Intensity Mapping in the Presence of Foregrounds and Correlated Continuum Emission

E. R. Switzer\textsuperscript{1}, C. J. Anderson\textsuperscript{1}, A. R. Pullen\textsuperscript{2}, and S. Yang\textsuperscript{2}

\textsuperscript{1} NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA; eric.r.switzer@nasa.gov
\textsuperscript{2} Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, New York, NY 10003, USA

Received 2018 April 30; revised 2018 December 15; accepted 2018 December 17; published 2019 February 12

Abstract

Intensity mapping has attracted significant interest as an approach to measuring the properties of the interstellar medium in typical galaxies at high redshift. Intensity mapping measures the statistics of surface brightness as a function of frequency, making it sensitive to not only all line emission of interest but also radiation from all other sources. Significant effort has gone into developing approaches that reject foreground contamination. Additionally, the target galaxies have multiple sources of emission that can complicate the interpretation of the line brightness. We describe the problem of jointly estimating correlated continuum emission and cleaning uncorrelated continuum emission, such as from the Milky Way. We apply these considerations to a cross-correlation of Planck data with BOSS quasars for a determination of [C II] for $2 < z < 3.2$. Intensity mapping surveys with few bands have unique challenges for treating foregrounds and avoiding bias from correlated continuum emission. We show how a future intensity mapping survey with many bands can separate line from continuum emission in cross-correlation.

Key words: cosmic background radiation – galaxies: evolution – large-scale structure of universe

1. Introduction

A major goal of modern astrophysics is to understand the evolution of galaxies in their cosmological context. Two epochal shifts of great interest are the formation of the galaxies that reionize the universe and the subsequent dramatic decline in the star formation rate (SFR) from $z \sim 2$ to the present (Madau & Dickinson 2014) despite the continued growth of dark matter halos. Studies of CO, [C II], and 21 cm line emission are particularly valuable as tracers of star formation and its precursors in the interstellar medium during these eras (Carilli & Walter 2013).

Most studies of line emission to date measure the properties of individual galaxies to draw broader conclusions about galaxy evolution. These galaxies are often selected from surveys targeting very luminous objects, such as quasars and luminous IR galaxies (Carilli & Walter 2013). Such catalogs provide a biased sample of gas tracers rather than measuring the average population. Even when line emission is detected blindly (e.g., Decarli et al. 2016), these detections may be limited to the brightest objects, so they are not representative of most galaxies. Blind surveys are often in cosmologically small volumes (especially at low redshift), making the source counts subject to cosmic variance (Robertson 2010). Furthermore, surveys for individual objects are expensive because they must detect galaxies at high significance, which requires large apertures or interferometers to gain high flux sensitivity and avoid confusion.

Intensity mapping (IM; Hogan & Rees 1979; Scott & Rees 1990; Madau et al. 1997; Suginoara et al. 1999; Chang et al. 2008; Wyithe et al. 2008; Visbal & Loeb 2010; Visbal et al. 2011) overcomes many of these challenges, so it is attracting increasing interest and investment, as detailed in the recent white paper by Kovetz et al. (2017). Rather than identifying individual objects as in a galaxy redshift survey, IM measures the integrated emission at a given (observed) frequency from galaxies or the intergalactic medium (IGM). It requires only modest aperture sizes to reach the smallest scales of cosmological interest, regardless of confusion. The method was originally developed for 21 cm radiation from reionization (Hogan & Rees 1979; Scott & Rees 1990; Madau et al. 1997) but has expanded to numerous lines and science goals. Observations of CO, [C II], and 21 cm can constrain the luminosity function of galaxies too faint to observe individually at high redshift. These observations will be critical for constraining the most uncertain aspects of current galaxy formation models, namely, the role of star formation and “feedback” from stars, supernovae, and active black holes.

Several measurements that bear on galaxy evolution have already been provided by IM. Cross-correlation of 21 cm emission and WiggleZ (Switzer et al. 2013) has determined the H I abundance at $z \sim 0.8$, which is consistent with feedback from active galactic nuclei (Villaescusa-Navarro et al. 2016; Padmanabhan et al. 2017). A preliminary indication of [C II] emission in cross-correlation between Planck and the Baryon Oscillation Spectroscopic Survey (BOSS) quasar sample for $2 < z < 3.2$ (Pullen et al. 2018) favors collisional excitation models at these redshifts that are different from local scaling relations. COPSS-II (Keating et al. 2016) observed CO to provide the first inferences of the molecular gas abundance $2.3 < z < 3.3$, ruling out several models. Croft et al. (2018) found diffuse Ly\alpha emission in cross-correlation with the Ly\alpha forest and quasars at a level that is consistent with hydrodynamic simulations. Kovetz et al. (2017) described several dedicated IM experiments, both ongoing and future.

An interpretation of an autonomous IM survey must argue that there is no bias from residual bright continuum radiation (Milky Way and extragalactic), interloper lines from other redshifts, or instrumental response. Cross-correlation with overdensities inferred from a galaxy redshift survey provides an approach to circumvent this foreground contamination. Here, uncorrelated foreground variance increases errors but does not bias the determination. Cross-correlation also allows breakdown by galaxy type or environment to probe the impact of those factors on the interstellar medium (ISM) (Wolz et al. 2017).

We describe several challenges and approaches related to cross-correlation in the specific context of the measurement in Pullen et al. (2018). Here the cross-power of Planck 545 GHz

\url{https://doi.org/10.3847/1538-4357/aa9a9b}
and BOSS quasars is used to constrain [C II] across $2 < z < 3.2$, and [C II] is a tracer of great interest for IM. Star formation excites the 157.7 μm (1900 GHz) $^2P_{3/2} \rightarrow ^2P_{1/2}$ fine-structure transition of [C II] in several phases of the ISM. Emitting as much as 0.65% of the typical $L_{\text{FIR}} \sim 10^2 L_\odot$ far-IR luminosity of the host (Stacey et al. 2010), [C II] is the brightest far-IR cooling line of star-forming galaxies. There is a well-established log-linear relation between [C II] and the SFR (De Looze et al. 2014; Pineda et al. 2014; Herrera-Camus et al. 2015) that becomes more complex at high redshift (Capak et al. 2015; Vallini et al. 2015; Lagache et al. 2018) and as a function of metallicity (Croxall et al. 2017; Smith et al. 2017).

The analysis in Pullen et al. (2018) estimates the [C II] surface brightness as $I_c = 6.6^{+5.0}_{-4.8} \times 10^4$ Jy sr$^{-1}$ but finds inconclusive support for a [C II] emission component from a Bayesian model test. This is a tantalizing suggestion of [C II] at $2 < z < 3.2$, and one of our goals here is to attempt to upgrade it to a secure detection or more stringent limit. The BOSS-North region is well out of the galactic plane, but the 545 GHz map does have significant Milky Way dust emission. This contamination suggests that additional effort toward suppressing the foregrounds may be fruitful. With this as a backdrop, we will describe the general problem of rejecting bright Milky Way continuum foregrounds while constraining correlated line and continuum emission.

Typically, IM is described in terms of one isolated line, such as [C II] above. However, the target galaxies also contribute a characteristic spectral energy distribution (SED), which includes bright dust or a synchrotron continuum contribution at most frequencies of interest for IM. The overdensity field traced by a companion galaxy redshift survey will, therefore, have nonzero cross-correlation with the intensity survey across many frequencies. The cross-correlation of the intensity survey at $v_i$ with galaxies at $z_i$ therefore measures the SED of galaxies at $z_i$ emitting into $v_i$, SED$(v_i, z_i)$ (Serra et al. 2014; Switzer 2017; Chiang et al. 2018; Pullen et al. 2018).

A determination of the line brightness must marginalize over the continuum part of the SED. Traditional foreground cleaning that removes uncorrelated continuum emission (such as from the Milky Way) will also suppress the correlated continuum radiation. The cross-correlation in a cleaned map no longer traces SED$(v_i, z_i)$ but has residual correlated continuum emission due to the slight differences in SED between correlated (largely Milky Way) and correlated continuum radiation. While foreground cleaning may be successful in revealing the redshifted line emission, it may muddle the correlated continuum and produce bias. We argue that IM analyses should apply a self-consistent approach to deweighting uncorrelated continuum foregrounds and estimating correlated continuum in the target galaxies.

Estimates of the emission of the Milky Way that are independent of extragalactic sources have utility in removing foregrounds without impacting the extragalactic correlated continuum signal. We find that existing spatial templates of reddening to stars (Green et al. 2015) and inferred from HI (Lenz et al. 2017) do not significantly aid the line amplitude estimation in the case of Planck 545 GHz × BOSS quasars. Similar templates may have great utility in the future for separating continuum terms.

A fundamental challenge in the Planck analysis is the number of available frequency channels relative to the number of estimated parameters. We consider a future tomographic spectral survey with many frequency bands and argue that low $k_\perp$ bins isolate the correlated continuum emission (barring instrumental effects). Cutting these bins has a minor impact except in reducing sensitivity at low $|k|$. These conclusions bode well for future spectral IM surveys (Kovetz et al. 2017), which anticipate numerous frequency channels. Throughout, we will use the example of [C II], but considerations here apply to 21 cm and other targets of IM.

Section 2 reviews Milky Way foregrounds in the Planck data and describes a simple map-space cleaning approach. This approach neglects the joint interpretation of uncorrelated continuum emission from the Milky Way and correlated continuum emission from the target galaxies. Section 3 develops some general results for IM with both correlated and uncorrelated continuum emission in a simplified setting. It argues for the utility of Milky Way–only templates, which we apply to the Planck × BOSS analysis in Section 5. Marginalizing over a correlated continuum model amounts to throwing out those degrees of freedom (dof) along the line of sight. An intensity survey should therefore have many more bands than correlated continuum model parameters. Section 6 considers this true tomographic case and shows that correlated continuum bias lies at low $k_\perp$.

2. Milky Way Foregrounds in the Planck Data

Pullen et al. (2018) presented a preliminary indication of [C II] emission from $2 < z < 3.2$ that is correlated with large-scale structure (LSS) traced by quasars in the BOSS survey. Planck High Frequency Instrument (HFI) data (Lamarre et al. 2010; Planck HFI Core Team et al. 2011) at 545 GHz serve as the intensity map, and measurements at 353 and 857 GHz help to isolate dust continuum and thermal Sunyaev-Zel’dovich (tSZ) emission from the [C II] of interest. Figure 1 shows significant contamination from the Milky Way in the BOSS-North region of the Planck 545 GHz map. This Milky Way emission is uncorrelated with the [C II] signal from $2 < z < 3.2$, so it contributes variance but not bias to the cross-correlation with quasars. Figure 2 shows that the [C II]...
signal from Pullen et al. (2018) remains $100\times$ lower than the total 545 GHz power spectrum, which is dominated by foregrounds at all scales.

Pullen et al. (2018) performed no foreground cleaning in the map domain, so the bright Milky Way emission contributes significant variance to each of the cross-powers of 353, 545, and 857 GHz with BOSS. Our primary goal here is to assess whether some form of foreground cleaning could upgrade the modest indication of [C II] in Pullen et al. (2018) to a secure detection of redshifted line emission. We will start with a standard approach of cleaning Milky Way (uncorrelated) continuum emission in map space and argue that this is complicated by correlated dust continuum emission from the target galaxies.

A simple approach to cleaning galactic emission is to form a linear combination of adjacent bands, as

$$\mathbf{x}_{\text{clean}} = \mathbf{x}_{545} - \kappa_{353}\mathbf{x}_{353} - \kappa_{857}\mathbf{x}_{857}. \tag{1}$$

The goal of finding linear combination parameters $\kappa_{353}$, $\kappa_{857}$ is to minimize the auto-power or rms variation of the cleaned map (and hence the error bars of the cross-correlation). Stack the 353 and 857 GHz maps into a matrix $\mathbf{A}$. Then the linear combination coefficients that minimize variance are given by

$$\kappa = (\mathbf{A}^T \mathbf{N}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{N}^{-1} \mathbf{x}_{545}. \tag{2}$$

where $\mathbf{N}$ is the covariance of the 545 GHz map. Figure 3 shows the map after linear combination cleaning, with some residual Milky Way dust emission still visible but greatly reduced.

Figure 4 shows the cross-power of quasars with the cleaned map compared to the input 545 GHz map. The dramatic change in the cross-power from linear combination cleaning motivates a reconsideration of foregrounds in the Pullen et al. (2018) analysis. Not only do the errors diminish, but the level of the cross-power also drops considerably. This is because much of the cross-correlation between Planck 545 GHz and BOSS quasars is from correlated continuum emission.

The linear combination that cleans the Milky Way from the 545 GHz map also removes much of the correlated continuum emission. If the dust continuum emission shared the same SED, the linear combination cleaning would be highly effective in removing Milky Way foregrounds and allow a unique determination of the line brightness independent of correlated continuum emission. The fact that the Milky Way dust SED (which drives the linear combination coefficients) is not identical to the correlated continuum SED means that the cleaned cross-power in Figure 4 has some poorly determined residual correlated continuum emission.

3 An internal linear combination can be biased by spurious correlations between signal and foreground, especially on large angular scales. Simulations show $<1\%$ bias for $\ell > 200$ (the range reported here), but as we will argue, this map-domain cleaning is primarily illustrative and is not used in parameter estimates.
dust continuum to separate the [C II] contribution. The clustering of [C II] emission with quasars in $2 < z < 3.2$ is localized to the 545 GHz band, so a linear combination of bands that cleans 545 GHz is an effective way to isolate [C II]. In contrast, the marginalization over the correlated dust continuum depends on information from 353 and 857 GHz, so it becomes entangled with attempts to use those bands to clean Milky Way emission at 545 GHz.

3. Sensitivity with Continuum Emission

This section builds up a set of analytic results for multitracer analysis of IM data with correlated and uncorrelated continuum emission contamination. It has become standard to clean continuum in IM data in the map domain and simulate the impact of the resulting signal loss on the power spectrum (e.g., Switzer et al. 2015 for single-dish instruments and Cheng et al. 2018 for interferometers). In contrast, the multitracer approach developed here jointly (1) deweights Milky Way foregrounds, (2) marginalizes over correlated continuum emission, and (3) estimates the line amplitude. In this context, signal loss does not need to be estimated separately because it is included self-consistently in the parameter estimate. Rather than calculating signal loss through simulations, expressions here will focus on the error of the line estimate and how continuum emission causes constraints to degrade from thermal noise limits.

For the expressions to remain tractable and illustrative, we will make several simplifying assumptions: (1) shot noise in the galaxy redshift survey is negligible, and the linear bias is known; (2) inter-bandpower correlations are negligible, so we can consider expressions for a single bandpower/multipole; and (3) there is no stochasticity between the density traced in the intensity and galaxy redshift survey. The line intensity and galaxy fields should trace the same underlying density on linear scales, making this approximation reasonable on large-area surveys such as used in Section 5.

This regime is applicable to single-dish surveys that can access linear modes, such as BINGO (Battye et al. 2016), GBT (Switzer et al. 2013), Parkes (Anderson et al. 2018), and MeerKAT/SKA (Bull et al. 2015) for 21 cm, and most of the proposed experiments for high-frequency lines: CCAT-prime (Vavagiakos et al. 2018), COMAP (Tveit Ihlé et al. 2018), CONCERTO (Dumitru et al. 2018), EXCLAIM (Padmanabhan 2018), SPHEREx (Doré et al. 2014), STARFIRE (Uzgil et al. 2014), and HEITDEX (Fonseca et al. 2017). Some of these simplifying assumptions can be easily relaxed in this framework. Appendix A describes the impact of stochasticity and shot noise, and Section 4.2 describes low $k$ modes. Additionally, the spherical harmonic approach here is well-matched to surveys on large (curved) regions of the sky.

3.1. Uncorrelated Foregrounds in a Multitracer Setting

Foregrounds that are uncorrelated with cosmological density will increase the errors of a determination of the line brightness from cross-correlation but not bias the result. This section quantifies the impact of uncorrelated foregrounds in a multitracer (McDonald & Seljak 2009) setting, reviewed in Appendix A.

Let $\delta$ (a vector) be a spatial map of the cosmological overdensity in a redshift slice that corresponds to a frequency channel in the intensity survey. Model the observed galaxy redshift survey map $x_\mu$ as a tracer with linear bias $b_\mu$ and shot noise $n_\mu$. Model the intensity map as a line surface brightness amplitude $S_L$ times overdensity $\delta$ plus noise $n_I$. (Here $S_L = I_L b_L$ refers to the line brightness $I_L$ multiplied by the linear cosmological bias $b_L$ of the line emission.) Emission from the Milky Way and galaxies not in the selected redshift slice will contribute continuum emission that is uncorrelated with $\delta$. Let the uncorrelated continuum foregrounds have a spatial template $x_u$ and spectral dependence $f(\nu)$. These appear in the intensity map at frequency $\nu_I$ as $f(\nu_I)x_u$. All vectors refer to 2D map slices at constant frequency or redshift. The observed galaxy redshift survey $(x_\mu)$ and intensity $(x_I)$ maps are

$$x_\mu = b_\mu \delta + n_\mu,$$

$$x_I = S_L \delta + f(\nu_I)x_u + n_I. \quad (3)$$

Form a data vector $d = \{C^{\ell\ell}_F, C^{\ell\ell}_G\}$ from the power spectrum of the galaxy redshift survey $C^{\ell\ell}_G$ and the cross-power between the intensity and galaxy redshift survey maps $C^{\ell\ell}_F$. Model the observed two-point functions as

$$C^{\ell\ell}_F = b_\mu S_L C^{\delta\delta}_\ell,$$

$$C^{\ell\ell}_G = \delta^2 C^{\delta\delta}_\ell + N_{\ell}, \quad (4)$$

where the galaxy bias and shot noise are assumed to be known, for simplicity. The model then has parameters $\theta = \{S_L, C^{\delta\delta}_\ell\}$, where $C^{\delta\delta}_\ell$ is the power spectrum of the underlying density $\delta$ fluctuations. For illustrative purposes, we consider the constraint based on a single multipole or bandpower. The log-likelihood is

$$2\mathcal{L} = \ln \det \Sigma + [d - \mu(\theta)]^T \Sigma^{-1}(d - \mu(\theta)). \quad (5)$$

The terms in the covariance $\Sigma$ under the assumption of Gaussian fluctuations in the data are given by

$$\text{Cov}(C_A^{\ell\ell}, C_B^{CD}) = \frac{1}{\nu_C} \delta_{\ell_C \ell_D} \left( C_A^{CD} C_B^{BC} + C_A^{AC} C_B^{BD} \right), \quad (6)$$

where $C_A^{BC}$ is the cross-power of fields $A$ and $B$, and $\nu_C$ is the number of modes in the measurement, which is roughly $(2\ell+1)f_{\text{sky}}$ for a 2D map on the fraction $f_{\text{sky}}$ of the sky and a bandpower of width $\Delta \ell$. The delta function $\delta_{\ell_C \ell_D}$ is exact for all-sky, and for simplicity, we assume that it is approximately the case for broad bandpowers on a partial sky.

Following the approach of Appendix A, the variance of a determination of the line amplitude $S_L$ from a single multipole $\ell$ is (marginalized over $C^{\delta\delta}_\ell$)

$$\sigma^2_{S_L} = \frac{N_{\ell} + f(\nu_I)^2 C^{\mu\mu}_B}{C^{\delta\delta}_\ell \nu_I}, \quad (7)$$

where $C^{\mu\mu}_B$ is the power spectrum of the uncorrelated emission (the map $x_u$), $N_{\ell}$ is the noise variance of the intensity map, and $C^{\delta\delta}_\ell$ is the power spectrum of the density fluctuations $\delta$. This expression is exact (neglecting shot noise), rather than an expansion about small-intensity map noise (Bernstein & Cai 2011). From Equation (7), foregrounds that are uncorrelated with the cosmological signal play the same role as noise in the intensity map.

Cosmic variance would appear in $\sigma^2_{S_L}$ as a term $\propto S^2_L$. As described in Appendix A, this is avoided by adding covariance to the galaxy auto-power $C^{\delta\delta}_\ell$ in the likelihood (assuming negligible stochasticity between the intensity and galaxy redshift survey).
3.2. Deweighting Uncorrelated Continuum Emission

Intensity measurements at two frequencies can separate uncorrelated continuum and line emission. Include a “veto” map \( x_\nu \) at a different frequency \( \nu_\nu \). The uncorrelated continuum foreground \( f(\nu_\nu) x_\nu \) in the intensity map will appear at this frequency as \( f(\nu_\nu) x_\nu \), where \( f(\nu_\nu) \) accounts for the SED and \( u' \), denotes the spatial pattern of uncorrelated foregrounds in the veto map which may not be fully coherent with the spatial pattern \( x_\nu \) in the intensity map. Model this set of maps as

\[
\begin{align*}
    x_g &= b_g \delta + h_g, \\
    x_\nu &= f(\nu_\nu) x_\nu + S_\nu \delta + n_\nu, \\
    x_V &= f(\nu_\nu) x_{\nu'} + n_{\nu'}. \tag{8}
\end{align*}
\]

The veto map will also contain line signal from a different redshift, given as \( S_\nu(z_\nu) \delta(z_\nu) \), but we will assume that this has a negligible correlation with \( \delta(z_\nu) \) in the primary intensity survey. That is an excellent approximation for, e.g., [C II] emission in Planck 353 GHz versus 545 GHz, which originates from considerably different redshifts. In this case, the redshifted line emission in the veto band amounts to uncorrelated noise, so it is accommodated in the assumed noise power \( N_\nu \). Two bands are sufficient to fit for a line amplitude and one continuum parameter.

Intensity surveys with greater adjacency in the bands must account for signal correlations in the likelihood. Standard codes (Challinor & Lewis 2011; Di Dio et al. 2013) can calculate interband signal correlations, and several analyses of LSS (Gaztañaga et al. 2012; Di Dio et al. 2014; Nicola et al. 2014; Salazar-Albornoz et al. 2017) have used similar approaches that could be adopted for IM. Appendices B and C describe the multiband case, which can accommodate more general SEDs of the continuum correlations and interband signal correlations. Sections 4.2 and 6 describe low-\( k_\| \) modes in the \( C_\ell \) and \( P(k_\|, k_\perp) \) context, respectively.

To characterize the lack of coherence between the spatial distribution of uncorrelated foreground contamination in the intensity map \( x_g \) and veto map \( x_\nu \), define “foreground” stochasticity \( r_\nu = c_\nu^{\text{off}} \int \frac{c_\nu^{\text{on}}}{c_\nu^{\text{on}}} c_\nu^{\text{off}} \), where \( c_\nu^{\text{off}} \) is the cross-power between \( x_g \) and \( x_{\nu'} \).

Expand the observation vector in Equation (5) to include the cross-power of the veto map and galaxy redshift survey \( \mathbf{d} = \{ C_\nu^{\ell \ell}, C_\nu^{\ell V}, C_\nu^{\ell g} \} \) and estimate the parameters \( \mathbf{\theta} = \{ S_\nu, C_\nu^{\ell g} \} \). In the limit that the veto band measures the foregrounds well, e.g., \( N_\nu \ll f(\nu_\nu)^2 C_\nu^{uu} \). Then,

\[
\sigma^2 \mathbf{x}_\nu = \frac{1}{\nu_\nu C_\nu^{\ell \ell}} \left[ N_\nu + \frac{f(\nu_\nu)^2}{f(\nu_\nu)^2} N_\nu \right] \frac{f(\nu_\nu)^2 C_\nu^{uu}}{\nu_\nu C_\nu^{\ell g}} (1 - r_\nu^2). \tag{9}
\]

To interpret this result, make a cleaned map that removes uncorrelated continuum emission using the veto map, as \( x_{\nu\text{clean}} = x_\nu - f(\nu_\nu)/f(\nu_\nu) x_\nu \). In the limit that the foreground in the veto band perfectly traces the intensity map \( r_\nu = 1 \), \( x_{\nu\text{clean}} = S_\nu \delta + n_\nu - f(\nu_\nu)/f(\nu_\nu) n_\nu \). The noise variance in this cleaned map is \( N_\nu + N_\nu f(\nu_\nu)^2/f(\nu_\nu)^2 \), which is the numerator of the first term in Equation (9). The first term therefore describes the impact of thermal noise in a cleaned map. To interpret the second term, note that the numerator is \( f(\nu_\nu)^2 C_\nu^{uu} \), the foreground covariance in the intensity map. If \( r_\nu = 0 \), the veto band provides no cleaning in the intensity band, and the variance is increased by the full foreground brightness analogous to Equation (7). Alternately, if \( r_\nu = 1 \), the veto band is a perfect tracer of uncorrelated foregrounds in the intensity map, and the second term goes to zero. In this limit, the line brightness can be determined without impact from the foreground variance.

We have not explicitly cleaned the map anywhere in the likelihood. The appearance of an underlying cleaned map \( x_{\nu\text{clean}} \) arises self-consistently with the parameter estimate through the action \( \Sigma^{-1} \mathbf{d} \) in the likelihood, Equation (5). Foregrounds common to both the intensity and veto maps show up as correlated noise between the cross-powers \( C_\nu^{UV} \) and \( C_\nu^{Vg} \). Specifically, these off-diagonal terms in \( \Sigma \) appear from the cross-variance \( C_\nu^{UV} = f(\nu_\nu) f(\nu_\nu) C_\nu^{uu} \). For a simple demonstration in the limit that \( r_\nu = 1 \) and there is no noise in the veto map \( N_\nu = 0 \), \( \Sigma^{-1} \) takes a linear combination of the cross-powers \( C_\nu^{UV} \) and \( C_\nu^{Vg} \) to form a cleaned cross-power analogous to the cleaned map, as

\[
C_\nu^{\ell g \text{clean}} = C_\nu^{\ell g} - \frac{f(\nu_\nu)}{f(\nu_\nu)} [f(\nu_\nu)] C_\nu^{\ell V}. \tag{10}
\]

An additional map cleaned through a linear combination in map space will not add information to the likelihood. Furthermore, taking a linear combination such as Equation (1) outside of the context of the likelihood can induce bias from, e.g., spurious correlation of foregrounds and signal. This necessitates signal loss simulations to account for the impact of any map operations done before the parameter likelihood in power-spectrum space (Switzer et al. 2015).

In summary, bright Milky Way continuum foregrounds can be deweighted through observations at multiple frequencies, limited by the coherence \( r_\nu \) between frequencies.

3.3. Uncorrelated and Correlated Continuum Emission

In addition to line emission, the host galaxies also emit a dust continuum that correlates with cosmological overdensity. To determine the line amplitude independently of the continuum in these galaxies, the continuum contribution must be modeled and marginalized over. Model this correlated continuum component as \( g(\nu) S_C \delta \), where \( g(\nu) \) describes the SED of the correlated continuum and \( S_C \) is the amplitude of the continuum. Extending the map model in Equation (8),

\[
\begin{align*}
    x_g &= b_g \delta + h_g, \\
    x_\nu &= f(\nu_\nu) x_\nu + S_\nu \delta + n_\nu, \\
    x_V &= f(\nu_\nu) x_{\nu'} + g(\nu_\nu) S_C \delta + n_{\nu'}. \tag{11}
\end{align*}
\]

Continue to use the observation vector \( \mathbf{d} = \{ C_\nu^{\ell \ell}, C_\nu^{\ell V}, C_\nu^{\ell g} \} \) but expand the parameters to include an estimate of the correlated continuum amplitude \( S_C \) as \( \mathbf{\theta} = \{ S_\nu, S_C, C_\nu^{\ell g} \} \). The variance of the estimate of \( S_C \) is not marginalized over the other parameters \( (F_S \Sigma_S)^{-1} \) for Fisher matrix \( F \) is the same as Equation (9). That is, we can determine the amplitude \( S_C \) in the intensity map as before. However, now the correlated amplitude is \( S_C = S_\nu + g(\nu_\nu) S_C. \) To determine the line brightness \( S_L \) independently of the correlated continuum \( g(\nu_\nu) S_C \), we need to...
marginalize over $S_C$, giving

$$
\sigma_{\delta L}^2 = \frac{1}{vT_C^2} \left[ N_C \varepsilon_{\ell} + \frac{g(\nu_l)^2}{g(\nu_l)^2} N_T \right] + \frac{C_{\ell}^\text{corr}}{vT_C^2} \left[ \frac{f(\nu_l) - f(\nu_l)g(\nu_l)}{g(\nu_l)} \right]^2 \right] \varepsilon_{\ell} \right] + \frac{2C_{\ell}^\text{corr}}{vT_C^2} \left( 1 - r_p \right) f(\nu_l) f(\nu_l) \frac{g(\nu_l)}{g(\nu_l)}.
$$

(12)

To interpret this result, make a cleaned map that removes the correlated continuum emission using the veto map, as

$$
\hat{x}_\text{clean} = x_l - \frac{g(\nu_l)}{g(\nu_l)} x_Y.
$$

Unlike the previous section, this map cleans the correlated continuum emission using a linear combination of bands. The first term in Equation (12) is just the thermal noise in this cleaned map. The term in brackets in the second line is the residual of the uncorrelated continuum foreground emission after the correlated emission has been cleaned. If both the correlated and uncorrelated continua have the same SED, $g(\nu_l) \propto f(\nu_l)$, both are deweighted equally. The third term is a cross-term and arises from the stochasticity between the spatial shape of uncorrelated foregrounds in maps $x_l$ and $x_Y$, which contributes variance to $\sigma_{\delta L}^2$ even if the correlated and uncorrelated continua have the same SED.

### 3.4. Bias from Incorrect Correlated Continuum Models

If the correlated continuum spectrum is not accurately modeled, residuals may be spuriously interpreted as line emission. In the linear model for two intensity bands (Equation (11)), let the true correlated continuum SED be $g'(\nu_l)$ and model it as $g(\nu_l)$. In this case, the estimated line amplitude is

$$
\hat{S}_L = S_L + S_C \left[ g'(\nu_l) - \frac{g(\nu_l)}{g(\nu_l)} g'(\nu_l) \right].
$$

(13)

which is biased by the difference between the true SED and the model.

In the case where the number of instrument bands $N_{\text{band}}$ equals the number of parameters $N_{\text{param}}$, the spectral model uses all dof, and there is no additional handle to identify bias in the line amplitude. The model may have an excellent goodness of fit because it is using $S_L$ to fit a residual from an insufficient continuum model. If $N_{\text{band}} \gg N_{\text{param}}$, channels that have no expected line emission correlation should be consistent with zero, providing a test.

In summary, in the case of relatively few bands, the line amplitude may be undetectably compromised by an incomplete continuum SED model and subject to variance from bright Milky Way emission that cannot be independently down-weighted. Section 4 develops a model to forecast an experiment with many bands in a $C_L$ approach. Section 6 describes correlated continuum emission in the context of $P(k_c, k_b)$ in a survey with many bands.

### 4. Extension to Many Bands

The preceding toy models give some intuition for the impact of uncorrelated and correlated continuum emission. The analytic Fisher matrix approach presented there becomes cumbersome to treat more than two intensity channels, requiring analytic $N_{\text{band}} \times N_{\text{band}}$ inverses for $N_{\text{band}}$ maps for each parameter, followed by an analytic $N_{\text{param}} \times N_{\text{param}}$ inverse over parameters to find the marginalized error on $S_L$.

Appendix C builds a linear model for the joint deweighting of uncorrelated continuum and fit to both correlated line and continuum emission. Here we model the cross-power at each $\ell$ as a set of correlated continuum and line parameters $\theta$ that linearly describe the measured cross-powers as a function of frequency $\ell_l$ as $\ell_l = M\theta + n_x$, where the bandpowers have noise $n_x$, described by covariance $N_x$. The parameters $\theta = (S_L, S_C)$, where $S_L$ is the line amplitude and $S_C|l = S_C$ is the amplitude of linear spectral templates $g(\nu_l)$ that describe the correlated continuum emission as a function of frequency $\nu_l$. The matrix $M = \{g_1, g_2, \ldots, g_n\}$ holds the spectral templates for the line correlation $\xi$ and $N_\text{comp}$ component spectral modes of the correlated continuum, and $\xi = \xi(\nu_l)$ describes the correlation of the line intensity at each frequency $\nu_l$ with the galaxies at redshift $z$. In the case of the Planck, the bands are widely spaced in redshift so that $\xi(\nu_l)$ is a delta function at $\nu_l = 545$ GHz to an excellent approximation. Section 4.2 describes the impact of correlations at low $k_l$ on an intensity survey with many narrow bands. Appendix B describes how the Milky Way produces strong, low-rank correlations in $N_x^{-1}$, and Appendix D describes the case of foregrounds that have the same spatial template at all frequencies.

The variance $\Sigma_\theta$ on the correlated line and continuum amplitude parameters $\theta$ in this simple linear model is

$$
\Sigma_\theta = (M^T N_x^{-1} M)^{-1}.
$$

(14)

Appendix C shows that this simple form reproduces the results from previous sections under a slightly more restrictive set of assumptions.

### 4.1. How Many Bands Are Needed?

The linear formulation in Equation (14) emphasizes an accounting for overall dof. Let the line signal have one dof (the line amplitude $S_L$), the correlated continuum model have $D_{\text{corr}}$ dof, and the uncorrelated continuum emission model have $D_{\text{uncorr}}$. Assuming these modes are mutually independent, the total number of dof needed to account for all terms is $D_{\text{corr}} + D_{\text{uncorr}} + 1$. If a survey has $D_{\text{corr}} + 1$ bands, it can separate correlated line and continuum emission, but it has no dof remaining to suppress uncorrelated (Milky Way) emission. In the model of Equation (11), there are two intensity map channels and one dof for the line and correlated continuum emission, respectively. This leaves no additional freedom to remove uncorrelated foregrounds, and these show up in the second term of Equation (12).

Equation (14) permits a rapid simulation of an instrument with $N_{\text{band}}$ uniform bands for a hypothetical $|C_\ell|$ survey from $2 < z < 2.5$ (543 and 633 GHz) to illustrate dof. For correlated emission, we fit an unknown amplitude for a $T_{\text{dust}} = 27.2$ K, $\beta = 1.5$ dust source at $z = 2.25$ ($T_{\text{dust}} = 1.0 + z$), where $B(T_{\text{dust}}, \nu)$ is the Planck law and for the amplitude of the derivative of intensity with respect to temperature $dI_{\text{dust}}/dT$, corresponding to a fit to the dust temperature to first order. There are two correlated dof to fit the continuum and one signal dof, which gives three dof in total. For Milky Way dust emission, we take emission at 19 K and $\beta = 1.5$ with an amplitude that is $10 \times$ the thermal noise (after marginalizing over the correlated continuum). We take the signal correlation...
to exist only in one spectral band, but Section 4.2 describes the impact of signal correlations along the line of sight. Figure 5 shows the inflation of errors (relative to thermal) as a function of the number of bands. In a survey with three bands, there is no remaining latitude to deweight uncorrelated continuum, and the error on the line amplitude is penalized by 10× thermal, consistent with the variance of the contamination. Adding one additional channel provides a dof to suppress the Milky Way, independent of the correlated continuum determination.

The simulation in this section is set up to emphasize the impact of dof in a general survey. Section 5.6 uses Equation (14) to interpret the Planck × BOSS and shows that there is not a strong penalty for the three bands in this particular case, given the measured amplitude of the Milky Way foreground and similarity in the correlated and uncorrelated spectral indices.

4.2. Line-of-sight Correlations

Cosmological structure is also correlated along the line of sight, so a slice of density from the galaxy redshift survey at redshift $z$ will correlate not only with line emission from redshift $z$ but also nearby redshifts $z'$. Correlations at low $k_\parallel$ (and consequently at low $\ell$) can extend over many spectral bins in the intensity map. In the case of Planck, this effect is negligible between the bands 353, 545, and 857 GHz (at the $\ell$ reported here), but the effect will be significant for future tomographic surveys with many narrow bands.

Section 6 develops a complete view of this effect by considering the 3D power spectrum of a tomographic survey. It shows that signal at low $k_\parallel$ is corrupted by correlated continuum emission. However, the linear framework in Equation (14) provides another approach to understand the impact of this effect. Appendix C describes line-of-sight signal correlations in terms of a correlation kernel $\xi(\nu_\gamma)$ between the line intensity at $\nu_\gamma$ and a galaxy survey at fixed redshift $z$. This kernel can be calculated as described in Challinor & Lewis (2011) and Di Dio et al. (2013).

Figure 6 demonstrates the inflation of errors at low $k_\parallel$, in the same simulation setup as Section 4.1 and taking 200 channels. At low $k_\parallel$, the clustered line emission is spectrally smooth. It is, therefore, less distinguishable from correlated and uncorrelated continuum emission, so its errors rapidly grow. Here modes with spatial wavenumbers $k_\parallel$ 10× higher than the largest modes in the volume gain immunity by varying rapidly spectrally compared to the continuum.

It has been recognized since Hogan & Rees (1979) that bright and spectrally smooth Milky Way emission contaminates long $k_\parallel$ modes. A new point emphasized here is that correlated continuum emission also contaminates these modes.

4.3. Application to Auto-power Analysis

The formalism and discussion here have applied to the cross-power between an intensity survey and galaxy redshift survey tracer. It is useful to contrast this with considerations for the intensity survey auto-power. In the cross-power, the uncorrelated continuum (Milky Way) adds variance, so the aim is simply to reduce the Milky Way variance. In the auto-power, residual Milky Way variance translates directly into a bias that is difficult to model because the Milky Way (1) is much brighter than the correlated continuum (and so instrumental response is even more critical), (2) is not statistically isotropic, and (3) does not have a uniform spectrum. Our discussion is therefore limited to the cross-power, with possible application to the auto-power deferred to future work.

4.4. Adding a Template Map for the Milky Way

The measurement of Planck × BOSS quasars is in the challenging regime where $N_{\text{band}} \approx N_{\text{param}}$. Given that this is the best intensity map for [C II] currently available, we can try to improve the foreground cleaning with a Milky Way template.

Extend the map model in Equation (11) to include a Milky Way template with amplitude $A_M$ and noise variance $N_{M}$, as

\begin{align*}
    x_d &= b_d \delta + n_d, \\
    x_f &= f(\nu_\gamma)x_u + g(\nu_\gamma)S_C \delta + S_f \delta + n_f, \\
    x_V &= f(\nu_\gamma)x_u + g(\nu_\gamma)S_C \delta + n_V, \\
    x_M &= A_M x_u + n_M.
\end{align*}

(15)
Equation (14) gives

\[ \sigma_{S}^2 = \frac{1}{\nu I C_{\ell}^{\text{CMB}}} \left[ N_{\ell} + \left( \frac{g(\nu_I)}{g(\nu)} \right)^2 N_{\nu} \right] + \frac{N_{\ell} A_{\ell}^2}{\nu I C_{\ell}^{\text{CMB}}} \left[ f(\nu_I) - f(\nu) \frac{g(\nu_I)}{g(\nu)} \right]^2. \]

This is just Equation (12) where the bright Milky Way variance \( C_{\ell}^{\text{MW}} \) has been traded with signal-to-noise ratio in the tracer map, \( N_{\ell}/A_{\ell}^2 \). Appendix E includes the impact of stochasticity \( r_M \) between the tracer of the Milky Way emission and the continuum emission in the observed bands. In this case, an additional term of the variance is

\[ \sigma_{S, \text{loc}}^2 = \frac{C_{\ell}^{\text{MW}}}{\nu I C_{\ell}^{\text{CMB}}} \left[ f(\nu_I) - f(\nu) \frac{g(\nu_I)}{g(\nu)} \right]^2 (1 - r_M^2). \]

If the Milky Way template traces the dust emission poorly, \( r_M \to 0 \), this reverts to the full foreground contamination in Equation (12). Appendix D describes the extension to many bands.

5. Milky Way Template Cleaning Applied to Planck \( \times \) BOSS

The previous section argued that Milky Way templates can provide leverage to suppress foregrounds, especially in the case where the number of spectral channels is similar to the number of correlated continuum parameters. The analysis of Planck \( \times \) BOSS for [C II] in Pullen et al. (2018) is in this regime. This section considers the application of Milky Way templates in this case to boost the signal-to-noise ratio of the correlated line emission.

5.1. Existing Milky Way Templates

There are approaches to using broadband measurements to model thermal dust emission in the Milky Way (Planck Collaboration et al. 2014a, 2016b; Meisner & Finkbeiner 2015). However, these galactic templates are prone to also containing extragalactic thermal dust emission, including correlated continuum emission. We instead seek models of thermal dust emission that do not rely on sensitivity to dust emission directly.

Thermal dust emission depends on both the column depth and the temperature. Existing inferences of temperature measure dust emission around the peak of the SED using broadband observations. These maps may, therefore, correlate with extragalactic radiation and cannot be used. Both reddening and H I abundance trace the dust column depth and can be inferred independently of extragalactic emission.

We use the direct inference of reddening to Pan-STARRS stars from Green et al. (2015), shown in Figure 7. This map is noisy, capturing only the brightest features in the BOSS-North field. To reach higher sensitivity, we also consider the reddening map from Lenz et al. (2017) inferred from HI4PI (HI4PI Collaboration et al. 2016) measurements of the H I column through 21 cm radiation (Figure 8). Both Green et al. (2015) and Lenz et al. (2017) are imperfect tracers of Milky Way dust emission. The H I tracers for dust density become biased in high-density regions where H2 has formed (Lenz et al. 2017), so the H I tracer is best in areas out of the galactic plane considered here. Reddening and H I are imperfect tracers of the dust column depth (Cardelli et al. 1989; Fitzpatrick & Massa 2007; Schlafly et al. 2016), and the dust column is an imperfect tracer of the thermal dust emission.

Figure 9 shows the reduction in bandpower errors when the Green et al. (2015) and Lenz et al. (2017) maps are projected out of the Planck 545 GHz channel. This linear combination in map space is only for visual demonstration. The Milky Way emission template from Lenz et al. (2017) decreases variance on the largest scales in the measured cross-power, while the template from Green et al. (2015) produces higher noise in the cross-power for \( \ell > 300 \).

5.2. Planck and BOSS Data

The Planck 353, 545, and 857 GHz maps have full width at half maximum (FWHM) of 4.41, 4.47, and 4/23, respectively, and are binned in HEALPix (Górski et al. 2005) pixelization with \( N_{\text{side}} = 2048 \). We use a point-source mask that is the union of published masks (Planck Collaboration et al. 2016a) across the three frequencies and apodized across 0.5° and a galactic emission mask with 2° apodization. In total, this leaves \( f_{\text{sky}} = 0.332 \) in the BOSS survey region. Throughout, we will work in MJy sr\(^{-1}\) units and convert from measurements in \( K_{\text{CMB}} \) units to MJy sr\(^{-1}\) using the mean coefficients 287.45 (353...
and redshift completeness $d > 0.9$ quasars over the SDSS-III (DR12). This procedure yields a catalog of 178,622 quasars, which consists of 862,735 galaxies over 10,229 deg$^2$. Additionally, requiring $0.43 < z < 0.7$ and pixel coverage $>90\%$ gives 777,202 quasars over 8294 deg$^2$ and an overlap with the Planck map of 6483 deg$^2$ with 75,244 quasars.

It is also beneficial to have a galaxy redshift sample that has no cross-correlation with the [C II] emission but does trace the CIB and SZ clustered emission as part of a parametric model for the components. Here we use the BOSS DR12 CMASS luminous red galaxy (LRG) sample (Alam et al. 2015, 2017; Reid et al. 2016), which consists of 862,735 galaxies over 9376 deg$^2$ and a mean $z = 0.57$ and is designed to be stellar mass--limited at $z > 0.45$. Sample selection here requires spectroscopic sectors (Aihara et al. 2011) with completeness $>70\%$ and redshift completeness $>80\%$. Additionally, requiring $0.43 < z < 0.7$ and pixel coverage $>90\%$ gives 777,202 galaxies over 10,229 deg$^2$.

In both the quasar and LRG samples, we form the overdensity $\delta = (n - \bar{n})/\bar{n}$ using systematic weights (Anderson et al. 2014). The LRG and CMASS mask regions are additionally apodized by a $0.5$ FWHM kernel.

5.3. Measured Cross-powers

Figure 10 shows the six angular cross-powers used to estimate [C II] emission and continuum nuisance parameters. These cross-spectra were performed over three Planck bands: 353, 545, and 857 GHz. These maps were cross-correlated with overdensity maps of quasars and LRGs cataloged by the BOSS survey. We perform the cross-correlations over angular scales $\ell = 100$–1000. These spectra are identical to those presented in Pullen et al. (2018).

We also consider two Milky Way templates to marginalize over contamination in the Planck maps. Specifically, we use the HI–derived dust map from Lenz et al. (2017) and the reddening map from Green et al. (2015). As observables, we cross-correlate these dust maps with both the quasar and LRG overdensity maps (Figure 11). Section 5.4 investigates evidence for correlation with LSS.

5.4. Models for CIB and C II Signals

We briefly review the CIB and [C II] emission models used in the [C II] constraints from Pullen et al. (2018). We begin with the angular cross-power spectrum between CIB emission in a Planck map at frequency $\nu$ and an LSS tracer overdensity...
where \( dn/dL \) is the infrared galaxy luminosity function and \( L_{\nu}(1+z) \) is a model for the CIB luminosity emitted at rest-frame frequency \( \nu(1+z) \). Here \( b_{\text{CIB}}(k,z) \) and \( dn/dL \) are predicted using the same halo model as in Shang et al. (2012). The model for the luminosity depends on several parameters; however, we only allow three parameters in our fit to vary: a luminosity amplitude \( L_0 \), a redshift evolution parameter \( \alpha \), and the dust temperature \( T_d \).

Following Pullen et al. (2018), we extend the CIB spectrum model in Shang et al. (2012) to include \([\text{C}\,\text{II}]\) emission in the host galaxies with amplitude \( A_{\text{C\,\text{II}}} \). We also add correlated thermal SZ emission, which we model with an amplitude \( A_{\text{SZ}} \) multiplying a spectral template. See Pullen et al. (2018) for more details.

The bandpower covariance that appears in the likelihood is described in Pullen et al. (2018) and includes \( \ell-\ell' \) correlations from masked sky regions.

### 5.5. CIB and \([\text{C}\,\text{II}]\) Parameter Estimates

We perform our Markov chain Monte Carlo (MCMC) analysis using CosmoMC (Lewis & Bridle 2002) for the six parameters

\[
\Xi \equiv \{ T_d, \alpha, L_0, A_{\text{C\,\text{II}}}, A_{\text{SZ}}, b_{\text{QSO}} \}.
\]

As in Pullen et al. (2018), we fit the parameters using the six cross-power spectra between the three high-frequency Planck bands and the quasars and LRGs presented in Figure 10; these are measured in nine bins over \( 100 < \ell < 1000 \). We label this fit the “Pullen + 2018” fit. We include the full covariance matrix between all cross-power spectra. Also as before, we fit the mean levels of the CIB emission in the three Planck bands (Béthermin et al. 2012), as well as 10 SFR density measurements (Madau & Dickinson 2014).

The Milky Way templates show some evidence of correlation with the quasar overdensity. This could bias the determination of the correlated \([\text{C}\,\text{II}]\) amplitude, so we additionally marginalize over a nuisance clustering anisotropy based on the CIB \( \times \) quasar clustering model with amplitude \( \alpha \).

Figure 12 shows the parameter fits before and after including Milky Way emission templates. The posteriors for the CIB parameters and thermal SZ amplitude do not change significantly from those of Pullen et al. (2018). The posterior distribution of \( A_{\text{C\,\text{II}}} \) is only modestly impacted by marginalizing over the Lenz et al. (2017) template. We also find that the \( A_{\text{C\,\text{II}}} \) measurement is not degenerate with \( \alpha \) for either Milky Way template. The MCMC constrains \(-0.092 < \alpha_{\text{Lenz}} < -0.025 \) and \(-0.121 < \alpha_{\text{Green}} < -0.042 \) at 95% confidence. This indication of a correlation between the Milky Way–only tracers and the quasar overdensity warrants future investigation. Synchrotron emission from the quasars could correlate with the anisotropy of the synchrotron background in the 21 cm maps.

### 5.6. Interpreting the MCMC Results

We can interpret the lack of impact of Milky Way templates in Planck \( \times \) BOSS quasar measurements using the multiband formalism in Section 4. To do this, recast the relevant parts of the full nonlinear model as linear modes that fit the correlated continuum emission as in Equation (14). Let the data vector be the cross-powers of the Planck 353, 545, and 857 GHz maps and BOSS quasars. Take a simple model for the SED of the

![Figure 11. Angular cross-power spectra between the Milky Way templates from Green et al. (2015) and Lenz et al. (2017) for both the quasars and LRGs from the BOSS survey. The Lenz et al. (2017) map is an HI column density map scaled to the dust emission in the 545 GHz Planck band. The Green et al. (2015) map is an extinction map multiplied by \( 10^6 \).](attachment:figure11.png)
correlated continuum with \( L_{\nu}^{\text{dust}} = \nu^\beta B(T_{\text{dust}}, (1 + z)\nu) \), where \( B(T_{\text{dust}}, \nu) \) is the Planck law, and we fix \( \beta = 1.5 \). To build a linear model, let \( M \) have a mode that is the best-fit dust with \( T_{\text{dust}} = 27.2 \) K, evaluated at \( z = 2.25 \) (near the peak of the quasar number density). To model changes in temperature, also add a mode \( d\nu T_{\text{dust}} \).

For the covariance in Equation (14), use the bandpower covariance of the Planck data. Figure 13 shows this principal mode of the bandpower covariance, which is consistent with emission from the Milky Way \( (\beta = 1.5, T_{\text{dust}} = 19 \) K from \( z = 0 \)). The principal mode is spectrally similar to emission with \( T_{\text{dust}} = 27.2 \) K from \( z = 2.25 \) in the correlated continuum model. The similarity in spectral indices means that the operation that marginalizes over correlated continuum emission is also very effective at suppressing the Milky Way emission. In the setting of Equation (16), the correlated and uncorrelated continuum terms have similar SEDs, so \( f(\nu) \approx g(\nu) \) and the second term of the variance are highly suppressed. The value of a Milky Way template map is diminished in this case.

In the linear model, the Lenz et al. (2017) template increases the signal-to-noise ratio in the line amplitude estimate by only 8%. Adding a derivative mode \( dT_{\text{dust}}/dT \) (as a linearized temperature parameter) reduces the improvement of the Lenz et al. (2017) template to 3%. Most of this improvement comes from \( \ell < 500 \), where the Milky Way template is best correlated to Planck 545 GHz (Figure 14). In contrast, if the extragalactic dust temperature were the same as the Milky Way’s \( \approx 19 \) K (but redshifted to \( z = 2.25 \)), the spectral match between the correlated SED modes and the galactic dust would not be nearly as good. For a 19 K emitter at \( z = 2.25 \), the dust spectrum crests in the 545 GHz band, while at \( z = 0 \), the dust emission in 857 GHz is brighter. In this scenario, the Lenz et al. (2017) template would lead to a 27% improvement in the line constraint. Hence, the Planck \( \times \) BOSS quasar measurements benefit from a lucky coincidence between the galactic and redshifted extragalactic emission SED. This similarity allows the marginalization over the extragalactic SED to also effectively deweight the Milky Way emission.
6. A True Tomographic Survey

Section 4 argues that an intensity survey needs at least as many frequencies as dof in the correlated continuum mode and argued that suppression of continuum contamination is equivalent to throwing out $k_\parallel$ (line-of-sight wavevector) information. This section describes how correlated continuum emission appears in the full 3D power spectrum $P_s(k_\perp, k_\parallel)$ of tomographic surveys with numerous spectral bins.

Here we will mock up a future tomographic survey to show the correlated continuum and line emission in the full 3D power spectrum $P_s(k_\perp, k_\parallel)$ for wavenumbers perpendicular, $k_\perp$, and parallel, $k_\parallel$, to the line of sight. Correlated continuum emission enters at low $k_\parallel$. Cutting $k_\parallel < k_\parallel^\text{cut}$ recovers the line correlation independently of the continuum but results in some loss of sensitivity at low $|k|$. 

6.1. Impact of Correlated Continuum Emission in the 3D Cross-power

Before doing a numerical simulation, we can get intuition for the impact of the SED on the IM cross-power by working in the Hubble approximation. Here the SED is a direct convolution of the overdensity field in the line-of-sight direction, which becomes a multiplication in $k_\parallel$ space.

The surface brightness at frequency $\nu$ is the integral of comoving specific emission intensity $j(\nu, z)$ across all redshifts in a tomographic survey slice,

$$I_\nu = \int \frac{d\chi}{1 + z} \frac{dz}{d\chi} j(\nu, z),$$

(21)

where $\chi(z)$ is the comoving distance. For a thin tomographic redshift slice $\delta z_\parallel$, the brightness of the line emission will not depend on $\delta z_\parallel$. In contrast, the correlated component of the continuum emission will be the integral in Equation (21) over the redshift slice, so it will depend linearly on the redshift width $\delta z_\parallel$.

Let the continuum SED at frequency $\nu$ from galaxies at redshift $z$ be $\Theta(\nu(1 + z))$ and take a graybody dust (Blain et al. 2002). Evaluate the SED on the comoving grid points and find its discrete Fourier transform $\tilde{\Theta}(k_\parallel)$, normalized such that $\tilde{\Theta}(k_\parallel = 0) = 1$. If the overdensity in 3D $k$-space is $\delta$, then the convolution with the full line plus continuum SED is

$$\delta_\nu^{BM} = S_L (1 + \Delta f_\text{cont} \tilde{\Theta}(k_\parallel)) \delta,$$

(22)

where $\Delta z$ is the total redshift range, or $N_{\text{chan}} \delta z$ for $N_{\text{chan}}$ channels, and $f_\text{cont}$ is the fraction of continuum to line intensity from galaxies per unit redshift. Relating this to the Planck–BOSS cross-correlation model in Section 5.4, $f_\text{cont} = \frac{25}{A_{CIB}}$. Here $S_L$ is the line brightness times the cosmological linear bias $I_{CIB} b_{CIB}$, and we assume that the dust emission shares the same bias for simplicity.

The cross-power $\delta_\nu^{BM} \delta_\nu^{BM}$ binned onto $k_\parallel$ and $k_\parallel$ is then

$$P_s(k_\perp, k_\parallel) = S_L b_{CIB} (1 + \Delta f_\text{cont} \tilde{\Theta}(k_\parallel)) P_s(k_\perp, k_\parallel),$$

(23)

where $P_s(k_\perp, k_\parallel)$ is the power spectrum of the underlying dark matter field $\delta$. This is just the expected IM cross-power $P_s(k_\perp, k_\parallel) = S_L b_{CIB} P_s(k_\perp, k_\parallel)$ plus a term that multiplies the $k_\parallel$ direction. In the case where galaxies have a constant SED over this spectral range, $\Theta(k_\parallel)$ is a delta function, and the continuum only appears at $k_\parallel = 0$. In practice, continuum emission will contaminate higher $k_\parallel$ because the SED has additional spectral structure, the Hubble approximation will break down, and there may be mixing between $k_\parallel$ modes because of the survey geometry.

We simulate the effect of correlated continuum emission on the cross-power spectrum by drawing Gaussian sample maps from a $z = 2.25$ power spectrum (Di Dio et al. 2013). The maps are 500 Mpc/$h$ cubes with a pixel size of 1 Mpc/$h$, which corresponds to $2 < z < 2.5$ in the line-of-sight direction. The continuum SED is then simulated by adding a graybody dust SED whose magnitude is proportional to the
overdensity at each voxel and redshifted according to that voxel’s redshift. To illustrate the sensitivity of a future survey, we take $A_{CII}$ a factor of 20 smaller than the posterior in Section 5.5. Since our focus is only on the ratio of correlated continuum to line emission, we do not apply the multiplicative prefactor $S_r b_d$ (Equation (23)) to our simulated intensity maps. Therefore, the units of the power spectra in Figures 15 and 16 are $Mpc^3 h^{-3}$. Figure 15 shows the two-dimensional power spectrum of one realization of this simulated model. (The impact of the continuum convolution is partly stochastic, so averaging over many simulations also averages down part of the correlated continuum.)

Only the two lowest nonzero $k_\parallel$ modes are noticeably biased by the continuum SED. This is shown more clearly in Figure 16, in which we plot the power spectrum averaged over $k_\perp$ as a function of $k_\parallel$. Figure 17 shows the decrease in the number of modes for $|k|$ by cutting the two lowest nonzero $k_\parallel$ modes from the binned one-dimensional power spectrum.

We can also revisit the question from Section 4.1 about the number of bands that a survey needs to treat correlated continuum emission without significant signal loss. In the tomographic case, the number of $k_\parallel$ modes corresponds roughly to the number of frequency channels in the spectrometer. Removing the five lowest $k_\parallel$ bins rejects the majority of the correlated continuum. Schematically, if there are a total of 500 channels, this results in a $5/500 = 1\%$ loss in signal. Conversely, if there are 10 channels, and five $k_\parallel$ modes must be removed, 50% of the signal is lost. The calculations here have made simple assumptions for the number of dof in the correlated continuum SED. In practice, both the correlated and uncorrelated continua may be more complex through their astrophysics or interaction with the instrument.

7. Discussion

Cross-correlation between IM data sets and galaxy redshift surveys has become a gold standard for robustness against foreground contamination. This cross-correlation is sensitive to not only the line emission in target galaxies but also the complete, correlated SED. The inference of line brightness needs to marginalize over nonlinear contributions in the SED. The influence of correlated continuum emission has only recently been appreciated (Serra et al. 2014; Switzer 2017; Chiang et al. 2018; Pullen et al. 2018). Here we have expanded the consideration to a more general survey to understand the circumstances when correlated continuum emission can be a significant pitfall for IM.

Correlated continuum emission should be estimated jointly (self-consistently) with bright Milky Way foreground cleaning. For most IM surveys, a synchrotron or dust continuum dominates the correlated SED. This SED is similar but not identical to the bright continuum emission from the Milky Way. We show that a properly formed likelihood will deweight foregrounds in $C_\ell$ space analogously to map-space cleaning through a linear combination. Doing the deweighting as part of the parameter estimation from the power spectra self-consistently treats the uncorrelated and correlated continuum emission. In contrast, an analysis that cleans in map space
before the likelihood analysis requires simulations to quantify residual correlated continuum emission and signal loss.

The IM observations with a limited number of bands have unique challenges that we consider in the case of Planck × BOSS (Pullen et al. 2018). When the number of channels is approximately the number of correlated continuum parameters, there is limited latitude to (1) deweight the bright uncorrelated foregrounds while also marginalizing over the correlated continuum and (2) verify that the line amplitude estimate is not biased because of an incomplete continuum model.

We reconsider the analysis of Planck × BOSS with the addition of Milky Way–only templates. These templates have the potential to provide leverage to separate and clean Milky Way emission independently of the correlated dust continuum. However, in practice, we find that existing templates (Green et al. 2015; Lenz et al. 2017) do not yet provide significant benefit, due both to the stochasticity with which they trace the Milky Way and to a similarity in the correlated and Milky Way dust continuum spectra.

Our analysis of Planck × BOSS will have shortcomings for future IM data, which move from simple detection to astrophysical characterization. Rather than taking a parametric model for the clustered emission, future work could directly solve for the SED(z, ν) and anisotropy structure Cℓ of the correlation, similar to a “CMB-only” approach (Dunkley et al. 2013). Direct inference of the clustering anisotropy could measure strong scale-dependent bias and stochasticity of the tracers that may not be understood well in a simple parametric model. Proposed IM experiments (Kovetz et al. 2017) have numerous frequency channels and should permit a unique measurement of the clustering of line and continuum radiation as a function of redshift.

We acknowledge David Spergel for recommendations for cleaning correlated foregrounds. This work was supported under NASA 16-ADAP16-0033, and is based on observations obtained with Planck (http://www.esa.int/Planck), an ESA science mission with instruments and contributions directly funded by ESA Member States, NASA, and Canada. Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is http://www.sdss3.org/.

SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

Appendix A The Multitracer Approach

Recent work (Alonso & Ferreira 2015; Bull et al. 2015; Fonseca et al. 2015; Kovetz et al. 2017; Switzer 2017) argues that the multitracer approach (McDonald & Seljak 2009; Seljak 2009; Bernstein & Cai 2011) can apply fruitfully to IM and allow a determination of line amplitudes without cosmic variance. This appendix describes this approach in map and power-spectrum space (Abramo & Leonard 2013; Blake et al. 2013) without foregrounds.

Following the definitions of Section 3, a model of the observed galaxy redshift survey and intensity map, respectively, is xg = bgδ + ng and xℓ = bℓδ + nℓ. We will start with the more standard derivation of the Fisher matrix in map space (e.g., Bernstein & Cai 2011). Per ℓ mode, the covariance of the galaxy redshift survey xg,ℓ and intensity map xi,ℓ is

$$\Sigma_{gℓ} = \begin{pmatrix} b_{gℓ}^2 C_{gℓ}^{bb} + N_{g,ℓ} & b_{gℓ} S_{ℓ} C_{gℓ}^{bg} \\ b_{gℓ} S_{ℓ} C_{gℓ}^{bg} & S_{ℓ}^2 C_{gℓ}^{bb} + N_{i,ℓ} \end{pmatrix} . \quad (24)$$

The Fisher matrix element for the determination of the covariance parameters θi and θj given the observables {xg, xi} is then (Tegmark et al. 1997)

$$F_{θi,θj} = \frac{1}{2} \text{Tr} \left( \Sigma_{gℓ}^{-1} \frac{dSigma_{gℓ}}{dθ_i} \Sigma_{gℓ}^{-1} \frac{dSigma_{gℓ}}{dθ_j} \right) . \quad (25)$$

In this setup, the parameters to constrain are θ = {Sℓ, Cℓbb, Ni}. From the inverse of the Fisher matrix, the minimum variance of the determination of Sℓ for each multipole (suppressing the ℓ label) marginalized over Cℓbb and Ni yields

$$\sigma^2_{Sℓ} = \frac{1}{\nuℓ} \left( \frac{N_i}{C_{ℓ}^{bb}} + \frac{N_g S^2_L}{b^2_g} + \frac{N_i N_g}{(C_{ℓ}^{bb})^2 b^2_g} + 2 \frac{N_i^2}{(C_{ℓ}^{bb})^2} S^2_L \right) . \quad (26)$$

where νℓ is the number of modes in the measurement, which is roughly (2ℓ + 1)fsky for a 2D map on the fraction fsky of the sky. This uses no expansion in the approximation of small noise (e.g., Bernstein & Cai 2011).

Sample variance produces a term linear in Sℓ2, but here all factors of Sℓ appear with orders of Ng/Cℓbb. To simplify expressions throughout to provide some analytic insight, we will assume that galaxy survey shot noise is negligible, or Ng/Cℓbb ≪ 1 (taking no terms beyond order zero). In this case,

$$\sigma^2_{Sℓ} = \frac{N_i}{(\nuℓ C_{ℓ}^{bb})} .$$

This has a simple interpretation as the signal-to-noise ratio per mode, divided by the number of modes.

The standard discussion above aligns well with a likelihood analysis of pixel space. Pixel-space likelihoods are expensive in practical data analysis, requiring an Npix × Npix matrix inverse for Npix pixels. We instead prefer to formulate the analytic Fisher matrix in power-spectrum space, where the likelihood for parameters in Section 5 is performed. Working with a likelihood of two-point information also allows more flexibility in the parameter constraint. For example, a likelihood using Equation (24) has access to information from both the cross-power bgSℓCℓbb and the auto-power S2ℓCℓbb + Ni of the intensity survey. In practice, the auto-power may not be usable because of additive bias from residual foregrounds, but it can simply be excluded from the two-point likelihood.
Here the two-point functions are
\[ C^{H}_H = S_H^2 C^{BB}_H + N_H, \]
\[ C^{L}_L = b_g S_L C^{BB}_L, \]
\[ C^{BB}_L = b_g^2 C^{BB}_L + N_g. \]  
(27)

Let the observable be the vector \( \mu = \{ C^{H}_H, C^{L}_L, N_g \} \) and estimate the parameters \( \{ S_L, C^{BB}_L, N_H \} \). In this case, the Fisher matrix for determination of the mean parameters (Tegmark et al. 1997) is
\[ F_{0,0} = \frac{1}{2} \text{Tr} \left( \text{Cov}(\mu) \right) \left( \begin{array}{c} d \mu_1 \, d \mu_2 \\ d \mu_1 \, d \mu_2 \\ d \mu_1 \, d \mu_2 \end{array} \right). \]  
(28)

where \( \text{Cov}(\mu) \) is the covariance of the two-point functions in the observation vector \( \mu \). The matrix elements of this covariance matrix are given by Equation (6).

Marginalizing over \( \{ C^{BB}_L, N_H \} \) and ignoring galaxy redshift survey shot noise, one recovers the estimate of \( S_L \) with variance \( \sigma^2_{S_L} = N_H / (\nu_T C^{BB}_L) \), as above. Sample variance is absorbed in \( C^{BB}_L \) as the explanatory variable, allowing \( S_L \) to be determined independently. Indeed, when the observable is only the cross-power \( \mu = \{ C^{L}_L \} \), and the line amplitude \( S_L \) is the only parameter, the variance \( \sigma^2_{S_L} = (2S_L^2 + N_H / C^{BB}_L) / \nu_T \) has a term \( \propto N_H^2 \).

To simplify the analytics throughout and avoid use of the intensity auto-power, we reduce the observables to the cross-powers and the galaxy redshift survey auto-power, or \( \mu = \{ C^{L}_L, C^{BB}_L \} \) here. This provides two equations allowing a fit of two parameters \( \{ S_L, C^{BB}_L \} \), so it assumes that \( N_H \) is fixed and known. In this case, the variance on \( S_L \) marginalized over \( C^{BB}_L \) is again \( \sigma^2_{S_L} = N_H / (\nu_T C^{BB}_L) \). This setup is a starting point for Section 3.

Both approaches above assume that the galaxy redshift survey and line emission are perfect tracers of the same overdensity. In practice, there is some stochasticity \( r \) between the two tracers, yielding \( C^{BB}_L = r b_g S_L C^{BB}_L \). In either the map or power-spectrum space approach, and assuming negligible shot noise, the line amplitude constraint with stochasticity between the tracers becomes
\[ \sigma^2_{S_L} = \frac{N_H}{r^2 \nu_T C^{BB}_L} + \frac{S_L^2}{\nu_T} \left( \frac{1}{r^2} - 1 \right). \]  
(29)

Note that this does contain cosmic variance terms \( \propto S_L^2 \) when \( r \neq 1 \). When there is stochasticity between the populations, the component of variance that is not correlated between the tracers adds cosmic variance. Furthermore, the thermal noise term \( N_H / (r^2 \nu_T C^{BB}_L) \) scales as \( 1/r^2 \). On the large angular scales considered in Section 5, the galaxy redshift survey and intensity trace the same underlying linear perturbations, where \( r \approx 1 \) is a reasonable approximation.

Evasion of cosmic variance in the line amplitude estimate can be understood in both map and power-spectrum space. In map space, the galaxy redshift survey provides a spatial template of overdensity whose amplitude can be fit in the intensity survey without cosmic variance (it is a determination of a mean amplitude rather than of variance). In power-spectrum space, cosmic variance is correlated between the intensity and galaxy redshift survey. Fitting jointly for \( C^{BB}_L \) and \( S_L \) makes \( C^{BB}_L \) the explanatory variable of the cosmic covariance. Also, \( S_L \) only appears as an amplitude in the intensity survey, so it is determined independently of the cosmic covariance.

**Appendix B**

**A Multiband Model with Separable Foregrounds**

Extend the model above to two intensity bands, \( A \) and \( B \), at frequencies \( \nu_A \) and \( \nu_B \). Let the amplitude of the correlated signal in maps \( A \) and \( B \) be \( \phi \) and \( \psi \). Add foreground covariance with spatial anisotropy \( C^{BB} \) and amplitude \( f(\nu) \) to both intensity bands. The two-point functions are
\[ C^{AA}_A = \phi^2 C^{BB}_L + f(\nu_A)^2 C^{uu}_A + N_A, \]
\[ C^{BB}_B = \psi^2 C^{BB}_L + f(\nu_B)^2 C^{uu}_B + N_B, \]
\[ C^{AB}_A = \phi \psi C^{BB}_L + f(\nu_A)f(\nu_B) C^{uu}, \]
\[ C^{BG}_B = \phi^2 f(\nu_B) C^{BB}_B, \]
\[ C^{BG}_B = \psi^2 f(\nu_A) C^{BB}_L. \]  
(30)

In the limit of negligible shot noise, the covariance of determinations \( \phi \) and \( \psi \) is
\[ \text{Cov}(\phi, \psi) = f(\nu_B)^2 C^{uu}_B + N_B, \]
\[ \text{Cov}(\phi, \psi) = f(\nu_A)f(\nu_B) C^{uu}, \]
\[ \text{Cov}(\phi, \psi) = f(\nu_A)^2 C^{uu}_A + N_A. \]  
(31)

Extending this to many bands gives covariance the form
\[ N = \frac{1}{\nu_T C^{BB}_L} (N_{nh} + C^{uu} f f^T), \]  
(32)

where \( N_{nh} \) is the diagonal matrix of thermal noise in each band (\( N_A \) and \( N_B \) above) and \( f = f(\nu) \) is the foreground spectrum evaluated in each of the bands. In summary, if the Milky Way adds common spatial structure to the intensity maps, it will appear as a rank-1 term in the bandpower covariance as a function of frequency.

**Appendix C**

**Multiband SED Reconstruction**

Here we develop a simple linear model for an intensity survey with many channels. Let the multiband IM survey have frequencies labeled \( \nu_i \) and \( \nu_B \). Let the amplitude of the correlated signal in maps \( A \) and \( B \) be \( \phi \) and \( \psi \). Add foreground covariance with spatial anisotropy \( C^{BB} \) and amplitude \( f(\nu) \) to both intensity bands. The two-point functions are
\[ C^{AA}_A = \phi^2 C^{BB}_L + f(\nu_A)^2 C^{uu}_A + N_A, \]
\[ C^{BB}_B = \psi^2 C^{BB}_L + f(\nu_B)^2 C^{uu}_B + N_B, \]
\[ C^{AB}_A = \phi \psi C^{BB}_L + f(\nu_A)f(\nu_B) C^{uu}, \]
\[ C^{BG}_B = \phi^2 f(\nu_B) C^{BB}_B, \]
\[ C^{BG}_B = \psi^2 f(\nu_A) C^{BB}_L. \]  
(33)

In the case of Planck, quasars in \( 2 < z < 3.2 \) only correlate with one frequency channel, so \( \xi(\nu_i) = 1 \) in \( \nu_i = 545 \) GHz and \( \xi(\nu) = 0 \) at the other Planck frequencies. In a more typical intensity survey with narrow spectral channels, the clustering of LSS at low \( k \) produces correlations between channels in the kernel \( \xi(\nu) \) (Challinor & Lewis 2011; Di Dio et al. 2013).

In the illustrative limit of negligible shot noise, the galaxy redshift survey auto-power fixes \( C^{BB} \), including cosmic variance, making the model for the cross-powers in
Appendix D

Geometry of the Joint Correlated and Uncorrelated Continuum Estimation

We can now consider the effect of Milky Way foregrounds by letting the covariance $N^{-1}_u$ be the rank-1 foreground model Equation (32), equivalent to a spatial pattern common across all frequencies. In practice, Milky Way foregrounds will have higher rank, which can be included in $N^{-1}_u$. In this simplified setting, we can continue analytically by applying the Sherman–Morrison formula twice to give

$$\Sigma_\theta = \frac{1}{\nu \ell \ell} \left( \Sigma_{th} + C^{uu}_\ell \alpha_f \alpha_f^T \right),$$

where

$$\Delta \equiv f^T N^{-1}_u (f - M \alpha_f),$$
$$\Sigma_{th} = (M^T N^{-1}_u M)^{-1},$$
$$\alpha_f \equiv \Sigma_{th} M^T N^{-1}_u f.$$ (40)

Here $\Sigma_{th}$ is the variance for the estimate of $\theta$ with only thermal noise and no foregrounds, and $\alpha_f$ is the optimal linear estimator, with thermal noise only, applied to the foreground SED. Hence, $C^{uu}_\ell \alpha_f \alpha_f^T$ is the amplitude of the foregrounds in the $\theta$ parameter basis. Finally, $\Delta$ is a discriminant that describes how much information about the uncorrelated foregrounds is left after fitting for the correlated spectral components in $M$.

We can now consider the geometry of the joint correlated and uncorrelated continuum separation problem. The key quantity here is the discriminant $\Delta$. This is a measure of the remaining rms of the uncorrelated foregrounds $f$ after fitting out the correlated continuum modes $g_i$. (To see this, note that $\alpha_f$ projects $f$ onto the parameter basis $\theta$, and then $M$ projects onto a spectrum differenced with $f$ in $\Delta$.)

If $M$ is a complete spectral basis, then $\Delta = 0$ regardless of the spectral shape of $f$. This will be the case any time the number of independent parameters in the correlated SED model is equal to the number of bands, such as in the case of Planck × BOSS. In this regime,

$$\Sigma_\theta = \frac{1}{\nu \ell \ell} \left( \Sigma_{th} + C^{uu}_\ell \alpha_f \alpha_f^T \right).$$

These two terms correspond to the thermal noise and foreground components in Equation (12) with $r_F = 1$. The second term represents foregrounds that contaminate the line amplitude estimate after marginalizing the correlated continuum spectral modes $g_i$. An observation with few bands may not provide sufficient dof to deweight uncorrelated continuum foregrounds and model the correlated continuum foreground.

Adding a template leads to a qualitative improvement in the foreground cleaning. When $N_{\text{band}} = N_{\text{param}}$, Equation (41) shows that the line amplitude variance always has a term proportional to $C^{uu}_\ell$. By adding a template, $C^{uu}_\ell \Delta$ in the denominator of Equation (39) can remain large. In the limit of large foregrounds,

$$\Sigma_\theta = \frac{1}{\nu \ell \ell} \left( \Sigma_{th} + \alpha_f^T \alpha_f \Delta \right).$$

This linear form recovers our primary results so far. Equation (38) recovers Equation (9) in the case where $M = [[1, 0]^T]$ (only $S_t$ is fit from the data) and $r_F = 1$. Here $N_{th} = C^{uu}_\ell [f(\nu_1), f(\nu_2)]^T [f(\nu_1), f(\nu_2)]$ and $N_{bg} = \text{diag}(N_{th, bg})$. Adding one correlated SED component as $M = [[1, 0]^T, [g(\nu_1), g(\nu_2)]^T]$ recovers Equation (12), again with $r_F = 1$. 

Form a new data vector $d^\ell_\ell$ from the measured cross-powers at frequency $\nu_i$ divided by $b_\ell C_\ell^{D\delta}(z)$ inferred from the galaxy redshift survey

$$d^\ell_\ell(\nu_i) = \hat{C}_\ell(\nu_i) x_s(z)/b_\ell C_\ell^{D\delta}(z).$$ (34)

This is an estimator for $S_\ell(\xi(\nu_i) + S_\ell(\nu_i))$. Model the SED of the correlated continuum as the sum of linear components $g_i(\nu_i)$ and amplitudes $S_{C,j}$,

$$S_C(\nu_i) = \sum_j S_{C,j} g_j(\nu_i).$$ (35)

Pack each spectral component into a vector $g_i | = g_i(\nu_i)$ and the signal correlation kernel into a vector $\xi | = \xi(\nu_i)$. In the case of widely separated bins in the Planck analysis, this is a delta function at 545 GHz. Pack the line signal and $N_{\text{comp}}$ correlated continuum modes into $M = [\xi, g_1, ..., g_{N_{\text{comp}}}^\ell]$. The linear amplitude parameter vector is then $\theta = [S_\ell, S_C]$, where $S_C$ are the $S_{C,j}$, linear parameters for the SED. With these substitutions, the model for multiband cross-powers in Equation (33) is

$$d^\ell_\ell = M \theta + n_s,$$ (36)

where $n_s$ is the noise of the cross-power measurements drawn from covariance $N_s$ that includes both thermal noise ($N_{th}$) and foreground ($N_{bg}$) covariance as

$$N_s = \frac{1}{\nu \ell \ell} (N_{th} + N_{bg}).$$ (37)

(Note that the data vector defined in Equation (34) is an estimator for the amplitudes for $\phi$ and $\psi$ in Appendix B but applied to many bands. This results in the same form for the covariance as Equation (32)).

This model solves jointly for the line amplitude $S_\ell$ and spectral baseline amplitudes $S_C$ that describe the correlated continuum emission (Switzer 2017). The linear estimate $\hat{\theta}$ and its covariance $\Sigma_\theta$ are

$$\hat{\theta} = \Sigma_\theta M^T N^{-1} d^\ell_\ell,$$
$$\Sigma_\theta = (M^T N^{-1}_u M)^{-1}.$$ (38)

We have written this expression for a single multipole $\ell$ for simplicity, but the model could be extended as a likelihood on the cross-powers at all $\ell$ in practice.

This linear form recovers our primary results so far. Equation (38) recovers Equation (9) in the case where $M = [[1, 0]^T]$ (only $S_t$ is fit from the data) and $r_F = 1$. Here $N_{th} = C^{uu}_\ell [f(\nu_1), f(\nu_2)]^T [f(\nu_1), f(\nu_2)]$ and $N_{bg} = \text{diag}(N_{th, bg})$. Adding one correlated SED component as $M = [[1, 0]^T, [g(\nu_1), g(\nu_2)]^T]$ recovers Equation (12), again with $r_F = 1$. 

Equation (33) approximately linear in the amplitude $[S_\ell(\xi(\nu_i) + S_\ell(\nu_i))]$. We can therefore use a simple analytic model of the cross-power as a function of frequency to understand the joint constraint of correlated line and continuum emission. In practice with BOSS quasars, shot noise is not negligible, but the linear approach gives intuition for the geometry of the joint analysis of correlated and uncorrelated continua. Future IM experiments would be complemented well by galaxy redshift surveys with small shot noise on the scales of interest.
Now the foreground brightness $C_{\ell \nu}^{uu}$ drops out, and the numerator $\alpha_r \alpha_r^T$ is suppressed by $\Delta$. Again in this limit, the variance of the $S_r$ estimate is not compromised by bright foregrounds, making this a promising approach. This is the extension of Equation (16) to multiple bands.

**Appendix E**

The Impact of Stochasticity between Milky Way Tracers and Dust Emission

Extend the template model of Equation (15) to allow for imperfect correlation $r_M = C_{\ell \nu}^{Ma}/\sqrt{C_{\ell \nu}^{MM} C_{\ell \nu}^{uu}}$ with the actual foreground $x_\nu$ in the intensity map. The correlations between this map and itself (MM), the intensity map (MI), and veto map (MV) are

$$C_{\ell}^{MM} = A_M^2 C_{\ell \nu}^{uu} + N_M,$$

$$C_{\ell}^{MI} = r_M A_M f(\nu_1) C_{\ell \nu}^{uu},$$

$$C_{\ell}^{MV} = r_M A_M f(\nu_2) C_{\ell \nu}^{uu}.$$  

For simplicity, let $N_M \to 0$, so the template has no noise but is still an imperfect tracer of the true foreground. Then,

$$\sigma_{x_\ell}^2 = \frac{C_{\ell \nu}^{uu}}{v_I C_{\ell}^{NN}} \left[ f(\nu_1) - f(\nu_2) \frac{g(\nu_1)}{g(\nu_2)} \right]^2 (1 - r_M^2) + \frac{1}{v_I C_{\ell}^{NN}} N_I + \frac{g(\nu_1)}{g(\nu_2)} N_v.$$  

**Appendix F**

Redshift-space Distortions on the Limber-approximated Angular Power Spectrum

In this paper, we argue that redshift-space distortions (RSD) can be neglected for the bulk of our analysis. To prove this, we first derive the angular power spectrum, which is directly related to the CIB-LSS angular cross-power spectrum we use in our analysis, under the Limber approximation (Loverde & Afshordi 2008). We begin with the full expression for the angular power spectrum in redshift space given by $C_{\ell} = C_{\ell}^{00} + 2\beta C_{\ell}^{0r} + \beta^2 C_{\ell}^{rr}$, where $\beta$ is the RSD parameter given by the growth rate-clustering bias ratio,

$$C_{\ell}^{00} = \frac{2}{\pi} \int dk k^2 P(k) |W^0_{\ell}(k)|^2,$$

$$C_{\ell}^{0r} = \frac{2}{\pi} \int dk k^2 P(k) W^0_{\ell}(k) W^r_{\ell}(k),$$

$$C_{\ell}^{rr} = \frac{2}{\pi} \int dk k^2 P(k) |W^r_{\ell}(k)|^2,$$  

where $P(k)$ is the matter power spectrum at redshift $z = 0$ and the window function $W_{\ell}(k) = W^0_{\ell}(k) + W^r_{\ell}(k)$. Given the selection function $\phi(r)$, we can write (Padmanabhan et al. 2007)

$$W^0_{\ell}(k) = \int dr \phi(r) j_0(k r)$$

$$W^r_{\ell}(k) = \int dr \phi(r) \left[ \frac{2\ell^2 + 2\ell - 1}{(2\ell - 1)(2\ell + 3)} j_0(k r) \right. - \ell(\ell - 1) \frac{j_{\ell - 2}(k r)}{(2\ell - 1)(2\ell + 1)} - \ell(\ell + 2) \left. \frac{j_{\ell + 2}(k r)}{(2\ell + 1)(2\ell + 3)} \right].$$  

(48)

We wish to perform operations similar to those in Loverde & Afshordi (2008) to get the Limber approximation for this expression. Using Equations (7)–(11) in Loverde & Afshordi (2008), we can show that the Limber approximation at first order is

$$\sqrt{\frac{2}{\pi}} \int dr \phi(r) j_\ell(k r) \approx \frac{1}{k \sqrt{\ell + 1/2}} \phi(\frac{\ell + 1/2}{k})$$

$$\approx \frac{1}{k \sqrt{\ell}} \phi(\frac{\ell}{k}),$$  

(49)

where we approximate $\ell + 1/2 \approx \ell$ to make the following equations less cluttered. Inserting this into Equation (47), we can write

$$C_{\ell}^{00} = \frac{1}{\ell} \int dk \phi^2 \left( \frac{\ell}{k} \right) P(k),$$

$$C_{\ell}^{0r} = \frac{1}{\sqrt{\ell}} \int dk \phi \left( \frac{\ell}{k} \right) F(\ell, k) P(k),$$

$$C_{\ell}^{rr} = \int dk F^2(\ell, k) P(k),$$  

(50)

where

$$F(\ell, k) = \frac{2\ell^2 + 2\ell - 1}{(2\ell - 1)(2\ell + 3)} \frac{\phi(\ell/k)}{\sqrt{\ell}} - \frac{\ell(\ell - 1)}{(2\ell - 1)(2\ell + 1)} \frac{\phi([\ell - 2]/k)}{\sqrt{\ell - 2}} - \frac{(\ell + 1)(\ell + 2)}{(2\ell + 1)(2\ell + 3)} \frac{\phi([\ell + 2]/k)}{\sqrt{\ell + 2}}.$$  

(51)

Now we wish to use this expression to show when RSD can be neglected. In this argument, we will assume $\phi(r)$ is a top-hat distribution centered at $r = R$ with full width $\Delta R$. In this case, it is evident that if $\ell/k - R \ll \Delta R$, then $\phi(\ell/k) \approx \phi(\ell/k)$. Additionally, assuming $\ell$ is large enough such that $\sqrt{\ell + 1/2} \approx \sqrt{\ell}$, we can show that in this case, $F(\ell, k) \approx 1/(4\ell^{5/2})$, which decreases rapidly to zero, eliminating the RSD effect. Taking the approximation that $\ell/k \sim R$, we argue that all the $\phi$s are nonzero when $2/k < \Delta R/2$, which implies that the RSD effect vanishes when $\ell \gtrsim 4R/\Delta R$. For the quasars, $4R/\Delta R \sim 17$, while for the LRGs, $4R/\Delta R \sim 10$. Thus, we are justified in neglecting RSD for our analysis.

**References**

Abramo, L. R., & Leonard, K. E. 2013, MNRAS, 432, 318

Aihara, H., Allende Prieto, C., An, D., et al. 2011, ApJS, 193, 29

Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, ApJS, 219, 12

Alam, S., Ata, M., Bailey, S., et al. 2017, MNRAS, 470, 2617

Alonso, D., & Ferreira, P. G. 2015, PhRvD, 92, 063525

Anderson, C. J., Luciw, N. I., Li, Y.-C., et al. 2018, MNRAS, 476, 3382
