Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **526,** 5494–5519 (2023) https://doi.org/10.1093/mnras/stad3107
Advance Access publication 2023 October 12

The FLAMINGO project: revisiting the S_8 tension and the role of baryonic physics

Ian G. McCarthy [1]★ Jaime Salcido [1] Joop Schaye [2] Juliana Kwan,[1] Willem Elbers [3] Roi Kugel [2] Matthieu Schaller [2,4] John C. Helly,[3] Joey Braspenning,[2] Carlos S. Frenk,[3] Marcel P. van Daalen [2] Bert Vandenbroucke [2] Jonah T. Conley,[1] Andreea S. Font[1] and Amol Upadhye[1]

[1]*Astrophysics Research Institute, Liverpool John Moores University, Liverpool, L3 5RF, UK*
[2]*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*
[3]*Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK*
[4]*Lorentz Institute for Theoretical Physics, Leiden University, PO box 9506, NL-2300 RA Leiden, the Netherlands*

Accepted 2023 October 8. Received 2023 October 2; in original form 2023 September 12

ABSTRACT

A number of recent studies have found evidence for a tension between observations of large-scale structure (LSS) and the predictions of the standard model of cosmology with the cosmological parameters fit to the cosmic microwave background (CMB). The origin of this 'S_8 tension' remains unclear, but possibilities include new physics beyond the standard model, unaccounted for systematic errors in the observational measurements and/or uncertainties in the role that baryons play. Here, we carefully examine the latter possibility using the new FLAMINGO suite of large-volume cosmological hydrodynamical simulations. We project the simulations onto observable harmonic space and compare with observational measurements of the power and cross-power spectra of cosmic shear, CMB lensing, and the thermal Sunyaev-Zel'dovich (tSZ) effect. We explore the dependence of the predictions on box size and resolution and cosmological parameters, including the neutrino mass, and the efficiency and nature of baryonic 'feedback'. Despite the wide range of astrophysical behaviours simulated, we find that baryonic effects are not sufficiently large to remove the S_8 tension. Consistent with recent studies, we find the CMB lensing power spectrum is in excellent agreement with the standard model, while the cosmic shear power spectrum, tSZ effect power spectrum, and the cross-spectra between shear, CMB lensing, and the tSZ effect are all in varying degrees of tension with the CMB-specified standard model. These results suggest that some mechanism is required to slow the growth of fluctuations at late times and/or on non-linear scales, but that it is unlikely that baryon physics is driving this modification.

Key words: methods: numerical – galaxies: clusters: general – galaxies: formation – large-scale structure of Universe – cosmology: theory.

1 INTRODUCTION

The standard model of cosmology, the so-called ΛCDM model, is based on the Friedmann–Lemaître–Robertson–Walker solution to Einstein's field equations for an isotropic and homogeneous universe. The standard model contains only six free parameters but describes a wealth of large-scale cosmological data remarkably well, including the temperature and polarization anisotropies in the cosmic microwave background (CMB), measurements of baryon acoustic oscillations (BAOs) in the clustering of galaxies, the redshift–distance relations of supernovae (Sn) Type Ia, and measurements of the growth of large-scale structure (LSS) including the abundance of galaxy clusters, galaxy clustering, cosmic shear and CMB lensing, and the thermal Sunyaev-Zel'dovich (tSZ) effect (see e.g. Planck Collaboration VI 2020a). Fits to these data point to a Universe, which is spatially flat and whose present-day energy density is dominated by dark matter and dark energy. However, the physical nature of these components has so far remained elusive.

A fruitful avenue for exploration into the nature of dark matter and dark energy, and to test the standard cosmological paradigm generally, is to look for signs of deviations in cosmological data sets from the predictions of the standard model and its possible extensions. On this front, there has been much activity in the past few years as the fidelity of cosmological data sets has rapidly increased. Interestingly, even though the standard model describes many cosmological data sets extremely well, the best-fitTING parameter values from different observables do not always appear fully consistent with each other. The most notable example of this is the so-called 'Hubble tension', which is that measurements of the local expansion rate of space yield a value for Hubble's constant, H_0, that is larger than predicted by the standard model of cosmology when it is fit to the CMB temperature and polarization anisotropies and BAO data (e.g. Planck Collaboration VI 2020a; Riess et al. 2022). This tension has now reached the $\approx 5\sigma$ level (Riess et al. 2022).

Another notable tension, which we focus on in the present study, is the so-called 'S_8 tension', where S_8 is defined as $\sigma_8\sqrt{\Omega_{\rm m}/0.3}$, where

★ E-mail: i.g.mccarthy@ljmu.ac.uk

Ω_m is the present-day matter density and σ_8 is the linearly evolved variance of the present-day matter density field filtered on a 8 Mpc h^{-1} scale. Note that S_8 best describes the combination of σ_8 and Ω_m that is constrained by cosmic shear (weak lensing) data. In this case, the best-fitting value of S_8 from several low-redshift observations of LSS, including cosmic shear (e.g. Heymans et al. 2021; Abbott et al. 2022) and other probes, appears to be in mild ($\approx$1–3σ) tension with the predictions of the standard model with parameter values specified by the CMB and BAO and, as we discuss below, also CMB lensing. While a tension of this magnitude is often not regarded as being statistically compelling, it is worth highlighting here that the tension has been persistent for nearly a decade now, since the first Planck data release, and spans several independent probes, each of which appear to show tensions with this level of significance and in the same direction (e.g. see fig.1 of McCarthy et al. 2018).

Various possible solutions have been put forward to reconcile the low-redshift LSS observations with the primary CMB + BAO combination. This includes mischaracterized systematic uncertainties in the LSS observations (e.g. photometric redshift, galaxy shape, and intrinsic alignment uncertainties in cosmic shear measurements, non-linear biasing in galaxy clustering, halo mass biasing in cluster counts, etc.), or possibly even in the primary CMB measurements (e.g. Addison et al. 2016; Planck Collaboration LI 2017; but see Rosenberg, Gratton & Efstathiou 2022). On the theory side, LSS tests of cosmology often probe the non-linear regime and therefore require cosmological simulations, or models that have been calibrated on such simulations, to predict the clustering of matter on small scales and at late times. Furthermore, part and parcel of this non-linear evolution is that matter collapses along filaments forming 'haloes' at the nodes. Here, the densities reach sufficiently high values that radiative cooling of the gas becomes efficient, leading to further collapse and eventually galaxy formation (e.g. White & Frenk 1991). With this comes a variety of energetic feedback processes associated with the formation of stars and the accretion of matter onto supermassive black holes.

It is straightforward to show that the energy released by the accretion of matter onto supermassive black holes can be of cosmological significance if it is able to efficiently couple to the gas (e.g. Silk & Rees 1998). That is, the energy is sufficient to expel baryons from galaxy groups (e.g. McCarthy et al. 2010, 2011), which will also lead to a back reaction on the dark matter halo (e.g. Van Daalen et al. 2011). In short, in the presence of energetic feedback, we expect the clustering of matter to be significantly affected on non-linear scales, and the results of full cosmological hydrodynamical simulations (e.g. Van Daalen et al. 2011; Mummery et al. 2017; Springel et al. 2018; Van Daalen, McCarthy & Schaye 2020; Salcido et al. 2023) as well as analytic halo models that use the observed baryon content of massive haloes as input (e.g. Debackere, Schaye & Hoekstra 2020) back up this physical intuition. Recent studies have shown that if such processes are not accounted for, they will lead to significant biases in the recovered cosmological parameters in forthcoming surveys (e.g. Semboloni et al. 2011; Semboloni, Hoekstra & Schaye 2013; Schneider et al. 2020; Castro et al. 2021; for a recent review see Chisari et al. 2019).

But what role, if any, do unaccounted for (or mischaracterized) baryonic effects have on the current S_8 tension? Based on the BAHAMAS simulations (McCarthy et al. 2017), we have previously argued that the effects are likely to be too small to explain the current tension (McCarthy et al. 2018). Consistent with this are the findings of several recent analyses of cosmic shear data, which have made marginal detections of the impact of baryons on the matter clustering and find that its magnitude aligns well with the predictions of cosmological hydrodynamical simulations such as BAHAMAS (see, e.g, Tröster et al. 2022; Aricò et al. 2023; Chen et al. 2023). However, Amon & Efstathiou (2022) and Preston, Amon & Efstathiou (2023) have recently challenged this conventional wisdom, showing that the existing constraints on the impact of baryons are sensitive to the adopted priors on the baryon parameters and that with a wider set of priors, allowing for much more aggressive feedback beyond what is typically simulated, it may be possible after all to reconcile the primary CMB(+BAO + CMB lensing) measurements with low-redshift LSS measurements.

In the present study, we revisit the S_8 tension and the role that baryons play. We use the new FLAMINGO suite of large-volume cosmological simulations (Kugel et al. 2023; Schaye et al. 2023) that includes variations in box size, resolution, cosmology (including massive neutrino cosmologies), and, importantly, a careful, systematic variation of the efficiencies and nature of feedback from star formation and active galactic nuclei (AGNs). We project the simulations onto observable harmonic space and make predictions for the power and cross-power spectra of cosmic shear, CMB lensing, and the tSZ effect, which sample fluctuations over a very wide range of physical scales and redshifts and have different sensitivities to halo mass. While there is a significant degree of overlap in terms of the scales between the various auto- and cross-power spectra (allowing for important consistency checks), there is also a great deal of complementarity. We compare the FLAMINGO predictions with the latest measurements of these quantities from the KiDS 1000 (Tröster et al. 2022) and DES Y3 survey (Doux et al. 2022) cosmic shear data, Planck and South Pole Telescope (SPT) tSZ data (Bolliet et al. 2018 and Reichardt et al. 2021, respectively), and Planck, ACT, and SPT CMB lensing data (Wu et al. 2019, Planck Collaboration VIII 2020b, and Qu et al. 2023, respectively). Despite the wide range of astrophysical behaviours simulated, we will show that baryonic effects and their uncertainties encapsulated within the FLAMINGO simulations are not sufficiently large to significantly alter the current S_8 discussion.

The present study is organized as follows. In Section 2, we describe the FLAMINGO suite of cosmological hydrodynamical simulations and our approach to projecting these simulations onto observable harmonic space. In Section 3, we present our main results, including an examination of the box size, resolution, cosmological, and feedback dependencies of the auto- and cross-spectra involving cosmic shear, CMB lensing, and tSZ effect data. We also compare to the most recent measurements of these quantities. In Section 4, we summarize our main findings and conclude.

2 SIMULATIONS AND COMPUTATION OF COSMOLOGICAL OBSERVABLES

2.1 Description of simulations

We provide here a brief summary of the FLAMINGO simulations, referring the reader to Schaye et al. (2023) and Kugel et al. (2023) for in depth presentations.

The FLAMINGO simulations were performed using SWIFT (Schaller et al. 2023), a fully open-source coupled cosmology, gravity, hydrodynamics, and galaxy formation code.[1] Gravitational forces are computed using a 4$^{\rm th}$-order fast multipole method (Greengard & Rokhlin 1987; Cheng, Greengard & Rokhlin 1999; Dehnen 2014) on

[1] Publicly available, including the version used for these simulations, at www.swiftsim.com

Table 1. FLAMINGO hydrodynamical simulations. The first four lines list the simulations that use the fiducial galaxy formation model and assume the fiducial cosmology (D3A) but use different volumes and resolutions. The remaining lines list the model variations, which all use a 1 Gpc box and intermediate resolution. The columns list the simulation identifier (where m8, m9, and m10 indicate $\log_{10}$ of the mean initial baryonic particle mass and correspond to high, intermediate, and low resolution, respectively; absence of this part implies m9 resolution); the number of standard deviations by which the observed stellar masses are shifted before calibration, Δm_*; the number of standard deviations by which the observed cluster gas fractions are shifted before calibration, $\Delta f_{\rm gas}$; the AGN feedback implementation (thermal or jets); the comoving box side length, L; the number of baryonic particles, $N_{\rm b}$ (which equals the number of CDM particles, $N_{\rm CDM}$); the number of neutrino particles, N_ν; the initial mean baryonic particle mass, $m_{\rm g}$; the mean CDM particle mass, $m_{\rm CDM}$; the Plummer-equivalent comoving gravitational softening length, $\epsilon_{\rm com}$; the maximum proper gravitational softening length, $\epsilon_{\rm prop}$; and the assumed cosmology which is specified in Table 2.

Identifier	Δm_* (σ)	$\Delta f_{\rm gas}$ (σ)	AGN	L (cGpc)	$N_{\rm b}$	N_ν	$m_{\rm g}$ ($\rm M_\odot$)	$m_{\rm CDM}$ ($\rm M_\odot$)	$\epsilon_{\rm com}$ (ckpc)	$\epsilon_{\rm prop}$ (pkpc)	Cosmology
L1_m8	0	0	thermal	1	3600^3	2000^3	1.34×10^8	7.06×10^8	11.2	2.85	D3A
L1_m9	0	0	thermal	1	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A
L1_m10	0	0	thermal	1	900^3	500^3	8.56×10^9	4.52×10^{10}	44.6	11.40	D3A
L2p8_m9	0	0	thermal	2.8	5040^3	2800^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A
fgas+2σ	0	+2	thermal	1	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A
fgas−2σ	0	−2	thermal	1	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A
fgas−4σ	0	−4	thermal	1	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A
fgas−8σ	0	−8	thermal	1	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A
M*−σ	−1	0	thermal	1	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A
M*−σ_fgas−4σ	−1	−4	thermal	1	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A
Jet	0	0	jets	1	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A
Jet_fgas−4σ	0	−4	jets	1	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A
Planck	0	0	thermal	1	1800^3	1000^3	1.07×10^9	5.72×10^9	22.3	5.70	Planck
PlanckNu0p24Var	0	0	thermal	1	1800^3	1000^3	1.06×10^9	5.67×10^9	22.3	5.70	PlanckNu0p24Var
PlanckNu0p24Fix	0	0	thermal	1	1800^3	1000^3	1.07×10^9	5.62×10^9	22.3	5.70	PlanckNu0p24Fix
LS8	0	0	thermal	1	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	LS8

small scales and a particle-mesh method solved in Fourier space on large scales, following the force splitting approach of Bagla & Ray (2003). The hydrodynamic equations are solved using the smoothed-particle hydrodynamics (SPH) method (for a review, see Price 2012), in particular the SPHENIX flavour of SPH (Borrow et al. 2022), which was designed specifically for simulations of galaxy formation.

The suite consists of the 16 hydrodynamical simulations, listed in Table 1 (reproduced from table 2 of Schaye et al. 2023), and 12 gravity-only simulations. We examine only the hydro simulations in the present study. The majority of the runs adopt a $(1\ \rm Gpc)^3$ cubic volume, denoted by 'L1' in the simulation identifier, although one run has a volume of $(2.8\ \rm Gpc)^3$ ('L2p8'). The simulations span three different resolutions ('m10', 'm9', and 'm8', where the number indicates the rounded logarithm base 10 of the baryonic particle mass), with the mass (spatial) resolution between consecutive resolutions changing by a factor of 8 (2). Most runs adopt intermediate resolution ('m9'), which corresponds to an (initial) mean baryonic particle mass of $\approx 1 \times 10^9\ \rm M_\odot$, a mean cold dark matter particle mass of $\approx 6 \times 10^9\ \rm M_\odot$, and a maximum proper gravitational softening length of 5.7 kpc, all of which are similar to our previous BAHAMAS simulations (McCarthy et al. 2017) but within considerably larger volumes than BAHAMAS. At $z > 2.91$ the softening length is held constant in comoving units at 22.3 kpc. All runs use equal numbers of baryonic and dark matter particles, while the number of neutrino particles is a factor 1.8^3 smaller. Table 1 provides the parameter values specifying the numerical resolution the various runs.

The values of the cosmological parameters for our fiducial model are the maximum-likelihood values from the Dark Energy Survey Year Three (DES Y3; Abbott et al. 2022) '3 × 2pt + All Ext.' ΛCDM cosmology ('D3A' in Table 2). These values assume a spatially flat universe and are based on the combination of constraints from DES Y3 '3 × 2-point' correlation functions: cosmic shear, galaxy clustering, and galaxy-galaxy lensing, with constraints from external data from BAO, redshift-space distortions, SN Type Ia, and Planck observations of the CMB (including CMB lensing), Big-Bang nucleosynthesis, and local measurements of the Hubble constant (see Abbott et al. 2022 for details). Our fiducial cosmology, D3A, uses the minimum neutrino mass allowed by neutrino oscillation experiments of $\sum m_\nu = 0.06$ eV (Esteban et al. 2020; de Salas et al. 2021), which is consistent with the 95 per cent confidence upper limit of 0.13 eV from DES Y3. Note that the simulations include neutrino particles using the new δf method of Elbers et al. (2021).

For the purposes of the present study, it is important to highlight the inclusion of the Planck primary CMB constraints in the D3A cosmology. As already discussed in Section 1, there is a mild tension between some LSS observables, such as cosmic shear, and the primary CMB, such that the latter prefers a larger value of S_8. Hence, a joint fit to these observables will generally result in a value of S_8 that will be larger than preferred by cosmic shear alone, and the statistical precision of the Planck data set is such that the joint value is closer to that preferred by the primary CMB. This will become relevant later on, when we compare the predictions of the simulatons to LSS observables.

An important aspect of our hydrodynamical simulations is the calibration of parameters which characterise the efficiencies of stellar and AGN feedback. Following our approach in BAHAMAS (McCarthy et al. 2017), the subgrid models for BH accretion and for stellar and AGN feedback are calibrated to the observed $z = 0$ galaxy stellar mass function (SMF), gas mass fractions within R_{500c} for galaxy groups and clusters at $z \approx 0.1$–0.3 from a combination of X-ray and weak lensing data, and the $z = 0$ relation between BH mass and stellar mass. Our choice of calibration observables is motivated by the fact that the impact of baryon physics on LSS is highly correlated with the baryon fractions of galaxy groups and clusters (e.g. Semboloni et al. 2011; Semboloni, Hoekstra & Schaye 2013; Schneider et al. 2019; Van Daalen, McCarthy & Schaye 2020; Salcido et al. 2023), as these objects dominate the matter

Table 2. The values of the cosmological parameters used in different simulations. The columns list the prefix used to indicate the cosmology in the simulation name (note that for brevity the prefix 'D3A' that indicates the fiducial cosmology is omitted from the simulation identifiers); the dimensionless Hubble constant, h; the total matter density parameter, Ω_m; the dark energy density parameter, Ω_Λ; the baryonic matter density parameter, Ω_b; the sum of the particle masses of the neutrino species, $\sum m_\nu$; the amplitude of the primordial matter power spectrum, A_s; the power-law index of the primordial matter power spectrum, n_s; the amplitude of the initial power spectrum parametrized as the r.m.s. mass density fluctuation in spheres of radius $8\,h^{-1}$ Mpc extrapolated to $z = 0$ using linear theory, σ_8; the amplitude of the initial power spectrum parametrized as $S_8 \equiv \sigma_8\sqrt{\Omega_\mathrm{m}/0.3}$; the neutrino matter density parameter, $\Omega_\nu \cong \sum m_\nu/(93.14\,h^2\,\mathrm{eV})$. Note that the values of the Hubble and density parameters are given at $z = 0$. The values of the parameters that are listed in the last three columns have been computed from the other parameters.

Prefix	h	Ω_m	Ω_Λ	Ω_b	$\sum m_\nu$	A_s	n_s	σ_8	S_8	Ω_ν
-	0.681	0.306	0.694	0.0486	0.06 eV	2.099×10^{-9}	0.967	0.807	0.815	1.39×10^{-3}
Planck	0.673	0.316	0.684	0.0494	0.06 eV	2.101×10^{-9}	0.966	0.812	0.833	1.42×10^{-3}
PlanckNu0p24Var	0.662	0.328	0.672	0.0510	0.24 eV	2.109×10^{-9}	0.968	0.772	0.807	5.87×10^{-3}
PlanckNu0p24Fix	0.673	0.316	0.684	0.0494	0.24 eV	2.101×10^{-9}	0.966	0.769	0.789	5.69×10^{-3}
LS8	0.682	0.305	0.695	0.0473	0.06 eV	1.836×10^{-9}	0.965	0.760	0.766	1.39×10^{-3}

clustering signal (e.g. van Daalen & Schaye 2015; Mead et al. 2020). For FLAMINGO we use a systematic Bayesian approach to the fitting that has previously been applied to the semi-analytic model GALFORM (Bower et al. 2010; Rodrigues, Vernon & Bower 2017) and to a variety of cosmological emulators based on gravity-only simulations (e.g. Heitmann et al. 2014; Lawrence et al. 2017; Euclid Collaboration 2019). As described in Kugel et al. (2023), we employ machine learning to fit the subgrid parameters to the calibration data. We use Gaussian process emulators trained on 32-node Latin hypercubes of simulations. The 32 nodes are distributed approximately randomly in the hypercube so that the minimum distance between the nodes is maximized. A hydrodynamical simulation is run for each node and we then build a separate emulator for each observable based on all 32 simulations. The SMF emulator takes as input the stellar mass, M_*, and the subgrid parameter vector, $\boldsymbol{\theta}$, and it predicts the SMF, $f(M_*)$. The inputs for the gas fraction emulator are the total group/cluster mass, $M_{500\mathrm{c}}$ (i.e. the mass inside the radius $R_{500\mathrm{c}}$ within which the mean density is 500 times the critical density), and the subgrid parameters $\boldsymbol{\theta}$. It outputs the gas fraction as a function of mass, $f_{\mathrm{gas,\,500c}}(M_{500\mathrm{c}})$.

We re-calibrate the feedback model as the resolution of the simulations is varied. This is motivated by the fact that a higher-resolution simulation resolves smaller scales and higher gas densities and will therefore, for example, yield different radiative losses and different BH accretion rates (and hence different AGN feedback), which will, in turn, change the structure of the interstellar medium even on scales resolved by the lower-resolution run (e.g. Schaye et al. 2015). Another novel aspect of our approach is that the calibration takes into account the expected observational errors and biases. We impose random errors on the simulated stellar masses to account for Eddington bias. During the calibration of the fiducial intermediate-resolution model we fit for systematic errors in the SMF due to cosmic variance, bias in the inferred stellar mass, and for hydrostatic mass bias in the cluster gas fractions inferred from X-ray observations. The best-fitting bias factors, which are negligible for cosmic variance and stellar mass, and which is consistent with the literature for the hydrostatic mass bias, are then applied to the calibration data for all resolutions and models.

Note that the emulators are not only used to design simulations that reproduce the observations, but also to create models in which the SMF and/or cluster gas fractions are shifted to higher/lower values. This allows us to specify model variations in terms of the number of σ by which they deviate from the calibration data, which is more intuitive and useful than specifying simulations solely by the values of subgrid parameters that are not directly observable. FLAMINGO includes four models in which cluster gas fractions are varied (by $+2$, -2, -4, and -8σ, respectively) while keeping the SMF unchanged, one model in which the SMF is reduced by decreasing the stellar masses by the expected systematic error (0.14 dex; Behroozi et al. 2019) while keeping gas fractions fixed, and two models that simultaneously vary the gas fractions and the SMF. Finally, two models use jet-like AGN feedback rather than the fiducial isotropic and thermal AGN feedback, one of which is calibrated to the fiducial data and one to gas fractions shifted down by 4σ. Comparison of these last two models with the corresponding fiducial ones enables estimates of the uncertainty due to differences in the implementation of AGN feedback for a common calibration.

FLAMINGO includes four intermediate-resolution hydrodynamical simulations with the fiducial calibration of the subgrid physics in $(1\ \mathrm{Gpc})^3$ volumes that vary the cosmological parameters. Three of the alternative cosmologies we consider are variations on Planck Collaboration VI (2020a): their best-fitting ΛCDM model with the minimum allowed neutrino mass, $\sum m_\nu = 0.06$ eV ('Planck'); a model with a high neutrino mass, $\sum m_\nu = 0.24$ eV, (allowed at 95 per cent confidence by Planck) in which the other cosmological parameters take their corresponding best-fitting values from the Planck MCMC chains ('PlanckNu0p24Var'); and a model with the same high neutrino mass, $\sum m_\nu = 0.24$ eV, that keeps all other parameters fixed to the values of model Planck, except for Ω_CDM, which was reduced in order to keep Ω_m constant ('PlanckNu0p24Fix'). Note that for the latter model we fix the primordial power spectrum amplitude, A_s, rather than S_8. All models with $\sum m_\nu = 0.24$ eV use three massive neutrino species of 0.08 eV. Finally, we include the 'lensing cosmology' from Amon et al. (2023) ('LS8'). This model has a lower amplitude of the power spectrum, $S_8 = 0.766$, compared with 0.815 and 0.833 for D3A and Planck, respectively. Amon et al. (2023) show that the lensing cosmology is consistent with observations of galaxy clustering from BOSS DR12 (Reid et al. 2016) and galaxy-galaxy lensing from D3A (Abbott et al. 2022), HSC Y1 (Aihara et al. 2018) and KiDS 1000 (Kuijken et al. 2019) over a wide range of scales, $0.15 - 60\ h^{-1}$ Mpc, if allowances are made for theoretical uncertainties associated with baryonic feedback and assembly bias. By contrast, they show that the Planck cosmology does not fit the same data on small scales. We note that Heymans et al. (2021) showed that the LS8 model is also consistent with KiDS-1000 cosmic shear measurements.

The simulations are initialized at $z = 31$, using multifluid third-order Lagrangian perturbation theory (3LPT) ICs generated with the MONOFONIC code (Hahn et al. 2020; Michaux et al. 2021). The

ICs accurately reproduce the relative growth of the individual fluid components. For FLAMINGO, a modified version of MONOFONIC was used that implements the effects of massive neutrinos.[2] We use the prescriptions for 3-fluid ICs with CDM, baryons, and massive neutrinos outlined in Elbers et al. (2022), which builds on the 2-fluid formalism of Rampf, Uhlemann & Hahn (2021) and Hahn, Rampf & Uhlemann (2021). CDM and baryon particles are set up in a two-stage process. First, the combined mass-weighted CDM + baryon fluid is initialized with single-fluid 3LPT, accounting for the presence of neutrinos. This single fluid is then split into separate components with distinct transfer functions by perturbing the masses and velocities in accordance with the first-order compensated mode. Hahn, Rampf & Uhlemann (2021) showed that discreteness errors can be suppressed by perturbing particle masses rather than displacements, thereby eliminating spurious growth of the compensated mode (see also Bird et al. 2020; Liu et al. 2023). The underlying Gaussian random fields were chosen from subregions of PANPHASIA to facilitate future zoom-in resimulations (Jenkins 2013). To limit cosmic variance without compromising the ability to do zooms, we used partially fixed ICs (Angulo & Pontzen 2016), setting the amplitudes of modes with $(kL)^2 < 1025$ to the mean variance, where k is the wavenumber and L is the side length of the simulation box. For a more in depth description of the ICs, please see Schaye et al. (2023).

2.2 Projecting to cosmological observables

Below, we describe how the FLAMINGO data set accompanying each simulation is projected onto cosmological observables. In the present study, we focus on observables in spherical harmonics space, namely the angular power spectrum, and reserve a configuration-space analysis for a future study.

2.2.1 *Fiducial Limber 1D analysis*

The observed 2D angular power spectrum, C_ℓ, between two fields at a multipole moment ℓ can be derived employing the Limber approximation (i.e. fluctuations are only in the plane of the sky) and integrating the relevant weighted 3D power spectrum along the line of sight (e.g. Kaiser 1992; Tröster et al. 2022):

$$C_\ell = \int_0^{\chi(z_{\rm max})} \frac{W^{\rm A}(\chi) W^{\rm B}(\chi)}{\chi^2} P_{\rm A,B}\left(\frac{\ell+\frac{1}{2}}{\chi}, z(\chi)\right) {\rm d}\chi \ , \qquad (1)$$

where $P(k, z)$ is the relevant 3D power (or cross-power) spectrum, $W^{\rm A}$ and $W^{\rm B}$ are the window functions (or weighting kernels) of the two fields, and the integral is taken over comoving distance, χ, back to $\chi(z_{\rm max})$ where $z_{\rm max}$ is the maximum redshift, which we specify below for each of the power spectra.

Our focus in the present study will be on the angular auto- and cross-spectra between cosmic shear.[3], CMB lensing, and the tSZ effect. As we have already investigated the cross-spectrum between CMB lensing and the tSZ effect in Schaye et al. (2023), we will concentrate here on the shear, CMB lensing, and tSZ effect autopower spectra and the shear–tSZ and shear–CMB lensing cross-spectra. Taken together with the CMB lensing–tSZ effect cross in Schaye et al. (2023), we will therefore have examined all possible auto- and cross-spectra of these three observables. Note that when examining auto- and cross-spectra that involve cosmic shear and CMB lensing but not the tSZ effect, the relevant 3D power spectrum, $P_{\rm A,B}$, in equation (1) is the matter power spectrum, $P_{\rm m,m}(k, z)$, which is computed using the matter overdensity field and has units of Mpc^3. For the tSZ effect auto power spectrum, $P_{\rm A,B}$ is the 3D electron pressure power spectrum, $P_{\rm e,e}(k, z)$, which is computed using the electron pressure field and has units of eV^2 cm^{-6}. For the cosmic shear–tSZ effect cross, $P_{\rm A,B}$ is the 3D matter–electron pressure cross-spectrum, $P_{\rm m,e}(k, z)$ and has units of Mpc^3 eV cm^{-3}. We calculate the shot noise-subtracted auto- and cross-power spectra of the hydrodyamical simulations following a procedure that is equivalent to that described in the appendix of Mead et al. (2020).

To ensure accurate computation of the observable power spectra via the Limber approximation, 3D power spectra are output on the fly with a high redshift cadence, particularly at low redshifts where non-linear and baryonic effects are most evident. Specifically, we adopt an output frequency of $\Delta z = 0.05$ between $z = 0$ and $z = 3$ (60 outputs), $\Delta z = 0.25$ between $z = 3$ and $z = 12$ (36 outputs), $\Delta z = 0.5$ between $z = 12$ and $z = 20$ (16 ouputs), and $\Delta z = 1$ between $z = 20$ and $z = 30$ (10 outputs). Note that the fine sampling of $\Delta z = 0.05$ below $z = 3$ was deliberately chosen to match that of the background source redshift distributions of the KiDS and DES data sets. We have tested the convergence of our calculations by using only half of the 3D power spectra (every second one) and find that resulting power spectra agree with those from our full calculation to typically better than a per cent accuracy over the range of scales examined here.

In the present study, we limit our analysis to multipoles with $\ell > 100$ (corresponding to $\theta \approx 1^\circ$), for which our assumption of a flat sky, which is implicit in equation (1), is highly accurate. This multipole limit is also motivated by the fact that our 2D light cone-based maps (described below) have been constructed assuming a flat sky and that for some tests that we examine which have a significant high-redshift contribution (e.g. CMB lensing), the simulation box size prevents us from probing very large angular scales.

In the flat-sky limit, the window function of the shear field of the i-th source sample, $W^{\gamma_i}(\chi)$, may be written as (e.g. Bartelmann & Schneider 2001)

$$W^{\gamma_i}(\chi) = \frac{3}{2}\left(\frac{H_0}{c}\right)^2 \Omega_{\rm m} \frac{\chi}{a(\chi)} \int_z^{z_{\rm max}} n_i(z') \left[1 - \frac{\chi(z)}{\chi(z')}\right] {\rm d}z' \ , \qquad (2)$$

where c is the speed of light, $a(\chi)$ is the scale factor at comoving distance χ, and $n_i(z)$ is the source redshift distribution of the sample i, which is normalized such that its integration from $z = 0$ to $z = z_{\rm max}$ is unity. For auto- and cross-spectra involving cosmic shear, $z_{\rm max}$ is set by the maximum redshift of the observed source redshift distribution. Here, we compare to KiDS 1000 and DES Y3 samples, for which $z_{\rm max} = 3$.

The window function of the CMB lensing convergence field, $W^{\kappa_{\rm CMB}}$, can be derived from equation (2) by replacing the source redshift distribution, $n(z)$, with the Dirac delta function (i.e. a single source plane) and integrating to yield

$$W^{\kappa_{\rm CMB}} = \frac{3}{2}\left(\frac{H_0}{c}\right)^2 \Omega_{\rm m} \frac{\chi}{a(\chi)} \left(1 - \frac{\chi}{\chi_{\rm CMB}}\right) \ , \qquad (3)$$

where $\chi_{\rm CMB}$ is the comoving distance to the surface of last-scattering, assumed to be at $z_{\rm CMB} = 1100$. For the CMB lensing auto power spectrum, which has a very extended window in redshift space (i.e. is sensitive to fluctuations over a wide range of distances), we integrate equation (1) back to the initial conditions of the simulations, corresponding to $z_{\rm max} = 31$. We have checked that contributions from higher redshifts are negligible by using the linear power spectrum beyond this maximum redshift. Note that when comparing to observational measurements below, we use both the CMB lensing

[2] https://github.com/wullm/monofonic

[3] Specifically E-mode shear, which we refer to here as just shear.

convergence ($\kappa_{\rm CMB}$) and deflection potential (ϕ) which, in terms of their angular power spectra, are related via $\phi_\ell = 2\kappa_{{\rm CMB},\,\ell}/[\ell(\ell+1)]$.

The window function of the tSZ effect, $W^y(\chi)$, is

$$W^y(\chi) = \frac{\sigma_{\rm T}}{m_{\rm e}c^2}\frac{1}{a^2(\chi)} \ , \tag{4}$$

where $\sigma_{\rm T}$ is the Thompson scattering cross-section and $m_{\rm e}$ is the electron rest mass. As for the case of the CMB lensing power spectrum, we integrate equation (1) for the tSZ power spectrum back to $z_{\rm max} = 31$. However we note that, as the tSZ effect emerges from the inverse-Compton scattering of CMB photons by hot free electrons, there is essentially no contribution from redshifts greater than $z = 7.8$, which corresponds to the redshift of reionization of the simulations. Furthermore, over the range of scales accessible to current observations, the tSZ effect power spectrum is mainly sensitive to massive, relatively nearby clusters (e.g. Komatsu & Seljak 2002; Battaglia et al. 2012; McCarthy et al. 2014) and the integrated signal is converged beyond $z \approx 3$.

2.2.2 *Map-based 2D analysis*

We describe here an alternative, map-based (2D) analysis of the simulations, which we will compare to the fiducial 1D analysis described above. This will provide a consistency check of our results but also allows us to get a handle on the role of cosmic variance, as our lightcones are constructed for multiple observer locations. Note that since our light cones in most cases are restricted to a maximum redshift of $z = 3$, we do not attempt to compute the CMB lensing auto power spectrum with a 2D analysis, as a non-negligible fraction of the signal comes from beyond this redshift for that statistic.

As described in the appendix of Schaye et al. (2023), tSZ effect maps are constructed on the fly by accumulating the Compton y values of individual particles crossing the light cone onto HEALPIX maps over fixed intervals in redshift. To construct a total (integrated) Compton y map, we simply sum these maps along the line of sight back to $z = 3$, which is sufficient for the tSZ power spectrum and the tSZ–cosmic shear cross-spectra that we consider.

To construct cosmic shear and CMB lensing convergence maps, we follow the method described in McCarthy et al. (2018), which employs the so-called Born approximation (i.e. light ray paths are approximated as straight lines). In short, for each HEALPIX total mass map (of which there are 60 per light cone back to $z = 3$, also produced on the fly), we compute a projected (2D) overdensity map, $\delta(\chi, \boldsymbol{\theta})$. The maps are then integrated along the line of sight weighted by the window function (lensing kernel) to yield the total convergence map:

$$\kappa(\boldsymbol{\theta}) = \int_0^{\chi(z_{\rm max})} W^{\gamma_i,\,\kappa_{\rm CMB}}(\chi)\,\delta(\chi,\boldsymbol{\theta}){\rm d}\chi \ , \tag{5}$$

where $W^{\gamma_i,\,\kappa_{\rm CMB}}(\chi)$ is the window function corresponding to either the i-th galaxy sample (as in equation (2)) for cosmic shear or the CMB lensing single-source plane (as in equation (3)), and $z_{\rm max} = 3$. Note that $z_{\rm max} = 3$ is sufficient for cross-correlations between CMB lensing, cosmic shear, and the tSZ effect.

We use the NAMASTER [4] package (Alonso et al. 2019) to compute the auto- and cross-spectra of the dimensionless scalar (spin-0) quantities y and κ. To save computational effort, the HEALPIX maps here have been downsampled from $N_{\rm side} = 16384$ to $N_{\rm side} = 4096$, corresponding to an angular resolution of ≈0.86 arcmin, which is sufficient for the comparisons to observations presented below. When computing the spectra, we initially use a multipole moment resolution (bandpower) of $\Delta\ell = 8$ but then employ a Savitzky-Golay filter of order 3 and window size of 15 to further smooth the simulated spectra. Note that the smoothing is applied for visual (plotting) purposes only. Unlike for the fiducial Limber 1D analysis, we do not quantitatively compare the smoothed 2D map-based power spectra to the observed power spectra (e.g. compute a goodness of fit). We deconvolve the $N_{\rm side} = 4096$ pixel window function from the computed cross-spectra using the *pixwin* function within the HEALPIX package.

We note that within the flat-sky limit adopted here, the C_ℓ's for (E-mode) shear are the same as those for the convergence, κ (e.g. Kilbinger et al. 2017; Wei et al. 2018). Thus, for our 2D cosmic shear analysis, there is no need to convert the convergence field, $\kappa(\boldsymbol{\theta})$, in equation (5) into a shear field, $\gamma_i(\boldsymbol{\theta})$, before computing the shear power spectra[5]

3 RESULTS

In this section, we present the main results of our analyses. We begin by examining the power spectra of cosmic shear, tSZ effect, and CMB lensing, before examining the cross-spectra between these observables. For each case, we make use of the full FLAMINGO suite of variations, exploring the dependence of the signals on cosmology, the efficiency and nature of feedback, and simulation volume and resolution. We compare the simulations with the most recent observational measurements of these quantities.

3.1 Effective length scales and redshifts probed by different observables

To help aid the interpretation of the results presented below and to assess consistency between the different tests, we show in Fig. 1 the C_ℓ-weighted mean k-scale ($k_{\rm eff}$; top panel) and redshift ($z_{\rm eff}$; bottom panel) of the various auto- and cross-power spectra considered here. To compute $k_{\rm eff}$ and $z_{\rm eff}$, we use the fiducial 1D Limber integration [equation (1)] with 3D power spectra from the largest FLAMINGO hydro run (L2p8_m9). More specifically, to compute $k_{\rm eff}$, we modify equation (1) to include an additional multiplicative term, $k \equiv (\ell + 1/2)/\chi$ and then integrate the modified equation over χ. We then divide the result by the integration of the unmodified equation (1), to yield the C_ℓ-weighted k scale, $k_{\rm eff}$. The calculation of $z_{\rm eff}$ is performed in an analogous way, but using $z(\chi)$ as the additional multiplicative term. Note that for cosmic shear ($\gamma_{\rm E} - \gamma_{\rm E}$), we use source redshift distributions of the 5 tomographic bins of the KiDS 1000 survey (Tröster et al. 2022; as discussed below in Section 3.2), apart from the shear–CMB lensing cross, for which we use the combined (single bin) distribution from Robertson et al. (2021), as described in Section 3.6.

The top panel of Fig. 1 shows that $k_{\rm eff}$ generally rises with increasing ℓ as expected, though the gradient is significantly more shallow for spectra involving tSZ y relative to pure lensing power spectra. In general, the cosmic shear power spectrum, tSZ effect power spectrum, and their cross-spectrum probe the smallest scales (largest $k_{\rm eff}$'s) at a given multipole, followed by the KiDS 2D cosmic shear–CMB lensing cross and CMB lensing–y cross, which probe

[4] https://namaster.readthedocs.io/en/latest/

[5] We have verified this to very high precision by converting the spin-0 convergence field into a spin-2 shear field using the method of Kaiser & Squires (1993) and then computing the shear power spectra using NAMASTER.

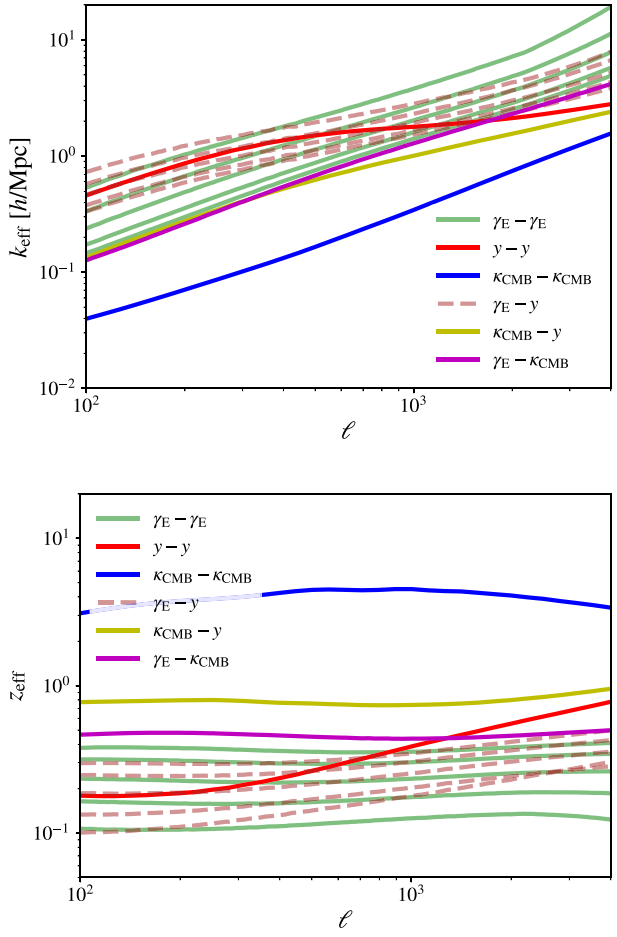

Figure 1. C_ℓ-weighted k-scale ($k_{\rm eff}$; top panel) and redshift ($z_{\rm eff}$; bottom panel) of the auto- and cross-power spectra involving cosmic shear, the tSZ effect, and CMB lensing considered in this paper. For cosmic shear power spectra ($\gamma_{\rm E}$-$\gamma_{\rm E}$), we plot only the autospectra for the five tomographic bins of the KiDS 1000 survey (i.e. 1–1, 2–2,..., 5–5, shown in light green curves), where the 1–1 power spectrum has the smallest effective redshift and the largest effective k. Similarly, for the cosmic shear–tSZ effect cross-spectrum (dashed brown curves), the 1-y cross has the smallest $z_{\rm eff}$ and the largest $k_{\rm eff}$. The top panel shows that $k_{\rm eff}$ rises with increasing ℓ, though the gradient is significantly shallower for spectra involving tSZ y relative to pure lensing power spectra. In general, the cosmic shear power spectrum, tSZ effect power spectrum, and their cross-probe the smallest scales (largest $k_{\rm eff}$'s) at a given ℓ, with the KiDS 2D cosmic shear–CMB lensing cross and CMB lensing–y cross-probing larger scales, and the CMB lensing power spectrum probing the largest scales. The effective redshift of the auto- and cross-spectra are rank ordered in the opposite sense to k, with the cosmic shear power spectrum, tSZ effect power spectrum, and their cross-spectrum being sensitive to the lowest redshifts (generally $z_{\rm eff} \approx 0.1$–0.4, apart from the tSZ effect power spectrum on small angular scales), followed by the KiDS 2D cosmic shear–CMB lensing cross ($z_{\rm eff} \approx 0.5$), the CMB lensing–y cross ($z_{\rm eff} \approx 0.8$), and the CMB lensing power spectrum sampling the highest redshifts ($z_{\rm eff} \approx 3$–4).

larger scales, and the CMB lensing power spectrum which samples the largest scales at a given multipole.

The bottom panel of Fig. 1 shows the effective redshifts of the auto- and cross-spectra are rank ordered in the opposite sense to k, with the cosmic shear power spectrum, tSZ effect power spectrum, and their cross-spectrum being sensitive to the lowest redshifts (generally $z_{\rm eff} \approx 0.1$–0.4, apart from the tSZ effect power spectrum on small angular scales that probes somewhat larger redshifts), followed by the KiDS 2D cosmic shear–CMB lensing cross ($z_{\rm eff} \approx 0.5$), the CMB lensing–y cross ($z_{\rm eff} \approx 0.8$) explored in Schaye et al. (2023), and the CMB lensing power spectrum being sensitive to the highest redshifts ($z_{\rm eff} \approx 3$–4). Thus, a simultaneous analysis of all these statistics will sample fluctuations over very wide ranges of redshift and physical scale, and in this respect, there is a significant degree of overlap but also complementarity between the different observables.

Another relevant dimension that we do not consider in Fig. 1, but which would be fruitful to examine in future studies in the context of the halo model, is the halo mass and radial dependence of the various auto- and cross-spectra. For example, while the cosmic shear and tSZ effect power spectra may probe similar physical scales and redshifts, they are known to depend quite differently on halo mass and this is likely to be important in the context of potential baryon/feedback effects. When describing the impact of baryons on the various power spectra below, we will provide a qualitative link to the role of halo mass, leaving a quantitative exploration for future work.

3.2 Cosmic shear power spectrum

We begin by comparing the FLAMINGO simulations to measurements of the cosmic shear power spectrum from the KiDS 1000 survey[6] (Kuijken et al. 2019; Heymans et al. 2021) from the recent study of Tröster et al. (2022). The background source galaxies are divided into five tomographic bins based on their photometric redshifts, which are derived from nine-band imaging data spanning optical to infrared wavelengths (Hildebrandt et al. 2021). The tomographic bins, labelled 1 through 5, have selection windows of $0.1 < z_{\rm B} < 0.3, 0.3 < z_{\rm B} < 0.5, 0.5 < z_{\rm B} < 0.7, 0.7 < z_{\rm B} < 0.9$, and $0.9 < z_{\rm B} < 1.2$, respectively, where $z_{\rm B}$ corresponds to the maximum in the redshift posterior probability distribution for individual galaxies. The source redshift distributions, $n_i(z)$, for the five tomographic bins are shown in fig. 1 of Tröster et al. (2022). We use these distributions when computing the shear window function [equation (2)] in order to project the simulation 3D power spectra (or 2D maps) onto shear power spectra for comparison to the KiDS 1000 measurements. Tröster et al. (2022) measure the auto- and cross-power spectra between the five tomographic bins. Following their analysis, we adopt angular scale cuts of $100 < \ell < 1500$, corresponding to the range over which the KiDS cosmic shear methodology has been validated. Tröster et al. (2022) have made the measurements, covariance matrices, source redshift distributions, and their analysis software publicly available.[7] These authors have also measured the shear–tSZ effect cross-spectrum, which we compare to in Section 3.5 below.

In Fig. 2, we examine the box size and resolution dependence of the predicted cosmic shear power spectra. The numbers in the top-right corner of each panel indicate the tomographic bins being used (e.g. 3–2 indicates a cross-spectrum between the third and second tomographic bins, which is equivalent to 2–3). The open circles correspond to the KiDS 1000 measurements of Tröster et al. (2022), and the error bars shown correspond to the diagonal components only of the covariance matrix. The solid coloured curves correspond to the predicted spectra for the FLAMINGO calibrated hydro simulations with the fiducial D3A cosmology while the box size and resolution are varied. Following the same colour coding, the numbers on the left of each panel indicate χ^2 for the L1_m9 run and the $\Delta\chi^2$ (in

[6] https://kids.strw.leidenuniv.nl/
[7] https://github.com/tilmantroester/KiDS-1000xtSZ

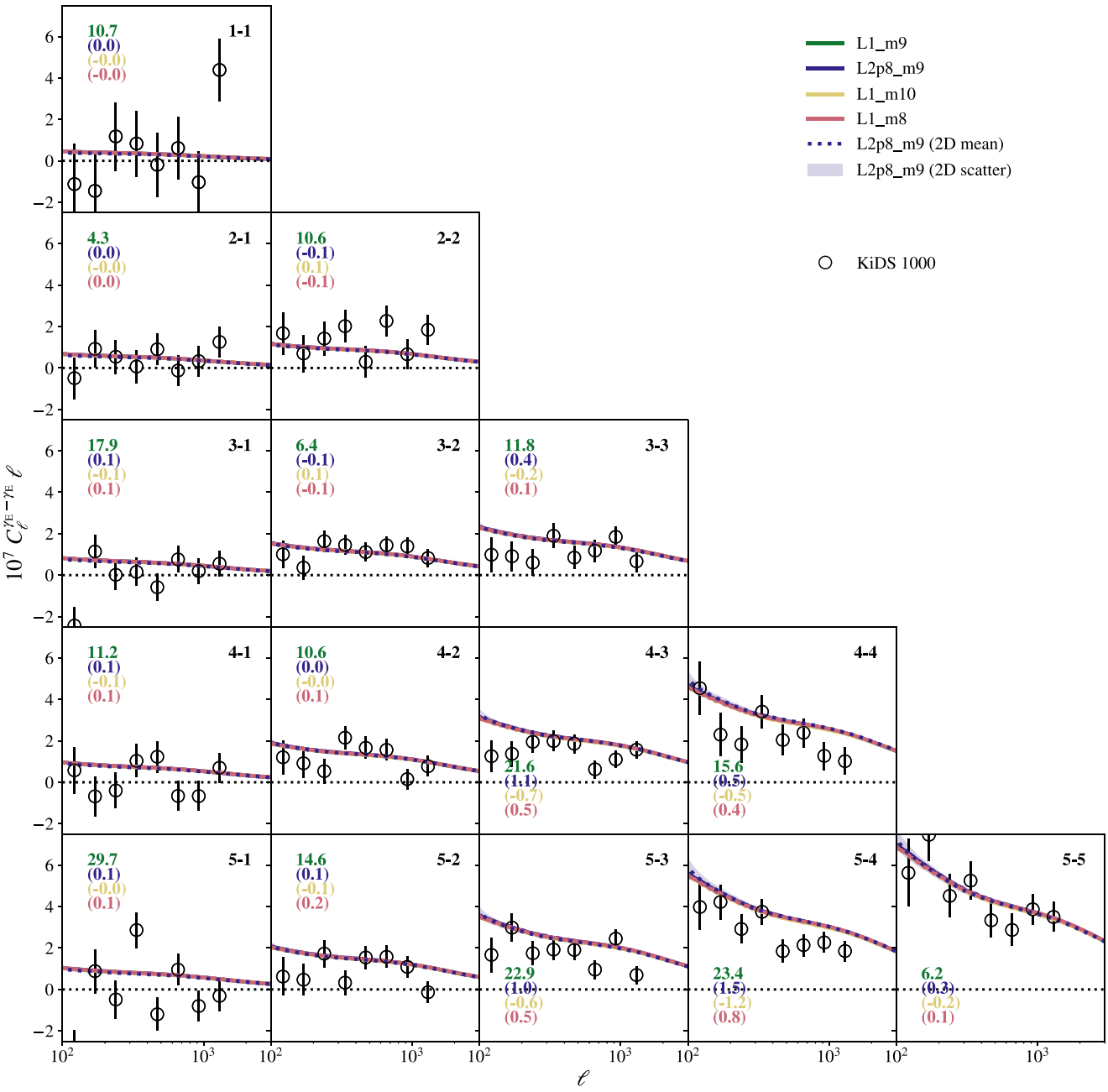

Figure 2. Dependence of the predicted KiDS 1000 cosmic shear power spectrum on simulation box size and resolution. The numbers in the top right corner of each panel indicate the tomographic bins being used (e.g. 3–2 indicates a cross-spectrum between the third and second tomographic bins, which is equivalent to 2–3). The open circles correspond to the KiDS 1000 measurements of Tröster et al. (2022) and the error bars correspond to the diagonal components of the covariance matrix. The solid coloured curves correspond to the predicted spectra for the FLAMINGO simulations with the fiducial D3A cosmology as the box size and resolution are varied. Following the same colour coding, the numbers on the left of each panel indicate χ^2 for the L1_m9 run and the $\Delta\chi^2$ (in parentheses) of the other runs with respect to L1_m9. Note that the χ^2's are computed here using the diagonal elements of the covariance matrix and we compute them for the fiducial 1D Limber method only (solid curves). A negative value for the $\Delta\chi^2$ indicates a better match to the data. The blue dashed curve and shaded region correspond to the mean power spectra and scatter from the 2D map-based analysis of the 8 light cones for the L2p8_m9 run, which has the same resolution, cosmology, and calibrated feedback model as L1_m9. The simulation predictions are converged with box size and resolution, but tend to predict spectra with slightly higher amplitude than observed (i.e. the S_8 tension), particularly for the spectra involving the higher tomographic bins.

parentheses) of the other runs with respect to L1_m9. Note that the χ^2's are computed here adopting the diagonal elements of the covariance matrix and we compute them for the fiducial 1D Limber method only (solid curves). The blue dashed curve and shaded region correspond to the mean power spectra and scatter from the 2D map-based analysis of the 8 light cones for the L2p8_m9 run, which has the same resolution, cosmology, and calibrated feedback model as L1_m9. Generally speaking, the predicted signal is strongest for spectra involving the higher tomographic bin numbers, simply as a result of there being a longer path length along the line

of sight from the observer to the galaxy samples and thus more lensing.

From Fig. 2, we conclude that the predictions of the simulations are robust to variations in box size and resolution, as well as analysis method (1D Limber versus 2D map-based). There is perhaps a hint that the larger 2.8 Gpc volume has slightly more power on the largest scales ($\ell \lesssim 500$) for the spectra involving the most distant tomographic bins. However, these differences are generally small compared to variations of the simulations with respect to the KiDS 1000 measurements, variations between neighbouring data points, and the estimated uncertanties.

Also evident from Fig. 2 is the tendency of the simulations to predict power spectra that are elevated with respect to the KiDS 1000 measurements, particularly for the higher tomographic bins from which much of the signal originates. This is confirmation of the well-known S_8 tension, noting again that the D3A cosmology includes Planck CMB constraints which pull the preferred value of S_8 to higher values than favoured by cosmic shear alone. However, we highlight here that our conclusions are only qualitative, as we have not taken into account the uncertainties in the D3A cosmology, nor have we marginalized over relevant systematic uncertainties in the lensing measurements (e.g. intrinsic alignments, photo-z uncertainties, etc.).

In Fig. 3, we examine the FLAMINGO cosmological variations in the context of the fiducial calibrated hydro model. Specifically, in 1 Gpc volumes, we compare the fiducial D3A cosmology with the maximum-likelihood Planck 2018 cosmology ('Planck'), two Planck-based cosmologies where the summed mass of neutrinos is raised from the minimum value of 0.06–0.24 eV (see Table 2), and the 'lensing' cosmology (LS8) of Amon et al. (2023), which uses CMB data to inform the mass densities of baryons and CDM as well as the primordial power spectrum shape (n_s), but uses cosmic shear to set the amplitude (S_8, or A_s).

The dependence on cosmology is particularly evident in the spectra involving the higher tomographic bins, with the Planck cosmology yielding spectra that are elevated with respect to the fiducial D3A cosmology (and thus in slightly stronger tension with the KiDS 1000 measurements), whereas increasing the summed neutrino mass suppresses the power over all scales sampled here. Note that increasing the neutrino mass and lowering S_8 have similar effects on the cosmic shear power spectra. This is a consequence of our approach of fixing A_s (i.e. using the CMB to specify it) when increasing the summed neutrino mass, as the main effect of increasing the neutrino mass in this case is to suppress the clustering amplitude at late times (i.e. lower S_8). The LS8 'lensing' cosmology, with $S_8 = 0.766$, predicts the lowest power and is in best agreement with the measurements, which is essentially by construction. Our findings appear consistent with those of Tröster et al. (2022), who obtain $S_8 \approx 0.75 \pm 0.02$ using the HMx model of Mead et al. (2020) to model the cosmic shear power spectrum and cosmic shear–tSZ cross-spectrum. Note that HMx includes a model for marginalizing over the impact of baryons which was calibrated on our previous BAHAMAS simulations.

We now turn to the impact of variations in baryon physics on the cosmic shear power spectrum. In Fig. 4 we show how varying the gas fractions of groups and clusters (as mediated through variations in stellar and primarily AGN feedback) affects cosmic shear. Even though the feedback variations span a wide range ($+2\sigma$ to -8σ about the observed gas mass fraction–halo mass relation; see Kugel et al. 2023; Schaye et al. 2023), they result in only relatively minor effects on the power spectra, which are most evident at $\ell \gtrsim 700$. In particular, reducing the gas fractions slightly improves the match to the measurements of the high tomographic bin data. Note, however, that the improvement is generally small compared to that which is obtained from variations in the baseline cosmology. Indeed, the slight preference for increased feedback with respect to the fiducial calibrated model is most likely driven by the fact that there is an offset with respect to the baseline D3A cosmology generally (that is, increased feedback is partially compensating for a difference in cosmology). We can test this by assuming the impact of cosmology variations is separable from the impact of baryon variations (Van Daalen et al. 2011; Mummery et al. 2017; Van Daalen, McCarthy & Schaye 2020). Specifically, we compute the 'suppression function' of the fgas-8σ run with respect to the fiducial hydro model (both in the fiducial D3A cosmology) by simply taking the ratio of their power spectra (e.g. Semboloni et al. 2011). Assuming this ratio is independent of cosmology, we multiply it with the LS8 power spectrum (which uses the fiducial calibrated hydro model). This procedure approximates the impact of running the LS8 cosmology but with stronger feedback. Comparing the unmodified LS8 and its enhanced feedback variant to the observational measurements, we find that the preference for stronger feedback largely goes away. For example, in the 5–3 and 5–4 cross-spectra cases, which have the strongest preference for increased feedback in the fiducial D3A cosmology, the $\Delta\chi^2$'s between the LS8 model and its increased feedback (fgas-8σ) variant are only -0.9 and -1.5, compared to -2.6 and -4.2 in the D3A cosmology in Fig. 4. Thus, we confirm that cosmology and baryon feedback can be degenerate when fitting to the observational measurements, which underscores the importance of external data sets (e.g. group baryon fractions) in constraining baryonic feedback.

We examine other baryonic feedback scenarios in Fig. A1 in Appendix A, namely the impact of variations in the SMF (both at fiducial and reduced gas fractions) and the calibrated and stronger jet model of AGN feedback. Similar to what was concluded above, we find that these variations in the feedback models have generally a minor impact on the cosmic shear power spectrum. Variations in the SMF at fixed gas fraction have a negligible impact on all scales, which is not unexpected as the baryon fractions of groups and clusters are generally dominated by the hot gas. The jet models yield similar results to the fiducial thermal AGN model at fixed gas fraction.

The DES Y3 release provides an independent data set with similar statistical precision to the KiDS 1000 survey against which we can compare the simulations to test for consistency (or lack thereof). In Appendix A, we compare the FLAMINGO simulations with the DES Y3 harmonic space measurements of Doux et al. (2022). As shown in Fig. A2, the lensing LS8 cosmology yields a somewhat better fit to the data relative to the fiducial D3A cosmology (particularly amongst the majority of the higher tomographic bins), whereas a Planck CMB cosmology yields a worse fit for virtually all bins. Also consistent with the KiDS 1000 comparison above, increasing the efficiency of feedback slightly improves the fit to the DES measurements, but is less significant than the improvement that results from lowering of S_8. We conclude that while there is evidence for (mild) tension of the DES Y3 measurements with the Planck CMB cosmology, it is of slightly lower significance than for the KiDS 1000 survey. This is consistent with findings of Doux et al. (2022), who infer $S_8 = 0.784 \pm 0.026$ from the DES Y3 power spectra, representing a 1.5σ tension with Planck. We note that KiDS 1000 and DES Y3 shear results are consistent with each other to within 1σ (e.g. Abbott et al. 2023).

For completeness, in Table 3, we present the χ^2 values of the predicted cosmic shear power spectra with respect to the observational KiDS 1000 and DES Y3 measurements. The values are summed over the 15 and 10 auto- and cross-spectra from KiDS and

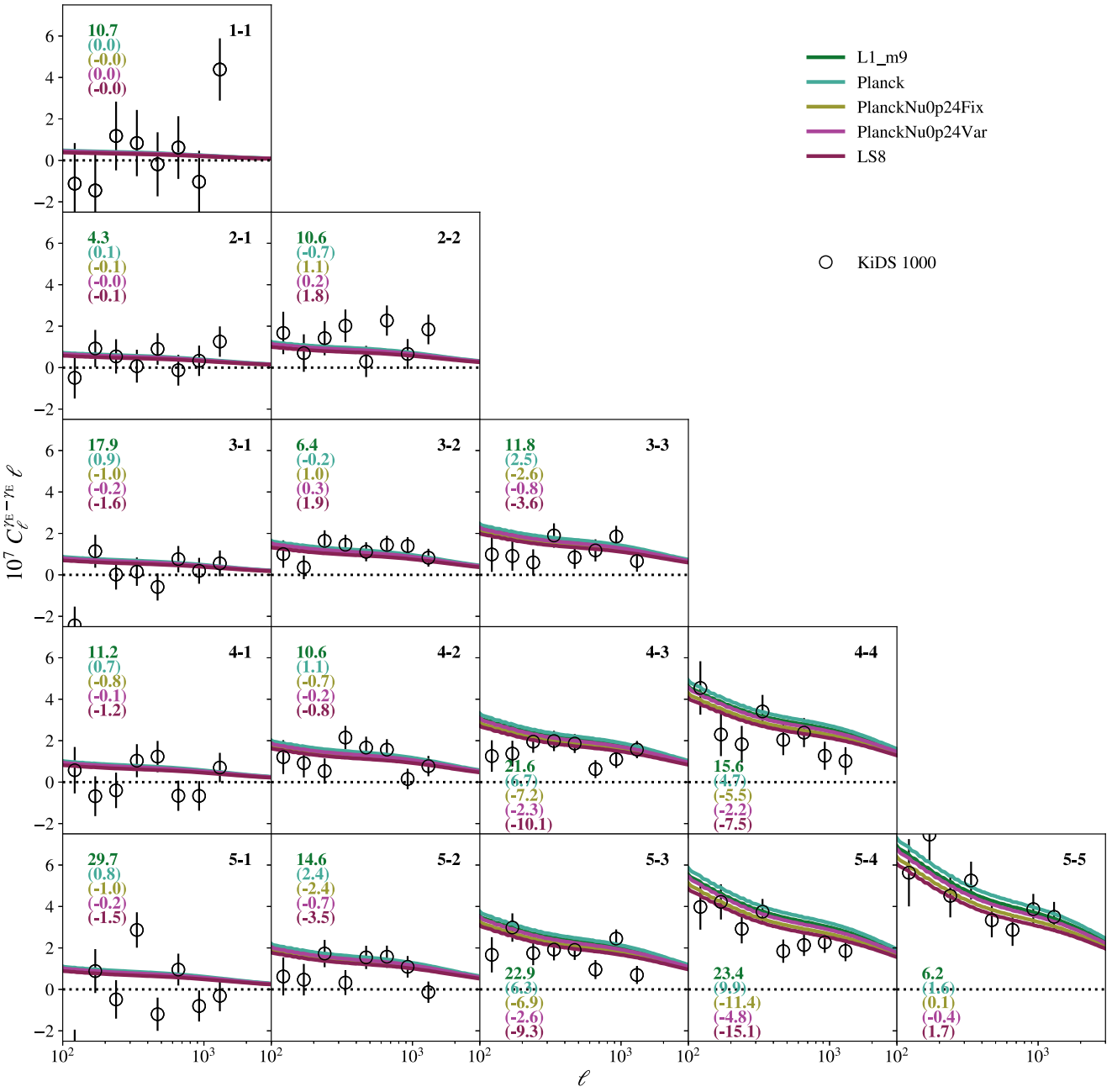

Figure 3. As Fig. 2, but showing the dependence of the predicted KiDS 1000 cosmic shear power spectrum on cosmology. The solid coloured curves correspond to the predicted spectra for the FLAMINGO simulations as the background cosmology is varied. A Planck maximum-likelihood cosmology yields spectra with elevated power relative to the fiducial D3A (and is thus in stronger tension with the lensing measurements), whereas increasing the summed neutrino mass from 0.06 eV (fiducial) to 0.24 eV suppresses the power over all scales sampled here. The LS8 'lensing' cosmology predicts the lowest power and is in best agreement with the measurements, particularly for the high tomographic bin data (except 5–5).

DES, respectively, and are shown for both the cases where only the diagonal errors are adopted and when the full covariance matrix is used. The results confirm those shown in the figures, although the differentiability of the models is reduced when taking into account the full covariance of the measurements (as expected).

3.3 tSZ effect power spectrum

We now turn our attention to the tSZ effect power spectrum. The tSZ effect is induced from the inverse Compton scattering of CMB photons by hot, free electrons in the intracluster medium (ICM) of galaxy groups and clusters (Sunyaev & Zeldovich 1972; Birkinshaw 1999). The effect appears as a decrement in CMB temperature maps at radio wavelengths and an increment at millimetre scales (for a review see Carlstrom, Holder & Reese 2002). As its amplitude is proportional to both the ICM electron number density and temperature, it is particularly strong for massive galaxy clusters. Indeed, the self-similar expectation is that the integrated tSZ flux scales with halo mass to the 5/3 power (e.g. White, Hernquist & Springel 2002; Planck Collaboration XI 2013) and previous studies

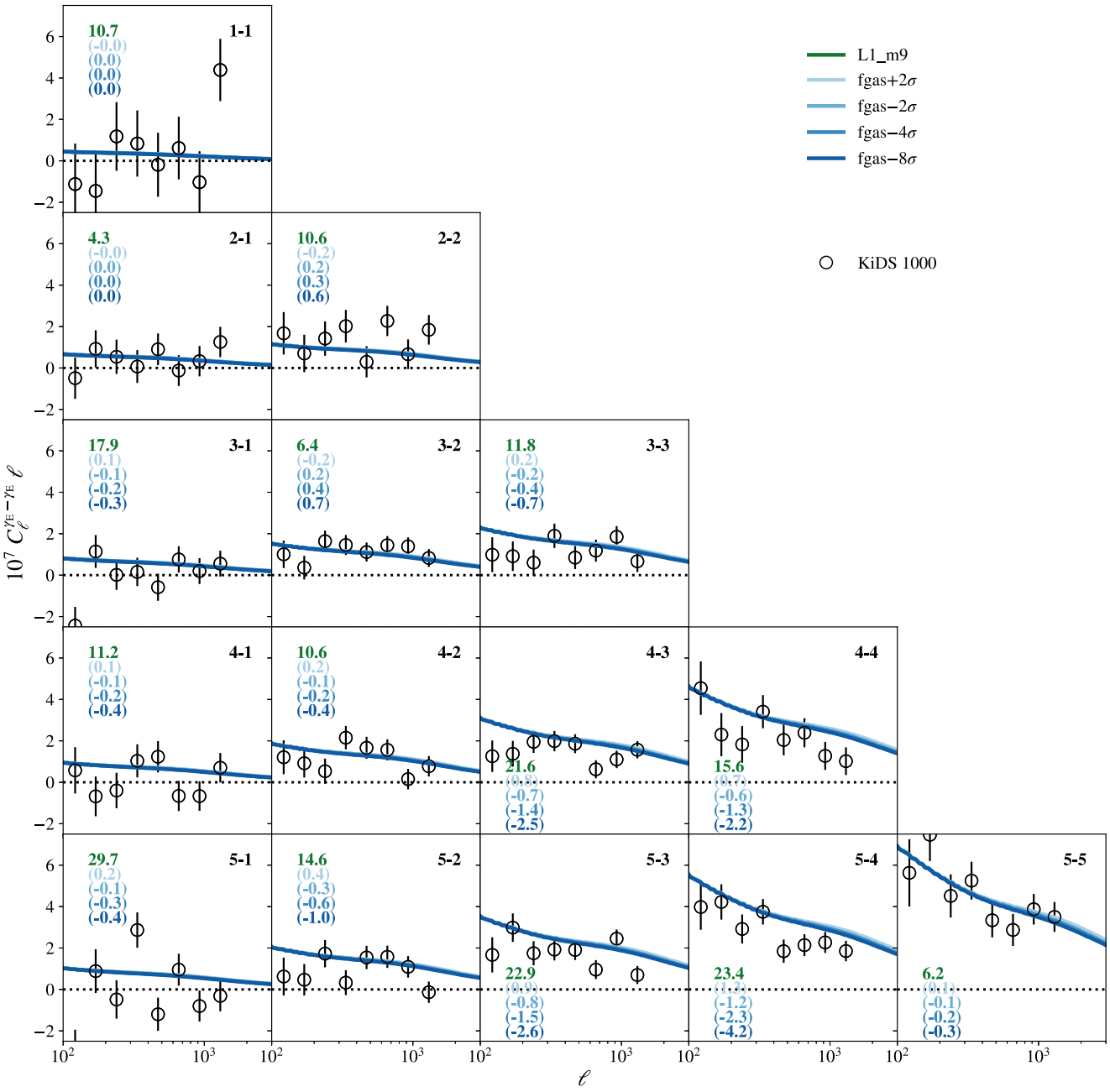

Figure 4. As Fig. 2, but showing the dependence of the predicted KiDS 1000 cosmic shear power spectrum on baryon physics, namely variations in the gas fractions of groups and clusters that are mediated primarily through variations in the AGN feedback strength. The solid coloured curves correspond to the predicted spectra for the FLAMINGO simulations as the gas fractions are varied from $+2\sigma$ to -8σ with respect to the observed gas fraction–halo mass relation (see Kugel et al. 2023; Schaye et al. 2023). Lowering the gas fractions (increasing the feedback strength) relative to the fiducial calibrated model results in a slightly improved match to the data, though the improvement is generally small compared to that from the investigated changes in the baseline cosmology (Fig. 3).

have shown that massive clusters tend to dominate the tSZ effect power spectrum (e.g. Komatsu & Seljak 2002; Battaglia et al. 2012; McCarthy et al. 2014). As the abundance of massive clusters is a sensitive probe of Ω_m and particularly σ_8, the tSZ effect, which is proportional to the square of the tSZ flux, is even more sensitive to these cosmological parameters (e.g. Komatsu & Seljak 2002; Shaw et al. 2010; Millea et al. 2012). For example, the amplitude of the tSZ effect power spectrum scales approximately as $\sigma_8^{8.3}$ (Shaw et al. 2010).

We compare the FLAMINGO simulations to the latest tSZ effect power spectrum measurements, namely the Planck-based measurements reported in Bolliet et al. (2018) and the SPT data in Reichardt et al. (2021). Note that Bolliet et al. (2018) present an improved re-analysis of the Planck 2015 tSZ data set from Planck Collaboration XXII (2016) by taking into account the tri-spectrum in the covariance matrix and placing physical constraints on the amplitudes of foreground contaminants (particularly radio and infrared point sources and the clustered infrared background, or

Table 3. χ^2 values of the predicted KiDS 1000 and DES Y3 cosmic shear power spectra with respect to the measurements of Tröster et al. (2022) and Doux et al. (2022), respectively. There are 120 (270) independent data points summed over the 15 (10) KiDS 1000 (DES Y3) auto- and cross-spectra. Values in parentheses indicate the $\Delta\chi^2$ with respect to the L1_m9 run. χ^2 values using either the diagonal elements alone or the full covariance matrix are provided.

Prefix	KiDS 1000		DES Y3	
	$\chi^2_{\rm diag}$	$\chi^2_{\rm covar}$	$\chi^2_{\rm diag}$	$\chi^2_{\rm covar}$
L1_m9	217.5	183.2	355.2	324.3
L2p8_m9	(5.1)	(2.2)	(11.3)	(3.5)
L1_m10	(−3.7)	(−1.7)	(−6.5)	(−2.8)
L1_m8	(2.8)	(1.4)	(3.9)	(1.2)
fgas+2σ	(4.7)	(2.4)	(5.6)	(1.3)
fgas−2σ	(−4.0)	(−2.0)	(−4.5)	(−1.0)
fgas−4σ	(−7.8)	(−3.8)	(−8.4)	(−1.7)
fgas−8σ	(−13.7)	(−6.7)	(−14.0)	(−2.6)
M*−σ	(−2.6)	(−1.3)	(−2.8)	(−0.5)
M*−σ_fgas−4σ	(−9.3)	(−4.6)	(−9.6)	(−1.7)
Jet	(0.8)	(0.6)	(−0.5)	(−0.7)
Jet_fgas−4σ	(−14.8)	(−7.2)	(−16.9)	(−4.2)
Planck	(36.9)	(16.1)	(76.4)	(24.2)
PlanckNu0p24Fix	(−37.3)	(−16.3)	(−43.9)	(−13.1)
PlanckNu0p24Var	(−14.2)	(−6.3)	(−21.7)	(−7.5)
LS8	(−48.8)	(−21.2)	(−34.4)	(−8.5)

CIB). We use the tabulated tSZ power spectrum measurements and total diagonal uncertainties from table 4 of Bolliet et al. (2018). Note that total uncertainties include the non-Gaussian contribution from the tri-spectrum that dominates on large scales (Komatsu & Seljak 2002). The full covariance matrix was not tabulated, though as shown in fig. 3 of Bolliet et al. (2018), the diagonal uncertainties tend to dominate.

In the top-left panel of Fig. 5, we examine the dependence of the tSZ effect power spectrum on simulation box size and resolution. The solid coloured curves correspond to the predicted spectra for the FLAMINGO simulations with the fiducial D3A cosmology as the box size and resolution are varied. We see that the simulation predictions are largely converged with resolution at fixed box size (1 Gpc), but the 1 Gpc boxes are missing a small amount of power compared to the 2.8 Gpc volume on all scales. We attribute this offset, which is larger than for the case of cosmic shear, as being due to the disproportionate influence of very massive, rare clusters on this statistic. Consistent with this, interpretation is the relatively large degree of cosmic variance in the 2D map-based analysis (shaded blue region), which shows significant variation between the eight lightcones, particularly for $\ell \lesssim 300$. We note, however, that the mean power spectrum from the 2D method (dashed blue curve) is generally in very good agreement with the fiducial 1D Limber calculation.

Comparing the simulations to the observational measurements, it is immediately obvious from the top-left panel of Fig. 5 that the simulations predict far too much power on all scales relative to the Planck and SPT measurements. However, as already noted, the tSZ effect power spectrum is very sensitive to the adopted cosmology. In the top-right panel of Fig. 5, we therefore explore the cosmological variations in FLAMINGO. A Planck maximum-likelihood cosmology yields spectra with slightly elevated power relative to the fiducial D3A cosmology (and is thus in stronger tension with the measurements), whereas increasing the summed neutrino mass from 0.06 eV (fiducial) to 0.24 eV suppresses the power over all scales sampled here, though not sufficiently to bring the simulations into agreement with the data. The LS8 'lensing' cosmology predicts the lowest power and is in best agreement with the tSZ power spectrum measurements, although it still predicts slightly too much power relative to the Planck measurements at intermediate scales and particularly with respect to the SPT measurements on small angular scales.

In the bottom panels of Fig. 5, we examine the feedback dependence of the tSZ effect. The bottom-left panel explores variations in the gas mass fractions of groups and clusters (again mediated primarily through AGN feedback variations), while the bottom-right panel explores variations in the SMF (at the fiducial and reduced gas fractions) and variations in the nature of the AGN feedback implementation (thermal versus jet). We conclude from these comparisons that the tSZ power spectrum is generally insensitive to even large variations in the baryon physics on the large angular scales probed by Planck, which is consistent with the findings of previous studies that used cosmological hydro simulations (e.g. Battaglia et al. 2012; McCarthy et al. 2014). Note that while the feedback variations we have explored are generally unable to liberate baryons from the very massive haloes and large physical scales that dominate the tSZ effect, baryon physics can still in principle alter the power spectrum on large scales through variations in the efficiency of star formation, as star formation siphons off the hot gas reservoir that gives rise to the tSZ effect (see da Silva et al. 2001 for a dramatic example). However, current observational measurements of the baryon fractions of massive clusters indicate that they have approximately their full cosmological complement of baryons (with $f_b \approx \Omega_b/\Omega_m$) and that the hot gas dominates (e.g. Gonzalez et al. 2013; Akino et al. 2022), implying that feedback has strongly curtailed star formation in these systems and their progenitors. Such behaviour is also effectively enforced in the FLAMINGO simulations through calibration to the observed $z = 0$ galaxy SMF. We further highlight that the fiducial FLAMINGO simulation predicts a tSZ effect–halo mass scaling relation that is in excellent agreement with Planck tSZ cluster measurements (see fig. 15 of Schaye et al. 2023).

The situation changes on the smaller scales probed by SPT (a few arcminutes), which are more sensitive to group-mass haloes. Here, the variations in feedback can give rise to relatively large effects on the tSZ power, although none of the variations we have explored here can reproduce the low amplitude of the SPT measurements for the fiducial cosmology. It is possible that some combination of cosmological and feedback modifications (e.g. similar to the 'lensing' cosmology but with stronger-than-fiducial feedback, or a lower amplitude cosmology with fiducial feedback) could reconcile these measurements, but we leave that as an open question for future work.

In summary, similar to the cosmic shear comparison in Section 3.2, but with higher significance, the tSZ effect power spectrum prefers a low S_8 cosmology compared to the D3A and Planck cosmologies. This conclusion is qualitatively consistent with some previous studies of the tSZ effect that also used cosmological hydrodynamical simulations (e.g. McCarthy et al. 2014, 2018). It is also consistent with the halo model-based analyses[8] of Planck Collaboration XXI (2014),

[8]We note that the tSZ power spectrum in Planck Collaboration XXII (2016) is of somewhat higher amplitude than in the re-analysis by Bolliet et al. (2018), which is likely the reason why the inferred value of S_8 from the former study is slightly larger than one would anticipate based on the comparison in Fig. 5. In addition, the uncertainties are larger in the latter study due to the inclusion of the tri-spectrum in the covariance matrix. See Bolliet et al. (2018) for further discussion.

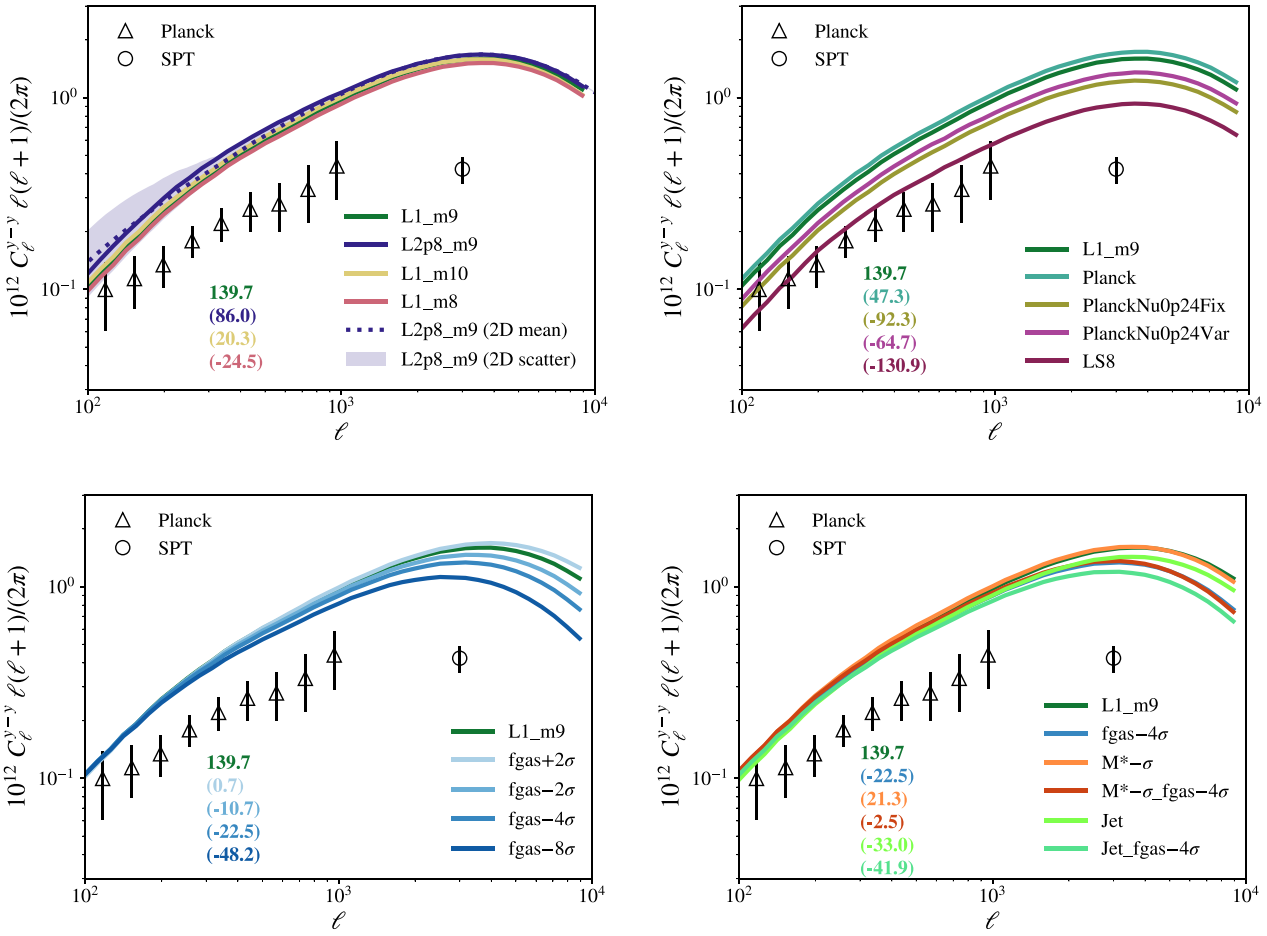

Figure 5. The tSZ effect angular power spectrum. The open triangles correspond to the Planck tSZ measurements of Bolliet et al. (2018) while the open circle corresponds to the SPT measurements of Reichardt et al. (2021). Top left: Dependence on simulation box size and resolution. Note that the χ^2 values displayed in the bottom left are computed with respect to the Planck tSZ measurements are take into account only the diagonal elements of the covariance matrix. The simulation predictions are largely converged with resolution, though the 1 Gpc boxes are missing a small amount of power compared to the 2.8 Gpc volume. Top right: Dependence on cosmology. A Planck maximum-likelihood cosmology yields spectra with elevated power relative to the fiducial D3A cosmology (and is thus in stronger tension with the measurements), whereas increasing the summed neutrino mass from 0.06 eV (fiducial) to 0.24 eV suppresses the power over all scales sampled here. The LS8 'lensing' cosmology predicts the lowest power and is in best agreement with the tSZ power spectrum measurements. Bottom left: Dependence of the predicted tSZ power spectrum on baryon physics, namely variations in the gas fractions of groups and clusters which are mediated primarily through variations in the AGN feedback strength. Feedback effects are most evident at small angular scales ($\ell \gtrsim 2000$) and cannot reconcile the offset from the large-scale Planck measurements. Bottom right: Dependence on other baryon variations, including variations in the stellar mass function (both at the fiducial and reduced gas fractions) and the calibrated and strong jet models of AGN feedback. The effects of variations in the stellar mass function are generally negligible compared to those of variations in the gas fractions.

Planck Collaboration XXII (2016), and Bolliet et al. (2018) who infer $S_8 \approx 0.78 \pm 0.02$ and $S_8 \approx 0.75 \pm 0.04$, respectively, when adopting a hydrostatic mass bias consistent with weak lensing observations and the predictions of simulations. (By contrast, Reichardt et al. 2021 require $S_8 \approx 0.69 \pm 0.03$ to match their SPT measurements when using the halo model of Shaw et al. 2010, although as already noted these scales can be significantly affected by feedback.) Note, however, that in the context of the halo model one can boost the best-fitting S_8 by appealing to a larger halo mass bias, but only at the expense of agreement with simulation predictions and observational weak lensing mass constraints. One advantage of the comparison in the present study is that we go directly from either 3D power spectra or 2D maps to a tSZ power spectrum prediction without the intermediate step of defining and counting haloes and choosing a mass bias. The comparison here is therefore more direct.

While the discussion of the impact of baryon feedback on current cosmic shear power spectrum constraints remains an open discussion due to the degeneracy between cosmology and baryon feedback (e.g. Amon & Efstathiou 2022; Schneider et al. 2022; Tröster et al. 2022; Aricò et al. 2023; Chen et al. 2023), it is much more difficult to appeal to baryons as a solution to the tSZ effect power offset on large scales, owing to the fact that this statistic is dominated by very massive clusters which are observed to be 'baryonically closed' and dominated by hot gas.

3.4 CMB lensing power spectrum

The final auto power spectrum that we consider is the CMB lensing power spectrum. As already noted in Section 2.2, CMB lensing receives contributions from matter fluctuations over a wide range of redshifts, with a C_ℓ-weighted mean redshift of $z_{\rm eff} \approx 3 - 4$ (see Fig. 1). For a review of CMB lensing, see Lewis & Challinor (2006). Current measurements of this statistic, which are essentially derived from four-point measurements of the CMB temperature and

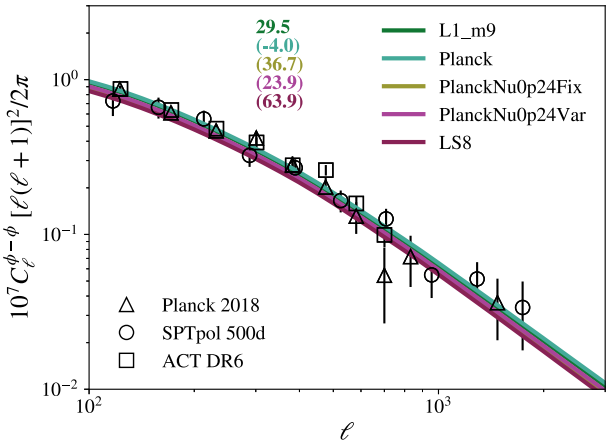

Figure 6. Dependence of the CMB lensing angular power spectrum on cosmology. The open triangles, circles, and squares correspond to the Planck 2018, SPTpol 500d, and ACT DR6 measurements of Planck Collaboration VIII (2020b), Wu et al. (2019), and Qu et al. (2023), respectively. The error bars correspond to the diagonal components of the covariance matrices. Note that the χ^2 values displayed in the top left are computed by summing over the three data sets and take into account only the diagonal errors. A Planck maximum-likelihood cosmology yields spectra with slightly elevated power relative to the fiducial D3A (with a similarly good match to the measurements), whereas increasing the summed neutrino mass from 0.06 eV (fiducial) to 0.24 eV suppresses the power and worsens the agreement with the data. The LS8 'lensing' cosmology predicts the lowest power, but in contrast with galaxy lensing and the tSZ effect tests, this model yields the worst agreement with the CMB lensing measurements.

polarization maps (Hu & Okamoto 2002), are generally restricted to relatively large angular scales of $\ell \lesssim 2000$. Given that much of the signal arises from high redshifts, existing measurements typically probe linear scales, which considerably simplifies the modelling the CMB lensing power spectrum. Indeed, in Appendix A, we confirm that this statistic is insensitive to variations in box size, resolution, and baryon physics. This situation will change in the near future, however, as forthcoming CMB lensing experiments such as the Simons Observatory (Ade et al. 2019) and CMB-S4 (Abazajian et al. 2019) will begin to probe scales that are more sensitive to non-linear evolution and baryonic physics (McCarthy, Hill & Madhavacheril 2022; Upadhye et al. 2023). Here, we focus on the cosmological dependence of the CMB lensing power spectrum.

In Fig. 6, we explore the cosmology variations in FLAMINGO and compare with the latest measurements from Planck 2018 (Planck Collaboration VIII 2020b), SPTpol 500d (Wu et al. 2019), and ACT DR6 (Qu et al. 2023). The quoted χ^2 values sum over the three data sets and are calculated with respect to the diagonal elements only of the respective covariance matrices. While the full covariance matrices are available, combining them in a rigorous way is non-trivial due to the spatial overlap of the surveys. Furthermore, as the CMB lensing power spectrum probes mainly linear scales, the off-diagonal uncertainties are expected to be small. Indeed, Qu et al. (2023) find the off-diagonal correlations to typically be less than 10 per cent of the total uncertainty.

In agreement with the above studies, we find that a Planck-like cosmology yields an excellent match to the CMB lensing measurements. The Planck maximum-likelihood cosmology yields spectra with slightly elevated power relative to the fiducial D3A, but with a similarly good match to the measurements. Increasing the summed neutrino mass from 0.06 eV (fiducial) to 0.24 eV suppresses the power and worsens the agreement with the observed power spectrum. This is an important result, since one of the proposed ways of reconciling the primary CMB fluctuations with the apparent suppression of the growth of LSS is to appeal to the fact that neutrinos do not cluster significantly on small scales. Some previous studies have suggested values of $\sum m_\nu \approx 0.2$–0.4 eV (e.g. Battye & Moss 2014; Beutler et al. 2014; Wyman et al. 2014; McCarthy et al. 2018) could reconcile most of the tension. However, here we see that raising the summed mass of neutrinos worsens the agreement with the observed CMB lensing power spectrum, by suppressing the predicted amplitude below what is measured. Thus, CMB lensing observations play a critical role in challenging massive neutrinos as a solution to the S_8 tension. Lastly, the LS8 'lensing' cosmology predicts the lowest power and in contrast with cosmic shear and the tSZ effect tests discussed above, this model yields the worst agreement with the CMB lensing measurements.

In the most recent analysis using data from ACT DR6, Madhavacheril et al. (2023) find that the combination of ACT lensing + Planck lensing + BAO yields a tight constraint of $S_8 = 0.83 \pm 0.02$, in excellent agreement with the Planck CMB cosmology and the fiducial D3A cosmology. Note that CMB lensing alone actually best constrains the parameter combination $\sigma_8\Omega_m^{0.25}$ (rather than S_8) and a fit to the ACT DR6 measurements yields 0.61 ± 0.02 for this quantity (Qu et al. 2023). For context, our fiducial D3A, Planck, and LS8 cosmologies have $\sigma_8\Omega_m^{0.25} = 0.60$, 0.61, and 0.56. Thus, CMB lensing alone agrees extremely well with the Planck CMB and D3A cosmologies and is in a moderate degree of tension with the LS8 cosmology. Note that the two 0.24 eV neutrino cosmologies yield values of 0.61 and 0.59 for the Var and Fix variants, respectively, which is consistent with the CMB lensing-only constraints of (Qu et al. 2023). However, it is clear from the χ^2 values in Fig. 6 that the neutrino cosmologies fit the measurements worse than the D3A and Planck cosmologies, which may imply that the $\sigma_8\Omega_m^{0.25}$ parameter does not capture the full cosmological dependence of the CMB lensing-only constraints.

Assuming no significant systematic errors have been neglected in the three autopower spectra comparisons above (either on the observational or theoretical sides), which would alter the conclusions drawn, then based on Fig. 1, there are two possible generic ways to reconcile the three probes explored so far: (i) a modification to the redshift evolution of matter fluctuations, such that the fluctuations grow more slowly at late times than predicted in the fiducial cosmology (noting that CMB lensing probes higher redshifts than cosmic shear or the tSZ effect; see Fig. 1, bottom panel); and/or (ii) a modification on non-linear scales, such that the fluctuations on these scales grow more slowly than expected (cosmic shear and tSZ effect probe non-linear scales, whereas CMB lensing probes linear scales; see Fig. 1, top panel). Preston, Amon & Efstathiou (2023; see also Nguyen, Huterer & Wen 2023) recently came to the same general conclusions. Here, however, we argue using the FLAMINGO hydrodynamical simulations that baryonic feedback is unlikely to be the physics driving the required modifications.

3.5 Cosmic shear–tSZ effect cross-spectrum

Having examined the auto power spectra of the three observables (shear, tSZ, and CMB lensing) in the previous sections, we now examine the cross-spectra between these variables. Importantly, the cross-spectra contain additional information about these observables. For example, current measurements of the cosmic shear power spectrum are most sensitive to the clustering (two-halo) and structure (one-halo) of group-mass haloes ($M \sim 10^{13-14}$ $M_\odot$) at a distance

roughly half way to the background source population, whereas the tSZ effect power spectrum is mostly sensitive to the structure of low-redshift, very massive haloes ($M \sim 10^{15}$ $M_\odot$; Komatsu & Seljak 2002). The cross-spectrum, however, is sensitive to both the clustering and structure of haloes with redshifts and masses intermediate between these regimes (i.e. $M \sim 10^{14-15}$ $M_\odot$), as shown in Mead et al. (2020). This is simply because the cross-spectrum picks out (only those) structures that contribute significantly to both observables.

Another important benefit of cross-spectra measurements is that they have different biases than auto power spectra. For example, the noise (e.g. shot noise, detector noise) in two independent maps, such as cosmic shear and tSZ, is expected to be uncorrelated and therefore will not contribute to the cross-spectrum between those maps, whereas if unaccounted for, noise can significantly bias both observational and simulation measurements of auto power spectra. In the case of simulations, for example, particle shot noise must be subtracted from the power spectra. Other forms of bias (e.g. in galaxy shape estimation, the separation of the tSZ from the clustered infrared background, etc.) will enter into the auto power spectrum differently than for the cross-spectrum. For example, multiplicative forms of bias will enter into the power spectrum squared, but only linearly in the cross-spectrum. Thus, a simultaneous examination of the auto- and cross-power spectra gives an important cross-check on possible biases that could be affecting both measurements.

In the context of the impact of baryons on LSS, the cosmic shear–tSZ effect cross-spectrum is also interesting for another reason. Specifically, this cross-spectrum is essentially a measurement of how the hot gas (in particular its thermal energy density) traces that of the underlying matter field. Thus, in addition to yielding another cosmological test, this statistic also provides a valuable opportunity to assess the realism of feedback models when it is examined on small scales, as feedback is known to strongly alter the hot gas properties of groups and clusters (e.g. McCarthy et al. 2010; Le Brun et al. 2014; Planelles et al. 2014; Henden et al. 2018; Oppenheimer et al. 2021).

Tröster et al. (2022) recently performed a spatial cross-correlation analysis of the KiDS 1000 tomographic data set with tSZ effect maps constructed from the Planck 2015 data set (Planck Collaboration XXII 2016) and the ACT DR4 data set (Mallaby-Kay et al. 2021). The ACT data have lower noise than the Planck y map, but the overlap between the ACT DR4 and KiDS 1000 surveys is only partial. Thus, in practice the cross-spectrum derived using the Planck data set is better constrained and we therefore focus our comparison on that data set. Note that in the comparisons presented in the previous sections, the effects of the beam and pixel window function were deconvolved from the observed power spectra, whereas Tröster et al. (2022) have not deconvolved either effect (electing instead to convolve the theory). For consistency with the previous comparisons, we deconvolve the Planck beam and the $N_{\rm side} = 2048$ pixel window functions from the measured cross-spectra of Tröster et al. (2022).

In Fig. 7, we present a comparison of the full suite of FLAMINGO simulations with the measurements of Tröster et al. (2022), processed as described above. The columns are organized by cosmic shear tomographic bin (1–5; low redshift to high redshift from left to right). The rows follow our previous comparisons (from top to bottom): (i) impact of simulation resolution, box size, and analysis method (1D versus 2D); (ii) dependence on cosmology; (iii) dependence on group/cluster gas fraction; and (iv) other baryonic variations.

In the top row of Fig. 7, we see that the predicted cross-spectra are relatively well converged with resolution at fixed box size (1 Gpc). The high-resolution simulation (L1_m8) predicts slightly less power than the fiducial (L1_m9) and low res (L1_m10) runs, owing to a slightly higher star formation efficiency in groups/clusters in the high res. simulation (this was also visible in the tSZ effect power spectrum in Fig. 5). Comparing the 1 Gpc and 2.8 Gpc fiducial resolution runs, the former are missing a small amount of power on the largest angular scales, which is most evident for the cross-spectra involving the higher tomographic bins. Similar to the tSZ effect power spectrum, we attribute this slight offset as being due to a non-negligible contribution of massive, rare haloes. There is good agreement between the fiducial 1D Limber methodology (solid curves) and the map-based 2D analysis (dashed curve), the latter of which shows that cosmic variance becomes relevant on scales of $\ell \lesssim 500$ (shaded region). It is also clear from the top row of Fig. 7 that the fiducial calibrated FLAMINGO simulations in the fiducial D3A cosmology predicts too much power compared to the observational measurements on all but the smallest scales. This is particularly evident for the highest tomographic bins.

In the second row from the top, we examine the cosmology dependence of the shear–tSZ cross. Here, we see that increasing the neutrino mass, whose main effect is to lower S_8 (or, more generally, the clustering amplitude), or directly lowering S_8 at a fixed minimal neutrino mass (LS8), lowers the amplitude of the predicted cross-spectra, yielding an improved match to the data. Consistent with the cosmic shear and tSZ effect power spectrum comparisons in Sections 3.2 and 3.3, respectively, we find the 'lensing' cosmology (LS8) yields the best agreement with the data, although it is still somewhat elevated with respect to the observations (similar to the tSZ effect power spectrum in Fig. 5). The Planck cosmology, on the other hand, yields a higher amplitude than the fiducial D3A cosmology and is therefore in slightly stronger tension with the observational measurements. Note that the shear–tSZ cross-spectrum best constrains a combination of σ_8 and Ω_m that differs from the definition of S_8. Tröster et al. (2022) find that the parameter $\Sigma_8^{0.2} \equiv \sigma_8(\Omega_m/0.3)^{0.2}$ describes the degeneracy well for the shear–tSZ cross. They determine $\Sigma_8^{0.2} \approx 0.72 \pm 0.04$. For comparison, the D3A, Planck, and LS8 cosmologies used here have $\Sigma_8^{0.2} = 0.81$, 0.82, and 0.76, respectively. Because cosmic shear and the shear–tSZ cross have different dependencies on Ω_m and σ_8, a joint analysis helps to break the degeneracy between these parameters (see fig. 6 of Tröster et al. 2022; see also Fang et al. 2023). As already mentioned in Section 3.2, jointly modelling these two probes, Tröster et al. (2022) find $S_8 \approx 0.75 \pm 0.02$, representing a $\approx 3\sigma$ tension with the Planck CMB cosmology.

In the bottom two rows of Fig. 7, we examine the feedback dependence of the shear–tSZ cross-spectrum. Variations in feedback lead to noticeable differences in the predictions on scales of $\ell \gtrsim 400$ and the spread in the predictions becomes comparable to the spread due to cosmological variations on scales of $\ell \gtrsim 2000$, making this statistic more sensitive to feedback variations than the cosmic shear power spectrum on these scales. Nevertheless, on the basis of Fig. 7, we generally conclude that the feedback variations we have explored cannot reconcile the offset between the data and the fiducial D3A cosmology, as many of the bins are on angular scales that are not significantly impacted by baryons. This is consistent with the findings of Tröster et al. (2022) who, even though the impact of baryons has been marginalized over using the halo model of Mead et al. (2020), still find a $\approx 3\sigma$ tension with the Planck CMB cosmology, similar to that derived from cosmic shear alone.

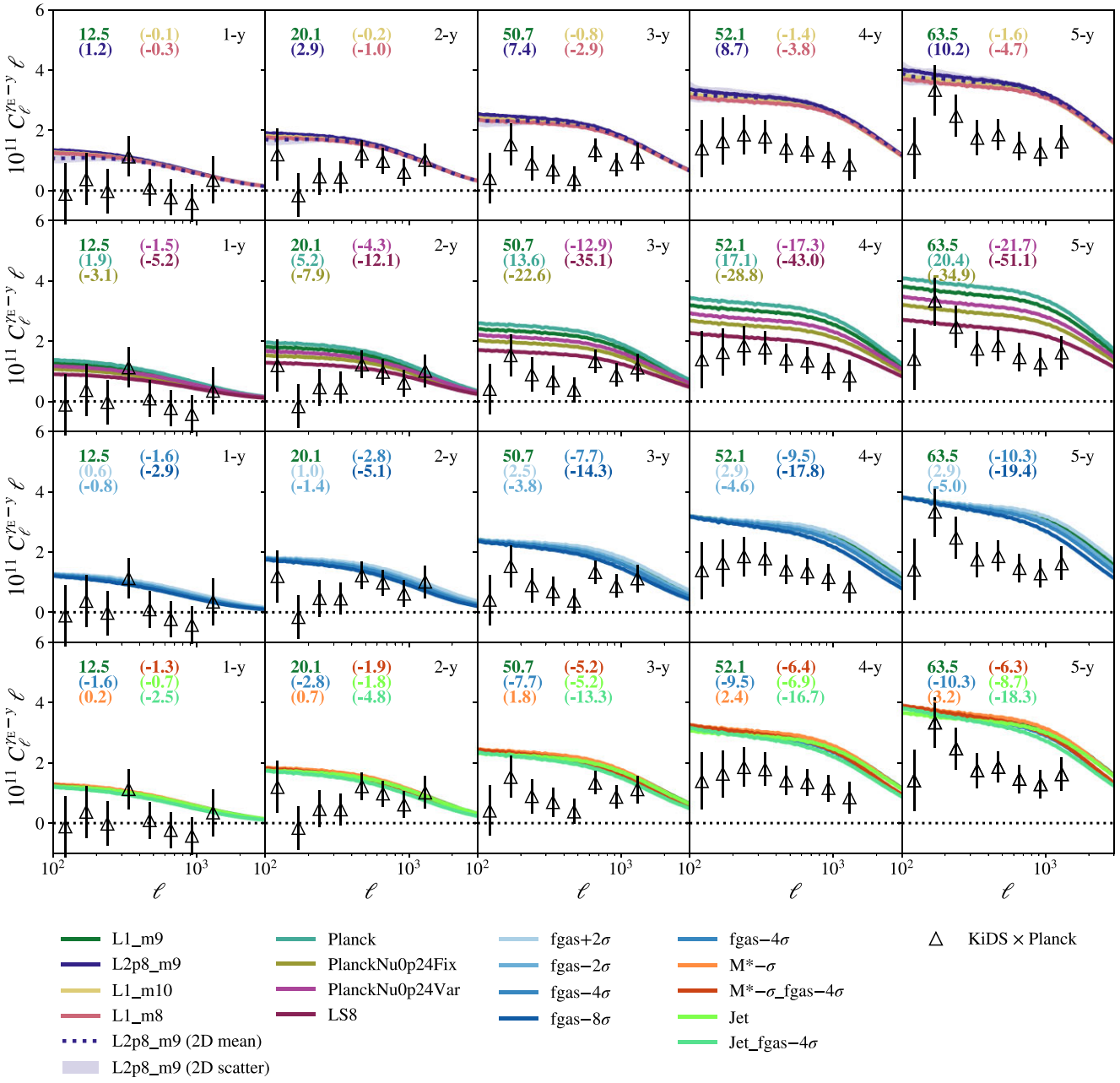

Figure 7. The cosmic shear–tSZ effect angular cross-power spectrum. The open triangles correspond to the KiDS1000 × Planck measurements of Tröster et al. (2022). The different columns correspond to the cross-spectrum between different KiDS tomographic bins (1–5) and the tSZ data. Top row: Dependence on simulation box size and resolution. The simulation predictions are largely converged with resolution, though the 1 Gpc boxes are missing a small amount of power compared to the 2.8 Gpc volume. Second row: Dependence on cosmology. A Planck maximum-likelihood cosmology yields spectra with elevated power relative to the fiducial D3A (and is thus in stronger tension with the measurements), whereas increasing the summed neutrino mass from 0.06 eV (fiducial) to 0.24 eV suppresses the power over all scales sampled here. The LS8 'lensing' cosmology predicts the lowest power and is in best agreement with the measurements, though it still predicts slightly too much power. Third row: Dependence on baryon physics, namely variations in the gas fractions of groups and clusters which are mediated primarily through variations in the AGN feedback strength. Feedback effects are significant on scales of $\ell \gtrsim 500$ but cannot reconcile the offset at larger scales. Bottom row: Dependence on other baryon variations, including variations in the stellar mass function (both at the fiducial and reduced gas fractions) and the fiducial and strong jet models of AGN feedback. The effects of variations in the stellar mass function are generally negligible compared to those of variations in the gas fractions.

In Table 4, we list the χ^2 values of the cosmic shear–tSZ cross-power spectra. The values are summed over the five cross-spectra and are listed for both the cases where only the diagonal errors are adopted and when the full covariance matrix is used. The tabulated values confirm the visual trends in Fig. 7, although the differentiability of the models is reduced when taking into account the full covariance of the measurements.

In Appendix A, we present a comparison to an independent measurement of the shear–tSZ cross-spectrum from Hojjati et al. (2017). Those authors performed a spatial cross-correlation analysis between the Planck 2015 tSZ map and cosmic shear data from the RCSLenS survey (Hildebrandt et al. 2016). While the KiDS 1000 survey covers a larger area and is significantly deeper than RCSLenS, it is nevertheless useful to check whether the conclusions

Table 4. χ^2 values of the predicted KiDS 1000 cosmic shear–Planck tSZ cross-power spectra with respect to the measurements of Tröster et al. (2022), for which there are 40 independent data points summed over the five cross-spectra. Values in parentheses indicate the $\Delta\chi^2$ with respect to the L1_m9 run. χ^2 values using either the diagonal elements alone or the full covariance matrix are provided.

Prefix	$\chi^2_{\rm diag}$	$\chi^2_{\rm covar}$
L1_m9	198.9	78.9
L2p8_m9	(30.4)	(7.7)
L1_m10	(−4.1)	(−2.0)
L1_m8	(−12.7)	(−3.0)
fgas+2σ	(9.9)	(3.9)
fgas−2σ	(−15.5)	(−5.5)
fgas−4σ	(−31.8)	(−11.0)
fgas−8σ	(−59.5)	(−19.9)
M*−σ	(8.3)	(1.7)
M*−σ_fgas−4σ	(−21.0)	(−8.3)
Jet	(−23.4)	(−6.5)
Jet_fgas−4σ	(−55.6)	(−18.8)
Planck	(58.1)	(16.2)
PlanckNu0p24Fix	(−97.4)	(−26.9)
PlanckNu0p24Var	(−57.7)	(−16.3)
LS8	(−146.5)	(−39.8)

are consistent. In short, as shown in Fig. A4, four of the five multipole bins from Hojjati et al. (2017) have amplitudes lower than predicted by the fiducial FLAMINGO model in the fiducial D3A cosmology, in general agreement with the findings above. However, these also correspond to relatively small angular scales of $\ell \sim 1000$ where the impact of feedback is non-negligible. Thus, while there is general consistency with the KiDS 1000 × Planck tSZ results above, the significance of the tension with the D3A cosmology is less compelling for the RCSLenS-based comparison owing to the uncertainties of baryonic modelling.

3.6 Cosmic shear–CMB lensing cross-spectrum

The final statistic we examine is the cross-spectrum between cosmic shear and CMB lensing. Specifically, in Fig. 8, we compare the predictions of the FLAMINGO simulations to the recent measurements of Robertson et al. (2021). Those authors performed a spatial cross-correlation analysis between the KiDS 1000 cosmic shear data and the CMB lensing maps from the Planck 2018 and ACT DR4 releases. We use their combined Planck/ACT measurements and uncertainties. The uncertainties correspond to the diagonal elements of the covariance matrix only, as the full covariance matrix is not publicly available. Note that Robertson et al. (2021) used a single large tomographic bin for KiDS data ($0.1 < z_{\rm B} < 1.2$), and we use the source redshift distribution, $n(z)$, shown in Fig. 1 of their paper. As we demonstrated in Fig. 1 of the present study, this particular cross-spectrum generally probes larger scales and higher redshifts than cosmic shear alone. We find that the measurements of Robertson et al. (2021) are sensitive to both linear scales ($k \approx 0.1$ [h Mpc^{-1}]) at $\ell \approx 100$ and non-linear scales ($k \approx 3$ [h Mpc^{-1}]) at $\ell \approx 2000$, with a mean (C_ℓ-weighted) redshift of $z_{\rm eff} \approx 0.5$.

In the top-left panel of Fig. 8, we examine the dependence of the predicted shear–CMB lensing cross-spectrum on simulation box size and resolution, as well as analysis method (1D versus 2D). We find this cross-spectrum is converged with respect to box size and resolution. Furthermore, there is excellent consistency between the fiducial 1D Limber method and the 2D map-based method for analysing the simulations. The map-based analysis demonstrates that there is considerable cosmic variance between the different light cones on angular scales of $\ell \lesssim 500$, owing to the fact that this corresponds to large physical scales that are sampled with relatively few modes even in the large 2.8 Gpc box. As highlighted previously, the simulation ICs employ mode fixing on large physics scales and therefore the fiducial 1D Limber predictions ought to predict the mean shear–CMB lensing power spectrum to high accuracy for our chosen cosmology.

As in the previous comparisons (with the clear exception of the CMB lensing power spectrum in Section 3.4), the predicted shear–CMB lensing cross-spectrum for the fiducial calibrated FLAMINGO hydro simulation in the fiducial D3A cosmology has too much power compared to the observational measurements. In the top-right panel of Fig. 8, we examine the cosmology dependence of the cross-spectrum. Consistent with the previous statistics explored above, we find that lowering the value of S_8 (whether directly, or via an increase in the summed neutrino mass) improves the match to the measurements, with the LS8 cosmology yielding the best match. Even the LS8 cosmology is somewhat elevated with respect to the measurements, however. This is consistent with the findings of Robertson et al. (2021), who used linear theory + HALOFIT (Smith et al. 2003; Takahashi et al. 2012) to predict the cross-spectrum and measured $S_8 = 0.64 \pm 0.08$. A more recent configuration-space analysis by Chang et al. (2023), using galaxy lensing and galaxy clustering data from DES Y3 cross-correlated with Planck + SPT CMB lensing, found a somewhat larger amplitude of $S_8 \approx 0.73 \pm 0.03$, though still low compared to the D3A or Planck CMB cosmologies. For comparison, the D3A, Planck, and LS8 cosmologies have S_8 values of 0.815, 0.833, and 0.766, respectively.

The bottom panels of Fig. 8 examine the feedback dependence of the predicted shear–CMB lensing cross. Minor effects are visible at $\ell \gtrsim 500$, but these do not alter the conclusions drawn above.

Going forward, there is considerable promise in measurements of this particular cross-spectrum, as high signal-to-noise tomographic analysis will be possible by combining Euclid and LSST shear measurements with Planck, Advanced ACT, and Simons Observatory CMB lensing measurements, allowing one to bridge the gap in scale and redshift between current cosmic shear-only and CMB lensing-only measurements.

3.7 Summary of comparison to observational data

We provide here an overall summary of the comparison of the FLAMINGO suite of simulations to the various observed auto- and cross-spectra discussed above, along with the CMB lensing–tSZ effect cross-spectrum presented in Schaye et al. (2023). In particular, in Table 5, we present the χ^2 values and number of standard deviations, N_σ, of each simulation with respect to the observed power spectra. For the cosmic shear power spectrum ($\gamma_{\rm E}$-$\gamma_{\rm E}$) and cosmic shear–tSZ effect cross-spectrum ($\gamma_{\rm E}$-y), we use the full covariance matrices to compute χ^2, whereas for the other cases, we use the diagonal elements of the covariance matrix only, as the covariance matrices were either unavailable or could not be straightforwardly combined when multiple data sets were used. The numbers displayed in bold face correspond to the models that are within 3σ of the observational measurements.

We see from Table 5 that the LS8 model performs the best for all the comparisons apart from the CMB lensing power spectrum, where it is clearly disfavoured, and the DES Y3 cosmic shear power spectrum (PlanckNu0p24Fix performs marginally better). However, the cosmic shear–CMB lensing ($\gamma_{\rm E}$-$\kappa_{\rm CMB}$) and CMB lensing–tSZ

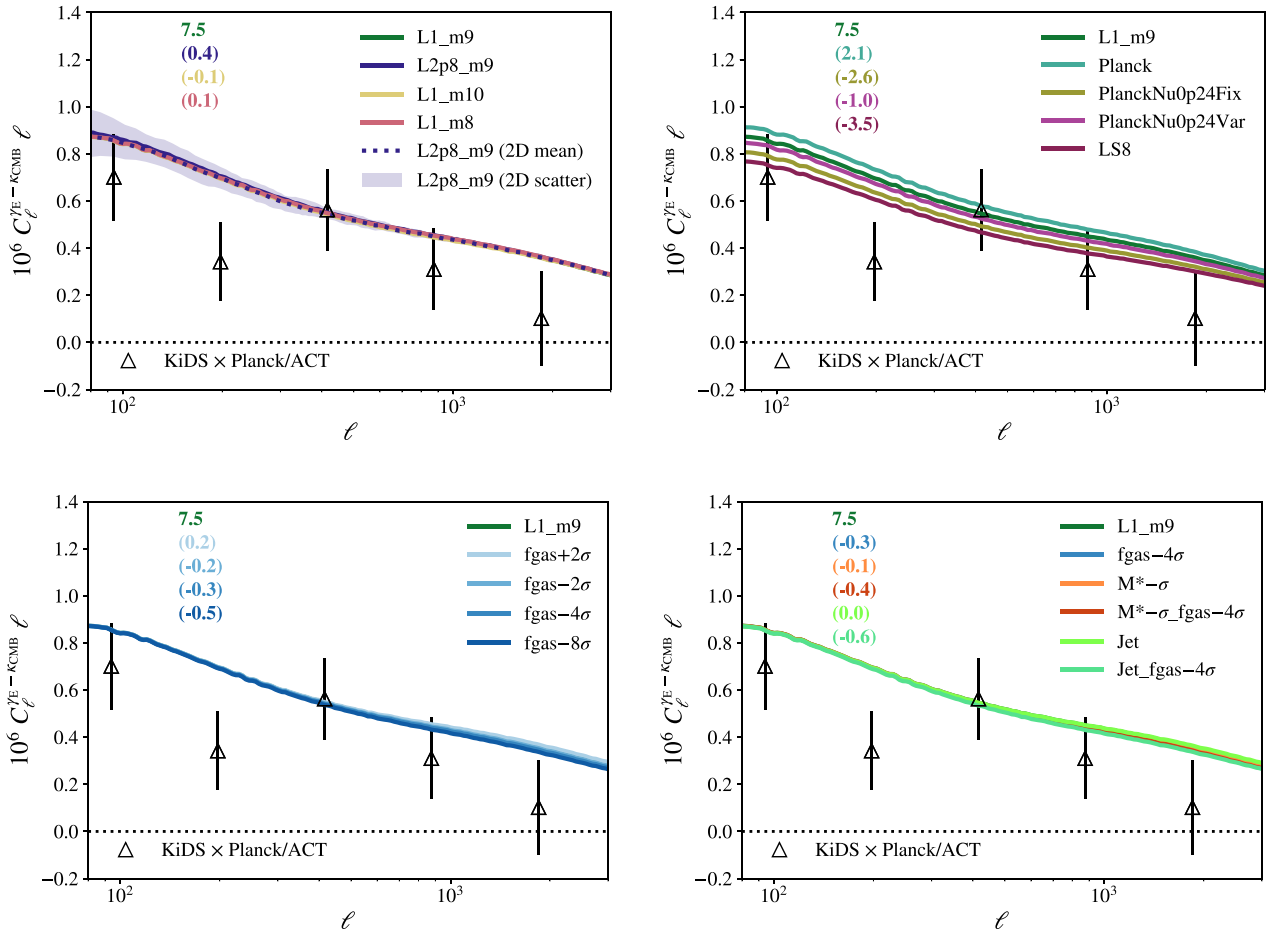

Figure 8. The cosmic shear–CMB lensing angular cross-spectrum. The open triangles correspond to the KiDS 1000 × Planck/ACT CMB lensing cross-spectrum measurements of Robertson et al. (2021). The χ^2 values displayed in the top left of each panel take into account only the diagonal elements of the covariance matrix. Top left: Dependence on simulation resolution and box size. The simulation predictions are converged with resolution, box size, and analysis method (1D; solid curves versus 2D; dashed curve). The shaded region shows the scatter between the different light cones for the 2D map-based method. Top right: Dependence on cosmology. A Planck maximum-likelihood cosmology yields spectra with elevated power relative to the fiducial D3A (and is thus in stronger tension with the measurements), whereas increasing the summed neutrino mass from 0.06 eV (fiducial) to 0.24 eV suppresses the power over all scales sampled here. The LS8 'lensing' cosmology predicts the lowest power and is in best agreement with the cross-power spectrum measurements. Bottom left: Dependence on baryon physics, namely variations in the gas fractions of groups and clusters which are mediated primarily through variations in the AGN feedback strength. Feedback effects are most evident at small angular scales ($\ell \gtrsim 1000$) and cannot reconcile the offset from the large-scale measurements. Bottom right: Dependence on other baryon variations, including variations in the stellar mass function (both at the fiducial and reduced gas fractions) and the fiducial and strong jet models of AGN feedback. The effects of variations in the stellar mass function are generally negligible compared to variations in the gas fractions.

effect[9] (κ_{CMB}-y) cross-spectra are not particularly constraining at present, and the Planck and D3A cosmologies are only in mild tension with the data for these tests. The strongest evidence for a S_8 tension comes from the tSZ effect power spectrum (y-y), the cosmic shear–tSZ effect cross-spectrum (γ_E-y), and the cosmic shear power spectrum (γ_E-γ_E) in order of decreasing tension level. This highlights the importance of the tSZ effect for constraining cosmology. We also highlight that when the small-scale SPT measurements are included in the tSZ power spectrum comparison, all models are formally rejected. As noted previously, some combination of a somewhat reduced value of S_8 and more extreme feedback may be able to accommodate the SPT measurements, but we leave that for future work.

While Table 5 provides a convenient way to compare the models and to demonstrate which tests are most constraining, there are two caveats to bear in mind, particularly with regard to the level of tension reported. First, we have not marginalized over any possible sources of systematic error in the observations, such as photometric redshift uncertainties and intrinsic alignments in the case of cosmic shear, CIB leakage in the tSZ effect, and so on. Indeed, this likely explains why the LS8 model does not provide a particularly good

[9]Just prior to submission of this paper, McCarthy & Hill (2023) presented an updated measurement of the CMB lensing-tSZ effect cross-spectrum using Planck data. For their highest signal-to-noise measurement, they find an amplitude of $A = 0.82 \pm 0.21$ compared to a Planck cosmology with expectation of $A = 1$. Hill & Spergel (2014) found $A = 1.10 \pm 0.22$ with respect to the same cosmology. In the analysis of Schaye et al. (2023) and in Table 5, we scaled the amplitude of Hill & Spergel (2014) down by 20 per cent based on Hurier (2015), which would correspond to $A = 0.92 \pm 0.18$, which is consistent with the new findings of McCarthy & Hill (2023). The slightly lower amplitude of the new measurements implies a slightly stronger tension with the Planck cosmology, though only at the $\approx 1\sigma$ level (compared to 0.6σ in Table 5).

Table 5. The χ^2 values and number of standard deviations, N_σ, of each simulation with respect to the observed auto- and cross-power spectra. For the cosmic shear power spectrum ($\gamma_{\rm E}$-$\gamma_{\rm E}$) and cosmic shear–tSZ effect cross-spectrum ($\gamma_{\rm E}$-y), we use the full covariance matrices to compute χ^2, whereas for the other cases, we use the diagonal errors only. We compute χ^2 values with respect to the following observational data sets: the KiDS 1000 cosmic shear measurements of Tröster et al. (2022) [$\gamma_{\rm E}$-$\gamma_{\rm E}$ (KiDS)], the DES Y3 cosmic shear measurements of Doux et al. (2022) [$\gamma_{\rm E}$-$\gamma_{\rm E}$ (DES)], the Planck tSZ effect power spectrum measurements of Bolliet et al. (2018) [y-y], the Planck + SPT tSZ effect power spectrum measurements of Bolliet et al. (2018) and Reichardt et al. (2021) [y-y (+ SPT)], the Planck, SPT, and ACT CMB lensing power spectrum measurements of Planck Collaboration VIII (2020b), Wu et al. (2019), and Qu et al. (2023), respectively [$\kappa_{\rm CMB}$-$\kappa_{\rm CMB}$], the KiDS 1000 cosmic shear–Planck tSZ effect cross-spectrum of Tröster et al. (2022) [$\gamma_{\rm E}$-y], the KiDS 1000 cosmic shear–ACT/Planck CMB lensing cross-spectrum of Robertson et al. (2021) [$\gamma_{\rm E}$-$\kappa_{\rm CMB}$], and the Planck CMB lensing–tSZ effect cross-spectrum of Hill & Spergel (2014) [$\kappa_{\rm CMB}$-y]. The numbers displayed in bold face correspond to the models that are within 3σ of the observational measurements.

Prefix	$\gamma_{\rm E}$-$\gamma_{\rm E}$ (KiDS)		$\gamma_{\rm E}$-$\gamma_{\rm E}$ (DES)		y-y		y-y (+ SPT)		$\kappa_{\rm CMB}$-$\kappa_{\rm CMB}$		$\gamma_{\rm E}$-y		$\gamma_{\rm E}$-$\kappa_{\rm CMB}$		$\kappa_{\rm CMB}$-y	
	χ^2	N_σ	χ^2	N_σ	χ^2	N_σ	χ^2	N_σ	χ^2	N_σ	χ^2	N_σ	χ^2	N_σ	χ^2	N_σ
L1_m9	183.2	3.6	324.3	**2.3**	139.7	10.5	435.3	19.9	29.5	**0.4**	78.9	3.6	7.5	**1.2**	2.8	**0.2**
L2p8_m9	185.4	3.7	327.7	**2.4**	225.7	13.9	568.4	22.9	27.1	**0.1**	86.6	4.1	8.0	**1.3**	3.2	**0.4**
L1_m10	181.5	3.5	321.5	**2.2**	160.0	11.4	455.3	20.4	29.8	**0.5**	76.9	3.5	7.4	**1.2**	2.8	**0.2**
L1_m8	184.6	3.7	325.5	**2.3**	115.2	9.4	372.0	18.2	29.4	**0.4**	75.9	3.4	7.7	**1.3**	2.6	**0.1**
fgas+2σ	185.6	3.7	325.5	**2.3**	140.4	10.6	470.6	20.7	29.4	**0.4**	82.8	3.9	7.7	**1.3**	2.7	**0.2**
fgas−2σ	181.2	3.5	323.3	**2.2**	129.0	10.0	368.0	18.1	29.5	**0.4**	73.5	3.2	7.4	**1.2**	2.8	**0.2**
fgas−4σ	179.4	3.4	322.5	**2.2**	117.2	9.5	302.8	16.3	29.6	**0.4**	67.9	**2.8**	7.2	**1.1**	2.8	**0.2**
fgas−8σ	176.6	3.3	321.6	**2.2**	91.5	8.1	198.4	12.8	29.7	**0.4**	59.0	**2.0**	7.0	**1.1**	3.0	**0.3**
M*−σ	181.9	3.6	323.8	**2.2**	161.0	11.5	468.9	20.7	29.5	**0.4**	80.7	3.7	7.4	**1.2**	3.1	**0.3**
M*−σ_fgas−4σ	178.6	3.4	322.6	**2.2**	137.2	10.4	333.1	17.2	29.6	**0.4**	70.6	3.0	7.2	**1.1**	3.2	**0.4**
Jet	183.8	3.7	323.6	**2.2**	106.7	8.9	327.4	17.1	29.6	**0.4**	72.4	3.1	7.6	**1.2**	2.3	**0.0**
Jet_fgas−4σ	176.1	3.3	320.1	**2.1**	97.8	8.5	231.1	14.1	29.8	**0.5**	60.1	**2.1**	6.9	**1.1**	3.4	**0.4**
Planck	199.3	4.4	348.5	3.2	187.0	12.5	552.9	22.6	25.4	**0.1**	95.2	4.7	9.6	**1.7**	3.8	**0.6**
PlanckNu0p24Fix	166.9	**2.8**	311.2	**1.8**	47.4	5.2	184.4	12.3	66.2	4.0	52.0	**1.4**	5.0	**0.6**	1.8	**0.3**
PlanckNu0p24Var	176.9	3.3	316.8	**2.0**	75.0	7.1	258.1	14.9	53.4	**2.9**	62.6	**2.3**	6.5	**1.0**	1.9	**0.2**
LS8	162.0	**2.6**	315.7	**1.9**	8.8	**0.4**	63.2	6.2	93.4	5.8	39.2	**0.1**	4.0	**0.2**	2.4	**0.0**

fit to the KiDS 1000 cosmic shear power spectrum, with $N_\sigma = 2.6$ in that case (although we note that approximately 1σ of this tension comes from a single cross, the 5–1 bin, which shows large point-to-point fluctuations). Marginalizing over these uncertainties will tend to decrease the magnitude of the reported deviations. Secondly, we quote the χ^2 and N_σ values with respect to a specific set of cosmological parameters. For example, for the Planck simulation, we use the Planck 2018 maximum-likelihood cosmological parameters and the statistics we report are therefore with respect to that specific choice of parameters. The level of tension we report does not factor in the uncertainties in the cosmological parameters from the Planck primary CMB data set, it is only with respect to the maximum-likelihood cosmology. Thus, our estimates will be an upper limit for the level of tension between the primary CMB and LSS observables in this case.

4 DISCUSSION AND CONCLUSIONS

We have used the new FLAMINGO simulations to explore the spatial clustering and cross-correlation signals of three large-scale tracers: cosmic shear, CMB lensing, and the tSZ effect. Our analysis, carried out in harmonic space, included an exploration of the dependence of these signals on the choice of cosmological parameters including neutrino mass, the efficiency and nature of so-called 'baryonic feedback' (specifically stellar and AGN feedback), and simulation box size, resolution, and method of analysis (1D Limber integration versus a 2D map-based analysis). Note that the stellar and AGN feedback were calibrated using machine-learning methods to reproduce the observed galaxy stellar mass function and the gas fractions of galaxy groups and clusters. However, FLAMINGO also includes extreme variations with respect to the observed baryon fractions, obtained by varying the feedback efficiencies. We compared the full FLAMINGO suite of hydro simulations to the latest measurements of the auto- and cross-power spectra involving the three observables and commented on the evidence for an 'S_8 tension' and its robustness to uncertainties in baryonic modelling.

The main conclusions of this study may be summarized as follows:

(i) The auto- and cross-power spectra involving cosmic shear, CMB lensing, and the tSZ effect include contributions from a very wide range of physical scales and redshifts (Fig. 1). While there is considerable overlap between the different two-point functions in terms of physical scale and redshift (allowing for consistency tests), there is also a large degree of complementarity, with the CMB lensing power spectrum probing generally linear scales and higher redshifts while the cosmic shear and tSZ effect power spectra probe low redshifts and into the non-linear regime though they are sensitive to different halo masses. Cross-spectra between the different observables are typically sensitive to physical scales, redshifts, and halo masses that are intermediate between the power spectra of the individual fields, thus adding additional information.

(ii) The predicted cosmic shear power spectrum is robust to variations in box size, resolution, and method of analysis (Fig. 2). Consistent with recent studies (e.g. Heymans et al. 2021; Tröster et al. 2022; Amon et al. 2023), we find that a Planck CMB cosmology, and the fiducial cosmology that combines DES Y3 constraints with a variety of external data sets, including the primary CMB (D3A), predicts power spectra that are elevated with respect to the KiDS 1000 measurements and to a lesser extent the DES Y3 measurements. This is in the context of our fiducial calibrated hydro simulation. We find that adopting a lower value of S_8, as suggested by recent cosmic shear measurements (LS8 run), yields an improved match to the measurements, particularly for the higher tomographic bins that contain much of the integrated signal (Fig. 3). Increasing the feedback efficiency with respect to the fiducial calibrated model, which has the effect of lowering the gas fractions of groups and clusters, also marginally improves the quality of the fit (Fig. 4) in the context of the D3A cosmology, but this preference largely

evaporates when adopting a lower S_8. Even though it contains extreme variations in feedback, FLAMINGO predicts that current cosmic shear measurements are only marginally sensitive to the impact of baryon physics.

(iii) The predicted tSZ effect power spectrum is generally robust to variations in resolution, box size, and method of analysis (Fig. 5). As much of the signal is produced by local, massive clusters, very large simulation boxes ($\gtrsim 1$ Gpc) are required to accurately predict the power. Another consequence of this dependence on massive, nearby clusters is that cosmic variance uncertainties are important on large scales (particularly $\ell \lesssim 300$). The tSZ effect power spectrum is extremely sensitive to the amplitude of fluctuations (σ_8) and, on the angular scales probed by Planck, is generally insensitive to variations in the efficiency of feedback (owing to the fact that it is dominated by 'baryonically closed' massive clusters). The situation changes on smaller angular scales probed by SPT and ACT, which are more sensitive to group-mass haloes and therefore more susceptible to feedback effects. Current measurements of the tSZ effect on large scales (Bolliet et al. 2018) are in strong tension with the predictions of the hydrodynamical simulations in either a D3A or Planck CMB cosmology (see Table 5). Adopting a lower value of S_8, however, yields a significantly improved match to the measurements, consistent with the cosmic shear power spectrum analysis.

(iv) We find that the CMB lensing power spectrum is insensitive to simulation box size, resolution, method of analysis, and feedback implementation (see Fig. A3). This is a direct consequence of it probing mainly linear scales and high redshifts. In contrast to the cosmic shear and tSZ effect power spectra, the predicted CMB lensing power spectrum for both the D3A and Planck CMB cosmologies is in excellent agreement with recent measurements from Planck, SPT, and ACT, whereas the cosmologies with an increased neutrino mass and (particularly) LS8 are disfavoured (Fig. 6).

(v) The predicted cosmic shear–tSZ effect cross-spectrum is generally robust to variations in resolution, box size, and method of analysis (Fig. 7). Similar to the tSZ effect power spectrum, much of the signal is produced by local, massive clusters, thus requiring large simulation boxes to accurately predict this statistic. Consistent with Tröster et al. (2022), we find that the predicted cross-spectrum in both the D3A or Planck CMB cosmologies is elevated with respect to the observed cross-spectrum between KiDS 1000 and Planck and ACT tSZ on scales of $\ell \lesssim 1500$. Decreasing the gas fractions of groups and clusters by increasing the efficiency of feedback leads to a marginal improvement in the fit but cannot accommodate the observed offset on large angular scales ($\ell \lesssim 1000$). Adopting a lower S_8 improves the match to the data.

(vi) The predicted cosmic shear–CMB lensing cross-spectrum is very robust to variations in resolution, box size, and method of analysis (Fig. 8). The impact of baryons is also minimal compared to current measurement uncertainties. Consistent with Robertson et al. (2021), we find that the predicted cross-spectrum in either the D3A or Planck CMB cosmologies is generally elevated with respect to the observed cross-spectrum between KiDS 1000 and Planck and ACT CMB lensing. Adopting a lower S_8 improves the match to the data.

(vii) We summarize the level of tension of the simulations with respect to the observations for all of the auto- and cross-power spectra we have considered in Table 5.

Our findings are largely consistent with those of the individual studies that presented the measurements we compare to, at least in terms of qualitative conclusions. This was not guaranteed to be the case, as the individual studies all use very different theoretical frameworks, which is a crucial part of cosmological parameter inference. For example, some of the studies employ linear theory (e.g. Madhavacheril et al. 2023; Qu et al. 2023) while others use the halo model (e.g. Bolliet et al. 2018; Tröster et al. 2022) or gravity-only emulators (e.g. Robertson et al. 2021; Doux et al. 2022) to compute the non-linear evolution. With regards to the potential impact of baryons, in some studies, it was not taken into account (e.g. Robertson et al. 2021; Madhavacheril et al. 2023; Qu et al. 2023) while others employed the halo model and marginalized over the uncertainties (e.g. Bolliet et al. 2018; Tröster et al. 2022) while still others discarded small-scale measurements (e.g. Doux et al. 2022) in an attempt to avoid biasing due to baryons. Here, we have taken a single suite of hydro simulations and projected them into different observables, thus allowing a direct comparison between the different tests. From this comparison, three major conclusions are: (i) power spectra that probe late times and/or non-linear scales are in tension with the predictions of the standard LCDM model with parameters set by the primary CMB, BAO, and CMB lensing; (ii) while increasing the neutrino mass can reduce the tension between the CMB + BAO and LSS, measurements of the CMB lensing power spectrum disfavour this solution; and (iii) the effects of baryon physics and, importantly, its uncertainties, are generally insufficient to reconcile these tensions.

Our results do not preclude baryons from playing some minor role in the current discussion, which, perhaps when combined with other factors (e.g. systematic errors in observational measurements), could help to resolve the tension. Indeed recent cosmic shear analyses claim to have marginally detected deviations from the predictions of gravity-only simulations and at a level that is consistent with that predicted by calibrated simulations such as BAHAMAS and FLAMINGO (e.g. Aricò et al. 2023; Chen et al. 2023). And while including baryons in the modelling tends to boost the inferred value of S_8, in general, the value is shifted by less than 1σ. Thus, both calibrated simulations and current cosmic shear measurements point to a fairly benign role for baryons at present. However, this situation will change radically for forthcoming Stage IV surveys, including LSST, Euclid, and DESI, since the statistical precision of the measurements will be significantly higher and the measurements will extend to smaller scales, thus requiring a much more careful accounting of the impact of baryon physics.

Returning to the recent studies of Amon & Efstathiou (2022) and Preston, Amon & Efstathiou (2023), these authors showed that, given sufficient flexibility in the incorporated baryonic modelling, it is possible to obtain a satisfactory fit to both the primary CMB and cosmic shear measurements. The required suppression of the matter power spectrum is, however, typically much stronger than predicted by simulations such as BAHAMAS and FLAMINGO, for which the baryon physics has been calibrated to reproduce the baryon fractions of galaxy groups and clusters. Recent work has shown that the suppression is strongly tied to the baryon fractions (e.g. Semboloni et al. 2011; Semboloni, Hoekstra & Schaye 2013; Schneider et al. 2019; Van Daalen, McCarthy & Schaye 2020; Salcido et al. 2023), which was a motivating factor in the calibration strategies of these simulations. Thus, appealing to a much stronger suppression of the matter power spectrum in order to reconcile the primary CMB and LSS would appear to require baryon fractions that violate observational constraints on the baryon fractions of groups and clusters. Grandis et al. (2023) have recently come to similar conclusions using weak lensing-calibrated cluster gas and stellar fractions together with the baryonification methods of Schneider et al. (2019) and Aricò et al. (2021) to predict the suppression of the matter

power spectrum. However, we note that the level of feedback required to reconcile the primary CMB and the cosmic shear measurements depends on the cosmic shear data set employed. Preston, Amon & Efstathiou (2023) find that relatively weaker feedback, which is more comparable to that in BAHAMAS and FLAMINGO, is required to reconcile the DES Y3 measurements with the CMB, whereas stronger feedback is required when using the KiDS 1000 measurements. This is likely driven by the fact that the tension in S_8 between the DES Y3 measurements and the CMB is relatively mild to begin with, before baryonic effects are considered.

In terms of using baryon fractions to constrain baryonic effects on the matter power spectrum, a caveat that is worth further consideration is that carefully accounting for the X-ray selection function of galaxy groups is non-trivial (e.g. Pearson et al. 2017). Also, the present cosmic shear data are sensitive to LSS over a wider range of redshifts than for which we presently have useful observational constraints on group/cluster baryon fractions. Thus, if the simulations significantly underestimate the efficiency of feedback at higher redshifts ($z \gtrsim 0.3$), then they could underestimate the impact of baryons on the cosmic shear power spectrum. Observations of groups/clusters at higher redshift and with a well-defined selection function would be highly valuable. The recent kinetic SZ (kSZ) effect stacking measurements of SDSS eBOSS galaxies at $z \approx 0.5$ in Schaan et al. (2021) appear to be well suited as an independent test of feedback models (e.g. Schneider et al. 2022). We emphasize, however, that the most strongly discrepant power spectrum we examined is the Planck tSZ effect power spectrum, which is primarily sensitive to low-redshift, massive clusters which are baryonically closed, dominated by hot gas, and for which X-ray and tSZ surveys are typically highly complete. Increasing feedback at higher redshifts will not significantly impact this metric.

If neither feedback nor unaccounted for (or mischaracterized) systematic errors are behind the tension (though it will require additional work to conclusively demonstrate the latter), then the exciting implication would appear to be that new physics, perhaps in the dark sector, is required. Specifically, new physics that preferentially impacts non-linear scales and/or late times in order to retain consistency with the CMB lensing power spectrum, as proposed by Amon & Efstathiou (2022) and Preston, Amon & Efstathiou (2023) as an alternative interpretation (to baryon physics) of the required suppression. Many suggestions in this vein have recently been put forward, including a contribution from ultra-light axions to the dark matter (e.g. Rogers et al. 2023), mild baryon–dark matter scattering (He et al. 2023), and invoking interactions (in the form of a frictional drag) between dark matter and dark energy (Poulin et al. 2023). Testing these extensions will require forthcoming Stage IV surveys, which will measure the power and cross-power spectra with much higher statistical precision and will allow for much finer binning in redshift, physical scale, and halo mass. With the increased statistical precision comes the requirement for a very careful consideration of baryonic effects, which may be degenerate with the additional degrees of freedom in the dark sector extensions.

Finally, to make stronger statements about tensions and possible resolutions thereof in the context of cosmological hydrodynamical simulations and which simultaneously accounts for relevant observational systematic uncertainties, we will require a way to quickly span a wide range of baryon feedback scenarios and background cosmologies in order to incorporate the predictions of such simulations into cosmological pipelines. The latter (cosmological variations) has already been achieved but only in the context of gravity-only simulations, through emulators constructed from grids of simulations that span some range of cosmological parameters (e.g. Heitmann et al. 2014; Lawrence et al. 2017; Euclid Collaboration 2019; McClintock et al. 2019; Euclid Collaboration 2021). The former has just recently been achieved for the first time in volumes of sufficient size for LSS applications, in Salcido et al. (2023). Those authors produced a suite of 400 cosmological hydro simulations in 100 Mpc h^{-1} volumes (the Antilles simulations) and have developed an emulator for the relative impact of baryons on the matter power spectrum that takes the mass-dependent baryon fractions of groups and clusters as its input and predicts the suppression of the matter power spectrum. In the context of current cosmic shear measurements, this is the most accurate emulator currently available and can be easily applied in existing pipelines.

However, for the next generation of measurements, which will have significantly improved statistical precision, the assumed separability of baryonic effects and cosmological effects in the approach of Salcido et al. (2023) and other approximate methods for incorporating baryons (e.g. HMcode; Mead et al. 2020) will need to be revisited. Ideally, a single emulator based on a grid of cosmological hydrodynamical simulations that *simultaneously* varies the relevant astrophysical and cosmological parameters should be the basis of cosmological pipelines. Furthermore, to take advantage of the wide variety of complementary LSS observables, including weak lensing, galaxy clustering, redshift-space distortions, the tSZ and kSZ effects, cluster counts, etc., box sizes of $\sim$1 Gpc are required. With FLAMINGO, we have taken an important step forward to show that it is possible to carry out such simulations in a careful way, using machine-learning-based emulators as part of the calibration. The next step is to extend this approach to a simultaneous exploration of cosmology and astrophysics, which is the subject of ongoing work.

ACKNOWLEDGEMENTS

The authors thank the referee for a prompt and constructive report. IGM thanks Tilman Tröster and Cyrille Doux for assistance with KiDS and DES cosmic shear measurements and Alexandra Amon for helpful comments on a draft version of the paper. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 769130). CSF acknowledges support from the European Research Council (ERC) through Advanced Investigator grant DMIDAS (GA 786910). This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1, and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure.

DATA AVAILABILITY

The data supporting the plots within this article are available on reasonable request to the corresponding author. The FLAMINGO simulation data will eventually be made publicly available, though we note that the data volume (several petabytes) may prohibit us from simply placing the raw data on a server. In the meantime, people interested in using the simulations are encouraged to contact the corresponding author.

REFERENCES

Abazajian K. et al., 2019, preprint (arXiv:1907.04473)
Abbott T. M. C. et al., 2022, Phys. Rev. D, 105, 023520

Abbott T. M. C. et al., 2023, preprint (arXiv:2305.17173)
Addison G. E., Huang Y., Watts D. J., Bennett C. L., Halpern M., Hinshaw G., Weiland J. L., 2016, ApJ, 818, 132
Ade P. et al., 2019, J. Cosmol. Astropart. Phys., 2019, 056
Aihara H. et al., 2018, PASJ, 70, S8
Akino D. et al., 2022, PASJ, 74, 175
Alonso D., Sanchez J., Slosar A., LSST Dark Energy Science Collaboration, 2019, MNRAS, 484, 4127
Amon A., Efstathiou G., 2022, MNRAS, 516, 5355
Amon A. et al., 2023, MNRAS, 518, 477
Angulo R. E., Pontzen A., 2016, MNRAS, 462, L1
Aricò G., Angulo R. E., Contreras S., Ondaro-Mallea L., Pellejero-Ibañez M., Zennaro M., 2021, MNRAS, 506, 4070
Aricò G., Angulo R. E., Zennaro M., Contreras S., Chen A., Hernández-Monteagudo C., 2023, A&A, 678, A109
Bagla J. S., Ray S., 2003, New Astron., 8, 665
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Battaglia N., Bond J. R., Pfrommer C., Sievers J. L., 2012, ApJ, 758, 75
Battye R. A., Moss A., 2014, Phys. Rev. Lett., 112, 051303
Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143
Beutler F. et al., 2014, MNRAS, 444, 3501
Bird S., Feng Y., Pedersen C., Font-Ribera A., 2020, J. Cosmol. Astropart. Phys., 2020, 002
Birkinshaw M., 1999, Phys. Rep., 310, 97
Bolliet B., Comis B., Komatsu E., Macías-Pérez J. F., 2018, MNRAS, 477, 4957
Borrow J., Schaller M., Bower R. G., Schaye J., 2022, MNRAS, 511, 2367
Bower R. G., Vernon I., Goldstein M., Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2010, MNRAS, 407, 2017
Carlstrom J. E., Holder G. P., Reese E. D., 2002, ARA&A, 40, 643
Castro T., Borgani S., Dolag K., Marra V., Quartin M., Saro A., Sefusatti E., 2021, MNRAS, 500, 2316
Chang C. et al., 2023, Phys. Rev. D, 107, 023530
Chen A. et al., 2023, MNRAS, 518, 5340
Cheng H., Greengard L., Rokhlin V., 1999, J. Comput. Phys., 155, 468
Chisari N. E. et al., 2019, Open J. Astrophys., 2, 4
da Silva A. C., Kay S. T., Liddle A. R., Thomas P. A., Pearce F. R., Barbosa D., 2001, ApJ, 561, L15
Debackere S. N. B., Schaye J., Hoekstra H., 2020, MNRAS, 492, 2285
de Salas P. F., Forero D. V., Gariazzo S., Martínez-Miravé P., Mena O., Ternes C. A., Tórtola M., Valle J. W. F., 2021, J. High Energy Phys., 2021, 71
Dehnen W., 2014, Comput. Astrophys. Cosmol., 1, 1
Doux C. et al., 2022, MNRAS, 515, 1942
Elbers W., Frenk C. S., Jenkins A., Li B., Pascoli S., 2021, MNRAS, 507, 2614
Elbers W., Frenk C. S., Jenkins A., Li B., Pascoli S., 2022, MNRAS, 516, 3821
Esteban I., Gonzalez-Garcia M. C., Maltoni M., Schwetz T., Zhou A., 2020, J. High Energy Phys., 2020, 178
Euclid Collaboration, 2019, MNRAS, 484, 5509
Euclid Collaboration, 2021, MNRAS, 505, 2840
Fang X. et al., 2023, preprint (arXiv:2308.01856)
Gonzalez A. H., Sivanandam S., Zabludoff A. I., Zaritsky D., 2013, ApJ, 778, 14
Grandis S., Arico' G., Schneider A., Linke L., 2023, preprint (arXiv:2309.02920)
Greengard L., Rokhlin V., 1987, J. Comput. Phys., 73, 325
Hahn O., Michaux M., Rampf C., Uhlemann C., Angulo R. E., 2020, Astrophysics Source Code Library, p. ascl
Hahn O., Rampf C., Uhlemann C., 2021, MNRAS, 503, 426
He A., Ivanov M. M., An R., Gluscevic V., 2023, ApJ, 954, L8
Heitmann K., Lawrence E., Kwan J., Habib S., Higdon D., 2014, ApJ, 780, 111
Henden N. A., Puchwein E., Shen S., Sijacki D., 2018, MNRAS, 479, 5385
Heymans C. et al., 2021, A&A, 646, A140
Hildebrandt H. et al., 2016, MNRAS, 463, 635
Hildebrandt H. et al., 2021, A&A, 647, A124
Hill J. C., Spergel D. N., 2014, J. Cosmol. Astropart. Phys., 2014, 030
Hojjati A. et al., 2017, MNRAS, 471, 1565
Hu W., Okamoto T., 2002, ApJ, 574, 566
Hurier G., 2015, A&A, 575, L11
Jenkins A., 2013, MNRAS, 434, 2094
Kaiser N., 1992, ApJ, 388, 272
Kaiser N., Squires G., 1993, ApJ, 404, 441
Kilbinger M. et al., 2017, MNRAS, 472, 2126
Komatsu E., Seljak U., 2002, MNRAS, 336, 1256
Kugel R. et al., 2023, preprint (arXiv:2306.05492)
Kuijken K. et al., 2019, A&A, 625, A2
Lawrence E. et al., 2017, ApJ, 847, 50
Le Brun A. M. C., McCarthy I. G., Schaye J., Ponman T. J., 2014, MNRAS, 441, 1270
Lewis A., Challinor A., 2006, Phys. Rep., 429, 1
Liu X., Emberson J. D., Buehlmann M., Frontiere N., Habib S., 2023, MNRAS, 522, 3631
Madhavacheril M. S. et al., 2023, preprint (arXiv:2304.05203)
Mallaby-Kay M. et al., 2021, ApJS, 255, 11
McCarthy F., Hill J. C., 2023, preprint (arXiv:2308.16260)
McCarthy I. G. et al., 2010, MNRAS, 406, 822
McCarthy I. G., Schaye J., Bower R. G., Ponman T. J., Booth C. M., Dalla Vecchia C., Springel V., 2011, MNRAS, 412, 1965
McCarthy I. G., Le Brun A. M. C., Schaye J., Holder G. P., 2014, MNRAS, 440, 3645
McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, MNRAS, 465, 2936
McCarthy I. G., Bird S., Schaye J., Harnois-Deraps J., Font A. S., van Waerbeke L., 2018, MNRAS, 476, 2999
McCarthy F., Hill J. C., Madhavacheril M. S., 2022, Phys. Rev. D, 105, 023517
McClintock T. et al., 2019, ApJ, 872, 53
Mead A. J., Tröster T., Heymans C., Van Waerbeke L., McCarthy I. G., 2020, A&A, 641, A130
Michaux M., Hahn O., Rampf C., Angulo R. E., 2021, MNRAS, 500, 663
Millea M., Doré O., Dudley J., Holder G., Knox L., Shaw L., Song Y. S., Zahn O., 2012, ApJ, 746, 4
Mummery B. O., McCarthy I. G., Bird S., Schaye J., 2017, MNRAS, 471, 227
Nguyen N.-M., Huterer D., Wen Y., 2023, Phys. Rev. Lett., 131, 111001
Oppenheimer B. D., Babul A., Bahé Y., Butsky I. S., McCarthy I. G., 2021, Universe, 7, 209
Pearson R. J. et al., 2017, MNRAS, 469, 3489
Planck Collaboration XI, 2013, A&A, 557, A52
Planck Collaboration XXI, 2014, A&A, 571, A21
Planck Collaboration XXII, 2016, A&A, 594, A22
Planck Collaboration LI, 2017, A&A, 607, A95
Planck Collaboration VI, 2020a, A&A, 641, A6
Planck Collaboration VIII, 2020b, A&A, 641, A8
Planelles S., Borgani S., Fabjan D., Killedar M., Murante G., Granato G. L., Ragone-Figueroa C., Dolag K., 2014, MNRAS, 438, 195
Poulin V., Bernal J. L., Kovetz E. D., Kamionkowski M., 2023, Phys. Rev. D, 107, 123538
Preston C., Amon A., Efstathiou G., 2023, MNRAS, 525, 5554
Price D. J., 2012, J. Comput. Phys., 231, 759
Qu F. J. et al., 2023, preprint (arXiv:2304.05202)
Rampf C., Uhlemann C., Hahn O., 2021, MNRAS, 503, 406
Reichardt C. L. et al., 2021, ApJ, 908, 199
Reid B. et al., 2016, MNRAS, 455, 1553
Riess A. G. et al., 2022, ApJ, 934, L7
Robertson N. C. et al., 2021, A&A, 649, A146
Rodrigues L. F. S., Vernon I., Bower R. G., 2017, MNRAS, 466, 2418

Rogers K. K., Hložek R., Laguë A., Ivanov M. M., Philcox O. H. E., Cabass G., Akitsu K., Marsh D. J. E., 2023, J. Cosmol. Astropart. Phys., 2023, 023
Rosenberg E., Gratton S., Efstathiou G., 2022, MNRAS, 517, 4620
Salcido J., McCarthy I. G., Kwan J., Upadhye A., Font A. S., 2023, MNRAS, 523, 2247
Schaan E. et al., 2021, Phys. Rev. D, 103, 063513
Schaller M. et al., 2023, preprint (arXiv:2305.13380)
Schaye J. et al., 2015, MNRAS, 446, 521
Schaye J. et al., 2023, preprint (arXiv:2306.04024)
Schneider A., Teyssier R., Stadel J., Chisari N. E., Le Brun A. M. C., Amara A., Refregier A., 2019, J. Cosmol. Astropart. Phys., 2019, 020
Schneider A., Stoira N., Refregier A., Weiss A. J., Knabenhans M., Stadel J., Teyssier R., 2020, J. Cosmol. Astropart. Phys., 2020, 019
Schneider A., Giri S. K., Amodeo S., Refregier A., 2022, MNRAS, 514, 3802
Semboloni E., Hoekstra H., Schaye J., van Daalen M. P., McCarthy I. G., 2011, MNRAS, 417, 2020
Semboloni E., Hoekstra H., Schaye J., 2013, MNRAS, 434, 148
Shaw L. D., Nagai D., Bhattacharya S., Lau E. T., 2010, ApJ, 725, 1452
Silk J., Rees M. J., 1998, A&A, 331, L1
Smith R. E. et al., 2003, MNRAS, 341, 1311
Springel V. et al., 2018, MNRAS, 475, 676
Sunyaev R. A., Zeldovich Y. B., 1972, Comm. Astrophys. Space Phys., 4, 173
Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152
Tröster T. et al., 2022, A&A, 660, A27
Upadhye A. et al., 2023, preprint (arXiv:2308.09755)
van Daalen M. P., Schaye J., 2015, MNRAS, 452, 2247
Van Daalen M. P., Schaye J., Booth C. M., Dalla Vecchia C., 2011, MNRAS, 415, 3649
Van Daalen M. P., McCarthy I. G., Schaye J., 2020, MNRAS, 491, 2424
Wei C. et al., 2018, ApJ, 853, 25
White S. D. M., Frenk C. S., 1991, ApJ, 379, 52
White M., Hernquist L., Springel V., 2002, ApJ, 579, 16
Wu W. L. K. et al., 2019, ApJ, 884, 70
Wyman M., Rudd D. H., Vanderveld R. A., Hu W., 2014, Phys. Rev. Lett., 112, 051302

APPENDIX A: ADDITIONAL COMPARISONS

In this appendix, we present additional analyses referred to in the main text.

In Fig. A1, we examine the dependence of the predicted KiDS 1000 cosmic shear power spectrum on baryon physics, specifically variations in the SMF (both at the fiducial and reduced gas fractions) and the fiducial and strong jet models of AGN feedback. We conclude that variations in the SMF are generally negligible compared to that of variations in the gas fractions.

In Fig. A2, we compare selected FLAMINGO runs to the DES Y3 cosmic shear power spectra data from Doux et al. (2022). The DES Y3 background galaxy population is split among four tomographic bins shown in fig. 2 of Doux et al. (2022). We use these source redshift distributions in equation (2) to compute their respective window functions. As we are particularly interested in the role of baryons, we do not include any scale cuts to the DES Y3 power spectra. Examining Fig. A2, we see that the lensing LS8 cosmology yields a somewhat better fit to the data relative to the fiducial D3A cosmology (particularly among the majority of the higher tomographic bins), whereas a Planck CMB cosmology yields a worse fit for virtually all bins. Increasing the efficiency of feedback also slightly improves the fit, but less than the improvement due to lowering S_8. Consistent with findings in the literature (e.g. Abbott et al. 2022, 2023), we conclude that while there is some evidence for tension of the DES Y3 measurements with the Planck CMB cosmology, it is of slightly lower significance than for the KiDS 1000 survey.

In Fig. A3, we examine the dependence of the predicted CMB lensing angular power spectrum on simulation box size and resolution (top panel) and baryon physics (bottom panels). Over the range of scales examined here, the predicted CMB lensing power spectrum is converged (i.e. unaltered) with respect to these variations.

In Fig. A4, we compare the FLAMINGO cosmology and gas fraction variation runs with the cosmic shear–tSZ effect angular cross-power spectrum measurements of Hojjati et al. (2017), using data from RCSLenS (shear) and Planck (tSZ). As RCSLenS is a relatively shallow survey, Hojjati et al. (2017) combined the data into a single large tomographic bin. We use the source redshift distribution from that study to compute the predicted shear–tSZ effect cross. Consistent with the KiDS 1000 × Planck cross-examined in the main text, we see that four of the five multipole bins have amplitudes lower than that predicted by the fiducial FLAMINGO model in the fiducial D3A cosmology. However, as these four bins sample relatively small angular scales of $\ell \sim 1000$, where the impact of feedback is non-negligible, the significance of the tension with the D3A cosmology is clearly less pronounced than for the KiDS 1000 comparison in the main text.

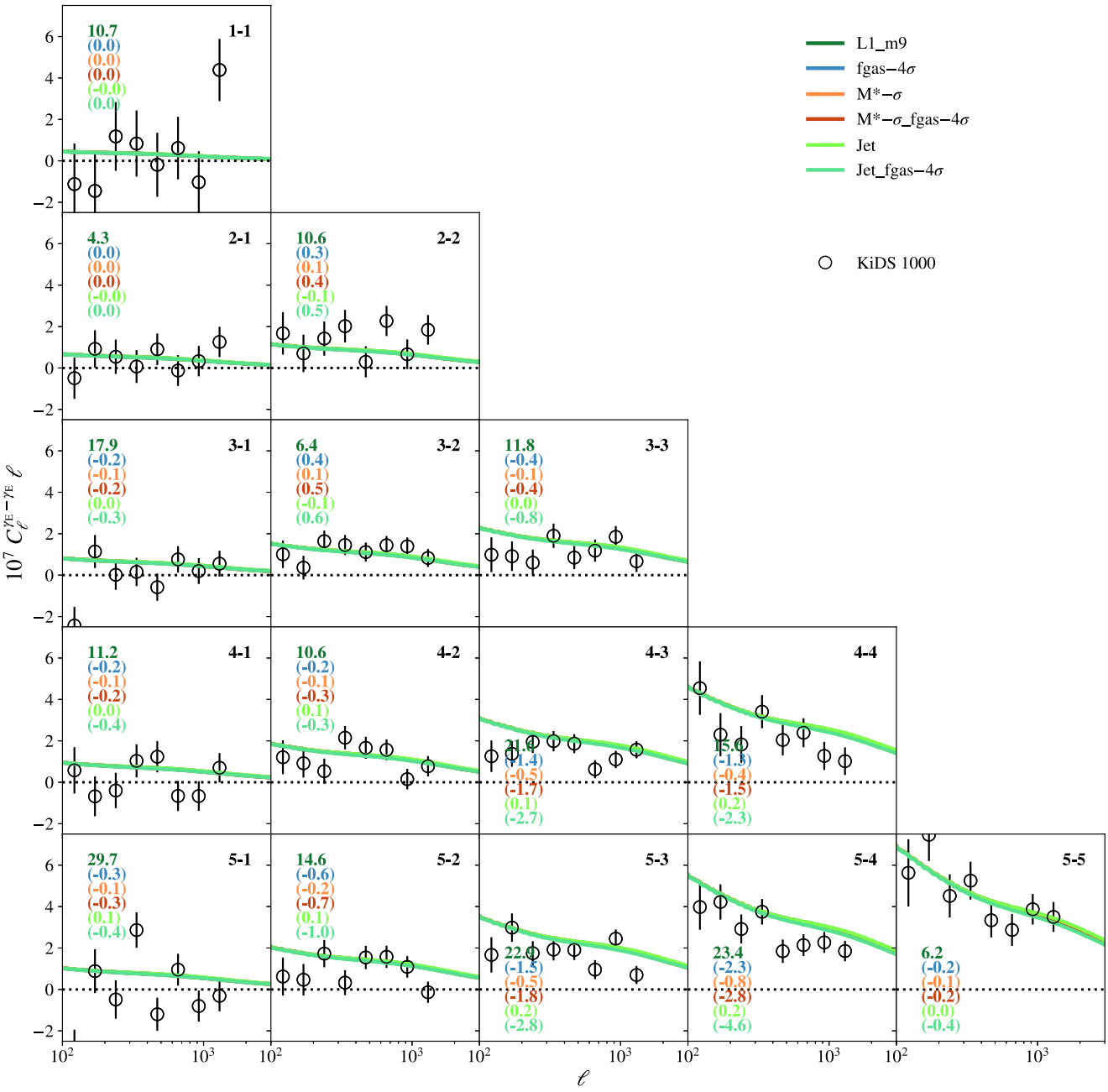

Figure A1. As Fig. 2, but showing the dependence of the predicted KiDS 1000 cosmic shear power spectrum on baryon physics, namely variations in the stellar mass function (both the fiducial and reduced cluster gas fractions) and the fiducial and strong jet models of AGN feedback. The solid coloured curves correspond to the predicted spectra for the FLAMINGO simulations as baryon models are varied. The effect of variations in the stellar mass function is generally negligible compared to that of variations in the gas fractions.

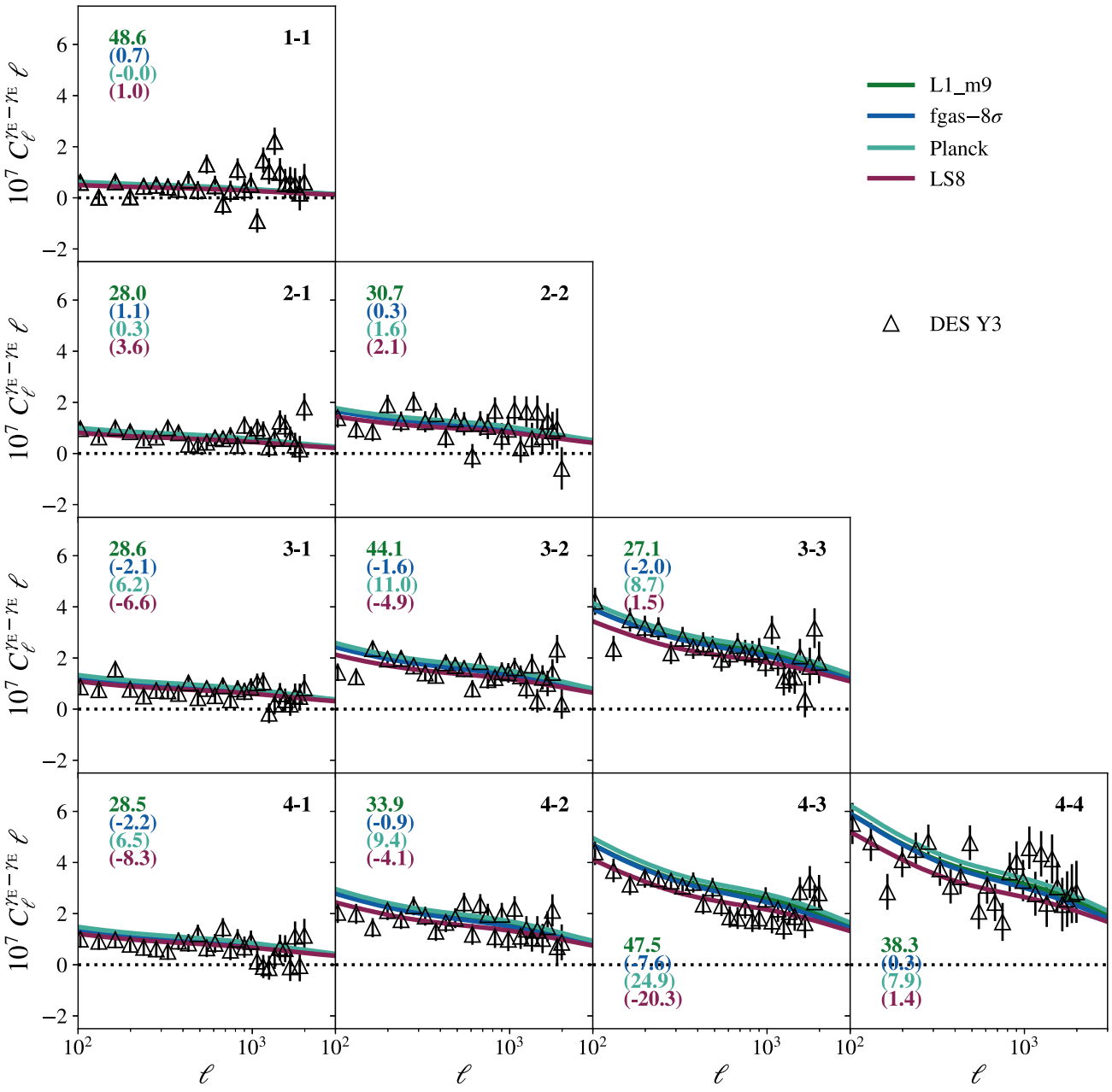

Figure A2. As Fig. 2, but showing selected predictions for the DES Y3 cosmic shear power spectrum. The open triangles correspond to the DES Y3 measurements of Doux et al. (2022) and the error bars correspond to the diagonal components of the covariance matrix. Note that the χ^2's are computed here adopting the diagonal errors only. A lensing LS8 cosmology yields somewhat better fit to the data relative to the fiducial D3A cosmology, whereas a Planck CMB cosmology yields a worse fit. Increasing the efficiency of feedback also slightly improves the fit, but less than the improvement due to lowering S_8.

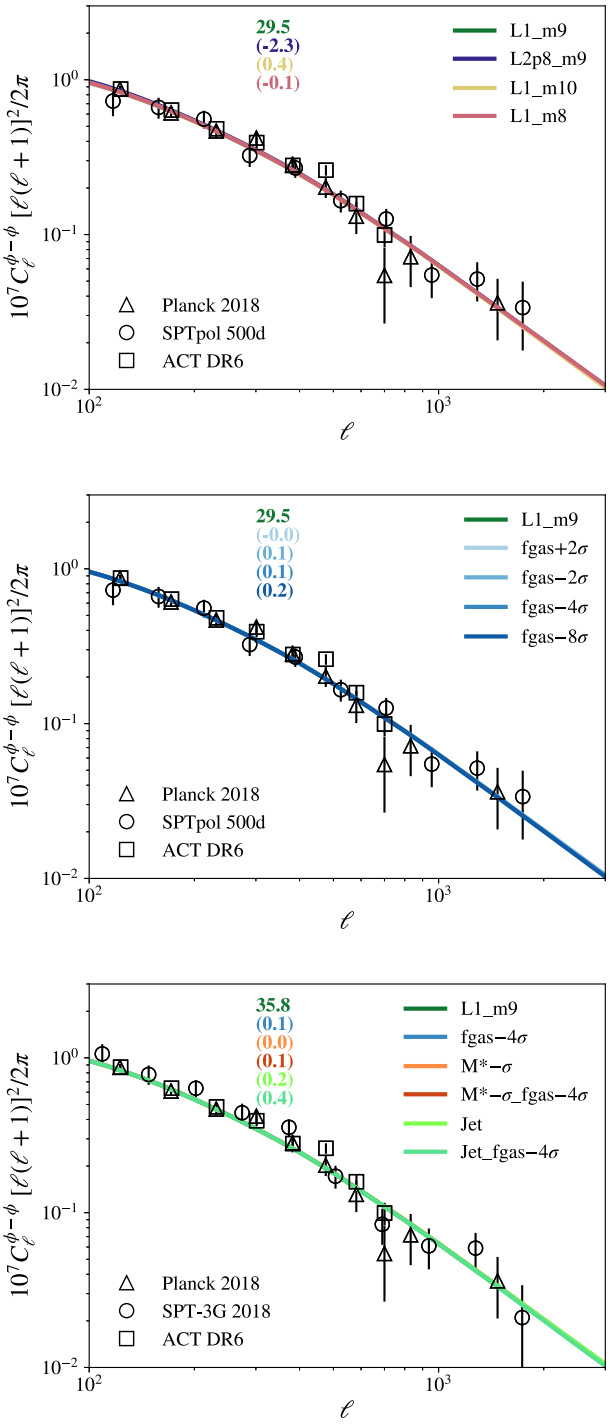

Figure A3. As Fig. 6, but showing the dependence of the predicted CMB lensing angular power spectrum on simulation box size and resolution (top panel) and baryon physics (middle and bottom panels). Over the range of scales examined here, the predicted CMB lensing power spectrum is converged (i.e. unaltered) with respect to these variations.

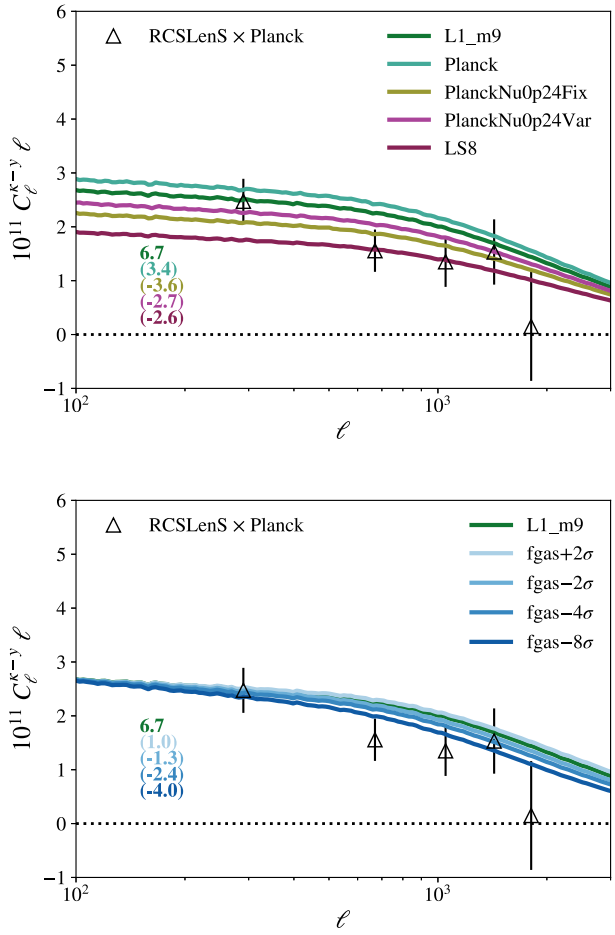

Figure A4. The cosmic shear–tSZ effect angular cross-power spectrum. The open triangles correspond to the RCSLenS × Planck tSZ measurements of Hojjati et al. (2017). Top: Dependence on cosmology. The solid coloured curves correspond to the predicted spectra for the FLAMINGO simulations for different background cosmologies. Bottom: Dependence on baryon physics, namely variations in the gas fractions of groups and clusters, which are mediated primarily through variations in the AGN feedback strength. The solid coloured curves correspond to the predicted spectra for the FLAMINGO simulations as the gas fractions are varied from $+2\sigma$ to -8σ with respect to the observed gas fraction–halo mass relation.

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