UC Berkeley
UC Berkeley Previously Published Works

Title

Multi-tracer intensity mapping: cross-correlations, line noise & decorrelation

Permalink

Journal

ISSN

Authors

Schaan, Emmanuel
White, Martin

Publication Date

DOI

Peer reviewed

Multi-tracer intensity mapping: Cross-correlations, Line noise & Decorrelation

Emmanuel Schaan[a,b] **and Martin White**[a,b,c]

[a]Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA
[b]Berkeley Center for Cosmological Physics, UC Berkeley, CA 94720, USA
[c]Department of Physics, University of California, Berkeley, CA 94720, USA

E-mail: eschaan@lbl.gov, mwhite@berkeley.edu

Abstract. Line intensity mapping (LIM) is a rapidly emerging technique for constraining cosmology and galaxy formation using multi-frequency, low angular resolution maps. Many LIM applications crucially rely on cross-correlations of two line intensity maps, or of intensity maps with galaxy surveys or galaxy/CMB lensing. We present a consistent halo model to predict all these cross-correlations and enable joint analyses, in 3D redshift-space and for 2D projected maps. We extend the conditional luminosity function formalism to the multi-line case, to consistently account for correlated scatter between multiple galaxy line luminosities. This allows us to model the scale-dependent decorrelation between two line intensity maps, a key input for foreground rejection and for approaches that estimate auto-spectra from cross-spectra. This also enables LIM cross-correlations to reveal astrophysical properties of the interstellar medium inacessible with LIM auto-spectra. We expose the different sources of luminosity scatter or "line noise" in LIM, and clarify their effects on the 1-halo and galaxy shot noise terms. In particular, we show that the effective number density of halos can in some cases exceed that of galaxies, counterintuitively. Using observational and simulation input, we implement this halo model for the Hα, [O<small>III</small>], Lyman-α, CO and [C<small>II</small>] lines. We encourage observers and simulators to measure galaxy luminosity correlation coefficients for pairs of lines whenever possible. Our code is publicly available at https://github.com/EmmanuelSchaan/HaloGen/tree/LIM. In a companion paper, we use this halo model formalism and code to highlight the degeneracies between cosmology and astrophysics in LIM, and to compare the LIM observables to galaxy detection for a number of surveys.

Contents

1 Introduction

Our ability to measure the large-scale structure of the Universe is improving rapidly, providing a promising means to study a wide variety of topics from theories of the early Universe to the formation and evolution of galaxies within the evolving cosmic web. The high redshift Universe, in particular, combines a large cosmological volume for precise statistical inference with a more linear density field which can be more reliably modeled and which is better correlated with the initial conditions.

Line intensity mapping (LIM) has emerged as a promising means of efficiently mapping the high z Universe in 3D with a high density of tracers over a large fraction of the sky [1, 2]. LIM refers to a low resolution, spectroscopic survey to measure integrated flux from unresolved sources on large areas of sky at different frequencies. LIM surveys capture the largest elements of the cosmic web and map out the distribution of matter in very large cosmological volumes in a fast and efficient manner, with good radial resolution. In this paper we will focus on LIM in optical/infra-red and sub-mm bands. More specifically, we implement a halo model for the following lines: Hα (656.28 nm), Lyman-α (121.6 nm), [OIII] (495.9 nm and 500.7 nm), [CII] (158 μm) and CO 1-0 (2.6 mm). Although we are leaving aside the field of radio-frequency LIM (for a recent survey of this area, see refs. [3, 4]), the formalism presented here applies identically to the post-reionization 21 cm.

While they can be efficient, LIM surveys are not without their challenges. Confusion among different lines from different redshifts observed at the same frequency provides a fundamental challenge to tomographic measurements. To tackle these issues, many methods have been proposed (see [1, 5] for a summary). Continuum interlopers can be controled by masking Fourier modes that vary slowly along the line of sight (LOS), corresponding to smooth spectral energy distributions. Line interlopers can be reduced in some cases by masking bright voxels, secondary line identification, cross-correlation with a contaminant template (such as a catalog of unwanted sources). Interlopers at a different redshift can cause a detectable anisotropy in the power spectrum, analogous to the Alcock-Paczynski effect [6, 7]. In many cases, cross-correlations of different lines [7], or of a line with a galaxy survey or with CMB lensing, allow one to suppress uncorrelated foregrounds [8–12]. This way, cross-correlations may be the path towards unbiased astrophysical and cosmological constraints at high redshift.

In this case, an accurate model of auto and cross power spectra is required. Galaxy luminosities in two different lines are generally not perfectly correlated. As a result, two intensity maps from the same galaxies in two different lines are not perfectly correlated either. The level of decorrelation determines the ability to use one of the lines as a proxy for the other, as in foreground cleaning. Accurately modeling the level of decorrelation is also crucial for estimating auto-spectra from observed cross-spectra, and to use LIM cross-correlations as astrophysical probes. Such modeling is generally achieved with semi-analytical models [13–19], frequently based on insights from hydrodynamical simulations [20–23] and spectroscopic surveys [24–28]. In this paper, we extend the conditional luminosity function (CLF [29]) formalism to the multivariate case, in order to correctly model the correlated scatter between different lines from the same galaxy. We build a new halo model [30, 31] upon this multi-line CLF, to predict the auto- and cross-spectra of intensity maps in different lines, including any potential decorrelation. This formalism consistently predicts cross-correlations of LIM with galaxy surveys, and galaxy and cosmic microwave background lensing (Appendix. B.2).

Because the galaxy populations at high redshifts are currently highly uncertain, predic-

tions based on different observed or simulated galaxy luminosity functions can differ wildly. In this paper, we thus do not mean to improve the accuracy of existing models for the LIM auto-spectra. Our predicted power spectra may differ significantly from others in the literature relying on different input. Instead, our goal is to provide a unified formalism, to consistently convert luminosity function inputs into joint predictions of the LIM auto-spectra, cross-spectra, and their cross-correlations with galaxy surveys and galaxy and CMB lensing.

We finish with a brief outline of the paper. In §2, we derive the general halo model based on the multivariate CLF formalism, to consistently predict the mean intensity and power spectrum of LIM and galaxy clustering in redshift-space and projected maps. We highlight the importance of scatter in galaxy and halo luminosities, and their effects on the auto and cross-power spectra. In §3, we review existing observational constraints on the LIM model derived in §2 for optical, UV and far IR lines. Specifically, we focus on the Hα, [OIII], Lyman-α, CO and [CII] lines. Finally, we use the multi-line halo model to quantify the decorrelation between intensity maps in different lines, a key quantity for several methods of line deconfusion. In the companion paper [32], we apply our formalism to highlight the degeneracies between astrophysics and cosmology with LIM. There, we compare the LIM observables to galaxy detection as probes of faint galaxies and tracers of the matter density field.

Throughout this paper, we assume a flat ΛCDM cosmology from ref. [33], with $\Omega_{\mathrm{CDM}} = 0.267$, $\Omega_b = 0.0493$, $H_0 = 67.12\,\mathrm{km/s/Mpc}$, $A_S = 2.3 \times 10^{-9}$, $n_s = 0.9624$, $N_{\mathrm{eff}} = 3.046$ with massless neutrinos. All distances, volumes and wavevectors are quoted in comoving units, and typically in $h^{-1}\mathrm{Mpc}$ or $h\,\mathrm{Mpc}^{-1}$ units unless otherwise specified. Halo masses refer to virial masses in $h^{-1}M_\odot$, defined as the mass enclosed within the virial radius, where the density is a factor $\Delta_{\mathrm{vir\ crit}}(z) = 18\pi^2 + 82\,[\Omega_m(z) - 1] - 39\,[\Omega_m(z) - 1]^2$ higher than the critical density $\rho_{\mathrm{crit}}(z)$ [34]. In all the halo model calculations, we make use of the Sheth & Tormen halo mass distribution function and linear bias [35]. We summarize the key symbols used in this paper in Table 1.

2   Multi-line halo model formalism

In this section, we present our halo model formalism, based on the conditional luminosity function. In the companion paper [32], we use this formalism to quantify the detectability of various line power spectra, disentangle the cosmological and astrophysical information, and compare LIM to galaxy detection.

2.1   Origin of nebular emission lines & line correlations

The lines of interest to us come from hot ($10^4\,\mathrm{K}$) clouds of interstellar gas that are heated, excited and ionized by ultraviolet photons (and the associated photoelectrons) from nearby hot stars or white dwarfs. These lines have long served as a major source of information about astrophysical phenomena in galaxies, such as star formation rate, metallicity, gas density and temperature [19, 36–38]. However, understanding the physical origin of nebular emission lines is also crucial for cosmology. Indeed, it determines which galaxies and which halos, are most effective at producing a given line. This in turn determines whether the corresponding line intensity map is a useful tracer of the matter density field, through the value of its bias and shot noise (of galaxies and halos). The physical origin of nebular emission lines also informs the level of correlation between a galaxy's luminosity in two different lines. Indeed, two lines produced in the same regions (e.g., HII regions) will likely be more correlated than lines

produced in distinct regions (e.g., photodissociation region and supernova remnant). As we show in this paper, this has implications for, e.g., our ability to undo the contamination from line interlopers, and thus for constraining cosmology from LIM. For this reason, we briefly summarize the physical origin of nebular emission lines, based on existing thorough reviews [36, 39–41], and illustrate it in Fig. 1.

At nebular densities, collisions are largely unimportant for the H and He lines we see, which are recombination lines. In many nebulae, all Lyman line photons are absorbed by other atoms (case B recombination). In this scenario the number of optical, Balmer photons (e.g. Hα and Hβ) equals the number of ionizing photons emitted by the stars. The level populations depend weakly on electron density and temperature so the line ratios are largely fixed by atomic physics, though observed line luminosities are additionally sensitive to absorption and dust attenuation within the nebula and in the interstellar medium. The situation is more complex for other lines. For example the strength of the collisionally ionized [OII] and [OIII] optical lines depends upon metallicity and excitation state of the gas, and is more sensitive to dust extinction. The relative strength of the IR CO transitions is sensitive to a nearly degenerate combination of H$_2$ gas density and temperature, with higher J lines being associated with warmer and denser gas. The strength of collisionally excited heavy element lines, relative to hydrogen recombination lines, is determined by the hardness of the ionizing radiation field (see §7.4 and Eq. 11 in [39]). In addition, there can be multiple gas components within a galaxy, each with differing metallicity and dust redenning, impacting the line ratios of optical and UV lines (e.g. Hα, [OII] and [OIII]) [42, 43], and far IR lines (e.g. the various CO transitions) [44]. As expected, metallicity has a strong effect on the relative flux of lines of different atomic species, and dust has a large effect on the relative flux of lines widely separated in frequency. Finally, it is expected that gas fraction, density, temperature, pressure, metallicity and the intensity and hardness of the radiation field will evolve strongly with redshift.
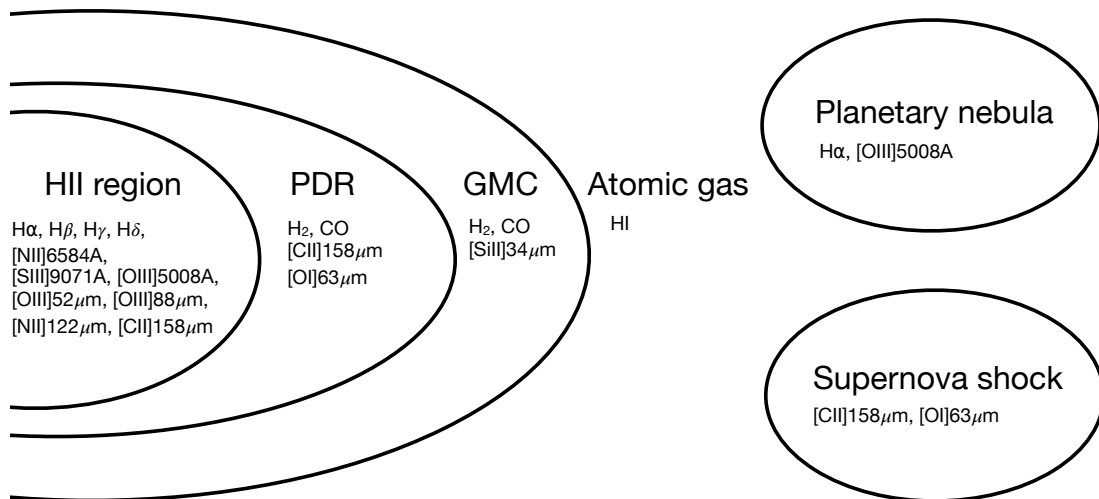

Figure 1. The line emission of a given galaxy is the sum of the contributions from the atomic gas, giant molecular clouds (GMC), photodissociation regions (PDR), HII regions around bright star, as well planetary nebulae and supernova remnants. The luminosities of lines coming from the same regions inside a galaxy are likely to be highly correlated.

2.2 Multi-line CLF: Jointly modeling multiple line intensities and power spectra

One particularly clear way of formalizing the evolution of galaxy properties and their dependence on the host halo mass is through the use of the "conditional luminosity function" (CLF; [29]), which we generalize to allow multiple luminosities (for the continuum and each line of interest). We introduce the multivariate conditional luminosity function $\phi(L_1, ..., L_n|m)$ such that $\phi(L_1, ..., L_n|m)\, dL_1...dL_n$ is the mean number of galaxies in one halo of mass m with luminosities $L_1, ..., L_n$ in lines $1, ..., n$. They satisfy:

$$\phi(L_1, ..., L_{n-1}|m) = \int dL_n \ \phi(L_1, ..., L_n|m). \tag{2.1}$$

The multivariate CLF describes the distribution of galaxy line luminosities inside halos. It determines the evolution of the galaxy population with halo mass and redshift.

The usual (unconditional) galaxy luminosity function Φ, frequently found in the literature, can be recovered from the CLF as

$$\Phi(L_1, ..., L_n) = \int dm \ n(m)\phi(L_1, ..., L_n|m), \tag{2.2}$$

such that $\Phi(L_1, ..., L_n)dL_1...dL_n$ is the mean number density of galaxies with luminosities $L_1, ..., L_n$ in lines $1, ..., n$, regardless of host halo mass.

In this paper, we focus on the mean line intensity and the line intensity auto- and cross-spectra, i.e. we discard the information from higher order correlations (e.g. the bispectrum [45] or voxel intensity distribution [46]). We give a step-by-step derivation in App. A.1 and A.3, and simply summarize and discuss the results here. The case of cross-correlating a line intensity map with a traditional galaxy or quasar survey is discussed in App. A.5 while the specialization to angular clustering is discussed in App. B. In what follows we shall neglect the distinction between satellites and centrals primarily due to a paucity of data for constraining such a model. The full case is described in the appendices and we discuss this simplification further later in the text. Finally, we ignore radiative transfer once the light is emitted by the galaxies but note that this is issue can be particularly delicate and important for the (resonant) Lyman-α line [47–49].

2.2.1 Mean intensity

The mean intensity, $\bar{I}_j$, in line j is determined by the univariate CLF, $\phi(L_j|m)$. Indeed, the zero-th moment of the CLF gives the mean number of galaxies in a halo of mass m, $N_{\rm gal}(m)$, and the total comoving number density of galaxies:

$$\begin{cases} N_{\rm gal}(m) = \int dL_1 \ \phi_1(L_1|m) \\ \bar{n}^{\rm gal} = \int dm \ n(m)N_{\rm gal}(m) \end{cases}, \tag{2.3}$$

where $n(m)$ is the halo mass function, the number of dark matter halos per volume per mass.

The first moment of the CLF gives the mean line j luminosity of a halo of mass m, $L_j(m)$, the total luminosity density per unit comoving volume, $\mathcal{L}_j$, the mean luminosity of a galaxy in a halo of mass m, $L_j^{\rm gal}(m)$, and the mean luminosity of a galaxy in the universe,

$n(m,z)$	Halo mass function
$b(m,z)$	Halo bias
$\mathrm{SFR}(m,z)$	Total star formation rate in a halo of mass m at redshift z
$\phi\left(L_1,...,L_n|m\right)$	Multi-line conditional luminosity function
$\Phi\left(L_1,...,L_n\right)$	Multi-line (unconditional) luminosity function
$\phi\left(L_j|m\right)$	Single-line conditional luminosity function
$\Phi\left(L_j\right)$	Single-line (unconditional) luminosity function
$N_{\mathrm{gal}}(m)$	Mean number of galaxies in a halo of mass m
$\bar{n}^{\mathrm{gal}}$	Total comoving number density of galaxies
$L_j(m)$	Mean line-j luminosity of a halo of mass m
$\mathcal{L}_j$	Total luminosity density per unit comoving volume
$L_j^{\mathrm{gal}}(m)$	Mean luminosity of a galaxy in a halo of mass m
$\bar{L}_j^{\mathrm{gal}}$	Mean luminosity of a galaxy in the universe
$\bar{I}_j$	mean specific intensity in line j
$P_{i,j}(k,\mu,z)$	Redshift-space cross-spectrum of lines i and j
$P_{i,j}^{\text{2-halo}}(k,\mu,z)$	2-halo term for lines i and j
$b_j(k,\mu,z)$	Scale-dependent intensity bias
$F(k,\mu,z)$	Effective growth rate of structure
$P_{i,j}^{\text{1-halo}}(k,\mu,z)$	1-halo term for lines i and j
$\bar{n}_{i,j}^{\mathrm{h\ eff}}(z)$	Effective mean number density of halos for the 1-halo term
$U_{i,j}^2(k,\mu,z)$	Effective squared halo profile for the 1-halo term.
$\sigma_{\mathrm{h}\ i,j}^2$	Halo "line noise" enhancing the 1-halo term
$P_{i,j}^{\mathrm{shot}}(z)$	Shot noise power spectrum for lines i and j
$\bar{n}_{i,j}^{\mathrm{gal\ eff}}$	Effective mean number density of galaxies for the shot noise term.
$\sigma_{\mathrm{gal}\ i,j}^2$	Galaxy "line noise" enhancing the shot noise term

Table 1. List of the main symbols used in this paper.

$\bar{L}_j^{\mathrm{gal}}$:

$$\begin{cases} L_j(m) = \int dL_j\ \phi(L_j|m)L_j \\[2mm] \mathcal{L}_j \equiv \int dm\ n(m)L_j(m) \\[2mm] L_j^{\mathrm{gal}}(m) = L_j(m)/N_{\mathrm{gal}}(m) \\[2mm] \bar{L}_j^{\mathrm{gal}} = \mathcal{L}_j/\bar{n}^{\mathrm{gal}} \end{cases} . \tag{2.4}$$

The mean halo luminosity $L_j(m)$, the key ingredient of many halo models in the literature, can thus be derived from the CLF.

The mean specific intensity in line j is

$$\bar{I}_j = \frac{1}{4\pi\nu_j^0}\frac{c}{H(z)}\ \int dL_j\ \Phi(L_j)L_j, \tag{2.5}$$

where ν_j^0 is the rest-frame frequency of line j. More intuitively

$$\boxed{\bar{I}_j = \frac{1}{4\pi\nu_j^0}\frac{c}{H(z)}\ \mathcal{L}_j,} \tag{2.6}$$

which simply converts the total galaxy luminosity per unit volume into an intensity at a given redshift and frequency.

While the mean intensity is determined by the univariate CLF, $\phi(L_1|m)$, the power spectrum also requires knowledge of the bivariate CLF, $\phi(L_1, L_2|m)$. The total power spectrum, a function of the comoving wave vector modulus k, cosine of the angle to the line of sight μ and redshift z, is made of three contributions:

$$P_{i,j}(k, \mu, z) = P_{i,j}^{\text{2-halo}}(k, \mu, z) + P_{i,j}^{\text{1-halo}}(k, \mu, z) + P_{i,j}^{\text{shot}}(z), \tag{2.7}$$

which we now describe in detail.

2.2.2 Power spectrum: 2-halo

On large scales, the number density of halos traces the underlying linear matter density field. This gives rise to the 2-halo term in the intensity power spectrum:

$$\boxed{P_{i,j}^{\text{2-halo}}(k, \mu, z) = \bar{I}_i \bar{I}_j \left[b_i + F\mu^2\right] \left[b_j + F\mu^2\right] P_{\text{lin}}.} \tag{2.8}$$

The standard assumption of a deterministic, linear, scale-independent but mass-dependent halo bias $b(m)$, combined with a (normalized) halo profile $u(k, m)$ distorted by the finger-of-god (FOG) effect, gives rise to a scale-dependent intensity bias:

$$b_j(k, \mu, z) \equiv \mathcal{L}_j^{-1} \int dm \; n(m) \; L_j(m) \; b(m) \; u(k, m) e^{-k^2\mu^2\sigma_d^2(m)/2}, \tag{2.9}$$

where $\sigma_d = \sigma_{v\,\text{1D}}/aH$ is the dispersion of the spurious displacement in redshift space due to the random LOS motion within halos $\sigma_{v\,\text{1D}}$. Following ref. [50], we approximate the LOS velocity dispersion as that of a singular isothermal sphere of mass m and radius r_{vir}, i.e. $\sigma_{v\,\text{1D}}^2 = \mathcal{G}m/2r_{\text{vir}}$. This simple analytical estimate is a good match to the simulation results reported in ref. [51]. On large scales, where $u(k, m)e^{-k^2\mu^2\sigma_d^2/2} \longrightarrow 1$, the intensity bias becomes scale-independent, and converges to the halo-luminosity-weighted averaged halo bias, shown in Fig. 2.

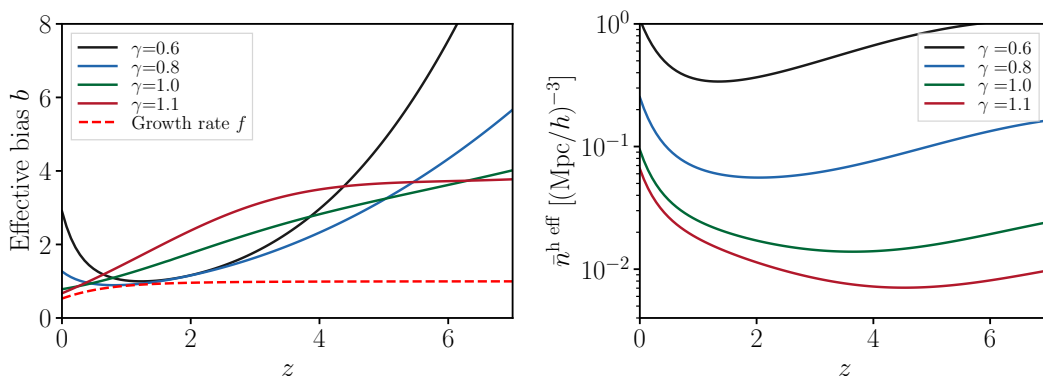**Figure 2.** Redshift evolution of the effective halo bias at $k = 0$ (left), which determines the amplitude of the 2-halo term. Redshift evolution of the effective halo number density (Eq. 2.13; right), which determines the amplitude of the 1-halo term. Different curves correspond to halo luminosities $L(M) \propto \text{SFR}^\gamma$, with $\gamma \simeq 1$ for optical and UV lines, and $\gamma = 0.6 - 1.1$ for far IR lines [52, 53].

In redshift-space, the effective growth rate appears:

$$F(k, \mu, z) \equiv f \int dm \, n(m) \left(\frac{m}{\bar{\rho}} \right) u(k, m) e^{-k^2 \mu^2 \sigma_d^2(m)/2}. \qquad (2.10)$$

Here the weighting is simply by halo mass, rather than halo luminosity, since the Kaiser effect is due to the large scale velocities, which are sourced by the mass. On large scales, where the FOG-corrected halo profile tends to one, we recover the usual logarithmic growth rate $F \longrightarrow f \equiv d \ln D / d \ln a$, as long as the matter density consistency relation $\int dm \, n(m) \, (m/\bar{\rho}) = 1$ is satisfied[1]. As shown in Fig. 2, the growth rate of structure is comparable to the effective halo bias at the lower redshifts, implying that redshift-space distortions are an important effect.

2.2.3 Power spectrum: 1-halo

On intermediate scales, the assumed Poisson distribution of halos results in a halo shot noise, called the 1-halo term:

$$P_{i,j}^{\text{1-halo}}(k, \mu, z) = \left(\frac{c}{4\pi H(z)} \right)^2 \frac{1}{\nu_i^0 \nu_j^0} \int dm \, n(m) \, |u(k, m)|^2 \, e^{-k^2 \mu^2 \sigma^2} L_i(m) L_j(m). \qquad (2.11)$$

Although the CLF formalism can describe the distinction between central and satellite galaxies (see §A.2), we ignore this distinction here and assume that all galaxies follow the NFW profile $u(k, m)$, with $u(k \to 0, m) \to 1$. The 1-halo term can be rewritten more transparently as:

$$\boxed{P_{i,j}^{\text{1-halo}}(k, \mu, z) = \bar{I}_i \bar{I}_j \frac{U_{i,j}^2(k, \mu, z)}{\bar{n}_{i,j}^{\text{h eff}}}.} \qquad (2.12)$$

As expected for a halo shot noise, the 1-halo term scales as the inverse mean number density of halos. The relevant, effective mean number density of halos, $\bar{n}_{i,j}^{\text{h eff}}$, depends on the lines i and j considered and properly counts the halos by taking into account their luminosities in these lines:

$$\bar{n}_{i,j}^{\text{h eff}} = \frac{\left(\int dm \, n(m) L_i(m) \right) \left(\int dm \, n(m) L_j(m) \right)}{\int dm \, n(m) L_i(m) L_j(m)}. \qquad (2.13)$$

This expression can be understood intuitively in several simple cases. For instance, if line luminosities were independent of halo mass above some cutoff (i.e. $L_i(m)$ and $L_j(m)$ do not depend on m), then we simply recover $\bar{n}_{i,j}^{\text{h eff}} = \int_{\text{min}} dm \, n(m)$, i.e. the total number of halos. More realistically, if the halo luminosities scale as the star formation rate in a halo of mass m, $\text{SFR}(m, z)$, we see that the halo shot noise scales as the mean squared halo star formation rate:

$$P_{i,j}^{\text{1-halo}} \propto \frac{1}{\bar{n}_{i,j}^{\text{h eff}}} \propto \int dm \, n(m) \text{SFR}^2(m). \qquad (2.14)$$

The effective halo number density is shown in Fig. 2 for cases where $L \propto \text{SFR}^\gamma$ for $\gamma = 0.6 - 1.1$, as is typical for optical, UV and IR lines.

---

[1]In practice, for most mass functions, recovering this consistency relation requires integrating down to unrealistically low halo masses. As is usual in the halo model, we "correct" this issue by dividing Eq. (2.10) by $\int dm \, n(m) \, (m/\bar{\rho})$. Such a correction is only necessary for Eq. (2.10), not Eqs. (2.9,2.11), where the luminosity or squared luminosity weighting provides an effective cutoff at low halo masses. The convergence of the mean intensity, bias and 1-halo term at low masses is shown in Fig. 7 for $\gamma = 1$. We verified it for $\gamma = 0.6 - 1.1$.

The scale and angular dependence of the 1-halo term is determined by the effective squared halo profile:

$$U_{i,j}^2(k, \mu, z) \equiv \frac{\int dm \; n(m) \, |u(k,m)|^2 \, e^{-k^2\mu^2\sigma^2} L_i(m) L_j(m)}{\int dm \; n(m) L_i(m) L_j(m)}. \qquad (2.15)$$

This is simply the average squared halo profile, weighted by the product of the halo luminosities in lines 1 and 2. On large-enough scales, where the FOG-corrected halo profile goes to unity, this takes the simple limit $U_{1,2}^2 \to 1$, such that the 1-halo term is simply:

$$P_{i,j}^{\text{1-halo}}(k \to 0, \mu, z) \to \frac{\bar{I}_i \bar{I}_j}{\bar{n}_{i,j}^{\text{h eff}}} = \left(\frac{c}{4\pi H(z)}\right)^2 \frac{1}{\nu_i^0 \nu_j^0} \int dm \; n(m) L_i(m) L_j(m), \qquad (2.16)$$

This last equation shows that it is the correlation between halo luminosities in different lines at the same mass that matters to the amplitude of the 1-halo term. The corresponding 1-halo amplitude in the $k \to 0$ limit is shown as a function of redshift in the right panel of Fig. 2.

In the halo model, the 1-halo term generically tends to a non-zero constant on large scales. This behavior appears unphysical for the matter density, where the low-k power spectrum is expected to approach the linear power spectrum $P_{\text{lin}}(k) \propto k$ [54], and any perturbation conserving mass and momentum on large scales should scale as $\delta P(k) \propto k^4$ [55–57]. This may be corrected by including the effect of halo exclusion [58] or compensated halo profiles [54], but the complete resolution of this problem remains an open question. In what follows, we therefore focus only on the small-scale behavior of the 1-halo term, and leave its large-scale behavior unspecified.

2.2.4 Power spectrum: shot noise

On the smallest scales, the Poisson fluctuations in the number of galaxies in a given halo produce a galaxy shot noise, which enhances the intensity power spectrum:

$$P_{i,j}^{\text{shot}}(z) = \left(\frac{c}{4\pi H(z)}\right)^2 \frac{1}{\nu_i^0 \nu_j^0} \int dL_i dL_j \; \Phi(L_i, L_j) L_i L_j. \qquad (2.17)$$

This galaxy shot noise term is the only one which truly requires the bivariate CLF, i.e. a knowledge of the correlation between the luminosity in different lines from the same galaxy. This can be rephrased, similarly to the 1-halo term:

$$\boxed{P_{i,j}^{\text{shot}}(z) = \frac{\bar{I}_i \bar{I}_j}{\bar{n}_{i,j}^{\text{gal eff}}},} \qquad (2.18)$$

where the effective number density of galaxies properly takes into account the contribution from each galaxy to the intensity:

$$\bar{n}_{i,j}^{\text{gal eff}} = \frac{\left(\int dL_i \; \Phi(L_i) L_i\right) \left(\int dL_j \; \Phi(L_j) L_j\right)}{\int dL_i dL_j \; \Phi(L_i, L_j) L_i L_j}. \qquad (2.19)$$

Because galaxies are point sources for our purposes, this power spectrum is scale-independent. Its amplitude scales as the inverse mean number density of galaxies, as expected for a galaxy shot noise. In Fig. 3, we predict the effective galaxy number density from various luminosity

functions in the literature, described in §3.1. Different observational determinations of the luminosity functions lead to quite different $\bar{n}_{\mathrm{gal,eff}}(z)$, though in all cases $\bar{n}_{\mathrm{gal,eff}}$ is comparable to the effective density of halos. This suggests that observations of small-scale clustering in LIM surveys could provide valuable information about the form and evolution of the underlying luminosity functions.
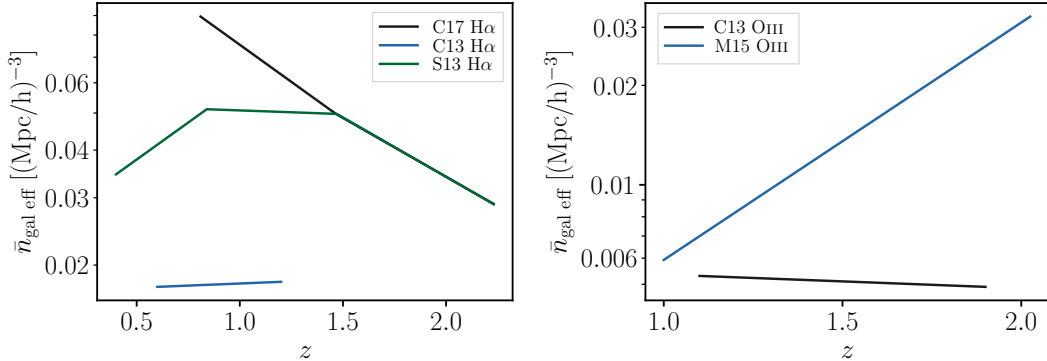

Figure 3. Comparison of the effective galaxy number densities, $\bar{n}_{\mathrm{gal,eff}}$ (Eq. 2.19), as a function of redshift, for Hα (left) and [O III] (right), as predicted from several observed luminosity functions (S13 [59]; C13 [60]; M15 [61]; C17 [62]). The various references imply very different $\bar{n}_{\mathrm{gal,eff}}$, indicating that a measurement of the galaxy shot noise would greatly increase our knowledge of the galaxy luminosity functions. In all cases, the effective number density of galaxies is similar to that of halos (Fig. 2), indicating that the shot noise and 1-halo terms are comparable in amplitude for Hα and [O III].

Looking at the mean intensity and the various power spectrum terms, we see that all the information is encoded in the multi-line CLF. The mean intensity is entirely determined by the first moment of the (unconditional) luminosity function, i.e. the mean luminosity density in the Universe $\mathcal{L}_j$. The 2-halo and 1-halo power spectra are determined by the first moment of the single-line CLF, i.e. $L_j(m)$. Finally, the shot noise requires the second moment of the two-line (unconditional) LF $\int dL_i dL_j \; \Phi(L_i, L_j) L_i L_j$. The LIM literature often uses the quantities $\mathcal{L}_j$, $L_j(m)$ and $\int dL_i \; \Phi(L_i) L_i^2$ as their starting point. Our formalism thus connects to the literature, extends it to cross-correlations, and shows how these quantities all derive from the multi-line CLF.

We combine the different components to form the observed power spectrum, in redshift space, in figure 4. For illustration figure 4 shows the Hα power spectrum for $\mu = 0$ (i.e. the real-space or tranverse power spectrum) and for $\mu = 0.5$ at $z = 0.81$, 1.47 and 2.23 with the different contributions to the total indicated by different line types. The power spectrum evolves only slowly with redshift over the interval shown, since the evolution of the matter clustering is largely compensated by the evolution of the halo-luminosity connection. One can see that the two halo term dominates on large scales (low k) with the 1-halo term dominating at intermediate scales with the shot-noise dominating at very small scales (high k). For modes close to along the line of sight direction ($\mu \approx 1$) the 2-halo and 1-halo terms are more strongly damped near $k \simeq 1 \, h \, \mathrm{Mpc}^{-1}$ while for transverse modes the 1-halo term has power to significantly higher k. Comparing the $\mu = 0$ and $\mu = 0.5$, the FOGs act as a switch, turning the 1-halo term on across the LOS and off along the LOS. Inaccuracies in our modeling of the FOG only change the scale at which this switch occurs, but not its effect.

The contribution of different halo masses to the total Hα power spectrum is shown in figure 5, again for $\mu = 0$ and $\mu = 0.5$ at $z = 0.81$. The left column shows the differential
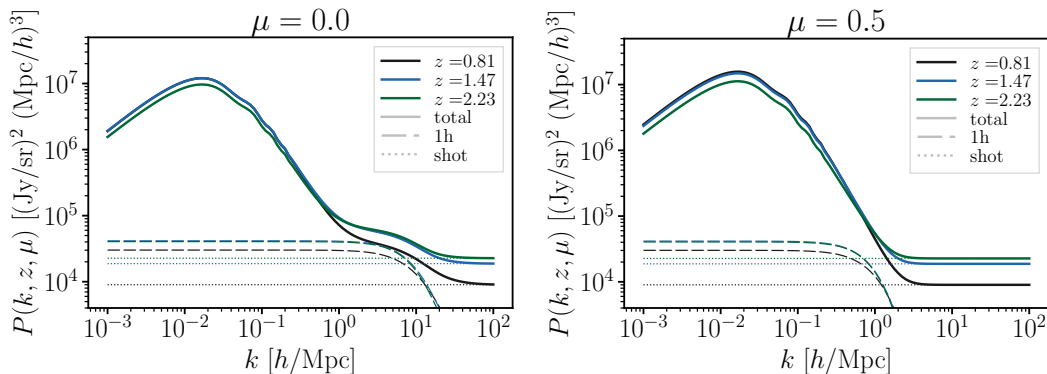
Figure 4. The total power spectrum (solid), 1-halo contribution (dashed) and shot noise (dotted) for the Hα power spectrum as a function of redshift (line colors), for $\mu = 0$ (left) and $\mu = 0.5$ (right). This assumes the Hα luminosity function from ref. [62].

contribution while the right column shows the cumulative contribution. For this line, at this redshift, we see the bulk of the power spectrum is contributed by halos in the range $10^{11} - 10^{12} \, h^{-1} M_\odot$. This is expected based on the relationship between Hα luminosity and star-formation rate, the relationship between $\dot{M}_\star$ and $M_\star$ and the stellar-mass–halo-mass relation which we discuss further in §3. While different halo mass ranges do not contribute identically to the 2-halo, 1-halo and shot-noise terms the correspondence is in fact quite good as can be seen by comparing the ordering of the lines at low and high k in the left column of figure 5. In this respect having access to the μ-dependence of the clustering signal provides significant additional information as can be seen by comparing the upper and lower panels in the left column of figure 5 near $k \simeq 1 \, h \, \mathrm{Mpc}^{-1}$.

2.2.5 Origin of galaxy and halo shot noise

The halo shot noise, which corresponds to the 1-halo term in the halo model, is determined by the effective number density of halos $\bar{n}^{\mathrm{h,eff}}$. According to Eq. 2.13, this is the integral of the mass function conditioned by the manner in which the mean line luminosity in a halo depends upon halo mass. In particular we expect star formation to be ineffective in low mass halos, making these halos essentially invisible to the LIM experiment and raising the 1-halo term. A measurement of the amplitude of the 1-halo term provides an integral constraint on the way in which luminosity is produced in halos as a function of halo mass.

By contrast the galaxy shot noise term, determined by $\bar{n}^{\mathrm{gal,eff}}$, is a sum of the fluctuations in galaxy line luminosity within halos of fixed mass and so is a probe of the stochasticity of the luminosity generation mechanism. One way to think of this stochasticity is in terms of "line noise", a concept we introduce below.

Finally, the existence of these two shot noise terms (1-halo term and galaxy shot noise) results from the assumptions that the number of galaxies in a given halo is Poisson distributed. This assumption is broken in a toy model where every halo contains either zero or one galaxy. In this case, the 1-halo and galaxy shot noise terms are identical, and should not be double counted. However, our assumption holds in the more realistic case where the halo luminosity is dominated by one galaxy, and most halos only contain zero or one galaxy, but some contain more, following a Poisson statistics.
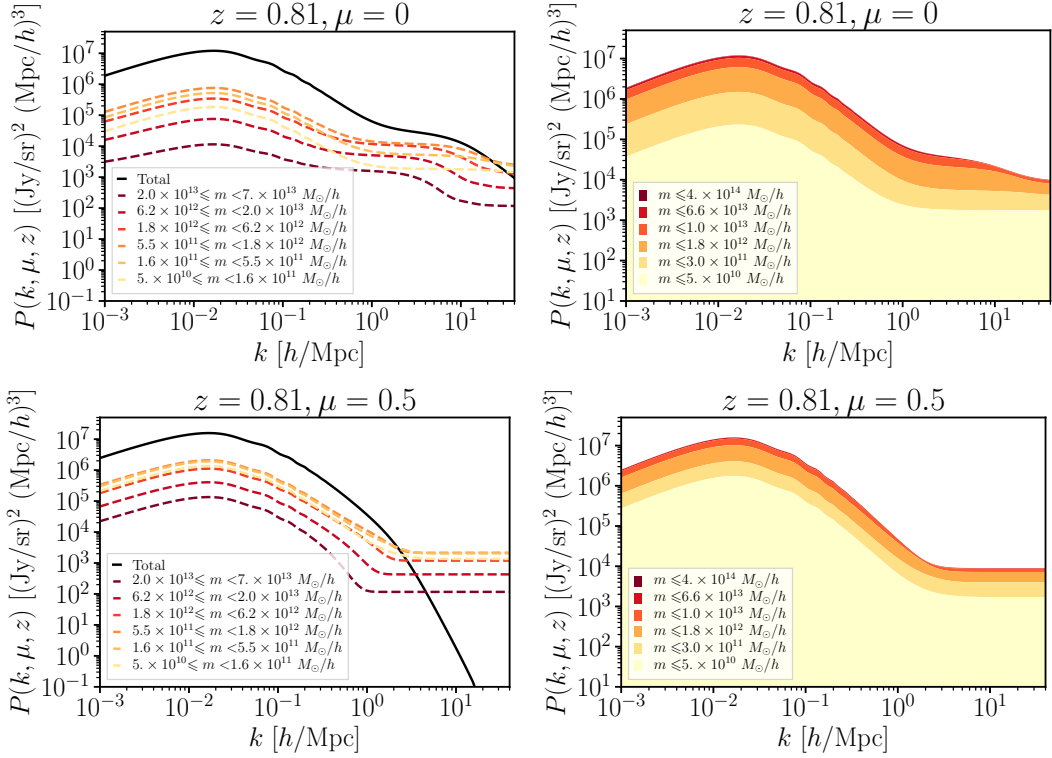
Figure 5. Contribution from different halo mass to the redshift-space power spectrum of Hα. The power spectrum is shown for wave vectors across the LOS ($\mu = 0$, top) and for wave vectors 60° away from the LOS ($\mu = 0.5$, bottom). The mass contributions are shown for each mass interval (left) and cumulatively (right). This assumes the Hα luminosity function from ref. [62].

2.3 Enhanced 1-halo term and shot noise from halo and galaxy "line noises"

The above formalism presents two interesting differences with the more familiar, galaxy-based halo model, leading to modifications to the power spectrum which we shall refer to as "line noise". These noise sources can be understood by looking at the expressions for $\bar{n}_{ij}^{\mathrm{h,eff}}$ and $\bar{n}_{ij}^{\mathrm{gal,eff}}$ which enter the 1-halo and shot-noise power spectra, respectively. In both cases, the line noise reduces the effective halo and galaxy number densities compared to our naive expectations, thus enhancing the 1-halo and shot noise power spectra.

As is apparent from Eq. (2.13), if the mass dependence of the mean halo luminosity in line i is significantly different from that in line j, $\bar{n}_{ij}^{\mathrm{h,eff}}$ can differ significantly from our naive expectations. Let us formally introduce a "halo line noise" $\sigma_{\mathrm{h}\ i,j}^2$ as the fractional covariance of halo luminosities in lines i and j across all halos:

$$\sigma_{\mathrm{h}\ i,j}^2 \equiv \frac{\langle L_i(m)L_j(m)\rangle_h - \langle L_i(m)\rangle_h\langle L_j(m)\rangle_h}{\langle L_i(m)\rangle_h\langle L_j(m)\rangle_h} \tag{2.20}$$

where the expectation value of any function X over all halos is defined as[2]:

$$\langle X \rangle_h \equiv \frac{1}{\bar{n}_h} \int dm \; n(m) \; X(m) \quad \text{with} \quad \bar{n}_h = \int dm \; n(m). \tag{2.21}$$

One can then rephrase the 1-halo power spectrum as:

$$P^{\text{1-halo}}_{i,j}(k \to 0, \mu, z) \to \frac{\bar{I}_i \bar{I}_j}{\bar{n}^{\text{h eff}}_{i,j}} = \frac{\bar{I}_i \bar{I}_j}{\bar{n}_h} \left[1 + \sigma^2_{\text{h } i,j} \right]. \tag{2.22}$$

This intuitive expression shows that any variance in the halo line luminosities reduces the effective number density of halos and enhances the 1-halo term.

A similar situation arises in the galaxy shot-noise term. This is the only term that depends upon the bivariate CLF, and so is sensitive to galaxy-line decorrelations. As Eq. (2.19) makes clear, $\bar{n}^{\text{gal,eff}}_{ij}$ depends upon how $\langle L_i L_j \rangle$ compares to $\langle L_i \rangle \langle L_j \rangle$ with the averages being over the CLF and halo mass function as shown explicitly in Eq. (2.19). To bring out this point more clearly, and to connect to Appendix A.3, let us introduce "galaxy line noise", $\sigma^2_{\text{gal } i,j}$, as the fractional covariance of lines i and j luminosities across all galaxies:

$$\sigma^2_{\text{gal } i,j} \equiv \frac{\langle L_i L_j \rangle_{\text{gal}} - \langle L_i \rangle_{\text{gal}} \langle L_j \rangle_{\text{gal}}}{\langle L_i \rangle_{\text{gal}} \langle L_j \rangle_{\text{gal}}} \tag{2.23}$$

where the expectation value over all galaxies is defined as:

$$\langle X \rangle_{\text{gal}} \equiv \frac{1}{\bar{n}_{\text{gal}}} \int dL_i dL_j \; \Phi(L_i, L_j) X, \tag{2.24}$$

allowing the shot noise power spectrum to be written as:

$$P^{\text{shot}}_{i,j} = \frac{\bar{I}_1 \bar{I}_2}{\bar{n}^{\text{gal eff}}_{i,j}} = \frac{\bar{I}_1 \bar{I}_2}{\bar{n}_{\text{gal}}} \left[1 + \sigma^2_{\text{gal } i,j} \right]. \tag{2.25}$$

The "galaxy line noise" is thus indeed an additional noise term, which enhances the galaxy shot noise. Ref. [63] recently used the a similar formalism to show the effect of line noise on the galaxy shot noise. Here, we have shown that the same formulation applies to the 1-halo term as well.

We quantify the importance of the halo and galaxy line noises below, and discuss the potential decorrelation they produce between the two intensity maps in line i and j in §4.2.

2.4 Special cases and ansatz for the CLF

In order to gain some intuition for these noise sources, and for $\bar{n}^{\text{h,eff}}_{ij}$ and $\bar{n}^{\text{gal,eff}}_{ij}$, let us consider some special forms of the CLF and see how they affect the amplitudes of the 2-halo, 1-halo and shot-noise terms.

---

[2]A subtlety with this formulation is that both $\sigma^2_{\text{h } i,j}$ and $\bar{n}_h = \int dm \; n(m)$ are "cutoff dependent", i.e. dependent on the arbitrary choice of the lower bound in the mass integrals. The expression for the 1-halo power spectrum remains correct for any sufficiently small choice of lower mass bound though, and the intuitive meaning of halo line noise remains.

Equal galaxy luminosities If all galaxies in the Universe have the same luminosity, i.e. if the CLF is

$$\phi\left(L_1, ..., L_n | m\right) \equiv \prod_{i=i}^{n} \delta^D\left(L_i - \bar{L}_i\right) N_{\rm gal}(m), \qquad (2.26)$$

then we recover the standard HOD description of the galaxy clustering (with $N_{\rm gal}$ including both the central and satellite galaxies, which we are not differentiating between). In particular, as expected, the galaxy line noise $\sigma^2_{\rm gal\,\,i,j}$ is zero in this case, and the effective galaxy number density is simply $\bar{n}^{\rm gal\,eff}_{i,j} = \bar{n}_{\rm gal}$, without any luminosity weighting.

However, intuitively, there is a non-zero halo line noise, since the halo luminosity $L_i(m) = \bar{L}_i N_{\rm gal}(m)$ varies with mass. Formally:

$$\sigma^2_{\rm h\,\,i,j} = \frac{\langle N^2_{\rm gal}(m)\rangle_h}{\langle N_{\rm gal}(m)\rangle_h^2} - 1. \qquad (2.27)$$

Further, if $N_{\rm gal}(m) \propto m$ with a large proportionality constant such that the galaxy shot noise is negligible, we simply recover the matter power spectrum (in redshift space), up to a multiplicative factor.

Our ansatz for the CLF In what follows, we will assume the following form for the CLF of the lines of interest:

$$\phi\left(L_1, ..., L_n | m\right) \equiv \frac{\Phi\left(L_1, ..., L_n\right)}{\int dL_1...dL_n\,\,\Phi\left(L_1, ..., L_n\right)} N_{\rm gal}\left(m\right). \qquad (2.28)$$

In effect, we have separated the dependences in halo mass and galaxy luminosity, a reasonable simplification given the scarcity of observational and simulation constraints on the CLF. The CLF ϕ, on the left-hand side, is expressed as a function of the usual (unconditional) luminosity function Φ on the right-hand side. The ratio can be interpreted as the joint probability distribution function for the luminosities $L_1, ..., L_n$ in the lines $1, ..., n$. However, this model is sufficient to encode the potential correlation between different line luminosities. As we describe in §4.1.1, ref. [61] provides a fitting function for the joint galaxy luminosities in Hα and [O III]. This bivariate LF is particularly valuable, as it allows us to compute the correlation coefficient between these two lines, a crucial input to our ansatz for predicting the LIM cross-power spectra. The halo mass dependence is completely encoded in the mean number of galaxies $N_{\rm gal}\left(m\right)$ per halo. In particular, the mean halo luminosity is simply $L_i(m, z) = \bar{L}_i N_{\rm gal}(m, z)$. This effectively assumes that the galaxy populations in all halos are identical, except for the number of galaxies per halo. It also means that the halo line ratios $L_i(m, z)/L_j(m, z)$ and galaxy line ratios $\bar{L}^{\rm gal}_i(z)/\bar{L}^{\rm gal}_j(z)$ are independent of mass. This also ignores the distinction between central and satellite galaxies inside halos, in order to reduce the number of free parameters in the model, given the sparsity of observational constraints. However, in Appendix A.2, we show that the CLF formalism naturally describes this distinction, and we derive the general expressions there. In practice, we model the number of galaxies per halo as:

$$N_{\rm gal}(m, z) \equiv \frac{\bar{n}_{\rm gal}(z)\,{\rm SFR}^\gamma(m, z)}{\int dm'\,\,n(m', z)\,{\rm SFR}^\gamma(m', z)}. \qquad (2.29)$$

This corresponds to the mass scaling $N_{\rm gal}(m) \propto \rm SFR^\gamma(m)$, customary in the LIM literature[3]. For the halo mass function and linear bias, we implement the fitting function from ref. [35].

To evaluate the star formation rate in a halo of given mass and redshift, we rely on the mean halo-to-stellar mass relation from ref. [67] (their Eqs. 2, 11-14 and Table 1) and the mean relation between stellar mass and star formation rate [68]. Refs. [67, 68] find a lognormal scatter of 0.15 dex (respectively 0.2 dex) in the ref. [67] (respectively ref. [68]) relation. However, as we show in Appendix A.3, the mean intensity, 1-halo and 2-halo terms only depend on the mean $\rm SFR(m)$ relation, not its scatter. While this is not true for the galaxy shot noise, this term is computed directly from the line LF, rather than the $\rm SFR(m)$ relation. In conclusion, we do not need to include the scatter in the $\rm SFR(m)$ relation to accurately predict the mean intensity, 1-halo, 2-halo and galaxy shot noise terms in our formalism. Note that this approach only yields the SFR of the central galaxy in the halo, whereas one should in principle include the sum of the SFRs of all galaxies within halos. However, as we show in §2.6, the LIM observables are dominated by low-mass halos ($m \sim 10^{12} M_\odot/h$), in which the number and luminosities of satellite galaxies is expected to be small. Indeed, ref. [69] shows that the mean number of satellites is only a few percent of the mean number of centrals in these low mass halos (see their Fig. 3). The mass integral is simply $\int dm \, n(m) \, N_{\rm gal}(m) = \bar{n}_{\rm gal}(z)$, as expected. Here, the halo line noise is entirely due to the mass dependence of the halo star formation rate:

$$\sigma^2_{{\rm h}\,i,j} \sim \frac{\langle \rm SFR^{\gamma_i+\gamma_j}(m)\rangle}{\langle \rm SFR^{\gamma_i}(m)\rangle \langle \rm SFR^{\gamma_j}(m)\rangle} - 1. \tag{2.30}$$

In practice the mass average above is ill-defined in the absence of a mass cutoff, since the mean number density of halos of all masses diverges. For this reason we have used a $\sim$ in the expression above, indicating that it is only a formal relation, but we believe it is still useful in highlighting the importance of the relative scatter in the halo SFR.

With our CLF ansatz, the galaxy line noise is simply given by the covariance of the individual galaxy luminosities:

$$P^{\rm shot}_{i,j} = \frac{\bar{I}_i \bar{I}_j}{\bar{n}_{\rm gal}} \left[1 + \sigma^2_{{\rm gal}\,i,j}\right] \quad \text{with} \quad \sigma^2_{{\rm gal}\,i,j} = \frac{{\rm Cov}\,[L_i, L_j]}{\bar{L}_i \bar{L}_j}. \tag{2.31}$$

---

[3]Taking $L \propto \rm SFR^\gamma$ is a common assumption, also made in refs. [6, 7, 52, 64] as just a few recent examples. However, recent simulation work [65, 66] suggests that such a simple model may not be an accurate extrapolation at the lower halo masses.

In summary, with this ansatz, the halo model expressions simplify to:

$$
\begin{cases}
\bar{I}_j = \dfrac{1}{4\pi\nu_j^0}\dfrac{c}{H(z)}\ \displaystyle\int dL_j\ \Phi(L_j)L_j, \\[2mm]
P_{i,j}^{\text{2-halo}}(k,\mu,z) = \bar{I}_i\bar{I}_j\left(b_i + F\mu^2\right)\left(b_j + F\mu^2\right)\ P_{\text{lin}} \\[2mm]
\text{with}\quad b_j(k,\mu,z) = \dfrac{\int dm\ n(m)\ \text{SFR}^{\gamma_j}(m)\ b(m)\ u(k,m)e^{-k^2\mu^2\sigma_d^2(m)/2}}{\int dm\ n(m)\ \text{SFR}^{\gamma_j}(m)} \\[2mm]
\text{and}\quad F(k,\mu,z) = f\displaystyle\int dm\ n(m)\left(\dfrac{m}{\bar{\rho}}\right)u(k,m)e^{-k^2\mu^2\sigma_d^2(m)/2}, \\[2mm]
P_{i,j}^{\text{1-halo}}(k,\mu,z) = \left(\dfrac{c}{4\pi H(z)}\right)^2\dfrac{\bar{I}_i\bar{I}_j}{\nu_i^0\nu_j^0}\ \dfrac{\int dm\ n(m)\text{SFR}^{\gamma_i+\gamma_j}(m)\ |u(k,m)|^2\ e^{-k^2\mu^2\sigma^2}}{\int dm\ n(m)\text{SFR}^{\gamma_i+\gamma_j}(m)}, \\[2mm]
P_{i,j}^{\text{shot}}(z) = \left(\dfrac{c}{4\pi H(z)}\right)^2\dfrac{1}{\nu_i^0\nu_j^0}\displaystyle\int dL_i dL_j\ \Phi(L_i,L_j)L_iL_j.
\end{cases}
\tag{2.32}
$$

In what follows, we gather observational and simulation constraints on $N_{\text{gal}}(m)$ and $\phi(L_i, L_j)$, in order to evaluate the LIM observables.

2.5   Relative size of the 1-halo and shot noise terms

We have seen that the 1-halo and galaxy shot noise terms scale are $I^2/\bar{n}^{\text{h eff}}$ and $I^2/\bar{n}^{\text{gal eff}}$ on large scales, respectively. Since halos can host several galaxies, one might expect the effective number density of galaxies to be higher than that of halos. One would thus expect the 1-halo term to be larger than the galaxy shot noise. In this section, we show that this is not necessarily the case, and clarify the regime in which this occurs.

From our Appendix A.3, we can write schematically the 1-halo and shot noise terms as:

$$
P = \bar{n}_h\left[\underbrace{\bar{N}_g^2\langle L_g\rangle^2}_{\text{1-halo}} + \underbrace{\bar{N}_g\langle L_g^2\rangle}_{\text{shot noise}}\right],
\tag{2.33}
$$

where P is the LIM power spectrum, $\bar{n}_h$ the mean number density of halos, $\bar{N}_g$ the mean number of galaxies per halo, and L_g an individual (random) galaxy luminosity. So the comparison is:

$$
\text{1-halo} \geq \text{shot noise}\quad\Leftrightarrow\quad \bar{N}_g \geq \frac{\langle L_g^2\rangle}{\langle L_g\rangle^2} = 1 + \sigma_g^2.
\tag{2.34}
$$

The 1-halo term is thus larger than the shot noise in the regime where halos contain many galaxies, and the fluctuations in galaxy luminosity are small.

The case of galaxy clustering, instead of intensity mapping, is obtained by setting $L_g = 1$ for each galaxy, such that galaxies are number-weighted, as opposed to luminosity-weighted. The comparison between 1-halo term and shot noise is thus simply:

$$
\text{1-halo} \geq \text{shot noise}\quad\Leftrightarrow\quad \bar{N}_g \geq 1.
\tag{2.35}
$$

Thus, for clustering, the 1-halo term is larger than the shot noise if halos host on average more than one galaxy. In practice, none of these inequalities have to be satisfied. Thus

the 1-halo term may be larger or smaller than the shot noise. In terms of the CLF, the comparison can be expressed as:

$$1\text{-halo} \geq \text{shot noise} \quad \Leftrightarrow$$
$$\int dm \; n(m) \; \left(\int dL \; \Phi(L|m)L \right)^2 \geq \int dm \; n(m) \; \int dL \; \Phi(L|m)L^2. \tag{2.36}$$

In the particular case of our CLF ansatz, this becomes:

$$1\text{-halo} \geq \text{shot noise} \quad \Leftrightarrow \quad \frac{\int dm \; n(m) \; \text{SFR}^{2\gamma}(m)}{\left(\int dm \; n(m) \; \text{SFR}^{\gamma}(m) \right)^2} \geq \frac{\int dL_i \; \phi(L_i)L_i^2}{\left(\int dL_i \; \phi(L_i)L_i \right)^2}. \tag{2.37}$$

Given the observed LF and the SFR(m) relation at the corresponding redshift, this inequality may or may not be true. We illustrate both cases in §3.1 (see Fig. 11).

2.6 Halo masses probed

A question of particular interest is how LIM compares to galaxy catalogs, in terms of the halo masses probed. For instance, galaxies selected from their emission lines are known to reside in lower mass halos than luminous red galaxies [69–71], and preferentially in sheets and filaments [72] where their star formation has not been quenched.

Before continuing it is worth reopening the issue of satellite vs. central galaxies, a distinction we have neglected in the formalism in the main text since we are unable to meaningfully constrain the more complex model at this point. For many of our inferences, neglecting this distinction is a reasonable approximation for LIM, since much of the emission-weighted signal comes from quite low mass halos (e.g. Fig. 5) where the satellite contribution is expected to be small. Further, as we move to higher redshift (where LIM becomes increasingly competitive) the number of high mass halos continues to drop but the manner in which galaxies occupy halos evolves much less dramatically [67] so the relative contribution of satellites is reduced. Neglecting the satellite-central distinction is however likely to impact our estimates for the FOG effect. For this reason we regard the estimates presented here as less reliable than many of our other inferences, and we would require more observational constraints in order to be able to properly refine them. Conversely LIM measurements could provide the much-needed constraints on a model including such a central-satellite split if they were able to tightly constrain the 2D power spectrum in the region where FoG effects dominate.

As shown in Fig. 6, the mean intensity and the 2-halo term (through the squared bias) receive contribution from a broad range of halo masses. Masses within $10^{10} - 10^{13} \, h^{-1} M_\odot$ contribute more than 80% of these quantities. On the other hand, the 1-halo term receives contributions from a narrower and heavier range of mass: 80% of it comes from $10^{12} - 10^{13} \, h^{-1} M_\odot$ halos. This is expected, since the 1-halo term weighs halos by their squared luminosity, thus up-weighting higher mass halos. In Fig. 6, the smooth cutoff at low masses is due to the SFR(m) relation, whereas the sharper cutoff at high masses is due to the exponential cutoff in the halo mass function. Galaxies selected by luminosity, stellar mass or line emission tend to only be found in halos with mass $m \geq 10^{11} \, h^{-1} M_\odot$ [69], especially at higher z. Through the mean intensity and 2-halo term, LIM may thus enable probing slightly lower mass halos, down to $10^{10} \, h^{-1} M_\odot$.
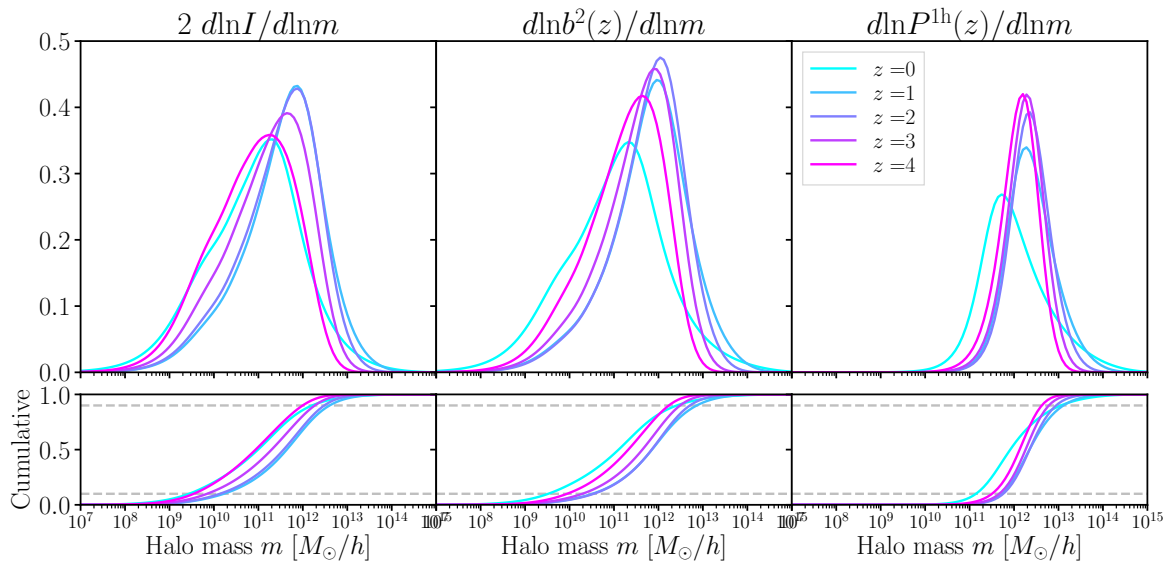
Figure 6. Contribution of each halo mass to the mean intensity (left), the 2-halo term (middle, via the squared effective halo bias) and 1-halo terms (right) at various redshifts. Here, we assumed that the halo luminosity scales as SFR, i.e. $\gamma = 1$, as is the case for most UV and optical lines. The mean intensity and 2-halo term probe similarly low halo masses, whereas the 1-halo term is sensitive to more massive halos. The grey dashed lines in the cumulative plots (bottom panels) indicate the 10th and 90th percentiles.

2.7 Galaxy luminosities probed

A key astrophysics goal of LIM is to learn about the distribution of luminosities of the galaxies which make up the intensity map. Our halo model predicts the relative weight of each galaxy luminosity bin to the LIM observables, as shown in Fig. 7.

3 LIM auto-spectra: single-line luminosity functions

The key quantity in our halo model is the galaxy CLF. The single-line CLF entirely determines the LIM auto-spectrum and the multi-line CLF predicts the LIM cross-spectrum. These CLFs can be measured in galaxy surveys or from simulations. In this section, we present several such constraints on the CLF.

3.1 Observational & simulation constraints

As we show, modeling the LIM power spectrum requires knowledge not only of the mean, but also of the scatter in the galaxy luminosities. Line luminosities can vary by large factors from galaxy to galaxy, driven in large part by the scatter in star formation rate [19, 36–38]. This scatter in line luminosity enhances the shot noise power spectrum.

Below we model the Hα, [OIII], Ly-α, [CII] and CO line luminosities using empirical fits to observed LFs. Ref. [62] presents Schechter model fits to the Hα LF (table 2). These fits are based on the High-redshift(Z) Emission Line Survey (HiZELS) survey [59], including additional and near-infrared narrow-band data from the Canada France Hawaii Telescope [73]. They cover three redshift bins over $z = 0.8 - 2.2$, and are corrected for the contamination
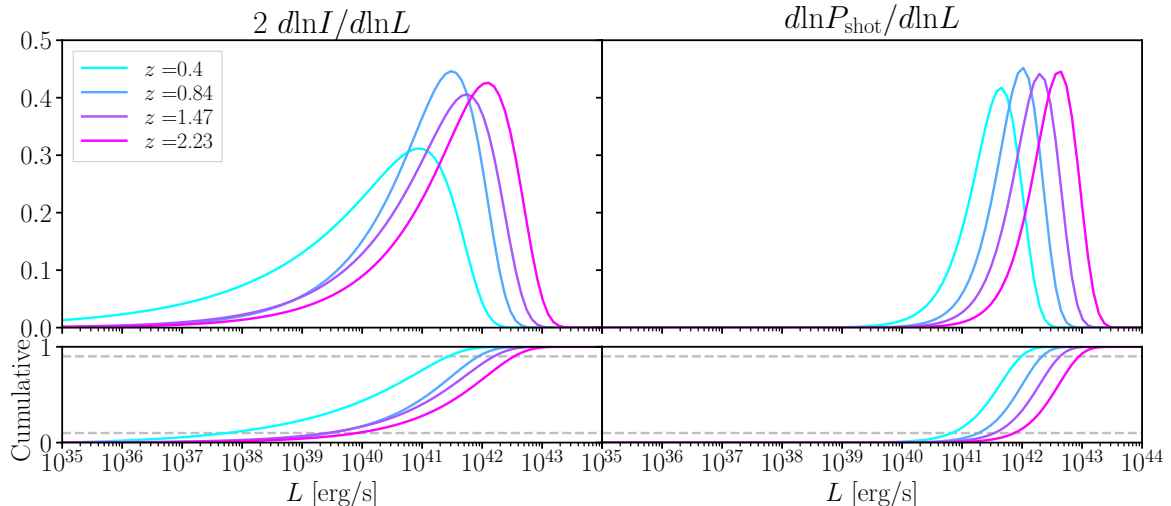
Figure 7. Contribution from each galaxy luminosity to the Hα mean intensity (left) and shot noise power spectrum (right), using LFs from ref. [59]. The mean intensity is sensitive to much lower luminosities than the shot noise. The grey dashed lines in the cumulative plots (bottom panels) indicate the 10th and 90th percentiles.

from [N II] lines at $654.8\,\mathrm{nm}$ and $658.4\,\mathrm{nm}$. These LFs also include a correction for dust attenuation, assuming a uniform 1 mag of dust extinction. We undo this correction in order to obtain the observed, dust attenuated, luminosity distribution. Ref. [60] presents Schechter model fits to the observed Hα and [O III] LFs from the WFC3 Infrared Spectroscopic Parallel Survey (WISP) program. The Hα LF fits describe two redshift bins over $z = 0.3 - 1.5$. The LF of the higher redshift bin is consistent with ref. [59], while that of the lower redshift bin is found to be slightly lower. The [O III] LFs is fit in four redshift bins spanning $z = 0.7 - 2.3$. Using data from the WISP program, ref. [61] further estimates the joint bivariate LF in Hα and [O III], providing a fitting function in one redshift bin covering $z = 0.8 - 1.2$. This is the observed, dust attenuate LF relevant to our LIM forecast. These measured Hα LFs have been used as a reference for a number of numerical and semi-analytical models, e.g. refs. [28, 64, 74–77]. Using HiZELS, ref. [78] measure the Hβ+[O III] and [O II] LFs in five redshift bins over $z = 0.8 - 5$. They compare their Schechter fits to existing measurements, including refs. [59, 60]. At $z = 1.47$, their Hβ+[O III] LF is consistent with the [O III] LF from [60] at the bright end, where the [O III]-emitting galaxies dominate over the Hβ-emitting population, but they diverge at the faint end, where the Hβ galaxies dominate. At the same redshift, their [O III] LF is consistent with [59]. We implement the Schechter fits to the [C II] and CO LFs predicted by the semi-analytical model of [79] over $z = 0 - 6$. The [C II] LF was also measured in the local Universe by [80]. Finally, we implement the Ly-α LF from [81] over $z = 2 - 5$.

The various LFs are shown in Fig. 8. There is some disagreement in the Hα luminosity function at $z \simeq 0.8$ but the individual estimates agree very well at higher redshift. The [O III] luminosity function shows more diversity, especially at the bright end and at higher redshift ($z \simeq 1.9$).

We compute the first two moments of these LFs, needed to compute the LIM mean intensity and power spectrum. The first moment gives the mean intensity (Fig. 9), which also
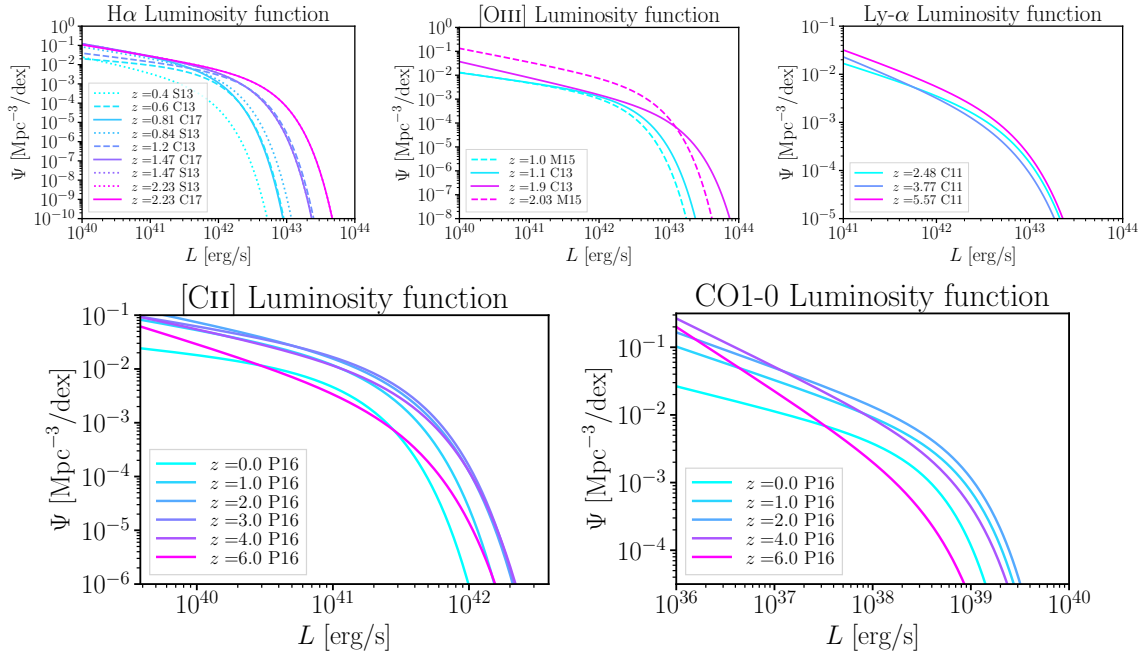
Figure 8. Galaxy luminosity functions $\Psi(L) \equiv dN/d\log_{10} L\, dV$ for Hα, [OIII], Ly-α, [CII] and CO1-0 from the various references (S13 [59], C13 [60], M15 [61], C17 [62], P16 [79] and C11 [81]). The Hα LFs from C17 and S13 are identical at $z = 1.47$ and 2.23.

normalizes the LIM power spectrum. As might have been expected from the disagreement in the luminosity functions themselves, the amplitude and evolution of the mean intensity is not well constrained at present, especially for [OIII]. The relative sizes of the mean intensities in different lines can be understood by comparing their LFs and rest-frame frequencies, since $\bar{I} \propto \left(\int dL\, \Phi(L)L \right)/\nu^0$ (Eq. (2.5)). For instance, the mean intensity in CO and optical lines are similar, even though galaxy CO luminosities are typically $\sim 10{,}000$ fainter. This is compensated by the CO rest-frame frequency at ~ 100 GHz, about $\sim 10{,}000$ smaller than for optical lines. On the other hand, the mean intensity is [CII] is much larger than in optical lines. This can be understood since the galaxy luminosities in [CII] are only somewhat fainter than in the optical lines, and the $\sim$ THz frequency of the [CII] line is $\sim 1{,}000$ smaller than for the PHz optical lines. The second moment determines the shot noise term (Fig. 10), through the effective number density of galaxies (Fig. 3 and Eq. 2.19). The level of agreement is qualitatively similar to that displayed by the first moment, even though the luminosity weightings are of course different in the two. The wide discrepancies between the predicted mean intensities (Fig. 9) and shot noises (Fig. 10) from the various measured luminosity functions is striking.

In Fig. 9, 10, 11, we only show the redshift ranges where luminosity functions from actual observations are available. We do not show the predictions from simulations such as EGG.

3.2 LIM auto-spectra: theoretical uncertainty

We show the predicted LIM power spectra from these LFs in Fig. 11. We show a wide range of scales, where each of the 2-halo, 1-halo and shot noise terms successively dominate. However, the model should certainly not be trusted out to these extremely small scales.
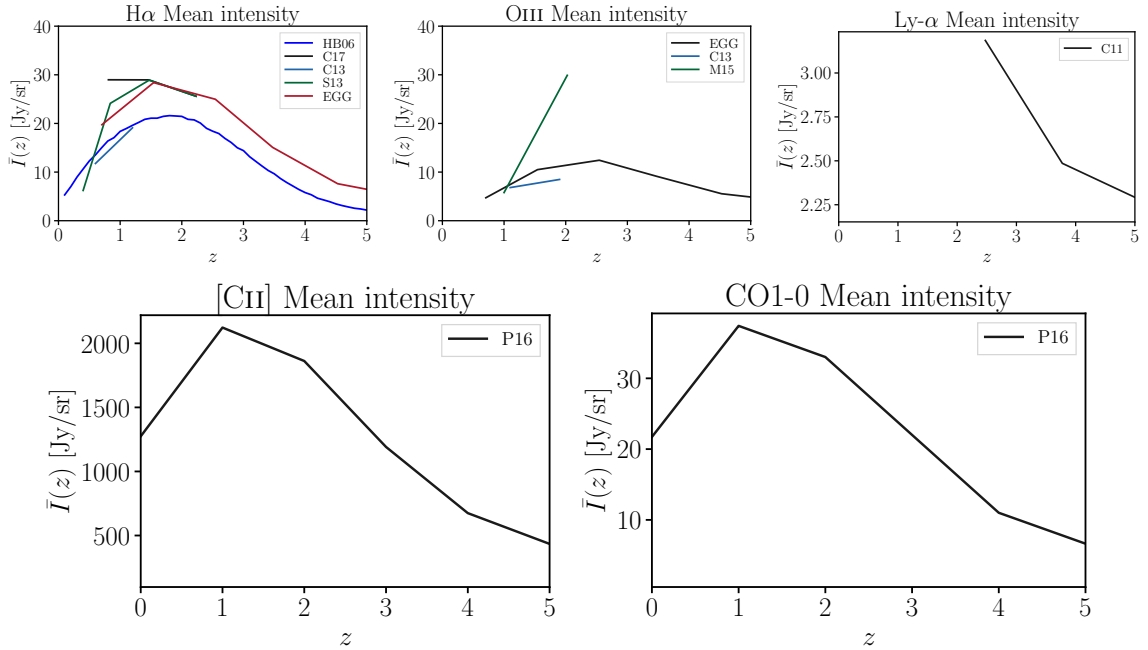
Figure 9. Comparison of the mean intensity as a function of redshift, for Hα, [OIII], Ly-α, [CII] and CO1-0, as predicted from the various observed luminosity functions (see Fig. 8) as well as the EGG simulator [28]. This first moment of the galaxy luminosity function also determines the overall normalization of the LIM power spectrum. The relative sizes of the mean intensities in the various lines can be understood by comparing their LFs (Fig. 8) and rest-frame frequencies.

Comparing the power spectra of the various lines, the relative size of the 2-halo and 1-halo term is determined by the index γ in the relation between mass and SFR. The overall normalization and its redshift evolution are further determined by the redshift evolution of the line luminosity function. Overall, the 1-halo and 2-halo terms appear to generally increase out to $z = 1 - 2$, then decrease at higher redshift. Because the redshift evolutions of the 1-halo and 2-halo terms are not identical, this corresponds to a complex redshift and scale-dependent bias. Finally, we note that the relative size of the 1-halo and shot noise terms depends crucially on our assumption that the galaxies are NFW-distributed within halos. If all galaxies were instead centrals, the 1-halo term would be much reduced, subdominant compared to the shot noise. This may be the case for instance for CO and [CII] at high redshift.

We compare these power spectra to previous predictions in the literature. Many useful references for comparison exist, including [82] for many of these lines, and [83–88] for CO and [CII]. For the Hα power spectrum, we compare with Fig. 25 in ref. [89]. There, the $z = 2$ power spectrum matches our prediction in the 1-halo and shot noise regimes, to about a factor two. Comparing with Fig. 5 in ref. [64] at $z = 1 - 2$, the 2-halo terms agree again to a factor of a few, but the shot noise in our analysis is larger by factors $10 - 100$. For [OIII], comparing with Fig. 25 in ref. [89] shows a 2-halo term at $z = 2$ consistent with ours from the LF of ref. [61], and larger than those from the LF of ref. [60]. We compare our Ly-α power spectrum to that in Fig. 6 of ref. [90]. Although the redshifts are slightly different ($z = 7$ instead of $z = 5.57$), the 2-halo terms agree within a factor about two, whereas our shot noise is larger by a a factor 100. This is probably explained in part by the difference in
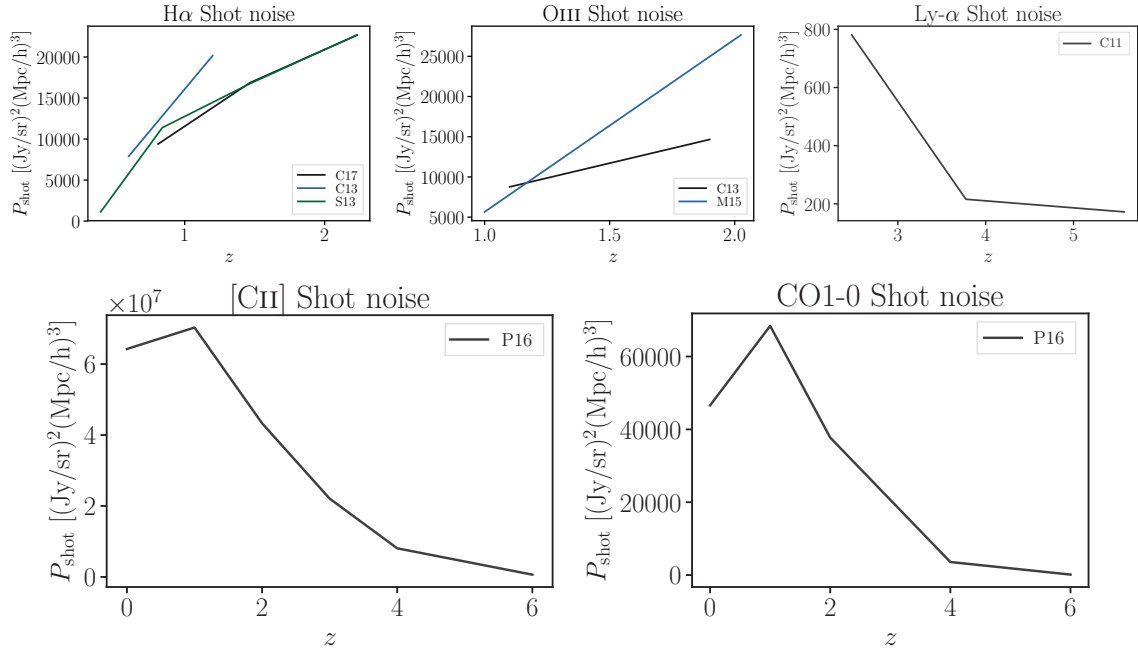
Figure 10. Shot noise power spectrum as a function of redshift, for Hα, [OIII], Ly-α, [CII] and CO1-0, as predicted from the various observed luminosity functions (see Fig. 8). The shot-noise power is determined by the second moment of the galaxy luminosity function.

redshifts, and in part by the theoretical uncertainty, estimated to be a factor 100 in Fig. 25 of ref. [89] in this range of redshifts. For the CO power spectrum, we compare to ref. [14]. In their Fig. 6 and 10, they show a range of predictions from the literature and from various relations between SFR and CO luminosity. These span a factor of 1000 at $z = 2.4 - 2.8$. Our prediction is within this range: in the upper end for the 2-halo term and in the middle for the shot noise term. For the [CII] power spectrum, we compare our prediction to Fig. 16 in ref. [86] at $z = 6$. This figure shows a factor of about 1000 of theoretical uncertainty. Our predicted 2-halo term is a factor two lower than their lowest prediction, and our 1-halo term is more than 10 times smaller than their lowest estimate. Given the large theoretical uncertainty, this is not completely unexpected. It may be explained by our choice of the LF of ref. [79]. Indeed, refs. [15, 79] mention that the simulated [CII] in[79] is lower than most semi-analytical models.

Thus, in general, the model predictions from different observed LFs or semi-analytical models differ significantly. This highlights the extreme theoretical uncertainty on the LIM power spectra, especially at high redshift where the LFs are not well known. This implies that measuring the LIM mean intensity and power spectrum will provide an important further constraint on the galaxy luminosity function, and the combination of the measurements should allow us to better constrain the underlying LFs.

4 LIM cross-spectra & decorrelation: multi-line luminosity functions

4.1 Observational & simulation constraints

Quantifying the correlation between the luminosities in different lines is crucial to correctly predict their cross-spectra in the shot noise regime, and in particular to assess the degree
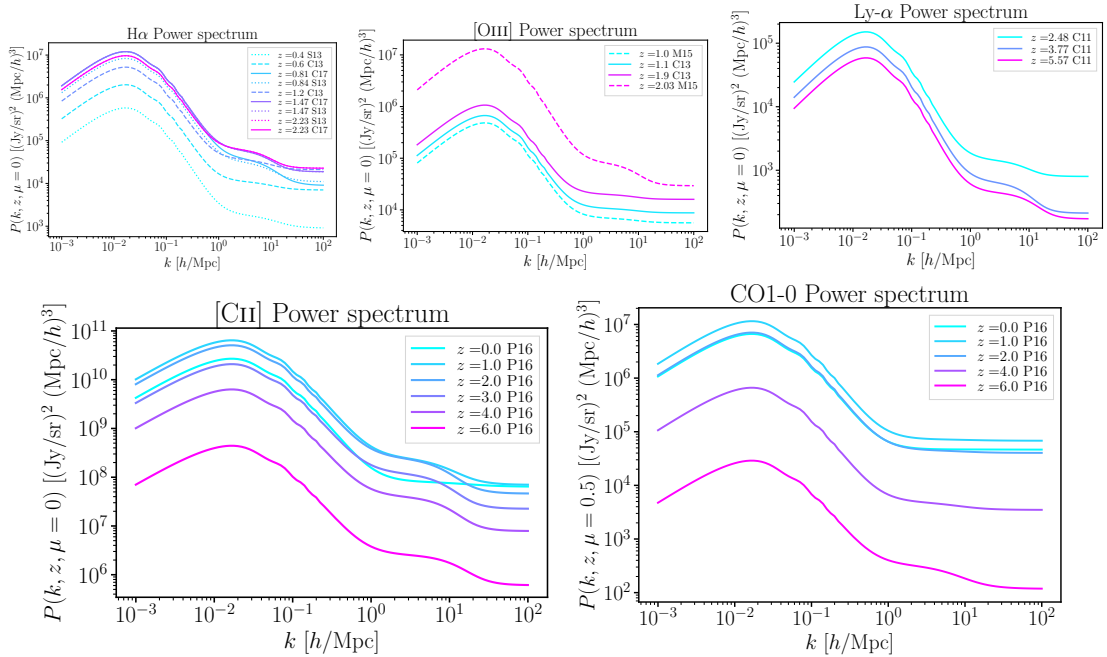
Figure 11. Power spectra of the Hα, OIII, Ly-α, [CII] and CO lines, predicted from the various observed luminosity functions (S13 [59], C13 [60], M15 [61], C17 [62], C11 [81] and P16 [79]). We show the model predictions out to extremely small scales ($k \sim 100 h/\mathrm{Mpc}$) to illustrate that the 2-halo term dominates on the largest scales, the 1-halo term on intermediate scales, and the shot noise on the smallest scales. However, our model should not be relied upon on such small scales.

of correlation of the two intensity maps. As described in §2.1, several physical parameters can affect the line luminosity ratios [36]. Metallicity, dust redenning and the ionization state of the gas determine the line ratios of optical and UV lines, such as Hα, [OII] and [OIII] [42], and far IR lines such as the various CO transitions [44]. Naturally, line ratios involving high excitation lines such as [OIII] or higher J CO lines are most sensitive to the ionization state [44]. This ionization state is determined by the local interstellar radiation field, which itself traces the local star formation rate density. In practice, accurately fitting the spectral line energy distribution of galaxies can require including multiple gas components inside the galaxy, with different ionization states. On the other hand, ratios of lines from different species, such as [OII]/Hα are more sensitive to metallicity [42]. The impact of dust reddening on line ratios should depend mostly on the separation in frequency of the two lines. Since metallicity, dust properties and ionization state vary from galaxy to galaxy and even inside a single galaxy, they will cause a partial decorrelation of the luminosities in the various lines. This degree of decorrelation determines the the amplitude of the line cross-power spectrum in the shot noise regime.

Observationally, measurements of the multi-line LFs, i.e. joint measurements of the luminosities in several lines for the same galaxy sample, are needed to assess this degree of correlation.

4.1.1 Mehta model

Using data from the Hubble WFC3 Infrared Spectroscopic Parallel Survey (WISP), ref. [61] provides a fitting function for the joint galaxy luminosities in Hα and [OIII] over the redshift

range $z = 0.8 - 1.2$. They model the [OIII] luminosity function with a realistic Schechter form, and model the probability distribution for the Hα of a galaxy, given its [OIII] luminosity, as lognormal. From this fitting function, we compute the correlation coefficient between the line luminosities in Hα and [OIII], finding $r = 0.65$ in that redshift bin. Such bivariate LF measurements are crucially needed in order to predict the LIM cross-power spectra.

4.1.2 GOALS survey

We make use of the publicly available catalog of line fluxes from the Great Observatories All-sky Luminous Infrared Galaxy Survey (GOALS) [27]. This catalog contains [CII] 158μm, [NII] 122μm, [OI] 63μm and [OIII] 88μm luminosities for 240 galaxies, observed with the Herschel Photodetector Array Camera and Spectrometer (PACS). These lines are some of the most important atomic fine-structure lines in the far IR/sub-mm range of $50 - 200\mu$m. In this sample, 239 galaxies have measured [CII] line fluxes, 77 galaxies have [NII], 239 have [OI] and 161 have [OIII].

These galaxies include[4] approximately 180 Luminous Infrared Galaxies (LIRGs) with total IR luminosity $L_{IR} \in [10^{11}L_\odot, 10^{12}L_\odot]$ and more than 20 Ultra-Luminous Infrared Galaxies (ULIRGs) with $L_{IR} \geq 10^{12}L_\odot$. They were selected from the IRAS Revised Bright Galaxy Sample (RBGS; [91]) with the aim of "provid[ing] an unbiased picture of the processes responsible for enhanced infrared emission in the local Universe, and [to be] excellent analogs for comparisons with infrared and sub-millimeter selected galaxies at high-redshift." The RBGS sample includes 629 of the brightest sources at $60\,\mu$m, with flux densities above $5.24\,$Jy and Galactic latitude above $5°$. "The LIRGs and ULIRGs targeted in GOALS span the full range of nuclear spectral types (type-1 and type-2 AGN, LINERs, and starbursts) and interaction stages (major mergers, minor mergers, and isolated galaxies)." The sample has a median redshift $z_{med} = 0.0215$, and spans $z_{min} = 0.0030$ to $z_{max} = 0.0918$. From this data, we estimate the line covariances and their correlation coefficients, shown in Fig. 12. We neglect the scatter from measurement error, which was found to be smaller [27].

4.1.3 EGG simulator

In order to estimate the degree of correlation we might expect between different lines, we also make use of the "empirical galaxy generator" (EGG[5]) [28]. EGG works by sampling galaxy stellar masses and redshifts from observed stellar mass functions for quiescent and star forming galaxies. For each galaxy, the stellar mass, redshift and star forming or quiescent nature then determine the star formation rate, the UVJ colors and the panchromatic broadband of the galaxy, through empirical prescriptions. The following emission lines are then simulated: CII $157\,\mu$m, NII $205\,\mu$m, CI $609\,\mu$m, CO1-0 $2.6\,$mm, CO2-1 $1.3\,$mm, CO3-2 $867\,\mu$m, CO4-3 $650\,\mu$m, CO5-4 $520\,\mu$m, CO6-5 $434\,\mu$m, CO7-6 $372\,\mu$m, Hα $656.3\,$nm, Hβ $486.1\,$nm, Hγ $434.1\,$nm, Hδ $410.2\,$nm, NII $658.4\,$nm, NII $654.8\,$nm, OIII $500.7\,$nm, OIII $495.9\,$nm, OII $372.7\,$nm, Lyα $121.6\,$nm. This includes all the lines that are relevant for LIM with SPHEREx, specifically: Hα $656.3\,$nm, Hβ $486.1\,$nm, [OIII] $500.7\,$nm and Lyα $121.6\,$nm.

The emission line modeling is described in ref. [92]. The Hα luminosity is assumed to scale with the star formation rate of the galaxy, with a fixed random scatter in the Kennicutt relation [93]. The other lines in the Balmer series (Hβ, Hγ, Hδ) are then inferred from Hα using fixed line ratios from case-B recombination, without scatter. Luminosities in [OIII] and [NII] are then inferred from the Hα and Hβ luminosities using empirical relations, with a

---

[4]See http://goals.ipac.caltech.edu/ for more details.
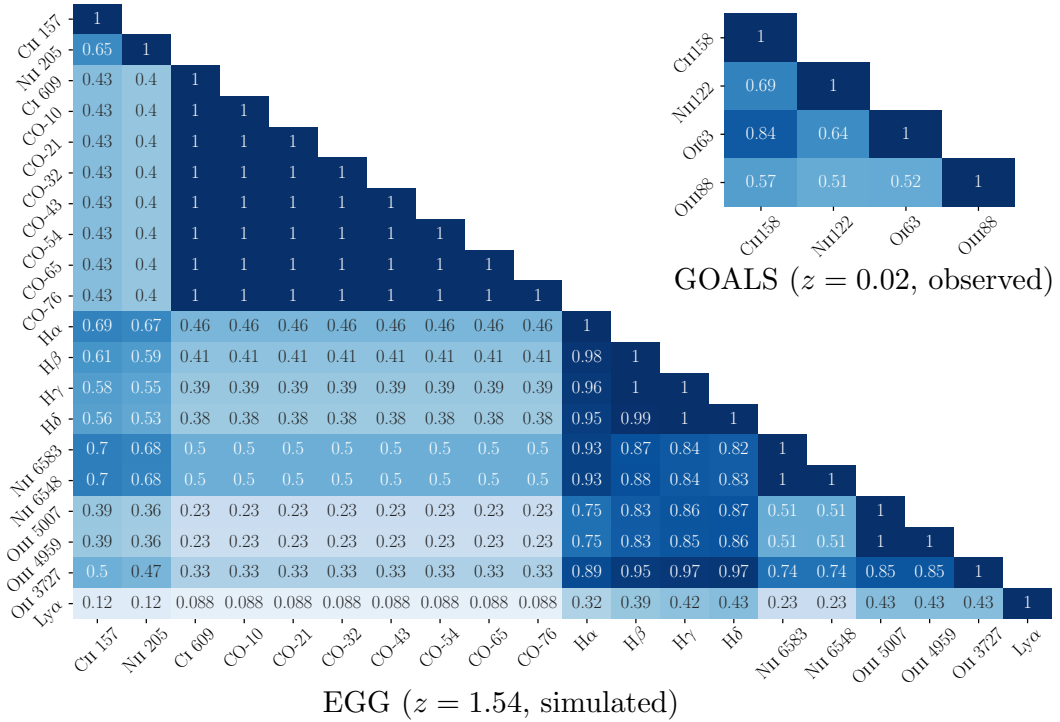
[5]http://cschreib.github.io/egg/

Figure 12. Galaxies emits light in various different lines. This figure shows the correlation coefficient between the line luminosities of a given galaxy in two different lines. These correlation coefficients are a function of redshift only, and were estimated from the GOALS survey at $z = 0.02$ (top right) and the EGG simulator at $z = 1.54$ (bottom left). The EGG correlation pattern is qualitatively similar out to $z = 4$. These correlation coefficients at the individual galaxy level are a key input to determine the scale-dependent correlation coefficient between two line intensity maps.

mean and scatter dependent on the galaxy metallicity. These relations were built to match the observed BPT diagrams [94], where available, between redshifts 0 and 2. A similar approach is adopted for [OII], inferred from Hβ via an empirical relation where the mean and scatter are functions of metallicity. For Lyα, the line luminosity is chosen proportional to Hα, with a fixed proportionality constant, but is subsequently attenuated with a variable escape fraction. This modeling produces a scatter in the various line luminosities, with a reasonable level of correlation, and is thus appropriate for our study.

By contrast, the modeling of the far IR lines is more simplified. For instance, the [CII] and [NII] luminosities scale with the total IR luminosity of the galaxy. The CO and CI lines scale with the gas mass in the galaxy. While no scatter is introduced between the various CO transitions, some scatter in present between the CO, [CII] and [NII] luminosities.

To simulate the galaxy population seen by a LIM experiment, we run EGG with the following options:

```
egg-gencat verbose area=1 mmin=8 mmax=14 zmin=0.05 zmax=10
```

This corresponds to a stellar mass limited sample, with $M_\star \in [10^8 M_\odot, 10^{14} M_\odot]$, over the redshift range $z \in [0.05, 10]$ and a sky area of $1\,\mathrm{deg}^2$. This produces about one million galaxies split between 109 redshift slices.

4.2 LIM cross-spectra and decorrelation

One may want to cross-correlate different line lintensity maps, e.g. to suppress line interlopers. In this case, knowing the amount of correlation between the two intensity maps is crucial. This is quantified by the scale-dependent correlation coefficient:

$$r_{1,2}(k,\mu,z) \equiv \frac{P_{1,2}(k,\mu,z)}{\sqrt{P_{1,1}(k,\mu,z)P_{2,2}(k,\mu,z)}} \tag{4.1}$$

On large scales where the 2-halo term dominates, this ratio is unity $r_{1,2}(k,\mu,z) = 1$ and the two intensity maps are perfectly correlated. On smaller scales, where the 1-halo or shot noise term dominates, the power spectrum is of the form

$$P = \frac{\bar{I}_1\bar{I}_2}{\bar{n}^{\text{eff}}} = \bar{I}_1\bar{I}_2\frac{1}{\bar{n}}\left(1+\sigma_{1,2}^2\right) \tag{4.2}$$

where $\bar{n}^{\text{eff}}$ and $\bar{n}$ are the effective and mean number density of halos or galaxies and $\sigma_{1,2}^2$ the halo or galaxy line noise contributions in the 1-halo and shot noise regimes, respectively. As explained in §2.3, these quantities are "cutoff dependent" and must be defined with care. If the line noise is negligible, $\sigma_{1,2}^2 \ll 1$, then the two intensity maps are again perfectly correlated: $r_{1,2}(k,\mu,z) = 1$. However, if the noises are large, $\sigma_{1,1}^2, \sigma_{1,2}^2, \sigma_{2,2}^2 \gg 1$, then the correlation coefficient drops to

$$r_{1,2}(k,\mu,z) = \frac{\sqrt{\bar{n}_{1,1}^{\text{eff}}\ \bar{n}_{2,2}^{\text{eff}}}}{\bar{n}_{1,2}^{\text{eff}}} \simeq \frac{\sigma_{1,2}^2}{\sigma_{1,1}\sigma_{2,2}}, \tag{4.3}$$

i.e. the correlation coefficient between the two intensity maps is equal to the correlation coefficient of the halo (galaxy) luminosities in lines 1 and 2, in the 1-halo (shot noise) regime. Thus, a large scatter in the line luminosities of halos or galaxies can cause a decorrelation between two intensity maps.

In the 1-halo regime, assuming $L_i(m) \propto \text{SFR}^{\gamma_i}(m,z)$, the correlation coefficient between lines 1 and 2 becomes:

$$r_{1,2}(k \to 0,\mu,z) = \frac{\int dm\ n(m)\ \text{SFR}^{\gamma_1+\gamma_2}(m)}{\sqrt{\int dm\ n(m)\ \text{SFR}^{2\gamma_1}(m)\int dm\ n(m)\ \text{SFR}^{2\gamma_2}(m)}}. \tag{4.4}$$

In the 1-halo regime, all lines with the same power law index $\gamma = 1$ relating luminosity to SFR will thus be perfectly correlated. This is typically the case of the UV and optical lines. For far IR lines like NII $122\,\mu$m, NIII $58\,\mu$m, CII $158\,\mu$m, which scale as star formation rate to the power $0.78 - 1.02$ according to ref. [53], the correlation coefficient is $r > 0.96$ out to redshift 5. The decorrelation between these lines in the 1-halo regime is thus negligible. However, CO has $\gamma = 0.6$, leading to a significant decorrelation with the other lines considered in the 1-halo regime.

In the shot noise regime, the joint luminosity function for Hα and [OIII] in ref. [61] predicts a correlation coefficient between these lines of $r = 0.65$. In Fig. 12, we show the correlation coefficients of line luminosities for an individual galaxy, as predicted by the EGG simulator and the as measured in the GOALS data. The correlation coefficients for line luminosities from individual galaxies are $r = 0.088 - 1$ from EGG and $r = 0.5 - 1$ from GOALS, depending on the lines considered. Astrophysically, these correlation coefficients are

a direct consequences of the physical processes inside galaxies responsible for the produceing the emission lines (Sec. 2.1).

Given the correlation coefficients at the level of individual galaxies, we can compute the scale-dependent correlation coefficients between any pair of intensity maps. We show them for a pair of highly-correlated lines (Hα, [OIII]) and for a pair of low-correlation lines (Ly-α, CO) in Fig. 13.

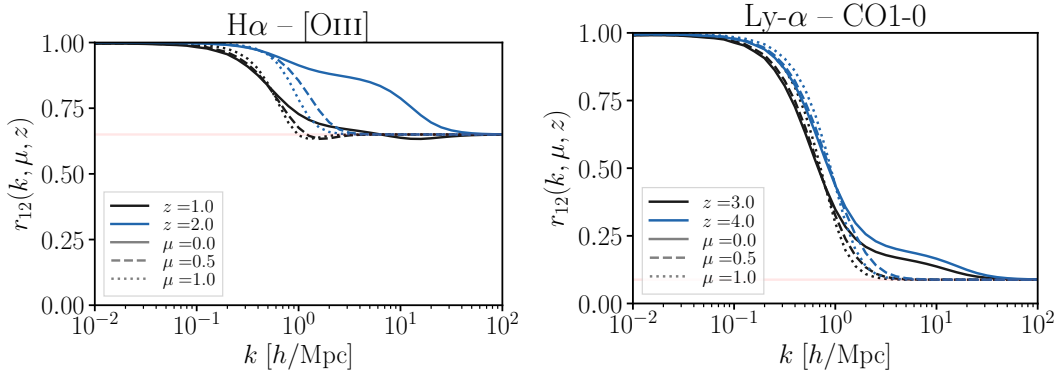**Figure 13**. Generic evolution of the correlation coefficient between line intensity maps. We consider the correlation betwewen Hα [62] and [OIII] [61] (left panel), and between Ly-α [81] and CO [79] (right panel). On large scales, where the 2-halo term dominates for both lines, the two maps are perfectly correlated. On small scales, where both maps are shot noise dominated, we recover the line luminosity correlation coefficient (red line) $r = 0.65$ (left, inferred from [61]) or $r = 0.088$ (right, from [28]). On intermediate scales, if the 1-halo terms dominate, the correlation coefficient is close to unity. This occurs at $z = 2$ for Hα and [OIII]. If instead the 1-halo and shot noise terms are comparable, e.g. at $z = 1$ and $k = 1h/\text{Mpc}$ for Hα – [OIII] (left panel), the correlation coefficient of the sum of 1-halo and shot noise can actually be lower than that of the shot noise alone ($r = 0.65$). In both panels, the FOG effect suppresses the 1-halo term along the LOS, bringing the correlation coefficient down to the shot noise value. This occurs at $k = 1 - 20h/\text{Mpc}$ both for Hα – [OIII] (left panel) and Ly-α – CO1-0 (right panel).

4.3 Implications

The value of the correlation coefficient computed above has several implications:

1. **Estimating the auto-spectrum from cross-spectra only.** Cross-spectra allow reduced foreground contamination. In ref. [12], cross-spectra of 21cm, [CII] and [OIII] lines are therefore combined to estimate the 21cm power spectrum:

$$\hat{P}_{aa} \equiv \frac{P_{ab}P_{ac}}{P_{bc}} = P_{aa}\frac{r_{ab}r_{ac}}{r_{bc}}. \tag{4.5}$$

Ref. [12] models the line emission from simulated halos to estimate the level of decorrelation between these lines, and how it limits the method. Here, we instead use an analytical halo model to predict the line correlation coefficients, including not only the effect of individual halos (1-halo term) but also that of individual galaxies (shot noise term). We further include the effect of RSD on the correlation coefficient. Additional noise or foregrounds, uncorrelated between intensity maps a, b and c, will drop out of the cross-correlations. As a result, the relevant correlation coefficients in Eq. (4.5) are

still the noiseless, foreground free ones. On the other hand, any foreground component which contaminates two maps, e.g., a and b, will bias the cross-correlation P_{ab}. This will thus bias the estimated auto-spectrum $\hat{P}_{aa}$.

2. **Using a line at a different frequency as a foreground tracer.** Consider a simple model where the measured line intensity at any frequency is the sum of a target (t) line and a foreground (f) line. We can produce intensity maps at two frequencies ν_1 and ν_2, such that the foreground emission in ν_1 comes from the same redshift as the target emission in ν_2. In other words:

$$
\begin{cases}
I_{\nu_1} = t(z_0) + f(z_1) \\
I_{\nu_2} = t(z_1) + \underbrace{f(z_2)}_{\text{ignore}} .
\end{cases}
\tag{4.6}
$$

How well can one subtract the foreground emission in I_{ν_1} using I_{ν_2}? We use a linear combination:

$$
\begin{aligned}
\hat{t}(z_0) &= I_{\nu_1} - \alpha I_{\nu_2} \\
&= t(z_0) + \underbrace{[f(z_1) - \alpha t(z_1)]}_{\text{to be reduced}}
\end{aligned}
\tag{4.7}
$$

where the coefficient α that minimizes the power spectrum of the square bracket above is $\alpha = P_{12}/P_{22} = P_{tf}(z_1)/P_{tt}(z_1)$. The residual power spectrum of the square bracket is then $P_{ff}(z_1)\left[1 - r_{tf}^2(z_1)\right]$. In other words, the maximum reduction of the foreground power spectrum achievable is entirely determined by $1 - r_{tf}^2(z_1)$. Equivalently, the total power spectrum of the estimated target line is $P_{\hat{t}} = P_{11}\left[1 - r_{12}^2\right]$, where the maximum power reduction is determined by $\left[1 - r_{12}^2\right]$. We show the corresponding reduction factors in Fig. 14.
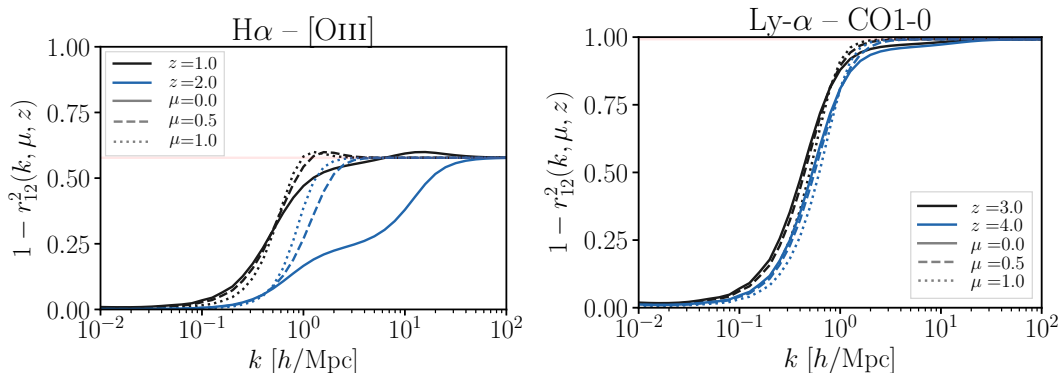

Figure 14. Factor $1 - r^2$, which determines the ability to use one line as a tracer of the other, in order to subtract foregrounds/interlopers. Values near zero indicate good subtraction while values near unity indicate ineffective subtraction.

3. **Astrophysics of star formation** The lines considered in this paper all trace the star formation rate in galaxies. However, they are produced in different regions of the interstellar medium, with different densities, temperatures, ionization fractions and chemical compositions. Thus, the correlation coefficient between two different line

luminosities, for a given galaxy, informs us about the correlated abundance of these ISM regions, in a way that would be impossible by measuring each line separately. This idea was explored in [82, 95, 96]. By connecting the correlation coefficient between various at the galaxy level to the correlation coefficient observable at the intensity map level, our formalism enables extracting this astrophysical information from LIM.

5 Conclusions

Line intensity mapping (LIM) has become a new frontier for studying the Universe, with early observations giving way to dedicated surveys. In this paper, we have provided a new halo model formalism, consistently capturing the various sources of galaxy and halo luminosity fluctuations and their effects on the LIM power spectrum.

To this end we have generalized the conditional luminosity function (CLF) formalism of ref. [29] to a multi-line CLF (§2.2). This extension produces a new, general and flexible halo model, which predicts not only the mean intensity and auto-spectrum of a line intensity map, but also cross-spectra of different lines, and the cross-spectra of LIM with galaxy catalogs and galaxy and CMB lensing (Appendices. A and B). We have provided the corresponding expressions for the mean, 2-halo, 1-halo and shot noise terms, in 3D redshift-space and for 2D projected maps (Appendices A-B).

The formalism highlights the similarity between the 1-halo and shot noise terms, both arising from the (approximately) Poissonian shot noise of halos and galaxies, respectively. In galaxy clustering, they are simply given by the inverse number density of halos and galaxies. In LIM, the CLF-based halo model shows that they are enhanced by a term we refer to as halo and galaxy "line noises", respectively (§2.3). We showed that the halo line noise is entirely sourced by luminosity fluctuations across halo masses, and is insensitive to luminosity fluctuations at fixed halo mass. The galaxy line noise is instead sourced by all galaxy luminosity fluctuations, i.e. $\langle L_{\rm gal}^2 \rangle - \langle L_{\rm gal} \rangle^2$. We thus defined LIM-relevant halo and galaxy "effective number densities", accounting for these effects. By tracking the sources of luminosity scatter, we showed that the amplitude of the 1-halo term does not necessarily exceed that of the shot noise on large scales, for Poissonian halos and galaxies (§2.5). We showed examples of lines for either case. This has implications for interloper rejection methods relying on the sparsity of the sources [97] – halos in the 1-halo regime and galaxies in the shot noise regime [32].Because it encodes the imperfect correlation between different line luminosities from the same galaxy, the multi-line CLF allows us to quantify the level of decorrelation between two LIMs from the same redshift (§4.1). This is a crucial modeling improvement, since many LIM analyses rely on cross-spectra to subtract interlopers or foregrounds, or to estimate auto-spectra.Furthermore, by connecting LIM cross-correlations to the galaxy-level properties, the multi-line CLF allows us to constrain properties of the galaxy interstellar media inaccessible with auto-spectra alone.

To evaluate the halo model, we have relied on observations and simulations (§3). We used observed luminosity functions for Hα [59, 60, 62], [O III] [60, 61], Lyman-α [81], CO [79] and [C II] [79]. While most observations and simulations focus on the single-line LF or CLF, modeling LIM cross-spectra crucially requires knowledge of the multi-line CLF. To estimate the correlation between different lines (§4.1), we relied on measurements from [61] for Hα and [O III] and the GOALS survey [27] for far infrared lines, as well as simulation results from EGG [28, 92]. We encourage semi-analytic models (SAMs), simulation and observational studies to output their multi-line CLFs whenever possible. Fitting functions for them, from simulations

such as [66], would be a valuable input for the halo model cross-spectra. In return, these simulations could be used to check the halo model prediction.

Other applications of the CLF are possible. A valuable extension would be to extend this formalism to also predict higher-point correlators of LIM, or joint multi-line voxel intensity distributions [46].

We have made our code publicly available at https://github.com/EmmanuelSchaan/HaloGen/tree/LIM. This modular code computes the LIM mean intensity, auto- and cross-spectra in 3d redshift-space and 2d projections. A key input to this code is the set of luminosity functions for the various lines. The uncertainty in these LFs at high redshift is very large at present, leading to a large uncertainty in the predicted power spectra. Power spectra from this code should thus not be considered ground truth, and may differ significantly from predictions in the literature based on different input luminosity functions. For this reason, the code was made highly modular, allowing the user to easily switch input (luminosity functions, star formation rate, etc.), making it a useful prediction tool for LIM forecasts.

Acknowledgments

A   Halo model

A.1   Pedagogical derivation of observables and a halo model

Since there are many expressions and approximations made in different papers in the literature, in this appendix we present a pedagogical derivation of our basic observables and an implementation of the halo model tailored to LIM. We have attempted to be clear about the approximations and assumptions made at each stage. Our approach builds upon the extensive earlier work in this subject, including refs. [7, 50, 52, 98–103] and the papers cited therein.

For a galaxy with line luminosity L_1 ([W]) in emission line 1, the observed line flux ([W m^{-2}]), integrated over the frequency width of the line, is $L_1/\left(4\pi D_L^2\right) = L_1/\left(4\pi\chi^2(1+z)^2\right)$, where have assumed a spatially flat Universe with comoving radial distance $\chi(z)$.

The sky specific intensity I_ν ([W m^{-2} sr^{-1} Hz^{-1}]), measures the sky flux per spatial resolution element $d\Omega$ ([sr]) and frequency resolution element $d\nu$ ([Hz]) of the LIM experiment. The narrow emission lines are typically not spectrally resolved, so the observed specific

---

[6]https://l2s2.sciencesconf.org/

intensity at $\nu = \nu_1^0/(1+z)$ includes the full line luminosity L_1 from all galaxies at redshift z:

$$I_1 = \underbrace{\frac{1}{\int d\nu d\Omega} \int d\nu d\Omega}_{\substack{\text{average in spatial and} \\ \text{spectral resolution element}}} \underbrace{\frac{d\chi}{d\nu} \int dL_1 \frac{dN_{\text{gal}}}{d\Omega d\chi dL_1}}_{\substack{\text{galaxy number per} \\ \text{resolution element} \\ [\text{sr}^{-1}\ \text{Hz}^{-1}]}} \underbrace{\frac{L_1}{4\pi D_L}}_{\substack{\text{galaxy line flux} \\ [\text{W m}^{-2}]}} \qquad [\text{W m}^{-2}\ \text{sr}^{-1}\ \text{Hz}^{-1}]$$

$$= \frac{dV}{d\Omega d\chi} \frac{d\chi}{d\nu} \int dL_1 \frac{dN_{\text{gal}}}{dV dL_1} \frac{L_1}{4\pi D_L} \qquad\qquad\qquad\qquad (\text{A.1})$$

$$= \chi^2 \frac{c\,(1+z)^2}{\nu_1^0 H(z)} \int dm\ n(m) \int dL_1\ \phi(L_1|m) \frac{L_1}{4\pi\chi^2(1+z)^2}$$

$$= \frac{1}{4\pi\nu_1^0} \frac{c}{H(z)} \int dm\ n(m) \int dL_1\ \phi(L_1|m)\ L_1.$$

where in the last two lines the first integrals are over halo mass, m, and the halo mass function, $n(m)$ – the number of halos in mass range $(m; dm)$ per comoving volume – and $\phi(L_1|m)$ is the conditional luminosity function defined in §2.2. Hence the mean specific intensity is

$$\bar{I}_1 = \frac{1}{4\pi\nu_1^0} \frac{c}{H(z)}\ \bar{n}_{\text{gal}} \bar{L}_1 \qquad [\text{W m}^{-2}\ \text{sr}^{-1}\ \text{Hz}^{-1}], \qquad\qquad (\text{A.2})$$

with:

$$\begin{cases} \bar{n}_{\text{gal}} = \displaystyle\int dm\ n(m) \int dL_1\ \phi(L_1|m) & \text{Mean galaxy number density} \qquad [\text{cMpc}^{-3}] \\[2mm] \bar{L}_1 = \dfrac{1}{\bar{n}_{\text{gal}}} \displaystyle\int dm\ n(m) \int dL_1\ \phi(L_1|m)\ L_1 & \text{Mean line 1 luminosity} \qquad\qquad\qquad [\text{W}] \end{cases}$$

$$(\text{A.3})$$

In this halo model, the local number density of halos $n(m,x)$ is a biased, Poisson sampling of the underlying linear matter density field. This produces the 2-halo and 1-halo terms in the power spectrum:

$$\langle \delta n(m,k) \delta n(m',k') \rangle = (2\pi)^3 \delta^D(k+k')\,\bar{n}(m)\bar{n}(m') \left[\underbrace{b(m)b(m')P_{\text{lin}}(k)}_{\text{2-halo term}} + \underbrace{\delta^D(m-m')\frac{1}{\bar{n}(m)}}_{\text{1-halo term}} \right],$$

$$(\text{A.4})$$

where the 1-halo term is nothing but the shot noise of halos, and the 2-halo term is related to the linear matter density field through a deterministic, linear, scale independent and mass dependent halo bias. This bias assumption is expected to be valid on large scales where the 2-halo term is significant.

We further assume that within a given halo, the number density of galaxies is a Poisson sampling of the (normalized) NFW profile [104], $u(k,m)$, of the halo. We currently lack enough observational constraints on satellite vs. central galaxy line emission to include this distinction in our models, and we will therefore ignore this distinction here. However, distinguishing satellites and centrals can be done easily in our formalism, as we show in

Appendix A.2. This Poisson process produces a galaxy shot noise term:

$$
\begin{aligned}
\langle \bar{n}(m)\delta\phi(L_1, L_2, k|m)\, \bar{n}(m')\delta\phi(L_1', L_2', k'|m')\rangle &= (2\pi)^3\, \delta^D(k + k') \\
&\times \delta^D(m - m')\delta^D(L_1 - L_1')\delta^D(L_2 - L_2') \\
&\times \underbrace{\bar{n}(m)\phi(L_1, L_2|m)}_{\text{galaxy shot noise}}.
\end{aligned}
\tag{A.5}
$$

Finally, we compute the specific intensity power spectrum in redshift space, following [7, 50, 52, 98–103]. We take into account the effect of FOG on the halo profile by substituting:

$$
u^{(s)}(k, \mu, m, z) = u(k, m)e^{-k^2\mu^2\sigma^2(m)/2}
\tag{A.6}
$$

where the velocity dispersion σ is possibly mass-dependent. The Kaiser effect follows from the conservation of the luminosity in redshift-space:

$$
\frac{\delta I_1^{(s)}(k)}{\bar{I}_1} = -\nabla \cdot \psi^{(s)}(k) = -\nabla \cdot \psi(k) + f\mu^2\delta_m(k) = \frac{\delta I_1(k)}{\bar{I}_1} + f\mu^2\delta_m(k).
\tag{A.7}
$$

Here we introduced the Lagrangian displacement, ψ, which sources the matter density fluctuations $(\delta_m + \nabla \cdot \psi = 0)$ and its redshift-space counterpart $\psi^{(s)} = \psi + f\,(\psi \cdot \hat{n})\,\hat{n}$.

With all the ingredients above, we start with the 2-halo term, where the halo number density fluctuations trace the matter density field. This 2-halo term corresponds to the first term in Eq. (A.4), and can be computed by taking the power spectrum of the "2-halo" field:

$$
\begin{aligned}
\delta I_1^{\text{2h}} &= \frac{1}{4\pi\nu_1^0}\frac{c}{H(z)} \\
&\times \left(\int dm\, n(m) \left[b(m) + \underbrace{f\mu^2}_{\text{Kaiser}} \right] u(k, m)\underbrace{e^{-k^2\mu^2\sigma^2(m)/2}}_{\text{FOG}} \int dL_1\phi(L_1|m)\, L_1 \right) \delta_{\text{lin}}(z) \\
&= \bar{I}_1 \left[b_1(k, \mu, z) + F(k, \mu, z)\mu^2 \right] \delta_{\text{lin}},
\end{aligned}
\tag{A.8}
$$

and therefore

$$
P_{1,2}^{(\text{s})\,\text{2-halo}}(k, \mu, z) = \bar{I}_1\bar{I}_2 \left[b_1 + F\mu^2 \right] \left[b_2 + F\mu^2 \right] P_{\text{lin}}
\tag{A.9}
$$

where we have defined the relevant physical parameters:

$$
\begin{cases}
b_j(k, \mu, z) \equiv \dfrac{1}{\bar{n}_{\text{gal}}\bar{L}_j} \displaystyle\int dm\, n(m)\, N_{\text{gal}}(m)L_j(m)\, b(m)\, u(k, m)e^{-k^2\mu^2\sigma^2/2} \\[2ex]
F(k, \mu, z) \equiv f \displaystyle\int dm\, n(m)\, \left(\dfrac{m}{\bar{\rho}} \right) u(k, m)e^{-k^2\mu^2\sigma^2/2}
\end{cases}
\tag{A.10}
$$

On scales larger than σ, such that the FOG effect is negligible, and on scales larger than the halo size, such that the halo profile $u(k, m) \simeq 1$, the bias b_j becomes the scale-independent large-scale bias of linear theory, and F tends to the usual logarithmic growth rate f. On smaller scales, both the bias b_j and effective growth rate F become scale-dependent and anisotropic.

Focusing now on the 1-halo term, due to the Poisson shot noise of halos, we use Eq. (A.4) to obtain:

$$P_{1,2}^{\text{(s) 1-halo}}(k, \mu, z) = \int dm\, n(m)\, |u(k,m)|^2\, e^{-k^2 \mu^2 \sigma^2(m)} \left[\int dL_1 \phi(L_1, m) L_1 \right] \left[\int dL_2 \phi(L_2, m) L_2 \right]$$
$$= \bar{I}_1 \bar{I}_2 \frac{U_{1,2}^2(k, \mu, z)}{\bar{n}_{1,2}^{\text{h,eff}}},$$
(A.11)

where we further introduced the line-pair-weighted, squared halo profile

$$U_{1,2}^2(k, \mu, z) \equiv \frac{\int dm\, n(m)\, |u(k,m)|^2\, e^{-k^2 \mu^2 \sigma^2} L_1(m) L_2(m)}{\int dm\, n(m)\, L_1(m) L_2(m)}$$
(A.12)

which tends to 1 as $k \to 0$ and the total halo number density,

$$\bar{n}_{1,2}^{\text{h,eff}} \equiv \frac{\left(\int dm\, n(m) L_1(m) \right) \left(\int dm\, n(m) L_2(m) \right)}{\int dm\, n(m) L_1(m) L_2(m)} \qquad [\text{cMpc}^{-3}] \quad .$$
(A.13)

Finally, the galaxy shot noise, coming from Poisson fluctuations in the number of galaxies within a halo, Eq. (A.5), becomes:

$$P_{1,2}^{\text{shot}}(z) = \frac{1}{(4\pi)^2 \nu_1^0 \nu_2^0} \left(\frac{c}{H(z)} \right)^2 \int dm\, n(m) \int dL_1 dL_2\, \phi(L_1, L_2, m)\, L_1 L_2$$
$$= \frac{\bar{I}_1 \bar{I}_2}{\bar{n}_{1,2}^{\text{gal,eff}}} \quad ,$$
(A.14)

with

$$\bar{n}_{1,2}^{\text{gal eff}} = \frac{\left(\int dm\, n(m) L_1(m) \right) \left(\int dm\, n(m) L_2(m) \right)}{\int dm\, n(m) \int dL_1 dL_2\, \phi(L_1, L_2|m) L_1 L_2} \qquad [\text{cMpc}^{-3}] \quad .$$
(A.15)

A.2 Halo occupation distribution: centrals and satellites

In this appendix, we show that our halo model based on the CLF naturally describes the distinction between satellite and central galaxies. As shown in [105], the particular choice of halo occupation distribution for the line-emitting galaxies can have a significant impact on the 1-halo LIM power spectrum. The central galaxies are assumed to be at the centers of halos, with no velocity with respect to the halo, and therefore no FOG effect, whereas the satellites are assumed to sample the NFW matter density profile of the halo, and are alone responsible for the FOG. We define their numbers $N_{\text{cen}}(m, z), N_{\text{sat}}(m, z)$ and mean luminosities $L_j^{\text{cen}}(m, z) L_j^{\text{sat}}(m, z)$ in a halo of mass m, at redshift z. The mean intensity is then simply the sum of the satellite and central luminosities:

$$\bar{I}_j = \frac{1}{4\pi \nu_1^0} \frac{c}{H(z)}\, \left[L_j^{\text{cen}} N_{\text{cen}} + L_j^{\text{sat}} N_{\text{sat}} \right].$$
(A.16)

The 2-halo term is given by:

$$P_{1,2}^{\text{2-halo}}(k, \mu, z) = \bar{I}_1 \bar{I}_2 \left(b_1 + F\mu^2 \right) \left(b_2 + F\mu^2 \right)\, P_{\text{lin}},$$
(A.17)

where the bias is

$$b_j(k, \mu, z) \equiv \mathcal{L}_j^{-1} \int dm\, n(m)\, b(m)\, \left[L_j^{\text{cen}} N_{\text{cen}} + L_j^{\text{sat}} N_{\text{sat}} u(k, m, z) e^{-k^2 \mu^2 \sigma_d^2(m)/2} \right]$$
(A.18)

and the effective growth rate of structure is

$$F(k, \mu, z) \equiv f \int dm \, n(m) \, \left(\frac{m}{\bar{\rho}} \right) u(k, m) e^{-k^2 \mu^2 \sigma_d^2(m)/2}. \tag{A.19}$$

The 1-halo term becomes:

$$P_{i,j}^{\text{1-halo}}(k, \mu, z) = \left(\frac{c}{4\pi H(z)} \right)^2 \frac{1}{\nu_i^0 \nu_j^0} \int dm \, n(m) \, \left[L_j^{\text{sat}\,2} N_{\text{sat}}^2 \, |u(k, m, z)|^2 \, e^{-k^2 \mu^2 \sigma_d^2(m)} \right.$$
$$\left. +2N_{\text{sat}} N_{\text{cen}} L_j^{\text{sat}} L_j^{\text{cen}} e^{-k^2 \mu^2 \sigma_d^2(m)/2} \right]. \tag{A.20}$$

Finally, the shot noise term is unchanged, since it is not affected by the spatial distribution of the galaxies (centrals vs. satellites) nor by their peculiar velocities.

A.3 Impact of luminosity fluctuations on the halo model

The distribution of galaxy luminosities has a non-zero scatter around its mean. How does this scatter affect the 2-halo, 1-halo and shot noise terms? In this appendix, we show explicitly that the 2-halo and 1-halo terms depend only on the mean galaxy luminosity, whereas the shot noise term is enhanced by the scatter in galaxy luminosities.

To better understand this, let us consider one pixel in configuration space. For now, we consider only halos of a fixed mass, with number N_h^{pix} within the observed pixel. We want to compute the variance of the pixel luminosity L, to show how the 2-halo, 1-halo and galaxy shot noise terms arise, and to show that the fluctuations in galaxy luminosities only affect the galaxy shot noise term, not the 1-halo term. The statistics of the number of halos in the pixel is assumed to be Poissonian:

$$\mathcal{P}(N_h^{\text{pix}}) \sim \text{Poiss}(N_h^{\text{pix}} | \bar{N}_h^{\text{pix}}) \tag{A.21}$$

Each halo contains a certain number N_g of galaxies. For a given halo, the number of galaxies in the halo is Poissonian:

$$\mathcal{P}(N_g) \sim \text{Poiss}(N_g | \bar{N}_g) \tag{A.22}$$

The total number of galaxies N_g^{pix} in the pixel is the sum of the number of galaxies in all of the halos in the pixel. A sum of a fixed number of Poisson variables is Poissonian. However, the sum of a Poisson number of Poisson variables is not Poissonian! In other words:

$$\mathcal{P}(N_g^{\text{pix}}) \neq \text{Poiss}(N_g^{\text{pix}} | \bar{N}_h^{\text{pix}} \bar{N}_g)$$
$$\text{but} \quad \mathcal{P}(N_g^{\text{pix}} | N_h^{\text{pix}}) \sim \text{Poiss}(N_g^{\text{pix}} | N_h^{\text{pix}} \bar{N}_g) \tag{A.23}$$

Here the presence or absence of a bar over N_h^{pix} matters. Finally, each galaxy in a halo has a luminosity L_g sampled independently from a probability distribution $\mathcal{P}_1(L_g)$. This probability distribution is the same object as ϕ, but normalized to integrate to unity. For the power spectrum or the variance, all we care about is that this distribution has a mean $\bar{L}_g$ and a variance which we write as $\bar{L}_g^2 \sigma_g^2$.

Finally, the total luminosity L in the pixel is the sum of the luminosities L_g of each of the N_g^{pix} galaxies in the pixel. Its probability distribution is complex. However, at fixed N_g^{pix}, it is simply $\mathcal{P}_1^{*N_g^{\text{pix}}}$, i.e. the convolution of $\mathcal{P}_1$ with itself N_g^{pix} times. For the power

spectrum, all we care about is that $\mathcal{P}_1^{*N_g^{\mathrm{pix}}}$ has mean $N_g^{\mathrm{pix}}\bar{L}_g$ and variance $N_g^{\mathrm{pix}}\bar{L}_g^2\sigma_g^2$. Again, the absence of a bar on N_g^{pix} matters here. In other words:

$$
\begin{aligned}
\langle L^2|N_g^{\mathrm{pix}}\rangle &= \langle L|N_g^{\mathrm{pix}}\rangle^2 + \mathrm{var}[L|N_g^{\mathrm{pix}}] \\
&= \bar{L}_g^2\left[N_g^{\mathrm{pix}\,2} + N_g^{\mathrm{pix}}\sigma_g^2\right]
\end{aligned}
\tag{A.24}
$$

Putting all this together, we can compute the variance of the pixel luminosity L:

$$
\begin{aligned}
\langle L^2\rangle &= \sum_{N_g^{\mathrm{pix}}} \mathcal{P}(N_g^{\mathrm{pix}})\langle L^2|N_g^{\mathrm{pix}}\rangle \\
&= \bar{L}_g^2 \sum_{N_g^{\mathrm{pix}}} \mathcal{P}(N_g^{\mathrm{pix}})\left[N_g^{\mathrm{pix}\,2} + N_g^{\mathrm{pix}}\sigma_g^2\right] \\
&= \bar{L}_g^2 \sum_{N_h^{\mathrm{pix}}} \mathcal{P}(N_h^{\mathrm{pix}}) \sum_{N_g^{\mathrm{pix}}} \mathcal{P}(N_g^{\mathrm{pix}}|N_h^{\mathrm{pix}})\left[N_g^{\mathrm{pix}\,2} + N_g^{\mathrm{pix}}\sigma_g^2\right] \\
&= \bar{L}_g^2 \sum_{N_h^{\mathrm{pix}}} \mathcal{P}(N_h^{\mathrm{pix}}) \sum_{N_g^{\mathrm{pix}}} \mathrm{Poiss}(N_g^{\mathrm{pix}}|N_h^{\mathrm{pix}}\bar{N}_g)\left[N_g^{\mathrm{pix}\,2} + N_g^{\mathrm{pix}}\sigma_g^2\right] \\
&= \bar{L}_g^2 \sum_{N_h^{\mathrm{pix}}} \mathcal{P}(N_h^{\mathrm{pix}})\left[(N_h^{\mathrm{pix}}\bar{N}_g)^2 + N_h^{\mathrm{pix}}\bar{N}_g + N_h^{\mathrm{pix}}\bar{N}_g\sigma_g^2\right] \\
&= \bar{L}_g^2 \sum_{N_h^{\mathrm{pix}}} \mathrm{Poiss}(N_h^{\mathrm{pix}}|\bar{N}_h^{\mathrm{pix}})\left[(N_h^{\mathrm{pix}}\bar{N}_g)^2 + N_h^{\mathrm{pix}}\bar{N}_g + N_h^{\mathrm{pix}}\bar{N}_g\sigma_g^2\right] \\
&= \bar{L}_g^2\left[(\bar{N}_h^{\mathrm{pix}}\bar{N}_g)^2 + \bar{N}_h^{\mathrm{pix}}\bar{N}_g^2 + \bar{N}_h^{\mathrm{pix}}\bar{N}_g + \bar{N}_h^{\mathrm{pix}}\bar{N}_g\sigma_g^2\right] \\
&= \bar{L}_g^2\left[\underbrace{(\bar{N}_h^{\mathrm{pix}}\bar{N}_g)^2}_{\text{2-halo}} + \underbrace{\bar{N}_h^{\mathrm{pix}}\bar{N}_g^2}_{\text{1-halo}} + \underbrace{\bar{N}_h^{\mathrm{pix}}\bar{N}_g(1+\sigma_g^2)}_{\text{shot noise + line noise}}\right]
\end{aligned}
\tag{A.25}
$$

This confirms that the fluctuations in galaxy luminosities only appear in the galaxy shot noise term. For the 1-halo and 2-halo terms, everything is as if each halo had the same luminosity $\bar{L}_g\bar{N}_g$, with no variation from halo to halo.

A.4 Multi-line CLF: tracer auto-spectrum

In order to compare the value of galaxy number counts and LIM as tracers of the matter density field, we need to consistently predict their power spectra, using the same CLF halo model. We assume that the galaxy sample is selected based on line luminosity, such that the number of satellites and centrals per halo are determined by the respective CLFs. For example, for galaxies selected to have a line luminosity above a threshold L_{min}, we have $N_{\mathrm{gal}}(m,z) = \int_{L_{\mathrm{min}}}^{\infty} dL\,\Phi(L|m,z)$. Here, we simply review the expressions for the clustering power spectrum in this halo model. For generality, we include the distinction between centrals and satellites. The mean number density of galaxies is simply $\bar{n}_g = \int dm\,n(m)\left[N_{\mathrm{cen}}(m) + N_{\mathrm{sat}}(m)\right]$. The 2-halo term in the power spectrum is given by:

$$
P_g^{\text{2-halo}}(k,\mu,z) = \left(b_g + F\mu^2\right)^2 P_{\mathrm{lin}},
\tag{A.26}
$$

where the bias is

$$
b_g(k,\mu,z) \equiv \int \frac{1}{\bar{n}_g}dm\,n(m)\,b(m)\left[N_{\mathrm{cen}} + N_{\mathrm{sat}}u(k,m,z)e^{-k^2\mu^2\sigma_d^2(m)/2}\right]
\tag{A.27}
$$

and the effective growth rate of structure is

$$F(k, \mu, z) \equiv f \int dm\, n(m) \left(\frac{m}{\bar{\rho}}\right) u(k, m) e^{-k^2 \mu^2 \sigma_d^2(m)/2}. \tag{A.28}$$

The 1-halo term becomes:

$$P_g^{\text{1-halo}}(k, \mu, z) = \frac{1}{\bar{n}_g^2} \int dm\, n(m) \left[N_{\text{sat}}^2 |u(k, m, z)|^2 e^{-k^2 \mu^2 \sigma_d^2(m)} \right. \\ \left. + 2 N_{\text{sat}} N_{\text{cen}} e^{-k^2 \mu^2 \sigma_d^2(m)/2} \right]. \tag{A.29}$$

Finally, the shot noise power spectrum is simply $P_g^{\text{shot}} = \bar{n}_g^{-1}$.

A.5 Multi-line CLF: cross-correlation with a tracer

One important observable in LIM surveys is the cross-correlation with a discrete tracer of the mass, e.g., individually detected quasars (QSO). Previous papers model these cross-correlations (e.g., [106]). Here, we describe them in terms of the CLF. To describe such cross-correlation, we restrict the multi-line CLF to count only sources of a given type. In what follows, we focus on QSO for concreteness, but the formalism applies identically to any tracer of the matter density. We thus introduce the QSO multi-line CLF $\phi(L_1, ..., L_n|m, \text{QSO})$ such that $\phi(L_1, ..., L_n|m, \text{QSO})dL_1...dL_n$ is the mean number of QSOs in a halo of mass m, with luminosity L_i in each line i. In particular, the QSO CLF describes the full complexity of the QSO selection function. If the QSO selection function is only based on fluxes in several bands, the QSO CLF can be simply obtained by adding these bands to the standard CLF, and multiplying it by the selection function. As above, we can then define the mean number density of QSOs per halo or per unit comoving volume:

$$\begin{cases} N_{\text{QSO}}(m) = \int dL_1\, \phi(L_1|m, \text{QSO}) \\ \bar{n}^{\text{QSO}} = \int dm\, n(m) N_{\text{QSO}}(m) \end{cases}, \tag{A.30}$$

and the mean QSO luminosity in line j:

$$\begin{cases} L_j^{\text{QSO}}(m) = \int dL_j\, \phi(L_j|m, \text{QSO}) L_j \\ \mathcal{L}_j^{\text{QSO}} \equiv \int dm\, n(m) L_j^{\text{QSO}}(m) \end{cases}. \tag{A.31}$$

The QSO selection function, described by the QSO CLF, typically includes a minimum luminosity in some bands, corresponding to the detection threshold of the survey considered. This provides an effective lower luminosity (and halo mass) cutoff in the integrals above.

The cross-power spectrum between the intensity in line 1 I_1 and the number density of QSO n_{QSO} is again the sum of three contributions:

$$P_{1,\text{QSO}}(k, \mu, z) = P_{1,\text{QSO}}^{\text{2-halo}}(k, \mu, z) + P_{1,\text{QSO}}^{\text{1-halo}}(k, \mu, z) + P_{1,\text{QSO}}^{\text{shot}}(z), \tag{A.32}$$

The 2-halo term is given by

$$\boxed{P_{1,\text{QSO}}^{\text{2-halo}}(k, \mu, z) = \bar{I}_1 \bar{n}^{\text{QSO}} \left(b_1 + F\mu^2\right) \left(b_{\text{QSO}} + F\mu^2\right) P_{\text{lin}},} \tag{A.33}$$

where the QSO bias term is

$$b_j(k, \mu, z) \equiv \frac{1}{\bar{n}^{\text{QSO}}} \int dm \; n(m) \; N_{\text{QSO}}(m) \; b(m) \; u(k,m) e^{-k^2 \mu^2 \sigma^2/2}. \tag{A.34}$$

The 1-halo term becomes

$$\boxed{P_{i,\text{QSO}}^{\text{1-halo}}(k, \mu, z) = \bar{I}_i \bar{n}^{\text{QSO}} \frac{U_{i,\text{QSO}}^2(k, \mu, z)}{\bar{n}_{i,\text{QSO}}^{\text{h eff}}},} \tag{A.35}$$

where the effective halo profile and number densities are given by:

$$\begin{cases} U_{i,\text{QSO}}^2(k, \mu, z) \equiv \dfrac{\int dm \; n(m) \, |u(k,m)|^2 \, e^{-k^2 \mu^2 \sigma^2} L_i(m) N_{\text{QSO}}(m)}{\int dm \; n(m) L_i(m) N_{\text{QSO}}(m)} \\[4ex] \bar{n}_{i,\text{QSO}}^{\text{h eff}} \equiv \dfrac{\left(\int dm \; n(m) L_i(m) \right) \left(\int dm \; n(m) N_{\text{QSO}}(m) \right)}{\int dm \; n(m) L_i(m) N_{\text{QSO}}(m)} \end{cases} \tag{A.36}$$

Finally, the shot noise cross-power spectrum is given by

$$\boxed{P_{i,\text{QSO}}^{\text{shot}}(z) = \frac{\bar{I}_i \bar{n}^{\text{QSO}}}{\bar{n}_{i,\text{QSO}}^{\text{gal eff}}},} \tag{A.37}$$

with the effective galaxy number density

$$\bar{n}_{i,j}^{\text{gal eff}} = \frac{\left(\int dm \; n(m) L_i(m) \right) \left(\int dm \; n(m) N_{\text{QSO}}(m) \right)}{\int dm \; n(m) \int dL_i \; \phi(L_i|m, \text{QSO}) L_i}. \tag{A.38}$$

This formalism thus connects the observable cross-correlations between a tracer and intensity maps to the underlying CLF of the tracer. By doing so, it provides astrophysical insights into the abundance of the prevalence of the interstellar medium regions responsible for each line inside the tracers. Refs. [107, 108] indeed explore these cross-correlations, in order to learn about the gas abundances in galaxies and the properties of feedback in star formation at high redshift.

B Projected angular power spectrum

One of the primary advantages of LIM over purely photometric surveys is enhanced redshift precision. However there is also an intermediate regime, where the redshifts are not highly precise and redshift errors may have an unknown distribution, for which an analysis of angular clustering in multiple redshift bins is more appropriate. Our formalism generalizes easily to this situation, and we include the relevant expressions here for completeness.

$$I_X(\hat{n}) = \int d\chi \; W_X(\chi) I(\chi \hat{n}) \tag{B.1}$$

with χ the comoving radial distance and W_X defining the shell. The angular power spectrum of the cross-correlation between two shells is

$$C_\ell^{XY} = \frac{2}{\pi} \int d\chi_1 \, d\chi_2 \; W_X(\chi_1) W_Y(\chi_2) \int k^2 \, dk \; P_{XY}(k; \chi_1, \chi_2) j_\ell(k\chi_1) j_\ell(k\chi_2) \tag{B.2}$$

with j_ℓ the spherical Bessel function of order ℓ. Except on very large scales, where linear theory is appropriate, we can approximate $P_{XY}(k; \chi_1, \chi_2) \simeq P_{XY}(k, \mu, z(\chi_1))$.

If we now decompose our power spectrum at z into

$$P(k, \mu, z) = P^{2-\text{halo}} + P^{1-\text{halo}} + P^{\text{shot}} \tag{B.3}$$

we find significant simplifications. The shot-noise term is k-independent, and so can be pulled out of the k integral. The integral over k then gives a δ-function so that

$$C_\ell^{\text{shot}} = \int \frac{d\chi}{\chi^2} \; W_X(\chi) W_Y(\chi) \frac{\bar{I}_\chi \bar{I}_\chi}{\bar{n}_{\text{gal,eff}}} \tag{B.4}$$

If the bin width is significantly larger than the velocity dispersion of halos dominating the emission we can set $\mu = 0$ in the 1-halo term. Since this term typically contributes mostly on small-scales, we can also use the Limber approximation ([109, 110]; valid for $k\Delta\chi \gg 1$, with $\Delta\chi$ the width of the shell) to write

$$C_\ell^{1-\text{halo}} = \int \frac{d\chi}{\chi^2} \; W_X(\chi) W_Y(\chi) \frac{\bar{I}_\chi \bar{I}_\chi}{\bar{n}_{\text{h,eff}}} U^2 \left(k = \frac{\ell + 1/2}{\chi}, \mu = 0 \right) \tag{B.5}$$

This leaves only the 2-halo term. In the Limber approximation

$$C_\ell^{2-\text{halo}} = \int \frac{d\chi}{\chi^2} \; W_X(\chi) W_Y(\chi) \bar{I}_\chi \bar{I}_\chi b^2 \left(k = \frac{\ell + 1/2}{\chi}, \mu = 0 \right) P_{\text{lin}} \left(k = \frac{\ell + 1/2}{\chi} \right) \tag{B.6}$$

while at large scales one must perform the double integral in Eq. B.2 using Eq. 2.8.

B.1 Thin shell LIM

Finally we comment upon a limit in which the results become particularly simple and intuitive. If we take W to be a top-hat centered at χ_0 and of width $\Delta\chi \ll \chi_0$, normalized to unit integral, and we work in the Limber approximation,

$$C_\ell \approx \mathcal{V}^{-1} \; P \left(k = \frac{\ell + 1/2}{\chi_0}, \mu = 0, z \right) \tag{B.7}$$

with $\mathcal{V} = \chi_0^2 \Delta\chi$ the volume per steradian. In this limit the 2D power spectrum equals the 3D power spectrum for modes of wavelength $\lambda = 2\pi/k$ subtending an angle $\theta = 2\pi/\ell$ at distance χ_0 once accounting for the volume per solid angle.

Similarly, for higher order n-point functions $P^{(n)}$, the Limber approximation

$$P_\ell^{(n)} = \int \frac{d\chi}{\chi^{2(n-1)}} \; W_X(\chi)^n \; P^{(n)}(\boldsymbol{k} = \frac{\ell + 1/2}{\chi}) \tag{B.8}$$

simplifies in the thin shell regime to:

$$P_\ell^{(n)} = \mathcal{V}^{-(n-1)} P^{(n)}(\boldsymbol{k} = \frac{\ell + 1/2}{\chi}). \tag{B.9}$$

In this limit non-overlapping redshift shells are uncorrelated, so all of the information is contained in the angular auto-spectra. This makes it particularly easy to see what information is obtained from what parts of the clustering signal.

B.2 Cross-correlation with CMB & galaxy lensing

To compute the cross-power spectrum of LIM and galaxy or CMB lensing convergence κ, we introduce the lensing kernel, for a source located at comoving radial distance χ_S.

$$W^\kappa(\chi, \chi_S) = \frac{3}{2} \left(\frac{H_0}{c}\right)^2 \Omega_m^0 \frac{\chi}{a(\chi)} \left(1 - \frac{\chi}{\chi_S}\right). \tag{B.10}$$

For CMB lensing, the source distance is the surface of last scatter, at a distance $\chi_{\rm SLS}$, corresponding to $z_{\rm SLS} \sim 1100$:

$$W^{\kappa_{\rm CMB}}(\chi) = W^\kappa(\chi, \chi_{\rm SLS}). \tag{B.11}$$

For galaxy lensing, we average the lensing kernel $W^\kappa(\chi, \chi_S)$ over the redshift distribution $dn_S/d\chi_S$ of the source galaxies:

$$W^{\kappa_{\rm gal}}(\chi) = \frac{1}{n_{\rm S}} \int d\chi_S \frac{dn_{\rm S}}{d\chi_S} W^\kappa(\chi, \chi_S), \quad \text{with } n_S = \int d\chi \frac{dn_S}{d\chi}. \tag{B.12}$$

In the Born and Limber approximations, we can thus predict the cross-spectrum of LIM with the (galaxy or CMB) lensing convergence. The power spectrum is the sum of a 1-halo term

$$C_\ell^{1-\rm halo} = \int \frac{d\chi}{\chi^2} W^{\rm LIM}(\chi) W^\kappa(\chi) \left(\frac{c}{4\pi H(z)}\right) \frac{1}{\nu_{\rm LIM}^0}$$
$$\int dm \, n(m) \left| u(k = \frac{\ell + 1/2}{\chi}, m) \right|^2 e^{-\left(\frac{\ell+1/2}{\chi}\right)^2 \mu^2 \sigma^2} L_i(m) \frac{m}{\bar{\rho}_m} \tag{B.13}$$

and a 2-halo term

$$C_\ell^{2-\rm halo} = \int \frac{d\chi}{\chi^2} W^{\rm LIM}(\chi) W^\kappa(\chi) \bar{I}_{\rm LIM} \left(b_{\rm LIM} + F\mu^2\right) \left(b_m + F\mu^2\right) P_{\rm lin}. \tag{B.14}$$

Here, $I_{\rm LIM}$ and $b_{\rm LIM}$ are the scale and redshift-dependent mean intensity and bias for LIM, used throughout this paper. The effective growth rate of structure is the same as Eq. (2.10). The quantity b_m is the usual matter bias in the halo model:

$$b_m(k, \mu, z) \equiv \int dm \, n(m) \frac{m}{\bar{\rho}_m} b(m) \, u(k, m) e^{-k^2 \mu^2 \sigma_d^2(m)/2}. \tag{B.15}$$

References

[1] E. D. Kovetz, M. P. Viero, A. Lidz, L. Newburgh, M. Rahman, E. Switzer et al., *Line-Intensity Mapping: 2017 Status Report, ArXiv e-prints* (2017) [1709.09066].

[2] E. Kovetz, P. C. Breysse, A. Lidz, J. Bock, C. M. Bradford, T.-C. Chang et al., *Astrophysics and Cosmology with Line-Intensity Mapping*, in BAAS, vol. 51, p. 101, May, 2019, 1903.04496.

[3] Cosmic Visions 21 cm Collaboration, R. Ansari, E. J. Arena, K. Bandura, P. Bull, E. Castorina et al., *Inflation and Early Dark Energy with a Stage II Hydrogen Intensity Mapping Experiment, arXiv e-prints* (2018) arXiv:1810.09572 [1810.09572].

[4] Z. Ahmed, D. Alonso, M. A. Amin, R. Ansari, E. J. Arena, K. Bandura et al., *Research and Development for HI Intensity Mapping, arXiv e-prints* (2019) arXiv:1907.13090 [1907.13090].

[5] A. R. Pullen, O. Doré and J. Bock, *Intensity Mapping across Cosmic Times with the Lyα Line*, ApJ **786** (2014) 111 [1309.2295].

[6] A. Lidz and J. Taylor, *On Removing Interloper Contamination from Intensity Mapping Power Spectrum Measurements*, ApJ **825** (2016) 143 [1604.05737].

[7] Y. Gong, X. Chen and A. Cooray, *Cosmological constraints from line intensity mapping with interlopers*, arXiv e-prints (2020) arXiv:2001.10792 [2001.10792].

[8] S. R. Furlanetto and A. Lidz, *The Cross-Correlation of High-Redshift 21 cm and Galaxy Surveys*, ApJ **660** (2007) 1030 [astro-ph/0611274].

[9] A. Lidz, O. Zahn, S. R. Furlanetto, M. McQuinn, L. Hernquist and M. Zaldarriaga, *Probing Reionization with the 21 cm Galaxy Cross-Power Spectrum*, ApJ **690** (2009) 252 [0806.1055].

[10] E. Visbal and A. Loeb, *Measuring the 3D clustering of undetected galaxies through cross correlation of their cumulative flux fluctuations from multiple spectral lines*, J. Cosmology Astropart. Phys. **2010** (2010) 016 [1008.3178].

[11] E. Visbal, H. Trac and A. Loeb, *Demonstrating the feasibility of line intensity mapping using mock data of galaxy clustering from simulations*, J. Cosmology Astropart. Phys. **2011** (2011) 010 [1104.4809].

[12] A. Beane, F. Villaescusa-Navarro and A. Lidz, *Measuring the EoR Power Spectrum without Measuring the EoR Power Spectrum*, ApJ **874** (2019) 133 [1811.10609].

[13] D. T. Chung, M. P. Viero, S. E. Church and R. H. Wechsler, *Forecasting [C II] line-intensity mapping measurements between the end of reionization and the epoch of galaxy assembly*, arXiv e-prints (2018) arXiv:1812.08135 [1812.08135].

[14] T. Y. Li, R. H. Wechsler, K. Devaraj and S. E. Church, *Connecting CO Intensity Mapping to Molecular Gas and Star Formation in the Epoch of Galaxy Assembly*, ApJ **817** (2016) 169 [1503.08833].

[15] G. Popping, D. Narayanan, R. S. Somerville, A. L. Faisst and M. R. Krumholz, *The art of modelling CO, [C I], and [C II] in cosmological galaxy formation models*, MNRAS **482** (2019) 4906 [1805.11093].

[16] S. Saito, S. de la Torre, O. Ilbert, C. Dubois, K. Yabe and J. Coupon, *The synthetic Emission Line COSMOS catalogue: Hα and [O II] galaxy luminosity functions and counts at 0.3 < z < 2.5*, MNRAS **494** (2020) 199 [2003.06394].

[17] K. N. Hainline, R. E. Hviding, M. Rieke, I. Shivaei, R. Endsley, E. Curtis-Lake et al., *Simulating JWST/NIRCam Color Selection of High-redshift Galaxies*, ApJ **892** (2020) 125 [2003.04326].

[18] J. Chevallard and S. Charlot, *Modelling and interpreting spectral energy distributions of galaxies with BEAGLE*, MNRAS **462** (2016) 1415 [1603.03037].

[19] J. Gutkin, S. Charlot and G. Bruzual, *Modelling the nebular emission from primeval to present-day star-forming galaxies*, MNRAS **462** (2016) 1757 [1607.06086].

[20] M. Hirschmann, S. Charlot, A. Feltre, T. Naab, E. Choi, J. P. Ostriker et al., *Synthetic nebular emission from massive galaxies - I: origin of the cosmic evolution of optical emission-line ratios*, MNRAS **472** (2017) 2468 [1706.00010].

[21] K. Moriwaki, N. Yoshida, I. Shimizu, Y. Harikane, Y. Matsuda, H. Matsuo et al., *The distribution and physical properties of high-redshift [O III] emitters in a cosmological hydrodynamics simulation*, MNRAS **481** (2018) L84 [1805.07062].

[22] A. Pallottini, A. Ferrara, D. Decataldo, S. Gallerani, L. Vallini, S. Carniani et al., *Deep into the structure of the first galaxies: SERRA views*, MNRAS **487** (2019) 1689 [1905.08254].

[23] H. Katz, T. P. Galligan, T. Kimm, J. Rosdahl, M. G. Haehnelt, J. Blaizot et al., *Probing cosmic dawn with emission lines: predicting infrared and nebular line emission for ALMA and JWST*, MNRAS **487** (2019) 5902 [1901.01272].

[24] I. De Looze, M. Baes, G. J. Bendo, L. Cortese and J. Fritz, *The reliability of [C II] as an indicator of the star formation rate*, MNRAS **416** (2011) 2712 [1106.1643].

[25] L. Spinoglio, K. M. Dasyra, A. Franceschini, C. Gruppioni, E. Valiante and K. Isaak, *Far-IR/Submillimeter Spectroscopic Cosmological Surveys: Predictions of Infrared Line Luminosity Functions for z < 4 Galaxies*, ApJ **745** (2012) 171 [1110.4837].

[26] I. De Looze, D. Cormier, V. Lebouteiller, S. Madden, M. Baes, G. J. Bendo et al., *The applicability of far-infrared fine-structure lines as star formation rate tracers over wide ranges of metallicities and galaxy types*, A&A **568** (2014) A62 [1402.4075].

[27] T. Díaz-Santos, L. Armus, V. Charmandaris, N. Lu, S. Stierwalt, G. Stacey et al., *A Herschel/PACS Far-infrared Line Emission Survey of Local Luminous Infrared Galaxies*, ApJ **846** (2017) 32 [1705.04326].

[28] C. Schreiber, D. Elbaz, M. Pannella, E. Merlin, M. Castellano, A. Fontana et al., *EGG: hatching a mock Universe from empirical prescriptions⋆*, A&A **602** (2017) A96 [1606.05354].

[29] X. Yang, H. J. Mo and F. C. van den Bosch, *Constraining galaxy formation and cosmology with the conditional luminosity function of galaxies*, MNRAS **339** (2003) 1057 [astro-ph/0207019].

[30] U. Seljak, *Analytic model for galaxy and dark matter clustering*, MNRAS **318** (2000) 203 [astro-ph/0001493].

[31] J. A. Peacock and R. E. Smith, *Halo occupation numbers and galaxy bias*, MNRAS **318** (2000) 1144 [astro-ph/0005010].

[32] E. Schaan and M. White, *Astrophysics & Cosmology from Line Intensity Mapping vs Galaxy Surveys*, arXiv e-prints (2021) arXiv:2103.01971 [2103.01971].

[33] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont et al., *Planck 2015 results. XIII. Cosmological parameters*, A&A **594** (2016) A13 [1502.01589].

[34] G. L. Bryan and M. L. Norman, *Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons*, ApJ **495** (1998) 80 [astro-ph/9710107].

[35] R. K. Sheth and G. Tormen, *Large-scale bias and the peak background split*, MNRAS **308** (1999) 119 [astro-ph/9901122].

[36] D. E. Osterbrock and G. J. Ferland, *Astrophysics of gaseous nebulae and active galactic nuclei*. 2006.

[37] E. M. Levesque, L. J. Kewley and K. L. Larson, *Theoretical Modeling of Star-Forming Galaxies. I. Emission-Line Diagnostic Grids for Local and Low-Metallicity Galaxies*, AJ **139** (2010) 712 [0908.0460].

[38] N. Byler, J. J. Dalcanton, C. Conroy and B. D. Johnson, *Nebular Continuum and Line Emission in Stellar Population Synthesis Models*, ApJ **840** (2017) 44 [1611.08305].

[39] G. J. Ferland, *Quantitative Spectroscopy of Photoionized Clouds*, ARA&A **41** (2003) 517.

[40] M. Peimbert, A. Peimbert and G. Delgado-Inglada, *Nebular Spectroscopy: A Guide on Hii Regions and Planetary Nebulae*, PASP **129** (2017) 082001 [1705.06323].

[41] B. T. Draine, *Physics of the Interstellar and Intergalactic Medium*. 2011.

[42] J. Moustakas, J. Kennicutt, Robert C. and C. A. Tremonti, *Optical Star Formation Rate Indicators*, ApJ **642** (2006) 775 [astro-ph/0511730].

[43] I. Shivaei, M. Kriek, N. A. Reddy, A. E. Shapley, G. Barro, C. Conroy et al., *The MOSDEF Survey: The Strong Agreement between Hα and UV-to-FIR Star Formation Rates for z ~2 Star-forming Galaxies*, ApJ **820** (2016) L23 [1603.02284].

[44] E. Daddi, H. Dannerbauer, D. Liu, M. Aravena, F. Bournaud, F. Walter et al., *CO excitation of normal star-forming galaxies out to z = 1.5 as regulated by the properties of their interstellar medium*, A&A **577** (2015) A46 [1409.8158].

[45] A. Beane and A. Lidz, *Extracting Bias Using the Cross-bispectrum: An EoR and 21 cm-[C II]-[C II] Case Study*, ApJ **867** (2018) 26 [1806.02796].

[46] P. C. Breysse, E. D. Kovetz, P. S. Behroozi, L. Dai and M. Kamionkowski, *Insights from probability distribution functions of intensity maps*, MNRAS **467** (2017) 2996 [1609.01728].

[47] C. Behrens, C. Byrohl, S. Saito and J. C. Niemeyer, *The impact of Lyman-α radiative transfer on large-scale clustering in the Illustris simulation*, A&A **614** (2018) A31 [1710.06171].

[48] C. Byrohl, S. Saito and C. Behrens, *Radiative transfer distortions of Lyman α emitters: a new Fingers-of-God damping in the clustering in redshift space*, MNRAS **489** (2019) 3472 [1906.02173].

[49] S. Gurung-López, S. Saito, C. M. Baugh, S. Bonoli, C. G. Lacey and Á. A. Orsi, *Determining the systemic redshift of Lyman α emitters with neural networks and improving the measured large-scale clustering*, MNRAS **500** (2021) 603 [2005.12931].

[50] M. White, *The redshift-space power spectrum in the halo model*, MNRAS **321** (2001) 1 [astro-ph/0005085].

[51] A. E. Evrard, J. Bialek, M. Busha, M. White, S. Habib, K. Heitmann et al., *Virial Scaling of Massive Dark Matter Halos: Why Clusters Prefer a High Normalization Cosmology*, ApJ **672** (2008) 122 [astro-ph/0702241].

[52] Y.-T. Cheng, T.-C. Chang, J. Bock, C. M. Bradford and A. Cooray, *Spectral Line De-confusion in an Intensity Mapping Survey*, ApJ **832** (2016) 165 [1604.07833].

[53] J. Fonseca, M. B. Silva, M. G. Santos and A. Cooray, *Cosmology with intensity mapping techniques using atomic and molecular lines*, MNRAS **464** (2017) 1948 [1607.05288].

[54] A. Cooray and R. Sheth, *Halo models of large scale structure*, Phys. Rep. **372** (2002) 1 [astro-ph/0206508].

[55] Y. B. Zeldovich, *Survey of Modern Cosmology*, *Advances in Astronomy and Astrophysics* **3** (1965) 241.

[56] P. J. E. Peebles, *The large-scale structure of the universe*. 1980.

[57] L. Mercolli and E. Pajer, *On the velocity in the Effective Field Theory of Large Scale Structures*, J. Cosmology Astropart. Phys. **2014** (2014) 006 [1307.3220].

[58] T. Baldauf, U. Seljak, R. E. Smith, N. Hamaus and V. Desjacques, *Halo stochasticity from exclusion and nonlinear clustering*, Phys. Rev. D **88** (2013) 083507 [1305.2917].

[59] D. Sobral, I. Smail, P. N. Best, J. E. Geach, Y. Matsuda, J. P. Stott et al., *A large Hα survey at z = 2.23, 1.47, 0.84 and 0.40: the 11 Gyr evolution of star-forming galaxies from HiZELS*, MNRAS **428** (2013) 1128 [1202.3436].

[60] J. W. Colbert, H. Teplitz, H. Atek, A. Bunker, M. Rafelski, N. Ross et al., *Predicting Future Space Near-IR Grism Surveys Using the WFC3 Infrared Spectroscopic Parallels Survey*, ApJ **779** (2013) 34 [1305.1399].

[61] V. Mehta, C. Scarlata, J. W. Colbert, Y. S. Dai, A. Dressler, A. Henry et al., *Predicting the Redshift 2 Hα Luminosity Function Using [OIII] Emission Line Galaxies*, ApJ **811** (2015) 141 [1505.07843].

[62] R. K. Cochrane, P. N. Best, D. Sobral, I. Smail, D. A. Wake, J. P. Stott et al., *The H α luminosity-dependent clustering of star-forming galaxies from z ∼ 0.8 to ∼2.2 with HiZELS*, MNRAS **469** (2017) 2913 [1704.05472].

[63] Z. Chen, L. Wolz, M. Spinelli and S. G. Murray, *Extracting HI Astrophysics from Interferometric Intensity Mapping*, arXiv e-prints (2020) arXiv:2010.07985 [2010.07985].

[64] Y. Gong, A. Cooray, M. B. Silva, M. Zemcov, C. Feng, M. G. Santos et al., *Intensity Mapping of Hα, Hβ, [OII], and [OIII] Lines at z < 5*, ApJ **835** (2017) 273 [1610.09060].

[65] A. Lupi, A. Pallottini, A. Ferrara, S. Bovino, S. Carniani and L. Vallini, *Predicting FIR lines from simulated galaxies*, MNRAS **496** (2020) 5160 [2004.06118].

[66] S. Yang, R. S. Somerville, A. R. Pullen, G. Popping, P. C. Breysse and A. S. Maniyar, *Multitracer cosmological line intensity mapping mock lightcone simulation*, arXiv e-prints (2020) arXiv:2009.11933 [2009.11933].

[67] B. P. Moster, T. Naab and S. D. M. White, *Galactic star formation and accretion histories from matching galaxies to dark matter haloes*, MNRAS **428** (2013) 3121 [1205.5807].

[68] J. S. Speagle, C. L. Steinhardt, P. L. Capak and J. D. Silverman, *A Highly Consistent Framework for the Evolution of the Star-Forming "Main Sequence" from z ~0-6*, ApJS **214** (2014) 15 [1405.2041].

[69] E. Jimenez, N. Padilla, S. Contreras, I. Zehavi, C. Baugh and A. Orsi, *The assembly bias of emission line galaxies*, arXiv e-prints (2020) arXiv:2010.08500 [2010.08500].

[70] G. Favole, J. Comparat, F. Prada, G. Yepes, E. Jullo, A. Niemiec et al., *Clustering properties of g-selected galaxies at z ∼ 0.8*, MNRAS **461** (2016) 3421 [1507.04356].

[71] V. Gonzalez-Perez, J. Comparat, P. Norberg, C. M. Baugh, S. Contreras, C. Lacey et al., *The host dark matter haloes of [O II] emitters at 0.5 ¡ z ¡ 1.5*, MNRAS **474** (2018) 4024 [1708.07628].

[72] V. Gonzalez-Perez, W. Cui, S. Contreras, C. M. Baugh, J. Comparat, A. J. Griffin et al., *Do model emission line galaxies live in filaments at z ∼ 1?*, MNRAS **498** (2020) 1852 [2001.06560].

[73] D. Sobral, J. Matthee, P. N. Best, I. Smail, A. A. Khostovan, B. Milvang-Jensen et al., *CF-HiZELS, an ∼10 deg² emission-line survey with spectroscopic follow-up: Hα, [O III] + Hβ and [O II] luminosity functions at z = 0.8, 1.4 and 2.2*, MNRAS **451** (2015) 2303 [1502.06602].

[74] A. Merson, Y. Wang, A. Benson, A. Faisst, D. Masters, A. Kiessling et al., *Predicting Hα emission-line galaxy counts for future galaxy redshift surveys*, MNRAS **474** (2018) 177 [1710.00833].

[75] A. Merson, A. Smith, A. Benson, Y. Wang and C. Baugh, *Linear bias forecasts for emission line cosmological surveys*, MNRAS **486** (2019) 5737 [1903.02030].

[76] Z. Zhai, A. Benson, Y. Wang, G. Yepes and C.-H. Chuang, *Prediction of H α and [O III] emission line galaxy number counts for future galaxy redshift surveys*, MNRAS **490** (2019) 3667 [1907.09680].

[77] Z. Zhai, C.-H. Chuang, Y. Wang, A. Benson and G. Yepes, *Clustering in the Simulated Hα Galaxy Redshift Survey from Nancy Grace Roman Space Telescope*, arXiv e-prints (2020) arXiv:2008.09746 [2008.09746].

[78] A. A. Khostovan, D. Sobral, B. Mobasher, P. N. Best, I. Smail, J. P. Stott et al., *Evolution of the H β + [O III] and [O II] luminosity functions and the [O II] star formation history of the Universe up to z ∼ 5 from HiZELS*, MNRAS **452** (2015) 3948 [1503.00004].

[79] G. Popping, E. van Kampen, R. Decarli, M. Spaans, R. S. Somerville and S. C. Trager, *Sub-mm emission line deep fields: CO and [C II] luminosity functions out to z = 6*, MNRAS **461** (2016) 93 [1602.02761].

[80] S. Hemmati, L. Yan, T. Diaz-Santos, L. Armus, P. Capak, A. Faisst et al., *The Local [C II] 158 µm Emission Line Luminosity Function*, ApJ **834** (2017) 36 [1611.03092].

[81] P. Cassata, O. Le Fèvre, B. Garilli, D. Maccagni, V. Le Brun, M. Scodeggio et al., *The VIMOS VLT Deep Survey: star formation rate density of Lyα emitters from a sample of 217 galaxies with spectroscopic redshifts 2 ≤ z ≤ 6.6*, A&A **525** (2011) A143 [1003.3480].

[82] G. Sun, B. S. Hensley, T.-C. Chang, O. Doré and P. Serra, *A Self-consistent Framework for Multiline Modeling in Line Intensity Mapping Experiments*, ApJ **887** (2019) 142 [1907.02999].

[83] P. Serra, O. Doré and G. Lagache, *Dissecting the High-z Interstellar Medium through Intensity Mapping Cross-correlations*, ApJ **833** (2016) 153 [1608.00585].

[84] B. Yue and A. Ferrara, *Studying high-z galaxies with [C II] intensity mapping*, MNRAS **490** (2019) 1928 [1909.11656].

[85] D. T. Chung, M. P. Viero, S. E. Church and R. H. Wechsler, *Forecasting [C II] Line-intensity Mapping Measurements between the End of Reionization and the Epoch of Galaxy Assembly*, ApJ **892** (2020) 51 [1812.08135].

[86] The CONCERTO collaboration, P. Ade, M. Aravena, E. Barria, A. Beelen, A. Benoit et al., *A wide field-of-view low-resolution spectrometer at APEX: instrument design and science forecast*, arXiv e-prints (2020) arXiv:2007.14246 [2007.14246].

[87] H. Padmanabhan, *Constraining the CO intensity mapping power spectrum at intermediate redshifts*, MNRAS **475** (2018) 1477 [1706.01471].

[88] H. Padmanabhan, *Constraining the evolution of [C II] intensity through the end stages of reionization*, MNRAS **488** (2019) 3014 [1811.01968].

[89] O. Doré, J. Bock, M. Ashby, P. Capak, A. Cooray, R. de Putter et al., *Cosmology with the SPHEREX All-Sky Spectral Survey*, arXiv e-prints (2014) arXiv:1412.4872 [1412.4872].

[90] M. B. Silva, M. G. Santos, Y. Gong, A. Cooray and J. Bock, *Intensity Mapping of Lyα Emission during the Epoch of Reionization*, ApJ **763** (2013) 132 [1205.1493].

[91] D. B. Sanders, J. M. Mazzarella, D. C. Kim, J. A. Surace and B. T. Soifer, *The IRAS Revised Bright Galaxy Sample*, AJ **126** (2003) 1607 [astro-ph/0306263].

[92] C. Schreiber, L. Miller, C. Duncan and collaborators, *The biasing effect of photometric noise on estimation of the point spread function for the Euclid weak lensing survey*, in prep (2020) .

[93] J. Kennicutt, Robert C., *The Global Schmidt Law in Star-forming Galaxies*, ApJ **498** (1998) 541 [astro-ph/9712213].

[94] J. A. Baldwin, M. M. Phillips and R. Terlevich, *Classification parameters for the emission-line spectra of extragalactic objects.*, PASP **93** (1981) 5.

[95] P. C. Breysse and M. Rahman, *Feeding cosmic star formation: exploring high-redshift molecular gas with CO intensity mapping*, MNRAS **468** (2017) 741 [1606.07820].

[96] C. Heneka, A. Cooray and C. Feng, *Probing the Intergalactic Medium with Lyα and 21 cm Fluctuations*, ApJ **848** (2017) 52 [1611.09682].

[97] Y.-T. Cheng, T.-C. Chang and J. J. Bock, *Phase-space Spectral Line Deconfusion in Intensity Mapping*, ApJ **901** (2020) 142 [2005.05341].

[98] U. Seljak, *Redshift-space bias and β from the halo model*, MNRAS **325** (2001) 1359 [astro-ph/0009016].

[99] X. Kang, Y. P. Jing, H. J. Mo and G. Börner, *An analytical model for the non-linear redshift-space power spectrum*, MNRAS **336** (2002) 892 [`astro-ph/0201124`].

[100] J. L. Tinker, *Redshift-space distortions with the halo occupation distribution - II. Analytic model*, MNRAS **374** (2007) 477 [`astro-ph/0604217`].

[101] T. Kanemaru, C. Hikage, G. Hütsi, A. Terukina and K. Yamamoto, *What can we learn from higher multipole power spectra of galaxy distribution in redshift space?*, Phys. Rev. D **92** (2015) 023523 [`1501.04372`].

[102] D. Sarkar and S. Bharadwaj, *Modelling redshift space distortion in the post-reionization H I 21-cm power spectrum*, MNRAS **476** (2018) 96 [`1801.07868`].

[103] D. Sarkar and S. Bharadwaj, *Redshift-space distortions of the H I 21-cm intensity mapping signal due to the internal motions within galaxies*, MNRAS **487** (2019) 5666 [`1906.07032`].

[104] J. F. Navarro, C. S. Frenk and S. D. M. White, *A Universal Density Profile from Hierarchical Clustering*, ApJ **490** (1997) 493 [`astro-ph/9611107`].

[105] L. Wolz, S. G. Murray, C. Blake and J. S. Wyithe, *Intensity mapping cross-correlations II: HI halo models including shot noise*, MNRAS **484** (2019) 1007 [`1803.02477`].

[106] G. Sun, T.-C. Chang, B. D. Uzgil, J. Bock, C. M. Bradford, V. Butler et al., *Probing Cosmic Reionization and Molecular Gas Growth with TIME*, arXiv e-prints (2020) arXiv:2012.09160 [`2012.09160`].

[107] L. Wolz, C. Blake and J. S. B. Wyithe, *Determining the H I content of galaxies via intensity mapping cross-correlations*, MNRAS **470** (2017) 3220 [`1703.08268`].

[108] P. C. Breysse and R. M. Alexandroff, *Observing AGN feedback with CO intensity mapping*, MNRAS **490** (2019) 260 [`1904.03197`].

[109] D. N. Limber, *The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field.*, ApJ **117** (1953) 134.

[110] M. Loverde and N. Afshordi, *Extended Limber approximation*, Phys. Rev. D **78** (2008) 123506 [`0809.5112`].