

A forecast for large-scale structure constraints on Horndeski gravity with CO line intensity mapping

Bryan R Scott¹,^{*} Kirit S. Karkare^{2,3} and Simeon Bird¹

¹*Department of Physics & Astronomy, University of California, Riverside, Riverside, CA 92521, USA*

²*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*

³*Fermi National Accelerator Laboratory, MS209, P.O. Box 500, Batavia, IL 60510, USA*

Accepted 2023 May 1. Received 2023 April 5; in original form 2022 September 28

ABSTRACT

We consider the potential for line intensity mapping (LIM) of the rotational CO(1-0), CO(2-1), and CO(3-2) transitions to detect deviations from General Relativity from $0 < z < 3$ within the framework of a very general class of modified gravity models, called Horndeski's theories. Our forecast assumes a multitracers analysis separately obtaining information from the matter power spectrum and the first two multipoles of the redshift space distortion power spectrum. To achieve ± 0.1 level constraints on the slope of the kinetic gravity braiding and Planck mass evolution parameters, a mm-wave LIM experiment would need to accumulate $\approx 10^8$ – 10^9 spectrometre-hours, feasible with instruments that could be deployed in the 2030s. Such a measurement would constrain the parameters of Horndeski's theory at a level at worst competitive to and at best an order of magnitude tighter than existing constraints from the CMB and LSS. Our modelling code is publicly available.

Key words: gravitation – large-scale structure of Universe – cosmology: observations.

1 INTRODUCTION

The theory of General Relativity (GR) has withstood attempts at revision on theoretical and experimental grounds for more than a century. In light of the non-renormalizability of GR and the need to explain the observed change in the expansion rate of the universe, there is now a rich taxonomy of theories that revise standard GR, including $f(R)$, Horava–Lifshitz, and scalar-tensor theories (for a thorough review, see Clifton et al. 2012). Despite stringent experimental limits on deviations from GR on small scales, measurements of the Hubble constant (Riess et al. 1998; Schmidt et al. 1998; Perlmutter et al. 1999), the Cosmic Microwave Background [CMB; Aghanim et al. (2020)] and Baryonic Acoustic Oscillations [BAOs; Wang (2006)] all point to an accelerated expansion of the universe. Although the most minimal explanation is arguably a cosmological constant, other potential solutions include a new coupling to the matter sector or a modification of the gravity theory itself.

Astrophysical and cosmological tests of GR are also worth pursuing even if they do not seek to explain the expansion of the Universe, as they are able to probe regimes inaccessible to Solar System probes.

Horndeski's theories are the most general theories which include a scalar field and have second-order equations of motion (Horndeski 1974). They are interesting because they include a number of previously studied classes of model as subcases, including Brans–Dicke, $f(R)$, and Galileon models (Bellini & Sawicki 2014). Here,

we interpret the scalar field in Horndeski's gravity as a generic model for dark energy candidates.¹

Modified gravity effects can be observed both through changes to the background expansion and large-scale geometry, or through measurements of large-scale structure (LSS) that probe changes to the Poisson equation and the physics of galaxy formation and evolution [Li & Koyama (2019) and references therein]. BAO and CMB measurements probe geometry, while lensing, redshift space distortions (RSD), and biased tracers of the matter power spectrum probe structure formation. It is typical to combine multiple probes to improve constraining power and perform consistency checks (Troxel et al. 2018), with specific consistency conditions for Horndeski's theories derived in Hojjati, Pogosian & Zhao (2011) and Peirone et al. (2018). Measurements of galaxy cluster abundance and the linear growth rate of perturbations have placed limits on modifications to gravity, especially in the dark energy equation of state (w) – linear growth rate parameter (γ) plane [cf. Mantz et al. (2008); Rapetti, Allen & Mantz (2008)]. Modified gravity effects have also been pursued in the context of relieving cosmological tensions (Pogosian et al. 2022).

As a general framework of modified gravity, Horndeski's theories have a variety of operators which can be constrained by different cosmological experiments. In the context of effective field theory parametrizations of modified gravity, Kreisch & Komatsu (2018) find that modified gravity constraints are primarily driven by a large amplitude modification of the ISW in CMB measurements.

¹However, scalar fields are also a possible candidate for beyond the standard model descriptions of dark matter [for one such model, see Magaña & Matos (2012)].

* E-mail: bscot002@ucr.edu

Mancini et al. (2019) pursue lensing and clustering constraints on Horndeski’s model parameters from KiDS + GAMA. Noller & Nicola (2019) combine measurements of the CMB, RSD and SDSS $P_m(k)$ measurements from the LRG catalog to derive constraints on the Horndeski model parameters, with results similarly driven by the large effect of modified gravity parameters at low ℓ in Planck C_ℓ^{TT} measurements. However, large uncertainties, especially from galaxy biasing and degeneracies in the measurement of $P_m(k)$, limit its constraining power beyond the CMB-only result. In addition, the recent simultaneous detection of gravitational and electromagnetic waves from neutron star mergers constrains the speed of gravity to match the speed of light to high accuracy (Baker et al. 2017; Arai & Nishizawa 2018). Interactions between dark energy and gravitational waves can also be used to constrain a linear combination of Horndeski’s model parameters (Creminelli et al. 2020; Noller 2020). Future theoretical work as well as next-generation surveys [e.g. Vera C. Rubin Observatory/Legacy Survey of Space and Time (Ivezić et al. 2019), Nancy Grace Roman Space Telescope (Spergel et al. 2015), and Euclid (Laureijs et al. 2011)] are likely to improve measurements of the matter power spectrum and will improve constraints subject to uncertainties in the biasing of LSS tracers.

In this paper, we will make forecasts for an alternative avenue for constraining modified gravity based on line intensity mapping (LIM). LIM is a promising technique for constraining both astrophysics and cosmology in large cosmological volumes (Breysse, Kovetz & Kamionkowski 2014; Pullen, Doré & Bock 2014; Kovetz et al. 2017; Creque-Sarbinowski & Kamionkowski 2018; Karkare & Bird 2018; Moradinezhad Dizgah, Keating & Fialkov 2019; Gong, Chen & Cooray 2020). LIM uses moderate resolution observations to detect LSS in aggregate by integrating over discrete sources (Kovetz et al. 2017). By using a spectrometer to target a spectral line with known rest-frame wavelength, the redshift of the source emission is known and 3D maps of cosmic structure can be constructed. LIM is particularly advantageous for measuring a wide redshift range, including high redshifts where individual galaxies become too faint to be detected in a traditional galaxy survey. This makes LIM an attractive way to probe the evolution of structure during both the epochs of matter and dark energy domination and thus provide powerful constraints on modified gravity scenarios.

Numerous atomic and molecular lines are now being targeted by LIM experiments across a wide range of wavelengths. High-redshift measurements of the 21 cm neutral hydrogen spin-flip transition in the intergalactic medium are expected to probe the timing and sources of reionization (e.g. HERA DeBoer et al. 2017), while similar measurements of HI within galaxies at lower redshifts will constrain dark energy (Cosmic Visions 21 cm Collaboration et al. 2018; Wu & Zhang 2022). At shorter wavelengths, far-infrared (IR) lines such as the CO $J \rightarrow J - 1$ rotational transitions and the ionized-carbon [CII] fine structure line trace star formation and cold gas, are known to be bright in early galaxies, and are being targeted at redshifts $0 < z < 10$ by instruments observing in the cm-THz range (Righi, Hernández-Monteagudo & Sunyaev 2008; Breysse et al. 2014; Silva et al. 2015; Yue et al. 2015; Fonseca et al. 2016; Li et al. 2016; Padmanabhan 2017). Finally, lines such as H α and Ly α are targets for optical LIM experiments; for example, SPHEREx targets galactic astrophysics and reionization (Gong et al. 2011, 2017; Silva et al. 2013, 2017). Current and near-future far-IR LIM experiments aim to constrain the star formation rate and galaxy formation history over a wide range of redshifts. Those targeting CO include COPSS (Keating et al. 2016) and mmIME (Keating et al. 2020; Breysse et al. 2022), which have both reported first detections of CO shot noise power, COMAP (Li et al. 2016), EXCLAIM (Padmanabhan et al. 2022),

YTLA (Ho et al. 2009) and SPT-SLIM (Karkare et al. 2022). TIME (Crites et al. 2014) and CONCERTO target [CII] (Ade et al. 2020). Recently, the COMAP collaboration has placed upper limits on the CO(1-0) signal at clustering scales (Ihle et al. 2022). While the first generation of mm-wave LIM experiments will primarily constrain the astrophysics of the emission line, future generations have the potential to deliver competitive constraints on cosmology, including early/dynamical dark energy and neutrino masses (Karkare & Bird 2018; Moradinezhad Dizgah et al. 2022). In particular, compact mm-wave spectrometers are now being demonstrated that could enable future surveys with orders of magnitude more sensitivity than current experiments (Karkare et al. 2020).

In this paper, we investigate the constraining power of future ground-based wide-bandwidth mm-wave LIM experiments targeting multiple rotational CO transitions over the redshift range $z \approx 0-3$. In addition to large accessible cosmological volumes, this extends constraints on modified gravity to higher redshifts than are available in current large optical surveys. In Section 2, we review Horndeski’s gravity and the application to LSS through the matter power spectrum and redshift space distortions. Then, in Section 3, we introduce the formalism of LSS measurements with LIM. In Section 4, we investigate the range of accessible scales and required survey integration times to achieve competitive constraints on the linear theory parameters, accounting for the atmosphere, astrophysical continuum, and interloper lines. In Section 5, we discuss implications of these results. We conclude in Section 6.

Although modifications to GR generally imply a different expansion history, we assume that all deviations are small and only affect linear structure formation around a Λ CDM background. As such, where necessary, we assume a flat Λ CDM-like cosmology with $h = 0.678$, $\Omega_b h^2 = 0.0224$, $\Omega_c h^2 = 0.12$, and $\Omega_\Lambda = 1 - \Sigma \Omega_i$.

2 HORNDESKI’S GRAVITY

Horndeski’s theories construct a relativistic theory of gravity from a Lagrangian, including a metric tensor and a scalar field, and lead to second-order equations of motion (Horndeski 1974). In this section, we review the features of Horndeski’s theory relevant to this work and refer the interested reader to Bellini & Sawicki (2014), which develops the formalism employed here, and its application to the Einstein–Boltzmann solver CLASS (Blas, Lesgourgues & Tram 2011) to produce the extended Horndeski in Linear Cosmic Anisotropy Solving System (HI_CLASS; Zumalacárregui et al. 2017; Bellini, Sawicki & Zumalacárregui 2020).

In the linear regime, solving the perturbed Einstein equation allows for the construction of four functions of time, denoted $\alpha_i(t)$, that translate the functional degrees of freedom in the action into four time-dependent parametric degrees of freedom (Bellini & Sawicki 2014). The Horndeski action, the background relations, and prescriptions for the $\alpha_i(t)$ fully determine the evolution of perturbations in the linear regime and hence LSS. There are four functions, two of which are in principle measurable by LIM (α_B and α_M) and two of which are not (α_K and α_T). They have the following physical interpretations:

(i) α_B encodes mixing between the scalar and metric perturbations that arises from the clustering of the Horndeski scalar field and appears as perturbations to T_{0i} . $\alpha_B = 0$ in Λ CDM + GR. We treat α_B as a free parameter to be constrained by the LIM experiment.

(ii) α_M rescales the Planck mass, representing a change in the strength of gravity. While a constant rescaling of the strength of gravity does not affect structure formation, its time evolution generates anisotropic stress. Since α_M parametrizes the evolution of

the Planck mass with time, $\alpha_M = 0$ in Λ CDM + GR. We treat α_M as a free parameter to be constrained by the LIM experiment.

(iii) α_K represents perturbations to the energy-momentum tensor $T_{\mu\nu}$ arising directly from the action. These can be thought of as perturbations in an additional fluid connected with the modification to gravity. However, α_K affects only scales close to the cosmological horizon, far larger than those measured by LIM or other LSS probes [for a discussion of this to second order, see Bellini & Sawicki (2014)]. While $\alpha_K = 0$ represents the value in Λ CDM + GR and is therefore a natural choice, we choose $\alpha_K = 1$ to ensure that our models easily satisfy the condition for avoiding ghosts in the scalar mode: $\alpha_K + 3/2\alpha_B^2 > 0$.

(iv) α_T gives the tensor speed excess, potentially inducing anisotropic stress, even in the absence of scalar perturbations. $\alpha_T = 0$ in Λ CDM + GR. We set $\alpha_T = 0$, as it is well constrained by measurements of the speed of gravitational waves (Baker et al. 2017; Arai & Nishizawa 2018).

These expressions are implemented in the Einstein–Boltzmann solver Horndeski in Cosmic Linear Anisotropy Solver (HI_CLASS), which we use to predict the matter power spectrum $P_m(k)$ under an assumed Λ CDM background, and to vary the free functions according to the parametrizations described in the next section.

2.1 Parametrizations

In the linear regime of cosmological perturbation theory, we assume that all perturbations are small and taken around a flat background space time,

$$ds^2 = -(1 + 2\Psi)dt^2 + (1 - 2\Phi)\delta_{ij}dx^i dx^j, \quad (1)$$

where Ψ and Φ are small metric perturbations. In the case of fluid scalar perturbations and general theories of gravity, LSS observations such as the galaxy power spectrum, weak lensing shear field, or RSD probe a small number of combinations of these potentials. In Horndeski’s theory, the potentials are also complicated functions of the α_i , arbitrary functions that represent the maximal amount of information available from cosmology to constrain the dynamics of this class of models. The evolution of the flat background itself can be determined from the Friedmann equations.

The functional freedom to pick the Horndeski α_i allows any evolution for the background space time to be realized. LSS alone cannot pick out either the expansion history or a unique form for the α_i . To reduce this freedom, we begin by first noting that geometric measurements are consistent with the universe being nearly Λ CDM, which we select as our model for the background evolution. Once a background is chosen, it is necessary to define a functional form to parametrize how modifications to gravity evolve with time.

Parametrization I: A natural choice in a nearly Λ CDM universe is to parametrize the modified gravity effect as proportional to the cosmological constant density, Ω_Λ . As this term grows with redshift, it ‘turns on’ modified gravity effects at late times and during the epoch of dark energy domination. Thus, for our first parametrization, we assume that α_B and α_M are linear functions of Ω_Λ :

$$\alpha_{B,M} = c_{B,M}\Omega_\Lambda. \quad (2)$$

Here, we have adopted notation from Noller & Nicola (2019) and Kreisch & Komatsu (2018) and refer to this as parametrization I.

Parametrization II: To evaluate the sensitivity of our probes to observations at high redshift, we use an alternate parametrization where the effect of gravity modification is linearly proportional to the

scale factor (Zumalacárregui et al. 2017). This allows the modified gravity to become important at early times, before the onset of dark energy domination. We thus have

$$\alpha_{B,M} = c_{B,M}a. \quad (3)$$

In motivating our choice of parametrizations, we have implicitly assumed that the Horndeski Lagrangian is interpreted as a Dark Energy like scalar field. Other interpretations of the scalar field can motivate alternative parametrizations, including additional powers of the scale factor that are considered in Kreisch & Komatsu (2018) and Noller & Nicola (2019).

Fig. 1 shows the relative deviation of the matter power spectrum at a fixed scale ($k = 0.05 \text{ h Mpc}^{-1}$) as the Planck mass rescaling c_M and braiding c_B parameters are allowed to vary. Large deviations from Λ CDM are possible for extreme values of the α functions. As noted in Noller & Nicola (2019), curves that intersect the Λ CDM prediction exhibit a degeneracy between c_M and c_B for the matter power spectrum with $c_B \approx 1.8c_M$ in parametrization II ($\alpha_i \propto a$). By design, the effect of modifying gravity is largest for the late-time universe, near the end of matter domination. One implication of this evolution is that achieving robust constraints on these linear theory parameters and simultaneously constraining deviations from GR in both parametrizations requires an experiment that targets a large range in redshift.

By selecting a fiducial k -scale to summarize the effects of varying α_M and α_B , we have ignored the k -dependence introduced by the modification to gravity. A well-known generic feature of these models is a turnover in the power spectrum, where an excess on large-scales becomes a deficit on small scales (with respect to Λ CDM). However, this turnover occurs on scales near the cosmological horizon and is thus extremely difficult to measure with LSS measurements.²

At the intermediate scales measured by a LIM experiment, the characteristic feature of modified gravity models relative to Λ CDM + GR is a uniform excess in the power spectrum. The size and behaviour with varying c_M, c_B of the effect depends strongly on the choice of parametrization, with a < 1 per cent difference in Λ CDM at $z = 3$ in parametrization I and a few per cent difference at $z = 0.5$, even for extreme values of the c_M, c_B . The effect of modified gravity on the power spectrum is larger in parametrization II, approaching instability in the theory when c_M is small and c_B is large. We therefore expect greater sensitivity to the c_M, c_B in parametrization II than in parametrization I.

To summarize, we have two modified gravity functions that scale with the background evolution of the space time that we seek to constrain. The c_B and c_M parameters govern respectively the evolution of the braiding α_B (clustering of dark energy) and the Planck mass run rate α_M (the large-scale strength of gravity). Λ CDM differs from the models we consider in that the large-scale strength of gravity is fixed and dark energy does not cluster. We assume two functional forms for the scaling of these parameters with the background evolution: one parametrization that scales with the effective dark energy component density Ω_{DE} and one that scales with the scale factor. **We fix the remaining two functions (α_K, α_T) to be constant and assign them values of unity and zero, respectively.** We specifically forecast for the uncertainties $\sigma(c_B), \sigma(c_M)$.

²The turnover is an unambiguous signature of modified gravity and would constrain the braiding scale, a function of α_M and α_B .

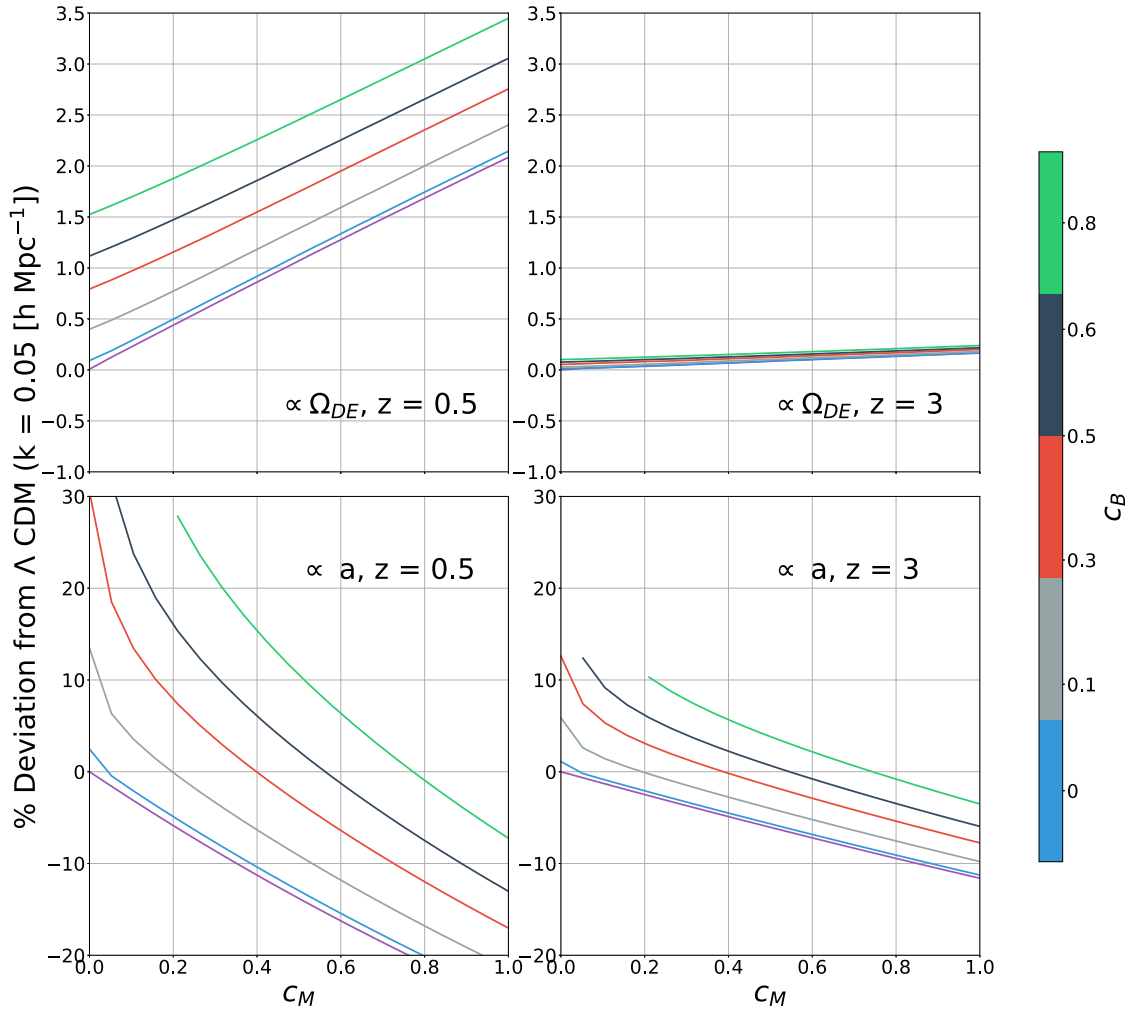


Figure 1. Relative deviation of the matter power spectrum for fixed $k = 0.05 \text{ h Mpc}^{-1}$ at $z = 0.5$ (Left) and $z = 3$ (Right) as a function of c_M , with curves labelled by their value of c_B . Top row shows parametrization I, and bottom row shows parametrization II. The c_B and c_M parameters are allowed to vary over the range 0–1. In parametrization II ($\alpha_i \propto a$), we have truncated the results due to gradient instabilities when c_M is small and c_B is large.

3 LINE INTENSITY MAPPING

In this section, we discuss the details of the LIM observables used in our projections. We then discuss experimental effects that limit the scales accessible in the power spectrum, including the effects of interloper lines and Galactic foregrounds. Possible atmospheric effects are discussed in Appendix B.

3.1 Line power spectrum

Emission lines targeted by LIM experiments originate in galaxies that are biased with respect to the underlying matter overdensity field. On the intermediate and large scales we consider here, outside the non-linear regime, we can parametrize clustering with a scale-independent clustering bias $b(z)$ that varies with redshift. Since the target lines we consider are correlated with galaxy properties (e.g. star formation rate and metallicity) that evolve with redshift, the line intensity $I(z)$ is also redshift-dependent. The LIM clustering power spectrum is

$$P_{\text{clust}}(k, z) = b^2(z) I^2(z) P_m(k, z). \quad (4)$$

Here, $P_m(k, z)$ is the underlying matter power spectrum that contains the cosmological information (Section 2.1). We show

the matter power spectrum for a range of choices of the c_M and c_B in Fig. 2, and the product of the matter power spectrum normalization and linear growth function, $f(z)\sigma_8(z)$, in Fig. 3. We assume the bias is scale-independent and varies linearly with redshift, $b = (1 + z)$, which is a reasonably close approximation to the Sheth–Tormen-inspired (Sheth, Mo & Tormen 2001) bias evolution from Moradinezhad Dizgah et al. (2022).

We assume that the line evolution is given by the line models in Delabrouille et al. (2019). Line intensities are estimated from the specific luminosity density $\rho(z)$. Redshift-dependent luminosity densities are a function of the halo mass ($\text{dn/dM}(M, z)$) or line luminosity functions ($\text{dn/dL}(z)$), which are obtained through empirical scaling relations through the dependence of $L(M)$ with $\text{SFR}(M, z)$ or $\text{SFRD}(M, z)$. In the Delabrouille et al. (2019) models, the line luminosities are derived from the Eagle simulation and are uncertain from a factor of a few to 10. For a recent review of line intensity modelling, see Bernal & Kovetz (2022). The observable effect of the modification to gravity is a constant excess or deficit in $P_m(k)$. The degeneracy between the line intensity, $I(z)$, and the matter power spectrum introduces a degeneracy between the modified gravity effect and the evolution of the bias and line intensity. However, $f(z)\sigma_8(z)$, which

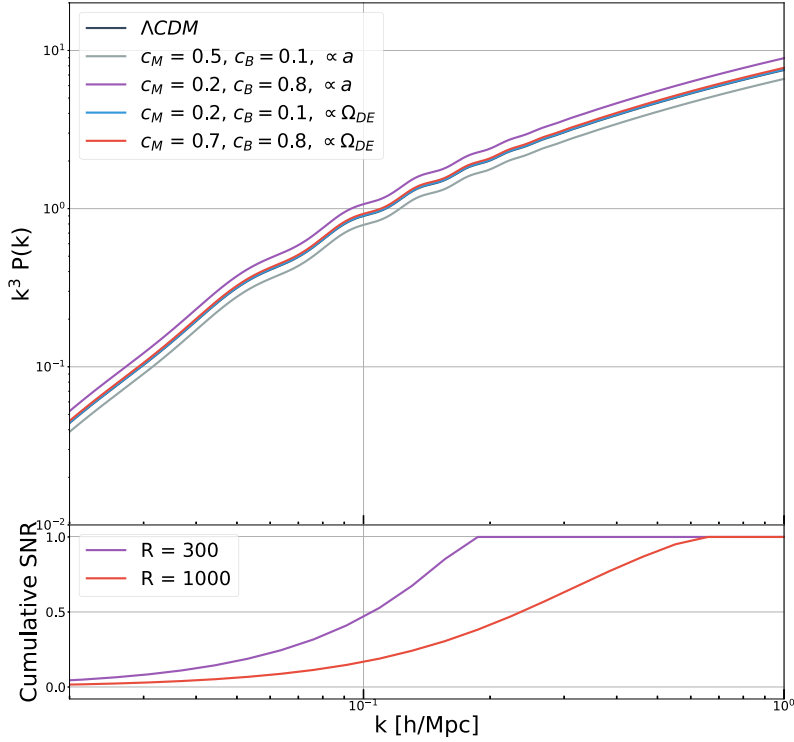


Figure 2. Top panel shows the matter power spectrum at $z = 0.5$ in both parametrizations. We have chosen values of c_B and c_M representative of the range of deviations in $P_m(k)$ that we constrain. Bottom panel shows the cumulative constraining power as a function of scale assuming a spectral resolution of $R = 300$ or 1000 in the baseline $f_{\text{sky}} = 40$ per cent case. The Signal to Noise Ratio (SNR) saturates once the scales probed are below the spectral resolution of the LIM experiment. The lower spectral resolution with $R = 300$ causes the SNR to saturate at a smaller k than in the $R = 1000$ case.

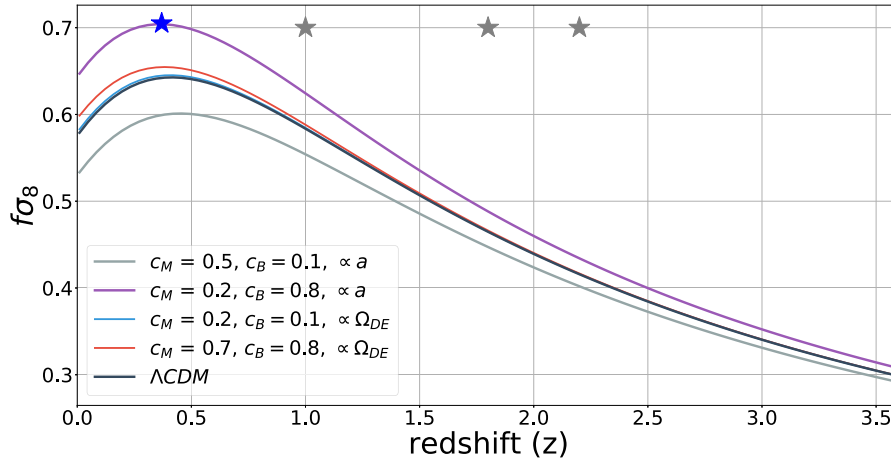


Figure 3. Evolution of the power spectrum normalization $f(z)\sigma_8(z)$ over the range of redshifts accessible to the experiment we forecast for. We have also indicated the approximate redshift where $f\sigma_8$ (blue star, $z \approx 0.3$) and the CO(1-0) (grey star, $z \approx 1$), CO(2-1) (grey star, $z \approx 1.8$), and CO(3-2) (grey star, $z \approx 2.2$) lines that we target in our forecast are brightest within the fiducial experiment bandwidth. The evolution of the line brightness differs from that of the power spectrum normalization, both in the peak and evolution with redshift. This allows an experiment that measures LSS at a range of redshifts to disentangle the evolution of a modified gravity effect on $P_m(k)$ from that of $I(z)$.

controls the power excess on the scales we observe, has a distinctly different form of redshift evolution from the astrophysics-dependent term $I(z)$. Observing a continuing increase relative to the Λ CDM expectation as the line intensity decreases (or vice versa) as a function of redshift can therefore break this degeneracy,

allowing us to potentially recover a signal from modified gravity effects.

Further, several approaches to breaking the line intensity-bias-power spectrum degeneracy have been pursued in the literature. For example, taking ratios of the $\mu = 0$ and $\mu = 2$ multipole moments

Table 1. Line frequencies, target redshifts, P_{shot} estimates, and line temperatures used in this forecast. Unlisted rotational transitions up to CO(9-8) are assumed to contribute interloper power but are not included as targets as they are an order of magnitude smaller in line brightness temperature.

Line	ν_{rest} (GHz)	ν_{obs} (GHz)	z	P_{shot} (μK^2)	Temp (μK)
(1-0)	115	95	0.21	2.66	0.14
CO(2-1)	230	95	1.42	8.54	0.75
CO(3-2)	345	95	2.63	2.98	0.60
CO(2-1)	230	150	0.53	81.8	0.24
CO(3-2)	345	150	1.3	100	0.46
CO(3-2)	345	245	0.4	295	0.14

of auto- and cross-power spectra can yield an estimate of the line bias and be used to infer the line intensity. In addition to higher order moments of the power spectrum, Breyse (2022) has shown that, once clustering information in the Voxel Intensity Distribution is accounted for, a VID joint analysis can break the degeneracy in measurements of $P(k)$ alone. As this relies on the use of clustering information in both $P(k)$ and the VID, it is important to account for the covariance between them (Sato-Polito & Bernal 2022). We leave a full accounting of a combined VID-power spectrum search for modified gravity effects to future work and instead consider an analysis of the 2-pt statistics of a line intensity map.

An LIM experiment measures the clustering power, shot noise due to the discrete nature of the emitting galaxies, and instrumental noise:

$$P_{\text{obs}}(k, z) = P_{\text{clust}}(k, z) + P_{\text{shot}}(z) + P_{\text{N}}. \quad (5)$$

The uncertainty in the power spectrum measurement from an LIM experiment depends both on the number of observed modes and on the instrumental noise. We write the number of statistically independent Fourier modes at a scale k , in bins of width Δk , in a total volume V_s as

$$N_m(k) = \frac{k^2 \Delta k V_s}{4\pi^2} \quad (6)$$

and the variance $\sigma(k)$ on a measurement of $P(k)$ at a scale k is

$$\sigma^2(k, z) = \frac{P_{\text{obs}}^2(k, z)}{N_m(k)}. \quad (7)$$

It is common to write equation (7) with an additional factor of $1/2$ that arises from considering only the upper half of the Fourier plane for a real-valued field. We follow Li et al. (2016) and include this factor of $1/2$ in our definition of $N_m(k)$ in equation (6). Estimates for P_{shot} are given in Table 1, while estimates of P_{N} are discussed in Section 3.3.

3.2 Redshift space distortions

Observations of LSS are not made in the isotropic comoving space in which the matter power spectrum is defined, but in the $2 + 1$ dimensional space of angles and redshift. Since the redshift of an emitter has components due to both the Hubble flow and its peculiar velocity, the power spectrum in redshift space is distorted relative to comoving space (Hamilton 1998).

Because the inferred transverse and line-of-sight coordinates are affected differently by the RSD, it is necessary to consider the full anisotropic power spectrum in the space of parallel (k_{\parallel}) and perpendicular (k_{\perp}) modes. The RSD power spectrum can be expressed in (k, μ) coordinates, where the cosine of the angle is denoted $\mu = \hat{z} \cdot \hat{k}$. Then, in the linear plane-parallel approximation,

the anisotropic matter power spectrum is

$$P_{\text{obs}}(k, \mu, z) = [b(z)^2 I(z)^2 + f(z)^2 I(z)^2 \mu^2] P_m(k). \quad (8)$$

Here, $f(z)$ is the linear growth rate of structure, which is sensitive to modifications of gravity. We show the redshift evolution of $f\sigma_8$ for the range of Horndeski's theories we consider in Fig. 3.

Noller & Nicola (2019) consider constraints on the Horndeski's theory parameters c_B and c_M from both anisotropic clustering measurements of the growth factor D in SDSS and $f\sigma_8$ at $z = 0.57$ and in the 6dF survey at $z = 0.067$. While the power spectrum adds little constraining power directly, the RSD constraint improves the posterior uncertainties, especially on the c_M parameter, when compared to the CMB-only constraint.

Although one can infer the value of $f\sigma_8$ directly from the full shape of the anisotropic matter power spectrum in (k, μ, z) space, it is simpler to consider constraints from the non-vanishing $\ell = 0, 2$, and 4 moments obtained by convolving equation (8) with the Legendre polynomials \mathcal{L}_l :

$$P_l(k) = \frac{2l+1}{2} \int_{-1}^1 \mathcal{L}_l(\mu) P(k, \mu) d\mu \quad (9)$$

Explicit expressions for the monopole and quadrupole are

$$\begin{aligned} P_0(k) &= \left(1 + \frac{2}{3}\beta(bI)^2 + \frac{1}{5}(bI)^2\beta^2\right) P_m(k) \\ P_2(k) &= \left(\frac{4}{3}(bI)^2\beta + \frac{4}{7}(bI)^2\beta^2\right) P_m(k). \end{aligned} \quad (10)$$

Here, we neglect the $l = 4$ moment since the hexadecapole is both difficult to measure and contains little information not present in the first two multipoles (Chung et al. 2019). For consistency with the literature, we also work with $\beta = f/b$ rather than $f\sigma_8$ directly. These expressions allow us to compute the moments of the redshift space distortions from the isotropic power spectrum $P_m(k)$. The variance between the multipole moments can be computed explicitly:

$$\text{Cov}_{l,l'}(k) = \frac{(2l+1)(2l'+1)}{N_m} \int_{-1}^1 \mathcal{L}_l(\mu) \mathcal{L}_{l'}(\mu) (P_{\text{obs}}(k, \mu))^2 d\mu. \quad (11)$$

Taruya, Nishimichi & Saito (2010) (eq. C2-C4) gives explicit expressions for $\text{Cov}_{l,l'}$ (where we have here combined their shot noise term with our P_{N} notation). For the monopole:

$$\begin{aligned} \text{Cov}_{0,0}(k) &= \frac{2}{N_k} \left[\left(1 + \frac{4}{3}\beta + \frac{6}{5}\beta^2 + \frac{4}{7}\beta^3 + \frac{1}{9}\beta^4\right) \right. \\ &\quad \times (bI)^2 P_m(k)^2 + 2P_{\text{N}} \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right) \\ &\quad \left. \times (bI)^2 P_m(k) + P_{\text{N}}^2 \right]. \end{aligned} \quad (12)$$

For the monopole-quadrupole cross-term:

$$\begin{aligned} \text{Cov}_{0,2}(k) &= \frac{2}{N_k} \left[\left(\frac{8}{3}\beta + \frac{24}{7}\beta^2 + \frac{40}{21}\beta^3 + \frac{40}{99}\beta^4\right) \right. \\ &\quad \left. \times (bI)^2 P_m(k)^2 + 2P_{\text{N}} \left(\frac{4}{3}\beta + \frac{4}{7}\beta^2\right) (bI)^2 P_m(k) \right]. \end{aligned} \quad (13)$$

Finally, for the quadrupole:

$$\begin{aligned} \text{Cov}_{2,2}(k) = & \frac{2}{N_k} \left[\left(5 + \frac{220}{21}\beta + \frac{90}{7}\beta^2 + \frac{1700}{231}\beta^3 + \frac{2075}{1287}\beta^4 \right) \right. \\ & \times (bI)^2 P_m(k)^2 \\ & \left. + 2P_N \left(5 + \frac{220}{21}\beta + \frac{30}{7}\beta^2 \right) (bI)^2 P_m(k) + 5P_N^2 \right]. \end{aligned} \quad (14)$$

The above expressions are exact in the case of flat P_N , and are approximately correct on intermediate and large scales where the finite spatial and spectral resolution induce only small attenuation in the signal.

3.3 Target lines, redshifts, and noise estimates

Our forecasts focus on measuring the LIM power spectrum from $0 < z < 3$, where ground-based CO experiments are most sensitive. In addition to the rotational CO transitions, the OI, OIII, and NII, especially CII lines, are promising targets for future LIM experiments. Of these, only the CII line is comparable in expected line temperature to the low $J_{up} = 3, 2, 1$ rotational CO transitions at the redshifts where we expect the signature of modified gravity to be potentially detectable. CII at $z \approx 7$ may be detectable at the edge of our fiducial survey bandwidth with comparable brightness to the CO(4-3) transition, roughly an order of magnitude dimmer than the transitions we forecast for. For an experiment targeting CII, the expected line temperature lies between the CO(4-3) and CO(3-2) transitions, and owing to both the line temperature and small amplitude of the power spectrum excess relative to GR, should yield no better sensitivity to the dimmest CO transition we consider in this work.

In parametrization I ($\alpha_i \propto \Omega_\Lambda$), the excess near the turnover in the matter power spectrum at $k = 0.05 \text{ h Mpc}^{-1}$ is ~ 1 per cent at $z = 3$ and an order of magnitude larger at $z = 0.5$. The evolution in the effect of modified gravity for parametrization II ($\alpha \propto a$) is comparable in magnitude but begins at earlier redshifts, as shown in Fig. 1.

We consider an experiment measuring the CO $J \rightarrow J - 1$ rotational transitions, which emit at rest-frame frequencies of 115J GHz. CO offers several advantages compared to other LIM targets: it is a known tracer of molecular gas and is therefore indicative of star formation (which peaked at $z \sim 2$), it has been detected in individual galaxies at high redshift using ground-based telescopes observing in the millimetre band, and the multiple transitions allow a wide range of redshifts to be detected in a modest instrumental bandwidth. Our forecasts use the CO line amplitudes from Delabrouille et al. (2019).

To detect the CO power spectrum, we consider ground-based mm-wave LIM surveys observing roughly from 75 to 310 GHz. Technology for this frequency range has seen significant recent development for large-format CMB arrays: focal planes featuring dense arrays of background-limited detectors are now common (BICEP2 Collaboration 2015), and current instruments have demonstrated wide-band optics that can measure the 1–3 mm band in a single receiver (Nadolski et al. 2020). Current-generation mm-wave spectrometers are significantly larger than their broadband counterparts since they generally use a physically large apparatus (e.g. grating, Fourier Transform, or Fabry–Perot) for spectral separation. However, *on-chip* spectrometer technology is rapidly progressing (Shirokoff et al. 2012) and instruments are now being planned to demonstrate LIM with dense spectrometre arrays that approach CMB packing efficiency. Our forecasts anticipate that this technology can be scaled

over the next 10 yr in the same manner as CMB instruments leading up to CMB-S4 (Abazajian et al. 2016).

The oxygen line at 118 GHz and the water line at 183 GHz naturally divide up the 75–310 GHz mm-wave band into three windows: 75–115, 120–175, and 190–310 GHz. We discuss our approach to estimating noise power in Appendix A. To account for the frequency dependence of the line temperature, we calculate an effective redshift and line strength for each target by averaging over the window. Target line frequencies, temperatures, and redshifts are given in Table 1. Additional contributions to the noise model are discussed in the following section.

3.4 Finite resolution and foregrounds

3.4.1 Instrument resolution

The scales accessible to a LIM experiment are determined by the finite spatial and frequency resolutions. In the frequency direction, the smoothing scale is characterized by the spectrometre resolution $\delta\nu$, while in the transverse direction, the smoothing is a function of the beamwidth θ_b . These correspond to comoving smoothing scales at redshift z in the transverse σ_\perp and parallel (to the line of sight) direction σ_\parallel . In the perpendicular (spatial) direction, the smoothing scale is

$$\sigma_\perp = \frac{\theta_b R(z)}{\sqrt{8 \ln 2}}, \quad (15)$$

where θ_b is the full width at half-maximum of the beam. The smoothing scale in the parallel (frequency) direction is a function of the target frequency, resolution, and the Hubble scale $H(z)$,

$$\sigma_\parallel = \frac{c \delta\nu (1+z)}{H(z) v_{\text{obs}}}. \quad (16)$$

The noise power spectrum $P_N(k)$ is the product of the white-noise level P_N and a factor accounting for the finite spectral and spatial resolutions of the instrument,

$$P_N(k) = P_N e^{k^2 \sigma_\perp^2} \int_0^1 d\mu e^{\mu^2 k^2 (\sigma_\parallel^2 - \sigma_\perp^2)} \quad (17)$$

where μ is the cosine of the angle between the wavevector k and the line-of-sight direction, and the integral averages over all such angles μ to yield the spherically averaged 3D power spectrum. Here, we treat the signal $P(k)$ as fixed, while the finite resolution of the survey causes the noise to become inflated at small scales. This differs from the physical situation, in which instruments generally have flat noise properties as a function of k (or its angular counterpart, ℓ), above a scale ℓ_{knee} . In fact, it is the inherent signal that is attenuated and not the noise.

Each point in the 2D $k_\parallel - k_\perp$ Fourier plane (averaged over the angular directions) has some attenuation factor due to the instrument resolution. The RSD introduce some phase dependence into the signal as structures move in the redshift direction only, that are picked out by the RSD operator. However, the attenuated 2D noise does not change due to the velocity-induced distortion of the signal. On large and intermediate scales, and for small values of P_N the attenuation factor contributes negligibly to the noise on measurements of the multipole moments. This allows us to use equations (12)–(14) even in the case of finite instrument resolution.

3.4.2 Interloper lines

For observations at fixed redshift and target line frequency, line confusion arises because the emission from sources at various

redshifts overlaps in observed frequency. Without additional information, observed power at a given target redshift cannot be easily distinguished from power at a different redshift that has the same observed frequency. This effect can be large. For example, for [CII] experiments targeting $z \sim 7$, CO rotational transitions between $z = 0.45$ and $z = 1.8$ act as foregrounds with power larger than the target line. For the low- J transitions of CO that we target, higher rotational transitions are the main source of interloper confusion. To model this scale-dependent effect, we modify the numerator of equation (7) to

$$\sum_i b_i(z) I_i(z) P_m(k, z_i) + P_{\text{shot}} + P_N. \quad (18)$$

For the RSD multipoles, we similarly modify equation (11) to sum over the RSD power spectrum at each redshift,

$$\text{Cov}_{l,l'}(k) = \frac{(2l+1)(2l'+1)}{N_m} \times \int_{-1}^1 \mathcal{L}_l(\mu) \mathcal{L}_{l'}(\mu) (\sum_i P_m(k, \mu, z_i) + P_N)^2 d\mu. \quad (19)$$

In other words, we assume that the interloper power adds to the noise and does not contribute signal to the estimate of $P_m(k, z_i)$. In fact, the interloper contributions themselves carry cosmological information similar to the information that we will consider from the brightest lines. For example, in a wide-band experiment observing a large range in redshift and different rotational CO transitions simultaneously, internal cross-correlations may be able to extract the underlying matter power spectrum from each CO line and add them coherently to the signal.

A wide variety of techniques has been proposed for interloper deconfusion. For surveys targeting higher redshifts, masking techniques – i.e. removing brighter pixels that are more likely to come from lower redshifts Breyse, Kovetz & Kamionkowski (2015); Yue et al. (2015), or using an external catalog of bright interloping galaxies Sun et al. (2018) can significantly reduce the interloper contribution. Cross-correlations between lines can also reconstruct a high percentage of the true underlying map (Chung et al. 2019; Cheng, Chang & Bock 2020). Finally, geometric tests for interloper deconfusion were introduced in Silva et al. (2015) and Lidz & Taylor (2016).

3.4.3 Galactic continuum foregrounds

Galactic continuum emission can be a significant foreground for both CMB and LIM experiments. For CO, thermal dust can significantly eclipse the line brightness temperatures at frequencies above 50 GHz while non-thermal synchrotron emission dominates at lower frequencies. By fitting a smooth, low-order polynomial to a foreground that slowly varies in frequency, this can be subtracted and removed, leaving only the underlying cosmological signal. However, residuals from fitting these broadband terms can lead to spuriously inferred excess matter power at large scales that is a function of the residuals after continuum subtraction (McQuinn et al. 2006).

We used NBODYKIT to combine the linear matter power spectrum with Galactic dust-continuum, and studied recovery of the power spectrum. In order to study the scale dependent behaviour, we initially assume a degree of map level cleaning has left a continuum component comparable in magnitude to the cosmology sourced intensity fluctuations. We then generate mock sightlines consisting of a residual component from the fit to the combined cosmological and continuum intensity fluctuations. We do this by first adding a mock LIM signal to a typical Galactic dust spectrum and then removing a series of low-order polynomials, which are then projected onto a random field with weights determined from the combination of the

fit residuals and the underlying cosmological signal. We find that the continuum residuals after fitting have a scale-dependent effect with the largest impact on large scales ($\lesssim 10^{-2}$ h Mpc $^{-1}$). These scales contribute little weight to the overall constraint (due to the small number of available modes), and are additionally impacted by atmospheric noise.

In addition to a scale-dependent effect on the recovered power spectrum, residuals to the continuum fit could introduce an additional photon shot noise term in equation (5). From an average of an ensemble of sightlines, we estimate the worst-case size of this term to be ~ 20 per cent for $k \approx 0.1$ h Mpc $^{-1}$, which are those that Fig. 2 shows as most constraining of modified gravity. However, it should be possible to do substantially better by, for example, masking the most contaminated areas of the sky using high resolution *Planck* maps. Since the magnitude of the potential photon shot noise from galactic foregrounds is difficult to estimate accurately in advance, we simply note that continuum estimation is an important part of the data analysis but do not include it in our forecasts.

4 RESULTS

In this section, we review the Fisher matrix formalism used to derive constraints, and describe the specifics of the survey we forecast, motivated by the relevant scales needed to constrain modified gravity. We then present constraints for a future mm-wave LIM experiment as a function of sensitivity and account for various systematic effects.

4.1 Fisher matrix formalism

Fisher matrix methods are a standard way of estimating the precision of future experiments (Albrecht et al. 2006). Beginning from an assumption of Gaussian errors, by Taylor expanding about the true parameter values, we have

$$\exp\left(-\frac{1}{2}\chi^2\right) \propto \exp\left(-\frac{1}{2}F_{jk}\delta p_j\delta p_k\right) \quad (20)$$

where the matrix F_{jk} is called the Fisher matrix and can be evaluated as

$$F_{jk} = \sum_b \frac{N_b}{\sigma_b^2} \frac{\delta f_b}{\delta p_j} \frac{\delta f_b}{\delta p_k}. \quad (21)$$

The Fisher matrix is equivalent to the inverse of the covariance matrix. Equation 21 instructs us to estimate the covariance matrix by computing derivatives of the observable quantity in bins labelled by b and with corresponding error σ_b . Inverting F_{jk} then yields the variance and covariance of the model parameters. In our case, we use the binned power spectrum, $P_b(k)$, and estimate the error per bin σ_b using equations (7), (17), and (18). An explicit expression of the Fisher matrix in terms of the modified gravity parameters is

$$F_{M,B} = \sum_k \frac{N_k}{\sigma_k^2} \frac{\delta P(k)}{\delta c_{M,B}} \frac{\delta P(k)}{\delta c_{M,B}}. \quad (22)$$

Equation (21) is sufficient for estimating the covariance in the c_B and c_M parameters from a single measurement of the power spectrum. In the case where multiple emission lines are independently used to constrain the power spectrum shape, the combined Fisher matrix is given by the sum of the independent Fisher matrices for each line, $F^{A+B} = F^A + F^B$.

As the RSD multipole moments are not statistically independent, we compute the joint constraint from the full 5. (inverse) covariance

matrix:

$$F_{M,B} = \sum_k \sum_{l,l'} \frac{\delta P_l(k)}{\delta c_{M,B}} c_{l,l'}^{-1} \frac{\delta P_{l'}(k)}{\delta c_{M,B}}, \quad (23)$$

where l, l' run over the 0 and 2 RSD multipole moments. As the derivatives in equations (21) and (23) depend on the values of c_M , c_B , we assign them fiducial values from Noller & Nicola (2019) for our forecasted results in Sections 4.3 and 4.4.

4.2 Survey definition and accessible scales

Constraints on the modified gravity models considered here will search for a *nearly scale-invariant change* in the matter power spectrum from Λ CDM for $k \gtrsim 10^{-2} \text{ h Mpc}^{-1}$. This implies that the astrophysical line emission terms in equation (4) need to be known to better than the \sim few per cent deviations in P_m that we are considering. A LIM survey's sensitivity to the power spectrum falls off at the largest scales due to foreground filtering, atmospheric noise, and the decreasing number of Fourier modes in a finite survey volume. A heuristic for the sensitivity of a survey to an observable, e.g. the matter power spectrum on a given k -scale, is to count the number of accessible modes accessible at that scale.

The number of observable modes can be improved by increasing either the spectral or angular resolution or survey sky fraction. The LIM surveys we consider here are mismatched in angular and spectral resolution; while the arcmin scales accessible with 5–10m class dishes correspond to $k \sim 1\text{--}10 \text{ h Mpc}^{-1}$, current mm-wave spectrometers have only been demonstrated up to $R \sim 300$ corresponding to $k \sim 0.2 \text{ h Mpc}^{-1}$. However, a factor of \sim several improvement in resolution should be possible with technology developments in the near future. In Fig. 2, we show the matter power spectrum ($k^3 P_m(k)$) and cumulative signal-to-noise ratio on the power spectrum deviation as a function of k for representative values of c_M , c_B in both parametrizations at $z = 0.5$, focusing on the difference between the $R = \frac{\delta\nu}{\nu} = 300$ and 1000 cases. Increasing the spectral resolution increases the number of modes in the survey, leading to improved sensitivity even when a survey is unable to resolve the smallest scale structures. Most of the constraining power in the $R = 300$ case occurs around $k \sim 0.1 \text{ h Mpc}^{-1}$ due to the larger number of modes after accounting for both the spectral resolution and number of modes contained in the survey volume. The scale at which non-linear growth affects the power spectrum at the 2 per cent level is $k \sim 0.1 \text{ h Mpc}^{-1}$ at $z \approx 0$, and $k \sim 0.25 \text{ h Mpc}^{-1}$ at $z = 3$ (the mean redshift of the lines considered is shown in Table 1). Thus for $R = 1000$ about 50 per cent of the constraining power comes from weakly non-linear scales. The exact level of non-linear growth expected in Horndeski's gravity is uncertain, but experiments with HALOFIT (Smith et al. 2003) suggest a moderate increase in the matter power spectrum, and thus moderately improved sensitivity. To be conservative, we assume the predictions of linear theory. Another source of non-linear biasing is the relationship between CO emissivity and dark matter power, which depends on the CO luminosity function (Breyse et al. 2014) and exhibits non-linear effects at $k \sim 0.2 \text{ h Mpc}^{-1}$ (Einasto et al. 2019). The remaining uncertainties in the scale at which non-linear biasing becomes important will be decreased with future small-scale detection of CO shot noise.

As a baseline survey definition, we consider a survey over 40 per cent of the sky observing 75–310 GHz with $R = 300$. The sky fraction is set in part by the physical limits of telescopes and optics that often restrict observing to elevation angles $\geq 40\text{--}50$ deg. Bright emission from the Galactic centre can further restrict accessible sky

fractions by another ≈ 10 per cent. This survey geometry corresponds to a range of accessible scales between $\approx 2 \times 10^{-3} \leq k \leq 5 \times 10^{-1} \text{ h Mpc}^{-1}$. The maximum scale is set by the resolution in the frequency direction, while the minimum scale is set by the assumed sky fraction. Increasing the sky fraction to 70 per cent improves access to the largest scales by about a factor of 2, while the smallest scales remain limited by the resolution in the frequency direction. Atmospheric and galactic thermal continuum foregrounds can also limit sensitivity to the largest scale modes.

Fixing the sky fraction and bandwidth allows us to make the estimates of the noise power given in Table 1. The white-noise contribution arises from incident photon power from the atmosphere, telescope, and detector (equations (A1), (A2)). Within each of the mm-wave atmospheric windows ($\approx 75\text{--}115$, $125\text{--}175$, and $180\text{--}310$ GHz), we use an effective noise equivalent temperature (NET) in which all of the frequency channels within each band are nverse variance weighted. We then calculate the voxel volume using equation (A4) for the minimum spatial and frequency scales, set by the telescope's angular resolution and spectrometer spectral resolution, respectively. We convert between NET and integration time via equation (A3).

4.3 Fiducial analysis

For our fiducial survey, we assume the experiment described in the previous section, with $R = 300$ spectral resolution and $f_{\text{sky}} = 40$ per cent. In Table 1, we summarize our target lines, redshifts, and shot noise estimates. Combined constraints are obtained by summing the Fisher matrices over the full bandwidth, assuming statistical independence of each target line and redshift. We focus on the effects of survey definition for the sensitivity of the baseline survey to c_M , c_B , leaving the effects of interlopers and atmospheric foregrounds for the next section.

Fig. 4 shows the 1σ posterior widths, $\sigma(c_b)$ and $\sigma(c_M)$, in parametrization II ($\alpha_i \propto a$, top row) and parametrization I ($\alpha_i \propto \Omega_\Lambda$, bottom row), as a function of *spectrometre-hours*, the product of the number of spectrometers and integration time. In addition to the matter power spectrum constraint, we also show the results from the RSD power spectrum monopole and quadrupole separately. The fiducial experiment ($f_{\text{sky}} = 40$ per cent, $R = 300$) achieves similar sensitivity to the RSD monopole and matter power spectrum, while the quadrupole is significantly less constraining (Chung et al. 2019).

Sensitivity can be improved by increasing the number of modes through larger survey volumes or improved spectral resolution. In the centre and right-hand panels of Fig. 4, we forecast for increasing the spectral resolution from $R = 300$ to $R = 1000$ at fixed sky fraction and for increasing the sky fraction from $f_{\text{sky}} = 40$ per cent to $f_{\text{sky}} = 70$ per cent. We find a factor of a few improvement in sensitivity, achieving a ± 0.1 level constraint on each of the c_B , c_M in parametrization II at 10^8 spectrometre-hours, with improved sensitivity with longer integration times in both the $R = 1000$ and $f_{\text{sky}} = 70$ per cent experiments.

Increasing the spectral resolution provides larger returns on sensitivity than going to higher sky fractions at fixed spectrometre-hours, with sensitivity approaching the ± 0.01 level in $\approx 10^8$ spectrometre-hours. However, this result depends on the assumed shot noise for each target line. In a test where the shot noise was assumed to take its $z = 2$ values from Moradinezhad Dizgah et al. (2022), sensitivity saturates near ± 0.1 at 10^8 spectrometre-hours in the $R = 1000$ experiment in both parametrizations. Less sensitivity to the c_M , c_B is achieved in parametrization I, regardless of survey definition.

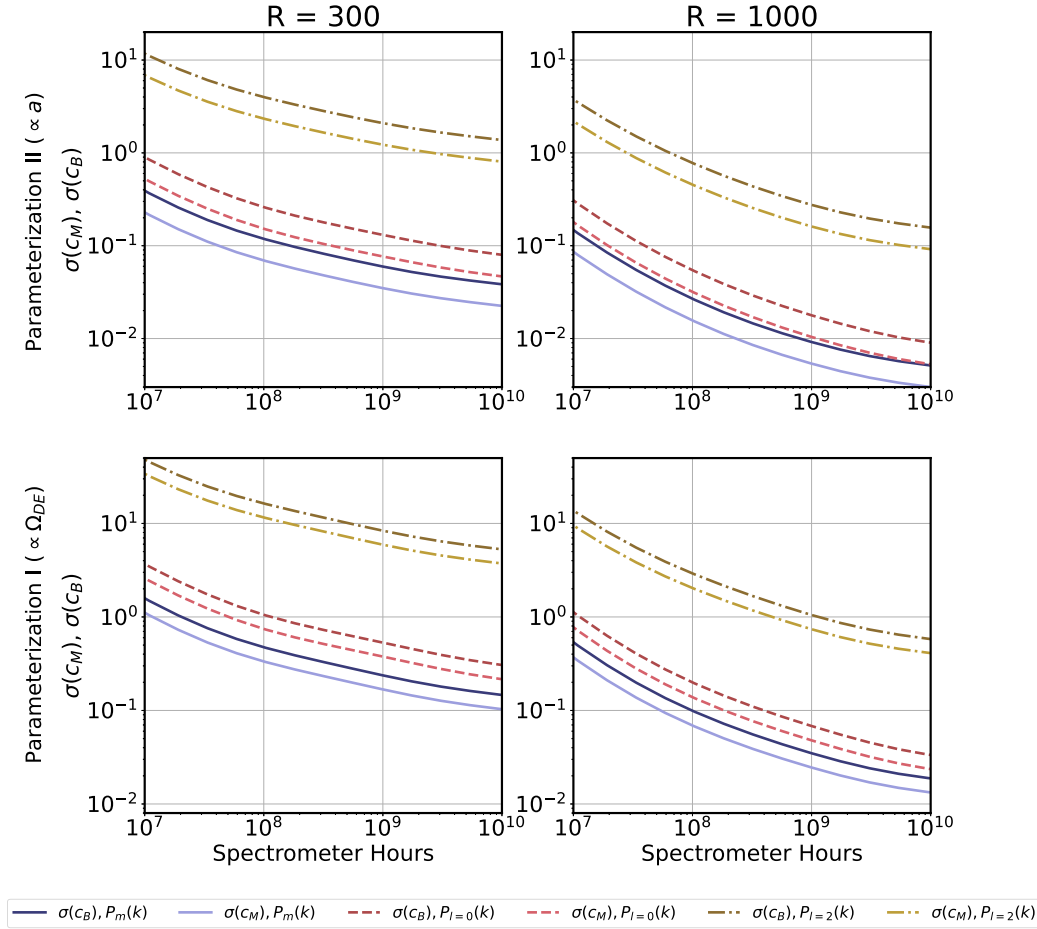


Figure 4. Differences in forecasted sensitivity to the c_M , c_B parameters is a factor of ≈ 2 , independent of sky fraction or spectral resolution. Here, we show forecasted sensitivity (posterior width) as a function of spectrometre-hours for the c_M and c_B parameters from the matter power spectrum and redshift space distortion multipoles in the baseline ($R = 300$, $f_{\text{sky}} = 40$ per cent) and increased spectral resolution ($R = 1000$) case. At fixed spectral resolution, sensitivity scales with sky fraction as the square root of the sky fraction ratio. Top panels are for parametrization II and the bottom panel is for parametrization I.

4.4 Accounting for interlopers and low-frequency noise

In order to quantify the effect of different analysis choices on sensitivity to the power spectrum, we now consider the impact that interlopers and low-frequency noise have on measurements of the c_M and c_B parameters. We begin by modifying the fiducial analysis and baseline survey according to the discussion in Section 3.3. Both the survey geometry and atmospheric scale limit the maximum accessible scales. Since the signal-to-noise ratio on measurements of the power spectrum decreases significantly on the largest scales, the atmospheric parameters ℓ_{knee} and α will only significantly impact the constraint if they differ substantially from the scale set by the survey geometry. As discussed in Section B, we assume that the atmospheric noise will be similar to that observed at the south pole. As the relevant observable scales are above ℓ_{knee} (see Fig. 2), the choice of ℓ_{knee} has little impact on our results.

We further consider the effect of interloper lines that mimic redshift-dependent intensity fluctuations and therefore pose a potentially more serious problem. Interlopers can mimic a modified gravity effect since at fixed redshift, a change in the intensity bias is degenerate with a change in the growth function. In Fig. 5, we plot the sensitivity as a function of spectrometre-hours for both the baseline case ($f_{\text{sky}} = 40$ per cent, $R = 300$) and a case including

interlopers and low-frequency noise. We treat the interlopers following Section 3.4.2, where interloper lines are assumed to contribute noise but not signal to the measurement of the modified gravity parameters. Here, we consider the sensitivity from $P_m(k)$ and the sum of $P_{l=0}(k)$ and $P_{l=2}(k)$ computed using equations (12)–(14) and (23). Interlopers are treated as in equation (19).

The inclusion of interlopers can significantly reduce constraining power. While we obtain ± 0.1 level constraints in the fiducial survey for parametrization II in 10^8 spectrometre-hours, this now requires 10^9 spectrometre-hours or more. In parametrization I, a ± 0.1 constraint is no longer obtained in our range of spectrometre-hours. While such a measurement would still allow for characterization of the size of modified gravity effect over a range in redshift, an interloper-contaminated LIM measurement would add only a very limited amount of information as compared to the existing CMB and LSS measurements.

When there is clean separation between interloper and target lines (the ‘interloper free’ baseline case), $P_m(k)$ is more sensitive to the values of the c_M , c_B than the sum of the information from $P_{l=0}(k)$ and $P_{l=2}(k)$. However, in the case of poor line separation, this situation is reversed, where the RSD multipole moments retain more of the sensitivity that is lost in the matter power spectrum. That is, the difference between the interloper and interloper-free cases is smaller.

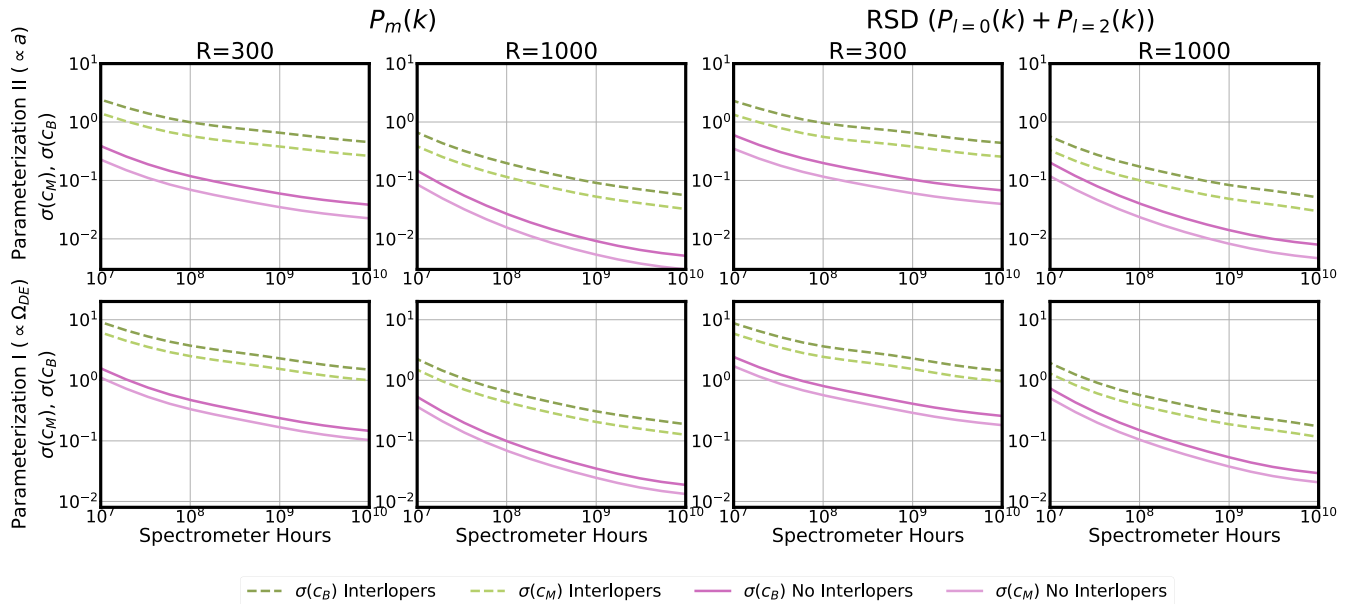


Figure 5. Including interlopers leads to a decrease in sensitivity both from $P_m(k)$ and the sum of $P_{l=0}(k) + P_{l=2}(k)$, with a reduced sensitivity gap for the redshift space measurements. As before, we plot forecasted sensitivity (posterior width) as a function of spectrometer hours for the c_M and c_B parameters when interlopers are included or excluded in the baseline survey. Left-hand panels show sensitivity from the matter power spectrum and right-hand panels for RSD multipoles. Top panels show parametrization II, and bottom panels show parametrization I.

This result is anticipated by the close relationship between Alcock–Paczynski tests, which can be used to achieve line separation, and the redshift space distortion. The two effects are degenerate with the matter power spectrum and require an assumed background cosmology to fully isolate from one another (Samushia, Percival & Raccanelli 2012). The degree of line separation working in redshift space depends on the linear growth factor f and ratios of volumes between the target and interloper line redshifts.

The two cases we have considered here (foreground/interloper-free and interloper-contaminated) roughly bound the range of expected sensitivity. The interloper-contaminated case we have considered is unrealistically pessimistic, where no attempt is made to remove interlopers before performing a cosmological analysis. Numerous techniques have been proposed in the literature for reducing their contributions (see Section 3.4.2). Although outside the scope of this work, one complication for LIM measurements of modified gravity is that several interloper mitigation schemes (e.g. geometric methods) depend on an assumed near- Λ CDM expansion history. Fully quantifying the effect of various assumptions on the recovery of the signatures of modified gravity is left for future work.

5 DISCUSSION

The observable signature of Horndeski’s gravity on LSS is a scale-independent change in the normalization of the matter power spectrum for $k > 10^{-3} \text{ h Mpc}^{-1}$, observable in either comoving or redshift space. Noller & Nicola (2019) found that the inclusion of $f\sigma_8$ from BOSS DR11 CMASS and 6dF led to increased sensitivity relative to the CMB-only and CMB + mPk cases. We therefore also forecast for an experiment targeting the RSD monopole and quadrupole, which carry information about the velocity field.

As both the matter power spectrum and redshift space distortions probe the structure formation history, which depend directly on how the strength of gravity is changing, we find more sensitivity to the c_M parameter as compared to the scalar field-metric coupling parameter

c_B . We find similar sensitivity to c_B , c_M from combining the first two moments of the RSD power spectrum and the matter power spectrum alone. Consistent with the expectations from Chung et al. (2019), the quadrupole contributes limited sensitivity compared to the monopole-only result. This is because the uncertainties on the quadrupole power spectrum are a factor of $\approx \sqrt{2l+1}/2$ larger than the uncertainties on the monopole. We assume fiducial models for the line biases and temperatures as discussed in Section 3.1 under the assumption that both will be well constrained by future experiments, for example, through multiple-line cross correlations or via cross-correlation with galaxy surveys Chung et al. (2019).

The ± 0.1 sensitivity for both c_B , c_M we find in both parametrizations at 10^8 – 10^9 spectrometre-hours are competitive with the $\sigma(c_B) \approx 0.5$ – 1.1 and $\sigma(c_M) \approx 0.3$ – 0.8 found in (Noller & Nicola 2019) from the combination of CMB + LSS. In a combined shear-lensing-clustering analysis, Mancini et al. (2019) placed tighter $\sigma(c_B)$, $\sigma(c_M) \approx 0.2$ – 0.3 constraints from LSS in the parametrization I. Our forecasted sensitivity is therefore, at worst, competitive with these existing measurements and, at best, represent a significant (up to order of magnitude) improvement in sensitivity.

Unmitigated interloper emission can reduce the sensitivity at fixed integration time by roughly an order of magnitude. This is expected since the primary effect of modifying gravity in our parametrizations mimics a change to the growth function with redshift. As interloper lines add noise power from a range of redshifts, this adds scatter to the inferred growth function, or equivalently the overall amplitude of the power spectrum on a range of scales.

The inclusion of interlopers in our baseline surveys leads to a reduction in sensitivity to the Horndeski’s linear theory α functions. We expect the sensitivity of a future experiment to lie somewhere between the no interlopers and interlopers cases shown in Fig. 5. Although a large number of methods to mitigate the effect of interlopers on cosmological analyses have been studied in the literature, this motivates future work to understand how interlopers may bias future measurements in cosmology.

An LIM experiment will produce a set of redshift-dependent power spectrum amplitudes that are weighted by the line temperatures and bias factors. Under a fixed background cosmology and assumed evolution of the line intensities, internal and external cross correlations can be used to both disentangle the interloper contributions and significantly reduce degeneracies between the astrophysics- and cosmology-dependent terms.

Our constraints make use of only the average line intensity across the target band and not its redshift evolution. However, the line intensity is expected to trace star formation and therefore peak at $z \approx 2$, while the modified gravity power spectrum excess is expected to grow monotonically with redshift. Therefore, the evolution of the two effects is expected to generically differ with an overall change in the line evolution as compared to Λ CDM. This provides another potential avenue for an LIM experiment to probe modified gravity directly from the redshift evolution. Making use of this information will require improvements in our understanding of the line evolution ($I(z)$) models and scaling relations that link these models to the SFR.

While direct constraints on Horndeski's gravity from galaxy surveys have been challenged by limited cosmological volumes and uncertainty in the galaxy bias, next-generation galaxy surveys will probe larger volumes, allowing for joint analysis and cross-correlations that can break degeneracies between multiple probes. Galaxy-LIM cross correlations and multiline LIM cross-correlations, for example, can separate the line bias and intensities even in the presence of interlopers (Schaan & White 2021).

6 CONCLUSIONS

In this work, we investigated the ability of a wide-bandwidth ground-based LIM experiment targeting rotational CO transitions to constrain the linear theory parameters of Horndeski models. We consider two parametrizations for the evolution of these parameters, governing the braiding and running of the Planck mass, where both are allowed to evolve with the effective dark energy density Ω_{DE} or with the scale factor a . Both parametrizations predict larger effects at low redshift, with excesses in apparent power at small scales and deficits in power at large scales for a large part of this 2D parameter space.

With observations in three atmospheric bands from 75 to 310 GHz, we find that the bright rotational CO transitions from redshifts 0–3 yield posterior widths for these parameters approaching the sensitivity of CMB and existing galaxy survey constraints at 10^8 – 10^9 spectrometre-hours. This result is robust to the presence of continuum foreground and atmospheric effects, being primarily driven by information obtained from intermediate scales and therefore mainly limited by the degree of interloper contamination. Models in which the modified gravity effect is proportional to the scale factor rather than Ω_{DE} yield constraints that are about an order of magnitude larger at fixed integration time, a result that is consistent with past measurements. There is a significant uncertainty about what limits the sensitivity of future experiments in the space of noise, astrophysical, and cosmological modelling uncertainties, and our results should therefore be viewed as a preliminary estimate of the performance of a real instrument. None the less, these results show that future LIM experiments could place competitive constraints on the space of modified gravity theories.

Horndeski's theories represent a general class of modified gravity models that add scalar-coupled terms to the gravitational Lagrangian. As discussed in Bellini & Sawicki (2014), measuring values of the α functions therefore constrains the parameter space of viable modifi-

cations to GR. These include metric $f(R)$, Kinetic Gravity Braiding, Galileon, Brans–Dicke, Palatini, and Gauss–Bonnet models.

LIM experiments with CMB heritage could potentially reach 10^8 – 10^9 spectrometre-hours over the next 10–15 yr. On similar time-scales, space-based spectro-polarimeters operating in the far-IR are expected to become feasible (Delabrouille et al. 2019). A space-based instrument would trade angular spatial resolution for increased sensitivity to the integrated line emission through wider bandwidth and reduced large-scale noise due to a lack of atmosphere. Combined with a larger f_{sky} , this would enable a range of complementary CMB and galaxy cluster science. As the signature of a modified gravity effect is largely scale-independent on intermediate scales, such an experiment would be able to improve constraints on deviations from GR through both direct measurement of the matter power spectrum and through multitracers analyses similar to the one that we consider here.

Measurement of modified gravity effects will require improvements in our knowledge of target line biases and intensities to break parameter degeneracies. While analysis and modelling methods for LIM remain in their infancy compared to well-developed methods for CMB and galaxy survey measurements, LIM experiments targeting rotational CO benefit from both this heritage and bright line temperatures. This makes these transitions promising targets for constraining modified gravity theories. Our results show that future LIM experiments can achieve constraints on the linear parameters of Horndeski's theories that are competitive with the current state of the art.

ACKNOWLEDGEMENTS

We thank Garrett Keating for shot noise estimates used in this forecast and for helpful conversations. We are also grateful to Johannes Noller for helpful comments on modified gravity constraints from gravitational waves and the BOSS power spectrum normalization, as well as Yu Feng, for assistance with estimates of the impact of the dust continuum. Finally, we thank an anonymous reviewer for detailed comments.

SB is supported by NSF grant AST-1817256. KSK is supported by an NSF Astronomy and Astrophysics Postdoctoral Fellowship under award AST-2001802.

DATA AVAILABILITY

All data are available publicly at <https://github.com/bscot/Horndeski-LIM>

REFERENCES

- Abazajian K. N. et al., 2016, preprint (arXiv:1610.02743)
- Ade P. et al., 2018, *Phys. Rev. Lett.*, 121, 221301
- Ade P. et al., 2020, *A&A*, 642, A60
- Aghanim N. et al., 2020, *A&A*, 641, A6
- Albrecht A. et al., 2006, preprint (arXiv:astro-ph/0609591)
- Arai S., Nishizawa A., 2018, *Phys. Rev. D*, 97, 104038
- BICEP2 Collaboration, 2015, *ApJ*, 812, 176
- Baker T., Bellini E., Ferreira P. G., Lagos M., Noller J., Sawicki I., 2017, *Phys. Rev. Lett.*, 119, 251301
- Bellini E., Sawicki I., 2014, *J. Cosmol. Astropart. Phys.*, 2014, 050
- Bellini E., Sawicki I., Zumalacárregui M., 2020, *J. Cosmol. Astropart. Phys.*, 2020, 008
- Bernal J. L., Kovetz E. D., 2022, *A&AR*, 30, 5
- Blas D., Lesgourgues J., Tram T., 2011, *J. Cosmol. Astropart. Phys.*, 2011, 034

- Breyse P. C., 2022, preprint (arXiv:2209.01223)
- Breyse P. C., Kovetz E. D., Kamionkowski M., 2014, *MNRAS*, 443, 3506
- Breyse P. C., Kovetz E. D., Kamionkowski M., 2015, *MNRAS*, 452, 3408
- Breyse P. C., Yang S., Somerville R. S., Pullen A. R., Popping G., Maniyar A. S., 2022, *ApJ*, 929, 30
- Cheng Y.-T., Chang T.-C., Bock J. J., 2020, *ApJ*, 901, 142
- Chung D. T. et al., 2019, *ApJ*, 872, 186
- Clifton T., Ferreira P. G., Padilla A., Skordis C., 2012, *Phys. Rep.*, 513, 1
- Cosmic Visions 21 cm Collaboration et al., 2018, Inflation and Early Dark Energy with a Stage II Hydrogen Intensity Mapping Experiment, preprint (arXiv:1810.09572)
- Cremineilli P., Tambalo G., Vernizzi F., Yingcharoenrat V., 2020, *J. Cosmol. Astropart. Phys.*, 2020, 002
- Creque-Sarbinowski C., Kamionkowski M., 2018, *Phys. Rev. D*, 98, 063524
- Crites A. T. et al., 2014, in Holland W. S., Zmuidzinas J., eds, SPIE Conf. Ser. Vol. 9153, Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII. SPIE, Bellingham, p. 91531W
- DeBoer D. R. et al., 2017, *PASP*, 129, 045001
- Delabrouille J. et al., 2019, preprint (arXiv:1909.01591)
- Einasto J., Liivmägi L. J., Suhonenko I., Einasto M., 2019, *A&A*, 630, A62
- Fonseca J., Silva M. B., Santos M. G., Cooray A., 2016, *MNRAS*, 464, 1948
- Gong Y., Cooray A., Silva M., Santos M. G., Bock J., Bradford C. M., Zemcov M., 2011, *ApJ*, 745, 49
- Gong Y., Cooray A., Silva M. B., Zemcov M., Feng C., Santos M. G., Dore O., Chen X., 2017, *ApJ*, 835, 273
- Gong Y., Chen X., Cooray A., 2020, *ApJ*, 894, 152
- Hamilton A. J. S., 1998, *Astrophysics and Space Science Library*. Springer Netherlands, p. 185
- Ho P. T. P. et al., 2009, *ApJ*, 694, 1610
- Hojjati A., Pogossian L., Zhao G.-B., 2011, *J. Cosmol. Astropart. Phys.*, 2011, 005
- Horndeski G. W., 1974, *Int. J. Theor. Phys.*, 10, 363
- Ihle H. T. et al., 2022, *ApJ*, 933, 185
- Ivezić Ž. et al., 2019, *ApJ*, 873, 111
- Karkare K. S., Bird S., 2018, *Phys. Rev. D*, 98, 043529
- Karkare K. S. et al., 2020, *J. Low Temp. Phys.*, 199, 849
- Karkare K. S. et al., 2022, *J. Low Temp. Phys.*, 209, 758
- Keating G. K., Marrone D. P., Bower G. C., Leitch E., Carlstrom J. E., DeBoer D. R., 2016, *ApJ*, 830, 34
- Keating G. K., Marrone D. P., Bower G. C., Keenan R. P., 2020, *ApJ*, 901, 141
- Kovetz E. D. et al., 2017, preprint (arXiv:1709.09066)
- Kreisch C., Komatsu E., 2018, *J. Cosmol. Astropart. Phys.*, 2018, 030
- Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
- Li B., Koyama K., 2019, Modified Gravity. WORLD SCIENTIFIC. Available at: <https://www.worldscientific.com/doi/abs/10.1142/11090>
- Li T. Y., Wechsler R. H., Devaraj K., Church S. E., 2016, *ApJ*, 817, 169
- Lidz A., Taylor J., 2016, *ApJ*, 825, 143
- McQuinn M., Zahn O., Zaldarriaga M., Hernquist L., Furlanetto S. R., 2006, *ApJ*, 653, 815
- Magaña J., Matos T., 2012, *J. Phys.: Conf. Ser.*, 378, 012012
- Mancini A. S. et al., 2019, *MNRAS*, 490, 2155
- Mantz A., Allen S. W., Ebeling H., Rapetti D., 2008, *MNRAS*, 387, 1179
- Moradinezhad Dizgah A., Keating G. K., Fialkov A., 2019, *ApJ*, 870, L4
- Moradinezhad Dizgah A., Keating G. K., Karkare K. S., Crites A., Choudhury S. R., 2022, *ApJ*, 926, 137
- Nadolski A. et al., 2020, *Appl. Opt.*, 59, 3285
- Noller J., 2020, *Phys. Rev. D*, 101, 063524
- Noller J., Nicola A., 2019, *Phys. Rev. D*, 99, 103502
- Padmanabhan H., 2017, *MNRAS*, 475, 1477
- Padmanabhan H., Breyse P., Lidz A., Switzer E. R., 2022, *MNRAS*, 515, 5813
- Paine S., 2022, The am atmospheric model (12.0). Zenodo. Available at: <https://zenodo.org/record/5794521>
- Peirone S., Koyama K., Pogossian L., Raveri M., Silvestri A., 2018, *Phys. Rev. D*, 97
- Perlmutter S. et al., 1999, *ApJ*, 517, 565
- Pogossian L. et al., 2022, *Nature Astron.*, 6, 1484
- Pullen A. R., Doré O., Bock J., 2014, *ApJ*, 786, 111
- Rapetti D., Allen S. W., Mantz A., 2008, *MNRAS*, 388, 1265
- Riess A. G. et al., 1998, *AJ*, 116, 1009
- Righi M., Hernández-Monteagudo C., Sunyaev R. A., 2008, *A&A*, 489, 489
- Samushia L., Percival W. J., Raccanelli A., 2012, *MNRAS*, 420, 2102
- Sato-Polito G., Bernal J. L., 2022, preprint (arXiv:2202.02330)
- Schaan E., White M., 2021, *J. Cosmol. Astropart. Phys.*, 2021, 067
- Schmidt B. P. et al., 1998, *ApJ*, 507, 46
- Sheth R. K., Mo H. J., Tormen G., 2001, *MNRAS*, 323, 1
- Shirokoff E. et al., 2012, in Holland W. S., Zmuidzinas J., eds, SPIE Conf. Ser. Vol. 8452, Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VI. SPIE, Bellingham, p. 84520R
- Silva M. B., Santos M. G., Gong Y., Cooray A., Bock J., 2013, *ApJ*, 763, 132
- Silva M. B., Santos M. G., Cooray A., Gong Y., 2015, *ApJ*, 806, 209
- Silva M. B., Zaroubi S., Koostra R., Cooray A., 2017, *MNRAS*, 475, 1587
- Smith R. E. et al., 2003, *MNRAS*, 341, 1311
- Spergel D. et al., 2015, preprint (arXiv:1503.03757)
- Sun G. et al., 2018, *ApJ*, 856, 107
- Taruya A., Nishimichi T., Saito S., 2010, *Phys. Rev. D*, 82, 063522
- Troxel M. et al., 2018, *Phys. Rev. D*, 98, 043528
- Wang Y., 2006, *ApJ*, 647, 1
- Wu P.-J., Zhang X., 2022, *J. Cosmol. Astropart. Phys.*, 2022, 060
- Yue B., Ferrara A., Pallottini A., Gallerani S., Vallini L., 2015, *MNRAS*, 450, 3829
- Zumalacárregui M., Bellini E., Sawicki I., Lesgourgues J., Ferreira P. G., 2017, *J. Cosmol. Astropart. Phys.*, 2017, 019

APPENDIX A: ESTIMATING NOISE POWER

Noise power can be expressed in terms of the NET or the noise equivalent flux density. For consistency with our parametrization of the line luminosities in terms of line temperature (μK), we choose to work in NET. We begin by assuming a dual polarization instrument and calculate the noise equivalent power (NEP) from the incident photon load Q for each detector:

$$\text{NEP}_{\text{ph}} = 2h\nu Q + \frac{1}{N_{\text{modes}}} \frac{2Q^2}{\Delta\nu}, \quad (\text{A1})$$

where ν is the detector center frequency and $\Delta\nu$ is the bandwidth. Q is the sum of power arriving at the detector from the atmosphere and emission from the telescope:

$$Q_{\text{tot}} = Q_{\text{atm}} + Q_{\text{tel}}. \quad (\text{A2})$$

A detector observing a load of temperature T with optical efficiency η sees photon power $Q \approx 2\eta kT\Delta\nu$ (for $h\nu \ll kT$). We use the *am* atmospheric modelling software to calculate the typical atmospheric temperature at each frequency for the South Pole winter (Paine 2022). The telescope emission is assumed to be at the ambient South Pole temperature $\sim 250\text{ K}$ with an emission $\epsilon = 0.01$, as measured for the *South Pole Telescope*. Finally, we assume that each detector has an NEP of $\sim 10^{-18} \text{ W}/\sqrt{\text{Hz}}$, which is added in quadrature to the incident photon NEP.

After converting the NEP to a white noise level $\sigma_{\text{rms}} (\approx 481 \mu\text{K} \cdot \sqrt{s})$, the noise power spectrum for a given integration time per pixel t_{pix} is

$$P_{\text{N}} = V_{\text{vox}} \frac{\sigma_{\text{rms}}^2}{t_{\text{pix}}}, \quad (\text{A3})$$

where the voxel volume is

$$V_{\text{vox}} = r(z)^2 \frac{\lambda(1+z)^2}{H(z)} \Omega_{\text{pix}} \Delta\nu. \quad (\text{A4})$$

Here $r(z)$ is the comoving radial distance, λ is the wavelength, $H(z)$ is the Hubble parameter, and Ω_{pix} is the pixel size.

APPENDIX B: ATMOSPHERIC FLUCTUATIONS

Atmospheric fluctuations generate scale-dependent noise. A frozen pattern of 2D fluctuations blowing across the field of view at fixed height above the instrument produces a $1/f$ or Kolmogorov spectrum with an approximate form of $P_N(\ell) \propto \ell^{-8/3}$. Here we have introduced ℓ , the angular counterpart to k , which is the Fourier transform pair of the angle θ on the sky. Since atmospheric fluctuations are local to the instrument, it is common to express their effects in ℓ rather than through the redshift-dependent mapping to k . Since the fluctuations are finite in size, they only affect the largest accessible scales, with the cut-off used to define the parameter ℓ_{knee} , such that

$$P_N(\ell) = P_N \left(1 + \left(\frac{\ell}{\ell_{\text{knee}}} \right)^\alpha \right). \quad (\text{B1})$$

The values of ℓ_{knee} and α are determined empirically from fits to observed band powers at fixed scan rate. In our forecasts, we fix $\alpha = -2.8$ and $\ell_{\text{knee}} = 200$ to be consistent with measured values from contemporary fast-scanning CMB experiments in temperature (Ade et al. 2018). These values are scan strategy-dependent and should be viewed as approximate. Moreover, it is possible that LIM measurements will have improved noise properties due to the ability to excise atmospheric lines in the spectroscopic measurement. Near-future pathfinder experiments will provide more detailed atmospheric characterization suitable for LIM forecasts and better inform estimates of the largest scales at which LIM is sensitive to cosmology.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.