FOREGROUND AND SENSITIVITY ANALYSIS FOR BROAD BAND (2D) 21 cm–Lyα AND 21 cm–Hα CORRELATION EXPERIMENTS PROBING THE EPOCH OF REIONIZATION

Abraham R. Neben[1], Brian Stalder[2], Jacqueline N. Hewitt[1], John L. Tonry[2]

[1]MIT Kavli Institute, Massachusetts Institute of Technology, Cambridge, MA 02139 and
[2]Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822
Draft version April 27, 2017

ABSTRACT

A detection of the predicted anticorrelation between 21 cm and either Lyα or Hα from the Epoch of Reionization (EOR) would be a powerful probe of the astrophysics of the EOR. While 3D intensity maps isolate foregrounds in the lowest $k_\parallel$ modes, infrared surveys cannot yet match the field of view and redshift resolution of radio intensity mapping experiments. In contrast, 2D (i.e., broad band) infrared intensity maps can be measured with current experiments and are not sensitivity limited. We find that 2D experiments can measure of order 100% of the 3D fluctuation power at $k < 0.2\,\mathrm{Mpc}^{-1}$ while preserving the correlation properties. However, we show that foregrounds pose two challenges: (1) simple geometric effects result in percent-level correlations between fluxes in these two bands, even if their luminosities are uncorrelated; and (2) the largely uncorrelated radio and infrared foreground residuals contribute a sample variance noise to the cross spectrum. The first challenge demands better foreground masking and subtraction, while the second requires measurements over large fields of view with many independent samples to average down uncorrelated radio and infrared power. Using radio and infrared observations from the Murchison Widefield Array (MWA) and the Asteroid Terrestrial-impact Last Alert System (ATLAS) we set an upper limit on the 21 cm–Lyα power spectrum at $z \sim 7$ of $\Delta^2 < 181\,(\mathrm{kJy/sr \cdot mK})$ (95%) at $\ell \sim 800$. We predict the levels of foreground flux correlation and the sample variance noise in several possible future experiments, and show that higher resolution radio and infrared surveys such as LOFAR, SKA-LOW, Hubble, Dark Energy Survey can start to probe optimistic models of the 21 cm–Lyα EOR cross spectrum.

Subject headings: cosmology: observations — dark ages, reionization, first stars — infrared: diffuse background

1. INTRODUCTION

Deep radio and infrared observations are nearing detection of the first stars and galaxies from the cosmic dawn. As such sources form, they are thought to blow out ionized bubbles, eventually merging and reionizing the universe. See Furlanetto et al. (2006); Morales & Wyithe (2010); Pritchard & Loeb (2012); Mesinger (2016) for reviews. First generation 21 cm observatories such the Murchison Widefield Array (MWA) (Tingay et al. 2013; Bowman et al. 2013) and the Precision Array for Probing the Epoch of Reionization (PAPER) (Parsons et al. 2014; Ali et al. 2015; Pober et al. 2015; Jacobs et al. 2015) are setting ever tighter limits on redshifted neutral hydrogen emission from the neutral regions between these bubbles, and the now-underway Hydrogen Epoch of Reionization Array (HERA) (DeBoer et al. 2017; Neben et al. 2016; Ewall-Wice et al. 2016; Patra et al. 2017) is expected to detect and characterize the EOR power spectrum in the coming years. Similar efforts are underway by the Low Frequency Array (Yatawatta et al. 2013; van Haarlem et al. 2013). Ultimately, the Square Kilometer Array (SKA) (Hall 2005; Dewdney & Lazio 2008; Dewdney et al. 2009; bij de Vaate et al. 2011) will image the EOR over redshift, revealing the detailed hydrogen reionization history of the universe.

At the same time, new galaxy surveys are beginning to constrain the reionizing sources themselves. Deep galaxy surveys (Bowler et al. 2017; Roberts-Borsani et al. 2016; Bowler et al. 2015; Wilkins et al. 2016; Bouwens et al. 2015, 2014; Dunlop et al. 2013; Illingworth et al. 2013; Ellis et al. 2013; Robertson et al. 2013; Oesch et al. 2013; Grogin et al. 2011; Bouwens et al. 2011) and cluster lensing surveys (Livermore et al. 2017; Repp et al. 2016; McLeod et al. 2016; Bouwens et al. 2016a,b; Coe et al. 2015; Huang et al. 2016; McLeod et al. 2015; Atek et al. 2015) are finding hundreds of galaxy candidates at $6 < z < 10$ down to UV magnitudes of $M_{AB} \sim -17$ (Finkelstein 2016), and extremely wide surveys are searching for the rare bright galaxies (Bernard et al. 2016; Calvi et al. 2016; Schmidt et al. 2014; Bradley et al. 2012; Trenti et al. 2011) from the reionization epoch. However, current models require the ionizing contribution of far fainter galaxies down to $M_{AB} \sim -13$ (Bouwens 2016; Alvarez et al. 2012) in order for reionization to be complete by the time that CMB optical depth measurements (Planck Collaboration et al. 2016) say it must be. Deeper observations with the James Webb Space Telescope (JWST) (Gardner et al. 2006), expected to probe down to $M_{AB} \sim -15.5$ in order hundred-hour integrations (Finkelstein 2016), will be needed to begin to probe this crucial faint population directly.

Infrared intensity mapping offers several advantages compared to galaxy surveys. Power spectrum analyses can be sensitive to an EOR component even if the signal-to-noise in individual pixels is small, and instead of being limited to the brightest galaxies, intensity mapping is sensitive to the cumulative light from *all* sources. The expected bright Lyα (e.g. Amorín et al. 2017) and Hα

(e.g. Smit et al. 2014) radiation from EOR galaxies at $z \sim 6 - 8$ is motivating intensity mapping at micron-scale wavelengths. Working around foregrounds is challenging, though. While early studies suggested angular fluctuations in infrared intensity maps traced EOR galaxies (e.g., Kashlinsky et al. 2005, 2007, 2012), Helgason et al. (2016) find that given current constraints on the EOR, this is unlikely. Intrahalo light Cooray et al. (2012); Zemcov et al. (2014) and Galactic dust Yue et al. (2016) have been proposed as more likely explanations for the larger than expected fluctuations, though Mitchell-Wynne et al. (2015) show that much of this excess can be removed with higher resolution measurements using Hubble. In any case, all these components are likely present at some level and degeneracies coupled with imperfect foreground knowledge make isolating the EOR contribution difficult.

For these reasons, cross correlation with 21 cm maps may in fact be the *only* way to extract the EOR component of the near infrared background. The synergy is clear: the galaxies sourcing reionization generate strong Lyα emission, while the neutral regions between them glow at rest frame 21 cm. On typical ionized bubble scales, bright spots in IR maps likely correspond to ionized regions, and thus, 21 cm dark spots, and vice versa. This effect is expected to source an anticorrelation, seen in simulations by Silva et al. (2013); Heneka et al. (2016) and modeled analytically by Feng et al. (2017); Mao (2014).

A similar large scale anticorrelation is found by Lidz et al. (2009); Park et al. (2014) in simulations of 21 cm cross correlation with galaxy redshift surveys. However, conducting redshift surveys both wide and deep enough to cross correlate with 21 cm maps is challenging due to the hugely different spatial scales probed by 21 cm experiments and spectroscopic galaxy surveys. For instance, the $\sim 3'$ angular resolution of the MWA is nearly equal to the field of view of the Hubble Deep Field and the James Webb Space Telescope (JWST). In a source-by-source manner, (Beardsley et al. 2015) show that this anticorrelation may be studied by inspecting the 21 cm brightness temperatures at the locations of JWST sources, but as discussed above source detections will be limited to the brightest and rarest objects.

Intensity mapping thus holds great promise to extend EOR science, and 3D intensity mapping (i.e., with a redshift dimension) have the advantage of avoiding the majority of continuum emission from intermediate redshift galaxies. This foreground emission is expected to contaminate only the lowest few line of sight fourier modes (Gong et al. 2017), while line interlopers at intermediate redshifts are easily masked (Gong et al. 2014, 2017; Pullen et al. 2014; Comaschi et al. 2016). 3D power spectra are also easier to understand theoretically as they quantify emission from a fundamentally 3D volume, and early demonstrations of this type of analysis are given by Chang et al. (2010); Masui et al. (2013) who detected the cross correlation between 21 cm emission and a galaxy redshift survey at $z \sim 1$. However, near infrared intensity mapping in 3D likely requires space-based observations to avoid atmospheric OH lines (e.g. Sullivan & Simcoe 2012), as well as fine spectral resolution to match the redshift resolution of typical 21 cm experiment.

For instance, Pober et al. (2014) show that with moder-

ate foreground assumptions, HERA should achieve $> 5\sigma$ detections of the 21 cm power spectrum over $0.2 < k_\parallel < 0.5$ at $z \sim 8$, corresponding to redshift scales of $0.03 < \Delta z < 0.1$. Resolving these same line of sight modes of the Lyα field at the same redshift requires a spectral resolution[1] of $80 < R < 250$. Achieving this high spectral resolution over a large enough field of view to match a 21 cm survey is challenging. The proposed SPHEREx mission (Doré et al. 2014, 2016) would image the entire sky in the near infrared with $R = 40$ spectroscopy for a cost of roughly $100M, and the concept Cosmic Dawn Intensity Mapper (Cooray et al. 2016) mission would achieve R=200–300 over 10 square degrees for nearly ten times the cost.

In contrast, 2D (i.e, broad band) intensity mapping enables similar science with far shallower and cheaper observations (Fernandez et al. 2014; Mao 2014), though here the main challenge is imperfect foreground removal. Even if radio and infrared foreground residuals are uncorrelated, they leak power into the cross correlation analysis which averages down only over sufficiently large fields of view. As OH emission from the atmosphere is relatively smooth over few degree scales (High et al. 2010), there is hope that these observations can be conducted from the ground for a further reduction in cost. A number of new ground-based wide field surveys surveys are coming online such as the Dark Energy Survey (Dark Energy Survey Collaboration et al. 2016), Pan-STARRS (Tonry et al. 2012), and the Asteroid Terrestrial-impact Last Alert System (ATLAS) (Tonry 2011). Further, the Transiting Exoplanet Survey Satellite (Ricker et al. 2014) will survey the entire sky over 600–1000 nm band, and the Wide-Field Infrared Survey Telescope (WFIRST) (Spergel et al. 2013) will observe instantaneous deep fields $200\times$ larger those of Hubble and JWST. It is important to note that a large, uniform focal plane greatly facilitates intensity mapping, lest structures on relevant angular scales be lost in the calibration of many independent regions of a segmented focal plane, such as that of Pan-STARRS.

In this paper we study the foregrounds in broad band 21 cm–Lyα and 21 cm–Hα intensity mapping correlation experiments targeting the EOR. We begin in Sec. 2 with a review of our fourier transform and power spectrum conventions. In Sec. 3 we present the MWA and ATLAS observations we use and discuss processing these data into images. In Sec. 4 we characterize the bright radio and infrared point source foregrounds and demonstrate that geometric effects introduce slight positive correlations which will overpower the cosmological signal unless significant masking and subtraction are conducted. In Sec. 5, we study how best to mask and subtract radio and infrared foregrounds in real world radio and infrared images and quantify the foreground residuals. We set the first limits on the broad band 21 cm–Lyα cross spectrum at $z = 7$ using data from the MWA and ATLAS, and predict the sensitivities of future experiments and compare them with the expected levels of geometric foreground correlation, illustrating what it will take to realize this measurement.

---

[1] The redshift resolution of a spectral line intensity mapping experiment observing emission at redshift z is given by $\Delta z = (1 + z)/R$, where R is the spectral resolving power.

2. POWER SPECTRUM AND CORRELATION CONVENTIONS

2.1. *Power spectrum definitions*

We define the 3D power spectrum $P(\vec{k})$ of the image cube $I(\vec{x})$ as

$$P(\vec{k}) = \frac{\langle |\tilde{I}(\vec{k})|^2 \rangle}{V} \tag{1}$$

where $\tilde{I}(\vec{k})$ is given by

$$\tilde{I}(\vec{k}) = dV \sum_{\vec{x}} I(\vec{x}) e^{i\vec{k}\cdot\vec{x}}, \tag{2}$$

Here V is the survey volume, and dV is the voxel size. Note that $P(\vec{k})$ has units of $[I]^2 \cdot \mathrm{Mpc}^3$, and that we often plot instead the more intuitive quantity $\Delta(\vec{k}) = [k^3 P(\vec{k})/2\pi^2]^{1/2}$.

Similarly, over narrow fields of view, the angular power spectrum $C(\vec{\ell})$ of a 2D (e.g, broad band) image $I(\vec{\theta})$ can be shown to be approximately

$$C(\vec{\ell}) = \frac{\langle |\tilde{I}(\vec{\ell})|^2 \rangle}{\Omega} \tag{3}$$

where $\tilde{I}(\vec{\ell})$ is given by

$$\tilde{I}(\vec{\ell}) = d\Omega \sum_{\vec{\theta}} I(\vec{\theta}) e^{i\vec{\ell}\cdot\vec{\theta}}, \tag{4}$$

Ω is the survey solid angle, and $d\Omega$ is the pixel size. Thus, over a narrow field of view, we need only evaluate a fourier transform to estimate the angular power spectrum. Writing this out in detail gives[2]

$$C(\ell(a,b)) = \left\langle \left| \sum_{m,n} I(m,n) \exp\left(-\frac{2\pi i}{N}(am+bn) \right) \right|^2 \right\rangle \frac{d\theta^2}{N^2} \tag{5}$$

where $d\theta = d\theta_x = d\theta_y$ is the pixel size, $N \equiv N_x = N_y$ is the number of pixels on a side of a square image, $\ell^2 = \ell_x^2 + \ell_y^2$, and

$$\ell_x = 2\pi a/N d\theta \tag{6}$$
$$\ell_y = 2\pi b/N d\theta \tag{7}$$

Note that $C(\vec{\ell})$ has the units of $[I]^2 \cdot d\theta^2$, and we often work with $\Delta(\vec{\ell}) = [\ell(\ell+1)C(\vec{\ell})/2\pi]^{1/2}$ which has the same units[3] as I.

The 3D 21 cm power spectrum is often cylindrically binned from 3D $\vec{k}$ space to 2D $(k_\perp, k_\parallel)$ space where $k_\perp^2 \equiv k_x^2 + k_y^2$ represents modes perpendicular to the line of sight, and $k_\parallel = k_z$ represents modes along the line of sight. We show in Appendix B that this cylindrically binned power spectrum is related to the angular power spectrum of a broad band image (over a narrow field of view) as

$$P(k_\perp, k_\parallel = 0) = D_c^2 \Delta D_c C(\ell(k_\perp)). \tag{8}$$

Here $\ell = D_c k_\perp$, where D_c is the comoving line of sight distance to the center of the cube, and ΔD_c is the comoving depth of the cube.

2.2. *Cross spectrum vs. coherence*

The 3D and 2D cross spectra are defined, extending Eqns. 1 and 3 to the cross spectrum as

$$P_{12}(\vec{k}) = \frac{\langle \tilde{I}_1^*(\vec{k}) \tilde{I}_2(\vec{k}) \rangle}{V} \tag{9}$$

$$C_{12}(\vec{\ell}) = \frac{\langle \tilde{I}_1^*(\vec{\ell}) \tilde{I}_2(\vec{\ell}) \rangle}{\Omega} \tag{10}$$

where 1 and 2 denote the 21 cm and the IR fields, respectively. The cross spectrum is a quantity which ranges between $\pm(C_1(\vec{\ell})C_2(\vec{\ell}))^{1/2}$ in the 2D case, depending on how correlated or uncorrelated the two fields are. It is thus often renormalized as

$$c_{12}(\vec{\ell}) \equiv \frac{C_{12}(\vec{\ell})}{\sqrt{C_1(\vec{\ell})C_2(\vec{\ell})}} \tag{11}$$

where c is known as the coherence and is insensitive to a simple rescaling of either field. However, uncorrelated foreground residuals in either field will substantially bias the coherence towards zero (Lidz et al. 2009; Furlanetto & Lidz 2007), whereas they merely contribute a zero mean noise to the cross spectrum. Slight foreground correlations, of course, will bias the cross spectrum as well, as we explore later.

3. OBSERVATIONS AND IMAGING

3.1. *21 cm Observations*

The MWA is a low frequency radio interferometer in Western Australia consisting of 128 phased array tiles, each with $\sim 30° \times (150\,\mathrm{MHz}/f)$ beams (full-width-at-half-maximum) and steerable in few degree increments with a delay line beamformer. We use low frequency observations of a quiet field centered at (RA,Dec)=$(0°,-27°)$ J2000 recorded over 30.72 MHz bandwidth centered at 186 MHz, corresponding to $z = 6.0 - 7.3$ for rest frame 21 cm.

The MWA observations are recorded as 2 min "snapshots" which are flagged for RFI using COTTER (Offringa et al. 2015), then calibrated and imaged using Fast Holographic Deconvolution[4] (Sullivan et al. 2012). Model visibilities are simulated from a foreground model of diffuse (Beardsley et al. 2016) and point source (Carroll et al. 2016) emission in the field, and used for both calibration and foreground subtraction. For each snapshot, FHD produces naturally weighted data and model image cubes as well as primary and synthesized beam cubes. FHD outputs these "cubes" in HEALPix format per frequency. Note that this processing is performed in parallel on "odd" and "even" data cubes whose data are interleaved at a 2 sec cadence for the purpose of eliminating the noise bias in power spectrum estimation.

---

[2] Note that the normalization of $d\theta^2/N^2$ has been misstated as $1/N^2$ by Zemcov et al. (2014) (Eqn. 3 of their supplement) and $d\theta^2$ by Cooray et al. (2012) (Eqn. 1 of their supplement).

[3] Over small fields of view, we must necessarily work at large ℓ, implying that $\ell(\ell+1) \approx \ell^2$, which has units of inverse steradians in view of Eqns. 6 and 7

[4] https://github.com/EoRImaging/FHD

We rotate these HEALPix maps so the MWA field center lies at the north pole, then project the pixels onto the xy plane to obtain naturally weighted image space cubes of the raw data (I_{nat}), the model data ($I_{nat,mod}$), the synthesized beam (I_w) (i.e., the fourier transform of the uv weights), and the primary beam, all in orthographic projection. We flag the upper and lower 80 kHz channels in each of 24 coarse channels across the band to mitigate aliasing, average each cube over frequency to make broad band images, then apply uniform weighting as in Dillon et al. (2015) using

$$I_{uni}(\vec{\theta}) = \frac{10^{-26}\lambda^2}{k_B} \sum_{\vec{u}} \frac{\tilde{I}_{nat}(\vec{u})}{\tilde{I}_w(\vec{u})} e^{-2\pi i \vec{\theta}\cdot\vec{u}} d^2 u \qquad (12)$$

where $\tilde{I}_i(\vec{u}) = \sum_{\vec{\theta}} I_i(\vec{\theta}) e^{2\pi i \vec{\theta}\cdot\vec{u}} d\Omega$ for $i = $ nat$, w$. The units of I_{nat} and I_w are most easily seen from their approximate definitions of $I_{nat}(\vec{\theta}) \approx \sum_j V(\vec{u}_j) e^{-2\pi i \vec{u}_j \cdot \vec{\theta}} du^2$ and $I_w(\vec{\theta}) \approx \sum_j e^{-2\pi i \vec{u}_j \cdot \vec{\theta}} du^2$, where the sums are over all measured visibilities $V(\vec{u}_j)$ in units of Jy. These definitions are only approximate because FHD performs corrections to account for w terms.
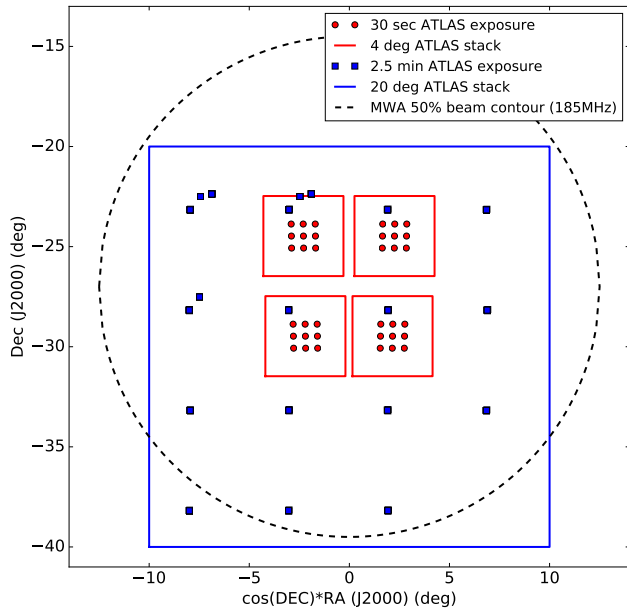
3.2. IR Observations

FIG. 1.— The MWA deep integration field (black dashed circle) is shown relative to our two ATLAS surveys. Blue markers show the observation centers of our wide ATLAS survey aimed at studying foregrounds, and the large blue square shows the stacked image. Red circle markers show observation centers for our slightly deeper survey, and red squares show the four stacked images we generate. Note that the ATLAS field of view is 5.5°.

ATLAS is a system of multiple 0.5 m f/2 wide-field telescopes (Tonry 2011) designed for near-earth asteroids detection and tracking. Two telescopes are currently in operation located on the islands of Maui and Hawaii. Each telescope has a single 10560×10560 STA1600 CCD sensor, with a pixel scale of 1.86″, giving a field of view of 5.5° on a side. We observe in the Johnson I band (810 nm center with 150 nm full-width-at-half-max) with

the Maui telescope, corresponding to $z = 5.1 - 6.3$ for Lyα. While this redshift range doesn't exactly match that of our radio observations, it overlaps sufficiently for our purpose of characterizing the noise and foregrounds in 21cm–Lyα cross correlation experiments. The automatic processing pipeline provides single sky-flattened (using a median of the nightly science exposures) images registered to nominal 2MASS astrometry.

We perform two separate observing campaigns, which we illustrate in Fig. 1. We first perform a wide survey to best characterize bright foregrounds. We raster scan a roughly $20° \times 20°$ grid with 5° spacing over the MWA field (dashed black circle), integrating for 2.5 min at each pointing (blue square markers). The observations were conducted between 2016/09/07 22:00 and 2016/09/08 00:50 Hawaii-Aleutian Standard Time, when the moon was 36% illuminated. We use `swarp`[5] (Bertin et al. 2002) to stack all these frames over a 20° orthographic field centered on (RA,Dec) = (0°, −30°) (blue square) with 1.86″ resolution, using the default background subtraction settings to mitigate temporal and spatial background variation.

Our second campaign is a slightly deeper survey designed to better mitigate airglow fluctuations and CCD systematics for the purpose of studying faint foregrounds below the detection limit. We select four 5 deg fields positioned around the MWA beam peak for best cross correlation precision: (RA,Dec) = (−2.5°, −24.5°), (2.5°, −24.5°), (−2.5°, −29.5°), (2.5°, −29.5°) (J2000). We raster scan a 3×3 grid of 30 sec observations within each field (red circle markers) intended to mitigate slight amplifier non-uniformities across the CCD array. The observations were conducted on 2016/11/02 between 21:47 and 23:11 Hawaii-Aleutian Standard Time, when the moon was 5% illuminated.

We stack the frames in each of the four deep fields using `swarp` over only the central $4° \times 4°$ region over which all nine 30 sec frames overlap (red squares). Otherwise slight background discontinuities would be introduced by the different temporal coverage of different regions of the stack. In this stacking we disable background subtraction for the purpose of studying the effects of airglow-induced diffuse backgrounds.

4. POINT SOURCE FOREGROUNDS

In this section we show with data and simulations that geometric effects can introduce slight correlations between radio and infrared foreground fluxes in broadband surveys, and we quantify how these correlations vary with source masking depth. As the foregrounds are so much brighter than the EOR emission, even slight foreground correlations can bury the predicted EOR anticorrelation, though the sign difference will help to identify which effect has been detected.

4.1. Catalogs

To characterize the bright sources relevant to broad band 21 cm–Lyα and 21 cm–Hα intensity mapping correlation measurements, we calculate the correlations between catalogs at 185 MHz, 850 nm, and 4.5 μm as a function of mask depth. These bands correspond roughly to 21 cm, Lyα, and Hα respectively from $z \sim 6 - 7$.
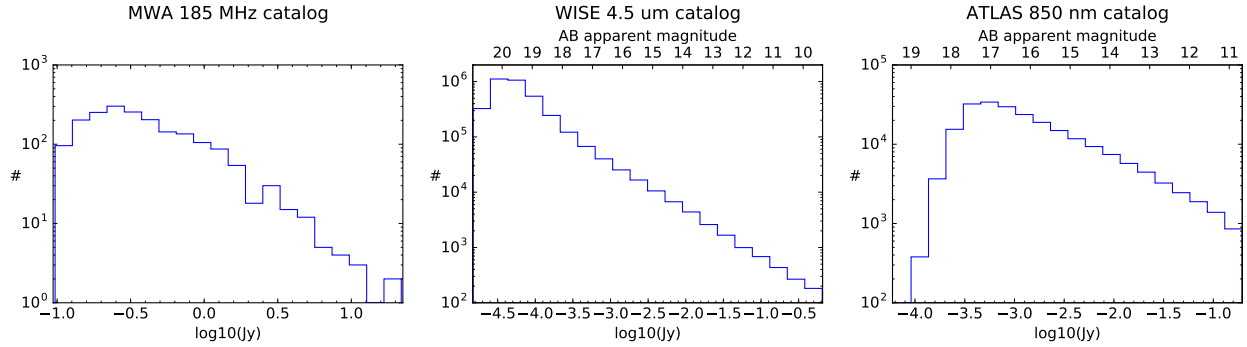
---

[5] http://www.astromatic.net/software/swarp

FIG. 2.— Histogram of source fluxes in the 185 MHz catalog (left), the 4.5 μm catalog (center), and the 850 nm catalog (right). The catalogs are complete to roughly 250 mJy, 0.09 mJy, and 0.5 mJy respectively.

We use the 185 MHz catalog reduced from deep observations of the MWA field depicted in Fig. 1 by Carroll et al. (2016). Fig. 2 (left panel) shows a histogram of source fluxes in this field. The survey depth varies somewhat over the MWA field due to the varying primary beam, resulting in a catalog which is 50% complete down to 150 mJy and 95% complete down to 250 mJy. These completeness levels are shallower than the 70 mJy completeness quoted by Carroll et al. (2016) due to our large rectangular field. The MWA's intrinsic astrometry is at the $2' - 3'$ level, though (Carroll et al. 2016) cross-match with higher frequency catalogs to achieve order $10''$ astrometry.

We use the W2 band of ALLWISE (Wright et al. 2010; Cutri et al. 2013) as our 4.5 μm catalog. We download the list of sources within the MWA field using the All Sky Search on the NASA/IPAC Infrared Science Archive[6], and plot the histogram of source fluxes in Fig. 2 (center panel). This ALLWISE band is specified to be 95% complete down to 88 μJy (15.7 AB mag), though it has slight sky coverage non-uniformities due to satellite coverage.

Lastly, we run SExtractor[7] (Bertin & Arnouts 1996) on our wide $20°$ ATLAS composite image to generate an 850 nm catalog. We allow local background bias and noise estimation, set pixel saturation at 20,000 counts to avoid artifacts, and use the AUTO aperture profile. We extract sources down to 3σ above the background, in order to achieve the most complete point source mask. Given that our ATLAS observations have been calibrated and imaged through a preliminary pipeline, we cross match these sources with sources closer than $1''$ in the AAVSO[8] Photometric All Sky Survey[9] (Henden et al. 2016), finding matches for $\sim 20\%$ of ATLAS detections. Fig. 3 (top) shows a 2D histogram of APASS versus ATLAS magnitude as a function of ATLAS magnitude. We fit a gaussian to the relative magnitude for sources brighter than 13 mag, and find that our roughly calibrated ATLAS sources are too bright by 0.279 ± 0.003 mag. Applying this correction, we plot a histogram of ATLAS source fluxes in Fig. 2) (right panel), finding that our survey is complete to roughly 0.5 mJy.
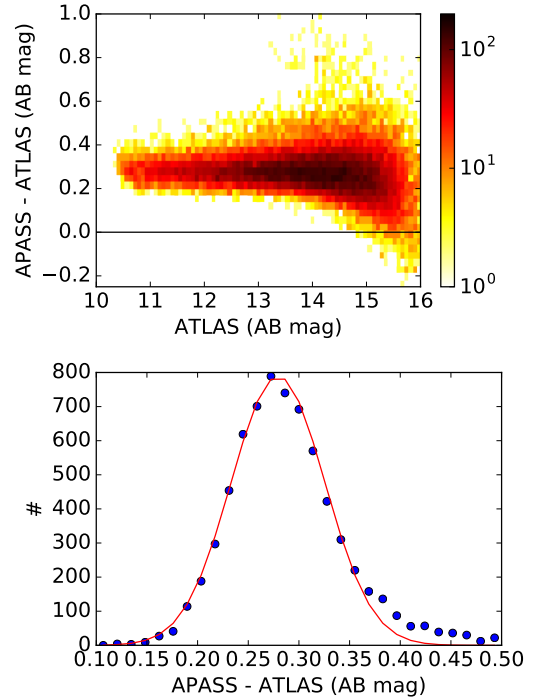
4.2. Catalog radio–infrared flux correlations

FIG. 3.— To improve the rough initial ATLAS calibration, we cross match ATLAS sources with those from APASS, and plot a 2D histogram (top) of the relative magnitude as a function of ATLAS magnitude, then fit a gaussian to the magnitude offset for sources brighter than 13 mag. We find that our roughly calibrated ATLAS sources are too bright by 0.279 ± 0.003 mag.

Having prepared catalogs of point source foregrounds in our three bands, we proceed to study how they manifest in intensity mapping correlation experiments. Traditionally, radio/infrared correlations have been studied by cross matching high frequency radio detections with infrared sources coincident within a few arcseconds, then plotting radio versus infrared luminosity. Such studies have revealed the well known radio–far infrared correlation thought to be due to massive star formation (e.g. Helou et al. 1985; de Jong et al. 1985; Yun et al. 2001; Xu et al. 1994; Mauch & Sadler 2007; Willott et al. 2003). Massive stars blow out ionized bubbles, generating radio free-free emission correlated with the ionizing flux. Some fraction of these ionizing photons are absorbed by dust clouds and reprocessed into far-infrared emission (Xu et al. 1994). At radio frequencies lower than ~ 10 GHz, synchrotron dominates over free-free emission, and the

correlation is thought to arise from the acceleration of cosmic ray electrons in these stars' supernovae.

Our approach is different. For all the advantages of broad band intensity mapping, sources cannot be localized to specific redshifts, meaning that it is foreground fluxes, not luminosities, whose correlations are of interest. Of course compact foregrounds may be masked or subtracted to some residual level, but any correlation of these residual foreground fluxes could bury the EOR correlation. We begin in this section by analyzing foreground fluxes as a function of masking depth, and in the next section turn to the foregrounds in residual images below the detection limit of these catalogs.

We begin by gridding all three catalog fluxes in Jy to the $20° \times 20°$ grid centered at $(\mathrm{RA,Dec}) = (0, -30°)$ depicted in Fig. 1 at $5'$ resolution, and calculating the zero lag correlations between the images as

$$c = \frac{\langle I_{\mathrm{rad}} I_{\mathrm{IR}} \rangle - \langle I_{\mathrm{rad}} \rangle \langle I_{\mathrm{IR}} \rangle}{\sqrt{(\langle I_{\mathrm{rad}}^2 \rangle - \langle I_{\mathrm{rad}} \rangle^2)(\langle I_{\mathrm{IR}}^2 \rangle - \langle I_{\mathrm{IR}} \rangle^2)}} \qquad (13)$$

where $\langle \rangle$ denotes an average over pixels, and the uncertainty due to sample variance is $\Delta c \approx N_{\mathrm{pix}}^{-1/2}$, where N_{pix} is the total number of pixels in the image. Between MWA and WISE catalogs we find $c = -0.003 \pm 0.005$ and between MWA and ATLAS catalogs we find $c = 0.001 \pm 0.005$. Both are consistent with zero, as expected, as the brightest sources in both infrared catalogs are likely stars, whose radio emission is expected to be negligible. As an experiment, we recalculate these correlations after excluding the brightest 10% of sources in all three catalogs, effectively masking down to $10^{-3.75}$ Jy at $4.5\,\mu\mathrm{m}$ and 10^{-2} Jy at $850\,\mathrm{nm}$, and find $c_{\mathrm{MWA-WISE}} = 0.031 \pm 0.005$ and $c_{\mathrm{MWA-ATLAS}} = 0.0086 \pm 0.005$. The former is a 6σ detection, and merits some investigation. How does this apparent correlation depend on the flux cut? What is it due to? And what does it mean for broad band correlation experiments? Further, does the MWA–ATLAS correlation remain consistent with zero at stricter flux cuts?

To begin to answer these questions, we plot in Fig. 4 the 185 MHz–$4.5\,\mu\mathrm{m}$ correlation (top left) and 185 MHz–850 nm correlation (top right) as a function of the masking depth (i.e., the maximum flux of remaining sources). We plot the SNRs of these correlation measurements, taking the noise to be $N_{\mathrm{pix}}^{-1/2}$ as described above, in the next row. Note that adjacent cells in the correlation matrix plots are somewhat correlated, so a consistent positive sign is not in and of itself evidence of significance. We assess significance by comparing of each correlation measurement individually with the expected noise (the SNR), as well as by checking that the correlation vanishes when the 185 MHz image is flipped (bottom two rows).

The 185 MHz and $4.5\,\mu\mathrm{m}$ catalogs exhibit a positive correlation peaking at 0.0332 ± 0.005 after masking infrared sources down to 10^{-4} Jy (18.9 mag) and radio sources down to 1 Jy, and remains significant down to the completeness limits of these catalogs. There is no significant correlation detection after flipping the 185 MHz image, indicating this detection is not an artifact of the analysis or of primary beam or vignetting effects. The 185 MHz and 850 nm catalogs exhibit a marginal 3σ cor-

relation after masking infrared sources down to 10^{-3} Jy (16.4 mag) and radio sources down to 0.3 Jy, though it does not appear significant in comparison to the level of correlation noise in the flipped image.

To understand these findings, we begin by investigating which $4.5\,\mu\mathrm{m}$ sources are responsible for this correlation. We select the subset of sources detected in the WISE $3.4\,\mu\mathrm{m}$, $4.5\,\mu\mathrm{m}$, and $12\,\mu\mathrm{m}$ bands and plot them (Fig. 5, left panel) in the $W_{23} \equiv [4.6\,\mu\mathrm{m}] - [12\,\mu\mathrm{m}]$ versus $W_{12} \equiv [3.4\,\mu\mathrm{m}] - [4.6\,\mu\mathrm{m}]$ color-color space used by Wright et al. (2010) to illustrate the separation between different types of sources. Nikutta et al. (2014) study more quantitatively how sources separate in this space, finding that stars are isolated in the region $W_{12} = -0.04 \pm 0.03$, $W_{23} = 0.05 \pm 0.04$ (1σ). In the right panel, we plot the faintest 90% of sources (fainter than 18.25 mags at $4.6\,\mu\mathrm{m}$) in the same color-color space and observe this cut effectively cleanly excludes nearly all the stars. This explains the detection of a 185 MHz–$4.5\,\mu\mathrm{m}$ correlation only after masking the brightest 10% of sources.

To further probe which mid infrared sources are responsible for this correlation, we make a rough cut to separate quasars and active galactic nuclei (AGN) $(W_{12} > 0.16)$ from starforming galaxies (SF) $(W_{12} < 0.16)$ (Nikutta et al. 2014; Kurcz et al. 2016). In Fig. 6 we plot the power spectrum of 185 MHz sources (left panel), $4.5\,\mu\mathrm{m}$ AGN (center panel, blue points), and $4.5\,\mu\mathrm{m}$ starforming galaxies (center panel, red points). In the right panel we plot the coherence (i.e., normalized cross spectrum) of the 185 MHz catalog with AGN (blue) and with starforming galaxies (red). The AGN cut exhibits no significant correlation with the 185 MHz sources, while the starforming galaxy cut exhibits a significant correlation rising from a few percent at $\ell \sim 7000$ to 16% at $\ell \sim 300$.

The fall of the correlation towards high ℓ is likely due in part to the MWA's $3'$ resolution at 185 MHz, corresponding to a maximum ℓ of roughly 4000, and in part to the similar fall in the $4.5\,\mu\mathrm{m}$ catalog power spectrum. Both the falling $4.5\,\mu\mathrm{m}$ catalog power spectrum and the relatively flat 185 MHz power spectrum are functions of the detailed properties of these surveys. Tegmark et al. (2002); Dodelson et al. (2002) show that the galaxy angular power spectrum $C(\ell)$ is approximately equal to the 3D matter power spectrum $P(k(\ell))$ convolved with with a window function which depends on the redshift coverage and flux limit of the sample. The matter power spectrum is known to rise as k^1 for $k \lesssim 0.02\,\mathrm{h/Mpc}$ before falling as k^{-3}. Galaxy surveys typically probe the regime just after the turnover where the slope is transitioning from 0 to -3 (Tegmark & Zaldarriaga 2002). In order to maximize its sensitivity to low surface EOR 21 cm emission, the MWA was designed as a relatively compact array in comparison to higher resolution radio interferometers such as the Very Large Array. This low resolution makes the MWA catalog severely flux limited (Carroll et al. 2016), which in turn effectively masks many galaxies which would otherwise be seen. This large masked volume translates into a wide fourier convolution kernel, explaining why the MWA catalog power spectrum is so flat.

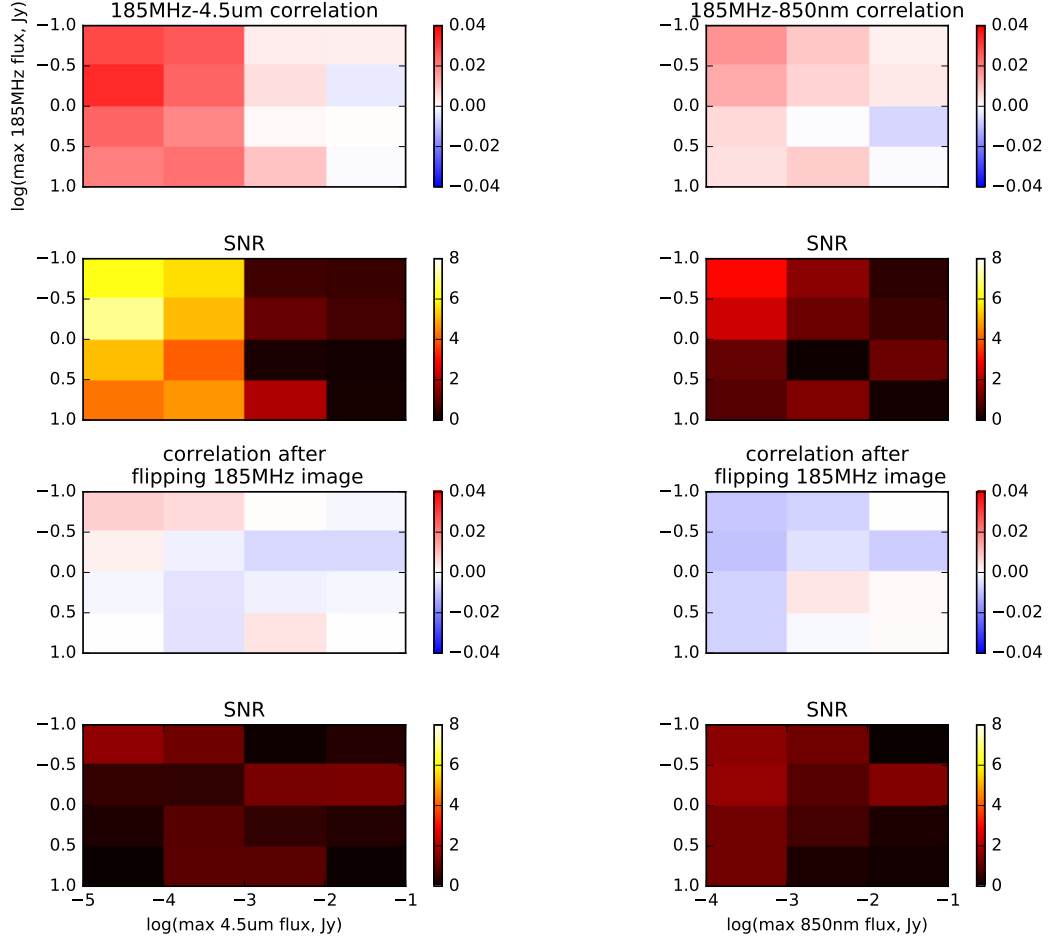Lastly, we hypothesize that the absence of an observed 185 MHz–850 nm correlation is due to the larger fraction

FIG. 4.— Image space correlation coefficient between 185 MHz and 4.5 μm sources (top left), and 185 MHz and 850 nm sources (top right), both as a function of radio and infrared flux cuts. In the second row we calculate the SNR in each bin. The bottom two rows show the correlations and SNRs after flipping the 185 MHz image about its vertical axis before the correlation calculation, giving an independent estimate of the noise. A significant correlation between 185 MHz and 4.5 μm sources appears after masking infrared sources down to 1 mJy, which we show in Fig. 5 cleanly removes the stars from the sample. The 185 MHz and 850 nm sources exhibit a marginal correlation after masking down to 1 mJy, though it does not appear significant in comparison to the noise correlation seen after flipping the radio image. We show in Fig. 7 that this is due to increased stellar contamination at 850 nm than at 4.5 μm.
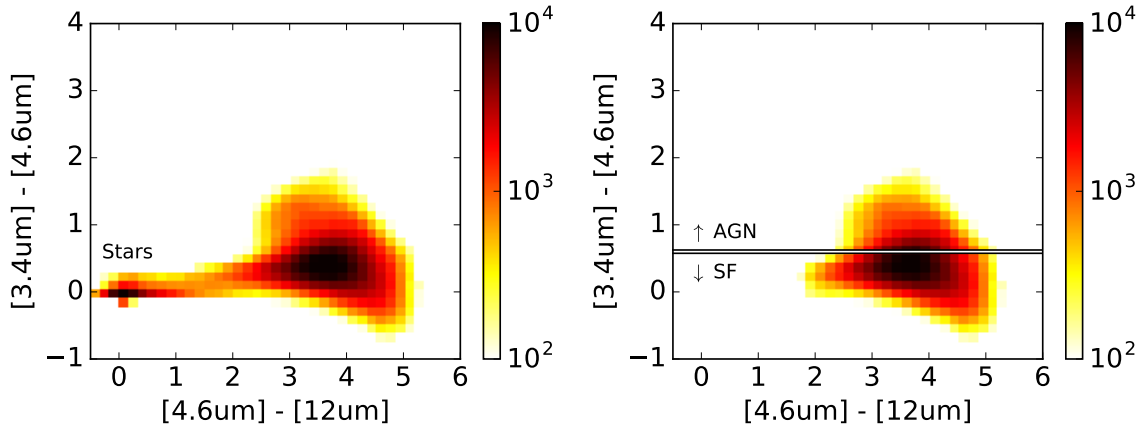

FIG. 5.— The ALLWISE sources are plotted in the color–color space of Wright et al. (2010) prior to any flux cuts (left panel), showing cluster of stars near (0,0). Then after cutting out the brightest 10% of sources (fainter than 18.25 AB mag), the stars are cleanly removed. We roughly split the remaining sources into AGNs ($W_{12} > 0.16$) and starburst galaxies (SF) ($W_{12} < 0.16$) (Nikutta et al. 2014; Kurcz et al. 2016)

.

of stars in $850\,\mathrm{nm}$ images than at $4.5\,\mu\mathrm{m}$ images. To check this, we study the fraction of stars and galaxies in a similar Galactic field observed by Sloan Digital Sky Survey (SDSS) (Eisenstein et al. 2011) DR13 (SDSS Collaboration et al. 2016). SDSS catalog[10] objects are conveniently classified as either stars or galaxies using *ugriz* photometry, though unfortunately SDSS doesn't reach the declination of our MWA field at $(\mathrm{RA,Dec})=(0,-27°)$. We thus use a $5°$ field centered at $(\mathrm{RA,Dec})=(205°,22°)$, which has the same galactic longitude but is flipped to the other side of the galactic plane, giving a similar line of sight through the galaxy.

In Fig. 7 we plot a histogram of the fluxes of the objects marked as stars (red) and galaxies (blue) in this field at the SDSS i band $(762\pm65)\,\mathrm{nm}$. We observe that stars dominate the field above $10^{-4}\,\mathrm{Jy}$, this is a factor of a few below the survey depth of our wide ATLAS $850\,\mathrm{nm}$ image, and thus none of the flux cuts explored above were deep enough to reveal the extragalactic sources. This is consistent with the fact that the $185\,\mathrm{MHz}$–$4.6\,\mu\mathrm{m}$ correlation only appeared after the stars were removed, a procedure easier in the mid-infrared than the near-infrared.

4.3. Simulations of distance-induced flux correlation

Let us now consider why the $4.5\,\mu\mathrm{m}$ SF sample is 5–15% correlated with the $185\,\mathrm{MHz}$ catalog, while the AGN sample is not. Of course some slight correlation is expected simply because brighter AGN typically reside in more massive galaxies, which are typically brighter in stars (see, for instance, Fig. 1 in (Seymour et al. 2007) or Fig. 4 in (Willott et al. 2003)), but Mauch & Sadler (2007) find no correlation between radio and *near*-infrared luminosities. As discussed above, though, broadband correlation intensity mapping experiments are affected not only by luminosity correlations but by flux correlations as well. We show in this section that fluxes in two different bands may appear correlated due to geometric effects even when their intrinsic luminosities are completely independent of each other. By geometric effects we refer to to the fact that more distant objects are generally weaker in all bands than nearer objects.

We first make a few approximations to build intuition, then simulate the effect as a function of source masking depth. Consider a sky survey with fixed field of view of a set of objects with uncorrelated infrared and radio emission. By uncorrelated we mean that the infrared and radio luminosities are independent random variables determined by the infrared and radio luminosity functions, respectively. Assume that the objects are uniformly distributed in space out to $z\sim0.5$, and work in cartesian space for simplicity. We are interested in the effective correlation between radio and infrared fluxes in the same sky pixels, but let us approximate this by calculating the correlation between source fluxes in the two bands. Starting from Eqn. 13, we have

$$c = \frac{\langle F_{\mathrm{rad}}F_{\mathrm{IR}}\rangle - \langle F_{\mathrm{rad}}\rangle\langle F_{\mathrm{IR}}\rangle}{\sqrt{(\langle F_{\mathrm{rad}}^2\rangle - \langle F_{\mathrm{rad}}\rangle^2)(\langle F_{\mathrm{IR}}^2\rangle - \langle F_{\mathrm{IR}}\rangle^2)}} \quad (14)$$

where $\langle\rangle$ denote an average over sources in the catalog. Then making the approximation that the radio and infrared luminosity functions are independent of line of

sight distance, we can rewrite this equation in terms of moments of these luminosity functions and the distribution of object distances using $F_i = L_i/4\pi d^2$ for $i =\mathrm{radio,IR}$.

$$c = \frac{\beta - 1}{\sqrt{(\beta\alpha_{\mathrm{rad}} - 1)(\beta\alpha_{\mathrm{IR}} - 1)}} \approx \frac{1}{\sqrt{\alpha_{\mathrm{rad}}\alpha_{\mathrm{IR}}}} \quad (15)$$

where $\beta \equiv \langle d^{-4}\rangle/\langle d^{-2}\rangle^2$ and $\alpha_i = \langle L_i^2\rangle/\langle L_i\rangle^2$, and again $\langle\rangle$ denote an average over sources in the catalog. The approximation in the last part of this equation results from the fact that for typical parameters and luminosity functions (see below) $\beta \gg 1$ and $\alpha > 1$.

For a survey of a fixed angular field of view, uniform spatial distribution of objects, and cartesian spacetime, the distribution of object distances d grows as d^2, giving

$$\beta = \frac{d_{\max}^3 - d_{\min}^3}{3d_{\max}d_{\min}(d_{\max} - d_{\min})} \quad (16)$$

We observe that the radio–infrared flux correlation c for some type of objects is a function of their radio and infrared luminosity functions. In fact, we can see immediately that if the luminosity distributions are wide, their α's are large, and c is small. Conversely, if the luminosity functions are narrow, then the distance to the sources plays a more significant role in determining their fluxes, and so c is larger.

To quantify whether this effect can explain our measured radio–infrared correlation in SF galaxies and the lack of one in AGN, we use AGN and SF luminosity functions at $1.4\,\mathrm{GHz}$ from Mauch & Sadler (2007) (Fig. 8, right panel) and at $8\,\mu\mathrm{m}$ from Fu et al. (2010) (left panel). The former describe galaxies at $z < 0.3$, while the latter describe galaxies at $z \sim 0.6$. In principle we should use luminosity functions at our actual radio and infrared bands, $185\,\mathrm{MHz}$ and $4.5\,\mu\mathrm{m}$, and of course this analysis could be extended using proper redshift-dependent luminosity functions, though we find that our simplified analysis suffices to explain our earlier correlation measurements. We leave a more detailed study for future work. Indeed Prescott et al. (2016) find that the AGN and starforming galaxy radio luminosity functions at lower frequencies, specifically at $325\,\mathrm{MHz}$, closely follow those at $1.4\,\mathrm{GHz}$ up to an overall scaling which cancels out of our correlation coefficient. We use approximately the same range of luminosities used by Mauch & Sadler (2007) and Fu et al. (2010), and adjust the minimum luminosities slightly to achieve the same number density of each type of object in both radio and infrared surveys. The number density of AGN is $\sim 0.0020/\mathrm{Mpc}^3$ while that of SF is $\sim 0.00011/\mathrm{Mpc}^3$. In the end we find that our results are only weakly sensitive to these luminosity minima as their faint ends become less and less significant in real, flux-limited samples.

We pick fiducial survey parameters of $d_{\min} = 20\,\mathrm{Mpc}$ and $z_{\max} = 0.75$ ($3000\,\mathrm{Mpc}$), giving $\beta \approx 54$. Using the above luminosity functions, we find $\alpha_{\mathrm{SF,IR}} = 1.474$, $\alpha_{\mathrm{SF,rad}} = 14.56$, $\alpha_{\mathrm{AGN,IR}} = 22.97$, $\alpha_{\mathrm{AGN,rad}} = 257.5$. These values agree with qualitative observation that the AGN luminosity function is wider than the SF luminosity function in both radio and infrared bands (Fig. 8). These values give a predicted radio–infrared correlation of 0.21 for SF and 0.01 for AGN agreeing with our finding of a significant radio–infrared correlation for SF and
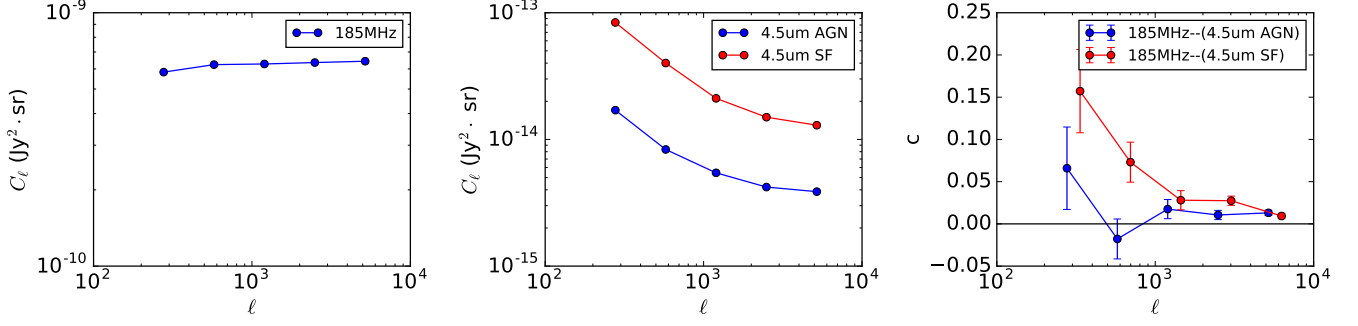
FIG. 6.— Power spectrum of 185 MHz sources (left panel), $4.5\,\mu$m sources (center panel), and coherence between 185 MHz and $4.5\,\mu$m sources (right panel). We roughly separate the $4.5\,\mu$m sources into starforming galaxies (SF) and active galactic nuclei (AGN) as illustrated in Fig. 5. We find that the 185 MHz–$4.5\,\mu$m correlation observed in Fig. 4 holds only for the starforming galaxies in the infrared sample.
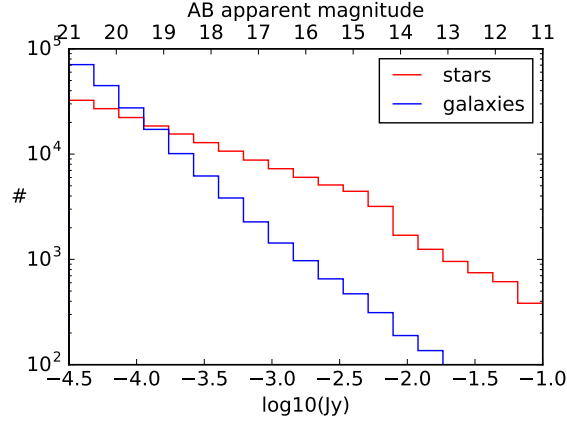

FIG. 7.— Histogram of fluxes of stars and galaxies in a $5°$ SDSS field at the same galactic longitude as the MWA field, but flipped to the other side of the galactic plane. These fluxes are in the SDSS i band (762 ± 65) nm.
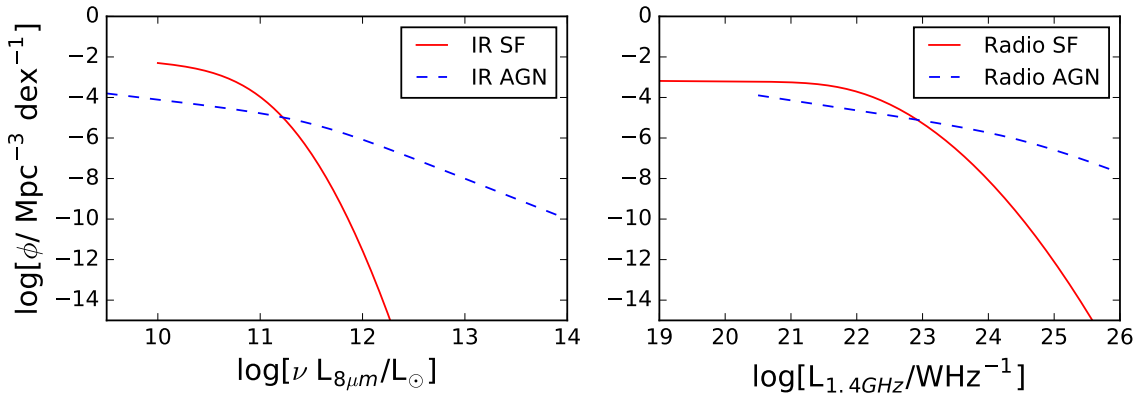

FIG. 8.— AGN and SF luminosity functions at $8\,\mu$m from Fu et al. (2010) (left panel) and at 1.4 GHz from Mauch & Sadler (2007) (right panel).

near-zero correlation for AGN. The exact values deviate from our measurements for a number of reasons. The MWA and WISE catalogs are not matched in depth or redshift coverage, and thus don't survey an exactly overlapping set of radio and infrared sources. Further, real world luminosity functions can exhibit redshift evolution. Of course, our measurement in Sec. 4.2 did not even split up the MWA catalog into separate AGN and SF subsets as such detailed characterization of low frequency radio foregrounds remains an active area of research. Lastly, the limited MWA resolution pushes the observed correlation with infrared images to zero at high-ℓ, suppressing the overall correlation computed in image space.

Can this unwanted radio–infrared foreground correlation be mitigated by masking the brightest sources? Using the luminosity functions presented above, we simulate radio and infrared surveys for each of AGN and SF. We begin by generating the mock radio catalogs of AGN and SF, choosing a poisson random number of each in each of 400 logarithmic luminosity bins. We distribute the objects uniformly over a volume $D_{\max} = cz_{\max}/H_0 = 3212\,\mathrm{Mpc}$ deep and $\theta_{\mathrm{FOV}} D_{\max} = 1121\,\mathrm{Mpc}$ wide, then pick a random infrared luminosity for each radio object from the appropriate infrared luminosity function. Finally we plot in Fig. 9, along the lines of Fig. 4, the predicted $1.4\,\mathrm{GHz}$–$8\,\mu m$ correlation of our mock AGN and SF catalogs after masking down to a maximum radio and infrared flux. As we saw above, without any flux cut we find a roughly 20% radio–infrared correlation for AGN and negligible correlation for SF. As we mask fainter and fainter sources, the AGN correlation generally increases to the 5–10% level, while the SF correlation first increases, then decreases after masking down to 10^{-4} Jy. With increasing mask depth, these correlations do not necessarily approach zero monotonically, and more detailed modeling of effective foreground flux correlations will be necessary in real world intensity mapping correlation experiments probing the EOR. In the next section we move beyond the bright sources and study the magnitudes and correlation properties of the residual radio and infrared foregrounds in our MWA and ATLAS observations.

5. RESIDUAL FOREGROUNDS AND CROSS SPECTRUM LIMITS

In this section we characterize the power spectra and correlation properties of the residual 185 MHz and 850 nm foregrounds after subtracting and masking the the bright sources identified by the surveys discussed in the previous section.

5.1. *Residual 21 cm foregrounds*

We begin by quantifying the 185 MHz foreground residuals in angular power spectrum measurements. We ask how much observation time is required to achieve the best foreground subtraction, with an eye towards the large survey areas required to mitigate the sample variance noise due to the uncorrelated component[11] of radio and

---

[11] Of course, the slight geometric correlations discussed in Sec. 4 between radio and infrared fluxes introduce additional problems. However, we will show in Sec. 5.4 that the leading order challenge in such intensity mapping correlation measurements is averaging down the sample variance noise in the cross spectrum due to the uncorrelated component radio and infrared foregrounds.

infrared foregrounds in cross spectrum analyses. In fact, while 3D 21 cm intensity mapping measurements are limited by radio thermal noise and demand order thousand hour integrations to reveal the cosmological signal (e.g. Beardsley et al. 2013; Pober et al. 2014), noise in broadband measurements is subdominant to foreground residuals after much shorter integration times. The MWA's uv plane is essentially full after just 3 hours of rotation synthesis, and compounding days together serves only to reduce noise in 3D power spectra.

Working with the MWA image products presented in Sec. 3.1, we compute the angular power spectrum as

$$C(\ell) = \frac{\sum_{\vec{\ell}} |\tilde{I}_{\mathrm{uni}}(\vec{\ell})\, \tilde{I}_w(\vec{\ell})|^2}{\sum_{\vec{\ell}} |\tilde{I}_w(\vec{\ell})|^2} \frac{d\theta^2}{N^2} \qquad (17)$$

summing over all the $\vec{\ell}$ values in each ℓ bin, where N is the number of pixels of size $d\theta$ on each side of the square image. Note that $\ell = 2\pi u$. We estimate the thermal noise power spectrum by computing the power spectrum of the difference cube which contains only thermal noise.

We plot in Fig. 10 the power spectra of 185 MHz broad band images from 3 hour data selections, spread over 1 (red solid), 5 (green solid), and 19 (blue solid) nights. We also make a broad band image of the same 32 hour data set used by Beardsley et al. (2016) (black hashed). We plot the power spectra of the raw (pre foreground subtraction), residual (post foreground subtraction), and noise (difference between successive integrations) images out to $\ell = 2600$, corresponding to a maximum baseline length of ~ 700 m. Beyond this, the uv coverage becomes sparse, introducing artifacts in the application of gridding and uniform weighting, though a more sophisticated analysis could likely use these longer baselines. We cross-check our imaging and power spectrum analysis by comparing the $k_\parallel = 0$ bin of the 3D power spectrum of Beardsley et al. (2016) converted to an angular power spectrum using Eqn. 8 (cyan and magenta dashed).

We find that the raw power spectra of the different 3 hour data sets agree with each other and with that of the 32 hour data set, as expected given that foregrounds overwhelm thermal noise in a broad band image. Interestingly, the residual power spectrum decreases as the 3 hours are spread over more and more nights until it reaches the level of the deep 32 hour integration, a factor of ~ 4 lower in power than in the single night analysis. These findings suggest that slight ionosphere-related errors limit the accuracy of each night's calibration, and thus, of its foreground subtraction. Thus while we use the 32 hour integration for convenience in this analysis, it is by no means necessary to use such a deep integration in a broad band analysis, especially when we will find later that survey speed is essential for such broad band correlation experiments. Note that as expected, the thermal noise of the deep integration is a factor of ~ 10 in power below that of the 3 hour integrations, which are themselves at least a factor of ~ 100 below the foreground residuals in these broad band images.

5.2. *Residual IR foregrounds*

We proceed to generate foreground-masked 850 nm images of each of the four deep ATLAS integration fields shown in Fig. 1. Each of these fields is a stack of
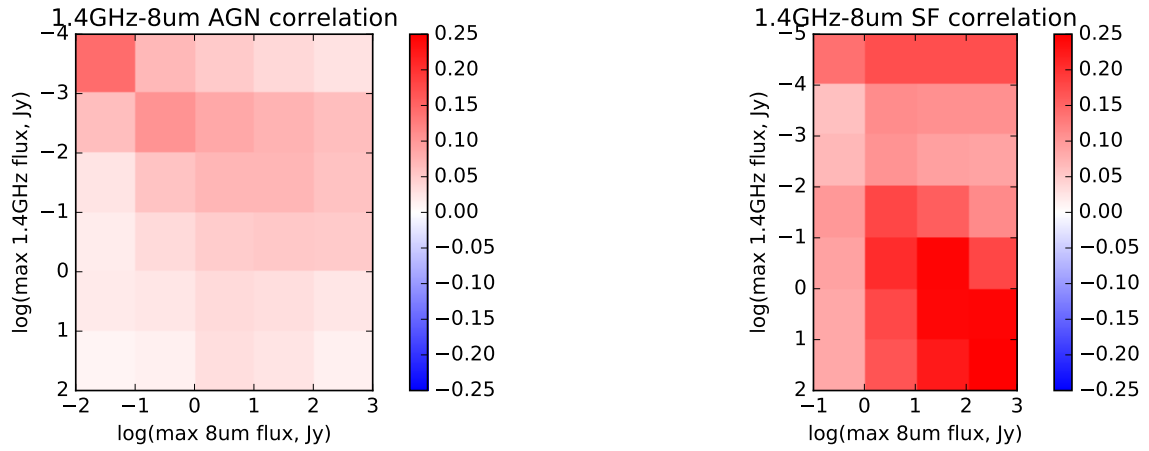
FIG. 9.— Apparent correlation of radio and mid-infrared fluxes from a mock survey with independent radio and mid-infrared luminosities using the luminosity functions from Fig. 8 for AGN and SF. Without any flux cuts (lower right corner of each plot) we observe a significant correlation between radio and mid-infrared fluxes for SF objects, and a near-zero correlation for AGN. This agrees with our measurements on real 185 MHz and 4.5 μm sources in Fig. 4. As fainter and fainter objects are masked, we observe that the AGN correlation gradually strengthens, while the SF correlation weakens somewhat.
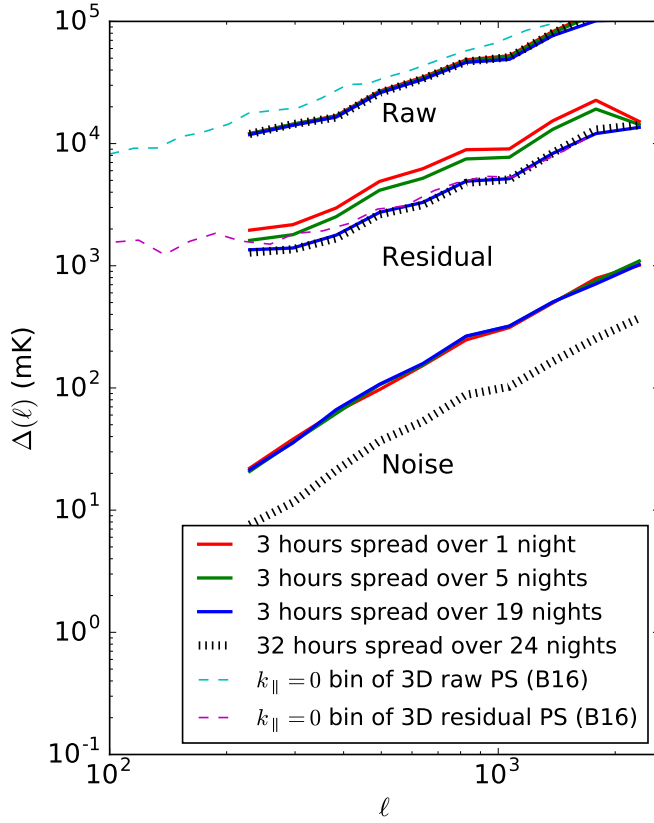
FIG. 10.— Raw, residual (post foreground subtraction), and noise power spectra of 185 MHz broad band images from various 3 hour data selections (solid lines) and from the 32 hour data selection used by Beardsley et al. (2016) (black hashed). The uv plane is nearly filled after only 3 hours, yet we find that spacing these ~ 100 two-minute integrations over many independent nights reduces the foreground residuals by a factor of up to ~ 4 in power. As a cross check on our analysis, we plot the $k_\parallel = 0$ bin of the 3D power spectrum of Beardsley et al. (2016) converted to an angular power spectrum using Eqn. 8 (cyan and magenta dashed).

9 30 sec exposures with $5°$ field of view, dithered such that the overlap is a $\sim 4°$ region. By confining ourselves to this overlap region, we avoid the background discontinuities which affect many nominally wide field infrared image datasets whose mosaicing introduces significant background patchiness. Mitchell-Wynne et al. (2015) have demonstrated that complex fitting (Fixsen et al. 2000) can help reduce such patchiness in small mosaics (tens of arcminutes), but further work is required to study whether such techniques can be applied to much larger fields.

Our approach is to mask each image at ATLAS's native $1.86''$ resolution, then coarse grid down to $3.5'$, approximately the resolution of the MWA, taking each coarse pixel's value as the average of all unmasked fine pixels within it. If fewer than 10% of its fine pixels remain unmasked, we consider the whole coarse pixel masked to avoid introducing too much noise variation between different coarse pixels. We proceed in the following four stages, which we illustrate in Fig. 11 for the ATLAS field centered at (RA,Dec)= $(2.74°, -24.79°)$ (the top right red box in Fig. 1). Each row shows the result of an additional masking stage, as outlined below. The left column shows a typical $9'$ field to illustrate the masking up close; the center column shows the resulting coarse binned im-

age with $3.5'$ resolution; and the right column shows the FFT of the center image (plotted as $\log[\Delta(\vec{\ell})/(\mathrm{kJy/sr})]$) in order to identify detector systematics.

1. **Mask saturated regions.** (Fig. 11, row 1) Nearly saturated pixels are associated with nearby bright stars which would dominate fluctuation measurements, we thus mask $4'$ around all pixels within 30% of saturation (white regions in left and center columns). This wide mask removes the broad wings of the PSF revealed by these extremely bright sources. Roughly 92% of fine pixels remain unmasked after this step, and 96% of coarse pixels remain.

2. **Mask sources to 5σ.** (Fig. 11, row 2) We mask circular regions with radius equal to five times the profile RMS along the minor axis of each source as measured by SExtractor (see Sec. 4.1). The reason we use the minor axis RMS is that vertical charge leakage in the CCD array results in unrealistically large major axis RMS measurements for bright sources. After this stage $\sim 87\%$ of fine pixels remain unmasked, and the fraction of coarse pixels remaining is unaffected.

3. **Mask sources to 12σ and mask other emission above 70 kJy/sr over the background.** (Fig. 11, row 3) We use a larger masking radius to remove the PSF wings around the 5σ source masks. We also determine that the vertical CCD charge leakage can be isolated by looking for emission brighter than 70 kJy/sr over the background, and that it can be flagged without cutting into the shot noise between sources. 58% of fine pixels remain unmasked after this step, and again the fraction of coarse pixels remaining is unaffected.

4. **Mask horizontal and vertical fourier modes.** (Fig. 11, row 4) The previous two stages revealed detector artifacts within $\Delta\ell \sim 100$ of $\ell_x = 0$ and $\ell_y = 0$. These compact fourier systematics correspond to slight horizontal and vertical discontinuities in the center image due to imperfect gain matching between 16 different amplifiers which process different rectangular regions of the CCD array. We conservatively mask fourier modes with $|\ell_x| < 200$ or $|\ell_y| < 200$ to eliminate this effect.

Note that these 850 nm deep observations were recorded during near new moon conditions, and we find that the mean air glow in source-free regions is $\sim 3 \times 10^3$ kJy/sr, of order 19 AB mag/arcsec^2. For comparison, Sullivan & Simcoe (2012) measure a 1020 nm continuum air glow brightness (i.e., after spectrally masking the OH lines) of 20 ± 0.5 AB mag/asec^2 far away from the moon.

We proceed to characterize the residual infrared fluctuations in power spectrum space, using the optimal quadratic estimator to properly account for the masking of coarse pixels. This estimator was introduced to astronomy by Tegmark (1997) to recover CMB power spectra from maps with arbitrary survey geometries and noise properties, and has recently been revived for 3D power spectrum analysis of 21 cm data by Dillon et al. (2014, 2015); Liu & Tegmark (2011); Dillon et al. (2013);
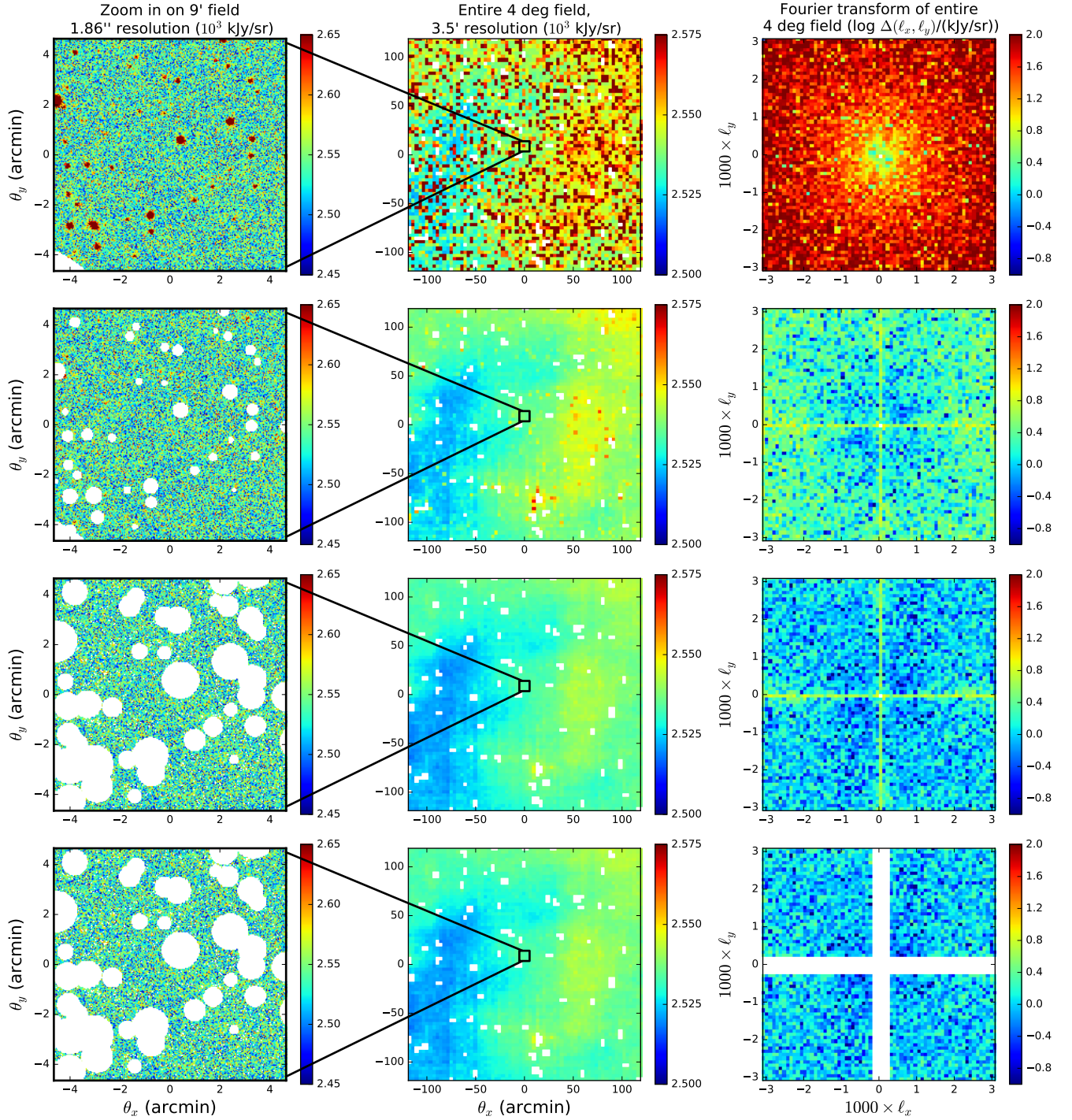
FIG. 11.— The rows of this figure illustrate our four stages of foreground removal on ATLAS images, as detailed in Sec. 5.2. We apply this process to all four 4° deep ATLAS fields shown in Fig. 1, but here show the one centered at (RA,Dec)= 2.74°, −24.79°) for illustration. The first row shows the results of masking 4′ around all nearly saturated regions; the second row shows the results of masking out to 5σ; the third row shows the results after masking out to 12σ and all pixels above 70 kJy/sr over the background, and the last row shows the results after masking the nearly horizontal or vertical fourier modes. In each row, we show the central 9′ field at 1.86″ resolution (left), the entire 4° field after coarse gridding to 3.5′ (center), and the FFT of the coarse gridded field to highlight systematics (right). Note that the left and center panels of the bottom row are identical to those in the row above it, illustrating that the image space mask is the same, only the fourier mask is different.

Ali et al. (2015). We employ this estimator in a manner more similar to the original CMB case, using it to account for pixel masking in broad band power spectrum estimation. We briefly summarize the estimator here, and refer to Dillon et al. (2014) for a more detailed description.

We label the normalized estimator of the power in bin α as p_α, related to the unnormalized estimator q_α as $\mathbf{p} = \mathbf{M}\mathbf{q}$. Bold lower-case letters are vectors, while bold upper-case letters are matrices. The unnormalized estimator is given by

$$q_\alpha = \frac{1}{2}(\mathbf{x} - \langle \mathbf{x} \rangle)^t \mathbf{C}^{-1} \mathbf{C}_{,\alpha} \mathbf{C}^{-1} (\mathbf{x} - \langle \mathbf{x} \rangle) \qquad (18)$$

where $\mathbf{x}$ is a column vector containing all $N_{\rm pix}$ pixel measurements, $\mathbf{C}$ is the pixel-pixel covariance matrix, and $\mathbf{C}_{,\alpha} \equiv d\mathbf{C}/dp_\alpha$ is the derivative of the covariance with respect to the power in bin α. Note that $\mathbf{C}_{,\alpha} = \mathbf{A}^\dagger \mathbf{A}$, where $\mathbf{A}$ is a $N_\alpha \times N_{\rm pix}$ with elements $A_{ij} = \exp(i\vec{\theta}_j \cdot \vec{\ell}_i)$. Here $\vec{\ell}_i$ refers to the i'th out of N_α $\vec{\ell}$ modes in bin α. Note that ^{t} denotes a transpose, and † donates a conjugate transpose.

The matrix of window functions (i.e., horizontal error bars) of the band powers p_α, defined such that $\mathbf{p}_{\rm estimated} = \mathbf{W}\mathbf{p}_{\rm true}$ is given by $\mathbf{W} = \mathbf{M}\mathbf{F}$, where $\mathbf{F}$ is the Fisher matrix and $\mathbf{M}$ is an arbitrary invertible normalization function encoding the compromise between horizontal and vertical error bars. The covariance between the measured p_α values is given by $\boldsymbol{\Sigma} = \mathbf{M}\mathbf{F}\mathbf{M}^t$. Dillon et al. (2014) argue that taking $\mathbf{M} \propto \mathbf{F}^{-1/2}$ is a good compromise between small horizontal error bars and small vertical error bars. For simplicity, we take $\mathbf{M} = \mathbf{F}^{-1/2}$, and correct the normalization at the end by dividing each element of $\mathbf{p}$ by the peak of the appropriate row of $\mathbf{W}$. In this case, the bandpower variances are given by the reciprocals of the peaks of the window functions used to normalize the bandpowers. We find that using the sum instead of the peak significantly biases the recovered bandpowers downward when the power spectrum is non-flat, as in our case. Lastly, the elements of the Fisher matrix are given by

$$\mathbf{F}_{\alpha\beta} = \frac{1}{2}\mathrm{tr}\left(\mathbf{C}^{-1}\mathbf{C}_{,\alpha}\mathbf{C}^{-1}\mathbf{C}_{,\beta}\right) \qquad (19)$$

Now we turn to application of this formalism to power spectrum estimation from our masked IR images. Later we will adapt it to estimation of the 21 cm–IR cross spectrum. We take $\mathbf{x}$ to be a vector of all IR coarse (3.5′) pixel values, with masked pixels set to zero. After gridding the high resolution images to reach this resolution, photon shot noise is negligible, and the image space covariance is the sum of the signal covariance $\mathbf{C}_{\rm signal}$ and the masking covariance $\mathbf{C}_{\rm mask}$. $\mathbf{C}_{\rm mask}$ is a diagonal matrix with ∞ for masked pixels, and 0 otherwise. In practice, we replace ∞ with a number 10^7 times larger than the largest eigenvalue of $\mathbf{C}_{\rm signal}$, finding that the results are not sensitive to this parameter. The signal covariance is easily obtained by writing it in fourier space, where it is a diagonal matrix with a guess of the true power spectrum on the diagonal, then fourier transforming it into image space with fourier transform matrix $\mathcal{F}$.

$$\mathbf{C} = \mathcal{F}^\dagger \mathbf{C}_{\rm ft} \mathcal{F} + \mathbf{C}_{\rm mask} \qquad (20)$$

Note that $\mathcal{F}$ is an $N_{\rm pix} \times N_{\rm pix}$ matrix with elements $\mathcal{F}_{ij} = \exp(-i\vec{\theta}_i \cdot \vec{\ell}_j)$, where i runs over all pixels and j runs over all fourier cells. Said differently, a guess of the power spectrum is necessary to optimally downweight the sample variance noise on the estimated power spectrum. If the accuracy of the guess were in question, one could always iterate by feeding the estimated power spectrum back into the quadratic estimator, though in practice we find this is not necessary.

Using this estimator, we calculate the power spectrum after each stage of masking and plot the mean spectrum over all four deep ATLAS fields in Fig. 12 (left panel). We conservatively take the standard deviation of each bandpowers over the four fields as its error bar, due to sample variance. The power spectrum of all 850 nm sources, masking only saturated regions, rises proportionally to ℓ (red dots), as expected for poisson source counts when the power spectrum is plotted as $\Delta = \sqrt{\ell^2 C_\ell / 2\pi}$. Masking sources out to 5σ removes two orders of magnitude in power (green dots), and masking out to 12σ and above 70 kJy/sr over the background removes a factor of a few more in power (blue dots). Lastly, excluding modes with $|\ell_x| < 200$ or $|\ell_y| < 200$ gives our final 850 nm anisotropy spectrum (black dots). We note that increasing any of the masking parameters introduced here in this section does not significantly alter these results. Increasing the masking radius around saturated pixels, for instance, results in masking of more coarse pixels, but removes negligible additional power from these residual spectra. After only the first masking stage, we use a flat power spectrum in $C(\ell)$ as our guess for the quadratic estimator formalism, and after the other three masking steps we use the 1.1 μm power spectrum from Zemcov et al. (2014).

In Fig. 12 (right panel) we compare our final residual power spectrum measurement with other measurements[12] in the literature. Our 850 nm ATLAS spectrum (black dots) agrees very well with the $(1.1 \pm 0.25)\,\mu$m CIBER spectrum (blue dots), with much smaller error bars at the ℓ modes we can access ($\ell \lesssim 4000$) due to ATLAS's larger field of view.

Zemcov et al. (2014) argue that this spectrum is limited by foregrounds at all ℓ: by diffuse Galactic light (DGL) (i.e., dust) at $\ell \lesssim 1000$, by intrahalo light (IHL) (cyan dashed line) at $\ell \sim 1000$, and by galaxy number counts below the flux limit at larger ℓ. ATLAS's 2″ resolution is only a factor of 3 finer than CIBER's 7″ resolution, so perhaps it is not surprising that the level of galaxy number counts is not appreciably different. The ATLAS points at $\ell > 1000$ are in fact slightly above CIBER's, though the measurements were conducted over different bands. Mitchell-Wynne et al. (2015) demonstrate a dramatic improvements in foreground reduction $\ell \gtrsim 10^3$ in 10′ Hubble mosaics with 0.1″ resolution. Their power spectrum (red points) is a factor of ~ 10 below the previous IHL model, and their new fitting suggests that CIBER was in fact limited almost exclusively by DGL (red solid line). For comparison, we plot their new IHL model (red dashed) and EOR model (grey area). These results show that finer angular resolution can help reduce the foregrounds in the $\ell \sim 10^2 - 10^4$ modes that can be

---

[12] Note that $I_f/({\rm kJy/sr}) \approx 0.3\lambda I_\lambda/({\rm nW/m^2/sr})$ at $\lambda = 1\,\mu$m.
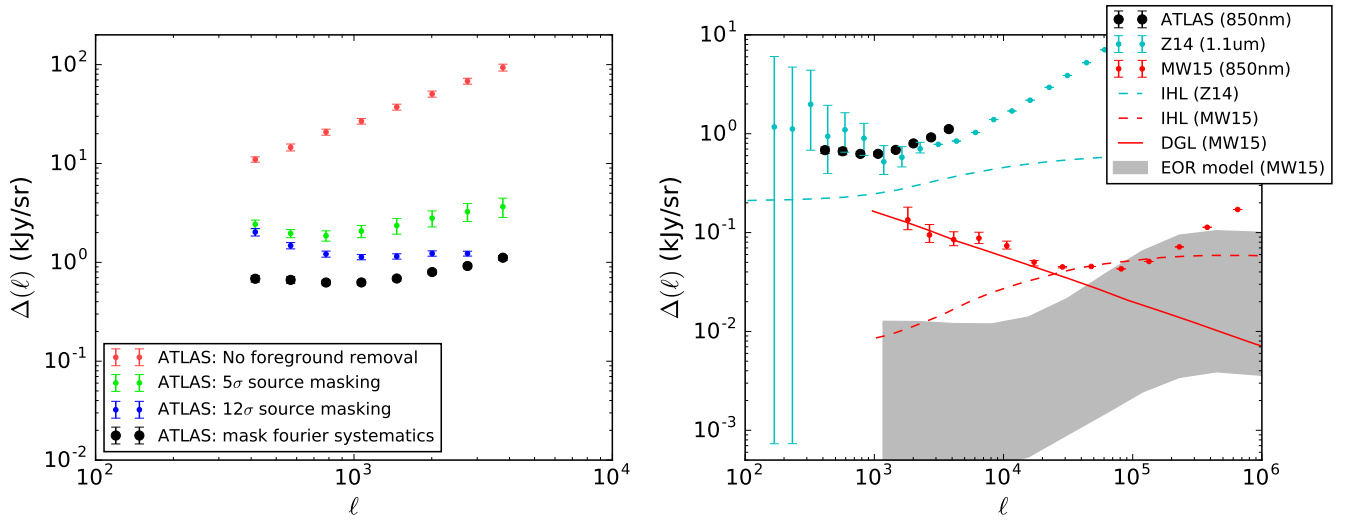
Fig. 12.— **Left panel:** Power spectra of our broad band ATLAS images after various stages of foreground removal as described in Sec. 5.2. The error bars show the sample variance noise estimated by the standard deviation over our four 4° fields (see Fig. 1). Black dots show our final ATLAS 850 nm power spectrum after all stages of foreground removal. **Right panel:** We compare our ATLAS power spectrum (black dots) to the 1.1 μm power spectrum of Zemcov et al. (2014) (Z14) (cyan dots) using the CIBER experiment and the 850 nm power spectrum of Mitchell-Wynne et al. (2015) (MW15) using 10′ Hubble mosaics. We show the intrahalo light (IHL) models of ZI4 (cyan dashed line) and MW15 (red dashed line), as well as the diffuse galactic light (DGL) model of the latter authors. These results show that finer angular resolution can help reduce the foregrounds in the $\ell \sim 10^2 - 10^4$ modes that can be cross correlated with 21 cm EOR intensity mapping experiments. The gray area shows the predicted EOR contribution to the infrared anisotropies from MW15.

cross correlated with 21 cm EOR intensity mapping experiments.

5.3. *Modeling the 21 cm–Lyα cross spectrum*

Before turning to 21 cm–Lyα cross spectrum measurements with our 185 MHz and 850 nm images, we generate optimistic and pessimistic theoretical cross spectra for comparison. We simulate 21 cm and Lyα cubes using 21 cm power spectra from Pober et al. (2014), Lyα power spectra from Gong et al. (2014), and the coherence between the two fields from Heneka et al. (2016). Combining simulations from all these sources allows us to better estimate the modeling uncertainty. Future work is needed to more self-consistently model these fields and their correlation over a range of possible reionization scenarios, and to infer astrophysical parameters from observed cross spectra.

Gong et al. (2014) model the Lyα cross spectrum and plot an uncertainty region over a range of likely values of escape fraction of ionizing photons, fraction of radiation emitted at Lyα, star-forming rate, and IGM clumping factor (their Fig. 1). We take the upper and lower edges of their uncertainty region as our optimistic and pessimistic power spectra, respectively.

Similarly, Pober et al. (2014) simulate 21 cm power spectra over a range of reionization scenarios with various values of ionizing efficiency (ζ) (including the escape fraction), the minimum halo virial temperature (T_{vir}), and the ionizing photon mean free path in the IGM (R_{mfp}). Whether the signal at our redshift of interest ($z = 7$) is large or not depends mostly on whether reionization is already largely finished by that time or not. So we take as our pessimistic model the $z = 8$ power spectrum of the ($\zeta = 31.5$, $T_{\mathrm{vir}} = 1.5 \times 10^4$ K, $R_{\mathrm{mfp}} = 30$ Mpc) scenario, whose reionization midpoint is $z = 9.5$. We take as our optimistic model a late reionization scenario with $T_{\mathrm{vir}} = 3 \times 10^5$ K (other parameters unchanged), whose reionization midpoint is $z = 5.5$.

From these power spectra and coherence functions, we generate approximate 21 cm and Lyα cubes assuming gaussian statistics, a rough approximation given the 21 cm field is expected to become somewhat less gaussian as reionization proceeds (Wyithe & Morales 2007). The simulated cubes have $(1 \,\mathrm{Mpc})^3$ resolution over a $(218 \,\mathrm{Mpc})^3$ volume at $z = 7$, corresponding to $\Delta z = 0.6$ and $0.4'$ angular resolution. Our 21 cm and Lyα cubes have units of mK and kJy/sr, respectively, and we average them in the line of sight direction to produce broad band images.

We plot the original 3D spherically averaged power spectra and coherence function as solid lines in Fig. 13, and the computed 2D power spectra from the line-of-sight averaged cubes, as well as their coherence as dashed lines. The 3D power spectra are plotted as $\Delta(k)$ and $\Delta(\ell)$ as defined in Sec. 2.1. As expected, line of sight averaging tends to remove power on short spatial scales (compared to the line of sight depth of the cube), but it acts similarly on both cubes, so the coherence function is preserved. Note that the cross spectrum Δ_{12} is defined in terms of the coherence as $\Delta_{12}^2 = c[\Delta_1^2 \Delta_2^2]^{1/2}$ for both the 3D and 2D cases, and we use the same coherence function from Heneka et al. (2016) in the optimistic and pessimistic scenarios.

5.4. *Limits on the 21 cm–Lyα cross spectrum and an experimental design study*

Having characterized the residual sky power spectrum at 185 MHz and 850 nm after the best foreground masking and subtraction permitted by our datasets, we now search for a correlation between these radio and foreground residuals. To account for the non-uniform *uv* sampling of the MWA image as well as for the image space masking of the ATLAS image, we again use the optimal quadratic estimator. In Appendix C we show that with the approximation that the correlation between the two images is small, the proper extension of the optimal quadratic estimator from the auto spectrum case to the cross spectrum case is given by

$$q_\alpha \approx (\mathbf{x}_{21} - \langle \mathbf{x}_{21} \rangle)^t \mathbf{C}_{21}^{-1} \mathbf{C}_{,\alpha} \mathbf{C}_{\mathrm{IR}}^{-1} (\mathbf{x}_{\mathrm{IR}} - \langle \mathbf{x}_{\mathrm{IR}} \rangle) \quad (21)$$

$$\mathbf{F}_{\alpha\beta} \approx \mathrm{tr} \left(\mathbf{C}_{21}^{-1} \mathbf{C}_{,\alpha} \mathbf{C}_{\mathrm{IR}}^{-1} \mathbf{C}_{,\beta} \right) \quad (22)$$

where $\mathbf{C}_{,\alpha}$ is the same matrix used in Sec. 5.2, $\mathbf{x}_{21}$ and $\mathbf{x}_{\mathrm{IR}}$ are the vectors of 21 cm and IR pixel values. As before we use normalization $\mathbf{M} = \mathbf{F}^{-1/2}$ in $\mathbf{p} = \mathbf{Mq}$, and perform a final normalization using the peaks of the window functions $\mathbf{W} = \mathbf{MF}$. We use the same IR covariance matrix given in Eqn. 20, and use power spectrum of the residual MWA broadband images to construct the radio covariance matrix.

We apply this formalism to each of the four $4°$ deep ATLAS fields shown in Fig. 1, pairing each IR image with the overlapping region of the MWA image cropped from the naturally weighted image. We crop the image space synthesized beam over the same field of view, use it to apply uniform weighting to the cropped MWA image using Eqn. 12. In Fig. 14 we plot the resulting cross spectrum (red markers) averaged over all four fields with 2σ error bars. All are consistent with zero, and roughly half the estimated bandpowers are negative (triangles) and positive (circles), as expected if there is no measured correlation. Our bins are evenly spaced in $\log \ell$, and it is beneficial to overlap them slightly since our normalization matrix $M \sim F^{-1/2}$ ensures bin errors are always uncorrelated. We take the tops of the error bars as 95% upper limits on the cosmological cross spectrum of 21 cm and Lyα emission from the EOR, achieving a tightest upper limit of $\Delta^2 < 181 \,(\mathrm{kJy/sr \cdot mK})$ (95%) at $\ell \sim 800$. For comparison, we plot optimistic (black solid) and pessimistic (black dashed) model cross spectra as discussed in the previous section.

Also in Fig. 14 we plot the radio–infrared foreground cross spectrum that a 1% flux correlation would produce (green line), as motivated by our results in Sections 4.3 and 4.2. Such a slight geometric flux correlation is thus still predicted to be slightly below the sensitivity of our MWA–ATLAS experiment, but is it significant for other classes of experiments? To study this we predict for a range of current and future radio–IR surveys both the foreground sample variance noise and the foreground cross spectrum for a slight geometric correlation.

We describe the radio and IR surveys we study here, and summarize them side by side in Table 1.

Radio surveys

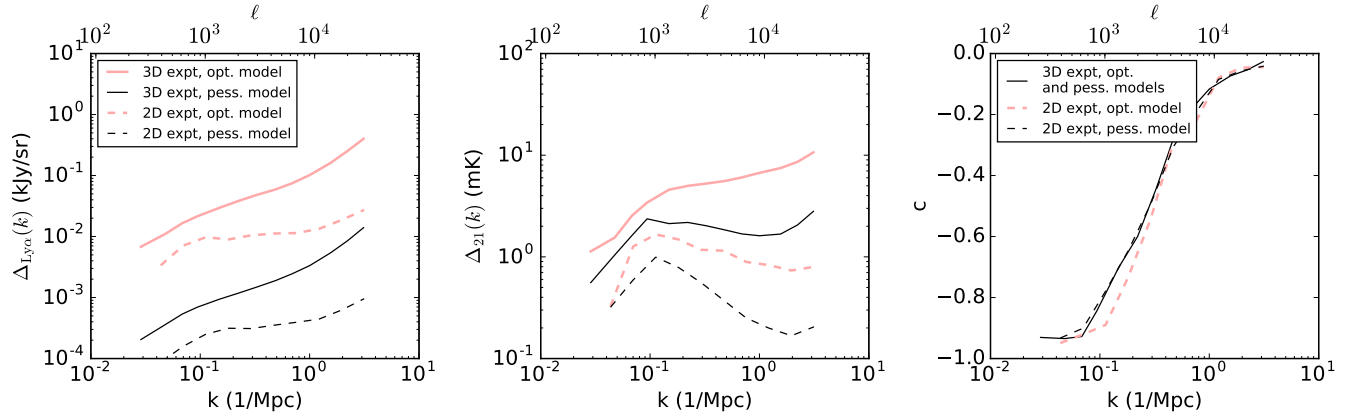- *MWA*. We use the $6'$ resolution permitted by our MWA analysis, corresponding to a maximum base-

FIG. 13.— As described in Sec. 5.3, we identify optimistic (red solid lines) and pessimistic (black solid lines) Lyα (left panel) and 21 cm (center panel) 3D power spectra from Gong et al. (2014) and Pober et al. (2014), respectively, and use the coherence function from Heneka et al. (2016) (right panel, black solid) for both scenarios. We then simulate approximate cubes assuming gaussian statistics, average them over frequency, and plot the angular power spectra of Lyα (left panel, dashed lines) and 21 cm emission (center panel, dashed lines), and their coherence functions (right panel, dashed lines). Power on short angular scales ($\ell \gtrsim 10^4$) is suppressed by a factor of ~ 100 (in power units), while power on longer angular scales ($\ell \lesssim 10^2$) is suppressed by only a factor of ~ 2 compared to the 3D power spectrum. This indicates that there remains significant observable signal in broad band 21 cm–Lyα cross spectrum observations.
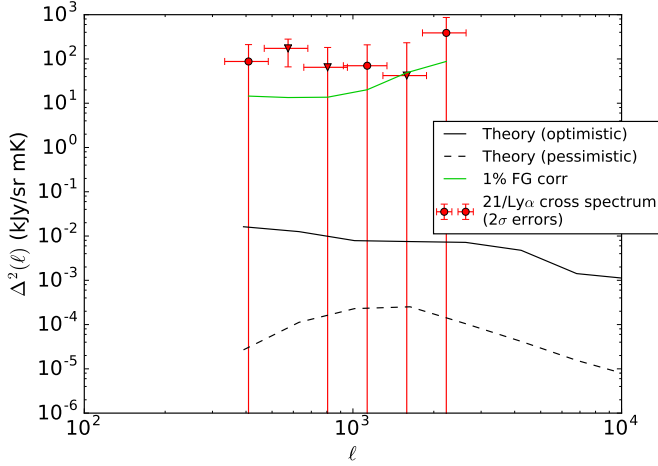
FIG. 14.— Measured cross spectrum between 185 MHz and 850 nm images after our best foreground subtraction and masking (red markers). All are consistent with zero, and roughly half are positive (circles) and negative (triangles), as expected if there is no measured correlation. Error bars show 2σ uncertainties, and we take the tops of these error bars as upper limits on the cross spectrum of 21 cm–Lyα emission from the EOR. We achieve an upper limit of $\Delta^2 < 181\,(\mathrm{kJy/sr \cdot mK})$ (95%) at $\ell \sim 800$. We also plot the radio–infrared foreground cross spectrum that a 1% flux correlation would produce (green line), and find that it is below the sensitivity of our MWA–ATLAS experiment.

TABLE 1
CURRENT AND FUTURE RADIO AND IR SURVEYS

Radio Survey	Resolution	$S_{\mathbf{max}}$
MWA	6′	150 mJy
MWA (GMRT catalog)	6′	25 mJy
HERA (GMRT catalog)	12′	25 mJy
GMRT	40″	25 mJy
LOFAR	20″	3 mJy
SKA	10″	0.1 mJy

IR Survey	Field of View	$f_{\mathbf{IR}}$[a]
ATLAS	8°	1.0
WideATLAS	40°	1.0
Hubble mosaic	4°	0.01
DES mosaic	40°	0.01

[a]Fraction of IR foreground power remaining relative to our AT-LAS analysis.

line length of 700 m, and a survey depth at 50% completeness of 150 mJy (see Sec. 4.1).

- *HERA*. Primarily designed for 3D power spectrum measurements which are severely thermal noise limited, HERA is a compact redundant array, and even the outriggers only improve the resolution to 12′. We assume the GMRT catalog is used for source subtraction (see below) DeBoer et al. (2017).

- *GMRT*. We assume a maximum baseline length of 10 km, within which the uv plane is filled relatively uniformly, this corresponds to a 40″ resolution. Intema et al. (2017) demonstrate a catalog depth of 25 mJy at 50% completeness.

- *LOFAR*. Similarly, we use a maximum baseline length of 20 km, to maintain a relatively filled uv plane, corresponding to a 20″ resolution. Yatawatta et al. (2013) report a catalog depth of 3 mJy.

- *SKA*. Prandoni & Seymour (2015) report a nominal SKA-LOG confusion-limited survey depth of 0.1 mJy at 10″ resolution.

IR surveys

- *ATLAS*. Our four ATLAS stacking field give a total field of view of 8°, though we are planning a future wider survey with a 40° field of view.

- *Hubble mosaic*. Mitchell-Wynne et al. (2015) have demonstrated high fidelity mosaicing using the technique of Fixsen et al. (2000), producing 10′ wide mosaics from the $\sim 2'$ Hubble field of view. We assume that a 4° Hubble field can be used through some combination of mosaicing (making larger IR images) and tiling (calculating the cross spectrum between each of many small Hubble fields and the radio image, then averaging the results). Motivated by the results of Mitchell-Wynne et al. (2015), we assume that the fine resolution of Hubble permits removal of IR foregrounds by a factor of ~ 100.

- *DES mosaic*. Along these lines, we assume a similar analysis can be conducted using the Dark Energy SurveyDark Energy Survey Collaboration et al. (2016) over a field of view which is $10\times$ larger on a side, due to the instrument's correspondingly larger field of view. We assume the same improvement in foreground removal as in the Hubble mosaic discussed previously.

Neglecting the small correlation, the sample variance noise due to uncorrelated foregrounds is

$$\Delta_{\mathrm{noise}}^2(\ell) = \sqrt{\frac{\Delta_{\mathrm{IR}}^2(\ell)\Delta_{\mathrm{rad}}^2(\ell)}{2N_\ell}}$$

$$= \sqrt{\frac{\Delta_{\mathrm{ATLAS}}^2(\ell)\Delta_{\mathrm{MWA}}^2(\ell)f_{\mathrm{rad}}f_{\mathrm{IR}}}{2N_\ell}}$$

$$\times \left(\frac{\theta_{\mathrm{FOV,ATLAS}}}{\theta_{\mathrm{FOV}}}\right)\left(\frac{d\theta}{d\theta_{\mathrm{MWA}}}\right) \quad (23)$$

where f_{IR} and f_{rad} are the fractions of IR and radio foreground power remaining relative to those in our ATLAS and MWA analyses. Using the empirical formula for radio source counts given in Di Matteo et al. (2002), f_{rad} scales as $f_{\mathrm{rad}} \propto S_{\mathrm{max}}^{1.25}$, where S_{max} is the flux limit of the subtraction. Scaling this relative to our MWA analysis gives $f_{\mathrm{rad}} = (S_{\mathrm{max}}/S_{\mathrm{max,MWA}})^{1.25}$.

Similarly, then the foreground cross spectrum for a correlation c is given by

$$\Delta_{\mathrm{FG}}^2(\ell) = c\sqrt{\Delta_{\mathrm{IR}}^2(\ell)\Delta_{\mathrm{rad}}^2(\ell)}$$

$$= c\sqrt{\Delta_{\mathrm{ATLAS}}^2(\ell)\Delta_{\mathrm{MWA}}^2(\ell)f_{\mathrm{rad}}f_{\mathrm{IR}}} \quad (24)$$

In Fig. 15 we plot the foreground sample variance noise versus foreground cross spectrum for a range of current and future radio and infrared survey pairs. We calculate these for the $\ell = 400$ bin and assume a 1% geometric foreground correlation. The diagonal black line separates the regions within which each of these effects dominates.

Legend:
- ● MWA/ATLAS
- ★ MWA (GMRT catalog)/WideATLAS
- ☆ HERA (GMRT catalog)/WideATLAS
- ★ GMRT/WideATLAS
- ▼ GMRT/Hubble (4deg mosaic)
- ■ GMRT/DES (40deg mosaic)
- ★ LOFAR/WideATLAS
- ▼ LOFAR/Hubble (4deg mosaic)
- ■ LOFAR/DES (40deg mosaic)
- ☆ SKA/WideATLAS
- ▽ SKA/Hubble (4deg mosaic)
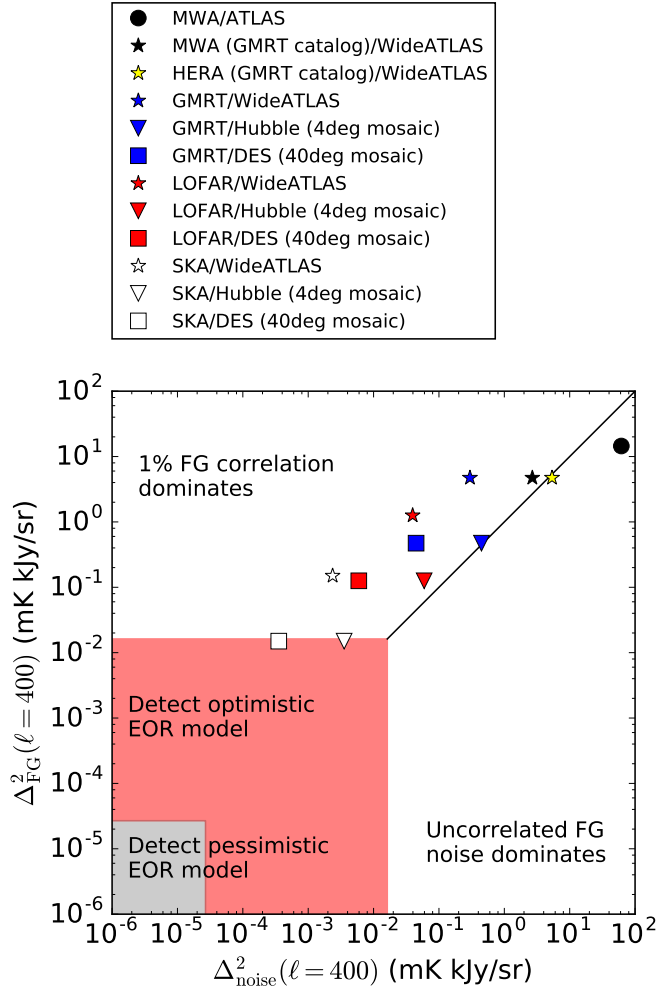- □ SKA/DES (40deg mosaic)

FIG. 15.— Foreground cross spectrum due to percent-level geometric correlations (y-axis) due to sample variance noise due to the uncorrelated component of foregrounds (x-axis) for a number of possible radio–IR intensity mapping survey pairs. This work (black circle) is the only one limited by sample variance noise, even a nominal improvement using the deeper GMRT catalog becomes foreground correlation-limited. Deeper radio (e.g., LOFAR and SKA) and IR (e.g., Hubble and DES mosaics) are needed to give the foreground removal required to suppress the geometric foreground correlation below the optimistic hexpected EOR signal.

We observe that our MWA/ATLAS analysis (black circle) is the only experiment limited by sample variance noise of uncorrelated foregrounds. An improved analysis using the GMRT catalog for radio source subtraction from MWA images and widening the ATLAS survey to a wide 40° (black star) falls barely in the foreground correlation limited regime. A similar analysis using HERA instead of the MWA (yellow star) produces similar results, as HERA is not optimized for the wide field, high resolution radio surveys that broad band correlation experiments require.

LOFAR reaches within a factor of a few in power of the optimistic EOR cross spectrum model when correlated against the Hubble (red triangle) or DES (red square) mosaics, and the SKA coupled with the same IR surveys achieves a convincing detection of the optimistic model (white square and white triangle). None of the experiments considered can reach the pessimistic model, but excluding the optimistic model already begins to exclude

late reionization scenarios, and performing the same correlation experiment at higher redshifts can likely give more leverage on earlier ones.

Interestingly, we note that virtually all the intermediate experiments between this work and the ultimate detection will be limited by the geometric foreground correlations. Motivated by our results in Sec. 4.2, we have assumed here that the coherence remains at the percent-level irrespective of mask depth, though more work is needed to model this effect more completely and self-consistently. Luckily, the negative sign of the predicted 21 cm–Lyα correlation from the EOR should clearly distinguish itself from this foreground effect, meaning there is still hope to make the measurements simulated by Mao (2014); Fernandez et al. (2014).

6. DISCUSSION

Radio and infrared surveys are on the cusp of direct observations of the two sides of the EOR. Deep infrared surveys are beginning to probe the bright end of the ionizing galaxy population itself, while low frequency radio surveys are constraining the 21 cm brightness of the neutral IGM. New radio and infrared instruments are all but assured to yield the first direct constraints on the astrophysics of the EOR, and combining these measurements will be crucial to confirm these first detections and extend their science reach. In particular, measurements of the predicted anticorrelation between redshifted radio and Lyα emission can shed light on properties and statistics of the ionizing sources which 21 cm measurements are not directly sensitive to, while simultaneously yielding redshift information on the near infrared background for the first time.

3D correlation measurements are an ultimate goal, but making infrared maps with sufficient spectral resolution over large enough fields to match the comoving volumes probed by low frequency 21 cm experiments is challenging and expensive. In contrast, 2D (i.e., broad band) infrared maps can be measured by many existing ground- and space-based observatories, and are a natural first step towards 3D experiments.

We have shown that foregrounds pose two significant challenges for 2D intensity mapping correlation experiments: (1) simple geometric effects result in percent-level correlations between source fluxes in these two bands, even if their luminosities are uncorrelated; and (2) the largely uncorrelated radio and infrared foreground residuals contribute a sample variance noise to the cross spectrum. The first challenge demands better foreground masking and subtraction, while the second requires measurements over large fields of view with many independent samples to average down uncorrelated radio and infrared power.

In the first part of this paper we searched for correlations between radio and infrared catalogs fluxes. We began with the 185 MHz MWA catalog and the 4.5 μm WISE catalog; any foreground correlation between these bands would limit 21cm–Hα correlation analyses at $z \sim 7$. We detect a $\sim 5\%$ correlation at $> 6\sigma$ significance after masking 4.5 μm sources down to 1 mJy, which corresponds to the flux below which extragalactic sources dominate the catalog. We reproduce this observed 185 MHz–4.5 μm correlation in simulations, confirming that it is sourced by starforming galaxies rather

than AGN. The narrow radio and infrared luminosity functions of the former make distance a much stronger effect than for AGN in calculating fluxes from luminosities.

We then performed 850 nm observations of the MWA field using the ATLAS telescope, and searched for correlations between 185 MHz and 850 nm catalog fluxes in MWA and ATLAS catalogs. Any correlation between these bands would limit 21 cm–Lyα analyses. We observed a weak correlation at the $\sim 3\sigma$ level which grew stronger towards our deepest 850 nm mask depth of 0.1 mJy. This is consistent with our finding from Sloan Digital Sky Survey data that at 850 nm, stars still dominate source counts above this flux. Deeper near infrared studies are needed to study how these correlations depend on mask depth to quantify what level of foreground removal is required to mitigate them. For now, we assume that radio–infrared flux correlations are at the percent-level for the purpose of weighing foreground flux correlations versus foreground sample variance noise in our experimental design study.

In the second part of this paper we analyzed the residual 185 MHz and 850 nm foregrounds in MWA and ATLAS images after subtracting radio sources and masking infrared sources. We computed power spectra of the band averaged MWA cubes and found they agree with the $k_{\parallel} = 0$ bin of the 3D power spectra of Beardsley et al. (2016) after conversion from 3D to 2D units. Despite the fact that the MWA image is not thermal noise-limited, we find that foreground subtraction improves significantly (by a factor of 2 in power) when the observations are spread over two weeks instead of a single night, consistent with ionosphere-induced calibration errors. This implies that if multiple fields are to be observed, the observing time should be rotated between them on a nightly cadence.

We then used the optimal quadratic estimator of Tegmark (1997) to compute the power spectrum of our 850 nm ATLAS images after masking sources at 2″ resolution. These results agree well over $300 < \ell < 4000$ with the 1.1 μm power spectrum measured by (Zemcov et al. 2014) using CIBER, a sounding rocket with 7″ resolution and $\sim 1°$ field of view. Our larger field of view permits significantly improved precision at $\ell < 1000$, and while our slightly increased resolution *should* permit better masking, the lack of foreground reduction indicates that the power spectrum has hit a floor at these angular scales. Mitchell-Wynne et al. (2015) show that mild improvement is possible using the Hubble's significantly improved 0.1″ resolution, but argue that further improvement at these angular scales is stymied by Galactic dust. They demonstrate that significant improvement is only possible at larger ℓ, angular scales largely inaccessible to 21 cm observatories focusing on power spectrum estimation such as PAPER, the MWA, and HERA. Taking advantage of these infrared foreground reductions to reduce foreground noise and correlations in 21 cm–Lyα correlation experiments will require the higher resolution images of LOFAR and SKA-LOW.

Turning to cross spectrum measurements, we simulate the signal loss in 2D versus 3D EOR intensity mapping experiments. We find that 2D experiments can recover of order 100% the signal at $k < 0.2\,\mathrm{Mpc}^{-1}$, but that power on shorter length scales suppressed by a factor of $10 - 10^2$ in power due to averaging out of of $k_{\parallel} > 0$ modes.

Finally, we used our foreground subtracted and masked MWA and ATLAS images to set the first limits on the 21 cm–Lyα cross spectrum from $z \sim 7$. We adapted the optimal quadratic estimator to the cross spectrum case to properly account for both the non-uniform uv sampling of radio interferometers and the non-uniform image sampling of masked infrared images. Our results are consistent with zero correlation, and our strictest upper limit on the EOR cross spectrum is $\Delta^2 < 181\,(\mathrm{kJy/sr} \cdot \mathrm{mK})$ (95%) at $\ell \sim 800$.

A percent-level correlation between radio and infrared foreground fluxes, of the sort suggested by our catalog correlation analyses, remains below this upper limit, but will be crucial for future experiments. We weigh the impact of a percent-level radio–infrared foreground correlation against the sample variance noise due to their uncorrelated part for several possible future experiments. The simplest improvement we consider is increasing the ATLAS survey area by a factor of ~ 10 and subtracting radio fluxes below the MWA's confusion limit using the GMRT catalog. We predict that the sensitivity boost of these improvements should reveal the cross spectrum floor due to the percent-level geometric flux correlations. Pushing down this floor requires foreground reduction due to higher resolution radio surveys such as LOFAR and SKA-LOW and infrared surveys such as Hubble and the Dark Energy Survey, though future work is needed to better understand these geometric flux correlations at very deep flux cuts.

We conclude that detecting the predicted EOR anticorrelation between redshifted 21 cm and Lyα emission in 2D (i.e., broad band) images is close to being within reach of current surveys, and future analyses using SKA-LOW can begin to significantly constrain EOR parameter space. These 2D experiments will be valuable complements to future 3D correlation analyses using infrared cubes measured using the proposed SPHEREx and Cosmic Dawn Intensity Mapper telescopes. These 3D surveys exchange foreground challenges for sensitivity and cost challenges, but will eventually probe the short spatial scales inaccessible to 2D surveys. The decade of direction observation of the EOR is upon us, and correlation experiments will help us move from the era of detection to the era of astrophysics.

APPENDIX

A. POWER SPECTRUM OF PHOTON SHOT NOISE

In Sec. 5.2 we measure the maximum airglow to be $I_{\mathrm{air}} = 5 \times 10^3$ kJy/sr, and in this appendix we calculate the power spectrum of this photon shot noise. We must observe that the mean number of photons collected by a pixel during

each observation is $\langle N_{\mathrm{ph}} \rangle = I_{\mathrm{air}} A t_{\mathrm{int}} \Delta f d\theta^2 / hf$, where $A = (0.5\,\mathrm{m})^2$ is the collecting area of ATLAS, $t_{\mathrm{int}} = 30\,\mathrm{sec}$, Δf and f are the frequency bandwidth and center frequency of I band, and $d\theta$ is the pixel size. The passband has $\Delta\lambda = 150\,\mathrm{nm}$ and $\lambda = 800\,\mathrm{nm}$.

The shot noise contribution to the power spectrum is given by

$$C_{\mathrm{IR,shot}}(\vec{\ell}) = \left\langle \left| \sum_{m,n} I_{\mathrm{shot}}(m,n) e^{-2\pi i(ma+nb)/N} \right|^2 \right\rangle \frac{d\theta^2}{N^2} \tag{A1}$$

where $I_{\mathrm{shot}}(m,n) \equiv I(m,n) - \langle I(m,n) \rangle$ denotes the photon shot noise contribution to pixel (m,n), and N is the number of pixels on each side of the square image. Then using the fact that the shot noise is uncorrelated between different pixels, we find

$$C_{\mathrm{IR,shot}}(\vec{\ell}) = \sum_{m,n} \langle I_{\mathrm{shot}}^2(m,n) \rangle \frac{d\theta^2}{N^2} \tag{A2}$$

Note that $I(m,n) = N_{\mathrm{ph}}(m,n) hf / \Delta f A t_{\mathrm{int}} d\theta^2$ and $\langle N_{\mathrm{ph}}^2 \rangle = \langle N_{\mathrm{ph}} \rangle$, so we have

$$C_{\mathrm{IR,shot}}(\vec{\ell}) = \langle N_{\mathrm{ph}} \rangle \left(\frac{hf}{\Delta f A t_{\mathrm{int}} d\theta^2} \right)^2 d\theta^2 \tag{A3}$$

$$C_{\mathrm{IR,shot}}(\vec{\ell}) = \frac{I_{\mathrm{air}} h\lambda}{\Delta\lambda A t_{\mathrm{int}}} \tag{A4}$$

which gives $\Delta(\ell = 10^3) = 6 \times 10^{-2}\,\mathrm{kJy/sr}$ for our deep fields with $t_{\mathrm{int}} = 2.5\,\mathrm{min}$.

B. RELATION BETWEEN THE POWER SPECTRUM OF IMAGE CUBES AND BROADBAND IMAGES

We focus in this paper on the spherical power spectrum of broadband images, C_ℓ, instead of that of image cubes, $P(\vec{k})$, as 21 cm observations have focused on. Here we work out the approximate relation between the two over small fields of view (i.e., for large ℓ) to facilitate comparison with past 21 cm power spectrum results. In particular, we calculate the scaling factor B relating the purely transverse modes of the power spectrum $P(k_\perp, k_\parallel = 0)$ of a image cube $I(\theta_x, \theta_y, f)$ to the spherical power spectrum of a broad band image C_ℓ as

$$P(k_\perp, k_\parallel = 0) = B C_{\ell(k_\perp)} \tag{B1}$$

Using the fourier transform convention discussed in Sec. 2, the left side of the equation is given by

$$P(k_\perp, k_\parallel = 0) = \frac{1}{N_\perp^2 N_\parallel dV} \langle |\tilde{I}(k_x, k_y, k_\parallel = 0)|^2 \rangle \tag{B2}$$

where $N_\perp \equiv N_x = N_y$ is the number of pixels in each of the two transverse dimensions of the image cube, and $N_\parallel$ is the number of pixels in the line of sight (ie, frequency) dimension. The comoving pixel volume is $dV = (D_c d\theta)^2 (\Delta D_c / N_\parallel)$, where D_c is the line of sight comoving distance from the present day to the center of the cube, and ΔD_c is the comoving line of sight thickness of the cube. Lastly, recall that $k_\perp$ is related to k_x and k_y as $k_\perp = \sqrt{k_x^2 + k_y^2}$.

Now substituting the definition of the fourier transform, we find

$$P(k_\perp, k_\parallel = 0) = \frac{1}{N_\perp^2 N_\parallel dV} \left\langle \left| dV \sum_{\theta_x, \theta_y, f} I(\theta_x, \theta_y, f) e^{iD_c(k_x\theta_x + k_y\theta_y)} \right|^2 \right\rangle \tag{B3}$$

Simplifying and writing this in terms of the broadband image $I_{\Delta f}(\theta_x, \theta_y) \equiv \frac{1}{N_\parallel} \sum_f I(\theta_x, \theta_y, f)$, we find

$$P(k_\perp, k_\parallel = 0) = (D_c^2 \Delta D_c) \frac{d\theta^2}{N_\perp^2} \left\langle \left| \sum_{\theta_x, \theta_y} I_{\Delta f}(\theta_x, \theta_y) e^{iD_c(k_x\theta_x + k_y\theta_y)} \right|^2 \right\rangle \tag{B4}$$

Now denote $k_x = a \cdot dk$, $k_y = b \cdot dk$, $\theta_x = m \cdot d\theta$, and $\theta_y = n \cdot d\theta$, where $dk = 1/N_\perp D_c d\theta$.

$$P(k_\perp(\ell(a,b)), k_\parallel = 0) = (D_c^2 \Delta D_c) \frac{d\theta^2}{N_\perp^2} \left\langle \left| \sum_{m,n} I_{\Delta f}(m,n) e^{2\pi i(am+bn)/N_\perp} \right|^2 \right\rangle \tag{B5}$$

Comparing with Equations 5, 6, and 7, we see that $B \equiv P(k_\perp, k_\parallel = 0)/C_{\ell(k_\perp)} = D_c^2 \Delta D_c$ and $\ell = D_c k_\perp$.

C. EXTENDING THE OPTIMAL QUADRATIC POWER SPECTRUM ESTIMATOR TO THE CROSS SPECTRUM CASE

The optimal quadratic estimator formalism presented in Sec. 5.2 was constructed to estimate the power spectrum of an image with arbitrary pixel sampling and noise properties. In this section we extend this formalism to achieve the same advantages in cross spectrum measurements.

Consider measurements at two bands over the same set of pixels on the sky, $\mathbf{x}_1$ and $\mathbf{x}_2$, each column vectors with N elements. Let us combine these together into a single column vector containing all measurements as $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$, whose covariance is given by

$$\mathbf{C} \equiv \langle \mathbf{x}\mathbf{x}^\dagger \rangle - \langle \mathbf{x} \rangle \langle \mathbf{x}^\dagger \rangle = \begin{pmatrix} \mathbf{C}_1 & \mathbf{C}_{12} \\ \mathbf{C}_{12}^\dagger & \mathbf{C}_2 \end{pmatrix} \tag{C1}$$

and

$$\frac{d\mathbf{C}}{dp_{12}^\alpha} = \begin{pmatrix} \mathbf{0} & \mathbf{C}_{,\alpha} \\ \mathbf{C}_{,\alpha}^\dagger & \mathbf{0} \end{pmatrix} \tag{C2}$$

$\mathbf{C}_1$ and $\mathbf{C}_2$ depend only on the auto power spectra of the different fields; only $\mathbf{C}_{12}$ depends on the cross spectrum. Said another way, $C_{,\alpha}$ is the same matrix used in Sec. 5.2, but used here in the off-diagonal parts of $d\mathbf{C}/dp_{12}^\alpha$ so as to capture the cross products between the two fields[13]. And as before, the unnormalized estimator q_α of the power in band α is given by

$$q_\alpha = \frac{1}{2}(\mathbf{x} - \langle \mathbf{x} \rangle)^t \mathbf{C}^{-1} \frac{d\mathbf{C}}{dp_{12}^\alpha} \mathbf{C}^{-1}(\mathbf{x} - \langle \mathbf{x} \rangle) \tag{C3}$$

and the elements of the Fisher matrix are

$$\mathbf{F}_{\alpha\beta} = \frac{1}{2} \mathrm{tr}\left(\mathbf{C}^{-1} \frac{d\mathbf{C}}{dp_{12}^\alpha} \mathbf{C}^{-1} \frac{d\mathbf{C}}{dp_{12}^\beta} \right) \tag{C4}$$

In the case where the correlation between the two fields is expected to be weak, and we are primarily interested in setting an upper limit, we can get a significant speedup by approximating $\mathbf{C}_{12} \approx 0$ in our guess covariance. The expressions for q_α and $F_{\alpha\beta}$ then simplify to:

$$q_\alpha \approx (\mathbf{x}_1 - \langle \mathbf{x}_1 \rangle)^t \mathbf{C}_1^{-1} \mathbf{C}_{,\alpha} \mathbf{C}_2^{-1}(\mathbf{x}_2 - \langle \mathbf{x}_2 \rangle) \tag{C5}$$

$$\mathbf{F}_{\alpha\beta} \approx \mathrm{tr}\left(\mathbf{C}_1^{-1} \mathbf{C}_{,\alpha} \mathbf{C}_2^{-1} \mathbf{C}_{,\beta} \right) \tag{C6}$$

REFERENCES

Ali, Z. S., et al. 2015, ApJ, 809, 61
Alvarez, M. A., Finlator, K., & Trenti, M. 2012, ApJ, 759, L38
Amorín, R., et al. 2017, Nature Astronomy, 1, 0052
Atek, H., et al. 2015, ApJ, 814, 69
Beardsley, A. P., Morales, M. F., Lidz, A., Malloy, M., & Sutter, P. M. 2015, ApJ, 800, 128
Beardsley, A. P., et al. 2013, MNRAS, 429, L5
—. 2016, ArXiv e-prints
Bernard, S. R., et al. 2016, ApJ, 827, 76
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bertin, E., Mellier, Y., Radovich, M., Missonnier, G., Didelon, P., & Morin, B. 2002, in Astronomical Society of the Pacific Conference Series, Vol. 281, Astronomical Data Analysis Software and Systems XI, ed. D. A. Bohlender, D. Durand, & T. H. Handley, 228
bij de Vaate, J., et al. 2011, in General Assembly and Scientific Symposium, 2011 XXXth URSI, 1–4

Bouwens, R. 2016, in Astrophysics and Space Science Library, Vol. 423, Understanding the Epoch of Cosmic Reionization: Challenges and Progress, ed. A. Mesinger, 111
Bouwens, R. J., Oesch, P. A., Illingworth, G. D., Ellis, R. S., & Stefanon, M. 2016a, ArXiv e-prints
Bouwens, R. J., et al. 2011, ApJ, 737, 90
—. 2014, ApJ, 793, 115
—. 2015, ApJ, 803, 34
—. 2016b, ApJ, 830, 67
Bowler, R. A. A., Dunlop, J. S., McLure, R. J., & McLeod, D. J. 2017, MNRAS, 466, 3612
Bowler, R. A. A., et al. 2015, MNRAS, 452, 1817
Bowman, J. D., et al. 2013, PASA, 30, e031
Bradley, L. D., et al. 2012, ApJ, 760, 108
Calvi, V., et al. 2016, ApJ, 817, 120
Carroll, P. A., et al. 2016, MNRAS, 461, 4151
Chang, T.-C., Pen, U.-L., Bandura, K., & Peterson, J. B. 2010, Nature, 466, 463
Coe, D., Bradley, L., & Zitrin, A. 2015, ApJ, 800, 84
Comaschi, P., Yue, B., & Ferrara, A. 2016, MNRAS, 463, 3193
Cooray, A., et al. 2012, Nature, 490, 514
—. 2016, ArXiv e-prints
Cutri, R. M., et al. 2013, Explanatory Supplement to the AllWISE Data Release Products, Tech. rep.
Dark Energy Survey Collaboration et al. 2016, MNRAS, 460, 1270
de Jong, T., Klein, U., Wielebinski, R., & Wunderlich, E. 1985, A&A, 147, L6
DeBoer, D. R., et al. 2017, Publications of the Astronomical Society of the Pacific, 129, 045001
Dewdney, P., & Lazio, T. 2008, Radio Science Bulletin, 5

[13] One might object to our form of $d\mathbf{C}/dp_{12}^\alpha$, arguing that artificially limiting ourselves to cross products between the two fields is tantamount to throwing a significant amount of the information contained in the data sets, and thus our estimator cannot be optimal. There are certainly situations in which this would be the case. If we had some theory of how each field was related to the matter density field of the universe, then both the auto-products and cross-products contain similar information. However we take an empirical approach where we assume we know nothing about either field and want only to know their correlation properties. Only the cross products contain that information.

Dewdney, P. E., Hall, P. J., Schilizzi, R. T., & Lazio, T. J. L. 2009, Proceedings of the IEEE, 97, 1482

Di Matteo, T., Perna, R., Abel, T., & Rees, M. J. 2002, ApJ, 564, 576

Dillon, J. S., Liu, A., & Tegmark, M. 2013, Phys. Rev. D, 87, 043005

Dillon, J. S., et al. 2014, Phys. Rev. D, 89, 023002

—. 2015, Phys. Rev. D, 91, 123011

Dodelson, S., et al. 2002, ApJ, 572, 140

Doré, O., et al. 2014, ArXiv e-prints

—. 2016, ArXiv e-prints

Dunlop, J. S., et al. 2013, MNRAS, 432, 3520

Eisenstein, D. J., et al. 2011, AJ, 142, 72

Ellis, R. S., et al. 2013, ApJ, 763, L7

Ewall-Wice, A., et al. 2016, ApJ, 831, 196

Feng, C., Cooray, A., & Keating, B. 2017, ArXiv e-prints

Fernandez, E. R., Zaroubi, S., Iliev, I. T., Mellema, G., & Jelić, V. 2014, MNRAS, 440, 298

Finkelstein, S. L. 2016, PASA, 33, e037

Fixsen, D. J., Moseley, S. H., & Arendt, R. G. 2000, ApJS, 128, 651

Fu, H., et al. 2010, ApJ, 722, 653

Furlanetto, S., Oh, S., & Briggs, F. 2006, Physics Reports, 433, 181

Furlanetto, S. R., & Lidz, A. 2007, ApJ, 660, 1030

Gardner, J. P., et al. 2006, Space Science Reviews, 123, 485

Gong, Y., Cooray, A., Silva, M. B., Zemcov, M., Feng, C., Santos, M. G., Dore, O., & Chen, X. 2017, ApJ, 835, 273

Gong, Y., Silva, M., Cooray, A., & Santos, M. G. 2014, ApJ, 785, 72

Grogin, N. A., et al. 2011, ApJS, 197, 35

Hall, P. J. 2005, The Square Kilometre Array: An Engineering Perspective

Helgason, K., Ricotti, M., Kashlinsky, A., & Bromm, V. 2016, MNRAS, 455, 282

Helou, G., Soifer, B. T., & Rowan-Robinson, M. 1985, ApJL, 298, L7

Henden, A. A., Templeton, M., Terrell, D., Smith, T. C., Levine, S., & Welch, D. 2016, VizieR Online Data Catalog, 2336

Heneka, C., Cooray, A., & Feng, C. 2016, ArXiv e-prints

High, F. W., Stubbs, C. W., Stalder, B., Gilmore, D. K., & Tonry, J. L. 2010, PASP, 122, 722

Huang, K.-H., et al. 2016, ApJ, 817, 11

Illingworth, G. D., et al. 2013, ApJS, 209, 6

Intema, H. T., Jagannathan, P., Mooley, K. P., & Frail, D. A. 2017, A&A, 598, A78

Jacobs, D. C., et al. 2015, ApJ, 801, 51

Kashlinsky, A., Arendt, R. G., Ashby, M. L. N., Fazio, G. G., Mather, J., & Moseley, S. H. 2012, ApJ, 753, 63

Kashlinsky, A., Arendt, R. G., Mather, J., & Moseley, S. H. 2005, Nature, 438, 45

—. 2007, ApJ, 657, L131

Kurcz, A., Bilicki, M., Solarz, A., Krupa, M., Pollo, A., & Małek, K. 2016, A&A, 592, A25

Lidz, A., Zahn, O., Furlanetto, S. R., McQuinn, M., Hernquist, L., & Zaldarriaga, M. 2009, ApJ, 690, 252

Liu, A., & Tegmark, M. 2011, Phys. Rev. D, 83, 103006

Livermore, R. C., Finkelstein, S. L., & Lotz, J. M. 2017, ApJ, 835, 113

Mao, X.-C. 2014, ApJ, 790, 148

Masui, K. W., et al. 2013, ApJ, 763, L20

Mauch, T., & Sadler, E. M. 2007, MNRAS, 375, 931

McLeod, D. J., McLure, R. J., & Dunlop, J. S. 2016, MNRAS, 459, 3812

McLeod, D. J., McLure, R. J., Dunlop, J. S., Robertson, B. E., Ellis, R. S., & Targett, T. A. 2015, MNRAS, 450, 3032

Mesinger, A., ed. 2016, Astrophysics and Space Science Library, Vol. 423, Understanding the Epoch of Cosmic Reionization

Mitchell-Wynne, K., et al. 2015, Nature Communications, 6, 7945

Morales, M. F., & Wyithe, J. S. B. 2010, Annual Reviews of Astronomy and Astrophysics, 48, 127

Neben, A. R., et al. 2016, ApJ, 826, 199

Nikutta, R., Hunt-Walker, N., Nenkova, M., Ivezić, Ž., & Elitzur, M. 2014, MNRAS, 442, 3361

Oesch, P. A., et al. 2013, ApJ, 773, 75

Offringa, A. R., et al. 2015, PASP, 32, e008

Park, J., Kim, H.-S., Wyithe, J. S. B., & Lacey, C. G. 2014, MNRAS, 438, 2474

Parsons, A. R., et al. 2014, ApJ, 788, 106

Patra, N., et al. 2017, ArXiv e-prints

Planck Collaboration et al. 2016, A&A, 594, A13

Pober, J. C., et al. 2014, ApJ, 782, 66

—. 2015, ApJ, 809, 62

Prandoni, I., & Seymour, N. 2015, Advancing Astrophysics with the Square Kilometre Array (AASKA14), 67

Prescott, M., et al. 2016, MNRAS, 457, 730

Pritchard, J. R., & Loeb, A. 2012, Reports on Progress in Physics, 75, 086901

Pullen, A. R., Doré, O., & Bock, J. 2014, ApJ, 786, 111

Repp, A., Ebeling, H., & Richard, J. 2016, MNRAS, 457, 1399

Ricker, G. R., et al. 2014, in Proc. SPIE, Vol. 9143, Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave, 914320

Roberts-Borsani, G. W., et al. 2016, ApJ, 823, 143

Robertson, B. E., et al. 2013, ApJ, 768, 71

Schmidt, K. B., et al. 2014, ApJ, 786, 57

SDSS Collaboration et al. 2016, ArXiv e-prints

Seymour, N., et al. 2007, ApJS, 171, 353

Silva, M. B., Santos, M. G., Gong, Y., Cooray, A., & Bock, J. 2013, ApJ, 763, 132

Smit, R., et al. 2014, ApJ, 784, 58

Spergel, D., et al. 2013, ArXiv e-prints

Sullivan, I. S., et al. 2012, ApJ, 759, 17

Sullivan, P. W., & Simcoe, R. A. 2012, PASP, 124, 1336

Tegmark, M. 1997, Phys. Rev. D, 55, 5895

Tegmark, M., & Zaldarriaga, M. 2002, Phys. Rev. D, 66, 103508

Tegmark, M., et al. 2002, ApJ, 571, 191

Tingay, S. J., et al. 2013, PASA, 30

Tonry, J. L. 2011, PASP, 123, 58

Tonry, J. L., et al. 2012, ApJ, 750, 99

Trenti, M., et al. 2011, ApJ, 727, L39

van Haarlem, M. P., et al. 2013, A&A, 556, A2

Wilkins, S. M., Bouwens, R. J., Oesch, P. A., Labbé, I., Sargent, M., Caruana, J., Wardlow, J., & Clay, S. 2016, MNRAS, 455, 659

Willott, C. J., Rawlings, S., Jarvis, M. J., & Blundell, K. M. 2003, MNRAS, 339, 173

Wright, E. L., et al. 2010, AJ, 140, 1868

Wyithe, J. S. B., & Morales, M. F. 2007, MNRAS, 379, 1647

Xu, C., Lisenfeld, U., & Voelk, H. J. 1994, A&A, 285, 19

Yatawatta, S., et al. 2013, A&A, 550, A136

Yue, B., Ferrara, A., & Salvaterra, R. 2016, ArXiv e-prints

Yun, M. S., Reddy, N. A., & Condon, J. J. 2001, ApJ, 554, 803

Zemcov, M., et al. 2014, Science, 346, 732