Cosmic Dawn simulations

sharing the data

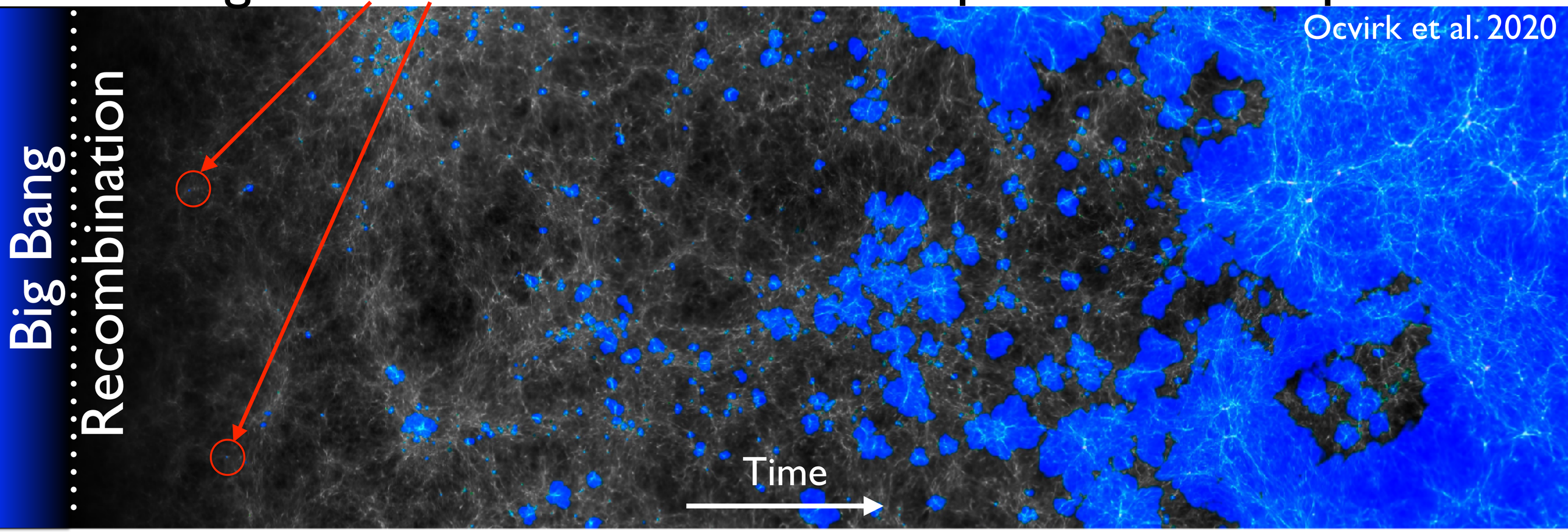
P. Ocvirk and Cosmic Dawn collaboration
https://coda-simulation.github.io/
Observatoire astronomique de Strasbourg
Universite de Strasbourg

Cosmic Dawn Collaborators

- **P. Ocvirk, D. Aubert**, **J. Lewis** (Observatoire astronomique de Strasbourg)
- **J. Sorce** (ICs, **CLUES**)
- **R. Teyssier**, T. Stranex, (University of Zurich) (code + sci. exp.)
- **Y. Dubois** (IAP) (SN feedback, dust model)
- **F. Roy, Y. Rasera** (Observatoire de Paris) (large scale deployment, pfof)
- **M. Gronke** (University of California, Santa Barbara) (sci exp)
- **H. Park** (sci exp)
- **P. R. Shapiro**, T. Dawoodbhoy, J. Lee (University of Texas, Austin) (sci. exp.)
- **I. Iliev**, **L. Conaboy**, A. Hosein (University of Sussex) (sci. exp.)
- **S. Gottloeber** (Leibniz Institute for Astrophysics, Potsdam) (ICs + sci. exp., **CLUES**)
- **G. Yepes**, A. Knebe (Universidad Autonoma de Madrid) (ICs + sci. exp., **CLUES**)
- **Y. Hoffmann** (Hebrew University of Jerusalem) (ICs, **CLUES**)

I - The Epoch of Reionization

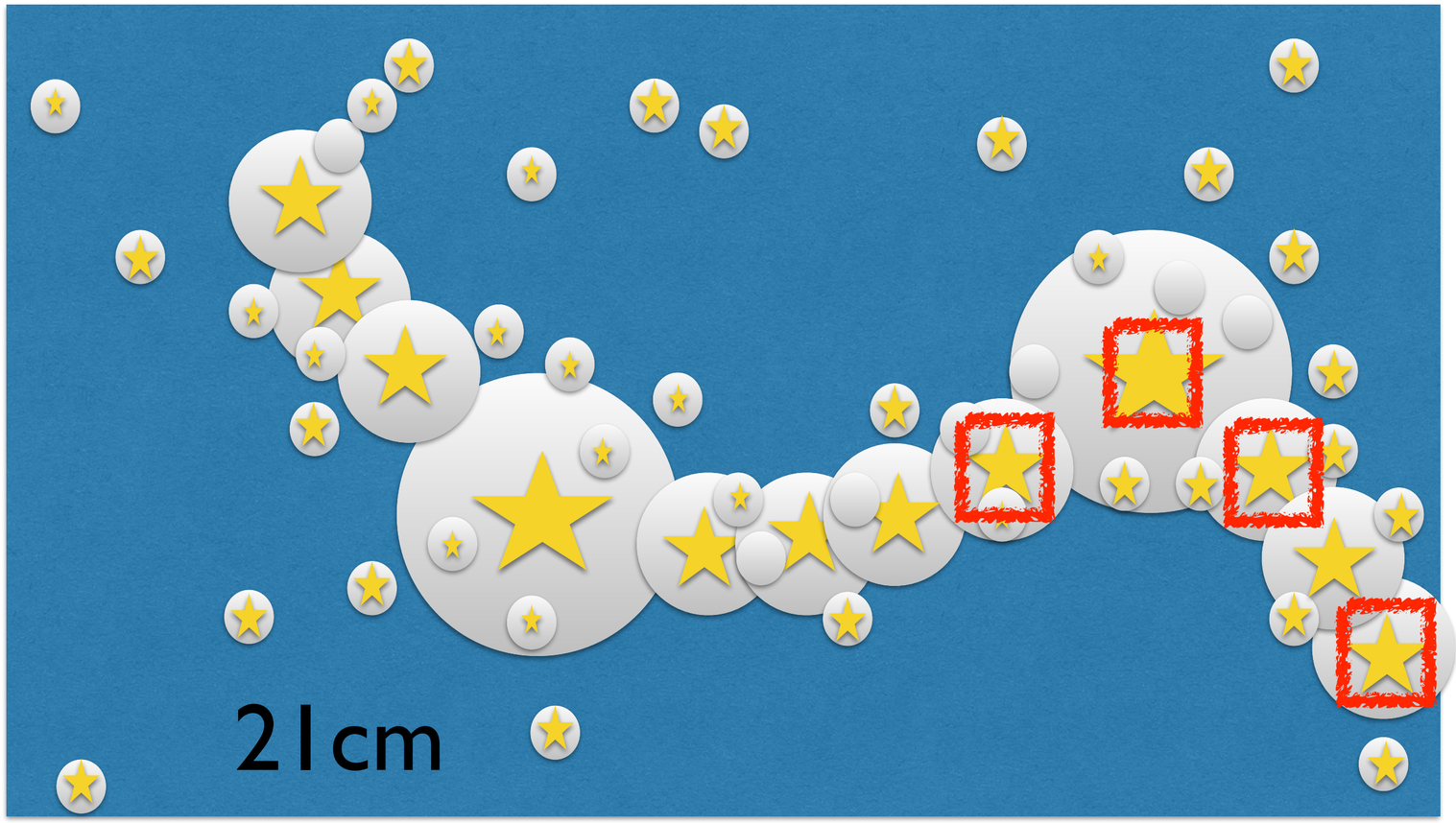
Dark ages | First stars | Partial overlap | Overlap

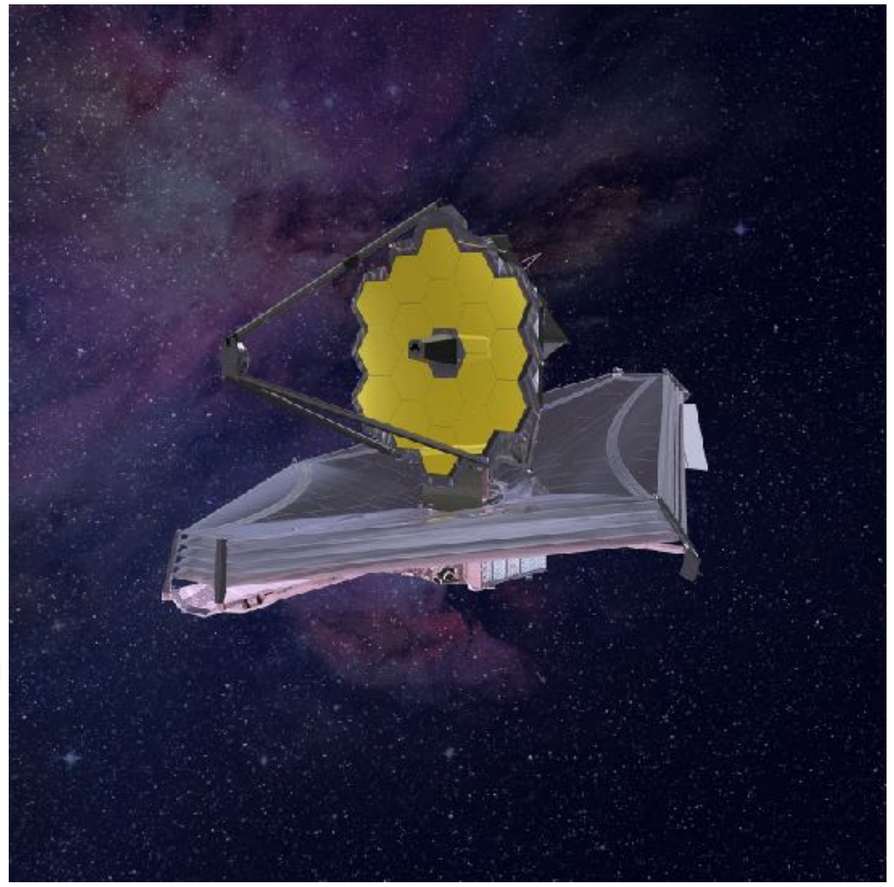


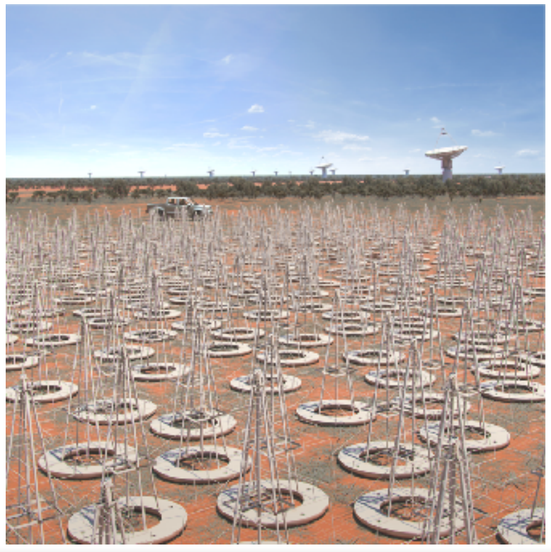
z=1100 | z=30-15? | z=7 | z=6

Neutral | 50% ionised | Ionized

The Epoch of Reionization







LOFAR
NenuFAR
HERA
SKA, …

JWST 2021
NGRST
Euclid

II - Dark ages / EoR open questions

- When did dark ages / reionization start / finish?
- Ionising sources?
  - Galaxies (high/low mass?)
  - Ionising UV Escape fraction?
  - Black Holes (stellar / supermassive)
  - X-ray Binaries?
- Radiative feedback on early galaxies?
- mass limit for star formation? => impact on reionization history?

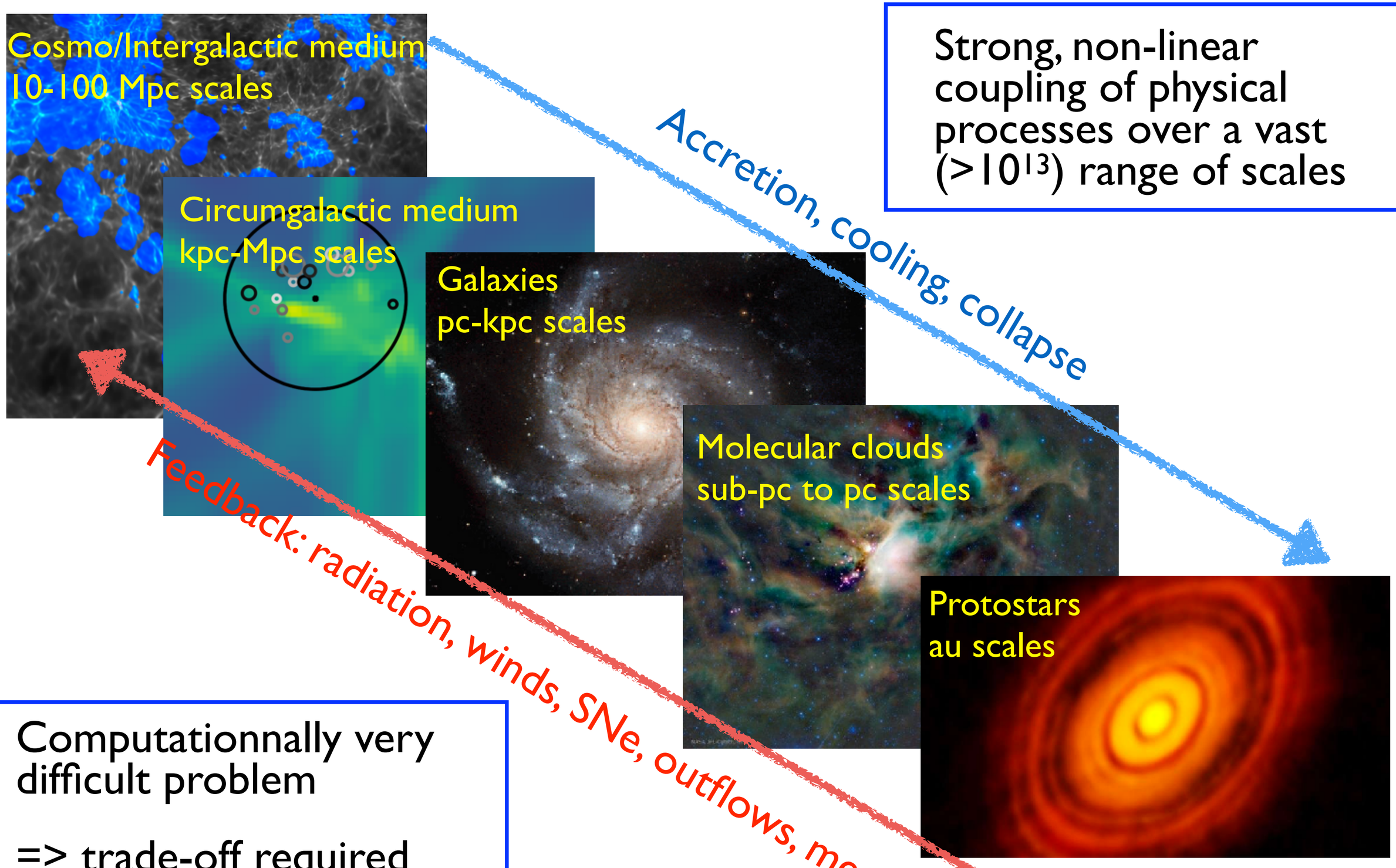
II - Fully coupled Radiation-hydro with RAMSES-CUDATON (Ocvirk+2016)

- **RAMSES** (Teyssier 2002): **CPU**
  - gravity (PM) + hydrodynamics
  - star formation + SN thermal + kinetic feedback
  - chemical enrichment, dust production + destruction (Lewis+2022)

T, x_{HI} ↑ ↓ $T, \rho_{HI}, stars, \rho_{dust}$

- **ATON** (Aubert 2008): UV Radiative Transfer,
  - photon propagation, H ionization
  - H Photo-heating + cooling
  - dust absorption

II - The tyranny of scales



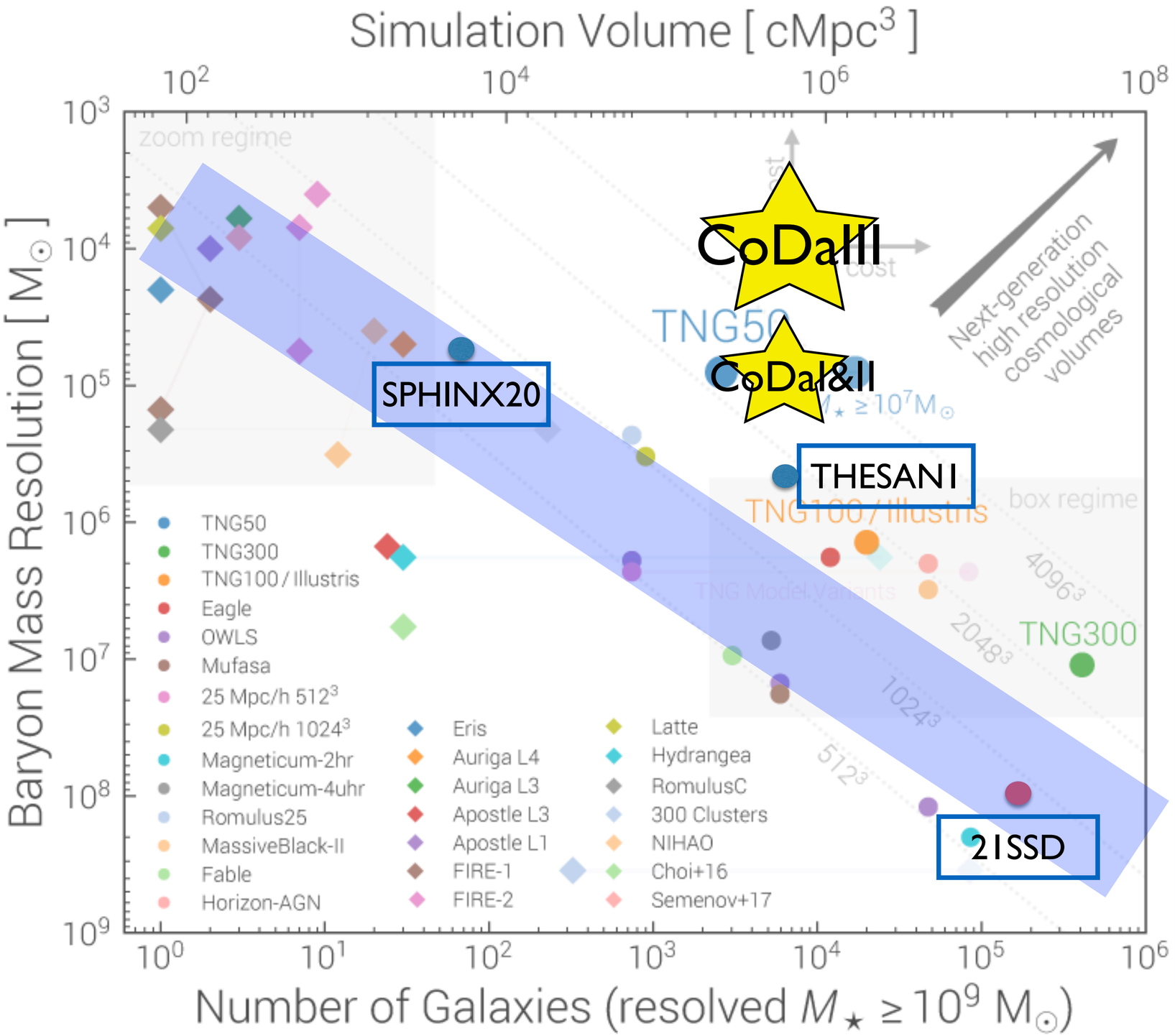
Strong, non-linear coupling of physical processes over a vast (>10^{13}) range of scales

Computationnally very difficult problem

=> trade-off required

(Inspired from J. Rosdahl's slide)

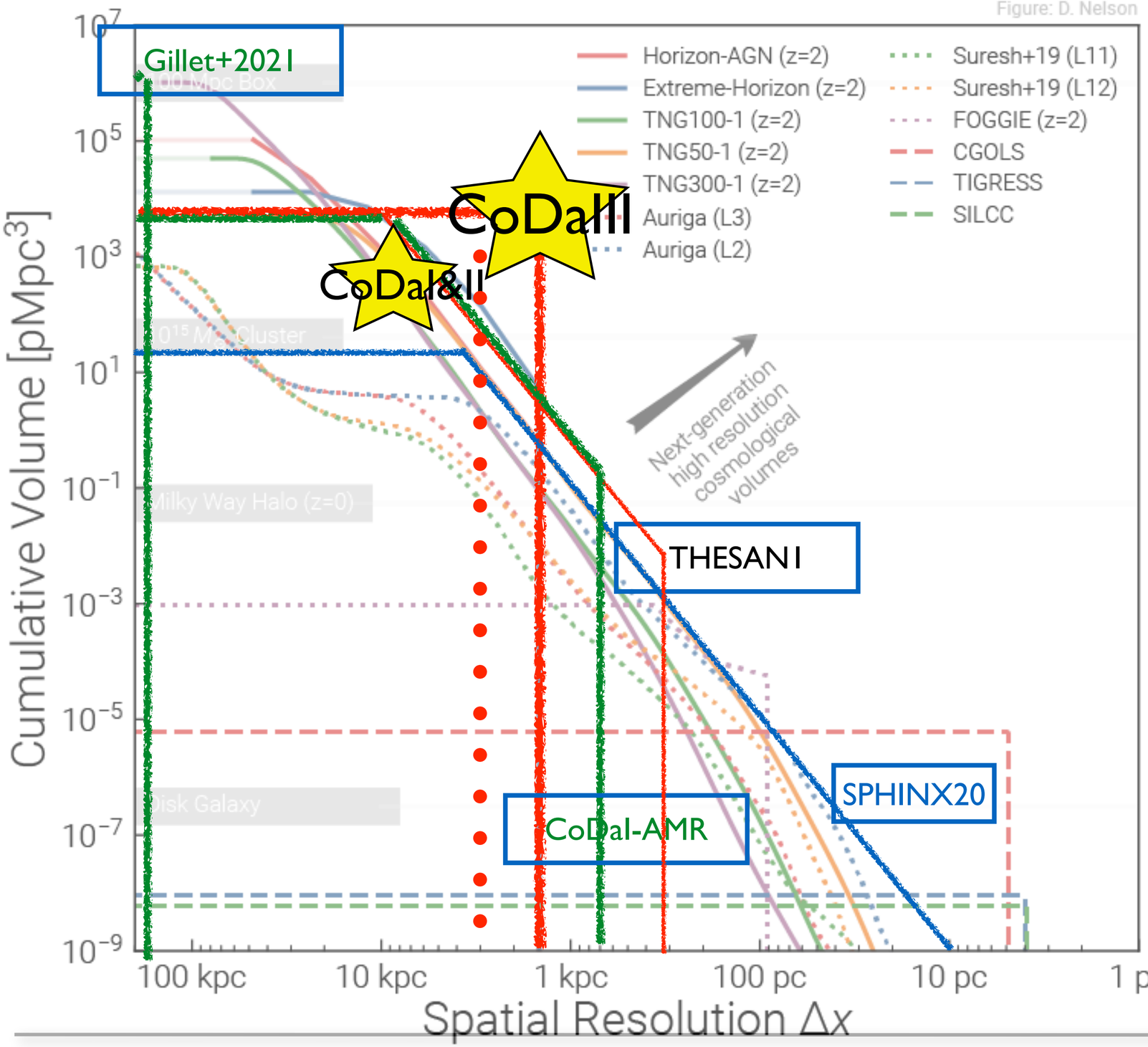
II - The "main sequence" of galaxy formation simulations



- trade-off between size and resolution
- highlighted = recent EoR, fully coupled Radiation-hydrodynamical sims
- Cosmic Dawn sims stands out thanks to:
  - raw power
  - GPU optimization
- Intermediate resolution:
  - Large volume
  - ~coarse description of low mass haloes

 (plot from tng-project.org)

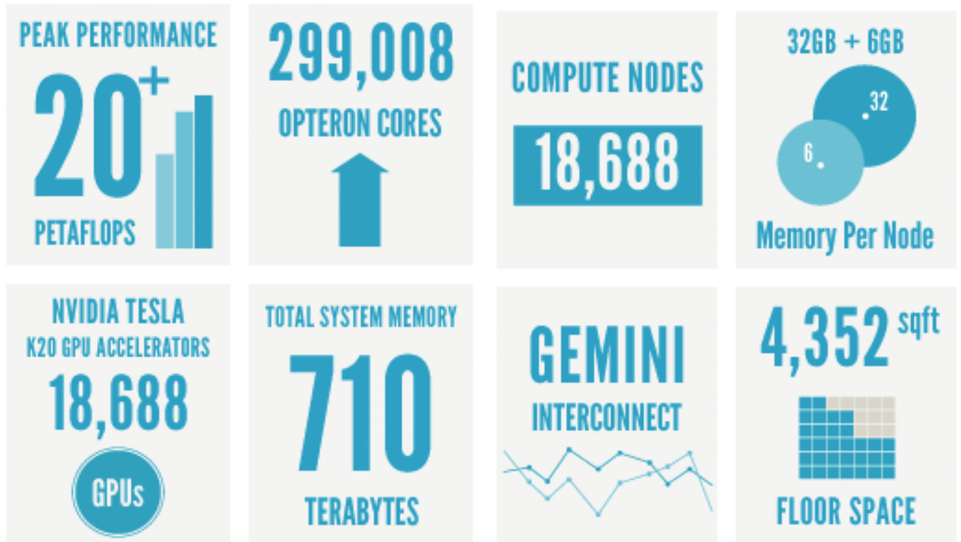
II - The "main sequence" of galaxy formation simulations



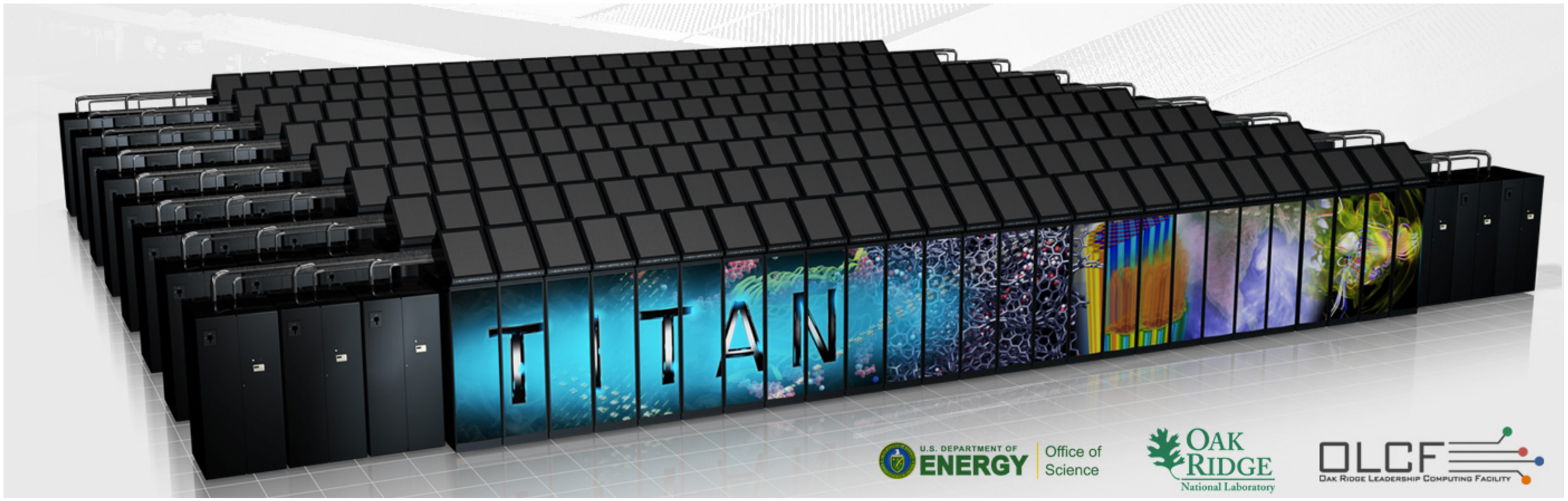
- EoR sims footprints are shown for z=6
- trade-off between size and resolution
- inclined tracks = AMR/SPH sims
- AMR/SPH sims can reach very high res
- but use a reduced speed of light c=0.2-0.02
- Again, Cosmic Dawn simulations stand out as very large, intermediate resolution, fixed grid

 (plot from tng-project.org)

TITAN at Oak Ridge National Laboratory (CoDa I (2012 &II (2016))



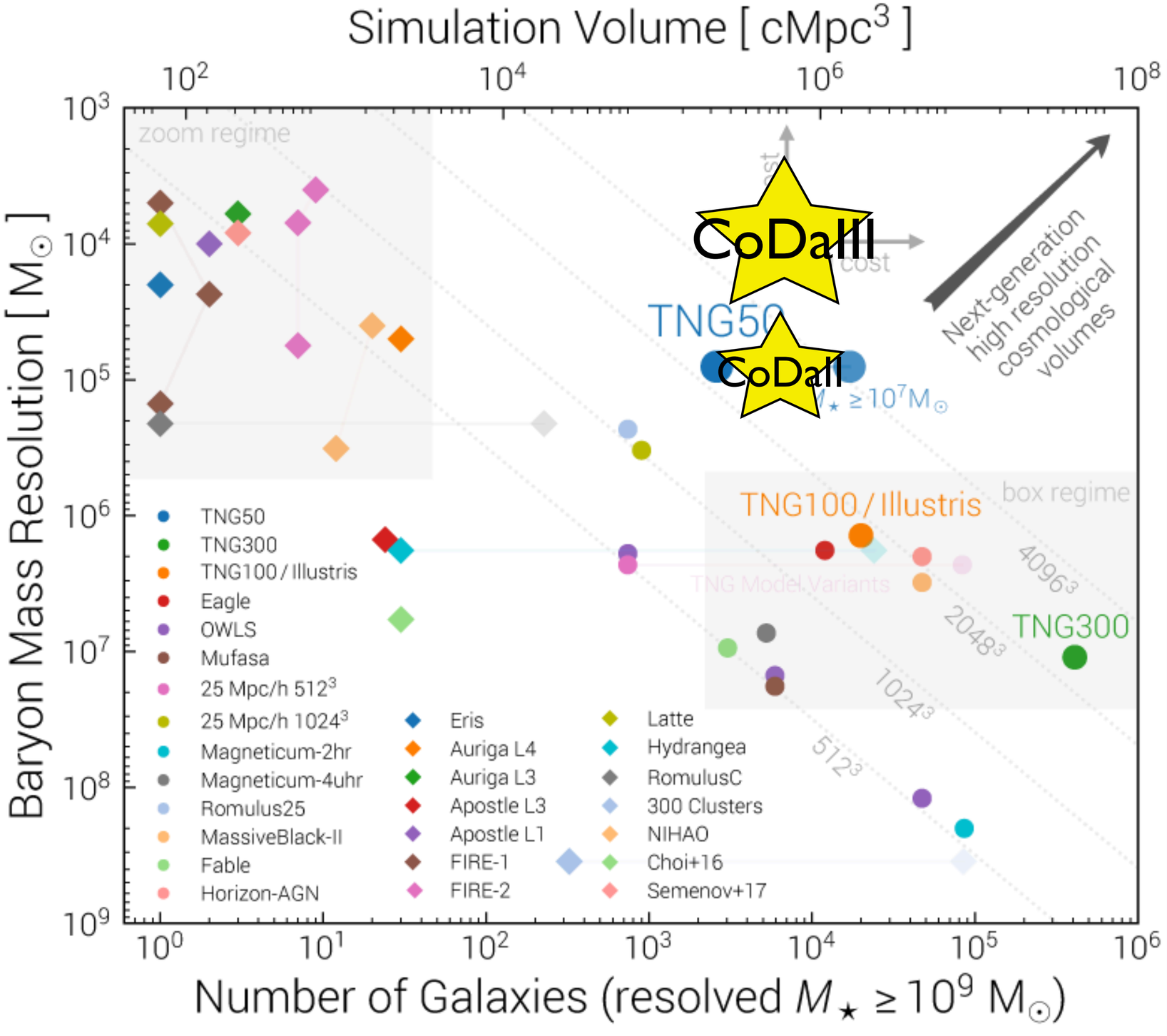
- **18,688 GPUs**
- Cosmic Dawn I, II: $64h^{-1}Mpc$ / 4096^3
- Cosmic Dawn I-AMR (EMMA)



Summit (Cosmic Dawn III (2020))

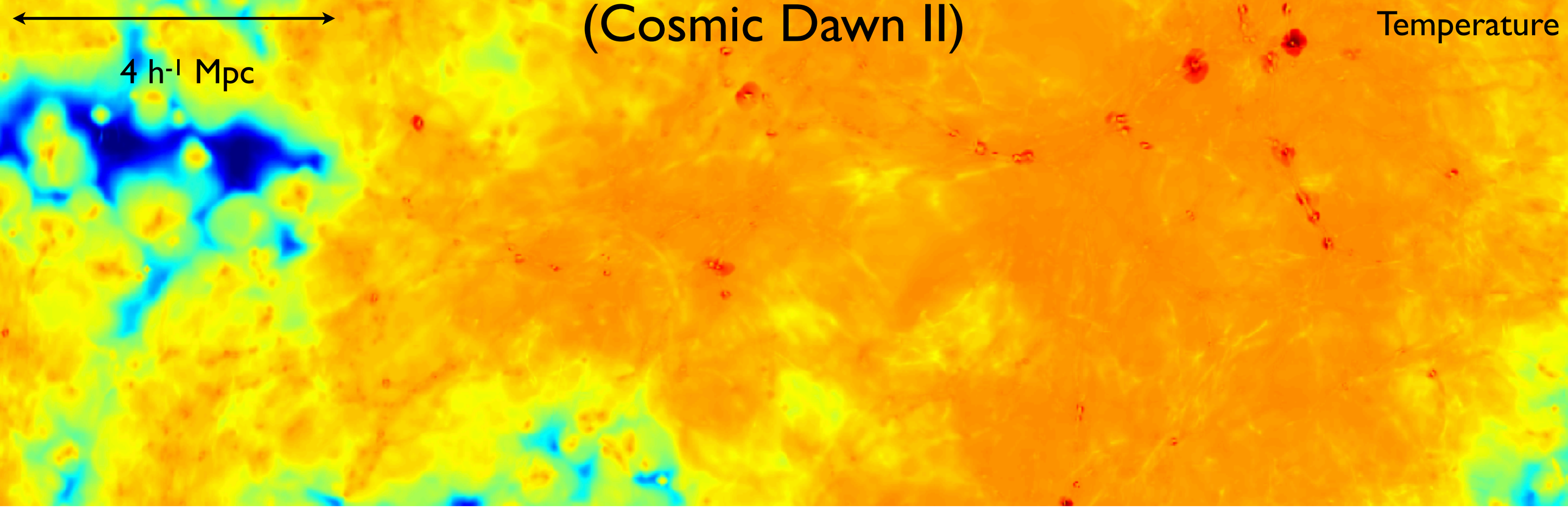
Oak Ridge Leadership Computing Facility

Setup: Cosmic Dawn III on Summit



(from tng website)

- 131 072 CPUs
- 24 576 GPUs
- 94Mpc box (64 h^{-1} Mpc)
- 8192^3 grid
- $Mhalo_{min}$ ~2 x 10^7 $M_\odot$
- Δx ~ 11.5 kpc comoving (< 2 kpc physical)
- z_{end}=4.7
- ~ 10 days runtime on Summit (finished Feb.21)
- 20 PB data
- OLCF/Summit (top 1-4)



Galaxy populations
LFs & rad. FeedBack

Photon budget of galaxies

21cm

Reionization history of galaxies/pairs/LG

IGM transmission (Lyα, LyC)

Ocvirk 2016,2020, Dawoodbhoy 2018, Lewis 2020, Aubert 2018, Sorce 2022, Gronke 2020, Park 2021,Park 2022, Lewis 2022, Gillet 2021, Hosein 2022

Analysis workflow

- CoDaIII: originally 120 snaps (~10 Myr between z=150 and z~5) for a total of ~20PB. High frequency required for:
  - efficient (?) restarts
  - merger trees
- For scientific exploitation a "reduced" dataset is produced:
  - amr2cube : => 512 sub-cubes of 1024^3 for gas properties (implies data duplication) (mostly J. Lewis)
  - starfiles: small-ish (700 million stars)
  - FoF halo catalogs (L. Conaboy)
  - reionization maps (very small)

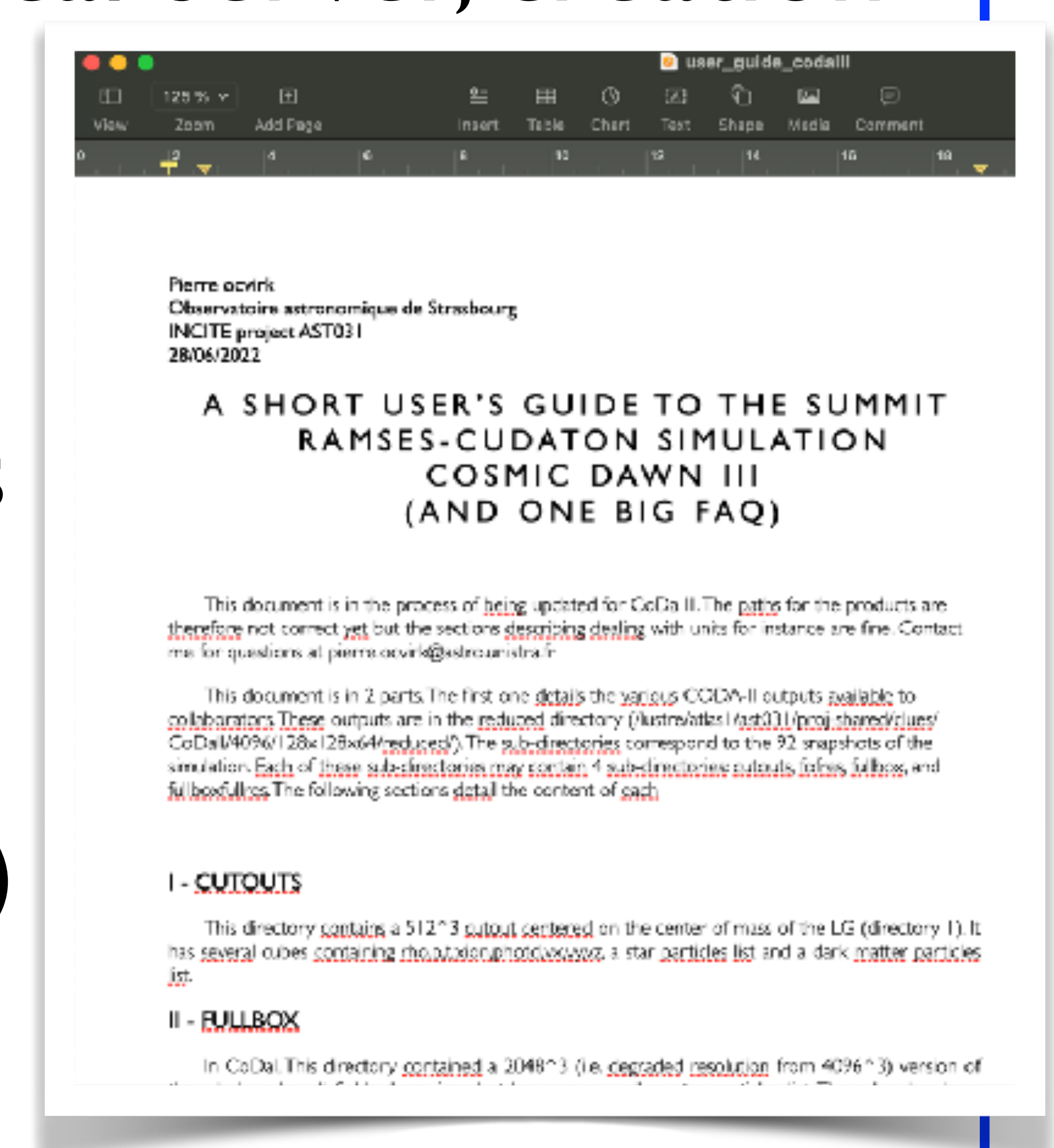
Data analysis: different needs for different science cases

Galaxy populations LFs & rad. FeedBack	z=5,6,7,8,9,10,12,15 FoF halo cats, star files (+ rho, Z, dust fields)
Photon budget of galaxies	+ xHI for f_{esc}
IGM transmission (Lyα, LyC)	z=4.6,4.7,….6.0,6.5,7.0 rho,T,xHI,vz
21cm	z=6,7,8,…30 rho_{1MPC}, starfiles

- strongly non-exhaustive (e.g. Aubert 2018, Sorce 2022)
- perimeter of data = strong constraint on possible science
- some things we could not do: full follow up of f_{esc} evolution of galaxies:
  - rockstar failure
  - requires outputs at max res at max frequency for z=5-20

Sharing Cosmic Dawn data WITHIN collaboration

- Huge simulations => raw & reduced data sharing essential for scientific exploitation, within collaboration
  - Full access for collaborators at Computing Center (Oak Ridge)
  - In Strasbourg: fraction of reduced data kept on local server, creation of accounts for collaborators
  - At Collaborators' institutes
- Requires regular communication & collab meetings
- User guide and status document
- Transfer is a hassle (dozens of TB across the globe)
  - BUT globus is so very awesome!
  - BUT not endorsed / accepted everywhere / not free=> alternatives?


Pierre ocvirk
Observatoire astronomique de Strasbourg
INCITE project AST031
28/06/2022

A SHORT USER'S GUIDE TO THE SUMMIT RAMSES-CUDATON SIMULATION COSMIC DAWN III (AND ONE BIG FAQ)

This document is in the process of being updated for CoDa II. The paths for the products are therefore not correct yet but the sections describing dealing with units for instance are fine. Contact me for questions at pierre.ocvirk@astro.unistra.fr

This document is in 2 parts. The first one details the various CODA-II outputs available to collaborators. These outputs are in the reduced directory (/lustre/atlas1/ast031/proj-shared/clues/CoDaII/4096/128x128x64/reduced/). The sub-directories correspond to the 92 snapshots of the simulation. Each of these sub-directories may contain 4 sub-directories: cutouts, fofres, fullbox, and fullboxfullres. The following sections detail the content of each.

I - CUTOUTS

This directory contains a 512^3 cutout centered on the center of mass of the LG (directory 1). It has several cubes containing rho, ... , a star particles list and a dark matter particles list.

II - FULLBOX

In CoDaI. This directory contained a 2048^3 (i.e. degraded resolution from 4096^3) version of

Sharing Cosmic Dawn data OUTSIDE collaboration

- Simulation life time argument:
  - CoDaI: prod 2012, but Dawoodbhoy 2018
  - CoDaII: prod 2016, but Sorce et al. 2022, Park et al. 2022
- Mostly, interested collaborators will contact me and gain access to Oak Ridge Computing Center or server in Strasbourg.
- Currently only modest attempt at data publication through VizieR@CDS:
  - suited to small-ish, high-level data products (e.g. galaxy catalogs)
  - very simple interaction with CDS
  - fully interoperable, accessible through python/astroquery, pyvo
  - perennial
  - generates DOI

Summary and outlook/questions

- Beyond VizieR for high-level products, for sharing lower level data products:
  - obvious difficulty due to volumetry => reduced data set
  - Lossy compression for “reduced” data products (outputs of amr2cube)
  - Galactica?
- Transfer & share: “french” globus? (CCIN2P3?)
- …