The origin and evolution of the galaxy mass–metallicity relation

Xiangcheng Ma,[1][*] Philip F. Hopkins,[1] Claude-André Faucher-Giguère,[2] Nick Zolman,[1] Alexander L. Muratov,[3] Dušan Kereš[3] and Eliot Quataert[4]

[1]*TAPIR, MC 350-17, California Institute of Technology, Pasadena, CA 91125, USA*
[2]*Department of Physics and Astronomy and CIERA, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA*
[3]*Department of Physics, Center for Astrophysics and Space Sciences, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA*
[4]*Department of Astronomy and Theoretical Astrophysics Center, University of California Berkeley, Berkeley, CA 94720, USA*

ABSTRACT

We use high-resolution cosmological zoom-in simulations from the Feedback in Realistic Environment (FIRE) project to study the galaxy mass–metallicity relations (MZR) from $z = 0$–6. These simulations include explicit models of the multiphase ISM, star formation, and stellar feedback. The simulations cover halo masses $M_{\rm halo} = 10^9$–$10^{13}\,{\rm M_\odot}$ and stellar masses $M_* = 10^4$–$10^{11}\,{\rm M_\odot}$ at $z = 0$ and have been shown to produce many observed galaxy properties from $z = 0$–6. For the first time, our simulations agree reasonably well with the observed mass–metallicity relations at $z = 0$–3 for a broad range of galaxy masses. We predict the evolution of the MZR from $z = 0$–6, as $\log(Z_{\rm gas}/{\rm Z_\odot}) = 12 + \log({\rm O/H}) - 9.0 = 0.35[\log(M_*/{\rm M_\odot}) - 10] + 0.93 \exp(-0.43z) - 1.05$ and $\log(Z_*/{\rm Z_\odot}) = [{\rm Fe/H}] + 0.2 = 0.40[\log(M_*/{\rm M_\odot}) - 10] + 0.67 \exp(-0.50z) - 1.04$, for gas-phase and stellar metallicity, respectively. Our simulations suggest that the evolution of MZR is associated with the evolution of stellar/gas mass fractions at different redshifts, indicating the existence of a universal metallicity relation between stellar mass, gas mass, and metallicities. In our simulations, galaxies above $M_* = 10^6\,{\rm M_\odot}$ are able to retain a large fraction of their metals inside the halo, because metal-rich winds fail to escape completely and are recycled into the galaxy. This resolves a longstanding discrepancy between 'subgrid' wind models (and semi-analytic models) and observations, where common subgrid models cannot simultaneously reproduce the MZR and the stellar mass functions.

Key words: galaxies: evolution – galaxies: formation – cosmology: theory.

1 INTRODUCTION

The galaxy mass–metallicity relation (MZR) is one of the most fundamental properties observed in galaxies. In the local Universe, there is a tight correlation between galaxy stellar mass and gas-phase oxygen abundance for star-forming galaxies (e.g. Tremonti et al. 2004), with an intrinsic scatter of only 0.1 dex in log(O/H). This relation has been extended to local dwarf galaxies and found to be a uniform, tight correlation over five orders of magnitude in stellar mass, from $M_* = 10^6$–$10^{11}\,{\rm M_\odot}$ (Lee et al. 2006). Many different groups have confirmed the MZR to exist not only in the local Universe but also at high redshifts up to $z \sim 2.3$ (e.g. Savaglio et al. 2005; Erb et al. 2006; Zahid, Kewley & Bresolin 2011; Zahid et al. 2012; Andrews & Martini 2013; Henry et al. 2013a,b; Steidel et al. 2014; Yabe et al. 2014; Sanders et al. 2015). Zahid et al. (2013) compiled a number of the observed MZR from

$z = 0$–2.3 and found that the MZR evolves with redshift, with higher metallicity at low redshift for a given stellar mass. The MZR is also found at $z \gtrsim 3$ (e.g. Maiolino et al. 2008; Mannucci et al. 2009), despite the fact that the results are obtained from very small samples.

Gas-phase metallicities represent the 'current' state of chemical enrichment in the galaxies, while stellar metallicities reflect the 'time-averaged' galactic metallicity across the whole star formation history. Similarly, an MZR is also found in stellar metallicities. Gallazzi et al. (2005) derived the stellar metallicities for ∼44 000 galaxies from the Sloan Digital Sky Survey (SDSS) and found a tight correlation between stellar mass and stellar metallicity for galaxies of stellar masses 10^9–$10^{12}\,{\rm M_\odot}$. Kirby et al. (2013) measured the metallicities of individual stars in a sample of dwarf galaxies within the Local Group and found the SDSS stellar MZR can be continually extended down to $10^3\,{\rm M_\odot}$. Despite the fact that stellar metallicity is challenging to measure at high redshifts, the stellar MZR provides very important and complimentary insights on the chemical evolution of galaxies, especially for massive quiescent

*E-mail: xchma@caltech.edu

galaxies and satellite galaxies in the Local Group where the gas-phase metallicities are hard to measure due to their low gas content.

Simple analytic models of galactic chemical evolution, such as the 'closed-box', 'leaky-box' and 'accreting-box' models (e.g. Schmidt 1963; Talbot & Arnett 1971; Searle & Sargent 1972; Edmunds 1990), are often quoted to illustrate the qualitative behaviour of the MZR. More complicated models have also been developed to work in cosmological contexts and to connect gas inflows, outflows, and star formation to galactic chemical evolution (e.g. Dalcanton 2007; Finlator & Davé 2008; Davé, Finlator & Oppenheimer 2012; Lilly et al. 2013; Lu, Blanc & Benson 2015a). These models indicate that the existence of MZR is the consequence of an interplay between star formation efficiency, metal loss from gas outflows, and gas recycling and accretion. For example, the stellar mass–halo mass relation (e.g. Behroozi, Wechsler & Conroy 2013; Moster, Naab & White 2013) indicates that the star formation efficiency (fraction of baryons turned into stars) is lower in low-mass galaxies than in more massive galaxies, suggesting that low-mass galaxies should be less metal-enriched. Meanwhile, galactic winds are ubiquitous (see e.g. Veilleux, Cecil & Bland-Hawthorn 2005, for a recent review), carrying metals away from galaxies. Low-mass galaxies have shallow potential wells, so they tend to lose a significant fraction of their gas and metals, while massive galaxies have potential wells deep enough to prevent material from escaping the galaxy (e.g. Dekel & Silk 1986). On the other hand, gas inflows bring the metal-poor gas in the galactic halo and/or in the intergalactic medium (IGM) inwards, diluting the metal content in the interstellar medium (ISM) and supplying new material for star formation (e.g. Kereš et al. 2005; Faucher-Giguère, Kereš & Ma 2011). During this process, a considerable fraction of the gas and metals that have been formerly ejected via outflows eventually come back to the galaxy (e.g. Bertone, De Lucia & Thomas 2007; Oppenheimer et al. 2010). Galaxy mergers and AGN activity could also be important, in the sense that they can trigger violent starburst, drive intensive gas outflows, and ultimately quench the star formation in the galaxy (e.g. Springel, Di Matteo & Hernquist 2005; Hopkins et al. 2013a).

Analytical models usually rely on simplified assumptions such as perfect mixing and adopt simple analytic prescriptions describing star formation, gas accretion, and outflows. In reality, these physical processes are tightly connected to each other and therefore must be treated self-consistently to understand the complete picture of galactic chemical evolution. Semi-analytic models (SAMs) of galaxy formation follow cosmological halo growth and halo mergers and include physically and/or empirically motivated prescriptions of heating and cooling, star formation, metal enrichment, gas accretion and outflows, recycling, and AGN feedback (e.g. Croton et al. 2006; Somerville et al. 2008; Lu et al. 2011, 2014; Benson 2012; Guo et al. 2013; Henriques et al. 2013, 2015; Yates et al. 2013; Lu, Mo & Lu 2015b). They are much less computationally expensive to run than hydrodynamic simulations and are able to reproduce a number of galaxy properties for a broad range of stellar mass. However, one major challenge for SAMs is *simultaneously* reproducing observed stellar masses, star formation rates (SFRs), and metallicities. The metallicities of low-mass galaxies are particularly sensitive to the galactic wind model because strong outflows are required to suppress star formation in low-mass systems (see e.g. Lu et al. 2014, for a detailed comparison and discussion). Moreover, even though different SAMs have been successfully tuned to match the $z = 0$ stellar mass function (SMF), many of them fail to match the observed the SMFs at high redshifts (e.g. Somerville & Davé 2014). At the same time, these models typically fail to match high-redshift MZR measurements and also diverge from one another in

their MZR predictions. Nonetheless, it is encouraging that recently improved SAMs are able to reconcile stellar masses, colours, and SFRs of galaxies from $z = 0$–3 (e.g. Henriques et al. 2013, 2015).

Large-volume cosmological hydrodynamic simulations produce large samples of galaxies and are powerful tools for statistical studies of galaxy properties (e.g. Bertone et al. 2007; Davé, Finlator & Oppenheimer 2011b; Torrey et al. 2014; Schaye et al. 2015). These simulations however usually have relatively poor mass and spatial resolution. They cannot explicitly resolve the multiphase structure of the ISM, when and where star formation takes place, how galactic winds are launched by stellar feedback, and how the winds interact with the circumgalactic medium (CGM). Approximate, empirical 'sub-grid' models of the ISM structure, star formation, and stellar feedback are required and used. For example, Davé, Oppenheimer & Finlator (2011a) implemented a momentum-driven wind model, with wind mass loading factors and velocities prescribed as a function of bulk galaxy properties. In their implementation, hydrodynamic interactions are temporarily suppressed as gas from the ISM is 'kicked' into the galactic wind. Simulations using such simple prescriptions reveal similar problems to the SAMs. Torrey et al. (2014) found a steeper slope than observed at the low-mass end of the MZR. These authors attributed it to the low metal retention efficiency in the presence of strong outflows, which were required in their model in order to prevent low-mass galaxies from forming too many stars. They further emphasized the tension between suppressing star formation and retaining enough metals in low-mass galaxies. Furthermore, the star formation histories in these simulations are very different and not all consistent with observations at high redshifts. Many cosmological simulations tend to form *too many* stars at early times (e.g. Davé et al. 2011a; Fiacconi, Feldmann & Mayer 2015; Sparre et al. 2015; for a review, see Somerville & Davé 2014). Such problems are also common in SAMs. They are likely the result of imperfect star formation and stellar feedback models implemented in those simulations (cf. Scannapieco et al. 2012). Consequently, these simulations predict very different evolution of the MZR.

Therefore, when using cosmological hydrodynamic simulations to understand the MZR and its evolution, one is required to capture the 'correct' behaviour of star formation, stellar feedback, gas outflows, and the mixing and interaction of galactic winds with the CGM on all relevant scales. Encouragingly, Obreja et al. (2014) presented a suite of cosmological zoom-in simulations from the MaGICC project using an improved star formation and supernovae (SNe) feedback model. Their model includes an empirical prescription to approximate the effects of stellar feedback mechanisms operating before the first SNe explode. These authors showed that their simulations match the stellar mass–halo mass relation and the observed MZR from $z = 0$–3, for the eight galaxies in their sample. In this work, the first of a series, we will study the chemical evolution of galaxies using the FIRE (Feedback in Realistic Environment) simulations (Hopkins et al. 2014). The FIRE project[1] is a series of cosmological zoom-in simulations that are able to follow galaxy merger history, interactions of galaxies with the IGM, and many other important processes. These simulations include a full set of realistic physical models and explicitly resolve the multiphase structure of the ISM, star formation, and stellar feedback, with *no* need to tune parameters. The FIRE simulations successfully reproduce many observed galaxy properties, including the stellar mass–halo mass relation, star formation histories, the

---

[1] FIRE project, website: http://fire.northwestern.edu

Kennicutt–Schmidt law, and the star-forming main sequence, from $z = 0$–6, for a broad range of galaxy masses in $M_* = 10^4$–10^{11} M$_\odot$ (Hopkins et al. 2014). Also, the FIRE simulations predict reasonable covering fractions of neutral hydrogen in the haloes of $z = 2$–3 Lyman Break Galaxies (LBGs; Faucher-Giguère et al. 2015) and self-consistently generate galactic winds with velocities and mass loading factors broadly consistent with observational requirements (Muratov et al. 2015). These results further justify the reliability to study galactic chemical evolution using the FIRE simulations.

This paper focuses on the galaxy mass–metallicity relation. In companion papers, we will also study the stellar metallicity distribution functions and [α/Fe] abundance ratio variation in dwarf galaxies, metallicity gradients and their origins, metal outflows and recycling. We start by describing the simulations in Section 2 and present the mass–metallicity relation at different redshifts in Section 3. In Section 4, we discuss the key processes that drive the shape and evolution of the MZR. We summarize and conclude in Section 5.

2 THE SIMULATIONS

2.1 Simulation details

This work is part of the FIRE project. A full description of the numerical methods and physics included in our simulations is presented in (Hopkins et al. 2014, and references therein). We summarized their main features here. All the simulations use the newly developed GIZMO code (Hopkins 2014) in 'P-SPH' mode. P-SPH adopts a Lagrangian pressure-entropy formulation of the smoothed particle hydrodynamics (SPH) equations (Hopkins 2013), which eliminates the major differences between SPH, moving-mesh, and grid codes, and resolves many well-known issues in traditional density-based SPH formulations. The gravity solver is a heavily modified version of the GADGET-3 code (Springel 2005), and P-SPH also includes substantial improvements in the artificial viscosity, entropy diffusion, adaptive time-stepping, smoothing kernel, and gravitational softening algorithm.

We use the multiscale 'zoom-in' initial conditions generated with the MUSIC code (Hahn & Abel 2011), using second-order Lagrangian perturbation theory. The first set of simulations have been run down to $z = 0$ and cover halo masses 10^9–10^{13} M$_\odot$ and stellar masses 10^4–10^{11} M$_\odot$ at $z = 0$ (mxx series). Most of them have been presented in Hopkins et al. (2014). The simulations m09 and m10 are isolated dwarfs, constructed using the method from Oñorbe et al. (2014). Simulations m10v, m11, m11v, m12q, m12i, and m13 are chosen to match the initial conditions from the AGORA project (Kim et al. 2014), which will enable future comparisons with a wide range of simulation codes and physics implementations. Simulation m12v is based on the initial conditions studied in Kereš & Hernquist (2009) and Faucher-Giguère & Kereš (2011). The simulations with a label 'v' have relatively violent merger histories at $z < 2$, while the rest have more typical merger histories. The resolution of these simulations is chosen to scale with the mass of the system to ensure we are able to resolve the giant molecular clouds (GMCs). We also include a separate set of simulations run to $z = 2$ (z2hxxx series), which are presented in (Faucher-Giguère et al. 2015). Their main haloes are chosen to host LBG and cover halo masses 1.9×10^{11}–1.2×10^{12} M$_\odot$ at $z = 2$. Finally, we include another series of simulations only run to $z \sim 6$, but with extremely high resolutions (z5mxx series). These simulations are presented in Ma et al. (2015). Their main haloes cover halo masses from 7.7×10^8–5.6×10^{10} M$_\odot$ at $z = 6$, and these galaxies are believed to contribute most to the

cosmic reionization (e.g. Kuhlen & Faucher-Giguère 2012; Robertson et al. 2013). The initial conditions of all the simulations are summarized in Table 1.

In our simulations, gas follows an ionized+atomic+molecular cooling curve from 10–10^{10} K, including metallicity-dependent fine-structure and molecular cooling at low temperatures and high-temperature metal-line cooling followed species-by-species for 11 separately tracked species (H, He, C, N, O, Ne, Mg, Si, S, Ca, and Fe; see Wiersma et al. 2009a). At each timestep, the ionization states and cooling rates are determined from a compilation of CLOUDY runs, including a uniform but redshift-dependent photoionizing background tabulated in Faucher-Giguère et al. (2009), and photoionizing and photoelectric heating from local sources. Gas self-shielding is accounted for with a local Jeans-length approximation, which is consistent with the radiative transfer calculations in Faucher-Giguère et al. (2010).

Star formation is allowed only in dense, molecular, and self-gravitating regions with hydrogen number density above some threshold $n_{\rm th} = 10$–100 cm^{-3} (Hopkins, Narayanan & Murray 2013b). Stars form at 100 per cent efficiency per free-fall time when the gas meets these criteria. The self-gravity criterion is physically required to obtain the correct spatial star formation distribution in galaxies (Hopkins et al. 2013b; Padoan & Nordlund 2011), but the galaxy-averaged star formation efficiency is regulated by feedback at much lower values ($\sim$1 per cent per dynamical time; e.g. Faucher-Giguère, Quataert & Hopkins 2013). We stress that changing these parameters in a reasonable range only yields small and random variations to the global star formation history, as long as feedback is active (see Hopkins, Quataert & Murray 2011, 2012).

Once a star forms, it inherits the metallicity of each tracked species from its parent gas particle. Every star particle is treated as a single stellar population with known mass, age, and metallicity. Then all the feedback quantities, including ionizing photon budgets, luminosities, stellar spectra, SNe rates, mechanical luminosities of stellar winds, metal yields, etc., are directly tabulated from the stellar population models in STARBURST99 (Leitherer et al. 1999), assuming a Kroupa (2002) initial mass function (IMF) from 0.1–100 M$_\odot$.[2] We account for several different stellar feedback mechanisms, including: (1) local and long-range momentum flux from radiative pressure; (2) energy, momentum, mass and metal injection from SNe and stellar winds; and (3) photoionization and photoelectric heating. We follow Wiersma et al. (2009b) and include the metal yields from Type II SNe, Type I SNe, and stellar winds. We note that the Type II SNe yield table from Woosley & Weaver (1995) adopted in our simulations produce Mg roughly $\sim$0.4 dex below the typical values in modern models (e.g. Nomoto et al. 2006). This will have little effect on the galaxy properties in our simulations, as Mg is not an important coolant. Nevertheless, we will add 0.4 dex to the Mg abundance to correct this in the analysis below.

All simulations adopt a standard flat ΛCDM cosmology with cosmological parameters consistent with $H_0 = 70.2$ km s^{-1} Mpc^{-1}, $\Omega_\Lambda = 0.728$, $\Omega_{\rm m} = 1 - \Omega_\Lambda = 0.272$, $\Omega_{\rm b} = 0.0455$, $\sigma_8 = 0.807$ and $n = 0.961$ (e.g. Hinshaw et al. 2013; Planck Collaboration 2013).

---

[2] In principle, the 'IMF-averaged' approximation does not hold for the ultrahigh resolution simulations in the z5mxx series, where the mass of a star particle is only 10–100 M$_\odot$. Nevertheless, we confirmed that these simulations predict similar global galaxy properties to those of much poorer resolutions (see Ma et al. 2015).

Table 1. Simulation initial conditions.

Name	$M_{\rm halo}$ (M$_\odot$)	$m_{\rm b}$ (M$_\odot$)	$\epsilon_{\rm b}$ (pc)	$m_{\rm dm}$ (M$_\odot$)	$\epsilon_{\rm dm}$ (pc)	Merger history	Notes
m09	2.5e9	2.6e2	1.4	1.3e3	30	normal	isolated dwarf
m10	0.8e10	2.6e2	3.0	1.3e3	30	normal	isolated dwarf
m10lr	0.8e10	2.1e3	2.1	1.0e4	35	normal	low resolution
m10v	0.8e10	2.1e3	7.0	1.0e4	70	violent	–
m11	1.4e11	7.0e3	7.0	3.5e4	70	quiescent	–
m11v	3.3e11	5.6e4	7.0	3.0e5	140	violent	–
m12v	6.3e11	3.9e4	10	2.0e5	140	violent	several $z < 2$ mergers
m12q	1.2e12	7.1e3	10	2.8e5	140	late merger	–
m12i	1.1e12	5.0e4	14	2.8e5	140	normal	large ($\sim 10\,R_{\rm vir}$) box
m13	6.0e12	3.6e5	21	2.2e6	210	normal	small group mass
z2h350	7.9e11	5.9e4	9	2.9e5	143	normal	–
z2h400	7.9e11	5.9e4	9	2.9e5	143	quiescent	–
z2h450	8.7e11	5.9e4	9	2.9e5	143	normal	–
z2h506	1.2e12	5.9e4	9	2.9e5	143	violent	–
z2h550	1.9e11	5.9e4	9	2.9e5	143	quiescent	–
z2h600	6.7e11	5.9e4	9	2.9e5	143	violent	–
z2h650	4.0e11	5.9e4	9	2.9e5	143	normal	–
z2h830	5.4e11	5.9e4	9	2.9e5	143	normal	–
z5m09	7.6e8	16.8	0.14	81.9	5.6	quiescent	ultrahigh resolution
z5m10	1.3e10	131.6	0.4	655.6	7	normal	ultrahigh resolution
z5m10mr	1.4e10	1.1e3	1.9	5.2e3	14	normal	–
z5m11	5.6e10	2.1e3	4.2	1.0e4	14	normal	–

Notes. Parameters describing the initial conditions for our simulations (units are physical):

(1) Name: simulation designation.

(2) $M_{\rm halo}$: approximate mass of the main halo at $z = 0$ (mxx series), $z = 2$ (z2hxxx series), or $z = 6$ (z5mxx series).

(3) $m_{\rm b}$: Initial baryonic (gas and star) particle mass in the high-resolution region.

(4) $\epsilon_{\rm b}$: Minimum baryonic force softening (minimum SPH smoothing lengths are comparable or smaller. Force softening is adaptive (mass resolution is fixed).

(5) $m_{\rm dm}$: dark matter particle mass in the high-resolution region.

(6) $\epsilon_{\rm dm}$: minimum dark matter force softening (fixed in physical units at all redshifts).

2.2 Halo identification, stellar mass and metallicity

We use the Amiga Halo Finder (AHF; Gill, Knebe & Gibson 2004; Knollmann & Knebe 2009) to identify galactic haloes and galaxies in our simulations. The AHF code uses the adaptive mesh refinement method and identifies haloes and subhalos as groups of *bound* particles (dark matter, gas, and stars). In this work, we only consider those 'well-resolved' haloes that include more than 10^5 bound particles, have at most 10 per cent of their mass contaminated by low-resolution particles, and contain at least 100 gas and 100 star particles, respectively. These criteria are somewhat arbitrary, but varying these numbers within a reasonable range will have little effect on our conclusions. If none of the haloes meets these criteria in a snapshot (this happens in some snapshots at high redshifts ($z \sim$ 6), where the galaxy progenitors are too small to contain so many particles), we will take the most massive halo in the high-resolution region in our analysis. We do not include subhalos/satellite galaxies in this work. The centre of a halo is located at the centre of mass of the finest refinement level. We adopt the virial overdensity from Bryan & Norman (1998), which evolves with cosmic time.

We only consider the main galaxy in each halo. To remove the contamination of satellite galaxies, we exclude any gas/star particle that is bound to a subhalo in the analysis below. We measure the galaxy stellar mass (M_*) by summing over the mass of all star particles that belong to the main galaxy. Then, we define its stellar metallicity (as well as the abundance of each tracked species) as mass-averaged metallicity (abundance) of all star particles. To separate halo gas and the ISM, we apply a simple temperature cri-

teria and select all gas particles below 10^4 K as the ISM. In our simulations, this is equivalent to selecting gas above some density threshold of a few $0.1\,{\rm cm}^{-3}$ (we explicitly check the gas distribution in the density–temperature plane), which is comparable to the mean gas density within a few stellar effective radii. It naturally picks warm ionized and cold neutral gas. We define the gas-phase metallicity as the mass-weighted metallicity of all gas particles that belong to the ISM (we compare and discuss three different definitions of gas-phase metallicity in Appendix A).[3]

In this work, we use $Z_{\rm gas}$ and Z_* to refer to the mass fraction of all heavy elements in gas and stars, respectively. In Section 3, we will primarily use oxygen abundance $12 + \log({\rm O/H})$ to present gas-phase metallicities, which is defined in terms of number ratio of oxygen to hydrogen atoms, in order to directly compare with observations. For stellar metallicity, we will primarily use Z_* in the rest of this work. In the literature, gas-phase metallicity and stellar metallicity are also sometimes referred as $Z_{\rm gas}$ and iron abundance [Fe/H] (in solar units), respectively. For these reasons, we provide the conversion between these quantities for our simulated galaxies. We will show the calibration in Appendix B but directly give the

---

[3] In many cosmological simulations with 'sub-grid' models, gas-phase metallicity is usually defined as SFRs-averaged metallicity. However, our simulations explicitly resolve multiphase ISM structures and include realistic models of star formation and feedback. Individual gas particles are very sensitive to local feedback processes. For these reasons, we do not apply SF-averaged gas-phase metallicity to our simulations.
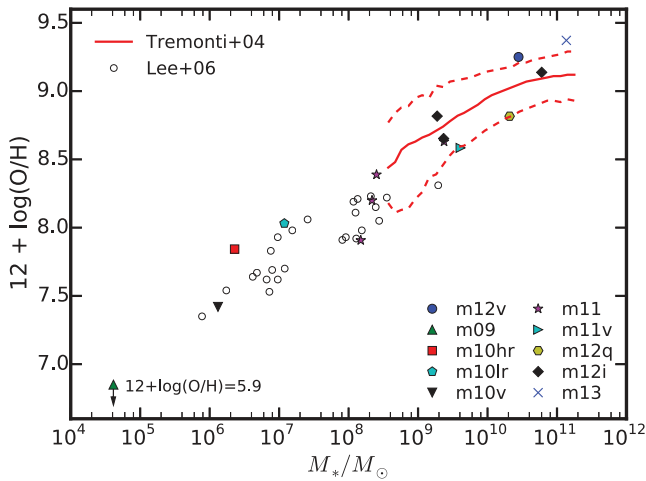
Figure 1. Stellar mass–gas-phase oxygen abundance relation at $z = 0$. The red solid and dashed curves represent the median and 2σ dispersion of the SDSS MZR at $z \sim 0.1$ (Tremonti et al. 2004). The open circles denote the data of the dwarf galaxy sample from Lee et al. (2006). Our simulations are in good agreement with observations from $M_* = 10^6$–$10^{11}\,M_\odot$.

results here. For a solar metallicity of 0.02 and a solar iron abundance 0.001 73 (both in mass fraction), we obtain $12 + \log(\mathrm{O/H}) = \log(Z_{\mathrm{gas}}/Z_\odot) + 9.00$ and $[\mathrm{Fe/H}] = \log(Z_*/Z_\odot) - 0.20$. We emphasize that these relations may suffer from systematic uncertainties that originate from: (1) Type II and Type I SNe rates; (2) metal yields of tracked species from different channels; and (3) the solar abundances we adopt in our simulations. However, the shape and evolution of the MZR should be robust to these uncertainties.

3 THE MASS–METALLICITY RELATION

In this section, we present both the gas-phase and stellar MZR from $z = 0$–6 and compare our results with observations and other simulations. We will further explore the most important factors that shape the MZR and drive its evolution in the Section 4.

3.1 The MZR at $z = 0$

We begin by showing the gas-phase MZR at $z = 0$. In Fig. 1, we present the stellar mass–gas-phase oxygen abundance relation for our mxx series simulations at $z = 0$. For comparison, we also present the median and 2σ dispersion of the SDSS MZR from Tremonti et al. (2004, red solid and dashed lines) and the data of individual local dwarf galaxies compiled in Lee et al. (2006, open circles) in Fig. 1. We remind the reader that these observed gas-phase oxygen abundances are derived from the relative strength of strong nebulae emission lines produced by photoionization from young massive stars, so that the observed gas-phase MZR only holds for star-forming galaxies. Also, we emphasize that the overall shape of gas-phase MZR strongly depends on which empirical calibration it uses and the normalization of this relation differs by at most 0.7 dex between different calibrations (Kewley & Ellison 2008, see also Fig. 6). For these reasons, we do not apply any correction to these observed data but keep them in their original forms.

In general, our simulations agree reasonably well with observations across stellar mass from $M_* = 10^6$–$10^{11}\,M_\odot$. However, our simulations do not show evidence for flattening at the high-mass end of the gas-phase MZR. The gas-phase metallicity increases with stellar mass up to $M_* \sim 10^{11}\,M_\odot$ in our sample. The simula-

tions predict slightly higher metallicities than the observed relation from Tremonti et al. (2004) at $M_* = 10^{11}\,M_\odot$. The most significant discrepancy is due to our m13 run, which is a somewhat lower resolution simulation of a massive galaxy and which did not include the possible effects of AGN feedback. For example, as it has been shown in Hopkins et al. (2014), the main galaxy in **m13** have the cooling flow problem and never quenches at low redshift. The SFR of this galaxy is $3\,M_\odot\,\mathrm{yr}^{-1}$, which is fairly low in its star formation history, but significantly higher than observationally inferred values below $z \sim 1$. Consequently, this galaxy might be overenriched at low redshift. If so, this suggests that additional physics, such as AGN feedback, is probably required to fully understand the chemical evolution in massive galaxies, at least in the sense of quenching star formation. Alternatively, it has also been proposed that the observed MZR could continue to rise at the high-metallicity end when using new metallicity diagnostics that account for non-equilibrium electron energy distributions (see e.g. Dopita et al. 2013; Nicholls et al. 2013). Furthermore, we note that the 'flatness' of MZR at the high-mass end behaves very differently when applying different empirical calibrations (e.g. Kewley & Ellison 2008). Therefore, we do not further quantitatively discuss the discrepancy between our simulations and observations at the massive-end of MZR, but rather focus on galaxies below $M_* = 10^{11}\,M_\odot$ where our simulations are most robust. A larger sample of simulations with improved resolution at the massive end is required to make a robust comparison.

Most of our simulated galaxies are still forming stars (at least very weakly) at $z = 0$, except for m09. The m09 is a low-mass isolated dwarf galaxy (comparable to the ultrafaint dwarfs around the Milky Way), in which star formation has been shut down since $z = 3$ by cosmic reionization (Oñorbe et al. 2015). At $z = 0$, this galaxy has lost almost all metals it produced (see Section 4). Although its gas-phase metallicity is an order of magnitude lower than the extrapolation of the observed MZR down to $M_* = 10^4\,M_\odot$, it is not contradictory to observations in the sense that the gas-phase metallicity of such galaxies cannot be measured due to lack of strong nebular emission lines.

In Fig. 2, we show the stellar mass–stellar metallicity relationship at $z = 0$ and compare our simulations with the SDSS sample from Gallazzi et al. (2005, red solid and dashed curves) and the dwarf
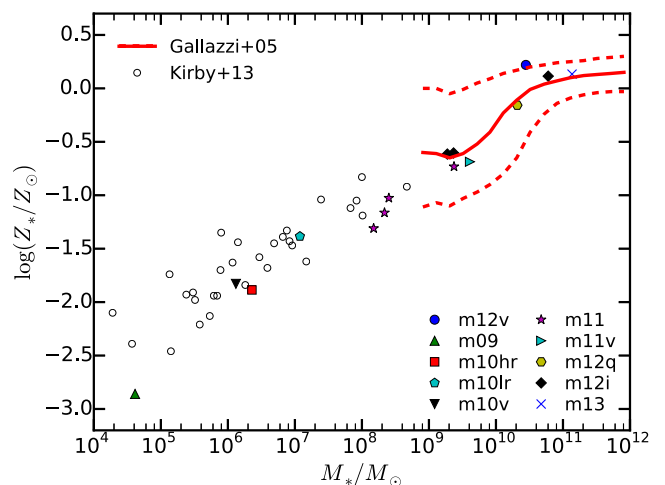
Figure 2. Stellar mass–stellar metallicity relation at $z = 0$. The red solid and dashed curves are the median and 1σ dispersion of the SDSS MZR in the local Universe (Gallazzi et al. 2005). The open circles represent the values of [Fe/H] of individual dwarfs from Kirby et al. (2013). Again, the agreement is good from 10^4–$10^{11}\,M_\odot$.
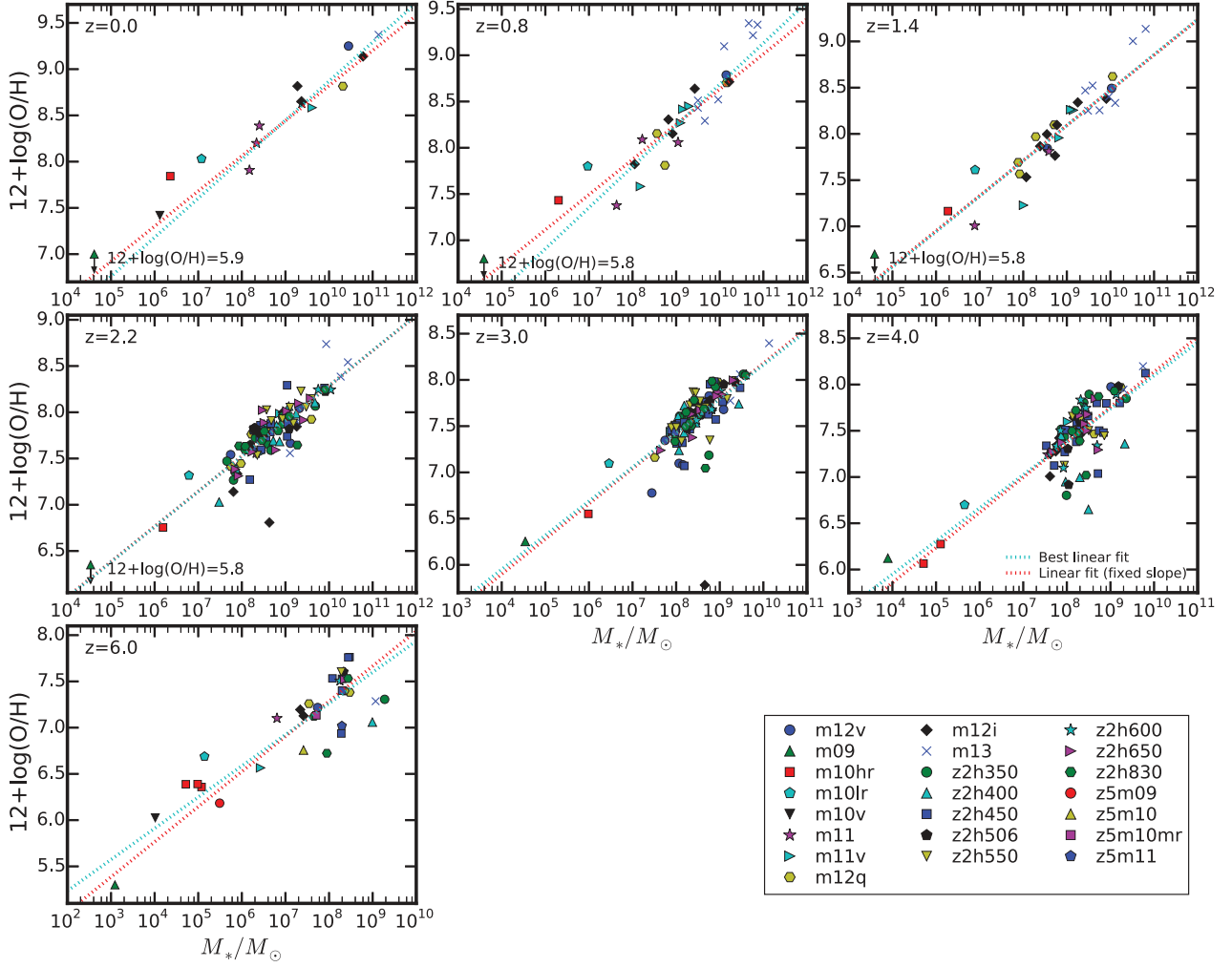
Figure 3. Stellar mass–gas-phase metallicity relation at all redshifts. Cyan dotted lines show the best linear fit $\log(Z_{gas}/Z_{\odot}) = 12 + \log(O/H) - 9.0 = \gamma_g[\log(M_*/M_{\odot}) - 10] + Z_{g,10}$. The red dotted lines show the best fit for a fixed slope $\gamma_g = 0.35$. Note that a constant slope provides a very good fit, where the zero-point evolves by ~ 1 dex from $z = 0$–6.

galaxies from Kirby et al. (2013, open circles). Note that the stellar metallicities from Gallazzi et al. (2005) are measured from absorption features of galaxy-integrated spectra (mostly Mg and Fe lines), while the metallicities from Kirby et al. (2013) are derived from Fe abundances of individual stars. The conversion between different methods and their systematic uncertainties are complex and beyond the scope of this paper. For our purposes, we avoid any correction to these observations but present them in their original values.[4]

Our simulations match these observations quite well over the whole mass coverage from $M_* = 10^4$–10^{11} M$_{\odot}$. The simulated sample shows a flatness in the stellar MZR around $M_* = 10^{11}$ M$_{\odot}$ at $z = 0$, consistent with the observed SDSS MZR from Gallazzi et al. (2005). This is the consequence of the fact that the growth of the more massive galaxies in our simulations is dominated by mergers and accretion of low-mass metal-poor satellites rather than *in situ* star formation at low redshifts. Therefore, the average stellar metal-

licities do not strongly evolve despite the fact that the stellar masses may grow considerably at low redshifts (see also Choi et al. 2014).

3.2 Evolution of the mzr

Figs 3 and 4 show the gas-phase and stellar MZR, respectively, from $z = 0$–6. We note that for $z \gtrsim 2$ and $z = 6$, we include the z2hxxx and z5mxx simulations in our analysis. The stellar MZR is tighter than the gas-phase MZR, i.e. the gas-phase MZR has larger scatter than stellar MZR at fixed stellar mass. This is because in our simulations, especially at high redshifts, star formation is dominated by multiple bursts, which drives bursts of gas outflows (Muratov et al. 2015). As a consequence, instantaneous gas-phase metallicities may have considerable time fluctuations associated with gas inflows, outflows, and mergers. This effect is larger at high redshifts when the galaxy progenitors are of much lower masses and galaxy mergers are more common, resulting in some outliers that deviate from the main MZR at high redshifts. Despite the short-time-scale fluctuations, both the gas-phase and stellar metallicities increase with time on cosmological time-scales. At all times, gas-phase metallicities are higher than stellar metallicities, since gas-phase metallicities represent the current state of metal enrichment in the galaxies, while

---

[4] In Fig. 2, we plot the values of [Fe/H] from Kirby et al. (2013), avoiding the complicated conversion between [Fe/H] and $Z_*/Z_{\odot}$ for the observed sample.
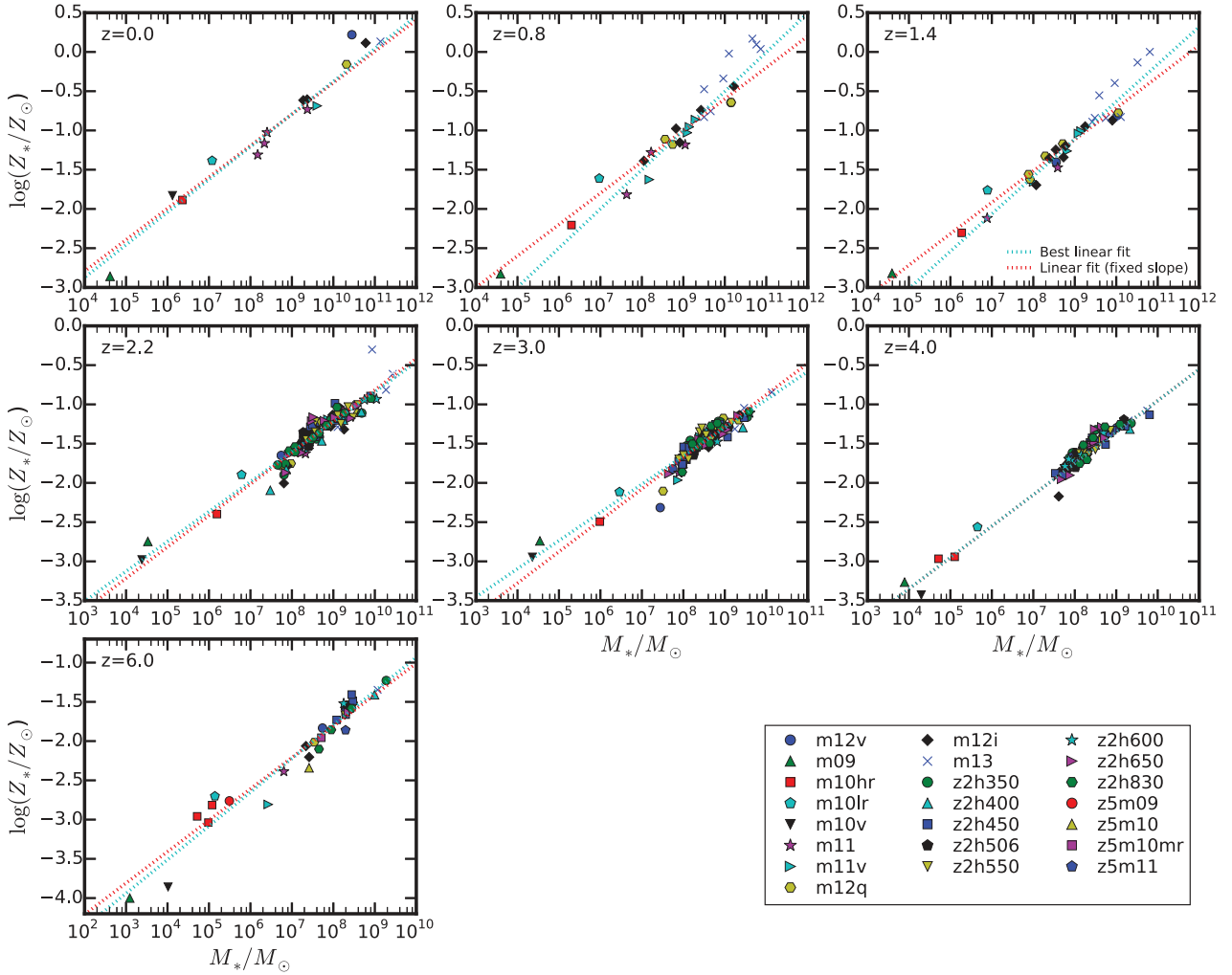
Figure 4. Stellar mass–stellar metallicity relation at all redshifts. Cyan dotted lines show the best linear fit at each redshift $\log(Z_*/Z_\odot) = [Fe/H] + 0.2 = \gamma_*[\log(M_*/M_\odot) - 10] + Z_{*,10}$. The red dotted lines show the best fit for a fixed slope $\gamma_* = 0.40$. Again, the slope is approximately constant, while the normalization evolves by ~ 1 dex.

stellar metallicities reflect the average galactic metallicities across the whole time. Both metallicities should converge at high redshifts.

To illustrate this quantitively, we fit the gas-phase and stellar MZR at different redshifts for our simulated galaxies with simple linear functions $\log(Z_{gas}/Z_\odot) = 12 + \log(O/H) - 9.0 = \gamma_g[\log(M_*/M_\odot) - 10] + Z_{g,10}$ and $\log(Z_*/Z_\odot) = [Fe/H] + 0.2 = \gamma_*[\log(M_*/M_\odot) - 10] + Z_{*,10}$, where γ_g and γ_* are the slopes and $Z_{g,10}$ and $Z_{*,10}$ represent the typical gas-phase metallicity and stellar metallicity at $M_* = 10^{10}\,M_\odot$. Although simple linear function do not capture the flatness of stellar metallicity above $M_* \sim 10^{11}\,M_\odot$ at $z < 1$, it is sufficient for our purposes here. We use least-squares fitting to obtain the best fit (the cyan dotted lines in Figs 3 and 4). In principle, both the slopes and zero-points should be functions of redshift. Nevertheless, the MZR at different redshifts have very similar slopes. For simplicity, we pick the mean slope of each relation and redo the linear fit using fixed slopes. We choose $\gamma_g = 0.35$ and $\gamma_* = 0.40$ (red dotted lines in Figs 3 and 4) and confirm that both the best linear fit and the fixed-slope fit describe the simulations reasonably well. We then attribute the evolution of MZR to the evolution of $Z_{g,10}$ and $Z_{*,10}$ with redshift, which we show in Fig. 5. We fit these parameters as a function of redshift by an exponential function
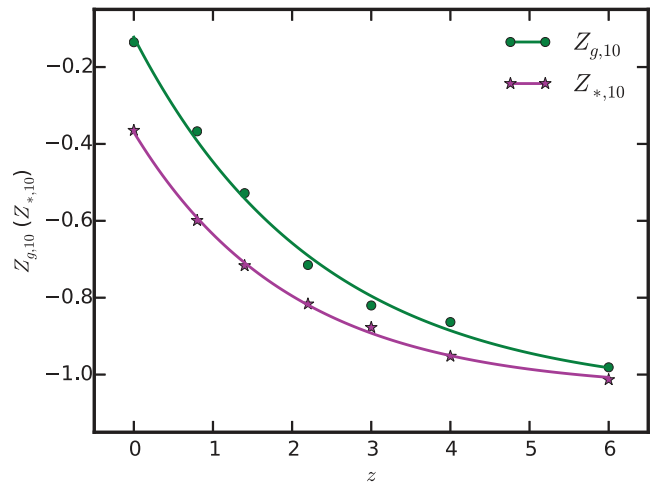
Figure 5. The gas-phase and stellar metallicity at $M_* = 10^{10}\,M_\odot$, $Z_{g,10}$ and $Z_{*,10}$ as a function of redshift. The solid lines are the best fit of exponential functions $Z_{g,10} = 0.93\exp(-0.43z) - 1.05$ and $Z_{*,10} = 0.67\exp(-0.50z) - 1.04$.
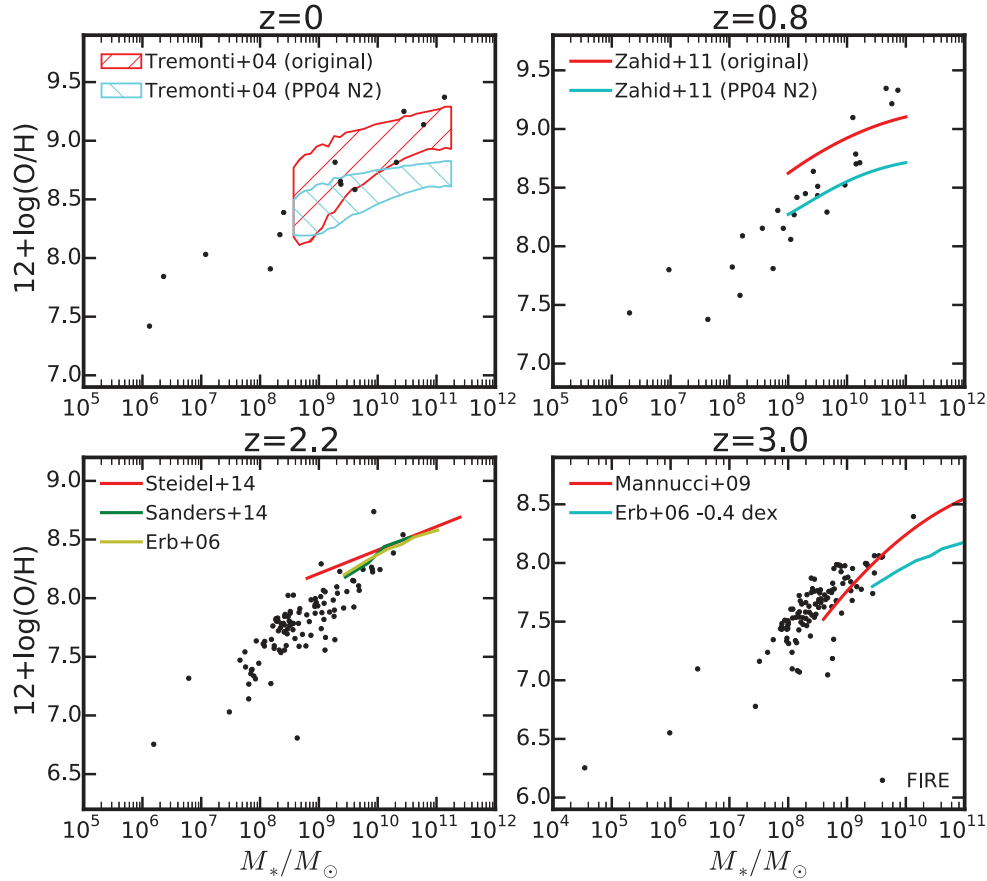
Figure 6. Stellar mass–gas-phase oxygen abundance relations at $z = 0$, 0.8, 2.2, and 3.0, as compared with a number of observations at these redshifts. In the upper panels, we show both the original relations (red lines) from Tremonti et al. (2004, $z \sim 0$) and Zahid et al. (2011, $z \sim 0.8$), and the relations converted to PP04 N2 calibration (cyan lines) following Kewley & Ellison (2008). In the lower-left panel, we show the observed MZR at $z \sim 2.3$ from Steidel et al. (2014, the red line), Sanders et al. (2015, the green line), and Erb et al. (2006, the yellow line). In the lower-right panel, we show the best fitting from Mannucci et al. (2009, $z \sim 3.1$). We also shift the Erb et al. (2006) data downwards by 0.4 dex for a comparison as motivated by fig. 5 in Mannucci et al. (2009). Our simulations are broadly consistent with observations over a wide range of stellar mass from $z = 0$–3, given the significant systematic uncertainties observational determinations of metallicities.

$F(z) = A \exp(-Bz) + C$. The best fit gives $Z_{g, 10} = 0.93 \exp(-0.43z) - 1.05$ and $Z_{*, 10} = 0.67 \exp(-0.50z) - 1.04$, respectively (the green and magenta lines in Fig. 5). These give the gas-phase and stellar MZR from $z = 0$–6 as $\log(Z_{gas}/Z_{\odot}) = 12 + \log(O/H) - 9.0 = 0.35 \quad [\log(M_*/M_{\odot}) - 10] + 0.93 \exp(-0.43z) - 1.05$ and $\log(Z_*/Z_{\odot}) = [Fe/H] + 0.2 = 0.40 \quad [\log(M_*/M_{\odot}) - 10] + 0.67 \exp(-0.50z) - 1.04$, respectively.

In general, the fitting functions above represent the gas-phase and stellar MZR fairly well for our simulated galaxies, except for the flattening of the stellar MZR above $M_* \sim 10^{11} \, \mathrm{M_{\odot}}$ at $z = 0$. We emphasize that these results have systematic uncertainties from Type II and Type Ia SNe rates, the solar abundance, and the metal yield tables we implement in our simulations. When using these fitting functions, one should notice the uncertainties and make adjustments accordingly.

3.3 Comparison with observations and other models

In Fig. 6, we compare the gas-phase MZR between our simulations and a number of observations at multiple redshifts. We show the observed MZR at $z \sim 0$ (Tremonti et al. 2004), $z \sim 0.8$ (Zahid et al. 2011), $z \sim 2.2$ (Erb et al. 2006; Steidel et al. 2014; Sanders et al. 2015), and $z \sim 3.1$ (Mannucci et al. 2009). We recall that these

observed relations are originally obtained using different calibrations, and the systematic uncertainty between different metallicity diagnostics could be up to 0.7 dex (Kewley & Ellison 2008). To illustrate this point, we also convert all the observed relation to the N2 calibration from Pettini & Pagel (2004, hereafter PP04) unless their original data are already presented using this calibration. For Tremonti et al. (2004) and Zahid et al. (2011), we do the conversion following the formula from Kewley & Ellison (2008, table 3 therein). In either case, we present both their original relations and the converted relations using PP04 N2 calibration in Fig. 6. At $z \sim 2.2$, the observed relations are at already presented in PP04 N2 calibration (e.g. Erb et al. 2006; Steidel et al. 2014; Sanders et al. 2015). Mannucci et al. (2009) adopted a very different metallicity calibration, which is established using $z \gtrsim 3$ galaxy samples only. Fig. 5 in Mannucci et al. (2009) suggests that the MZR evolves by ~0.4 dex from $z \sim 3.1$–~2.2. Motivated by their results, we also move the $z \sim 2.2$ MZR from Erb et al. (2006) downwards by 0.4 dex for a comparison (lower-right panel in Fig. 6).

In general, our simulations are in reasonable agreement with these observations in a broad range of stellar mass at $z = 0$–3, especially when the observed relations are in their original forms. We emphasize that the empirical calibrations developed from the local Universe are not necessarily valid for high-redshift galaxies (e.g. Steidel et al. 2014; Kewley et al. 2015). Given the large
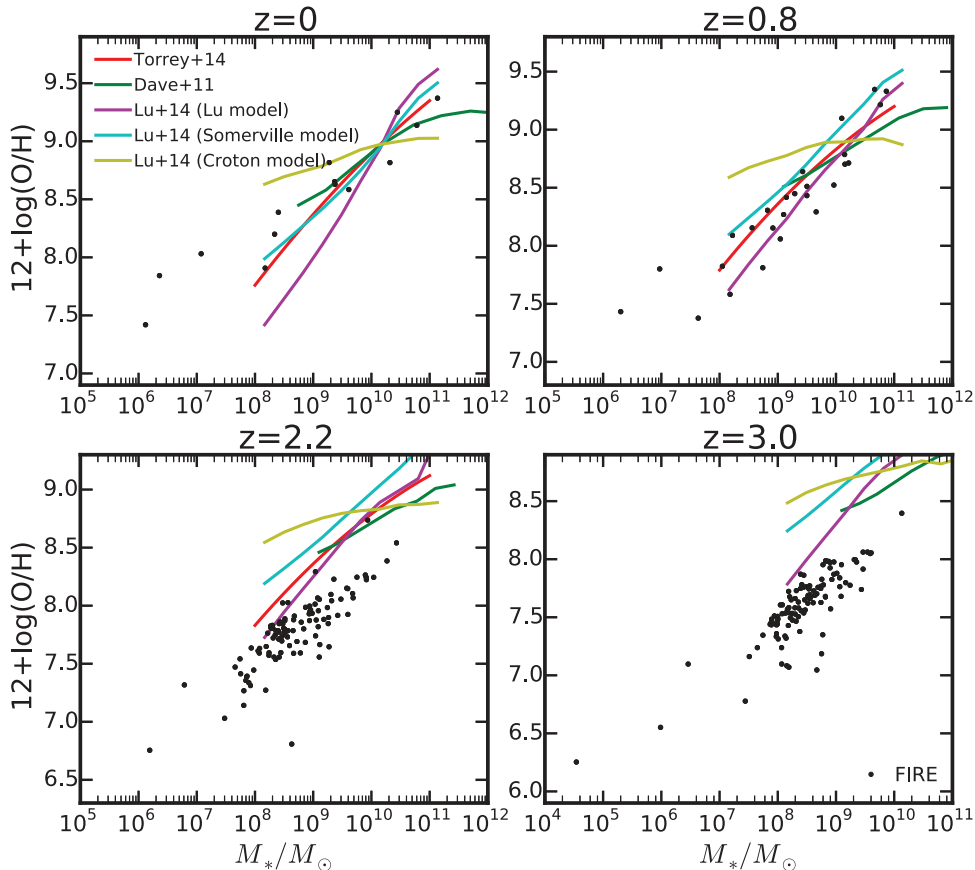
Figure 7. Stellar mass–gas-phase oxygen abundance relation at $z = 0$, 0.8, 2.2, and 3.0, as compared with other numerical simulations and SAMs. We renormalize other works to $12 + \log(O/H) = 8.9$ at $M_* = 10^{10}\,M_\odot$ at $z = 0$ with respect to our simulations. Red and green lines show the results from cosmological simulations presented in Torrey et al. (2014) and Davé et al. (2011b), respectively, which used popular 'sub-grid' models for galactic winds. Magenta, cyan, and yellow lines show the predictions of three SAMs from Lu et al. (2014, the Lu model, the Somerville model, and the Croton model, respectively). All of these models reproduce the correct $z = 0$ SMF, but none of them correctly reproduces the slope or the redshift evolution of the MZR.

systematic uncertainties, we do not provide a detailed quantitative discussion of the discrepancies between our simulations and observations. Our results on the evolution of the MZR in Section 3.2 are predictions that can be tested more accurately as our understanding of the observational systematic uncertainties improves.

In Fig. 7, we also compare the MZR from our simulations with other cosmological simulations and SAMs. We compare our results with two other simulations, Torrey et al. (2014, red lines) and Davé et al. (2011b, green lines), and three SAMs from Lu et al. (2014, the Lu model, magenta; the Somerville model, cyan; the Croton model, yellow). These models adopt 'sub-grid' empirical models of galactic winds and stellar feedback, which couple some fraction of energy and/or momentum from SNe to the gas, and force certain amount of the gas to escape the galaxy. Note that the metal yields and solar abundance used in different works are not exactly the same, we renormalize all the $z = 0$ MZR to $12 + \log(O/H) = 8.9$ at $M_* = 10^{10}\,M_\odot$ for comparison. At $z = 0$, Torrey et al. (2014) and the Lu model show steeper slopes at the low-mass end, due to the low metal retention efficiency in low-mass galaxies, a consequence of invoking strong outflows to suppress star formation in these galaxies.[5] Some models predict higher metallicities at the most

massive end. Furthermore, these models show significant discrepancies at $z \gtrsim 2$. Our simulations predict much stronger evolution of MZR from $z = 3$–0 than any other models. Particularly, the Somerville model and the Croton model predict inverse evolution trends – the gas-phase metallicity decreases at lower redshifts at fixed stellar mass – in contrast with observations and other models. We recall that although these models are tuned to match the observed SMF at $z = 0$, they tend to predict systematically higher SMFs than the observed ones for $M_* \lesssim 10^{11}\,M_\odot$ at $z > 0$ (Somerville & Davé 2014), a consequence of the fact that galaxies in these models form *too* many stars at early time (e.g. Davé et al. 2011a; Fiacconi et al. 2015; Sparre et al. 2015). In Section 4.4, we further explore how the different star formation histories between these models cause the discrepancies in the MZR at high redshifts.

---

[5] This can be simply illustrated using the 'leaky box' model (e.g. Schmidt 1963). Assuming the outflow rate is proportional to the SFR ($\dot{M}_{out} = \eta \cdot$ SFR, where η is the mass loading factor), the metallicity is

inversely proportional to $1 + \eta$. Low-mass galaxies are very efficient in driving outflows and thus have high-mass loading factors compared to massive galaxies. In SAMs and some simulations with 'sub-grid' feedback models, it is often assumed that either the metals are well mixed in the system or that the outflowing gas has a metallicity comparable to the metallicity in the ISM. As a consequence, low-mass galaxies tend to lose a large fraction not only of their gas but also of their metals, and therefore to end up with very low metallicities.

Figure 8. Metal retention fraction for our simulated galaxies at $z = 0$. M_Z ($< R_{vir}$) is the total amount of metals retained (in both gas and stars) within the virial radius. yM_* (y is the mean effective yield) is the total metal mass produced by stars in the galaxies. The retained fraction of metal in the halo increases with stellar mass, from 30 per cent at $M_* = 10^6\,M_\odot$ to about unity at $M_* > 10^{10}\,M_\odot$. However, the ultrafaint dwarfs (e.g. m09) are only able to retain 2 per cent of their metals in the halo.

4 DISCUSSION

We showed above that the gas-phase and stellar MZR in our simulations agree broadly with available observations at different redshifts. We also found that our predictions diverge significantly from those of several large-volume cosmological hydrodynamical simulations and SAMs. In this section, we explore the key factors that drive the shape and evolution of the MZR and discuss why our predictions differ from some other models.

4.1 Where are the metals?

Our simulations produce much higher metallicities for galaxies of stellar mass $M_* < 10^9\,M_\odot$ than Torrey et al. (2014) and the Lu model in Lu et al. (2014), indicating that our low-mass galaxies retain more metals compared to those models, despite the fact that these galaxies have high wind mass loading factors up to 100. To explicitly show this, we present in Fig. 8 the metal mass fraction retained within R_{vir} as a function of stellar mass for the simulated sample at $z = 0$. The numbers are obtained as follows. First, we estimate the effective yield y for every simulation as the ratio between total metal mass (in both gas and stars) and the total stellar mass in the whole simulation volume. Then, the metal retention fraction for a galaxy is simply the ratio between the total metal mass within the virial radius, $M_Z(< R_{vir})$, and yM_*, where M_* is the galaxy stellar mass as defined in Section 2.2. Thus, yM_* represents the total amount of metal ever produced by the stars in the galaxy. As shown in Fig. 8, the metal retention fraction generally increases with stellar mass. In our simulated sample, galaxies above $M_* = 10^{10.5}\,M_\odot$ are able to keep almost all metals they have produced. At much lower masses ($M_* = 10^6$–$10^7\,M_\odot$), they can still retain at least 30 per cent to a half of their metals within the halo. In contrast, the ultrafaint dwarf in our sample, m09 ($M_* = 4 \times 10^4\,M_\odot$), only retains 2 per cent of its metals within R_{vir} at $z = 0$.

To quantify in more detail how metals are retained in galaxy haloes, we show in Fig. 9 the cumulative metal retention fraction, as a function of radius, for different gas phases (cool gas with $T < 10^4$ K

and warm gas with $10^4\,\text{K} < T < 4 \times 10^5\,\text{K}$).[6] At $z = 0$ (top row), low-mass galaxies such as m10 ($M_* = 2 \times 10^6\,M_\odot$) have most of their metals in the warm CGM, while in massive galaxies like m12i ($M_* = 6 \times 10^{10}\,M_\odot$), the majority of the metals are found in stars. This trend is qualitatively consistent with the empirical halo metal budget presented in Peeples et al. (2014, Fig. 6). In most cases, we find that only a small fraction of the total metal mass is found in hotter ($T > 4 \times 10^5\,\text{K}$) gas. Our results are in contrast with the large-volume simulations of Ford et al. (2015) based on a parametrized galactic wind model, in which stars, ISM, and the cool CGM contain comparable metal masses for haloes of mass similar to our m12i run.

Regarding the spatial distribution of metals, in m11 ($M_* = 2 \times 10^9\,M_\odot$), over 60 per cent of the metals are concentrated in the central 0.1 R_{vir} (mostly the galaxy) and only a small fraction ($\lesssim 40$ per cent) of metals are in the CGM or lost into the IGM. In massive systems such as m12i ($M_* = 6 \times 10^{10}\,M_\odot$), almost all the metals are in the central 0.1 R_{vir}. In low-mass galaxies like m10 ($M_* = 2 \times 10^6\,M_\odot$), metals are more evenly distributed among the galaxy, the CGM, and the IGM. In ultrafaint dwarfs like m09 ($M_* = 4 \times 10^4\,M_\odot$), most of the metals it has ever produced are lost in the IGM by $z = 0$. This is consistent with the fact that outflows in low-mass galaxies are more efficient (they have much higher mass loading factor) and can propagate more easily to large radii than in massive systems (Muratov et al. 2015). It also shows that the metals are far from well mixed in the halo of more massive galaxies with stellar mass $M_* > 10^9\,M_\odot$ (or in terms of halo mass $M_{vir} > 10^{11}\,M_\odot$).

For a comparison, we also show the cumulative metal distribution for the progenitors of these galaxies at $z = 3$ (the bottom panel in Fig. 9). Similar to $z = 0$, a significant fraction of metals are still retained in R_{vir} at $z = 3$, although metals are more uniformly distributed from the centre to a few virial radii. These galaxies have much lower mass than their low-redshift decedents, and thus they are more efficient in driving gas outflows from star-forming regions throughout the halo.

4.2 Circumgalactic O VI

Although this paper is primarily focused on the metallicity of gas and stars inside galaxies, it is useful to check whether our simulations are consistent with observed CGM metal absorption. In addition to the overall metal budget discussed above, the COS-Haloes program has provided useful measurements of O VI absorption around $\sim L^*$ galaxies at $z \approx 0.1 - 0.4$ (Tumlinson et al. 2011). Fig. 10 shows the O VI column density map around our m12i simulated halo at $z = 0$. For this comparison, we assume that a fraction $f_{\text{O VI}} = 0.2$ of the oxygen is in O VI and only include warm and hot gas ($T > 10^4$ K) in the halo. $f_{\text{O VI}} = 0.2$ is the maximum expected if the oxygen is in collisional ionization equilibrium, though it is possible that O VI is also photoionized and/or subject to non-equilibrium effects (e.g. Oppenheimer & Davé 2009; Oppenheimer & Schaye 2013) so that this ionization fraction is not a strict upper limit. The figure shows that for m12i the characteristic $N_{\text{O VI}}$ drops from $\sim 10^{15}\,\text{cm}^{-2}$ at impact parameter $b = 20$ kpc from the central galaxy to $\sim 10^{13.5}\,\text{cm}^{-2}$ at $b = 200$ kpc. The simulation agrees well with the O VI columns measured by Tumlinson et al. (2011)

---

[6] In our simulations, most of the diffuse ($n_H < 0.1\,\text{cm}^{-3}$) gas has temperature $T > 10^4$ K so a temperature cut at $T = 10^4$ K also effectively separates ISM and CGM gas, justifying our approach of using gas with $T < 10^4$ K to evaluate gas-phase ISM metallicities.
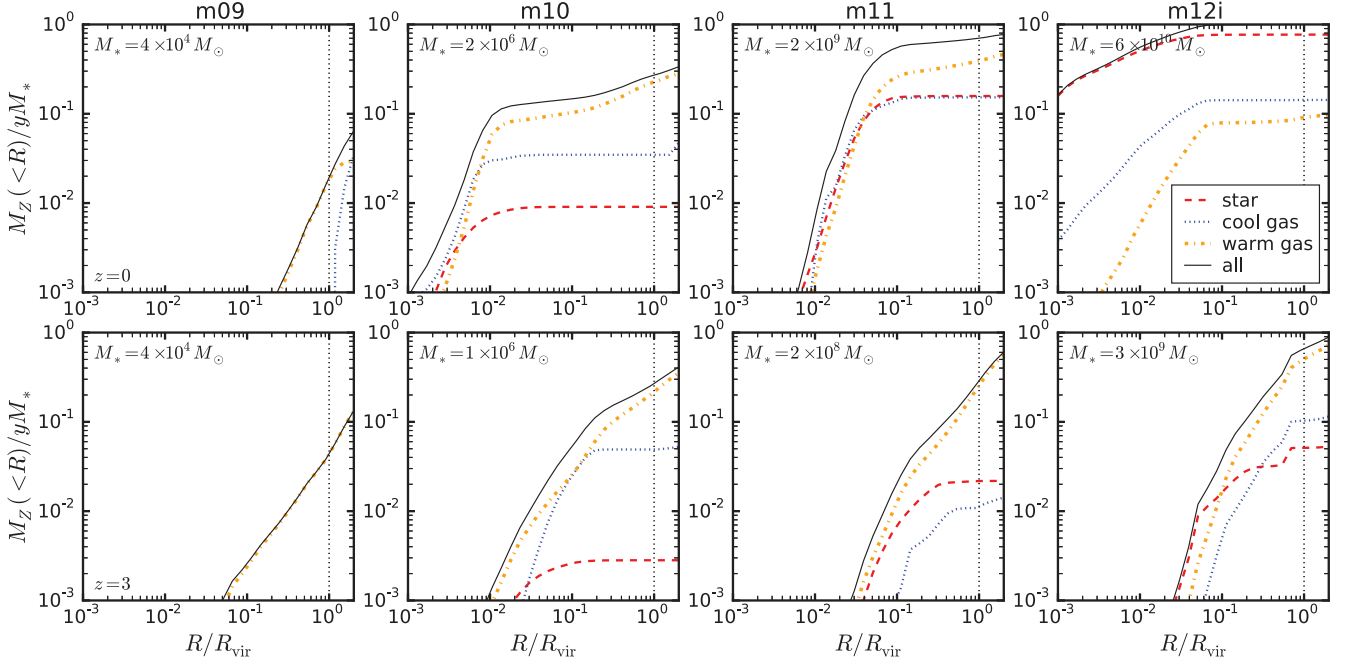
Figure 9. Cumulative metal mass in selected simulated haloes at $z = 0$ (top) and $z = 3$ (bottom), normalized by the total metal mass produced by stars (yM_*). The red dashed, blue dotted, orange dash–dotted, and black solid lines show the metal mass in stars, cool gas ($T < 10^4$ K), warm gas (10^4 K $< T < 4 \times 10^5$ K), and total, respectively. The stellar mass of each galaxy is indicated at the top left corner of each panel and the black dotted lines show the virial radius. At $z =$ 0, most of the metals in our more massive simulated galaxies such as m12i are in stars and within 0.1 R_{vir} of the halo centre, while in low-mass galaxies, the majority of metals are found in the warm CGM. In low-mass galaxies at $z = 0$ and in high-redshift galaxies, a larger fraction of the metals are found at larger radii from the halo centre, consistent with the fact that galactic outflows are more powerful in these systems.
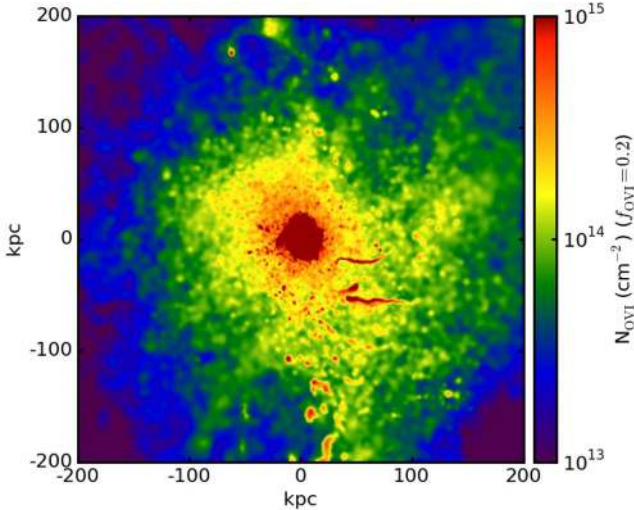
Figure 10. O VI column density map for the m12i halo at $z = 0$. We crudely assume that a fraction $f_{O\,VI} = 0.2$ of the oxygen is in O VI and only include warm and hot gas ($T > 10^4$ K) in the halo. The characteristic $N_{O\,VI}$ drops from $\sim 10^{15}$ cm^{-2} at impact parameter $b = 20$ kpc from the central galaxy to $\sim 10^{13.5}$ cm^{-2} at $b = 200$ kpc. The simulation agrees well with the O VI columns measured by COS-Haloes Tumlinson et al. (2011) around low-redshift $\sim L^*$ star-forming galaxies at impact parameters $b < 50$ kpc but appears to underestimate O VI columns by a factor of a few at larger impact parameters. Overall the agreement with observed O VI columns is reasonable given the uncertainties in ionization correction.

around low-redshift $\sim L^*$ star-forming galaxies at impact parameters $b < 50$ kpc but appears to underestimate O VI columns by a factor of a few at larger impact parameters. Overall the agreement with observed O VI columns is reasonable given the uncertainties in ionization correction. More systematic and detailed comparisons of CGM metal statistics from the FIRE simulations with observations will be reported in future papers (Hafen et al., in preparation).

4.3 Metal outflows, inflows, and recycling

SAMs and large-volume cosmological simulations require 'sub-grid' models of galactic winds, which often incorporate fairly crude approximations. In this subsection, we further examine the metal inflow and outflow rates and the metallicities of gas inflows and outflows in our simulations and compare with the assumptions of common 'sub-grid' models.

We follow Faucher-Giguère et al. (2011) and Muratov et al. (2015) and define the gas outflow rates, metal outflow rates, and metallicities of outflow gas as

$$\frac{\partial M}{\partial t} = \sum_i \boldsymbol{v} \cdot \frac{\boldsymbol{r}}{|r|} M_i / \mathrm{d}L, \qquad (1)$$

$$\frac{\partial M_{\mathrm{metal}}}{\partial t} = \sum_i \boldsymbol{v} \cdot \frac{\boldsymbol{r}}{|r|} Z_i \, M_i / \mathrm{d}L, \qquad (2)$$

$$Z_{\mathrm{outflow}} = \frac{\partial M_{\mathrm{metal}}}{\partial t} \bigg/ \frac{\partial M}{\partial t}, \qquad (3)$$
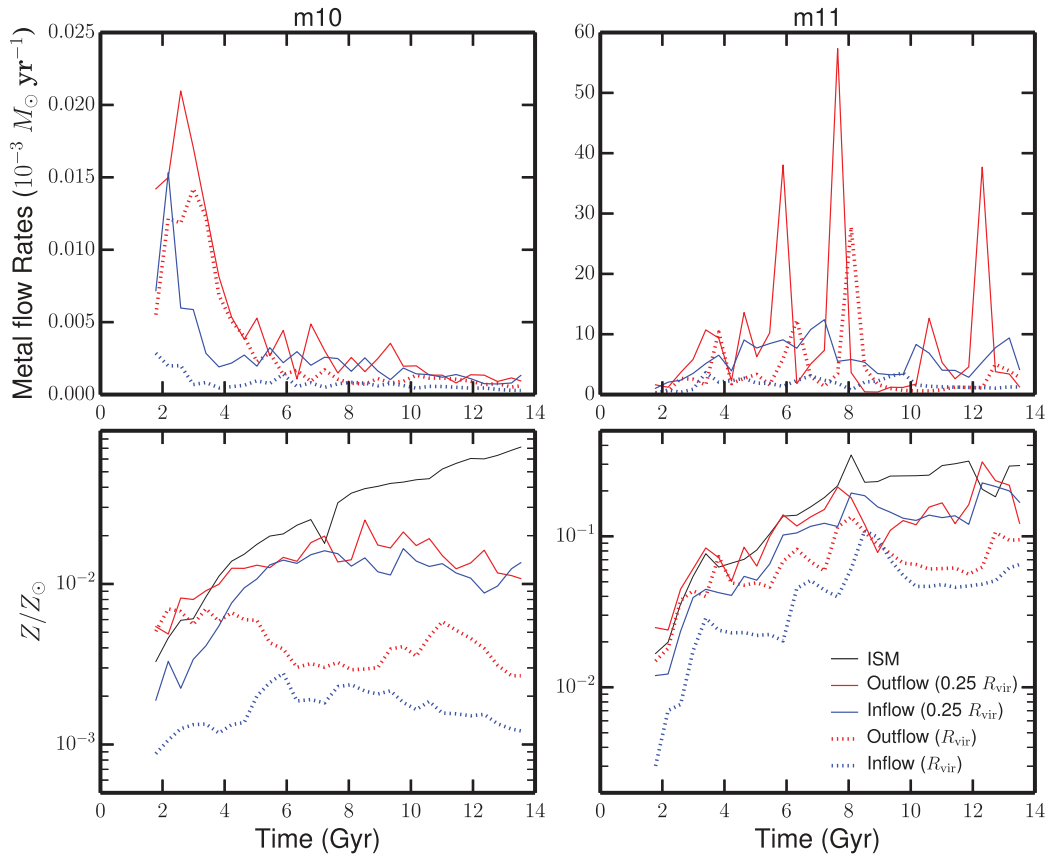
Figure 11. Upper: metal inflow (blue) and outflow rates (red) from $z = 0$–4. Solid and dotted lines show the metal inflow/outflow rates measured at 0.25 R_{vir} and R_{vir}, respectively. Bottom: metallicities of inflowing/outflowing gas. The black line shows the metallicity of the ISM. All quantities are averaged over a time-scale of 400 Myr. Metals are efficiently ejected in fountains reaching 0.25 R_{vir}, but they do not usually reach R_{vir} – they are either deposited in the halo or recycled efficiently in galactic fountains. Outflowing gas that escapes from the halo at R_{vir} tends to be less enriched than the gas in the ISM.

where M_i and Z_i are the mass and metallicity of the ith gas particle within the shell of thickness $dL = 0.1 R_{vir}$ with radial velocity outwards $\boldsymbol{v} \cdot \frac{\boldsymbol{r}}{|\boldsymbol{r}|} > 0$. The inflow rates and inflow metallicities are defined in the same way but for gas particles with inward radial velocity $\boldsymbol{v} \cdot \frac{\boldsymbol{r}}{|\boldsymbol{r}|} < 0$. The upper panels in Fig. 11 show that the metal inflow/outflow rates at 0.25 R_{vir} (blue/red solid lines) and at R_{vir} (blue/red dotted lines) for our m10 (left) and m11 (right) simulations. We average the inflow/outflow rates on a time-scale of 400 Myr. In either case, the net metal outflow rates are considerably lower at R_{vir} than at 0.25 R_{vir}, indicating that the metals are either deposited in the halo or returned back to the ISM. At high redshifts, metals ejected in outflows can be more easily driven to R_{vir} than at low redshifts. At 0.25 R_{vir}, metal inflow rates are comparable to metal outflow rates, suggesting a high efficiency of metal recycling. The lower panels in Fig. 11 show the average metallicities of inflows and outflows at both 0.25 R_{vir} and at R_{vir}, as compared to the metallicity of the ISM (black solid lines). The outflow metallicities are much lower at R_{vir} than at 0.25 R_{vir} (and even more than in the ISM) because outflowing gas sweeps up and mixes with more metal-poor gas in the halo as it propagates outwards. This is particularly important for low-mass galaxies, such as m10 ($M_* = 2 \times 10^6$ M$_\odot$), which can have wind mass loading factors up to ~100, yet retain a large fraction of the metals they produced in their haloes.

Our analysis calls into question a number of assumptions and approximations often adopted in analytic, semi-analytic, and large-volume cosmological hydrodynamic models of galaxy formation.

First of all, unlike often assumed in analytic and SAMs, metals are generally not well mixed in galaxy haloes (e.g. Fig. 9). In particular, in many 'sub-grid' galactic wind models, wind gas is assumed to have a metallicity directly related to the ISM metallicity (e.g. Davé et al. 2011a; Torrey et al. 2014), an assumption that oversimplifies the complex mass and metal loading that takes places in our more explicit simulations. Our simulations also indicate that metal reaccretion on to galaxies (recycling) is important on small scales, an effect which is not well captured in SAMs and in 'sub-grid' models that either assume that the ejected gas never returns to the galaxy, or which ignore hydrodynamical interactions between the wind and the gas close to the galaxy.

Recently, Lu et al. (2015b) compared three different SAM feedback models – one including only gas ejection, one including both gas ejection and recycling, and the other including a model of 'preventive' feedback. Lu et al. (2015b) found that none of these models could *simultaneously* reproduce the MZR, the distribution of metals in different phases, and the SFR observed at $z = 0$–3. This finding is consistent with the picture suggested by our high-resolution simulations that the chemical evolution of galaxies is a complex process and that it is necessary to self-consistently model galaxy–halo interactions in order to capture it faithfully. It is encouraging that our cosmological simulations with explicit stellar feedback and hydrodynamical interactions tracked at all times appear to produce a low-mass-end slope of the MZR that is closer to observations than most previous models, without the need for parameter tuning. Our results are broadly consistent with those of Brook et al. (2014), who
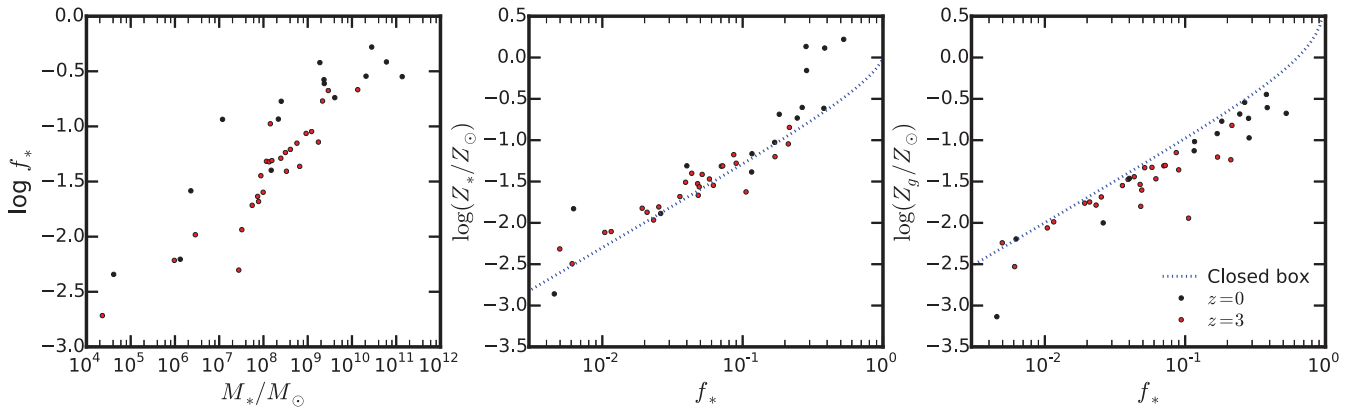
Figure 12. Left: stellar mass fraction $f_* = M_*/(M_{gas} + M_*)$ as a function of stellar mass. M_{gas} here is the total gas mass in the *halo* (not only in the galaxy). Middle: stellar metallicity Z_* as a function of f_*. Right: gas-phase metallicity Z_{gas} as a function of f_*. For consistency, Z_{gas} here is the average metallicity of *all* gas in the halo (including *both* the ISM *and* the halo gas). Black points and red points show the primary FIRE simulations at $z = 0$ and $z = 3$, respectively. Blue dotted lines show the simple 'closed box' model predictions assuming an effective metal yield of $y = 0.02$. The $z = 0$ and $z = 3$ galaxies share the same Z_*–f_* and Z_{gas}–f_* relations, but the f_*–M_* relation evolves by ~0.5 dex from $z = 3$–0. This indicates that the evolution of the MZR is associated with the evolution of f_* (at a fixed stellar mass) at different redshifts. The major offset between our simulations and the predictions of the 'closed box' model is largely due to the fact that the metals are not perfectly mixed throughout the halo. Especially in massive galaxies, gas tends to be more metal-enriched in the central star-forming regions than in the outer halo, so stellar metallicities tend to be higher and gas-phase metallicities (including the halo gas) are lower than the predictions of the 'closed box' model.

also highlighted the importance of metal mixing with the CGM and recycling for explaining the MZR. The simulations of Brook et al. (2014) also provide a fair match to the observed MZR at $z = 0$–3 (Obreja et al. 2014).

4.4 Why the MZR evolves with redshift?

Another major difference between our simulations and other theoretical work is we predict much stronger evolution of the MZR from $z = 3$–0 (e.g. the stellar metallicity increases by 0.5 dex at fixed stellar mass, see Fig. 5). Observations and some theoretical models suggest a fundamental metallicity relation (FMR) between stellar mass, SFR, and metallicity that holds for star-forming galaxies both in the local Universe and at high redshifts (e.g. Mannucci et al. 2010; Mannucci, Salvaterra & Campisi 2011; Lilly et al. 2013; Cullen et al. 2014; Obreja et al. 2014; Zahid et al. 2014). Motivated by these results, we attempt to qualitatively illustrate what might be the primary factor that drives the evolution of MZR in this section. We start by reviewing the simplest chemical evolution model, i.e. the 'closed box' model, which predicts the stellar and gas-phase metallicities as a function of stellar mass fraction, $f_* = M_*/(M_{gas} + M_*)$ as the following

$$Z_* = y \left[\frac{1 - f_*}{f_*} \ln(1 - f_*) + 1 \right], \tag{4}$$

$$Z_g = -y \ln(1 - f_*), \tag{5}$$

where y is the effective metal yield (e.g. Schmidt 1963; Talbot & Arnett 1971; Searle & Sargent 1972). The parameter f_* describes the fraction of baryons that have been turned into stars, and $1 - f_*$ is the 'gas fraction'. In Fig. 12, we show the relation between stellar and gas-phase metallicities and f_*, respectively (the middle- and right-hand panels), for our mxx series simulations at $z = 0$ and $z = 3$ (black and red points). We emphasize that we account for *both* the halo gas *and* the ISM in the total gas mass when calculating f_*, since halo gas is actively involved in supplying star formation and metal exchange in most cases. For consistency, the gas-phase metallicities shown in Fig. 12 is the average metallicity of *all* gas in the halo. For

illustrative purpose, we also show the simple predictions from the 'closed box' model, assuming an effective metal yield of $y = 0.02$ (blue dotted lines in Fig. 12).

The simulated data at $z = 0$ and 3 overlap with each other in the Z_*–f_* and Z_{gas}–f_* diagrams. In the left-hand panel of Fig. 12, we also show the relation between f_* and M_* for these galaxies at both redshifts. There is a systematic offset (~0.5 dex) in the f_*–M_* relation between galaxies at $z = 0$ and 3. Note that in the limit of $f_* \ll 1$, one has $Z_*, Z_{gas} \propto f_*$. Therefore, the 0.5 dex offset in f_*–M_* relation propagates to the 0.5 dex evolution of the MZR from $z = 3$ to 0. This suggests that the evolution of the MZR is associated with the evolution of f_* (at a fixed stellar mass) within the *halo* at different redshifts, providing a first hint of a universal metallicity relation between stellar mass, gas mass, and metallicities (cf. Bothwell et al. 2013; Zahid et al. 2014, for observational evidences). In simulations with 'sub-grid' feedback models and SAMs, where the $z = 0$ SMFs are tuned to match observations, galaxies tend to form a large fraction of their stars at high redshift and therefore their evolution is weaker at lower redshift (e.g. Somerville & Davé 2014), as opposed to observations and our simulations. In other words, these models produce higher f_* than our simulations at fixed stellar mass at $z > 0$ and an f_*–M_* relation barely evolving from $z = 3$–0. Therefore, galaxies in those models are more metal-enriched at high redshifts, and the evolution of the MZR is weaker than our simulations.

Our simulations are qualitatively consistent with the simple 'closed box' predictions applied to *halo* quantities.[7] This is not unreasonable because a large fraction (order unity) of metals is retained within the virial radius at both redshifts (see e.g. Fig. 9). However, we emphasize that one should not think our simulated

---

[7] We emphasize that in the analog of Fig. 12 where we measure f_* using *only* the gas in the galaxy (i.e. excluding the halo gas), all the galaxies are well below the predictions of the closed box model and there is no well-defined relation, indicating that galaxies themselves are far from closed boxes. This suggests the necessity of accounting for halo gas as reservoirs in galaxy evolution.

galaxies are closed boxes, because the metals are not perfectly well mixed in the galactic halo. This explains the major offset between the 'closed box' model and our simulations (middle and right-hand panels in Fig. 12), especially in the most massive systems where this effect is stronger. Since gas in the centre of the galaxy tends to be more metal-enriched than gas in the outer halo and stars preferentially form in the central region, stellar metallicities tend to be higher and the gas-phase metallicities (including the halo gas) are lower than the predictions of the closed box model (applied to halo quantities). The mixing of metals is very complex and associated with galactic fountains on different scales. Although the 'closed box' model gives a natural relation between stellar mass, gas mass, and the metallicities, the parametrization of a universal metallicity relation for galactic quantities (i.e. excluding the halo) is more complicated than the simple model. This is worth further investigation in more detail in future work.

5 CONCLUSION

We use a series of high-resolution cosmological zoom-in simulations spanning halo masses 10^9–10^{13} M$_\odot$ and stellar masses 10^4–10^{11} M$_\odot$ at $z = 0$ from the FIRE project to study the galaxy mass–metallicity relations at $z = 0$–6. These simulations include explicit models of multiphase ISM, star formation, and stellar feedback. As has been shown in previous papers, these simulations successfully reproduce many observed galaxy properties, including the stellar mass–halo mass relation, star-forming main sequence, the Kennicutt–Schmidt law, star formation histories, etc., for a wide range of galaxies at many redshifts (Hopkins et al. 2014). These simulations also predict reasonable covering fractions of neutral hydrogen in the haloes of $z = 2$–3 LBGs (Faucher-Giguère et al. 2015) and self-consistently generate galactic winds with velocities and mass loading factors broadly consistent with observational requirements (Muratov et al. 2015). These simulations adopt 'standard' stellar population models and metal yield tables from Type I and Type II SNe and stellar winds, following species-by-species for 11 separately tracked elements. Our key conclusions include the following.

(i) The simulations predict galaxy mass–metallicity relations that agree reasonably well with a number of observations from $z = 0$–3 for a broad range of stellar masses. Both gas-phase and stellar metallicities evolve monotonically from $z = 0$–6, with higher metal abundance at low redshifts at fixed stellar mass. The best linear fits of the MZR for our simulated galaxies as a function of redshift are $\log(Z_{gas}/Z_\odot) = 12 + \log(O/H) - 9.0 = 0.35 \quad [\log(M_*/M_\odot) - 10] + 0.93 \exp(-0.43z) - 1.05$ and $\log(Z_*/Z_\odot) = [Fe/H] + 0.2 = 0.40 \quad [\log(M_*/M_\odot) - 10] + 0.67 \exp(-0.50z) - 1.04$, for gas-phase metallicity and stellar metallicity, respectively. We emphasize that the normalizations may have systematic uncertainties that originate from the SNe rates, yield tables, and solar abundance we adopt, but the evolution of the MZR is robust to these uncertainties.

(ii) The stellar MZR becomes flat around $M_* \sim 10^{11}$ M$_\odot$ since $z = 0$, because the most massive galaxies in our simulations evolve via mergers and accretion of satellites rather than *in situ* star formation at low redshifts. Therefore, the stellar metallicity does not increase despite the fact that the stellar mass grows considerably. We do not see the flatness in the gas-phase MZR at the high-mass end seen in observations because gas continues to be enriched by non-negligible star formation. This apparent discrepancy may be due to the more limited resolution in our m13 run or to the lack

of AGN feedback in our simulations. AGN might be required to quench star formation below $z \sim 1$ in such massive galaxies.

(iii) The evolution of MZR is associated with the evolution of the gas/stellar mass fraction within the inner halo (not just inside the galaxy effective radius) at different redshifts. This provides a first hint of a universal metallicity relation between stellar mass, gas mass, and metallicities, but its parametrization for galactic quantities (as opposed to for halo quantities, which behave more like a closed box) is much more complicated than simple analytic models. We will investigate this in more detail in future work.

(iv) Galaxies above 10^6 M$_\odot$ can retain a large fraction of their metals in the halo even up to $z = 3$. The net metal outflow rates near the virial radius are always lower than those near the galaxy, indicating that the metals either get deposited in the halo or return back to the ISM. The high metal inflow rates and the high metallicity of inflowing gas at $0.25\,R_{vir}$ suggest a high efficiency of metal recycling (a finding that we have confirmed using particle tracking; Anglés-Alcázar et al., in prep.). On average, the outflows at outer radii are much less metal-enriched than those at the inner radius. This effect helps resolve the tension between the need for strong gas outflows and high metal retention fractions in low-mass galaxies.

(v) These differential recycling and metal retention effects are not properly accounted for in most semi-analytic and early generation of 'sub-grid' feedback models that are popular in cosmological simulations. As a result, these simplified models cannot simultaneously reproduce the galaxy mass function and the slope and redshift evolution of the MZR. By explicitly resolving the 'missing physics' in these models, we reconcile the longstanding discrepancy, and provide a clear way forward to improve the sub-grid and SAMs.

Nevertheless, our simulations are still limited in sample size. In the near future, we will expand our simulations to include more dwarf galaxies covering halo mass from $M_{halo} = 10^8$–10^{11} M$_\odot$ and to enlarge our sample at the most massive end to better understand whether the flattening of the MZR is real and what drives the flatness. This may depend critically on AGN feedback. We will provide quantitative analysis on metal outflow rates, outflow metallicities, metal recycling, and their relation with galaxy properties in future work (Muratov et al., in preparation; Anglés-Alcázar et al, in preparation).

REFERENCES

Andrews B. H., Martini P., 2013, ApJ, 765, 140
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Benson A. J., 2012, New Astron., 17, 175
Bertone S., De Lucia G., Thomas P. A., 2007, MNRAS, 379, 1143
Bothwell M. S., Maiolino R., Kennicutt R., Cresci G., Mannucci F., Marconi A., Cicone C., 2013, MNRAS, 433, 1425
Brook C. B., Stinson G., Gibson B. K., Shen S., Macciò A. V., Obreja A., Wadsley J., Quinn T., 2014, MNRAS, 443, 3809
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Chan T. K., Kereš D., Oñorbe J., Hopkins P. F., Muratov A. L., Faucher-Giguère C.-A., Quataert E., 2015, MNRAS, 455-PLX2, 2981
Choi J. et al., 2014, ApJ, 792, 95
Croton D. J. et al., 2006, MNRAS, 365, 11
Cullen F., Cirasuolo M., McLure R. J., Dunlop J. S., Bowler R. A. A., 2014, MNRAS, 440, 2300
Dalcanton J. J., 2007, ApJ, 658, 941
Davé R., Oppenheimer B. D., Finlator K., 2011a, MNRAS, 415, 11
Davé R., Finlator K., Oppenheimer B. D., 2011b, MNRAS, 416, 1354
Davé R., Finlator K., Oppenheimer B. D., 2012, MNRAS, 421, 98
Dekel A., Silk J., 1986, ApJ, 303, 39
Dopita M. A., Sutherland R. S., Nicholls D. C., Kewley L. J., Vogt F. P. A., 2013, ApJS, 208, 10
Edmunds M. G., 1990, MNRAS, 246, 678
Erb D. K., Shapley A. E., Pettini M., Steidel C. C., Reddy N. A., Adelberger K. L., 2006, ApJ, 644, 813
Faucher-Giguère C.-A., Kereš D., 2011, MNRAS, 412, L118
Faucher-Giguère C.-A., Lidz A., Zaldarriaga M., Hernquist L., 2009, ApJ, 703, 1416
Faucher-Giguère C.-A., Kereš D., Dijkstra M., Hernquist L., Zaldarriaga M., 2010, ApJ, 725, 633
Faucher-Giguère C.-A., Kereš D., Ma C.-P., 2011, MNRAS, 417, 2982
Faucher-Giguère C.-A., Quataert E., Hopkins P. F., 2013, MNRAS, 433, 1970
Faucher-Giguère C.-A., Hopkins P. F., Kereš D., Muratov A. L., Quataert E., Murray N., 2015, MNRAS, 449, 987
Fiacconi D., Feldmann R., Mayer L., 2015, MNRAS, 446, 1957
Finlator K., Davé R., 2008, MNRAS, 385, 2181
Ford A. B. et al., 2015, preprint (arXiv:1503.02084)
Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, MNRAS, 362, 41
Gill S. P. D., Knebe A., Gibson B. K., 2004, MNRAS, 351, 399
Guo Q. et al., 2013, MNRAS, 428, 1351
Hahn O., Abel T., 2011, MNRAS, 415, 2101
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R. E., Guo Q., Lemson G., Springel V., 2013, MNRAS, 431, 3373
Henriques B. M. B., White S. D. M., Thomas P. A. Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663
Henry A., Martin C. L., Finlator K., Dressler A., 2013a, ApJ, 769, 148
Henry A. et al., 2013b, ApJ, 776, L27
Hinshaw G. et al., 2013, ApJS, 208, 19
Hopkins P. F., 2013, MNRAS, 428, 2840
Hopkins P. F., 2014, MNRAS, 450, 53
Hopkins P. F., Quataert E., Murray N., 2011, MNRAS, 417, 950
Hopkins P. F., Quataert E., Murray N., 2012, MNRAS, 421, 3522
Hopkins P. F., Cox T. J., Hernquist L., Narayanan D., Hayward C. C., 2013a, Murray, N. MNRAS, 430, 1901
Hopkins P. F., Narayanan D., Murray N., 2013b, MNRAS, 432, 2647
Hopkins P. F., Keres D., Onorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581
Kereš D., Hernquist L., 2009, ApJ, 700, L1
Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2
Kewley L. J., Ellison S. L., 2008, ApJ, 681, 1183
Kewley L. J., Zahid H. J., Geller M. J., Dopita M. A., Hwang H. S., Fabricant D., 2015, ApJ, 812, L20
Kim J.-h. et al., 2014, ApJS, 210, 14

Kirby E. N., Cohen J. G., Guhathakurta P., Cheng L., Bullock J. S., Gallazzi A., 2013, ApJ, 779, 102
Knollmann S. R., Knebe A., 2009, ApJS, 182, 608
Kroupa P., 2002, Science, 295, 82
Kuhlen M., Faucher-Giguère C.-A., 2012, MNRAS, 423, 862
Lee H., Skillman E. D., Cannon J. M., Jackson D. C., Gehrz R. D., Polomski E. F., Woodward C. E., 2006, ApJ, 647, 970
Leitherer C. et al., 1999, ApJS, 123, 3
Lilly S. J., Carollo C. M., Pipino A., Renzini A., Peng Y., 2013, ApJ, 772, 119
Lu Y., Kereš D., Katz N., Mo H. J., Fardal M., Weinberg M. D., 2011, MNRAS, 416, 660
Lu Y. et al., 2014, ApJ, 795, 123
Lu Y., Blanc G. A., Benson A., 2015a, ApJ, 808, 129
Lu Y., Mo H. J., Lu Z., 2015b, preprint (arXiv:1504.02109)
Ma X., Kasen D., Hopkins P. F., Faucher-Giguère C.-A., Quataert E., Kereš D., Murray N., 2015, MNRAS, 453, 960
Maiolino R. et al., 2008, A&A, 488, 463
Mannucci F. et al., 2009, MNRAS, 398, 1915
Mannucci F., Cresci G., Maiolino R., Marconi A., Gnerucci A., 2010, MNRAS, 408, 2115
Mannucci F., Salvaterra R., Campisi M. A., 2011, MNRAS, 414, 1263
Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121
Muratov A. L., Keres D., Faucher-Giguere C.-A., Hopkins P. F., Quataert E., Norman N., 2015, MNRAS, 454, 2691
Nicholls D. C., Dopita M. A., Sutherland R. S., Kewley L. J., Palay E., 2013, ApJS, 207, 21
Nomoto K., Tominaga N., Umeda H., Kobayashi C., Maeda K., 2006, Nucl. Phys. A, 777, 424
Obreja A., Brook C. B., Stinson G., Domnguez-Tenreiro R., Gibson B. K., Silva L., Granato G. L., 2014, MNRAS, 442, 1794
Oñorbe J., Garrison-Kimmel S., Maller A. H., Bullock J. S., Rocha M., Hahn O., 2014, MNRAS, 437, 1894
Oñorbe J., Boylan-Kolchin M., Bullock J. S., Hopkins P. F., Kereš D., Faucher-Giguère C.-A., Quataert E., Murray N., 2015, MNRAS, 454, 2092
Oppenheimer B. D., Davé R., 2009, MNRAS, 395, 1875
Oppenheimer B. D., Schaye J., 2013, MNRAS, 434, 1043
Oppenheimer B. D., Davé R., Kereš D., Fardal M., Katz N., Kollmeier J. A., Weinberg D. H., 2010, MNRAS, 406, 2325
Padoan P., Nordlund Å., 2011, ApJ, 730, 40
Peeples M. S., Werk J. K., Tumlinson J., Oppenheimer B. D., Prochaska J. X., Katz N., Weinberg D. H., 2014, ApJ, 786, 54
Pettini M., Pagel B. E. J., 2004, MNRAS, 348, L59 (PP04)
Planck Collaboration XVI, 2013, A&A, 571, A16
Robertson B. E. et al., 2013, ApJ, 768, 71
Sanders R. L. et al., 2015, ApJ, 799, 138
Savaglio S. et al., 2005, ApJ, 635, 260
Scannapieco C. et al., 2012, MNRAS, 423, 1726
Schaye J. et al., 2015, MNRAS, 446, 521
Schmidt M., 1963, ApJ, 137, 758
Searle L., Sargent W. L. W., 1972, ApJ, 173, 25
Somerville R. S., Davé R., 2014, ARA&A, 53, 51
Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, MNRAS, 391, 481
Sparre M. et al., 2015, MNRAS, 447, 3548
Springel V., 2005, MNRAS, 364, 1105
Springel V., Di Matteo T., Hernquist L., 2005, ApJ, 620, L79
Steidel C. C. et al., 2014, ApJ, 795, 165
Talbot R. J., Jr, Arnett W. D., 1971, ApJ, 170, 409
Torrey P., Vogelsberger M., Genel S., Sijacki D., Springel V., Hernquist L., 2014, MNRAS, 438, 1985
Tremonti C. A. et al., 2004, ApJ, 613, 898
Tumlinson J. et al., 2011, Science, 334, 948
Veilleux S., Cecil G., Bland-Hawthorn J., 2005, ARA&A, 43, 769
Wiersma R. P. C., Schaye J., Smith B. D., 2009a, MNRAS, 393, 99
Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, MNRAS, 399, 574

Woosley S. E., Weaver T. A., 1995, ApJS, 101, 181

Yabe K. et al., 2014, MNRAS, 437, 3647

Yates R. M., Henriques B., Thomas P. A., Kauffmann G., Johansson J., White S. D. M., 2013, MNRAS, 435, 3500

Zahid H. J., Kewley L. J., Bresolin F., 2011, ApJ, 730, 137

Zahid H. J., Bresolin F., Kewley L. J., Coil A. L., Davé R., 2012, ApJ, 750, 120

Zahid H. J., Geller M. J., Kewley L. J., Hwang H. S., Fabricant D. G., Kurtz M. J., 2013, ApJ, 771, L19

Zahid H. J., Dima G. I., Kudritzki R.-P., Kewley L. J., Geller M. J., Hwang H. S., Silverman J. D., Kashino D., 2014, ApJ, 791, 130

APPENDIX A: DIFFERENT DEFINITIONS OF GAS-PHASE METALLICITY

In this work, the gas-phase metallicity is defined as the mass-weighted average metallicity of all gas particles below 10^4 K, which we refer as the ISM gas. In principle, there are many alternative approaches to define gas-phase metallicities. In this section, we discuss three definitions and compare them with each other: (1) the average metallicity of all gas particles below 10^4 K in the galaxy (our default definition); (2) the average metallicity of all gas particles within 0.1 $R_{\rm vir}$; and (3) the average metallicity of all gas particles with temperature between 7 000–15 000 K and density above 0.5 cm^{-3}. In Fig. A1, we compare definition (1) and (2) in the left-hand panel and (1) and (3) in the right-hand panel for all galaxies presented in Fig. 3.

Definition (1) is designed to automatically select all the warm ionized gas and cold neutral gas (the ISM), definition (2) aims to pick the gas in the star-forming regions, and definition (3) is observationally motivated to select the nebular gas which produce the strong nebular emission lines in star-forming galaxies. In general, these definitions are consistent with each other. Most of the galaxies lie very close to the $y = x$ relation in each panel of Fig. A1. However, there are a few outliers in these diagrams. Definition (2) can be problematic in merging systems, where the halo centre may deviate far from the stellar bulk, and thus 0.1 $R_{\rm vir}$ does not necessarily probe the star-forming region. Definition (3) is largely affected by abundance variance between gas particles, since there are usually not many gas particles at any single instant that meet the temperature and density criteria. However, a time-averaged version of definition (3) removes most of the outliers. Therefore, we argue that our default definition is more adaptive and flexible than other definitions.
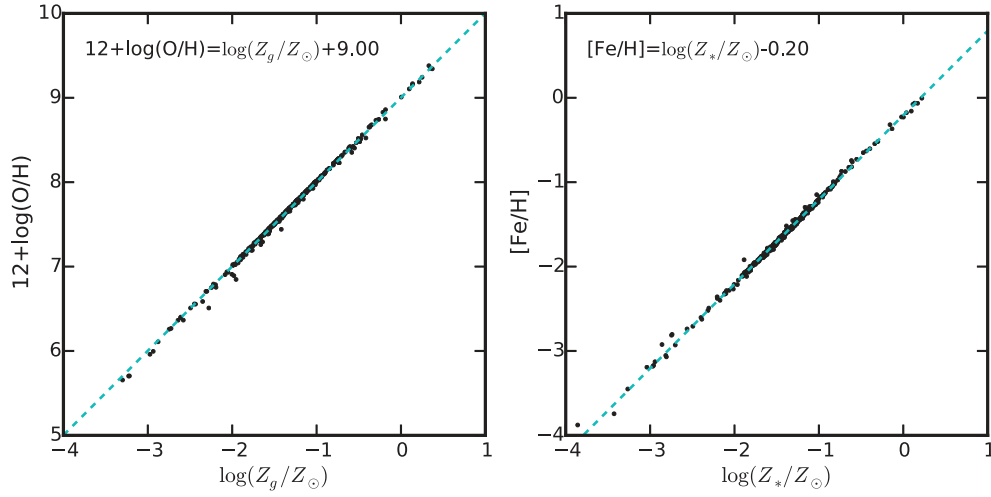
APPENDIX B: METALLICITIES IN DIFFERENT FORMS

In this work, we primarily use $12 + \log(\mathrm{O/H})$ and Z_* to present gas-phase metallicity and stellar metallicity, respectively. In the literature, gas-phase metallicity and stellar metallicity are sometimes presented in terms of $Z_{\rm gas}$ and [Fe/H]. Therefore, we also provide the conversion between these different forms of metallicities for comparison. We emphasize these conversions are obtained from our simulations only and there are systematic uncertainties originating from the uncertain relative metal yields between species and solar abundances we adopt.

In Fig. B1, we show the relations between $12 + \log(\mathrm{O/H})$ and $\log(Z_{\rm gas}/Z_\odot)$ (left-hand panel) and the relation between [Fe/H] and $\log(Z_*/Z_\odot)$ (right-hand panel), where we adopt a solar metallicity $Z_\odot = 0.02$ and a solar iron abundance of 0.001 73, both in mass fraction. In both panels, we collect data of all the simulated galaxies at all epochs we present earlier in this paper. Both relations are extremely tight and have slope unity, which ensures the validity, at least to the first order, to use either quantity to represent metallicities interchangeably. The best fits for our simulations are $12 + \log(\mathrm{O/H}) = \log(Z_{\rm gas}/Z_\odot) + 9.0$ and [Fe/H] $= \log(Z_*/Z_\odot) - 0.20$. We emphasize that there relations may suffer from systematic uncertainties that originate from: (1) Type II and Type I SNe rates, (2) metal yields of tracked species from different channels, and (3) the solar abundances we adopt in our simulations.
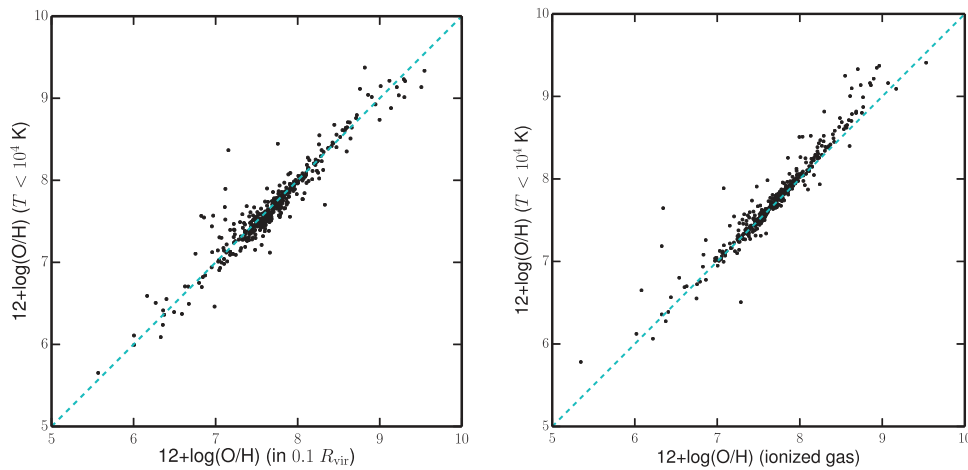
Figure A1. Gas-phase oxygen abundances in different definitions. Left: the relation of gas oxygen abundances between definition (1) the average metallicity of all gas particles below 10^4 K and (2) the average metallicity of all gas particles within 0.1 $R_{\rm vir}$. Right: The relation of gas oxygen abundances between definition (1) and (3) the average metallicity of all gas particles with temperature between 7 000 and 15 000 K and density above 0.5 cm^{-3}. The cyan dashed lines show the $y = x$ relation. The black points show all the data presented in Fig. 3. Different definitions agree well, and have *no* qualitative effect on any of our conclusions. Most of the 'outliers' are caused by transient, stochastic time variability.

Figure B1. Relations between different forms of metallicities. *Left*: Gas-phase oxygen abundance $12 + \log(\mathrm{O/H})$ versus gas-phase metallicity (mass fraction of all metals) Z_{gas}. *Right*: Stellar iron abundance [Fe/H] versus stellar metallicity Z_*. Black dots collect all the data points presented in this work. The cyan lines represent the best fits of these relations with slope unity. These definitions give essentially identical results, and are equivalent, for all of our results in this paper.

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