Physical Models of Galaxy Formation in a Cosmological Framework

Rachel S. Somerville[1] and Romeel Davé[2]

[1]Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854; email: somerville@physics.rutgers.edu

[2]Department of Physics, University of the Western Cape, Cape Town 7530, South Africa, South African Astronomical Observatories, Cape Town 7935, South Africa, and African Institute for Mathematical Sciences, Cape Town 7945, South Africa; email: romeeld@gmail.com

Annu. Rev. Astron. Astrophys. 2015. 53:51–113

First published online as a Review in Advance on April 16, 2015

The *Annual Review of Astronomy and Astrophysics* is online at astro.annualreviews.org

This article's doi: 10.1146/annurev-astro-082812-140951

Keywords

galaxy formation, galaxy evolution, numerical simulations, cosmology

Abstract

Modeling galaxy formation in a cosmological context presents one of the greatest challenges in astrophysics today due to the vast range of scales and numerous physical processes involved. Here we review the current status of models that employ two leading techniques to simulate the physics of galaxy formation: semianalytic models and numerical hydrodynamic simulations. We focus on a set of observational targets that describe the evolution of the global and structural properties of galaxies from roughly cosmic high noon ($z \sim 2$–3) to the present. Although minor discrepancies remain, overall, models show remarkable convergence among different methods and make predictions that are in qualitative agreement with observations. Modelers have converged on a core set of physical processes that are critical for shaping galaxy properties. This core set includes cosmological accretion, strong stellar-driven winds that are more efficient at low masses, black hole feedback that preferentially suppresses star formation at high masses, and structural and morphological evolution through merging and environmental processes. However, all cosmological models currently adopt phenomenological implementations of many of these core processes, which must be tuned to observations. Many details of how these diverse processes interact within a hierarchical structure formation setting remain poorly understood. Emerging multiscale simulations are helping to bridge the gap between stellar and cosmological scales, placing models on a firmer, more physically grounded footing. Concurrently, upcoming telescope facilities will provide new challenges and constraints for models, particularly by directly constraining inflows and outflows through observations of gas in and around galaxies.

1. INTRODUCTION

SED: spectral energy distribution

SFR: star-formation rate

SF: star formation

CDM: cold dark matter

SMBH: supermassive black hole

The past decade has seen remarkable progress in measuring the properties of galaxies across the electromagnetic spectrum and over the majority of cosmic history. Wide-field surveys have collected samples of millions of nearby galaxies, spanning roughly six orders of magnitude in galaxy mass and a rich range of galaxy types and environments, from isolated galaxies in voids to rich clusters. Medium-deep surveys have collected samples of tens of thousands of galaxies out to $z \sim 6$, and ultradeep surveys have identified samples of hundreds to thousands of galaxy candidates at $z \sim$ 6–8, with a few candidates identified (mainly behind lensing clusters) at $z \sim 9$–10 (for an overview of recent surveys see Madau & Dickinson 2014). The panchromatic wavelength coverage enabled by a suite of ground- and space-based telescopes has allowed construction of detailed spectral energy distributions (SEDs) for large samples of galaxies. These SEDs may be used to obtain estimates of photometric redshifts for galaxies that are too faint for spectroscopic measurements and estimates of intrinsic parameters, such as stellar masses and star-formation rates (SFRs).

In addition, high spatial resolution imaging, primarily from the *Hubble Space Telescope* (HST), and spectroscopy, including data from an increasing number of surveys using integral field spectrographs, have enabled study of galaxies' internal structure and kinematics. In particular, the Wide Field Camera 3 on the HST has made it possible to study galaxy structure and morphology in the rest-frame optical back to cosmic noon—the peak of cosmic star formation (SF) and black hole (BH) accretion activity at $z \sim 2$–3 (Madau & Dickinson 2014). We are truly living in a golden age of facilities and databases for studying how galaxies formed and evolved.

Concurrently over the past decade, advances in numerical methodologies and computing speed have allowed extraordinary progress in our ability to simulate the formation of structure within the paradigm of the cold dark matter (CDM) model (e.g., Springel et al. 2005c, Klypin et al. 2011). A variety of techniques have been developed for computing detailed predictions for the expected observable properties of galaxies on the basis of ab initio (albeit approximate) treatments of the physical processes expected to be important in shaping galaxy formation and evolution. The project of genuinely ab initio computational simulation of galaxy formation is beyond current capabilities; this is due to the vast range of spatial scales involved, from the subparsec scales of individual stars and supernovae (SNe), and accretion disks of supermassive BHs (SMBHs), to the super-megaparsec scales of the cosmic web, and to the wide array of poorly understood physical processes. However, by zeroing in on different scales through many different approaches, models are providing fundamental insights into the physical processes that are responsible for molding galaxy properties. Cosmological galaxy formation models have now matured into an essential tool for understanding galaxy evolution and, hence, it is timely to review this topic.

We focus specifically on state-of-the-art physically motivated cosmological models of galaxy formation and ask a series of questions: (*a*) How well are these models able to predict or reproduce the observed distribution functions of global galaxy properties, such as stellar mass, and the evolution of these functions? (*b*) How well do the models reproduce global scaling relations, such as correlations between stellar mass, cold gas fraction, SFR, and metallicity? (*c*) What do these models predict for the demographics of different types of galaxies (e.g., star-forming versus quiescent, or disk-dominated versus spheroid-dominated)? (*d*) Are models able to reproduce observed structural scaling relations—such as those relating mass with radial size, density, and internal velocity—and the evolution of these relations for different types of galaxies? (*e*) With regard to all these questions, what insights have we gained into the process of galaxy formation from the successes and failures of our current models?

The plan for the rest of this review is as follows. In Section 1.1, we give a broad overview of the observational results that we target in our review. In Section 1.2, we discuss the cosmological

background, mainly pointing the reader to other sources. In Section 1.3, we give a brief overview of the physical processes that are included in most state-of-the-art models of galaxy formation, and in Section 1.4, we introduce different tools for modeling galaxy properties. In Section 2, we give a more detailed description of the methods used in the models that we discuss in the remainder of the review, which include numerical hydrodynamic simulations and semianalytic models. In Section 3, we discuss the "subgrid" modeling connected with physical processes that are not directly resolved in cosmological simulations, including the formation of stars and SMBHs, and the impact of feedback from these objects on forming galaxies. In Section 4, we discuss the predictions of current models and how they measure up to observations for global properties of galaxies (Section 4.1) and galaxy internal structure and kinematics (Section 4.2). We conclude with a summary and outlook in Section 5.

SMF: stellar mass function

1.1. Observational Targets

In this review, we focus on the global and structural properties of the stellar components of galaxies from roughly cosmic high noon ($z \sim 2$–3) to the present. We acknowledge that there are many important observations that provide crucial constraints on models that lie beyond this scope. The summary here is quite brief; we refer the reader to other recent reviews and papers for a more comprehensive overview.

1.1.1. Global properties: distribution functions. Multiwavelength imaging surveys complemented with photometric or spectroscopic redshifts yield estimates of familiar global galaxy properties, such as the luminosity and color at various rest-frame wavelengths from the UV to far-IR. In recent years, it has become popular to estimate stellar masses by fitting galaxy SEDs with simple parametric models of galaxy SF histories combined with stellar population models (Walcher et al. 2011, Conroy 2013). SFRs are also estimated using SED modeling, or similarly using extinction-corrected rest-UV measures, but more reliably by adding mid- to far-IR photometry and/or nebular emission lines, such as Hα. We refer to the comoving number density of galaxies as a function of a global property, such as luminosity or stellar mass, as a distribution function. It has long been known that galaxy distribution functions typically have a characteristic shape often described by a Schechter function (Schechter 1976), which is parameterized by a normalization, a turnover, and an asymptotic slope to low masses. Examples of luminosity functions, stellar mass functions (SMFs), and the cold gas (atomic hydrogen) mass function of nearby galaxies from recent large surveys are presented in the review by Blanton & Moustakas (2009).

To higher redshifts, galaxy rest-frame optical–near-IR luminosity functions and SMFs have been measured from medium-deep surveys out to $z \sim 4$. At higher redshifts, SMF estimates exist but rely on stellar mass estimates from rest-UV fluxes, which are likely less robust. These measurements have yielded a number of important insights into galaxy assembly: (*a*) Galaxies appear to be continuously building up their mass over cosmic time, in accord with the hierarchical formation picture and inconsistent with monolithic collapse (Madau & Dickinson 2014). (*b*) The number density of massive galaxies ($m_{\rm star} > M_{\rm char}$, in which $m_{\rm star}$ is the stellar mass and $M_{\rm char}$ is the characteristic mass in the Schechter function) increases rapidly from $z \sim 4$–2 but then stays nearly constant or increases slowly from $z \sim 2$–0, indicating that massive galaxies formed and assembled their stars relatively early (Marchesini et al. 2009, Moustakas et al. 2013, Muzzin et al. 2013). (*c*) The comoving number density of low-mass galaxies ($m_{\rm star} < M_{\rm char}$) increases more rapidly than that of more massive galaxies at $z = 1$–2, indicating that low-mass galaxies formed their stars later and over a longer timescale. This result is sometimes called "mass assembly downsizing" (Cimatti et al. 2006).

It has long been known that the color-luminosity distribution of galaxies is strongly bimodal (e.g., Baldry et al. 2004), with most galaxies falling onto a relatively narrow (in optical colors) "red

sequence" or a broader "blue cloud." Spectroscopic indicators of stellar population age as well as UV and IR photometry have confirmed that in the local Universe, the red sequence is largely composed of quiescent galaxies with predominantly old stellar populations, whereas the blue cloud represents star-forming galaxies with younger stellar populations and significant ongoing SF (Kauffmann et al. 2003, Brinchmann et al. 2004, Salim et al. 2007, Schiminovich et al. 2007). Because of the strongly bimodal nature of the population, it has become common to draw a line in either the color-luminosity or color-mass plane, or in the specific SFR (sSFR $\equiv SFR/m_{\rm star}$) versus $m_{\rm star}$ plane, and to speak of red and blue galaxies or star-forming and quiescent galaxies.

sSFR: specific star-formation rate

MZR: mass-metallicity relation

Recent deep surveys have shown that these two populations (star-forming and quiescent) can be clearly identified at least up to $z \sim 2$, and perhaps up to higher redshifts of $z \sim 3$–4 (Brammer et al. 2011, Muzzin et al. 2013). Intriguingly, the comoving number and mass density of quiescent galaxies have been increasing over time since $z \sim 2$, whereas the number and mass density of star-forming galaxies have stayed roughly constant or decreased during this same interval (Bell et al. 2004, 2007; Faber et al. 2007; Brammer et al. 2011; Muzzin et al. 2013). Given that it is the star-forming population that is expected to be growing more massive due to the birth of new stars, this constancy or decrease has profound and unexpected implications—it implies that more and more star-forming galaxies must be having their SF extinguished or quenched as cosmic time progresses.

1.1.2. Global properties: scaling relations. Galaxies show many correlations among their global properties. We refer to such a correlation as a scaling relation when the conditional value of galaxy property y for a fixed value of another property x has a relatively small scatter. Stellar mass is often used as the x variable in galaxy scaling relations. Certain well-known examples of global scaling relations with $m_{\rm star}$ are the SFR for star-forming galaxies, sometimes known as the star-forming main sequence (Noeske et al. 2007, Wuyts et al. 2011), the mean fraction of cold gas ($f_{\rm gas} \equiv m_{\rm gas}/m_{\rm star}$) in the interstellar medium (ISM; Baldry et al. 2008, Peeples & Shankar 2011), and the metallicity of stars or ISM gas (mass-metallicity relation, MZR; Tremonti et al. 2004, Gallazzi et al. 2005, Zahid et al. 2013). Furthermore, some of the tightest known scaling relations in astronomy are those between galaxy properties and the mass of the SMBH they harbor (for a comprehensive review see Kormendy & Ho 2013).

Deep multiwavelength surveys have provided constraints on the evolution of these scaling relations. The normalization of the star-forming main sequence has declined by a factor of ∼20 since $z \sim 2$ (Speagle et al. 2014 and references therein), and a fairly tight sequence appears to be in place up to $z \sim 6$ (Salmon et al. 2014, Steinhardt et al. 2014). The MZR has evolved in the sense that galaxies of a given mass had lower gas-phase metallicities at high redshift (Savaglio et al. 2005, Erb et al. 2006, Zahid et al. 2013, Steidel et al. 2014, Wuyts et al. 2014). There is evidence from measurements of CO (an indirect tracer of molecular hydrogen, H_2) in fairly massive high-redshift galaxies that the gas fraction of galaxies has decreased significantly over cosmic time since $z \sim 2$ (Bothwell et al. 2013b; Saintonge et al. 2013; Tacconi et al. 2010, 2013; Genzel et al. 2014). Indirect estimates of cold gas fractions from inverted SF densities, assuming a fixed relationship between SF density and cold gas density, also indicate a rapid decrease in cold gas fraction from $z \sim 2$ to the present (Erb et al. 2006, Popping et al. 2012; G. Popping, P.S. Behroozi, M.S. Peeples, in preparation).

Some scaling relations show clear second-parameter dependences, in the sense that the scatter about a given relation is correlated with some other galaxy property. For instance, the MZR may show a second-parameter dependence on SF, in the sense that more rapidly star-forming galaxies at a given mass have lower metallicities (Lara-López et al. 2010, Mannucci et al. 2010). The cold gas content shows a similar correlation, with high-HI-mass galaxies having lower metallicities (Bothwell et al. 2013a, Lara-López et al. 2013).

1.1.3. Demographics: correlations with galaxy type. Since the original discovery of fuzzy nebulae, it has been known that galaxies come in different morphological types (Hubble 1926). There are many different methods for quantifying and classifying galaxy morphology, and this subject is reviewed by Conselice (2014); see also the discussion by Buta (2013) focused on the nearby Universe. Although galaxy morphology encompasses many complex facets of galaxy structure, including the presence of bulges, thin and thick disks, bars, spiral arms, etc., for the purposes of this review, we focus on a single simplified metric: the fraction of a galaxy's light or mass contributed by a flattened, rotationally supported disk, and that contained in an oblate or triaxial, pressure supported bulge or spheroid (often denoted by the bulge-to-disk ratio B/D or bulge-to-total ratio B/T). The bulge-to-disk ratio is broadly correlated with classical Hubble type (Simien & de Vaucouleurs 1986). We further simplify much of our discussion by referring to just two classes of galaxies, disk-dominated and spheroid-dominated. (In this review, we use the term spheroid to mean galaxies or galaxy components that structurally and kinematically resemble classic giant ellipticals, not to be confused with dwarf spheroidals or spheroidal galaxies. See Section 1.1.4 for further explanation.) Unfortunately, there is no standard value for the critical value of B/T used to divide these populations; values from $0.3 < (B/T)_{\rm crit} < 0.7$ are used in the literature. As it is difficult to robustly decompose the light of observed galaxies into a spheroid and disk component, other metrics such as the concentration (the ratio of the radius containing 90% of the light to the radius containing 50% of the light) or the Sérsic index (n_s; another measure of the slope of the light profile; e.g., Blanton & Moustakas 2009) are frequently used as rough proxies for morphology. We deliberately avoid using the terms early type and late type as they are used sometimes to refer to galaxy classes divided by morphology and sometimes to those divided according to their stellar populations (star forming versus quiescent).

***B/T*:** bulge-to-total ratio

Regardless of how galaxies are classified, there are robust demographic trends for disk-dominated versus spheroid-dominated galaxies. There is a very strong trend between morphology and color or SF activity, such that disk-dominated galaxies are predominantly blue and star forming, whereas spheroid-dominated galaxies are largely red and quiescent, with nearly uniformly old stellar populations (e.g., Roberts & Haynes 1994, Kauffmann et al. 2003, Blanton & Moustakas 2009). This trend appears to hold up to $z \sim 2$, with the caveat that red optical color becomes a less robust tracer of old stellar populations, as many star-forming galaxies at high redshift are reddened by dust. The characteristic Schechter function mass $M_{\rm char}$ is larger for quiescent or spheroid-dominated galaxies, and the slope is much shallower (e.g., Bernardi et al. 2010). Put another way, the fraction of spheroid-dominated galaxies increases strongly with stellar mass and luminosity. Furthermore, as emphasized by Binggeli et al. (1988), different types of galaxies can have luminosity functions that deviate considerably from the Schechter form.

A number of studies have shown that the probability for a galaxy to be quiescent depends on both its stellar mass and large-scale environment or halo mass (Balogh et al. 2004, Hogg et al. 2004, Peng et al. 2010, Woo et al. 2013). Recent work has shown that the correlation between quiescence and other internal properties, such as spheroid fraction, velocity dispersion, and central density, is even stronger than that with stellar mass (Bell et al. 2012, Cheung et al. 2012, Bluck et al. 2014, Lang et al. 2014).

1.1.4. Structural scaling relations. Both disks and spheroids exhibit correlations between their stellar mass or luminosity, their radial size, and their internal velocity (Faber & Jackson 1976, Kormendy 1977, Tully & Fisher 1977, Shen et al. 2003, Courteau et al. 2007, Bernardi et al. 2010). For disk-dominated galaxies, the radial size is usually characterized by the scale radius r_s [the scale radius in the exponential function characterizing the radial light profile; e.g., Mo et al. (2010), their equation 2.29 on p. 50], and the characteristic velocity is the rotation velocity at the

GR: General Relativity

maximum of the rotation curve $V_{\rm rot}$, which usually occurs at around $2r_s$. For spheroid-dominated galaxies, the radial size is characterized by the half-light radius or effective radius r_e (the radius that contains half of the total luminosity), and the internal velocity is characterized by the (line-of-sight) velocity dispersion σ. Several of these relationships have names, such as the Tully-Fisher ($L - V_{\rm rot}$; Tully & Fisher 1977) relation for disks and the Faber-Jackson ($L - \sigma$; Faber & Jackson 1976), and Kormendy ($L - r_e$; Kormendy 1977) relations for spheroids. A combination of these three quantities forms a fundamental plane; i.e., galaxies populate a relatively thin plane in *L*-*r*-*V* space or rescaled versions of these variables (Djorgovski & Davis 1987, Faber et al. 1987, Bender et al. 1992, Burstein et al. 1997). The familiar named bivariate relations are projections of this plane.

The slope, scatter, and evolution of these structural scaling relationships for spheroids and disks carry important clues about the formation history and relationships between these objects. For example, (*a*) the size-mass relationship is considerably steeper for spheroids than for disks at all redshifts (Shen et al. 2003, Bernardi et al. 2010, van der Wel et al. 2014); (*b*) since $z \sim 2$, the size-mass relation for spheroids has evolved much more rapidly than that for disks (Trujillo et al. 2006, van der Wel et al. 2014); (*c*) the size distribution at fixed mass is narrower for spheroids than for disks (van der Wel et al. 2014); and (*d*) the evolution of the Tully-Fisher and Faber-Jackson relations has been relatively mild (Kassin et al. 2007; Cappellari et al. 2009; Cenarro & Trujillo 2009; Miller et al. 2011, 2012). We note that many high-redshift studies present the scaling relations for galaxies divided according to whether they are star forming or quiescent, rather than spheroid or disk dominated, but this seems to make little difference to the qualitative results (van der Wel et al. 2014).

Another illustration of the importance of structural-kinematic scaling relations is demonstrated by the distinction between classical bulges and pseudobulges (Kormendy & Kennicutt 2004, Kormendy 2013). Classical bulges have centrally concentrated light profiles with Sérsic indices $n_s \sim 2$–3 (where a de Vaucouleur profile has $n_s = 4$) and lie on an extension of the fundamental plane for giant ellipticals. Pseudobulges have more extended light profiles that are similar to those of disks ($n_s \sim 1$) and lie on a different fundamental plane from classical bulges and giant ellipticals (Kormendy et al. 2009). Furthermore, classical bulges and pseudobulges have different correlations with SMBH mass (Kormendy & Ho 2013). Similarly, the dwarf galaxies that are confusingly termed dwarf spheroidals and dwarf ellipticals obey very different fundamental plane relations than do classical bulges and ellipticals of all luminosities (Kormendy et al. 2009, Kormendy & Bender 2012). In fact, dwarf spheroidals and dwarf ellipticals are indistinguishable from dwarf irregulars in their structural parameter correlations. These diverse scaling relations hint at different formation mechanisms for these objects, as reviewed by Kormendy (2013).

1.2. Cosmological Background

Our modern theory of cosmology is based on the ansatz that the Universe is homogeneous and isotropic on large scales (the cosmological principle) and on Einstein's theory of General Relativity (GR), which says that the structure of space-time is determined by the mass and energy content of the Universe. Together these allow us to derive equations that describe the evolution of the scale factor (or characteristic size and density) of the Universe in terms of the parameters specifying the mass and energy density. Observations have shown that the Universe started from a much denser, hotter, and nearly homogeneous state and has been expanding for approximately the past thirteen and a half billion years (e.g., Mo et al. 2010).

In this standard picture, quantum fluctuations in the very early Universe were processed during a period of very rapid expansion called inflation to create the small inhomogeneities that are detected via temperature fluctuations in the cosmic microwave background. These tiny fluctuations,

viewed at the time when free electrons combined with nuclei to form neutral atoms at a redshift of $z \simeq 1{,}100$, have now been studied in exquisite detail with many experiments, including the *Wilkinson Microwave Anisotropy Probe* and *Planck* satellites. When combined with other observations, such as the distance-redshift relation from Type Ia SNe, abundances of galaxy clusters, constraints on the present-day expansion rate (Hubble parameter H_0) from nearby Cepheid stars, and galaxy clustering (e.g., baryon acoustic oscillations), these measurements yield stringent constraints on the fundamental cosmological parameters (Hinshaw et al. 2013, Planck Collab. et al. 2014).

ΛCDM: cold dark matter with a cosmological constant (Λ)

These combined observations point to a Universe that is geometrically flat and dominated by dark matter and dark energy, which together account for more than 95% of the energy density of the Universe. The physical nature of both of these mysterious substances is unknown, although there are numerous theories. In the most popular variant of the standard model, which we refer to as ΛCDM, the dark matter (DM) is cold and collisionless and makes up ∼25% of the cosmic mass-energy density, and the dark energy is in the form of a cosmological constant Λ (as expected in the most general form of Einstein's equations of GR), comprising ∼70%. The remaining ∼5% is baryons (which in this context includes leptons), i.e., normal atoms that make up stars, gas, and heavy elements (metals). Although these cosmological parameters are still uncertain by up to perhaps 10%, for the purposes of understanding how galaxies form and evolve, this level of uncertainty is largely irrelevant.

With the initial conditions specified, if we neglect baryonic physics, it is relatively straightforward to compute how the density field of the dominant DM component evolves as the Universe expands. If we imagine the matter density field as a mountain range, the landscape in the CDM picture is extremely craggy, with many small-scale peaks superimposed on top of the medium- and large-scale peaks and valleys. As the Universe expands, the background density decreases. When a peak exceeds a critical overdensity relative to the background, the region within that peak stops expanding and becomes gravitationally self-bound. Numerical *N*-body techniques have been used to extensively study and characterize the growth of structure in dissipationless (DM only) ΛCDM simulations, as we discuss in Section 2.1.1. The gravitationally bound structures that form in these simulations are commonly referred to as DM halos, and the abundance, internal structure, shape, clustering, and angular momentum of these halos over cosmic time has been thoroughly quantified (see Mo et al. 2010, chapters 6 and 7 and references therein). Based on these DM only simulations, the standard ΛCDM paradigm has been extremely successful at explaining and reproducing observations on scales larger than a few kiloparsecs (e.g., Primack 2005), thereby providing a robust framework upon which to build models of galaxy formation and evolution.

1.3. Overview of Physical Processes

In this section, we briefly overview the main physical processes that are commonly included in current models of galaxy formation. We discuss these processes and their implementation in more detail in Sections 2 and 3.

- **Gravity**. Gravity plays a crucial role in building the skeleton for galaxy formation. The shape and amplitude of the primordial power spectrum of density fluctuations depend on the cosmological parameters and the properties of DM. This spectrum, processed by gravity, determines the number of DM halos of a given mass that have collapsed at any given time and how quickly these halos grow over cosmic time via merging and accretion. It also determines how DM halos cluster in space. In the standard paradigm, every galaxy is born within one of these dark halos. When halos merge, each containing its own central galaxy, gravity and dynamical friction gradually cause the orbits to decay until the galaxies merge. Mergers can have important effects on galaxies, including triggering bursts of

star formation and accretion onto central SMBHs, and transforming galaxy structure and morphology.

GMC: giant molecular cloud

- **Hydrodynamics and thermal evolution**. When an overdense region composed of gas and DM collapses, strong shocks form, increasing the entropy of the gas. The subsequent evolution of the gas is then determined by how efficiently the gas can cool and radiate away its thermal energy. The primary cooling processes relevant for galaxy formation over most of cosmic history are two-body radiative processes. Gas that is hotter than $T \geq 10^7$ K is fully collisionally ionized and cools predominantly via bremsstrahlung (free-free emission). In the temperature range of $10^4 < T < 10^7$ K, collisionally ionized atoms can decay to their ground state, and electrons can recombine with ions. Below temperatures of 10^4 K, cooling occurs through collisional excitation/de-excitation of heavy elements (metal-line cooling) and molecular cooling.

  Following collapse and shock heating, if radiative cooling is inefficient, a pressure-supported quasi-hydrostatic gaseous halo may form. This gas then gradually cools in what is often referred to as a cooling flow. This is also sometimes referred to as hot mode accretion. Once the gas cools and loses pressure support, it collapses until it is supported by its angular momentum. If the cooling time of the gas is short compared with the dynamical time, the gas may accrete directly onto the protogalaxy without ever forming a hot quasi-hydrostatic halo (White & Frenk 1991, Birnboim & Dekel 2003). Cosmological hydrodynamic simulations have shown that this sort of cold mode accretion tends to occur when gas flows in along relatively cold, dense filaments (Kereš et al. 2005).
- **Star formation**. Once gas has collapsed into the central regions of the halo, it may become self-gravitating, i.e., dominated by its own gravity rather than that of the DM. Because gas cools more rapidly the higher its density, if cooling dominates over heating, then a runaway process can ensue whereby giant molecular cloud (GMC) complexes form, and eventually some dense cloud cores within these complexes collapse and reach the extreme densities necessary to ignite nuclear fusion. However, many details of this process remain poorly understood. Moreover, most cosmological simulations are not able to resolve even the scales on which GMCs form, much less individual cores. Therefore all existing cosmological simulations implement empirical subgrid recipes to model SF.
- **Black hole formation and growth.** The first seed BH may have formed in the early Universe by a variety of mechanisms. They may be the remnants of Population III (metal free) stars, or they may form via direct collapse of very low angular momentum gas, or via stellar dynamical processes (Volonteri 2010). These seed BHs may grow by either accreting gas that has negligible angular momentum or an accretion disk that drains the gas of angular momentum via viscosity (Netzer 2013). These processes are, again, poorly understood and virtually impossible to model explicitly in cosmological simulations, so they are modeled via subgrid recipes.
- **Star-formation feedback**. Observations show that less than 10% of the global baryon budget today is in the form of stars. However, in CDM models without some sort of feedback (or suppression of cooling and SF), we would expect most of the gas to have cooled and formed stars by the present day. Even the pioneers of the earliest models of galaxy formation within a CDM framework recognized this overcooling problem and suggested that energy generated by SN explosions could heat gas and perhaps blow it out of galaxies, making SF inefficient (White & Rees 1978, Dekel & Silk 1986, White & Frenk 1991). It is now recognized that there are many processes associated with massive stars and SNe (e.g., photoheating, photoionization, winds) that could contribute to making SF inefficient and driving large-scale winds that reduce the baryon fractions in galaxies (for an overview see Hopkins et al.

2012b). Once again, most cosmological simulations cannot resolve these physical processes in detail, so nearly all current models implement subgrid recipes to attempt to capture their effect on galaxy scales.

- **Active galactic nucleus feedback**. There is strong observational evidence that most or perhaps all spheroid-dominated galaxies (which comprise the majority of all massive galaxies) contain an SMBH (for a recent review see Kormendy & Ho 2013). A simple calculation indicates that the amount of energy that must have been released in growing these BHs must exceed the binding energy of the host galaxy, suggesting that it could have a very significant effect on galaxy formation (Silk & Rees 1998); however, it is still uncertain how efficiently this energy can couple to the gas in and around galaxies. Observational signatures of feedback associated with active galactic nuclei (AGNs) include high-velocity winds, which may be ejecting the cold ISM from galaxies, and hot bubbles apparently generated by giant radio jets, which may be heating the hot halo gas (for recent reviews see Fabian 2012, Heckman & Best 2014). AGN feedback is also treated using subgrid recipes in current cosmological simulations.
- **Stellar populations and chemical evolution**. To make direct comparisons between models and observations, many modelers convolve their predicted SF histories with simple stellar population models, which provide the UV–near-IR SEDs for stellar populations of a single age and metallicity (Conroy 2013), folding in an assumed stellar initial mass function (IMF). [Note that all cosmological simulations to date, as far as we are aware, have assumed that the IMF is universal. However, there is mounting evidence that this assumption may not be valid (see the recent review by Bastian et al. 2010), which could have important implications for galaxy formation.] Many models now include the important contribution of gas recycling from stellar mass loss self-consistently within simulations (Leitner & Kravtsov 2011). In addition, as stars evolve and go SN, they produce and distribute heavy elements throughout the gas that surrounds galaxies, evidently polluting the intergalactic medium (IGM) out to fairly large distances from galaxies. Chemical evolution is a critical part of galaxy formation for several reasons: (*a*) Cooling rates at intermediate temperatures are highly enhanced in metal-enriched gas; (*b*) the luminosity and color of stellar populations of a given age are sensitive to metallicity; and (*c*) heavy elements produce dust, which dims and reddens galaxies in the UV and optical bands and reradiates the absorbed energy in the mid- to far-IR. Most cosmological models of galaxy formation now include a treatment of chemical evolution.
- **Radiative transfer**. The radiation emitted by stars and AGNs can have an important impact on galaxy formation. Radiation can directly heat gas and can also modify cooling rates (especially for metal-enriched gas) by changing the ionization state of the gas. Moreover, the transmission of radiation of different wavelengths through scattering by dust can greatly impact the measured total luminosity, color, and observationally determined morphological and structural properties of galaxies, especially in the rest-frame UV and optical, which are often all that is available at high redshift. Most current cosmological simulations that are run to low redshift ($z \leq 6$) do not include radiative transfer self-consistently owing to the added computational expense. However, with sufficiently high resolution, radiative transfer through a dusty ISM can be computed in postprocessing to estimate the observed panchromatic properties of galaxies (e.g., Jonsson et al. 2010) and their line emissions (e.g., Narayanan et al. 2008).

IMF: initial mass function

IGM: intergalactic medium

1.4. Overview of Basic Tools

Theorists have developed a wide range of different tools for modeling galaxy formation and evolution. Here, we briefly summarize the most commonly used tools and highlight some significant

differences between them. We provide a more detailed description of the methods used in the modeling tools in Section 2 of this review.

SHAM: subhalo abundance matching

SAM: semianalytic model

There is a popular class of models, including halo occupation distribution models (e.g., Berlind & Weinberg 2002, Zheng et al. 2005), conditional luminosity function models (Yang et al. 2003, van den Bosch et al. 2007), and subhalo abundance matching (SHAM) models and related techniques (e.g., Tasitsiomi et al. 2004, Conroy et al. 2006, Behroozi et al. 2010, Moster et al. 2010b). These techniques derive mappings between observable properties of galaxies and predicted properties of DM halos, and in general they contain no actual modeling of physical processes. Although this family of techniques is extremely useful for gaining insights into the required connection between observable galaxies and DM halos, we do not discuss these types of models in detail in this review.

The most explicit way to model galaxy formation is using numerical hydrodynamic techniques, in which the equations of gravity, hydrodynamics, and thermodynamics are concurrently solved for particles and/or grid cells representing DM, gas, and stars. The advantage of these techniques is that, within the limitations of the adopted numerical resolution, one obtains predictions of the density of each of these three components (as well as that of heavy elements) over cosmic time. One also obtains predictions for the velocities of the stars and DM, and the temperature of the gas. Thus the structure and kinematics of galaxies as well as their global properties and spatial distribution can be studied in great detail (see **Figures 1** and **2** for examples). The main limitation of these techniques is that computational exigencies restrict the dynamic range that can be explicitly simulated. This, combined with our still imperfect understanding of the physics that governs small-scale processes, such as SF, BH growth, and feedback, means that (as already discussed) many important processes must be treated using uncertain and somewhat arbitrary subgrid recipes. Moreover, computational limitations have historically made it difficult to experiment extensively with different subgrid recipes or to explore the multidimensional space of the variables that parameterize these recipes.

The other widely used technique for modeling galaxy formation in a cosmological context is semianalytic models (SAMs). This method does not explicitly solve fundamental equations for particles or grid cells but rather adopts a set of simplified flow equations for bulk components (see Baugh 2006, Benson 2010 for reviews). For example, a typical SAM tracks how much gas accretes into halos, how much hot gas cools and turns into stars, how feedback processes remove cold gas from the galaxy or heat the halo gas, and how mergers transform disks into spheroids, etc. **Figure 3** shows graphical representations of some of the quantities that can be tracked in an SAM for several example halo merger trees.

The computational requirements of these models are enormously reduced compared with fully numerical simulations. This makes it possible to make predictions for very large volumes or simulate galaxies over a larger range of halo mass, and also to extensively explore different subgrid recipes treating the most uncertain aspects of galaxy formation. Recently, several groups have coupled SAMs with a Bayesian inference approach and used Markov chain Monte Carlo techniques to sample the posterior probability distribution of the multidimensional space of the model parameters (Henriques et al. 2009, Lu et al. 2011). This is a powerful approach for exploring parameter degeneracies and obtaining more rigorous statistical assessments of the goodness of fit of specific models or model families with observational data (for an alternative approach using Bayesian emulator methods see Bower et al. 2010). As well, the less explicit nature of SAMs has allowed modelers to bypass some of the numerical issues that for many years caused difficulties in reproducing basic properties of galaxies in numerical simulations.

The field has now reached an interesting point where numerical simulations are able to reproduce fundamental observations at a similar level as SAMs. Interestingly, much of this success

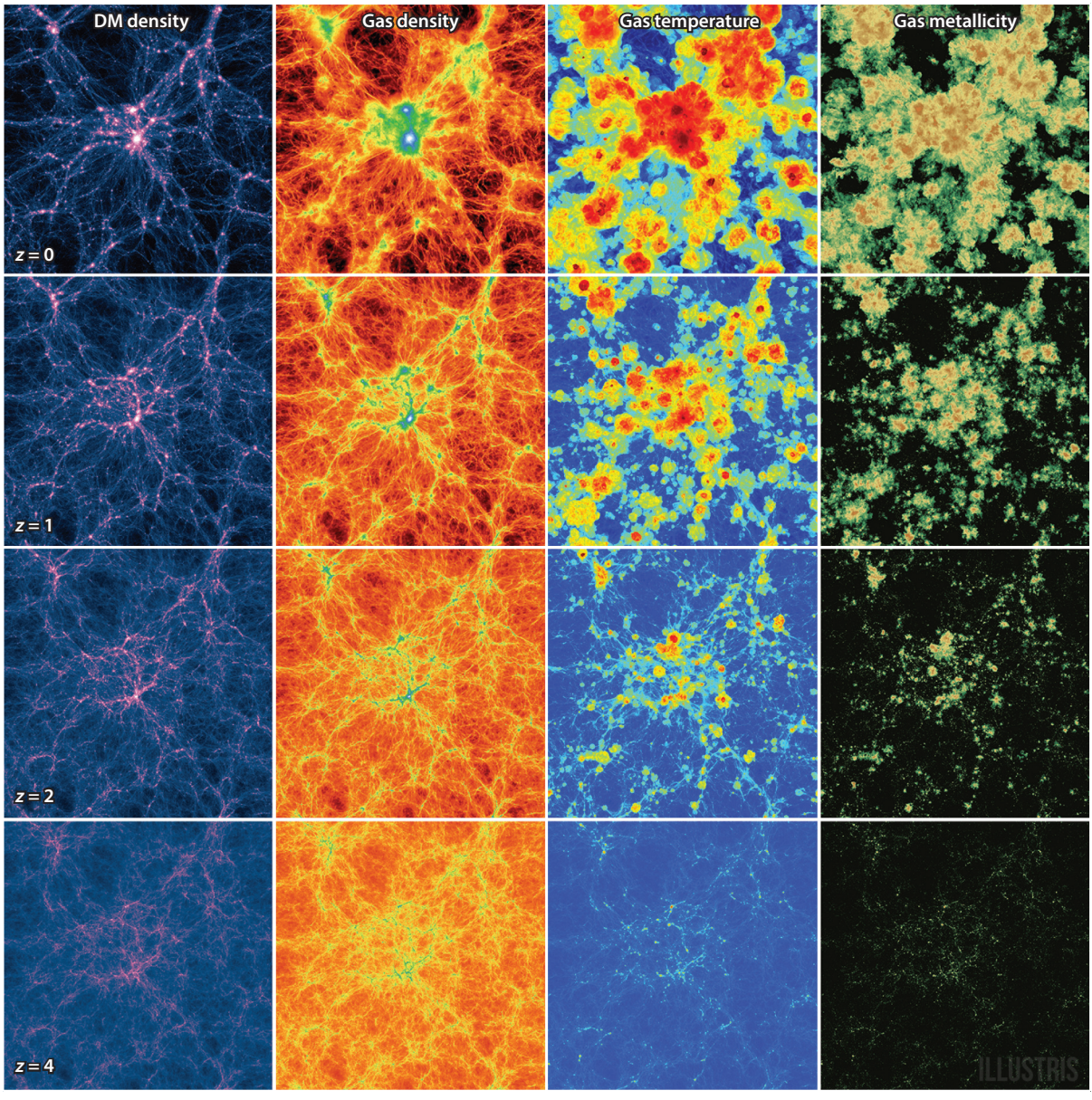

Figure 1

Visualization of representative quantities computed by numerical hydrodynamic simulations, from the Illustris project. From left to right, the dark matter (DM) density, gas density, gas temperature, and gas metallicity are shown at different cosmic times (from *top* to *bottom*: $z = 0, z = 1, z = 2, z = 4$). The slice shown has a projected thickness of 21.3 cMpc (comoving megaparsec) and shows the whole Illustris simulation box, which is 106.5 cMpc on a side. Reprinted from Vogelsberger et al. (2014a) with permission.

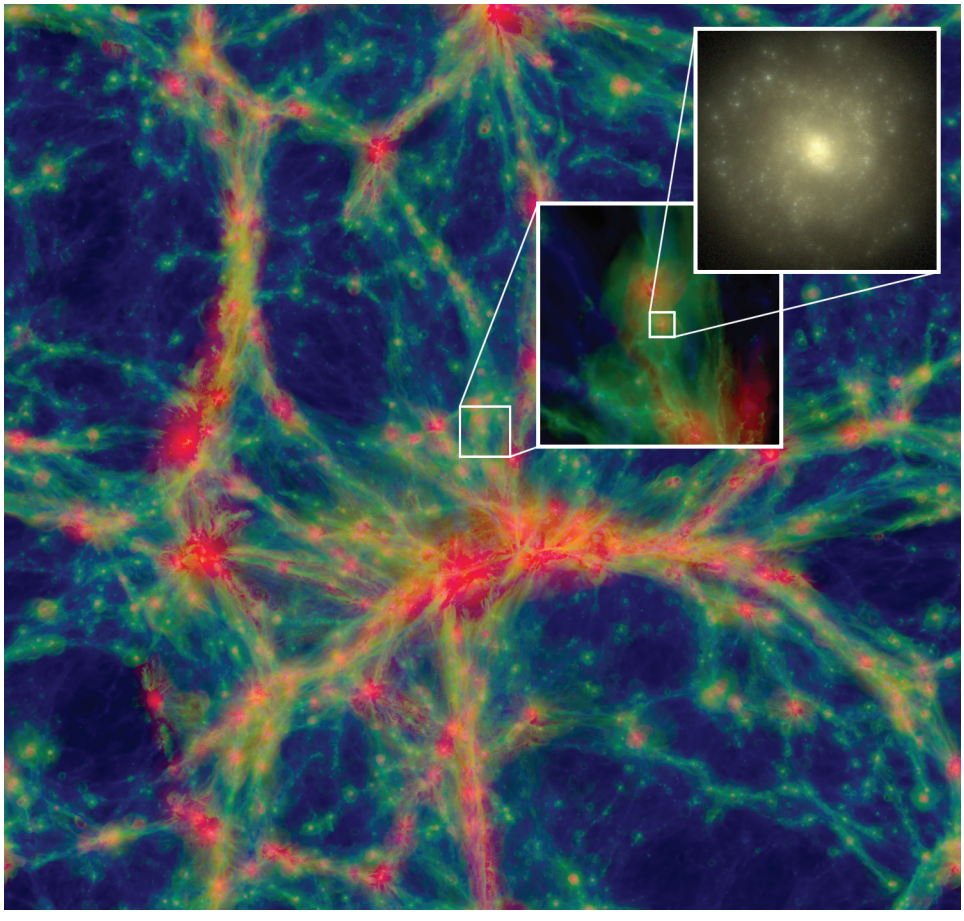

Figure 2

A 100 × 100 × 20 cMpc slice through the EAGLE simulation, illustrating the dynamic range that is attainable with state-of-the-art numerical hydrodynamic simulations. The intensity represents the gas density, whereas the color indicates the gas temperatures (*blue* through *green* through *red* indicate cooler to hotter). The inset shows a region 10 cMpc (comoving megaparsec) and 60 ckpc (comoving kiloparsec) on a side. The zoom into an individual galaxy with stellar mass 3×10^{10} $M_\odot$ shows the optical band stellar light. Reprinted from Schaye et al. (2015) with permission.

has been achieved by adopting a similar approach to the one that has long been used by SAMs, namely, (*a*) parameterizing the physical processes that can't be simulated explicitly and tuning these parameters to match a subset of observations, and (*b*) experimenting with different subgrid recipes to achieve the best match to a set of observations. Even the recipes themselves are often very similar to the ones that are commonly implemented in SAMs. Encouragingly, the two techniques have arrived at the same qualitative conclusions about galaxy formation and evolution for all the topics that we discuss in this review. For this reason, we structure this review largely in terms of the physical processes and the general insights into how they shape galaxy formation, giving examples from both SAMs and numerical simulations.

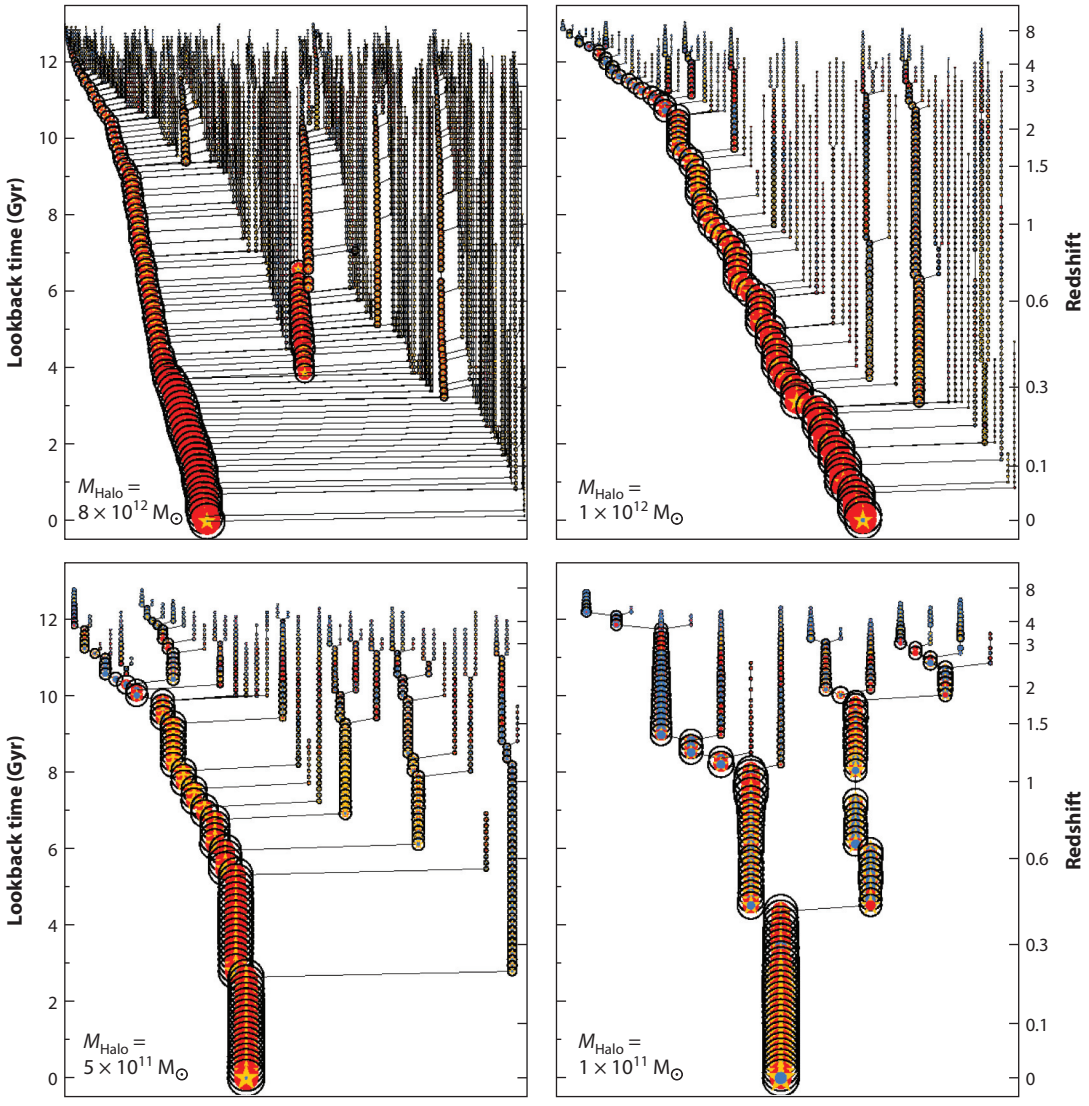

Figure 3

Visualization of representative predictions from a semianalytic model. Symbol sizes represent the mass of the host dark matter halo; the *x*-axis is arbitrary. Symbols connected by lines represent halo mergers. Colors represent the mass of different galaxy components (*red*, hot gas; *blue*, cold gas; *yellow*, stars). Several different final host halo masses are shown as indicated on the figure panels. Halos with the same virial mass can have a diversity of merger histories (not shown). Reprinted from Hirschmann et al. (2012a) with permission.

2. TOOLS AND METHODS

2.1. Gravity

Gravity is the driver of cosmological structure formation, and computing the evolution of clustering and collapse of matter under gravity forms the backbone for galaxy formation models, be they SAMs or hydrodynamic simulations. *N*-body gravity solvers can be used in combination with

hydrodynamic codes or can be used to compute DM halo merger trees, which form the basis for SAMs.

2.1.1. Numerical *N*-body methods. Gravity solvers, or *N*-body codes, must determine the force on each mass element from all others by solving Poisson's equation. Numerically solving Poisson's equation to evolve large-scale structure has a long and storied history that has been extensively reviewed elsewhere (e.g., Bertschinger 1998, Bagla 2005, Dehnen & Read 2011), so we greatly limit our discussion here.

The basic approach is to subdivide a representative portion of the Universe into many particles, compute the forces on these particles from all others, and evolve the system forward in discrete time-steps. In cosmological *N*-body simulations, the equations are solved within a comoving frame, and the volume is typically evolved with periodic boundaries, under the assumption that there is a space-filling set of identical volumes that approximately represent the larger-scale matter distribution. The expansion rate of the comoving frame is computed using the Friedmann equation (obtained from the Einstein equations within GR; see, e.g., Mo et al. 2010, ch. 3.2), but the equations actually solved are the familiar Newtonian versions because GR corrections are generally negligible.

N-body methods are particle based, mesh based, or a hybrid. In galaxy formation, the most popular particle-based approach is the tree code (Barnes & Hut 1986), in which the forces from distant groups of particles are approximated via their multipole moments. The particle-mesh (PM) method, in contrast, computes the potential on a grid via a Fourier transform of the density field and moves particles along potential gradients (Hockney & Eastwood 1988). Both scale with particle number N as $\mathcal{O}(\mathcal{N} \log \mathcal{N})$, though PM is considerably faster. Moreover, PM codes intrinsically account for all periodic replicas of the volume, whereas tree codes must use add-on techniques, such as Ewald summation (Hernquist et al. 1991).

The advantage of tree codes is that the forces on particles can be accurately represented down to the chosen force softening length ϵ, whereas PM codes are limited in resolution to their cell size. The ratio of the box length to ϵ defines the dynamic range of the calculation. A hybrid tree-PM approach is thus a popular method to increase dynamic range, in which the small-range forces are more accurately calculated using a tree, whereas large-range (and periodic) forces are computed via a faster PM method (e.g., GADGET-2; Springel 2005). The largest *N*-body simulations today evolve $\sim 10^{12}$ particles, with a dynamic range exceeding a million. With the advent of new computing technologies, such as graphics processing units, there is the potential for even larger computations if such highly threaded, cache-limited hardware can be effectively utilized; so far this has proved challenging, but progress is being made.

2.1.2. Dark matter halos and subhalos. A basic ansatz of our current picture of galaxy formation is that galaxies form within DM halos. Identifying these objects in *N*-body simulations is the first step in constructing the merger trees (see below) that form the gravitational backbone for SAMs. On-the-fly halo finding is carried out within some hydrocodes as well to use halo properties for subgrid recipes.

The halo mass (or virial mass) is usually defined as the mass within a sphere that encloses an average density $\Delta_{\rm vir}$ relative to the background density of the Universe. Similarly, the virial radius is defined as the radius within which the overdensity is equal to this critical value. The actual value of $\Delta_{\rm vir}$ is unfortunately not standardized and is based on a simple model of the collapse of a uniform spherical overdensity. In an Einstein-de Sitter Universe, after collapse and virialization, such a uniform spherical perturbation would have an average density $\simeq 178$ times that of the background (or critical) density (Mo et al. 2010, ch. 5.1). Many works use a fixed value of $\Delta_{\rm vir} = 200$, which

is just a rounding up of 178; some apply it relative to the critical density and some relative to the background density. Some works use redshift- and cosmology-dependent values of $\Delta_{\rm vir}$, as given by the fitting function from Bryan & Norman (1998). These different conventions introduce redshift-dependent differences, of as much as a factor of two in halo virial masses, radii, and internal velocities, that readers must be alert to when comparing results from the literature.

One of the generic features of the ΛCDM paradigm is that halos have a great deal of substructure. This substructure arises from objects that collapse and become bound at an earlier time, then get subsumed into a larger virialized structure. A subhalo is a halo that was once a distinct halo but is now contained within another virialized halo.

Methods used to identify halos in *N*-body simulations include friends-of-friends, spherical overdensity, and 6D phase-space-based methods. Knebe et al. (2011) present a comprehensive description and comparison of the results of different halo finders. Different halo finders tend to agree fairly well (within ∼10%) for basic halo properties, such as mass and peak circular velocity of distinct halos; the cumulative $z = 0$ halo mass function differs by $\pm 10\%$ across the 16 halo finders tested by Knebe et al. (2011). However, Klypin et al. (2011) point out that much larger differences between friends-of-friends and spherical overdensity-based finders can arise at high redshift. Identifying substructure is more halo-finder dependent; here 6D phase-space-based halo finders, such as ROCKSTAR (Behroozi et al. 2013b), were found to perform significantly better.

2.1.3. Merger trees. In semianalytic models, the formation of structure through gravitational instability in an expanding Universe is represented via merger trees. A merger tree records the masses of DM halos and the times at which these progenitor halos merge together to form a larger halo (see **Figure 3**). Merger trees may be either extracted from N-body simulations or constructed using semianalytic methods.

The first proposed methods for constructing merger trees semianalytically (Kauffmann et al. 1993, Cole et al. 1994, Somerville & Kolatt 1999) used statistical methods based on the extended Press-Schechter model (Lacey & Cole 1993). More recent methods apply empirical corrections to achieve better agreement with numerical simulations (e.g., Parkinson et al. 2008). These methods provide an important complement to merger trees extracted from *N*-body simulations, as they are extremely flexible, and can be used to efficiently explore different cosmologies and power spectra and large dynamic ranges in halo mass. Moreover, *N*-body-based merger trees have their own limitations, as discussed below.

Extracting merger trees from an *N*-body simulation appears straightforward on the face of it—one identifies DM halos at a series of redshifts or output times and then identifies which halos at earlier times are progenitors of a given halo identified at some later time. In practice, however, there are complications. (*a*) Results may be sensitive to the method used for identifying halos, as discussed above. (*b*) The definition of progenitor is not unique, because the particles from a halo at some time t_1 may end up in different halos at a later time t_2. (*c*) A halo may be identified as a subhalo in one time-step, then move outside of the virial radius of the host again at some later time. (*d*) Subhalos are tidally stripped as they orbit within their host halos and eventually become difficult to identify—most halo finders can no longer robustly identify subhalos when they drop below 30–40 particles (Knebe et al. 2011).

For semianalytic merger trees and to track substructure once the subhalo can no longer be identified in the *N*-body simulation, most SAMs include a procedure to estimate the time for a satellite's orbit to decay due to dynamical friction. A variation of the Chandrasekhar formula (Mo et al. 2010, their section 12.3.1) is generally used for this purpose. Many SAMs use refined versions of this formula based on numerical simulations (e.g., Boylan-Kolchin et al. 2008). Some subhalos may be tidally destroyed before they reach the center and their stars added to a diffuse stellar halo.

Satellites that survive until they reach the center are assumed to merge with the central galaxy. SAMs then implement a set of recipes for the effect of the merger, which generally includes an enhanced burst phase of SF as well as some sort of morphological transformation (e.g., moving stars from the disk to the spheroid component).

SPH: smoothed particle hydrodynamics

EC-SPH: entropy-conserving SPH

2.2. Hydrodynamics: Numerical Techniques

To directly model the visible component of the Universe requires modeling gas physics, i.e., solving the equations of hydrodynamics and evolving them concurrently with the chosen gravity solver. Doing so enormously increases the complexity of the code, resulting in longer calculations with greater intrinsic uncertainties. Most hydrocodes are based on solving the Euler equations (e.g., Mo et al. 2010, p. 366), representing mass, momentum, and energy conservation, typically closed by assuming a nonrelativistic ideal gas equation of state. The Euler equations are a form of the Navier-Stokes equations assuming no viscosity or conduction. In most cases, it is necessary to add an artificial viscosity term to properly handle convergent flows and shocks. Some experimentation has also been done with adding other physics whereby it is necessary to solve the Navier-Stokes equations directly; see Springel (2010b) for more discussion.

2.2.1. Lagrangian methods. In galaxy formation, historically the most popular Lagrangian method is smoothed particle hydrodynamics (SPH; see reviews by Monaghan 1992, Springel 2010b). Briefly, in SPH, the particles themselves carry the information about the fluid, which is obtained via a kernel-weighted sum over neighboring particles closer than a smoothing length (h):

$$X_i = \Sigma_j m_j (X_j/\rho_j) W(|\mathbf{r}_i - \mathbf{r}_j|, h_i, h_j). \quad (1)$$

Here, X_i is the quantity to be estimated, m and ρ are the particles' mass and density, and W is the kernel, which is some spherical function of the distance between particles in units of the smoothing length. X_i can also be a gradient of a quantity, in which case the gradient propagates through to the kernel, which becomes ∇W. The efficiency and simplicity of evaluating the Euler equations based on these local kernel-smoothed quantities give SPH many of its key advantages, including natural spatial adaptivity and trivial implementation in three dimensions.

Classic SPH (e.g., Hernquist & Katz 1989, Monaghan 1992) evaluates the density first as a kernel-smoothed average over nearby masses, then evaluates the thermal energy to update the pressure, and then evaluates the hydrodynamic acceleration. A variant of this method is used in the code GASOLINE (Wadsley et al. 2004). A key drawback is that this method does not explicitly conserve energy and entropy in adiabatic flows in the case of variable smoothing lengths. Entropy-conserving (EC)-SPH (Springel 2005) mitigated this flaw by explicitly including variational terms in h as derived from the Lagrangian and was formulated using entropy as the evolved variable. EC-SPH is employed in the widely used code GADGET-2 (Springel 2005). Subsequently, it was noted that a side effect of EC-SPH is the creation of an artificial pressure between cold and hot phases, resulting in a surface tension that causes, for example, cold clumps moving through a hot medium to be significantly more resistant to disruption than they are in grid-based codes (Agertz et al. 2007). Ironically, classic SPH performs somewhat better in this regard (at the cost of increased particle interpenetration), but all these versions fail to realistically capture Kelvin-Helmholtz instabilities.

Several promising approaches have been developed recently to mitigate these issues. Read & Hayfield (2012) proposed SPHS in which they showed that the error in the momentum equation in classic SPH can be reduced by using a different kernel shape with a larger number of neighbors, along with a modification of artificial viscosity to include a higher-order dissipation switch that

anticipates shocks. SPHS can yield much better results for a range of surface instability tests, but the required increase in the number of SPH neighbors (442 versus ~40) slows the calculation and lowers the resolution.

DI-SPH: density-independent SPH

A different approach was pursued by Saitoh & Makino (2013), following on Ritchie & Thomas (2001). They argued that difficulties in classic or EC-SPH arose from the requirement that the density distribution be differentiable, which is violated at contact discontinuities. They proposed a new formulation that was density-independent, DI-SPH, which used kernel sums to separately obtain the energy density and internal energy, from which the density is inferred. DI-SPH was shown to remove the artificial surface tension and enable improved treatment of surface instabilities (among other tests). Hopkins (2013) reformulated DI-SPH in terms of entropy to incorporate the improved conservation properties of EC-SPH. This pressure-entropy SPH provides much improved handling of surface instabilities versus classic SPH, with fewer numbers of neighbors than for SPHS. A modified treatment of artificial viscosity has also been widely implemented and helps improve the performance of SPH in this regard (e.g., Hu et al. 2014). Thanks to such improvements, current formulations of SPH can now track surface instabilities and associated phenomena to an accuracy that, not long ago, was widely regarded to be challenging for SPH.

2.2.2. Eulerian methods. A time-honored approach to solving hydrodynamics is to discretize the fluid onto grid cells and then compute the advection of properties across the cell boundaries. This is the basis of Eulerian approaches, which formulate the solution to the Euler equations in the fixed frame rather than in the fluid frame as in the Lagrangian approach.

Most current cosmological Eulerian hydrocodes employ a high-order Godunov scheme. Here, the Riemann problem is solved across cell faces, which yields a pressure at each cell face, thereby giving the force on the fluid across the cell. The fluid, with all its associated properties, is then advected across the cell face. If the cell is assumed to have uniform properties within it, this is called a (first-order) Godunov solver. Modern codes employ parabolic interpolation, known as the piecewise parabolic method. Note that while higher-order interpolation provides a more accurate solution, it requires using information from neighboring cells that effectively lowers the spatial resolution.

Given the dynamic range involved in modeling galaxy formation, a key development was the implementation of adaptive mesh refinement (AMR). Here, cells satisfying some local criteria (typically based on mass) are split into subcells, enabling improved resolution in those regions. This effectively achieves some of the Lagrangian methods' primary advantage of being naturally adaptive in space and time. Current AMR hydrocodes for galaxy formation include Enzo (Bryan et al. 2014), RAMSES (Teyssier 2010), FLASH (**http://flash.uchicago.edu/site/flashcode/**), and Hydro-Adaptive Refinement Tree (H-ART; Kravtsov et al. 1997).

2.2.3. Arbitrary Lagrangian-Eulerian methods. Optimally, one would like to unite the advantages of piecewise parabolic method in handling shocks and contact discontinuities with SPH's natural adaptivity. One approach is to use a deformable mesh, in which the mesh follows the fluid. Such arbitrary Lagrangian-Eulerian codes have historically not played a large role in astrophysics (see, e.g., Pen 1998), but this has recently changed with the introduction of AREPO (Springel 2010a).

AREPO uses a Voronoi tessellation to subdivide space around particles. A Voronoi tessellation is a space-filling set of polyhedral cells in which the space within a given cell is closer to one particle than any other. The Riemann problem is then solved across the cell faces to obtain the force on the particle. The mesh is regenerated as the particles move. In this way, AREPO is able to naturally

follow the fluid like a Lagrangian code while retaining the advantages of Godunov solvers, such as excellent handling of contact discontinuities, surface instabilities, and shocks, and the lack of artificial viscosity.

EoR: epoch of reionization

2.2.4. Advantages and disadvantages. Traditionally, Eulerian methods have enjoyed superiority in handling strong shocks and surface instabilities, whereas Lagrangian methods like SPH are more adaptive and provide increased dynamic range for a given computational expense. However, in recent times the gaps have been closing in both directions. AREPO has some important advantages over both Lagrangian and Eulerian methods, in particular EC-SPH (Vogelsberger et al. 2012).

An advantage of a particle-based approach such as SPH is that the movement of mass is directly tracked. This makes it more straightforward to follow the mass as it assembles into galaxies and to track where material ejected from galaxies ends up. It is also straightforward to implement kinetic winds, which, as we discuss below, has had substantial success as a subgrid prescription for galactic outflows. Nonetheless, in mesh codes it is possible to use tracer particles for these purposes. For instance, AREPO has implemented kinetic winds by spawning particles that are decoupled from the hydrocode mesh and then later rejoin.

AMR offers the key advantage that the mesh can be refined to arbitrarily high resolution, whereas particle-based methods are limited in resolution by the particle mass. This allows individual systems to be examined in great detail, albeit at great computational cost. For example, Enzo merger simulations by Kim et al. (2009) and H-ART cosmological zoom simulations by Ceverino et al. (2014) both achieved a dynamic range of $\sim 10^6$, whereas the most ambitious current SPH simulations can only achieve $\sim 10^5$.

More broadly, because all modern codes generally yield similar answers in basic tests relevant to galaxy formation in which the answer is approximately known, at this stage it is difficult to identify one code or methodology that is clearly superior to the others. For most properties, differences in subgrid prescriptions yield much larger variations than differences in hydrodynamical techniques.

2.3. Thermal Evolution

The key difference between baryons and DM in galaxy formation is that baryons can dissipate their potential energy via radiative processes. Radiative cooling and photoionization heating are thus implemented in essentially all codes, whereas radiation transport is a growing subfield with specific applications to the epoch of reionization (EoR) and line emission. In the next two subsections, we discuss how thermal evolution is modeled in numerical simulations and SAMs.

2.3.1. Cooling and heating in numerical simulations. Early simulations included radiative cooling only via primordial gas (H+He), but most simulations today also include cooling from metal-line emission, which dominates particularly at $10^5 \leq T \leq 10^7$ K for typical warm-hot gas metallicities. Early works employed cooling rates assuming collisional ionization equilibrium (Sutherland & Dopita 1993), but more recent work by Wiersma et al. (2009a) better accounts for the photoionization of metals by the metagalactic radiation field.

Simulations focusing on the post-EoR Universe typically account for photoionization heating by assuming all the gas is optically thin and in ionization equilibrium with a spatially uniform metagalactic radiation field (e.g., Faucher-Giguère et al. 2009, Haardt & Madau 2012). During the EoR, these assumptions break down, and continuum radiative transfer is necessary to properly model the feedback from photoionization heating on galaxy growth. Two approaches are used: applying radiative transfer in postprocessing to existing density distributions (Iliev et al. 2006), which is useful for evolving large volumes to study the final stages of EoR; and full radiative hydrodynamic codes that evolve the ionizing field together with the baryons, including modeling

SF to self-consistently predict the properties of the sources (Iliev et al. 2009, Finlator et al. 2011, Pawlik & Schaye 2011, Wise & Abel 2011). Given that this review focuses on the postreionization Universe, we do not discuss this further here.

2.3.2. Cooling and cosmological accretion in semianalytic models. Most SAMs implement some variant of the self-similar cooling flow model originally proposed by White & Frenk (1991) to track the thermal evolution of gas. As the gas enters the halo, it is assumed to be shock-heated to the virial temperature $T_{\rm vir} = 35.9[V_{\rm vir}/({\rm km\ s^{-1}})]^2$ K, in which $V_{\rm vir}$ is the halo virial velocity. One may then calculate the cooling time, which is the time it would take for the gas to radiate away all its energy:

$$t_{\rm cool} = \frac{\frac{3}{2}\mu m_{\rm p} kT}{\rho_{\rm g}(r)\Lambda(T, Z_{\rm h})}. \quad (2)$$

Here, $\mu m_{\rm p}$ is the mean molecular mass, T is the temperature of the gas, $\rho_{\rm g}(r)$ is the radial density profile of the gas, $\Lambda(T, Z_{\rm h})$ is the temperature and metallicity dependent cooling function (e.g., Sutherland & Dopita 1993), and $Z_{\rm h}$ is the metallicity of the hot halo gas.

The hot gas is assumed to be distributed with a smooth spherically symmetric density profile. Most models assume that the density profile is described by a singular isothermal sphere [$\rho_{\rm g}(r) \propto r^{-2}$], although some use different density profiles, such as a Navarro-Frenk-White profile (Navarro et al. 1997) or a cored Navarro-Frenk-White profile (Cole et al. 2000).

One can then solve for the cooling radius, within which gas has had time to dissipate all its thermal energy by cooling. To do this, one must adopt a timescale over which cooling is assumed to have taken place. Common choices for this timescale are the time since the halo has experienced a major (at least 2:1) merger (e.g., Somerville & Primack 1999) or the halo dynamical time $t_{\rm dyn} = r_{\rm vir}/V_{\rm vir}$ (e.g., Springel et al. 2001). It may happen that the model predicts $r_{\rm cool} > r_{\rm vir}$, indicating that the cooling time is shorter than the dynamical time, corresponding to the cold flow regime described in Section 1.3. In this case, most modelers generally assume that gas can flow into the halo within a dynamical time. Although this model is very simple, several studies have shown that the predicted cooling and accretion rates are in surprisingly good agreement with those from numerical hydrodynamic simulations (Benson et al. 2001, Yoshida et al. 2002, Hirschmann et al. 2012a, Monaco et al. 2014).

2.4. Chemical Evolution

Tracking the enrichment of gas with heavy elements is important for cooling calculations, and for predictions of galactic chemical evolution. Most numerical hydrodynamic simulations now include a model for chemical enrichment. Early models tracked only Type II SN enrichment, which is closely related to the oxygen abundance. To track other key elements, such as carbon and iron, it is necessary to model asymptotic giant branch (AGB) stars whose ejecta dominate the present-day carbon budget and Type Ia SN that produce most of the iron in stellar-dominated systems. Such delayed feedback sources are now included in most codes, which track a suite of individual elements (Oppenheimer & Davé 2008, Wiersma et al. 2009b). The dominant uncertainty typically comes from the metal yield models from SN and stellar evolution, particularly at low metallicities and high masses. Hence at present, absolute abundance predictions should be considered accurate to only a factor of two, but relative trends of metallicity versus other galaxy properties, such as stellar mass, are likely more robust.

Most SAMs use a simple instantaneous recycling approximation in which a yield y of heavy elements is produced by stars in each time-step: $\mathrm{d}M_Z = y\,\mathrm{d}m_{\rm star}$, where $\mathrm{d}M_Z$ is the mass of metals produced and $\mathrm{d}m_{\rm star}$ is the mass of stars formed. In general these metals are deposited into the

cold ISM, although some models deposit some of the metals directly in the hot halo gas. Metals may then be ejected from the cold gas reservoir by winds, and are deposited in either the IGM or the hot gas halo. Most SAMs treat the yield y as a free parameter rather than taking it from SN yield calculations and neglect enrichment by Type Ia SNe and AGB stars (so again, the predicted metallicities most closely trace α elements such as oxygen). However, some SAMs in recent years have incorporated more detailed treatments of chemical enrichment, tracking multiple individual elements, and the finite timescales for enrichment and gas recycling from AGB stars, Type Ia, and Type II SNe (Nagashima et al. 2005, Arrigoni et al. 2010, Yates et al. 2013).

2.5. Initial Conditions and Zoom Simulations

The generation of standard cosmological initial conditions involves (*a*) generating a linear matter power spectrum via a transfer function (e.g., Eisenstein & Hu 1999); (*b*) Gaussian-random sampling of the power spectrum for modes within the simulation volume; and (*c*) evolving the modes forward in the linear regime via the Zel'dovich approximation (see Bertschinger 1998 for more details). This generates particle positions and velocities sampling the matter field within a specified volume for a specified cosmology, at some specified high redshift that is optimally just before structure within the volume first goes nonlinear (see, e.g., MUSIC; Hahn & Abel 2011).

Zoom simulations are an increasingly popular and useful technique. In zooms, a subvolume within a cosmologically representative region is evolved at much higher resolution, together with surrounding regions of coarser resolution that provide the tidal field from large-scale structure. After an initial coarse-grained run, a halo or region of interest is selected, and its particles are tracked back to the original initial conditions to define the zoom region. Particles within the zoom region are sampled to finer resolution, including the requisite small-scale power, and the entire volume is run again, typically with hydrodynamics turned on only in the zoom region. In this way, zooms provide an increased dynamic range at a manageable computational cost, albeit only for a single galaxy or halo and its environs.

Simulations of idealized isolated galaxies, or mergers thereof, provide a valuable test bed to explore detailed physical processes, particularly in the ISM. Initial conditions are typically created in a stable disk configuration (Hernquist 1993a), and then dynamical perturbations grow from tides induced by either a merger or internal stochasticity. Such models can achieve extremely high resolution (by cosmological standards) and can serve to isolate physics of particular interest; hence they remain useful even if they do not fully represent the cosmological baryon cycle.

3. SUBGRID PHYSICS

3.1. Star Formation and the Interstellar Medium

A huge body of observations from UV through near-IR light traces the emission from stars. To make contact with these observations, models must attempt to compute how gas in galaxies turns into stars. The ISM is a complex place, with multiple gas phases coexisting at very different densities and temperatures (McKee & Ostriker 1977). Cosmological simulations of more than a single galaxy are still orders of magnitude away from capturing the spatial scales, temperatures, and densities where stars actually form. Moreover, physical processes that are not typically included or captured well in cosmological simulations, such as magnetic fields and turbulence, are thought to play important roles on the scales of dense molecular cloud cores and protostars (McKee & Ostriker 2007). However, advances in our theoretical understanding of SF as well as better observational characterization of key scaling relations (for a review, see Kennicutt & Evans 2012) have enabled the development of empirical recipes that smooth over much of the small-scale complexity.

Stars are observed to form in the dense, cold, molecular phase of the ISM, and current observations support a (nearly) universal SF efficiency in molecular gas, where about 1% of the gas is converted into stars per free-fall time (Bigiel et al. 2008, 2011; Krumholz et al. 2012; Leroy et al. 2013). Thus the ability to track where molecular gas forms should lead to a more physical approach to modeling SF. The ISM is observed to become H_2-dominated at ~1–100 atoms cm^{-3}. Because gravitational instability is thought to be one of the driving forces in the formation of GMCs (Dobbs et al. 2014), simply requiring a density threshold for SF of a few atoms per cubic centimeter may be a good first approximation. However, this also requires high-enough resolution (~100 pc) to attain these densities, which is currently achievable only in zoom simulations.

In more detail, H_2 formation is catalyzed by dust and destroyed by Lyman-Werner radiation, so one would expect that H_2 production is thus roughly proportional to metallicity, whereas destruction depends on the ability to self-shield against interstellar radiation. Some zoom simulations now include a simplified phenomenological treatment of chemical networks and H_2 dust- and self-shielding (Gnedin et al. 2009, Christensen et al. 2012). Fitting functions that attempt to capture the essence of H_2 formation and dissociation and the resulting dependence of H_2 fraction f_{H_2} on gas density, metallicity, and local UV background have been presented based on zoom simulations and on idealized (noncosmological) disk simulations and analytic models (Krumholz et al. 2009, McKee & Krumholz 2010, Gnedin & Kravtsov 2011, Gnedin & Draine 2014).

An alternative approach for partitioning gas into HI and H_2 is to use the empirical relationship between f_{H_2} and the disk midplane pressure, pointed out by Blitz & Rosolowsky (2004; BR). They found that the molecular fraction $R_{mol} \equiv \Sigma_{H_2}/\Sigma_{HI}$ was correlated with the disk hydrostatic midplane pressure P: $R_{mol} = (\frac{P}{P_0})^{\alpha_{BR}}$, where P_0 and α_{BR} are free parameters that are obtained from a fit to the observational data. The hydrostatic pressure as a function of radius in the disk can be estimated from the cold gas surface density, the stellar surface density, and the ratio of the vertical velocity dispersions of the gas and stars (Elmegreen 1989). This approach can be used to estimate f_{H_2} either self-consistently (see below) or in postprocessing in numerical simulations or SAMs (Obreschkow et al. 2009, Duffy et al. 2012).

3.1.1. Numerical implementation. The basic recipe for SF in many cosmological simulations has not changed markedly from the pioneering work of Katz (1992). Gas that is dense and converging is assigned an SFR following a Schmidt (1959) law, namely

$$\dot{\rho}_* = \frac{\varepsilon_* \rho_{gas}}{t_{ff}} \propto \rho_{gas}^{1.5}, \tag{3}$$

where the last proportionality arises because the local free-fall time $t_{ff} \propto \rho^{-0.5}$. The free parameter ε_* is typically calibrated to match the amplitude of the observed Kennicutt (1998) relation in simulations of idealized, isolated disks. Long-term SF histories are generally insensitive to the precise value of ε_* within reasonable variations (Katz et al. 1996, Schaye et al. 2010), because as discussed later, globally, SF is driven primarily by gas accretion and, over cosmological timescales, is not limited by the rate of conversion of gas into stars in the ISM. A somewhat different approach was proposed by Schaye & Dalla Vecchia (2008): they analytically recast the Kennicutt-Schmidt relation as a function of pressure rather than density, assuming a self-gravitating disk.

Stars are generally only allowed to form when the density exceeds some critical value, the choice of which is another free parameter. Springel & Hernquist (2003) incorporated a density threshold based on where the Jeans mass became lower than the particle mass, at which point a subgrid ISM model is required; this value turned out to be ≈0.1 cm^{-3} for typical mass resolutions adopted in cosmological volumes at the time.

This simple SF prescription applied to individual disk galaxies was found to quickly collapse gas down to the (artificial) Jeans scale in the simulations, which produced highly clumpy disks that looked nothing like local grand-design spirals. The solution, introduced in cosmological simulations by Springel & Hernquist (2003), was to artificially overpressurize the ISM by implementing a subgrid ISM model based on work by McKee & Ostriker (1977) that tracked the balance between SN energy input and cooling within a multiphase ISM. The temperature of the ISM gas (defined as gas above the SF threshold density) is then raised as high as 10^6 K. Robertson et al. (2004) extended this model to an arbitrary ISM effective equation of state and showed that, with appropriate overpressurization, this approach can reproduce smooth, stable, gas-rich spirals as observed today. Ironically, as we discuss further in Section 4.2.1, it turns out that simulations with no or minimal ISM pressure (Ceverino et al. 2010, 2014) do well at reproducing the clumpy disks that are now known to be common at high redshift ($z \sim 2$; Genzel et al. 2011), though simulations with pressurization can also reproduce these (Genel et al. 2012b).

It is clear that real stars do not form at densities of $\sim$0.1 atoms cm^{-3}. Moreover, because the Kennicutt relation is only observed to hold when the ISM is averaged over scales of 0.5–1 kpc, once simulations resolve smaller scales, it becomes dubious to use an SF prescription that is calibrated to match this relation. Thus much recent effort has been directed toward incorporating more realistic treatments of the ISM into cosmological simulations. High-resolution zoom simulations that adopt a higher SF threshold ($\sim$5 atoms cm^{-3}) and efficient stellar-driven winds (see Section 4.2.1 for further discussion) show marked improvement in their ability to produce realistic disks (e.g., Governato et al. 2007, Guedes et al. 2011). Other simulators (e.g., Kuhlen et al. 2012, Agertz & Kravtsov 2014) have incorporated subgrid recipes to compute the density of molecular hydrogen ρ_{H_2} and then use that result in an equation similar to Equation 3 in place of ρ_{gas}—no arbitrary density threshold need then be applied.

An exciting development is that cosmological zoom simulations are starting to be able to resolve the Jeans mass/length of gas, corresponding to the scale of molecular cloud complexes, allowing more direct modeling of the multiphase ISM (e.g., the FIRE simulations; Hopkins et al. 2014a). Concurrently, ISM simulations including detailed treatments of nonequilibrium chemistry and turbulence are pushing outward in scale to start bridging the gap with the cosmological runs (e.g., Walch et al. 2011, Mac Low & Glover 2012). Continuing interactions between the galaxy formation and ISM/SF communities should soon allow us to place our subgrid recipes on a more secure physical foundation.

3.1.2. Implementation in semianalytic models. The usual approach to modeling SF in SAMs is very similar to the approach used in numerical hydrodynamic simulations, described above. Gas that has cooled according to the cooling model described in Section 2.3.2 loses its pressure support and collapses further, until it is supported by its angular momentum, forming a disk. The initial angular momentum of the halo gas can then be used to estimate the radial size of the disk, as described in Section 4.2.4. Some SAMs track the radial structure of the disk in cylindrically symmetric annuli (Kauffmann 1996a, Avila-Reese et al. 1998, Dutton & van den Bosch 2009, Fu et al. 2010), whereas most models assume the disk radial surface density distribution to be an exponential, as is generally the case in observed disk galaxies.

Different SAMs use different but roughly physically equivalent variants of Equation 3. Early SAMs typically used an expression of the form

$$\dot{m}_* = \varepsilon_* \frac{m_{\rm cold}}{\tau_*},$$

where $\dot{m}_*$ is the total SFR in the galaxy, $m_{\rm cold}$ is the total cold gas mass in the galaxy, τ_* is a characteristic timescale for SF, and ε_* is a parameter representing the global SF efficiency (SFE). The SF timescale is often assumed to scale with the dynamical time of the DM halo, $\tau_* \propto \tau_{\rm dyn} \propto r_{\rm H}/V_{\rm H}$, where $r_{\rm H}$ is the characteristic halo radius and $V_{\rm H}$ is the characteristic halo circular velocity. However, it was quickly realized that an SF law of this form could not reproduce the observed trend of increasing cold gas fractions with decreasing stellar mass in the low-redshift Universe. Thus, modelers either introduced an SF threshold, such that only the fraction of the cold gas above this threshold was eligible to participate in SF, or made τ_* an explicit function of halo properties, e.g., of $V_{\rm H}$ (Cole et al. 2000), such that the SF timescale is made longer in lower mass galaxies.

SFE: star-formation efficiency

Models that track disk structure in more detail are able to use empirical laws that are closer to what is actually observed. For example, Somerville et al. (2008b) adopted a Kennicutt-like expression, in which the SFR surface density of the disk is calculated according to $\Sigma_{\rm SFR} = A_{\rm SF}\Sigma_{\rm gas}^{N_{\rm SF}}$ for $\Sigma_{\rm gas} > \Sigma_{\rm crit}$ (and zero otherwise). The parameters $A_{\rm SF}$ and $N_{\rm SF}$ are taken directly from observations (e.g., Kennicutt 1998), and $\Sigma_{\rm crit}$ is treated as a free parameter. A similar approach, but with a radius and circular velocity–dependent $\Sigma_{\rm crit}$ based on the Toomre condition for gravitational instability, is adopted in the SAMs developed by the group at the Max-Planck-Institut für Astrophysik (MPA; e.g., Kauffmann et al. 1999, Croton et al. 2006, Guo et al. 2011).

Several groups have recently developed SAMs that attempt to track atomic and molecular gas separately (Fu et al. 2010, Lagos et al. 2011b, Popping et al. 2014b, Somerville et al. 2015). Various recipes for H_2 formation, employing either the metallicity-based fitting functions of Krumholz et al. (2009) and Gnedin & Kravtsov (2011) or alternatively the empirical pressure-based recipe of Blitz & Rosolowsky (2004), have been implemented in these SAMs. Again, the computed density of H_2 may then be used in an empirically calibrated SF law with no need to assume a density threshold, essentially removing all free parameters from the SF recipe (within the observational uncertainties on the slope and normalization of the relationship between $\Sigma_{\rm SFR}$ and $\Sigma_{\rm H_2}$). Overall, it appears that the main predictions of SAMs, especially for stellar properties of galaxies, are quite insensitive to the details of the gas partitioning recipe (Fu et al. 2010, Lagos et al. 2011b, Berry et al. 2014, Popping et al. 2014b, Somerville et al. 2015).

It is well known that galaxy interactions and mergers can trigger starbursts with enhanced SFE, and most SAMs implement a burst mode of SF in galaxies that have experienced a recent merger. Studies based on hydrodynamic simulations of binary galaxy mergers have shown that the enhancement in the SFE above that in an isolated galaxy is a fairly strong function of the mass ratio of the merger. Many SAMs implement the fitting function introduced by Cox et al. (2008), who parameterized the burst efficiency as $e_{\rm burst} = e_{\rm burst,0}\mu^{\gamma}$, where μ is the merger mass ratio, and $e_{\rm burst}$ is defined as the fraction of the total gas reservoir that is consumed in the burst.

Subsequent studies have shown that $e_{\rm burst}$ and the burst timescale also depend on the implementation of stellar feedback and the treatment of the ISM (Robertson et al. 2006b, Cox et al. 2008). Hopkins et al. (2009a) found that the burst efficiency depends strongly on the cold gas fraction in the progenitors, with lower burst efficiencies found in mergers with higher progenitor gas fractions. However, Moster et al. (2011) did not find a strong correlation with the progenitor cold gas fraction when they included a hot halo in the merger progenitors. Although there have been numerous studies of SF enhancement in mergers using numerical hydrodynamic simulations of binary mergers (e.g., Mihos & Hernquist 1996; Springel 2000; Cox et al. 2006, 2008), these simulations are not in a cosmological context and therefore must assume idealized initial conditions. Furthermore, most have not included cosmological accretion or cooling from a hot gas halo. To our knowledge, there has not been a systematic exploration of the enhancement of SF activity in mergers using cosmologically situated hydrodynamic simulations.

SAMs predict that burst-mode SF makes a relatively minor contribution to the overall global SFR density at any epoch (e.g., Baugh et al. 2005, Somerville et al. 2008b), in agreement with observations (Rodighiero et al. 2011, Schreiber et al. 2014) and cosmological hydrodynamic simulations (Murali et al. 2002, Kereš et al. 2005). However, merger-triggered bursts may be important for producing certain populations, such as ultraluminous IR galaxies and high-redshift submillimeter-detected galaxies (Niemi et al. 2012, Hayward et al. 2013), in agreement with observations that suggest a strong connection between major mergers and starbursts (e.g., Sanders & Mirabel 1996, Kormendy et al. 2009).

3.2. Black Hole Growth

The first seed BHs may have been left behind after the explosions of massive stars that were formed out of primordial gas in the early Universe. These Pop III seed BHs are expected to have masses of ~ 100 $M_\odot$; however, such seeds cannot grow into the 10^9-$M_\odot$ BHs required to power observed quasars at $z \sim 6$–7 if their growth is Eddington limited. Recently, several mechanisms for creating more massive seed BHs (10^4–10^6 $M_\odot$) have been proposed (for a review see Volonteri 2010). However, in cosmological simulations, the usual approach is to place seed BHs by hand in halos above a critical mass ($M_{\rm H} \gtrsim 10^{10}$–$10^{11}$ $M_\odot$). In some cases, seeds of a fixed mass are used; in others, the seed mass is chosen in order to place the BH on the local $M_{\rm BH} - \sigma$ relation. The results that we discuss here are generally insensitive to the details of the seeding procedure.

One must then calculate how rapidly these seed BHs accrete gas and grow in mass. The currently predominant model relies on the idea that BH growth is limited by Bondi accretion of mass within the sphere of influence (Bondi 1952), given by

$$\dot{M}_{\rm Bondi} = \alpha \frac{4\pi\ G^2\ M_{\rm BH}^2\, \rho}{(c_{\rm s}^2 + v^2)^{3/2}}, \tag{4}$$

where $M_{\rm BH}$ is the mass of the BH, $c_{\rm s}$ is the sound speed of the gas, v is the bulk velocity of the BH relative to the gas, ρ is the density of the gas, and α is a boost parameter included because models typically lack the spatial resolution to resolve the Bondi radius (Booth & Schaye 2009, Johansson et al. 2009a). Early models took α to be constant (typically $\sim$100), but some simulators make α a function of density (e.g., Booth & Schaye 2009) and some recent simulations resolve the Bondi radius and therefore adopt $\alpha = 1$. Typically, the accretion rate is capped at the Eddington rate. As galaxies merge, their BHs are assumed to merge when they come within some distance of each other, typically a softening length (thereby ignoring GR timescales for BH inspiral).

The Bondi accretion model predicts fairly low accretion rates when galaxies are undisturbed, but when strong torques drive gas toward the nucleus as in a major merger, accretion rates can be boosted to levels sufficient to power quasars (Di Matteo et al. 2005, Springel et al. 2005b). This is consistent with the observation that local ultraluminous IR galaxies, which are mostly major mergers, also show strong AGN activity (Sanders & Mirabel 1996). In one paradigm, low accretion rates ($\lesssim 0.01 \dot{M}_{\rm Edd}$, in which $\dot{M}_{\rm Edd}$ is the Eddington rate) are associated with radiatively inefficient accretion, as in an advection-dominated accretion flow (Narayan & Yi 1994, Blandford & Begelman 1999). In this case, most of the energy is advected into the BH, and little emerges as radiation. BHs powered at higher accretion rates are radiatively efficient and give rise to the population of observed X-ray, UV, and optically luminous quasars and AGNs.

Assuming Bondi accretion requires the accompaniment of strong feedback to obtain BHs that follow the $M_{\rm BH} - \sigma$ relation, as this simple argument demonstrates (Anglés-Alcázar et al. 2013a). Consider two BHs of mass $M_{\rm a}$ and $M_{\rm b}$. If they grow according to the general prescription

$\dot{M}_{\rm BH} = D(t) M_{\rm BH}^{p}$, then

$$\frac{\rm d}{{\rm d}t}\left(\frac{M_{\rm a}}{M_{\rm b}}\right) = D(t)\frac{M_{\rm a}^{p}}{M_{\rm b}}\left[1-\left(\frac{M_{\rm a}}{M_{\rm b}}\right)^{1-p}\right]. \quad (5)$$

It is easy to show that the two masses will diverge if $p > 1$, and they will converge if $p < 1$. For Bondi accretion, $p = 2$; hence for BHs to converge onto an $M_{\rm BH} - \sigma$ relation, some strongly self-regulating feedback process must counteract Bondi accretion and make p effectively less than unity. We discuss possible feedback processes in Section 3.3.3, but in general such tuned self-regulation is not so straightforward to arrange, for the usual reason that outward energetic processes tend to escape through paths of least resistance, whereas inflows typically arrive through the dense, harder-to-disrupt gas.

It is worth emphasizing that the widely used Bondi model implicitly assumes that the accreting gas has negligible angular momentum, which is unlikely to be a good assumption in general. Recently, the problem of dissipating angular momentum to enable BH accretion has received more attention in the cosmological milieu. Hopkins & Quataert (2010, 2011) studied angular momentum transport in disks with nonaxisymmetric perturbations both analytically and in simulations, showing that such secular processes can significantly fuel BH growth, as also suggested by Bournaud et al. (2011b) and Gabor & Bournaud (2013). Implementing this analytic work into zooms and cosmological simulations, Anglés-Alcázar et al. (2013, 2014) showed that this torque-limited accretion behaves qualitatively differently than Bondi accretion, because in the Hopkins & Quataert (2011) model, the exponent of BH growth is $p = \frac{1}{6}$. Although this model also must incorporate feedback, such feedback does not have to strongly couple to the inflow to achieve self-regulation.

BH accretion in SAMs is of necessity more schematic. In one of the first semianalytic models that incorporated BH growth in the framework of a cosmological model of galaxy formation, Kauffmann & Haehnelt (2000) assumed that all BH growth is triggered by major mergers. Following such an event, they assumed that some fraction of the cold gas was accreted by the BH, with this fraction being a function of the halo circular velocity. A similar recipe is incorporated into the later generations of MPA-SAMs (e.g., Croton et al. 2006, De Lucia & Blaizot 2007, Guo et al. 2011, Henriques et al. 2013). Other SAMs additionally trigger accretion following minor mergers and disk instabilities (Bower et al. 2006, Somerville et al. 2008b, Hirschmann et al. 2012b). Some models allow an additional growth channel through a Bondi-like accretion from the hot halo (Somerville et al. 2008b, Fanidakis et al. 2011). In the Santa Cruz SAMs (Somerville et al. 2008b), BH growth is parameterized on the basis of the results of hydrodynamic binary merger simulations (Cox et al. 2006, 2008; Robertson et al. 2006b) as characterized by Hopkins et al. (2005b). In these models, rapid BH accretion is triggered following a major or minor merger. The BH accretes at the Eddington rate until the BH reaches a critical mass, in which the energy being radiated is sufficient to halt further accretion. The accretion rate then declines in a power-law blow out phase until the BH switches off (Hopkins et al. 2005a).

All SAMs and numerical cosmological hydrodynamic simulations that explicitly include BH growth use the local $M_{\rm BH} - \sigma$ or $M_{\rm BH} - M_{\rm spheroid}$ relation to calibrate the free parameters in the BH accretion recipes. A wide variety of BH growth recipes appear to be able to successfully reproduce this relationship.

3.3. Feedback Processes

Feedback can be divided into two general classes, preventive and ejective. Preventive feedback retards SF by stopping gas from accreting into the ISM, whereas ejective feedback describes

processes that remove the gas from the ISM after it has been accreted. Current wisdom suggests that preventive feedback dominates when the majority of halo gas is near the halo's virial temperature (as in very small dwarfs or massive galaxies), whereas ejective feedback dominates when most of the halo's gas is well below its virial temperature (as in typical star-forming galaxies). However, individual physical processes can potentially act in both ejective and preventive ways.

3.3.1. Squelching: photoionization suppression. Photons above 13.6 eV that ionize hydrogen typically add an approximately electronvolt-scale amount of latent heat, corresponding to a temperature increase of $\sim 10^4$ K. Hence the postreionization optically thin IGM has a temperature around this value, which means that gas in halos whose virial temperatures are comparable with 10^4 K should be unable to cool their gas. This temperature corresponds to a halo mass of $\sim 10^8$ $M_\odot$, implying that photoionization will strongly reduce the baryon content and hence suppress galaxy formation in halos below this mass. This suppression has sometimes been called squelching (Somerville 2002).

Squelching can have a residual impact on halos much larger than 10^8 $M_\odot$, because they are hierarchically assembled in part from squelched halos. Gnedin (2000) showed that the characteristic mass below which halos contain substantially less than their fair share of baryons is well represented by a filtering scale that smooths the baryonic perturbations. Hence one can define a filtering mass, which describes the halo mass that on average contains half the cosmic fraction of baryons.

The filtering mass depends on the intricate interplay between photoionization, cooling, and hierarchical growth, which is challenging to model. Early work suggested a roughly constant circular velocity, below which baryon accretion is suppressed, of around 30–50 km s^{-1} (e.g., Quinn et al. 1996, Thoul & Weinberg 1996). If extrapolated to today, this would imply halos up to several times 10^{10} $M_\odot$ would be significantly suppressed in baryon content (Gnedin 2000). More recent simulations by Okamoto et al. (2008) found a smaller filtering mass scale, $M_F \sim 4 \times 10^9$ $M_\odot$ today, but these simulations still assumed ionization equilibrium, did not include metal-line cooling, and adopted a uniform metagalactic ionizing background. Observations of late-type dwarfs with circular velocities $\lesssim$42 km s^{-1} suggest that their baryon content is much smaller than expected based on scaling relations from larger galaxies (Kormendy & Freeman 2014), thus providing direct constraints on the filtering mass.

An additional complication can arise when galactic outflows are included along with squelching, as the two can combine to produce an amplification of suppression that is stronger than the product of the individual effects (Pawlik & Schaye 2009, Finlator et al. 2012). The magnitude of the effect depends on the outflow model implemented, but can be up to a 60% amplification during the EoR. Unfortunately, the high expense of these calculations that include radiative transfer while resolving very small halos prohibits their evolution down to $z = 0$; hence it is not clear how significant this effect is at later epochs.

In SAMs, photoionization squelching is generally implemented by assuming that reionization occurs instantaneously throughout the Universe at a fixed input redshift. At all later times, the gas that is allowed to accrete into halos is reduced by a factor $f_{\rm coll}(M_H, z)$. This function is parameterized on the basis of the results of numerical hydrodynamic simulations and is expressed as a function of the filtering mass (Gnedin 2000, Kravtsov et al. 2004, Okamoto et al. 2008).

3.3.2. Star-formation feedback. Stars, massive ones in particular, deposit copious amounts of energy and momentum into the ISM during their lives and deaths. Stellar feedback is invoked to explain two kinds of inefficiencies in galaxies: (*a*) The efficiency of the conversion of gas into stars within GMCs is puzzlingly low, only approximately 1% per free-fall time (Krumholz et al. 2012); (*b*) the stellar and baryon fraction within galactic-sized halos is much smaller than the

universal value, ranging from a few to 20% (Behroozi et al. 2010, Moster et al. 2010b). The first inefficiency has been ascribed to turbulence generated by stars and SNe within GMCs (e.g., Krumholz et al. 2012 and references therein). For purposes of cosmological simulations, because observations suggest that this efficiency is nearly universal, this efficiency can largely be folded into the normalization of the SF recipe.

The second inefficiency is generally ascribed to large-scale galactic outflows powered by massive stars and SNe. Signatures of such outflows, with mass-loss rates likely of the same order as the SFR or larger, are ubiquitously observed in star-forming galaxies (Veilleux et al. 2005). Modeling galactic outflows has therefore become a central challenge for recent simulations.

Early work attempted to model stellar feedback via the deposition of thermal energy from SNe in the surrounding gas (e.g., Katz et al. 1996). It was quickly realized that this had almost no effect, because the short cooling times meant that the energy radiated away very quickly, adding negligible ISM pressure, let alone driving an outflow. Since then, most cosmological models have adopted some subgrid prescription to enable effective ejective feedback that typically involves either implementing ad hoc tricks, such as turning off cooling for some time or superheating the gas, that attempt to mimic the ISM processes that allow stellar-driven winds to develop in real galaxies or else implementing outflows via kinetic energy injection.

A variant on the ISM heating model called blast wave feedback was developed by Stinson et al. (2006) and has been extensively used in GASOLINE and RAMSES (Bournaud et al. 2010). Here, after the gas is heated, radiative cooling is shut off for the lifetime of the SN-driven blast wave as predicted by a spherical Sedov solution. This enables the gas to feel the higher pressure and develop a coherent large-scale outflow. Although successful in many regards, this model still predicted too much early SF, so Stinson et al. (2013) added early stellar feedback intended to mimic the energy input from young stellar winds and radiation.

Another variant of a purely thermal stellar feedback model was proposed by Dalla Vecchia & Schaye (2012) and implemented in the EAGLE simulations (Schaye et al. 2015). Instead of turning off cooling, the trick for preventing the energy from immediately cooling away involves making the energy deposition stochastic. The mean amount of energy injected per mass of stars formed is set by stellar population models and SN energetics. The temperature jump of particles receiving a boost is specified ($\Delta T = 10^{7.5}$ K, typically), and a parameter $f_{\rm th}$ determines the probability that a given SPH particle in the vicinity of a star-forming particle will get heated. Hence the gas is heated to much higher temperatures than would be the case if the same amount of energy were continuously added to all the SPH neighbors, increasing the cooling time and mitigating energy losses. The overall efficiency of the feedback can be adjusted by varying $f_{\rm th}$. Schaye et al. (2015) made $f_{\rm th}$ a function of the local gas metallicity and density, as they found that this most successfully reproduced the observed SMFs and galaxy sizes.

A popular approach introduced by Navarro & White (1993) and implemented into GADGET-2 by Springel & Hernquist (2003) is to simulate outflows by giving gas kicks, rather than trying to overpressurize ISM gas by adding thermal energy. Such kinetic outflows are less directly tied to the physics generating outflows, but this approach enables greater control over outflow parameters in order to both mimic observed outflows more closely and assess the impact of varying the outflow parameters. In such models, hydrodynamics is sometimes shut off (decoupled) for some interval to mimic the collective power of SNe blowing a chimney through the ISM; it is unclear whether this provides a more physical description of outflow propagation through the ISM, but it generally does result in better resolution convergence (Dalla Vecchia & Schaye 2008). These models are parameterized by a mass loading factor $\eta \equiv \dot{M}_{\rm out}/\dot{M}_*$ and a wind velocity $v_{\rm wind}$, which together determine how many particles to kick and how hard to kick them. Springel & Hernquist (2003) assumed a constant mass loading factor and constant wind velocity and showed that this yielded

a cosmic SF history in much better agreement than a model without outflows. Oppenheimer & Davé (2006) showed that adopting scalings motivated by analytic momentum-driven wind models (Murray et al. 2005) produced better agreement with many galaxy and IGM properties, including the galaxy MZR, the enrichment history of the IGM, and the galaxy SMF (see also Finlator & Davé 2008; Oppenheimer & Davé 2008; Davé et al. 2011b, 2013). For momentum-driven winds, the mass loading factor scales as $\eta \propto \sigma^{-1}$, and the wind velocity scales as $v_{\rm wind} \propto \sigma$, where σ is the velocity dispersion of the galaxy. The Illustris simulations (Vogelsberger et al. 2014a) also employ kinetic winds by creating and launching decoupled wind particles and rejoining them back into the gas mesh after recoupling. They adopt scalings expected for energy-driven winds, namely $\eta \propto \sigma^{-2}$.

Most semianalytic models parameterize SF feedback in a similar manner based on the approach introduced by White & Frenk (1991) and Kauffmann et al. (1993). In each time-step, the SAM computes the rate at which cold gas is ejected from the disk by a wind:

$$\dot{m}_{\rm ej} = \epsilon_{\rm w} \left(\frac{V_0}{V_{\rm c}} \right)^{\alpha_{\rm w}} \dot{m}_*,$$

where $\dot{m}_*$ is the SFR in the galaxy, $V_{\rm c}$ is the circular velocity of the galaxy, V_0 is an arbitrary normalization parameter, and $\epsilon_{\rm w}$ and $\alpha_{\rm w}$ are treated as tunable free parameters. For $\alpha_{\rm w} = 1$ or $\alpha_{\rm w} = 2$, this is equivalent to the momentum-driven or energy-driven wind scalings discussed above. One must then decide what happens to the ejected gas, and here different modelers diverge more widely. Some fraction of the ejected gas may escape the DM halo and may be tracked in an ejected reservoir from which it is allowed to accrete into the halo again over a longer timescale. Otherwise, the ejected gas is added to the halo hot gas reservoir. SAMs generally implement some sort of model, of varying complexity, to arrange that the fraction of ejected gas that escapes the halo is larger at lower halo $V_{\rm H}$, and asymptotes to unity above $V_{\rm H} \simeq 120$–150 km s^{-1} (or a halo mass of a few $\times 10^{12}$ M$_\odot$).

3.3.3. Active galactic nucleus feedback. Observational phenomena associated with accreting BHs include electromagnetic radiation, relativistic jets, and less-collimated nonrelativistic outflows (Krolik 1999). There are several different physical mechanisms whereby the large amounts of energy and momentum produced by AGNs can couple with the gas in and around galaxies, possibly regulating the growth of the BH itself and potentially suppressing cooling and SF on galactic scales. At the most basic level, AGNs can heat gas up (thermal feedback), drive winds that eject gas (kinetic feedback), and ionize or photodissociate gas (radiative feedback). The main heating mechanisms are Compton, photoionization, and photo-electric heating. Radiation may also drive winds via pressure on spectral lines, free electrons, or dust. These winds may originate in the torus or accretion structure near the BH, the broad-line region (BLR), larger nuclear scales (about kiloparsec), or all of the above. Winds arising on small (BLR/accretion disk) scales may drive galaxy-scale winds by shocking and sweeping up ISM gas—or they may simply vent out of the galaxy without ejecting much mass. In addition, highly relativistic giant radio jets may heat the intracluster medium through bubbles, weak shocks, and sound waves (McNamara & Nulsen 2007, Fabian 2012).

Focusing first on the processes associated with the radiatively efficient radiative mode (sometimes called quasar mode or bright mode) of BH growth, one of the major dynamical questions is whether AGN-driven winds are primarily energy driven or momentum driven. As in the case of stellar driven winds, the question is how quickly and efficiently the thermal energy that is generated when the wind shocks the surrounding gas is radiated away. Momentum cannot, of course,

be radiated away, and so if most of the thermal energy is quickly dissipated, we term the wind momentum driven. If radiative losses are negligible, we term it energy driven. Clearly, real winds may often be somewhere in between. The significance of this distinction is that the momentum flux of swept-up material in an energy-conserving outflow is boosted as a result of work done by the hot shocked gas (an effect familiar from the Sedov-Taylor phase in SN remnants).

It has been argued that in the dense cold gas that must surround rapidly accreting BHs, cooling times are short and winds must be predominantly momentum-driven (King 2005, Ostriker et al. 2010, Debuhr et al. 2011). However, observations of AGN-driven outflows suggest boost factors of $\dot{p}/\dot{p}_{\rm rad} \sim 2$–30 (e.g., Moe et al. 2009, Sturm et al. 2011), with an average probably around 10, in which $\dot{p}_{\rm rad} = L_{\rm AGN}/c$ is the radiative momentum flux output by the AGN. Faucher-Giguère & Quataert (2012) argued recently on the basis of analytic calculations that AGN-driven outflows are likely to be largely energy-conserving in many situations relevant to observed systems, particularly for fast ($v_{\rm w} \sim 10{,}000$–$30{,}000$ km s^{-1}) winds.

One of the earliest 3D simulations of AGN feedback in galaxies was presented by Springel et al. (2005b) and Di Matteo et al. (2005). Here, the BH accretion rate was modeled using the Bondi approach outlined above, and the resulting bolometric luminosity was assumed to be proportional to the BH accretion rate. A fixed fraction of the bolometric luminosity was deposited into the neighboring gas particles as thermal energy. These simulations did not use cosmological initial conditions but considered binary mergers of idealized disk galaxies without hot gas halos. This work showed that deposition of approximately 5% of the bolometric luminosity was able to drive strong outflows that eventually halted further accretion onto the BH and also removed nearly all cold gas from the galaxy, resulting in quenching of SF (Springel et al. 2005a). Furthermore, the models produced self-regulated BH growth, leading to a tight $M_{\rm BH} - \sigma$ relationship in agreement with the observed one. A similar approach has been used in many subsequent studies. Although these studies, taken at face value, suggest that purely energy-driven winds can regulate BH growth and drive large-scale outflows, it is likely that radiative losses were artificially suppressed owing to the highly pressurized ISM model adopted in these simulations.

Moreover, these simulations neglected the expected momentum deposition. Several recent works have implemented momentum-driven winds in hydrodynamic simulations via radiation pressure on dust (Debuhr et al. 2010, 2011) and via BLR winds (Choi et al. 2012, 2014a) and found that these winds can play a significant role in modulating the growth of the BH and the galaxy. In addition, the SF remains quenched over a much longer timescale in the simulations that include momentum-feedback, because the density of hot gas near the center of the halos is significantly reduced (Choi et al. 2014b).

The other important class of feedback processes is connected with highly collimated jets of relativistic particles (jet mode or radio mode; see the recent reviews by Fabian 2012 and Heckman & Best 2014). The kinetic energy in these jets can exceed the total bolometric luminosity of the AGN by several orders of magnitude. Although jets may be observed at many wavelengths, there is a class of sources detected at radio wavelengths that do not exhibit the classic signatures of radiatively efficient AGNs—no UV, X-ray, or IR excess, and no highly ionized emission lines. Optically, these objects resemble normal massive early-type galaxies. They are associated with radiatively inefficient accretion, and with extremely low accretion rates onto the central BH. The radio jets are observed to correspond, in many cases, with bubbles visible in X-ray images, regions filled with hot plasma presumably heated by shocks from the jet's interaction with the intracluster medium. Studies of the bubble energetics have shown that there is easily enough energy deposited in the intracluster medium to offset cooling; in fact, in groups and low-mass clusters the energy probably exceeds the requirements for balancing cooling by up to an order of magnitude. Radio

galaxies are common in massive early type galaxies in groups and clusters, and bubbles and/or radio sources are seen in 95% of cool core clusters (clusters with short central cooling times).

Once again, the energetics are such that one expects this jet mode feedback to have a significant impact on galaxy formation, but many details of the physics remain unclear. The main puzzle is how such highly collimated bipolar jets can nearly isotropically heat a large volume of intragroup or cluster gas (Vernaleo & Reynolds 2006). The bubbles provide an important clue—these bubbles rise buoyantly in the hot atmosphere, reaching fairly large radii. Heating may occur via turbulent mixing of bubbles with the intracluster medium (Scannapieco & Brüggen 2008), viscous dissipation of weak shocks (Ruszkowski et al. 2004), or cosmic ray heating (Sharma et al. 2009). Although some recent simulations that attempt to explicitly model jet heating in 3D have claimed greater success at averting the cooling flow problem (Gaspari et al. 2011, Li & Bryan 2014), all these simulations neglect a cosmological formation history, with merging and accretion, as well as SF and stellar feedback. A detailed physical understanding of how the jets couple to the surrounding hot gas and how effective they are in regulating cooling flows over long timescales remains lacking (see also Babul et al. 2013, Cielo et al. 2014).

Sijacki et al. (2007) were the first to attempt to include both the radiative and jet modes of AGN feedback in numerical cosmological simulations, albeit in a simplified way, necessitated by the relatively coarse numerical resolution. Above a critical BH accretion rate ($\sim$0.01 times the Eddington rate), the AGN was assumed to be radiatively efficient and a fraction of the AGN bolometric luminosity was deposited in the gas as thermal energy. Below the critical accretion rate, the AGN is assumed to be radiatively inefficient and to produce jets that inflate bubbles—however, they do not directly simulate the jet. Instead they insert bubbles by hand, with energy and radius scaled to the BH mass as motivated by analytic models for radio cocoon expansion.

Similar approaches have now been implemented in a few sets of cosmological simulations. The Illustris simulations, using the AREPO moving mesh code, also use the Bondi accretion model and a similar feedback scheme to that of Sijacki et al. (2007). In addition, the Illustris simulations include a simplified treatment of photoionization and photoheating arising from the AGN radiation field. A somewhat different approach is taken in the EAGLE (Schaye et al. 2015) and OWLS (Schaye et al. 2010) simulations—these adopt a variant of the stochastic thermal feedback model used for SF feedback, described in Section 3.3.2, in which an average energy injection rate is given by $E_{\rm BH} \propto \dot{m}_{\rm acc} c^2$, where $\dot{m}_{\rm acc}$ is the accretion rate onto the BH and c is the speed of light. The injected energy is stored by each BH until it can stochastically heat some minimum number of particles with a temperature increase ΔT. The value of ΔT may depend on resolution and is higher than for the stellar feedback model $\Delta T \sim 10^{8.5}$–10^9 K (Schaye et al. 2015). Other simulations do not explicitly follow BH growth and associated feedback but include heuristic quenching mechanisms based on surrogate galaxy or halo properties (Gabor et al. 2011, Gabor & Davé 2012).

Several groups have implemented radiative mode AGN-driven winds and jet mode AGN heating in SAMs. Although the details differ from model to model, there are a number of common elements that are widely adopted: (*a*) A distinction is made between BH fueling via cold gas (which is typically assumed to be driven into the nucleus by mergers and/or disk instabilities; see Section 3.2) and by hot gas, which is generally assumed to accrete via a cooling flow. (*b*) BH accretion fueled by the merger/disk instability–driven mode is associated with radiatively efficient accretion at a significant fraction of the Eddington rate; accretion fueled by hot gas is assumed to lead to very sub-Eddington, radiatively inefficient accretion associated with the jet mode. (*c*) The jet mode is assumed to be activated only in the presence of a quasi-hydrostatic hot halo, i.e., when the halo is predominantly accreting via the hot mode discussed earlier. (*d*) The jet mode is able to extract a certain fraction of the BH mass in the form of energy, which is used to offset cooling or assumed to be able to establish a heating-cooling balance when the BH mass exceeds a critical value.

For example, in the Croton et al. (2006) model, the jet mode accretion rate is modeled as

$$\dot{m}_{\rm BH,R} = \kappa_{\rm AGN} \left(\frac{m_{\rm BH}}{10^8\ {\rm M}_\odot} \right) \left(\frac{f_{\rm hot}}{0.1} \right) \left(\frac{V_{\rm vir}}{200\,{\rm km\,s^{-1}}} \right)^3,$$

where $f_{\rm hot}$ is the fraction of the total halo mass in the form of hot gas, $m_{\rm BH}$ is the mass of the BH, and $V_{\rm vir}$ is the virial velocity of the halo. The cooling rate computed as described in Section 2.3.2 is offset by a heating term, such that the effective cooling rate is

$$\dot{m}_{\rm cool,eff} = \dot{m}_{\rm cool} - \frac{L_{\rm AGN}}{\frac{1}{2} V_{\rm vir}^2},$$

where $L_{\rm AGN} = \epsilon_{\rm rad} \dot{m}_{\rm BH} c^2$ with $\epsilon_{\rm rad} = 0.1$, which is the conversion of accreted rest mass into energy. Other SAMs use similar scalings, some of which attempt to make a more explicit connection with the invoked physical processes and/or with observations (e.g., Monaco et al. 2007, Somerville et al. 2008b), but these appear to produce similar results at $z = 0$, even for the redshift evolution of massive galaxies (Fontanot et al. 2009).

In addition to jet mode feedback, some SAMs implement AGN-driven winds. Somerville et al. (2008b) adopted momentum-driven wind scalings associated with radiative mode AGN activity:

$$\frac{{\rm d}M_{\rm out}}{{\rm d}t} = \epsilon_{\rm wind} \epsilon_{\rm rad} \frac{c}{V_{\rm esc}} \dot{m}_{\rm acc},$$

where $\epsilon_{\rm wind}$ represents the effective coupling efficiency, $V_{\rm esc}$ is the escape velocity of the galaxy, and $\dot{m}_{\rm acc}$ is the BH accretion rate in the radiatively efficient mode. See also Fontanot et al. (2006) for an alternate implementation of radiative mode wind feedback in SAMs.

Examples of SAMs that do not follow BH growth explicitly, but instead implement more heuristic halo-based quenching, include Cattaneo et al. (2006) and Lu et al. (2011). In these models, cooling is simply switched off when the halo mass exceeds a critical value, which may depend on redshift.

4. RESULTS FROM CURRENT MODELS: INSIGHTS AND PUZZLES

We return for a moment to **Figures 1** and **2** to illustrate some general insights into the process of galaxy formation and evolution in the ΛCDM framework that have arisen from numerical simulations. Starting with the left column of **Figure 1**, we see that structure formation in the DM component proceeds via the formation of sheets or giant walls, which form filaments where the sheets intersect. DM halos form at the intersection of filaments, which funnel DM and gas into halos like tributaries flowing into a lake. Comparing the first and second columns of **Figure 1**, one can see that there is a very strong correspondence between the DM and gas density fields on large scales. This illustrates that gas flows on large scales are dominated by gravity. Moving to the third column of **Figure 1**, we can see that the gas surrounding massive halos is hot, and larger regions become heated as time progresses. This heating is in part due to shock heating as halos collapse, but in these simulations it is also in large part due to SF and AGN feedback. Finally, examining the rightmost column of **Figure 1**, we see that metals are dispersed to quite large distances from galaxies, and polluted regions again fill a larger comoving volume over time. **Figure 2** shows how filaments of relatively cold gas can sometimes penetrate some distance into hot halos—these supply the cold mode accretion discussed earlier (sometimes called stream fed accretion). The inset in **Figure 2** emphasizes how small galaxies are compared with the structures seen in the cosmic web.

4.1. Global Properties

LF: luminosity function

In this subsection, we discuss model predictions for global properties of galaxies and how they compare with observations. In Section 4.1.1, we discuss distribution functions; in Section 4.1.2, we discuss scaling relations; and in Section 4.1.3, we discuss demographics of star-forming and quiescent galaxies.

4.1.1. Stellar mass assembly over cosmic time. A fundamental observational target for modelers is reproducing the statistical distributions of global properties for galaxy populations at different cosmic epochs, such as luminosity functions (LFs), SMFs, and cold gas mass functions. It has been realized for some time that the observed local LF or SMF is not naturally reproduced by galaxy formation models based within the ΛCDM paradigm: CDM models generically predict that the slope of the mass function of DM halos has a slope of $\alpha_H \sim -2$, whereas the slope of the observed galaxy SMF locally is much shallower ($\alpha_g \simeq -1.3$). A number of authors have suggested that SN feedback could flatten out the low-mass slope by suppressing SF in low-mass halos (Larson 1974, Dekel & Silk 1986, White & Frenk 1991). Furthermore, although ΛCDM predicts an exponential cutoff or knee in the halo mass function, with a similar functional form to that of the observed SMF, the halo mass function turnover is at much larger masses. Although the cooling times in these massive-, group-, and cluster-sized halos are predicted to be somewhat longer than in low-mass halos (Rees & Ostriker 1977, Blumenthal et al. 1984), this turns out to be insufficient to explain the very inefficient SF required to reconcile the abundance of massive galaxies with that of DM halos.

After decades of effort, theoretical models of galaxy formation are now fairly successful at reproducing the SMF of galaxies at $z \sim 0$ by invoking a plausible, if still in most cases schematic, set of physical processes. **Figure 4** shows a compilation of predictions of recent numerical hydrodynamic simulations and SAMs for the SMF from $z = 4$ to $z \sim 0$. These models are all taken directly from the original publications, and no attempt has been made to calibrate them to the same set of observations or to correct for the slight differences in cosmology. [The stellar masses in the GALFORM models have been multiplied by a factor of 1.23 to convert from a Kennicutt to a Chabrier IMF (Mitchell et al. 2013).] This success has been obtained by tuning not only free parameters but also the recipes associated with the subgrid physics (SF, stellar feedback, AGN feedback). Predictions of the buildup of stellar mass over cosmic time, with these recipes and parameters held fixed, present a more stringent test of the models.

In a broad-brush sense the model predictions are generally encouraging. A very general prediction of ΛCDM-based models is that galaxies built up their stellar mass gradually over time, which is supported by observations. All models predict efficient early SF ($z \gtrsim 4$) in low-mass halos and steep stellar mass and rest-UV luminosity functions at these early epochs, in agreement with observations. Models including AGN feedback or heuristic quenching predict that massive galaxies formed earlier and more rapidly than lower-mass galaxies, again in qualitative agreement with observations. Most models even demonstrate very good quantitative agreement, within the errors on stellar mass estimates, between predicted and observed SMFs and LFs for massive galaxies ($m_* > M_{\rm char}$). Note that in **Figure 4**, the expected uncertainties in the stellar masses that are inherent in the observational estimates have not been accounted for in most of the model comparisons. Including these in a simplified manner brings the model predictions into better agreement with the observations on the massive end (e.g., Henriques et al. 2013, Lu et al. 2013), as shown here for the MPA SAM as an illustration. For a more detailed study of this issue see Mitchell et al. (2014).

As can be seen as well in **Figure 4**, models currently have greater difficulties reproducing the abundances and assembly histories of low-mass galaxies at intermediate redshifts. Fontanot

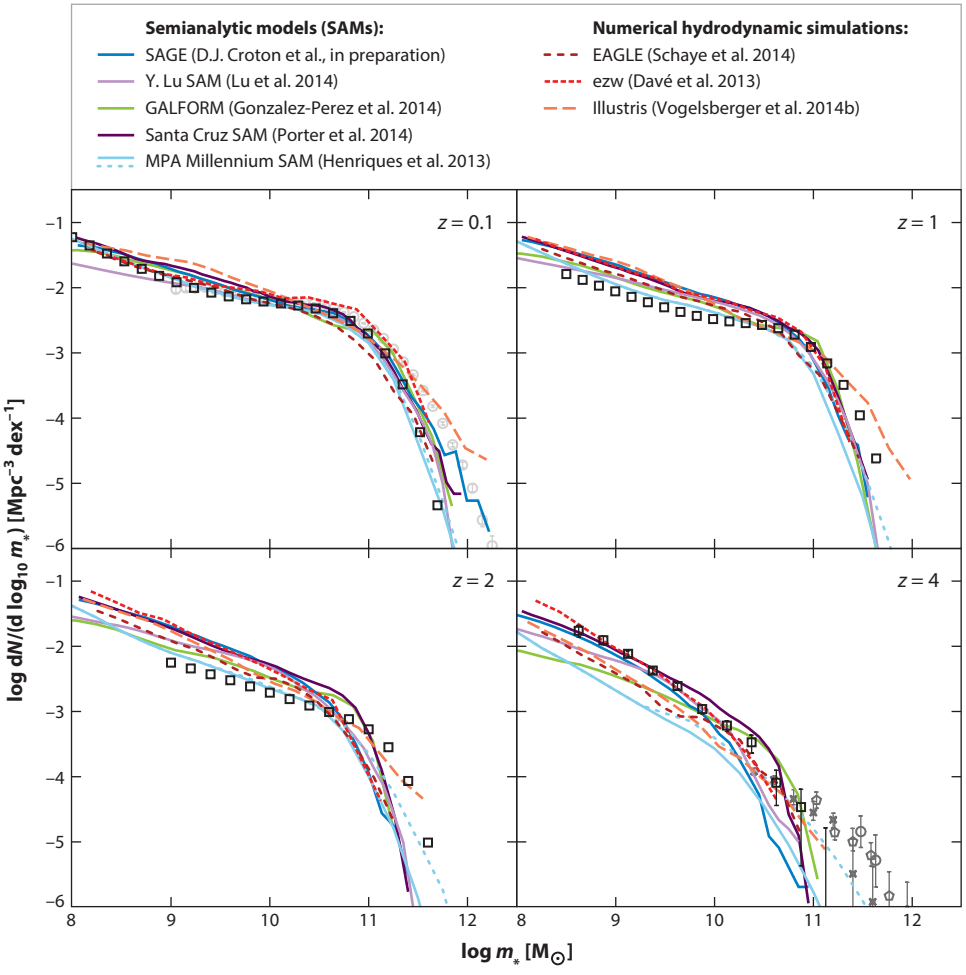

Figure 4

Galaxy stellar mass function at redshifts $z \sim 0$–4. In the $z = 0.1$, $z = 1$, and $z = 2$ panels, black square symbols show a double-Schechter fit to a compilation of observational estimates. Observations included in the fit are: $z = 0.1$, Baldry et al. (2008), Moustakas et al. (2013); and $z = 1$ and $z = 2$, Tomczak et al. (2014), Muzzin et al. (2013). The fits shown at $z = 1$ and $z = 2$ are interpolated to these redshifts from adjacent redshift bins in the original published results. The formal quoted 1σ errors on the estimates shown in these three panels are comparable with the symbol size and are not shown for clarity (the actual uncertainties are much larger but are difficult to estimate accurately). In the $z = 0.1$ panel, the estimates by Bernardi et al. (2013) are also shown (*open gray circles*). In the $z = 4$ panel, we show estimates from Duncan et al. (2014, *squares*), Caputi et al. (2011, *crosses*), Marchesini et al. (2010, *circles*, for $z = 3$–4), and Muzzin et al. (2013, *pentagons*, for $z = 3$–4). Solid colored lines show predictions from semianalytic models (SAMs): SAGE (D.J. Croton, A.R.H. Stevens, C. Tonini, T. Garel, M. Bernyk, et al., in preparation, *dark blue*), Y. Lu SAM (Lu et al. 2014, *magenta*), GALFORM (Gonzalez-Perez et al. 2014, *green*), the Santa Cruz SAM (Porter et al. 2014, *purple*), and the MPA Millennium SAM (Henriques et al. 2013, *light blue*). The dotted light blue line shows the Henriques et al. (2013) SAM with observational errors convolved. Colored dashed lines show predictions from numerical hydrodynamic simulations: EAGLE simulations (Schaye et al. 2015, *dark red*), ezw simulations of Davé and collaborators (Davé et al. 2013, *bright red*), and the Illustris simulations (Vogelsberger et al. 2014b, *orange*).

et al. (2009) demonstrated that three independently developed SAMs overproduce galaxies with $m_* \lesssim 10^{10}$ $M_\odot$ by a factor of ~2–3 over the redshift range of $4 \lesssim z \lesssim 0.5$. Weinmann et al. (2012) showed that a qualitatively similar problem exists for SAMs and for hydrodynamic simulations. This problem appears to persist even in the latest state-of-the-art cosmological hydrodynamic simulations, as seen in **Figure 4**, and already pointed out in the case of Illustris by Torrey et al. (2014). As discussed by Fontanot et al. (2009), several different sets of observations suggest that massive galaxies form early and rapidly, whereas low-mass galaxies form later and with a more extended timescale—the phenomenon that is often referred to as downsizing or staged galaxy formation (Noeske et al. 2007a). The overproduction of low-mass galaxies is a symptom of the failure of current models to reproduce this mass dependence in the SF histories of galaxies. Low-mass DM halos actually have earlier formation times than high-mass halos—the opposite of the trend seen in observations (Conroy & Wechsler 2009). In current simulations, the SF histories closely trace the DM mass accretion histories, thus similarly failing to reproduce the observed trend.

It seems clear that the subgrid recipes controlling SF and/or stellar feedback need to be modified to solve this problem. Henriques et al. (2013) found that making the stellar feedback stronger and modifying the timescale for the reaccretion of ejected gas led to significant improvement in the MPA-SAM for the predicted abundances of low-mass galaxies as well as other observed properties at $z \lesssim 3$. White et al. (2014) investigated several classes of empirical solutions to this problem, including modifying the efficiency of stellar-driven galaxy outflows, modifying the timescale for gas to turn into stars, and modifying the timescale for gas to be accreted (or reaccreted) into galaxies. They concluded that solutions that modified the outflow efficiencies and accretion timescales were the most promising. Moreover, Torrey et al. (2014) experimented with changing the coupling strength and velocity of the stellar-driven winds and found that this can change the normalization of the SMF at the low-mass end but cannot change the evolutionary shape, which is what is required to solve this problem.

A convenient way to assess the success of a cosmological simulation in reproducing the galaxy SMF or LF is via empirical constraints on the relationship between stellar mass (or luminosity) and halo mass, as derived by galaxy-halo mapping techniques, such as SHAM and HOD, and other methods such as galaxy-galaxy lensing, clustering, satellite kinematics, and X-ray observations (Behroozi et al. 2010, 2013a; Moster et al. 2010b, 2013 and references therein). Different methods and groups are generally all in broad agreement that SF feedback plays a crucial role in shaping this relationship for halos with $M_H \lesssim 10^{12}$ $M_\odot$, with photoionization squelching perhaps also playing a significant role below halo masses of approximately a few 10^{10} $M_\odot$ (this mass scale remains uncertain; see Section 3.3.1). At larger halo masses, $M_H \gtrsim 10^{12}$ $M_\odot$, there is a general consensus that AGN feedback probably plays an important role, although other processes (such as gravitational heating) may contribute as well (Khochfar & Ostriker 2008, Johansson et al. 2009b, Birnboim & Dekel 2011). To reproduce the slope of the stellar mass–halo mass (m_*–M_H) relation at low masses, most models adopt stellar feedback recipes that either assume or result in mass loading factors that increase fairly strongly with decreasing halo mass or circular velocity, similar to the energy- or momentum-driven wind scalings discussed in Section 3.3.2.

There is a broad though not universal consensus that AGN feedback implemented purely via deposition of thermal energy associated with the radiatively efficient mode of BH growth (as in, e.g., Di Matteo et al. 2005) does not by itself suppress cooling and SF in massive halos enough (or on long enough timescales) to satisfy observational constraints. Although thermal energy deposition can temporarily slow or halt cooling, after several gigayears, the gas starts to recool and form stars (Gabor & Davé 2012, Choi et al. 2014b). An exception is the stochastic thermal feedback model implemented in the EAGLE simulations, which reproduces the observed stellar fractions very well, though there is still tension between the predicted temperatures of the hot gas

in group- and cluster-sized halos and X-ray observations (Schaye et al. 2015). Another solution on the high-mass end is nearly constant injection of energy via jet mode feedback, although as discussed in Section 3.3.3, implementations of this process in cosmological simulations remain schematic. Inclusion of the momentum deposition associated with the radiatively efficient mode also appears to be able to suppress cooling for longer periods of time (Choi et al. 2014b).

4.1.2. Global scaling relations: gas, star formation, and metals. Galaxies are composed of stars, gas, metals, BHs, and DM. The scaling relations between these properties as a function of mass and redshift provide crucial constraints on galaxy growth and are in principle among the most direct ways to constrain baryon cycling processes.

The basic origin of many scaling relations can be understood in a simple framework based on mass balance in the ISM (alternately called an equilibrium, bathtub, or gas regulator model):

$$\dot{M}_{\rm inflow} = \dot{M}_* + \dot{M}_{\rm outflow} + \dot{M}_{\rm gas}, \tag{6}$$

where the terms are the baryonic mass inflow rate, SFR, mass outflow rate, and rate of change of the mass in the gas reservoir. When averaged over cosmological timescales, $\dot{M}_{\rm gas}$ is expected to be small compared with the other terms (Finlator & Davé 2008, Davé et al. 2012, Dekel & Mandelker 2014), though its evolution can have important effects (Lilly et al. 2013). Inflow into halos is driven primarily by gravitational accretion from the IGM (Dekel et al. 2009a, Kereš et al. 2005). The rate at which DM halos grow, the halo mass accretion rate ($\dot{M}_H$), is well characterized in ΛCDM and roughly given by $\dot{M}_H \propto M_H\,(1+z)^{2.5}$ (Dekel et al. 2009a, Faucher-Giguère et al. 2011). However, preventive feedback within galaxy halos can retard gas accretion into the ISM, and outflows can remove fuel for SF even after it enters the ISM, so $\dot{M}_{\rm inflow}$ may not trace $\dot{M}_H$.

We can rewrite Equation 6 as

$$\mathrm{sSFR} = \frac{\zeta(1+z)^{2.5}}{(m_*/M_H)\times(1+\eta)}, \tag{7}$$

where ζ is the fraction of material entering the virial radius that makes it into the ISM, and $\eta \equiv \dot{M}_{\rm outflow}/\dot{m}_*$ is the outflow mass loading factor. The dependence of sSFR on m_* and z therefore reflects the evolving combination of accretion and feedback.

Figure 5 shows a comparison of SFR versus m_* for the SAMs and simulations shown in **Figure 4**, along with a compilation of recent observational determinations as described in the figure caption. All models generally reproduce the near-unity slope at all redshifts. Most models match the amplitude at $z \sim 0$, although the turnover, due to quenching, at high masses can vary significantly (and can be sensitive to the definition of which galaxies are considered star-forming galaxies), and models tend to predict a steeper trend at low masses. By $z \sim 1$–2, it is clear that most models fall below the observations, a long-standing discrepancy first highlighted by Daddi et al. (2007). The redshift dependence of the sSFR is generically difficult to match in models because it differs strongly in the intermediate redshift regime ($4 \lesssim z \lesssim 0.5$) from the dependence predicted by the DM mass accretion rate (Davé 2008, Sparre et al. 2014). By $z = 4$, some models are able to match the data, though others continue to fall substantially short. The normalization of the predicted SFR versus m_* relation depends on resolution and the calibration of the subgrid parameters—e.g., Schaye et al. (2015) show (in their figure 11) that a higher resolution simulation in the EAGLE suite, recalibrated to the SMF, predicts a higher SFR at $m_* \lesssim 10^{10}$, in better agreement with the observations. However, the redshift dependence of the sSFR is roughly unchanged (Furlong et al. 2014).

These trends can be generally understood in the mass balance framework. From abundance matching, the m_*–M_H relation is constrained to evolve mildly with redshift (Behroozi et al. 2013a,

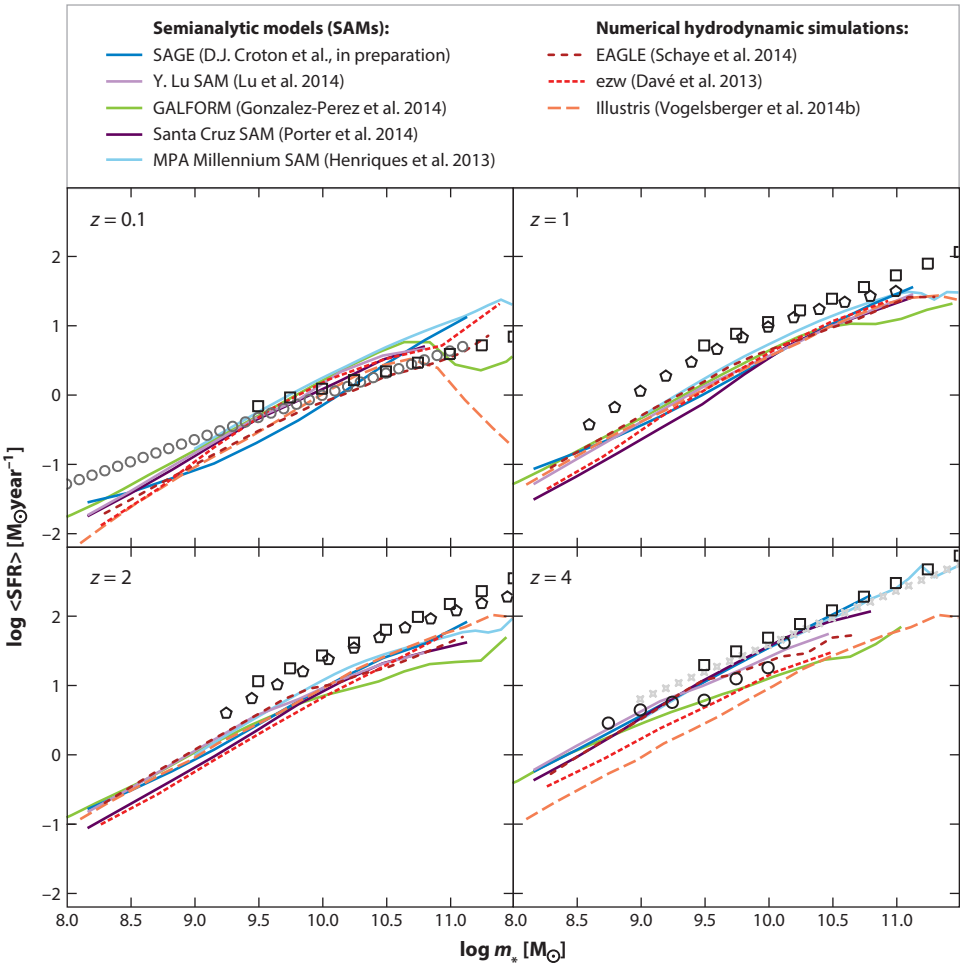

Figure 5

The average star-formation rate in bins of stellar mass for redshift bins from $z \cong 0$–4. Gray and black symbols show observational estimates: $z = 0.1$, Salim et al. (2007, *open circles*); $z = 1$ and $z = 2$, Whitaker et al. (2014, *pentagons* interpolated in redshift from the published results); $z = 4$, Steinhardt et al. (2014, *crosses*), Salmon et al. (2014, *circles*); all panels, fit to data compilation from Speagle et al. (2014, *squares*). Colored lines show predictions from semianalytic models and numerical hydrodynamic simulations; key is the same as in **Figure 4**. Note that the observational estimates shown are for star-forming galaxies; different methods have been used to isolate the star-forming sequence from quiescent galaxies. Some of the modelers have applied a cut to select star-forming galaxies, but some have not.

Moster et al. 2013). If ζ and η also evolve slowly, then sSFR should evolve as $\sim (1+z)^{2.5}$. This is indeed roughly the evolution observed out to $z \sim 2$ (e.g., Lilly et al. 2013, Speagle et al. 2014, Whitaker et al. 2014). However, to higher redshift the evolution slows, suggesting either that the $m_* - M_H$ relation or η is higher or that ζ is lower. The assumption of $\dot{M}_{\rm gas} \approx 0$ may be faulty at very early epochs if the inflowing gas cannot be processed in the ISM fast enough, which observations suggest may be the case at $z \gtrsim 4$ (Papovich et al. 2011). It had been suggested that the efficiency of converting ISM gas into stars is reduced owing to the lower metallicity at early epochs, which results in less efficient formation of molecular gas (Krumholz & Dekel 2012), but hydrodynamic

simulations and SAMs incorporating H_2 formation modeling suggest that this effect is not large enough to solve the problem at the observed mass scales (Christensen et al. 2012, Somerville et al. 2015). Moreover, the EAGLE simulations also include a metallicity-dependent density threshold for SF as proposed by Schaye (2004), representing the same physical effect, but still suffer the same problem. Hence although it is encouraging that most models are now able to predict the sSFR evolution to within a factor of 2–3 and predict roughly the right qualitative trend, such discrepancies, if real, could be pointing to a need to revise our basic understanding of the physical processes regulating SF at these epochs.

The mass dependence of the sSFR also poses interesting challenges to models. In detail, the halo mass accretion rate has a superlinear dependence on M_H, which would naïvely translate into a positive slope for sSFR(m_*). Observations indicate a subunity slope, becoming shallower with time (see **Figure 5**). Part of this may be because for $M_H \geq 10^{11}$ $M_\odot$, an increasing amount of halo inflow is gravitationally shocked into hydrostatic equilibrium (Birnboim & Dekel 2003, Gabor & Davé 2012, Kereš et al. 2005). Simulations show this is sufficient to explain the mildly negative slope in moderate-sized halos (Davé et al. 2011b, Faucher-Giguère et al. 2011) but is insufficient to explain the rapid increase in quenched galaxies at high masses, which requires additional feedback, likely associated with AGNs.

Moreover, models with outflows tuned to reproduce the observed SMF (hence m_*–M_H) predict a flat or positive slope for sSFR (m_*) at $M_H \leq 10^{11}$ $M_\odot$, whereas observations show a negative slope. The stochasticity of SF in dwarf galaxies (Tolstoy et al. 2009) may result in a duty cycle whereby observed samples preferentially select dwarfs that are in a high sSFR state, but observations show that essentially all isolated dwarfs in the nearby Universe are star forming (Geha et al. 2012). This is an aspect of the dwarf galaxy conundrum highlighted by Weinmann et al. (2012) and White et al. (2014) and discussed above, which remains a puzzle: In current models that are normalized to fit the present-day SMF, not only do dwarf galaxies form their stars too early (resulting in the low-mass excess at intermediate redshift as seen in **Figure 4**), they also have sSFRs that are well below the observed values (see also Torrey et al. 2014). At higher redshifts, the mass dependence is in good agreement with existing observations, but deeper near-IR data is needed to test if a similar discrepancy occurs in mass-selected samples of dwarfs at $z \gtrsim 1$.

Another key scaling relation is the stellar mass–gas phase metallicity relation (i.e., MZR), which can also be understood from Equation 6. Given a metal yield y per unit SF, the metallicity should be the yield times the SFR, divided by the amount of accreting gas, i.e.,

$$Z = \frac{y\dot{M}_*}{\dot{M}_{\rm inflow}} \approx \frac{y}{1+\eta}, \tag{8}$$

where this approximation is valid in the limit of no recycled (i.e., previously enriched) accretion into the ISM (Finlator & Davé 2008). Wind recycling is generally more rapid in more massive galaxies (Oppenheimer et al. 2010), which tends to make the MZR steeper. Also, outflows that are significantly more enriched than the ISM can result in a lower metallicity. Hence the MZR primarily reflects galactic outflows, modulated by wind recycling and the metal loading factor (Peeples & Shankar 2011).

Figure 6 shows the predicted MZR in our suite of SAMs and simulations compared with observations. We emphasize that, due to uncertainties in the theoretical yields of at least a factor of ∼2 and differences of ∼30% in the adopted value of solar metallicity in different simulations, the absolute normalizations of the predicted MZR should not be given as much weight as the trends with mass and redshift. We also show a recent compilation of observational estimates. Gas-phase abundance measures are sensitive to calibration (Kewley & Ellison 2008), but it is usually the case that relative abundances are more consistent among various indices. Hence the slope of the

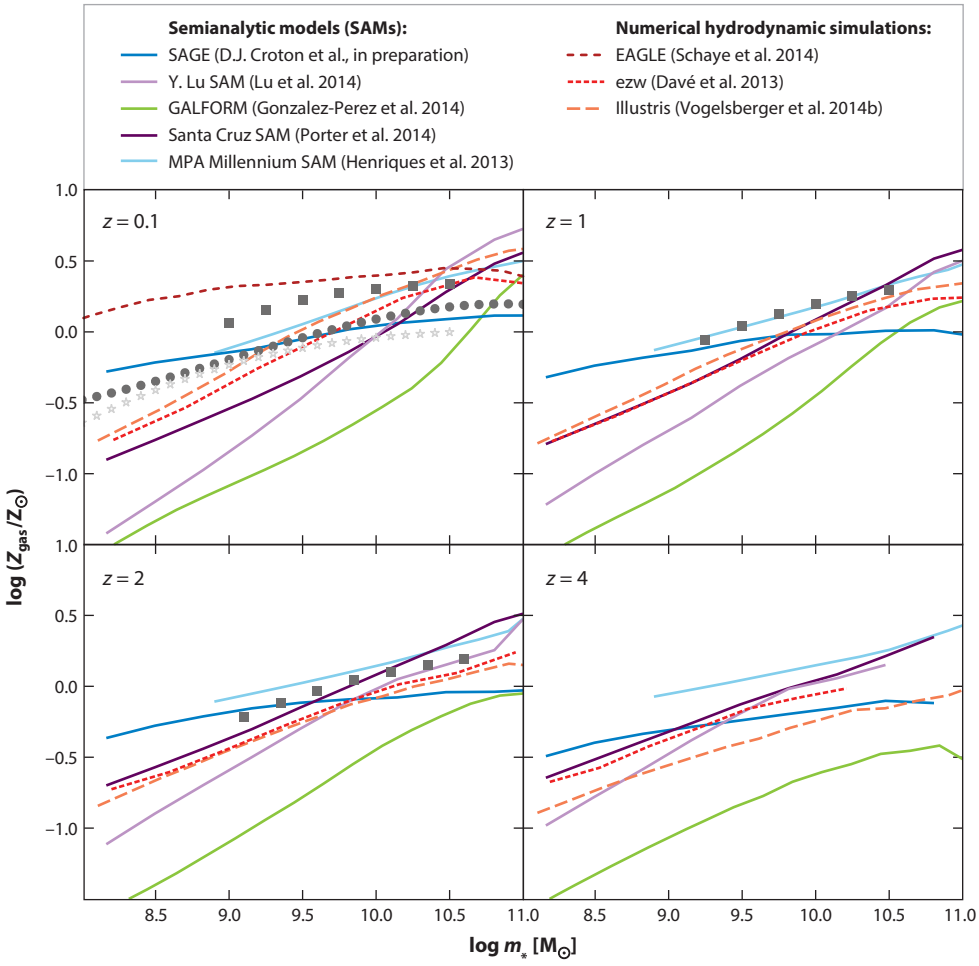

Figure 6

The average metallicity of cold gas in bins of stellar mass for redshift bins from $z \cong 0$–4. Gray and black symbols show observational estimates: $z = 0.1$, Peeples et al. (2014, *filled circles*) and Andrews & Martini (2013, *stars*). In all panels, the filled squares show the compilation by Zahid et al. (2013). Colored lines show predictions from semianalytic models and numerical hydrodynamic simulations; key is the same as in **Figure 4**.

observed MZR is more robustly known than the amplitude, though the amplitude should still be accurate to within a factor of 2–3. We show MZR determinations at $z = 0.1$, $z = 1$, and $z = 2$, converted to the same calibration, from Zahid et al. (2013). We also show the local MZR from Peeples et al. (2014), which uses the average of all the calibrations presented by Kewley & Ellison (2008), and the local direct method MZR from Andrews & Martini (2013).

At $z = 0$, most models produce roughly the correct metallicity for galaxies with stellar masses of a few $\times$ 10^{10} $M_\odot$, but predicted MZRs are typically steeper than the observed relations to low masses and have a less pronounced turnover to high masses (EAGLE, which produces a very shallow MZR, is a notable exception). To higher redshifts, models generally predict slow evolution, approximately a factor of two at a given stellar mass from $z = 2 \rightarrow 0$, which is roughly consistent with available observations.

To explore the origin of the slope discrepancy, note that Equation 8 shows that (in the absence of recycling and metal-enriched outflows), when $\eta \gg 1$ as is generally the case at low masses in these models, the observed MZR $Z \propto m_*^{0.3}$ (Tremonti et al. 2004) implies $\eta \propto m_*^{-0.3}$. When such a scaling (which is similar to the momentum-driven wind scaling) is implemented in hydrodynamic simulations, this produces good agreement with the observed MZR (Finlator & Davé 2008, Davé et al. 2011a), but the predicted SMF is somewhat too steep at the faint end (Davé et al. 2011b). Ameliorating this by incorporating a steeper mass dependence of η results in an MZR that is too steep (Davé et al. 2013). Accounting for wind recycling does not help this problem—Oppenheimer et al. (2010) highlighted the importance of wind recycling in shaping the SMF at intermediate masses, but in general wind recycling is more important at higher masses, which further steepens the MZR. In general, current simulations have difficulty simultaneously matching the low-mass ends of the SMF and the MZR, suggesting that enrichment in low-mass galaxies is not fully understood. This problem was also discussed in the context of the Illustris simulations by Torrey et al. (2014), who speculated that this tension suggests that preventative, rather than ejective, feedback is dominant in low-mass galaxies.

Cold gas scaling relations provide information on the fuel for SF. CO measurements are currently the best tracer of molecular gas content, although there remain significant uncertainties in the conversion factor from CO to H_2 (X_{CO}; Bolatto et al. 2013), particularly to higher redshifts. Observations show that low-mass galaxies are more gas rich, with $f_{gas} \propto m_*^{-0.57}$ (Peeples & Shankar 2011). Direct estimates of the H_2 fraction of galaxies to high redshift from CO and dust-based methods (corrected for selection effects) indicate a rise in $m_{H_2}/(m_{H_2} + m_*)$ back in cosmic time to $z \sim 2$, then a plateau or possibly a slight decline (Geach et al. 2011, Saintonge et al. 2013, Tacconi et al. 2013, Genzel et al. 2014, Scoville et al. 2014). Empirical estimates of H_2 and total gas fraction from extended SHAM modeling (Popping et al. 2014a) indicate a similar behavior.

Figure 7 shows a comparison of the cold gas fractions ($\equiv m_{cold}/m_*$) in models, in which we defined cold gas in the numerical simulations as that having a hydrogen number density $n_H > 0.13$ cm^{-3}. At $z = 0$, models generally reproduce the steeply rising gas fractions to low masses, though some have gas fractions significantly below the observed ones at the lowest masses. Gas fractions tend to be fairly sensitive to the prescription used to turn that gas into stars, which varies significantly among models. This generic trend of gas fraction with galaxy mass in the models arises from two physical effects that make the global SFE lower in low-mass galaxies: stronger stellar feedback (Brooks et al. 2007) and less efficient formation of H_2 (Christensen et al. 2012, Popping et al. 2014b). At a given mass, models generally predict rising gas fractions to earlier epochs, which is in broad agreement with observational and SHAM-based estimates out to $z \sim 2$, though gas fractions from the a priori models tend to be lower than the empirical SHAM predictions at $z \sim 2$–1; perhaps this is another manifestation of the dwarf galaxy conundrum discussed above. Models that track H_2 formation generally predict that galaxies become increasingly dominated by H_2 at higher masses and at high redshift (Fu et al. 2010, Lagos et al. 2011a, Popping et al. 2014b).

Atomic HI can be detected in emission in nearby galaxies and in distant galaxies via absorption. Because HI represents a transient phase of accretion from the ionized IGM to the molecular ISM, it is necessary to include both self-shielding and molecular gas formation physics in order to model it, neither of which is straightforward at typical cosmological, or even zoom, resolutions. Nonetheless, SAMs and simulations can broadly reproduce HI mass functions and scaling relations (Obreschkow et al. 2009; Popping et al. 2009, 2014b; Lagos et al. 2011a; Duffy et al. 2012; Davé et al. 2013). HI may be a particularly good tracer of environmental processes, including satellite quenching (Lagos et al. 2014, Rafieferantsoa et al. 2014), because it usually arises in the more loosely bound outskirts. In addition, SAMs and numerical simulations are being used to study the nature of HI seen in absorption (Lyman-limit and damped Lyman-α systems) and its connection with galaxies identified

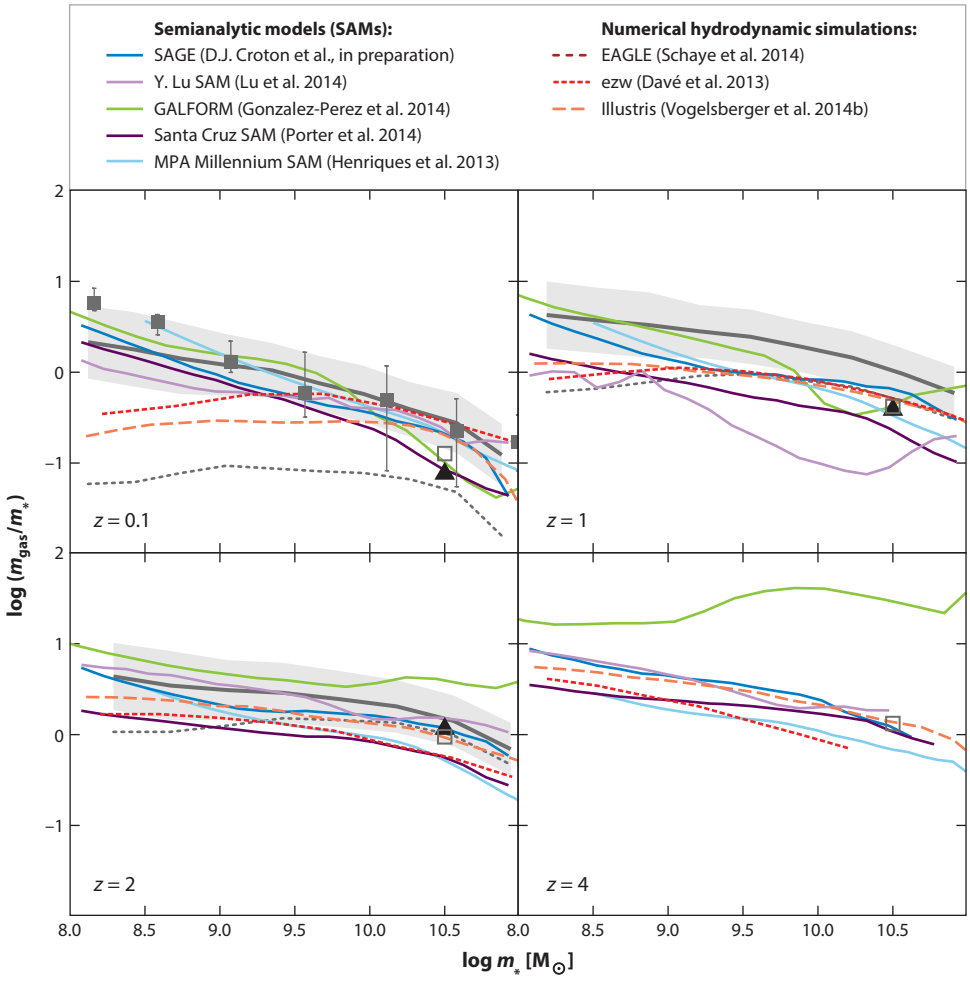

Figure 7

The average cold gas fraction in bins of stellar mass for redshift bins from $z \cong 0$–4. Gray and black lines and symbols show observational estimates: $z = 0.1$, binned results from the compilation by Peeples et al. (2014, *filled squares*). In all panels, the open squares show the predictions of the equilibrium model presented by Saintonge et al. (2013), which are in good agreement with their data compilation and with the estimates by Genzel et al. (2014, *solid triangles*), extending up to $z \sim 3$. The solid gray lines and light gray shaded areas show the empirical total cold gas fraction ($(M_{\rm HI} + M_{\rm H_2})/m_*$) estimates from Popping et al. (2014a). Dotted gray lines show the molecular fraction estimates ($M_{\rm H_2}/m_*$) from Popping et al. (2014a). Note that the $z = 0$ observational estimates shown are for HI + H_2, whereas the $z > 0$ estimates are based on CO and most closely trace H_2. Colored lines show predictions for the total cold gas fraction from semianalytic models and numerical hydrodynamic simulations; color key is the same as in **Figure 4**.

in emission (Rahmati et al. 2013, Berry et al. 2014, Bird et al. 2014, Rahmati & Schaye 2014). These studies provide important complementary constraints on disk formation and feedback processes.

So far, we have only considered mean scaling relations, which can be understood in terms of the average accretion rate into the ISM. In the accretion-driven scenario, galaxies fluctuate around the scaling relations, and the timescale to return to the mean is comparable with that required to double the mass of the galaxy. Hence the scatter of the scaling relations reflects the frequency

and efficacy of perturbing events. In particular, mergers can drive significant departures from the mean scalings. For example, galaxies that lie significantly above the SFMS are observed to have concentrated, spheroid-like (high Sérsic) light profiles (Wuyts et al. 2011), as expected if they are driven by major mergers. Reproducing the scatter in the observed scaling relations over cosmic time is a stringent challenge that models are only beginning to tackle (e.g., Sparre et al. 2014).

For the MZR, the scatter is seen to be well correlated with SFR, in the sense that galaxies at a given mass with low metallicity have high SFRs (Lara-López et al. 2010, Mannucci et al. 2010) and high HI content (Bothwell et al. 2013a, Lara-López et al. 2013). This is a natural outcome of the accretion rate fluctuation scenario, because a galaxy that undergoes an uptick in accretion increases its SFR and gas content, owing to a larger gas supply, and lowers its metallicity because the accreted gas (or infalling galaxy) tends to have lower metallicity (Davé et al. 2011a). This so-called fundamental metallicity relation has two aspects, namely this second-parameter trend and the claim by Mannucci et al. (2010) that it is invariant with redshift from $z \sim 0$–2.5. However, calibration uncertainties in metallicity measures, owing to evolving ISM conditions (Kewley et al. 2013), make the redshift independence difficult to robustly confirm, and even the existence of this second-parameter trend with SFR is not as clear at higher redshifts (Sanders et al. 2015, Steidel et al. 2014).

4.1.3. Demographics of star-forming and quiescent galaxies. The existence of quiescent galaxies, which almost entirely ceased forming stars many billions of years ago, is an additional indication of the need for some sort of quenching mechanism—processes that prevent gas from cooling and/or forming stars. Peng et al. (2010) coined the terms mass quenching and environmental quenching. Because of the strong correlations between quiescence and other galaxy internal properties (see Section 1.1), we prefer the terms internal quenching and environmental quenching. Some of the discussion here mirrors that in Section 4.1.1; however, the requirements for producing the correct internal and environmental statistical correlations for quiescent galaxies are more stringent than simply reproducing the SMF—models that reproduce the latter are not guaranteed to reproduce the former.

The massive galaxies that are predominantly early type and quiescent in the observed Universe are expected to reside within massive DM halos ($\gtrsim 10^{12}$ $M_\odot$). These halos are expected theoretically, and known observationally through X-ray observations, to be filled with hot gas at virial temperatures of a few $\times$ 10^6–10^8 K that is gravitationally shock-heated on infall and enriched to about a third of solar. Assuming hydrostatic equilibrium, this gas should be cooling fairly rapidly, at rates of hundreds to thousands of solar masses per year. The absence of the signatures of gas cooling below about one-third of the virial temperature in clusters, along with the absence of large amounts of cold gas or young stars, constitutes the classic cooling flow problem (McNamara & Nulsen 2007 and references therein). This problem has its counterpart in theoretical models in that it has proven difficult to find plausible physical mechanisms that can suppress cooling and keep galaxies hosted by massive halos as quiescent as they are observed to be.

Simulations without any sort of quenching mechanism (such as AGN feedback) produce inverted color-magnitude relations (more massive and luminous galaxies are more likely to be blue and star forming) without any hint of bimodality (Somerville et al. 2008, Gabor et al. 2011). The first generation of SAMs that included AGN feedback were able to qualitatively reproduce the observed bimodality of the color and sSFR distribution and the fraction of quiescent galaxies as a function of stellar mass (Bower et al. 2006, Croton et al. 2006, Somerville et al. 2008, Kimm et al. 2009); certainly, including AGN feedback greatly improved the results relative to the old models. In these models, the mechanism that was primarily or entirely responsible for quenching was the jet mode type of feedback described in Section 3.3.3, in which SF dies out because the hot

gas halo is continually heated so the supply of new cold gas is cut off. More heuristic models, in which cooling is simply shut off when the DM halo exceeded a certain critical mass, performed nearly as well as models that explicitly implemented jet mode AGN feedback (Cattaneo et al. 2006, Kimm et al. 2009). Some recent SAMs reproduce the observed tighter correlation of the quiescent fraction with B/T rather than with stellar mass, whereas others do not, suggesting that this could provide constraints on quenching mechanisms (Lang et al. 2014).

Springel et al. (2005a) showed that including thermal AGN feedback in hydrodynamic simulations of isolated binary mergers was able to drive powerful winds that evacuated most of the cold gas from the galaxy, leading to strong quenching of SF. These results motivated semiempirical models positing that quenching associated with mergers, rapid BH growth, and radiative mode AGN feedback could explain the growth of the quiescent early-type population (Hopkins et al. 2008a,b). However, subsequent work with SAMs and cosmologically based hydrodynamic simulations suggested that thermal feedback associated with the radiative mode of BH accretion, when implemented using algorithms similar to those of Springel et al. (2005b), is not able to produce long-lived quiescent galaxies, because it fails to prevent subsequent accretion, which reactivates SF within a few billion years (see, e.g., Gabor et al. 2011, Choi et al. 2014b).

Following the approach presented by Sijacki et al. (2007), the Illustris simulations explicitly included both local thermal heating associated with BH accretion above a critical rate (representing radiative mode) and more distributed heating associated with low BH accretion rates (representing jet mode). They produced a bimodal color-magnitude diagram, with a red galaxy fraction as a function of stellar mass and environment in good agreement with observations by Peng et al. (2010) and others (Vogelsberger et al. 2014a). The observed red-sequence colors have proven difficult to reproduce quantitatively in all types of models (Guo et al. 2011, Gabor & Davé 2012, Vogelsberger et al. 2014a), but significant uncertainties remain in this regime in the stellar population models that are used to predict such colors from models. In contrast, the stochastic thermal AGN feedback model as implemented in the EAGLE simulations does not explicitly have two distinct modes and can still reproduce quenched galaxy observations at a similar level (Schaye et al. 2015). Cosmological zoom-in simulations including fast momentum-driven AGN winds also appear to be able to quench and maintain quiescence over long timescales without any explicit jet mode type feedback (Choi et al. 2014b). Conversely, Gabor & Davé (2015) suggested that the presence of a gaseous halo kept hot by AGN feedback is sufficient to quench a galaxy, without the need for additional radiative mode feedback, showing that this reproduces both internal and environmental quenching as observed. Hence there remains much debate over the relative importance of these two AGN feedback modes, whether one or both are required, and even whether they are distinct.

Reproducing the observed patterns of environmental quenching has provided another challenge to models. Peng et al. (2012) showed that when the Sloan Digital Sky Survey galaxies were identified as "satellites" or "centrals" using a group catalog, the fraction of quiescent centrals depended only on stellar mass, whereas the fraction of quiescent satellites depended on both mass and environment. Certainly there are many candidate processes that could preferentially quench satellites, such as harassment, tidal stripping, or ram pressure stripping. In many SAMs, galaxies are not allowed to accrete any new gas from the hot halo or the IGM once they become satellites (sometimes called strangulation). This is known to produce far too high a fraction of quiescent satellites (Weinmann et al. 2006b, Font et al. 2008, Kimm et al. 2009). Instead, satellite quenching seems to take a surprisingly long time, perhaps many gigayears (Wetzel et al. 2012). Hydrodynamic simulations indeed show that infalling satellites remain star forming for at least a gigayear (Simha et al. 2009), as it takes time for the hot gas and DM from the halo in which the satellite galaxy was born to be stripped away. Including this delayed stripping of the hot gas halo, without including any other environmental effects (e.g., tidal or ram pressure stripping of the cold gas in

satellites), improves satellite statistics in SAMs (Font et al. 2008, Weinmann et al. 2010), though some tension with observations remains (Hirschmann et al. 2014). A particularly curious observational result is galaxy conformity, in which halos with red central galaxies preferentially have red satellites (Weinmann et al. 2006a). This effect extends even beyond the virial radius to surrounding centrals, and it is not reproduced at the observed level in SAMs (Kauffmann et al. 2013). It can be reproduced in age abundance matching models, an extension of abundance matching that uses halo formation times to assign SFRs or colors (Hearin et al. 2014), but the physics that drives conformity remains unclear. Satellite and environmental quenching has not yet been extensively investigated in self-consistent cosmological simulations, but there is clearly much to be learned by doing so, and this is an area in which much progress can be made in the near future.

4.2. Internal Structure and Kinematics

The internal structure and kinematics of galaxies provide perhaps even more stringent constraints on models of galaxy formation than global properties. We separately discuss the predominant theoretical ideas regarding the formation of galactic disks and spheroids and then discuss structural scaling relations and demographics of galaxies divided in terms of morphology.

4.2.1. Formation of galactic disks. The long-standing conventional paradigm to explain the origin of galactic disks posits that gas accreting from the halo conserves its specific angular momentum **j**, thereby settling into a disk (Fall & Efstathiou 1980, Mo et al. 1998). Although modern cosmological simulations support this basic paradigm, they suggest that the full story is much more complicated.

The average specific angular momentum of galactic disks is indeed comparable with that expected from conserving **j** from the halo (Dalcanton et al. 1997, Dutton & van den Bosch 2012). However, the distribution of **j** within disks predicted from simple infall is strongly inconsistent with observations, in the sense that observed galaxies have a strong deficit of low-**j** gas, a mild deficit of high-**j** gas, and a large excess of intermediate **j** gas (Bullock et al. 2001, van den Bosch et al. 2001). This suggests that some process removes low-**j** gas and deposits it at intermediate **j**.

Early numerical hydrodynamic simulations of disk galaxy formation suffered an even more severe angular momentum catastrophe, as they produced disks with much lower average **j** than the halo, indicating that a large amount of **j** was being lost during the formation process. For many years, simulations were only able to produce very compact disks with large spheroids and were unable to produce spirals even as late type as the Milky Way (Sommer-Larsen et al. 1999, Steinmetz 1999, Navarro & Steinmetz 2000). Moreover, these galaxies exhibited centrally peaked rotation curves in disagreement with observed flat rotation curves, did not lie on the observed Tully-Fisher relation, and formed far too large a fraction of their available baryons into stars. It was gradually realized that the origin of this angular momentum catastrophe lay in too-efficient SF and gas consumption in small objects at high redshift. These then assembled into low redshift galaxies via relatively gas-poor mergers, which are very efficient at dissipating angular momentum and building spheroids (Maller & Dekel 2002).

Implementing more efficient SF feedback has proven to be the key to solving all these problems (e.g., Governato et al. 2007, Guedes et al. 2011). Stellar-driven winds preferentially remove low–angular momentum gas from the centers of galaxies and deposit it in the disk outskirts after retorquing in the halo (Brook et al. 2012, Übler et al. 2014). Feedback also makes SF less efficient, keeping galaxies gas rich, which makes disks more resilient to mergers (Robertson et al. 2006a, Governato et al. 2009). Finally, the baryonic mass of small infalling satellites, particularly at early epochs, is greatly reduced, thereby mitigating early spheroid growth via merging.

Figure 8 shows a state-of-the-art high-resolution zoom-in simulation of a disk galaxy using the GASOLINE code (Christensen et al. 2012). One can see that it is now possible to form very late-type and even bulgeless galaxies (Christensen et al. 2014). The same simulations predict a $z = 0$ m_*–M_H relation in agreement with observational constraints (Munshi et al. 2013). In dwarf galaxies, the same blast wave feedback model can impulsively heat the DM, removing the central cusp generically predicted in DM simulations, thereby producing rotation curves in better agreement with observations (Governato et al. 2010, Oh et al. 2011, Pontzen & Governato 2012). As a bonus, Brooks & Zolotov (2014) have shown that destroying the central cusps in dwarf galaxies leads to enhanced tidal stripping of satellites. The combined effects of energetic stellar feedback and enhanced stripping produce satellites with internal kinematics that agree with observed dwarf spheroidal galaxies in the Local Group, plausibly resolving the too-big-to-fail problem pointed out by Boylan-Kolchin et al. (2011). Although these successes may be specific to a particular subgrid model for feedback that may or may not be fully accurate, nonetheless it suggests that it is possible to form disk galaxies with realistic properties within a ΛCDM Universe provided that the subgrid treatment of stellar feedback and the ISM possess certain key features. First, stellar feedback must be effective at keeping galaxy-wide SFEs low, and stellar winds must preferentially remove low–angular momentum material. Second, SF should occur only in very dense, highly clustered environments like those that are expected to form GMCs, not smoothly distributed over the whole disk, which helps to make stellar feedback more efficient because the SF is highly clustered as in real galaxies.

Recent observations of disks during cosmic noon have presented new challenges for models. Disks at $z \sim 2$ are observed to be substantially puffier, having rotation velocity $V_{\rm rot}$ divided by gas dispersion σ of ∼3 as opposed to ∼10 for today's disks (Förster Schreiber et al. 2009). Many $z \sim 2$ disks also have large, bright clumps that comprise a substantial portion of the disk SF (Guo et al. 2012), though significantly less of the stellar mass (Wuyts et al. 2012), and are generating outflows (Genzel et al. 2011). Although most of these objects are sufficiently massive by $z \sim 2$ to likely evolve into ellipticals today, lower-mass objects that will evolve into today's disks generally have even higher $V_{\rm rot}/\sigma$. Understanding the origin of these properties and the evolution of the population to $z = 0$ has become a major cottage industry.

As discussed earlier, ISM pressurization was introduced in simulations to stabilize disks against fragmentation (Robertson et al. 2004) and allow them to form spiral galaxies like those observed today. With the discovery of clumpy disks at $z \sim 2$ (Elmegreen & Elmegreen 2005, Förster Schreiber et al. 2009), the abundant clumps that formed in simulations without ISM pressurization began to be touted as a feature (Bournaud et al. 2007, Dekel et al. 2009a, Ceverino et al. 2010). However, such models produced overly high stellar fractions, typically $\gtrsim 50\%$, whereas abundance matching constraints suggest a value of ∼10–20% (Behroozi et al. 2010, Moster et al. 2010b, Wake et al. 2011). After implementing more efficient stellar feedback due to radiation pressure from young stars, Ceverino et al. (2014) reduced the stellar fractions by a factor of 2–3, but they are still somewhat high (Moody et al. 2014). Models with significant ISM pressurization, either imposed (Genel et al. 2012) or self-consistently generated (Faucher-Giguère et al. 2013), matched stellar fraction constraints and still produced massive clumps, but these were less prominent and quicker to disrupt.

The clump formation can be understood analytically in the context of the Toomre Q parameter (Toomre 1964), $Q \equiv c_s \Omega / \pi G \Sigma$, in which c_s is the sound speed, Ω is the angular speed, and Σ is the local surface density. If $Q < 1$, gravitational collapse can overcome shearing disruption, and the region is unstable. Clumps are regions with $Q \ll 1$ (Genzel et al. 2011), and simulations indicate that clumps are unstable regions self-regulated by gravity (Ceverino et al. 2010). The characteristic mass scale for instability is $m_{\rm clump} \lesssim 10^9$ $M_\odot$, in good agreement with observations

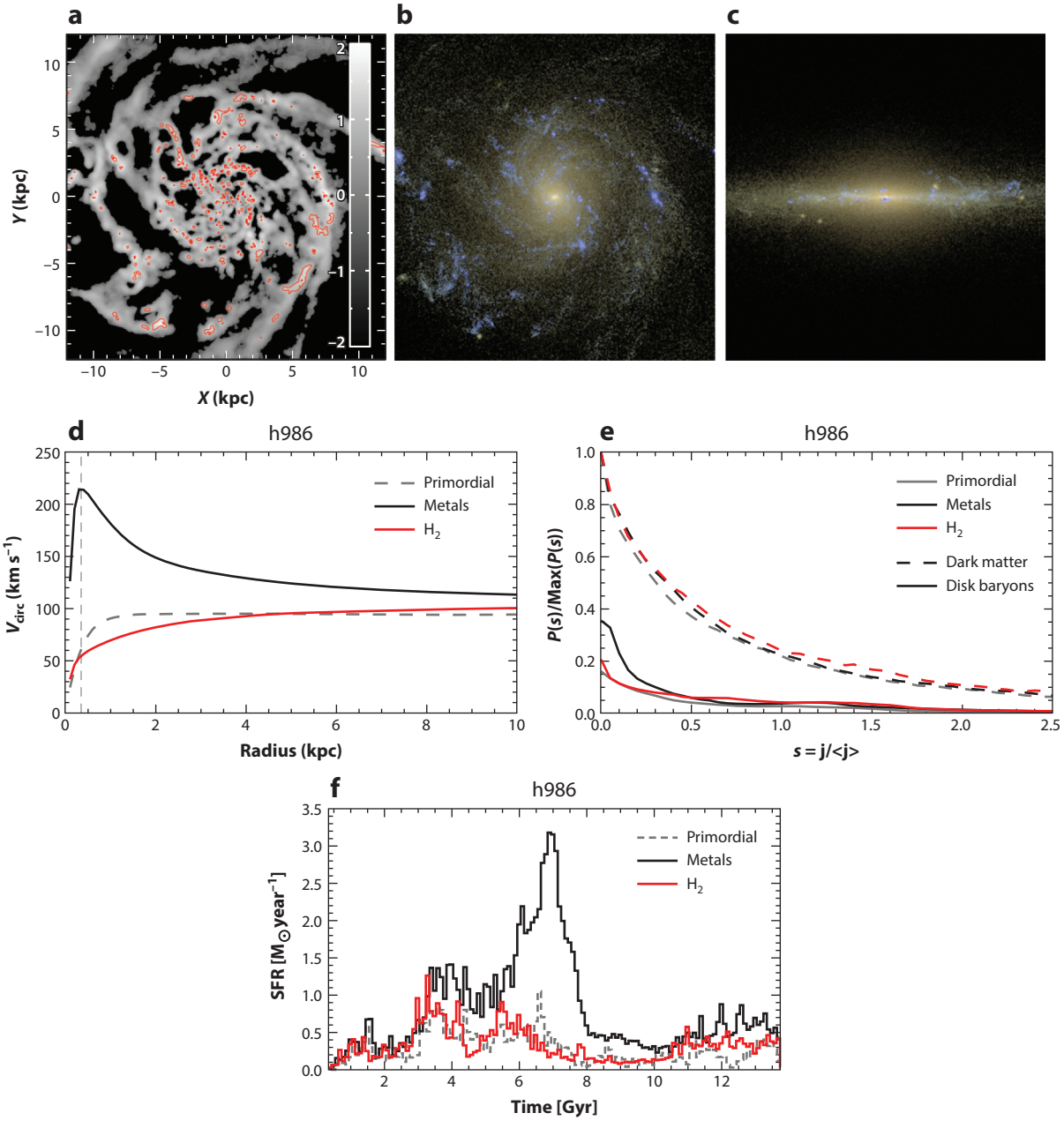

Figure 8

(*a*,*b*,*c*) Gas surface density and simulated observations of a spiral galaxy formed in a high-resolution zoom-in simulation. (*a*) HI gas surface density (*gray*) and H_2 surface density (*red*). (*b*,*c*) Optical images of stellar emission showing the galaxy in face-on and edge-on orientations. (*d*) Rotation curve for the same initial conditions but with different subgrid physics for treating cooling (primordial only versus metal-line) and the interstellar medium (H_2-based cooling and star formation). (*e*) Distribution of scaled specific angular momentum for the dark matter and baryons in these same simulations. (*f*) Star-formation histories for the same three models shown in the middle panels. Reprinted from Christensen et al. (2014) with permission.

that suggest clumps up to these masses are present in high-redshift galaxies (Genzel et al. 2011). If this is the basic origin of clumps, then they are expected to become less prominent in disks at later epochs, reducing in mass as disks settle (Dekel et al. 2009a). Simulations show that disks do indeed settle toward $z = 0$, in accord with observations (Kassin et al. 2014). If cosmological accretion drives turbulence in the ISM, then the settling may be due to the decreasing accretion rate at lower redshifts (Genel et al. 2012a), though Hopkins et al. (2013a) argue that accretion does not drive the turbulence in galactic disks. It remains to be demonstrated that a single subgrid ISM model can simultaneously reproduce the $\sim 10^9$ $M_\odot$ clumps in turbulent high-z disks along with thin disks with $\lesssim 10^6$ $M_\odot$ clumps today like the Milky Way. Nonetheless there is at least a plausible description for the evolution of clumps in disks across cosmic time.

4.2.2. Formation of spheroid-dominated galaxies. Since the seminal work by Toomre (1977), it has been recognized that nearly equal mass (major) mergers can efficiently remove angular momentum from stellar disks, producing dispersion-dominated spheroids (Toomre 1977; Barnes 1988, 1992; Hernquist 1992, 1993b; Mihos & Hernquist 1996). Unequal mass (minor) mergers down to mass ratios of ~1:10 can thicken disks and build up the spheroid component of galaxies (Walker et al. 1996, Moster et al. 2010a). Mergers are expected to be ubiquitous in the hierarchical CDM paradigm. Thus the most basic picture of the origin of the two dominant classes of galaxy morphologies is that smooth accretion of gas produces disks, and mergers destroy disks and build spheroid-dominated galaxies. A merger-driven formation mechanism for spheroid-dominated galaxies has been implemented in most SAMs since the earliest such models (Kauffmann et al. 1993, Baugh et al. 1996, Kauffmann 1996b, Somerville & Primack 1999), motivated by the studies based on binary mergers simulated with numerical hydrodynamics. These early works and others over the past decade (e.g., De Lucia et al. 2006, Parry et al. 2009) have shown that this picture can qualitatively reproduce many of the observed correlations pertaining to galaxy morphology; namely, spheroid-dominated galaxies are predicted to be more massive, to be more common in massive halos, and to have older stellar populations. In empirical support of this picture, it has been shown that the observed rate of mergers derived from pair counts and visually identified interacting galaxies is in plausible statistical agreement with the buildup of the quiescent, spheroid-dominated population (Hopkins et al. 2008a, Robaina et al. 2010).

More recent work has led to a refinement of the merger picture. Numerical hydrodynamic simulations showed that gas-rich mergers do not drive efficient angular momentum loss and therefore lead to reformation of disk-dominated galaxies (Hopkins et al. 2009a; Robertson et al. 2004, 2006a). In addition, following the formation of a spheroid via a merger, newly accreted gas can reform a disk. Thus, a picture has developed in which morphological transformation and morphological demographics are intimately linked with feedback and quenching. It is known that the massive early-type galaxies in the local Universe formed most of their stars at least 8–10 Gyr ago, around $z \sim 2$–4 (Trager et al. 2000, Thomas et al. 2005). We also know that the average massive star-forming galaxies at $z \sim 2$ are quite gas rich (Tacconi et al. 2013, Genzel et al. 2014). Thus, to produce a spheroid-dominated population at $z = 0$, some process had to consume or remove much of the gas from their progenitors before they merged, and prevent significant amounts of new gas from cooling. This appears to point qualitatively toward a combination of ejective and preventative feedback, perhaps linked with two different modes of AGN feedback.

It has been suggested that spheroids may also form and grow in situ as a result of internal gravitational instabilities. There are two different kinds of internal processes that may grow bulges; they are frequently grouped together under the term disk instabilities but they are physically quite distinct and thought to produce fundamentally different kinds of bulges. We have already discussed the formation of giant clumps in Toomre-unstable disks (see Section 4.2.1), sometimes called

violent disk instabilities (Dekel et al. 2009a). If these clumps survive and migrate to the galaxy center, they may form a classical bulge (Elmegreen et al. 2008, Dekel et al. 2009a, Bournaud et al. 2011a). However, there remains some debate about the importance of clumps in feeding spheroid growth. Simulations implementing kinetic feedback that were able to match m_*–M_H constraints suggested that clumps mostly disrupt before reaching the center (Genel et al. 2012b). Hopkins et al. (2012a) also found that in simulations of isolated disks (not cosmological) with a suite of physically motivated stellar feedback physics, even large clumps mostly blow themselves apart while in the disk, thereby only modestly contributing to spheroid growth. However, some recent simulations suggest that clumps can survive substantially longer than a disk dynamical time and grow a spheroid (Bournaud et al. 2014, Mandelker et al. 2014). Comparing stellar and SFR maps, Wuyts et al. (2012) showed that clump lifetimes are ~100–200 Myr, which would suggest disruption unless inward migration can occur on a single dynamical time or less, but radial age gradients of clumps suggest somewhat longer lifetimes (Genzel et al. 2011). Additionally, as the giant clumps orbit within the disk, even if they disrupt before reaching the center, they may drive inflows of gas into the galaxy nucleus via the same sort of physics as merger-induced nuclear inflows (Bournaud et al. 2011b).

The other process that is referred to as a disk instability is not really a (global) instability at all. It involves the secular transfer of mass into a compact, dynamically hot component via the formation of a bar (Toomre 1964, Hohl 1971, Ostriker & Peebles 1973, Combes et al. 1990). The topic of galactic bars is largely outside of the scope of this review, but a few points are worth briefly noting. First, viewed side-on, bars may be identified as boxy bulges (our Galaxy is a familiar example), but if viewed face-on these structures would not be identified as bulges (Combes et al. 1990). It is generally impossible to robustly distinguish bars from bulges in distant galaxies. Second, secular processes can redistribute angular momentum and mass within the disk, building a pseudobulge (Kormendy 2013, Kormendy & Kennicutt 2004). In contrast to the violent disk instabilities described above, the disk essentially remains in dynamical equilibrium during this secular evolution. The fundamental differences between classical bulges and pseudobulges are briefly summarized in Section 4.2, and a much more complete discussion is given by Kormendy & Kennicutt (2004). The stronger correlation between BH mass and classical bulge mass, recently emphasized by Kormendy & Ho (2013), is presumably evidence that the processes that build classical bulges (mergers and violent disk instabilities) are most closely connected with BH fueling.

4.2.3. Demographics of spheroid- and disk-dominated galaxies. Explaining the demographics of galaxies of different morphologies is another challenge for theory. Detailed quantitative statistical comparisons between the predictions of cosmological simulations and observations of galaxy morphological demographics are difficult, because up until now, most observational studies of galaxy morphology have used classifiers that are not straightforward for models to predict. SAMs predict the fraction of stellar mass or light in a spheroid component (B/T), whereas most observational studies use visual morphological classification or statistics such as Sérsic index or concentration. A few observational studies have carried out decompositions into spheroid and disk contributions (Gadotti & Kauffmann 2009, Simard et al. 2011, Bluck et al. 2014), but there are large uncertainties in these decompositions as well (see, e.g., Benson et al. 2007, Tasca & White 2011). There is a large dispersion in observational estimates of galaxy morphological demographics derived from different surveys and classification methods.

A number of studies have compared the predictions of SAMs with observational estimates of luminosity or SMFs divided by galaxy morphology or with the fraction of disk- or spheroid-dominated galaxies as a function of stellar mass (e.g., Benson et al. 2007, Parry et al. 2009, Guo et al. 2011, Porter et al. 2014). These studies all found fairly good agreement between

the predictions of these different SAMs and the observations, but interestingly the dominant mechanism that drives spheroid growth is different in different models, as we discuss further below.

Although the details of the prescriptions differ, all SAMs that attempt to track galaxy morphology assume that mergers destroy disks and build spheroids. However SAM-based studies have found, to varying degrees, that non-merger-related mechanisms for spheroid growth may be needed. The most commonly invoked alternative to mergers is a disk instability mode as described above. This is assumed to occur when the mass in the disk exceeds a critical value that depends on the angular momentum of the disk material. The implementation of this process varies widely among models, leading to significantly differing conclusions about its importance. The GALFORM SAMs assume that when a disk becomes unstable, all the stars and gas in the disk are moved to a spheroid component. They find that these disk instabilities are the dominant channel for spheroid growth except at the highest stellar masses (Parry et al. 2009). Other SAMs (De Lucia & Blaizot 2007, Guo et al. 2011, Porter et al. 2014) make a more moderate assumption, that just enough stars or stars and gas are moved from the disk to the spheroid to return the system to stability. These models find that disk instabilities are needed to reproduce the observed numbers of spheroid-dominated galaxies at intermediate masses (Porter et al. 2014) but are subdominant in driving spheroid growth at all masses. An important and apparently robust prediction is that models with a disk instability–driven channel for spheroid growth appear to form massive spheroids earlier than models in which spheroids form only via mergers (De Lucia et al. 2011, Porter et al. 2014). These predictions can now be confronted with observations from the new generation of medium-deep surveys with HST (Brennan et al. 2015).

The flip side of producing enough spheroid-dominated galaxies is the challenge of producing galaxies that are close to pure disks, which are perhaps surprisingly frequently observed in the real Universe (Kormendy et al. 2010, Fisher & Drory 2011). If mergers destroy disks and build spheroids, and nearly all halos of all masses have experienced mergers during their formation history, as predicted by ΛCDM, is it possible to reconcile the existence of these objects with the ΛCDM picture? Hopkins et al. (2009) showed that including the suppression of disk destruction in mergers with high gas fraction progenitors alleviates this problem, bringing predictions into agreement with observations in an SAM—the majority of mergers occur at high redshift when galaxy gas fractions are expected to have been fairly high. Furthermore, Moster et al. (2010a, 2012) showed that accounting for the presence of both cold gas in the disk and hot gas in the halo decreases disk heating due to minor mergers by a factor of 2–3 relative to previous calculations that included dissipationless components (stars and DM) only. However, Porter et al. (2014) showed that adding a disk instability–driven channel for spheroid formation, tuned to reproduce the abundances of spheroid-dominated galaxies, may leave behind too few objects with extremely low $B/T \lesssim 0.2$. Detailed studies with larger samples of galaxies simulated at high resolution in a full cosmological context are required to determine whether this is truly a fundamental problem for ΛCDM, but it remains a serious concern.

Extensive detailed predictions on morphological demographics from numerical cosmological simulations have not yet appeared in the literature. Such studies should be possible with the new generation of simulations, and detailed analysis of these simulations should help shed light on the physical mechanisms that are responsible for shaping galaxy morphology.

4.2.4. Structural scaling relations. The existence of structural scaling relations for galaxies, the relationship between the structure of disks and spheroids at a given mass scale, and the evolution of these relations over cosmic time encode crucial information about galaxy formation and provide stringent constraints for models.

What physics determines the internal structure of galaxies? The most basic picture is that DM and diffuse gas acquire angular momentum through tidal torques and mergers (Peebles 1969, Vitvitska et al. 2002), leading to DM and gaseous halos with a broad log-normal distribution of spin parameters. The dimensionless spin parameter is usually defined as

$$\lambda \equiv \frac{J|E|^{1/2}}{GM^{5/2}},$$

where M, J, and E are the mass, angular momentum, and total energy of the system, respectively (Mo et al. 2010, p. 502). If we assume, perhaps naïvely, that the halo gas conserves its angular momentum as it cools and collapses to form a disk, and that the postcollapse disk surface density profile has an exponential form, then the disk scale radius is given by

$$r_s = \frac{1}{\sqrt{2}} \lambda r_H F_R^{-1} F_E^{-1/2},$$

where r_H is the virial radius of the halo, and F_R and F_E are functions that account for the initial density profile of the DM halo and the contraction of the inner halo as a result of the increased gravitational force after the gas falls in (Mo et al. 1998). The rotation velocity can then be calculated by adding the contribution of the exponential disk and the contracted halo in quadrature.

In spite of its simplicity, this model does remarkably well at reproducing the size-mass relation for disk galaxies and its evolution since $z \sim 2$ (Somerville et al. 2008, Firmani & Avila-Reese 2009, Dutton et al. 2011). Recently, high-resolution numerical hydrodynamic simulations have also been quite successful at reproducing the size-mass relation for galactic disks and its evolution (Brooks et al. 2011, Aumer et al. 2013). As discussed above, hydrodynamic simulations have only recently been able to successfully reproduce disk sizes, and including strong SF driven outflows that preferentially remove the low–angular momentum material appears to be a crucial component of this success. Recent simulations suggest that accretion by cold streams may bring most of the gas into galaxies, with an average specific angular momentum that is a factor of ∼2–3 higher than that of the DM halo (Stewart et al. 2013, Kimm et al. 2011). About a factor of 2–3 of this angular momentum is then lost via torques within the misaligned disk and via outflows (Danovich et al. 2014). The success of the simple model for predicting disk sizes may therefore be simply a happy accident.

Much of the convincing evidence for the importance of mergers in producing spheroid-dominated galaxies comes from the success of merger simulations in reproducing structural properties of classical bulges and elliptical galaxies. For example, early work (Barnes 1992, Hernquist 1992) showed that mergers transform rotationally supported disks with exponential light profiles into slowly rotating remnants with luminosity profiles that are well described by an $r^{1/4}$ form over a large radial range. More recently, it has been shown that remnants of binary disk mergers lie on the observed fundamental plane (Robertson et al. 2006b, Hopkins et al. 2009b).

A striking recent observation is that, at fixed stellar mass, spheroid-dominated galaxies at $z \sim 2$ have much smaller sizes and central densities higher by orders of magnitude compared with today's (e.g., Trujillo et al. 2006; van Dokkum et al. 2008, 2014; Barro et al. 2013; van der Wel et al. 2014). For dissipationless (dry) mergers on parabolic orbits (hence with small orbital energy), energy conservation and the virial theorem can be used to show that, given a progenitor mass ratio $\eta \leq 1$ and ratio of squares of their velocity dispersions of $\epsilon \leq 1$, the ratio of final to initial radius is given by

$$\frac{r_f}{r_i} = \frac{(1+\eta)^2}{1+\epsilon\eta} \tag{9}$$

(Naab et al. 2009). For a 1:1 merger, $\eta = \epsilon = 1$, hence $r_f/r_i = 2$, which leads to a modest surface density reduction of a factor of four. One can show that, for a given total mass increase, the size is increased much more by a series of minor mergers than by a single major one. Numerical simulations confirm such a size increase in dissipationless mergers, which generally move galaxies along the mass-size relation (Boylan-Kolchin et al. 2005). This can reproduce the observed size increase and central density reduction (Naab et al. 2009, Oser et al. 2012) since $z \sim 2$ for cosmologically plausible merger histories (Gabor & Davé 2012), via minor mergers depositing material predominantly in the outskirts.

If disks (star-forming galaxies) are continuously being transformed into spheroids (quiescent galaxies), as the demographic observations indicate (see Section 1.1), how then can we understand the very different slopes and evolution of the size-mass relationship for disks and spheroids? Several recent works have pointed out that accounting for the effects of dissipation in gas-rich mergers can lead to important changes in the scaling relations (Covington et al. 2008, 2011; Hopkins et al. 2009b, 2010; Shankar et al. 2010, 2013; Porter et al. 2014). In the presence of gas, energy is dissipated, which can lead to merger remnants that are smaller and denser than their progenitors. Porter et al. (2014) implemented self-consistently within the Santa Cruz SAM a recipe for computing spheroid sizes and velocity dispersions, based on a simple analytic model that includes the effects of gas dissipation. They tested their model and derived form factors based on a suite of binary merger simulations. They showed that without any tuning of free parameters the model predicts rapid size evolution of spheroid-dominated galaxies since $z \sim 2$, along with weaker evolution in the Faber-Jackson relation, in very good quantitative agreement with the observed structural relations.

In this picture, dissipation plays a major role in explaining the different slope, scatter, and evolution of the size-mass relation for spheroid-dominated (quiescent) galaxies relative to disks. Lower-mass spheroids have lower-mass progenitors, which have higher gas fractions at all redshifts. More gas means more dissipation and smaller remnants, thus a steeper size-mass relation. Progenitors at higher redshifts have higher gas fractions than those at lower redshift, so the size-mass relation for spheroids tilts away from that for disks more, contributing to more rapid size evolution especially for the lower-mass spheroids. The decrease in scatter occurs because disks with higher angular momentum have larger radii and lower gas densities, resulting in less efficient SF. These large radius disks therefore end up with higher gas fractions and experience more dissipation when they merge, producing smaller remnants. Similarly, the observed tilt of the fundamental plane can be explained by the expected trends in galaxy gas content with mass and redshift and the physics of gas dissipation in mergers (Hopkins et al. 2009b, Covington et al. 2011, Porter et al. 2014).

5. SUMMARY AND OUTLOOK

Galaxy formation models set within the hierarchical CDM paradigm have made remarkable progress over the past decade. In this review, we focused on the methods and phenomenology of models that attempt to track astrophysical processes and predict galaxy properties within a cosmological framework. We identified a set of key observations that current models strive to reproduce—models which describe the assembly of galaxies from cosmic noon ($z \sim 2$–3) to the present. These observations include distribution functions of global properties, such as SMFs, and global scaling relations, such as those between stellar mass and SFR, gas fraction, and ISM metallicity. In addition, observations are now starting to provide measurements of galaxy demographics—how the breakdown of the galaxy population in terms of star-forming and quiescent, and in terms of disk- and spheroid-dominated objects has evolved over this time period. The observed

relationships between global and structural properties (such as light profile shape, size or internal density, and kinematics) and their evolution provide even stronger constraints on models. We described how well current state-of-the-art galaxy formation models are able to reproduce these observations and what we have learned from their successes and failures about the physics of galaxy formation.

Although many discrepancies with observations remain, overall we would give today's suite of galaxy formation models a passing grade. Summarizing the scorecard we have discussed in detail in this review:

- Qualitatively, hierarchical models correctly predict the buildup of stellar mass over cosmic time, with massive galaxies forming earlier and more rapidly than low-mass galaxies. Quantitatively, most models agree with observed galaxy number densities from $z \sim 4$–0 at least at the factor of 2–3 level. However, models tend to predict that galaxies have nearly self-similar SF histories, whereas observations imply a stronger mass dependence for these histories (sometimes known as downsizing). This is part of a set of linked discrepancies connected with low-mass galaxies that we termed the dwarf galaxy conundrum, which remains an open puzzle for models.
- Models predict qualitatively the right slope and evolution of mean global scaling relations between stellar mass and SFR, gas fraction, or gas phase metallicity. These linked correlations can be understood at the most basic level via a very simple flow model describing an approximate equilibrium between galactic inflows and outflows. Quantitatively, compared with our current observational estimates, models tend to predict a star-forming main sequence and MZR that are too steep, and possibly gas fractions that are too low at intermediate redshift, in galaxies with low stellar masses. These are additional symptoms of the dwarf galaxy conundrum mentioned above. Also, models have difficulty reproducing the observed redshift dependence of the sSFR at any mass, indicating that real galaxies deviate from the simple equilibrium model.
- Models can qualitatively explain the existence of two basic morphological types, disks and spheroids, via two different assembly modes. Disks are formed via smooth accretion of diffuse gas, which largely conserves its angular momentum, whereas spheroids are formed via gas-poor mergers that efficiently transfer angular momentum. Recently, numerical simulations demonstrated the ability to form pure disks in at least some cases—this is a major achievement because previous generations of simulations were only able to form spheroid-dominated galaxies. The strong feedback in such models also results in rotation curves and Local Group satellite demographics in better agreement with observations, which had previously been identified as a fundamental challenge to cold and collisionless DM. However, it is still unclear how well models match observed morphological demographics and their evolution in detail. There is still much debate about how efficiently mergers can build spheroids, how this depends on the parameters of the merger and the gas fraction of the progenitors, and the role of other processes such as secular evolution and violent disk instabilities.
- Models predict the correct qualitative trends between stellar population demographics (the fraction of SF and quiescent galaxies) and internal properties such as stellar mass: Galaxies with higher stellar masses, higher spheroid fractions, and higher central densities have a higher probability of being quiescent. Some models correctly reproduce the dependence of quiescent fraction on environmental parameters such as large-scale density as well, but the physics specific to the quenching of satellite galaxies remains imperfectly understood. Quantitatively, models still have difficulty reproducing observed color or sSFR distributions in detail. Models have not yet extensively confronted the emerging measurements of stellar population demographics at high redshift.

- Many models are able to at least qualitatively reproduce the observed sizes and internal velocities of observed galaxies and scaling relations such as the Kormendy (size-mass), Tully-Fisher (mass-velocity), and fundamental plane relations. Reproducing observed disk sizes in numerical simulations has been a multidecade struggle, and the solution has emerged through a combination of greatly increased resolution, more physical treatment of the ISM, and the effective implementation of stellar winds that preferentially remove low–angular momentum gas. Correctly reproducing the structural scaling relations and their evolution for both disks and spheroids, as well as the correct overall evolution of the number densities of these two populations, remains an open challenge for models.

Although there remains a wide range of models, and a healthy diversity of computational methods, virtually all models implement a qualitatively similar set of core physical processes. Although it is possible that all models are being led down the garden path due to their reliance on phenomenology, the concordance among models using different methods is encouraging, and strongly suggests that we are making fundamental progress in at least identifying the main physical players involved. Some of the core processes identified include the prevalence of cold smooth accretion in building disks and fueling star-forming galaxies, the ubiquity and efficiency of SF-driven outflows, the importance of BH-related feedback in quenching SF in massive galaxies, merger-driven morphological evolution that depends on the gas content of progenitors, and various physical processes that uniquely impact satellite galaxies once they fall into a larger halo containing hot gas. In addition, the convergence toward a similar qualitative view of the types of processes that are needed in different circumstances, based on more empirical considerations (e.g., preventative versus ejective feedback, internal versus environmental quenching, etc.), is also encouraging.

Many of these processes connect stellar scales to cosmological scales, making ab initio modeling nearly impossible and forcing models to rely on phenomenological prescriptions to describe subgrid physics, which must be calibrated in some way by observations. It is clear that many model results are sensitive to the details of these subgrid recipes and their implementation, leading to a valid concern that these models may have little genuine predictive power (Haas et al. 2013a,b). There are perhaps two ways to combat this concern. First, although the subgrid recipes and their parameters are tuned to match a subset of observations, the current suite of available observations is diverse and rich enough that, by confronting models with as wide as possible a set of complementary constraints and by exploring different subgrid recipes and implementations, one can isolate the approach that satisfies the broadest set of constraints. Second, by studying small-scale simulations (for example, of the ISM and the formation of individual stars or regions near SMBHs), one may hope to place the subgrid recipes used in our cosmological simulations on a physically grounded foundation. Zoom techniques are now enabling simulations that are starting to bridge the gap between the scales of individual stars and SMBHs and galactic scales. Although it will not be feasible to simulate cosmological volumes with these techniques in the near future, they will allow us to learn much about the interface between the microscales of stars and BHs and the macroscales of galaxies.

In addition, there are physical processes that may be important in regulating galaxy formation but which are not commonly included in current mainstream models. These include turbulence, magnetic fields, cosmic rays, and self-consistent radiative transfer. It is important to carry out experiments to determine the importance of these processes in shaping the observable properties of galaxies, and there has been significant recent progress on this front as well (e.g., Scannapieco & Brüggen 2010, Kotarba et al. 2011, Wise & Abel 2011, Mendygral et al. 2012, Hanasz et al. 2013, Pfrommer 2013).

Ideally, we would like to obtain direct observational confirmation (or refutation) of the set of core processes that models currently invoke. However, in many cases this is challenging. Smooth

gas accretion (i.e., in small enough lumps that adiabatically add to the fuel supply without disrupting galactic structure) is expected to be very diffuse and in a phase that is difficult ($T \sim 10^4$ K) to nearly impossible ($T \sim 10^5$ K) to detect. The key parameter characterizing outflow efficiency in models is the mass-loss rate, but because outflows are highly multiphasic it is difficult to account for all the mass (Veilleux et al. 2005). We observe the signatures of BH activity in the form of AGNs and jets associated with massive galaxies, but it is difficult to observationally constrain how efficiently this energy couples to surrounding gas to enact quenching. We can observe signposts and signatures of mergers in the form of close pairs and morphologically disturbed galaxies, but their rate is difficult to quantify precisely and their effect is difficult to directly constrain observationally. We can measure the statistics of galaxies in different environments, but it has been difficult with existing samples to disentangle the correlations between environment and internal properties and to locate the environments at high redshift that are the progenitors of typical groups and clusters in the local Universe.

However, there are several important observational developments taking place now, or on the horizon, that will challenge and help to refine our models of galaxy formation. First, a new generation of submillimeter and radio interferometers [including ALMA (Atacama Large Millimeter/Submillimeter Array), NOEMA (Northern Extended Millimeter Array), JVLA (Jansky Very Large Array), Apertif, ASKAP (Australian Square Kilometre Array Pathfinder), MeerKAT, and the SKA (Square Kilometer Array)] will literally revolutionize our ability to characterize the cold gas in the ISM of galaxies out to high redshifts (Carilli & Walter 2013). Second, high-resolution spectroscopy in the rest-frame UV is now able to probe the diffuse gas and metals in the circumgalactic medium of galaxies for galaxy-targeted sightlines spanning a diverse range of galaxy types, from nearby galaxies to $z \sim 2$–3 (e.g., Rudie et al. 2012, Prochaska et al. 2013, Tumlinson et al. 2013, Peeples et al. 2014). This provides constraints on the gas and metals that have been ejected by the winds invoked by our models, which probably comprise a much larger fraction of the halo baryon budget than the stars and cold ISM within galaxies. Third, integral field unit spectrographs on ground-based telescopes and on *James Webb Space Telescope* will allow us to better characterize stellar and AGN-driven winds and to study spatially resolved stellar population parameters and kinematics for large samples of nearby and high-redshift galaxies. Finally, high-resolution, wide-field multiwavelength imaging, such as will be possible with Wide-Field Infrared Survey Telescope, will enable us to study galaxy internal properties and demographics over a much larger range of environments, allowing us to better disentangle internal and environmental forces and to accumulate better statistics for rare events, such as mergers and luminous AGNs.

We thus live in interesting times in which modelers are now offering some specific and nontrivial challenges to observers to go out and confirm, or rule out, key physical processes. Just because a given mechanism is not observed does not mean it is not occurring; one must carefully assess whether that mechanism is expected to be observable. A general trend is that models make the most direct predictions about gas-related processes, particularly inflows and outflows in the baryon cycle, with the growth of stellar and BH components being almost a side effect. Hence, in principle, observations that directly trace gas processes offer the greatest potential for new advances and constraints. Modelers and observers must work together to identify key tests that can be conducted with present and upcoming facilities to constrain the core physical processes. The emerging interplay between galaxy formation models and state-of-the-art telescopes is the hallmark of a healthy and vibrant area of research.

The way forward for galaxy formation models is fairly clear but immensely challenging. As a blueprint, consider the Lyα forest: Several decades ago, studying the interplay of gas dynamics with cosmological structure formation led to a revolution in our understanding that eventually resulted in the Lyα forest becoming a pillar of precision cosmology. Our goal should be to equivalently turn

galaxy formation into a precision field, in which parameterized recipes are tied to the physics of small-scale processes in such a way that the parameters no longer need to be empirically tuned but are constrained by our physical understanding of those processes (e.g., stellar evolution models, or BH accretion disk models). Numerical simulations on different scales (zooms and cosmological volumes) and SAMs have crucial and complementary roles to play in this process, helping to better understand the physics in detail as well as synthesize and parameterize it within a ΛCDM context. It is almost surely the case that the physical processes included in models so far will not be sufficient to fully describe galaxy evolution, and there will be many twists and surprises forthcoming. Hence there is much work to be done, but it appears that cosmological models of galaxy formation are on a secure foundation for the exciting journey ahead.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

It would take many more pages to thank all the colleagues who have provided valuable insights and participated in discussions that have shaped this work, and we apologize for the inevitable choices we had to make to review this vast topic while conforming to page limits. But we would particularly like to thank Andrew Benson, Richard Bower, Rob Crain, Darren Croton, Michelle Furlong, Violeta Gonzalez-Perez, Bruno Henriques, Yu Lu, Joop Schaye, Paul Torrey, Mark Vogelsberger, and their collaborators for providing the data from their models and simulations and for constructive comments on this review. We also thank Avishai Dekel, Thorsten Naab, and Gergö Popping for comments. We especially thank our Scientific Editor, John Kormendy, for his thorough reading of the paper, and for comments and suggestions that improved the review. R.S.S. gratefully acknowledges the generous support of the Downsbrough family. R.D. acknowledges support from the South African Research Chairs Initiative and the South African National Research Foundation. This work was supported in part by NASA grant NNX12AH86G.

LITERATURE CITED

Agertz O, Kravtsov AV. 2014. arXiv:1404.2613
Agertz O, Moore B, Stadel J, et al. 2007. *MNRAS* 380:963
Andrews BH, Martini P. 2013. *Ap. J.* 765:140
Anglés-Alcázar D, Özel F, Davé R. 2013. *Ap. J.* 770:5
Anglés-Alcázar D, Özel F, Davé R, Oppenheimer BD. 2014. *MNRAS* 782:84
Arrigoni M, Trager SC, Somerville RS, Gibson BK. 2010. *MNRAS* 402:173
Aumer M, White SDM, Naab T, Scannapieco C. 2013. *MNRAS* 434:3142
Avila-Reese V, Firmani C, Hernández X. 1998. *Ap. J.* 505:37
Babul A, Sharma P, Reynolds CS. 2013. *Ap. J.* 768:11
Bagla JS. 2005. *Curr. Sci.* 88:1088
Baldry IK, Glazebrook K, Brinkmann J, et al. 2004. *Ap. J.* 600:681
Baldry IK, Glazebrook K, Driver SP. 2008. *MNRAS* 388:945
Balogh ML, Baldry IK, Nichol R, et al. 2004. *Ap. J. Lett.* 615:L101
Barnes J, Hut P. 1986. *Nature* 324:446
Barnes JE. 1988. *Ap. J.* 331:699
Barnes JE. 1992. *Ap. J.* 393:484

Barro G, Faber SM, Pérez-González PG, et al. 2013. *Ap. J.* 765:104
Bastian N, Covey KR, Meyer MR. 2010. *Annu. Rev. Astron. Astrophys.* 48:339
Baugh CM. 2006. *Rep. Prog. Phys.* 69:3101
Baugh CM, Cole S, Frenk CS. 1996. *MNRAS* 283:1361
Baugh CM, Lacey CG, Frenk CS, et al. 2005. *MNRAS* 356:1191
Behroozi PS, Conroy C, Wechsler RH. 2010. *Ap. J.* 717:379
Behroozi PS, Wechsler RH, Conroy C. 2013a. *Ap. J.* 770:57
Behroozi PS, Wechsler RH, Wu HY, et al. 2013b. *Ap. J.* 763:18
Bell EF, van der Wel A, Papovich C, et al. 2012. *Ap. J.* 753:167
Bell EF, Wolf C, Meisenheimer K, et al. 2004. *Ap. J.* 608:752
Bell EF, Zheng XZ, Papovich C, et al. 2007. *Ap. J.* 663:834
Bender R, Burstein D, Faber SM. 1992. *Ap. J.* 399:462
Benson AJ. 2010. *Phys. Rep.* 495:33
Benson AJ, Dzanovic D, Frenk CS, Sharples R. 2007. *MNRAS* 379:841
Benson AJ, Pearce FR, Frenk CS, Baugh CM, Jenkins A. 2001. *MNRAS* 320:261
Berlind AA, Weinberg DH. 2002. *Ap. J.* 575:587
Bernardi M, Meert A, Sheth RK, et al. 2013. *MNRAS* 436:697
Bernardi M, Shankar F, Hyde JB, et al. 2010. *MNRAS* 404:2087
Berry M, Somerville RS, Haas MR, et al. 2014. *MNRAS* 441:939
Bertschinger E. 1998. *Annu. Rev. Astron. Astrophys.* 36:599
Bigiel F, Leroy A, Walter F, et al. 2008. *Astron. J.* 136:2846
Bigiel F, Leroy AK, Walter F, et al. 2011. *Ap. J. Lett.* 730:L13
Binggeli B, Sandage A, Tammann GA. 1988. *Annu. Rev. Astron. Astrophys.* 26:509
Bird S, Vogelsberger M, Haehnelt M, et al. 2014. *MNRAS* 445:2313
Birnboim Y, Dekel A. 2003. *MNRAS* 345:349
Birnboim Y, Dekel A. 2011. *MNRAS* 415:2566
Blandford RD, Begelman MC. 1999. *MNRAS* 303:1
Blanton MR, Moustakas J. 2009. *Annu. Rev. Astron. Astrophys.* 47:159
Blitz L, Rosolowsky E. 2004. *Ap. J. Lett.* 612:L29
Bluck AFL, Mendel JT, Ellison SL, et al. 2014. *MNRAS* 441:599
Blumenthal GR, Faber SM, Primack JR, Rees M. 1984. *Nature* 311:517
Bolatto AD, Wolfire M, Leroy AK. 2013. *Annu. Rev. Astron. Astrophys.* 51:207
Bondi H. 1952. *MNRAS* 112:195
Booth CM, Schaye J. 2009. *MNRAS* 398:53
Bothwell MS, Maiolino R, Kennicutt R, et al. 2013a. *MNRAS* 433:1425
Bothwell MS, Smail I, Chapman SC, et al. 2013b. *MNRAS* 429:3047
Bournaud F, Chapon D, Teyssier R, et al. 2011a. *Ap. J.* 730:4
Bournaud F, Dekel A, Teyssier R, et al. 2011b. *Ap. J. Lett.* 741:L33
Bournaud F, Elmegreen BG, Elmegreen DM. 2007. *Ap. J.* 670:237
Bournaud F, Elmegreen BG, Teyssier R, Block DL, Puerari I. 2010. *MNRAS* 409:1088
Bournaud F, Perret V, Renaud F, et al. 2014. *Ap. J.* 780:57
Bower RG, Benson AJ, Malbon R, et al. 2006. *MNRAS* 370:645
Bower RG, Vernon I, Goldstein M, et al. 2010. *MNRAS* 407:2017
Boylan-Kolchin M, Bullock JS, Kaplinghat M. 2011. *MNRAS* 415:40
Boylan-Kolchin M, Ma CP, Quataert E. 2005. *MNRAS* 362:184
Boylan-Kolchin M, Ma CP, Quataert E. 2008. *MNRAS* 383:93
Brammer GB, Whitaker KE, van Dokkum PG, et al. 2011. *Ap. J.* 739:24
Brennan R, Pandya V, Somerville RS, et al. 2015. *MNRAS.* Accepted. arXiv:1501.06840
Brinchmann J, Charlot S, White SDM, et al. 2004. *MNRAS* 351:1151
Brook CB, Stinson G, Gibson BK, et al. 2012. *MNRAS* 419:771
Brooks AM, Governato F, Booth CM, et al. 2007. *Ap. J. Lett.* 655:L17
Brooks AM, Solomon AR, Governato F, et al. 2011. *Ap. J.* 728:51
Brooks AM, Zolotov A. 2014. *Ap. J.* 786:87

Bryan GL, Norman ML. 1998. *Ap. J.* 495:80
Bryan GL, Norman ML, O'Shea BW, et al. 2014. *Ap. J. Suppl.* 211:19
Bullock JS, Dekel A, Kolatt TS, et al. 2001. *Ap. J.* 555:240
Burstein D, Bender R, Faber S, Nolthenius R. 1997. *Astron. J.* 114:1365
Buta RJ. 2013. In *Secular Evolution of Galaxies*, ed. J Falcón-Barroso, JH Knapen, pp. 155–258. Cambridge, UK: Cambridge Univ. Press
Cappellari M, di Serego Alighieri S, Cimatti A, et al. 2009. *Ap. J. Lett.* 704:L34
Caputi KI, Cirasuolo M, Dunlop JS, et al. 2011. *MNRAS* 413:162
Carilli CL, Walter F. 2013. *Annu. Rev. Astron. Astrophys.* 51:105
Cattaneo A, Dekel A, Devriendt J, Guiderdoni B, Blaizot J. 2006. *MNRAS* 370:1651
Cenarro AJ, Trujillo I. 2009. *Ap. J. Lett.* 696:L43
Ceverino D, Dekel A, Bournaud F. 2010. *MNRAS* 404:2151
Ceverino D, Klypin A, Klimek ES, et al. 2014. *MNRAS* 442:1545
Cheung E, Faber SM, Koo DC, et al. 2012. *Ap. J.* 760:131
Choi E, Naab T, Ostriker JP, Johansson PH, Moster BP. 2014a. *MNRAS* 442:440
Choi E, Ostriker JP, Naab T, Johansson PH. 2012. *Ap. J.* 754:125
Choi E, Ostriker JP, Naab T, Oser L, Moster BP. 2014b. *MNRAS* 442:440
Christensen C, Quinn T, Governato F, et al. 2012. *MNRAS* 425:3058
Christensen CR, Governato F, Quinn T, et al. 2014. *MNRAS* 440:2843
Cielo S, Antonuccio-Delogu V, Macciò AV, Romeo AD, Silk J. 2014. *MNRAS* 439:2903
Cimatti A, Daddi E, Renzini A. 2006. *Astron. Astrophys.* 453:29
Cole S, Aragón-Salamanca A, Frenk CS, Navarro JF, Zepf SE. 1994. *MNRAS* 271:781
Cole S, Lacey CG, Baugh CM, Frenk CS. 2000. *MNRAS* 319:168
Combes F, Debbasch F, Friedli D, Pfenniger D. 1990. *Astron. Astrophys.* 233:82
Conroy C. 2013. *Annu. Rev. Astron. Astrophys.* 51:393
Conroy C, Wechsler RH. 2009. *Ap. J.* 696:620
Conroy C, Wechsler RH, Kravtsov AV. 2006. *Ap. J.* 647:201
Conselice CJ. 2014. *Annu. Rev. Astron. Astrophys.* 52:291–337
Courteau S, Dutton AA, van den Bosch FC, et al. 2007. *Ap. J.* 671:203
Covington M, Dekel A, Cox TJ, Jonsson P, Primack JR. 2008. *MNRAS* 384:94
Covington MD, Primack JR, Porter LA, et al. 2011. *MNRAS* 415:3135
Cox TJ, Jonsson P, Primack JR, Somerville RS. 2006. *MNRAS* 373:1013
Cox TJ, Jonsson P, Somerville RS, Primack JR, Dekel A. 2008. *MNRAS* 384:386
Croton DJ, Springel V, White SDM, et al. 2006. *MNRAS* 365:11
Daddi E, Dickinson M, Morrison G, et al. 2007. *Ap. J.* 670:156
Dalcanton J, Spergel D, Summers F. 1997. *Ap. J.* 482:659
Dalla Vecchia C, Schaye J. 2008. *MNRAS* 387:1431
Dalla Vecchia C, Schaye J. 2012. *MNRAS* 426:140
Danovich M, Dekel A, Hahn O, Ceverino D, Primack J. 2014. arXiv:1407.7129
Davé R. 2008. *MNRAS* 385:147
Davé R, Finlator K, Oppenheimer BD. 2011a. *MNRAS* 416:1354
Davé R, Finlator K, Oppenheimer BD. 2012. *MNRAS* 421:98
Davé R, Katz N, Oppenheimer BD, Kollmeier JA, Weinberg DH. 2013. *MNRAS* 434:2645
Davé R, Oppenheimer BD, Finlator K. 2011b. *MNRAS* 415:11
Debuhr J, Quataert E, Ma CP. 2011. *MNRAS* 412:1341
Debuhr J, Quataert E, Ma CP, Hopkins P. 2010. *MNRAS* 406:55
Dehnen W, Read JI. 2011. *Eur. Phys. J. Plus* 126:55
Dekel A, Birnboim Y, Engel G, et al. 2009a. *Nature* 457:451
Dekel A, Mandelker N. 2014. *MNRAS* 444:2071
Dekel A, Sari R, Ceverino D. 2009b. *Ap. J.* 703:785
Dekel A, Silk J. 1986. *Ap. J.* 303:39
De Lucia G, Blaizot J. 2007. *MNRAS* 375:2
De Lucia G, Fontanot F, Wilman D, Monaco P. 2011. *MNRAS* 414:1439

De Lucia G, Springel V, White SDM, Croton D, Kauffmann G. 2006. *MNRAS* 366:499
Di Matteo T, Springel V, Hernquist L. 2005. *Nature* 433:604
Djorgovski S, Davis M. 1987. *Ap. J.* 313:59
Dobbs CL, Krumholz MR, Ballesteros-Paredes J, et al. 2014. In *Protostars and Planets VI*, ed. H Beuther, R Klessen, C Dullemond, T Henning pp. 3–26. Tucson: Univ. Ariz. Press. arXiv:1312.3223
Duffy AR, Kay ST, Battye RA, et al. 2012. *MNRAS* 420:2799
Duncan K, Conselice CJ, Mortlock A, et al. 2014. *MNRAS* 444:2960
Dutton AA, van den Bosch FC. 2009. *MNRAS* 396:141
Dutton AA, van den Bosch FC. 2012. *MNRAS* 421:608
Dutton AA, van den Bosch FC, Faber SM, et al. 2011. *MNRAS* 410:1660
Eisenstein DJ, Hu W. 1999. *Ap. J.* 511:5
Elmegreen BG. 1989. *Ap. J.* 338:178
Elmegreen BG, Bournaud F, Elmegreen DM. 2008. *Ap. J.* 688:67
Elmegreen BG, Elmegreen DM. 2005. *Ap. J.* 627:632
Erb DK, Shapley AE, Pettini M, et al. 2006. *Ap. J.* 644:813
Faber SM, Dressler A, Davies RL, Burstein D, Lynden-Bell D. 1987. In *Nearly Normal Galaxies. From the Planck Time to the Present*, ed. SM Faber, p. 175. New York: Springer
Faber SM, Jackson RE. 1976. *Ap. J.* 204:668
Faber SM, Willmer CNA, Wolf C, et al. 2007. *Ap. J.* 665:265
Fabian AC. 2012. *Annu. Rev. Astron. Astrophys.* 50:455
Fall SM, Efstathiou G. 1980. *MNRAS* 193:189
Fanidakis N, Baugh CM, Benson AJ, et al. 2011. *MNRAS* 410:53
Faucher-Giguère CA, Kereš D, Ma CP. 2011. *MNRAS* 417:2982
Faucher-Giguère CA, Lidz A, Zaldarriaga M, Hernquist L. 2009. *Ap. J.* 703:1416
Faucher-Giguère CA, Quataert E. 2012. *MNRAS* 425:605
Faucher-Giguère CA, Quataert E, Hopkins PF. 2013. *MNRAS* 433:1970
Finlator K, Davé R. 2008. *MNRAS* 385:2181
Finlator K, Davé R, Özel F. 2011. *Ap. J.* 743:169
Finlator K, Oh SP, Özel F, Davé R. 2012. *MNRAS* 427:2464
Firmani C, Avila-Reese V. 2009. *MNRAS* 396:1675
Fisher DB, Drory N. 2011. *Ap. J. Lett.* 733:L47
Font AS, Bower RG, McCarthy IG, et al. 2008. *MNRAS* 389:1619
Fontanot F, De Lucia G, Monaco P, Somerville RS, Santini P. 2009. *MNRAS* 397:1776
Fontanot F, Monaco P, Cristiani S, Tozzi P. 2006. *MNRAS* 373:1173
Förster Schreiber NM, Genzel R, Bouché N, et al. 2009. *Ap. J.* 706:1364
Fu J, Guo Q, Kauffmann G, Krumholz MR. 2010. *MNRAS* 409:515
Furlong M, Bower RG, Theuns T, Schaye J, Crain RA, et al. 2014. arXiv:1410.3485
Gabor JM, Bournaud F. 2013. *MNRAS* 434:606
Gabor JM, Davé R. 2012. *MNRAS* 427:1816
Gabor JM, Davé R. 2015. *MNRAS* 447:374
Gabor JM, Davé R, Oppenheimer BD, Finlator K. 2011. *MNRAS* 417:2676
Gadotti DA, Kauffmann G. 2009. *MNRAS* 399:621
Gallazzi A, Charlot S, Brinchmann J, White SDM, Tremonti CA. 2005. *MNRAS* 362:41
Gaspari M, Melioli C, Brighenti F, D'Ercole A. 2011. *MNRAS* 411:349
Geach JE, Smail I, Moran SM, et al. 2011. *Ap. J. Lett.* 730:L19
Geha M, Blanton MR, Yan R, Tinker JL. 2012. *Ap. J.* 757:85
Genel S, Dekel A, Cacciato M. 2012a. *MNRAS* 425:788
Genel S, Naab T, Genzel R, et al. 2012b. *Ap. J.* 745:11
Genzel R, Newman S, Jones T, et al. 2011. *Ap. J.* 733:101
Genzel R, Tacconi LJ, Lutz D, et al. 2014. arXiv:1409.1171
Gnedin NY. 2000. *Ap. J.* 542:535
Gnedin NY, Draine BT. 2014. *Ap. J.* 795:37
Gnedin NY, Kravtsov AV. 2011. *Ap. J.* 728:88

Gnedin NY, Tassis K, Kravtsov AV. 2009. *Ap. J.* 697:55
Gonzalez-Perez V, Lacey CG, Baugh CM, et al. 2014. *MNRAS* 439:264
Governato F, Brook C, Mayer L, et al. 2010. *Nature* 463:203
Governato F, Brook CB, Brooks AM, et al. 2009. *MNRAS* 398:312
Governato F, Willman B, Mayer L, et al. 2007. *MNRAS* 374:1479
Guedes J, Callegari S, Madau P, Mayer L. 2011. *Ap. J.* 742:76
Guo Q, White S, Boylan-Kolchin M, et al. 2011. *MNRAS* 413:101
Guo Y, Giavalisco M, Ferguson HC, Cassata P, Koekemoer AM. 2012. *Ap. J.* 757:120
Haardt F, Madau P. 2012. *Ap. J.* 746:125
Haas MR, Schaye J, Booth CM, et al. 2013a. *MNRAS* 435:2931
Haas MR, Schaye J, Booth CM, et al. 2013b. *MNRAS* 435:2955
Hahn O, Abel T. 2011. *MNRAS* 415:2101
Hanasz M, Lesch H, Naab T, et al. 2013. *Ap. J. Lett.* 777:L38
Hayward CC, Narayanan D, Kereš D, et al. 2013. *MNRAS* 428:2529
Hearin AP, Watson DF, van den Bosch FC. 2014. arXiv:1404.6524
Heckman T, Best P. 2014. *Annu. Rev. Astron. Astrophys.* 52:589
Henriques BMB, Thomas PA, Oliver S, Roseboom I. 2009. *MNRAS* 396:535
Henriques BMB, White SDM, Thomas PA, et al. 2013. *MNRAS* 431:3373
Hernquist L. 1992. *Ap. J.* 400:460
Hernquist L. 1993a. *Ap. J. Suppl.* 86:389
Hernquist L. 1993b. *Ap. J.* 409:548
Hernquist L, Bouchet FR, Suto Y. 1991. *Ap. J. Suppl.* 75:231
Hernquist L, Katz N. 1989. *Ap. J. Suppl.* 70:419
Hinshaw G, Larson D, Komatsu E, et al. 2013. *Ap. J. Suppl.* 208:19
Hirschmann M, De Lucia G, Wilman D, et al. 2014. *MNRAS* 444:2938
Hirschmann M, Naab T, Somerville RS, Burkert A, Oser L. 2012a. *MNRAS* 419:3200
Hirschmann M, Somerville RS, Naab T, Burkert A. 2012b. *MNRAS* 426:237
Hockney RW, Eastwood JW. 1988. *Computer Simulation Using Particles*. Bristol: Hilger
Hogg DW, Blanton MR, Brinchmann J, et al. 2004. *Ap. J. Lett.* 601:L29
Hohl F. 1971. *Ap. J.* 168:343
Hopkins PF. 2013. *MNRAS* 428:2840
Hopkins PF, Bundy K, Hernquist L, Wuyts S, Cox TJ. 2010. *MNRAS* 401:1099
Hopkins PF, Cox TJ, Kereš D, Hernquist L. 2008a. *Ap. J. Suppl.* 175:390
Hopkins PF, Cox TJ, Younger JD, Hernquist L. 2009a. *Ap. J.* 691:1168
Hopkins PF, Hernquist L, Cox TJ, et al. 2005a. *Ap. J.* 630:716
Hopkins PF, Hernquist L, Cox TJ, Kereš D. 2008b. *Ap. J. Suppl.* 175:356
Hopkins PF, Hernquist L, Cox TJ, Kereš D, Wuyts S. 2009b. *Ap. J.* 691:1424
Hopkins PF, Hernquist L, Martini P, et al. 2005b. *Ap. J. Lett.* 625:L71
Hopkins PF, Kereš D, Murray N. 2013a. *MNRAS* 432:2639
Hopkins PF, Kereš D, Murray N, Quataert E, Hernquist L. 2012a. *MNRAS* 427:968
Hopkins PF, Kereš D, Onorbe J, et al. 2014. *MNRAS* 445:581
Hopkins PF, Quataert E. 2010. *MNRAS* 407:1529
Hopkins PF, Quataert E. 2011. *MNRAS* 415:1027
Hopkins PF, Quataert E, Murray N. 2012b. *MNRAS* 421:3522
Hopkins PF, Somerville RS, Cox TJ, et al. 2009. *MNRAS* 397:802
Hu CY, Naab T, Walch S, Moster BP, Oser L. 2014. *MNRAS* 443:1173
Hubble EP. 1926. *Ap. J.* 64:321
Iliev IT, Ciardi B, Alvarez MA, et al. 2006. *MNRAS* 371:1057
Iliev IT, Whalen D, Mellema G, et al. 2009. *MNRAS* 400:1283
Johansson PH, Burkert A, Naab T. 2009a. *Ap. J. Lett.* 707:L184
Johansson PH, Naab T, Ostriker JP. 2009b. *Ap. J. Lett.* 697:L38
Jonsson P, Groves BA, Cox TJ. 2010. *MNRAS* 403:17
Kassin SA, Brooks A, Governato F, Weiner BJ, Gardner JP. 2014. *Ap. J.* 790:89

Kassin SA, Weiner BJ, Faber SM, et al. 2007. *Ap. J. Lett.* 660:L35
Katz N. 1992. *Ap. J.* 391:502
Katz N, Weinberg DH, Hernquist L. 1996. *Ap. J. Suppl.* 105:19
Kauffmann G. 1996a. *MNRAS* 281:475
Kauffmann G. 1996b. *MNRAS* 281:487
Kauffmann G, Colberg JM, Diaferio A, White SDM. 1999. *MNRAS* 303:188
Kauffmann G, Haehnelt M. 2000. *MNRAS* 311:576
Kauffmann G, Heckman TM, White SDM, et al. 2003. *MNRAS* 341:54
Kauffmann G, Li C, Zhang W, Weinmann S. 2013. *MNRAS* 430:1447
Kauffmann G, White SDM, Guiderdoni B. 1993. *MNRAS* 264:201
Kennicutt RC Jr. 1998. *Ap. J.* 498:541
Kennicutt RC Jr, Evans NJ. 2012. *Annu. Rev. Astron. Astrophys.* 50:531
Kereš D, Katz N, Weinberg DH, Davé R. 2005. *MNRAS* 363:2
Kewley LJ, Dopita MA, Leitherer C, et al. 2013. *Ap. J.* 774:100
Kewley LJ, Ellison SL. 2008. *Ap. J.* 681:1183
Khochfar S, Ostriker JP. 2008. *Ap. J.* 680:54
Kim J-H, Wise JH, Abel T. 2009. *Ap. J. Lett.* 694:L123
Kimm T, Devriendt J, Slyz A, et al. 2011. arXiv:1106.0538
Kimm T, Somerville RS, Yi SK, et al. 2009. *MNRAS* 394:1131
King A. 2005. *Ap. J. Lett.* 635:L121
Klypin AA, Trujillo-Gomez S, Primack J. 2011. *Ap. J.* 740:102
Knebe A, Knollmann SR, Muldrew SI, et al. 2011. *MNRAS* 415:2293
Kormendy J. 1977. *Ap. J.* 218:333
Kormendy J. 2013. In *Secular Evolution in Disk Galaxies*, ed. J Falcón-Barroso, JH Knapen, pp. 1–159. Cambridge, UK: Cambridge Univ. Press. arXiv:1311.2609
Kormendy J, Bender R. 2012. *Ap. J. Suppl.* 198:2
Kormendy J, Drory N, Bender R, Cornell ME. 2010. *Ap. J.* 723:54
Kormendy J, Fisher DB, Cornell ME, Bender R. 2009. *Ap. J. Suppl.* 182:216
Kormendy J, Freeman KC. 2014. arXiv:1411.2170
Kormendy J, Ho LC. 2013. *Annu. Rev. Astron. Astrophys.* 51:511
Kormendy J, Kennicutt RC Jr. 2004. *Annu. Rev. Astron. Astrophys.* 42:603
Kotarba H, Lesch H, Dolag K, et al. 2011. *MNRAS* 415:3189
Kravtsov AV, Gnedin OY, Klypin AA. 2004. *Ap. J.* 609:482
Kravtsov AV, Klypin AA, Khokhlov AM. 1997. *Ap. J. Suppl.* 111:73
Krolik JH. 1999. *Active Galactic Nuclei: From the Central Black Hole to the Galactic Environment*. Princeton, NJ: Princeton Univ. Press
Krumholz MR, Dekel A. 2012. *Ap. J.* 753:16
Krumholz MR, Dekel A, McKee CF. 2012. *Ap. J.* 745:69
Krumholz MR, McKee CF, Tumlinson J. 2009. *Ap. J.* 699:850
Kuhlen M, Krumholz MR, Madau P, Smith BD, Wise J. 2012. *Ap. J.* 749:36
Lacey C, Cole S. 1993. *MNRAS* 262:627
Lagos CDP, Baugh CM, Lacey CG, et al. 2011a. *MNRAS* 418:1649
Lagos CDP, Davis TA, Lacey CG, et al. 2014. *MNRAS* 443:1002
Lagos CDP, Lacey CG, Baugh CM, Bower RG, Benson AJ. 2011b. *MNRAS* 416:1566
Lang P, Wuyts S, Somerville RS, et al. 2014. *Ap. J.* 788:11
Lara-López MA, Cepa J, Bongiovanni A, et al. 2010. *Astron. Astrophys.* 521:53
Lara-López MA, Hopkins AM, López-Sánchez AR, et al. 2013. *MNRAS* 433:35
Larson RB. 1974. *MNRAS* 169:229
Leitner SN, Kravtsov AV. 2011. *Ap. J.* 734:48
Leroy AK, Walter F, Sandstrom K, et al. 2013. *Astron. J.* 146:19
Li Y, Bryan GL. 2014. *Ap. J.* 789:54
Lilly SJ, Carollo CM, Pipino A, Renzini A, Peng Y. 2013. *Ap. J.* 772:119
Lu Y, Mo HJ, Weinberg MD, Katz N. 2011. *MNRAS* 416:1949

Lu Y, Wechsler RH, Somerville RS, et al. 2014. *Ap. J.* 597:123
Mac Low MM, Glover SCO. 2012. *Ap. J.* 746:135
Madau P, Dickinson M. 2014. *Annu. Rev. Astron. Astrophys.* 52:415
Maller AH, Dekel A. 2002. *MNRAS* 335:487
Mandelker N, Dekel A, Ceverino D, et al. 2014. *MNRAS* 443:3675
Mannucci F, Cresci G, Maiolino R, Marconi A, Gnerucci A. 2010. *MNRAS* 408:2115
Marchesini D, van Dokkum PG, Förster Schreiber NM, et al. 2009. *Ap. J.* 701:1765
Marchesini D, Whitaker KE, Brammer G, et al. 2010. *Ap. J.* 725:1277
McKee CF, Krumholz MR. 2010. *Ap. J.* 709:308
McKee CF, Ostriker EC. 2007. *Annu. Rev. Astron. Astrophys.* 45:565
McKee CF, Ostriker JP. 1977. *Ap. J.* 218:148
McNamara BR, Nulsen PEJ. 2007. *Annu. Rev. Astron. Astrophys.* 45:117
Mendygral PJ, Jones TW, Dolag K. 2012. *Ap. J.* 750:166
Mihos JC, Hernquist L. 1996. *Ap. J.* 464:641
Miller SH, Bundy K, Sullivan M, Ellis RS, Treu T. 2011. *Ap. J.* 741:115
Miller SH, Ellis RS, Sullivan M, et al. 2012. *Ap. J.* 753:74
Mitchell PD, Lacey CG, Baugh CM, Cole S. 2013. *MNRAS* 435:87
Mo H, van den Bosch FC, White S. 2010. *Galaxy Formation and Evolution*. Cambridge, UK: Cambridge Univ. Press
Mo HJ, Mao S, White SDM. 1998. *MNRAS* 295:319
Moe M, Arav N, Bautista MA, Korista KT. 2009. *Ap. J.* 706:525
Monaco P, Benson AJ, De Lucia G, et al. 2014. *MNRAS* 441:2058
Monaco P, Fontanot F, Taffoni G. 2007. *MNRAS* 375:1189
Monaghan JJ. 1992. *Annu. Rev. Astron. Astrophys.* 30:543
Moody CE, Guo Y, Mandelker N, et al. 2014. *MNRAS* 444:1389
Moster BP, Macciò AV, Somerville RS, Johansson PH, Naab T. 2010a. *MNRAS* 403:1009
Moster BP, Macciò AV, Somerville RS, Naab T, Cox TJ. 2011. *MNRAS* 415:3750
Moster BP, Macciò AV, Somerville RS, Naab T, Cox TJ. 2012. *MNRAS* 423:2045
Moster BP, Naab T, White SDM. 2013. *MNRAS* 428:3121
Moster BP, Somerville RS, Maulbetsch C, et al. 2010b. *Ap. J.* 710:903
Moustakas J, Coil AL, Aird JA, et al. 2013. *Ap. J.* 767:50
Munshi F, Governato F, Brooks AM, et al. 2013. *Ap. J.* 766:56
Murali C, Katz N, Hernquist L, Weinberg DH, Davé R. 2002. *Ap. J.* 571:1
Murray N, Quataert E, Thompson TA. 2005. *Ap. J.* 618:569
Muzzin A, Marchesini D, Stefanon M, et al. 2013. *Ap. J.* 777:18
Naab T, Johansson PH, Ostriker JP. 2009. *Ap. J. Lett.* 699:L178
Nagashima M, Lacey CG, Okamoto T, et al. 2005. *MNRAS* 363:31
Narayan R, Yi I. 1994. *Ap. J. Lett.* 428:L13
Narayanan D, Cox TJ, Kelly B, et al. 2008. *Ap. J. Suppl.* 176:331
Navarro JF, Frenk CS, White SDM. 1997. *Ap. J.* 490:493
Navarro JF, Steinmetz M. 2000. *Ap. J.* 538:477
Navarro JF, White SDM. 1993. *MNRAS* 265:271
Netzer H. 2013. *The Physics and Evolution of Active Galactic Nuclei*. Cambridge, UK: Cambridge Univ. Press
Niemi SM, Somerville RS, Ferguson HC, et al. 2012. *MNRAS* 421:1539
Noeske KG, Faber SM, Weiner BJ, et al. 2007a. *Ap. J.* 660:47
Noeske KG, Weiner BJ, Faber SM, et al. 2007b. *Ap. J. Lett.* 660:L43
Obreschkow D, Croton D, De Lucia G, Khochfar S, Rawlings S. 2009. *Ap. J.* 698:1467
Oh SH, Brook C, Governato F, et al. 2011. *Astron. J.* 142:24
Okamoto T, Gao L, Theuns T. 2008. *MNRAS* 390:920
Oppenheimer BD, Davé R. 2006. *MNRAS* 373:1265
Oppenheimer BD, Davé R. 2008. *MNRAS* 387:577
Oppenheimer BD, Davé R, Kereš D, et al. 2010. *MNRAS* 406:2325
Oser L, Naab T, Ostriker JP, Johansson PH. 2012. *Ap. J.* 744:63

Ostriker JP, Choi E, Ciotti L, Novak GS, Proga D. 2010. *Ap. J.* 722:642
Ostriker JP, Peebles PJE. 1973. *Ap. J.* 186:467
Papovich C, Finkelstein SL, Ferguson HC, Lotz JM, Giavalisco M. 2011. *MNRAS* 412:1123
Parkinson H, Cole S, Helly J. 2008. *MNRAS* 383:557
Parry OH, Eke VR, Frenk CS. 2009. *MNRAS* 396:1972
Pawlik AH, Schaye J. 2009. *MNRAS* 396:46
Pawlik AH, Schaye J. 2011. *MNRAS* 412:1943
Peebles PJE. 1969. *Ap. J.* 155:393
Peeples MS, Shankar F. 2011. *MNRAS* 417:2962
Peeples MS, Werk JK, Tumlinson J, et al. 2014. *Ap. J.* 786:54
Pen UL. 1998. *Ap. J. Suppl.* 115:19
Peng Y-j, Lilly SJ, Kovac K, et al. 2010. *Ap. J.* 721:193
Peng Y-j, Lilly SJ, Renzini A, Carollo M. 2012. *Ap. J.* 757:4
Pfrommer C. 2013. *Ap. J.* 779:10
Planck Collab., Ade PAR, Aghanim N, et al. 2014. *Astron. Astrophys.* 571:16
Pontzen A, Governato F. 2012. *MNRAS* 421:3464
Popping A, Davé R, Braun R, Oppenheimer BD. 2009. *Astron. Astrophys.* 504:15
Popping G, Behroozi PS, Peeples MS. 2014a. arXiv:1409.1574
Popping G, Caputi KI, Somerville RS, Trager SC. 2012. *MNRAS* 425:2386
Popping G, Somerville RS, Trager SC. 2014b. *MNRAS* 442:2398
Porter LA, Somerville RS, Primack JR, Johansson PH. 2014. *MNRAS* 444:942
Primack JR. 2005. *New Astron. Rev.* 49:25
Prochaska JX, Hennawi JF, Simcoe RA. 2013. *Ap. J. Lett.* 762:L19
Quinn T, Katz N, Efstathiou G. 1996. *MNRAS* 278:49
Rafieferantsoa M, Davé R, Anglés-Alcazar D, et al. 2014. arXiv:1408.2531
Rahmati A, Pawlik AH, Raicevic M, Schaye J. 2013. *MNRAS* 430:2427
Rahmati A, Schaye J. 2014. *MNRAS* 438:529
Read JI, Hayfield T. 2012. *MNRAS* 422:3037
Rees MJ, Ostriker JP. 1977. *MNRAS* 179:541
Ritchie BW, Thomas PA. 2001. *MNRAS* 323:743
Robaina AR, Bell EF, van der Wel A, et al. 2010. *Ap. J.* 719:844
Roberts MS, Haynes MP. 1994. *Annu. Rev. Astron. Astrophys.* 32:115
Robertson B, Bullock JS, Cox TJ, et al. 2006a. *Ap. J.* 645:986
Robertson B, Cox TJ, Hernquist L, et al. 2006b. *Ap. J.* 641:21
Robertson B, Yoshida N, Springel V, Hernquist L. 2004. *Ap. J.* 606:32
Rodighiero G, Daddi E, Baronchelli I, et al. 2011. *Ap. J. Lett.* 739:L40
Rudie GC, Steidel CC, Trainor RF, et al. 2012. *Ap. J.* 750:67
Ruszkowski M, Brüggen M, Begelman MC. 2004. *Ap. J.* 615:675
Saintonge A, Lutz D, Genzel R, et al. 2013. *Ap. J.* 778:2
Saitoh TR, Makino J. 2013. *Ap. J.* 768:44
Salim S, Rich RM, Charlot S, et al. 2007. *Ap. J. Suppl.* 173:267
Salmon B, Papovich C, Finkelstein SL, et al. 2014. arXiv:1407.6012
Sanders DB, Mirabel IF. 1996. *Annu. Rev. Astron. Astrophys.* 34:749
Sanders RL, Shapley AE, Kriek M, et al. 2015. *Ap. J.* 799:138
Savaglio S, Glazebrook K, Le Borgne D, et al. 2005. *Ap. J.* 635:260
Scannapieco E, Brüggen M. 2008. *Ap. J.* 686:927
Scannapieco E, Brüggen M. 2010. *MNRAS* 405:1634
Schaye J. 2004. *Ap. J.* 609:667
Schaye J, Crain RA, Bower RG, et al. 2015. *MNRAS* 446:521
Schaye J, Dalla Vecchia C. 2008. *MNRAS* 383:1210
Schaye J, Dalla Vecchia C, Booth CM, et al. 2010. *MNRAS* 402:1536
Schechter P. 1976. *Ap. J.* 203:297
Schiminovich D, Wyder TK, Martin DC, et al. 2007. *Ap. J. Suppl.* 173:315

Schmidt M. 1959. *Ap. J.* 129:243
Schreiber C, Pannella M, Elbaz D, et al. 2014. arXiv:1409.5433
Scoville N, Aussel H, Sheth K, et al. 2014. *Ap. J.* 783:84
Shankar F, Marulli F, Bernardi M, et al. 2010. *MNRAS* 403:117
Shankar F, Marulli F, Bernardi M, et al. 2013. *MNRAS* 428:109
Sharma P, Chandran BDG, Quataert E, Parrish IJ. 2009. *Ap. J.* 699:348
Shen S, Mo HJ, White SDM, et al. 2003. *MNRAS* 343:978
Sijacki D, Springel V, di Matteo T, Hernquist L. 2007. *MNRAS* 380:877
Silk J, Rees MJ. 1998. *Astron. Astrophys.* 331:1
Simard L, Mendel JT, Patton DR, Ellison SL, McConnachie AW. 2011. *Ap. J. Suppl.* 196:11
Simha V, Weinberg DH, Davé R, et al. 2009. *MNRAS* 399:650
Simien F, de Vaucouleurs G. 1986. *Ap. J.* 302:564
Somerville RS. 2002. *Ap. J. Lett.* 572:L23
Somerville RS, Barden M, Rix H, et al. 2008a. *Ap. J.* 672:776
Somerville RS, Hopkins PF, Cox TJ, Robertson BE, Hernquist L. 2008b. *MNRAS* 391:481
Somerville RS, Kolatt TS. 1999. *MNRAS* 305:1
Somerville RS, Popping G, Trager SC. 2015. *MNRAS*. In press. arXiv:1503.00755
Somerville RS, Primack JR. 1999. *MNRAS* 310:1087
Sommer-Larsen J, Gelato S, Vedel H. 1999. *Ap. J.* 519:501
Sparre M, Hayward CC, Springel V, et al. 2014. arXiv:1409.0009
Speagle JS, Steinhardt CL, Capak PL, Silverman JD. 2014. *Ap. J. Suppl.* 214:15
Springel V. 2000. *MNRAS* 312:859
Springel V. 2005. *MNRAS* 364:1105
Springel V. 2010a. *MNRAS* 401:791
Springel V. 2010b. *Annu. Rev. Astron. Astrophys.* 48:391
Springel V, Di Matteo T, Hernquist L. 2005a. *Ap. J. Lett.* 620:L79
Springel V, Di Matteo T, Hernquist L. 2005b. *MNRAS* 361:776
Springel V, Hernquist L. 2003. *MNRAS* 339:289
Springel V, White SDM, Jenkins A, et al. 2005c. *Nature* 435:629
Springel V, White SDM, Tormen G, Kauffmann G. 2001. *MNRAS* 328:726
Steidel CC, Rudie GC, Strom AL, et al. 2014. *Ap. J.* 795:165
Steinhardt CL, Speagle JS, Capak P, et al. 2014. *Ap. J.* 791:25
Steinmetz M. 1999. *Ap. Space Sci.* 269:513
Stewart KR, Brooks AM, Bullock JS, et al. 2013. *Ap. J.* 769:74
Stinson G, Seth A, Katz N, et al. 2006. *MNRAS* 373:1074
Stinson GS, Brook C, Macciò AV, et al. 2013. *MNRAS* 428:129
Sturm E, González-Alfonso E, Veilleux S, et al. 2011. *Ap. J. Lett.* 733:L16
Sutherland RS, Dopita MA. 1993. *Ap. J. Suppl.* 88:253
Tacconi LJ, Genzel R, Neri R, et al. 2010. *Nature* 463:781
Tacconi LJ, Neri R, Genzel R, et al. 2013. *Ap. J.* 768:74
Tasca LAM, White SDM. 2011. *Astron. Astrophys.* 530:106
Tasitsiomi A, Kravtsov AV, Wechsler RH, Primack JR. 2004. *Ap. J.* 614:533
Teyssier R. 2010. RAMSES: a new N-body and hydrodynamical code. *Astrophys. Source Code Libr.*, ascl:1011.007 (Abstr.)
Thomas D, Maraston C, Bender R, Mendes de Oliveira C. 2005. *Ap. J.* 621:673
Thoul AA, Weinberg DH. 1996. *Ap. J.* 465:608
Tolstoy E, Hill V, Tosi M. 2009. *Annu. Rev. Astron. Astrophys.* 47:371
Tomczak AR, Quadri RF, Tran KVH, et al. 2014. *Ap. J.* 783:85
Toomre A. 1964. *Ap. J.* 139:1217
Toomre A. 1977. In *Evolution of Galaxies and Stellar Populations*, ed. BM Tinsley, RBG Larson, pg. 401. New Haven, CT: Yale Univ. Observ.
Torrey P, Vogelsberger M, Genel S, et al. 2014. *MNRAS* 438:1985
Trager SC, Faber SM, Worthey G, González JJ. 2000. *Astron. J.* 119:1645

Tremonti CA, Heckman TM, Kauffmann G, et al. 2004. *Ap. J.* 613:898
Trujillo I, Förster Schreiber NM, Rudnick G, et al. 2006. *Ap. J.* 650:18
Tully RB, Fisher JR. 1977. *Astron. Astrophys.* 54:661
Tumlinson J, Thom C, Werk JK, et al. 2013. *Ap. J.* 777:59
Übler H, Naab T, Oser L, et al. 2014. *MNRAS* 443:2092
van den Bosch FC, Burkert A, Swaters RA. 2001. *MNRAS* 326:1205
van den Bosch FC, Yang X, Mo HJ, et al. 2007. *MNRAS* 376:841
van der Wel A, Franx M, van Dokkum PG, et al. 2014. *Ap. J.* 788:28
van Dokkum PG, Bezanson R, van der Wel A, et al. 2014. *Ap. J.* 791:45
van Dokkum PG, Franx M, Kriek M, et al. 2008. *Ap. J. Lett.* 677:L5
Veilleux S, Cecil G, Bland-Hawthorn J. 2005. *Annu. Rev. Astron. Astrophys.* 43:769
Vernaleo JC, Reynolds CS. 2006. *Ap. J.* 645:83
Vitvitska M, Klypin AA, Kravtsov AV, et al. 2002. *Ap. J.* 581:799
Vogelsberger M, Genel S, Springel V, et al. 2014a. *MNRAS* 444:1518
Vogelsberger M, Genel S, Springel V, et al. 2014b. *Nature* 509:177
Vogelsberger M, Sijacki D, Kereš D, Springel V, Hernquist L. 2012. *MNRAS* 425:3024
Volonteri M. 2010. *Astron. Astrophys. Rev.* 18:279
Wadsley JW, Stadel J, Quinn T. 2004. *New Astron.* 9:137
Wake DA, Whitaker KE, Labbé I, et al. 2011. *Ap. J.* 728:46
Walch S, Wünsch R, Burkert A, Glover S, Whitworth A. 2011. *Ap. J.* 733:47
Walcher J, Groves B, Budavári T, Dale D. 2011. *Ap. Space Sci.* 331:1
Walker IR, Mihos JC, Hernquist L. 1996. *Ap. J.* 460:121
Weinmann SM, Kauffmann G, von der Linden A, De Lucia G. 2010. *MNRAS* 406:2249
Weinmann SM, Pasquali A, Oppenheimer BD, et al. 2012. *MNRAS* 426:2797
Weinmann SM, van den Bosch FC, Yang X, Mo HJ. 2006a. *MNRAS* 366:2
Weinmann SM, van den Bosch FC, Yang X, et al. 2006b. *MNRAS* 372:1161
Wetzel AR, Tinker JL, Conroy C. 2012. *MNRAS* 424:232
Whitaker KE, Franx M, Leja J, et al. 2014. *Ap. J.* 795:104
White CE, Somerville RS, Ferguson HC. 2014. arXiv:1407.1850
White SDM, Frenk CS. 1991. *Ap. J.* 379:52
White SDM, Rees MJ. 1978. *MNRAS* 183:341
Wiersma RPC, Schaye J, Smith BD. 2009a. *MNRAS* 393:99
Wiersma RPC, Schaye J, Theuns T, Dalla Vecchia C, Tornatore L. 2009b. *MNRAS* 399:574
Wise JH, Abel T. 2011. *MNRAS* 414:3458
Woo J, Dekel A, Faber SM, et al. 2013. *MNRAS* 428:3306
Wuyts E, Kurk J, Förster Schreiber NM, et al. 2014. *Ap. J. Lett.* 798:40
Wuyts S, Förster Schreiber NM, Genzel R, et al. 2012. *Ap. J.* 753:114
Wuyts S, Förster Schreiber NM, van der Wel A, et al. 2011. *Ap. J.* 742:96
Yang X, Mo HJ, van den Bosch FC. 2003. *MNRAS* 339:1057
Yates RM, Henriques B, Thomas PA, et al. 2013. *MNRAS* 435:3500
Yoshida N, Stoehr F, Springel V, White SDM. 2002. *MNRAS* 335:762
Zahid HJ, Geller MJ, Kewley LJ, et al. 2013. *Ap. J. Lett.* 771:L19
Zheng Z, Berlind AA, Weinberg DH, et al. 2005. *Ap. J.* 633:791

RELATED RESOURCES

MeerKAT: **http://www.ska.ac.za/meerkat/index.php**
NOEMA: **http://iram-institute.org/EN/noema-project.php**