Leveraging cross-correlations and linear covariance-based filtering for line-intensity map reconstructions at linear scales

Dongwoo T. Chung[*]

Canadian Institute for Theoretical Astrophysics, University of Toronto,
60 St. George Street, Toronto, Ontario M5S 3H8, Canada
and Dunlap Institute for Astronomy and Astrophysics, University of Toronto,
50 St. George Street, Toronto, Ontario M5S 3H4, Canada

We explore the possible application of linear covariance-based (LCB) filtering to line-intensity mapping (LIM) signal reconstructions. Originally introduced for reconstruction of the integrated Sachs-Wolfe effect in the cosmic microwave background, the LCB filter is an optimal map estimator that extends the simple Wiener filter by leveraging external correlated data. Given a detectable strong LIM-galaxy or LIM-LIM cross power spectrum, we show recovery of high-redshift, large-scale line-intensity fluctuations—even in the presence of bright interloper emission—in simulations of a futuristic [C II] LIM survey as well as simulated future iterations of the CO Mapping Array Project. With sufficient galaxy abundances or low LIM survey noise, normalized cross-correlation between the LCB reconstruction and the true signal reaches 70%–90% on large, linear comoving scales corresponding to $k \sim 0.1$ Mpc^{-1}. This suggests the possible use of such signal reconstructions in astrophysical or cosmological contexts that require identifying the locations of line emissivity peaks and voids, although clear shortcomings exist on smaller scales. The successful application of the LCB filter in simulated LIM contexts highlights the importance of cross-correlations to studies of the reionizing and reionized high-redshift universe with LIM and other large-scale structure surveys.

I. INTRODUCTION

Line-intensity mapping (LIM; for recent reviews see Refs. [1–3]) will provide a view of the cosmic web not through resolving populations of galaxies or other discrete sources, but by tracing the spatial-spectral fluctuations of unresolved, integrated emission in a given atomic or molecular line species. The relevant ideas were first developed for tomography of the high-redshift universe with the 21 cm neutral hydrogen line [4,5], but recent activity has spread to theoretical explorations and construction of experiments targeting emission in carbon monoxide (CO) and singly ionized carbon ([C II]) lines [6–24]. Observations in those lines could probe the cosmic history of star formation and molecular gas and strongly complement other surveys of large-scale structure (LSS), including 21 cm tomography.

Despite the name, line-intensity mapping is not necessarily about high-fidelity mapping of the redshift-space fluctuations in line intensity, but about measuring the statistics of these fluctuations. Most literature has focussed on recovery of the power spectrum. Some work also explores one-point statistics [25–27] and higher-order

statistics, although the latter more so in the context of 21 cm intensity mapping [28–30]. Reconstructing the large-scale line-intensity fluctuations themselves has not been as much of a focus of research.

Yet future LIM experiments will reach sensitivities that make such analyses feasible. Such analyses could potentially also be desirable in many contexts, in terms of being able to designate overdensities or peaks for follow-up, probe underdensities or voids, diagnose the environment of the interstellar medium in a global sense, and so on.

In principle, the Wiener filter (abbreviated at points in this work as the WF) provides a simple, minimum-variance linear estimator of the original signal, but only provided that there is sufficient prior knowledge of the signal power spectrum and provided that the signal fluctuations dominate over noise at relevant scales. This will often not be the case even after removal of continuum foregrounds. For instance, [C II] surveys will see prominent interloper line emission from lower redshifts (which would be considered part of the noise in this picture). This emission is expected to be so bright that discriminating it from the [C II] signal is a significant portion of the [C II] LIM theoretical literature [31–35].

Previous literature has also contemplated image processing techniques that use convolutional neural

[*]dongwooc@cita.utoronto.ca

networks in order to resolve line confusion and reconstruction (see, e.g., Refs. [36,37]). But the expected signal is not well understood (which to be fair is what motivates these surveys in the first place). When experiments forecast the detectability of CO and [C II] LIM signal from the epoch of reionization (e.g., Refs. [17,38]), they show models for the power spectrum that span at least one order of magnitude. Inadequate anticipation of signal and noise easily confounds neural networks without appropriate training (see, e.g., Ref. [39]). Possibly such uncertainty around the signal has also more generally discouraged interest in exploring LIM signal reconstruction.

One area of significant interest for LIM research, on the other hand, has been the use of cross-correlations involving LIM (or other datasets containing unresolved emission) to improve high-redshift astrophysical and cosmological inferences [40–48], a line of work that traces all the way back in real-world application to the first detections of 21 cm intensity fluctuations in cross-correlation with spectroscopic galaxy surveys [49–51]. This interest is a natural inclination. LIM surveys and other LSS surveys all trace the same cosmic web while their systematics and noise will often be disjoint, and cross-correlations will yield scientific output not possible with the autocorrelations in isolation. LIM analyses may also exploit cross power spectra between different parts of its dataset to improve estimation of what is actually an autopower spectrum (see, e.g., Ref. [27]). So, in general, cross-correlations are an invaluable asset for LIM surveys.

One may naturally ask after a reconstruction technique that behaves as linearly and understandably as the Wiener filter but also leverages the power of cross-correlations, and this already exists in the form of the linear covariance-based (LCB) filter. First introduced by Barreiro et al. [52], the LCB filter has mostly been used for reconstruction of maps of the integrated Sachs-Wolfe (ISW) effect, distinguished from the primordial cosmic microwave background by cross-correlation with large-scale structure [52–56]. This work applies the same formalism to mock LIM data and examines resulting reconstructions in simulated CO and [C II] LIM surveys cross-correlated with galaxy surveys or other LIM data.

To be clear, the central contention of this paper is *not* that the LCB filter is necessarily the best estimator to use. Other reconstruction techniques relying on forward modeling, physical priors (e.g., cosmic tidal shear), Bayesian statistics, information theory, or some combination of these are available [57–61]. But for the purposes of this work, we want to demonstrate the extent to which cross-correlations can indeed improve reconstruction of LIM signals relative to relying on autocorrelation alone, with the LCB filter and Wiener filter representing the optimal solutions in the linear regime in each case.

More specifically, we ask: Can the LCB filter successfully reconstruct the structure of line-intensity fluctuations on large linear scales by exploiting
 (i) cross-correlations with surveys of discrete sources and/or
 (ii) cross-correlation with a measurement of a different, correlated line-intensity signal?

The paper is structured as follows. We provide an overview of the LCB filter itself in Sec. II, reviewing its derivation and considering cases relevant to LIM reconstruction. After a brief interlude in Sec. III about metrics for "successful reconstruction," we then consider a proof of concept for applying the LCB filter to LIM with simulations of [C II] signal reconstruction in Sec. IV and undertake additional case studies in the context of the CO Mapping Array Project (COMAP) in Sec. V. We end by drawing conclusions from those results in Sec. VI.

We use base-10 logarithms throughout unless otherwise specified, as well as Λ cold dark matter cosmologies with parameters $\{\Omega_m, \Omega_\Lambda, \Omega_b, h \equiv H_0/(100 \, \mathrm{km \, s^{-1} \, Mpc^{-1}}), \sigma_8, n_s\}$ stated near the start of each section as appropriate. Distances carry an implicit h^{-1} dependence throughout, which also propagates through to masses (all based on virial halo masses $\propto h^{-1}$) and volume densities ($\propto h^3$).

II. THE LCB FILTER

We first recap the derivation of Barreiro et al. [52], but rephrased in terms of the isotropic power spectrum $P(k)$ describing a random Gaussian field, instead of specifically in a spherical multipole expansion. This in and of itself is largely a superficial change as mathematically the two phrasings are equivalent. However, we provide the derivation already generalized to reconstructing the desired signal from cross-correlation with n other tracers.

Note also that Manzotti and Dodelson [53] derive an ISW estimator based on likelihood optimization, but state that it will also reduce to the LCB estimator. We verify this explicitly in Appendix A. Thus, the LCB filter describes an optimal estimator for the signal that combines cross-correlation and autocorrelation measurements across all observables.

Following the general derivation provided in Sec. II A, we consider in Sec. II B the $n = 1$ case originally described by Barreiro et al. [52] and then consider aspects of practical implementation for matter tracers in Sec. II C.

A. General case

We begin by noting that the auto- and cross power spectra by definition describe the covariance of the Fourier modes of the random Gaussian fields,

$$P_{ij}^{(S)}(k) = \langle s_i(k) s_j^*(k) \rangle. \tag{1}$$

Note that we have assumed $\langle s_i(k) \rangle = 0$ for all k and i, implying that any nonzero mean that may have existed has already been subtracted in mapping the field.

Assume mode coupling is negligible, i.e., the covariance matrix of the signal fields is block diagonal with all cross-k elements equal to zero and described fully by $P_{ij}^{(S)}(k)$. The covariance will also be symmetric, as $P_{ij}^{(S)}(k) = P_{ji}^{(S)}(k)$. Assume furthermore that we obtain measurements d_i of each field with uncorrelated additive Gaussian noise,

$$d_i = s_i + n_i. \tag{2}$$

If the power spectrum of each noise component is thus some constant $P_i^{(N)}$, then the total covariance matrix is given by

$$C_{ij}(k) = P_{ij}^{(S)}(k) + \delta_{ij} P_i^{(N)}, \tag{3}$$

with δ_{ij} being the Kronecker delta. Going forward, we write the autopower spectra as $P_i^{(S)}(k) \equiv P_{ii}^{(S)}(k)$ for brevity.

Then given measurements $\{d_1, d_2, \ldots, d_n\}$ of tracers $\{s_1, s_2, \ldots, s_n\}$, we wish to obtain a reconstruction of s_{n+1} from data d_{n+1}. For brevity we notate $p \equiv n+1$, and we will further abbreviate the signal and noise power spectra for this last observable as $P^{(S)} \equiv P_p^{(S)}$ and $P^{(N)} \equiv P_p^{(N)}$.

Consider for a moment the covariance of the vector $\mathbf{f} = (d_1, d_2, \ldots, d_n, s_p)$, i.e., the covariance between the observations of all n tracers that we are not reconstructing and the intrinsic signal of the tracer s_p that we do want to reconstruct. This then is given by

$$D_{ij}(k) = \begin{cases} P_{ij}^{(S)}(k) + \delta_{ij} P_i^{(N)} & \text{if } i \le n \quad \text{and} \quad j \le n, \\ P_{ij}^{(S)}(k) & \text{otherwise.} \end{cases} \tag{4}$$

with δ_{ij} being the Kronecker delta.

We are interested in the Cholesky decomposition of this matrix, i.e., the identification of a lower triangular matrix $L(k)$ such that $D = LL^T$ at each k. This is because once we obtain L, we can claim that some hidden Gaussian random vector $\mathbf{h}$, whose elements are all uncorrelated with diagonal unit covariance, has generated our correlated observations and signals f_i such that

$$\mathbf{f}(k) = L(k)\mathbf{h}(k). \tag{5}$$

What we then want to do is express the signal s_p in terms of the hidden $\mathbf{h}$ and the other data d_i, as well as the lower triangular matrix L. At each k,

$$s_p = \sum_{i=1}^{p} L_{pi} h_i$$

$$= L_{pp} h_p + \sum_{i=1}^{n} L_{pi} h_i$$

$$= L_{pp} h_p + \sum_{i=1}^{n} \left(L_{pi} \sum_{j=1}^{n} (L^{-1})_{ij} d_j \right), \tag{6}$$

noting that we write the row index first when indexing the matrices. For the last step we recall that $L^{-1}\mathbf{f} = \mathbf{h}$, and note that L^{-1} is also a lower triangular matrix so that $(L^{-1})_{ip} = 0$ for all $i \le n$.

Note that no knowledge of the hidden fields is necessary to calculate the second term, which is purely defined in terms of the data-signal covariances and the data for tracers 1 through n. At this point, Barreiro *et al.* [52] propose that, to estimate the scaled hidden field $L_{pp} h_p$, one should Wiener filter the difference between the actual data and the second term of Eq. (6), i.e.,

$$\bar{d} \equiv d_p - \sum_{i=1}^{n} \left(L_{pi} \sum_{j=1}^{n} (L^{-1})_{ij} d_j \right) = L_{pp} h_p + n_p. \tag{7}$$

Then the linear covariance-based estimator for s_p simply substitutes the above (with a Wiener filter factor) into Eq. (6). With a slight rearrangement of the double sum, we can write the estimator as

$$\hat{s}_p = \frac{L_{pp}^2}{L_{pp}^2 + P^{(N)}} \left(d_p - \sum_{j=1}^{n} \sum_{i=1}^{n} L_{pi} (L^{-1})_{ij} d_j \right)$$
$$+ \sum_{j=1}^{n} \sum_{i=1}^{n} L_{pi} (L^{-1})_{ij} d_j. \tag{8}$$

This is analogous to the expression given by Refs. [54,55].

The form of Eq. (8) also allows us to easily see the power spectrum of the reconstructed $\hat{s}_p$. By definition, $\bar{d}$ is uncorrelated with the second term of Eq. (6) and should have a power spectrum given by $L_{pp}^2 + P^{(N)}$. The power spectrum of all of $\hat{s}_p$ should then be

$$P^{(\hat{s})} = \frac{L_{pp}^4}{L_{pp}^2 + P^{(N)}}$$
$$+ \sum_{j=1}^{n} \sum_{j'=1}^{n} \sum_{i=1}^{n} \sum_{i'=1}^{n} L_{pi} L_{pi'} (L^{-1})_{ij} (L^{-1})_{i'j'} D_{jj'}. \tag{9}$$

However, Eq. (8) can also be clearly rewritten as

$$\hat{s}_p = \frac{L_{pp}^2}{L_{pp}^2 + P^{(N)}} d_p$$
$$+ \left(1 - \frac{L_{pp}^2}{L_{pp}^2 + P^{(N)}} \right) \sum_{j=1}^{n} \sum_{i=1}^{n} L_{pi} (L^{-1})_{ij} d_j. \tag{10}$$

This provides an alternate intuition for what is going on: the estimator starts with filtered data for the observable p and then introduces additional fluctuations based on correlations between the observables. Indeed, with $n = 0$ (i.e., without anything to cross-correlate with d_p) we would simply recover the Wiener filter. However, to delve further into this kind of intuition, it is useful to return to the context of the original derivation given by Barreiro *et al.* [52], where the signal to be reconstructed is only correlated with one additional tracer.

B. Special case: $n = 1$

Suppose we only have two observables, d_1 and d_2, and we wish to obtain the estimator $\hat{s}_2$ for the signal underlying the latter observation. In this case, the upper triangular matrix $L(k)$ from the Cholesky decomposition is straightforwardly found at each k,

$$L_{11} = (P_1^{(S)} + P_1^{(N)})^{1/2}, \tag{11}$$

$$L_{21} = P_{12}^{(S)}(P_1^{(S)} + P_1^{(N)})^{-1/2}, \tag{12}$$

$$L_{22} = \left(P^{(S)} - \frac{P_{12}^{(S)2}}{P_1^{(S)} + P_1^{(N)}}\right)^{1/2}, \tag{13}$$

and of course $L_{12} = 0$. General, recursive expressions to evaluate L from D exist for $n > 1$ also and are given in Appendix B.

As far as the inverse of this matrix is concerned, only the element $(L^{-1})_{11}$ will be relevant in evaluating Eq. (10) given that $n = 1$, and it is extremely straightforward to find

$$(L^{-1})_{11} = 1/L_{11} = (P_1^{(S)} + P_1^{(N)})^{-1/2}. \tag{14}$$

Substituting the above into Eq. (10) yields

$$\hat{s}_2 = \frac{L_{22}^2}{L_{22}^2 + P^{(N)}}d_2 + \left(1 - \frac{L_{22}^2}{L_{22}^2 + P^{(N)}}\right)\left(\frac{L_{12}}{L_{11}}d_1\right)$$

$$= \frac{L_{22}^2}{L_{22}^2 + P^{(N)}}d_2$$

$$+ \left(1 - \frac{L_{22}^2}{L_{22}^2 + P^{(N)}}\right)\frac{P_{12}^{(S)}}{P_1^{(S)} + P_1^{(N)}}d_1, \tag{15}$$

where

$$\frac{L_{22}^2}{L_{22}^2 + P^{(N)}} = \frac{(P_1^{(S)} + P_1^{(N)})P^{(S)} - P_{12}^{(S)2}}{(P_1^{(S)} + P_1^{(N)})(P^{(S)} + P^{(N)}) - P_{12}^{(S)2}}. \tag{16}$$

It is interesting to consider some limiting cases of this. If the noise in d_1 is overwhelmingly large, then the weights for d_1 go to zero and the best estimator for s_2 is simply the

Wiener-filtered data d_2. On the other hand, if the noise in d_2 is overwhelmingly large (i.e., $P^{(N)} \gg L_{22}^2$), then the weights for d_2 go to zero instead so that the best estimator for s_2 is actually simply a filtered version of d_1, except the weights have the cross power spectrum $P_{12}^{(S)}$ in the numerator, where the Wiener filter weights would have had $P_1^{(S)}$. Where $\hat{s}_2$ lands between these extremes simply depends on the relative signal and noise power spectrum amplitudes.

The power spectrum of the estimator is given by setting $n = 1$ in Eq. (9),

$$P^{(\hat{s})} = \frac{[(P_1^{(S)} + P_1^{(N)})P^{(S)} - P_{12}^{(S)2}]^2/(P_1^{(S)} + P_1^{(N)})}{(P_1^{(S)} + P_1^{(N)})(P^{(S)} + P^{(N)}) - P_{12}^{(S)2}}$$

$$+ \frac{P_{12}^{(S)2}}{P_1^{(S)} + P_1^{(N)}}. \tag{17}$$

The values of this expression in the limit of large $P_1^{(N)}$ or large $P^{(N)}$ are each consistent with expectations for power spectra of Wiener-filtered data.

As with the Wiener filter, the LCB filter suppresses fluctuations relative to the original signal,

$$P^{(S)} - P^{(\hat{s})} = \frac{[(P_1^{(S)} + P_1^{(N)})P^{(S)} - P_{12}^{(S)2}]P^{(N)}}{(P_1^{(S)} + P_1^{(N)})(P^{(S)} + P^{(N)}) - P_{12}^{(S)2}}. \tag{18}$$

The cross power spectrum magnitude cannot exceed the geometric mean of the total autopower spectra, i.e., we can define

$$\alpha \equiv \frac{|P_{12}^{(S)}|^2}{(P_1^{(S)} + P_1^{(N)})(P^{(S)} + P^{(N)})}, \tag{19}$$

which must satisfy $0 \leq \alpha \leq 1$. Then

$$P^{(S)} - P^{(\hat{s})} = \frac{(P^{(S)} - \frac{\alpha}{1-\alpha}P^{(N)})P^{(N)}}{P^{(S)} + P^{(N)}}. \tag{20}$$

In the case of $\alpha = 0$, where there is no correlation between observables 1 and 2, we simply recover the power spectrum bias of the Wiener filter for observable 2 alone. However, it is clear that nonzero correlation (i.e., nonzero values of α) would reduce the bias down from this limit, potentially significantly depending on the relative amplitudes of the signal and noise power spectra. This would suggest that cross-correlations allow recovery of additional fluctuations in the underlying signal.

C. Special case: $s_i = B_i s$ for all i

Suppose all our observations are of the same underlying signal, but subject to a different linear bias. One might

consider, for example, that fluctuations in line brightness temperature δT_{line} trace the matter density contrast δ_m scaled by the luminosity-weighted average temperature-bias product $\langle Tb \rangle_{\text{line}}$. In a halo model that relates halo mass M_h and redshift z to a line luminosity $L(M_h, z)$, we may calculate this quantity as

$$\langle Tb \rangle_{\text{line}} = \frac{c^3 (1 + z)}{8 \pi k_B \nu_{\text{rest}}^3 H(z)} \int dM_h \frac{dn}{dM_h} L(M_h, z) b(M_h),$$
(21)

given the linear bias $b(M_h)$ for halos of virial mass M_h, the halo mass function dn/dM_h, the rest-frame line emission frequency ν_{rest}, the Hubble parameter $H(z)$ for the assumed cosmology, the speed of light c, and the Boltzmann constant k_B. [When working in units of flux density rather than brightness temperature, replacing the multiplier in front of the integral with simply $c/[4\pi\nu_{\text{rest}}H(z)]$ in Eq. (21) gives the mean intensity-bias product. See, e.g., Ref. [62]].

Then on large linear scales, the matter power spectrum P_m and the line-intensity power spectrum P_{line} are related by

$$P_{\text{line}}(k) \simeq \langle Tb \rangle_{\text{line}} P_m(k),$$
(22)

where we ignore the Poissonian shot noise that begins dominating at nonlinear scales. This is hardly a novel insight, and the vast majority of LIM forecasts rely on halo models with linear bias. Galaxy number density contrast as measured by a galaxy survey will of course have a similar effective bias b_{gal} such that the galaxy density contrast has a clustering power spectrum $P_{\text{gal}}(k) \simeq b_{\text{gal}} P_m(k)$. The key point is that this bias parameter, be it $\langle Tb \rangle_{\text{line}}$ or b_{gal}, differs across the range of atomic lines, molecular lines, and galaxy selections. If we had only observations of the exact same tracer, the LCB filter would simply reduce to an inverse variance-weighted Wiener filter, which we show explicitly in Appendix C.

Return to a more general picture for a moment. Instead of dealing specifically with δ_m and some combination of $\langle Tb \rangle_{\text{line}}$ or b_{gal} parameters, suppose we have an underlying field s and observable fields s_i for $i \in \{1, …, n, p \equiv n+1\}$ all tracing s with some linear bias such that $s_i = B_i s$ for all i. Then the signal auto- and cross power spectra are all proportional to some common power spectrum. Call this $P(k)$, so that $P_{ij}^{(S)}(k) = B_i B_j P(k)$.

As before we would like to go about reconstructing s_p. Defining the appropriate matrix D, we have

$$D_{ij}(k) = \begin{cases} B_i B_j P(k) + \delta_{ij} P_i^{(N)} & \text{if } i \leq n \quad \text{and} \quad j \leq n, \\ B_i B_j P(k) & \text{otherwise.} \end{cases}$$
(23)

With prior knowledge of B_p and $P(k)$, application of the LCB filter is then straightforward. However, what if we were to plead ignorance of either B_p or $P(k)$? Even in the absence of correlated noise, we would not be able to claim knowledge of $B_p^2 P(k)$ unbiased by $P_p^{(N)}$ from one dataset d_p by itself.

In the case where $s_i = s_j$ for all i and j (e.g., all $B_i = 1$), every cross spectrum is effectively an unbiased estimator of the signal power spectrum. As we mentioned above, Appendix C shows that this case simply reduces to the Wiener-filtered reconstruction of the inverse noise variance-weighted average of d_i.

If the signals are biased by factors B_i that are not necessarily equal, estimation of $P^{(S)}$ is less straightforward. In fact, if $n = 1$ (i.e., we have two independent observations), we would only have one cross spectrum, and this would not be sufficient for unbiased estimation of $B_p^2 P(k)$.

However, $n \geq 2$ is sufficient to estimate $B_p^2 P(k)$ based on the cross spectra. It is straightforward to see that, in the absence of correlated noise, we expect

$$\hat{P}_p^{(S)}(k) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{P_{ip}^{(S)}(k) P_{jp}^{(S)}(k)}{P_{ij}^{(S)}(k)}.$$
(24)

Having an estimator in this vein then allows us to define D purely in terms of total auto- and cross power spectra. Denoting the total autopower spectrum of d_i as $P_i^{(T)} \equiv P_i^{(S)} + P_i^{(N)}$, we have

$$D_{ij}(k) = \begin{cases} P_i^{(T)}(k) & \text{if } i = j \leq n, \\ P_{ij}^{(S)}(k) & \text{if } i \neq j, \\ \hat{P}_p^{(S)}(k) & \text{if } i = j = p. \end{cases}$$
(25)

Note that none of this actually requires $B_i \neq B_j$ for all i and j. So for instance, we could leverage two independent observations of one line to reconstruct a third independent observation of a different line tracing the same matter density fluctuations.

III. INTERMEZZO: QUANTIFYING THE SUCCESS OF OPTIMAL ESTIMATORS

Before moving to consider potential applications of the LCB filter, we need to address how this work will quantify successful reconstruction of the structure of line-intensity fluctuations.

It is clear that the LCB filter results in optimal estimation of the target signal, but with a power spectrum that is suppressed relative to that of the original signal (although less so than estimation with the Wiener filter). Furthermore, the assumptions behind the design of the LCB filter ultimately almost all run up against the reality of the

LIM signal. The signal for many atomic and molecular lines will be intrinsically highly non-Gaussian if originating from later cosmic epochs and principally tracing galaxies (whose statistics have long been understood to behave less like a Gaussian random field and more like Poissonian draws from an underlying log-normal field—see, e.g., Coles and Jones [63]).

Furthermore, even if the line-intensity field in actual comoving space were a Gaussian random field, what a LIM survey will actually be able to observe is the *pseudo-intensity* field convolved with instrumental response and survey transfer functions. Thus, the output of the LCB estimator will be a *pseudo-pseudointensity* field, a suppressed version of the fluctuations already suppressed by the observation.

The central aim of this paper that was stated in the Introduction is not to show that the LCB filter reconstructs the amplitudes of line-intensity fluctuations. Rather, we will show that it allows reconstruction of the *structure* of fluctuations on large linear scales, in a way that improves over the Wiener filter alone. The morphology of line-intensity fluctuations on large scales, separate from their amplitude, is still of interest for many of the reasons we discussed in the Introduction.

Of course, we would like to show that the LCB filter improves over the Wiener filter in recovery of $P(k)$ amplitudes as well. However, to quantify the success of a given signal estimator $\hat{s}$ in recovering structure regardless of amplitude, we will calculate the normalized cross-correlation between the estimator and the original signal, as a function of k. Specifically, if the original signal has a power spectrum $P^{(S)}(k)$, the reconstruction has a power spectrum $P^{(\hat{s})}(k)$, and the signal-reconstruction cross power spectrum is $P^{(S \times \hat{s})}(k)$, the normalized cross-correlation is

$$r(k) \equiv \frac{P^{(S \times \hat{s})}(k)}{\sqrt{P^{(S)}(k)P^{(\hat{s})}(k)}}. \tag{26}$$

Perfect recovery of the phases of the Fourier modes, regardless of amplitude, corresponds to $r = 1$.

IV. SIMULATIONS OF [C II] LIM RECONSTRUCTION: PROOF OF CONCEPT

Having considered the formalism around the LCB filter and how to quantitatively establish successful reconstruction, we are now ready to consider applications to LIM reconstruction. Our first simulated case study is a futuristic [C II] LIM survey that has a similarly futuristic catalog of $z = 5.5$–6.6 Lyman-alpha emitters (LAEs) available for cross-correlation. We will begin by explaining the scientific context behind [C II] LIM in Sec. IV A, before defining the survey model in Sec. IV B and the line emission signal and interloper models in Sec. IV C. We

will then be prepared to consider reconstruction results in simulations in Sec. IV D.

For this section, we assume cosmological parameter values of $\Omega_m = 0.307$, $\Omega_\Lambda = 0.693$, $\Omega_b = 0.048$, $h = 0.678$, $\sigma_8 = 0.829$, and $n_s = 0.96$, to maintain consistency with the Small MultiDark-Planck (SMDPL) simulation [64] that we will use below. This cosmology is also consistent with the Planck 2015 results [65].

A. Context: Why ionized carbon and Lyman-alpha emitters?

The [C II] line, emitting at a rest-frame frequency of 1900.5 GHz, is of interest at these high redshifts as a potential signpost for star formation and diffuse interstellar gas in the late epoch of reionization. [C II] is a key cooling line for the interstellar media of galaxies, accounting for as much as 1% of the total far-infrared luminosity [66], and correlates with star-formation rate in population studies, both in nearby galaxies [67,68] and at redshifts reaching into the tail end of the epoch of reionization [69]. [C II] also traces ionized and diffuse neutral phases of the interstellar medium, as shown by detailed line diagnostics in nearby galaxies [70] and extended emission at high redshift [71–75], which has motivated use of [C II] as a tracer of neutral or H I gas in galaxies across cosmic history [76,77]. Whether it traces the assembly of the first star-forming galaxies assembling postreionization or neutral gas accreting into galaxies, the [C II] LIM signal has strong complementarity with 21 cm tomography and provides a key probe of the late epoch of reionization.

Searches for [C II] emission from $z \simeq 2.5$ using cross-correlation of Planck 545 GHz data with quasar positions have yielded tentative results [43,46] but depend heavily on accurate models of correlated continuum emission due to the broadband nature of the intensity data. This is a problem not shared to the same extent by LIM surveys with many ($\gtrsim 10^2$) channels, which should confine correlated continuum emission to the lowest-k line-of-sight modes and thus be theoretically capable of rejecting correlated continuum with minimal signal loss [78]. (We will neglect correlated continuum contamination for the simulations in the remainder of this work, which for our purposes should be equivalent to assuming that the channel count of these surveys allows appropriate debiasing of signal power spectra to correctly scale the reconstructed line intensity.) This has motivated the development of numerous [C II] LIM pathfinder experiments [8,17,24].

However, LIM observations of [C II] from $z \sim 6$ (so at $\simeq 270$ GHz) will see significant interloper line emission from the rotational transitions of CO at lower redshift. The $CO(J \to J-1)$ lines, which have rest-frame frequencies of $\simeq 115 \cdot J$ GHz, trace cold molecular gas that fuels star formation and are thus interesting in their own right, and we will discuss CO LIM in Sec. V. However, the interloper CO(3–2) through CO(7–6) emission components are a

significant concern for [C II] surveys as they originate from redshifts 0.3–2.0, parts of cosmic history that are full of much closer, more metal-rich, comparatively more massive, and more actively star-forming galaxies compared to $z \sim 6$. We may also consider the neutral carbon [C I](1–0) line (rest-frame frequency of 492.16 GHz) as an interloper along the same lines as the CO lines, as it too acts as a tracer of cold molecular gas [79–81]. Previous literature has extensively explored different ways of reducing interloper bias in the measured [C II] power spectrum [31–35], but here we consider the even more ambitious possibility of reconstructing the [C II] fluctuations themselves without interloper bias by leveraging cross-correlation against a hypothetical LAE catalog.

We choose to simulate a LAE catalog because previous surveys (e.g., Refs. [82,83]) suggest LAEs are abundant at these redshifts, with Ouchi *et al.* [83], for example, finding > 50 LAEs per square degree per $\Delta z \simeq 0.1$. Combining better photometry with next-generation integral field units or massively multiplexed spectrographs, future surveys could identify thousands of LAEs across wide fields and wide continuous redshift ranges.

However, note that for the purposes of this work, what matters most is not the exact galaxy selection for the most part, but rather the source density, the correlation of source density contrast with the signal to be reconstructed, and the effective linear bias of the galaxy sample. This last factor affects the amplitude of the cross power spectrum, thus affecting detectability for fixed survey sensitivity. Ultimately, even if a wide-field flux-limited LAE survey is not feasible, any spectroscopic survey with similar source abundances that selects for similarly star-forming galaxies would serve equally well as a cross-correlation target for some future LIM survey.

B. Survey parameters and Gaussian noise model

The futuristic [C II] survey that we simulate here takes its cues from the proposed road map of Karkare *et al.* [84], which suggests that future stages of [C II] surveys should increase geometrically in sensitivity. The futuristic survey concept considered in Ref. [62] fits in stage 3 of the road map of Karkare *et al.* [84], with $N_{\mathrm{spec}} t_{\mathrm{surv}} = 1.8 \times 10^8$ spectrometer hours covering 1024 square degrees of sky.

We follow Ref. [62] and calculate the noise-equivalent intensity per detector as

$$\sigma_{\mathrm{spec}} = 10^6 \ \mathrm{Jy \ sr^{-1} \ s^{1/2}} \left(\frac{\delta \nu}{2.5 \ \mathrm{GHz}} \right)^{-1/2} \left(\frac{\epsilon_{\mathrm{sys}}}{0.05} \right), \quad (27)$$

given channel bandwidth $\delta \nu$ and system emissivity ϵ_{sys}, under the implicit assumption that the detection bandwidth is equal to the frequency channel width. Unlike Ref. [62], we will assume $\epsilon_{\mathrm{sys}} = 0.05$ for simplicity, and we will assume $\delta \nu = 312.5 \ \mathrm{MHz}$ for added futurism.

Assuming adequate control of atmospheric and other sources of frequency-dependent Earth-bound noise, the final "thermal" map noise per voxel should then be

$$\sigma_N = \frac{\sigma_{\mathrm{spec}}}{\sqrt{N_{\mathrm{spec}} t_{\mathrm{surv}} \Omega_{\mathrm{pix}} / \Omega_{\mathrm{surv}}}}. \quad (28)$$

While we assume the LIM survey covers a total of 1024 square degrees of sky, we conservatively assume only a square 9 deg^2 subfield can be cross-correlated with an overlapping LAE catalog and thus reconstructed with the LCB filter. This 3×3 deg^2 patch will be simulated with 512×512 pixels, for a solid angle per pixel of $\Omega_{\mathrm{pix}} \simeq 1.05 \times 10^{-8}$ sr. With all appropriate substitutions made, we find $\sigma_N \simeq 1.9 \times 10^4 \ \mathrm{Jy \ sr^{-1}}$.

We also assume a Gaussian beam profile with a full width at half maximum (FWHM) of 50 arc sec, taken to be achromatic for simplicity. This approximately corresponds to the diffraction limit for a 6 m telescope observing at 250 GHz, so it is quite plausible to intentionally degrade the observation to that common resolution across all frequencies after the fact (e.g., to avoid mode mixing where possible).

C. Simulation parameters and line emission models

For [C II] simulations, our starting point is the SMDPL simulation, which is a cosmological N-body box of comoving size $400 h^{-1}$ Mpc and mass resolution $9.6 \times 10^7 h^{-1} M_\odot$. More precisely, the starting point relevant to this work is the catalog of dark matter halos identified in the simulation across space and time. Given the mass resolution, the halo population should be reasonably complete down to virial masses of $M_h \simeq 10^{10} M_\odot$; we set this as the minimum halo mass for nonzero line luminosity.

We use publicly available halo catalogs (online at time of writing at https://www.peterbehroozi.com/data.html) that have already been processed through the UniverseMachine framework of Behroozi *et al.* [85]. This framework provides an empirical galaxy model based on associating halo merger histories with star formation and, as a result, assigns galaxy properties such as star-formation rate (SFR) and dust extinction (A_{UV}) to every halo in each simulation snapshot. The UniverseMachine release includes code to generate light cones from these halo catalogs, so we generate ten light cones spanning 3×3 deg^2 and $z = 0$–9. These will be useful for all kinds of LIM predictions, especially given the mass resolution of the SMDPL simulation, but we leave further exploration to future work.

We convert the galaxy properties generated by UniverseMachine into line luminosities relevant to our [C II] survey in the frequency range of 250–290 GHz, which include not only the [C II] line from redshifts $z = 5.5$–6.6, but also CO rotational transitions as well as the atomic carbon [C I] line. We use correlations between SFR and line luminosities summarized by Ref. [62], which in turn adapts

all of the relevant relations from the work of Sun *et al.* [86]. We briefly resummarize them here:

(i) The [C II] luminosity is assumed on average to be linearly proportional to the SFR (via UV luminosity), with the proportionality derived from combining values given by Refs. [86,87],

$$\log\left(\frac{L_{[\text{C\,II}]}}{L_\odot}\right) = \log\left(\frac{\text{SFR}}{M_\odot\,\text{yr}^{-1}}\right) + 7.34. \quad (29)$$

In addition, we assume a log-normal scatter around this average with a standard deviation of 0.4 (in units of dex).

(ii) The CO transitions and the [C I] line are all assumed to correlate perfectly with the CO(1–0) luminosity, which in turn is assumed on average to be related to the SFR (via IR luminosity) through a power law,

$$\log\left(\frac{L_{\text{CO}(1-0)}}{L_\odot}\right) = \alpha_{\text{IR–CO}}^{-1}\left[\log\left(\frac{\text{SFR}}{M_\odot\,\text{yr}^{-1}}\right) + 9.76 - \beta_{\text{IR-CO}}\right] - 4.31, \quad (30)$$

where $\alpha_{\text{IR-CO}} = 1.27$ and $\beta_{\text{IR-CO}} = -1.0$. We also assume a log-normal scatter around this average with a standard deviation of 0.4 (again in units of dex). From this scattered CO(1–0) luminosity, the luminosities for the CO($J \to J-1$) lines can be calculated as follows:

$$L_{\text{CO}(J\to J-1)} = r_J J^3 L_{\text{CO}(1-0)}, \quad (31)$$

where we use the values of $r_3 = 0.73$, $r_4 = 0.57$, $r_5 = 0.32$, $r_6 = 0.19$, and $r_7 = 0.1$. We essentially calculate the [C I] luminosity in the same way, treating it as if it were another CO line with $J_{[\text{C\,I}]} \equiv \nu_{\text{rest},[\text{C\,I}]}/\nu_{\text{rest},\text{CO}(1-0)} = 4.27$,

$$L_{[\text{C\,I}]} = r_{[\text{C\,I}]} J_{[\text{C\,I}]}^3 L_{\text{CO}(1-0)}, \quad (32)$$

where $r_{[\text{C\,I}]} = 0.18$.

All line ratios relative to the CO(1–0) luminosity are taken from Sun *et al.* [86], who in turn refer the reader to Refs. [80,81,88] for observational literature probing the correlations of the individual line species with IR luminosity.

In addition, we also need to estimate Lyman-alpha luminosities in order to simulate a flux-limited (and therefore luminosity-limited) LAE sample. We once again look to Sun *et al.* [86], who provide an approximate relation between the SFR and the Lyman-alpha luminosity assuming equilibrium between ionization and recombination, with the Lyman-alpha photons coming from recombination events only. Given $f_\gamma = 4000$ ionizing photons produced per stellar baryon, an escape fraction $f_{\text{esc}} = 0.1$ of ionizing photons that thus do not lead to recombination events, a fraction $f_{\text{Ly}\alpha} = 0.67$ of recombination events producing Lyman-alpha emission, and an escape fraction $f_{\text{esc}}^{\text{Ly}\alpha} = 0.6$ of Lyman-alpha photons that reach the observer,

$$L_{\text{Ly}\alpha} = \frac{f_\gamma \text{SFR}/(10^{A_{\text{UV}}/2.5})}{m_p/(1-Y)}(1 - f_{\text{esc}})f_{\text{Ly}\alpha}f_{\text{esc}}^{\text{Ly}\alpha}E_{\text{Ly}\alpha}. \quad (33)$$

Here, m_p is the proton mass, $Y = 0.24$ is the helium mass fraction, A_{UV} is the dust extinction (recalling that UniverseMachine calculates A_{UV} alongside SFR for each halo), and $E_{\text{Ly}\alpha} = hc/(1215\,\text{Å})$ is the energy per Lyman-alpha photon. As Sun *et al.* [86] acknowledge, the resulting model still ignores environmental factors that affect the distribution of LAEs, but matches luminosity functions and tracer bias constraints in the literature derived from a 14–21 deg^2 Subaru Hyper-Suprime Cam (HSC) survey [83]. This is sufficient for our purposes as we mostly want to reproduce realistic source densities and biases rather than exact source selections, at least in the context of the present work.

We use LIMLAM_MOCKER (available at time of writing at https://github.com/georgestein/limlam_mocker) to load each simulation light cone, generate line luminosities for each halo, and bin the [C II] and interloper line luminosities into cubes of flux density (in units of Jy sr^{-1}) based on the redshift-space locations of the halos. The LIM survey instrument is assumed to have a Gaussian beam on the sky with 50 arc sec FWHM, so we convolve the flux density cube with the appropriate Gaussian kernel. We also subtract the mean from all flux density cubes, as LIM surveys tend to remove the mean level in analysis while subtracting continuum foregrounds. We then have a pseudointensity cube with 128 frequency channels spanning 250–290 GHz and 512 pixels along each angular dimension, to which we then add Gaussian noise with standard deviation σ_N derived in Sec. IV B.

At the same time, we generate a mock LAE catalog corresponding to each flux density cube by selecting only halos with an associated Lyman-alpha luminosity of $L_{\text{Ly}\alpha} > 10^{42.5}$ erg s^{-1}. This is somewhat more conservative than the cut of $10^{42.2} - 10^{42.3}$ erg s^{-1} used by Sun *et al.* [86] for their mock LAE survey. Note, however, that we are hoping for a wide-field survey to generate this LAE catalog across the whole redshift range, which likely will not be possible with the same source density as the Subaru HSC narrow-band survey of Refs. [83,89] (which only spans two redshift bands of $\Delta z \simeq 0.1$ each). The resulting catalogs provide between 7700 and 8600 LAE positions across our ten light cones.

We bin the LAEs into a density contrast cube, obtaining the source count in each voxel but then dividing by the mean and subtracting 1. We may now obtain total auto- and
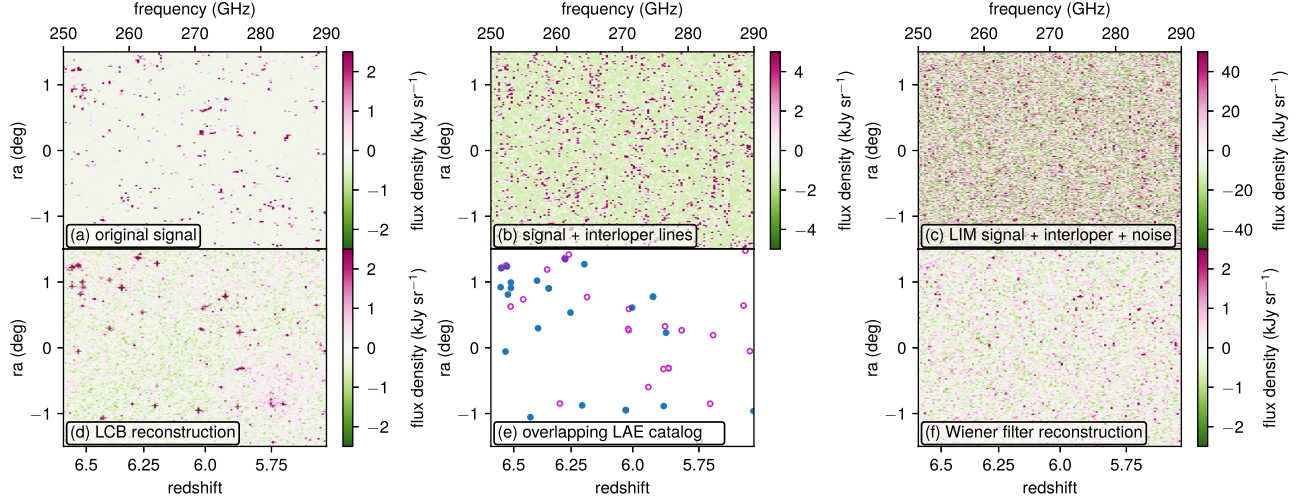
FIG. 1. Visualization of a slice of an example [C II] LIM simulation, showing six outputs: (a) the mean-subtracted [C II] signal in isolation; (b) the mean-subtracted [C II] signal summed with the mean-subtracted interloper CO and [C I] emission; (c) the total mean-subtracted signal and interloper emission plus Gaussian noise as described in Sec. IV B; (d) the LCB-filtered data; (e) the LAE positions contained in the slice shown of the data cube, including both sources properly contained in the slice's declination bin (filled circles) and sources in neighboring declination bins (unfilled circles); and (f) the Wiener-filtered data. Both the LCB and Wiener filter reconstructions assume full knowledge of the signal power spectrum.

cross spectra with this LAE density contrast cube and the LIM data cube, for each one of our ten semi-independent realizations.

The LCB filter, like the Wiener filter, requires knowledge of the signal power spectrum. We obtain a form for the pseudointensity power spectrum (i.e., the power spectrum of the [C II] flux density field after convolution with the instrument beam and mean subtraction) by averaging across all ten light cones, basically assuming that there is a reasonable estimate of at least the shape of the [C II] power spectrum from the wider LIM survey.

D. Simulation results

Before considering the filtering results, for completeness we note the signal-to-noise ratio summed across all k for both the [C II] auto and [C II]-LAE cross power spectra (assuming successful removal of interloper line emission in the former case). If N_m modes are available to the observation at wave number k, the variance for each power spectrum is given by

$$\sigma^2[P^{(S)}] = \frac{(P^{(S)} + P^{(N)})^2}{N_m}, \qquad (34)$$

$$\sigma^2[P_{12}^{(S)}] = \frac{(P_1^{(S)} + P_2^{(N)})(P^{(S)} + P^{(N)}) + P_{12}^{(S)2}}{2N_m}. \qquad (35)$$

with d_1 and d_2 respectively being the LAE and LIM observations. Note that in this case $P^{(N)}$ comes from the sum of the thermal noise described in Sec. IV B and the interloper line emission.

The total signal-to-noise ratio for a given power spectrum $P(k)$ is then given by the signal-to-noise ratio in each k-bin summed in quadrature,

$$(\mathrm{SNR})[P(k)] = \left(\sum_k \frac{P^2(k)}{\sigma^2[P(k)]}\right)^{1/2}. \qquad (36)$$

Across our ten different realizations, this total signal-to-noise ratio ranges from 91 to 117 for the [C II] autopower spectrum in the 9 deg^2 field alone (recalling that we have imagined a survey that spans over 1000 deg^2 in total) and from 312 to 337 for the [C II]-LAE cross power spectrum. While detectability forecasting is not the focus of the present work, this calculation shows that there is significant information content in the cross power spectrum of our mock surveys that the LCB filter will use but the Wiener filter will simply ignore.

1. Filtering with full knowledge of signal power spectrum

To start with, we assume that the power spectrum $P^{(S)}$ of the signal is fully known—possibly from the wider LIM survey—and used to design the LCB filter, as well as a Wiener filter for comparison. We first find the average signal power spectrum across all ten light cones, then use it while iterating through the light cones again to design the filters in each simulation, along with the actual observed total auto- and cross power spectra from that simulation. We show an example set of observables and reconstructions under this implementation in Fig. 1.

Note first the strength of the total interloper line-intensity contrast—which significantly exceeds that of the [C II] intensity contrast. The Wiener filter, lacking the

information needed to distinguish between the signal and interloper lines, ends up reconstructing neither component. The LCB filter, however, is able to leverage the cross-correlation between the signal component and the LAE catalog to isolate certain bits of flux density contrast that likely originate from $z \sim 6$ [C II]. The resulting reconstruction is not perfect, with some spurious bright spots and failure to recover bright spots without a colocated LAE overdensity. However, the large-scale fluctuations certainly appear soundly isolated from the interloper emission as well as the noise.

Moving from first impressions to quantitative results, we show in Fig. 2 the range across all ten simulations of relevant power spectra and normalized cross-correlation. Before considering the reconstructions, it is worth noting that in this futuristic LIM survey, the limiting factor in sensitivity to the [C II] power spectrum is not the thermal Gaussian noise, but the interloper line emission (which is what accounts for the difference between the total observed power spectrum and the combination of the true signal and thermal noise power spectra). In the presence of thermal noise alone, the Wiener filter would likely actually succeed in reconstructing at least the largest scales. However, with the interloper emission power spectrum being 10 times that of the signal power spectrum, the Wiener filter reconstruction is inevitably suppressed in amplitude by the corresponding amount. The normalized cross-correlation with the true signal is also poor, mostly staying near 30%–40%.

Using the measured [C II]-LAE cross-correlation, the LCB filter clearly improves reconstruction of the [C II] LIM signal. Compared to the Wiener filter, the LCB filter greatly reduces the power spectrum bias. More importantly, in the range of $k = 0.02$–1 Mpc^{-1} (between the lowest k available given the survey volume and the highest meaningful k given the survey resolution), the normalized cross-correlation between the LCB filter reconstruction and the true signal never falls below 50%, and reaches 80%–90% at the linear scales of $k \lesssim 0.1$ Mpc^{-1}.

Unsurprisingly, the normalized cross-correlation between the LCB reconstruction and the true signal is very similar to that between the LAE catalog and the true signal. This matches our first impressions from visual inspection of Fig. 1, in that we largely isolate the portion of the [C II] signal that correlates with the LAE galaxies. Knowledge of the expected [C II] power spectrum still aids the LCB filter, but does not contribute as much to the reconstruction as the cross-correlation, at least in this case.

2. Filtering with knowledge of shape but not amplitude of signal power spectrum

Now we consider a more realistic scenario where the power spectrum $P^{(S)}$ is not in fact fully known. In principle, if [C II] traces matter density fluctuations, the shape of
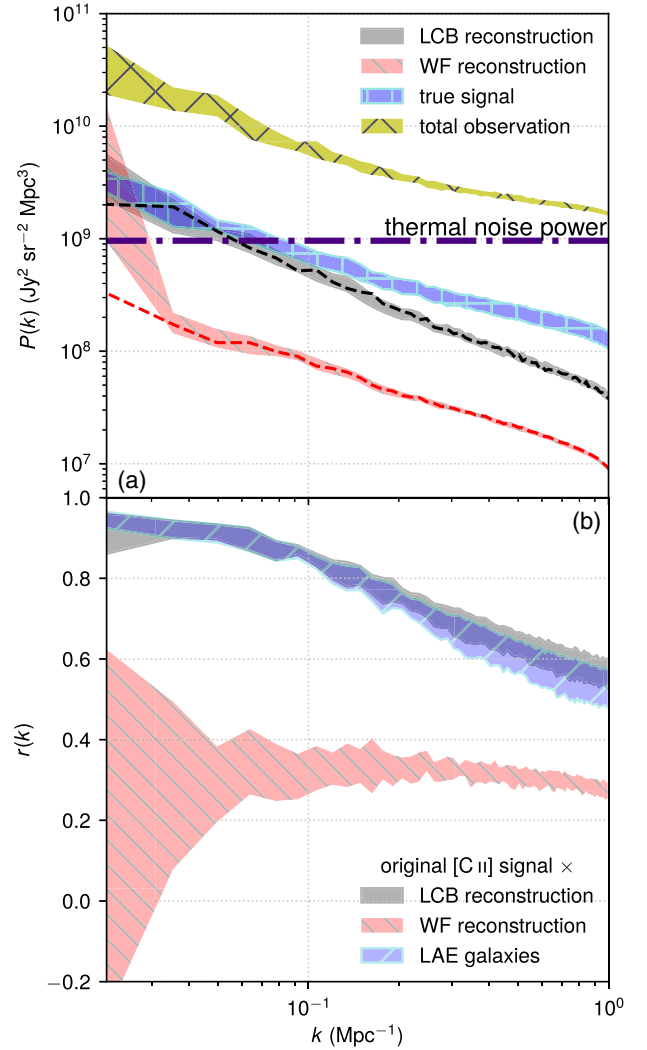

FIG. 2. The full interval of [C II] reconstruction results across our ten semi-independent light cones. We show (a) the obtained power spectra of the LCB and Wiener filter reconstructions alongside their expected values (dashed lines), as well as the original input signal, the total observation including noise and interloper emission, and the thermal noise power spectrum by itself (dash-dotted line). We also show (b) the normalized cross-correlation $r(k)$ between the original input signal and the two reconstructions, as well as between the [C II] emission and the LAE density contrast.

the [C II] power spectrum could be known beforehand to a great extent. The amplitude, however, could be imperfectly determined. The presence of interloper bias and other sources of additive noise could result in overestimation, but overcorrecting for these sources of biases could result in underestimation.

We thus consider calculating LCB and Wiener filter coefficients, based on an assumed $P^{(S)}(k)$ that is off by some multiplicative factor from the true $P^{(S)}(k)$, and examining the resulting LCB and WF reconstructions. We show the resulting change in the reconstruction
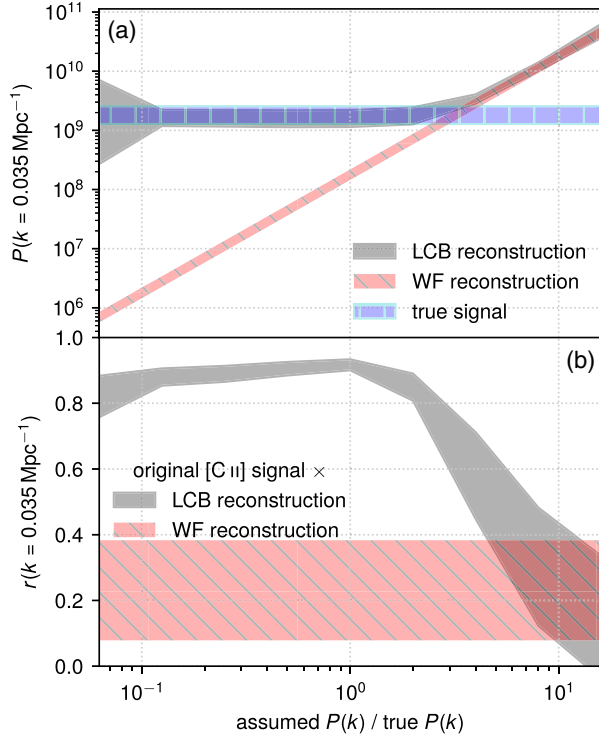
FIG. 3. Effect of calculating LCB and Wiener filter reconstructions of the [C II] signal with an assumed power spectrum $P^{(S)}(k)$ for the signal that differs from the true $P^{(S)}(k)$ by some multiplicative offset indicated along the x-axis. As a function of this offset, we show at $k = 0.035$ Mpc^{-1} the range across our ten simulated observations of (a) the value of the power spectra of the LCB and WF reconstructions (with the true signal power plotted for comparison) and (b) the normalized cross-correlation of each reconstruction with the true [C II] signal.

at $k = 0.035$ Mpc^{-1} in Fig. 3, with the multiplicative offset ranging from 1/16 to 16.

Neither filter is robust against overestimation of $P^{(S)}(k)$. This is sensible given the definitions of the filter coefficients, as noisy data will not be sufficiently down-weighted if we overestimate the contribution of the target signal to the total measured power spectrum.

However, the LCB filter *is* robust against underestimation of $P^{(S)}(k)$. Whereas the Wiener filter would naïvely down-weight what is in fact perfectly serviceable data, the LCB filter is anchored by the presence of external data and cross-correlation of the signal against those data. Therefore, even if we lose trust in the [C II] autopower spectrum, the [C II]-galaxy cross spectrum and the galaxy autospectrum together limit the resulting bias in the reconstruction.

This distinction is visible not only in the power spectrum of the LCB and WF reconstructions, but in their normalized cross-correlation against the true signal. Regardless of the exact coefficients, the Wiener filter is simply reweighting the same data, and so there is no reason for $r(k)$ to change
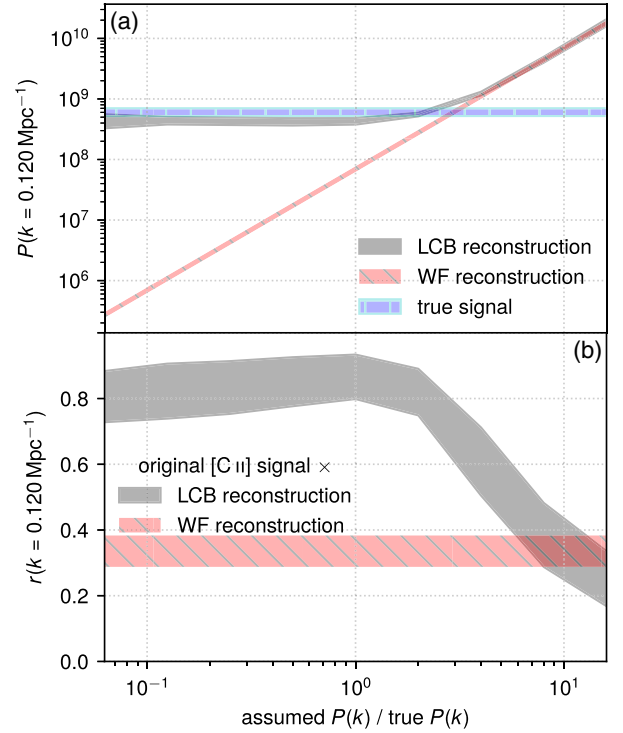
FIG. 4. Same as Fig. 3, showing evolution with signal power spectrum misestimation of (a) power spectra of reconstructions (with the true signal power plotted for comparison) and (b) normalized cross-correlation, but all at a higher wave number of $k = 0.120$ Mpc^{-1}.

with the filter coefficients. More generally, this suggests that, in the presence of strong interloper emission, a Wiener filter will never reconstruct the Fourier modes of the [C II] fluctuations with high fidelity.

However, while overestimation of the power spectrum by beyond a factor of 2 rapidly degrades the LCB reconstruction as well, underestimation by as much as a factor of 8 has relatively little effect, with $r(k = 0.035$ Mpc$^{-1})$ against the true signal staying above 80% thanks to the LAE cross-correlation filling the vacuum left by the down-weighted LIM data. We show in Fig. 4 the same results at a slightly higher $k = 0.120$ Mpc^{-1} to show some degree of k independence in these qualitative conclusions about the merits of the LCB filter.

Given its robust performance, the LCB filter would likely equally excel at recovering the interloper emission components individually, but we leave exploration of this possibility to future work.

V. POTENTIAL APPLICATIONS IN CO LIM

Different LIM experiments have different sets of foregrounds and interloper emission components, and here we will consider different possible future configurations of the CO Mapping Array Project as the focus of our remaining

TABLE I.   Key parameters for COMAP Pathfinder-like and COMAP-ERA simulations, either taken from [38] or newly assumed for this work.

Parameter		COMAP Pathfinder-like Ka band	COMAP-ERA Ku band	COMAP-ERA Ka band
Frequency coverage	(GHz)	26–34	13–17	26–34
Beam FWHM	(arcmin)	4.5	3.9	4.5
Pixel size Ω_{pix}	(arcmin^2)	4×4	4×4	4×4
Science channelization $\delta\nu$	(MHz)	15.625	7.8125	15.625
Nominal system temperature T_{sys}	(K)	44	20	44
Spectrometer count per dish N_{spd}		19	38	19
Solid angle per field Ω_{obs}	(deg^2)	2×2	2×2	2×2
Dish hours per field $N_{dish}t_{obs}$	(hr)	12500	57000	110000
Noise per voxel σ_N	(μK)	11.42	2.55	3.85
Number of fields per survey		3	3	3

simulations. We provide the context behind COMAP first in Sec. V A, then consider potential use cases for the LCB filter in near-term and future phases of COMAP operations respectively in Secs. V B and V C.

For this section, we assume cosmological parameter values of $\Omega_m = 0.286$, $\Omega_\Lambda = 0.714$, $\Omega_b = 0.047$, $h = 0.7$, $\sigma_8 = 0.82$, and $n_s = 0.96$, to maintain consistency with previous simulations used by Ref. [27].

A. Context: The COMAP Pathfinder-like survey, the COMAP expanded reionization array concept, and a fiducial CO model

1. COMAP road map and survey parameters

As we noted in Sec. IV A, the rotational transition lines of CO trace the cold molecular gas that fuels star formation [90]. Therefore, CO LIM has the potential to trace the evolution of the early star-forming environments from the late epoch of reionization to the epoch of peak star formation and galaxy assembly.

COMAP is currently the only operating LIM experiment with dedicated single-dish instrumentation targeting the low-J CO lines, having deployed a Pathfinder experiment in 2019 and published early science results from the first observing season as summarized by Ref. [21]. While the current Pathfinder phase of COMAP operates in the Ka band and covers 26–34 GHz in observing frequency, this corresponds to not only CO(1–0) emission from $z = 2.4$–3.4 (the main science target of the COMAP Pathfinder), but also to CO(2–1) emission from $z = 5.8$–7.9. Future phases of COMAP as described in Ref. [38] will deploy Ku-band instruments observing at 12–20 GHz, thus measuring CO(1–0) at $z = 4.8$–8.6 and allowing cross-correlation between Ka- and Ku-band observations to better constrain cosmic molecular gas content at $z \sim 7$.

Table I describes the parameters we assume for two phases of COMAP.

(i) The first is a COMAP Pathfinder-like survey, with slightly increased integration time per field compared to the fiducial Pathfinder survey design. This sensitivity could be readily achieved with extended Pathfinder observations and the addition of a second instrument would further facilitate matters.

(ii) The second is the COMAP Expanded Reionization Array (COMAP-ERA), a hypothetical stage 2 experiment (if we index the Pathfinder as stage 0) with most of the survey parameters as outlined in Ref. [38].

In both cases, we alter the voxel size so that the pixel size is roughly the beam size and the channel width significantly exceeds the native $\simeq 2$ MHz resolution of the COMAP spectrometers. Also in both cases, the noise per voxel is calculated as

$$\sigma_N = \frac{T_{sys}}{\sqrt{\delta\nu N_{spd} N_{dish} t_{obs} \Omega_{pix}/\Omega_{obs}}}. \tag{37}$$

2. CO model and simulation parameters

We must still define a model for CO emission and apply it to an ensemble of simulations. Unlike in Sec. IV, where we used UniverseMachine light cones from a single cosmological simulation with empirically determined star-formation rates for each halo, here we will make use of an ensemble of independent approximate N-body simulations using the peak-patch method [91]. Without merger histories for peak-patch halos, we will simply associate the halo mass M_h and cosmological redshift z

to an average star-formation rate and the star-formation rate to an average CO line luminosity.

For the CO luminosity model, we use the model of Li *et al.* [11] as modified by Keating *et al.* [16], assigning luminosities to all halos with virial mass above $10^{10}M_\odot$.

(i) The model begins by using the empirical model of Behroozi *et al.* [92,93] to interpolate the average star-formation rate for a given halo mass and redshift. Each halo is then assigned its appropriate star-formation rate with log-normal scatter of $\sigma = 0.3$ dex (although we implement the scatter in a way that preserves the linear mean).

(ii) The star-formation rate is assumed to be linearly proportional to the bolometric IR luminosity, via a factor of $10^{10}L_\odot(M_\odot\ \mathrm{yr}^{-1})^{-1}$.

(iii) Empirical fits to observations relate the IR and average $CO(J \to J - 1)$ luminosities,

$$\log\frac{L_{\mathrm{IR}}}{L_\odot} = \alpha_J\left(\log\frac{L_{\mathrm{CO},J}}{L_\odot} + 4.31\right) + \beta_J. \quad (38)$$

Specifically, following fits to a local sample of galaxies from Kamenetzky *et al.* [88], we use $\alpha_1 = 1.27$, $\beta_1 = -1.0$, $\alpha_2 = 1.11$, and $\beta_2 = -0.6$.

(iv) Once again, there is log-normal scatter of $\sigma = 0.3$ dex around this average CO luminosity for fixed IR luminosity.

Using LIMLAM_MOCKER once again, we "paint" CO luminosities so modeled onto two separate sets of peak-patch simulations, one for the $z \sim 3$ CO(1–0) signal and another for the $z \sim 7$ CO(1–0) and CO(2–1) signals.

(i) For the $z \sim 3$ signal, we use 9 of the 161 independent peak-patch light cone simulations generated for Ref. [27]. The simulations have a box size of $L_{\mathrm{box}} = 1140$ Mpc and a resolution of $N_{\mathrm{cells}} = 4096^3$. This is sufficient to span 26–34 GHz in CO(1–0) observing frequency and $9.6° \times 9.6°$ on the simulated sky, while resolving halos as small as $M_h \simeq 2.5 \times 10^{10}M_\odot$ at high completeness. We correct peak-patch halo masses via abundance matching to the halo mass function of Tinker *et al.* [94]. We divide each box into 16 subfields of $2° \times 2°$ each, for a total of 144 simulations of CO(1–0) at $z \sim 3$ as would be observed by COMAP.

(ii) For the $z \sim 7$ signal, we use 16 light cones out of a new set of peak-patch simulations designed specifically with reionization-epoch COMAP signals in mind. These simulations have $L_{\mathrm{box}} = 960$ Mpc and a resolution of $N_{\mathrm{cells}} = 5640^3$, thus spanning $6° \times 6°$ on the simulated sky as well as the CO(1–0) and CO(2–1) observing frequencies relevant for COMAP-ERA, while resolving halos slightly below $M_h \simeq 10^{10}M_\odot$. We again correct peak-patch halo masses via abundance matching to the halo mass

function of Tinker *et al.* [94], incorporating high-redshift corrections from Appendix G of Ref. [92]. We divide each box into nine subfields of $2° \times 2°$ each, for a total of 144 simulations of CO lines at $z \sim 7$ as would be observed by COMAP.

Here, we apply the two-tier line-broadening model of Ref. [95] to all simulations. All halos with virial mass $M_h > 10^{11}M_\odot$ are assigned a random inclination i and, based on this i as well as the virial velocity v_{vir} of the halo, also assigned a line profile FWHM given by $v_{\mathrm{vir}}\sin i/0.866$. These halos are then binned in 16 linearly spaced bins of line width, while the halos with virial mass $M_h < 10^{11}M_\odot$ are assumed to have negligible line width compared to the COMAP science channel bandwidth. The CO temperature cube calculated from each bin of halos is convolved with a Gaussian kernel of appropriate scale along the frequency dimension, so that the final line-broadened CO cube is the sum of the Gaussian-filtered 17 CO temperature cubes.

We randomly pair the $144 z \sim 3$ simulations with the $144 z \sim 7$ simulations, ending up with 144 simulated $2° \times 2°$ fields. In all cases the Ka-band cube is the sum of the $z \sim 3$ CO(1–0) and $z \sim 7$ CO(2–1) signals (with the instrument beam and survey transfer function applied to both as detailed for each survey below) with appropriately scaled Gaussian noise, while the Ku-band cube is the sum of the $z \sim 7$ CO(1–0) signal (again with the beam and transfer function applied) with appropriately scaled Gaussian noise.

B. COMAP Pathfinder-like survey: Reconstructing CO(1–0) with external cross-correlations

As noted above, the principal target for the Pathfinder survey is $z \sim 3$ CO(1–0) emission from the epoch of galaxy assembly and peak star formation. Here we consider the possibility of reconstructing this $z \sim 3$ signal, with the help of an overlapping galaxy catalog.

1. Methods

First, we further clarify the generation of mock observations. After summing the $z \sim 3$ CO(1–0) and $z \sim 7$ CO(2–1) signals, we convolve the total Ka-band brightness temperature into a Gaussian filter with the appropriate kernel for the beam size given in Table I for the Pathfinder-like observations, as before. However, we will simulate the effect of foreground subtraction in more detail than the simple mean subtraction that we used for simulations in Sec. IV.

Based on the heuristic form of the COMAP transfer function given by Ref. [96], and a measurement of $\sim 70\%$ main beam efficiency used for both COMAP high-redshift and Galactic analyses [97,98], we approximate the portion of the power spectrum transfer function not included in the Gaussian filter as

$$\mathcal{T}_{\rm hp}(k_\perp, k_\parallel) = \frac{0.49}{(1 + e^{5 - 100\ {\rm Mpc} \cdot k_\perp})(1 + e^{5 - 200\ {\rm Mpc} \cdot k_\parallel})}.$$

(39)

We multiply the Fourier transform of the beam-convolved CO cube by the square root of this function, and then invert the Fourier transform to obtain a high-pass-filtered signal scaled appropriately by the beam efficiency. These steps transform the signal from a physical brightness temperature into a pseudotemperature, which for convenience we will notate as T. All power spectra for the remainder of this whole section are implicitly pseudopower spectra.

By contrast, we will not consider the overlapping galaxy survey in nearly as much detail, given the limited information about each halo in the peak-patch simulations. We mainly select mock "galaxies" from the halos in the $z \sim 3$ simulation with virial mass above a certain $M_{h,\min}$, with additional selection possible if desired after that. To gauge the effect of increasing source density and decreasing galaxy bias, we try three different types of selection:

(i) We suppose that the spectroscopic galaxy survey catalogs the locations of all halos with virial mass above $M_{h,\min}/M_\odot = 6 \times 10^{12}$.

(ii) We suppose that the spectroscopic galaxy survey catalogs the locations of all halos with virial mass above $M_{h,\min}/M_\odot = 3 \times 10^{12}$.

(iii) We mock a $z \sim 3$ LAE survey similar to the Hobby-Eberly Telescope Dark Energy Experiment (HETDEX; see Refs. [99–101]) by using a slightly modified version of the LAE model of Ref. [96], which assumed that 5% of halos with virial mass above $9.3 \times 10^{10} M_\odot$ would be LAE hosts. Here, we first select all halos with virial mass above $M_{h,\min}/M_\odot = 10^{11}$, randomly select 5% of these halos (mimicking a Lyman-alpha emission duty cycle), and then apply a regular mask across the whole survey field that leaves unmasked $50'' \times 50''$ squares with centers spaced apart by $100''$ in both angular directions. This masking step culls 3/4 of the mock LAEs from the catalog and is meant to mimic the sparse sampling of the HETDEX survey, although the actual HETDEX window function will be more complicated. Note that both the 5% duty cycle and sparse sampling will induce additional shot noise uncorrelated with the CO signal.

The resulting source counts per patch for each selection are 718^{+103}_{-98}, 2872^{+312}_{-318}, and 5460^{+247}_{-268} (90% intervals) across 144 simulations. Source densities of $\sim 10^3$ deg^{-1} per $\Delta z = 1$ at this redshift range are eminently plausible from currently operating experiments like the aforementioned HETDEX LAE survey, which expects to lleverage integral field spectroscopy to catalog roughly 2000 LAE positions per square degree spanning $z = 2$–3.5 even with sparse coverage filling only 1/4.5 of the nominal survey area.

Future facilities will also readily achieve such source densities, including ground-based highly multiplexed multifiber spectrographs like DESI-II or MegaMapper targeting $z > 2$ Lyman-break galaxies [102,103] and the Nancy Grace Roman Space Telescope whose reference survey strategy includes surveying 2×10^6 [O III] emitters at $z = 2$–3 across 2000 deg^2 [104].

We simulate the observations in 48 batches of three fields, mimicking a single COMAP "survey." Each survey shares the same LCB and WF coefficients, designed as follows:

(i) We mimic the feed-feed pseudo-cross-spectrum (FPXS) estimator used by Ref. [97] for the CO power spectrum in each field. After generating 38 different mock data cubes with the same signal but with independently generated Gaussian noise with standard deviation per voxel of $\sigma_N \sqrt{38}$ (assumed to be 38 disjoint splits, e.g., in detector and elevation range), we obtain 19×18 cross spectra between these cubes and average them to obtain the FPXS estimate for the CO autospectrum in that field. We then use the average of these FPXS estimators across all fields in the survey as our estimate for $P^{(S)}(k)$ purely from the noisy mock data rather than directly from the noiseless input signals. [This method will slightly bias the signal power spectrum due to the common presence of background CO(2–1) emission, but this background is sufficiently subdominant that the extent of overestimation will be in the range that the LCB filter has already shown to tolerate well in Sec. IV D.]

(ii) We obtain the galaxy power spectrum in each field and average across fields. The three-field average estimates the value of $P_1^{(T)}(k) = D_{11}$ required by Eq. (25) for the LCB filter.

(iii) We now average the mock cubes into a single cube with noise per voxel of σ_N and obtain cross spectra for each field. The average of the cross spectra across all fields is what we use for $D_{12} = D_{21} = P_{12}^{(S)}(k)$ as required by Eq. (25) for the LCB filter.

With the relevant auto- and cross power spectra estimated for the whole survey, we apply the LCB and WF reconstructions to each field, measuring reconstruction power spectra as well as normalized cross-correlation against the true (pseudo)signal on a per-field basis.

2. Results

As before, we note expected detectability briefly. We predict strong detections of CO-galaxy cross spectra with this Pathfinder-like survey. With either smaller catalogs of ~ 700 mass-selected objects or ~ 5500 mock LAEs for each 4 deg^2 field, the average total (all-k) signal-to-noise ratio for the cross power spectrum is 21, rising to 27 for larger mass-selected catalogs of ~ 3000 objects per field.
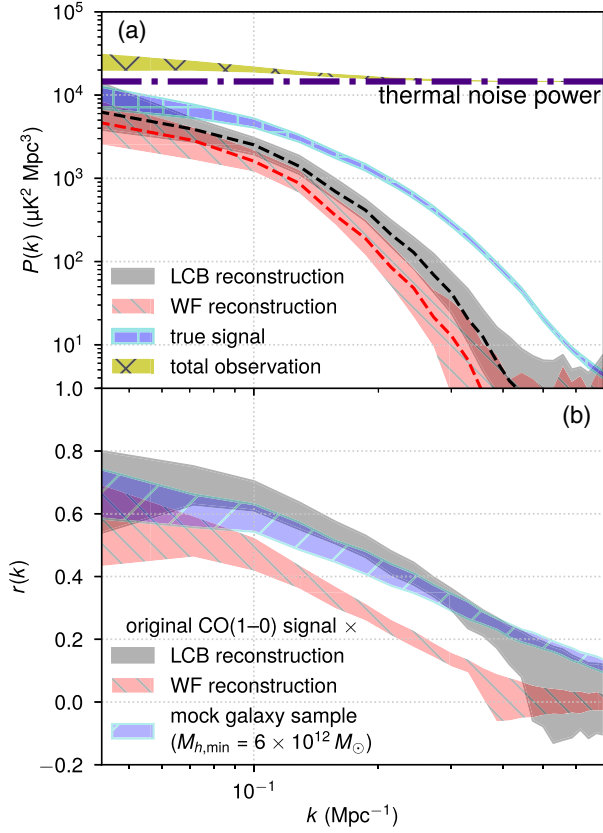
FIG. 5. 90% intervals of reconstruction summary statistics across our $144 = 48 \times 3$ simulated Pathfinder-like survey fields, given an overlapping galaxy sample simulated with $M_{h,\text{min}}/M_\odot = 6 \times 10^{12}$. We show (a) the obtained power spectra of the LCB and Wiener filter reconstructions alongside their expected values (dashed lines), as well as the original input signal, the total observation including noise and interloper emission, and the thermal noise power spectrum by itself (dash-dotted line). We also show (b) the normalized cross-correlation $r(k)$ between the original input signal and the two reconstructions, as well as between the noiseless CO(1–0) pseudosignal and the galaxy density contrast.
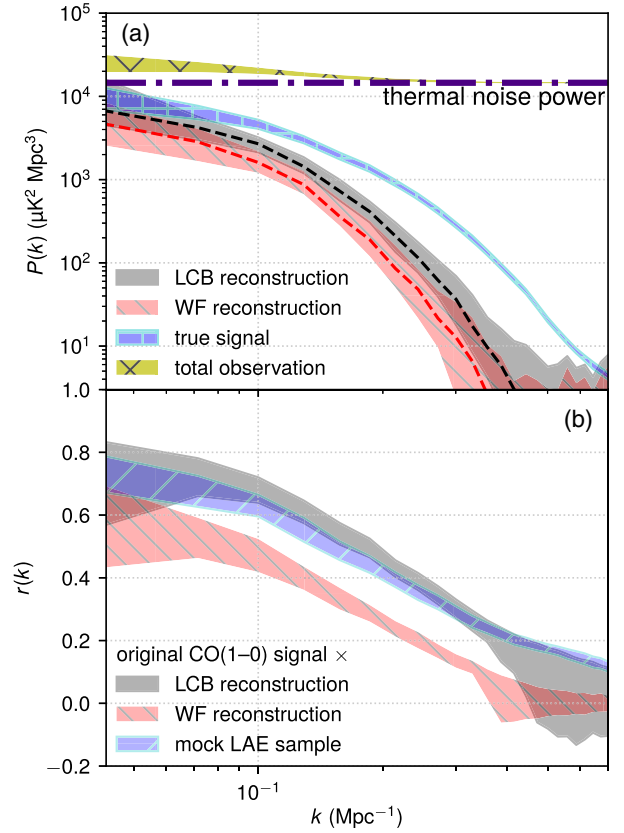
FIG. 6. Same as Fig. 5, but now for the mock LAE sample with sparse sampling. We again show 90% intervals across $144 = 48 \times 3$ simulated Pathfinder-like survey fields of (a) power spectra for signal, noise, and reconstructions, as well as (b) normalized cross-correlations between the noiseless CO(1–0) pseudosignal and either a reconstruction or the mock galaxy sample.

The CO(1–0) power spectrum is detectable with a total signal-to-noise ratio of 10 per field on average. This is consistent with the signal-to-noise forecasts from the COMAP Collaboration [96] to within a factor of order unity, once we account for various factors such as the number of fields, the inclusion of the contribution of sample variance to noise, differences in voxelization of the survey volume, approximations made in our transfer function and noise models, calculation details of line broadening, and so on. Again, since the focus of this work is on introducing the reconstruction method rather than forecasting detectability, we consider the level of agreement adequate.

Figure 5 summarizes results for the less abundant mass-selected mock galaxy sample with $M_{h,\text{min}}/M_\odot = 6 \times 10^{12}$. Both LCB and WF reconstructions result in significantly suppressed power spectrum amplitudes, but the LCB filter is able to leverage cross-correlations to significantly improve the normalized cross-correlation against the true pseudosignal, meaning that we recover the correct structure of fluctuations down to smaller scales. Note that at lower k, $r(k)$ is actually slightly higher on average between the LCB reconstruction and the true signal than between the mock galaxies and the true signal. The LCB filter in practice leverages not only the CO-galaxy cross-correlation but also the knowledge of expected structure at large scales, filling in large-scale fluctuations not necessarily completely mapped by our relatively low-density galaxy sample.

Results are similar with our mock LAE sample, as sown in Fig. 6. Despite a sample just over 7.5 times more abundant than the $M_{h,\text{min}}/M_\odot = 6 \times 10^{12}$ sample, the lower tracer bias and noise-inducing selection of this sample results in similar detectability and thus similar reconstruction improvement to the lower-abundance sample. But compared to that sample, the mock LAE sample does trace the large-scale modes with higher fidelity, resulting in slightly (although not significantly) improved $r(k)$ and slightly reduced $P(k)$ bias at low k. Of course, this
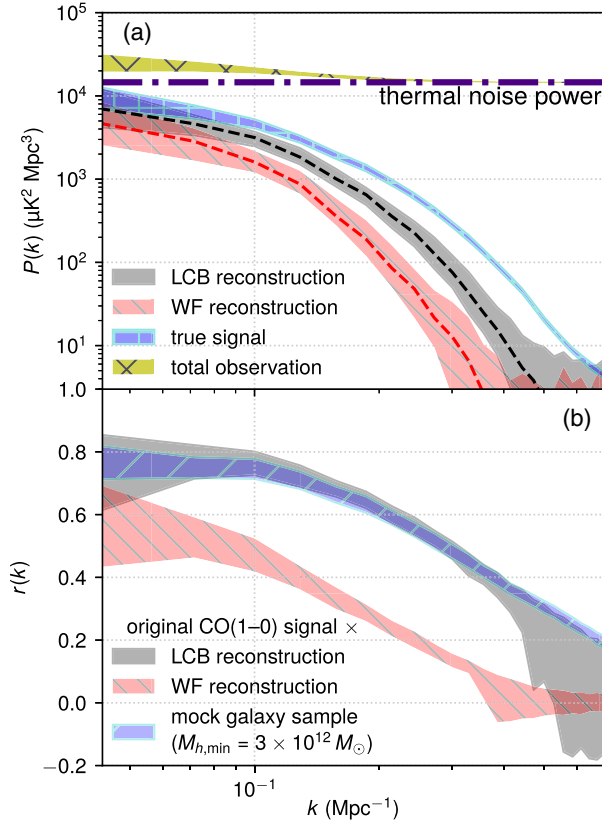
FIG. 7. Same as Fig. 5 or Fig. 6, but now for an overlapping galaxy sample with $M_{h,\mathrm{min}}/M_\odot = 3 \times 10^{12}$, leading to high source abundances but lower shot noise. We still show 90% intervals across $144 = 48 \times 3$ simulated Pathfinder-like survey fields of (a) power spectra for signal, noise, and reconstructions, as well as (b) normalized cross-correlations between the noiseless CO(1–0) pseudosignal and either a reconstruction or the mock galaxy sample.

in turn slightly increases the advantage over the Wiener filter reconstruction.

The improvement of the LCB filter over the Wiener filter is even more significant with a more abundant mass-selected mock source catalog with $M_{h,\mathrm{min}}/M_\odot = 3 \times 10^{12}$. Figure 7 shows significantly reduced suppression of the power spectrum of the LCB-filtered data and even more marked improvement in the normalized cross-correlation, which now stays above 50% for all $k \lesssim 0.3~\mathrm{Mpc}^{-1}$. The Wiener-filtered data only reach this $r(k)$ range for $k \lesssim 0.1~\mathrm{Mpc}^{-1}$, where the LCB-filtered data reach typical $r(k)$ values of 70%–80%. With a more abundant galaxy sample, we also see a reduced difference compared to Fig. 5 between the normalized cross-correlations of the true signal against the LCB reconstruction versus the galaxy sample. This suggests a scenario more similar to that of Sec. IV where the galaxy sample is sufficiently abundant that the cross-correlation drives the reconstruction at all scales.

We conclude consideration of the Pathfinder-like scenario by showing in Fig. 8 slices of a typical simulation of

signals, observations, and reconstruction, with a source catalog given by $M_{h,\mathrm{min}}/M_\odot = 3 \times 10^{12}$. Note that the $z \sim 3$ CO(1–0) signal clearly dominates over the $z \sim 7$ CO(2–1) background, but noise clearly dominates the total observation. As a result, the Wiener filter reconstructs a much weaker set of fluctuations compared to the original signal, whereas the LCB filter as before reconstructs something much closer to the original line-intensity fluctuations where the CO contrast coincides with galaxy density contrast.

C. COMAP-ERA: Isolating CO(2–1) with internal cross-correlations

The $z \sim 3$ CO(1–0) emission dominating over $z \sim 7$ CO(2–1) emission presents a boon for the Pathfinder-like survey, but an interloper and a nuisance for COMAP-ERA, which needs to access the reionization-epoch signal. However, as we discussed in Sec. VA, the deployment of lower-frequency instrumentation will allow observations of CO(1–0) at the same redshifts. Thus cross-correlations internal to the COMAP-ERA survey (as opposed to with external datasets) will isolate the common $z \sim 7$ CO signal.

1. Methods

We simulate fields in 48 batches of three, thus mocking whole surveys at once as in Sec. VB. In fact, most of the simulation workflow is the same as in Sec. VB, so we enumerate the main differences.

First, we omit simulation of the FPXS estimation. However, we also "split" the Ku-band data in two, generating two data cubes with the same signal but noise per voxel scaled up by $\sqrt{2}$ and treating them as two independent observations (which they may be at some level if they are obtained from, e.g., different subsets of COMAP-ERA receivers). Thus for the LCB filter we have $n = 2$, instead of $n = 1$ as in the previous case studies. This Ku-band data split would not be useful if we were simply trying to reconstruct the $z \sim 7$ CO(1–0) signal as the result would reduce to the Wiener filter as discussed in Appendix C, but is a handy construction for use with the Ka-band observation to obtain an unbiased estimate of the CO(2–1) power spectrum, as described in Sec. II C. As before, this and all other total auto- and cross power spectra are averaged across the fields in each survey, with the survey-wide averages used for LCB filter calculations in each field of the survey.

We also adjust the power spectrum transfer function slightly to account for the changed relations between angular or frequency scales with comoving lengths at $z \sim 7$,

$$\mathcal{T}_{\mathrm{hp}}(k_\perp, k_\parallel) = \frac{0.49}{(1 + e^{5 - 138~\mathrm{Mpc} \cdot k_\perp})(1 + e^{5 - 144~\mathrm{Mpc} \cdot k_\parallel})}. \quad (40)$$

As the transfer function is calculated as a function of k, we apply it equally to Ku- and Ka-band CO cubes.
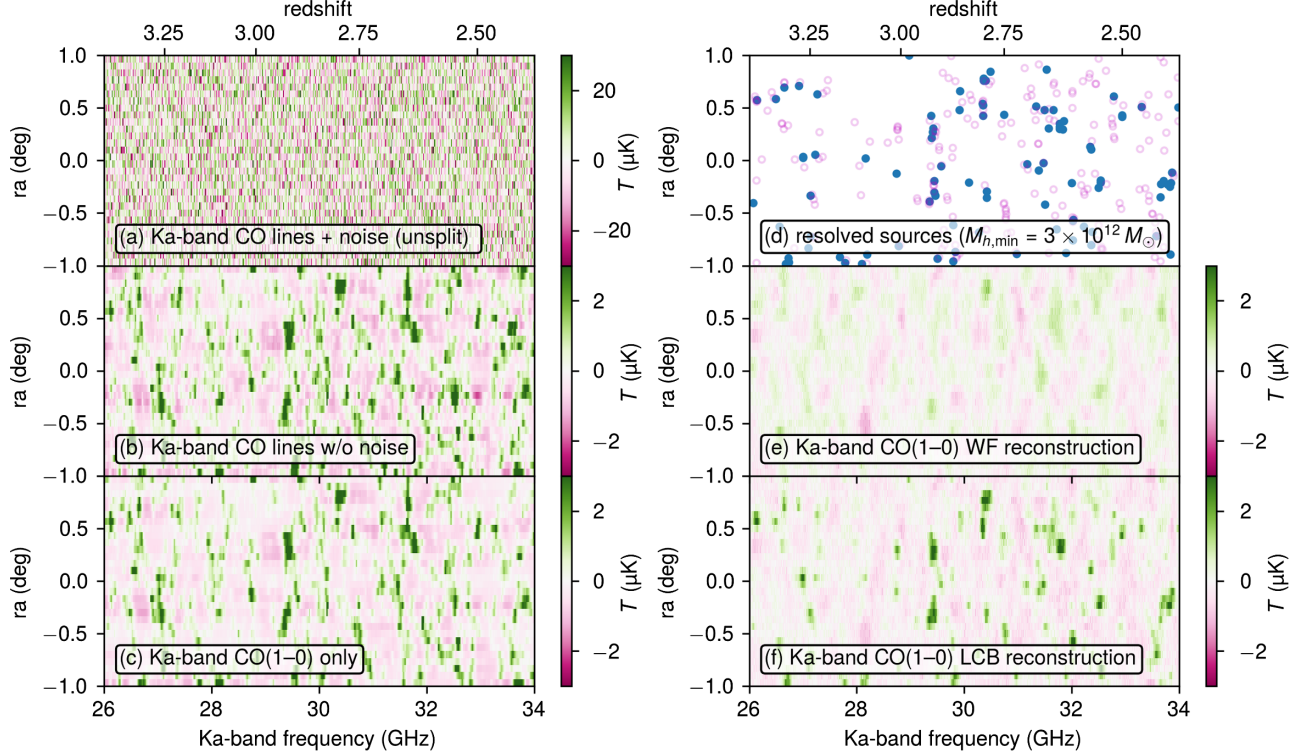
FIG. 8. Visualization of a slice of an example COMAP Pathfinder-like simulation, showing six outputs: (a) the total simulated Ka-band observation, including the CO(1–0) signal from $z \sim 3$, the CO(2–1) background from $z \sim 7$, and noise; (b) the sum of the noiseless filtered signal and background CO emission; (c) the noiseless $z \sim 3$ CO(1–0) pseudotemperature in isolation; (d) mock galaxy positions contained in the slice shown of the data cube, including both sources properly contained in the slice's declination bin (filled circles) and sources in neighboring declination bins (unfilled circles); (e) the Wiener-filtered data; and (f) the LCB-filtered data. The mock galaxy sample is generated based on a minimum halo mass of $3 \times 10^{12} M_\odot$.

For $z \sim 7$ CO(2–1), we offer the Wiener filter reconstruction the best possible chance to succeed by using the true signal pseudopower spectrum for each field with the total Ka-band autopower spectrum to Wiener filter the data. However, this advantage is entirely irrelevant in the presence of strong interloper emission, which the results of Sec. IV D should already suggest.

We also cut Fourier modes with $k < 0.01$ or $k > 0.7$ from the LCB reconstruction, as we expect cross-correlation information at those scales to be spurious at best due to the COMAP beam and pipeline transfer functions simulated here. This step is not necessary for the Wiener filter, as again we provide it with the true signal power spectrum.

2. Results

We briefly advertise average signal-to-noise ratio expectations, which are 52 for the Ku-band CO(1–0) signal, 12 for the Ka-band CO(2–1) signal, and 28 for the cross-band CO power spectrum. All these values are broadly in line with the projections of [38] for the same CO emission model and suggest that the cross power spectrum has significant information to aid in reconstruction with the LCB filter.

Figure 9 shows the key results across our simulations. The qualitative results are largely the same as for the [C II]

survey simulated in Sec. IV D. The principal difference is that here the cross-correlation target is not an external sample of discrete galaxies, but other line-intensity mapping data. This distinction is irrelevant to the LCB filter, which recovers the lowest-k modes with $r(k)$ exceeding 80% and generally recovers linear-scale CO(2–1) fluctuations much better than the Wiener filter.

3. Limitations of reconstructions

At this point, we have repeatedly established the ability of the LCB filter to robustly reconstruct the true signal with significant improvements over the Wiener filter. We now demonstrate some limitations of these reconstructions, in the context of examining the relation between CO(2–1) and CO(1–0) emission surveyed by COMAP-ERA across the same volume.

The CO(2–1)/CO(1–0) intensity ratio is a key probe of the environments traced by CO-bright molecular gas, being associated with the temperatures and densities of molecular clouds as well as heating from the interstellar radiation field, and the consequent excitation of the CO rotational transitions [105–110]. A global measurement of this line ratio should be a statistical diagnostic for the dynamics and chemistry of the earliest molecular clouds and galactic dust
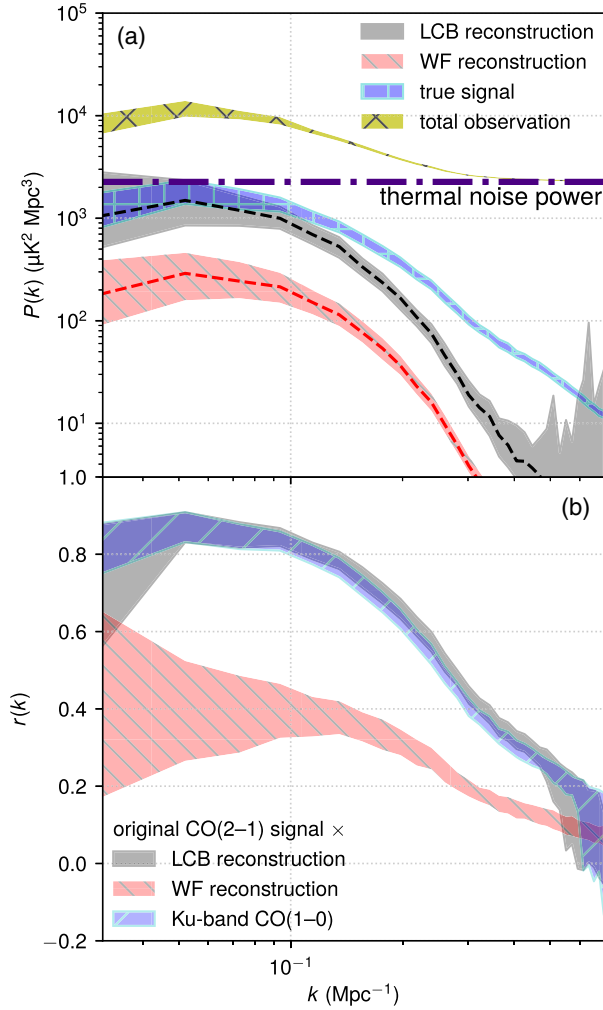
FIG. 9.   90% intervals of reconstruction summary statistics for Ka-band data across our $144 = 48 \times 3$ simulated COMAP-ERA survey fields. We show (a) the obtained power spectra of the LCB and Wiener filter reconstructions alongside their expected values (dashed lines), as well as the original input signal, the total Ka-band observation including noise and interloper emission, and the thermal noise power spectrum by itself (dash-dotted line). We also show (b) the normalized cross-correlation $r(k)$ between the original input signal and the two reconstructions, as well as between the noiseless CO(1–0) pseudosignal and the galaxy density contrast.
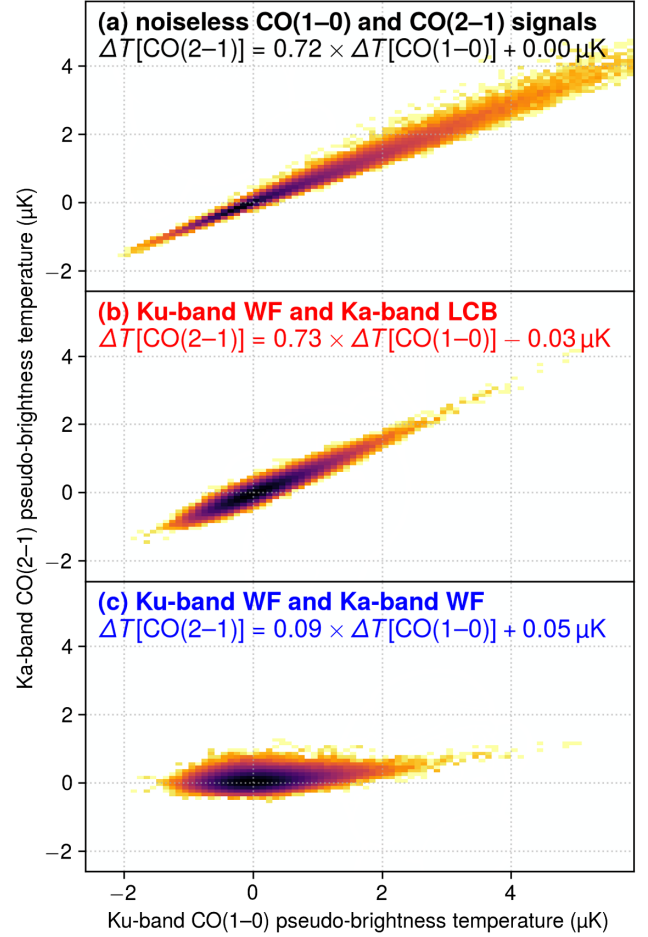

FIG. 10.   Joint probability distributions, shown using a log color scale (darker colors indicate higher probability), of CO(1–0) and CO(2–1) temperature fluctuations at $z \sim 7$. We show distributions for a representative realization of (a) the pair of noiseless pseudosignals, (b) simulated COMAP-ERA reconstructions using the Wiener filter or the LCB filter as appropriate, and (c) simulated COMAP-ERA reconstructions using only Wiener filters for each band. The LCB reconstruction, although failing to help recover the correct shape of the joint probability distribution, allows recovery of the correct slope of the correlation and thus the correct global line ratio in $\langle Tb \rangle$ to within a few percent, modulo bias from observational distortions. Using only Wiener filter reconstructions fails to recover even the correct slope of the correlation line.

in cosmic history and would perhaps improve prospects of recovering quantities like the cosmic molecular gas density in a physically motivated fashion.

Since LIM analyses will typically discard information about the mean intensity in cleaning the data of continuum foregrounds, the best that one can infer is the ratio of CO(2–1) temperature fluctuations to CO(1–0) temperature fluctuations. For fluctuations on linear scales, the result will not be the ratio of the line temperatures, but the ratio of the line temperature-bias products $\langle Tb \rangle_{\rm line}$, which we briefly mentioned in Sec. II C as the conversion between matter density contrast and CO temperature

contrast. That said, in principle, we expect these lines to correlate strongly (see, e.g., [20]) and trace the underlying matter density with similar biases. This expectation is certainly the case for the CO model used in this work. At $z = 6.68$ (corresponding to the midpoints of the COMAP observing frequency bands), the CO(2–1)/CO(1–0) ratio in $\langle Tb \rangle$ [calculated via Eq. (21), assuming the halo bias model of Ref. [111] and the same halo mass function used for peak-patch mass adjustments in Sec. V A 2] is 0.74. The CO(2–1)/CO(1–0) ratio in $\langle T \rangle$, calculated simply by removing $b(M_h)$ from the integrand of Eq. (21), is only around 5% lower at 0.71.

We should be able to find the CO(2–1)/CO(1–0) ratio in $\langle Tb \rangle$ from simulated COMAP-ERA data in two ways. The first is by taking the ratio of the CO(1–0)–CO(2–1) cross power spectrum to the CO(1–0) autopower spectrum on linear scales. Simulated surveys recover a ratio of 0.75 ± 0.03 (68% interval) on average across 48×3 patches, so recovery is successful and unbiased. The second is by finding a linear fit to the reconstructed pseudotemperature fields, namely the LCB-filtered Ka-band CO(2–1) data and the Wiener-filtered Ku-band CO(1–0) data. We show the correlation between the reconstructions in Fig. 10 through their joint probability distribution, alongside the actual correlation of the original, noiseless pseudotemperature fields.

The original pseudosignals have a CO(2–1)/CO(1–0) ratio of 0.72 in $\langle Tb \rangle$ with very little scatter from realization to realization, showing a slight bias relative to the true value resulting from observational effects (in particular the smaller beam FWHM and thus reduced beam dilution in the Ku-band observation compared to the Ka-band observation). But once we accept this bias between the true ratio and the pseudosignal ratio, we note that the reconstructions (which, again, are effectively pseudo-pseudosignals) successfully recover the correct pseudosignal ratio, obtaining

0.73 ± 0.02 (68% interval) across our 48×3 simulated patches. Using only Wiener filter reconstructions naturally fails to recover the same line ratio, due to strong suppression of the Ka-band reconstruction by the overwhelming $z \sim 3$ interloper emission.

But while we recover the correct overall slope of the correlation, the reconstructions evidently do not recover the actual shape of the joint probability distribution. The original pseudosignals show a tight correlation at low pseudotemperatures that broadens for the brightest CO peaks. These trends are sensible given our model, which assumes a log-normal scatter in CO(1–0) or CO(2–1) luminosity for fixed star-formation rate. Because the scatter is log-normal, at low luminosities and thus for small temperature fluctuations, the absolute difference between the CO(1–0) and CO(2–1) signals will be smaller than for high luminosities and large temperature fluctuations. The original joint probability distribution, even with distortion by observational effects, thus contains interesting information about the stochasticity of CO emission and its relation to CO luminosity. No such information is recovered by the reconstructions, which do show skewed distributions but not the correct size or shape of skew. This shortcoming is to
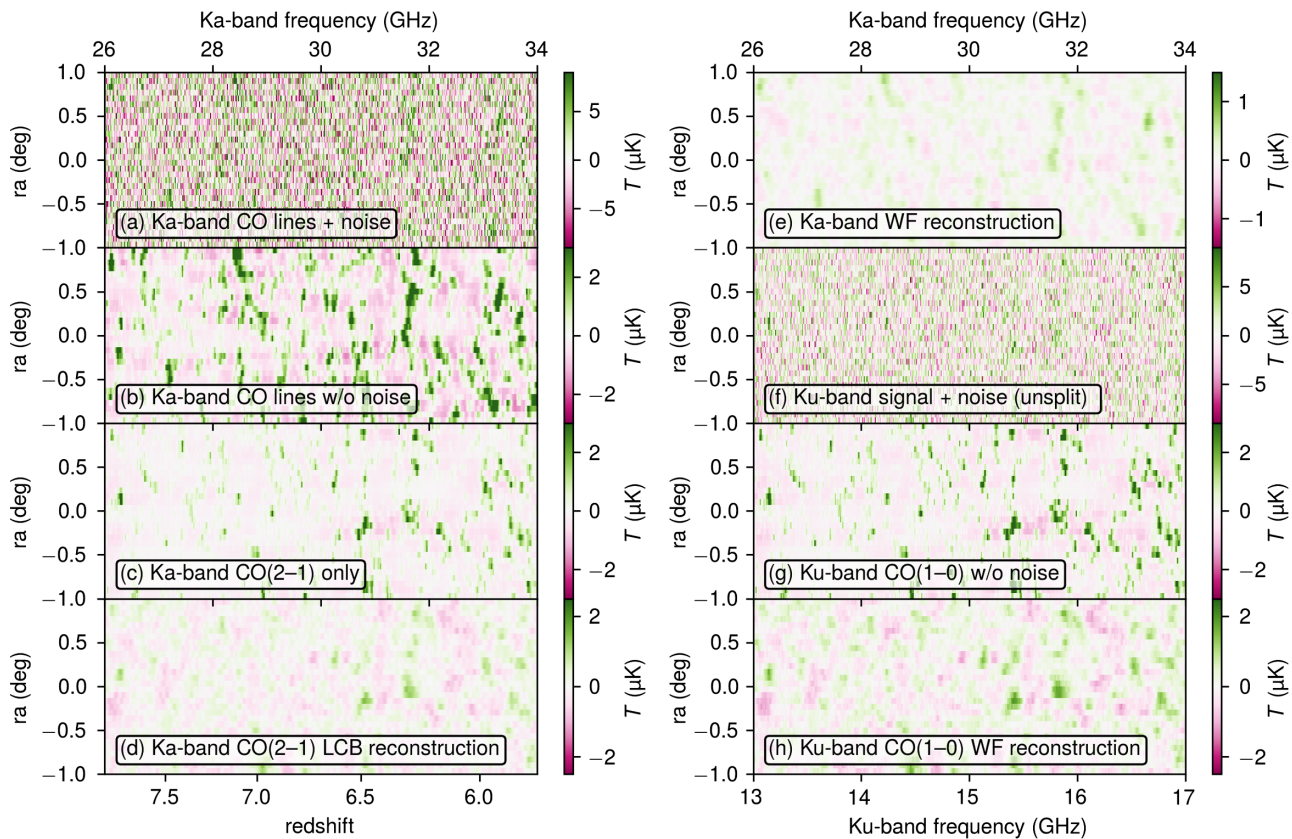

FIG. 11.   Visualization of a slice of an example COMAP-ERA simulation, showing six outputs: (a) the total simulated Ka-band observation, including the CO(1–0) interloper from $z \sim 3$, the CO(2–1) signal from $z \sim 7$, and noise; (b) the sum of the noiseless filtered signal and interloper CO emission; (c) the noiseless $z \sim 7$ CO(2–1) pseudotemperature in isolation; (d) the LCB-filtered Ka-band data, based on cross-correlation with the Ku-band measurement of CO(1–0); (e) the Wiener-filtered Ka-band data; (f) the total simulated Ku-band observation including noise; (g) the noiseless Ku-band CO(1–0) signal; and (h) the Wiener-filtered Ku-band data.

be expected from applying a linear estimator to two highly non-Gaussian fields. In this context, the LCB and Wiener filters recover large-scale fluctuations well, but small-scale statistics less so.

Ultimately, the LCB filter is valuable in translating the cross-correlation between correlated signals into a redshift-space map that robustly rejects disjoint systematics and noise, but will not recover information about global properties beyond the power spectra used to calculate the filter. More sophisticated reconstruction techniques, potentially with sensible informative inbuilt priors, would likely significantly improve recovery of local statistics of and between observables.

In concluding discussion of COMAP-ERA simulation results, we plot slices of various mock observations and reconstructions from a typical realization of a survey field in Fig. 11, serving as a qualitative graphical recap.

VI. CONCLUSIONS

We can now decisively answer the questions we posed in the Introduction:

(i) "Can the LCB filter successfully reconstruct the structure of line-intensity fluctuations on large linear scales by exploiting cross-correlations with surveys of discrete sources?" Yes. We showed this both in the context of a futuristic [C II] LIM survey and in the context of near-term observations feasible with the COMAP Pathfinder. Both normalized cross-correlation and estimator power spectrum bias relative to the true signal improve significantly over the Wiener filter, across a range of different assumptions about the information available to calculate the filter.

(ii) "Can the LCB filter successfully reconstruct the structure of line-intensity fluctuations on large linear scales by cross-correlation with a measurement of a different, correlated line-intensity signal?" Yes. Simulations of the proposed COMAP-ERA survey show strong recovery of $z \sim 7$ CO(2–1) from Ka-band data through cross-correlation with Ku-band data measuring $z \sim 7$ CO(1–0), despite the presence of bright interloper emission in CO(1–0) from $z \sim 3$ in the Ka band. The LCB filter should thus enable COMAP-ERA to robustly map molecular gas in multiple lines across large volumes spanning the late epoch of reionization.

As we have stated multiple times—as early as the Introduction, in fact—the LCB filter is not necessarily the best estimator to use, given the strong non-Gaussianity of a line-intensity field that principally traces the gas content of galaxies. We saw this in the COMAP-ERA linear reconstructions, which failed to recover the correct skew of the joint probability distribution between CO(1–0) and CO(2–1) temperature fluctuations. Furthermore, the shot noise components of different tracers could become a significant source of small-scale reconstruction variance in this formalism. However, at large linear scales where the signal *is* approximately Gaussian, the LCB filter appears successful, suggesting that it or a similarly straightforward technique may be sufficient for some use cases. For instance, we hope future work will investigate identification of large ($\gtrsim 10$ Mpc diameter) voids and peaks in the context of COMAP-ERA reconstructions. (A detailed consideration is beyond the scope of this work but a preliminary consideration of the [C II] simulations from Sec. IV, as outlined in Appendix D, suggests reason for optimism.) Factoring in redshift-space distortions and anisotropic information, which we have neglected here, would also likely improve reconstruction.

We also hope that the possibility of robust signal reconstruction through cross-correlation further motivates coordination between future LIM surveys *and* between LIM and other LSS surveys. Such collaboration will enable looking beyond statistical constraints on the early cosmic history of star formation and galaxy assembly, toward directly mapping these processes across large cosmic volumes in many different shades.

APPENDIX A: PROOF OF EQUIVALENCE OF LCB FILTERING WITH THE ESTIMATOR OF MANZOTTI AND DODELSON [53]

We show here explicitly that the LCB filter is equivalent to the optimal map estimator of Manzotti and Dodelson [53]. Although this was shown explicitly for $n = 1$ by Weaverdyck *et al.* [56], we show it for all n, with the only assumption being that there is zero covariance between different wave numbers k or between spherical harmonic multipoles ℓ in the use case of Manzotti and Dodelson [53].

First, we note that the matrix D is defined in much the same way as we define it in this work, e.g., in Eq. (4). However, they allocate the observable of interest to index 1; we will adapt any of their expressions to our convention of allocating index $p = n + 1$. There is an additional covariance for which Manzotti and Dodelson [53] use the notation C, but to denote the covariance of $d_p - s_p$, i.e., the sum of the covariances of all components of the observable p that we are *not* interested in reconstructing (noise, other uncorrelated interlopers, and so on). So their C is really our $P^{(N)}$.

Manzotti and Dodelson [53] also define a further additional variance N, and we reproduce the definition of their Eq. (8) here (albeit with necessary adaptations as described above),

$$N^{-1} = (P^{(N)})^{-1} + (D^{-1})_{pp}. \tag{A1}$$

Given this N, we now reproduce the estimator of Manzotti and Dodelson [53] as defined by their Eq. (9) (with some adaptations and omitting the dependence on k for brevity),

$$\hat{s}_p = N \left[(P^{(N)})^{-1} d_p - \sum_{j=1}^{n} (D^{-1})_{pj} d_j \right]. \tag{A2}$$

First, recall that our formalism involves a Cholesky decomposition of D such that $D = LL^T$, with L being a lower triangular matrix such that $L_{ij} = 0$ if $i < j$. Then since L^{-1} is also a lower triangular matrix (and its transpose L^{-T} is upper triangular),

$$1 = (L^{-1}L)_{pp} = \sum_{i=1}^{p} (L^{-1})_{pi} L_{ip} = (L^{-1})_{pp} L_{pp}, \tag{A3}$$

and

$$(D^{-1})_{pp} = \sum_{i=1}^{p} (L^{-T})_{pi} (L^{-1})_{ip}$$
$$= \sum_{i=1}^{p} (L^{-1})_{ip}^2 = (L^{-1})_{pp}^2 = (L_{pp})^{-2}. \tag{A4}$$

More generally,

$$(D^{-1})_{pj} = \sum_{i=1}^{p} (L^{-T})_{pi} (L^{-1})_{ij} = (L^{-1})_{pp} (L^{-1})_{pj}$$
$$= (L_{pp})^{-1} (L^{-1})_{pj}. \tag{A5}$$

Additionally, since $LL^{-1} = I$ (the identity matrix),

$$\delta_{pj} = (LL^{-1})_{pj} = \sum_{i=1}^{p} L_{pi} (L^{-1})_{ij}$$
$$= L_{pp} (L^{-1})_{pj} + \sum_{i=1}^{n} L_{pi} (L^{-1})_{ij}. \tag{A6}$$

Then for all $j < p$, i.e., for all j from 1 to n, putting together Eqs. (A5) and (A6) gives us

$$(D^{-1})_{pj} = -\frac{1}{L_{pp}^2} \sum_{i=1}^{n} L_{pi} (L^{-1})_{ij}. \tag{A7}$$

Substituting this result, Eqs. (A1) and (A4) all at once into Eq. (A2), we find

$$\hat{s}_p = \frac{L_{pp}^2 P^{(N)}}{L_{pp}^2 + P^{(N)}} \left(\frac{d_p}{P^{(N)}} + \frac{1}{L_{pp}^2} \sum_{j=1}^{n} \sum_{i=1}^{n} L_{pi} (L^{-1})_{ij} d_j \right)$$
$$= \frac{L_{pp}^2}{L_{pp}^2 + P^{(N)}} d_p + \frac{P^{(N)}}{L_{pp}^2 + P^{(N)}} \sum_{j=1}^{n} \sum_{i=1}^{n} L_{pi} (L^{-1})_{ij} d_j. \tag{A8}$$

One may readily see that this is entirely equivalent to Eq. (10), which is what was to be shown.

APPENDIX B: FURTHER EXPRESSIONS FOR USE WITH THE LCB FILTER WHEN $n > 1$

It is straightforward to show that, in general, each element of the ith row of the Cholesky decomposition of D may be expressed recursively in terms of the $(i - 1) \times (i - 1)$ submatrix preceding it as well as the preceding elements of that row,

$$L_{ii} = \sqrt{D_{ii} - \sum_{j=1}^{i-1} L_{ij}^2}, \tag{B1}$$

$$L_{ij} = \frac{1}{L_{jj}} \left(D_{ij} - \sum_{k=1}^{j-1} L_{ik} L_{jk} \right), \tag{B2}$$

where the latter is for $j < i$ only.

Specifically, in the case of $n = 2$, where D is a 3×3 matrix, we obtain the following elements of L:

$$L_{11} = \sqrt{D_{11}}, \tag{B3}$$

$$L_{21} = D_{21}/\sqrt{D_{11}}, \qquad (B4)$$

$$L_{22} = \sqrt{D_{22} - D_{21}^2/D_{11}}, \qquad (B5)$$

$$L_{31} = D_{31}/\sqrt{D_{11}}, \qquad (B6)$$

$$L_{32} = (D_{32} - D_{21}D_{31}/D_{11})/L_{22}, \qquad (B7)$$

$$L_{33} = \sqrt{D_{33} - \frac{D_{31}^2}{D_{11}} - L_{32}^2}. \qquad (B8)$$

Note that we do not expand L_{22} and L_{32} in the expressions for L_{32} and L_{33} respectively, for brevity. It should be clear that the first three of these expressions are the same as for a 2×2 matrix and evaluate to Eqs. (11)–(13) given D as defined in Eq. (4).

In particular, note that since

$$L_{pp}^2 = D_{pp} - \sum_{j=1}^{n} L_{pj}^2 = P^{(S)} - \sum_{j=1}^{n} L_{pj}^2, \qquad (B9)$$

it is always the case that $L_{pp}^2 + P^{(N)}$ can be written purely in terms of the total power spectrum $P^{(S)} + P^{(N)}$ of d_p and the upper left $n \times n$ submatrix of L, which in turn only depends on the total power spectra of the other observables.

APPENDIX C: EQUIVALENCE OF THE LCB FILTER FOR $s_i = s_j$ FOR ALL i AND j WITH A WIENER FILTER FOR THE INVERSE NOISE VARIANCE-WEIGHTED AVERAGE OF ALL DATA

Suppose now that all our observations are of the same underlying signal, i.e., $s_i = s_j$ for all i and j, and call this signal $s \equiv s_i$ for all i. The observations still have independent, uncorrelated noise, possibly even different levels of noise. However, importantly, the expected values of the cross power spectra are the same as the expected value of the signal autopower spectrum, unbiased by noise. So the relevant covariance matrix is given by

$$D_{ij}(k) = \begin{cases} P^{(S)}(k) + \delta_{ij}P_i^{(N)} & \text{if } i \leq n \text{ and } j \leq n, \\ P^{(S)}(k) & \text{otherwise.} \end{cases} \qquad (C1)$$

For intuition, consider again the special case of $n = 1$. Then Eq. (16) becomes

$$\frac{L_{22}^2}{L_{22}^2 + P_2^{(N)}} = \frac{P^{(S)}P_1^{(N)}}{(P^{(S)} + P_1^{(N)})(P^{(S)} + P_2^{(N)}) - P^{(S)2}}, \qquad (C2)$$

and we can rewrite Eq. (15) as

$$\hat{s} = \frac{P^{(S)}(P_1^{(N)}d_2 + P_2^{(N)}d_1)}{P^{(S)}(P_1^{(N)} + P_2^{(N)}) + P_1^{(N)}P_2^{(N)}}$$

$$= \frac{P^{(S)}d_*}{P^{(S)} + P^{(N*)}}, \qquad (C3)$$

where d_* denotes the noise inverse variance-weighted average

$$d_* = \frac{d_1/P_1^{(N)} + d_2/P_2^{(N)}}{1/P_1^{(N)} + 1/P_2^{(N)}}, \qquad (C4)$$

whose power spectrum will be the sum of the signal power spectrum, which will still be $P^{(S)}$, and a noise power spectrum that one may readily find equal to

$$P^{(N*)} \equiv \frac{1}{1/P_1^{(N)} + 1/P_2^{(N)}}. \qquad (C5)$$

Written in this way, we can easily see that the LCB-filtered estimator for the signal from the covariance defined in Eq. (C1) is equivalent to applying a Wiener filter to the inverse variance-weighted d_* in the case of $n = 1$.

Now consider the more general case. We will consider the estimator in the form given by Manzotti and Dodelson [53], which we handily proved equivalent to the LCB filter in Appendix A. It is straightforward to find the inverse of the covariance matrix D, given its simple form. Namely, the upper left $n \times n$ submatrix of D^{-1} is given by

$$(D^{-1})_{ij} = \delta_{ij}/P_i^{(N)}, \qquad (C6)$$

for $i \leq n$ and $j \leq n$. Meanwhile, for $i < p$, the final row and final column are both given by

$$(D^{-1})_{jp} = (D^{-1})_{pj} = -1/P_j^{(N)}, \qquad (C7)$$

and finally,

$$(D^{-1})_{pp} = 1/P^{(S)} + \sum_{j=1}^{n} 1/P_j^{(N)}. \qquad (C8)$$

Note that the upper left $n \times n$ submatrix is completely irrelevant to the estimator, but we have provided it here for completeness.

In any case, the variance N associated with $\hat{s}_p$ as given by Eq. (A1) is

$$N^{-1} = 1/P^{(N)} + (D^{-1})_{pp} = 1/P^{(S)} + \sum_{j=1}^{p} 1/P_j^{(N)}, \qquad (C9)$$

such that the estimator as given by Eq. (A2) evaluates to

$$\hat{s}_p = \frac{\sum_{j=1}^{p} d_j/P_j^{(N)}}{1/P^{(S)} + \sum_{i=j}^{p} 1/P_j^{(N)}}. \qquad (C10)$$

But in this general case,

$$d_* = \frac{\sum_{j=1}^{p} d_j/P_j^{(N)}}{\sum_{i=j}^{p} 1/P_j^{(N)}}, \qquad (C11)$$

and

$$1/P^{(N*)} = \sum_{i=j}^{p} 1/P_j^{(N)}, \qquad (C12)$$

so that the estimator is

$$\hat{s}_p = \frac{d_*/P^{(N*)}}{1/P^{(S)} + 1/P^{(N*)}} = \frac{P^{(S)} d_*}{P^{(S)} + P^{(N*)}}, \qquad (C13)$$

which is equivalent to applying a Wiener filter to the inverse variance-weighted d_*, just as we sought to show for all n. This is hardly surprising as both the Wiener filter and the LCB filter are optimal map estimators, and it should not be possible to obtain extra information from the same data simply by cross-correlating splits.

APPENDIX D: A PRELIMINARY CONSIDERATION OF IMPROVEMENTS IN PEAK IDENTIFICATION WITH LCB RECONSTRUCTIONS

Due to the varied ways in which one may identify and qualify peaks and voids in the cosmic web, a detailed consideration of improvements in peak or void identification is well beyond the scope of the present work. However, with an extremely basic peak identification workflow, we may demonstrate probable cause to expect a significant improvement in the ability of a survey to identify line-intensity peaks through a LCB reconstruction when compared to the Wiener-filtered reconstruction.

For this, we return to the ten SMDPL/UniverseMachine light cones and their corresponding [C II] simulations used for Sec. IV. Immediately, based on the size of the simulation, we note that the survey only measures a limited number of modes at scales of ~100 Mpc, so we discard modes with $k < 0.05$ Mpc^{-1} from both the LCB and Wiener filter reconstructions. We also blur the data cube with a 3D Gaussian filter with its profile defined by a comoving rms width of 4.2 Mpc in all directions. This corresponds to a FWHM of 10 Mpc, reflecting our qualitative conclusion that the fidelity of the

TABLE II. 68% interval and minima/maxima across ten realizations of the number of peaks identified (out of the brightest 200) in the [C II] LIM signal that has a partner peak in either the reconstruction of the signal (through either the Wiener filter or the LCB filter) or the overlapping LAE density contrast cube (again identifying partner peaks from the brightest or most overdense 200 peaks in the mock observation used).

	Partner peaks	
Data cube searched	68% interval	Extrema
WF reconstruction	(34, 51)	(33, 65)
LAE cube	(52, 61)	(43, 64)
LCB reconstruction	(61, 69)	(52, 76)

LCB reconstruction is best suited to identification of $\gtrsim 10$ Mpc diameter features.

From this preprocessed data cube we identify the brightest voxel as the center of a peak. Assuming the peak must have a spherical profile, we count the total flux contained in spherical shells expanding away from the peak center and consider the extent of the peak to end at the first shell where the total contained flux ceases to increase. We then search for the next-brightest peak excluding all voxels within this extent of the previous peak, repeating until we have identified the 200 brightest peaks in the [C II] reconstruction. For comparison, we also identify the 200 most overdense peaks in LAE density contrast in an analogous fashion.

We then compare against the 200 brightest peaks identified analogously in the true [C II] signal and see how many peaks have a partner peak identified in the reconstructions (or in the LAE density contrast cube), as defined by being within ten voxels of the true peak. We tabulate the results in Table II. The Wiener-filtered data clearly skew toward identifying fewer partnered peaks, while both the LAE data and the LCB reconstruction skew toward identifying more partnered peaks. By the metric, the LCB reconstruction somewhat outperforms the LAE cube and significantly outperforms the WF reconstruction.

These results will vary significantly depending on the number of peaks we choose to identify in each cube, the definition of the peak extent and selection criterion, the smoothing scale for preprocessing the data cubes, and so on. Therefore, what we demonstrate here should hardly be seen as the limit of what these data and reconstruction methods can do in recovering cosmic structure. Nonetheless, the overall outcome demonstrates on a preliminary level that the LCB reconstruction may truly provide an improvement in identifying cosmic web features compared to either the LIM dataset alone or its cross-correlation target alone (the LAE catalog in this case).

[1] E. D. Kovetz *et al.*, arXiv:1709.09066.
[2] E. Kovetz, P. C. Breysse, A. Lidz, J. Bock, C. M. Bradford, T.-C. Chang, S. Foreman, H. Padmanabhan, A. Pullen, D. Riechers, M. B. Silva, and E. Switzer, Bull. Am. Astron. Soc. **51**, 101 (2019), https://baas.aas.org/pub/2020n3i101/release/1.
[3] J. L. Bernal and E. D. Kovetz, Astron. Astrophys. Rev. **30**, 5 (2022).
[4] P. Madau, A. Meiksin, and M. J. Rees, Astrophys. J. **475**, 429 (1997).
[5] T.-C. Chang, U.-L. Pen, J. B. Peterson, and P. McDonald, Phys. Rev. Lett. **100**, 091303 (2008).
[6] A. Lidz, S. R. Furlanetto, S. P. Oh, J. Aguirre, T.-C. Chang, O. Doré, and J. R. Pritchard, Astrophys. J. **741**, 70 (2011).
[7] Y. Gong, A. Cooray, M. Silva, M. G. Santos, J. Bock, C. M. Bradford, and M. Zemcov, Astrophys. J. **745**, 49 (2012).
[8] A. T. Crites, J. J. Bock, C. M. Bradford, T. C. Chang, A. R. Cooray, L. Duband, Y. Gong, S. Hailey-Dunsheath, J. Hunacek, P. M. Koch, C. T. Li, R. C. O'Brient, T. Prouve, E. Shirokoff, M. B. Silva, Z. Staniszewski, B. Uzgil, and M. Zemcov, in *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII*, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 9153, edited by W. S. Holland and J. Zmuidzinas (SPIE, Bellingham, 2014), p. 91531W.
[9] M. Silva, M. G. Santos, A. Cooray, and Y. Gong, Astrophys. J. **806**, 209 (2015).
[10] B. Yue, A. Ferrara, A. Pallottini, S. Gallerani, and L. Vallini, Mon. Not. R. Astron. Soc. **450**, 3829 (2015).
[11] T. Y. Li, R. H. Wechsler, K. Devaraj, and S. E. Church, Astrophys. J. **817**, 169 (2016).
[12] G. K. Keating, D. P. Marrone, G. C. Bower, E. Leitch, J. E. Carlstrom, and D. R. DeBoer, Astrophys. J. **830**, 34 (2016).
[13] H. Padmanabhan, Mon. Not. R. Astron. Soc. **475**, 1477 (2018).
[14] H. Padmanabhan, Mon. Not. R. Astron. Soc. **488**, 3014 (2019).
[15] S. Dumitru, G. Kulkarni, G. Lagache, and M. G. Haehnelt, Mon. Not. R. Astron. Soc. **485**, 3486 (2019).
[16] G. K. Keating, D. P. Marrone, G. C. Bower, and R. P. Keenan, Astrophys. J. **901**, 141 (2020).
[17] M. Aravena *et al.* (CCAT-Prime Collaboration), Astrophys. J. Suppl. Ser. **264**, 7 (2023).
[18] E. Schaan and M. White, J. Cosmol. Astropart. Phys. 05 (2021) 067.
[19] E. Schaan and M. White, J. Cosmol. Astropart. Phys. 05 (2021) 068.
[20] S. Yang, G. Popping, R. S. Somerville, A. R. Pullen, P. C. Breysse, and A. S. Maniyar, Astrophys. J. **929**, 140 (2022).
[21] K. A. Cleary *et al.* (COMAP Collaboration). Astrophys. J. **933**, 182 (2022).
[22] H. Padmanabhan, P. Breysse, A. Lidz, and E. R. Switzer, Mon. Not. R. Astron. Soc. **515**, 5813 (2022).
[23] M. Bethermin *et al.*, Astron. Astrophys. **667**, A156 (2022).
[24] A. Fasano *et al.*, Proc. SPIE Int. Soc. Opt. Eng. **12190**, 375 (2022).
[25] P. C. Breysse, E. D. Kovetz, P. S. Behroozi, L. Dai, and M. Kamionkowski, Mon. Not. R. Astron. Soc. **467**, 2996 (2017).

[26] P. C. Breysse, C. J. Anderson, and P. Berger, Phys. Rev. Lett. **123**, 231105 (2019).
[27] H. T. Ihle *et al.* (COMAP Collaboration). Astrophys. J. **871**, 75 (2019).
[28] S. Majumdar, J. R. Pritchard, R. Mondal, C. A. Watkinson, S. Bharadwaj, and G. Mellema, Mon. Not. R. Astron. Soc. **476**, 4007 (2018).
[29] A. Beane and A. Lidz, Astrophys. J. **867**, 26 (2018).
[30] P. La Plante, A. Lidz, J. Aguirre, and S. Kohn, Astrophys. J. **899**, 40 (2020).
[31] P. C. Breysse, E. D. Kovetz, and M. Kamionkowski, Mon. Not. R. Astron. Soc. **452**, 3408 (2015).
[32] A. Lidz and J. Taylor, Astrophys. J. **825**, 143 (2016).
[33] Y.-T. Cheng, T.-C. Chang, J. Bock, C. M. Bradford, and A. Cooray, Astrophys. J. **832**, 165 (2016).
[34] G. Sun, L. Moncelsi, M. P. Viero, M. B. Silva, J. Bock, C. M. Bradford, T. C. Chang, Y. T. Cheng, A. R. Cooray, A. Crites, S. Hailey-Dunsheath, B. Uzgil, J. R. Hunacek, and M. Zemcov, Astrophys. J. **856**, 107 (2018).
[35] Y.-T. Cheng, T.-C. Chang, and J. J. Bock, Astrophys. J. **901**, 142 (2020).
[36] K. Moriwaki, N. Filippova, M. Shirasaki, and N. Yoshida, Mon. Not. R. Astron. Soc. **496**, L54 (2020).
[37] K. Moriwaki, M. Shirasaki, and N. Yoshida, Astrophys. J. Lett. **906**, L1 (2021).
[38] P. C. Breysse *et al.* (COMAP Collaboration). Astrophys. J. **933**, 188 (2022).
[39] D. N. Pfeffer, P. C. Breysse, and G. Stein, arXiv:1905.10376.
[40] E. Visbal and A. Loeb, J. Cosmol. Astropart. Phys. 11 (2010) 016.
[41] A. R. Pullen, T.-C. Chang, O. Doré, and A. Lidz, Astrophys. J. **768**, 15 (2013).
[42] P. Serra, O. Doré, and G. Lagache, Astrophys. J. **833**, 153 (2016).
[43] A. R. Pullen, P. Serra, T.-C. Chang, O. Doré, and S. Ho, Mon. Not. R. Astron. Soc. **478**, 1911 (2018).
[44] A. Beane, F. Villaescusa-Navarro, and A. Lidz, Astrophys. J. **874**, 133 (2019).
[45] G. Sun, B. S. Hensley, T.-C. Chang, O. Doré, and P. Serra, Astrophys. J. **887**, 142 (2019).
[46] S. Yang, A. R. Pullen, and E. R. Switzer, Mon. Not. R. Astron. Soc. **489**, L53 (2019).
[47] C. J. Anderson, E. R. Switzer, and P. C. Breysse, Mon. Not. R. Astron. Soc. **514**, 1169 (2022).
[48] R. P. Keenan, G. K. Keating, and D. P. Marrone, Astrophys. J. **927**, 161 (2022).
[49] T.-C. Chang, U.-L. Pen, K. Bandura, and J. B. Peterson, Nature (London) **466**, 463 (2010).
[50] K. W. Masui, E. R. Switzer, N. Banavar, K. Bandura, C. Blake, L. M. Calin, T. C. Chang, X. Chen, Y. C. Li, Y. W. Liao, A. Natarajan, U. L. Pen, J. B. Peterson, J. R. Shaw, and T. C. Voytek, Astrophys. J. Lett. **763**, L20 (2013).
[51] E. R. Switzer, K. W. Masui, K. Bandura, L. M. Calin, T. C. Chang, X. L. Chen, Y. C. Li, Y. W. Liao, A. Natarajan, U. L. Pen, J. B. Peterson, J. R. Shaw, and T. C. Voytek, Mon. Not. R. Astron. Soc. **434**, L46 (2013).
[52] R. B. Barreiro, P. Vielva, C. Hernandez-Monteagudo, and E. Martinez-Gonzalez, IEEE J. Sel. Top. Signal Process. **2**, 747 (2008).

[53] A. Manzotti and S. Dodelson, Phys. Rev. D **90,** 123009 (2014).

[54] P. A. R. Ade *et al.* (Planck Collaboration). Astron. Astrophys. **594,** A21 (2016).

[55] L. Bonavera, R. B. Barreiro, A. Marcos-Caballero, and P. Vielva, Mon. Not. R. Astron. Soc. **459,** 657 (2016).

[56] N. Weaverdyck, J. Muir, and D. Huterer, Phys. Rev. D **97,** 043515 (2018).

[57] F. S. Kitaura and T. A. Enßlin, Mon. Not. R. Astron. Soc. **389,** 497 (2008).

[58] U.-L. Pen, R. Sheth, J. Harnois-Deraps, X. Chen, and Z. Li, arXiv:1202.5804.

[59] N. Oppermann, M. Selig, M. R. Bell, and T. A. Enßlin, Phys. Rev. E **87,** 032136 (2013).

[60] H.-M. Zhu, U.-L. Pen, Y. Yu, X. Er, and X. Chen, Phys. Rev. D **93,** 103504 (2016).

[61] C. Modi, M. White, A. Slosar, and E. Castorina, J. Cosmol. Astropart. Phys. 11 (2019) 023.

[62] D. T. Chung, Mon. Not. R. Astron. Soc. **513,** 4090 (2022).

[63] P. Coles and B. Jones, Mon. Not. R. Astron. Soc. **248,** 1 (1991).

[64] A. Klypin, G. Yepes, S. Gottlöber, F. Prada, and S. Heß, Mon. Not. R. Astron. Soc. **457,** 4340 (2016).

[65] P. A. R. Ade *et al.* (Planck Collaboration). Astron. Astrophys. **594,** A13 (2016).

[66] G. J. Stacey, N. Geis, R. Genzel, J. B. Lugten, A. Poglitsch, A. Sternberg, and C. H. Townes, Astrophys. J. **373,** 423 (1991).

[67] I. De Looze *et al.*, Astron. Astrophys. **568,** A62 (2014).

[68] R. Herrera-Camus *et al.*, Astrophys. J. **800,** 1 (2015).

[69] D. Schaerer *et al.*, Astron. Astrophys. **643,** A3 (2020).

[70] K. V. Croxall, J. D. Smith, E. Pellegrini, B. Groves, A. Bolatto, R. Herrera-Camus, K. M. Sandstrom, B. Draine, M. G. Wolfire, L. Armus, M. Boquien, B. Brandl, D. Dale, M. Galametz, L. Hunt, J. Kennicutt, R., K. Kreckel, D. Rigopoulou, P. van der Werf, and C. Wilson, Astrophys. J. **845,** 96 (2017).

[71] S. Carniani, R. Maiolino, R. Smit, and R. Amorín, Astrophys. J. Lett. **854,** L7 (2018).

[72] S. Fujimoto, M. Ouchi, A. Ferrara, A. Pallottini, R. J. Ivison, C. Behrens, S. Gallerani, S. Arata, H. Yajima, and K. Nagamine, Astrophys. J. **887,** 107 (2019).

[73] J. Matthee, D. Sobral, L. A. Boogaard, H. Röttgering, L. Vallini, A. Ferrara, A. Paulino-Afonso, F. Boone, D. Schaerer, and B. Mobasher, Astrophys. J. **881,** 124 (2019).

[74] M. Rybak, G. Calistro Rivera, J. A. Hodge, I. Smail, F. Walter, P. van der Werf, E. da Cunha, C.-C. Chen, H. Dannerbauer, R. J. Ivison, A. Karim, J. M. Simpson, A. M. Swinbank, and J. L. Wardlow, Astrophys. J. **876,** 112 (2019).

[75] R. A. Meyer, F. Walter, C. Cicone, P. Cox, R. Decarli, R. Neri, M. Novak, A. Pensabene, D. Riechers, and A. Weiss, Astrophys. J. **927,** 152 (2022).

[76] K. E. Heintz, D. Watson, P. A. Oesch, D. Narayanan, and S. C. Madden, Astrophys. J. **922,** 147 (2021).

[77] K. E. Heintz, P. A. Oesch, M. Aravena, R. J. Bouwens, P. Dayal, A. Ferrara, Y. Fudamoto, L. Graziani, H. Inami, L. Sommovigo, R. Smit, M. Stefanon, M. Topping, A. Pallottini, and P. van der Werf, Astrophys. J. Lett. **934,** L27 (2022).

[78] E. R. Switzer, C. J. Anderson, A. R. Pullen, and S. Yang, Astrophys. J. **872,** 82 (2019).

[79] Q. Jiao, Y. Zhao, M. Zhu, N. Lu, Y. Gao, and Z.-Y. Zhang, Astrophys. J. Lett. **840,** L18 (2017).

[80] F. Valentino, G. E. Magdis, E. Daddi, D. Liu, M. Aravena, F. Bournaud, A. Cibinel, D. Cormier, M. E. Dickinson, Y. Gao, S. Jin, S. Juneau, J. Kartaltepe, M.-Y. Lee, S. C. Madden, A. Puglisi, D. Sanders, and J. Silverman, Astrophys. J. **869,** 27 (2018).

[81] N. P. H. Nesvadba, R. Cañameras, R. Kneissl, S. Koenig, C. Yang, E. Le Floc'h, A. Omont, and D. Scott, Astron. Astrophys. **624,** A23 (2019).

[82] E. C. Herenz, L. Wisotzki, R. Saust, J. Kerutt, T. Urrutia, C. Diener, K. B. Schmidt, R. A. Marino, G. de la Vieuville, L. Boogaard, J. Schaye, B. Guiderdoni, J. Richard, and R. Bacon, Astron. Astrophys. **621,** A107 (2019).

[83] M. Ouchi *et al.*, Publ. Astron. Soc. Jpn. **70,** S13 (2018).

[84] K. S. Karkare, A. Moradinezhad Dizgah, G. K. Keating, P. Breysse, and D. T. Chung, arXiv:2203.07258.

[85] P. Behroozi, R. H. Wechsler, A. P. Hearin, and C. Conroy, Mon. Not. R. Astron. Soc. **488,** 3143 (2019).

[86] G. Sun *et al.*, Astrophys. J. **915,** 33 (2021).

[87] G. Sun and S. R. Furlanetto, Mon. Not. R. Astron. Soc. **460,** 417 (2016).

[88] J. Kamenetzky, N. Rangwala, J. Glenn, P. R. Maloney, and A. Conley, Astrophys. J. **829,** 93 (2016).

[89] A. Konno *et al.*, Publ. Astron. Soc. Jpn. **70,** S16 (2018).

[90] C. L. Carilli and F. Walter, Annu. Rev. Astron. Astrophys. **51,** 105 (2013).

[91] G. Stein, M. A. Alvarez, and J. R. Bond, Mon. Not. R. Astron. Soc. **483,** 2236 (2019).

[92] P. S. Behroozi, R. H. Wechsler, and C. Conroy, Astrophys. J. **770,** 57 (2013).

[93] P. S. Behroozi, R. H. Wechsler, and C. Conroy, Astrophys. J. Lett. **762,** L31 (2013).

[94] J. Tinker, A. V. Kravtsov, A. Klypin, K. Abazajian, M. Warren, G. Yepes, S. Gottlöber, and D. E. Holz, Astrophys. J. **688,** 709 (2008).

[95] D. T. Chung *et al.* (COMAP Collaboration). Astrophys. J. **923,** 188 (2021).

[96] D. T. Chung *et al.* (COMAP Collaboration). Astrophys. J. **933,** 186 (2022).

[97] H. T. Ihle *et al.* (COMAP Collaboration). Astrophys. J. **933,** 185 (2022).

[98] T. J. Rennie *et al.* (COMAP Collaboration). Astrophys. J. **933,** 187 (2022).

[99] G. J. Hill, K. Gebhardt, E. Komatsu, N. Drory, P. J. MacQueen, J. Adams, G. A. Blanc, R. Koehler, M. Rafal, M. M. Roth, A. Kelz, C. Gronwall, R. Ciardullo, and D. P. Schneider, ASP Conf. Ser. **399,** 115 (2008), http://aspbooks.org/custom/publications/paper/399-0115.html.

[100] G. J. Hill *et al.*, Astron. J. **162,** 298 (2021).

[101] K. Gebhardt *et al.*, Astrophys. J. **923,** 217 (2021).

[102] D. J. Schlegel *et al.*, arXiv:2209.04322.

[103] D. J. Schlegel *et al.*, arXiv:2209.03585.

[104] Y. Wang, Z. Zhai, A. Alavi, E. Massara, A. Pisani, A. Benson, C. M. Hirata, L. Samushia, D. H. Weinberg, J. Colbert, O. Doré, T. Eifler, C. Heinrich, S. Ho, E. Krause, N. Padmanabhan, D. Spergel, and H. I. Teplitz, Astrophys. J. **928,** 1 (2022).

[105] T. Hasegawa, IAU Symp. **170**, 39 (1997), https://ui.adsabs.harvard.edu/abs/1997IAUS..170...39H/abstract.

[106] T. Hasegawa, J. Morino, K. Sorai, T. Handa, K. Sato, T. Oka, M. Seta, S. Sakamoto, M. Hayashi, R. Booth, L. K. Nyman, L. Bronfman, M. Rubio, and P. Shaver, in *Diffuse Infrared Radiation and the IRTS*, Astronomical Society of the Pacific Conference Series Vol. 124, edited by H. Okuda, T. Matsumoto, and T. Rollig (Astronomical Society of the Pacific, San Francisco, 1997), p. 244.

[107] K. Sorai, T. Hasegawa, R. S. Booth, M. Rubio, J. I. Morino, L. Bronfman, T. Handa, M. Hayashi, L. Å. Nyman, T. Oka, S. Sakamoto, M. Seta, and K. S. Usuda, Astrophys. J. **551**, 794 (2001).

[108] J. Koda, N. Scoville, T. Hasegawa, D. Calzetti, J. Donovan Meyer, F. Egusa, R. Kennicutt, N. Kuno, M. Louie, R. Momose, T. Sawada, K. Sorai, and M. Umei, Astrophys. J. **761**, 41 (2012).

[109] J. Koda *et al.*, Astrophys. J. Lett. **890**, L10 (2020).

[110] J. S. den Brok *et al.*, Mon. Not. R. Astron. Soc. **504**, 3221 (2021).

[111] J. L. Tinker, B. E. Robertson, A. V. Kravtsov, A. Klypin, M. S. Warren, G. Yepes, and S. Gottlöber, Astrophys. J. **724**, 878 (2010).