Title

THE AGORA HIGH-RESOLUTION GALAXY SIMULATIONS COMPARISON PROJECT. II. ISOLATED DISK TEST

Journal

The Astrophysical Journal, 833(2)

ISSN

0004-637X

Authors

Kim, Ji-hoon
Agertz, Oscar
Teyssier, Romain
et al.

Publication Date

2016-12-20

DOI

10.3847/1538-4357/833/2/202

Peer reviewed

DRAFT VERSION DECEMBER 20, 2016
Preprint typeset using LaTeX style emulateapj v. 12/16/11

THE *AGORA* HIGH-RESOLUTION GALAXY SIMULATIONS COMPARISON PROJECT. II: ISOLATED DISK TEST

JI-HOON KIM[1,2,3,4], OSCAR AGERTZ[5,6], ROMAIN TEYSSIER[7], MICHAEL J. BUTLER[8,†,‡], DANIEL CEVERINO[9,†,‡], JUN-HWAN CHOI[10,†,‡], ROBERT FELDMANN[7,11,†], BEN W. KELLER[12,†,‡], ALESSANDRO LUPI[13,†,‡], THOMAS QUINN[14,†,‡], YVES REVAZ[15,†,‡], SPENCER WALLACE[14,†], NICKOLAY Y. GNEDIN[16,17,18,‡], SAMUEL N. LEITNER[19,‡], SIJING SHEN[20,‡], BRITTON D. SMITH[21,‡], ROBERT THOMPSON[22,‡], MATTHEW J. TURK[23,‡], TOM ABEL[1,2], KENZA S. ARRAKI[24], SAMANTHA M. BENINCASA[12], SUKANYA CHAKRABARTI[25], COLIN DEGRAF[20], AVISHAI DEKEL[26], NATHAN J. GOLDBAUM[22], PHILIP F. HOPKINS[3], CAMERON B. HUMMELS[3], ANATOLY KLYPIN[24], HUI LI[27], PIERO MADAU[28,13], NIR MANDELKER[29,26], LUCIO MAYER[7], KENTARO NAGAMINE[30,31], SARAH NICKERSON[7], BRIAN W. O'SHEA[32], JOEL R. PRIMACK[33], SANTI ROCA-FÀBREGA[26], VADIM SEMENOV[17], IKKOH SHIMIZU[30], CHRISTINE M. SIMPSON[34], KEITA TODOROKI[35], JAMES W. WADSLEY[12], AND JOHN H. WISE[36] FOR THE *AGORA* COLLABORATION[37]

[1]Kavli Institute for Particle Astrophysics and Cosmology, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
[2]Department of Physics, Stanford University, Stanford, CA 94305, USA
[3]Department of Astronomy, California Institute of Technology, Pasadena, CA 91125, USA
[4]Einstein Fellow, me@jihoonkim.org
[5]Department of Physics, University of Surrey, Guildford, Surrey, GU2 7XH, United Kingdom
[6]Lund Observatory, Department of Astronomy and Theoretical Physics, Lund University, SE-22100 Lund, Sweden
[7]Centre for Theoretical Astrophysics and Cosmology, Institute for Computational Science, University of Zurich, Zurich, 8057, Switzerland
[8]Max-Planck-Institut für Astronomie, D-69117 Heidelberg, Germany
[9]Zentrum für Astronomie der Universität Heidelberg, Institut für Theoretische Astrophysik, 69120 Heidelberg, Germany
[10]Department of Astronomy, University of Texas, Austin, TX 78712 , USA
[11]Department of Astronomy, University of California at Berkeley, Berkeley, CA 94720, USA
[12]Department of Physics and Astronomy, McMaster University, Hamilton, ON L8S 4M1, Canada
[13]Institut d'Astrophysique de Paris, Sorbonne Universites, UPMC Univ Paris 6 et CNRS, 75014 Paris, France
[14]Department of Astronomy, University of Washington, Seattle, WA 98195, USA
[15]Institute of Physics, Laboratoire d'Astrophysique, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland
[16]Particle Astrophysics Center, Fermi National Accelerator Laboratory, Batavia, IL 60510, USA
[17]Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA
[18]Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA
[19]Department of Astronomy, University of Maryland, College Park, MD 20742, USA
[20]Kavli Institute for Cosmology, University of Cambridge, Cambridge, CB3 0HA, United Kingdom
[21]Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh, EH9 3HJ, United Kingdom
[22]National Center for Supercomputing Applications, University of Illinois, Urbana, IL 61801, USA
[23]School of Information Sciences, Department of Astronomy, University of Illinois, Urbana, IL 61801, USA
[24]Department of Astronomy, New Mexico State University, Las Cruces, NM 88001, USA
[25]School of Physics and Astronomy, Rochester Institute of Technology, Rochester, NY 14623, USA
[26]Center for Astrophysics and Planetary Science, Racah Institute of Physics, The Hebrew University, Jerusalem, 91904, Israel
[27]Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA
[28]Department of Astronomy and Astrophysics, University of California at Santa Cruz, Santa Cruz, CA 95064, USA
[29]Department of Astronomy, Yale University, New Haven, CT 06520, USA
[30]Department of Earth and Space Science, Graduate School of Science, Osaka University, Toyonaka, Osaka, 560-0043, Japan
[31]Department of Physics and Astronomy, University of Nevada, Las Vegas, NV 89154, USA
[32]Department of Computational Mathematics, Science and Engineering, Department of Physics and Astronomy, National Superconducting Cyclotron Laboratory, Michigan State University, Lansing, MI 48824, USA
[33]Department of Physics, University of California at Santa Cruz, Santa Cruz, CA 95064, USA
[34]Heidelberger Institut für Theoretische Studien, 69118 Heidelberg, Germany
[35]Department of Physics and Astronomy, University of Kansas, Lawrence, KS 66045, USA
[36]Center for Relativistic Astrophysics, School of Physics, Georgia Institute of Technology, Atlanta, GA 30332, USA
[37]http://www.AGORAsimulations.org/
[†]These authors, in alphabetical order, contributed to the article by leading the effort within each code group to perform and analyze simulations. and
[‡]These authors, in alphabetical order, contributed to the article by developing GRACKLE and implementing its interface with participating codes.

Draft version December 20, 2016

ABSTRACT

Using an isolated Milky Way-mass galaxy simulation, we compare results from 9 state-of-the-art gravito-hydrodynamics codes widely used in the numerical community. We utilize the infrastructure we have built for the *AGORA* High-resolution Galaxy Simulations Comparison Project. This includes the common disk initial conditions, common physics models (e.g., radiative cooling and UV background by the standardized package GRACKLE) and common analysis toolkit `yt`, all of which are publicly available. Subgrid physics models such as Jeans pressure floor, star formation, supernova feedback energy, and metal production are carefully constrained across code platforms. With numerical accuracy that resolves the disk scale height, we find that the codes overall agree well with one another in many dimensions including: gas and stellar surface densities, rotation curves, velocity dispersions, density and temperature distribution functions, disk vertical heights, stellar clumps, star formation rates, and Kennicutt-Schmidt relations. Quantities such as velocity dispersions are very robust (agreement within a few tens of percent at all radii) while measures like newly-

formed stellar clump mass functions show more significant variation (difference by up to a factor of $\sim$3). Systematic differences exist, for example, between mesh-based and particle-based codes in the low density region, and between more diffusive and less diffusive schemes in the high density tail of the density distribution. Yet intrinsic code differences are generally small compared to the variations in numerical implementations of the common subgrid physics such as supernova feedback. Our experiment reassures that, if adequately designed in accordance with our proposed common parameters, results of a modern high-resolution galaxy formation simulation are more sensitive to input physics than to intrinsic differences in numerical schemes.

Keywords: cosmology: theory – galaxies: formation – galaxies: evolution – galaxies: kinematics and dynamics – ISM: structure – methods: numerical – hydrodynamics

1. INTRODUCTION

Decades of strenuous effort by computational astrophysicists have propelled numerical experiments to become one of the most widely used tools in theorizing how galaxies form in the Universe. Numerical experiments are often the only means to put our theory to a test, the result of which we can compare with observational data to validate the model's feasibility. Since the success of galaxy formation theory is predicated on robust numerical experiments, it is only reasonable that we apply the same scientific standard of reproducibility to galaxy formation simulations. In other words, it should be considered as a fundamental principle that researchers must not establish findings from a single numerical experiment as scientific knowledge. Only after the result is reproduced independently by other researchers and proven not to be an isolated incidence can we build any conclusive theory about how galaxies actually form in the Universe.

However, the task of replicating galaxy simulations or, equivalently, comparing simulations between codes, has not received high priority.[1] Instead, the task is considered complex and time-consuming because one needs to ensure that identical physics is used in an identical initial condition with identical runtime settings. This is sometimes perceived as tedious and unrewarding for early-career researchers. In fact, the lack of reproducibility checks is not unique to the field of numerical galaxy formation (e.g., Open Science Collaboration 2015; Nature Survey 2016). And its cause is not simply an unwillingness of only computational astrophysicists, either (Everett & Earp 2015). Rather, addressing the system (or the lack thereof) which checks the reproducibility of simulations would require a collective action by the entire community. It cannot be simply about asking individual researchers to release their data dumps, but it should be about building a system that incentivizes simulations published in a reproducible manner. It should also be about assembling an infrastructure that reduces the cost of reproducibility checks, on which simulations are verified routinely and effortlessly (Nosek et al. 2015; Begley & Ioannidis 2015).

The *AGORA* High-resolution Galaxy Simulations Comparison Project (*Assembling Galaxies Of Resolved Anatomy*) is the collective response by the numerical galaxy formation community to such a challenge. Since its first meeting in 2012 at the University of California at Santa Cruz, the *AGORA* Collaboration has aimed to compare galaxy-scale numerical experiments on a variety of code platforms with state-of-the-art resolution. Our shared goal is to ensure that physical assumptions are responsible for any success in galaxy formation simulations, rather than artifacts of particular implementations. Through a multi-platform approach from the beginning, we strive to improve all our codes by "increasing the level of realism and predictive power of galaxy simulations and the understanding of the feedback processes that regulate galaxy metabolism" (Kim et al. 2014), and by doing so to find solutions to long-standing problems in galaxy formation. Because the interplay between numerical resolution and subgrid modelings of stellar physics is crucial in galaxy-scale simulations, we require that simulations be designed with state-of-the-art resolution, $\lesssim$ 100 pc, which is currently allowed within realistic computational cost bounds.

In the Project's flagship paper, Kim et al. (2014), we explained the philosophy behind the Project and detailed the publicly available Project infrastructure we have put together. We also described the proof-of-concept test, in which we field-tested our infrastructure with a dark matter-only cosmological zoom-in simulation, finding a robust convergence between participating codes. More than 140 researchers from over 60 academic institutions worldwide have since agreed to take part in the Collaboration, many of whom having been actively engaged in working groups and sub-projects.[2] The cohort of numerical codes participating in the Project currently include, but are not limited to in future studies: the Lagrangian smoothed particle hydrodynamics codes (SPH; Gingold & Monaghan 1977; Lucy 1977; Monaghan 1992) CHANGA, GADGET, GASOLINE, and GEAR, and the Eulerian adaptive mesh refinement codes (AMR; Berger & Oliger 1984; Berger & Colella 1989) ART-I, ART-II, ENZO, and RAMSES, and the mesh-free finite-volume Godunov code GIZMO (see Section 5 for information on each code).

In this second report of our continuing endeavor, we use an isolated Milky Way-mass galaxy simulation to compare 9 widely used state-of-the-art gravito-hydrodynamics codes. As in all comparison studies in *AGORA*, the participating codes share the common initial condition (i.e., generated by MAKEDISK; see Section 2), common physics models (e.g., radiative cooling and UV background provided by the standardized package GRACKLE; see Section 3.1; Smith et al. 2016; Bryan et al. 2014; Kim et al. 2014),[3] and common analysis platform (i.e, `yt` toolkit; Turk et al. 2011).[4] We adopt spatial resolution of 80 pc that resolves the scale height of the disk. This helps the codes to be less dependent on phenomenological prescriptions of sub-resolution processes which are inevitably introduced in low-resolution ($>$ kpc) simulations. As modern galaxy formation simulations with state-of-the-art resolution and physics prescriptions become more and more computationally expensive, it is timely that we compare high-

[1] Code comparisons in the astrophysical community have previously been undertaken, albeit with simplified physics in a different scale (e.g., Frenk et al. 1999; O'Shea et al. 2005), or focusing only on hydrodynamics solvers (e.g., Agertz et al. 2007; Tasker et al. 2008).

[2] See the Project website at `http://www.AGORAsimulations.org/` for more information on the Project including its membership, and its task-oriented and science-oriented working groups.

[3] The website is http://grackle.readthedocs.org/.

[4] The website is http://yt-project.org/.

Table 1
Initial Condition Characteristics

	Dark matter halo	Stellar disk	Gas disk	Stellar bulge
Density profile	Navarro et al. (1997)	Exponential	Exponential	Hernquist (1990)
Structural properties	$M_{200} = 1.074 \times 10^{12}\,\mathrm{M}_\odot$, $v_{\mathrm{c},200} = 150\,\mathrm{km\,s^{-1}}$, $R_{200} = 205.5$ kpc, $c = 10$, $\lambda = 0.04$	$M_{\mathrm{d},\star} = 3.438 \times 10^{10}\,\mathrm{M}_\odot$, $r_\mathrm{d} = 3.432$ kpc, $z_\mathrm{d} = 0.1\,r_\mathrm{d}$	$M_{\mathrm{d,gas}} = 8.593 \times 10^{9}\,\mathrm{M}_\odot$, $f_\mathrm{gas} = 0.2$	$M_{\mathrm{b},\star} = 4.297 \times 10^{9}\,\mathrm{M}_\odot$, $M_{\mathrm{b},\star}/M_\mathrm{d} = 0.1$
Number of particles	10^5	10^5	10^5	1.25×10^4
Particle mass	$m_\mathrm{DM} = 1.254 \times 10^7\,\mathrm{M}_\odot$	$m_{\star,\mathrm{IC}} = 3.437 \times 10^5\,\mathrm{M}_\odot$	$m_{\mathrm{gas,IC}} = 8.593 \times 10^4\,\mathrm{M}_\odot$	$m_{\star,\mathrm{IC}} = 3.437 \times 10^5\,\mathrm{M}_\odot$

resolution isolated disk simulations to check how successfully these galaxies are reproduced by their peers.[5] Readers should note that our intention is *not* to identify a "correct" or "incorrect" code, but to focus instead on juxtaposing the codes for physical insights and learn how much scatter one should expect among modern numerical tools in the field (see Section 7 for more discussion).

The remainder of this article is organized as follows. In Section 2 we explain the isolated disk initial condition used in the study. The common input physics and runtime parameters required in the participating codes are discussed in Section 3 and 4, respectively. Then Section 5 describes 9 hydrodynamics codes that participated in this comparison. Section 6 presents the results of our comparison focusing on similarities and discrepancies discovered in various multi-dimensional analyses. Finally in Section 7 we summarize our findings and conclude the paper with remarks on future work. We will also stress the importance of collaborative and reproducible research in the numerical galaxy formation community the *AGORA* Project strives to promote.

2. INITIAL CONDITION

In this section we describe the Milky Way-mass isolated initial condition (IC) we adopt in this study. While this IC is part of a set of disk ICs generated for *AGORA* simulations that were first introduced in Section 2.2 of the Project flagship paper (Kim et al. 2014), we briefly explain its important structural properties for completeness.[6]

The disk galaxy IC with properties characteristic of Milky Way-mass galaxies at redshift $z \sim 1$ is generated with a privately shared version of MAKEDISK (Springel et al. 2005).[7,8] The IC has the following components (see also Table 1 and Figure 1): (1) a dark matter halo with $M_{200} = 1.074 \times 10^{12}\,\mathrm{M}_\odot$, $R_{200} = 205.5$ kpc and circular velocity of $v_{\mathrm{c},200} = 150\,\mathrm{km\,s^{-1}}$ that follows the Navarro-Frenk-White (NFW; Navarro et al. 1997) profile with concentration parameter $c = 10$ and spin parameter $\lambda = 0.04$, (2) an exponential disk with $M_\mathrm{d} = 4.297 \times 10^{10}\,\mathrm{M}_\odot$, scale length $r_\mathrm{d} = 3.432$ kpc and scale height $z_\mathrm{d} = 0.1\,r_\mathrm{d}$ that is composed of 80% stars and 20% gas in mass (i.e., $f_\mathrm{gas} = M_{\mathrm{d,gas}}/M_\mathrm{d} = 0.2$), (3) a stellar bulge with $M_{\mathrm{b},\star} = 4.297 \times 10^9\,\mathrm{M}_\odot$ that follows the Hernquist profile ($M_{\mathrm{b},\star}/M_\mathrm{d} = 0.1$; Hernquist 1990). The disk or bulge stars in the IC do not contribute to the feedback budget.

Among the three resolution choices provided in Kim et al. (2014) here we employ a "low-resolution" IC which has 10^5 particles each for the halo, the stellar disk, and the gas disk, and 1.25×10^4 particles for the bulge. The initial gas temperature in the disk is set to 10^4 K, not to the specific internal energy computed by MAKEDISK. The initial metal fraction in the gas disk is 0.02041.[9] When the gas disk is initialized on mesh-based codes (ART-I, ART-II, ENZO, and RAMSES), instead of using the particles provided by MAKEDISK, we require that the participants use an analytic density profile of

$$\rho_{\mathrm{d,gas}}(r, z) = \rho_0 e^{-r/r_\mathrm{d}} \cdot e^{-|z|/z_\mathrm{d}} \quad (1)$$

with $\rho_0 = M_{\mathrm{d,gas}}/(4\pi r_\mathrm{d}^2 z_\mathrm{d})$, where r is the cylindrical radius and z is the vertical height from the disk plane. To set up a disk in a centrifugal equilibrium we also ask that the participants utilize the rotational velocity profile binned from an actual gas particle distribution within $|z| < z_\mathrm{d}$.[10] In mesh-based codes, we additionally include a uniformly low density gas halo with $n_\mathrm{H} = 10^{-6}\,\mathrm{cm^{-3}}$ and zero initial velocity, since they cannot have cells with zero density. The halo is initially set to 10^6 K and zero metallicity. Note that this gaseous halo does not exist in particle-based codes (SPH codes or GIZMO).

It is worth to note one point about the common disk IC adopted here. As readers may find in Section 6.6, in our experiment that includes radiative cooling, star formation and feedback (*"Sim-SFF"*; see Section 3), only $\sim 10^9\,\mathrm{M}_\odot$ additional stars form in 500 Myr in all codes on average. When compared with the stellar and gas components present in the IC, this means only a $\sim 3\%$ increase in stellar mass, and a $\sim 12\%$ decrease in gas mass. As the discussion in Section 6 should make clear, given this relatively small change in stellar mass, it is expected that the stellar feedback in our experiment will be very inefficient.

[5] Comparisons of cosmological zoom-in simulations are also in the making to test the robustness of the code suite over 13.8 Gyr of evolution. See the Project's flagship paper, (Kim et al. 2014), for more information.

[6] The public Dropbox link is http://goo.gl/8JzbIJ.

[7] MAKEDISK is an earlier realization of a code similar to GALIC (Yurin & Springel 2014). GALIC is publicly available, and its website is http://www.h-its.org/tap-software-en/galic-code/.

[8] While the Milky Way's f_gas is $\sim 10\%$, typical galaxies with the Milky Way stellar mass at $z \sim 0$ have $f_\mathrm{gas} \sim 20\%$ (Catinella et al. 2010). In this regard, one can say that we model a more typical galaxy at $z \sim 0$ than the Milky Way.

[9] This fractional value 0.02041 corresponds to 1 $Z_\odot$ for GRACKLE v2.0, but to 1.5761 $Z_\odot$ for GRACKLE v2.1 or above. It is because the solar metallicity unit $Z_\odot$ was updated from 0.02041 to 0.01295 in GRACKLE v2.1. Since cooling rates pre-tabulated by CLOUDY are at 1 $Z_\odot$, not at specific *metal fraction* value, the cooling rates in GRACKLE's equilibrium cooling mode will differ depending on which GRACKLE version is adopted (see Section 3.1 for more on GRACKLE). For example, in the current study, the codes using GRACKLE v2.1 (CHANGA, GASOLINE, GADGET-3, and GIZMO) show slightly enhanced cooling rates than the ones using GRACKLE v2.0 or below (ART-I, ART-II, ENZO, RAMSES, and GEAR). Generally speaking, initial gas metallicity should be set up so that it is consistent with the chosen GRACKLE version interfacing with the code. We refer interested readers to the GRACKLE v2.1 release note at https://goo.gl/BNRfwJ.

[10] This *actual* initial velocity profile, provided in `vcirc_SPH.dat` in our public Dropbox link, is different from the file `vcirc.dat` produced by MAKEDISK itself. The difference is $\sim 5\%$ in the central few kpc.

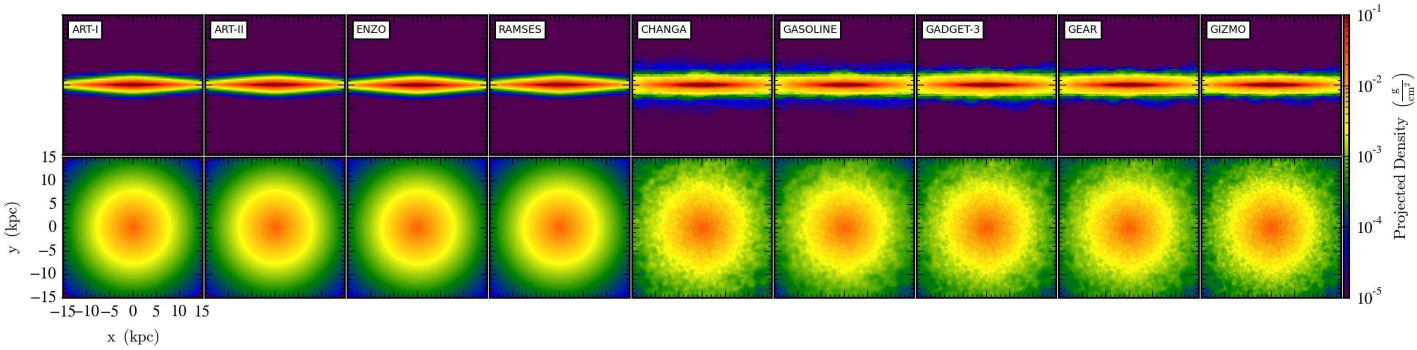

Figure 1. The 0 Myr snapshots of the isolated Milky Way-mass galaxy simulations by 9 participating codes. Disk gas surface densities in a 30 kpc box, edge-on *(top)* and face-on *(bottom)*, produced with the common analysis toolkit `yt`. For visualizations of the particle-based codes hereafter (Figures 1-3, 14-15, 32, 34, 35) – but not in any other analyses except these figures – `yt` uses an in-memory octree on which gas particles are deposited using smoothing kernels. Comparing 0 Myr snapshots – dumped immediately after each code reads in the IC – is to check the exact identity of ICs interpreted by each code. See Section 6.1 for more information on this figure, and Section 5 for descriptions of participating codes in this comparison. The full color version of this figure is available in the electronic edition. The high-resolution versions of this figure and article are available at the Project website, http://www.AGORAsimulations.org/.

3. COMMON PHYSICS: SIM-NOSF AND SIM-SFF

We now describe the common physics employed in our experiment including equilibrium gas cooling, metagalactic UV background, star formation, and energy and metal yields by supernovae. Note that the common physics adopted here is a variation of the common physics model recommended in all *AGORA* simulations by default; see Section 3 of the Project flagship paper (Kim et al. 2014). For the present study all participating code groups are asked to run two simulations starting from the identical IC: (1) *"Sim-noSF"* with radiative gas cooling but without star formation or feedback, and (2) *"Sim-SFF"* with radiative cooling, star formation and feedback. In Section 3.1, we first list the gas physics that are common in both *Sim-noSF* and *Sim-SFF*. Then in Section 3.2, the subgrid prescriptions of stellar physics for *Sim-SFF* are explained.

3.1. *Gas Physics: Radiative Cooling, UV Background, and Pressure Floor*

The rate at which the gas in our galaxy radiatively cools is determined by *AGORA*'s standard chemistry and cooling library GRACKLE (Smith et al. 2016; Bryan et al. 2014; Kim et al. 2014).[3] For this study, the *equilibrium* cooling version of GRACKLE is interfaced with each participating code, either via GRACKLE's original interface or via N. Gnedin's auxiliary API.[11] In the chosen equilibrium cooling mode, GRACKLE follows tabulated cooling rates pre-computed by the photoionization code CLOUDY (Ferland et al. 2013).[12] The pre-computed look-up table also includes metal cooling rates for solar abundances, 1 $Z_{\odot}$, as a function of gas number density and temperature. These metal cooling rates are then scaled linearly with metallicity which is followed in our simulations as a separate passive scalar.[13] We also adopt metagalactic UV background radiation at $z = 0$ by Haardt & Madau (2012) provided by GRACKLE. For the difference between the chosen UV background model and previous calculations such as Haardt & Madau (1996) or Faucher-Giguère et al. (2009), we refer the readers to Section 3.3 of Kim et al. (2014).

Lastly, a non-thermal Jeans pressure floor is applied to stabilize the scales of the smoothing length (particle-based codes) or the finest cell (mesh-based codes) against unphysical collapse and to avoid artificial fragmentation due to unresolved pressure gradient (Truelove et al. 1997; Robertson & Kravtsov 2008). In practice, it is achieved by enforcing that the local Jeans length $\lambda_{\rm Jeans}$ be sufficiently resolved with the finest resolution elements at all times. That is,

$$\lambda_{\rm Jeans} = N_{\rm Jeans}\Delta x \tag{2}$$

where $\Delta x = 80$ pc is the adopted spatial resolution (finest cell size or softening length; see Section 4.1) and $N_{\rm Jeans} = 4$ is the Jeans number adopted from Truelove et al. (1997). This gives the required pressure floor value as

$$P_{\rm Jeans} = \frac{1}{\gamma\pi} N_{\rm Jeans}^2 G\rho_{\rm gas}^2 \Delta x^2 \tag{3}$$

where G is the gravitational constant, $\gamma = 5/3$ is the adiabatic index, and $\rho_{\rm gas}$ is the gas density. Note that $N_{\rm Jeans}$ is not necessarily equal to the parameter controlling the pressure support in each code. For actual parameter choices for selected codes, see Appendix A. For implementations using polytropes in ART-II and RAMSES, see Sections 5.2 and 5.4, respectively.

3.2. *Stellar Physics: Star Formation, and Energy, Mass, Metal Yields From Core-Collapse Supernovae*

In addition to the gas physics described in the previous section, *Sim-SFF* incorporates subgrid models for star formation and supernova feedback. First, a parcel of gas above the threshold $n_{\rm H,\,thres} = 10\ {\rm cm}^{-3} = \rho_{\rm gas,\,thres}/m_{\rm H}$ produces stars at a rate that follows the local Schmidt law as

$$\frac{d\rho_\star}{dt} = \frac{\varepsilon_\star \rho_{\rm gas}}{t_{\rm ff}} \tag{4}$$

where $\rho_\star$ is the stellar density, $t_{\rm ff} = (3\pi/(32G\rho_{\rm gas}))^{1/2}$ is the local free-fall time, and $\varepsilon_\star = 1\%$ is the star formation efficiency per free-fall time. We caution that the parameter $\varepsilon_\star$ is not necessarily equal to the star formation efficiency parameter found in each code (e.g., for CHANGA and GASOLINE, see Appendix B). For a new star particle to spawn, it should have at least the mass of a gas particle in the IC, $m_{\rm gas,\,IC} = 8.593 \times 10^4\,{\rm M}_\odot$. Note that $n_{\rm H,\,thres}$ adopted in this experiment is for this particular run only, and represents where the Jeans polytrope intersects with a typical $T - \rho$ equation of state in

[11] The website is https://bitbucket.org/gnedin/agora_api/.
[12] The website is http://www.nublado.org/.
[13] See, however, footnote 9 on how a different version of GRACKLE may affect the cooling rates for the gas with the same *metal fraction* (but not the same *metallicity* interpreted by CLOUDY).

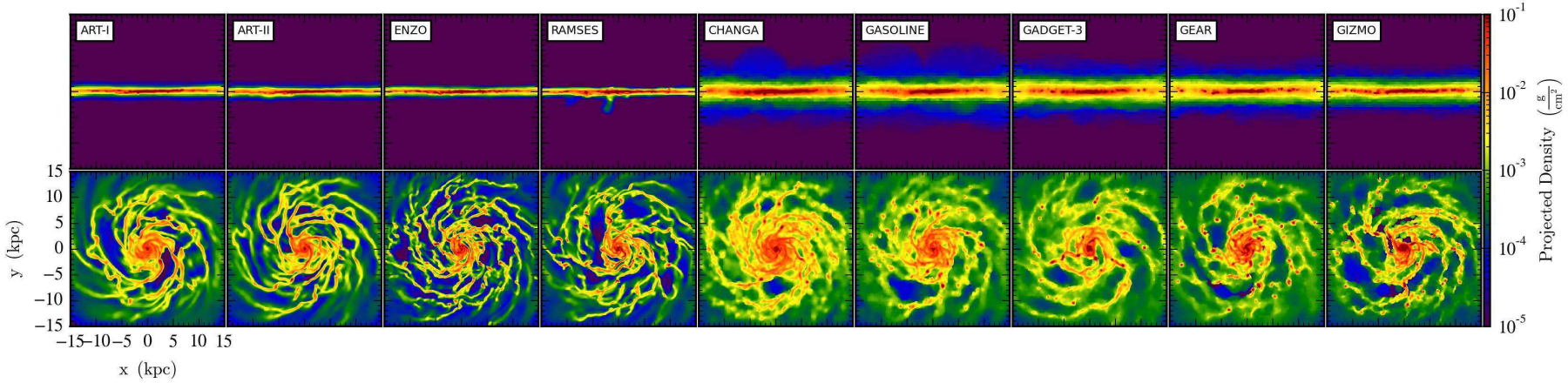

Figure 2. The 500 Myr composite of gas surface densities from *Sim-noSF* with radiative gas cooling but without star formation or supernova feedback. Each frame is centered on the galactic center – location of maximum gas density within 1 kpc from the center of gas mass. For visualizations of the particle-based codes hereafter (Figures 1-3, 14-15, 32, 34, 35) – but not in any other analyses except these figures – `yt` uses an in-memory octree on which gas particles are deposited using smoothing kernels. See Section 5 for descriptions of participating codes in this comparison, and Section 6.1 for a detailed explanation of this figure. Compare with Figure 14. Simulations performed by: Daniel Ceverino (ART-I), Robert Feldmann (ART-II), Mike Butler (ENZO), Romain Teyssier (RAMSES), Spencer Wallace (CHANGA), Ben Keller (GASOLINE), Jun-Hwan Choi (GADGET-3), Yves Revaz (GEAR), and Alessandro Lupi (GIZMO). The full color version of this figure is available in the electronic edition. The high-resolution versions of this figure and article are available at the Project website, http://www.AGORAsimulations.org/.

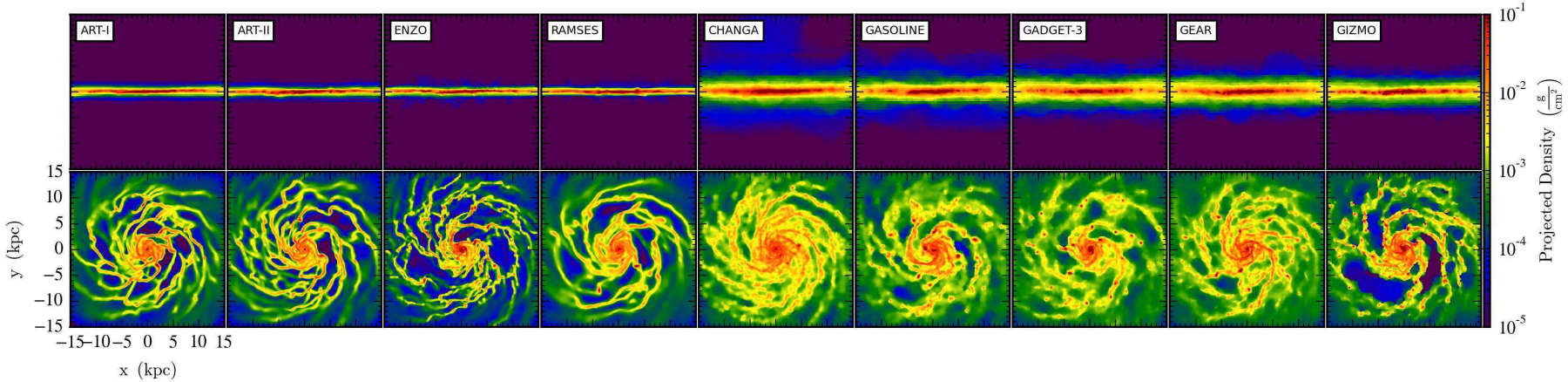

Figure 3. Same as Figure 2 but for *Sim-SFF* with star formation and feedback. See Section 6.1 for a detailed explanation of this figure. See also Section 3.2 for the common star formation prescription and the guideline for supernova feedback, and Section 5 for the exact deposit scheme of thermal feedback energy implemented in each code. Compare with Figures 15, 21, 29, 32, 34, and 35.

our disks (see Sections 5.2 and 5.4).[14]

New star particles inject energy, mass, and metals back into the interstellar medium (ISM) through core-collapse (Type II) supernovae. Assuming the *AGORA* standard Chabrier (2003) initial mass function (IMF) and that stars with masses between 8 and 40 $M_\odot$ explode as Type II supernovae, one Type II supernova occurs per every 91 $M_\odot$ stellar mass formed (see Section 3.5 of Kim et al. 2014). With the *AGORA* recommended fitting formulae Eqs. (4)-(6) of Kim et al. (2014) and the assumed IMF, this single burst is found to release 2.63 $M_\odot$ of metals[15] and 14.8 $M_\odot$ of gas (including metals). Per every 91 $M_\odot$ stellar mass, these metal and mass are instantaneously deposited into its surrounding after a delay time of 5 Myr, along with a net thermal energy of 10^{51} ergs.

We note that exact deposit schemes for energy, mass, and metals are left at each participant's discretion. We do not intend to overly specify a single *common* deposit scheme which will need to be inevitably different from one code to another (e.g., between mesh-based codes and particle-based codes), as we argued in Section 3.8 of the Project flagship paper (Kim et al. 2014). Nevertheless, for all mesh-based codes (ART-I, ART-II, ENZO, and RAMSES), the same strategy was chosen: thermal energy, mass, and metals are added to the cell where a 5 Myr old star particle sits at the time of explosion, and to this cell only. For particle-based codes (CHANGA, GASOLINE, GADGET-3, GEAR, and GIZMO), each code's deposit scheme is discussed in detail in Section 5. In future *AGORA* projects, we plan to calibrate different feedback schemes against observations and against one another. We refer the readers to Section 7 for more discussion on this future work.

4. COMMON RUNTIME PARAMETERS

Here we review the runtime parameters each group is required to adopt, such as gravitational softening and hydrodynamic smoothing lengths for particle-based codes and refinement thresholds for mesh-based codes.

4.1. *Gravitational Softening Length and Finest Mesh Size*

For all codes the gas mass resolution in hydrodynamics needs to be set as close as possible to $m_{\rm gas,IC} = 8.593 \times 10^4\,M_\odot$. Assuming that we wish to resolve a self-gravitating clump with 64 of these resolution elements, the corresponding Jeans length scale becomes

$$\lambda_{\rm Jeans} = 2\left(\frac{64\,m_{\rm gas,IC}}{(4\pi/3)\,n_{\rm H,thres}}\right)^{1/3} = 348.7\,{\rm pc}\,, \qquad (5)$$

and therefore, from Eq. (2) we choose a spatial resolution of 80 pc. This value is used as the finest cell size Δx for mesh-based codes, and as the gravitational softening length $\varepsilon_{\rm grav}$ for particle-based codes. For all particle-based codes taking part in the present study, gravity is softened according to the cubic spline kernel (e.g., Eq. (A1) of Hernquist & Katz 1989). For readers interested in the actual parameter choices, in Appendix C we examine the meanings of relevant parameters in different particle-based codes.

[14] As noted in Kim et al. (2014), star formation prescription parameters such as $n_{\rm H,thres}$ or $\varepsilon_\star$, the initial mass of star particles, and the stochasticity of star formation, are all highly dependent on numerical resolution. An idealized test like the disk simulation presented here is essential to tune up such parameters for computationally expensive cosmological simulations.

[15] Per unit stellar mass formed, the total fractional ejected metal masses (oxygen and iron combined) is $M_Z = 2.09\,M_{\rm O} + 1.06\,M_{\rm Fe} = 2.09 \times 0.0133 + 1.06 \times 0.0011 = 2.9\%$.

4.2. *Minimum Hydrodynamical Smoothing Length*

For particle-based codes (including GIZMO; see Section 5.9 and footnote 28), we require that the hydrodynamical smoothing lengths for collisional particles do not drop below 20% of the gravitational softening lengths. Unlike the gravitational softening kernel, exact smoothing kernel choices differ from code to code, and are detailed for each of the particle-based codes in Section 5. We also refer the readers interested in the actual parameter choices to Appendix C again.

4.3. *Refinement Strategy*

We recommend to mesh-based code groups that a cell be split into 8 child cells once the cell contains more mass than $m_{\rm gas,IC} = 8.593 \times 10^4\,M_\odot$ (1 gas particle mass in the IC of particle-based codes), or 8 collisionless particles (disk/bulge star particles in the IC with $m_{\star,\rm IC} = 3.437 \times 10^5\,M_\odot$, or dark matter particles with $m_{\rm DM} = 1.254 \times 10^7\,M_\odot$). This causes the grids to be refined in a fashion similar to the Lagrangian behavior of particle-based codes, and keeps the ratio of collisionless particle numbers to gas cells approximately unity on average. However, exact refinement strategies differ slightly from code to code, and are detailed for each of the mesh-based codes in Section 5.[16] We continue to refine the grids down to the resolution limit $\Delta x = 80$ pc (see Section 4.1) where the non-thermal pressure floor kicks in (see Section 3.1).

5. PARTICIPATING CODES

In this section we introduce the 9 gravito-hydrodynamics codes taking part in this test, focusing in particular on hydrodynamics solvers, refinement schemes for mesh-based codes (ART-I, ART-II, ENZO, and RAMSES), and supernova feedback implementations for particle-based codes (CHANGA, GASOLINE, GADGET-3, GEAR, and GIZMO). We leave out details that are commonly adopted across platforms such as gas cooling (Section 3.1) or star formation (Section 3.2), or that were included in the *AGORA* flagship paper such as gravitational dynamics (Section 5 of Kim et al. 2014). We also point out that the codes involved in future *AGORA* studies are not necessarily limited to the ones described herein.

5.1. ART-I

In ART-I, differential equations of fluid dynamics are integrated using a shock-capturing Eulerian method described in Khokhlov (1998). It uses a 2nd-order accurate Godunov solver (Godunov 1959) that evaluates Eulerian fluxes by solving the Riemann problem at every cell interface (Colella & Glaz 1985). Left and right states of the Riemann problem are obtained by piecewise linear interpolation (van Leer 1979). In contrast to other versions of ART (see Section 5.2.1 of Kim et al. 2014), ART-I with distinctive star formation and feedback recipes (e.g., Ceverino & Klypin 2009; Ceverino et al. 2014) have been developed by A. Klypin and collaborators.

The octree-based, multi-level adaptive mesh allows users to control the grid structure at the individual cell level. For this comparison, the ART-I group uses a 128^3 root grid covering a $(1.304\ {\rm Mpc})^3$ box, then achieves an ~ 80 pc cell size at maximum 7 levels of refinement. The mass thresholds above

[16] Given differences in refinement machineries among mesh-based codes it is impractical, if not impossible, to impose an exactly identical refinement criterion across all codes. We instead adopt a trial-and-error approach within the guideline presented in Section 4.3, which resulted in all mesh-codes eventually converging to a similar overall grid structure.

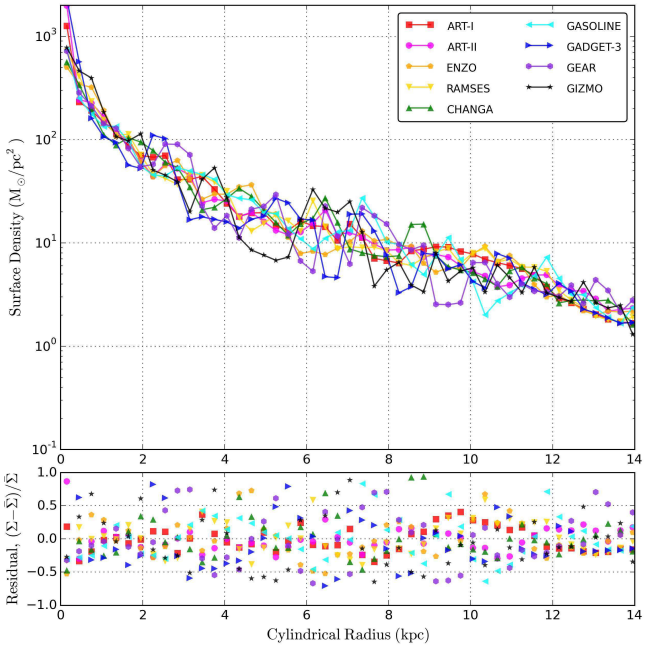

Figure 4. Cylindrically-binned gas surface density profiles at 500 Myr for *Sim-noSF* without star formation or feedback. The cylindrical radius is from the galactic center – location of maximum gas density within 1 kpc from the center of gas mass. In all analyses for particle-based codes hereafter – except the graphical visualizations such as Figures 1-3 – raw particle fields are used, not the interpolated or smoothed fields constructed in `yt`. Shown in the bottom panel is the fractional deviation from the mean of these profiles. See Section 6.1 for more information on this figure. The *y*-axis range of the top panel is kept identical among Figures 4-7 and 22 for easier comparison.

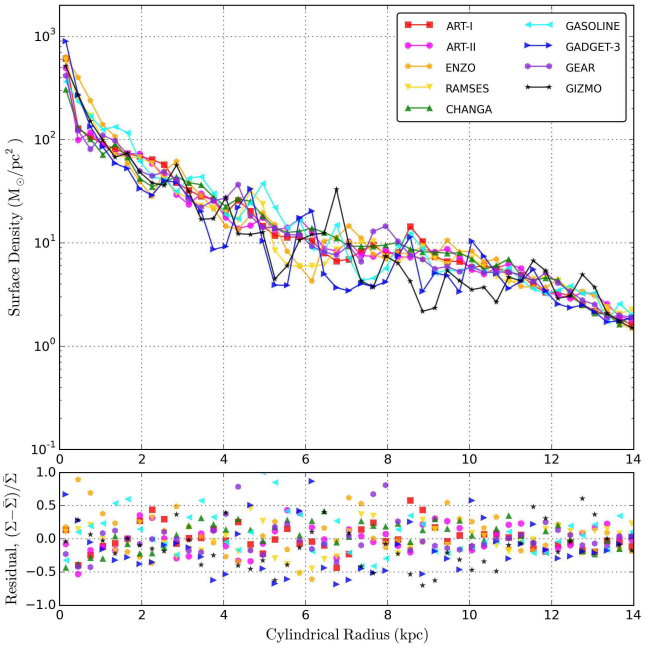

Figure 5. Same as Figure 4 but for *Sim-SFF* with star formation and feedback. Compare with Figures 22 and 27.

which a cell is adaptively refined into an oct of 8 child cells are $m_{\rm gas,IC} = 8.593 \times 10^4 \, {\rm M}_\odot$ and $m_{\star,\rm IC} = 3.437 \times 10^5 \, {\rm M}_\odot$ for gas and collisionless particles, respectively.[17] For the supernova feedback scheme to deposit the thermal energy adopted by all mesh-based codes, we refer the readers to Section 3.2.

5.2. ART-II

ART-II solves the gravito-hydrodynamics equations using a particle-mesh + Eulerian AMR approach. ART-II features MPI parallelization for distributed memory machines, flexible time-stepping hierarchy, and a variety of unique physics modules (e.g., Gnedin & Kravtsov 2011; Agertz et al. 2013) developed by N. Gnedin, A. Kravtsov and collaborators.

For the present study, starting from a uniform 128^3 root grid covering $(1.311 \text{ Mpc})^3$, cells are refined up to 7 additional levels to reach the finest size of 80 pc. Spherical regions of 4 (6, 10) root grid cells radius around the box center are always refined to at least 3 (2, 1) additional levels relative to the root grid. The (de-)refinement procedure consists of three steps. First, cells are marked for refinement if the gas mass in the cell exceeds $0.6 m_{\rm gas,IC} = 0.6 \times 8.593 \times 10^4 {\rm M}_\odot$, or if the cell contains 2 or more dark matter and/or star particles that were present in the IC. We then use a diffusion step to also mark neighboring cells for refinement and thus smooth the shape of the regions to be refined. By contrast, cells with gas masses below $0.2 m_{\rm gas,IC}$ or without particles are marked for de-refinement provided they also satisfy a number of additional constraints. Finally, cells are refined (de-refined) by splitting them into 8 (by merging 8 child cells).

The pressure floor implemented in ART-II affects cells at the highest level of refinement by modifying the gas pressure values that enter the Riemann solver (i.e., not the actual pressure or temperature fields) with

$$P_{\rm cell} = \texttt{max}(P_{\rm Jeans}, \, P_{\rm gas}) \quad (6)$$

$$= \texttt{max}(n_{\rm H} k_{\rm B} T_{\rm Jeans}, \, P_{\rm gas}) \quad (7)$$

where $P_{\rm cell}$ is the value entering the Riemann solver, $P_{\rm gas}$ is the gas pressure field in the simulation, $k_{\rm B}$ is the Boltzmann constant, $n_{\rm H} = \rho_{\rm gas}/m_{\rm H}$, and $T_{\rm Jeans} = T_{\rm J}(n_{\rm H}/n_{\rm H,J})$ with $T_{\rm J} = 1800$ K and $n_{\rm H,J} = 8 \text{ cm}^{-3}$. This polytrope choice is designed to match the common prescription Eq. (3) with $N_{\rm Jeans} \simeq 4$. For the supernova feedback scheme to deposit the thermal energy adopted by all mesh-based codes, see Section 3.2.

5.3. ENZO

ENZO is a block-structured adaptive mesh code, developed by an open-source, community-driven approach (Bryan & Norman 1997; O'Shea et al. 2004; Bryan et al. 2014).[18] Among a variety of solver choices, for this comparison the 3rd-order accurate piecewise parabolic method (PPM) is selected to reconstruct the left and right states of the Godunov problem (Colella & Woodward 1984; Bryan et al. 1995), along with a Harten-Lax-van Leer with Contact (HLLC) Riemann solver (Toro et al. 1994). A maximum 30% of the required Courant-Friedrichs-Lewy (CFL) timestep is used to advance fluid elements; i.e., CFL safety factor = 0.3. In addition

[17] We note that ART-I and ENZO cannot refine cells by particle numbers, but only by particle masses. By contrast, in the reported runs, ART-II and RAMSES refine cells by particle numbers. The refinement criteria are chosen to ensure an agreement among mesh-based codes in overall grid structures. See also footnote 16.

[18] The website is http://enzo-project.org/.

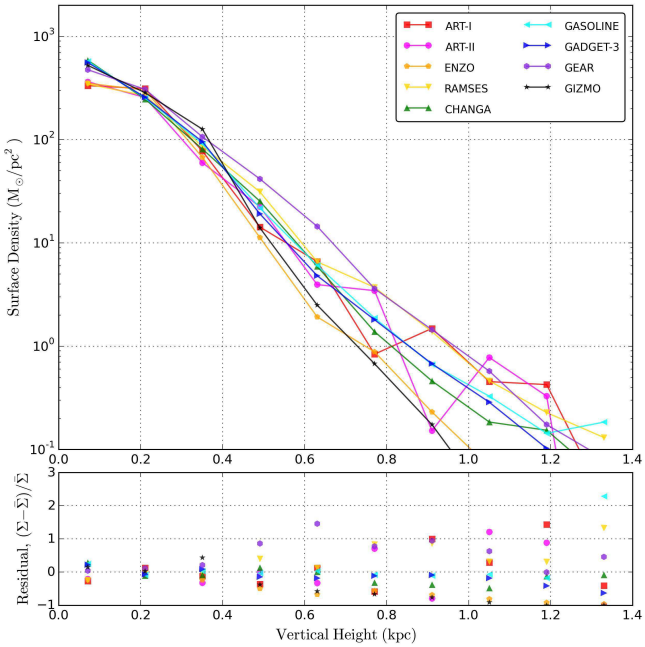

Figure 6. Vertically-binned gas surface density profiles at 500 Myr for *Sim-noSF* without star formation or feedback. The height is the absolute vertical distance from the $x-y$ disk plane centered on the galactic center, $|z_i - z_{\rm center}|$. Shown in the bottom panel is the fractional deviation from the mean of these profiles. The y-axis range of the top panel is kept identical among Figures 4-7 and 22 for easier comparison.

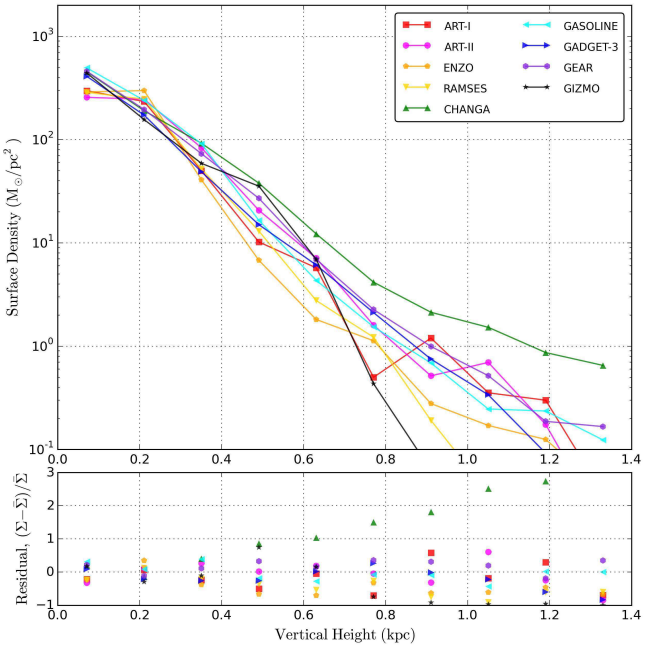

Figure 7. Same as Figure 6 but for *Sim-SFF* with star formation and feedback.

to solving the conservation equations for mass, momentum and energy, the equation for internal energy is also solved in parallel, and the conservative or non-conservative formulation is adaptively selected based on a local estimate of the energy truncation errors. This ensures that the gas temperature remains physical, even in highly supersonic regions.

The ENZO group uses a 64^3 initial root grid covering a $(1.311\ {\rm Mpc})^3$ simulation box, then achieves 80 pc resolution with maximum 8 levels of refinement. The mass thresholds above which a cell is refined by factors of two in each axis are $m_{\rm gas,IC} = 8.593 \times 10^4 {\rm M}_\odot$ and $8\,m_{\star,{\rm IC}} = 8 \times 3.437 \times 10^5 {\rm M}_\odot$ for gas and collisionless particles, respectively.[17] The non-thermal pressure floor Eq. (3) is used to modify the gas pressure inside the Riemann solver, but not to alter the actual gas energy field. For the supernova feedback scheme adopted by all mesh-based codes, we refer the readers to Section 3.2.

5.4. RAMSES

RAMSES is an octree-based adaptive mesh code featuring an unsplit 2nd-order accurate Monotone Upstream-centered Scheme for Conservation Laws (MUSCL) Godunov scheme for the gaseous component (Teyssier 2002).[19] For this comparison, RAMSES group uses a ideal gas equation of state with $\gamma = 5/3$, along with the HLLC Riemann solver (Toro et al. 1994) and the MinMod slope limiter (Roe 1986). The CFL safety factor for controlling the time step is set to 0.5. The dual energy formalism adopted in ENZO simulations (Section 5.3) is also used in RAMSES runs.

For this study, starting from a uniform 128^3 root grid covering $(320\ {\rm kpc})^3$, cells are refined up to 5 additional levels to achieve an ~ 80 pc cell size. The refinement process works as follows. First, new refinement is triggered on a cell-by-cell basis if the baryonic mass (gas + newly formed stars) exceeds $m_{\rm gas,IC} = 8.593 \times 10^4 {\rm M}_\odot$, or if the number of dark matter and/or star particles that are present in the IC exceeds 8.[20] We then mark additional cells by performing a mesh smoothing operation, expanding the initial area by one cell width in every direction. When new cells are created or old cells destroyed, density, momentum and internal energy are used as averaging and interpolating variables, thereby preventing a grid point with spurious temperature.

In RAMSES, the gas pressure field includes the non-thermal pressure support term given by a temperature polytrope $T_{\rm Jeans} = \mu T_{\rm J}(n'_{\rm H}/n_{\rm H,J})$ with mean molecular weight μ, $n'_{\rm H} = \rho_{\rm gas} X_{\rm H}/m_{\rm H}$, $T_{\rm J} = 1800$ K, $n_{\rm H,J} = 8\ {\rm cm}^{-3}$ and $X_{\rm H} = 0.76$.[21] As in ART-II, this polytrope approximately matches the common pressure support prescription Eq. (3). Newly created star particles in RAMSES have a fixed mass of $m_{\rm gas,IC}$, but they are spawned with a Poisson probability distribution whose parameters are designed to mimic the local Schmidt law, Eq. (4). Lastly, for the common supernova feedback scheme adopted by all mesh-based codes, we refer the readers to Section 3.2.

5.5. CHANGA

CHANGA is a reimplimentation of GASOLINE (see Section 5.6) in the CHARM++ runtime system.[22] CHARM++ (Kale & Krishnan 1996)[23] enables the overlap of computation and communication and provides adaptive load balancing infrastructure, allowing CHANGA to scale to hundreds of thousands

[19] The website is http://www.itp.uzh.ch/∼teyssier/Site/RAMSES.html.

[20] Readers should notice subtle differences here in refinement strategies between RAMSES and other mesh-based codes. Newly formed stars are considered as part of the baryonic fluid, so they do not change the particle refinement based solely on collisionless particles in the IC.

[21] This means that in order to retrieve gas internal energy or temperature (e.g., Section 6), the pressure support term needs to be subtracted out from RAMSES's pressure field, which is the only field being tracked.

[22] The website is http://www-hpcc.astro.washington.edu/tools/changa.html.

[23] The website is http://charm.cs.uiuc.edu/.

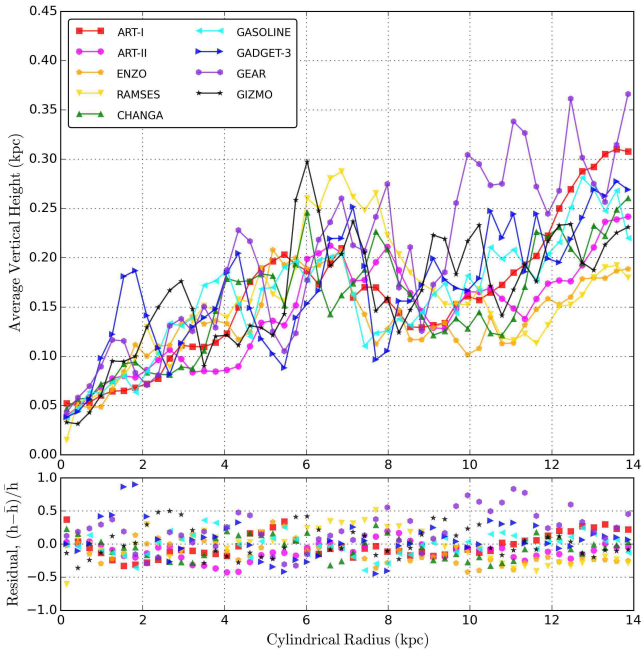

Figure 8. Cylindrically-binned, mass-weighted averages of gas vertical heights for *Sim-noSF* without star formation or feedback. The height is the absolute vertical distance from the $x-y$ disk plane centered on the galactic center, $|z_i - z_{center}|$. Shown in the bottom panel is the fractional deviation from the mean of these profiles. See Section 6.1 for an explanation on how this figure is made.

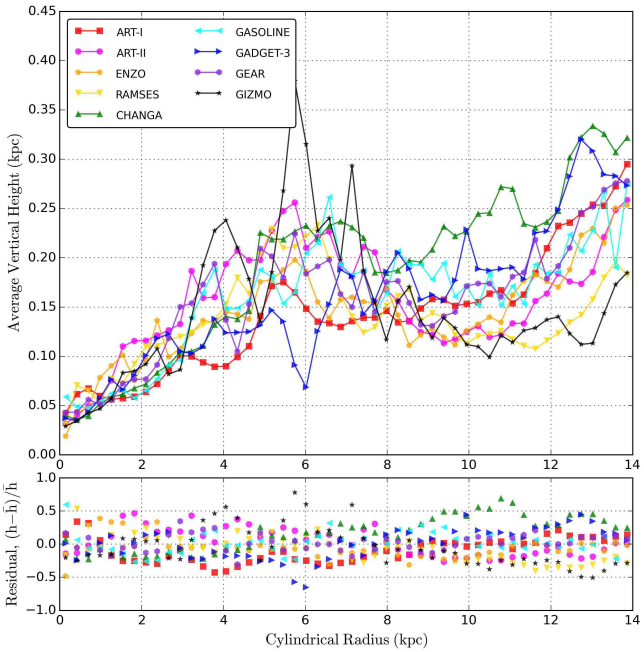

Figure 9. Same as Figure 8 but for *Sim-SFF* with star formation and feedback.

of processor cores (Menon et al. 2015). The hydrodynamics in CHANGA closely follows that of GASOLINE. SPH forces are calculated using the method of Ritchie & Thomas (2001), and energy is diffused using the scheme of Shen et al. (2010), both of which providing a more accurate treatment of multiphase ISM. Timesteps are determined by the minimum of an acceleration and a CFL criterion. Furthermore, the timesteps of neighbors are kept within a factor of 2 of each other as in Saitoh & Makino (2009) in order to accurately integrate highly supersonic flows.

For this work, a *k*-th nearest neighbor algorithm is used to find the $N_{\rm smooth} = 64$ nearest neighbors which are smoothed with the Wendland C4 kernel (Dehnen & Aly 2012) to determine hydrodynamic properties. Unlike conventional versions of CHANGA or GASOLINE, the supernovae thermal energy, mass, and metals are directly distributed to the 64 neighboring gas particles.[24] Gas particles that are neighbors of particles that will explode as a supernova in their next timestep are put on timesteps suitable for their post-supernova thermal energy, preventing them from being on a much smaller timestep required in the CFL condition. GRACKLE cooling is implemented but it does not self-consistently account for the *PdV* work or other external sources of energy, a requirement for CHANGA and GASOLINE's energy integration. Therefore, we split the energy integration into a half timestep of GRACKLE cooling, then a full timestep of *PdV* heating, and finally a second half timestep of cooling.

[24] The feedback prescription used in this experiment needed a reimplementation of the feedback routine normally used in CHANGA and GASOLINE (e.g., Stinson et al. 2006). In particular, in previous work, the supernovae rate determined by the stellar age and IMF is converted to an energy injection "rate" which is then incorporated into the thermal energy integration of neighboring gas particles. By contrast, for this study, a supernova event occurs instantaneously, making the rate an ill-defined quantity in their existing energy integration machinery.

5.6. GASOLINE

GASOLINE is a massively parallel SPH code, first described in Wadsley et al. (2004), that has subsequently been updated with modern SPH features. It contains a subgrid model for turbulent mixing of metals and energy (e.g., Shen et al. 2010), a timestep limiter by Saitoh & Makino (2009) (see Section 5.5), and a geometric density estimator for SPH force expressions (see Section 2.4 of Keller et al. 2014, for a latest detailed description of the code and its performance).

For the current work, the GASOLINE group uses a Wendland C4 smoothing kernel (Dehnen & Aly 2012) with $N_{\rm smooth} = 200$ neighbors.[25] The same feedback scheme as CHANGA's (Section 5.5) is implemented, smoothed with the Wendland C4 kernel over 64 neighbors (not $N_{\rm smooth} = 200$) to better match the amount of mass heated by feedback events with other particle-based codes. Gas particles that receive feedback compute their required CFL timestep at the timestep prior to receiving feedback, which helps to prevent numerical instability and overcooling. GRACKLE cooling is implemented by applying a half timestep of cooling, then a full timestep of external *PdV* heating, followed by a final half timestep of cooling, as in CHANGA (Section 5.5).

5.7. GADGET-3

GADGET-3 is an updated version of GADGET-2, a cosmological tree-particle-mesh (TPM) SPH code that was originally developed by V. Springel (Springel et al. 2001; Springel 2005).[26] GADGET-3 has important updates from GADGET-

[25] Note the difference in $N_{\rm smooth}$ from CHANGA in Section 5.5. Dehnen & Aly (2012) showed that this kernel can use larger neighbor numbers without the pairing instability which may effectively remove resolution, and that doing this improves performance on a number of basic hydrodynamics tests.

[26] The website is http://www.h-its.org/tap-software-en/gadget-code/ or http://www.mpa-garching.mpg.de/gadget/.

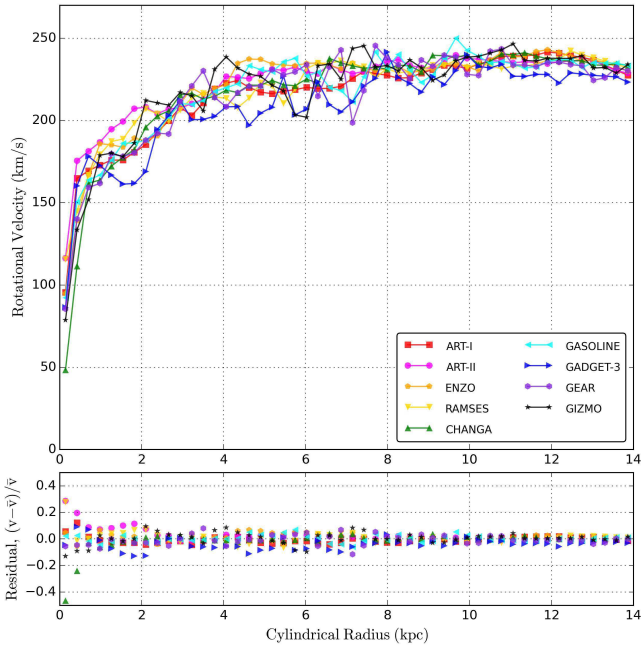

Figure 10. Gas rotation velocity curves at 500 Myr for *Sim-noSF* without star formation or feedback. The cylindrical radius and rotational velocity are with respect to the galactic center – location of maximum gas density within 1 kpc from the center of gas mass. Shown in the bottom panel is the fractional deviation from the mean of these profiles. See Section 6.2 for a detailed explanation on how this figure is made.

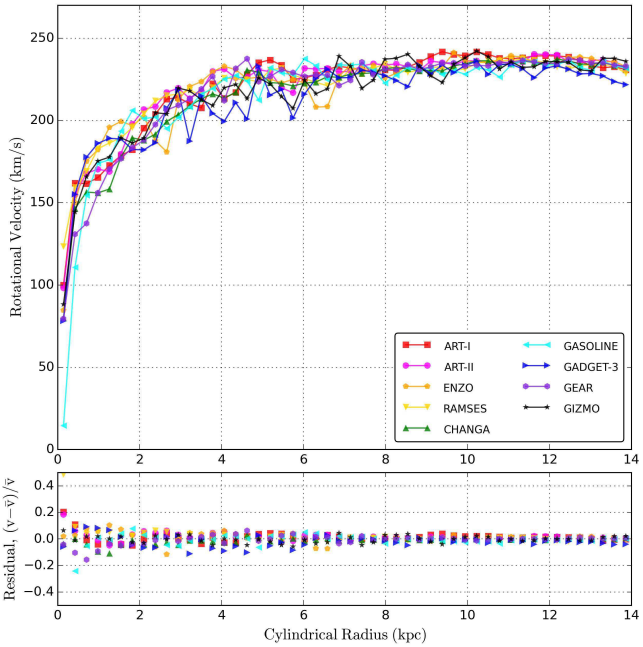

Figure 11. Same as Figure 10 but for *Sim-SFF* with star formation and feedback. Compare with Figure 24.

2, such as domain decomposition and dynamic tree reconstruction which may slightly alter the *N*-body dynamics. The GADGET-3 code used in this comparison is a modified version of the original GADGET-3 by K. Nagamine and his collaborators, which includes pressure-entropy formulation by Hopkins (2013), time-dependent artificial viscosity, variable smoothing lengths, among others (e.g., Choi & Nagamine 2012; Thompson et al. 2014; Aoyama et al. 2016).

For the present study, the GADGET-3 group adopts a quintic spline smoothing kernel (Morris 1996) with $N_{\rm ngb} = 64$. The implementation of supernova feedback is based on an updated version of Todoroki (2014) that largely follows a Sedov-Taylor blast wave method outlined in Stinson et al. (2006, 2013, but not their cooling shutoff model). The exact model used in the current work is fully described in Aoyama et al. (2016), but, in brief, the implementation comprises the following steps. Every time a star particle explodes, we compute the "shock radius" based on Chevalier (1974) and McKee & Ostriker (1977), and then find the gas particles within the radius. We then inject thermal energy and metal yields into the identified gas particles within the shock radius, weighted by the SPH spline kernel. Finally, we note that the results of this version of GADGET-3 are not representative of all the GADGET-3 codes in the community, because some of the results are strongly dependent on the detailed implementations of baryonic physics, such as star formation and feedback.

5.8. GEAR

GEAR is a self-consistent, fully parallelized, chemodynamical tree SPH code (Revaz & Jablonka 2012) which is built on the publicly available GADGET-2 code (see Section 5.7; Springel 2005). The simulations reported here are run with the improvements dicussed in Revaz et al. (2016), including the pressure-entropy formulation proposed by Hopkins (2013), individual and adaptive time-stepping schemes (Durier & Dalla Vecchia 2012), artificial viscosity (Monaghan & Gingold 1983) supplemented with the Balsara switch (f_{ij} from Balsara 1995), and particle-based time-dependent viscosity coefficient (Rosswog et al. 2000).

For this study, the standard cubic spline smoothing kernel (Monaghan & Lattanzio 1985) in GADGET-2 is used with $N_{\rm ngb} = 50$. The feedback energy, mass, and metal injection into the ISM is implemented following the standard SPH scheme. The implementation comprises the following steps. Every time a star particle explodes, we first find the nearest gas particles, according to the *weighted* number of neighbors as defined in Springel & Hernquist (2002). A *desired* number of neighbors $N_{\rm ngb} = 50$ is used. Then we inject thermal energy and yields into the neighboring gas particles, weighted by the SPH spline kernel. GRACKLE cooling is performed after the kick step, once gas particles have eventually received supernova feedback energy and once the size of the next timestep is known. The adiabatic cooling/heating is first applied, and then the radiative one provided by GRACKLE.

5.9. GIZMO

GIZMO (Hopkins 2015) is a new mesh-free Godunov code based on discrete tracers, aimed at capturing the advantages of both Lagrangian and Eulerian techniques.[27] The numerical scheme implemented in GIZMO, initially proposed by Lanson & Vila (2008), follows the implementation of Gaburov & Nitadori (2011) and relies on the discretization of the Euler equations of hydrodynamics among a set of discrete tracers. Unlike in the moving mesh technique, where the volume is partitioned by a Voronoi tessellation, GIZMO distributes the volume fraction assigned to the tracers through a kernel function. For the current work, GADGET's standard cubic spline smoothing kernel is used with $N_{\rm ngb} = 32$. Note, unlike SPH

[27] The website is http://www.tapir.caltech.edu/∼phopkins/Site/GIZMO.html.

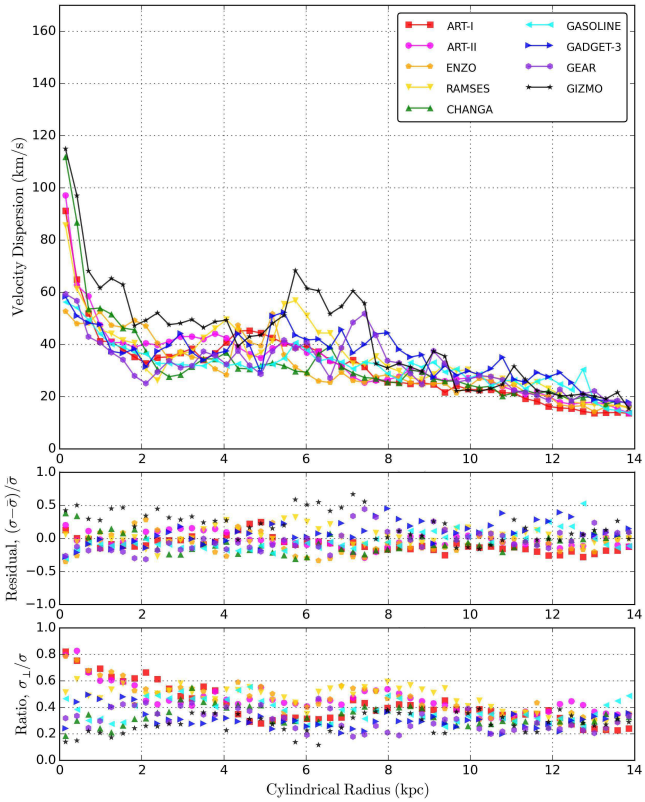

Figure 12. Gas velocity dispersion curves at 500 Myr for *Sim-noSF* without star formation or feedback. The velocity dispersion is the square root of mass-weighted averages of $(v_i - v_{rot}(r))^2$. Shown in the middle panel is the fractional deviation from the mean of these profiles. In the bottom panel we plot the ratio of *vertical* velocity dispersion (*z*-direction) to *total* velocity dispersion. See Section 6.2 for a detailed explanation on how this figure is made. The *y*-axis range of the top panel is kept identical among Figures 12-13 and 25 for easier comparison.

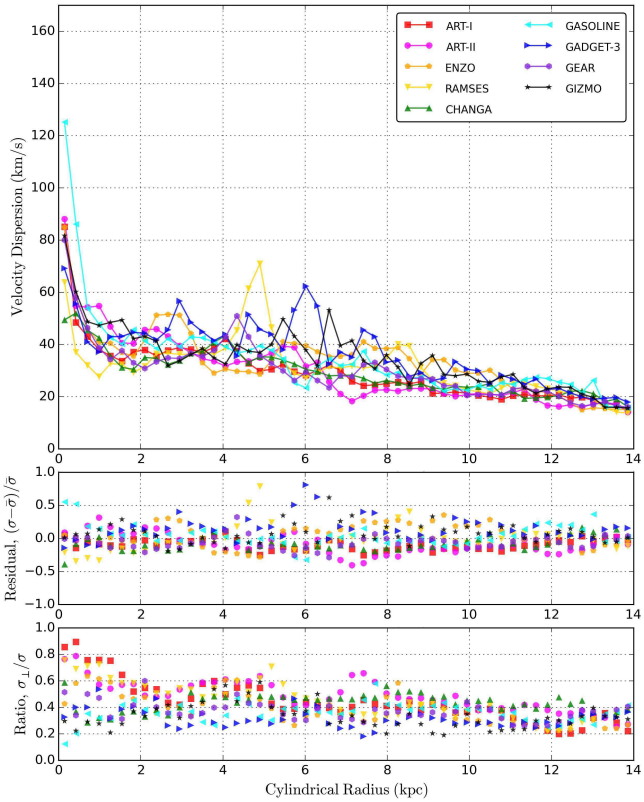

Figure 13. Same as Figure 12 but for *Sim-SFF* with star formation and feedback. Compare with Figure 25.

codes, these tracers only represent unstructured cells, sharing an "effective face" with the neighboring cells.[28] The Riemann problem is then solved across these faces using a Godunov method as in mesh-based codes, to accurately resolve shocks without artificial dissipation terms. Unlike mesh-based codes, these cells are not fixed in space and time, resulting in the scheme's Lagrangian behavior with intrinsically adaptive resolution. When the time evolution of the common face between two cells is considered, we use the 2nd-order accurate Meshless Finite Mass method (MFM; described in Hopkins 2015). GIZMO's time-stepping scheme is fully adaptive, and closely follows GADGET-3 or AREPO (Springel 2010). It also includes a timestep limiter by Saitoh & Makino (2009) (see Section 5.5).

GIZMO's gravity solver is based on the tree algorithm inherited from GADGET-3, itself descending from GADGET-2 (see Section 5.7; Springel 2005). Gravitational softenings in GIZMO can be fixed or fully adaptive, but in the reported runs fixed softening length is used matching SPH codes. To model supernova feedback, the GIZMO simulations shown here adopts a similar feedback strategy used in GEAR (Section 5.8). That is, energy, mass and metals are distributed among the neighboring gas particles/cells in a kernel-weighted fashion, but with $N_{ngb} = 32$. For star particles, timesteps are constrained to prevent supernovae from exploding in the timestep when the stars formed. Lastly, we caution that different feedback implementations using GIZMO in the literature adopt different algorithms to distribute supernova feedback energy (e.g., Hopkins et al. 2014, by the *FIRE* Collaboration).

[28] It is worth noting that the kernel size in GIZMO (what is called the "smoothing length" in SPH) does not play any role in the dynamics, but is simply related to each cell's "effective volume." For $N_{ngb} = 32$, a radius enclosing approximately this many neighbors is used to estimate the effective volume per particle at 2nd-order accuracy, with most of the "effective volume" coming from the region within a single inter-particle separation length around the tracer.

6. RESULTS

In this section, we lay out the results of the first isolated disk galaxy comparison by the *AGORA* Collaboration. We focus on similarities and discrepancies discovered by comparing 500 Myr snapshots of the participating simulations in two setups: *Sim-noSF* and *Sim-SFF*. As defined in Section 3, *Sim-noSF* refers to a run with radiative gas cooling but without star formation or feedback, and *Sim-SFF* refers to a run with radiative cooling, star formation and supernova feedback.

In our simulation analyses, a key role has been played by the *AGORA* recommended community-driven analysis platform `yt` (Turk et al. 2011; Turk & Smith 2011; Turk 2013).[4] It natively processes data from all 9 participating simulation codes discussed in this paper, plus many other modern astrophysics codes such as ATHENA (Stone et al. 2008), FLASH (Fryxell et al. 2000), GADGET-3-SPHS (Read & Hayfield 2012), NYX (Almgren et al. 2013), and ORION (Truelove et al. 1998), to name a few. Interested readers may try a unified, publicly available `yt` script employed in the present anal-

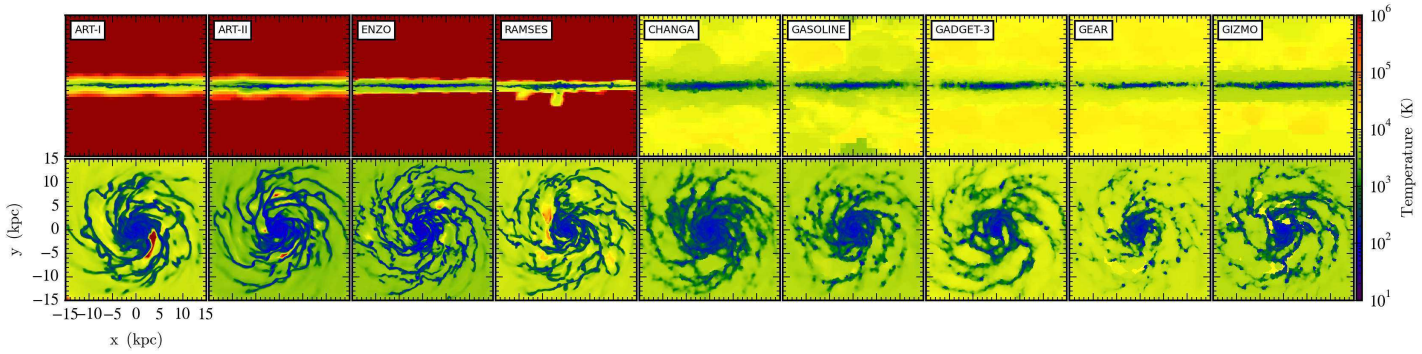

Figure 14. The 500 Myr composite of density-square-weighted gas temperature projections, edge-on *(top)* and face-on *(bottom)*, for *Sim-noSF* with radiative gas cooling but without star formation or supernova feedback. See Section 5 for descriptions of participating codes in this comparison, and Section 6.3 for a detailed explanation of this figure. Compare with Figure 2. The full color version of this figure is available in the electronic edition. The high-resolution versions of this figure and article are available at the Project website, http://www.AGORAsimulations.org/.

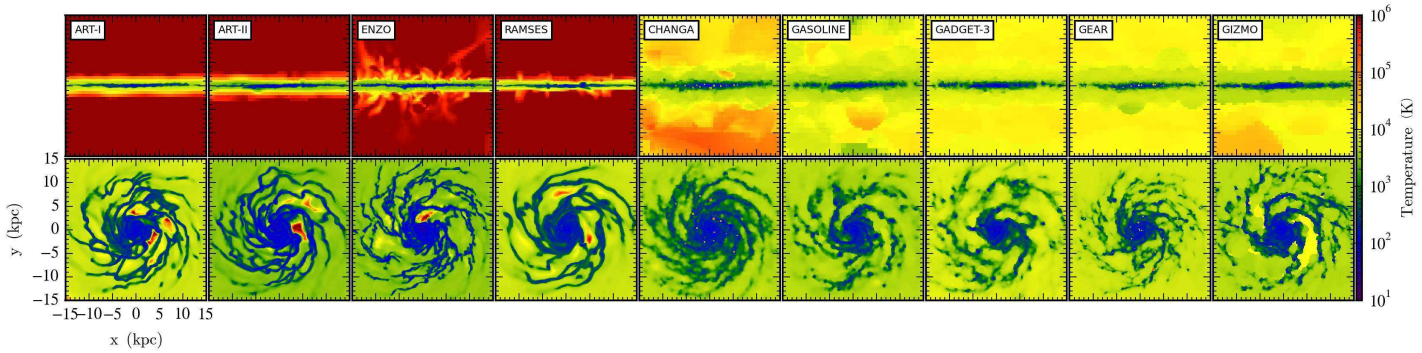

Figure 15. Same as Figure 14 but for *Sim-SFF* with star formation and feedback. See Section 6.3 for a detailed explanation of this figure. See also Section 3.2 for the common star formation prescription and the guideline for supernova feedback, and Section 5 for the exact deposit scheme of thermal feedback energy implemented in each code. Compare with Figure 3.

yses that has been developed throughout the progress of this study.[29,30] We also plan to make datasets used in the present study publicly available in the near future (see Section 7 for more information).

6.1. *Gas Disk Morphology*

We first examine the morphology of gas disks evolved in each of the codes in our experiments. In Figures 1 to 3, we compile 9 panels that exhibit the results of the isolated disk galaxy simulations, first with radiative gas cooling but without star formation or feedback, *Sim-noSF*, and second with star formation and feedback, *Sim-SFF*, by the 9 participating codes. Each panel displays the disk gas surface density in a 30 kpc box centered on the location of maximum gas density within 1 kpc from the center of gas mass. This centering criterion is adopted in all subsequent figures and plots. For visualizations of the particle-based codes hereafter (Figures 1-3, 14-15, 32, 34, 35) – but not in any other analyses except these figures – `yt` uses an in-memory octree on which gas particles are deposited using smoothing kernels. The resolution of this octree governs the resolution of produced images. If more than 8 particles are in an oct, that oct is refined into 64 child octs (i.e., `yt` parameters `n_ref` $= 8$ and `over_refine_factor` $= 2$), providing compatible or better image resolution than a typical SPH visualization. The densities are assigned to the octree in a scatter step. That is, we first calculate a particle's smoothing length, and then add the particle's density contribution to all cell centers of the octree cells that are within the particle's smoothing sphere.

We asked every code to output the state of the simulation immediately after it was initialized – so-called "0 Myr snapshot" – to allow ourselves to directly compare whether the initial condition (IC) generation was successful and consistent among codes. This exercise has been strenuously carried out for all the analyses items presented in Sections 6.1-6.3 and 6.7, enabling us to correct inconsistently initialized simulations early in the study. One such example, the surface density comparison of 0 Myr snapshots, is shown in Figure 1. A clear distinction in gas disk initialization between mesh-based and particle-based codes can be seen in this figure. To model the gas disk, SPH particles or GIZMO's discrete tracers are generated by drawing random numbers from the distribution function given by an analytic density profile. This by definition results in Poisson noise in the disk surface density shown here. Readers can also observe slight differences between mesh-based and particle-based codes in how the density field is represented in their calculations. By the nature of reconstructing the density from the positions of particles, the particle-based codes may smooth out the strong density contrast in the IC at the edge of the initial gas disk.

Interestingly, in Figures 2 and 3 at 500 Myr, there are other subtle differences noticeable between mesh-based and particle-based codes as well as within these sub-groups. While the peak densities and filamentary structures in the disks are very similar across all codes, it is noticeable that the mesh-based codes typically show lower densities in the inter-arm regions of the disks. The typical densities in those inter-

[29] The website is http://bitbucket.org/mornkr/agora-analysis-script/.

[30] `yt` version 3.3 or later is required for the script to reproduce our analyses. For the figures and plots in Section 6, the `yt-dev` changeset d7f213e1752e is used.

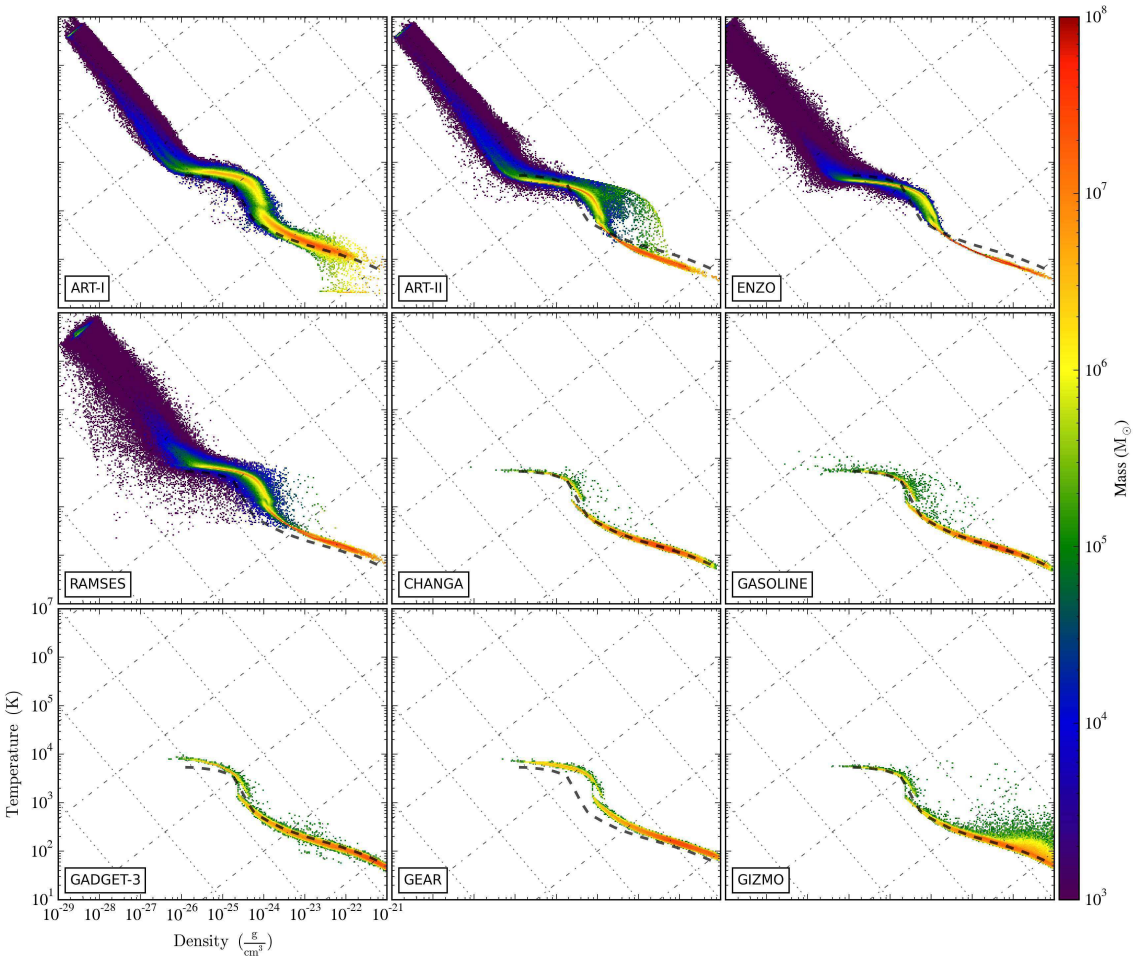

Figure 16. The 500 Myr composite of 2-dimensional probability distribution function of density and temperature for the gas within 15 kpc from the galactic center in *Sim-noSF* without star formation or feedback. Colors represent the total gas mass in each 2-dimensional bin. A gaseous halo – low density, high temperature gas in the upper left corner of each panel – exists only in mesh-based codes, but not in particle-based codes (SPH codes or GIZMO). To guide readers' eyes, we use a thick dashed line in each panel to plot the mean temperature in each density bin for CHANGA. The thin dotted diagonal lines denote the slope of constant pressure process, and the thin dot-dashed diagonal lines that of constant entropy process. Note that different versions of GRACKLE are interfaced with different codes (GRACKLE v2.1 in CHANGA, GASOLINE, GADGET-3 and GIZMO, versus GRACKLE v2.0 or below in ART-I, ART-II, ENZO, RAMSES and GEAR). See Section 6.3 for more information on this figure. The full color version of this figure is available in the electronic edition. The high-resolution versions of this figure and article are available at the Project website, http://www.AGORAsimulations.org/.

arm regions – while not containing much mass – may differ by as much as an order of magnitude between mesh-based and particle-based codes (see also Section 6.7 and Figure 35 for a related discussion on spatial resolution). Another distinguishing aspect among the participating codes is the number of dense clumps formed. This is true in both the simulations with and without star formation and feedback (see also Section 6.4 and Figures 21 and 23 for a related comparison of newly-formed stellar clumps).

Figures 4 and 5 are the cylindrically-binned gas surface density profiles for *Sim-noSF* and *Sim-SFF*, respectively. The cylindrical radius is defined as the distance from the galactic center. Raw particle fields are used for profiles of the particle-based codes, not the interpolated or smoothed fields constructed in `yt`. In other words, the total mass in each cylindrical annulus is divided by its area so that each gas particle contributes only to a bin in which its center falls. As noted earlier, the gas surface densities show a high degree of correspondence across all codes in both *Sim-noSF* and *Sim-SFF*. All 9 profiles agree very well within a factor of a few at all radii (averaged fractional deviation for $2 < r < 10$ kpc in *Sim-SFF* is 32.2% or 0.121 dex),[31] and can be approximated by exponential profiles at radii > 1.5 kpc. Note that due to the gas consumed by star formation, the gas surface density slightly decreases from Figure 4 to Figure 5, the latter of which could be compared with Figures 22 (newly-formed stellar surface density) and 27 (star formation rate surface density). Aside from this small decrease in density, there is little impact of supernova feedback from contrasting *Sim-noSF* and *Sim-SFF*. The inefficiency of stellar feedback is partly related to a only small increase in stellar mass in the first 500 Myr of evolution (see Sections 2 and 6.6 for more discussion).

Displayed in Figures 6 and 7 are the vertically-binned gas surface density profiles for *Sim-noSF* and *Sim-SFF*, respectively. The height is defined as the absolute vertical distance from the $x-y$ disk plane centered on the galactic center, i.e., $|z_i - z_{\rm center}|$. Again, for particle-based codes raw particle fields are used to produce the profiles, not the interpolated or smoothed fields in `yt`. Note a smaller range in x-axes than in the previous figure (only one tenth of Figures 4 and 5). The

[31] Defined as $N_i^{-1} \sum_{2<r_i<10} \left[N_{\rm code}^{-1} \sum_{\rm code} \left\{ (f_{i,\rm code}/\overline{f_i}) - 1 \right\}^2 \right]^{1/2}$.

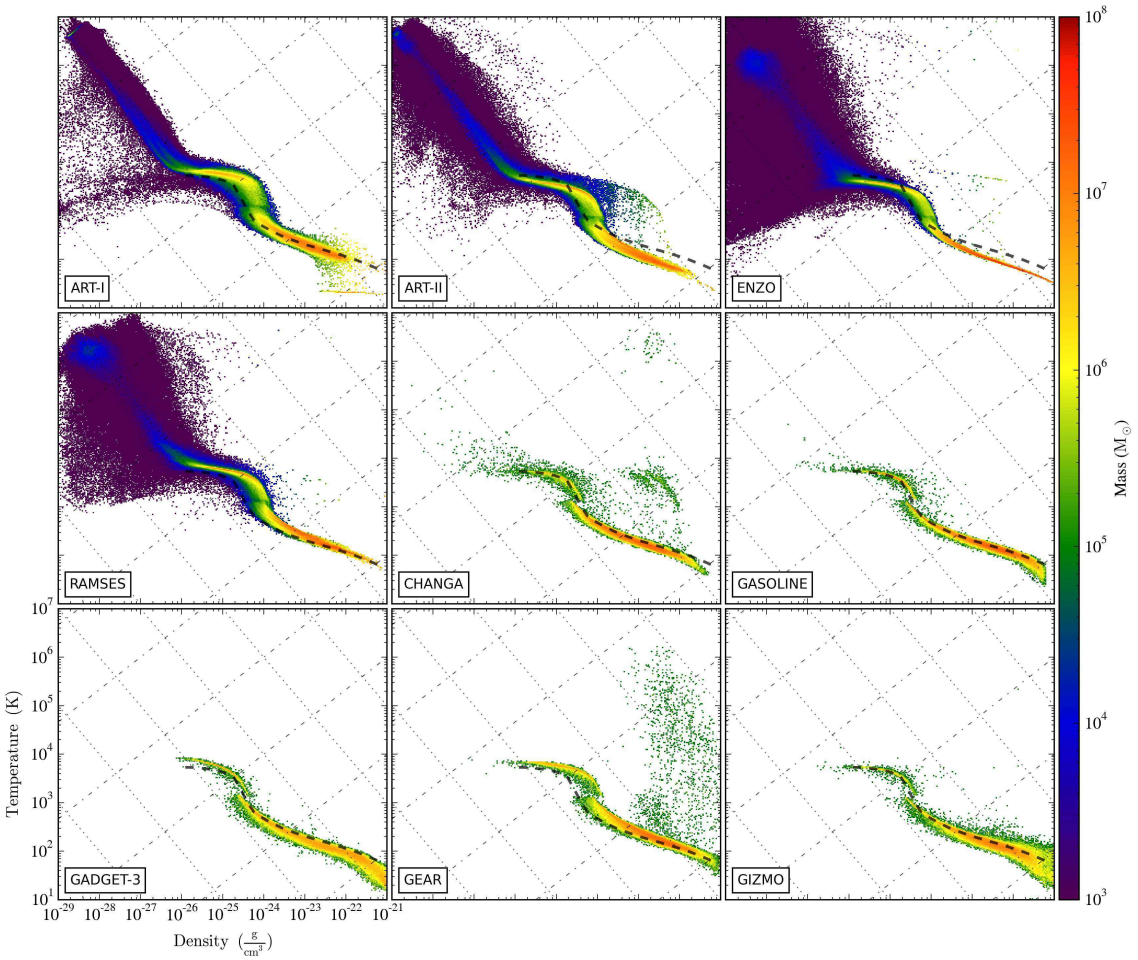

Figure 17. Same as Figure 16 but for *Sim-SFF* with star formation and feedback. We use a thick dashed line in each panel to plot the mean temperature in each density bin in CHANGA's *Sim-noSF* run (same as in Figure 16). See Section 6.3 and the caption of Figure 16 for a detailed explanation of this figure. Compare with Figure 33.

surface densities begin to diverge above ∼0.6 kpc from the disk plane, but below that substantial agreement appears (averaged fractional deviation for $z < 0.6$ kpc in *Sim-SFF* is 30.4 % or 0.115 dex). There is no systematic difference between mesh-based and particle-based codes, similar to what we find in radial density profiles, Figures 4 and 5.

Figures 8 and 9 show the cylindrically-binned, mass-weighted average of gas vertical heights for *Sim-noSF* and *Sim-SFF*, respectively. As defined before, the height is an absolute vertical distance from the disk plane. Thus, each line in Figures 8 and 9 represents the averaged $|z_{\rm i} - z_{\rm center}|$ as a function of cylindrical radius. Combined with Figures 6 and 7, these plots provide insight into the thicknesses of the gas disks in each of the codes. Again, no systematic difference is found between mesh-based and particle-based codes.

6.2. *Gas Disk Kinematics*

Here we examine the kinematics of gas disks that are evolved using each of the participating codes. First, in Figures 10 and 11 we show the gas rotation velocity curves for *Sim-noSF* and *Sim-SFF*, respectively. The curve reveals a mass-weighted average of gas rotational velocity of gas cells/particles, as a function of cylindrical radius. The cylindrical radius and rotational velocity are defined with respect to the galactic center (see Section 6.1 for our adopted centering scheme). For mesh-based codes, only the dense enough gas cells ($\rho_{\rm gas} > 10^{-25}\,{\rm g\,cm^{-3}}$) are considered in order to minimize the contribution of the gaseous halo which is present only in mesh-based codes (see Section 2 for more information about the halo gas distribution in our IC). From these two figures it is clear that there exists a very good agreement on gas kinematics in all the codes, as good as within a few percent at ∼ 10 kpc for both *Sim-noSF* and *Sim-SFF* (averaged fractional deviation for $2 < r < 10$ kpc in *Sim-SFF* is 2.8% or 0.012 dex). Discrepancies in the central region (< 1.5 kpc) are a result of determining the galactic center that may constantly shift its position.

Figures 12 and 13 reveal the gas velocity dispersion curves for *Sim-noSF* and *Sim-SFF*, respectively. The velocity dispersion quantifies the residual velocity components of gas other than the rotational velocity found in Figures 10 and 11. In other words, each line in Figures 12 and 13 denotes a square root of a mass-weighted average of $(v_{\rm i} - v_{\rm rot}(r))^2$, as a function of cylindrical radius. As in Figures 10 and 11, for mesh-based codes only dense enough cells are used to compute the dispersion. Again a good agreement is found in the velocity dispersion between all codes within a few tens of percent for both *Sim-noSF* and *Sim-SFF* (averaged fractional deviation for $2 < r < 10$ kpc in *Sim-SFF* is 17.8% or 0.071 dex). Larger variations in the central region (< 1.5 kpc) are partly due to the center determination, just like in rotation velocity curves, Figures 10 and 11. The discrepancies are also pro-

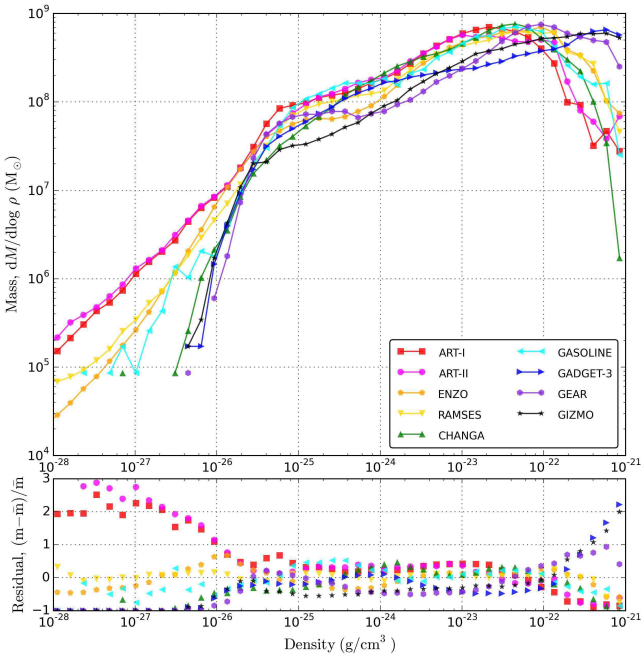

Figure 18. Gas density probability distribution function at 500 Myr for *Sim-noSF* without star formation or feedback. Note that a gaseous halo – low density tails towards the left side of this plot – exists only in mesh-based codes, but not in particle-based codes (SPH codes or GIZMO). Shown in the bottom panel is the fractional deviation from the mean of these profiles. See Section 6.3 for more information on this figure.

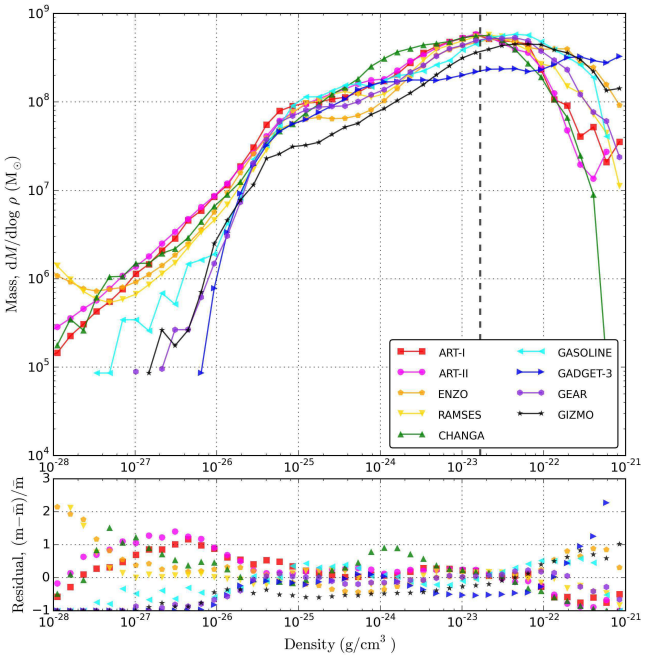

Figure 19. Same as Figure 18 but for *Sim-SFF* with star formation and feedback. The thick dashed line denotes the star formation threshold density, $n_{\rm H,thres} = 10\ {\rm cm}^{-3}$.

duced by vertical movement of gas in the inner disk, which is captured in mesh-based codes (ART-I, ART-II, ENZO, and RAMSES) but not as well in particle-based codes (CHANGA, GASOLINE, GADGET-3, GEAR, and GIZMO) given our very particular choice of 80 pc resolution. This can be clearly seen in the bottom panels of Figures 12 and 13, in which we plot the ratio of *vertical* velocity dispersion (z-direction) to *total* velocity dispersion to illustrate the contribution by vertical movement of gas. From these figures, we find that a significant portion of gas velocity dispersion in the inner disk measured in mesh-based codes are driven by vertical movement, but not in particle-based codes (see Section 6.7 and Figure 35 for a related discussion on spatial resolution).

6.3. *Thermal Structure of the Interstellar Medium*

In Figures 14 and 15, we compile 9 panels of the density-square-weighted gas temperature projections for *Sim-noSF* and *Sim-SFF*, respectively.[32] Each panel is for the central 30 kpc box (see Section 6.1 for our adopted centering scheme). Readers should note that, by design only mesh-based codes (ART-I, ART-II, ENZO, and RAMSES) initially include a gaseous halo with low density and high temperature ($T_{\rm gas} = 10^6$ K; colored dark red), but not the particle-based codes (CHANGA, GASOLINE, GADGET-3, GEAR, and GIZMO; see Section 2). All of these codes show similar features without star formation and feedback in Figure 14. We continue a related discussion on *Sim-noSF* using Figures 16 and 18 later in this section.

But slight differences in supernova feedback implementation – even within a common guideline – may cause different features in the temperature map, Figure 15. For example, small hot bubbles are visible in CHANGA and GEAR, but not in GASOLINE or GADGET-3 in which the common supernova

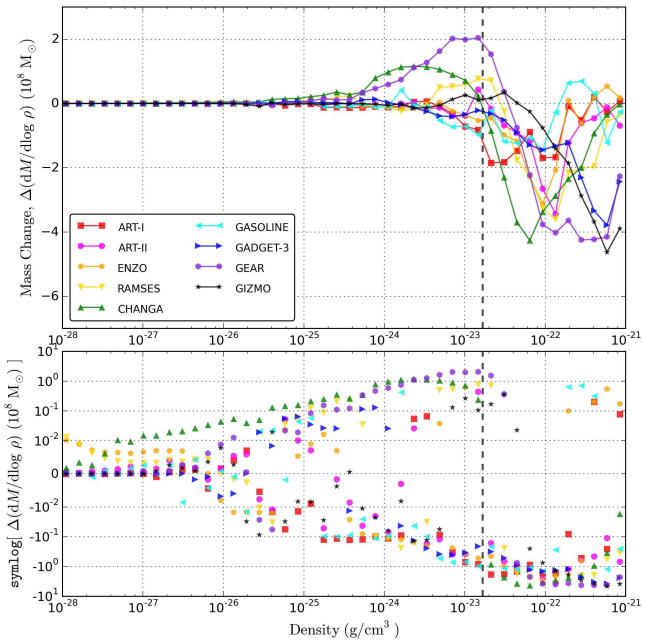

Figure 20. Code-by-code mass change in each density bin from Figure 18 to 19. This plot measures the effect of star formation and feedback by subtracting the density probability distribution function of *Sim-noSF* from that of *Sim-SFF*. The y-axis spans from -7×10^8 M$_\odot$ to $+3\times10^8$ M$_\odot$ in a linear scale. Shown in the bottom panel is the sign-preserving logarithm of the mass difference in order to make smaller changes to stand out.[ii] The thick dashed line denotes the star formation threshold density, $n_{\rm H,thres} = 10\ {\rm cm}^{-3}$. See Section 6.3 for more information on this figure.

[32] Temperature information may not be readily available in some codes in which only internal energy (T/μ; GADGET-3 and GEAR) or pressure ($\rho T/\mu$; RAMSES) is tracked instead. In this case, we use the $T-\mu$ table derived from GRACKLE v2.0 to acquire the temperature.

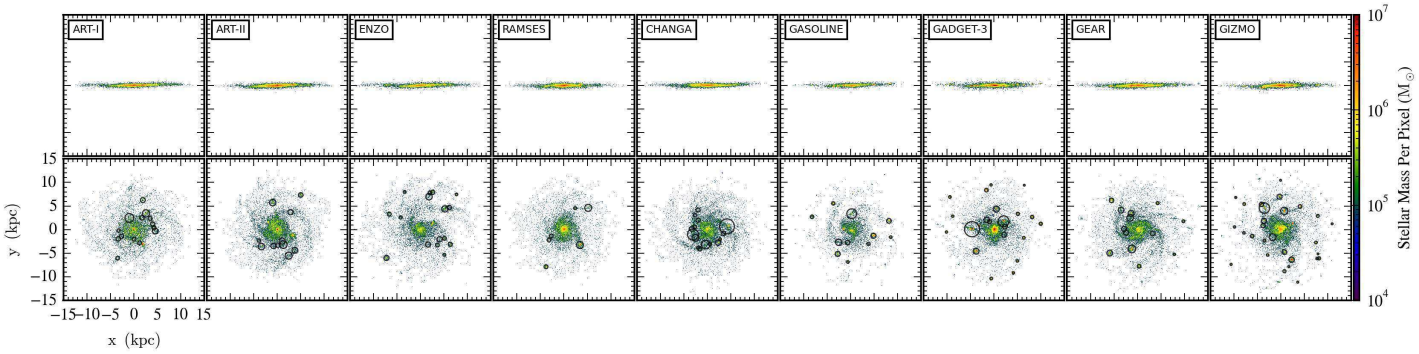

Figure 21. The 500 Myr composite of newly-formed star particle distributions, edge-on *(top)* and face-on *(bottom)*, for *Sim-SFF* with star formation and feedback. Only the newly-formed star particles are drawn, not the disk or bulge stars that were present in the IC. Colors represent the total newly-formed stellar mass in each 2-dimensional bin. We also highlight the clumps of newly-formed star particles identified by the Friends-of-Friends (FOF) algorithm with circles. Clumps with masses below $2.6 \times 10^6\,M_\odot$ and the most massive clump found by FOF (always associated with the stellar bulge at the galactic center) are excluded. See Section 6.4 for a detailed explanation of this figure. Compare with Figures 3, 15, and 35. The full color version of this figure is available in the electronic edition. The high-resolution versions of this figure and article are available at the Project website, http://www.AGORAsimulations.org/.

feedback schemes are implemented slightly differently.[33] We also note the chimneys of hot gas departing from the disk in ENZO and RAMSES, clearly visible in the edge-on maps of Figure 15. This gas ejecta from the disk as a result of supernova feedback is not as evident in ART-I, ART-II, or particle-based codes, although some hot bubbles are seen in CHANGA and GEAR. We caution that the spatial resolution employed in this paper is only 80 pc. It is not as high as the resolutions in some of the modern zoom-in cosmological simulations, and may not be enough for particle-based codes to resolve the chimney-like structure above and below the disk (see Section 6.7 and Figure 35 for a related discussion on spatial resolution). We refer the readers to a continued discussion on *Sim-SFF* using Figures 17 and 19 later in this section.[34]

To better understand what we find in Figures 14 and 15, we show the 2-dimensional probability distribution functions (PDFs) of gas density and temperature in Figures 16 and 17 for *Sim-noSF* and *Sim-SFF*, respectively. We consider gas within 15 kpc from the galactic center. Colors represent the total gas mass in each 2-dimensional bin. As explained in Section 6.1, raw particle fields are used for the PDFs of particle-based codes, not the smoothed fields constructed by `yt`. Note that a gaseous halo, represented by low density, high temperature gas in the upper left corner of each panel, is designed to exist only in mesh-based codes, but is absent in particle-based codes (SPH codes and GIZMO; see Section 2). It is therefore *not* the intended scope of this paper to compare this hot halo or circumgalactic medium between codes. To guide readers' eyes, we plot the mean temperature in each density bin from CHANGA's *Sim-noSF* run with a thick dashed line in each panel (in both Figures 16 and 17; in the range of $[10^{-26}, 10^{-21}]\,\mathrm{g\,cm^{-3}}$). CHANGA's mean profile is close to the "mean of means" of these 9 codes, thus helps to compare the PDFs' relative positions between codes. The thin dotted diagonal lines denote the slope of constant pressure process,

[33] Both CHANGA and GASOLINE spread supernova feedback energy to 64 neighboring particles, but they use different neighbor numbers for smoothing in hydrodynamics: $N_{\mathrm{smooth}} = 64$ in CHANGA versus 200 in GASOLINE (see Sections 5.5 and 5.6). In addition, unlike GEAR, GADGET-3 adopts the Sedov-Taylor blast wave method (see Section 5.7). A slight change in the details of shock radius estimation is shown to cause a large difference.

[34] In Figures 15 and 32 readers may notice hemispherical shapes of size ~ 5 kpc in some particle-based codes, e.g., GASOLINE or GEAR. These are not expanding blast waves driven by supernova feedback, but a visualization effect due to `yt`'s smoothing kernel for SPH particles or GIZMO's discrete tracers. This feature is also clearly seen in Figure 32, and barely seen in Figure 3. Higher-resolution simulations would minimize this artifact.

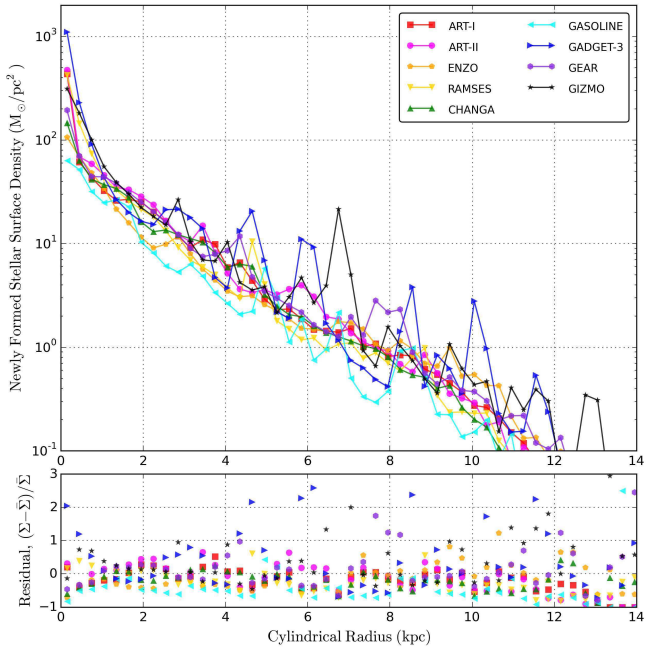

Figure 22. Cylindrically-binned newly-formed stellar surface density profiles at 500 Myr for *Sim-SFF* with star formation and feedback. Only the newly-formed star particles are considered, not the disk or bulge stars that were present in the IC. Shown in the bottom panel is the fractional deviation from the mean of these profiles. See Section 6.4 for a detailed explanation of this figure. Compare with Figures 5 and 27. The *y*-axis range of the top panel is kept identical among Figures 4-7 and 22 for easier comparison.

and the thin dot-dashed diagonal lines that of constant entropy process.

Overall, all codes exhibit similar behaviors in Figure 16 when without star formation and feedback, just like the broad similarity observed in Figure 14. A clear branch of gas is visible in all codes extending towards higher density, lower temperature, owing to the common treatment of cooling by the GRACKLE library (see also Figure 18, and a related discussion later in this section). But, as noted in Section 3.1, due to varying GRACKLE versions participating codes are interfaced with (GRACKLE v2.1 in CHANGA, GASOLINE, GADGET-3 and GIZMO, versus GRACKLE v2.0 or below in ART-I, ART-II, ENZO, RAMSES and GEAR), the cooling rates differ slightly from code to code even at the same density, temperature, and metal fraction. See Section 2 (footnote 9) or Section 3.1 (footnote 13) for more information. Any remaining discrepancy is

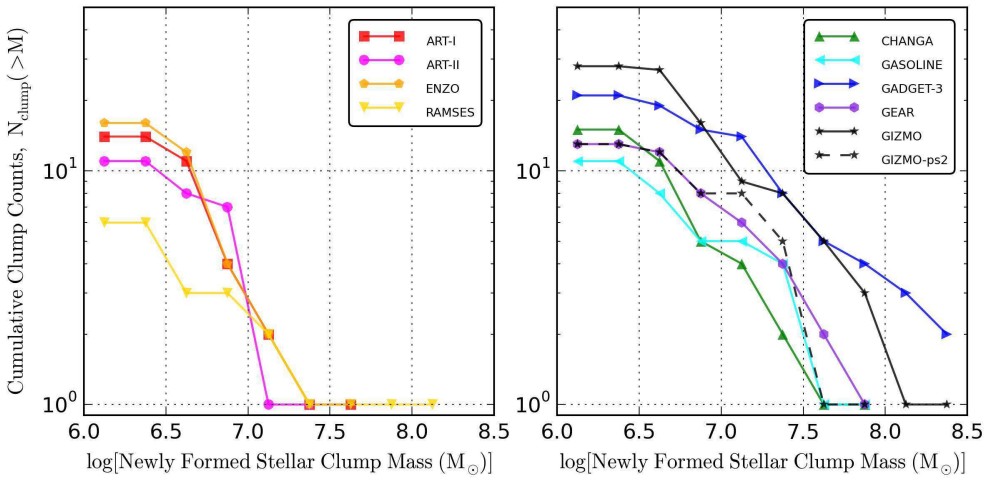

Figure 23. The cumulative mass function of newly-formed stellar clumps at 500 Myr for *Sim-SFF*. Clumps with masses below $2.6 \times 10^6\,M_\odot$ and the most massive clump found by FOF (always associated with the stellar bulge at the galactic center) are excluded. See Section 6.4 for more information on this figure, including an additional test GIZMO-PS2 for which GIZMO's Jeans pressure support is increased by a factor of 2.

attributable to the difference in how each code sub-cycles its cooling module in the hydrodynamics calculation.

Now we compare the density−temperature PDFs when star formation and supernova feedback are included in Figure 17. All codes successfully lower the fraction of low temperature, high density gas by forming stars and then injecting their feedback energy (see also Figures 18-20, and a related quantitative discussion later in this section). However, notable differences exist between codes as to how gas reacts to the supernova feedback. For example, in CHANGA and GEAR some gas is leaving the aforementioned high density branch towards higher temperature (up to 10^6 K) due to supernova feedback, but not in GASOLINE, GADGET-3 and GIZMO. The high temperature, high density gas seen in CHANGA and GEAR is associated with the small hot bubbles discussed in Figure 15. As explained earlier, these discrepancies in particle-based codes are attributed to different numerical implementations of the common feedback physics. Also noticeable is the hot gas being ejected from the disk as a result of feedback particularly in ENZO and RAMSES, seen as a broader distribution of hot gas in Figure 17 compared to Figure 16.

For a more quantitative comparison of the ISM thermal structure, shown in Figures 18 and 19 are the gas mass distributions along the density axis (density probability distribution function) for *Sim-noSF* and *Sim-SFF*, respectively, simply derived from Figures 16 and 17 above. Note again that a gaseous halo, represented by low density tails towards the left side of these plots, is by design included only in the mesh-based codes, but not in the particle-based codes. In Figure 18, when without star formation and feedback all codes show similar distributions within a factor of a few difference in the density range $[10^{-25}, 10^{-22}]\,\mathrm{g\,cm^{-3}}$. A notable deviation is that three particle-based codes, GADGET-3, GEAR and GIZMO, hold more mass at density above $10^{-22}\,\mathrm{g\,cm^{-3}}$ than the rest of the codes do. However, in Figure 19, now in a more realistic setup including star formation and feedback, no clear systematic difference exists between mesh-based and particle-based codes. While the agreement in the range $[10^{-25}, 10^{-22}]\,\mathrm{g\,cm^{-3}}$ is again within a factor of a few (averaged fractional deviation in $10^{-25} < \rho < 10^{-22}\,\mathrm{g\,cm^{-3}}$ is 28.6% or 0.109 dex), the codes diverge from one another up to more than an order of magnitude at density above $10^{-22}\,\mathrm{g\,cm^{-3}}$, translating into differences in clumping properties (Figure 23) and star formation rates (Figure 26).

Then, Figure 20 plots the code-by-code mass change in each density bin from Figure 18 to Figure 19. This plot aims to measure the effect of star formation and supernova feedback by subtracting the density probability distribution of *Sim-noSF* from that of *Sim-SFF*. As noted in our discussion of Figure 17, *Sim-SFF* lowers the fraction of high density gas by forming stars at above $n_{\mathrm{H,\,thres}} = 10\ \mathrm{cm^{-3}} = \rho_{\mathrm{gas,\,thres}}/m_{\mathrm{H}}$ (see Section 3.2) and then injecting their feedback energy. This impact is more evident in the bottom panel of Figure 20, where we show the sign-preserving logarithm, `symlog()`, of the mass change, making smaller changes more discernible.[35] Without exception, gas masses at density above $\rho_{\mathrm{gas,\,thres}} = 1.67 \times 10^{-23}\,\mathrm{g\,cm^{-3}}$ (thick dashed line) are reduced by star formation and feedback, even if inefficiently. The gas is either consumed by star formation or redistributed by thermal supernova feedback to a less dense region below $\rho_{\mathrm{gas,\,thres}}$ away from star-forming sites.

6.4. *Stellar Disk Morphology*

In this section, we study the morphology of stellar disks formed in *Sim-SFF*. Figure 21 shows the distributions of newly-formed star particles in *Sim-SFF* for each of the 9 codes. Only the newly-formed star particles are drawn, not the disk or bulge stars that were present in the IC. Star particles present in the IC are excluded from all "stellar" particle analyses hereafter. Each frame is centered on the galactic center, defined in Section 6.1 as the location of peak gas density within 1 kpc from the center of gas mass. This center almost always coincides approximately with the center of the most massive stellar clump. Colors represent the total newly-formed stellar masses in each 2-dimensional bin. The bottom rows also highlights clumps of newly-formed stars (more discussion on stellar clumps later in this section).

Figure 22 depicts surface densities of newly-formed star particles – excluding star particles present in the ICs – for *Sim-SFF*, calculated in cylindrically symmetric radial bins. While there are differences up to a factor of a few among some codes (e.g., between GADGET-3 and GASOLINE, due to their different rates of galaxy-wide star formation shown in Figure 26; averaged fractional deviation for $2 < r < 10$ kpc is 53.9% or

[35] $\mathrm{symlog}(x) = \mathrm{sgn}(x) \times \log|x|$.

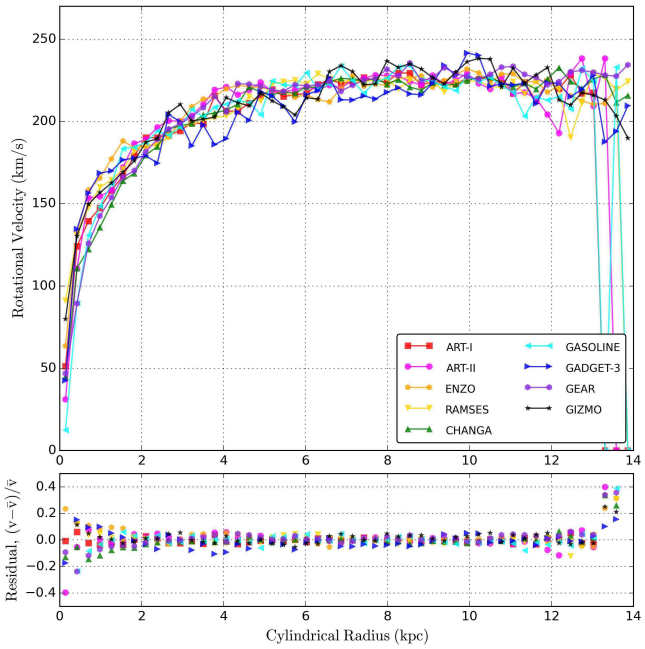

Figure 24. Stellar rotation velocity curves at 500 Myr for *Sim-SFF* with star formation and feedback. The cylindrical radius and rotational velocity are with respect to the galactic center – location of maximum gas density within 1 kpc from the center of gas mass. Only the newly-formed star particles are considered. See Section 6.5 for a detailed explanation on how this figure is made. Compare with Figure 11. The *y*-axis range of the top panel is kept identical among Figures 10-11 and 24 for easier comparison.

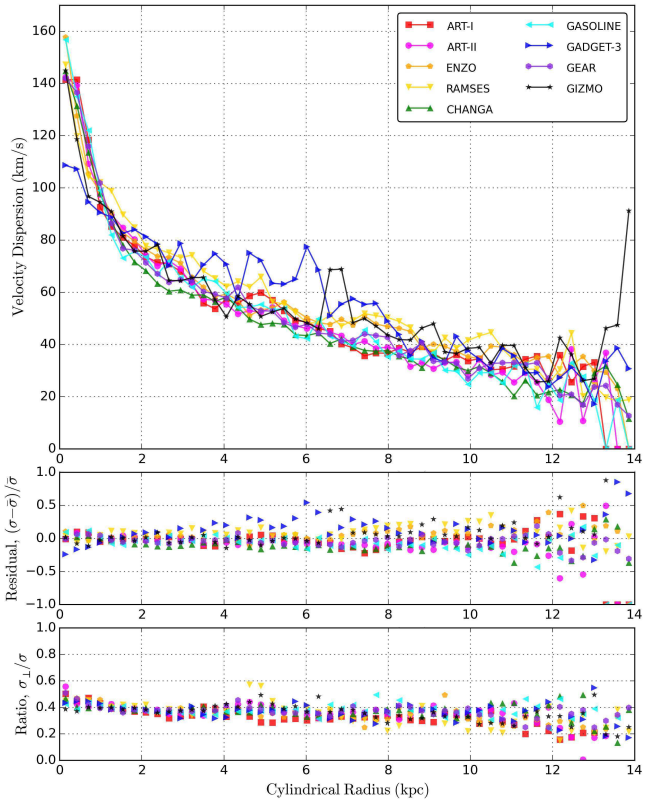

Figure 25. Stellar velocity dispersion curves at 500 Myr for *Sim-SFF*. Shown in the middle panel is the fractional deviation from the mean of these profiles. In the bottom panel we plot the ratio of *vertical* velocity dispersion (*z*-direction) to *total* velocity dispersion. Compare with Figure 13. The *y*-axis range of the top panel is kept identical among Figures 12-13 and 25 for easier comparison.

0.187 dex), all the lines can be well fit by an exponential disk profile at radii > 1.5 kpc. Occasional fluctuations visible in the profiles (e.g., at ~ 6.5 kpc in GIZMO) are due to dense stellar clumps located at these particular radii.

In order to compare the distribution of newly-formed star particles and the level of disk fragmentation between different codes, we identify clumps in the distribution of newly-formed star particles using the Friends-of-Friends (FOF) algorithm (Efstathiou et al. 1985).[36] We only consider clumps with newly-formed stellar masses above $2.6 \times 10^6\,M_\odot$ (equivalent to 30 times the mass of star particles present in the IC, $30\,m_{\star,\mathrm{IC}}$). The most massive clump found by FOF is excluded since it is always associated with the stellar bulge at the galactic center, but we do not explicitly remove gravitationally unbound clumps (which may have overestimated the number of clumps). Identified clumps are marked with circles in the bottom row of Figure 21, where the radius of each circle indicates the clump's virial radius. We also show the cumulative mass functions of newly-formed stellar clumps in Figure 23. It is worth to note that very similar clump distributions and mass functions are discovered when we identify clumps using the HOP algorithm (Eisenstein & Hut 1998) in lieu of FOF. The general trends for clumps discussed below are largely independent of the clump finder.

In all codes, the majority of newly-formed stellar clumps have masses below $10^{7.5}\,M_\odot$, and there is a relatively sharp decline in the number of clumps towards higher masses.[37] From these figures, it is also clear that nearly-formed stellar clumps are the most prevalent in the GIZMO run, and less so in the RAMSES run than in other codes. The relatively large number of stellar clumps in GIZMO is not a transient feature, but consistently observed across snapshots until 1 Gyr. This is related to the fact that GIZMO produces the most clumpy gas disk among the codes even with star formation and feedback in *Sim-SFF* (see Figure 3). While preserving all the common elements in comparison (such as pressure floor or resolution; described in Sections 3 and 4), the GIZMO group has carried out extensive tests with other simulation parameters to check what most dictates the level of fragmentation (e.g., smoothing kernel, N_{ngb}, GRACKLE version, slope limiter, dual energy formalism; but always within conventional norms), finding that the different parameter choices do not qualitatively alter the mass function. However, we note that increasing the Jeans pressure floor by as little as 100% – well within the uncertainty in the geometry prefactor of Jeans pressure support equation, Eq. (3) – and reverting to GRACKLE v2.0 – where metal cooling rates are slightly lower compared to v2.1

[36] For the FOF machinery, it is implicitly assumed that all newly-formed star particles have the same mass. The actual mass difference is no more than a factor of 2. The FOF finder also implicitly demands that particles are located in a periodic box, whose size we manually set to 3.25^3 Mpc (largest box size used by one of the groups). Given the periodic boundary condition, we use a linking length equal to 0.25% of the average inter-particle distance.

[37] These clumps are not to be confused with the giant clumps of masses between 10^8 and $10^9\,M_\odot$ observed in star-forming galaxies at redshift $z \sim 2$. These $z \sim 2$ clumps are expected to form in disks with gas fractions of 40-50%, much higher than the initial 20% gas fraction in our experiment.

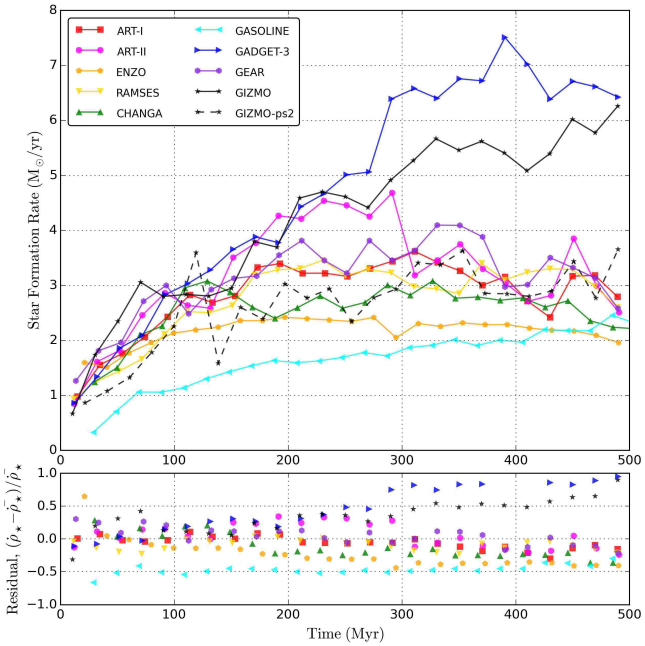

Figure 26. Galaxy-wide star formation rates by 500 Myr for *Sim-SFF*. Shown in the bottom panel is the fractional deviation from the mean. See Section 6.6 for a detailed explanation on this figure, including an additional test GIZMO-PS2 for which GIZMO's Jeans pressure support is increased by a factor of 2.

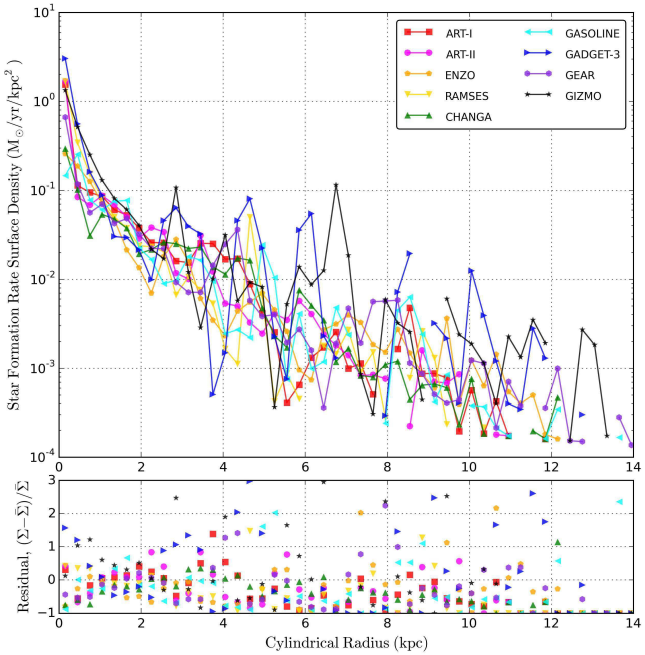

Figure 27. Cylindrically-binned star formation rate surface density profiles at 500 Myr for *Sim-SFF*. Star formation rates are estimated using the newly-formed star particles that are younger than 20 Myr old. Shown in the bottom panel is the fractional deviation from the mean of these profiles. See Section 6.6 for a detailed explanation on how this figure is made. Compare with Figures 5 and 22.

because of the different solar metallicity definition – entirely removes the discrepancy (see the black dashed line in Figure 23 labeled GIZMO-PS2). A possible explanation for the stronger fragmentation in GIZMO is that the "effective" gravitational resolution in GIZMO is slightly higher compared to other codes (as observed by Few et al. 2016), resulting from a combination of different choices in their implementation (e.g., slope limiter, gradient estimator, density estimator).[38]

6.5. *Stellar Disk Kinematics*

Following the analysis shown in Section 6.2 for the gas disk kinematics, here we study the kinematics of stellar disks formed by each code in *Sim-SFF*. In Figure 24 we show the rotation velocity curves for newly-formed star particles in *Sim-SFF*. As in gas rotation velocity curves of Figures 11, each line represents a mass-weighted average of stellar rotational velocities as a function of cylindrical radius. As in Section 6.4, only the newly-formed star particles are considered for these profiles, not the disk or bulge stars in the IC. Just as in the gas rotation curves, the stellar rotation velocities show a high degree of similarity among the different codes, as good as within a few percent at certain radii (averaged fractional deviation for $2 < r < 10$ kpc is 2.5% or 0.011 dex). Disagreements seen at radius > 12 kpc for some codes such as ART-II and GASOLINE are attributed to a small number statistics near the edge of a stellar disk.

[38] Careful readers may notice that the stronger fragmentation in GIZMO conflicts what Mayer et al. (2016) found. We caution that their setup is different from ours. For example, Mayer et al. (2016) employed a different pressure floor prescription that depends on the kernel size, and more effective stellar feedback based on Stinson et al. (2006). They modeled a low-metallicity massive gas rich galaxy without considering metal cooling, which was shown by the GIZMO group to significantly affect the fragmentation. As discussed, a slight variation in GIZMO parameters is enough to erase their discrepancy with other codes. In Mayer et al. (2016), this role may be played by the different subgrid models and the absence of metal cooling.

In Figure 25 we analyze the velocity dispersion curves for newly-formed star particles in *Sim-SFF*. As in gas velocity dispersion curves of Figures 13, this plot quantifies the residual velocity components other than the rotational velocity computed in Figure 24. Once again, a good agreement is found that all codes lie within a few tens of percent from one another at radii < 10 kpc (averaged fractional deviation for $2 < r < 10$ kpc is 11.2% or 0.046 dex). When compared with Figures 12 and 13, the agreement is particularly better in the central region (< 1.5 kpc). Note that when we plot the ratio of *vertical* velocity dispersion to *total* velocity dispersion in the bottom panel of Figure 25, the systematic discrepancy found in Figure 13 between mesh-based and particle-based codes at radii < 1.5 kpc no longer exists. This confirms our assertion in Section 6.2 that the said discrepancy in Figure 13 is due to vertical gas movement captured only in mesh-based codes.

6.6. *Star Formation Relation*

In this section, we compare the star formation rates (SFR) of different codes in *Sim-SFF* and check whether we reproduce the observed relation between gas surface density and SFR surface density. In the following discussion, SFRs are time-averaged over the past 20 Myrs and derived based on the ages of newly-formed star particles in the 500 Myr snapshots.

Figure 26 displays the evolution of the galaxy-wide SFR of each run by 500 Myr. For most codes, SFRs increase at early times, reach a maximum at 200-300 Myr after the start of the simulation, and then plateaus at later times. The SFRs evolve smoothly without evidence for strong bursts. The SFRs are within a factor of $\sim$3 from one another at all times (averaged fractional deviation for $50 < t < 500$ Myr is 32.8% or 0.123 dex), and for most codes the values settle at 2-4 $M_\odot\,yr^{-1}$ over most of the simulated time. Some differences are noteworthy. For example, GASOLINE and ENZO predict somewhat lower SFRs, especially at intermediate times, while the SFR

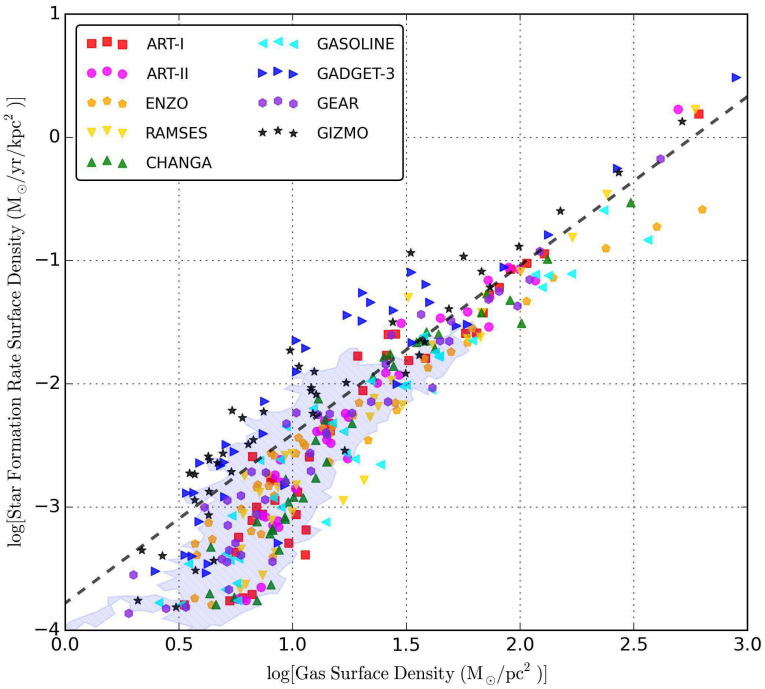

Figure 28. The Kennicutt-Schmidt relation for *Sim-SFF* at 500 Myr using the azimuthally-averaged gas surface densities (Figure 5) and SFR surface densities (Figure 27). The thick black dashed line denotes a best observational fit by Kennicutt et al. (2007). The blue hatched contour marks the observed sub-kpc patches in nearby galaxies by Bigiel et al. (2008), where their *hydrogen* surface density is multiplied by 1.36 to match the *total* gas surface density in our simulations. See Section 6.6 for a detailed explanation on how this figure is made. The full color version of this figure is available in the electronic edition.

for GIZMO never plateaus or begins to decline, but reaches a maximum of $\sim 6\ \mathrm{M_\odot\,yr^{-1}}$ at 500 Myr. GADGET-3 produces the most stellar mass in this time period, but its SFR does not further grow after $\sim$ 300 Myr. The total stellar mass formed in 500 Myr ranges from $0.8\times10^9\,\mathrm{M_\odot}$ in GASOLINE to $2.4\times10^9\,\mathrm{M_\odot}$ in GADGET-3. GIZMO's efficient and clumpy star formation is discussed in detail in Section 6.4 and Figure 23. We note again that increasing the Jeans pressure floor in GIZMO by a factor of 2 – well within the uncertainty in the geometry prefactor of Jeans pressure support equation, Eq. (3) – and reverting to GRACKLE v2.0 entirely removes GIZMO's discrepancy (see the black dashed line in Figure 26 labeled GIZMO-PS2; compare with Figure 23). In addition to the systematic tests the GIZMO group performed, the GADGET-3 group has also tried a run with a slight variation in the treatment of supernova feedback that would increase the number of gas particles inside the "shock radius" (Section 5.7). We find that this slight variation indeed produces significantly lower SFR for GADGET-3, closer to ENZO's value (results not shown here). Our tests strongly suggest that the SFR evolution is highly sensitive to the details of the numerical implementation of the common subgrid physics, including pressure floor and feedback prescriptions.

Now, to better understand the differences in SFR, we plot in Figure 27 the SFR surface densities as functions of cylindrical distance from the galactic center. Again, SFRs are estimated using the newly-formed star particles that are younger than 20 Myr old. The agreement between the codes is generally encouraging, especially outside the central 0.5 kpc. We note, however, that the SFR within 0.5 kpc constitutes a large fraction of the total galactic SFR. ENZO and GASOLINE have significantly lower SFRs in the central region of the galaxy, thus explaining the difference seen in Figure 26. In contrast, much of the excess star formation in GADGET-3 and GIZMO takes place at large radii and is likely related to the formation of larger numbers of stellar clumps, as seen in Figures 21 and 23. For example, a $\sim$ 6.5 kpc peak in GIZMO's SFR profile in Figure 27 coincides with a peak at the same radius in its stellar surface density profile in Figure 22.

When measured on galactic scales ($\sim$ kpc), the gas surface density and SFR surface density are tightly correlated (e.g., Schmidt 1959; Kennicutt 1989, 1998; Kennicutt et al. 2007; Bigiel et al. 2008). Consequently, the relation between these two observables, the Kennicutt-Schmidt (KS) relation, is frequently used to calibrate the modeling of star formation in galaxy-scale simulations. In Figure 28 we show the KS relation for *Sim-SFF*, where gas and SFR surface densities are measured within cylindrical annuli, as computed in Figures 5 and 27, respectively. Only annuli with nonzero averaged SFR surface densities are considered. The thick black dashed line denotes a best observational fit by Eq. (8) in Kennicutt et al. (2007), $\log\Sigma_{\rm SFR} = 1.37\log\Sigma_{\rm gas} - 3.78$, for a spatially-resolved patches in M51a. The blue hatched contour marks the observed sub-kpc patches in nearby galaxies taken from Figure 8 of Bigiel et al. (2008), where their *hydrogen* surface density is multiplied by 1.36 to account for helium to match the *total* gas density in our simulations (see their Section 2.3.1).[39]

As Figure 28 reveals, all participating codes predict a KS relation that agrees well with one another within a factor of a few, and with observed nearby disk galaxies in Bigiel et al. (2008). In particular, by combining star formation and thermal supernova feedback, most codes match both the normalization of the observed relation and the characteristic "threshold" value of the gas surface density ($\sim 10\,\mathrm{M_\odot\,pc^{-2}}$) below which star formation becomes less efficient. However, it

[39] We caution that galaxies in the Bigiel et al. (2008) sample are in nearby universe with relatively low gas fractions. These are slightly different from the initial condition of our experiment that is modeled as a disk galaxy at $z \sim 1$ with 20% gas fraction (see Section 2).

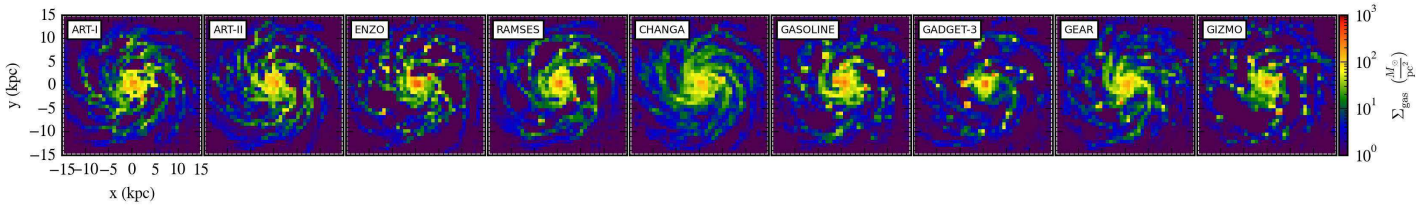

Figure 29. The 500 Myr snapshot of face-on gas surface densities for *Sim-SFF* at 750 pc resolution. For particle-based codes, surface densities are estimated by depositing gas particles via the Cloud-In-Cell (CIC) scheme on to a 2-dimensional uniform grid with 750 pc resolution. This image could be considered as a degraded version of Figure 3 although a different deposit algorithm is used for particle-based codes. See Section 6.6 for a detailed explanation on how this figure is made. The full color version of this figure is available in the electronic edition.

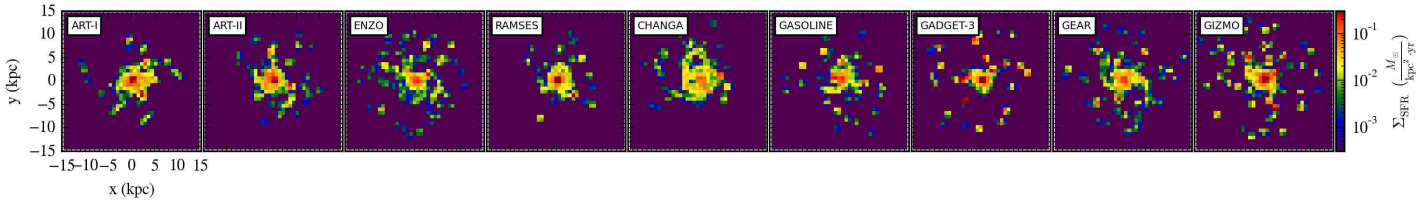

Figure 30. The 500 Myr snapshot of face-on star formation rate surface densities for *Sim-SFF* at 750 pc resolution. SFR surface densities are estimated by depositing the newly-formed star particles that are younger than 20 Myr old on to a uniform grid with 750 pc resolution. See Section 6.6 for a detailed explanation on how this figure is made. The dynamic range of the color axis (3 orders of magnitude) are kept identical among Figures 29 and 30 to help see if the gas depletion times are similar among pixels.

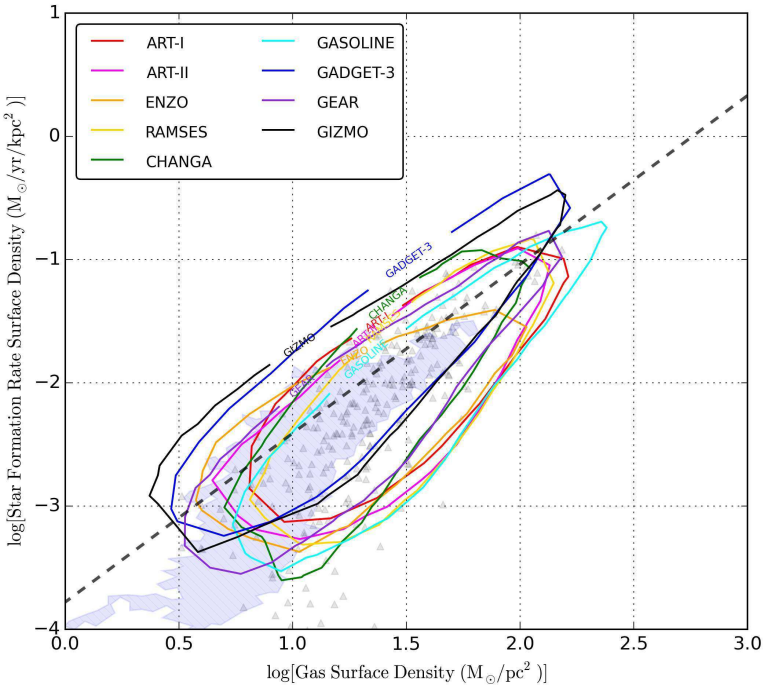

Figure 31. Same Kennicutt-Schmidt plane as Figure 28 but now with gas surface densities (Figure 29) and SFR surface densities (Figure 30) averaged in 750×750 pc patches at 500 Myr. As an example, shown as gray triangles are the patches with nonzero SFR surface density found in CHANGA. For all other codes, only 80% percentile contours are drawn. The thick black dashed line denotes a best observational fit by Kennicutt et al. (2007). The blue hatched contour marks the observed sub-kpc patches in nearby galaxies by Bigiel et al. (2008). See Section 6.6 for a detailed explanation on how this figure is made. The axes ranges are kept identical among Figures 28 and 31 for easier comparison. The full color version of this figure is available in the electronic edition.

should be noted that our simulations do not include multi-phase gas physics that explicitly models the transition between atomic and molecular hydrogen at $\sim 10\,{\rm M}_\odot\,{\rm pc}^{-2}$ (e.g., Krumholz et al. 2009). More investigation may be needed to check how the apparent change in slope seen here is affected by our choice of subgrid physics (e.g., star formation efficiency $\epsilon_\star$, or thermal feedback energy budget; see Section 3.2). Figure 28 also highlights that there are some differences between the codes. SFR surface densities of GASOLINE and ENZO lie slightly below the other codes at a given gas surface density, while GADGET-3 and GIZMO show higher SFRs than the rest of the codes. These differences are generally in line with what was observed in global SFR of Figure 26.

In order to better match the observational technique by Bigiel et al. (2008), one may consider using sub-kpc patches to generate the KS relation rather than cylindrical annuli. In Figures 29 and 30 we present mock observations of gas surface densities and SFR surface densities for *Sim-SFF* at 500 Myr. For mesh-based codes in Figure 29, their panels in Figure 3 are degraded to 750 pc resolution. For gas particles for

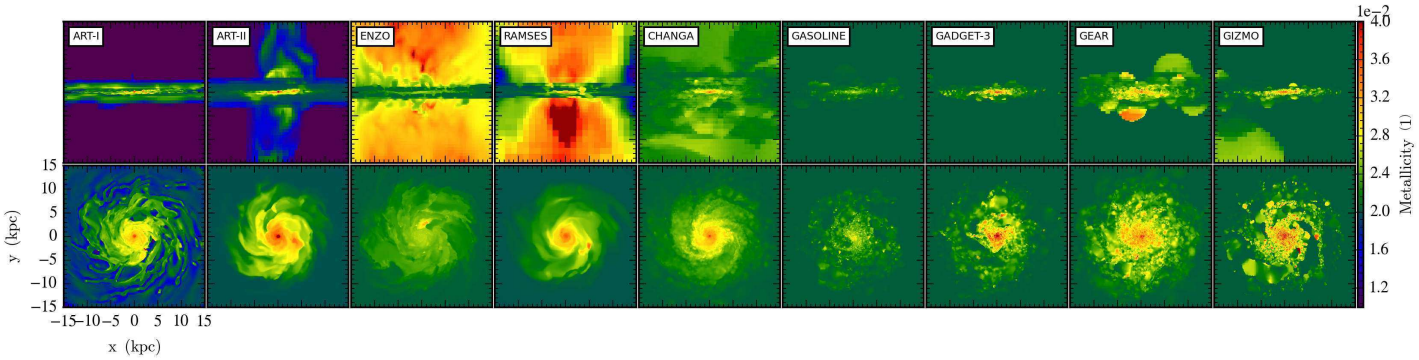

Figure 32. The 500 Myr composite of density-square-weighted gas metal fraction projections, edge-on *(top)* and face-on *(bottom)*, for *Sim-SFF* with star formation and feedback. Colors represent the ratio of metal density to total gas density (not in units of $Z_\odot$). The color axis spans from 0.01 to 0.04 in a linear scale. See Section 6.7 for a detailed explanation of this figure. Compare with Figures 3, 15, and 21. The full color version of this figure is available in the electronic edition. The high-resolution versions of this figure and article are available at the Project website, http://www.AGORAsimulations.org/.

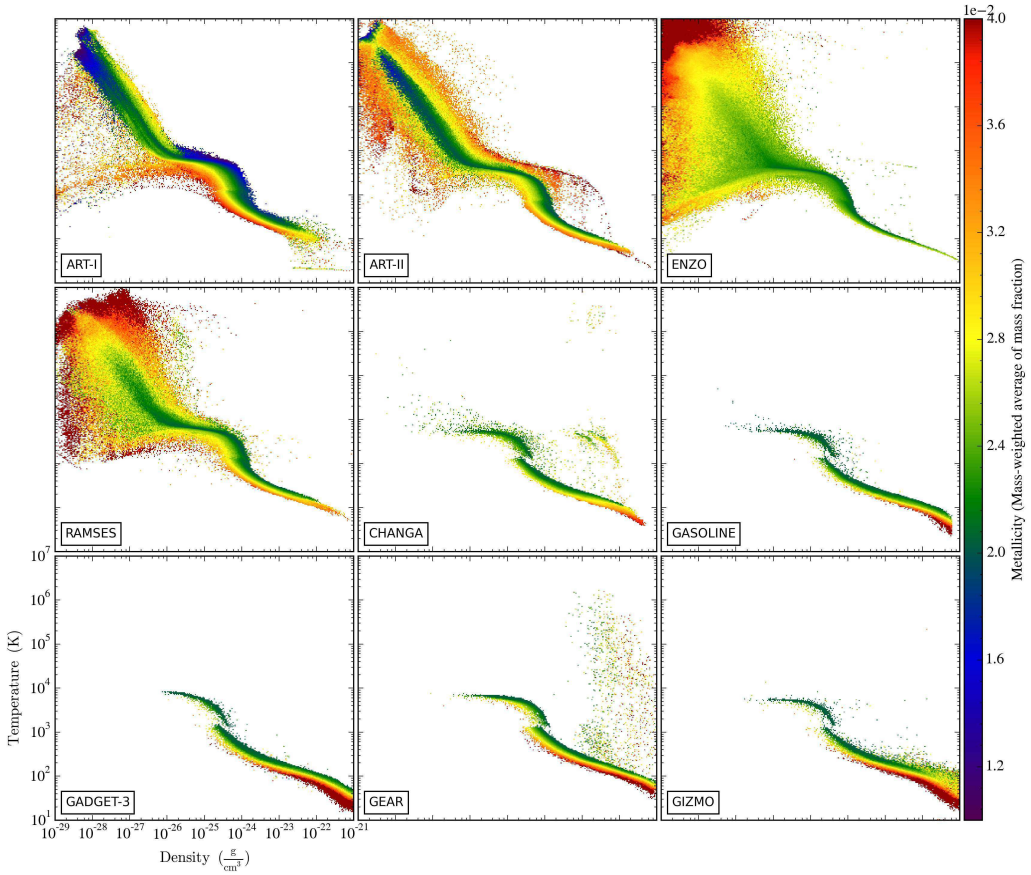

Figure 33. The 500 Myr composite of the mass-weighted averages of gas metal fraction on a density−temperature plane for the gas within 15 kpc from the galactic center in *Sim-SFF* with star formation and feedback. The metal fraction is simply the ratio of metal density to total gas density (not in units of $Z_\odot$). Note that a gaseous halo – low density, high temperature gas in the upper left corner of each panel – exists only in mesh-based codes, but not in particle-based codes (SPH codes or GIZMO). Compare with Figure 17. See Section 6.7 for more information on this figure. The full color version of this figure is available in the electronic edition. The high-resolution versions of this figure and article are available at the Project website, http://www.AGORAsimulations.org/.

particle-based codes in Figure 29 and young star particles of age less than 20 Myr in Figure 30, we use the Cloud-In-Cell (CIC) scheme to deposit particle masses on to a uniform 2-dimensional grid with 750 pc resolution.[40] This resolution matches Bigiel et al. (2008)'s reported working resolution. The dynamic range of the color axis are set to 3 orders of magnitude in both Figures 29 and 30, in order to help readers to see if the gas depletion times are similar among pixels.

Then we identify all 750×750 pc patches with nonzero SFR surface density, and plot them in Figure 31 on the same KS plane as Figure 28. As an example, all such patches found in the CHANGA run are shown as gray triangles. For all other codes, only 80% percentile contours are drawn. Again, all participating codes reproduce the slope and normalization of the observed KS relation well. But slight differences in SFR surface densities exist. Contours for GASOLINE and ENZO lie below the pack, while GADGET-3 and GIZMO's contours

[40] Thus, overall, Figure 29 could be considered as a "degraded" version of Figure 3 although, in Figure 3, smoothing kernels – not CIC – are used to deposit gas particles in particle-based codes on to an octree.

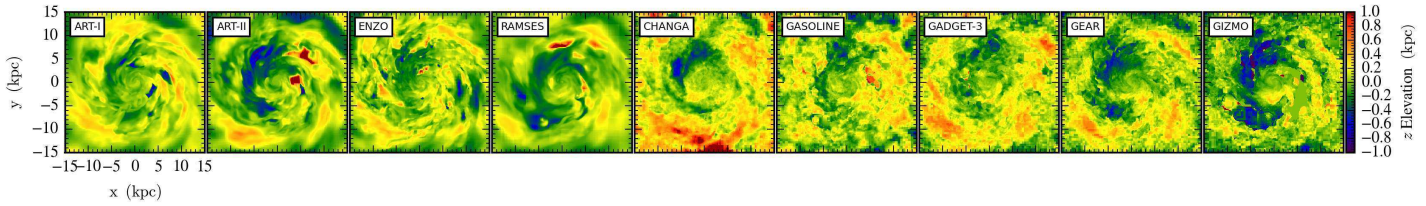

Figure 34. The 500 Myr face-on composite of density-weighted averages of gas elevations for *Sim-SFF* with star formation and feedback. See Section 6.7 for a detailed explanation on how this figure is made. Compare with Figures 3 and 15. The full color version of this figure is available in the electronic edition. The high-resolution versions of this figure and article are available at the Project website, http://www.AGORAsimulations.org/.

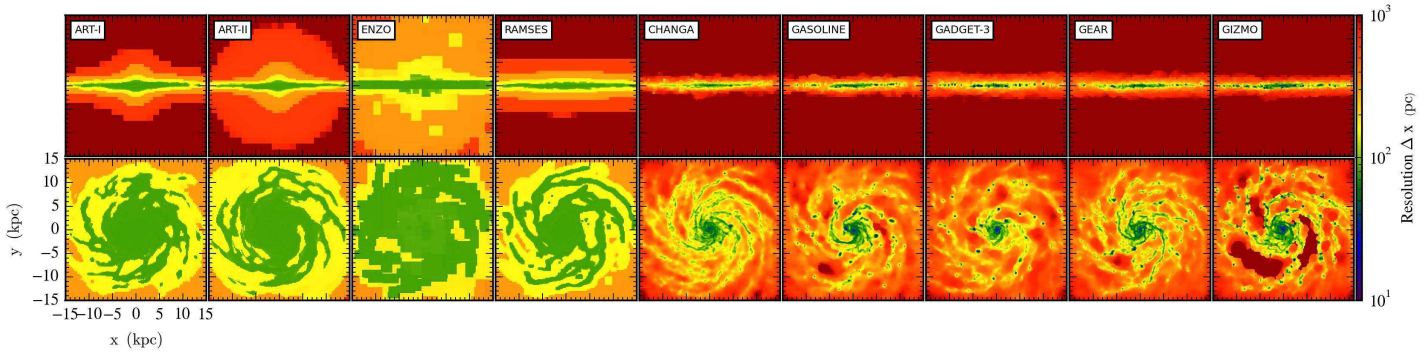

Figure 35. The 500 Myr composite of the size of resolution elements along different lines of sight, edge-on *(top)* and face-on *(bottom)*, for *Sim-SFF* with star formation and feedback. The color axis spans from 10 to 10^3 pc in a logarithmic scale, with highest resolution shown in dark blue. See Section 6.6 for a detailed explanation on how this figure is made. Compare with Figures 3, 15, and 21. The full color version of this figure is available in the electronic edition.

sit above all other codes at all gas surface densities. These findings are consistent with what we see in Figures 26 and 28.

6.7. *Other Comparisons: Metal Fraction, Disk Elevation, and Spatial Resolution*

In Figure 32 we present the projections of density-square-weighted gas metal fraction for *Sim-SFF* with star formation and feedback. The metal fraction we show here is simply the ratio of metal density to total gas density, and it is not in units of $Z_\odot$, in order to minimize any confusion caused by GRACKLE 2.0 versus 2.1 implementations (see footnotes 9 and 13). The color axis spans from 0.01 to 0.04 in a linear scale. The edge-on views of mesh-based codes, particularly ENZO and RAMSES, show high metallicity filaments flowing out of the disk, carrying metals into the embedding halo (see also Figure 15). However, as noted in Section 2, a gaseous halo exists only in mesh-based codes, but not in particle-based codes (SPH codes and GIZMO). Therefore, it is *not* the intended scope of this paper to compare the metal content of the halo which, by design, is captured only in mesh-based codes. In the face-on views within the disk we find qualitatively similar results across codes, with denser gas (corresponding to star forming regions) tending to be more metal-enriched. Significant differences in the morphology of metal distribution exist, however. Differences in numerical implementations of the stellar feedback model are responsible for such discrepancies (see Section 6.3 for more discussion).[34]

Figure 33 shows the mass-weighted averages of gas metal fraction in *Sim-SFF* on a 2-dimensional density−temperature plane for gas within 15 kpc of the galactic center. Raw particle fields are used for particle-based codes, not the interpolated or smoothed fields constructed in `yt`. Note again that a gaseous halo represented by low density, high temperature gas in the upper left corner of each panel exists only in mesh-based codes. For particle-based codes, high density, low temperature gas has higher metallicity because of its correspondingly higher star formation rate and thus metal enrichment. For mesh-based codes, on the other hand, high metal fraction is found in low density, high temperature gas as well, which is contaminated by hot metal-enriched materials dispersed by supernova feedback.[41] These two observations are related to how well metals get mixed in each code. Readers may have noticed the difference between mesh-based and particle-based codes already in Figure 32 by focusing on mixing of the metals. With neither the halo gas nor a sophisticated metal mixing scheme in place, metal-enriched gas in particle-based codes tends to stay near the dense star-forming sites that provided the metals.

In Figure 34 is the density-weighted average of gas elevations from the $x-y$ disk plane for *Sim-SFF* with star formation and feedback. That is, averages of $(z_\mathrm{i} - z_\mathrm{center})$ such that a positive (negative) value indicates the gas along that line of sight is located above (below) the disk plane on average. For particle-based codes, we use the reconstructed density field from `yt`'s in-memory octree on which gas particles are deposited using smoothing kernels (see Section 6.1). This figure helps to visualize and estimate the warping of the gas disk (e.g., Levine et al. 2006). All participating codes produce largely flat disks, with vertical offsets less than ± 1 kpc. Yet, it is also true that all codes show some levels of coherent warping or flaring along the disk plane. This strongly suggests that all these codes are able to resolve vertical instabilities.

Finally, Figure 35 compares the sizes of spatial resolution elements each code imposes on the galactic disk of *Sim-SFF* at 500 Myr. For mesh-based codes, this is a projection of gas cell sizes along different lines of sight, weighted by (cell volume)$^{-2}$ so that the maximal resolution element along that

[41] Readers may notice that ART-I does not show strong outflows. Rather, the continued inflow of gas from the halo provides zero metallicity gas mixed in to the disk (see the initial halo setup in Section 2). ART-I's discrepancies in Figures 32-33 are manifestations of weaker stellar feedback than other mesh-based codes, which are exaggeratedly visible only because mesh-based codes include a large reservoir of zero metallicity gas around the disk.

Table 2
Relative Differences Between Codes in Main Observables (for *Sim-SFF* run)

	Level of agreement between codes	Figures	Relevant sections
Gas surface density (cylindrically-binned)	averaged fractional deviation for $2 < r < 10$ kpc = 32.2% (0.121 dex)[ee]	Figure 5	Section 6.1
Gas surface density (vertically-binned)	averaged fractional deviation for $z < 0.6$ kpc = 30.4% (0.115 dex)	Figure 7	Section 6.1
Gas average vertical height	averaged fractional deviation for $2 < r < 10$ kpc = 19.1% (0.076 dex)	Figure 9	Section 6.1
Gas rotation velocity	averaged fractional deviation for $2 < r < 10$ kpc = 2.8% (0.012 dex)	Figure 11	Section 6.2
Gas velocity dispersion	averaged fractional deviation for $2 < r < 10$ kpc = 17.8% (0.071 dex)	Figure 13	Section 6.2
Gas density probability distribution	averaged fractional deviation in $10^{-25} < \rho < 10^{-22}\,\mathrm{g\,cm^{-3}}$ = 28.6% (0.109 dex), up to more than an order of magnitude difference at $\rho > 10^{-22}\,\mathrm{g\,cm^{-3}}$	Figure 19	Section 6.3
Newly-formed stellar surface density	averaged fractional deviation for $2 < r < 10$ kpc = 53.9% (0.187 dex)	Figure 22	Section 6.4
Newly-formed stellar clump mass function	all data points lie within a factor of ∼3 from the mean at each mass	Figure 23	Section 6.4
Newly-formed stellar rotation velocity	averaged fractional deviation for $2 < r < 10$ kpc = 2.5% (0.011 dex)	Figure 24	Section 6.5
Newly-formed stellar velocity dispersion	averaged fractional deviation for $2 < r < 10$ kpc = 11.2% (0.046 dex)	Figure 25	Section 6.5
Galaxy-wide star formation rate	averaged fractional deviation for $50 < t < 500$ Myr = 32.8% (0.123 dex)	Figure 26	Section 6.6
KS relation (azimuthally-averaged)	all data points lie within a factor of ∼3 from the mean at each $\Sigma_{\rm gas}$	Figure 28	Section 6.6
KS relation (patch-averaged)	all data points lie within a factor of ∼3 from the mean at each $\Sigma_{\rm gas}$	Figure 31	Section 6.6

line of sight could stand out. For particle-based codes, this is a projection of gas particle sizes, defined as $(m_{\rm gas}/\rho_{\rm gas})^{1/3}$, smoothed on to `yt`'s octree (the same octree `yt` used in other edge-on/face-on visualizations; see Section 6.1 for more information on the octree), weighted by (particle volume)$^{-2}$, defined as $(m_{\rm gas}/\rho_{\rm gas})^{-2}$. The color axis spans from 10 to 10^3 pc in a logarithmic scale, with highest resolution shown in dark blue. The green color in mesh-based codes marks the finest mesh size permitted, 80 kpc, while the blue color in the spirals and clumps of particle-based codes can be associated with the minimum smoothing length permitted, 0.2×80 pc. Because of its Lagrangian nature, particle-based codes best demonstrate their strengths in dense clumps and spirals. Meanwhile, with its flexible refinement strategy, mesh-based codes may best utilize their strengths in the contact regions with high density contrast, such as above and below the disk. For example, in the edge-on views of Figure 35, the green-colored high resolution region covering the disk is thicker in mesh-based codes than in particle-based codes.[42]

7. DISCUSSION AND CONCLUSION

Using an isolated Milky Way-mass galaxy simulation, we compared results from 9 state-of-the-art gravito-hydrodynamics codes widely used in the numerical community (Section 5). We utilized the infrastructure we have built for the *AGORA* High-resolution Galaxy Simulations Comparison Project. For the common initial conditions for these isolated galaxy simulations we used the ones generated by MAKEDISK (Section 2).[7] We also adopted the common physics models (e.g., radiative cooling and UV background by the standardized package GRACKLE; Section 3.1)[3] and common analysis toolkit `yt`,[4] both of which are publicly available. Subgrid physics such as pressure floor, star formation prescription, supernova feedback energy, and metal production have been meticulously constrained across participating codes (Sections 3.1 and 3.2). Strenuous efforts have also been made to ensure the consistency between the parameters that control resolutions of the codes (Section 4).

[42] From the face-on views of Figure 35, readers may distinguish the differences in grid construction machineries of mesh-based codes: octree structures in ART-I, ART-II, and RAMSES, versus block structures in ENZO.

With numerical accuracy that resolves the disk scale height – high-order numerical methods in modern simulation codes combined with high spatial resolution – we find that the codes overall agree well with one another in many dimensions, including: gas and stellar surface densities, gas and stellar rotation curves and velocity dispersions, gas density and temperature distribution functions, disk vertical heights, newly-formed stellar clumps, SFRs, and Kennicutt-Schmidt relations (Section 6). Quantities such as velocity dispersions are very robust (e.g., gas and newly-formed stellar velocity dispersions agree within a few tens of percent at all radii) while other measures like newly-formed stellar clump mass functions show more significant variation (differ by up to a factor of ∼3). In Table 2 we summarize the relative differences between codes for the main observables studied in this report. Some discrepancies can be understood as systematic differences between codes, for example, between mesh-based and particle-based codes in the low density region (Figures 2-3 and Section 6.1), and between more diffusive and less diffusive schemes in the high density tail of the density distribution (Figure 19 and Section 6.3). The latter translates into differences in clumping properties (Figure 23) and star formation rates (Figure 26) of different codes. These intrinsic code differences are not as serious as some might have mistakenly extrapolated from previous code comparisons (e.g., Scannapieco et al. 2012), and are generally small compared to the variations in numerical implementations of the common subgrid physics such as supernovae feedback. Our experiment reveals the remarkable level of agreement between different modern simulation tools despite their codebases having evolved largely independently for many years. It is also reassuring that our computational tools are more sensitive to input physics than to intrinsic differences in numerical schemes, and that predictions made by the participating numerical codes are reproducible and likely reliable. If adequately designed in accordance with our proposed common parameters (e.g., cooling, metagalactic UV background, stellar physics, resolution; see Sections 3 and 4), results of a modern high-resolution galaxy formation simulation are likely robust.

It is worth briefly noting a few points about our study presented in this paper: (1) During the course of the present study, we have developed and field-tested important pieces of

the *AGORA* infrastructure such as the common initial condition, common physics models, and common analysis toolkit. In particular, it should be noted that all the analyses in Section 6 are carried out with common `yt` scripts that are nearly independent of simulation codes. This common analysis platform approach has repeatedly proven its strength in *AGORA* comparisons including this study, significantly reducing the cost needed to hack the codes that any one researcher might not be familiar with, and allowing moving straight to *science-driven* comparisons of underlying physical properties. (2) While we find that the 500 Myr snapshot we used for the comparison is representative of each simulation, we caution that any similarity or discrepancy found here may not be universally the case at every single epoch. In an ongoing study using the same suite of simulations presented here, but multiple snapshots up to 2 Gyr, we are systematically checking if any conclusion drawn in this work is challenged by the fact that we compared a snapshot of a galaxy at a single epoch.[43] (3) Comparison studies in the *AGORA* Collaboration including this work are not intended to decide which numerical implementation is a "correct" one. The problem we are trying to numerically solve, i.e., galaxy formation, does not have a well-defined solution at a given resolution that every code is expected to converge to. Thus, it is never our intention to identify a "correct" or "incorrect" code, nor even a "better" or "worse" code. Instead, we aim to determine how much scatter one should expect among different numerical implementations in a particular problem of galaxy formation, given nominally similar physics and runtime parameters.

We plan to further investigate our isolated disk galaxy simulations in other interesting dimensions, such as disk stability, bulge-to-disk decomposition, spiral and bar formation, and mass inflow and outflow, among others. As mentioned in Section 3.2, we also intend to calibrate feedback schemes against observations (e.g., metrics such as galactic fountain, outflows, mass-loading, fraction of hot/warm/cold gas, and main sequence star formation) and against one another. While we complete analyses for these ongoing efforts, we are aiming to publicly release the 500 Myr snapshots used in the present article from all participating codes in January 2017 (tentatively) through the *AGORA* Project website.[2] This is to allow any interested party in the numerical galaxy formation community to be able to compare their own simulations with the *AGORA* snapshots, using our publicly available common `yt` script if needed.[29]

Finally, we emphasize the role the *AGORA* Project played in promoting *collaborative and reproducible research* in the numerical galaxy formation community. Over the past four years we have collaboratively formed a one-of-a-kind platform where members of the numerical community can work together and verify one another's work. Not only have we successfully built a common, publicly available infrastructure fully encompassing all the components to run galaxy-scale simulations in a reproducible manner – initial conditions, physics packages, calibrated runtime parameters, analysis pipeline, and data storage – but we also have founded an open forum where members could talk to and learn from one another. This Project has become a great experiment in itself in which it was continuously shown how beneficial a platform like this could be for any scientific community. Through workshops and teleconferences, and via common languages and infrastructure built together, Project participants were able to better understand other codes, and improve their own. Participants found an optimal set of simulation parameters that makes their code to be best compatible with others. We came to understand how seemingly identical parameters differ in their meanings in different codes, and how seemingly different parameters have in fact identical meanings. In some comparisons, numerical errors were discovered and fixed in participating codes. The *AGORA* framework, now tested with the common physics and subgrid models, are serving as a launchpad to initiate *astrophysically-motivated* comparisons aimed at raising the predictive power of galaxy simulations, especially as we run the zoom-in cosmological simulations outlined in our flagship paper (Kim et al. 2014). In the coming years, we expect *AGORA* to continue to provide a sustainable and fertile platform on which numerical experiments are readily validated and cross-calibrated, and ambitious multi-platform collaborations are forged.

The authors of this paper thank the members of the *AGORA* Collaboration who are not on the author list but have provided helpful suggestions throughout the progress of the paper, including John Forbes. We also thank Volker Springel for providing the original versions of GADGET-3 and MAKEDISK to be used in the *AGORA* Project. We gratefully acknowledge the financial and logistical support from the University of California High-Performance AstroComputing Center (UC-HiPACC) during the annual *AGORA* Workshops held at the University of California Santa Cruz from 2012 to 2016. This research also used resources of the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The publicly available ENZO and `yt` codes used in this work are the products of collaborative efforts by many independent scientists from numerous institutions around the world. Their commitment to open science has helped make this work possible. Ji-hoon Kim acknowledges support from NASA through an Einstein Postdoctoral Fellowship, grant PF4-150147, and support from the Moore Center for Theoretical Cosmology and Physics at Caltech. A part of his computing time was provided by Extreme Science and Engineering Discovery Environment (XSEDE) allocation TG-AST140064. XSEDE is supported by National Science Foundation (NSF) grant No. ACI-1053575. He is also grateful for the support from the computational team at SLAC National Accelerator Laboratory during the usage of the clusters for the simulation analysis. Oscar Agertz acknowledges support from STFC consolidated grant ST/M000990/1 and the Swedish Research Council grant 2014-5791. Daniel Ceverino acknowledges support from the European Research Council (ERC) via the ERC Advanced Grant STARLIGHT Project No. 339177. Robert Feldmann acknowledges support in part by NASA through Hubble Fellowship grant HF2-51304.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555, in part by the Theoretical Astrophysics Center at UC Berkeley, and by NASA ATP grant 12-ATP-120183. Alessandro Lupi acknowledges support by the ERC Project No. 267117 (PI J. Silk). Tom Quinn acknowledges partial support by the

[43] In fact, the evolution of the same *AGORA* isolated IC adopted in this work has been studied at different epochs already using many of the participating codes, albeit with different input physics and runtime parameters (e.g., Agertz et al. 2013; Keller et al. 2014; Goldbaum et al. 2015, 2016; Aoyama et al. 2016; Semenov et al. 2016).

NSF through grants No. AST-1514868 and AST-1311956 and NASA Hubble grant HST-AR-13264. CHANGA simulations where run on resources provided by XSEDE and NASA Pleiades. Robert Thompson, Matthew Turk and Nathan Goldbaum acknowledge support by the Gordon and Betty Moore Foundation's Data-Driven Discovery Initiative through grant GBMF4561 (PI M. Turk), and by the NSF through grant No. ACI-1535651. Tom Abel acknowledges partial support by the Kavli Foundation. Sukanya Chakrabarti acknowledges support by the NSF through grant No. AST-1517488. Piero Madau acknowledges support by the NSF through grant No. AST-1229745 and by NASA through grant NNX12AF87G. Kentaro Nagamine and Ikkoh Shimizu acknowledge support from the JSPS KAKENHI grant No. JP26247022. Some of the GADGET-3 simulations were carried out on the XC30 machine at the Center for Computational Astrophysics, National Astronomical Observatory of Japan. Brian O'Shea acknowledges support from NASA through grants NNX12AC98G, NNX15AP39G, and NASA Hubble theory grants HST-AR-13261.01-A and HST-AR-14315.001-A. He was also supported in part by the sabbatical visitor program at the Michigan Institute for Research in Astrophysics at the University of Michigan in Ann Arbor, and gratefully acknowledges their hospitality. Joel Primack acknowledges support from STScI through grant HST-GO-12060.12-A-004, and NASA Advanced Supercomputing for Pleiades time on which ART-I simulations were run. Christine Simpson acknowledges support from the ERC under ERC-StG grant EXAGAL-308037 and from the Klaus Tschira Foundation. John Wise acknowledges support by the NSF through grants No. AST-1333360 and AST-1614333 and NASA Hubble theory grants HST-AR-13895 and HST-AR-14326.

APPENDIX

A. PRESSURE FLOOR PARAMETERS IN SELECTED CODES

We caution that the definition of $N_{\rm Jeans}$ in our Jeans pressure floor recommendation, Eq. (3),

$$P_{\rm Jeans} = \frac{1}{\gamma\pi} N_{\rm Jeans}^2 G\rho_{\rm gas}^2 \Delta x^2 \tag{3}$$

is different from another widely used formula in particle-based codes, Eq. (1) of Hopkins et al. (2011),

$$P_{\rm Jeans} = \frac{1.2}{\gamma} N_{\rm Jeans}^{2/3} G\rho_{\rm gas}^2 h_{\rm sml}^2. \tag{A1}$$

Our choice of $N_{\rm Jeans} = 4$ in Eq. (3) (see Section 3.1) is equivalent to $N_{\rm Jeans} = 8.75$ in Eq. (A1) if the smoothing length $h_{\rm sml}$ is replaced with a fixed number $\Delta x = 80$ pc. For GEAR, the runtime parameter `JeansMassFactor` controls the Jeans pressure support and it equates to $N_{\rm Jeans}$ in Eq. (A1), to be set to 8.75. For CHANGA and GASOLINE, the relevant runtime parameter is `dResolveJeans`, but it is not equal to $N_{\rm Jeans}$ in Eq. (A1), but to the entire prefactor $1.2 N_{\rm Jeans}^{2/3} \gamma^{-1}$. Therefore, to follow the *AGORA* recommendation in Section 3.1, `dResolveJeans` should be set to $1.2 \times (8.75)^{2/3} \times (5/3)^{-1} = 3.06$, along with replacing $h_{\rm sml}$ with 80 pc. For some codes, one of the pressure floor formulae is simply hardcoded without a tunable runtime parameter: Eq. (A1) in GADGET-3, and Eq. (3) in GIZMO. The parameter `MinimumPressureSupportParameter` in ENZO refers to $N_{\rm Jeans}^2$ in Eq. (3) above (see `Grid_SetMinimumSupport.C`). Thus, to follow the *AGORA* recommendation it should be set to 16. For pressure floor implementations using polytropes in ART-II and RAMSES, see Sections 5.2 and 5.4, respectively.

B. STAR FORMATION EFFICIENCY PARAMETERS IN CHANGA AND GASOLINE

Unlike in other codes, the parameter which controls the star formation efficiency in CHANGA and GASOLINE, `CStar` ($c_\star$ defined in Eq. (3a) of Katz 1992), is not equal to $\varepsilon_\star$ in our Eq. (4) of Section 3.2. This is due to a difference in the denominator of the Schmidt formula: $t_{\rm g} = (1/(4\pi G\rho_{\rm gas}))^{1/2}$ in Eq. (3a) of Katz (1992), versus $t_{\rm ff} = (3\pi/(32 G\rho_{\rm gas}))^{1/2}$ in our Eq. (4). Therefore, in order to follow the *AGORA* recommendation in Section 3.2, `CStar` should be set to $\varepsilon_\star(t_{\rm g}/t_{\rm ff}) = 0.01 \times (32/12\pi^2)^{1/2} = 0.0052$ in CHANGA and GASOLINE.

C. SOFTENING AND SMOOTHING PARAMETERS IN PARTICLE-BASED CODES

Throughout this paper we define the gravitational softening length $\varepsilon_{\rm grav}$ (Section 4.1) as the equivalent Plummer softening length. This equates to the parameters `Softening[Gas/Halo/Disk/Stars]` in GADGET-3, GEAR and GIZMO. It is however different from the typical definition of spline size h beyond which the gravitational force becomes exactly Newtonian (h as defined in Eq. (4) of Springel 2005). For GADGET-3, GEAR and GIZMO, the spline size h is equal to $2.8\varepsilon_{\rm grav} = 2.8 \times 80 = 224$ pc (see `ForceSoftening[i]` in `gravtree.c`). Note that the parameter `MinGasHsmlFractional` that controls the minimum hydrodynamical smoothing length (Section 4.2) is defined as a fraction of h, not of $\varepsilon_{\rm grav}$. Therefore, in order to set the minimum smoothing length to $0.2\varepsilon_{\rm grav}$, `MinGasHsmlFractional` should be set to $0.2/2.8 = 0.0714$.

By contrast, the spline size parameter `dSoft` in CHANGA and GASOLINE is not equal to the equivalent Plummer softening length, but to $h/2$ in the Springel (2005) definition of h above.[44] Therefore, in order to have $\varepsilon_{\rm grav} = 80$ pc in CHANGA or GASOLINE runs, `dSoft` should be set to $2.8\varepsilon_{\rm grav}/2 = 224/2 = 112$ pc. One should keep in mind that the parameter which controls the minimum hydrodynamical smoothing length, `dhMinOverSoft`, is defined as a fraction of `dSoft`, not of $\varepsilon_{\rm grav}$. Therefore, in order to set the minimum smoothing length to $0.2\varepsilon_{\rm grav}$, `dhMinOverSoft` should be set to $0.2 \times (2/2.8) = 0.143$.

REFERENCES

Agertz, O., Kravtsov, A. V., Leitner, S. N., & Gnedin, N. Y. 2013, ApJ, 770, 25
Agertz, O., Moore, B., Stadel, J., Potter, D., Miniati, F., Read, J., Mayer, L., Gawryszczak, A., Kravtsov, A., Nordlund, Å., Pearce, F., Quilis, V., Rudd, D., Springel, V., Stone, J., Tasker, E., Teyssier, R., Wadsley, J., & Walder, R. 2007, MNRAS, 380, 963
Almgren, A. S., Bell, J. B., Lijewski, M. J., Lukić, Z., & Van Andel, E. 2013, ApJ, 765, 39
Aoyama, S., Hou, K.-C., Shimizu, I., Hirashita, H., Todoroki, K., Choi, J.-H., & Nagamine, K. 2016, arXiv:1609.07547
Balsara, D. S. 1995, Journal of Computational Physics, 121, 357
Begley, C. G., & Ioannidis, J. P. 2015, Circulation Research, 116, 116
Berger, M. J., & Colella, P. 1989, Journal of Computational Physics, 82, 64
Berger, M. J., & Oliger, J. 1984, Journal of Computational Physics, 53, 484
Bigiel, F., Leroy, A., Walter, F., Brinks, E., de Blok, W. J. G., Madore, B., & Thornley, M. D. 2008, AJ, 136, 2846

[44] It is worth to point out that `dSoft` is indeed equal to h defined in other papers, e.g., Eq. (14) of Katz et al. (1996) or Eq. (21) of the original Monaghan & Lattanzio (1985).

Bryan, G. L., & Norman, M. L. 1997, arXiv:9710187
Bryan, G. L., Norman, M. L., O'Shea, B. W., Abel, T., Wise, J. H., Turk, M. J., Reynolds, D. R., Collins, D. C., Wang, P., Skillman, S. W., Smith, B., Harkness, R. P., Bordner, J., Kim, J.-h., Kuhlen, M., Xu, H., Goldbaum, N., Hummels, C., Kritsuk, A. G., Tasker, E., Skory, S., Simpson, C. M., Hahn, O., Oishi, J. S., So, G. C., Zhao, F., Cen, R., Li, Y., & Enzo Collaboration. 2014, ApJS, 211, 19
Bryan, G. L., Norman, M. L., Stone, J. M., Cen, R., & Ostriker, J. P. 1995, Computer Physics Communications, 89, 149
Catinella, B., Schiminovich, D., Kauffmann, G., Fabello, S., Wang, J., Hummels, C., Lemonias, J., Moran, S. M., Wu, R., Giovanelli, R., Haynes, M. P., Heckman, T. M., Basu-Zych, A. R., Blanton, M. R., Brinchmann, J., Budavári, T., Gonçalves, T., Johnson, B. D., Kennicutt, R. C., Madore, B. F., Martin, C. D., Rich, M. R., Tacconi, L. J., Thilker, D. A., Wild, V., & Wyder, T. K. 2010, MNRAS, 403, 683
Ceverino, D., & Klypin, A. 2009, ApJ, 695, 292
Ceverino, D., Klypin, A., Klimek, E. S., Trujillo-Gomez, S., Churchill, C. W., Primack, J., & Dekel, A. 2014, MNRAS, 442, 1545
Chabrier, G. 2003, ApJL, 586, L133
Chevalier, R. A. 1974, ApJ, 188, 501
Choi, J.-H., & Nagamine, K. 2012, MNRAS, 419, 1280
Colella, P., & Glaz, H. M. 1985, Journal of Computational Physics, 59, 264
Colella, P., & Woodward, P. R. 1984, Journal of Computational Physics, 54, 174
Dehnen, W., & Aly, H. 2012, MNRAS, 425, 1068
Durier, F., & Dalla Vecchia, C. 2012, MNRAS, 419, 465
Efstathiou, G., Davis, M., White, S. D. M., & Frenk, C. S. 1985, ApJS, 57, 241
Eisenstein, D. J., & Hut, P. 1998, ApJ, 498, 137
Everett, J., & Earp, B. 2015, Frontiers in Psychology, 6, 1152
Faucher-Giguère, C.-A., Lidz, A., Zaldarriaga, M., & Hernquist, L. 2009, ApJ, 703, 1416
Ferland, G. J., Porter, R. L., van Hoof, P. A. M., Williams, R. J. R., Abel, N. P., Lykins, M. L., Shaw, G., Henney, W. J., & Stancil, P. C. 2013, Revista Mexicana de Astronomia y Astrofisica, 49, 137
Few, C. G., Dobbs, C., Pettitt, A., & Konstandin, L. 2016, MNRAS, 460, 4382
Frenk, C. S., White, S. D. M., Bode, P., Bond, J. R., Bryan, G. L., Cen, R., Couchman, H. M. P., Evrard, A. E., Gnedin, N., Jenkins, A., Khokhlov, A. M., Klypin, A., Navarro, J. F., Norman, M. L., Ostriker, J. P., Owen, J. M., Pearce, F. R., Pen, U.-L., Steinmetz, M., Thomas, P. A., Villumsen, J. V., Wadsley, J. W., Warren, M. S., Xu, G., & Yepes, G. 1999, ApJ, 525, 554
Fryxell, B., Olson, K., Ricker, P., Timmes, F. X., Zingale, M., Lamb, D. Q., MacNeice, P., Rosner, R., Truran, J. W., & Tufo, H. 2000, ApJS, 131, 273
Gaburov, E., & Nitadori, K. 2011, MNRAS, 414, 129
Gingold, R. A., & Monaghan, J. J. 1977, MNRAS, 181, 375
Gnedin, N. Y., & Kravtsov, A. V. 2011, ApJ, 728, 88
Godunov, S. K. 1959, Matematicheskii Sbornik, 89, 271
Goldbaum, N. J., Krumholz, M. R., & Forbes, J. C. 2015, ApJ, 814, 131
—. 2016, ApJ, 827, 28
Haardt, F., & Madau, P. 1996, ApJ, 461, 20
—. 2012, ApJ, 746, 125
Hernquist, L. 1990, ApJ, 356, 359
Hernquist, L., & Katz, N. 1989, ApJS, 70, 419
Hopkins, P. F. 2013, MNRAS, 428, 2840
—. 2015, MNRAS, 450, 53
Hopkins, P. F., Kereš, D., Oñorbe, J., Faucher-Giguère, C.-A., Quataert, E., Murray, N., & Bullock, J. S. 2014, MNRAS, 445, 581
Hopkins, P. F., Quataert, E., & Murray, N. 2011, MNRAS, 417, 950
Kale, L. V., & Krishnan, S. 1996, in Parallel Programming using C++, ed. G. V. Wilson & P. Lu (MIT Press), 175–213
Katz, N. 1992, ApJ, 391, 502
Katz, N., Weinberg, D. H., & Hernquist, L. 1996, ApJS, 105, 19
Keller, B. W., Wadsley, J., Benincasa, S. M., & Couchman, H. M. P. 2014, MNRAS, 442, 3013
Kennicutt, Jr., R. C. 1989, ApJ, 344, 685
—. 1998, ApJ, 498, 541
Kennicutt, Jr., R. C., Calzetti, D., Walter, F., Helou, G., Hollenbach, D. J., Armus, L., Bendo, G., Dale, D. A., Draine, B. T., Engelbracht, C. W., Gordon, K. D., Prescott, M. K. M., Regan, M. W., Thornley, M. D., Bot, C., Brinks, E., de Blok, E., de Mello, D., Meyer, M., Moustakas, J., Murphy, E. J., Sheth, K., & Smith, J. D. T. 2007, ApJ, 671, 333
Khokhlov, A. 1998, Journal of Computational Physics, 143, 519
Kim, J.-h., Abel, T., Agertz, O., Bryan, G. L., Ceverino, D., Christensen, C., Conroy, C., Dekel, A., Gnedin, N. Y., Goldbaum, N. J., Guedes, J., Hahn, O., Hobbs, A., Hopkins, P. F., Hummels, C. B., Iannuzzi, F., Keres, D., Klypin, A., Kravtsov, A. V., Krumholz, M. R., Kuhlen, M., Leitner, S. N., Madau, P., Mayer, L., Moody, C. E., Nagamine, K., Norman, M. L., Onorbe, J., O'Shea, B. W., Pillepich, A., Primack, J. R., Quinn, T., Read, J. I., Robertson, B. E., Rocha, M., Rudd, D. H., Shen, S., Smith, B. D., Szalay, A. S., Teyssier, R., Thompson, R., Todoroki, K., Turk, M. J., Wadsley, J. W., Wise, J. H., Zolotov, A., & for the AGORA Collaboration. 2014, ApJS, 210, 14
Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009, ApJ, 699, 850
Lanson, N., & Vila, J.-P. 2008, SIAM J. Numer. Anal., 46, 1912
Levine, E. S., Blitz, L., & Heiles, C. 2006, ApJ, 643, 881
Lucy, L. B. 1977, AJ, 82, 1013
Mayer, L., Tamburello, V., Lupi, A., Keller, B., Wadsley, J., & Madau, P. 2016, ApJL, 830, L13
McKee, C. F., & Ostriker, J. P. 1977, ApJ, 218, 148
Menon, H., Wesolowski, L., Zheng, G., Jetley, P., Kale, L., Quinn, T., & Governato, F. 2015, Computational Astrophysics and Cosmology, 2, 1
Monaghan, J. J. 1992, ARAA, 30, 543
Monaghan, J. J., & Gingold, R. A. 1983, Journal of Computational Physics, 52, 374
Monaghan, J. J., & Lattanzio, J. C. 1985, Astronomy & Astrophysics, 149, 135
Morris, J. P. 1996, PASA, 13, 97
Nature Survey. 2016, Nature, 533, 452
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493
Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Paluck, E. L., Simonsohn, U., Soderberg, C., Spellman, B. A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E. J., Wilson, R., & Yarkoni, T. 2015, Science, 348, 1422
Open Science Collaboration. 2015, Science, 349, aac4716
O'Shea, B. W., Bryan, G., Bordner, J., Norman, M. L., Abel, T., Harkness, R., & Kritsuk, A. 2004, arXiv:0403044
O'Shea, B. W., Nagamine, K., Springel, V., Hernquist, L., & Norman, M. L. 2005, ApJS, 160, 1
Read, J. I., & Hayfield, T. 2012, MNRAS, 422, 3037
Revaz, Y., Arnaudon, A., Nichols, M., Bonvin, V., & Jablonka, P. 2016, Astronomy & Astrophysics, 588, A21
Revaz, Y., & Jablonka, P. 2012, Astronomy & Astrophysics, 538, A82
Ritchie, B. W., & Thomas, P. A. 2001, MNRAS, 323, 743
Robertson, B. E., & Kravtsov, A. V. 2008, ApJ, 680, 1083
Roe, P. L. 1986, Annual Review of Fluid Mechanics, 18, 337
Rosswog, S., Davies, M. B., Thielemann, F.-K., & Piran, T. 2000, Astronomy & Astrophysics, 360, 171
Saitoh, T. R., & Makino, J. 2009, ApJL, 697, L99
Scannapieco, C., Wadepuhl, M., Parry, O. H., Navarro, J. F., Jenkins, A., Springel, V., Teyssier, R., Carlson, E., Couchman, H. M. P., Crain, R. A., Dalla Vecchia, C., Frenk, C. S., Kobayashi, C., Monaco, P., Murante, G., Okamoto, T., Quinn, T., Schaye, J., Stinson, G. S., Theuns, T., Wadsley, J., White, S. D. M., & Woods, R. 2012, MNRAS, 423, 1726
Schmidt, M. 1959, ApJ, 129, 243
Semenov, V. A., Kravtsov, A. V., & Gnedin, N. Y. 2016, ApJ, 826, 200
Shen, S., Wadsley, J., & Stinson, G. 2010, MNRAS, 407, 1581
Smith, B. D., Bryan, G. L., Glover, S. C. O., Goldbaum, N. J., Turk, M. J., Regan, J., Wise, J. H., Schive, H.-Y., Abel, T., Emerick, A., O'Shea, B. W., Anninos, P., Hummels, C. B., & Khochfar, S. 2016, arXiv:1610.09591
Springel, V. 2005, MNRAS, 364, 1105
—. 2010, MNRAS, 401, 791
Springel, V., Di Matteo, T., & Hernquist, L. 2005, MNRAS, 361, 776
Springel, V., & Hernquist, L. 2002, MNRAS, 333, 649
Springel, V., Yoshida, N., & White, S. D. M. 2001, New Astronomy, 6, 79
Stinson, G., Seth, A., Katz, N., Wadsley, J., Governato, F., & Quinn, T. 2006, MNRAS, 373, 1074
Stinson, G. S., Brook, C., Macciò, A. V., Wadsley, J., Quinn, T. R., & Couchman, H. M. P. 2013, MNRAS, 428, 129
Stone, J. M., Gardiner, T. A., Teuben, P., Hawley, J. F., & Simon, J. B. 2008, ApJS, 178, 137
Tasker, E. J., Brunino, R., Mitchell, N. L., Michielsen, D., Hopton, S., Pearce, F. R., Bryan, G. L., & Theuns, T. 2008, MNRAS, 390, 1267
Teyssier, R. 2002, Astronomy & Astrophysics, 385, 337
Thompson, R., Nagamine, K., Jaacks, J., & Choi, J.-H. 2014, ApJ, 780, 145

Todoroki, K. 2014, Master's thesis, University of Nevada, Las Vegas
Toro, E. F., Spruce, M., & Speares, W. 1994, Shock Waves, 4, 25
Truelove, J. K., Klein, R. I., McKee, C. F., Holliman, II, J. H., Howell, L. H., & Greenough, J. A. 1997, ApJL, 489, L179
Truelove, J. K., Klein, R. I., McKee, C. F., Holliman, II, J. H., Howell, L. H., Greenough, J. A., & Woods, D. T. 1998, ApJ, 495, 821
Turk, M. J. 2013, arXiv:1301.7064
Turk, M. J., & Smith, B. D. 2011, arXiv:1112.4482
Turk, M. J., Smith, B. D., Oishi, J. S., Skory, S., Skillman, S. W., Abel, T., & Norman, M. L. 2011, ApJS, 192, 9
van Leer, B. 1979, Journal of Computational Physics, 32, 101
Wadsley, J. W., Stadel, J., & Quinn, T. 2004, New Astronomy, 9, 137
Yurin, D., & Springel, V. 2014, MNRAS, 444, 62