# Early structure formation in the THESAN radiation-magneto-hydrodynamics simulations

**E. Garaldi[1]** , **R. Kannan[2], A. Smith[3]†, V. Springel[1], R. Pakmor[1],**
**M. Vogelsberger[3] and L. Hernquist[2]**

[1]Max-Planck Institute for Astrophysics, Karl-Schwarzschild-Str. 1,
D-85741 Garching, Germany

[2]Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge,
MA 02138, USA

[3]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

**Abstract.** The formation of the first galaxies in the Universe is the new frontier of both galaxy formation and reionization studies. This creates a fierce new challenge, i.e. to simultaneously understand in a unique and coherent picture the processes of galaxy formation and reionization, and – crucially – their connection. To this end, we present the THESAN suite of cosmological radiation-magneto-hydrodynamical simulations. They are unique since they: (i) cover a very broad range of spatial and temporal scales; (ii) include an unprecedentedly-broad range of physical processes for simulations of such scales and resolution; (iii) exploit knowledge accumulated at low redshift to minimize the number of free parameters in the physical model; (iv) use a variance-suppression technique in the production of initial conditions to increase their statistical fidelity. Finally, the THESAN suite includes multiple runs of the same initial conditions, exploring current unknowns in the physics of dark matter and ionizing sources.

**Keywords.** galaxies: high-redshift – cosmology: large-scale structure of universe – radiative transfer – methods: numerical

## 1. Introduction

The Epoch of Reionization (EoR) is a landmark in the evolution of the Universe. During this time, the hydrogen residing in the intergalactic medium (IGM) between galaxies was ionised by the energetic radiation produced by stars within primeval galaxies. As such, the EoR links the high-redshift cosmic structures with the Universe today, providing ground for assessing our understanding of galaxy formation (which comes mostly from low-redshift observations) and playing a key role in shaping the formation and evolution of galaxies. For these reasons, the formation of galaxies at redshift $z \gtrsim 6$ is the new frontier in the study of both galaxy formation and cosmic reionization.

Despite its importance, the EoR is among the least understood epochs in galaxy evolution. The reasons are twofold. On the one hand, observationally probing such a remote epoch is remarkably challenging, and only recently have surveys (e.g. REBELS and ALPINE) and instruments (e.g. `ALMA`, `JWST`, `CCAT-p`, `SPHEREx`, `LOFAR`, `MWA`, `HERA` and `SKA`) started to provide detailed information on the $z \gtrsim 6$ Universe. On the other hand, theoretical investigation is hindered by the multi-scale nature of the physical processes involved. These difficulties have not, however, prevented us from developing a relatively

† NHFP Einstein Fellow.

coherent picture of the EoR in the last decades. In particular, its timeline is constrained by a number of measurements to the redshift range $5.5 \lesssim z \lesssim 10$ and its tail-end (where individual ionized regions coalesce) to occur at $5 \lesssim z \lesssim 6$. These measurements come from the evolution of Lyman$\alpha$ – hereafter Ly$\alpha$ – emitters and the number density of Lyman break galaxies (Ota et al. 2008; Ono et al. 2012; Pentericci et al. 2014; Choudhury et al. 2015; Tilvi et al. 2014; Mesinger et al. 2015), from the dark pixel statistics (McGreer et al. 2011, McGreer et al. 2015, Lu et al. 2020), from GRB afterglow (Totani et al. 2006; Chornock et al. 2006), from the Doppler widths of Ly$\alpha$ absorption lines in the quasar near zones (Bolton et al. 2012), from the quasar damping wing (Mortlock et al. 2011, Schroeder et al. 2013, Greig et al. 2017, Greig et al. 2019, Wang et al. 2020), from CMB modeling (Robertson et al. 2013), from the Gunn-Peterson optical depth (Fan et al. 2006; Davies et al. 2018; Yang et al. 2020; Bosman et al. 2021), from the angular correlation function of Ly$\alpha$ emitters (Sobacchi & Mesinger 2015), from the rest-frame UV continuum of galaxies (Schenker et al. 2014), from the detection of Ly$\alpha$ emission in Lyman break galaxies (Mason et al. 2018, Mason et al. 2019, Hoag et al. 2019, Jung et al. 2020), from a combination of Ly$\alpha$ luminosity function, clustering and line profile (Ouchi et al. 2010) and from the Ly$\alpha$ visibility (Dijkstra et al. 2011). This process was most likely powered by a large number of star-forming galaxies with stellar mass $M_{\rm star} \lesssim 10^8 \, \rm M_\odot$ and relatively-high escape fractions of ionizing photons (see e.g. Haardt & Madau 2012).

Numerical studies of the EoR are among the most challenging in astrophysics, as they simultaneously necessitate large volumes, adequate resolution to capture sub-galactic processes, and an accurate treatment of Radiation Transport (RT). While a handful of simulations with at least some of these characteristics have become available in the last few years (Aurora, SPHINX, Renaissance, Obelisk, CoDa, CoDaII, CROC), they often compromise on one of them. Most importantly, these simulations are only calibrated and tested against observations at $z \gtrsim 5$, which are not only scarce and of lower quality compared to $z = 0$ ones, but also are degenerate with the astrophysical processes these simulations aim at studying.

In this proceedings we will introduce THESAN, a new suite of simulations designed to bypass the aforementioned problems and provide a comprehensive view of a large volume of the high-$z$ Universe. We intend that this suite will provide a firm ground upon which future investigations can be built. For this reason, we will make the data public in the near future and welcome proposals for collaborative work.

## 2. Design and calibration

The THESAN simulations are run using the AREPO code (Springel 2010), which solves the equations of magneto-hydrodynamics on a moving mesh, generated as the Voronoi tessellation of a set of mesh-generating points that approximately follow the gas flow. Gravity is computed using a hybrid Tree-PM approach, which separately computes long-range (via a particle mesh approach) and short-range (through a hierarchical oct-tree) forces. The time integration of these equations is carried out in a hierarchical fashion.

One of the main characteristics of THESAN is the fact that it features a suite of simulations exploring the (quite uncertain) high-redshift physics. For computational reasons, such exploration is carried out using a slightly lower resolution compared to the main run (named THESAN-1). These complementary runs explore the effects of numerical resolution (THESAN-2), of the galaxy mass-dependence of photon escape (THESAN-HIGH-2 and THESAN-LOW-2), and of the nature of dark matter (THESAN-SDAO-2). In addition, we performed simulations using the original IllustrisTNG model (THESAN-TNG-2 and THESAN-TNG-SDAO-2), without any radiation (THESAN-NORT-2), as well as dark-matter only runs that extend to $z = 0$ (THESAN-DARK-1 and THESAN-DARK-2). The main features of these runs are reported in Table 1. All simulations share

**Table 1.** Overview of the THESAN simulation suite. From left to right the columns indicate the name of the simulation, initial particle number, the mass of the dark matter and gas particles, the minimum softening length of gas, star and dark matter particles, the minimum cell size at $z = 5.5$, the final redshift of the simulation, and the escape fraction of ionizing photons from the birth cloud (if applicable).

| Name | $N_{\rm part}$ | $m_{\rm DM}$ [$M_\odot$] | $m_{\rm gas}$ [ckpc] | $\epsilon$ [pc] | $r_{\rm cell}^{\rm min}$ [pc] | $z_{\rm end}$ | $f_{\rm esc}$ |
|---|---|---|---|---|---|---|---|
| THESAN-1 | $2 \times 2100^3$ | $3.12 \times 10^6$ | $5.82 \times 10^5$ | 2.2 | $\sim 10$ | 5.5 | 0.37 |
| THESAN-2 | $2 \times 1050^3$ | $2.49 \times 10^7$ | $4.66 \times 10^6$ | 4.1 | $\sim 35$ | 5.5 | 0.37 |
| THESAN-WC-2 | $2 \times 1050^3$ | $2.49 \times 10^7$ | $4.66 \times 10^6$ | 4.1 | $\sim 35$ | 5.5 | 0.43 |
| THESAN-HIGH-2 | $2 \times 1050^3$ | $2.49 \times 10^7$ | $4.66 \times 10^6$ | 4.1 | $\sim 35$ | 5.5 | 0.8 |
| THESAN-LOW-2 | $2 \times 1050^3$ | $2.49 \times 10^7$ | $4.66 \times 10^6$ | 4.1 | $\sim 35$ | 5.5 | 0.95 |
| THESAN-SDAO-2 | $2 \times 1050^3$ | $2.49 \times 10^7$ | $4.66 \times 10^6$ | 4.1 | $\sim 35$ | 5.5 | 0.55 |
| THESAN-TNG-2 | $2 \times 1050^3$ | $2.49 \times 10^7$ | $4.66 \times 10^6$ | 4.1 | $\sim 35$ | 5.5 | $-$ |
| THESAN-TNG-SDAO-2 | $2 \times 1050^3$ | $2.49 \times 10^7$ | $4.66 \times 10^6$ | 4.1 | $\sim 35$ | 5.5 | $-$ |
| THESAN-NORT-2 | $2 \times 1050^3$ | $2.49 \times 10^7$ | $4.66 \times 10^6$ | 4.1 | $\sim 35$ | 5.5 | $-$ |
| THESAN-DARK-1 | $2 \times 2100^3$ | $3.70 \times 10^6$ | $-$ | 2.2 | $-$ | 0.0 | $-$ |
| THESAN-DARK-2 | $2 \times 1050^3$ | $2.96 \times 10^7$ | $-$ | 4.1 | $-$ | 0.0 | $-$ |

the same initial conditions, and follow the evolution of a patch of the Universe of size $L_{\rm box} = 95.5\, h^{-1}$ Mpc.

## 2.1. *The galaxy formation model*

To maximise the physical fidelity of THESAN across the entire history of the Universe, we employ as sub-resolution physics prescriptions the IllustrisTNG model (Weinberger et al. 2017; Pillepich et al. 2018), which has been extensively investigated at low redshift and proved to match many of the observed galaxy properties. This choice ensures that (i) the THESAN simulations have a galaxy formation model that is plausible throughout the entire history of the Universe, and (ii) the simulations have a single free parameter to be tuned (the unresolved stellar escape fraction of ionising photons, $f_{\rm esc}$), since all other parameters are kept fixed to the original IllustrisTNG values. The IllustrisTNG model includes, among others, the following: stochastic star formation, mass, energy and metal return from supernovae and AGB stars, explicit tracking of 9 metal species, stellar winds, black holes accretion and (bimodal) feedback, and magnetic fields originating from a uniform cosmological seed field.

The only modification to the original IllustrisTNG model is the replacement of its spatially-uniform UV background with a self-consistent RT scheme, which allows us to properly capture the EoR and that we describe in the following section. Finally, we also include in THESAN a model for dust evolution (following McKinnon et al. 2017), which includes production from stellar evolution and ISM growth, and destruction from supernovae, astration and thermal sputtering.

## 2.2. *Self-consistent radiation transport*

The propagation of radiation is followed using a moment-based approach, where the first two moments of the RT equation are solved employing the M1 closure relation (Levermore 1984), thanks to the AREPO-RT (Kannan et al. 2019) extension of AREPO. To reduce the computational load, only the UV part of the source spectrum is tracked employing three energy bins defined by the following thresholds: $[13.6, 24.6, 54.4, \infty)\, {\rm eV}$, in order to bracket the ionization energies of hydrogen and helium. Each resolution element tracks, for each bin, the comoving photon number density and flux. Within each of these bins, radiation is assumed to follow the spectral *shape* of a 2 Myr old, quarter-solar metallicity stellar population, as predicted by the BPASS library
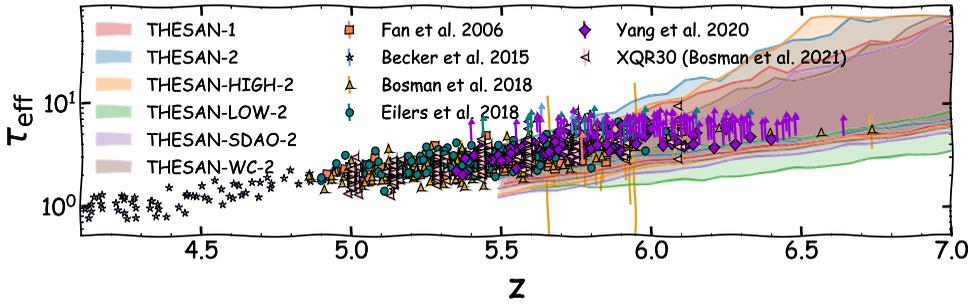
**Figure 1.** Evolution of the effective Ly$\alpha$ optical depth ($\tau_{\mathrm{eff}}$). The shaded regions show the central 95% of the data computed from synthetic spectra, while symbols report individual observations as indicated in the legend.

(Eldridge et al. 2017), which we use to compute the source spectra including the contribution from binary stellar systems. The predicted photon flux from the BPASS library is attenuated by a constant stellar escape fraction $f_{\mathrm{esc}}$ that captures the effect of the (unresolved) birth cloud of the star particles. The value of $f_{\mathrm{esc}}$ is approximately tuned to obtain the desired reionization history, which we match to the so-called 'late' reionization (i.e. an EoR ending at $z \lesssim 5.5$).

## 3. First results

We begin by showing the evolution of the Ly$\alpha$ effective optical depth ($\tau_{\mathrm{eff}}$) in a selection of THESAN runs in Fig. 1. Almost all THESAN runs agree well with the measurements available in the redshift range where they overlap ($5.5 \leq z \lesssim 6.5$). The only exception is THESAN-LOW-2, where only galaxies hosted by haloes of mass $M_{\mathrm{halo}} \leq 10^{10}\,\mathrm{M}_\odot$ are allowed to have non-zero escape fraction. This run completes reionization much earlier than the others (since small galaxies are mostly dominant at higher redshift), and therefore has a significantly more transparent IGM than the other THESAN runs. When investigating the distribution of $\tau_{\mathrm{eff}}$ in THESAN-1 at a given redshift (not shown), we find that it is slightly shifted to more transparent values with respect to observations, suggesting a (slightly) too early reionization.

### 3.1. *The IGM – galaxy connection*

The combination of large volume, high resolution and realistic galaxy formation model renders THESAN a unique tool to study the connection between the IGM properties and the high-$z$ galaxy population. In this Section, we do so by investigating how the Ly$\alpha$ flux in quasars (QSOs) depends on the distance between the line of sight and nearby galaxies. In particular, we follow the approach originally suggested in Kakiichi et al. (2018) and study the quantity $\langle T(r) \rangle / \bar{T} - 1$, where $T(r)$ is the transmissivity (i.e. the transmitted flux divided by the continuum value at the same location) at a given distance $r$ between a pixel in the QSO spectrum and a galaxy, the angular brackets indicate an average over all galaxy – pixel pairs at distance $r$, and $\bar{T}$ indicates the average over all pairs, regardless of their distance.

In Fig. 2 we show how the (normalised) transmissivity depends on distance for different galaxy selections at $z = 5.5$. In particular, lines in different shades of grey correspond to selections based on the mass of the dark matter halo hosting the galaxy, the purple line displays the effect of selecting galaxies with stellar mass $M_{\mathrm{star}} \geq 10^{10}\,\mathrm{M}_\odot$ and the orange line shows the result of identifying galaxies based on their CIV absorption along the same QSO sightline. We compare our results to two observational campaigns. Unlike previous
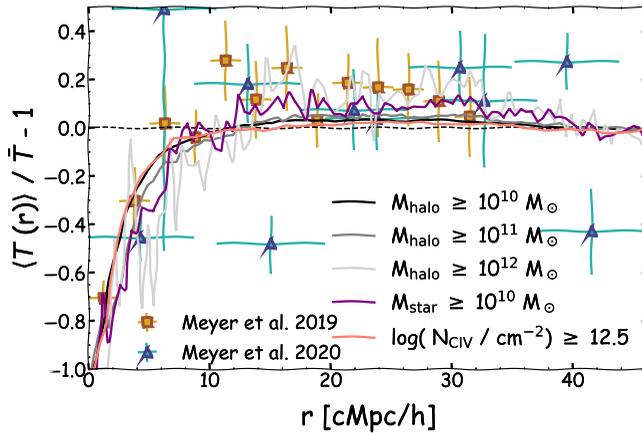
**Figure 2.** Average (normalised) transmitted flux $\langle T(r) \rangle$ as a function of the galaxy-spectrum distance $r$, normlised by its global average $\bar{T}$, for different galaxy-selection criteria at $z = 5.5$.
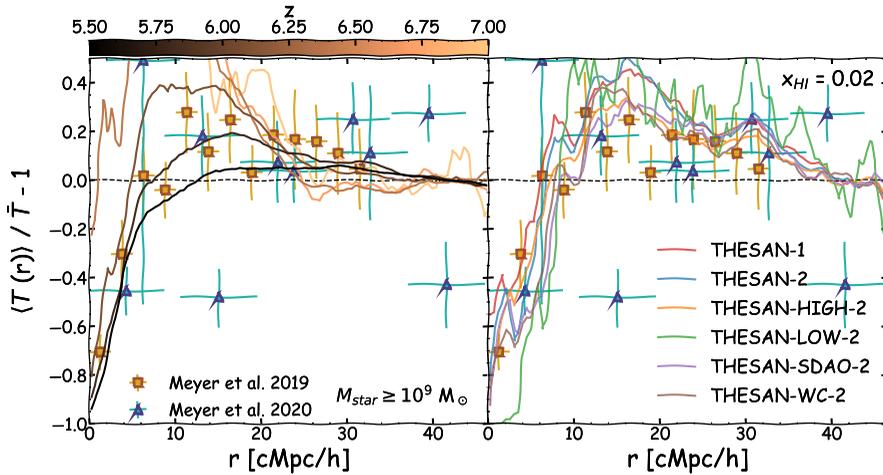


**Figure 3.** Same as Fig. 2, but only showing the $M_{\mathrm{star}} \geq 10^9\,\mathrm{M}_\odot$ selection as a function of redshift (left panel) and of photon escape and dark matter model (right). Notice that in the right panel, we matched the volume-averaged neutral fraction $x_{\mathrm{HI}} = 0.02$ rather than the redshift in the different THESAN runs, in order to remove dependency on the reionization history.

studies, THESAN is able to recover for the first time the flux modulation as a function of $r$. However, the flux enhancement at intermediate scales $10 \lesssim r/(h^{-1}\,\mathrm{Mpc}) \lesssim 30$ is reproduced only by the most extreme of our galaxy selections (i.e. $M_{\mathrm{halo}} \geq 10^{12}\,\mathrm{M}_\odot$). This is, however, in tension with the estimated host mass of the observed galaxy population by Meyer et al. (2019), which is estimate to be $M_{\mathrm{halo}} \gtrsim 10^{10}\,\mathrm{M}_\odot$. Additionally, trying to mimic the selection procedure of Meyer et al. (2019) by selecting galaxies based on their CIV absorption results in a much smaller flux enhancement than observed.

In order to investigate the origin of this discrepancy between THESAN and the observations, we show in Fig. 3 how $\langle T(r) \rangle / \bar{T}$ depends on the redshift (left) and on some of the high-$z$ physics (right), for the $M_{\mathrm{star}} \geq 10^9\,\mathrm{M}_\odot$ selection. The left panel shows that this measure is very sensitive to the redshift, with the flux enhancement moving to smaller $r$ and larger values with increasing redshift. This is fully compatible with its interpretation as a proximity effect. The ionised bubbles around galaxies grow (on average) with time, producing a more extended region of flux enhancement (thanks to the suppressed neutral

fraction). Simultaneously, $\bar{T}$ strongly decreases with increasing redshift outside of these regions, following the evolution of the average neutral fraction in the universe. Hence, for a given flux within the ionised bubbles the ratio $\langle T(r) \rangle / \bar{T}$ is much higher at higher redshift.

The left panel of Fig. 3 also shows the sensitivity of $\langle T(r) \rangle / \bar{T}$ to the timing of reionization. Hence, it can be used as promising tool to constrain the latter, provided it is not (too) sensitive on other properties of the EoR. We test this in the right panel of Fig. 3, where we report the flux modulation in the different THESAN runs, hence exploring different photon escape and dark matter models. Notice that we have factored out the different reionization histories of these runs by matching the volume-averaged neutral fraction ($x_{HI} = 0.02$) in the runs instead of their redshift. We note here that there will be some residual effect due to the simulations being at slightly different stages of the structure formation process, but the reionization histories are similar enough to render this residual difference negligible for all but the THESAN-LOW-2 run, which completes reionization significantly earlier than the others. The panel shows that $\langle T(r) \rangle / \bar{T}$ is quite insensitive to the specific radiation escape and dark matter models employed, making it possible to use the flux excess as a way to constrain the timing of reionization. Hence, comparing the values obtained with available observations, we conclude that reionization occurs slightly too early in THESAN, despite it being designed to simulate a 'late' reionization history ending at $z \lesssim 5.5$. This is consistent with the conclusion obtained from the analysis of the effective optical depth distribution.

## Discussion

*Christian Boily*: Regarding the optical depth, you said your modelling matches the data points. It seems to me that your models were actually significantly below the lower [observational] values.

*Enrico Garaldi*: Of course it depends on the source model. What I am referring to is the red band in the plot, which corresponds to our fiducial run. It goes through the data fairly well. I did not have the time to show it, but what you can do is to take thin redshift slices and check the distribution of optical depth at any given redshift. We match the observations well, although we are slightly too transparent at every redshift, which indicates that an even-later reionization better matches current data, but the difference is not large. This green band instead is THESAN-LOW-2, the model where only low-mass galaxies contribute to reionization, and it is instead in significant tension with the data. There is of course the caveat that all these models have an escape fraction included.

*Mahavir Sharma*: I wonder if you have a prediction on the 21cm power spectrum and a comparison to the MWA or LOFAR data.

*Enrico Garaldi*: There is definitely in the papers, but I have here only one plot. These are the predictions for different redshifts, and the straight lines are the predicted sensitivities for LOFAR, HERA and SKA-LOW.

## References

Ota K., et al., 2008, ApJ, 677, 12
Ono Y., et al., 2012, ApJ, 744, 83
Pentericci L., et al., 2014, ApJ, 793, 113
Choudhury T. R., Puchwein E., Haehnelt M. G., Bolton J. S., 2015, MNRAS, 452, 261
Tilvi V., et al., 2014, ApJ, 794, 5

Mesinger A., Aykutalp A., Vanzella E., Pentericci L., Ferrara A., Dijkstra M., 2014, MNRAS, 446, 566

McGreer I. D., Mesinger A., Fan X., 2011, MNRAS, 415, 3237

McGreer I. D., Mesinger A., D'Odorico V., 2015, MNRAS, 447, 499

Lu T.-Y., et al., 2020, ApJ, 893, 69

Totani T., Kawai N., Kosugi G., Aoki K., Yamada T., Iye M., Ohta K., Hattori T., 2006, PASJ, 58, 485

Chornock R., Berger E., Fox D. B., Lunnan R., Drout M. R., Fong W.-f., Laskar T., Roth K. C., 2013, ApJ, 774, 26

Bolton J. S., Becker G. D., Raskutti S., Wyithe J. S. B., Haehnelt M. G., Sargent W. L. W., 2012, MNRAS, 419, 2880

Mortlock D. J., et al., 2011, Nature, 474, 616

Schroeder J., Mesinger A., Haiman Z., 2013, MNRAS, 428, 3058

Greig B., Mesinger A., Haiman Z., Simcoe R. A., 2017, MNRAS, 466, 4239

Greig B., Mesinger A., Bañados E., 2019, MNRAS, 484, 5094

Wang F., et al., 2020, ApJ, 896, 23

Robertson B. E., et al., 2013, ApJ, 768, 71

Fan X., et al., 2006, AJ, 132, 117

Davies F. B., et al., 2018, ApJ, 864, 142

Yang J., et al., 2020, ApJ, 904, 26

Bosman S. E. I., et al., 2021, arXiv e-prints, p. arXiv:2108.03699

Sobacchi E., Mesinger A., 2015, MNRAS, 453, 1843

Schenker M. A., Ellis R. S., Konidaris N. P., Stark D. P., 2014, ApJ, 795, 20

Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2018, ApJ, 856, 2

Mason C. A., et al., 2019, MNRAS, 485, 3947

Hoag A., et al., 2019, ApJ, 878, 12

Jung I., et al., 2020, ApJ, 904, 144

Ouchi M., et al., 2010, ApJ, 723, 869

Dijkstra M., Mesinger A., Wyithe J. S. B., 2011, MNRAS, 414, 2139

Haardt F., Madau P., 2012, ApJ, 746, 125

Springel V., 2010, MNRAS, 401, 791

Weinberger R., Springel V., Hernquist L., Pillepich A. et al., 2017, MNRAS, 465, 3291

Pillepich A., Nelson D., Hernquist L., Springel V., Pakmor R., et al., 2018, MNRAS, 473, 4077

McKinnon R., Torrey P., Vogelsberger M., Hayward C. C., Marinacci F., 2017, MNRAS, 468, 1505

Levermore C. D., 1984, J. Quant. Spectrosc. Radiative Transfer, 31, 149

Kannan R., Vogelsberger M., Marinacci F., McKinnon R., et al. 2019, MNRAS, 485, 117

Eldridge J. J., Stanway E. R., Xiao L.,et al. 2017, Publ. Astron. Soc. Australia, 34, e058

Fan X., Strauss M. A., Becker R. H. White R. L. et al., 2006, AJ, 132, 117

Becker G. D., Bolton J. S., Madau P., Pettini M. et al., 2015, MNRAS, 447, 3402

Bosman S. E. I., Fan X., Jiang L., Reed S., Matsuoka Y. et al., 2018, MNRAS, 479, 1055

Bosman S. E. I., Davies F. B., Becker G., Keating L. et al., 2021, arXiv e-prints (2108.03699)

Eilers A.-C., Davies F. B., Hennawi J. F., 2018, ApJ, 864, 53

Yang J., Wang F., Fan X., Hennawi J., Davies F. et al., 2020, ApJ, 904, 26

Kakiichi K., Ellis R. S., Laporte N., Zitrin A., Eilers A.-C. et al., 2018, MNRAS, 479, 43

Meyer R. A., Bosman S. E. I., Kakiichi K., Ellis R. S., 2019, MNRAS, 483,19

Meyer R. A., Kakiichi K., Bosman S. E. I., Ellis, R. S. et al., 2020, MNRAS, 494, 1560