



# High-redshift predictions from IllustrisTNG – III. Infrared luminosity functions, obscured star formation, and dust temperature of high-redshift galaxies

Xuejian Shen<sup>1</sup>,<sup>1</sup>★ Mark Vogelsberger<sup>2</sup>,<sup>2</sup> Dylan Nelson<sup>3</sup>,<sup>3</sup> Sandro Tacchella<sup>4</sup>,<sup>4</sup> Lars Hernquist,<sup>4</sup> Volker Springel<sup>5</sup>,<sup>5</sup> Federico Marinacci<sup>6</sup>,<sup>6</sup> and Paul Torrey<sup>7</sup>,<sup>7</sup>

<sup>1</sup>TAPIR, California Institute of Technology, Pasadena, CA 91125, USA

<sup>2</sup>Department of Physics, Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>3</sup>Universität Heidelberg, Zentrum für Astronomie, Institut für theoretische Astrophysik, Albert-Ueberle-Str 2, D-69120 Heidelberg, Germany

<sup>4</sup>Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA, 02138, USA

<sup>5</sup>Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str 1, D-85741 Garching, Germany

<sup>6</sup>Department of Physics & Astronomy ‘Augusto Righi’, University of Bologna, via Gobetti 93/2, I-40129 Bologna, Italy

<sup>7</sup>Department of Astronomy, University of Florida, 211 Bryant Space Sciences Center, Gainesville, FL 32611 USA

Accepted 2021 December 22. Received 2021 November 24; in original form 2021 April 26

## ABSTRACT

We post-process galaxies in the IllustrisTNG simulations with SKIRT radiative transfer calculations to make predictions for the rest-frame near-infrared (NIR) and far-infrared (FIR) properties of galaxies at  $z \geq 4$ . The rest-frame  $K$ - and  $z$ -band galaxy luminosity functions from TNG are overall consistent with observations, despite  $\sim 0.5$  dex underprediction at  $z = 4$  for  $M_K \lesssim -25$  and  $M_z \lesssim -24$ . Predictions for the *JWST* MIRI observed galaxy luminosity functions and number counts are given. Based on theoretical estimations, we show that the next-generation survey conducted by *JWST* can detect 500 (30) galaxies in F1000W in a survey area of 500 arcmin<sup>2</sup> at  $z = 6$  ( $z = 8$ ). As opposed to the consistency in the UV, optical, and NIR, we find that TNG, combined with our dust modelling choices, significantly underpredicts the abundance of most dust-obscured and thus most luminous FIR galaxies. As a result, the obscured cosmic star formation rate density (SFRD) and the SFRD contributed by optical/NIR dark objects are underpredicted. The discrepancies discovered here could provide new constraints on the sub-grid feedback models, or the dust contents, of simulations. Meanwhile, although the TNG predicted dust temperature and its relations with IR luminosity and redshift are qualitatively consistent with observations, the peak dust temperature of  $z \geq 6$  galaxies are overestimated by about 20 K. This could be related to the limited mass resolution of our simulations to fully resolve the porosity of the interstellar medium (or specifically its dust content) at these redshifts.

**Key words:** methods: numerical – galaxies: evolution – galaxies: formation – galaxies: high-redshift – infrared: galaxies.

## 1 INTRODUCTION

The  $\Lambda$ CDM model (e.g. Planck Collaboration 2016, 2020) is the standard theoretical paradigm for structure formation. In this framework, initial small density perturbations grow via gravitational instability and produce the large-scale structures, as well as the bound dark matter haloes where galaxies form. Based on this, the theory of galaxy formation (e.g. White & Rees 1978; Blumenthal et al. 1984; Cole et al. 2000) makes predictions that can be tested by observed galaxy populations. Aside from the well-studied constraints in the local Universe, galaxies formed in the early Universe provide a new testing ground for galaxy formation theories (see reviews of Shapley 2011; Stark 2016; Dayal & Ferrara 2018, and references therein), with open questions related to star formation in dense molecular clouds and stellar/supernovae feedback, the metal enrichment, and dust formation in early environments, the seeding of massive black

holes, the triggering of active galactic nucleus (AGN) activity, etc.

Due to the limited wavelength coverage of the *Hubble Space Telescope* (*HST*) and the insufficient sensitivity of infrared (IR) instruments, the observation of high-redshift galaxies was mainly performed in the rest-frame ultraviolet (UV; e.g. Bouwens et al. 2003; Wilkins et al. 2010; McLure et al. 2013; Finkelstein et al. 2015; Oesch et al. 2018; Bouwens et al. 2019). However, UV observations are inadequate for revealing the entire galaxy population. It is known that the cosmic star formation rate density (SFRD) at low redshift is dominated by dusty star-forming galaxies (DSFGs; Magnelli et al. 2011; Casey et al. 2012; Gruppioni et al. 2013) that are heavily obscured in optical and UV while bright at far-infrared (FIR) wavelengths. At high redshift, owing to instrumental limitations, the abundance of such galaxies and their contribution to cosmic star formation are still highly uncertain (e.g. Casey et al. 2018). In recent years, ALMA has been identifying some FIR-bright but UV-faint galaxies at  $z \gtrsim 3$  (e.g. Simpson et al. 2014; Wang et al. 2019; Yamaguchi et al. 2019; Franco et al. 2020; Dudzevičiūtė

\* E-mail: xshen@caltech.edu

et al. 2020) and measuring the dust continuum emission from these galaxies. These observations reveal galaxies that were hidden in previous optical/near-infrared (NIR) selections and may be the tip-of-the-iceberg of the highly obscured high-redshift galaxy population. Several future sub-millimeter/millimeter instruments have been proposed, including the TolTEC camera on the Large Millimeter Telescope (LMT; Bryan et al. 2018), the Origins Space Telescope (OST; Battersby et al. 2018) and the Chajnantor Sub-Millimeter Survey Telescope (CSST; Golwala 2018), which will help to reveal the demographics of DSFGs at high redshift.

Meanwhile, the James Webb Space Telescope (*JWST*; Gardner et al. 2006) will be in operation soon. As discussed in Vogelsberger et al. (2020, Paper I of this series), the NIRCam of *JWST* will push the detection of galaxies in the UV to the fainter end and help reveal lower mass galaxies, which have a significant contribution to the cosmic SFRD, as well as more heavily obscured DSFGs. In addition, the NIRCam and the Mid-Infrared Instrument (MIRI) will provide photometric and spectroscopic access to the rest-frame UV to mid-infrared (mid-IR) spectral energy distributions (SEDs) of galaxies at high redshift. Rest-frame optical and NIR observations would be particularly useful for an unbiased measurement of galaxy stellar mass and constraints on the star formation histories of galaxies. Mid-IR observations of dust emission lines would shed light on the physical properties of dust in high-redshift galaxies. Undoubtedly, the next-generation galaxy surveys will provide a much deeper and broader spectral coverage of galaxy emission at high redshift, particularly at IR wavelengths. The advancement will hopefully provide a more complete picture of galaxy formation in the early Universe, aside from the UV emission from the unobscured young stellar populations.

In parallel to the observational efforts, theoretical predictions are also necessary for the study of high-redshift galaxies. Several attempts have been made with semi-analytical models of galaxy formation (e.g. Clay et al. 2015; Liu et al. 2016; Cowley et al. 2018; Tacchella et al. 2018; Yung et al. 2019a,b) paired with simple prescriptions for dust extinction, either empirical dust corrections based on observational relations or simple dust shell models that link extinction with integrated optical depth. Similar prescriptions have been adopted in cosmological simulations (e.g. Cullen et al. 2017; Wilkins et al. 2017; Ma et al. 2018). However, such treatments are great simplifications of the scattering and absorption processes of dust grains and effectively neglect the complicated dust geometry. The predictive power for the dust continuum emission at IR wavelengths is also limited. Alternatively, radiative transfer calculations have been introduced for many galaxy simulations to model dust (or neutral gas) absorption and emission. For example, Cen & Kimm (2014) ran radiative transfer calculations for a sample of 198 galaxies in cosmological zoom-in simulations at  $z \sim 7$ , with predictions for IR properties of ALMA galaxies; Camps et al. (2016) and Trayford et al. (2017) performed radiative transfer calculations on the EAGLE simulation (Crain et al. 2015; Schaye et al. 2015) and made predictions for UV-to-sub-millimeter SEDs of galaxies in the local Universe; Ma et al. (2019) applied radiative transfer calculations to the FIRE-2 (Hopkins et al. 2018) simulations and studied dust extinction and emission in  $z \geq 5$  galaxies. Similar techniques have also been used to study the physical origin and variations of the IRX- $\beta$  relation and dust attenuation curves (e.g. Safarzadeh, Hayward & Ferguson 2017; Narayanan et al. 2018; Liang et al. 2020; Schulz et al. 2020). Several radiation-hydrodynamical simulations (with on-the-fly radiative transfer calculations; e.g. Rosdahl et al. 2013; Kimm & Cen 2014; Ocvirk et al. 2016; Kimm et al. 2017; Rosdahl et al. 2018) have been performed to study cosmic reionization

on large scales and the escape of ionizing photons from early galaxies.

The IllustrisTNG project is a series of large, cosmological magneto-hydrodynamical simulations of galaxy formation (Nelson et al. 2019b; Pillepich et al. 2019). The TNG simulations have been calibrated and tested by numerous low-redshift observables (e.g. Genel et al. 2018; Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Pillepich et al. 2018b; Springel et al. 2018) and offer an unprecedented stand point for the study of high-redshift galaxy populations. In Paper I, we developed a radiative transfer post-processing pipeline to calculate SEDs and images of galaxies at  $z \geq 2$  in the TNG simulations. The pipeline was calibrated based on the UV luminosity functions that are well constrained by observations. In Paper I and Shen et al. (2020, Paper II of the series), we made predictions for the rest-frame UV luminosity functions, optical emission line luminosity functions, UV continuum, and optical spectral indices. We found good agreement with observations, except for a missing population of heavily obscured, UV red galaxies in the TNG simulations. Here, we aim to adapt the pipeline for predictions for the IR properties of galaxies. The major advantages of TNG compared to other cosmological simulations are its representative box size to allow statistical predictions for galaxy populations (even at high redshift) and its reasonable mass and spatial resolution to describe the multiphase interstellar medium (ISM), star formation, and feedback processes. Unlike some simulations dedicated for high-redshift studies, TNG has been evolved to low redshift and tested against various low-redshift observables.

This paper is organized as follows: In Section 2, we briefly describe the IllustrisTNG simulation suite and state the numerical parameters of TNG50, TNG100, and TNG300 in detail. In Section 3, we describe the method we used to derive the dust attenuated broadband photometry and SEDs of galaxies in the simulations. The main results are presented in Section 4, where we make various predictions and comparisons with observations. The summary and conclusions are presented in Section 5. In Appendix A, we present additional tests of different SKIRT configurations and numeric convergence.

## 2 SIMULATION

The analysis here is based on the IllustrisTNG simulation suite (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Pillepich et al. 2018b; Springel et al. 2018), including the newest addition, TNG50 (Nelson et al. 2019b; Pillepich et al. 2019) with the highest numerical resolution in the suite. The IllustrisTNG simulation suite is the follow-up project to the Illustris simulations (Genel et al. 2014; Vogelsberger et al. 2014a,b; Nelson et al. 2015; Sijacki et al. 2015). The simulations were conducted with the moving-mesh code AREPO (Springel 2010; Pakmor et al. 2016; Weinberger, Springel & Pakmor 2020) and adopted the IllustrisTNG galaxy formation model (Weinberger et al. 2017; Pillepich et al. 2018a), which is an update of the Illustris galaxy formation model (Vogelsberger et al. 2013; Torrey et al. 2014). The IllustrisTNG simulation suite consists of three primary simulations – TNG50, TNG100, and TNG300 (Nelson et al. 2019a), covering three different periodic, uniformly sampled volumes, roughly  $\approx 50^3, 100^3, 300^3$  Mpc<sup>3</sup>. For simplicity, in the following, we will refer to TNG50-1, TNG100-1, and TNG300-1 as TNG50, TNG100, and TNG300, respectively. The simulations employ the following cosmological parameters (Planck Collaboration 2016):  $\Omega_m = 0.3089$ ,  $\Omega_b = 0.0486$ ,  $\Omega_\Lambda = 0.6911$ ,  $H_0 = 100 h$  km s<sup>-1</sup> Mpc<sup>-1</sup> = 67.74 km s<sup>-1</sup> Mpc<sup>-1</sup>,  $\sigma_8 = 0.8159$ , and  $n_s = 0.9667$ . The numerical parameters of the simulations are summarized in Table 1.

**Table 1.** IllustrisTNG simulation suite. The table shows the basic numerical parameters of the three primary IllustrisTNG simulations: simulation volume side length, number of gas cells ( $N_{\text{gas}}$ ), number of dark matter particles ( $N_{\text{dm}}$ ), baryon mass resolution ( $m_b$ ), dark matter mass resolution ( $m_{\text{dm}}$ ), Plummer-equivalent maximum physical softening length of dark matter and stellar particles ( $\epsilon_{\text{dm, stars}}$ ), and the minimal comoving cell softening length  $\epsilon_{\text{gas}}^{\text{min}}$ . In the following, we will refer to TNG50-1, TNG100-1, and TNG300-1 as TNG50, TNG100, and TNG300, respectively.

IllustrisTNG Simulation	Run	Volume side length ( $h^{-1}\text{Mpc}$ )	$N_{\text{gas}}$	$N_{\text{dm}}$	$m_b$ ( $h^{-1}M_{\odot}$ )	$m_{\text{dm}}$ ( $h^{-1}M_{\odot}$ )	$\epsilon_{\text{dm, stars}}$ ( $h^{-1}\text{kpc}$ )	$\epsilon_{\text{gas}}^{\text{min}}$ ( $h^{-1}\text{kpc}$ )
TNG300	TNG300(-1)	205	2500 <sup>3</sup>	2500 <sup>3</sup>	$7.4 \times 10^6$	$4.0 \times 10^7$	1.0	0.25
TNG100	TNG100(-1)	75	1820 <sup>3</sup>	1820 <sup>3</sup>	$9.4 \times 10^5$	$5.1 \times 10^6$	0.5	0.125
TNG50	TNG50(-1)	35	2160 <sup>3</sup>	2160 <sup>3</sup>	$5.7 \times 10^4$	$3.1 \times 10^5$	0.2	0.05

### 3 GALAXY IR LUMINOSITIES AND SEDS

In this section, we introduce the approach we adopt to calculate dust-attenuated/intrinsic IR SEDs and band luminosities/magnitudes of galaxies in the IllustrisTNG simulations. We follow the Model C procedure introduced in Paper I and adapt it for IR predictions. In this section, we will briefly review the procedure and refer the readers to Paper I for more details.

In this work, we define a galaxy as being either a central or satellite galaxy as identified by the SUBFIND algorithm (Springel et al. 2001; Dolag et al. 2009). For all the following analysis, we impose a stellar mass cut for galaxies. We only consider galaxies with a stellar mass larger than 100 times the baryonic mass resolution,  $100 \times m_b$ , within twice the stellar half mass radius. Galaxies resolved with a lower number of resolution elements will not be considered, since we assume that their structure is not reliably modelled.

The calculation of galaxy SEDs consists of two steps: (I) characterize the radiation source and determine the spatial and wavelength distribution of the intrinsic emission; (II) perform dust radiative transfer calculations, including dust absorption, dust self-absorption, and dust emission. In the first step, we assigned intrinsic emission to stellar particles in the simulations according to their ages and metallicities using the stellar population synthesis method. To be specific, we adopt the Flexible Stellar Population Synthesis (FSPS) code (Conroy, Gunn & White 2009; Conroy & Gunn 2010) to model the intrinsic SEDs of old stellar particles ( $t_{\text{age}} > 10\text{ Myr}$ ) and the MAPPINGS-III SED library (Groves et al. 2008) to model those of young stellar particles ( $t_{\text{age}} < 10\text{ Myr}$ ). The MAPPINGS-III SED library self-consistently considers the dust attenuation in the birth clouds of young stars which cannot be properly resolved in the simulations. In the second step, we perform the full Monte Carlo dust radiative transfer calculations using a modified version of the publicly available SKIRT (version 8)<sup>1</sup> code (Baes et al. 2011; Camps, Baes & Saffly 2013; Saffly, Baes & Camps 2014; Camps & Baes 2015). Modifications were made to incorporate the FSPS SED templates into SKIRT. Photon packages are randomly released based on the source distribution characterized by the positions and SEDs of stellar particles. The emitted photon packages will further interact with the dust in the ISM. To determine the distribution of dust in the ISM, we select cold, star-forming gas cells (with  $\text{SFR} > 0$  or temperature  $< 8000\text{ K}$ ) from the simulations and calculate the metal mass distribution based on the metallicities of selected cells. We assume that dust is traced by metals in the ISM and turn the metal mass distribution into the dust mass distribution with a constant, averaged dust-to-metal (DTM) ratio of all galaxies at a fixed redshift. To be specific, this DTM ratio depends on redshift as  $0.9 \times (z/2)^{-1.92}$ , which has been calibrated based on the observed UV luminosity functions at  $z = 2-10$  as introduced in Paper I.

The dust mass distribution is then mapped on to an adaptively refined octree grid for radiative transfer calculations with the refinement criterion chosen to match the spatial resolution of the simulations. Ultimately, after photons fully interact with dust in the galaxy and escape, they are collected by a mock detector  $1\text{ pMpc}^2$  away from the simulated galaxy, along the positive  $z$ -direction of the simulation coordinates. The integrated galaxy flux is then recorded, which provides us with the dust-attenuated SED of the galaxy in the rest frame. Galaxy SEDs without the resolved dust attenuation are derived in the same way with no resolved dust distribution included. We note that we use the term ‘without the resolved dust’ because the unresolved dust component in the MAPPINGS-III SED library is always present. Compared with the resolved dust attenuation, the impact of the unresolved dust attenuation on galaxy continuum emission is limited. For rest-frame broadband photometry, galaxy SEDs are convolved with the transmission curves using the SEDPY<sup>3</sup> code. For the calculation of apparent band magnitudes, the rest-frame flux is redshifted, corrected for intergalactic medium absorption (Madau 1995; Madau et al. 1996) and converted to the observed spectra. In addition to the rest-frame optical/UV bands and NIRCcam bands of *JWST* studied in Paper I, we add rest-frame NIR bands (J, H, Ks bands of the 2MASS survey Skrutskie et al. 2006) and the MIRI bands of *JWST*.

In Paper I and Paper II, we limited our predictions to rest-frame UV/optical properties of galaxies. To extend the predictions to IR wavelengths, the set-up of the radiative transfer calculations has been modified in the following aspects:

(i) The wavelength grid is adapted to have 101 points with a wider wavelength coverage. We first evenly sample 51 points from  $0.1$  to  $1000\mu\text{m}$  as the base grid. To better resolve the emission lines in mid-IR wavelengths, we create a refined grid in the mid-IR, sampling 61 points from  $2$  to  $30\mu\text{m}$ . Combining the base grid and the refined grid in the mid-IR, we get the final grid consisting of 101 points. Following Paper I, by default, the number of photon packages per wavelength grid is set to be the number of bound stellar particles  $N_{\text{star}}$  in the galaxy, with  $10^2$  ( $10^5$ ) as the minimum (maximum) number. In empirical tests, we find that this choice of photon package numbers gives converged galaxy UV-to-FIR SEDs, except for galaxies that are close to the selection criteria ( $N_{\text{star}} \sim 100$ , see the stellar mass cut we adopt above). To account for this, for calculations on TNG50, we increase the number of photon packages and the minimum number by a factor of three to achieve better convergence in low-mass, poorly resolved galaxies. Further increasing the number of photon packages will not lead to differences in galaxy SEDs. Poorly resolved galaxies in TNG100 and TNG300 can be resolution corrected based on well-resolved TNG50 galaxies, if the results show significant differences.

<sup>1</sup>[http://www.skirt.ugent.be/root/\\_landing.html](http://www.skirt.ugent.be/root/_landing.html)

<sup>2</sup>Physical Mpc.

<sup>3</sup><https://github.com/bd-j/sedpy>

Details of the convergence tests and discussion of the choice of photon package numbers are shown in Appendix A.

(ii) Non-local thermal equilibrium is considered. Small dust grains are allowed to be stochastically heated and decouple from local thermal equilibrium. In order to trace grains of different sizes separately, we switch our dust model from the Draine et al. (2007) model to the Zubko, Dwek & Arendt (2004) multigrain model, which has been adopted in Camps et al. (2016), Trayford et al. (2017), and Schulz et al. (2020). Similar to the Draine et al. (2007) model adopted in Paper I, the Zubko et al. (2004) model includes a composition of graphite, silicate, and polycyclic aromatic hydrocarbon grains. The size distributions and the relative amount of the dust grains are chosen so as to reproduce the dust properties of the Milky Way. Different from the model in Paper I [using the average dust properties of the Draine et al. (2007) dust mixture], the Zubko et al. (2004) model traces dust grains of different sizes separately. We adopt 10 bins for grain sizes for each type of dust and further increasing the number of bins to 15 does not lead to differences in photometric predictions. The impact of the dust model on galaxy SEDs is illustrated in Fig. A1 in Appendix A.

(iii) In practice, we find that switching to the Zubko et al. (2004) multigrain model (see above) leads to  $\sim 0.2$ – $0.3$  mag underprediction on the UV luminosities of galaxies. To compensate for that, we decrease the DTM ratio in the new runs by 25 per cent compared to the calibrated model in Paper I. After this modification, the UV magnitudes of galaxies are consistent with the results of Paper I with  $\leq 0.1$  mag differences and the total absorbed luminosities integrated from UV to optical are consistent with Paper I with  $\leq 0.1$  dex differences.

(iv) Dust self-absorption and re-emission are included. The dust emission procedure is carried out iteratively until the total luminosity absorbed by dust converges at a  $< 3$  per cent level. This mainly affects the peak of the FIR continuum and has little impact in the UV, optical, and NIR, as shown in Fig. A1 in Appendix A.

(v) The inclusion of dust self-absorption significantly increases the computational cost of the radiative transfer calculations. Therefore, for this work, we limit the radiative transfer calculations to three selected snapshots corresponding to  $z = 4, 6, 8$ , rather than the redshift range  $z = 2$ – $10$  covered by Paper I, Paper II. For  $z = 4$ , we perform the calculations for all three simulations. For  $z = 6$  and  $z = 8$ , we perform the calculations only for TNG100 and TNG300, the dynamical ranges of which are sufficient to match IR observations of galaxies.

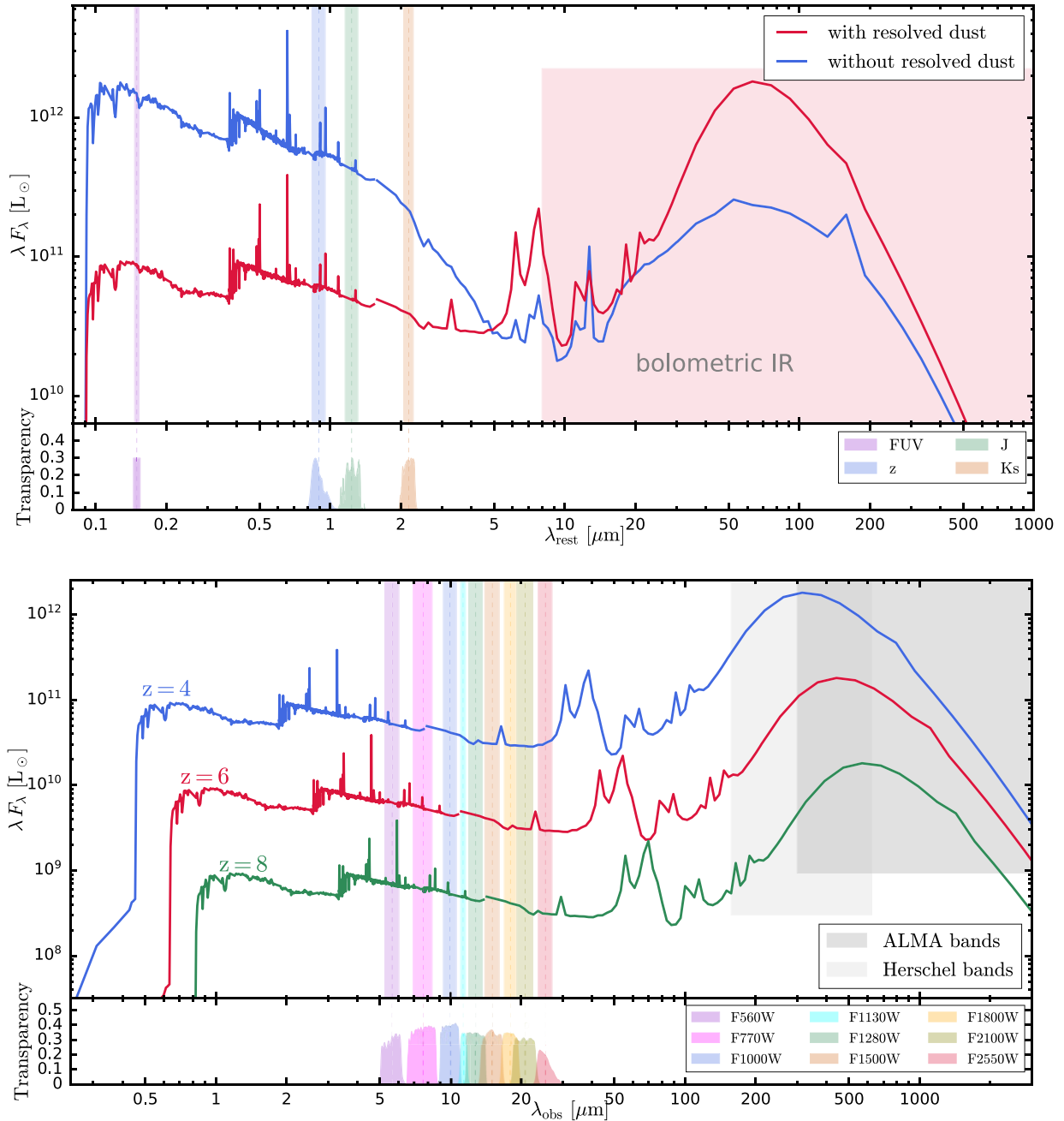
In the top panel of Fig. 1, we show integrated SEDs from the radiative transfer post-processing of a star-forming galaxy in TNG100. In the top panel, we show the rest-frame SEDs with/without resolved dust attenuation. The ‘resolved’ dust refers to the dust resolved by the simulations and involved in radiative transfer calculations, and the ‘unresolved’ dust refers to the one associated with the MAPPINGS-III SED model. It is encouraging that the UV to optical SED derived in Paper I (with better wavelength resolution and no dust emission included) can be smoothly connected to (without renormalization) the IR SED derived in this work at  $\lambda \sim 1.7 \mu\text{m}$ . In the figure, the rest-frame bands involved in this work are shown for comparison. The rest-frame FUV band ( $1500 \text{ \AA}$ ) is sensitive to the young stellar populations and thus the on-going star formation in the galaxy. The rest-frame NIR bands ( $z$ ,  $J$ ,  $K_s$ ) are sensitive to the emission of old stellar populations and thus the integrated star formation in the past. Notably, the luminosities in these bands are less affected by dust attenuation and serve as good indicators for galaxy stellar mass. The bolometric IR luminosity is defined as the integrated

flux in the wavelength range  $8$ – $1000 \mu\text{m}$ , which is dominated by dust continuum emission and indicates the total amount of energy absorbed by dust at short wavelengths. In the bottom panel, we show the apparent SEDs (all with resolved dust attenuation) in observer’s frame assuming that the galaxy is at  $z = 4, 6, 8$ . For reference, we show the *JWST* MIRI bands and the wavelength coverage of Herschel and ALMA bands at FIR. The comparison demonstrates MIRI’s promise for measuring the rest-frame optical and NIR emission of galaxies at high redshift. Paired with Herschel/ALMA observation at longer wavelengths and NIRCам observation in rest-frame UV, a fairly complete coverage of galaxy rest-frame SED can be achieved.

## 4 RESULTS

### 4.1 NIR band luminosity functions

Rest-frame NIR luminosities of galaxies are sensitive to the old stellar populations and are less affected by dust absorption or emission, on-going star formation or the choice of population synthesis model. Therefore, the NIR luminosity function is an ideal indicator for the stellar mass assembly history of galaxies in the Universe. The rest-frame  $K$ -band (including the  $K_s$  variant) centred around  $2.2 \mu\text{m}$  has been widely used for such studies at high redshift (e.g. Drory et al. 2003; Pozzetti et al. 2003; Caputi et al. 2006; Saracco et al. 2006; Cirasuolo et al. 2010; Mortlock et al. 2017). In Fig. 2, we present the galaxy rest-frame  $K$ -band luminosity functions at  $z = 4, 6, 8$  predicted from the IllustrisTNG simulations and compare them with observations (Cirasuolo et al. 2010; Mortlock et al. 2017). The predictions from TNG at different redshifts are shown in solid lines with different colours. Luminosity functions from TNG100 and TNG300 are combined here following the procedure described in Paper I. We note that resolution corrections are not applied here, since we find that the rest-frame NIR band luminosity functions from TNG50, TNG100, and TNG300 agree well in their shared dynamical ranges at the redshifts considered. Similar agreement is found for the MIRI apparent band luminosity functions. Binned estimations from observations are shown with open markers. The best-fitting Schechter functions in these observational studies are shown in dashed lines (truncated at the observational limit). One set of measurements in Mortlock et al. (2017) was originally performed at  $z \simeq 3.25$ . To compare it with our predictions at  $z = 4$ , we correct the binned estimations for the decrease in the number density normalization by linearly extrapolating their best-fitting  $\phi^*$  in single Schechter fits at  $z \leq 3.25$  to  $z = 4$ . The Schechter fit at  $z = 4$  from Mortlock et al. (2017) is also obtained by linearly extrapolating their best-fitting  $\phi^*$  to  $z = 4$  while keeping  $M^*$  and  $\alpha$  the same as those at  $z = 3.25$ . The galaxy  $K$ -band luminosity function at  $z \gtrsim 2$  can be well characterized by a single Schechter function (Mortlock et al. 2017). Previous measurements of the  $K$ -band luminosity function at  $z \gtrsim 1$  (Caputi et al. 2006; Saracco et al. 2006; Cirasuolo et al. 2010) suggested a shallow faint-end slope of  $\alpha \simeq -1$ , independent of redshift. However, updated measurements (Mortlock et al. 2017), which probed much fainter luminosities than previous studies, revealed a steep faint-end slope of  $-2 \lesssim \alpha \lesssim -1.5$  at  $z \gtrsim 2$  and argued that previous studies were incomplete at faint luminosities. Overall, the TNG predictions agree well with the observational constraints and the faint-end slopes are as steep as suggested by Mortlock et al. (2017). However, compared to the observational results, the simulation prediction at  $z = 4$  exhibits a ‘bump’ at  $M_K \sim -22.5$  mag. Although being consistent with Cirasuolo et al. (2010) results, the simulation prediction at  $z = 4$  falls below the more recent Mortlock et al. (2017) measurements at  $M_K \sim -25$  mag by about  $0.5$  dex. In terms of the redshift evolution

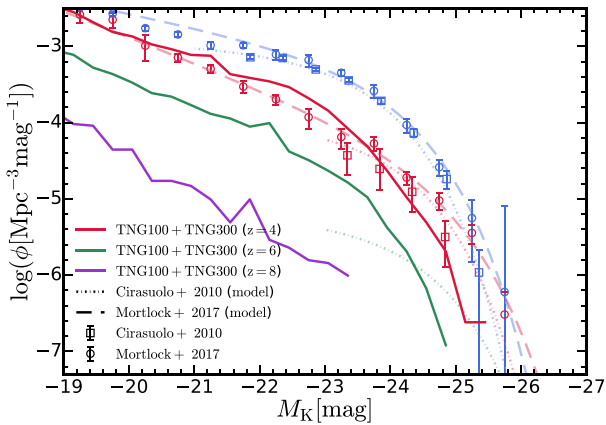


**Figure 1.** Galaxy SEDs in rest-frame or observer's frame. The SEDs are derived from the radiative transfer post-processing of a star-forming galaxy in TNG100. The UV to optical SED derived in Paper I can be smoothly connected to (without renormalization) the IR SED derived in this work at  $\lambda \sim 1.7 \mu\text{m}$ . In the top panel, we show the rest-frame SEDs with/without resolved dust attenuation and emission, along with the rest-frame bands involved in this work. The dashed lines show pivot wavelengths and the shaded regions indicate effective bandwidths. The bolometric IR luminosity is defined as the integrated flux in  $8\text{--}1000 \mu\text{m}$  as shown by the red-shaded region. The lower sub-panel shows the transmission curves of the bands. In the bottom panel, we show the SEDs (all with resolved dust attenuation and emission) in the observer's frame assuming that the galaxy is located at  $z = 2, 4, 6$ . The *JWST* MIRI broad bands and the wavelength coverage of ALMA and Herschel bands are shown for comparison. The transmission curves of the MIRI bands are also shown in the lower sub-panel.

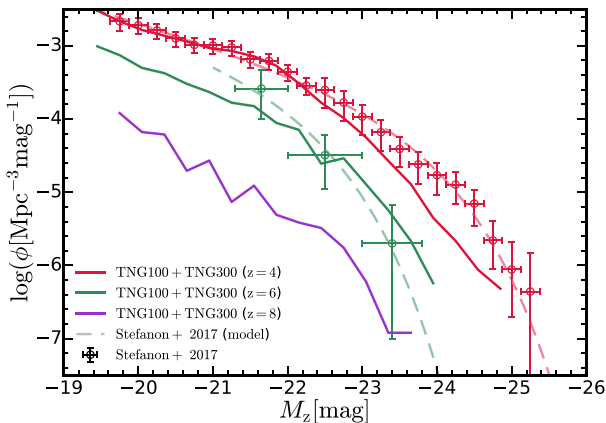
of the *K*-band luminosity function at  $z \gtrsim 2$ , TNG predicts that the number density normalization continuously decreases towards higher redshift while the faint-end slope remains the same. Approaching  $z \sim 2$ , the evolution of the bright end luminosity function stalls, which is likely related to the quenching of star formation in massive galaxies.

The rest-frame *z*-band centred around  $0.9 \mu\text{m}$  is also used in studies of NIR luminosity functions and measurements of galaxy

stellar mass functions (e.g. Stefanon et al. 2017). It is located at shorter wavelengths than the *K*-band. It can thus be probed to higher redshift with current instruments and it remains free from contamination of nebular emission lines and the Balmer break. In Fig. 3, we present the galaxy *z*-band luminosity function at  $z = 4, 6, 8$  from the IllustrisTNG simulations and compare them with observations (Stefanon et al. 2017). Both the binned estimations



**Figure 2.** Galaxy  $K$ -band luminosity function. The galaxy rest-frame  $K$ -band luminosity functions at  $z = 4, 6, 8$  from the IllustrisTNG simulations are presented in solid lines as labelled. Binned estimations from observations (Cirasuolo et al. 2010; Mortlock et al. 2017) are shown with open markers. The best-fitting Schechter functions in these work are shown in dashed and dotted lines. Note that the  $z = 6$  result from Cirasuolo et al. (2010) is based on a model-dependent extrapolation of the observations at  $z \lesssim 4$ . Observational constraints at  $z = 2, 4, 6$  are shown in blue, red, and green colours, respectively. The predicted  $K$ -band luminosity functions are consistent with observations, especially the steep faint-end slope as revealed by most recent observations.



**Figure 3.** Galaxy  $z$ -band luminosity function. The galaxy rest-frame  $z$ -band luminosity functions at  $z = 4, 6, 8$  from the IllustrisTNG simulations are presented in solid lines as labelled. Binned estimations and Schechter fits from observations (Stefanon et al. 2017) are shown with open markers and dashed lines. Observational constraints at  $z = 4, 6$  are shown in red and green colours, respectively. The predicted  $z$ -band luminosity functions are consistent with observations. The evolutionary pattern is similar to that of the  $K$ -band luminosity function.

and the Schechter fits from the observations are presented. The predicted luminosity functions are overall consistent with observations, despite that the predicted number density at  $M_z \sim -24$  at  $z = 4$  is about 0.5 dex lower than observations. Such discrepancy is not found at  $z = 6$  or in the  $K$ -band luminosity functions. The redshift evolution of the  $z$ -band luminosity function is similar to that of the  $K$ -band luminosity function presented above. At  $z \geq 4$ , the number density normalization continuously decreases towards higher redshift while the faint-end slope remains the same. Since passive galaxies have very limited contribution to the NIR luminosity

**Table 2.** JWST MIRI wide filter characteristics and detection limits. The table contains the pivot wavelengths, bandwidths, and detection limits of the JWST MIRI wide filters involved in this work. The detection limits are calculated assuming  $10^4$ s and  $10^5$ s exposure time with target signal-to-noise ratios of 10 and 5 based on the JWST ETC. For a given exposure time, the detection limits with target signal-to-noise ratios of 5–10 can simply be obtained through a constant shift of  $\sim 0.77$  mag.

Filter	Wavelength ( $\mu\text{m}$ )	Bandwidth ( $\mu\text{m}$ )	$m_{\text{SNR}=10, T_{\text{exp}}=10^4\text{s}}^{\text{lim}}$ (mag)	$m_{\text{SNR}=5, T_{\text{exp}}=10^5\text{s}}^{\text{lim}}$ (mag)
F560W	5.6	1.2	26.38	28.33
F770W	7.7	2.2	25.68	27.61
F1000W	10.0	2.0	24.75	26.69
F1130W	11.3	0.7	23.74	25.68
F1280W	12.8	2.4	23.96	25.88
F1500W	15.0	3.0	23.36	25.35 <sup>a</sup>
F1800W	18.0	3.0	22.33 <sup>a</sup>	24.12 <sup>a</sup>
F2100W	21.0	5.0	<22 <sup>b</sup>	<24 <sup>b</sup>
F2500W	25.0	4.0	<22 <sup>b</sup>	<24 <sup>b</sup>

<sup>a</sup>The number of exposures per specification is doubled while the number of groups per integration is halved, to avoid saturation of background exposure (see text for details).

<sup>b</sup>The saturation of background exposure cannot be avoided by tuning the observational strategy, so rough upper limits are given.

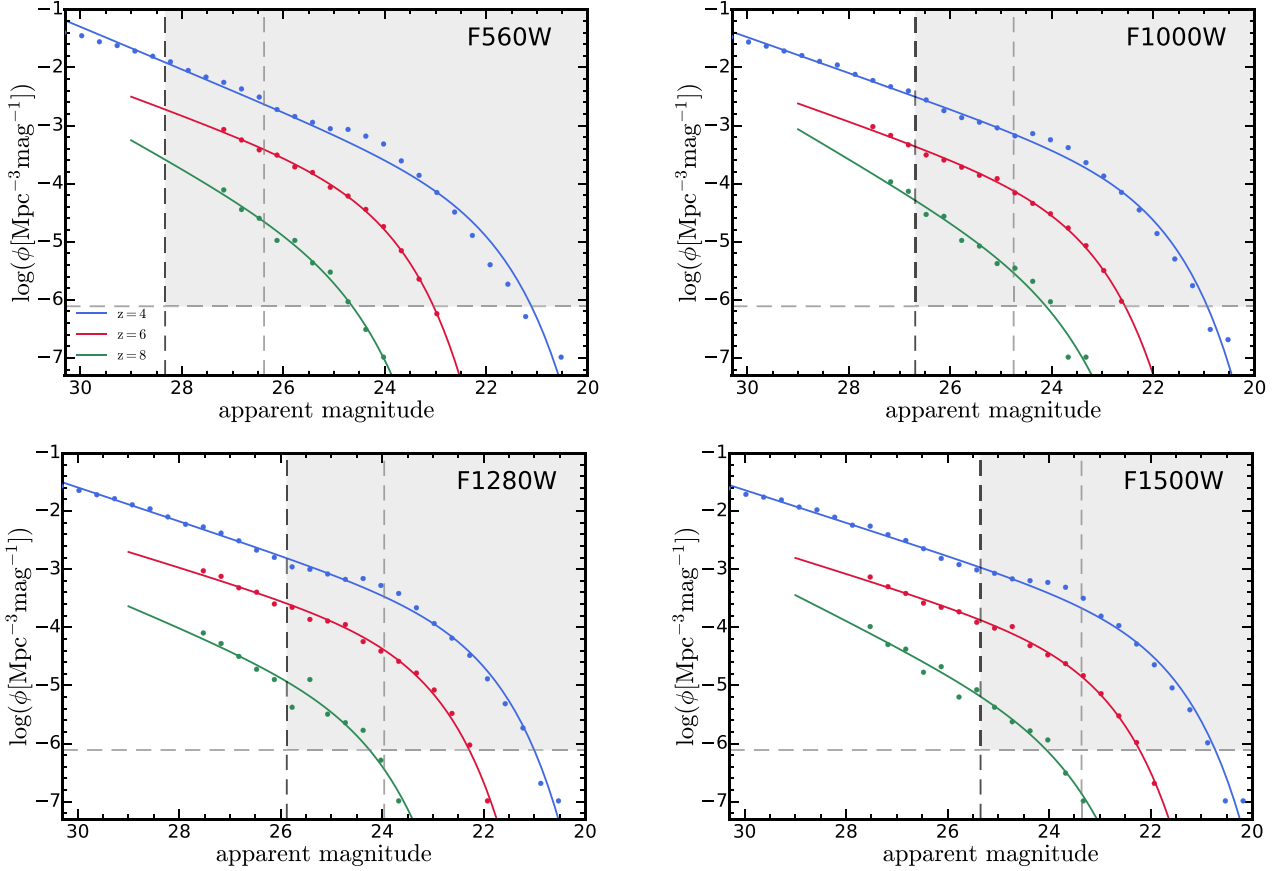
functions (or stellar mass functions) of galaxies at high redshift (e.g. Mortlock et al. 2017), the evolutionary pattern of the NIR luminosity functions reflects the mass assembly of star-forming galaxies at high redshift. The weak evolution of the faint-end slope indicates that the logarithmic increase of galaxy luminosities with cosmic time is roughly uniform across luminosities, and the mass growth rates (mass doubling times) of high-redshift galaxies are roughly independent of their masses.

## 4.2 JWST MIRI bands apparent luminosity functions

The JWST MIRI provides imaging and spectroscopic observing modes from 4.9 to 28.8  $\mu\text{m}$  (Rieke et al. 2015; Wright et al. 2015), which could be used to identify high-redshift galaxies. In this section, we will make predictions for the apparent band luminosity functions of MIRI and numbers of galaxies expected in a survey volume.

Basic information of the MIRI broad bands are listed in Table 2, along with the detection limit of each band. The detection limits are calculated using the JWST Exposure Time Calculator (ETC)<sup>4</sup> with the following configuration details. Sources are treated as point sources, and the exposure time is set to either  $10^4$ s or  $10^5$ s as indicated in the table. Furthermore, the readout pattern is set to SLOW, which yields a high signal-to-noise ratio and can efficiently reach a maximum survey depth. For the  $10^4$ s exposure time, we employ the full sub-array with 18 groups per integration, 1 integration per exposure, and 24 exposures per specification. For the  $10^5$ s exposure time, we employ 42 groups per integration, 1 integration per exposure, and 100 exposures per specification. For some long wavelength bands, to avoid saturation of background exposure, we double the number of exposures per specification while halving the groups per integration. The aperture used for imaging is circular with radius 0.1 arcsec. The background subtraction is performed with a sky annulus with inner and outer radii of 1.2 and 1.98 arcsec<sup>2</sup>. The ETC background model includes celestial sources (zodiacal light, ISM, and cosmic IR background), along with telescope thermal and scattered light. This

<sup>4</sup><https://jwst.etc.stsci.edu/>



**Figure 4.** *JWST* MIRI band apparent luminosity functions. Binned estimations from simulations are shown in solid circles and the best-fitting Schechter functions are shown in solid lines. The horizontal dashed line indicates the number density corresponding to one galaxy in a survey area of  $500 \text{ arcmin}^2$  with survey depth  $\Delta z = 1$  centred around  $z = 6$ . Here, we have assumed that all the galaxies above the detection limit can be detected and selected with 100 per cent completeness. This reference line shifts only slightly changing the survey centre to  $z = 4$  or  $z = 8$ . The vertical dashed lines indicate the detection limits we calculated assuming  $\text{SNR} = 10$ ,  $T_{\text{exp}} = 10^4 \text{ s}$  for the grey line and  $\text{SNR} = 5$ ,  $T_{\text{exp}} = 10^5 \text{ s}$  for the black line.

background model varies with the target coordinates (RA, Dec.) and time of year. Here, we choose position on the sky to the Hubble Ultra-Deep Field at RA =  $03^{\text{h}}32^{\text{m}}39^{\text{s}}.00$ , Dec. =  $-27^{\circ}47'29''.0$  and choose the background configuration to be ‘low’. For each band, we then set up all the ETC parameters as described and then vary the apparent magnitude of the source until a target signal-to-noise ratio is reached. This then sets the corresponding apparent magnitude detection limit for this band for various exposure times and target signal-to-noise ratios.

From the simulation, we derive the apparent band luminosity of galaxies following the procedure in Paper I. The luminosity functions in each band from different simulations are combined and the resulting combined luminosity functions can be well reproduced by the Schechter function (Schechter 1976)

$$\phi(M) = \frac{dn}{dM} = \frac{0.4 \ln(10) \phi^*}{10^{0.4(M-M^*)}(\alpha+1)} e^{-10^{-0.4(M-M^*)}}, \quad (1)$$

where  $M$  is the rest-frame (apparent) magnitude when describing rest-frame (apparent) luminosity functions. The Schechter function can also be expressed as

$$\phi(L) = \frac{dn}{d \log L} = \phi^* \left( \frac{L}{L_*} \right)^{\alpha+1} e^{-L/L_*}, \quad (2)$$

when describing the number density of galaxies per dex of luminosity, where  $\phi^*$  is the number density normalization,  $L_*$  is the break

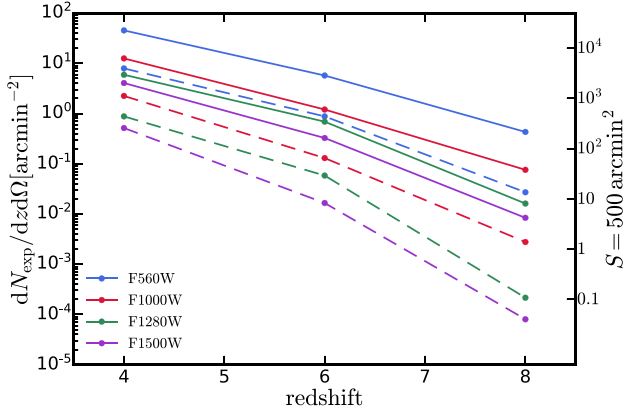
luminosity and  $\alpha$  is the faint-end slope. The best-fitting Schechter function parameters are shown in Appendix B. In Fig. 4, we compare the binned estimations and Schechter fits of the apparent band luminosity functions of four selected *JWST* MIRI broadbands to the detection limits of MIRI in Table 2. This comparison demonstrates the promise of MIRI for the identification of galaxies up to  $z \sim 8$ , assuming a survey area<sup>5</sup> of  $\sim 500 \text{ arcmin}^2$  and depth of  $\Delta z = 1$ .

Based on the Schechter fit of the luminosity function, the cumulative number density of galaxies can be calculated by integrating the Schechter function as

$$\begin{aligned} \phi_{\text{cum}}(< M^{\text{lim}}) &= \int_{L^{\text{lim}}}^{\infty} \phi^* \left( \frac{L}{L_*} \right)^{\alpha} \exp\left(-\frac{L}{L_*}\right) \frac{dL}{L_*} \\ &= \phi^* \Gamma_{\text{inc}}(\alpha + 1, 10^{-0.4(M^{\text{lim}}-M^*)}), \end{aligned} \quad (3)$$

where  $\Gamma_{\text{inc}}(a, z) = \int_z^{\infty} t^{a-1} e^{-t} dt$  is the incomplete upper Gamma function and  $M^{\text{lim}}$  is the magnitude limit of integration. Then the expected number of galaxies above the detection limit ( $m^{\text{lim}}$ ) in a

<sup>5</sup>For reference, the survey area of the Spitzer Extended Deep Survey (SEDS; Ashby et al. 2013) in the IRAC 3.6 and  $4.5 \mu\text{m}$  bands is about  $1.46 \text{ deg}^2$ , which was followed by the Spitzer-Cosmic Assembly Deep Near-infrared Extragalactic Legacy Survey (S-CANDELS; Ashby et al. 2015) with a smaller area of  $0.16 \text{ deg}^2$  ( $576 \text{ arcmin}^2$ ) and increased depth.



**Figure 5.** Expected number counts of galaxies (theoretically estimated) in *JWST* MIRI bands. We show the number of galaxies per unit area of field of view and per unit survey depth in the MIRI bands as a function of redshift. The numbers on the right-hand side assume a survey area of  $500 \text{ arcmin}^2$ . The dashed lines show predictions for the relatively shallow detection limit (assuming  $\text{SNR} = 10$ ,  $T_{\text{exp}} = 10^4 \text{ s}$ ) and the solid lines show predictions for the deep detection limit (assuming  $\text{SNR} = 5$ ,  $T_{\text{exp}} = 10^5 \text{ s}$ ). We refer to the results as ‘theoretically estimated’ since we have assumed effectively 100 per cent completeness in detection and selection of galaxies.

survey volume can be calculated as

$$N_{\text{exp}} \simeq \phi_{\text{cum}}(< m_{\text{lim}}) \frac{dV_{\text{com}}}{d\Omega dz}(z) \Delta\Omega \Delta z, \quad (4)$$

where  $\Delta\Omega$  is the solid angle corresponding to the area of the survey,  $\Delta z$  is the redshift coverage of the survey and  $dV_{\text{com}}/d\Omega dz$  is the differential comoving volume element at the redshift of the survey

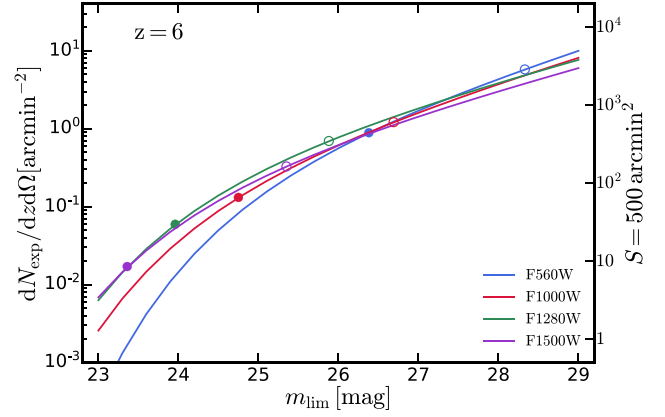
$$\frac{dV_{\text{com}}}{d\Omega dz}(z) = \frac{c(1+z)^2 d_A(z)^2}{H_0 E(z)}, \quad (5)$$

where  $d_A(z)$  is the angular diameter distance and  $H(z) = H_0 E(z)$  is the Hubble parameter at redshift  $z$ . We note that the calculations here effectively assume that all the galaxies above the magnitude limit  $m_{\text{lim}}$  can be detected and selected in a real imaging survey. However, in reality, the completeness correction is necessary to recover the physical abundance of galaxies, the expected number of galaxies detected should be

$$N_{\text{exp}} \simeq \left( \int_{-\infty}^{+\infty} dm \frac{d\phi_{\text{cum}}(< m)}{dm} P(m, z) \right) \frac{dV_{\text{com}}}{d\Omega dz}(z) \Delta\Omega \Delta z, \quad (6)$$

where  $P(m, z)$  is the completeness function for a specific observation. Our calculations above effectively assume  $P(m, z) = 0$  when  $m > m_{\text{lim}}$  and  $P(m, z) = 1$  when  $m \leq m_{\text{lim}}$ . But in a real observation, the transition of  $P(m, z)$  at  $m_{\text{lim}}$  would be smooth. The shape of the completeness function depends on the details of the observational configuration, background noises and selection criteria. Therefore, the expected numbers calculated here should be interpreted as a theoretical estimate and one needs to be cautious comparing them with real observational results.

The expected number counts of galaxies (theoretically estimated) in the four selected MIRI broadbands at  $z = 4-8$  are shown in Fig. 5. At  $z = 6$ , assuming a target  $\text{SNR} = 5$  and  $T_{\text{exp}} = 10^5 \text{ s}$ ,  $\sim 3000$  ( $\sim 500$ ) galaxies are expected in F560W (F1000W) with a survey area of  $\sim 500 \text{ arcmin}^2$  and depth of  $\Delta z = 1$ . Even at  $z = 8$  in F1500W (which has the poorest sensitivity among the four bands selected),  $\sim 5$  galaxies are expected to be detected with the same observational configuration. Previously, the deepest large NIR galaxy surveys (e.g. Ashby et al. 2013; Steinhardt et al. 2014; Ashby et al. 2015) have

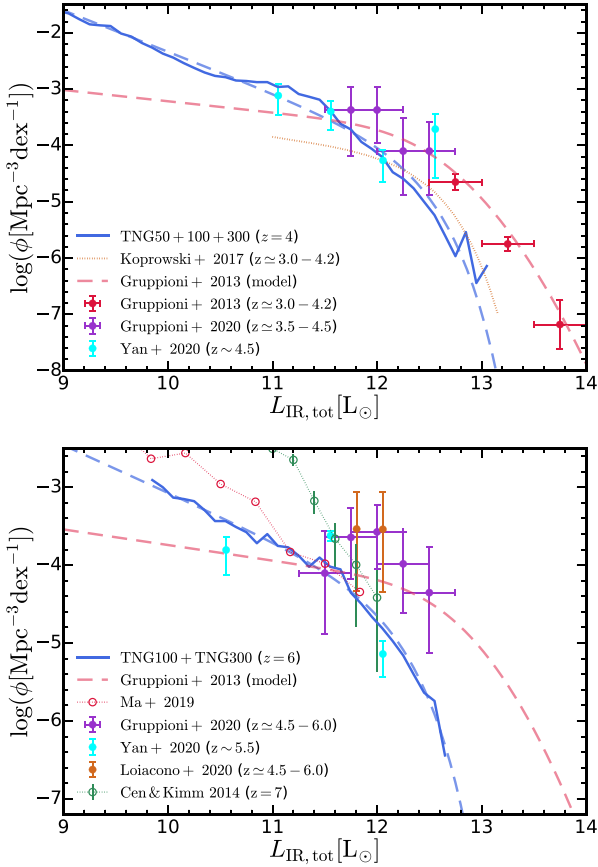


**Figure 6.** Expected number counts of galaxies (theoretically estimated) in *JWST* MIRI bands versus detection limit. We show the number of galaxies per unit area of field of view and per unit survey depth in the MIRI bands as a function of detection limit at  $z = 6$ . The detection limit is expressed as the limiting apparent magnitude. The labelling is the same as Fig. 5. The solid (open) circles indicate the number counts at the detection limit assuming  $\text{SNR} = 10$  and  $T_{\text{exp}} = 10^4 \text{ s}$  ( $\text{SNR} = 5$  and  $T_{\text{exp}} = 10^5 \text{ s}$ ). We note that the detection limit does not scale with the total exposure time in a trivial way and also depends on the details of the observational configuration.

been mainly conducted by the Infrared Array Camera (IRAC; Fazio et al. 2004) on the *Spitzer Space Telescope* (Werner et al. 2004). In these surveys, the  $5-\sigma$  survey depth is around 25 mag in the IRAC 3.6 and  $4.5 \mu\text{m}$  bands (Steinhardt et al. 2014; Mortlock et al. 2017; Stefanon et al. 2017) and around 24.5 mag in the IRAC 5.8 and  $8.0 \mu\text{m}$  bands (deepest data in the GOODS-N field as discussed in Stefanon et al. 2017). Compared to the detection limits of *JWST* MIRI with comparable exposure time ( $10^5 \text{ s} \sim 30 \text{ h}$ ), the detection limit at  $5-10 \mu\text{m}$  can be pushed deeper by *JWST* by about 2 mag and the number of galaxies selected at  $z \gtrsim 6$  with complete photometric data at rest-frame UV to NIR  $8 \mu\text{m}$  will be boosted from order ten (Stefanon et al. 2017) to a few hundred. In Fig. 6, we show the expected number counts of galaxies at  $z = 6$  detected in the four MIRI broadbands as a function of detection limit. Relevant numbers can be read out when the detection limits vary with other observational configurations. For shallow detection limits, the number counts are lower in MIRI bands located at shorter wavelengths due to stronger dust attenuation at shorter wavelengths in rest-frame NIR. The difference diminishes for deep detection limits since faint, low-mass galaxies dominate the total number counts and they are less affected by dust attenuation.

### 4.3 Bolometric IR luminosity functions and obscured SFRD

The FIR luminosities of galaxies are dominated by dust continuum emission, which is reprocessed from the UV emission of young stellar populations. Therefore, FIR luminosities are sensitive to on-going star formation in galaxies. Since the IR SEDs of galaxies peak at FIR wavelengths, the bolometric IR luminosity integrated at  $8-1000 \mu\text{m}$  has often been used as an indicator for star formation (e.g. Kennicutt 1998; Murphy et al. 2011). In particular, IR based measurements can reveal the star formation in heavily dust obscured galaxies which may be too faint to be observed in the rest-frame UV and optical. The IR measurements could reflect a significant portion of the total amount of star formation in galaxies. Therefore, the constraints of the bolometric IR luminosity functions, as well as the abundance of



**Figure 7.** Bolometric IR luminosity function. The galaxy bolometric IR luminosity functions at  $z = 4$  and  $z = 6$  from the IllustrisTNG simulations are presented in solid lines. Binned estimations and Schechter fits from observations (Gruppioni et al. 2013; Koprowski et al. 2017; Gruppioni et al. 2020) are shown with solid markers and dashed lines. Predictions from theoretical works (Cen & Kimm 2014; Ma et al. 2019) are shown with open markers and dotted lines. Compared to observations, TNG underpredicts the abundance of most luminous IR galaxies at both redshifts and predicts a steeper faint-end slope.

the most luminous IR emitters (usually accompanied by strong dust obscuration) are important to constrain the cosmic SFRD.

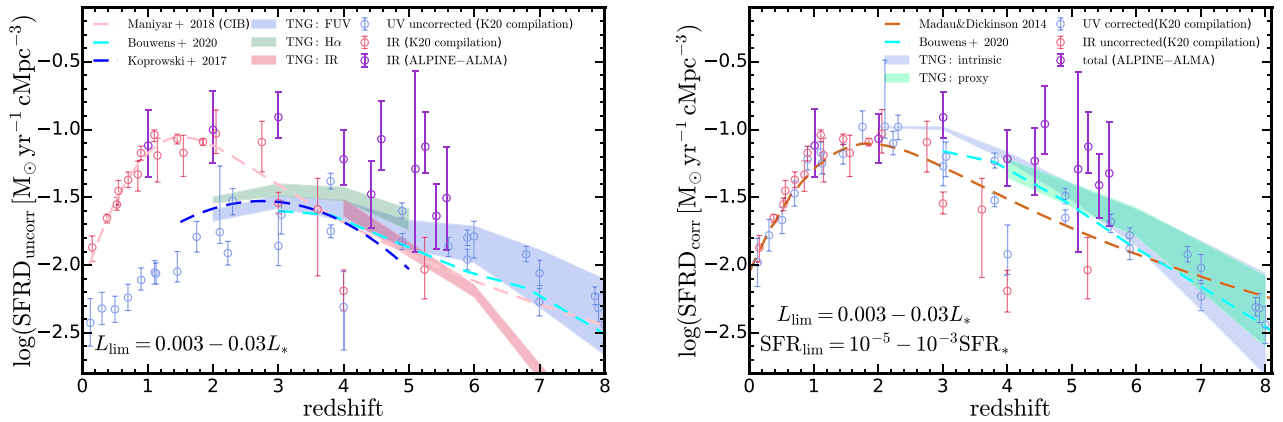
#### 4.3.1 Bolometric IR luminosity function

In Fig. 7, we present the galaxy bolometric IR luminosity function at  $z = 4$  and  $z = 6$  from the IllustrisTNG simulations. The luminosity functions from different simulations are combined together and resolution corrections are not applied since we find that the TNG50, TNG100, and TNG300 results are consistent with each other in shared dynamical ranges. We compare the simulation predictions with the following observations: the Herschel luminosity functions (e.g. Gruppioni et al. 2013); the SCUBA-2 luminosity functions (e.g. Koprowski et al. 2017); the ALPINE-ALMA luminosity functions (e.g. Gruppioni et al. 2020). We also include the bolometric IR luminosity functions converted from the [CII] luminosity functions of UV-selected galaxies in Yan et al. (2020), Loiacono et al. (2020) by Gruppioni et al. (2020). At  $z \approx 4$ , the Herschel and the ALPINE-ALMA luminosity function measurements are mutually consistent with each other and cover complementary luminosity ranges. However, the SCUBA-2 luminosity function

is systematically lower, potentially due to the different method in calculating the IR luminosities and incompleteness in sample selection. The TNG prediction is in agreement with the Herschel and ALPINE-ALMA luminosity functions at  $L_{\text{IR}} \lesssim 10^{12} L_{\odot}$  while underpredicting the abundance of galaxies at the bright end with  $L_{\text{IR}} \gtrsim 10^{13} L_{\odot}$ . The actual number of IR luminous galaxies is important for the cumulative IR luminosity density of galaxies, since the faint-end slopes of IR luminosity functions are usually shallow ( $\alpha \gg -2$ ). In addition, at the faint end, TNG predicts steeper luminosity functions than the extrapolation of observational based models. If we translate the typical IR luminosity ( $L_{\text{IR}} \sim 10^{13} L_{\odot}$ ) of these ‘missing’ galaxies in TNG simulations to SFR, it would roughly be  $\sim 1000 M_{\odot}/\text{yr}$  (Kennicutt 1998; Murphy et al. 2011). Assuming the averaged properties of high-redshift ( $z \gtrsim 3$ ) sub-millimeter galaxies with a typical specific star formation rate of  $\log s\text{SFR} \sim -8.5$  (Magnelli et al. 2012; da Cunha et al. 2015; Miettinen et al. 2017), the implied stellar masses of these ‘missing’ galaxies will be  $M_{*} \sim 10^{11.5} M_{\odot}$ .

At  $z \approx 6$ , the TNG prediction is consistent with the observations at  $L_{\text{IR}} \lesssim 10^{11.5} L_{\odot}$  while it is lower at  $L_{\text{IR}} \gtrsim 10^{12} L_{\odot}$  by about an order of magnitude. The difference of the redshift bins in observations and simulations will likely not lead to such a large discrepancy, since the evolution of the number density normalization and the break luminosity is slow at  $z \gtrsim 2$  (Gruppioni et al. 2020). At the bright end, similar to the discrepancy at  $z = 4$ , the simulation underpredicts the abundance of galaxies at  $L_{\text{IR}} \gtrsim 10^{12} L_{\odot}$  and predicts a steeper faint-end slope compared to the evolutionary model constrained in Gruppioni et al. (2013). We also compare our results with theoretical predictions from Cen & Kimm (2014; radiative transfer calculations on 198 galaxies with  $5 \times 10^8 M_{\odot} < M_{*} < 3 \times 10^{10} M_{\odot}$  at  $z = 7$ ) and Ma et al. (2019; radiative transfer on FIRE-2 high-redshift simulations using SKIRT). Their predictions are in good agreement with TNG at  $L_{\text{IR}} \sim 10^{12} L_{\odot}$  while suggesting much steeper faint-end luminosity functions, which deviate from observations further. At the bright end, the simulations by Cen & Kimm (2014) and Ma et al. (2019) lose predictive power since the most massive, heavily obscured galaxies were not sampled. If we extrapolate their IR luminosity functions to brighter luminosities (by either a power law or exponential cut-off), both of them will infer a lower abundance of luminous IR galaxies compared to observations, similar to the TNG prediction.

In observational studies, merger-driven starbursts have been considered as the classical explanation for the extremely high IR luminosities of sub-millimeter galaxies (e.g. Chakrabarti et al. 2008; Engel et al. 2010; Narayanan et al. 2010; Hayward et al. 2011). Since our galaxy selection and radiative transfer calculations are all performed based on sub-haloes identified by the SUBFIND algorithm, it is possible that we mistreat merging systems (with high intrinsic SFRs and IR luminosities) as distinct, individual merging galaxies and therefore underpredict the abundance of IR-luminous systems. To test this scenario, we first select all the galaxies with  $\text{SFR} \geq 100 M_{\odot}/\text{yr}$  (which roughly corresponds to  $L_{\text{IR}} \gtrsim 10^{12} L_{\odot}$ ) and  $M_{*} \geq 100 \times m_b$  at  $z = 4$  in TNG300. Based on this sample of galaxies, we start from the most massive galaxy (as the host galaxy) and link galaxy companions with distance to host  $d \leq d_{\text{lim}} = 50 \text{ pkpc}$  to this host galaxy. We repeat the same process to the remaining set of galaxies that have not been linked to other galaxies. Finally, after all the galaxies are properly linked to their hosts, the SFRs of galaxies are recalculated by summing the SFRs of all the galaxy companions linked to the host galaxies. We find that the typical enhancement in SFR of the host galaxies is  $\Delta(\log \text{SFR}) = 0.2^{+0.08}_{-0.10}$  while the abundance of galaxies with  $\text{SFR} \geq 1000 M_{\odot}/\text{yr}$  (which roughly corresponds to  $L_{\text{IR}} \gtrsim 10^{13} L_{\odot}$ ) does not change at all. Even



**Figure 8.** Evolution of the cosmic SFRD. In the left-hand panel, we show the uncorrected SFRD estimated from the UV,  $H\alpha$ , and IR luminosity densities of galaxies in the IllustrisTNG simulations. The UV and  $H\alpha$  luminosities are based on results in Paper I and Paper II. The luminosity densities are derived by integrating the luminosity functions to  $0.003\text{--}0.03 L_*$ , where  $L_*$  is the break luminosity. The luminosity densities are converted to SFRD assuming the conversion factors in Murphy et al. (2011). We compare the simulation results with observations compiled in Khusanova et al. (2020). The UV (IR) observations are shown with blue (red) points. Specifically, we show the most recent IR-based estimations from Khusanova et al. (2020), Loiacono et al. (2020), and Gruppioni et al. (2020) in purple points. For reference, we show the evolution of SFRD constrained by UV observations from Koprowski et al. (2017) and Bouwens et al. (2020) in dashed lines. In the right-hand panel, we show the total SFRD (corrected for dust attenuation). The light and dark blue dashed lines show constraints from Bouwens et al. (2020) and Koprowski et al. (2017) based on compiled UV observations. In the right-hand panel, we show the total SFRD as a function of redshift. From simulations, we calculate it through two approaches: the summation of SFRDs inferred by UV and IR indicators (labelled with ‘proxy’), and the SFRD measured from the instantaneous SFR of gas cells in simulations (labelled with ‘intrinsic’). For observations, we show the SFRD measurements corrected for dust attenuation. The predicted SFRD from simulations is lower than the results of the ALPINE-ALMA survey, in particular the obscured SFRD.

if we increase  $d_{\text{lim}}$  to  $1\text{ pMpc}$ , the abundance of galaxies with  $\text{SFR} \geq 1000 M_\odot/\text{yr}$  only increases by about  $0.2\text{ dex}$  which is still far from explaining the underprediction we found in the bolometric IR luminosity function. So we conclude that the underprediction is unlikely related to the definition and selection of galaxies in post-processing.

This underpredicted abundance of luminous IR galaxies in TNG is consistent with the underpredicted UV dust attenuation in massive galaxies and the underpredicted abundance of heavily obscured UV red systems found in Paper II. Hayward et al. (2021) has also reported the underpredicted counts of sub-millimeter galaxies in TNG. Since our model is calibrated based on the dust-attenuated UV luminosity functions, the solution to the tension would not only require a higher abundance of dust (stronger dust attenuation) but also higher intrinsic UV emission (either stronger star formation or additional radiation sources). One plausible explanation is resolution effects. Though resolution corrections have been considered in our model, the predictions at the bright (massive) end are still primarily determined by TNG100 and TNG300, which may fall short of resolving star formation and metal enrichment in the dense ISM. A similar effect has been investigated in Lim et al. (2020), where the star formation efficiency of proto-clusters in TNG300 is much lower than observations and part of it can be attributed to resolution effects. However, the luminosity functions (of all the bands we studied) of TNG100 and TNG300 at  $z \geq 4$  are roughly identical at the bright (massive) end (Vogelsberger et al. 2020), indicating that convergence (of galaxy bulk luminosities and masses) is reached in massive galaxies at the resolution level in the redshift range. Another possible explanation is the sub-grid stellar/AGN feedback model adopted in TNG. In Hayward et al. (2021), it is shown that the original Illustris simulation with an older feedback model predicts the correct abundance of sub-millimeter galaxies, and massive galaxies in TNG may be quenched too early by feedback. Though it is hard to isolate the exact cause, the abundance of bright IR galaxies remains an appealing channel to constrain the feedback model in cosmological simulations.

### 4.3.2 Obscured star formation at high redshift

To relate the luminosity functions to the obscured SFRD in the Universe, we first perform fits to the bolometric IR luminosity functions with the Schechter function (equation 1, the best-fitting parameters are shown in Appendix B) and integrate best-fitting Schechter functions to derive the cumulative IR luminosity density

$$\begin{aligned} \rho_{\text{IR}}(<L_{\text{IR}}^{\text{lim}}) &= \int_{L_{\text{IR}}^{\text{lim}}}^{\infty} \phi^* L_* \left( \frac{L_{\text{IR}}}{L_*} \right)^{\alpha+1} \exp\left(-\frac{L_{\text{IR}}}{L_*}\right) \frac{dL_{\text{IR}}}{L_*} \\ &= \phi^* L_* \Gamma_{\text{inc}}(\alpha+2, L_{\text{IR}}^{\text{lim}}/L_*), \end{aligned} \quad (7)$$

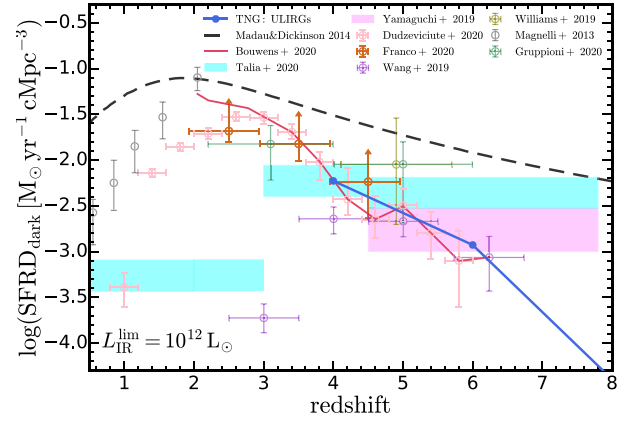
where  $\Gamma_{\text{inc}}(a, z)$  is the incomplete upper Gamma function as in equation (3) and  $L_{\text{IR}}^{\text{lim}}$  is the limit of integration which will be discussed later. The derived cumulative IR luminosity density is converted to the  $\text{SFRD}_{\text{IR}}$  (the SFRD measured in IR) using the calibration in Murphy et al. (2011), which assumed the Chabrier (2003) initial mass function consistent with the choice in TNG. We apply the same procedure to the predicted UV luminosity functions in Paper I and the  $H\alpha$  luminosity functions in Paper II to derive  $\text{SFRD}_{\text{UV}}$  and  $\text{SFRD}_{H\alpha}$ . We finally derive the SFRD traced by the three indicators and we sum up  $\text{SFRD}_{\text{IR}}$  and  $\text{SFRD}_{\text{UV}}$  to get the total SFRD inferred from indicators.<sup>6</sup>

In the left-hand panel of Fig. 8, we compare these predictions with the observational constraints compiled in Khusanova et al. (2020, see references therein). These observational results have been converted to the Chabrier (2003) initial mass function. These observations have been divided into three groups: blue circles, the SFRD derived from FUV observations without dust attenuation corrections; red

<sup>6</sup>In principle, both UV and  $H\alpha$  light trace the unobscured star formation in galaxies. The dust attenuation of them could be different due to the geometry of radiation source distribution. Since the majority of the energy absorbed by dust is in UV, it is better to pair  $\text{SFRD}_{\text{UV}}$ , rather than  $\text{SFRD}_{H\alpha}$ , with  $\text{SFRD}_{\text{IR}}$  to calculate total SFRD.

circles, SFRD derived from IR and sub-millimeter observations; purple circles, SFRD<sub>IR</sub> derived from the recent ALPINE-ALMA survey (Gruppioni et al. 2020; Khusanova et al. 2020; Loiacono et al. 2020). The predictions from TNG<sup>7</sup> are presented as shaded regions with the lower (upper) boundary corresponding to  $0.03$  ( $0.003$ )  $L_*$  as the integration limits, where  $L_*$  is the break luminosity of the best-fitting Schechter function. These integration limits are commonly adopted in literature. The SFRD<sub>UV</sub> predicted from TNG is in agreement with the majority of the UV observations at  $z = 2-8$ . The lower boundary of the prediction agrees almost perfectly with the most recent constraint from Bouwens et al. (2020), where  $0.03 L_*$  was adopted as the integration limit. The uncertainty induced by the integration limit is small compared to the observational uncertainties at  $z \lesssim 5$ , but becomes significant at  $z \gtrsim 6$  due to the steep faint-end slope of the luminosity function there. In addition, the SFRD<sub>H $\alpha$</sub>  predicted from TNG agrees with the SFRD<sub>UV</sub> with  $\lesssim 0.2$  dex differences. Such differences could be attributed to the uncertainties in modelling the emission line intensity of young stellar populations as discussed in Paper II. The SFRD<sub>IR</sub> predicted from TNG agrees with the model in Maniyan, Béthermin & Lagache (2018), derived based on the cosmic infrared background anisotropies, at  $z \lesssim 6$  but is lower at  $z \gtrsim 6$ . Compared to the recent ALPINE-ALMA observations, the TNG prediction is about half an order of magnitude lower. The deficiency is related to the underpredicted abundance of luminous IR galaxies in TNG, compared to the IR luminosity functions constrained by the ALPINE-ALMA survey, as shown above. The prediction from TNG supports the picture that the obscured SFRD dominates at low redshift, starts to decline at  $z \sim 2$  and eventually becomes subdominant compared to the unobscured SFRD at  $z \gtrsim 5$ .

In the right-hand panel of Fig. 8, we show the total SFRD predicted from the IllustrisTNG simulations and compare it with the SFRD measured in the IR and UV that have been corrected for dust attenuation. To calculate the total SFRD from simulations, the first way is to simply sum up SFRD<sub>UV</sub> and SFRD<sub>IR</sub> derived via cumulative luminosity densities. An alternative way is to measure the SFRD from the instantaneous star formation in gas cells in simulations. We determine the SFR of each TNG galaxy by summing up the SFRs in gas cells within a 30 pkpc aperture of the galaxy. The choice of aperture is consistent with the box size we set for radiative transfer post-processing and our definition for other galaxy bulk properties. We perform resolution corrections to the SFRs of galaxies in TNG100 and TNG300 using the method described in Paper I and Paper II. With the resolution corrected SFRs of galaxies, we construct SFR functions and combine the SFR functions of TNG50, TNG100, and TNG300 as in Paper I and Paper II. Finally, we fit the combined SFR functions with the Schechter function and integrate the Schechter function to the integration limit,  $10^{-3}$  ( $10^{-5}$ ) SFR<sub>\*</sub>, where SFR<sub>\*</sub> is the break SFR of the SFR function. The limit is smaller compared to the integration limits we choose when integrating over the luminosity function, because the break SFR of the SFR functions is relatively larger and empirically we find these limits make the SFRD consistent with observations simultaneously at  $z = 2$  and  $z = 8$ . As shown in the right-hand panel of Fig. 8, the total SFRDs derived by the two methods agree remarkably well and give similar level of uncertainties induced by integration limits. Compared to the observational constraints, the total SFRD from TNG is larger than the Madau & Dickinson (2014) model at  $2 \lesssim z \lesssim 6$  while it gives similar



**Figure 9.** Cosmic SFRD contributed by the optical-NIR-dark galaxies. We show the SFRD contributed by galaxies with  $L_{\text{IR}} > 10^{12} L_{\odot}$  (ULIRGs) in TNG through integration over the predicted bolometric IR luminosity function. This serves as an upper limit for the SFRD of optical-NIR-dark galaxies. We compare the results with the observational constraints compiled in Bouwens et al. (2020), as well as the constraints from the ALPINE-ALMA survey (Gruppioni et al. 2020; Talia et al. 2020), as labelled. The TNG prediction is in agreement with some of the observations at  $z \geq 4$  while being  $\sim 1\sigma$  lower than the estimations from Williams et al. (2019), Gruppioni et al. (2020), and Talia et al. (2020).

prediction at  $z \gtrsim 7$ . The lower boundary of the prediction agrees well with the recent estimation in Bouwens et al. (2020; assuming  $L_{\text{lim}} = 0.03 L_*$ ) where an updated dust correction was applied and the contribution of heavily obscured, optical/NIR dark galaxies was accounted for. However, the predicted SFRD is still  $\sim 0.2$  dex lower than what is suggested by the ALPINE-ALMA survey.

To illustrate the contribution of heavily obscured, luminous IR galaxies to the cosmic SFRD, we show the TNG prediction and several direct/indirect observational constraints of the SFRD in optical/NIR dark galaxies in Fig. 9. We show the fiducial model adopted in Bouwens et al. (2020) and the compiled observational data therein, including the SFRD of ULIRGs in Magnelli et al. (2013; integrated to  $L_{\text{IR}} = 10^{12} L_{\odot}$ ) and Dudzeviciute et al. (2020; integrated to  $S_{870} = 1 \text{ mJy}$ ), constraints of NIR dark galaxies in Yamaguchi et al. (2019), Williams et al. (2019), Wang et al. (2019), and Franco et al. (2020). In addition, we include constraints from the ALPINE-ALMA survey (Gruppioni et al. 2020; Talia et al. 2020). The TNG prediction is derived by integrating the bolometric IR luminosity function to  $L_{\text{IR}} = 10^{12} L_{\odot}$ , which roughly corresponds to SFR  $> 100 M_{\odot}/\text{yr}$ . The criterion matches the property of typical DSFGs that are missed from optical and NIR observations (e.g. Wang et al. 2019; Yamaguchi et al. 2019; Talia et al. 2020). Since this essentially assumes that all the galaxies with  $L_{\text{IR}} = 10^{12} L_{\odot}$  are optical/NIR dark, the prediction should be taken as an upper limit. Among observations, despite the consensus on the steady decline of the SFRD contributed by luminous IR galaxies at  $z \gtrsim 2$ , the quantitative contribution varies in the literature, with roughly an order of magnitude variation at  $z \gtrsim 4$ . The TNG prediction is in agreement with some of the observations at  $z \geq 4$  while in tension with the results from Williams et al. (2019), Gruppioni et al. (2020), and Talia et al. (2020) at a  $1\sigma$  level. Considering that the TNG prediction is an upper limit for the SFRD of optical/NIR dark galaxies, the discrepancy is even larger than revealed by this comparison. The discrepancy is related to the underprediction of luminous IR galaxies in TNG as discussed above and demonstrates that the difference in

<sup>7</sup>We note that the continuous evolution of the IR luminosity density, as well as the corresponding SFRD is obtained through interpolation of the simulation and post-processing results at  $z = 4, 6, 8$ .

the total SFRD between TNG predictions and observations is mainly due to a deficiency of luminous IR galaxies in TNG.

#### 4.4 Dust temperature

Most of the IR and sub-millimeter emission from galaxies is produced through thermal emission of dust grains in the ISM, heated by the UV radiation of young stellar populations (if not accounting for AGN activity and CMB heating). The volume averaged thermal emission of dust is often described by a modified blackbody (MBB) spectrum,  $f_\nu \sim \epsilon_\nu B_\nu(T_{\text{dust}})$ , where  $B_\nu(T_{\text{dust}})$  is the blackbody spectrum and  $\epsilon_\nu \sim \nu^\beta$  is the volume averaged emissivity function that accounts for the dust opacity (in the optical thin limit). At FIR wavelengths, the emissivity index  $\beta$  takes the value about 2 according to dust scattering theory (Weingartner & Draine 2001) and consistent with observational constraints (e.g. Dunne et al. 2000; Draine et al. 2007).  $T_{\text{dust}}$  represents the temperature of the relatively cool dust component which dominates the emission at long wavelengths (as opposed to the warm dust that dominates mid-IR emission). The peak of a MBB spectrum ( $L_\nu$ ) is related to the so-called ‘peak’ dust temperature as

$$\lambda_{\text{peak}} = 96.64 \left( \frac{T_{\text{peak}}}{30\text{K}} \right)^{-1} \mu\text{m}, \quad (8)$$

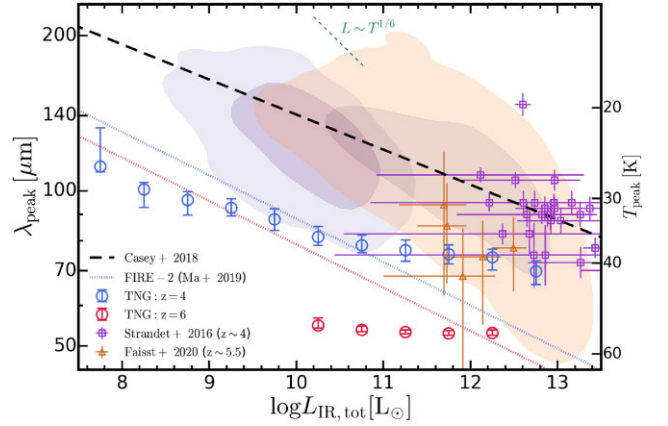
assuming  $\beta = 2$ . The relation has the same form as the Wien’s displacement law, but here the peak is defined as where  $L_\nu$  reaches a maximum. This is a direct and less model-dependent way to measure dust temperature based solely on the observed FIR spectrum of galaxies and is widely used in observational studies (e.g. da Cunha et al. 2013; Schreiber et al. 2018; Faisst et al. 2020). To determine the peak of the FIR spectrum ( $L_\nu$ ), we first find where the maximum flux is reached in the SKIRT output spectrum. Since the wavelength grid for SKIRT calculations is sparse, to determine the peak more accurately, we interpolate the galaxy spectrum with a cubic spline and determine the peak based on the interpolated function.

A more robust way in determining the dust temperature is through SED modelling. Considering the warm dust emission in the mid-IR and dust opacity (not in the optically thin limit), the dust continuum emission can be described as (Blain, Barnard & Chapman 2003; Casey 2012)

$$\begin{aligned} L_\nu(\lambda) &= N_{\text{bb}} f(\lambda, \beta, T_{\text{sed}}) + N_{\text{pl}} (\lambda/\lambda_c)^\alpha e^{-(\lambda/\lambda_c)^2}, \\ N_{\text{pl}} &= N_{\text{bb}} f(\lambda_c, \beta, T_{\text{sed}}), \\ f(\lambda, \beta, T_{\text{sed}}) &= \frac{(1 - e^{-(\lambda_0/\lambda)^\beta})(\lambda/c)^3}{e^{hc/\lambda k T_{\text{sed}}} - 1}, \end{aligned} \quad (9)$$

where  $\alpha$  is the mid-IR spectral slope,  $\lambda_c$  is related to  $\alpha$  (as defined in Casey 2012, and taking a typical value of  $\sim 50\mu\text{m}$ ),  $N_{\text{bb}}$  is a normalization factor,  $\lambda_0$  is the wavelength where the optical depth is unity,  $T_{\text{sed}}$  is the SED temperature. The free parameters of the model are  $\alpha$ ,  $\beta$ ,  $T_{\text{sed}}$ , and  $N_{\text{bb}}$ . These parameters can be determined by fitting the IR spectrum from SKIRT calculations with the model.  $T_{\text{sed}}$  is usually larger than  $T_{\text{peak}}$  especially when the optically thin assumption is no longer valid (Casey 2012). We note that both  $T_{\text{sed}}$  and  $T_{\text{peak}}$  should be considered as light-weighted dust temperatures, as opposed to the mass-weighted temperature of cold dust emitting at the Rayleigh–Jeans tail of the FIR spectrum, and are more appropriate for comparing to observational constraints. The CMB is an additional source of dust heating, especially at high redshift when the CMB temperature is close to the dust temperature, which is not captured by the simulations. To account for that, a CMB correction of dust temperatures is implemented as (e.g. da Cunha et al. 2013)

$$T_{\text{dust}} = (T_{\text{dust},i}^{4+\beta} + T_{\text{CMB}}^{4+\beta}(z) - T_{\text{CMB}}^{4+\beta}(0))^{1/(4+\beta)}. \quad (10)$$

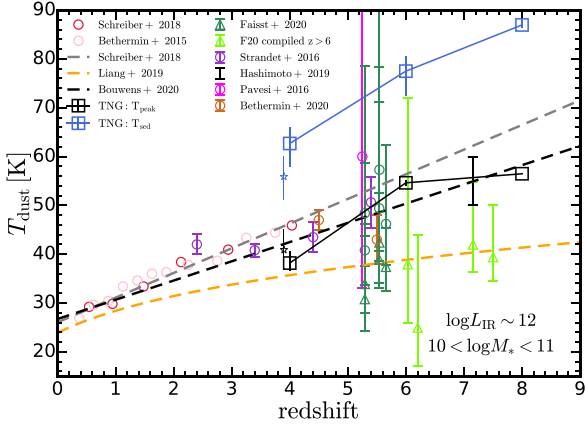


**Figure 10.** Peak of IR spectrum (dust peak temperature) versus bolometric IR luminosity. We show the peak of the IR spectrum versus bolometric IR luminosity from TNG in open circles, with error bars indicating  $1.5\sigma$  dispersion (87 per cent inclusion) in each bin. The peak of the spectrum is related to the peak dust temperature  $T_{\text{peak}}$  with equation (8). We compare the results with observational constraints based on: H-ATLAS samples at  $z < 0.1$  in Valiante et al. (2016; purple shaded region, 1 and  $2\sigma$  contour); COSMOS samples at  $0.5 < z < 2$  from Lee et al. (2013; orange shaded region); SPT-detected DSFGs at  $z \sim 4$  in Strandet et al. (2016; purple squares); ALMA observations of four main-sequence galaxies at  $z \sim 5.5$  from Faisst et al. (2020; orange triangles). The former three are compiled by Casey et al. (2018) and a suggested redshift-independent relation from this work is shown in the black dashed line. For reference, theoretical predictions from the FIRE-2 simulations (Ma et al. 2019) are shown in dotted lines. The TNG predicted IR peak wavelengths of  $z \geq 4$  galaxies are shorter than the low-redshift galaxies and thus disfavour a redshift-independent  $\lambda_{\text{peak}}-L_{\text{IR}}$  relation. The peak wavelength anticorrelates with IR luminosity until a plateau of dust temperature is reached at  $L_{\text{IR}} \gtrsim 10^{11} L_\odot$ . TNG predictions at  $z = 4$  are compatible with the SPT-selected galaxies at  $z \sim 4$  while the predictions at  $z = 6$  give higher peak dust temperatures than suggested by  $z \sim 5.5$  galaxies in Faisst et al. (2020).

where  $T_{\text{dust},i}$  is the intrinsic dust temperature and  $T_{\text{CMB}}(z) = T_{\text{CMB}}(0)(1+z) = 2.73\text{K}(1+z)$  is the CMB temperature at  $z$ . In addition, since the dust continuum of high-redshift galaxies is always measured against the CMB background, the detectability, as well as the measured FIR fluxes are affected by CMB. However, this effect has already been corrected in most of the observational studies.

##### 4.4.1 $T_{\text{dust}}$ versus $L_{\text{IR}}$

In Fig. 10, we show the peak of the IR spectrum (and  $T_{\text{peak}}$  equivalently) as a function of bolometric IR luminosity of TNG galaxies. The peak of the spectrum is translated to the peak dust temperature  $T_{\text{peak}}$  using equation (8). Only galaxies with  $M_* > 1000 m_b$  ( $m_b$  is the baryon mass resolution) are considered in this analysis to reduce the random scatter caused by poor sampling of radiation sources. We compare the results with observational constraints computed by Casey et al. (2018) based on the H-ATLAS samples at  $z \lesssim 0.1$  from Valiante et al. (2016), the COSMOS samples at  $0.5 \lesssim z \lesssim 2$  from Lee et al. (2013) and the *South Pole Telescope* (SPT) detected DSFGs at  $z \sim 4$  from Strandet et al. (2016). In addition, we include results from ALMA observations of four main-sequence galaxies at  $z \sim 5.5$  reported in Faisst et al. (2020). Theoretical predictions from the FIRE-2 simulations (Ma et al. 2019) are shown for reference. Our results disfavour the redshift-independent  $\lambda_{\text{peak}}-L_{\text{IR}}$  relation proposed by Casey et al. (2018). Similar to the predictions from FIRE-2, at low IR luminosities, the peak wavelengths at  $z = 4$  in TNG



**Figure 11.** Dust temperature versus redshift. We show the dust temperature as a function of redshift. The TNG predictions are the median  $T_{\text{peak}}$  and  $T_{\text{sed}}$  of galaxies with  $10^{10} M_{\odot} < M_{*} < 10^{11} M_{\odot}$  in all three simulations. The error bars indicate  $1.5\sigma$  scatter. With open stars, we show the median and scatter of only TNG50 galaxies at  $z = 4$ . For comparison, we show observational constraints from Schreiber et al. (2018; including the Béthermin et al. 2015 samples), Strandet et al. (2016), Faisst et al. (2020; and  $z > 6$  galaxies in Knudsen et al. 2016; Hashimoto et al. 2019; Bakx et al. 2020 re-measured). In addition, we include constraints from Pavesi et al. (2016), Hashimoto et al. (2019), and Béthermin et al. (2020). Observations classified as peak dust temperature measurements are shown in triangles. Others are shown in circles. For reference, we show relations proposed by Schreiber et al. (2018) and Bouwens et al. (2020) in grey and black dashed lines. We show theoretical predictions from Liang et al. (2019) in the orange dashed line.

are about two times lower than the value in the local Universe. The peak wavelength anticorrelates with IR luminosity but the slope of the relation is shallower than the  $\lambda_{\text{peak}} \sim T_{\text{peak}}^{-1} \sim L_{\text{IR}}^{-1/(4+\beta)}$  law (assuming  $\beta = 2$ ) due to increasing abundance of dust in more luminous IR galaxies. When  $L_{\text{IR}}$  reaches  $\sim 10^{11} L_{\odot}$ , a plateau of  $T_{\text{peak}} \simeq 40$  K is reached. This is likely due to the increasing dust opacity in more IR luminous galaxies which makes the hot dust hidden from observation and effectively decreases the light-weighted dust temperature. This plateau feature makes the TNG prediction more consistent with the SPT-detected galaxies at  $z \sim 4$ , compared to the FIRE-2 results. The TNG predicted peak dust temperature at  $z = 6$  is higher than that at  $z = 4$  by about 20 K and is higher than the dust temperatures of galaxies observed at  $z \sim 5.5$  (Faisst et al. 2020). In addition, the scatter of the predicted dust temperature is also smaller than galaxies in observations. This is likely caused by the sparseness of the wavelength grid in SKIRT calculations, which makes the spectral peaks cluster around grid points.

#### 4.4.2 $T_{\text{dust}}$ versus redshift

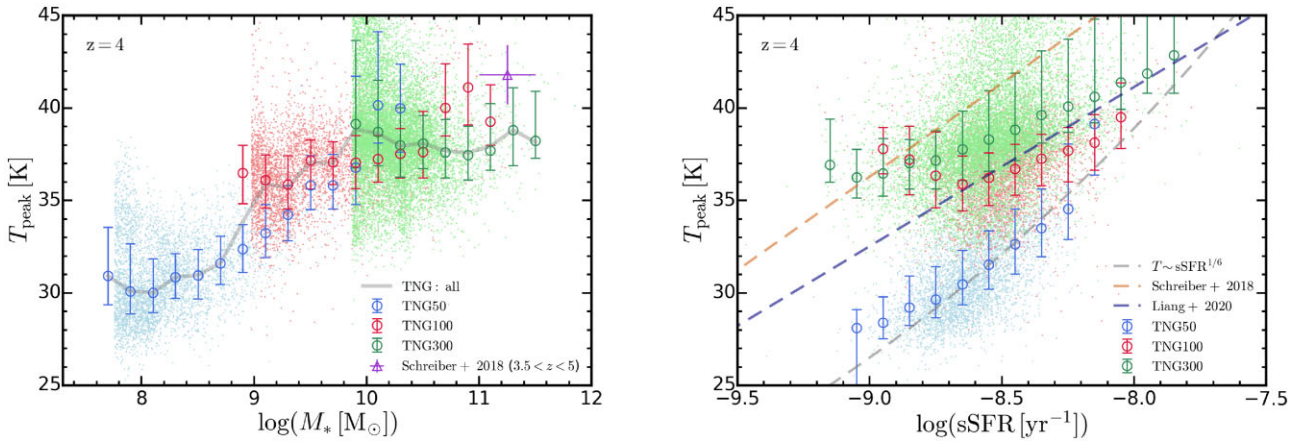
In Fig. 11, we show the dust temperature as a function of redshift. The TNG predictions are the  $T_{\text{peak}}$  and  $T_{\text{sed}}$  of galaxies with  $10^{10} M_{\odot} < M_{*} < 10^{11} M_{\odot}$  in all three simulations. The mass range roughly corresponds to galaxies with  $L_{\text{IR}} \sim 10^{12} L_{\odot}$  and is chosen to match the typical value of observed samples at high redshift (e.g. Schreiber et al. 2018; Bouwens et al. 2020; Faisst et al. 2020). Results derived purely from TNG50 galaxies at  $z = 4$  are also shown and they do not differ considerably from the results of all TNG galaxies. For comparison, we show observational constraints at low redshift from Schreiber et al. [2018; and the results based on Béthermin et al. (2015) samples], SPT-selected galaxies at  $z = 2-6$

in Strandet et al. (2016; compiled by Bouwens et al. 2020). In addition to these, we include constraints based on the ALPLINE-ALMA galaxies in Béthermin et al. (2020), the Lyman break galaxy in Pavesi et al. (2016), the four main-sequence ALMA galaxies in Faisst et al. (2020; having  $T_{\text{sed}}$  determined by a  $\beta$  free fit but also having  $T_{\text{peak}}$  measured),  $z \gtrsim 6$  galaxies from Knudsen et al. (2016), Hashimoto et al. (2019), Bakx et al. (2020) re-measured by Faisst et al. (2020), and the  $z \sim 7$  galaxy in Hashimoto et al. (2019) with their original temperature measurement. Among these observational constraints, only a set of measurements in Faisst et al. (2020) can be clearly classified as peak dust temperatures (they are marked as triangles in the Figure). Other observations all involve some level of SED fitting with various assumptions on the emissivity index  $\beta$  and fitting functions, which could result in model-dependent uncertainties. It is also worth noting that only Faisst et al. (2020) assumed the same multicomponent SED model as ours with a general description of emission opacity. Therefore, the  $T_{\text{sed}}$  we obtain here forms an apple-to-apple comparison only with their results. When compared to other observational studies listed here, one should be cautious in interpreting the differences of the fitted dust temperatures, since results can be dramatically affected by the assumptions of the SED model (see the tests in Casey et al. 2012, for example). For reference, we include the models proposed by Schreiber et al. (2018) and Bouwens et al. (2020) based on observational constraints and the  $T_{\text{peak}}-z$  relation from the FIRE-2 simulations (Liang et al. 2019).

In general, the dust temperatures in both observations and simulations are higher at high redshift than in the local Universe. This can be understood by the correlation of dust temperature and the specific star formation rate (see also Magnelli et al. 2014; Ma et al. 2019)

$$\begin{aligned}
 T_{\text{dust}} &\sim \left( \frac{L_{\text{IR}}}{M_{\text{dust}}} \right)^{1/(4+\beta)} \sim \left( \frac{\text{SFR}}{M_{\text{dust}}} \right)^{1/(4+\beta)} \\
 &\sim \left( \frac{\text{SFR}}{M_{*}} \frac{M_{*}}{M_{\text{metal}}} \frac{M_{\text{metal}}}{M_{\text{dust}}} \right)^{1/(4+\beta)} \\
 &\sim \left[ \text{sSFR} \frac{1}{M_{\text{metal}}/M_{*}} \left( \frac{1}{\text{DTM}} - 1 \right) \right]^{1/(4+\beta)}, \quad (11)
 \end{aligned}$$

where sSFR is the specific star formation rate and DTM is the DTM ratio. We have used  $L_{\text{IR}} \sim \text{SFR}$  and  $\text{DTM} \equiv M_{\text{dust}}/(M_{\text{dust}} + M_{\text{metal}})$  in the derivation above. The average sSFR increases at high redshift and reaches a plateau at  $z \sim 4$  (e.g. Tomczak et al. 2016; Santini et al. 2017). In addition, the metal-to-stellar mass ratio decreases at high redshift owing to the shift to lower mass on the mass-metallicity relation and the potentially decreasing normalization of the mass-metallicity relation at high redshift (Ma et al. 2016). Moreover, the DTM could also be lower at high redshift, as found in the calibration procedure in Paper I and other studies (e.g. Inoue 2003; McKinnon, Carilli & Beasley 2016; Aoyama et al. 2017; Behrens et al. 2018). These factors will drive the dust temperature to be higher at high redshift, which is consistent with the phenomena shown in Fig. 10. Compared to observations, the TNG predicted  $T_{\text{peak}}$  at  $z = 4$  is consistent with observations while the  $T_{\text{sed}}$  from TNG is significantly higher than both observational constraints and  $T_{\text{peak}}$ . When optically thick dust self-absorption is considered in SED fitting, it is actually expected that  $T_{\text{sed}}$  will be considerably larger than  $T_{\text{peak}}$ . For example,  $T_{\text{peak}} \simeq 40$  K could correspond to  $T_{\text{sed}} \simeq 60$  K as shown in Casey (2012) and Liang et al. (2019). At  $z \geq 6$ , the SED dust temperature from TNG is still about 20 K higher than the temperature measured in observations. At this redshift range, the peak dust temperature from TNG follows the models in Schreiber et al. (2018) and Bouwens et al. (2020), which linearly rise up at higher redshift. However, such high  $T_{\text{peak}}$  is in tension with the peak dust temperature measurements



**Figure 12.** Scaling relation of dust temperature. *Left:* dust temperature versus galaxy stellar mass. We show the galaxies from TNG50/100/300 with  $M_* \geq 1000m_b$ . The median and  $1.5\sigma$  dispersion are shown in circles with error bars for each simulation. The median relation derived based on all TNG galaxies is shown with the grey dashed line. The peak dust temperature shows a positive dependence on galaxy stellar mass at  $M_* \lesssim 10^{10} M_\odot$  while reaches a plateau at larger masses. The temperature at the massive end is consistent with observational constraints, including the measurement from stacked galaxy SEDs (Schreiber et al. 2018) shown here and other measurements at  $z \simeq 4$  shown in Fig. 11. *Right:* dust temperature versus galaxy specific star formation rate. The notation is the same as the left-hand panel. The grey dashed line shows the relation  $T \sim \text{sSFR}^{1/(4+\beta)}$  ( $\beta = 2$ ) for reference. The TNG50 galaxies and the envelope of all TNG galaxies follow the reference relation. Galaxies in TNG100/300 have systematically hotter dust than TNG50 at fixed sSFR, due to the mass dependence of dust temperature shown in the left-hand panel. The results are also compared to theoretical predictions from the FIRE simulations (Liang et al. 2020) and the observational study (Schreiber et al. 2018).

of  $z \gtrsim 5$  galaxies compiled in Faisst et al. (2020), which suggests a plateau of  $T_{\text{peak}}$  at  $z \gtrsim 4$ . This behaviour is in agreement with the FIRE-2 predictions which are based on simulations of higher spatial and mass resolution. In conclusion, these comparisons here suggest that the TNG predicted dust temperatures are higher than observations at  $z \geq 6$ .

#### 4.4.3 $T_{\text{dust}}$ versus $M_*$ and sSFR

In the left-hand panel of Fig. 12, we show the peak dust temperature as a function of galaxy stellar mass at  $z = 4$ . The binned estimations derived from TNG50/100/300 are shown explicitly along with scatter plot of individual galaxies. For each simulation, we only select galaxies with  $M_* > 1000m_b$  for analysis. In general, the dust temperature increases with galaxy stellar mass when  $M_* \lesssim 10^{10} M_\odot$  and reaches a plateau when  $M_* \gtrsim 10^{10} M_\odot$ . The plateau feature is similar to what we found in the  $T_{\text{peak}}-L_{\text{IR}}$  relation and is consistent with the weak dependence of  $T_{\text{peak}}$  on  $M_*$  of  $M_* \gtrsim 10^9 M_\odot$  galaxies found in Liang et al. (2020). For massive galaxies, the predicted dust temperature is roughly consistent with the stacked SED measurement of stellar mass-binned samples in Schreiber et al. (2018) and other measurements at  $z \simeq 4$  shown in Fig. 11. Results from different simulations are consistent with each other in their shared dynamical ranges, but simulation with poorer resolution always exhibits larger scatter and slight shift towards higher temperature. This marks the potential overprediction of dust temperature in galaxies with poor sampling of radiation sources (and dust attenuating gas cells as well). In the right-hand panel of Fig. 12, we show the peak dust temperature as a function of galaxy sSFR. For reference, the grey dashed line shows the relation  $T \sim \text{sSFR}^{1/(4+\beta)}$  ( $\beta = 2$ ) as in equation (11). Although TNG galaxies show significant dispersion in this plane, the low temperature envelope of the distribution roughly follows the reference relation. TNG50 galaxies follow the reference relation while TNG100/300 galaxies have systematically hotter dust component at fixed sSFR. This is related to the mass dependence of dust temperature shown in the left-hand panel. We compare

our results with relations reported in the observational (Schreiber et al. 2018) and theoretical studies (Liang et al. 2020). Liang et al. (2020) found the correlation between  $\Delta T_{\text{dust}} \equiv T_{\text{dust}} - T_{\text{dust}}^{\text{med}}$  and the star formation burstiness  $SB \equiv \text{sSFR} - \text{sSFR}^{\text{med}}$ , where ‘med’ indicates the median value of their sample. We convert the relation to the  $T_{\text{dust}}$ -sSFR plane using the reported median values therein. Similarly, Schreiber et al. (2018) found the  $\Delta T_{\text{dust}}-SB$  correlation but the reference points are the main sequence dust temperature and sSFR, for which we use the  $T_{\text{dust}}^{\text{MS}}(z)$  therein and  $\log \text{sSFR}^{\text{MS}}(z = 4) \simeq -8.5$ , which is consistent with observational constraints (e.g. Salmon et al. 2015; Tomczak et al. 2016; Santini et al. 2017), as well as TNG galaxies shown here. All of the sampled galaxies in Liang et al. (2020) at  $z = 4$  have  $M_* \gtrsim 10^9 M_\odot$ . As expected, their relation is quite consistent with TNG100/300 galaxies with comparable stellar masses. The Schreiber et al. (2018) samples are even more massive galaxies ( $M_* \gtrsim 10^{10} M_\odot$ ). Their relation lies slightly above the TNG300 predictions and is consistent with the same level of difference in Fig. 11.

#### 4.4.4 Discussions

Equation (11) indicates that the dust temperature is related to the dust mass when  $L_{\text{IR}}$  is fixed. Since most of the theoretical works have dust models calibrated based on UV luminosities of galaxies, the total amount of energy absorbed and re-emitted at IR wavelengths should match observations by construction, regardless of the dust model adopted. However, the dust temperature offers another channel to test theoretical predictions since it anti-correlates with the actual amount of dust predicted when  $L_{\text{UV}}^{\text{abs}} \sim L_{\text{IR}}$  is fixed. Here, TNG predicts higher peak dust temperatures than some observations at  $z \geq 6$ , as well as results from simulations of higher resolution indicating that the dust mass could be underpredicted. In Vogelsberger et al. (2020), we found that the average DTM ratio follows a  $(z/2)^{-1.92}$  law at  $z \geq 2$ , which implies  $\text{DTM} \lesssim 0.1$  at  $z \gtrsim 6$  much lower than the value in the local Universe  $\sim 0.4$  (Dwek 1998). Such a low predicted dust abundance could be attributed to the limited spatial and mass resolution of high-

redshift galaxies in the TNG simulations. As discussed in Cen & Kimm (2014), if the porosity (and clumpiness) of the ISM is not well-resolved, the dust opacity of radiation sources can be overestimated. In the calibration procedure, less dust will be required to produce the same amount of attenuation. Cen & Kimm (2014) has artificially allowed a fraction  $f_{\text{esc}}$  of the intrinsic radiation leak from their  $z \sim 7$  galaxies to correct for this effect and the best-fitting DTM rose to 0.4 with  $f_{\text{esc}} = 0.1$ , as opposed to the DTM of 0.06 they found originally. Another potential caveat of the dust model is that the DTM is set to a constant for all the galaxies at a given redshift. The dust abundance of a small sub-population of massive star-forming galaxies, which are completely dark in UV/NIR searches, could be heavily under-estimated in this calibration process. This could serve as a potential explanation to the missing IR bright galaxies in TNG without involving the galaxy formation physics models of the simulations.

In the meantime, if the dust-attenuating medium is not well-resolved, the optical depth of the hot dust component heated at the vicinity of radiation sources could be underestimated (i.e. a single layer of gas cells covering a radiation source are heated while their dust emission are not shielded at all), which will result in higher light-weighted dust temperatures as well. Such effect appears in the left-hand panel of Fig. 12. At similar stellar mass, the result from low-resolution simulation exhibits larger scatter and systematic shift towards higher temperature. For each simulation, the scatter is also larger at smaller stellar mass and, in a small fraction of poorly resolved galaxies (at the edge of our stellar mass cut), dust can be heated to  $\gtrsim 3\sigma$  higher than the median temperature.

The measurements or assumptions of dust temperature could also affect the estimation of the bolometric IR luminosity. For example, if the observational data only cover the Rayleigh–Jeans tail (long wavelengths) of the IR SED, the inferred bolometric luminosity will be sensitive to the dust temperature assumed. Given the same flux measurement at long wavelengths, the higher the dust temperature assumed, the higher the bolometric luminosity will be determined. However, in terms of the discrepancy, we found for the IR luminosity function in Section 4.3.1, this effect would actually exaggerate the mismatch at the bright end, since the observational studies typically assume lower dust temperatures than the simulation results (i.e. increasing the dust temperature will lead to even higher inferred bolometric IR luminosities and make it even more challenging for the TNG simulations to match).

## 5 CONCLUSIONS

In this paper, we have expanded the theoretical predictions of high-redshift galaxies from IllustrisTNG to IR wavelengths, using SKIRT radiative transfer calculations. The analysis pipeline has been adapted from Paper I to self-consistently model dust emission and self-absorption at IR wavelengths. The pipeline produces a catalog of high-redshift galaxies with their NIR-to-FIR SEDs calculated in this paper, along with their UV-optical high resolution SEDs calculated in Paper I. Based on this, we make various NIR and FIR predictions of galaxy populations at  $z \geq 4$ . Our findings can be summarized as follows:

(i) The predicted rest-frame  $K$ -band and  $z$ -band luminosity functions at  $z \geq 4$  are presented and compared with existing observations. The predictions are consistent with observations, despite a slight underprediction at the bright end of the  $z = 4$   $K$ -band and  $z$ -band luminosity functions. The luminosity functions in both bands are tracers of galaxy stellar mass assembly. The faint-end luminosity

functions evolve in a self-similar way at  $z \geq 4$  with a roughly constant faint-end slope. This indicates that the star-forming galaxies at this epoch have roughly the same mass doubling time.

(ii) Assuming 100 per cent survey completeness, we make theoretical predictions for the *JWST* MIRI apparent band luminosity functions and number counts. At  $z = 6$ ,  $\sim 3000$  ( $\sim 500$ ) galaxies are expected in the F560W (F1000W) band assuming a survey area of 500 arcmin<sup>2</sup> and depth of  $\Delta z = 1$ . Large NIR galaxy surveys conducted with MIRI can be about 2 mag deeper than the current best observations.

(iii) We make predictions for the bolometric IR luminosity functions at  $z \geq 4$ . The results are mainly affected by the predicted FIR dust continuum emission. A better agreement with observations is achieved at the faint end compared to previous theoretical attempts. However, the abundance of most luminous IR galaxies ( $L_{\text{IR}} \sim 10^{13} L_{\odot}$ ) is significantly underpredicted ( $\sim 1$  dex deficiency) by TNG. The discrepancy consistently shows up at  $z = 4$  and  $z = 6$ , and is potentially related with the underpredicted counts of sub-millimeter galaxies reported in previous works based on TNG (Hayward et al. 2021). The tension cannot be resolved by considering merging systems in TNG.

(iv) By integrating over the bolometric IR luminosity function, we make predictions for the obscured cosmic SFRD at  $z \geq 4$ . Combining the rest-frame UV luminosity function in Paper I and the  $H\alpha$  luminosity function in Paper II, we are able to compare the cosmic SFRD traced by different indicators. The predicted unobscured SFRD (traced by UV and  $H\alpha$ ) is consistent with observations. The total SFRD derived by summing up  $\text{SFRD}_{\text{UV}}$  and  $\text{SFRD}_{\text{IR}}$  agrees remarkably well with the instantaneous SFRD traced by gas cells in simulations. The obscured SFRD predicted from TNG suggests that it becomes sub-dominant (contributes less than 50 per cent to the total SFRD) at  $z \gtrsim 5$  and diminishes at higher redshift. Such a prediction is in tension with the recent ALPINE-ALMA survey which suggests a significant contribution of unobscured SFRD at  $4 \lesssim z \lesssim 6$ .

(v) Specifically, we check the SFRD contributed by the most obscured and thus most luminous IR galaxies ( $L_{\text{IR}} \gtrsim 10^{12} L_{\odot}$ ) in simulations and compare it to the SFRD of optical/NIR dark galaxies revealed in IR/sub-millimeter observations. The SFRD in such galaxies in simulations is about  $1\sigma$  lower than the results from the ALPINE-ALMA survey.

(vi) We make predictions for the dust temperature of high-redshift galaxies. We find that the dust temperature positively correlates with both galaxy bolometric IR luminosity and redshift. The predicted peak dust temperature in typical IR galaxies is about 40 K (60 K) at  $z = 4$  ( $z = 6$ ). The SED dust temperature is systematically higher than  $T_{\text{peak}}$  by about 20 K. The predicted  $T_{\text{peak}}-L_{\text{IR}}$  relation is shifted from the local relation to higher temperatures while remaining consistent with the sub-millimeter galaxies selected at  $z \simeq 4$ . However, at  $z \geq 6$ , we find that TNG overpredicts the peak dust temperature of galaxies by about 20 K. This overprediction of dust temperature could be related to the low DTM ratio of our model, and could be attributed to the limitation of the simulations in resolving the porosity and clumpiness of the ISM.

In conclusion, similar to the previous comparisons in rest-frame UV/optical in Paper I and Paper II, the NIR properties of high-redshift galaxies in TNG are quite consistent with observations. Predictions for the *JWST* MIRI instrument are calculated, which leads to a complete *JWST* galaxy multiband photometric catalog when combined with the NIRCам predictions in Paper I. However, we find large, systematic discrepancies of FIR properties of most

luminous IR galaxies in TNG compared to recent observations. The abundance of luminous IR galaxies and their contribution to the cosmic SFRD are underpredicted by TNG. The solution to this would require both higher intrinsic on-going star formation and stronger metal and dust enrichment in high-redshift galaxies. The discrepancy found here could serve as a constraint on the sub-grid feedback model of cosmological simulations.

## ACKNOWLEDGEMENTS

MV acknowledges support through the National Aeronautics and Space Administration (NASA) Astrophysics Theory Program (ATP) grants 16-ATP16-0167, 19-ATP19-0019, 19-ATP19-0020, 19-ATP19-0167, and the National Science Foundation (NSF) grants AST-1814053, AST-1814259, AST-1909831, and AST-2007355. ST is supported by the Smithsonian Astrophysical Observatory through the Center for Astrophysics (CfA) Fellowship. PT acknowledges support from NSF grant AST-2008490. FM acknowledges support through the Program ‘Rita Levi Montalcini’ of the Italian MUR. The primary TNG simulations were realized with compute time granted by the Gauss Centre for Super-computing (GCS): TNG50 under GCS Large-Scale Project GCS-DWAR (2016; PIs Nelson/Pillepich), and TNG100 and TNG300 under GCS-ILLU (2014; PI Springel) on the GCS share of the supercomputer Hazel Hen at the High Performance Computing Center Stuttgart (HLRS).

## DATA AVAILABILITY

The simulation data of the IllustrisTNG project is publicly available at <https://www.tng-project.org/data/>. The analysis data of this work was generated and stored on the super-computing system Cannon at Harvard University. The data underlying this article can be shared on a reasonable request to the corresponding author.

## REFERENCES

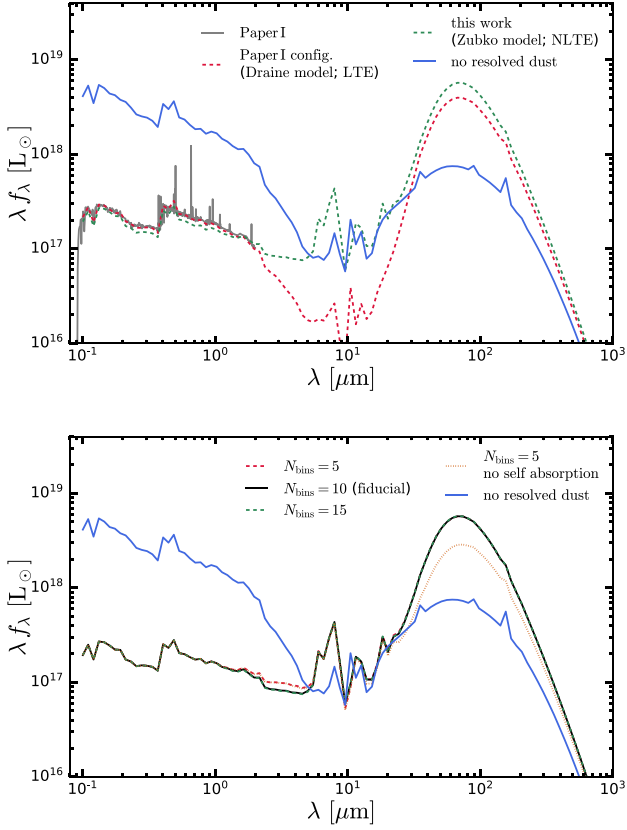
- Aoyama S., Hou K.-C., Shimizu I., Hirashita H., Todoroki K., Choi J.-H., Nagamine K., 2017, *MNRAS*, 466, 105
- Ashby M. L. N. et al., 2013, *ApJ*, 769, 80
- Ashby M. L. N. et al., 2015, *ApJS*, 218, 33
- Baes M., Verstaappen J., De Looze I., Fritz J., Saftly W., Vidal Pérez E., Stalevski M., Valcke S., 2011, *ApJS*, 196, 22
- Bakx T. J. L. C. et al., 2020, *MNRAS*, 493, 4294
- Battersby C. et al., 2018, *Nature Astron.*, 2, 596
- Behrens C., Pallottini A., Ferrara A., Gallerani S., Vallini L., 2018, *MNRAS*, 477, 552
- Béthermin M. et al., 2015, *A&A*, 573, A113
- Béthermin M. et al., 2020, *A&A*, 643, A2
- Blain A. W., Barnard V. E., Chapman S. C., 2003, *MNRAS*, 338, 733
- Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, *Nature*, 311, 517
- Bouwens R. et al., 2020, *ApJ*, 902, 112
- Bouwens R. J. et al., 2003, *ApJ*, 595, 589
- Bouwens R. J., Stefanon M., Oesch P. A., Illingworth G. D., Nanayakkara T., Roberts-Borsani G., Labbé I., Smit R., 2019, *ApJ*, 880, 25
- Bryan S. et al., 2018, in Zmuidzinas J., Gao J.-R., eds, *SPIE Conf. Ser. Vol. 10708, Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy IX*. SPIE, Bellingham, p. 107080J
- Camps P., Baes M., 2015, *Astron. Comput.*, 9, 20
- Camps P., Baes M., Saftly W., 2013, *A&A*, 560, A35
- Camps P., Trayford J. W., Baes M., Theuns T., Schaller M., Schaye J., 2016, *MNRAS*, 462, 1057
- Caputi K. I., McLure R. J., Dunlop J. S., Cirasuolo M., Schael A. M., 2006, *MNRAS*, 366, 609
- Casey C. M. et al., 2012, *ApJ*, 761, 140
- Casey C. M. et al., 2018, *ApJ*, 862, 77
- Casey C. M., 2012, *MNRAS*, 425, 3094
- Cen R., Kimm T., 2014, *ApJ*, 782, 32
- Chabrier G., 2003, *PASP*, 115, 763
- Chakrabarti S., Fenner Y., Cox T. J., Hernquist L., Whitney B. A., 2008, *ApJ*, 688, 972
- Cirasuolo M., McLure R. J., Dunlop J. S., Almaini O., Foucaud S., Simpson C., 2010, *MNRAS*, 401, 1166
- Clay S. J., Thomas P. A., Wilkins S. M., Henriques B. M. B., 2015, *MNRAS*, 451, 2692
- Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, *MNRAS*, 319, 168
- Conroy C., Gunn J. E., 2010, *ApJ*, 712, 833
- Conroy C., Gunn J. E., White M., 2009, *ApJ*, 699, 486
- Cowley W. I., Baugh C. M., Cole S., Frenk C. S., Lacey C. G., 2018, *MNRAS*, 474, 2352
- Crain R. A. et al., 2015, *MNRAS*, 450, 1937
- Cullen F., McLure R. J., Khochfar S., Dunlop J. S., Dalla Vecchia C., 2017, *MNRAS*, 470, 3006
- da Cunha E. et al., 2013, *ApJ*, 766, 13
- da Cunha E. et al., 2015, *ApJ*, 806, 110
- Dayal P., Ferrara A., 2018, *Phys. Rep.*, 780, 1
- Dolag K., Borgani S., Murante G., Springel V., 2009, *MNRAS*, 399, 497
- Draine B. T. et al., 2007, *ApJ*, 663, 866
- Drory N., Bender R., Feulner G., Hopp U., Maraston C., Snigula J., Hill G. J., 2003, *ApJ*, 595, 698
- Dudzevičiūtė U. et al., 2020, *MNRAS*, 494, 3828
- Dunne L., Eales S., Edmunds M., Ivison R., Alexander P., Clements D. L., 2000, *MNRAS*, 315, 115
- Dwek E., 1998, *ApJ*, 501, 643
- Engel H. et al., 2010, *ApJ*, 724, 233
- Faisst A. L., Fudamoto Y., Oesch P. A., Scoville N., Riechers D. A., Pavesi R., Capak P., 2020, *MNRAS*, 498, 4192
- Fazio G. G. et al., 2004, *ApJS*, 154, 10
- Finkelstein S. L. et al., 2015, *ApJ*, 810, 71
- Franco M. et al., 2020, *A&A*, 643, A30
- Gardner J. P. et al., 2006, *Space Sci. Rev.*, 123, 485
- Genel S. et al., 2014, *MNRAS*, 445, 175
- Genel S. et al., 2018, *MNRAS*, 474, 3976
- Golwala S., 2018, The Chajnantor Sub/Millimeter Survey Telescope, Atacama Large-Aperture Submm/mm Telescope (AtLAST). Zenodo, p. 46
- Groves B., Dopita M. A., Sutherland R. S., Kewley L. J., Fischera J., Leitherer C., Brandl B., van Breugel W., 2008, *ApJS*, 176, 438
- Gruppioni C. et al., 2013, *MNRAS*, 432, 23
- Gruppioni C. et al., 2020, *A&A*, 643, A8
- Hashimoto T. et al., 2019, *PASJ*, 71, 71
- Hayward C. C. et al., 2021, *MNRAS*, 502, 2922
- Hayward C. C., Kereš D., Jonsson P., Narayanan D., Cox T. J., Hernquist L., 2011, *ApJ*, 743, 159
- Hopkins P. F. et al., 2018, *MNRAS*, 480, 800
- Inoue A. K., 2003, *PASJ*, 55, 901
- Kennicutt Robert C. J., 1998, *ARA&A*, 36, 189
- Khusanova Y. et al., 2021, *A&A*, 649, A152
- Kimm T., Cen R., 2014, *ApJ*, 788, 121
- Kimm T., Katz H., Haehnelt M., Rosdahl J., Devriendt J., Slyz A., 2017, *MNRAS*, 466, 4826
- Knudsen K. K., Richard J., Kneib J.-P., Jauzac M., Clément B., Drouart G., Egami E., Lindroos L., 2016, *MNRAS*, 462, L6
- Koprowski M. P., Dunlop J. S., Michałowski M. J., Coppin K. E. K., Geach J. E., McLure R. J., Scott D., van der Werf P. P., 2017, *MNRAS*, 471, 4155
- Lee N. et al., 2013, *ApJ*, 778, 131
- Liang L. et al., 2019, *MNRAS*, 489, 1397
- Liang L., Feldmann R., Hayward C. C., Narayanan D., Çatmabacak O., Kereš D., Faucher-Giguère C.-A., Hopkins P. F., 2021, *MNRAS*, 502, 3210
- Lim S., Scott D., Babul A., Barnes D. J., Kay S. T., McCarthy I. G., Rennehan D., Vogelsberger M., 2021, *MNRAS*, 501, 1803
- Liu C., Mutch S. J., Angel P. W., Duffy A. R., Geil P. M., Poole G. B., Mesinger A., Wyithe J. S. B., 2016, *MNRAS*, 462, 235

- Loiacono F. et al., 2021, *A&A*, 646, A76
- Ma X. et al., 2018, *MNRAS*, 478, 1694
- Ma X. et al., 2019, *MNRAS*, 487, 1844
- Ma X., Hopkins P. F., Faucher-Giguère C.-A., Zolman N., Muratov A. L., Kereš D., Quataert E., 2016, *MNRAS*, 456, 2140
- Madau P., 1995, *ApJ*, 441, 18
- Madau P., Dickinson M., 2014, *ARA&A*, 52, 415
- Madau P., Ferguson H. C., Dickinson M. E., Giavalisco M., Steidel C. C., Fruchter A., 1996, *MNRAS*, 283, 1388
- Magnelli B. et al., 2012, *A&A*, 539, A155
- Magnelli B. et al., 2013, *A&A*, 553, A132
- Magnelli B. et al., 2014, *A&A*, 561, A86
- Magnelli B., Elbaz D., Chary R. R., Dickinson M., Le Borgne D., Frayer D. T., Willmer C. N. A., 2011, *A&A*, 528, A35
- Maniyar A. S., Béthermin M., Lagache G., 2018, *A&A*, 614, A39
- Marinacci F. et al., 2018, *MNRAS*, 480, 5113
- McKinnon M., Carilli C., Beasley T., 2016, in Hall H. J., Gilmozzi R., Marshall H. K., eds, *The next generation very large array, Ground-based and Airborne Telescopes VI. Proceedings of the SPIE*, p. 990627
- McLure R. J. et al., 2013, *MNRAS*, 432, 2696
- Miettinen O. et al., 2017, *A&A*, 606, A17
- Mortlock A., McLure R. J., Bowler R. A. A., McLeod D. J., Mármol-Queraltó E., Parsa S., Dunlop J. S., Bruce V. A., 2017, *MNRAS*, 465, 672
- Murphy E. J. et al., 2011, *ApJ*, 737, 67
- Naiman J. P. et al., 2018, *MNRAS*, 477, 1206
- Narayanan D., Hayward C. C., Cox T. J., Hernquist L., Jonsson P., Younger J. D., Groves B., 2010, *MNRAS*, 401, 1613
- Narayanan D., Davé R., Johnson B. D., Thompson R., Conroy C., Geach J., 2018, *MNRAS*, 474, 1718
- Nelson D. et al., 2015, *Astron. Comput.*, 13, 12
- Nelson D. et al., 2018, *MNRAS*, 475, 624
- Nelson D. et al., 2019a, *Comput. Astrophys. Cosmol.*, 6, 2
- Nelson D. et al., 2019b, *MNRAS*, 490, 3234
- Ocvirk P. et al., 2016, *MNRAS*, 463, 1462
- Oesch P. A., Bouwens R. J., Illingworth G. D., Labbé I., Stefanon M., 2018, *ApJ*, 855, 105
- Pakmor R., Springel V., Bauer A., Mocz P., Munoz D. J., Ohlmann S. T., Schaal K., Zhu C., 2016, *MNRAS*, 455, 1134
- Pavesi R. et al., 2016, *ApJ*, 832, 151
- Pillepich A. et al., 2018a, *MNRAS*, 473, 4077
- Pillepich A. et al., 2018b, *MNRAS*, 475, 648
- Pillepich A. et al., 2019, *MNRAS*, 490, 3196
- Planck Collaboration, Planck 2015 results. XIII. Cosmological parameters 2016, *A&A*, 594, A13
- Planck Collaboration, Planck 2018 results. VI. Cosmological parameters 2020, *A&A*, 641, A6
- Pozzetti L. et al., 2003, *A&A*, 402, 837
- Rieke G. H. et al., 2015, *PASP*, 127, 584
- Rosdahl J. et al., 2018, *MNRAS*, 479, 994
- Rosdahl J., Blaizot J., Aubert D., Stranex T., Teyssier R., 2013, *MNRAS*, 436, 2188
- Safarzadeh M., Hayward C. C., Ferguson H. C., 2017, *ApJ*, 840, 15
- Saftly W., Baes M., Camps P., 2014, *A&A*, 561, A77
- Salmon B. et al., 2015, *ApJ*, 799, 183
- Santini P. et al., 2017, *ApJ*, 847, 76
- Saracco P. et al., 2006, *MNRAS*, 367, 349
- Schaye J. et al., 2015, *MNRAS*, 446, 521
- Schechter P., 1976, *ApJ*, 203, 297
- Schreiber C., Elbaz D., Pannella M., Ciesla L., Wang T., Franco M., 2018, *A&A*, 609, A30
- Schulz S., Popping G., Pillepich A., Nelson D., Vogelsberger M., Marinacci F., Hernquist L., 2020, *MNRAS*, 497, 4773
- Shapley A. E., 2011, *ARA&A*, 49, 525
- Shen X. et al., 2020, *MNRAS*, 495, 4747
- Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, *MNRAS*, 452, 575
- Simpson J. M. et al., 2014, *ApJ*, 788, 125
- Skrutskie M. F. et al., 2006, *AJ*, 131, 1163
- Springel V. et al., 2018, *MNRAS*, 475, 676
- Springel V., 2010, *MNRAS*, 401, 791
- Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, *MNRAS*, 328, 726
- Stark D. P., 2016, *ARA&A*, 54, 761
- Stefanon M., Bouwens R. J., Labbé I., Muzzin A., Marchesini D., Oesch P., Gonzalez V., 2017, *ApJ*, 843, 36
- Steinhardt C. L. et al., 2014, *ApJ*, 791, L25
- Strandet M. L. et al., 2016, *ApJ*, 822, 80
- Tacchella S., Bose S., Conroy C., Eisenstein D. J., Johnson B. D., 2018, *ApJ*, 868, 92
- Talia M., Cimatti A., Giuliatti M., Zamorani G., Béthermin M., Faisst A., Le Fèvre O., Smolčić V., 2021, *ApJ*, 909, 23
- Tomczak A. R. et al., 2016, *ApJ*, 817, 118
- Torrey P., Vogelsberger M., Genel S., Sijacki D., Springel V., Hernquist L., 2014, *MNRAS*, 438, 1985
- Trayford J. W. et al., 2017, *MNRAS*, 470, 771
- Valiante E. et al., 2016, *MNRAS*, 462, 3146
- Vogelsberger M. et al., 2014a, *MNRAS*, 444, 1518
- Vogelsberger M. et al., 2014b, *Nature*, 509, 177
- Vogelsberger M. et al., 2020, *MNRAS*, 492, 5167
- Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, *MNRAS*, 436, 3031
- Wang T. et al., 2019, *Nature*, 572, 211
- Weinberger R. et al., 2017, *MNRAS*, 465, 3291
- Weinberger R., Springel V., Pakmor R., 2020, *ApJS*, 248, 32
- Weingartner J. C., Draine B. T., 2001, *ApJ*, 548, 296
- Werner M. W. et al., 2004, *ApJS*, 154, 1
- White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
- Wilkins S. M., Bunker A. J., Ellis R. S., Stark D., Stanway E. R., Chiu K., Lorenzoni S., Jarvis M. J., 2010, *MNRAS*, 403, 938
- Wilkins S. M., Feng Y., Di Matteo T., Croft R., Lovell C. C., Waters D., 2017, *MNRAS*, 469, 2517
- Williams C. C. et al., 2019, *ApJ*, 884, 154
- Wright G. S. et al., 2015, *PASP*, 127, 595
- Yamaguchi Y. et al., 2019, *ApJ*, 878, 73
- Yan L. et al., 2020, *ApJ*, 905, 147
- Yung L. Y. A., Somerville R. S., Finkelstein S. L., Popping G., Davé R., 2019a, *MNRAS*, 483, 2983
- Yung L. Y. A., Somerville R. S., Popping G., Finkelstein S. L., Ferguson H. C., Davé R., 2019b, *MNRAS*, 490, 2855
- Zubko V., Dwek E., Arendt R. G., 2004, *ApJS*, 152, 211

## APPENDIX A: TEST WITH DIFFERENT SKIRT CONFIGURATIONS

Compared to the procedure in Paper I, the dust model and several parameters in the SKIRT calculations have been modified for the predictions in the IR. In this section, we will test how these modifications impact the resulting galaxy SED and discuss our parameter choices.

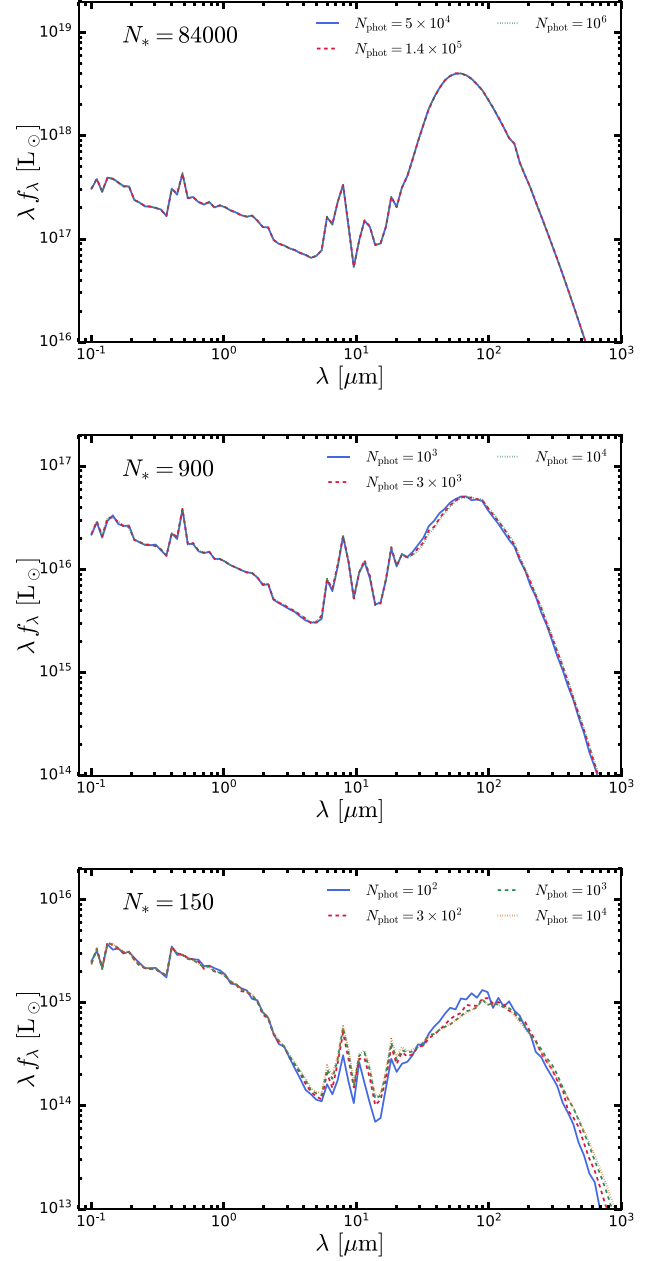
The first difference in the SKIRT set-up is that we have switched to the Zubko et al. (2004) multigrain dust mixture (from the average dust property of the Draine et al. (2007) dust mixture), in order to trace grains of small sizes independently and consider their decoupling from local thermal equilibrium. In the top panel of Fig. A1, we show that the main impact of this switch on galaxy SEDs is the enhancement of flux in the mid-IR and FIR. Compared to the UV-to-optical SED produced in Paper I, the attenuation in the UV is slightly stronger, so we decide to decrease the DTM ratio to compensate this effect and maintain consistency of the UV predictions with observations. In the bottom panel of Fig. A1, we show SEDs calculated if we vary the set-up (wavelength grid, dust model, dust self-absorption, and number of bins of grain sizes) from the fiducial one introduced in the main text (see Section 3). The



**Figure A1.** Galaxy SED calculated by different SKIRT configurations. *Top:* we show the SEDs of a well-resolved galaxy (with  $\sim 1.4 \times 10^5$  stellar particles) at  $z = 4$  in TNG100. The UV-to-NIR SED calculated in Paper I is shown in the grey solid line. The SED without resolved dust attenuation is shown in blue as a reference. The SED (that extends to IR wavelengths) calculated with the original dust model in Paper I, which takes the average dust properties of the Draine et al. (2007) dust mixture, is shown in red. The SED calculated with the Zubko et al. (2004) multigrain dust model adopted in this work, including the non-local thermal equilibrium of small grains, is shown in green. The major difference between the new model and the old one is the enhanced flux at mid-IR and FIR wavelengths. The UV-to-optical attenuation has also been enhanced due to the model switch, which has been compensated by tuning the DTM ratio as discussed in the main text. *Bottom:* we show the SEDs of the same galaxy as the top panel but with variations in SKIRT configurations from the fiducial setup. The fiducial results with number of grain size bins  $N_{\text{bins}} = 10$  is shown in black. The SED calculated with  $N_{\text{bins}} = 5$  ( $N_{\text{bins}} = 15$ ) is shown in red (green). The comparison demonstrates the convergence of the SED when  $N_{\text{bins}} \geq 10$ . In addition, we show the SED calculated without dust self-absorption and the major change induced by the self-absorption is the enhanced flux at the FIR peak.

number of bins for dust grain sizes has impact on the mid-IR SED but the results converge when  $N_{\text{bins}} \geq 10$ . The dust self-absorption is important for the prediction at the FIR peak and the flux there can be enhanced by  $\sim 0.3$  dex compared to the one without self-absorption.

In addition, due to various changes in the SKIRT set-up (see Section 3), the requirement on the number of photon packages,  $N_{\text{phot}}$ , to reach convergence may vary from the one in Paper I. To test this, we choose several galaxies with different numbers of stellar particles from TNG100 and perform SKIRT calculations with different  $N_{\text{phot}}$ . In Fig. A2, we show the results of three galaxies as examples and we note that the behaviour of all tested galaxies is similar. For the well-resolved galaxy with  $N_* = 84\,000$ , the SED is converged with  $N_{\text{phot}} = 50\,000$ , which is smaller than  $N_*$ . For the galaxy with  $N_* = 900$ ,



**Figure A2.** Galaxy SEDs calculated with different number of photon packages. For presentation, we pick three galaxies selected at  $z = 4$  in TNG100 as examples. *Top:* a well-resolved galaxy with  $N_* = 84\,000$  in the input of SKIRT calculation. The SED is converged with  $N_{\text{phot}} = 50\,000 < N_*$ . *Middle:* a resolved galaxy with  $N_* = 900$  and the SED is converged with  $N_{\text{phot}} = 1000 \sim N_*$ , despite tiny differences at  $\lambda \sim 40 \mu\text{m}$ . *Bottom:* a poorly resolved galaxy with  $N_* = 150$ . The SED with  $N_{\text{phot}} = 100$  underpredicts the flux in the mid-IR and overpredicts the flux at the rising edge of the FIR peak (by  $\lesssim 0.1$  dex), but the SED is converged with  $N_{\text{phot}} \geq 300 \sim 2 \times N_*$ .

the SED is also converged with  $N_{\text{phot}} \gtrsim N_*$ , despite tiny differences at  $\lambda \sim 40 \mu\text{m}$ . For the poorly resolved galaxy with  $N_* = 150$ , the SED is converged with  $N_{\text{phot}} \gtrsim 2 \times N_*$  while having a steeper rising edge of the FIR peak when  $N_{\text{phot}} = 100$ . As discussed in the main text, we set  $N_{\text{phot}} = N_*$  for radiative transfer calculations of galaxies in TNG100 and TNG300. For low mass, poorly resolved galaxies in these two simulations, the uncertainties from the SKIRT calculations and from the unresolved physical processes of the simulations are

degenerate, but these uncertainties can be captured to first order by the resolution correction procedure with TNG50 as a reference. Meanwhile, we choose  $N_{\text{phot}} = 3 \times N_*$  for TNG50 galaxies to ensure the convergence of the SEDs of poorly resolved galaxies in this simulation, which does not have a simulation in the TNG suite with higher resolution for correction. The comparisons shown in Fig. A2 justify our parameter choices in this work.

## APPENDIX B: BEST-FITTING SCHECHTER FUNCTION PARAMETERS

In Table B1, we present the best-fitting Schechter function parameters for the rest-frame FUV/bolometric IR luminosity functions and the *JWST* MIRI band apparent luminosity functions. The fitting functions have been described in Section 4.2. The FUV results are taken from Paper I.

**Table B1.** Parameters of the best-fitting Schechter functions to the luminosity functions. The table contains the best-fitting faint-end slope  $\alpha$ , number density normalization  $\phi^*$ , and break luminosity  $L^*/\text{break magnitude } M^*$  for the rest-frame FUV/bolometric IR luminosity functions and the *JWST* MIRI band apparent luminosity functions from the TNG simulations.

Band	Redshift	$\alpha$	$M^*$ (mag)	$\log \phi^*$ [1/Mpc <sup>3</sup> /mag]
FUV (rest-frame)	2	−1.58	−20.45	−2.65
	4	−1.80	−21.10	−3.04
	6	−2.04	−21.31	−3.61
	8	−2.45	−21.44	−4.71
Band	redshift	$\alpha$	$m^*$ (mag)	$\log \phi^*$ [1/Mpc <sup>3</sup> /mag]
F560W	4	−1.91	22.50	−4.01
	6	−1.80	24.53	−3.89
	8	−2.20	25.27	−5.00
F1000W	4	−1.77	22.50	−3.75
	6	−1.77	23.91	−4.15
	8	−2.28	24.39	−5.38
F1280W	4	−1.73	22.61	−3.71
	6	−1.67	23.71	−4.08
	8	−1.91	24.83	−5.12
F1500W	4	−1.70	22.30	−3.77
	6	−1.68	23.53	−4.25
	8	−2.10	24.21	−5.51
Band	redshift	$\alpha$	$\log L^*$ ( $L_\odot$ )	$\log \phi^*$ (1/Mpc <sup>3</sup> /dex)
Bolometric IR (rest-frame)	4	−1.73	12.25	−4.35
	6	−1.61	12.08	−4.70
	8	−1.79	12	−5.95

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.