



Mitigating Internal Instrument Coupling for 21 cm Cosmology. II. A Method Demonstration with the Hydrogen Epoch of Reionization Array

Nicholas S. Kern¹ , Aaron R. Parsons¹, Joshua S. Dillon^{1,22} , Adam E. Lanman², Adrian Liu³ , Philip Bull⁴ , Aaron Ewall-Wice⁵ , Zara Abdurashidova¹, James E. Aguirre⁶ , Paul Alexander⁷, Zaki S. Ali¹, Yanga Balfour⁸, Adam P. Beardsley^{9,22} , Gianni Bernardi^{8,10,11} , Judd D. Bowman⁹ , Richard F. Bradley¹², Jacob Burba², Chris L. Carilli¹³ , Carina Cheng¹, David R. DeBoer¹, Matt Dexter¹, Eloy de Lera Acedo⁷, Nicolas Fagnoni⁷, Randall Fritz⁸, Steve R. Furlanetto^{17,18} , Brian Glendenning¹³, Deepthi Gorthi¹, Bradley Greig^{15,16}, Jasper Grobbelaar⁸, Ziyaad Haldaj⁸, Bryna J. Hazelton^{17,18} , Jacqueline N. Hewitt¹⁹, Jack Hickish¹, Daniel C. Jacobs⁹ , Austin Julius⁸, Joshua Kerrigan² , Piyanat Kittiwisit⁹ , Saul A. Kohn⁶ , Matthew Kolopanis⁹ , Paul La Plante⁶ , Telalo Lekalake⁸, David MacMahon¹, Lourence Malan⁸, Cresshim Malgas⁸, Matthys Maree⁸, Zachary E. Martinot⁶, Eunice Matsetela⁸, Andrei Mesinger²⁰, Mathakane Molewa⁸, Miguel F. Morales¹⁷ , Tshegofalang Mosiane⁸, Steven G. Murray⁹ , Abraham R. Neben¹⁹ , Aaron R. Parsons¹, Nipanjana Patra¹ , Samantha Pieterse⁸, Jonathan C. Pober² , Nima Razavi-Ghods⁷, Jon Ringuette¹⁷, James Robnett¹³, Kathryn Rosie⁸, Peter Sims², Craig Smith⁸, Angelo Syce⁸, Nithyanandan Thyagarajan^{9,13,23} , Peter K. G. Williams²¹ , and Haoxuan Zheng¹⁹

¹ Department of Astronomy, University of California, Berkeley, CA, USA; nkern@berkeley.edu

² Department of Physics, Brown University, Providence, RI, USA

³ Department of Physics and McGill Space Institute, McGill University, Montreal, Canada

⁴ School of Physics & Astronomy, Queen Mary University of London, London, UK

⁵ Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

⁶ Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA, USA

⁷ Cavendish Astrophysics, University of Cambridge, Cambridge, UK

⁸ SKA-SA, Cape Town, South Africa

⁹ School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA

¹⁰ Department of Physics and Electronics, Rhodes University, P.O. Box 94, Grahamstown, 6140, South Africa

¹¹ INAF-Istituto di Radioastronomia, via Gobetti 101, I-40129 Bologna, Italy

¹² National Radio Astronomy Observatory, Charlottesville, VA, USA

¹³ National Radio Astronomy Observatory, Socorro, NM, USA

¹⁴ Department of Physics and Astronomy, University of California, Los Angeles, CA, USA

¹⁵ School of Physics, University of Melbourne, Parkville, VIC 3010, Australia

¹⁶ ARC Centre of Excellence for All-Sky Astrophysics in 3 Dimensions (ASTRO 3D), University of Melbourne, VIC 3010, Australia

¹⁷ Department of Physics, University of Washington, Seattle, WA, USA

¹⁸ eScience Institute, University of Washington, Seattle, WA, USA

¹⁹ Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA

²⁰ Scuola Normale Superiore, I-56126 Pisa, PI, Italy

²¹ Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA

Received 2019 July 5; revised 2019 November 15; accepted 2019 December 1; published 2020 January 9

Abstract

We present a study of internal reflection and cross-coupling systematics in Phase I of the Hydrogen Epoch of Reionization Array (HERA). In a companion paper, we outlined the mathematical formalism for such systematics and presented algorithms for modeling and removing them from the data. In this work, we apply these techniques to data from HERA's first observing season as a method demonstration. The data show evidence for systematics that, without removal, would hinder a detection of the 21 cm power spectrum for the targeted Epoch of Reionization (EoR) line-of-sight modes in the range $0.2 h^{-1} \text{Mpc}^{-1} < k_{\parallel} < 0.5 h^{-1} \text{Mpc}^{-1}$. In particular, we find evidence for nonnegligible amounts of spectral structure in the raw autocorrelations that overlaps with the EoR window and is suggestive of complex instrumental effects. Through systematic modeling on a single night of data, we find we can recover these modes in the power spectrum down to the integrated noise floor, achieving a dynamic range in the EoR window of 10^6 in power (mK^2 units) with respect to the bright galactic foreground signal. Future work with deeper integrations will help determine whether these systematics can continue to be mitigated down to EoR levels. For future observing seasons, HERA will have upgraded analog and digital hardware to better control these systematics in the field.

Unified Astronomy Thesaurus concepts: [Reionization \(1383\)](#); [Cosmology \(343\)](#); [Astronomy data analysis \(1858\)](#)

1. Introduction

The Epoch of Reionization (EoR) marks a fundamental phase transition in cosmic history, where neutral hydrogen filling the intergalactic medium (IGM) was ionized by a radiation field

thought to originate from the formation of the first generation of stars and galaxies in the universe (for reviews, see Furlanetto et al. 2006; Loeb & Furlanetto 2013; Mesinger 2016). One of the only direct probes of the IGM throughout the entirety of the EoR is neutral hydrogen's 21 cm line. A hyperfine transition of neutral hydrogen, 21 cm emission is a three-dimensional, tomographic probe of the IGM's density, ionization, and temperature structure.

²² NSF Astronomy and Astrophysics Postdoctoral Fellow.

²³ Jansky Fellow.

Low-frequency radio surveys promise to revolutionize our understanding of the IGM by using the 21 cm line to map out its morphology during EoR and place constraints on the sources responsible for its heating and eventual reionization.

Over the past decade, experiments like the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER; Parsons et al. 2014; Jacobs et al. 2015; Cheng et al. 2018; Kolopanis et al. 2019), the Murchison Widefield Array (MWA; Dillon et al. 2014; Beardsley et al. 2016; Ewall-Wice et al. 2016b; Barry et al. 2019), the Low Frequency Array (LOFAR; Patil et al. 2017; Gehlot et al. 2019), the Giant Metre Wave Radio Telescope (GMRT; Paciga et al. 2013), and the Long Wavelength Array (LWA; Eastwood et al. 2019) have placed increasingly competitive limits on the 21 cm power spectrum. These experiments face the challenge of separating a weak cosmological signal from foreground emission that is generally 10^5 times brighter in order to characterize the EoR. Instrumental systematics further complicate this effort, which can cause foreground signal to contaminate Fourier modes in the data that would otherwise only be noise-limited. As such, many of the current upper limits on the 21 cm power spectrum have been limited by instrumental systematics. Current and future experiments like the Hydrogen Epoch of Reionization Array (HERA; DeBoer et al. 2017) and the Square Kilometre Array (SKA; Mellema et al. 2013) are nominally forecast to provide high-significance characterizations of the 21 cm signal and place constraints on IGM properties and the sources driving reionization (Pober et al. 2014; Greig & Mesinger 2015, 2018; Greig et al. 2015; Ewall-Wice et al. 2016a; Liu & Parsons 2016; Kern et al. 2017). However, these forecasts neglect the impact of systematic contamination, which can significantly hamper an experiment’s overall sensitivity and parameter constraining ability. Precise modeling and separation of instrumental systematics is therefore necessary for 21 cm experiments to make robust detections of the cosmological 21 cm signal.

Systematic contamination comes in a variety of forms, including calibration errors, ionospheric faraday rotation, primary beam ellipticity, analog signal chain imperfections (such as impedance mismatches), and others. In a companion paper, Kern et al. (2019a), we presented techniques for modeling and removing systematics specifically due to internal instrument coupling, such as signal chain reflections and antenna cross-coupling (e.g., crosstalk). In that paper, we described the phenomenology of internal instrument systematics in the interferometric visibilities, proposed algorithms for removing them from the data, and demonstrated their performance against numerical simulations. In this work, we investigate data from HERA Phase I for internal instrument systematics and apply our systematic modeling algorithms as a proof of concept.

The structure of this paper is as follows. In Section 2 we describe the data and observations used for this analysis. In Section 3 we examine the data for signal chain reflections and demonstrate reflection calibration on HERA autocorrelation visibilities. In Section 4 we present a study of cross-coupling systematics in the HERA system and demonstrate cross-coupling removal performance on a few select baselines. In Section 5 we perform joint reflection and cross-coupling systematic removal for baselines across the entire HERA array and compute power spectra. Lastly, in Section 6 we summarize our results.

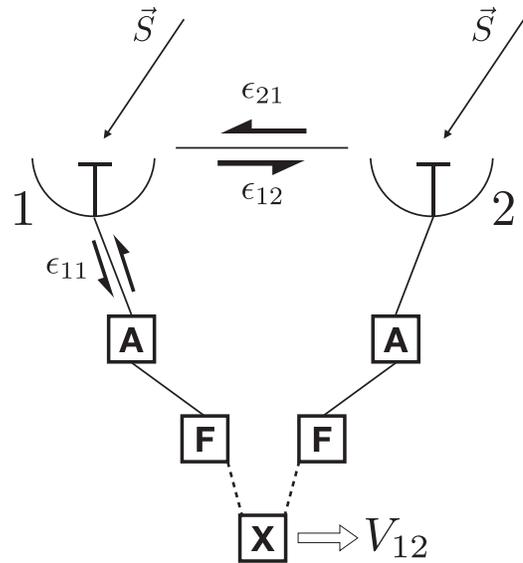


Figure 1. Schematic of HERA signal chains for two antennas, 1 and 2. Sky signal (\mathcal{S}) enters each antenna’s dish and feed, where it is converted into a voltage and travels down a 150 m coaxial cable to a processing node holding a post-amplification module (A) before being directed through a 20 m cable to an engine that digitizes and Fourier transforms the signal (F) and then sent to the correlator (X) to produce the visibility V_{12} . A possible cable reflection in antenna 1’s signal chain is marked as ϵ_{11} , traversing up and down the cable connecting the feed to the node. A possible source of feed-to-feed coupling is marked as ϵ_{12} and ϵ_{21} , where a signal is reflected off of antenna 1’s feed and into antenna 2’s feed, or vice versa. The dashed line from F to X denotes a signal pathway after digitization, where reflection is not a concern.

2. Observations

Data were taken during HERA Phase I, which observed from 2017 to 2018 while undergoing active construction (DeBoer et al. 2017). The Phase I instrument was a hybrid HERA-PAPER system, taking the signal chains and correlator from the PAPER experiment (Parsons et al. 2010; Ali et al. 2015) and attaching them to new HERA antennas. The HERA antenna is a parabolic dish spanning 14 m in diameter, with a focal height designed to minimize reflections within the dish (Ewall-Wice et al. 2016; Neben et al. 2016; Thyagarajan et al. 2016; Patra et al. 2018). The feed uses the PAPER sleeved dipole as the active element, which, in the Phase I instrument, has been optimized for the HERA antenna (DeBoer 2015; Fagnoni & De Lera Acedo 2016). An active balun or front-end module (FEM) is connected to the feed and houses a low-noise amplifier. After initial amplification, the signals are sent through a 150 m coaxial cable (first cable in Figure 1) to a node unit in the field holding a post-amplifier module (PAM; box A in Figure 1) and then another coaxial cable of about 20 m in length (second cable in Figure 1) to a container holding ROACH2 boards²⁴ (Parsons et al. 2008; Hickish et al. 2016) that digitize the signals and then Fourier transform them into the frequency domain (box F in Figure 1). Finally, a graphics processing unit (GPU) correlator cross-multiplies the signals between all antenna pairs to form interferometric visibilities that are integrated for 10.7 s before being written to disk (box X in Figure 1).

The observations presented in this work come from a single night spanning 8 hr of local sidereal time (LST) on Julian Date 2458101. At that time, the array consisted of 46 operational

²⁴ <https://casper.ssl.berkeley.edu/wiki/ROACH2>

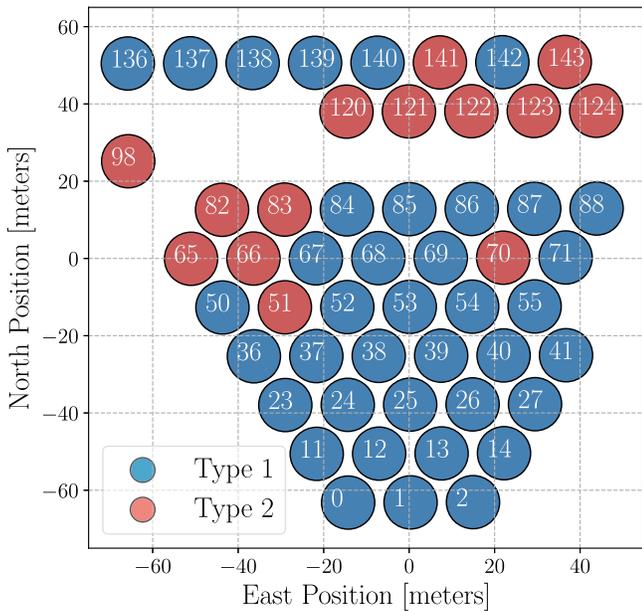


Figure 2. HERA array configuration at the time of observations on Julian Date 2458101 with roughly 46 operational antennas, showing which fall into type 1 (blue) and type 2 (red) signal chain categories.

antennas, each with dual-polarization dipole feeds (Figure 2). Additionally, the signal chains of the array were split into two categories: type 1, which used newly manufactured FEMs, PAMs, and coaxial cables specifically for HERA Phase I, and type 2, which repurposed the PAPER FEMs, PAMs, and coaxial cables (blue and red in Figure 2, respectively). In this analysis, we only use north–south (“YY”) linear dipole polarization data, although all four auto- and cross-feed polarization data products are recorded by the correlator. Additional observational parameters are tabulated in Table 1.

The data have been preprocessed with part of the HERA reduction and calibration pipeline. Specifically, the data are first flagged for radio frequency interference (RFI) using a median filter and watershed algorithm operating on the cross-correlation visibilities (Kerrigan et al. 2019). In this work, we also enact two additional steps for RFI flagging. The first takes stacked autocorrelation visibilities and differences them across time and frequency, normalizes them by their median absolute deviation, and flags the residual at the 4σ level. Our second step runs a delay-based, iterative deconvolution on a subset of the autocorrelation visibilities, which attempts to deconvolve the discontinuous windowing function created by flagged data. This is similar in concept to the image-based CLEAN deconvolution (Högbom 1974), except applied to the frequency and delay domains rather than the uv and lm domains and with the missing data coming from RFI rather than incomplete uv sampling. We then normalize the filtered residual in frequency space by its median absolute deviation and again enact RFI cuts at the 4σ level. Flags from each of the three independent steps are combined with a logical OR operator and then broadcast across time and/or frequency if a 15% flagged threshold is met for any individual time bin or frequency channel. An example of the fairly aggressive resultant visibility flagging mask is shown in Figure 3. In total, roughly 30% of the data volume is flagged, although this likely contains a decent amount of overflagging.

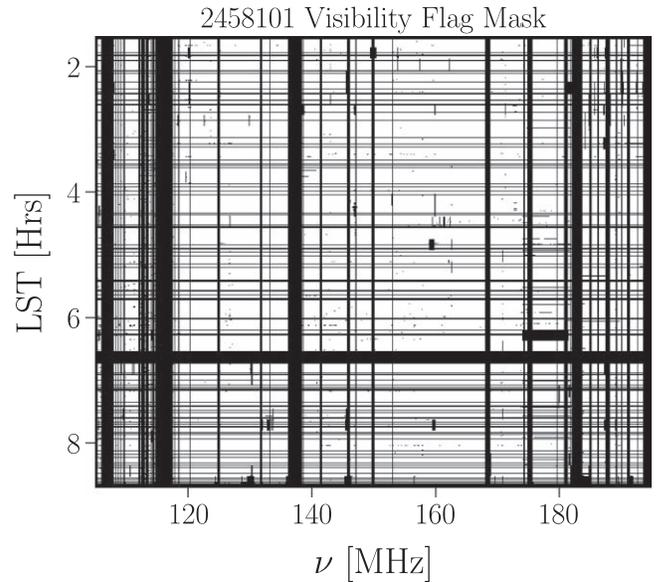


Figure 3. Aggressive RFI visibility mask as a function of time and frequency after three rounds of flagging. Flags are broadcast across time and/or frequency if a 15% flagged threshold is met per time bin and frequency channel. For this particular night, $\sim 30\%$ of the total data volume is flagged, which is suboptimal in that it is likely a significant overflagging but with the benefit of being more aggressive in flagging low-level and repeating RFI.

Table 1
HERA Observation Parameters

| Parameter | Value |
|---------------------------|---|
| Observation date | 13 December 2017 |
| Array coordinates | $-30^{\circ}7\text{S}, 21^{\circ}4\text{E}$ |
| JD range | 2458101.27–2458101.61 |
| LST range | 1.5–9.6 hr |
| Integration time | 10.7 s |
| Frequency range | 100–200 MHz |
| Channel width | 97.65 kHz |
| Dish diameter | 14 m |
| Feed type | PAPER dipole |
| Instrumental polarization | North–south (“YY”) |
| Cable type | 150 and 20 m coaxial |

Note. For the 2017–2018 observation, the HERA correlator used the convention that the X dipole points east–west while the Y dipole points north–south, which is not the standard Hamaker & Bregman (1996) definition.

Next, we calibrate the data using a highly simplified antenna-based calibration. The full HERA calibration pipeline computes complex antenna gains for each time integration over the entire night from a combination of redundant calibration (J. Dillon et al. 2019, in preparation) and a constrained absolute calibration with the resultant gains smoothed across time and frequency (Kern et al. 2019b). In this work, we take the gains derived from these steps and (1) average them across the entire night into a single spectrum, (2) average their amplitude across frequency to a single number, and (3) fit for a phase slope across frequency (i.e., a single antenna delay). We are left with a single amplitude and delay for each antenna, which we apply to all times of the night. This has the effect of properly setting the flux scale of the data and also calibrates out the antenna cable delay but ensures that the gain itself we apply to the data has little to no spectral structure.

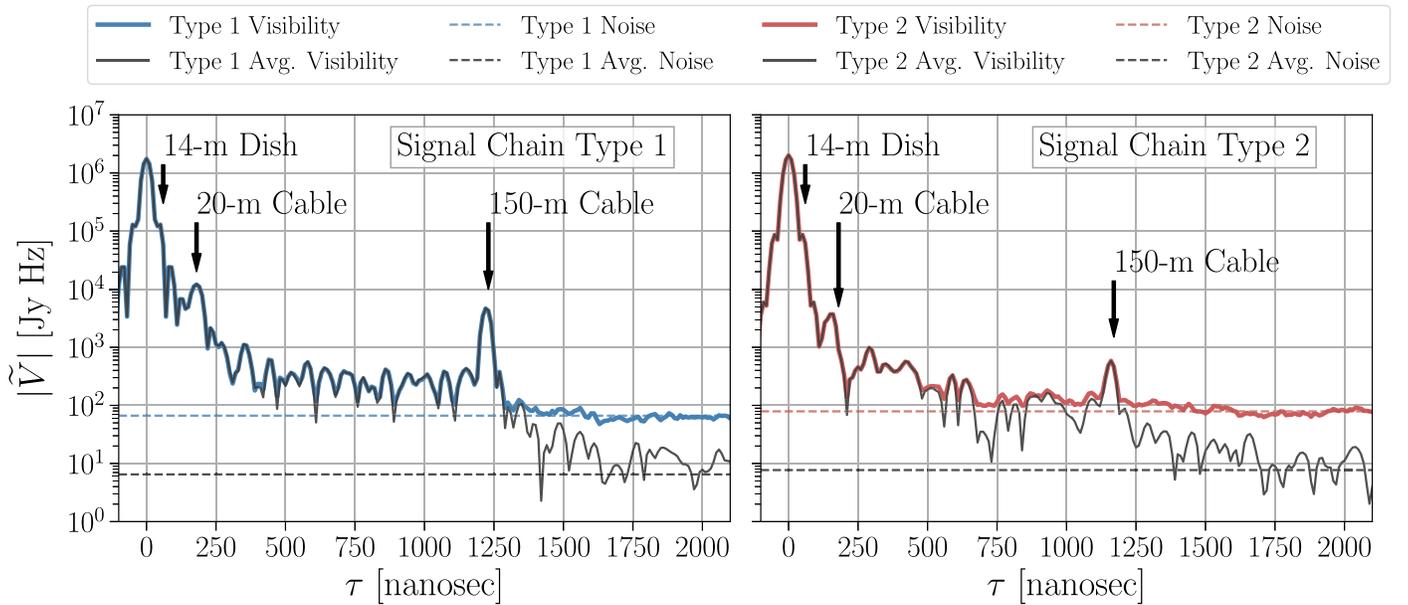


Figure 4. Autocorrelation visibilities for signal chain type 1 (left) and type 2 (right) with absolute (blue and red) and complex (black) time averaging and their associated noise floors (dashed). Antennas 84 and 121 were used for the two autocorrelation visibilities. Delays for relevant length scales in the analog system are marked with arrows. Resonances in the dish and reflections in the cables tend to be worse for signal chain type 1. Additionally, we see evidence for a systematic tail in both signal chain types spanning a wide range of delays that does not integrate down like noise.

Because of our highly simplified calibration, the instrumental bandpass is not corrected for and still exists in the data. Calibration, being multiplicative in frequency space, can be thought of as a convolution in delay space. The true response of the visibilities in delay space is therefore initially convolved by the bandpass kernel upon measurement by the telescope. Assuming the bandpass is composed primarily of large-scale modes, its impact will be a slight smoothing-out of the true sky delay response and features created by systematics. Bandpass calibration performed beforehand may therefore sharpen systematics in delay space and actually make it easier to model and remove them. Antenna-based calibration for HERA in the context of redundant and absolute calibration is explored in Kern et al. (2019b) and J. Dillon et al. (2019, in preparation).

3. Signal Chain Reflections

In this section, we inspect the data for evidence of signal chain reflections. To do this, we take the autocorrelation visibility from each antenna and look for peaks in delay space (see Kern et al. 2019a for a summary of the algorithm). The calibrated data are filled with flags due to RFI (Figure 3) and thus nulled to zero at the flagged channels. This is not ideal for inspecting the data in delay space, as the Fourier transform of such a discontinuous windowing function creates strong side lobes. To mitigate this, we employ the same delay-based, iterative deconvolution algorithm from before to subtract these side lobes, effectively interpolating across the nulled gaps in the data due to RFI (Parsons & Backer 2009). We allow the deconvolution to place model components out to delays of $|\tau| < 1600$ ns and iterate until the process reaches $5 \times$ the noise floor of the data. We then make a copy of the data, and with the first copy, we average the absolute value of the deconvolved visibilities in delay space across a few hours of LST. With the second copy, we average the full complex-valued, deconvolved visibilities across the same time range, which will have a lower noise floor due to the complex average. Finally, we apply a

Blackman window to the CLEANed visibilities before taking its Fourier transform.

Figure 4 shows these data products for the type 1 (left) and type 2 (right) signal chain, with the absolute time-averaged data shown in blue or red and the complex time-averaged data shown in solid black. Additionally, the thermal noise floors of each data product are plotted as dashed lines, which are estimated from the data via adjacent time and frequency differencing and then divided by $1/\sqrt{N_{\text{avg}}}$, where N_{avg} is the number of complex averages performed on the data. We find that type 2 signal chains achieve a better overall impedance match with the analog system, leading to slightly less structure in the autocorrelations across a wide range of delays. Nonetheless, we do see evidence for reflections from both the 20 and 150 m cables, with reflection amplitudes in the range of roughly 3×10^{-3} and 1×10^{-3} , respectively. Of major concern is the tail of the autocorrelation response, which starts at low delays and slopes down to the noise floor out to the 150 m cable delay. This tail is over an order of magnitude larger than that predicted by simulations of the HERA dish and feed (Ewall-Wice et al. 2016).

In this case, where the noise floor has been integrated down (solid black), we see that delays outside the 150 m cable delay seem to effectively integrate down with the noise, while delays inside the 150 m cable delay do not. This means that the features at low and intermediate delays are coherent on long timescales of at least a few hours. The abrupt change at ~ 1250 ns is also possibly suggestive that a tailed response might in part be originating within the 150 m cable. A possible mechanism for this could be subreflections within the cable due to intrinsic cable imperfections or environmental wear and damage along the cable. Another explanation is the effect of mutual coupling between neighboring antennas, which we explore in more detail in cross-correlation visibilities in the following section. It is not easy to distinguish between these two effects in the autocorrelation visibilities alone. Direct electromagnetic simulations of mutual

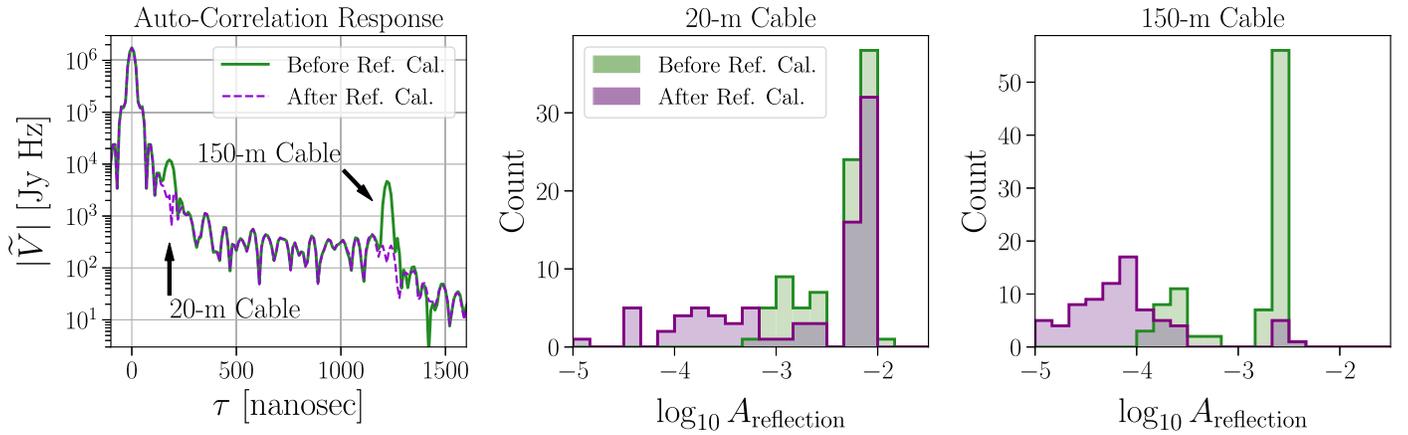


Figure 5. Reflection calibration performed over the full band (120–180 MHz) and applied to the autocorrelation visibilities. Left: autocorrelation response before (green) and after (purple) calibration demonstrating suppression of reflection systematics by roughly an order of magnitude in the visibility. Middle: histogram of derived 20 m reflection amplitudes before and after calibration. In the majority of cases, we only see suppression by a factor of a few. Right: histogram of derived 150 m reflection amplitudes before and after calibration. In the majority of cases, we see suppression by at least an order of magnitude. Less suppression for the 20 m cable is likely attributable to more significant frequency evolution in the reflection parameters.

coupling in the HERA system provide mixed evidence, predicting it to appear at a similar amplitude and slope in the autocorrelations but also predicting it to truncate at lower delays of ~ 600 ns (Fagnoni et al. 2019).

The fact that the autocorrelations show a systematic tail that, for $\tau > 300$ ns or $k_{\parallel} > 0.2 h \text{ Mpc}^{-1}$, shows only 3–4 orders of magnitude of dynamic range is concerning, given that fiducial EoR amplitudes are generally assumed to lie at or below 5 orders of magnitude in dynamic range in the visibility for similar k (Thyagarajan et al. 2016). Furthermore, the observed systematic tail extends over a wide range of delays that covers essentially all of the k_{\parallel} modes of interest ($0.2 h \text{ Mpc}^{-1} < k < 0.6 h \text{ Mpc}^{-1}$). These systematics need to be well understood and mitigated if the data are to be used for stringent EoR limits.

Next, we attempt to model some of these features and calibrate them out. One needs to proceed carefully when doing this because calibrating out structure that is inherent to the true data will actually create systematics. To be conservative, we only target the two features that we know to correlate with the expected delays of the 20 and 150 m coaxial cables at ~ 200 and ~ 1250 ns. We use the method described in Kern et al. (2019a) to derive reflection parameters across the full bandwidth, excluding the band edges (120–180 MHz), and then apply them to the data in frequency space. Figure 5 shows the result, demonstrating the delay response of an autocorrelation before (green) and after (purple) reflection calibration and showing the derived reflection amplitudes of the 20 and 150 m cable reflections before and after calibration. We find that, in general, we can suppress the 150 m cable reflection by a couple orders of magnitude (in the visibility), whereas for the 20 m cable reflection, we get, on average, only a factor of a few suppression.

Often a limiting factor in reflection modeling is frequency evolution of the reflection parameters (Ewall-Wice et al. 2016b). In Figure 6, we plot the autocorrelation response, having taken the Fourier transform of the data over a low band (120–150 MHz; blue) and a high band (150–180 MHz; gold) plotted in decibels relative to their peak value. Between the split subbands, we observe only slight evolution for the 150 m cable bump (labeled) but see more evolution in the 20 m cable bump (labeled). Furthermore, compared to the 150 m bump

feature, the low-level fluctuations from 200 to 1200 ns are also substantially different between the low- and high-band spectral windows. Recall that the noise floor of the data in this case is around -50 dB, meaning that the observed fluctuations at -35 dB are over an order of magnitude above the noise floor. This is likely at least part of the reason why we achieve less suppression for the 20 m cable reflection and suggests that to mitigate reflections to higher dynamic range, we will need to perform reflection calibration at the subband level. Because we find that the suppression achieved by modeling these reflections across the full band is sufficient for this analysis (Section 5), we defer subband reflection modeling to future studies.

An immediate concern one might have about this technique is the fact that we are applying a calibration with spectral structure at the same or similar delays we hope to use for measuring the EoR power spectrum (Grobler et al. 2014; Mouri Sardarabadi & Koopmans 2019), which may lead to signal loss (Cheng et al. 2018). In our companion paper, we study signal loss in this same scenario with simulated reflection systematics, and we find that although the autocorrelation visibilities may sustain low levels of signal loss, the cross-correlation visibilities show resistance to signal loss across all delays (Kern et al. 2019a). This is in part due to the subspace that reflection calibration spans relative to the EoR signal: reflection calibration spans a direction-independent, antenna-based space, while EoR is fundamentally a baseline-dependent measurement. This is further compounded by the fact that our reflection calibration method only uses the autocorrelation visibilities to derive reflection parameters. We refer the reader to Kern et al. (2019a) for a more detailed description of the algorithm and our signal loss simulations.

4. Antenna Cross-couplings

Next, we turn our attention to HERA’s cross-correlation visibilities in order to probe for antenna cross-coupling systematics. Specifically, we look at the north–south instrumental polarization (also denoted as “YY”) for baselines (11, 12), (11, 13), and (11, 14), which are three east–west baselines with lengths of 15, 29, and 44 m, respectively (Figure 2). These baselines display some of the strongest cross-coupling systematics seen in the data but are otherwise fairly nominal baselines.

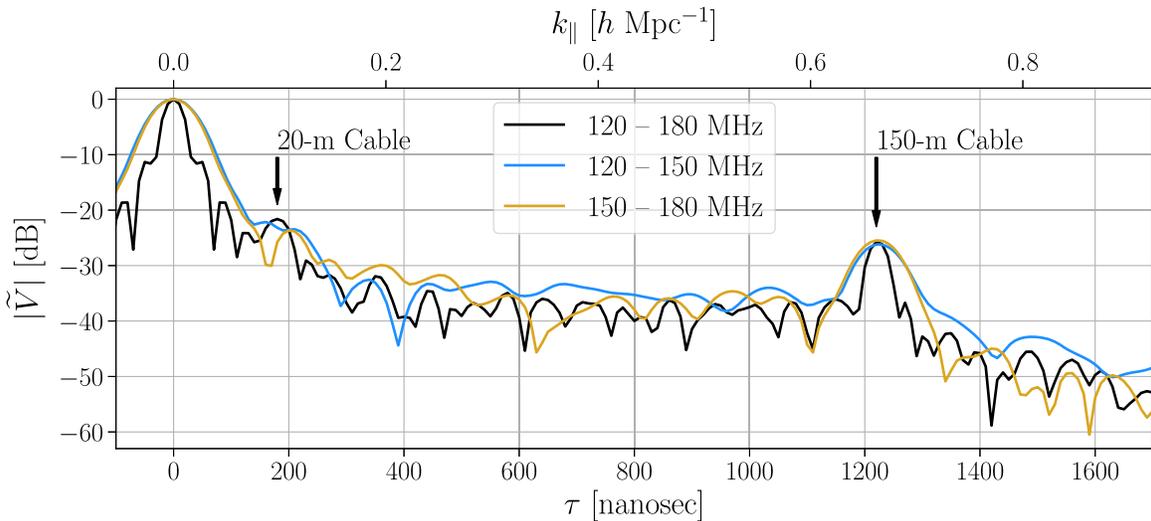


Figure 6. Autocorrelation visibility after complex time averaging, transformed over the full band (120–180 MHz; black), just the low side of the band (120–150 MHz; blue), and just the high side of the band (150–180 MHz; gold) for a type 1 signal chain. The 150 m cable reflection parameters are fairly consistent between both sides of the band, while the 20 m cable reflection shows significantly more frequency evolution. The smaller peaks along the systematic tail also shows significant frequency evolution.

In a similar fashion as before, we perform a delay-space deconvolution to fill in missing data due to RFI flags and suppress its side lobes in the delay domain. We allow the deconvolution to set model components out to $|\tau| < 1600$ ns and iterate down to $5\times$ the noise floor of the visibilities. Figure 7 shows visibility spectra from the three baselines of interest after deconvolution. We can clearly see a fast ripple on all baselines with a spectral scale of roughly 1 MHz. We also see larger-scale ripples (particularly at the lower half of the band) that decrease in spectral scale with increasing baseline length. As we will see below, the former is likely a combination of a cross-coupling and reflection systematic, while the latter may also be a form of cross-coupling systematic.

Next, we window the visibilities from 120 to 180 MHz with a Blackman–Harris function (Blackman & Tukey 1958) to limit spectral leakage and then Fourier transform the visibilities to delay space. At the moment, we are only interested in diagnosing systematics, so we do not square the Fourier amplitudes as we would in forming power spectra, meaning the visibilities are in units of Jy Hz. Figure 8 shows the result for the 15 (blue), 29 (orange), and 44 (green) m baselines. Also plotted as dashed vertical lines are the geometric horizons for each baseline. The nearly symmetric peaks at each baseline’s geometric horizon could be due to the “pitchfork” effect predicted to exist for wide-field radio interferometers (Thyagarajan et al. 2015). The pitchfork effect is not a systematic in the context of this work; it is a natural phenomenon from diffuse foregrounds and explained as the boosting of measured diffuse sky power near the horizon, where sky signal shows up in the visibilities with delays of the baseline’s geometric horizon. While HERA has a more compact primary beam compared to other low-frequency 21 cm experiments (e.g., MWA, PAPER), the pitchfork effect was nonetheless predicted to exist from simulations of the HERA dish and feed (Thyagarajan et al. 2016). However, these features could also be due to sky emission reflecting off the feed of one antenna and entering the feed of a neighboring antenna (i.e., feed-to-feed reflections or mutual coupling), which is a form of antenna cross-coupling that we would also expect to appear at the delay of each baseline’s geometric horizon. While both are expected to produce power at a baseline’s geometric horizon,

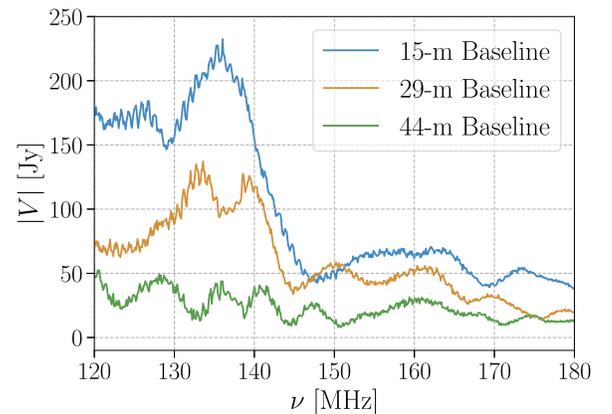


Figure 7. Cross-correlation visibility amplitudes in frequency space for three east–west-oriented baselines increasing in length from 15 up to 44 m at an LST of ~ 6 hr. In addition to a broad-scale ripple that decreases in spectral scale with increasing baseline length (most apparent at lower frequencies), we can also see a fast ripple at roughly a 1 MHz scale in all baselines that is likely due to a cross-coupling systematic.

both are also expected to be slowly time-variable, meaning that they will occupy similar modes in the delay and fringe-rate Fourier domains.²⁵

In Figure 9, we compare the data against a simulated diffuse foreground visibility from Kern et al. (2019a), which uses the Global Sky Model (GSM; de Oliveira-Costa et al. 2008) as the foreground model and a simulated direction-dependent primary beam response for HERA (Fagnoni et al. 2019). While we do see evidence for a slight pitchfork effect in the simulated data at the geometric horizon delay, its amplitude is considerably weaker than what is observed in the data. There is also some total power missing from the $\tau = 0$ mode, which is likely due to our exclusion of point sources in the simulation. The simulated

²⁵ Cross-coupling produces slowly time-variable signals in the visibility because it inserts a copy of the autocorrelation, which is slowly time-variable. The pitchfork mechanism is a mimicking of the autocorrelation at declinations near the horizon; thus, we expect it to have a slow time variability like the autocorrelation.

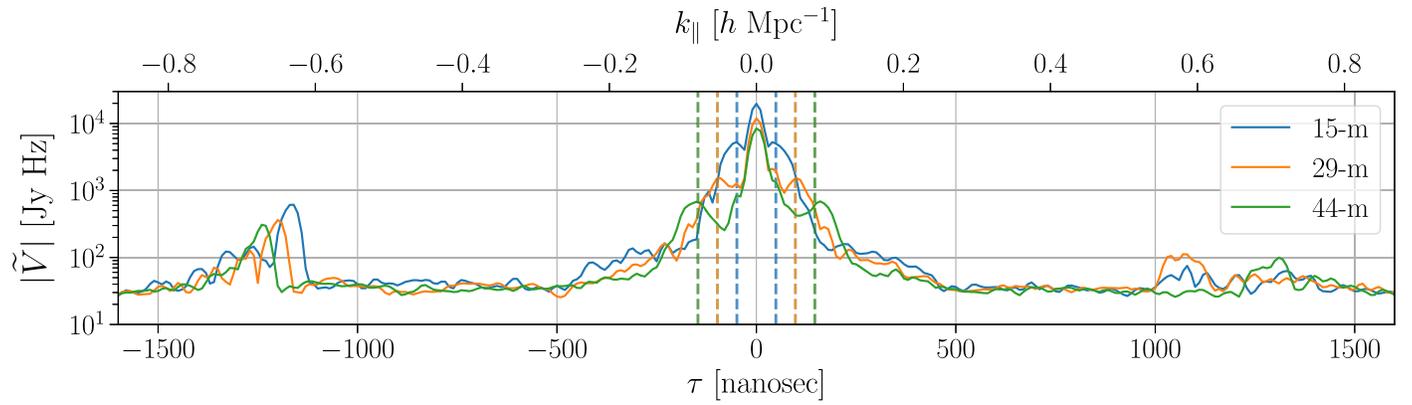


Figure 8. HERA cross-correlation visibilities averaged in amplitude across LST for three east–west baselines of increasing length: 15, 29, and 44 m (blue, orange, and green, respectively). The dashed vertical lines represent the geometric delay of the horizon for each baseline, within which foreground emission is nominally bounded. We see spikes in amplitude at the geometric horizon (“low-delay spikes”) and also at higher delays of $|\tau| > 700$ ns (“high-delay spikes”). The low-delay spikes are thought to be either a pitchfork effect, as predicted by Thyagarajan et al. (2015), or antenna cross-coupling. Evidence suggests the high-delay features to be some kind of cross-coupling systematic.

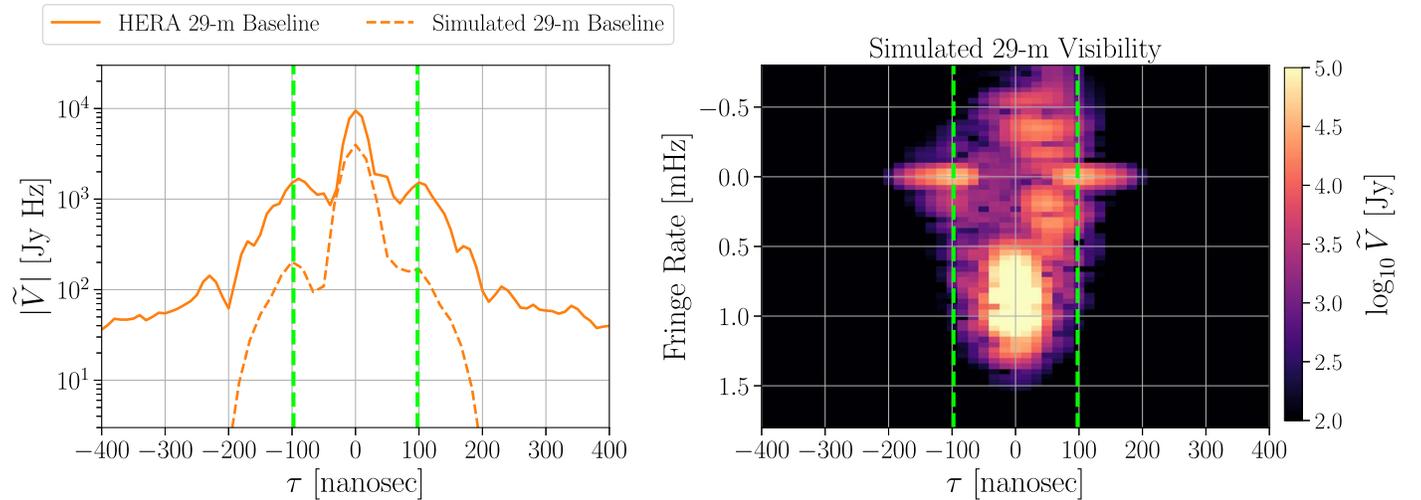


Figure 9. Comparison of HERA data with a simulated foreground visibility using the diffuse GSM sky for a 29 m east–west baseline. Left: averaged HERA cross-correlation visibility amplitude in delay space (solid) with an equivalent data product from a simulated foreground visibility with a matching LST range (dashed). The geometric baseline horizon is shown at ~ 100 ns (dashed green). While we see some evidence for a slight pitchfork-like structure in the simulated visibility, it is significantly weaker than the power bumps at equivalent delays in the real data. Right: simulated visibility transformed to fringe rate and delay space, with the geometric baseline horizon overplotted (dashed green). We can more clearly see the existence of the pitchfork effect in this plot, which is centered at $f = 0$ MHz, extends out to the geometric horizon, and falls off after.

pitchfork can be seen more clearly when transforming the simulated visibility into fringe rate and delay space (right panel of Figure 9), where indeed we see the pitchfork occupying $f \sim 0$ MHz modes, as expected. This comparison needs further study to better understand the nature of excess power at the horizon delay: the pitchfork feature is highly dependent on the adopted primary beam response at the horizon, which is typically the least accurate aspect of the simulated primary beam response and also hard to characterize in the field (Lanman et al. 2019). Future work using a combination of perturbed primary beam simulations and empirical beam constraints will help us better disentangle these effects in HERA data.

We also see evidence in Figure 8 for nonnegligible amounts of spillover of foreground emission (or suprahorizon emission) beyond the baseline’s geometric horizon, which has also been observed by other 21 cm experiments (e.g., Pober et al. 2013b; Beardsley et al. 2016). Suprahorizon emission can come naturally from the intrinsic spectral structure of the foregrounds.

It can also be created by the chromaticity of the instrumental gain that pushes out structure inherently contained within the geometric horizon or from low-level artifacts in the data that have a similar effect (Offringa et al. 2019). As noted above, the antenna-based gains we apply to the data are simplified to a single flux scaling and delay, meaning that part of this suprahorizon emission may be due to uncalibrated instrumental gain terms, which we do not explore in this work. For a foreground-avoidance approach to estimating the 21 cm power spectrum, the presence of suprahorizon emission is highly concerning because it limits our ability to measure the low k modes that, in theory, probe the EoR at the highest signal-to-noise ratio. The upside is that if suprahorizon emission is slowly time-variable (as are both the pitchfork effect and antenna cross-coupling systematics), then regardless of its origin, we can mitigate it by filtering it off in Fourier space. Indeed, this is exactly the principle that cross-coupling subtraction algorithms are founded upon.

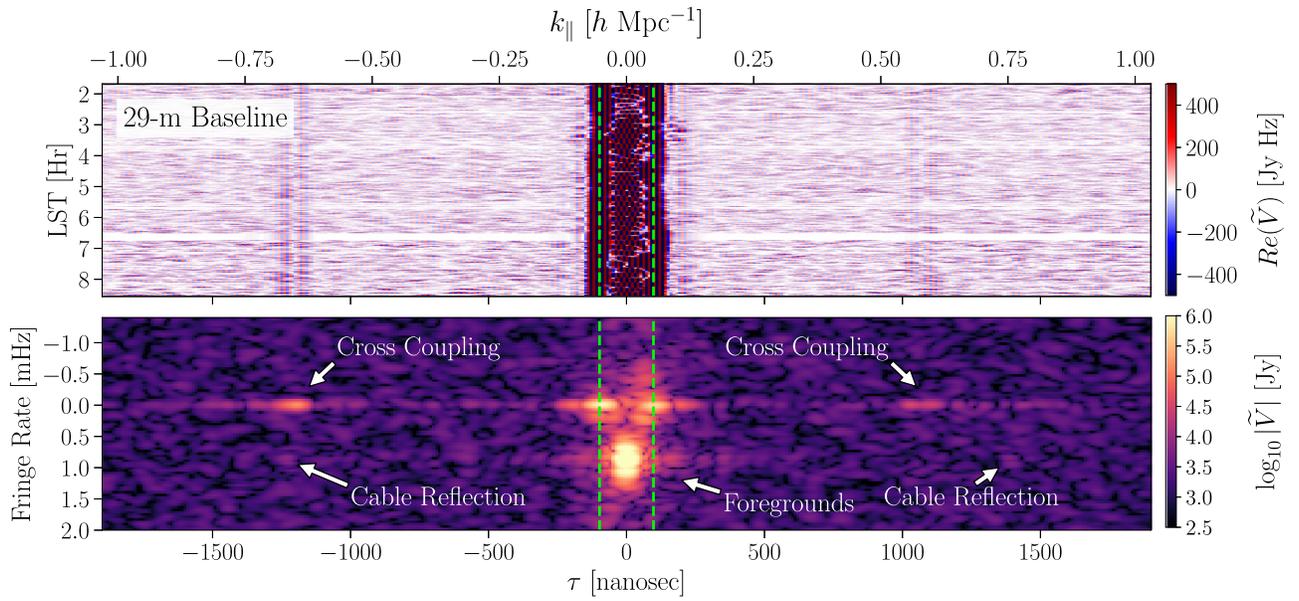


Figure 10. HERA cross-correlation visibility showing foregrounds, cable reflections, and cross-coupling systematics. Top: real component of the visibility in time and delay space, showing foreground power falling within the geometric horizon (green dashed). Notice that power well within the horizon fringes quickly as a function of time, while power near the geometric horizon shows much slower time variability and has spillover to outside the baseline’s horizon. Bottom: visibility amplitude in fringe rate and delay space. Here we can see the slowly time-variable systematics confined to $f \sim 0$ MHz fringe-rate modes, while foreground power is boosted to positive fringe rates. In addition, although not visible in the top plot, we can see the cable reflection just barely visible from the background noise, which appears at positive fringe rates because it is merely a copy of the intrinsic foreground signal.

Another striking feature in Figure 8 is the large amount of excess power above the noise floor at high delay ($|\tau| > 700$ ns). These features, which we refer to as the “high-delay” spikes, exhibit some very peculiar behavior. First, these features seem to be highly baseline-dependent: the three baselines shown in this section are all tied to antenna 11, yet their structures do not seem to be significantly correlated between the baselines. Second, their profile as a function of delay does not show isolated, individual peaks, as one might expect from one or a few feed-to-feed reflections, but rather a wide range of delays corrupted by excess power. Third, while the structures show up roughly near the delays where we would expect reflections from the 150 m cable to appear, they also show up at delays significantly smaller, enough to necessitate a considerably shorter cable length than 150 m, which is unlikely. The high-delay spikes exhibit slow time variability, with their power centered at $f = 0$ MHz, as we would expect from a cross-coupling systematic. Figure 10 shows the cross-correlation visibility from the 29 m baseline in time and delay space (top), as well as in fringe rate and delay space (bottom), where we recall that the latter is merely the Fourier transform of the former across time. We can clearly see that the high-delay structures are slowly variable, both by their slow movement as a function of time in the top plot and the fact that their power is centered at $f = 0$ MHz in the bottom plot. This is in contrast to the foreground power centered at $\tau = 0$ ns, which oscillates rapidly as a function of time and is therefore boosted to positive fringe rates, with the exception of the power at the baseline’s geometric horizon (dashed green), which, like the systematics at high delay, exhibits slow time variability centered at $f = 0$ MHz.

What we cannot see by looking at the visibility in time and delay but can barely begin to discern when we transform to the fringe-rate domain are the cable reflections at $|\tau| \sim 1300$ ns. As we saw in Figure 4, the measured reflection amplitudes are roughly 3×10^{-3} times the peak power in the visibility.

Because the high-delay spikes at $f = 0$ MHz also show up at similar delays and are stronger in amplitude, we cannot see the cable reflections in Figure 8 or in the top panel of Figure 10 buried under the other systematics. Reflections have the same time structure as the unreflected signal, so by transforming to fringe-rate space, we can isolate them from the slowly time-variable systematics, and indeed, we can just barely see them above the noise floor of the cross-correlation visibilities at roughly 3×10^{-3} times the main foreground power, as expected. Figure 10 also shows evidence for the suprahorizon emission having two distinct components: one that has fast time variability, like foregrounds from the main lobe of the primary beam, and another that is slowly fringing, like a cross-coupling systematic or pitchfork effect, and both extend considerably beyond the baseline’s geometric horizon.

Currently, there is no single physical model for the origin of the high-delay spikes that can explain all of the behavior observed in the data. In Appendix A, we explore some simple physical models for the systematic and show that we can tentatively rule them out; however, further work is needed to more fully understand their origin. Nonetheless, their temporal behavior is suggestive of some kind of antenna cross-coupling that occurs at some point along the signal chain. At present, what we can say with certainty is that their time dependence is highly inconsistent with an EoR signal, and as such, we can suppress it by filtering the data in fringe-rate space before forming power spectra (see Kern et al. 2019a for details on why this is inconsistent with an EoR signal).

With that in mind, Figure 11 shows the result of running an singular value decomposition (SVD)-based cross-coupling model (Kern et al. 2019a) on the 29 m baseline data, which decomposes the matrix shown in the top panel of Figure 10 into orthogonal time eigenmodes (T), orthogonal delay eigenmodes (D), and their singular values (S). Before taking the SVD, we apply a band-stop window on the data matrix that assigns zero weight to all delay

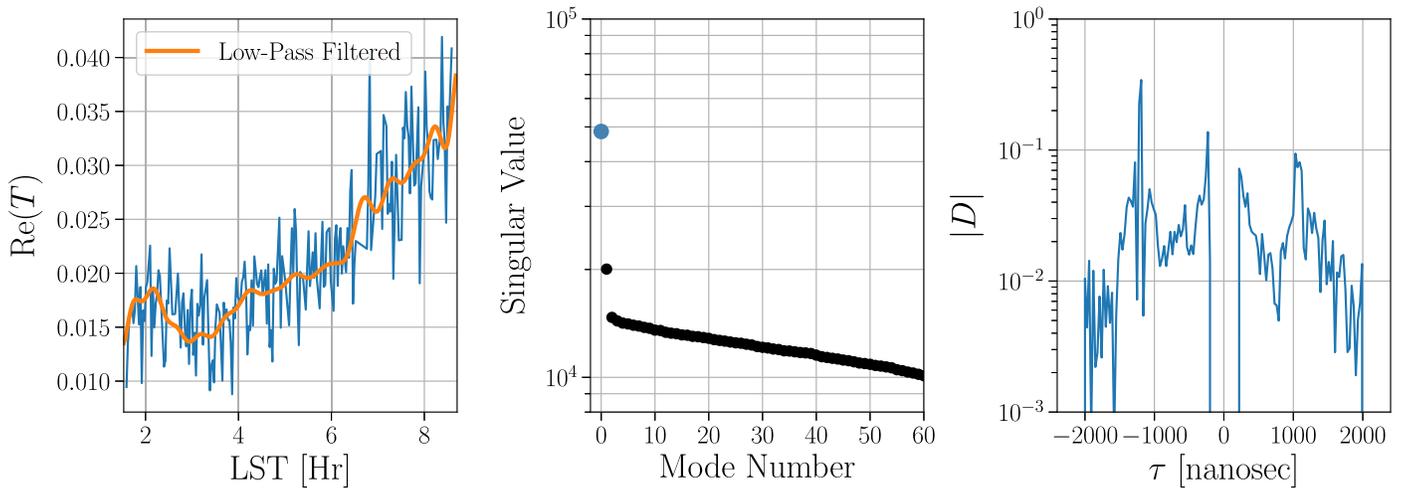


Figure 11. Singular value decomposition of the 29 m east-west baseline visibility from Figure 10. Left: first T eigenvector across time, showing its raw (blue) and low-pass filtered (orange) forms, having filtering out modes with $f > 0.46$ MHz with a Gaussian process model (Kern et al. 2019a). Middle: first 60 singular values, showing that most of the variance in the systematic-prone regions can be described with a handful of modes before a noise plateau is reached. Right: first D eigenvector across delay, showing it picking up on the slowly variable structure at large delays ($|\tau| \sim 1200$ ns) and also some structure near the baseline horizon ($|\tau| \sim 200$ ns).

modes outside of the range $200\text{ns} < |\tau| < 2000$ ns, which was chosen to encompass most of the observed cross-coupling systematics and reject the foregrounds at very low delays. The left panel plots the first T eigenmode across time, showing the raw eigenmode (blue) and the eigenmode after low-pass filtering it out to $f_{\text{max}} = 0.46$ MHz (orange). We use the Gaussian process-based filter explored in Kern et al. (2019a) to low-pass filter these time modes. The middle panel shows the first 60 singular values, giving us a sense of how much the information content is isolated into the first few eigenmodes. We find that most of the structure can be described with only a handful of modes before reaching a plateau. In forming the systematic model, we keep the top 30 modes out of ~ 1000 and truncate the rest. We choose 30 based on inspection of Figure 11, which gets the strongest first few modes but also tries to get some of the modes after the plateau, as they may be picking up on the systematic at a low level. Lastly, the right panel shows the first D eigenmode across delay, showing it picking up the high-delay cross-coupling systematic and some of the suprahorizon emission at low delay. In addition to picking up on the systematic, the SVD will pick up on the noise of the data as well. However, because we keep only a small fraction of the eigenmodes and additionally smooth them across time, we do not suspect that we are subtracting a significant component of the noise in the process of systematic removal. For the 29 m baseline, the number of Fourier modes kept in the low-pass smoothing filter is $\sim 1/80$, the total number of Fourier modes in the data.

We repeat this for the other baselines at hand, low-pass filtering the T basis vectors from the 15 and 44 m baselines with $f_{\text{max}} = 0.14$ and 0.83 MHz, respectively, using a Gaussian process-based smoothing for the low-pass filter. See Table 1 and Appendix B of Kern et al. (2019a) for more details on this process. Figure 12 shows the baselines in Figure 8 after cross-coupling subtraction, with the vertical dashed lines showing the minimum delay of the cross-coupling model at $\tau = 200$ ns. The top panel shows only cross-coupling subtraction, where we see significant suppression of the high-delay spikes and the outer edge of the low-delay spikes. As expected, after subtracting the strong cross-coupling terms at high delay, we are left with the appearance of localized bumps that mark the cable reflections

(gray shaded regions), which, we recall, were not subtracted out with the cross-coupling because they occupy fringe-rate modes that were filtered out of the systematic model in the process of smoothing. The bottom panel shows the data after applying reflection calibration from Section 3 and cross-coupling subtraction, showing that the data are now consistent with a scale-independent thermal noise floor for all delays outside $|\tau| > 500$ ns.

There is, however, still a slight slope in the data at intermediate delays of $200\text{ns} < |\tau| < 500$ ns, which is part of the suprahorizon emission we observed earlier. To ensure that this tail is not coming from the cross-coupling component that we attempted to filter out, we can plot the systematic-subtracted data in fringe rate and delay space, which is shown in Figure 13, with the blue dashed region showing the region of Fourier space where cross-coupling subtraction was performed. Figure 13 confirms that the excess signal in the range $200 < |\tau| < 500$ ns observed in Figure 12 does not come from modes that should have been subtracted in the process of cross-coupling removal and originates from the second suprahorizon component at higher fringe rates. As discussed above, this suprahorizon emission can come from uncalibrated bandpass terms or low-level artifacts in the data, which push foregrounds out in delay that were intrinsically contained within the geometric horizon. These effects can be somewhat mitigated with better bandpass calibration and data flagging, but they are still active areas of research in the literature. Additionally, the slight overlap of low fringe-rate power inside the dashed region at $|\tau| = 200$ ns is produced by a windowing function applied to the data before taking its Fourier transform.

Signal loss is a principal concern when applying any baseline-dependent operation to the data, as we have done with cross-coupling subtraction. In Kern et al. (2019a), we vetted our cross-coupling modeling algorithms for EoR signal loss against numerical visibility simulations of the HERA Phase I system. We show that by low-pass filtering the systematic model along time (Figure 11), we can harden our systematic model against EoR signal loss to an almost arbitrary level. In our case, we chose the fringe-rate bounds above by adopting a

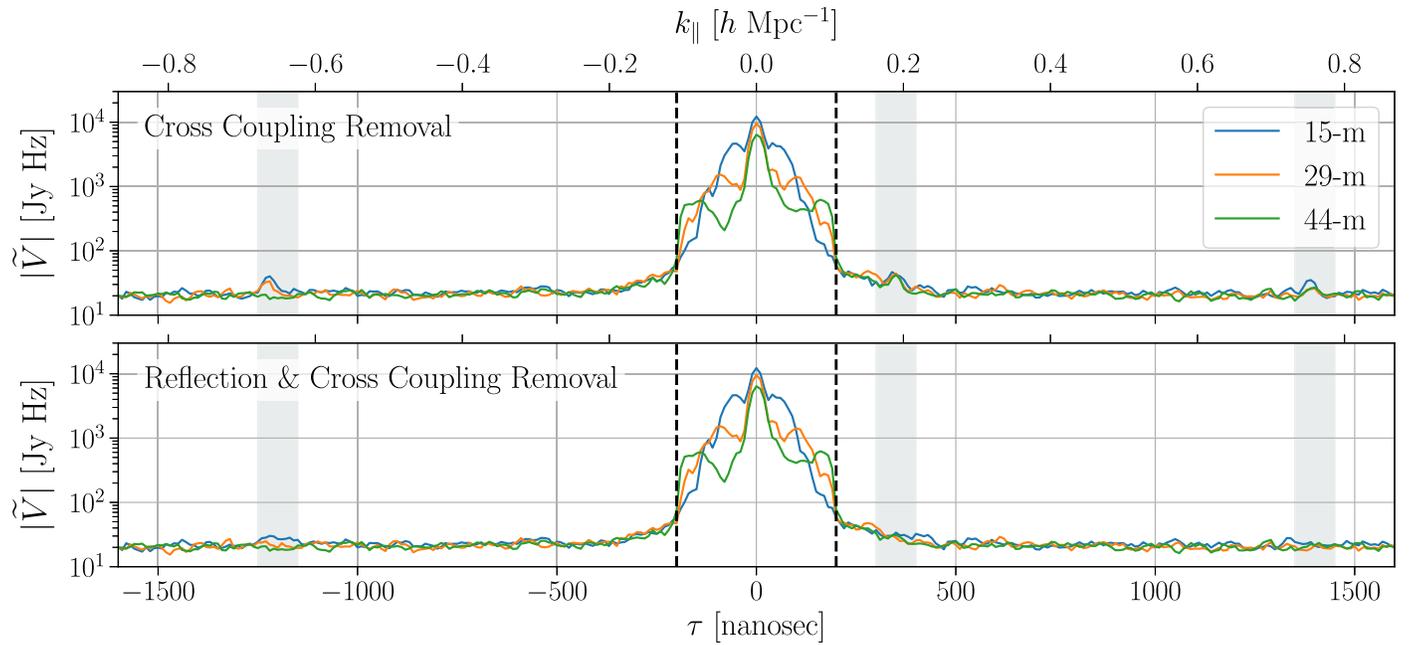


Figure 12. HERA cross-correlation visibilities from Figure 8 after cross-coupling subtraction but before reflection calibration (solid) and after both cross-coupling subtraction and reflection calibration (dashed). The black dashed lines represent the lower delay boundary of the cross-coupling model. Gray shaded regions indicate expected delays for reflection systematics having inspected the autocorrelations for peaks. Joint systematic suppression yields cross-correlations visibly free of systematics at the level of the per-baseline noise floor.

signal loss tolerance of 1% in EoR power, which is below the expected measurement error of the full HERA array. We refer the reader to our analysis and discussion in that paper for more details on signal loss quantification in the context of cross-coupling removal.

5. Power Spectrum Estimation

Now that we have demonstrated that we can suppress reflection and cross-coupling systematics for a few baselines down to their individual noise floors, we would like to prove that we can similarly do this for baselines across the entire array and confirm that these systematics are a nonlimiting factor in the power spectrum even after redundant baseline averaging. We will focus on the same three baseline orientations (14, 29, and 44 m east–west baselines) but now look at all baselines within the array that fall within each baseline group.

To estimate the three-dimensional 21 cm power spectrum, $P_{21}(\mathbf{k})$, we use the delay spectrum estimator (Parsons et al. 2012a, 2014; Liu et al. 2014). The delay spectrum is a per-baseline, visibility-based power spectrum estimator that relies on the Fourier transform of the visibility across frequency into the delay (τ) domain,

$$\tilde{V}(\mathbf{u}, \tau) = \int d\nu e^{2\pi i \nu \tau} V(\mathbf{u}, \nu), \quad (1)$$

where $\mathbf{u} = \mathbf{b}/\lambda$ is the baseline vector divided by the observing wavelength. The “delay transform” of the visibility is not a direct measurement of the line-of-sight cosmological k_{\parallel} mode, due to an interferometer’s inherent chromaticity (Morales et al. 2012). Approximating it as such is known as the “delay approximation,” which was shown to be a good approximation for short baselines and is one of the motivating factors behind HERA’s compact design (Parsons et al. 2012a; Dillon & Parsons 2016). We refer the reader to Morales et al. (2019) for a broader

discussion of various 21 cm power spectrum estimators. The delay spectrum approximation of the 21 cm power spectrum is then the square of the delay-transformed visibility with the appropriate scaling factors,

$$\hat{P}_{21}(k_{\perp}, k_{\parallel}) \approx |\tilde{V}(\mathbf{u}, \tau)|^2 \frac{X^2 Y}{\Omega_{pp} B_p} \left(\frac{c^2}{2k_B \bar{\nu}^2} \right)^2, \quad (2)$$

where X and Y are redshift-dependent scalings converting sky angles and frequencies to cosmological length scales, Ω_{pp} is the sky integral of the squared antenna primary beam response, $\bar{\nu}$ is the delay transform center frequency, and B_p is the delay transform bandwidth, as defined in Appendix B of Parsons et al. (2014). The factors relating the \mathbf{u} and τ Fourier domains inherent to the telescope to the cosmological Fourier domains of k_{\perp} and k_{\parallel} are

$$\begin{aligned} k_{\parallel} &= \frac{2\pi}{X} \tau, \\ k_{\perp} &= \frac{2\pi}{Y} \frac{b}{\lambda}, \end{aligned} \quad (3)$$

where $X = c(1+z)^2 \nu_{21}^{-1} H(z)^{-1}$, $Y = D(z)$, $\nu_{21} = 1.420$ GHz, $H(z)$ is the Hubble parameter, $D(z)$ is the transverse comoving distance, b is the baseline length, and λ is the observing wavelength (Parsons et al. 2012b; Liu et al. 2014). In this work, we adopt cosmological parameters from the *Planck* mission’s 2015 analysis (Planck Collaboration et al. 2016), namely $\Omega_{\Lambda} = 0.6844$, $\Omega_b = 0.04911$, $\Omega_c = 0.26442$, and $H_0 = 67.27$ km s $^{-1}$ Mpc $^{-1}$.

Cross-multiplying a visibility with itself in Equation (2) to form a delay spectrum will result in an overall bias in power due to the noise present in the data. To avoid this, we take visibility spectra adjacent to each other in LST separated by 10.7 s and apply a phasing term to align their phase centers

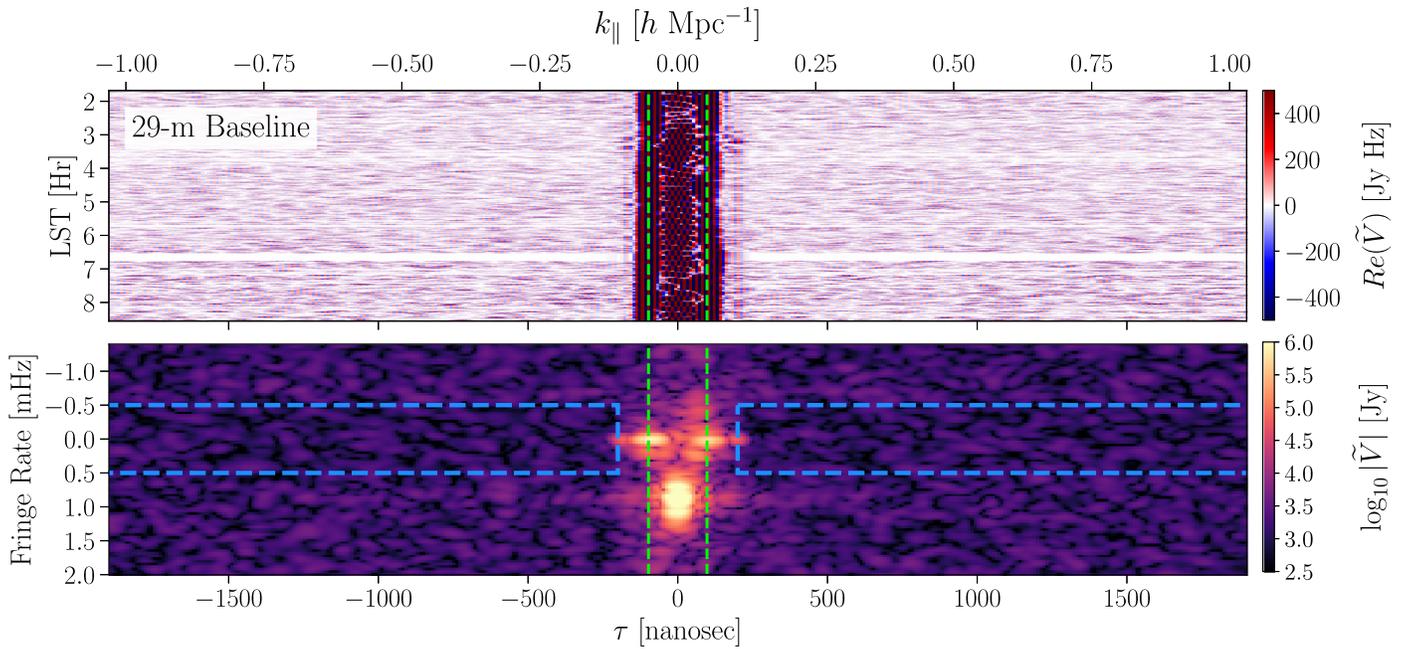


Figure 13. Same 29 m visibility in fringe rate and delay space as shown in Figure 10 but now with reflection and cross-coupling systematics removed. The blue dashed region shows where the cross-coupling algorithm modeled and removed systematics, and the green dashed lines mark the baseline’s geometric horizon.

before cross-multiplication (Pober et al. 2013a). This means that the two visibilities to leading order measure the same cosmological mode on the sky but have uncorrelated noise realizations, such that they do not produce a noise bias upon cross-correlation.

Thermal noise in interferometric visibilities is mean-zero, Gaussian-distributed, and statistically uncorrelated on all time and frequency scales; however, it generally is nonstationary and will have an amplitude dependence as a function of LST and frequency. A signal chain’s system temperature is proportional to the total amount of noise power received by the analog system and is the sum of the sky and receiver noise,

$$T_{\text{sys}}(\nu, t) = T_{\text{sky}}(\nu, t) + T_{\text{rcvr}}(\nu, t) \quad (\text{K}). \quad (4)$$

In practice, antenna signal chains will have variable system temperatures due to different angular primary beam responses and receiver properties. A visibility-based system temperature can therefore be estimated, which is the system temperature as measured by a particular baseline. This can be estimated by taking differences of adjacent pixels in time and frequency and relating the rms to a system temperature via the radiometer equation,

$$\sigma_{\text{rms}}^{ij} = \frac{2k_b\nu^2}{c^2\Omega_p} \frac{T_{\text{sys}}^{ij}}{\sqrt{\Delta\nu\Delta t}}, \quad (5)$$

where σ_{rms}^{ij} is the rms of the visibility between antennas i and j in jansky, k_b is the Boltzmann constant, ν is the average observing frequency, Ω_p is the angular integral of the peak-normalized primary beam response in steradians, $\Delta\nu$ is the correlator channel width in hertz, and Δt is the correlator integration time in seconds (Thompson et al. 2017). Another estimate of the noise comes directly from the autocorrelation visibility, which itself is a measurement of the total power received by a particular antenna. For a cross-correlation visibility between antennas i and j , we can estimate the

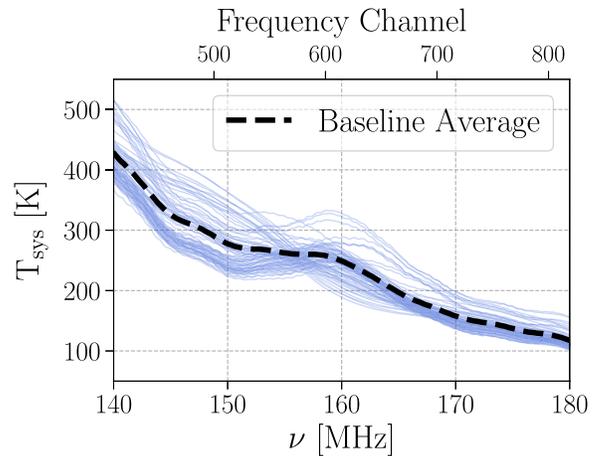


Figure 14. System temperature curves for all baselines used in the power spectral analysis (blue) and their average (black dashed). Delay spectra presented in this section are formed between channels 450 and 650 (144–163 MHz) with an effective system temperature of ~ 270 K.

baseline’s system temperature as

$$\sqrt{V_{ii}V_{jj}} = \frac{2k_b\nu^2}{c^2\Omega_p} T_{\text{sys}}^{ij}, \quad (6)$$

where V_{ii} is the autocorrelation visibility of antenna i . While both methods are comparable, we defer to using the autocorrelations, which, in practice, generally lead to more stable and cleaner noise models.

Figure 14 shows system temperature estimates for each baseline participating in the analysis (blue) and the averaged system temperature, which is each baseline’s system temperature averaged in quadrature. Again, because we have not corrected for the bandpass structure of the gains, the large-scale fluctuations in Figure 14 are not unexpected and would be smoothed out after solving for and applying the appropriate

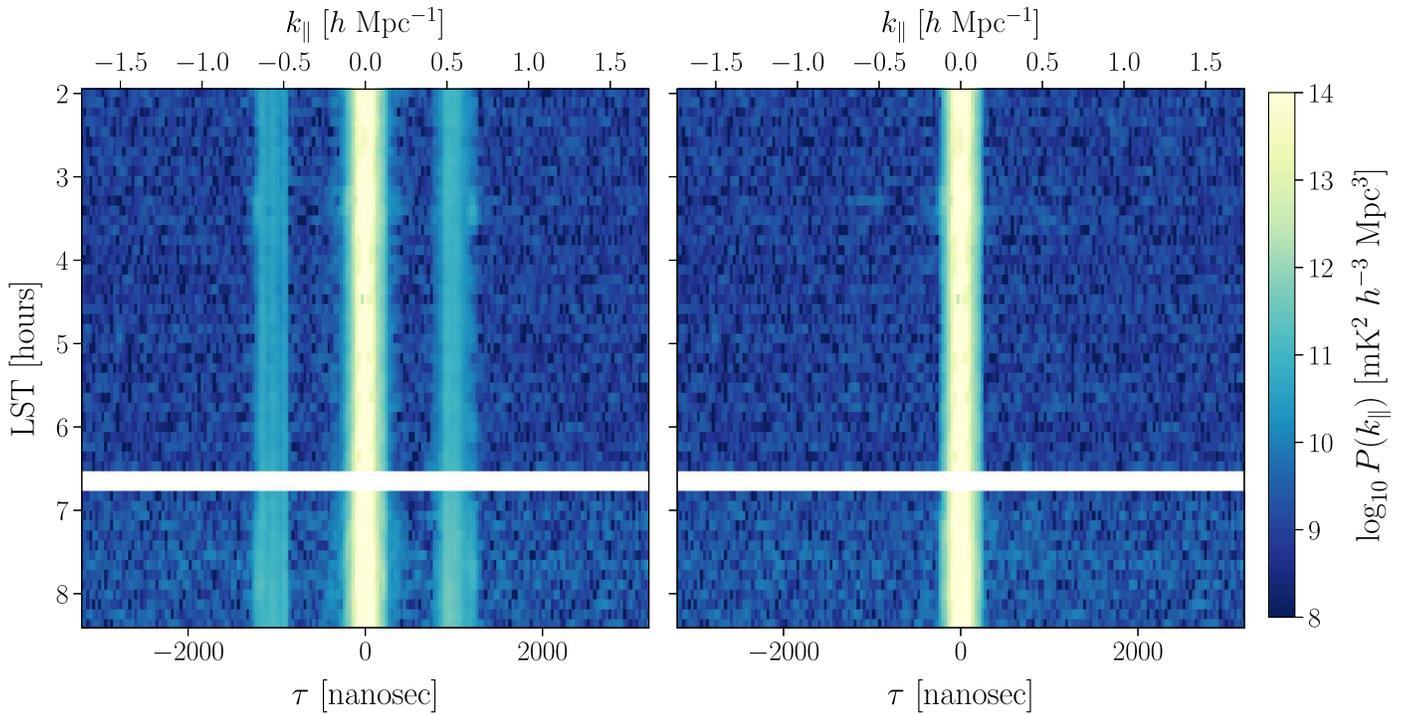


Figure 15. Averaged power spectrum waterfall of the east–west 15 m group showing the absolute value of the real component of the power spectra, having first incoherently averaged 35 separate baseline pairs in the group. We plot the data with systematics in (left) and with systematics removed (right).

instrumental gains. The presence of such structure in the noise curves does not change the fundamental results of this section. Power spectra presented in this section are formed between channels 450 and 650 (144–163 MHz), with an effective system temperature of ~ 270 K.

With an understanding of the noise properties of our data, we can compute a theoretical estimate of the noise power spectrum, P_N , which is equivalent to the rms of the power spectrum if the only component in the data were noise. This is one way to measure the uncertainty on the estimated power spectra but also represents the theoretical amplitude of the power spectra in the limit that they are noise-dominated (as opposed to signal- or systematic-dominated). This is given in Cheng et al. (2018) as

$$P_N = \frac{X^2 Y \Omega_{\text{eff}} T_{\text{sys}}^2}{t_{\text{int}} N_{\text{coherent}} \sqrt{2 N_{\text{incoherent}}}}, \quad (7)$$

where the X and Y scalars are the same as before, T_{sys} is the system temperature in millikelvin, t_{int} is the correlator integration time in seconds, N_{coherent} is the number of sample averages done at the visibility level (i.e., before visibility squaring), and $N_{\text{incoherent}}$ is the number of sample averages done at the power spectrum level (i.e., after visibility squaring). Here Ω_{eff} is the effective beam area given by $\Omega_{\text{eff}} = \Omega_p^2 / \Omega_{pp}$, where Ω_p is the integral of the beam across the sky in steradians, and Ω_{pp} is the integral of the squared-beam across the sky in steradians (Pober et al. 2013a; Parsons et al. 2014). We calculate P_N for each redundant group using the baseline-averaged system temperature.

The data are natively sampled at a 10.7 s cadence. Before forming power spectra, we coherently average each visibility across LST for 3.6 minutes (20 samples), applying a fringe stop in each averaging window to limit sky signal attenuation. We

select a wide spectral window between channels 400 and 700 (139–168 MHz) and apply a Blackman–Harris windowing function before transforming to Fourier space. Because the cosmological signal undergoes nonnegligible evolution within such a bandwidth, we would not normally use such a wide bandwidth for setting upper limits; however, we do this to achieve better resolution in delay space for diagnostic purposes. We then cross-multiply the visibilities and apply the necessary normalization factors as per Equation (2). For simplicity, we only form power spectra by cross-multiplying baselines with themselves (at adjacent times) and do not cross-correlate different baselines within redundant groups. Then we average the power spectra within each redundant group (i.e., an incoherent average). For the 15, 29, and 44 m groups, this involves averaging 35, 28, and 20 independent baselines, respectively.

What we are left with is a single complex-valued power spectrum waterfall for each redundant group as a function of LST and delay, consisting of 60 leftover time bins and 200 delay bins. In Figure 15, we show this for the 15 m group with and without systematic removal (right and left). In our final step, we take the real component of each power spectrum waterfall and average its absolute value over the remaining time bins with uniform weighting for each bin. This is done to make a higher signal-to-noise measurement of the noise floor at the level of the power spectrum waterfall: we could have gained more sensitivity by not taking the absolute value before averaging, but our point here is to make a visually clearer comparison with the known noise level rather than gain increased sensitivity. Figure 16 shows the power spectra of the data without systematic removal (blue) and with systematic removal (orange) and also shows the theoretical noise level given our visibility noise estimates and taking into account the various forms of averaging before and after squaring the

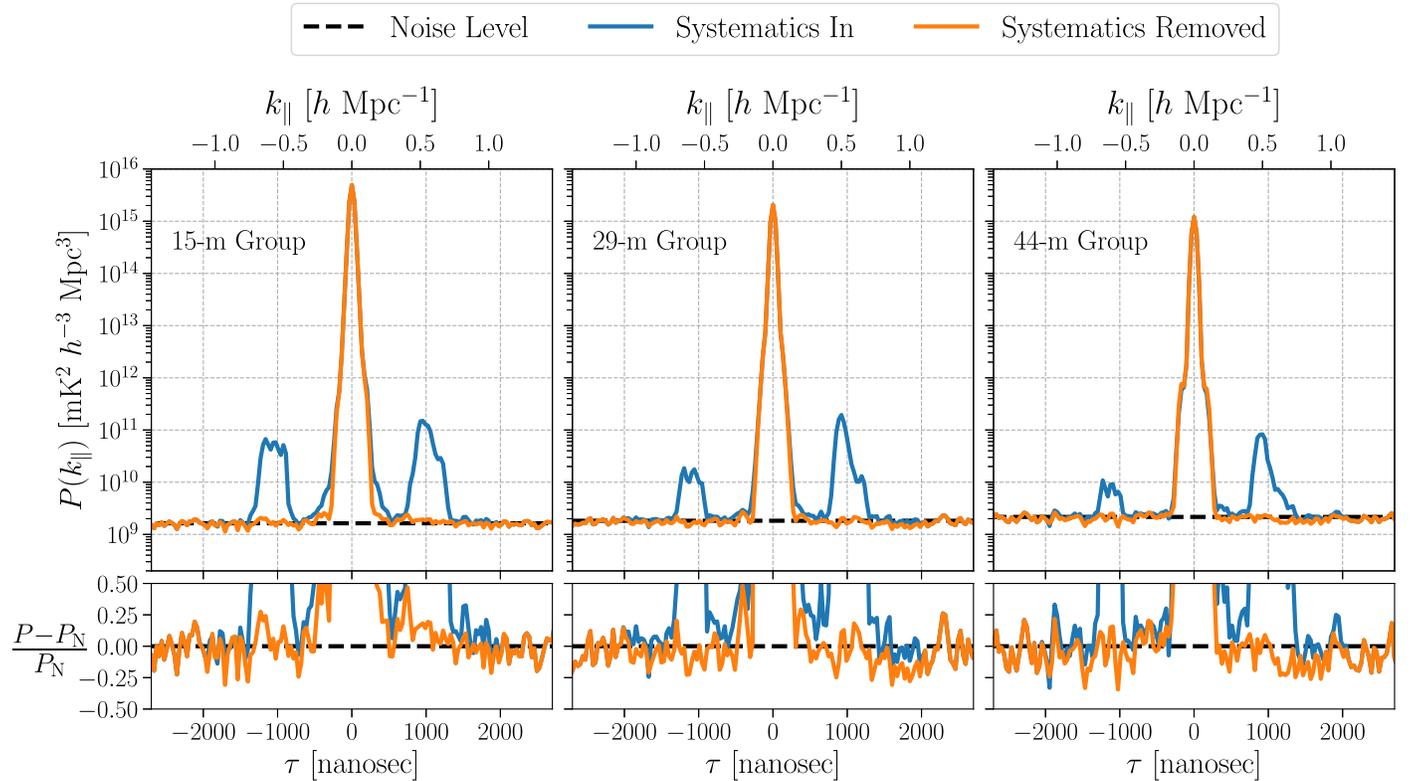


Figure 16. Delay spectra for three unique baseline lengths oriented along the east–west axis without systematic removal (blue) and with systematic removal (orange). The power spectra are formed directly from the visibilities for each baseline in the array and incoherently averaged within each redundant group; then, their absolute value is averaged across the remaining bins in LST. We see suppression of high-delay systematics down to the integrated noise floor and get some suppression of suprahorizon power at low delay.

visibilities (black dashed). In this case, the systematic removal includes both cross-coupling subtraction and reflection calibration. We find that we can suppress the observed systematics by roughly 2 orders of magnitude in power, enabling us to achieve 6 orders of magnitude in dynamic range with respect to the peak foreground power for $|k_{\parallel}| > 0.2 h \text{ Mpc}^{-1}$.

The power spectra in Figure 16 show general agreement with our prediction of the thermal noise floor for delays considerably outside of the foreground wedge. The bottom panels show the fractional offset of the data with respect to the analytic noise, showing that P_N is broadly consistent with the data at the $\sim 15\%$ level. While the 29 and 44 m groups seem to exhibit slight systematic offsets between the observed and predicted P_N , this offset is contained within the majority of the random fluctuations seen in the noise floor; without noise propagation trials, this offset is hard to quantify rigorously, which we defer to future work. Although the geometric horizon for these short baselines is on the order of 50–150 ns, the Blackman–Harris windowing function pushes this out by about +100 ns, such that their effective horizon is on the order of 150–250 ns. However, we can still see some amount of positive power near the transition region, particularly for the 15 m group. This could be due to uncalibrated bandpass terms in the data, low-level artifacts in the data missed by RFI flagging, or residual reflection and cross-coupling systematics. More complete gain calibration and deeper integrations will allow us to investigate this at higher signal-to-noise ratio levels.

A. Gosh et al. (2019, in preparation) also propose methods for subtracting systematics observed in the HERA Phase I instrument using a Gaussian process–based model. With their

model, they find good subtraction of the systematic down to similar dynamic ranges (10^6 in power), at the cost of possible signal loss at the $\sim 10\%$ level. Systematics of a similar nature were also observed in the HERA-19 commissioning array (Kohn et al. 2019). However, a direct comparison with this work is difficult because the array was reconfigured en route to the Phase I configuration.

As a final note, we would like to clarify how we came to the noise level plotted in Figure 16. Noise in the interferometric visibility is a complex Gaussian random variable, meaning that when we form power spectra by squaring the visibilities, we are left with a noise component that is drawn from a complex normal-product distribution. A real-valued normal-product distribution can be shown to be described by a modified Bessel function of the second kind of order zero (Wells et al. 1962; Cui et al. 2016). A complex-valued normal-product random variable is simply the sum of two real-valued normal-product random variables, which means its probability density function (PDF) is a convolution of the Bessel function with itself and turns out to be a double-sided exponential distribution. Therefore, after squaring the visibilities, noise in the power spectrum is drawn exponentially.

However, most power spectrum pipelines will average the data after squaring the visibilities (i.e., incoherent averaging), which will re-Gaussianize the data due to the central limit theorem. Indeed, to create Figure 15, we perform a few dozen incoherent averages across redundant baselines after squaring the visibilities, meaning that it is fair to assume the noise in our power spectrum is Gaussian-distributed. However, in order to collapse our data along the LST axis to form Figure 16, we took the absolute value of the real component of the power spectrum

before averaging. The absolute value operation transforms the noise from a normally distributed, mean-zero random variable into a random variable drawn from a half-normal distribution, which is no longer mean-zero and has an expectation value of $\sigma\sqrt{2/\pi}$. Recall from Equation (7) that P_N tells us the expected rms of the real (or imaginary) component of the complex power spectrum due to thermal noise. Therefore, the act of taking the absolute value of the real component of the power spectra and averaging across LST means we need to multiply our final P_N estimate by a factor of $\sqrt{2/\pi}$, which is what is actually plotted in Figure 16 as the black dashed line.

6. Summary

In this work, we investigate data from HERA Phase I for signal chain reflection and antenna cross-coupling systematics. We find cable reflections on the order of $\sim 10^{-3}$ in amplitude and a systematic tail in the autocorrelation visibilities straddling the EoR window at roughly the 10^{-3} – 10^{-4} level, which is considerably larger than that expected from simulations of the HERA dish and feed. If not mitigated, the systematic tail observed in the autocorrelations may prevent HERA Phase I from setting competitive upper limits on the EoR, let alone detecting it. We show that reflection calibration can help to suppress some of these features by about an order of magnitude in the visibility at specific k_{\parallel} modes. The presence of the systematic tail in the autocorrelation may be indicative of highly complex cable subreflections that will be hard to calibrate out down to EoR levels, even with the methods demonstrated here.

We also inspect the data for antenna cross-coupling systematics and find that they contaminate the data at high delays near the edge of the targeted EoR modes at $k_{\parallel} \sim 0.5 h \text{ Mpc}^{-1}$. We also find evidence for excess emission at each baseline’s geometric horizon that is likely due to either (1) a pitchfork effect (Thyagarajan et al. 2016) or (2) feed-to-feed mutual coupling. These features produce nonnegligible spillover into the EoR window and thus need to be controlled for foreground-avoidance power spectrum approaches. We investigate three east–west baselines of increasing length (15, 29, and 44 m) that exhibit particularly strong systematics and find that we can model and suppress both of the contaminating components in the EoR window down to the noise floor of the nightly observation.

We then form power spectra from three redundant groups for baselines across the entire array. We show that by combining reflection calibration and cross-coupling subtraction on specific baseline orientations, we can suppress systematics for $k_{\parallel} > 0.2 h \text{ Mpc}^{-1}$ down to the integrated noise floor of the array for a single nightly observation, with the exception of a weak suprahorizon tail at low k that merits further investigation through improved bandpass calibration and RFI flagging. Instrumental bandpass calibration for HERA Phase I is explored in Kern et al. (2019b) and J. Dillon et al. (2019, in preparation).

This work shows that the immediate systematics seen in the HERA Phase I system can be modeled and dealt with down to a dynamic range of 10^6 in the power spectrum against the peak foreground power, even with an extremely simple approach to direction-independent, antenna-based calibration. While this is reassuring, fiducial EoR levels are expected to appear at dynamic ranges of $\sim 10^{10}$ in the power spectrum for low- k modes (Thyagarajan et al. 2016). Future work with deeper integrations will help determine whether these systematics can continue to be mitigated down to EoR levels.

This material is based upon work supported by the National Science Foundation under grant Nos. 1636646 and 1836019 and institutional support from the HERA collaboration partners. This research is funded in part by the Gordon and Betty Moore Foundation. HERA is hosted by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Innovation. A.E.L. and J.C.P. would like to acknowledge NASA grant 80NSSC18K0389. A.L. acknowledges support from a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant and a Discovery Launch Supplement, as well as the Canadian Institute for Advanced Research (CIFAR) Azrieli Global Scholars program. Parts of this research were supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D) through project No. CE170100013. G.B. acknowledges funding from the INAF PRIN-SKA 2017 project 1.05.01.88.04 (FORECaST), support from the Ministero degli Affari Esteri della Cooperazione Internazionale—Direzione Generale per la Promozione del Sistema Paese Progetto di Grande Rilevanza ZA18GR02 and the National Research Foundation of South Africa (grant No. 113121) as part of the ISARP RADIOSKY2020 Joint Research Scheme, from the Royal Society and the Newton Fund under grant NA150184, and from the National Research Foundation of South Africa (grant No. 103424).

Appendix A

Physical Models for the Observed Cross-coupling

Here we explore the feasibility for some simple physical models as the origin of the “high-delay” cross-coupling systematics investigated in Section 4. In summary, we cannot find a single model that explains all of the observed behavior of the systematics, but we can tentatively rule out some simplistic models. In what follows, we adopt the mathematical conventions in Section 2 of Kern et al. (2019a) when discussing voltage spectra, visibilities, and coupling coefficients. Specifically, for two antennas, 1 and 2, with intrinsic voltage spectra v_1 and v_2 , we can write the voltage of antenna 1 corrupted by a cable reflection as

$$v'_1 = v_1(1 + \epsilon_{11}), \quad (8)$$

where ϵ_{11} is the cable reflection coefficient, and we can write the voltage of antenna 1 corrupted by cross-coupling from antenna 2 as

$$v'_1 = v_1 + \epsilon_{21}v_2, \quad (9)$$

where ϵ_{21} is the cross-coupling of antenna 2’s voltage into antenna 1’s voltage. If the uncorrupted cross-correlation visibility is $V_{12} = v_1v_2^*$, then the visibility corrupted by a reflection from antenna 1 can be written as

$$V'_{12} = v'_1v_2^* = v_1v_2^*(1 + \epsilon_{11}), \quad (10)$$

and the visibility corrupted by cross-coupling can be written as

$$V'_{12} = v'_1v_2^* = v_1v_2^* + \epsilon_{21}v_2v_2^*. \quad (11)$$

A.1. A Noise Source in the Field

A cross-coupling-like signal can be generated by a stable noise source in the field, which will not fringe over time. We can rule this out as the systematic mechanism simply based on

the fact that we observe cross-coupling systematics on short baselines at delays of ≥ 1000 ns: any (unreflected) source situated in the field will have a maximum achievable delay corresponding to the baseline's geometric horizon, which for short baselines is from 50 to 150 ns.

A.2. Mutual Coupling Boosted by Cable Reflections

One way to get cross-coupling at high delays is to take cross-coupling at low delays (e.g., mutual coupling) and boost it to high delays via a cable reflection. If antenna 1 observes cross-coupling from antenna 2 that then travels down and gets reflected in the cables of antenna 1, we can write the final measured visibility as

$$\begin{aligned} V'_{12} &= (v_1 + \epsilon_{21}v_2)(1 + \epsilon_{11})v_2^* \\ &= v_1v_2^* + v_1v_2^*\epsilon_{11} + \epsilon_{21}v_2v_2^* + \epsilon_{21}v_2v_2^*\epsilon_{11}. \end{aligned} \quad (12)$$

On the right-hand side of Equation (12), we recognize the first term as the uncorrupted visibility, the second term as the cable-reflected visibility, the third term as the first-order cross-coupling systematic at low delay, and the last term as the cross-coupling systematic boosted to high delay. What we find is that the systematic can only be boosted to specific delays, determined by the product $\epsilon_{21}\epsilon_{11}$. What we see in the data (specifically, the right side of Figure 10) are cross-coupling systematics at delays that are not consistent with this expectation. Furthermore, the high-delay systematics do not look like a shifted version of the low-delay systematics, which is also a prediction of this model.

A.3. A Broadcasting Antenna

This model is a hybrid of the first two models and states that a single antenna, say, antenna 3, receives sky signal that traverses down its signal chain, is reflected back up one of its cables, and is rebroadcast out into the field and then picked up by neighboring antennas, mimicking a stable noise source in the field that has acquired a large delay lag due to the cable reflection in antenna 3's signal chain. We can write the visibility between antennas 1 and 2 in the presence of this signal as

$$V'_{12} = (v_1 + \epsilon_{31}\epsilon_{33}v_3)(v_2 + \epsilon_{32}\epsilon_{33}v_3)^* \quad (13)$$

$$= v_1v_2^* + v_1\epsilon_{32}\epsilon_{33}^*v_3^* + \epsilon_{31}\epsilon_{33}v_3v_2^* + \epsilon_{31}\epsilon_{33}v_3\epsilon_{32}^*\epsilon_{33}^*v_3^*. \quad (14)$$

We recognize the first term on the right-hand side as the uncorrupted visibility and the fourth term as a standard cross-coupling term (due to the autocorrelation nature of $v_3v_3^*$) that has had its large delay canceled out due to $\epsilon_{33}\epsilon_{33}^*$ and thus does not appear at high delays. Only the second and third terms will appear at high delays, but we can see that these terms are actually fringing terms because they contain products like $v_1v_3^*$ rather than $v_3v_3^*$ and thus will not appear centered at a fringe rate of 0 MHz, as we observe in the data.

A.4. Summary

While we have tentatively ruled out a few simple physical models, we still cannot point to a single mode that seems to explain the wide variety of behavior observed in the high-delay systematics. What we can say is that the high-delay $f \sim 0$ MHz terms seem to be physically disconnected from the $f \sim 0$ MHz terms at low delays (i.e., at each baseline's geometric horizon). Regardless of its origin, we do know that the high-delay features do not look like an EoR signal and can therefore be

filtered out of the data. Work is currently underway to assess whether these systematics appear in the upgraded HERA Phase II system, and if so, what can be done in the field to mitigate their presence in the field.

Appendix B Software

The analysis presented in this work relies heavily on the Python programming language (<https://www.python.org>) and Python software developed by HERA collaboration members. Here we provide a list of these packages and their version or Git hash: `aipy` [v2.1.12] (<https://github.com/HERA-Team/aipy>), `healvis` [v1.0.0] (<https://github.com/RadioAstronomySoftwareGroup/healvis>; Lanman & Pober 2019), `hera_cal` [v2.0] (https://github.com/HERA-Team/hera_cal), `hera_sim` [v0.0.1] (https://github.com/HERA-Team/hera_sim), `pyuvdata` [v1.3.6] (<https://github.com/RadioAstronomySoftwareGroup/pyuvdata>; Hazelton et al. 2017), and `uvtools` [v0.1.0] (<https://github.com/HERA-Team/uvtools>). These packages, in turn, rely heavily on other publicly available software packages, including `astropy` [v2.0.14] (<https://astropy.org>; The Astropy Collaboration et al. 2013), `healpy` [v1.12.9] (<https://github.com/healpy/healpy>), `h5py` [v2.8.0] (<https://www.h5py.org/>), `matplotlib` [v2.2.4] (<https://matplotlib.org>), `numpy` [v1.16.2] (<https://www.numpy.org>), `scipy` [v1.2.1] (<https://scipy.org>), and `scikit-learn` [v0.19.2] (<https://scikit-learn.org>).

ORCID iDs

Nicholas S. Kern  <https://orcid.org/0000-0002-8211-1892>
 Joshua S. Dillon  <https://orcid.org/0000-0003-3336-9958>
 Adrian Liu  <https://orcid.org/0000-0001-6876-0928>
 Philip Bull  <https://orcid.org/0000-0001-5668-3101>
 Aaron Ewall-Wice  <https://orcid.org/0000-0002-0086-7363>
 James E. Aguirre  <https://orcid.org/0000-0002-4810-666X>
 Adam P. Beardsley  <https://orcid.org/0000-0001-9428-8233>
 Gianni Bernardi  <https://orcid.org/0000-0002-0916-7443>
 Judd D. Bowman  <https://orcid.org/0000-0002-8475-2036>
 Chris L. Carilli  <https://orcid.org/0000-0001-6647-3861>
 Steve R. Furlanetto  <https://orcid.org/0000-0002-0658-1243>
 Bryna J. Hazelton  <https://orcid.org/0000-0001-7532-645X>
 Daniel C. Jacobs  <https://orcid.org/0000-0002-0917-2269>
 Joshua Kerrigan  <https://orcid.org/0000-0002-1876-272X>
 Piyanat Kittiwisit  <https://orcid.org/0000-0003-0953-313X>
 Saul A. Kohn  <https://orcid.org/0000-0001-6744-5328>
 Matthew Kolopanis  <https://orcid.org/0000-0002-2950-2974>
 Paul La Plante  <https://orcid.org/0000-0002-4693-0102>
 Miguel F. Morales  <https://orcid.org/0000-0001-7694-4030>
 Steven G. Murray  <https://orcid.org/0000-0003-3059-3823>
 Abraham R. Neben  <https://orcid.org/0000-0001-7776-7240>
 Nipanjana Patra  <https://orcid.org/0000-0002-9457-1941>
 Jonathan C. Pober  <https://orcid.org/0000-0002-3492-0433>
 Nithyanandan Thyagarajan  <https://orcid.org/0000-0003-1602-7868>
 Peter K. G. Williams  <https://orcid.org/0000-0003-3734-3587>

References

- Ali, Z., Parsons, A. R., Zheng, H., et al. 2015, *ApJ*, 809, 61
 Barry, N., Wilensky, M., Trott, C. M., et al. 2019, *ApJ*, 884, 1
 Beardsley, A., Hazelton, B. J., Sullivan, I. S., et al. 2016, *ApJ*, 833, 102

- Blackman, R. B., & Tukey, J. W. 1958, *BSTJ*, 37, 185
- Cheng, C., Parsons, A. R., Kolopanis, M., et al. 2018, *ApJ*, 868, 26
- Cui, G., Yu, X., Iommelli, S., & Kong, L. 2016, *ISPL*, 23, 1662
- DeBoer, D. 2015, HERA Phase I Feed Design, Tech. Rep., HERA Memo 11 (Berkeley, CA: Department of Astronomy, Univ. California), <http://reionization.org/wp-content/uploads/2015/01/feedP1.pdf>
- DeBoer, D., Parsons, A. R., Aguirre, J. E., et al. 2017, *PASP*, 129, 045001
- de Oliveira-Costa, A., Tegmark, M., Gaensler, B. M., et al. 2008, *MNRAS*, 388, 247
- Dillon, J., & Parsons, A. 2016, *ApJ*, 826, 181
- Dillon, J., Liu, A., Williams, C. L., et al. 2014, *PhRvD*, 89, 23002
- Eastwood, M. W., Anderson, M. M., Monroe, R. M., et al. 2019, *AJ*, 158, 84
- Ewall-Wice, A., Bradley, R., DeBoer, D., et al. 2016, *ApJ*, 831, 196
- Ewall-Wice, A., Dillon, J. S., Hewitt, J. N., et al. 2016b, *MNRAS*, 460, 4320
- Ewall-Wice, A., Hewitt, J., Mesinger, A., et al. 2016a, *MNRAS*, 458, 2710
- Fagnoni, N., & De Lera Acedo, E. 2016, An Updated Electromagnetic Simulation of the HERA Antenna with CST and Comparison with Measurements, Tech. Rep., HERA Memo 21 (Cambridge: Department of Physics, Univ. Cambridge), http://reionization.org/wp-content/uploads/2013/03/HERA_memo_21_CST_simulation_of_HERA_and_comparison_with_measurements.pdf
- Fagnoni, N., de Lera Acedo, E., DeBoer, D. R., et al. 2019, arXiv:1908.02383
- Furlanetto, S., Oh, S., & Briggs, F. 2006, *PhR*, 433, 181
- Gehlot, B. K., Mertens, F. G., Koopmans, L. V. E., et al. 2019, *MNRAS*, 488, 4271
- Greig, B., & Mesinger, A. 2015, *MNRAS*, 449, 4246
- Greig, B., & Mesinger, A. 2018, in *IAU Symp. 333, Peering towards Cosmic Dawn* (Cambridge: Cambridge Univ. Press), 18
- Greig, B., Mesinger, A., & Koopmans, L. 2015, arXiv:1509.03312
- Grobler, T. L., Nunhokee, C. D., Smirnov, O. M., van Zyl, A. J., & de Bruyn, A. G. 2014, *MNRAS*, 439, 4030
- Hamaker, J. P., & Bregman, J. D. 1996, *A&AS*, 117, 161
- Hazelton, B. J., Jacobs, D. C., Pober, J. C., & Beardsley, A. P. 2017, *JOSS*, 2, 140
- Hickish, J., Abdurashidova, Z., Ali, Z., et al. 2016, *JAI*, 5, 1641001
- Högbom, J. A. 1974, *A&AS*, 15, 417
- Jacobs, D., Pober, J. C., Parsons, A. R., et al. 2015, *ApJ*, 801, 51
- Kern, N. S., Dillon, J. S., Parsons, A. R., et al. 2019b, arXiv:1910.12943
- Kern, N. S., Liu, A., Parsons, A. R., Mesinger, A., & Greig, B. 2017, *ApJ*, 848, 23
- Kern, N. S., Parsons, A. R., Dillon, J. S., et al. 2019a, *ApJ*, 884, 105
- Kerrigan, J., La Plante, P., Kohn, S., et al. 2019, *MNRAS*, 488, 2605
- Kohn, S. A., Aguirre, J. E., La Plante, P., et al. 2019, *ApJ*, 882, 58
- Kolopanis, M., Jacobs, D. C., Cheng, C., et al. 2019, *ApJ*, 883, 133
- Lanman, A. E., & Pober, J. C. 2019, *MNRAS*, 487, 5840
- Lanman, A. E., Pober, J. C., Kern, N. S., et al. 2019, arXiv:1910.10573
- Liu, A., & Parsons, A. 2016, *MNRAS*, 457, 1864
- Liu, A., Parsons, A., & Trott, C. 2014, *PhRvD*, 90, 23018
- Loeb, A., & Furlanetto, S. 2013, *The First Galaxies in the Universe* (Princeton, NJ: Princeton Univ. Press)
- Mellema, G., Koopmans, L. V. E., Abdalla, F. A., et al. 2013, *ExA*, 36, 235
- Mesinger, A. 2016, *Understanding the Epoch of Cosmic Reionization* (Cham: Springer), 423
- Morales, M., Hazelton, B., Sullivan, I., & Beardsley, A. 2012, *ApJ*, 752, 137
- Morales, M. F., Beardsley, A., Pober, J., et al. 2019, *MNRAS*, 483, 2207
- Mouri Sardarabadi, A., & Koopmans, L. V. E. 2019, *MNRAS*, 483, 5480
- Neben, A. R., Bradley, R. F., Hewitt, J. N., et al. 2016, *ApJ*, 826, 199
- Offringa, A. R., Mertens, F., & Koopmans, L. V. E. 2019, *MNRAS*, 484, 2866
- Paciga, G., Albert, J. G., Bandura, K., et al. 2013, *MNRAS*, 433, 639
- Parsons, A., Backer, D., Siemion, A., et al. 2008, *PASP*, 120, 1207
- Parsons, A., Liu, A., Aguirre, J. E., et al. 2014, *ApJ*, 788, 106
- Parsons, A., Pober, J., Aguirre, J., et al. 2012a, *ApJ*, 756, 165
- Parsons, A., Pober, J., McQuinn, M., Jacobs, D., & Aguirre, J. 2012b, *ApJ*, 753, 81
- Parsons, A. R., & Backer, D. C. 2009, *AJ*, 138, 219
- Parsons, A. R., Backer, D. C., Foster, G. S., et al. 2010, *AJ*, 139, 1468
- Patil, A., Yatawatta, S., Koopmans, L. V. E., et al. 2017, *ApJ*, 838, 65
- Patra, N., Parsons, A. R., DeBoer, D. R., et al. 2018, *ExA*, 45, 177
- Planck Collaboration et al. 2016, *A&A*, 594, A13
- Pober, J., Parsons, A. R., Aguirre, J. E., et al. 2013a, *ApJL*, 768, L36
- Pober, J., Parsons, A. R., DeBoer, D. R., et al. 2013b, *AJ*, 145, 65
- Pober, J., Liu, A., Dillon, J. S., et al. 2014, *ApJ*, 782, 66
- The Astropy Collaboration, A., et al. 2013, *A&A*, 558, A33
- Thompson, A. R., Moran, J. M., & Swenson, G. W., Jr. 2017, *Interferometry and Synthesis in Radio Astronomy* (3rd ed.; Cham: Springer)
- Thyagarajan, N., Jacobs, D. C., Bowman, J. D., et al. 2015, *ApJ*, 804, 14
- Thyagarajan, N., Parsons, A. R., DeBoer, D. R., et al. 2016, *ApJ*, 825, 9
- Wells, W. T., Anderson, R. L., & Cell, J. W. 1962, *Ann. Math. Stat.*, 33, 1016