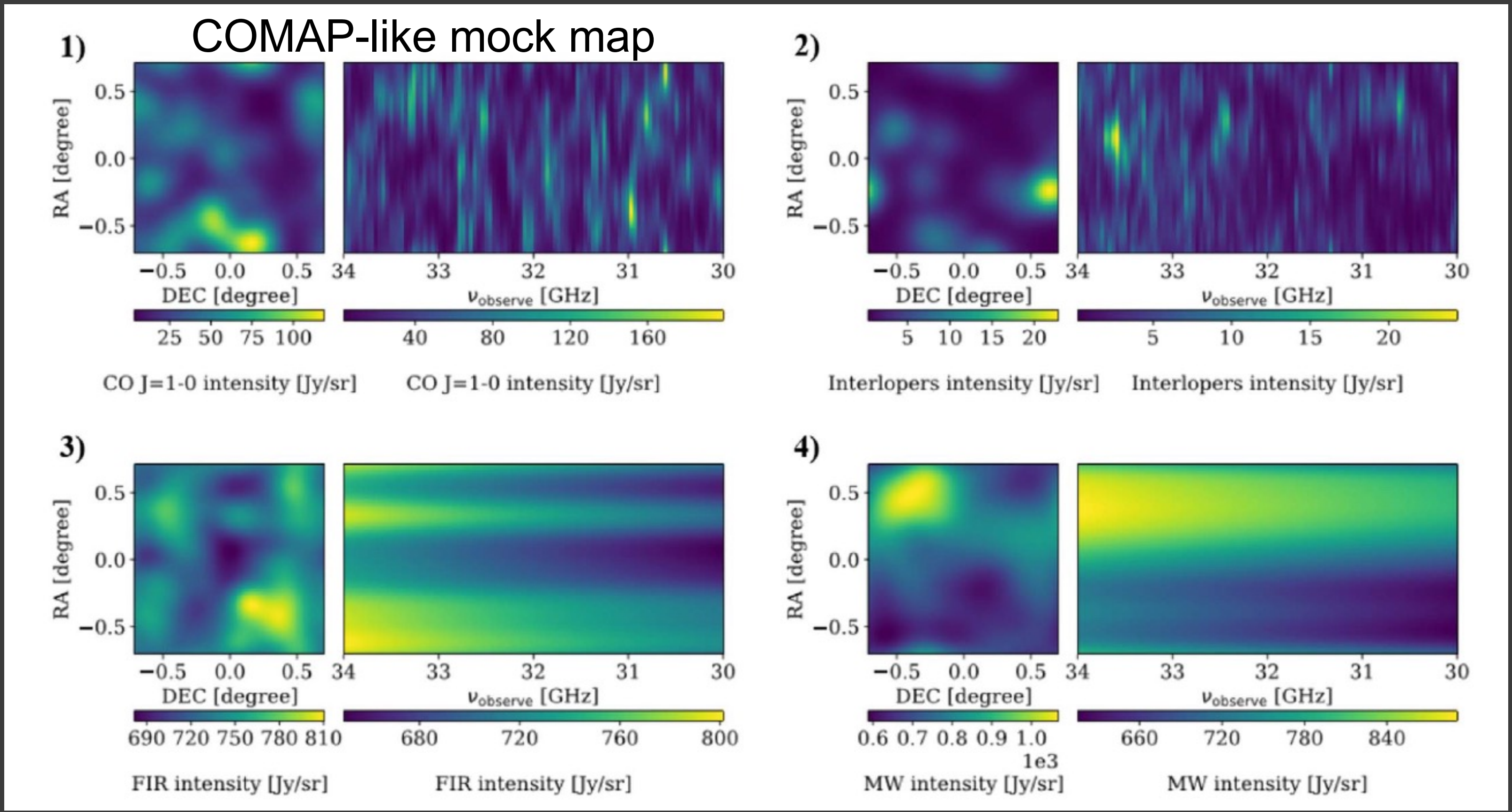
the same models are able to fit scaling relations for both CO and [CII] – places constraints on small-scale structure of the ISM



building multi-tracer mock intensity maps for next generation experiments

-2x2 sq. deg. lightcones from N-body sims & populate with galaxies using the Santa Cruz SAM
-predict all CO lines, [CII], [CI]

Shengqi Yang

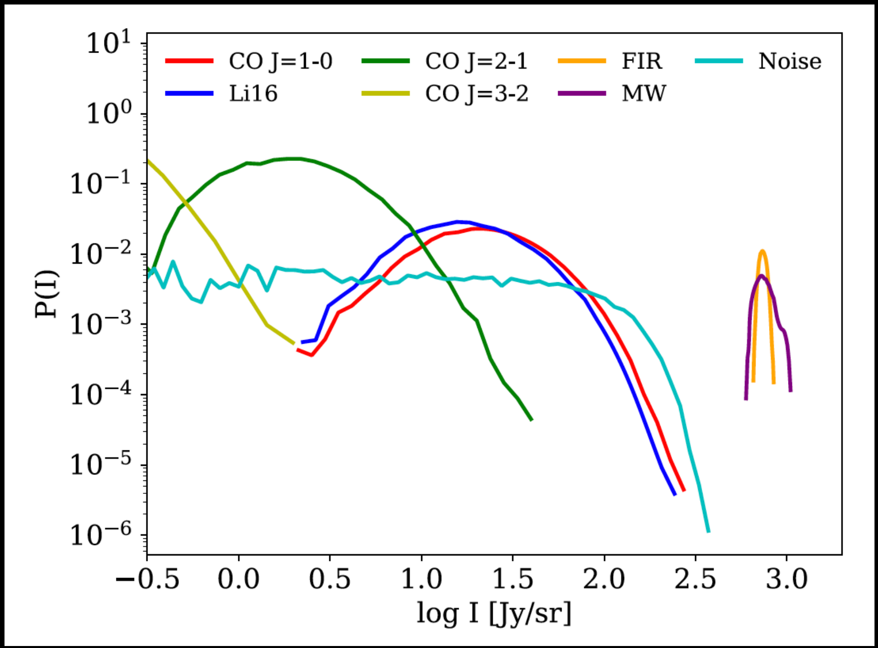


Yang et al. 2021

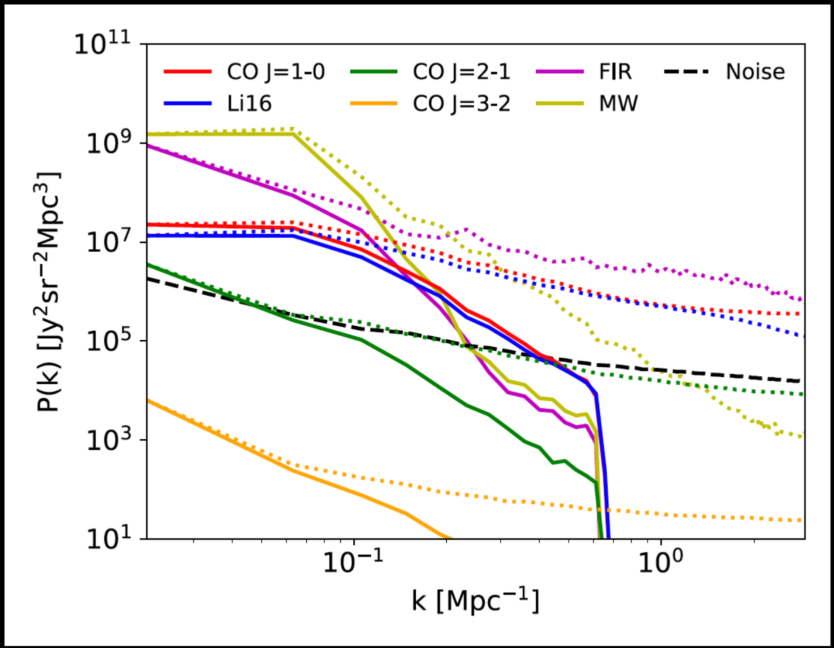
see paper for EXCLAIM-like [CII] map

COMAP forecast

Voxel Intensity distribution



power spectrum



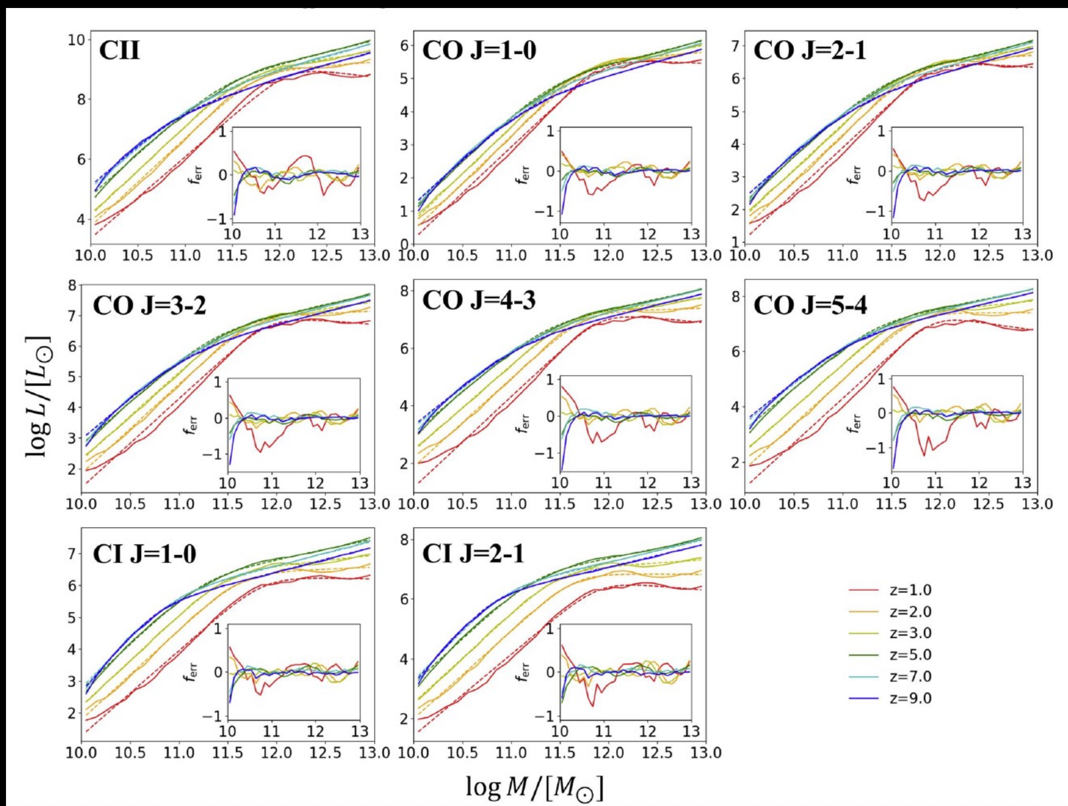
see paper for EXCLAIM ([CII]) forecasts

Yang et al. 2021

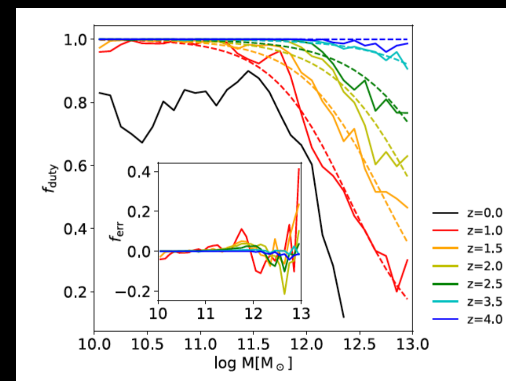
empirical halo model
description of physics
based model

$$\frac{L}{[L_\odot]} = 2N\frac{M}{[M_\odot]}\left[\left(\frac{M/[M_\odot]}{M_1}\right)^{-\alpha} + \left(\frac{M/[M_\odot]}{M_1}\right)^{\beta}\right]^{-1},$$

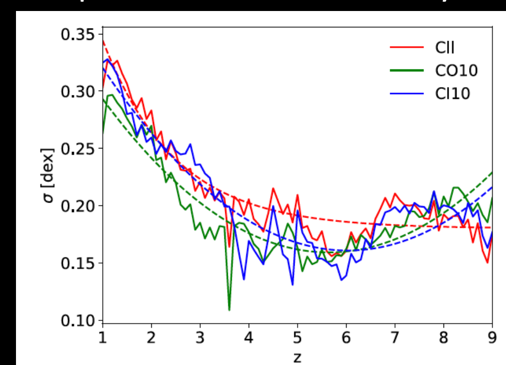
line luminosity vs. halo mass



fraction of line-emitting galaxies



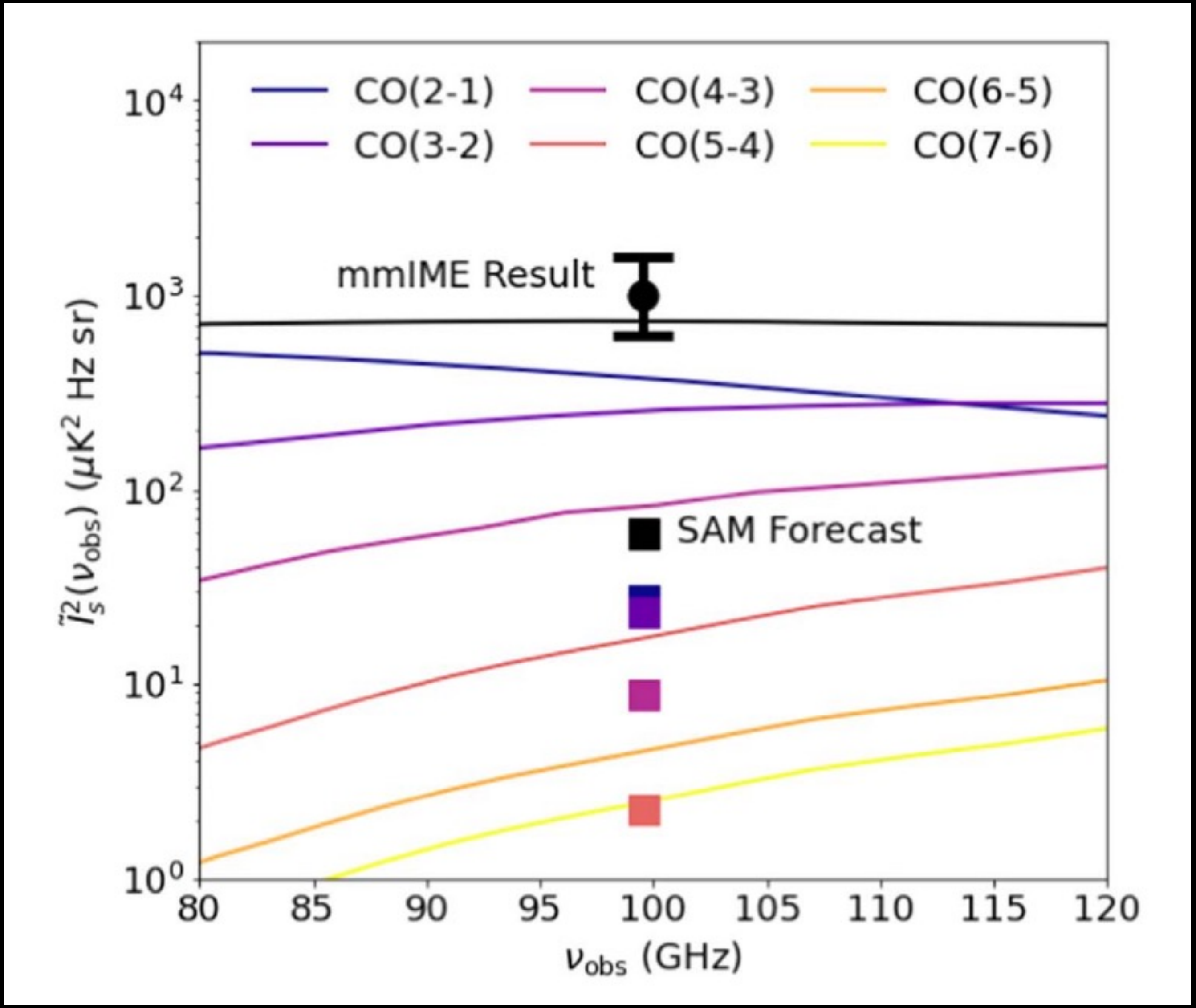
dispersion in line luminosity



https://users.flatironinstitute.org/~rsomerville/Data_Release/LIM/

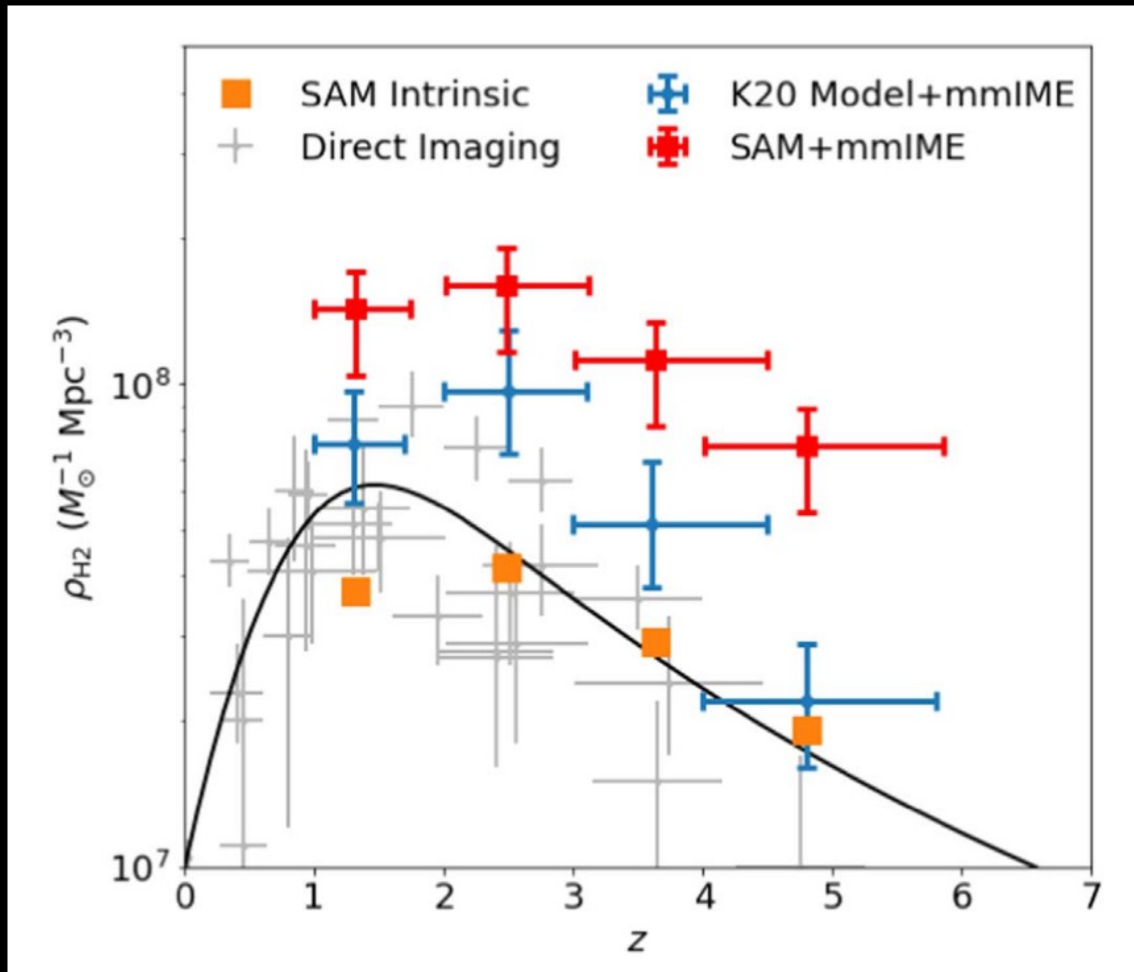
Yang et al. 2022

early results: Santa Cruz + sub-mm SAM mmIME forecast is low



Breysse et al. 2022

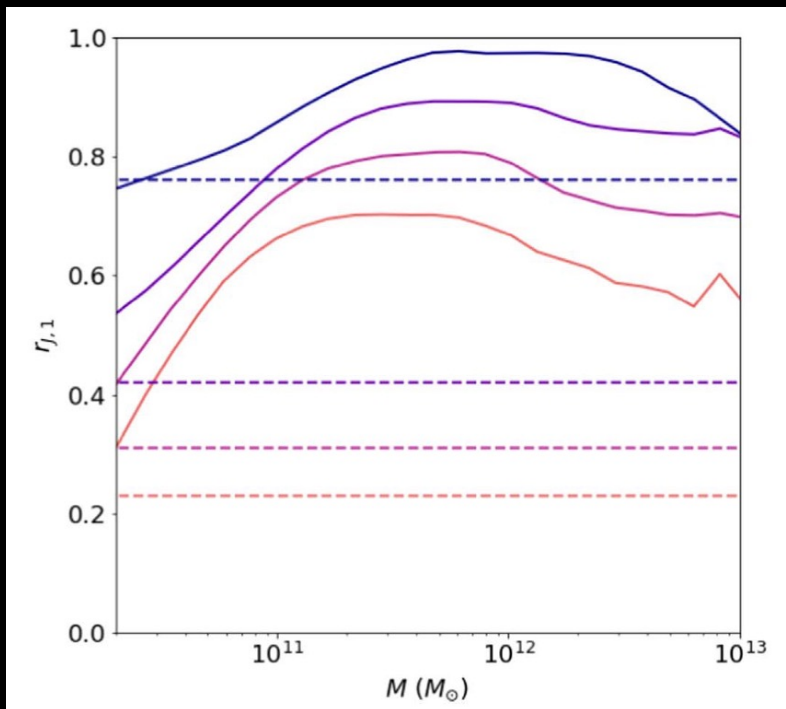
inference for physical quantities (e.g. ρ_{H2}) depend on underlying modeling assumptions



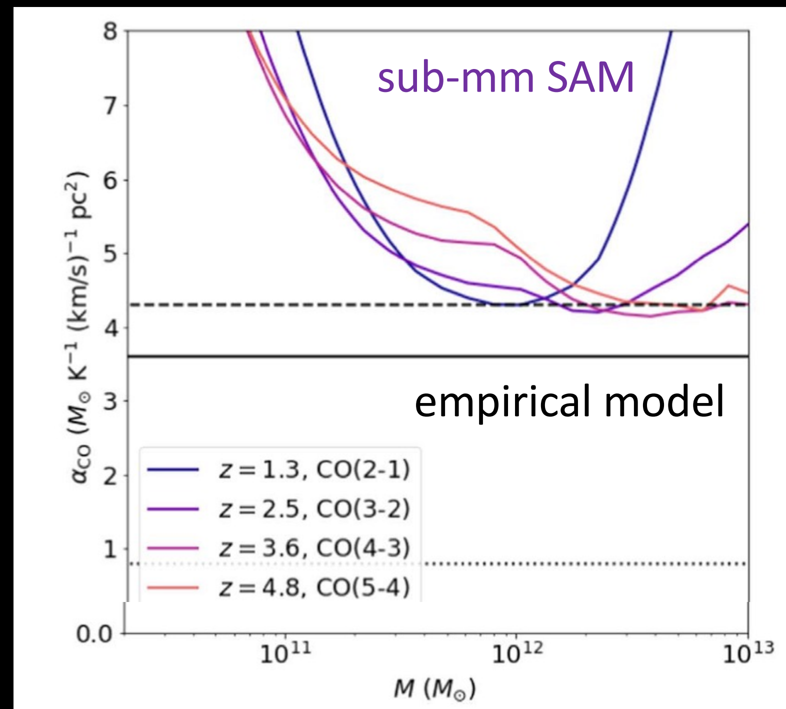
Breysse et al. 2022

traced to different underlying assumptions about how CO line emission traces H_2

CO SLED $r_{J,1} = L'_J/L'(1\text{-}0)$



$\alpha_{CO} = M_{H2}/L'_{CO(1\text{-}0)}$

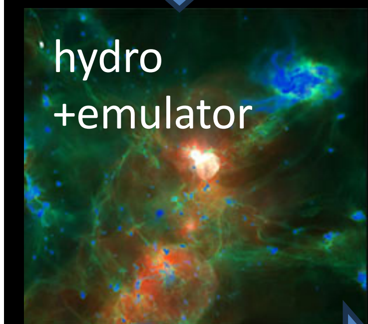


halo mass

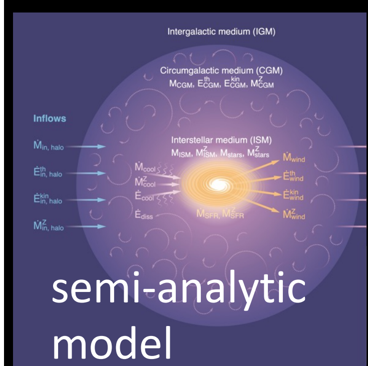
Breysse et al. 2022

building a framework for multi-tracer joint inference

ICs, parameters

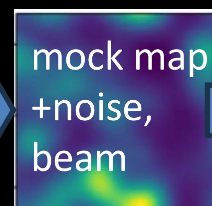


or



physical properties (m_*, ρ_{gas}, Z, SFR, ...)

line luminosity relation emulator $L_i(\mathbf{p})$

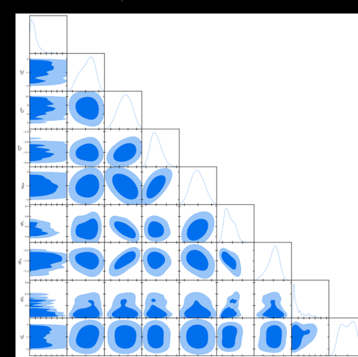


field level or summary

Observational Constraints

I_ν^i $\quad P^i(k)$ $\quad P^{ij}(k)$ $\quad P^{ig}(k)$

inference



joint posterior

e.g. Zhang, Pullen, rss *et al* 2023

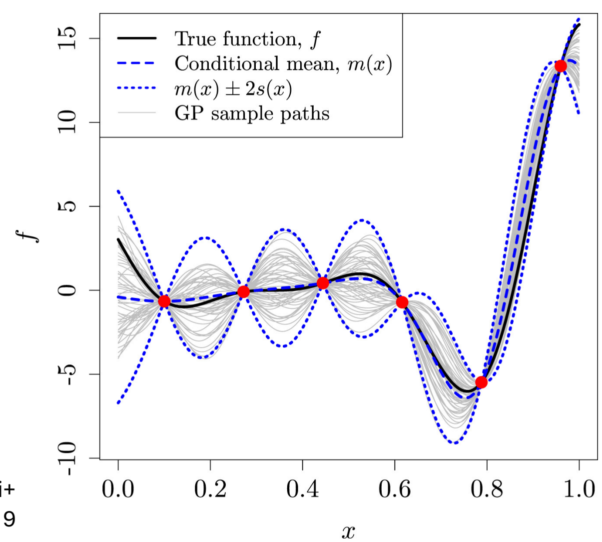
Line luminosity emulator using Gaussian Processes

Want to evaluate

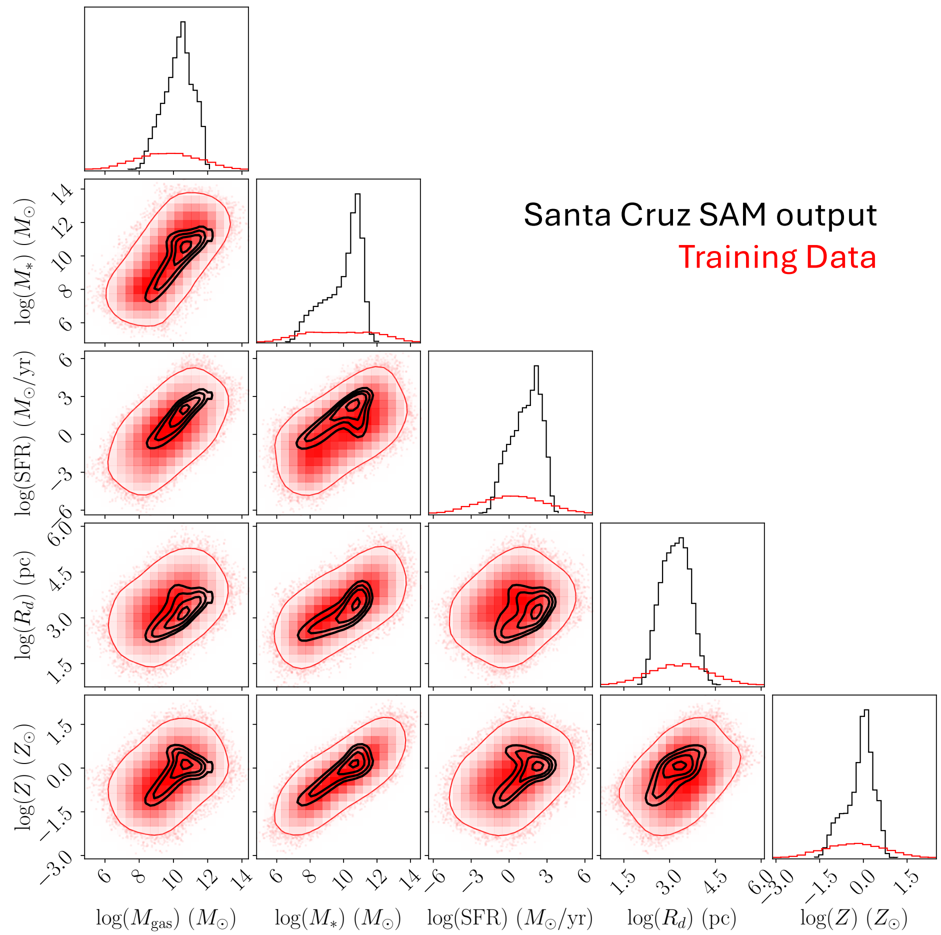
$$L_{\text{Line}}(M_{gas}, M_*, \text{SFR}, R_{\text{disk}}, Z)$$

Assume each value of L_{Line} is a correlated random variable, find the maximum probability

$$P(L_{\text{Line}}^{\text{Test}} | L_{\text{Line}}^{\text{Train}})$$

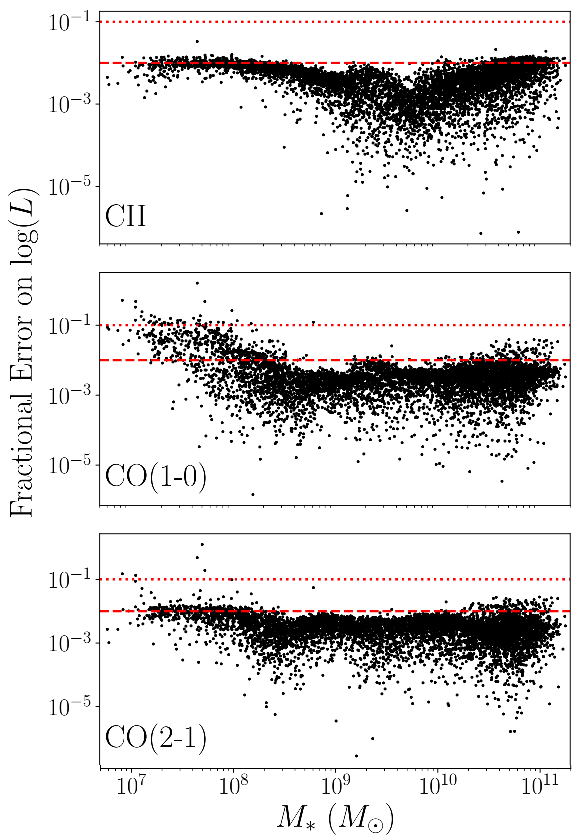


Mohammadi+ 2019

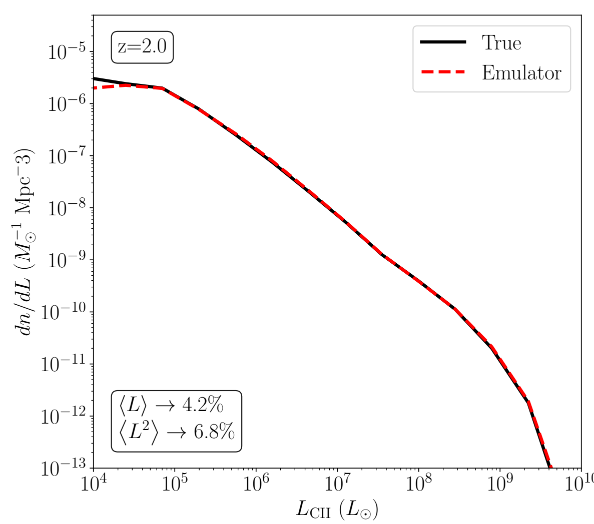


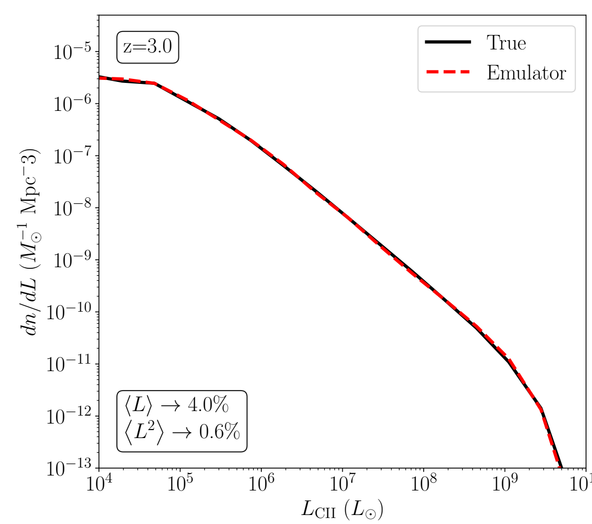
Breysse et al. in prep.

Line Luminosity Relation Emulator Performance



- ~1000x speedup on CPU (faster GPU version coming!)
- ~1% accuracy on log(L), 10% accuracy on LIM integrals
- Currently includes CII and CO(1-0) through (5-4)

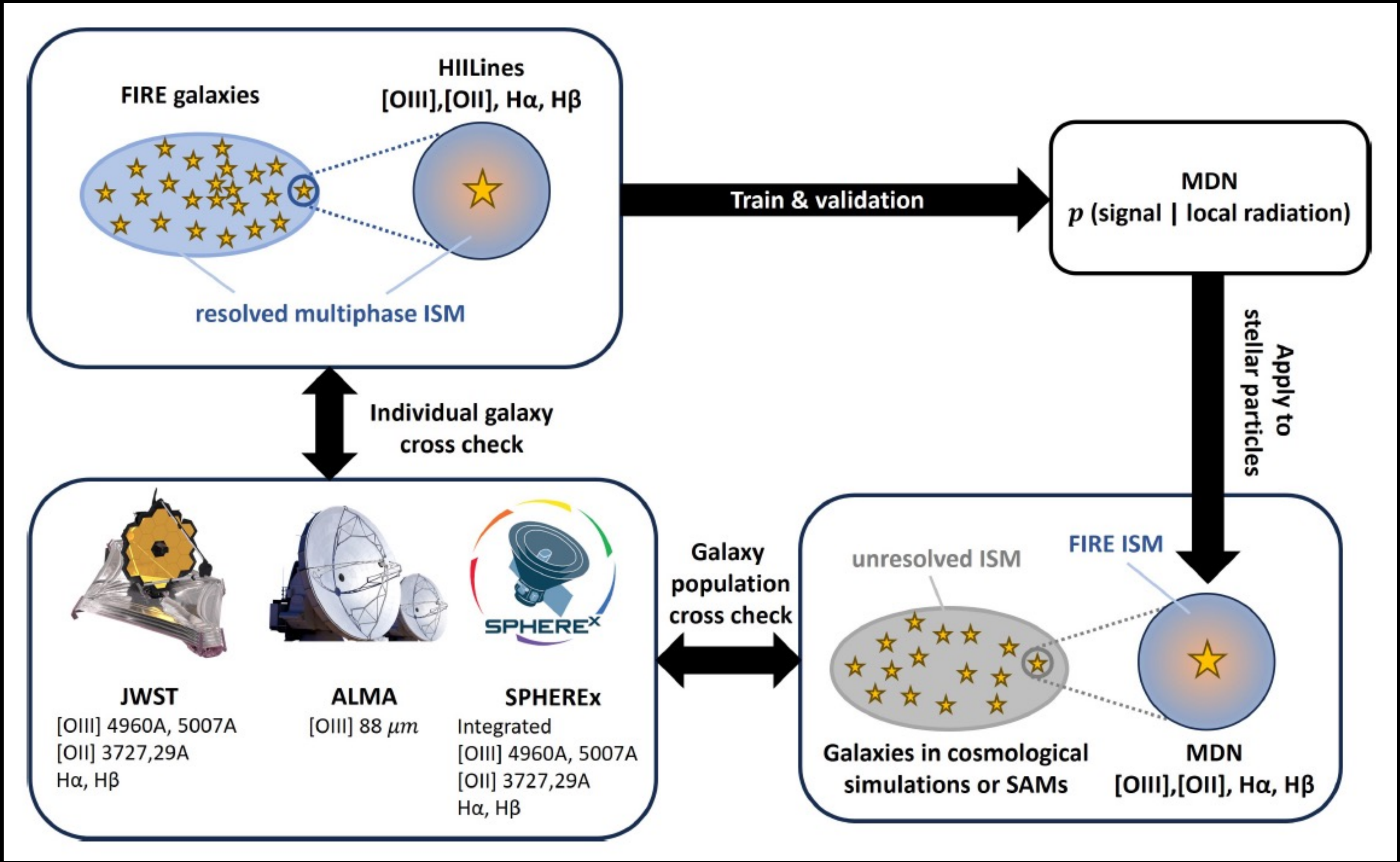




Breysse et al. in prep.

see also Garcia et al. 2024

line emulator for Hα, Hβ, [OII], [OIII] with mixture density networks trained on FIRE



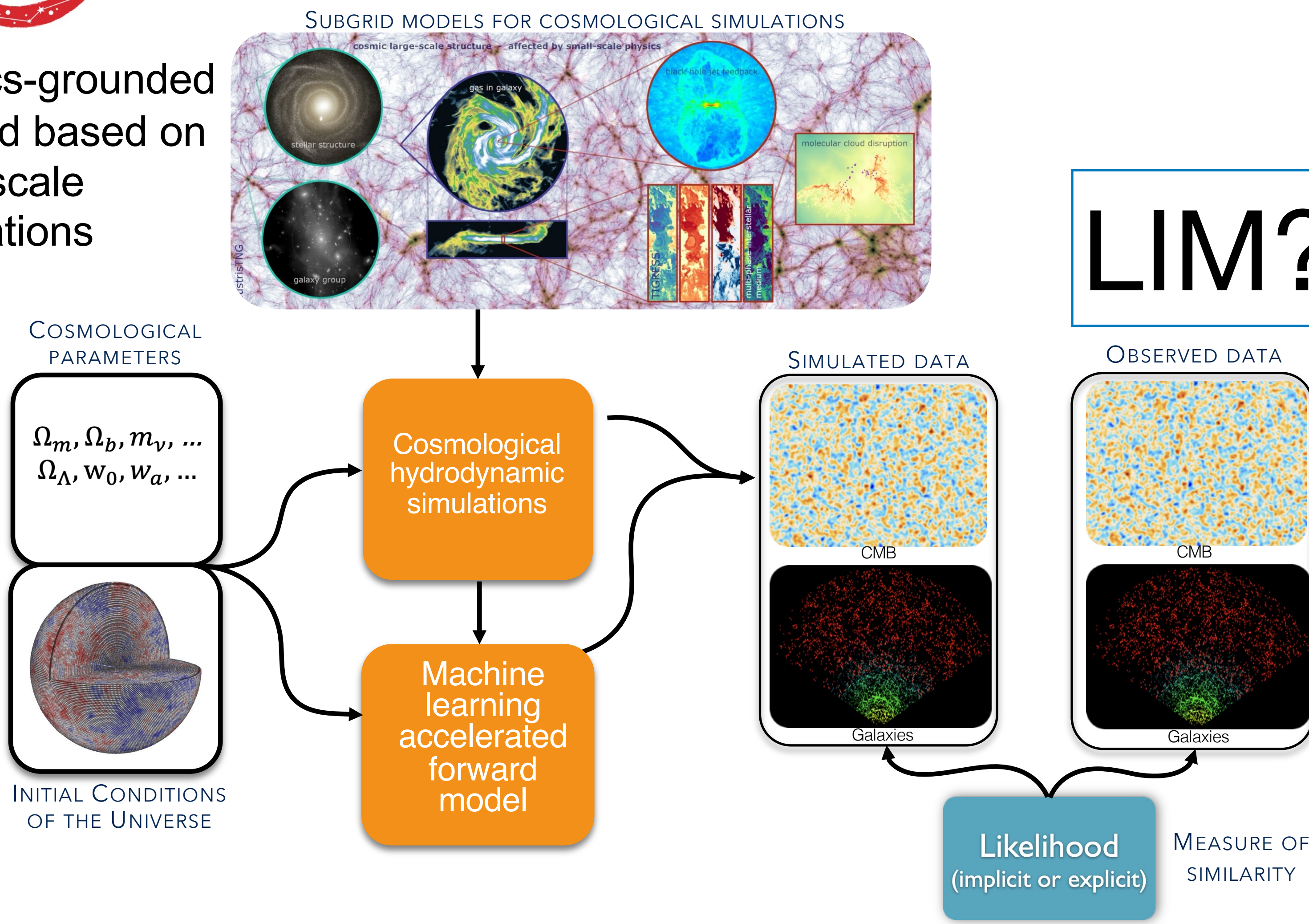
Yang et al. 2025

see also Tolgay talk

Simons Collaboration on Learning the Universe

physics-grounded subgrid based on multi-scale simulations

LIM?



http://learning-the-universe.org/

CAMELS -SAM

Lucia Perez

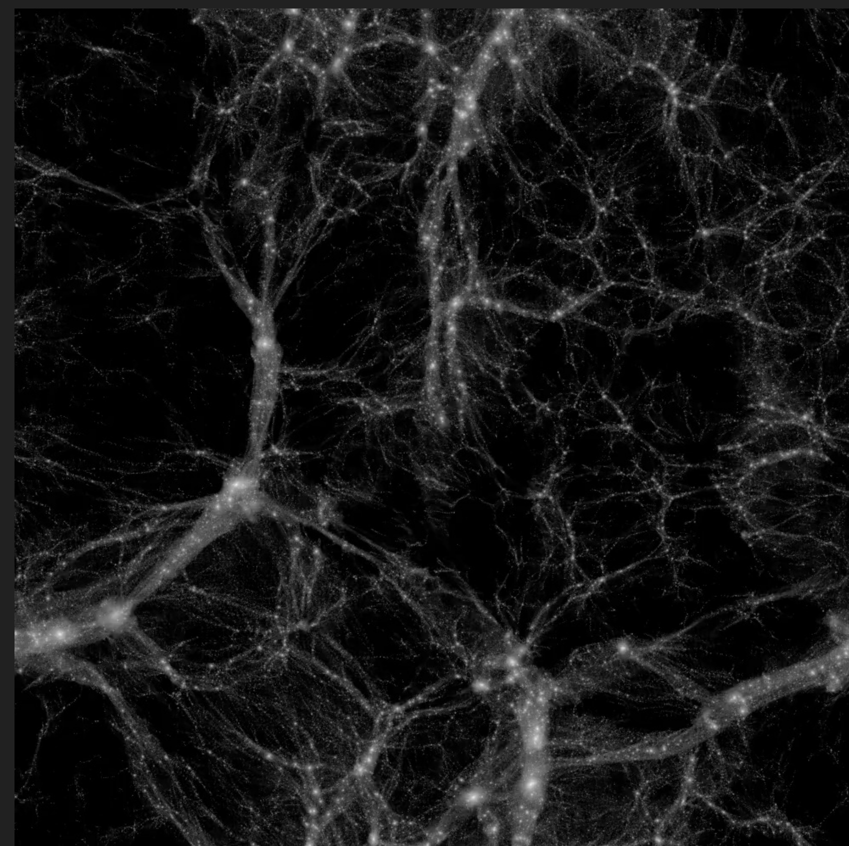
> ➢ 1000+ DM only Gadget-III simulations
> ➢ $(100\ h^{-1}\ \mathrm{Mpc})^3$ in volume
> ➢ $N=640^3$ particles of $\sim 1\text{-}6 \times 10^8\ h^{-1}\ M_{sol}$
> ➢ 100 snapshots, ROCKSTAR catalogs, and ConsistentTrees merger trees

- Cosmological parameters:
  - $\mathbf{\Omega_m}$ from 0.1 to 0.5 | $\boldsymbol{\sigma_8}$ from 0.6 to 1

under development: larger volume, scaled DM simulations (R. Stiskalek)

Latin-hypercube exploration of ~10 galaxy formation parameters using Santa Cruz SAM

photometry & emission lines



SIMONS

Simons Collaboration on Learning the Universe sapphire

modeling galaxies as complex dynamical systems

with automatic differentiation and parallelization

Viraj Pandya

a new physical framework for accelerated Bayesian inference using galaxies and their gas

Multi-CPU/GPU Parallelization

Automatic Differentiation

CGM

BHs

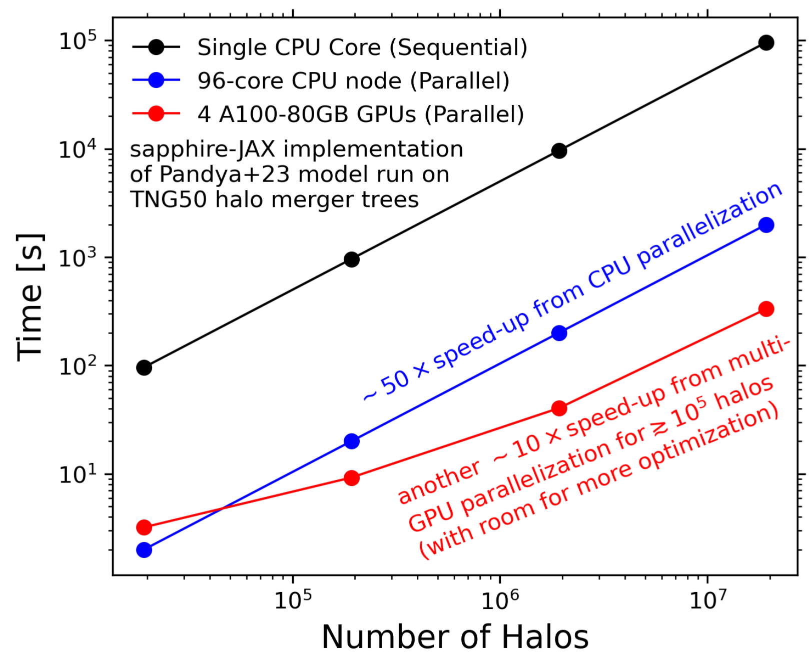
Satellites

SF/ISM

Implicit Likehood Inference

accelerated exploration of parameter variations for millions of model realizations in parallel

Pandya et al. (in prep.)



Pandya et al. 2023; Pandya et al. in prep

SIMONS

Simons Collaboration
on Learning the Universe

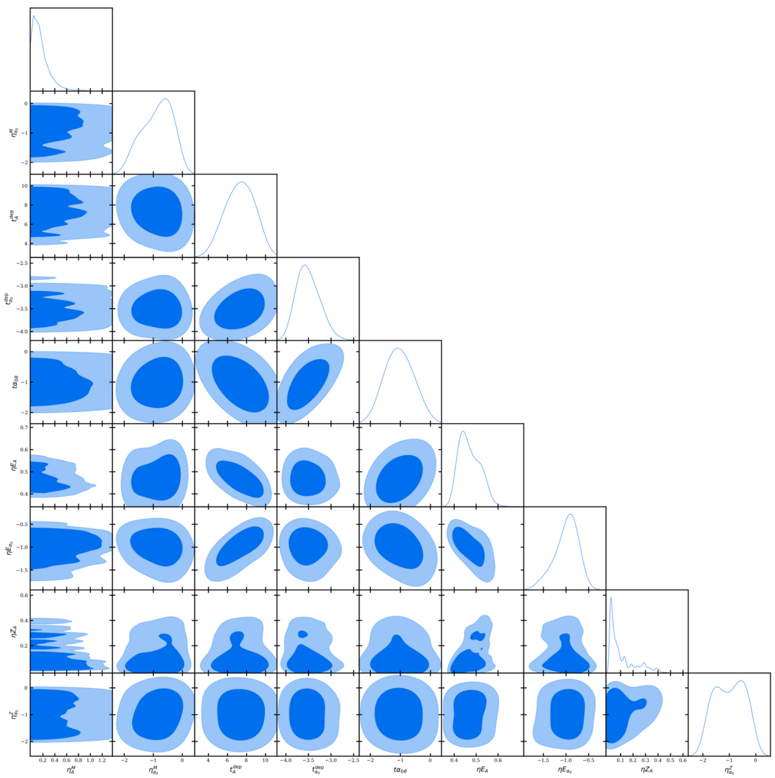
Viraj Pandya

sapphire

modeling galaxies as complex dynamical systems

Implicit Likelihood Inference

train a neural network to learn the mapping between parameters & outputs (e.g. Ho et al. 2024) → directly constrain the multi-dimensional posterior

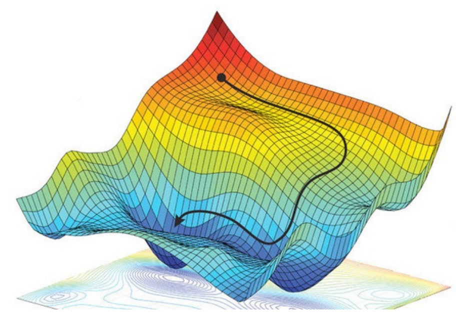


astrophysics parameters

JAX

with automatic differentiation and parallelization

hamiltonian monte carlo



use gradients to speed up parameter space exploration

summary

- upcoming LIM experiments have great potential to constrain uncertainties in galaxy formation models
- physics-based galaxy formation models offer useful insights for interpreting LIM results
- new techniques (next-generation SAMs, ML-based emulators) are enabling physics-grounded multi-tracer inference

extra slides