Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **509,** 1323–1341 (2022) https://doi.org/10.1093/mnras/stab3088
Advance Access publication 2021 October 25

From EMBER to FIRE: predicting high resolution baryon fields from dark matter simulations with deep learning

M. Bernardini [1]★ R. Feldmann [1] D. Anglés-Alcázar,[2,3] M. Boylan-Kolchin [4] J. Bullock [5] L. Mayer[1] and J. Stadel[1]

[1]*Center for Theoretical Astrophysics and Cosmology, Institute for Computational Science, University of Zurich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland*
[2]*Department of Physics, University of Connecticut, 196 Auditorium Road, U-3046, Storrs, CT 06269-3046, USA*
[3]*Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA*
[4]*Department of Astronomy, The University of Texas at Austin, 2515 Speedway, Stop C1400, Austin, TX 78712, USA*
[5]*Department of Physics and Astronomy, University of California, 4129 Reines Hall, Irvine, CA 92697, USA*

Accepted 2021 October 20. Received 2021 October 20; in original form 2021 July 20

ABSTRACT

Hydrodynamic simulations provide a powerful, but computationally expensive, approach to study the interplay of dark matter and baryons in cosmological structure formation. Here, we introduce the **EM**ulating **B**aryonic **E**n**R**ichment (EMBER) Deep Learning framework to predict baryon fields based on dark matter-only simulations thereby reducing computational cost. EMBER comprises two network architectures, U-Net and Wasserstein Generative Adversarial Networks (WGANs), to predict 2D gas and H I densities from dark matter fields. We design the conditional WGANs as stochastic emulators, such that multiple target fields can be sampled from the same dark matter input. For training we combine cosmological volume and zoom-in hydrodynamical simulations from the *Feedback in Realistic Environments* (FIRE) project to represent a large range of scales. Our fiducial WGAN model reproduces the gas and H I power spectra within 10 per cent accuracy down to ∼10 kpc scales. Furthermore, we investigate the capability of EMBER to predict high resolution baryon fields from low resolution dark matter inputs through upsampling techniques. As a practical application, we use this methodology to emulate high-resolution H I maps for a dark matter simulation of a $L = 100\,\mathrm{Mpc}\,h^{-1}$ comoving cosmological box. The gas content of dark matter haloes and the H I column density distributions predicted by EMBER agree well with results of large volume cosmological simulations and abundance matching models. Our method provides a computationally efficient, stochastic emulator for augmenting dark matter only simulations with physically consistent maps of baryon fields.

Key words: methods: numerical – methods: statistical – galaxies: haloes – dark matter – large-scale structure of Universe.

1 INTRODUCTION

The fundamental source of cosmological structure formation and its dynamics is the cosmic density field and its non-linear evolution. The details of the origin and evolution of this structure and its distribution over a variety of scales depends on the physics of the individual matter components – baryons and dark matter – and their mutual gravitational interaction. Overdense regions of dark matter, termed dark matter haloes, form the building blocks of large-scale structure as they define the landscape of potential wells in which baryonic matter flows to form individual groups and clusters of galaxies (e.g. Guo et al. 2010; Reddick et al. 2013; Somerville & Davé 2015; Wechsler & Tinker 2018; Feldmann, Faucher-Giguère & Kereš 2019). Understanding this coupled evolution in the linear regime is relatively straightforward, but becomes challenging when the evolution transitions into a highly non-linear regime on small scales (Weinberg 1972). Accurately modelling the non-linear interactions on those scales is challenging as they are often intractable for purely analytical approaches.

For this reason, various statistical and semi-analytical models have been proposed to quickly augment dark matter only simulations with information about baryonic components (Somerville & Primack 1999; Kravtsov et al. 2004; Behroozi, Wechsler & Conroy 2013). Halo Occupation Distribution (HOD) models and abundance matching approaches construct a mapping between the masses of dark matter haloes to the properties of the baryons residing inside them (Peacock & Smith 2000; Kravtsov et al. 2004; Schneider et al. 2019). The mapping is thus governed primarily by the halo mass, neglecting any environmental information such as e.g. the clustering of structure. Semi-analytical models (SAMs) describe the baryonic evolution in more detail (Croton, Gao & White 2007; Benson 2012). They use pre-calculated merger trees of dark matter haloes and model the key processes of galaxy formation by a set of coupled differential equations (e.g. Somerville & Primack 1999; Cole et al. 2000; Cora et al. 2018). They are able to take dynamical information into account and can model different gas phases and their interactions. However, SAMs generally do not follow the dynamical interaction of dark matter and baryons and require extensive parameter calibrations (see

★ E-mail: mauro.bernardini@uzh.ch

e.g. Knebe et al. 2015; Chuang et al. 2015, for a detailed comparison of different models).

To date, numerical hydrodynamical simulations offer the most principled approach to model and study in depth the intrinsic physical properties of systems comprised of dark matter and baryons (e.g. Bryan & Norman 1998; Springel & Hernquist 2003; Kereš et al. 2005; Springel et al. 2005; Vogelsberger et al. 2012; Hopkins et al. 2014; Feldmann & Mayer 2015; Wetzel & Nagai 2015; Feldmann et al. 2016; Wetzel et al. 2016; Feldmann et al. 2017; Pillepich et al. 2017; Hopkins et al. 2018; Davé et al. 2019). Simulations that consider only the physics of gravity are straightforward; it is baryonic physics, and its backreaction on the dark matter distribution that presents the most substantial challenge at present. The brute-force computation offers a better understanding of the dark matter and gas dynamics compared to SAMs (Hirschmann et al. 2011). However, their computational cost, being the main limiting factor, currently prohibits simulations of very large volumes with arbitrarily high resolution (Schaye et al. 2010, 2015; Vogelsberger et al. 2014; Khandai et al. 2015; Davé, Thompson & Hopkins 2016; Feng et al. 2016; Nelson et al. 2017, 2019). The trade-off between simulated box size and particle mass resolution is important, since it limits the range of scales a single simulation can cover (e.g. Katz & White 1993; Knebe & Domínguez 2003; Sirko 2005; Romeo et al. 2008). Cosmological zoom-in simulations try to mitigate this problem by preselecting a collapse region, which is then enriched in resolution (e.g. Bertschinger 2001; Naab, Johansson & Ostriker 2009; Feldmann, Carollo & Mayer 2011; Hahn & Abel 2011; Anglés-Alcázar et al. 2014; Hopkins et al. 2014; Onorbe et al. 2014). In this way, very high resolution simulations of individual haloes of different masses are possible, but the technique still suffers from large computational costs and data storage. The major aim of this work lies in exploring a methodology based on Deep Learning models to overcome this numerical trade-off by enriching cosmological simulations of dark matter with high resolution baryonic information at much reduced computational cost.

Feedback processes (e.g. stellar and AGN feedback) regulate star formation by expelling gas back into the surroundings of galaxies (Anglés-Alcázar et al. 2017a; Biernacki & Teyssier 2018; Hopkins et al. 2018; Li et al. 2018; Valentini et al. 2019). As a result, the complicated phase-space and temperature distribution of gas around galaxies inherently contains information about the feedback physics (Barnes et al. 2018; Chabanier et al. 2020). Thus, studying absorption signatures of neutral hydrogen and metal lines in the absorption spectra of background quasars is important to reveal the major physical processes that drive galaxy formation.

Galaxies accrete large quantities of fresh metal-poor gas from the intergalactic medium (IGM) to form new stars. This cosmological gas supply has a strong dependence on redshift and halo mass. Gas in massive haloes is shock heated and requires a long time to cool and settle into the galaxy disc whereas cold gas streams can reach the disc directly in less massive haloes (e.g. Kereš et al. 2005; Dekel & Birnboim 2006; Brooks et al. 2009; Faucher-Giguère, Kereš & Ma 2011; Woods et al. 2014; Ho, Martin & Turner 2019; Stern et al. 2020). This connection between galaxies and gas reservoirs within their parent halo is therefore an important aspect in galaxy formation models.

Modelling the evolution of galaxies requires to understand the evolution of the two main baryonic constituents, stars, and gas. Both simulations and observations have led to progress in understanding the evolution of stellar properties such as e.g. the star formation rate (SFR) and the main sequence over cosmic times (Karim et al. 2011; Guglielmo et al. 2015; Hwang, Shin & Song 2019; Feldmann 2020; Tacconi, Genzel & Sternberg 2020). The understanding of the dense molecular phase of the interstellar medium (ISM) has also improved through observational surveys of H_2 abundances with CO tracing techniques (e.g. Bolatto, Wolfire & Leroy 2013; Pavesi et al. 2018; Tacconi et al. 2018; Decarli et al. 2019).

In contrast, much less is known about atomic hydrogen (H I), especially at intermediate to high redshifts. Simulations predict that a significant fraction of the accreted gas in haloes is relatively cold and thus contains large amounts of atomic hydrogen (Kereš et al. 2005; Faucher-Giguère & Kereš 2011; Fumagalli et al. 2011, 2013; Nelson et al. 2013). The column densities of H I typically increase towards galaxy centres, which makes absorbers with high H I column densities better tracers of the gas in the near vicinity of galaxies (Bird et al. 2014; Rahmati et al. 2015; Crain et al. 2016; Diemer et al. 2019; Stern et al. 2021). Unfortunately, due to current observational constraints only galaxies residing in relatively massive haloes ($\geq 10^{12}$ $M_{\odot}$) can be observed, which requires large boxes or many individual zoom-in runs to simulate a statistically sound sample of such systems (Altay et al. 2011; Faucher-Giguère et al. 2015, 2016; Barnes et al. 2018).

The H I distribution in galaxies of the local Universe is measured through observations of emissions in the 21-cm line (e.g. Kirby et al. 2012; Reeves et al. 2015). This method is currently limited to nearby galaxies. After reionization ($z \sim 6$) the observation of neutral gas is currently only possible through absorption signatures in the spectra of bright background sources (e.g. quasars; Altay et al. 2011; Morganti & Oosterloo 2018; Glowacki et al. 2019; Weltman et al. 2020). Future 21-cm observations with significantly improved sensitivity such as e.g. the Square Kilometer-Array (e.g. Weltman et al. 2020) will be able to map the distribution of neutral hydrogen at high redshifts, thus enabling to study the formation processes of stars and galaxies in the young Universe (Mellema et al. 2013; Koopmans et al. 2015; Pritchard et al. 2015).

Machine learning and especially deep learning algorithms have recently become a promising tool to capture and learn high dimensional relations related to physical processes in cosmology (Ntampaka et al. 2019; Cohen et al. 2020; Villaescusa-Navarro et al. 2020a,b). They offer a valuable alternative between the computationally much cheaper but low resolution semi-analytical models and the much more expensive hydrodynamical simulations. A major advantage of machine learning methods is that predictions can be produced on time-scales typically much smaller than simulations, mitigating a major bottleneck.

Recently, a wide variety of deep learning algorithms have been deployed to accelerate the generation of cosmic matter fields. Applications range from generating dark matter density fields of different cosmologies (Feder, Berger & Stein 2020; Perraudin et al. 2020), to super-resolution maps (Kodi Ramanah et al. 2020; Li et al. 2020) to weak lensing convergence maps (Tamosiunas et al. 2020). Another interesting application is to learn the mapping between two matter components. A prominent example is to link the distribution of dark matter to specific baryonic fields like galaxies (Agarwal, Davé & Bassett 2018; Jo & Kim 2019; Zhang et al. 2019; Moster et al. 2020), neutrinos (Giusarma et al. 2019) as well as various gas fields (Tröster et al. 2019; Zamudio-Fernandez et al. 2019; Dai & Seljak 2020; Thiele et al. 2020; Wadekar et al. 2020; Harrington et al. 2021; Lovell et al. 2021; Prelogovic et al. 2021).

T he advancements in adversarial training of neural networks propose an interesting aspect for modelling physical systems that show stochastic variations. The key advantage that Generative Adversarial Networks (GANs) offer lies in the probabilistic nature of their predictions. GANs model the underlying distribution of the data by generating samples according to a high-level metric which is learned

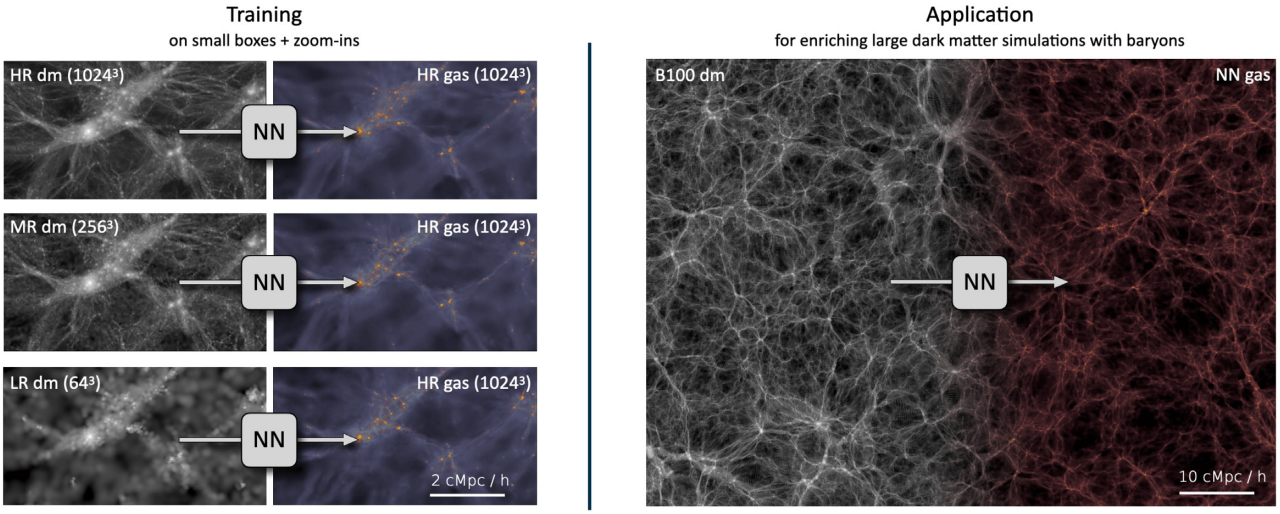

Figure 1. Illustration of our machine learning pipeline. We train neural networks on small cosmological volumes and zoom-in simulations with high resolution to predict baryonic counterparts from dark matter inputs. We investigate the upsampling capabilities of the networks by training individually on different dark matter input resolutions (indicated on the left in the training figure), while the target fields are always fixed to the highest resolution (see Section 4.5 for details). As indicated on the right, the trained neural networks can then be applied to large dark matter only simulations (e.g. the 100 h^{-1} Mpc box used in this work) to enrich them with the specified baryon fields at low computational cost.

Table 1. Summary overview of all simulations snapshots that are used to produce the network data. N denotes the total initial number of dark matter and gas particles in the simulation volume. $M_{\rm vir}^{\rm max}$ is the virial mass of the largest dark matter halo present in that simulation at $z = 2$. FIREbox^{PF}(hydro) and the zoom-in simulations are used for training and testing the algorithm. The dark matter-only FIREBOX runs are used during the resolution study and B100 is solely used for applying our method to a large-scale dark matter only simulation.

Simulation	Note	L (h^{-1} cMpc)	N	$m_{\rm b}$ (h^{-1} M$_\odot$)	$m_{\rm dm}$ (h^{-1} M$_\odot$)	$M_{\rm vir}^{\rm max}$ (h^{-1} M$_\odot$)	h	$\Omega_{\rm m}$	$\Omega_{\rm b}$	Ω_Λ	σ_8
FIREBOX^{PF}	Hydro	15	2×1024^3	4.23×10^4	2.27×10^5	2.87×10^{12}	0.6774	0.3089	0.0486	0.6911	0.8159
FIREBOX	Dm-only	15	1024^3	0	2.69×10^5	4.53×10^{12}	0.6774	0.3089	0	0.6911	0.8159
FIREBOX	Dm-only	15	256^3	0	1.72×10^7	3.03×10^{12}	0.6774	0.3089	0	0.6911	0.8159
A1	Hydro, zoom	100	7.50×10^7	2.32×10^4	1.19×10^5	1.98×10^{12}	0.6970	0.2821	0.0461	0.7179	0.817
A2	Hydro, zoom	100	2.33×10^8	2.23×10^4	1.19×10^5	2.56×10^{12}	0.6970	0.2821	0.0461	0.7179	0.817
A4	Hydro, zoom	100	1.34×10^8	2.32×10^4	1.19×10^5	2.13×10^{12}	0.6970	0.2821	0.0461	0.7179	0.817
A8	Hydro, zoom	100	2.92×10^8	2.32×10^4	1.19×10^5	2.56×10^{12}	0.6970	0.2821	0.0461	0.7179	0.817
B100	Dm-only	100	1024^3	0	7.9×10^7	4.83×10^{13}	0.6774	0.3089	0	0.6911	0.8159

in the training process. This is a key advantage compared to neural networks that are trained on low-level metrics (e.g. mean-squared error), since the high-level metric is encoded by an entire network itself. When trained correctly, GANs act as stochastic emulators generating samples that are statistically consistent with the data-set.

In this work we explore this methodology by combining small high resolution cosmological boxes with zoom-in simulations. We show that adversarial learning offers a promising pathway for modelling fully non-linear relations between cosmological matter fields. In particular, we use recent advancements in adversarial training of neural networks to predict high resolution baryonic fields from dark matter inputs. We also investigate the upsampling capabilities of the networks by training on different dark matter input resolutions. The trained neural networks can then be used to predict the baryonic counterparts of large dark matter only simulations, which constitutes the primary advantage of this methodology (we show a summary overview in Fig. 1).

This paper is structured as followed. In Section 2, we describe the simulations used in this work and explain the pipeline to produce the data samples in Section 3. In Section 4, we briefly revisit some key aspects in generative learning as well as recent research developments. We introduce the theoretical aspects of the network type used in this work and discuss its key advantages. We also formulate the mappings that the neural networks learn and give a detailed overview of the architectures and further details regarding the training of the networks. The results are presented and discussed in Section 5. Finally, we propose future applications of our method and conclude in Section 6.

2 SIMULATIONS

The simulations used in this work are part of the Feedback in Realistic Environments (FIRE[1]) project (Hopkins et al. 2014, 2018). In the following, we give a brief overview of the simulation details and show a summary of the most important parameters in Table 1.

We use the $z = 2$ snapshots of the FIREbox^{PF} and MASSIVEFIRE simulations run with the FIRE-2 physics model (Hopkins et al. 2018) for creating our data sample. Our simulations are run using GIZMO (Hopkins 2015),[2] a multimethod gravity plus hydrodynamics

[1]See the official FIRE project website: https://fire.northwestern.edu
[2]A public version of GIZMO is available at http://www.tapir.caltech.edu/~phopkins/Site/GIZMO.html

code. Initial conditions were generated using the multiscale initial condition tool MUSIC (Hahn & Abel 2011) where the random seed is fixed. The simulations are run with Planck 2015 cosmology (Ade et al. 2016) where $H_0 = 67.74\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\mathrm{M} = 0.3089$, $\Omega_\Lambda = 0.6911$, $\Omega_\mathrm{b} = 0.0486$, $\sigma_8 = 0.8159$, and $n_s = 0.9667$. Here, we summarize only the most important aspects of the simulations and refer the interested reader to corresponding work for further details.

FIREbox^{PF} is a high resolution hydrodynamical cosmological simulation with box length of $15\,h^{-1}$ cMpc (Feldmann et al., in preparation). The box contains initially 1024^3 dark matter and 1024^3 gas particles. Dark matter and baryon masses are $m_\mathrm{dm} = 2.27 \times 10^5$ and $m_\mathrm{gas} = 4.23 \times 10^4\,h^{-1}\mathrm{M}_\odot$. The softening lengths for dark matter particles is 80 pc and gas particles have a minimum softening length of 1.5 pc.

To augment the number of dark matter haloes in the high-mass end, we add four zoom-in simulations (A1, A2, A4, and A8) selected from the original MASSIVEFIRE (Feldmann et al. 2016, 2017) suite and re-simulated with FIRE-2 physics and massive black holes (Anglés-Alcázar et al. 2017b), which have a similar resolution as FIREbox^{PF}. The particle masses are $m_\mathrm{dm} = 1.19 \times 10^5$ and $m_\mathrm{gas} = 2.23 \times 10^4\,h^{-1}\mathrm{M}_\odot$. The gravitational softening lengths of dark matter and gas particles are 143 and 9 pc.

For this work we make use of the total gas and neutral hydrogen abundances from the simulations. Briefly, gas cooling follows an implicit algorithm described in Hopkins et al. (2018) that includes various processes (free–free, photo-ionization/recombination, Compton, photoelectric, metal-line, molecular, fine-structure, dust collisional, and cosmic ray physics). The relevant metal ionization states are tabulated from CLOUDY simulations (Ferland et al. 1998) where the process of self-shielding is accounted for via a local Sobolev/Jeans-length approximation calibrated from radiative transfer experiments (Faucher-Giguère et al. 2010; Rahmati et al. 2013). We refer to Hopkins et al. (2018) for a full description of the simulated gas physics.

We furthermore use a pure dark matter-only simulation (B100) with boxsize of $L = (100\,h^{-1}\,\mathrm{cMpc})^3$ and dark matter mass resolution of $m_\mathrm{dm} = 7.9 \times 10^7\,h^{-1}\mathrm{M}_\odot$ to demonstrate the general applicability of our proposed machine learning approach. We provide a summary overview of all simulations in Table 1.

3 DATA GENERATION

The task of the neural networks is to predict 2D gas and H I mass maps from 2D dark matter inputs. We use a combination of the `smooth` and `tipgrid` algorithms for the pixelization of the 2d input and target fields.[3] For every particle, `smooth` computes a smoothing length, which is defined as half of the distance to the nth neighbour particle. We find that setting $n = 80$ works well for our approach. Next, we divide the simulation region into 10 equally spaced slabs for each of the 3 spatial directions. `tipgrid` then interpolates the particles in the same slab on to a 2D grid by depositing the particles mass with a spherically symmetric kernel according to the smoothing lengths computed beforehand. The grid resolution we chose is 4096^2, such that one pixel resolves roughly $3.6\,h^{-1}$ckpc. We found this combination to be the best for our application. For training the networks we then dynamically create samples with dimensions of 512^2. Although a higher grid resolution would help in resolving even more small-scale information, the input dimensions of our neural networks would increase drastically. The resulting set of 30 maps represents the mass of the deposited matter fields over slabs with depth $1.5\,h^{-1}$ Mpc. Note that the largest haloes in our simulations all have $R_\mathrm{vir} \sim 350\,h^{-1}$ckpc and therefore the probability of splitting a halo between two slabs is negligible for our approach.

A similar procedure is applied to the zoom-in simulations, but here we crop the innermost region containing 75 per cent of all high resolution particles. Note that for the zoom-in simulations we manually fix the halo centre to be at the centre of the slabs, and then generate a total of 10 projections according to varying angles. This guarantees that we only extract the density field within the high resolution region.

We use the maps from the x and y direction of FIREbox^{PF} (hydro) as our training set and retain the projections along the z-direction for testing the algorithm. Similarly, we augment the training set with 2/3 of the projections from the zoom-in simulations. The network training uses tiles from the zoom-in simulations and FIREbox^{PF} in the exact same way. The maps from the FIREbox 1024^3 dm-only simulation simply act as an additional testing set for reasons described in the following paragraph.

For our application it is important to notice that the dark matter field in the dark matter only simulations differs from the one in the corresponding hydrodynamical run. We will refer to these fields as dark matter only (`dmo`) and dark matter hydro (`dmh`) from now on. The `dmo` and `dmh` fields differ on scales where baryonic processes affect the dark matter distribution, such as e.g. adiabatic contraction resulting in deeper halo potential wells due to the cooling of gas in the halo centers (Blumenthal et al. 1986; Jesseit, Naab & Burkert 2002; Gnedin et al. 2004). Feedback processes can change the underlying dark matter structure as well (Navarro, Eke & Frenk 1996; Governato et al. 2012; Chan et al. 2015; Oñorbe et al. 2015; Lazar et al. 2020). Another important effect that is induced by the presence of baryons is that the `dmh` halo morphology tends to be more spherically symmetric compared to the `dmo` haloes (see e.g. Tissera & Domínguez-Tenreiro 1998; Bett et al. 2010; Kazantzidis, Abadi & Navarro 2010; Butsky et al. 2016; Chua et al. 2019; Cataldi et al. 2020). However, for our application and training algorithm the largest effect is that the exact location of the dark matter haloes is different in the two dark matter fields. Since the gas concentration is typically highest in the halo centres, the `dmo` dark matter haloes are offset compared to the gas peaks in the target fields. This factor limits the usefulness of the `dmo` as the input field in the training process, since the neural networks were not designed to learn this shear effect. We therefore use the `dmh` version for the traditional training and testing of the networks and retain the `dmo` version for pure external testing purposes. We then compare the ability of the trained networks to produce the exact power spectrum of the target fields when predicting from either the `dmo` or the `dmh` version, despite being only trained on `dmh` data. We artificially downsample the 1024^3 simulation to 256^3, 64^3, and 16^3 by randomly selecting particles and adjusting their masses accordingly. The deposition of the downsampled simulations on to the grid is then exactly the same as in the high resolution case (same pixel resolution of 4096^2). We use the suffix `ds` to indicate downsampled simulations.

The downsampled dark matter maps represent simulations that are in principle different than a simulation of that native resolution, because some high-level modes might survive the downsampling process. To understand the impact of training from downsampled dark matter inputs and then applying the networks to dark matter simulations of that native resolution, we also compare all summary statistics to the FIREbox^{PF} runs with native resolution 256^3, respectively.

[3] https://github.com/N-BodyShop/smooth

For H I we also compute the column density distribution function (CDDF) $f(N_{\rm H\,I})$ for the true and predicted maps. The CDDF is a pixel based quantity often used in observational studies of H I, which is defined such that $f(N_{\rm H\,I})\,{\rm d}N_{\rm H\,I}\,{\rm d}X$ is the number of absorbers per unit column density bin and unit absorption length ${\rm d}X$. Following Rahmati et al. (2013) we write the CDDF as

$$f(N_{\rm H_I}, z) \equiv \frac{{\rm d}^2 n}{{\rm d}N_{\rm H_I}\,{\rm d}X}, \tag{1}$$

where ${\rm d}X(z)$ is the absorption distance which is related to the box size ${\rm d}L$ as ${\rm d}X = (H_0/c)(1+z)^2\,{\rm d}L$ (see Appendix A for details).

4 EMBER

EMBER (**EM**ulating **B**aryonic **En****R**ichment) is a framework of several neural networks that we train to map dark matter to baryons. The task of the neural networks is to predict 2D gas counterparts from 2D dark matter inputs. The pixel resolution of the maps in our data-set is $\sim$3.6 ckpc h^{-1}, a length-scale where astrophysical processes have a large impact on the exact gas configurations.

To understand the importance of modeling such small cosmological scales, we investigate two different methodologies. First, a purely deterministic approach (U-Net), and secondly, a probabilistic approach (WGAN) that is able to capture the small-scale variations in our data-set. Our trained models can then be used as emulators to enrich dark matter simulations with their corresponding gas fields.

In the following sections, we describe the theoretical background as well as the implementation of these two methodologies that comprise the EMBER framework. We first introduce both algorithms on a theoretical level, and then discuss implementation and training aspects in more detail. Note that throughout the following sections we refer to the dark matter input maps as x and the target gas fields as y with their corresponding underlying distributions p_x and p_y.

4.1 U-Net

For the implementation of the neural networks we make use of the U-Net architecture, which was first introduced by Ronneberger, Fischer & Brox (2015) to solve bio-medical image segmentation tasks but has been shown to perform well in regression scenarios as well (e.g. Thiele et al. 2020; Wadekar et al. 2020). Our U-Net architecture is a fully convolutional auto-encoder consisting of two main branches, an encoding and a decoding part. As in the general auto-encoder case, the information extraction and compression is realized by convolutional blocks followed by information pooling. We use strided convolutions in our implementation. In order to recover the original input dimensions, the decoding units consist of an initial upsampling layer followed by consecutive convolution operations, which results in higher resolution feature maps. In this process the network extracts advanced features, but loses the information of where those features are located in the image. To this end, Ronneberger et al. (2015) introduced skip connections that copy and concatenate the information from the corresponding encoder level with the up-sampled data from the decoder part. In this manner, the spatial information from the contraction path is directly transferred to the expanding branch without being passed through the bottleneck. The skip connections are then concatenated after the upsampling operation with tensors holding information that emerges from deeper parts of the network.

4.2 Generative adversarial networks

Generative adversarial networks (GANs) are a framework first introduced by Goodfellow et al. (2014) where two networks, a generator G and a discriminator D, compete in an adversarial game. G is trained to generate data with a distribution close to the true data distribution p_y, where the in- and output are vectors with arbitrary dimensions. D is optimized to distinguish real from fake samples by mapping vectors from the true and generated data domains to [0, 1]. A value of 1 implies that D marks a given sample as real. The training objective of G is to maximize the misclassification of D. This setup corresponds to a min–max algorithm where both neural networks try to outperform their corresponding opponent. To learn an approximation to the true data distribution, one defines an input noise variable $\eta \sim p_\eta$. The generator G learns p_g, an approximation to p_y, through encoding of the latent variable η. A successfully trained generator can then be used to produce new predictions by sampling the latent space by varying η.

Conditional GANs (cGANs) incorporate additional information $x \sim p_x$ (Mirza & Osindero 2014). In practice the conditional information x is an additional feature vector which is used for training both the generator and critic. The min–max game is then expressed in terms of objective functions $\mathcal{L}_D$ for the discriminator and $\mathcal{L}_G$ for the generator (Goodfellow et al. 2014) as

$$\mathcal{L}_D^{\rm adv} = +\mathbb{E}_\eta\left[D(G(\eta|x)|x))\right] - \mathbb{E}_y\left[D(y|x)\right], \tag{2}$$

$$\mathcal{L}_G^{\rm adv} = -\mathbb{E}_\eta\left[D(G(\eta|x)|x))\right], \tag{3}$$

where y and x are samples from the true data distributions p_y and p_x. For a successfully trained cGAN the distributions p_g and p_y will be very similar ($p_g \simeq p_y$). Hence, one obtains a model that is capable of sampling from the joint distribution (Mirza & Osindero 2014; Feder et al. 2020)

$$p(y, x) = p_x(x)p_y(y|x). \tag{4}$$

A major challenge while training GANs is the situation when p_g and p_y have little overlap or are disjoint. Prior work (see e.g. Salimans et al. 2016; Arjovsky & Bottou 2017) pointed out that this scenario invokes an inherent instability of GAN training. If the overlap of the two distributions is small, training a discriminator that perfectly separates real and fake samples is relatively easy compared to the task of the generator which exposes a major dilemma. On one hand, the discriminator must perform well enough for the generator to have accurate feedback. On the other hand, the gradients for the generator vanish in case of a perfect discriminator. Various improvements to the training of GAN models have been presented to artificially increase the amount of overlap between p_g and p_y (see e.g. Jenni & Favaro 2019), such as Instance Noise (Sønderby et al. 2016) and one-sided label smoothing and flipping (Salimans et al. 2016).

4.3 Wasserstein GANs

Arjovsky, Chintala & Bottou (2017) introduced WGANs to mitigate the aforementioned problems by using the Wasserstein metric as a measure of similarity between p_g and p_y. In this framework, the critic D learns an approximation of the Wasserstein distance. This approach requires D to be Lipschitz continuous over the input domain (see e.g. Arjovsky et al. 2017). Various approaches have been proposed to satisfy this constraint such as weight clipping (Arjovsky et al. 2017) or total variational regularization (Zhang, Zhang & Gao 2018). In this work, we adopt the gradient penalty scheme introduced by Gulrajani et al. (2017), where D is constrained as $|\nabla D| = 1$, to fulfill Lipschitz

continuity. The total critic objective then becomes

$$\mathcal{L}_D = \mathcal{L}_D^{\rm adv} + \gamma \cdot \mathbb{E}_{\boldsymbol{y}} \left[(||\nabla_{\boldsymbol{y}} D(\boldsymbol{y}|x)||_2 - 1)^2 \right], \quad (5)$$

where $\boldsymbol{y}$ are samples drawn from the distribution $p_{\boldsymbol{y}}$, which smoothly interpolates p_g and p_y, and γ is a hyperparameter. We refer the interested reader to Gulrajani et al. (2017) for a detailed description of the gradient-penalty scheme.

Arjovsky et al. (2017) showed this setup does not suffer from vanishing gradients for the generator in the case of a well-performing critic network. Moreover, in case of optimal training, the critic network is fully converged in every iteration step and propagates the most meaningful gradients back to the generator (Arjovsky & Bottou 2017; Arjovsky et al. 2017).

4.4 On the difficulty of synthesizing high resolution images

The successful training of networks that are able to synthesize high resolution images of large dimensions is a notoriously difficult task. The major bottleneck lies in the training instability of GANs due to the passage of uninformative gradients from the discriminator to the generator as described above. Karras et al. (2017) proposed a new technique by progressively adding higher resolution layers throughout the training process. Gradually adding higher resolution information is a viable technique to mitigate the missing overlap problem as the network first learns to match the distribution in lower dimensions and slowly migrates towards the full distribution by incorporating higher dimensional information. In this algorithm, whenever a new layer is added it is gradually faded in such that the training progress of the pre-trained part is retained. Even though this technique has been shown to produce state-of-the-art performance, the entire training process remains difficult due to various selections of hyperparameters.

Karnewar & Wang (2019) proposed a new method termed MSG-GAN that is based on the idea of gradually matching distributions on multiple resolution scales. Opposed to any progressive adding of layers, in MSG-GAN the entire network composed of generator and critic is trained simultaneously on all resolution levels. The key ingredient lies in multiscale connections between the generator and critic layers of the same resolution, which allow the gradients to be backpropagated into the various levels directly. Karnewar & Wang (2019) find that networks including these connections are less sensitive to the choice of hyperparameters or loss functions. The method is robust to different network architectures and drastically helps in the training stability over a wide variety of data sets. This property is especially useful for this work, since we train multiple WGANs with fixed hyperparameters across different data-sets.

4.5 Formulating the network mappings

We train a collection of neural networks to predict 2D total gas and H I mass maps from dark matter inputs derived from four different resolution levels: 1024^3 (HR), 256^3 (MR), 64^3 (LR), and 16^3 (VLR). The 2D target fields, however, are always derived from the 1024^3 (HR) simulation. The data are generated from the FIREbox^{PF} and 4 MASSIVEFIRE zoom-in simulations at fixed redshift $z = 2$. The dark matter input x and the target gas field samples y implicitly define the joint probability distribution $p(y, x)$ in equation (4). The marginal distribution $p(x)$ is the probability of a certain dark matter configuration that is computed in the simulation. In our application, the WGANs learn the distribution $p_g(y|x)$, which is modelled implicitly as it is not constrained with any prior information. Since a single simulation run only probes a sub-collection of all possible samples x, the learned distribution by the WGANs $p_g(y|x)$ is an approximated version of the true underlying distribution $p(y|x)$.

We train a total of 10 separate neural networks, 8 WGANs and 2 U-Nets, to investigate key questions regarding the mappings depending on the resolution level of the dark matter input. We give a detailed overview of the network architectures in Section 4.6.

(i) *HR* → *HR*: In the case of HR input, we train 4 networks, two WGANs and two U-Nets (see Section 4.6 for a detailed comparison). Each pair either predicts the target of total gas or H I. We conduct this comparison between U-Net and WGAN to understand the impact of the generative part when being trained on data-sets that exhibit small-scale features emerging from fully non-linear baryonic effects in the simulation.

(ii) *MR* → *HR*: For the MR input we train two WGANs, one for total gas and one for H I. This application is different, because a substantial amount of dark matter information is missing in the input fields. The WGAN needs to fill in the information about the missing small-scale modes in a physically consistent way. To demonstrate the use-case of this method we apply it to a large dark matter only simulation of boxsize $100\,h^{-1}$ Mpc (B100) with the same cosmological parameters and mass resolution as the training set.

(iii) *LR* / *VLR* → *HR*: For the LR/VLR case we repeat the same exercise as in the MR case to understand to what level the WGAN is capable of generating accurate target fields when being presented with dark matter information containing only the very largest modes. The conditional information from x is minimal in this application and the learning of the distributions is mainly driven through the direct feedback from the critic network.

4.6 Network Implementation

We implement the U-Net version in the `tensorflow` neural network API (Abadi et al. 2015).[4] A detailed schematic of the network is shown in Fig. 2. The individual convolution blocks are constructed by two convolutions with filter size (3×3) and stride 1. The first network layer has $n_{\rm f} = 8$ convolution filters that are activated by a LeakyReLU. After the first convolution block the data are copied and split along two different paths. The first path is a skip connection that concatenates to the up-flowing data on the reascending network part. Along the second path the data is downsampled by another convolution layer with $n_{\rm f}$ filters, kernel size (3×3) and stride 2 before it flows into the next block. All convolution layers are initialized with the glorot-normal initialization scheme (Glorot & Bengio 2010).

As the data descend the network reaching deeper levels the number of filters increases by a factor of 2 for each subsequent block. Overall there are 8 convolution blocks. Within deeper layers the network becomes more sensitive to large-scale structures in the input image since the strided convolutions reduce the dimensions by a factor of 2 for each additional layer. Since one cell represents a physical size of $\sim$3.6 ckpc h^{-1}, the downsampling operations increase the receptive field by a total factor of $2^7 = 128$. In the upsampling branch all features scales are used together with the spatial information provided by the skip connections. We use (1×1) convolutions at the end of each upsampling block to obtain the intermediate outputs, which are activated by a native ReLu function as the target fields only contain values greater than 0.

Our fiducial WGAN model, is constructed by augmenting the U-Net architecture with techniques from (Karras et al. 2017; Karras,

[4]Full code can be found at the official github repository.

[h]

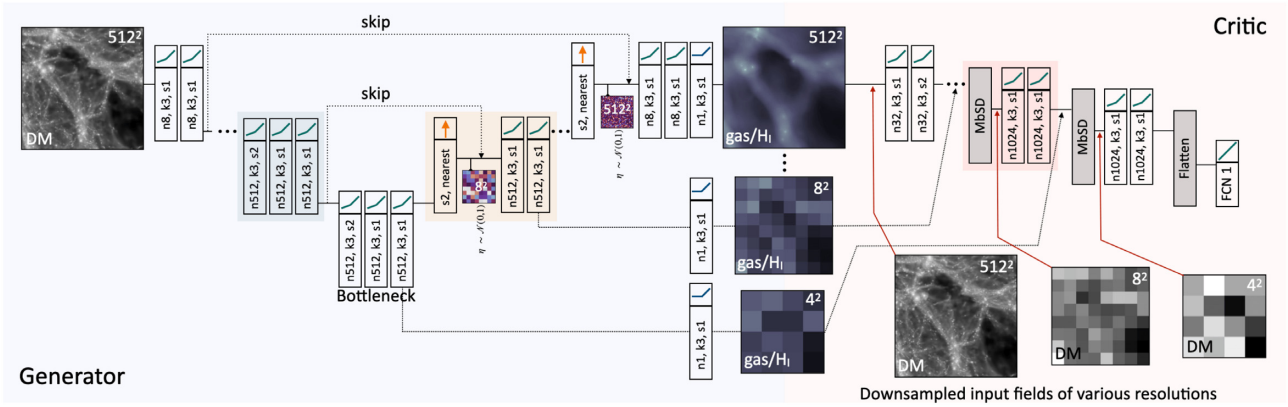

Figure 2. The architecture of the WGAN and U-Net. The generator network shown on the left in light blue is the U-Net with multiscale outputs whereas the critic is shown on the right in light red. n, k, and s denote number of filters, kernel size, and stride for each convolution filter in the network, which are all using same padding and are generally activated using LeakyReLUs except for the multiresolution connections (ReLU) and the final dense layer of the critic (linear). The green, orange, and red shaded blocks of operations are examples of a generator convolution, upsampling, and critic convolution block. The noise injection follows the upsampling operation in the decoding part of the WGAN. The artificially downsampled input fields are concatenated in the positions in the critic network as described in Section 4.6. The pure U-Net also studied in this work corresponds to the same generator architecture shown above but without any noise inclusion layers and no critic network. The exact same network architecture is used for all trainings and tests of EMBER.

Laine & Aila 2018; Karnewar & Wang 2019). We mainly follow Karnewar & Wang (2019) in our implementation and complement the U-Net with our own critic model which we design to be fully convolutional except for the final dense neuron. Similar to the descending part of the U-Net, the critic is constructed with convolution blocks. A single block contains two convolution layers with filter size (3×3) where the second filter has stride 2 for downsampling and for activations we use LeakyReLU. Following Karnewar & Wang (2019), we let the critic network extract as much information as possible from all information available at a single iteration step. In order to feed the intermediate scale information from the generator to the critic, we convolve the tensors at the corresponding level with a single (1×1) filter (Karnewar & Wang 2019). These connections are a vital part as they allow the gradients to penetrate the generator layers on every resolution scale simultaneously. This results in the behaviour observed by Karnewar & Wang (2019) where the deepest parts of the generator are optimized first and the synchronization of higher resolution levels follows in a bottom-to-top fashion. The inputs in the first critic layer are the original input field (dark matter) field as well as the target field (total gas or H I). The idea is that the critic network should learn whether or not a certain combination of input and target field is realistic or not on all resolution scales.

The inputs of the lower resolution blocks consist of three fields in total: the information coming from the higher critic block, a downsampled version of the dark matter and the target field itself. We use Minibatch Standard Deviation (MbSD; Karras et al. 2017) on the concatenated tensors coming from the higher level and the downsampled target, but concatenate the downscaled dark matter only after this layer. We empirically found that concatenating the dark matter tensor after the standard deviation layer generally results in better performance, presumably because the dark matter information smooths the difference between the outputs of the MbSD layer for a batch of fake and real samples.

The architecture comprised of a U-Net and a critic model defines a prediction pipeline that is deterministic in nature. To promote the generator to a stochastic model, we include an additional noise layer in each upsampling block of the generator to build our fiducial WGAN models. This additional noise then allows the generator to learn a distribution conditioned on the input field and the network is allowed to block any additional noise input, if it does not improve the realism of the generated images. To this end, we follow a similar strategy as Karras et al. (2018). Instead of concatenating the upflowing tensors with the skip connections directly, the data coming from lower network parts are multiplied with Gaussian noise $\eta \sim \mathcal{N}(0, 1)$ which is controlled by a trainable parameter ω. This parameter can be different for each level as the successful inclusion of noise is in principle resolution dependent. The feature maps f are then modified according to

$$f' = f + \omega \odot \eta. \tag{6}$$

We find that the WGANs make use of the additional noise inputs to produce small-scale details in the generated fields as discussed in Section 5.

4.7 Training details

The training of neural networks is simplified when the numerical values of the data are of $\mathcal{O}(1)$. Given that the physical value range of our input and target fields is over nine orders of magnitude, the choice of the data normalization scheme plays a crucial role for our task. We tried many different normalization schemes and found that depending on the cumulative distribution function of the target field the following mixture of a power-law and a log-transform works best

$$\tilde{x} = \frac{1}{k} \log \left[\left(\frac{x}{x_0} \right)^q + 1 \right], \tag{7}$$

where x is the field to transform. The values of the free parameters in this scheme are manually tuned, such that the high end tail of pixel values is retained while keeping intermediate and small pixel values in a reasonable interval as well. The exact parameters are given in Table 2. We found that the appropriate normalization scheme plays a crucial role for achieving accurate predictions for the power-spectrum and bispectrum.

Since the evaluation metrics of the neural network are primarily driven by the high end tail of the pixel distribution, choosing the right loss function for optimizing the networks is a crucial ingredient. In

Table 2. Summary overview of the manually tuned hyperparameters in the normalization schemes. The corresponding parameters are the same for normalizing the same field across different resolutions.

Field	k	$x_0/\mathrm{M}_\odot$	q
dmh	3.0	10^7	1.0
dmo	3.0	10^7	1.0
Gas	3.0	1.5×10^6	1.0
H I	0.9	10^4	0.2

Table 3. Summary table of the chosen hyperparameters used in the total loss function. Since we keep the network parameters fixed across all training runs, (α, β, γ) represent the important hyperparameters in our approach. For the U-Nets we normalize the losses to the DSSIM (α). The β parameters vary as well since the PDF of the target fields is different. In the WGAN case, the losses are normalized to the adversarial loss.

	α	β	γ
U-Net (gas)	1	2	–
U-Net ($\mathrm{H_I}$)	1	200	–
WGAN (gas)	10	0.4	10
WGAN ($\mathrm{H_I}$ I)	5	0.02	50

most regression applications, a pure pixel-based loss is enough. We find that adding a loss term measuring the structure similarity of two pictures helps in the training process. We use DSSIM (Wang et al. 2004), a metric that alleviates the pixel-by-pixel comparison as it is evaluated upon sub-windows over the image by comparing various moments (mean, standard deviation, and covariance) connected to the image morphology. For maximizing the performance on the high pixel values of density peaks, we combine the DSSIM with a pure pixel-based loss in the form of a mean-square-log error (MSLE). The total perceptual loss on every resolution scale is then aggregated to account for the multiscale (MS) nature of the mapping

$$\mathcal{L}_\mathrm{p} = \alpha \sum_\mathrm{MS} \mathrm{DSSIM}(t, p) + \beta \sum_\mathrm{MS} \mathrm{MSLE}(t, p), \tag{8}$$

where t and p denote true and predicted maps. Since we only have the generator in the U-Net case, the loss function only contains the perceptual loss. In the WGAN case the perceptual loss is simply added to the adversarial loss for the generator, i.e.

$$\mathcal{L}_\text{U-Net} = \mathcal{L}_\mathrm{p} \quad \text{and} \quad \mathcal{L}_G = \mathcal{L}_\mathrm{p} + \gamma \mathcal{L}_G^\mathrm{adv}. \tag{9}$$

The prefactors (α, β, γ) account for weighting the different loss contributions (see Table 3). We note that even though the adversarial loss could in principle take care of any pixel-based loss by simply encoding it, we find that keeping $\mathcal{L}_\mathrm{p}$ helps with convergence and stability especially in the beginning parts of training. Once the perceptual loss saturates, the gradient penalty and adversarial losses become the dominant measure of the WGAN performance.

During training time we dynamically create batches of 8 images with size 512^2 from the training set. The optimizer we use is Adam (Kingma & Ba 2017) with parameters $\beta_1 = 0$ and $\beta_2 = 0.99$ where the learning rates are reduced in a polynomial fashion. The networks were trained for 3×10^5 iterations, which means that every generator network has seen approximately 2.5 million data samples. The training was performed on a single Tesla V-100 GPU card and took 9 h in the case of the U-Nets and 96 h for the WGANs.

During prediction time, we only make use of the trained generator to produce new samples. Since the generator architecture is fully convolutional, the network can make predictions on inputs with arbitrary input sizes. Predicting all projections (of size 4096^2) of a single simulation box takes approximately 60 s on the same V-100 GPU card.

5 COMPARISON AND DISCUSSION

In this section, we give a detailed overview of various statistics that we use to measure the network capabilities. To evaluate the predictive power of the networks we compare the following summary statistics: total mass, pixel probability density, 2D power spectrum, and bispectrum between the true and predicted mass maps. Furthermore, we conduct a halo-based analysis by comparing true and predicted gas masses inside dark matter haloes. In the H I case we also compare the column density distribution function (CDDF) $f(N_\mathrm{H\,I})$ between the true and predicted maps.

We observe that in the final stages of training, the predicted power spectrum can fluctuate especially on scales $\sim$10 ckpc h^{-1}. We use the power spectrum of the prediction on the training set as the primary metric to determine the best networks. We then identify for each U-Net and WGAN the point to stop training, and take this exact checkpoint for further analysis.

For each of the different input resolutions, we use the WGANs to predict a total of 128 test boxes (from `dmo ds`, from `dmh ds` and in the 256 case also from the `dmo native`) to quantify the amount of internal scatter in the summary statistics. Note that each realization corresponds to a different latent vector η as described in Section 4.2. The U-Nets predict only one box per input as the mapping is deterministic. The statistics of those predicted maps are always compared with the high resolution gas fields (*HR*, 1024^3), since we want to quantify to what extent we can use lower resolution *N*-body simulations to achieve similar accuracy compared to the high resolution results.

5.1 WGAN versus U-Net

Figs 3 and 4 are a visual summary showcasing the network capabilities of generating realistic samples across different scales. The WGANs have learned to encode the additional noise inputs resulting in very diverse small-scale structures. The last column shows the U-Net predictions, which resemble a smoothed out version of the WGANs. The absence of a critic network paired with no noise inclusion results in very smooth gas maps that lack the high resolution small-scale features. From a conceptual point of view this behaviour is understandable because the U-Net performs a regression of the mean of all possible gas configurations.

In Table 4, we show the fractional deviation of the total mass in the box between predicted and true gas maps for the eight networks. Generally, the two mappings from either `dmo` and `dmh` result in very similar deviations. For the total gas mass, the U-Net and WGANs perform equally well, but the WGANs outperform the U-Nets across all resolutions in the case of the more difficult H I mapping. In this case we found that the U-Nets perform worse and generally overpredict the mass in the box. In Figs 5 and 6 we show the predicted PDF, CDDF, power spectra, and bispectra for the *HR* case (top row). As expected, the U-Nets fail to reproduce the correct power on small scales whereas the WGAN models show very good agreement for all statistics.

5.2 WGAN for extreme upsampling

In the section, before we described the performances of the network when trained on HR dark matter input. In this section, we explore the

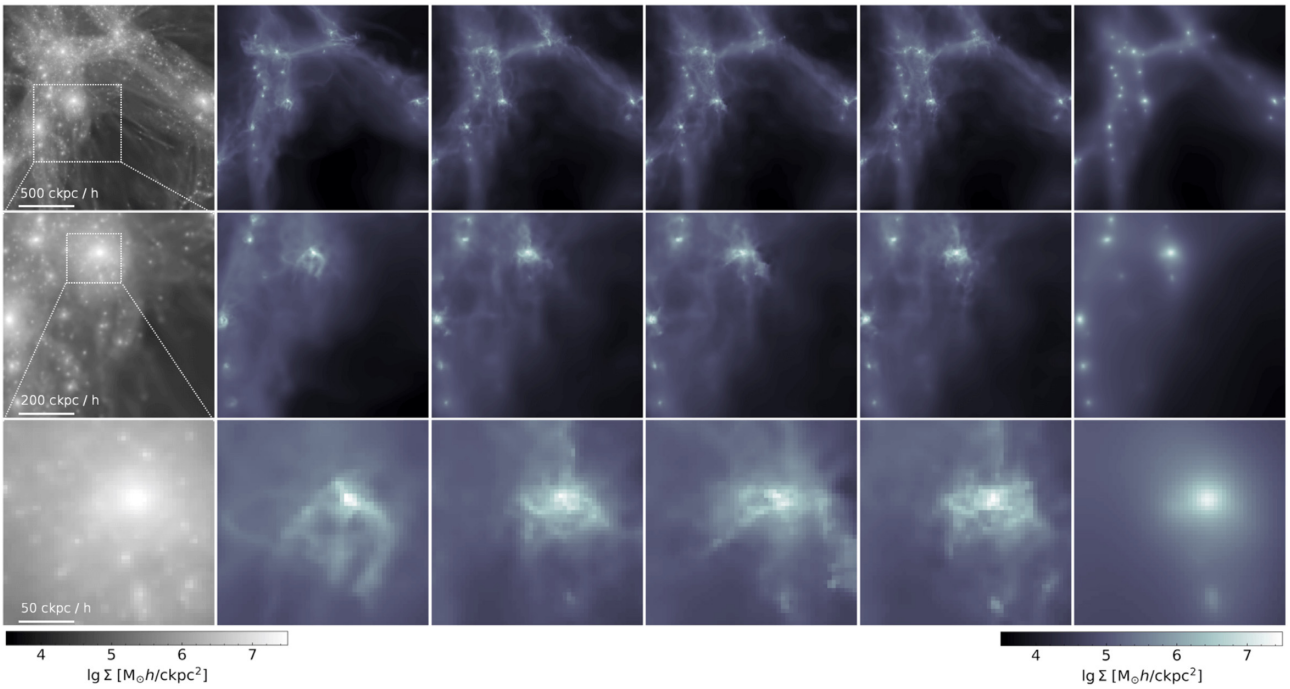

Figure 3. We show a summary grid of the network capabilities regarding visual feature richness for the HR case. The first column is the HR `dmh` input of the network, whereas the second column shows the corresponding target gas distribution in the FIREbox^{PF} simulation. The third, fourth, and fifth column are three samples produced by the WGAN and the last column is the U-Net prediction. The tiles in the first row have pixel dimensions of 512^2. We also show zoomed-in regions in the second and third row to highlight the rich features produced by the WGAN on smaller scales.

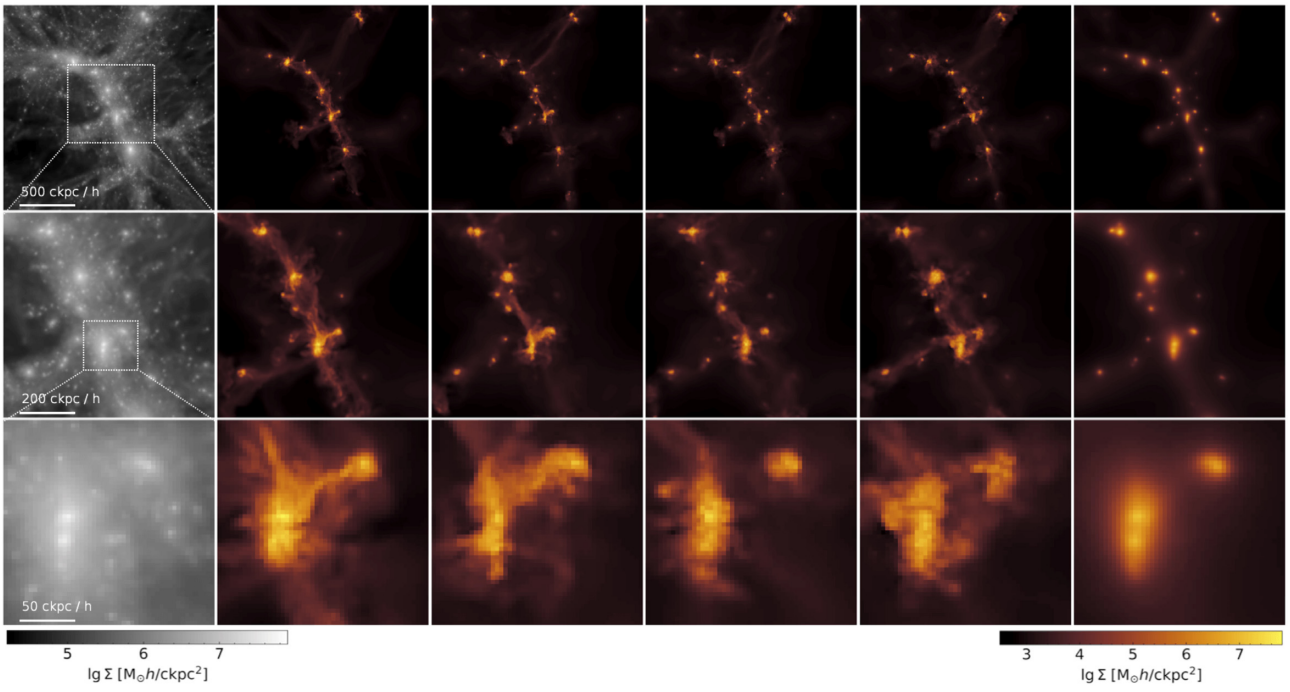

Figure 4. Same as Fig. 3 but for the H I predictions. The log-scaling of the colour scheme is the same as in Fig. 3 to emphasize the stronger clustering of H I.

upscaling capabilities of the WGANs when trained on lower resolution dark matter inputs and discuss the impact on the predictive power.

For the MR case we show the same summary statistics in Figs 5 and 6. In general, the mappings between dark matter and gas become more difficult in the case of lower resolution inputs because small scale features are absent in the input fields. Therefore, an accurate encoding of individual features through the noise inputs is necessary to obtain accurate summary statistics. Figs 5 and 6 show that the WGAN successfully reproduce power spectra within ∼10 per cent and bispectra within ∼20 per cent down to scales of ∼10 ckpc h^{-1}. Interestingly, in the extreme upsampling cases from *LR/VLR* → *HR* the WGAN still performs well on the power and

Table 4. Summary table of the median fractional error (given in per cent) of the predicted total mass in the box for the total of eight networks across all tested simulation resolutions and input fields. We denote downsampled dark matter inputs with 'ds', whereas otherwise the native resolution is taken as input. All masses are computed and compared to the (ground truth) FIREbox^{PF} (1024^3) test data set. The fiducial WGAN models achieve uncertainties of ∼2 per cent for total gas and ∼5 per cent for H I. Generally, predicting the targets from lower resolutions results in larger errors and predicting H I masses is systematically more difficult than total gas masses.

Network	Map	Input res.	Δ_m [per cent]
WGAN	dmo/dmh → gas	1024^3 native	1.92
U-Net	dmo/dmh → gas	1024^3 native	−1.13/−0.56
WGAN	dmo → gas	256^3 native	1.71
WGAN	dmo/dmh → gas	256^3 ds	1.45/1.22
WGAN	dmo/dmh → gas	64^3 ds	0.86/0.86
WGAN	dmo/dmh → gas	16^3 ds	−8.12/−8.12
WGAN	dmo/dmh → H I	1024^3 native	5.54/5.54
U-Net	dmo/dmh → H I	1024^3 native	−40.84/−37.33
WGAN	dmo → H I	256^3 native	2.53
WGAN	dmo/dmh → H I	256^3 ds	4.85/3.9
WGAN	dmo/dmh → H I	64^3 ds	−14.34/−14.34
WGAN	dmo/dmh → H I	16^3 ds	−75.76/−75.76

bispectra down to ∼50 ckpc h^{-1}. However, the pixel based statistics as the PDF and the CDDF show larger deviations, presumably because the power and bispectra are predominantly determined by the high density pixels, whereas the PDF and CDDF depend on the entire pixel range. In particular, our method fails to reproduce any sensible prediction of the CDDF in the extreme upsampling case of *VLR* → *HR*.

5.3 Predictive power of EMBER

In this section we quantify the predictive power of EMBER by exploring how well the networks perform when tested on physically motivated metrics. Furthermore, we apply the WGAN model to a large dark matter only simulation (B100) to extend the metrics and compare them to related work.

5.3.1 Halo based analysis

The WGAN methodology defines a framework that is completely halo-free. It is therefore interesting to conduct a halo-based analysis despite the model not knowing the notion of a dark matter halo. For this we compare the projected gas masses within one projected virial radius to the corresponding dark matter masses of individual haloes.

First, we identify dark matter haloes with `Amiga Halo Finder` (Knollmann & Knebe 2009).[5] We then use the information of the virial radii to construct 2D masks and compute the included dark matter and gas masses from the projected density fields. We note that this approach computes the masses over the entire 1.5 h^{-1} cMpc which would in principle overestimate H I masses by including gas outside R_{vir}. However, as shown in Feldmann et al. (in preparation), at $z = 2$ almost all H I resides inside dark matter haloes and the amount of H I beyond one virial radius is negligible. We refer to Appendix B for a more detailed discussion.

We show the results of this analysis in Fig. 7. To quantify the intrinsic scatter in the data set we bin the dark matter masses and compute the median gas masses for each bin, whereas the lower panel indicates the true and predicted amount of intrinsic scatter. Interestingly, the deterministic U-Net shows non-zero scatter across all halo masses and the WGAN prediction is in even better agreement with the simulation. Furthermore, Fig. 7 exhibits a very interesting aspect. The exact amount of gas does not just simply depend on the halo mass alone, but also on the dark matter environment. Dark matter haloes of the same mass but at different locations in the cosmic web can contain varying amounts of gas. For a fixed halo mass bin the scatter in the gas and H I counterparts can be large depending on the cosmic environmental density as well as the surrounding gas reservoir on Mpc-scale. Since the networks have access to this large-scale information, the predictions can account for the dark matter environment. To emphasize this aspect, mass bins are colour coded according to their median 2D number density in Fig. 7, which is computed by counting the number of neighbouring haloes within a $r_0 = 1\,h^{-1}$ Mpc disc, centred at the halo of interest, i.e.

$$n_{\mathrm{halo,2D}} = \frac{N(< r_0)}{\pi r_0^2}. \tag{10}$$

Furthermore, the halo based analysis suggests that at fixed redshift the dark matter cosmic web contains sufficient information to reconstruct the gas mass distribution (see e.g. Kraljic et al. 2019). Apparently, the required information from the halo growth history is encoded in the environment such that including information from previous redshifts is not strictly necessary.

The FIREbox^{PF} simulation shows large scatter in $M_{\mathrm{H\,I}}$ for haloes below 10^{11} M$_\odot$. We find a scatter of order 0.5 dex, with an extreme case of 2 dex around 10^{10} M$_\odot$. The halo mass to H I mass relation has previously been modelled with abundance matching (AM) techniques (see e.g. Papastergis et al. 2013; Padmanabhan & Kulkarni 2017; Padmanabhan, Refregier & Amara 2017; Spina, Porciani & Schimd 2021). Typically, these models target halo masses above ∼10^{10} M$_\odot$ as they are constrained from observational data. The large scatter for low-mass haloes indicates that simple AM models might break down here as more information about the dark matter halo is necessary to accurately predict the contained H I mass. We show this behaviour in Fig. 7 where we compare our results to the AM predictions of Padmanabhan & Kulkarni (2017) at $z = 2$. For large halo masses, our sample shows a small scatter and follows the theoretical linear relation very closely. However for halo masses below 10^{10} M$_\odot$ a linear relation is a poor description of the data as there are subgroups that tend to populate distinct places in the mass plane. We argue that our neural network approach is better suited to model the mapping in these regimes because it accounts for the higher complexity of the relation. At fixed halo mass the H I mass correlates with the environmental density. Fig. 7 shows that the H I content in haloes with different environments (colours indicating different halo number densities) as well as the predicted scatter, are almost perfectly reconstructed by the neural network indicating that the environment is necessary for learning the mean and scatter of the relation. Overall, our WGAN model makes more accurate predictions than the U-Net, which underestimates the scatter for halo masses 10^8–10^{10} M$_\odot$. We therefore conclude that the WGAN offers a valuable approach to model the dark matter to H I mass relation across the entire range probed by the training data and especially in the low-mass end where environmental effects become noticeably more important.

[5]Code available at: popia.ft.uam.es/AHF/Download.html.

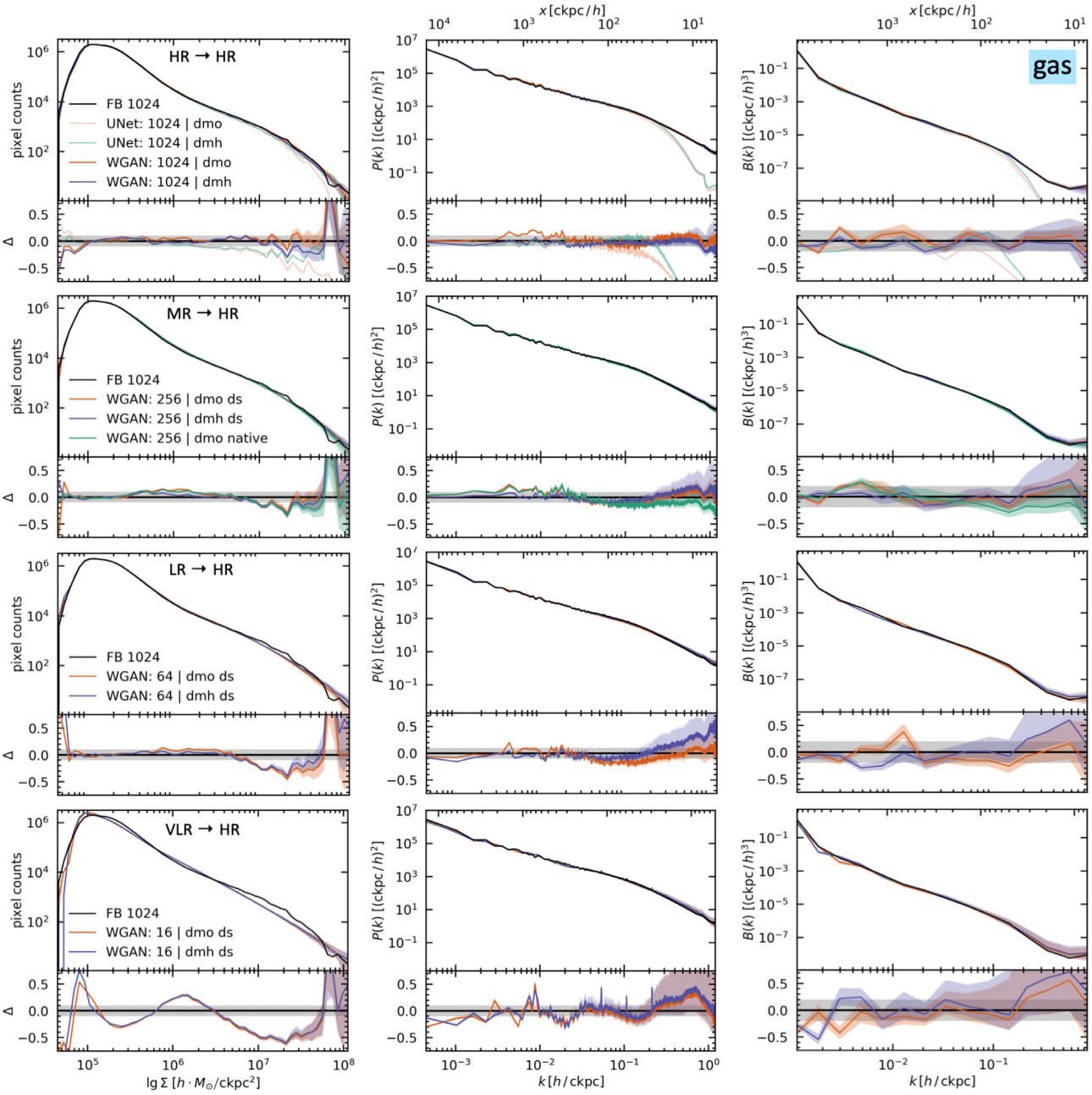

Figure 5. PDF, power spectra, and bispectra (from left to right) for the predicted gas projections for WGAN and U-Net when mapping from dmh and dmo of the FIREbox^{PF} simulation. The bispectra are shown for the case of equilateral triangles. The four rows represent the statistics for the *HR*, *MR*, *LR*, and *VLR* mappings. FB 1024 denotes the FIREbox^{PF} *HR* statistics. Note that e.g. (1024 | dmo) indicates that the 2D maps from the 1024^3 dark matter only simulation are used as network input. The lower panels show the fractional error when compared to the ground truth. Note that the error is computed on the quantities directly and not on the log of the quantities. The shaded bands denote the 10 or 20 per cent fractional error limit, depending on the statistic that is shown. For the WGANs we plot quantiles (16, 50, and 84) derived from the 128 predicted boxes as a true scatter estimate of the WGAN. All curves show predictions on the test set.

5.3.2 Application to B100

Fig. 6 shows that the WGAN prediction on the power and bispectrum is very similar across the different input fields, despite the model being trained on dmh ds as input. This behaviour is crucial when applying the model to enrich dark matter only simulations with larger box sizes. To demonstrate the pipeline, we predict the HI maps for the B100 simulation, which has very similar mass resolution properties as the *MR* training set (see Table 1 for details). Since the generator network is fully convolutional, we can predict entire HI slices simultaneously, eliminating the problem of edge effects. However, due to the large memory consumption, this operation is currently only possible on CPUs. The prediction of one HI mass map takes ∼1 h. Fig. 8 displays the summary statistics of the B100 predictions, which nicely extend and match the *HR* FIREbox^{PF} and *MR* WGAN prediction on large and small scales. We show an example slice of the projected box in Fig. 9.

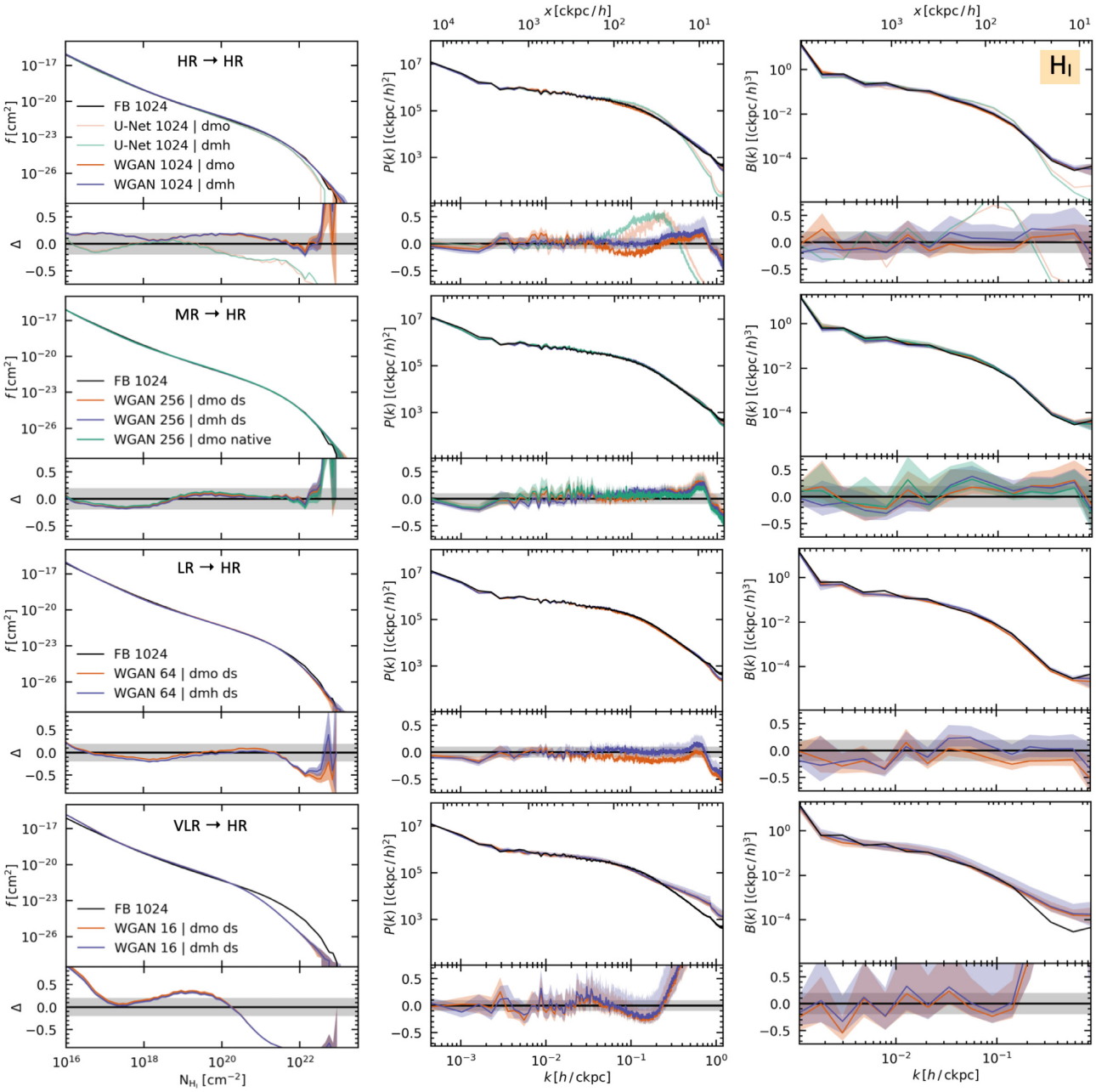

Figure 6. Similar to Fig. 5 but showing the column density distribution function, power spectra, and bispectra (from left to right) of the H I gas component. The shaded bands denote the 10 or 20 per cent fractional error limit, depending on the statistic that is shown. As for the gas case we plot quantiles (16, 50, and 84) derived from the 128 predicted boxes as a true scatter estimate of the WGAN. The H I maps are predicted from the test data sets of the hydrodynamical (dmh) and the dark matter-only (dmo) FIREbox^{PF} simulation.

The CDDF is a measure of the number of H I absorbers per unit column density. Fig. 8 demonstrates how the network extrapolates the CDDF for B100 with respect to FIREbox^{PF} when probing a larger volume. In all our experiments we observe a steepening of the CDDF beyond log $N_{HI} \geq 21$, which then linearly extends towards higher density systems. This result is in good agreement with related work in Rahmati et al. (2015) up to 10^{21} cm^{-2}. Rahmati et al. (2013) showed that in order to have a converged CDDF up to 10^{22} cm^{-2} one requires a box size of at least 50 Mpc h^{-1}. Furthermore, when identifying sightlines with grid cells, a coarser grid will introduce an artificial smoothing, effectively shifting the CDDF to lower number densities. This is an additional advantage of our method, as it allows to compute column densities over high resolution grids for large boxes. The upsampling capabilities in our approach are especially interesting for creating mock observations for applications where a broad range of resolved halo masses is necessary. One such application is Intensity Mapping (IM; Kovetz et al. 2019; Padmanabhan 2019), where a significant contribution to the H I signal is expected to come from galaxies residing in lower mass haloes of $\approx 10^9\ h^{-1}\ M_\odot$ (Cunnington et al. 2018; Villaescusa-Navarro et al. 2018). Simultaneously, realistic mock observations also require simulations of large box sizes – hundreds of Mpc to Gpc – to supply the wide and deep light cones relevant to IM studies. The EMBER approach attempts to bridge these two extreme regimes by construction.

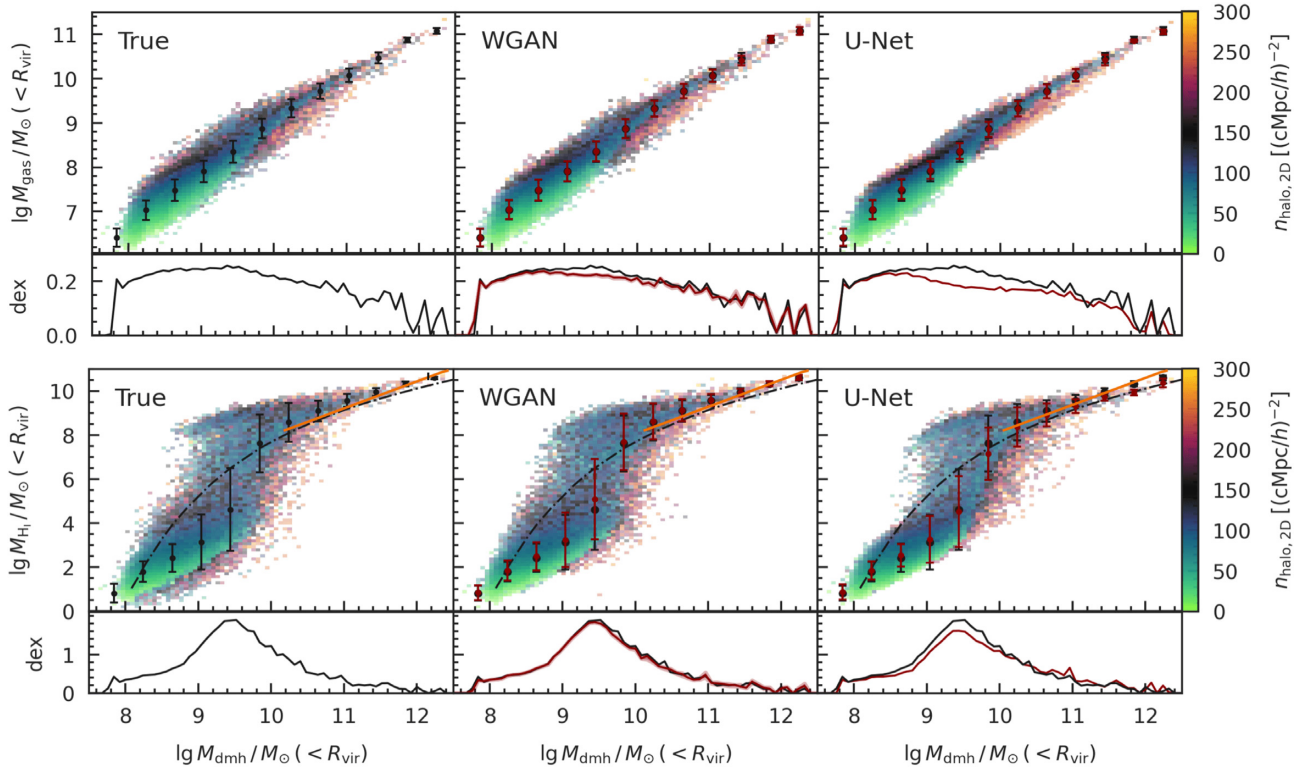

Figure 7. Summary figure of the halo analysis showing the relation between the projected gas (upper panel), H I (lower panel) and dark matter mass within one projected virial radius. Bins are colour coded by the median halo number density as a proxy for the halo environment where a higher transparency indicates fewer data points. We also show quantiles (16, 50, and 84) computed on the binned data. Also shown is the H I abundance matching result at $z = 2$ from Padmanabhan & Kulkarni (2017) in orange and the fit from Villaescusa-Navarro et al. (2018) as the black dash–dotted line. The curves in the lower panels indicate the scatter across the range of halo masses. Red lines with shaded area denote the (16, 50, and 84) quantiles for the 128 predicted boxes with WGAN whereas only the median is shown in the U-Net case. Black symbols denote true data points, which are also shown in the WGAN and U-Net panels for better visual comparison. Red symbols denote the corresponding predictions.

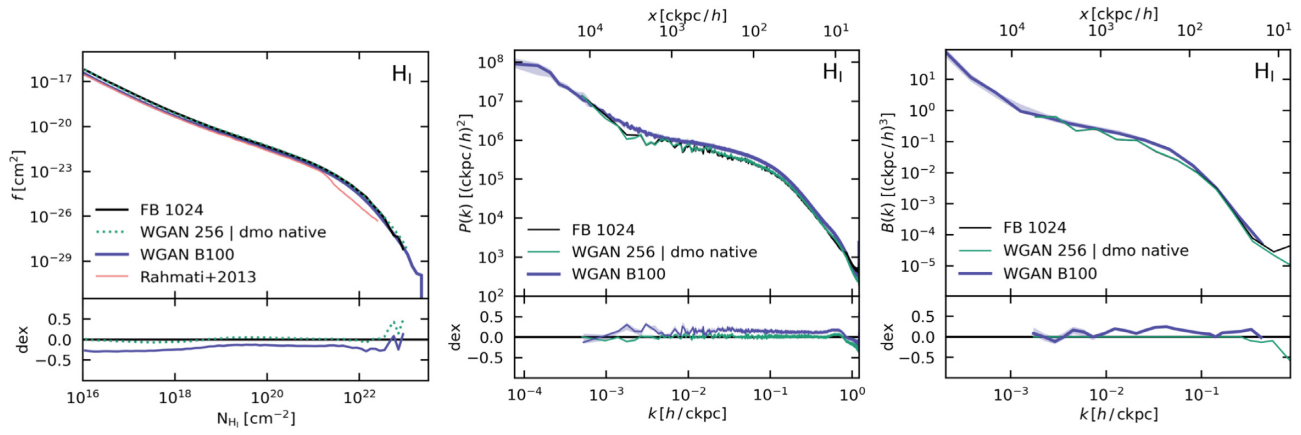

Figure 8. Here we show the CDDF, power spectrum and bispectrum of the B100 prediction. We also overplot the statistics from the FIREbox^{PF} simulation (*HR*, labelled as FB1024) and the WGAN prediction on the native dark matter-only *MR* FIREbox (labelled as WGAN 256 | dmo native) to emphasize how the WGAN model can be used on large dark matter simulations to extend the summary statistics, e.g. probing lower H I column densities. Shaded violet areas denote the 10 per cent error limit, representing the internal scatter of the WGAN predictions.

We also show the power and bispectrum of the B100 prediction compared to the FIREbox^{PF} simulation and the WGAN prediction on the *MR* `dmo native`. The statistics of the power and bispectrum are in good agreement between the three different predictions but the B100 statistics slightly differ on intermediate scales. On one hand, cosmic variance might be one possible explanation of this behaviour as we only have one FIREbox^{PF} realization to compare against. On the other hand, the effects from large-scale modes might be important as well. van Daalen, McCarthy & Schaye (2019) found that the most massive haloes have an effect on the power spectrum

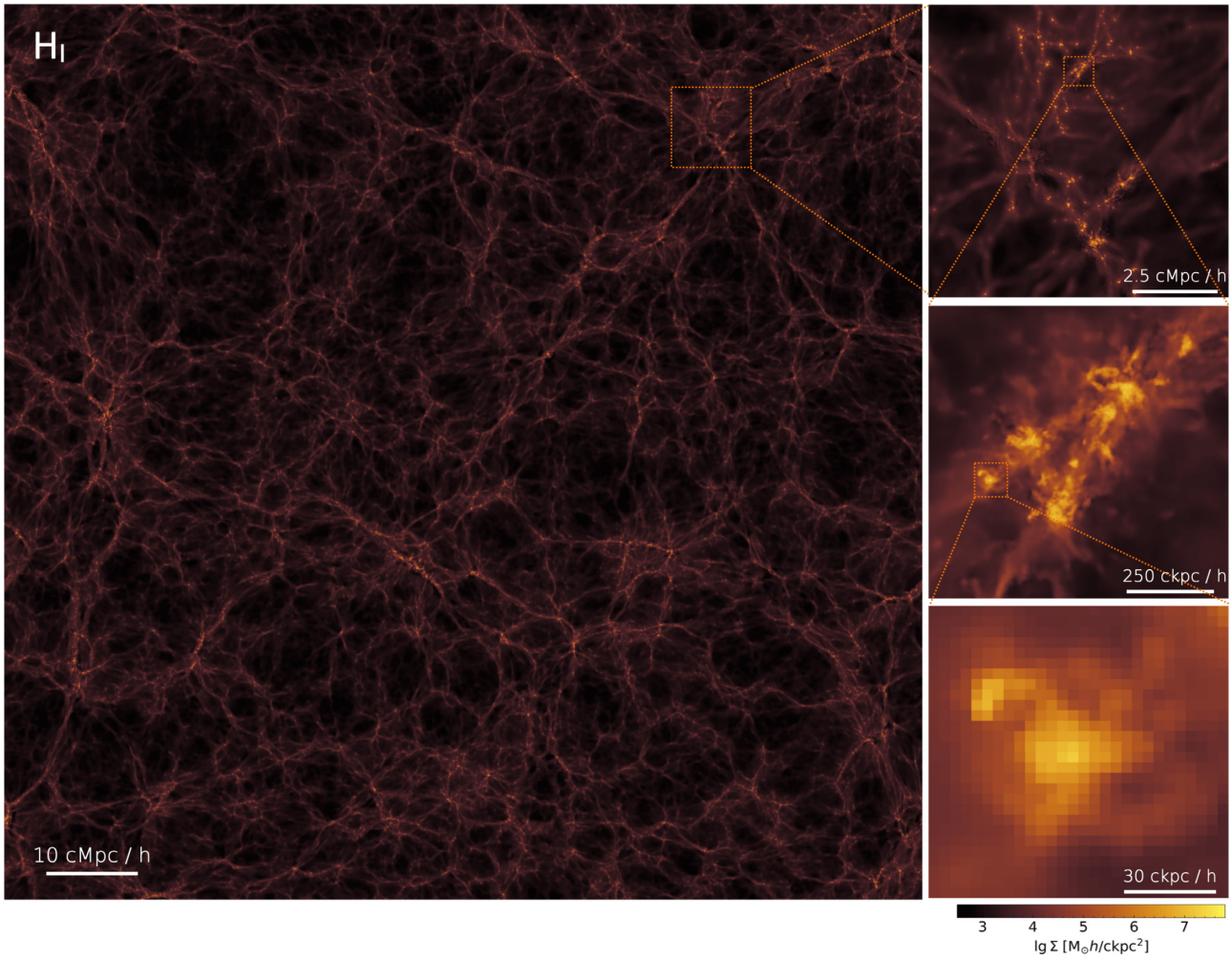

Figure 9. Illustration of an individual slice of 1.5 cMpc h^{-1} depth and 100 cMpc h^{-1} in size of the emulated H I field using our fiducial WGAN H I model. The full image shown is reduced in pixel resolution compared to the original version (due to filesize limitations) that consists of 27307^2 pixels, where one pixel resolves ∼3.6 ckpc h^{-1}. The full map can be found at the official github repository.

as they contribute to the power on scales $k \geq 10^{-2}\, h\,\mathrm{ckpc}^{-1}$. This aspect was pointed out before by Chisari et al. (2018), who compared power spectra from sub-volumes drawn from their fiducial $100\,h^{-1}$ cMpc simulation. They found significant variation between samples, depending on whether a massive object was present in a given sub-volume or not. For the comparison of the B100 with FIREbox^{PF} in Fig. 8, these results indicate that the slight difference in power is a combination of cosmic variance and the more massive haloes affecting intermediate scales.

Baryons modify the full density field and thus affect the entire hierarchy of higher order statistics beyond the power spectrum (Arico et al. 2020). Foreman et al. (2020) showed that baryonic effects on the bispectrum of hydrodynamical simulations carry additional information with respect to the power spectrum. As showed in Foreman et al. (2020) and Arico et al. (2020), massive haloes contribute to baryonic effects more in the bispectrum than in the power spectrum. Consequently, the bispectrum measured in relatively small boxes is different at small scales with respect to larger boxes. This effect is similar to the one for the power spectrum. To mitigate this issue we included four zoom-in simulations of massive haloes such that the training data represented the modes of individual massive haloes.

6 SUMMARY AND CONCLUSIONS

We have presented EMBER, a novel deep-learning-based framework to emulate baryonic maps, specifically for gas and atomic hydrogen, from dark matter data alone. Our training data are based on hydrodynamical simulations run with the FIRE-2 physics model at high numerical resolution ($m_\mathrm{b} = 3$–$6 \times 10^4\,\mathrm{M}_\odot$). This simulation suite includes a $(15\,\mathrm{cMpc}/h)^3$ cosmological volume simulation (FIREbox^{PF}) and several zoom-in simulations of massive haloes. We emphasize that combining small cosmological boxes with high-resolution zoom-in simulations is a crucial ingredient for applying our model to predict baryon fields for dark matter simulations of larger box sizes. This methodology ensures that the entire range of scales is represented in the training set.

By applying EMBER to test data, we found that it is able to reproduce important physical statistics such as power and bispectra as well as the H I CDDF. In particular, FIREbox^{PF} predicts that the relationship between halo mass and H I galaxy mass is environmentally dependent and breaks dramatically with significant scatter around a characteristic halo mass scale $\sim 5 \times 10^9\,\mathrm{M}_\odot$. Such a trend would be difficult to capture in a traditional halo-mass-based approach, while EMBER is able to reproduce the relationship with remarkable accuracy using only the dark matter field (see Fig. 7). Furthermore,

we showed that EMBER is capable of emulating high resolution baryon information from low resolution dark matter inputs through upsampling techniques. This is an extremely attractive property of our approach, as it allows to populate large, but low resolution, dark matter simulations with baryon fields at the high resolution level of the training data. In the following, we summarize our main findings in more detail.

(i) We showed that a stochastic WGAN architecture is able to capture and learn the feature distribution on very small scales better than a U-Net model. We conclude that the additional variance that the WGAN offers is necessary to accurately model the mapping from dark matter to baryons especially on scales $\leq$100 ckpc h^{-1}.

(ii) In particular, we found that the H I U-Net struggles when comparing the total mass in the box. Wadekar et al. (2020) used a second network trained just on the high density pixels to mitigate this problem. However, we found that the WGANs do not suffer from this problem and adapt well to the different target fields.

(iii) The WGAN models are able to reproduce the power spectra on the corresponding gas targets with $\sim$10 per cent accuracy down to $\sim$10 ckpc h^{-1}. Furthermore, we found very good agreement on the pixel PDF and the CDDF. The bispectra of the predictions are within 20 per cent accuracy indicating that the adversarial networks are capable of capturing even higher order moments in the data set such as the filamentary structures between haloes.

(iv) We conducted a halo-based analysis to compare the true and predicted gas masses within one virial radius of parent dark matter haloes. The network predictions agree very well with FIREbox^{PF} and the analytical AM relation from Padmanabhan & Kulkarni (2017), indicating that the network has learned to retrieve features in the surrounding of the dark matter haloes to determine the contained gas mass. This is a big advantage of EMBER as it defines a halo-free method (compare e.g. with Lovell et al. 2021) that is sensitive to the dark matter environment on a large range of scales and can thus be used to extend the result of current AM techniques down to very low halo masses. Furthermore, the reproduced scatter in the gas mass at fixed halo masses is very close to the intrinsic scatter in the data set despite the fact that the model inputs have no dynamical information. We will investigate this direction in the future to better understand the main drivers that determine the gas masses and scatter in haloes and how it depends on the underlying dark matter morphology.

(v) We investigated the case of predicting the target fields from maps that are derived from lower resolution dark matter simulations and compared them with the high resolution targets. From the analysis, we conclude that the networks were still able to make accurate predictions in terms of the summary statistics when reducing the input resolution by two levels to 256^3, i.e. a factor 64 in mass.

(vi) We conducted the same analysis for the extreme upsampling cases by reducing the dark matter resolution by four levels (64^3) and six levels (16^3) and found reasonably accurate predictions (typically $\sim$20 per cent error) on the gas and H I power spectrum and bispectrum down to scales $\sim$50 ckpc h^{-1}. However, the extreme upscaling from 16^3 to 1024^3 results in errors far beyond 10 per cent when predicting the PDF, CDDF and the total mass in the box.

(vii) We quantitatively illustrated the application of how the WGAN models can be used to predict gas maps for larger cosmological volumes. More specifically, we applied the WGAN H I network trained on our hydrodynamical simulation suite to emulate H I mass maps for a 100 cMpc h^{-1} dark matter simulation. On small scales, the predicted gas and H I power spectrum and CDDF agree well with those in our simulation suite, while on large scales they are in good agreement with the results by Rahmati et al. (2015).

Overall, our fiducial WGAN approach shows very good agreement for all the tested metrics and has excellent upsampling capabilities when presented with lower resolution information.

Future work

In the current implementation, EMBER regresses only one target field from one single input channel. One promising extension would be to add more dark matter inputs, e.g. the velocity field or the dispersion thereof to include dynamical information. Using the velocity dispersion as an additional feature might allow to identify merging systems and subsequently information on the recent assembly history and the gas content.

Furthermore, the network architecture is easily extended to predict additional baryon fields such as temperature, pressure, or stellar densities. Since the noise input in the WGAN models is responsible for generating the small-scale features, training on multiple target fields simultaneously ensures that predictions of different baryon fields are coherent when sampling. This is an important advantage as opposed to training multiple networks that each predict one individual field.

Since our networks were trained on data from one redshift, this approach currently works for this specific epoch. In principle one can train the same network architectures on any specific redshift. Furthermore, it would be interesting to allow redshift interpolation by training on all redshift slices. We will investigate this path in future work.

Currently, our methodology operates on 2D projection maps that aggregate information over a slab thickness of 1.5 h^{-1} Mpc. This approach effectively produces tomographic maps of the underlying 3D volume. Tomographic approaches have recently been used to reconstruct from observations 3D density maps to compute Lyman alpha flux profiles (e.g. Lee et al. 2018; Newman et al. 2020; Li, Horowitz & Cai 2021). Our approach is well matched with current line-of-sight resolution limits of spectrographs such as e.g. MUSE (Bacon et al. 2010), which are in fact of order $\sim$Mpc (see e.g. Ravoux et al. 2020, and references therein).

We also conducted a halo-based analysis to understand to what extent different dark matter environments influence the prediction of the gas masses contained in the haloes. In future work, we will investigate a framework to quantify the importance of the dark matter environment for the predictions, since this is the major advantage that our methodology offers compared to halo-based models. Finally, the methodology of EMBER presented in this work promises an exciting pathway for fast and accurate enrichment of dark matter only simulations with high resolution baryon information.

ACKNOWLEDGEMENTS

MB thanks Tomasz Kacprzak for fruitful discussions and Darren Reed for technical support for the GPU training. Furthermore the authors want to thank Claude-André Faucher-Giguère, Hugues Lascombes, Romain Teyssier, and Aurel Schneider for helpful comments and insights that helped improve this work. RF acknowledges financial support from the Swiss National Science Foundation (grant no 157591 and 194814). DAA acknowledges support by NSF grant AST-2009687 and by the Flatiron Institute, which is supported by the Simons Foundation. MBK acknowledges support from NSF CAREER award AST-1752913, NSF grant AST-1910346, NASA grant NNX17AG29G, and *HST*-AR-15006, *HST*-AR-15809, *HST*-GO-15658, *HST*-GO-15901, *HST*-GO-15902, *HST*-AR-16159, and *HST*-GO-16226 from the Space Telescope Science Institute, which

is operated by AURA, Inc., under NASA contract NAS5-26555. We acknowledge PRACE for awarding us access to MareNostrum at the Barcelona Supercomputing Center (BSC), Spain. This research was partly carried out via the Frontera computing project at the Texas Advanced Computing Center. Frontera is made possible by National Science Foundation award OAC-1818253. This work was supported in part by a grant from the Swiss National Supercomputing Centre (CSCS) under project IDs s697 and s698. We acknowledge access to Piz Daint at the Swiss National Supercomputing Centre, Switzerland under the University of Zurich's share with the project ID uzh18. This work made use of infrastructure services provided by S3IT (http://www.s3it.uzh.ch), the Service and Support for Science IT team at the University of Zurich. Finally, we thank the referee Simeon Bird for his valuable input that helped improve this work.

DATA AVAILABILITY STATEMENT

We provide the source code, networks and maps at the official repository: https://github.com/maurbe/ember.

REFERENCES

Abadi M. et al., 2016, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), p. 265
Ade P. A. R. et al., 2016, A&A, 594, A13
Agarwal S., Davé R., Bassett B. A., 2018, MNRAS, 478, 3410
Altay G., Theuns T., Schaye J., Crighton N. H. M., Dalla Vecchia C., 2011, ApJ, 737, L37
Anglés-Alcázar D., Davé R., Özel F., Oppenheimer B. D., 2014, ApJ, 782, 84
Anglés-Alcázar D., Faucher-Giguère C.-A., Kereš D., Hopkins P. F., Quataert E., Murray N., 2017a, MNRAS, 470, 4698
Anglés-Alcázar D., Faucher-Giguère C.-A., Quataert E., Hopkins P. F., Feldmann R., Torrey P., Wetzel A., Kereš D., 2017b, MNRAS, 472, L109
Aricò G., Angulo R. E., Hernández-Monteagudo C., Contreras S., Zennaro M., Pellejero-Ibañez M., Rosas-Guevara Y., 2020, MNRAS, 495, 4800
Arjovsky M., Bottou L., 2017, preprint (arXiv:1701.04862)
Arjovsky M., Chintala S., Bottou L., 2017, ICML
Bacon R. et al., 2010, in McLean I. S., Ramsay S. K., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III. SPIE, Bellingham, p. 773508
Bahcall J. N., Peebles P. J. E., 1969, ApJ, 156, L7
Barnes D. J., Kannan R., Vogelsberger M., Marinacci F., 2020, MNRAS, 494, 1143
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Benson A. J., 2012, NewA, 17, 175
Bertschinger E., 2001, ApJS, 137, 1
Bett P., Eke V., Frenk C. S., Jenkins A., Okamoto T., 2010, MNRAS, 404, 1137
Biernacki P., Teyssier R., 2018, MNRAS, 475, 5688
Bird S., Vogelsberger M., Haehnelt M., Sijacki D., Genel S., Torrey P., Springel V., Hernquist L., 2014, MNRAS, 445, 2313
Blumenthal G. R., Faber S. M., Flores R., Primack J. R., 1986, ApJ, 301, 27
Bolatto A. D., Wolfire M., Leroy A. K., 2013, ARA&A, 51, 207
Brooks A. M., Governato F., Quinn T., Brook C. B., Wadsley J., 2009, ApJ, 694, 396
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Butsky I. et al., 2016, MNRAS, 462, 663
Cataldi P., Pedrosa S., Tissera P., Artale C., 2020, MNRAS, 501, 5679
Chabanier S., Bournaud F., Dubois Y., Palanque-Delabrouille N., Yèche C., Armengaud E., Peirani S., Beckmann R., 2020, MNRAS, 495, 1825
Chan T. K., Kereš D., Oñorbe J., Hopkins P. F., Muratov A. L., Faucher-Giguère C. A., Quataert E., 2015, MNRAS, 454, 2981
Chisari N. E. et al., 2018, MNRAS, 480, 3962
Chuang C.-H. et al., 2015, MNRAS, 452, 686
Chua K. T. E., Pillepich A., Vogelsberger M., Hernquist L., 2019, MNRAS, 484, 476
Cohen A., Fialkov A., Barkana R., Monsalve R. A., 2020, MNRAS, 495, 4845
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Cora S. A., Hough T., Vega-Martínez C. A., Orsi A. A., 2018, MNRAS, 483, 1686
Crain R. A. et al., 2016, MNRAS, 464, 4204
Croton D. J., Gao L., White S. D. M., 2007, MNRAS, 374, 1303
Cunnington S., Harrison I., Pourtsidou A., Bacon D., 2018, MNRAS, 482, 3341
Dai B., Seljak U., 2020, Proc. Natl. Acad. Sci., 118, 2020324118
Davé R., Thompson R., Hopkins P. F., 2016, MNRAS, 462, 3265
Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827
Decarli R. et al., 2019, ApJ, 882, 138
Dekel A., Birnboim Y., 2006, MNRAS, 368, 2
Diemer B. et al., 2019, MNRAS, 487, 1529
Faucher-Giguère C.-A., Kereš D., 2011, MNRAS, 412, L118
Faucher-Giguère C.-A., Kereš D., Dijkstra M., Hernquist L., Zaldarriaga M., 2010, ApJ, 725, 633
Faucher-Giguère C.-A., Kereš D., Ma C.-P., 2011, MNRAS, 417, 2982
Faucher-Giguère C.-A., Hopkins P. F., Kereš D., Muratov A. L., Quataert E., Murray N., 2015, MNRAS, 449, 987
Faucher-Giguère C.-A., Feldmann R., Quataert E., Kereš D., Hopkins P. F., Murray N., 2016, MNRAS, 461, L32
Feder R. M., Berger P., Stein G., 2020, Phys. Rev. D, 102, 103504
Feldmann R., 2020, Commun. Phys., 3, 226
Feldmann R., Mayer L., 2015, MNRAS, 446, 1939
Feldmann R., Carollo C. M., Mayer L., 2011, ApJ, 736, 88
Feldmann R., Hopkins P. F., Quataert E., Faucher-Giguère C.-A., Kereš D., 2016, MNRAS, 458, L14
Feldmann R., Quataert E., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., 2017, MNRAS, 470, 1050
Feldmann R., Faucher-Giguère C.-A., Kereš D., 2019, ApJ, 871, L21
Feng Y., Di-Matteo T., Croft R. A., Bird S., Battaglia N., Wilkins S., 2016, MNRAS, 455, 2778
Ferland G. J., Korista K. T., Verner D. A., Ferguson J. W., Kingdon J. B., Verner E. M., 1998, PASP, 110, 761
Foreman S., Coulton W., Villaescusa-Navarro F., Barreira A., 2020, MNRAS, 498, 2887
Fumagalli M., Prochaska J. X., Kasen D., Dekel A., Ceverino D., Primack J. R., 2011, MNRAS, 418, 1796
Fumagalli M., Hennawi J. F., Prochaska J. X., Kasen D., Dekel A., Ceverino D., Primack J., 2013, ApJ, 780, 74
Giusarma E., Reyes Hurtado M., Villaescusa-Navarro F., He S., Ho S., Hahn C., 2019, preprint (arXiv:1910.04255)
Glorot X., Bengio Y., 2010, Understanding the difficulty of training deep feedforward neural networks In Proc. of 13th International Conf. on Artificial Intelligence and Statistics (AISTATS'10) 249
Glowacki M. et al., 2019, MNRAS, 489, 4926
Gnedin O. Y., Kravtsov A. V., Klypin A. A., Nagai D., 2004, ApJ, 616, 16
Goodfellow I. J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y., 2014, Adv. Neural Inf. Process. Sys. 3
Governato F. et al., 2012, MNRAS, 422, 1231
Guglielmo V., Poggianti B. M., Moretti A., Fritz J., Calvi R., Vulcani B., Fasano G., Paccagnella A., 2015, MNRAS, 450, 2749
Gulrajani I., Ahmed F., Arjovsky M., Dumoulin V., Courville A., 2017, Proceedings of the 31st International Conference on Neural Information Processing Systems, p. 5769
Guo Q., White S., Li C., Boylan-Kolchin M., 2010, MNRAS, 404, 1111
Hahn O., Abel T., 2011, MNRAS, 415, 2101
Harrington P., Mustafa M., Dornfest M., Horowitz B., Lukic Z., 2021, preprint (arXiv:2106.12662)
Hirschmann M., Naab T., Somerville R. S., Burkert A., Oser L., 2011, MNRAS, 419, 3200
Hopkins P. F., 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. et al., 2018, MNRAS, 480, 800
Ho S. H., Martin C. L., Turner M. L., 2019, ApJ, 875, 54
Hwang H. S., Shin J., Song H., 2019, MNRAS, 489, 339
Jenni S., Favaro P., 2019, preprint (arXiv:1906.04612)
Jesseit R., Naab T., Burkert A., 2002, ApJ, 571, L89
Jo Y., Kim J.-H., 2019, MNRAS, 489, 3565
Karim A. et al., 2011, ApJ, 730, 61
Karnewar A., Wang O., 2019, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 7796
Karras T., Aila T., Laine S., Lehtinen J., 2017, preprint (arXiv:1710.10196)
Karras T., Laine S., Aila T., 2018, preprint (arXiv:1812.04948)
Katz N., White S. D. M., 1993, ApJ, 412, 455
Kazantzidis S., Abadi M. G., Navarro J. F., 2010, ApJ, 720, L62
Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2
Khandai N., Di Matteo T., Croft R., Wilkins S., Feng Y., Tucker E., DeGraf C., Liu M.-S., 2015, MNRAS, 450, 1349
Kingma D. P., Ba J., 2015, CoRR
Kirby E. M., Koribalski B., Jerjen H., Lopez-Sanchez A., 2012, MNRAS, 420, 2924
Knebe A. et al., 2015, MNRAS, 451, 4029
Knebe A., Domínguez A., 2003, Publ. Astron. Soc. Aust., 20, 173
Knollmann S. R., Knebe A., 2009, ApJS, 182, 608
Kodi Ramanah D., Charnock T., Villaescusa-Navarro F., Wandelt B. D., 2020, MNRAS, 495, 4227
Koopmans L. V. E. et al., 2015, preprint (arXiv:1505.07568)
Kovetz E. et al., 2019, Bull. Am. Astron. Soc., 51, 101
Kraljic K. et al., 2019, MNRAS, 491, 4294
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35
Lazar A. et al., 2020, MNRAS, 497, 2393
Lee K.-G. et al., 2018, ApJS, 237, 31
Li Y.-P. et al., 2018, ApJ, 866, 70
Li Y., Ni Y., Croft R. A. C., Matteo T. D., Bird S., Feng Y., 2020, Proc. Natl. Acad. Sci., 118, 2022038118
Li Z., Horowitz B., Cai Z., 2021, ApJ, 919, 20
Lovell C. C., Wilkins S. M., Thomas P. A., Schaller M., Baugh C. M., Fabbian G., Bahé Y., 2021, preprint (arXiv:2106.04980)
Mellema G. et al., 2013, Exp. Astron., 36, 235
Mirza M., Osindero S., 2014, preprint (arXiv:1411.1784)
Morganti R., Oosterloo T., 2018, A&AR, 26
Moster B. P., Naab T., Lindström M., O'Leary J. A., 2021, MNRAS, 507, 2115
Naab T., Johansson P. H., Ostriker J. P., 2009, ApJ, 699, L178
Nagamine K., Springel V., Hernquist L., 2004, MNRAS, 348, 421
Navarro J. F., Eke V. R., Frenk C. S., 1996, MNRAS, 283, L72
Nelson D., Vogelsberger M., Genel S., Sijacki D., Kereš D., Springel V., Hernquist L., 2013, MNRAS, 429, 3353
Nelson D. et al., 2017, MNRAS, 475, 624
Nelson D. et al., 2019, Computational Astrophysics and Cosmology, 6, 2
Newman A. B. et al., 2020, ApJ, 891, 147
Ntampaka M. et al., 2019, BAAS, 51, 14
Onorbe J., Garrison-Kimmel S., Maller A. H., Bullock J. S., Rocha M., Hahn O., 2014, MNRAS, 437, 1894
Oñorbe J., Boylan-Kolchin M., Bullock J. S., Hopkins P. F., Kereš D., Faucher-Giguère C.-A., Quataert E., Murray N., 2015, MNRAS, 454, 2092
Padmanabhan H., 2019, preprint (arXiv:1910.14059)
Padmanabhan H., Kulkarni G., 2017, MNRAS, 470, 340
Padmanabhan H., Refregier A., Amara A., 2017, MNRAS, 469, 2323
Papastergis E., Giovanelli R., Haynes M. P., Rodríguez-Puebla A., Jones M. G., 2013, ApJ, 776, 43
Pavesi R. et al., 2018, ApJ, 864, 49
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Perraudin N., Marcon S., Lucchi A., Kacprzak T., 2020, Frontiers in Artificial Intelligence, 4, 66
Pillepich A. et al., 2017, MNRAS, 473, 4077
Prelogovic D., Mesinger A., Murray S., Fiameni G., Gillet N., 2021, preprint (arXiv:2107.00018)
Pritchard J. et al., 2015, preprint (arXiv:1501.04291)
Rahmati A., Schaye J., Pawlik A. H., Raicevic M., 2013, MNRAS, 431, 2261
Rahmati A., Schaye J., Bower R. G., Crain R. A., Furlong M., Schaller M., Theuns T., 2015, MNRAS, 452, 2034
Ravoux C. et al., 2020, J. Cosmol. Astropart. Phys., 2020, 010
Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, ApJ, 771, 30
Reeves S. N., Sadler E. M., Allison J. R., Koribalski B. S., Curran S. J., Pracy M. B., 2015, MNRAS, 450, 926
Romeo A. B., Agertz O., Moore B., Stadel J., 2008, ApJ, 686, 1
Ronneberger O., Fischer P., Brox T., 2015, Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 234
Salimans T., Goodfellow I., Zaremba W., Cheung V., Radford A., Chen X., 2016, Proceedings of the 30th International Conference on Neural Information Processing Systems 2234
Schaye J. et al., 2010, MNRAS, 402, 1536
Schaye J. et al., 2015, MNRAS, 446, 521
Schneider A., Teyssier R., Stadel J., Chisari N. E., Brun A. M. L., Amara A., Refregier A., 2019, J. Cosmol. Astropart. Phys., 2019, 020
Sirko E., 2005, ApJ, 634, 728
Somerville R. S., Davé R., 2015, ARA&A, 53, 51
Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087
Spina B., Porciani C., Schimd C., 2021, MNRAS, 505, 3492
Springel V. et al., 2005, Nature, 435, 629
Springel V., Hernquist L., 2003, MNRAS, 339, 289
Stern J. et al., 2021, MNRAS, 507, 2869
Stern J., Fielding D., Faucher-Giguère C.-A., Quataert E., 2020, MNRAS, 492, 6042
Sønderby C. K., Caballero J., Theis L., Shi W., Huszár F., 2016, preprint (arXiv:1610.04490)
Tacconi L. J. et al., 2018, ApJ, 853, 179
Tacconi L. J., Genzel R., Sternberg A., 2020, ARA&A, 58, 157
Tamosiunas A., Winther H. A., Koyama K., Bacon D. J., Nichol R. C., Mawdsley B., 2020, MNRAS, 506, 3049
Thiele L., Villaescusa-Navarro F., Spergel D. N., Nelson D., Pillepich A., 2020, ApJ, 902, 129
Tissera P. B., Domínguez-Tenreiro R., 1998, MNRAS, 297, 177
Tröster T., Ferguson C., Harnois-Déraps J., McCarthy I. G., 2019, MNRAS, 487, L24
Valentini M. et al., 2019, MNRAS, 491, 2779
van Daalen M. P., McCarthy I. G., Schaye J., 2019, MNRAS, 491, 2424
Villaescusa-Navarro F. et al., 2018, ApJ, 866, 135
Villaescusa-Navarro F. et al., 2020a, ApJ, 915, 71
Villaescusa-Navarro F., Wandelt B. D., Anglés-Alcázar D., Genel S., Zorrilla Mantilla J. M., Ho S., Spergel D. N., 2020b, preprint (arXiv:2011.05992)
Vogelsberger M., Sijacki D., Kereš D., Springel V., Hernquist L., 2012, MNRAS, 425, 3024
Vogelsberger M. et al., 2014, MNRAS, 444, 1518
Wadekar D., Villaescusa-Navarro F., Ho S., Perreault-Levasseur L., 2020, ApJ, 916, 42
Wang Z., Bovik A., Sheikh H., Simoncelli E., 2004, IEEE Trans. Image Process., 13, 600
Wechsler R. H., Tinker J. L., 2018, ARA&A, 56, 435
Weinberg S., 1972, Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity, John Wiley & Sons, Inc.
Weltman A. et al., 2020, PASA, 37, 2
Wetzel A. R., Nagai D., 2015, ApJ, 808, 40
Wetzel A. R., Hopkins P. F., Kim J.-h., Faucher-Giguère C.-A., Kereš D., Quataert E., 2016, ApJ, 827, L23
Woods R. M., Wadsley J., Couchman H. M. P., Stinson G., Shen S., 2014, MNRAS, 442, 732
Zamudio-Fernandez J., Okan A., Villaescusa-Navarro F., Bilaloglu S., Derin Cengiz A., He S., Perreault Levasseur L., Ho S., 2019, preprint (arXiv:1904.12846)
Zhang L., Zhang Y., Gao Y., 2018, preprint (arXiv:1812.00810)
Zhang X., Wang Y., Zhang W., Sun Y., He S., Contardo G., Villaescusa-Navarro F., Ho S., 2019, preprint (arXiv:1902.05965)

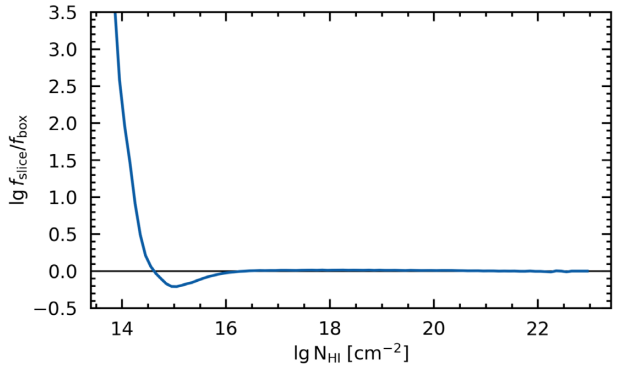

Figure A1. We show the ratio of the H I CDDF in the FIREbox^{PF} simulation when computing it by accumulating pixels over individual slabs compared to the CDDF of the projection of the total box. The comparison indicates that for systems above $N_{HI} > 10^{16}$cm^{-2} the statistics are identical.

APPENDIX A: CDDF

In this section, we discuss important aspects regarding the computation of the CDDF for simulation data with finite box size L. The absorption length $X(z)$ depends on the redshift z via the following integral expression (Bahcall & Peebles 1969; Nagamine, Springel & Hernquist 2004)

$$X(z) = \int_0^z (1+x)^2 \frac{H_0}{H(x)}\,\mathrm{d}x. \tag{A1}$$

The comoving distance L to redshift z is determined by the cosmology through

$$L = c\int_0^z \frac{\mathrm{d}x}{H(x)} \quad \text{with} \quad \mathrm{d}L = \frac{c}{H(z)}\,\mathrm{d}z. \tag{A2}$$

The differential absorption length ΔX is then computed as following:

$$\begin{aligned}\frac{\mathrm{d}X}{\mathrm{d}L} &= \frac{\mathrm{d}}{\mathrm{d}L}\int_0^z (1+x)^2\frac{H_0}{H(x)}\mathrm{d}x \\ &= \frac{H(z)}{c}\frac{\mathrm{d}}{\mathrm{d}z}\int_0^z (1+x)^2\frac{H_0}{H(x)}\mathrm{d}x = \frac{H_0}{c}(1+z)^2.\end{aligned} \tag{A3}$$

The CDDF is a pixel-based statistic computed over 2D projection maps. The exact shape of the CDDF therefore depends on the projection width of individual slabs, because individual systems can in principle overlap. We check that the CDDF has converged for our use-case by computing it once for the H I maps projected over the entire box (f_{totalbox}) and once for the CDDF accumulated over all slabs f_{slabs} (shown in Fig. A1). The comparison study clearly shows that the CDDF is identical for the two approaches above $N_{HI} > 10^{16}$cm^{-2}. Hence, systems with lg $N_{HI} > 16$ rarely overlap in the box. Since we are not interested in lower column density systems in this study, we conclude that either approach is applicable for our prediction pipeline.

APPENDIX B: HALO BASED ANALYSIS

In Fig. B1, we show a more detailed comparison between dark matter halo masses computed from AHF and the projected masses that we used for our halo based approach. The figure compares these two masses and shows how the offset changes with respect to the grid resolution and mass scale. Clearly, the projected halo masses are systematically larger than the AHF masses. This trend becomes stronger for smaller halo masses (and thus smaller virial radii). The explanation of this effect is two-fold. First, the grid resolution plays an important role as higher resolution masks can approximate the exact virial radius better and thus get a better estimate of the contained mass. We show this analysis for different pixel resolutions and conclude that at our resolution level of $3.6\,h^{-1}$ckpc this relation is converged. Secondly, projecting the mass over an entire slab will overestimate the masses within individual haloes. This effect becomes stronger for smaller haloes as the contamination becomes higher. The largest differences are of order 0.5 dex and for halo masses $\log(M_{\mathrm{dmh}}/\mathrm{M}_\odot)(<R_{\mathrm{vir}}) > 10.5$ the two masses are almost identical.

In Fig. B2, we show the halo counts in the FIREbox^{PF} simulation for individual mass bins as described in Section 5.3. For the gas the distribution is fit with a broken power law with breaking point at $\log(M_{\mathrm{dmh}}/M_\odot)_{\mathrm{break}} = 10.46(2)$ and slope parameters $\alpha_1 = 1.16(1)$ and $\alpha_2 = 0.89(1)$.

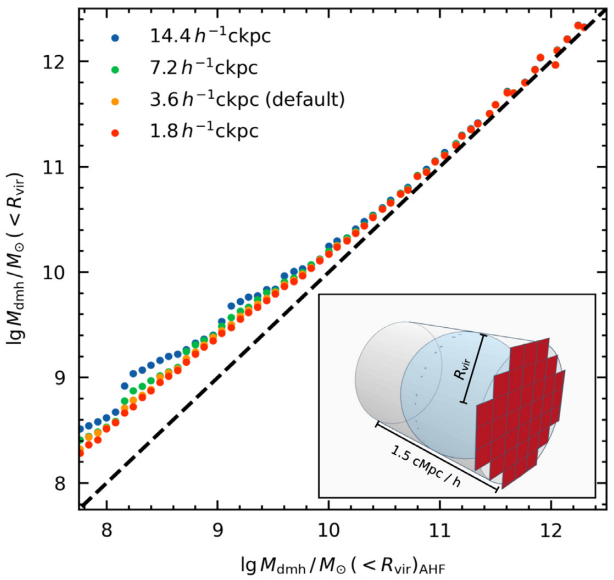

Figure B1. This figure shows a comparison between the median dark matter masses within one virial radius and the projected halo masses over one slab thickness of $1.5\,h^{-1}$ cMpc. The inset figure illustrates our approach. The mass within one virial radius as computed by AHF is shown in blue. The cylinder over which the masses are projected in our approach is shown in grey. Since our approach operates upon grids, we also show an example of a pixelized halo mask in red. The x-axis corresponds to the true AHF mass in blue while the y-axis is the projected mass shown in red. We also show this relation for different pixel resolutions indicated in the legend. Note that $3.6\,h^{-1}$ ckpc corresponds to the resolution that is used throughout this work.

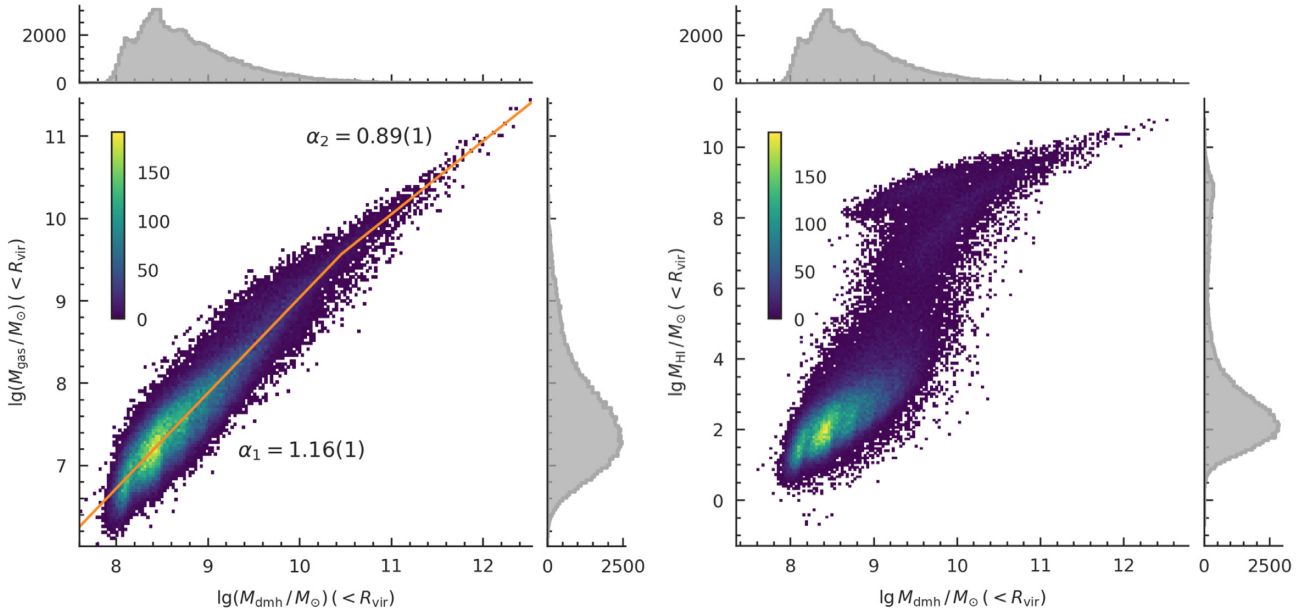

Figure B2. Joint distributions showing the projected dark matter and gas masses in individual haloes corresponding to Fig. 7. We provide the underlying data at the official github repository https://github.com/maurbe/ember.

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