

SDR Pathfinder for Understanding Transient and Noise-level Interference in the Karoo (SPUTNIK)

Alec T. Josaitis*, Eloy de Lera Acedo*, David R. DeBoer†

*Astrophysics Group - Cavendish Laboratory, University of Cambridge, JJ Thompson Avenue, Cambridge CB3 0HE

†Astronomy Department, University of California - Berkeley, 425 Campbell Hall, University Dr., Berkeley 94720, USA

Abstract—The SDR Pathfinder for Understanding Transient and Noise-level Interference in the Karoo (SPUTNIK) is presented. We describe how a low-cost RFI monitoring system, using solely off-the-shelf components, directly contributes to the analysis efforts of a precision 21cm cosmology instrument. A SPUTNIK system overview is provided, as well as a generalized software-defined radio (SDR) internal calibration technique to achieve wideband, ± 1.5 dBm-level accuracy and a measured dynamic range of ≥ 40 dB.

I. INTRODUCTION

Efforts are underway to construct low-frequency radio telescopes, such as the Hydrogen Epoch of Reionization Array (HERA) [1] and the Square Kilometre Array (SKA) [2], to observe the red-shifted 21cm line of neutral hydrogen. These new experiments intend to study the evolution of the IGM before, during, and after reionization using observations of fluctuations in the 21cm power spectrum. Such precision cosmology experiments require multi-prong data analysis and RFI-flagging techniques to claim detection of the cosmological 21cm signal from beneath a background of smooth astrophysical foregrounds and even stronger, highly chromatic terrestrial radio frequency interference (RFI). We present here a low-cost, single-antenna RFI monitoring system, using solely off-the-shelf digital signal processing (DSP) hardware, which satisfies one of the HERA collaboration's three primary forms of RFI flagging. This paper describes the RFI monitor's role in a global flagging strategy for a 21cm experiment, and details the system design and calibration.

II. THE IMPACT OF RFI MONITORING ON 21CM DATA ANALYSIS

A. Delay Spectrum Analysis and the Consequences of RFI

HERA is a 350-element interferometer, built in South Africa's Karoo Radio Astronomy Reserve specifically to characterize the evolution of the 21cm signal from cosmic dawn ($z \approx 30$) through the full reionization of the IGM ($z \approx 6$). Each element of the interferometer is a 14-m parabolic dish, capable of observing from 50-250 MHz. Active electronic components are currently being installed to the Vivaldi-style dish feeds; the most notable of these electronics being the Front-End Modules (FEMs) and Post-Amplification Modules (PAMs) [3]. HERA seeks to measure the evolution of the global 21cm signal's power spectrum. Fluctuations are expected to have been introduced in the power spectrum as the Universe's first stars and black holes heated and ionized the IGM [1]. These fluctuations are scale-dependent, with most of the power



Fig. 1. A satellite image of the HERA interometer, which is currently being commissioned. SPUTNIK serves as one of the three primary approaches for understanding the RFI environment at the HERA site.

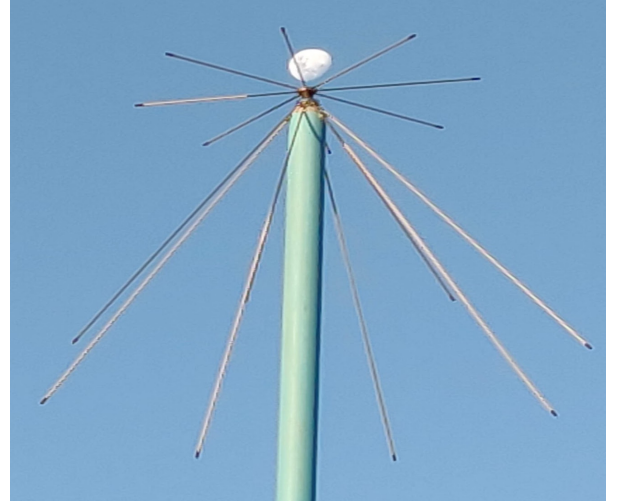


Fig. 2. A photo of the SPUTNIK discone antenna at the HERA site in the Karoo Radio Astronomy Reserve.

at angles characteristic of ionized regions of the IGM – which are predicted to be tens of arc-minutes across [4]. The dishes of HERA, then, must produce beams smaller than this characteristic size, so that the characteristic fluctuations are not averaged across the sky, as is done in a global 21cm experiment. Imperfect beam patterns, along with imperfect

instrument calibration, will produce chromatic effects in the otherwise smooth foreground spectrum. HERA surmounts this issue by designing its hexagonal array with redundant baselines, which improve instrument calibration and sensitivity to the EoR power spectrum when the data is analyzed using the ‘delay spectrum’ technique [1]. Chromatic effects contaminate a different subset of delay space (the Fourier transform of the spectral domain) than the expected “window” of EoR signal. This contamination subspace is called the “wedge”, and has been extensively studied by the HERA collaboration [1]. The wedge is a direct consequence of the chromatic response of HERA’s interferometer.

RFI adds spectral structure in an otherwise smooth polarized foreground emission; miscalibration of RFI-contaminated foregrounds pollute subsets of the spectral domain expected to contain the EoR window. One of the key reasons that the wedge approach to HERA’s 21cm analysis is thought to work is that it effectively discriminates between spectrally-smooth foregrounds and an EoR signal with spectral structure – the structure being caused by the cosmic dawn and the gradual reionization of the IGM [1]. The spectral structure of RFI, if not removed, could be considered as part of the spectral structure of the EoR signal itself. Such a spectral distortion, along with other consequences of not accounting for RFI, are summarized below: .

- *Introduction of Non-Linearity into Analog Signal Chains:* RFI, which can be many orders of magnitude higher in power than astrophysical signals, can drive active components in the analog signal chain (e.g. amplifiers) out of their range of linear operation, thus corrupting the analog signal by making the output unpredictable or systematically incorrect.
- *Over-Saturation of Analog to Digital Conversion (ADC):* Even if the analog system chain is sufficiently robust so as to not be driven non-linear from RFI, the power of the RFI may be high enough so as to saturate the bit resolution of the signal chain’s ADC, thus corrupting the signal.
- *Contamination in the Spectral Domain:* Analog to digitally-converted channels which are not saturated may be contaminated so frequently in time that it is not practical to use the corresponding, scientifically-relevant frequency during data analysis.
- *Contamination of the EoR Window:* The spectral structure of the RFI, if not removed, could manifest itself outside of the wedge in delay space and into the expected EoR window, thus being considered part of the structure of the EoR signal itself and distorting the scientific product of the experiment.

B. Removing RFI from HERA Data

HERA takes a three-pronged approach to removing RFI from its data, comparing the RFI found by various subsets of the entire instrument with an independent RFI monitor. First, data products created by the HERA instrument (e.g. raw and absolutely calibrated visibilities from auto and cross

correlations of interferometric elements), are processed by the watershed XRFI algorithm, which uses modified z-score thresholds to identify outliers in power as a function of both time and frequency [5]. The data input into XRFI comes from the interferometer as it was intended to operate – all subsystems on, the array observing as designed.

Second, the active electronics of the analog system chain (the chain of powered components interfacing with voltage coming from HERA’s interferometric elements prior to the signal being digitally converted) are tested to see whether they are injecting chromatic features into the HERA digital (post-analog) signal chain, which would skew RFI flagging statistics. In this second technique, the FEM is removed from a reference Vivaldi feed and replaced with a passive balun, which is used to direct the balanced electromagnetic radiation from the feed onto an unbalanced coaxial cable, thus removing the effects of the FEM’s active amplifiers and “RF over Fiber” (RFoF) transition from coaxial cable to optical fiber. Data collected from this reference, FEM-less element is then processed by XRFI.

Third, an independent RFI monitor, using none of the same hardware or antenna as HERA, is deployed at the HERA site and its data products are also flagged for RFI. This independent RFI monitor is the SDR Pathfinder for Understanding Transient and Noise-level Interference in the Karoo (SPUTNIK). The beam pattern of the SPUTNIK discone antenna, described in the next section, is significantly more sensitive to horizon-level RFI than HERA’s Vivaldi feed, which has an azimuth-centered gain [3]. Sidelobe studies of the Vivaldi feed are ongoing, but the most recent analysis demonstrates that mutual coupling between Vivaldi feeds and neighboring element dishes can produce high sidelobes at the horizon level [3]. These simulations further suggest that reflections of horizon emissions from the furthest away elements in the HERA array considerably spread the interferometer’s system time response, resulting in a direction-dependent chromatic response of the array [3]. Horizon-dominated RFI detected by HERA, even if at low-level, can impede the Collaboration’s field-specific (not simulated) characterization of Vivaldi side lobe chromaticity. Having a smoothly changing, horizon-centered beam pattern in an independent RFI monitor improves our ability to decouple the Vivaldi sidelobe chromaticity from the dynamic RFI environment found at the HERA site.

III. THE SPUTNIK SYSTEM

A. System Overview

Fig. 3 provides an overview of SPUTNIK. A passive discone antenna sits at the HERA site, outside the array, at a height comparable to the array’s Vivaldi feeds. The signal is carried via LMR-195 grade SMA cable into an RF-tight enclosure, amplified with a flat frequency response via a Minicircuits ZKL-1R5+ coaxial amplifier, and then input into an Ettus N200 SDR¹. Calibrated and uncalibrated Fourier spectra from the SDR are stored on a unix-based networked computer in

¹<https://www.ettus.com>

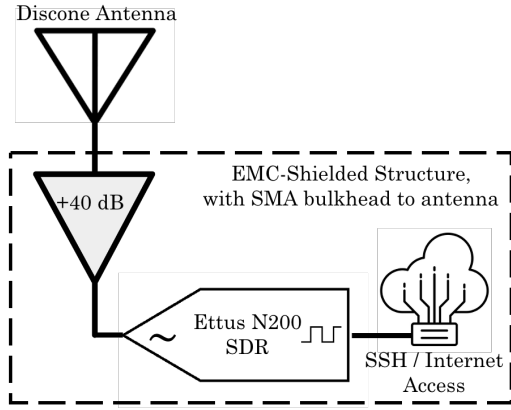


Fig. 3. A system-level overview of SPUTNIK. A discone antenna connects to an electromagnetically compliant (EMC), shielded preamplifier and SDR. The SDR outputs calibrated and uncalibrated Fourier spectra and shares these files with the HERA librarian via a networked computer. The SDR may be remotely reprogrammed via an ssh connection.



Fig. 4. A photo of the SPUTNIK discone and electromagnetically compliant (EMC), shielded structure in the Karoo. The discone is mounted atop, but electrically-insulated from, the structure.

the RF enclosure, and uploaded to the HERA Librarian [6] via an SSH- tunneled rsync connection.

Data shared by SPUTNIK and the passive Vivaldi feed chain described in Section (II.B) are structured in the RF Interference Data System (RIDS) format, and fully-compatible with RIDS analysis scripts [7]. Daily RFI statistics are calculated via RIDS and made viewable to the collaboration via an online commissioning dashboard. The SDR may be reprogrammed remotely via SSH. The purpose of the RF-tight enclosure housing the SDR and preamplifier is two-fold. First, to avoid generating RFI on site which could damage the array; second, to avoid detecting RFI generated by the SDR and preamplifier. Currently, the antenna is mounted above and outside the RF enclosure, and has been tested to be electromagnetically compliant with site, ensuring that spurious signals from the SDR and preamplifier do not couple to the SMA cable and radiate from the antenna.

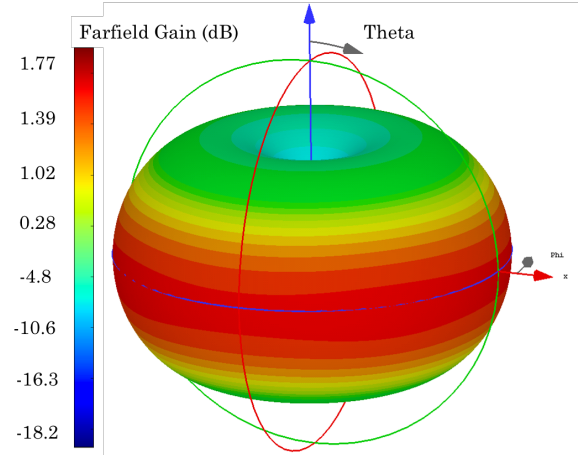


Fig. 5. Farfield gain simulation of the SPUTNIK discone antenna at 100MHz. The SPUTNIK discone response is significantly smoother at the horizon than HERA's Vivaldi feed (see Fig. 6, allowing for an achromatic detection of RFI. This improves our ability to decouple the Vivaldi sidelobe chromaticity from the dynamic RFI environment found at the HERA site.

B. The Discone Antenna

We use CST Microwave Studio's time-domain solver to simulate the SPUTNIK discone antenna, both with and without the EMC-shielded structure contaminating the beam. Fig. 5 contains the farfield gain simulation at 100 MHz without the EMC structure. The SPUTNIK discone has a significantly smoother beam response at the horizon than the HERA Vivaldi feed across the entire measurement band. A simulation of the discone's S11 reflection coefficient is plotted in Fig. 7 with and without the EMC structure, for the current SPUTNIK configuration. The structure has a negligible effect on the antenna impedance in the range of the simulation.

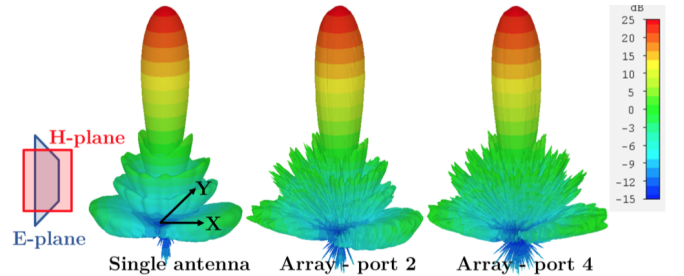


Fig. 6. Gain patterns at 150 MHz (amplitude in dB) of the HERA Vivaldi Feed with and without mutual coupling [3]

C. The Software-Defined Radio

SPUTNIK uses an SDR-based, mobile, sweep- spectrum analyzer. Differing from conventional radios which have purpose-built hardware components in an analog signal processing chain, SDRs digitally process a signal by programming the effects of analog components, such as bandpass filters and demodulators, onto a microprocessor or FPGA by means of embedded system software. SPUTNIK uses an Ettus Research N200 Universal Software Radio Peripheral (USRP) with

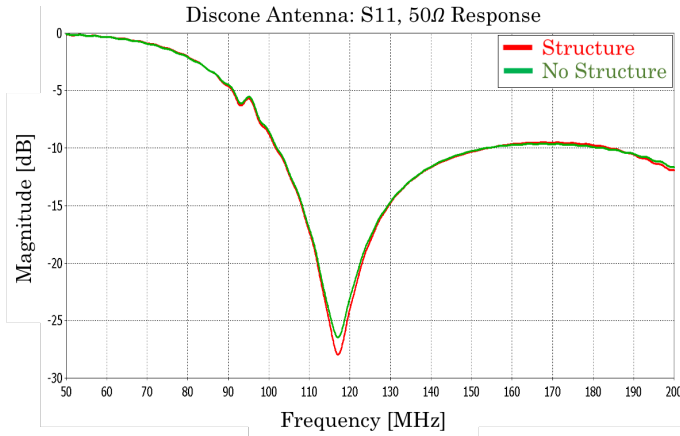


Fig. 7. S11 reflection coefficient of the SPUNTIK discone antenna, simulated using CST, with and without the EM- shielded structure.

UBX40 daughterboard. We communicate with and control the Ettus SDR using Gnuradio, an open-source development toolkit, which provides “blocks” of signal processing software to implement SDRs and signal-processing systems.

IV. ABSOLUTE CALIBRATION OF SDR-BASED SWEEP SPECTRUM ANALYZERS

A. Technique Overview and Application to RFI Monitoring

This is a generalized overview of how to absolutely, internally calibrate SDRs, a useful step for commissioning RFI monitoring systems at a radio astronomy site, in our case HERA, and in the HF, VHF, and UHF frequency range of 20-500MHz. The technique has been used to calibrate three separate systems to within ± 1.5 dBm, each system being based on a different make and model of SDR (HackRF One², Airspy R2³, Ettus Research N200 USRP). SDRs may be programmed to behave like a spectrum analyzer, tuned to a given center frequency f_c , where spectra with range $[f_c - \frac{f_{samp}}{2}, f_c + \frac{f_{samp}}{2}]$ accumulate for some integration time and are then averaged to become one ‘measurement’ spectrum. Depending on the ADC sample rate and bandpass filters used in an SDR, the bandwidth of a single spectrum may be smaller than the science requirements of the spectrum analyzer’s application (in our case, site RFI in the range 20-500 MHz). Thus, one must program a sweep spectrum analyzer, where multiple measurement-spectra are created by re-tuning to different values of f_c , and making a single, averaged measurement spectra (also of width f_{samp}) centered around the f_c . To avoid roll-off effects from the bandpass filters inside the Ettus N200 and UBX 40 daughterboard, we ignore 2.5MHz of data from each side of a measured spectra, thus reducing our effective sampling rate by 5MHz. Note that these 5MHz frequency bands are not skipped in the data measurement, rather this band reduction necessitates an increase in the total number of

required, recorded spectra in order to sweep across the entire 20-500 MHz band.

B. Absolutely Calibrate a Subset of Channels to Self-Consistently Calibrate All Channels

SDR-based spectrum analyzers bin their measurements in discrete Fast Fourier Transform (FFT) channels. The amplitude-squared of the FFT output represents power in a channel, in arbitrary units of the SDR. These arbitrary powers must be converted into physically-meaningful units, in our case dBm. Assuming that every channel behaves linearly in power, every channel of the FFT would ideally be calibrated separately. In principle, one could use a signal generator to inject a single-frequency (tone) signal of known power into just one frequency channel of the SDR and measure the arbitrary FFT power value. Calibrating any channel at one power, in a given input power range corresponding to a linear SDR response, is sufficient to calibrate the entire channel’s linear response. One would repeat this method for every channel. This explicit, channel-by-channel calibration can be cumbersome for RFI monitoring systems operating across a wide frequency range, in our case, 2048 channels per 25 MHz in ranges between 20-500 MHz. If we not only assume that all FFT channels used by the SDR are linear in power, but also know to within an acceptable tolerance (for SPUNTIK, $> \pm 1.5$ dBm between channels) the power across channels of a continuous source, and absolutely calibrate at least one point per spectra (e.g. per 2048 channels), then we can calibrate all FFT channels to within the acceptable tolerance. The overall calibration method will now be described in subsection detail. Applying this calibration to an uncalibrated SDR noise source response (IV.D), we recover the reference noise measurement (IV.C) to within ± 1.5 dBm across the entire sweep spectrum range (see Appendix I). Signals across this band may be recovered with this precision down to at least -85 dBm, the minimum settable power of the Agilent 83752A Synthesized Sweeper we used to inject signals⁴. Considering the reference calibration measurement (IV.C) is made upwards of -45dBm, the dynamic range of the internal calibration is at least 40 dB. This is a tested and conservative bound on dynamic range, for we have also calibrated spectra in the lab using reference power levels ≥ 5 dBm over that presented in (IV.C).

C. Reference Measurement of Noise Source Response

A reference, wideband noise source is measured across the entire measurement bandwidth (20 – 500 MHz) which is to be calibrated. We measure the noise source using a Keysight N5247A network analyzer⁵. The resolution bandwidth (RBW) of the Keysight is adjusted such that the power measured by the network analyzer, across the entire measurement bandwidth, was in the range (-55, -45) dBm, corresponding to ≈ 5 dBm higher than the expected maximum RFI power at the HERA site. Experience suggests it is valuable to calibrate SDRs at a higher power than what is expected to be measured,

²<https://greatscottgadgets.com/hackrf/one>

³<https://airspy.com/airspy-r2/>

⁴<https://www.agilent.com>

⁵<https://www.keysight.com>

for higher power calibrations tend to have a smaller fractional error, making measurements of lower power (e.g. RFI) more precise. In order to ensure a smooth power change across frequency, to within a ± 1.5 dBm tolerance between neighboring-channel samples, we use n_{sample} number of samples in our Keysight reference measurement to minimize the quantity $\frac{f_{stop} - f_{start}}{n_{sample} * RBW}$, while maintaining a sufficiently large RBW so as to maintain the (-55,-45 dBm) output power requirement.

D. Uncalibrated SDR, Noise Source Response

The same wideband noise source used to make the absolute reference measurement in (IV.C) is connected to the input of our SDR, and a sweep-spectrum measurement is recorded across the entire measurement frequency bandwidth. While we do assume that any individual channel of an SDR is linear, we do not assume a flat input gain response across neighboring channels. Thus, one must ensure the same SDR channels always correspond to the same physical frequencies. We ensure this by fixing our SDR programmable settings, such as f_{start} , f_{stop} , and the SDR sample rate for the remainder of our SDR calibration.

E. Calibrating SDR Reference Channels

For at least one frequency in each spectra of the SDR sweep spectrum range, we use an Agilent 83752A Synthesized Sweeper to inject a given (frequency, power) pair recorded previously in the absolute noise source reference measurement (Section IV.C). We measure the SDR responses to these reference inputs, in arbitrary SDR units.

F. One-time, Per-Channel SDR Coefficient Calculation

For each channel in each spectrum, we calculate the ratio of the SDR response of the wideband noise power (IV.D) to the response at the reference channel in that spectra (IV.E). For example, if the reference channel and response for a given spectra were (n_{ref}, P_{ref}) , then we wish to calculate $\frac{P_i}{P_{ref}}$, $i \in [0, Length_{FFT})$. Note that these per-channel calibration coefficients are functions of SDR sweep parameters, which were fixed in (IV.D).

SPUTNIK System Specifications	
Frequency Range	20-500 MHz, 12.2 kHz RBW
Measured Noise Floor	-85 dBm internal, see (IV.B)
Precision	± 1.5 dBm
Dynamic Range	≥ 40 dB

V. EXTERNAL, FIELD-BASED CALIBRATION

A. Methodology

With the SDR internally calibrated, SPUTNIK must also be externally calibrated to be commissioned in the Karoo. The most important external factors to consider are cable losses in the system and the renormalization of the analog system chain impedance to the discone antenna. Both factors are discussed below. Appendix II summarizes the different stages of external calibration hereunder described.

B. Cable Losses

There are three sources of attenuation due to cable loss in SPUTNIK, all of which are LMR-195 grade SMA cables. First, a 30' cable connects the discone antenna to the exterior bulkhead of the EMC-shielded structure. Second, a 2' cable connects the interior bulkhead to the ZKL-1R5+ amplifier. Third, a 0.5' cable connects the amplifier to the SDR. All cable losses are considered during calibration.

C. Renormalization of Antenna System Impedance

The only active component in the SPUTNIK analog system chain, between the discone antenna output and SDR input, is the ZKL-1R5+ amplifier. The only passive components are the LMR-195 grade SMA cables described in (V.B). To calculate power measured at the antenna, the S-parameters of the amplifier must be renormalized from the factory-provided 50 Ω reference measurement to the measured antenna impedance. In the field, we find the antenna to be poorly matched to the amplifier below the FM band. For this reason, we use the simulated, rather than a field-measured, S11 to calculate the External Calibration Solution reported in Appendix II.

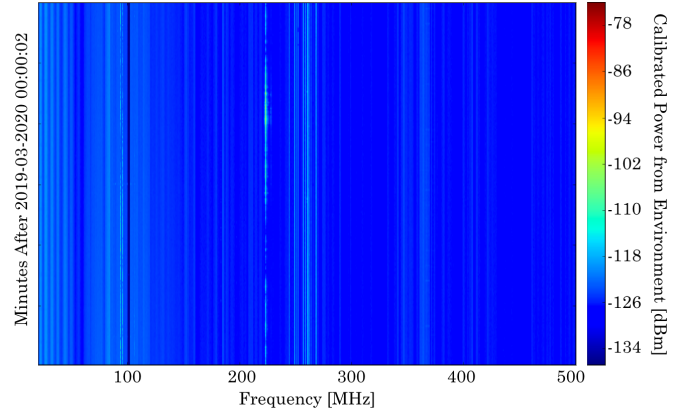


Fig. 8. Example internally and externally calibrated waterfall, one of the final data products of SPUTNIK, which is compatible with RIDS.

VI. CONCLUSION

SPUTNIK demonstrates how off-the-shelf SDR technology may directly contribute to the overall RFI monitoring strategy of a medium or large scale precision cosmology instrument. By having a horizon-centered antenna beam pattern, this independent RFI monitor can characterize emissions which appear only in the sidelobes of HERA's mutually-coupled, Vivaldi feed, interferometric elements. According to simulation, the HERA system's time response is spread considerably due to horizon emissions from the furthest away elements in the HERA array, resulting in a direction-dependent chromatic response of the array. By detecting and flagging low-level horizon RFI, SPUTNIK can improve the characterization of the HERA array's time response. SPUTNIK's internal calibration technique has been used to achieve ± 1.5 dBm accuracy over a wideband sweep spectrometer. The remote, RF-shielded

system can remotely store data and be reprogrammed via an SSH-tunneled rsync connection.

VII. ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant Nos. 1636646 and 1836019 and institutional support from the HERA collaboration partners. This research is funded in part by the Gordon and Betty Moore Foundation. HERA is hosted by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Innovation.

REFERENCES

- [1] D. R. DeBoer, A. R. Parsons *et al.*, “Hydrogen Epoch of Reionization Array (HERA),” , vol. 129, no. 974, p. 045001, Apr. 2017.
- [2] G. Mellema, L. V. E. Koopmans *et al.*, “Reionization and the Cosmic Dawn with the Square Kilometre Array,” *Experimental Astronomy*, vol. 36, no. 1-2, pp. 235–318, Aug. 2013.
- [3] N. Fagnoni, E. de Lera Acedo *et al.*, “Understanding the HERA Phase I receiver system with simulations and its impact on the detectability of the EoR delay power spectrum,” , Oct. 2020.
- [4] A. Liu, J. R. Pritchard *et al.*, “Global 21 cm signal experiments: A designer’s guide,” , vol. 87, no. 4, p. 043002, Feb. 2013.
- [5] J. R. Kerrigan, J. C. Pober *et al.*, “Improved 21 cm Epoch of Reionization Power Spectrum Measurements with a Hybrid Foreground Subtraction and Avoidance Technique,” , vol. 864, no. 2, Sep. 2018.
- [6] P. Williams and P. L. Plante, “HERA Librarian,” 2016. [Online]. Available: <https://github.com/HERA-Team/librarian>
- [7] D. DeBoer, “RF Interference Data System (RIDS),” 2018. [Online]. Available: <https://github.com/HERA-Team/rids>

APPENDIX I

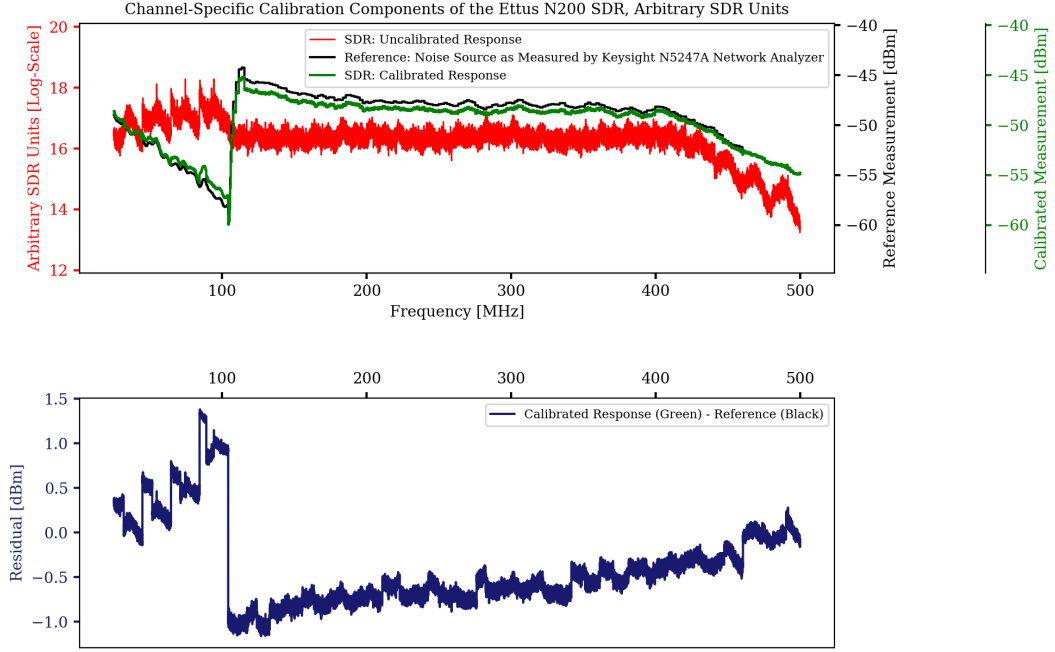


Fig. 9. A plot summarizing the SDR internal calibration method described in Section IV. A noise source, with a power output *approx* 5 dBm higher than what is expected to be measured in the Karoo is measured by a reference network analyzer across the entire frequency range (see Section IV.C, plotted in black). The same noise source is then injected into an uncalibrated SDR (Section IV.D); the response is recorded (red). A subset of the reference data (IV.C) is then used to calibrate each spectrum in the SDR sweep range (IV.E). Applying this calibration (IV.F) to the raw SDR noise response, we recover the reference noise measurement (IV.C) to within ± 1.5 dBm across the entire sweep range (blue). Note that the Ettus N200 SDR has two substantively different, internal, modes of operation: one below and one above $f \approx 110$ MHz. To our knowledge, this is intrinsic to the Ettus N200 SDR; a user cannot force one mode of operation to be used both above and below the threshold frequency. Nevertheless, both modes of operation can be calibrated to within ± 1.5 dBm.

APPENDIX II

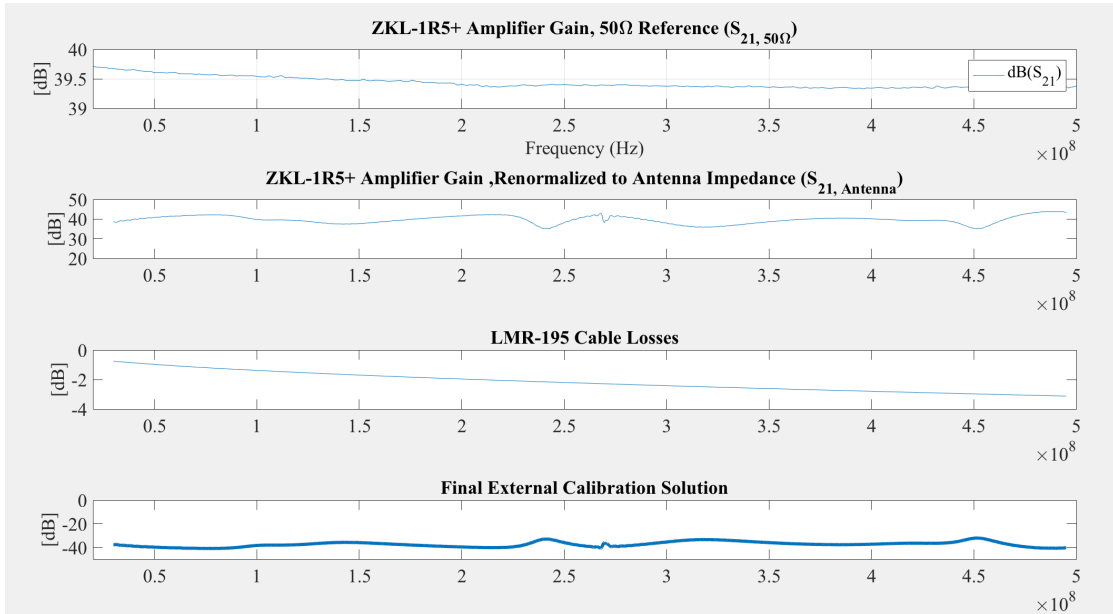


Fig. 10. An overview of the external, field-specific calibration of SPUTNIK. See Section V for details regarding impedance renormalization and cable losses.