Automated Detection of Antenna Malfunctions in Large-N Interferometers: A Case Study with the Hydrogen Epoch of Reionization Array

Dara Storer[1], Joshua S Dillon[2], Daniel Jacobs[3], Miguel Morales[1], Bryna J Hazelton[1], Aaron Ewall-Wice[4], Zara Abdurashidova[2], James Aguirre[5], Paul Alexander[6], Zaki S Ali[2], Yanga Balfour[7], Adam Beardsley[8], Gianni Bernardi[9], Tashalee S Billings[5], Judd Bowman[3], Richard F Bradley[10], Philip Bull[11], Jacob Burba[12], Steven Carey[6], Chris Carilli[13], Carina Cheng[2], David R. DeBoer[4], Eloy de Lera Acedo[6], Matt Dexter[2], Scott Dynes[14], John Ely[6], Nicolas Fagnoni[6], Randall Fritz[7], Steven R Furlanetto[15], Kingsley Gale-Sides[6], Brian Glendenning[13], Deepthi Gorthi[2], Bradley Greig[16], Jasper Grobbelaar[7], Ziyaad Halday[7], Jacqueline Hewitt[17], Jack Hickish[2], Tian Huang[6], Alec Josaitis[6], Austin Julius[7], MacCalvin Kariseb[7], Nicholas S Kern[14], Joshua Kerrigan[12], Piyanat Kittiwisit[18], Saul A Kohn[5], Matthew Kolopanis[3], Adam Lanman[12], Paul La Plante[2], Adrian Liu[19], Anita Loots[7], David MacMahon[2], Lourence Malan[7], Cresshim Malgas[7], Zachary E Martinot[5], Andrei Mesinger[20], Mathakane Molewa[7], Tshegofalang Mosiane[7], Steven G Murray[3], Abraham Richard Neben[21], Bojan Nikolic[6], Chuneeta Devi Nunhokee[2], Aaron Parsons[22], Robert Pascua[19], Nipanjana Patra[2], Samantha Pieterse[7], Jonathan C Pober[12], Nima Razavi-Ghods[6], Daniel Riley[14], James Robnett[13], Kathryn Rosie[7], Mario G Santos[18], Peter Sims[19], Saurabh Singh[19], Craig Smith[7], Jianrong Tan[19], Nithyanandan Thyagarajan[23], Peter K. G. Williams[24], and Haoxuan Zheng[14]

[1]University of Washington
[2]University of California, Berkeley
[3]Arizona State University
[4]University of California
[5]University of Pennsylvania
[6]University of Cambridge
[7]South African Radio Astronomy Observatory
[8]Winona State University
[9]Harvard-Smithsonian Center for Astrophysics
[10]NRAO Technology Center
[11]Queen Mary University
[12]Brown University
[13]National Radio Astronomy Observatory
[14]Massachusetts Institute of Technology
[15]University of California, Los Angeles
[16]University of Melbourne
[17]MIT
[18]University of Western Cape

[19]McGill University
[20]Scuola Normale Superiore
[21]MIT Kavli Institute for Astrophysics and Space Research
[22]U.C. Berkeley
[23]Commonwealth Scientific and Industrial Research Organisation
[24]Center for Astrophysics — Harvard & Smithsonian

November 22, 2022

Abstract

We present a framework for identifying and flagging malfunctioning antennas in large radio interferometers. Using data from 105 antennas in the Hydrogen Epoch of Reionization Array (HERA) as a case study, we outline two distinct categories of metrics designed to detect outliers along known failure modes of the array: cross-correlation metrics, based on all antenna pairs, and auto-correlation metrics, based solely on individual antennas. We define and motivate the statistical framework for all metrics used, and present tailored visualizations that aid us in clearly identifying new and existing systematics. Finally, we provide a detailed algorithm for implementing these metrics as flagging tools on real data sets.

Automated Detection of Antenna Malfunctions in Large-N Interferometers: A Case Study with the Hydrogen Epoch of Reionization Array

Dara Storer[1], Joshua S. Dillon[2,†], Daniel C. Jacobs[3], Miguel F. Morales[1], Bryna J. Hazelton[1,4], Aaron Ewall-Wice[2], Zara Abdurashidova[2], James E. Aguirre[6], Paul Alexander[7], Zaki S. Ali[2], Yanga Balfour[8], Adam P. Beardsley[3,9,†], Gianni Bernardi[10,11,8], Tashalee S. Billings[6], Judd D. Bowman[3], Richard F. Bradley[12], Philip Bull[13,14], Jacob Burba[15], Steven Carey[7], Chris L. Carilli[16], Carina Cheng[2], David R. DeBoer[17], Eloy de Lera Acedo[7], Matt Dexter[17], Scott Dynes[5], John Ely[7], Nicolas Fagnoni[7], Randall Fritz[8], Steven R. Furlanetto[18], Kingsley Gale-Sides[7], Brian Glendenning[16], Deepthi Gorthi[2], Bradley Greig[19], Jasper Grobbelaar[8], Ziyaad Halday[8], Jacqueline N. Hewitt[5], Jack Hickish[17], Tian Huang[7], Alec Josaitis[7], Austin Julius[8], MacCalvin Kariseb[8], Nicholas S. Kern[2,5], Joshua Kerrigan[15], Piyanat Kittiwisit[14], Saul A. Kohn[6], Matthew Kolopanis[3], Adam Lanman[15], Paul La Plante[2,6], Adrian Liu[21], Anita Loots[8], David MacMahon[17], Lourence Malan[8], Cresshim Malgas[8], Zachary E. Martinot[6], Andrei Mesinger[22], Mathakane Molewa[8], Tshegofalang Mosiane[8], Steven G. Murray[3], Abraham R. Neben[5], Bojan Nikolic[7], Chuneeta Devi Nunhokee[2], Aaron R. Parsons[2], Robert Pascua[2,21], Nipanjana Patra[2], Samantha Pieterse[8], Jonathan C. Pober[15], Nima Razavi-Ghods[7], Daniel Riley[5], James Robnett[16], Kathryn Rosie[8], Mario G. Santos[8,14], Peter Sims[21], Saurabh Singh[21], Craig Smith[8], Jianrong Tan[21], Nithyanandan Thyagarajan[23,1‡,24], Peter K. G. Williams[25,26], Haoxuan Zheng[5]

[1]Department of Physics, University of Washington, Seattle, WA
[2]Department of Astronomy, University of California, Berkeley, CA
[†]NSF Astronomy and Astrophysics Postdoctoral Fellow
[3]School of Earth and Space Exploration, Arizona State University, Tempe, AZ
[4]eScience Institute, University of Washington, Seattle, WA
[5]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA
[6]Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA
[7]Cavendish Astrophysics, University of Cambridge, Cambridge, UK
[8]South African Radio Astronomy Observatory, Black River Park, 2 Fir Street, Observatory, Cape Town, 7925, South Africa
[9]Department of Physics, Winona State University, Winona, MN
[10]Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown, 6140, South Africa
[11]INAF-Istituto di Radioastronomia, via Gobetti 101, 40129 Bologna, Italy
[12]National Radio Astronomy Observatory, Charlottesville, VA
[13]Queen Mary University London, London E1 4NS, UK
[14]Department of Physics and Astronomy, University of Western Cape, Cape Town, 7535, South Africa
[15]Department of Physics, Brown University, Providence, RI
[16]National Radio Astronomy Observatory, Socorro, NM
[17]Radio Astronomy Lab, University of California, Berkeley, CA
[18]Department of Physics and Astronomy, University of California, Los Angeles, CA
[19]School of Physics, University of Melbourne, Parkville, VIC 3010, Australia
[21]Department of Physics and McGill Space Institute, McGill University, 3600 University Street, Montreal, QC H3A 2T8, Canada
[22]Scuola Normale Superiore, 56126 Pisa, PI, Italy
[23]National Radio Astronomy Observatory, Socorro, NM 87801, USA
[‡]National Radio Astronomy Observatory Jansky Fellow
[24]CSIRO, Space and Astronomy, P. O. Box 1130, Bentley, WA 6102, Australia
[25]Center for Astrophysics, Harvard & Smithsonian, Cambridge, MA
[26]American Astronomical Society, Washington, DC

Corresponding author: Dara Storer, `dstorer@uw.edu`

Abstract

We present a framework for identifying and flagging malfunctioning antennas in large radio interferometers. Using data from 105 antennas in the Hydrogen Epoch of Reionization Array (HERA) as a case study, we outline two distinct categories of metrics designed to detect outliers along known failure modes of the array: cross-correlation metrics, based on all antenna pairs, and auto-correlation metrics, based solely on individual antennas. We define and motivate the statistical framework for all metrics used, and present tailored visualizations that aid us in clearly identifying new and existing systematics. Finally, we provide a detailed algorithm for implementing these metrics as flagging tools on real data sets.

1 Introduction

The Hydrogen Epoch of Reionization Array (HERA; (DeBoer et al., 2017)) is a many element radio interferometer designed to observe large scale structures during the Epoch of Reionization (EoR) using the redshifted 21 cm signal from neutral hydrogen. Study of the EoR through detection and observation of the 21 cm line will provide critical insights into the formation of the earliest structures of the universe, and help inform understanding of the underlying physics behind galaxy formation and the intergalactic medium (Furlanetto et al., 2006; Morales & Wyithe, 2010; Pritchard & Loeb, 2012). Other interferometric arrays aimed at detecting the 21 cm signal include the the Precision Array for Probing the Epoch of Reionization (PAPER) (Parsons et al., 2010), the Giant Metrewave Radio Telescope (GMRT; Paciga et al. (2011)), the Murchison Widefield Array (MWA; Tingay et al. (2013)), the LOw Frequency ARray (LOFAR; van Haarlem et al. (2013)), and the Canadian Hydrogen Intensity Mapping Experiment (CHIME; Newburgh et al. (2014)), as well as the upcoming Square Kilometer Array (SKA; Mellema et al. (2013)) and the upcoming Hydrogen Intensity and Real-time Analysis eXperiment (HIRAX; Saliwanchik et al. (2021)).

The 21 cm fluctuation signal is very faint; typical models forecast signal amplitudes in the tens of millikelvin, making the signal four to five orders of magnitude fainter than the bright radio foregrounds (Santos et al., 2005; Bernardi et al., 2010). Attempts to measure the power spectrum using radio interferometers must therefore be executed with high sensitivity and precision analysis techniques in order to realistically achieve a detection (Liu & Shaw, 2020). Achieving sufficient sensitivity requires an interferometer with a large number of antennas observing for months, which introduces a high level of complexity to the system. For example, when completed HERA will have 350 individual dishes each with a dual-polarization signal chain including several analog and digital subcomponents. Therefore, the need for high sensitivity and precision results in thousands of interconnected subsystems that must be commissioned by a relatively small number of people, which poses a significant challenge. Additionally, due to the faintness of the signal, low level systematics that might be deemed negligible in other astronomical applications can have the potential to leak into the power spectrum and obscure the 21 cm signal. Therefore, systematics must either be resolved, methodically avoided, or directly removed in order to achieve sufficiently clean data. Some examples of contaminants common in these types of interferometers include adverse primary beam effects (Beardsley et al., 2016a; Ewall-Wice et al., 2016; Fagnoni et al., 2020; Joseph et al., 2019; Chokshi et al., 2021), internal reflections (Ewall-Wice et al., 2016; Beardsley et al., 2016b; Kern et al., 2019; Kern, Parsons, et al., 2020; Kern, Dillon, et al., 2020), radio frequency interference (RFI) (Wilensky et al., 2020; Whitler et al., 2019), and any analog or digital systematics resulting from the specific design and configuration of the array and its component electronics (Benkevitch et al., 2016; de Gasperin et al., 2019; Star, 2020).

In this work we focus on any systematics arising from a malfunction in an individual antenna, component, or subsystem. While there are some systematics we can avoid using clever analysis techniques (see Kern, Parsons, et al. (2020) for example), we manage most systematics by directly removing the affected antennas from the raw data. This requires us to identify and flag any data exhibiting a known malfunction, and develop methodologies for catching new or previously unidentified systematic effects. While the primary goal of flagging data is to produce the cleanest possible data for analysis, it has the added benefit of providing information regarding the scope and character of prevalent issues to the commissioning team, which is essential to our ultimate goal of finding and resolving the source of the problem. The purpose of this work is to outline a framework for identifying and flagging malfunctioning antennas.

While manual inspection of all HERA data would be an effective approach to antenna flagging, for a 105 element array it would involve assessing 22,155 baselines, each of which has 1024 frequency bins and thousands of time integrations. Therefore, the hands-on time involved is neither practical nor reproducible, and so an automated approach is preferred. In this paper we present an automated approach to antenna quality assessment and flagging. Our approach is to design a set of statistical metrics based on common failure modes of the instrument. We also optimize the metrics to use a limited fraction of the data so they are usable in a real time pipeline. We break these metrics into two categories: cross-correlation metrics (per-antenna values calculated using all baselines), and auto-correlation metrics (per-antenna values calculated using only the auto-correlations). For the duration of this paper we define cross-correlations as correlations between two different antennas, and auto-correlations as the correlation of an antenna with itself. These two methods have complementary advantages. The cross-correlation metrics require a larger data volume, but give us insight into the performance of the whole array and all component subsystems, whereas the auto-correlation metrics are optimized to use a small amount of data, and help assess functionality of individual array components. We outline how each of our metrics is designed to catch one or more known failure modes in the smallest amount of data possible and validate that the automation procedure flags these failures effectively. We also use tools such as simulated noise and comparisons with manual flags to aid in validating our procedure. While these metrics were designed based on HERA data, it is important to note that both the approach and the metrics themselves are applicable to any large interferometric array.

HERA is currently observing while under construction. When completed, the array will have 350 dual-polarization antennas. The data used in this paper were collected on September 29, 2020 (JD 2459122) when there were 105 antennas online, shown graphically in Figure 1. Note that this data is from the second phase of the HERA array, which uses Vivaldi feeds rather than dipoles, along with other changes, and differs significantly from the phase one data analyzed in HERA Collaboration et al. (2021). The HERA receivers are distributed throughout the array in nodes which contain modules for post-amplification, filtering, analog to digital conversion, and a frequency Fourier transform. Each node serves up to 12 antennas. Node clocks are synchronized by a White Rabbit timing network (Moreira et al., 2009). Figure 1 illustrates the node architecture overlain with antenna cataloging developed in this paper. These flags were produced using almost ten hours of data from this night. The high fraction of malfunctioning antennas was partly attributable to limited site access due to the COVID-19 pandemic. HERA has no moving parts and performs a drift scan observation of $\sim 10^\circ$ patch around zenith. The portion of the sky observed on JD 2459122 is shown overlaid on the radio sky in Figure 2.

This paper is organized as follows. In Section 2 we outline the two cross-correlation metrics, providing details of their calculation and a demonstration of their utility. We also examine the distribution of the primary cross-correlation metric across the array, and investigate whether systematics are affecting its statistics. In Section 3 we introduce four auto-correlation metrics, explaining their necessity, describing their precise statistical formulation, and giving examples of typical and atypical antennas. Finally, in Section 4 we summarize our methods and results.

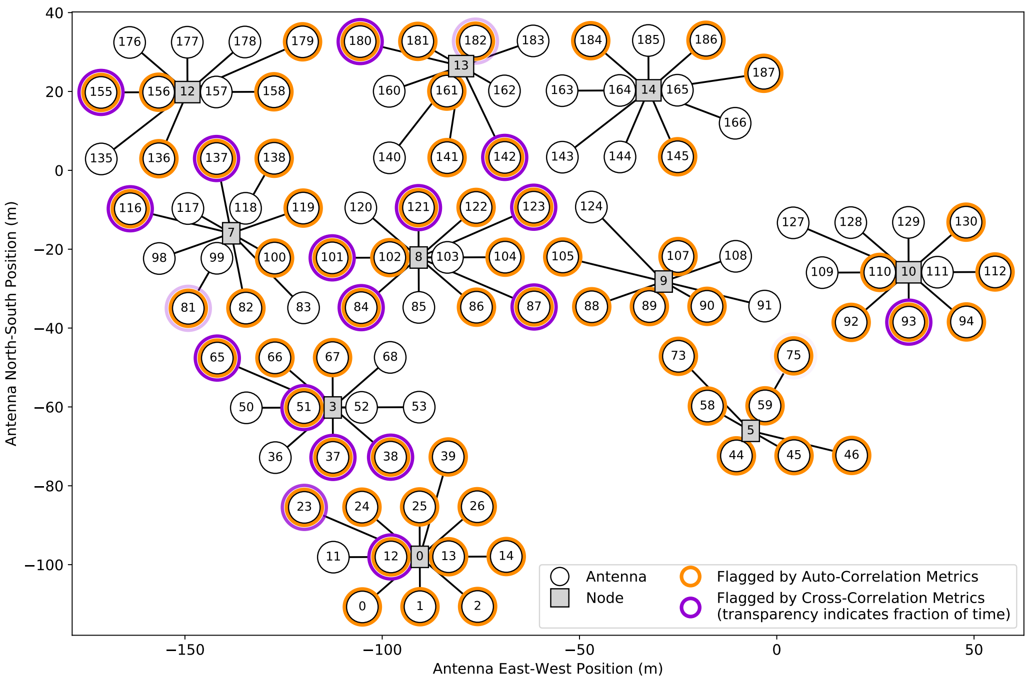

Figure 1. Array layout and antenna quality statuses on Sept 29, 2020 (JD 2459122) as determined by the algorithms laid out in Sections 2 and 3. In HERA, each antenna is connected to a node, which contains amplifiers, digitizers, and the F-engine. Node connections are denoted here by solid black lines. Most of the elements are in the Southwest sector of the split-hexagonal array configuration, with a few in the Northwest and East sectors (Dillon & Parsons, 2016; DeBoer et al., 2017). Only actively instrumented antennas are drawn; many more dishes had been built by this point.

2 Cross-Correlation Metrics

Flagging of misbehaving antennas is necessary in preventing them from impacting calibration, imaging, or power spectrum calculation steps. Here we define a misbehavior to be any feature which makes an antenna unusual when compared to others. In practical terms, the pathologies of antenna malfunction are not limited to the signal chain at the antenna, but could manifest anywhere in the system up to the output of the correlator. Depending on where along the signal chain the pathology lies, we might see evidence of it in the either the auto-correlations, the cross-correlations, or both. For example, if an antenna's timing was out of sync with another's, its auto-correlations might look fine, but its cross-correlations would highlight this systematic. In particular, as an interferometric array grows in size, it is vital to track the health of the entire array, not just the auto-correlations or the cross-correlations in isolation.

In Section 2.1 we define a new cross-correlation metric that is aimed at quantifying how well each antenna is correlating with the rest of the array, and we validate this metric with a simulation. Next, in Section 2.2 we utilize this correlation metric to identify cross-polarized antennas. Finally, in Section 2.3 we outline our specific algorithm

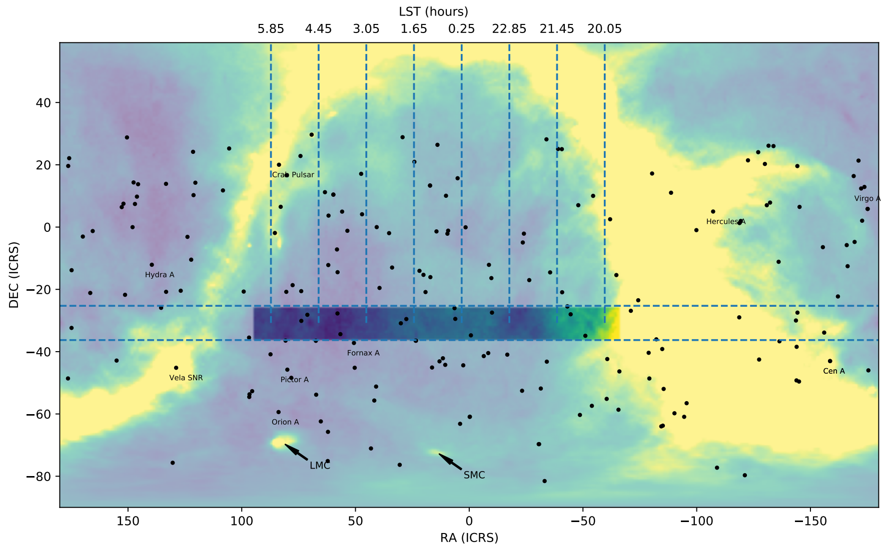

Figure 2. Map of the radio sky (Remazeilles et al., 2015), with the HERA observation band for JD 2459122 shaded, based on a Full Width Half Max of 12 degrees. Individual sources shown are those included in the GLEAM 4Jy catalog (White et al., 2020) with a flux greater than 10.9Jy.

for identifying and removing problematic antennas using the cross-correlation metric framework.

2.1 Identifying Antennas That Do Not Properly Correlate

Our most generalized metric for assessing antenna function tests how well antennas correlate with each other. There are many reasons antennas might not correlate: one of the gain stages might be broken, cables might be hooked up incorrectly, or not phase-aligned with other functional antennas. Assessment of cross-correlations in uncalibrated data is challenging because the correlations can vary widely depending on the baseline length and sky configuration. In particular, one must be able to tell the difference between baselines that include both the expected sky signal and noise versus baselines that include only noise. A metric which is robust against these and other challenges is the normalized and averaged correlation matrix.

$$C_{ij} \equiv \left\langle \frac{V_{ij}^{\text{even}} V_{ij}^{\text{odd}*}}{\left|V_{ij}^{\text{even}}\right| \left|V_{ij}^{\text{odd}}\right|} \right\rangle_{t,\nu} \tag{1}$$

where $\langle\rangle_{t,\nu}$ represents an average over time and frequency, and V_{ij}^{even} and V_{ij}^{odd} are pairs of measurements of the same sky with independent noise. This holds for any correlator outputs separated by timescales short enough that the sky will not rotate appreciably, so that we can assume that time adjacent visibilities are observing the same sky signal but with independent noise realizations.

Division by the visibility amplitude in Equation 1 minimizes the impact of very bright RFI that might differ between even and odd visibilities and dominate the statistics. We experimented with alternative statistics like a maximum and a median to compress across time and frequency but found that with the normalized correlation a simple average was sufficiently robust.

In HERA's case we are able to utilize our specific correlator output to construct even and odd visibilities that are interleaved on a 100 ms timescale. To explain this, we digress briefly into the output of the HERA correlator. In its last stage of operation, antenna voltage spectra are cross-multiplied and accumulated over 100 ms intervals. These visibilities can be averaged over the full 9.6 second integration before being written to disk. However, in order to improve our estimate of noise and to aid in the estimation of power spectra without a thermal noise bias, we split these 96 spectra into two interleaved groups, even and odd, and sum them independently before writing them to disk. Thus, each is essentially 4.8 seconds of integrated sensitivity, spread over 9.6 seconds of observation.

Due to our chosen normalization, the correlation metric measures the phase correlation between visibilities, and is unaffected by overall amplitudes. If the phases are noise-like, the antennas will be uncorrelated and this value will average down to zero. If V_{ij}^{even} and V_{ij}^{odd} are strongly correlated, we expect this statistic to be near one. The normalization in Equation 1 is particularly useful in mitigating the effects of RFI and imperfect power equalization between antennas.

We can visualize the correlation matrix C_{ij} with each baseline pair ij as an individual pixel, such that the auto-correlations fall along the diagonal. A schematic of this visualization is shown in Figure 3. To emphasize any patterns related to electronic connectivity, antennas are organized by their node connection, and within that by their sub-node level electronic connections. Node boundaries are denoted by light blue lines. While the nodal structure used here is specific to HERA, the principal of organizing by electronic connectivity is a generalizable technique for highlighting patterns that may be due to systematics in particular parts of the system. Additionally, plotting the matrices in this way allows us to assess the system health on an array-wide level and on an individual antenna level all in one plot, which is increasingly useful as the size of an array grows.

To study the performance of any single antenna it is useful to form a per-antenna cross-correlation metric C_i by averaging over all baselines that include a given antenna:

$$C_i \equiv \frac{1}{N_{\text{ants}} - 1} \sum_{j \neq i} C_{ij}. \tag{2}$$

We calculate this metric separately for all four instrumental visibility polarizations: NN, EE, EN, NE. The panels below each matrix in Figure 3 show this per-antenna average correlation metric C_{ij}.

Next, Figure 4 shows a visualization of C_{ij} for all four polarizations, using data from a representative subset of the HERA array for simplicity. Here the values have a bimodal distribution (most obvious in the East-East and North-North polarizations), where most antennas are either showing a consistently low metric value, or are close to the array average. This bimodality is also clear in the lower panels showing the per-antenna metric C_i. Here we see more clearly that there is a fairly stable array-level average metric value for each polarization, with a handful of antennas appearing as outliers. The dashed line in the lower panels shows the threshold that is used for antenna flagging, with the points below the threshold marked in red - see Section 2.3 for more on this. There are two primary features to note in Figure 4. First, we see that antennas 51 and 87 are lower than the array average in the North-North and East-East polarizations, but are higher than average in the other two polarizations. Thses points are marked in cyan in the lower panel. The reason for this pathology is that antennas 51 and 87 are cross-polarized, meaning that the cables carrying the East and North polarizations are swapped somewhere along the cable path - this will be discussed further in Section 2.2. Second, we note that there appears to be a slight increase in the average metric power for baselines within the

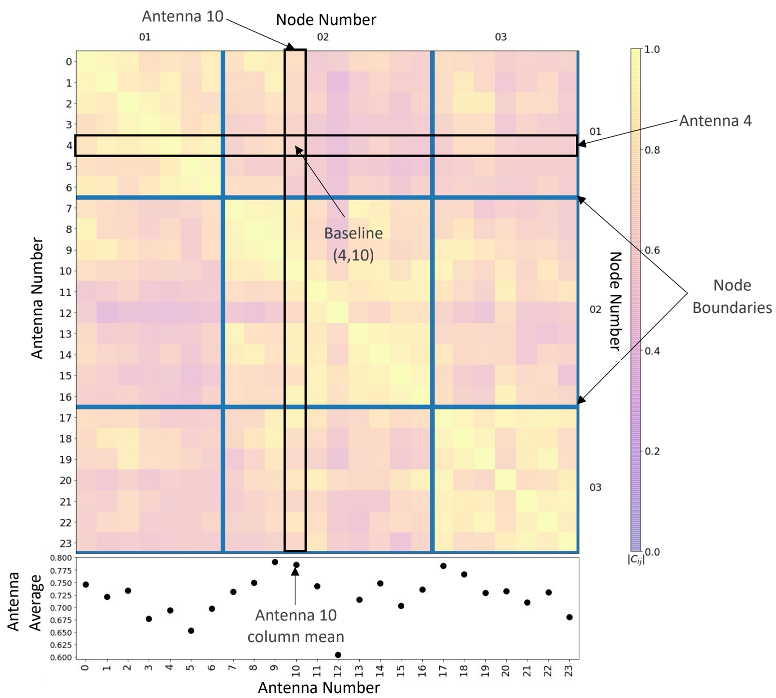

Figure 3. Schematic showing how we visually represent the matrix C_{ij} and the per-antenna metric C_i. Each pixel in the matrix represents an individual baseline ij, identified by the two antennas that pixel corresponds to. The light blue lines denote the node boundaries, and antennas within each node are additionally sorted by their sub-node level electronic connections. The panel below the matrix shows the per-antenna average, calculated as the column mean for each antenna. (Note that in practice this average is computed iteratively - see Section 2.3.)

same node compared over baselines to antennas in different nodes. We explore this effect in the next section.

2.1.1 Understanding the Correlation Metric with Simulations

Figure 4 shows that there is a significant amount of structure in the correlation matrices, specifically related to node connections. Baselines within a node appear to have larger values of C_{ij} than baselines between nodes. We have previously noticed instances of severe node-based structure when there are timing mismatches between nodes due to a failure of the clock distribution system. Figure 5 is an example from an observation when the timing system was known to be broken, and we see clearly that timing mismatches depress the correlation metric. This causes much clearer node structure than the more common structure seen in Figure 4. Therefore, one wonders: are the larger C_{ij} values on the intra-node baselines due to some milder form of this clock distribution issue—perhaps a small error in timing—or is this structure otherwise explicable or even expected?

Put another way, what is the expectation value of C_{ij} as defined in Equation (1)? We can make the assumption that $\langle V_{ij}^{\text{even}}\rangle = \langle V_{ij}^{\text{odd}}\rangle \equiv V_{ij}^{\text{true}}$ and that the two only differ by their noise, n_{ij}, with mean 0 and variance σ_{ij}^2. Ignoring time and frequency dependence, then we can use Equation (1) to first order (ignoring correlations between the numerator and denominator) to find that

$$\langle C_{ij}\rangle = \left\langle \frac{(V_{ij}^{\text{true}} + n_{ij}^{\text{even}})(V_{ij}^{\text{true}} + n_{ij}^{\text{odd}})^*}{\left|V_{ij}^{\text{true}} + n_{ij}^{\text{even}}\right| \left|V_{ij}^{\text{true}} + n_{ij}^{\text{odd}}\right|} \right\rangle \approx \frac{\left|V_{ij}^{\text{true}}\right|^2}{\left|V_{ij}^{\text{true}}\right|^2 + \sigma_{ij}^2}. \tag{3}$$

This approximate expectation value shows us the importance of the signal-to-noise ratio (SNR). At high SNR, $\langle C_{ij}\rangle$ goes to 1, assuming the two even and odd signal terms are actually the same—i.e. that the array is correlating. At low SNR, $\langle C_{ij}\rangle$ goes to 0.

It follows then that the apparent node-based structure in C_{ij} might actually be the impact of the the relationship between SNR and baseline length. Inspecting the array configuration (see Figure 1) we see that baselines within the same node tend to be shorter then baselines involving two nodes. Shorter baselines are dominated by diffuse galactic synchrotron emission, which means that they tend to have a higher signal than longer baselines. Since all baselines have similar noise levels and since higher SNR leads to larger values of C_{ij}, this could account for the effect.

In order to confirm that our node structure is explicable as a baseline length effect rather than some other systematic, we can implement a simple simulation with thermal noise. We calculate V_{ij}^{true} from our data as $(V_{ij}^{\text{even}}+V_{ij}^{\text{odd}})/2$, and take this as a reasonable stand-in for the sky signal, in lieu of a more sophisticated simulation, since it should have approximately the right relative power and should largely average out the instrumental noise. To each visibility V_{ij}^{true} we then add independent Gaussian-distributed thermal noise, with variance given by

$$\sigma_{ij}^2 = \frac{|V_{ii}V_{jj}|}{\Delta t \Delta \nu}, \tag{4}$$

where Δt is the integration time and $\Delta\nu$ is the channel width. This noise is uncorrelated between baselines, times, and frequencies. We then calculate C_{ij}. We compare the C_{ij} with simulated noise to the observed C_{ij} in Figure 6. We can see clearly that the node-based structure we observed in the original correlation matrices is completely reproduced when using a Gaussian noise estimate. This conclusion helps confirm that apparent node-based structure in C_{ij} is is driven by sky feature amplitude, which sets the SNR, rather than systematics.

Finally, in Figure 7 we confirm that our metric distribution is representative of the sky by plotting C_{ij} versus baseline length for all four polarizations. We color each base-

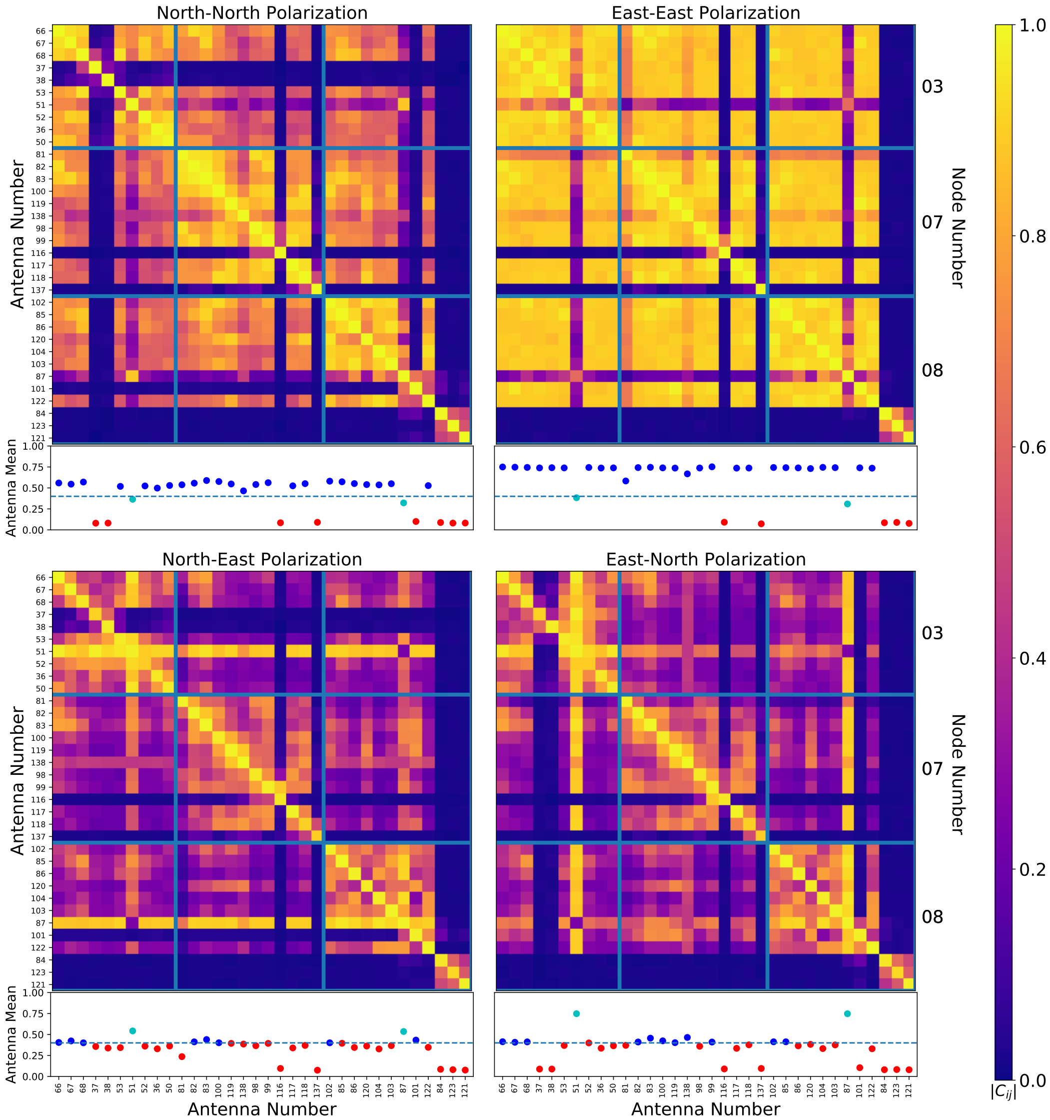

Figure 4. The correlation metric C_{ij} as calculated in Equation 1. Light blue lines denote the boundaries between nodes. The per-antenna average metric C_i as calculated in Equation 2 plotted below each matrix. This average is clearly bimodal and suggests a useful division into good and bad antennas using the blue dashed line as a threshold. Note that the points in cyan, which were were flagged for being cross-polarized, are close to the threshold, but will be flagged using a different metric and therefore do not pose a concern (see Section 2.2 and Figure 8 for more details on this). Also note that we do not use the North-East and East-North polarizations for antenna flagging.

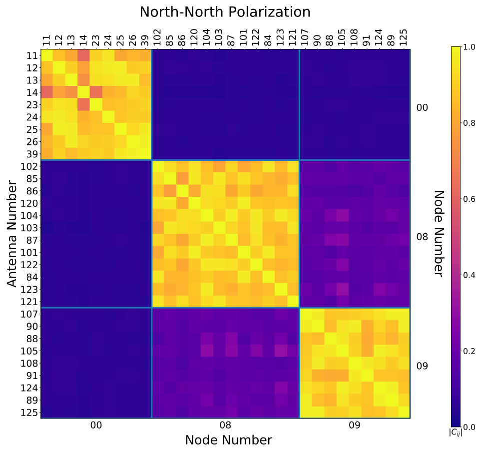

Figure 5. A correlation matrix for a single polarization of HERA phase II data from October 21, 2019 (JD 2458778), taken at a time when the timing system was malfunctioning and antennas between different nodes were not correlating, showing a clear block-diagonal along node lines. This is a sample case where the auto-correlations are nominally acceptable, and investigation of the cross-correlations is necessary to see this type of failure mode.

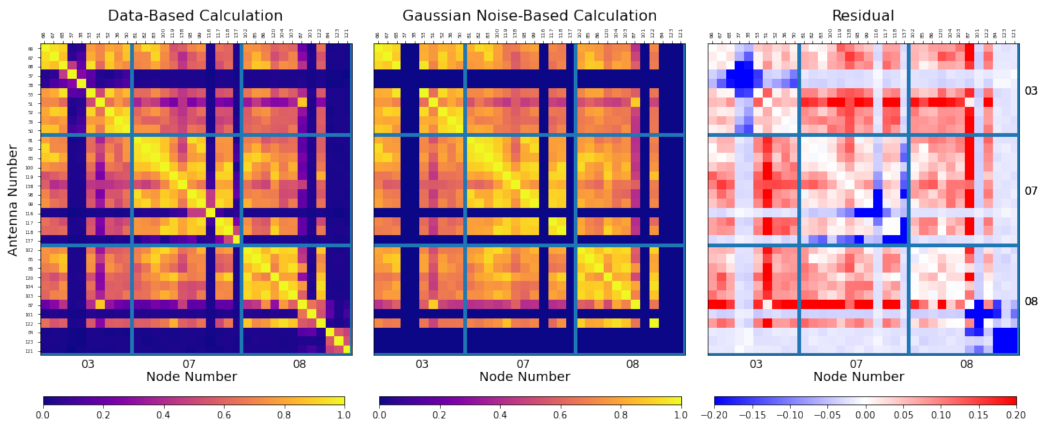

Figure 6. Comparison of the correlation metric computed using true noise from the data (left) and simulated Gaussian thermal noise calculated using the auto-correlations (middle), along with the residual (right). We see here clearly that the node-related structure observed in Figure 8 is fully reproduced using simulated Gaussian noise in lieu of the measured noise used in the original calculation.

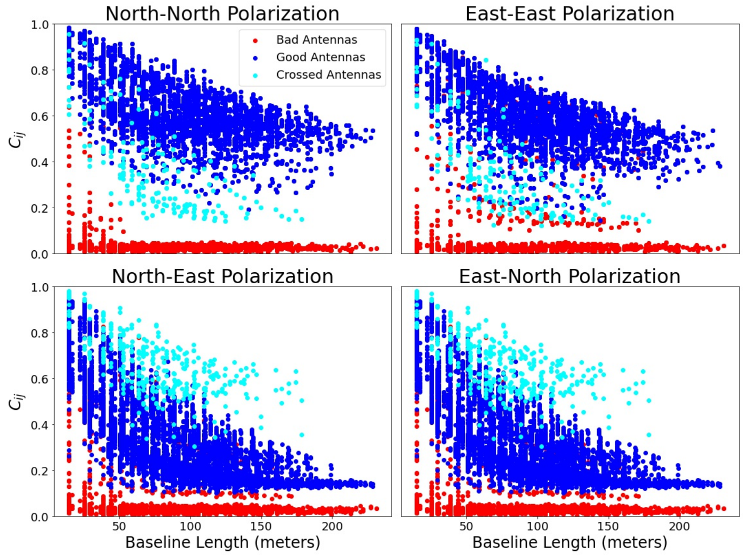

Figure 7. The correlation metric C_{ij} plotted versus baseline length for all four polarizations, with red points representing baselines including at least one antenna that was flagged by this metric, cyan points representing baselines including at least one antenna that was identified as cross-polarized (see section 2.2), and blue points representing baselines where neither constituent antenna is flagged. We observe that nominally well-functioning antennas follow an expected power law shape for galactic emission as a function of baseline length, and we note that cross-polarized antennas are clearly identifiable as having excess power in the North-East and East-North polarizations.

line by whether both constituent antennas were unflagged (blue), at least one was flagged for having a low correlation metric (red), or at least one was flagged for being cross-polarized (cyan). We clearly see the smooth distribution we would expect from sky features, with clearly distinguishable sub-groups by flagging categorization. We would expect a power law slope for galactic emission with strong variation as a function of baseline azimuthal angle, while the point source component should be independent of baseline length or angle, and noise should be similar to point sources (Byrne et al., 2021). Notably, the nominally good antennas generally follow this pattern, with a strong increase towards shorter baselines. Additionally, cross-polarized baselines show a potential transition between galactic domination to point source or noise domination around 100 meters. At frequencies near the middle of the HERA band this corresponds to 1.5 degrees, which is roughly the scale at which point sources are commensurate with galactic emission (Byrne et al., 2021). Given the significant agreement between our measured and expected distributions of C_{ij}, we are confident in our conclusion that the observed structure in Figure 4 is driven by sky features rather than instrumental systematics.

2.2 Identifying Cross-Polarized Antennas

As we have already seen in passing, the correlation metric C_{ij} clearly identifies cross-polarized antennas. Here, cross-polarized means that the physical cables carrying the East and North polarization measurements got swapped in the field. When things are hooked up correctly, we expect to see a stronger correlation between matching polarizations (i.e. EE[1] and NN), and a weaker correlation between different polarizations. Cross polarized antennas have the opposite situation, with stronger correlation in EN and NE.

We identify this situation automatically with a cross-polarization metric formed from the difference between four polarization combinations in the per-antenna correlation metric:

$$C_i^{P_\parallel - P_\times} \equiv \frac{1}{N_{\mathrm{ants}} - 1} \sum_{j \neq i} (C_{ij}^{P_\parallel} - C_{ij}^{P_\times}). \tag{5}$$

where $P_\parallel$ is either the EE or NN polarization, and $P_\times$ is either the NE or EN polarization.

We then calculate our cross-polarization metric as the maximum of the four combinations of same-polarization and opposite-polarization visibilities:

$$R_i = \max\left\{C_i^{NN-NE}, C_i^{NN-EN}, C_i^{EE-NE}, C_i^{EE-EN}\right\} \tag{6}$$

We take the maximum because it's possible to get negative values for some of the $C_i^{P_\parallel - P_\times}$ when one polarization is dead and the other is not. However, when all four values are negative (ie a negative maximum), then the antenna is likely cross-polarized. In Figure 8 we show each of the four differences of C_{ij}. Two antennas, 51 and 87, show negative values in all four combinations, indicating swapped cables. Three other antennas—37, 38, and 101—show up negative in two polarizations, which indicate a single dead polarization, rather than a swap.

2.3 Identifying and Removing Antennas in Practice

Using our correlation metric C_i defined in Equation 2 and our cross-polarization statistic R_i defined in Equation 6 we can implement an iterative algorithm to flag and remove broken and cross-polarized antennas. In Figure 4 we clearly saw that dead antennas have a value of C_i very near zero. As a result, when we calculate C_i for functional antennas by averaging over all constituent baselines, the low correlation between a functional and a dead antenna will decrease the overall value of C_i for the functional antenna. In the case where only a couple of antennas are broken among the whole array this may be tolerable, but it is possible for this bias to cause functional antennas to look much worse than they are, and to potentially drop below the flagging threshold.

To prevent the expected value of our metric from being biased by dead antennas, we implement an iterative metric calculation and flagging approach, outlined in Algorithm 1. First, we calculate C_i for all antennas and identify any that are completely dead (i.e. C_i=0) and remove them. Then, recalculate C_i and R_i for all antennas, identify and remove the worst antenna if it falls below the threshold. We continue with this recalculation and reassessment of the metrics until all remaining antennas are above the threshold in both metrics. Figure 9 shows a comparison between the values of C_i calculated by directly averaging C_{ij} for each antenna versus using the iterative algorithm. We see clearly from this figure that implementing an iterative approach brings our data into a truly bimodal realm where establishing a threshold is straightforward. Based on the observed values, we set an empirical threshold of $C_i = 0.4$, such that any antennas below that value will be flagged and removed. Note that the two antennas marked in cyan are both cross-polarized, so their value near the threshold is not worrisome. As noted in section 2.2, these points are flagged for having a maximum value of R_i below zero. This iterative approach to flagging is robust against varying proportions of broken antennas, which is essential for flagging during the commissioning phase of an array.

[1] HERA dipoles, being fixed, are referred to by their cardinal directions. This avoids much confusion.

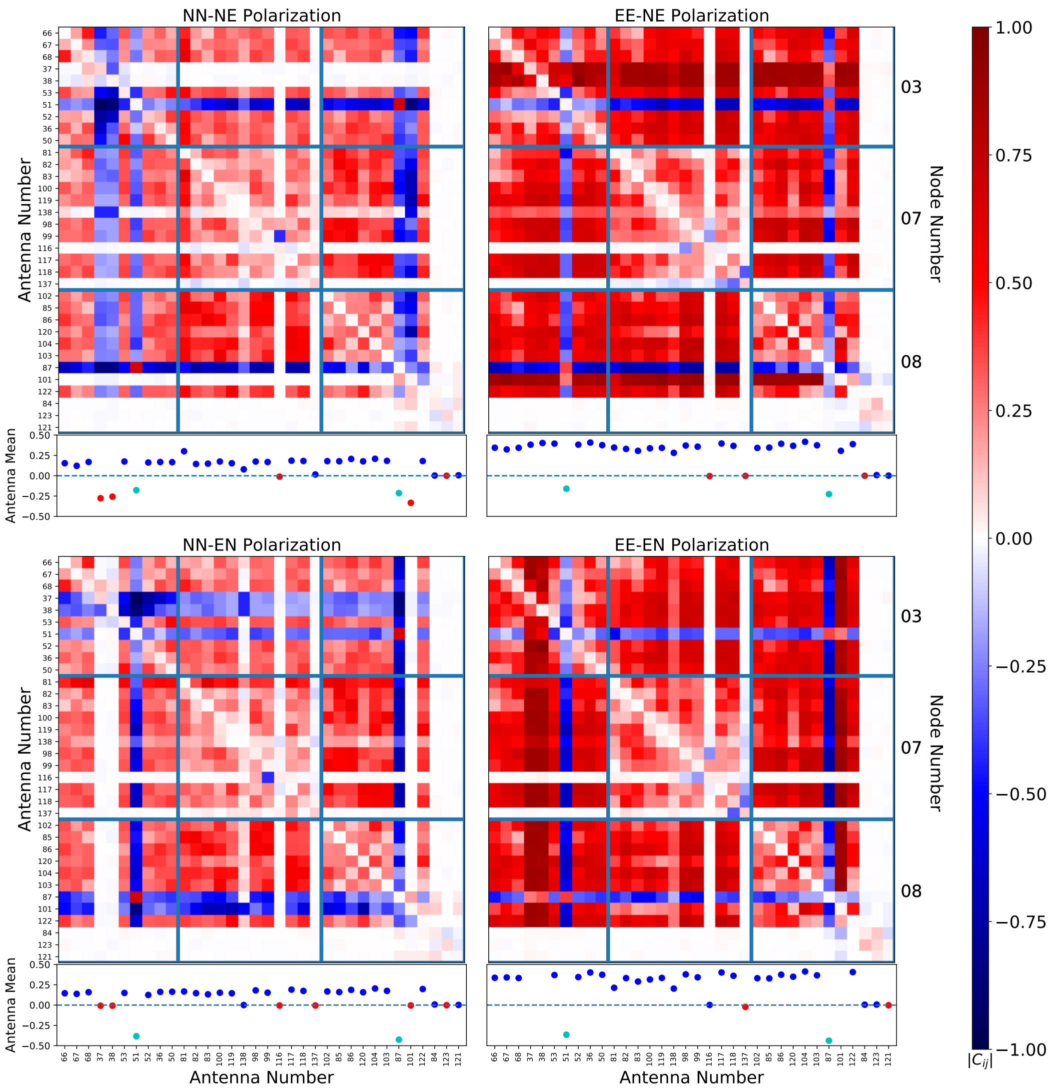

Figure 8. The cross-polarization metric defined in Equation 5. Any antenna with an average metric R_i that is negative in all four polarization combinations is deemed cross-polarized and marked in the lower panels in cyan. Antennas with a positive antenna mean are marked with blue dots, and those with a negative mean are marked with red dots. Here antennas 51 and 87 are cross-polarized. Antennas 37, 38, and 101 are negative in two of the four panels each because they have one dead polarization.

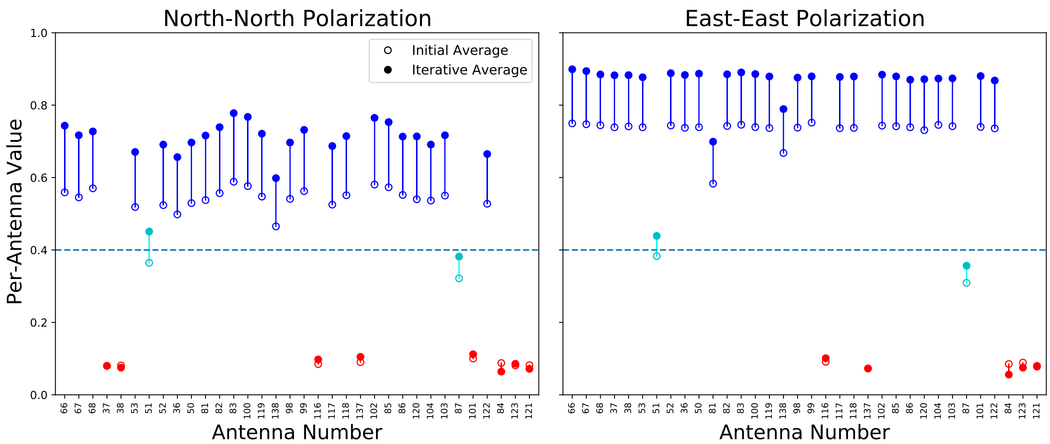

Figure 9. Comparison of the final value of C_i calculated for each antenna using a direct average over C_{ij} versus the iterative calculation outlined in Algorithm 1. We see that using the iterative algorithm helps create a clearer boundary between functional and nonfunctional antennas. Antennas marked in cyan are those that were flagged for being cross-polarized. This plot uses the same representative subset of antennas used in Figures 4 and 6.

3 Auto-Correlation Metrics

While the correlation metrics provide an absolute check on data quality of a particular antenna, not all effects will be caught by this approach. For example, if one antenna has a bandpass structure completely unlike the rest—an effect that might be calibratable—it is useful to identify it and flag it as a symptom of some deeper malfunction in the array. It is useful, therefore, to assess antennas for ways in which they deviate from others, assuming that the plurality of antennas will be well-behaved.[2]

Identification of misbehavior is more difficult with a new system. A newly-built telescope system with novel combinations of technologies means that we lack an a-priori model for how signal chains might malfunction. In early commissioning we observed broadband temporal and spectral instabilities in visibilities which motivated a metric that examines whole nights of data.

We choose to focus on auto-correlations V_{ii} for two reasons. The first is data volume. The number of auto-correlations scales with $N_{\rm ant}$ while the number of visibilities scales with $N_{\rm ant}^2$—far too big to load into memory at once for a whole night of data. Second, because our goal is to identify malfunctioning antennas before calibration, we focus on auto-correlations because they are easier to compare without calibration. Comparison between visibilities measuring the same baseline separation requires at minimum a per-antenna delay calibration to flatten phases. That term in autocorrelations cancels out, leaving each $V_{ii}^{\rm obs} = |g_i|^2 V_{\rm auto}^{\rm true}$. Since most bandpass gains should be similar, auto-correlations can be sensibly compared to one other to look for outliers before calibrating. Even if $|g_i|^2$ differs between antennas, that is something we would like to know and perhaps rectify in the field.

[2] Even when the majority of antennas are malfunctioning, our iterative techniques for outlier detection can still be robust when the malfunctions are multi-causal. To crib from *Anna Karenina*, all happy antennas are alike, but every unhappy antenna is unhappy in its own way.

Historically, auto-correlations from radio interferometers are seldom used. For example, at the VLA the autos are usually discarded (Taylor & Rupen, 1999). The usual reasons given for this are that auto-correlations have a noise bias and that gain variations are assumed to not correlate between antennas. However, given HERA's sensitivity to calibration stability, this assumption is worth re-considering.

Each antenna's auto-correlation stream can be reduced statistically across an entire observation to a single metric spectrum which can then be quickly compared to all other spectra to search for outliers. For HERA, a drift-scan telescope which operates continuously each night for months at a time, one full night's observation time is a useful averaging time range. We focus on four factors motivated by antenna failure modes noted in manual inspection of hundreds of antenna-nights of autocorrelation data: bandpass shape (Section 3.1), overall power level (Section 3.2), temporal variability (Section 3.3), and temporal discontinuities (Section 3.4). The purpose of this section is to develop quantitative metrics that capture these qualitative concerns in a rigorous way, attempting to reduce antenna "badness" along each of these dimensions to a single number. In Section 3.5 we show how these four statistics together produce a useful summary of per-antenna performance (see Figure 15).

Each of these four statistics comes in two flavors. The first is a median-based statistic which is more robust against transient or narrow-band outliers in each time vs. frequency plot or "waterfall", like RFI. The second is a more sensitive mean-based statistic. Our basic approach, outlined in pseudocode in Algorithm 2, is to remove the worst antennas with the robust statistics, then flag RFI, then flag the more subtly bad antennas with the mean-based statistics. In the following sections, we offer a more precise definition of the calculations and the algorithmic application.

3.1 Outliers in Bandpass Shape

Our first metric is designed to identify and flag antennas with discrepant bandpass structures. This often indicates a problem in the analog signal chain. As we mention in Algorithm 2, we first reduce the auto-correlation for antenna i, polarization p to a single spectrum $S(\nu)$ as follows.

$$S^{\text{med}}_{i,p}(\nu) \equiv \frac{\text{med}\left\{V_{ii,pp}(t,\nu)\right\}_t}{\text{med}\left\{V_{ii,pp}(t,\nu)\right\}_{t,\nu}} \tag{7}$$

where $\text{med}\left\{\right\}_t$ indicates a median over time while $\text{med}\left\{\right\}_{t,\nu}$ indicates a median over both time and frequency. This gives us a notion of the average bandpass shape while normalizing the result to remove differences between antennas due to overall power. The reduction from waterfall to spectrum only needs to be computed once per antenna.

We can now compute the median difference between each antenna's spectrum and the median spectrum with the same polarization p according to the following formula:

$$D^{\text{med}}_{i,p} \equiv \text{med}\left\{\left|S^{\text{med}}_{i,p}(\nu) - \text{med}\left\{S^{\text{med}}_{j,p}(\nu)\right\}_j\right|\right\}_\nu, \tag{8}$$

where j indexes over all unflagged antennas. To determine which antenna to flag, if any, we convert each $D^{\text{med}}_{i,p}$ into a modified z-score by comparing it to the overall distribution of distances. These modified z-scores are defined as

$$\begin{aligned} z^{\text{mod}}_{i,p} &\equiv \frac{\sqrt{2}\text{erf}^{-1}(0.5)\left(D_{i,p} - \text{med}\left\{D_{j,p}\right\}_j\right)}{\text{MAD}\left\{D_{j,p}\right\}_j} \\ &\approx 0.67449\left(\frac{D_{i,p} - \text{med}\left\{D_{j,p}\right\}_j}{\text{MAD}\left\{D_{j,p}\right\}_j}\right), \end{aligned} \tag{9}$$

where $\text{MAD}\left\{\right\}_j$ is the median absolute deviation over antennas and $\text{erf}^{-1}(x)$ is the inverse error function. The factor of $\sqrt{2}\text{erf}^{-1}(0.5)$ normalizes the modified z-score so that the expectation value of a z^{mod} of a sample drawn from a Gaussian distribution is the same as its standard z-score.[3]

Having computed modified z-scores for every antenna and every polarization, we iteratively remove the antenna with the worst modified z over all metrics and both polarizations. When one polarization is flagged, we flag the whole antenna. We then recompute $D^{\mathrm{med}}_{i,p}$ and $z^{\mathrm{mod}}_{i,p}$ and continue flagging antennas until none have a modified z over a chosen threshold, in our case 8.0.

Next we perform a simple RFI flagging, analogous to the algorithm used in HERA Collaboration et al. (2021), but performed on a single auto-correlation waterfall averaged over all remaining antennas. This process includes a search for local outliers after median filtering and then mean filtering, which are flagged as RFI. Finally, a thresholding algorithm is performed that throws out entire channels or entire integrations which are themselves significant outliers after analogous 1D filtering. The results of this process are shown in Figure 10. This process flags 12.6% of the data, excluding band-edges, and leaves 11.3% of channels and 1.0% of all times completely flagged. This is likely an under-count of RFI; the algorithm is to designed to flag the most egregious outliers that might skew the statistics described below, rather than to find and remove RFI for the purpose of making sensitive 21 cm power spectrum measurements.

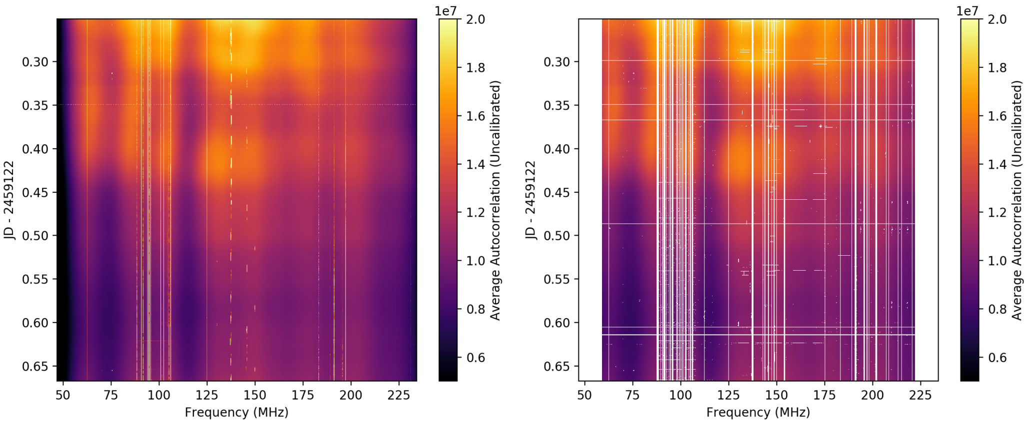

Figure 10. Auto-correlation averaged over good antennas, before and after RFI flags. RFI is excised using local median and mean filters to search for outliers, followed by 1D thresholding. This is a simplified version of the algorithm used in HERA Collaboration et al. (2021) with the exception that it is sped up by performing it on a single waterfall averaged over unflagged antennas.

After RFI flagging, we next compute shape metric spectra with mean-based statistics. Analogously to Equation 7 this case,

$$S^{\mathrm{mean}}_{i,p}(\nu) \equiv \frac{\langle V_{ii,pp}(t,\nu)\rangle_t}{\langle V_{ii,pp}(t,\nu)\rangle_{t,\nu}}, \tag{10}$$

where $\langle\rangle_t$ indicates a weighted-mean over the time dimension, giving zero weight to times and frequencies flagged for RFI. Likewise, these spectra are reduced to scalar distance

[3] Were the distribution of distance metrics Gaussian (it is generally not), then one could think of modified z-score of 8 as an "8σ outlier." This kind of language is imprecise, but often useful for building intuition.

metrics as

$$D_{i,p}^{\text{mean}} \equiv \left\langle \left| S_{i,p}^{\text{mean}}(\nu) - \left\langle S_{j,p}^{\text{mean}}(\nu) \right\rangle_j \right| \right\rangle_\nu , \tag{11}$$

where again averages are performed over unflagged antennas, times, and frequencies. Just as before, we compute modified z-scores to iteratively flag the worst antenna outlier, recalculating $D_{i,p}^{\text{mean}}$ after each antenna is flagged. This proceeds until no antennas exceed a z-score of 4; half that used during the first round median cut.

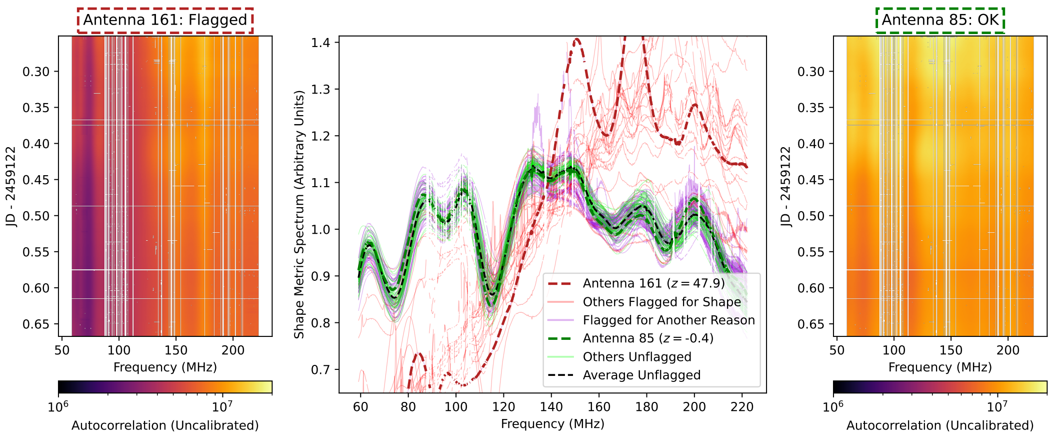

Figure 11. Here we show the shape metric spectra, defined in Equation 10, for all North/South-polarized antennas in the array (center panel). Outliers (red lines) are defined as has having a modified z-score greater than 4.0 in their scalar distance metric (Equation 11) compared compared to the average good antenna (black dashed line) and the distribution of good antennas (light green lines). Note that this figure includes flagging by three other metrics causing some antennas to be flagged even though they look ok here. We highlight two example antennas and show their full auto-correlation waterfalls, one flagged (161; left panel and dark red dashed line) and one functioning normally (85; right panel and dark green dashed line). Antenna 161's bandpass is notably low at low frequency, making it a clear outlier in the distribution of bandpasses in the center panel.

In Figure 11 we show the the results of this operation with example waterfalls and metric spectra for antennas that were and were not flagged by our modified z-score cut of 4.0. In general, we find that the metric robustly identifies antennas with metric spectra discrepant from the main group of antennas. Almost everything in red in Figure 11 is a pretty clear outlier. Where exactly to draw the line is tricky, and likely requires some manual inspection of metric spectra and waterfalls for antennas near the cutoff. Note that this figure includes flagging by all four metrics. Some moderate outliers in shape were not flagged for shape but were flagged for other reasons, indicating that this metric and the other three discussed below are not completely independent.

3.2 Outliers in Bandpass Power

We next turn to looking for outliers in bandpass power. High power might indicate incorrect amplifier settings while a signal chain malfunction might cause anomalously low power. Our approach for finding outliers is very similar to the one for finding outliers in bandpass shape in power laid out in Section 3.1. Here we lay out the mathematical approach, highlighting and motivating differences between the two.

Once again, we begin by defining median-based metric spectra which collapse each antenna's waterfall down to a single number per frequency. For bandpass power, that is simply

$$S_{i,p}^{\text{med}}(\nu) \equiv \text{med}\left\{V_{ii,pp}(t,\nu)\right\}_t \,. \tag{12}$$

This is simply an unnormalized version of Equation 7. However, instead of directly comparing each antenna's spectrum with the median spectrum, we instead compare their logarithms:

$$D_{i,p}^{\text{med}} \equiv \text{med}\Big\{\Big|\log\left(S_{i,p}^{\text{med}}(\nu)\right) - \log\Big(\text{med}\left\{S_{j,p}^{\text{med}}(\nu)\right\}_j\Big)\Big|\Big\}_\nu, \tag{13}$$

This logarithmic distance measure reflects the fact that gains are multiplicative and that the optimal ranges for amplifier and digitization are themselves defined in decibels. We take the absolute value of the difference of the logs because we want to penalize both antennas with too little power, which may indicate a malfunction, and antennas with too much power, which may cause a nonlinear response to the sky signal.

After RFI flagging as described in the previous section, we next proceed with outlier detection using modified mean-based statistics, which are straightforward adaptations of Equations 12 and 13:

$$S_{i,p}^{\text{mean}}(\nu) \equiv \left\langle V_{ii,pp}(t,\nu)\right\rangle_t \,, \tag{14}$$

$$D_{i,p}^{\text{mean}} \equiv \Big\langle\Big|\log\left(S_{i,p}^{\text{mean}}(\nu)\right) - \log\Big(\left\langle S_{j,p}^{\text{mean}}(\nu)\right\rangle_j\Big)\Big|\Big\rangle_\nu \,. \tag{15}$$

Once again, as we can see in Figure 12, this metric picks a number of antennas that are clearly behaving differently than the main group. As we saw in the previous section

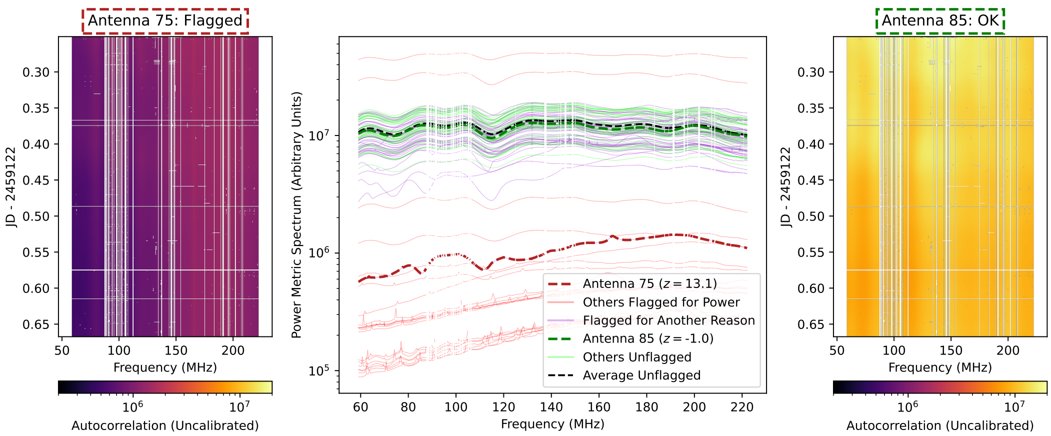

Figure 12. Here we show bandpass power metric spectra, defined in Equation 14, for all North/South-polarized antennas in the array (center panel). Just as in Figure 11 we show flagged and unflagged antennas, highlighting example auto-correlation waterfalls of good (85; right panel) and bad (75; left panel) antennas, as defined by the modified z-score of their distance metric (Equation 15). While antenna 75's bandpass structure is similar to the normal antennas, its autocorrelation has roughly an order of magnitude less power. This makes us suspicious that the amplifiers in the signal chain are not operating properly.

we see there are some antenna which appear to be "in family" according to this metric but are flagged for other reasons. But now we can start to see why this might be. A few of the flagged antennas appear to be fine according to their bandpass shape but are significantly lower or higher in power than the rest.

3.3 Outliers in Temporal Variability

We now turn to the the question of searching for outliers in the *temporal structure* of the antenna response. While the metrics follow a similar pattern—median-based spectra and distances, followed by mean-based spectra and distances—they are mathematically quite different from those in Sections 3.1 and 3.2.

During observing and subsequent inspection analysis sharp discontinuities were observed in the auto-correlations. Often, though not always, these are rapid changes occurring within a single integration. Sometimes they are accompanied with apparent changes in the bandpass shape or power. Sometimes the effects are relatively localized in frequency; sometimes they are broadband. Sometimes they are frequent jumps; sometimes there are just a handful of discontinuities followed by minutes or hours of stability. Developing a physical understanding of the origin of these effects is an ongoing research effort outside the scope of this paper. Absent that understanding—and a hardware fix to prevent the effects—we have to consider this behavior suspicious and therefore meriting flagging.

Here and in Section 3.4 we present two metrics for automatically identifying temporal effects. In general, we are looking for forms of temporal structure of the auto-correlations that cannot be explained by the sky transiting overhead. The first looks for high levels of temporal variability throughout the night. To distinguish temporal variability due to sky-rotation from anomalous temporal structure, our metrics are based on a comparison of each antenna's auto-correlation waterfall with an average waterfall over all antennas. For our first round of median statistics, we use the median absolute deviation of the waterfall along the time axis after dividing out the median waterfall over antennas. Thus,

$$S_{i,p}^{\mathrm{med}}(\nu) \equiv \mathrm{MAD}\left\{ \frac{V_{ii,pp}(t,\nu)}{\mathrm{med}\left\{V_{jj,pp}(t,\nu)\right\}_j} \right\}_t . \tag{16}$$

to produce a single spectrum for each antenna that can be reasonably interpreted as the standard deviation over time of each channel with respect to the mean over time.

The significance of each spectrum we estimate as the median over frequency of the extent to which any antenna's temporal variability metric spectrum exceeds the median metric spectrum over all antennas:

$$D_{i,p}^{\mathrm{med}} \equiv \mathrm{med}\left\{ S_{i,p}^{\mathrm{med}}(\nu) - \mathrm{med}\left\{S_{j,p}^{\mathrm{med}}(\nu)\right\}_j \right\}_\nu . \tag{17}$$

Note that we do not take the absolute value of the difference; while shape and power mismatches are penalized both for being too low and for being too high, we do not penalize antennas for varying less that the median. These simply become negative z-scores –indicating that an antenna has less temporal variation than the median signal– and do not result in flags.

Our mean-based metrics are a straightforward adaptation of Equations 16 and 17:

$$S_{i,p}^{\mathrm{mean}}(\nu) \equiv \left[\left\langle \left(\frac{V_{ii,pp}(t,\nu)}{\langle V_{jj,pp}(t,\nu)\rangle_j} \right)^2 \right\rangle_t - \left\langle \frac{V_{ii,pp}(t,\nu)}{\langle V_{jj,pp}(t,\nu)\rangle_j} \right\rangle_t^2 \right]^{1/2}, \tag{18}$$

$$D_{i,p}^{\mathrm{mean}} \equiv \left\langle S_{i,p}^{\mathrm{mean}}(\nu) - \left\langle S_{j,p}^{\mathrm{mean}}(\nu)\right\rangle_j \right\rangle_\nu . \tag{19}$$

In theory, the denominator of Equations 16 and 18 should change each time an antenna is thrown out and the distance measures and modified z-scores are recomputed. This can be computationally expensive when a large fraction of the array needs flagging, as has sometimes been the case during HERA commissioning. In practice, we take a shortcut. During the median-statistics round, we simply neglect this effect, relying on the fact that the median statistics are relatively insensitive to the set of antennas that are flagged. During the next round using mean-based statistics, we iteratively remove antennas until no antennas remain above our modified z-score cut. Only then do we recompute the metric spectra in Equation 18. In general, this has the effect of making the metric spectra more sensitive to temporal variability, since the mean spectrum will include fewer anomalously variable antennas. The standard procedure of removing antennas and recalculating each $D_{i,p}^{\text{mean}}$ (but not each $S_{i,p}^{\text{mean}}(\nu)$) is repeated. This loop continues until no more more antennas are flagged after recalculating $S_{i,p}^{\text{mean}}(\nu)$ one final time.

In Figure 13 we show the the resulting mean-based metric spectra after iteratively removing outliers. While there are some very clear outliers that are successfully iden-

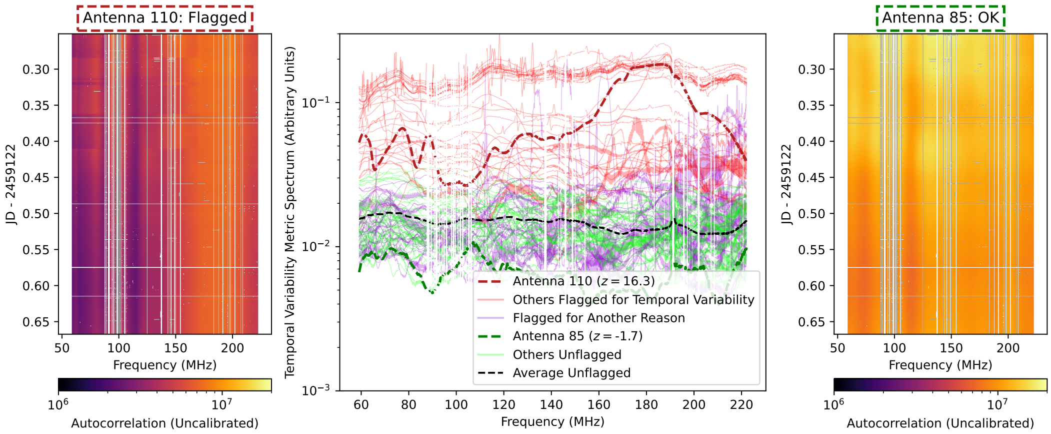

Figure 13. Here we show temporal variability metric spectra, defined in Equation 18, for all North/South-polarized antennas in the array (center panel). Just as in Figure 11 we show flagged and unflagged antennas, highlighting example auto-correlation waterfalls of good (85; right panel) and bad (110; left panel) antennas, as defined by the modified z-score of their distance metric (Equation 19). The malfunction in antenna 110—chunks of time where the waterfall shape and amplitude varies discontinuously—is subtle. It is easiest to see in the waterfall at low frequencies during the first half of the night. These sorts of effects are often more visible in metric spectra and in renormalized waterfalls, as demonstrated in Figure 15.

tified, the precise line between what should be considered good and what should be considered bad is ambiguous. Clearly the pathology seen in Antenna 110 is worthy of flagging and the metric successfully identifies it as having high variability relative to the average waterfall. Likewise, most of what is identified as good appears to be behaving like most of the other antennas. Just as with the previous metrics, some level of inspection of antennas near the cutoff is warranted.

3.4 Outliers in Temporal Discontinuities

Though a range of temporal variation pathologies were noted during the observing and data inspection phase one that stood out was abrupt changes occurring faster than the integration time and lasting minutes to hours. Our second metric for anomalous temporal structure looks for such sharp discontinuities, which also cannot be explained by sky rotation. As with our metric for overall temporal variability (see Section 3.3), our metric is based on examining each antenna's waterfall after dividing out the average waterfall of unflagged antennas. Instead of using the median absolute deviation or the standard deviation, which are measures of variability on any timescale, we instead want to detect variability on the shortest timescale—which is the hardest to explain with antenna-to-antenna primary beam variations (Dillon et al., 2020).

Beginning with the auto-correlation scaled by the median over antennas, we compute the discrete difference along the time axis, and then collapse that waterfall (which is only one integration shorter than the original) along the time axis to a metric spectrum. In our first round of flagging using median statistics, this becomes:

$$S_{i,p}^{\text{med}}(\nu) \equiv \text{med}\left\{ \left| \frac{V_{ii,pp}(t+\Delta t,\nu)}{\text{med}\left\{V_{jj,pp}(t+\Delta t,\nu)\right\}_j} - \frac{V_{ii,pp}(t,\nu)}{\text{med}\left\{V_{jj,pp}(t,\nu)\right\}_j} \right| \right\}_t, \tag{20}$$

where Δt is our integration time (9.6 s in this data set). Our distance measure, designed to penalize only excessive levels of temporal discontinuities, is the same as in Equation 17:

$$D_{i,p}^{\text{med}} \equiv \text{med}\left\{ S_{i,p}^{\text{med}}(\nu) - \text{med}\left\{S_{j,p}^{\text{med}}(\nu)\right\}_j \right\}_\nu . \tag{21}$$

The adaption to mean-based statistics is straightforward:

$$S_{i,p}^{\text{mean}}(\nu) \equiv \left\langle \left| \frac{V_{ii,pp}(t+\Delta t,\nu)}{\left\langle V_{jj,pp}(t+\Delta t,\nu)\right\rangle_j} - \frac{V_{ii,pp}(t,\nu)}{\left\langle V_{jj,pp}(t,\nu)\right\rangle_j} \right| \right\rangle_t, \tag{22}$$

$$D_{i,p}^{\text{mean}} \equiv \left\langle S_{i,p}^{\text{mean}}(\nu) - \left\langle S_{j,p}^{\text{mean}}(\nu)\right\rangle_j \right\rangle_\nu . \tag{23}$$

In Figure 14, we show metric spectra for all antennas for a single polarization and examples of nominal and abnormal waterfalls. Antennas flagged as bad show a wide variety of strange behavior: some show broadband effects, others are more localized. Antenna 89 and one other even shows spectrally oscillatory levels of temporal discontinuities; we currently have no explanation for this effect. Perhaps these features provide further clues to the ongoing system integration and debugging efforts.

The good antennas are fairly tightly clustered around the average, which is spectrally flat. That behavior is expected if the integration-to-integration differences are purely attributable to thermal noise. Normalizing each waterfall by the average good waterfall should cancel out the spectral and temporal dependence of the noise. Given that theoretical expectation this might be the easiest of all the metrics to set an absolute cut, rather than a relative one based on the modified z-score. However, the wide variety of poorly-understood malfunctions combined with the possibility that low-level RFI might still contaminate these metrics complicates that picture.

3.5 Assessing Individual Antenna Quality in Practice

One advantage of the `auto_metrics` framework is that it is straightforwardly applicable to new combinations of metric spectra and distance measures. For example, it should be noted that the anomalous temporal structure metrics in Sections 3.3 and 3.4 are not exhaustive. By averaging over the whole night, they privilege frequent or persistent effects over infrequent ones. For example, a strong jump in the waterfall like we

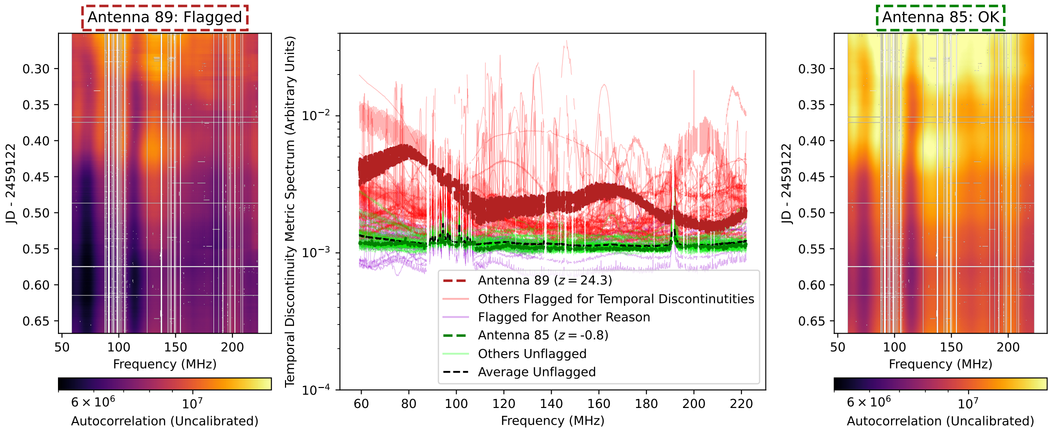

Figure 14. Here we show temporal discontinuity metric spectra, defined in Equation 22, for all North/South-polarized antennas in the array (center panel). Just as in Figure 11 we show flagged and unflagged antennas, highlighting example auto-correlation waterfalls of good (85; right panel) and bad (89; left panel) antennas, as defined by the modified z-score of their distance metric (Equation 23). The discontinuities are often hard to perceive without a very careful inspection of the waterfall. Once again, these sorts of effects are often more visible in metric spectra and in renormalized waterfalls, as demonstrated in Figure 15.

see in Antenna 110 in Figure 13 that then quickly reverts to "standard" behavior and does not repeat might not be caught by either metric. One could imagine other ways of computing $S(\nu)$ or D that up-weight rare excursions from normality. While we continue to assess antenna malfunctions and develop other metrics, it is worthwhile to continue the visual inspection of auto-correlation waterfalls normalized by the average of nominally good antennas to identify other modalities of malfunction.

In particular, we find it useful to produce a suite of per-antenna visualizations of the different metric spectra and the corresponding auto-correlation waterfalls. In Figure 15 we show three such examples: one clearly malfunctioning (Antenna 0), one nominal (Antenna 85), and one borderline case that we ultimately flagged (Antenna 24). For each, we show their metric spectra compared to all unflagged antennas, along with the z-scores, highlighting which antennas were automatically flagged. These plots synthesize the information about how discrepant each antenna is along the four axes considered here and help clarify why.

In Figure 15 we also show both the waterfalls and the normalized waterfalls, which are divided by the average good waterfall (Figure 10) and then normalized to average to 1. We find it particularly useful to look closely at these normalized waterfalls, especially in borderline cases like Antenna 24. Antenna 24's bandpass shape is sufficiently discrepant with the others to merit an automatic flag, though this does not necessarily mean that it is uncalibratable. More concerning are the abrupt discontinuities at high frequency around 2459122.3 and around 2459122.5. This is precisely the kind of issue we worried about: a strong but rare temporal feature that just barely misses the threshold. Examples like this motivate by-eye inspect of borderline antennas. This is what we have done with recent HERA data. The automatic pipeline produces `jupyter` notebooks with plots like Figure 15 for all antennas, sorting them by the single highest z-score metric. This makes it easy to find the borderline antennas and decide whether to flag them on a case-by-case basis.

4 Summary

Construction and commissioning of HERA is currently under way with hundreds of antennas aiming to reach the extreme dynamic range necessary to detect and characterize the neutral hydrogen signal from the epoch of reionization. Separating that signal from foregrounds four to five orders of magnitude stronger requires both large volumes of data and the swift and reliable identification of malfunctions that adversely affect data quality. In this work, we report on new metrics which sensitively detect various pathologies reported anecdotally during manual data inspection and reliably classify them. In some cases, the precise underlying mechanism (e.g. an antenna with swapped cables for its two polarizations) is known. In others, a physical explanation requires lab and field tests that are beyond the scope of this paper. Armed with per-antenna classifications the instrument teams can more effectively triage issues according to their occurrence rate. By inspecting the nightly analysis and dashboard reports the team can quickly assess the impact of hardware changes. Meanwhile, the definition of metric *spectra* provides a physically meaningful signature which can be exploited by instrument engineers to identify characteristics like reflections, clipping, interference, and more.

The definition of metrics which isolate features of interest and standard ways of displaying them routinely is crucial to managing a large array with a small team. As digital and analog systems grow in capability, arrays will continue to grow in antenna count. Arrays like OVRO-LWA-III, DSA-2000 (Hallinan et al., 2021), HIRAX (Saliwanchik et al., 2021), CHORD (Vanderlinde et al., 2019), PUMA (Castorina et al., 2020), SKA-Low (Mellema et al., 2013) and more will use hundreds to thousands of elements. Ultimately the maintenance time per-antenna imposes a significant design pressure on large arrays. This kind of pressure can also affect arrays with fewer antennas but with more elaborate receivers or wider geographic distributions. A prime example of this regime is the proposed ngVLA. With 244 antennas distributed across New Mexico, Arizona, and Mexico, along with outriggers extending to VLBA sites across north America and six cryogenic receivers, operation will require careful minimization of maintenance time.[4]. Quick identification of subtle systematic errors using semi-automatic systems like we describe here are expected to be essential.

In 21 cm cosmology experiments, the reliability and precision of arrays will continue to be the dominant factor affecting sensitivity. Identifying, flagging, and ultimately fixing subtle instrument issues will continue to be the first line of defense. Further work in this area is needed, for example, using simulations to replace the detection of relative outliers with absolute thresholds or to replace iterative flagging with a single analysis step. That said, a system like the one presented here will be necessary for triaging malfunctions and extracting science-quality data to form the basis for future cosmology results.

Acknowledgments
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 by the Gordon and Betty Moore Foundation through grant GBMF5215 to the Massachusetts Institute of Technology. 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. JSD gratefully acknowledges the support of the NSF AAPF award #1701536. A. Liu acknowledges support from the New Frontiers in Research Fund Exploration grant program, the Canadian Institute for Advanced Research (CIFAR) Azrieli Global Scholars program, a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant

[4] See ngVLA memo 020.10.05.00.00-0002-PLA, S6.2 at `https://ngvla.nrao.edu/page/projdoc`

and a Discovery Launch Supplement, the Sloan Research Fellowship, and the William Dawson Scholarship at McGill. D. Storer acknowledges that this material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1762114. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. MGS and PK acknowledge support from the South African Radio Astronomy Observatory (SARAO) and National Research Foundation (Grant No. 84156). G.B. acknowledges 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 Number 113121) as part of the ISARP RADIOSKY2020 Joint Research Scheme. Data used in this paper is not yet publicly available.

References

Beardsley, A. P., Hazelton, B. J., Sullivan, I. S., Carroll, P., Barry, N., Rahimi, M., ... Wyithe, J. S. B. (2016a, dec). FIRST SEASON MWA EOR POWER SPECTRUM RESULTS AT REDSHIFT 7. *The Astrophysical Journal*, *833*(1), 102. Retrieved from `https://doi.org/10.3847/1538-4357/833/1/102` doi: 10.3847/1538-4357/833/1/102

Beardsley, A. P., Hazelton, B. J., Sullivan, I. S., Carroll, P., Barry, N., Rahimi, M., ... et al. (2016b, Dec). First season mwa eor power spectrum results at redshift 7. *The Astrophysical Journal*, *833*(1), 102. Retrieved from `http://dx.doi.org/10.3847/1538-4357/833/1/102` doi: 10.3847/1538-4357/833/1/102

Benkevitch, L. V., Rogers, A. E. E., Lonsdale, C. J., Cappallo, R. J., Oberoi, D., Erickson, P. J., & Baker, K. A. V. (2016). *Van vleck correction generalization for complex correlators with multilevel quantization.*

Bernardi, G., de Bruyn, A. G., Harker, G., Brentjens, M. A., Ciardi, B., Jelić, V., ... et al. (2010, Nov). Foregrounds for observations of the cosmological 21cm line. *Astronomy & Astrophysics*, *522*, A67. Retrieved from `http://dx.doi.org/10.1051/0004-6361/200913420` doi: 10.1051/0004-6361/200913420

Byrne, R., Morales, M. F., Hazelton, B., Sullivan, I., Barry, N., Lynch, C., ... Jacobs, D. C. (2021). *A map of diffuse radio emission at 182 mhz to enhance epoch of reionization observations in the southern hemisphere.*

Castorina, E., Foreman, S., Karagiannis, D., Liu, A., Masui, K. W., Meerburg, P. D., ... White, M. (2020, February). Packed Ultra-wideband Mapping Array (PUMA): Astro2020 RFI Response. *arXiv e-prints*, arXiv:2002.05072.

Chokshi, A., Line, J. L. B., Barry, N., Ung, D., Kenney, D., McPhail, A., ... Webster, R. L. (2021, Jan). Dual polarization measurements of mwa beampatterns at 137 mhz. *Monthly Notices of the Royal Astronomical Society*, *502*(2), 1990–2004. Retrieved from `http://dx.doi.org/10.1093/mnras/stab156` doi: 10.1093/mnras/stab156

de Gasperin, F., Dijkema, T. J., Drabent, A., Mevius, M., Rafferty, D., van Weeren, R., ... Williams, W. (2019, February). Systematic effects in LOFAR data: A unified calibration strategy. A&A, *622*, A5. doi: 10.1051/0004-6361/201833867

DeBoer, D. R., Parsons, A. R., Aguirre, J. E., Alexander, P., Ali, Z. S., Beardsley, A. P., ... Zheng, H. (2017, April). Hydrogen Epoch of Reionization Array (HERA). PASP, *129*(4), 045001. doi: 10.1088/1538-3873/129/974/045001

Dillon, J. S., Lee, M., Ali, Z. S., Parsons, A. R., Orosz, N., Nunhokee, C. D., ... et al. (2020, Oct). Redundant-baseline calibration of the hydrogen epoch of reionization array. *Monthly Notices of the Royal Astronomical Society*, *499*(4), 5840–5861. Retrieved from `http://dx.doi.org/10.1093/mnras/staa3001` doi: 10.1093/mnras/staa3001

Dillon, J. S., & Parsons, A. R. (2016, August). Redundant Array Configurations for 21 cm Cosmology. ApJ, *826*(2), 181. doi: 10.3847/0004-637X/826/2/181

Ewall-Wice, A., Bradley, R., Deboer, D., Hewitt, J., Parsons, A., Aguirre, J., ... Wirt, B. (2016, November). The Hydrogen Epoch of Reionization Array Dish. II. Characterization of Spectral Structure with Electromagnetic Simulations and Its Science Implications. ApJ, *831*(2), 196. doi: 10.3847/0004-637X/831/2/196

Ewall-Wice, A., Dillon, J. S., Hewitt, J. N., Loeb, A., Mesinger, A., Neben, A. R., ... et al. (2016, May). First limits on the 21 cm power spectrum during the epoch of x-ray heating. *Monthly Notices of the Royal Astronomical Society*, *460*(4), 4320–4347. Retrieved from `http://dx.doi.org/10.1093/mnras/stw1022` doi: 10.1093/mnras/stw1022

Fagnoni, N., de Lera Acedo, E., DeBoer, D. R., Abdurashidova, Z., Aguirre, J. E., Alexander, P., ... et al. (2020, Oct). Understanding the hera phase i receiver system with simulations and its impact on the detectability of the eor delay power spectrum. *Monthly Notices of the Royal Astronomical Society*, *500*(1), 1232–1242. Retrieved from `http://dx.doi.org/10.1093/mnras/staa3268` doi: 10.1093/mnras/staa3268

Furlanetto, S. R., Peng Oh, S., & Briggs, F. H. (2006, Oct). Cosmology at low frequencies: The 21cm transition and the high-redshift universe. *Physics Reports*, *433*(4-6), 181–301. Retrieved from `http://dx.doi.org/10.1016/j.physrep.2006.08.002` doi: 10.1016/j.physrep.2006.08.002

Hallinan, G., Ravi, V., & Deep Synoptic Array Team. (2021, January). The DSA-2000: A Radio Survey Camera. In *American astronomical society meeting abstracts* (Vol. 53, p. 316.05).

HERA Collaboration, Abdurashidova, Z., Aguirre, J. E., Alexander, P., Ali, Z. S., Balfour, Y., ... Zheng, H. (2021). *First results from hera phase i: Upper limits on the epoch of reionization 21 cm power spectrum.*

Joseph, R. C., Trott, C. M., Wayth, R. B., & Nasirudin, A. (2019, 12). Calibration and 21-cm power spectrum estimation in the presence of antenna beam variations. *Monthly Notices of the Royal Astronomical Society*, *492*(2), 2017-2028. Retrieved from `https://doi.org/10.1093/mnras/stz3375` doi: 10.1093/mnras/stz3375

Kern, N. S., Dillon, J. S., Parsons, A. R., Carilli, C. L., Bernardi, G., Abdurashidova, Z., ... Zheng, H. (2020, February). Absolute Calibration Strategies for the Hydrogen Epoch of Reionization Array and Their Impact on the 21 cm Power Spectrum. ApJ, *890*(2), 122. doi: 10.3847/1538-4357/ab67bc

Kern, N. S., Parsons, A. R., Dillon, J. S., Lanman, A. E., Fagnoni, N., & de Lera Acedo, E. (2019, October). Mitigating Internal Instrument Coupling for 21 cm Cosmology. I. Temporal and Spectral Modeling in Simulations. ApJ, *884*(2), 105. doi: 10.3847/1538-4357/ab3e73

Kern, N. S., Parsons, A. R., Dillon, J. S., Lanman, A. E., Liu, A., Bull, P., ... Zheng, H. (2020, January). Mitigating Internal Instrument Coupling for 21 cm Cosmology. II. A Method Demonstration with the Hydrogen Epoch of Reionization Array. ApJ, *888*(2), 70. doi: 10.3847/1538-4357/ab5e8a

Liu, A., & Shaw, J. R. (2020, Apr). Data analysis for precision 21 cm cosmology. *Publications of the Astronomical Society of the Pacific*, *132*(1012), 062001. Retrieved from `http://dx.doi.org/10.1088/1538-3873/ab5bfd` doi: 10.1088/1538-3873/ab5bfd

Mellema, G., Koopmans, L. V. E., Abdalla, F. A., Bernardi, G., Ciardi, B., Daiboo, S., ... Zaroubi, S. (2013, August). Reionization and the Cosmic Dawn with the Square Kilometre Array. *Experimental Astronomy*, *36*(1-2), 235-318. doi: 10.1007/s10686-013-9334-5

Morales, M. F., & Wyithe, J. S. B. (2010, September). Reionization and Cosmology with 21-cm Fluctuations. ARA&A, *48*, 127-171. doi: 10.1146/annurev-astro-081309-130936

Moreira, P., Serrano, J., Wlostowski, T., Loschmidt, P., & Gaderer, G. (2009). White rabbit: Sub-nanosecond timing distribution over ethernet. *2009 International Symposium on Precision Clock Synchronization for Measurement, Control and Communication*, 1-5.

Newburgh, L. B., Addison, G. E., Amiri, M., Bandura, K., Bond, J. R., Connor, L., ... et al. (2014, Jul). Calibrating chime: a new radio interferometer to probe dark energy. *Ground-based and Airborne Telescopes V*. Retrieved from `http://dx.doi.org/10.1117/12.2056962` doi: 10.1117/12.2056962

Paciga, G., Chang, T.-C., Gupta, Y., Nityanada, R., Odegova, J., Pen, U.-L., ... Sigurdson, K. (2011, 04). The GMRT Epoch of Reionization experiment: a new upper limit on the neutral hydrogen power spectrum at z$\sim$ 8.6. *Monthly Notices of the Royal Astronomical Society*, *413*(2), 1174-1183. Retrieved from `https://doi.org/10.1111/j.1365-2966.2011.18208.x` doi: 10.1111/j.1365-2966.2011.18208.x

Parsons, A. R., Backer, D. C., Foster, G. S., Wright, M. C. H., Bradley, R. F., Gugliucci, N. E., ... Werthimer, D. J. (2010, April). The Precision Array for Probing the Epoch of Re-ionization: Eight Station Results. AJ, *139*(4), 1468-1480. doi: 10.1088/0004-6256/139/4/1468

Pritchard, J. R., & Loeb, A. (2012, Jul). 21 cm cosmology in the 21st century. *Reports on Progress in Physics*, *75*(8), 086901. Retrieved from `http://dx.doi.org/10.1088/0034-4885/75/8/086901` doi: 10.1088/0034-4885/75/8/086901

Remazeilles, M., Dickinson, C., Banday, A. J., Bigot-Sazy, M. A., & Ghosh, T. (2015). *An improved source-subtracted and destriped 408 mhz all-sky map.*

Saliwanchik, B. R. B., Ewall-Wice, A., Crichton, D., Kuhn, E. R., Ölçek, D., Bandura, K., ... Wulf, D. (2021). *Mechanical and optical design of the hirax radio telescope.*

Santos, M. G., Cooray, A., & Knox, L. (2005, Jun). Multifrequency analysis of 21 centimeter fluctuations from the era of reionization. *The Astrophysical Journal*, *625*(2), 575–587. Retrieved from `http://dx.doi.org/10.1086/429857` doi: 10.1086/429857

Star, P. (2020). *Implementation of van vleck correction for the mwa.* Retrieved from `https://github.com/EoRImaging/Memos/blob/master/PDFs/007_Van_Vleck_A.pdf`

Taylor, & Rupen, M. P. (1999, January). *Synthesis Imaging in Radio Astronomy II* (Vol. 180).

Tingay, S. J., Goeke, R., Bowman, J. D., Emrich, D., Ord, S. M., Mitchell, D. A., ... Wyithe, J. S. B. (2013, January). The Murchison Widefield Array: The Square Kilometre Array Precursor at Low Radio Frequencies. PASA, *30*, e007. doi: 10.1017/pasa.2012.007

van Haarlem, M. P., Wise, M. W., Gunst, A. W., Heald, G., McKean, J. P., Hessels, J. W. T., ... van Zwieten, J. (2013, August). LOFAR: The LOw-Frequency ARray. A&A, *556*, A2. doi: 10.1051/0004-6361/201220873

Vanderlinde, K., Liu, A., Gaensler, B., Bond, D., Hinshaw, G., Ng, C., ... Kaspi, V. (2019, October). The Canadian Hydrogen Observatory and Radio-transient Detector (CHORD). In *Canadian long range plan for astronomy and astrophysics white papers* (Vol. 2020, p. 28). doi: 10.5281/zenodo.3765414

White, S. V., Franzen, T. M. O., Riseley, C. J., Wong, O. I., Kapińska, A. D., Hurley-Walker, N., ... et al. (2020). The gleam 4-jy (g4jy) sample: I. definition and the catalogue. *Publications of the Astronomical Society of Australia*, *37*. Retrieved from `http://dx.doi.org/10.1017/pasa.2020.9` doi: 10.1017/pasa.2020.9

Whitler, L. R., Beardsley, A., & Jacobs, D. (2019, January). The Effects of RFI on

21-cm Measurements of the Epoch of Reionization. In *American astronomical society meeting abstracts #233* (Vol. 233, p. 349.17).
Wilensky, M. J., Barry, N., Morales, M. F., Hazelton, B. J., & Byrne, R. (2020, 08). Quantifying excess power from radio frequency interference in Epoch of Reionization measurements. *Monthly Notices of the Royal Astronomical Society*, *498*(1), 265-275. Retrieved from https://doi.org/10.1093/mnras/staa2442 doi: 10.1093/mnras/staa2442

Appendix A Algorithms

Algorithm 1 provides pseudocode for the iterative antenna flagging algorithm described in Section 2.3. Algorithm 2 provides pseudocode for the auto-correlation based flagging algorithm described in Section 3.

Algorithm 1: `corr_metrics` pseudocode.

```
// Compute the correlation metric as defined in Equation 1 for every
   baseline.
// Identify completely dead antennas
for ant ∈ ants do
    if per-ant median corr metric == 0 then
        Flag and remove ant.
    else
        Continue.
    end
end
// Iteratively flag and recalculate metric
for iteration do
    Recalculate corr and cross metrics using only unflagged antennas.
    if Worst dead ant is worse than worst crossed ant and below flag threshold
      then
        Flag ant as dead and remove
    else if Worse crossed ant is worse than worst dead ant and below flag
      threshold then
        Flag ant as crossed and remove
    else
        break
    end
end
```

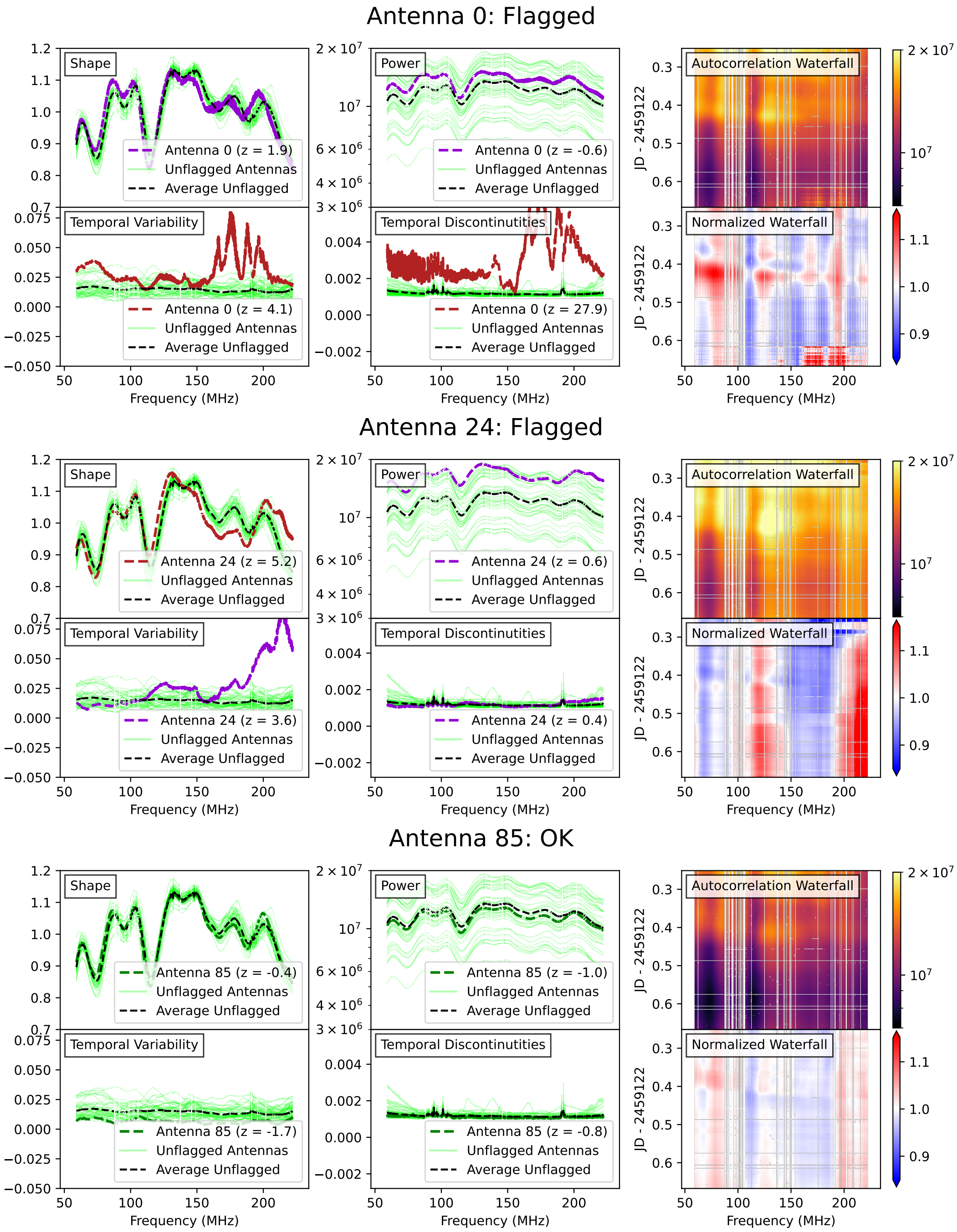

Figure 15. An example of the summary dashboard used to inspect antenna metrics showing three cases –one for a clearly malfunctioning antenna, one for a borderline flagged antenna, and one for a good antenna. In each we show the metric spectra of the individual antenna compared to all good antennas in light green, helping us to easily see whether the antenna is an outlier. We also show the full auto-correlation waterfalls, both raw and fractional deviation from the antenna average (Figure 10). The effects detected by our metrics can generally be seen in either the raw or normalized waterfall.

Algorithm 2: Pseudocode for `auto_metrics`, our method for identifying outliers in antenna auto-correlations.

```
// First look for outliers with median-based statistics.
for V_ii ∈ auto-correlation waterfalls do
    for metric ∈ all auto_metrics do
        Compute median-based statistical metric spectrum from V_ii.
    end
end
while True do
    for metric ∈ all auto_metrics do
        Compute median metric spectrum over all V_ii for each polarization.
        for spectrum ∈ all spectra do
            Compute a median-based distance between each spectrum and the
             median over antennas.
        end
        for distance metric ∈ all distance metrics do
            Compute a modified z-score from the distribution of distances metrics
             of unflagged antennas.
        end
    end
    if ∃ modified z-scores above a given threshold then
        Flag the antenna with the largest single z-score over all metrics.
    else
        Break.
    end
end
// Next excise RFI.
Average all unflagged antennas into a single auto-correlation waterfall.
Flag all outlier pixels, channels, or integrations with the XRFI algorithm.
// Now look for outliers with mean-based statistics.
for V_ii ∈ auto-correlation waterfalls do
    for metric ∈ all auto_metrics do
        Compute mean-based statistical metric spectrum from V_ii, excluding RFI
         flags.
    end
end
while True do
    for metric ∈ all auto_metrics do
        Compute mean metric spectrum over all V_ii for each polarization.
        for spectrum ∈ all spectra do
            Compute a mean-based distance between each spectrum and the mean
             over antennas.
        end
        for distance metric ∈ all distance metrics do
            Compute a modified z-score from the distribution of distances metrics
             of unflagged antennas.
        end
    end
    if ∃ modified z-scores above a given threshold then
        Flag the antenna with the largest single z-score over all metrics.
    else
        Break.
    end
end
```