

Advances in constraining intrinsic alignment models with hydrodynamic simulations

S. Samuroff¹,¹★ R. Mandelbaum¹ and J. Blazek^{2,3}

¹*McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

²*Department of Physics, Northeastern University, Boston, MA 02115, USA*

³*Laboratory of Astrophysics, École Polytechnique Fédérale de Lausanne, CH-1290 Versoix, Switzerland*

Accepted 2021 September 1. Received 2021 August 13; in original form 2020 September 22

ABSTRACT

We use galaxies from the ILLUSTRISTNG, MASSIVEBLACK-II, and ILLUSTRIS-I hydrodynamic simulations to investigate the behaviour of large scale galaxy intrinsic alignments. Our analysis spans four redshift slices over the approximate range of contemporary lensing surveys $z = 0-1$. We construct comparable weighted samples from the three simulations, which we then analyse using an alignment model that includes both linear and quadratic alignment contributions. Our data vector includes galaxy–galaxy, galaxy–shape, and shape–shape projected correlations, with the joint covariance matrix estimated analytically. In all of the simulations, we report non-zero IAs at the level of several σ . For a fixed lower mass threshold, we find a relatively strong redshift dependence in all three simulations, with the linear IA amplitude increasing by a factor of ~ 2 between redshifts $z = 0$ and $z = 1$. We report no significant evidence for non-zero values of the tidal torquing amplitude, A_2 , in TNG, above statistical uncertainties, although MBII favours a moderately negative $A_2 \sim -2$. Examining the properties of the TATT model as a function of colour, luminosity and galaxy type (satellite or central), our findings are consistent with the most recent measurements on real data. We also outline a novel method for constraining the TATT model parameters directly from the pixelized tidal field, alongside a proof-of-concept exercise using TNG. This technique is shown to be promising, although comparison with previous results obtained via other methods is non-trivial.

Key words: gravitational lensing: weak – methods: numerical – large-scale structure of Universe – cosmology: theory.

1 INTRODUCTION

It is now well established that the weak lensing of distant galaxies by foreground mass provides a relatively clear window on to the large-scale structure of the Universe. This is true whether that foreground mass is in the form of discrete matter concentrations, as traced by galaxies (i.e. galaxy–galaxy lensing; Mandelbaum et al. 2013; Leauthaud et al. 2017; Joudaki et al. 2018; Prat et al. 2018; Blake et al. 2020), massive dark matter haloes (cluster lensing; Melchior et al. 2017; Dark Energy Survey Collaboration 2020), or the continuous large-scale matter distribution (cosmic shear; Heymans et al. 2013; Dark Energy Survey Collaboration 2016; Troxel et al. 2018; Chang et al. 2019; Hamana et al. 2020; Hildebrandt et al. 2020; Amon et al. 2021; Asgari et al. 2021; Secco, Samuroff et al. 2021). Though the measurement method and the exact form of the theory predictions differ slightly in the three cases, they are all fundamentally probes of the growth of structure at low redshift. Similarly cross-correlations between galaxy lensing and other observables can be powerful probes in their own right; recent examples include galaxy lensing \times CMB lensing (Schaan et al. 2017), voids correlated with CMB lensing (Vielzeuf et al. 2021), and galaxy weak lensing crossed with gamma-ray emission (Ammazzalorso et al. 2020), each of which provide probes of dark matter with slightly different sensitivities. A

measurement of cosmological weak lensing, however, is subject to a range of systematic effects; that is, observational effects that mimic a cosmological lensing signal, and so bias cosmological inference if one neglects them. Depending on the systematic in question, the most effective mitigation strategy may be quite different. In broad terms, however, the standard approach is to either (i) mitigate systematics where possible, either by applying a calibration to the data, or discarding the data points most strongly affected or (ii) marginalize over them with a parametric model. Often a combination of the two is appropriate, and the prior used in (ii) is informed by additional data or simulations, and detailed testing of the calibration step in (i).

This work focuses on one particular source of systematic bias, which enters all of the weak lensing measurements described above: galaxy intrinsic alignments (IAs). The fact that the projected shapes of galaxies residing in the same local region of the cosmic web are correlated has been known for many years now (Catelan, Kamionkowski & Blandford 2001; Heymans & Heavens 2003). For pairs of galaxies at the same redshift, the physically localized intrinsic shape–shape correlations can persist even on relatively large angular scales. Fortunately, in practice this signal, commonly referred to as the II contribution, is typically weak; it is also absent, by construction, from a measurement of galaxy–galaxy lensing, which reduces the sensitivity to II further in the context of a multiprobe analysis. Often more dangerous are what are known as GI correlations, which arise

* E-mail: ssamuroff@cmu.edu

due to the fact that foreground mass causes both local gravitational interactions in foreground galaxies and lensing in background objects (Hirata & Seljak 2004).

Unfortunately, many of the avenues available for understanding other lensing systematics are not feasible in the case of intrinsic alignments. For example image simulations, which have become an invaluable tool for quantifying shear calibration errors (Mandelbaum et al. 2018; Zuntz et al. 2018; Kannawadi et al. 2019; Sánchez et al. 2020) cannot be used for understanding IAs due to their fundamentally astrophysical nature. For quite different reasons, the various sophisticated methods that the lensing community has developed for calibrating photometric redshift errors in recent years (e.g. Choi et al. 2016; Gruen & Brimiouille 2017; Gatti et al. 2018, 2021; Prat et al. 2019; Alarcon et al. 2020; Myles et al. 2021) have limited potential for cross-use as IA mitigation tools. Although direct mitigation methods have been proposed in the literature (Heymans & Heavens 2003; Joachimi & Schneider 2010), to date these have been limited in their applicability, in large part because they tend to rely on having good single-galaxy redshift information. They also often focus on the (typically subdominant) II contribution (although the Joachimi & Schneider 2010 method can include both). The standard approach in cosmological lensing studies is to model IAs using a (semi-) physically motivated parametric model, and marginalize over its (typically 2–5) parameters with wide flat priors. Given this background, hydrodynamic simulations are one of a small number of possible routes to understanding intrinsic alignments in cosmological lensing surveys, either for model building, or deriving informative priors for the existing models. Although analytic models are relatively well motivated on very large physical scales, this is much less true on small to intermediate scales. Extending beyond this regime, then, either requires simulations or the addition of extra terms to the model, controlled by new parameters (Schneider & Bridle 2010; Blazek, Vlah & Seljak 2015; Blazek et al. 2019; Fortuna et al. 2021). Although not the focus of this paper, another route is to use real galaxies to make a direct IA measurement (see e.g. Hirata et al. 2007; Joachimi et al. 2011; Mandelbaum et al. 2011; Blazek et al. 2012; Singh, Mandelbaum & More 2015; Johnston et al. 2019). This approach avoids questions about the realism of simulations. It does, however, have its own challenges, not least the need for accurate per-galaxy redshift information, and the typically fairly restricted galaxy selections (often bright, red, low redshift samples).

Although a substantial amount of literature exists on the subject of IAs in hydrodynamic simulations, it is fair to say that there is significant variation in focus and methodology. For example, a series of studies by a group working on the HORIZON-AGN simulation have looked in detail at the alignment of 2D and 3D subhalo shapes with their local large scale structure and the cosmic web (e.g. Dubois et al. 2014; Codis et al. 2015a; Soussana et al. 2020). Intriguingly, Codis et al. (2015a) found hints that blue galaxy IAs could survive in projection at a level detectable by future surveys. A number of papers based on MASSIVEBLACK-II (e.g. Tenneti et al. 2015a,b; Bhowmick et al. 2020) and ILLUSTRIS-I (Hilbert et al. 2017) have explored similar themes. Minor discrepancies in the details of the IA signal, and its dependence on galaxy properties, have been uncovered; thus far, however, the interpretation of these differences has been complicated by both the relatively low signal-to-noise on large scales, and methodological differences. In a more recent addition to this literature, Zjupa, Schäfer & Hahn (2020) explore IAs in the ILLUSTRISTNG simulations. In that work, they fit linear and quadratic alignment models to elliptical and spiral galaxies, over the range $z = 0–1$ and as a function of environment, with results consistent with those presented in this paper.

This work is intended as a step towards a more complete understanding of intrinsic alignments in hydrodynamic simulations, building on these earlier studies. We present a unified analysis of samples from various recent simulations, with measurement methods and selection functions matched in order to make a meaningful quantitative comparison. Unlike many previous studies, we focus on two-point intrinsic-galaxy and intrinsic–intrinsic alignment statistics w_{g+} and w_{++} , which are commonly used in observational studies; this is primarily because one can derive well-defined analytic predictions for them, which directly correspond to the IA modelling used in cosmological lensing analyses. This is less true of statistics like the 3D EE and ED correlations (e.g. Chisari et al. 2015; Tenneti et al. 2015a), and halo misalignment statistics (Codis, Pichon & Pogosyan 2015b; Tenneti et al. 2015b), all of which have been used in many simulation-based studies. In this work we perform a simultaneous analysis of these w_{g+} and w_{++} , alongside the equivalent galaxy–galaxy correlations, in order to fully exploit the large-scale IA information in these simulated data sets.

The paper is structured as follows. In Section 2 we outline the properties of the three simulated data sets used in this work, ILLUSTRISTNG, MASSIVEBLACK-II, and ILLUSTRIS-I, and describe the selection used to construct comparable object catalogues. Section 3 then sets out the pipeline taking us from public (stellar and dark matter) particle data and SUBFIND group tables to shape catalogues, and eventually to two-point measurements. The theory calculations, which we use to connect these measurements to IA models, are described in Section 4. In Section 5 we present the results of our baseline-likelihood analyses using the two-point alignment data, and then in Section 6 we discuss a series of extensions. We fit one of the more sophisticated alignment models in the literature, and consider the dependence of its parameters on various galaxy properties. In addition to the two-point constraints, Section 7 presents a novel method for extracting alignment information directly from the simulated matter field. We develop the basic principles, and present an example using ILLUSTRISTNG. Finally, we conclude and briefly discuss our results in the context of the field in Section 8.

2 DATA

We consider three discrete cosmological simulation volumes in this study. Of these, ILLUSTRISTNG is chronologically the most recent, and so benefits from the improvements derived from the analysis of earlier simulation efforts. The simulation runs are evolved according to Newtonian dynamics and assume similar but non-identical cosmologies, which are set out in Table 1, with particles evolved from a set of initial conditions at high redshift. In each redshift snapshot, groups are identified using the SUBFIND friends-of-friends (FoF) group finding algorithm (Springel et al. 2001).

2.1 MASSIVEBLACK-II

MASSIVEBLACK-II has been used in various previous studies, and is described in a number of existing publications; details about the approximations and modelling can be found in Khandai et al. (2015) and Di Matteo et al. (2012). The simulation has a comoving volume of $(100 h^{-1} \text{ Mpc})^3$, and was generated using P-GADGET, which is a version of GADGET3 (Springel 2005). Initial conditions were generated with a transfer function generated by CMBFAST at $z = 159$. Star formation is modelled as a binary phase process, triggered when a region of gas reaches some threshold density.

Table 1. Properties of the simulation volumes used in this work. The particle mass quoted in the right-most column is the mean of gas particles. Note that in MASSIVEBLACK-II all particles are equally weighted, while ILLUSTRISTNG they cover a range (see Nelson et al. 2019). σ_8 is shown in parentheses as it is a derived parameter.

Simulation	Volume / $h^{-3} \text{ Mpc}^3$	Cosmology	Mean gas particle Mass/ $10^6 M_\odot$
ILLUSTRISTNG	205 ³	$A_s = 2.13 \times 10^{-9}$ $\Omega_m = 0.31$ $\Omega_b = 0.05$ $n_s = 0.97$ $h = 0.68$ ($\sigma_8 = 0.816$)	11.0
MBII	100 ³	$A_s = 2.43 \times 10^{-9}$ $\Omega_m = 0.28$ $\Omega_b = 0.05$ $n_s = 0.97$ $h = 0.70$ ($\sigma_8 = 0.816$)	2.2
ILLUSTRIS-1	75 ³	$A_s = 2.23 \times 10^{-9}$ $\Omega_m = 0.27$ $\Omega_b = 0.05$ $n_s = 0.96$ $h = 0.70$ ($\sigma_8 = 0.809$)	1.3

Stellar particles are generated randomly from gas particles with a probability determined by their star formation rate. Stellar winds are modelled using the parametrization of Hernquist & Springel (2003). AGN feedback, which is particularly relevant in high mass galaxy populations, where IAs are also strong, is also included; details of the black hole growth and AGN feedback models see Khandai et al. (2015)’s Section 2.3. MASSIVEBLACK-II is several years old now, and has a number of known deficiencies. Particularly, there are relatively few spiral galaxies, and an overabundance of extended ellipticals compared with data (Tenneti, Mandelbaum & Di Matteo 2016). The AGN feedback prescription is also known to be weak, meaning that there are more massive galaxies at low redshift than data suggests (Khandai et al. 2015).

2.2 Illustris

ILLUSTRIS-1 is another hydrodynamic simulation whose data are public. The smallest of the three considered in this work, the box has a total comoving volume of $V = (75 h^{-1} \text{ Mpc})^3$, which was evolved using the moving mesh grid code, AREPO (Weinberger, Springel & Pakmor 2020). The various physical processes approximated in ILLUSTRIS-1, in brief, include radiative cooling (both primordial and due to heavy elements) with self-shielding corrections; star-formation in dense regions of gas; stellar evolution with associated metal enrichment; supernova feedback and quasar-mode, radio-mode, and radiative mode AGN feedback. The above prescriptions have ~ 15 tunable parameters, which were fixed to values obtained using a significantly smaller volume, higher resolution, set of simulations. Details of these models can be found in Vogelsberger et al. (2014). Although it is useful as a comparison here, ILLUSTRIS-1 is known to have flaws; in particular the stellar and AGN feedback prescription is extreme, making it something of an outlier compared with other simulations in the literature such as MASSIVEBLACK-II, Eagle and Horizon-AGN (see Nelson et al. 2015 for a summary).

2.3 ILLUSTRISTNG

ILLUSTRISTNG is the most recent hydrodynamic simulation included here. The particle and group data are described in the release papers (Springel et al. 2018; Nelson et al. 2019), and are available for download.¹ The ILLUSTRISTNG data are generated using AREPO. A Monte Carlo tracer particle scheme is used to follow the Lagrangian evolution of baryonic matter. The hydrodynamic element comprises prescriptions for a handful of different physical processes, including emission line radiative cooling; stochastic star formation; supernova feedback, and AGN feedback. The latter has two modes (referred to as ‘quasar’ and ‘kinetic wind’ modes), depending on the accretion rate. Details of these prescriptions can be found in Pillepich et al. (2018a). It is worth remarking that ILLUSTRISTNG is tuned explicitly to match observations at $z = 0$ using a number of statistics; specifically the galaxy stellar mass function, the total gas mass content within the virial radius of massive groups, the stellar mass–stellar size, and the black hole–galaxy mass relations, and the overall shape of the cosmic star formation rate density at high redshift.

2.4 Sample selection

2.4.1 Fiducial catalogues

To obtain a galaxy sample from which we can draw useful conclusions for each of the simulated data sets, we impose additional quality cuts. Although our measurements are not subject to the usual observational biases (due, for example, to fitting ellipticities in the presence of pixel noise, or imperfect PSF modelling), they are affected by convergence bias (e.g. Chisari et al. 2015). That is, subhaloes with an insufficient number of particles to provide a meaningful shape measurement alter the ensemble ellipticity distribution of the sample. To avoid such effects, we impose a selection based on the number of particles in a galaxy (dark matter and stellar). The exact choice of cut here is based on the convergence tests of Tenneti et al. (2015b). This translates into a slightly different mass cut for each simulation due to the respective mass resolutions of the three data sets. We thus additionally impose a direct cut on stellar mass, such that the samples all have the same lower bound on M_* . The final selection is then:

$$\begin{aligned} n_{\text{DM}} &> 1000 \\ n_* &> 300 \\ M_* &> 1.6 \times 10^9 h^{-1} M_\odot. \end{aligned} \quad (1)$$

This leaves a total of $\sim 15\,000$, $35\,000$, and $170\,000$ usable galaxies in ILLUSTRIS-1, MASSIVEBLACK-II, and ILLUSTRISTNG samples, respectively. Note that the cut in equation (1) is imposed on each snapshot independently, resulting in the per-redshift numbers shown in Table 2.

In Fig. 1, we show the ellipticity distribution and stellar mass function for each of the samples. ‘Ellipticity’ in this context is defined as the magnitude of the spin-2 complex ellipticity defined in Section 3.1. As discussed there, the exact value for a given galaxy is dependent on the details of the measurement method (i.e. the relative weighting of stellar matter at different radii). Given that the measurement pipeline is applied consistently to the different simulations, however, Fig. 1 does allow a meaningful comparison. The striking discrepancy in the upper panel has been noted elsewhere (see for example, Tenneti et al. 2016’s fig. 2); galaxies in ILLUSTRIS-1 are significantly rounder than

¹<http://www.tng-project.org/data>

Table 2. Physical properties of the galaxy samples considered in this work. The object selection is as set out in Section 2.4, and is applied independently at each redshift. Here n_c (fourth column) is the comoving galaxy number density of the sample. The methods used to separate red/blue and satellite/central galaxies are described in Sections 6.2.4 and 6.2.1, respectively.

Simulation	Redshift	Number of galaxies	$n_c/h^3\text{Mpc}^{-3}$	Red fraction	Satellite fraction	Mean stellar mass/ $10^9 M_\odot$
ILLUSTRISTNG	0.00	171 684	0.020	0.34	0.33	20.0
ILLUSTRISTNG	0.30	168 399	0.020	0.22	0.32	18.8
ILLUSTRISTNG	0.62	159 925	0.019	0.18	0.30	17.5
ILLUSTRISTNG	1.00	145 394	0.017	0.12	0.27	16.0
MASSIVEBLACK-II	0.00	33 578	0.033	N/A	0.45	15.0
MASSIVEBLACK-II	0.30	34 646	0.035	N/A	0.46	13.2
MASSIVEBLACK-II	0.62	35 523	0.036	N/A	0.48	11.6
MASSIVEBLACK-II	1.00	35 482	0.036	N/A	0.49	10.0
ILLUSTRIS-I	0.00	18 489	0.044	N/A	0.32	17.6
ILLUSTRIS-I	0.30	17 203	0.041	N/A	0.31	16.8
ILLUSTRIS-I	0.62	15 181	0.036	N/A	0.29	16.1
ILLUSTRIS-I	1.00	12 881	0.031	N/A	0.27	15.0

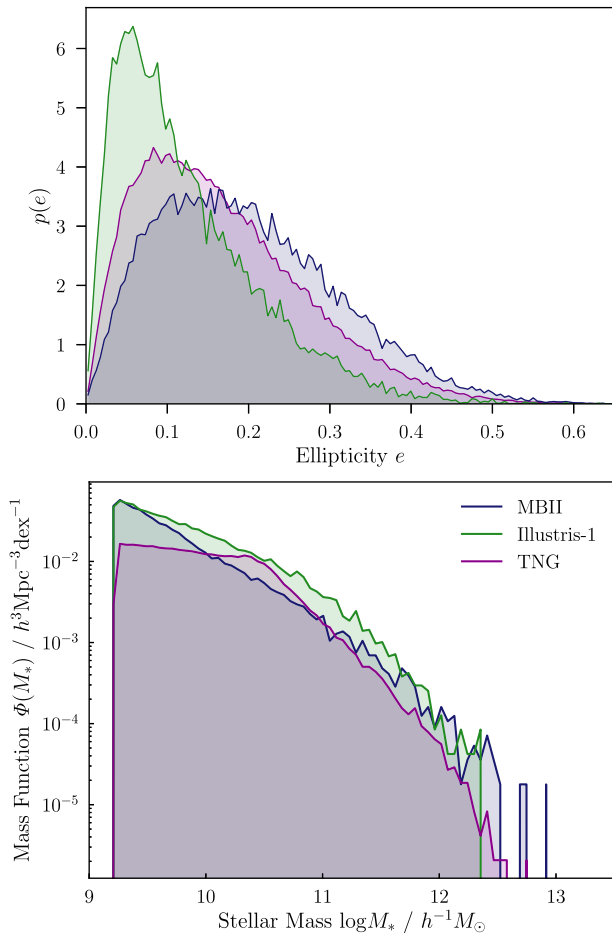


Figure 1. Upper: Normalized distributions of projected ellipticity for the $z = 0$ samples used in this work, with the cuts described in Section 2.4. Shown are ILLUSTRISTNG (purple), MASSIVEBLACK-II (dark blue), and ILLUSTRIS-I (green). Lower: the stellar mass functions for the same samples.

both comparable simulations and real data. The differences in the mass function (lower panel) mean that, even with a common lower bound, the mean stellar mass of the samples differs slightly (see the right-hand column in Table 2). At given redshift, the mean masses are ordered (descending) ILLUSTRISTNG, ILLUSTRIS-I, MASSIVEBLACK-II. The differences in the low mass end of the stellar mass function

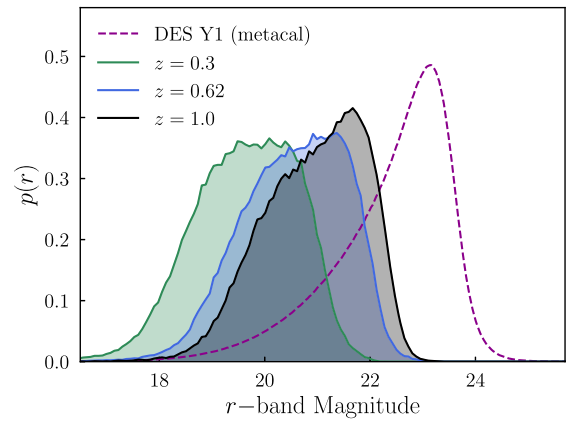


Figure 2. Normalized distributions of r -band apparent magnitude for our ILLUSTRISTNG sample. We convert the absolute magnitude in the catalogues to apparent magnitude at each snapshot, assuming the correct cosmology of the simulation. A detailed description of how the simulated absolute magnitudes are computed can be found in Nelson et al. (2018). A power-law approximation for the SED is used to compute the k -corrections for the apparent magnitudes; although this is not rigorously correct, it is sufficient for our purposes, given that the k -corrections are comfortably subdominant to the distance modulus, and that we are only attempting a qualitative comparison here. For reference, the unfilled curve shows the equivalent distribution for the fiducial Dark Energy Survey Year 1 shape catalogue, after quality cuts (c.f. Zuntz et al. 2018's fig. 3).

are thought to come primarily from changes in how stellar-driven winds are implemented in ILLUSTRISTNG. For further discussion and comparison with data, see Pillepich et al. (2018b).

ILLUSTRISTNG is unusual amongst hydrodynamic simulations, in the sense that it has realistic galaxy magnitudes, integrated over a number of different pass bands. We include the SDSS *griz* band magnitudes in our processed catalogues, and will use them in the following sections. Briefly, these are evaluated by summing the luminosity of star particles in a particular subhalo, and the appropriate filter band-pass is applied. More detail on this calculation can be found in Nelson et al. (2018)'s section 3. The distribution of apparent r -band magnitudes in three ILLUSTRISTNG snapshots is shown in Fig. 2. For reference, the observed magnitude distribution from the DES Y1 METACALIBRATION catalogue is also included (dashed purple). It is worth remembering here that, unlike the simulated data, DES is a flux-limited imaging survey, with galaxies distributed

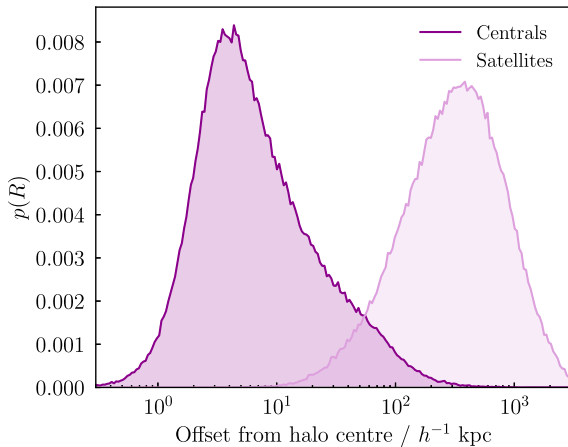


Figure 3. Normalized distributions of galaxy offsets from the centre of mass of their host haloes, in the ILLUSTRISTNG simulation at $z = 0$. The central flag used is defined in Section 2.4.2. Note that the distribution labelled ‘satellites’ is boosted by a factor of 15 for visibility.

across a range of redshifts (ensemble median redshift $z \sim 0.59$; Zuntz et al. 2018), and so direct comparison is not useful; they are shown here to illustrate that the simulated galaxy samples here are not representative of those in a typical lensing survey, but are a brighter subset.

2.4.2 Central flagging

Key to halo model-based descriptions of galaxy alignments is the ability to split galaxies cleanly into satellites and centrals (see Fortuna et al. 2021 for a recent example). Galaxies residing at the centres of their halo tend to be older and more massive than the satellites in the same halo; in the halo model picture, the clustering and shape properties of these two sets of galaxies is fundamentally different. For this reason it is, then, interesting to explore the behaviour of satellites and centrals separately. For this work we simply designate the most massive galaxy in each FoF group as the central.² Although noisy, this definition is less prone to misclassification than one based on geometry, particularly in high mass groups in which the region around the bottom of the potential well is relatively crowded. We show the distribution of galaxy–halo separations for centrals and satellites at $z = 0$ in Fig. 3. Although not shown here, a similar pattern is seen in the higher redshift snapshots. That the mass based classifier is a strong indicator of galaxy position in the halo offers some reassurance that the central flagging is, in fact, literally selecting central galaxies. This is a relatively old problem, and various previous studies have explored different ways to flag central galaxies (see for example, Rykoff et al. 2016).

2.4.3 Galaxy colours

There is much evidence in the literature to indicate that IAs are strongly dependent on galaxy colour (Joachimi et al. 2011; Heymans et al. 2013; Singh et al. 2015; Johnston et al. 2019; Samuroff et al. 2019). Clearly photometric colour is a proxy for a host of other physical properties, which ultimately determine how strongly the galaxy sample is aligned, and one could equivalently use other

properties such as morphology and bulge/disc ratio. Although crude, a binary type split is often useful, given that mixed galaxy samples commonly exhibit a clear bimodality in colour (or colour–magnitude) space (e.g. Baldry et al. 2004; Valentini et al. 2018), and that this maps roughly on to differences in IA properties. That said, the IA signal in the simulations (or indeed any galaxy sample) is a complex function of many correlated quantities (e.g. colour, morphology, dynamical properties). Although it is useful to study IAs in subpopulations defined using proxies, it is worth proceeding with care, and bearing in mind that the full picture is more complicated.

Whereas quantities like stellar mass and subhalo shapes are relatively simple to obtain from hydrodynamic simulations, mapping them on to observable quantities like fluxes and colours is non-trivial. This has historically been a challenging problem, and there are documented deficiencies in the galaxy photometry for MASSIVEBLACK-II and ILLUSTRIS-I; the equivalent quantities for ILLUSTRISTNG are, however, thought to be fairly realistic (see e.g. Nelson et al. 2018). In brief, in ILLUSTRISTNG a stellar synthesis model is used to predict the stellar population of each particle in a subhalo as a function of metallicity and age. This process includes basic models for dust emission and nebular line emission. The predicted stellar spectrum is multiplied by the SDSS optical/near-IR *ugriz* band-passes (airmass 1.3), producing a magnitude in each filter. The per-particle magnitudes are then summed over the ensemble bound to the subhalo. This process is explained in more detail in Nelson et al. (2018)’s section 3 (see their ‘Model (A)’).

We inspect the colour–magnitude diagrams and make a linear division in $g - i$ colour space

$$(g - i) = m_{gi} \times r + c_{gi}, \quad (2)$$

to roughly mimic the green valley division. The colour magnitude diagram evolves with redshift, and so we carry out this process independently in each snapshot, giving $\mathbf{m}_{gi} = (0.045, 0.045, 0.055, 0.022)$, $\mathbf{c}_{gi} = (1.84, 1.84, 1.95, 1.19)$. The red fraction resulting from this split at each redshift is shown in Table 2. The numbers here are roughly consistent with those seen in real data, and change with redshift in an intuitively correct way (i.e. the low redshift Universe has a larger abundance of massive red elliptical galaxies compared with $z = 1$). The $r - i$ colour magnitude diagram for our split sample is shown in Fig. 4. Given that the split is imposed in $g - i$, it is somewhat reassuring that we see clearly defined well separated samples in this space.

3 MEASUREMENTS

3.1 Galaxy shapes

In a 3D cosmological volume, the most natural way to quantify a galaxy’s shape is via its inertia tensor. Analogous to projected ellipticities, which are constructed from the moments of a galaxy light profile, the most general form for the inertia tensor is:

$$I_{ij} = \frac{1}{W} \sum_{k=1}^{N_p} w_k x_{i,k} x_{j,k}, \quad (3)$$

where the indices ij indicate one of the three spatial coordinate axes $i, j \in (x, y, z)$, and the sum runs over the number of particles within the subhalo. For our purposes, this means star particles, but one could equivalently estimate the shape of the dark matter subhalo using the same equation. The prefactor w_k is the weight allocated to particle k , and W is the sum of the weights; in the case of MASSIVEBLACK-II, all of the star particles have the same mass, and so the weights

²In the nomenclature of the TNG data release, the central in each group is identified using the ‘GroupFirstSub’ flag.

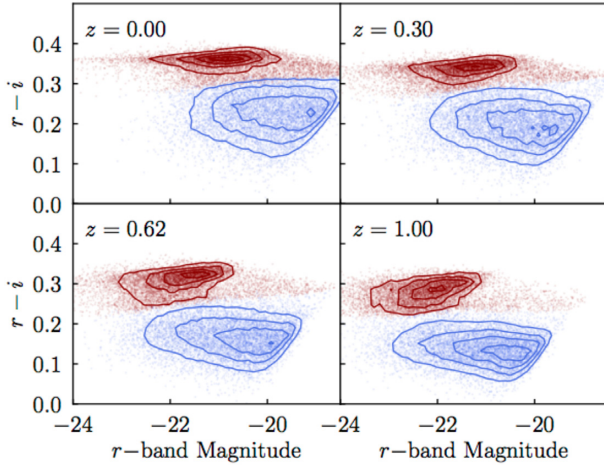


Figure 4. Colour magnitude diagrams for our ILLUSTRISTNG sample at four redshifts (labelled upper left-hand panel). The two sets of contours show the distributions of the red and blue samples, as outlined in the text. Note that the split is imposed in $g-i$ versus r space; by construction it separates the red/blue peaks in this space, but this is not true for other colour definitions (such as the one shown here). The fine points show a randomly downsampled selection of galaxies from each population. Note that unlike in Fig. 2, the magnitudes used here (including for estimating colours) are absolute, not apparent, ones.

are flat. In ILLUSTRISTNG and ILLUSTRIS-1, this is not the case, and each particle is weighted by its mass. An alternative, known as the reduced inertia tensor (see Chisari et al. 2015; Tenneti et al. 2016), weights particles by their inverse square distance from the subhalo centroid. This process is known to bias the measured ellipticities low, necessitating a further iterative correction procedure. Although we mention this here for context, since it has been used a handful of times in the literature, it is not used in this work. Note also that we have reason to think the IA signal is, in reality, dependent on the radial weighting of the shape measurements, an effect that has been observed in real data (Singh & Mandelbaum 2016).

By performing an eigenvalue decomposition on I , one can obtain 3D axis vectors and lengths, which in turn can be projected into the 2D second moments Q_{xx} , Q_{yy} , Q_{xy} . The recipe is set out by Piras et al. (2018) (see their equations 13–15), and we refer the reader to that paper for the mathematical detail. Although for technical reasons our pipeline goes via 3D shapes, it is also worth noting that one could also simply measure the projected 2D moments of a subhalo directly. With the projected moments, one can then construct the spin-2 ellipticity of a galaxy as

$$(e_1, e_2) = \frac{(Q_{xx} - Q_{yy}, 2Q_{xy})}{Q_{11} + Q_{22} + 2\sqrt{|Q|}}. \quad (4)$$

It is worth bearing in mind that there are in fact two common ellipticity definitions used for weak lensing. The one defined above is equivalent to an ellipticity magnitude, written in terms of (projected) axis ratios, $e = (a-b)/(a+b)$; for detailed discussion of both this and the alternative ellipticity definition, and their respective advantages, see Melchior & Viola (2012). Note that this is a Cartesian projection along one axis of the simulation box, not a light-cone projection with conversion to angular coordinates. The positive and negative e_1 direction, then, is defined by the x , y coordinate directions of the square simulation volume. Although this measurement does not correspond directly to what one could do in reality, the difference is not thought to be significant, given the statistical size and other limitations of the samples considered in this work.

3.2 Two-point correlations

All correlation functions used in this paper are computed using the public HALOTOOLS package³⁴ (Hearin et al. 2017). The most straightforward (and highest signal-to-noise) two-point measurement one could make is that of galaxy clustering in three dimensions. We adopt a common estimator of the form (Landy & Szalay 1993):

$$\xi_{gg}^{ij}(r_p, \Pi) = \frac{D^i D^j - D^i R^j - D^j R^i + R^i R^j}{R^i R^j}, \quad (5)$$

where DD , RR , and DR are weighted counts of galaxy–galaxy, random–random, and galaxy–random pairs, binned in perpendicular and line-of-sight separation, r_p and Π . The indices i, j denote a pair of catalogues (either galaxy positions, or random points), which are correlated together. In both cases above, R represents the positions of a set of random points drawn from a flat distribution within the simulation volume.

The cross-correlation of galaxy positions and intrinsic ellipticities, $\xi_{g+}(r_p, \Pi)$, can similarly be estimated, as a function of r_p and Π . We use a modified Landy–Szalay estimator of the form:

$$\xi_{g+}^{ij}(r_p, \Pi) = \frac{S_+^i D^j - S_+^i R^j}{R^i R^j} \quad (6)$$

(see Mandelbaum et al. 2011). One can similarly measure the shape–shape correlation:

$$\xi_{++}^{ij}(r_p, \Pi) = \frac{S_+^i S_+^j}{R^i R^j}. \quad (7)$$

The terms in the numerator represent shape correlations and are defined as

$$S_+^i D^j \equiv \frac{1}{2} \sum_{\alpha \neq \beta} w_\alpha w_\beta e_+(\beta|\alpha), \quad (8)$$

$$S_+^i S_+^j \equiv \frac{1}{4} \sum_{\alpha \neq \beta} w_\alpha w_\beta e_+(\alpha|\beta) e_+(\beta|\alpha), \quad (9)$$

where the indices α, β run over galaxies and $e_+(\beta|\alpha)$ is the tangential ellipticity of galaxy β , rotated into the coordinate system defined by the separation vector with galaxy α . For the fiducial catalogue, ILLUSTRISTNG, the weights are equal and normalized to the number of galaxies. In order to make a direct comparison of the different samples, galaxies in the MASSIVEBLACK-II and ILLUSTRIS-1 catalogues are assigned weights, such that the host halo mass distributions of the three match. For detail about the weighting scheme, which we refer to as halo-mass reweighting, and discussion about the impact on our results, we refer the reader to Appendix C. IAs are known to be dependent on cosmology and the host halo mass distribution, and this process should remove differences due to discrepancies in these factors. It is also true, however, that other properties such as the details of the galaxy–halo connection and the properties of the galaxies themselves also potentially have an impact. Such differences represent a systematic uncertainty (since we cannot say with certainty which of the simulations, if any, represents reality, nor straightforwardly homogenize them), and so any resulting differences should be treated as such.

In lensing studies it is also common to assign galaxies something approximating inverse variance (shape noise + measurement uncertainty) weights (see for example, Zuntz et al. 2018). Since this weighting tends to upweight bright, high S/N galaxies, it seems likely it would also boost the IA signal. That said, in practice lensing

³https://github.com/duncandc/halotools_ia

⁴<https://halotools.readthedocs.io/v0.7>

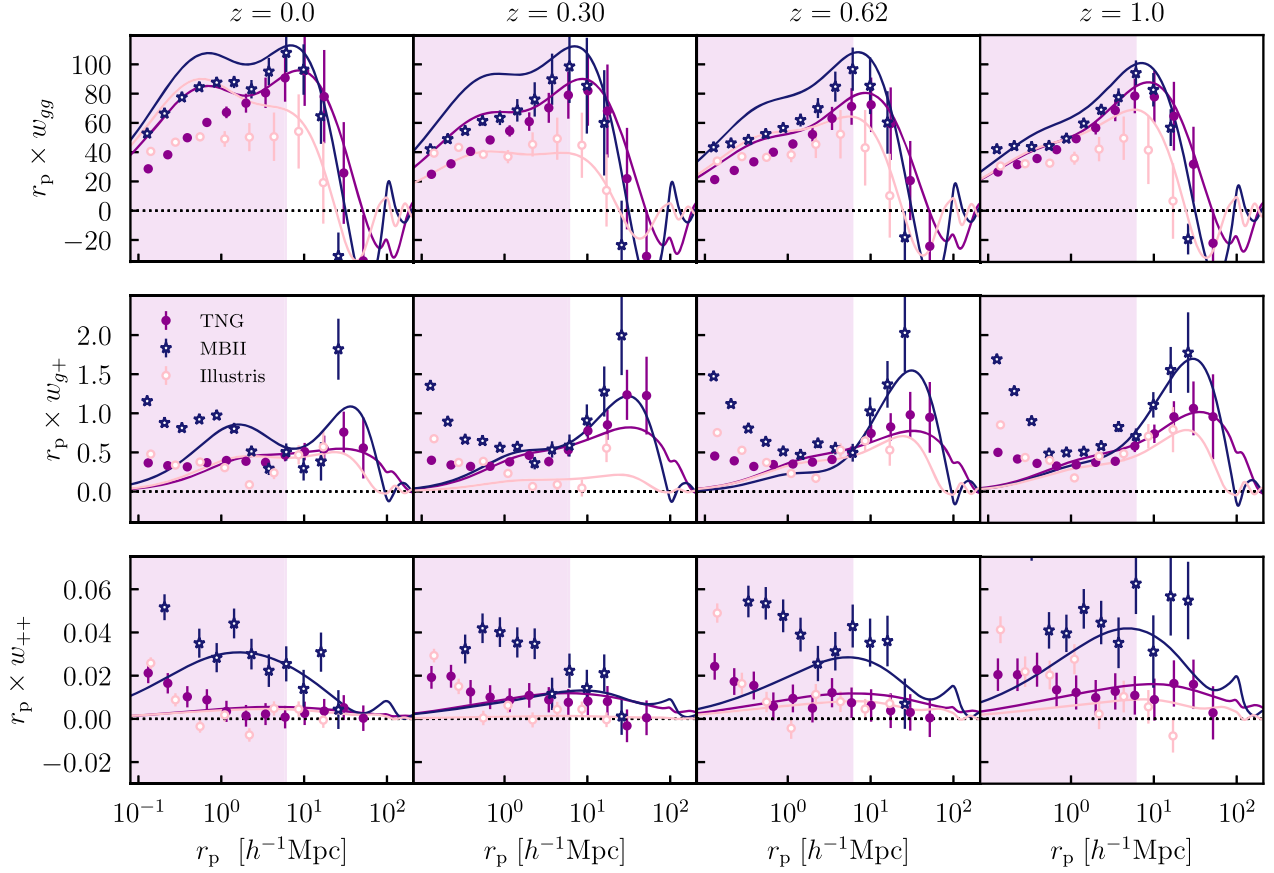


Figure 5. The fiducial data vectors used in this work. Shown from top are galaxy–galaxy, galaxy–shape, and shape–shape two-point correlations, at four discrete redshifts (left to right, as indicated). The point styles indicate measurements made on MASSIVEBLACK-II (dark blue stars), ILLUSTRIS-1 (pink open circles) and ILLUSTRISTNG (purple filled circles). The solid lines are the theory predictions, evaluated at the best fitting point in the TATT parameter space for each data set. Scales within the shaded regions ($r_p < 6 h^{-1} \text{ Mpc}$) are excluded from the fits, using both the TATT and NLA models.

weights tend to be shape noise dominated, and so relatively uniform across the sample, meaning the magnitude of this effect is expected to be small.

From these 3D measurements, obtaining the 2D projected correlations is a case of integrating along the line of sight. One has,

$$w_{ab}(r_p) = \int_{-\Pi_{\max}}^{\Pi_{\max}} d\Pi \xi_{ab}(r_p, \Pi). \quad (10)$$

Here the lower indices ab denote a type of two-point correlation, $a, b \in (g, +)$. Π_{\max} is an integration limit, which is set by the simulation volume. For this study we adopt a value equal to a third of the box size, or $\Pi_{\max} = 68 h^{-1} \text{ Mpc}$ for ILLUSTRISTNG, $\Pi_{\max} = 33 h^{-1} \text{ Mpc}$ for MASSIVEBLACK-II, and $\Pi_{\max} = 25 h^{-1} \text{ Mpc}$ for ILLUSTRIS-1. In practice, for our purposes we wish to maximize Π_{\max} ⁵; although it is true that very long baselines will eventually harm the signal-to-noise by including uncorrelated pairs, on scales of a few tens of Mpc we are well within the regime where extending Π_{\max} helps us to access

additional large scale signal modes (see Joachimi et al. 2011’s App. A2 for further discussion).

The measurements described above, for the three simulations, are shown in Fig. 5 (left to right ascending in redshift, top to bottom different correlations). Although by-eye comparison is complicated by the differing box size (which translates into an effective k limit that alters the relative amplitudes of the measured correlation functions, albeit in a way that is included in the model) there are a few notable features here. On large scales, the shape and amplitude of the large scale galaxy–galaxy correlation roughly matches between ILLUSTRISTNG and MASSIVEBLACK-II (although ILLUSTRIS-1 is somewhat lower and flatter). This is much less true of w_{g+} ; noticeably MASSIVEBLACK-II exhibits a much stronger scale-dependent signal on scales approaching the one halo regime. On large scales too there are hints at systematic differences in shape, although we are somewhat limited here by cosmic variance.

3.3 Tidal and shape fields

In addition to the two-point measurements described above, we also implement a new method to derive IA constraints at the field level. We refer the reader to Section 7 for details, but the method involves deriving constraints on IA parameters via a comparison of the (pixelized) 3D tidal field and the intrinsic galaxy shape field (see also Hilbert et al. 2017, who also use the tidal field directly to measure IAs, albeit via two-point functions, and Bate et al. 2020, who use

⁵The upper limit here is imposed to avoid artefacts due to the finite size of the simulations. In principle, one could use even smaller Π_{\max} , which would have two effects: it would suppress the amplitude of the signal due to the loss of correlated galaxy pairs, and also increase shot and shape noise. The former can be incorporated into the model, and so is not an issue; the latter, on the other hand, inescapably reduces our ability to draw conclusions from the measurements.

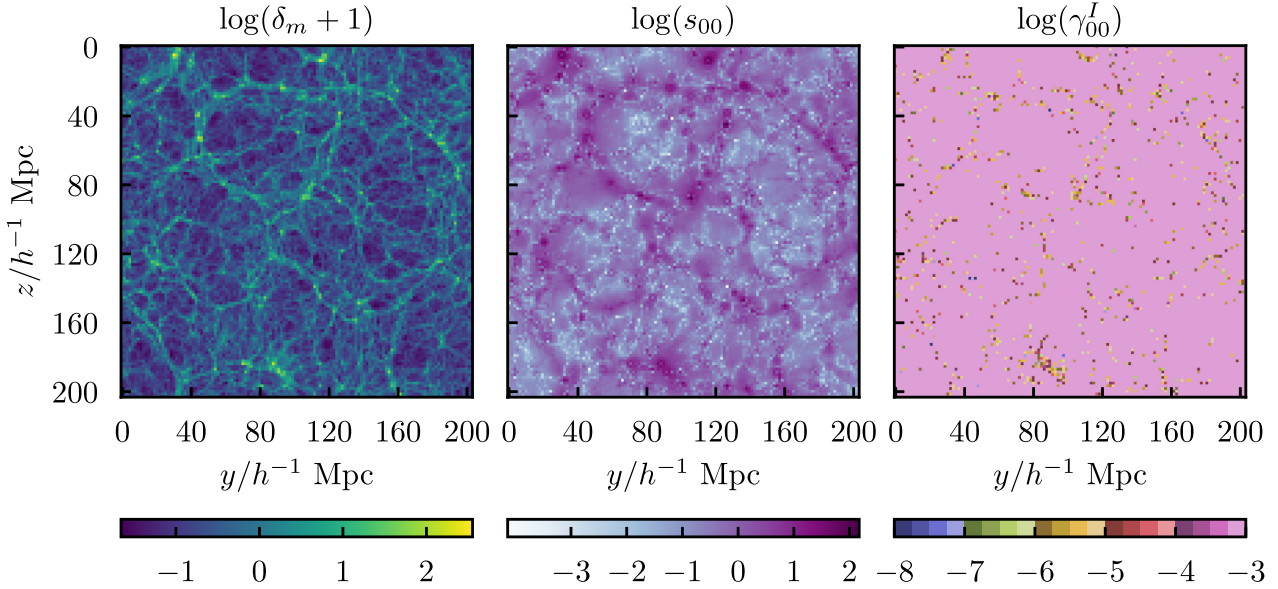


Figure 6. The $z = 0$ dark matter overdensity field and associated quantities from ILLUSTRISTNG. Here we show (left to right) the matter overdensity $\delta_m + 1$, the 0,0 component of the 3×3 dark matter tidal tensor, and the same component of the 3×3 galaxy shape tensor γ^I . The pixel resolution is 128/side, resulting in a physical pixel scale of $1.59 h^{-1}$ Mpc, which is at the finer end of the range of pixel scales presented in this work. While the full simulation box is clearly 3D, for illustrative purposes we choose here to show a 2D slice through the centre. In the right-hand panel, white pixels indicate those containing no galaxies that pass cuts into our final ILLUSTRISTNG shape catalogue.

similar measurements to explore the alignment of elliptical galaxy progenitors). To this end, we need an estimate of that tidal tensor as a function of position; we obtain this from the gridded particle data as follows.

Starting with the table of particle positions at fixed redshift, we divide the simulation box into 3D cubic pixels. The pixel size L is an unconstrained analysis variable, and affects the physical interpretation of the eventual results. We choose to perform our measurements using three different scales, 16 pixels across ($L \sim 12.8 h^{-1}$ Mpc), 32 pixels across ($L \sim 6.4 h^{-1}$ Mpc), and 64 pixels ($L \sim 3.2 h^{-1}$ Mpc). Within each pixel p in the grid, centred at position \mathbf{x}_p , we measure the overdensity of matter and stars,

$$\delta(\mathbf{x}_p) = \frac{N_p}{\langle N_p \rangle_p} - 1. \quad (11)$$

That is, the total number of dark matter particles in pixel p , divided by the mean occupation across all pixels. In the case of dark matter, all particles in ILLUSTRISTNG are weighted equally, and the values in the equation above are raw number counts, rather than sums of masses. Using the Fourier space version of the Poisson equation, one can show that the traceless tidal tensor can be obtained from the overdensity field as:

$$s_{ij}(\mathbf{k}) = \left(\frac{k_i k_j}{k^2} - \frac{1}{3} \delta_{ij} \right) \delta(\mathbf{k}), \quad (12)$$

where $k^2 = k_1^2 + k_2^2 + k_3^2$. For more details about the mathematics see Catelan & Porciani (2001), Alonso, Hadzhiyska & Strauss (2016). The two indices here ij denote a single element of the 3×3 tensor matrix within pixel p .

We also obtain a noisy estimate for the intrinsic shear in pixel p , $\gamma^I_{ij}(\mathbf{x}_p)$, by averaging the trace-free inertia tensors of galaxies within it. That is,

$$\gamma^I_{ij}(\mathbf{x}_p) = \left\langle I_{ij,k} - \frac{1}{3} \delta_{ij} \text{Tr}[\mathbf{I}_k] \right\rangle_k, \quad (13)$$

where the subscript k denotes a particular galaxy from pixel p , and the angle brackets $\langle \rangle_k$ indicate averaging over those galaxies. We estimate the per-element variance of the 3×3 matrix γ^I directly by computing the RMS over all galaxies; that is, we assume that shape noise dominates, such that the covariance matrix is diagonal, and can be written as $C_{ij,p}^{-1} = \delta_{ij} \sigma_{\text{SN}\mu}^{-2}$, or the inverse square shape variance for component $\mu \in (1, 2)$. Note that this is a global quantity, computed across pixels and applied to each of them. We confirm that the covariance scales with pixel size roughly as one might expect from geometric arguments as $\sigma_{\text{SN}\mu} \propto L^{-3/2}$. A 1D slice of the three fields described here, as measured in the $z = 0$ ILLUSTRISTNG snapshot, can be seen in Fig. 6. Shown are (left to right): dark matter overdensity, the upper diagonal element of the dark matter tidal tensor, and the smoothed galaxy shape field. It is apparent from Fig. 6 that there is an obvious qualitative correspondence between the raw matter field and the tidal tensor (compare the left-most and middle panels). The sampling of galaxies is much sparser, which is evidenced by the amount of white space in the right-most panel. Depending on the pixel scale, the fraction of unoccupied pixels is between 20 per cent and 80 per cent. Although striking in this figure, and worth noting, the impact of this sampling is explicitly incorporated into our IA modelling, as described in Section 4.

4 THEORY

Our analysis pipeline is built within COSMOSIS⁶ (Zuntz et al. 2015). The new modules introduced in this paper has been validated against older free-standing code. Although we will not discuss this process in detail here, a longer discussion can be found in Appendix A. Sampling is performed using MULTINEST (Feroz et al. 2019), and in the subset of chains where the Bayesian evidence is needed, we also run using POLYCHORD (Handley, Hobson & Lasenby 2015), with

⁶<https://bitbucket.org/joezuntz/cosmosis>, v1.6; master branch

Table 3. Free parameters for the IA model fits implemented in this work. All fits are performed on a single snapshot, with the same priors applied irrespective of redshift. Note that the linear galaxy bias b_g is not an IA parameter (i.e. it does not enter either the GI or II power spectra), but it is included in the modelling and so is shown here. The choice of priors here is designed to be conservative, and well clear of the posterior edges. We discuss the possible impact of this choice, and demonstrate robustness to projection effects in Section 5.

Model	Parameter	Prior
NLA	A_1	U[−6, 6]
	b_g	U[0.05, 8]
TATT	A_1	U[−6, 6]
	A_2	U[−6, 6]
	b_{TA}	U[−6, 6]
	b_g	U[0.05, 8]

more stringent accuracy settings.⁷ In all cases, we fix the cosmology to the input for the relevant simulation, with the parameters given in Table 1 and zero neutrino mass. The matter power spectrum is generated using CAMB with non-linear modifications from HALOFIT (Takahashi et al. 2012). A simulation of finite box size (i.e. any simulation) has an effective k limit, at which the power spectra are truncated (see Power & Kneib 2006 and Bagla, Prasad & Khandai 2009 for discussion and quantification), an effect that primarily impacts large physical scales, but potentially has ramifications at smaller separations too. In order to avoid biasing our results, we explicitly include this truncation in our modelling. Given the box sizes, the actual effective small- k cutoff is at $k_{\min} = 2\pi/L$, or $\sim 0.03 h^{-1} \text{ Mpc}$, $\sim 0.06 h^{-1} \text{ Mpc}$, and $\sim 0.08 h^{-1} \text{ Mpc}$ in the cases of ILLUSTRIS, MASSIVEBLACK-II, and ILLUSTRIS-I, respectively. We assess the impact of this detail by repeating our fits with fixed $k_{\min} = 0.06 h^{-1} \text{ Mpc}$ for the three simulations. The resulting biases, arising from ignoring the small- k cut off, is potentially quite significant (~ 20 per cent) in both the galaxy bias and IA parameters.

Our fiducial analysis includes physical scales in the range $6 < r_p < L/3 h^{-1} \text{ Mpc}$, where L is the length of the simulation box. Unlike in real survey data, an upper cut is necessary to avoid edge effects due to the finite simulation size. The lower cut follows several other studies (Joachimi et al. 2011; Singh et al. 2015; Johnston et al. 2019), and is intended to be conservative in removing data affected by non-linear bias. We explicitly test this choice in Section 6.1.

4.1 Modelling intrinsic alignments

We consider two different IA scenarios in our fits, discussed in more detail below. While it is useful to think of these as entirely separate models, and indeed we will refer to them as such, it is worth bearing in mind that they are nested. That is, the more complex model reverts to the simpler one when a subset of its parameters are zero. For reference, the free parameters in each of these models and the associated priors in each case are shown in Table 3.

4.1.1 Non-linear alignment model

One common predictive IA model is the non-linear alignment (NLA) model; in essence, it is an empirically motivated modification (see Bridle & King 2007; Hirata et al. 2007) to a physically motivated (at least partially, in certain regimes) prescription known as the Linear

Alignment (LA) model (Catelan et al. 2001; Hirata & Seljak 2004; Hirata & Seljak 2010; Blazek, McQuinn & Seljak 2011). Under the assumption of linear alignments, one can write the intrinsic shape of a galaxy in terms of the background gravitational potential at the time of galaxy formation as:

$$(e_+^I, e_-^I) = -\frac{\bar{C}_1}{4\pi G} \left(\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2}, 2\frac{\partial^2}{\partial x \partial y} \right) \phi(\chi_*), \quad (14)$$

where \bar{C}_1 is a normalization constant, typically fixed at a value of $5 \times 10^{-14} M_\odot^{-1} h^{-2} \text{ Mpc}^3$ (Brown et al. 2002). Following Hirata & Seljak (2004), the GI and II power spectra have the form:

$$P_{\text{GI}}(k) = -\frac{\bar{C}_1 \bar{\rho}(z)}{D(z)} a^2(z) P_\delta^{\text{lin}}(k) \quad (15)$$

and

$$P_{\text{II}}(k) = \left(\frac{\bar{C}_1 \bar{\rho}(z)}{D(z)} \right)^2 a^4(z) P_\delta^{\text{lin}}(k). \quad (16)$$

Here $\bar{\rho}$ is the (spatially averaged) mean matter density of the Universe and D is the linear growth function. The model also predicts higher order contributions, as well as non-zero B modes arising from galaxy clustering, though these are typically neglected in implementations of the NLA model (Hirata & Seljak 2004; Blazek et al. 2015; see the next section for further discussion). We follow many previous analyses in fixing \bar{C}_1 to Brown et al. (2002)’s value, and parametrizing deviations in strength of alignment from this baseline with a free amplitude, such that $P_{\text{GI}} \rightarrow A_1 P_{\text{GI}}$ and $P_{\text{II}} \rightarrow A_1^2 P_{\text{II}}$.

The feature that defines the NLA is the substitution of the linear power spectrum in equations (15) and (16) for the non-linear version. The rationale for this change is as an attempt to capture the non-linear tidal field, and indeed it does appear to improve the performance on small to intermediate scales (see for example Bridle & King 2007; Blazek et al. 2015; Singh et al. 2015), even if it is not necessarily internally consistent.

4.1.2 Tidal alignment and tidal torque model

Our second IA model, referred to as the Tidal Alignment + Tidal Torque (TATT) model, was first proposed by Blazek et al. (2019) and has been employed a number of times in the context of cosmic shear analyses in the recent past (see Troxel et al. 2018; Samuroff et al. 2019). We will provide a brief overview of the theory, and refer the reader to those papers a more detailed description.

In this framework, a galaxy’s intrinsic shape⁸ is written as an expansion in the trace-free tidal field tensor s_{ij} :

$$\gamma_{ij}^I = \underbrace{C_1 s_{ij}}_{\text{Tidal Alignment}} + \underbrace{C_{18}(\delta \times s_{ij})}_{\text{Density Weighting}} + \underbrace{C_2 \left[\sum_{k=0}^2 s_{ik} s_{kj} - \frac{1}{3} \delta_{ij} s^2 \right]}_{\text{Tidal Torquing}} + \dots, \quad (17)$$

with both sides of the equation evaluated at a position \mathbf{x} , which may be either a Lagrangian or an Eulerian position. The two amplitudes C_1 and C_2 describe the magnitude of alignment due to tidal alignment and tidal torquing, respectively. It is worth bearing in mind, however,

⁸The intrinsic shape here is defined in an analogous way to the projected ellipticity; it is the trace-free component of the moment matrix in three dimensions, or equivalently, the eigenvector matrix of the 3D inertia tensor. As noted in Blazek et al. (2019), it is not a uniquely defined quantity, and depends on the radial weighting of the measurement algorithm.

⁷live_points = 250, tolerance = 0.1, num_repeats = 60

that these terms can absorb IAs due to other mechanisms when fit to real data; for example, an effective non-zero C_1 can in principle arise in a pure tidal torquing IA scenario, when combined with non-linear growth (Hui & Zhang 2002). The term, with the coefficient $C_{1\delta}$, is a so-called density weighting contribution, and arises from the fact that one can only measure galaxy shapes in a position where there is actually a galaxy (see e.g. Hirata & Seljak 2004 and Blazek et al. 2015 for further discussion). Also note that the product of the matter overdensity and tidal fields δs_{ij} implicitly assumes a smoothing scale, a detail we will return to in Section 7. The real-space dark matter tidal tensor is a 3×3 matrix, defined in relation to the overdensity field in equation (12). If the tidal tensor is computed using the non-linear matter field, then the leading term in equation (17) is equivalent to the NLA prediction. If the TATT model parameters are varied together, however, they can enter the data in potentially degenerate ways, meaning that the A_1 part of the full TATT space will not necessarily match the NLA fit to the same data, if $A_2 \neq 0$ is preferred. One then has:

$$C_1 = -A_1 \bar{C}_1 \frac{\Omega_m \rho_{\text{crit}}}{D(z)}, \quad (18)$$

$$C_2 = 5A_2 \bar{C}_1 \frac{\Omega_m \rho_{\text{crit}}}{D^2(z)}. \quad (19)$$

and

$$C_{1\delta} = -A_{1\delta} \bar{C}_1 \frac{\Omega_m \rho_{\text{crit}}}{D(z)}. \quad (20)$$

The constant \bar{C}_1 is the same as the one discussed in the previous section. The IA power spectra (GI and II) are derived from perturbation theory and are given by integrals over the matter power spectrum; for details see sections A–C of Blazek et al. (2019). Our version of the TATT model is identical to that of Troxel et al. (2018), Blazek et al. (2019) and Samuroff et al. (2019). It makes use of the FAST-PT code (McEwen et al. 2016; Fang et al. 2017), and is implemented within COSMOSIS.

Following Blazek et al. (2015), we do not vary $A_{1\delta}$ directly, but rather assume the density weighting term is related to the tidal alignment amplitude via a coefficient (i.e. $C_{1\delta} = b_{\text{TA}} C_1$). The original motivation for this parametrization was that IA correlations scaling with δs_{ij} were generated by the density weighting of the IA field, which can only be observed where galaxies are located (see Blazek et al. 2015 for a more detailed discussion). As with the other terms, $C_{1\delta}$ can be thought of more generally as describing any alignment physics with large-scale correlations that depend on δs_{ij} , and so does not necessarily correspond directly to the galaxy bias constrained by w_{gg} , as per the simple density weighting picture. Indeed, in a linear and ‘local Lagrangian’ picture of IA formation, in which intrinsic galaxy shapes are a linear function of the local tidal field initially present where the halo (and galaxy) form, a $C_{1\delta} \sim C_1$ term will be generated by the advection of galaxies between the Lagrangian and Eulerian frames Schmitz et al. (2018). Given the potential for other physical effects to be captured by the same term, it is safest to allow it to vary as a free parameter over a similar range to the other IA parameters (see Table 3). Previous studies have chosen to fix it to unity (Troxel et al. 2018; Blazek et al. 2019; Samuroff et al. 2019), based on physical arguments. In these cases, however, the density weighting term has been subdominant, allowing only very broad constraints on b_{TA} ; Samuroff et al. (2019), show that the decision to fix it was not a significant source of uncertainty in the context of DES Y1 3×2 pt cosmology. This is likely to be less true for our direct IA measurements.

Finally, we note that the TATT model predicts a non-zero IA-induced B-mode term, which enters the II power spectrum, and is sensitive to $C_{1\delta}$ and C_2 (see Blazek et al. 2019, equations 37–39). These contributions are included in our modelling of w_{++} . Again, we demonstrated in Samuroff et al. (2019) (Appendix C) that this choice has negligible impact on parameter constraints in the context of a DES Y1 3×2 pt analysis. This is not trivially true for the type of measurement considered in this work, and so we include the extra B-mode terms when fitting the TATT model here.

4.1.3 Modelling two-point correlations

Given an IA power spectrum from either of the models described, one can predict the projected correlation functions at fixed redshift via Hankel transforms. Under the Limber approximation one has:

$$w_{g+}^{z_s}(r_p) = -b_g^{z_s} \int \frac{dkk}{2\pi} J_2(kr_p) P_{\text{GI}}(k, z = z_s), \quad (21)$$

with the z_s indicating a particular redshift (snapshot), and J_2 being a second-order Bessel function of the first kind. We assume linear galaxy bias, $b_g \equiv \delta_g/\delta$, which is marginalized with a wide prior (Table 3). The range $b_g = [0.05, 8]$ is intended to be conservative, and the bias is always well constrained within these bounds. An important thing to note here, however, is that in a high dimensional parameter space typical of cosmological analyses such wide priors can cause shifts in the 2D constraints via projection effects (see e.g. Joachimi et al. 2020; Secco, Samuroff et al. 2021 for discussion); in our relatively simple setup we do not expect this to be an issue. We verify this in our fiducial ILLUSTRATING TATT analysis by reducing the b_g prior width to $[0.05, 4]$, and confirm it does not alter our results. A similar exercise, halving the volume of the prior on the less well constrained b_{TA} again has no significant impact.

In real data one would also need to evaluate an integral over a redshift kernel, defined by the sample’s redshift distribution (Mandelbaum et al. 2011’s Appendix A); in our case this reduces to evaluating $P_{\text{GI}}(k)$ at a particular redshift z_s . The other two-point correlations follow by analogy as:

$$w_{gg}^{z_s}(r_p) = b_g^{z_s} b_g^{z_s} \int \frac{dkk}{2\pi} J_0(kr_p) P_{\delta}(k, z = z_s), \quad (22)$$

and

$$w_{++}^{z_s}(r_p) = \int \frac{dkk}{2\pi} [J_0(kr_p) + J_4(kr_p)] P_{\text{II}}(k, z = z_s). \quad (23)$$

In the case where we are including an IA induced B-mode contribution, the above becomes a sum of two integrals (see e.g. Blazek et al. 2015, equation 2.8):

$$\begin{aligned} w_{++}^{z_s}(r_p) &= w_{++}^{z_s, \text{EE}}(r_p) + w_{++}^{z_s, \text{BB}}(r_p) \\ &= \int \frac{dkk}{2\pi} [J_0(kr_p) + J_4(kr_p)] P_{\text{II}}^{\text{EE}}(k, z = z_s) \\ &\quad + \int \frac{dkk}{2\pi} [J_0(kr_p) - J_4(kr_p)] P_{\text{II}}^{\text{BB}}(k, z = z_s). \end{aligned} \quad (24)$$

The TATT E and B mode power spectra here are given by Blazek et al. (2019)’s equations (38) and (39). In the NLA case $P_{\text{II}}^{\text{BB}} = 0$, and equation (24) reduces to equation (23).

Although we do not compute the 3D correlations $\xi_{ab}(r_p, \Pi)$, we do factor in the fact that the line of sight integral in the measurement has a finite limit Π_{max} (e.g. equation 10). The effect of this is to suppress the signal slightly, as correlated pairs are cut off. We can test the magnitude of this by comparing our observables at a fiducial point in parameter space with an external modelling code,

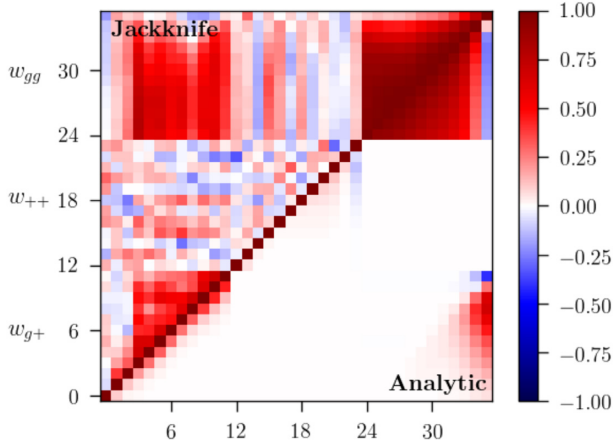


Figure 7. The correlation matrix for our fiducial ILLUSTRISTNG two-point measurements, as estimated using jackknife resampling (upper left-hand triangle), and an analytic Gaussian approximation (lower right). Note that the two covariance matrices are symmetric about the diagonal; the triangle configuration is shown here for illustrative purposes only.

which explicitly includes Π_{\max} . Since the impact is found to be independent of r_p on large scales, to the level of ~ 0.5 per cent, we incorporate it into our modelling as a single multiplicative factor μ , which we compute for each correlation function, at each redshift (i.e. 12 numbers per simulation). In the case of ILLUSTRISTNG, Π_{\max} is relatively large ($68 h^{-1}$ Mpc), and so the signal damping is only 1 – 2 per cent (which is comfortably subdominant to uncertainty). For MASSIVEBLACK-II and ILLUSTRIS-I ($\Pi_{\max} = 33$ and $\Pi_{\max} = 25 h^{-1}$ Mpc, respectively), however, μ is somewhat larger, which shifts the IA parameters upwards slightly. Although our qualitative conclusions are robust even without this correction, omitting it is seen to bias the A_1 and A_2 towards low values by ~ 10 –50 per cent.

4.2 Covariance matrix of two-point functions

In order to derive robust parameter constraints from our measurements we need a representative, numerically stable, estimate for the covariance matrix of those measurements. The full data vector consists of three two-point measurements for each of four snapshots; this gives us $N = 4 \times 3 \times N_p$ data points for each simulated galaxy sample (96, 144, and 192 in the case of ILLUSTRIS-I, MASSIVEBLACK-II, and ILLUSTRISTNG, respectively). Our fiducial covariance estimate is calculated analytically, a detail of this analysis that differs from many previous studies, most of which have opted for an internal covariance estimator such as jackknife resampling. The analytic approach has a number of advantages, not least the ability to extend to large scales where jackknife estimates break down. We show a comparison of our fiducial correlation matrix, calculated using the method described below, and a jackknife estimate in Fig. 7.

Although in principle the covariance has higher order contributions resulting from mode mixing (e.g. Krause, Eifler & Blazek 2016), given the limited statistical power of the simulations, and the fact that shot and shape noise tend to dominate on the scales we fit, the dominant Gaussian contribution is considered sufficient for our purposes. In the Gaussian approximation, a given element is the sum of a noise term and a cosmic variance contribution:

$$\text{Cov} [w_{\alpha\beta}^{z_s}(r_{p,i}), w_{\delta\gamma}^{z_s}(r_{p,j})] = C_{\alpha\beta\delta\gamma}^{\text{SN}, z_s z_s, kj} + C_{\alpha\beta\delta\gamma}^{\text{CV}, z_s z_s, kj}, \quad (25)$$

where the Greek indices denote correlation types i.e. $\alpha, \beta, \delta, \gamma \in (g, +)$; z_s identifies a particular redshift slice and j, k are

comoving scale bins. The cross-correlations between snapshots is potentially complicated, given that the galaxy properties are strongly (but not fully) correlated. However, since we will not attempt a fully simultaneous analysis across redshifts, but rather restrict our inference to one snapshot at a time, we will neglect these additional covariance terms. One can write each element as:

$$\begin{aligned} \text{Cov} [w_{\alpha\beta}^{z_s}(r_{p,i}), w_{\delta\gamma}^{z_s}(r_{p,j})] \\ = \delta_{ij} \frac{2\pi}{\mathcal{A}_p r_{p,i} \Delta r_p} \int k dk \Theta_{\alpha\beta}(kr_{p,i}) \Theta_{\delta\gamma}(kr_{p,i}) \\ \times [\tilde{P}_{\alpha\delta}(k, z_s) \tilde{P}_{\beta\gamma}(k, z_s) + \tilde{P}_{\alpha\gamma}(k, z_s) \tilde{P}_{\beta\delta}(k, z_s)], \end{aligned} \quad (26)$$

with the kernels

$$\Theta_{\mu\nu}(x) = \begin{cases} J_2(x) & \mu\nu = g+ \\ J_0(x) & \mu\nu = gg \\ J_0(x) + J_4(x) & \mu\nu = ++ \end{cases}, \quad (27)$$

where J_ν is a Bessel function of the first kind of order ν . In IA measurements on real data \mathcal{A}_p is a function of redshift, and accounts for the survey mask; in our case it is simply the cross-sectional area of the simulation box in h^{-2} Mpc $^{-2}$. One should also note that the power spectra here are subject to a noise contribution,

$$\tilde{P}_{\alpha\beta}(k, z_s) = P_{\alpha\beta}(k, z_s) + N_{\alpha\beta}^{z_s}, \quad (28)$$

where $N_{\alpha\beta}^{z_s} = 1/n_c^{z_s}$ for $\alpha\beta = gg$, $N_{\alpha\beta} = (\sigma_e^{z_s})^2/n_c^{z_s}$ for $\alpha\beta = ++$, and $N_{\alpha\beta} = 0$ for $\alpha\beta = g+$. The denominator n_c is the comoving volume number density of the sample at $z = z_s$ in h^3 Mpc $^{-3}$, and $\sigma_e^{z_s}$ is the projected ellipticity dispersion.

As is apparent from the above, the analytic covariance matrix is sensitive to some extent on the input parameter values (cosmology, galaxy bias, and IAs). As stated before, cosmological parameters are fixed to those appropriate for the simulation in question, as per Table 1. For the other (IA and bias) parameters, we generate the fiducial matrix for each sample using an iterative procedure similar to that of Krause et al. (2017). That is, we repeatedly fit the data to obtain IA and galaxy bias parameter constraints, update the covariance matrix and fit again. Our convergence criteria are that (i) the marginalized 1D parameter posteriors are not systematically different between iterations, and (ii) the χ^2 and evidence values are stable to within a few percent. In all samples, the covariance matrix converges within 2–4 iterations.

We also test our fiducial analytic covariance matrix against a version computed using jackknife resampling. In brief, the jackknife method involves dividing the data into N spatial subregions, and repeating the measurement N times, each time removing one of them. The validity of this approach relies on various (potentially strong) assumptions; not least it assumes the subregions are statistically independent (see Anderson 2003; Hartlap, Simon & Schneider 2007 for discussion), and that the scales of interest are much smaller than scale of the subregions. These factors, combined with the relatively small number of subregions allowed by even ILLUSTRISTNG (the largest of the simulations considered here), are the primary reason we consider jackknife as an approximate test of, and not a viable alternative to, our analytic predictions. In the fiducial case (ILLUSTRISTNG), we divide the 3D box into $N_{\text{jk}} = 4^3 = 64$ cubic subvolumes. A visual comparison of the correlation matrices can be found in Fig. 7. We also compare the root diagonals of the two covariance matrices (see Fig. A2). Although there is approximate agreement between the two, the jackknife method tends to underestimate the variance on virtually all scales in the three correlations. On the relevant scales for our fits ($> 6 h^{-1}$ Mpc), the differences are at the level of up to ~ 25 –50 per cent in w_{g+} .

5 IA CONSTRAINTS FROM TWO-POINT MEASUREMENTS

As discussed, our baseline methodology is to fit the joint data vector of w_{gg} , w_{g+} and w_{++} simultaneously for a given simulation and at a given redshift. In this section we present the results of these likelihood analyses. This approach is analogous to cosmological inference using 3×2 pt data, with the significant difference that our parameter space is several times smaller (and does not include cosmological parameters). It carries a number of advantages, not least benefitting from some level of complementarity in the degeneracies of the different data vector elements.

We perform our IA model fits to each of the four redshift snapshots independently, a choice primarily driven by the covariance matrix; unlike in real data, where each galaxy can be assigned (albeit not necessarily correctly) to a single tomographic bin, here we effectively have one realization of the galaxy field, which is evolved with redshift. The galaxy population, the shape noise, and the cosmic variance are, then, potentially heavily correlated between redshifts, which makes a fully simultaneous analysis complicated. Modelling such correlations is non-trivial, and not considered a valuable exercise within the scope of this paper.

As discussed, our baseline methodology is to fit the joint data vector of w_{gg} , w_{g+} and w_{++} simultaneously for a given simulation and at a given redshift. In this section we present the results of these likelihood analyses. This approach is analogous to cosmological inference using 3×2 pt data, with the significant difference that our parameter space is several times smaller (and does not include cosmological parameters). It carries a number of advantages, not least benefitting from some level of complementarity in the degeneracies of the different data vector elements.

We perform our IA model fits to each of the four redshift snapshots independently, a choice primarily driven by the covariance matrix; unlike in real data, where each galaxy can be assigned (albeit not necessarily correctly) to a single tomographic bin, here we effectively have one realization of the galaxy field, which is evolved with redshift. The galaxy population, the shape noise and the cosmic variance are, then, potentially heavily correlated between redshifts, which makes a fully simultaneous analysis complicated. Modelling such correlations is non-trivial, and not considered a valuable exercise within the scope of this paper. That said, the results presented here are the first time a robust comparison has been attempted with a homogenized sample, using w_{g+} and w_{++} simultaneously, and with an analytic covariance matrix that is numerically stable on large scales.

5.1 NLA & TATT

The posteriors from NLA model fits to the various simulations at $z = 0$ are shown in the upper panel of Fig. 8. As described in Section 3.2, the MASSIVEBLACK-II and ILLUSTRIS-I samples are reweighted, such that the halo mass distributions match (see also the discussion in Appendix C, where we demonstrate the importance of this reweighting). This process is designed to allow meaningful comparison between simulations by ensuring that differences in halo mass distribution are not driving the offset in the IA-bias parameter space. As noted above, the halo mass weighting is not guaranteed to eliminate all differences due to sample composition arising from differences in the galaxy halo connection in the simulations. For clarity, we do not show the three other snapshots at $z > 0$, but note that very similar qualitative trends are seen out to $z = 1$.

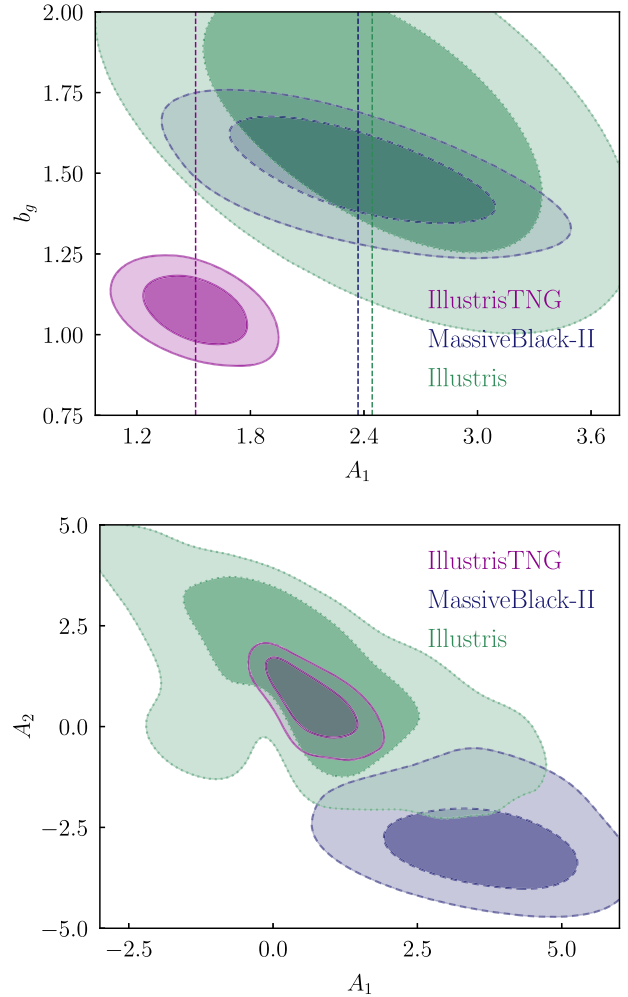


Figure 8. 1σ and 2σ confidence contours from our NLA (top) and TATT (bottom) model fits to various hydrodynamic simulations at $z = 0$. Shown are ILLUSTRISTNG (purple, solid), ILLUSTRIS-I (green, dotted), and MASSIVEBLACK-II (blue, dashed). Note that the MASSIVEBLACK-II and ILLUSTRIS-I samples here are weighted, such that the distributions of host halo masses match between the simulations, in order to allow meaningful comparison with ILLUSTRISTNG (see Section 3.2). The three vertical lines in the NLA panel show the A_1 values of the respective posterior peaks; these best-fitting values are $A_1^{\text{TNG}} = 1.51$, $A_1^{\text{MBII}} = 2.37$, and $A_1^{\text{I}} = 2.44$. The three simulations are consistent in the NLA space to $\sim 1\sigma$, although some mild disagreement is seen in the case of the more complex model.

The galaxy bias (horizontal axis, upper panel) differs slightly between the different simulations, at the $1-2\sigma$ level. Although not shown for TATT, the marginalized posterior on bias is close to independent of the choice of IA parametrization, primarily because w_{gg} dominates the constraint. The relative agreement between the detected NLA signal in the different simulations here is interesting, in the context of existing literature. It has been observed (Chisari et al. 2016; Tenneti et al. 2016) that MASSIVEBLACK-II tends to prefer a slightly stronger IA amplitude than ILLUSTRIS-I. This conclusion is supported at some level here; in the NLA case MASSIVEBLACK-II favours slightly larger A_1 values than either ILLUSTRISTNG or ILLUSTRIS-I in all but the lowest redshift snapshot (Fig. 9 below).

The joint posteriors on the TATT model amplitudes A_1 , A_2 are shown in the lower panel of Fig. 8 (see also Appendix B for the full

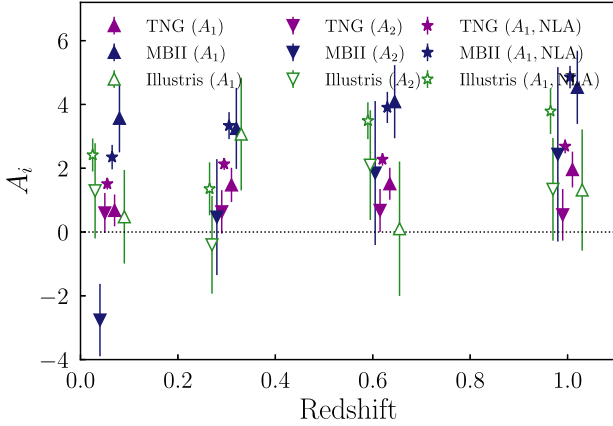


Figure 9. The redshift evolution of the two TATT model amplitudes in our three simulated data sets. The fits were performed on each redshift slice independently. The simpler NLA fits are also shown for reference (stars). The results from MASSIVEBLACK-II (dark blue), ILLUSTRIS-I (green), and ILLUSTRISTNG (purple) are shown. Note that the galaxy samples for a given simulation at different redshifts strongly overlap, and so the errors are potentially highly correlated, to an extent not reflected in the 1σ error bars in this figure.

TATT posteriors from the three simulations). These amplitudes can be thought of as controlling the strength of different IA contributions, which are linear and quadratic in the tidal field, respectively. Note that the TATT fits also include additional parameters (b_{TA} and linear galaxy bias b_g), which are marginalized in this 2D representation (see Section 4.1.2 and Table 3). In this limited parameter space we do not believe our marginalized results to be significantly affected by prior volume effects (e.g. the discussion in Joachimi et al. 2020). We confirm that rerunning the TATT chains with a reduced b_{TA} prior $U[0, 6]$ does not qualitatively change the TATT posteriors. In the case of MASSIVEBLACK-II and ILLUSTRIS-I, the constraint is weaker compared with ILLUSTRISTNG, to the extent that quite different TATT IA scenarios are allowed within 1σ . That ILLUSTRIS-I offers little-to-no constraint on the extended model is unsurprising; indeed we are fitting a small handful of relatively noisy points in the $>6 h^{-1}$ Mpc range, which provide no real information on the shape of the correlation function.

We now see some notable disagreement between the different simulations; whereas ILLUSTRISTNG favours a region of parameter space that resembles NLA (i.e. $A_2 \sim 0$), MASSIVEBLACK-II prefers $A_2 < 0$ at 3σ . While this could be a sign of a real alignment signal, generated by the physics models of MASSIVEBLACK-II, it is worth being cautious here; the TATT model will respond to any structure in the data, regardless of physical origin, and MASSIVEBLACK-II has known limitations.⁹ Inspecting the data vector (Fig. 5) more closely, it seems that the $A_2 < 0$ is driven by the gradual slope in the shape of the correlation functions between 1 and $10 h^{-1}$ Mpc. This feature is seen in both w_{g+} and w_{++} , and it does indeed seem to be relatively

well fit by the quadratic alignment contribution. It is also notable that there is no corresponding feature at around the same scale in w_{gg} .

In all cases we note that the data favour low values of b_{TA} , albeit with relatively large uncertainties. That said, the region of parameter space where one could reasonably interpret the TATT tidal alignment bias as a pure physical galaxy bias are still encompassed within the $\sim 1\sigma$ confidence bounds, with $b_{TA}^{TNG} = 0.73 \pm 1.75$, $b_{TA}^{MBII} = 0.25 \pm 0.73$. Interestingly, in the upper redshift bins MASSIVEBLACK-II prefers a weakly negative b_{TA} (Table B1), the physical interpretation of which is not immediately clear. Given the sample selection, and the limitations of the simulations, it is not obvious that the low b_{TA} values transfer to real lensing data, but it is interesting, in the sense that the data are (mostly) showing a preference for the simpler IA scenario.

From the ILLUSTRISTNG fits, the final posterior mean TATT parameter values at $z = 0$ are:

$$\begin{aligned} A_1^{TNG} &= 0.68 \pm 0.49, & A_2^{TNG} &= 0.60 \pm 0.62, \\ b_{TA}^{TNG} &= 0.73 \pm 1.75. \end{aligned} \quad (29)$$

The A_1 constraint here is consistent with the equivalent NLA amplitude from the two-parameter fits, a conclusion that largely holds across the three simulations. That is, switching to TATT leads to a degradation in the uncertainty on A_1 (by roughly 50 per cent for ILLUSTRISTNG at $z = 0$), but no significant shift in the favoured value. Remarkably, although MASSIVEBLACK-II favours negative $A_2 = -2.3 \pm 1.0$ at the level of $\sim 2 - 3\sigma$, ILLUSTRIS-I and ILLUSTRISTNG are consistent with zero across the redshift range. The small A_2 values differ slightly from recent studies on DES data (Troxel et al. 2018; Samuroff et al. 2019), which report a preference for $A_2 < 0$ (although our ILLUSTRISTNG constraint is still at most $\sim 2\sigma$ from the DES Y1 mixed sample; Samuroff et al. 2019 fig. 12). Note however that in such analyses on photometric data like the studies cited above, where the two-point functions are measured in broad redshift bins as a function of angular scale, a significant amount of mode-mixing can occur. Although this is included in the modelling, and so is not a cause for concern, it does mean the mapping between minimum θ and a minimum physical scale is not one-to-one. In addition to this, it is worth bearing in mind that no analysis on real data can ever be perfect; despite various robustness tests and validation carried out for DES Y1, we cannot altogether rule out the leakage of other modelling errors (e.g. in the photometric redshift distributions) into the IA constraints. For these, amongst other, reasons that it is not trivial to extrapolate from our results to comment on the detectability of higher order IA contributions in real data.

The lack of a clear detection of higher order alignment terms is not altogether surprising, given the relatively conservative scale cuts implemented here ($r_p > 6 h^{-1}$ Mpc; see also Section 6.1). Given the difference in physical scaling, naturally the alignment of galaxies on very large scales should resemble the tidal alignment scenario ($A_1 > 0$, $A_2 = 0$). Although we do not have a strong first-principles prediction of the scales on which the quadratic terms should become significant, we can make a rough estimate. Based on theory predictions, in scenarios that are consistent with previous observations (Samuroff et al. 2019), the regime where the tidal torquing terms are not totally subdominant to tidal alignment is somewhere on the scale of a few h^{-1} Mpc (see Blazek et al. 2015; Blazek et al. 2019). This places our fits in the marginal regime, where it is possible, but not certain that we might detect a non-NLA-like alignment signal.

⁹In particular, there is a lack of realistic spiral type galaxies, and a relative overabundance of diffuse elliptical objects compared with data. Due to relatively weak AGN feedback, MASSIVEBLACK-II produces an overpredicts the number of massive galaxies at low redshift (Khandai et al. 2015). Another manifestation of this is seen in the impact of baryons on the non-linear matter power spectrum, which is significantly different from that in other recent hydrodynamic simulations compared by Huang et al. (2019), as shown in their fig. 1.

5.2 Evolution with redshift

To illustrate the redshift evolution of the various IA parameters, we show the marginalized best fits and 1σ uncertainties in Fig. 9. As before, we show all three simulations in purple/blue/green. It is worth keeping in mind here that there is significant overlap between samples at different redshifts, meaning the shape noise is potentially quite strongly correlated. The interpretation of the various trends shown in this figure, then, are not trivial. That said, the basic patterns noted above are seen to hold across the redshift range. That is, with the partial, weak, exception of MASSIVEBLACK-II, the A_1 values obtained in the NLA and the TATT analyses are consistent with each other for a given simulation (compare the stars with the triangles in Fig. 9). The TA alignment amplitude rises more or less monotonically in ILLUSTRISTNG and MASSIVEBLACK-II, although the latter predicts a slightly higher NLA amplitude at all redshifts. With the extra freedom of the TATT model, we see some more striking differences, with MASSIVEBLACK-II favouring a higher A_1 by a factor of $\sim 3-4$ (although the upwards trend with redshift persists). This seems to fit with an underlying assumption of the linear alignment model: that IAs are frozen into a population of galaxies at early times (see Kirk et al. 2012; Kiessling et al. 2015; and Schmitz et al. 2018; particularly their App. A and references therein). As the underlying large-scale structure evolves and haloes grow, the subhalo mass distribution shifts upwards. In our case, then, the fixed stellar mass cut is more stringent, and removes a larger fraction of weakly aligned objects at high redshift than at low redshift. The net effect of this is an increase in the measured IA signal with increasing z . Though physically interesting, we reiterate that the flat lower mass cut at each redshift is not representative of the selection function in a real lensing catalogue. In real data with realistic flux- and shape-based quality cuts, the changes in composition with redshift will have a significant bearing on how the effective IA amplitude evolves. A step in this direction (albeit still not capturing the full complexity of a redshift-dependent selection in real data) would be to evolve a mass cut imposed at a given redshift by tracing the population through the simulation merger tree. Bhowmick et al. (2020) attempt such an exercise for MASSIVEBLACK-II, with results that are qualitatively consistent with this study. Although in the consistently traced sample (SAMPLE-TREE in their terminology) increasing halo-satellite misalignment tends to wash out alignments at high z , the impact of the changing population opposes, and largely outweighs this trend.

In contrast, A_2 is more or less constant with redshift in all simulations (the downward triangles in Fig. 9). Notably MASSIVEBLACK-II's preference for $A_2 < 0$ is not seen to persist across snapshots, although the interpretation of this is non-trivial. Particularly in the higher redshift slices, the MASSIVEBLACK-II posteriors exhibit significant bimodality, which appears to arise from a degeneracy between A_2 and b_{TA} . Although positive and negative A_2 result in quite different w_g predictions, all other parameters held fixed, the combination $b_{TA} \sim 0$, $A_2 \sim -3.5$ and $b_{TA} \sim -1$, $A_2 \sim 2.5$ both produce theory curves that fit the $z = 1$ data adequately on scales $r_p > 6$ (see Fig. B2). The theory predictions differ somewhat on smaller scales, suggesting that pushing below our fiducial scale cut could potentially help to break this degeneracy. This distorts the 1D point representation in Fig. 9; since what is shown in this figure is a single Gaussian approximation, the presence of bimodality both shifts the posterior mean (the points) and broadens the 1σ standard deviation significantly, relative to either of the peaks.

5.3 Tensions and model comparison

Beyond simple posterior constraints, one can also gauge the ability of the data to support the extended modelling in a quantitative way. A number of goodness-of-fit metrics exist in the literature, and we consider a subset of those here. Since ILLUSTRIS-I is relatively unconstraining, and is known to have flaws (in the sense that it overpredicts the strength of baryonic feedback, which is known to interact with IAs; Soussana et al. 2020), we compare the results using ILLUSTRISTNG and MASSIVEBLACK-II only. The simplest metric is the raw shift in χ^2 when switching between models (see for example, Krause et al. 2016); in the ILLUSTRISTNG case, that is $\Delta\chi^2 = -0.33$, marginally favouring the extended model, with similar values obtained at higher redshift. A somewhat stronger preference is seen in MASSIVEBLACK-II, which gives $\Delta\chi^2 = -12.92$. One slightly more sophisticated indicator of model fit is the Bayesian Information Criterion (BIC; Arevalo, Cid & Moya 2017), which effectively balances reducing the theory-data residuals against the extra complexity of the model. For ILLUSTRISTNG, $\Delta BIC = 4.6$, which translates into a 'positive' preference for NLA. That is, by this indicator, the data do not warrant the additional parameters. In contrast, the MASSIVEBLACK-II data, which we recall showed a preference for non-zero A_2 , gives $\Delta BIC = -3.8$, this time in favour of the TATT model. Considering finally the Bayes factor (Marshall, Rajguru & Slosar 2006), we see a similar picture: $B = Z_{TATT}/Z_{NLA} = 0.02$ for ILLUSTRISTNG, which indicates that the data favour the simpler model (or rather, the extra TATT parameters do not provide a sufficiently better fit to outweigh the added model complexity). Again, in the case of MASSIVEBLACK-II, the results are slightly clearer, with $B = 5.65$, which (just) falls into the category of 'substantial' evidence on the Jeffreys Scale. In summary, these numerical exercises bear out the qualitative picture we saw earlier; while ILLUSTRISTNG, on the relatively large scales considered, shows no evidence that the NLA model is insufficient, MASSIVEBLACK-II does show hints.

A different, but related, question one could ask is: given our results, and assuming a particular underlying model, to what extent can we say that there is disagreement between the simulations? Do the hints at non-zero TATT parameters in MASSIVEBLACK-II point to systematic tension between the underlying physical alignment models, or are they in fact consistent with realizations of the same model? We reiterate here that the samples are weighted, such that differences in the underlying halo mass distribution should not be responsible for any differences between the simulations. Again, there are a number of metrics available, suited to different scenarios with different caveats (see Lemos et al. 2020 for discussion), and we will not attempt a comprehensive comparison. For our purposes, we adopt a slightly different form of the Bayes ratio (see equation V.3, Dark Energy Survey Collaboration 2018),

$$R = \frac{p(\mathbf{D}_{TNG}, \mathbf{D}_{MBII} | M_{IA})}{p(\mathbf{D}_{TNG} | M_{IA}) p(\mathbf{D}_{MBII} | M_{IA})}. \quad (30)$$

The numerator here is the Bayesian evidence obtained from jointly analysing the two-point data from the two simulations. The lower terms are those from the separate analyses of ILLUSTRISTNG and MASSIVEBLACK-II in isolation. Note that in the joint analysis, we assume the two data sets are independent, with no cross-covariance. In the TATT case, we find $R < 0.1$, which constitutes strong evidence for tension on the Jeffreys Scale. Again, this is implied by the differences we saw in the marginalized credibility contours, but it is interesting that it is borne out by the numerical metric.

6 EXTENSIONS BEYOND THE FIDUCIAL TWO-POINT ANALYSIS

In this section, we discuss a series of modifications to our baseline analysis, with the aim of exploring the basic results above in more depth. This includes a series of analyses with less stringent cuts, probing scales down to $1 h^{-1}$ Mpc. We also examine the dependence of the signal on various physical properties, including colour, type (central or satellite) and luminosity.

6.1 Exploring smaller physical scales

As we have seen in Section 5, our fits to the large scale ILLUSTRISTNG correlation functions are consistent with the NLA scenario (i.e. pure tidal alignment). While there is a detectable IA signal, the parameters controlling deviations from NLA are consistent with zero. At least in principle, however, there exists a regime where the higher order corrections are significant (and thus necessary to model the data adequately), but one halo contributions are still subdominant (see Blazek et al. 2019’s fig. 1). It is this that motivates us to extend our fits below the fiducial cut off at $6 h^{-1}$ Mpc.

The fiducial cut follows Joachimi et al. (2011) and, as discussed there, is conservative by design, intended to be well clear of the scales on which non-linear bias enters the data. The precise scales on which the linear approximation breaks down is, however, somewhat dependent on the galaxy selection, as well as the statistical precision of the measurement. One benefit of using simulated data, however, is that we have access to the dark matter field directly; it is, then, possible to check where exactly non-linear galaxy bias begins to manifest in our particular measurements. A longer discussion can be found in Appendix D, but in brief we estimate the effective scale-dependent bias as a function of r_p as the ratio $b_g = (w_{gg}/w_{\delta\delta})^{1/2}$. Based on this exercise, within ILLUSTRISTNG’s statistical uncertainties, we see that the linear bias assumption holds well down to $\sim 1 h^{-1}$ Mpc. Motivated by this finding, we repeat our fiducial analysis, sequentially relaxing the lower scale cut down to $r_p > 1 h^{-1}$ Mpc. The results can be found in Fig. 10 (see also Table 4).

As we can see in Fig. 10, all the way down to $1 h^{-1}$ Mpc, the higher order TATT model parameters favoured by the ILLUSTRISTNG data are consistent with zero. This includes the density weighting term b_{TA} , as well as the quadratic amplitude A_2 . Although there appears to be information on the smaller scales, evidenced by the reduction in the size of the posteriors and the slight change in the degeneracy direction, there is no clear sign of deviations from NLA. The added constraining power is particularly clear in the case of the A_2 amplitude, although we also see a modest tightening of the uncertainties on A_1 and b_{TA} about their central values. This translates into best-fitting theory data vectors that differ only negligibly compared with the error bars on the scales included in the fits. It seems reasonable to draw from this that although we have physical reason to think that the additional TATT contributions exist in the Universe, they are small enough on the scales we use to be undetectable, given the statistical precision of ILLUSTRISTNG. The higher order terms scale rapidly with r_p , and so it is quite possible that they dominate in a similar regime to non-linear galaxy bias. This is also consistent with the conclusions one might draw from naively looking at the data vectors in Fig. 5; the purple points are reasonably fitted by the purple lines (the best fitting NLA model), even down to scales $\sim 1 h^{-1}$ Mpc. This is true of both w_{g+} and w_{++} and, while deviations do exist, they are at slightly smaller scales ($< \sim 1 h^{-1}$ Mpc). Fitting IAs on even smaller scales, where non-linear bias becomes non-negligible, is possible, given that perturbation theory predicts

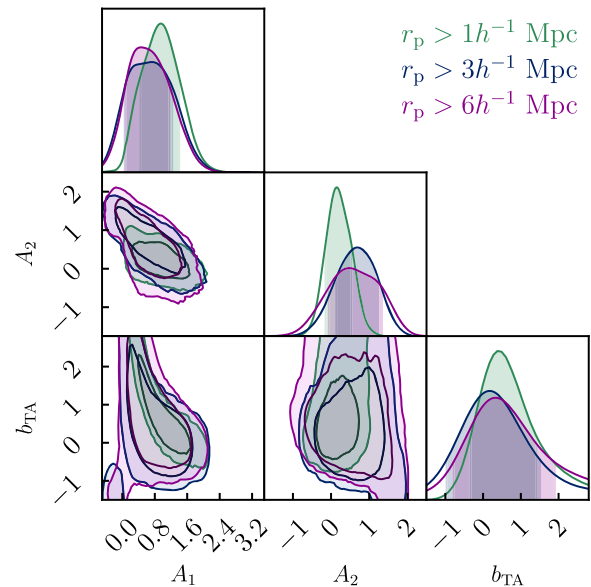


Figure 10. TATT parameter constraints from our $z = 0$ ILLUSTRISTNG sample with a selection of lower scale cuts (as labelled). The three analyses favour approximately the same A_1 , with slightly varying precision. Even in the case of the least stringent cuts, the results are consistent with $A_2 = 0$.

higher order bias contributions in much the same way as the higher order IA contributions in TATT. It is, however, complicated by the presence of non-linear bias – non-linear IA cross terms, which we cannot safely assume are negligible. Although we do not attempt such an analysis here, implementing a consistent perturbative model, including the cross terms, is the focus of ongoing work. Although for simplicity, we focus on the lowest redshift, inspection of Fig. 5 suggests potentially larger deviations at higher redshifts (where the TATT model predictions based on fits to $> 6 h^{-1}$ Mpc appear to extrapolate less well down to smaller scales).

We perform a similar exercise with MASSIVEBLACK-II, fitting the $z = 0$ correlation functions down to $1 h^{-1}$ Mpc. Again, the constraints tighten significantly; now, however, the contours shift in the negative A_2 , positive A_1 direction ($A_1 = 5.1 \pm 0.6$, $A_2 = -3.9 \pm 0.3$, $b_{TA} = -0.1 \pm 0.1$). This shift appears to be driven by a feature in the two-point measurements, which can be seen in the middle left-hand panel of Fig. 5, peaking at $\sim 1 h^{-1}$ Mpc. The physical explanation for this is not immediately clear, but it appears to persist at some level across the redshift range.

6.2 Dependence on galaxy properties

In this section, we impose a series of catalogue level splits, with the aim of understanding how our results depend on galaxy properties. For two main reasons, we only consider the fiducial ILLUSTRISTNG catalogues in this section. First, the larger volume allows some leeway, such that sub-divisions can be made without degrading the constraining power beyond the point of usefulness. Secondly, and more importantly, only in ILLUSTRISTNG do we have sufficiently realistic galaxy photometry (see Section 2.4.3). Although some of the properties considered here are correlated, we seek to disentangle the impact of each in so far as we can. For each of the cases discussed below, the new data vectors are recomputed using the same pipeline as before. For each subsample, we also repeat the iterative covariance matrix calculation discussed in Section 4.2 with the appropriate galaxy densities and ellipticity dispersions.

Table 4. Quality metrics for TATT model fits to ILLUSTRISTNG at $z = 0$. The second column, labelled N_{pts} indicates the total number of points included in the joint fit to w_{gg} , w_{g+} and w_{++} , after scale cuts.

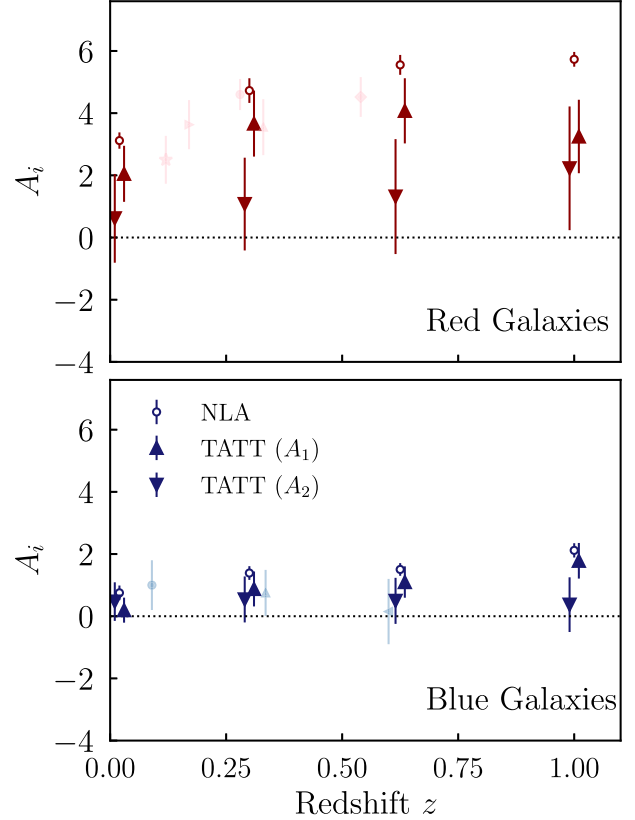
Cut/ h^{-1} Mpc	Model	N_{pts}	A_1	A_2	b_{TA}	b_g
$r_p > 6$	NLA	14	1.51 ± 0.18	0.0	0.0	1.08 ± 0.07
$r_p > 6$	TATT	14	0.68 ± 0.50	0.60 ± 0.62	0.73 ± 1.75	1.05 ± 0.08
$r_p > 3$	TATT	17	0.70 ± 0.53	0.65 ± 0.50	0.44 ± 2.03	1.05 ± 0.05
$r_p > 2$	TATT	20	0.61 ± 0.51	0.69 ± 0.49	0.46 ± 2.32	1.05 ± 0.05
$r_p > 1$	TATT	23	0.95 ± 0.44	0.20 ± 0.32	1.03 ± 1.41	0.91 ± 0.02

6.2.1 Galaxy colour

The first split we examine is in colour–magnitude space. The ability to perform a colour cut, and retain a significant number of red and blue objects, is a marked difference between this work and previous direct IA measurements on real data, which have focused on bright red samples at low redshift. We recompute the correlation functions and covariance matrices for the red and blue subsamples described in Section 2.4.3. As in all of our large scale fits, the full unsplit catalogue is used for the density part of the correlations. This gives us an analogous two new data vectors, $\mathbf{D}^{\text{red}} = (\mathbf{w}_{++}^{RR}, \mathbf{w}_{g+}^R, \mathbf{w}_{gg})$ and $\mathbf{D}^{\text{blue}} = (\mathbf{w}_{++}^{BB}, \mathbf{w}_{g+}^B, \mathbf{w}_{gg})$, with the superscripts R and B denoting the red and the blue samples. Note that the density tracer sample is not split, and so w_{gg} here is the same in the two data vectors (and the same as that analysed in Section 5). We fit both IA models using each data vector, with the results shown in Fig. 11. For the sake of clarity and to aid comparison, rather show the full parameter contours, we have condensed the IA amplitude parameters into 1D posterior means and 68 per cent error bounds. While this is useful for illustrative purposes, it can be reductive in cases where the posterior is non-Gaussian, as we will discuss below.

As before, the single NLA amplitude approximately agrees with the A_1 amplitude from the TATT fits in almost all cases; the exception to this is the high z red sample, which favours a combination with non-zero TT contribution and a correspondingly lower TA amplitude, although the significance of $A_2 \neq 0$ is still only $\sim 1 - 2\sigma$. Although the details of the redshift distribution and the sample selection make direct comparison non-trivial, it is interesting to note that this disagrees mildly with the findings of Samuroff et al. (2019), which are based on fits to real cosmological lensing measurements from DES Y1, where positive values of A_2 in a red source sample were disfavoured at the level of $\sim 2\sigma$ (see their fig. 16). We also plot a number of previous direct IA measurements in Fig. 11 (the pastel coloured points in both panels), from BOSS LOWZ (Singh et al. 2015), KiDS, GAMA, and SDSS (Johnston et al. 2019), MegaZ (Joachimi et al. 2011), and WiggleZ (Mandelbaum et al. 2011). Although red galaxy measurements are more numerous, there are a handful of comparable studies on blue galaxies. As one can see from Fig. 11, our fits on ILLUSTRISTNG are largely consistent with the measurements on data. The only slight deviation from this is WiggleZ, which is lower than our results at equivalent redshift (albeit only by $\sim 2\sigma$). It is, however, worth bearing in mind that WiggleZ is atypical in terms of sample, comprising a bright starburst population, rather than a simple colour-selected blue sample.

A notable, and perhaps worrying, feature of Fig. 11 is the relatively strong IA signature in blue galaxies. The amplitude of w_{g+}^B , while significantly lower w_{g+}^R , is persistently non-zero at $z > 0.5$. To aid in understanding this observation, we repeat the two-point measurements and NLA fits on the upper redshift snapshot, with an additional mass cut, considering only galaxies in the lower 25 per cent, $M_* < 2.1 \times 10^{10} h^{-1} M_\odot$ (mean stellar mass $M_* = 1.3 \times 10^{10} h^{-1} M_\odot$). Even here, we see non-zero alignments at several

**Figure 11.** Best-fitting IA model parameters as a function of redshift for our colour split ILLUSTRISTNG samples. Note that two TATT model amplitudes are fit simultaneously for each of the two samples. The pink points in the upper panel are measurements of the NLA model amplitude in red galaxies from the literature. Specifically, we show SDSS main sample ($z = 0.12$; Johnston et al. 2019), BOSS LOWZ ($z = 0.28$; Singh et al. 2015), GAMA red sample ($z = 0.17$ and $z = 0.33$; Johnston et al. 2019) and MegaZ ($z = 0.54$; Joachimi et al. 2011). Similarly, the light blue points in the lower panel represent published blue galaxy constraints: SDSS main sample (at $z = 0.09$; Johnston et al. 2019), GAMA Z2B ($z = 0.34$; Johnston et al. 2019), and WiggleZ ($z = 0.51$; Mandelbaum et al. 2011).

σ , $A_1 = 2.1 \pm 0.6$. Although lower than both the blue and unsplit samples at $z = 1$, it is still a relatively strong signal. Remarkably, we find that the high redshift blue IA feature persists under further mass splitting, down to $M_* < 1.9 \times 10^9 h^{-1} M_\odot$; at this point, there are only ~ 1000 blue galaxies in the shape sample, such that although the measurement is consistent with null, the error bars still encompass significant non-zero values. Although interesting, it is not clear whether this is a function of the relatively stringent convergence cut ($M_* > 1.6 \times 10^9 h^{-1} M_\odot$), and if so how far down in mass the alignment signal continues. It is also not obvious whether this transfers to a significant IA lensing contaminant in a more realistic

setup; implementing a redshift-dependent selection function, typical of real cosmic shear is a topic we will explore in future work.

Given that the role of a galaxy within its halo is a significant factor in determining its alignment, we also repeat the high z blue measurements with an additional satellite/central split. The results here are less ambiguous: the residual blue galaxy signal is generated almost entirely by central galaxies. That is w_{g+} , as measured using satellite galaxy shapes is consistent with zero on scales $r_p > 6 h^{-1}$ Mpc. This result seems to support, at least in our case (which is simplified relative to real data in a number of ways), the findings of Johnston et al. (2019), which suggest colour alone is an imperfect determinant of IA properties. Singh et al. (2015) also note similar, although consider only LRGs (that is, their results were a statement on the relative homogeneity of IAs in red sequence galaxies of given luminosity, rather than on the efficacy of colour based splits). In the absence of blue high z alignment measurements in real data, it is difficult to say whether this is a fault in the simulations, generating an artificially strong IA signal in blue centrals, or a real feature of the Universe.

6.2.2 Stellar mass

Although stellar mass is not, in general, an observable quantity it is an important one; this is true both in that IAs (and other galaxy properties) significantly depend on it, and that it is a proxy for actual observables. Indeed, this link is key to approaches such as Halo Occupation Distribution (HOD) modelling. We compute each galaxy's stellar mass as an unweighted sum over the stellar particles assigned to its subhalo. Unlike with colour and centrals/satellites there is no natural dividing line for this split, and so we choose to divide galaxies into equal number mass bins. For the moment we will consider a simple upper/lower mass division, but will consider a more complex binning in what follows. Again, the split is applied to the shape sample only, leaving the density tracer intact (and so w_{gg} unchanged).

Though the satellite fraction is not systematically changed by the division in any of the snapshots, we do see a shift in the abundance of red galaxies. That is, the red fraction of the high mass sample is boosted relative to the full sample, from ~ 35 per cent to ~ 60 per cent at $z = 0$ and from ~ 12 per cent to 24 per cent at $z = 1$. This qualitative trend, that the red fraction increases with mass, and declines with redshift, is consistent with the patterns seen in real data (see e.g. Prescott et al. 2011). In the case of the NLA constraints we have a relatively simple picture from the mass-split reanalysis; at a given redshift, high mass galaxies are both more biased, and more strongly aligned; the direction of the shift in bias-IA amplitude parameter space when going from the high to low mass is roughly the same, irrespective of redshift. The redshift trend can perhaps be understood as follows: if we are to believe the basic LA model premise, then intrinsic alignments are imprinted at early times, and persist into the low redshift universe. In this picture, at least, if high mass red galaxies at $z = 0$ are strongly aligned then it follows that the high-redshift objects that will become bright red high mass galaxies are also strongly aligned. In other words, as redshift increases, even if the mean subhalo mass declines (which it does, in Table 2) the more massive, redder section of the galaxy population will be strongly aligned relative to the rest. Some evidence for this picture can be seen in Bhowmick et al. (2020).

What does stay fixed, however, is the lower mass threshold we impose on our catalogues. As the whole mass distribution shifts downwards, then, we are preferentially cutting more of the lower

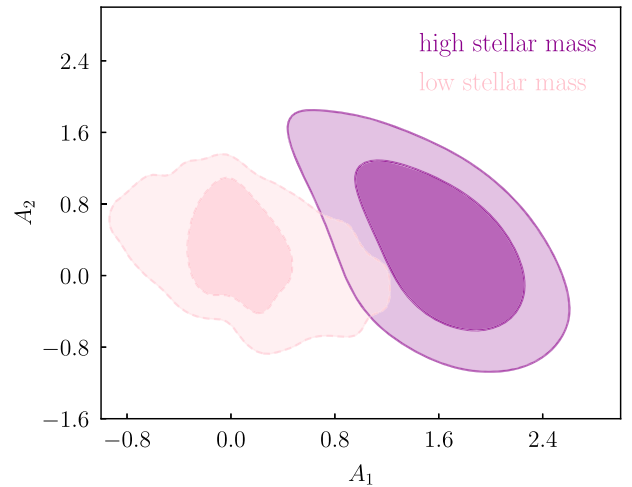


Figure 12. TATT model posterior constraints from ILLUSTRISTNG, under a binary high/low mass split. The sample is divided about the median stellar mass, $M_* = 7.8 \times 10^9 h^{-1} M_\odot$, and the two subsamples are fit independently.

part of the mass distribution, and so discarding a larger fraction of weakly aligned objects. The gradual evolution in IA and bias parameters covers the range from virtually unaligned low mass galaxies at $z = 0$ to $A_1 \sim 3.5$ in the high mass high redshift bin. It is worth bearing in mind that, although physically interesting, this pattern does not trivially carry over into real data, because in such cases other observational effects become relevant. The fact that we typically use flux-limited galaxy samples for lensing measurements, for example, means that the mean stellar mass tends to increase with redshift, not decline as in our case. Fully separating out these effects, of sample composition and evolution of intrinsic alignments, would require a more careful exploration using merger trees of the sort presented by Bhowmick et al. (2020).

We also rerun our TATT analysis on the mass-split data vectors, giving the marginalized parameter constraints shown in Fig. 12. For clarity, we show only $z = 0$ here, but find similar patterns in all snapshots. As before, A_1 gradually increases over the range $z = 0-1$, and the 1σ contour encompasses $A_2 = 0$ in all cases. That said, there is a relatively strong anticorrelation between the two IA amplitudes, such that a range of scenarios with $A_2 > 0$, combined with slightly reduced A_1 , are also equally favoured.

6.2.3 Luminosity

In addition to the binary mass cut above, we also consider directly the luminosity dependence of the measured IA signal, defining four equal-number bins in r -band luminosity. To separate actual luminosity dependence and changes in the red fraction between bins, we impose the red/blue colour split described in Section 6.2.1. The signal-to-noise in the upper red bin is particularly high, which motivates a further equal number subdivision, slightly extending our luminosity coverage. We then have the luminosity bins $L_{r, \text{red}}/L_0 = [(0.030 - 0.351), (0.351 - 0.575), (0.575 - 0.977), (0.977 - 48.091)]$ and $L_{r, \text{blue}}/L_0 = [(0.028 - 0.131), (0.131 - 0.224), (0.224 - 0.424), (0.424 - 16.341)]$, which roughly, but not exactly, correspond to mass bins. In each one we recompute the correlation functions and the covariance matrix, then fit using both IA models. As before we impose the split only on the shape sample, which is correlated with the full density sample. The results, as a function of r - and luminosity, are shown for NLA and TATT respectively in

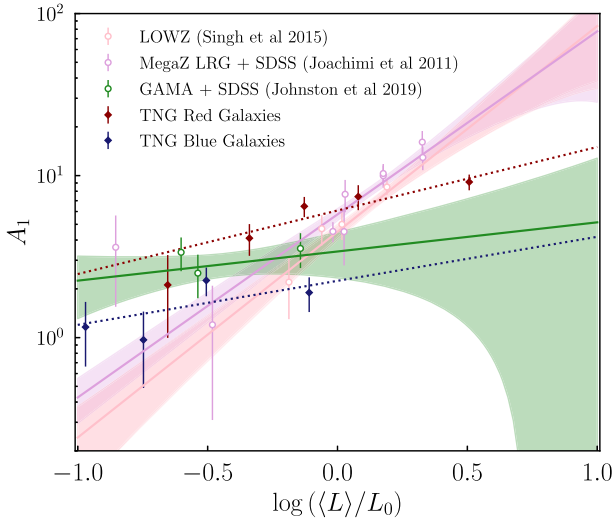


Figure 13. Luminosity dependence of the measured NLA intrinsic alignment amplitude. The red/blue diamonds show the two colour subsamples of ILLUSTRISTNG at $z = 0.3$, and the dotted lines of the same colour show power-law fits to these data. For reference, we also show comparable measurements from MegaZ + SDSS LRG + L4 + L3 (Joachimi et al. 2011), BOSS LOWZ (Singh et al. 2015) and KiDS×GAMA+SDSS (Johnston et al. 2019) in purple, pink, and green. The shaded bands represent their fits and the corresponding uncertainties.

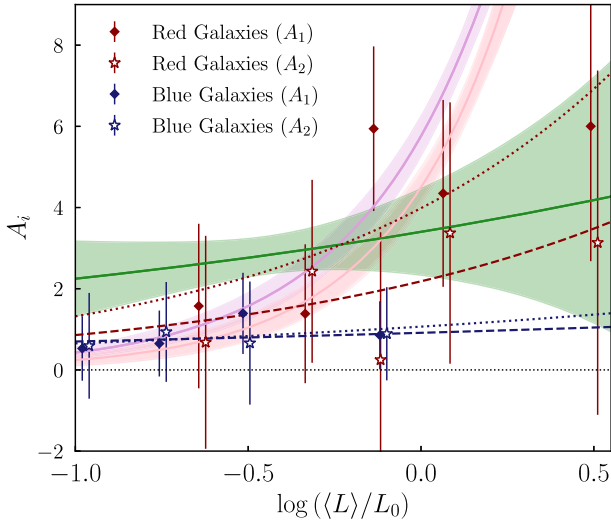


Figure 14. Luminosity dependence of the measured TATT model intrinsic alignment parameters. As in Fig. 13 the shaded bands show fits to MegaZ, LOWZ, and SDSS+GAMA (points now omitted). The diamonds show the tidal alignment amplitude A_1 , while the stars show the tidal torquing contribution A_2 . We show the best-fitting power laws, parametrized $A_i(L) = A_{i,z}(L/L_0)^{\beta_i}$, for A_1 (dotted) and A_2 (dashed). The numerical values of the power-law slopes are quoted in Section 6.2.3.

Figs 13 and 14. For the purposes of comparison with the literature, we consider the second snapshot, $z = 0.3$ only here. This choice does not significantly change the conclusions of this section.

Fig. 13 shows the NLA case, with open points indicating previous constraints on the IA luminosity relation using data (Joachimi et al. 2011; Johnston et al. 2019; see also Fortuna et al. 2021’s fig. 5 and Singh et al. 2015’s fig. 10). All of these represent direct IA measurements at low redshift, using relatively bright red samples.

The MegaZ + SDSS LRG + L4 + L3 fit (Joachimi et al. 2011) in particular has been widely used in the literature to extrapolate the luminosity dependence to fainter samples (see e.g. Krause et al. 2016). In addition, we show our new results from ILLUSTRISTNG, both red and blue samples. The blue simulated subsample lies towards the fainter end of this plot, going fainter than any of the data measurements. The red, on the other hand, covers a wider luminosity range, spanning both KiDS×GAMA+SDSS (Johnston et al. 2019) and the MegaZ + SDSS LRG + L4 + L3 fit (Joachimi et al. 2011) points. It is worth remarking here that unlike Fig. 11, the points each represent a different set of galaxies. Whereas there the different snapshots strongly overlap, and so are subject to highly correlated errors, the noise realizations should now be independent, making fitting a trend relatively simple. The luminosity dependence is parametrized as

$$A_1(L, z) = A_z \left(\frac{L}{L_0} \right)^{\beta_1}, \quad (31)$$

where L_0 is a pivot luminosity, corresponding to an absolute magnitude $M_r = -22$. The amplitude A_z and power-law index β_1 are left as free parameters in our fits. Doing a simple least-squares fit to equation (31), we obtain $\beta_{1,\text{red}} = 0.38 \pm 0.08$, $A_{z,\text{red}} = 6.3 \pm 0.5$ for the ILLUSTRISTNG red sample. Notably, this is somewhat shallower than both MegaZ + SDSS LRG + L4 + L3 ($\beta_{1,\text{MegaZ}} = 1.13^{+0.25}_{-0.27}$, shaded purple in Fig. 13) and LOWZ ($\beta_{1,\text{LOWZ}} = 1.27 \pm 0.27$; pink shaded); it is slightly steeper than, but consistent to $\sim 1\sigma$ with, the KiDS×GAMA+SDSS red sample ($\beta_{1,\text{GAMA}} = 0.18^{+0.20}_{-0.22}$, green shaded). For the most part, this fits with the broken power-law picture painted by the existing data sets (i.e. a relatively steep slope at high L , turning into a much flatter function below $L/L_0 \sim 0.8$). Our uppermost L bin, however, indicates something slightly different; the measured IA amplitude in this bin is both relatively well constrained, and below the extrapolated MegaZ power-law prediction by several σ . Taken together with the third point $A_1(L/L_0 = 0.74) = 6.6 \pm 0.9$, which is slightly above the data, this seems to hint at some level of disagreement between simulations and data.

One caveat here is that the x -axis positions are point estimates from luminosity distributions, which have finite width. In the case of the blue sample, the distributions are relatively compact and Gaussian; in the case of the higher L red sample luminosity bin, this is not the case, and the $p(L)$ distribution is broad, with a trailing upper tail, reaching a maximum luminosity of $\log(L/L_0) = 1.6$. Using the modal luminosity as our point estimate, the rightmost point in Fig. 13 shifts slightly to the left, thereby reducing the apparent tension with the earlier results.

Another complicating factor here is the evolving satellite fraction in both our, and the published, samples. In our case, the ILLUSTRISTNG red sample satellite fraction changes significantly from $f_s = 0.7$ in the lowest luminosity bin ($\log(L)/L_0 = -0.53$) to $f_s = 0.18$ in the brightest bin ($\log(L)/L_0 = 0.5$). In contrast, the satellite fraction of the ILLUSTRISTNG blue sample is quite stable at $f_s \sim 0.3$ across the L range. We explore the impact of this directly by repeating the measurements using red centrals only; as one might expect from the numbers above, the amplitude in the lower bins shifts upwards slightly (to $A_1 = 3.9 \pm 1.0$), to a value which is consistent with the central only GAMA measurement (the yellow point in Fortuna et al. 2021’s fig. 14). The upper luminosity bins are almost completely unchanged (since the IA signal in those bins is heavily dominated by centrals anyway). That is, the central/satellite trend does not seem to be sufficient on its own to explain the discrepancy between our results on ILLUSTRISTNG and the steeper slope seen in LOWZ and MegaZ.

The trend in blue galaxies, $\beta_{1,\text{blue}} = 0.24 \pm 0.21$, $A_{z,\text{blue}} = 2.5 \pm 0.7$, is also interesting, particularly given the lack of existing blue sample measurements. Our results are consistent with no luminosity evolution in blue galaxies, at least at the faint end of the luminosity function. More troubling, perhaps, is the persistence of a relatively strong blue IA signal. This is in accordance with our previous findings, but it is particularly striking here that even in the faintest blue galaxies at $z = 0.3$, there is a non-negligible IA signal, $A_1 = 1\text{--}2$. Although the topic clearly warrants caution, and further investigation in real data, if it bears out this could have significant consequences for future cosmic shear analyses.

An important point to bear in mind here is the choice of pivot luminosity. By convention, and to facilitate comparison with previous results on real data, we choose a pivot L_0 corresponding to an r -band absolute magnitude of $M_r = -22$ (see e.g. Joachimi et al. 2011; Singh et al. 2015; Johnston et al. 2019). This is appropriate for those studies, and for our red galaxy sample, in that L_0 is more or less in the centre of the luminosity range. In the case of the blue ILLUSTRISTNG sample, however, the bulk of the sample is below $L/L_0 = 1$. Although this is a valid analysis choice, and indeed useful for comparison with the literature, it does mean that A_z and β are likely non-trivially correlated in this case.

Finally, we repeat this exercise using the TATT model, again at $z = 0.3$, with the results shown in Fig. 14. Again, we show the best fits to MegaZ, LOWZ, and KiDS×GAMA+SDSS, but now for clarity we omit the corresponding data points. Although we fit power-law slopes as before, the constraints are degraded relative to the NLA case. Fitting to the red sample, we find $\beta_{1,\text{red}} = 0.31 \pm 0.31$, $\beta_{2,\text{red}} = 0.36 \pm 0.35$. As before, the red galaxy TA amplitude A_1 increases from faint to bright galaxies ($\beta_{1,\text{red}} > 0$), as one would naively expect. In all but the brightest two bins in the red galaxy sample, the TT amplitude A_2 remains consistent with zero to $\sim 1\sigma$. Although we report weak positive β_2 here, the fits are extremely noisy, such that very different scenarios are allowed within the uncertainties. At the current precision, then, there is little hope of distinguishing between power law and non-power-law forms of luminosity evolution, at least for A_2 . In the blue sample, we find $\beta_{1,\text{blue}} = 0.14 \pm 0.31$, $\beta_{2,\text{blue}} = 0.34 \pm 0.14$, consistent with no coherent variation with L across the range.

6.2.4 Centrals and satellites

As well as luminosity and colour, the IA signal is known to depend on galaxy type, and so we next consider a satellite/central split. We divide the ILLUSTRISTNG shape catalogues into centrals and satellites using the method described in Section 2.4.2, before remeasuring the two-point functions and repeat the covariance matrix calculation. This gives us two new data vectors $\mathbf{D}^c = (\mathbf{w}_{++}^{cc}, \mathbf{w}_{g+}^c, \mathbf{w}_{gg})$, $\mathbf{D}^s = (\mathbf{w}_{++}^{ss}, \mathbf{w}_{g+}^s, \mathbf{w}_{gg})$. In all cases the shape part of the correlation is either central or satellite, and the density part uses the unsplit catalogue. Note that we repeat this exercise with split density samples, and confirm that we return consistent (albeit slightly degraded) IA constraints when fitting on large scales.

The NLA analysis on these new satellite/central disaggregated data vectors are shown in Fig. 15. As expected, the galaxy bias is consistent between the two, and constrained primarily by w_{gg} , which is the same in the two data vectors. By construction the large scale fits to these data are each sensitive to a particular combination of two halo IA power spectra. Specifically \mathbf{D}^c is sensitive to $(P_{\text{GI}}^{2h,s}, P_{\text{II}}^{2h,ss})$, and \mathbf{D}^s probes $(P_{\text{GI}}^{2h,c}, P_{\text{II}}^{2h,cc})$. Again, we assume that on two halo scales, the satellite/central composition of the density tracer is not

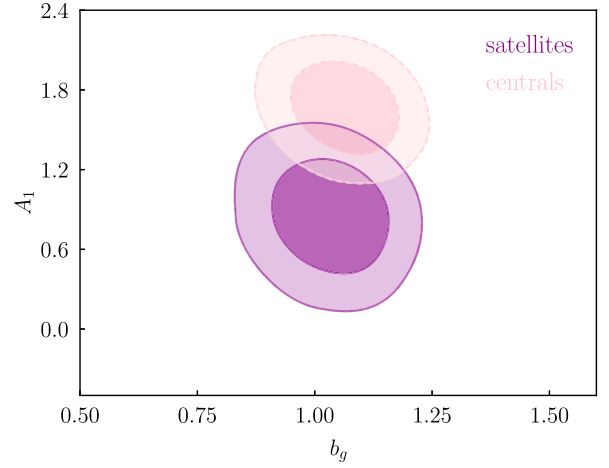


Figure 15. NLA model constraints on our ILLUSTRISTNG sample, split into satellites and centrals. In both cases the split is imposed on the shape sample; the density tracer sample used in w_{g+} and w_{gg} is the full $z = 0$ ILLUSTRISTNG catalogue. As shown, the large intrinsic alignment of satellites is weaker than that of centrals by a factor of ~ 2 .

relevant. Notably, the amplitude of large-scale central alignments in Fig. 15 is stronger than that of satellites by a factor of ~ 2 , at the level of a few σ . The subject of satellite alignments has been discussed quite extensively in the literature, and the overall picture fits with our results here. A number of theoretical studies point to satellite IAs being dominated by tidal torque induced radial alignments within their haloes (Faltenbacher et al. 2007; Knebe et al. 2008; Pereira, Bryan & Gill 2008), which scale rapidly with separation, and tend to wash out on very large scales. There is also now evidence from various observations on both cluster and galaxy scales supporting the same picture (Sifón et al. 2015; Singh et al. 2015; Huang et al. 2018). This, again, is consistent with Johnston et al. (2019) and Fortuna et al. (2021), who suggest satellite shapes are effectively random on sufficiently large scales. Centrals, on the other hand, tend to align with the host halo, and so trace the large scale correlations in the background large-scale structure (Catelan et al. 2001; Kiessling et al. 2015). Although not shown in Fig. 15, it is also worth noting that the central galaxies show a clear monotonic increase in IA amplitude with redshift, a trend which is not replicated in satellite galaxies. There is also some evidence to suggest not only that satellites are distributed in a significantly anisotropic way within their haloes (Zentner et al. 2005; Piscionere et al. 2015; Butsky et al. 2016), but also that this anisotropy can have a non-trivial impact on the IA signal, even on large scales (Huang et al. 2016; Samuroff, Mandelbaum & Di Matteo 2020).

We next repeat our analysis of the 4 split data vectors, but now using the three parameter TATT model instead of NLA. Fig. 16 shows the marginalized parameter constraints at four redshifts. As in the simpler fits above, the central IA signal is stronger than that in satellite galaxies by a factor of a few, although the constraints are degraded to the extent that it is difficult to draw meaningful predictions from this. Again, there is no clear evidence of non-zero favoured values of either the tidal torquing amplitude A_2 , or the density term $A_\delta = b_{\text{TA}}A_1$, in either the satellite or the central population.

We also present the correlation functions of centrals and satellites in Fig. 17. This is a worthwhile exercise for a variety of reasons, not least that there is information on the small-scale IA signal missed in the large-scale fits. While we cannot, at the present time, fit the IA signal on scales $< \sim 1 h^{-1} \text{ Mpc}$, the qualitative comparison can

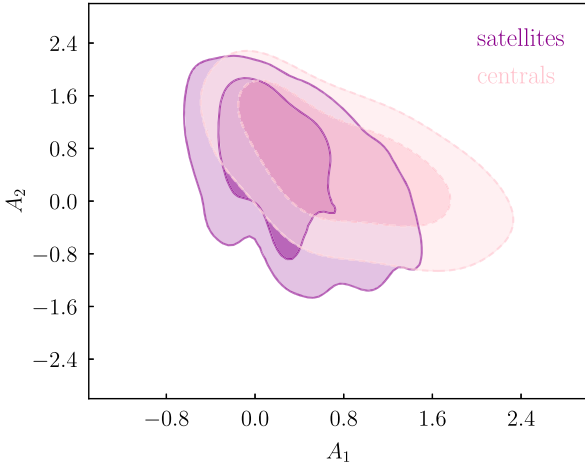


Figure 16. TATT model constraints from ILLUSTRISTNG at $z = 0$, after decomposing the catalogue into central and satellite galaxies. The metric used to define the two classes is outlined in Section 6.2.4. As in Fig. 15, the density sample used here is the full, unsplit catalogue, and the satellite/central split is imposed only on the shape sample.

be instructive. Given that there is some evidence that they behave differently, we consider blue and red satellites/centrals separately here. We also drop the full sample density tracer, and instead use one of the four (red/blue, satellite/central) subsamples. The motivation here is that, while on large scales, the density tracer is effectively just that: a probe of the large-scale matter distribution multiplied by a linear galaxy bias, on scales approaching the one halo regime this no longer holds. Fig. 17 shows these new data vectors. As shown we measure both w_{g+} and w_{++} , and recompute the covariance matrices with the appropriate densities. For reference, the dark red crosses also show the equivalent satellite/central red galaxy w_{g+} correlations from KiDS×GAMA here (c.f. Johnston et al. 2019 fig. 7, red points/band). On large scales at least, our ILLUSTRISTNG red sample is consistent with their measurements. There are a few interesting features here to note, however. First, we see a relatively strong red galaxy 1h contribution on scales $< 1 h^{-1}$ Mpc. Although the general trends match the real data, with ss and to a lesser extent cs exhibiting strong scale-dependent IAs in this regime, the magnitude is somewhat higher in our sample. This is particularly interesting, given that our sample characteristics are similar ($\langle L \rangle / L_0 = 0.91$ and 0.34 for our red and blue samples, respectively, compared with their ~ 0.99 and 0.50). As discussed briefly in Section 6.2.1, we observe a persistent non-zero IA signal in blue galaxies on large scales; here we can see it is dominated by the cc correlation, with a smaller contribution from sc . Also notable is that, in line with some previous direct IA measurements (e.g. Singh et al. 2015), the large-scale satellite correlations do not appear to vanish on large scales. Focusing on the right-hand panels, the purple and pink points are consistently positive and non-zero. While small compared with the red central terms, and consistent with the dark red points from GAMA, there appears to be a detectable signal at the precision allowed by ILLUSTRISTNG.

Now considering the lower panel, we see shape–shape correlations involving satellites do indeed appear to be zero on large scales, irrespective of colour. Indeed, the large scale w_{++} is driven primarily by the cc component, with all other subsets of the data apparently consistent with null signal at $> 6 h^{-1}$ Mpc. As before, we see no significant 1h cc term, down to $\sim 0.1 h^{-1}$ Mpc (a result which should

be true by construction, since each halo contains only one central galaxy).

7 DIRECT IA CONSTRAINTS FROM 3D FIELDS

In earlier sections, we set out an analysis based on measuring and modelling the two-point functions of intrinsic galaxy shapes. This is the most common method for deriving information about intrinsic alignments from data, be it simulated or real (see e.g. Hirata et al. 2007; Chisari et al. 2015; Singh et al. 2015). This section outlines an alternative approach, which exploits the fact that cosmological simulations allow direct access to the underlying matter field. The basic idea is that, with a suitable choice of smoothing scale, one can measure the components of equation (17) directly and perform a linear fit to obtain constraints on the various amplitudes. Bypassing two-point correlations in this way has several advantages, not least that it is potentially less susceptible to noise.

The method for obtaining the tidal tensor and the intrinsic shape field is described in Section 3.3. In brief, the process involves pixelizing the simulation volume at given redshift, and so building smoothed 3D shape and density fields. We can then compute the tidal field by Fourier transforming the density field (see equation 12). One important thing to bear in mind is that we are free to choose the pixel scale, a choice which has some bearing on the physical interpretation of the result.

With these ingredients in hand, we can proceed to fit for the amplitudes in equation (17). By varying C_1 , C_2 , and b_{TA} , we seek to minimize

$$\chi^2(\mathbf{p}_{IA}) = \sum_{i,j,p} \left[\gamma_{ij,p}^I - \gamma_{ij,p}^{I,model}(\mathbf{p}_{IA}) \right] \mathbf{C}_{ij,p}^{-1} \times \left[\gamma_{ij,p}^I - \gamma_{ij,p}^{I,model}(\mathbf{p}_{IA}) \right]. \quad (32)$$

Here the IA model parameters are $\mathbf{p}_{IA} = (C_1, C_2, b_{TA})$. The indices i, j indicate an element of the 3×3 shape tensor, and p identifies a pixel, within which galaxy shapes are averaged. The theory prediction $\gamma_{ij,p}^{I,model}$ is obtained by evaluating equation (17). We will refer to this technique for constraining IA parameters, in contrast to the earlier two-point methodology, as Direct Alignment Field Fitting (DAFF). As in the two-point analysis, the likelihood sampling is performed using COSMOSIS, using MULTINEST; one can find the modules for this here: https://github.com/ssamuroff/direct-ia-theory/tree/master/likelihood/field_fit.

Our DAFF TATT constraints, using a range of pixel scales, are presented in Fig. 18. For reference, the light purple contours also show the equivalent $z = 0$ TATT model posterior from the ILLUSTRISTNG two-point analysis.

Notably, on smaller smoothing scales particularly, there is a significant gain in signal-to-noise. Although, perhaps unsurprisingly, $L > 12 h^{-1}$ Mpc offers little information on the higher order IA contributions, with a suitable choice of scale, the TATT posterior volume is reduced quite considerably. Although this is highly promising in terms of the DAFF method’s future utility, we should point out a few caveats in the comparison.

It is perhaps worth remembering here that these are not independent data sets. The underlying galaxy field, and the shape noise are the same in each, albeit smoothed on different scales. This is also true of the matter tidal field. For such comparisons, it is difficult to gauge the significance of parameter shifts, given that the confidence contours do not account for these correlations.

The second consideration, which muddies the comparison, is that of the scales probed. The earlier analyses of w_{g+} and w_{++}

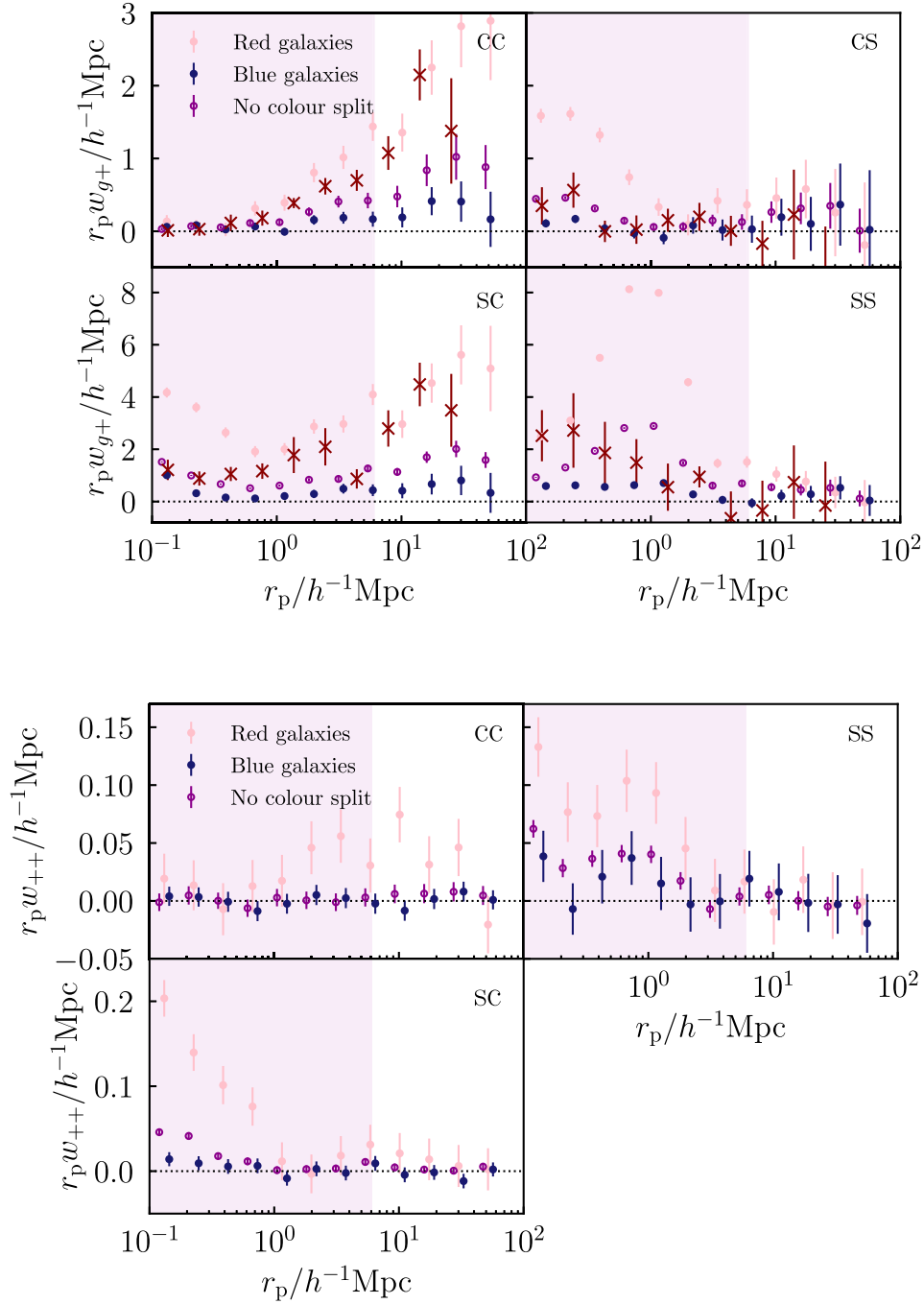


Figure 17. Upper: Projected galaxy-shape correlation functions w_{g+} , for various subsamples of the $z = 0$ ILLUSTRISTNG sample. Shown clockwise from the top right are the central autocorrelation; the satellite shape – central position correlation; the satellite autocorrelation; and the central shape – satellite position cross correlation. In addition to the satellite/central split, galaxies are split by colour (shown by the different colour points). For reference, the shaded region shows scales excluded in the fiducial analysis. The red crosses are analogous split measurements on real KiDS+GAMA data (Johnston et al. 2019; see their fig. 7). Lower: The same, but for shape-shape correlations w_{++} .

have an explicit window of sensitivity determined by our choice of scale cuts, $6 < r_p < 68 h^{-1} \text{Mpc}$. Within that window, however, all scales are fit simultaneously (albeit with unequal weight). The TATT implementation used in the two-point fits does not set an explicit smoothing scale, which some previous incarnations of NLA have, to suppress galaxy-scale fluctuations; rather the filter is included as

an implicit element of the model. The various IA amplitudes are effectively renormalized to account for the impact of small scale processes on mid-to-large scale modes (see Blazek et al. 2019, Section F for discussion). The DAFF approach, in contrast, does include a smoothing scale, in a way that is inherent and unavoidable. The various fields are explicitly pixelized, and averaged on a fixed

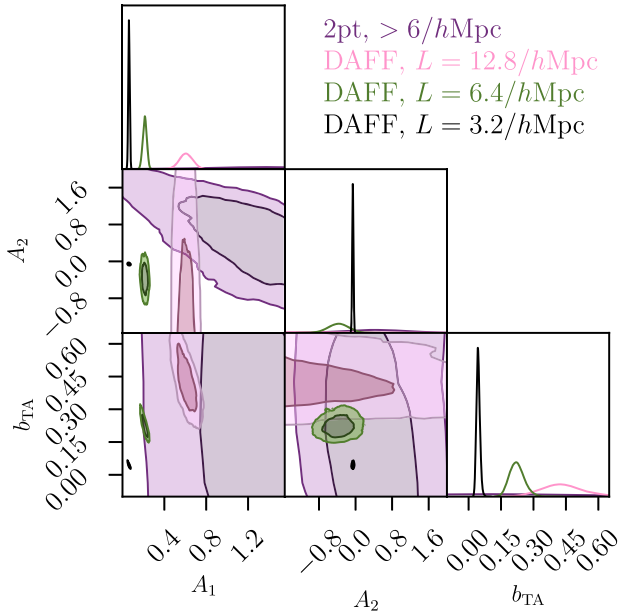


Figure 18. Posterior IA parameter constraints obtained via the DAFF fits described in Section 7. Shown are results using three different pixel scales (indicated in the legend); for reference, we also include the equivalent 1D constraints from $w_{gg} + w_{g+} + w_{++}$, with a scale cut-off at $r_p > 6h^{-1}$ Mpc.

scale with cubic pixels.¹⁰ While this allows some level of control over the physical scales probed, it complicates direct comparison with the two-point results.

One can, and people historically have, adopt a method similar to DAFF in the case of galaxy bias, as discussed in some detail by Desjacques, Jeong & Schmidt (2018) (see Section 4.2, pages 85–93). In order to correctly interpret the results there are corrections of the order of σ_L^2 , (i.e. the variance of linear density field on scale L), which convert between an N -point bias, and that of the moments/scatter. These corrections are complicated to compute, and are the focus of ongoing work. While it is necessary to have a robust estimate of these terms in order to use the DAFF method to make constraints to a precision of better than a factor of a few, we set out here only to present a proof of concept, and so defer calculation of these (order of unity on scales $\sim 6h^{-1}$ Mpc) additional terms to a future work. That said, these corrections should alter both the centring of the IA posteriors and the width by roughly the same factor, such that the signal to noise is approximately conserved. Based on this reasoning, we expect an improvement in the signal-to-noise (posterior mean divided by the 1σ marginalized error) on A_1 of a factor of ~ 4 relative to the two-point constraints.

With these caveats firmly in mind, it is apparent that when one goes to smaller smoothing scales, the favoured A_2 starts to deviate from zero. While the level of significance is still only $\sim 2 - 3\sigma$ at most, this pattern makes sense.

8 DISCUSSION AND CONCLUSIONS

We have presented a detailed study of galaxy intrinsic alignments using two-point measurements from three of the most recent public

¹⁰The use of cubic pixels is an explicit modelling choice in our DAFF pipeline. One could conceivably apply e.g. Gaussian smoothing on top of the pixelization.

hydrodynamic simulations. After halo mass reweighting of the samples to simplify the comparison, we find mild differences in the alignment signal in the three simulations, with MASSIVEBLACK-II favouring a somewhat stronger IA signal. Our key results are summarized below.

(i) We analysed each sample using the NLA model, which assumes linearity in the tidal field. All three of the simulations considered show strong evidence for a non-zero NLA amplitude. The results from the three are consistent within $\sim 2\sigma$, although MASSIVEBLACK-II consistently displays a slightly stronger IA signal relative to the other two across all snapshots.

(ii) For the first time, we fit the more complex TATT model to these various simulated data sets. On scales $r_p > 6h^{-1}$ Mpc we find no clear indication of A_2 , $b_{TA} \neq 0$ in ILLUSTRISTNG ($A_2 = 0.4 \pm 0.6$), and so no strong evidence for deviation from the NLA scenario, at least within the (relatively large) statistical uncertainty of the measurement. MASSIVEBLACK-II, on the other hand, shows a mild preference for negative values ($A_2 = -2.3 \pm 1.0$). In all cases, the best fitting A_1 from the TATT fits is consistent with the amplitude from the NLA only fits on the same data, albeit with greater uncertainty. There is also some level of degeneracy between the two TATT amplitudes, such that combinations with non-zero A_2 , combined with a slightly reduced but still non-zero A_2 are also allowed.

(iii) We discussed a series of fits extending to smaller scales. We justified this by comparing galaxy–galaxy and matter–matter correlations, finding the linear bias assumption to hold in our case down to $\sim 1h^{-1}$ Mpc. Even with these relaxed cuts, we do not report a statistically significant detection of A_2 or b_{TA} in ILLUSTRISTNG.

(iv) We presented a colour split IA analysis on ILLUSTRISTNG, an exercise enabled by the relatively realistic bimodal colour distribution it exhibits. As expected, the red sample displays a strong alignment signal across redshifts. Our results on blue galaxies are consistent with observations at low redshifts; at higher redshifts, $z > 0.5$, where direct constraints on real data are lacking, we detect a non-zero IA signal $A_1 \sim 2$. We explored the origins of this blue IA feature, reporting that it persists even in relatively faint blue subsamples (down to $M_* < 1.9 \times 10^9 h^{-1} M_\odot$), and is generated almost entirely by blue centrals.

(v) We examined the luminosity dependence in red galaxies, reporting results consistent with those of Johnston et al. (2019); the red ILLUSTRISTNG sample favours a marginally shallower slope than Joachimi et al. (2011) and Singh et al. (2015), whose fits are dominated by brighter galaxies. We have also reported the constraints on the luminosity dependence of the quadratic tidal torquing amplitude A_2 . In the TATT fits, A_1 and A_2 exhibit consistent luminosity evolution, which is significant at the level of $1 - 2\sigma$.

(vi) We presented a similar exercise for blue galaxies, which extend into the faint regime where there is a relative paucity of constraints from real data. Our fits prefer a power-law index $\beta_1 \sim 0.24$, which is consistent with the equivalent fit to the red sample. In the TATT parameter space the constraining power is degraded, yielding β_1 and β_2 values consistent with no luminosity dependence.

(vii) We also fitted disaggregated central and satellite correlations. On large scales, our fits favour a weak but non-vanishing, satellite shape alignment signal. Centrals show a stronger signal, and a similar trend with redshift to the mixed sample. The TT and density weighting components of the TATT model do not differ systematically between satellites and centrals, with both A_2 and b_{TA} consistent with zero. At the correlation function level, we see the satellite alignment signal is dominated by the red cs correlation; it

persists in the mixed sample, on $r_p > 6 h^{-1}$ Mpc, albeit subdominant by a factor of several to the red central terms.

(viii) We have outlined a new method for recovering intrinsic alignment information from hydrodynamic simulations, which we refer to as DAFF. Although the results depend significantly on the choice of smoothing scale (in part because we have omitted small corrections, the computation of which is left for future work), the approach potentially offers a significant boost in constraining power relative to an equivalent two-point analysis, and greater control over the physical scales probed.

This work is one of a relatively small number that focus on deriving parametric IA constraints from hydrodynamic simulations (Chisari et al. 2015; Codis et al. 2015a; Tenneti et al. 2015a; Hilbert et al. 2017; Bhowmick et al. 2020); it is the first to attempt a comprehensive analysis of directly comparable samples from multiple simulations, including the current state of the art (ILLUSTRISTNG). Unlike most previous studies, we perform a simultaneous analysis, modelling a joint $w_{gg} + w_{g+} + w_{++}$ data vector. Our analysis also includes an analytic covariance matrix, which is both numerically stable and avoids the (potentially limiting) assumptions of internal estimators such as jackknife.

Although our findings are a building block in our understanding of the behaviour of large-scale galaxy alignments, we urge caution in applying our findings directly to cosmological measurements. That is, our samples are comparable with each other, but are not tailored to match the more complex selection redshift-dependent function of a typical lensing shape sample used in cosmic shear and galaxy–galaxy lensing measurements. Another useful exercise would be to use the observed trends with luminosity, colour and galaxy type to extrapolate out a mock IA signal, more representative of the contamination in real lensing data; in turn, this can be used to test our IA models in a cosmological context. This is a relatively straightforward extension of the results presented here, and is the focus of future work.

Although our exploration of the dependence of IAs on galaxy properties is extensive, it is by no means exhaustive. We have not, for example, considered the question of galaxy morphology, and how directly colour maps on to disc/bulge classification, in terms of IA properties.

The novel DAFF method, which does not involve measuring two-point correlations, is to our knowledge the first implementation in the literature. Although an analogous idea exists in the literature for galaxy bias (see Desjacques et al. 2018, Section 4.2 and the references therein), it has never been discussed in the context of IAs before. The analysis on the $z = 0$ ILLUSTRISTNG snapshot should be seen as a proof-of-concept exercise; while promising, there are still gaps in our interpretation (see Section 7), which are the subject of ongoing work, but beyond the scope of this paper.

It is now well established that intrinsic alignments exist in the Universe, and must be accounted for at some level to avoid biasing cosmological analyses based on cosmic shear and galaxy–galaxy lensing. IAs have been included in cosmic shear analyses for as long as shear has been a competitive cosmological probe (Heymans et al. 2013; Dark Energy Survey Collaboration 2016; Jee et al. 2016; Troxel et al. 2018; Chang et al. 2019; Hikage et al. 2019; Hamana et al. 2020; Hildebrandt et al. 2020; Asgari et al. 2021). Only recently, however, have the lensing data been of sufficient volume to potentially incur biases due to model insufficiency (see Krause et al. 2016, and the Stage IV forecasts of Fortuna et al. 2021 and the tests in the context of DES Y3 Secco, Samuroff et al. 2021; see also Joachimi et al. 2020 for an interesting counter discussion). Developing a

fuller understanding of intrinsic alignments, then, will be crucial for, arguably, the current generation of cosmological surveys, and certainly the next. This paper is one small step in this direction, providing the first detailed analysis at the level of model constraints on the best available cosmological hydrodynamic simulations. Our results, of course, come with a number of caveats. Most notably, our selection function is not intended to accurately match current or future lensing surveys. This is in part because recreating the complex redshift-dependent selection function in a real lensing sample, which would typically be based on a number of correlated observables, is a difficult task; it is also, however, a function of our aim in this study. We wish to understand the behaviour of IAs at a physical level, in order to feed into understanding IAs and model building efforts, rather than make a detailed prediction or robustness test for a particular survey. The behaviour of intrinsic alignments on small physical scales is an important topic for future investigation, and one that could conceivably be addressed using hydrodynamic simulations; indeed, due to the larger number of measureable modes, the signal to noise on small scales is relatively high. The TATT approach allows some hope of pushing to smaller scales (though not into the regime of $< \sim 1 h^{-1}$ Mpc, where one would need an explicit model for 1h alignment contributions). Unfortunately, a number of other poorly understood effects enter on small scales, particularly non-linear galaxy bias and baryonic physics. In order to pursue IA constraints on such scales, it is likely that one would need to consider both higher order bias terms and the interplay with the higher order IA terms.

ACKNOWLEDGEMENTS

We are grateful to Duncan Campbell, Maria Cristina Fortuna, François Lanusse, Jolanta Zjupa, Ami Choi, Scott Dodelson, and the DES WLWG for valuable discussions at various points in the writing of this paper. We would also like to thank Yesukhei Jagvaral for his help debugging the measurement pipeline, Sukhdeep Singh for lending his code to help validate our pipeline and Harry Johnston for kindly allowing us access to his KiDS+GAMA measurements. Finally, thanks to our anonymous referee, for a constructive and thoughtful set of comments. This work made use of the Coma cluster, which is hosted by the McWilliams Center, at CMU. We also used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231. SS and RM are supported by the US National Science Foundation (NSF) under Grant No. 1716131.

DATA AVAILABILITY STATEMENT

The simulated particle data used in this paper is available via the ILLUSTRISTNG data release page: <https://www.tng-project.org/data/downloads/TNG300-1>. Derived data products such as shape catalogues, covariances, and two-point files are also available on request (contact the corresponding author). The theory analysis code and measurement pipeline are hosted on public repositories: https://github.com/ssamuroff/direct_ia_theory and https://github.com/ssamuroff/hydro_ias.

REFERENCES

- Alarcon A., Sánchez C., Bernstein G. M., Gaztañaga E., 2020, *MNRAS*, 498, 2614
- Alonso D., Hadzhiyska B., Strauss M. A., 2016, *MNRAS*, 460, 256

- Ammazzalorso S. et al., 2020, *Phys. Rev. Lett.*, 124, 101102
- Amon A. et al., 2021, preprint ([arXiv:2105.13543](https://arxiv.org/abs/2105.13543))
- Anderson T. W., 2003, *An Introduction to Multivariate Statistical Analysis*, 3 edn. Wiley-Interscience, Hoboken, NJ
- Arevalo F., Cid A., Moya J., 2017, *Eur. Phys. J. C*, 77, 565
- Asgari M. et al., 2021, *A&A*, 645, A104
- Bagla J. S., Prasad J., Khandai N., 2009, *MNRAS*, 395, 918
- Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, *ApJ*, 600, 681
- Bate J., Chisari N. E., Codis S., Martin G., Dubois Y., Devriendt J., Pichon C., Slyz A., 2020, *MNRAS*, 491, 4057
- Bhowmick A. K., Chen Y., Tenneti A., Di Matteo T., Mandelbaum R., 2020, *MNRAS*, 491, 4116
- Blake C. et al., 2020, *A&A*, 642, A158
- Blazek J., McQuinn M., Seljak U., 2011, *J. Cosmol. Astropart. Phys.*, 2011, 010
- Blazek J., Mandelbaum R., Seljak U., Nakajima R., 2012, *J. Cosmol. Astropart. Phys.*, 2012, 041
- Blazek J., Vlah Z., Seljak U., 2015, *J. Cosmol. Astropart. Phys.*, 2015, 015
- Blazek J. A., MacCrann N., Troxel M. A., Fang X., 2019, *Phys. Rev. D*, 100, 103506
- Bridle S., King L., 2007, *New J. Phys.*, 9, 444
- Brown M. L., Taylor A. N., Hambly N. C., Dye S., 2002, *MNRAS*, 333, 501
- Butsky I. et al., 2016, *MNRAS*, 462, 663
- Catelan P., Porciani C., 2001, *MNRAS*, 323, 713
- Catelan P., Kamionkowski M., Blandford R. D., 2001, *MNRAS*, 320, L7
- Chang C. et al., 2019, *MNRAS*, 482, 3696
- Chisari N. et al., 2015, *MNRAS*, 454, 2736
- Chisari N. et al., 2016, *MNRAS*, 461, 2702
- Choi A. et al., 2016, *MNRAS*, 463, 3737
- Codis S. et al., 2015a, *MNRAS*, 448, 3391
- Codis S., Pichon C., Pogosyan D., 2015b, *MNRAS*, 452, 3369
- Dark Energy Survey Collaboration, 2016, *Phys. Rev. D*, 94, 022001
- Dark Energy Survey Collaboration, 2018, *Phys. Rev. D*, 98, 043526
- Dark Energy Survey Collaboration, 2020, *Phys. Rev. D*, 102, 023509
- Desjacques V., Jeong D., Schmidt F., 2018, *Phys. Rep.*, 733, 1
- Di Matteo T., Khandai N., DeGraf C., Feng Y., Croft R. A. C., Lopez J., Springel V., 2012, *ApJ*, 745, L29
- Dubois Y. et al., 2014, *MNRAS*, 444, 1453
- Faltenbacher A., Li C., Mao S., van den Bosch F. C., Yang X., Jing Y. P., Pasquali A., Mo H. J., 2007, *ApJ*, 662, L71
- Fang X., Blazek J. A., McEwen J. E., Hirata C. M., 2017, *J. Cosmol. Astropart. Phys.*, 2017, 030
- Feroz F., Hobson M. P., Cameron E., Pettitt A. N., 2019, *Open J. Astrophys.*, 2, 10
- Fortuna M. C., Hoekstra H., Joachimi B., Johnston H., Chisari N. E., Georgiou C., Mahony C., 2021, *MNRAS*, 501, 2983
- Gatti M. et al., 2018, *MNRAS*, 477, 1664
- Gatti M. et al., 2021, *MNRAS*, 504, 4312
- Gruen D., Brimiouille F., 2017, *MNRAS*, 468, 769
- Hamana T. et al., 2020, *PASJ*, 72, 16
- Handley W. J., Hobson M. P., Lasenby A. N., 2015, *MNRAS*, 453, 4384
- Hartlap J., Simon P., Schneider P., 2007, *A&A*, 464, 399
- Hearin A. P. et al., 2017, *AJ*, 154, 190
- Hernquist L., Springel V., 2003, *MNRAS*, 341, 1253
- Heymans C., Heavens A., 2003, *MNRAS*, 339, 711
- Heymans C. et al., 2013, *MNRAS*, 432, 2433
- Hikage C. et al., 2019, *PASJ*, 71, 43
- Hilbert S., Xu D., Schneider P., Springel V., Vogelsberger M., Hernquist L., 2017, *MNRAS*, 468, 790
- Hildebrandt H. et al., 2020, *A&A*, 633, A69
- Hirata C. M., Seljak U. c. v., 2010, *Phys. Rev. D*, 82, 049901
- Hirata C. M., Seljak U., 2004, *Phys. Rev. D*, 70, 063526
- Hirata C. M., Mandelbaum R., Ishak M., Seljak U., Nichol R., Pimbblet K. A., Ross N. P., Wake D., 2007, *MNRAS*, 381, 1197
- Huang H.-J., Mandelbaum R., Freeman P. E., Chen Y.-C., Rozo E., Rykoff E., Baxter E. J., 2016, *MNRAS*, 463, 222
- Huang H.-J., Mandelbaum R., Freeman P. E., Chen Y.-C., Rozo E., Rykoff E., 2018, *MNRAS*, 474, 4772
- Huang H.-J., Eifler T., Mandelbaum R., Dodelson S., 2019, *MNRAS*, 488, 1652
- Hui L., Zhang J., 2002, preprint ([arXiv:astro-ph/0205512](https://arxiv.org/abs/astro-ph/0205512))
- Jee M. J., Tyson J. A., Hilbert S., Schneider M. D., Schmidt S., Wittman D., 2016, *ApJ*, 824, 77
- Joachimi B., Schneider P., 2010, preprint ([arXiv:1009.2024](https://arxiv.org/abs/1009.2024))
- Joachimi B., Mandelbaum R., A*alla F. B., Bridle S. L., 2011, *A&A*, 527, A26
- Joachimi B. et al., 2021, *A&A*, 646, A129
- Johnston H. et al., 2019, *A&A*, 624, A30
- Joudaki S. et al., 2018, *MNRAS*, 474, 4894
- Kannawadi A. et al., 2019, *A&A*, 624, A92
- Khandai N., Di Matteo T., Croft R., Wilkins S., Feng Y., Tucker E., DeGraf C., Liu M.-S., 2015, *MNRAS*, 450, 1349
- Kiessling A. et al., 2015, *Space Sci. Rev.*, 193, 67
- Kirk D., Rassat A., Host O., Bridle S., 2012, *MNRAS*, 424, 1647
- Knebe A., Draganova N., Power C., Yepes G., Hoffman Y., Gottlöber S., Gibson B. K., 2008, *MNRAS*, 386, L52
- Krause E., Eifler T., Blazek J., 2016, *MNRAS*, 456, 207
- Krause E. et al., 2017, preprint ([arXiv:1706.09359](https://arxiv.org/abs/1706.09359))
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Leauthaud A. et al., 2017, *MNRAS*, 467, 3024
- Lemos P. et al., 2021, *MNRAS*, 505, 6179
- Mandelbaum R. et al., 2011, *MNRAS*, 410, 844
- Mandelbaum R., Slosar A., Baldauf T., Seljak U., Hirata C. M., Nakajima R., Reyes R., Smith R. E., 2013, *MNRAS*, 432, 1544
- Mandelbaum R. et al., 2018, *MNRAS*, 481, 3170
- Marshall P., Rajguru N., Slosar A., 2006, *Phys. Rev. D*, 73, 067302
- McEwen J. E., Fang X., Hirata C. M., Blazek J. A., 2016, *J. Cosmol. Astropart. Phys.*, 2016, 015
- Melchior P. et al., 2017, *MNRAS*, 469, 4899
- Melchior P., Viola M., 2012, *MNRAS*, 424, 2757
- Myles J. et al., 2021, *MNRAS*, 505, 4249
- Nelson D. et al., 2015, *Astron. Comput.*, 13, 12
- Nelson D. et al., 2018, *MNRAS*, 475, 624
- Nelson D. et al., 2019, *Comput. Astrophys. Cosmol.*, 6, 2
- Pereira M. J., Bryan G. L., Gill S. P. D., 2008, *ApJ*, 672, 825
- Pillepich A. et al., 2018a, *MNRAS*, 473, 4077
- Pillepich A. et al., 2018b, *MNRAS*, 475, 648
- Piras D., Joachimi B., Schäfer B. M., Bonamigo M., Hilbert S., van Uitert E., 2018, *MNRAS*, 474, 1165
- Piscionere J. A., Berlind A. A., McBride C. K., Scoccimarro R., 2015, *ApJ*, 806, 125
- Power C., Knebe A., 2006, *MNRAS*, 370, 691
- Prat J. et al., 2018, *Phys. Rev. D*, 98, 042005
- Prat J. et al., 2019, *MNRAS*, 487, 1363
- Prescott M. et al., 2011, *MNRAS*, 417, 1374
- Rykoff E. S. et al., 2016, *ApJS*, 224, 1
- Samuroff S. et al., 2019, *MNRAS*, 489, 5453
- Samuroff S., Mandelbaum R., Di Matteo T., 2020, *MNRAS*, 491, 5330
- Sánchez J. et al., 2020, *MNRAS*, 497, 210
- Schaan E., Krause E., Eifler T., Doré O., Miyatake H., Rhodes J., Spergel D. N., 2017, *Phys. Rev. D*, 95, 123512
- Schmitz D. M., Hirata C. M., Blazek J., Krause E., 2018, *J. Cosmol. Astropart. Phys.*, 2018, 030
- Schneider M. D., Bridle S., 2010, *MNRAS*, 402, 2127
- Secco L. F. et al., 2021, preprint ([arXiv:2105.13544](https://arxiv.org/abs/2105.13544))
- Sifón C., Hoekstra H., Cacciato M., Viola M., Köhlinger F., van der Burg R. F. J., Sand D. J., Graham M. L., 2015, *A&A*, 575, A48
- Singh S., Mandelbaum R., 2016, *MNRAS*, 457, 2301
- Singh S., Mandelbaum R., More S., 2015, *MNRAS*, 450, 2195
- Soussana A. et al., 2020, *MNRAS*, 492, 4268
- Springel V., 2005, *MNRAS*, 364, 1105
- Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, *MNRAS*, 328, 726
- Springel V. et al., 2018, *MNRAS*, 475, 676

- Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, *ApJ*, 761, 152
- Tenneti A., Singh S., Mandelbaum R., di Matteo T., Feng Y., Khandai N., 2015a, *MNRAS*, 448, 3522
- Tenneti A., Mandelbaum R., Di Matteo T., Kiessling A., Khandai N., 2015b, *MNRAS*, 453, 469
- Tenneti A., Mandelbaum R., Di Matteo T., 2016, *MNRAS*, 462, 2668
- Troxel M. A. et al., 2018, *Phys. Rev. D*, 98, 043528
- Valentini M., Bressan A., Borgani S., Murante G., Girardi L., Tornatore L., 2018, *MNRAS*, 480, 722
- Velliscig M. et al., 2015, *MNRAS*, 453, 721
- Vielzeuf P. et al., 2021, *MNRAS*, 500, 464
- Vogelsberger M. et al., 2014, *MNRAS*, 444, 1518
- Weinberger R., Springel V., Pakmor R., 2020, *ApJS*, 248, 32
- Zentner A. R., Kravtsov A. V., Gnedin O. Y., Klypin A. A., 2005, *ApJ*, 629, 219
- Zjupa J., Schäfer B. M., Hahn O., 2020, preprint ([arXiv:2010.07951](https://arxiv.org/abs/2010.07951))
- Zuntz J. et al., 2015, *Astron. Comput.*, 12, 45
- Zuntz J. et al., 2018, *MNRAS*, 481, 1149

APPENDIX A: PIPELINE AND COVARIANCE MATRIX VALIDATION

The likelihood pipeline used in this work is built from public code, developed within the COSMOSIS framework. During the process of developing this code base, we implemented a series of validation exercises, intended to ensure our results are both accurate and repeatable.

The first step in this process is a data vector-level comparison between different theory codes. We generate a non-linear matter power spectrum using COSMOSIS, which is then fed into (i) our theory pipeline, which is used for inference in this work, and (ii) an external code developed by an independent group, and used in Singh et al. (2015). The two codes produce projected correlations $w_{gg}(r_p)$, $w_{g+}(r_p)$, and $w_{++}(r_p)$. The r_p sampling is slightly different, and so we interpolate to a comparable set of values. The result is shown in Fig. A1; we can see here that the two agree relatively well. Though the residuals in the two IA correlations are non-zero and roughly scale-independent, the difference is comfortably within ~ 0.5 per cent.

Though it is reassuring that the two codes are consistent with each other at some (relatively sensible) set of input parameter values, this is not in itself a rigorous demonstration that our pipeline is unbiased. Using the Singh et al. (2015) code we then generate a fiducial data vector, w_{gg} , w_{g+} , w_{++} at four redshifts $z = (0.0, 0.30, 0.625, 1.00)$. Using these mock data, we run our inference pipeline with the fiducial (analytic) covariance matrices obtained through the process described in Section 4.2. We report that we can recover the input parameters to comfortably within 0.5σ .

We also compare our analytic covariance matrix with an alternative, obtained by jackknife resampling. The jackknife covariance is generated by dividing the ILLUSTRISTNG box into $4^3 = 64$ subvolumes, and iteratively remeasuring our data vector. Although the comparison is useful as a cross test, it is worth bearing in mind that the jackknife estimator relies on a number of assumptions that do not strictly hold in our case (Hartlap et al. 2007). That is, although order-of-magnitude differences are not expected, we have first principles reasons to trust our fiducial covariance.

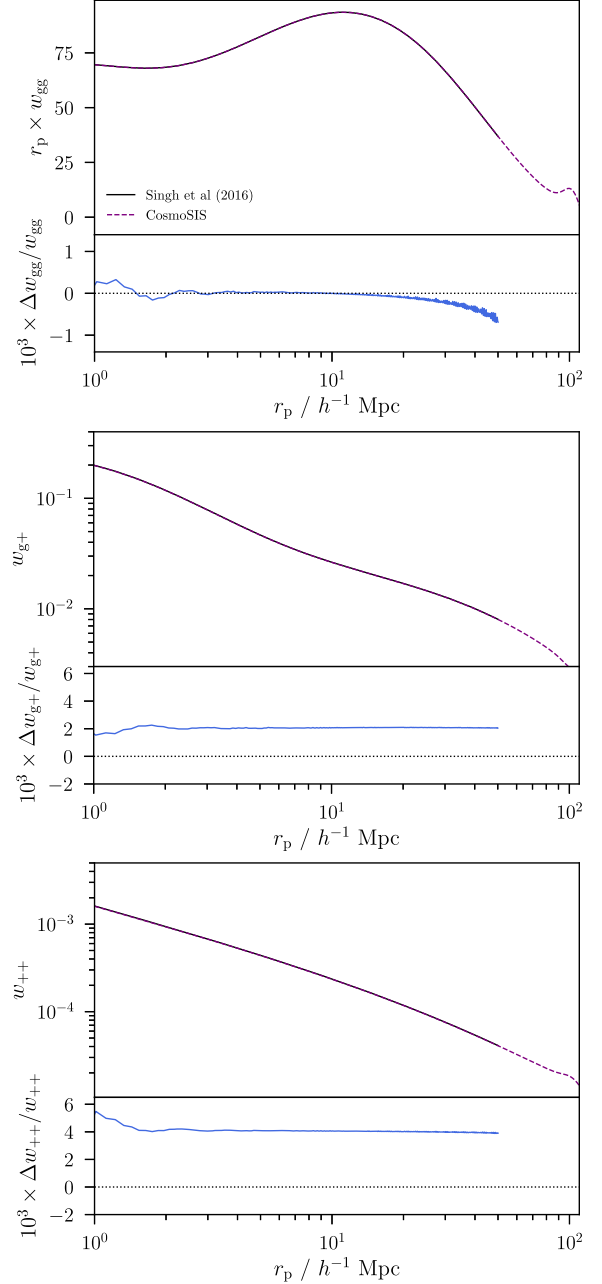


Figure A1. A comparison of theory data vectors produced by two independent codes. The dashed purple and black (solid) lines show the outputs of the COSMOSIS module produced for this work, and an external theory code used in Singh et al. (2015). In the lower panel we show the fractional residual between the two.

The numerical comparison of the diagonals can be seen in Fig. A2. As can be seen, the differences are significant, on all scales considered. In most cases (all but w_{gg} on small scales), jackknife tends to underestimate the uncertainties at the level of 25 per cent or more.

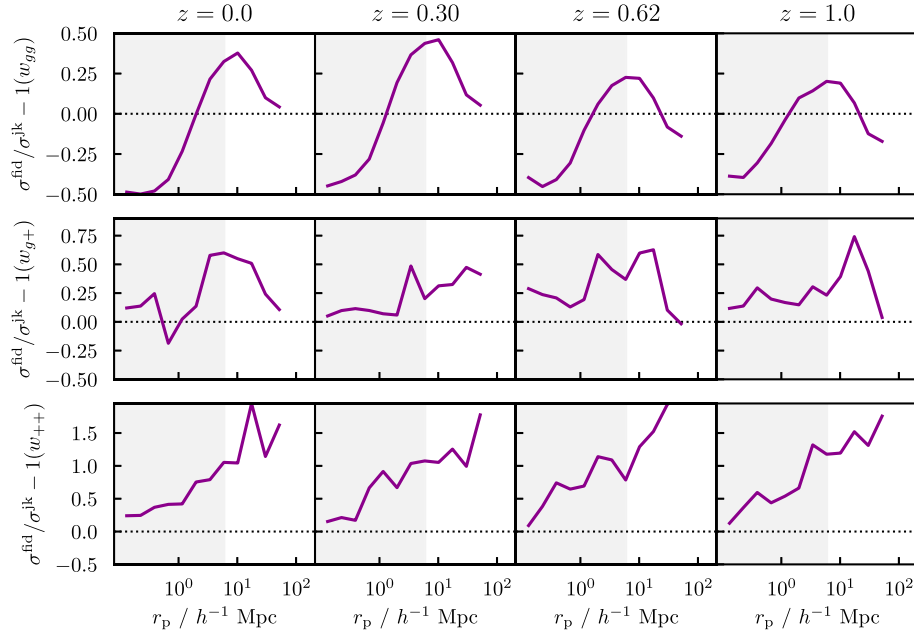


Figure A2. A comparison of jackknife and analytic covariance matrices for ILLUSTRISTNG. Here we show the square root of the diagonal of the two covariances, for four snapshots and three two-point functions, as labelled.

APPENDIX B: POSTERIOR CONSTRAINTS FROM MASSIVEBLACK-II & ILLUSTRIS-1

In this appendix, we present the full posterior constraints on our three simulated samples. In the main body of this work we presented only a selection of these to emphasize our most interesting findings. For completeness, they are shown in Figs B1 and B2. This represents the baseline TATT analysis on our three simulations at $z = 0$ and $z =$

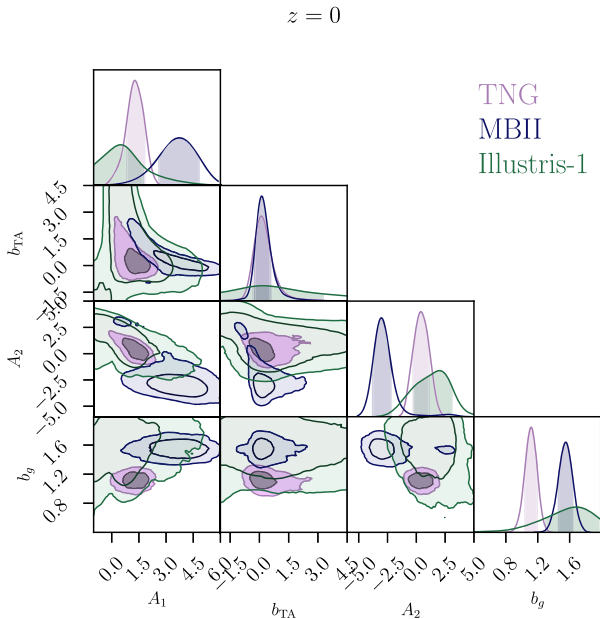


Figure B1. Posterior TATT model constraints from the lowest redshift snapshot of ILLUSTRISTNG, MASSIVEBLACK-II, and ILLUSTRIS-1.

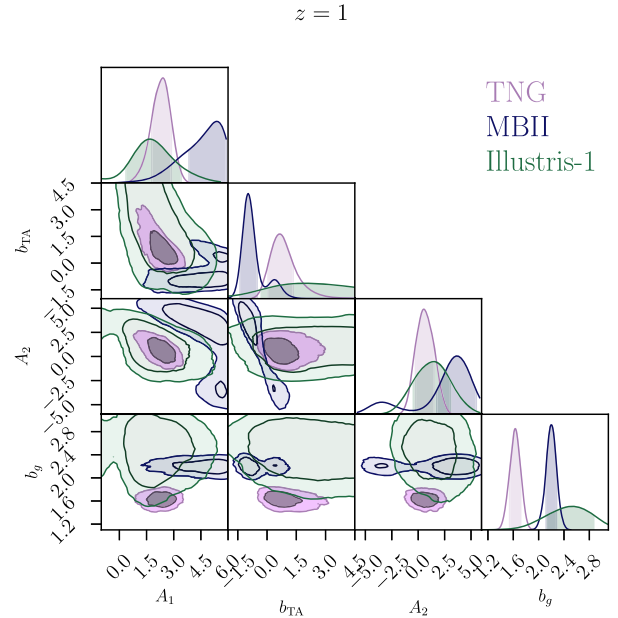


Figure B2. The same as B1, but at $z = 1$.

Table B1. The best-fitting TATT parameters and 1σ posterior uncertainties from all samples/redshifts considered in this work.

Simulation	Redshift	A_1	A_2	b_{TA}
TNG	0.0	0.68 ± 0.50	0.60 ± 0.62	0.73 ± 1.75
TNG	0.3	1.48 ± 0.54	0.63 ± 0.68	0.36 ± 0.80
TNG	0.62	1.51 ± 0.50	0.68 ± 0.67	0.63 ± 0.88
TNG	1.0	1.96 ± 0.56	0.54 ± 0.81	0.66 ± 0.88
MBII	0.0	3.57 ± 1.07	-2.76 ± 1.14	0.26 ± 0.63
MBII	0.3	3.24 ± 1.27	0.47 ± 1.82	-0.28 ± 0.75
MBII	0.62	4.09 ± 1.15	1.85 ± 2.26	-0.94 ± 0.58
MBII	1.0	4.53 ± 1.15	2.44 ± 2.73	-0.70 ± 0.61

Table B1 – *continued*

Simulation	Redshift	A_1	A_2	b_{TA}
Illustris	0.0	0.48 ± 1.47	1.29 ± 1.49	-0.05 ± 2.83
Illustris	0.3	3.07 ± 1.77	-0.40 ± 1.53	-1.03 ± 1.24
Illustris	0.62	1.03 ± 2.11	2.10 ± 1.72	-0.09 ± 3.62
Illustris	1.0	1.32 ± 1.90	1.34 ± 1.61	1.34 ± 3.14

1. As described, the fiducial analysis has four free parameters (A_1 , A_2 , b_{TA} , b_g), includes the joint data vector $w_{gg} + w_{g+} + w_{++}$, and scale cuts $r_p > 6 h^{-1}$ Mpc for all correlation functions.

APPENDIX C: IMPACT OF GALAXY WEIGHTS

To allow a meaningful comparison of galaxy samples from the three simulations included in this paper, we derive a set of galaxy weights for our MASSIVEBLACK-II and ILLUSTRIS-1 catalogues. Our comparison in Section 5 makes use of this scheme. The idea here is to weight the galaxy samples such that the distribution of host halo masses match exactly. The weight assigned to galaxy i from simulation X is then:

$$w_i^X = p^{\text{TNG}}(M_h^i) / p^X(M_h^i), \quad (\text{C1})$$

where $p^X(M_h)$ is the normalized histogram of host halo masses in simulation X , and j is a mass bin to which galaxy i belongs.

The impact of this weighting on our IA constraints is shown in Fig. C1. Although for clarity we show only the contours from the lowest redshift, we find very similar behaviour in the other three snapshots. It has been established elsewhere that there is a relatively tight relation between host halo mass and bias, at least on large physical scales; it is, then, intuitively correct that the weighting should bring the galaxy bias constraints (upper panel) from the two simulations into relatively close agreement.

The reweighting scheme described above is motivated by the fact that IAs are known to be dependent on galaxy (subhalo) mass, and so differences in the stellar mass function can very easily translate into differences in w_{g+} measurements (see e.g. Velliscig et al. 2015’s figs 2 and 3). Although this does not account for inherent

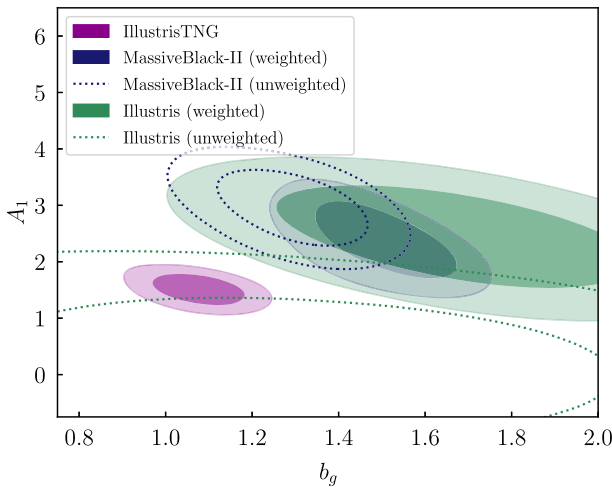


Figure C1. Demonstration of the impact of galaxy reweighting on the posterior IA constraints. Here we show the $z = 0$ constraints on the NLA model amplitude and linear galaxy bias from MASSIVEBLACK-II and ILLUSTRIS-1, with and without weighting. By construction ILLUSTRISTNG (shown in purple) is unaffected by weighting.

differences in the galaxy–halo connection and the subgrid physics of the simulations, such effects are both more interesting (in the sense that the impact on IAs is less well understood at a basic level), and more difficult to correct for.

APPENDIX D: VALIDITY OF THE LINEAR GALAXY BIAS APPROXIMATION

Since our perturbative TATT model includes higher order terms, there is some value in seeking to push to slightly smaller scales. It is also true, however, that as one does so, one eventually enters the regime in which non-linear galaxy bias also starts to become relevant. Such higher order bias contributions, and the cross-IA terms, are complex to model and not fully implemented in our analysis; if it exists, then, we would ideally like to identify a range of scales below our fiducial cut at $r_p = 6 h^{-1}$ Mpc, on which the linear bias approximation is valid (or, at least, deviations from it are subdominant to other uncertainties).

We can obtain an estimate for the effective scale-dependent galaxy bias in ILLUSTRISTNG as the ratio of the galaxy–galaxy projected correlation, and the matter–matter equivalent:

$$b'_g(r_p|z) = \sqrt{\frac{w_{gg}}{w_{\delta\delta}}}. \quad (\text{D1})$$

We refer to this as an ‘effective’ bias because, strictly speaking, the galaxy bias is defined in terms of the 3D density field $b_g = \delta_g / \delta$ (or equivalently in terms of 3D power spectra). Converting from 3D power spectra to projected correlations w_{gg} and w_{g+} involves an integral over k (e.g. equations 21 and 22), and if b_g is scale-dependent, it no longer separates cleanly from that integral. What we measure, then, is an effective bias b'_g , which is not quite the same as the true 3D galaxy bias $b_g(k)$.

The discrete snapshots in the simulation allow a relatively clean measurement of b'_g at a given redshift. For this exercise, we use the measured w_{gg} correlation in a particular snapshot. Although, of course, this includes some level of statistical noise, the signal-to-noise is relatively high. For the matter–matter part it is sufficient to use the theory prediction at the input ILLUSTRISTNG cosmology. While HALOFIT is subject to its own uncertainties on small scales (~ 5 per cent at $k < 1 h^{-1}$ Mpc; Takahashi et al. 2012), we do not expect them to alter the conclusions of our approximate calculations.

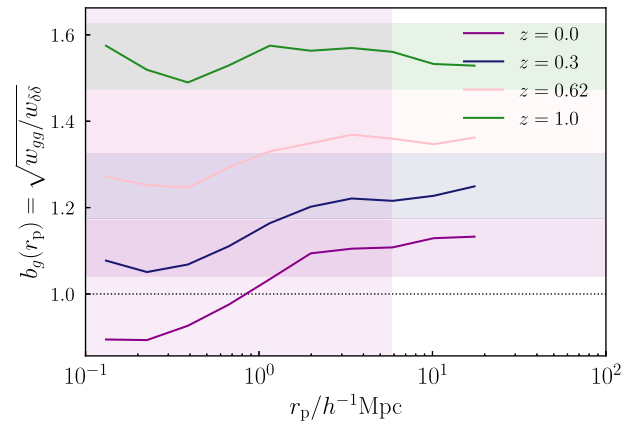


Figure D1. Galaxy bias as a function of physical scale. The bias is estimated as the ratio of the matter–matter and galaxy–galaxy projected correlations. The horizontal shaded bands show the best-fitting linear bias values and the 1σ uncertainties, as obtained from fits to the large scale w_{gg} correlations.

The resulting effective scale-dependent bias estimates are shown in Fig. D1. While the fiducial cut at $6 h^{-1}$ Mpc (the pink shaded region) does indeed effectively exclude scales on which the bias cannot be captured by a single coefficient, we can also see that it is relatively conservative. That is, there is a region from $r_p \sim 1 h^{-1}$ Mpc upwards, in which (within uncertainties) the bias is linear, but which are excluded by the fiducial cut. On the basis of these results, we carry out fits (see Section 6.1) with lower cutoffs as low as $1 h^{-1}$ Mpc.

At $z = 0$, we see that the linear bias assumption begins to break down in our ILLUSTRISTNG sample at $1 h^{-1}$ Mpc. It is worth bearing

in mind that in three dimensions, since one is measuring the actual galaxy bias, rather than b'_g , the approximation likely becomes invalid at some larger scale. By nature, the projected correlations at given r_p mix contributions from larger 3D separations, which bring them closer to linear bias. The breaking scale seems to shift downwards at high z , which is perhaps as a result of the growth of structure (i.e. bias is closer to linear at high redshift at a given r_p).

This paper has been typeset from a \LaTeX file prepared by the author.