

Model-independent test for gravity using intensity mapping and galaxy clustering

Muntazir M. Abidi,¹ Camille Bonvin¹,¹ Mona Jalilvand,^{1,2,3} and Martin Kunz¹

¹*Université de Genève, Département de Physique Théorique and Centre for Astroparticle Physics, 24 quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland*

²*Department of Physics, McGill University, 3600 rue University, Montreal, Quebec H3A 2T8, Canada*

³*McGill Space Institute, McGill University, 3550 rue University, Montreal, Quebec H3A 2A7, Canada*



(Received 30 August 2022; accepted 16 January 2023; published 9 March 2023)

We propose a novel method to measure the E_G statistic from clustering alone. The E_G statistic provides an elegant way of testing the consistency of General Relativity by comparing the geometry of the Universe, probed through gravitational lensing, with the motion of galaxies in that geometry. Current E_G estimators combine galaxy clustering with gravitational lensing, measured either from cosmic shear or from CMB lensing. In this paper, we construct a novel estimator for E_G , using only clustering information obtained from two tracers of the large-scale structure: intensity mapping and galaxy clustering. In this estimator, both the velocity of galaxies and gravitational lensing are measured through their impact on clustering. We show that with this estimator, we can suppress the contaminations that affect other E_G estimators and consequently test the validity of General Relativity robustly. We forecast that with the coming generation of surveys like HIRAX and *Euclid*, we will measure E_G with a precision of up to 7% (3.9% for the more futuristic SKA2).

DOI: [10.1103/PhysRevD.107.063514](https://doi.org/10.1103/PhysRevD.107.063514)

I. INTRODUCTION

One of the main goals of the coming generation of large-scale structure surveys is to test the consistency of General Relativity (GR) at cosmological scales. Since the observation of the accelerated expansion of the Universe in 1998 [1,2], a large number of theories of gravity beyond General Relativity have been constructed. Testing these theories one by one by confronting them with observations is not anymore a feasible option. This complexity in the landscape of models beyond GR has led the community to build model-independent tests of gravity, i.e., tests that can be applied to data without relying on particular modeling and whose outcome will either confirm or rule out the validity of GR (e.g., [3–10]).

One particularly useful test is the so-called E_G statistic, first proposed by Zhang *et al.* [11]. The idea of this test is to combine galaxy-lensing correlations with galaxy-velocity correlations to test the relation between the sum of the metric potentials, $\Phi + \Psi$, and the galaxy peculiar velocity V ;

$$E_G = \frac{\text{galaxy} - \text{lensing}}{\text{galaxy} - \text{velocity}} \propto \frac{\langle \delta_g(\Phi + \Psi) \rangle}{\langle \delta_g V \rangle}. \quad (1)$$

In GR, these two quantities are related via Einstein's equations so that E_G takes a specific scale-independent value. In modified theories of gravity, however, the relation between Φ and Ψ is generically modified, as is the growth rate of structures f , which governs the evolution of peculiar

velocities [12]; E_G is consequently modified and becomes potentially scale dependent. Measurements of E_G provide, therefore, a direct test of the validity of GR. This test has the advantage of being independent of galaxy bias b , since the galaxy density appears both in the numerator and denominator of Eq. (1), and also of the initial conditions, which cancel out in the ratio [4].

Different methods have been used to measure E_G in practice. First, since peculiar velocities are not straightforward to measure, the galaxy-velocity correlation in the denominator has been replaced by the product of the galaxy-galaxy correlations and the parameter $\beta = f/b$, which can be measured directly from redshift-space distortions (RSD) [13]. Concerning the numerator, two observables have been used: 1) the correlations between galaxy clustering and cosmic shear, and 2) the correlations between galaxy clustering and CMB lensing.

The first measurement of E_G was carried out by Reyes *et al.* [14] using the first method applied to luminous red galaxies from the Sloan Digital Sky Survey (SDSS) [15]. They measured $E_G = 0.39 \pm 0.06$, confirming the Λ CDM predictions, on the scale of tens of Mpc. Later, Amon *et al.* [16] measured E_G , again with the first method, by combining deep imaging data from the Kilo-Degree Survey [17] with the overlapping spectroscopic 2-degree Field Lensing Survey [18], the Baryon Oscillation Spectroscopic Survey [19] and the Galaxy Mass Assembly Survey [20] and found some tension of their results with the GR predictions. Pullen *et al.* [21] used the second method, cross-correlating

the Planck CMB lensing map with the SDSS III CMASS galaxy samples and found an almost 2.6σ deviation from the GR prediction. Other works on the E_G measurements found similar tensions [22–24].

Measuring E_G more precisely with the coming generation of surveys will reveal whether these tensions persist, possibly indicating a breakdown of GR at cosmological scales. In this context, minimizing the impact of systematic effects on measurements of E_G is crucial. Recently, it has been shown that one important contamination, which is negligible for current surveys, will contaminate E_G at high redshift, invalidating its use to test the consistency of GR with the next generation of surveys [25]. This contamination is the contribution of lensing magnification to galaxy clustering. Since lensing magnification correlates strongly with cosmic shear (method 1) and with CMB lensing (method 2), it inevitably leads to an extra contribution in the numerator of Eq. (1). At $z = 1.5$ this contamination reaches 25–40%, leading to an E_G which is neither scale independent nor bias independent. In [26] a method has been proposed to remove this contamination by measuring additional correlations, namely shear-shear correlations and shear-CMB lensing correlations, and subtracting them from E_G .

In this paper, we propose an alternative way of measuring E_G *without contamination*, using only clustering information. We use two different tracers of the large-scale structure (LSS); galaxy clustering and 21 cm intensity mapping (IM). Intensity mapping is a novel technique to map the LSS by measuring the intensity fluctuations of some emission line (typically the 21 cm line emitted by neutral hydrogen) with radio surveys. These fluctuations, which follow the dark matter distribution, can be used as a new tracer of the LSS. Although the presence of large foregrounds make IM autocorrelation measurements very challenging, this problem is mitigated in cross-correlation measurements of IM with galaxy clustering, which have already been successfully performed [27–35].

In [36], we built a new observable, called GIMCO, which combines IM with galaxy clustering to obtain a direct measurement of the galaxy-lensing correlation. Here we propose to use GIMCO in the numerator of E_G . Instead of measuring the galaxy-lensing correlation from cosmic shear or CMB lensing, we measure it directly from clustering. As we will see, this method has the strong advantage of being unaffected by the lensing contamination described above, which affects both galaxy-shear correlations and galaxy-CMB lensing correlations. Moreover, since it relies only on clustering information, it is unaffected by potential inconsistencies between low redshift and high redshift data sets (as is the case for galaxy-CMB lensing correlations) or systematics affecting the measurement of cosmic shear. It provides, therefore, a robust way of testing the consistency of GR.

The rest of the paper is organized as follows: In Sec. II we introduce our new method to measure E_G using

intensity mapping. In Sec. III we study the contaminations to E_G , and we determine the optimal redshift binning to reduce these contaminations to a negligible level. In Sec. IV we forecast the precision with which E_G will be measured with the coming generation of galaxy and IM surveys and the constraints expected on modifications of gravity. We conclude in Sec. V.

II. E_G FROM INTENSITY MAPPING

We build E_G using two different biased tracers of the matter clustering: the fluctuations in galaxy number counts, Δ_g , and the fluctuations in the 21 cm brightness temperature, Δ_{HI} . The fully relativistic expressions at linear order for these two quantities have been derived in [37–40]. Here we consider only the terms that dominate in the angular power spectrum of thick redshift bins. For the galaxy number counts, we have

$$\Delta_g(\mathbf{n}, z) = b_g(z)\delta(\mathbf{n}, z) + (5s(z) - 2)\kappa(\mathbf{n}, z), \quad (2)$$

where \mathbf{n} denotes the direction of observation, z is the redshift, b_g is the linear galaxy bias, δ is the matter density contrast, and s is the magnification bias. The first term in Eq. (2) is the density contribution, and the second term is the lensing magnification, proportional to the lensing convergence κ ,

$$\kappa(\mathbf{n}, z) = \frac{1}{2} \int_0^{\chi(z)} d\chi' \frac{\chi(z) - \chi'}{\chi(z)\chi'} \nabla_\Omega^2(\Phi + \Psi)(\mathbf{n}, \chi'), \quad (3)$$

where ∇_Ω^2 denotes the angular Laplacian.¹

Contrary to galaxy number counts, the fluctuations in the brightness temperature are not affected by lensing magnification at linear order in perturbation theory due to conservation of surface brightness [40], and we have

$$\Delta_{\text{HI}}(\mathbf{n}, z) = b_{\text{HI}}(z)\delta(\mathbf{n}, z), \quad (4)$$

where b_{HI} is the bias of neutral hydrogen. In Eqs. (2) and (4) we have neglected redshift-space distortions since they are subdominant for thick redshift bins. In our forecasts, we will choose the binning such that this contribution is always negligible and does not contaminate E_G (see Sec. III).

We expand the galaxy number counts and brightness temperature in spherical harmonics

$$\Delta_g(\mathbf{n}, z) = \sum_{\ell m} a_{\ell m}^g(z) Y_{\ell m}(\mathbf{n}), \quad (5)$$

$$\Delta_{\text{HI}}(\mathbf{n}, z) = \sum_{\ell m} a_{\ell m}^{\text{HI}}(z) Y_{\ell m}(\mathbf{n}), \quad (6)$$

¹Note that we use the following metric convention $ds^2 = a(\eta)^2[-(1 + 2\Psi)d\eta^2 + (1 - 2\Phi)\delta_{ij}dx^i dx^j]$, where η denotes conformal time, so that $\chi = \eta_0 - \eta$ for the conformal time η_0 .

and consider their angular power spectra

$$C_{\ell}^{XY}(z, z') = \langle a_{\ell m}^X(z) a_{\ell m}^{*Y}(z') \rangle, \quad (7)$$

for $X, Y = g, \text{HI}$. We can then build E_G from these spectra in the following way: the numerator of E_G is given by the galaxy-lensing correlation. Since galaxy number counts are affected by lensing magnification whereas intensity mapping is not, the following estimator, proposed for the first time in [36], is directly proportional to the galaxy-lensing correlation

$$\begin{aligned} E_{\ell}^{\times}(z_f, z_b) &\equiv C_{\ell}^{\text{HIg}}(z_f, z_b) - C_{\ell}^{\text{gHI}}(z_f, z_b) \\ &= (5s(z_b) - 2)b_{\text{HI}}(z_f)C_{\ell}^{\delta\kappa}(z_f, z_b) \\ &\quad - (5s(z_f) - 2)b_{\text{HI}}(z_b)C_{\ell}^{\kappa\delta}(z_f, z_b) \\ &\quad + [b_g(z_b)b_{\text{HI}}(z_f) - b_g(z_f)b_{\text{HI}}(z_b)]C_{\ell}^{\delta\delta}(z_f, z_b). \end{aligned} \quad (8)$$

The term we are interested in is the one in the second line, $C_{\ell}^{\delta\kappa}(z_f, z_b)$, which represents the lensing magnification of background galaxies generated by a foreground density at z_f . The other lensing term in the third line, $C_{\ell}^{\kappa\delta}(z_f, z_b)$, is always negligible since it is due to correlations of foreground lensing magnification with background density. Finally, the density term in the last line is suppressed by two effects: first by the bias difference, which would exactly vanish if the biases were redshift independent, and second by the fact that density correlations, $C_{\ell}^{\delta\delta}(z_f, z_b)$, quickly decrease with redshift separation. In Sec. III C we will choose the redshift bins such that this density contamination is negligible. This estimator, called GIMCO [36], provides therefore a robust way of isolating lensing magnification, and we will use it in the numerator of E_G

$$E_{\ell}^{\times}(z_f, z_b) \simeq (5s(z_b) - 2)b_{\text{HI}}(z_f)C_{\ell}^{\delta\kappa}(z_f, z_b). \quad (9)$$

In the denominator of E_G we need the galaxy-velocity correlation. In [41], this correlation was replaced by the product of $\beta_g(z_f) = f(z_f)/b_g(z_f)$ and the galaxy-galaxy correlations. In our case, instead of the galaxy-galaxy correlations, we use the galaxy-intensity mapping correlations, such that the HI bias in Eq. (9) cancels. We therefore have

$$\begin{aligned} E_G &= \frac{E_{\ell}^{\times}(z_f, z_b)}{\beta_g(z_f)C_{\ell}^{\text{gHI}}(z_f, z_f)} = \frac{C_{\ell}^{\text{HIg}}(z_f, z_b) - C_{\ell}^{\text{gHI}}(z_f, z_b)}{\beta_g(z_f)C_{\ell}^{\text{gHI}}(z_f, z_f)} \\ &= \frac{(5s(z_b) - 2)C_{\ell}^{\delta\kappa}(z_f, z_b)}{f(z_f)C_{\ell}^{\delta\delta}(z_f, z_f)}. \end{aligned} \quad (10)$$

We see from Eq. (10) that we only need two types of correlators to measure E_G ; the cross-correlation between galaxy clustering and intensity mapping and the

autocorrelation of galaxy clustering, from which β_g is measured. This is particularly interesting to test deviations from GR; Eq. (10) tests the relation between density, velocity and lensing potential measured from the same correlators. Therefore, the outcome of this test is not subject to potential inconsistencies between different data sets, as can be the case for standard versions of E_G that rely on galaxy clustering and shear or on galaxy clustering and CMB lensing.

As discussed above, the numerator in Eq. (10) is not affected by the lensing contamination computed in [25]; the density-magnification contribution is the *signal* in our estimator, whereas the magnification-magnification contamination is absent since intensity mapping is insensitive to lensing magnification. The denominator in Eq. (10) contains density-magnification contamination and an RSD contamination. In Sec. III we will show that we can choose the redshift binning such that these two contaminations remain negligible.

Note that an E_G estimator using intensity mapping has already been proposed in [42]. However, this estimator differs from ours; in [42] the lensing signal is not measured from clustering as we do here, but from CMB lensing cross-correlated with intensity mapping. This estimator, like ours, is unaffected by magnification contamination. It relies, however on the autocorrelation of intensity mapping to measure β_{HI} as well as C_{ℓ}^{HIHI} . The autocorrelations of intensity mapping will be very challenging to measure due to the difficulty of subtracting the foregrounds accurately. Our estimator relies instead only on cross-correlations of intensity mapping with galaxy clustering, which are unaffected by intensity mapping foregrounds and have already been measured [27–35].

III. SIGNAL-TO-NOISE AND CONTAMINATION

We now study the detectability of E_G with the coming generation of intensity mapping and galaxy surveys. For intensity mapping, we consider the 21 cm intensity mapping survey HIRAX, which will measure the neutral hydrogen distribution in the redshift range of $z = 0.775$ to 2.55 covering 15,000 square degrees of the southern sky [43,44]. For the galaxy survey, we study two examples, one modeled on SKA phase 2, covering 30,000 square degrees, based on the specifications in [45,46], and a 15,000 square-degree *Euclid*-like survey [47], based on [48,49]. We perform separate forecasts for the spectroscopic and photometric samples for the *Euclid*-like scenario. The spectroscopic sample has the advantage of providing an accurate measurement of β_g , through RSD, whereas the photometric sample will measure β_g with relatively large error bars. On the other hand, the photometric sample has a significantly lower shot noise due to the much larger number of galaxies detected and extends to higher redshifts. It is, therefore, interesting to study how the measurement of E_G differs in these two samples. The overlapping redshift range between

HIRAX and the galaxy surveys are $z \in [0.9, 1.8]$ for *Euclid*-like spectroscopic, $z \in [0.78, 2.16]$ for *Euclid*-like photometric and $z \in [0.78, 2.0]$ for SKA2-like. We assume that *Euclid*-like has a 2/3 aerial overlap with HIRAX, leading to $f_{\text{sky}} = 0.242$, whereas the SKA2-like survey fully overlaps with HIRAX leading to $f_{\text{sky}} = 0.363$.

For HIRAX correlated with the *Euclid*-like spectroscopic and the SKA2-like surveys, we consider top-hat redshift bins since the accuracy in the redshift determination is excellent. For the *Euclid*-like photometric surveys, on the other hand, we use the bins given in [49] (see Fig. 3), which we approximate with Gaussian window functions according to Table 4 in [50]. Since we are using the code CLASS [51,52] to compute the angular power spectra, and it does not allow us to choose two different windows at two different redshifts, we also use Gaussian windows for HIRAX in our forecasts in this case.

The signal depends on the galaxy magnification bias, s . For *Euclid*-like photometric, we use the magnification bias measured from the flagship simulation, see Table 1 of [49]. For SKA2, we use the expression from [53], and for *Euclid*-like spectroscopic, we use the model developed in [54]. The signal is by construction independent of galaxy bias, however, the variance of E_G depends on the galaxy bias. Explicit expressions for the biases used to model the different surveys can be found in Appendix B.

A. Limber approximation

E_G depends on the foreground and background redshifts, z_f and z_b . However, in Limber approximation, the background dependence goes away, and E_G directly probes the deviations from GR at redshift z_f . We first relate the metric potentials (that enter in the convergence) to the matter density contrast, δ , using the Poisson equation. Since we want to test the validity of GR, we allow for deviations, encoded in the parameter Σ , see e.g., [3,55–57]

$$k^2(\Phi + \Psi) = -3H_0^2\Omega_m(1+z)\Sigma(z)\delta(k, z). \quad (11)$$

Here Ω_m denotes the matter density parameter today, and H_0 represents the present-time value of the Hubble parameter. In Λ CDM, $\Sigma = 1$ on the scales and redshifts of interest for us. Deviations from Λ CDM can be encoded into a modification to the Poisson equation, and a difference between the two metric potentials, which combine into a Σ generically different from 1 [3,55–57]. Plugging this into the convergence (3), and using Limber approximation [58] we obtain the density-magnification correlation, see e.g., [49]

$$C_\ell^{\delta\kappa}(z_f, z_b) = \frac{\ell(\ell+1)}{(\ell+1/2)^2} \frac{3H_0^2\Omega_m}{2} \int_{z_f^{\min}}^{z_f^{\max}} dz n_{\text{HI}}(z)\Sigma(z) \times (1+z)P_{\delta\delta}\left[\frac{\ell+1/2}{\chi(z)}, z\right] \int_{z_b^{\min}}^{z_b^{\max}} dz' n_g(z') \frac{\chi(z') - \chi(z)}{\chi(z)\chi(z')}, \quad (12)$$

where $n_g(z)$ is the normalized galaxy distribution, $n_{\text{HI}}(z)$ is the normalized intensity distribution and the integrals run over the size of the redshift bins. Note that the ℓ -dependent coefficient in front of the integral can be set to one at large ℓ , where Limber approximation is valid. Moreover, we assume that Σ is independent of k , which is a common assumption in large-scale structure analyses (e.g., [59]), and is motivated by the fact that various theories of modified gravity obey this assumption in the quasistatic approximation, e.g., [60–62].

The density-density correlation can also be simplified using Limber approximation, leading to

$$C_\ell^{\delta\delta}(z_f, z_f) = \int_{z_f^{\min}}^{z_f^{\max}} dz \frac{H(z)}{\chi^2(z)} n_g(z) n_{\text{HI}}(z) P_{\delta\delta}\left[\frac{\ell+1/2}{\chi(z)}, z\right]. \quad (13)$$

Inserting Eqs. (12) and (13) in E_G (10) and assuming that the functions $\Sigma(z)$, $\chi(z)$, $H(z)$, and $n_g(z)$ vary slowly with redshift inside each redshift bin, we obtain

$$E_G(\ell, z_f, z_b) = \Gamma(z_f, z_b) \frac{\Omega_m \Sigma(z_f)}{f(z_f)}, \quad (14)$$

where

$$\begin{aligned} \Gamma(z_f, z_b) &\equiv (5s(z_b) - 2) \frac{3H_0^2}{2H(z_f)} (1 + z_f) \\ &\times \int_{z_f^{\min}}^{z_f^{\max}} dz n_{\text{HI}}(z) \int_{z_b^{\min}}^{z_b^{\max}} dz' n_g(z') \frac{\chi(z') - \chi(z)}{\chi(z)\chi(z')} \\ &\times \left(\int_{z_f^{\min}}^{z_f^{\max}} dz n_g(z) n_{\text{HI}}(z) \right)^{-1} \\ &\simeq (5s(z_b) - 2) \frac{3H_0^2}{2H(z_f)} (1 + z_f) \\ &\times \frac{n_g(z_b)}{n_g(z_f)} \frac{1}{\chi(z_f)} \int_{z_b^{\min}}^{z_b^{\max}} dz \frac{\chi(z) - \chi(z_f)}{\chi(z)}. \end{aligned} \quad (15)$$

The second expression corresponds to top-hat bins, while the first one allows for more general redshift bins.

The coefficient Γ depends on the background cosmology, the galaxy distribution and the magnification bias. A common assumption in large-scale structure analyses is to fix the background to Λ CDM (since the CMB has extremely well constrained it) and to test for deviations from General Relativity at the level of the perturbations. This is, for example, the strategy adopted in RSD analyses to measure β_g in a model-independent way, see e.g., [63]. Here we follow the same procedure. The galaxy distribution and the magnification bias are quantities that can directly be measured for a given population of galaxies and a given survey. Consequently, Γ is a parameter we can predict. Note that we do not include Ω_m in Γ as it cannot be measured in a model-independent way when allowing for general dark

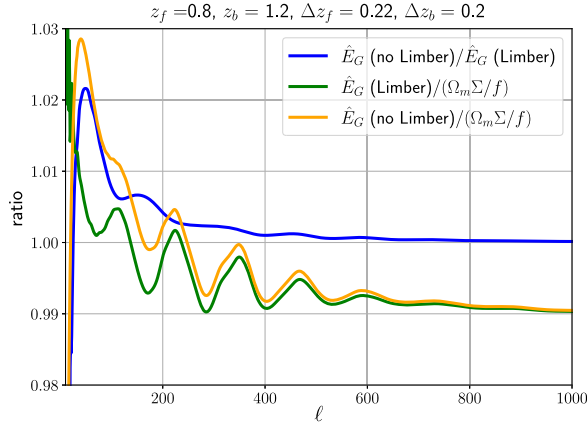


FIG. 1. Accuracy of the approximations used in our estimator: in blue, we show the ratio of \hat{E}_G computed without and with Limber approximation; in green the ratio with and without redshift evolution, and in orange the ratio without and with both approximations.

energy or modified-gravity scenarios [4,64]. We then define a rescaled E_G variable, by dividing it with $\Gamma(z_f, z_b)$

$$\begin{aligned} \hat{E}_G(\ell, z_f, z_b) &\equiv \frac{E_\ell^\times(z_f, z_b)}{\Gamma(z_f, z_b)\beta_g(z_f)C_\ell^{\text{gHI}}(z_f, z_f)} \\ &= \frac{\Omega_m \Sigma(z_f)}{f(z_f)}. \end{aligned} \quad (17)$$

We have tested the validity of the approximations used to obtain Eq. (17). In Fig. 1 we compare \hat{E}_G using the full-sky angular power spectra, with the result using Limber approximation, and with the second line of Eq. (17), which further neglects redshift evolution within each bin. Limber approximation induces a subpercent error on \hat{E}_G for ℓ above 100 (and less than 0.1% for $\ell \geq 400$). Neglecting redshift evolution induces a $\sim 1\%$ percent error at all ℓ 's. This is well below the statistical uncertainties on \hat{E}_G , as we will see in Sec. IV. The model-independent and ℓ -independent function $\Gamma(z_f, z_b)$ defined in Eq. (16) is therefore perfectly adapted and allows us to define \hat{E}_G in a model-independent way, in contrast to the function $C_\Gamma(\ell)$ used in [21], which depends on the density power spectrum and is thus model dependent.²

B. Signal-to-noise ratio

We now compute the signal-to-noise ratio (SNR) of \hat{E}_G for the three combinations of surveys. Since the signal is independent of the background redshift, as discussed above, we sum over all pairs at fixed z_f ;

²Note that to properly test the validity of Eq. (17), it is necessary to use the same distribution functions, n_g and n_{HI} , in Eqs. (12), (13), and (16). In our case, we use top-hat window functions approximated by a smooth function in CLASS, with a sharpness that can be adjusted.

$$\begin{aligned} \left(\frac{S}{N}\right)^2(z_f) &= \sum_{\ell, \ell'=\ell_{\min}}^{\ell_{\max}} \sum_{z_b, z'_b=z_{\min}}^{z_{\max}} \hat{E}_G(\ell, z_f, z_b) \\ &\times \text{cov}^{-1}[\hat{E}_G(\ell, z_f, z_b), \hat{E}_G(\ell', z_f, z'_b)] \\ &\times \hat{E}_G(\ell', z_f, z'_b). \end{aligned} \quad (18)$$

The sum runs over all multipoles used in the analysis and all background redshift bins. In Sec. III C we will select those such that the contamination from RSD, lensing and density are negligible.

The full expression for the covariance of \hat{E}_G is given in Appendix A. It depends on the covariance of C_ℓ^{gHI} , on the variance of β_g and the covariance between the two. The covariance of C_ℓ^{gHI} is given by

$$\begin{aligned} \text{cov}[C_\ell^{\text{gHI}}(z_1, z_2), C_\ell^{\text{gHI}}(z_3, z_4)] \\ &= \frac{1}{f_{\text{sky}}(2\ell + 1)} \\ &\times [C_\ell^{\text{gg}}(z_1, z_3)C_\ell^{\text{HIHI}}(z_2, z_4) + C_\ell^{\text{gHI}}(z_1, z_4)C_\ell^{\text{HIg}}(z_2, z_3)]. \end{aligned} \quad (19)$$

All terms are affected by cosmic variance. In addition, the galaxy-galaxy correlation is affected by shot noise when $z_1 = z_3$

$$C_\ell^{\text{gg sn}}(z_1, z_3) = \frac{\delta_{z_1, z_3}}{\bar{N}(z_1)}, \quad (20)$$

where $\bar{N}(z_1)$ denotes the mean number of galaxies per redshift bin and steradian. The HI – HI correlation is affected by shot noise and interferometer noise; however, it has been shown in [65,66] that the former is always subdominant with respect to the latter. In [36], we derived an expression for interferometer noise for HIRAX, based on an analytical expression derived in [67], adapted to the outcome of numerical simulations for HIRAX [65,66].

Finally, the cross-correlation between the galaxy and intensity mapping is not affected by shot noise or interferometer noise.³

As shown in Appendix A, the covariance of \hat{E}_G is also affected by the variance of β_g . This quantity is measured from the multipoles of the correlation function (or the power spectrum), which are directly sensitive to RSD. Spectroscopic surveys are designed to measure β_g very precisely and are expected to reach a precision of 1% [69].

³Note that, as shown in [68], there is a residual shot noise contribution in the cross-correlation due to the galaxies which contribute to both the number counts and the intensity mapping signal. This contribution is however negligible in the covariance, since it adds to the covariance from HI-HI, which is dominated by interferometer noise, that is always significantly larger than shot noise [65,66].

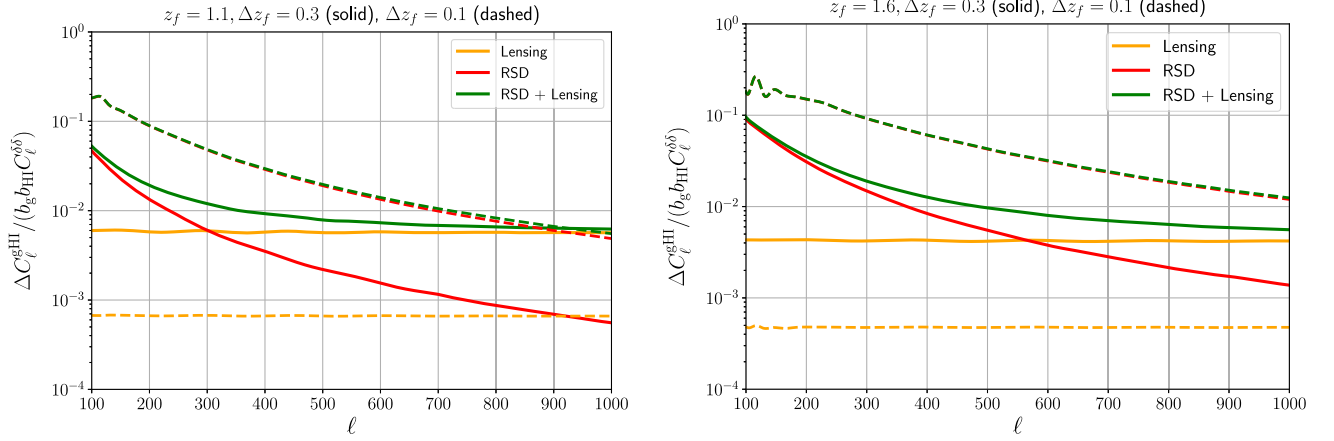


FIG. 2. Relative contamination to $C_\ell^{\text{gHI}}(z_f, z_f)$ due to lensing magnification and RSD. The left panel shows relative contaminations at $z_f = 1.1$ and the right panel at $z_f = 1.6$. For each panel there are two sets of curves: $\Delta z_f = 0.1$ (dashed) and $\Delta z_f = 0.3$ (solid). For the narrower choice of Δz_f , the RSD contribution dominates, so that the green and red dashed curves nearly coincide.

Photometric surveys can also measure β_g but with much less precision, of the order of 10% [70]. We use these two values for the spectroscopic and photometric forecasts. Note that, as shown in Appendix A, the variance of β_g generates nondiagonal contributions, with $\ell \neq \ell'$, to the covariance of \hat{E}_G . This is because β_g is a parameter measured from the full correlation function in each redshift bin. Its error is, therefore, fully correlated for different values of the angular multipoles ℓ and ℓ' . Neglecting these correlations, as has been done in [71], would underestimate the β_g contribution to the variance by a factor of roughly $2\ell + 1$.

Finally, the covariance of \hat{E}_G depends on the covariance between C_ℓ^{gHI} and β_g . Even though this covariance is not precisely zero (since β_g and C_ℓ^{gHI} can be measured from the same volume), it is highly suppressed and can be neglected; as has been shown in [72], C_ℓ^{gHI} is mainly insensitive to small radial modes, due to the size of the redshift bins, which washes out small-scale information, whereas β_g is measured from small radial modes. This leads to a negligible covariance between these two types of measurements.

C. Contaminations

From Eq. (18) we see that the SNR depends on the range of multipoles used in the analysis and the redshift bins. We determine those to reduce the contaminations to \hat{E}_G . As discussed previously, the numerator in \hat{E}_G is not affected by the lensing contamination that affects standard \hat{E}_G estimators [25]. However, when z_f is close to z_b , the numerator is affected by density-density contamination, given by the last line in Eq. (8). In order to reduce this contamination to acceptable levels, once the redshift bins have been fixed, we will compute the contamination for each pair and remove those for which it is too high.

The denominator in \hat{E}_G is affected by contamination from lensing magnification and by contamination from RSD. The first one has been discussed in [25], whereas the second one is usually ignored. We aim to choose the widths of the redshift bins to minimize these two contaminations without compromising the SNR. The contribution from RSD and lensing magnification to the galaxy-HI spectrum at equal redshift can be expressed as

$$\begin{aligned} \Delta C_\ell^{\text{gHI}}(z_f, z_f) &= b_{\text{HI}}(z_f)(5s(z_f) - 2)C_\ell^{\delta\kappa}(z_f, z_f) \\ &+ [b_g(z_f) + b_{\text{HI}}(z_f)]C_\ell^{\text{RSD}}(z_f, z_f) \\ &+ C_\ell^{\text{RSD RSD}}(z_f, z_f). \end{aligned} \quad (21)$$

The term in the first line is the lensing magnification contamination, and the two terms in the second line are the contaminations from RSD. In Fig. 2 we show the relative contamination from lensing magnification and RSD for $z_f = 1.1$ and $z_f = 1.6$ and for two values of the width: $\Delta z_f = 0.1$ and $\Delta z_f = 0.3$. We choose here the *Euclid*-like spectroscopic survey specifications. We see that for thin bins (dashed curves), the contamination due to lensing is less than 0.1% for both redshifts, $z_f = 1.1$ and $z_f = 1.6$, while RSD contributes more than 1% on a wide range of scales. For thick bins (solid curves), the situation is different. Here RSD is still the dominant contamination on large scales, but on small scales ($\ell \gtrsim 300$ for $z_f = 1.1$ and $\ell \gtrsim 550$ for $z_f = 1.6$) the lensing magnification contamination dominates. We see that this contamination does not increase with redshift. This is because, in Limber approximation, only the intra-bin lensing contributes. This contribution decreases with redshift roughly at the same rate as the density-density contribution. However, the bias increases faster with redshift than the magnification bias factor $5s - 2$

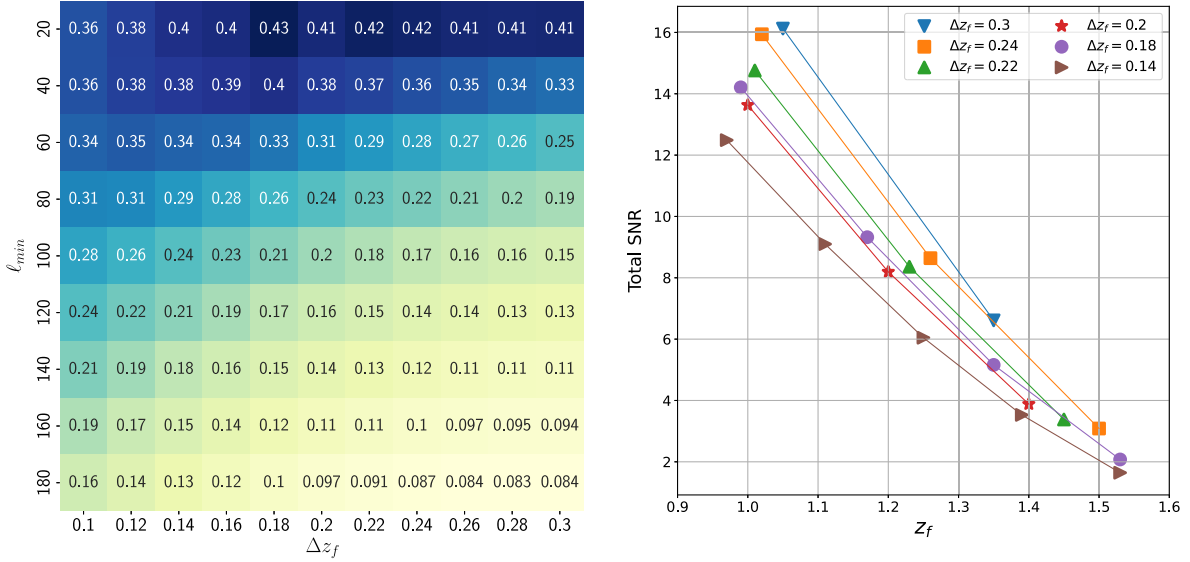


FIG. 3. *Left panel*: SNR contamination from RSD and lensing magnification for a *Euclid*-like spectroscopic survey as a function of ℓ_{\min} and Δz_f , for $z_f = 1.1$, $z_b = 1.6$, and $\Delta z_b = 0.22$. *Right panel*: Total SNR as a function of z_f for different values of Δz_f . Note that in each case, the size of the last redshift bin is adapted to cover the whole available redshift range.

(see Fig. 6 in Appendix B), leading, overall, to a slight decrease with redshift of the relative lensing contamination. Overall the wider redshift bin choice is clearly better in terms of total contamination, which remains below 3% for $\ell \gtrsim 100$ for $z_f = 1.1$ and $\ell \gtrsim 200$ for $z_f = 1.6$.

Based on Fig. 2 we see that for each foreground redshift z_f , we can choose a width Δz_f that would minimize the total contamination. Here we have compared the contamination with the density contribution $b_g b_{\text{HI}} C_\ell^{\delta\delta}$. In practice, however, what will determine if our estimator is biased or not is the size of the contamination with respect to the variance; as a rule of thumb, if the contamination leads to a contribution in \hat{E}_G which is of the same order as its variance, we expect a bias of the order of 1σ on the best-fit parameters extracted from \hat{E}_G . We define, therefore, the following SNR associated with the contaminations

$$\left(\frac{\Delta S}{N}\right)^2(z_f) \equiv \sum_{\ell, \ell' = \ell_{\min}}^{\ell_{\max}} \sum_{z_b, z'_b = z_{\min}}^{z_{\max}} \Delta \hat{E}_G(\ell, z_f, z_b) \times \text{cov}^{-1}[\hat{E}_G(\ell, z_f, z_b), \hat{E}_G(\ell', z_f, z'_b)] \Delta \hat{E}_G(\ell', z_f, z'_b). \quad (22)$$

Here $\Delta \hat{E}_G$ is the difference between the contaminated signal, and the uncontaminated one,

$$\Delta \hat{E}_G(\ell, z_f, z_b) = \left\{ \frac{E_\ell^\times(z_f, z_b)}{\beta_g(z_f)(b_g(z_f)b_{\text{HI}}(z_f)C_\ell^{\delta\delta}(z_f, z_f) + \Delta C_\ell^{\text{gHI}}(z_f, z_f))} - \frac{E_\ell^\times(z_f, z_b)}{\beta_g(z_f)b_g(z_f)b_{\text{HI}}(z_f)C_\ell^{\delta\delta}(z_f, z_f)} \right\} \frac{1}{\Gamma(z_f, z_b)}, \quad (24)$$

where $\Delta C_\ell^{\text{gHI}}$ contains the contamination from RSD and lensing, defined in Eq. (21).

Our strategy to choose optimal redshift bins is based on two factors: minimizing the contamination (22) and maximizing the total signal to noise (18). There are four variables that we need to optimize: ℓ_{\min} , ℓ_{\max} , Δz_f , and Δz_b . We choose $\ell_{\max} = 1000$, since above this value, the nonlinear modeling of lensing magnification and density fluctuations is uncertain. In Fig. 3 (left panel) we plot the SNR contamination as a function of ℓ_{\min} and Δz_f , for the redshift pair ($z_f = 1.1$, $z_b = 1.6$), and for fixed $\Delta z_b = 0.22$. We show the results for the case of the *Euclid*-like spectroscopic survey. We see that at fixed ℓ_{\min} , the contamination decreases as we increase the bin size. This is because the RSD contamination decreases for thick bins. The lensing contamination, on the other hand, increases with the size of the bin, but it never overcomes the contamination from RSD once we sum over multipoles. Moreover, the contamination decreases quickly when increasing ℓ_{\min} . This is again due to RSD, which peaks at low ℓ as can be seen from Fig. 2.

To optimally choose Δz_f and ℓ_{\min} , we need to put Fig. 3 in balance with the SNR for \hat{E}_G . Clearly, decreasing ℓ_{\min} will increase the SNR. A good compromise is to choose $\ell_{\min} = 100$. This is also motivated by the fact that IM surveys like HIRAX may not have good enough calibration to precisely measure lower multipoles [67]. Concerning the size of the bins; the SNR at each z_f increases with the size of the bins, but if we want to test the evolution of \hat{E}_G with redshift, we want to have a reasonable number of foreground bins. In Fig. 3 (right panel), we plot the SNR summed over all background bins as a function of foreground redshift z_f and

TABLE I. Optimal redshift bin configurations for *Euclid*-like spectroscopic, photometric, and SKA2-like surveys.

<i>Euclid</i> spectro		<i>Euclid</i> photo		SKA2	
z	Δz	z	Δz	z	Δz
1.02	0.24	0.84	0.2	1.02	0.24
1.26	0.24	1.00	0.2	1.26	0.24
1.50	0.24	1.14	0.2	1.50	0.24
1.71	0.18	1.30	0.2	1.74	0.24
		1.44	0.2	1.93	0.14
		1.62	0.2		
		1.78	0.5		
		1.91	0.5		

 TABLE II. SNR contamination from density for the *Euclid*-like spectroscopic survey.

z_f	$z_b = 1.26$	$z_b = 1.50$	$z_b = 1.71$
1.02	0.09	3×10^{-16}	7×10^{-16}
1.26		0.006	4×10^{-16}
1.50			0.04

TABLE III. SNR contamination from density for the SKA2-like survey.

z_f	$z_b = 1.26$	$z_b = 1.50$	$z_b = 1.74$	$z_b = 1.91$
1.02	1.40	10^{-16}	10^{-16}	10^{-16}
1.26		0.94	10^{-16}	10^{-16}
1.50			0.60	10^{-17}
1.74				1.54

for different sizes of the bins. We see that $\Delta z_f = 0.24$ is optimal; it allows us to measure \hat{E}_G for three foreground redshifts (versus two for $\Delta z_f = 0.3$) and has a total SNR of 18.4, larger than for all other cases. The thinner choice, $\Delta z_f = 0.14$, is interesting since it allows us to measure \hat{E}_G in a higher number of redshift bins. However, the overall SNR (summed over redshift) is 8 percent lower than for $\Delta z_f = 0.24$. We therefore choose $\ell_{\min} = 100$ and $\Delta z_f = 0.24$, which lead to a small SNR contamination of

0.17. Note that if we reduce ℓ_{\min} to 60, the SNR increases only very slightly from 18.4 to 18.9, whereas the contamination increases from 0.17 to 0.28. We conclude therefore that $\ell_{\min} = 100$ is a better choice. This procedure needs to be performed for each value of the foreground redshift z_f . We find similar results at all redshifts, with a maximum SNR contamination of 0.21 for the pair $z_f = 1.02$, and $z_b = 1.71$, which is still perfectly acceptable.

We then repeat the same procedure for the *Euclid*-like photometric and the SKA2-like surveys. In Table I, we list the optimal binning for each case.

Once the binning has been fixed, we compute the SNR contamination from density in the numerator of \hat{E}_G [due to the last line in Eq. (8)], for each pair (z_f, z_b) . The results for the *Euclid*-like spectroscopic survey are shown in Table II. This SNR contamination is at most 9% (for neighboring pairs). We can therefore keep all pairs in our forecast. For SKA2, the density contamination for adjacent pairs is more significant, as seen in Table III. This is because the variance of \hat{E}_G is smaller for SKA2 due to the broader sky coverage and lower shot noise. Consequently, the contamination from density relative to the variance is larger for SKA2 than for *Euclid* spectroscopic. To avoid bias in our estimator, we remove adjacent pairs for SKA2.

For *Euclid*-like photometric, the situation is different since we are using Gaussian window functions. In this case, there is still a significant overlap of nonconsecutive bins, especially at high redshift, leading to non-negligible density contamination, as seen from Table IV. Based on these results, we remove all pairs with contamination larger than 0.25. With this, we ensure that the bias from such contamination is below 0.25σ , which is acceptable. We see that with this criteria, we have only three foreground bins for *Euclid*-like photometric. Note that in reality, we expect the contamination to be smaller; here, we have used a Gaussian window for HIRAX as well, due to the limitation of CLASS to have two different window functions. In practice, however, HIRAX can select bins with sharp edges for which the bins overlap, and consequently, the density contamination will be reduced.

Overall, our choice of binning is such that the total contamination from RSD, lensing and density is at most 0.3σ for *Euclid*-like spectroscopic (for $z_f = 1.02$), 0.6σ for SKA2-like (for $z_f = 1.02$), and 0.4σ for *Euclid*-like

 TABLE IV. SNR contamination from density for the *Euclid*-like photometric survey.

z_f	$z_b = 1.0$	$z_b = 1.14$	$z_b = 1.30$	$z_b = 1.44$	$z_b = 1.62$	$z_b = 1.78$	$z_b = 1.91$
0.84	40.8	3.7	0.25	0.004	10^{-6}	10^{-3}	10^{-3}
1.00		22.6	1.8	0.15	10^{-3}	0.1	0.03
1.14			27.1	4.3	0.08	0.24	0.14
1.30				12.8	1.4	2.7	1.3
1.44					5.7	10.3	5.8
1.62						66.8	51.8
1.78							28.7

photometric (for $z_f = 1.14$). This means that our estimator is robust and will lead to a measurement of $\Omega_m \Sigma / f$ that is not significantly biased (half a sigma at most). Note that one possibility to reduce the bias further would be to model the contamination using a reference cosmology, and to add it to the model. This ‘reference contamination’ can then be kept fixed. This is not correct (since the contamination depends on cosmology), but since it is already subdominant, the bias generated by keeping it fixed would be very small.

IV. MODIFIED GRAVITY FORECASTS

From Eq. (17), we see that constraints on $\hat{E}_G(z_f)$ directly translate into constraints on the combination of parameters $\Omega_m \Sigma(z_f) / f(z_f)$. For each z_f , the Fisher element for this combination reads

$$\begin{aligned} \mathcal{F} &= \sum_{\ell, \ell' = \ell_{\min}}^{\ell_{\max}} \sum_{z_b, z'_b = z_{\min}}^{z_{\max}} \frac{d\hat{E}_G(\ell, z_f, z_b)}{d(\Omega_m \Sigma / f)} \\ &\quad \times \text{cov}^{-1}[\hat{E}_G(\ell, z_f, z_b), \hat{E}_G(\ell', z_f, z'_b)] \frac{d\hat{E}_G(\ell', z_f, z'_b)}{d(\Omega_m \Sigma / f)} \\ &= \sum_{\ell, \ell' = \ell_{\min}}^{\ell_{\max}} \sum_{z_b, z'_b = z_{\min}}^{z_{\max}} \text{cov}^{-1}[\hat{E}_G(\ell, z_f, z_b), \hat{E}_G(\ell', z_f, z'_b)]. \end{aligned} \quad (25)$$

In Fig. 4 we show the relative error on $\Omega_m \Sigma / f$ for the three combinations of surveys, using the optimal redshift bins defined in Table I. The constraints expected from a *Euclid*-like spectroscopic survey are very good at low redshift, around 7%. The precision decreases as the foreground redshift increases, reaching 32% at $z_f = 1.50$. The fact that the constraints are significantly better at low redshift is due to two reasons; firstly, when z_f is small, we can correlate it with a larger number of background redshift

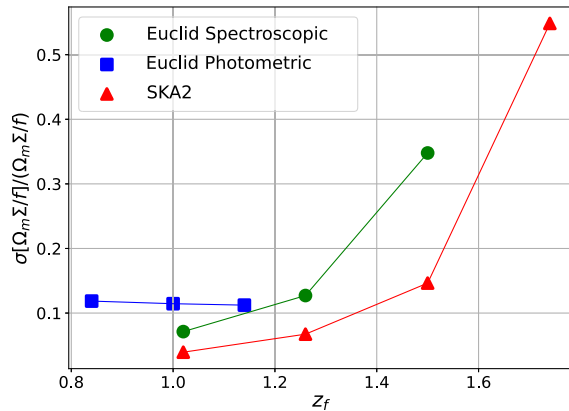


FIG. 4. Relative uncertainty on $\Omega_m \Sigma / f$ as function of foreground redshift for *Euclid*-like spectroscopic, *Euclid*-like photometric and SKA2-like. In the Fisher analysis we use $\ell_{\min} = 100$, $\ell_{\max} = 1000$, and the redshift bins given in Table I.

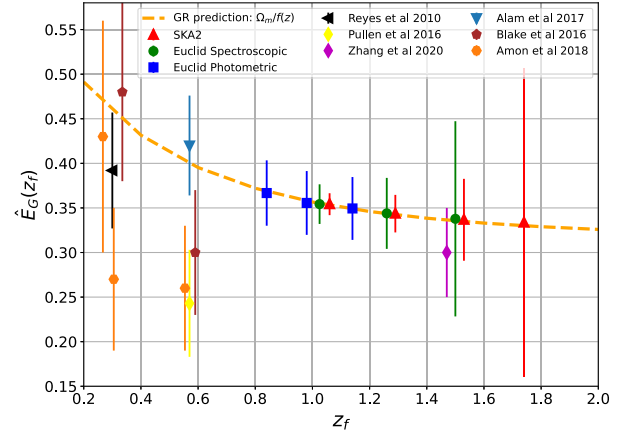


FIG. 5. \hat{E}_G and its error bars, as function of foreground redshift, z_f . We compare our results from Fisher forecasts with measurements from [14,16,21–23,74]. The dashed orange line is the GR prediction, $\hat{E}_G(z) = \Omega_m / f(z)$ calculated with our fiducial cosmology.

bins, which increases the precision of the measurement. Secondly, at low redshift, the number density for a *Euclid*-like spectroscopic survey is significantly larger than at high redshift, leading to a strong suppression of shot noise.

On the other hand, a *Euclid*-like photometric survey has a precision of $\sim 11\%$ for all foreground redshifts. The primary source of error for photometric surveys is the error on β_g (of 10%) since this parameter cannot be well constrained from a photometric survey. This consequently limits the precision with which \hat{E}_G can be measured. To model this uncertainty correctly, it is crucial to include the β_g covariance fully correlated in ℓ as in (A2). As shown in Appendix A, this nondiagonal contribution follows from the fact that β_g is measured by combining all available scales (from the multipoles of the power spectrum) and not independently for each value of ℓ . Neglecting these nondiagonal terms in the covariance leads to an uncertainty on \hat{E}_G that is smaller than the uncertainty on β_g alone, like for example in [71], where \hat{E}_G is predicted to be measured with a precision of 1% with LSST, even though LSST can measure β_g with a precision of 10% only, which is not possible. Similarly, in [73], \hat{E}_G is forecasted to be measured with a precision of 1% with LSST and photometric *Euclid*, using CMB lensing, assuming that β_g can be measured with a 10% uncertainty. Again, this is not possible: \hat{E}_G cannot be measured better than β_g is. As before, the error in these forecasts comes from the fact that the error on β_g is wrongly assumed to be uncorrelated in ℓ [see their Eq. (17)].

The constraints predicted for SKA2 are tighter by a factor of 2 with respect to the constraints from *Euclid*-like spectroscopic, reaching 3.9% at low redshift. This is mainly due to the higher number density of SKA2, the broader sky coverage, and the larger redshift range.

TABLE V. Previous measurements of \hat{E}_G .

Authors	z	$\hat{E}_G(z)$	Agreement
Reyes <i>et al.</i> [14]	0.37	0.392 ± 0.065	Agrees with GR
Blake <i>et al.</i> [18]	0.32	0.48 ± 0.10	Agrees with GR
	0.57	0.30 ± 0.07	1.4σ deviation
Amon <i>et al.</i> [16]	0.27	0.43 ± 0.13	Agrees with GR
	0.31	0.27 ± 0.08	2.3σ deviation
	0.55	0.26 ± 0.07	2σ deviation
Alam <i>et al.</i> [22]	0.57	0.42 ± 0.056	Agrees with GR
Pullen <i>et al.</i> [41]	0.57	0.24 ± 0.060	2.6σ deviation

However, the constraints are quite poor in the last redshift bin, of 54% only. This is due to the large shot noise at high redshift, which significantly degrades the constraints, and to the fact that only one pair can be used for this case.

In Fig. 5, we compare our forecasted constraints with current measurements of \hat{E}_G from [14,16,21,23]. The orange dashed line represents the GR prediction, $\hat{E}_G(z) = \Omega_m/f(z)$, computed with our fiducial choice of cosmology.⁴ The best measurement for \hat{E}_G is the one from [22] at $z_f = 0.57$ which has a precision of 13%. We see from Fig. 5 that our method to measure \hat{E}_G from clustering of galaxies and intensity mapping has the potential to significantly tighten these constraints over a wide range of redshifts, allowing us to test the validity of GR robustly on cosmological scales.

V. CONCLUSION

Testing the consistency of GR at cosmological scales is one of the main goals of modern cosmology. Since a large number of theories have been constructed where gravity is modified or where a dark energy component has been added, confronting each theory individually with observations is no longer feasible. It is, therefore, necessary to build model-independent tests. In this paper, we focused on one of these tests: the \hat{E}_G statistic, which compares the evolution of the sum of the potentials with that of the velocity.

The \hat{E}_G statistic has been successfully measured with various data sets, showing mild tensions with the GR predictions in some cases. It is therefore of great interest to see if future surveys will confirm or not these tensions. However, the currently used estimators of \hat{E}_G suffer from two problems; first, they are affected by lensing

⁴Note that the measurements depend on the choice of background cosmology through Γ and can therefore not directly be compared with our prediction to assess if there is a tension or not. In Table V we provide for this reason a summary of the tension (or agreement) quoted from the respective studies.

magnification, a contamination that will become important for future surveys and may invalidate their use. Second, they mix different observables (clustering and shear, or clustering and CMB lensing), making them sensitive to tensions between these different data sets, due for example to different systematics.

In this paper, we have proposed an alternative estimator for \hat{E}_G , which relies *only on clustering measurements* from two different tracers: galaxies and intensity mapping. We have shown that this estimator is very robust; it only uses cross-correlations that are less affected by systematics, and by choosing the binning in which \hat{E}_G is measured appropriately, the contaminations are below 0.3σ for a *Euclid*-like spectroscopic survey, below 0.6σ for a SKA2-like survey, and below 0.4σ for a *Euclid*-like photometric survey. Moreover, our estimator depends on a choice of fiducial cosmology only through background parameters, namely through $H(z)$ and distances. Contrary to some estimators used in the literature, it does not depend on the evolution of perturbations, like in [21].

Using Fisher forecasts, we have found that \hat{E}_G , and consequently deviations from GR, can be measured over a wide range of redshifts and with a precision of up to 7% with a *Euclid*-like spectroscopic survey and 3.9% with a SKA2-like survey. We have also explicitly shown that a *Euclid*-like photometric survey cannot reach such a good precision on \hat{E}_G . The reason is that for a photometric survey, β_g is measured at best with a precision of 10%, which intrinsically limits the precision on \hat{E}_G . This point was not properly accounted for in some previous forecasts, which claim a 1% precision on \hat{E}_G using photometric surveys like *Euclid* or LSST [71,73]. The error in these forecasts comes from the covariance matrix, which does not properly include correlations in ℓ 's generated by the fact that β_g is measured from a combination of all scales, and not independently for each values of ℓ .

Our study shows therefore that the combination of intensity mapping and galaxy clustering provides an excellent strategy to measure \hat{E}_G robustly and precisely. This will dramatically improve current constraints and extend them to higher redshifts, allowing us to test deviations from GR very accurately.

ACKNOWLEDGMENTS

It is a pleasure to thank Patrick Simon for useful comments on the manuscript. M. A. and C. B. acknowledge financial support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 863929; project title "Testing the law of gravity with novel large-scale structure observables"). M. J. and M. K. acknowledge financial support from the Swiss National Science Foundation.

APPENDIX A: COVARIANCE OF \hat{E}_G

We compute the covariance of $\hat{E}_G(\ell, z_f, z_b)$ for a fixed value of z_f since we measure it independently in each foreground redshift bin. However, since we combine different background redshift bins in the measurements, we need to account for the covariance of \hat{E}_G between different background redshifts z_b and z'_b . Moreover, since β_g , which enters into \hat{E}_G , is measured by combining all scales (and not independently for each value of ℓ),

it will induce correlations between \hat{E}_G at different values of ℓ .

In general, for $X = f(A, B)$ and $X' = f(A', B')$, where f is a generic function, the covariance of X and X' can be expressed as

$$\text{Cov}(X, X') = \sum_{\substack{\alpha=A, B \\ \beta=A', B'}} \frac{\partial X}{\partial \alpha} \text{Cov}(\alpha, \beta) \frac{\partial X'}{\partial \beta}. \quad (\text{A1})$$

Using this formula, the covariance of \hat{E}_G becomes

$$\begin{aligned} \text{Cov}[\hat{E}_G(\ell, z_f, z_b), \hat{E}_G(\ell', z_f, z'_b)] &= \left\{ \frac{\text{Cov}[E_\ell^\times(z_f, z_b), E_{\ell'}^\times(z_f, z'_b)]}{E_\ell^\times(z_f, z_b) \hat{E}_\ell^\times(z_f, z'_b)} + \frac{\text{Var}[C_\ell^{\text{gHI}}(z_f, z_f)]}{(C_\ell^{\text{gHI}}(z_f, z_f))^2} - \frac{\text{Cov}[E_\ell^\times(z_f, z_b), C_{\ell'}^{\text{gHI}}(z_f, z_f)]}{E_\ell^\times(z_f, z_b) C_{\ell'}^{\text{gHI}}(z_f, z_f)} \right. \\ &\quad \left. - \frac{\text{Cov}[E_\ell^\times(z_f, z'_b), C_\ell^{\text{gHI}}(z_f, z_f)]}{E_\ell^\times(z_f, z'_b) C_\ell^{\text{gHI}}(z_f, z_f)} \right\} E_G(\ell, z_f, z_b) E_G(\ell', z_f, z'_b) \delta_{\ell, \ell'}^{\text{D}} \\ &\quad + \hat{E}_G(\ell, z_f, z_b) \hat{E}_G(\ell', z_f, z'_b) \frac{\text{Var}[\beta_g]}{\beta_g^2}, \end{aligned} \quad (\text{A2})$$

where $\delta_{\ell, \ell'}^{\text{D}}$ represents the Dirac delta function. Here, we have neglected the covariance between β_g and both C_ℓ^{gHI} and E_ℓ^\times , as discussed in the main text.

The covariance of GIMCO, which enters in the second line of Eq. (A2) is given by

$$\begin{aligned} &\text{Cov}[E_\ell^\times(z_f, z_b), E_{\ell'}^\times(z_f, z'_b)] \\ &= \text{Cov}[C_\ell^{\text{gHI}}(z_b, z_f), C_{\ell'}^{\text{gHI}}(z'_b, z_f)] \\ &\quad - \text{Cov}[C_\ell^{\text{gHI}}(z_b, z_f), C_{\ell'}^{\text{gHI}}(z_f, z'_b)] \\ &\quad - \text{Cov}[C_\ell^{\text{gHI}}(z_f, z_b), C_{\ell'}^{\text{gHI}}(z'_b, z_f)] \\ &\quad + \text{Cov}[C_\ell^{\text{gHI}}(z_f, z_b), C_{\ell'}^{\text{gHI}}(z_f, z'_b)], \end{aligned} \quad (\text{A3})$$

where the covariance of C_ℓ^{gHI} is given in Eq. (19).

Similarly, the covariance of E_ℓ^\times with $C_{\ell'}^{\text{gHI}}$, which enters in the fourth and fifth lines of Eq. (A2) can be written as

$$\begin{aligned} &\text{Cov}[\hat{E}_\ell^\times(z_f, z_b), C_{\ell'}^{\text{gHI}}(z_f, z_f)] \\ &= \text{Cov}[C_\ell^{\text{gHI}}(z_b, z_f), C_{\ell'}^{\text{gHI}}(z_f, z_f)] \\ &\quad - \text{Cov}[C_\ell^{\text{gHI}}(z_f, z_b), C_{\ell'}^{\text{gHI}}(z_f, z_f)]. \end{aligned} \quad (\text{A4})$$

APPENDIX B: GALAXY AND INTENSITY MAPPING SURVEYS

This paper uses three example surveys for galaxy clustering: *Euclid*-like spectroscopic, *Euclid*-like photometric, and SKA2-like, and one intensity mapping survey, HIRAX.

The bias of the intensity mapping from HIRAX is given by [36]

$$b_{\text{HI}}(z) = 0.677(1 + 3.8 \times 10^{-1}z + 6.7 \times 10^{-2}z^2). \quad (\text{B1})$$

For the *Euclid*-like photometric survey, we use the bias and magnification bias that have been measured from the flagship simulation in each of the redshift bins we are using, see Table 1 of [49].

For the SKA2-like survey, we use the bias, and magnification bias modeled in [45]:

$$b_g(z) = 0.5887 \exp(0.813z), \quad (\text{B2})$$

$$s(z) = s_0 + s_1z + s_2z^2 + s_3z^3, \quad (\text{B3})$$

with $s_0 = -0.1068$, $s_1 = 1.359$, $s_2 = -0.620008$, and $s_3 = 0.1885$.

Finally, for the *Euclid*-like spectroscopic survey we use the bias given in [48] (Table 3). Since we are not using the same bins as there, we interpolate this bias with the following function

$$b_g(z) = -1.094 + 1.813\sqrt{1+z}. \quad (\text{B4})$$

The magnification bias for *Euclid*-like spectroscopic has never been modeled nor measured in simulations. We use, therefore, for simplicity, the expression given in [54] (which is another model for SKA2)

$$s(z) = 0.9329 - 1.562 \exp(-2.437z). \quad (\text{B5})$$

Galaxy and magnification biases are plotted in Fig. 6.

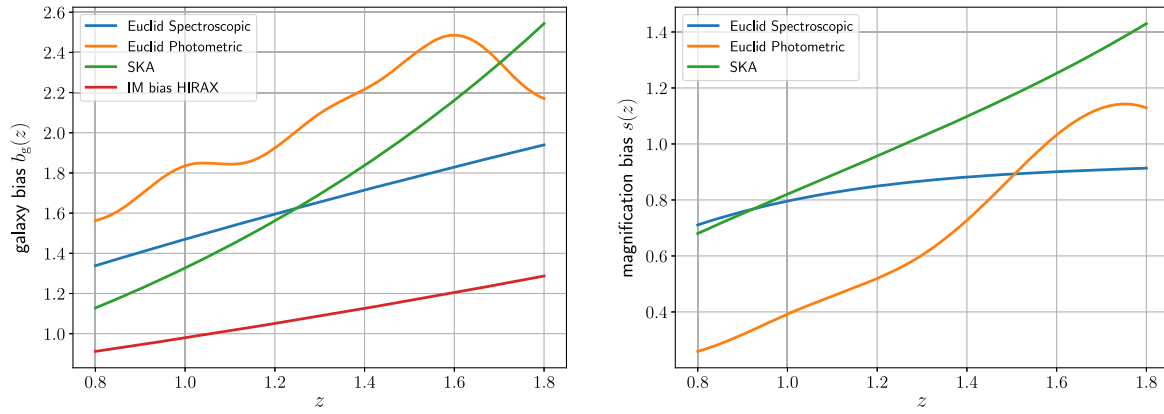


FIG. 6. *Left panel:* Galaxy bias $b_g(z)$ for *Euclid*-like spectroscopic, *Euclid*-like photometric and SKA2-like surveys. We also show intensity mapping bias for HIRAX. *Right panel:* magnification bias for *Euclid*-like spectroscopic, *Euclid*-like photometric and SKA2-like.

APPENDIX C: WINDOW FUNCTIONS

We use the top-hat window function for spectroscopic surveys and a Gaussian window for the photometric survey. The reason is that in spectroscopic surveys, we can measure precisely the redshift, while in photometric surveys, we have larger errors on the redshift. For the top-hat, we choose a smooth version of the top-hat filter,

$$W_{\text{TH}}(z) = \begin{cases} \frac{1}{2} \left[1 + \tanh\left(\frac{z-\mu+\Delta z/2}{\alpha}\right) \right] & \text{if } z \leq \mu \\ \frac{1}{2} \left[1 + \tanh\left(-\frac{(z-\mu-\Delta z/2)}{\alpha}\right) \right] & \text{otherwise,} \end{cases} \quad (\text{C1})$$

where μ is the mean of the redshift bin, Δz is the width of the bin, and α is the speed of the transition which we choose to be $\alpha = \Delta z/10$. The expression for the Gaussian window function is given by

$$W_{\text{G}}(z) = \frac{1}{\Delta z \sqrt{2\pi}} \exp\left[-\frac{(z-\mu)^2}{2\Delta z^2}\right]. \quad (\text{C2})$$

-
- [1] Adam G. Riess *et al.*, Observational evidence from supernovae for an accelerating universe and a cosmological constant, *Astron. J.* **116**, 1009 (1998).
- [2] S. Perlmutter *et al.*, Measurements of Ω and Λ from 42 high redshift supernovae, *Astrophys. J.* **517**, 565 (1999).
- [3] Martin Kunz, The phenomenological approach to modeling the dark energy, *C. R. Phys.* **13**, 539 (2012).
- [4] Luca Amendola, Martin Kunz, Mariele Motta, Ippocratis D. Saltas, and Ignacy Sawicki, Observables and unobservables in dark energy cosmologies, *Phys. Rev. D* **87**, 023501 (2013).
- [5] Ana Marta Pinho, Santiago Casas, and Luca Amendola, Model-independent reconstruction of the linear anisotropic stress η , *J. Cosmol. Astropart. Phys.* **11** (2018) 027.
- [6] Felipe Oliveira Franco, Camille Bonvin, and Chris Clarkson, A null test to probe the scale-dependence of the growth of structure as a test of general relativity, *Mon. Not. R. Astron. Soc.* **492**, L34 (2020).
- [7] Camille Bonvin, Felipe Oliveira Franco, and Pierre Fleury, A null test of the equivalence principle using relativistic effects in galaxy surveys, *J. Cosmol. Astropart. Phys.* **08** (2020) 004.
- [8] Renan Boschetti, L. Raul Abramo, and Luca Amendola, Fisher matrix for multiple tracers: All you can learn from large-scale structure without assuming a model, *J. Cosmol. Astropart. Phys.* **11** (2020) 054.
- [9] Daniel Sobral-Blanco and Camille Bonvin, Measuring anisotropic stress with relativistic effects, *Phys. Rev. D* **104**, 063516 (2021).
- [10] Levon Pogonian, Marco Raveri, Kazuya Koyama, Matteo Martinelli, Alessandra Silvestri, Gong-Bo Zhao, Jian Li,

- Simone Peirone, and Alex Zucca, Imprints of cosmological tensions in reconstructed gravity, *Nat. Astron.* **6**, 1484 (2022).
- [11] Pengjie Zhang, Michele Liguori, Rachel Bean, and Scott Dodelson, Probing Gravity at Cosmological Scales by Measurements which Test the Relationship between Gravitational Lensing and Matter Overdensity, *Phys. Rev. Lett.* **99**, 141302 (2007).
- [12] Levon Pogosian and Alessandra Silvestri, The pattern of growth in viable $f(R)$ cosmologies, *Phys. Rev. D* **77**, 023503 (2008); **81**, 049901(E) (2010).
- [13] Nick Kaiser, Clustering in real space and in redshift space, *Mon. Not. R. Astron. Soc.* **227**, 1 (1987).
- [14] Reinabelle Reyes, Rachel Mandelbaum, Uros Seljak, Tobias Baldauf, James E. Gunn, Lucas Lombriser, and Robert E. Smith, Confirmation of general relativity on large scales from weak lensing and galaxy velocities, *Nature (London)* **464**, 256 (2010).
- [15] Donald G. York *et al.*, The sloan digital sky survey: Technical summary, *Astron. J.* **120**, 1579 (2000).
- [16] A. Amon *et al.*, KiDS + 2dFLenS + GAMA: Testing the cosmological model with the E_G statistic, *Mon. Not. R. Astron. Soc.* **479**, 3422 (2018).
- [17] Konrad Kuijken *et al.*, Gravitational lensing analysis of the kilo degree survey, *Mon. Not. R. Astron. Soc.* **454**, 3500 (2015).
- [18] Chris Blake *et al.*, The 2-degree field lensing survey: Design and clustering measurements, *Mon. Not. R. Astron. Soc.* **462**, 4240 (2016).
- [19] Kyle S. Dawson *et al.*, The Baryon oscillation spectroscopic survey of SDSS-III, *Astron. J.* **145**, 10 (2013).
- [20] S. P. Driver *et al.*, Galaxy and mass assembly (GAMA): Survey diagnostics and core data release, *Mon. Not. R. Astron. Soc.* **413**, 971 (2011).
- [21] Anthony R. Pullen, Shadab Alam, Siyu He, and Shirley Ho, Constraining gravity at the largest scales through CMB lensing and galaxy velocities, *Mon. Not. R. Astron. Soc.* **460**, 4098 (2016).
- [22] Shadab Alam, Hironao Miyatake, Surhud More, Shirley Ho, and Rachel Mandelbaum, Testing gravity on large scales by combining weak lensing with galaxy clustering using CFHTLenS and BOSS CMASS, *Mon. Not. R. Astron. Soc.* **465**, 4853 (2017).
- [23] Chris Blake *et al.*, RCSLenS: Testing gravitational physics through the cross-correlation of weak lensing and large-scale structure, *Mon. Not. R. Astron. Soc.* **456**, 2806 (2016).
- [24] S. de la Torre *et al.*, The VIMOS public extragalactic redshift survey (VIPERS). Galaxy clustering and redshift-space distortions at $z = 0.8$ in the first data release, *Astron. Astrophys.* **557**, A54 (2013).
- [25] Azadeh Moradinezhad Dizgah and Ruth Durrer, Lensing corrections to the $E_g(z)$ statistics from large scale structure, *J. Cosmol. Astropart. Phys.* **09** (2016) 035.
- [26] Basundhara Ghosh and Ruth Durrer, The observable E_g statistics, *J. Cosmol. Astropart. Phys.* **06** (2019) 010.
- [27] Tzu-Ching Chang, Ue-Li Pen, Kevin Bandura, and Jeffrey B. Peterson, Hydrogen 21-cm intensity mapping at redshift 0.8, *Nature (London)* **466**, 463 (2010).
- [28] K. W. Masui *et al.*, Measurement of 21 cm brightness fluctuations at $z \sim 0.8$ in cross-correlation, *Astrophys. J. Lett.* **763**, L20 (2013).
- [29] E. R. Switzer *et al.*, Determination of $z \sim 0.8$ neutral hydrogen fluctuations using the 21 cm intensity mapping auto-correlation, *Mon. Not. R. Astron. Soc.* **434**, L46 (2013).
- [30] C. J. Anderson *et al.*, Low-amplitude clustering in low-redshift 21-cm intensity maps cross-correlated with 2dF galaxy densities, *Mon. Not. R. Astron. Soc.* **476**, 3382 (2018).
- [31] Denis Tramonte and Yin-Zhe Ma, The neutral hydrogen distribution in large-scale haloes from 21-cm intensity maps, *Mon. Not. R. Astron. Soc.* **498**, 5916 (2020).
- [32] Lincheng Li, Lister Staveley-Smith, and Jonghwan Rhee, An HI intensity mapping survey with a phased array feed, *Res. Astron. Astrophys.* **21**, 030 (2021).
- [33] Laura Wolz *et al.*, HI constraints from the cross-correlation of eBOSS galaxies and Green Bank Telescope intensity maps, *Mon. Not. R. Astron. Soc.* **510**, 3495 (2022).
- [34] Mandana Amiri *et al.*, Detection of cosmological 21 cm emission with the Canadian hydrogen intensity mapping experiment, [arXiv:2202.01242](https://arxiv.org/abs/2202.01242).
- [35] Steven Cunnington *et al.*, H I intensity mapping with MeerKAT: Power spectrum detection in cross-correlation with WiggleZ galaxies, *Mon. Not. R. Astron. Soc.* **518**, 6262 (2022).
- [36] Mona Jalilvand, Elisabetta Majerotto, Camille Bonvin, Fabien Lacasa, Martin Kunz, Warren Naidoo, and Kavilan Moodley, New Estimator for Gravitational Lensing Using Galaxy and Intensity Mapping Surveys, *Phys. Rev. Lett.* **124**, 031101 (2020).
- [37] Jaiyul Yoo, A. Liam Fitzpatrick, and Matias Zaldarriaga, A new perspective on galaxy clustering as a cosmological probe: General relativistic effects, *Phys. Rev. D* **80**, 083514 (2009).
- [38] Camille Bonvin and Ruth Durrer, What galaxy surveys really measure, *Phys. Rev. D* **84**, 063505 (2011).
- [39] Anthony Challinor and Antony Lewis, The linear power spectrum of observed source number counts, *Phys. Rev. D* **84**, 043516 (2011).
- [40] Alex Hall, Camille Bonvin, and Anthony Challinor, Testing general relativity with 21-cm intensity mapping, *Phys. Rev. D* **87**, 064026 (2013).
- [41] Anthony R. Pullen, Shadab Alam, Siyu He, and Shirley Ho, Constraining gravity at the largest scales through CMB lensing and galaxy velocities, *Mon. Not. R. Astron. Soc.* **460**, 4098 (2016).
- [42] Alkistis Pourtsidou, Testing gravity at large scales with hI intensity mapping, *Mon. Not. R. Astron. Soc.* **461**, 1457 (2016).
- [43] L. B. Newburgh *et al.*, HIRAX: A probe of dark energy and radio transients, *Proc. SPIE Int. Soc. Opt. Eng.* **9906**, 99065X (2016).
- [44] Devin Crichton *et al.*, Hydrogen intensity and real-time analysis experiment: 256-element array status and overview, *J. Astron. Telesc. Instrum. Syst.* **8**, 011019 (2022).
- [45] Eleonora Villa, Enea Di Dio, and Francesca Lepori, Lensing convergence in galaxy clustering in Λ CDM and beyond, *J. Cosmol. Astropart. Phys.* **04** (2018) 033.

- [46] Philip Bull, Extending cosmological tests of general relativity with the square kilometre array, *Astrophys. J.* **817**, 26 (2016).
- [47] R. Scaramella *et al.*, Euclid preparation—I. The Euclid wide survey, *Astron. Astrophys.* **662**, A112 (2022).
- [48] A. Blanchard *et al.*, Euclid preparation: VII. Forecast validation for Euclid cosmological probes, *Astron. Astrophys.* **642**, A191 (2020).
- [49] F. Lepori *et al.* (Euclid Collaboration), Euclid preparation: XIX. Impact of magnification on photometric galaxy clustering, *Astron. Astrophys.* **662**, A93 (2022).
- [50] Viraj Nistane, Mona Jalilvand, Julien Carron, Ruth Durrer, and Martin Kunz, An estimator for the lensing potential from galaxy number counts, *J. Cosmol. Astropart. Phys.* **06** (2022) 024.
- [51] Diego Blas, Julien Lesgourgues, and Thomas Tram, The cosmic linear anisotropy solving system (CLASS) II: Approximation schemes, *J. Cosmol. Astropart. Phys.* **11** (2011) 034.
- [52] Enea Di Dio, Francesco Montanari, Julien Lesgourgues, and Ruth Durrer, The CLASSgal code for relativistic cosmological large scale structure, *J. Cosmol. Astropart. Phys.* **11** (2013) 044.
- [53] Goran Jelic-Cizmek, Francesca Lepori, Camille Bonvin, and Ruth Durrer, On the importance of lensing for galaxy clustering in photometric and spectroscopic surveys, *J. Cosmol. Astropart. Phys.* **04** (2021) 055.
- [54] Francesco Montanari and Ruth Durrer, Measuring the lensing potential with tomographic galaxy number counts, *J. Cosmol. Astropart. Phys.* **10** (2015) 070.
- [55] Martin Kunz and Domenico Sapone, Dark Energy versus Modified Gravity, *Phys. Rev. Lett.* **98**, 121301 (2007).
- [56] Luca Amendola, Martin Kunz, and Domenico Sapone, Measuring the dark side (with weak lensing), *J. Cosmol. Astropart. Phys.* **04** (2008) 013.
- [57] Levon Pogosian, Alessandra Silvestri, Kazuya Koyama, and Gong-Bo Zhao, How to optimally parametrize deviations from general relativity in the evolution of cosmological perturbations?, *Phys. Rev. D* **81**, 104023 (2010).
- [58] D. Nelson Limber, The analysis of counts of the extragalactic nebulae in terms of a fluctuating density field. II, *Astrophys. J.* **119**, 655 (1954).
- [59] P. A. R. Ade *et al.*, Planck 2015 results. XIV. Dark energy and modified gravity, *Astron. Astrophys.* **594**, A14 (2016).
- [60] C. Brans and R. H. Dicke, Mach's principle and a relativistic theory of gravitation, *Phys. Rev.* **124**, 925 (1961).
- [61] G. R. Dvali, Gregory Gabadadze, and Massimo Porrati, 4-D gravity on a brane in 5-D Minkowski space, *Phys. Lett. B* **485**, 208 (2000).
- [62] E. Babichev, C. Deffayet, and R. Ziour, k-Mouflage gravity, *Int. J. Mod. Phys. D* **18**, 2147 (2009).
- [63] Florian Beutler *et al.*, The clustering of galaxies in the completed SDSS-III Baryon oscillation spectroscopic survey: Anisotropic galaxy clustering in fourier-space, *Mon. Not. R. Astron. Soc.* **466**, 2242 (2017).
- [64] Martin Kunz, The dark degeneracy: On the number and nature of dark components, *Phys. Rev. D* **80**, 123001 (2009).
- [65] J. Richard Shaw, Kris Sigurdson, Ue-Li Pen, Albert Stebbins, and Michael Sitwell, All-sky interferometry with spherical harmonic transit telescopes, *Astrophys. J.* **781**, 57 (2014).
- [66] J. Richard Shaw, Kris Sigurdson, Michael Sitwell, Albert Stebbins, and Ue-Li Pen, Coaxing cosmic 21 cm fluctuations from the polarized sky using m-mode analysis, *Phys. Rev. D* **91**, 083514 (2015).
- [67] Philip Bull, Pedro G. Ferreira, Prina Patel, and Mario G. Santos, Late-time cosmology with 21 cm intensity mapping experiments, *Astrophys. J.* **803**, 21 (2015).
- [68] L. Wolz, C. Blake, and J. S. B. Wyithe, Determining the HI content of galaxies via intensity mapping cross-correlations, *Mon. Not. R. Astron. Soc.* **470**, 3220 (2017).
- [69] Luca Amendola *et al.*, Cosmology and fundamental physics with the Euclid satellite, *Living Rev. Relativity* **21**, 2 (2018).
- [70] Jacobo Asorey, Martin Crocce, and Enrique Gaztanaga, Redshift-space distortions from the cross-correlation of photometric populations, *Mon. Not. R. Astron. Soc.* **445**, 2825 (2014).
- [71] Alkistis Pourtsidou, Testing gravity at large scales with HI intensity mapping, *Mon. Not. R. Astron. Soc.* **461**, 1457 (2016).
- [72] Peter L. Taylor and Katarina Markovič, Covariance of photometric and spectroscopic two-point statistics: Implications for cosmological parameter inference, *Phys. Rev. D* **106**, 063536 (2022).
- [73] Anthony R. Pullen, Shadab Alam, and Shirley Ho, Probing gravity at large scales through CMB lensing, *Mon. Not. R. Astron. Soc.* **449**, 4326 (2015).
- [74] Yucheng Zhang *et al.*, Testing general relativity on cosmological scales at redshift $z \sim 1.5$ with quasar and CMB lensing, *Mon. Not. R. Astron. Soc.* **501**, 1013 (2021).