Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **493,** 305–319 (2020) doi:10.1093/mnras/staa281
Advance Access publication 2020 February 5

# The accuracy of weak lensing simulations

Stefan Hilbert,[1,2★] Alexandre Barreira,[3★] Giulio Fabbian ,[4,5] Pablo Fosalba,[6,7] Carlo Giocoli ,[8,9,10] Sownak Bose ,[11] Matteo Calabrese,[12] Carmelita Carbone,[13] Christopher T. Davies,[14] Baojiu Li ,[14] Claudio Llinares[15] and Pierluigi Monaco[16,17,18,19]

[1]*Exzellenzcluster Universe, Boltzmannstr. 2, D-85748 Garching, Germany*
[2]*Ludwig-Maximilians-Universität, Universitäts-Sternwarte, Scheinerstr. 1, D-81679 München, Germany*
[3]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85748 Garching, Germany*
[4]*Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK*
[5]*Institut d'Astrophysique Spatiale, CNRS (UMR 8617), Univ. Paris-Sud, Université Paris-Saclay, Bât. 121, F-91405 Orsay, France*
[6]*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
[7]*Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*
[8]*Dipartimento di Fisica e Astronomia, Alma Mater Studiorum Università di Bologna, via Gobetti 93/2, I-40129 Bologna, Italy*
[9]*INAF, Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Gobetti 93/3, I-40129 Bologna, Italy*
[10]*INFN, Sezione di Bologna, viale Berti Pichat 6/2, I-40127 Bologna, Italy*
[11]*Center for Astrophysics, Harvard and Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA*
[12]*Astronomical Observatory of the Autonomous Region of the Aosta Valley (OAVdA), Loc. Lignan 39, I-11020 Nus (AO), Italy*
[13]*INAF - Institute of Space Astrophysics and Cosmic Physics (IASF) Via Corti 12, I-20133 Milano (MI), Italy*
[14]*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*
[15]*Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth PO1 3FX, UK*
[16]*Università di Trieste, Dipartimento di Fisica, sezione di Astronomia, via Tiepolo 11, I-34143 Trieste, Italy*
[17]*INAF-Osservatorio Astronomico di Trieste, via Tiepolo 11, I-34143 Trieste, Italy*
[18]*INFN, Sezione di Trieste, via Valerio 2, I-34127 Trieste, Italy*
[19]*IFPU - Institute for the Fundamental Physics of the Universe, Via Beirut 2, I-34014 Trieste, Italy*

Accepted 2020 January 24. Received 2020 January 17; in original form 2019 October 30

## ABSTRACT

We investigate the accuracy of weak lensing simulations by comparing the results of five independently developed lensing simulation codes run on the same input *N*-body simulation. Our comparison focuses on the lensing convergence maps produced by the codes, and in particular on the corresponding PDFs, power spectra, and peak counts. We find that the convergence power spectra of the lensing codes agree to $\lesssim 2$ per cent out to scales $\ell \approx 4000$. For lensing peak counts, the agreement is better than 5 per cent for peaks with signal-to-noise $\lesssim 6$. We also discuss the systematic errors due to the Born approximation, line-of-sight discretization, particle noise, and smoothing. The lensing codes tested deal in markedly different ways with these effects, but they none-the-less display a satisfactory level of agreement. Our results thus suggest that systematic errors due to the operation of existing lensing codes should be small. Moreover their impact on the convergence power spectra for a lensing simulation can be predicted given its numerical details, which may then serve as a validation test.

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

## 1 INTRODUCTION

The images of distant galaxies are weakly sheared due to the differential deflection of their light by the gravity of the intervening cosmic large-scale structure. This gravitational lensing effect is commonly referred to as weak lensing or cosmic shear (see e.g. Bartelmann & Schneider 2001; Kilbinger 2015; Mandelbaum 2018, for reviews), and it carries information about both the space–time geometry and the large-scale matter distribution of the Universe. Cosmic shear observations therefore prove extremely useful in tests of cosmological models and constraints on cosmological parameters, as the analyses of the Kilo Degree Survey (KiDS; Hildebrandt et al. 2017, 2020), the Dark Energy Survey (DES; Abbott et al. 2018; Troxel et al. 2018), and the Subaru Hyper

★ E-mail: hilbert@mpa-garching.mpg.de(SH); barreira@mpa-garching.mpg.de(AB)

Suprime-Cam (HSC) survey (Hamana et al. 2019; Hikage et al. 2019) have recently demonstrated.

Upcoming wide-field imaging surveys such as the *Euclid* satellite (Laureijs et al. 2011),[1] the Large Synoptic Survey Telescope (LSST Dark Energy Science Collaboration 2012)[2] (LSST), or the *Wide Field Infrared Survey Telescope* (Spergel et al. 2013)[3] (*WFIRST*) will allow us to measure the cosmic shear signal over a wide range of angular scales and redshifts with very small statistical uncertainties (e.g. Laureijs et al. 2011). To fully exploit this unprecedented statistical power, we require not only a thorough understanding of the systematic errors inherent in the measurements themselves, but also very accurate theoretical predictions for cosmic structure formation and its associated cosmic shear signal. Numerical simulations are an extremely valuable and widespread tool in gravitational lensing analysis. They can be used to compute predictions for gravitational lensing observables accurately in the non-linear/small-scale regime of structure formation (which can be used to calibrate faster semi-analytical methods to compute gravitational lensing predictions), as well as to build synthetic mock lensing data to be used in tests of different methods to measure the lensing shear signal from observations. Lensing simulations are, however, also subject to statistical and systematic errors themselves, and these must be well understood in order to appropriately apply and interpret their output.

Various methods for simulating gravitational lensing observations have been developed over the past few decades. They differ in the intended application, which naturally leads to differences in, for instance, the way the deflector mass is distributed and modelled, or in the method to compute the gravitational light deflection and distortion caused by the deflecting mass. For example, stars as deflectors in microlensing studies have been modelled as point masses (e.g. Paczynski 1986; Wambsganss, Paczynski & Katz 1990; Kochanek 2004). Analytic extended mass profiles have been used to represent the lensing mass of galaxies and clusters in strong lensing simulations (e.g. Blandford & Kochanek 1987; Grossman & Narayan 1988; Bezecourt, Pello & Soucail 1998; Metcalf & Madau 2001; Oguri 2002; Giocoli et al. 2012; Xu et al. 2015; Despali et al. 2018). Other simulations of the strong lensing signal caused by galaxies and clusters employ 'non-parametric' mass distributions extracted from *N*-body and hydrodynamical simulations, either alone or in combination with analytic mass profiles (e.g. Bartelmann & Weiss 1994; Meneghetti et al. 2000, 2017; Horesh et al. 2005, 2011; Puchwein et al. 2005; Hilbert et al. 2007, 2008; Puchwein & Hilbert 2009; Xu et al. 2009).

There is also a variety of methods to simulate the weak lensing signal caused by the large-scale distribution of matter in the Universe, which is the signal we focus on in this paper. Some simulations adopt analytical prescriptions (e.g. Kainulainen & Marra 2011; Giocoli et al. 2017) or realizations of lognormal distributed fields (e.g. Xavier, Abdalla & Joachimi 2016) to describe the large-scale mass distribution. A more common approach has been to use the matter distribution generated by *N*-body or hydrodynamical simulations of cosmic structure formation. Among these, many methods employ a multiple-plane (e.g. Wambsganss, Cen & Ostriker 1998; Jain, Seljak & White 2000; Hamana & Mellier 2001; Hilbert et al. 2009; Harnois-Déraps, Vafaei & Van Waerbeke 2012; Petkova, Metcalf & Giocoli 2014; Petri 2016) or multiple-sphere algorithm (e.g. Das & Bode 2008; Fosalba et al. 2008; Becker 2013; Fabbian, Calabrese & Carbone 2018; Gouin et al. 2019), in which the continuous mass distribution is projected on to a set of discrete 2D mass distributions that act as deflectors along the line-of-sight. There are also methods that bypass the line-of-sight discretization and make use of the 3D mass distribution of the *N*-body simulation outputs (e.g. Couchman, Barber & Thomas 1999; Vale & White 2003; Carbone et al. 2008; Kiessling et al. 2011; Li et al. 2011; Barreira et al. 2016; Breton et al. 2019)

The different weak-lensing simulation methods have different advantages and disadvantages in terms of speed and memory requirements, but importantly, the type and number of approximations made in them can also have an impact on their accuracy. Even for numerical codes that adopt overall the same lensing simulation method, there can still be differences in the final result, as different implementations may handle different approximations differently. Published work on weak lensing simulations usually contains tests of the correctness and accuracy of the numerical algorithms, but what is exactly tested and the way it is reported can vary widely, which makes comparisons across the literature hard. The importance to establish benchmark tests of the accuracy of these methods is therefore hard to overstate, specially given the ever increasing precision of the observational data (Krause et al. 2017).

Specifically in this paper, we investigate the current level of accuracy of weak lensing simulations by comparing the results of different lensing simulation codes ran on the same output of an *N*-body simulation of cosmic structure formation. We compare five codes: HILBERT (Hilbert et al. 2007, 2009) and MAPSIM (Giocoli et al. 2015), which are post-processing multiple-plane lensing codes; MICE (Fosalba et al. 2008, 2015) and LENS$^2$HAT (Fabbian & Stompor 2013; Calabrese et al. 2015), which are post-processing multiple-sphere codes; and RAY-RAMSES (Barreira et al. 2016), which runs on-the-fly with the simulation and uses the 3D distribution directly. We focus the comparison on the lensing convergence maps, their associated power spectra, and probability distribution functions (PDFs), as well as lensing peaks counts. We also comment on systematic errors such as those associated with the Born approximation, particle noise, smoothing, and the line-of-sight discretization for multiple-plane/sphere methods. Based on these results, we outline and discuss a procedure to help validate the output of weak lensing simulations and quantify their accuracy.

The code comparison we present in this paper adds to a large body of work on the validation of numerical *N*-body codes and algorithms to extract cosmological information from their output. Some of the codes/methods already subject to similar testing include: gravity-only *N*-body algorithms (Schneider et al. 2016), including non-standard gravity (Winther et al. 2015), galaxy formation codes (Scannapieco et al. 2012), methods to identify haloes/subhaloes (Knebe et al. 2011; Onions et al. 2012), galaxies (Knebe et al. 2013), voids (Colberg et al. 2008), and tidal debris (Elahi et al. 2013) from the output of *N*-body simulations, codes to construct halo merger trees (Srisawat et al. 2013) and fast generation of halo catalogues (Chuang et al. 2015), including comparing the covariances of their two- (Blot et al. 2019; Lippich et al. 2019) and three-point statistics (Colavincenzo et al. 2019). The importance of validation analyses such as these is crucial to identify and mitigate sources of systematic errors in our theoretical predictions. This serves as the main motivation for the weak-lensing code comparison analysis carried out here.

This paper is organized as follows. In Section 2, we review the main theoretical aspects of weak gravitational lensing and how accurately can one expect numerical simulation methods to operate.

[1] http://www.euclid-ec.org
[2] http://www.lsst.org
[3] https://wfirst.gsfc.nasa.gov

In Section 3, we describe the setup of our code comparison analysis, including the underlying $N$-body simulation of cosmic structure and the operation of the lensing simulation codes that participate in the comparison. Our main comparison results are presented and discussed in Section 4, where we also quantify systematic errors at play in lensing simulations. We summarize and conclude in Section 5.

## 2 THEORY

In this section, we display some of the main equations needed to understand weak-lensing observables, with emphasis on the calculation of two-point statistics, including how it can be affected by numerical resolution issues in lensing simulations (such as discretization of the line-of-sight and particle shot-noise).

### 2.1 Weak gravitational lensing

While travelling from source to observer, light will be deflected by the gravity of intervening matter structures (see e.g. Bartelmann & Schneider 2001; Kilbinger 2015; Mandelbaum 2018, for reviews). In a weakly perturbed Friedmann–Lemaître–Robertson–Walker universe, the angular (unobserved) position $\boldsymbol{\beta}$ of a source at comoving line-of-sight distance $\chi_{\rm s}$ and redshift $z_{\rm s} = z(\chi_{\rm s})$ and with (observed) image position $\boldsymbol{\theta}$ on the sky is given to very high accuracy by:

$$\boldsymbol{\beta}(\boldsymbol{\theta}, z_{\rm s}) = \boldsymbol{\theta} - \frac{2}{{\rm c}^2}\int_0^{\chi_{\rm s}} {\rm d}\chi_{\rm d}\, \frac{f_{\rm ds}}{f_{\rm d} f_{\rm s}} \nabla_{\boldsymbol{\beta}} \Phi\big(\boldsymbol{\beta}(\boldsymbol{\theta}, \chi_{\rm d}), \chi_{\rm d}, z_{\rm d}\big). \tag{1}$$

Here, c denotes the speed of light, $f_{\rm ds} = f_K(\chi_{\rm s} - \chi_{\rm d})$, $f_{\rm d} = f_K(\chi_{\rm d})$, and $f_{\rm s} = f_K(\chi_{\rm s})$, where $f_K(\chi)$ denotes the comoving angular diameter distance for comoving line-of-sight distance $\chi$, and $z_{\rm d} = z(\chi_{\rm d})$ the redshift corresponding to comoving line-of-sight distance $\chi_{\rm d}$. Furthermore, $\nabla_{\boldsymbol{\beta}}$ denotes the gradient w.r.t. the angular position $\boldsymbol{\beta}$, and $\Phi\big(\boldsymbol{\beta}, \chi_{\rm d}, z_{\rm d}\big)$ denotes the Newtonian gravitational potential at position $(\boldsymbol{\beta}, \chi_{\rm d})$ and redshift $z_{\rm d}$. On a flat sky, the Jacobian[4]

$$\frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2-\omega \\ -\gamma_2+\omega & 1-\kappa+\gamma_1 \end{pmatrix} \tag{2}$$

defines the lensing convergence $\kappa$, lensing shear $\gamma = \gamma_1 + {\rm i}\gamma_2$, and the lensing rotation $\omega$. From equation (1) it follows that

$$\frac{\partial \beta_i(\boldsymbol{\theta}, z_{\rm s})}{\partial \theta_j} = \delta_{ij} - \frac{2}{{\rm c}^2}\int_0^{\chi_{\rm s}} {\rm d}\chi_{\rm d}\, \frac{f_{\rm ds}}{f_{\rm d} f_{\rm s}} \times \frac{\partial^2 \Phi\big(\boldsymbol{\beta}(\boldsymbol{\theta}, \chi_{\rm d}), \chi_{\rm d}, z_{\rm d}\big)}{\partial\beta_i \partial\beta_k} \frac{\partial \beta_k\big(\boldsymbol{\theta}, \chi_{\rm d}\big)}{\partial \theta_j}, \tag{3}$$

where $\delta_{ij}$ is the Kronecker delta symbol. Replacing $\boldsymbol{\beta}$ on the RHS by the unperturbed position $\boldsymbol{\theta}$ in equation (3) yields the first-order (in $\Phi$) approximation:

$$\frac{\partial \beta_i(\boldsymbol{\theta}, z_{\rm s})}{\partial \theta_j} = \delta_{ij} - \frac{2}{{\rm c}^2}\int_0^{\chi_{\rm s}} {\rm d}\chi_{\rm d}\, \frac{f_{\rm ds}}{f_{\rm d} f_{\rm s}} \frac{\partial^2 \Phi\big(\boldsymbol{\theta}, \chi_{\rm d}, z_{\rm d}\big)}{\partial\theta_i \partial\theta_j}, \tag{4}$$

which is sometimes called the Born approximation.[5] Note that the rotation $\omega$ vanishes in this approximation.

If one makes use of the Poisson equation for the gravitational potential $\Phi$ and also neglects boundary terms at the observer and source, one obtains the following approximation for the convergence:

$$\kappa(\boldsymbol{\theta}, z_{\rm s}) = \int_0^{\chi_{\rm s}} {\rm d}\chi_{\rm d}\, q_{\rm ds}\, \delta_{\rm m}\big(\boldsymbol{\theta}, \chi_{\rm d}, z_{\rm d}\big) \tag{5}$$

with the lensing efficiency factor

$$q_{\rm ds} = \frac{3H_0^2\Omega_{\rm m}}{2{\rm c}^2}(1+z_{\rm d}) f_{\rm d} \frac{f_{\rm ds}}{f_{\rm s}}, \tag{6}$$

where $\delta_{\rm m}\big(\boldsymbol{\theta}, \chi_{\rm d}, z_{\rm d}\big)$ is the matter density contrast.

Assuming statistical isotropy, the two-point correlation $\xi_\kappa(\vartheta, z_{\rm s})$ of the convergence $\kappa$ for sources at angular separation $\vartheta$ and redshift $z_{\rm s}$ can then be written as:

$$\begin{aligned} \xi_\kappa(|\boldsymbol{\vartheta}|, z_{\rm s}) &= \langle \kappa(\boldsymbol{\theta}, z_{\rm s}) \kappa(\boldsymbol{\theta} + \boldsymbol{\vartheta}, z_{\rm s})\rangle \\ &= \int_0^{\chi_{\rm s}} {\rm d}\chi_{\rm d}\, q_{\rm ds} \int_0^{\chi_{\rm s}} {\rm d}\chi'_{\rm d}\, q'_{\rm ds} \\ &\quad \times \left\langle \delta_{\rm m}\big(\boldsymbol{\theta}, \chi_{\rm d}, z_{\rm d}\big) \delta_{\rm m}\big(\boldsymbol{\theta} + \boldsymbol{\vartheta}, \chi'_{\rm d}, z'_{\rm d}\big)\right\rangle. \end{aligned} \tag{7}$$

Here, $\langle \ldots \rangle$ denotes the expectation for a given (statistically homogeneous and isotropic) ensemble of universes.

Assuming $\delta_{\rm m}$ is slowly evolving with redshift and matter correlations are short-ranged compared to $\chi_{\rm s}$, one can apply a Limber-type approximation to obtain:

$$\xi_\kappa(\vartheta, z_{\rm s}) = \int_0^{\chi_{\rm s}} {\rm d}\chi_{\rm d}\, q_{\rm ds}^2\, \xi_{\rm m}^\perp\big(\vartheta\chi_{\rm d}, z_{\rm d}\big), \tag{8}$$

where

$$\xi_{\rm m}^\perp\big(R, z\big) = \int_{-\infty}^{\infty} {\rm d}L\, \xi_{\rm m}\big(\sqrt{R^2 + L^2}, z\big), \tag{9}$$

is the line-of-sight projection of the 3D two-point correlation function $\xi_{\rm m}(r, z)$ of the matter density contrast at comoving separation $r$ at redshift $z$. The corresponding approximation to the spherical harmonic power spectrum $C_\kappa(\ell)$ of the convergence as a function of harmonic wavenumber $\ell$ reads:

$$C_\kappa(\ell) = \int_0^{\chi_{\rm s}} {\rm d}\chi_{\rm d}\, q_{\rm ds}^2\, P_{\rm m}(\ell/\chi_{\rm d}, z_{\rm d}), \tag{10}$$

where $P_{\rm m}(k, z_{\rm d})$ denotes the 3D matter power spectrum for wavenumber $k$ at redshift $z_{\rm d}$.

In cosmic-shear surveys, one aims to estimate the shear from the observed shapes of galaxy images. Given a direction $\boldsymbol{\vartheta}$ on the sky, one may define a tangential and a cross shear component as

$$\gamma_{\rm T}(\boldsymbol{\theta}, z_{\rm s}, \boldsymbol{\vartheta}) = \Re\left[-{\rm e}^{-2{\rm i}\varphi(\boldsymbol{\vartheta})}\gamma(\boldsymbol{\theta}, z_{\rm s})\right], \tag{11a}$$

$$\gamma_\times(\boldsymbol{\theta}, z_{\rm s}, \boldsymbol{\vartheta}) = \Im\left[-{\rm e}^{-2{\rm i}\varphi(\boldsymbol{\vartheta})}\gamma(\boldsymbol{\theta}, z_{\rm s})\right], \tag{11b}$$

where $\varphi(\boldsymbol{\vartheta})$ is the position angle of $\boldsymbol{\vartheta}$. Using these, one may define the shear correlation functions

$$\begin{aligned} \xi_\pm(|\boldsymbol{\vartheta}|, z_{\rm s}) &= \langle \gamma_{\rm T}(\boldsymbol{\theta}, z_{\rm s}, \boldsymbol{\vartheta}) \gamma_{\rm T}(\boldsymbol{\theta} + \boldsymbol{\vartheta}, z_{\rm s}, \boldsymbol{\vartheta})\rangle \\ &\quad \pm \langle \gamma_\times(\boldsymbol{\theta}, z_{\rm s}, \boldsymbol{\vartheta}) \gamma_\times(\boldsymbol{\theta} + \boldsymbol{\vartheta}, z_{\rm s}, \boldsymbol{\vartheta})\rangle. \end{aligned} \tag{12}$$

[4] For simplicity, we discuss these equations for a flat sky with $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ as 2D Cartesian coordinate vectors. For a spherical sky, the partial derivatives w.r.t. the angular positions have to be replaced by covariant derivatives on the sphere (Becker 2013).

[5] The term 'Born approximation' is not used uniformly throughout the literature, and sometimes instead refers to an equation obtained by just replacing $\partial_{\beta_i}\partial_{\beta_k}\Phi(\boldsymbol{\beta}, \chi_{\rm d}, z_{\rm d})$ by $\partial_{\theta_i}\partial_{\theta_k}\Phi(\boldsymbol{\theta}, \chi_{\rm d}, z_{\rm d})$ in equation (3), but keeping the factor $\partial_{\theta_j}\beta_k$.

Within the first-order approximation, we have that

$$\xi_+(\vartheta, z_s) = \xi_\kappa(\vartheta, z_s) \tag{13a}$$

$$\xi_-(\vartheta, z_s) = \xi_\kappa(\vartheta, z_s) + \int_0^\vartheta \mathrm{d}\vartheta' \left( \frac{4\vartheta'}{\vartheta^2} - \frac{12\vartheta'^3}{\vartheta^4} \right) \xi_\kappa(\vartheta', z_s). \tag{13b}$$

The shear $\gamma$ transforms like a spin-2 field and thus can be decomposed into rotationally invariant E and B modes (Stebbins 1996; Kamionkowski, Kosowsky & Stebbins 1997; Zaldarriaga & Seljak 1997). When employing the flat-sky, first-order (Born), and Limber approximations, the shear E- $C_\gamma^{(\mathrm{EE})}(\ell)$ and B-mode $C_\gamma^{(\mathrm{BB})}(\ell)$ power spectra obey the following relations:

$$C_\gamma^{(\mathrm{EE})}(\ell) = C_\kappa(\ell), \tag{14a}$$

$$C_\gamma^{(\mathrm{BB})}(\ell) = C_\omega(\ell) = 0, \tag{14b}$$

where $C_\omega(\ell)$ is the lensing rotation power spectrum.

Corrections to the flat-sky approximation for the convergence and shear power spectra become relevant only on very large scales with $\ell < 10^2$ (e.g. Hu 2000; Castro, Heavens & Kitching 2005; Becker 2013; Kilbinger et al. 2017). For example, the relation between the convergence and shear power spectrum is modified by a factor $[(\ell+2)(\ell-1)]/[\ell(\ell+1)]$ on a spherical sky, which differs by $< 1$ per cent from unity for $\ell \geq 15$. Corrections to the Limber approximation also only become relevant on very large scales. For $\ell \gtrsim 10^2$, such corrections to the convergence and shear power spectra are expected to be below 1 per cent (Angulo & Hilbert 2015; Kilbinger et al. 2017).

Second- and higher order contributions in the gravitational potential $\Phi$ to the convergence and shear yield small corrections to their power spectra. These are expected to be at least two orders of magnitude below the leading order terms for $10^2 \lesssim \ell \lesssim 10^4$ (see e.g. Cooray & Hu 2002; Hirata & Seljak 2003; Shapiro & Cooray 2006; Hilbert et al. 2009; Krause & Hirata 2010; Petri, Haiman & May 2017a). Note, however, that beyond-Born corrections may be more significant for other weak-lensing quantities. In particular, the galaxy–galaxy lensing shear profiles (e.g. Ziour & Hui 2008; Hilbert et al. 2009; Ghosh, Durrer & Sellentin 2018; Simon & Hilbert 2018), higher order moments of the convergence field (Petri et al. 2017a) and lensing bispectrum (Pratten & Lewis 2016) may require one to go beyond first order and undeflected light rays (Petri, Haiman & May 2017b).

Higher order terms also cause a non-vanishing rotation $\omega$, and create a small B-mode component in the shear field that is likely below the detection threshold even for upcoming galaxy lensing surveys. The amplitude of these corrections depends however on the distance to the lensing sources (the farther the sources the higher the *number of deflections*), which has been motivating studies of its importance for future analysis of the lensing of the cosmic microwave background (Lewis & Pratten 2016; Marozzi et al. 2016; Pratten & Lewis 2016; Fabbian et al. 2018).

### 2.2 Weak lensing power spectra in simulations

The convergence and shear power spectra measured from simulated weak-lensing observations may differ from the pure theory predictions (10) and (14a,b) for a number of reasons. First, there is sample variance because every simulation set covers only a finite number of realizations of a given area fraction and depth. The impact of sample variance can however be estimated analytically, by resampling techniques, or by generating many realizations of the simulated lensing maps.

Secondly, there is the issue of the intrinsic accuracy of equations (10) and (14a,b), which assume a flat sky and use the Born and the Limber approximation. As we noted in the previous subsection, corrections from going beyond these approximations on the convergence and shear power spectra are however expected to be well below 1 per cent for $10^2 \lesssim \ell \lesssim 10^4$, which is sufficiently small even for surveys like *Euclid*, LSST, or *WFIRST*. Hence, taking the validity of these approximations as established thus allows one to actually use them in internal self-consistency tests: codes ran in and out of these approximations should return spectra that differ comfortably by less than 1 per cent on the relevant scales.

Thirdly, the methods to measure the convergence and shear from the simulations may introduce biases. In this work, we do not consider possible issues due to, e.g. differences between shear $\gamma$ and reduced shear $g = \gamma/(1-\kappa)$ (which is more closely related to observed galaxy ellipticities) or noisy and biased image shape measurements. Instead we assume that we have bias- and noise-free measurements of the convergence (except in Section 4.3, when we also add Gaussian random noise to the simulated convergence maps before we count lensing peaks).

Fourthly, numerical approximations (smoothing, line-of-sight projections, particle discreteness, etc.) employed in the lensing simulations may impact the simulated convergence and shear power spectra in various ways. However, with sufficient knowledge of the numerical details of the simulations, one may be able to account for some of these effects in a modified theory prediction for the convergence and shear power spectra. One may then compare these modified predictions with the measured power spectra as part of a validation procedure for the lensing simulations (see e.g. Section 3 of Fosalba et al. 2008).

As an example, consider a lensing simulation based on an $N$-body simulation of cosmic structure formation. In that case, the matter power spectrum $P_\mathrm{m}^{(\mathrm{sim})}(k, z)$ going into the lensing simulation can be described by a continuous component and a shot noise component due to sampling of the density field by a finite number of particles/mass elements in the simulation:

$$P_\mathrm{m}^{(\mathrm{sim})}(k, z) = P_\mathrm{m}^{(\mathrm{cont})}(k, z) + P_\mathrm{m}^{(\mathrm{sn})}(k, z). \tag{15}$$

The continuous component $P_\mathrm{m}^{(\mathrm{cont})}(k, z)$ should closely resemble the theoretical power spectrum $P_\mathrm{m}(k, z)$ at least on the scales represented well by the $N$-body simulation, but deviates more strongly on very large and very small scales due to the finite simulation box size and resolution. The shot-noise term $P_\mathrm{m}^{(\mathrm{sn})}(k, z)$ is usually sufficiently well described by white noise with an amplitude given by the inverse particle density, $P_\mathrm{m}^{(\mathrm{sn})}(k, z) = V N_\mathrm{p}^{-1}$, where $V$ denotes the simulation volume, and $N_\mathrm{p}$ denotes the simulation particle number (assumed to have the same mass).

As a first step to adjust the prediction for the convergence power spectra from the simulation, one may replace the theoretical power spectrum with the power spectrum measured from the $N$-body simulation:

$$C_\kappa^{(\mathrm{sim})}(\ell) = \int_0^{\chi_s} \mathrm{d}\chi_\mathrm{d}\, q_\mathrm{ds}^2\, P_\mathrm{m}^{(\mathrm{sim})}(\ell/\chi_\mathrm{d}, z_\mathrm{d}). \tag{16}$$

Further, if the matter distribution of the simulation is only available at a finite number of snapshots at redshifts, $z_\mathrm{d}^{(i)}$, $i = 1, \ldots,$ one can write

$$C_\kappa^{(\mathrm{sim})}(\ell) = \sum_i \int_{\chi_\mathrm{d}^{(i,\mathrm{lo})}}^{\chi_\mathrm{d}^{(i,\mathrm{hi})}} \mathrm{d}\chi_\mathrm{d}\, q_\mathrm{ds}^2\, P_\mathrm{m}^{(\mathrm{sim})}\left(\ell/\chi_\mathrm{d}, z_\mathrm{d}^{(i)}\right), \tag{17}$$

where $\chi_\mathrm{d}^{(i,\mathrm{lo})}$ and $\chi_\mathrm{d}^{(i,\mathrm{hi})}$ denote the lower and upper boundary of the slice of the observer's backward light-cone filled by the matter

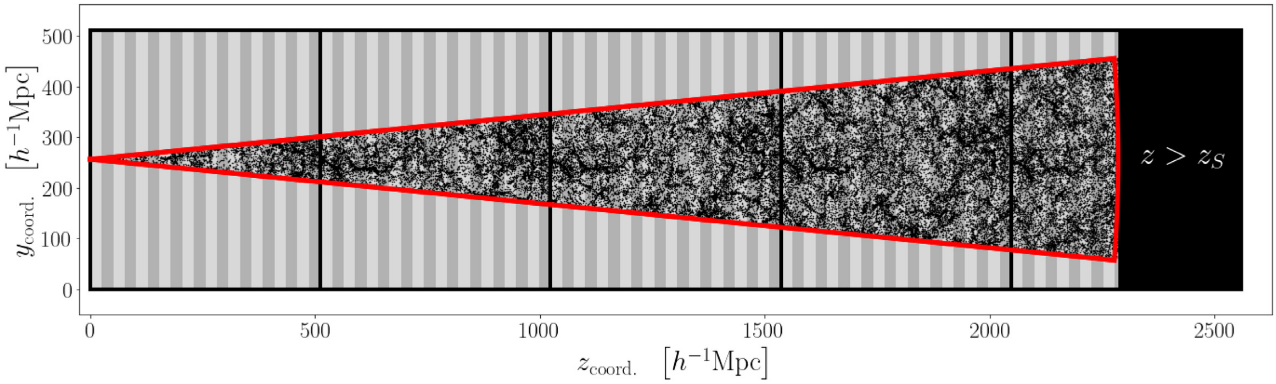


**Figure 1.** Sketch of the light-cone setup. The black solid lines represent the five 512 Mpc h$^{-1}$ simulation boxes that are tiled to encompass the 10 × 10 deg$^2$ light-cone that extends out to $z_{\rm s} = 1$. The light-cone geometry is depicted by the red solid line. The grey shaded slices mark the comoving distance (in the (0,0,1) ($z_{\rm coord.}$) direction) covered by each of the 90 simulation snapshots. The comoving volume of the right-most simulation box that corresponds to $z > z_{\rm s}$ (black) is never used in the calculations. In this work, all five simulation boxes correspond to the same *N*-body simulation, and hence, there is repetition of structure along the line-of-sight (this is however not important for our code comparison results).

distribution from snapshot $i$ at redshift $z_{\rm d}^{(i)}$. If the lensing simulation employs lens planes at distances $\chi_{\rm d}^{(i)}$ with one plane per snapshot and with the matter projected on to lens planes with parallel projection, then we can write

$$C_{\kappa}^{(\rm sim)}(\ell) = \sum_i \left(\chi_{\rm d}^{(i,{\rm hi})} - \chi_{\rm d}^{(i,{\rm lo})}\right) \times q_{\rm ds}^{(i)2}\, P_{\rm m}^{(\rm sim)}\left(\ell/\chi_{\rm d}^{(i)}, z_{\rm d}^{(i)}\right), \tag{18}$$

where $q_{\rm ds}^{(i)}$ is the lensing efficiency at $z_{\rm d}^{(i)}$. Finally, if all of the binning, interpolation, and smoothing schemes employed by the lensing simulations in their various steps can be effectively described by some convolution of the 3D matter distribution with a window function $W$ with a fixed comoving smoothing scale, then the prediction for the simulated convergence power spectrum can be written as:

$$C_{\kappa}^{(\rm sim)}(\ell) = \sum_i \left(\chi_{\rm d}^{(i,{\rm hi})} - \chi_{\rm d}^{(i,{\rm lo})}\right) \times q_{\rm ds}^{(i)2}\, \hat{W}\left(\ell/\chi_{\rm d}^{(i)}\right) P_{\rm m}^{(\rm sim)}\left(\ell/\chi_{\rm d}^{(i)}, z_{\rm d}^{(i)}\right), \tag{19}$$

where $\hat{W}(k)$ denotes the Fourier transform of filter $W$ (see Section 4.4.3 for a specific application).

## 3 METHODS

In this section, we describe the numerical methods that participate in this comparison project. This includes the lensing simulation codes themselves (HILBERT, LENS$^2$HAT, MAPSIM, MICE, and RAY-RAMSES), as well as the $N$-body simulation that provides the common realization of large-scale structure on which they perform their calculations.

### 3.1 *N*-body simulation and light-cone geometry

We base our lensing simulation code comparison on a common $N$-body simulation of cosmic structure formation in a spatially flat standard cold-dark-matter cosmology with a cosmological constant. The assumed cosmological parameters are: a matter density parameter $\Omega_{\rm m} = 0.32$, a baryon density parameter $\Omega_{\rm b} = 0.049$, an amplitude parameter $\sigma_8(z = 0) = 0.83$, spectral index $n_{\rm s} = 0.96$ for the initial density power spectrum, and a Hubble constant $H_0 = h100\,{\rm km\,s^{-1}\,Mpc^{-1}}$ with $h = 0.67$. In this work, we ignore the effects of the energy density of radiation, as well as massive neutrinos.

The simulation is carried out with the adaptive-mesh-refinement (AMR) code RAMSES in a cubic box of side length $L = 512\,h^{-1}$ Mpc with $1024^3$ matter tracer particles from $z = 99$ to redshift $z = 0$. The interpolation between the particle positions and the positions of the AMR mesh cells (needed to construct the density on the AMR grid and also evaluate the forces at the particle positions) is done with a cloud-in-cell (CIC) scheme. The cell refinement criterion adopted is 8, i.e. an AMR cell is split into eight child-cells if the particle number in the cell exceeds 8. The initial conditions are generated with second-order Lagrangian perturbation theory using the routines implemented in the PINOCCHIO code (Monaco, Theuns & Taffoni 2002; Munari et al. 2017), which follow closely those of the initial conditions code 2LPTIC (Scoccimarro 1998). The initial conditions are generated using the linear power spectrum computed by the CAMB code (Lewis, Challinor & Lasenby 2000) at $z = 0$, with its amplitude rescaled to $z = 99$ using the linear growth factor of our fiducial cosmology. This ensures that the non-negligible contribution of radiation at $z = 99$, that CAMB captures but which we omit from the $N$-body simulation, does not cause discrepancies at the lower redshifts of interest.

In this work, we consider a single galaxy source redshift $z_{\rm s} = 1$ and the particle information in the simulation is outputted at 90 redshift values between $z = z_{\rm s} = 1$ and $z = 0$. The output times are equally spaced in comoving distance $\chi(z)$ with intervals $\Delta\chi = 25.6\,h^{-1}$Mpc. By tiling up these 90 simulation snapshots one can construct a realization of the time evolution of cosmic large-scale structure that encompasses a light-cone with 10 × 10 deg$^2$ out to $\chi_{\rm s} = \chi(z_{\rm s}) \approx 2286\,h^{-1}$ Mpc. We place the observer at ($L$/2, $L$/2, 0) in Cartesian box coordinates, with the central line of sight of the observer's backward light-cone chosen along the (0,0,1) direction. This lensing light-cone geometry is depicted in Fig. 1.

All 90 snapshots are from the same $N$-body simulation, and as a result, the same large-scale structures appear at different epochs along the light-cone (most notably closer to the centre of the field-of-view). This issue can be avoided by running five different $N$-body simulations, each providing 20 snapshots along the light-cone (cf. five black boxes drawn in Fig. 1). When comparing results using just one $N$-body simulation to results using five simulations for tests, we do not detect any significant impact on our results,

**Table 1.** Summary of the key features of the lensing simulation codes compared in this paper. The entries 'w/ development' indicate that the code versions used do not immediately admit the corresponding feature, but that there is no impediment for it to be implemented with further development.

| Name | HILBERT | LENS$^2$HAT | MAPSIM | MICE | RAY-RAMSES |
|---|---|---|---|---|---|
| Code paper | Hilbert et al. (2009) | Fabbian et al. (2018) | Giocoli et al. (2015) | Fosalba et al. (2008) | Barreira et al. (2016) |
| Code type | Post-process (multiple plane) | Post-process (multiple sphere) | Post-process (multiple plane) | Post-process (multiple sphere) | On the fly |
| LOS projection | $\parallel$ to central LOS | Radial | Radial | Radial | Radial |
| LOS resolution | Particle outputs | Particle outputs | Particle outputs | Particle outputs | RAMSES time steps |
| Ray grid scheme | Regular grid | HEALPIX[6] | Regular grid | HEALPIX | Regular grid |
| Full-sky maps | w/ development | ✓ | w/ development | ✓ | w/ development |
| Beyond-Born | ✓ | ✓ | w/ development | w/ development | w/ development |

[6]Despite adopting a HEALPIX grid for this work, LENS$^2$HAT supports ray-tracing on arbitrary isolatitudinal grids on the sphere that are symmetric with respect to the equator. See Fabbian & Stompor (2013) and references therein for more details.

which focus on the differences between numerical methods for a fixed realization of the large-scale structure (with the exact degree of realism of such a realization being of secondary importance). We thus report only the results using a single $N$-body simulation for brevity.

We note also that our main goal is to assess the accuracy of the various lensing codes on their lensing predictions for small angular scales, for which our simulated field of view is sufficient. Our field of view is appreciably smaller than the expected area for Euclid (100 deg$^2$ versus 15 000 deg$^2$), but our comparison analysis remains of interest for such wide-field imaging surveys since a significant portion of the constraining power comes from the smallest scales. Further, on larger angular scales, calculations based on linear theory are sufficient and the lensing codes tested here are not strictly needed to obtain theoretical predictions.

### 3.2 Lensing simulations

We refer to the lensing simulation codes that participate in this comparison project as the HILBERT (Hilbert et al. 2007, 2009), LENS$^2$HAT (Fabbian & Stompor 2013; Calabrese et al. 2015; Fabbian et al. 2018), MAPSIM (Giocoli et al. 2015), MICE (Fosalba et al. 2008, 2015), and RAY-RAMSES (Barreira et al. 2016) codes. The main data produced by these codes for this analysis are lensing convergence maps with 2048$^2$ pixels on a regular mesh that covers a $10 \times 10$ deg$^2$ field of view. This default pixel resolution corresponds to an angular resolution of $\approx 18$ arcsec (due to the details of their operation, the LENS$^2$HAT and MICE codes will work at slightly lower resolution).

The main goal of this paper is to evaluate the level of agreement between these maps for a number of different summary statistics. In the remainder of this section, we describe, in alphabetical order, the main aspects of the operation of these lensing simulation codes. We shall be succinct in the description and we refer the interested reader to the relevant cited literature for more details on their operation. Table 1 summarizes the main key features of the lensing simulation codes.

#### 3.2.1 HILBERT

The HILBERT lensing simulation code is described in Hilbert et al. (2007, 2009). The code implements a multiple-lens-plane algorithm in the flat-sky approximation, and it is capable of computing convergence and shear fields both in full ray-tracing mode with multiple light deflections and lens–lens coupling, and within the Born approximation with unperturbed ray trajectories and without lens–lens coupling.

In the first step of the code operation, the backward light-cone up to the source redshift $z_s = 1$ is divided into 90 redshift slices of 25.6 $h^{-1}$ Mpc comoving thickness. Each slice is filled with the matter distribution of a snapshot of the $N$-body simulation. The matter of each slice is projected parallel to the (0,0,1) direction on to a lens plane located at the centre of the slice. From the matter on the lens planes, the 2D lensing potential and its first and second derivatives are computed using a particle-mesh particle-mesh (PMPM) method. On each lens plane, a coarse mesh with 16 384$^2$ mesh points and side length 512 $h^{-1}$ Mpc spans the whole cross-section of the box. One or more finer meshes with 8192$^2$ mesh points and 5 $h^{-1}$ kpc mesh spacing are used to cover the intersection of the light-cone with each lens plane. The simulation particles are assigned to the 2D mesh points using CIC assignment. The matter distribution on the meshes is smoothed further with a Gaussian kernel with a constant comoving width $\sigma_G = 10\,h^{-1}$ kpc per dimension. Thus the smoothing is well captured by a kernel $\hat{W}$ in comoving harmonic space that is a product of the (circularly averaged) CIC mass assignment kernel and a Gaussian smoothing kernel:

$$\hat{W}(k) = \mathrm{sinc}\left(\frac{d_{\mathrm{fine}} k}{2}\right)^4 \exp\left(-\frac{\sigma_G^2 k^2}{2}\right), \tag{20}$$

where $k$ denotes the comoving wavenumber, and $d_{\mathrm{fine}} = 5\,h^{-1}$ kpc denotes the fine mesh spacing.

The coarse meshes are used to compute a long-range low-pass filtered version of the lensing potential from the projected density using Fast Fourier Transforms (FFT). From that long-range potential, the first and second derivatives on the mesh points are computed using finite differencing. Similarly, the fine meshes are used to compute the complementary high-pass filtered short-ranged part of the lensing potential and its derivatives.

Light rays are then traced back from the observer through the series of lens planes to the source plane. In full ray-tracing mode, deflection angles at each lens plane are computed by bilinear interpolation of the first derivatives of the lensing potential at the mesh points on to the ray position. The lensing Jacobians of the rays are updated using the second derivatives of the lensing potential. In Born mode, light rays are not deflected and the lensing Jacobians of the rays are computed by sums of the second derivatives with appropriate lensing efficiency weights.

#### 3.2.2 LENS$^2$HAT

LENS$^2$HAT implements a multiple lens ray-tracing algorithm in spherical coordinates on the full sky using the approaches of Fosalba et al. (2008) and Das & Bode (2008), and was originally developed to perform high-resolution CMB lensing simulations (Fabbian & Stompor 2013; Fabbian et al. 2018). The current version of the code reconstructs the full-sky backward light-cone around the observer using the particle snapshots produced by an $N$-body simulation out to the comoving distance of the highest redshift available from the snapshots following Calabrese et al. (2015). Because of the finite size of an $N$-body simulation box, the code replicates the box volume the necessary number of times in space to fill the entire observable volume between the observer and the source plane. In order to minimize systematics arising from the box replica (and thus by the repetition of the same structures along the line of sight) the code can randomize the particle positions as described in Carbone et al. (2008), Carbone et al. (2009). However, this randomization is not performed here to ensure the LENS$^2$HAT code 'sees' the same large-scale structures as the other codes. Furthermore, the results reported here focus only of the $10 \times 10\,\mathrm{deg}^2$ portion of the sky that is common to all codes (cf. Fig. 1).

The backward light-cone is sliced into 90 full-sky spherical shells such that the median comoving distance spanned by each shell coincides with the comoving distance at the redshift of each $N$-body snapshot. The particles inside each of these shells are projected on to spheres. The surface mass density $\Sigma$ on each sphere is defined on a 2D grid. For each pixel of the $i$-th sphere one has

$$\Sigma^{(i)}(\boldsymbol{\theta}) = \frac{nm}{A_{\mathrm{pix}}}\,, \tag{21}$$

where $n$ is the number of particles in the pixel, $A_{\mathrm{pix}}$ is the pixel area in steradians, and $m$ is the particle mass of the $N$-body simulation in each pixel.

For this work, the LENS$^2$HAT code produces a full-sky convergence map on a HEALPIX[7] grid (Górski et al. 2005) with $N_{\mathrm{side}} = 8192$, which corresponds to a pixel resolution of 26 arcsec. The lensing convergence of a source plane at redshift $z_{\mathrm{s}}$ in the Born approximation is computed as the lensing-efficiency-weighted sum of the surface mass density:

$$\kappa(\boldsymbol{\theta}, \chi_{\mathrm{s}}) = \frac{4\pi \mathrm{G}}{\mathrm{c}^2}\frac{1}{f_{\mathrm{s}}}\sum_i \left(1 + z_{\mathrm{d}}^{(i)}\right)\frac{f_{\mathrm{ds}}^{(i)}}{f_{\mathrm{d}}^{(i)}}\left[\Sigma^{(i)}(\boldsymbol{\theta}) - \bar{\Sigma}^{(i)}\right], \tag{22}$$

where $\Sigma^{(i)}$ denotes the angular surface mass density, $\bar{\Sigma}^{(i)}$ is the mean angular surface mass density of the $i$-th shell, and $f_{\mathrm{ds}}^{(i)}$ and $f_{\mathrm{d}}^{(i)}$ are the corresponding distances at the redshift of the $i$-th shell. The angular position of the centre of each HEALPIX pixel coincides with the direction of propagation of the rays in the Born approximation.

The common sky patch for the code comparison is extracted from the HEALPIX map with a Lambert azimuthal equal area projection using 18 arcsec pixels.[8] We correct for the effect of projecting the HEALPIX map to the higher resolution flat-sky map in the convergence power spectrum estimation by multiplying $C_\kappa(\ell)$ by a pixel window function $w_\ell^2$ estimated as follows. We first synthesize 100 Gaussian realizations of a convergence field on an $N_{\mathrm{side}} = 8192$ HEALPIX grid from a theoretical power spectrum $C_\kappa^{\mathrm{MC}}(\ell)$ at $z_{\mathrm{s}} = 1$ using the HEALPIX `synfast` routine. These realizations are then similarly projected to a flat-sky and their power spectrum measured. The pixel window function $w_\ell$ is then defined as

$$w_\ell^{-2} = \frac{\langle C_\kappa^{\mathrm{MC,proj.}}(\ell)\rangle}{C_\kappa^{\mathrm{MC}}(\ell)}, \tag{23}$$

where the angular brackets denote the average over the power spectrum measured from each projected Gaussian field $C_\kappa^{\mathrm{MC,proj.}}(\ell)$.

LENS$^2$HAT can also be used to propagate the lensing Jacobian beyond the Born approximation (Fabbian et al. 2018). In this case the 2D lensing potential and its derivatives required by the multiple-lens algorithm are computed in the harmonic domain by solving the Poisson equation, and later resampled on a higher resolution ECP pixelization (Muciaccia, Natoli & Vittorio 1997) that can reach the arcsec resolution. The perturbed ray trajectories are then computed using a nearest grid point interpolation scheme.

### 3.2.3 MAPSIM

The MAPSIM code has been developed by Giocoli et al. (2015) and it works in two main steps termed i-MAPSIM and ray-MAPSIM. In the first step i-MAPSIM, the particle positions in the simulation snapshots that lie within the desired field-of-view are projected on to different lens planes located along the line of sight. Each particle is placed in the nearest lens plane maintaining angular positions. The mass density is then interpolated from the projected particle positions to a 2D grid using a triangular shaped cloud (TSC) scheme. The grid pixels are chosen to have the same angular size on all lens planes and no particle randomization is performed to ensure MAPSIM calculates the lensing signal using the same large-scale structure as the other codes. The angular surface mass density $\Sigma^{(i)}$ on the $i$-th plane is computed as in equation (21).

In the second step (as done by Petri, Haiman & May 2016; Giocoli et al. 2017, 2018; Petri et al. 2017b; Castro et al. 2018)[9] ray-MAPSIM constructs the lensing convergence map in the Born approximation by simply summing up the surface mass density from each plane along the line-of-sight, weighted appropriately by the lensing efficiency kernels as in equation (22), except now $i$ labels planes perpendicular to the (0,0,1) direction instead of concentric spheres. Since the grid pixels have the same angular resolution, they are also the direction of propagation of the light rays in the Born approximation. There is thus no need to interpolate the projected density defined on the grid to the exact light ray position.

### 3.2.4 MICE

The methodology of the MICE code presented in Fosalba et al. (2008, 2015) guided the development of the LENS$^2$HAT, and hence, the two codes work very similarly. The $N$-body data are sliced and projected on to spherical shells (called the 'onion universe' in Fosalba et al. 2008). Each such shell is used to define surface mass density fields on HEALPIX grids. Finally, the surface mass is summed along the line-of-sight, weighted by the appropriate weak lensing efficiency factors.

The MICE lensing maps produced for this comparison project take the HEALPIX data from LENS$^2$HAT as input. The lensing convergence is then calculated *independently of* LENS$^2$HAT employing the Born approximation. The common sky patch is extracted from the HEALPIX map with a Lambert azimuthal equal area projection using 18 arcsec pixels, and the power spectra measured from the patch are corrected for the projection.

[7] http://healpix.sourceforge.net

[8] We use the `azeqview` routine of the HEALPIX python package HEALPY.

[9] The mass maps produced in this first step i-MAPSIM can also be used as input to the GLAMER lensing code (Metcalf & Petkova 2014; Petkova et al. 2014) to perform ray-tracing calculations (including beyond the Born approximation).

#### 3.2.5 RAY-RAMSES

The RAY-RAMSES code is described in detail in Barreira et al. (2016). The computation of the lensing quantities is based on the original ideas of White & Hu (2000), Li et al. (2011) and it is done on-the-fly during the RAMSES $N$-body simulation that produced the 90 snapshots that serve as input to the other codes. This code therefore does not rely on any discretization of the density field along the line-of-sight, or more precisely, it retains the full line-of-sight (or time) resolution attained by the $N$-body simulation itself. Likewise, RAY-RAMSES also bypasses the need to choose a density assignment scheme to construct 2D projected density planes from the 3D density distribution, i.e. the transverse spatial resolution is directly specified by the AMR grid structure of the RAMSES simulation.

In RAY-RAMSES, light rays are initialized at the curved surface of constant $z_\mathrm{s} = 1$ and then subsequently followed in unperturbed trajectories until they reach the observer at $z = 0$. In each simulation particle time-step, the rays are moved by the distance light would travel during that time interval. The size of the time-steps depends on the refinement of the grid in RAMSES: rays located in refined regions are integrated more often than rays in unrefined regions. The lensing integral accumulated during a time-step is the sum of the lensing integral associated with each crossed cell:[10]

$$\kappa_\mathrm{ts} = \frac{1}{\mathrm{c}^2} \sum_{\mathrm{cells}} \int_{\chi_\mathrm{d}^\mathrm{cell,end}}^{\chi_\mathrm{d}^\mathrm{cell,start}} \mathrm{d}\chi_\mathrm{d} \frac{f_\mathrm{d} f_\mathrm{ds}}{f_\mathrm{s}} \nabla_\mathrm{2D}^2 \Phi, \tag{24}$$

where $\chi_\mathrm{d}^\mathrm{cell,start}$ and $\chi_\mathrm{d}^\mathrm{cell,end}$ are the comoving distances between the observer and ray at the start and end of its trajectory inside a given mesh cell, and the sum runs over all of the cells crossed by the ray during the time-step. The total lensing convergence is the sum of the convergence accumulated during all of the simulation time-steps $\kappa = \sum_{ts} \kappa_\mathrm{ts}$.

The quantity $\nabla_\mathrm{2D}^2 \Phi = \nabla_1 \nabla^1 \Phi + \nabla_2 \nabla^2 \Phi$ is the 2D Laplacian of the gravitational potential ($\nabla_1$ and $\nabla_2$ represent the curved-sky generalizations of $\partial/\partial\theta_1$ and $\partial/\partial\theta_2$ in Section 2.1), which is related to the 3D second-derivative $\nabla_i \nabla_j \Phi$ ($i, j = x, y, z$) via geometrical factors determined by the direction of motion of the ray. The values of $\nabla_i \nabla_j \Phi$ are evaluated by finite-differencing the potential in neighbouring cells, analogously to how standard RAMSES computes the force. In the calculations presented in this paper, the values of $\nabla_i \nabla_j \Phi$ are treated as constant inside each cell.[11] RAY-RAMSES can evaluate also the shear components $\gamma_1$ and $\gamma_2$ by replacing $\nabla_\mathrm{2D}^2 \Phi$ in equation (24) with $\nabla_1 \nabla^1 \Phi - \nabla_2 \nabla^2 \Phi$ and $2\nabla_1 \nabla^2 \Phi$, respectively. The calculation of these quantities for a given cell involves the information from a larger number of neighbouring cells compared to the calculation of $\nabla_\mathrm{2D}^2 \Phi$ for $\kappa$. This constitutes an effective form of smoothing that explains some differences between the convergence and shear power spectra of RAY-RAMSES. We will return to this point below in Section 4.4.4.

As it runs on-the-fly with the simulation, the generation of lensing maps using RAY-RAMSES with certain specifications modified (e.g. increased number of pixels) requires rerunning the $N$-body simulation. All of the RAY-RAMSES results shown below correspond to the default $2048^2$ ($\approx 18$ arcsec) map resolution.

[10]This is integration 'method B' in Barreira et al. (2016).

[11]RAY-RAMSES allows also to evaluate $\nabla_i \nabla_j \Phi$ inside each cell via trilinear interpolation using the quantity's value reconstructed at the cell vertices (see Barreira et al. 2016, for the details).

## 4 RESULTS

In this section, we compare the various lensing codes by analysing a number of statistics of the lensing maps: their PDF, the convergence power spectrum, and lensing peak counts. We also study the impact that a number of variations in code setups (e.g. Born versus beyond-Born approximation, smoothing schemes, line-of-sight resolution, convergence versus shear power spectra) can have on the results.

### 4.1 Convergence maps

The convergence maps produced by the different lensing simulation codes are compared in Fig. 2. This comparison serves as a basic sanity check that the codes were successfully run on the same light-cone geometry and cosmic large-scale structure. This is confirmed by the good agreement between the position of high-$\kappa$ and low-$\kappa$ regions, as well as the small differences to the mean map of the codes.

The convergence one-point distributions of the maps are shown in Fig. 3. In the absence of any smoothing performed on the maps (left-hand panels), the PDF of the convergence from RAY-RAMSES has a noticeably higher peak at $\kappa \approx -0.01$, and it decays more sharply towards more negative $\kappa$ values. The RAY-RAMSES PDF also has higher probability of large $\kappa > 0.06$ values than the other codes. This indicates a stronger smoothing of the matter by RAY-RAMSES in low-density regions and a higher resolution in high-density regions compared to the other codes. This is as one would expect from the underlying adaptive 3D mesh employed in RAY-RAMSES compared to the non-adaptive 2D grids of the other codes.

After smoothing the maps with a Gaussian kernel with size 1 arcmin (centre panels), the shapes of the PDFs of all five codes are brought closer together, but slight horizontal shifts in the corresponding curves remain. This indicates that the codes produce slightly different mean values of the convergence across the field of view $\langle\kappa\rangle_\mathrm{fov}$. When the mean is taken out (right-hand panels), the PDFs agree to $\approx 1$ per cent in the range of $\kappa$ values where the PDFs are sizeable. We do not investigate further the origin of remaining differences on the tail of the distribution given the many different details in the operation of the codes. For example, the differences in the exact way the projection on to discrete planes/spheres is done in the codes (projection along line-of-sight versus along $z_\mathrm{coord}$, different density assignment schemes, etc.) can cause small differences that would appear exacerbated in relative differences of a (small) PDF. We note also that the slight difference in the values of $\langle\kappa\rangle_\mathrm{fov}$ is not worrying as it has little impact on lensing observables. For example, in the remainder of this section, we compare statistics measured from maps without the mean subtracted and we will find very good agreement between the codes. Furthermore, analyses based on shear statistics are more closely related to the actually observed galaxy ellipticities and are not sensitive to shifts in the mean signal across the field of view.

### 4.2 Convergence power spectra

The power spectra of the convergence maps of the HILBERT, LENS$^2$HAT, MAPSIM, MICE, and RAY-RAMSES codes are shown in Fig. 4. The power spectra are calculated with Fourier transforms assuming the flat-sky approximation, which is valid for the small field-of-view our comparison is based on.[12] All spectra were

[12]In cases where the curvature of the sky cannot be ignored, the spectra need to be computed using spherical harmonic decompositions instead.

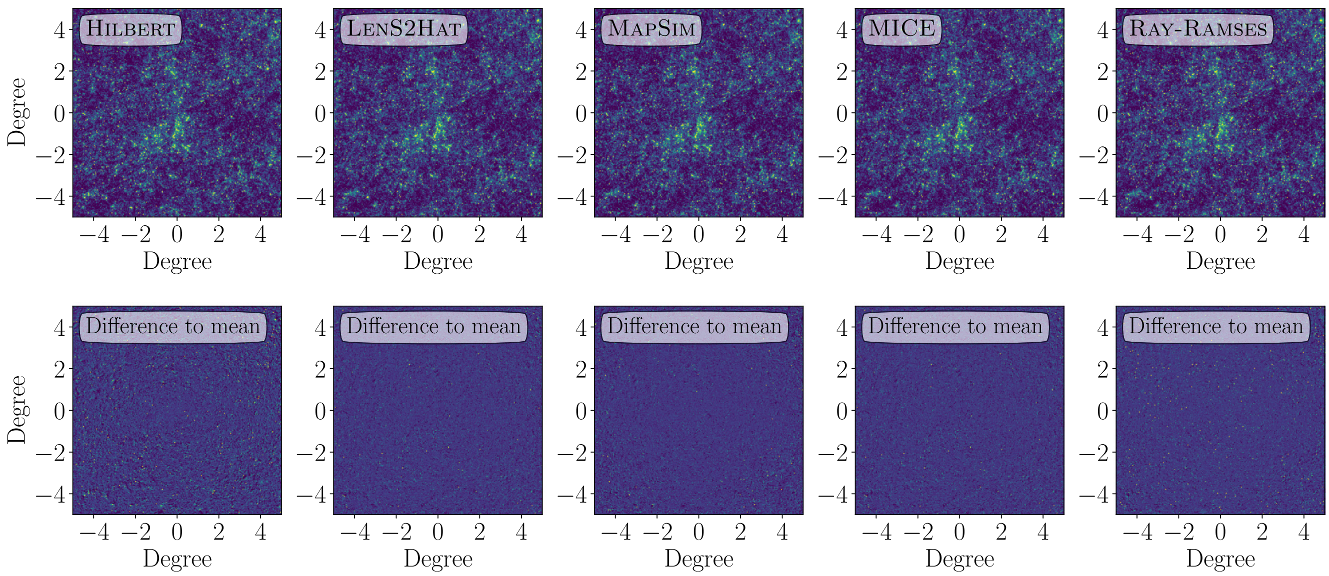


**Figure 2.** Lensing convergence maps produced by the lensing simulation codes, as labelled ($10 \times 10\,\mathrm{deg}^2$ with $2048 \times 2048$ resolution). The upper panels show the convergence maps as obtained by the codes. The lower panels show the corresponding difference to the mean of the codes. The colour scale is the same in all panels, ranging from $\kappa = -0.02$ (dark blue) to $\kappa = 0.1$ (bright yellow).

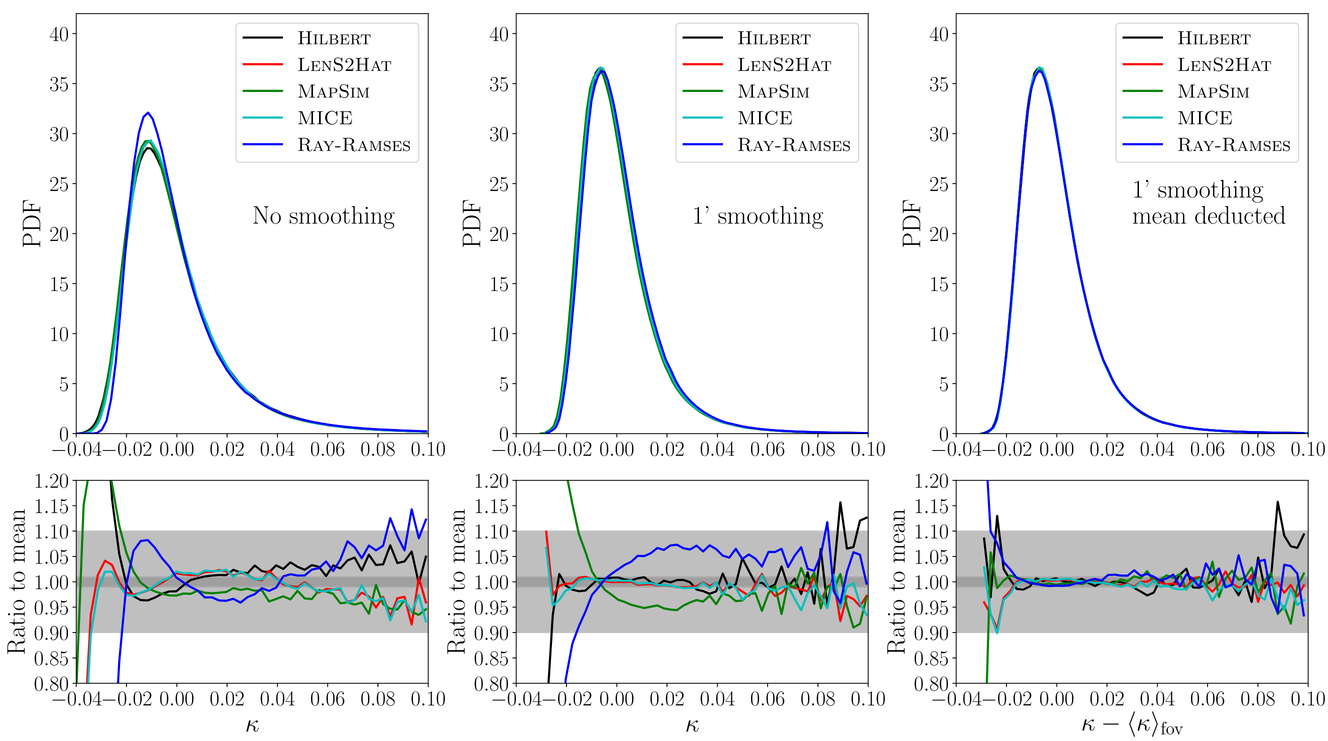


**Figure 3.** Probability density functions (PDF) of the convergence maps. The upper panels show the PDFs obtained by the lensing codes without smoothing (left-hand panel), after smoothing using a Gaussian kernel with width 1 arcmin (centre), and after smoothing and deducting the field mean value (right-hand panel). The lower panels show the corresponding difference to the PDF given by the mean of the codes. The light and dark grey shaded areas indicate 10 per cent and 1 per cent fractional errors. In all of the upper panels, the read and cyan lines are practically indistinguishable; in the right upper panel, all curves are indistinguishable.

evaluated using the routines in the publicly available LENSTOOLS software (Petri 2016).[13] The resulting spectra are then subsequently averaged in logarithmically spaced $\ell$-bins. Fig. 4 also shows the result obtained by integrating the 3D matter power spectrum measured directly from the simulation snapshots according to equation (16), as well as integrating the power spectrum given by the HALOFIT fitting formula (Smith et al. 2003; Takahashi et al. 2012).

[13] We checked that our results are identical using an independent power spectrum calculation code.

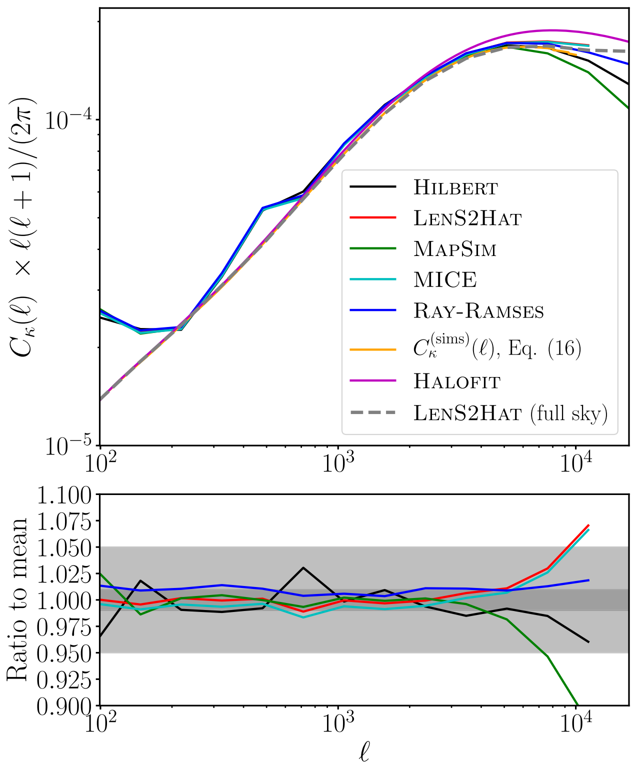


**Figure 4.** Power spectrum of the lensing convergence maps obtained by the lensing simulation codes, as labelled (upper panel). The result obtained by integrating the 3D non-linear matter power spectrum measured in the simulations according to equation (16) is also shown for reference (orange line). The magenta curve shows the convergence power spectrum obtained by integrating the matter power spectrum given by the HALOFIT formula. In the upper panel, the red and cyan lines are nearly indistinguishable. The lower panel shows the ratio to the mean of the codes. The light and dark grey shaded areas indicate 5 per cent and 1 per cent fractional errors.

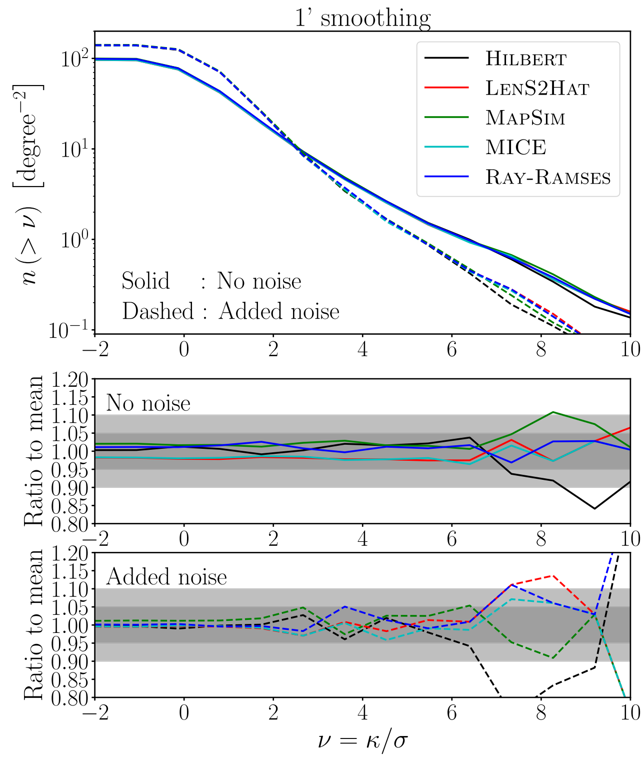


**Figure 5.** Cumulative count of lensing convergence peaks. The result is shown for the maps obtained by the various codes, with (dashed) and without (solid) shape noise added to the maps, as labelled. The result shown is for maps smoothed with a Gaussian kernel with the size of 1 arcmin. The lower panels show the ratio to the mean result of the codes. The light and dark grey shaded areas indicate 10 per cent and 5 per cent fractional errors.

The lower panel of Fig. 4 shows the ratio of the individual code results to their mean. The LENS$^2$HAT, MAPSIM, MICE, and RAY-RAMSES codes agree with the mean of the codes to $\lesssim 2$ per cent on scales $\ell \lesssim 4000$. The same holds for the HILBERT code, when ignoring the fluctuations of $\approx 3$ per cent at $\ell \approx 100$ and $\ell \approx 700$. These larger differences are likely due to the parallel projection (i.e. along the $z_{\mathrm{coord.}}$) used by HILBERT in contrast to the radial projection employed by the other codes, which causes slight differences in what structures appear where in the field of view. Note also that on scales $\ell \gtrsim 4000$, where the code differences become larger (10 per cent for $\ell = 10^4$ in some cases), uncertainties associated with the modelling of baryonic processes (most notably stellar and AGN feedback) are expected to be sources of larger systematic errors (Semboloni et al. 2011; Barreira et al. 2019; Gouin et al. 2019; Huang et al. 2019). Overall, we therefore conclude that the lensing codes tested display a satisfactory level of agreement, which is a valuable cross-check in preparation for the analysis of future surveys.

On scales $\ell \gtrsim 3000$, the codes systematically underpredict the amplitude of the power spectrum predicted by the HALOFIT result. This reflects the lower resolution of the *N*-body simulation used to construct the convergence maps of the tested codes. This is confirmed by the fact that the result of equation (16), which uses the power spectrum directly measured from the snapshots, displays the same level of suppression on small scales relative to the HALOFIT curve.

The code results also differ from equation (16) on scales $\ell \lesssim 1000$. On these angular scales, the results are severely affected by sample variance as they correspond to a single realization of a rather small field of view; it is in fact a reassuring cross-check that the codes all agree on this peculiar shape of the spectra. Increasing the size of the field of view and/or the number of realizations of our $10 \times 10\,\mathrm{deg}^2$ field of view (and then taking the mean) would reduce the impact of sample variance. As a sanity check, the grey dashed line shows the spectra measured from the full-sky map produced by the LENS$^2$HAT code, i.e. without specifying to the common $10 \times 10\,\mathrm{deg}^2$ field of view. This effectively increases the number of modes sampled (though not all independent because of box replication), which brings the result closer to the large-scale prediction of equation (16).

As an additional test, we also investigate the power spectrum of the log-transformed convergence (Neyrinck, Szapudi & Szalay 2009; Seo et al. 2011, 2012; McCullagh et al. 2013; Llinares & McCullagh 2017), which is sensitive to higher order statistics of the convergence field itself. We find that the power spectra of $\log_\kappa(0.1 + \kappa)$ (the value of 0.1 ensures the argument is always positive for our maps) measured from the maps of the codes agree with their mean to better than $\approx 2$ per cent on scales $\ell \lesssim 3000$.

### 4.3 Peak counts

Lensing peak counts contain information beyond the power spectrum, and their inclusion in data analyses has been advocated to be able to yield significantly improved cosmological parameter constraints (Dietrich & Hartlap 2010; Hilbert et al. 2012; Marian et al. 2012; Liu et al. 2015, 2016; Shan et al. 2018; Davies, Cautun & Li 2019). Fig. 5 shows the cumulative number density of lensing convergence peaks found in the maps of the different codes. Peaks

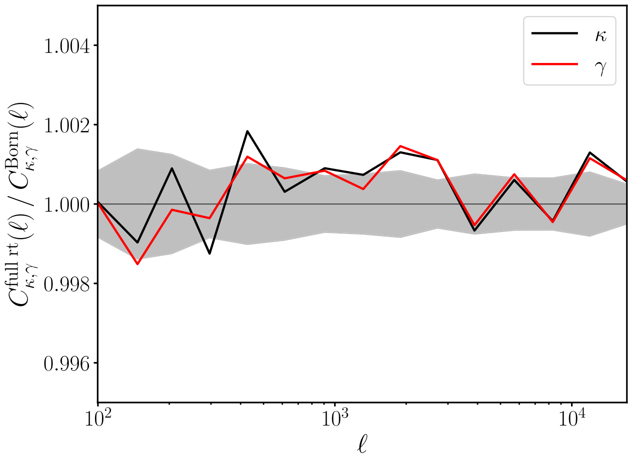


**Figure 6.** Ratio of the convergence and shear power spectra from the HILBERT code run with full ray-tracing to that assuming the Born approximation. The lines show ratios of the mean over 16 $10 \times 10\,\mathrm{deg}^2$ fields. The grey region indicates the statistical error on these ratios (in terms of standard deviation) estimated from the field-to-field variation.

are identified as pixels whose amplitude is higher than that of all its eight neighbours and the result is plotted in terms of the peak height (or signal-to-noise) $\nu = \kappa/\sigma$, where $\sigma$ is the standard deviation of all pixels. We use the $\kappa$ and $\sigma$ values of each code to define their peak height. We also count peaks on maps with their mean convergence subtracted (cf. Section 4.1). The result shown is for peaks counted on maps smoothed with a Gaussian filter of size 1 arcmin, and with and without Gaussian shape noise added (assuming a variance $\sigma_e^2 = 0.31^2$ for the shape noise in the shear estimate per galaxy and an effective number density of source galaxies $\bar{n}_{\rm eff} = 30\,\mathrm{arcmin}^{-2}$; the values of both $\kappa$ and $\sigma$ in $\nu = \kappa/\sigma$ are computed for maps with and without added noise). The assumption of Gaussian distributed shape noise is not critical to the code comparison.

The codes all display very similar lensing peak count predictions. In particular, for both the cases with and without shape noise, the codes agree with the mean prediction to better than 5 per cent for peaks with $\nu \lesssim 6$. This level of agreement is naturally related to that observed in Section 4.1 for the convergence maps PDFs. For higher $\nu$, the relative differences increase as the cumulative peak count becomes more sensitive to small changes. Should current or future real data analyses require better than $\sim 20$ per cent precision for $\nu \sim 10$, then these code differences should be understood better.

For completeness, we have also compared the codes for two additional cases (not shown): (i) peak counts on maps without the mean convergence deducted; (ii) using the values of $\sigma$ found in the map of one of the codes to define $\nu$ for all codes. We found that using a common $\sigma$ value yields effectively the same level of agreement and skipping subtracting the mean convergence yields an agreement of 10 per cent for $\nu \lesssim 6$.

## 4.4 Systematic errors

In this section, we quantify various possible sources of systematic errors in lensing simulations. Unless specified otherwise, the results described here are produced by the HILBERT code, and correspond to the average over 16 realizations of the $10 \times 10\,\mathrm{deg}^2$ field-of-view, obtained by picking 16 different observer orientations in the simulation box. This reduces the statistical noise due to the finiteness of the field of view (by up to a factor of four, assuming the realizations are statistically independent).

### *4.4.1 Born approximation*

As mentioned in Section 2.1, the impact of adopting the Born approximation (i.e. computing the lensing signal along unperturbed ray trajectories) on the convergence or shear power spectrum is expected to lie comfortably below the 1 per cent level for angular scales relevant for current and future surveys. This can be explicitly checked by comparing full beyond-Born ray-tracing with Born-approximation results. Conversely, assuming that the Born approximation is valid to better than 1 per cent on a given range of angular scales, such a comparison can also be used as a self-consistency check of a lensing simulation code that can perform both types of calculations.

Fig. 6 shows the ratio of the convergence and shear power spectra obtained without adopting the Born approximation to those obtained adopting it. The relative differences are $\sim 0.1$ per cent out to $\ell = 10^4$ and are likely due to numerical noise[14] rather than a direct consequence of the Born approximation. This corroborates past similar conclusions in the literature (e.g. Jain et al. 2000; Hilbert et al. 2009; Giocoli et al. 2016; Fabbian, Lewis & Beck 2019), but recall, as noted in Section 2.1, the degree of validity of the Born approximation can be different for other lensing statistics.

### *4.4.2 Line-of-sight discretization*

The users of multiple-plane and multiple-sphere algorithms must decide how many planes or spheres the algorithm employs to represent the matter distribution along the line-of-sight. Employing more planes/spheres naturally incurs on higher computational, data storage, and data transfer costs. Fewer planes/spheres may lead to larger line-of-sight discretization errors.

Fig. 7 illustrates for the HILBERT code, how the number of planes affects the resulting convergence power spectra. Reducing the number of planes from 90 to 30 or even to 10 may cause deviations in the measured power by up to 5 per cent. This is in rough accordance with earlier results by Vale & White (2003). However, the differences we find are within the expected statistical error for such a finite field, and bear no strong indication for a systematic difference. Further, recall that the HILBERT code projects the matter on to planes along the (0,0,1) direction, which exacerbates the impact of reducing the number of planes, compared to the other codes which do radial projections.

We can also use equation (19) to estimate the effect of a finite number of planes. Results are shown in Fig. 8 for different numbers $n$ of planes evenly spaced in comoving distance between the observer and the source. Employing $n = 90$ planes yields results that are practically indistinguishable from the limit $n \to \infty$. Even when using only $n = 10$ planes, which corresponds to a plane-to-plane distance of $\approx 200$ Mpc, the resulting convergence power spectrum deviates by less than 1 per cent.

We note however, that this small impact of the number of planes is partly due to our specific choice of source redshift and cosmology. In this case, the integrand in equation (10) is very symmetric for most $\ell$-values of interest, and the integral can be well approximated

[14]The main source of noise is that, for our finite fields of view, the light-cone in full-ray tracing contains slightly different matter structures than the one in Born mode due to the presence/absense of light deflections.

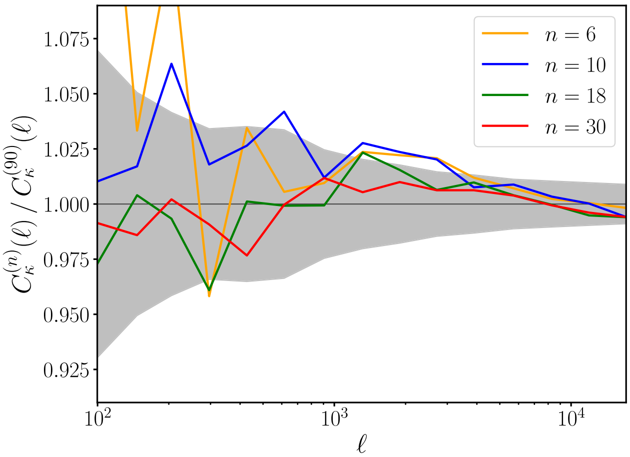


**Figure 7.** Ratio of the convergence power spectra $C_\kappa^{(n)}(\ell)$ for different number $n$ of planes to that using $n = 90$ planes for the HILBERT code. The results shown correspond to the mean of the 16 $10 \times 10\ \mathrm{deg}^2$ fields. The grey region indicates the relative statistical error (in terms of standard deviation) on the mean of the convergence power spectrum measured from these fields for the $n = 90$ case.

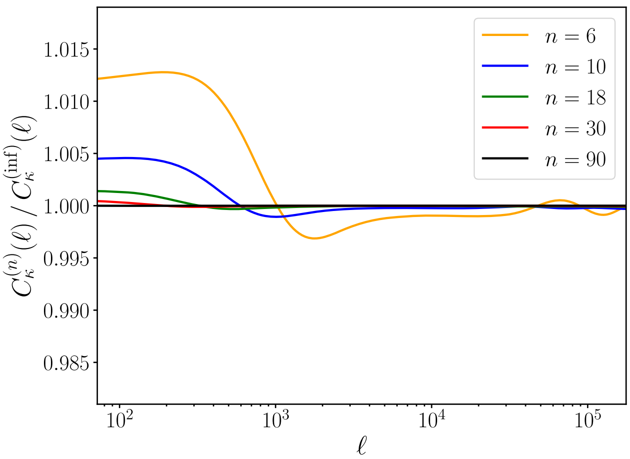


**Figure 8.** Ratio of the convergence power spectra $C_\kappa^{(n)}(\ell)$ for a finite number $n$ of planes to that in the limit $n \to \infty$, according to equation (19).

by a sum of the integrand values at just a few evenly spaced points between the observer and the source. For very different source redshifts (including realistic extended distributions) or cosmologies, the integrand may be more skewed, and the resolution along the line-of-sight may become more important.

### 4.4.3 Particle noise and smoothing

The 3D matter distribution in the underlying $N$-body simulation lacks power on very large and very small scales due to the finite box size and finite spatial/mass resolution. As mentioned in Section 4.2, this lack of power naturally propagates to the convergence and shear power spectra of the lensing simulations. Moreover, the lensing simulations are affected by additional smoothing, either inherent in some of their processing steps, e.g. when employing meshes of finite resolution, or applied explicitly, e.g. to reduce the impact of shot noise due to finite number of particles in the $N$-body simulation.

The effects of particle shot noise and smoothing are illustrated in Fig. 9 for the HILBERT code. Integrating the 3D matter power

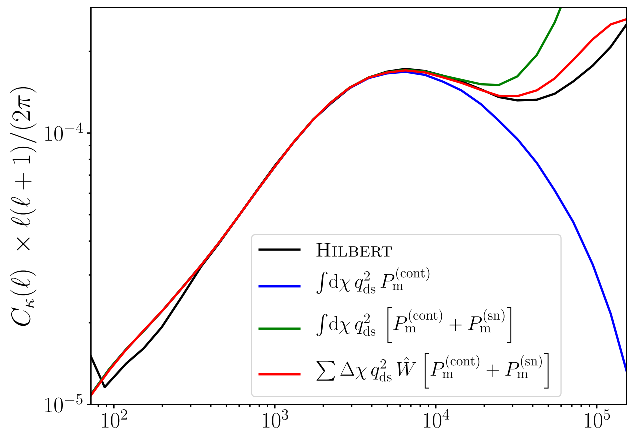


**Figure 9.** Comparison of convergence power spectra from HILBERT (black) with the prediction from equation (16) only considering the continuous component $P_{\mathrm{m}}^{(\mathrm{cont})}$ in equation (15) (blue), the prediction from equation (16) also accounting for shot noise (green), as well as the prediction from equation (19) taking into account line-of-sight discretization, shot noise, and smoothing (red).

spectrum directly without accounting for particle shot noise or smoothing yields much smaller values for the convergence power spectrum than the lensing simulation for $\ell \gtrsim 10^4$. Further, accounting for shot noise, but not for smoothing overestimates the convergence power spectra of the lensing simulation.

When the smoothing is taken into account (in addition to the line-of-sight discretization) assuming the kernel of equation (20), the prediction based on equation (19) agrees with the power spectrum directly measured from the convergence maps to better than 1 per cent for scales $300 \lesssim \ell \lesssim 20\,000$. This underlines the fact that a good understanding of the amount of shot noise and smoothing carried out by a given lensing code can prove useful in tests of its accuracy.

### 4.4.4 Shear power spectra

We also compare code predictions for the shear power spectrum (not just convergence), since the shear is more directly related to the actually observed galaxy image ellipticities. As mentioned in Section 2.1, $C_\gamma^{(\mathrm{EE})}(\ell) = C_\kappa(\ell)$ for a flat sky;[15] this relation holds to high accuracy beyond the Born approximation as well (cf. Fig. 6). In Fig. 10 we show the ratios of the shear and convergence power spectra for the HILBERT, MAPSIM, and RAY-RAMSES codes.[16] The deviations of the measured ratios from unity seen for $\ell \leq 1000$ can be attributed to the finite size of the field of view.[17] On scales $\ell > 1000$, the RAY-RAMSES result underestimates the shear power spectrum in a non-negligible way: the shear power is smaller by

[15]On a spherical sky, they differ by a known wavenumber-dependent factor that is close to unity for sufficiently large $\ell$.

[16]The HILBERT and RAY-RAMSES codes compute the lensing shear directly by projecting the corresponding second-derivatives of the lensing potential. The shear power spectrum from MAPSIM is calculated from a $\gamma$ map obtained from the $\kappa$ map by Fourier transforming $\tilde{\gamma}(\boldsymbol{\ell}) = (\tilde{\gamma}_1(\boldsymbol{\ell}), \tilde{\gamma}_2(\boldsymbol{\ell})) = \left(\frac{\ell_1^2-\ell_2^2}{\ell_1^2+\ell_2^2}, \frac{2\ell_1\ell_2}{\ell_1^2+\ell_2^2}\right)\tilde{\kappa}(\boldsymbol{\ell})$, where $\boldsymbol{\ell} = (\ell_1, \ell_2)$.

[17]For example, different matter structures outside the finite field of view may contribute to the shear seen within the field, but not to the convergence. This can explain why the shear and convergence spectra do not react in the same way when the orientation of the finite field-of-view changes.

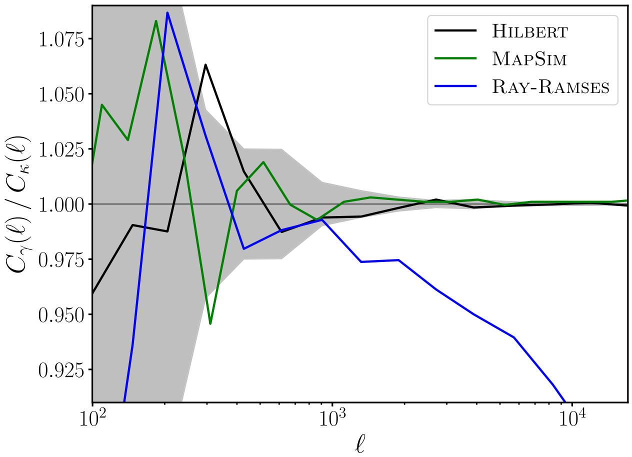


**Figure 10.** Ratio of shear and convergence power spectra for the HILBERT, MAPSIM, and RAY-RAMSES codes (computed from one $10 \times 10\,\mathrm{deg}^2$ field). The shaded area indicates the expected variance (estimated from the 16 HILBERT fields). For $\ell \leq 1000$, the deviations from unity can be attributed to the finite field size. The deviations for $\ell > 1000$ for RAY-RAMSES are due to different effective smoothings for shear and convergence.

approximately 6 per cent for $\ell \approx 4000$. This difference between the shear and convergence power spectrum can be traced back to an effective larger smoothing that exists in the calculation of the second derivatives of the potential that are integrated to calculate $\gamma_1$ and $\gamma_2$ (cf. Section 3.2.5). This suppression is therefore expected to remain even if the simulation is performed at higher mass resolution, although the scale at which $C_\gamma^{(\mathrm{EE})}(\ell) \neq C_\kappa(\ell)$ would be pushed to higher $\ell$ values.

This subtlety in the operation of the RAY-RAMSES code stresses further the importance to understand the impact of smoothing in lensing simulation codes. The smoothing in RAY-RAMSES is adaptive, controlled mostly by the resolution of the *N*-body simulation, and as shown here can yield different shear and convergence spectra (even though they should be the same). On the other hand, the smoothing schemes can be partly specified by the user in the other codes (e.g. in the smoothing or mass assignment scheme used to construct the multiple planes/spheres), and should therefore be subject to careful numerical convergence tests. For the LENS$^2$HAT code, the relation $C_\gamma^{(\mathrm{EE})}(\ell) = C_\kappa(\ell)$ has been verified to hold to sub-per cent precision, both in the Born approximation and beyond it (Fabbian et al. 2018).

## 5 SUMMARY AND CONCLUSIONS

In this paper, we have studied the relative accuracy of different weak lensing simulation codes by comparing the statistics of lensing convergence maps that each produced from a common underlying simulation of cosmic large-scale structure. The five codes HILBERT, LENS$^2$HAT, MAPSIM, MICE and RAY-RAMSES we compare (cf. Table 1 and Section 3.2) were developed independently from each other, and a significant number of results in the literature featuring these codes exists already. The comparison analysis we carried out in this paper thus serves the purpose of cross-checking the validity of these past results, as well as checking the extent to which any systematic difference between codes can affect the analysis of large weak lensing surveys.

A major difference between the codes is how they integrate the rays along the line-of-sight: HILBERT and MAPSIM project the deflector mass field on to planes perpendicular to the central line-of-sight of the field-of-view, MICE and LENS$^2$HAT project it on to concentric spherical shells around the observer, and RAY-RAMSES carries out the lensing integrations using the 3D distribution of the mass without projecting it. Other specifications such as interpolation schemes to reconstruct the deflector mass on regular grids, and additional smoothing applied on these grids, also differ between the codes.

The lensing simulation codes carried out their calculations on the output of the same *N*-body simulation performed with the RAMSES code. Specifically, the RAY-RAMSES code ran on-the-fly with the *N*-body calculation and produced the particle snapshots that the remaining codes took as input. We considered a light-cone geometry with area $10 \times 10\,\mathrm{deg}^2$, extending out to a single source redshift $z_\mathrm{s} = 1$ (cf. Fig. 1). This small field of view is not representative of the total area of large surveys like *Euclid* ($\approx 15\,000\,\mathrm{deg}^2$), but is sufficient to compare the code results on small angular scales, which are the scales for which we rely on numerical simulations to resolve non-linear structure formation.

The main results of the code comparison are:

(i) The PDFs of the maps agree to $\approx$ 1 per cent on $\kappa$ values where the PDF is sizeable, but only after applying 1 arcmin smoothing and subtracting the mean convergence over the field of view (cf. Fig. 3).

(ii) The convergence power spectra predicted by the codes agree to 2 per cent for $\ell \lesssim 4000$ (cf. Fig. 4). At $\ell = 10^4$, the differences can be as large as 10 per cent, mainly due to differences in the smoothing of the matter field.

(iii) The code predictions for lensing peak counts agree to better than 5 per cent for peaks with signal-to-noise $\nu \lesssim 6$, both for maps with and without shape noise added (cf. Fig. 5).

Corroborating previous results in the literature, we confirmed the validity of the Born approximation in the convergence power spectrum from lensing simulations. Following the rays along unperturbed trajectories has an impact that is smaller than 0.2 per cent for $\ell < 10^4$ (see Fig. 6). Note, however, that the Born approximation can have a stronger impact on other observables such as galaxy–galaxy lensing, higher order statistics or CMB lensing cross-correlations (Hilbert et al. 2009; Fabbian et al. 2019).

We also showed that reducing the number of lens planes from 90 (plane thickness $\sim$20 Mpc h$^{-1}$) to 10 (plane thickness $\sim$200 Mpc h$^{-1}$) can impact the resulting convergence power spectrum of the HILBERT code at the 5 per cent level (see Fig. 7), albeit with no clear systematic trend. This value is likely exacerbated by the parallel projection this code adopts. Analytic predictions based on matter power spectra suggest that the systematic error due to the line-of-sight discretization is smaller. Moreover, the good agreement between using 90 lens planes or the full *N*-body resolution used by RAY-RAMSES is telling that the discretization along the line-of-sight is not a critical source of systematic error (at least for slices with width $\lesssim$ 25 Mpc h$^{-1}$). Further, we noted that the shear power spectrum predicted by the RAY-RAMSES code is suppressed relative to that of the convergence on small scales ($\approx$ 6 per cent on $\ell \approx 4000$; see Fig. 10). This is due to an effectively larger smoothing that goes into the calculation of the integrand of $\gamma$ in this code, compared to $\kappa$.

We also compared the convergence power spectrum measured directly from the lensing simulation maps with predictions obtained by analytically integrating the non-linear 3D matter power spectrum measured from the *N*-body simulation snapshots. Smoothing effects and particle shot noise can be appropriately taken into account analytically. When doing so, e.g. for the HILBERT code, the analytical prediction agrees with the lensing simulation spectra to better than 1 per cent out to $\ell = 20\,000$ (cf. Fig. 9). These are scales already well below the scales that our *N*-body simulation could accurately resolve, but an appropriate analytical calculation can capture that loss of resolution and thus be used in self-consistency tests of the codes.

Overall, the comparison of the HILBERT, LENS$^2$HAT, MAPSIM, MICE, and RAY-RAMSES codes did not reveal any significant systematic errors in the statistical quantities we analysed. Further, in comparisons in which larger than a few per cent differences were observed, e.g. power spectrum on $\ell \gtrsim 4000$, there are other known sources of larger uncertainty such as the modelling of baryonic processes, or even the accuracy of *N*-body methods in gravity-only simulations. We thus conclude that the current accuracy of weak-lensing simulation codes is acceptable for applications to current and near future data analyses.

In accordance with previous works (e.g. Fosalba et al. 2008), we find that the convergence and shear power spectra measured from the lensing simulations can be accurately predicted analytically on the relevant scales, which suggests a method to validate lensing codes and simulations. As outlined in Section 2.2 and exemplified in Section 4.4.3, one adjusts the first-order predictions for the convergence and shear power spectra to account for the peculiarities of the input matter distribution, any line-of-sight discretization, particle noise, and smoothing according to the numerical parameters of cosmic structure simulation and the lensing simulation. One then measures the power spectra from the output of the lensing simulation and compares them to the adjusted first-order predictions. Any significant deviations (e.g. larger than expected from sample variance due to finite area and depth) may then indicate a problem with the lensing simulation code or the quantitative understanding of its numerical properties.

Throughout this paper, we refrained from drawing considerations on the numerical performance of the codes as it is hard to find objective points of comparison. The main distinction in terms of numerical resources concerns the post-processing or on-the-fly nature of the codes. The post-processing codes require relatively low numerical resources in the calculation of the lensing quantities *per se*, but the *N*-body data that they analyse (and which is more expensive to generate) is assumed to be pre-existent. Lensing calculations performed on the fly with the *N*-body calculation have the advantage of requiring in principle fewer data storage resources, but are less flexible to changes in the lensing setup adopted (e.g. changes in source redshift may require a new *N*-body simulation). The choices of which method/code will also in general be determined by the specific application in mind, based on the specific features of each code (summarized in Table 1). A main conclusion of this work is that, regardless of which method/code is chosen, the accuracy of the result should be within the level of agreement shown in this paper.

As future improvements to the analysis we carried out in this paper, one may extend the comparison to full-sky lensing simulations. Of the codes tested here, only LENS$^2$HAT and MICE can currently carry these out. There is in principle no impediment to make HILBERT, MAPSIM, and RAY-RAMSES capable of that too, but that would involve further code development. Additionally, it would be valuable to carry out a similar comparison of numerical codes such as PINOCCHIO (Monaco et al. 2002; Munari et al. 2017), PEAKPATCH (Stein, Alvarez & Bond 2019), ICE-COLA (Izard, Fosalba & Crocce 2018), or that of Giocoli et al. (2017), which are capable of generating fast (yet approximate) realizations of the deflector mass distribution and compute their lensing properties.


## ACKNOWLEDGEMENTS

We thank Peter Schneider for useful comments. SH acknowledges support by the DFG cluster of excellence 'Origin and Structure of the Universe' (`www.universe-cluster.de`). AB acknowledges support from the starting grant (ERC-2015-STG 678652) 'GrInflaGal' of the European Research Council. GF acknowledges support from the CNES postdoctoral fellowship and support by the UK STFC grant ST/P000525/1 as well as by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) ERC Grant Agreement No. [616170]. PF acknowledges support from MINECO through grant ESP2017-89838-C3-1-R, the European Union H2020-COMPET-2017 grant Enabling Weak Lensing Cosmology, and Generalitat de Catalunya through grant 2017-SGR-885. CG acknowledges the grants ASI n.I/023/12/0, ASI-INAF n. 2018-23-HH.0, and PRIN MIUR 2015 Cosmology and Fundamental Physics: illuminating the Dark Universe with Euclid'. MC carried out part of this work while supported by a grant of the EU-ESF, the Autonomous Region of the Aosta Valley, and the Italian Ministry of Labour and Social Policy (CUP B36G15002310006), and by a 2019 'Research and Education' grant from Fondazione CRT. The OAVdA is managed by the Fondazione Clément Fillietroz-ONLUS, which is supported by the Regional Government of the Aosta Valley, the Town Municipality of Nus, and the 'Unité des Communes valdôtaines Mont-Émilius'. CTD is funded by a Science and Technology Facilities Council (STFC) PhD studentship through grant ST/R504725/1. BL is supported by an ERC Starting Grant, ERC-StG-PUNCA-716532, and is additionally supported by the STFC Consolidated Grants (ST/P000541/1).

The simulations and some of the analyses in this work used the DiRAC facility, managed by the Institute for Computational Cosmology, Durham University on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1, and ST/S002502/1, Durham University, and STFC operation grant ST/R000832/1. DiRAC is part of the UK National e-Infrastructure.


## REFERENCES


Abbott T. M. C. et al., 2018, Phys. Rev. D, 98, 043526
Angulo R. E., Hilbert S., 2015, MNRAS, 448, 364
Barreira A., Llinares C., Bose S., Li B., 2016, J. Cosmol. Astropart. Phys., 5, 001
Barreira A., Nelson D., Pillepich A., Springel V., Schmidt F., Pakmor R., Hernquist L., Vogelsberger M., 2019, MNRAS, 488, 2079
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Bartelmann M., Weiss A., 1994, A&A, 287, 1
Becker M. R., 2013, MNRAS, 435, 115
Bezecourt J., Pello R., Soucail G., 1998, A&A, 330, 399
Blandford R. D., Kochanek C. S., 1987, ApJ, 321, 658
Blot L. et al., 2019, MNRAS, 485, 2806
Breton M.-A., Rasera Y., Taruya A., Lacombe O., Saga S., 2019, MNRAS, 483, 2671
Calabrese M., Carbone C., Fabbian G., Baldi M., Baccigalupi C., 2015, J. Cosmol. Astropart. Phys., 3, 049
Carbone C., Springel V., Baccigalupi C., Bartelmann M., Matarrese S., 2008, MNRAS, 388, 1618
Carbone C., Baccigalupi C., Bartelmann M., Matarrese S., Springel V., 2009, MNRAS, 396, 668
Castro P. G., Heavens A. F., Kitching T. D., 2005, Phys. Rev. D, 72, 023516
Castro T., Quartin M., Giocoli C., Borgani S., Dolag K., 2018, MNRAS, 478, 1305
Chuang C.-H. et al., 2015, MNRAS, 452, 686

Colavincenzo M. et al., 2019, MNRAS, 482, 4883
Colberg J. M. et al., 2008, MNRAS, 387, 933
Cooray A., Hu W., 2002, ApJ, 574, 19
Couchman H. M. P., Barber A. J., Thomas P. A., 1999, MNRAS, 308, 180
Das S., Bode P., 2008, ApJ, 682, 1
Davies C. T., Cautun M., Li B., 2019, MNRAS, 488, 5833
Despali G., Vegetti S., White S. D. M., Giocoli C., van den Bosch F. C., 2018, MNRAS, 475, 5424
Dietrich J. P., Hartlap J., 2010, MNRAS, 402, 1049
Elahi P. J. et al., 2013, MNRAS, 433, 1537
Fabbian G., Stompor R., 2013, A&A, 556, A109
Fabbian G., Calabrese M., Carbone C., 2018, J. Cosmol. Astropart. Phys., 2, 050
Fabbian G., Lewis A., Beck D., 2019, J. Cosmol. Astropart. Phys., 2019, 57
Fosalba P., Gaztañaga E., Castander F. J., Manera M., 2008, MNRAS, 391, 435
Fosalba P., Gaztañaga E., Castander F. J., Crocce M., 2015, MNRAS, 447, 1319
Ghosh B., Durrer R., Sellentin E., 2018, J. Cosmol. Astropart. Phys., 6, 008
Giocoli C., Meneghetti M., Bartelmann M., Moscardini L., Boldrin M., 2012, MNRAS, 421, 3343
Giocoli C., Metcalf R. B., Baldi M., Meneghetti M., Moscardini L., Petkova M., 2015, MNRAS, 452, 2757
Giocoli C. et al., 2016, MNRAS, 461, 209
Giocoli C. et al., 2017, MNRAS, 470, 3574
Giocoli C., Moscardini L., Baldi M., Meneghetti M., Metcalf R. B., 2018, MNRAS, 478, 5436
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Gouin C. et al., 2019, A&A, 626, A72
Grossman S. A., Narayan R., 1988, ApJ, 324, L37
Hamana T., Mellier Y., 2001, MNRAS, 327, 169
Hamana T. et al., 2019, preprint (arXiv:1906.06041)
Harnois-Déraps J., Vafaei S., Van Waerbeke L., 2012, MNRAS, 426, 1262
Hikage C. et al., 2019, PASJ, 71, 43
Hilbert S., White S. D. M., Hartlap J., Schneider P., 2007, MNRAS, 382, 121
Hilbert S., White S. D. M., Hartlap J., Schneider P., 2008, MNRAS, 386, 1845
Hilbert S., Hartlap J., White S. D. M., Schneider P., 2009, A&A, 499, 31
Hilbert S., Marian L., Smith R. E., Desjacques V., 2012, MNRAS, 426, 2870
Hildebrandt H. et al., 2017, MNRAS, 465, 1454
Hildebrandt H. et al., 2020, A&A, 633, A69
Hirata C. M., Seljak U., 2003, Phys. Rev. D, 68, 083002
Horesh A., Ofek E. O., Maoz D., Bartelmann M., Meneghetti M., Rix H.-W., 2005, ApJ, 633, 768
Horesh A., Maoz D., Hilbert S., Bartelmann M., 2011, MNRAS, 418, 54
Hu W., 2000, Phys. Rev. D, 62, 043007
Huang H.-J., Eifler T., Mandelbaum R., Dodelson S., 2019, MNRAS, 672
Izard A., Fosalba P., Crocce M., 2018, MNRAS, 473, 3051
Jain B., Seljak U., White S., 2000, ApJ, 530, 547
Kainulainen K., Marra V., 2011, Phys. Rev. D, 84, 063004
Kamionkowski M., Kosowsky A., Stebbins A., 1997, Phys. Rev. D, 55, 7368
Kiessling A., Heavens A. F., Taylor A. N., Joachimi B., 2011, MNRAS, 414, 2235
Kilbinger M., 2015, Rep. Prog. Phys., 78, 086901
Kilbinger M. et al., 2017, MNRAS, 472, 2126
Knebe A. et al., 2011, MNRAS, 415, 2293
Knebe A. et al., 2013, MNRAS, 428, 2039
Kochanek C. S., 2004, ApJ, 605, 58
Krause E., Hirata C. M., 2010, A&A, 523, A28
Krause E. et al., 2017, preprint (arXiv:1706.09359)
Laureijs R. et al., 2011, preprint (arXiv:e-print)
Lewis A., Pratten G., 2016, J. Cosmol. Astropart. Phys., 12, 003
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Li B., King L. J., Zhao G.-B., Zhao H., 2011, MNRAS, 415, 881
Lippich M. et al., 2019, MNRAS, 482, 1786
Liu X. et al., 2016, Phys. Rev. Lett., 117, 051101
Liu J., Petri A., Haiman Z., Hui L., Kratochvil J. M., May M., 2015, Phys. Rev. D, 91, 063507
Llinares C., McCullagh N., 2017, MNRAS, 472, L80
LSST Dark Energy Science Collaboration, 2012, preprint (arXiv:1211.0310)
Mandelbaum R., 2018, ARA&A, 56, 393
Marian L., Smith R. E., Hilbert S., Schneider P., 2012, MNRAS, 423, 1711
Marozzi G., Fanizza G., Di Dio E., Durrer R., 2016, J. Cosmol. Astropart. Phys., 9, 028
McCullagh N., Neyrinck M. C., Szapudi I., Szalay A. S., 2013, ApJ, 763, L14
Meneghetti M., Bolzonella M., Bartelmann M., Moscardini L., Tormen G., 2000, MNRAS, 314, 338
Meneghetti M. et al., 2017, MNRAS, 472, 3177
Metcalf R. B., Madau P., 2001, ApJ, 563, 9
Metcalf R. B., Petkova M., 2014, MNRAS, 445, 1942
Monaco P., Theuns T., Taffoni G., 2002, MNRAS, 331, 587
Muciaccia P. F., Natoli P., Vittorio N., 1997, ApJ, 488, L63
Munari E., Monaco P., Sefusatti E., Castorina E., Mohammad F. G., Anselmi S., Borgani S., 2017, MNRAS, 465, 4658
Neyrinck M. C., Szapudi I., Szalay A. S., 2009, ApJ, 698, L90
Oguri M., 2002, ApJ, 573, 51
Onions J. et al., 2012, MNRAS, 423, 1200
Paczynski B., 1986, ApJ, 301, 503
Petkova M., Metcalf R. B., Giocoli C., 2014, MNRAS, 445, 1954
Petri A., 2016, Astron. Comput., 17, 73
Petri A., Haiman Z., May M., 2016, Phys. Rev. D, 93, 063524
Petri A., Haiman Z., May M., 2017a, Phys. Rev. D, 95, 123503
Petri A., Haiman Z., May M., 2017b, Phys. Rev. D, 95, 123503
Pratten G., Lewis A., 2016, J. Cosmol. Astropart. Phys., 8, 047
Puchwein E., Hilbert S., 2009, MNRAS, 398, 1298
Puchwein E., Bartelmann M., Dolag K., Meneghetti M., 2005, A&A, 442, 405
Scannapieco C. et al., 2012, MNRAS, 423, 1726
Schneider A. et al., 2016, J. Cosmol. Astropart. Phys., 2016, 047
Scoccimarro R., 1998, MNRAS, 299, 1097
Semboloni E., Hoekstra H., Schaye J., van Daalen M. P., McCarthy I. G., 2011, MNRAS, 417, 2020
Seo H.-J., Sato M., Dodelson S., Jain B., Takada M., 2011, ApJ, 729, L11
Seo H.-J., Sato M., Takada M., Dodelson S., 2012, ApJ, 748, 57
Shan H. et al., 2018, MNRAS, 474, 1116
Shapiro C., Cooray A., 2006, J. Cosmol. Astropart. Phys., 3, 007
Simon P., Hilbert S., 2018, A&A, 613, A15
Smith R. E. et al., 2003, MNRAS, 341, 1311
Spergel D. et al., 2013, preprint (arXiv:1305.5422)
Srisawat C. et al., 2013, MNRAS, 436, 150
Stebbins A., 1996, preprint (astro-ph/9609149)
Stein G., Alvarez M. A., Bond J. R., 2019, MNRAS, 483, 2236
Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152
Troxel M. A. et al., 2018, Phys. Rev. D, 98, 043528
Vale C., White M., 2003, ApJ, 592, 699
Wambsganss J., Paczynski B., Katz N., 1990, ApJ, 352, 407
Wambsganss J., Cen R., Ostriker J. P., 1998, ApJ, 494, 29
White M., Hu W., 2000, ApJ, 537, 1
Winther H. A. et al., 2015, MNRAS, 454, 4208
Xavier H. S., Abdalla F. B., Joachimi B., 2016, MNRAS, 459, 3693
Xu D. D. et al., 2009, MNRAS, 398, 1235
Xu D., Sluse D., Gao L., Wang J., Frenk C., Mao S., Schneider P., Springel V., 2015, MNRAS, 447, 3189
Zaldarriaga M., Seljak U., 1997, Phys. Rev. D, 55, 1830
Ziour R., Hui L., 2008, Phys. Rev. D, 78, 123517

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