

# Toward accurate modeling of line-intensity mapping one-point statistics: Including extended profiles

José Luis Bernal 

*Instituto de Física de Cantabria (IFCA), CSIC-Universidad de Cantabria,  
Avenida de los Castros s/n, E-39005 Santander, Spain*

 (Received 22 September 2023; accepted 18 October 2023; published 14 February 2024)

Line-intensity mapping (LIM) is quickly attracting attention as an alternative technique to probe large-scale structure and galaxy formation and evolution at high redshift. LIM one-point statistics are motivated because they provide access to the highly non-Gaussian information present in line-intensity maps and contribute to break degeneracies between cosmology and astrophysics. Now that promising surveys are underway, an accurate model for the LIM probability distribution function (PDF) is necessary to employ one-point statistics. We consider the impact of extended emission and limited experimental resolution in the LIM PDF for the first time. We find that these effects result in a lower and broader peak at low intensities and a lower tail towards high intensities. Focusing on the distribution of intensities in the observed map, we perform the first model validation of LIM one-point statistics with simulations and find good qualitative agreement. We also discuss the impact on the covariance, and demonstrate that if not accounted for, large biases in the astrophysical parameters can be expected in parameter inference. These effects are also relevant for any summary statistic estimated from the LIM PDF, and must be implemented to avoid biased results. The comparison with simulations shows, however, that there are still deviations, mostly related with the modeling of the clustering of emitters, which encourage further development of the modeling of LIM one-point statistics.

DOI: [10.1103/PhysRevD.109.043517](https://doi.org/10.1103/PhysRevD.109.043517)

## I. INTRODUCTION

Line-intensity mapping (LIM) is an emerging observational technique that aims to obtain three-dimensional maps of the Universe using the integrated flux of bright spectral lines over cosmological scales [1–4]. Therefore, LIM probes the aggregate line emission by all sources at a given redshift. This grants access to otherwise undetectable faint populations of emitters and makes LIM the optimal tracer of the large-scale structure at high redshift in the high-noise or high-confusion regimes [5,6]. Furthermore, dropping the requirement of resolved detection of individual emitters enables the use of low-aperture telescopes and quick scans of large portions of the sky.

Besides tracing the underlying matter density fluctuations, line-intensity fluctuations depend on the astrophysical phenomena that trigger the line emission. Hence, combining different spectral lines, LIM also probes galaxy formation and evolution across cosmic times, complementing luminosity functions from galaxy surveys. LIM is attracting increasing attention, with numerous experiments and surveys currently observing [7–13] and many others that will see first light in the forthcoming years [14–20]. Preliminary detections (see, e.g., Refs. [21–29]) give rise to optimism regarding the prospects for this technique and its potential to bridge between low-redshift galaxy surveys and early

Universe probes like the cosmic microwave background (see, e.g., Refs. [30–33]).

Most commonly, the main type of summary statistics employed to analyze line-intensity maps are 2-point statistics like the correlation function or its Fourier counterpart, the power spectrum. They benefit from some robustness against smooth, uncorrelated observational contaminants and build upon the comprehensive formalism developed for galaxy surveys. However, line-intensity fluctuations, which trace the nonlinear, non-Gaussian large-scale structure and are modulated by nontrivial astrophysical processes, are very non Gaussian. Therefore, a significant part of the information contained in line-intensity maps is not captured by power spectrum measurements. Furthermore, LIM intrinsic sensitivity to both cosmology and astrophysics may hinder the interpretation of the LIM power spectra. For instance, the mean intensity and the bias relating matter and source fluctuations present a perfect degeneracy in the power spectrum at large scales. Although the degeneracy can be partially broken using smaller scales in the mildly nonlinear regime [34,35], the power spectrum is only sensitive to the first two moments of the luminosity function, which result in degenerate astrophysical and cosmological parameters (see, e.g., Refs. [36,37] for a discussion on the degeneracies between astrophysics and cosmology for the power spectrum).

Accessing non-Gaussian information and breaking degeneracies between cosmology and astrophysics motivate the development of alternative summary statistics. Non-Gaussian information can be obtained using higher-order statistics like the bispectrum and trispectrum. However, one-point statistics, which depend directly on the LIM probability distribution function (PDF), hence on the full distribution of (non-Gaussian) intensity fluctuations and the whole line-luminosity function, offer preferable properties.

The simplest one-point statistics is the actual estimator of the PDF, known as voxel intensity distribution (VID) in the context of LIM [38,39]. The VID is very complementary to the power spectrum because while the latter depends on clustering and the first two moments of the line-luminosity function, the former depends on subsequent convolutions of the luminosity function and zero-lag moments of clustering. The potential of the combination of both summary statistics has been demonstrated in the literature [40,41]. Furthermore, several studies have highlighted the sensitivity of the VID not only to astrophysical parameters but also to beyond- $\Lambda$ CDM cosmologies and physics beyond the Standard Model, either directly or through reducing degeneracies when combined with the power spectrum [42–46].

First instances of the LIM PDF formalism relied on small modifications to the probability-of-deflection techniques (see, e.g., Refs. [47–50]): they considered the conditional probability of finding a given intensity in a voxel given number of emitters contained in it [38,39]. Reference [41] explicitly modeled the dependence on the local emitter overdensity to derive the first analytic model of the VID supersample covariance and the VID-power spectrum covariance. Recently, Ref. [51] adapted the formalism presented in Refs. [52,53] to compute the LIM PDF using the halo model, emphasizing the ability of the VID to break the degeneracy between the mean intensity and the emitter bias.

Nonetheless, all previous work assumes that each emitter just contributes to the voxel in which it is contained. This in an unphysical approximation: whether it is due to extended sources, line broadening caused by the internal peculiar velocity of the gas or the limited experimental resolution, the observed emission from each emitter is redistributed to a finite volume that extends beyond the voxel in which each emitter is located.

In this work we build upon the halo-model based formalism from Ref. [51] and model this effect in the analytic derivation of the LIM PDF. We apply the Ergodic hypothesis and obtain the PDF of the intensity in a voxel for a single emitter from the spatial profile of its observed intensity. From this starting point we can model the LIM PDF including the smearing of the observed signal due to extended sources, line broadening, and limited experimental resolution. For the first time, we show a comparison between the theoretical analytic prediction of the LIM PDF

and the results for simulations to validate our model. We find qualitative good agreement, satisfactorily capturing the effects of the extended intensity profiles. Including extended observed intensity profiles significantly changes the VID prediction, and they must be taken into account to ensure unbiased parameter inference. We also find, as expected, that accounting for the fact that an emitter contributes to more than one voxel increases the correlation between intensity bins. However, other parts of the modeling, especially those related to matter clustering, must be improved to match the results from simulations. The updated VID modeling is now included in the publicly available code LIM.<sup>1</sup>

While in this work we focus on the VID, extended emission and limited experimental resolution must also be included in the derivation of any summary statistic that relies on the LIM PDF. Examples beyond the VID include 1-point cross correlations of different datasets, such as the conditional VID [54] or the deconvolved density estimator [55,56].

We assume the standard  $\Lambda$ CDM cosmology, with best-fit parameter values from the full *Planck* analysis without external observations [57]. In addition, we employ the following convention for the Fourier transforms. For a  $d$ -dimensional Fourier-space variable  $\mathbf{u}$  being the conjugate of a configuration-space variable  $\mathbf{v}$ , the direct and inverse Fourier transforms of a function  $f$  and its Fourier counterpart  $\tilde{f}$  are given by

$$\begin{aligned}\tilde{f}(\mathbf{u}) &= \int d^d\mathbf{v} f(\mathbf{v}) e^{-i\mathbf{v}\mathbf{u}}, \\ f(\mathbf{v}) &= \int \frac{d^d\mathbf{u}}{(2\pi)^d} \tilde{f}(\mathbf{u}) e^{i\mathbf{v}\mathbf{u}},\end{aligned}\quad (1)$$

where the tilde denote Fourier-space functions. Given the large dynamical range of the variables considered for Fourier transform (spanning orders of magnitude, positive and negative values), we compute the Fourier transform using nonuniform fast Fourier transform routines as implemented in FINUFFT<sup>2</sup> [58,59].

This article is structured as follows. First, we review the PDF modeling from Ref. [51] and introduce the effects of extended intensity profiles in Sec. II. We discuss limited experimental resolution and line broadening as causes of the extended profiles and their impact in the PDF prediction in Sec. III. We compare theoretical predictions with simulations in Sec. IV. We discuss the relevance of modeling the extended emission estimating the bias introduced in parameter inference when point sources are considered in Sec. V. We summarize and conclude in Sec. VI.

<sup>1</sup><https://github.com/jl-bernal/lim>.

<sup>2</sup><https://finufft.readthedocs.io/en/latest/>.

## II. LIM 1-POINT PDF

In this section we review the modeling to compute the line-intensity 1-point PDF and extend it to account for extended intensity profiles. We build upon the modeling for point sources presented in Ref. [51] using the halo model, which in turn adapted the formalism introduced in Refs. [52,53] for the thermal Sunyaev-Zel'dovich effect and weak lensing convergence, respectively.

### A. LIM signal

The specific line intensity per unit frequency emitted at a given position  $\mathbf{x}$  is given by

$$I_\nu(\mathbf{x}) = \frac{c}{4\pi\nu H(z)} \rho_L(\mathbf{x}), \quad (2)$$

where  $c$  is the speed of light,  $\nu$  is the rest-frame frequency of the spectral line of interest,  $H(z)$  is the Hubble parameter at redshift  $z$  corresponding to position  $\mathbf{x}$  and  $\rho_L$  is the line luminosity density. Using the Rayleigh-Jeans relation, the brightness temperature can be defined from the specific intensity as

$$T(\mathbf{x}) = \frac{c^3(1+z)^2}{8\pi k_B \nu^3 H(z)} \rho_L(\mathbf{x}) = X_{\text{LT}} \rho_L(\mathbf{x}), \quad (3)$$

where  $k_B$  is the Boltzmann constant and we have defined  $X_{\text{LT}}$  as a redshift-dependent multiplicative factor to simplify the expressions. During this work we will use the brightness temperature as variable to describe line-intensity maps, but our approach is equally applicable to specific intensities.

We can model the spatial distribution of the line luminosity associated to a single source located at  $\mathbf{x}_s$ , without loss of generality, as

$$\frac{dL_s}{d\mathbf{x}}(\mathbf{x}|\boldsymbol{\theta}_s) = L_{s,\text{tot}}(\boldsymbol{\theta}_s) \varrho(\mathbf{x} - \mathbf{x}_s|\boldsymbol{\theta}_s), \quad (4)$$

where  $L_{s,\text{tot}}$  is the total line luminosity of the source (i.e., integrated over its line width and point spread function) and  $\varrho$  is a three-dimensional spatial emission profile with inverse-volume units that is normalized to unity:  $\int d^3\mathbf{x} \varrho(\mathbf{x}|\boldsymbol{\theta}_s) = 1$ . Any experimental or astrophysical effect that results in a spatial smoothing of the observed signal in the final map can be embedded in  $\varrho$ , which is generally anisotropic. Both  $L_{s,\text{tot}}$  and  $\varrho$  depend on a set of astrophysical properties of the source included in the parameter vector  $\boldsymbol{\theta}_s$ . Then, the temperature in a specific point can be expressed as

$$T(\mathbf{x}) = X_{\text{LT}} \sum_s \frac{dL_s}{d\mathbf{x}}(\mathbf{x}|\boldsymbol{\theta}_s), \quad (5)$$

where the sum is over all sources, indexed by  $s$ . We build upon this expression to derive the PDF for extended profiles.

### B. The PDF for extended profiles

Since brightness temperature is an additive quantity, the PDF of the aggregate emission is the convolution of the PDF of each emitter. This calculation is much more tractable in Fourier space applying the convolution theorem. Let us define  $\tau$  as the Fourier conjugate of the brightness temperature. The Fourier transform of a given PDF  $\mathcal{P}$  of the brightness temperature,

$$\begin{aligned} \tilde{\mathcal{P}}(\tau) &= \int dT \mathcal{P}(T) e^{-iT\tau} = \langle e^{-iT\tau} \rangle \\ &= \frac{1}{V_{\text{obs}}} \int_{V_{\text{obs}}} d^3\mathbf{x} e^{-iT(\mathbf{x})\tau}, \end{aligned} \quad (6)$$

is the characteristic function, and angle brackets denote average over realizations. The characteristic function is dimensionless, and can also be obtained invoking the ergodic hypothesis taking the average over the observed volume  $V_{\text{obs}}$ , as done in the last equality above.

Let us explicitly separate the halo mass  $M$  from  $\boldsymbol{\theta}$  to ease the readability. Consider first an infinitesimal mass bin centered at  $M$ , for which the brightness temperature PDF is

$$\begin{aligned} \mathcal{P}^{(M)}(T) &= \int d\boldsymbol{\theta} \mathcal{P}(\boldsymbol{\theta}|M) \\ &\times \left[ \mathcal{P}_{N=0}^{(M,\boldsymbol{\theta})} \delta_D(T) + \mathcal{P}_{N=1}^{(M,\boldsymbol{\theta})} \mathcal{P}_1^{(M,\boldsymbol{\theta})}(T) \right], \end{aligned} \quad (7)$$

where we marginalize over the conditional multi-dimensional distribution  $\mathcal{P}(\boldsymbol{\theta}|M)$  of astrophysical properties given a halo mass (see, e.g., Ref. [60]),  $\mathcal{P}_{N=x}^{(M,\boldsymbol{\theta})}$  is the PDF of having  $x$  emitters (halos) of mass  $M$  and set of properties  $\boldsymbol{\theta}$  contributing to a specific point, and  $\mathcal{P}_x^{(M,\boldsymbol{\theta})}(T)$  is the PDF of finding a temperature  $T$  in a point in the space receiving contributions from  $x$  emitters with such properties. If there is no emitter (i.e.,  $N=0$ ), then there is no signal and  $\mathcal{P}_0(T) = \delta_D(T)$  is the Dirac delta. For an infinitesimal mass bin  $\mathcal{P}_{N>1}^{(M,\boldsymbol{\theta})} = 0$ , hence  $\mathcal{P}_{N=0}^{(M,\boldsymbol{\theta})} = 1 - \mathcal{P}_{N=1}^{(M,\boldsymbol{\theta})}$ .

The characteristic function can therefore be expressed as

$$\tilde{\mathcal{P}}^{(M)}(\tau) = 1 + \int d\boldsymbol{\theta} \mathcal{P}(\boldsymbol{\theta}|M) \mathcal{P}_{N=1}^{(M,\boldsymbol{\theta})} \left( \tilde{\mathcal{P}}_1^{(M,\boldsymbol{\theta})}(\tau) - 1 \right). \quad (8)$$

Although the profile  $\varrho$  of the extended emission extends arbitrarily in space, in practice we can truncate it at some distance large enough that there is no sizable signal loss. Under this assumption, the signal profile covers a finite volume  $V_{\text{prof}}$ , which can depend on  $M$  and  $\boldsymbol{\theta}$ , so that

$$\tilde{\mathcal{P}}_1^{(M,\boldsymbol{\theta})}(\tau) = \int d^3\mathbf{x} dL \mathcal{P}(\mathbf{x}|M, \boldsymbol{\theta}) \mathcal{P}(L|M, \boldsymbol{\theta}, \mathbf{x}) e^{-iT\tau}, \quad (9)$$

where we explicitly marginalize over the position [for which  $\mathcal{P}(\mathbf{x}|M, \boldsymbol{\theta})$  is uniform over  $V_{\text{prof}}$  and zero otherwise]

and  $\mathcal{P}(L|M, \boldsymbol{\theta}, \mathbf{x})$  accounts for any distribution of the line luminosity given the halo mass, the astrophysical properties, and the distance to the center of the profile.<sup>3</sup> Equation (9) is equivalent to Eq. (6) for a single source. Note that the temperature factor in the exponential depends on all the quantities that are marginalized over.

Finally, assuming Poisson statistics, and neglecting clustering for now,

$$\mathcal{P}_{N=1}^{(M, \boldsymbol{\theta})} = dM \frac{dn}{dM} V_{\text{prof}}(M, \boldsymbol{\theta}), \quad (10)$$

where  $dn/dM$  is the halo mass function. For infinitesimal bins,  $\mathcal{P}_{N=1}^{(M, \boldsymbol{\theta})} \ll 1$ , so that we can interpret Eq. (8) as the linear expansion of the exponential

$$\tilde{\mathcal{P}}^{(M)}(\tau) = \exp \left\{ \int d\boldsymbol{\theta} \mathcal{P}(\boldsymbol{\theta}|M) \mathcal{P}_{N=1}^{(M, \boldsymbol{\theta})} (\tilde{\mathcal{P}}_1^{(M, \boldsymbol{\theta})}(\tau) - 1) \right\}. \quad (11)$$

Now we need to extend this to all halos. Using the convolution theorem, the characteristic function for the whole population is the product of the individual characteristic functions. Thus, the characteristic function  $\tilde{\mathcal{P}}^{(u)}$  for the whole population without accounting for clustering is

$$\begin{aligned} \tilde{\mathcal{P}}^{(u)}(\tau) &= \prod \tilde{\mathcal{P}}^{(M)}(\tau) \\ &= \exp \left\{ \int dM d\boldsymbol{\theta} \mathcal{P}(\boldsymbol{\theta}|M) \frac{dn}{dM} V_{\text{prof}}(M, \boldsymbol{\theta}) \right. \\ &\quad \left. \times (\tilde{\mathcal{P}}_1^{(M, \boldsymbol{\theta})}(\tau) - 1) \right\}, \end{aligned} \quad (12)$$

where we directly have substituted the sum of exponents by the integral limit in the last equality.

From the expression above, and assuming that the astrophysical properties are uncorrelated with clustering, it is trivial to include the effect of clustering. Clustering varies in scales much larger than the observed intensity profiles of the sources. Therefore, for a specific realization (or position), we can include the effects of clustering by adding the halo overdensity field  $\delta_h \equiv (n_h - \langle n_h \rangle) / \langle n_h \rangle$  to the halo mass function, where  $n_h$  is the local halo number density. Therefore, we promote

$$\frac{dn}{dM} \rightarrow \frac{dn}{dM} (1 + \delta_h(\mathbf{x}, M)). \quad (13)$$

Then, the characteristic function  $\tilde{\mathcal{P}}^{(\delta)}$  accounting for clustering for a *single* realization is

$$\begin{aligned} \tilde{\mathcal{P}}^{(\delta)}(\tau) &= \exp \left\{ \int dM d\boldsymbol{\theta} \mathcal{P}(\boldsymbol{\theta}|M) \frac{dn}{dM} V_{\text{prof}}(M, \boldsymbol{\theta}) \right. \\ &\quad \left. \times (\tilde{\mathcal{P}}_1^{(M, \boldsymbol{\theta})}(\tau) - 1) \delta_h \right\} \tilde{\mathcal{P}}^{(u)}. \end{aligned} \quad (14)$$

The global characteristic function is then the average over the realizations of  $\tilde{\mathcal{P}}^{(\delta)}$ . Since  $\tilde{\mathcal{P}}^{(u)}$  does not depend on the overdensities, we can take it out of the average, and we are left with the exponential of the term including  $\delta_h$ . We invoke the moment-generating function, which states that for a random variable  $X$ ,

$$\langle e^X \rangle = \exp \left\{ \sum_{p=1}^{\infty} \langle X^p \rangle / p! \right\}. \quad (15)$$

By definition,  $\langle \delta_h \rangle = 0$ , and all terms with  $p > 2$  vanish for a Gaussian distribution. Although gravitational collapse induces non Gaussianities and higher-order terms should be included, we take a first approximation and truncate the moment-generating function at  $p = 2$ . Furthermore, we relate  $\delta_h$  to the underlying matter density field  $\delta_m$  with a linear, mass-dependent halo bias  $b_h$  so that  $\delta_h^{(M)} = b_h(M) \delta_m$  and the second cumulant of the halo distribution is

$$\langle \delta_h^{(M)}(\mathbf{x}) \delta_h^{(M')}(\mathbf{x}) \rangle \equiv b_h(M) b_h(M') \sigma^2, \quad (16)$$

where  $\sigma^2$  is the zero-lag variance of the matter distribution.

Finally, the overall characteristic function is given by

$$\begin{aligned} \tilde{\mathcal{P}}(\tau) &= \exp \left\{ \left[ \int dM d\boldsymbol{\theta} \mathcal{P}(\boldsymbol{\theta}|M) \frac{dn}{dM} V_{\text{prof}}(M, \boldsymbol{\theta}) \right. \right. \\ &\quad \left. \left. \times (\tilde{\mathcal{P}}_1^{(M, \boldsymbol{\theta})}(\tau) - 1) b_h(M) \right]^2 \frac{\sigma^2}{2} \right\} \tilde{\mathcal{P}}^{(u)}. \end{aligned} \quad (17)$$

We can obtain the PDF of the brightness temperature in a point by computing the inverse Fourier transform of the characteristic function above. Sometimes it can be useful to consider the PDF of brightness temperature fluctuations  $\Delta T \equiv T - \bar{T}$  as a crude approximation for foreground subtraction, where  $\bar{T}$  is the mean brightness temperature. This can be obtained multiplying the characteristic function by  $e^{i\bar{T}\tau}$ .

### C. Voxelized volume and practical considerations

The derivation in the subsection above returns the PDF of the brightness temperature in a specific point. In practice, however, we measure the brightness temperature from observations in a discretized map, a data cube in which each cell or voxel (three-dimensional pixels) corresponds to a comoving volume  $V_{\text{vox}}$ . Discretizing the map is relevant at three stages of the derivation above: the volume of the

<sup>3</sup>The innermost integral in Eq. (9) is equivalent to  $\mathcal{P}(T|M)$  in Ref. [51].

profiles, the characteristic function of a single specific emitter, and the cumulants of the matter overdensity field.

After discretizing the observed volume, the measured temperature  $T_i$  in a voxel centered at  $\mathbf{x}_i$  is

$$T_i = \frac{X_{\text{LT}}}{V_{\text{vox}}} \sum_s \int_{V_{\text{vox},i}} d^3\mathbf{x} \frac{dL_s}{d\mathbf{x}}(\mathbf{x}|\boldsymbol{\vartheta}_s), \quad (18)$$

where the sum is over all sources indexed by  $s$ .

For isotropic intensity profiles and spherical voxels, the right hand side of Eq. (9) can be easily evaluated inverting the relationship between temperature and position for a given emitter (see Appendix B in Ref. [53]). However, this is not possible in a more general setup. In particular, the observed signal profiles in LIM surveys are anisotropic due to different angular and spectral resolutions, to line broadening only affecting the profile in the direction along the line of sight, etc. Moreover, although the voxelization can be arbitrary it is more optimal to follow the experimental resolutions, which do not need to correspond to similar distances in the direction along and transverse to the line of sight.

Instead, we take a different approach and explicitly compute the spatial integral in Eq. (9), adapted for a voxelized space. First, we consider a three-dimensional space and locate the emitter of interest in its center. We grid the space with cells with the same comoving size as the voxels and compute the total luminosity on each voxel with the integral of Eq. (18). We truncate the emission profile at a minimum relative luminosity  $L_{\text{rel}}^{\text{min}}$  with respect to the voxel with maximum luminosity [e.g., keeping only the  $N_{\text{vox}}^{\text{prof}}$  voxels with  $L_{\text{vox}}/\max(L_{\text{vox}}) \geq L_{\text{rel}}^{\text{min}}$ ]. We normalize the luminosity on each voxel so that their sum is again  $L_{s,\text{tot}}$ , and consider  $V_{\text{prof}} = N_{\text{vox}}^{\text{prof}} V_{\text{vox}}$ .

Then, we take advantage of the scaling property of the Fourier transform. Let us consider some scale temperature  $T_0$  so that any  $T = CT_0$ . We can assume any PDF  $\mathcal{P}_{T_0}$  for  $T_0$  to account for any scatter in the conditional relations considered in the previous subsections. We assume a log-normal distribution

$$\mathcal{P}_{T_0}(T; T_0) = \frac{\exp\left\{\frac{-\log_{10}\left(\frac{T}{T_0}\right) - \frac{\sigma_{\text{scat}}^2 \log(10)}{2}}{2\sigma_{\text{scat}}^2}\right\}}{\sqrt{2\pi T \sigma_{\text{scat}} \log(10)}} \quad (19)$$

with mean  $T_0$  and scatter  $\sigma_{\text{scat}}$  in dex, so that for any  $T' \equiv CT_0$  we have  $\mathcal{P}_{T'}(T; CT_0) = \mathcal{P}_{T_0}(T/C; T_0)/C$ , which fulfills  $\tilde{\mathcal{P}}_{T'}(\tau; CT_0) = \tilde{\mathcal{P}}_{T_0}(C\tau; T_0)$ . Grouping all  $N_{\text{vox}}^{(i)}$  voxels with the same temperature  $T_i$ , the characteristic function for a halo with mass  $M$  and properties  $\boldsymbol{\vartheta}$  is

$$\tilde{\mathcal{P}}_1^{(M,\boldsymbol{\vartheta})}(\tau) = \sum_i \frac{N_{\text{vox}}^{(i)}}{N_{\text{vox}}^{\text{prof}}} \tilde{\mathcal{P}}_{T_0}(T_i\tau/T_0; T_0), \quad (20)$$

which corresponds to the weighted average of the characteristic function for each position in the spatial profile, i.e., a discretized version of Eq. (9).

Finally, the discretization also affects the impact of clustering in the PDF. Voxels are the basic unit of information we have access to, hence halo overdensities are smoothed over scales of the size of the voxel. This can be modeled by convolving the overdensity field with the window function  $W_{\text{vox}}$  of the voxel, usually a normalized top-hat function with the extent of the voxel, leaving

$$\delta_h^{\text{v}}(\mathbf{x}) = \int d^3\mathbf{x}' W_{\text{vox}}(\mathbf{x} - \mathbf{x}') \delta_h(\mathbf{x}'), \quad (21)$$

where  $\int d^3\mathbf{x} W_{\text{vox}} = 1$ . We substitute  $\delta_h$  by  $\delta_h^{\text{v}}$ , and similarly for  $\delta_m$ , in all their instances in the previous subsection. This change can be summarized with a slight modification in Eq. (17), where the cumulant of matter perturbations must be

$$\sigma_{\text{vox}}^2 = \int \frac{d^3\mathbf{k}}{(2\pi)^3} \tilde{W}_{\text{vox}}^2(\mathbf{k}) P_m^{(s)}(k), \quad (22)$$

$P_m^{(s)}$  is the nonlinear matter power spectrum in redshift space and  $\tilde{W}_{\text{vox}}$  is the Fourier transform of the voxel window function. We estimate the nonlinear power spectrum as modeled by HMcode [61]. In this work, we model redshift-space distortions with the Kaiser factor and a phenomenological Lorentzian suppression to model the fingers of God (following, e.g., Ref. [36]) with characteristic scale

$$\sigma_{\text{FoG}} = \frac{4\pi}{3} \int \frac{dk}{(2\pi)^3} P_m^{\text{lin}}(k), \quad (23)$$

where  $P_m^{\text{lin}}$  is the linear matter power spectrum in real space.

Still, the PDF is a continuous quantity which cannot be directly measured from the observations. Instead, the VID  $\mathcal{B}$ , a histogram of the measured brightness temperature in each voxel normalized by the total number of voxels, can be used as an estimator of the PDF. The relation between the VID and the PDF of the signal (i.e., in the absence of experimental thermal noise) is<sup>4</sup>

<sup>4</sup>Previous studies do not normalize the VID by the number of voxels observed, and therefore include a factor  $N_{\text{vox}}$  multiplying the integral in Eq. (24). We prefer to normalize the VID to deal with an intensive quantitative, the value of which does not depend on the size of the survey. This approach allows for a more intuitive understanding of the VID values, as well as easier comparisons between experiments.

$$\mathcal{B}^{(s)}(\Delta T_i) = \int_{\Delta T_i} d\Delta T \mathcal{P}(\Delta T), \quad (24)$$

where the integral is limited to the temperature interval centered on  $\Delta T_i$ . Nonetheless, there is no perfect experiment without noise, and the total VID  $\mathcal{B}$  is connected to the total PDF  $\mathcal{P}_{\text{tot}}(\Delta T) = (\mathcal{P} * \mathcal{P}_{\text{noise}})(\Delta T)$  through Eq. (24), where  $\mathcal{P}_{\text{noise}}$  is the instrumental noise PDF and “\*” denotes the convolution operator.

### III. EFFECTS FROM EXTENDED EMISSION

The derivation above is general enough to account for any three-dimensional signal profile and any source that can be embedded in the halo model formalism. Suppose a series of effects that cause extended signal which independently correspond to three-dimensional profiles (or window functions)  $W_i$ , normalized to unity. Then, the final extended observed profile is given by the convolution of all of them:

$$\varrho(\mathbf{x}) = (W_1 * W_2 * \dots * W_n)(\mathbf{x}). \quad (25)$$

As an example, we limit our study to emission coming from halos and to two of the most prevalent causes of extended profiles: experimental resolution and line broadening. The former relates to the fact that any experiment has limited resolution, which prevents to access small scale information and smears the observed maps; the latter is a physical effect caused by the Doppler broadening of the emission line due to the peculiar velocities of the gas where the emission originates.<sup>5</sup>

#### A. Limited resolution and line broadening

The small-scale information in line-intensity maps is limited by the angular and spectral resolutions of the experiment. The exact characterization of the beam profiles and the line-spread function depend on each experiment. Here, we will consider a Gaussian beam with full-width half maximum  $\theta_{\text{FWHM}}$  and a Gaussian line-spread function with standard deviation given by the channel width  $\delta\nu$  of the experiment. The standard deviation of the Gaussian window functions for these resolutions correspond to comoving distance scales transverse to and along the line of sight given by

$$\sigma_{\perp} = D_M(z) \frac{\theta_{\text{FWHM}}}{\sqrt{8 \log 2}}, \quad \sigma_{\parallel} = \frac{c\delta\nu(1+z)}{H(z)\nu_{\text{obs}}}, \quad (26)$$

respectively, where  $D_M$  is the comoving angular diameter distance, and  $\nu_{\text{obs}}$  is the observed frequency. The window function  $W_{\text{res}} = W_{\text{res}}^{\perp} W_{\text{res}}^{\parallel}$  for the resolution is composed of

<sup>5</sup>In the case of the Lyman- $\alpha$  line, radiative transfer also affects the observed signal profile; we leave the study of this effect for future work.

$$W_{\text{res}}^{\perp}(\mathbf{x}_{\perp}) = \frac{\exp\{-(x_{\perp,1}^2 + x_{\perp,2}^2)/2\sigma_{\perp}^2\}}{\sqrt{2\pi}\sigma_{\perp}},$$

$$W_{\text{res}}^{\parallel}(x_{\parallel}) = \frac{\exp\{-x_{\parallel}^2/2\sigma_{\parallel}^2\}}{\sqrt{2\pi}\sigma_{\parallel}}, \quad (27)$$

where the subscript “1” or “2” denote the directions of the axis transverse to the line of sight.

We assume that the bulk of the line broadening is due to the rotation of the gas in the halo, which in turn only depends on the halo mass. Depending on the line of interest, this may not be the case, especially if the line is not emitted by hot gas but by colder gas in the galactic disk. In the latter case, the width of the line also depends on the disk mass. In this work, we limit our focus to the halo-mass dependent model proposed in Ref. [62] and leave the study of further dependencies to future work.

Let us consider that the broadening follows a Gaussian profile with full-width half maximum determined by the rotation velocity  $v(M)$  in units of physical velocity. The standard deviation of the Gaussian profile in comoving space is

$$\sigma_v(M) = \frac{v(M)(1+z)}{2\sqrt{2} \log(2) H(z)}. \quad (28)$$

Line broadening effectively reduces the map resolution along the line of sight. Therefore, the window function capturing its effect is the same  $W_{\text{res}}^{\parallel}$  from Eq. (27), but using  $\sigma_v$  instead of  $\sigma_{\parallel}$ .

However, a strictly Gaussian profile would consider that all galaxies have the same angle of inclination with respect to the line of sight. A much more reasonable assumption involves randomly oriented emitters, as considered in Ref. [62]. The actual standard deviation for a galaxy with an inclination angle  $\varphi$  is  $2\sigma_v \sin \varphi / \sqrt{3}$ , assuming that the previously calculated  $\sigma_v$  corresponds to galaxies with median inclination over random inclinations. The average profile can be obtained marginalizing over the inclination of the galaxy, which follows a uniform distribution on  $\cos \varphi \equiv \mu_{\varphi}$ :

$$W_{\text{broad}}(x_{\parallel}, M) = \int_{-1}^1 d\mu_{\varphi} \frac{\exp\{-3x_{\parallel}^2/8\sigma_v^2(M)(1-\mu_{\varphi}^2)\}}{4\sigma_v(M)\sqrt{2\pi(1-\mu_{\varphi}^2)/3}}$$

$$= \frac{\sqrt{3\pi/2}}{4\sigma_v} \text{Erfc}\left(\sqrt{3x_{\parallel}^2/8\sigma_v^2}\right), \quad (29)$$

where Erfc is the complementary error function.

If we assume that these are the only contributions to extended signal, applying Eq. (25) we find

$$\begin{aligned} \varrho(\mathbf{x}, M) &= W_{\text{res}}(\mathbf{x}) * W_{\text{broad}}(x_{\parallel}, M) \\ &= W_{\text{res}}^{\perp}(\mathbf{x}_{\perp}) \left( W_{\text{res}}^{\parallel}(x_{\parallel}) * W_{\text{broad}}(x_{\parallel}, M) \right). \end{aligned} \quad (30)$$

If we were to ignore the effect of random inclinations for the line broadening, then  $\varrho^{\parallel}$  would be a Gaussian with standard deviation  $\sqrt{\sigma_{\parallel}^2 + \sigma_v^2(M)}$ . However, we have not found an analytic expression for the convolution of a Gaussian and Eq. (29). Numerically, it can be quickly computed using the convolution theorem. For reference, the Fourier transform of Eq. (29) is

$$\tilde{W}_{\text{broad}}(k_{\parallel}) = \left( \sqrt{\frac{2}{3}} k_{\parallel} \sigma_v \right)^{-1} F \left( \sqrt{\frac{2}{3}} k_{\parallel} \sigma_v \right), \quad (31)$$

where  $F$  is the Dawson integral.

### B. Experimental case and astrophysical model

By design, the VID depends not only on the intrinsic signal but also on the experimental setup of the LIM survey, e.g., through the noise, the resolution, the voxel volume, etc. For instance, if the spectral resolution is bad enough, the effect of line broadening will be negligible. This is why we need to focus on a specific case and generalize the qualitative effect of extended signal in the VID. To augment the effect of line broadening, we choose good experimental resolution.

We consider the spectral line CO(1-0) observed at  $\nu_{\text{obs}} = 28.8$  GHz ( $z = 3$ ), with  $\theta_{\text{FWHM}} = 3$  arcmin and  $\delta\nu = 15$  MHz, and a noise-per-voxel standard deviation of  $\sigma_N = 2.5$   $\mu$ K. This ensures some effect of the line broadening and that the VID is only dominated by noise at small brightness temperatures. Since we are interested in the VID prediction, we assume a good understanding of the foregrounds, calibration and sky subtraction.

The optimal pixel size for a projected angular map, balancing the minimization of the correlation between different intensity bins and the number of pixels to reduce the statistical uncertainties corresponds to  $\theta_{\text{FWHM}}$  [63]. We find the same to be true for the direction along the line of sight. Therefore, we choose our voxel size in the direction transverse to and along the line of sight as the correspondent to  $\theta_{\text{FWHM}}$  and  $\delta\nu^{\text{FWHM}} \equiv \delta\nu\sqrt{8\log 2}$ , respectively.

We model the mean relationship between the total line luminosity  $L_{s,\text{tot}}$  and the halo mass using the fiducial COMAP model [64],

$$\frac{L_{\text{CO}}}{L_{\odot}}(M) = 4.9 \times 10^{-5} \frac{C}{(M/M_{\star})^A + (M/M_{\star})^B}, \quad (32)$$

with  $A = -2.85$ ,  $B = -0.42$ ,  $C = 10^{10.63}$ , and  $M_{\star} = 10^{12.3} M_{\odot}$ , obtained from a fit to results from Universe Machine [65], COLDz [66] and COPSS [22]. We apply a

mean-preserving scatter  $\sigma_{\text{scat}} = 0.42$  dex in the expression above, introduced in practice in Eq. (19). We assume the halo mass function and mass-dependent linear halo bias from Refs. [67,68], respectively.

Probably the best estimation of the gas dynamics in a halo comes from the maximum circular velocity calculated by a halo finder in a simulation. However, there is no given relationship between this velocity and the halo mass. We choose to use the virial velocity as a good approximation, given by

$$v(M) = 50 \text{ km/s} \left( \frac{M}{10^{10} M_{\odot}} \right)^{1/3}, \quad (33)$$

to compute the line width with Eq. (28). We refer the interested reader to Ref. [62] for a comprehensive discussion on line broadening, covering the estimation of  $v(M)$  from observations, and the modeling of the line broadening in simulations and in the prediction of the power spectrum multipoles.

### C. Changes in the PDF

Intuitively, we can expect that extended signal profiles smooths the PDF: as each source contributes to more than one voxel, there are less very faint or very bright voxels, since the signal gets redistributed and homogenized. This scenario corresponds to a PDF with a smoother, smaller peak at low brightness temperatures and a smaller tail extending to high brightness temperatures. In this section we check this intuition by studying the effects of experimental resolution and line broadening in the PDF.

First, let us study the spatial distribution of the signal along the line of sight. We compare the profile along the line of sight under different assumptions in the top panel of Fig. 1. We include results for a point source, accounting only for the limited resolution of the experiment, and including line broadening with and without marginalizing over the inclination of the emitter. Limited resolution spreads the signal over three voxels,<sup>6</sup> but line broadening results in an even more extended profile for the most massive halos. Marginalizing over random inclinations results in a slightly more peaked profile at the center, but very similar tails.

For our example, only halos with masses  $\gtrsim 4 \times 10^{12} M_{\odot}$  have a line broadening larger than the line-spread function. However, as can be seen in the figure, accounting for the two effects together already results in wider profiles for lighter halos. This, of course, depends on the voxel size and the specific  $v(M)$  relation considered. We can quantify the relative contribution to the intensity map from the emitters that extend beyond the resolution limits with the luminosity-weighted distribution of  $\sigma_v$  (bottom panel of Fig. 1).

<sup>6</sup>Note that for the more standard voxel size corresponding to  $\delta\nu$  instead of  $\delta\nu^{\text{FWHM}}$  the signal extends over 5–7 voxels.

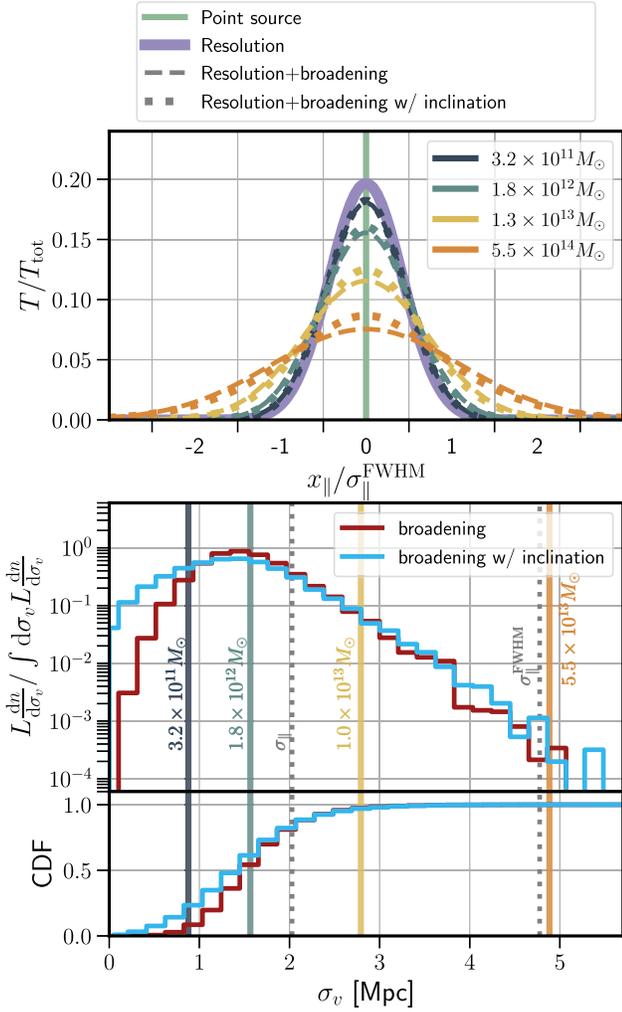


FIG. 1. Top panel: extended profiles in the direction along the line of sight for our example under different assumptions: we show a point source in light green, the effects of the limited resolution in purple, and adding line broadening with and without random inclination (dotted and dashed lines, respectively) with different colors for different halo masses. The grid in the horizontal axis corresponds to the voxel division. Bottom panel: luminosity-weighted distribution (top) and cumulative distribution (bottom) of the line broadening with and without random inclinations (blue and red, respectively). We also show the line broadening  $\sigma_v$  for edge-on emitters for the same masses considered in the top panel, the standard deviation  $\sigma_{\parallel}$  of the line-spread function and the voxel size  $\sigma_{\parallel}^{\text{FWHM}}$ .

The contribution of halos with emission lines broader than the line-spread function is small, because although they are the brightest emitters, their abundance is very small. We therefore expect a small effect in the VID with respect to considering only the experimental resolution. Finally, the random inclination of the emitters broadens the distribution of  $\sigma_v$ , but have a similar relative contribution to the line-intensity map (as seen with the cumulative distribution function).

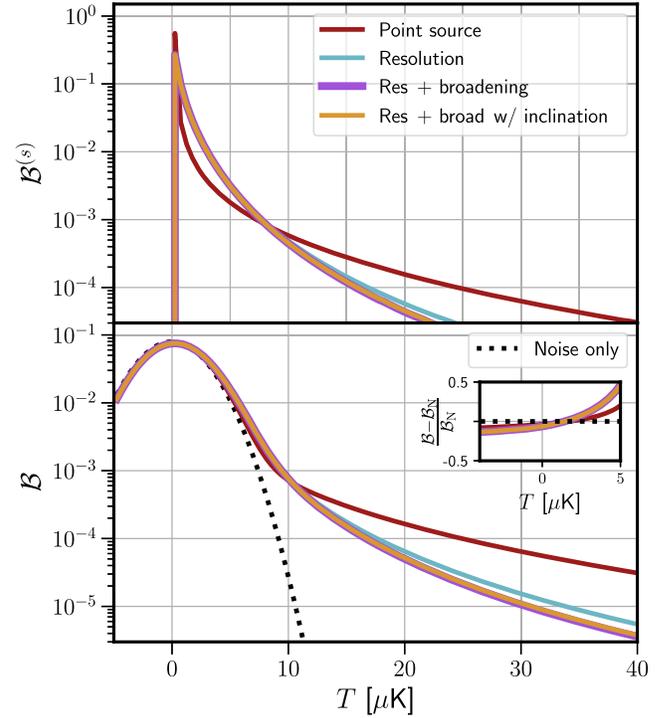


FIG. 2. VID of the signal (top) and including the contribution from the instrumental noise (bottom) for our example. We show the VID for point sources in red, including the effects from limited resolution in blue, adding line broadening for edge-on galaxies in magenta and marginalizing also over their inclination in orange. The inset enlarges the deviation over the noise-alone VID (black dotted).

We show the effects of the extended profiles in the PDF in Fig. 2, considering the same cases than in Fig. 1. We show the VID of the signal alone and including the noise, using 90 linearly spaced bins of  $0.5 \mu\text{K}$  width in the interval  $T \in [-5, 40] \mu\text{K}$ , as well as the VID corresponding to only noise. The effects induce a very significant change in the VID, following the intuition introduced at the beginning of the section. The extended emission narrows and smooths the VID, resulting in a lower tail towards higher brightness temperatures and a broader, lower peak at low brightness temperatures. Adding fractional contributions of several emitters to voxels which otherwise would be empty results in less very faint voxels. At the same time, distributing the signal of very bright voxels over their adjacent voxels imply that they are not that bright any more.

As one could expect from the results shown in Fig. 1, the impact of including the line broadening on top of the experimental resolution is small and limited only to the height of the tail at high temperatures. Marginalizing over the galaxy inclination has a negligible effect for this example, but it may have some impact for a different set up. Note that, although the VID at high brightness temperatures is significantly smaller after considering extended emission, we can expect better sensitivity to the signal of

interest, since there is a larger deviation from the noise-only signal at low brightness temperatures, where the statistical uncertainties are smaller. Figure 1 is a clear example that extended profiles must be taken into account when computing the LIM PDF.

The redistribution of signal to different voxels due to the extended profiles has another consequence for the interpretation of the VID. While the identification of bright voxels as those hosting more massive halos or brighter emitters is already nontrivial in the case of point sources (due to the unknown number of emitters in each voxel), the redistribution of signal further smears this connection. This results in a very broad distribution of halo masses without large differences between bright or faint voxels, as found in Ref. [69].

We show the dependence of the VID on clustering in Fig. 3. Higher matter clustering implies a higher clustering of galaxies and subsequently a larger variance in the brightness temperature fluctuations. Therefore, it increases the high-temperature tail of the VID, compensating at temperatures slightly above zero, as shown in the figure and in the inset, respectively. Since this effect can be degenerate with the  $L(M)$  relation, analyses must marginalize over  $\sigma_{\text{vox}}^2$  to avoid biased results.

Perhaps surprisingly, line broadening has a smaller impact on the VID than in the power spectrum multipoles (see Ref. [62]). This is because line broadening is more severe for brighter, more massive, but much less abundant halos. While they dominate the contributions to the shot

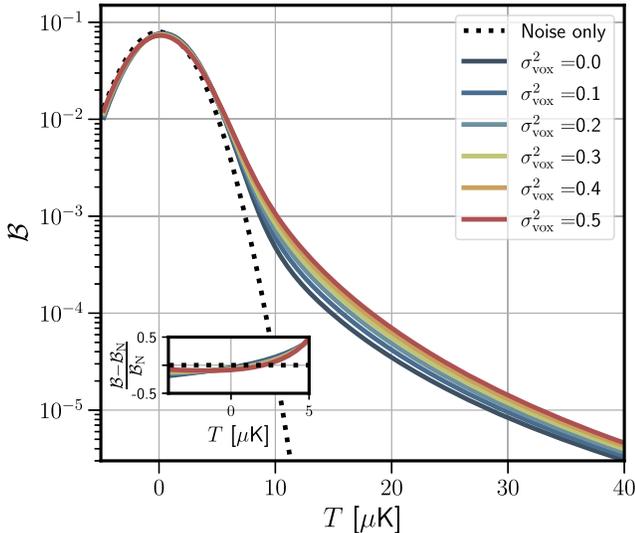


FIG. 3. VID including the contribution from the instrumental noise for our example and accounting for experimental resolution and the line broadening for randomly oriented galaxies. We show the prediction for several values of  $\sigma_{\text{vox}}^2$  with different colors, as indicated in the legend, and the noise-only VID with a black-dotted line. The inset in the bottom panel enlarges the deviation over the noise-alone VID.

noise of the power spectrum (which depends on the second moment of the luminosity function) and the linear power spectrum (since their fluctuations are more biased), they almost do not contribute to the overall luminosity (see Fig. 1). On top of that, the impact of the bias is smaller and very distributed over all masses in the VID [see Eq. (17)]. Furthermore, the quadrupole of the power spectrum traces the anisotropy of the clustering, hence it can be very affected by line broadening, while the VID is not sensitive to anisotropies.

#### IV. VALIDATION WITH SIMULATIONS

In this section we compare our predictions with simulations to test the accuracy of the modeling presented in Sec. II. We assume the astrophysical model and experimental set up described in Sec. III B.

We employ the publicly available code SIMPLE<sup>7</sup> [70], based on log-normal galaxy distribution simulations<sup>8</sup> [71], to compute the LIM log-normal simulations. We consider the luminosity-weighted average linear halo bias ( $b = 4.07$ ) corresponding to our model. We generate line-intensity maps for a volume of  $0.158 \text{ (Gpc}/h)^3$  at a single redshift  $z = 3$ , which corresponds to  $150^3$  cuboidlike voxels, defined as the full-width half maximum of the telescope beam and channel width. We run the simulations with a mesh three times finer (e.g.,  $450^3$  cells) in order to accurately capture the smoothing and the extended emission. We assign the intensity emitted by each galaxy to the mesh using the “nearest grid point” routine, apply as many filters as needed depending on the case to smooth the simulated line-intensity map, and then downsample to the voxel-size resolution using the same routine.<sup>9</sup>

For each realization, we consider four different cases: no smoothing applied (i.e., point sources); modeling the experimental resolution with an anisotropic Gaussian filter with standard deviations  $\sigma_{\parallel} = 1.37 \text{ Mpc}/h$  and  $\sigma_{\perp} = 1.62 \text{ Mpc}/h$  along and transverse to the line of sight, respectively; and including on top of it the line broadening with and without accounting for random inclinations.

We follow Ref. [62] to implement line broadening in the simulations. We first compute the line width  $\sigma_v$  for each galaxy. To do so, we modify SIMPLE to first assign a mass to each galaxy following the halo mass function. We can compute the line luminosity and line width associated with each mass using Eq. (32) with a mean-preserving

<sup>7</sup><https://github.com/mlujnie/simple>.

<sup>8</sup>[https://bitbucket.org/komatsu5147/lognormal\\_galaxies/src/master/](https://bitbucket.org/komatsu5147/lognormal_galaxies/src/master/).

<sup>9</sup>While “nearest grid point” introduces aliases and ringing for power spectrum measurements (see Ref. [72] for a thorough study), we have tested that higher-order mass-assignment kernels such as “clouds in cell” introduce spurious effects in the VID, even after compensating their effect on the map by deconvoluting the kernel. The finer grid also prevents from resolution-dependent artifacts related with the Fourier transform in the final map.

logarithmic scatter  $\sigma_{\text{scat}}$ , and Eqs. (33) and (28), respectively. Random galaxy inclinations are accounted for multiplying the edge-on  $\sigma_v$  by a factor  $2\sqrt{(1-\mu_\varphi^2)}/3$ , where  $\mu_\varphi$  is sampled from a uniform distribution within  $[-1, 1]$ . We bin galaxies in the simulation by  $\sigma_v$  and compute an intensity-weighted average  $\bar{\sigma}_v^i$  for each bin. Afterwards, we add the intensity from the galaxies within each bin to empty meshes and apply a Gaussian filter with  $\sigma_\perp = 0$  and  $\sigma_\parallel = \bar{\sigma}_v^i$ . We add all meshes and apply the resolution filter described above.

In all cases, we remove the mean from the maps, measure the VID for  $\Delta T$ , and compute their average

$$\bar{\mathcal{B}}_a = \frac{1}{N_r} \sum_i^{N_r} \mathcal{B}_a^{(i)} \quad (34)$$

and covariance

$$C_{ab} = \frac{1}{N_r} \sum_i^{N_r} \left\{ \left( \mathcal{B}_a^{(i)} - \bar{\mathcal{B}}_a \right) \left( \mathcal{B}_b^{(i)} - \bar{\mathcal{B}}_b \right) \right\} \quad (35)$$

for the  $\Delta T_a$  and  $\Delta T_b$  bins from  $N_r = 1008$  independent realizations. Note that we have defined  $\mathcal{B}_a \equiv \mathcal{B}(\Delta T_a)$ .

By definition, the galaxy bias in the log-normal simulations used in SIMPLE is linear and the same for all galaxies. We therefore assume  $b(M) = 4.07$  for the comparison between the theoretical prediction and the measured  $\bar{\mathcal{B}}$  from the log-normal simulations. Note that for this validation we use  $\sigma_{\text{vox}}^2 = 0.24$  computed from Eq. (22) using also Eq. (23), instead of fitting it to the measurements (see Fig. 3 for the effect of  $\sigma_{\text{vox}}^2$  in the VID).

We compare theoretical predictions and measurements from the simulations in Fig. 4, including  $\sqrt{C_{aa}}$  as a reference for the uncertainties. The model provides a good description of the mean measurements from the log-normal realizations, successfully capturing the effects of the smoothing.

There is a small absolute deviation between the prediction and the simulations at intermediate scales, present in all four cases. We show the relative difference between the theoretical prediction and the measurement from simulations in Fig. 5. Regarding the clustering, we consider the log-normal simulations introduced above, but also Poisson-sampled realizations without any clustering of emitters; regarding the intensity profiles, we consider point sources and extended profiles with effects from resolution and line broadening marginalizing over the inclination. The mean and covariance for the Poisson realizations are obtained from 100 independent realizations. Here we can appreciate that although our model qualitatively captures the effects of extended profiles, it does not provide an accurate prediction of the VID. The inaccuracy is most severe at intermediate brightness temperatures, where the relative deviation is maximum and the covariance is smaller.

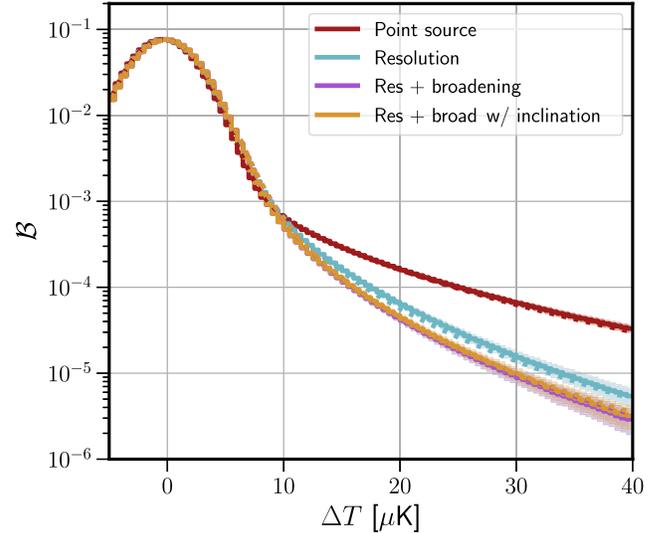


FIG. 4. Mean measured VID from 1008 independent log-normal realizations including instrumental noise (solid step lines) and the prediction using the formalism described in this work (dotted lines). Shaded regions correspond to the square root of the diagonal of the covariance. We show results for point sources in red, including the effects from resolution in blue, adding line broadening in purple and marginalizing over inclinations in orange.

We associate this deviation to two potential sources of error. First, we can see that even for point sources clustering is not properly captured. This is because the modeling of clustering is limited. Truncating the moment-generating function at second order removes any impact of non Gaussianity of the halo distribution in the PDF.<sup>10</sup> This could be captured including higher terms in the expansion of the cumulants of clustering in Eq. (15) [and subsequently in Eq. (17)]. Note that neglecting nonlinear bias is most likely a bigger issue, but our log-normal simulations assume linear bias by definition. Second, there is a small deviation introduced by the extended profiles even in the absence of clustering. This may be due to limitations of our model or to potential numerical artifacts in the implementation of the line broadening in the simulated line-intensity maps. While the impact of artifacts related with resolution and filters in the power spectrum have been thoroughly investigated, there is significantly less studies on their effects on the histogram of the map. Furthermore, these deviations may not depend trivially on the experimental setup and ratio between  $\sigma_\parallel$  and  $\sigma_v$ . We leave a dedicated study with a deeper study of the numerical artifacts for future work.

<sup>10</sup>Log-normal simulations exhibit a higher degree of non-Gaussianity in the galaxy distribution than  $N$ -body simulations (see, e.g., [73]). Therefore, deviations from the measurements in the log-normal realizations related to the truncation of the moment expansion are expected to be larger than in a more realistic case.

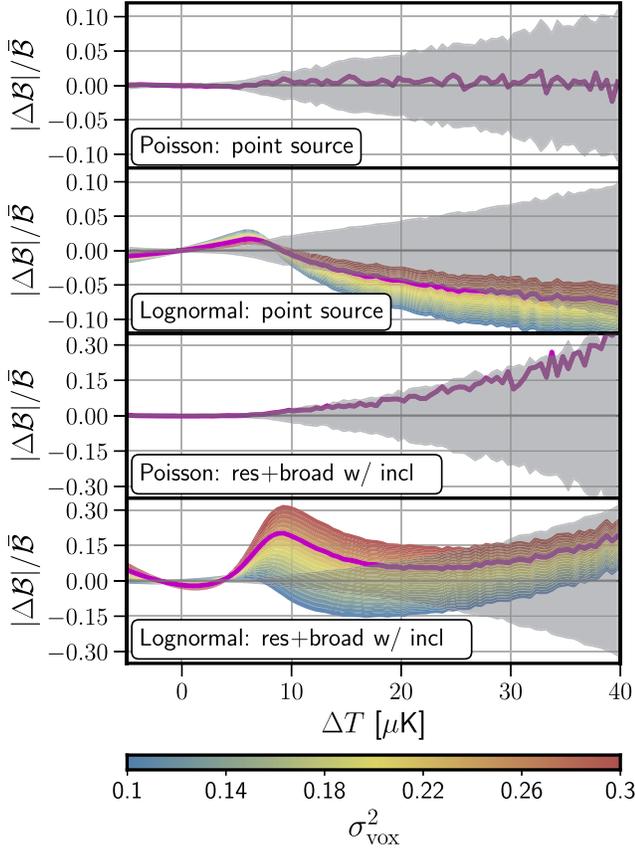


FIG. 5. Relative deviation between the theoretical prediction and the mean from Poisson realizations and log-normal simulations, considering point sources and extended profiles as indicated by the titles of the panels. For the cases with clustering (log-normal), we vary  $\sigma_{\text{vox}}^2$ , highlighting the fiducial value used throughout the rest of the study in magenta. Shaded gray areas show the relative uncertainties related with the diagonal of the covariance matrix. Note the change of scale in the y axis between the two top and the two bottom panels.

Finally, we study the impact of the extended profiles in the covariance matrix of the VID. An analytic estimation of the covariance can be obtained from the 2-point PDF, following the derivation described in Ref. [53] and adapting it to LIM. However, the covariance of the VID is usually assumed to be that of a multinomial distribution, since the VID is a histogram. A multinomial distribution presents a covariance

$$\mathcal{C}_{ab}^{\text{multinom}} = \begin{cases} \frac{\mathcal{B}_a}{N_{\text{vox}}} (1 - \mathcal{B}_a) & \text{if } a = b \\ -\frac{\mathcal{B}_a \mathcal{B}_b}{N_{\text{vox}}} & \text{if } a \neq b \end{cases}, \quad (36)$$

so that its correlation matrix, defined in general as  $\mathcal{R}_{ab} \equiv \mathcal{C}_{ab} / \sqrt{\mathcal{C}_{aa} \mathcal{C}_{bb}}$ , has off-diagonal terms given by  $\mathcal{R}_{ab}^{\text{multinom}} = -\sqrt{\mathcal{B}_a \mathcal{B}_b} / [(1 - \mathcal{B}_a)(1 - \mathcal{B}_b)]$ .

The diagonal of the numerical covariance computed using Eq. (35) agrees very well with the multinomial

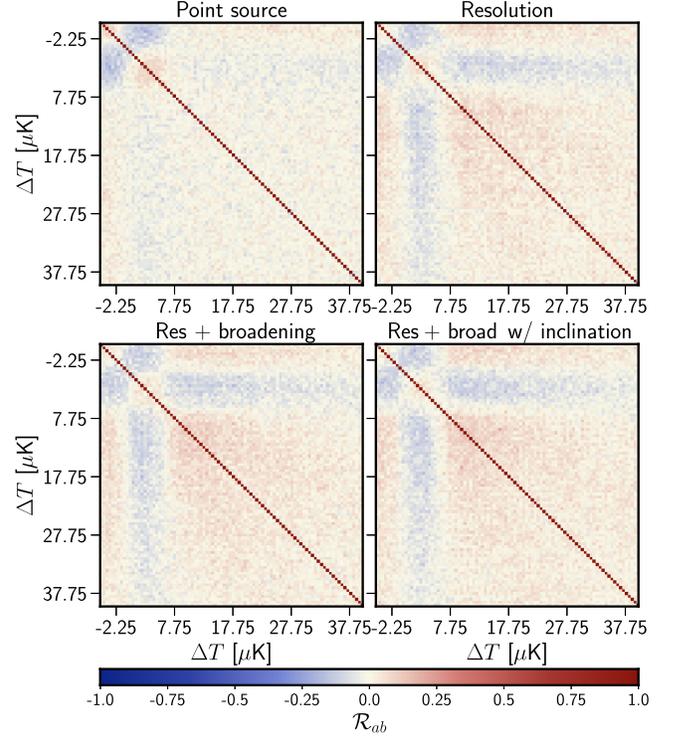


FIG. 6. Correlation matrices computed numerically from 1008 log-normal realizations. We consider four cases as indicated by the titles at the top of each panel.

variance for all cases, but the off-diagonal terms do not. We show the numerical correlation matrices computed from the log-normal realizations for the four cases in Fig. 6. We do not find the strong negative correlations expected for the multinomial distribution. On the contrary, we find negligible correlation for point sources except for  $\Delta T \lesssim 5 \mu\text{K}$ , which is dominated by the instrumental noise. Once emitters are allowed to contribute to more than one voxel, the off-diagonal correlation grows to  $\lesssim 0.1$ ; it is slightly higher once line broadening is considered. The deviation from the multinomial distribution off-diagonal covariance is due to physical effects and sources contributing to several voxels inducing correlations beyond the multinomial distribution. Note that the correlation would be larger for smaller voxels, as demonstrated in Ref. [63].

The covariance measured from the simulations does not include the supersample contributions. These contributions, derived for a previous formalism of the VID in Ref. [41], have been found to be generally smaller than statistical covariance from Poisson sampling of a histogram. Similarly, the off-diagonal physical covariance is also expected to be larger than the supersample covariance.

## V. PARAMETER-INFERENCE BIAS

We have shown that the VID accounting for extended emission is very different that the standard computation for

point sources. Therefore, an analysis assuming point sources is prone to obtain highly biased results.

We consider an analysis of the VID with the same experimental set up and model as in previous sections. We focus on the  $L(M)$  relation and take  $C$ ,  $A$ ,  $B$  and  $M_\star$  as free parameters [see Eq. (32)]. We also include  $\sigma_{\text{vox}}^2$  as nuisance parameter. Assuming Gaussian errors, the logarithm of the likelihood for the VID at position  $\mathbf{p}$  in the parameter space is

$$-2 \log \mathcal{L}(\mathbf{p}) = \sum_{ab} \Delta \mathcal{B}_a(\mathbf{p}) (C^{-1})_{ab} \Delta \mathcal{B}_b(\mathbf{p}), \quad (37)$$

where  $\Delta \mathcal{B}_a$  is the difference between the measured VID and the prediction. We use the covariance matrix numerically estimated in Sec. IV. However, the inverse of a covariance matrix estimated from a finite number of realizations is not unbiased; we apply the correction factor described in Ref. [74] to the inverse of the covariance matrix. In our case, the correction amounts to a  $\sim 10\%$  factor.

Let us assume, for this exercise, that the data is perfectly described by our modeling of the VID including the extended signal caused by experimental resolution and line broadening for random galaxy inclinations for a set of *true* parameter values  $\mathbf{p}_{\text{fid}}$ . An analysis for this model would then return best-fit parameter values  $\mathbf{p}_{\text{bf}}^{\text{sm}} = \mathbf{p}_{\text{fid}}$  and  $\Delta \mathcal{B}_a(\mathbf{p}_{\text{fid}}) = 0$ .

For an incorrect modeling—assuming point sources in our case—the best-fit  $\mathbf{p}_{\text{bf}}^{\text{ps}}$  inferred will be biased. This bias can be estimated using a Fisher-matrix analysis, through a linear approximation of the likelihood. This is not a good approximation for our case, since we expect large deviations in  $\mathcal{B}$  for the incorrect model (see Fig. 2). However, this exercise provides a qualitative estimate of the impact in parameter inference.

We denote the predictions for the two models with  $\mathcal{B}^{\text{sm}}$  and  $\mathcal{B}^{\text{ps}}$ , respectively, standing for “smoothed” and “point sources.” The estimated systematic bias introduced due to the poor modeling is [75]

$$\begin{aligned} \Delta \mathbf{p}_{\text{sys}} &\equiv \mathbf{p}_{\text{bf}}^{\text{ps}} - \mathbf{p}_{\text{fid}} \\ &= F_{\text{ps}}^{-1} \sum_{ab} \nabla_p \mathcal{B}_a^{\text{ps}}|_{\mathbf{p}_{\text{fid}}} (C_{\text{ps}}^{-1})_{ab} [\mathcal{B}_b^{\text{sm}}(\mathbf{p}_{\text{fid}}) - \mathcal{B}_b^{\text{ps}}(\mathbf{p}_{\text{fid}})], \end{aligned} \quad (38)$$

where

$$F = \sum_{ab} \nabla_p \mathcal{B}_a|_{\mathbf{p}_{\text{fid}}} (C^{-1})_{ab} (\nabla_p \mathcal{B}_b|_{\mathbf{p}_{\text{fid}}}) \quad (39)$$

is the Fisher matrix computed for the case of interest, and  $\nabla_p \mathcal{B}_a^{\text{ps}}|_{\mathbf{p}_{\text{fid}}}$  is the gradient in parameter space of the VID evaluated at  $\mathbf{p}_{\text{fid}}$ .

We show marginalized forecasted uncertainties for the two models including the estimation for the systematic bias

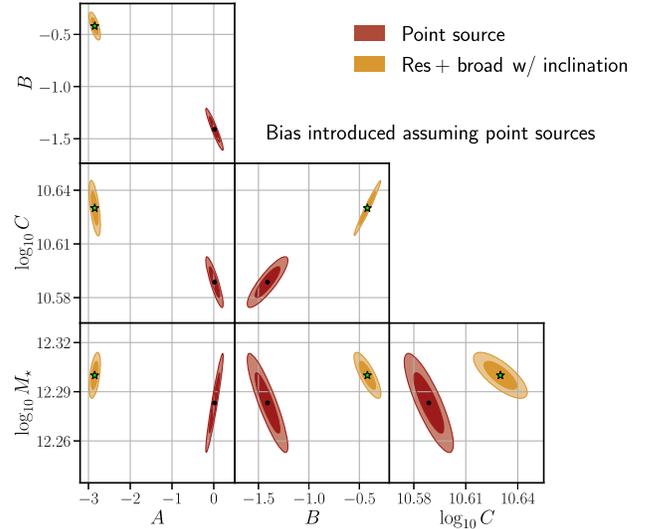


FIG. 7. 68% and 95% confidence level forecasted marginalized constraints on the parameters controlling the  $L(M)$  relation for the experimental set up described in Secs. III B and IV. We compare the results for point sources (red) and modeling the limited resolution of the experiment and line broadening marginalizing over the emitter inclination (orange), assuming that the latter is the correct description of the data. *True* parameter values are marked by a star, while the best fit assuming point sources is marked by a black dot.

in Fig. 7. All parameters present big systematic biases, except for  $\log_{10} M_\star$ , for which the marginalized one-dimensional bias is  $\sim 1.8\sigma$ . In addition, note that the constraining power and the degeneracies obtained if point sources are assumed are also different. Incomplete modeling may affect the conclusions obtained from a Fisher forecasts, which motivates the use of more accurate models even for qualitative forecast sensitivity [76].

## VI. CONCLUSIONS

Line-intensity mapping proposes an alternative tracer for the large-scale structure of the Universe, providing unprecedented sensitivity to high redshifts and faint emitters. Contrary to other tracers, LIM is also sensitive to astrophysical processes driving galaxy formation and evolution. This additional dependence adds another layer of complexity to the line-intensity fluctuation maps, which become very non Gaussian. Therefore, LIM power spectra can only capture a fraction of the information encoded in line-intensity maps. Furthermore, the power spectrum is affected by degeneracies between cosmology and astrophysics. These are the main motivations to explore one-point statistics, since they probe the whole distribution of intensity fluctuations and are more sensitive to features in the line-luminosity function.

So far, the modeling of LIM 1-point statistics has relied on an unphysical assumption: that each emitter contributes

just to the voxel that contains it. This assumption implies that experiments have perfect resolution and that all emitters are point sources with negligible line width.

In this work we have dropped this assumption and implemented for the first time the contribution of an emitter to several voxels in the LIM PDF theoretical prediction. We build upon previous work which use the halo model [51–53], noting the additional complication that LIM implies with respect to projected angular fields. We circumvent this complication invoking the ergodic hypothesis and integrating over the spatial distribution of the intensity profile when computing the intensity PDF of a single emitter.

We compare four different cases: point sources, adding the effects of limited resolution, adding on top it line broadening assuming that all galaxies are perfectly edge-on, and marginalizing over their inclination. Extended profiles significantly alter the LIM PDF prediction, with changes of more than one order of magnitudes. The resulting PDF has a broader, lower peak at low intensities and a lower, shorter tail towards high intensities. This follows our intuition: if an emitter contributes to several voxels, there will be fewer empty or very faint voxels and also less very bright voxels, respectively. We find that, for the example considered, the dominant effect is related to the limited resolution, while including the effects of line broadening adds a small modification. Since line broadening only acts in the direction along the line of sight and is more relevant for massive, rare emitters, it has a smaller effect on the VID than the experimental resolution, which acts in the three directions and affects all emitters equally. The qualitative behaviour agrees with our expectations, and although the LIM PDF depends a lot on the specific astrophysical model and experimental setup, we expect these results to be generally applicable.

We validate, for the first time, the theoretical prediction of the VID against simulations. We compare our implementation with log-normal simulations and find good qualitative agreement. Nonetheless, we still find deviations that are related with the clustering of emitters, as well as potential numerical artifacts from the implementation of line broadening in the simulations or limitations in our model to include extended profiles in the theoretical prediction. We leave further development of the modeling of clustering for the LIM PDF and a deeper study of the implementation of extended profiles in simulations and its impact on the measured VID for future work. Alternative frameworks like the theory of large deviations, which is being developed for the PDF of cosmic magnification (see e.g. Ref. [77] and references therein), provide a better prediction for the matter clustering PDF. Therefore, a

combination of both approaches, for instance combining the LIM PDF for unclustered sources with a PDF for the clustering of such sources, might lead to a more accurate prediction.

We also study the effects of extended profiles in the VID covariance. First, we see that while the diagonal of the covariance follows closely the variance for a multinomial distribution, as expected from a histogram, the off-diagonal terms significantly deviate. We find negligible correlation for point sources, instead of the strong negative correlations expected for a multinomial distribution. Second, we find that extended profiles increase the correlation between intensity bins in the VID, reaching correlations of around 10%. Finally, we demonstrate that very large systematic biases in parameter inference can be expected if pointlike intensity profiles are assumed.

One-point statistics, especially through its complementarity with the power spectrum, holds promise to break parameter degeneracies and significantly increase the constraining power of LIM surveys. In the case observational contaminants preclude the use of the VID, there are alternatives, combining different LIM observations or line-intensity maps and galaxy catalogs, which are more robust to observational systematics. However, since these alternative summary statistics still depend on the intensity PDF, they must account for the profiles of the observed signal from each emitter.

This work is but one of the first steps towards an accurate and efficient modeling for the LIM PDF. Further research is required to achieve a successful application to observations. Nonetheless, the promise of LIM one-point summary statistics motivates this effort, since only through their combination with LIM power spectra and other observations LIM surveys will unlock their full potential.

## ACKNOWLEDGMENTS

The author thanks Nickolas Kokron and Gabriela Sato-Polito for key discussions during the development of this work, and especially Patrick Breyse for important discussions and also sharing the implementation of the LIM PDF from his previous work. The author acknowledges funding from the Ramón y Cajal Grant No. RYC2021-033191-I, financed by Grant No. MCIN/AEI/10.13039/501100011033 and by the European Union “NextGenerationEU”/PRTR, and the computer resources provided by the Spanish Supercomputing Network (RES) node at Universidad de Cantabria and the Institute of Physics of Cantabria (IFCA-CSIC).

- [1] E. D. Kovetz *et al.*, Line-intensity mapping: 2017 status report, [arXiv:1709.09066](#).
- [2] J. R. Pritchard and A. Loeb, 21 cm cosmology in the 21st century, *Rep. Prog. Phys.* **75**, 086901 (2012).
- [3] A. Liu and J. R. Shaw, Data analysis for precision 21 cm cosmology, *Publ. Astron. Soc. Pac.* **132**, 062001 (2020).
- [4] J. L. Bernal and E. D. Kovetz, Line-intensity mapping: Theory review with a focus on star-formation lines, *Astron. Astrophys. Rev.* **30**, 5 (2022).
- [5] Y.-T. Cheng, R. de Putter, T.-C. Chang, and O. Dore, Optimally mapping large-scale structures with luminous sources, *Astrophys. J.* **877**, 86 (2019).
- [6] E. Schaan and M. White, Astrophysics & cosmology from line intensity mapping vs galaxy surveys, *J. Cosmol. Astropart. Phys.* **05** (2021) 067.
- [7] M. P. van Haarlem *et al.*, LOFAR: The LOw-Frequency ARray, *Astron. Astrophys.* **556**, A2 (2013).
- [8] K. Bandura *et al.*, Canadian hydrogen intensity mapping experiment (CHIME) pathfinder, *Proc. SPIE Int. Soc. Opt. Eng.* **9145**, 22 (2014).
- [9] D. R. DeBoer *et al.*, Hydrogen epoch of reionization array (HERA), *Publ. Astron. Soc. Pac.* **129**, 045001 (2017).
- [10] M. G. Santos *et al.* (MeerKLASS Collaboration), MeerKLASS: MeerKAT large area synoptic survey, in MeerKAT Science: On the Pathway to the SKA (2017), [arXiv:1709.06099](#).
- [11] K. A. Cleary *et al.*, COMAP early science: I. Overview, [arXiv:2111.05927](#).
- [12] P. Ade *et al.* (CONCERTO Collaboration), A wide field-of-view low-resolution spectrometer at APEX: Instrument design and scientific forecast, *Astron. Astrophys.* **642**, A60 (2020).
- [13] K. Gebhardt *et al.*, The Hobby–Eberly telescope dark energy experiment (HETDEX) survey design, reductions, and detections\*, *Astrophys. J.* **923**, 217 (2021).
- [14] M. Aravena *et al.* (CCAT-Prime Collaboration), CCAT-prime Collaboration: Science goals and forecasts with prime-cam on the fred Young submillimeter telescope, *Astrophys. J. Suppl. Ser.* **264**, 7 (2023).
- [15] G. Sun *et al.*, Probing cosmic reionization and molecular gas growth with TIME, *Astrophys. J.* **915**, 33 (2021).
- [16] E. R. Switzer *et al.*, Experiment for cryogenic large-aperture intensity mapping: Instrument design, *J. Astron. Telesc. Instrum. Syst.* **7**, 044004 (2021).
- [17] J. Vieira, J. Aguirre, C. M. Bradford, J. Filippini, C. Groppi, D. Marrone *et al.*, The terahertz intensity mapper (TIM): A next-generation experiment for galaxy evolution studies, [arXiv:2009.14340](#).
- [18] O. Doré *et al.*, Cosmology with the SPHEREX all-sky spectral survey, [arXiv:1412.4872](#).
- [19] L. V. E. Koopmans *et al.*, The cosmic dawn and epoch of reionization with the square kilometre array, *Proc. Sci. AASKA14* (2015) 001 [[arXiv:1505.07568](#)].
- [20] L. B. Newburgh *et al.*, HIRAX: A probe of dark energy and radio transients, *Proc. SPIE Int. Soc. Opt. Eng.* **9906**, 99065X (2016).
- [21] T.-C. Chang, U.-L. Pen, J. B. Peterson, and P. McDonald, Baryon acoustic oscillation intensity mapping as a test of dark energy, *Phys. Rev. Lett.* **100**, 091303 (2008).
- [22] G. K. Keating, D. P. Marrone, G. C. Bower, E. Leitch, J. E. Carlstrom, and D. R. DeBoer, COPSS II: The molecular gas content of ten million cubic megaparsecs at redshift  $z \sim 3$ , *Astrophys. J.* **830**, 34 (2016).
- [23] S. Yang, A. R. Pullen, and E. R. Switzer, Evidence for C II diffuse line emission at redshift  $z \sim 2.6$ , *Mon. Not. R. Astron. Soc.* **489**, L53 (2019).
- [24] G. K. Keating, D. P. Marrone, G. C. Bower, and R. P. Keenan, An intensity mapping detection of aggregate CO line emission at 3 mm, *Astrophys. J.* **901**, 141 (2020).
- [25] L. Wolz *et al.*, H i constraints from the cross-correlation of eBOSS galaxies and Green Bank Telescope intensity maps, *Mon. Not. R. Astron. Soc.* **510**, 3495 (2022).
- [26] M. L. Niemeyer *et al.*, Surface brightness profile of Lyman- $\alpha$  halos out to 320 kpc in HETDEX, *Astrophys. J.* **929**, 90 (2022).
- [27] S. Cunnington *et al.*, H i intensity mapping with MeerKAT: Power spectrum detection in cross-correlation with WiggleZ galaxies, *Mon. Not. R. Astron. Soc.* **518**, 6262 (2022).
- [28] M. L. Niemeyer *et al.*, Ly $\alpha$  Halos around [O iii]-selected Galaxies in HETDEX, *Astrophys. J. Lett.* **934**, L26 (2022).
- [29] S. Paul, M. G. Santos, Z. Chen, and L. Wolz, A first detection of neutral hydrogen intensity mapping on Mpc scales at  $z \approx 0.32$  and  $z \approx 0.44$ , [arXiv:2301.11943](#).
- [30] A. Moradinezhad Dizgah and G. K. Keating, Line intensity mapping with [CII] and CO(1-0) as probes of primordial non-Gaussianity, *Astrophys. J.* **872**, 126 (2019).
- [31] J. L. Bernal, P. C. Breyse, and E. D. Kovetz, Cosmic expansion history from line-intensity mapping, *Phys. Rev. Lett.* **123**, 251301 (2019).
- [32] G. Sato-Polito, J. L. Bernal, K. K. Boddy, and M. Kamionkowski, Kinetic Sunyaev-Zel’dovich tomography with line-intensity mapping, *Phys. Rev. D* **103**, 083519 (2021).
- [33] A. Moradinezhad Dizgah, G. K. Keating, K. S. Karkare, A. Crites, and S. R. Choudhury, Neutrino properties with ground-based millimeter-wavelength line intensity mapping, *Astrophys. J.* **926**, 137 (2022).
- [34] E. Castorina and M. White, Measuring the growth of structure with intensity mapping surveys, *J. Cosmol. Astropart. Phys.* **06** (2019) 025.
- [35] A. Moradinezhad Dizgah, F. Nikakhtar, G. K. Keating, and E. Castorina, Precision tests of CO and CII power spectra models against simulated intensity maps, *J. Cosmol. Astropart. Phys.* **02** (2022) 026.
- [36] J. L. Bernal, P. C. Breyse, H. Gil-Marín, and E. D. Kovetz, User’s guide to extracting cosmological information from line-intensity maps, *Phys. Rev. D* **100**, 123522 (2019).
- [37] S. Camera and H. Padmanabhan, Beyond  $\Lambda$ CDM with HI intensity mapping: Robustness of cosmological constraints in the presence of astrophysics, *Mon. Not. R. Astron. Soc.* **496**, 4115 (2020).
- [38] P. C. Breyse, E. D. Kovetz, and M. Kamionkowski, The high redshift star-formation history from carbon-monoxide intensity maps, *Mon. Not. R. Astron. Soc.* **457**, L127 (2016).
- [39] P. C. Breyse, E. D. Kovetz, P. S. Behroozi, L. Dai, and M. Kamionkowski, Insights from probability distribution functions of intensity maps, *Mon. Not. R. Astron. Soc.* **467**, 2996 (2017).

- [40] H. T. Ihle *et al.* (COMAP Collaboration), Joint power spectrum and voxel intensity distribution forecast on the CO luminosity function with COMAP, *Astrophys. J.* **871**, 75 (2019).
- [41] G. Sato-Polito and J. L. Bernal, Analytical covariance between the voxel intensity distribution and the power spectrum of line-intensity maps, *Phys. Rev. D* **106**, 103534 (2022).
- [42] J. L. Bernal, A. Caputo, and M. Kamionkowski, Strategies to detect dark-matter decays with line-intensity mapping, *Phys. Rev. D* **103**, 063523 (2021); **105**, 089901(E) (2022).
- [43] J. L. Bernal, A. Caputo, F. Villaescusa-Navarro, and M. Kamionkowski, Searching for the radiative decay of the cosmic neutrino background with line-intensity mapping, *Phys. Rev. Lett.* **127**, 131102 (2021).
- [44] J. L. Bernal, G. Sato-Polito, and M. Kamionkowski, Cosmic optical background excess, dark matter, and line-intensity mapping, *Phys. Rev. Lett.* **129**, 231301 (2022).
- [45] S. Libanore, C. Unal, D. Sarkar, and E. D. Kovetz, Unveiling cosmological information on small scales with line intensity mapping, *Phys. Rev. D* **106**, 123512 (2022).
- [46] T. Adi, S. Libanore, H. A. G. Cruz, and E. D. Kovetz, Constraining primordial magnetic fields with line-intensity mapping, *J. Cosmol. Astropart. Phys.* **09** (2023) 035.
- [47] P. A. G. Scheuer, A statistical method for analysing observations of faint radio stars, *Proc. Cambridge Philos. Soc.* **53**, 764 (1957).
- [48] X. Barcons, Confusion noise and source clustering, *Astrophys. J.* **396**, 460 (1992).
- [49] S. K. Lee, S. Ando, and M. Kamionkowski, The gamma-ray-flux probability distribution function from galactic halo substructure, *J. Cosmol. Astropart. Phys.* **07** (2009) 007.
- [50] G. Patanchon, P. A. R. Ade, J. J. Bock, E. L. Chapin, M. J. Devlin, S. R. Dicker *et al.*, Submillimeter number counts from statistical analysis of BLAST maps, *Astrophys. J.* **707**, 1750 (2009).
- [51] P. C. Breyse, Breaking the intensity-bias degeneracy in line intensity mapping, [arXiv:2209.01223](https://arxiv.org/abs/2209.01223).
- [52] L. Thiele, J. C. Hill, and K. M. Smith, Accurate analytic model for the thermal Sunyaev-Zel'dovich one-point probability distribution function, *Phys. Rev. D* **99**, 103511 (2019).
- [53] L. Thiele, J. C. Hill, and K. M. Smith, Accurate analytic model for the weak lensing convergence one-point probability distribution function and its autocovariance, *Phys. Rev. D* **102**, 123545 (2020).
- [54] P. C. Breyse, C. J. Anderson, and P. Berger, Canceling out intensity mapping foregrounds, *Phys. Rev. Lett.* **123**, 231105 (2019).
- [55] P. C. Breyse, D. T. Chung, and H. T. Ihle, Characteristic functions for cosmological cross-correlations, *Mon. Not. R. Astron. Soc.* **525**, 1824 (2023).
- [56] D. T. Chung, I. Bangari, P. C. Breyse, H. T. Ihle, J. R. Bond, D. A. Dunne *et al.* (COMAP Collaboration), The deconvolved distribution estimator: Enhancing reionization-era CO line-intensity mapping analyses with a cross-correlation analogue for one-point statistics, *Mon. Not. R. Astron. Soc.* **520**, 5305 (2023).
- [57] N. Aghanim *et al.* (Planck Collaboration), Planck 2018 results. VI. Cosmological parameters, *Astron. Astrophys.* **641**, A6 (2020); **652**, C4(E) (2021).
- [58] A. H. Barnett, J. Magland, and L. af Klinteberg, A parallel nonuniform fast Fourier transform library based on an “exponential of semicircle” kernel, *SIAM J. Sci. Comput.* **41**, C479 (2019).
- [59] A. H. Barnett, Aliasing error of the  $\exp(\beta\sqrt{1-z^2})$  kernel in the nonuniform fast Fourier transform, [arXiv:2001.09405](https://arxiv.org/abs/2001.09405).
- [60] Y. Zhang, A. R. Pullen, R. S. Somerville, P. C. Breyse, J. C. Forbes, S. Yang, Y. Li, and A. S. Maniyar, Characterizing the conditional galaxy property distribution using Gaussian mixture models, *Astrophys. J.* **950**, 159 (2023).
- [61] A. Mead, S. Brieden, T. Tröster, and C. Heymans, HMcode-2020: Improved modelling of non-linear cosmological power spectra with baryonic feedback, *Mon. Not. R. Astron. Soc.* **502**, 1401 (2021).
- [62] D. T. Chung *et al.* (COMAP Collaboration), A model of spectral line broadening in signal forecasts for line-intensity mapping experiments, *Astrophys. J.* **923**, 188 (2021).
- [63] T. Vernstrom, D. Scott, J. V. Wall, J. J. Condon, W. D. Cotton, E. B. Fomalont, K. I. Kellermann, N. Miller, and R. A. Perley, Deep 3 GHz number counts from a P(D) fluctuation analysis, *Mon. Not. R. Astron. Soc.* **440**, 2791 (2014).
- [64] D. T. Chung *et al.* (COMAP Collaboration), COMAP early science. V. Constraints and forecasts at  $z \sim 3$ , *Astrophys. J.* **933**, 186 (2022).
- [65] P. Behroozi, R. H. Wechsler, A. P. Hearin, and C. Conroy, UniverseMachine: The correlation between galaxy growth and dark matter halo assembly from  $z = 0-10$ , *Mon. Not. R. Astron. Soc.* **488**, 3143 (2019).
- [66] D. A. Riechers *et al.*, COLDz: Shape of the CO luminosity function at high redshift and the cold gas history of the universe, *Astrophys. J.* **872**, 7 (2019).
- [67] J. L. Tinker, A. V. Kravtsov, A. Klypin, K. Abazajian, M. S. Warren, G. Yepes, S. Gottlöber, and D. E. Holz, Toward a halo mass function for precision cosmology: The Limits of universality, *Astrophys. J.* **688**, 709 (2008).
- [68] J. L. Tinker, B. E. Robertson, A. V. Kravtsov, A. Klypin, M. S. Warren, G. Yepes, and S. Gottlöber, The large scale bias of dark matter halos: Numerical calibration and model tests, *Astrophys. J.* **724**, 878 (2010).
- [69] G. Sato-Polito, N. Kokron, and J. L. Bernal, A multi-tracer empirically-driven approach to line-intensity mapping light-cones, [arXiv:2212.08056](https://arxiv.org/abs/2212.08056).
- [70] M. L. Niemeyer, J. L. Bernal, and E. Komatsu, SIMPLE: Simple intensity map producer for line emission, [arXiv:2307.08475](https://arxiv.org/abs/2307.08475).
- [71] A. Agrawal, R. Makiya, C.-T. Chiang, D. Jeong, S. Saito, and E. Komatsu, Generating log-normal mock catalog of galaxies in redshift space, *J. Cosmol. Astropart. Phys.* **10** (2017) 003.
- [72] E. Sefusatti, M. Crocce, R. Scoccimarro, and H. Couchman, Accurate estimators of correlation functions in Fourier space, *Mon. Not. R. Astron. Soc.* **460**, 3624 (2016).
- [73] L. Blot *et al.*, Comparing approximate methods for mock catalogues and covariance matrices II: Power spectrum multipoles, *Mon. Not. R. Astron. Soc.* **485**, 2806 (2019).

- [74] J. Hartlap, P. Simon, and P. Schneider, Why your model parameter confidences might be too optimistic: Unbiased estimation of the inverse covariance matrix, *Astron. Astrophys.* **464**, 399 (2007).
- [75] J.L. Bernal, N. Bellomo, A. Raccanelli, and L. Verde, Beware of commonly used approximations. Part II. Estimating systematic biases in the best-fit parameters, *J. Cosmol. Astropart. Phys.* **10** (2020) 017.
- [76] N. Bellomo, J.L. Bernal, G. Scelfo, A. Raccanelli, and L. Verde, Beware of commonly used approximations. Part I. Errors in forecasts, *J. Cosmol. Astropart. Phys.* **10** (2020) 016.
- [77] A. Barthelemy, A. Halder, Z. Gong, and C. Uhlemann, Making the leap I: Modelling the reconstructed lensing convergence PDF from cosmic shear with survey masks and systematics, [arXiv:2307.09468](https://arxiv.org/abs/2307.09468).