Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **529,** 1862–1876 (2024) https://doi.org/10.1093/mnras/stae663
Advance Access publication 2024 March 7

Non-linear CMB lensing with neutrinos and baryons: FLAMINGO simulations versus fast approximations

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

[1]*Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK*
[2]*Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK*
[3]*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*
[4]*Lorentz Institute for Theoretical Physics, Leiden University, PO Box 9506, NL-2300 RA Leiden, the Netherlands*

Accepted 2024 March 2. Received 2024 February 29; in original form 2023 August 28

ABSTRACT

Weak lensing of the cosmic microwave background is rapidly emerging as a powerful probe of neutrinos, dark energy, and new physics. We present a fast computation of the non-linear CMB lensing power spectrum that combines non-linear perturbation theory at early times with power spectrum emulation using cosmological simulations at late times. Comparing our calculation with light-cones from the FLAMINGO 5.6 Gpc cube dark-matter-only simulation, we confirm its accuracy to 1 per cent (2 per cent) up to multipoles $L = 3000$ ($L = 5000$) for a $\nu\Lambda$CDM cosmology consistent with current data. Clustering suppression due to small-scale baryonic phenomena such as feedback from active galactic nuclei can reduce the lensing power by ~ 10 per cent. To our perturbation theory and emulator-based calculation, we add `SP(k)`, a new fitting function for this suppression, and confirm its accuracy compared to the FLAMINGO hydrodynamic simulations to 4 per cent at $L = 5000$, with similar accuracy for massive neutrino models. We further demonstrate that scale-dependent suppression due to neutrinos and baryons approximately factorize, implying that a careful treatment of baryonic feedback can limit biasing neutrino mass constraints.

Key words: neutrinos – cosmic background radiation – cosmological parameters – large-scale structure of Universe – cosmology: theory.

1 INTRODUCTION

Cosmological bounds on the sum of neutrino masses, $M_\nu = \sum m_\nu \leq 0.12$ eV, are rapidly converging on the lower bound $M_\nu = 0.06$ eV from laboratory experiments, promising a measurement in the next several years under restrictive assumptions about the cosmological model (Capozzi et al. 2018; de Salas et al. 2018; Aghanim et al. 2020a; Esteban et al. 2020; Palanque-Delabrouille et al. 2020; Alam et al. 2021; Abbott et al. 2022). However, relaxing these assumptions by allowing greater variation in the neutrino sector, dark energy, or scale-dependent bias, or by considering different combinations of probes, can substantially weaken cosmological bounds to $M_\nu \lesssim 0.5$ eV (Upadhye 2019; Di Valentino, Melchiorri & Silk 2020; Sgier et al. 2021; Di Valentino & Melchiorri 2022), while relatively cosmology-independent terrestrial experiments are consistent with up to $M_\nu \sim 2$ eV (Aker et al. 2019, 2022). Meanwhile, tensions in the Hubble expansion rate and small-scale clustering demand a deeper understanding of the interdependent constraints on neutrinos and other components of the standard cosmological model (Leauthaud et al. 2017; Abdalla et al. 2022; Cai et al. 2022; Amon et al. 2023).

Weak gravitational lensing of the cosmic microwave background (CMB), or CMB lensing, is a promising approach to measuring M_ν. It is immune to scale-dependent galaxy biasing, which has the potential to bias M_ν measurements from galaxy redshift surveys. Its source, the CMB, is well-understood, consisting of photons that last scattered over a very narrow range of redshifts around 1100. Additionally, unlensed CMB temperature and polarization perturbations are very accurately characterized using linear perturbation theory applied to adiabatic, nearly scale-invariant Gaussian random density fluctuations. Thus, CMB lensing evades biases due to intrinsic alignments and photometric redshift errors that affect lower redshift weak lensing surveys (Weinberg et al. 2013).

Next-generation CMB surveys are expected to improve upon the signal-to-noise ratio of the Planck survey by over an order of magnitude (Abazajian et al. 2016; Ade et al. 2019; Liu et al. 2022). Aside from mapping the CMB temperature on smaller scales, they will substantially advance our knowledge of the polarized CMB. Systematic uncertainties due to astrophysical foregrounds and atmospheric noise are significantly smaller for the polarization than the temperature, and the lack of a small-scale primordial B mode polarization reduces the uncertainty due to cosmic variance. Thus the next generation of experiments will be able to quantify CMB lensing to Legendre moments L of a few thousand, or angles below 10 arcmin.

★ E-mail: a.upadhye@ynu.edu.cn

Furthermore, whereas current data have significant uncertainties $\gtrsim 30$ per cent in the non-linear regime $L \gtrsim 1000$, CMB polarization measurements will allow Stage-IV surveys to make precision observations of CMB lensing. Abazajian et al. (2016) show that, in their Stage-IV forecast, the noise per mode of the lensing power spectrum from polarization measurements remains below unity past $L = 1000$, resulting in binned constraints at the 2 per cent accuracy level at $L \approx 2000$ and the 5 per cent level at $L \approx 3000$. Thus theoretical predictions must be at least this accurate. A further consideration is that a change of 0.06 eV in M_ν is associated with an ≈ 2 per cent reduction in the lensing power in the neutrino free-streaming regime, L greater than a few hundred, motivating a 1–2 per cent-level target theoretical accuracy. We demonstrate a CMB lensing computation capable of reaching this target, improving by a few percentage points at small scales relative to the commonly used `Halofit` computation.

Precision prediction of CMB lensing at these scales requires an understanding of non-linear corrections to the clustering of matter. Several approaches to non-linear clustering have been explored in recent years. Non-linear perturbation theory typically begins with the continuity and Euler equations of fluid dynamics, whose non-linear terms couple different Fourier modes together; see Crocce & Scoccimarro (2006), McDonald (2007), Taruya & Hiramatsu (2008), Matsubara (2008a, b), Pietroni (2008), and Lesgourgues et al. (2009). We focus here on the Time-Renormalization Group (Time-RG) perturbation theory of Pietroni (2008) and Lesgourgues et al. (2009), designed for massive-neutrino cosmologies with scale-dependent clustering growth, as implemented in the REDTIME code of Upadhye (2019). Another approach to non-linear corrections begins with the halo model of clustering, detailed in Ma & Fry (2000), Seljak (2000), and Cooray & Sheth (2002), tuned or supplemented by fitting functions to agree with large computer simulations, as in Smith et al. (2003), Bird, Viel & Haehnelt (2012), Takahashi et al. (2012), and Mead et al. (2015, 2020). As the `Halofit` function of Bird et al. (2012) was fit to neutrino simulations, we also consider it here. Finally, the most accurate estimates of non-linear clustering, within limited ranges of parameters and redshifts, are emulators based upon large suites of N-body simulations; see Heitmann et al. (2009), Lawrence et al. (2010), Knabenhans et al. (2021), and Moran et al. (2023). In this study, we use Euclid Emulator 2 of Knabenhans et al. (2021) and the Mira-Titan IV emulator of Moran et al. (2023).

As we are particularly interested in the small-scale ($k \gtrsim 0.1\ h\ \mathrm{Mpc}^{-1}$) suppression of clustering, hence CMB lensing, due to massive neutrinos, we must distinguish this suppression from the small-scale hydrodynamic effects of baryons. Cooling and clumping of baryons leads to the formation of supernovae, which expel baryonic matter from galaxies. Baryonic clustering at the centres of large haloes feeds active galactic nuclei (AGNs), which, in turn, heat the baryonic gas and expel baryonic particles. The combined effect of these phenomena is a suppression of clustering on $\sim$1 Mpc scales, while hydrodynamic models typically predict enhanced clustering on much smaller scales. We model hydrodynamic effects through an innovative fitting function, `SP(k)`, by Salcido et al. (2023), which uses the fact that the total hydrodynamic suppression is strongly correlated with the baryonic content of haloes of a characteristic mass;[1] see also van Daalen, McCarthy & Schaye (2020) and Pandey et al. (2023). We briefly explore a one-parameter generalization of `SP(k)`, showing that it covers a wide range of hydrodynamic models.

[1] `SP(k)` is publicly available at github.com/jemme07/pyspk.

Our goal in this work is a fast, accurate computation of CMB lensing in the non-linear regime, quantified by the power spectrum of the lensing potential ϕ, whose gradient determines the angle by which a CMB photon is deflected. We combine the Mira-Titan IV emulator at low redshifts with Time-RG perturbation theory at high redshifts to yield a rapid computation of the matter power spectrum that is accurate at the times and length-scales necessary for quantifying CMB lensing. Baryonic feedback effects are modelled using `SP(k)`. Running in under a second, our calculation converges to sub- per cent-level precision for Legendre moments $L \leq 10\,000$. We release our code, HYPHI, publicly at github.com/upadhye/hyphi.

Rigorously quantifying the errors in HYPHI requires a set of high-resolution, large-volume numerical simulations. The `FLAMINGO` simulation suite of Schaye et al. (2023) and Kugel et al. (2023) is an ideal testing ground for HYPHI CMB lensing calculations. With cosmological parameters chosen to match data from CMB probes and galaxy surveys, its CMB lensing power spectrum closely matches state-of-the-art measurements, as we show below. `FLAMINGO` includes the largest-particle-number hydrodynamic simulation reaching $z = 0$, which is necessary for covering the range of redshifts contributing the most to CMB lensing, and its relatively high-mass resolution of $7 \times 10^9\ \mathrm{M}_\odot$ provides a wealth of information on the impact of non-linear clustering. Further, the `FLAMINGO` suite independently varies the neutrino masses and hydrodynamic feedback, both of which suppress small-scale clustering, allowing us to investigate their separate effects on lensing.

We find close agreement between HYPHI and the `FLAMINGO` simulations across a wide range of neutrino masses, $0.06\ \mathrm{eV} \leq M_\nu \leq 0.24$ eV; source redshift bins spanning $0 \leq z \leq 25$; and a variety of feedback models. In particular, comparison with the `FLAMINGO` $M_\nu = 0.06$ eV 5.6 Gpc-box dark-matter-only (DMO) simulation demonstrates the accuracy of HYPHI to 1 per cent up to $L = 3000$ and 2 per cent up to $L = 5000$. Dividing the range $0 \leq z \leq 25$ into eight source mass bins, we find 5 per cent agreement to at least $L = 4000$ for each bin. `FLAMINGO` simulations with larger M_ν and hydrodynamic feedback are run in smaller 1 Gpc boxes, meaning that sample variance contributes more to the discrepancy between HYPHI and simulations, but even so, the two agree to 4 per cent up to $L = 4000$ in a model with $M_\nu = 0.24$ eV and the standard hydrodynamic feedback. Furthermore, we demonstrate that neutrino and baryonic suppression of the lensing power factorize, facilitating a marginalization over feedback parameters to constrain M_ν. We show that this result extends to a wide variety of feedback models, including those with jets, as well as those reducing the cluster gas fraction well below its best-fitting value.

This article is organized as follows. Section 2 provides overviews of CMB lensing, the `FLAMINGO` suite of simulations, and our modelling of baryonic feedback. DMO models are considered in Section 3, which studies the dependence of CMB lensing on the matter clustering, quantifies its suppression by massive neutrinos, and compares our fast computation to `FLAMINGO`. Baryonic effects are included in Section 4, which demonstrates that these two types of scale-dependent suppression factorize. Finally, Section 5 shows our conclusions.

2 BACKGROUND

2.1 CMB lensing

For a thorough, modern review of CMB lensing, see Lewis & Challinor (2006), which we briefly outline here. Lensing may be understood as deflecting a light ray from the surface of last scattering

as it passes through density perturbations along our past light-cone. The result is that a ray incident on a detector at angle $\hat{n}$ initially came from an angular position $\hat{n} + \vec{\epsilon}$ on the sky. CMB temperature perturbations $\Theta(\hat{n}) := \delta T(\hat{n})/\bar{T}$ on the last-scattering surface therefore are distorted by lensing to the observed perturbations $\tilde{\Theta}(\hat{n}) = \Theta(\hat{n} + \vec{\epsilon})$.

Qualitatively, the effects of lensing on the CMB temperature and polarization power spectra split into three regimes, defined by large, intermediate, and small scales. On large scales, with Legendre moments $L \ll 300$, lensing has only a small effect on perturbations. Acoustic peaks on intermediate scales $L \sim 1000$ are smeared out by lensing. On small scales, $L \gg 3000$, where diffusion damping sharply reduces the power of the unlensed CMB perturbations, lensing and other secondary anisotropies dominate the power.

At each position along a photon's path to us, its deflection is proportional to the gradient of the local gravitational potential. Thus its total deflection is the gradient of the line-of-sight integral of the gravitational potential, weighted by a lensing kernel g defined below. This integral, known as the lensing potential $\phi(\hat{n})$, sources the deflection $\vec{\epsilon} = \vec{\nabla}\phi$.

Gravitational lensing of the CMB may be quantified statistically using the power spectrum $C_L^{\phi\phi}$ of ϕ. In the Limber approximation of Limber (1953), appropriate to the scales of interest to us here, this is

$$C_L^{\phi\phi} \approx \int_0^{\chi_\star} d\chi \; P_\Phi\left(\frac{L}{\chi}, z(\chi)\right) \frac{g(\chi)^2}{\chi^6} \tag{1}$$

$$P_\Phi(k,z) = \left[\frac{3\Omega_{\rm m}(z)\mathcal{H}(z)^2}{2k^2}\right]^2 P_{\rm m}(k,z) = \frac{9\Omega_{\rm m,0}^2\mathcal{H}_0^4}{4a^2k^4} P_{\rm m}(k,z) \tag{2}$$

$$g(\chi) = 2\chi(1 - \chi/\chi_\star). \tag{3}$$

Here, P_Φ and $P_{\rm m}$ are, respectively, the power spectra of the gravitational potential and total matter in units of volume; $\chi(z)$ is the comoving distance to redshift z; $\chi_\star = \chi(z_\star)$; and $z_\star$ is the redshift of the baryon drag epoch. The cosmological parameters $\Omega_{\rm m,0}$ and $\mathcal{H}_0$ are the present-day values of the matter density as a fraction of the critical density $\bar{\rho}_{\rm m,0}/\bar{\rho}_{\rm crit,0}$, and the derivative of the logarithm of the scale factor with respect to conformal time $d\log(a)/d\tau$, respectively. The functions $\mathcal{H}(z) = d\log(a)/d\tau$ and $\Omega_{\rm m}(z) = \Omega_{\rm m,0}\mathcal{H}_0^2(1+z)/\mathcal{H}^2$ are the expansion rate and matter fraction at all redshifts.

In turn, the power spectrum of the lensed temperature perturbation $\tilde{\Theta}(\hat{n})$ is found from the appropriately weighted convolution of $C_L^{\phi\phi}$ with the unlensed power $C_\ell^{\Theta\Theta}$:

$$C_\ell^{\tilde{\Theta}\tilde{\Theta}} = \int \frac{d^2\vec{L}}{(2\pi)^2}\left[\vec{L}\cdot\left(\vec{\ell} - \vec{L}\right)\right]^2 C_\ell^{\Theta\Theta} C_L^{\phi\phi} + \left(1 - \ell^2 \int dL\, L^3 C_L^{\phi\phi}\right) C_\ell^{\Theta\Theta}. \tag{4}$$

Here, we work in the flat-sky limit in which $\vec{\ell}$ is the two-dimensional Fourier conjugate to the direction $\hat{n}$. Factors of L and $\ell - L$ arise from gradients of the lensing potential power spectrum. The first term on the right hand side describes the lensing-induced smearing of $C_\ell^{\Theta\Theta}$, while the second represents a smooth suppression of power due to lensing. Lensed CMB polarization power spectra are similarly found by convolution with $C_L^{\phi\phi}$; see Lewis & Challinor (2006).

Thus $C_L^{\phi\phi}$ is necessary for quantifying lensing of the CMB, and is useful for comparing the observed lensing to theoretical predictions. Our goal henceforth is a computation of $C_L^{\phi\phi}$, including non-linear clustering as well as power suppression by massive neutrinos and hydrodynamic effects.

2.2 The FLAMINGO simulations

The FLAMINGO simulations used here are thoroughly described in Schaye et al. (2023) and Kugel et al. (2023). They employed the SWIFT code of Schaller et al. (2023), which includes gravitation, hydrodynamic feedback, and subgrid models for unresolved physics relevant to galaxy formation, including metal-dependent radiative cooling, star formation, stellar evolution, and stellar and AGN feedback. Neutrinos were incorporated into the simulations through the δf method of Elbers et al. (2021), and included in the initial conditions of MONOFONIC, Hahn, Rampf & Uhlemann (2021) and Hahn et al. (2020), as detailed and implemented in Elbers (2022a, b) and Elbers et al. (2022).

Table 1 lists the code and cosmological parameters for the FLAMINGO simulations used in this work. We will use the large-volume run L5p6_m10_DMO to quantify the accuracy of our $C_L^{\phi\phi}$ computation at high redshifts; the $L_{\rm sim} = 2.8$ Gpc runs, L2p8_m9_DMO and L2p8_m9, to study the impact of baryonic feedback; and the several $L_{\rm sim} = 1$ Gpc runs to assess the impact of different neutrino masses and feedback processes on $C_L^{\phi\phi}$.

Light-cones are data products of the FLAMINGO simulations of Schaye et al. (2023). Light-cone maps apply the HEALPIX pixelization of Górski et al. (2005) to spherical shells, recording the total mass contained in each resulting volume element. Shell thicknesses in redshift space, $z_{i+1} - z_i$, are 0.05 up to $z = 3$, and then 0.25 up to $z = 5$, gradually increasing to 5 at $z = 15$, corresponding to homogeneous-universe comoving radii $\chi(z)$ between $\chi_i = \chi(z_i)$ and $\chi_{i+1} = \chi(z_{i+1})$. Each shell is divided into $12N_{\rm side}^2 = 3,221,225,472$ pixels, each of angular size 166 arcsec^2, with the HEALPIX parameter $N_{\rm side} = 16\,384$. For each HEALPIX angular pixel j, the lensing convergence κ_j is computed through a summation discretizing the line-of-sight integral over comoving distance:

$$\kappa_j = \frac{3}{2}\Omega_{\rm m,0}\mathcal{H}_0^2 \sum_i \Delta\chi_i \;\langle\chi_i\rangle(1+z_i)\left(1 - \frac{\langle\chi\rangle_i}{\chi_\star}\right)\frac{M_{ij} - \bar{M}_{ij}}{\bar{M}_{ij}}. \tag{5}$$

Here, $\Delta\chi_i := \chi_{i+1} - \chi_i$; $<\chi>_i$ is the mean comoving distance in redshift shell i; $\chi_\star$ is the comoving distance to the surface of last scattering; M_{ij} is the mass in shell i and HEALPIX pixel j; and $\bar{M}_{ij} = \frac{4\pi}{3}\bar{\rho}_{\rm m,0}(\chi_{i+1}^3 - \chi_i^3)/(12N_{\rm side}^2)$ is the homogenous-universe mass in i and j. We fix $\chi_\star = \chi(z_\star)$, with $z_\star = 1089.80$ as measured by Aghanim et al. (2020a).

Standard power spectrum computation codes such as the POLSPICE code of Szapudi et al. (2000) cannot process more than $2^{31} \approx 2 \times 10^9$ pixels, so we downsample the κ map to $N_{\rm side} = 8192$. Then we use POLSPICE with the `-pixelfile YES` option to compute its power spectrum, which we subsequently multiply by $4/L^4$ to yield the lensing potential power spectrum $C_L^{\phi\phi}$. We confirm by comparison to power spectra from lower resolution maps that our $C_L^{\phi\phi}$ with $N_{\rm side} = 8192$ has converged to better than 1 per cent (2 per cent) up to $L = 5000$ ($L = 6000$), which is sufficient for testing HYPHI.

The lensing potential power spectrum from the $L_{\rm sim} = 5.6$ Gpc FLAMINGO simulation, L5p6_m10_DMO, averaged over light-cones computed from eight different positions in the simulation volume, is compared with several recent observations in Fig. 1. Agreement across two orders of magnitude in the Legendre moment is impressive, demonstrating that the FLAMINGO simulation suite is an appropriate tool for quantifying non-linear CMB lensing. Since we are primarily concerned with neutrino and baryonic suppression effects, which are largest at small scales $L \gtrsim 1000$ where data error bars are large, we henceforth focus on FLAMINGO directly as a means of quantifying these small-scale effects and testing their predictions.

Fig. 2 shows FLAMINGO lensing potential power spectra for

Table 1. List of FLAMINGO simulations, with simulation parameters and cosmological parameters. $n_{\rm LC}$ is the number of high-redshift simulation light-cones produced; these cover the range $0 \leq z \leq 3$ for $L_{\rm sim} = 1000$ Mpc h^{-1}; $0 \leq z \leq 5$ for $L_{\rm sim} = 2800$ Mpc h^{-1}; and $0 \leq z \leq 25$ for $L_{\rm sim} = 5600$ Mpc h^{-1}. For feedback fits and standard deviations, as well as a description of the feedback models realizing the $f_{\rm gas}$ and SMF listed, see Schaye et al. (2023) and Kugel et al. (2023).

Name	$L_{\rm sim}$ (Mpc)	$N_{\rm CDM}^{1/3}$	$N_\nu^{1/3}$	$n_{\rm LC}$	$f_{\rm gas}$	SMF	M_ν (eV)	h	$\Omega_{\rm b,\,0}$	$\Omega_{\rm m,\,0}$	$10^9 A_s$	n_s
L1_m10_DMO	1000	900	500	1	N/A	N/A	0.06	0.681	0.0486	0.306	2.099	0.967
L1_m9_DMO	1000	1800	1000	1	N/A	N/A	0.06	0.681	0.0486	0.306	2.099	0.967
L1_m8_DMO	1000	3600	2000	1	N/A	N/A	0.06	0.681	0.0486	0.306	2.099	0.967
L2p8_m9_DMO	2800	5040	2800	8	N/A	N/A	0.06	0.681	0.0486	0.306	2.099	0.967
L5p6_m10_DMO	5600	5040	2800	8	N/A	N/A	0.06	0.681	0.0486	0.306	2.099	0.967
L11p2_m11_DMO	11 200	5040	2800	8	N/A	N/A	0.06	0.681	0.0486	0.306	2.099	0.967
Planck_DMO	1000	1800	1000	1	N/A	N/A	0.06	0.673	0.0494	0.316	2.101	0.966
PlanckNu0p12Var_DMO	1000	1800	1000	1	N/A	N/A	0.12	0.673	0.0496	0.316	2.113	0.967
PlanckNu0p24Fix_DMO	1000	1800	1000	1	N/A	N/A	0.24	0.673	0.0494	0.316	2.101	0.966
L1_m9	1000	1800	1000	1	fit	fit	0.06	0.681	0.0486	0.306	2.099	0.967
L1_m8	1000	3600	2000	1	fit	fit	0.06	0.681	0.0486	0.306	2.099	0.967
L2p8_m9	2800	5040	2800	8	fit	fit	0.06	0.681	0.0486	0.306	2.099	0.967
Planck	1000	1800	1000	1	fit	fit	0.06	0.673	0.0494	0.316	2.101	0.966
PlanckNu0p24Fix	1000	1800	1000	1	fit	fit	0.24	0.673	0.0494	0.316	2.101	0.966
fgas + 2σ	1000	1800	1000	1	fit+2σ	fit	0.06	0.681	0.0486	0.306	2.099	0.967
fgas-2σ	1000	1800	1000	1	fit−2σ	fit	0.06	0.681	0.0486	0.306	2.099	0.967
fgas-4σ	1000	1800	1000	1	fit−4σ	fit	0.06	0.681	0.0486	0.306	2.099	0.967
fgas-8σ	1000	1800	1000	1	fit−8σ	fit	0.06	0.681	0.0486	0.306	2.099	0.967
Jet	1000	1800	1000	1	fit[a]	fit[a]	0.06	0.681	0.0486	0.306	2.099	0.967
Jet_fgas-4σ	1000	1800	1000	1	fit−4σ[a]	fit[a]	0.06	0.681	0.0486	0.306	2.099	0.967
$M* - \sigma$	1000	1800	1000	1	fit	fit−1σ	0.06	0.681	0.0486	0.306	2.099	0.967
$M* - \sigma$_fgas-4σ	1000	1800	1000	1	fit−4σ	fit−1σ	0.06	0.681	0.0486	0.306	2.099	0.967

Note. [a]Jet and Jet_fgas-4σ realize the fit $f_{\rm gas}$ and SMF listed using momentum jets rather than thermal energy injection, as detailed in Schaye et al. (2023).

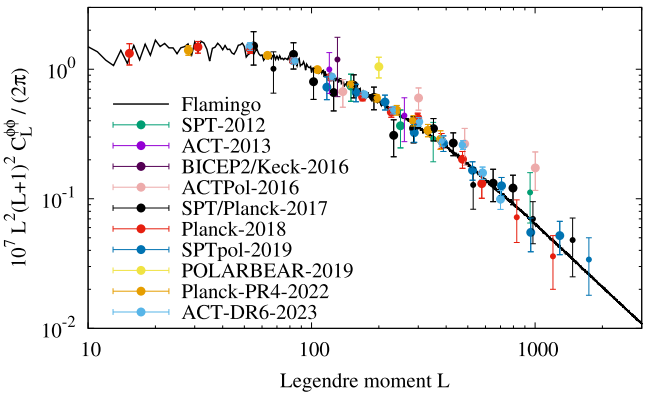

Figure 1. Lensing potential power spectrum from FLAMINGO model L5p6_m10_DMO of Table 1, integrated up to $z = 25$ and averaged over eight light-cones, compared with measurements of van Engelen et al. (2012), Das et al. (2014), Ade et al. (2016), Sherwin et al. (2017), Simard et al. (2018), Aghanim et al. (2020b), Wu et al. (2019), Aguilar Faúndez et al. (2020), Carron, Mirmelstein & Lewis (2022), and Qu et al. (2023). Small (large) data points show detections to 2σ ($\geq 3\sigma$).

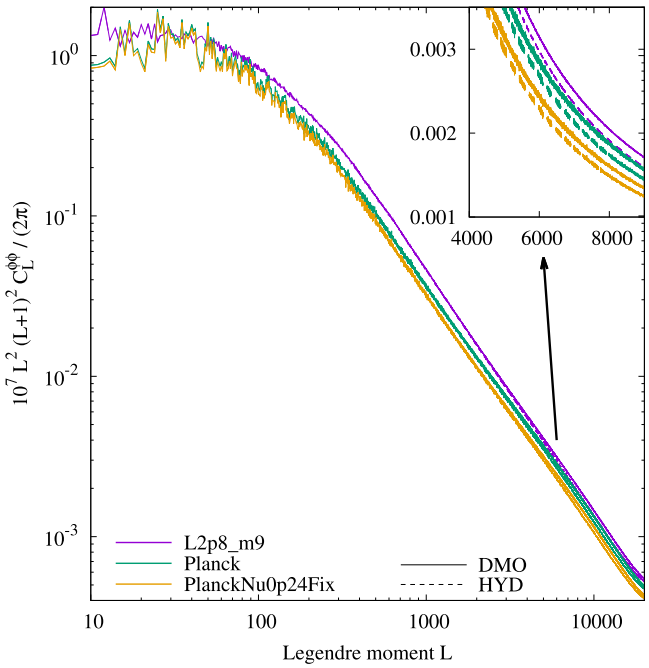

Figure 2. $C_L^{\phi\phi}$ from FLAMINGO L2p8_m9, Planck, and PlanckNu0p24Fix simulations (dashed) and their corresponding DMO simulations (solid) computed over the range of redshifts given in Table 1. (Inset) Baryonic suppression is evident at small scales relevant for next-generation probes.

several models. Although neutrinos and baryons both suppress $C_L^{\phi\phi}$ by ~ 10 per cent, these effects depend very differently upon scale. The neutrino suppression extends down to $L \approx 100$, while the baryonic suppression is visible only beyond $L = 1000$. Quantifying the difference between these effects is a major goal of this article.

The other obvious difference between L2p8_m9 and the Planck-like models, Planck and PlanckNu0p24Fix, in Fig. 2, is in their large-scale power. This difference is due to the fact the Planck-like models are simulated in smaller boxes whose light-cones are limited to $z \leq 3$, so our $C_L^{\phi\phi}$ computation based upon equation (5) is limited to that redshift range. A direct comparison of these power spectra to the data would first require some approximation of the higher z contribution to $C_L^{\phi\phi}$, such as the use of the FLAMINGO matter power spectra in the line-of-sight integral of equations (1)–(3). However,

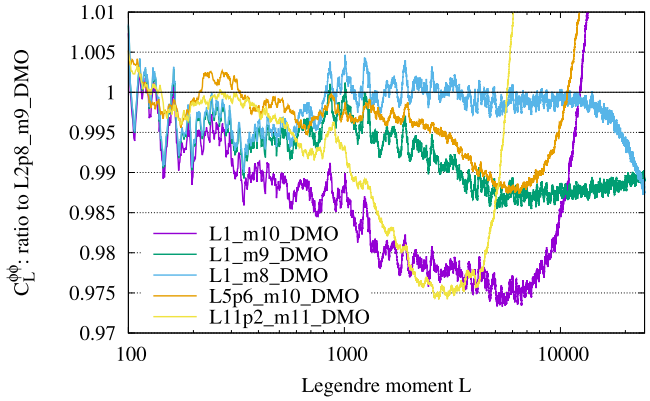

Figure 3. Ratios of $C_L^{\phi\phi}$ computed from the L1_m10_DMO, L1_m9_DMO, L1_m8_DMO, L5p6_m10_DMO, and L11p2_m11_DMO simulations to that from L2p8_m9_DMO, smoothed using a centred 100-point moving average. The higher resolution runs L1_m9_DMO, L1_m8_DMO, L2p8_m9_DMO, and L5p6_m10_DMO, with multiple box sizes, agree to 1 per cent up to $L = 3000$ and < 1.5 per cent beyond $L = 20\,000$.

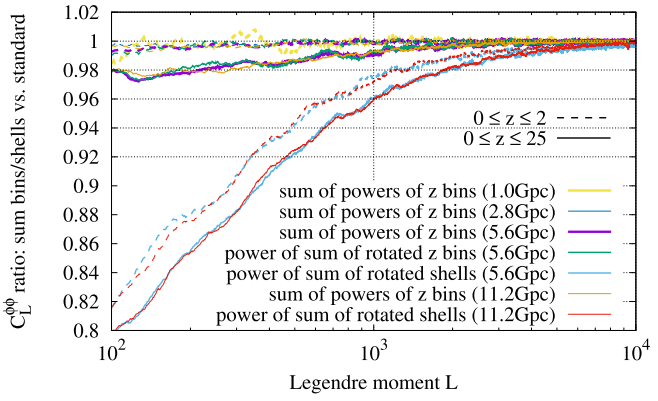

Figure 4. Effects of binning and bin/shell rotation on $C_L^{\phi\phi}$. For L5p6_m10_DMO, with a 5.6 Gpc box, methods (ii) (sum of powers of z bins), (iii) (power of sum of rotated z bins), and (iv) (power of sum of rotated z shells) are shown divided by the standard method (i), computed over the region $z \le 2$ (dashed) and $z \le 25$ (solid). Also shown are: methods (ii) and (iv) divided by (i) for L11p2_m11_DMO, whose large simulation volume makes it immune to volume replication errors; and method (ii) divided by (i) for L1_m9_DMO and L2p8_m9_DMO, whose small volumes make them especially prone to such errors.

since our goal is to test fast approximations using the light-cones, we take the opposite approach, and limit our perturbative and emulated computations to the redshift range covered by each light-cone.

One strength of the FLAMINGO simulation suite is that it contains multiple resolutions and box sizes for the same cosmology, allowing for convergence tests. Fig. 3 compares $C_L^{\phi\phi}$ integrated up to $z = 3$, computed from six different FLAMINGO simulations, with box sizes ranging from 1 to 11.2 Gpc and mean interparticle spacings ranging from 0.27 to 2.22 Mpc. Although the smaller box simulations might be expected to underpredict power on large scales, we see that this effect is under a per cent by $L = 100$. Similarly, although the simulation with the largest interparticle spacing, L11p2_m11_DMO, has significant small-scale errors, for L5p6_m10_DMO we see that these only appear for $L > 10\,000$. The runs L1_m9_DMO, L1_m8_DMO, L2p8_m9_DMO, and L5p6_m10_DMO all agree to 1 per cent up to $L = 3000$ and < 1.5 per cent beyond $L = 20\,000$, so we may use any of them for testing HYPHI. In particular, FLAMINGO simulations varying the neutrino mass and the hydrodynamic feedback parameters have the same resolution as L1_m9_DMO, so Fig. 3 confirms their suitability for our purposes.

We consider one further issue related to FLAMINGO light-cones, that of simulation volume replication, with L5p6_m10_DMO and L11p2_m11_DMO as particular examples. The radius of a light-cone extending to $z = 25$ exceeds the simulation box size $L_{\rm sim}$ for L5p6_m10_DMO, but not for L11p2_m11_DMO. Light-cones are generated by replicating the simulation volume consistently with its periodic boundary conditions. A photon propagating from $z = 25$ to $z = 0$ in L5p6_m10_DMO could pass through the same structures twice, and different portions of a large-z shell could see the same structures, introducing spurious correlations into our κ map. We consider several methods of mitigating such volume replication effects. We compute multiple κ maps, one for each z shell, and then combine them using one of four methods to obtain $C_L^{\phi\phi}$.

(i) Sum all κ maps along the entire line of sight as in equation (5), then compute $C_L^{\phi\phi}$ from the resulting κ map. This is our standard $C_L^{\phi\phi}$ computation, used in subsequent figures unless noted otherwise.

(ii) Sum the κ maps within redshift bins $z_i \le z \le z_{i+1}$, with the bin widths $\chi(z_{i+1}) - \chi(z_i)$ chosen to be of the order of the simulation box size. In practice, we choose $z_0 = 0$, $z_1 = 1$, $z_2 = 2$, $z_3 = 3$, $z_4 = 4$, $z_5 = 5$, $z_6 = 6$, $z_7 = 7.25$, $z_8 = 9.5$, $z_9 = 12.25$, $z_{10} = 15$, $z_{11} = 20$, and $z_{12} = 25$. We compute a separate κ map for each bin, and a $C_L^{\phi\phi}$ from each κ map, before summing to obtain the total $C_L^{\phi\phi}$.

(iii) Compute a κ map for each of the above z bins, then rotate each κ by a random angle. The rotated z-binned κ maps are then summed and the result used to compute $C_L^{\phi\phi}$.

(iv) Compute a κ map for each redshift shell, with $\Delta z = 0.05$ up to $z = 3$, and then randomly rotate each one. The rotated κ maps are summed and the result used to compute $C_L^{\phi\phi}$.

Using ρ to represent a random rotation, B a redshift binning, Σ a summation over bins or shells, and P the computation of an autopower spectrum, we may represent these as: (i) $P(\Sigma\kappa)$; (ii) $\Sigma P(B(\kappa))$; (iii) $P(\Sigma(\rho(B(\kappa))))$; and (iv) $P(\Sigma\ \rho(\kappa))$. Further, since the light-cone radius up to $z = 2$ is less than the simulation box size for L5p6_m10_DMO, we also compute $C_L^{\phi\phi}$ integrating only up to $z = 2$. This calculation, along with both $z \le 2$ and $z \le 25$ light-cones for L11p2_m11_DMO, are thus immune to volume replication effects. For further comparison, we also compute $C_L^{\phi\phi}$ up to $z = 2$ for the 1.0 Gpc-box L1_m9_DMO and the 2.8 Gpc-box L2p8_m9_DMO, which ought to be affected by volume replication.

Fig. 4 compares the second, third, and fourth methods to the first. Immediately apparent is the fact that the two largest box sizes agree to < 1 per cent at all L, even though the L11p2_m11_DMO simulation does not suffer from volume replication errors. Thus we see that these errors are negligible for our purposes, at least within the context of our power spectrum ratios smoothed with centred 100-point moving averages. This conclusion is strengthened over the $0 \le z \le 2$ range by comparing the two smaller box runs, with volume-replication errors, to the larger box runs without them; the two sets are again identical at the per cent level. Also apparent is the close agreement between methods (ii) and (iii), showing that bin rotation and separate-bin power spectrum computation decorrelate different z bins in very similar ways.

Next, consider small scales, $L \ge 1000$, which are most relevant to tests of the non-linear $C_L^{\phi\phi}$. The sum-of-bins calculation, method (ii), and the sum-of-rotated-bins calculation, method (iii), agree with the standard method (i) to 1 per cent across this entire range, for

all simulations and both light-cone sizes. Thus any of these three methods is sufficiently accurate for per cent-level tests of the non-linear $C_L^{\phi\phi}$. Meanwhile, method (iv) exhibits significant power loss, and should not be used for testing HYPHI.

Finally, consider $L < 1000$ in Fig. 4. Evidently, breaking the redshift range into a greater number of intervals, and rotating the resulting κ maps, leads to a greater power suppression; method (iv) uses 80 redshift shells, while methods (ii) and (iii) use 12 redshift bins each. Furthermore, the presence of this suppression even in the L11p2_m11_DMO power spectra and the $z \leq 2$ L5p6_m10_DMO power spectrum, for which the light-cone radii are less than $L_{\rm sim}$, shows that it is not due to volume replication effects.

The oscillatory nature of this low-L power suppression for method (iv), as well as its L-dependence, suggests that it is due to the exclusion of baryon acoustic oscillations (BAO) along the line of sight. Since $\chi_\star/2$ is the peak of the lensing kernel of equation (3), the Legendre moment L is approximately associated with the length-scale $\chi_\star/(2L) \sim 5000L^{-1}$ Mpc/h, which for $L \sim 50$ corresponds to the BAO scale. Thus, not only is map rotation unnecessary for suppressing volume replication effects, but excessive rotation erroneously throws out actual large-scale correlations which ought to be included in the power spectrum.

We conclude that κ maps from FLAMINGO light-cones, computed using method (i) above for the total power and methods (ii) or (iii) for tomographic redshift bins, are suitable for testing calculations of $C_L^{\phi\phi}$ in the non-linear regime $L \gtrsim 1000$ at the per cent level. At larger scales, $L \sim 100$, the redshift-binned methods (ii) and (iii) underpredict power by 2 per cent–3 per cent, an error of which we should be aware when testing the application of HYPHI to tomographic bins.

2.3 Baryonic feedback

Stellar and AGN feedback in the FLAMINGO simulations was calibrated using machine learning to two observables: the galaxy stellar mass function (SMF) at $z = 0$, measured in Driver et al. (2022); and cluster gas fractions, $f_{\rm gas}$, measured using X-ray and weak lensing observations, as compiled in Kugel et al. (2023). The SMF constrains the galaxy–halo connection, while $f_{\rm gas}$ correlates with the magnitude of the clustering suppression due to baryonic feedback (van Daalen et al. 2020). This approach allowed FLAMINGO to explore the parameter space of feedback models in an observationally relevant manner. For example, the high-gas-fraction model fgas + 2σ in Table 1 increased $f_{\rm gas}$ by twice the observational uncertainty, while the low-gas-fraction model fgas-2σ decreased $f_{\rm gas}$ by the same amount. For a complete description of feedback in FLAMINGO, see Schaye et al. (2023) and Kugel et al. (2023).

Fig. 5 demonstrates the impact of neglecting baryonic feedback above a redshift $z_{\rm b}$ on the lensing potential power spectrum. The figure compares FLAMINGO runs L2p8_m9_DMO and L2p8_m9 from Table 1, which are, respectively, DMO and hydrodynamic runs in 2.8 Gpc boxes. Neglecting feedback at all redshifts ($z_{\rm b} = 0$) leads to a maximum error of 8 per cent, which is reduced to 4 per cent (0.4 per cent) by including feedback for $z_{\rm b} \leq 1$ ($z_{\rm b} \leq 2$). Thus, per cent-level accuracy in the computation of $C_L^{\phi\phi}$ up to $L \gtrsim 1000$ requires the inclusion of baryonic feedback at least at low redshifts, $z \lesssim 2$.

Effects of baryons upon the clustering of matter have been investigated using hydrodynamic simulations by van Daalen et al. (2011, 2014), van Daalen et al. (2020), Hellwing et al. (2016), McCarthy et al. (2018), Springel et al. (2018), Chisari et al. (2018), and Pakmor et al. (2022), while their impact upon CMB lensing

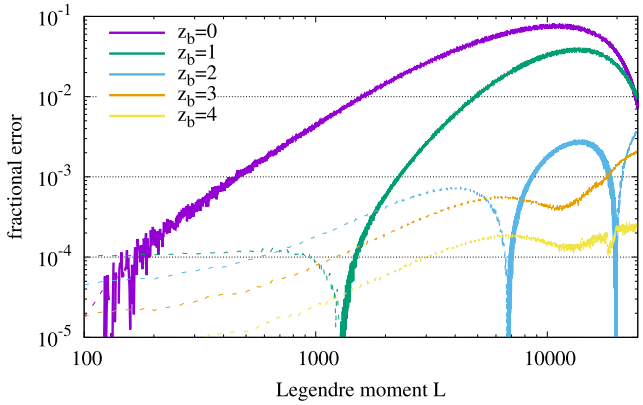

Figure 5. Impact of neglecting hydrodynamic suppression of the CMB lensing potential power spectrum for redshifts $z \geq z_{\rm b}$. In the $L \leq 25\,000$ range, $z_{\rm b} = 0$, 1, and 2 lead to errors of 8 per cent, 4 per cent, and 0.4 per cent, respectively, in the fiducial FLAMINGO hydrodynamic model L2p8_m9.

was considered by Chung, Foreman & van Engelen (2020) and McCarthy, Hill & Madhavacheril (2022). Recently, Salcido et al. (2023; see also Semboloni et al. 2011; van Daalen et al. 2020; Debackere, Schaye & Hoekstra 2021) demonstrated that, for the purpose of determining hydrodynamic suppression of the matter power spectrum, these phenomena may be reduced to their effect on the characteristic relative baryon fractions $\hat{f}_{\rm b}(z) = f_{\rm b}(\hat{M}, z)/(\Omega_{\rm b}/\Omega_{\rm m})$ of haloes of characteristic mass $\hat{M}(z)$ at redshift z. Given $\hat{f}_{\rm b}$, whether measured from observations or computed through simulations, the hydrodynamic suppression factor $P_{\rm m,\,hyd}/P_{\rm m,\,dmo}$ is fit to 2 per cent accuracy in the range $k \leq 12\ h$ Mpc^{-1} and $z \leq 3$ by the SP(k) function of that reference.

Given the full functional form of $\tilde{f}_{\rm b}(M, z)$, we could in principle compute the baryonic suppression directly using a halo model. However, this function is difficult to measure. Salcido et al. (2023) showed that, for each z, the k-dependent baryonic suppression is most strongly correlated with the baryon fraction at a single characteristic mass $\hat{M}(z)$. This correlation was shown to be independent of the strength of the subgrid feedback.

Our calculations in Section 4 will use $\tilde{f}_{\rm b}(M, z)$ measured directly from the FLAMINGO L1_m9 simulations, evaluated at the $\hat{M}(z)$ of Salcido et al. (2023). Qualitatively, $\tilde{f}_{\rm b}$ is large for haloes at high redshifts, before structures such as AGNs have formed; for haloes too small to form such structures; and for the largest haloes, which efficiently capture baryons, and are representative of the universe as a whole. The SP(k) fit covers the range $k \leq 12\ h$ Mpc^{-1} and $z \leq 3$ necessary for ≈ 2 per cent-level accuracy in $C_L^{\phi\phi}$. Above $z = 3$, we assume that the power spectrum suppression due to baryons is negligible, an approximation consistent with Fig. 5. For $k > 12\ h$ Mpc^{-1}, we again assume a negligible suppression, an approximation which we confirm to have a sub-per cent-level impact on $C_L^{\phi\phi}$ up to $L = 9000$.

3 NON-LINEAR CMB LENSING

3.1 Non-linear perturbation theory

In order to calculate the non-linear perturbative CDM + baryon (CB) power spectrum in the presence of massive neutrinos, we employ the Time-RG perturbation theory of Pietroni (2008) and Lesgourgues et al. (2009). Time-RG integrates the non-linear continuity and Euler equations of fluid dynamics over time for each wavenumber,

making it well-suited to massive neutrinos, which introduce a scale dependence into the growth factor, and to dark energy models with evolving equations of state.

Let $\eta = \ln(a/a_{\rm in})$ for some initial scale factor $a_{\rm in}$, and let primes denote derivatives with respect to η. Let the perturbation indices 0 and 1 correspond, respectively, to the density contrast $\delta = (\rho_{\rm CB} - \bar{\rho}_{\rm CB})/\bar{\rho}_{\rm CB}$ and the velocity divergence $\theta = -\vec{\nabla}\cdot\vec{v}/\mathcal{H}$; thus, for example, $P_{01}(k)$ represents $P_{\delta\theta}(k)$. Then at each k, the Time-RG equations of motion for the CB power spectra P_{ab} are

$$P'_{ab} = -\Xi_{ac}P_{cb} - \Xi_{bc}P_{ac} + I_{acd,bcd} + I_{bcd,acd}, \tag{6}$$

where

$$[\Xi_{ab}] = \begin{bmatrix} 0 & -1 \\ \dfrac{k^2\Phi}{\mathcal{H}^2\delta} & 1 + \dfrac{\mathcal{H}'}{\mathcal{H}} \end{bmatrix} \tag{7}$$

$$I_{acd,bef} = \int_q \gamma_{acd}^{\vec{k},\vec{q},\vec{p}} B_{bef}^{\vec{k},\vec{q},\vec{p}} \tag{8}$$

$$\gamma_{001}^{\vec{k},\vec{q},\vec{p}} = \gamma_{010}^{\vec{k},\vec{p},\vec{q}} = \frac{(\vec{q}+\vec{p})\cdot\vec{p}}{2p^2} \text{ and } \gamma_{acd}^{\vec{k},\vec{q},\vec{p}} = \frac{(\vec{q}+\vec{p})^2\vec{q}\cdot\vec{p}}{2q^2p^2} \tag{9}$$

$$I'_{acd,bef} = -\Xi_{bg}I_{acd,gef} - \Xi_{eg}I_{acd,bgf} - \Xi_{fg}I_{acd,beg} + 2A_{acd,bef} \tag{10}$$

$$A_{acd,bef} = \int_q \gamma_{acd}^{\vec{k},\vec{q},\vec{p}} \left[\gamma_{bgh}^{\vec{k},\vec{q},\vec{p}} P_{ge}^q P_{hf}^p + \gamma_{egh}^{\vec{q},-\vec{p},\vec{k}} P_{gf}^p P_{hb}^k + \gamma_{fgh}^{\vec{p},\vec{k},-\vec{q}} P_{gb}^k P_{he}^q\right] \tag{11}$$

with all other γ_{abc} zero, and summation over repeated indices implicit. Here, wavenumber superscripts denote functional dependence, so that $P_{ab}^{\vec{k}}$ denotes the power spectrum $P_{ab}(\vec{k})$ and $B_{abc}^{\vec{k},\vec{q},\vec{p}}$ the bispectrum $B_{abc}(\vec{k},\vec{q},\vec{p})$. We use $\int_q X^{\vec{k},\vec{q},\vec{p}}$ as shorthand for $\int \frac{d^3q}{(2\pi)^3}\frac{d^3p}{(2\pi)^3}(2\pi)^3\delta^{(\rm D)}(\vec{k}-\vec{q}-\vec{p})X(\vec{k},\vec{q},\vec{p})$ for any function $X(\vec{k},\vec{q},\vec{p})$, where $\delta^{(\rm D)}$ is the three-dimensional Dirac delta function. In Time-RG, the bipsectrum integrals $I_{acd,\,bef}$, initialized to zero, are the repositories of non-linear information, and are sourced by the mode-coupling integral $A_{acd,\,bef}$.

We may add to this system an evolution equation for the lensing potential power spectrum of equation (1). Substituting $k = L/\chi$, we recast the line-of-sight integral as an integral over our time coordinate η of

$$\frac{dC_L^{\phi\phi}}{d\eta} = \frac{9\Omega_{\rm m,0}^2\mathcal{H}_0^4}{4a^2L^4\mathcal{H}}\frac{g(\chi)^2}{\chi^2}P_{\rm m}\left(\frac{L}{\chi},z\right). \tag{12}$$

where a, z, $\mathcal{H}$, and χ are now functions of η. Since $P_{\rm cb}$, hence $P_{\rm m}$, is available at each time-step in our Time-RG integration, $C_L^{\phi\phi}$ may simply be added to our system of equations (6)–(11) with very little additional computational cost. Furthermore, if $P_{\rm m}$ contains a hydrodynamic suppression factor parametrized by some variables, we can compute a separate $C_L^{\phi\phi}$ for each of several choices of variables all at once, again with little additional computation.

3.2 Non-linear lensing power

Next, we use non-linear perturbation theory to consider the dependence of $C_L^{\phi\phi}$ upon clustering in different redshift ranges. The contribution of $z_{\rm min} \le z \le z_{\rm max}$ is found by integrating equation (12) from $\eta(z_{\rm max})$ to $\eta(z_{\rm min})$, where $\eta(z) = -\ln[(1+z)a_{\rm in}]$. We use $P_{\rm m} = (f_{\rm cb}P_{00}^{1/2} + f_\nu P_\nu^{1/2})^2$, as in Saito, Takada & Taruya (2008), where P_{00} is the Time-RG power spectrum of equation (6). We shall see in the remainder of this section that Time-RG is sufficiently accurate to provide a qualitative picture of the contributions of different redshifts.

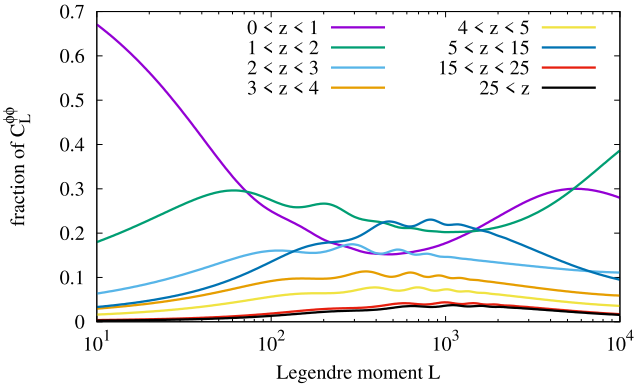

Figure 6. Fractional contributions to the total lensing potential power spectrum $C_L^{\phi\phi}$ due to clustering in different redshift bins. Low Ls are dominated by low-z clustering, while $L \sim 1000$ receives significant contributions from a range of redshifts.

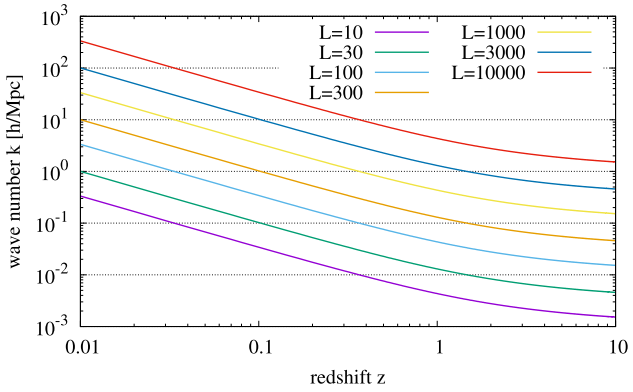

Figure 7. Mapping between L and $k = L/\chi(z)$ for a range of z.

Fig. 6 shows the contributions of several redshift bins to the total lensing potential power spectrum. The domination of the low-L lensing power by low-z clustering is explained by Fig. 7. Matter power spectra $P_{\rm m}(k, z)$ typically peak around $k = 0.01\ h\ {\rm Mpc}^{-1}$, a scale which contributes to $C_L^{\phi\phi}$ for $L = 10$ at $z \approx 0.5$, and for $L = 30$ at $z \approx 2$. Interestingly, at $L \sim 1000$, the total contribution from $z > 5$ in Fig. 6 is ≈ 30 per cent, since the lower z clustering contributing to this L is drawn from increasingly large k, for which the power spectrum rapidly declines.

Heat maps quantifying the fractional contributions of different wavenumbers and scale factors to $C_L^{\phi\phi}$ are shown in Fig. 8, which divides $dC_L^{\phi\phi}/d\eta$ from equation (12) by the final $a = 1$ lensing potential power. For $L = 1000$ (top) and $L = 3000$ (bottom), the contributions peak at (a, k) of $(0.60, 0.26\ h\ {\rm Mpc}^{-1})$ and $(0.79, 0.84\ h\ {\rm Mpc}^{-1})$, respectively, and are significant within factors of $\sim$3 of these values. Thus, the accurate computation of $C_L^{\phi\phi}$ up to L of several thousand requires reliable estimates of the matter power spectrum for $z \lesssim 5$ and $k \lesssim 2\ h\ {\rm Mpc}^{-1}$.

Since our focus is non-linear lensing at high L, we must also quantify the contribution of non-linear clustering to $C_L^{\phi\phi}$, which we may do by substituting the difference between non-linear and linear matter power spectra for $P_{\rm m}$ in equation (12). Fig. 9 shows the result. Across the entire range $L \le 10\,000$ considered, the total non-linear contribution for all $z > 3$ is ≤ 7 per cent of $C_L^{\phi\phi}$, while that for all $z > 5$ is ≤ 2 per cent. Thus our numerical tests of the non-linear

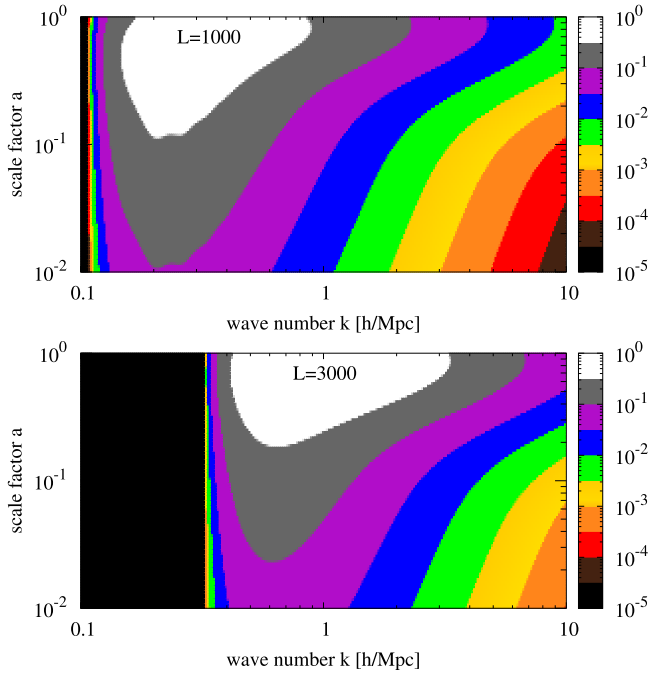

Figure 8. Heat map showing $dC_L^{\phi\phi}/d\eta/C_L^{\phi\phi}$, that is, the fractional contribution of each wavenumber and scale factor, to the total $a = 1$ lensing potential power spectrum. (Top) $L = 1000$. (Bottom) $L = 3000$. Large scales $k < L/\chi_*$ do not contribute to $C_L^{\phi\phi}$.

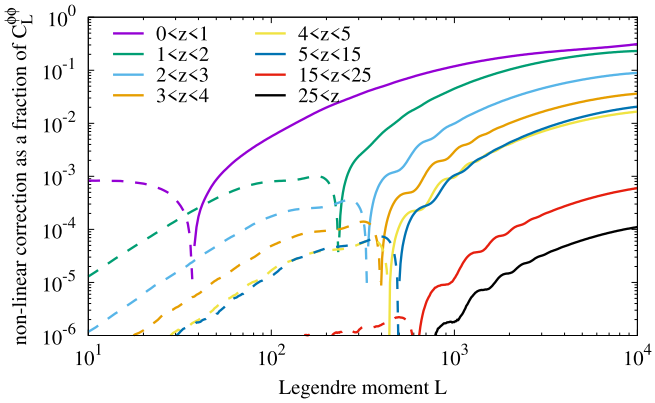

Figure 9. Fractional contributions of non-linear power to the total lensing potential power spectrum $C_L^{\phi\phi}$ due to clustering in different redshift bins. Solid (dashed) lines correspond to positive (negative) contributions.

convergence power spectrum in the remainder of this section should focus on $z \lesssim 3$–5.

3.3 *N*-body fits and emulation

Although Section 3.1 used the Time-RG non-linear perturbation theory to compute the matter power spectrum $P_{\rm m}(k)$, we may substitute any non-linear $P_{\rm m}(k)$ into equation (12). Perturbation theory accurately approximates $P_{\rm m}(k)$ at high redshifts but becomes increasingly inaccurate for $z < 1$, while approximations calibrated to *N*-body simulations have prioritized low-redshift accuracy.

This work considers three such simulation-based approximations. `Halofit`, inspired by the halo model, splits the non-linear power spectrum into a quasi-linear term, which approaches the linear power spectrum on large scales, and a non-linear term, which is fit to simulations; see Smith et al. (2003), Bird et al. (2012), and Takahashi et al. (2012). We implement the `Halofit` model of Bird et al. (2012) that has been fit for massive neutrino cosmologies.

The second approximation is the Mira-Titan IV (MT4) emulator of Moran et al. (2023) that uses Gaussian process modelling to interpolate a suite of 111 simulations. Chosen carefully to span a large parameter space, this suite includes 101 massive neutrino cosmologies with $0.00017 \leq \omega_\nu \leq 0.01$. The third approximation, Euclid Emulator 2 (EE2) of Knabenhans et al. (2021), covers a greater range of redshifts and wavenumbers than the MT4 emulator, at the cost of a smaller neutrino mass range, $\sum m_\nu \leq 0.15$ eV. `Halofit` has been calibrated up to $z = 3$, the MT4 emulator to $z = 2$, and EE2 to $z = 10$; at higher redshifts, we revert to Time-RG in equation (12).

MT4 and EE2 are based upon simulations with mass resolutions of 10^{10} and 10^9 $M_\odot$, respectively, compared with 7×10^9 $M_\odot$ for standard `FLAMINGO` simulations, and 8×10^8 $M_\odot$ for L1_m8_DMO. MT4 reaches $k \approx 7\,h$ Mpc^{-1} and EE2 reaches $\approx 9\,h$ Mpc^{-1}, compared with 17 h Mpc^{-1} for standard `FLAMINGO` simulations and 33 h Mpc^{-1} for L1_m8_DMO. Thus the `FLAMINGO` simulation suite is appropriate for testing emulator results.

Fig. 10 compares these non-linear methods at a range of redshifts. Evidently, the two emulators provide the most accurate power spectra at all redshifts for which they are available, with the MT4 emulator slightly more accurate at $z \geq 1$ and $1\,h$ Mpc$^{-1} \lesssim k \lesssim 7\,h$ Mpc^{-1}. As the MT4 emulator is restricted to the range $k \leq 5$/Mpc $\approx 7\,h$ Mpc^{-1}, we logarithmically extrapolate its power spectrum beyond that wavenumber, an approximation whose accuracy evidently diminishes around $z = 2$. For $z > 2$, Time-RG is somewhat more accurate than `Halofit` at large scales, $k \lesssim 2\,h$ Mpc^{-1}, and less accurate at small scales.

3.4 Tests of $C_L^{\phi\phi}$

Fig. 11 compares lensing potential power spectra from linear and Time-RG perturbation theories, `Halofit`, and the two emulators to the `FLAMINGO` lensing potential power spectrum, computed as described in Section 2.2. As might be expected from the comparison to a higher resolution simulation in Fig. 10, both perturbation theories underpredict small-scale power, with linear and Time-RG perturbation theory underpredicting by ≥ 3 per cent above $L \approx 300$ and $L \approx 2400$, respectively. `Halofit` is more accurate than Time-RG for $200 \lesssim L \lesssim 1000$ but underpredicts power by ≥ 3 per cent above $L \approx 1300$. In this and subsequent plots, the ratios to *N*-body power spectra are summed over all available light-cones and smoothed using centred 100-point moving averages.

Meanwhile, the emulator-based calculations are highly accurate. The two emulators, MT4 and EE2, agree at the per cent level at $L \lesssim 6000$, with MT4 somewhat more accurate for $1000 \lesssim L \lesssim 4000$ and less accurate for $L \gtrsim 5000$. Due to its greater neutrino mass range, and its slight accuracy advantage in Fig. 11, we will focus henceforth on MT4. Our standard $C_L^{\phi\phi}$ calculation (MT4 + TRG) will use the MT4 emulator for the matter power spectrum in the range $z \leq 2$ and Time-RG perturbation theory above that redshift.

The accuracy of this standard MT4 + TRG calculation for matter sources in several redshift bins is assessed in Fig. 12 (Top). Aside from low-L bumps consistent with Fig. 4, our MT4 + TRG calculation is accurate to ≤ 2 per cent up to $L = 3000$. For all but one redshift bin, $2 \leq z \leq 3$, MT4 + TRG is ≤ 5 per cent accurate for $L \leq 6000$. Meanwhile, for $2 \leq z \lesssim 5$, Euclid Emulator

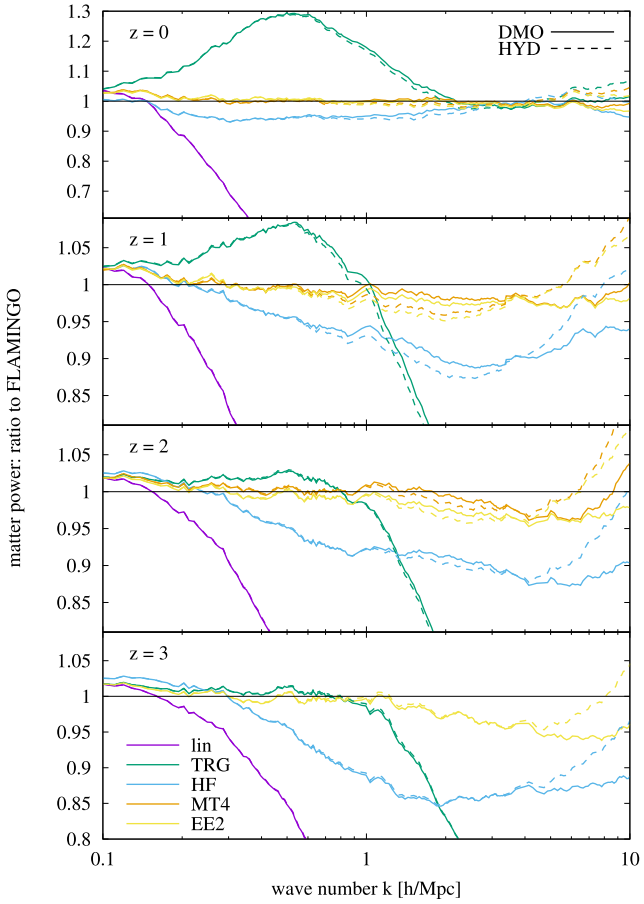

Figure 10. Matter power spectrum ratios to FLAMINGO for the models L1_m8_DMO (solid) and L1_m8 (dashed). Shown are power spectra from linear (lin) and Time-RG (TRG) perturbation theories; the Halofit (HF) model of Bird et al. (2012); the Mira-Titan IV emulator (MT4) of Moran et al. (2023); and Euclid Emulator 2 (EE2) of Knabenhans et al. (2021). SP(k) was used to model the baryonic suppression. MT4 is limited to $z \leq 2$.

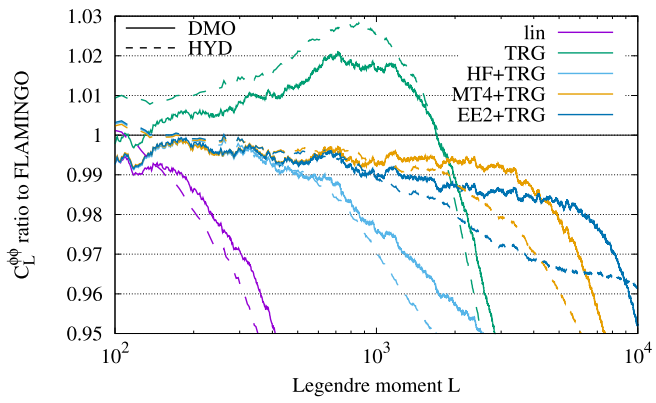

Figure 11. Lensing potential power spectrum computed using linear (lin) or Time-RG (TRG) perturbation theories, Halofit (HF), the Mira-Titan IV (MT4) emulator, or Euclid Emulator 2 (EE2), compared with the FLAMINGO power spectrum. HF is used up to $z = 3$, MT4 to $z = 2$, and EE2 to $z = 10$, above which $C_L^{\phi\phi}$ is computed using Time-RG. Solid lines compare DMO calculations to the FLAMINGO L5p6_m10_DMO $C_L^{\phi\phi}$ computed up to $z = 25$. Dashed lines, using SP(k) for the baryonic suppression, are compared with the FLAMINGO L2p8_m9 $C_L^{\phi\phi}$ computed to $z = 5$.

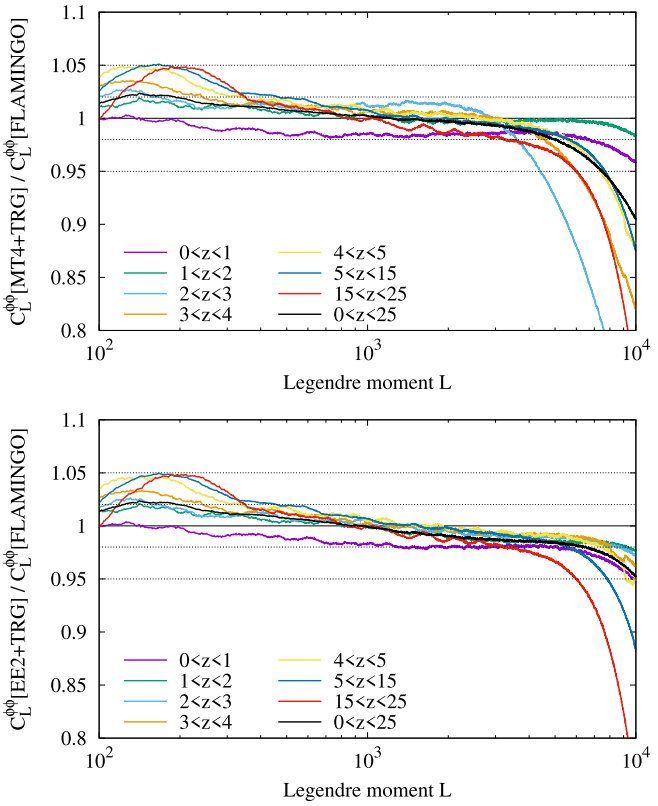

Figure 12. Accuracy of $C_L^{\phi\phi}$ contributions from several redshift bins. Our computation integrates equation (12) over the appropriate redshift range, using (Top) the Mira-Titan IV Emulator for $P_{\rm m}$ at $z \leq 2$ and Time-RG perturbation theory at $z > 2$; and (Bottom) Euclid Emulator 2 at $z \leq 10$ and Time-RG above that. These are compared with the FLAMINGO N-body $C_L^{\phi\phi}$ contributions from the same redshift bins. Inner and outer dotted lines show errors of 2 per cent and 5 per cent, respectively.

2 provides 5 per cent-accurate tomographic bins all the way to $L = 10\,000$, as seen in Fig. 12 (Bottom). Thus our calculation is accurate not just for the total CMB lensing power, but also for individual redshift bins, as needed for tomographic analyses such as Peacock & Bilicki (2018), Krolewski, Ferraro & White (2021), Chang et al. (2023), Abbott et al. (2023), and Wang et al. (2023).

3.5 Massive neutrino suppression

We begin by considering the effects of neutrinos on the matter power itself. Following the approach of Lesgourgues et al. (2009), Upadhye et al. (2014), and Upadhye (2019), we include neutrinos at the fully linear level, interpolating the ratio $\delta_\nu/\delta_{\rm cb}$ from CAMB (Lewis, Challinor & Lasenby 2000; Lewis & Bridle 2002). Thus, we include neutrinos in the gravitational potential $\Phi(k, z)$ of equation (7) as

$$k^2\Phi = -\frac{3}{2}\mathcal{H}^2\left(\Omega_{\rm cb}(z) + \Omega_\nu(z)\left.\frac{\delta_\nu}{\delta_{\rm cb}}\right|_{\rm lin}\right)\delta_{\rm cb}. \quad (13)$$

Ratios of matter power spectra differing only in M_ν, with $\Omega_{\rm m,\,0}$, $\Omega_{\rm b,\,0}$, h, n_s, A_s, w_0, and w_a held fixed, exhibit a characteristic 'spoon' feature shown in Fig. 13 that even non-linear perturbation theory fails to capture. Hannestad, Upadhye & Wong (2020) showed this to be a consequence of halo formation. Halofit, which was calibrated in

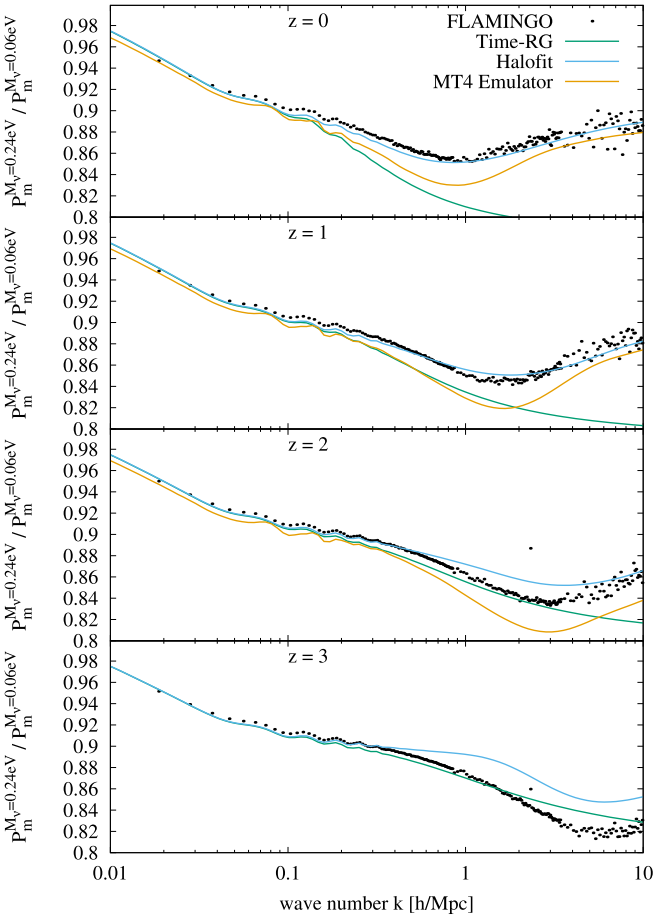

Figure 13. Spoon-like feature in the ratios of power spectra of two models, PlanckNu0p24Fix_DMO and Planck_DMO from Table 1, differing only in their neutrino mass sums M_ν. `Halofit` has been calibrated to fit low-redshift spoons accurately, while perturbation theory is qualitatively incapable of reproducing this non-perturbative effect. MT4 is limited to $z \leq 2$.

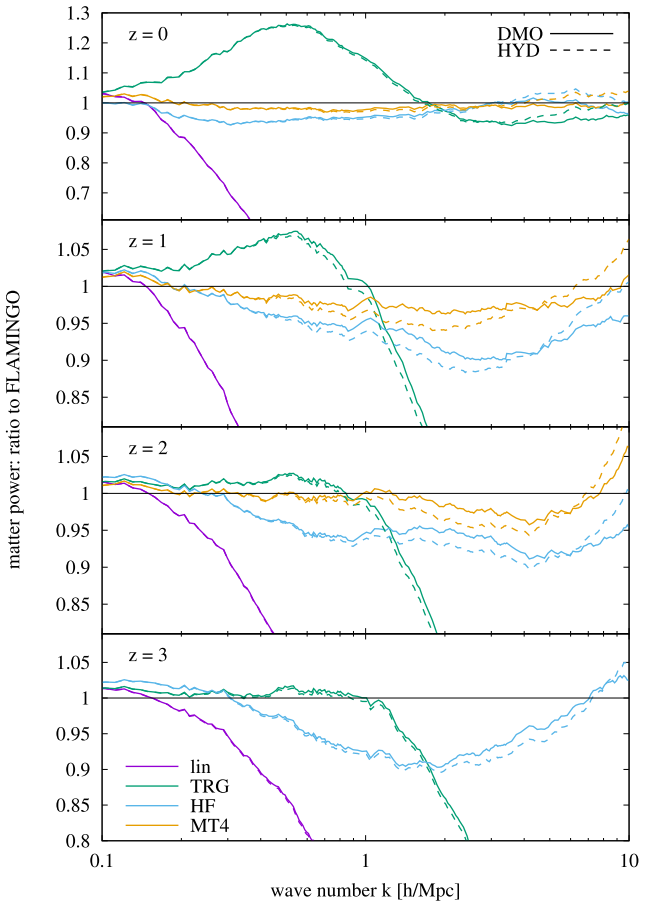

Figure 14. Matter power spectrum ratios to `FLAMINGO` for the models PlanckNu0p24Fix_DMO (solid) and PlanckNu0p24Fix (dashed), with $M_\nu = 0.24$ eV. Shown are power spectra from linear (lin) and Time-RG (TRG) perturbation theories, `Halofit` (HF), and the Mira-Titan IV emulator (MT4), with `SP(k)` used to model the baryonic suppression. MT4 is limited to $z \leq 2$.

Bird et al. (2012) to fit this feature, agrees closely with the N-body ratio, at least at low-z, while the Mira-Titan IV emulator is accurate to a few per cent.

Fig. 14 plots the matter power spectrum for a massive neutrino cosmology with $M_\nu = 0.24$ eV. In spite of `Halofit`'s accurate computation of the neutrino spoon for $z < 2$, we find that the matter power itself is more accurately computed for $z \leq 2$ by the Mira-Titan IV emulator, and for $z > 2$ at large scales, $k \leq 1\ h\ \mathrm{Mpc}^{-1}$, by Time-RG perturbation theory. Thus, our MT4 + TRG calculation continues to be the best approximation to the matter power spectrum for the k and z relevant to computing the CMB lensing power spectrum in $\nu\Lambda$CDM cosmologies.

Next, we return to our computation of the lensing potential power spectrum in massive neutrino models. We tested our fully linear neutrino approximation by allowing neutrinos to respond linearly to the non-linear CB growth, using the code of Chen, Upadhye & Wong (2021), and found the impact on $C_L^{\phi\phi}$ to be < 0.06 per cent even for neutrino fractions as large as $\Omega_{\nu,0}h^2 = 0.01$. Chen, Upadhye & Wong (2023) used a non-linear perturbation theory for massive neutrinos to show that the non-linear neutrino power is at most 2–3 times the linear response power for $\Omega_{\nu,0}h^2 \leq 0.005$, corresponding to $M_\nu = 0.47$ eV, so we may safely bound the impact of neutrino non-linearity to < 0.2 per cent over this range. This is a negligible source of error for near-future experiments, so we use the fully linear-ν approximation of equation (13) henceforth.

Errors in the CMB lensing potential power spectra for a model with $M_\nu = 0.24$ eV are shown in Fig. 15. The `FLAMINGO` simulation uses a 1000 Mpc box, for which we have light-cones up to $z = 3$, so the other calculations are integrated only to $z = 3$ for comparison. In accordance with our previous results, the MT4 + TRG calculation is accurate to < 2.5 per cent up to $L = 2000$ and to < 5 per cent up to $L = 6000$. These error estimates are limited by noise in the simulation itself, since only one high-z light-cone is available in the high-M_ν run.

4 HYDRODYNAMIC SUPPRESSION

4.1 Matter and convergence power spectra

Dashed curves in Fig. 10 compare the total matter power spectrum from the high-resolution `FLAMINGO` hydrodynamic simulation, L1_m8, to approximations applying the `SP(k)` fitting function of Salcido et al. (2023) to each of the following: linear and Time-RG perturbation theories, `Halofit`, and MT4 and EE2 emulators. Since `SP(k)` applied to P_m is a multiplicative correction, the ratio

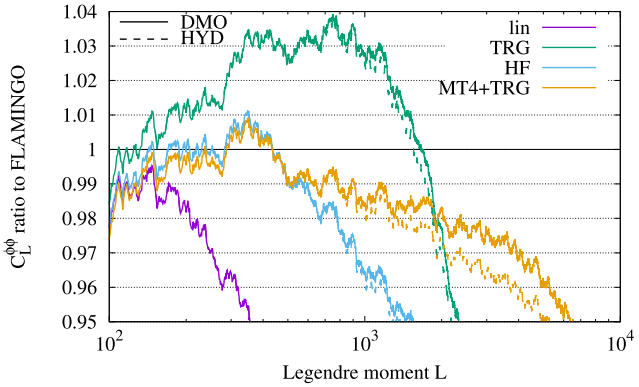

Figure 15. Lensing potential power spectra for $M_\nu = 0.24$ eV, computed using linear (lin) or Time-RG (TRG) perturbation theories, `Halofit` (HF), or the Mira-Titan IV (MT4) emulator, compared with `FLAMINGO` models PlanckNu0p24Fix_DMO and PlanckNu0p24Fix, computed using light-cone sources up to $z = 3$. For the orange curves, MT4 is used up to $z = 2$ and TRG above that. Solid lines compare DMO $C_L^{\phi\phi}$ to PlanckNu0p24Fix_DMO. Dashed lines compare $C_L^{\phi\phi}$ from the above methods, using `SP(k)` for the baryonic suppression, to PlanckNu0p24Fix.

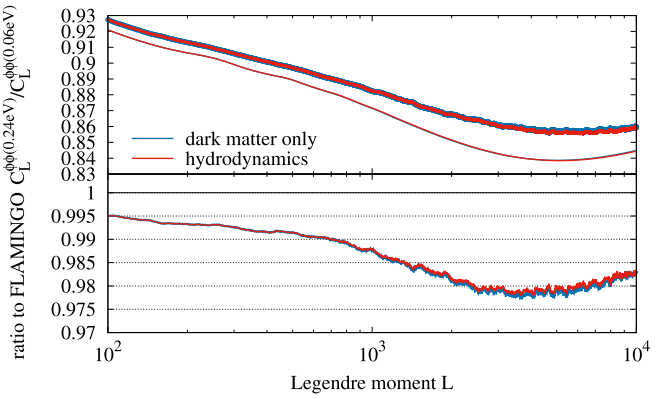

Figure 16. Spoon feature in the ratio of high-M_ν to low-M_ν power spectra. (Top) $C_L^{\phi\phi}$ ratio. Thick lines show the `FLAMINGO` (PlanckNu0p24Fix_DMO/Planck_DMO and PlanckNu0p24Fix/Planck) spoons, while thin lines show the corresponding MT4 + TRG spoons. (Bottom) Ratio of each MT4 + TRG spoon to its `FLAMINGO` counterpart.

of corresponding dashed and solid lines is the same in all cases, and the deviation between each dashed and solid line is due to error in `SP(k)`. At all redshifts, this error is ≤ 2 per cent up to $k \approx 5\ h$ Mpc^{-1}.

Aside from the small systematic overprediction of power at small scales, $k \gtrsim 5\ h$ Mpc^{-1}, for the hydro models, the DMO and hydro curves are qualitatively similar. Thus, our earlier conclusions apply as well to the hydro power spectra of Fig. 10: the two emulators predict the power spectrum accurately across their tested redshift ranges, while above $z = 2$, Time-RG perturbation theory accurately calculates the power up to $k \approx 1\ h$ Mpc^{-1} and `Halofit` above $k \approx 2\ h$ Mpc^{-1}.

Each of these methods, along with the `SP(k)` fitting function, has been applied to the computation of the CMB lensing potential power spectrum in the dashed curves of Fig. 11. They are compared to the `FLAMINGO` hydrodynamic simulation, L2p8_m9, whose light-cones cover redshifts $z \leq 5$. Our standard MT4+TRG + `SP(k)` computation is 2 per cent accurate up to $L = 3000$ and 5 per cent accurate to $L = 6000$, while the EE2 + `SP(k)` computation, which reaches smaller scales, is 4 per cent accurate all the way to $L = 10\,000$. Fig. 15 shows a similar accuracy level, 3 per cent to $L = 2000$ and 5 per cent to $L = 5000$, for a model with both baryonic feedback and massive neutrinos, PlanckNu0p24Fix.

Fig. 16 plots neutrino spoons for the lensing potential power spectra, analogous to Fig. 13 for the matter power. The spoon is not an observable; since we do not know the exact cosmological parameter values of the observed universe, we cannot divide its power spectrum by that of another model differing only in M_ν. Instead, we consider it as a test of systematic biases associated with variations of the neutrino mass alone. The power spectrum ratio is less sensitive to hydrodynamic effects than the power spectrum itself, which is why the DMO and hydrodynamic curves are nearly identical. Furthermore, this ratio also reduces errors present in perturbation theory and the emulator, so that the MT4 + TRG spoon is correct to < 2.5 per cent for all $L \leq 10\,000$.

Nevertheless, we may wonder about the impact of the systematic ≈ 2 per cent overprediction of the spoon depth seen in Fig. 16. This is indicative of a systematic overestimate of the neutrino suppression of $C_L^{\phi\phi}$ around $M_\nu = 0.24$ eV, a systematic underestimate around $M_\nu = 0.06$ eV, or some combination of those, with other cosmological parameters around their currently preferred values. Since this 2 per cent error for $L \gtrsim 2000$ is comparable to the binned error at $L = 2000$ forecast for a CMB Stage-IV survey in Abazajian et al. (2016), testing using mock analyses is warranted.

4.2 Factorizability of suppressions

Since neutrinos and baryonic feedback both suppress small-scale clustering, one may worry that errors in our feedback approximations will lead to significant biases in the neutrino mass determination. However, Mummery et al. (2017), using the `BAHAMAS` simulations of McCarthy et al. (2017, 2018), demonstrated for several clustering statistics that neutrino and baryonic suppressions factorize. That is, the combined effect of neutrino and baryon suppression is the product of the individual suppression factors. Here, we show that this factorization of neutrino and baryon effects also applies to the `FLAMINGO` CMB lensing potential power spectrum.

Consider the Planck-like cosmologies of Planck_DMO, PlanckNu0p24Fix_DMO, Planck, and PlanckNu0p24Fix from Table 1, which have either a minimal ($M_\nu = 0.06$ eV) or high ($M_\nu = 0.24$ eV) neutrino mass sum, and are simulated using dmo or a hydrodynamic (hyd) simulations. We may define three different suppression factors, capturing the effects of neutrinos, baryons, and both at the same time:

$$\mathcal{F}_\nu = \frac{C_L^{\phi\phi}(\text{dmo}, M_\nu = 0.24\text{eV})}{C_L^{\phi\phi}(\text{dmo}, M_\nu = 0.06\text{eV})} \tag{14}$$

$$\mathcal{F}_\text{b} = \frac{C_L^{\phi\phi}(\text{hyd}, M_\nu = 0.06\text{eV})}{C_L^{\phi\phi}(\text{dmo}, M_\nu = 0.06\text{eV})} \tag{15}$$

$$\mathcal{F}_{\nu+\text{b}} = \frac{C_L^{\phi\phi}(\text{hyd}, M_\nu = 0.24\text{eV})}{C_L^{\phi\phi}(\text{dmo}, M_\nu = 0.06\text{eV})} \tag{16}$$

Factorizability then implies that $\mathcal{F}_{\nu+\text{b}} = \mathcal{F}_\nu \times \mathcal{F}_\text{b}$.

Fig. 17 demonstrates the factorizability of neutrino and baryon suppressions in the CMB lensing potential power spectrum, using simulations as well as our standard MT4 + TRG + `SP(k)` calculation. While the simulations and emulators differ by ≈ 2 per cent on small scales, in keeping with our previous results, in each case factorizability of the neutrino and baryon suppressions is accurate to better than 1 per cent.

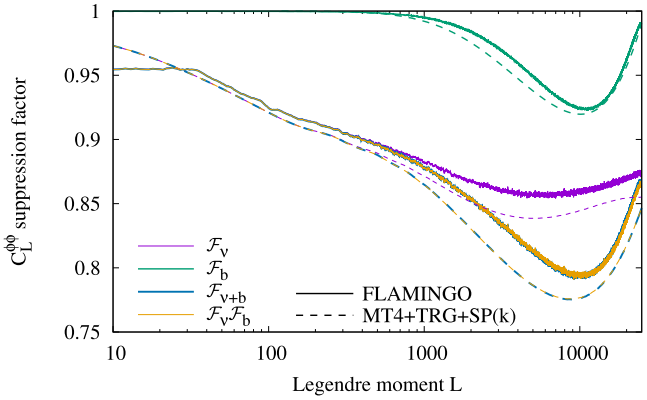

Figure 17. Factorizability of neutrino and hydrodynamic suppressions of the CMB lensing potential power spectrum. FLAMINGO simulations (solid) and MT4 + TRG calculations (dashed) of the neutrino suppression factor $\mathcal{F}_\nu$, the baryon suppression factor $\mathcal{F}_b$, the combined suppression factor $\mathcal{F}_{\nu+b}$, and the product of the first two, which accurately approximates the third.

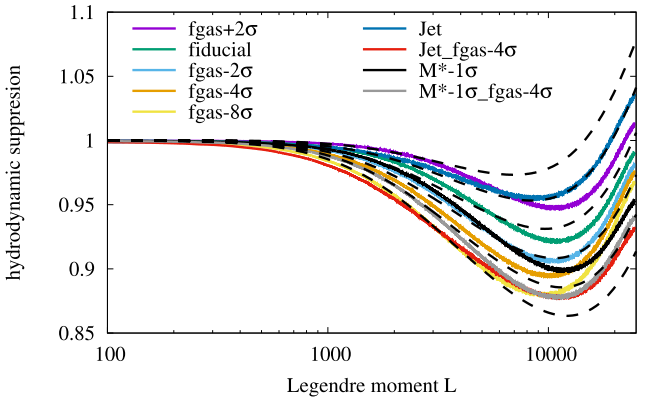

Figure 18. Model-dependent feedback suppression of $C_L^{\phi\phi}$. Our p_f feedback parametrization (dashed), which raises the L1_m9 baryon fraction $\hat{\tilde{f}}_b$ to the power p_f before applying the SP(k) fit of Salcido et al. (2023), is compared to a wide range of FLAMINGO hydrodynamic suppression factors (solid). Dashed curves range from $p_f = 0.5$ (uppermost curve) to $p_f = 1.5$ (lowest curve) in increments of $\Delta p_f = 0.2$.

This is due to the fact that the two suppressions depend very differently on length and time. Neutrino free-streaming suppresses clustering below the free-streaming length, corresponding to $k_{FS} \lesssim 0.1\ h\ \mathrm{Mpc}^{-1}$ for typical masses, which is manifested in a ~ 10 per cent suppression of $C_L^{\phi\phi}$ even at $L = 100$. Meanwhile, baryonic suppression is negligible below $L = 1000$ and is most significant around $L = 10\,000$. Further, the neutrino free-streaming length is larger at earlier times, and the resulting impact upon $C_L^{\phi\phi}$ occurs over a wide range of redshifts, as opposed to the baryonic suppression, which was shown in Fig. 5 to be dominated by $z \lesssim 1$.

The reader may wonder how well such factorizability applies to very different hydrodynamic feedback models. While SP(k) was calibrated to a wide range of feedback models (Salcido et al. 2023), a thorough exploration of the parameter space is beyond the scope of this article. For simplicity, consider a one-parameter family of generalizations that raise $\hat{\tilde{f}}_b(z) = f_b(\hat{M}, z)/(\Omega_b/\Omega_m)$ to a non-negative power p_f before applying SP(k). In this case, $p_f = 0$ implies no hydrodynamic feedback; $p_f < 1$ implies feedback that is weaker than that of the FLAMINGO fiducial model; and $p_f > 1$ to feedback that is stronger. Fig. 18 compares this approximation to a

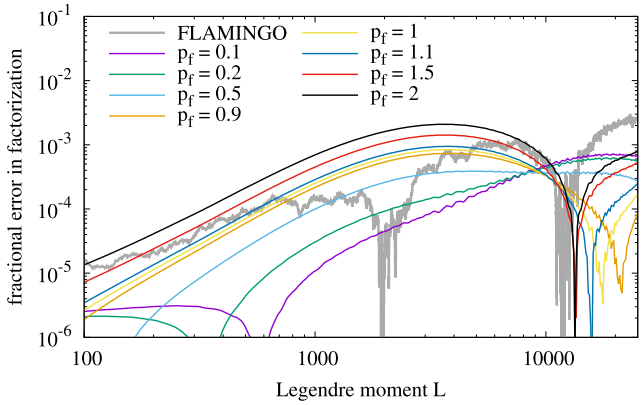

Figure 19. Error in $C_L^{\phi\phi}$ factorizability $|\mathcal{F}_\nu \times \mathcal{F}_b/\mathcal{F}_{\nu+b} - 1|$ versus p_f in the generalized feedback model raising $\tilde{f}_b(\hat{M}, z)$ to the power p_f in SP(k). Here, $p_f < 1$ ($p_f > 1$) corresponds to weaker (stronger) feedback than the FLAMINGO fiducial model. Also shown is the FLAMINGO factorizability error.

wide range of FLAMINGO feedback methods and finds $0.5 \leq p_f \leq 1.5$ to cover this range, to $L \approx 4000$, aside from one model, Jet_fgas-4σ.

Fig. 19 shows that factorizability of $C_L^{\phi\phi}$ holds to excellent precision for a wide range of p_f. While errors at $100 \lesssim L \lesssim 6000$ tend to rise with p_f, they remain ≤ 0.2 per cent all the way to $p_f = 2$, which has substantially stronger feedback, and are < 0.1 per cent at all L for the fiducial case, $p_f = 1$. Also shown is the factorization error in FLAMINGO itself, which is ≤ 0.3 per cent everywhere and has a similar functional form to HYPHI results. Factorizability opens up the possibility of a new SP(k)-like feedback fit applying directly to $C_L^{\phi\phi}(\mathrm{hyd})/C_L^{\phi\phi}(\mathrm{dmo})$, whose parameters could be treated as nuisance parameters whose marginalization would reveal the underlying neutrino suppression.

While Salcido et al. (2023) found that factorizability is a good approximation at the 2 per cent level with a history-independent baryonic correction model, the possibility remains that factorizability fails below this error threshold, or that baryonic correction is somewhat history-dependent. This limits the bound that we can place on the accuracy of factorizability, though we note that the direct factorizability measurement from FLAMINGO in Fig. 19 is consistent with our result of sub-per cent-level factorization errors.

4.3 Applicability to data constraints

Complete forecast constraints such as those of McCarthy et al. (2022), but using SP(k) through HYPHI, would first require forecast constraints on the halo baryon fraction $\hat{\tilde{f}}_b(z)$ as a function of redshift. As such, they are beyond the scope of this article. However, before concluding, we comment on the applicability of HYPHI to data and forecast constraints.

The simplest, but least powerful, approach is to parametrize $\tilde{f}_b(M, z)$ and then to treat these parameters as nuisance parameters over which to marginalize. Salcido et al. (2023) explore two- and three-parameter approximations to $\tilde{f}_b(M, z)$. A one-parameter approach, described in Section 4.2, begins with a representative $\tilde{f}_b(M, z)$ measured from a hydrodynamic simulation, and raises it to a power p_f prior to applying SP(k), with p_f marginalized over as a nuisance parameter. Fig. 18 demonstrates that the range $0.5 \leq p_f \leq 1.5$ covers a wide range of feedback models.

However, marginalizing over broad priors on $\tilde{f}_b(M, z)$ offers no significant advantages over directly parametrizing

$P_{\rm m,\,hyd}(k)/P_{\rm m,\,dmo}(k)$ as a function of subgrid feedback parameters, as done by Mead et al. (2020) in HMCODE. As such, we anticipate that the resulting forecast constraints will be similar to those of McCarthy et al. (2022).

The true strength of `SP(k)`-based power spectrum computations such as HYPHI is that $\tilde{f}_{\rm b}(M,z)$ is a physical observable. Measurements of the baryon fraction may be used to narrow the range over which $\tilde{f}_{\rm b}(M,z)$ are marginalized. For example, Salcido et al. (2023) parametrize $\tilde{f}_{\rm b}(M,z)$ and then provides parameter ranges consistent at the 2σ and 3σ level with the low-redshift measurements of Akino et al. (2022). As CMB lensing is sensitive to baryonic suppression at higher redshifts, $z\sim 1$, SZ cluster constraints such as those of Pandey et al. (2022) also provide useful priors on the baryon fraction, with the caveat that SZ surveys are sensitive to haloes with masses larger than the $\hat{M}(z)$ used in `SP(k)`.

5 CONCLUSION

Weak lensing of the CMB is becoming increasingly important as a galaxy-bias-independent constraint on the sum of neutrino masses M_ν. Our ability to interpret upcoming lensing measurements at small scales depends crucially on our ability to distinguish between the scale-dependent clustering suppressions due to baryons and neutrino free-streaming in the non-linear regime. Employing the FLAMINGO suite of simulations, the largest-particle-number hydrodynamic simulations reaching $z=0$ to date, we systematically compared several methods for computing the non-linear CMB lensing potential power spectrum, implemented in our HYPHI code.

Our preferred method combined the Mira-Titan IV DMO power spectrum emulator at $z\leq 2$ with Time-RG perturbation theory at higher redshifts. Comparing HYPHI with a 5.6 Gpc-box DMO simulation in Fig. 11, we found agreement to 1 per cent up to $L=3000$ and 2 per cent to $L=5000$. Fig. 12 shows that HYPHI attained 5 per cent accuracy to $L=4000$ in individual redshift bins, applicable to tomographic analyses. Including more massive neutrinos, as well as baryonic feedback through the `SP(k)` fitting function, individually or simultaneously, as in Figs 11 and 15, slightly degraded the accuracy of HYPHI, but even with both effects, we found MT4+TRG + `SP(k)` to be 4 per cent accurate up to $L=4000$. Finally, we demonstrated that neutrino and baryonic suppression effects on $C_L^{\phi\phi}$ factorize, potentially facilitating the marginalization over baryonic effects in the future.

ACKNOWLEDGEMENTS

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). WE acknowledges the Science and Technology Facilities Council (STFC) consolidated grant ST/X001075/1. RK acknowledges funding by Vici grant 639.043.409, Veni grant 639.041.751, and research programme Athena 184.034.002 from the Dutch Research Council (NWO). CSF acknowledges support from ERC Advanced Investigator grant, DMIDAS [GA 786910]. This paper makes use of the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (http://www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grants ST/H008519/1 and ST/K00087X/1, STFC DiRAC Operations grant ST/K003267/1, and Durham University. DiRAC is part of the National E-Infrastructure.

DATA AVAILABILITY

The HYPHI code is available at `github.com/upadhye/hyphi`. The authors of Schaye et al. (2023) intend to make the FLAMINGO simulation data public at a later date.

REFERENCES

Abazajian K. N. et al., 2016, preprint (arXiv:1610.02743)
Abbott T. M. C. et al., 2022, Phys. Rev. D, 105, 023520
Abbott T. M. C. et al., 2023, Phys. Rev. D, 107, 023531
Abdalla E. et al., 2022, J. High Energy Astrophys., 34, 49
Ade P. A. R. et al., 2016, ApJ, 833, 228
Ade P. et al., 2019, J Cosmol. Astropart. Phys., 02, 056
Aghanim N. et al., 2020a, A&A, 641, A6
Aghanim N. et al., 2020b, A&A, 641, A8
Aguilar Faúndez M. A. O. et al., 2020, ApJ, 893, 85
Aker M. et al., 2019, Phys. Rev. Lett., 123, 221802
Aker M. et al., 2022, Nat. Phys., 18, 160
Akino D. et al., 2022, PASJ, 74, 175
Alam S. et al., 2021, Phys. Rev. D, 103, 083533
Amon A. et al., 2023, MNRAS, 518, 477
Bird S., Viel M., Haehnelt M. G., 2012, MNRAS, 420, 2551
Cai R.-G., Guo Z.-K., Wang S.-J., Yu W.-W., Zhou Y., 2022, Phys. Rev. D, 106, 063519
Capozzi F., Lisi E., Marrone A., Palazzo A., 2018, Prog. Part. Nucl. Phys., 102, 48
Carron J., Mirmelstein M., Lewis A., 2022, J. Cosmol. Astropart. Phys., 09, 039
Chang C. et al., 2023, Phys. Rev. D, 107, 023530
Chen J. Z., Upadhye A., Wong Y. Y. Y., 2021, J. Cosmol. Astropart. Phys., 03, 065
Chen J. Z., Upadhye A., Wong Y. Y. Y., 2023, J. Cosmol. Astropart. Phys., 05, 046
Chisari N. E. et al., 2018, MNRAS, 480, 3962
Chung E., Foreman S., van Engelen A., 2020, Phys. Rev. D, 101, 063534
Cooray A., Sheth R. K., 2002, Phys. Rep., 372, 1
Crocce M., Scoccimarro R., 2006, Phys. Rev. D, 73, 063519
Das S. et al., 2014, J. Cosmol. Astropart. Phys., 04, 014
de Salas P. F., Forero D. V., Ternes C. A., Tortola M., Valle J. W. F., 2018, Phys. Lett. B, 782, 633
Debackere S. N. B., Schaye J., Hoekstra H., 2021, MNRAS, 505, 593
Di Valentino E., Melchiorri A., 2022, ApJ, 931, L18
Di Valentino E., Melchiorri A., Silk J., 2020, J. Cosmol. Astropart. Phys., 01, 013
Driver S. P. et al., 2022, MNRAS, 513, 439
Elbers W., 2022a, Astrophysics Source Code Library, record ascl:2212.004
Elbers W., 2022b, J. Cosmol. Astropart. Phys., 11, 058
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., 09, 178
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelman M., 2005, ApJ, 622, 759
Hahn O., Michaux M., Rampf C., Uhlemann C., Angulo R. E., 2020, Astrophysics Source Code Library, record ascl:2008.024
Hahn O., Rampf C., Uhlemann C., 2021, MNRAS, 503, 426
Hannestad S., Upadhye A., Wong Y. Y. Y., 2020, J. Cosmol. Astropart. Phys., 11, 062
Heitmann K., Higdon D., White M., Habib S., Williams B. J., Wagner C., 2009, ApJ, 705, 156
Hellwing W. A., Schaller M., Frenk C. S., Theuns T., Schaye J., Bower R. G., Crain R. A., 2016, MNRAS, 461, L11
Knabenhans M. et al., 2021, MNRAS, 505, 2840

Krolewski A., Ferraro S., White M., 2021, J. Cosmol. Astropart. Phys., 12, 028
Kugel R. et al., 2023, preprint (arXiv:1610.02743)
Lawrence E., Heitmann K., White M., Higdon D., Wagner C., Habib S., Williams B., 2010, ApJ, 713, 1322
Leauthaud A. et al., 2017, MNRAS, 467, 3024
Lesgourgues J., Matarrese S., Pietroni M., Riotto A., 2009, J. Cosmol. Astropart. Phys., 06, 017
Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511
Lewis A., Challinor A., 2006, Phys. Rep., 429, 1
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Limber D. N., 1953, ApJ, 117, 134
Liu J. et al., 2022, Sci. China Phys. Mech. Astron., 65, 109511
Ma C.-P., Fry J. N., 2000, ApJ, 543, 503
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
McDonald P., 2007, Phys. Rev. D, 75, 043514
Matsubara T., 2008a, Phys. Rev. D, 77, 063530
Matsubara T., 2008b, Phys. Rev. D, 78, 083519
Mead A., Peacock J., Heymans C., Joudaki S., Heavens A., 2015, MNRAS, 454, 1958
Mead A., Brieden S., Tröster T., Heymans C., 2020, MNRAS, 502, 1401
Moran K. R. et al., 2023, MNRAS, 520, 3443
Mummery B. O., McCarthy I. G., Bird S., Schaye J., 2017, MNRAS, 471, 227
Pakmor R., et al., 2023, MNRAS, 524, 2539
Palanque-Delabrouille N., Yèche C., Schöneberg N., Lesgourgues J., Walther M., Chabanier S., Armengaud E., 2020, J. Cosmol. Astropart. Phys., 04, 038
Pandey S. et al., 2022, Phys. Rev. D, 105, 123526
Pandey S. et al., 2023, MNRAS, 525, 1779
Peacock J. A., Bilicki M., 2018, MNRAS, 481, 1133
Pietroni M., 2008, J. Cosmol. Astropart. Phys., 10, 036
Qu F. J., et al., 2024, ApJ, 962, 112
Saito S., Takada M., Taruya A., 2008, Phys. Rev. Lett., 100, 191301
Salcido J., McCarthy I. G., Kwan J., Upadhye A., Font A. S., 2023, MNRAS, 523, 2247
Schaller M., et al., 2023, preprint (arXiv:2305.13380)
Schaye J., et al., 2023, MNRAS, 526, 4978
Seljak U., 2000, MNRAS, 318, 203
Semboloni E., Hoekstra H., Schaye J., van Daalen M. P., McCarthy I. J., 2011, MNRAS, 417, 2020
Sgier R., Lorenz C., Refregier A., Fluri J., Zürcher D., Tarsitano F., 2021, prepint (arXiv:2110.03815)
Sherwin B. D. et al., 2017, Phys. Rev. D, 95, 123529
Simard G. et al., 2018, ApJ, 860, 137
Smith R. E. et al., 2003, MNRAS, 341, 1311
Springel V. et al., 2018, MNRAS, 475, 676
Szapudi I., Prunet S., Pogosyan D., Szalay A. S., Bond J. R., 2000, ApJ, 548, L115
Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152
Taruya A., Hiramatsu T., 2008, ApJ, 674, 617
Upadhye A., 2019, J. Cosmol. Astropart. Phys., 05, 041
Upadhye A., Biswas R., Pope A., Heitmann K., Habib S., Finkel H., Frontiere N., 2014, Phys. Rev. D, 89, 103515
van Daalen M. P., Schaye J., Booth C. M., Vecchia C. D., 2011, MNRAS, 415, 3649
van Daalen M. P., Schaye J., McCarthy I. G., Booth C. M., Dalla Vecchia C., 2014, MNRAS, 440, 2997
van Daalen M. P., McCarthy I. G., Schaye J., 2020, MNRAS, 491, 2424
van Engelen A. et al., 2012, ApJ, 756, 142
Wang Z., Yao J., Liu X., Liu D., Fan Z., Hu B., 2023, MNRAS, 523, 3001
Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, Phys. Rep., 530, 87
Wu W. L. K. et al., 2019, ApJ, 884, 70

APPENDIX A: CODE USAGE AND PERFORMANCE

Our CMB lensing potential power spectrum code, HYPHI, is publicly available at `github.com/upadhye/hyphi` . Its precision and accuracy are respectively shown in Figs A1 and A2, for several combinations of the absolute and relative error tolerances $\epsilon_{\rm abs}$ and $\epsilon_{\rm rel}$, along with the running time for a single $C_L^{\phi\phi}$ computation on a standard desktop computer. Evidently, a highly accurate lensing potential power spectrum may be computed in a fraction of a second.

Fig. A2 also demonstrates the speed and accuracy of HYPHI for larger neutrino masses. Since only one high-redshift light-cone is available, the simulated $C_L^{\phi\phi}$ is noisier, preventing us from quantifying an accuracy better than 2 per cent − 3 per cent. Evidently HYPHI is $\lesssim 3$ per cent accurate to nearly $L = 4000$. The running time for each $\epsilon_{\rm abs}$ and $\epsilon_{\rm rel}$ combination is within 30 per cent of the values listed in Fig. A1.

A technical point in the PlanckNu0p24Fix_DMO computation deserves further comment. The FLAMINGO PlanckNu0p24Fix_DMO light-cone covers the redshift range $0 \leq z \leq 3$, so we compare it

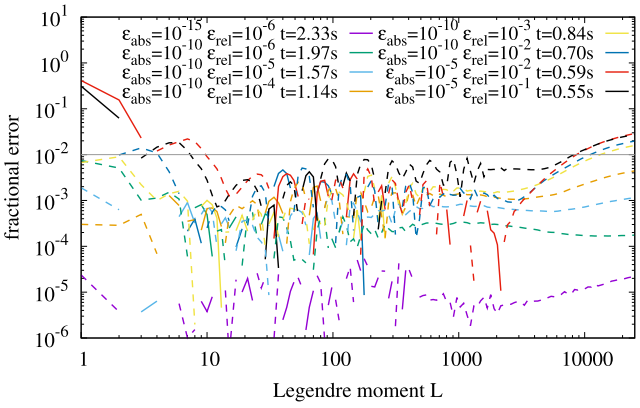

Figure A1. Numerical convergence of HYPHI $C_L^{\phi\phi}$ computation. Each $C_L^{\phi\phi}$ computation, labelled by its absolute and relative tolerances as well as its running time on a standard desktop computer, is compared to a high-quality run with absolute and relative error tolerances $\epsilon_{\rm abs} = 10^{-20}$ and $\epsilon_{\rm rel} = 10^{-12}$, respectively, whose running time is 26.16 s.

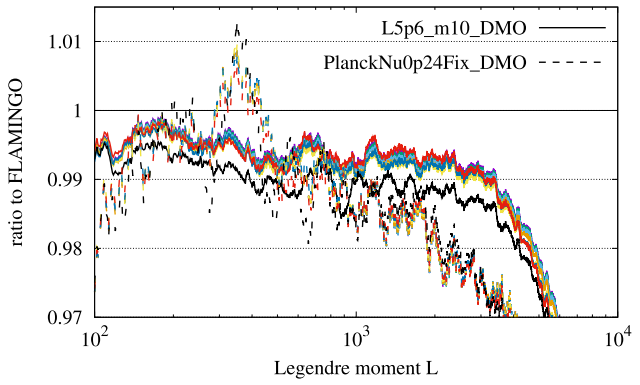

Figure A2. Accuracy of HYPHI $C_L^{\phi\phi}$ computation compared with the FLAMINGO L5p6_m10_DMO (solid) and PlanckNu0p24Fix_DMO (dashed) simulations. Absolute and relative tolerances for each line colour are the same as in Fig. A1.

with the HYPHI $C_L^{\phi\phi}$ including only lens masses from that range of redshifts. Although HYPHI can begin integrating $C_L^{\phi\phi}$ at $z = 3$, we find that doing so accurately requires a low absolute error tolerance, and correspondingly longer running times. Instead, in the dashed curves of Fig. A2, we have integrated $C_L^{\phi\phi}$ over all redshifts, but printed two separate power spectra at z of 3 and 0. These are then differenced to obtain the power spectra shown in the figure, allowing us to achieve running times comparable to those of L5p6_m10_DMO. This technique is also applicable to tomographic analyses. For example, printing $C_L^{\phi\phi}$ at redshifts 3, 2, 1, and 0, then differencing successive outputs, will provide fast and accurate power spectra in the $2 \leq z \leq 3$, $1 \leq z \leq 2$, and $0 \leq z \leq 1$ bins, respectively.

Each HYPHI run requires an input transfer function for the purpose of normalizing its matter power spectrum. Furthermore, the linear neutrino density as a function of time is interpolated from transfer functions over a range of redshifts. We find accurate results using transfer functions at the following 12 redshifts: 200, 100, 50, 20, 10, 5, 4, 3, 2, 1, 0.5, and 0. Currently, HYPHI is written to accept the 13-column transfer function files output by current versions of the CAMB code. Additionally, in its most accurate setting, HYPHI uses the Mira-Titan IV emulator, which must be compiled separately. Further instructions for compilation and usage of HYPHI are available at the web site listed above.

Two options are provided for implementing hydrodynamic corrections in HYPHI through `SP(k)`. First, the baryon fraction for the BAHAMAS simulations of McCarthy et al. (2017, 2018) is included with the code, and may be raised to a positive power p_f as in Section 4.2 prior to application of `SP(k)`. Secondly, a variant of the power-law baryon fraction approximation of Akino et al. (2022), as implemented in `SP(k)` by Salcido et al. (2023), is included in HYPHI. Its three parameters determine the normalization, mass dependence, and redshift dependence of the baryon fraction.

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