A Forecast for Large Scale Structure Constraints on Horndeski Gravity with Line Intensity Mapping

Bryan R Scott,1* Kirit S. Karkare,2,3 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, Illinois 60510, USA

Accepted XXX. Received YYY; in original form ZZZ

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 theories. Our forecast assumes a multi-tracer 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 ≈ 10^8 – 10^9 spectrometer hours, feasible with instruments that could be deployed in the 2030s. Such a measurement would constrain large portions of the remaining parameter space available to Scalar-Tensor modified gravity theories. Our modeling code is publicly available.

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

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 (BAO; 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 theories are the most general theories which include a scalar field and have 2nd 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).

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 theories derived in Hojjati et al. (2011); 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 (y) plane (cf. Mantz et al. (2008); Rapetti et al. (2008)).

As a general framework of modified gravity, Horndeski theories have a variety of operators which can be constrained by different cosmological experiments. In the context of effective field theory parameterizations 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. Mancini et al. (2019) pursue lensing and clustering constraints on Horndeski 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 k 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).

* E-mail: bryan.scott@email.ucr.edu

© 2022 The Authors
This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

© 2022 The Authors
This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.
ground-based wide-bandwidth mm-wave LIM experiments targeting multiple rotational CO transitions over the redshift range \(z = 0 - 3\). 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 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 \(\Lambda\)CDM background. As such, where necessary, we assume a flat \(\Lambda\)CDM-like cosmology with \(h = 0.678, \Omega_b h^2 = 0.0224, \Omega_c h^2 = 0.12\), and \(\Omega_{\Lambda} = 1 - \Omega_m\).

2 HORNDESKI GRAVITY

Horndeski 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 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 (Blaizot et al. 2011) to produce the extended Horndeski in Linear Cosmic Anisotropy Solving System (HI_CLASS) (Zumalacárregui et al. 2017; Bellini et al. 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 4 functions, two of which are in principle measurable by LIM (\(\alpha_B\) and \(\alpha_M\)) and two of which are not (\(\alpha_K\) and \(\alpha_T\)). They have the following physical interpretations:

- \(\alpha_B\) encodes mixing between the scalar and metric perturbations that arises from the clustering of the Horndeski scalar field, and appears as perturbations to \(\bar{T}_{0\mu}\). \(\alpha_B = 0\) in \(\Lambda\)CDM + GR. We treat \(\alpha_B\) as a free parameter to be constrained by the LIM experiment.
- \(\alpha_M\) 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 \(\alpha_M\) parameterizes the evolution of the Planck mass with time, \(\alpha_M = 0\) in \(\Lambda\)CDM + GR. We treat \(\alpha_M\) as a free parameter to be constrained by the LIM experiment.
- \(\alpha_K\) 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, \(\alpha_K\) 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 \(\Lambda\)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: \(\alpha_K + \frac{3}{2}a_T^2 > 0\).
- \(\alpha_T\) gives the tensor speed excess, potentially inducing anisotropic stress, even in the absence of scalar perturbations. \(\alpha_T = 0\)

In this paper, we investigate the constraining power of future
in $\Lambda$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 (HL_CLASS), which we use to predict the matter power spectrum $P_m(k)$ under an assumed $\Lambda$CDM background, and to vary the free functions according to the parameterizations described in the next section.

### 2.1 Parameterizations

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

$$ds^2 = -(1 + 2\Psi)dt^2 + (1 + 2\Phi)dx^i dx^j,$$

where $\Psi$ and $\Phi$ 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 theory the potentials are also complicated functions of the $\alpha_i$, 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 $\alpha_i$ allows any evolution for the background spacetime to be realized. LSS alone cannot pick out either the expansion history or a unique form for the $\alpha_i$. To reduce this freedom, we begin by first noting that geometric measurements are consistent with the universe being nearly $\Lambda$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 parameterize how modifications to gravity evolve with time.

**Parameterization I**: A natural choice in a nearly-$\Lambda$CDM universe is to parameterize the modified gravity effect as proportional to the cosmological constant density, $\Omega_\Lambda$. 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 parameterization we assume that $\alpha_B$ and $\alpha_M$ are linear functions of $\Omega_\Lambda$:

$$\alpha_{B,M} = c_{B,M} \Omega_\Lambda.$$  

(2)

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

**Parameterization II**: To evaluate the sensitivity of our probes to observations at high redshift, we use an alternate parameterization 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

$$\alpha_{B,M} = c_{B,M} a.$$  

(3)

Figure 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 $\Lambda$CDM are possible for extreme values of the $\alpha$ functions. As noted in Noller & Nicola (2019), curves that intersect the $\Lambda$CDM prediction exhibit a degeneracy between $c_M$ and $c_B$ for the matter power spectrum with $c_B \approx 1.8c_M$ in Parameterization II ($\alpha_T \ll a$). 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 parameterizations requires an experiment that targets a large range in redshift.

By selecting a fiducial $k$-scale to summarize the effects of varying $\alpha_M$ and $\alpha_B$, 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 $\Lambda$CDM). However, this turnover occurs on scales near the cosmological horizon and is thus extremely difficult to measure with LSS measurements.\(^1\)

At the intermediate scales measured by a LIM experiment, the characteristic feature of modified gravity models relative to $\Lambda$CDM + GR is a uniform excess in the power spectrum. The size and behavior with varying $c_M, c_B$ of the effect depends strongly on the choice of parameterization, with a $< 1\%$ difference in $\Lambda$CDM at $z = 3$ in Parameterization I and a few percent 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 Parameterization 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 Parameterization II than in Parameterization I.

To summarize, we have two modified gravity functions that scale with the background evolution of the spacetime that we seek to constrain. The $c_B$ and $c_M$ parameters govern respectively the evolution of the braiding $\alpha_B$ (clustering of dark energy) and the Planck mass run rate $\alpha_M$ (the large-scale strength of gravity). $\Lambda$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 parameterization that scales with the effective dark energy component density $\Omega_{DE}$ and one that scales with the scale factor. We fix the remaining two functions to be constant and assign them to unity. We specifically forecast for the uncertainties $\sigma(c_B), \sigma(c_M)$.

### 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, in addition to the effects of interloper lines and Galactic foregrounds.

#### 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 nonlinear regime, we can parameterize 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

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

(4)

Here $P_m(k, z)$ is the underlying matter power spectrum which 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 Figure 2.

\(^1\) The turnover is an unambiguous signature of modified gravity, and would constrain the braiding scale, a function of $\alpha_M$ and $\alpha_B$. 

MNRAS 000, 1–13 (2022)
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, 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.

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, 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.

and the product of the matter power spectrum normalization and linear growth function, $f(z)\sigma_8(z)$, in Figure 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-Torman-inspired (Sheth et al. 2001) bias evolution from Dizgah et al. (2021).

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 $(dn/dM(M,z))$ or line luminosity functions $(dn/dL(z))$, that are obtained through empirical scaling relations through the dependence of L(M) with SFR(M,z) or 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 ten. For a recent review of line intensity modeling, 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 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 $\Lambda$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.

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

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

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

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

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

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

Estimates for $P_{\text{shot}}$ are given in Table 1, while estimates of $P_N$ are discussed in Section 3.3.
Figure 2. Top panel shows the matter power spectrum at $z = 0.5$ in both parameterizations. 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 $f_{\text{sky}} = 40\%$ case. The 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 larger $k$ 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 the CO(1-0), CO(2-1), and CO(3-2) lines that we target in our forecast are brightest. 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 $I(z)$. 
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 $(k, \mu)$ 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

$$P_{\text{obs}}(k, \mu, z) = [b(z)^2 I(z)^2 + f(z)^2 I(z)^2 \mu^2] P_m(k).$$ \hspace{1cm} (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 theories we consider in Figure 3.

Noller & Nicola (2019) consider constraints on the Horndeski theory parameters $c_B$ and $c_M$ from both anisotropic clustering measurements of the growth factor $D$ in SDSS and $f \sigma_8$ at $z = 0.57$ 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 $(k, \mu, z)$ space, it is simpler to consider constraints from the non-vanishing $l = 0, 2, 4$ moments obtained by convolving Eq. 8 with the Legendre polynomials $\mathcal{L}_l$:

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

Explicit expressions for the monopole and quadrupole are

$$P_0(k) = \left(1 + \frac{2}{3} \beta (bl)^2 + \frac{1}{5} (bl)^2 \beta^2 \right) P_m(k)$$

$$P_2(k) = \frac{4}{7} (bl)^2 \beta^2 P_m(k).$$ \hspace{1cm} (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:

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

Taruya et al. (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:

$$\text{Cov}_{0,0}(k) = \frac{2}{N_k} \left( 1 + \frac{4}{3} \beta + \frac{6}{5} \beta^2 + \frac{4}{7} \beta^3 + \frac{4}{9} \beta^4 \right)$$

$$\times (bl)^2 P_m(k)^2 + 2P_N \left( 1 + \frac{2}{3} \beta + \frac{1}{5} \beta^2 \right) (bl)^2 P_m(k) + P_N^2.$$ \hspace{1cm} (12)

For the monopole-quadrupole cross-term:

$$\text{Cov}_{0,2}(k) = \frac{2}{N_k} \left( \frac{8}{3} \beta + \frac{24}{7} \beta^2 + \frac{40}{21} \beta^3 + \frac{40}{9} \beta^4 \right)$$

$$\times (bl)^2 P_m(k)^2 + 2P_N \left( \frac{4}{3} \beta + \frac{4}{7} \beta^2 \right) (bl)^2 P_m(k).$$ \hspace{1cm} (13)

Finally, for the quadrupole:

$$\text{Cov}_{2,2}(k) = \frac{2}{N_k} \left( \frac{5}{21} \beta + \frac{90}{7} \beta^2 + \frac{1700}{231} \beta^3 + \frac{2075}{1287} \beta^4 \right)$$

$$\times (bl)^2 P_m(k)^2 + 2P_N \left[ \frac{5}{21} \beta + \frac{30}{7} \beta^2 \right] (bl)^2 P_m(k) + 5P_N^2.$$ \hspace{1cm} (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 Parametrization I ($\alpha \approx \Omega_{\Lambda}$), the excess near the turnover in the matter power spectrum at $k = 0.01$ is $\sim 1\%$ at $z = 3$, and an order of magnitude larger at $z = 0.5$. The evolution in the effect of modified gravity for Parametrization II ($\alpha \approx \alpha$) is comparable in magnitude, but begins at earlier redshifts, as shown in Figure 1.

We consider an experiment measuring the CO $J = 1 \rightarrow 1$ rotational transitions, which emit at rest-frame frequencies of $115$ 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 $z < 2$), it has been detected in individual galaxies at high redshift using ground-based telescopes observing in the millimeter 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 $75$–$310$ 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 et al. 2015), and current instruments have demonstrated wideband 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 spectrometer arrays that approach CMB packing efficiency. Our forecasts anticipate that this technology can be scaled over the next ten years 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: $75$–$115$ GHz, $120$–$175$ GHz, 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...
Table 1. Line frequencies, target redshifts, $P_{\text{shot}}$ 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   | $v_{\text{rest}}$ [GHz] | $v_{\text{obs}}$ [GHz] | $z$  | $P_{\text{shot}}$ [$\mu$K$^2$] | Temp [$\mu$K] |
|--------|-------------------------|-------------------------|------|-------------------------------|---------------|
| CO(1-0)| 115                     | 95                      | 0.21 | 2.66                          | 0.14          |
| CO(2-1)| 230                     | 95                      | 1.42 | 8.54                          | 0.75          |
| CO(2-1)| 230                     | 95                      | 1.42 | 8.54                          | 0.75          |
| CO(2-1)| 230                     | 150                     | 0.53 | 81.8                          | 0.24          |
| CO(2-1)| 230                     | 150                     | 1.3  | 100                           | 0.46          |
| CO(3-2)| 345                     | 245                     | 0.4  | 295                           | 0.14          |
| CO(3-2)| 345                     | 245                     | 0.4  | 295                           | 0.14          |

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 spectrometer resolution $\delta v$, while in the transverse direction, the smoothing is a function of the beamwidth $\theta_b$. These correspond to comoving smoothing scales at redshift $z$ in the transverse $\sigma_{\perp}$ and parallel (to the line of sight) direction $\sigma_{\parallel}$. In the perpendicular (spatial) direction, the smoothing scale is

$$\sigma_{\perp} = \frac{\theta_b R(z)}{\sqrt{8\pi^2}}$$

(15)

where $\theta_b$ 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 $H(z)$,

$$\sigma_{\parallel} = \frac{c \delta v (1 + z)}{H(z)v_{\text{obs}}}$$

(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,

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

(17)

where $\mu$ is the cosine of the angle between the wavevector $k$ and the line-of-sight direction, and the integral averages over all such angles $\mu$ to yield the spatially-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, $\ell$), above a scale $\ell_{\text{knee}}$. 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 Eqs. 12 - 14 even in the case of finite instrument resolution.

3.4.2 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 $1/f$ Kolmogorov spectrum with an approximate form of $P_N(\ell) \propto \ell^{-8/3}$. Here we have introduced $\ell$, the angular counterpart to $k$, which is the Fourier transform pair of the angle $\theta$ on the sky. Since atmospheric fluctuations are local to the instrument, it is common to express their effects in $\ell$ 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 $\ell_{\text{knee}}$, such that

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

(18)

The values of $\ell_{\text{knee}}$ and $\alpha$ 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.

3.4.3 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 Eq. 7 to

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

(19)

For the RSD multipoles, we similarly modify Eq. 11 to sum over the RSD power spectrum at each redshift,

$$\text{Cov}_{l,l'}(k) = \frac{(2l+1)(2l'+1)}{N_m} \times \int_{-\ell}^{1} \mathcal{L}_l(\mu) \mathcal{L}_{l'}(\mu)(\Sigma_8 P_m(k, \mu, z_i) + P_N)^2 d\mu.$$  

(20)

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 wideband 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 Yue et al. (2015); Breysse 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 et al. 2020). Finally, geometric tests for interloper deconfusion were introduced in Silva et al. (2015) and Lidz & Taylor (2016).

3.4.4 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 the Galactic dust continuum, and studied recovery of the power spectrum. We began by adding a mock LIM signal to a typical Galactic dust spectrum, and then removed a series of low-order polynomials. We find that even under pessimistic assumptions, foreground removal only affects large scales ($k \sim 10^{-3} \text{ h/Mpc}$) that contribute little weight to the overall constraint (due to the small number of available modes), and which are additionally impacted by atmospheric noise. In a simple but more realistic model for the continuum fitting, the difference between the input and recovered spectrum is less than 1% on intermediate scales. We therefore neglect galactic continuum foregrounds in our forecast.

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

$$\exp\left(-\frac{1}{2}\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}\right) \propto \exp\left(-\frac{1}{2}F_{jk}\delta p_j \delta p_k\right)$$

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

$$F_{jk} = \sum_b \frac{N_b}{\sigma_b^2} \frac{\delta f_b}{\delta p_j} \frac{\delta f_b}{\delta p_k}.$$  \hspace{1cm} (22)

The Fisher matrix is equivalent to the inverse of the covariance matrix. Equation 22 instructs us to estimate the covariance matrix by computing derivatives of the observable quantity in bins labelled by $b$ and with corresponding error $\sigma_b$. 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 $\sigma_b$ using Equations 7, 17, and 19. An explicit expression of the Fisher matrix in terms of the modified gravity parameters is

$$F_{M,B} = \sum_k \frac{N_b}{\sigma_k^2} \frac{\delta P_k}{\delta c_{M,B}} \frac{\delta P_k}{\delta c_{M,B}}.$$  \hspace{1cm} (23)

Equation 22 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 covariance matrix:

$$F_{M,B} = \sum_l \sum_{l'} \frac{\delta P_l}{\delta c_{M,B}} \frac{\delta P_{l'}}{\delta c_{M,B}},$$  \hspace{1cm} (24)

where $l$, $l'$ run over the 0 and 2 RSD multipole moments.

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 $\Lambda$CDM for $k \gtrsim 10^{-2} \text{ h/Mpc}$. This implies that the astrophysical line emission terms in Eq. 4 need to be known to better than the ~few % 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 feasible 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 arcminute scales accessible with 5–10m class dishes correspond to $k \sim 1 - 10 \text{ h/Mpc}$, current mm-wave spectrometers have only been demonstrated up to $R \sim 300$ corresponding to $k \sim 0.2$. However, a factor of ~several improvement in resolution should be possible with technology developments in the near future. In Fig. 2 we show the matter power spectrum ($k^3P_m(k)$) 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 parameterizations 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 $k \sim 0.1 \text{ h/Mpc}$, 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% level is $k \sim 0.1 \text{ h/Mpc}$ at $z \approx 0$, and $k \sim 0.25 \text{ h/Mpc}$ at $z = 3$ (the mean redshift of the lines considered is shown in Table 1). Thus for $R = 1000$ about 50% of the constraining power comes from weakly non-linear scales. The exact level of non-linear growth expected in Horndeski 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 (Breyssse et al. 2014) and exhibits non-linear effects at $k \sim 0.2$ h/Mpc (Einasto et al. 2019). The remaining uncertainties in the scale at which nonlinear 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% 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 $\geq 40$–50 deg. Bright emission from the Galactic center can further restrict accessible sky fractions by another $\approx 10\%$. This survey geometry corresponds to a range of accessible scales between $\approx 2 \times 10^{-3} \leq k \leq 5 \times 10^{-1}$ h/Mpc. 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% improves access to the largest scales by about a factor of two, 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 ($\approx 75–115, 125–175,$ and $180–310$ GHz), we use an effective NET in which all of the frequency channels within each band are inverse 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 $f_{\text{sky}} = 40\%$. In Table 1 we summarize our target lines, redshifts, and shot noise of each target line and redshift. We focus on the effects of survey previous section, with integration time via Equation A3.

The degree of line separation working in redshift space depends on the linear growth factor $c_M$ and $c_B$ parameters. The inclusion of interlopers can significantly reduce constraining power. While we obtain $\pm 0.1$ level constraints in the fiducial survey for Parameterization II in $10^8$ spectrometer hours, this now requires $10^9$ spectrometer hours or more. In Parameterization I, a $\pm 0.1$ constraint is no longer obtained in our range of spectrometer-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. 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 et al. 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.

with sensitivity approaching the $\pm 0.01$ level in $\approx 10^8$ spectrometer 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 Dizgah et al. (2021), sensitivity saturates near $\pm 0.1$ at $10^8$ spectrometer-hours in the $R = 1000$ experiment in both parameterizations. Less sensitivity to the $c_M, c_B$ is achieved in Parameterization I, regardless of survey definition.

### 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 $\ell_{\text{knee}}$ and $\alpha$ will only significantly impact the constraint if they differ substantially from the scale set by the survey geometry. As discussed in Section 3.4.2, we assume that the atmospheric noise will be similar to that observed at the South Pole. As the relevant observable scales are above $\ell_{\text{knee}}$ (see Figure 2) the choice of $\ell_{\text{knee}}$ 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 Figure 5, we plot the sensitivity as a function of spectrometer-hours for both the baseline case ($\ell_{\text{knee}} = 40\%, R = 300$) and a case including interlopers and low frequency noise. We treat the interlopers following Section 3.4.3, 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 to 14 and 24. Interlopers are treated as in Equation 20.

The inclusion of interlopers can significantly reduce constraining power. While we obtain $\pm 0.1$ level constraints in the fiducial survey for Parameterization II in $10^8$ spectrometer hours, this now requires $10^9$ spectrometer hours or more. In Parameterization I, a $\pm 0.1$ constraint is no longer obtained in our range of spectrometer-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. 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 et al. 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.
Figure 4. Sensitivity to the $c_M$, $c_B$ parameters from the matter power spectrum or redshift space distortion monopole differs by a factor of $\approx 2$ independent of spectral resolution or sky fraction. Here we show forecasted sensitivity (posterior width) as a function of spectrometer hours for the $c_M$ and $c_B$ parameters in the baseline ($R = 300$, $f_{\text{sky}} = 40\%$), increased spectral resolution ($R = 1000$, $f_{\text{sky}} = 40\%$), and increased survey volume ($R = 300$, $f_{\text{sky}} = 70\%$) cases. Top panels are for Parameterization II and the bottom panel is for Parameterization I.

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 panels show sensitivity from the matter power spectrum, right panels for RSD multipoles. Top panels show Parameterization II, bottom panels show Parameterization I.
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 3.4.3). 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-$\Lambda$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 gravity on LSS is a scale-independent change in the normalisation of the matter power spectrum for $k > 10^{-3}$ h/Mpc, observable in either comoving or redshift space.

Noller & Nicola (2019) found that the inclusion of $f_{\sigma8}$ 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. 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 $\approx \sqrt{2I+1}/2$ 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 multiplex-line cross correlations or via cross-correlation with galaxy surveys Chung et al. (2019).

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 parameterizations 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 linear theory $\alpha$ functions. We expect the sensitivity of a future experiment to lie somewhere between the no interlopers and interlopers cases shown in Figure 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.

A 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 $\Lambda$CDM. This provides another potential avenue for a 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 ($f(z)$) models and scaling relations that link these models to the SFR.

While direct constraints on Horndeski 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 multi-line 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 parameterizations 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 $\Omega_{DE}$ or with the scale factor $a$. Both parameterizations 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 75–310 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 $10^8–10^9$ spectrometer-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 $\Omega_{DE}$ yield constraints that are about an order of magnitude larger at fixed integration time, a result that is consistent with past measurements. There is significant uncertainty about what limits the sensitivity of future experiments in the space of noise, astrophysical, and cosmological modeling uncertainties, and our results should therefore be viewed as a preliminary estimate of the performance of a real instrument. Nonetheless, these results show that future LIM experiments could place competitive constraints on the space of modified gravity theories.

Horndeski 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 $\alpha$ functions therefore constrains the parameter space of viable modifications to General Relativity. These include metric $f(R)$, Kinetic Gravity Braiding, Galileon, Brans-Dicke, Palantini, and Gauss-Bonnet models.

LIM experiments with CMB heritage could potentially reach $10^8–10^9$ spectrometer hours over the next 10–15 years. On similar timescales, 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

MNRAS 000, 1–13 (2022)
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 General Relativity through both direct measurement of the matter power spectrum and through multi-tracer analyses similar to the one 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 modeling 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 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 also thank Johannes Noller for helpful comments on both constraints from gravitational waves and the BOSS power spectrum normalization.

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 is available publicly at https://github.com/bscot/Horndeski-LIM

REFERENCES

Abazajian K. N., et al., 2016, arXiv e-prints, p. arXiv:1610.02743
Ade P., et al., 2018, Physical Review Letters, 121
Ade P., et al., 2020, Astronomy & Astrophysics, 642, A60
Aghanim N., et al., 2020, Astronomy & Astrophysics, 641, A6
Albrecht A., et al., 2006, arXiv e-prints, pp astro-ph/0609591
Arai S., Nishizawa A., 2018, Phys. Rev. D, 97, 104038
BICEP2 Collaboration et al., 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, Journal of Cosmology and Astroparticle Physics, 2014, 050–050
Bellini E., Sawicki I., Zamaladinegroui M., 2020, Journal of Cosmology and Astroparticle Physics, 2020, 008–008
Bernal J. L., Kovetz E. D., 2022, Line-Intensity Mapping: Theory Review, doi:10.48550/ARXIV.2206.15377, https://arxiv.org/abs/2206.15377
Blas D., Lesgourgues J., Tran T., 2011, Journal of Cosmology and Astroparticle Physics, 2011, 034–034
Breysse P. C., Kovetz E. D., Kamionkowski M., 2014, Monthly Notices of the Royal Astronomical Society, 443, 3506–3512
Breysse P. C., Kovetz E. D., Kamionkowski M., 2015, MNRAS, 452, 3408
Breysse P. C., Yang S., Somerville R. S., Pullen A. R., Popping G., Maniyar A. S., 2021, On estimating the cosmic molecular gas density from CO Line Intensity Mapping observations (arXiv:2106.14904)
Cheng Y.-T., Chang T.-C., Bock J. J., 2020, The Astrophysical Journal, 901, 142
Chung D. T., et al., 2019, The Astrophysical Journal, 872, 186
Clifton T., Ferreira P. G., Padilla A., Skordis C., 2012, Physics Reports, 513, 1–189
Cosmic Visions 21 cm Collaboration et al., 2019, Inflation and Early Dark Energy with a Stage II Hydrogen Intensity Mapping Experiment (arXiv:1810.09572)
Creminelli P., Tambalo G., Vernizzi F., Yngchoorenrat V., 2020, J. Cosmology Astropart. Phys., 2020, 002
Creque-Sarbinowski C., Kamionkowski M., 2018, Physical Review D, 98
Crites A. T., et al., 2014, in Holland W. S., Zmuidzinas J., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 9153, Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII, p. 9153W, doi:10.1117/12.2057207
DeBoer D. R., et al., 2017, Publications of the Astronomical Society of the Pacific, 129, 045001
Delabrouille J., et al., 2019, arXiv e-prints, p. arXiv:1909.01591
Dzigna A. M., Keating G. K., Karkare K. S., Crites A., Choudhury S. R., 2021, Neutrino Properties with Ground-Based Millimeter-Wavelength Line Intensity Mapping (arXiv:2110.00014)
Einasto J., Liivamägi L. J., Suhhonenko I., Einasto M., 2019, A&A, 630, A62
Fonseca J., Silva M. B., Santos M. G., Cooray A., 2016, Monthly Notices of the Royal Astronomical Society, 464, 1948–1965
Gong Y., Cooray A., Silva M., Santos M. G., Bock J., Bradford C. M., Zemcov M., 2011, The Astrophysical Journal, 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, The Astrophysical Journal, 894, 152
Hamilton A. J. S., 1998, in, Astrophysics and Space Science Library, Springer Netherlands, pp 185–275, doi:10.1007/978-94-011-4960-0_17
Ho P. T. P., et al., 2009, The Astrophysical Journal, 694, 1610–1618
Hojjati A., Pogosian L., Zhao G.-B., 2011, Journal of Cosmology and Astroparticle Physics, 2011, 005–005
Horndeski G. W., 1974, International Journal of Theoretical Physics, 10, 363
Ihle H. T., et al., 2021, COMAP Early Science: IV. Power Spectrum Methodology and Results, doi:10.48550/ARXIV.2111.05930, https://arxiv.org/abs/2111.05930
Ivezic Z., et al., 2019, ApJ, 873, 111
Karkare K. S., Bird S., 2018, Physical Review D, 98
Karkare K. S., et al., 2020, Journal of Low Temperature Physics, 199, 849–857
Karkare K. S., et al., 2022, Journal of Low Temperature Physics, 220, 1–189
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, arXiv e-prints, p. arXiv:1709.09066
Kreisch C., Komatsu E., 2018, Journal of Cosmology and Astroparticle Physics, 2018, 030–030
Laureijs R., et al., 2011, arXiv e-prints, p. arXiv:1110.3193
Li B., Koyama K., 2019, Modified Gravity. WORLD SCIENTIFIC (https://www.worldscientific.com/doi/pdf/10.1142/11090), doi:10.1142/11090, https://www.worldscientific.com/doi/abs/10.1142/11090
Li T. Y., Wechsler R. H., Devearaj K., Church S. E., 2016, The Astrophysical Journal, 817, 169
Lidz A., Taylor J., 2016, The Astrophysical Journal, 825, 143
Mancini A. S., et al., 2019, Monthly Notices of the Royal Astronomical Society, 490, 2155
Mantz A., Allen S. W., Ebeling H., Rapetti D., 2008, MNRAS, 387, 1179
McQuinn M., Zahn O., Zaldarriaga M., Hernquist L., Furlanetto S. R., 2006, The Astrophysical Journal, 653, 1477–1484
Moradinezhad Dzigha A., Keating G. K., Fialkov A., 2019, The Astrophysical Journal, 870, L4
Nadolski A., et al., 2020, Appl. Opt., 59, 3285
Noller J., 2020, Phys. Rev. D, 101, 063524
Noller J., Nicola A., 2019, Physical Review D, 99
Padmanabhan H., 2017, Monthly Notices of the Royal Astronomical Society, 475, 1477–1484
Padmanabhan H., Breysse P., Lidz A., Switzer E. R., 2021, Intensity mapping

MNRS 000, 1–13 (2022)
from the sky: synergizing the joint potential of [OIII] and [CII] surveys at reionization (arXiv:2105.12148)

Paine S., 2022, The am atmospheric model (12.0).

Peirone S., Koyama K., Pogosian L., Raveri M., Silvestri A., 2018, Physical Review D, 97

Pullen A. R., Doré O., Bock J., 2014, The Astrophysical Journal, 786, 111

Rapetti D., Allen S. W., Mantz A., 2008, Monthly Notices of the Royal Astronomical Society, 388, 1265–1278

Riess A. G., et al., 1998, AJ, 116, 1009

Righi M., Hernández-Monteagudo C., Sunyaev R. A., 2008, Astronomy & Astrophysics, 489, 489–504

Samushia L., Percival W. J., Raccanelli A., 2012, Monthly Notices of the Royal Astronomical Society, 420, 2102

Schaan E., White M., 2021, Journal of Cosmology and Astroparticle Physics, 2021, 067

Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1

Shirokoff E., et al., 2012, in Holland W. S., Zmuidzinas J., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 8452, Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VI. p. 84520R (arXiv:1211.1652), doi:10.1117/12.927070

Silva M. B., Santos M. G., Gong Y., Cooray A., Bock J., 2013, The Astrophysical Journal, 763, 132

Silva M. B., Santos M. G., Cooray A., Gong Y., 2015, Prospects for detecting CII emission during the Epoch of Reionization (arXiv:1410.4808)

Silva M. B., Zaroubi S., Kooistra R., Cooray A., 2017, Tomographic Intensity Mapping versus Galaxy Surveys: Observing the Universe in H-alpha emission with new generation instruments (arXiv:1711.09992)

Smith R. E., et al., 2003, MNRAS, 341, 1311

Sun G., et al., 2018, ApJ, 856, 107

Taruya A., Nishimichi T., Saito S., 2010, Physical Review D, 82

Troxel M., et al., 2018, Physical Review D, 98

Wang Y., 2006, The Astrophysical Journal, 647, 1–7

Wu P.-J., Zhang X., 2021, Prospects for measuring dark energy with 21 cm intensity mapping experiments (arXiv:2108.03552)

Yue B., Ferrara A., Pallottini A., Gallerani S., Vallini L., 2015, Monthly Notices of the Royal Astronomical Society, 450, 3829–3839

Zumalacárregui M., Bellini E., Sawicki I., Lesgourgues J., Ferreira P. G., 2017, Journal of Cosmology and Astroparticle Physics, 2017, 019–019

APPENDIX A: ESTIMATING NOISE POWER

Noise power can be expressed in terms of the Noise Equivalent Temperature (NET) or the Noise Equivalent Flux Density (NEFD). For consistency with our parameterization 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:

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

where $\nu$ 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:

$$Q_{\text{tot}} = Q_{\text{atm}} + Q_{\text{tel}}.$$

A detector observing a load of temperature $T$ with optical efficiency $\eta$ sees photon power $P \approx 2\eta kT \Delta \nu$ (for $h\nu \ll kT$). We use the am atmospheric modeling 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 ~ 250 K with an emission $\epsilon = 0.01$, as measured for the South Pole Telescope. Finally, we assume that each detector has an NEP of ~ $10^{-18}$ W/$\sqrt{\text{Hz}}$, which is added in quadrature to the incident photon NEP.

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

$$P_N = V_{\text{vox}} \frac{\sigma_{\text{rms}}^2}{t_{\text{pix}}},$$  \hspace{1cm} (A3)

where the voxel volume is

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

Here $r(z)$ is the comoving radial distance, $\lambda$ is the wavelength, $H(z)$ is the Hubble parameter, and $\Omega_{\text{pix}}$ is the pixel size.

This paper has been typeset from a TeX/LaTeX file prepared by the author.