

## Realistic simulations of galaxy formation in $f(R)$ modified gravity

Christian Arnold,<sup>1,\*</sup> Matteo Leo,<sup>1,2</sup> and Baojiu Li<sup>1</sup>

<sup>1</sup>*Institute for Computational Cosmology, Department of Physics,  
Durham University, South Road, Durham DH13LE, UK*

<sup>2</sup>*Institute for Particle Physics Phenomenology, Department of Physics,  
Durham University, South Road, Durham DH13LE, UK*

(Dated: July 8, 2019 - preprint, published in Nature Astronomy)

We have carried out a set of cosmological hydrodynamical simulations that follow galaxy formation in  $f(R)$  modified gravity models. Our simulations employ the Illustris-TNG full physics model and a new modified gravity solver in the AREPO code. For the first time we are able to investigate the degeneracy in the matter power spectrum between the effects of  $f(R)$ -gravity and feedback from active galactic nuclei (AGN), and the imprint of modified gravity on the properties of galaxies and on the distribution of dark matter, gas and stars in the universe.  $f(R)$ -gravity has an observable effect on the neutral hydrogen power spectrum at high redshift at a level of 20%. For both the F6 and F5 models, this is significantly larger than the predicted errors for the SKA1-MID survey, making this probe a powerful test of gravity on large scales. A similar effect is present in the power spectrum of the stars at high redshift. We also show that rotationally supported disc galaxies can form in  $f(R)$ -gravity, even in the partially screened regime. Our simulations indicate that there might be more disc galaxies in F6 compared to GR, and fewer in F5. Finally, we show that the back reaction between AGN feedback and modified gravity in the matter power spectrum is not important in the F6 model but has a sizeable effect in F5.

The physical nature of gravity and dark energy are central unsolved problems in modern physics. Einstein's General Relativity (GR) is generally accepted as the standard theory of gravity. GR has been empirically verified, to remarkably high precision, on small scales, but on cosmological scales the available tests are still not strongly constraining. Planned astronomical surveys will deliver data capable of distinguishing it from alternative theories of gravity. However, predictions for these observables require accurate simulations of the formation of cosmic structure in GR and alternatives, which must include the interactions between baryons and dark matter. Here, we present a set of cosmological hydrodynamical simulations of an attractive alternative to GR, the Hu & Sawicki  $f(R)$ -gravity, which is representative of the large class of chameleon-type screening modified gravity models. These simulations are the first to include galaxy formation processes, such as feedback from supernovae and super-massive black holes, which can affect observables at similar levels to the modifications to GR. We find that it is possible to form rotationally supported disk galaxies in  $f(R)$ -gravity and that back-reaction effects on the matter power spectrum are small. We also find that the stellar and neutral hydrogen power spectra are suppressed in  $f(R)$ -gravity to a level that is detectable with future neutral hydrogen surveys.

The results presented in this work were derived from the SHYBONE (Simulating HYdrodynamics BeyOND Einstein) simulations, a set of full-physics hydrodynamical simulations employing the Illustris-TNG model in Hu-Sawicki  $f(R)$ -gravity [1, 2], which is one of the most widely-studied models of modified gravity. The theory introduces an extra scalar degree of freedom which mediates a fifth force between matter

particles. This enhances standard gravity by up to 4/3 in low-density environments, while dense regions are screened from such enhancements via the chameleon screening mechanism [3]. This way, the model can pass the stringent constraints on gravity in the solar system [4] and still leave detectable signatures on large scales, making it an excellent tool to explore how possible deviations from GR would be observable.

$f(R)$ -gravity predicts identical propagation speeds of gravitational waves and photons, and so is compatible with the recent gravitational wave detection with optical counterparts [5–8]. With appropriate choices of model parameters (see Methods), it can explain the late-time accelerated expansion of our Universe without an explicit cosmological constant [2] (Note that the value of  $\Omega_\Lambda$  still enters the theory as a parameter). In addition to being theoretically well understood [9, 10],  $f(R)$ -gravity has also been studied extensively in dark matter only (DM-only) [11–20] and non-radiative hydrodynamical cosmological simulations [21–24]. These works were nevertheless unable to include a suitably calibrated baryonic feedback model at the same time, which is nontrivial and computationally expensive.

Such simulations with feedback have recently become possible thanks to a modified AREPO [26] code, which employs a new and optimized method to solve the fully nonlinear  $f(R)$ -gravity equations in the quasi-static limit (see Methods). In this work, this more efficient code is combined with the Illustris-TNG galaxy formation model [27–31] which incorporates prescriptions for gas-hydrodynamics, magnetic fields, star and black hole formation, feedback from supernovae and AGN, gas heating and cooling processes, as well as galactic winds. This new code made the SHYBONE simulations possible, which are dedicated to in-depth studies of the interplay of baryonic physics and modified gravity for the first time. A detailed description of the simulations used in our analysis can be found in Methods.

\* christian.arnold@durham.ac.uk

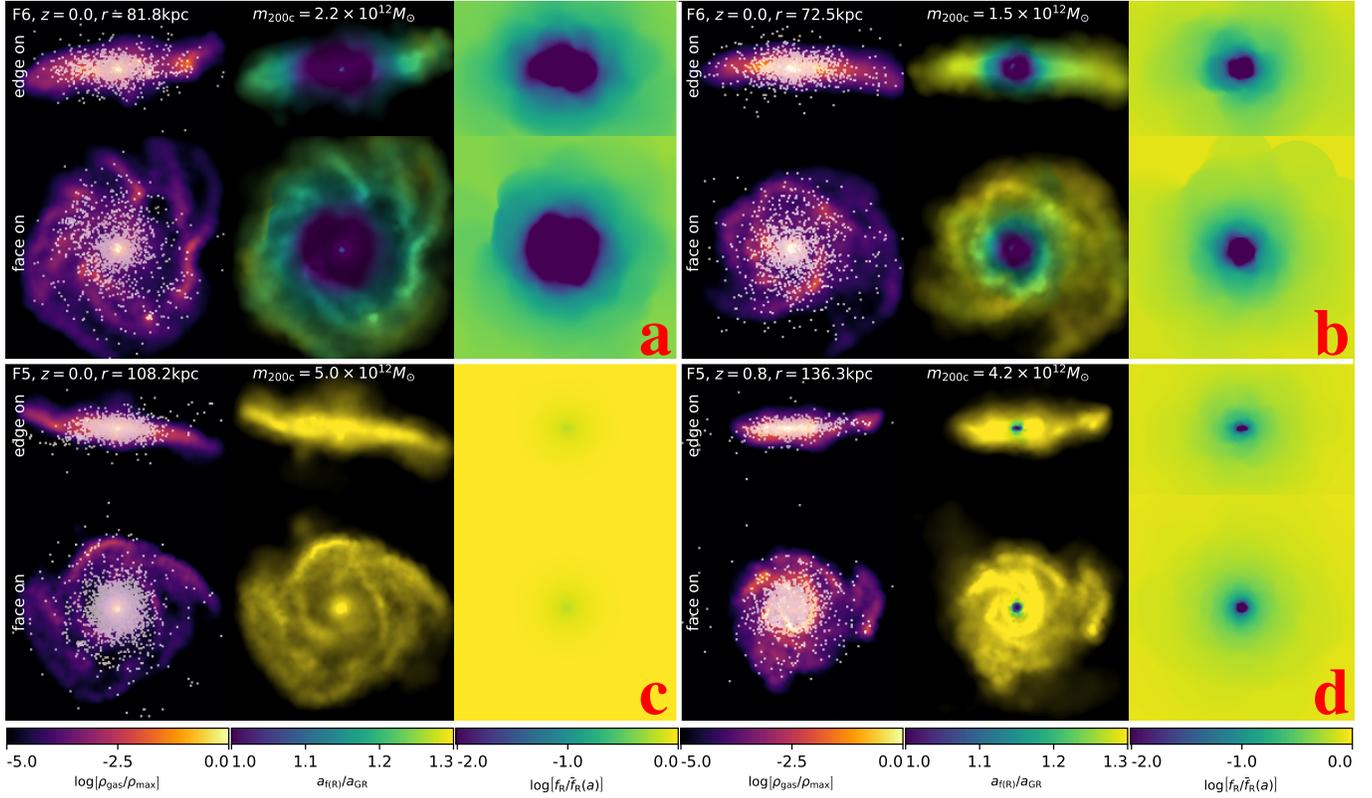


Figure 1. A selection of four galaxies from the F6 (top) and F5 (bottom) cosmology simulations at  $z = 0$  (top left, top right, and bottom left) and  $z = 0.8$  (bottom right). For each galaxy, the top panels show an edge-on view where the angular momentum of the stars is pointing up; the lower panels show a face on view. The left panels show the gas column density for the individual galaxies combined with the stars within the objects. The middle panels show the gas density (brightness) color-coded with the modified gravity forces within a thin slice through the center of the galaxies. Yellow coloured gas cells experience an enhanced total force  $F_{\text{tot}} = 4/3 F_{\text{GR}}$  while the dark blue regions do not experience a force enhancement due to modified gravity. The scalar field to background field ratio  $f_R/\bar{f}_R(a)$  (see methods) within the same slice through the central region is shown in the right hand panels. Again, yellow regions are unscreened with  $f_R \approx \bar{f}_R(a)$ . The dark blue regions are fully screened  $f_R < 10^{-2} \bar{f}_R(a)$ . We quote model, redshift, size of the plotted region from the center of the object and the total mass within  $r_{200\text{crit}}$  in the plots for each of the galaxies. Note that rotationally supported disk galaxies can form in  $f(R)$ -gravity despite the complicated force morphology in the partially screened regime.

## I. RESULTS

### A. Galaxies in $f(R)$ -gravity

Figure 1 shows a selection of galaxies from our full-physics simulations for the F6 (**a** and **b**) and F5 model (**c** and **d**) in the (total) mass range  $1.5 \times 10^{12} < M_{200c}/M_\odot < 5 \times 10^{12}$ . Three of the objects are at  $z = 0$  (**a-c**), while the galaxy in Panel **d** is at  $z = 0.8$  (only one disk galaxy is identified at  $z = 0$  for F5). The galaxies were selected to have a rotationally supported stellar disk, i.e., a rotational-to-total kinetic energy parameter  $\kappa = \sum (j_z/r)^2 / \sum v^2 > 0.57$  (where the  $z$ -direction is defined by the total angular momentum of the stars in the galaxy) [33].

It is evident from the central sub-panels of Figure 1 that both galaxies shown for the F6 model are partially screened. The gas in the inner regions is screened and experiences GR-like forces ( $a_{f(R)}/a_{\text{GR}} \approx 1$ ), while the gravitational force experienced by the gas in the outer regions

is enhanced by  $4/3$  (bright yellow regions in the gas). The size of the central screened region depends on the total mass of the galaxy and its host halo, and the results imply that objects similar to the Milky Way Galaxy are partially screened for F6. In the F5 model, in contrast, the chameleon screening is less efficient, and the objects shown in the bottom panels show either no (**c**) or very small (**d**) screened regions in the central panels. Gas and stars will thus experience fully enhanced gravity across the whole object, implying that the model is unlikely to pass local tests of gravity.

Interestingly, this complicated force morphology in F6 does not hinder galaxies from forming rotationally-supported disks within the gas and the stars. We find that there are more galaxies with  $\kappa > 0.57$  in F6 compared to GR and significantly fewer in F5. The latter is an interesting observation that might be related to the stability of the disks in the unscreened regime or an enhanced galaxy merger rate. However, we caution that the number of disk galaxies is small in all of our simulations (7, 16, 1 objects at  $z = 0$  and 34, 48, 22 objects at  $z = 0.8$  for GR, F6 and F5, in the large box, respectively and 56, 64 for

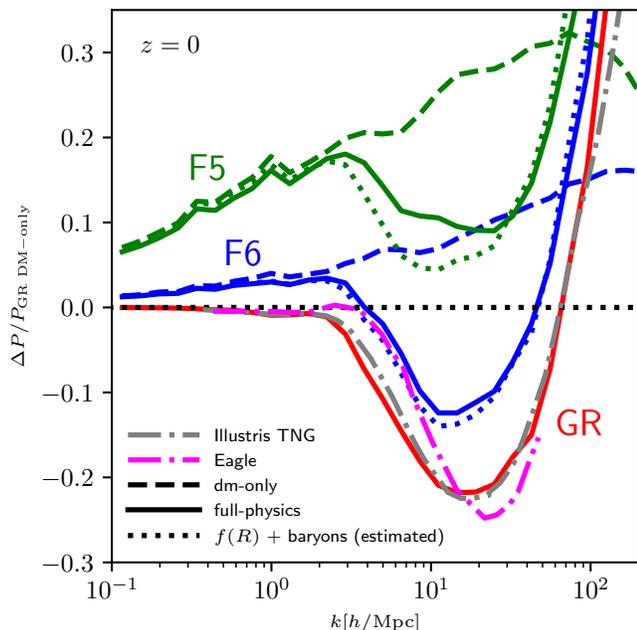


Figure 2. The relative difference of the matter power spectra with respect to the  $\Lambda$ CDM cosmology DM-only simulation for a  $\Lambda$ CDM universe (red), the F6 model (blue) and F5 (green). Dashed lines show relative differences for the DM-only  $f(R)$ -gravity runs, solid lines for the full-physics simulations using the same colours as above. An estimate for the combined effect of baryonic feedback and modified gravity (displayed as dotted lines) was obtained by adding the relative differences for the  $f(R)$ -gravity DM-only simulations to those from the GR full physics run. The gray and magenta dash-dotted lines indicate the impact of feedback on the matter power spectrum in the Illustris TNG [28] and Eagle [58] simulations, respectively. The horizontal black dotted line indicates equality.

GR and F6 at  $z = 1$  in the small box). The number counts of disk galaxies is therefore to be verified in future simulations.

### B. The matter power spectrum degeneracy

It is well-known that  $f(R)$ -gravity can cause enhancements in the matter power spectrum while baryonic effects, predominantly AGN feedback, act in the opposite direction [15, 32]. Previous simulation works were nevertheless unable to study both effects simultaneously and therefore could not model the possible back-reactions between them – a challenge that the full-physics simulations in this paper aim to tackle.

To inspect this degeneracy, in Figure 2 we plot the fractional changes of the total matter power spectrum with respect to results from the GR DM-only simulation. The dashed lines show the enhancement predicted by the  $f(R)$  DM-only simulations. Solid lines show the combined effects predicted by the full-physics  $f(R)$  simulations (for GR they show the impact of baryons only). The dotted lines are produced by simply adding up the relative changes due to modified gravity, as predicted by  $f(R)$  DM-only simulations, and due to AGN feedback, as predicted by the GR full-physics simulation.

As one can see from the plot, baryonic feedback suppresses power on scales  $k \gtrsim 2 h\text{Mpc}^{-1}$ , with the suppression reaching  $\sim 20\%$  at scale  $k \sim 10 h\text{Mpc}^{-1}$ . These results are in excellent agreement with previous works employing the same hydrodynamical model ([28]) and from the EAGLE simulation ([34]). Similar suppression at  $k \gtrsim 2 h\text{Mpc}^{-1}$  due to baryonic feedback can be seen for F6 and F5.

It is obvious from the plot that the effect of baryons on the power spectrum for the F6 model is negligible for  $k < 2 h\text{Mpc}^{-1}$ . The relative difference of the full-physics  $f(R)$ -gravity power spectrum is thus dominated by the modified gravity effect which increases power by a few percent at these scales. As for GR, the suppression due to baryons becomes active beyond  $k = 2 h\text{Mpc}^{-1}$ . Due to the  $f(R)$ -gravity effects, the minimum of the relative difference is nevertheless only 12% below the baseline. Interestingly, the blue dotted line in Figure 2 shows that the power spectrum of the full-physics F6 simulation can be predicted almost perfectly by an additive combination of the effects due to modified gravity and baryonic feedback. This suggests that there is negligible back-reaction between these effects for F6, and therefore computationally costly full-physics simulations for this model are not necessary as far as the matter power spectrum is concerned. This simplification does not hold for the F5 model, where the back-reaction effect is strong enough that a full-physics simulation is necessary if percent-accuracy predictions of the matter power spectrum are needed even at  $k \lesssim 1 h\text{Mpc}^{-1}$ . This will have important implications on the tests of this and similar models in surveys like Euclid.

The different strengths of interplay between modified gravity and AGN feedback in F6 and F5 can be explained by the chameleon screening mechanism. In the F6 model, the central regions of halos which are massive enough to carry AGN are screened, so that the flow of matter onto the central black hole (the AGN accretion) is not altered by modified gravity, resulting in similar AGN-feedback effects on the power spectrum as in GR. For F5, on the other hand, many of the AGN-hosting halos have become completely unscreened. The flow of matter towards central black holes therefore takes place in environments with by  $4/3$  enhanced gravity, leading to modified accretion and feedback efficiency, and in turn a non-negligible back-reaction between the modified gravity and feedback effects in the matter power spectrum (for details on the different types of AGN feedback see [35]).

### C. Power spectra of individual components

The effect on the total matter power spectrum seen above is a combination of effects on the different matter components. In Figure 3 we display the power spectra of these individual components for  $z = 2, 1$  and 0 (Panel a, b and c, respectively). The absolute spectra, which we find to agree with the original Illustris-TNG simulation [28], are displayed in the upper panels, the relative differences relative to GR in the lower panels.

The components react in very different ways to modified gravity. The dark matter distribution is to first order only affected by gravity, its power spectrum is thus enhanced in a

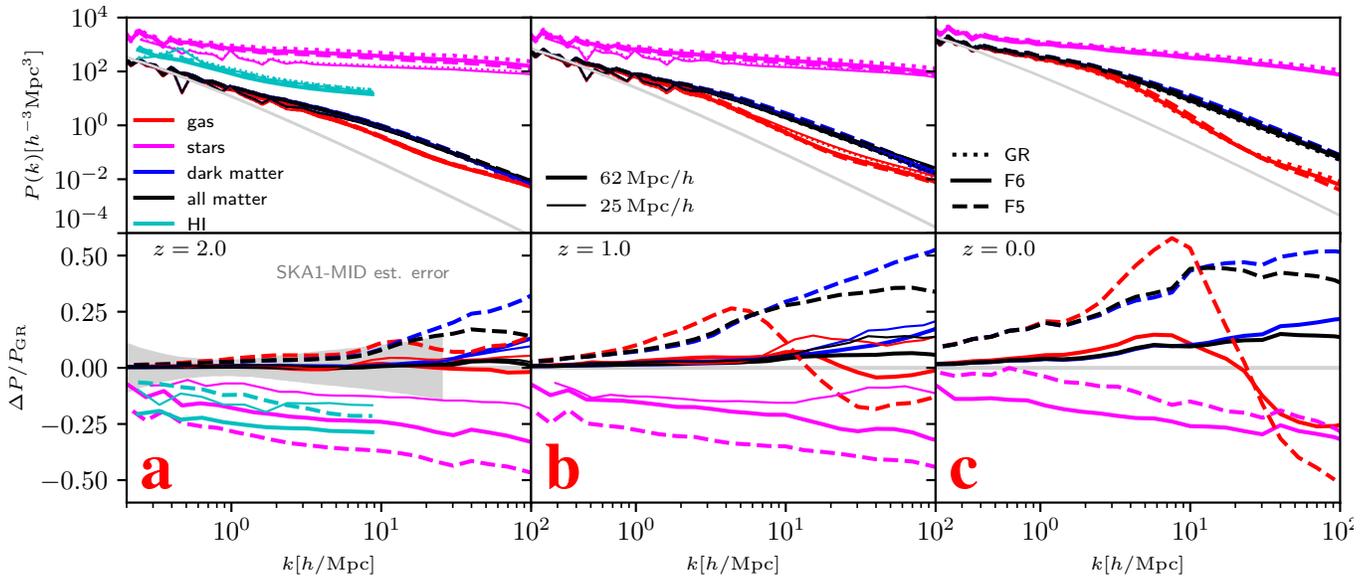


Figure 3. The 3D matter power spectrum of the different matter components in our simulations for redshift  $z = 2$  (a),  $z = 1$  (b) and  $z = 0$  (c). GR results are shown as dotted lines, results from our F6 and F5 simulations as solid and dashed lines, respectively. Absolute values are shown in the upper panels while the relative differences are indicated in the lower panels for each redshift. Thick lines indicate results from the  $62 h^{-1} \text{Mpc}$  simulation box, thin lines results from the  $25 h^{-1} \text{Mpc}$  box (these are only shown for  $z = 2$  and  $z = 1$ ). The power spectrum of gas is displayed in red, stars in magenta, dark matter in blue and the total matter spectrum in black. For  $z = 2$  and  $k < 10 h \text{Mpc}^{-1}$ , we show the neutral hydrogen power spectrum in cyan as well and compare it to predicted errors for SKA1-MID measurements of the HI power spectrum (gray shaded region; [38]). In the upper panels, the grey lines indicate the linearly evolved initial power spectrum. The horizontal grey lines in the lower panels indicate equality.

similar way as in DM-only simulations [e.g., 25, 32]. The dark matter and total matter power spectra are converged at  $k \lesssim 20 h \text{Mpc}^{-1}$  between the  $25 h^{-1} \text{Mpc}$  and  $62 h^{-1} \text{Mpc}$  boxes, suggesting that the results in Figure 2 can be trusted in this  $k$ -range.

The gas power spectrum is only mildly influenced by modified gravity at  $z = 2$ , and the effects are stronger at lower  $z$ . In particular, the gas is more clustered in F6 and F5 than in GR at intermediate ( $k < 10 h \text{Mpc}^{-1}$ ) scales. These effects are caused by the interplay of a number of processes that we only attempt to explain to first order here: the enhancement in gas clustering on intermediate scales is mainly a result of higher gas densities in the unscreened parts of galaxies – as Figure 1 shows, this will be primarily the case in low-mass unscreened objects or the outer regions of intermediate-mass objects. Comparing the enhancements in the gas power spectrum for F6 at  $z = 1$  for the small and the large simulation boxes, we see that the results are consistent at few-percent level for  $k < 10 h \text{Mpc}^{-1}$ .

Unlike dark matter and gas, the clustering of stars is more strongly effected by modified gravity at high redshift. At  $z = 2$ , the stellar power spectrum from the  $62 h^{-1} \text{Mpc}$  box is suppressed by  $\sim 20\text{-}25\%$  in F6 and  $\sim 30\text{-}40\%$  in F5 on scales  $k \sim 1\text{-}10 h \text{Mpc}^{-1}$  (thick magenta lines). The predicted suppression for F6 from the  $25 h^{-1} \text{Mpc}$  simulation (thin magenta line) is weaker, which is  $\sim 10\text{-}15\%$  in the same  $k$  range, indicating that the clustering of stars has not converged in the big box at this redshift. At  $z = 1$ , the two simulation boxes agree to a few percent for  $k < 5 h \text{Mpc}^{-1}$ .

A main reason for the suppression of the stellar power spectrum is the gas density in low-mass objects: as smaller halos are already unscreened at high redshift, the gas within these objects is denser and can cool more effectively, compared to the similar objects in GR. This leads to enhanced star formation inside small halos. We have checked that the fraction of  $m = 10^{10} M_{\odot}$  halos carrying stars is in fact 60% higher in F6 and 100% higher in F5 compared to GR at  $z = 2$ . These halos form from lower initial density peaks, which are intrinsically less clustered. This, together with the larger numbers of star-forming halos at high redshift in  $f(R)$ -gravity compared to GR, lead to a suppression of the stellar power spectrum in the former. As this process does not involve stellar dynamics, we expect it to be unaffected by potential self-screening of stars (which is not taken into account in our simulations).

#### D. The neutral hydrogen power spectrum

The enhanced gas cooling efficiency in low-mass objects in  $f(R)$  gravity at early times also affects the distribution of neutral hydrogen (HI), a large fraction of which resides in small halos. We quantify this using the HI power spectrum, calculated following the procedure outlined in [37]. The results from the  $25 h^{-1} \text{Mpc}$  boxes are plotted in Figure 3(a), which shows that at  $z = 2$  the HI power spectrum on scales  $k \sim 0.02\text{-}10 h \text{Mpc}^{-1}$  is suppressed by  $\sim 15\%$  for F6 (thin cyan line). The  $62 h^{-1} \text{Mpc}$  simulation boxes have not yet reached convergence as they cannot resolve small halos

( $m \lesssim 10^{10} M_{\odot}$ ) which host a significant fraction of the total HI. The small box, on the other hand, resolves halos down to  $m \sim 3 \times 10^8 M_{\odot}$ , which host the vast majority of HI [37]. For both box sizes we checked that the relative suppression of the HI power spectrum in  $f(R)$  gravity is almost identical to the suppression of the power spectrum of HI-hosting halos at  $k \lesssim 2 h\text{Mpc}^{-1}$ , showing that it is more sensitive to the clustering of halos than to baryonic feedback. Comparing our results to the expected uncertainty for an intensity mapping experiment similar to SKA1-MID (a future 21 cm intensity mapping survey, [38]), we find that with 1000 hours of observing time, SKA1-MID should be able to distinguish F6 and GR for  $z = 2$ . This is different from most other cosmological tests of this model, which are usually at lower redshifts.

Also interestingly, at  $z = 2$  the HI power spectrum for F5 is suppressed less than F6, which is because at this redshift F6 has a larger enhancement of the number of small halos while in F5 such small halos have merged to form larger ones. We checked that the F5 HI power spectrum is suppressed more at even higher redshifts.

While both the HI fraction and star formation rate in low-mass objects are sensitive to (stellar) wind feedback, we expect the impact by  $f(R)$ -gravity to be relatively robust against changes in the feedback model. This is because such changes cannot be arbitrary as that would otherwise lead to tensions between the simulations and observational data for the stellar observables (see Figure B1 in [35] for a detailed discussion on how different wind feedback implementations affect the stellar properties of galaxies at  $z = 0$ ). It is also known [39] that details in wind feedback are important in the strong Damped Lyman- $\alpha$  regime where HI column density is  $> 10^{21.5} \text{cm}^{-2}$ , while we have checked explicitly that modified gravity effects are non-negligible only at column densities  $\lesssim 10^{21.2} \text{cm}^{-2}$ .

### E. Galaxy properties

The SHYBONE simulations use the same hydrodynamical model for all three gravity theories. To check that a re-tuning of this model is not necessary for  $f(R)$  gravity, in Figure 4 we plot a few stellar observables which were used in the tuning of the Illustris-TNG model. Panel **a** displays the stellar mass fraction (SMF) at the centres of galaxies, as a function of the host halo mass from the  $62 h^{-1} \text{Mpc}$  box. We find that  $f(R)$ -gravity introduces a  $\lesssim 25\%$  change with respect to GR for both F6 and F5, which is smaller than the  $1\sigma$  errors of the observational data [40] shown for comparison. The absolute value of the SMF in our simulations does – as expected – not match the observations: The TNG model was tuned for a  $25 h^{-1} \text{Mpc}$  box with  $2 \times 512^3$  initial resolution elements, while simulations at lower resolution show lower star formation rates [35, see also the cyan lines in Panel **d**]. For comparison, we also show the SMF from TNG simulations at several resolutions (the cyan lines in Panel **a**). The mass resolution of our  $62 h^{-1} \text{Mpc}$  box is close to that in the  $25 h^{-1} \text{Mpc}$  TNG box with  $2 \times 128^3$  resolution elements, and so are the SMF results. Supplementary Material Figure 1 further confirms that the SMF from our  $25 h^{-1} \text{Mpc}$  box, which has the same reso-

lution as the highest-resolution TNG box, agrees with observational data at  $z = 1$ .

The galaxy stellar mass functions measured within 30 kpc from the halo centres are shown in panel **b**. Again, the results from the  $62 h^{-1} \text{Mpc}$  box fall in between the TNG runs with  $2 \times 128^3$  and  $2 \times 256^3$  resolution elements, due to the reduced star formation rate in low-resolution simulations. The relative differences between  $f(R)$ -gravity and GR are smaller than the uncertainties in the observational data (grey lines).

Panel **c** shows the galaxy gas fraction as a function of the host halo mass. Unlike the stellar properties, the gas fraction is not resolution dependent. The results from the simulations presented in this work are therefore consistent with the three  $25 h^{-1} \text{Mpc}$  TNG boxes. The differences between the gravity models are again much smaller than the spread in observational data from [44, 45] at  $M_{200c} \gtrsim 10^{13} M_{\odot}$ .

Finally, the cosmic star formation rate densities (SFRD) as a function of redshift are shown in panel **d** for the  $62 h^{-1} \text{Mpc}$  and  $25 h^{-1} \text{Mpc}$  (for  $z > 1$ ) boxes. The results from the three TNG boxes confirm that the SFRD is resolution dependent. Our  $25 h^{-1} \text{Mpc}$  simulations match the observational data of [40], and are comparable to the TNG test-box with the same resolution. The  $62 h^{-1} \text{Mpc}$  boxes, as expected, show a lower SFRD which falls between the two TNG test-boxes with  $2 \times 128^3$  and  $2 \times 256^3$  resolution elements. The relative difference between F6 and GR is nevertheless converged.

The effects of  $f(R)$ -gravity on galactic properties are the consequence of the interplay of different physical processes. The star formation rate depends on gas density and temperature, with an enhanced gravity resulting in higher gas densities within galaxies and consequently a higher SFRD, as shown in Figure 4 **d**. In particular, at high redshift the star formation rates in  $f(R)$ -gravity and GR are roughly the same because most objects are screened; towards lower redshifts, galaxies gradually become unscreened and the SFRD in the two  $f(R)$  models is enhanced with respect to GR. Due to the less efficient chameleon screening, in F5 this process happens earlier than in F6. At even later times, the relative enhancement of the SFRD in  $f(R)$ -gravity starts to decrease and becomes negative for F5 at  $z \lesssim 0.7$ . A reason for this is that the cold dense gas gets consumed faster in the  $f(R)$  models, leaving less material for star formation at later times. The increased SFRD causes a higher stellar mass function (Panel **b**), as well as making a larger number of small halos star forming, explaining the weaker clustering of stars.

In addition to the four stellar and gas properties shown in Figure 4, we have also checked our simulation predictions of the galaxy size and black hole mass, which were used to calibrate the Illustris-TNG baryonic model too. In all these cases, the relative difference induced by  $f(R)$ -gravity does not lead to tensions with observational data, and so we conclude that a re-tuning of the baryonic model is not necessary.

## II. DISCUSSION

Studying galaxy formation and its application to test gravity models with full-physics hydrodynamical simulations used to

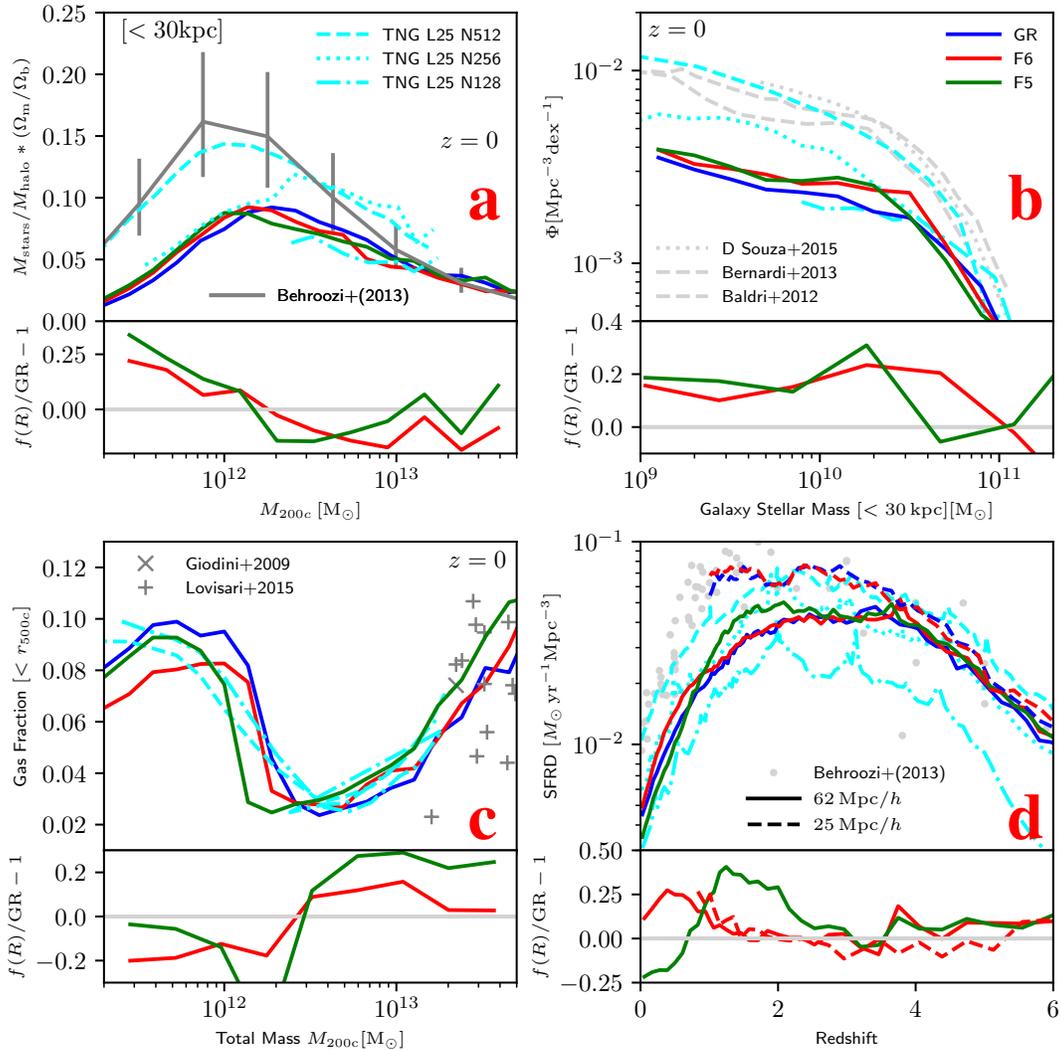


Figure 4. The stellar and gaseous properties of galaxies. Results from the simulations presented in this work are shown as blue lines for GR, red lines for F6 and green lines for F5 in the upper section for each of the four panels. The lower sections show the relative difference between  $f(R)$ -gravity and GR. The cyan lines show the results from [35] for the  $25 h^{-1} \text{Mpc}$  side-length Illustris-TNG test boxes with  $2 \times 128^3$  (dash-dotted lines),  $2 \times 256^3$  (dotted lines) and  $2 \times 512^3$  (dashed lines) initial resolution elements. Observational constraints from [40–45] are shown as grey lines and symbols. **a:** The galaxy stellar mass fraction within 30kpc from the halo centre as a function of halo mass in the  $62 h^{-1} \text{Mpc}$  box. **b:** The galaxy stellar mass function measured within 30kpc from the halo centre in the  $62 h^{-1} \text{Mpc}$  box. **c:** The galaxy gas fraction within  $r_{500c}$  as a function of total halo mass in the  $62 h^{-1} \text{Mpc}$  box. **d:** The cosmic SFRD as a function of redshift for both the  $25 h^{-1} \text{Mpc}$  (dashed lines, for  $z > 1$ ) and the  $62 h^{-1} \text{Mpc}$  (solid lines) simulation boxes.

be a formidable task, because of the large computational cost and the complexities in calibrating a realistic baryonic model. There were also doubts regarding whether meaningful cosmological signals can be cleanly extracted given the uncertainty in the subgrid physics model. In this work, we present a new suite of high-resolution simulations of  $f(R)$ -gravity which incorporate a large number of baryonic processes to reproduce a realistic population of galaxies as implemented in the Illustris-TNG model. While this model has been applied in  $f(R)$  simulations without changes, we have checked various galaxy gas and stellar properties, and found our simulation predictions in  $f(R)$ -gravity to agree with observations and the TNG simu-

lations of similar resolutions, so that a re-tuning is not necessary.

A key question addressed here is the degeneracy between baryonic feedback and modified gravity, which has important implications for physical quantities, such as the matter power spectrum, that are central to future galaxy surveys such as Euclid. We find that in weak models like F6 these two effects can be modelled separately and combined additively, with better-than-percent accuracy in the predicted matter power spectrum at  $k \lesssim 1 h \text{Mpc}^{-1}$ , while for stronger models like F5 there is a back-reaction between them, necessitating costly full-physics simulations if sub-percent accuracy is desired.

Despite the complicated screening profile in them, we find rotationally-supported disk galaxies can form even in partially screened halos. There is an indication that strong  $f(R)$  models such as F5 produce fewer disk galaxies than GR, which can be caused by the more frequent galaxy mergers in the former; if this is confirmed by future larger simulations, it may offer a possible test that is insensitive to baryonic physics (which has little impact on the merger rate). More interestingly, we find that the enhanced number of small dark matter halos in  $f(R)$ -gravity can significantly alter the spatial distributions of cool dense gas, stars and neutral hydrogen, and as a result the latter can be used as tracers for the clustering of small halos, which may offer useful model tests with future observations such as 21cm intensity mapping.

As a final remark, we note that though the resolution of the SHYBONE simulations is unprecedented in the case of  $f(R)$ -gravity, their relatively small boxes inevitably place a limit on a statistical analysis due to cosmic variance. Further, while the results look promising, in some cases, e.g., neutral hydrogen, research is still scarce even in GR, and in this sense we hope that our results will serve as a reality check and a call for more detailed studies in both GR and alternative models.

### III. METHODS

The results presented in this work were derived from a suite of cosmological hydrodynamical simulations of  $f(R)$ -gravity which, for the first time, allow to study the interplay between detailed baryonic feedback and modified gravity in the same calculation. The simulations were carried out with the AREPO [26] code, by combining the Illustris-TNG galaxy formation model [27–31] and a new, much more efficient, modified gravity solver in the code. Below we give an overview of the TNG model and  $f(R)$ -gravity, and introduce the new modified gravity solver. Further details about the code verification tests are included in the Supplementary Material.

The SHYBONE suite mainly consist of six simulations carried out for  $\Lambda$ CDM, F6 and F5 universes using the same initial condition. For each model, we carried out a full-physics hydrodynamical simulation and an associated DM-only run. These simulations initially contain  $512^3$  dark matter particles and the same number of gas cells (for the hydrodynamical simulations only) in a  $62 h^{-1} \text{Mpc}$ -a-side box, with mass resolutions  $m_{\text{DM}} = 1.28 \times 10^8 h^{-1} M_{\odot}$  and  $m_{\text{gas}} \approx 2.5 \times 10^7 h^{-1} M_{\odot}$  (the mass within the gas cells can vary by a factor of 2). For convergence test, two additional full-physics simulations (for F6 and GR) employing the same number of resolution elements in a  $25 h^{-1} \text{Mpc}$  box, with  $m_{\text{DM}} = 8.39 \times 10^6 h^{-1} M_{\odot}$  and  $m_{\text{gas}} \approx 1.6 \times 10^6 h^{-1} M_{\odot}$ , were run until  $z = 1$ . All simulation use the same starting redshift,  $z = 127$ . The softening lengths for DM particles and stars are 0.5 and  $1.25 h^{-1} \text{kpc}$  for the small and large boxes, respectively. The total run time is  $\sim 2.5$  million core hours, and all simulations use Planck 2016 [46] cosmological parameters  $\sigma_8 = 0.8159$ ,  $\Omega_m = 0.3089$ ,  $\Omega_B = 0.0486$ ,  $\Omega_{\Lambda} = 0.6911$ ,  $h = 0.6774$  and  $n_s = 0.9667$ .

Unless stated otherwise, we measure total masses of halos

in terms of  $m_{200c}$  which is the mass within a sphere enclosing 200 times the critical density of the Universe.

#### A. Baryonic physics and feedback

The Illustris TNG model, a successor of the Illustris galaxy formation model [47], incorporates prescriptions for a variety of physical processes necessary to reproduce a realistic galaxy population in cosmological simulations. The TNG model consists of a magneto-hydrodynamics solver on a moving mesh, a recipe for stellar formation, evolution and feedback, a prescription for black hole growth and AGN feedback [49], as well as an algorithm for the chemical enrichment and cooling of the gas. It has been tuned to reproduce observational results on the stellar mass function, cosmic star formation rate density (SFRD), the relation between black hole mass and galaxy stellar mass, galaxy gas fraction, stellar mass fraction and the size of galaxies, in  $\Lambda$ CDM simulations [35]. For a more detailed overview of the TNG model, see [27, 48].

For the simulations carried out for this project, we applied the same TNG model to all gravity theories considered. We are primarily interested in the question how astrophysical and cosmological observables in baryonic simulations are affected by modifications to gravity given a fixed model of galaxy formation. We explicitly checked that the changes to the observational gas and stellar properties against which the TNG model has been tuned are small compared to the uncertainties in the observations, so that a re-tuning is not necessary.

#### B. $f(R)$ -gravity

$f(R)$ -gravity [1] is an extension of Einstein's general relativity (GR) including an additional, scalar degree of freedom which induces an enhancement to the gravitational force in low-density environments. In dense regions like the Solar system, the model employs the chameleon screening mechanism [3] to screen the modifications to gravity and recover GR-like behaviour.

The theory is constructed by adding a scalar function  $f(R)$  of the Ricci scalar  $R$  to the Einstein-Hilbert action of GR:

$$S = \int d^4x \sqrt{-g} \left[ \frac{R + f(R)}{16\pi G} + \mathcal{L}_m \right], \quad (1)$$

where  $G$  is the gravitational constant,  $g$  is the determinant of the metric  $g_{\mu\nu}$  and  $\mathcal{L}_m$  is the matter Lagrangian density. Varying this action with respect to the metric leads to the Modified Einstein equations:

$$G_{\mu\nu} + f_{\text{R}} R_{\mu\nu} - \left( \frac{f}{2} - \square f_{\text{R}} \right) g_{\mu\nu} - \nabla_{\mu} \nabla_{\nu} f_{\text{R}} = 8\pi G T_{\mu\nu}, \quad (2)$$

where  $G_{\mu\nu}$  and  $R_{\mu\nu}$  denote the Einstein and Ricci tensor, respectively,  $T_{\mu\nu}$  the energy momentum tensor and  $\square \equiv$

$\nabla_\nu \nabla^\nu$ , where Einstein summation is used. The derivative of the scalar function,  $f_R \equiv df(R)/dR$ , becomes a new degree of freedom whose dynamics is determined by the trace of the modified Einstein equation.

In cosmological simulations one often works in the weak-field and quasi-static limit (see [50] for a discussion of its validity for  $f(R)$ -gravity), where Eq. (2) simplifies considerably: similar to the Newtonian limit of GR, the gravitational potential  $\Phi$  is given by a modified Poisson equation,

$$\nabla^2 \Phi = \frac{16\pi G}{3} \delta\rho - \frac{1}{6} \delta R, \quad (3)$$

where  $\delta\rho$  is the matter density perturbation, and  $\delta R = R - \bar{R}$  is the perturbation to the Ricci scalar  $R(f_R)$ , with  $f_R$  satisfying the following differential equation,

$$\nabla^2 f_R = \frac{1}{3} (\delta R - 8\pi G \delta\rho). \quad (4)$$

Throughout this work we adopt the model proposed by [2]:

$$f(R) = -m^2 \frac{c_1 \left(\frac{R}{m^2}\right)^n}{c_2 \left(\frac{R}{m^2}\right)^n + 1}, \quad (5)$$

$c$  denotes the speed of light here and  $m$  is a mass scale of the model given by  $m^2 \equiv \Omega_m H_0^2$ , with  $H_0$  being the Hubble constant, and  $c_1, c_2$  are model parameters. We choose  $n = 1$  in this work (This is somewhat arbitrary but the most convenient choice for the computations and widely used in simulations of  $f(R)$ -gravity). In addition to its relatively simple functional form, this model has two further advantages. First, it allows the so-called chameleon screening mechanism to suppress the modifications to GR in high-density environments, which is necessary to pass the very tight constraints on gravity within the Solar system [4]. Second, the model features a cosmic expansion history which is very close to that of a  $\Lambda$ CDM universe if one chooses [2]

$$\frac{c_1}{c_2} = 6 \frac{\Omega_\Lambda}{\Omega_m} \quad \text{and} \quad \frac{c_2 R}{m^2} \gg 1. \quad (6)$$

Given the functional form in Eq. (5), under the conditions of Eq. (6) the scalar field  $f_R$  can be approximated as

$$f_R \equiv \frac{df(R)}{dR} \approx -\frac{c_1}{c_2} \left(\frac{m^2}{R}\right)^2, \quad (7)$$

and consequently  $R$  can be written as a function of  $f_R$ . Instead of  $m, c_1, c_2$ , the theory can be more conveniently described by the parameters  $\Omega_m, \Omega_\Lambda$  and  $\bar{f}_{R0}$ , with  $\bar{f}_{R0}$  having the physical meaning of the present-day value of the background scalar field. With these parameters specified, the scalar field at arbitrary time can be expressed as

$$\bar{f}_R(a) = \bar{f}_{R0} \left[ \frac{\bar{R}_0}{R(a)} \right]^2, \quad (8)$$

where  $R_0$  is the value of the Ricci scalar today and  $R(a)$  is its value at scale factor  $a$ ,

$$\bar{R} = 3m^2 \left[ a^{-3} + 4 \frac{\Omega_\Lambda}{\Omega_m} \right]. \quad (9)$$

The threshold for the onset of chameleon screening is set by  $\bar{f}_{R0}$ . Current astrophysical constraints on the theory limit the background scalar field to values smaller than  $|\bar{f}_{R0}| = 10^{-6}$  (F6); cosmological constraints are weaker [51] ([52] find that the motions of galaxies in F6 predicted by DM-only simulations and sub-halo abundance matching (SHAM) are in tension with SDSS data. The authors of this paper nevertheless assume, that the same SHAM model can be applied to GR and  $f(R)$ -gravity if the mass of halos in the DM-only simulation, which is used as a parameter of the SHAM model in GR, is replaced by the dynamical mass for  $f(R)$ -gravity. The dynamical or effective mass is the mass a massive test particle would feel under the assumption of standard gravity. This mass measure accounts for the by 4/3 enhanced gravitational forces in unscreened objects and is the same as the true mass in screened objects. This assumption is to be validated against full-physics MG simulations to assess the robustness of the constraint). We also consider a model with  $|\bar{f}_{R0}| = 10^{-5}$  (F5) within this work as a toy model which features more prominent effects due to modified gravity in the simulations but is in tension with observations [51].

### C. A new modified gravity solver in AREPO

AREPO [26] is a massively-parallel, highly-optimized, cosmological simulation code. It employs a second-order Riemann solver using a moving Voronoi mesh to solve the Euler equations of hydrodynamics. The code also features state-of-the-art prescriptions of star formation, feedback from supernovae and active galactic nuclei, and magneto-hydrodynamics [35, 47, 49].

The original gravity solver of AREPO is based on the solver in P-GADGET3, which is itself based on GADGET2 [53]. The solver employs an oct-tree algorithm to obtain the short-range gravitational forces while the long-range forces are calculated using a particle-mesh algorithm to reduce the communication load when running in parallel. AREPO's gravity solver has recently been improved and extended in several ways but for the implementation of the modified gravity solver we (currently) stick to the original version.

Because of the similarity between the gravity algorithms in P-GADGET3 and AREPO, we use the modified gravity solver in MG-GADGET ([15], also based on P-GADGET3) as the starting point for the implementation of the  $f(R)$  gravity solver in AREPO. As the first step, we ported the solver from GADGET to AREPO, carrying out the necessary changes to adapt it to the code structure of AREPO and optimizing the data structures of the solver to reduce its memory footprint. We will refer to this implementation as the 'old' method in the following. In order to improve the convergence behaviour of the multigrid solver, we implemented a 'new' method to solve for the scalar field

based on the  $f(R)$  gravity solver in ECOSMOG [54] developed by [55].

In order to ensure that the new  $f(R)$  solver implemented in the code gives the correct result, we performed extensive comparison tests between the old and the new methods, as well as convergence tests for the MG-timestep criterion and the residual threshold on the individual grid levels. A selection of these tests are presented in the Supplementary Material.

### 1. AMR-grid construction and mass assignment

In order to solve for the gravitational forces one has to solve Eq. (4) first before the scalar field can be used to calculate the gravitational potential through Eq. (3). Due to the chameleon screening mechanism, which causes the scalar field value to drop by many orders of magnitude in dense environments with respect to the background value  $\bar{f}_R(a)$ , Eq. (4) is highly non-linear. The commonly-adopted method in cosmological simulation codes is therefore to solve the equation iteratively on a grid [15, 54, 56]. All modified gravity codes optimized for cosmological simulations employ an adaptively refining mesh (AMR-grid) for this purpose to allow for high spatial resolution in high-density regions while keeping the total computational cost down to a realistic level [25].

In AREPO we use an AMR-grid to solve for  $f_R$  as well. This is constructed from the gravitational oct-tree of the standard gravity solver in the same way as in MG-GADGET. Up to a maximum level, each tree node represents a mesh cell. In order to avoid a mesh that is too sparse on the finer AMR levels, all eight daughter tree nodes are created whenever a father node contains more than one simulation particle (this differs from the standard tree construction in AREPO; for more details see [15]). Because the tree is constructed from scratch for every time step in AREPO, there is no need to take care of tree-updates for this implementation. It is nevertheless necessary to construct the tree and solve for  $f_R$  before computing the long range gravity forces on the PM-grid as opposed to the standard gravity implementation in the code, where the tree is stored during the short-range force calculation only [26].

As in MG-GADGET, the particle masses are assigned to the mesh using a cloud-in-cell (CIC) assignment algorithm allowing to calculate the density contrast field for each mesh cell. The following step, to actually solve for the scalar field  $f_R$ , is where the old and new methods differ.

### 2. Scalar field solver

In the old method, Eq. (4) is rewritten in the following form by defining a new variable  $u = \ln(f_R/\bar{f}_R(a))$ ,

$$\nabla^2 e^u = -\frac{1}{3c^2 \bar{f}_R(a)} [\bar{R}(a) (1 - e^{-\frac{u}{2}}) + 8\pi G \delta\rho], \quad (10)$$

which is solved using the Newton-Raphson relaxation scheme with red black sweeps and multigrid acceleration (see [15] for more detail). This way unphysical positive values of

$f_R$  which might occur due to numerical errors during the iterations can be avoided.

The new method makes use of a different re-parametrisation of  $f_R$ . Following [55], we define

$$u \equiv \sqrt{\frac{f_R}{\bar{f}_R(a)}}, \quad (11)$$

which prevents positive values for  $f_R$  as well. This leads to a different form of the scalar field equation in terms of  $u$ ,

$$\nabla^2 (u^2) = \frac{1}{3} \left[ \frac{\bar{R}(a)}{\bar{f}_R(a)} \left( \frac{1}{u} - 1 \right) - \frac{8\pi G}{\bar{f}_R(a)} \delta\rho \right]. \quad (12)$$

This equation is then discretised on the AMR-grid constructed before. On a mesh of cell side-length  $l$ , the Laplace operator becomes

$$\begin{aligned} & \nabla^2 (u_{i,j,k}^2) \\ &= \frac{1}{l^2} [(u_{i+1,j,k}^2 + u_{i-1,j,k}^2 + u_{i,j+1,k}^2 + \dots) - 6u_{i,j,k}^2] \\ &\equiv \frac{1}{l^2} [L_{i,j,k}(u^2) - 6u_{i,j,k}^2], \end{aligned} \quad (13)$$

where we defined the operator  $L$ . Further defining

$$\alpha \equiv \frac{\bar{R}(a)}{18\bar{f}_R(a)} \quad \beta \equiv \frac{8\pi G}{18\bar{f}_R(a)}, \quad (14)$$

and

$$p \equiv -\frac{L_{i,j,k}(u^2)}{6} - \alpha l^2 - \beta l^2 \delta\rho, \quad (15)$$

$$q \equiv -\alpha l^2, \quad (16)$$

the discretised version of Eq. (12) becomes [55]

$$u_{i,j,k}^3 + p u_{i,j,k} + q = 0. \quad (17)$$

This cubic equation has a few advantages over the discretised equation in the old method. First, the result of the cubic equation can be found analytically for each mesh cell in each iteration step. The Newton-Raphson numerical update rule of the old method is thus avoided, leading to improved convergence speed and numerical stability of the solver. Second, the large number of expensive exponential and logarithm operations for each iteration step and mesh cell associated with the definition of  $u$  are avoided. Finally, the new method does not introduce extra nonlinearities – through the exponentials in Eq. (10) – in screened regions which slowed down the convergence of the solver in the old method, particularly at high redshift.

Applying Cardano's method in order to solve Eq. (17), one has to distinguish the three branches of solutions depending on

the values of  $p$  and  $q$ . In order to be consistent with [55], we define  $\Delta_0 \equiv -3p$  and  $\Delta_1 \equiv 27q$ . Independent of the density within a given cell and the field value of its neighbours,  $q < 0$  and thus  $\Delta_1 < 0$  in all cases. The behaviour of the solution depends on the sign of the discriminant  $D$ :

$$D \equiv \left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3. \quad (18)$$

**A:** For the case  $D > 0$ , Eq. (17) takes only one real solution which is given by

$$\tilde{u}_{i,j,k} = -\frac{1}{3} \left[ C + \frac{\Delta_0}{C} \right],$$

where 
$$C \equiv \left\{ \frac{1}{2} \left[ \Delta_1 + (\Delta_1^2 - 4\Delta_0^3)^{\frac{1}{2}} \right] \right\}^{\frac{1}{3}}. \quad (19)$$

We have used  $\tilde{u}_{i,j,k}$  to stress that the solution obtained is an approximation to the true solution  $u_{i,j,k}$  because the quantity  $L_{i,j,k}$  in  $p$ , cf. Eq. (15), depends on the values  $u_{i\pm 1, j\pm 1, k\pm 1}$  from the neighbouring cells. Note that this branch of the solution includes the case  $p = 0$ , for which Eq. (19) simplifies to  $\tilde{u}_{i,j,k} = \sqrt[3]{-q}$ .

**B:** For  $D < 1$ , Eq. (17) admits three real solutions:

$$\tilde{u}_{i,j,k} = -\frac{2}{3} \sqrt{\Delta_0} \cos \left( \frac{\Theta}{3} + \frac{2}{3} j \pi \right), \quad (20)$$

$$\text{where} \quad \Theta = \cos^{-1} \left( \frac{\Delta_1}{2\Delta_0^{\frac{2}{3}}} \right), \quad (21)$$

and  $j \in \{0, 1, 2\}$ . The cases  $j = 0$  and  $j = 2$  lead to  $\tilde{u}_{i,j,k} < 0$  which is in contradiction with our definition of  $u$ . The only physical solution thus is  $j = 1$ .

After the cubic equation is solved for a cell, the value of  $u$  therein is replaced by the above solution. In order to allow for efficient parallelisation on distributed-memory compute clusters and to allow for a faster convergence, we carry out the cell updates using a red-black-sweep pattern. We start from an initial guess of  $u = 1$  for the first time step of the simulation and use the result of the previous time step as an initial guess for all subsequent steps.

At the end of each iteration, an approximate error for  $u_{i,j,k}^2$  ( $\propto f_R$ ) is calculated as

$$e_{i,j,k} = \frac{r_{i,j,k}}{\bar{\rho}(a) l^2 \beta}, \quad (22)$$

where

$$r_{i,j,k} = \tilde{u}_{i,j,k}^3 + p \tilde{u}_{i,j,k} + q, \quad (23)$$

is the residual for a given approximate solution  $\tilde{u}_{i,j,k}$ . We consider the solution on a given AMR-level to be converged and stop the iterations once

$$\max(e_{i,j,k}) < 10^{-2}. \quad (24)$$

This criterion translates to an (approximate) maximum error of 1% on  $f_R/\bar{f}_R(a)$  on a given grid level. We note that the resulting fifth force errors are smaller than the random standard gravity errors originating from the tree walk.

### 3. Multigrid acceleration

The relaxation method used above successively reduces the error of the initial guess by iterations. It is well known that the decay of long wave-length modes of the error can be dramatically sped up by multigrid methods. For both the old and the new method we therefore use multigrid acceleration arranged in terms of V-cycles to speed up the convergence of the scalar field solver. Our V-cycle implementation for the old method is identical to MG-GADGET, and so we refer the reader to [15] for further details.

For the new method, we use the standard nonlinear full approximation storage (FAS) algorithm [57]. Eq. (17) is rewritten in terms of a non-linear operator

$$\mathcal{L}_l(u_l) = 0, \quad \text{where} \quad \mathcal{L}_l(u_l) \equiv u_l^3 + p_l u_l + q_l. \quad (25)$$

The subscript  $l$  denotes quantities and operators defined on the ‘fine’ grid level with cells of side-length  $l$  here, we will use  $\mathcal{L}$  for quantities and operators defined on the ‘coarse’ level with cells of side-length  $L = 2l$ . Eq. (23) can be written as

$$\mathcal{L}_l(\tilde{u}_l) = r_l, \quad (26)$$

for some approximate solution  $\tilde{u}_l$ . After a few pre-smoothing iterations to ensure that the short-wavelength modes in  $\tilde{u}_l$  have been reduced sufficiently, both sides of Eq. (26) are mapped to the coarse level using the restriction operator

$$\mathcal{R}\tilde{u}_l \equiv \sum_{i=0}^7 \frac{\tilde{u}_l^i}{8}, \quad (27)$$

where the summation is over the 8 daughter cells of the coarse cell. This leads to the following equation on the coarse level

$$\mathcal{L}_L(u_L) = \mathcal{L}_L(\mathcal{R}\tilde{u}_l) - \mathcal{R}r_l, \quad (28)$$

which can be solved either through iterations or further mapping to even coarser levels, to obtain an approximate coarse-level solution  $\tilde{u}_L$ . Using the prolongation operator  $\mathcal{P}\tilde{u}_L = \tilde{u}_l$  the coarse-level solution can then be used to correct the fine-level solution as

$$\tilde{u}_l^{\text{new}} = \tilde{u}_l + \mathcal{P}(\tilde{u}_L - \mathcal{R}\tilde{u}_l). \quad (29)$$

This correction is applied in our multigrid implementation as follows. If no initial guess for  $u$  is available, the solver starts to solve Eq. (17) iteratively on level 3 (the top-level node covering the whole simulation volume is level 1), and maps down this solution as the initial guess for the next finer level. This process is continued recursively until the finest level that covers the whole simulation volume (max-level-full-tree, MLFT) is reached. If an initial guess is available, the solver starts on MLFT directly.

All the finer levels (including MLFT) are then solved by V-cycles using corrections from the two next coarser grid levels. The solver would, for example, repeat the cycles 7-6-5-6-7 on level 7 until convergence is reached and then continue on level 8 with 8-7-6-7-8 V-cycles using the solution of level 7 as the initial guess on level 8. During the V-cycles, the solver carries out 4 pre- and 4 post-smoothing iterations before mapping to and after mapping back from the coarser level, to ensure that the short-wavelength modes of the error of  $\tilde{u}$  have sufficiently decayed. If the solver determines that the V-cycles do not lead to convergence on a given level, that level will be solved by iterations only. On the finest level which the mesh refinements extend to, the solution is obtained by iterations only for stability reasons.

#### 4. Force calculation

Both the old and new methods adopt the effective mass approach to calculate the total gravity force as in MG-GADGET, making use of the fact that the modified Poisson equation, (3), can be rewritten in terms of an effective density

$$\nabla^2\Phi = 4\pi G(\delta\rho + \delta\rho_{\text{eff}}), \quad (30)$$

where

$$\delta\rho_{\text{eff}} = \frac{\delta\rho}{3} - \frac{\delta R}{24\pi G}. \quad (31)$$

The advantage of this approach is that it allows to calculate the gravitational accelerations from Eq.(30) using the standard oct-tree – particle-mesh gravity solver readily implemented in AREPO.

#### 5. Local time stepping

To keep the number of computationally expensive modified gravity force calculation low, we adopt the local time-stepping scheme that was first implemented in MG-GADGET (for more details see [36]). This scheme takes advantage of the fact that the regions which require the smallest time steps, i.e., the regions with the highest accelerations, lie primarily within massive halos and are thus screened by the chameleon mechanism. For reasonable values of  $f_R$ , the maximum modified gravity acceleration in the simulation box at a given time is orders of magnitude smaller than the standard gravity acceleration.

Decoupling the modified gravity time step from the standard gravity and hydro time steps can therefore significantly reduce the computational cost of the simulations without loss of accuracy.

In practice, we decouple these time steps using a local time-stepping scheme already implemented in AREPO. The modified gravity time step is coupled to the long-range gravity PM time step global, while the short range gravitational forces and hydrodynamics are calculated on smaller time steps for the particles individually. In order to avoid very large time steps for the modified gravity solver, the global modified gravity time step is calculated according to

$$\Delta t_{\text{MG global}} = \min(\Delta t_{\text{MG } i}), \quad (32)$$

where  $\Delta t_{\text{MG } i}$  is the individual modified gravity time step for particle  $i$  assessed using an acceleration criterion. The global PM+modified gravity time step is finally determined by

$$\Delta t_{\text{MG+PM global}} = \min(\Delta t_{\text{MG global}}, \Delta t_{\text{PM global}}). \quad (33)$$

#### D. Power spectrum calculation

The power spectra of the total density field and the density field of the matter components are calculated using a module implemented within the AREPO code. The resolution elements are assigned to a Cartesian mesh using a CIC mass assignment in order to perform the required Fourier transforms [see 28, for a detailed description].

#### CORRESPONDING AUTHOR

Please address any correspondence related to this work to Christian Arnold (christian.arnold@durham.ac.uk).

#### ACKNOWLEDGEMENTS

We would like to thank the Illustris-TNG collaboration for allowing us to use their baryonic model to carry out the simulations presented in this work. We are particularly grateful to Volker Springel and Rainer Weinberger for their help with the AREPO code and for useful discussions on the results, and to Alejandro Benitez-Llambay for making Py-SPHViewer [59] available. Special thanks to Carlos Frenk and Jianhua He for their insightful comments on the results.

This work described in this paper is supported by the European Research Council through an ERC Starting Grant (ERC-StG-716532-PUNCA). BL is additionally supported by STFC Consolidated Grants ST/P000541/1 and ST/L00075X/1.

The cosmological simulations described in this work were run on the DiRAC Data Centric System at Durham University, United Kingdom, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS

National E-infrastructure capital grant ST/K00042X/1, STFC capital grants ST/H008519/1 and ST/K00087X/1, STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure.

The simulation code AREPO [26] is currently not publicly available. The analysis scripts used to analyze the simulation output can be made available to the reader on request.

### COMPETING INTERESTS

The authors declare no competing interests.

### AUTHOR CONTRIBUTIONS

C.A. and B.L. planned the project. C.A. developed, implemented and optimized (together with B.L.) the modified grav-

ity solver AREPO, ran the simulations and performed main part of the analysis. M.L. performed the analysis for the HI power spectrum. C.A., B.L. and M.L. interpreted the results. C.A. wrote the manuscript with contributions from M.L. and B.L.

### SUPPLEMENTARY MATERIAL

Supplementary material and the published version of this preprint are available from this website: <https://www.nature.com/articles/s41550-019-0823-y>

- 
- [1] Buchdahl, H. A. Non-linear Lagrangians and cosmological theory. *MNRAS* **150**, 1 (1970).
- [2] Hu, W. & Sawicki, I. Models of  $f(R)$  cosmic acceleration that evade solar system tests. *Phys. Rev. D* **76**, 064004 (2007). 0705.1158.
- [3] Khoury, J. & Weltman, A. Chameleon cosmology. *Phys. Rev. D* **69**, 044026 (2004). arXiv:astro-ph/0309411.
- [4] Will, C. M. The Confrontation between General Relativity and Experiment. *Living Reviews in Relativity* **17**, 4 (2014). 1403.7377.
- [5] Lombriser, L. & Taylor, A. Breaking a dark degeneracy with gravitational waves. *Journal of Cosmology and Astro-Particle Physics* **2016**, 031 (2016).
- [6] Sakstein, J. & Jain, B. Implications of the Neutron Star Merger GW170817 for Cosmological Scalar-Tensor Theories. *Physical Review Letters* **119**, 251303 (2017). 1710.05893.
- [7] Ezquiaga, J. M. & Zumalacárregui, M. Dark Energy After GW170817: Dead Ends and the Road Ahead. *Physical Review Letters* **119**, 251304 (2017).
- [8] Lombriser, L. & Lima, N. A. Challenges to self-acceleration in modified gravity from gravitational waves and large-scale structure. *Physics Letters B* **765**, 382–385 (2017).
- [9] Sotiriou, T. P. & Faraoni, V.  $f(R)$  theories of gravity. *Reviews of Modern Physics* **82**, 451–497 (2010). 0805.1726.
- [10] Joyce, A., Jain, B., Khoury, J. & Trodden, M. Beyond the cosmological standard model. *Phys. Rep.* **568**, 1–98 (2015). 1407.0059.
- [11] Schmidt, F. Dynamical masses in modified gravity. *Phys. Rev. D* **81**, 103002 (2010). 1003.0409.
- [12] Zhao, G.-B., Li, B. & Koyama, K. Testing Gravity Using the Environmental Dependence of Dark Matter Halos. *Physical Review Letters* **107**, 071303 (2011). 1105.0922.
- [13] Li, Y. & Hu, W. Chameleon halo modeling in  $f(R)$  gravity. *Phys. Rev. D* **84**, 084033 (2011). 1107.5120.
- [14] Lombriser, L., Li, B., Koyama, K. & Zhao, G.-B. Modeling halo mass functions in chameleon  $f(R)$  gravity. *Phys. Rev. D* **87**, 123511 (2013). 1304.6395.
- [15] Puchwein, E., Baldi, M. & Springel, V. Modified-Gravity-GADGET: a new code for cosmological hydrodynamical simulations of modified gravity models. *MNRAS* **436**, 348–360 (2013). 1305.2418.
- [16] Hellwing, W. A., Li, B., Frenk, C. S. & Cole, S. Hierarchical clustering in chameleon  $f(R)$  gravity. *MNRAS* **435**, 2806–2821 (2013). 1305.7486.
- [17] Zivick, P., Sutter, P. M., Wandelt, B. D., Li, B. & Lam, T. Y. Using cosmic voids to distinguish  $f(R)$  gravity in future galaxy surveys. *MNRAS* **451**, 4215–4222 (2015). 1411.5694.
- [18] Cautun, M. *et al.* The Santiago-Harvard-Edinburgh-Durham void comparison I: SHEDding light on chameleon gravity tests. *MNRAS* (2018). 1710.01730.
- [19] Mitchell, M. A., He, J.-h., Arnold, C. & Li, B. A general framework to test gravity using galaxy clusters - I. Modelling the dynamical mass of haloes in  $f(R)$  gravity. *MNRAS* **477**, 1133–1152 (2018).
- [20] Li, B. & Shirasaki, M. Galaxy-galaxy weak gravitational lensing in  $f(R)$  gravity. *MNRAS* **474**, 3599–3614 (2018). 1710.07291.
- [21] Arnold, C., Puchwein, E. & Springel, V. Scaling relations and mass bias in hydrodynamical  $f(R)$  gravity simulations of galaxy clusters. *MNRAS* **440**, 833–842 (2014). 1311.5560.
- [22] Arnold, C., Puchwein, E. & Springel, V. The Lyman  $\alpha$  forest in  $f(R)$  modified gravity. *MNRAS* **448**, 2275–2283 (2015). 1411.2600.
- [23] Hammami, A., Llinares, C., Mota, D. F. & Winther, H. A. Hydrodynamic effects in the symmetric and  $f(R)$ -gravity models. *MNRAS* **449**, 3635–3644 (2015). 1503.02004.
- [24] He, J.-h. & Li, B. Accurate method of modeling cluster scaling relations in modified gravity. *Phys. Rev. D* **93**, 123512 (2016). 1508.07350.
- [25] Winther, H. A. *et al.* Modified gravity N-body code comparison project. *MNRAS* **454**, 4208–4234 (2015). 1506.06384.
- [26] Springel, V. E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh. *MNRAS* **401**, 791–851 (2010). 0901.4107.
- [27] Pillepich, A. *et al.* Simulating galaxy formation with the IllustrisTNG model. *MNRAS* **473**, 4077–4106 (2018).
- [28] Springel, V. *et al.* First results from the IllustrisTNG simulations: matter and galaxy clustering. *MNRAS* **475**, 676–698

- (2018). 1707.03397.
- [29] Genel, S. *et al.* The size evolution of star-forming and quenched galaxies in the IllustrisTNG simulation. *MNRAS* **474**, 3976–3996 (2018).
- [30] Marinacci, F. *et al.* First results from the IllustrisTNG simulations: radio haloes and magnetic fields. *MNRAS* **480**, 5113–5139 (2018).
- [31] Nelson, D. *et al.* First results from the IllustrisTNG simulations: the galaxy colour bimodality. *MNRAS* **475**, 624–647 (2018). 1707.03395.
- [32] Arnold, C., Fosalba, P., Springel, V., Puchwein, E. & Blot, L. The modified gravity lightcone simulation project I: Statistics of matter and halo distributions. *MNRAS* **483**, 790–805 (2019).
- [33] Ferrero, I. *et al.* Size matters: abundance matching, galaxy sizes, and the Tully-Fisher relation in EAGLE. *MNRAS* **464**, 4736–4746 (2017).
- [34] Hellwing, W. A. *et al.* The effect of baryons on redshift space distortions and cosmic density and velocity fields in the EAGLE simulation. *MNRAS* **461**, L11–L15 (2016).
- [35] Pillepich, A. *et al.* First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies. *MNRAS* **475**, 648–675 (2018). 1707.03406.
- [36] Arnold, C., Springel, V. & Puchwein, E. Zoomed cosmological simulations of Milky Way-sized haloes in f(R) gravity. *MNRAS* **462**, 1530–1541 (2016). 1604.06095.
- [37] Villaescusa-Navarro, F. *et al.* Ingredients for 21 cm Intensity Mapping. *ApJ* **866**, 135 (2018). 1804.09180.
- [38] Santos, M. *et al.* Cosmology from a SKA HI intensity mapping survey. In *Advancing Astrophysics with the Square Kilometer Array (AASKA14)*, 19 (2015). 1501.03989.
- [39] Altay, G., Theuns, T., Schaye, J., Booth, C. M. & Dalla Vecchia, C. The impact of different physical processes on the statistics of Lyman-limit and damped Lyman  $\alpha$  absorbers. *MNRAS* **436**, 2689–2707 (2013). 1307.6879.
- [40] Behroozi, P. S., Wechsler, R. H. & Conroy, C. The Average Star Formation Histories of Galaxies in Dark Matter Halos from  $z = 0-8$ . *ApJ* **770**, 57 (2013). 1207.6105.
- [41] D’Souza, R., Vegetti, S. & Kauffmann, G. The massive end of the stellar mass function. *MNRAS* **454**, 4027–4036 (2015). 1509.07418.
- [42] Bernardi, M. *et al.* The massive end of the luminosity and stellar mass functions: dependence on the fit to the light profile. *MNRAS* **436**, 697–704 (2013). 1304.7778.
- [43] Baldry, I. K. *et al.* Galaxy And Mass Assembly (GAMA): the galaxy stellar mass function at  $z \leq 0.06$ . *MNRAS* **421**, 621–634 (2012). 1111.5707.
- [44] Giodini, S. *et al.* Stellar and Total Baryon Mass Fractions in Groups and Clusters Since Redshift 1. *ApJ* **703**, 982–993 (2009). 0904.0448.
- [45] Lovisari, L., Reiprich, T. H. & Schellenberger, G. Scaling properties of a complete X-ray selected galaxy group sample. *A&A* **573**, A118 (2015). 1409.3845.
- [46] Planck Collaboration *et al.* Planck 2015 results. XIII. Cosmological parameters. *A&A* **594**, A13 (2016). 1502.01589.
- [47] Vogelsberger, M. *et al.* Properties of galaxies reproduced by a hydrodynamic simulation. *Nature* **509**, 177–182 (2014). 1405.1418.
- [48] Vogelsberger, M. *et al.* Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe. *MNRAS* **444**, 1518–1547 (2014). 1405.2921.
- [49] Weinberger, R. *et al.* Simulating galaxy formation with black hole driven thermal and kinetic feedback. *MNRAS* **465**, 3291–3308 (2017).
- [50] Sawicki, I. & Bellini, E. Limits of quasistatic approximation in modified-gravity cosmologies. *Phys. Rev. D* **92**, 084061 (2015). 1503.06831.
- [51] Terukina, A. *et al.* Testing chameleon gravity with the Coma cluster. *JCAP* **4**, 013 (2014). 1312.5083.
- [52] He, J.-h., Guzzo, L., Li, B. & Baugh, C. M. No evidence for modifications of gravity from galaxy motions on cosmological scales. *Nature Astronomy* **2**, 967–972 (2018). 1809.09019.
- [53] Springel, V. The cosmological simulation code GADGET-2. *MNRAS* **364**, 1105–1134 (2005). arXiv:astro-ph/0505010.
- [54] Li, B., Zhao, G.-B., Teyssier, R. & Koyama, K. ECOSMOG: an Efficient COde for Simulating MOdified Gravity. *JCAP* **1**, 51 (2012). 1110.1379.
- [55] Bose, S. *et al.* Speeding up N-body simulations of modified gravity: chameleon screening models. *Journal of Cosmology and Astro-Particle Physics* **2017**, 050 (2017).
- [56] Llinares, C., Mota, D. F. & Winther, H. A. ISIS: a new N-body cosmological code with scalar fields based on RAMSES. Code presentation and application to the shapes of clusters. *A&A* **562**, A78 (2014). 1307.6748.
- [57] Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (eds.) *Numerical recipes* (Cambridge Univ. Press, Cambridge [u.a.], 2007), 3. ed. edn. Includes bibliographical references and index. - This printing is corrected to software version 3.0.
- [58] Schaye, J. *et al.* The EAGLE project: simulating the evolution and assembly of galaxies and their environments. *MNRAS* **446**, 521–554 (2015). 1407.7040.
- [59] Benitez-Llambay, A. py-sphviewer: Py-sphviewer v1.0.0 (2015). URL <http://dx.doi.org/10.5281/zenodo.21703>.