Astronomy and Computing 15 (2016) 72–89

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

Full length article

The EAGLE simulations of galaxy formation: Public release of halo and galaxy catalogues[☆]

S. McAlpine[a,*], J.C. Helly[a], M. Schaller[a], J.W. Trayford[a], Y. Qu[a], M. Furlong[a], R.G. Bower[a], R.A. Crain[b], J. Schaye[c], T. Theuns[a], C. Dalla Vecchia[d,e], C.S. Frenk[a], I.G. McCarthy[b], A. Jenkins[a], Y. Rosas-Guevara[f], S.D.M. White[g], M. Baes[h], P. Camps[h], G. Lemson[i]

[a] Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham, DH1 3LE, UK
[b] Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK
[c] Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands
[d] Instituto de Astrofísica de Canarias, C/ Vía Láctea s/n, 38205 La Laguna, Tenerife, Spain
[e] Departamento de Astrofísica, Universidad de La Laguna, Av. del Astrofísico Francisco Sánchez s/n, 38206 La Laguna, Tenerife, Spain
[f] Departamento de Astronomía, Universidad de Chile, Casilla 36-D, Las Condes, Santiago, Chile
[g] Max-Planck-Institut fur Astrophysik, Karl-Schwarzschild-Str. 1, D-85748 Garching, Germany
[h] Sterrenkundig Observatorium, Universiteit Gent, Krijgslaan 281, B-9000 Gent, Belgium
[i] Department of Physics & Astronomy, Johns Hopkins University, Baltimore, MD, 21218, USA

ARTICLE INFO

Article history:
Received 5 October 2015
Accepted 25 February 2016
Available online 17 March 2016

Keywords:
Cosmology: theory
Galaxies: formation
Galaxies: evolution
Method: numerical

ABSTRACT

We present the public data release of halo and galaxy catalogues extracted from the EAGLE suite of cosmological hydrodynamical simulations of galaxy formation. These simulations were performed with an enhanced version of the GADGET code that includes a modified hydrodynamics solver, time-step limiter and subgrid treatments of baryonic physics, such as stellar mass loss, element-by-element radiative cooling, star formation and feedback from star formation and black hole accretion. The simulation suite includes runs performed in volumes ranging from 25 to 100 comoving megaparsecs per side, with numerical resolution chosen to marginally resolve the Jeans mass of the gas at the star formation threshold. The free parameters of the subgrid models for feedback are calibrated to the redshift $z = 0$ galaxy stellar mass function, galaxy sizes and black hole mass–stellar mass relation. The simulations have been shown to match a wide range of observations for present-day and higher-redshift galaxies. The raw particle data have been used to link galaxies across redshifts by creating merger trees. The indexing of the tree produces a simple way to connect a galaxy at one redshift to its progenitors at higher redshift and to identify its descendants at lower redshift. In this paper we present a relational database which we are making available for general use. A large number of properties of haloes and galaxies and their merger trees are stored in the database, including stellar masses, star formation rates, metallicities, photometric measurements and mock *gri* images. Complex queries can be created to explore the evolution of more than 10^5 galaxies, examples of which are provided in the Appendix. The relatively good and broad agreement of the simulations with a wide range of observational datasets makes the database an ideal resource for the analysis of model galaxies through time, and for connecting and interpreting observational datasets.

1. Introduction

Galaxy formation is a complex, non-linear process that involves a wide range of physical and astrophysical phenomena, from the evolution of dark matter clustering to intricate feedback effects coupling gas cooling and outflows to star and black hole formation. Theoretical studies of galaxy formation thus require rigorous detailed modelling to link together these phenomena over a very wide range of scales. Two techniques have been developed for this purpose: semianalytic modelling (White and Frenk, 1991) and hydrodynamical simulations (Carlberg et al., 1990; Katz et al., 1992). Both techniques have been extensively developed over the past 25 years (e.g. Porter et al., 2014; Henriques et al., 2015; Lacey

☆ http://www.eaglesim.org/database.php.
* Corresponding author.
E-mail addresses: s.r.mcalpine@durham.ac.uk (S. McAlpine), j.c.helly@durham.ac.uk (J.C. Helly), matthieu.schaller@durham.ac.uk (M. Schaller).

http://dx.doi.org/10.1016/j.ascom.2016.02.004
2213-1337/

et al., 2015, for semi-analytic models) and (e.g. Oppenheimer et al., 2010; Puchwein and Springel, 2013; Dubois et al., 2014; Okamoto et al., 2014; Vogelsberger et al., 2014; Khandai et al., 2015, for hydrodynamical simulations).

Recently, the Virgo[1] Consortium's "Evolution and Assembly of GaLaxies and their Environments" simulation suite (EAGLE, Schaye et al., 2015; Crain et al., 2015) has been able to reproduce key observational datasets, such as the present-day stellar mass function of galaxies, the correlation of black hole mass and stellar mass and the dependence of galaxy sizes on stellar mass, with unprecedented fidelity. As well as reproducing these observations, which were used during the calibration of the simulation parameters, the simulation outputs match many other properties of the observed galaxy population and the intergalactic medium both at the present day and at earlier epochs, as we briefly discuss below. These simulations therefore provide a powerful resource for understanding the formation of galaxies and for linking and interpreting observational datasets.

The aim of this paper is to introduce and make available a relational database that can be queried using the Structured Query Language (SQL) to explore and exploit the halo and galaxy catalogues of the main EAGLE simulations. Columns containing integrated quantities describing the galaxies, such as stellar mass, star formation rates, metallicities and luminosities, are provided for more than 10^5 simulated galaxies and these can be individually followed through their evolution across cosmic time. This database is available at the address http://www.eaglesim.org/database.php.

The simulations follow the gravitational hydrodynamical equations, tracking the evolution of baryons and dark matter. The initial conditions reflect the small density fluctuations observed in the cosmic microwave background (CMB). By tracking the movement of baryon and dark matter particles, the simulations calculate how these fluctuations are amplified by gravity, and how pressure and radiative cooling of baryons separate these two matter components of the universe. The simulations include subgrid formulations to account for processes that cannot be directly resolved in the calculation and that describe how stars and black holes form and impact the matter distribution around them. EAGLE improves on previous hydrodynamical simulations of representative volumes, through the use of physically motivated subgrid source and sink terms as well as through the adoption of a clear strategy for the calibration of uncertain subgrid parameters (Crain et al., 2015) and by producing a galaxy population that reproduces many of the characteristics of the observed population over a wide range of redshifts.

The usability of the simulation data products is greatly enhanced when presented in a relational database, making it simple and quick to select galaxy samples based on multiple galaxy properties, to connect them to their halos and to follow their evolution over cosmic time (Lemson and Springel, 2006). Such databases were originally designed to host results from large surveys (e.g. the SDSS SkyServer Szalay et al., 2000) and later the halo catalogues from dark matter simulations and galaxy catalogues from semi-analytic models (applied to the *Millennium Simulation*, see Lemson and Virgo Consortium, 2006). They have since been expanded to include the wider range of data available from hydrodynamical simulations (e.g. Dolag et al., 2009; Khandai et al., 2015; Nelson et al., 2015). The database allows multiple indexing of the data that significantly enhances access speed and allows the selection of smaller data subsets that can be quickly analysed using simple scripting languages. This approach avoids the need for the user to copy the raw simulation data or even just the full galaxy catalogues, reducing data transfer volumes to a manageable level. The galaxy properties stored in the database can be compared to observations or to other models, whilst the physics of galaxy formation can be explored by tracking an individual galaxy's behaviour and environment through cosmic time.

This paper is intended as a reference guide for accessing the publicly available EAGLE database, and is laid out as follows. Section 2 presents a brief overview of the EAGLE simulation suite, including the list of simulations available in the database and the values of the subgrid parameters that vary, as well as an overview of the construction of the merger trees and database tables. A short tutorial describing how to access the data is presented in Section 3. We give some words of caution and some remarks on the simulations in Section 4 and conclude in Section 5. Some additional examples combining the PYTHON and SQL languages to access the data are given in Appendix A whilst the full list of galaxy and halo properties available in this data release is given in Appendix B together with a list of output redshifts in Appendix C and detailed equations given in Appendix D. Throughout this paper we quote magnitudes in the AB system and use 'h-free' units unless stated otherwise.

2. The EAGLE simulation suite

The EAGLE simulation suite is a set of cosmological hydrodynamical simulations in cubic, periodic volumes ranging from 25 to 100 comoving megaparsecs (cMpc) per side that track the evolution of both baryonic (gas, stars and massive black holes) and non-baryonic (dark matter) elements from a starting redshift of $z = 127$ to the present day. All simulations adopt a flat ΛCDM cosmology with parameters taken from the *Planck* mission (Planck Collaboration et al., 2014) results: $\Omega_\Lambda = 0.693$, $\Omega_m = 0.307$, $\Omega_b = 0.04825$, $\sigma_8 = 0.8288$, $n_s = 0.9611$, $Y = 0.248$ and $H_0 = 67.77$ km s^{-1} Mpc^{-1} (i.e. $h = 0.6777$). The initial conditions were generated using second-order Lagrangian perturbation theory (Jenkins, 2010) and the phase information is taken from the public PANPHAISA Gaussian white noise field (Jenkins, 2013). Full details of how the ICs were made are given in Appendix B of Schaye et al. (2015). The simulation suite was run with a modified version of the GADGET-3 Smoothed Particle Hydrodynamics (SPH) code (last described by Springel, 2005), and includes a full treatment of gravity and hydrodynamics. The modifications to the SPH method are collectively referred to as ANARCHY (Dalla Vecchia, (in prep.), see also Appendix A of Schaye et al., 2015; Schaller et al., 2015a), and use the $\mathcal{C}_2$ kernel of Wendland (1995), the pressure–entropy formulation of SPH of Hopkins (2013), the time-step limiters introduced by Durier and Dalla Vecchia (2012), the artificial viscosity switch of Cullen and Dehnen (2010) and a weak thermal conduction term of the form proposed by Price (2008). The effects of this state-of-the-art formulation of SPH on the galaxy properties are explored in detail by Schaller et al. (2015a).

2.1. Subgrid model

Processes not resolved by the numerical scheme are implemented as subgrid source and sink terms in the differential equations. For each process, schemes were adopted that are as simple as possible and that only depend on the local hydrodynamic properties. This last requirement differentiates EAGLE from most other cosmological, hydrodynamical simulation projects

[1] http://virgo.dur.ac.uk/.