

UNDERSTANDING POLARIZATION AS A FOREGROUND FOR HI EPOCH OF
REIONIZATION MEASUREMENTS

Saul Aryeh Kohn

A DISSERTATION

in

Physics and Astronomy

Presented to the Faculties of the University of Pennsylvania

in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2018

Supervisor of Dissertation

Graduate Group Chairperson

James E. Aguirre
Associate Professor of Physics and Astronomy

Whoever the graduate chair is
Professor of Physics and Astronomy

Dissertation Committee:

Adam Lidz, Associate Professor of Physics and Astronomy

Masao Sako, Associate Professor of Physics and Astronomy

another professor, Assistant Professor of Physics and Astronomy

another professor, Professor of Physics and Astronomy

for my grandparents, endless sources of inspiration

UNDERSTANDING POLARIZATION AS A FOREGROUND FOR HI EPOCH OF
REIONIZATION MEASUREMENTS

© COPYRIGHT

2018

Saul Aryeh Kohn

This work is licensed under the

Creative Commons Attribution

NonCommercial-ShareAlike 3.0

License

To view a copy of this license, visit

<http://creativecommons.org/licenses/by-nc-sa/3.0/>

Acknowledgments

Acknowledgements require a certain mindset to be written well.

ABSTRACT

UNDERSTANDING POLARIZATION AS A FOREGROUND FOR HI EPOCH OF
REIONIZATION MEASUREMENTS

Saul A. Kohn

James E. Aguirre

Abstracts are written last.

Contents

Title	i
Dedication	ii
Acknowledgments	iv
Abstract	v
List of Tables	ix
List of Figures	x
I Introduction & Mathematical Formalisms	1
1 The Epoch of Reionization	2
2 Astrophysical Polarization	3
3 Interferometry	4
4 Instrumental Polarization	5
5 Instruments	6
5.1 Instruments used in this work	7
5.1.1 The Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER)	7
5.1.1.1 PAPER-32 redundant array	7
5.1.1.2 PAPER-32 polarized imaging array	7
5.1.1.3 PAPER-64	7
5.1.1.4 PAPER-128	7
5.1.2 The Hydrogen Epoch of Reionization Array (HERA)	7
5.1.2.1 HERA-19 Commissioning Array	7
5.1.2.2 HERA-47	7

5.1.2.3	Future HERA Build-Outs	7
5.2	Other current and future low-frequency interferometers	7
5.2.1	The Low Frequency Array (LOFAR)	7
5.2.2	The Murchinson Widefield Array (MWA)	7
5.2.3	Square Kilometer Array – Low band (SKA-Low)	7
II	Outer space in Fourier space	8
6	Peering through the EoR Window	9
7	Data Preparation and Processing	10
7.1	Data Compression	10
7.1.1	Delay–Delay-Rate Filtering	11
7.1.2	Software Implementation	13
7.2	Radio Frequency Interference	15
7.2.1	PAPER-128	15
7.2.1.1	Average Properties	16
7.2.1.2	Individual Properties	21
7.2.1.3	Discussion	22
7.2.2	HERA-19 and PAPER-19	23
7.3	Crosstalk in PAPER-64	23
7.4	Pre-Redundant Calibration QA	23
7.5	Post-Redundant Calibration QA	23
8	Polarimetric Calibration	29
8.1	Redundant Calibration	29
8.2	Imaging Calibration	29
9	The Ionosphere	30
10	A view of the EoR window from the PAPER-32 imaging array	31
11	A view of the EoR window from the HERA-19 commissioning array	32
12	Deep integrations with PAPER-128	33
III	Expanding the potential of EoR measurements	34
13	Higher-order correlation functions between the kSZ and 21cm fields during the EoR	35
14	Deep Learning for 21cm Observations	36

15 Conclusions	37
Appendices	38
A Software	39
A.1 Astronomical Interferometry in Python (aipy)	39
A.2 Astronomy in Python (astropy)	40
A.3 Common Astronomy Software Applications (CASA)	40
A.4 Deep Learning packages	40
A.5 Hierarchical Equal Area isoLatitude Pixelization of the sphere (HEALPix)	40
A.6 pyuvdata	41
A.7 The Scientific Python Ecosystem (scipy)	41
Appendices	39
Bibliography	42

List of Tables

7.1	RFI frequencies and brief characterization for the averaged flags.	18
-----	--	----

List of Figures

7.1	The schema of the database used to organize and implement PAPER data compression.	14
7.2	A waterfall plot of RFI flags averaged over 150 days of PAPER-128 data.	24
7.3	The percentage of time that each frequency was flagged over the season. .	25
7.4	Possible FM radio contamination.	25
7.5	Flights from Cape Town to Johannesburg correspond to RFI in the 120.15 ± 0.35 MHz channels.	26
7.6	The temporal profile of the 5 RFI frequencies with unidentified causes. . .	27
7.7	Waterfalls of RFI flags for nights 2456732, 2456958 and 2457038.	28
7.8	Waterfalls of RFI flags for nights 2456898, 2456924 and 2456965	28

Part I

Introduction & Mathematical Formalisms

Chapter 1

The Epoch of Reionization

Chapter 2

Astrophysical Polarization

Chapter 3

Interferometry

Chapter 4

Instrumental Polarization

Chapter 5

Instruments

5.1 Instruments used in this work

5.1.1 The Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER)

5.1.1.1 PAPER-32 redundant array

5.1.1.2 PAPER-32 polarized imaging array

5.1.1.3 PAPER-64

5.1.1.4 PAPER-128

5.1.2 The Hydrogen Epoch of Reionization Array (HERA)

5.1.2.1 HERA-19 Commissioning Array

5.1.2.2 HERA-47

5.1.2.3 Future HERA Build-Outs

5.2 Other current and future low-frequency interferometers

7

5.2.1 The Low Frequency Array (LOFAR)

5.2.2 The Murchinson Widefield Array (MWA)

Part II

Outer space in Fourier space

Chapter 6

Peering through the EoR Window

Chapter 7

Data Preparation and Processing

The data volume of interferometric measurements inherently scale as the square of the number of antennas in the array (N_{ant}). Not only does the sheer volume of data from large- N_{ant} arrays pose a problem for data storage, but also it requires precise and efficient efforts to quality assure (QA) the data.

In this chapter, I will outline some of the efforts involved in data preparation, preprocessing and QA that are required for an EoR power spectrum estimate.

7.1 Data Compression

The PAPER-128 correlator produced 288 MIRIAD files per night. Each of these contained 8126 baselines, and each baseline contained visibilities over 1024 98 kHz frequency channels and 56 10 s time integrations. The four instrumental polarizations were in separate files. In sum, each file was 4.2 GB which meant that each night 1.2 TB of data were recorded.

In order to efficiently transport the data over Gigabit Ethernet from the Karoo Radio Quiet Zone (KRQZ) to Cape Town, and from Cape Town under transatlantic cables to Philadelphia, some compression was required. It was also required that such a compression, while lossy, did not effect the targeted cosmological signal.

7.1.1 Delay–Delay-Rate Filtering

The compression algorithm implemented for PAPER observations, Delay–Delay-Rate (DDR) filtering, was introduced in Parsons and Backer (2009) described in Parsons et al. (2014), and we briefly review it below.

The geometric delay of a celestial signal, originating from direction \hat{s} , incident on an interferometric baseline described by vector \vec{b} , is

$$\tau_g = |\vec{b} \cdot \hat{s}|/c \quad (7.1)$$

where c is the speed of light. This relationship implies that τ_g is bounded for a given baseline

$$-|\vec{b}|/c \leq \tau_g \leq |\vec{b}|/c \quad (7.2)$$

Equation 7.2 therefore gives the maximum value of $|\tau_g|$ physically meaningful for a given array – the maximum baseline length in that array, divided by c . For PAPER, the maximum baseline length is 300 m, corresponding to $\max(|\tau_g|) = 1\mu\text{s}$. As reviewed in Chapter 6, the delay axis may be accessed by Fourier transforming a visibility along the frequency axis. Once in delay space, power at delays larger in magnitude than $1\mu\text{s}$ could be removed. With a sufficiently large frequency bandwidth, this would not produce aliased signal, according to the critical Nyquist rate. By using the $1\mu\text{s}$ as a delay bound for all visibilities, the frequency axes of all compressed visibilities remained the same (reduced in number from 1024 to 203), which while sub-optimal from a compression point of view, allowed for ease of programming at later stages.

A similar geometric bound can be obtained by Fourier transforming the time axis of visibilities, provided that they were obtained in drift-scan mode (see Chapter 3). Parsons and Backer (2009) showed that the rate at which the geometric delay on an interferometric baseline changes is governed only by the position of the array on Earth, and the Earth’s rotation:

$$\dot{\tau}_g = -\frac{\omega_{\oplus} \cos \delta}{c} (b_x \sin \alpha + b_y \cos \alpha) \quad (7.3)$$

where ω_{\oplus} is the angular frequency of the Earth's rotation, α and δ are the hour-angle and declination of a point on the celestial sphere, respectively, and $\vec{b} = (b_x, b_y, b_z)$ is the baseline vector expressed in equatorial coordinates.

For arrays not close to the geographic poles, $|b_y| \gg |b_x|$, there is a maximum rate of change (corresponding to $(\alpha, \delta) = (0, 0)$), producing a bound on $\dot{\tau}_g$:

$$-\omega_{\oplus}|b_y|/c \leq \dot{\tau}_g \leq \omega_{\oplus}|b_y|/c \quad (7.4)$$

for a 300 m East-West baseline, the maximum delay-rate is approximately $\max(|\dot{\tau}_g|) = 0.07 \text{ ns s}^{-1}$. This delay-rate was not Nyquist sampled by a single PAPER file: requiring the previous and next files generated for that polarization to be appended on either side of each visibility's time axis to prevent aliasing from the decimation. For the large scale processing of months of data, this required a software pipeline described in Section 7.1.2.

There are also other issues with DDR compression, largely associated with instrument systematics. Delay transforms rely on the fact that the bright foregrounds that dominate the measured signal are spectrally smooth, and that the frequency response of the instrument is also spectrally smooth: this of course is the basis for the EoR window paradigm reviewed in Chapter 6. Likewise, delay-rate filtering assumes temporal smoothness. Radio Frequency Interference (RFI) signals created by human communications violate both models of smoothness, since they are typically confined to narrow bandwidths (creating sharp spikes along the frequency axis) and may be transient (creating sharp spikes along the time axis). This requires steadfast identification and flagging algorithms for RFI (see Section 7.2), and some variety of interpolation, fitting, or CLEANing across the flagged regions prior to compression.

By DDR filtering of PAPER-128 data using a 300 m baseline to set the width of the filters we were able to reduce the volume of the data by an approximate factor of 70.

7.1.2 Software Implementation

The first season of PAPER-128 data, due to a variety of circumstances, required compression on the computing cluster at the University of Pennsylvania. The raw data were stored on a high-volume drive that was able to connect with the cluster via a low-speed switch. The hardware capable of performing any sort of high-performance processing (i.e. holding the data in RAM) were ten “compute nodes” connected to the cluster via a high-speed switch, and mounted in an NFS architecture. The compute nodes could only hold ~ 10 PAPER-128 files in storage.

The processing stages for compression of a night of PAPER data, described below, required knowledge of the location and compression state of not only individual files, but also the neighbors-in-time of the file in question, in order to implement the DDR filter described above. To supervise the compression we created a MySQL database, which we interacted with via Shell and Python scripts. The database contained a table for the data files under processing and their compression state, a table of neighbor-relations, a table of file details, and a table of the processing nodes available. The schema of this database is shown in Figure 7.1.

To implement the compression, per file, the following steps were required:

1. Copying the file from the storage volume to the cluster. For a single night of data, this required roughly 8 hours.
2. Copying the file from the cluster to the compute node. This required roughly 5 minutes.
3. Generate copy of the file, with metadata corrections. This required roughly 1 minute.
4. Delete the raw file.
5. RFI-flag the high frequency-resolution data. This required roughly 2 minutes.
6. Delete the metadata-corrected file.

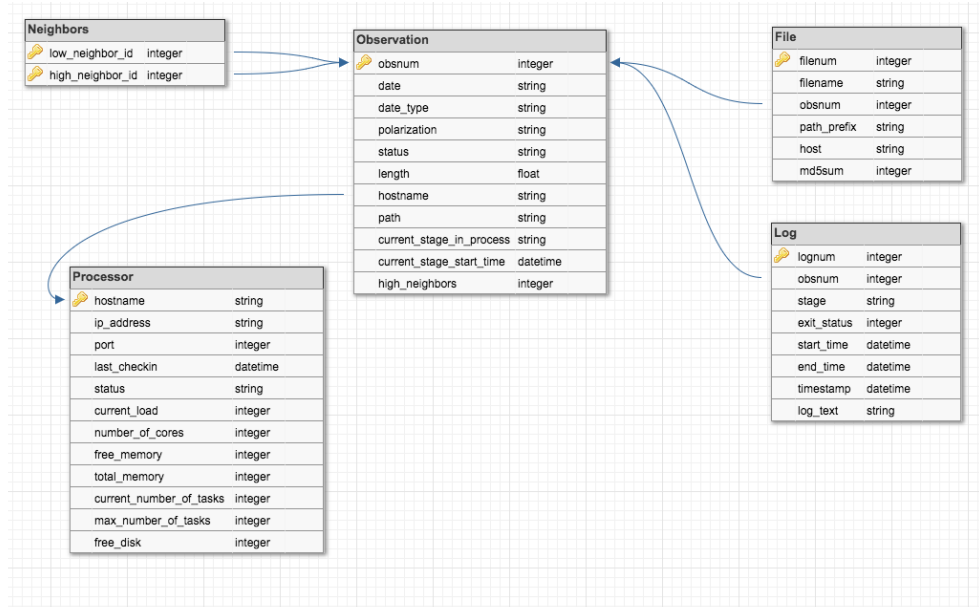


Figure 7.1: The schema of the database used to organize and implement PAPER data compression.

7. Acquire time-neighbors to the file in question, and bring them to the RFI-flagged stage. The time required for this stage varied with cluster activity, but usually required roughly 20 minutes.
8. DDR filter the RFI-flagged data, using an high-tolerance iterative CLEAN. This required roughly 20 minutes.
9. RFI flag the compressed data (coarse flagging), saving the flags to a separate file. This required roughly 1 minute.
10. Apply the coarse RFI flags to the *uncompressed*, RFI-flagged data. This required roughly one minute.
11. DDR filter the now twice-RFI-flagged data, using a low-tolerance iterative CLEAN. This required roughly 120 minutes.
12. Copy compressed data to the cluster.

13. Delete the twice-RFI-flagged data.
14. If the once-RFI-flagged data are not required as neighbors, delete them.
15. Delete the compressed data from the compute node.
16. If neighbors have already been compressed, delete them, otherwise begin their compression.
17. After all files are compressed, delete the uncompressed files from the cluster.

In total, this meant that across ten compute nodes, and efficient use of the fact that the neighbors could progress through the processing stages while the central file was being compressed, meant that it took roughly 20 to 24 hours to compress a night of observations.

7.2 Radio Frequency Interference

As noted above, RFI was able to introduce spectral and temporal structure that would cause ringing in the data during compression if it was not flagged. This meant that both identification and characterization of RFI was crucial to the scientific goals of the PAPER and HERA experiments. In Section 7.2.1, I present characterization of RFI in the second season of PAPER-128 data. By averaging flags in local time I was able to investigate “repeat offender” frequency bands and identify outlying “quiet” and “loud” days. In Section 7.2.2 I analyze RFI flags from the first Internal Data Release (IDR1) of HERA commissioning data, which contained 19 HERA feeds suspended above 5 m dishes in a close-packed hexagon, and 19 PAPER feeds in the same positions as the central dishes, allowing us to investigate the difference in flagging between feeds at different altitudes.

7.2.1 PAPER-128

The PAPER-128 2014 observation season ran from 18th June 2014 through the 30th April 2015. During this run, some 150 nights of data were recorded. A “night”, which I will

refer to using the JD at the start of observations, consists of twelve hours of observation from 6pm to 6am South African Standard Time (SAST). Observations as processed by the PAPER correlator are recorded in MIRIAD uv files. These files contain visibilities for each antenna pair in the array. Each integration is 20 seconds long over 1024 frequency bins from 100 to 200 MHz. Each uv file contains 56 integrations per antenna pair, and 72 uv files are recorded per linear polarization (xx, xy, yx, yy) per night.

Early in the PAPER data compression process, visibilities are flagged for RFI. This is accomplished by the aipy script *xrfi_simple.py*, which takes the derivative of the frequency axis of all baselines associated with a single antenna, and flags any frequencies with a derivative $\geq 6\sigma$ above the mean. We always flag the band-edges (~ 7 MHz on each side), since these frequencies are not useful to us, and always flag the 137 ± 0.6 MHz band associated with ORBCOMM satellite network transmissions. This process is repeated per integration within each uv file and stored in a Python numpy zip (npz) file. This means that any baseline associated with antenna 1 can contribute a flag to the resultant npz file, which in turn is applied to the data.

The result is 280 files of high-time and -frequency resolution files per night per linear polarization containing information about the RFI environment of the HERA site. I report on the properties of these flags in time- and frequency-space over the 2014 observation season. This section is organized it as follows: in Section 7.2.1.1, I analyse the average properties of RFI over the season by stacking flags in local time and normalizing appropriately. In Section 7.2.1.2 I address nights with particularly strange RFI properties. I discuss the implications of my findings in Section 7.2.1.3.

7.2.1.1 Average Properties

In order to assess the average properties of the RFI environment, I calculated a weighted average of flags over the season. Over 150 nights, one-time occurrences are washed-out beneath the 1% level, allowing me to assess persistent issues.

Nominally, each night should grant 3920 integrations-worth of flags over 1024 fre-

quency bins, per linear polarization. In reality, most of the time this holds true, but occasionally not all files are compressible (hence failing to generate flags) or observations fail to start at the correct time (so there are no data to flag). Also, in the event of an X-engine failure within the correlator, contiguous chunks of the band (in eighths, i.e. 25 MHz across) are flagged-out, usually for the rest of the night.

For this reason, I calculated a weighted average of the flags across the season, but neglected nights with correlator failures or late starts. Weights were simply the number of nights that contained that integration-bin in SAST. The resultant “flag density waterfall” is shown in Figure 7.2. The color scale is indicative of flagging frequency across the season, and line plots above and to the right of the waterfall showing the percentage of times and frequencies that were flagged, respectively.

A summary of the persistent (flagged $\geq 1\%$ of the time per channel) RFI frequencies can be found in Table 7.2.1.1. I have investigated each frequency and tried to find the most likely source for each. In most cases, this required looking at the properties in time as well as frequency. Others were more obvious from frequency alone, e.g. the 149.8 MHz transmission frequency from the International Space Station (ISS). Still others I could not track down a convincing explanation for, and these are listed with a ‘?’. A ‘?’ next to a possible cause indicates that the listed cause is the most prevalent at that frequency, but that the temporal properties of that cause do not necessarily make sense. Many of the characterizations arise from the South African Table of Frequency Allocations (SATFA; Staatskoerant (2008)).

Figure 7.3 shows the detail of the top panel of Figure 7.2. This figure highlights the broad swath of the band from roughly 150 to 180 MHz that was, on average, clear of RFI. This roughly corresponds to 21 cm redshifts $z = 6.9$ to 8.5 . This is one of the reasons that the Parsons et al. (2014) and Ali et al. (2015) limits on the 21 cm power spectrum concentrated on this redshift range – there were simply more unflagged data to average-down with. Furlanetto et al. (2006) show that the $z \sim 8$ universe can be considered roughly coeval over an ~ 8 MHz bandwidth. As such, the 30 MHz chunk could be used to

Table 7.1. RFI frequencies and brief characterization for the averaged flags.

ν MHz	Flagged %	Cause (Possible)	Notes or SAST Characterization
103 ± 3	100	BAND EDGE	Built-in to flagger.
107.25 ± 0.25	2.6	FM radio	Constant background at 2% level
107.55 ± 0.05	1.9	FM radio	Constant background at 2% level
108.1 ± 0.4	9	FM radio?	Rises with time, peaking at midnight and 4am
109 ± 0.4	11.5	FM radio?	Rises with time, peaking around 4am
112.8 ± 0.1	1.4	Aircraft?	Constant background at 1% level
114.05 ± 0.85	3.7	?1	Decreases till midnight; peak at 4am
116.55 ± 0.35	2.2	?2	Peak at midnight
120.15 ± 0.35	3.2	Aircraft	Roughly follows CPT↔JNB flight times
124.95 ± 0.35	5.5	Aircraft	Roughly follows CPT↔JNB flight times
130.25 ± 0.55	4.3	?3	Falls (7pm) and rises (3am) steeply
131.75 ± 0.35	10.3	Aircraft?	Peaks at 6:30, 7:30, 8:30, 9:30, 10 and then a steep falloff
136.05 ± 0.45	33.1	Radar?	Decreases over night
137.35 ± 0.85	100	ORBCOMM	
141.45 ± 0.35	2.1	Mobile phones?	High until 9pm, then at background 1% level
145.85 ± 0.45	10.7	Amateur radio	Strong 9pm-1am – this is the official downlink for ISS-HAM
149.75 ± 0.55	90.7	ISS	“Beeps”, but in stacked data peaks 2am
175.15 ± 0.35	20.5	VHF TV (video)	Channel 4. Peaks at 8:30pm, then falls to background 7%
181.15 ± 0.15	1.6	VHF TV (audio)	Channel 4. 2% level turns-off at 10pm
182.15 ± 0.35	75	?4	Decreases until 10pm (to 15%), when it begins a slow rise again
183.2 ± 0.5	89.7	VHF TV (video)	Channel 5. Rises throughout night.
186.25 ± 0.35	4.6	?5	Extreme turn-off at 9:45
189.15 ± 0.35	41.4	VHF TV (audio)	Channel 5. Rises throughout night.
189.9 ± 0.4	100	VHF TV	Channel 6. Built-in to flagger.
191.1 ± 0.3	100	VHF TV	Channel 7. Built-in to flagger.
196 ± 4	100	BAND EDGE	

create ~ 3 power spectra, as demonstrated in Jacobs et al. (2015). As we show below, the deactivation of VHF TV broadcasts could enable measurements up to the band edge.

FM Radio

SATFA lists the frequency band 87.5–108 MHz as available for FM radio broadcasts, leading me to postulate that the low-level RFI we observed in the 107.25 ± 0.25 and 107.55 ± 0.05 MHz bands had FM radio as the leading cause. The 108.1 ± 0.4 and 109 ± 0.4 MHz bands were outside of the official range, and exhibit odd temporal properties for human activity – two peaks at midnight and 4am – with a increasing number of flags throughout the average night (see Figure 7.4).

Aircraft communications

It was difficult to argue that the 112.1 ± 0.1 MHz signal is caused by aircraft communications since it maintained a constant background level. However, SATFA listed this frequency as reserved for aircraft communications and it has been used in the past as a calibration frequency for aircraft instruments Authority (2008).

The other aircraft frequencies were obvious, because they closely traced the 2-hour flight from Cape Town to Johannesburg¹. An example (120.15 ± 0.35 MHz) is shown in Figure 7.5. SATFA reserved frequencies 108–117.975 MHz for aeronautical radionavigation and 117.975–137 MHz for aeronautical mobile. In Table 7.2.1.1 I listed 131.75 ± 0.35 MHz as caused by aircraft since it falls in the aeronautical mobile band, but it does not follow the flight patterns as closely as the other bands.

¹Credit to Danny Jacobs for first spotting this and noting it in an internal PAPER circular in December 2009.

Orbital communications

ORBCOMM Inc.’s constellation of 29 LEO communication satellites is a well-known contaminant of the low-frequency sky, dominating over any astronomical signal at 137–138 MHz (although each satellite emits within a 20 kHz band). For this reason there was built-in flagging at 137.35 ± 0.85 MHz within the compression pipeline.

The largest contaminant without built-in flags in the pipeline were communications from the ISS. The 149.75 ± 0.55 MHz transmissions were semi-regular in time; they ‘beep’.

Onboard the ISS are HAM radio devices. Some countries have also launched satellites with these onboard, one of the purposes of which is to provide HAM radio operators something in space to communicate with. These devices are licensed to operate at 145.2 and 145.8 MHz, and SATFA listed the 144–146 MHz band as reserved for ‘Amateur–Satellite’ communications. We detected RFI at 145.85 ± 0.45 MHz, although strong signal across $\sim 10\%$ of the season that occurs 9pm–1am argues against human operation.

Mobile phones and VHF TV

A weak RFI signal at 141.45 ± 0.35 MHz was within the ‘mobile 1 BTX’ and aeronautical mobile band in SATFA, but other than this single listing I did not build a strong case for the signal’s cause.

VHF TV is broadcast over specifically-spaced video and audio frequencies. The strong signals at 183.2 ± 0.5 MHz and 189.15 ± 0.35 MHz had almost identical gradients for the percentage of flagging as a function of time of night. These frequencies corresponded exactly to Channel 5 of South African System I 625-line VHF TV signals for video and audio transmission, respectively. Similarly, the weaker signals at 175.15 ± 0.35 and 181.15 ± 0.15 MHz corresponded to Channel 4’s video and audio transmission, respectively, but they did not share the same temporal properties.

Unidentified sources

There were 5 RFI frequencies in the averaged data that I could not identify the sources of: weak emissions (flagged $< 5\%$ of the season) at 114.05 ± 0.85 , 116.55 ± 0.35 , 130.25 ± 0.55 and 186.25 ± 0.35 MHz, and one strong emission at 182.15 ± 0.35 MHz. The variation of each source with time is shown in Figure 7.6. The 186.25 ± 0.35 MHz had a sharp turn-off around 9:45pm each night, suggesting that it originated from some kind of automated device.

7.2.1.2 Individual Properties

Using the flags per night, I was able to assess the total number of flags as a percentage of the waterfall (i.e. $N_{\text{flags}} / (3920 \times 1024)$). The average flagging per night was $19.2 \pm 0.5\%$, which was dominated by the permanent flagging of ORBCOMM and band edges. Four nights deviated from the average by a $\geq 2\sigma$ excess: JDs 2456965, 2456732, 2456958 and 2457038. Their flag waterfalls are shown in Figure 7.7 (2456732, 2456958 and 2457038) and Figure 7.8 (2456965). While the strange nature of night 2456965 is discussed below, the three others followed the pattern of having strong contamination from FM and aircraft communication bands, but also had broadband ‘pulses’ up to about 20 minutes in length. The source of these broadband pulses is not well understood, although it was clear that ORBCOMM tends to spill outside of its allocated band on occasion.

JD 2456965 was easily the worst offender, and it exhibited a strange signal that wanders in frequency and time close to the ISS band. An event of note on this date (23rd November 2014) was a Soyuz FG launch that docked with the ISS – this may have been a signature of their transmissions². Similar signals were seen on 2456898 (28th August 2014; although only at the beginning of the night) and 2456924 (23rd September 2014). There was no listed orbital or suborbital activity for 2456898. There was an US ICBM

²The internet also suggests... less plausible explanations: <https://www.youtube.com/watch?t=11&v=VtZx8iP04zs>.

test off of the coast of Virginia on 2456924, but this was probably not the cause of the RFI. The flag waterfalls for these nights are shown in Figure ??.

Another property that the flag waterfalls in Figures 7.7 and 7.8 highlight is the presence of broadband RFI signals, typically present at frequencies lower than the ORB-COMM band. However, while we flagged at the low-end of the band (which had higher noise levels to begin with), it is likely that such broadband pulses dominated the band at those times, and that we failed to flag all of the integrations. Our flagging routine *xrft_simple.py* does contain a thresholding option for flagging the entire integration given some arbitrary number of frequencies flagged during that integration: some experimentation will be required to decide if that threshold should change.

7.2.1.3 Discussion

Based on my findings, I was able to recommend some actions that could be taken in the KRQZ to enable better measurements:

- Steps to reduce and ideally eliminate the VHF TV transmissions in the area would be very helpful, since these were clearly interfering with our measurements in the high-end of the band.
- The ISS 149.75 ± 0.55 MHz band should be permanently flagged within the compression pipeline.
- Pursuing re-routing of flight paths will not do much to help: we see aircraft signals for the duration of their flight, not just when they're over the Karoo.
- A lower threshold for identifying broadband RFI should be optimized.

A new, lower-frequency feed is currently under development by the HERA analog group. This would nominally allow measurements to be taken in the range 50–250 MHz, allowing science observations of the Dark Ages and the post-reionization Universe. It

should be noted that at the lowest frequencies FM radio will be a constant harassment to these measurements. At the higher frequencies, VHF TV will be the primary contaminant, but should be much easier to remove as it is both narrow-band and within the KRQZ's power to shut off.

7.2.2 HERA-19 and PAPER-19

7.3 Crosstalk in PAPER-64

7.4 Pre-Redundant Calibration QA

7.5 Post-Redundant Calibration QA

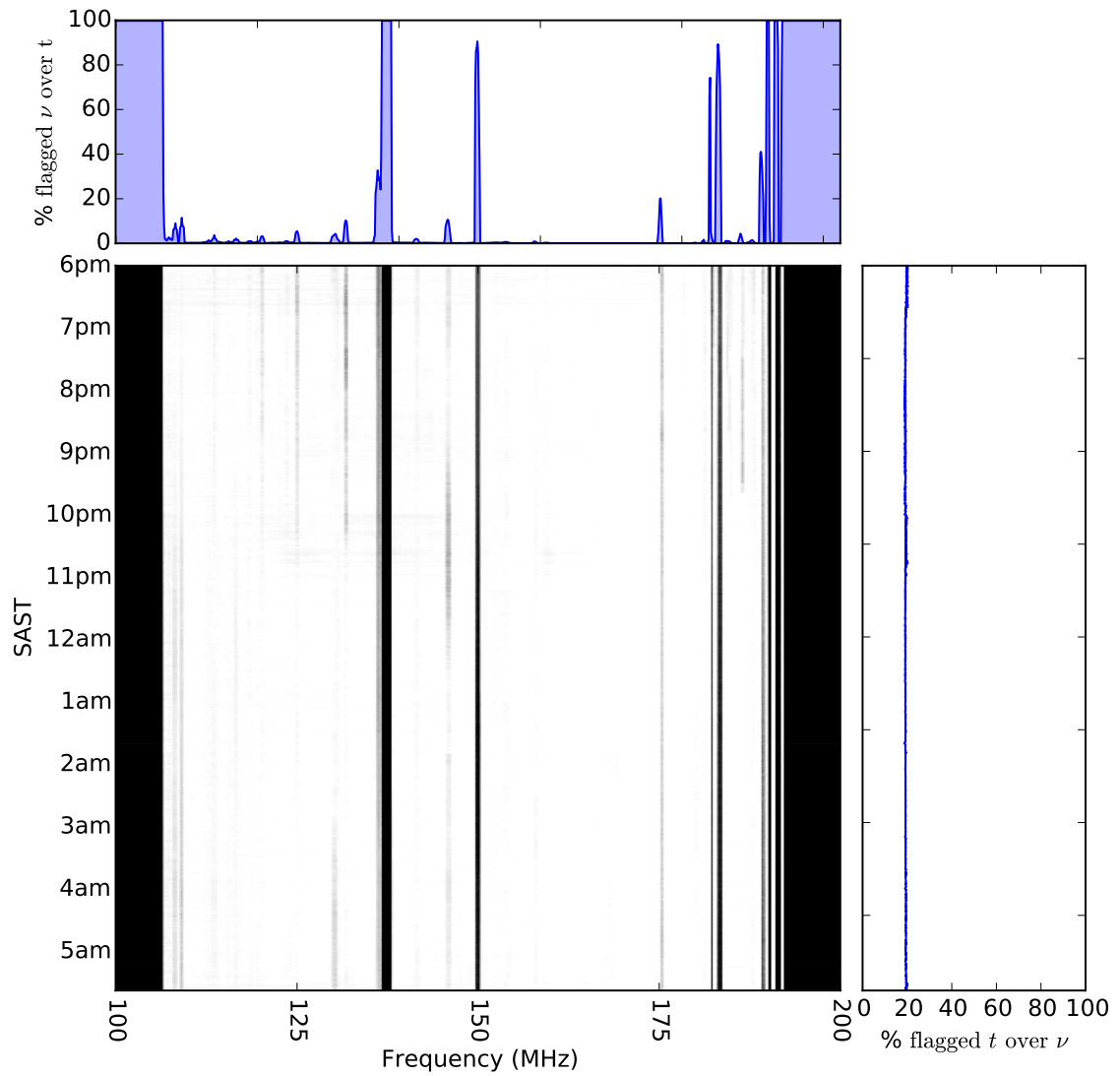


Figure 7.2: A waterfall plot of RFI flags averaged over 150 days of data. The gridding process is described in the text. Above the waterfall I show the percentage of the season each frequency is flagged, and to the right I show the percentage of frequencies that are flagged per integration.

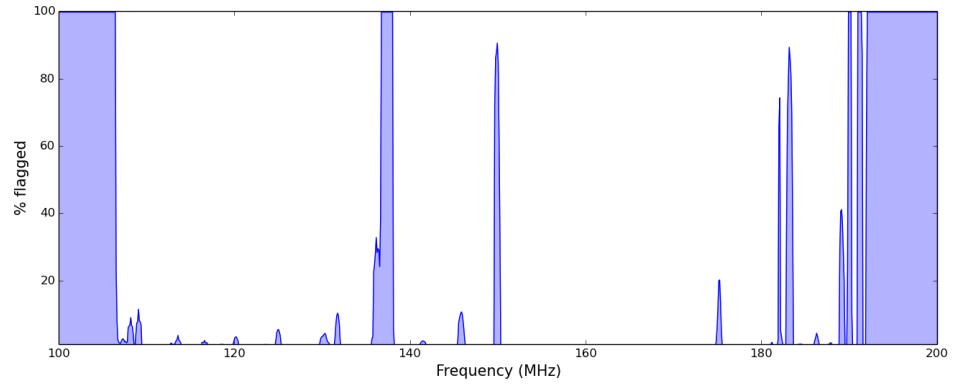


Figure 7.3: The percentage of time that each frequency was flagged over the season.

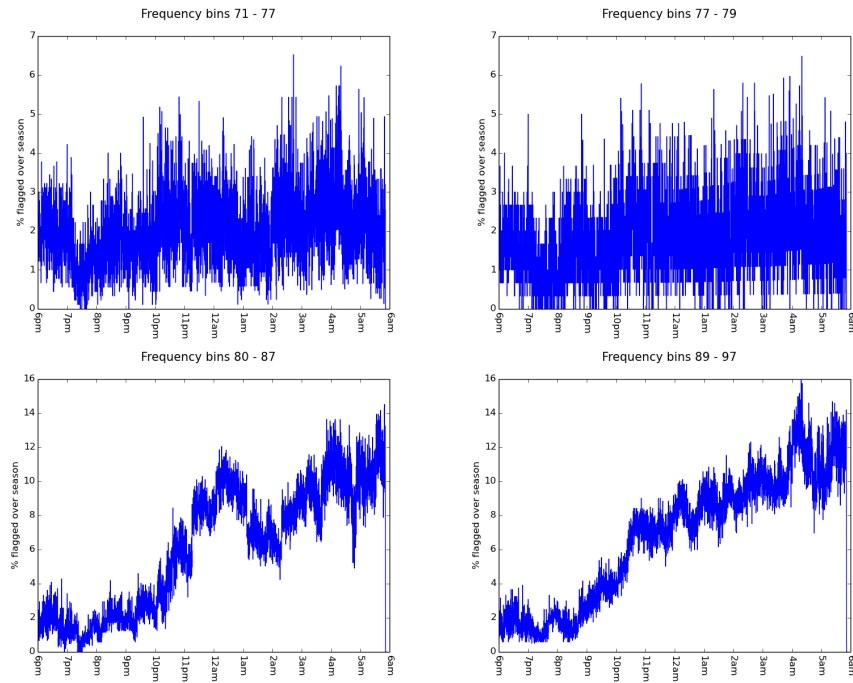


Figure 7.4: Possible FM radio contamination in the *Top*, left to right: 107.25 ± 0.25 and 107.55 ± 0.05 MHz bands, and *Bottom*, left to right: 108.1 ± 0.4 and 109 ± 0.4 MHz bands.

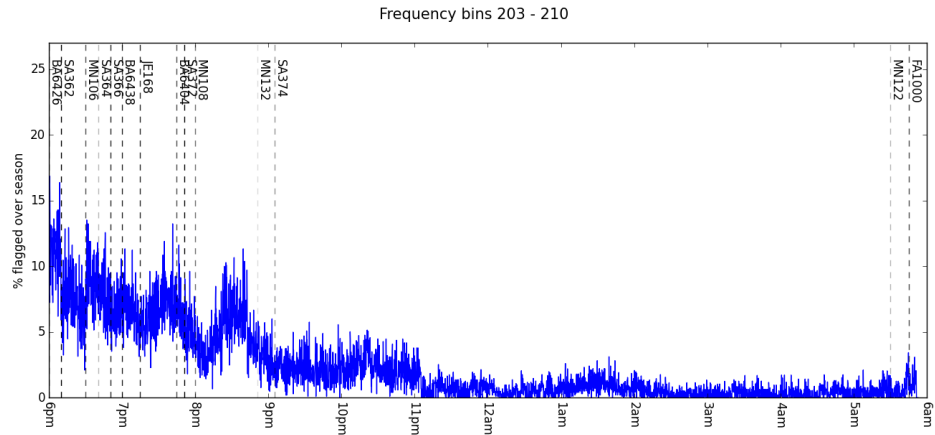


Figure 7.5: Flights from Cape Town to Johannesburg correspond to RFI in the 120.15 ± 0.35 MHz channels. Vertical dashed lines indicate a flight leaving Cape Town (flights from Johannesburg are roughly concurrent) and the flight code is listed. The transparency of a line is inversely proportional to how many days a week that flight is scheduled for. The flight is 2 to 2.5 hours long – and about 2 hours after the last flight of the day, the flags fall to background level (but notably, not always to zero).

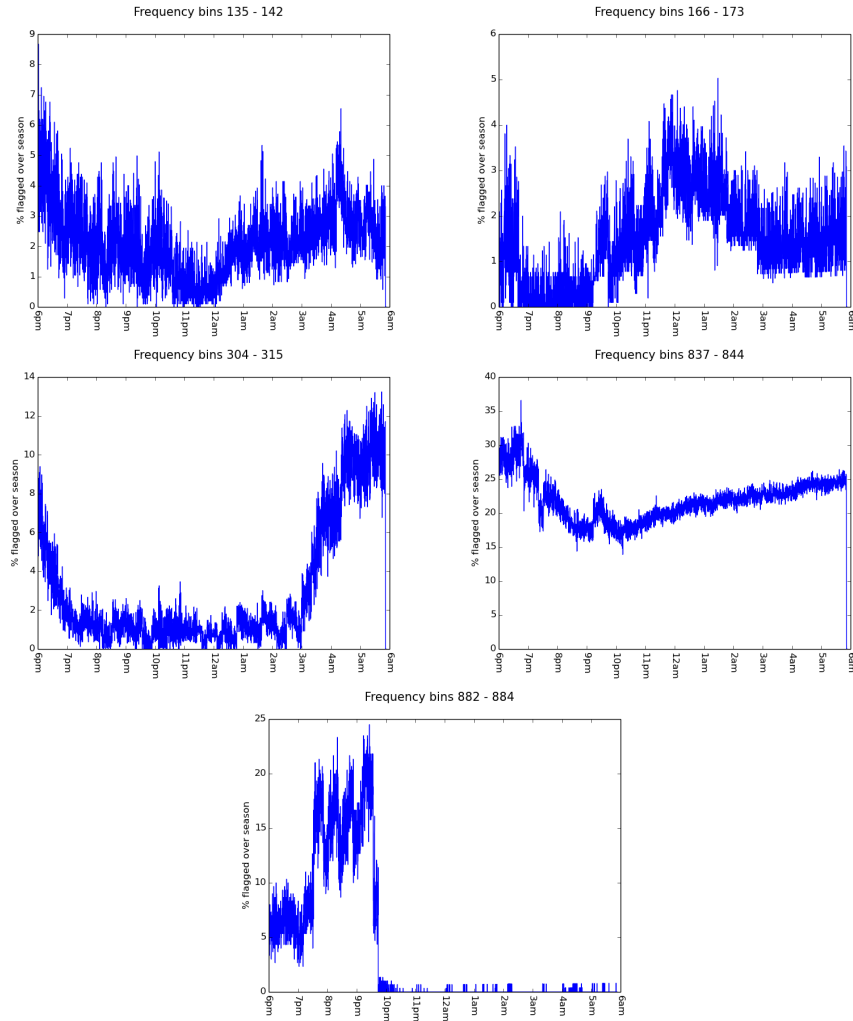


Figure 7.6: The temporal profile of the 5 RFI frequencies with unidentified causes. *Top, left to right:* 114.05 ± 0.85 and 116.55 ± 0.35 MHz. *Middle, left to right:* 130.25 ± 0.55 and 182.15 ± 0.35 MHz. *Bottom:* 186.25 ± 0.35 MHz. The 182.15 ± 0.35 MHz frequency is flagged a large amount of the time, making it our most-offending unidentified source.

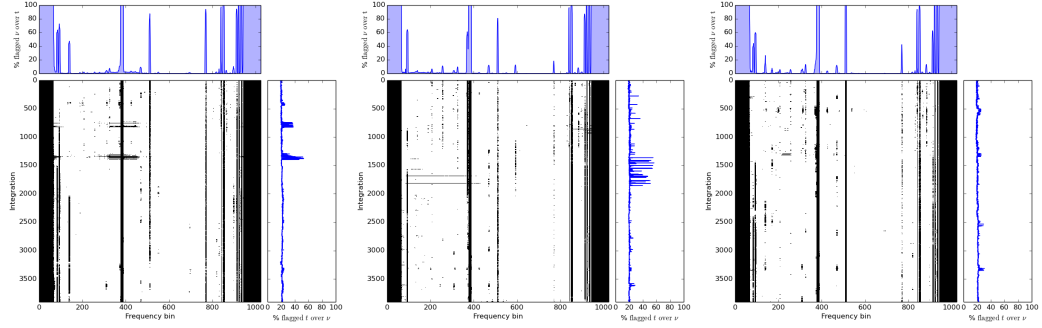


Figure 7.7: *Left to right:* waterfalls of flags for nights 2456732, 2456958 and 2457038. These three nights, along with 2456965, are $>20.2\%$ flagged; $> 2\sigma$ above the average flagging amount per night.

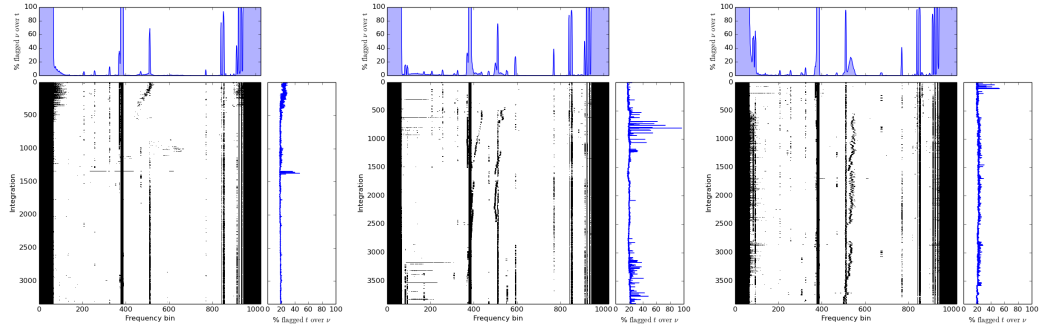


Figure 7.8: *Left to right:* waterfalls of flags for nights 2456898, 2456924 and 2456965. These three nights exhibit a strange behaviour of RFI that changes in frequency and time. JD 2456965 is by far the worst, and during this night as well as 2456898, we see a broadband ‘comb’ of flagged frequencies near the band edges

Chapter 8

Polarimetric Calibration

8.1 Redundant Calibration

8.2 Imaging Calibration

Chapter 9

The Ionosphere

Chapter 10

A view of the EoR window from the PAPER-32 imaging array

Chapter 11

A view of the EoR window from the HERA-19 commissioning array

Chapter 12

Deep integrations with PAPER-128

Part III

Expanding the potential of EoR measurements

Chapter 13

Higher-order correlation functions between the kSZ and 21cm fields during the EoR

Chapter 14

Deep Learning for 21cm Observations

Chapter 15

Conclusions

Appendices

Appendix A

Software

Software engineering and maintenance of existing codebases has been, generally speaking, historically undervalued and unappreciated (Muna et al. 2016). In this Appendix I would like to provide a brief description of the major software packages used in this work – without which, the work would not exist.

A.1 Astronomical Interferometry in Python (aipy)

The aipy software package (Parsons 2016) was developed by a team based largely at the University of California, Berkeley and led by Aaron Parsons. Developed under NSF funding for the PAPER experiment, it provides a Python API to interact with interferometric visibilities stored in the MIRIAD file format (Sault et al. 2011). It is able to efficiently query large MIRIAD files due to the API's closeness to the underlying C code. It also contains calibration, deconvolution, imaging and phasing code in Python, and interfaces with HEALPix (see Section A.5, below) as well as other astronomical Python packages.

aipy is maintained by the HERA software team, and can be found at: <https://github.com/HERA-Team/aipy>.

A.2 Astronomy in Python (astropy)

astropy is an open-source and community-developed core Python package for Astronomy, containing a host of extremely useful utility functions and objects (Astropy Collaboration et al. 2013).

A.3 Common Astronomy Software Applications (CASA)

CASA is under active development, with the primary goal of supporting the data post-processing needs of the next generation of radio telescopes. It is developed by an international consortium of scientists based at the National Radio Astronomical Observatory (NRAO), the European Southern Observatory (ESO), the National Astronomical Observatory of Japan (NAOJ), the CSIRO Australia Telescope National Facility (CSIRO/ATNF), and the Netherlands Institute for Radio Astronomy (ASTRON), under the guidance of NRAO (McMullin et al. 2007).

A.4 Deep Learning packages

Experimentation with deep learning analyses of 21 cm simulated observations took place in Keras (Chollet et al. 2015), PyTorch (Paszke et al. 2017) and Tensorflow (Abadi et al. 2016).

A.5 Hierarchical Equal Area isoLatitude Pixelization of the sphere (HEALPix)

The HEALPix software, and its Python wrapper healpy, provide a pixelization which subdivides a spherical surface into pixels which each cover the same surface area as every other pixel. Pixel centers occur on a discrete number of rings of constant latitude. This

scheme makes natively spherical measurements, such as angular power spectra and wide-field images, simple and efficient to interact with (Górski et al. 2005).

A.6 pyuvdata

pyuvdata provides a Python interface to interferometric data. It can read and write MIRIAD and UVFITS file formats, as well as read CASA measurement sets and FHD (Sullivan et al. 2012) visibility save files (Hazelton et al. 2017).

pyuvdata is maintained by the HERA software team, and can be found at: <https://github.com/HERA-Team/pyuvdata>.

A.7 The Scientific Python Ecosystem (scipy)

Many of the above tools require at least one of the many packages under the scipy ecosystem. It is truly foundational to almost any scientific analysis that takes place in Python (Jones et al. 2001).

Bibliography

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X.: 2016, *ArXiv e-prints*

Ali, Z. S., Parsons, A. R., Zheng, H., Pober, J. C., Liu, A., Aguirre, J. E., Bradley, R. F., Bernardi, G., Carilli, C. L., Cheng, C., DeBoer, D. R., Dexter, M. R., Grobbelaar, J., Horrell, J., Jacobs, D. C., Klima, P., MacMahon, D. H. E., Maree, M., Moore, D. F., Razavi, N., Stefan, I. I., Walbrugh, W. P., and Walker, A.: 2015, *ArXiv e-prints*

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A. M., Kerzendorf, W. E., Conley, A., Crighton, N., Barbary, K., Muna, D., Ferguson, H., Grollier, F., Parikh, M. M., Nair, P. H., Unther, H. M., Deil, C., Woillez, J., Conseil, S., Kramer, R., Turner, J. E. H., Singer, L., Fox, R., Weaver, B. A., Zabalza, V., Edwards, Z. I., Azalee Bostroem, K., Burke, D. J., Casey, A. R., Crawford, S. M., Dencheva, N., Ely, J., Jenness, T., Labrie, K., Lim, P. L., Pierfederici, F., Pontzen, A., Ptak, A., Refsdal, B., Servillat, M., and Streicher, O.: 2013, *A & A* **558**, A33

Authority, S. A. C. A.: 2008, *FLIGHT CALIBRATION DATES OF VOR's IN THE RE-*

- PUBLIC OF SOUTH AFRICA*, Technical Report 25.6, Aeronautical Information Circular, <http://209.203.9.244/resource%20center/AIC'S/25.6.pdf>
- Chollet, F. et al.: 2015, *Keras*, <https://github.com/fchollet/keras>
- Furlanetto, S. R., Oh, S. P., and Briggs, F. H.: 2006, *Physics Review* **433**, 181
- Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., and Bartelmann, M.: 2005, *ApJ* **622**, 759
- Hazelton, B., Beardsley, A., Pober, J., Jacobs, D., Ali, Z., and Lanman, A.: 2017, *HERA-Team/pyuvdata: Version 1.1*
- Jacobs, D. C., Pober, J. C., Parsons, A. R., Aguirre, J. E., Ali, Z. S., Bowman, J., Bradley, R. F., Carilli, C. L., DeBoer, D. R., Dexter, M. R., Gugliucci, N. E., Klima, P., Liu, A., MacMahon, D. H. E., Manley, J. R., Moore, D. F., Stefan, I. I., and Walbrugh, W. P.: 2015, *ApJ* **801**, 51
- Jones, E., Oliphant, T., Peterson, P., et al.: 2001, *SciPy: Open source scientific tools for Python*, [Online; accessed <today>]
- McMullin, J. P., Waters, B., Schiebel, D., Young, W., and Golap, K.: 2007, in R. A. Shaw, F. Hill, and D. J. Bell (eds.), *Astronomical Data Analysis Software and Systems XVI*, 376, p. 127, ASP, San Francisco, CA
- Muna, D., Alexander, M., Allen, A., Ashley, R., Asmus, D., Azzollini, R., Bannister, M., Beaton, R., Benson, A., Berriman, G. B., Bilicki, M., Boyce, P., Bridge, J., Cami, J., Cangi, E., Chen, X., Christiny, N., Clark, C., Collins, M., Comparat, J., Cook, N., Croton, D., Delberth Davids, I., Depagne, É., Donor, J., dos Santos, L. A., Douglas, S., Du, A., Durbin, M., Erb, D., Faes, D., Fernández-Trincado, J. G., Foley, A., Fotopoulou, S., Frimann, S., Frinchaboy, P., Garcia-Dias, R., Gawryszczak, A., George, E., Gonzalez, S., Gordon, K., Gorgone, N., Gosmeyer, C., Grasha, K., Greenfield, P., Grellmann, R., Guillochon, J., Gurwell, M., Haas, M., Hagen, A., Haggard, D., Haines,

T., Hall, P., Hellwing, W., Herenz, E. C., Hinton, S., Hlozek, R., Hoffman, J., Holman, D., Holwerda, B. W., Horton, A., Hummels, C., Jacobs, D., Juel Jensen, J., Jones, D., Karick, A., Kelley, L., Kenworthy, M., Kitchener, B., Klaes, D., Kohn, S., Konorski, P., Krawczyk, C., Kuehn, K., Kuutma, T., Lam, M. T., Lane, R., Liske, J., Lopez-Camara, D., Mack, K., Mangham, S., Mao, Q., Marsh, D. J. E., Mateu, C., Maurin, L., McCormac, J., Momcheva, I., Monteiro, H., Mueller, M., Munoz, R., Naidu, R., Nelson, N., Nitschelm, C., North, C., Nunez-Iglesias, J., Ogaz, S., Owen, R., Parejko, J., Patrício, V., Pepper, J., Perrin, M., Pickering, T., Piscionere, J., Pogge, R., Poleski, R., Pourtsidou, A., Price-Whelan, A. M., Rawls, M. L., Read, S., Rees, G., Rein, H., Rice, T., Riemer-Sørensen, S., Rusomarov, N., Sanchez, S. F., Santander-García, M., Sarid, G., Schoenell, W., Scholz, A., Schuhmann, R. L., Schuster, W., Scicluna, P., Seidel, M., Shao, L., Sharma, P., Shulevski, A., Shupe, D., Sifón, C., Simmons, B., Sinha, M., Skillen, I., Soergel, B., Spriggs, T., Srinivasan, S., Stevens, A., Streicher, O., Suchyta, E., Tan, J., Telford, O. G., Thomas, R., Tonini, C., Tremblay, G., Tuttle, S., Urrutia, T., Vaughan, S., Verdugo, M., Wagner, A., Walawender, J., Wetzel, A., Willett, K., Williams, P. K. G., Yang, G., Zhu, G., and Zonca, A.: 2016, *ArXiv e-prints*

Parsons, A.: 2016, *AIPY: Astronomical Interferometry in PYthon*, Astrophysics Source Code Library

Parsons, A. R. and Backer, D. C.: 2009, *AJ* **138**, 219

Parsons, A. R., Liu, A., Aguirre, J. E., Ali, Z. S., Bradley, R. F., Carilli, C. L., DeBoer, D. R., Dexter, M. R., Gugliucci, N. E., Jacobs, D. C., Klima, P., MacMahon, D. H. E., Manley, J. R., Moore, D. F., Pober, J. C., Stefan, I. I., and Walbrugh, W. P.: 2014, *ApJ* **788**, 106

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A.: 2017

- Sault, R. J., Teuben, P., and Wright, M. C. H.: 2011, *MIRIAD: Multi-channel Image Reconstruction, Image Analysis, and Display*, Astrophysics Source Code Library
- Staatskoerant: 2008, *South African Table of Frequency Allocations*, Technical Report 31264, ICASA, http://thornton.co.za/resources/31264_890_complete-1.pdf
- Sullivan, I. S., Morales, M. F., Hazelton, B. J., Arcus, W., Barnes, D., Bernardi, G., Briggs, F. H., Bowman, J. D., Bunton, J. D., Cappallo, R. J., Corey, B. E., Deshpande, A., deSouza, L., Emrich, D., Gaensler, B. M., Goeke, R., Greenhill, L. J., Herne, D., Hewitt, J. N., Johnston-Hollitt, M., Kaplan, D. L., Kasper, J. C., Kincaid, B. B., Koenig, R., Kratzenberg, E., Lonsdale, C. J., Lynch, M. J., McWhirter, S. R., Mitchell, D. A., Morgan, E., Oberoi, D., Ord, S. M., Pathikulangara, J., Prabu, T., Remillard, R. A., Rogers, A. E. E., Roshi, A., Salah, J. E., Sault, R. J., Udaya Shankar, N., Srivani, K. S., Stevens, J., Subrahmanyan, R., Tingay, S. J., Wayth, R. B., Waterson, M., Webster, R. L., Whitney, A. R., Williams, A., Williams, C. L., and Wyithe, J. S. B.: 2012, *ApJ* **759**, 17