Large-scale 21 cm signal predictions at cosmic dawn with calibrated subgrid galaxy formation

N Gillet, D Aubert, F Mertens, P Ocvirk

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **507,** 3179–3186 (2021)
Advance Access publication 2021 July 16

https://doi.org/10.1093/mnras/stab2010

Large-scale 21 cm signal predictions at cosmic dawn with calibrated subgrid galaxy formation

N. J. F. Gillet [1]★ D. Aubert,[1] F. G. Mertens[2,3] and P. Ocvirk [1]

[1] *Observatoire Astronomique de Strasbourg, Université de Strasbourg, CNRS UMR 7550, 11 Rue de l'Université, F-67000 Strasbourg, France*
[2] *Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700 AV Groningen, the Netherlands*
[3] *LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universite, F-75014 Paris, France*

Accepted 2021 July 12. Received 2021 July 12; in original form 2021 March 4

ABSTRACT

The process of heating and reionization of the Universe at high redshifts links small-scale structures/galaxy formation and large-scale intergalactic medium (IGM) properties. Even if the first one is difficult to observe, an observation window is opening on the second one, with the promising development of current and future radio telescopes. They will permit to observe the 21 cm brightness temperature global signal and fluctuations. The need for large-scale simulations is therefore strong to understand the properties of the IGM that will be observed. However, at the same time, the urge to resolve the structures responsible for those processes is important. We introduce in this study a simulation framework of the cosmic dawn and reionization, based on hydrodynamics and radiative transfer code and a simple subgrid galaxy formation process for 1 Mpc-resolution simulations. Here, this model is calibrated on the state-of-the-art simulation CoDaII. This scheme permits us to follow consistently dark matter, hydrodynamics, and radiative transfer evolution on large scales, while the subgrid model deals with the galaxy formation scale, in particular, taking into account the different feedback on the star formation. We process the simulation to produce a simulated 21 cm signal as close as possible to the observations.

Key words: galaxies: formation – radiative transfer – reionization.

1 INTRODUCTION

Despite the fact that the first billion years are full of events, the properties of the Universe between the emission of the cosmic microwave background (CMB) and redshift $z = 6$ are still poorly constrained by observations. It sees the cosmic dawn (CD), the birth and growth of the first structures, stars, and galaxies, as well as the cosmological change of properties of the intergalactic medium (IGM), from cold and neutral to hot and ionized during the epoch of heating (EoH) and epoch of reionization (EoR). Observational prospects seem promising, with, e.g. high-redshift galaxies probed by the *James Webb Spatial Telescope* or the avalanche of data from the current and in-development radio telescopes that will measure the IGM properties on large scales. In this work, a special focus will be put on this latter type of instruments such as LOw Frequency ARray (LOFAR; van Haarlem et al. 2013), which acquires data between 200 and 110 MHz (between redshifts 6 and 12), therefore focusing on the end of the EoH and the EoR. The LOFAR EoR Key Science Project has recently put upper limits on the power spectrum (PS) of the cosmic 21 cm signal at redshift 9.1 (Mertens et al. 2020). Another instrument, the New Extension in Nançay Upgrading loFAR (NENUFAR; Zarka et al. 2012), acquires data between 85 and 30 MHz (from redshifts 16 to 45) and therefore overlaps with the frequency range of claimed detection of the global signal at redshift 17 of the EDGES instrument (Bowman et al. 2018). In parallel, theoretical modelling of the physics and the signal is ongoing and aims at following the structure and galaxy formation on small scales and its impact on the properties of the IGM on large scales. The theoretical challenge is to do both: resolving the birth and properties of the first galaxies and their photons emission [Ly α, X-rays, and ultraviolet (UV) for example], and tracking the evolution of the IGM properties on cosmological distances.

Several groups address this challenge by using analytical models of galaxy formation (most often based on the local collapse mass fraction or halo mass function) and seminumerical treatments of the reionization (e.g. Visbal et al. 2012; Fialkov, Barkana & Visbal 2014; Park et al. 2019). These methods have the advantage to have a comprehensible set of galaxy formation parameters and be computationally efficient. Alternatively, others push to directly solve all scales with coupled hydrodynamics–radiative transfer simulations (e.g. Gnedin 2016; Ocvirk et al. 2016, 2020; Semelin et al. 2017). However, the trade-off between resolution and volume makes those simulations difficult to realize and costly while being still limited in the range of halo masses or cosmological scales that can be probed. A final alternative is to perform the radiative transfer in post-process on top of dynamics-only simulations. It uses high-resolution dark matter haloes to support a galaxy formation model while providing a realistic propagation of photons (e.g. Chardin, Puchwein & Haehnelt 2017; Kulkarni et al. 2019; Ross et al. 2019). The gain in computational time can be significant compared to fully coupled simulation but this method cannot probe the full extent of the respective feedbacks of matter and radiation, even if this coupling is weak at the scale involved here.

★ E-mail: nicolas.gillet@astro.unistra.fr

In this article, we present another alternative to produce large-scale simulations (>250 cMpc needed for proper IGM properties; Iliev et al. 2014; Kaur, Gillet & Mesinger 2020) in the context of current and future radio experiments of the CD.

It relies on a fully coupled radiative transfer–hydrodynamics code and a subgrid model of galaxy formation. The code permits to follow the hydrodynamics and the radiative transfer within the same framework, even if the coupling between the two is weak at the scale of interest here (∼1 cMpc). In practice, it means that we can process the simulation in one run, without having to store intermediate data and post-process it. The computational cost is small (compared to higher resolution), about 3000 cpuh for a simulation of 512^3 elements at 1 cMpc resolution. Furthermore, using this code in this regime provides the practical benefit to use the same framework for future zoom/high-resolution simulations in subvolumes, which will indeed require coupled radiative hydrodynamics and adaptive mesh refinement (AMR). Meanwhile, the subgrid model is necessary to take into account all the unresolved subgrid physics, mainly the star formation (SF), the radiative feedback on the SF, and absorption. In this study, we focus on subgrid SF modelling. Nasirudin, Iliev & Ahn (2020) have developed a robust methodology to generate subgrid mock halo population from which one can implement a galaxy formation model. Inspired by Nasirudin et al. (2020), we propose that the unresolved SF could be based on state-of-the-art high-resolution simulations of the EoR such as CoDaII (Ocvirk et al. 2020). This allows us to take into account, in the subgrid model, all the feedback on SF that is resolved in the CoDaII simulation.

This technique is implemented in the EMMA cosmological simulation code (Aubert, Deparis & Ocvirk 2015) and is demonstrated in the following sections. We show that this methodology leads to viable and consistent predictions of the 21 cm radio signal from the CD. We introduce the calibrated subgrid source formation model in Section 2 and then discuss the resulting large-scale 21 cm signal predictions in Section 3.

2 SOURCE FORMATION MODEL AND SIMULATION

In this study, we use the full physics cosmological simulation code for reionization EMMA (Aubert et al. 2015) in a large-scale/low-resolution mode. We extend the code with a new empirical source (star/galaxy) formation model based on the CoDaII simulation that provides more flexibility at high redshifts than standard methods and we also add simple prescriptions for the prediction of the 21 cm signal.

2.1 Sources

The challenge of large-scale/low-resolution simulations is to assign a production of ionizing photons per volume unit despite the lack of dense, non-linear structures in simulations. At high redshifts ($z > 6$), the main sources of UV photons are young massive stars: We have to find a way to assign to each resolution element a star formation rate (SFR), to follow the creation of stars, i.e. the sources of UV photons. A subgrid model has to be constructed, that would assign the production of photons as a function of the local structure formation.

Classically, semi-analytical galaxy formation models rely on the underlying dark matter collapse fraction and halo mass function (e.g. Fialkov et al. 2014; Park et al. 2019). It assumes that galaxies form in haloes and the SF depends on the halo's mass. In this study, we take an alternative approach. Instead of trying to resolve and follow dark matter structure formation and evolution, we simply suppose that a fraction of the gas will be star forming on megaparsec scales. We develop an empirical large-scale galaxy formation model based on the results of the state-of-the-art high-resolution, hydroradiative simulation of the reionization, CoDaII (Ocvirk et al. 2020).

2.1.1 SF in CoDaII

The CoDaII simulation has a box of $64\,h^{-1}$ cMpc side sampled on a Cartesian grid of 4096^3 elements. In a very standard manner, the production of stellar particles during the simulation is driven by an SFR density computed at each time-step, according to

$$\mathrm{SFR}^H_\varphi \propto \epsilon_* \rho^{1.5}, \text{ where } \rho > \rho_*. \tag{1}$$

In the CoDaII simulation, the SFR is directly proportional to the density at power 1.5, where $\epsilon_* = 0.02$ is the SF efficiency and $\rho_*/f_\Omega = \Delta_* = 50$ is the SF density threshold ($f_\Omega = \Omega_\mathrm{b}/\Omega_\mathrm{m}$ is the baryonic fraction).

In post-processing, we degrade the simulation outputs on a coarse grid of 64^3 cells corresponding to our $1\,h^{-1}$ cMpc goal resolution. In each cell, we compute a 'coarsened' density contrast ($\Delta = \rho/\langle\rho\rangle$) and a 'coarsened' SFR, using 10 snapshots between redshifts 5.7 and 15. This post-processed SFR density of CoDaII simulation is computed in each coarse $1\,h^{-1}$ cMpc cell as the sum of the stellar particle masses younger than 10 Myr, divided by 10 Myr. Hereafter, physical scale quantities are annotated with the letter 'φ' and comoving one with a 'c'. Low- and high-resolution quantities are annotated with 'L' and 'H', respectively, and refer to $1\,h^{-1}$ cMpc or the original CoDaII resolution ($15.625\,h^{-1}$ckpc). One might directly apply equation (1) on the low-resolution grid, but it results in a too permissive SF at high redshifts ($z \sim 30$). Indeed, the density contrast is smaller at low resolution at every redshift; therefore having one fixed density threshold would produce an almost flat cosmic SFR evolution with redshift, too much star at high redshift, or not enough at low redshift. We need to change the threshold parameter to mimic the subgrid collapse structures and control the SFR at high and low redshifts. Furthermore, the classical scheme applied at low resolution cannot take into account the subgrid quenching of the reionization on the smallest galaxies. To take this effect into account, we derive an empirical model based on the outputs of the CoDaII simulation.

2.1.2 Subgrid SFR

In the CoDaII case, each coarse cell ($1\,h^{-1}$ cMpc) is composed of 64^3 high-resolution cells. Each of them can be star forming (SFR $\propto \Delta^{1.5}$, cf. equation 1). Therefore, we derive that the low-resolution SFR on the comoving scale is defined as follows:

$$\mathrm{SFR}^L_\mathrm{c} = \bar{\epsilon_*} \Sigma^L_* \frac{1}{a^{1.5}_\mathrm{exp}}, \tag{2}$$

where $\bar{\epsilon_*}$ is the proportionality factor that absorbs all the constants. The expansion factor dependence comes from the physical to comoving transformation and the power 1.5 dependence to the density (cf. equation 1). Also, we define Σ_* as the star-forming gas density at the power 1.5 in each coarse cell: We call it the 'proxy to the star-forming gas: PSFG'. It is computed in the CoDaII simulation post-processed outputs as the sum of the density at power 1.5 of star-forming cells:

$$\Sigma^L_* = \sum \left(\Delta^H_i\right)^{1.5}, \text{ where } \Delta > \Delta*, \tag{3}$$

where the iterator i stands for each of the 64^3 high-resolution cells in a coarse cell of $1\,h^{-3}\,\mathrm{cMpc}^3$. Fig. 1 presents the PSFG for all coarse cells of $1\,h^{-3}\,\mathrm{cMpc}^3$ as a function of the overdensity. The

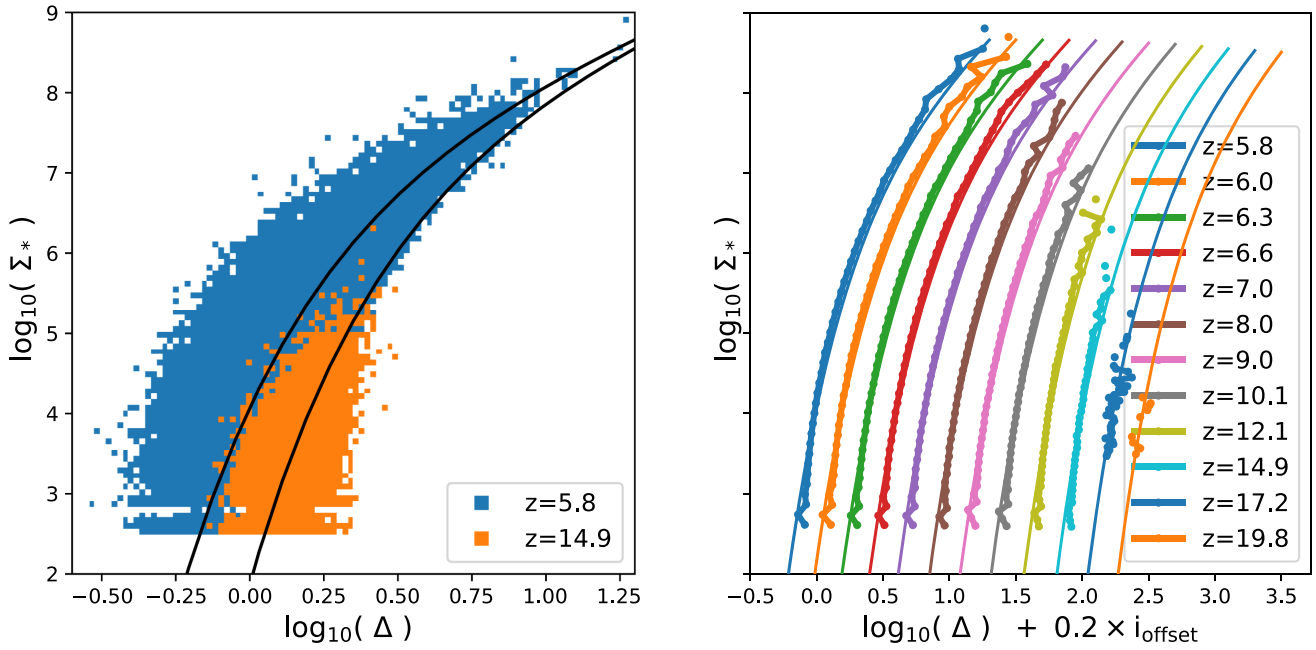

Figure 1. The PSFG density (Σ_*) as a function of density (Δ): On the left, the distribution of all coarse cells is shown at redshifts 6 and 15, with the fitted function in black, as in Shapiro et al. (in preparation). The right-hand panel presents the mean relations and their fits at all available redshifts. Each redshift relation is shifted by 0.2 dex for clarity. Note that at redshifts 17 and 20, we used the 2048^3 data cube instead of 4096^3 data cube, which explains the difference in resolution accessible in Σ_* for those two redshifts.

left-hand panel presents the distribution of (Δ^L, Σ_*^L) pairs in CoDaII at redshifts 6 and 15. At high density, Σ_* follows a power law as a function of the density contrast with a unit slope. The PSFG decreases sharply as the density contrast becomes smaller. Also, the scatter around the overall trend is large (for example, at $\Delta = 1$, Σ_* covers almost four orders of magnitude at redshift 6). The dispersion increases as the density decreases, and at the same time a hard minimum is set, imposed by the CoDaII simulation parameters ($\Sigma_{*,\min} = 50^{1.5}$, corresponding to a single high-resolution cell above the SF threshold in one coarse cell).

For the sake of simplicity, we model the mean behaviour of the (Δ^L, Σ_*^L) relation. Σ_* behaves as a power law for the density with an exponential cut-off at the low-density end. Large-volume simulation would likely have a larger density range than the calibration simulation, at the same resolution. At the low-density end, no special care is needed since the exponential cut-off ensures that SF is prevented. At the high-density end, we simply extrapolate the power law beyond the range of values in the calibration simulation. This model is purely empirical and does not take into account the dispersion induced by the variance in structure formation. However, it does take into account the underlying stellar and radiative feedback on the gas density implemented in the CoDaII simulation. The density of star-forming gas is parametrized as follows:

$$\Sigma_* = \epsilon_{\Sigma_*} \Delta 10^{-\Delta_{\Sigma_*}/\Delta}, \tag{4}$$

where ϵ_{Σ_*} is fitted at redshift 6 and kept constant at all higher redshifts, and Δ_{Σ_*} is adjusted at each redshift independently, as in Shapiro et al. (in preparation).

The evolution of the parameter Δ_{Σ_*} with redshift is obtained here for the CoDaII simulation. The mean evolution with redshift of the PSFG as function is shown in the right-hand panel of Fig. 1. Nevertheless, its evolution can be freely parametrized (empirically or physically) to explore or accommodate different scenarios and models of SF, for example, the inclusion of POP III stars, or more simply modulate the time evolution of the cosmic SFR. For the sake of simplicity, we consider a linear evolution of Δ_{Σ_*} with redshift, which is roughly consistent with the evolution given by CoDaII.

Fig. 1 is similar to the one presented in Nasirudin et al. (2020) (figs 6 and 7), where the number of haloes or the collapse mass fraction is presented instead of the PSFG. This is strongly expected as the star-forming gas should be in haloes; the more haloes there are , the greater will be the SFR. For the moment we model the average relationship between the density and the PSFG, we will take into account the dispersion around mean with a similar method as in Nasirudin et al. (2020).

2.1.3 *SF space distribution*

At this stage, every cell has a non-zero SFR. However, as we expect to have more SF in the densest regions, we also expect to have no SF in the most underdense ones and in between certain stochasticity. The left-hand panel of Fig. 2 illustrates the stochasticity by presenting the probability for a cell of $1\,h^{-3}$ cMpc^3 to have a non-zero SFR, as a function of the density contrast and redshift in the CoDaII simulation. The transition is smooth between high densities that always form stars and the low-density regions that do not. Also, this transition evolves with redshift, shifting towards low-density regions with time. At $z = 6$, a $1\,h^{-3}$ cMpc^3 volume with an average density has a 50 per cent probability to be star forming. Another way to visualize the stochasticity and the spatial distribution of the star-forming region is to look at the volume filling factor of star-forming cells, presented in the right-hand panel of Fig. 2. The blue line shows the SF volume filling factor of the CoDaII simulation, coarsened on scales of $1\,h^{-1}$ cMpc. The fraction of volume that forms stars

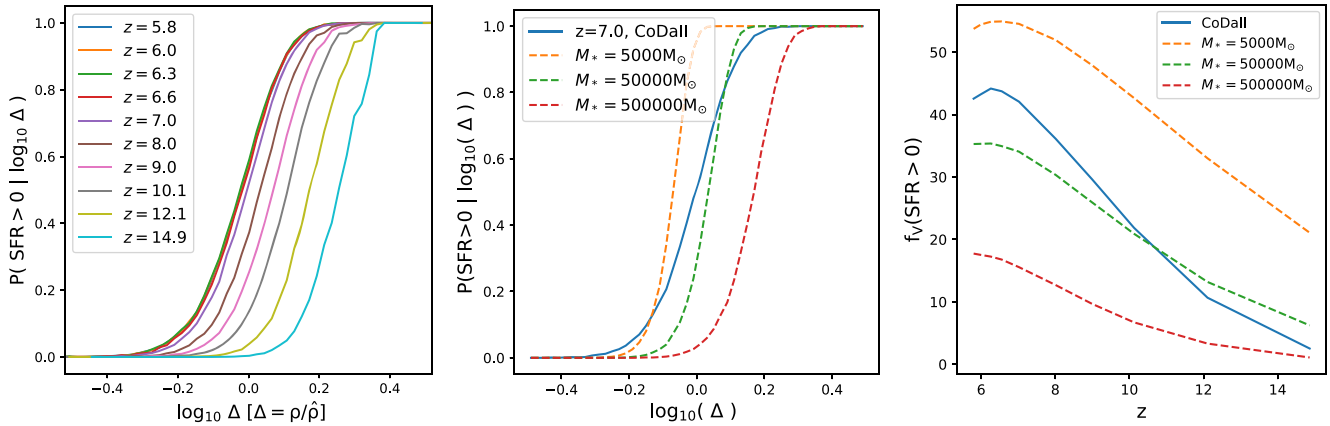

Figure 2. In the left-hand panel, the probability for a cell of $1\,h^{-3}$cMpc^3 of the CoDaII simulation to have a non-zero SFR as a function of density and redshift. Lines from redshifts 5.8 to 6.6 are almost identical and stack. In the middle panel, the same as the left, but only at redshift 7. The CoDaII simulation is the blue line, while the dashed orange, green, and red are the subgrid model with $M_* = 5000, 50\,000, 500\,000\,\mathrm{M}_\odot$, respectively. In the right-hand panel, the volume filling factor of non-zero SFR cells of $1\,h^{-3}$cMpc^3, for the CoDaII in blue and the subgrid models are the same as in the middle panel.

Table 1. Specifications of simulation: The cosmological parameters are extracted from Planck Collaboration VI (2020) table 2 last column (and $\Omega_\mathrm{b} = \Omega_\mathrm{b}h^2/h^2$).

Cosmology (Planck 18)	
Ω_Λ	0.6889
Ω_m	0.3111
Ω_b	0.04 897
h	0.6766
σ_8	0.8102
$\mathrm{n_{spec}}$	0.9665
Stars	
$\log_{10}\overline{\epsilon_*}$	−7.4
M_*	10^8 ($\mathrm{M}_\odot$)
t_*	10 (Myr)
z_{ON}	30
Radiation	
Stellar ionizing emissivity	4.32 10^{46} (ph s^{-1}M$_\odot^{-1}$)
f_esc	0.05
$\langle E_\mathrm{UV}\rangle$	20.65 (eV)
σ_UV	2.381 $\times 10^{-22}$ (m^2)
Speed of light	299 792 458 (m s^{-1})
Simulation specs	
Comoving resolution dx	1 (h^{-1} cMpc)
Dark matter particle mass	1.075 $\times 10^{11}$(M$_\odot$)

rises with redshift, with a maximum just below 50 per cent between redshifts 6 and 7. It means that, in the CoDaII simulation, almost half of the volume of the universe is star forming at redshift 7, smoothed on a scale of $1\,h^{-1}$ cMpc.

The local variations introduced above and the resulting SF spacial distribution will set the spatial evolution of the reionization process. It will affect the H II bubble size distribution and evolution, and the 21 cm temperature brightness PS too. Therefore, we introduce here one way to control the SF distribution in our simulations. We use a minimum stellar mass M_* and the SF process is discretized in stellar particles. With the same scheme as in CoDaII, the number of stellar particles created is drawn from a Poisson distribution. The mean SFR of a coarse cell is set by equation (2). Then, the mean stellar mass is obtained by multiplying by the time-step (dt), and the mean number of stellar particles is therefore obtained by dividing by the stellar mass particle $\bar{N}_* = \mathrm{SFR}_\mathrm{c}^L \times \mathrm{d}t/M_*$. In the end, the parameter M_* does the same as in high-resolution runs. It permits setting a minimum SFR in a cell and to cut SF where it is too low. However, the physical meaning of M_* is different. Here, it encompasses the local variations due to the SF and unresolved structure formation at the same time. We apply our new parametrization of the source formation to the outputs of the CoDaII simulation. The impact of M_* on the SF process is illustrated in the middle and right-hand panel of Fig. 2 with different M_* values: $5 \times 10^3\mathrm{M}_\odot$, $5 \times 10^4\mathrm{M}_\odot$, and $5 \times 10^5\mathrm{M}_\odot$ (orange, green, and red, respectively). This parameter controls the distribution of the SF as a function of the density, as shown in the middle panel of Fig. 2, which automatically translates to the volume filling factor, shown in the right-hand panel. Interestingly, as shown after, as the cosmic SF density is mostly set by the heaviest regions, this parameter does not affect the global SFR. Therefore, the global SFR and its spatial distribution are almost independent of this parametrization. The parameter M_* permits the choice between a 'diffuse' SFR distribution or a 'biased' SFR distribution. In the middle panel of Fig. 2, the differences in the slope between the CoDaII and the models come from the fact that only the average relationship between the density and the SFR is taken into account, and not the dispersion around the mean (see the left-hand panel of Fig. 1). We intend to develop the model in the future to take into account this dispersion (as in Nasirudin et al. 2020).

2.2 Simulation's set

The previously presented star formation model and an on-the-fly computation of the 21 cm signal (presented hereafter) have been added in the hydrodynamics–radiative transfer code EMMA (Aubert et al. 2015). It permits us to realize cosmological simulations of the CD, EoH, and EoR by coupling the evolution of dark matter, baryonic matter, source formation, and radiative transfer.

2.2.1 *Specifications*

We produce a ($512\,h^{-1}\,\mathrm{cMpc}^3$) simulation with a resolution of $1\,h^{-3}\,\mathrm{cMpc}^3$. The simulation's specifications are listed in Table 1. The source formation starts at redshift 30 and the actual speed

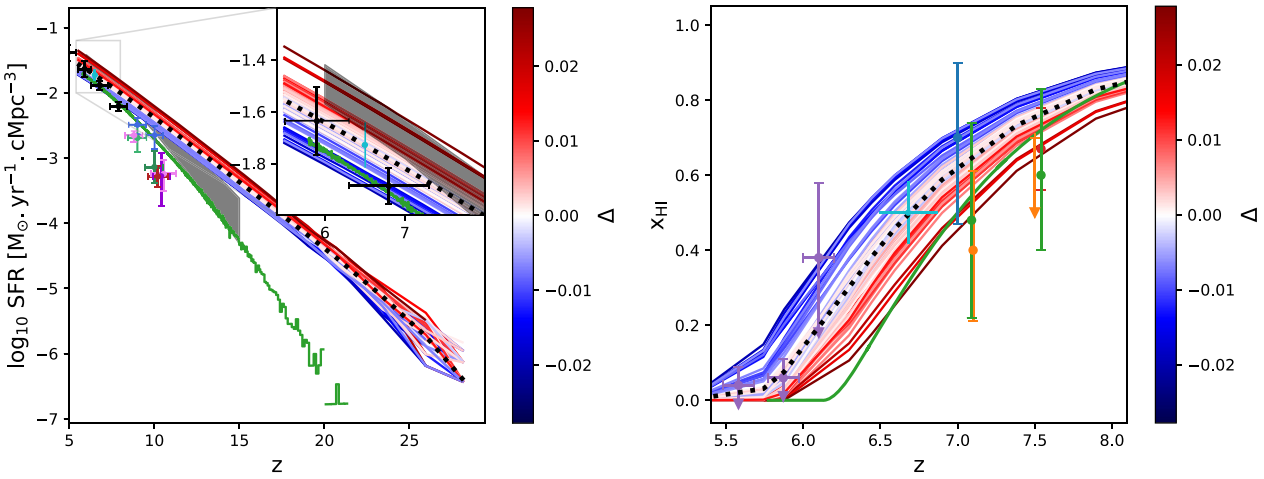

Figure 3. SFR and neutral fraction: In the left-hand panel, the evolution with redshift of the cosmic SFR, and on the right, the neutral fraction. The average of the whole volume is shown with the dotted black thick line. The blue and red lines present the cosmic SFR density and neutral fraction for subcubic volumes of 64 h^{-1} cMpc side (same volume as CoDaII), respectively, colour code for density of subregions. In both panels, the cyan error bar illustrates the 1-σ dispersion induced by the large-scale density fluctuations at redshift 6.5, and the green line is the average evolution of the CoDaII for reference. The observation points come from different probes. In the left-hand panel, the constraints on the cSFRD are: Bouwens et al. (2014, 2016) in black and violet, McLeod, McLure & Dunlop (2016) in blue, Oesch et al. (2013, 2014, 2018) in green and brown, Ishigaki et al. (2018) in pink, and the grey shaded area is Gillet, Mesinger & Park (2020). In the right-hand panel, the constraints on the neutral fraction are: McGreer, Mesinger & D'Odorico (2015) are in purple, Greig & Mesinger (2017) and Greig, Mesinger & Bañados (2019) are in orange, Davies et al. (2018) are in green, Bañados et al. (2018) is in red, and Wang et al. (2020) is in dark blue.

of light is used for the radiative transfer, to avoid artefacts as reported in Deparis et al. (2019) and Ocvirk et al. (2019). In this first simulation, X-rays are included but have a negligible impact on the gas temperature. Likewise, a self-consistent treatment of Ly α (usually performed in post-processing) is not yet included. The study of the effects of X-ray and Ly α will be done in the follow-up study (Gillet et al., in preparation).

2.2.2 *Results*

The cosmic SFR is calibrated to be roughly on or above of the observations at redshift 6 [3×10^{-2}($M_\odot$ yr^{-1} $cMpc^{-3}$)] and [10^{-6}($M_\odot$ yr^{-1} $cMpc^{-3}$))] at redshift 30. It accounts for the fact that the cosmic SFR predicted by the simulation contains the contribution of all the galaxies, while the observations are limited to a magnitude of -17. Fig. 3 presents the cosmic SFR in the left-hand panel and the neutral fraction in the right. The grey area presents the estimated total SFR (Gillet et al. 2020). In the simulation, the evolution of the cosmic SFR with redshift is induced by the evolution of the density distribution and the evolution of the parameter Δ_{Σ_*}. The ionization history is calibrated to have a mid-reionization between redshifts 6 and 7. The CoDaII averages are also shown in green for comparison. Even with its mass/spatial resolution, the CoDaII is not able to form stars at the early redshift ($z = 30$). Here, the new parametrization can form stars at the CD, while encompassing the subgrid feedback on SFR at later redshift.

Additionally, Fig. 3 presents the dispersion of the SFR and neutral fraction for subcubic volumes of $64h^{-1}$ cMpc side that can be compared to the volume of the CoDaII simulation that was used to calibrate the SF model. The overdensity of each subvolume is indicated in red and blue for the overdense and underdense regions, respectively. The dispersion in SFR is relatively constant between $z = 6$ and 30 and is comparable to the observations uncertainties (illustrated at redshift 6.5 with the cyan error bars). In the case of the neutral fraction, the dispersion at mid-reionization is slightly smaller than current observations estimation with ± 0.08 and the redshift dispersion is about ± 0.19 around the average mid-ionization redshift (illustrated with the cyan error bars). Overall, these results demonstrate that our new SF model can be made consistent with constraints during the EoR while providing a sustained SF during the CD.

3 21 CM SIGNAL

Additionally to the new SF prescription, we added in the code the computation of the 21 cm signal. The goal is to predict the possible 21 cm signal that could observe radio telescopes from the CD to the end of the reionization. For those kinds of observations, high resolution is not needed ($1\,h^{-1}$ cMpc of resolution is enough). However, a large volume is required to probe the largest mode that will be observed. The following results are presented for the largest box available in this study: $512h^{-1}$ cMpc .

3.1 Simulation of the signal

The formula of the 21 cm brightness temperature with respect to the CMB at a given redshift and point in space is given by

$$\delta T_{21} \approx 27(1 - x_{\rm HII})(1 + \delta)\left(1 - \frac{T_{\rm CMB}(z)}{T_{\rm s}}\right) C_{\rm cosmo}\ [\rm mK] \qquad (5)$$

$$C_{\rm cosmo} = \left(\frac{\Omega_{\rm b}}{0.044}\right)\left(\frac{h}{0.7}\right)\sqrt{\frac{0.27}{\Omega_{\rm m}}}\sqrt{\frac{1+z}{10}},$$

where $x_{\rm H\,II}$ is the ionized fraction of the gas, δ is its overdensity, $T_{\rm CMB}$ is the temperature of the CMB, and $T_{\rm s}$ is the spin temperature. We neglect the velocity gradient in this study. The spin temperature of the gas can be computed from

$$T_{\rm s} = \frac{1 + x_{\rm c} + x_\alpha}{T_{\rm CMB}^{-1} + x_{\rm c} T_{\rm K}^{-1} + x_\alpha T_{\rm c}^{-1}} \quad , \qquad (6)$$

where T_K is the kinetic temperature of the gas, T_c is the colour temperature of the radiation field at the Ly α transition, x_c is the collision coupling coefficient, and x_α is the coupling coefficient associated with Ly α pumping.

In this study, we do not include the Ly α radiative transfer therefore in the following we will consider two regimes. At first, we consider a uniform Ly α coupling factor rising with redshift due to a rising Ly A background: $[\log_{10}(x_\alpha) = -3/8\, z + 7.25]$ that mimics the average evolution from fig. 2 of Ross et al. (2019). By doing so, we can produce realistic global temperature evolution, but the PS cannot take into account the spatial fluctuations of x_α. We also consider the saturated regime, where we assume $x_\alpha \gg 1 + x_c$ everywhere and $T_s = T_c = T_K$.

Finally, the collision coupling coefficient accounts for the H–H, H–e$^-$, and H–H$^+$ collisions and is given by

$$x_c = \frac{T_{10}}{A_{10}} \frac{1}{T_{CMB}(z)} (n_{HI}\kappa_{HH} + n_p\kappa_{pH} + n_e\kappa_{eH}), \quad (7)$$

where κ_i are the spin de-excitation rates for each type of collisions, n_i are the densities, $T_{10} = 0.068$(K), and $A_{10} = 2.85 \times 10^{-15}$(s^{-1}) is the spontaneous emission rate. The de-excitation rates are taken into account as follows:

(i) κ_{HH} is interpolated from Zygelman (2005) table 2 column 4 for 1 K $\leq T_K \leq$ 300 K or $\kappa_{HH} = 3.1 \times 10^{-11} T_K^{0.357} e^{-32/T_K}$ (cm^3 s^{-1}) for 300 K $\leq T_K$ (Kuhlen, Madau & Montgomery 2006).

(ii) κ_{eH} is interpolated from Furlanetto & Furlanetto (2007a) table 1 for 1 K $\leq T_K \leq$ 10 000 K or $\log_{10}(\kappa_{eH}) \approx -8.0958$ for 10 000 K $\leq T_K$ (Liszt 2001).

(iii) κ_{pH} is interpolated from Furlanetto & Furlanetto (2007b) table 1 for 1 K $\leq T_K \leq$ 20 000 K or $\kappa_{pH} = 2\kappa_{HH}$ for 20 000 K $\leq T_K$.

The 21 cm signal is computed on the fly by the EMMA simulation code for the two Ly α regimes (saturated and average background). The PS of the simulated temperature brightness fields is computed using TOOLS21 CM (Giri, Mellema & Jensen 2020) in post-processing. The spherically average dimensionless PS $[\Delta^2(k)]$ is computed using

$$\Delta^2(k) = \frac{k^3}{2\pi^2} \langle P(\mathbf{k}) \rangle_{(k_x,k_y,k_z)}, \quad (8)$$

where $P(\boldsymbol{k})$ is the PS and k_i are the components of the wave vector along the simulation volume.

3.2 Observation of the signal

Having an 'ideal' noiseless 21 cm PS from the CD, we used PS_EOR [1] to take into account the UV coverage and the noise level due to the instrument. We focus on the NENUFAR (Zarka et al. 2012) observations as we are part of the NENUFAR CD key project. NENUFAR is a radio interferometer that will observe between 85 and 30 MHz, covering the CD epoch. Interestingly, it covers the 83–73 MHz band where the EDGES collaboration reported a signal detection (Bowman et al. 2018).

Radio interferometers may produce a three-dimensional data cube, two dimensional on the sky and the third dimension corresponding to the frequency that can be converted into distance/redshift/time assuming a cosmological model. To get as close as possible to the observations, we have to construct a data cube corresponding to the same coverage on the sky and depth in frequency. The observation specifications are listed in Table 2 and correspond to the ongoing CD observation program made with NENUFAR. We focus on the highest frequency band, centred on redshift 17 (corresponding to the EDGES claimed detection band). The shape of the observed volume is 2982.29 cMpc on the sky direction and 231.54 cMpc in depth. The volume is divided into 68^2 pixels on the sky and 51 along the line of sight. As the depth of the data cube is relatively small (231.54 cMpc), we neglect for the moment the increase of the size with the depth, as well as the time evolution along with the frequency (light-cone effects; Greig & Mesinger 2018): The simulation size (756 cMpc) is larger than the observational depth; a third of box is enough in depth. Conversely, on the sky's axes, the box is repeated ∼4 times. It should be noted that the observed modes are overwhelmingly due to $k_\parallel$, which correspond to the line of sight. The $k_\parallel$ modes are roughly one order of magnitude greater than $k_\perp$. Therefore, the result is not affected by the periodic repetition of the box. Once the mock data cube is filled by the simulation it is given to PS_EOR to compute the PS and the theoretical thermal noise level.

Table 2. Observation specifications: first, the frequency, secondly the sky, and finally the observation time information. The transformations to comoving distance are made at the central redshift; the data cube is considered as 'cubic'.

NENUFAR observations specs	
Band width	9.96 (MHz)
Channel width	195.3 (kHz)
Number of channels	51
Redshift at centre	17
Frequency at centre	78.91 (MHz)
Bandwidth limits	83.79–73.83 (MHz)
Bandwidth limit redshift	15.95–18.24
Depth	231.54 (cMpc)
Depth resolution	4.54 (cMpc)
Field of view	16 (°)
Field of view at centre	2982.29 (cMpc)
Number of pixels across the sky	68^2
Sky resolution	43.857 (cMpc)
Total obs time	1000 (h)
Time obs per day	8 (h)
Integration time	100 (s)

3.3 Simulated observations of the 21 cm

After the calibration of the SFR and ionization history (cf. 2.2.2), we analyse the 21 cm signal. Fig. 4 presents different quantities related to the 21 cm signal. In the left-hand panel, the global average brightness temperature is shown in red. The background colour shows the distribution of the brightness temperature with redshift (volume weighted). We note that the brightness temperature is bi-modal between redshifts 21 and 8, with a cold phase and a hot phase.We note that at high redshift, some cells are in emission, being heated and not fully ionized by close sources. This effect is discussed in Ross et al. (2019), who propose a subgrid two-phase medium model to solve this problem. We postpone the inclusion of this subgrid model to future studies. The middle panel presents the PS (for coeval cubes, i.e. not taking into account light-cone effects) at different redshifts and the right-hand panel presents the evolution of some specific $\Delta^2(k)$ with redshift. The PS presented here is qualitatively similar to simulated expectations (see Greig & Mesinger 2017; Ross et al. 2019; Reis, Fialkov & Barkana 2020, for examples). Note that above redshift 15 the PS is affected by the missing Ly α transfer. The uniform Ly α background reduces uniformly the power at every

[1] https://gitlab.com/flomertens/ps_eor

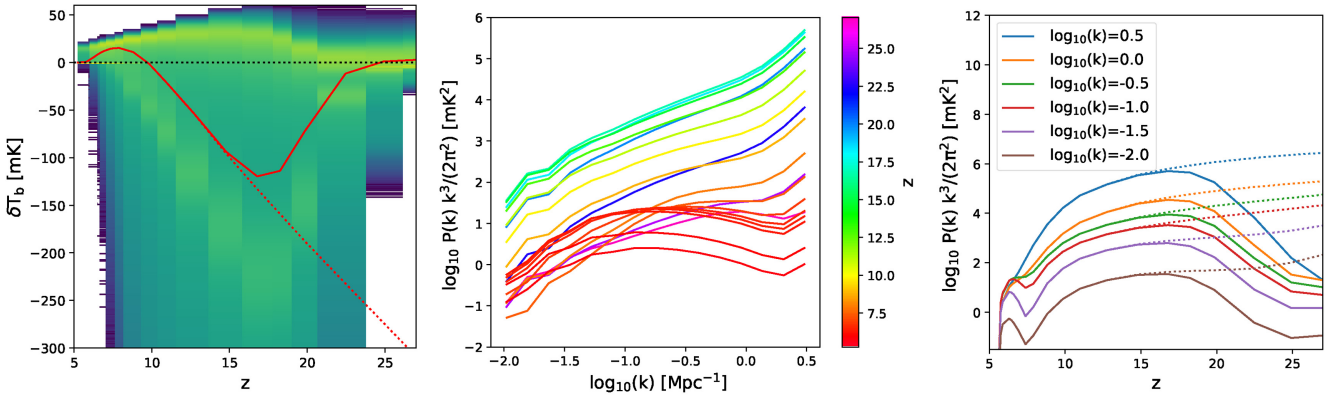

Figure 4. The 21 cm signal: In the left-hand panel, the distribution of the 21 cm temperature brightness function of redshift. The red full line presents the average evolution of the brightness temperature taking into account the uniform Ly α background, while the red dotted line assumes a fully coupled gas and spin temperature. The background colour code is for the volume-weighted distribution (the $\log_{10}$ of the cells count). In the middle panel, the PS at all scales at different redshifts (the colour code redshifts between 6 and 30) in the case of the Ly α uniform background. The right-hand panel presents the power evolution with redshift of some specific scales ($\log_{10} k = 0.5, 0, -0.5, -1$, and -1.5), for the Ly α uniform background and fully coupled approximation in full and dotted lines, respectively.

scale above redshift 15, illustrated in the right-hand panel of Fig. 4 with the full lines and the dotted lines illustrate a full Ly α coupling at all times. While the propagation of the Ly α photons through the IGM should induce spatial patterns and so different power evolution with redshift.

Finally, the main goal is to produce a 21 cm PS as close as possible to the future observed one. We process this cube through PS_EOR to take into account the UV coverage (see Section 3.2). In theory, the PS outputted by PS_EOR should be the same as the one obtained on the 'perfect' simulated cubes, in the range of scale well sampled, and in the absence of further distortion. In this study, we do not include other sources of noise subtraction or distortion on the signal, like wedge treatment or foreground residuals. The wedge is a portion of the Fourier space where the foreground signal due to the galaxy is dominant. There are two main strategies to extract the 21 cm cosmic signal. The first, the wedge avoidance, consists of cutting off the data where the Galactic foreground is too dominant. The resulting PS estimation should be foreground free, but some piece of the signal is lost as some data have been deleted. The second, the foreground modelling, consists of trying to keep all the data by modelling the foreground and substrate it. It has the advantage to conserve more data therefore more signals, but at the cost of some modelling dependences and foreground residuals that are difficult to quantify.

In Fig. 5, we present in blue the PS at redshift 17 and the 2σ theoretical error due to the thermal noise (dashed blue line). We also present the predicted PS at redshift 9 and the error for the LOFAR. In both cases, a detection is expected for wavenumber below $k = 0.1\,\mathrm{h\,cMpc^{-1}}$. The most recent upper limit at redshift 9 of $\log_{10}(\Delta^2) = 3.73$ at $k = 0.075\,\mathrm{h\,cMpc^{-1}}$ (Mertens et al. 2020) is 2 dex above our prediction and at redshift of 17 $\log_{10}(\Delta^2) = 782$ at $k = 0.15\,\mathrm{h\,cMpc^{-1}}$ (Gehlot et al. 2020) is 4 dex above our prediction (not added in Fig. 5).

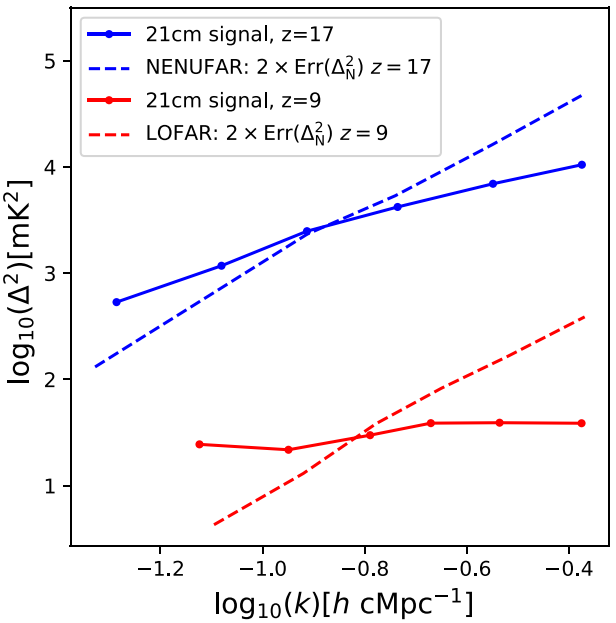

Figure 5. The 21 cm PS: The dotted full lines present the PS given by PS_EOR, which takes into account the resolution and the UV coverage of the instruments, NENUFAR and LOFAR, respectively, in blue and red at redshifts 17 and 9. The dashed lines present the expected 2σ sensitivity for 1000 h of observations.

4 CONCLUSIONS

In this paper, we introduce a new large-scale galaxy formation model in the fully coupled dark matter, hydrodynamics, radiative transfer code EMMA. This empirical model allows the efficient production of large-scale low-resolution simulations of the CD and EoR with a reduced and flexible set of parameters, based on the results of the state-of-the-art simulation of the reionization CoDaII. We ran a simulation using this model and predict the associated 21 cm signal. We process it up to the prediction of the PS with tools as close as possible to the one used to reduce the observational data. The resulting PS obtained on $512\,h^{-1}\,\mathrm{cMpc}^3/512^3$ elements of resolution fiducial simulation is qualitatively comparable to the state-of-the-art predictions.

We focused on the ongoing observations of the radio telescope NENUFAR, which is covering the CD. We predict that our fiducial model should be detected by NENUFAR at redshift 17 at a wavenum-

ber between $k = 0.1$ and $0.06\,\mathrm{h\,cMpc^{-1}}$ with 1000 h of observations. LOFAR should detect the signal at the same wavenumber at redshift 9.

While waiting for the data acquisition, reduction, and analysis, we plan to explore the parameter space. Specifically, the next step is to quantify how much a signal detection at $k = 0.1$ and $0.06\,\mathrm{h\,cMpc^{-1}}$ at redshift 17 may constrain our parameters, e.g. the SFR spatial distribution. A large number of points still have to be addressed, such as the inclusion of Ly α photons, which is essential for the computation of the 21 cm signal, or a subgrid treatment of the temperature to take into account the subcell multiphase of the gas (Ross et al. 2019).

ACKNOWLEDGEMENTS

We thank Anastasia Fialkov for fruitful discussions and sharing data to help the validation of the model. We thank the CoDa Collaboration for sharing the data of the CoDaII simulation. The CoDa II simulation was performed at Oak Ridge National Laboratory/Oak Ridge Leadership Computing Facility on the Titan supercomputer (INCITE 2016 award AST031). Processing was performed on the Eos and Rhea clusters. NG is supported by the University of Strasbourg IDEX post-doctoral grant 'Predicting with cosmological simulations the 21-cm signal from the Epoch of Reionization for future large radio observatories'. This work was granted access to the HPC resources of CINES under the allocations 2020-A0070411049 and 2021- A0090411049 'Simulation des signaux et processus de l'aube cosmique et Réionisation de l'Univers' made by GENCI. This research made use of ASTROPY, a community-developed core PYTHON package for astronomy (Astropy Collaboration 2018); MATPLOTLIB, a PYTHON library for publication-quality graphics (Hunter 2007); SCIPY, a PYTHON-based ecosystem of open-source software for mathematics, science, and engineering (Virtanen et al. 2020); NUMPY (Harris et al. 2020); and IPYTHON (Perez & Granger 2007).

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Astropy Collaboration et al., 2018, AJ, 156, 123
Aubert D., Deparis N., Ocvirk P., 2015, MNRAS, 454, 1012
Bañados E. et al., 2018, Nature, 553, 473
Bouwens R. J. et al., 2014, ApJ, 793, 115
Bouwens R. J. et al., 2016, ApJ, 833, 72
Bowman J. D., Rogers A. E. E., Monsalve R. A., Mozdzen T. J., Mahesh N., 2018, Nature, 555, 67
Chardin J., Puchwein E., Haehnelt M. G., 2017, MNRAS, 465, 3429
Davies F. B. et al., 2018, ApJ, 864, 142
Deparis N., Aubert D., Ocvirk P., Chardin J., Lewis J., 2019, A&A, 622, A142
Fialkov A., Barkana R., Visbal E., 2014, Nature, 506, 197
Furlanetto S. R., Furlanetto M. R., 2007a, MNRAS, 374, 547
Furlanetto S. R., Furlanetto M. R., 2007b, MNRAS, 379, 130
Gehlot B. K. et al., 2020, MNRAS, 499, 4158
Gehlot B. K., 2020, MNRAS, 499, 4158
Gillet N. J. F., Mesinger A., Park J., 2020, MNRAS, 491, 1980
Giri S., Mellema G., Jensen H., 2020, J. Open Source Softw., 5, 2363
Gnedin N. Y., 2016, ApJ, 821, 50
Greig B., Mesinger A., 2017, MNRAS, 472, 2651
Greig B., Mesinger A., 2018, MNRAS, 477, 3217
Greig B., Mesinger A., Bañados E., 2019, MNRAS, 484, 5094
Harris C. R. et al., 2020, Nature, 585, 357
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Iliev I. T., Mellema G., Ahn K., Shapiro P. R., Mao Y., Pen U.-L., 2014, MNRAS, 439, 725
Ishigaki M., Kawamata R., Ouchi M., Oguri M., Shimasaku K., Ono Y., 2018, ApJ, 854, 73
Kaur H. D., Gillet N., Mesinger A., 2020, MNRAS, 495, 2354
Kaur H. D., 2020, MNRAS, 495, 2354
Kuhlen M., Madau P., Montgomery R., 2006, ApJ, 637, L1
Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, MNRAS, 485, L24
Liszt H., 2001, A&A, 371, 698
McGreer I. D., Mesinger A., D'Odorico V., 2015, MNRAS, 447, 499
McLeod D. J., McLure R. J., Dunlop J. S., 2016, MNRAS, 459, 3812
Mertens F. G. et al., 2020, MNRAS, 493, 1662
Nasirudin A., Iliev I. T., Ahn K., 2020, MNRAS, 494, 3294
Ocvirk P. et al., 2016, MNRAS, 463, 1462
Ocvirk P. et al., 2020, MNRAS, 496, 4087
Ocvirk P., Aubert D., Chardin J., Deparis N., Lewis J., 2019, A&A, 626, A77
Ocvirk P., 2020, MNRAS, 496, 4087
Oesch P. A. et al., 2013, ApJ, 773, 75
Oesch P. A. et al., 2014, ApJ, 786, 108
Oesch P. A., Bouwens R. J., Illingworth G. D., Labbé I., Stefanon M., 2018, ApJ, 855, 105
Park J., Mesinger A., Greig B., Gillet N., 2019, MNRAS, 484, 933
Perez F., Granger B. E., 2007, Comput. Sci. Eng., 9, 21
Planck Collaboration VI, 2020, A&A, 641, A6
Reis I., Fialkov A., Barkana R., 2020, MNRAS, 499, 5993
Reis I., 2020, preprint (arXiv:2008.04914)
Ross H. E., Dixon K. L., Ghara R., Iliev I. T., Mellema G., 2019, MNRAS, 487, 1101
Semelin B., Eames E., Bolgar F., Caillat M., 2017, MNRAS, 472, 4508
van Haarlem M. P. et al., 2013, A&A, 556, A2
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Visbal E., Barkana R., Fialkov A., Tseliakhovich D., Hirata C. M., 2012, Nature, 487, 70
Wang F. et al., 2020, ApJ, 896, 23
Zarka P., Girard J. N., Tagger M., Denis L., 2012, in Boissier S., de Laverny P., Nardetto N., Samadi R., Valls-Gabaud D., Wozniak H., eds, SF2A-2012: Proc. Annu. Meeting French Soc. A&A. p. 687
Zygelman B., 2005, ApJ, 622, 1356

This paper has been typeset from a TEX/LATEX file prepared by the author.