HERA Memo #72: Explaining and Mitigating the Temporal Structure of Calibration Solutions

Max Lee[1] and Joshua S. Dillon[1]

[1]Department of Astronomy, University of California, Berkeley

September 17, 2019

1 Introduction

In an interferometer, each visibility is measured as the product of the true visibility on the sky with the complex gains of the two antennas involved. Namely,

$$V_{ij}^{obs}(\nu) = g_i(\nu)g_j(\nu)^* V_{ij}^{true}(\nu) + n_{ij}(\nu), \tag{1}$$

where V_{ij}^{obs} is the observed visibility, g_i is the gain of antenna i, n_{ij} is the noise on the measured visibility, and V_{ij}^{true} is the true visibility of interest. The process of calibrating a radio interferometer requires thorough knowledge of the antenna gains so that the true visibility can be inferred. In 21 cm cosmology, precise determination of these gains is essential to the isolation of 21 cm signal from smooth but 10^{4-5} times brighter astrophysical foregrounds. HERA is designed to be calibrated using redundant-baseline calibration (Dillon and Parsons, 2016), a technique that leverages multiple simultaneous measurements of the same V_{ij}^{true} by different pairs of antennas with the same physical separation. This allows solutions for both the gains and true visibilities simultaneously as a large over-determined system of equations (Liu et al., 2010). However, this technique cannot quite solve for all the degrees of freedom in the problem; there are a handful of unsolvable degeneracies per frequency that require reference to a sky-model or an externally-calibrated set of visibilities to resolve (Dillon et al., 2018). This process of absolute calibration (abscal) combined with redundant-calibration (redcal) from the backbone of calibration technique in HERA's first two Internal Data Releases (Dillon, 2019, 2018).

2 Temporal Structure of Calibration Solutions

We have known for some time that HERA's calibration solutions exhibit temporal structure inconsistent with thermal noise. While gain amplitudes and phases can certainly vary with time and environmental factors (e.g. ambient temperature), we can rule out the explanation that most of this variation is intrinsic or "real" by observing that the temporal structure in the gains repeats from day to day as a function of local sidereal time (LST), as we show in Figure 1 using calibration results from IDR 2.2.

Since the local time at a fixed LST varies by more than a hour over the 18 days of IDR 2.2, the effect must be attributable to variations in the sky as perceived by the instrument. If the instrument were truly redundant and the differences between antennas explicable by a single complex gain, the brightness of foregrounds and their distribution on the sky should not affect the calibration solution—explaining it requires invoking instrumental nonidealities.

This observation raises two key questions:

1. Can we use existing techniques for smoothing calibration solutions to reduce the temporal structure we see in the data, by filtering out timescales where we have no evidence for "real" temporal variation? If so, what timescale is appropriate for that?
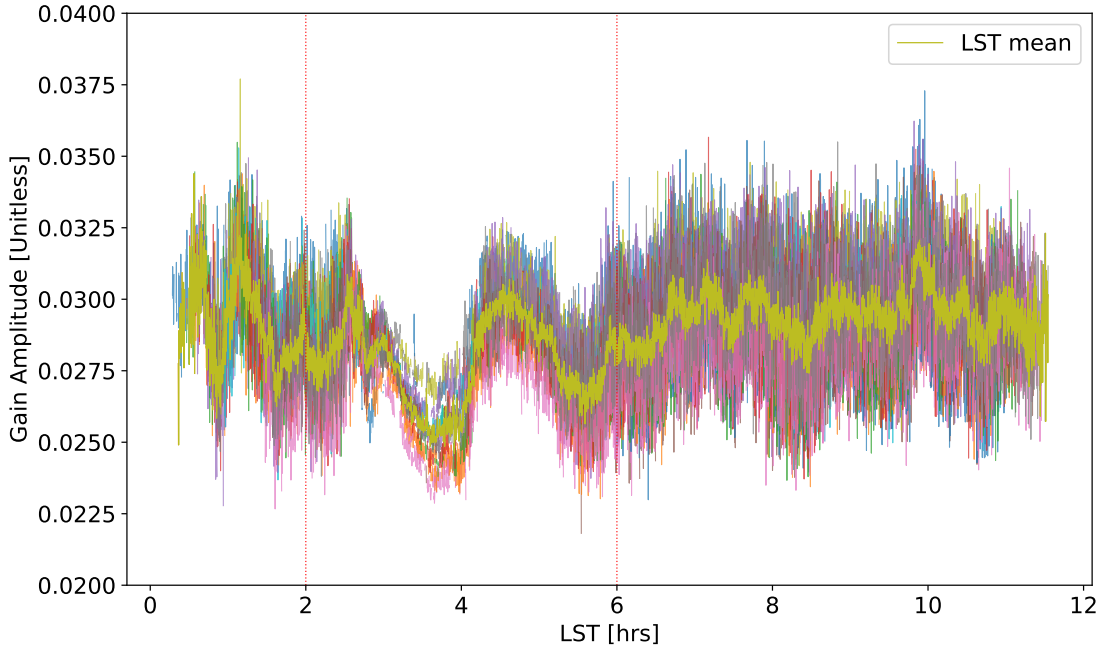
Figure 1: Here we show calibration gain amplitudes as a function of LST for a single antenna and polarization (in this case, Antenna 88x) and at a single frequency (∼125 MHz). Calibration is performed independently on every day, frequency, and 10.7 s integration. Each of the 18 background lines in the graph represents a day of gains from IDR 2.2 after redundant calibration, absolute calibration, and flagging. The magnitude of the fluctuations—at the ∼20% level—is typical of antennas and frequencies. Some of that fluctuation is expected from thermal noise, but the coherent fluctuations that repeat from day to day (especially around 4 hours) at the same LST indicates a systematic error. The foreground yellow line is an average calibration solution over all after removing two particularly flagged days (JD 2458104, 2458105, 2458109) and all times when the sun was above the horizon. This average can be used to estimate how much of the of the temporal variation is due to systematics—especially non-redundancy.

2. Can we explain the observed temporal structure in calibration solutions using a realistic sky model and plausible levels of non-redundancy?

In Section 3 we address the former question; in Section 4 we address the latter.

3   Estimating the Intrinsic Timescale of Calibration Variation

To understand how much of the temporal structure in our calibration solutions appears to be intrinsic or "real" and how much is attributable to effects that repeat night to night, we first want to calculate an average calibration solution over nights as a function of LST. For each day, we shift our calibration solutions to nearest grid points in LST and then average, taking into account different RFI flags on different days. JD 2458104, 2458105 and 2458109 were removed since they are substantially flagged. Likewise, times when the sun was above the horizon were also removed. The resultant average is shown in yellow in Figure 1.

In order to statistically assess the temporal structure of the calibration solutions by examining their temporal power spectra, we need to address the gaps in our data created by RFI flags. This is especially true of the night-by-night data, though we also see some flagged LSTs in our averaged data. We performed a low-pass filter on 60 s timescales to interpolate over RFI gaps using an iterative CLEAN algorithm. The result of this filter is demonstrated in Figure 2.

With the RFI gaps filled, we now calculate temporal power spectra of the gains. Our power spectra are
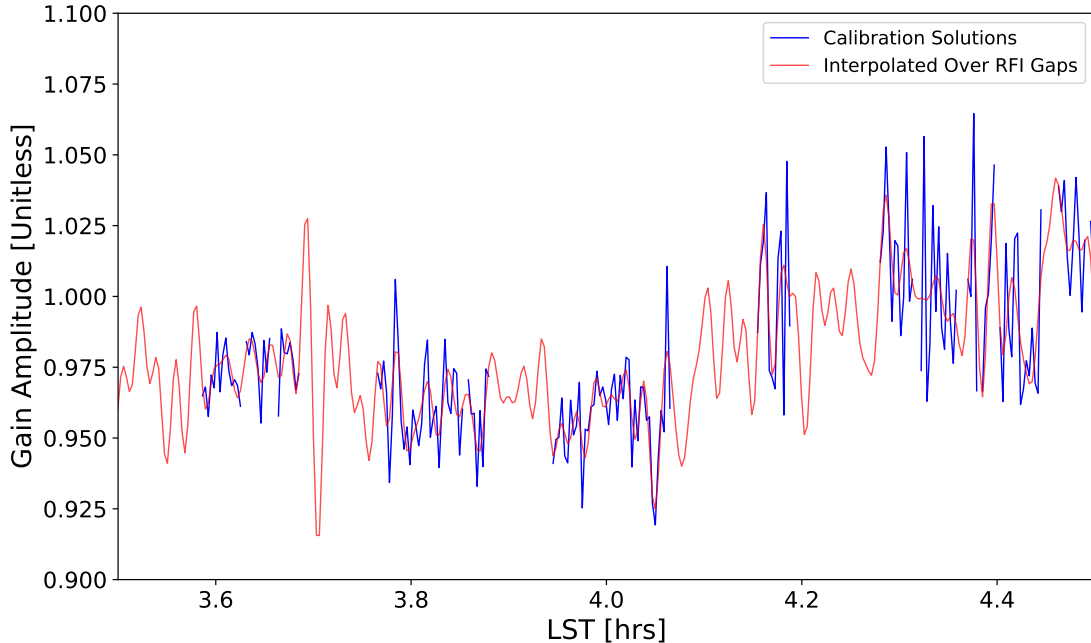
Figure 2: In order to eliminate the discontinuities in calibration solutions, we performed a low-pass Fourier filter, keeping temporal structure on timescales longer than 60 s. This allows us to then calculate temporal power spectra. Here we see the typical effect of the filter (*red line*) interpolating between the discontinuities in the data (*blue line*) on a single antenna and frequency (88x, $\nu \sim 125\,Hz$) for 2 hours LST on a given night (JD 2458098).

defined as

$$P_i(f) = |\text{FFT}\,(H(t)g_i(t))\,|^2, \qquad (2)$$

where f is fringe-rate, the Fourier dual to time of observation and $H(t)$ is a Hann window used to suppress sidelobes. We investigate the temporal power spectra of both our gain solutions and our gain solutions divided by the day-averaged solution (Figure 1), normalizing both to peak at 1 at $f = 0\,\text{Hz}$. This assumes that our gain solutions are separable as the product of a LST-dependent systematic and time-dependent intrinsic fluctuations. In Figure 3 we show the power spectra produced by separating the gains in this way. After dividing out our LST-mean calibration solution for every antenna from every day, we produce power spectra for each day and antenna. These power spectra are then averaged incoherently to form our best estimate of the intrinsic timescale of calibration fluctuations (the blue line in lower right panel). We also show the temporal power spectra of uncorrected gains averaged in the same way, which closely matches the average power spectra over all antennas of the LST-means for each antenna[1].

The main takeaway from this analysis is that most, if not all, of the temporal structure on timescales shorter than a few hours is attributable to the LST-locked gain variations that repeat from day to day. This means that, when smoothing our calibration solutions, we can comfortably filter out short timescales from our calibration solutions without losing real temporal structure. This analysis justifies the use of a 6-hour filtering timescale in IDR 2.2.

---

[1]These deviate at high fringe rate because the latter involves a coherent average over days and an incoherent average over antennas, rather than an incoherent average over both axes. As a result, this power spectrum hits a lower noise floor in the lower right panel.
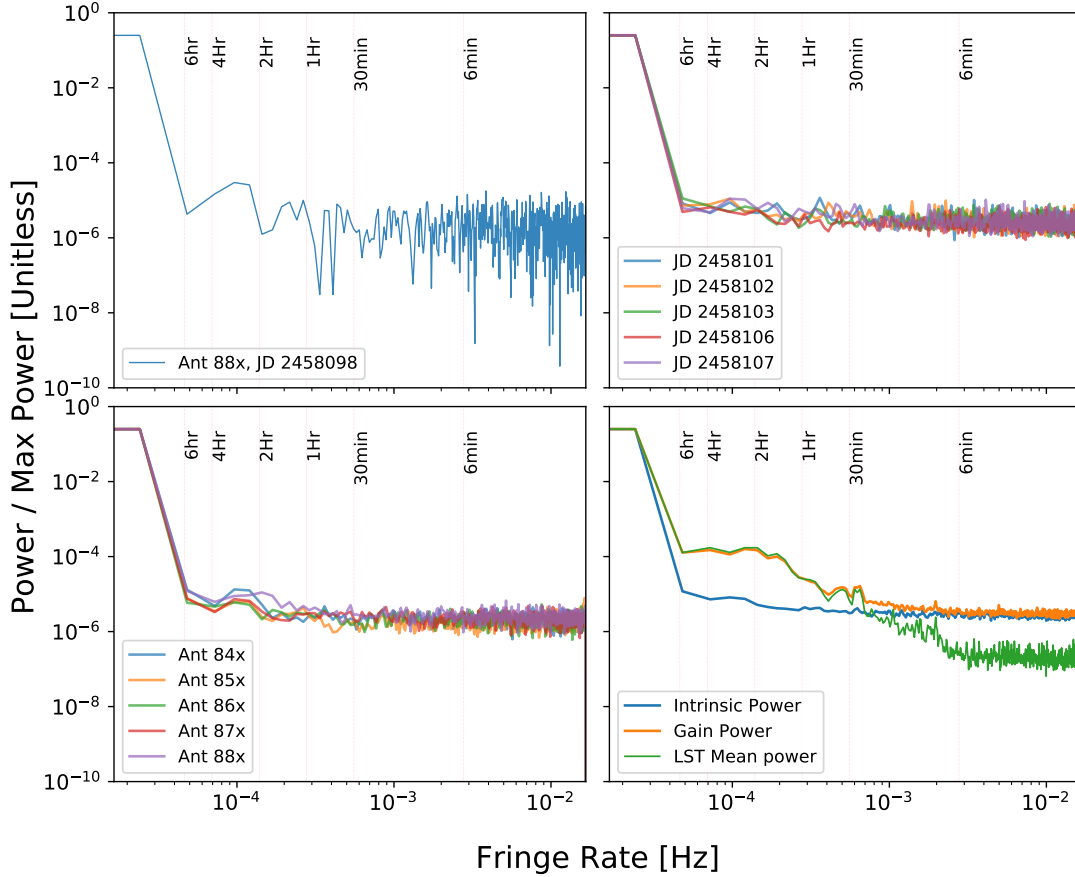
Figure 3: Here we show temporal power spectra of our calibration solutions. *Top left:* the renormalized power spectrum of a single antenna, 88x, on a single day, 2458098. *Top right:* Incoherently averaged power spectra over antennas for a subset of the days, with each line representing a different day. *Bottom left:* Incoherently averaged power spectra over days for a subset of the antennas, with each line representing a different antenna. *Bottom right:* Here we average incoherently over days and antennas, both before (orange line) and after (blue line) dividing out a per-antenna gain averaged over days (whose average power spectrum is shown in green). Our estimate for the intrinsic temporal power spectrum of the gain solutions quickly hits a noise floor and we see little evidence for real temporal structure on timescales shorter than ~6 hours.

4 Reproducing Sky-Locked Temporal Structure in Gain Solutions In Simulations with Non-Redundancy

We now turn to our second key question: what causes the observed LST-locked temporal structure in our calibration solutions? Our main suspect is non-redundancy, especially antenna-to-antenna variation of the primary beam. To see why, consider the Equation 1 for the observed visibility, but expanded to include the integral over the sky and the beam:

$$V_{ij}^{obs}(\nu) = g_i(\nu)g_j^*(\nu) \int d\Omega \, I(\hat{s}, \nu) B_{ij}(\hat{s}, \nu) \exp\left[-2\pi i \frac{\nu}{c}(\vec{b}_{ij} \cdot \hat{s})\right]. \tag{3}$$

Here I is the sky brightness, $B_{ij} = \sqrt{B_i B_j}$, the product of the voltage beams of antennas i and j, and $\vec{b}_{ij}$ is the baseline vector connecting them. To simulate a visibility, we approximate the sky as a set of discrete

sources,

$$I(\hat{s}, \nu) \approx \sum_n I_n(\nu) \delta(\hat{s} - \hat{s}_n), \tag{4}$$

yielding a sum over point sources at positions $\hat{s}_n$ with spectra $I_n(\nu)$:

$$V_{ij}(\vec{b}_{ij}, \nu) \approx g_i(\nu) g_j^*(\nu) \sum_n I_n(\nu) B_{ij}(\hat{s}_n, \nu) \exp\left[-2\pi i \frac{\nu}{c}(\vec{b}_{ij} \cdot \hat{s}_n)\right]. \tag{5}$$

Now consider the case where a single point source with flux I_0 dominates the observed sky. The observed visibility would be simply would simply be

$$V_{ij}(\vec{b}_{ij}, \nu) \approx \left[g_i(\nu)\sqrt{B_i(\hat{s}_0, \nu)}\right] \left[g_j^*(\nu)\sqrt{B_j(\hat{s}_0, \nu)}\right] I_0(\nu) \exp\left[-2\pi i \frac{\nu}{c}(\vec{b}_{ij} \cdot \hat{s}_0)\right]. \tag{6}$$

Here we have rearranged the terms to suggestively highlight the confusion that arises when beams vary from antenna to antenna. Namely, if we assume that all antennas have the same beam—an assumption key to redundant calibration and generally difficult to relax even in sky-based calibration—then we will misinterpret variations in the primary beam in a particular direction as a bright source moves overhead as variations in gain. This is consistent with our observation that the most dramatic temporal gain variations occur when Fornax A (RA= 3^h22^m) is in the main lobe of the primary beam or a close sidelobe. Of course, the effect can still occur with multiple bright sources, but it is harder to understand analytically.

To test this qualitative picture of temporal structure, we simulate a sky at 125 MHz over the same range of LSTs. We use the Global Sky Model (De Oliveira-Costa et al., 2008) with HEALPix $NSIDE = 128$ along with 1000 of the brightest sources closest to HERA's zenith from the GLEAM catalog (Hurley-Walker et al., 2017). We also add the brightest known radio sources listed in Table 2 of Hurley-Walker et al. (2017). While it does not appear in that paper, we also include Fornax A modeled as a single point source with 750 Jy at 150 MHz with spectra index $\alpha = -.81$ (McKinley et al., 2014).

Our noise-free simulated visibilities include all baselines in IDR 2.2. Following Orosz et al. (2019), we simulate non-redundancy by modeling each antenna with a perturbed position and with an Airy beam parametrized as

$$B_i(\vec{r}, \nu) = \left[\frac{2J_1\left(k\sqrt{a_x^2(x - x_{pc})^2 + a_y^2(y - y_{pc})^2}\right)}{k\sqrt{a_x^2(x - x_{pc})^2 + a_y^2(y - y_{pc})^2}}\right]^2, \tag{7}$$

where J_1 is a Bessel Function of the first kind, k is the wave number given by, $k = 2\pi\nu/c$, and x_{pc} and y_{pc} are the coordinates of the pointing center and a_x and a_y parameterize the shape of the beam. x_{pc} and y_{pc} both have mean 0 (i.e. pointing center at zenith) and a_x and a_y both have a mean of 6 m. However, each antenna has errors in these parameters, as well as their positions on the ground, drawn from a Gaussian distribution with standard deviations in Table 1.

Position Error: σ_{pos} [m]	Beam Size Error: σ_a [m]	Beam Pointing Error σ_{pc} [unitless]
0.03	0.0909	0.005135

Table 1: Antenna position, beam size, and beam pointing error levels. The latter two correspond to a .15° standard deviation the FWHM and a .28° standard deviation in the pointing. See Orosz et al. (2019) for more details.

Our simulation is generated with perfect calibration, but performing redundant-baseline calibration on a non-redundant array introduces temporal structure into the gain solutions. We then compute the temporal power spectrum of the gain solutions for each antenna in the simulation and average them together incoherently. The result, shown in Figure 4, reproduces the observed temporal power spectra extremely well at all timescales up to the point where the simulation dips below the noise floor in the data.

We can conclude that the observed LST-locked temporal structure in our gain solutions is largely, if not entirely, attributable to non-redundancy in the array and can therefore be mitigated. This mitigation can be done by making the array more redundant or finding ways to weight the visibilities in redundant calibration to down weight sidelobes where the most antenna-to-antenna variation is likely to occur.
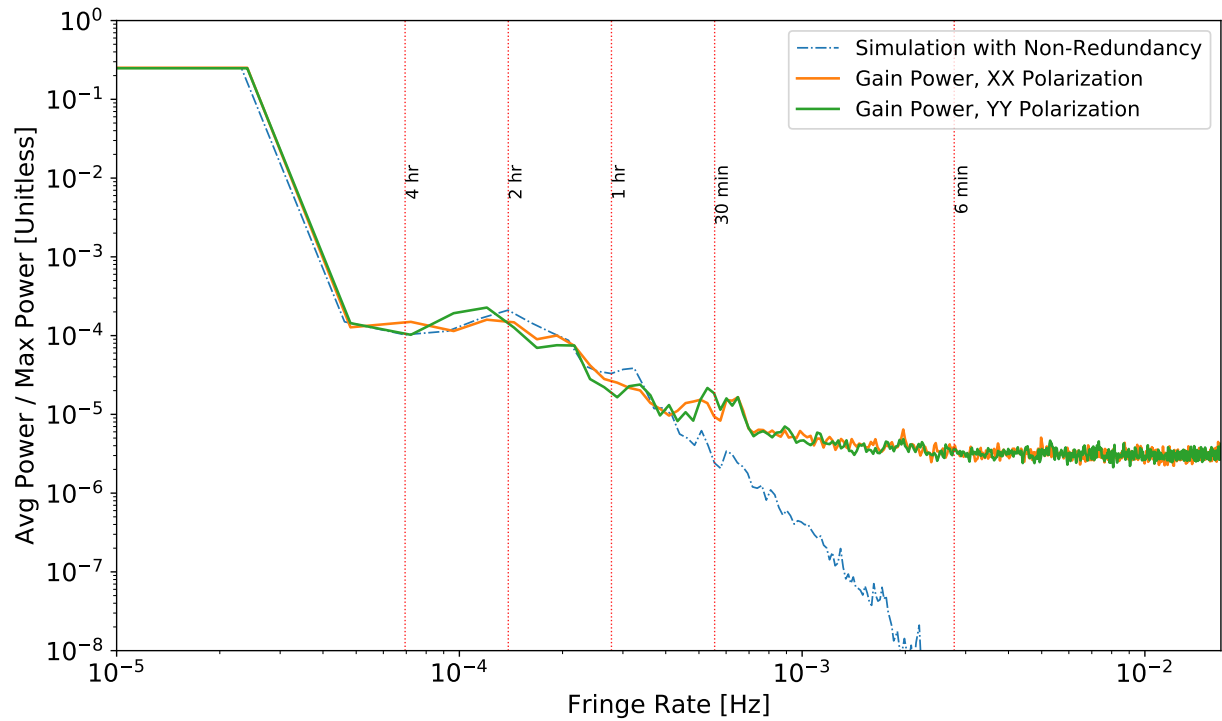
Figure 4: Here we compare the simulated power spectrum of the calibration gains with the actual power spectrum from IDR 2.2, using both instrumental polarizations (the orange line is the same as the one in the bottom right panel of Figure 3). The simulation faithfully reproduces the temporal statistics of the effect we saw using HERA data at basically all timescales until the data hit a noise floor in the data at a relative power level of $\sim 5 \times 10^{-6}$, lending credence to the hypothesis that the temporal structure we see is due to non-redundancy.)

References

De Oliveira-Costa, A., M. Tegmark, B. M. Gaensler, J. Jonas, T. L. Landecker, and P. Reich
2008. A model of diffuse galactic radio emission from 10 mhz to 100 ghz. *Monthly Notices of the Royal Astronomical Society*, 388(1):247–260.

Dillon, J. S.
2018. Hera memo 45: H1c idr2.1.

Dillon, J. S.
2019. Hera memo 69: H1c idr2.2.

Dillon, J. S., S. A. Kohn, A. R. Parsons, J. E. Aguirre, Z. S. Ali, G. Bernardi, N. S. Kern, W. Li, A. Liu, C. D. Nunhokee, and J. C. Pober
2018. Polarized redundant-baseline calibration for 21 cm cosmology without adding spectral structure. , 477:5670–5681.

Dillon, J. S. and A. R. Parsons
2016. Redundant Array Configurations for 21 cm Cosmology. , 826(2):181.

Hurley-Walker, N., J. R. Callingham, P. J. Hancock, T. M. O. Franzen, L. Hindson, A. D. Kapińska, J. Morgan, A. R. Offringa, R. B. Wayth, C. Wu, Q. Zheng, T. Murphy, M. E. Bell, K. S. Dwarakanath, B. For, B. M. Gaensler, M. Johnston-Hollitt, E. Lenc, P. Procopio, L. Staveley-Smith, R. Ekers, J. D. Bowman, F. Briggs, R. J. Cappallo, A. A. Deshpande, L. Greenhill, B. J. Hazelton, D. L. Kaplan, C. J. Lonsdale, S. R. McWhirter, D. A. Mitchell, M. F. Morales, E. Morgan, D. Oberoi, S. M. Ord, T. Prabu, N. U. Shankar, K. S. Srivani, R. Subrahmanyan, S. J. Tingay, R. L. Webster, A. Williams, and C. L. Williams
2017. GaLactic and Extragalactic All-sky Murchison Widefield Array (GLEAM) survey - I. A low-frequency extragalactic catalogue. , 464(1):1146–1167.

Liu, A., M. Tegmark, S. Morrison, A. Lutomirski, and M. Zaldarriaga
2010. Precision calibration of radio interferometers using redundant baselines. , 408(2):1029–1050.

McKinley, B., R. Yang, M. López-Caniego, F. Briggs, N. Hurley-Walker, R. B. Wayth, A. R. Offringa, R. Crocker, G. Bernardi, P. Procopio, B. M. Gaensler, S. J. Tingay, M. Johnston-Hollitt, M. McDonald, M. Bell, N. D. R. Bhat, J. D. Bowman, R. J. Cappallo, B. E. Corey, A. A. Deshpande, D. Emrich, A. Ewall-Wice, L. Feng, R. Goeke, L. J. Greenhill, B. J. Hazelton, J. N. Hewitt, L. Hindson, D. Jacobs, D. L. Kaplan, J. C. Kasper, E. Kratzenberg, N. Kudryavtseva, E. Lenc, C. J. Lonsdale, M. J. Lynch, S. R. McWhirter, D. A. Mitchell, M. F. Morales, E. Morgan, D. Oberoi, S. M. Ord, B. Pindor, T. Prabu, J. Riding, A. E. E. Rogers, D. A. Roshi, N. Udaya Shankar, K. S. Srivani, R. Subrahmanyan, M. Waterson, R. L. Webster, A. R. Whitney, A. Williams, and C. L. Williams
2014. Modelling of the spectral energy distribution of Fornax A: leptonic and hadronic production of high-energy emission from the radio lobes. *Monthly Notices of the Royal Astronomical Society*, 446(4):3478–3491.

Orosz, N., J. S. Dillon, A. Ewall-Wice, A. R. Parsons, and N. Thyagarajan
2019. Mitigating the effects of antenna-to-antenna variation on redundant-baseline calibration for 21 cm cosmology. , 487:537–549.