

FOREGROUND AND SENSITIVITY ANALYSIS FOR BROAD BAND 21 CM–LY α AND 21 CM–H α CORRELATION EXPERIMENTS PROBING THE EPOCH OF REIONIZATION

ABRAHAM R. NEBEN¹, BRIAN STALDER², JOHN L. TONRY², JACQUELINE N. HEWITT¹

¹MIT Kavli Institute, Massachusetts Institute of Technology, Cambridge, MA, 02139 USA and

²Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu HI 96822

Draft version February 1, 2017

ABSTRACT

NEED TO WRITE THE ABSTRACT

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 (Furlanetto et al. 2006; Morales & Wyithe 2010; Pritchard & Loeb 2012). 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. 2016) is expected to detect and characterize the EOR power spectrum in the coming years. Ultimately, the Square Kilometer Array (SKA) will image the EOR over redshift, revealing the detailed hydrogen reionization history of the universe (done).

At the same time, deeper galaxy redshift surveys are beginning to constrain the reionizing sources themselves. Hubble deep field observations (Bouwens et al. 2011; Illingworth et al. 2013; Dunlop et al. 2013) and cluster lensing surveys are finding tens of galaxies at $6 < z < 10$ down to UV magnitudes of $M_{AB} \sim -17$, and extremely wide surveys are searching for the rare bright ones (Schmidt et al. 2014; Trenti et al. 2011; Bradley et al. 2012). However, current models require the contribution of far fainter galaxies down to $M_{AB} \sim -13$ (Robertson et al. 2013; Alvarez et al. 2012) in order to agree with optical depth measurements (Planck Collaboration et al. 2016) from the cosmic microwave background. Deeper observations with the James Webb Space Telescope (JWST) (Gardner et al. 2006) and the Wide Field Infrared Space Telescope (WFIRST) (Spergel et al. 2013) will be needed to probe this crucial faint population (Atek et al. 2015).

Infrared intensity mapping offers several advantages compared to surveys. Power spectrum analyses can be sensitive to an EOR component even if the signal-to-noise in individual pixels small, and instead of being limited to the brightest galaxies, intensity mapping is sensitive to the cumulative light from *all* sources. Indeed, ionizing and Lyman-alpha radiation from EOR galaxies at $z \sim 6 – 8$ redshifts into the near infrared, 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. In fact, Cooray et al. (2012); Zemcov et al. (2014) argue that intrahalo light, consisting of tidally stripped stars dispersed throughout host halos, is the best explanation for the observed fluctuation excess over known EOR galaxy populations. This implies that even after significant foreground masking, EOR foreground emission in wide field infrared surveys is of order 10^4 times brighter than EOR emission at $1' – 60'$ scales.

Given these bright foregrounds, cross correlation with 21 cm maps may in fact be the *only* way to extract the diffuse EOR component of the near infrared background. The synergy is clear: the galaxies sourcing reionization generate strong Ly- α 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, sourcing an anticorrelation seen in EOR simulations by Silva et al. (2013); Heneka et al. (2016) and modeled analytically by Feng et al. (2017).

A similar anticorrelation on large scales is found by Lidz et al. (2009); Park et al. (2014) in simulations of 21 cm cross correlation with galaxy surveys, but 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 3' angular resolution of the MWA is of roughly the field of view of the Hubble Deep Field and the James Webb Space Telescope (JWST). It may be possible to cross-check the ionization environment of deep JWST sources by comparing the brightness temperature in the 21 cm map (Beardsley et al. 2015), but even after order ~ 100 hour integrations such detections will be near JWST limiting sensitivities Zackrisson et al. (2011).

In contrast, broad band intensity enables similar science with shallower observations (Fernandez et al. 2014; Mao 2014), though imperfect radio and infrared foreground subtraction will leak largely uncorrelated power into the cross correlation analysis which must be averaged out over large fields of view. Fortunately, the planned Transiting Exoplanet Survey Satellite (Ricker et al. 2014) and the proposed SPHEREx satellite mis-

mention class
the probe? "mission"

2 In intro it would be good to point out difference between 2D and 3D experiments and we are doing 2D in this paper.
Neben et al.

sion (Doré et al. 2014, 2016) would image the entire sky in the near infrared, and many ground-based near infrared surveys with few degree fields 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) are coming online. In the low frequency radio, the MWA has performed a deep survey of 400 square degrees at high galactic latitude (Beardsley et al. 2016), and HERA will survey ~ 2000 square degrees along a zenith strip (Dillon et al. 2015). 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.

With wide and deep near infrared and low frequency radio surveys happening now and on the horizon, we study in this paper the real world prospects of detecting the anti-correlation of diffuse 21 cm, Ly α , and H α emission from 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 in such measurements, demonstrating that distance variation of the sources combined with their finite luminosity distribution introduces slight positive correlations which overpower the cosmological signal unless significant masking and subtraction are done. In Sec. 5, we study how best to mask and subtract radio and infrared foregrounds on real world datasets and quantify the foreground residuals with their power spectra. We set the first limits on the broad band 21 cm–Ly α cross spectrum at $z = 7$ using data from the MWA and ATLAS, and compare the sensitivities of future experiments, illustrating what it will take to realize this measurement.

2. POWER SPECTRUM AND CORRELATION CONVENTIONS

2.1. Power spectrum definitions

We define the 3D power spectrum $P(\vec{k})$ of the image cube $I(\vec{x})$ following Ewall-Wice et al. (2014) as

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

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

$$\text{units are } k = \text{Mpc}^{-1} \quad \tilde{I}(\vec{k}) = dV \sum_{\vec{x}} I(\vec{x}) e^{i\vec{k} \cdot \vec{x}}, \quad (2)$$

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

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

$$\text{where? Zemcov Cooray} \quad C(\vec{\ell}) = \frac{\langle |\tilde{I}(\vec{\ell})|^2 \rangle}{\Omega} \quad \text{units of } \text{K ster} \quad (3)$$

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

$$\text{units are K ster} \quad \tilde{I}(\vec{\ell}) = d\Omega \sum_{\vec{\theta}} I(\vec{\theta}) e^{i\vec{\ell} \cdot \vec{\theta}}, \quad (4)$$

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, we find¹

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

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

$$\ell_x = 2\pi a / Nd\theta \quad (6)$$

$$\ell_y = 2\pi b / Nd\theta \quad (7)$$

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

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 A 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

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

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 co-moving 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

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

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

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 anti-correlated, uncorrelated, or correlated the two fields are. It is thus often renormalized as

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

where c is known as the coherence and is insensitive to a simple rescaling of either field. However, large foreground residuals in either field will substantially bias the

¹ Note that the normalization of $d\theta^2/N^2$ has been missstated 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).

I think before this section we need to motivate why measuring foreground correlation is important.

Foreground and sensitivity analysis for 21cm/Infrared studies

3

coherence towards zero (Lidz et al. 2009; Furlanetto & Lidz 2007), whereas they merely contribute a zero-mean noise to the cross spectrum.

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 $\sim 30^\circ \times (150 \text{ MHz}/f)$ full-width-at-half-max and steerable in few degree increments with a delay line beamformer. We use low frequency observations of a quiet field centered at $(\text{RA}, \text{Dec}) = (0^\circ, -27^\circ)$ J2000 recorded over 30.72 MHz bandwidth centered at 186 MHz, corresponding to $z = 6.0 - 7.3$. In Sec. 5.1 we examine several datasets from the second half of 2013 to study how much data is necessary to best mitigate foregrounds, and we summarize here the initial flagging, calibration, and imaging performed by the MWA pipeline prior to our processing.

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 Deconvolution² (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. We flag the upper and lower 80 kHz channels in each of 24 coarse channels across the band to mitigate aliasing, and average the remaining channels in frequency to make a broad band image.

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 in time to allow more leverage on estimating the magnitude of the noise and the noise bias. We average the cubes in frequency, rotate them so the EORO field center lies at the north pole, and project the pixels into the xy plane, resulting in an orthographic projection to the plane tangent to $(\text{RA}, \text{Dec}) = (0^\circ, -27^\circ)$.

3.2. IR Observations

ATLAS is a 0.5 m ($f/2.0$) telescope (Tonry 2011) in Hawaii designed to perform a wide field sky survey for near earth asteroids. The detector is a $10,500 \times 10,500$ STA-1600 CCD array, with a pixel scale of $1.86''$, with an overall field of view of 5.5° . We observe in the i band centered at 810 nm with full width at half max 150 nm, corresponding to $z = 5.1 - 6.3$. While this redshift range doesn’t exactly match that of our radio observations, it overlaps sufficiently for our purpose of characterizing the effects of noise foregrounds in 21cm-IR cross correlation experiments, and setting the first limits on the cross spectrum of residual foregrounds.

ASK BRIAN TO WRITE A FEW SENTENCES ABOUT THE ATLAS CALIBRATION AND IMAGING PIPELINE

We perform two separate observing campaigns, which we illustrate in Fig. 1. We first perform a wide survey

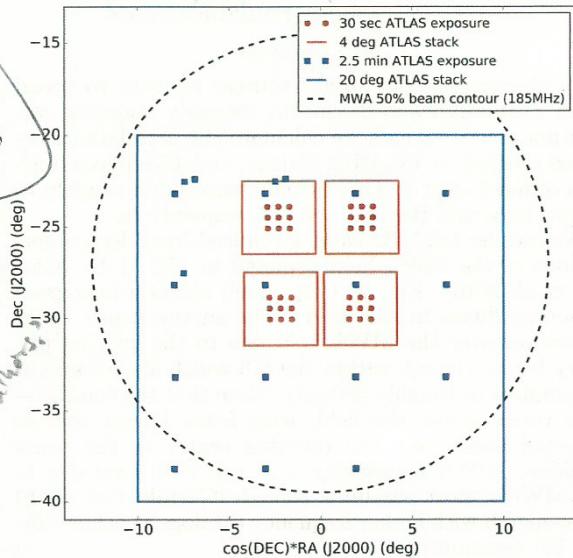


FIG. 1.— The MWA deep integration field (black dashed circle) is shown relative to our two ATLAS surveys. Blue markers show the observation centers of our wide ATLAS survey aimed at studying foregrounds, and the large blue square shows the stacked image. Red circle markers show observation centers for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

to best characterize bright foregrounds. We raster scan a roughly $20^\circ \times 20^\circ$ grid with 5° spacing over the MWA field (dashed black circle), integrating for 2.5 min at each pointing (blue square markers). The observations are conducted between 2016/09/07 22:00 and 2016/09/08 00:50 Hawaii-Aleutian Standard Time, when the moon was 36% illuminated. We then use `swarp`³ (Bertin et al. 2002) to stack all these frames over 20° orthographic field centered on $(\text{RA}, \text{Dec}) = (0^\circ, -30^\circ)$ (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. We select four 5° fields positioned around the MWA beam peak for best cross correlation precision: $(\text{RA}, \text{Dec}) = (-2.5^\circ, -24.5^\circ), (2.5^\circ, -24.5^\circ), (-2.5^\circ, -29.5^\circ), (2.5^\circ, -29.5^\circ)$ (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 $4^\circ \times 4^\circ$ 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 later on the effects of airglow-induced diffuse backgrounds.

² <https://github.com/EoRImaging/FHD>

³ <http://www.astromatic.net/software/swarp>

False correlate could
be caused by beam (radio)
+ vignetting (optical)

4. POINT SOURCE FOREGROUNDS

4.1. Catalogs

To characterize the bright sources relevant to broad band 21 cm–Ly α and 21 cm–H α intensity mapping correlation measurements we calculate the correlations between catalogs at 185 MHz, 850 nm, and 4.5 μ m as a function of mask depth. These bands correspond roughly to 21 cm, Ly α , and H α at $z \sim 6 - 7$, 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, though within the full-width-at-half-max it is complete to roughly 300 mJy. Note that the completeness varies across the field, with fewer fainter sources detected away from the pointing center as the beam weakens. MWA astrometry is at the 2'–3' level due to the MWA's short baselines, 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 μ m catalog. We download the list of sources within the MWA field using the All Sky Search on the NASA/IPAC Infrared Science Archive⁴, and plot the histogram of source fluxes in Fig. 2) (center panel). This ALLWISE band is specified to be 95% complete to 88 μ Jy (15.7 AB mag), though it has slight sky coverage non-uniformities due to satellite coverage.

Lastly, we run SExtractor⁵ (Bertin & Arnouts 1996) on our wide 20° wide 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 Photometric All Sky Survey⁶ (Henden et al. 2016), finding matches $\sim 10\%$ of the time. 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 out 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 1 mJy.

We find the catalog is complete to roughly a mJy (right panel in Fig. 2).

4.2. Catalog radio–infrared flux correlations

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

Neben et al.

Do you use
sources below the
completeness limit?
Is it true that cause the
correlation from the
same redshifts?

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). Massive stars blow out ionized bubbles, generating radio free-free emission correlated with the ionizing flux. Some fraction of these high energy photons are absorbed by dust clouds and reprocessed into far-infrared (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 mask the cosmological 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. A last comment on our approach is that though searching for radio–infrared correlations on a source-by-source level would be valuable cross check, we lack the radio astrometry to do so⁷.

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

$$\text{Brackets around } c = \frac{\langle I_{\text{rad}} I_{\text{IR}} \rangle - \langle I_{\text{rad}} \rangle \langle I_{\text{IR}} \rangle}{\sqrt{(\langle I_{\text{rad}}^2 \rangle - \langle I_{\text{rad}} \rangle^2)(\langle I_{\text{IR}}^2 \rangle - \langle I_{\text{IR}} \rangle^2)}} \quad (12)$$

What is the resolution (pixel size) of the grid?

where the uncertainty due to sample variance is approximately $dc = N_{\text{pix}}^{-1/2}$, 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 vanishingly small. 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 μ m and 10^{-2} Jy at 850 nm, and find $c_{\text{MWA-WISE}} = 0.031 \pm 0.005$ and $c_{\text{MWA-ATLAS}} = 0.0086 \pm 0.005$. 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 μ m correlation (top left) and 185 MHz–850 nm correlation (top right) as a function of the maximum flux percentile allowed in the data (masking depth).

⁷ In their cross matching study of GHz radio sources with optical detections, (McMahon et al. 2002) find that 99% of cross-matches within 10" of each other (see their Fig. 8), of order the positional accuracy quoted by Carroll et al. (2016).

⁴ <http://irsa.ipac.caltech.edu>

⁵ <http://www.astromatic.net/software/sextractor>

⁶ <https://www.aavso.org/download-apass-data>

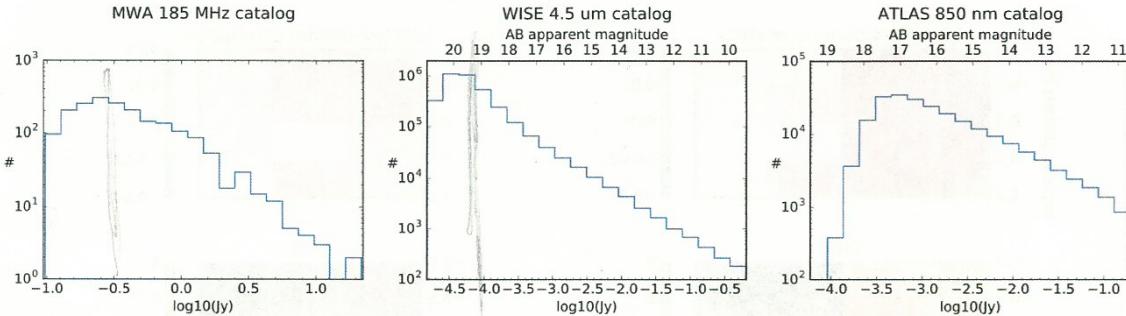


FIG. 2.— Histogram of source fluxes in the 185 MHz catalog (left), the 4.5 μm catalog (center), and the 850 nm catalog (right). The catalogs are complete to roughly 300 mJy, 0.05 mJy, and 0.3 mJy respectively.

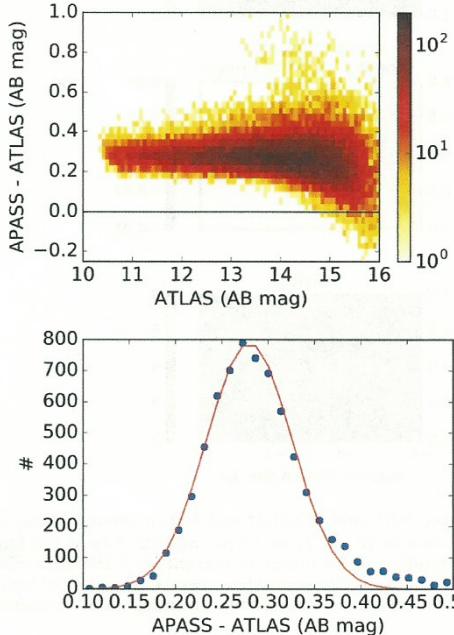


FIG. 3.— To improve the rough initial ATLAS calibration, we cross match ATLAS sources with those from APASS, and plot a 2D histogram (top) of the relative magnitude as a function of ATLAS magnitude, then fit a gaussian to the magnitude offset for sources brighter than 13 mag. We find that our roughly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

We plot the SNRs of these correlation measurements, taking the noise to be $1/\sqrt{N_{\text{pix}}}$ 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 μm 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. The 185 MHz and 850 nm catalogs exhibit a marginal 3σ correlation 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 digging deeper into the 4.5 μm catalog. We select the subset of sources detected in the WISE 3.4 μm , 4.5 μm , and 12 μm bands and plot them (Fig. 5, left panel) in the $W_{23} \equiv [4.6 \mu\text{m}] - [12 \mu\text{m}]$ versus $W_{12} \equiv [3.4 \mu\text{m}] - [4.6 \mu\text{m}]$ color-color space used by Wright et al. (2010) to illustrate the separation of different types of sources. Nikutta et al. (2014) study more quantitatively how these different sources separate in this color space, finding that stars separate clearly from all other sources 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 μm) in the same color-color space and observe this cut effectively cleanly excludes nearly all the stars. This explains why when masking this brightest 10% earlier, we observed zero correlation with 185 MHz sources, and why afterward the correlation suddenly emerges, after all, nearly all radio sources are non-stellar (i.e., extragalactic). Since

To further probe which mid infrared sources are responsible for this correlation, we make a rough cut to separate quasars and active galactic nuclei (AGNs) ($W_{12} > 0.16$) from starburst galaxies (SFs) ($W_{12} < 0.16$) (Nikutta et al. 2014; Kurcz et al. 2016). In Fig. 6 we plot the power spectrum of 185 MHz sources (left panel), the spectra of our AGN and SF cuts of 4.5 μm sources (center panel), and the coherence (i.e., correlation coefficient versus ℓ) of each of these cuts with the 185 MHz catalog (right panel). The AGN subset exhibits no significant correlation with the 185 MHz sources, while the SF cut exhibits a significant correlation rising from a few percent at $\ell \sim 7000$ to 13% at $\ell \sim 300$.

The fall of the correlation towards high ℓ is likely due in part to the MWA's 2' resolution at 185 MHz, corresponding to a maximum ℓ of roughly 6000, and in part to the similar fall in the 4.5 μm catalog power spectrum. Both the falling 4.5 μm 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_ℓ is approximately equal to the 3D matter power spectrum $P(k(\ell))$ convolved with a window function which depends on the redshift coverage

Does the correlation in extragalactic objects show up because they span a wider range of distances?

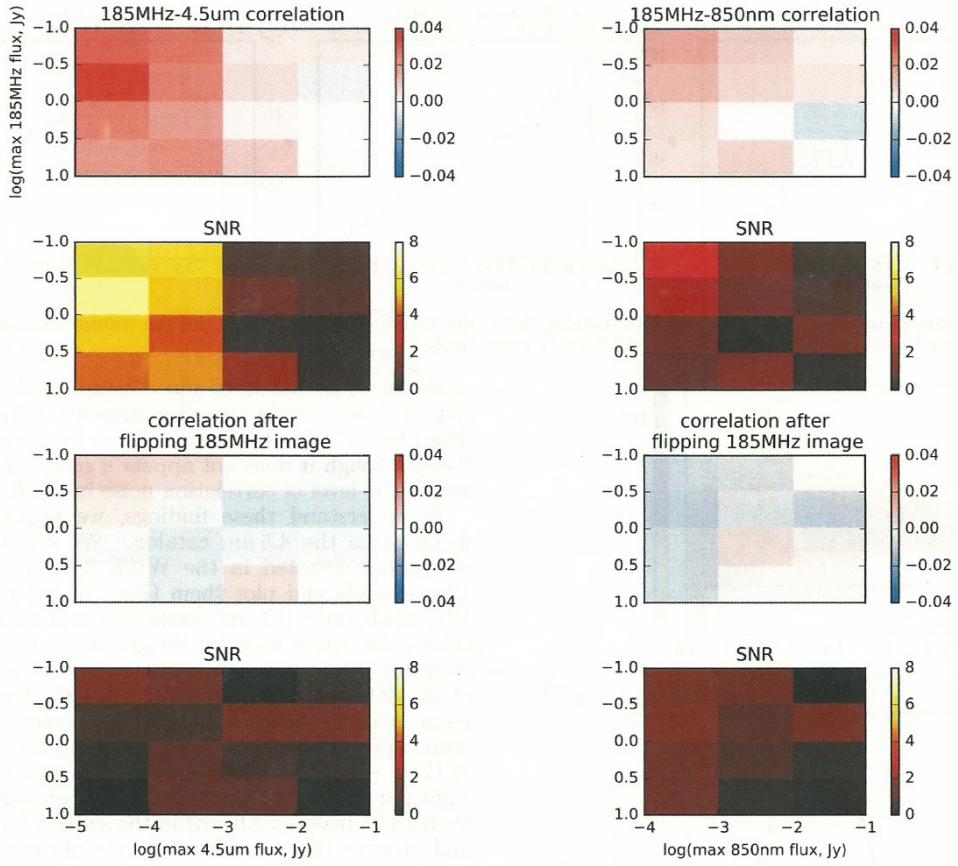


FIG. 4.— Image space correlation coefficient between 185 MHz and $4.5\mu\text{m}$ sources (top left), and 185 MHz and 850 nm sources (top right), both as a function of radio and infrared flux cuts. Excluding the top 10% of sources (down to $10^{-3.5}$ Jy at $4.5\mu\text{m}$ and 10^{-2} Jy at 850 nm) is a good proxy for removing the stars, below which we observe a 5% 185 MHz– $4.5\mu\text{m}$ correlation and a marginal percent level 185 MHz–850 nm correlation. In the second row we calculate the SNR in each bin, noting that neighboring bins are somewhat correlated. The bottom two rows show the correlations and SNRs after flipping the 185 MHz image in the calculation calculation, giving an independent estimate of the noise.

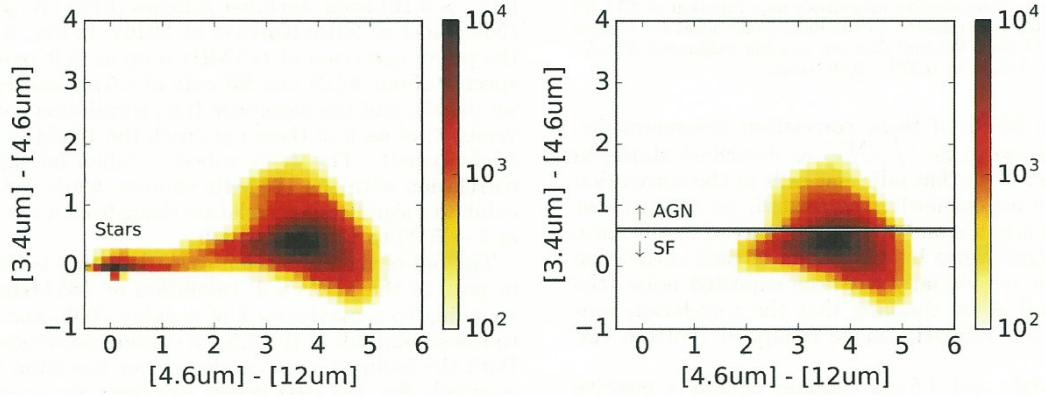


FIG. 5.— The ALLWISE sources are plotted in the color–color space of Wright et al. (2010) prior to any flux cuts (left panel), showing the stars near (0,0). Then after cutting out the brightest 10% of sources (fainter than 18.25 AB mag), the stars are cleanly removed. We roughly split the remaining sources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

and flux limit of the sample. The matter power spectrum is known to rise as k^1 for $k \lesssim 0.02 \text{ h/Mpc}$, then falls as k^{-3} . Galaxy surveys typically probe the regime just above 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 spatial mask translates into a wide fourier convolution kernel, explaining why the MWA catalog power spectrum is so flat.

4.3. Simulations of distance-induced flux correlation

Let us now turn to addressing why the $4.5 \mu\text{Jy}$ SF sample is 5–15% correlated with the 185 MHz catalog, while the AGN sample is not. Of course some slight correlation is expected at some level as brighter AGN typically reside in more massive galaxies, which are typically brighter in stars as well (see, for instance, Fig. 1 in (Seymour et al. 2007) or Fig. 4 in (Willott et al. 2003)) however Mauch & Sadler (2007) find no evidence of a radio/near-infrared luminosity correlation anywhere near as strong as that between radio and far-infrared emission. However, as discussed above, broadband intensity mapping correlation experiments are affected not only by luminosity correlations between different bands, but flux correlations as well. We thus attempt to quantify to what extent fluxes in two different bands may appear correlated due to distance effects even when their intrinsic luminosities are completely independent of each other. By distance effects we refer to the effect that more distant objects are generally weaker in all bands than nearer objects.

We first make a few approximations to get intuition for the effect, and then simulate the effect as a function of stricter and stricter flux cuts due to deeper and deeper foreground masking. Consider a sky survey over a fixed field of view, and 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 $z \sim 0.5$, 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. 12, we have

$$c = \frac{\langle F_{\text{rad}} F_{\text{IR}} \rangle - \langle F_{\text{rad}} \rangle \langle F_{\text{IR}} \rangle}{\sqrt{(\langle F_{\text{rad}}^2 \rangle - \langle F_{\text{rad}} \rangle^2)(\langle F_{\text{IR}}^2 \rangle - \langle F_{\text{IR}} \rangle^2)}} \quad (13)$$

Then writing this in terms of the radio and IR luminosities $F_i = L_i / 4\pi d^2$ for $i = \text{rad}, \text{IR}$, we find

$$c = \frac{\beta - 1}{\sqrt{(\beta \alpha_{\text{rad}} - 1)(\beta \alpha_{\text{IR}} - 1)}} \approx (\alpha_{\text{rad}} \alpha_{\text{IR}})^{-1/2} \quad (14)$$

where $\beta \equiv \langle d^{-4} \rangle / \langle d^{-2} \rangle^2$ and $\alpha_i = \langle L_i^2 \rangle / \langle L_i \rangle^2$. For a survey of a fixed angular field of view, uniform spatial

distribution of objects, and cartesian spacetime, the distribution of source distances d grows as d^2 , which gives

$$\beta \approx \frac{d_{\text{max}}^3 - d_{\text{min}}^3}{3d_{\text{max}} d_{\text{min}} (d_{\text{max}} - d_{\text{min}})} \quad (15)$$

We observe that the radio-infrared flux correlation 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 GHz from Mauch & Sadler (2007) (Fig. 7, right panel) and at $8 \mu\text{m}$ from Fu et al. (2010) (left panel). The former describe galaxies at $z < 0.3$, while the latter describe galaxies at $z \sim 0.6$. This analysis could be extended using proper redshift-dependent luminosity functions, possibly from simulations, though we find that despite these simplifications, our analysis suffices to explain our earlier measurements and we leave a more detailed study for future work. We use approximately the same range of luminosities used by Mauch & Sadler (2007) and Fu et al. (2010), adjust the minimum luminosities slightly to achieve the same number density of AGN ($\sim 0.0020/\text{Mpc}^3$) and SF ($\sim 0.00011/\text{Mpc}^3$) in both radio and infrared surveys. In the end we find that our results are only slightly sensitive to these luminosity minima as their faint ends become less and less significant in real, flux-limited surveys.

We pick fiducial survey parameters of $d_{\text{min}} = 20 \text{ Mpc}$ and $z_{\text{max}} = 0.75 \text{ Mpc}$, giving $\beta \approx 54$. Using the above luminosity functions, we find $\alpha_{\text{SF,IR}} = 1.474$, $\alpha_{\text{SF,rad}} = 14.56$, $\alpha_{\text{AGN,IR}} = 22.97$, $\alpha_{\text{AGN,rad}} = 257.5$. 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. 7). 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 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 above 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 pushed the observed correlation with infrared images to zero at high- ℓ , suppressing the overall correlation computed in image space from our simulation prediction.

To what extent 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

Modifying
for
185 MHz
luminosity
function

SEPARATE
SEGMENTATION
Motivate
it.

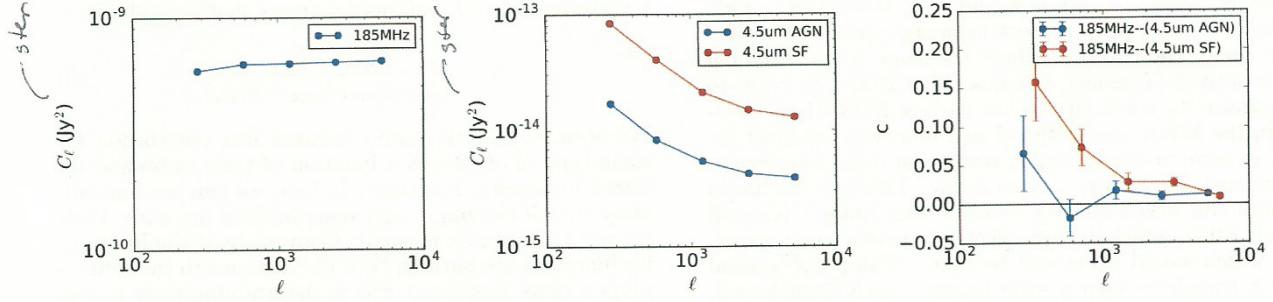


FIG. 6.— Power spectrum of 185 MHz sources (left panel), 4.5 μm sources (center panel), and coherence between 185 MHz and 4.5 μm sources (right panel). We roughly separate the 4.5 μm sources into starforming galaxies (SF) and active galactic nuclei (AGN) as illustrated in Fig. 5. We find that the 185 MHz–4.5 μm correlation observed in Fig. 4 holds only for the starforming galaxies in the infrared sample.

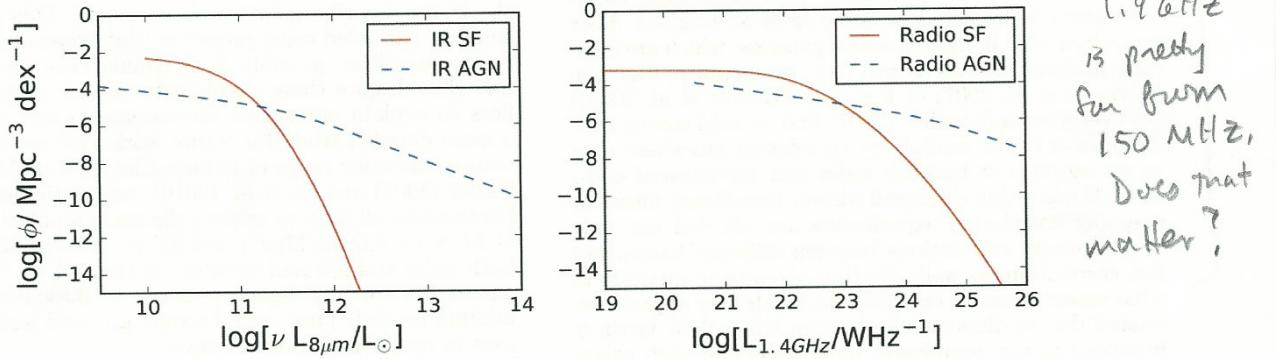


FIG. 7.— AGN and SF luminosity functions at 8 μm from Fu et al. (2010) (left panel) and at 1.4 GHz from Mauch & Sadler (2007) (right panel).

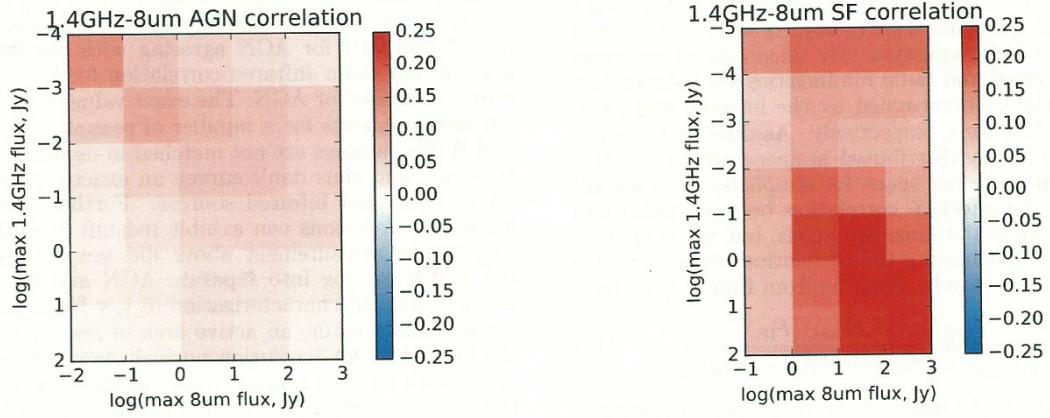


FIG. 8.— Apparent correlation of radio and mid-infrared fluxes from a mock survey with uncorrelated radio and mid-infrared luminosities using the luminosity functions from Fig. 7 for AGN and SF. Without any flux cuts (lower right corner of each plot) we observe a significant correlation between radio 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. We do not observe a monotonic trend, nor a general tendency towards zero as weaker and weaker sources are excluded.

son random number of each in each of 400 logarithmic luminosity bins. We distribute the objects uniformly over a volume $D_{\max} = cz_{\max}/H_0 = 3212$ Mpc deep and $\theta_{\text{FOV}} D_{\max} = 1121$ Mpc wide, then pick a random infrared luminosity for each radio object from the appropriate infrared luminosity function. Finally we plot in Fig. 8, along the lines of Fig. 4, the predicted 1.4 GHz– $8\mu\text{m}$ 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 change monotonically and they do not generally approach zero, 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 subtraction and masking of the the bright sources identified by the surveys discussed in the previous section.

5.1. Residual 21 cm foregrounds

In this section we characterize the 185 MHz foreground residuals in angular power spectrum measurements. We also quantify 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 uncorrelated radio and infrared foregrounds in cross spectrum analyses. In fact, while 21 cm intensity mapping measurements² in 3D are limited by radio thermal noise and demand ³order thousand hour integrations to reveal the cosmological signal (Beardsley et al. 2013; Pober et al. 2014), noise in broadband measurements is subdominant to foreground residuals after much shorter integration times.³ In fact, the *uv* plane is essentially full after only 3 hours of rotation synthesis, and compounding days together serves only to increase signal to noise.

The FHD outputs presented in Sec. 3.1 are naturally weighted image space cubes of the raw data (I_{nat}), the model data ($I_{\text{nat,mod}}$), the synthesized beam (I_w) (i.e., the fourier transform of the uv weights), and the primary beam. Each of these cubes has an *odd* and *even* version divided up at a 2 sec cadence. We average all these cubes over frequency to make broad band images, then apply uniform weighting as in Dillon et al. (2015) as

$$I_{\text{uni}}(\vec{\theta}) = \frac{10^{-26}}{2k_B\lambda^2} \sum_{\vec{u}} \frac{\tilde{I}_{\text{nat}}(\vec{u})}{\tilde{I}_w(\vec{u})} e^{-2\pi i \vec{\theta} \cdot \vec{u}} d^2 u \quad (16)$$

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 $i = \text{nat}, w$.

We then compute the angular power spectrum as

$$C(\ell) = \frac{\sum_{\vec{\ell}} |\tilde{I}_{\text{uni}}(\vec{\ell}) \tilde{I}_w(\vec{\ell})|^2}{\sum_{\vec{\ell}} |\tilde{I}_w(\vec{\ell})|^2} \quad (17)$$

summing over all the ℓ values in each ℓ bin. Note that $\ell = 2\pi u$. We estimate the thermal noise power spectrum by computing the power spectrum as above but using the odd minus even difference cube which contains only thermal noise.

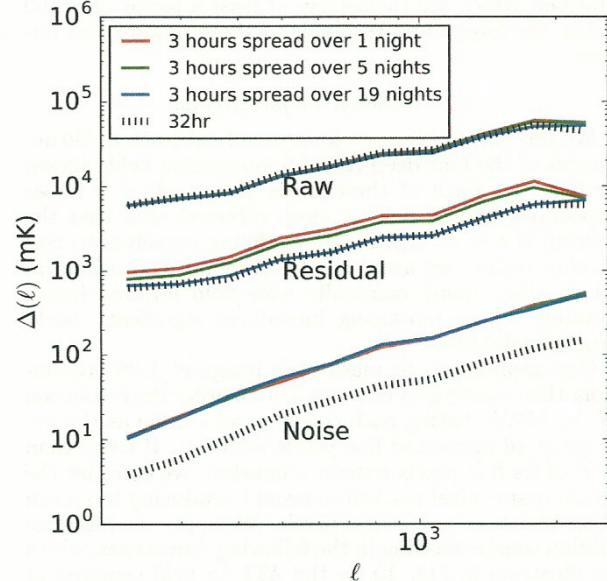


FIG. 9.— Raw, residual (post foreground subtraction), and noise power spectra of 185 MHz broad band images. We compare the 32 hour deep integration of Beardsley et al. (2016) to 3 hour integrations spaced over 1, 5, and 19 nights. The uv plane is nearly filled after only 3 hours, yet we find that spacing these ~ 100 two-minute integrations over more and more independent nights reduces the foreground residuals by a factor of up to ~ 4 in power. This bodes well for 21 cm-infrared broad band correlation experiments as we will find later that a very wide survey is essential to average down the sample variance noise in uncorrelated radio and infrared foregrounds.

We plot in Fig. 9 the power spectra of the raw, residual (post foreground subtraction), and noise images up to $\ell = 2600$, corresponding to a maximum baseline length of ~ 700 m. Beyond this, the uv coverage becomes quite sparse, introducing artifacts in the application of gridding and uniform weighting, though a more sophisticated analysis could likely use these long baselines. We compare the power spectra of the deep 32 hour integration of Beardsley et al. (2016) (black dashed) to those of 3 hour integrations spread over 1 (red), 5 (green), and 19 (blue) nights. Each of three datasets consists of ~ 100 two-minute integrations with a minimum spacing of 0, 5, and 24 minutes, all with occasional few-day gaps due to observing constraints and quality cuts.

We find that all four sets have nearly identical raw power spectra. However, 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 that of the