



# CONSTRAINING THE EPOCH OF REIONIZATION FROM THE OBSERVED PROPERTIES OF THE HIGH- $z$ UNIVERSE

EDUARD SALVADOR-SOLÉ<sup>1</sup>, ALBERTO MANRIQUE<sup>1</sup>, RAFAEL GUZMAN<sup>1,2</sup>, JOSÉ MIGUEL RODRÍGUEZ ESPINOSA<sup>3</sup>, JESÚS GALLEGÓ<sup>4</sup>, ARTEMIO HERRERO<sup>3</sup>, J. MIGUEL MAS-HESSE<sup>5</sup>, AND ANTONIO MARÍN FRANCH<sup>6</sup>

<sup>1</sup> Institut de Ciències del Cosmos, Universitat de Barcelona (UB-IEEC), E-08028 Barcelona, Spain; e.salvador@ub.edu

<sup>2</sup> Dept. of Astronomy, University of Florida, USA

<sup>3</sup> Instituto de Astrofísica de Canarias y Depto. de Astrofísica, Universidad de La Laguna, E-38205 La Laguna, Spain

<sup>4</sup> Depto. de Astrofísica, Facultad de Físicas, Universidad Complutense de Madrid, E-28040 Madrid, Spain

<sup>5</sup> Centro de Astrobiología—Depto. de Astrofísica (CSIC-INTA), E-28850 Madrid, Spain

<sup>6</sup> Centro de Estudios de Física del Cosmos de Aragón, E-44001 Teruel, Spain

Received 2016 March 23; revised 2016 October 14; accepted 2016 November 5; published 2016 December 29

## ABSTRACT

We combine observational data on a dozen independent cosmic properties at high- $z$  with the information on reionization drawn from the spectra of distant luminous sources and the cosmic microwave background (CMB) to constrain the interconnected evolution of galaxies and the intergalactic medium since the dark ages. The only acceptable solutions are concentrated in two narrow sets. In one of them reionization proceeds in two phases: a first one driven by Population III stars, completed at  $z \sim 10$ , and after a short recombination period a second one driven by normal galaxies, completed at  $z \sim 6$ . In the other set both kinds of sources work in parallel until full reionization at  $z \sim 6$ . The best solution with double reionization gives excellent fits to all the observed cosmic histories, but the CMB optical depth is  $3\sigma$  larger than the recent estimate from the *Planck* data. Alternatively, the best solution with single reionization gives less good fits to the observed star formation rate density and cold gas mass density histories, but the CMB optical depth is consistent with that estimate. We make several predictions, testable with future observations, that should discriminate between the two reionization scenarios. As a byproduct our models provide a natural explanation to some characteristic features of the cosmic properties at high- $z$ , as well as to the origin of globular clusters.

**Key words:** cosmology: dark ages, reionization, first stars – evolution – galaxies: formation – quasars: supermassive black holes – stars: Population III

## 1. INTRODUCTION

Reionization of the cosmic gas after it became essentially neutral at a redshift  $z \sim 1100$  plays a crucial role in the galaxy formation process. Unfortunately, the details of the epoch of reionization (EoR) are to a large extent unknown.

The absence of global absorption shortward of the rest-frame Lyman- $\alpha$  ( $\text{Ly}\alpha$ ) line in the spectra of quasars at  $z < 3$  made Gunn & Peterson (1965) realize that the hydrogen present in the nearby intergalactic medium (IGM) was ionized ( $\bar{x}_{\text{H I}} \lesssim 10^{-4}$ ) except for small intervening systems yielding discrete absorption lines, the so-called  $\text{Ly}\alpha$  forest. The Gunn–Peterson trough caused by neutral intergalactic hydrogen was finally found by Becker et al. (2001) and Djorgovski et al. (2001) in the spectra of quasars at  $z \sim 6$ , suggesting the value for redshift  $z_{\text{ion,H}}$  of full hydrogen ionization.

Several subsequent studies using (a) the mean opacity of the  $\text{Ly}\alpha$  forest (Fan et al. 2006); (b) the size of the proximity zone around quasars (Wyithe et al. 2005; Fan et al. 2006; Bolton & Haehnelt 2007a; Lidz et al. 2007; Maselli et al. 2007); (c) the detection of damping wing absorption by neutral IGM in quasar spectra (Mesinger & Haiman 2004, 2007; Mortlock et al. 2011b); (d) its non-detection in the spectra of gamma-ray bursts (Totani et al. 2006; McQuinn et al. 2007); (e) the abundance of  $\text{Ly}\alpha$  emitters (LAEs; Malhotra & Rhoads 2004; Haiman & Cen 2005; Furlanetto et al. 2006; McQuinn et al. 2007; Mesinger & Furlanetto 2008; Ouchi et al. 2010; Jensen et al. 2013); and (f) the covering fraction of dark pixels in the  $\text{Ly}\alpha$  and  $\text{Ly}\beta$  forests (McGreer et al. 2011) confirmed a value of  $z_{\text{ion,H}}$  of  $6.0^{+0.3}_{-0.5}$ .

Such a value is not inconsistent with the detection of LAEs beyond  $z = 6$ . Not only is the frequency of  $\text{Ly}\alpha$  photons shifted in origin due to outflow velocities of the emitting gas clouds, but they are also sufficiently redshifted when leaving the ionized bubbles around galaxies for not being absorbed by the neutral hydrogen present in the IGM. In any event, the LAE abundance rapidly drops beyond  $z = 6$  (Hayes et al. 2011), the most distant spectroscopically confirmed LAE lying at  $z = 6.96$  (Ota et al. 2008; but see Zitrin et al. 2015). This would imply that the volume filling factor of ionized regions,  $Q_{\text{H II}}(z)$ , is decreasing at  $z = 7$ .

An alternative way to estimate  $z_{\text{ion,H}}$  is from the analysis of the large-scale cosmic microwave background (CMB) anisotropies. Assuming instantaneous reionization, one can determine  $z_{\text{ion,H}}$  and the optical depth to Thomson scattering of free electrons by CMB photons,  $\tau$  (hereafter simply the CMB optical depth). The nine-year data gathered from the *Wilkinson Microwave Anisotropy Probe* (WMAP9; Hinshaw et al. 2013) led to  $z_{\text{ion,H}} = 10.6 \pm 1.2$  and  $\tau = 0.089 \pm 0.014$ . The estimates drawn from the temperature three-year *Planck* data combined with the polarization WMAP data were  $z_{\text{ion,H}} = 9.9^{+1.8}_{-1.6}$  and  $\tau = 0.078 \pm 0.014$  (Planck Collaboration 2016b), and the most recent data inferred taking into account the *Planck* data on the large-scale polarization anisotropies have decreased to  $z_{\text{ion,H}} = 8.8^{+0.9}_{-0.8}$  and  $\tau = 0.058 \pm 0.012$  (Planck Collaboration et al. 2016a).

The marked difference between the latter  $z_{\text{ion,H}}$  estimates and the previous ones drawn from the spectra of distant luminous sources seems to indicate that reionization, far from being instantaneous, was extended, and possibly non-monotonous

(Cen 2003; Chiu et al. 2003; Haiman & Holder 2003; Hui & Haiman 2003; Gnedin 2004; Naselsky & Chiang 2004; Sokasian et al. 2004; Wyithe & Loeb 2004; Furlanetto & Loeb 2005; Bolton & Haehnelt 2007b; Wyithe & Cen 2007). In this sense, the larger values of  $z_{\text{ion,H}}$  inferred assuming instantaneous reionization would just indicate that the ionized fraction was quite low by  $z \sim 10$ . Of course, the uncertainty in the associated  $\tau$  value is also large. However, provided the ionized fraction was small enough at  $z \gtrsim 15$ , the value of  $\tau$  should not be too biased (Holder et al. 2003).

The small-scale CMB anisotropies also inform on the duration  $\Delta z$  of reionization (Zahn et al. 2012; Planck Collaboration et al. 2016a). However, its derivation depends on the uncertain correction for dust emission and the dominant signal at the relevant spatial frequencies, and is very sensitive to the assumed morphological model of ionized bubbles.

Another integral quantity that constrains the EoR, similar to  $\tau$  but more sensitive to low redshifts, is the weak Comptonization distortion of the CMB radiation spectrum. Although its measured upper limit of  $y < 1.5 \times 10^{-5}$  (Mather et al. 1990) is a rather loose constraint, it has the advantage with respect to  $\tau$  of being inferred with no modeling.

But the only way to probe the EoR at every  $z$  is through 21 cm hyperfine line observations (see Morales & Wyithe 2010 for a comprehensive review). Unfortunately, the cosmological signal is five orders of magnitude smaller than that produced in galaxies, which greatly complicates the technique, and makes it necessary to have a previous good knowledge of the galaxy formation process (Iliev et al. 2008; Ichikawa et al. 2010). Presently, this approach has only allowed one to put a lower limit to  $\Delta z$  of 0.06 (Bowman & Rogers 2010).

Among the various mechanisms that might contribute to reionization (see e.g., Chen 2007), the most simple one is photo-ionization by luminous sources, namely active galactic nuclei (AGNs), normal galaxies with ordinary Population II (Pop II) stars, and first-generation metal-free Population III (Pop III) stars. The X-ray background demonstrates that AGNs emit one order of magnitude fewer UV photons than needed at  $z \sim 6$  (Worsley et al. 2005; Cowie et al. 2009; McQuinn 2012). This implies that either massive black holes (MBHs) are too scarce (Willott et al. 2010b), or their associated AGN are too obscured by dust (Treister et al. 2011), while more abundant mini-quasars associated with stellar black holes would have too-short duty cycles (Alvarez et al. 2009; Milosavljević et al. 2009). Likewise, the star formation rate (SFR) densities derived from the rest-frame UV luminosity functions (LFs) of normal bright galaxies (McLure et al. 2010; Bouwens et al. 2011, 2012a; Lorenzoni et al. 2011) show that these objects are insufficient to ionize the IGM by  $z \sim 6$ . Therefore, ionization would be achieved either as a result of a substantial abundance of faint star-forming galaxies (Bouwens et al. 2015; Ishigaki et al. 2015; Robertson et al. 2015) or of the still-undetected first-generation Pop III stars (Sokasian et al. 2004; Shin et al. 2008).

The situation is even more uncertain regarding He II reionization. The He II mean opacity inferred from the Ly $\alpha$  forest suggests that the redshift of complete He II reionization,  $z_{\text{ion,He}}$ , should be less than or approximately equal to 3 (Ricotti et al. 2000; Schaye et al. 2000; Dixon & Furlanetto 2009; McQuinn 2009; Becker et al. 2011), and probably no smaller than 2 as the AGN emission begins to decline at that redshift (Cowie et al. 2009). AGNs are indeed the most plausible

ionizing sources responsible of He II reionization as they emit more energetic photons than normal galaxies, and Pop III stars, the other sources of energetic photons, do not form at  $z < 6$ .

It is thus clear that to fully determine the EoR we must monitor the formation and evolution of the three kinds of luminous sources and their feedback since the dark ages. The problem with this approach is that some aspects of the physics involved are poorly known and must be treated through numerous parameters. This affects not only semi-analytic models (SAMs) but also cosmological hydrodynamic simulations. Indeed, simulations accurately deal with dark matter (DM), but most of the baryon physics is at the sub-resolution scale. Consequently, they use essentially the same recipes and parameters as SAMs in this regard. This problem is particularly hard for simulations because the huge CPU time they involve severely limits the coverage of the whole parameter space. Thus SAMs are the best option.

The observational data on the universe at high- $z$  are rapidly growing, and it might now be possible to address this problem successfully. That is, it might be possible to adjust all the free parameters of a SAM-like model such as the *Analytic Model of Intergalactic-medium and Galaxies* (AMIGA), specifically designed by Manrique et al. (2015) to deal with the interconnected process of galaxy formation and reionization. In the present paper we show that, combining the above-mentioned information on reionization with the current data on the evolution of the universe at  $z > 2$ , we can constrain, indeed, the freedom of this model, and have an unprecedented, detailed picture of the EoR.

The paper is organized as follows. In Section 2 we recall the basic characteristics of AMIGA, paying special attention to its free parameters. The observational data used to adjust those parameters are given in Section 3, and the fitting strategy used is described in Section 4. Our results are presented in Section 5, and discussed and summarized in Section 6.

Throughout the paper we assume the  $\Lambda$ CDM cosmology, with  $\Omega_{\Lambda} = 0.684$ ,  $\Omega_{\text{m}} = 0.316$ ,  $\Omega_{\text{b}} = 0.049$ ,  $h = 0.673$ ,  $n_{\text{s}} = 0.965$ , and  $\sigma_8 = 0.831$  (Planck Collaboration et al. 2016b).

## 2. THE MODEL

AMIGA is a very complete and detailed, self-consistent model of galaxy formation particularly well suited to monitor the coupled evolution of luminous sources and the IGM. All processes entering the problem are treated as accurately as possible according to our present knowledge. In particular, AMIGA is the only model of galaxy formation developed so far that includes Pop III stars, normal galaxies, and AGNs, together with an inhomogeneous multiphase IGM.

AMIGA deals with the properties of halos and their baryon content by interpolating them in grids of masses and redshifts that are progressively built, as halos merge and accrete, from trivial boundary conditions. This procedure is less computer memory-demanding than the usual method based on Monte-Carlo or  $N$ -body realizations of halo merger trees, and enables us to reach very high redshifts and very low masses while maintaining a good sampling over the entire range in redshift and mass, respectively. In this way we can readily integrate at every  $z$  the feedback of luminous sources for the halo mass function (MF), and accurately evolve the IGM.

The total number of free parameters in AMIGA is notably reduced in comparison to ordinary SAMs thanks to causally

linking physical processes that are usually treated as independent from each other. This greatly diminishes the degrees of freedom and renders the model predictions more reliable. The reader is referred to Manrique et al. (2015) for a full description of AMIGA. Here we provide a brief overview focusing on the model parameters.

### 2.1. IGM

Ionizing photons have a very short mean free path in a neutral hydrogenic gas (hereafter the gas outside halos is referred to as the “IGM”). For this reason, luminous sources ionize (and reheat) the IGM in small bubbles around them, which progressively grow and percolate. H II bubbles in turn harbor hotter He III sub-bubbles due to the most energetic fraction of UV photons emitted. Should the ionizing flux become less intense, the singly or doubly ionized bubbles stop stretching, and the surrounding H I or He II regions begin to grow again by recombination (or new dense H I or He II nodes begin to develop after complete reionization). In other words, the neutral, singly, and doubly ionized IGM phases remain well separated except for small residual fractions.

X-ray photons produced in supernovae (SNe) and to a lesser extent in Pop III stars, ordinary stars in binary systems, and AGNs have instead a large mean free path, and form a uniform background that reheats the IGM by Compton scattering.<sup>7</sup> On the contrary, adiabatic cooling due to cosmic expansion, collisional cooling in hot neutral regions, and Compton cooling by CMB photons oppose themselves to the heating. The latter mechanism is particularly effective at high- $z$  where CMB photons are more abundant.

The intergalactic gas falling into halos is shock-heated to their virial temperature. Thus the only halos able to trap gas at any given moment are those with virial temperatures higher than the time-varying temperature of the surrounding IGM. The accretion of gas onto galaxies is calculated in AMIGA as usual in SAMs: taking into account the cooling of the hot intrahalo gas and its free-fall to the halo center. (Hereafter the gas trapped within halos is simply referred to as “the hot gas,” and the gas inside galaxies as “the cold gas.”) However, such a classical scenario strictly holds only in the case of massive halos where the gas infall rate is limited by the cooling rate. In low-mass halos it is limited by the rate at which the halo accretes gas from outside. Since such an accretion is calculated in AMIGA independently of the symmetry of the problem, our treatment thus implicitly deals with accretion of cold gas onto galaxies according to the filament scenario (Dekel et al. 2009).

Not only do galaxies accrete gas, but they also lose it through SN- and AGN-driven winds. Thus the hot gas continuously incorporates metals through three channels: i) via halo mergers, through the gas within the merging halos; ii) via accretion, through the gas (if any) in small accreted halos and directly in the form of diffuse gas (inflows); and iii) via mass loss from the galaxies in the own halo (outflows). Halos themselves may lose the hot gas if it becomes unbound. AMIGA accurately takes into account all these mass exchanges between the different gaseous phases within ionized regions.

In neutral regions IGM remains unpolluted except for possible recombination periods after complete reionization.

Therefore, Pop III stars with metallicity below some threshold value  $Z_c$  for atomic cooling, in the range between  $10^{-5}$  and  $10^{-3}$  solar units, can only form through  $H_2$ -molecular cooling in such neutral, pristine regions, which are immediately photo-dissociated and photo-ionized around those stars.

The evolving extent of the different IGM phases is described by means of the volume filling factors of singly and doubly ionized regions,  $Q_{H II} \equiv \langle n_{H II} \rangle / \langle n_H \rangle$  and  $Q_{He III} \equiv \langle n_{He III} \rangle / \langle n_{He} \rangle$ , which result from solving the improved differential equations (Manrique & Salvador-Solé 2015)

$$\dot{Q}_{S II} = \frac{\dot{N}_{S II}}{\langle n_S \rangle} - \left[ \left\langle \frac{\alpha_{S I}(T)}{\mu^e} \right\rangle_{S II} \frac{C \langle n \rangle}{a^3} + \frac{d \ln \langle n_S \rangle}{dt} \right] Q_{S II} \quad (1)$$

where a dot means time-derivative, with trivial initial conditions in the dark ages. In Equation (1)  $t$  is the cosmic time, subscripts S, S I, and S II mean either H, H I, and H II or He, He II, and He III, respectively, angle brackets stand for the average over the region indicated by the subscript (or over the whole IGM in the absence of subscript),  $n$  is the comoving IGM baryon density,  $a$  is the cosmic scale factor, and  $\mu^e$  is the electronic contribution to the mean molecular weight, i.e.,  $(\mu^e)^{-1}$  is equal (neglecting metals) to zero,  $X + Y/4$ , or  $X + Y/2$ , in neutral, singly, or doubly ionized regions, respectively,  $X$  and  $Y$  being the usual hydrogen and helium mass fractions.  $\dot{N}_{S II}$  is the comoving metagalactic ionizing emissivity due to both luminous sources and recombinations, calculated according to Meiksin (2009),<sup>8</sup>  $\alpha_{S I}$  is the temperature-dependent recombination coefficient to the S I species, and  $C \equiv \langle n_{H II}^2 \rangle / \langle n_{H II} \rangle^2$ , is the so-called clumping factor giving the ratio between the real recombination factor and that in an ideal homogeneous universe at a given  $z$  (see below).

The term in claudators on the right of Equation (1) takes into account inflows/outflows, and the volume average in region S II of the function of temperature  $f(T) \equiv \alpha_{S I}(T)$  is calculated using the Taylor expansion  $f(T_i) + (d^2 f / dT_i^2) \sigma_{T_i}^2 / 2$  from the mean temperatures and variances,  $T_i$  and  $\sigma_{T_i}^2$ , in phases  $i = I, II$ , and III for neutral, singly, and doubly ionized regions, respectively, solutions of the differential equations (Manrique & Salvador-Solé 2015)

$$\frac{d \ln T_i}{d \ln(1+z)} = 2 + \frac{d \ln(\mu_i \varepsilon_i / n_i)}{d \ln(1+z)} \quad (2)$$

$$\frac{d \ln \sigma_{T_i}}{d \ln(1+z)} = \frac{d \ln T_i}{d \ln(1+z)}. \quad (3)$$

In Equations (2) and (3)  $\varepsilon_i$ ,  $n_i$ , and  $\mu_i$  are respectively the phase- $i$  comoving energy density, comoving baryon density, and volume average mean molecular weight, i.e.,  $\mu_i^{-1} = X + Y/4$ ,  $\mu_{II}^{-1} = \mu_I^{-1} (1 + Q_{H II})$ , and  $\mu_{III}^{-1} = \mu_{II}^{-1} + Y/(4X) Q_{He III}$  (neglecting metals). The first term equal to 2 on the right of Equation (2) arises from adiabatic cooling, while the second term includes the effects of Compton heating/cooling by CMB, Compton heating by X-rays, heating/cooling by ionization/recombination of the various hydrogen and helium species, cooling by collisional ionization and excitation, achievement of

<sup>7</sup> Compton scattering by CMB photons also heats the neutral phase until  $z \sim 150$ , when recombination is no more efficient, and the ionized fraction freezes out.

<sup>8</sup> For H I-ionizing photons, we include recombinations to He II and He I ground states. The contribution from He II Ly $\alpha$  recombinations is neglected for simplicity.



energy equipartition of newly ionized/recombined material, and inflows/outflows from halos.<sup>9</sup>

Thus, the evolution of luminous sources determines through Equations (1)–(3) the evolution of the IGM, and conversely, the evolution of the IGM determines in the way shown next the evolution of luminous sources, as well as the halo MF in regions with different ionization states (Manrique & Salvador-Solé 2015). Once the evolution of all these components has been determined we can calculate the CMB optical depth

$$\tau = \sum_{i=II}^{III} \int_0^\infty \sigma_T \frac{c(1+z)^2 dz}{\mu_i^e H(z)} \times \left[ n_i(z) + \int_{M_{\min,i}(z)}^\infty dM \frac{M_{\text{hg}}(M, z)}{m_p} N_i(M, z) \right], \quad (4)$$

and the Compton distortion  $y$ -parameter

$$y = \sum_{i=II}^{III} \int_0^\infty \sigma_T \frac{c(1+z)^2 dz}{\mu_i^e H(z)} \left[ n_i(z) \frac{kT_i(z)}{m_e c^2} + \int_{M_{\min,i}(z)}^\infty dM \frac{M_{\text{hg}}(M, z)}{m_p} N_i(M, z) \frac{kT_{\text{hg}}(M, z)}{m_e c^2} \right]. \quad (5)$$

In Equations (4) and (5)  $H(z)$  is the Hubble parameter,  $\sigma_T$  is the photon Thomson scattering cross-section,  $c$  is the speed of light,  $M_{\text{hg}}(M, z)$  and  $T_{\text{hg}}(M, z)$  are the average mass and temperature of the hot gas in halos with mass  $M$ , respectively,  $N_i(M, z) \equiv N_{\text{ion}}(M, z) Q_i(z) / (Q_{II}(z) + Q_{III}(z))$  is the halo comoving abundance in region  $i$ ,  $N_{\text{ion}}(M, z)$  being the halo MF in ionized regions, and  $M_{\min,i}(z)$  is the minimum halo mass in region  $i$  able to trap gas, all of them dependent on  $z$ .

## 2.2. Luminous Sources

Halos grow through major mergers and smooth accretion. AMIGA follows this growth in a well-contrasted analytic manner with no free parameters (Salvador-Solé et al. 1998, 2007; Raig et al. 2001). This allows us to accurately calculate, at any given moment, their abundance (Juan et al. 2014a, 2014b) and inner structure and kinematics (Salvador-Solé et al. 2012a, 2012b), which in turn sets the structure and temperature of their hot gas (Solanes et al. 2005). The amount and metallicity of such a hot gas and the properties of the central galaxy and its satellites are monitored in AMIGA from the individual halo aggregation history.

As long as the hot gas has a metallicity below  $Z_c$ , molecular cooling takes place (see Manrique et al. 2015). When the temperature reaches a minimum value that cannot be surpassed, the cold gas accumulates in a central disk until the Bonnor–Ebert mass is reached. The cloud then collapses and fragments, giving rise to a small Pop III star cluster of about  $1000 M_\odot$ .<sup>10</sup> The initial mass function (IMF) of Pop III stars is poorly known, but at the present stage we do not need it. We only need the mass fractions  $f_2$  and  $f_3$  of Pop III stars with masses in the ranges  $130 \leq M_*/M_\odot < 260$ , and  $M_* > 260 M_\odot$ , respectively. Indeed, stars in the former range explode in pair-instability SNe

(PISNe), release about half their mass in metals, and do not leave any black hole as remnant, while those in the most massive range essentially collapse into a black hole without producing metals (Heger & Woosley 2002). Thus, the yield of massive Pop III stars,  $p_{\text{III}}$ , and the mass fraction of a Pop III star cluster ending up in a coalesced mini-MBH,  $\beta_{\text{III}}$ , satisfy the relations<sup>11</sup>

$$p_{\text{III}} = 0.5 \times f_2 \quad (6)$$

$$\beta_{\text{III}} = 1.0 \times f_3. \quad (7)$$

On the other hand, the H I- and He II-ionizing photon emissivities of massive Pop III stars depend on the mass fraction  $f_2 + f_3$  in a well-known way (Schaerer 2002).<sup>12</sup> Last, Pop III stars with masses below  $130 M_\odot$  contributing to the mass fraction  $f_1 = 1 - f_2 - f_3$  behave like ordinary Pop II stars, except for their low metallicity. Therefore, the feedback of Pop III stars is fully determined by the mass fractions  $f_2$  and  $f_3$ .

When the metallicity of the hot gas exceeds  $Z_c$ , it undergoes atomic cooling. The gas then contracts, keeping the initial angular momentum, and settles in a rotationally supported disk. The structure of disks is calculated according to Mo et al. (1998) from the specific angular momentum of the gas at the cooling radius, which equals that of DM as provided by numerical simulations. This leads to their effective (half-mass) radii  $r_D$  and central surface density  $\Sigma(0) = M_D / (2\pi \cdot 0.3 r_D^2)$ , where  $M_D$  is the disk mass.<sup>13</sup>

If the disk is unstable, the cold gas is collected in a spheroid. Spheroids also form in galaxy mergers. Indeed, when a halo is captured by another one, it is truncated around the central galaxy, which becomes a satellite orbiting within the new halo. Satellites undergo dynamical friction and orbital decay, being eventually captured by the more massive central galaxy or merging with it if the mass ratio between the two objects is greater than 1:3. Satellites also interact between themselves and with the hot intrahalo gas. However, except for ram-pressure stripping, which is self-consistently calculated in AMIGA, these interactions (tidal stripping, gas starvation at the shock front of the hot gas, tidally induced disk-to-bulge mass transfer, ...) are neglected in the present work as they should play no significant role at high- $z$  when groups and clusters of galaxies are slightly developed and populated (but see Section 5.4).

Spheroids resulting from unstable disks or galaxy mergers have the Hernquist (1990) profile whose projection fits the  $r^{1/4}$  law. Collisions between gas clouds yield the dissipative contraction of recently assembled spheroids with mass  $M_B$  as stars form, according to the physically motivated differential equation for the effective radius  $r_B$  (Manrique et al. 2015)

$$r_B^2 \dot{r}_B = - \frac{Z_{\text{cg,B}}^{1/2} M_{\text{cg,B}}}{Z_\odot^{1/2} \rho_{\text{dis}} \bar{\tau}_{\text{acc}}}, \quad (8)$$

<sup>11</sup> If  $\beta_{\text{III}}$  times the mass of the star cluster is less than  $260 M_\odot$ , whether or not there is a mini-MBH of this latter mass is decided randomly using the Pop III star IMF (see Section 4).

<sup>12</sup> Schaerer (2002) presumes the Pop III star IMF with the usual Salpeter slope. But the results are not very sensitive to this assumption because the radiation emitted by all massive Pop III stars has essentially the same spectral energy distribution (SED), equal to that of a blackbody of  $\sim 10^5$  K.

<sup>13</sup> Caution with the notation used in the present paper for the effective radius of a galaxy component. It coincides with that used in (Manrique & Salvador-Solé 2015) for its length scale.

<sup>9</sup> Neglecting nonlinear effects (IGM is in linear and mildly nonlinear regimes), the volume average of the heating/cooling by gravitational compression/expansion of density fluctuations vanishes.

<sup>10</sup> The exact value depends slightly on  $z$  as the temperature of the gas cloud is bounded by the CMB temperature.

from the initial value given by the  $r_B$ – $M_B$  relation of non-contracted spheroids (see below) until the gas is exhausted or it reaches the density of molecular clouds ( $\sim 10^6$  particles  $\text{cm}^{-3}$ ). In Equation (8)  $M_{\text{cg},B}(t)$  and  $Z_{\text{cg},B}(t)$  are the mass and metallicity, respectively, of the cold gas in the spheroid at  $t$ ,  $\tau_{\text{acc}}$  is the time elapsed from the formation of the spheroid to the quenching of star formation by the enlightened central AGN (see below), and  $\rho_{\text{dis}}$  is a characteristic gas density for the dissipative contraction of spheroids.

When the cold metal-rich gas falls into a galactic component C, i.e., a bulge ( $C = B$ ) or a disk ( $C = D$ ), new Pop II stars form. The SFR satisfies the usual Schmidt–Kennicutt law

$$\dot{M}_{\text{sf},C} = \alpha_G \frac{M_{\text{cg},C}}{\tau_{\text{dyn},C}}, \quad (9)$$

where  $M_{\text{cg},C}(t)$  is the mass of cold gas available,  $\tau_{\text{dyn},C}$  is the dynamical timescale at the half-mass–radius of the galactic component, and  $\alpha_G$  is the star formation efficiency.

The amount of interstellar gas heated by SNe II in Pop II star formation episodes in the galactic component C is

$$M_{\text{th},C} = \epsilon_C \frac{2 \eta_{\text{SN}} E_{\text{SN}}}{V_C^2 - V_{\text{hg}}^2} M_{\text{sf},C}, \quad (10)$$

where  $V_C$  is the circular velocity at  $r_C$ ,  $V_{\text{hg}}$  is the typical thermal velocity of the hot gas in the halo where the heated gas is deposited,  $E_{\text{SN}} = 10^{51}$  erg is the typical energy liberated by one typical SN explosion,  $\eta_{\text{SN}} = 0.014 M_\odot^{-1}$  is the number of such explosions per unit stellar mass in a typical  $\sim 0.20$  Gyr duration starburst for the adopted IMF, and  $\epsilon_C$  is the SN heating efficiency of the component. Due to the different SFR, and the geometry and porosity of the gas in spheroids and disks,  $\epsilon_C$  is let to take different values in bulges ( $\epsilon_B$ ) and disks ( $\epsilon_D$ ). Disks smaller than the corresponding spheroids are, however, assumed to be oblate pseudo-bulges with  $\epsilon_B$ .

The amount of intrahalo gas heated through PISN during the explosion of very massive Pop III stars and leaving the halo is also calculated according to Equation (10), but with  $\epsilon_C = \epsilon_B$ ,  $E_{\text{SN}} = E_{\text{PISN}} = 10^{53}$  erg,  $\eta_{\text{SN}} = \eta_{\text{PISN}} = 0.0015 M_\odot^{-1}$  (Heger & Woosley 2002; Schaerer 2002),  $V_C$  equal to the halo circular velocity at half-mass–radius, and  $V_{\text{hg}} = 0$ .

The instantaneous emission at all the relevant wavelengths of normal galaxies is calculated according to their individual stellar formation and metallicity histories using the stellar population synthesis model (SPSM) by Bruzual & Charlot (2003) for the adopted IMF (see below) and taking into account the typical yield of Pop II stars (Vincenzo et al. 2016) and all the mass and metallicity exchanges between the different galaxy and halo components.

In particular, as long as halos accrete, the metal-enriched gas ejected from the central galaxy remains, due to viscosity, at the cooling radius where it is the next to cool. AMIGA assumes that, at the next major merger, all the hot gas of the progenitors is mixed up, except for their inner  $h_{\text{mix}}$  mass fraction, which retains all the reheated gas remaining in those progenitors. This means that metals ejected from central galaxies take at least two consecutive major mergers to get well mixed within the halo. As a consequence, the hot gas in the inner part of the halo is always somewhat more metal-rich than in the outer part.

MBHs originate as remnants of the most massive Pop III stars that, as mentioned before, are supposed to

coalesce in one mini-MBH per star cluster. When Pop III star clusters (or their remnants) are subsequently captured by normal galaxies and reach their spheroids, the corresponding mini-MBHs migrate by dynamical friction to their centers where they merge with the MBH lying there. MBHs at the center of spheroids progressively grow not only through mergers, but mainly through accretion of gas (and to a much lesser extent of stars). AMIGA does not assume any specific fraction of the gas fallen into spheroids to fuel the central MBH. Instead, it assumes that, each time new gas falls into a spheroid, part of it forms stars, and part fuels the MBH. The starburst lasts until it is quenched by the enlightened AGN that heats mechanically an amount of gas equal to

$$M_{\text{rh},B}^{\text{AGN}} = \epsilon_{\text{AGN}} \frac{2 \langle L \rangle \tau_{\text{acc}}}{V_B^2 - V_{\text{hg}}^2}, \quad (11)$$

and expels it back into the halo through super-winds. In Equation (11)  $\epsilon_{\text{AGN}}$  is the so-called quasar-mode AGN heating efficiency, and  $\langle L \rangle$  is the AGN bolometric luminosity  $L(t)$ , averaged over the typical duration ( $\tau_{\text{acc}} = 0.15$  Gyr; Kelly et al. 2010) of one such enlightening episodes. The bolometric light curve is modeled according to Hatziminaoglou et al. (2003) through the expression

$$L(t) = \dot{M}_{\text{BH}} V_{\text{last}}^2 / 2, \quad (12)$$

where  $V_{\text{last}}$  is the Keplerian velocity at the last marginally stable orbit (at 9.2 times the Schwarzschild radius) around the MBH, and  $\dot{M}_{\text{BH}}$  is the time-derivative of the MBH accretion curve,  $\dot{M}_{\text{BH}}(t)$ , assumed with a universal dimensionless bell-shaped form, whose maximum at  $\tau_{\text{acc}}$  sets the quenching of the starburst going on in the spheroid.

The fraction of H I- (and He II-) ionizing photons escaping from halos with a virial temperature below  $10^4$  K is self-consistently calculated by subtracting those photons ionizing the neutral intrahalo gas, taking into account recombinations. Above  $10^4$  K, a fixed value,  $f_{\text{esc}}$ , is adopted, distinguishing between galaxies,  $f_{\text{esc},G}$ , and AGNs,  $f_{\text{esc},\text{AGN}}$ .

The fraction of the SN explosion energy converted to X-ray photons through free–free emission of the SN remnant and inverse Compton scattering of CMB photons by relativistic electrons is about 1% (Oh & Haiman 2003), and the fraction of the AGN bolometric luminosity emitted in soft X-rays is 4% (Vasudevan & Fabian 2007). About half of those soft X-ray photons Compton heat the IGM in neutral regions due to the residual free electrons. The other half produce secondary ionizations and excitations (Oh & Haiman 2003), neglected in AMIGA.

### 2.3. Parameters

All quantities entering the previous equations are self-consistently evolved in AMIGA from trivial initial conditions. Only a few of them must be supplied as external inputs.

Some of those external inputs correspond to reasonably well-known functions of redshift, stellar mass, or galaxy mass. These are the following.

1. The clumping factor: Hydrodynamic simulations show that  $C(z)$  is equal to about 3 at  $z \sim 6$  (Finlator et al. 2012 and references therein), and evolves essentially as

$$C_F(z) = 9.25 - 7.21 \log(1 + z). \quad (13)$$

Strictly speaking, this fitting expression was derived by Finlator et al. (2012) from their simulations where reionization started at  $z \sim 13$ , and ended up at  $z_{\text{ion,H}} \sim 8$ . In the real universe,  $z_{\text{ion,H}}$  is instead closer to  $z \sim 6$ , and starts at a redshift  $z_{\text{start}}$  to be determined. We thus adopt the following generalization of expression (13),

$$C(z) = C_F(8) - \frac{[C_F(8) - C_F(13)]}{\log[(1 + z_{\text{start}})/7]} \log[(1 + z)/7], \quad (14)$$

with identical linear behavior but satisfying  $C(z_{\text{start}}) \equiv C_F(13) = 1$  and  $C(6) \equiv C_F(8) = 3$ .

2. The IMF of ordinary Pop II stars: Observations show that it can be approximated by a Salpeter IMF. In AMIGA we adopt the Salpeter slope,  $-2.35$ , for large masses up to  $130 M_\odot$ , and the slope  $-1$  for small masses in the range  $0.1 < M_*/M_\odot < 0.5$ . Such an IMF is consistent with the observed local one (Wilkins et al. 2008), and similarly top-heavy as the Chabrier (2003) IMF.
3. The  $\eta_B$ – $M_B$  relation for non-dissipatively contracted spheroids: Assuming for simplicity that it is a power law,  $\eta_B = AM_B^\gamma$ , and taking into account that spheroids at the two mass ends of the observed relations for nearby objects (Phillips et al. 1997; Shen et al. 2003, 2007) have not suffered dissipative contraction,<sup>14</sup> one is led to  $A \approx 0.049 \text{ Kpc } M_\odot^{-\gamma}$ , and  $\gamma \approx 0.20$ .

Even though there is some uncertainty in the preceding expressions, the results of the present study have been checked to be very robust against reasonable variations in the coefficients.

But most of the external inputs are constant or can be considered as such in a first approximation.<sup>15</sup> Many of them are either well determined, as in the case of the threshold mass ratio for galaxy captures with destruction, the yield of Pop II stars, and the energy and frequency of SNe, or they play an insignificant role in the results, as the fraction of the energies released by SNe and AGNs that are converted into X-rays, so they can safely be taken with the fixed values specified in Section 2.2. Others can also be taken with any reasonable fixed value because their uncertainty is absorbed in other parameters. This is the case of the factors 0.5 and 1.0 in Equations (6)–(7), the constant  $\tilde{\tau}_{\text{acc}}$  in Equation (8), and the product  $V_{\text{last}}^2 \tau_{\text{acc}}$  times the factor arising from the universal dimensionless MBH accretion curve in Equation (11).

However, all the remaining external inputs must be left free since they are poorly determined, and have a significant effect on the results. These are thus the real parameters of the model to be adjusted through the fit to the observational data. Their complete list is the following:

$Z_c$ :	threshold metallicity for atomic cooling,
$f_2$ :	mass fraction in intermediate Pop III stars,
$f_3$ :	mass fraction in massive Pop III stars,

$\rho_{\text{dis}}$ :	characteristic density for dissipative contraction,
$\alpha_G$ :	Pop I and II star formation efficiency,
$\epsilon_B$ :	SN heating efficiency in spheroids,
$\epsilon_D$ :	SN heating efficiency in disks,
$\epsilon_{\text{AGN}}$ :	AGN quasar-mode heating efficiency,
$h_{\text{mix}}$ :	mixing hot gas mass fraction,
$f_{\text{esc,G}}$ :	escape fraction from galaxies, and
$f_{\text{esc,AGN}}$ :	escape fraction from AGN.

As we will see in Section 4, there is some degeneracy between these 11 parameters. In addition, the unknown Pop III star IMF must satisfy some conditions that have not yet been enforced. As a consequence we will end up with only nine degrees of freedom. Although this may still seem quite a large number, it is similar to that found in all recent analytic models of the EoR<sup>16</sup> that find  $z_{\text{ion,H}} = 6$  from the fit to the observed SFR density of galaxies at  $z > 6$  (Bouwens et al. 2015; Ishigaki et al. 2015; Robertson et al. 2015). We remark, however, that our model is expected to recover not only the observed redshift of full hydrogen ionization, but the whole evolution of the universe from the fit to all the observed cosmic histories at  $z > 2$ . Such an achievement with so few parameters is possible thanks to the fact that our model is *self-consistent* so that many quantities can be calculated and need not to be considered as free parameters.

### 3. THE OBSERVED HIGH-Z UNIVERSE

All the observational data currently available on the evolution of the universe at  $z > 2$  (see Figure 1) refer to the following:

- (a) the cold gas mass density history (CGH),
- (b) the stellar mass density history (STH),
- (c) the MBH mass density history (MBHH),
- (d) the hot gas metallicity history (HGMH),
- (e) the cold gas metallicity history (CGMH),
- (f) the stellar metallicity history (STMH),
- (g) the IGM metallicity history (IGMMH),
- (h) the galaxy morphology history (GAMH),
- (i) the galaxy size history (GASH),
- (j) the SFR density history (SFH),
- (k) the H I-ionizing emissivity history (IEH), and
- (l) the IGM temperature history (IGMTH).

Taking into account that there are two independent IEHs, one for galaxies, and the other for AGNs, this represents 13 independent global, i.e., averaged or integrated, cosmic histories.<sup>17</sup> In addition, the following differential properties are also available at a few discrete redshifts (see e.g., Figures 8 and 11):

- (m) the galaxy stellar MFs (or UV LF), and
- (n) the MBH MFs (or AGN optical and X-ray LF).

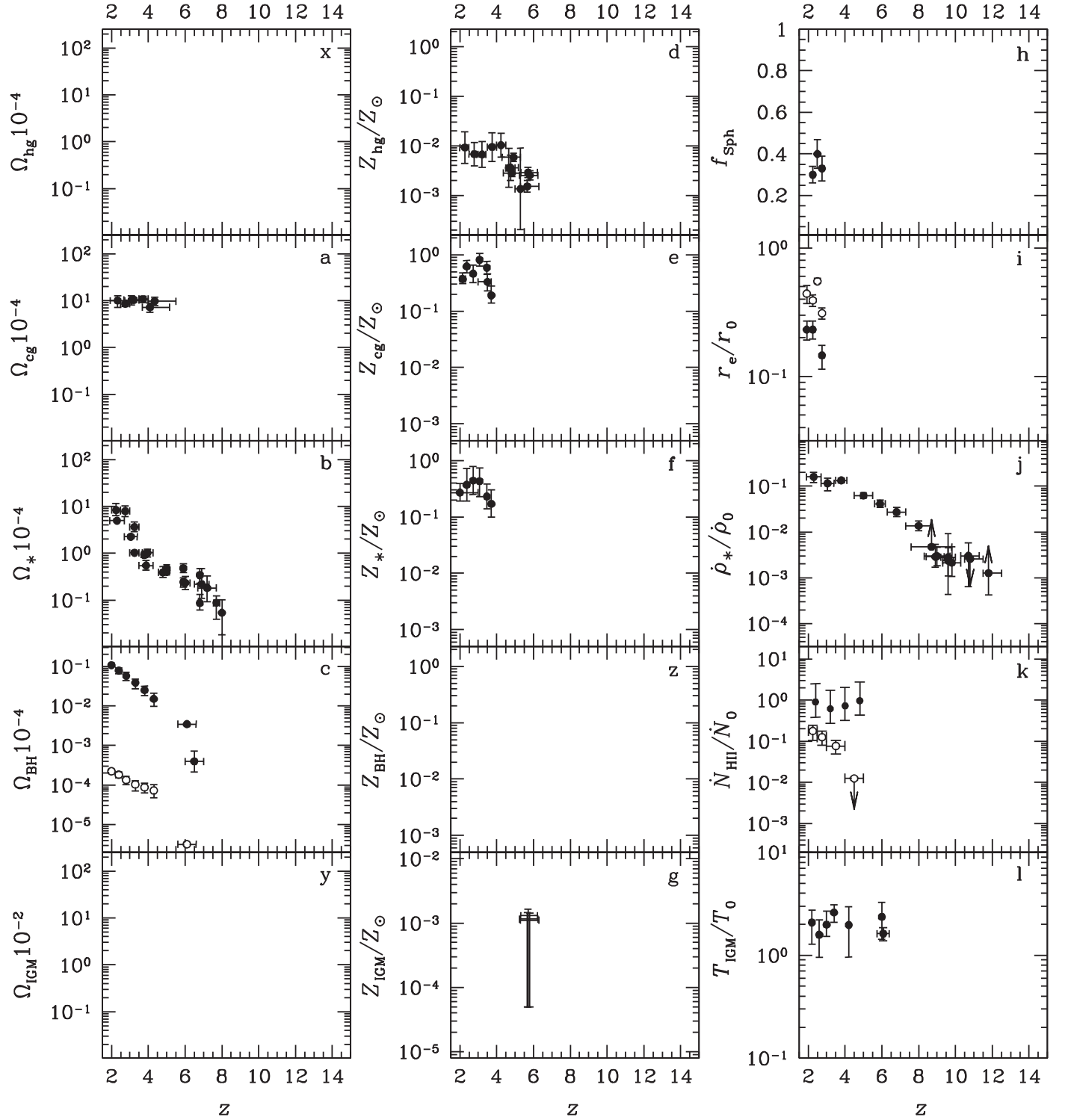
The mass density and emissivity estimates at different redshifts are derived from these MFs or LFs at the

<sup>14</sup> The most massive ellipticals have formed through dry major mergers, and spheroids with low enough mass reach the maximum density of molecular clouds.

<sup>15</sup> For instance, the escape fractions from galaxies and AGNs could vary with  $z$ . However, neither from a theoretical nor from an observational point of view is that dependence clear. Thus we assume them with fixed typical values, as usual.

<sup>16</sup> In all of them the clumping factor and the IMF of ordinary stars are also fixed, as well as the uncertain  $z$ -dependent dust attenuation.

<sup>17</sup> The STH is the time-integral of the SFH, but the two observables are inferred independently. In fact, as we will see, they turn out to be inconsistent with each other.



**Figure 1.** Cosmic histories with available data at  $z > 2$  (see text for the different symbols in some panels): (a) cold gas mass densities (Péroux et al. 2003; Prochaska et al. 2005); (b) stellar mass densities (Reddy & Steidel 2009; Stark et al. 2009; González et al. 2010, 2011; Labbé et al. 2010a, 2010b; Caputi et al. 2011; Mortlock et al. 2011a); (c) MBH mass densities (Kelly et al. 2010; Willott et al. 2010a; Treister et al. 2011); (d) hot gas metallicities (Songaila 2001; Ryan-Weber et al. 2009; Simcoe et al. 2011; D’Odorico et al. 2013); (e) cold gas metallicities (Maiolino et al. 2008; Sommariva et al. 2012; Cullen et al. 2014); (f) stellar metallicities (Halliday et al. 2008; Sommariva et al. 2012); (g) IGM metallicities (based on estimates by Ryan-Weber et al. 2009; Simcoe et al. 2011; D’Odorico et al. 2013); (h) fraction of spheroid-dominated galaxies with masses  $\gtrsim 10^{11} M_{\odot}$  (Buitrago et al. 2008; Bruce et al. 2014a); (i) median effective radii of spheroid-dominated galaxies (Buitrago et al. 2008); (j) SFR densities (Reddy & Steidel 2009; Lorenzoni et al. 2011; Bouwens et al. 2012b; Coe et al. 2013; Ellis et al. 2013; Oesch et al. 2013); (k) total ionizing emissivities (Becker & Bolton 2013) and the contribution from AGN (Cowie et al. 2009); and (l) IGM temperatures (Bolton et al. 2010, 2012; Lidz et al. 2010). Error bars are  $1\sigma$  deviations, except for panel (g) where they give upper and lower limits. The density parameters are for the *current value* of the critical cosmic density,  $\rho_c(z=0)$ , and the scaling factors in the right column are  $r_0 = r_e(z=0)$ ,  $\rho_0 = 1 M_{\odot} \text{ yr}^{-1} \text{ Mpc}^{-3}$ ,  $N_0 = 10^{51} \text{ photons s}^{-1} \text{ Mpc}^{-3}$ , and  $T_0 = 10^4 \text{ K}$ . Other cosmic histories with no data also included for the densities and metallicities in the left and middle columns to cover all baryonic components: (x) hot gas mass densities, (y) IGM mass densities, and (z) metallicities of matter fallen into MBHs.



corresponding  $z$ , extrapolated beyond the observational limits.<sup>18</sup> Therefore, these MFs or LF<sub>s</sub> harbor more information than the corresponding global properties above. However, they are harder to handle. Thus we will concentrate on trying to fit the former cosmic histories, and then check whether or not the galaxy and MBH MFs at a few representative redshifts are also well recovered.

We briefly discuss the techniques, approximations, and models used to infer all those data. All quantities given below refer to the Salpeter-like IMF adopted in AMIGA (applying the conversion factor from a strict Salpeter IMF provided by Wilkins et al. 2008) and to a Hubble constant of 67.3 km s<sup>-1</sup> Mpc<sup>-1</sup>.

- (a) Cold gas mass densities are computed from the distribution function of H I column densities per unit redshift in damped Ly $\alpha$  systems (DLAs) in the spectra of quasars up to  $z_{\text{ion,H}}$ , integrated from zero to infinity. The estimates shown in panel (a) of Figure 1 were obtained by Péroux et al. (2003) and Prochaska et al. (2005). Note that some amount of H<sub>2</sub> (and other molecules), not accounted for in these estimates, must also be present. However, the real gas densities are not expected to be substantially greater.
- (b) Stellar mass densities are inferred by fitting the SED of high- $z$  galaxies, obtained from several optical and infrared rest-frame bands, to template SEDs generated by SPSMs assuming an IMF and a dust absorption correction.<sup>19</sup> The resulting galaxy stellar MFs are extrapolated down to  $10^8 M_{\odot}$ . In panel (b) we show one illustrative sample of data sets obtained by several authors (Reddy & Steidel 2009; Stark et al. 2009; González et al. 2010, 2011; Labbé et al. 2010a, 2010b; Caputi et al. 2011; Mortlock et al. 2011a). The large scatter in the data is mostly due to the very uncertain correction for dust attenuation.
- (c) MBH masses at high redshifts are estimated by the method of reverberation mapping that assumes the broad line emitting clouds around AGNs are virialized. The largest uncertainty in those masses arises from the unknown geometry and inclination of the system, and is encapsulated in a virial factor ( $f_{\sigma}$  or  $f_{\text{FWHM}}$ , depending on whether the potential well is traced through the cloud velocity dispersion or the emission-line width, respectively) of the order of one. The MBH mass density estimates at  $z < 5$  and  $z = 6.1$  plotted as open circles in panel (c) have been computed using the MBH MFs derived by Kelly et al. (2010) and Willott et al. (2010a), respectively, from AGN LF<sub>s</sub> in rest-frame UV. These MFs are, however, greatly affected by AGN obscuration and inactivity, and are substantially lower than the filled circle at  $z = 6.5$  inferred by Treister et al. (2011) from the extrapolated AGN LF in X-rays, less affected by these effects.<sup>20</sup> On the contrary, this latter estimate is fairly aligned with the filled circles at the previous redshifts,

derived assuming a constant MBH-to-spheroid mass ratio,  $\mu$ , equal to  $8 \times 10^{-3}$  (Kormendy & Ho 2013, taking into account the change in IMF),<sup>21</sup> and the upper envelope of the stellar mass densities given in item (b).<sup>22</sup>

- (d) Absorption lines of metal ions, usually C IV, are observed outside the Ly $\alpha$  forest in the spectra of distant quasars. The absorbers have column densities  $N$  according to a distribution function at any given  $z$  well fitted by the expression  $f(N) \propto N^{-(1.32+0.0779z)}$  (D’Odorico et al. 2013), from which one can infer the average metal abundance of the absorbers. The metallicities plotted in panel (d), obtained by Songaila (2001), Ryan-Weber et al. (2009), Simcoe et al. (2011), and D’Odorico et al. (2013), have been homogenized to the widest column density range available ( $10^{12}$ – $10^{15}$  cm<sup>-2</sup>), converted to C abundances for a [C IV/C] abundance ratio of 0.05 appropriate to high column densities (Simcoe 2006), and then to metallicities assuming the C mass fraction of ejecta from ordinary Pop I and II stars given by Ryan-Weber et al. (2009). Indeed, the absorbers are believed to be neutral polluted gas regions around intervening halos, with density contrasts spanning from 10 to 100 in redshifts going from 6 to 2, respectively. Consequently, these metallicities should coincide with that of the hot gas in low-mass halos having accreted gas for the first time (i.e., with masses between the minimum trapping mass at that  $z$  and, say, 1 dex larger).
- (e) Gas-phase metallicities of high- $z$  star-forming galaxies are derived from emission lines in their composite spectra. The estimates obtained for different metal ions are calibrated against each other, and converted to the [O/H] abundance ratio, then to the global metallicity, using the relation provided by Sommariva et al. (2012). In panel (e) we plot typical values obtained for small samples of galaxies with stellar masses above  $\sim 5 \times 10^9 M_{\odot}$  at  $z$  around 3.5 and 2.2, respectively, inferred by Maiolino et al. (2008) and Cullen et al. (2014), and several individual values for a few galaxies with similar masses at the remaining redshifts inferred by Sommariva et al. (2012).
- (f) Stellar metallicities of high- $z$  star-forming galaxies are derived similarly to the gas-phase metallicities in item (e) but from ion absorption lines. In panel (f) we plot the values obtained for a small sample of galaxies with stellar masses above  $\sim 5 \times 10^9 M_{\odot}$  at  $z$  around 2.0 inferred by Halliday et al. (2008) and several individual values corresponding to a few galaxies with similar masses inferred by Sommariva et al. (2012).
- (g) There are no direct measures of the average IGM metallicities. However, we can put in panel (g) some upper and lower limits. Indeed, as mentioned in item (d), metal absorption lines outside the Ly $\alpha$  forest in the spectra of distant quasars would be produced by the polluted IGM around halos. Since the metallicity of those systems decreases with increasing  $z$  due to the increasing effect of galaxy ejecta, their values for the most distant systems ( $5 \leq z < 6$ ) are the best upper limits to the

<sup>18</sup> Given the asymptotic logarithmic slopes of those MFs or LF<sub>s</sub>, such extrapolations are not expected to have any dramatic effect (but see Section 5.3).

<sup>19</sup> Other quantities, such as the metallicity or the time elapsed since the last starburst, are usually varied within reasonable ranges and typical (or favorite) values are adopted.

<sup>20</sup> Only about 20%–30% of all AGN are believed to be Compton thick, and hence lost in X-rays (Castelló-Mor et al. 2013). On the other hand, most of them are active at  $z \gtrsim 6$  (Willott et al. 2010a) due to the high frequency of unstable disks (see Section 5).

<sup>21</sup> Observations show that the MBH-to-spheroid mass ratio is constant at least up to  $z \sim 2$ . The mild evolution seen beyond that redshift is likely due to the Malmquist bias (Kormendy & Ho 2013).

<sup>22</sup> We adopt the upper envelope because it is more consistent with the SFR density estimates of item (j).



average “primordial” (i.e., enriched by Pop III stars only) IGM metallicity we can have. For the present purposes, these metallicities are derived assuming a  $[\text{C IV}/\text{C}]$  abundance ratio of 0.5 (Ryan-Weber et al. 2009), appropriate to low column densities, and a carbon mass fraction of Pop III stars (model C in Schaerer 2002) as it corresponds to IGM polluted by Pop III stars only. On the other hand, the IGM metallicity can become lower than  $Z_c$  at low- $z$  due to the lower metallicity-to-ionizing photon contribution of normal galaxies into the IGM compared to that of Pop III stars that dominate at high- $z$ . A lower limit of  $0.5 \times 10^{-4} Z_\odot$  is thus an educated guess.

- (h) Morphological fractions are inferred from fitting a Sérsic (1968) law to galaxy surface-brightness profiles using circular apertures. This yields the Sérsic index,  $n$ , and the effective (half-light) radius,  $r_e$ , of the galaxy. Alternatively one can decompose the surface-brightness profile in a bulge and a disk. The fractions of spheroid-dominated star-forming galaxies with stellar masses  $\sim 10^{11} M_\odot$  plotted in panel (h) were obtained by Buitrago et al. (2008) for a threshold Sérsic index value of  $n > 2$ , and by Bruce et al. (2014a), at  $z = 2.5$ , for star-forming galaxies with bulge-to-total light ratios greater than 0.5. Both methods yield similar results.
- (i) Galaxy sizes depend on the galaxy morphology. The median effective radii of spheroid-dominated ( $n > 2$ ) and disk-dominated ( $n \leq 2$ ) objects scaled to the corresponding local values (Shen et al. 2003) for galaxies with stellar masses  $\sim 10^{11} M_\odot$  plotted in panel (i) as filled circles and open circles, respectively, were obtained by Buitrago et al. (2008) for the same galaxy sample as used to draw the morphological fraction in item (h). The estimates inferred by Bruce et al. (2014b) directly using the effective radii of bulges and disks yield substantially different results and have not been considered. On the other hand, we will concentrate, from now on, in the effective radii of spheroid-dominated galaxies, as those of disk-dominated galaxies depend on the relative spatial distribution of stars and gas in the disk, which is not modeled in AMIGA.
- (j) SFR densities,  $\dot{\rho}_*$ , are derived from the luminosity of observed galaxies in broadband or narrowband filters that traces active star formation, correcting for incompleteness from the extrapolated LFs and for dust absorption, applying IMF-dependent SFR calibrations. In panel (j) we plot the estimates inferred by Reddy & Steidel (2009), Lorenzoni et al. (2011), Bouwens et al. (2012b), Oesch et al. (2013), Coe et al. (2013), and Ellis et al. (2013). The dust-uncorrected values listed by Bouwens et al. (2012b) have been shifted upwards according to Jaacks et al. (2012) to correct for that effect. The resulting SFR densities are in very good agreement with the latest results published by McLeod et al. (2015).
- (k) Similarly to the escape fraction of ionizing (or Lyman-continuum, LyC) photons, which has two different versions, one for galaxies and the other for AGNs, we will distinguish between the H I-ionizing emissivity from normal galaxies and AGN. At high- $z$ , the LyC emissivity of normal galaxies is usually estimated from their SFR densities assuming a given escape fraction of ionizing photons (Richard et al. 2006; Stark et al. 2007; Bouwens

et al. 2011). These estimates are not considered here as they are equivalent to the SFR densities plotted in panel (h). An alternative method (Becker & Bolton 2013), *independent from the SFR densities* consists of using the Ly $\alpha$  effective optical depth inferred from the comparison of the observed Ly $\alpha$  forest with synthetic data obtained from hydrodynamic simulations, modeling photon mean free paths for the assumed ionizing emissivities. These are the estimates plotted as filled circles in panel (k). Regarding LyC emissivity from AGNs, this is estimated from UV observations of a large sample of X-ray AGNs. The results derived by Cowie et al. (2009) are plotted in the same panel as empty circles.

- (l) The temperature of the singly ionized IGM is usually derived from the small-scale structure of the Ly $\alpha$  forest, using hydrodynamic simulations and galaxy and quasar emission models. This is the case for the data at  $z < 5$  derived by Lidz et al. (2010) plotted in panel (l). We do not include the estimates inferred by Becker et al. (2011) as they greatly depend on the poorly known density-temperature relation. Unfortunately, all those measures are extremely challenging for  $z$  approaching 6. An alternative method that sidesteps such a difficulty consists of probing the IGM temperature within a proper distance of  $\sim 5$  Mpc of quasars by combining the cumulative probability distribution of Doppler broadening and synthetic Ly $\alpha$  spectra. The measurements plotted at  $z \sim 6$  were obtained in this way by Bolton et al. (2010) and Bolton et al. (2012), using one and seven quasars, respectively.
- (m) Galaxy stellar MFs are obtained from the galaxy LFs in rest-frame UV. The conversion involves the assumptions and procedures mentioned in item (b). For consistency, we adopt the MFs corrected for absorption inferred by González et al. (2011) at  $z = 3.8$  and  $z = 6.8$ , which were also used to derive the stellar mass densities plotted in panel (b) at those redshifts that typically bracket the range of interest.
- (n) Similar comments hold for the MBH MFs. Their derivation from optical AGN LFs involves the assumptions and procedures mentioned in item (c). For consistency, we adopt the MFs inferred by Vestergaard et al. (2008) and Willott et al. (2010a) at  $z = 3.3$  and  $z = 6.1$ , respectively, which were also used to infer the MBH mass densities plotted in panel (c) at those redshifts that bracket the range of interest.

#### 4. FITTING STRATEGY

The goal is to adjust the 11 free parameters mentioned in Section 2 through the fit to the 13 independent cosmic histories described in Section 3, enforcing the constraints on  $z_{\text{ion,H}}$  and  $Q_{\text{H II}}(z)$  mentioned in Section 1 as priors.

Scanning a whole 11 dimensional (11D) parameter space is not an easy task. Fortunately, we can take advantage of some particularities of the problem that greatly simplify the fitting procedure.

First, the characteristic dissipation density,  $\rho_{\text{dis}}$ , is fully constrained by the data, regardless of the value of any other parameter. Indeed, Equation (8) recovers the observed effective radius of massive spheroids at  $z = 2-3$  formed in major mergers of two galaxies with the typical gas content and

metallicity (see panels (i), (a), and (e) of Figure 1) provided only  $\log(\rho_{\text{dis}} M_{\odot}^{-1} \text{Kpc}^3) = 6.0 \pm 0.6$ .

Second, the properties of normal galaxies and MBHs decouple from those of Pop III stars (Manrique et al. 2015). Indeed, massive Pop III stars lose metals through PISN-driven winds into the ionized bubbles they form around. When this polluted gas falls into halos and cools, normal galaxies form. In principle, the properties of those galaxies should thus depend on the metallicity of the ionized IGM set by Pop III stars through the values of  $Z_c$ ,  $f_2$ , and  $f_3$ . But normal galaxies eject such large amounts of metals into the hot gas in halos through Type II SN- and AGN-driven winds that the hot gas rapidly loses the memory of its initial metallicity, and the properties of galaxies rapidly become independent of  $Z_c$ ,  $f_2$ , and  $f_3$ . Similarly, MBHs are seeded by the remnants of massive Pop III stars. But the dramatic growth of MBHs within spheroids, regulated by the amount of gas reaching them, causes MBHs to rapidly lose the memory of their seeds, and the MBH properties also become independent of  $Z_c$ ,  $f_2$ , and  $f_3$ . We can thus concentrate on adjusting, in a first step, the parameters referring to Pop III stars and, afterward, those referring to normal galaxies and MBHs. Notice that, even though the properties of the two kinds of sources are independent of each other, we cannot exchange the order of those fits. Some of the observables involving the parameters of the latter set refer to mean mass densities per unit volume, not per unit *ionized* volume, or to metallicities averaged over all regions, not over *ionized* regions. Consequently, they involve not only the properties of normal galaxies and MBHs, but also the volume filling factor of ionized regions, which depends on the mass fraction,  $f_2 + f_3$ , of massive Pop III stars. Nonetheless, once the parameters  $f_2$  and  $f_3$  are fixed, those observables will depend only on the parameters referring to normal galaxies and MBHs.

Third, most of the parameters in these two sets can be adjusted sequentially.

For a given value of  $Z_c$ , the ionization and photo-heating of the IGM at high- $z$  depend only on the mass fraction  $f_2 + f_3$  of massive Pop III stars, while the mass of metals ejected in ionized regions depends only on the mass fraction  $f_2$ . Thus it should be possible to adjust  $f_2 + f_3$  by fitting the IGMTH at high- $z$ , and then adjust  $f_2$  by fitting the IGMMH also at high- $z$ .

On the other hand, for a given couple of  $\epsilon_B$  and  $\epsilon_D$  values, the amount of gas in the disk or spheroid of a galaxy at any moment is equal to the amount of gas fallen into it, which depends on cosmology (through the halo characteristics, the hot gas cooling rate, and the galaxy merger rate),<sup>23</sup> minus the amount of gas used to form stars and ejected from the component (the latter depending on the former).<sup>24</sup> The inductive reasoning thus implies that the stellar masses of normal galaxies depend only on  $\alpha_G$ . Similarly, the mass of gas fueling the central MBH of a galaxy depends on the gas mass collected in the spheroid, which together with the stellar mass sets the dynamics, and hence, the mass of newly formed stars and of gas heated by SNe, the mass of the MBH determining its accretion rate, and the mass of gas heated by the enlightened AGNs. Therefore, the inductive reasoning implies that the MBH depends on both  $\alpha_G$  and  $\epsilon_{\text{AGN}}$ . Last, the metallicity of

the hot gas in the halo that determines, through the gas cooling and the metal-enriched gas ejected from galaxies, the cold gas, and stellar metallicities of those objects depends on cosmology, on  $\alpha_G$  and  $\epsilon_{\text{AGN}}$ , and on  $h_{\text{mix}}$  setting the mixing rate of the reheated gas in the halo. Finally, the emissivity of galaxies and AGNs depends on all the previous parameters setting the whole evolution of galaxies and AGNs and their respective intrinsic ionizing photon production rates and the values of  $f_{\text{esc}}$  from the two kinds of sources. Consequently, it should be possible to adjust  $\alpha_G$  by fitting the SFH (or STH), then  $\epsilon_{\text{AGN}}$  by fitting the MBHH, then  $h_{\text{mix}}$  by fitting the HGMH (or any of the CGMH and STMH), and last the  $f_{\text{esc}}$  values of galaxies and AGNs by fitting their corresponding IEHs with the suited priors on  $z_{\text{ion,H}}$  and  $d_z Q_{\text{H II}}$  (hereafter  $d_z$  means  $z$ -derivative).

Of course, this sequential fitting procedure should be carried out for every possible set of  $Z_c$ ,  $\epsilon_B$ , and  $\epsilon_D$  values in the corresponding three-dimensional space. But this is a minor difficulty compared to directly looking for acceptable solutions in the whole 11D parameter space.

Any “acceptable solution” should thus fit with acceptable  $\chi^2$  values the IGMMH, the IGMTH, the SFH (and the STH), the MBHH, the HGMH (and the CGMH and STMH), and the IEHs for galaxies and AGNs, with  $z_{\text{ion,H}} = 6.0^{+0.3}_{-0.5}$  and  $d_z Q_{\text{H II}}(z = 7) < 0$ . Moreover, since  $\epsilon_B$  and  $\epsilon_D$  govern the amount of cold gas remaining in disks and of stars forming in spheroids, the CGH and GAMH should hopefully also be fitted. And given the way  $\rho_{\text{dis}}$  is determined, the GASH should too. In principle, the solution so obtained should also recover the observed galaxy stellar and MBH MFs and, provided  $Z_c$  is not degenerate so that there is still one degree of freedom to play with, we may expect the solution to also fulfill the constraints on  $\tau$  and  $y$ . Thus the problem is not only well posed, but in principle also slightly overdetermined.

Unfortunately, the IGMMH and IGMTH at high- $z$  are not well constrained. The only available data on the IGM temperature refer to  $z \lesssim 6$  where Pop III stars play no significant role, and there are just a few loose bounds for the IGM metallicity. Consequently,  $f_2$  and  $f_3$  cannot be determined in the previous simple manner. Moreover,  $Z_c$  is degenerate with  $f_2$  and  $f_3$ . The good news is that this degeneracy reduces the degrees of freedom to 10, which facilitates the fit at the cost, of course, of making it even harder to find any acceptable solution.

Indeed, massive Pop III stars are the first source of ionizing photons, with emissivity equal to  $m_p \xi_{\text{III}} (f_2 + f_3) \dot{M}_{\text{III}}$ , where  $\dot{M}_{\text{III}}$  and  $\xi_{\text{III}}$  are the formation rate and LyC photon production rate of Pop III stars, respectively. Those stars also pollute with metals the ionized bubbles at the rate  $0.5f_2 \dot{M}_{\text{III}}$ . Therefore, the formation of normal galaxies is possible provided the metallicity of such bubbles,  $0.5f_2/[m_p \xi_{\text{III}} (f_2 + f_3)]$ , is greater than  $Z_c$ . But this condition is necessary although not sufficient. When normal galaxies form, they keep on ionizing and photo-heating the surrounding IGM, while they lose very few metals into it (they rather enrich the hot intrahalo gas). Consequently, the metallicity of ionized bubbles decreases, and their temperature increases, so the formation of new galaxies may be quenched, and reionization may be aborted. Only if the initial IGM metallicity is *sufficiently large* will the formation of normal galaxies last long enough for reionization to proceed successfully.

According to the previous discussion, the evolution of the universe for some values  $Z_c$ ,  $f_2$ , and  $f_3$  will thus be essentially

<sup>23</sup> Cooling is much more sensitive to the density of the hot gas than to its metallicity, dependent on  $h_{\text{mix}}$ .

<sup>24</sup> The gas mass fueling the MBH is negligible compared to the mass converted into stars.

the same (with  $Z_{\text{IGM}}$  in units of  $Z_c$ ) as for other values  $Z'_c$ ,  $f'_2$ , and  $f'_3$ , provided the IGM reionization and metal-enrichment histories driven by Pop III stars are the same, that is, provided the two sets of values satisfy the relations

$$f'_2 + f'_3 = f_2 + f_3 \quad (15)$$

$$\frac{f'_2}{(f'_2 + f'_3)Z'_c} = \frac{f_2}{(f_2 + f_3)Z_c}. \quad (16)$$

This leads to the following simple fitting procedure. For some fiducial values of  $Z_c$ , and  $f_2/(f_2 + f_3)$  warranting that reionization will not be aborted, we scan all possible values of  $f_3$ ,  $\epsilon_B$ , and  $\epsilon_D$ , and for each set of these three values we adjust  $\alpha_G$ ,  $\epsilon_{\text{AGN}}$ ,  $h_{\text{mix}}$ , and  $f_{\text{esc}}$  in the sequential manner explained above, with the priors on  $z_{\text{ion,H}}$  and  $d_z Q_{\text{H II}}$ . Then, given any acceptable solution, the relations (15) and (16) provide with any other combination of  $Z_c$ ,  $f_2$ , and  $f_3$  values leading to it. Last, we can check the values of  $\tau$  and  $y$  as well as the right behavior of the galaxy stellar and MBH MFs.

#### 4.1. Additional Constraints on the Pop III Star IMF

Were the Pop III star IMF known, we could readily calculate the mass fractions  $f_2$  and  $f_3$ , and, as a bonus, leave the above-mentioned degeneracy. Unfortunately, such an IMF is poorly determined. But we can still enforce some conditions such an IMF must satisfy in order to further constrain the problem and reduce the degrees of freedom to 9 (8 if we discount  $\rho_{\text{dis}}$ , directly determined by the data).

The IMF of Pop III stars must be top-heavier than the IMF of ordinary stars (Larson 1998).<sup>25</sup> Approximating it by a power law  $f(M_\star) \propto M_\star^{-\alpha_{\text{III}}}$  like the Salpeter IMF, this implies either that  $\alpha_{\text{III}}$  is smaller than the Salpeter slope, 2.35, or that the upper or lower stellar masses are larger than for ordinary stars, i.e.,  $M_{\text{III}}^{\text{lo}} > 0.5 M_\odot$  and  $M_{\text{III}}^{\text{up}} > 130 M_\odot$ ,<sup>26</sup> or both conditions at the same time.

But this is not all. The fact that the slope of the IMF for ordinary Pop I and II stars appears to be little sensitive to their metallicity indicates that  $\alpha_{\text{III}}$  should not be substantially smaller than 2.35;<sup>27</sup> i.e., it should be larger than 2.25 or at most larger than 2.15. On the other hand,  $M_{\text{III}}^{\text{lo}}$  should be smaller than  $260 M_\odot$ , otherwise Pop III stars would not enrich with metals the IGM, and  $M_{\text{III}}^{\text{up}}$  should be larger than  $260 M_\odot$ , otherwise Pop III stars would not seed MBHs, and smaller than  $500 M_\odot$  or at most smaller than  $1000 M_\odot$  (Heger & Woosley 2002; Schaerer 2002).

To account for all these restrictions parameters,  $f_2$  and  $f_3$  in the fitting procedure described above must be replaced by the equivalent ones (for a fixed  $\alpha_{\text{III}}$  value)  $M_{\text{III}}^{\text{lo}}$  and  $M_{\text{III}}^{\text{up}}$  parameters. Specifically, for some fiducial values of  $Z_c$  in the range  $[10^{-5} Z_\odot, 10^{-3} Z_\odot]$ , of  $\alpha_{\text{III}}$  in the range  $[2.25, 2.35]$  (or  $[2.15, 2.35]$ ), and of  $M_{\text{III}}^{\text{up}}$  in the range  $[260 M_\odot, 500 M_\odot]$  (or  $[260 M_\odot, 1000 M_\odot]$ ) warranting that reionization is not aborted, we scan all possible values of  $M_{\text{III}}^{\text{lo}}$  in the range  $[0.5 M_\odot, 130 M_\odot]$ , and of  $\epsilon_B$  and  $\epsilon_D$  in the range  $[0, 1] \times [0, 1]$ , and adjust in the sequential manner explained above the parameters  $\alpha_G$ ,

$\epsilon_{\text{AGN}}$ ,  $h_{\text{mix}}$ , and the two  $f_{\text{esc}}$  (see Figure 2), with the priors on  $z_{\text{ion,H}}$  and  $d_z Q_{\text{H II}}$ . Using the relations (15) and (16), we find all sets of  $Z_c$ ,  $\alpha_{\text{III}}$ ,  $M_{\text{III}}^{\text{up}}$ , and  $M_{\text{III}}^{\text{lo}}$  values leading to every acceptable solution. Last, we check whether or not the constraints on  $\tau$  and  $y$  are also fulfilled, and whether or not the observed galaxy stellar and MBH MFs are well recovered.

The result will depend, of course, on the somewhat uncertain masses ( $130 M_\odot$  and  $260 M_\odot$ ) delimiting the ranges where Pop III stars have distinct behaviors. But this is a small drawback compared to the gain of better constraining  $Z_c$  and the Pop III star IMF.

## 5. RESULTS

Following this fitting procedure, we have looked for any possible solution, in the sense of being compatible with the currently available data on the high- $z$  universe and the usual reionization constraints, and checked whether it has single or multiple reionization episodes.

The idea of double reionization was advocated by several authors (Cen 2003; Chiu et al. 2003; Haiman & Holder 2003; Hui & Haiman 2003; Naselsky & Chiang 2004; Sokasian et al. 2004; Furlanetto & Loeb 2005) to reconcile the apparent inconsistency between the value of the redshift of complete hydrogen ionization obtained from the spectra of distant sources ( $z_{\text{ion,H}} \sim 6$ ; Becker et al. 2001; Djorgovski et al. 2001) with that found from the CMB optical depth derived from the first-year WMAP data ( $z_{\text{ion,H}} > 15$  as implied by  $\tau = 0.17$ ; Kogut et al. 2003). Subsequent CMB-based measurements decreased this latter estimate until the present value of  $z_{\text{ion,H}}$  of about 10. In addition, it became clear that an *instantaneous reionization* event at  $z \sim 10$  was not necessarily inconsistent with *reionization being completed* by  $z_{\text{ion,H}} \sim 6$ . Consequently, the double reionization scenario lost its interest. However, the possibility that reionization is double still remains. As a matter of fact, this is quite natural to expect.

Indeed, Pop III star clusters have a typical mass of  $\sim 10^3 M_\odot$ , which is much lower than the mass  $\sim 10^6 M_\odot$  of the most abundant dwarf galaxies at high redshifts (see below). However, massive Pop III stars emit 40–80 times more ionizing photons than normal O, B stars, so both kinds of sources compete in the IGM reionization. What is the dominant contribution? Pop III stars are the first to form, so they necessarily drive the initial ionization phase. And, since neutral pristine regions progressively disappear, the final ionization phase will be driven by normal galaxies. The SFR of Pop III stars is fully determined by cosmology, and their more or less top-heavy IMF depends on the unknown value of  $M_{\text{III}}^{\text{lo}}$  (for the fiducial values of  $\alpha_{\text{III}}$  and  $M_{\text{III}}^{\text{up}}$ ). Whereas the IMF of Pop I & II stars, set by the physics of protostellar clouds, is well determined, and their more or less intense SFR depends on the unknown value of  $\alpha_G$ . Consequently, the answer to the question above depends on the relative values of the two independent parameters  $M_{\text{III}}^{\text{lo}}$  and  $\alpha_G$ . For the  $\alpha_G$  value giving good fits to the SFH of galaxies, the larger the value of  $M_{\text{III}}^{\text{lo}}$ , the earlier Pop III stars will begin to ionize the IGM, and the more probable will be the double reionization scenario.

### 5.1. Reionization History

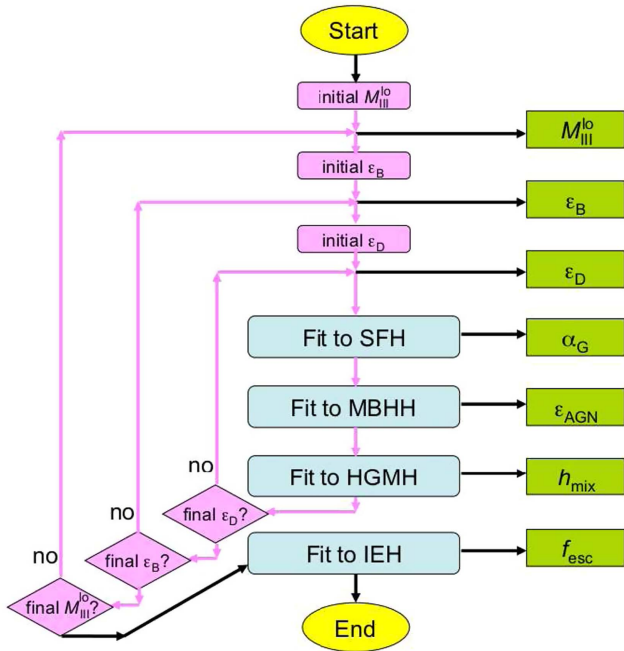
Figure 3 shows the reionization histories of the only acceptable solutions we find. They are concentrated in two narrow sets, one with one single hydrogen ionization episode at

<sup>25</sup> The rationale for a top-heavier Pop III IMF is that the fragmentation of protostellar clouds is favored by metals, absent in the pristine gas.

<sup>26</sup> The upper mass of Pop I and II stars is poorly determined. An upper limit of  $150 M_\odot$  is sometimes considered, although no stars more massive than  $130 M_\odot$  have ever been observed (Figer 2005; but see Crowther et al. 2010).

<sup>27</sup> This is the reason why in most studies dealing with the Pop III star IMF the value of 2.35 is adopted (e.g., Schaerer 2002).





**Figure 2.** Fitting procedure making use of the Pop III star IMF and some fiducial values of  $Z_c$ ,  $\alpha_{\text{III}}$ , and  $M_{\text{III}}^{\text{up}}$ .

$z \sim 6$ , and the other one with two ionization episodes at  $z \sim 10$  and  $z \sim 6$ , separated by a short recombination period. The best solutions of the two kinds, i.e., those that provide the best fit to the data on the SFH, MBHH, HGMH, IEHs, GAMH, and GASH with the priors  $z_{\text{ion,H}} = 6.0$  and  $d_z Q_{\text{H II}}(z = 7) < 0$ , are hereafter referred to as “S” (for single reionization) and “D” (for double reionization).

In all the solutions, the reionization of He II also shows two distinct phases separated by a short recombination period. The first phase leads to a maximum in the volume fraction of He II-ionized regions, although without complete ionization, at  $z \sim 10$  in the solutions with double hydrogen reionization, and at  $z \sim 6-7$  in the solutions with single hydrogen reionization. The first phase is also driven by Pop III stars, but the second one, completed at  $z_{\text{ion,He}} \sim 2.8$ , is driven by AGNs instead of normal galaxies. Thus, the He II reionization history confirms the idea that, when two independent ionizing sources with distinct timing are at play, it is natural to have a non-monotonous reionization.

Table 1 lists the values of  $M_{\text{III}}^{\text{lo}}$ ,  $\epsilon_B$ ,  $\epsilon_D$ ,  $\alpha_G$ ,  $\epsilon_{\text{AGN}}$ ,  $h_{\text{mix}}$ ,  $f_{\text{esc,G}}$ , and  $f_{\text{esc,AGN}}$  defining models S and D for the fiducial values of  $Z_c$ ,  $\alpha_{\text{III}}$ , and  $M_{\text{III}}^{\text{up}}$  given in Table 2. The errors in  $M_{\text{III}}^{\text{lo}}$  delimit the acceptable solutions around models S and D; those in  $\epsilon_B$  and  $\epsilon_D$  mark the deviations from the best values of these two parameters still leading to acceptable solutions; and the errors in the remaining parameters are  $1\sigma$  deviations from their best values in models S and D, estimated from the  $\chi^2$ -test of their respective fits. In principle, the degeneracy between  $Z_c$  and the Pop III star IMF could cause the value of  $M_{\text{III}}^{\text{lo}}$  to vary from that quoted in Table 1. But, for the reasons explained next, there is actually not much room for the possible values of  $Z_c$ ,  $\alpha_{\text{III}}$ , and  $M_{\text{III}}^{\text{up}}$ , which should thus be close to those given in Table 2.

All the combinations of  $Z_c$ ,  $\alpha_{\text{III}}$ ,  $M_{\text{III}}^{\text{up}}$ , and  $M_{\text{III}}^{\text{lo}}$  leading to any given solution can be obtained from one particular combination leading to that solution and Equations (15) and (16) through the relations between the couples of  $M_{\text{III}}^{\text{lo}}$  and  $M_{\text{III}}^{\text{up}}$ , and  $f_2$  and  $f_3$

values for a given  $\alpha_{\text{III}}$  value. That degeneracy is shown in Figure 4. See also Figure 5 for the degeneracy in  $Z_c$ ,  $f_2$ , and  $f_3$ . Since the minimum possible value of  $f_2/(f_2 + f_3)$  for reionization not to be aborted depends on  $Z_c$ , and, for a given  $\alpha_{\text{III}}$ ,  $f_2/(f_2 + f_3)$  is a function of  $M_{\text{III}}^{\text{up}}$  alone, there is also a one-to-one correspondence between the maximum possible value of  $M_{\text{III}}^{\text{up}}$  and  $Z_c$ . Notice that, for the reasons explained in Section 5.3, the maximum possible value of  $M_{\text{III}}^{\text{up}}$  for each  $Z_c$  is smaller in single than in double reionization. Specifically, it is equal to the maximum  $M_{\text{III}}^{\text{up}}$  value found in double reionization times 0.85/0.21, where 0.21 and 0.85 are the minimum  $f_2/(f_2 + f_3)$  values in single and double reionization, respectively.

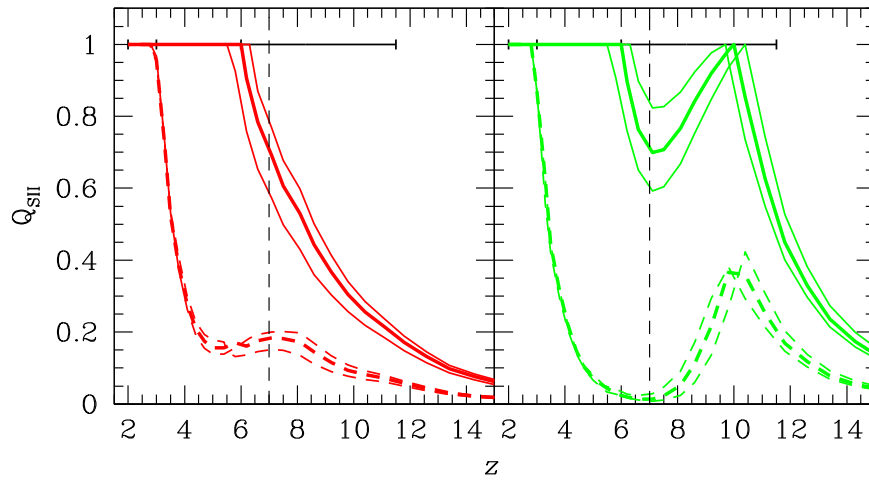
As shown in Figure 4, the upper limit of  $Z_c$  set by the lower limit of  $M_{\text{III}}^{\text{up}}$  equal to  $260 M_{\odot}$  is  $10^{-3.93} Z_{\odot}$  in double reionization, and  $10^{-4.53} Z_{\odot}$  in single reionization. Whereas the lower limit set by the upper limit of  $M_{\text{III}}^{\text{up}}$  equal to  $500 M_{\odot}$  ( $1000 M_{\odot}$ ) is  $10^{-4.17} Z_{\odot}$  ( $10^{-4.30} Z_{\odot}$ ) in double reionization, and  $10^{-4.77} Z_{\odot}$  ( $10^{-4.90} Z_{\odot}$ ) in single reionization. Therefore, the central value and allowed range in double reionization are:  $\log(Z_c/Z_{\odot}) = -4.0 \pm 0.1 Z_{\odot}$ , whereas in single reionization they are:  $\log(Z_c/Z_{\odot}) = -4.6 \pm 0.1$ . These are the fiducial values of  $Z_c$  adopted for the fit. Note that, in the case of double reionization, the observed properties of the high- $z$  universe imply a value of  $Z_c$  of  $10^{-4} Z_{\odot}$ , which coincides with that preferred on pure theoretical grounds (Santoro & Shull 2006; Smith et al. 2009; Schneider & Omukai 2010). Instead, in the case of single reionization, they imply a value of  $Z_c$  closer to the lower limit of  $10^{-5} Z_{\odot}$ .

In Figure 4 we also see that the degeneracy track associated with any given solution depends slightly on  $\alpha_{\text{III}}$ , and so does the exact upper limit of  $Z_c$ . But, for all reasonable values of  $\alpha_{\text{III}}$ , this dependence is only significant for unrealistic values of  $M_{\text{III}}^{\text{up}}$  close to  $1000 M_{\odot}$ . We can thus adopt  $\alpha_{\text{III}} = 2.35$  without any loss of generality, even though the true value of this parameter could be somewhat smaller.

Of course, all  $M_{\text{III}}^{\text{up}}$  values smaller than the maximum possible value for a given  $Z_c$  (i.e., on the left of the corresponding vertical line in Figure 4) together with the  $M_{\text{III}}^{\text{lo}}$  value associated with it along the degeneracy track yield essentially the same solution. (In Figure 5 all the couples  $f_2$  and  $f_3$  along the degeneracy track upward of that value of  $Z_c$  also lead to the same solution.) But the maximum possible value of  $M_{\text{III}}^{\text{up}}$  in both single and double reionization,  $300 M_{\odot}$ , is already close to the lower limit of  $260 M_{\odot}$ , so the true value of  $M_{\text{III}}^{\text{up}}$  must be close, indeed, to the fiducial one of  $300 M_{\odot}$ .

Having fixed the values  $Z_c$ ,  $\alpha_{\text{III}}$ , and  $M_{\text{III}}^{\text{up}}$ , the central and limiting values of  $M_{\text{III}}^{\text{lo}}$  given in Table 1 have been obtained as follows. In the single reionization case, the best solutions they define have the central and limiting values of  $z_{\text{ion,H}}$ , equal to  $6.0_{-0.5}^{+0.3}$ , the condition  $d_z Q_{\text{H II}} < 0$  at  $z = 7$  being automatically fulfilled in this case. In the double reionization case, the best solutions they define have  $z_{\text{ion,H}} = 6.0_{-0.5}^{+0.3}$ , but, in addition, the minimum of  $Q_{\text{H II}}(z)$  is at a redshift larger than, although as close as possible to,  $z = 7$ . This warrants  $d_z Q_{\text{H II}} < 0$  at  $z = 7$  as wanted, together with the minimum possible redshift of first complete ionization, which is very convenient to have the smallest possible value of  $\tau$  (see below).

The values of  $\epsilon_B$  and  $\epsilon_D$  are well constrained by the observed GAMH and GASH, as expected. In fact, those two properties appear to be extremely sensitive to small changes in the values of these (as well as other) parameters. Their values close to one



**Figure 3.** Only acceptable solutions with single (left panel) and double (right panel) reionization. In thick lines models S and D give the best solution of the respective types. In thin lines the solutions bracket the two sets. The error bars centered at  $z = 2.5$  and  $z = 8.8$  along  $Q_{\text{SII}} = 1.0$  give the estimated limits for the redshifts of complete helium and hydrogen ionizations, respectively (see Section 1); the vertical dotted black line marks the redshift  $z = 7$  where  $Q_{\text{HII}}$  is found to decrease with increasing  $z$ . (In the online journal, the solutions for single and double reionization in all the following figures in are in red and green, respectively.)

and zero, respectively, in the two models agree with the results of analytic and hydrodynamic studies on the effects of SN heating (e.g., Dekel & Silk 1986; Mac Low & Ferrara 1999; Strickland & Stevens 2000), and they are also consistent with some properties of the nearby universe such as the high abundance of  $\alpha$ -elements in the hot gas of halos or the correlations between the metallicity of the hot gas and the abundances of early- and late-type galaxies in galaxy clusters, indicative that most gas outflows take place from spheroids (e.g., Silk 1997).

The values of  $\alpha_G$  and  $\epsilon_{\text{AGN}}$  are also well constrained by the observed SFH and MBHH. The larger value of  $\alpha_G$  in model S compensates the lower volume filling factor of ionized regions at high redshifts in this model compared to model D. Indeed, the SFH is averaged over the whole space, while stars lie in ionized regions only. This then leads to a smaller value of  $\epsilon_{\text{AGN}}$  in model S so as to recover the right MBH-to-spheroid mass ratio despite the larger amount of stars forming in individual spheroids.

The value of  $h_{\text{mix}}$  is the least constrained. The reason for this is that the data on HGMH show a large scatter, and, even though the data on the CGMH and STMH are better determined, these cosmic histories are quite insensitive to  $h_{\text{mix}}$ . But this result, far from being negative for our purposes, is heartwarming as the mixing of metals in the hot gas is hard to model, and the hot gas metallicity estimates are not very reliable (see Section 3).

The values of  $f_{\text{esc,G}}$  and  $f_{\text{esc,AGN}}$  are also well constrained by the observed IEHs from galaxies and AGN and, more importantly, by the priors  $z_{\text{ion,H}} = 6.0^{+0.3}_{-0.5}$  and  $z_{\text{ion,He}} \sim 2.8$ . Notice that the uncertainty of about  $\pm 0.8$  in  $z_{\text{ion,He}}$  (Ricotti et al. 2000; Schaye et al. 2000; Dixon & Furlanetto 2009; McQuinn 2009; Becker et al. 2011) has not been taken into account since any change of that order in  $z_{\text{ion,He}}$  has no significant effect in the remaining cosmic properties. The values of  $f_{\text{esc,AGN}}$  are one order of magnitude smaller than those of  $f_{\text{esc,G}}$ , which is consistent with the extra absorption of ionizing photons produced in the circumnuclear region of galaxies. Regarding the values of  $f_{\text{esc,G}}$ , they agree with the observational estimate (spanning from 0.03 to 0.07) at redshifts out to  $z \sim 5$  (see e.g., Wyithe et al. 2010 and references

therein). But these empirical estimates are quite uncertain, while all recent theoretical studies of reionization based on the observed SFH find instead  $f_{\text{esc,G}} \geq 0.20$  (Bouwens et al. 2015; Ishigaki et al. 2015; Robertson et al. 2015). About half this difference arises from the fact that those studies neglect the effect of Pop III stars (AGN have a negligible contribution to the H I-ionizing emissivity at  $z > 6$ ), and the other half from the notably larger effective value of the LyC photon production rate from galaxies,  $\xi_{\text{ion}}$ , defined through the relation  $\dot{N}_{\text{HII}} = f_{\text{esc,G}} \xi_{\text{ion}} \dot{\rho}_*$ , found in our *self-consistent* modeling compared to that found in those studies (see Figure 6). As we will show in Section 5.3, that increased  $\xi_{\text{ion}}$  value is due to a notably larger (by a factor  $\sim 2$ ) contribution of low-mass galaxies with much lower metallicities (more than one order of magnitude lower than near solar).

## 5.2. Fits to the Observed Global Properties

In Figure 7 we show the fits to the global data that result in the best models S and D. As it can be seen, both solutions recover satisfactorily the observed evolution of all cosmic properties. The only exception is the STH (panel (b)) and, to a much lesser extent, the CGH (panel (a)). Since the STH is the time-integral of the SFH, and this latter curve is very well fitted (see panel (j)), the poor fit to the STH simply reflects the well-known inconsistency between the data sets of the two cosmic histories. We will come back to this point in Section 5.3. The possible origin of the slight excess in the predicted cold gas mass density at  $z < 3$  will also be discussed in Section 5.4.

The fits to the data are particularly good in the case of model D. The predicted SFH almost fully matches the data, while in model S it only does so marginally (see panel (j)). This is reflected in the respective  $\chi^2$  values of 1.07 and 2.65.<sup>28</sup> Likewise, the predicted CGH in model D is almost as shallow as the observed curve in the redshift range covered by the data, while that in model S is slightly steeper (see panel (b)); the respective  $\chi^2$  values are 1.90 and 2.35. The predicted HGMH, CGMH, and STMH curves for the metallicities averaged over

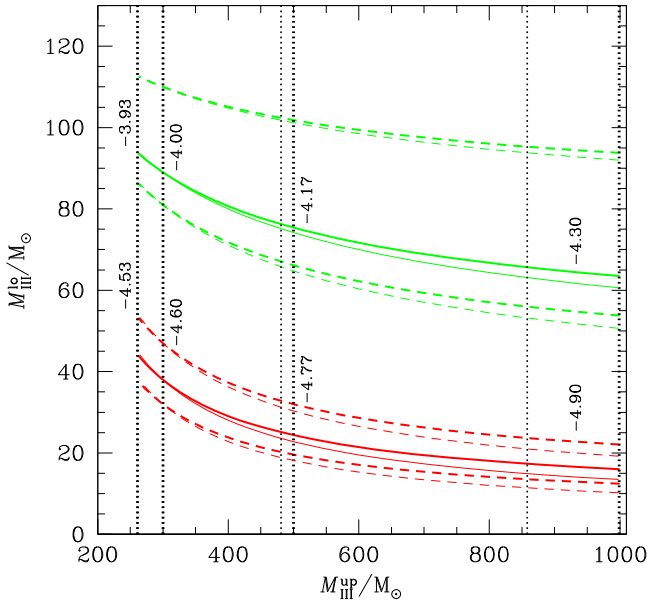
<sup>28</sup> All  $\chi^2$  values correspond to the logarithmic curves as plotted in Figure 7 and consider only points with well-measured errors bars (upper and lower bounds are excluded).

**Table 1**  
Values of the Adjusted Free Parameters (besides  $\rho_{\text{dis}} = 10^6 M_\odot \text{ Kpc}^{-3}$ ) for the Fiducial Values of the Degenerate Ones

Model	$M_{\text{III}}^{\text{lo}}/M_\odot$ [0.5, 260]	$\epsilon_B$ [0, 1]	$\epsilon_D$ [0, 1]	$\alpha_G$ [0, 1]	$\epsilon_{\text{AGN}}$ [0, 1]	$h_{\text{mix}}$ [0, 1]	$f_{\text{esc,G}}$ [0, 1]	$f_{\text{esc,AGN}}$ [0, 1]
S	$38_{-7}^{+5}$	$1.00_{-0.05}^{+0.00}$	$0.00_{-0.00}^{+0.05}$	$0.33 \pm 0.11$	$0.0016 \pm 0.0003$	$0.4 \pm 0.3$	$0.053 \pm 0.007$	$0.0053 \pm 0.0006$
D	$87_{-6}^{+9}$	$1.00_{-0.05}^{+0.00}$	$0.00_{-0.00}^{+0.05}$	$0.19 \pm 0.04$	$0.0025 \pm 0.0004$	$0.4 \pm 0.3$	$0.055 \pm 0.008$	$0.0053 \pm 0.0005$

**Table 2**  
Fiducial Values of the Degenerate Free Parameters

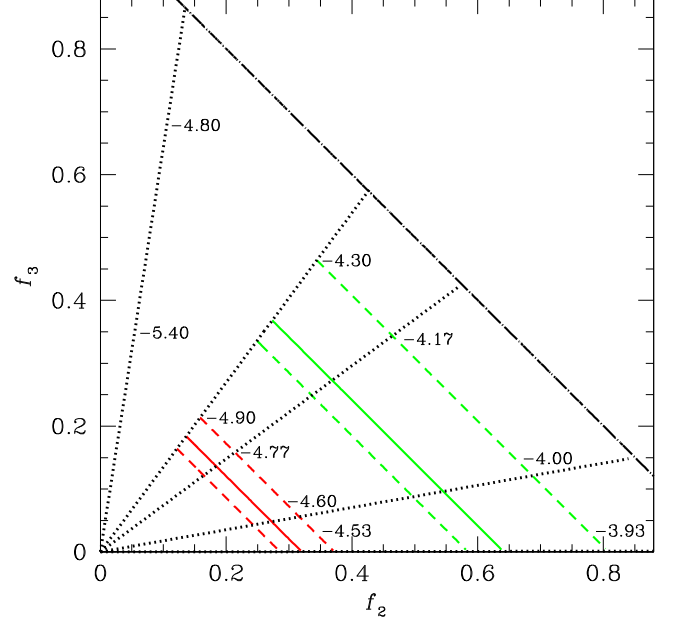
Model	$\log(Z_c/Z_\odot)$ [-5, -3]	$\alpha_{\text{III}}$ [2.15, 2.35]	$M_{\text{III}}^{\text{up}}/M_\odot$ [260, 1000]
S	$-4.6 \pm 0.1$	$2.35_{-0.10}^{+0.00}$	$300_{-40}^{+0}$
D	$-4.0 \pm 0.1$	$2.35_{-0.10}^{+0.00}$	$300_{-40}^{+0}$



**Figure 4.** Degeneracy in  $Z_c$ ,  $\alpha_{\text{III}}$ ,  $M_{\text{III}}^{\text{lo}}$ , and  $M_{\text{III}}^{\text{up}}$  leading to the same solutions with single (lower lines) and double (upper lines) reionization (solid lines for the best solutions and dashed lines for the respective bracketing solutions) for  $\alpha_{\text{III}}$  equal to 2.35 (thick lines) and 2.25 (thin lines). The vertical black dotted lines mark the maximum possible  $M_{\text{III}}^{\text{up}}$  values for the quoted logarithms of  $Z_c$  in double reionization (upper values), and single reionization (lower values).

all halos and galaxies are also substantially different in the two models. However, when the averages are restricted to the mass ranges included in the observations, the two models agree with the data (see panels (e)–(f)). Similarly, the predicted GAMH and GASH are somewhat different at high- $z$  in the two models, but they both match the data at  $z \sim 2$ –3 (see panels (h) and (i)). Last, the predicted IGMMH is larger in model D than in model S (see panel (g)). However, the empirical bounds on this quantity are not very reliable (see below). A full comprehensive picture of the interconnected behavior of all these cosmic histories will be given in Sections 5.3 and 5.4.

And what about the observational constraints on reionization? In Table 3 we give the values predicted in the best models S and D. The error bars mark those predicted in the models bracketing the two sets of acceptable solutions around them. The values of  $z_{\text{ion,H}}$  and  $d_z Q_{\text{H II}}(z=7)$  satisfy, of course, the observed constraints because they have been enforced as priors. But the values of  $\tau$  and  $y$  are checked for the first time. The values of the Compton distortion  $y$ -parameter predicted in models S and



**Figure 5.** Same degeneracy (with the solutions for single reionization on the right and those for double reionization on the left) as in Figure 4 in terms of  $Z_c$ ,  $f_2$ , and  $f_3$ . The black dot-dashed line, with equation  $f_2 + f_3 = 1$ , delimits the region with possible solutions. The diagonal black dotted lines mark the minimum possible  $f_2/(f_2 + f_3)$  values for the quoted logarithms of  $Z_c$  in double reionization (outer values), and single reionization (inner values).

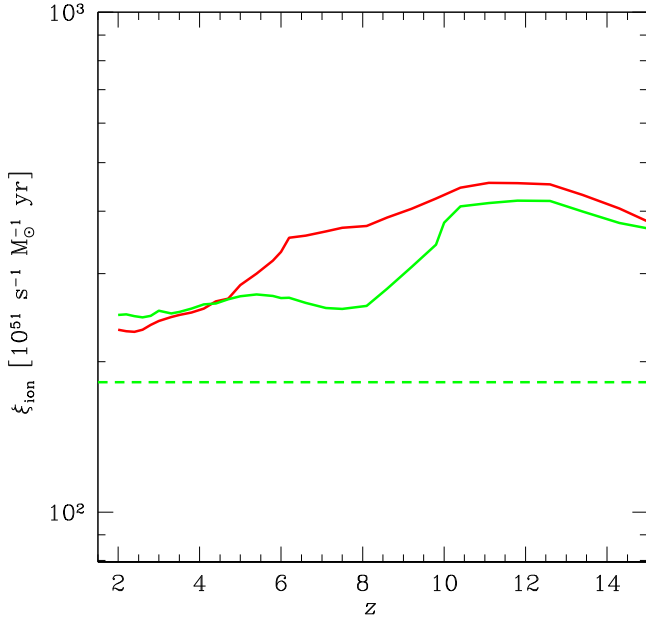
D pass the test without trouble: they are one order of magnitude lower than the observed upper limit. The predicted values of  $\tau$  are also consistent with the empirical estimates drawn from the *WMAP9* and *Planck3* data at the  $1\sigma$  level. However, while the value predicted in model S still agrees at the  $1\sigma$  level with the estimate of  $\tau = 0.058 \pm 0.012$  recently reported by the Planck team, that in model D only does at the  $3\sigma$  level. We remind the reader that the way the condition  $d_z Q_{\text{H II}}(z=7) < 0$  was enforced in this model warrants that the redshift of first complete reionization is as small as possible, so the only way to further decrease the predicted value of  $\tau$  is to relax that condition. This possibility cannot be discarded, indeed: the typical outflow velocity of gas clouds in LAEs could decrease with increasing  $z$  so that the absence of LAEs at  $z > 7$  would not imply that  $Q_{\text{H II}}(z)$  is decreasing at  $z = 7$ .<sup>29</sup> In any event, there is also the caveat that the value of  $\tau$  inferred from the CMB anisotropies under the instantaneous reionization approximation is quite uncertain.

### 5.3. Checking the Differential Properties

But the fact that models S and D fit all the observed global properties, in particular the SFH and MBHH, does not necessarily mean that they also recover the observed galaxy

<sup>29</sup> It could be increasing, and the outflow velocity could be decreasing rapidly enough to balance that increase.





**Figure 6.** Effective LyC photon production rates predicted in models S and D for galaxies with any mass and metallicity (solid lines) and for galaxies with stellar masses larger than  $\sim 5 \times 10^9 M_\odot$  and near solar metallicities only (dashed lines; the predictions from the two models superpose). The dashed line coincides, except for a factor 1.61 due to the different IMF used, with the LyC photon production rate predicted in the reionization models by Robertson et al. (2015), Ishigaki et al. (2015), and Bouwens et al. (2015).

stellar or MBH MFs. In the present section we check that, describing in detail the formation and properties of galaxies and MBHs of different masses.

At high- $z$  halos are barely massive and their concentration is high, so the hot gas cools very efficiently. But, for this to happen, halos must first trap gas, that is, their virial temperature must be higher than the IGM temperature,  $T_{\text{IGM}}$ . The minimum mass of halos trapping gas at  $z$  is

$$M_{\text{trap}}(z) \approx \left\{ \frac{[3k T_{\text{IGM}}(z)/G]^3}{4\pi \Delta(z) \bar{\rho}(z)/3} \right\}^{1/2}, \quad (17)$$

where we have taken into account that the inner mean density of halos at  $z$  is a factor  $\Delta(z)$  (close to 178 at high- $z$ ) times the mean cosmic density  $\bar{\rho}(z)$ . Therefore, *new* luminous objects (Pop III star clusters or galaxies) will form as long as *new* halos will trap gas (with undercritical or overcritical metallicity, respectively), that is, as long as  $M_{\text{trap}}(z)$  will decrease or increase less rapidly than the always increasing typical halo mass of collapse at  $z$ ,  $M_c(z)$ , equal to the mass for which the rms density perturbation equals the critical overdensity for ellipsoidal collapse at  $z$  (e.g., Juan et al. 2014b).  $M_c(z)$  is fully determined by cosmology, so the formation of new luminous objects depends only on the evolution of  $T_{\text{IGM}}(z)$ .

The temperature of the neutral cosmic gas decouples from the CMB temperature at  $z \sim 150$ , when Compton heating of residual electrons from CMB photons becomes negligible. Then, the cosmic neutral gas cools adiabatically as  $(1+z)^2$  until the first Pop III stars form. The X-ray background they produce (mostly through PISNe explosions) heats the neutral regions above the CMB temperature by Compton scattering from the diffuse residual electrons. However, Compton cooling

from CMB photons prevents such a heating from being too marked, and  $T_{\text{IGM}}(z)$  diminishes essentially as  $1+z$  until Compton cooling progressively declines by the dilution of CMB photons. Then, the temperature of neutral regions keeps on diminishing almost as  $(1+z)^2$  as X-ray heating is insufficient to balance adiabatic cooling (see Figure 7, panel (I)).

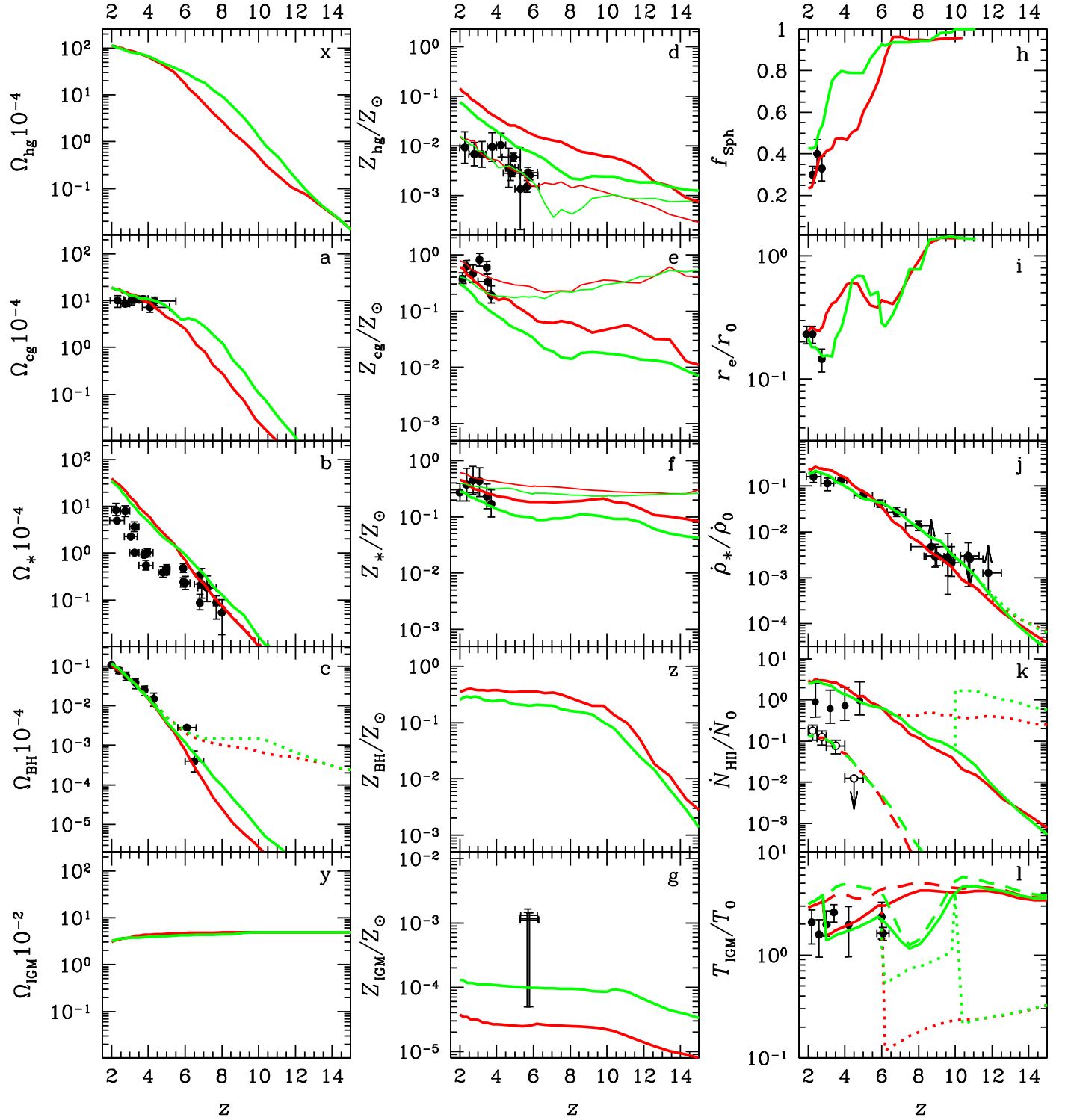
Such an evolution of  $T_{\text{IGM}}(z)$  in neutral pristine regions implies that  $M_{\text{trap}}(z)$  decreases roughly as  $(1+z)^{3/2}$  all the time except for a short intermediate period where it is kept essentially constant. Meanwhile,  $M_c(z)$  starts growing very rapidly, and progressively slows down from  $(1+z)^{-2}$  above  $z = 20$  to  $(1+z)^{-1}$  below  $z = 6$ , being proportional to  $(1+z)^{-3/2}$  at an intermediate redshift of about 10. As a consequence, Pop III star clusters are permanently forming in neutral *pristine* regions. The situation is different, however, in the neutral *polluted* nodes that develop in model D during recombination. There, the temperature is kept much more constant (see Figure 7, panel (I)),<sup>30</sup> causing  $M_{\text{trap}}(z)$  to increase roughly as  $(1+z)^{-3/2}$ , that is, more rapidly than  $M_c(z)$  since  $z \sim 10$ . Consequently, no Pop III stars will form in this model after that redshift.

More specifically,  $M_c(z)$  starts from a very tiny value, and does not reach what is believed to be the neutralino free streaming mass,  $10^{-6} M_\odot$ , until  $z \sim 23$ . This means that the only halos present at any higher redshift (necessarily more massive than the free streaming mass of DM particles) are much more massive than the typical collapse mass at that  $z$ . Thus their abundance is insignificant. In fact, the first mini-halo (of  $\sim 5 \times 10^5 M_\odot$ ) able to trap gas and form a Pop III star cluster (of  $\sim 2000 M_\odot$ ) within the *current* horizon is found at  $z \sim 48$ .

As the temperature of the neutral IGM decreases, the mass of mini-halos able to form Pop III star clusters diminishes, and, at  $z \sim 45$ , it is low enough for such halos to lose their metal-enriched gas into ionized bubbles when the most massive Pop III stars explode in PISNe. Then the IGM metallicity in ionized regions rapidly increases above  $Z_c$ . From that moment, when the ionized metal-enriched gas is trapped by halos, it rapidly undergoes atomic cooling, and gives rise to the formation of normal galaxies. But the IGM in ionized bubbles is also photo-heated to a much higher temperature than in neutral regions, so halos able to trap ionized gas are again very massive and rare. In fact, at those redshifts most galaxies form instead in slightly less massive halos (of  $\gtrsim 10^6$ – $10^7 M_\odot$ ) that retained part of the hot gas when Pop III stars exploded. Since the amount of gas remaining in such halos can be very small, most of those galaxies are “failed” in the sense that they have very low stellar masses (from  $10^2 M_\odot$  to  $10^5 M_\odot$ ).

As halos able to trap *fresh* metal-enriched ionized gas become increasingly abundant, more and more galaxies form with stellar masses above  $10^5 M_\odot$ . This leads to the development of a bimodal MF. But the formation of normal galaxies causes ionized regions to be photo-heated to an approximately constant temperature, so  $M_{\text{trap}}(z)$  finally increases at the same rate,  $(1+z)^{-3/2}$ , as  $M_c(z)$  at  $z \sim 10$ .

<sup>30</sup> This feature should allow one to tell the difference between single and double reionization by looking at the IGMTH. Although there is no temperature estimates at  $z > 6$ , the spin temperature of hydrogen atoms is already coupled the neutral gas temperature through the Wouthuysen–Field effect, so such a difference between models S and D should leave an unequivocal signal in 21 cm line observations.



**Figure 7.** Evolution of galaxies and IGM predicted in models S and D, compared to the observational data shown in Figure 1. We plot the contribution from normal galaxies (solid lines) and normal galaxies plus Pop III star clusters (dotted lines); in panel (k) we also plot the contribution of AGNs (dashed lines). In panels (d), (e), and (f), we plot the average metallicities in halos and galaxies within the mass ranges covered by the data (thin solid lines), as well as the average metallicities over all halos and galaxies (thick solid lines) with no corresponding data. In panels (h) and (i) predictions are only shown at the redshifts where galaxies with masses around  $10^{11} M_{\odot}$  are abundant enough. In panel (l) we show the average IGM temperatures in singly ionized (solid lines), doubly ionized (dashed lines), and neutral (dotted lines) regions, the latter shifted a factor 500 upward at  $z$  higher than the redshift of first (or unique) complete ionization. Note that the sum at any  $z$  of the densities in the left-hand panels gives the total baryon mass density at that  $z$  (scaled to the constant critical density at  $z = 0$ ), and the sum at any  $z$  of the product of the metallicities in the middle panels times the corresponding densities in the left panels gives the cumulative mass density of metals produced until that  $z$ .

Consequently, the formation of new galaxies with stellar masses above  $\sim 10^6 M_{\odot}$  (the lower mass limit is slowly increasing with decreasing  $z$ ) is halted roughly at that redshift.

From that moment, the evolution of luminous sources differs in the two models. In model S, Pop III star clusters keep on forming in neutral pristine regions, and galaxies keep on

**Table 3**  
Reionization Constraints

Model	$z_{\text{ion,H}}$ $6.0^{+0.3}_{-0.5}$	$d_z Q_{\text{H II}}(z = 7)$ <0	$0.084 \pm 0.017^a$	$y \times 10^5$ <5
S	$6.0^{+0.3}_{-0.5}$	$-0.19^{+0.02}_{-0.02}$	$0.072^{+0.004}_{-0.005}$	$0.25 \pm 0.01$
D	$6.0^{+0.3}_{-0.5}$	$-0.13^{+0.03}_{-0.02}$	$0.102^{+0.001}_{-0.001}$	$0.26 \pm 0.01$

**Note.**

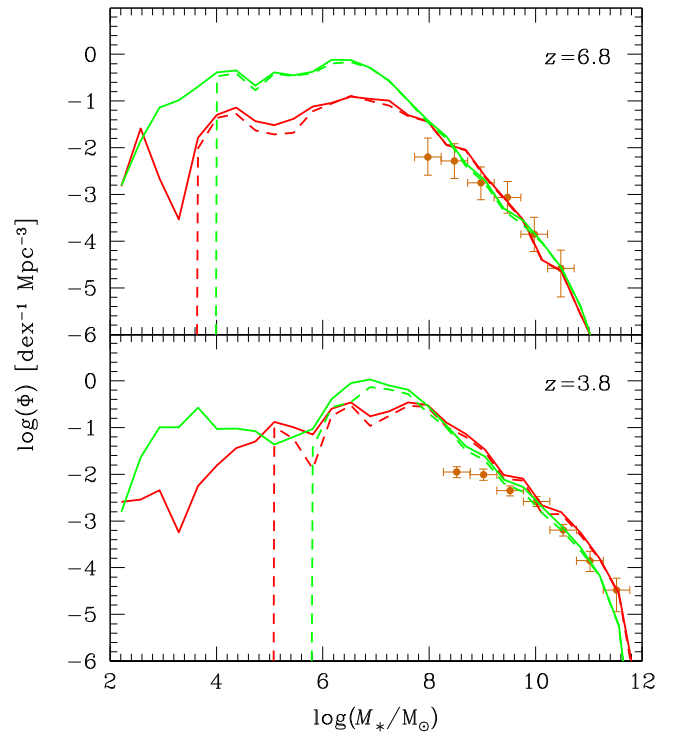
<sup>a</sup> This empirical value of  $\tau$  combines those inferred from the *WMAP* nine-year and the *Planck* three-year data.

growing, although not forming, in ionized ones. While, in model D, Pop III stars stop forming at  $z \sim 10$  when the first complete reionization takes place. (Remember that no Pop III stars form in the neutral polluted regions that develop during the subsequent recombination period.) There is thus a sudden lack of ionizing photons that could definitely interrupt the reionization process. Fortunately, recombination yields a big dip in the IGM temperature of ionized regions (see Figure 7, panel (l)) that allows halos with masses as small as  $10^8 M_\odot$  to trap fresh ionized gas. As explained next, the metallicity of this gas is still overcritical, so an abundant amount of normal galaxies then form that partially replace Pop III stars in the production of ionizing photons. This will allow reionization to recover by  $z \sim 7.2$ .

Indeed, the IGM metallicity in ionized regions is initially kept roughly constant due to the fixed proportion of metals ( $\propto f_2$ ) and ionizing photons ( $\propto f_2 + f_3$ ) provided by Pop III stars. But, once normal galaxies begin to form, there is a larger contribution of ionizing photons than of metals (remember that metals ejected from galaxies mainly stay within halos), so the metallicity of ionized regions begins to decrease. This does not prevent, of course, the *total average* IGM metallicity from growing (Figure 7, panel (g)) through the increasing weight of ionized regions. But the metallicity of ionized regions themselves can become undercritical. This is actually what happens in model S where the minimum value of  $f_2/(f_2 + f_3)$  just needs to warrant the overcritical metallicity when ionized bubbles form. (The undercritical metallicity of these regions at  $z < 10$  in model S has no effect in the reionization process because no galaxies can form anyway in this model since  $z \sim 10$ .) The situation is different, however, in model D where the IGM temperature decreases since  $z \sim 10$ . Provided the minimum value of  $f_2/(f_2 + f_3)$  is large enough (the maximum value of  $M_{\text{III}}^{\text{up}}$  low enough) to warrant that the IGM metallicity in ionized regions is still overcritical at  $z \sim 10$ , new galaxies will form, indeed, in this model from that redshift to about  $z \sim 7$  when the temperature of the ionized IGM increases again at a high enough rate for  $M_{\text{trap}}(z)$  to grow more rapidly than  $M_c(z)$ .<sup>31</sup> The minimum mass of trapping halos in model D during that period is the largest ever been, and the metallicity of the initial hot gas is the lowest, so the cooling rate reaches very low values. As a consequence, a large fraction of these second-generation normal galaxies have very small stellar masses, similar to those of the above-mentioned failed galaxies.

Since  $z \sim 10$  in model S or since  $z \sim 7$  in model D, no new galaxies form. Galaxies just grow at the center of halos through

<sup>31</sup> In fact, since the maximum possible value of  $M_{\text{III}}^{\text{up}}$  is close to the lower limit of  $260 M_\odot$ , the average IGM metallicity at  $z \sim 10$  cannot be much larger than  $Z_c$  either (see Figure 7, panel (g)). Therefore, another characteristic of double reionization is that the IGM metallicity at  $7 \lesssim z \lesssim 10$  is close to  $Z_c$ .

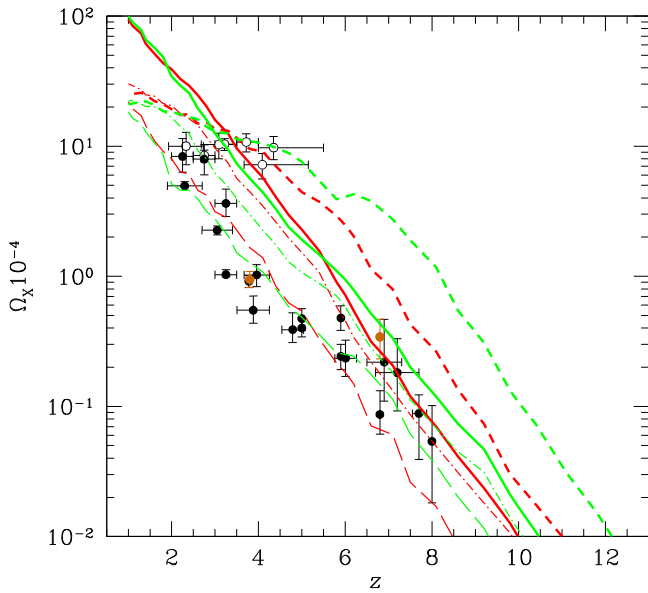


**Figure 8.** Galaxy stellar MFs predicted for all galaxies (solid lines) and central galaxies (dashed lines) in model S and model D at two representative redshifts. Dots with error bars give the MFs inferred at those redshifts by González et al. (2011) from galaxy UV LFs, shifted consistently with the stellar mass densities in model D (see text).

the accretion of new cooled gas and satellite captures. Of course, some of them freeze out as they become satellites of other central galaxies when their host halos are captured by other halos. Consequently, at any  $z$  lower than 7, there are in both models galaxies with the whole spectrum of stellar masses. The only difference is that, while in model S all the very low-mass satellites are failed galaxies, in model D many of them are second-generation galaxies formed at  $z < 10$ . Notice that, even though halos where those two kinds of galaxies form have similar masses,  $M_c(z)$  is much greater at  $z < 10$  than at the formation of failed galaxies, so those second-generation low-mass galaxies are more abundant than failed ones.

Figure 8 shows the galaxy stellar MFs that result in models S and D at redshifts  $z = 3.8$  and  $z = 6.8$ . To compare them with the empirical MFs inferred by González et al. (2011) at the same redshifts we need to account for the above-mentioned inconsistency between the SFH and STH data sets. Indeed, at high- and low- $z$ , the predicted STH curve, consistent with the empirical SFH one, is close to the upper envelope of the stellar mass density estimates, and between  $z \lesssim 6$  and  $z \sim 4$  it is even higher (see Figure 9). Wilkins et al. (2008) put forward that such an inconsistency could be due to a  $z$ -varying IMF. However, the largest discrepancy is found at an intermediate redshift range, which would require the IMF to have a strange non-monotonous trend. A simpler explanation is that the correction for dust attenuation achieved in the galaxy UV luminosities is typically underestimated. Galaxies in the redshift range  $4 \lesssim z \lesssim 6$  would be substantially brighter than believed, the effect being less marked at higher and lower redshifts because of the lower metallicity of galaxies and the more accurate correction achieved, respectively. But, whatever





**Figure 9.** Same Pop II STHs (thick solid lines) and CGHs (thick dashed lines) as in panels (b) and (a) of Figure 7 but extended down to  $z = 1$ . Also shown are the stellar mass densities in disks (thin long-dashed lines) and in spheroids (thin dot-dashed lines). To avoid confusion the cold gas mass density estimates are here plotted as open circles. The stellar mass density estimates obtained by González et al. (2011) from the galaxy stellar MFs shown in Figure 8 are plotted as brown dots.

the real cause of that inconsistency, provided all the observed galaxies are similarly affected, the *shape* of the MFs derived from the UV LFs would be right except for an horizontal shift. Specifically, according to the stellar mass density estimated inferred by González et al. (2011) at redshifts 3.8 and 6.8 and the theoretical STH curve consistent with the empirical SFH one, the MFs used by González et al. (2011) to infer those estimates should be shifted in model D (S) by a factor of about 6.0 (7.9) and 1.3 (0.8), respectively. Taking into account that the extrapolations of those empirical MFs were underestimated by factors of 1.6 (1.3) and 2.1 (1.4), respectively, the actual shifts to be applied at those redshifts are 9.5 (10.3) and 2.7 (1.2), respectively. As can be seen in Figure 9, these shifts make the empirical MFs fully match indeed the predicted ones except for the incomplete low-mass bins (see below). Thus the predicted galaxy stellar MFs are in very good agreement with the observations. The fact that the MF predicted at  $z = 3.8$  also matches the Schechter fit to the original (with no shift) galaxy stellar MF recently derived by Caputi et al. (2015) at  $3 \leq z \leq 4$  assuming a constant IMF also gives strong support to the interpretation that the inconsistency between the SFH and STH data sets is due to a deficient correction for dust attenuation. According to our results in model D (S), the UV luminosities of galaxies inferred by González et al. (2011) at  $z = 3.8$  and  $z = 6.8$  would be underestimated by 2.5 (2.7) mag and 1.0 (1.1) mag, respectively.

Regarding the behavior at lower masses of the predicted MFs, we find the turnover at  $\sim 3.5$  dex from the knee in model S, and at  $\sim 4.5$  dex in model D. These results are consistent with the latest UV LFs of galaxies observed at  $z \sim 3$  (Alavi et al. 2016; Livermore et al. 2016), which find that the turnover is not yet detected at  $\sim 6.5$  mag (or  $\sim 2.6$  dex in mass) from the knee. We remark that all previous models of galaxy formation predicted the turnover at just  $\sim 3.5$  mag ( $\sim 1.4$  dex in mass) from the knee (Jaacks et al. 2013; Kuhlen et al. 2013), which is

contradicted by these observations. In this sense, our model is the first to predict MFs consistent with the observations. According to it, such LFs should still go  $\sim 2.25$  mag or  $\sim 4.75$  mag deeper in models S or D, respectively, for the turnover to be finally detected.

Model D also predicts the existence of a second bump at very low masses in the galaxy stellar MFs at  $z = 3.8$  caused by the second-generation low-mass galaxies that become satellites soon after their formation. This bump is far from the current observational limits of the galaxy stellar MFs at those redshifts. However, given the small mass of those satellites, a large fraction of them should still survive at  $z = 0$ . In model S, there is no such a bump, although the abundance of very low-mass galaxies is also quite high. Thus a prediction of our model, regardless of the version, is that it should be possible to detect at  $z = 0$  very old satellites with masses between  $10^2 M_\odot$  and  $10^6 M_\odot$ . We describe next the expected metallicity, morphology, and size of those predicted relics.

The hot gas metallicity increases in parallel to the cold gas and stellar metallicities (see Figure 7, panel (d), (e) and (f)). The only effect that partly balances that increase is the accretion by halos of low-metallicity IGM in ionized regions. That increase is more marked in model S than in model D due to the steeper SFH caused by the larger  $\alpha_G$  value, and hence, the larger amount of stars that are permanently forming in galaxies. Another more subtle difference between the two models is the small dip in the hot gas metallicity in model D after  $z \sim 10$  due to the large amount of fresh gas that is being incorporated by halos during recombination owing to the dip in the temperature of the ionized IGM. This translates, of course, in less marked dips in the CGMH and STMH of model D. The stellar and cold gas metallicities of the second-generation galaxies formed in this model during recombination, similar to those of the primordial failed galaxies formed in both models, are of about  $10Z_c$ , i.e., significantly lower than average.<sup>32</sup> On the contrary, the metallicities of galaxies more massive than  $5 \times 10^9 M_\odot$  are, of course, larger than average (see also Figure 7, panels (e) and (f)) and show a well-understood behavior with  $z$ : at the highest redshift where those galactic masses are assembled, such galaxies are extraordinarily massive, having been formed in the merger of evolved high metallicity galaxies; then the metallicities slowly decrease as the masses become increasingly ordinary and such galaxies form through other channels (including accretion of hot gas with lower metallicity); and, when the masses of the galaxies approach the average galaxy mass at  $z \sim 4$ , they begin to increase just as the average metallicities do since  $z \sim 6$ .

The morphology of galaxies follows from the following general trends. When a galaxy forms, the gas it collects comes from the very central part of the halo, so it has little angular momentum, and it is piled up in a spheroid. Consequently, the failed galaxies in models S and D as well as the second-generation low-mass galaxies in model D are spheroids. But, soon after their formation, the morphology of galaxies evolves at the center of halos. At very high- $z$ , when cooling is very efficient, galaxies usually collect large amounts of cold gas that render disks unstable. In addition, halos are small, and galaxy mergers frequent, so whenever disks are stable they are short-lived and never reach large radii. But, as halos grow, cooling

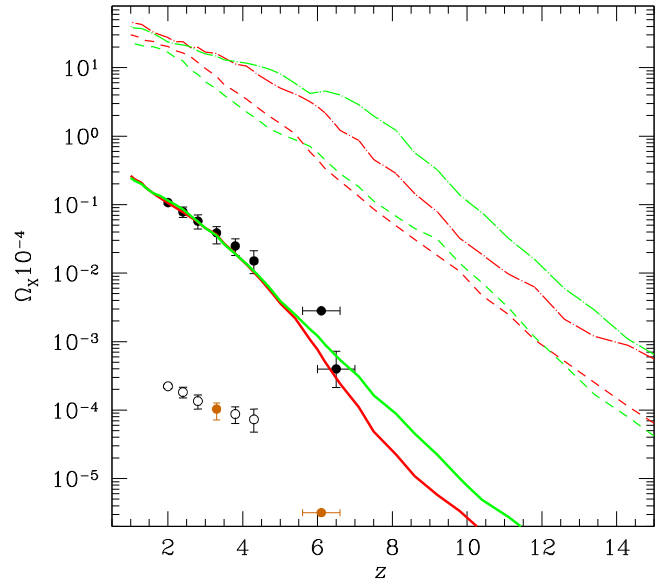
<sup>32</sup> These metallicities are nonetheless larger than the metallicity of the initial gas due to the rapid increase in the metallicity of the gas produced during star formation.

becomes less efficient and galaxy mergers less frequent, so disks grow larger and are longer-lived. As a consequence, the fraction of spheroid-dominated galaxies slowly decreases with decreasing  $z$ . On the other hand, at any given  $z$ , the most massive galaxies tend to be spheroids arising from galaxy mergers of massive objects, while disks preferentially lie in intermediate-mass galaxies. These trends explain the observed fraction of spheroid-dominated objects with masses above  $\sim 10^{11} M_{\odot}$  at  $z \sim 2-3$ . We remark, however, that about 25% of galaxies at those redshifts are red dead ellipticals with no star formation. These galaxies have been excluded from the results shown in panel (h) of Figure 7, as they would go undetected. Therefore, the real fraction of spheroid-dominated massive galaxies at those redshifts is  $\sim 0.65$  rather than  $\sim 0.40$ .

The size of galaxies increases in parallel to their mass from the beginning. In fact, when they form as spheroids from a gas cloud with small angular momentum, even though the metallicity of the gas is low (about  $Z_c$ ), they undergo a large dissipative contraction until they reach the density of molecular clouds. This leads to stellar systems with the typical density of globular clusters. The size of more evolved spheroids depends on the amount and metallicity of the gas they collect (see Equation (8)). The result is an evolution of the average equivalent radius (measured as the square root of the mass-weighted squared equivalent radii of the bulge and the disk, if any) of spheroid-dominated galaxies with masses above  $10^{11} M_{\odot}$  that recovers the observations at  $z \sim 2-3$  (Figure 7, panel (i)). Specifically, when those very massive galaxies appear for the first time in dry mergers of massive spheroids, their equivalent radius is even slightly larger than that of present-day slightly contracted spheroids of the same mass; this radius begins to decrease as the mass becomes increasingly ordinary and those galaxies form from unstable disks and wet mergers; it then increases again up to a secondary maximum caused by the increasing lack of gas followed by the increase of the gas metallicity at  $z \sim 5$  (see Figure 7, panel (e)); and it finally increases until  $z = 0$  due to the definite small amount of gas present in galaxies.

Thus the properties of very low-mass galaxies are essentially the same in both models. They are among the first objects with Pop II stars to form; they are spheroids with no gas; they are very compact, with the typical density of globular clusters; and their Pop II stars have very low metallicities (of about 10–100 times  $Z_c$ , with  $Z_c$  equal to  $\sim 0.25 \times 10^{-4} Z_{\odot}$  in model S, and  $\sim 10^{-4} Z_{\odot}$  in model D). On the other hand, those of them that froze out as satellites soon after formation have very large baryon-to-DM mass ratios as their original DM halos are truncated by the potential well of the also quite dense capturing halo near the region where all the stars are strongly concentrated. And the fraction of those satellites that will be captured a long time afterward (because of the slight dynamic friction they suffer) by evolved, much less dense galaxies will survive inside them as small independent dynamical entities.

Clearly, all these properties greatly resemble those of present-day globular clusters. Since such objects are often seen to freely orbit within halos, they should have formed outside galaxies and have a cosmological origin similar to them. It is thus not surprising that a self-consistent model such as AMIGA that monitors the formation and evolution of luminous objects since the dark ages does not miss them. Notice that, if this is indeed the origin of globular clusters, the

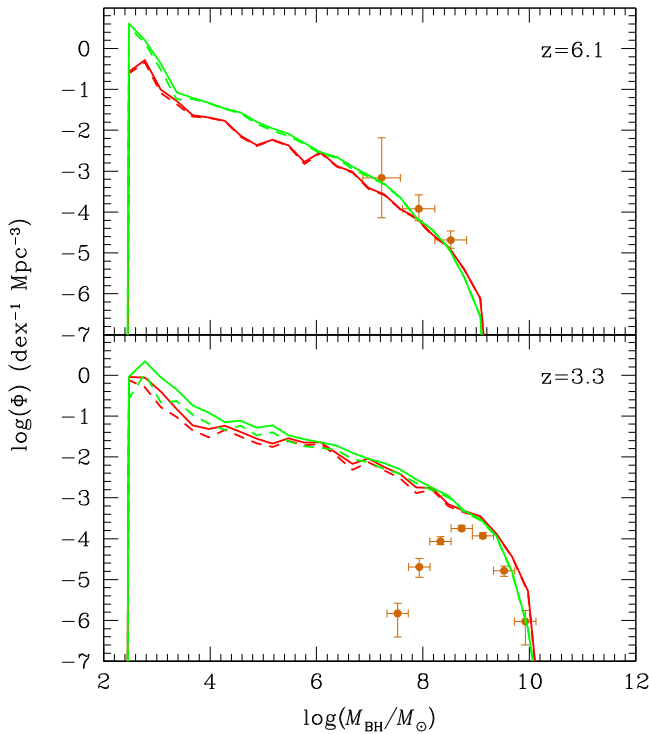


**Figure 10.** Same MBHs (thick solid lines) as in panel (c) of Figure 7 but extended down to  $z = 1$ . Also shown are the spheroid mass density history (thin dashed lines), and disk mass density history (thin dot-long-dashed lines). Empty circles give the MBH mass density estimates obtained from rough (incomplete and uncorrected for obscuration and inactivity) MBH MFs. Brown dots at  $z = 3.3$  and  $z = 6.1$  correspond to the estimates derived by Kelly et al. (2010) and Willott et al. (2010a), respectively, from the MFs used in Figure 11.

larger abundance of such objects compared to massive galaxies is more consistent with model D than with model S.

Let us now turn to the MFs of MBHs. MBHs grow since the remnant mini-MBHs of Pop III star clusters begin to develop at the center of galaxies (Figure 7, panel (c)). There, they merge with the MBHs brought by accreted or merged galaxies having reached the center of the galaxy by dynamical friction, and accrete new material. Indeed, when some amount of gas reaches the spheroid it fuels the MBH. The associated AGN is then lit, and begins to reheat the circumnuclear gas, quenching the star formation that is going on in the spheroid. The more massive the MBH, the larger both the AGN bolometric luminosity and the fraction of gas reheated and expelled from the spheroid back into the halo. This leads to a self-regulated growth of MBHs and stellar spheroids, which tends toward a constant MBH-to-spheroid mass ratio as observed (Figure 10). This thus gives strong support to the mechanism proposed by Di Matteo et al. (2005) included in AMIGA for the coupled growth of MBHs and spheroids.

The parallel evolution of spheroids and MBHs goes together with that of disks. Of course, when gas replenishment begins to decline by  $z \sim 7$  (see Section 5.4), the disk and spheroid mass density histories become slightly shallower. (The change of slope is a little less marked in spheroids than in disks due to their unchanged behavior in major mergers.) But this has a minor effect in the parallel evolution of the spheroid and MBH stellar mass densities because mergers become at the same time increasingly dry. Indeed, in dry mergers spheroids grow through the addition of stars belonging to the merging galaxies, while MBHs grow by coalescence of the respective MBHs and, since the MBH-to-spheroid mass ratio is the same in all the progenitors, it remains unchanged in the final object. As a consequence, the MBH and spheroid mass ratio will be kept essentially unaltered until  $z = 0$ .



**Figure 11.** MFs of MBH in central galaxies (dashed lines) and in all galaxies (solid lines) predicted in models S and D at two representative redshifts. Dots with error bars give the MBH MFs inferred at those redshifts by Willott et al. (2010a; top panel) and Vestergaard et al. (2008; bottom panel) from optical AGN LFs, shifted consistently with the MBH mass densities in model D and a well-suited  $f_{\text{FWHM}}$  value (see text).

Figure 11 shows the MBH MFs predicted at redshifts  $z = 3.3$  and  $6.1$ . As in the case of the galaxy stellar MFs, to compare these MFs to the empirical ones derived at the same redshifts by Willott et al. (2010a) and Vestergaard et al. (2008) we must take into account any effect that could affect the latter. As mentioned, the observed AGN LFs are greatly affected by incompleteness, obscuration, and inactivity, which cause the MBH mass density estimates derived from them to be much lower than expected from the observed constant MBH-to-spheroid mass ratio. Incompleteness affects in an uneven way the different mass bins of the MBH MFs, but their large mass end should be complete. Obscuration is also larger for low-mass objects (Ueda et al. 2003), but the most massive bins should be similarly affected. Thus a vertical shift by a factor  $s_1$  of a few hundreds should be enough to correct the empirical MBH MFs for obscuration and inactivity. In addition, there is the uncertainty in the MBH mass estimates themselves coming from the poorly determined virial factor,  $f_\sigma$  or  $f_{\text{FWHM}}$ . For instance, the MBH masses derived by Willott et al. (2010a) and Vestergaard et al. (2008) at  $z = 6.1$  and  $3.3$ , respectively, from emission-line reverberation assume  $f_{\text{FWHM}} = 1.17$  (Collin et al. 2006), while, according to Yong et al. (2016), this factor could be up to a factor of five smaller. Consequently, the empirical MBH MFs derived by those authors might also need to be shifted horizontally toward large masses by a factor  $s_2$  in the range  $1 \geq s_2 \geq 0.2$ . The value of  $s_2$  is unknown, but it should be independent of  $z$ . On the other hand, the shifts  $s_1/s_2$  these effects would cause to the MBH mass density estimates derived by Kelly et al. (2010) and Willott et al. (2010a) should make them match the predicted MBH curve consistent with the observed MBH-to-spheroid mass ratio, equal to  $\sim 400$  and

$\sim 250$  at  $z = 3.3$  and  $z = 6.1$ , respectively (see Figure 10).<sup>33</sup> Taking into account the different degrees of completion in the extrapolated MFs used in the calculation of those MBH mass density estimates, the vertical shifts to be applied to the corresponding MFs at  $z = 3.3$  and  $z = 6.1$  should be 1.58 and 1.70 times larger, respectively. Therefore, with the only freedom of the  $z$ -independent value of  $s_2$  (with  $1 \geq s_2 \geq 0.2$ ), it should be possible to shift the empirical MFs at the two redshifts so as to match those predicted at the two redshifts. As shown in Figure 11, this is indeed what we find for  $s_2 = 0.33$  (or  $f_{\text{FWHM}} = 0.39$ ). Notice that, as a consequence of the small correction in  $f_{\text{FWHM}}$ , the upper mass limit of MBHs at  $z = 3.3$  becomes  $\sim 10^{10} M_\odot$ , which is more reasonable than the original value of  $\sim 3 \times 10^{10} M_\odot$  if we compare with such an upper limit at  $z = 0$ .

#### 5.4. Reliability of the Model

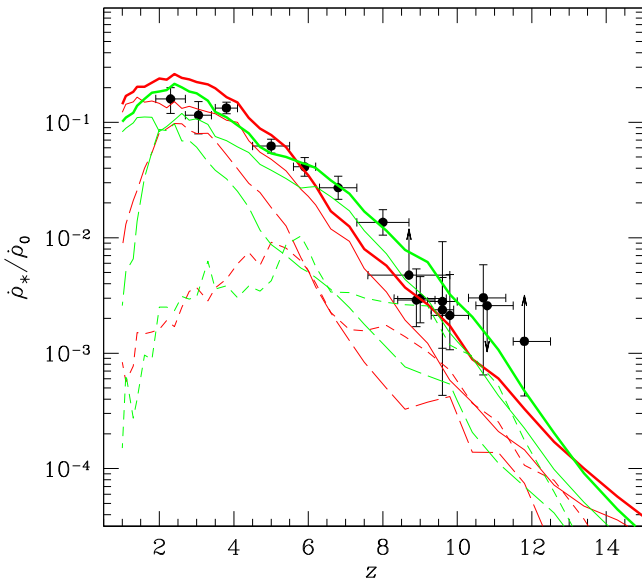
It might be argued that the model could be wrong and the solutions found for some appropriate values of the free parameters still recover the observed global and differential cosmic properties at  $z > 2$ . But such good results have been achieved with *less than one degree of freedom per curve*, which is unbelievable for any model that would not include the right physics. To see this we must simply think, for instance, in the fitting by means of a well-suited function with no physics of *one single monotonous curve* such as the observed SFH at  $z > 2$  requires four parameters (e.g., Robertson et al. 2015). Furthermore, not only does our model with only nine degrees of freedom provide very good fits to the dozen observed curves at  $z > 2$  (plus the galaxy stellar and MBH MFs), but it also automatically recovers all their extreme and saddle points at  $z \sim 2$  with no need to change the value of any parameter.

As mentioned, disks become increasingly abundant as  $z$  decreases, so the cold gas mass density increases (Figure 7, panel (a)). But cooling becomes increasingly inefficient, and by  $z \sim 7$  the rate of gas storage in disks begins to decline and the CGH becomes increasingly shallower. This effect is particularly marked in model D, where the CGH reaches a plateau between  $z \sim 5$  to  $z \sim 3$ . The observed CGH is also flat at those redshifts, but the plateau reaches  $z \sim 2$  (see Figure 7, panel (b)). There are several possible causes for that small discrepancy, such as the fact that the empirical cold gas mass density estimates do not include molecular gas or that observations may be biased toward large (massive) DLAs, which have the lowest gas content. But the most likely reason is that satellite interactions, neglected in the present work, begin to have a significant impact in the amount of gas present in disks. In any event, the predicted stagnation of cold gas around  $z \sim 4$  corresponds to a saddle point, not to a maximum, in agreement with observation (e.g., Péroux et al. 2003; Prochaska et al. 2005). At lower redshifts the temperature of the singly and doubly ionized IGM begins to diminish (see Figure 7, panel (l)), causing a rather constant slope in the HGMH, and the continuation of cooling.

Of course, the SFR density of ordinary (Pop II) stars increases in parallel to the cold gas mass density (see Figure 9), so when gas replenishment begins to decline by  $z \sim 7$ , the SFR density also slows down. Until that redshift, spheroids were much more abundant than disks, which never collected large amounts of gas. Nonetheless, disks were small, so their

<sup>33</sup> These values essentially hold for both models D and S.

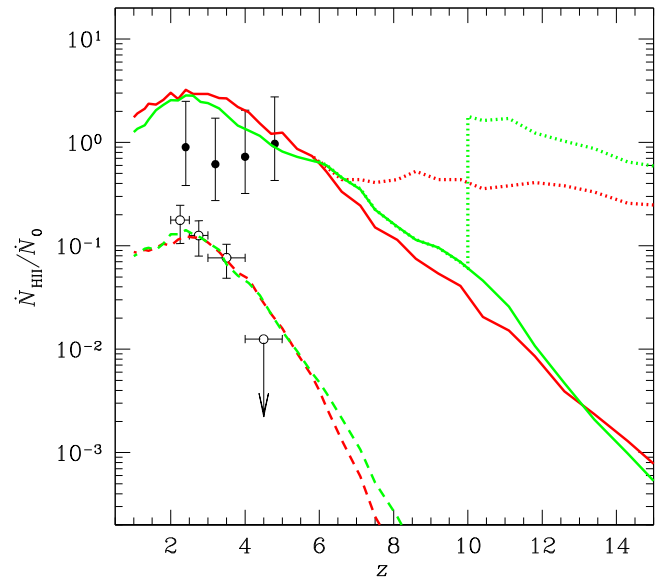




**Figure 12.** Same Pop II SFHs (thick solid lines) as in panel (i) of Figure 7 but extended down to  $z = 1$ . Also shown are the contributions from disks (thin solid lines) and spheroids arising from galaxy mergers (thin short-dashed lines) or direct infall of cold gas in unstable disks (thin long-dashed lines).

dynamical time was short, and the SFR density in disks was notable (see Figure 12). The small amount of gas in disks also prevents the SFR density in mergers from being too large. It is just comparable to that in unstable disks at high- $z$ , and reaches a maximum at  $z \sim 6$ . This does not mean that mergers are less frequent from that moment; they just become increasingly dry. As the cooling rate declines, the SFR in *individual* unstable disks also does, but the corresponding *global* SFR density keeps on increasing because of the increasing abundance of spheroids contributing to it. The same happens with disks: even though the size of disks increases and the amount of cold gas they harbor decreases with decreasing  $z$  so that the SFR in individual disks decreases, the global SFR density in disks keeps on increasing because of the increasing abundance of disks. As a result, the ratio between the SFR densities in disks and spheroids is kept roughly constant. However, the declining amount of cold gas in disks ultimately causes the SFR in both unstable and stable disks to reach a maximum at  $z \sim 2$ , and then to rapidly diminish toward  $z = 0$  (see Figure 12). Therefore, the origin of this maximum in the SFH is the same as for the plateau in the CGH: the lack of gas replenishment. However, in the SFH case, such a stagnation leads to a real maximum, not to a saddle point as in the CGH case. The drop in the SFR in disks and spheroids at  $z \sim 2$  is so marked that, even though at lower redshifts the cold gas mass density begins to increase again, the total SFR density will keep on decreasing until  $z = 0$ .

The stagnation of cold gas mass at  $z \lesssim 5$  has a similar effect on the SFH in models S and D. But the rest of the curve is slightly different in the two models. In model D, the SFH grows more steeply at high- $z$  despite the smaller value of  $\alpha_G$  (see Figure 7, panel (j)) because of the more rapid growth of ionized regions where normal galaxies lie. But, near  $z = 10$  the formation of Pop III stars declines and ionized regions stop growing, causing the SFH to become shallower. This yields a concave curvature in the SFH of model D that perfectly fits the data. Instead, the SFH in model S has a more constant slope. This difference in the SFH of both models combined with the



**Figure 13.** Same IEHs from normal galaxies (solid lines) and AGNs (dashed lines) as in panel (k) of Figure 7 but extended down to  $z = 1$ .

large abundance of spheroids with a very efficient SN heating also translates into a larger amount of hot gas in model D between  $z \sim 12$  to  $z \sim 5$  (see Figure 7, panel (x)). Unfortunately, there are no observational data on this quantity yet, so the behavior of the HGMH in the two models cannot be checked.

The behavior of the IEH from galaxies and AGNs follows from the evolution of those two kinds of sources described above. At very high- $z$ , Pop III stars have the dominant contribution. But this contribution diminishes as neutral pristine regions progressively disappear. In model D it completely vanishes at  $z \sim 10$  when Pop III stars stop forming, whereas in model S it continues until  $z \sim 6$ . On the other hand, the maximum contribution from AGNs is one order of magnitude smaller at  $z \sim 2$ . Consequently, the H I-ionizing emissivity at  $z \lesssim 6$  is always dominated by normal galaxies, and its evolution closely follows the SFH associated with ordinary stars. In particular, the IEH also reaches a maximum at  $z \sim 2$  (see Figure 13). What is less obvious is the maximum in both models S and D of the H I- and He II-ionizing (as well as the X-ray) emissivities from AGNs at that redshift (Figure 13): the MBH mass density is ever increasing at  $z < 2$  (see Figure 10) so we could naively expect their emissivities to do so also. What causes these maxima is the fact that the most massive MBHs stop being fueled by that redshift (their galaxies undergo dry mergers). Consequently, the bright AGNs stop radiating as quasars, and remain in the more quiescent radio phase. This has a major impact in the global AGN emissivities because bright AGNs have the largest bolometric luminosity (e.g., Hatziminaoglou et al. 2003). However, massive MBHs keep on growing through dry mergers and less massive ones through wet mergers of low-mass galaxies that still harbor gas.

## 6. SUMMARY AND DISCUSSION

Observational data on the high- $z$  universe are now numerous and accurate enough to severely constrain the interconnected evolution of galaxies and IGM. In fact, they would completely determine it provided the IGM metallicity and temperature were known at  $z \gtrsim 6$ . These pieces of information are needed,

indeed, to determine the threshold metallicity  $Z_c$  for atomic cooling and the Pop III star IMF that play a crucial role in the reionization process. Assuming that IMF is a power law, the conditions that the metallicity of ionized regions must be high enough for triggering sustained galaxy formation and Pop III stars massive enough for seeding MBHs greatly constrain those two properties.

The only acceptable solutions we find are in two narrow sets, one for small  $M_{\text{III}}^{\text{lo}}$  values (moderately top-heavy Pop III star IMFs) leading to single reionization, and another one for large  $M_{\text{III}}^{\text{lo}}$  values (top-heavier Pop III star IMFs) leading to double reionization. The latter solutions have a first ionization phase driven by Pop III stars completed at  $z \sim 10$ , and after a short recombination period a second ionization phase driven by normal galaxies completed at  $z \sim 6$ . In the former solutions with single reionization, both kinds of luminous sources contribute in parallel to the ionization of the IGM, completed at  $z \sim 6$ .

The best solution with double reionization gives excellent fits to all the observables. Instead, the best solution with single reionization predicts a too-monotonous SFH, and a slightly too steep CGH at  $z \sim 4$ . However, the CMB Thomson optical depth found in double reionization ( $\tau = 0.102^{+0.001}_{-0.001}$ ), though consistent with the observational estimates from the *WMAP* 9-year and *Planck* three-year data, is  $3\sigma$  larger than the most recent estimate ( $\tau = 0.058 \pm 0.014$ ) from the *Planck* data taking into account the large-scale polarization anisotropies. Instead, the CMB optical depth found in single reionization ( $\tau = 0.072^{+0.004}_{-0.005}$ ) is consistent with that estimate. There is thus some tension between the new estimate of  $\tau$  and the observed properties of the universe at high- $z$ . However, the empirical values of the CMB optical depth derived under the instantaneous reionization approximation must be taken with caution.

The fact that with only nine degrees of freedom it is possible to fit a dozen independent cosmic histories with non-trivial features (extreme and saddle points) confers strong reliability to our self-consistent model of the EoR.

As a byproduct, we have clarified the origin of some interesting features shown by the data on the high- $z$  universe: (1) the very small sizes of spheroids at  $z \sim 2$ –3; (2) the sudden drop of the SFH at  $z < 2$ ; (3) the similar behavior of the UV (and X-ray) emissivity from AGNs at the same redshift; (4) the plateau in the CGH between at  $1 < z < 5$ ; and (5) the rough constancy of the MBH-to-spheroid mass ratio. Our results also suggest that globular clusters are but the relics of the lowest mass galaxies with primordial origin or, more probably, formed during recombination in model D, which soon become frozen as satellite galaxies.

Finally, we have provided several predictions that should allow one to confirm the validity of the present model and to tell between single and double reionization. The most interesting one is that, in the case of double reionization, LAEs should be detectable near  $z = 10$ . This could explain the recent detection (to be confirmed) of a galaxy with Ly $\alpha$  emission at  $z = 8.68$  (Zitrin et al. 2015).

Funding for this work was provided by the Spanish MINECO under projects AYA2012-39168-C03-1, AYA2012-39168-C03-2, AYA2013-446724-P, MDM-2014-0369 of ICCUB (Unidad de Excelencia “María de Maeztu”), AYA2015-70498-C02-2-R (co-funded with FEDER funds),

and the Catalan DEC grant 2014SGR86. J.M.M.H. was supported by Spanish MINECO grant ESP2014-59789-P.

## REFERENCES

- Alavi, A., Siana, B., Richard, J., et al. 2016, *ApJ*, **832**, 56  
 Alvarez, M. A., Wise, J. H., & Abel, T. 2009, *ApJL*, **701**, L133  
 Becker, G. D., & Bolton, J. S. 2013, *MNRAS*, **436**, 1023  
 Becker, G. D., Bolton, J. S., Haehnelt, M. G., & Sargent, W. L. W. 2011, *MNRAS*, **410**, 1096  
 Becker, R. H., Fan, X., White, R. L., et al. 2001, *AJ*, **122**, 2850  
 Bolton, J. S., Becker, G. D., Raskutti, S., et al. 2012, *MNRAS*, **419**, 2880  
 Bolton, J. S., Becker, G. D., Wyithe, J. S. B., Haehnelt, M. G., & Sargent, W. L. W. 2010, *MNRAS*, **406**, 612  
 Bolton, J. S., & Haehnelt, M. G. 2007a, *MNRAS*, **381**, L35  
 Bolton, J. S., & Haehnelt, M. G. 2007b, *MNRAS*, **382**, 325  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2011, *ApJ*, **737**, 90  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2012a, *ApJL*, **752**, L5  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2012b, *ApJ*, **754**, 83  
 Bouwens, R. J., Smit, R., Labbe, I., et al. 2016, *ApJ*, **831**, 176  
 Bowman, J. D., & Rogers, A. E. E. 2010, *Natur*, **468**, 796  
 Bruce, V. A., Dunlop, J. S., McLure, R. J., et al. 2014a, *MNRAS*, **444**, 1001  
 Bruce, V. A., Dunlop, J. S., McLure, R. J., et al. 2014b, *MNRAS*, **444**, 1660  
 Bruzual, G., & Charlot, S. 2003, *MNRAS*, **344**, 1000  
 Buitrago, F., Trujillo, I., Conselice, C. J., et al. 2008, *ApJL*, **687**, L61  
 Caputi, K. I., Cirasuolo, M., Dunlop, J. S., et al. 2011, *MNRAS*, **413**, 162  
 Caputi, K. I., Ilbert, O., Laigle, C., et al. 2015, *ApJ*, **810**, 73  
 Castelló-Mor, N., Carrera, F. J., Alonso-Herrero, A., et al. 2013, *A&A*, **556**, A114  
 Cen, R. 2003, *ApJ*, **591**, 12  
 Chabrier, G. 2003, *PASP*, **115**, 763  
 Chen, X. 2007, *NuPhS*, **173**, 133  
 Chiu, W. A., Fan, X., & Ostriker, J. P. 2003, *ApJ*, **599**, 759  
 Coe, D., Zitrin, A., Carrasco, M., et al. 2013, *ApJ*, **762**, 32  
 Collin, S., Kawaguchi, T., Peterson, B. M., & Vestergaard, M. 2006, *A&A*, **456**, 75  
 Cowie, L. L., Barger, A. J., & Trouille, L. 2009, *ApJ*, **692**, 1476  
 Crowther, P. A., Schnurr, O., Hirschi, R., et al. 2010, *MNRAS*, **408**, 731  
 Cullen, F., Cirasuolo, M., McLure, R. J., Dunlop, J. S., & Bowler, R. A. A. 2014, *MNRAS*, **440**, 2300  
 D’Oro, V., Cupani, G., Cristiani, S., et al. 2013, *MNRAS*, **435**, 1198  
 Dekel, A., Birnboim, Y., Engel, G., et al. 2009, *Natur*, **457**, 451  
 Dekel, A., & Silk, J. 1986, *ApJ*, **303**, 39  
 Di Matteo, T., Springel, V., & Hernquist, L. 2005, *Natur*, **433**, 604  
 Dixon, K. L., & Furlanetto, S. R. 2009, *ApJ*, **706**, 970  
 Djorgovski, S. G., Castro, S., Stern, D., & Mahabal, A. A. 2001, *ApJ*, **560L**, 5  
 Ellis, R. S., McLure, R. J., Dunlop, J. S., et al. 2013, *ApJL*, **763**, L7  
 Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, *AJ*, **132**, 117  
 Figer, D. F. 2005, *Natur*, **434**, 192  
 Finlator, K., Oh, S. P., Özel, F., & Davé, R. 2012, *MNRAS*, **427**, 2464  
 Furlanetto, S. R., & Loeb, A. 2005, *ApJ*, **634**, 1  
 Furlanetto, S. R., Zaldarriaga, M., & Hernquist, L. 2006, *MNRAS*, **365**, 1012  
 Gnedin, N. Y. 2004, *ApJ*, **610**, 9  
 González, V., Labbé, I., Bouwens, R. J., et al. 2010, *ApJ*, **713**, 115  
 González, V., Labbé, I., Bouwens, R. J., et al. 2011, *ApJL*, **735**, L34  
 Gunn, J. E., & Peterson, B. A. 1965, *ApJ*, **142**, 1633  
 Haiman, Z., & Cen, R. 2005, *ApJ*, **623**, 627  
 Haiman, Z., & Holder, G. P. 2003, *ApJ*, **595**, 1  
 Halliday, C., Daddi, E., Cimatti, A., et al. 2008, *A&A*, **479**, 41  
 Hatziminaoglou, E., Mathez, G., Solanes, J.-M., Manrique, A., & Salvador-Solé, E. 2003, *MNRAS*, **343**, 692  
 Hayes, M., Schaerer, D., Östlin, G., et al. 2011, *ApJ*, **730**, 8  
 Heger, A., & Woosley, S. E. 2002, *ApJ*, **567**, 532  
 Hernquist, L. 1990, *ApJ*, **356**, 359  
 Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, *ApJS*, **208**, 19  
 Holder, G. P., Haiman, Z., Kaplinghat, M., & Knox, L. 2003, *ApJ*, **595**, 13  
 Hui, L., & Haiman, Z. 2003, *ApJ*, **596**, 9  
 Ichikawa, K., Barkana, R., Iliev, I. T., Mellema, G., & Shapiro, P. R. 2010, *MNRAS*, **406**, 2521  
 Iliev, I. T., Mellema, G., Pen, U.-L., Bond, J. R., & Shapiro, P. R. 2008, *MNRAS*, **384**, 863  
 Ishigaki, M., Kawamata, R., Ouchi, M., et al. 2015, *ApJ*, **799**, 12  
 Jaacks, J., Choi, J.-H., Nagamine, K., Thompson, R., & Varghese, S. 2012, *MNRAS*, **420**, 1606  
 Jaacks, J., Thompson, R., & Nagamine, K. 2013, *ApJ*, **766**, 94  
 Jensen, H., Laursen, P., Mellema, G., et al. 2013, *MNRAS*, **428**, 1366

- Juan, E., Salvador-Solé, E., Domènech, G., & Manrique, A. 2014a, *MNRAS*, **439**, 719
- Juan, E., Salvador-Solé, E., Domènech, G., & Manrique, A. 2014b, *MNRAS*, **439**, 3156
- Kelly, B. C., Vestergaard, M., Fan, X., et al. 2010, *ApJ*, **719**, 1315
- Kogut, A., Spergel, D. N., Barnes, C., et al. 2003, *ApJS*, **148**, 161
- Kormendy, J., & Ho, L. C. 2013, *ARA&A*, **51**, 511
- Kuhlen, M., Madau, P., & Krumholz, M. R. 2013, *ApJ*, **776**, 34
- Labbé, I., González, V., Bouwens, R. J., et al. 2010a, *ApJL*, **708**, L26
- Labbé, I., González, V., Bouwens, R. J., et al. 2010b, *ApJL*, **716**, L103
- Larson, R. B. 1998, *MNRAS*, **301**, 569
- Lidz, A., Faucher-Giguère, C.-A., Dall’Aglio, A., et al. 2010, *ApJ*, **718**, 199
- Lidz, A., McQuinn, M., Zaldarriaga, M., Hernquist, L., & Dutta, S. 2007, *ApJ*, **670**, 39
- Livermore, R. C., Finkelstein, S. L., & Lotz, J. M. 2016, arXiv:1604.06799
- Lorenzoni, S., Bunker, A. J., Wilkins, S. M., et al. 2011, *MNRAS*, **414**, 1455
- Mac Low, M.-M., & Ferrara, A. 1999, *ApJ*, **513**, 142
- Maiolino, R., Nagao, T., Grazian, A., et al. 2008, *A&A*, **488**, 463
- Malhotra, S., & Rhoads, J. E. 2004, *ApJL*, **617**, L5
- Manrique, A., & Salvador-Solé, E. 2015, *ApJ*, **803**, 103
- Manrique, A., Salvador-Solé, E., Juan, E., et al. 2015, *ApJS*, **216**, 13
- Maselli, A., Gallerani, S., Ferrara, A., & Choudhury, T. R. 2007, *MNRAS*, **376**, L34
- Mather, J. C., Cheng, E. S., Eplee, R. E., Jr., et al. 1990, *ApJL*, **354**, L37
- McGreer, I. D., Mesinger, A., & Fan, X. 2011, *MNRAS*, **415**, 3237
- McLeod, D. J., McLure, R. J., Dunlop, J. S., et al. 2015, *MNRAS*, **450**, 3032
- McLure, R. J., Dunlop, J. S., Cirasuolo, M., et al. 2010, *MNRAS*, **403**, 960
- McQuinn, M. 2009, *ApJL*, **704**, L89
- McQuinn, M. 2012, *MNRAS*, **426**, 1349
- McQuinn, M., Hernquist, L., Zaldarriaga, M., & Dutta, S. 2007, *MNRAS*, **381**, 75
- Meiksin, A. A. 2009, *RvMP*, **81**, 1405
- Mesinger, A., & Furlanetto, S. R. 2008, *MNRAS*, **386**, 1990
- Mesinger, A., & Haiman, Z. 2004, *ApJL*, **611**, L69
- Mesinger, A., & Haiman, Z. 2007, *ApJ*, **660**, 923
- Milosavljević, M., Bromm, V., Couch, S. M., & Oh, S. P. 2009, *ApJ*, **698**, 766
- Mo, H. J., Mao, S., & White, S. D. M. 1998, *MNRAS*, **295**, 319
- Morales, M. F., & Wyithe, J. S. B. 2010, *ARA&A*, **48**, 127
- Mortlock, A., Conselice, C. J., Bluck, A. F. L., et al. 2011a, *MNRAS*, **413**, 2845
- Mortlock, D. J., Warren, S. J., Venemans, B. P., et al. 2011b, *Natur*, **474**, 616
- Naselsky, P., & Chiang, L. 2004, *MNRAS*, **347**, 795
- Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2013, *ApJ*, **773**, 75
- Oh, S. P., & Haiman, Z. 2003, *MNRAS*, **346**, 456
- Ota, K., Iye, M., Kashikawa, N., et al. 2008, *ApJ*, **677**, 12
- Ouchi, M., Shimasaku, K., Furusawa, H., et al. 2010, *ApJ*, **723**, 869
- Péroux, C., McMahon, R. G., Storrie-Lombardi, L. J., & Irwin, M. J. 2003, *MNRAS*, **346**, 1103
- Phillips, A. C., Guzmán, R., Gallego, J., et al. 1997, *ApJ*, **489**, 543
- Planck Collaboration, Adam, R., Aghanim, N., et al. 2016a, *A&A*, **596**, A108
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016b, *A&A*, **594**, A13
- Prochaska, J. X., Herbert-Fort, S., & Wolfe, A. M. 2005, *ApJ*, **635**, 123
- Raig, A., González-Casado, G., & Salvador-Solé, E. 2001, *MNRAS*, **27**, 939
- Reddy, N. A., & Steidel, C. C. 2009, *ApJ*, **692**, 778
- Richard, J., Pelló, R., Schaerer, D., Le Borgne, J.-F., & Kneib, J.-P. 2006, *A&A*, **456**, 861
- Ricotti, M., Gnedin, N. Y., & Shull, J. M. 2000, *ApJ*, **534**, 41
- Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, *ApJL*, **802**, L19
- Ryan-Weber, E. V., Pettini, M., Madau, P., & Zych, B. J. 2009, *MNRAS*, **395**, 1476
- Salvador-Solé, E., Manrique, A., González-Casado, G., & Hansen, S. H. 2007, *ApJ*, **666**, 181
- Salvador-Solé, E., Serra, S., Manrique, A., & González-Casado, G. 2012a, *MNRAS*, **424**, 3129
- Salvador-Solé, E., Solanes, J. M., & Manrique, A. 1998, *ApJ*, **499**, 542
- Salvador-Solé, E., Viñas, J., Manrique, A., & Serra, S. 2012b, *MNRAS*, **423**, 2190
- Santoro, F., & Shull, J. M. 2006, *ApJ*, **643**, 26
- Schaerer, D. 2002, *A&A*, **382**, 28
- Schaye, J., Theuns, T., Rauch, M., Efstathiou, G., & Sargent, W. L. W. 2000, *MNRAS*, **318**, 817
- Schneider, R., & Omukai, K. 2010, *MNRAS*, **402**, 429
- Sérsic, J.-L. 1968, Atlas de Galaxias Australes (Córdoba: Obs. Astron.)
- Shen, B. S., Mo, H. J., White, S. D. M., et al. 2007, *MNRAS*, **379**, 400
- Shen, S., Mo, H. J., White, S. D. M., et al. 2003, *MNRAS*, **343**, 978
- Shin, M.-S., Trac, H., & Cen, R. 2008, *ApJ*, **681**, 756
- Silk, J. 1997, *ApJ*, **481**, 703
- Simcoe, R. A. 2006, *ApJ*, **653**, 977
- Simcoe, R. A., Cooksey, K. L., Matejek, M., et al. 2011, *ApJ*, **743**, 21
- Smith, B. D., Turk, M. J., Sigurdsson, S., O’Shea, B. W., & Norman, M. L. 2009, *ApJ*, **691**, 441
- Sokasian, A., Yoshida, N., Abel, T., Hernquist, L., & Springel, V. 2004, *MNRAS*, **350**, 47
- Solanes, J. M., Manrique, A., González-Casado, G., & Salvador-Solé, E. 2005, *ApJ*, **628**, 45
- Sommariva, V., Mannucci, F., Cresci, G., et al. 2012, *A&A*, **539**, A136
- Songaila, A. 2001, *ApJ*, **561**, 153
- Stark, D. P., Ellis, R. S., Bunker, A., et al. 2009, *ApJ*, **697**, 1493
- Stark, D. P., Ellis, R. S., Richard, J., et al. 2007, *ApJ*, **663**, 10
- Strickland, D. K., & Stevens, I. R. 2000, *MNRAS*, **314**, 511
- Totani, T., Kawai, N., Kosugi, G., et al. 2006, *PASJ*, **58**, 485
- Treister, E., Schawinski, K., Volonteri, M., Natarajan, P., & Gawiser, E. 2011, *Natur*, **474**, 356
- Ueda, Y., Akiyama, M., Ohta, K., & Miyaji, T. 2003, *MNRAS*, **598**, 886
- Vasudevan, R. V., & Fabian, A. C. 2007, *MNRAS*, **381**, 1235
- Vestergaard, M., Fan, X., Tremonti, C. A., Osmer, P. S., & Richards, G. T. 2008, *ApJL*, **674**, L1
- Vincenzo, F., Matteucci, F., Belfiore, F., & Maiolino, R. 2016, *MNRAS*, **455**, 418
- Wilkins, S. M., Trentham, N., & Hopkins, A. M. 2008, *MNRAS*, **385**, 687
- Willott, C. J., Albert, L., Arzoumanian, D., et al. 2010a, *AJ*, **140**, 546
- Willott, C. J., Delorme, P., Reylé, C., et al. 2010b, *AJ*, **139**, 906
- Worsley, M. A., Fabian, A. C., Bauer, F. E., et al. 2005, *MNRAS*, **357**, 1281
- Wyithe, J. S. B., & Cen, R. 2007, *ApJ*, **659**, 890
- Wyithe, J. S. B., Hopkins, A. M., Kistler, M. D., Yüksel, H., & Beacom, J. F. 2010, *MNRAS*, **401**, 2561
- Wyithe, J. S. B., & Loeb, A. 2004, *Natur*, **427**, 815
- Wyithe, J. S. B., Loeb, A., & Carilli, C. 2005, *ApJ*, **628**, 575
- Yong, S. Y., Webster, R. L., & King, A. L. 2016, *PASA*, **33**, e009
- Zahn, O., Reichardt, C. L., Shaw, L., et al. 2012, *ApJ*, **756**, 65
- Zitrin, A., Labbé, I., Belli, S., et al. 2015, *ApJL*, **810**, L12