Low Noise at Low Cost for Large Radio Astronomy Arrays

Thesis by
Kiran Arik Shila

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

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2025
Defended May 15, 2025

Kiran Arik Shila
ORCID: 0000-0003-4652-7038

I come with empty hands and the desire to unbuild walls

URSULA K. LE GUIN
THE DISPOSSESSED

ACKNOWLEDGEMENTS

In the winter of 2018, I found myself at a crossroads in my career. I was finishing my Bachelor's degree, had completed several internships and contracting jobs, and had done a bit of research at school. However, I was unsure about my next steps, whether to pursue something more academic, whether to focus on hardware or software, or even what I truly wanted to work on. Around that time, I was shown a posting for a summer job at Caltech, working with Sandy Weinreb on radio astronomy. I knew almost nothing about the field, but I seemed to have the "hands-on" radio experience Sandy was looking for, so I emailed him. Less than five minutes later, close to midnight in California, I received a reply asking if I could FaceTime him. We ended up talking for over two hours. That summer, I worked with Sandy and the team on hardware that would become part of the DSA-10 radio telescope, and I had more fun than I ever could have imagined. I would like to first and foremost thank Sandy for introducing me to this field and for providing invaluable mentorship over these years, forever altering the trajectory of my career.

A few years later, after completing my Master's degree, I was accepted to return to Caltech to pursue a PhD. Unfortunately, this was in the fall of 2020, at the height of the COVID-19 pandemic. As a result, I had to complete my first year of studies online, remotely from Florida, one of the most challenging educational experiences of my life. However, many of us quickly found a sense of community online with others around the world who were going through the same thing. Through this, I connected with the people who have since become some of my closest friends. Thank you to David, Elijah, Emma, Hannah, Ian, Isaac, Kyle, Nick, Quinn, Rhiannon, Sarah, Saren, Sasha, Steven, Talya, and Yoonsoo for including me in the world's smartest clown car.

Upon returning to Caltech, I was introduced to Professors Gregg Hallinan, Vikram Ravi, and Shri Kulkarni from the Astronomy department, all of whom provided invaluable guidance and support. I would like to thank Gregg, my advisor, who kept me focused on making an impact on our project's scientific goals. He also encouraged (and funded!) my travel to conferences around the world, where I met many collaborators who continue to inspire me. Vikram was instrumental in helping me bridge the gap between my engineering background and the world of astronomy. Shri proposed and funded my first project, the Galactic Radio Explorer (GReX), and appointed me as the project engineer, for which I am deeply grateful.

I would also like to especially acknowledge Steve Padin. One day in early 2022, I received a knock on my office door, and Steve introduced himself with the words, “I heard you need help.” Since then, Steve has spent countless hours in the lab with me, checking my work, keeping me on track, and working through problems alongside me. He has been a key advisor, particularly on all aspects of engineering, always offering his time and support whenever I needed it. I truly do not know if I would have been able to complete this degree without him, and for that, I am sincerely grateful.

Thank you to Liam Connor, who worked as project scientist for GReX. All the hours working on code and in the car driving back and forth to OVRO chatting about anything and everything is time I wouldn’t trade for anything.

I’d like to thank Schmidt Sciences for funding DSA-2000 and by extension my degree.

Thank you to James, Mark, Francois, Katie, Vinand, Jonas, Dave, Charlie, and everyone else at OVRO and the DSA-2000 team.

Thank you to David Hodge and Courtney Keeler for all the help in the lab.

Thank you to Jack, Jonathan, Mitch, Morag, Dan, Kathryn, and the rest of the CASPER collaboration.

Thank you to Glenn Price, Barb Caitlin, Scott Babcock, and the rest of the folks from Caltech’s performing and visual arts for the incredible time with the Wind and Symphonic Orchestra, Percussion Ensemble, and Jazz big band and improv group. You all gave me a much-needed artistic outlet and a way to engage with LA’s musical community.

Thank you to my committee, Professors Gregg Hallinan, Vikram Ravi, Steve Padin, Katie Bouman, Ali Hajimiri, and Alireza Marandi for helping me through candidacy and the final steps of my degree.

Finally, I’d like to thank my wonderful partner Ivy and our cat, Rocket, for moving across the country to support me in this crazy degree. Lastly, to my mother, who gave me love, support, and strength through this wild ride. I love you all.

ABSTRACT

The 2020s is the decade of survey instruments in astronomy. Radio astronomy is no exception, with Caltech's proposed DSA-2000 being the most powerful radio interferometer in the world, costing much less than competing instruments. Key to this achievement are two core breakthroughs: a completely ambient-temperature receiver and a "radio camera" backend that images the sky in real time. DSA-2000 will have record-breaking survey speed and sensitivity, enabled by these two key breakthroughs, giving astronomers all over the world open access to exquisite all-sky maps to enable the discovery of billions of new radio sources, precise timing of pulsars, and localization of fast radio bursts. The array will produce enough data to keep astronomers busy for a century.

In this thesis, we discuss the development of one of the key breakthroughs, the ambient-temperature receiver. Specifically, we focus on the design, testing, and implementation of the wideband, ambient-temperature low noise amplifier. We cover the design from analytic first principles through precision measurement of its performance. We follow this with a discussion of the design and implementation of the analog signal path, including a high performance, RF over fiber link. Finally, we discuss the Galactic Radio Explorer (GReX) instrument, designed as a global experiment probing the brightest radio transients in the local universe.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] K. A. Shila, “Computationally Efficient Design of an LNA Input Matching Network Using Automatic Differentiation,” *IEEE Journal of Microwaves*, May 7, 2025. DOI: 10.1109/JMW.2025.3568779,
K.A.S is the sole author, and designed the amplifier, performed the experiments, and wrote the paper. Minor edits were made to the published version to adapt to the thesis style.

[2] K. A. Shila, S. Niedbalski, L. Connor, *et al.*, “GReX: An Instrument Overview and New Upper Limits on the Galactic FRB Population.,” *Publications of the Astronomical Society of the Pacific*, 2025, In Review. arXiv: 2504.18680 [astro-ph.HE],
K.A.S is the first author and contributed the engineering details of the paper and the updated rate analysis. Minor edits were made to the published version to adapt to the thesis style.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 The Decade for Surveys

One would be hard-pressed to find a better time to be working in astronomy than at this moment. Amazing new instruments are coming online that are pushing the boundaries of what can be observed, giving astronomers a deluge of data to explore and new secrets of the universe to unlock. For engineers, every new instrument starts as a question: "What is possible?" Every year the answer changes as technology progresses, giving astronomers more compute power, more sensitivity, and lower costs.

This decade is especially exciting with the deployment of a slew of powerful new instruments. Specifically, we are seeing an emphasis on survey instruments, mapping out the sky across the entire EM spectrum. These types of experiments are not proposal-based where one would request to look at some interesting object, instead relying on the community at large to explore the data, searching for patterns, outliers, and really anything interesting. This is well-motivated, as most major discoveries in astronomy were serendipitous. Having such a large volume of data across parameter space enables discovery of the unexpected, assuming we are prepared [1]. Additionally, surveys democratize science by enabling everyone to explore the rich datasets, instead of allowing only a select few whose proposals were accepted to have access to high-quality data. All this is to say that surveys are a key tool in modern astronomy and are consistently ranked among the highest priority in science funding [2].

In the optical, the Vera Rubin Observatory in Chile will soon conduct a massive all-sky survey [3], with a projected first light in July 2025. The recently launched SPHEREx satellite [4] has just started its survey. In the radio, the VLA Sky Survey (VLASS) [5] is ongoing, imaging the whole sky visible to the VLA in New Mexico. VLASS builds off the success of previous VLA-based surveys[1], further improving sensitivity and resolution. However, while the survey continues to be successful, the speed at which it images the sky pales in comparison to the other modern surveys. Rubin images $10\,\mathrm{deg}^2$ every 15 s or $2\,400\,\mathrm{deg}^2$/hr (during the nighttime). SPHEREx

[1]Which are some of the most scientifically successful experiments that the VLA has performed.

will survey the entire sky once every six months, or a speed of 5 deg^2/hr. VLASS operates at only about 0.01 deg^2/hr, while splitting time between the survey and other planned observations. Even though VLASS (and previous radio surveys) are very successful, they are being executed on instruments poorly optimized for surveys.

This is where the Deep Synoptic Array 2000 (DSA-2000) [6] comes in. Rather than using existing radio telescopes to survey the sky, DSA-2000 is a bespoke survey instrument to match the scale and cadence of surveys from other wavelengths. This will result in a massive legacy dataset that will produce enough data to keep astronomers busy for decades to come.

1.2 Designing A Survey Instrument

In the design of a survey instrument, the key metric is the speed at which we can create images of the sky. The faster we can make high-sensitivity maps, the faster we can get these maps in the hands of astronomers to do science. We can quantify this measure via survey speed, defined as the time required to map one patch of sky down to some fixed noise level

$$\text{SS (deg}^2\text{/hr)} \approx 1.217\eta_a\Delta_\nu\left[\frac{\lambda S_\nu N D}{T_{\text{Sys}}}\right]^2 . \tag{1.1}$$

This equation for the survey speed of reflector-based radio telescope arrays (derived in Section B.1) depends on the number of antennas N, the system noise temperature T_{Sys}, the diameter of the dishes D, the aperture efficiency of the dishes η_a, the observation wavelength λ, the integration bandwidth Δ_ν, and the desired 1σ flux density S_ν in Jy. In essence, this metric combines the sensitivity of the instrument with how much of the sky it can see. One could also just ask about that sensitivity, given by

$$S_\nu\ (\mu\text{Jy @ 1 hr}) \approx 1852\frac{T_{\text{Sys}}}{\eta_a D^2 N\sqrt{\Delta_{\nu(\text{GHz})}}}, \tag{1.2}$$

which gives the RMS noise in μJy in one hour of integration. The ideal instrument would have a minimal noise and a large collecting area, which corresponds to a high survey speed.

Using these metrics, we can compare the planned DSA-2000 performance against other planned and deployed instruments. DSA-2000 will be composed of 2000, 5 m dishes operating at an aperture efficiency of about 0.7, observing from 0.7 GHz to 2 GHz with a system noise temperature of about 25 K. For this comparison, we will look at DSA-110 [7], ASKAP [8], MeerKAT [9], CHORD [10], VLA [11], and

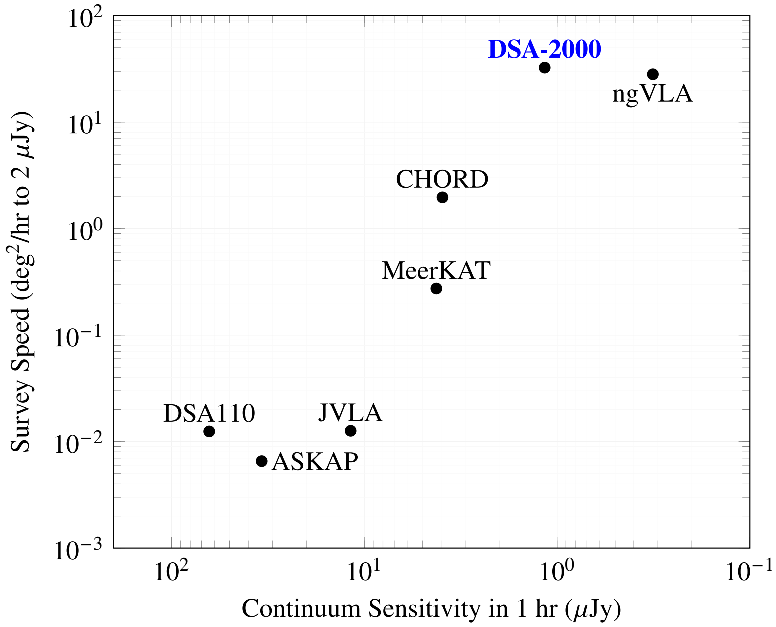


Figure 1.1: Comparison of radio telescopes at 1.4 GHz in survey speed and sensitivity.

ngVLA [12], shown in Figure 1.1. From these results, it is clear that the proposed DSA-2000 will outperform all instruments in this band in survey speed, with only the ngVLA having higher sensitivity. This is not too surprising as the ngVLA is a multi-billion dollar project with massive reflectors and cryogenically cooled receivers. If we quantify the science output of these instruments as proportional to their survey speed[2], and then try to predict science per dollar, we see a very different story.

The construction costs for a radio telescope array can be broken down into two major chunks: the processing hardware and the antennas themselves. For an N-antenna array, the cost of the compute scales with N^2. We will ignore operation cost and everything else just so we can get a grasp on the design-space. To keep costs low, it would seem that we would want to keep N low, but our survey speed *increases* by N^2, so these are certainly at odds. Currently, compute is relatively inexpensive for the kind of processing we need to do, at least compared to the cost of the antenna hardware itself. Our current estimates for DSA-2000 put the cost of the compute (again ignoring power, installation, operations, etc.) at around \$20M for 2,000 antennas. The antennas cost about \$30k each, for a total of \$60M before trenching, installation, etc. Compare this to ngVLA's much larger, cryogenically cooled antennas, which cost about \$7M each.

[2] A somewhat dubious claim, but good enough for this exercise.

To improve sensitivity, matching the ngVLA, we could add cryogenics to cool the DSA-2000 receiver to 20 K, which would cost about \$100k per antenna [13]. Simulations predict this will reduce our system noise temperature by 7 K to 18 K, as cooling only impacts the feed and LNA and there are still other sources of noise (the sky, spillover noise, etc.; see Figure 1.2). If each antenna costs around \$30k, adding cryogenics quadruples the cost, while not even doubling the survey speed.

All of this leads to the conclusion that if you can build high-performance/inexpensive antennas and receivers, build lots of them! And the key to an inexpensive antenna is not using cryogenics. As survey speed scales with T_{Sys}^{-2}, any reduction in system noise will dramatically increase the science per dollar of the instrument. The strong impact of reducing noise then motivates the work in this thesis. Of all the contributions to T_{Sys}, currently the LNA is the largest (as shown in Figure 1.2), which suggests where to spend effort to maximize the impact on the science.

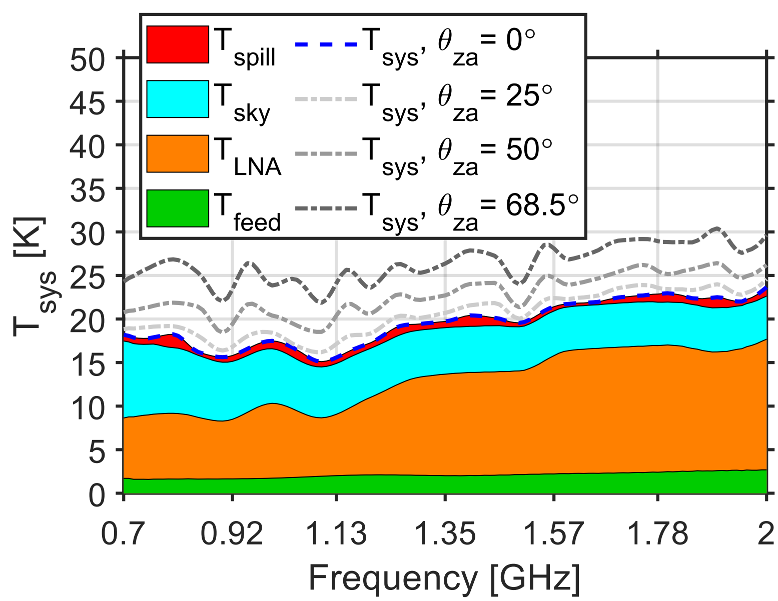


Figure 1.2: DSA-2000 system noise temperature contributions. Reproduced from [14] with permission from J. Flygare.

If it is so obvious to build a large N telescope, a reasonable question would be, "why haven't people done this before?" To answer this, we need to discuss a bit more about radio telescope interferometers.

Radio telescopes were initially built using a single, large aperture. A large collecting area increases sensitivity and resolution, but the cost of building a large reflector with diameter D is $\propto D^{2.5-2.7}$ [15]. Additionally, there are practical upper limits on the size of a reflector one could build, with the largest single-dish antenna being the FAST telescope in China at 500 m in diameter, with a 300 m effective aperture. At a typical observation frequency of 1.4 GHz, the diffraction-limited resolution of FAST is 3 arcmin. Compare this with a modest optical observatory such as Keck, with a 10 m mirror, where they achieve an angular resolution of better than 0.05 arcsec (assuming a modern, adaptive optics imaging system).

To overcome this resolution shortcoming, the concept of radio interferometry was developed by Ruby Payne-Scott and Joseph Pawsey in the 1940s. They used the interference pattern produced by combining radio signals reflecting off the ocean with an antenna on top of a cliff to pinpoint emission from sunspots. This essentially provided them with the angular resolution of a telescope the size of the distance from their antenna to the image of their antenna in the ocean. This concept was further developed by Martin Ryle and Antony Hewish in the 50s, where they used an array of antennas and Earth's rotation to synthesize an equivalent aperture beyond that provided by just the antenna baselines. This technique is so powerful, it was part of Ryle and Hewish's Nobel Prize in Physics in 1974[3]. Modern aperture synthesis uses a combination of spatially distributed antennas and Earth rotation to synthesize a single massive telescope. Very long baseline interferometry (VLBI) has receivers all over the world to synthesize an Earth-sized telescope, used to create the first images of a black hole [16].

The radio interferometer imaging procedure starts with the coherent combination of the time-averaged product of signals from pairs of antennas. As the signal from the sky is essentially incoherent noise, this result is a statistical correlation product between antennas. There exists a relationship between the collection of all these correlation products (known as visibilities) and the image of the sky. This relationship is known as the van Cittert-Zernike theorem[4] and states

$$V(\mathbf{r}_1, \mathbf{r}_2) \approx \iint_{-\infty}^{\infty} I(l, m) e^{-2\pi j (ul+vm)} dl dm = \mathcal{F}\{I(l, m)\}\,, \tag{1.3}$$

where $V(\mathbf{r}_1, \mathbf{r}_2)$ is the visibility or mutual spatial coherence function given by the time average product of the received signal (electric field) from the pair of antennas at two points in space ($\mathbf{r}_1$ and $\mathbf{r}_2$)

$$V(\mathbf{r}_1, \mathbf{r}_2) = \langle E(\mathbf{r}_1) E(\mathbf{r}_2)^* \rangle\,, \tag{1.4}$$

[3]Which, notably, did not include Jocelyn Bell, who actually discovered the first pulsar to which the Prize was attributed.

[4]Using a small angle approximation, which results in the Fourier relationship.

and u and v are the spatial differences between the two points[5]

$$\mathbf{r}_1 - \mathbf{r}_2 = \mathbf{u} = (x_1 - x_2, y_1 - y_2) = (u, v)\,, \tag{1.5}$$

and $I(l, m)$ is the intensity distribution (image) of the sky in a direction-cosine reference frame. All of this is to say that the inverse Fourier transform of the visibilities gives us the image of the sky.

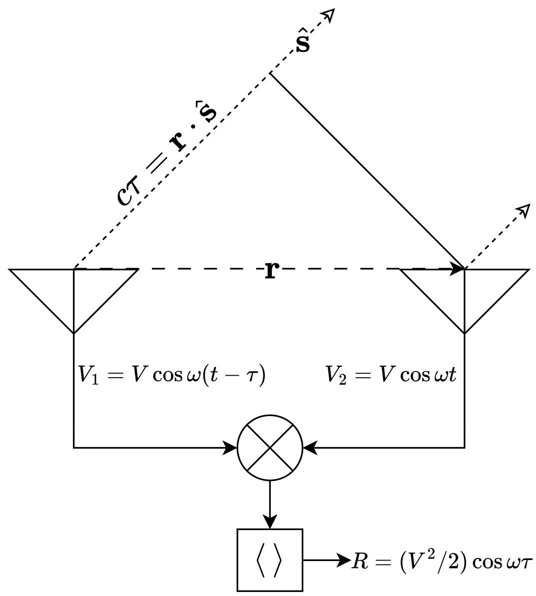
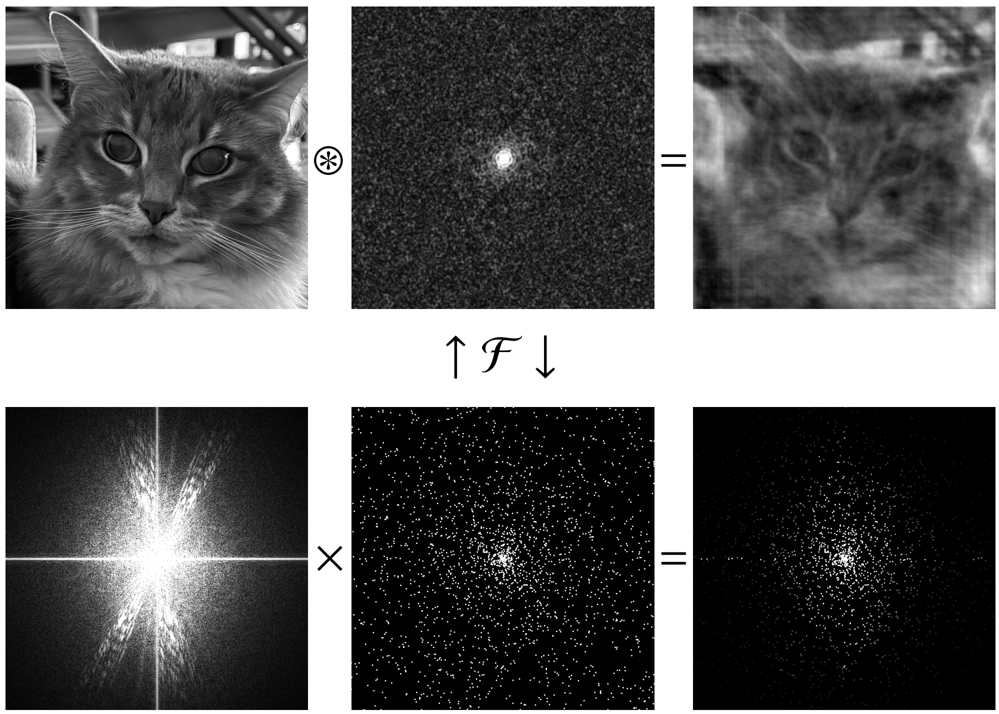


Figure 1.3: Diagram of a radio interferometer.

A problem then arises from the fact that we have not continuously sampled all the points in u and v as we have a discrete number of antennas and therefore a finite number of visibilities. Not only that, but there is a maximum distance that we can separate antennas due to cable lengths, land usage, Earth's diameter, and other practical considerations. Combining these two shortcomings, we essentially image the sky through a diffraction-limited, shattered mirror. This imaging procedure is shown in Figure 1.4 as Fourier pairs, where a sparse sampling of the Fourier-domain of the true image results in the measured, dirty image. As the dirty image is a distorted version of the true image, we must attempt to recover the true image algorithmically. However, due to the sparse sampling of the uv plane, this is an ill-posed inverse problem and does not have a simple solution. Various optimization-based strategies are used for this, perhaps the most popular being an iterative blind deconvolution algorithm called CLEAN [18].

[5]Ignoring changes in height z, which is something you might want to consider [17]

(a) True Image (b) Point-Spread Function (c) Dirty Image



(d) True Visibilities (e) Sampling Function (f) Sampled Visibilities

Figure 1.4: Radio interferometer Fourier pairs.

As we increase the number of antennas and the area they cover, we fill out the *uv* plane more and more, resulting in a better image. In the same manner as survey speed, this encourages us to build a telescope with bigger N. However, the tradition in radio astronomy is to distribute the raw visibilities as the data product. In sparse arrays like the Very Large Array, or the Event Horizon Telescope, the scientist performing the analysis might want full control over how the image is formed from the sparse data, making their own assumptions on how to interpret them. For DSA-2000, let us assume we observe 16,000 frequency channels every 10 seconds. If there are 2,000 dual-polarized antennas and each visibility is an 8+8-bit complex number, that works out to a visibility data rate of 25.6 GB per second. Over a year at this rate, we would accumulate 0.8 exabytes of data[6]. The entirety of the Internet Archive[7] stores around 100 petabytes of data for real-time, global access at a cost of \$2.5M per month. At our data rate, that totals \$300M per year, more than the construction cost of the array, just in storage! Not only that, but this rate would keep increasing as we accumulate more and more data as we continue to survey the sky. This cost is

[6]Or eight times the storage capacity of Data from Star Trek: The Next Generation.

[7]`https://archive.org/`

one of the primary reasons why the community has not built large N arrays.

But here is the trick; at some point as N increases, the uv plane becomes complete enough where a simple inverse Fourier transform of the sampled visibilities results in a dirty image that requires little to no "cleaning". Therefore, we could build a deterministic pipeline to capture images of the sky in real time and store those images instead of the raw visibilities, saving many orders of magnitude in storage costs. This is what we call the "Radio Camera" concept and is the enabling feature of the DSA-2000, breaking the cost curve in storage for large N arrays. The low-pass filtering effect due to the maximum separation of antennas is still in play, resulting in less sharp final images. This can be improved slightly through standard image post-processing such as sharpening algorithms or ML-based super resolution approaches [19].

1.3 Thesis Outline

Given the motivation to build an inexpensive, high-performance receiver, this thesis covers my work in squeezing out every drop of performance across the analog signal path for DSA-2000 to maximize sensitivity and survey speed, increasing our science impact per dollar. The thesis is broken into two major sections: my work on the DSA-2000 and my work on the Galactic Radio Explorer (GReX), which is a separate, smaller experiment aimed at finding exceptionally bright fast radio bursts (FRBs). These are two distinct projects and are presented out of order chronologically, with DSA-2000 content presented first, representing the bulk of my work. GReX acted as my introduction to the field of radio astronomy, starting early in my degree at Caltech. It was intended to be a pathfinder instrument for DSA-2000, but changed in scope as the project progressed.

The novel contributions of this thesis are:

- Development of an ambient-temperature low noise amplifier matching network using advanced algorithmic techniques (DSA-2000)
- Experiments with feed impedance to LNA noise matching and chip-scale Peltier micro-cooling (DSA-2000)
- Development of a low-cost, rapidly deployable radio telescope (GReX)

We start in Chapter 2 with an overview of analytic formalisms surrounding low noise amplifier design and optimization. Having a collection of closed-form results that encompass the various performance metrics for an LNA are required for designing an optimization framework. Chapter 3 is a reproduction of a published article that discusses the design of an improved room-temperature low noise amplifier for the DSA-2000, using the formalisms built in the previous chapter, and is my primary contribution. Chapter 4 discusses the complete analog signal path for DSA-2000, including the analysis, design, and implementation of a high-performance RF over fiber link and further enhancements to the LNA such as solid-state cooling and antenna impedance co-optimization. We finish with Chapter 5, a reproduction of an article that describes the GReX instrument and new upper limits on galactic FRB populations.

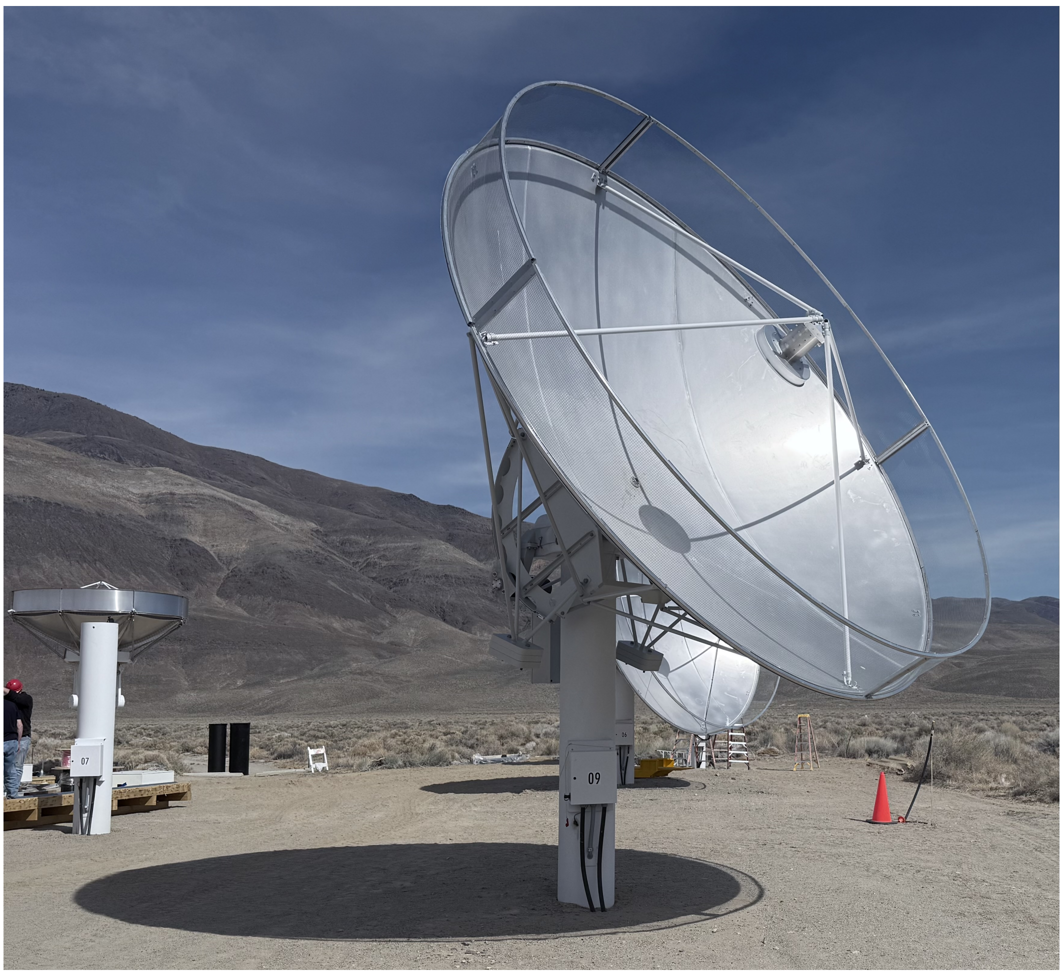


Figure 1.5: DSA-2000 Test Array under construction at OVRO

References

[1] R. P. Norris, "Discovering the Unexpected in Astronomical Survey Data," *Publications of the Astronomical Society of Australia*, vol. 34, e007, Jan. 2017. DOI: 10.1017/pasa.2016.63.

[2] E. National Academies of Sciences and Medicine, *Pathways to Discovery in Astronomy and Astrophysics for the 2020s*. Washington, DC: The National Academies Press, 2023. DOI: 10.17226/26141.

[3] F. B. Bianco, Ž. Ivezić, R. L. Jones, *et al.*, "Optimization of the Observing Cadence for the Rubin Observatory Legacy Survey of Space and Time: A Pioneering Process of Community-focused Experimental Design," *The Astrophysical Journal Supplement Series*, vol. 258, no. 1, p. 1, Dec. 2021. DOI: 10.3847/1538-4365/ac3e72.

[4] B. P. Crill, M. Werner, R. Akeson, *et al.*, "SPHEREx: NASA's near-infrared spectrophotometric all-sky survey," in *Space Telescopes and Instrumentation 2020: Optical, Infrared, and Millimeter Wave*, vol. 11443, SPIE, Dec. 13, 2020, pp. 61–77. DOI: 10.1117/12.2567224.

[5] M. Lacy, S. A. Baum, C. J. Chandler, *et al.*, "The Karl G. Jansky Very Large Array Sky Survey (VLASS). Science Case and Survey Design," *Publications of the Astronomical Society of the Pacific*, vol. 132, p. 035 001, Mar. 1, 2020. DOI: 10.1088/1538-3873/ab63eb.

[6] G. Hallinan, V. Ravi, S. Weinreb, *et al.*, "The DSA-2000 — A Radio Survey Camera," *Bulletin of the AAS*, vol. 51, no. 7, Sep. 30, 2019.

[7] V. Ravi and D.-1. Collaboration, "The DSA-110: Overview and first results," *Bulletin of the AAS*, vol. 55, no. 2, Feb. 2023.

[8] A. W. Hotan, J. D. Bunton, A. P. Chippendale, *et al.*, "Australian square kilometre array pathfinder: I. system description," *Publications of the Astronomical Society of Australia*, vol. 38, e009, Jan. 2021. DOI: 10.1017/pasa.2021.1.

[9] J. Jonas and the MeerKAT Team, "The MeerKAT Radio Telescope," in *Proceedings of MeerKAT Science: On the Pathway to the SKA*, vol. 277, SISSA Medialab, Feb. 2018, p. 001. DOI: 10.22323/1.277.0001.

[10] K. Vanderlinde, A. Liu, B. Gaensler, *et al.*, "The Canadian Hydrogen Observatory and Radio-transient Detector (CHORD)," Zenodo, Tech. Rep., Oct. 2019. DOI: 10.5281/zenodo.3765414.

[11] "VLA Observational Status Summary 2022B," NRAO Science Site. (), [Online]. Available: https://science.nrao.edu/facilities/vla/docs/manuals/oss2022B.

[12] NRAO, *ngVLA Performance Estimates (December 2021)*, https://ngvla.nrao.edu/page/performance.

[13] L. R. D'Addario, "Advanced Cryocoolers For Next Generation VLA," California Institute of Technology, ngVLA Memo 24, Sep. 2017.

[14] J. Flygare, "Wideband Low-Loss Feed Design for the DSA-2000 Ambient Temperature Array in Radio Astronomy," in *2024 IEEE International Symposium on Antennas and Propagation and INC/USNC-URSI Radio Science Meeting (AP-S/INC-USNC-URSI)*, Firenze, Italy: IEEE, Jul. 2024, pp. 5–6. DOI: 10.1109/AP-S/INC-USNC-URSI52054.2024.10686435.

[15] G. T. van Belle, A. B. Meinel, and M. P. Meinel, "The scaling relationship between telescope cost and aperture size for very large telescopes," in *Ground-Based Telescopes*, vol. 5489, SPIE, Sep. 28, 2004, pp. 563–570. DOI: 10.1117/12.552181.

[16] Event Horizon Telescope Collaboration, K. Akiyama, A. Alberdi, *et al.*, "First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole," *The Astrophysical Journal*, vol. 875, p. L1, Apr. 1, 2019. DOI: 10.3847/2041-8213/ab0ec7.

[17] T. J. Cornwell, K. Golap, and S. Bhatnagar, "W Projection: A New Algorithm for Wide Field Imaging with Radio Synthesis Arrays," *Astronomical Data Analysis Software and Systems*, 2005.

[18] J. A. Högbom, "Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines," *Astronomy and Astrophysics Supplement*, vol. 15, pp. 417–426, Jun. 1974.

[19] L. Connor, K. L. Bouman, V. Ravi, and G. Hallinan, "Deep radio-interferometric imaging with POLISH: DSA-2000 and weak lensing," *Monthly Notices of the Royal Astronomical Society*, vol. 514, no. 2, pp. 2614–2626, Aug. 2022. DOI: 10.1093/mnras/stac1329.

Chapter 2

OPTIMIZATION OF LOW NOISE AMPLIFIERS

In this chapter, we discuss the analytic formalisms around the design of low noise amplifiers and applications in optimization-based design. Too often, engineers quickly reach for the optimizer in the circuit simulator to achieve some desired result. Without a sense of fundamental limits and a grasp on the trade-space, this quickly leads to frustration when desired goals are not met. All the trade-offs that we make when designing amplifiers are sourced from analytic expressions. To design for a given point in the trade-space, we typically need to solve a nonlinear, constrained optimization problem. For many of these problems, the objective function is surprisingly convex[1], with some having an exact, analytic solution. We use these solutions along with appropriate numerical optimization to quickly visualize the design space for a given device and make informed decisions during amplifier design. This analytic background then provides motivation and bounds for building more complex amplifier optimization problems.

2.1 Noise Theory Background

No discussion around LNAs would be complete without an overview of noise in microwave circuits. Rather than give full derivations, which can be found in many excellent textbooks, we will focus on the important results.

Sources of Noise

Johnson-Nyquist noise [1] describes that a resistor at physical temperature T produces a noise power proportional to its temperature due to the physical agitation of carriers inside the conductor. For temperature T and Boltzmann's constant k_B, the power spectral density of the noisy resistor is

$$S_{\text{Nyquist}}\ (W/\text{Hz}) = k_B T\ . \tag{2.1}$$

This noise power is approximately white, with the power spectral density falling off at high frequencies following Planck's law. The approximation that $S_{\text{Planck}} \approx S_{\text{Nyquist}}$ is adequate up through terahertz frequencies at room temperature. This noise is

[1] Where there is exactly one local minimum, versus non-convex optimization where finding the global minimum is generally NP-Hard.

present in all electronics and can be the limiting factor for sensitivity at microwave frequencies. As all sources of thermal noise are uncorrelated, there is little we can do to reduce it apart from physical cooling, motivating the use of cryogenics in applications such as radio telescopes and quantum computers.

Shot noise, described by Schottky in 1918 [2], arises from random fluctuations in electric current due to the discretization of charge. In DC conditions, if one were to observe each carrier individually pass through a conductor, the exact arrival time of the next carrier is uncertain and follows a Poissonian process. The fluctuation in the count of carriers per unit time over the average velocity of the carriers is the shot noise. This process is also white, and given a current of I and electron charge q, shot noise has a power spectral density of

$$S_{\text{Shot}}\ (A^2/\text{Hz}) = 2qI\,. \tag{2.2}$$

Other sources of noise include $1/f$ noise, which combines many mechanisms with a $1/f$ power spectral density, and quantum noise, which describes the minimum noise from a physical linear amplifier. For microwave circuits, these sources of noise are usually dwarfed by thermal and shot noise. The one exception is for oscillators, as upconversion of $1/f$ noise can result in phase noise in microwave circuits.

Noise in Two Port Networks

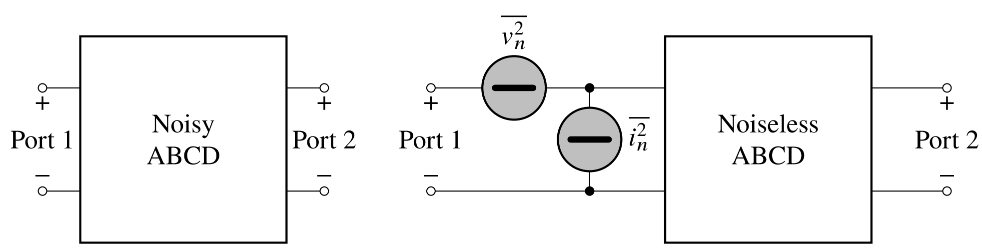


Figure 2.1: Equivalent noise representations of ABCD networks.

Given a linear two-port network with a collection of internal noise generators, Rothe and Dahlke [3] show us that we can represent this circuit as a noiseless two-port network with external, correlated noise generators. Figure 2.1 shows the equivalent circuits in an ABCD (cascade) network form, with the noise current and voltage sources at the input. These sources are correlated by some complex correlation coefficient, ρ. Given the magnitude of these noise sources and the correlation coefficient, one can draw a few conclusions about the behavior under varied source

(Port 1) terminations. For example, if the sources were perfectly correlated, we could present a source admittance Y_S that will generate a noise voltage from i_n to perfectly destructively interfere with v_n, resulting in zero noise. As we have seen earlier, there are many uncorrelated noise processes in real circuits, so in reality this correlation in the input-referred noise generators is some nonzero, less than unity quantity. The relationship between Y_S and the resulting noise implies that there exists a source impedance that minimizes the input-referred noise.

The noise dependence on the source admittance is the primary result that enables the development of low noise microwave amplifiers. Just as we can design matching networks to power match devices, we can design matching networks to present the optimal source impedance that minimizes noise.

2.2 Noise Circles and Constraint Equations

The noise figure of a two-port amplifier [4] is

$$F = F_{\text{Min}} + \frac{4r_n|\Gamma_{\text{S}} - \Gamma_{\text{Opt}}|^2}{(1 - |\Gamma_{\text{S}}|^2)|1 + \Gamma_{\text{Opt}}|^2}\,, \tag{2.3}$$

where F_{Min} is the minimum achievable noise temperature, r_n is the so-called noise equivalent resistance representing how quickly the noise changes as Γ_{S} moves from Γ_{Opt}, normalized to the characteristic impedance Z_0, Γ_{Opt} is the source reflection coefficient that achieves F_{Min}, and Γ_{S} is the source reflection coefficient. These three numbers, F_{Min}, r_n, and Γ_{Opt}, form the noise parameters of an amplifier and fully characterize the noise behavior of the device under any source impedance condition. These three numbers can be computed from the noise generators in Figure 2.1, $\overline{v_n^2}$, $\overline{i_n^2}$, and their correlation ρ.

If we were to attempt to solve Equation 2.3 for Γ_{S} such that we achieve a desired noise figure F, we start by separating the Γ_{S} components from the rest of the equation

$$\frac{|\Gamma_{\text{S}} - \Gamma_{\text{Opt}}|^2}{1 - |\Gamma_{\text{S}}|^2} = \frac{F - F_{\text{Min}}}{4r_n}|1 + \Gamma_{\text{Opt}}|^2\,. \tag{2.4}$$

As the right side of this equation is constant, we write

$$N = \frac{\Delta F}{4r_n}|1 + \Gamma_{\text{Opt}}|^2\,, \tag{2.5}$$

where ΔF is the desired noise penalty (in noise figure) from F_{Min}. The solution to this in Γ_{S} is a circle on the complex plane with center

$$C_F = \frac{\Gamma_{\text{Opt}}}{1 + N} \tag{2.6}$$

and radius

$$R_F = \frac{\sqrt{N^2 + N(1 - |\Gamma_{\text{Opt}}|^2)}}{1 + N} . \tag{2.7}$$

So, for a given set of noise parameters and desired noise (greater than or equal to the device's minimum noise), we draw a circle on the Smith chart of equal noise. If we were to design a matching network to build an amplifier for this noise, we have an infinite number of impedances to choose from.

As engineering is all about tradeoffs, we need to start introducing additional constraints apart from noise. If we did not, one would rightly ask, "Why not simply match to Γ_{Opt}?" Primarily, the answer to this arises from the unfortunate reality that Γ_{Opt} is distinct from the conjugate input reflection coefficient of the device, Γ_{In}^*. This implies that as you attempt to match a device for optimum noise, you do not optimally deliver power to it. In other words, the optimum noise impedance has a reflection or gain penalty.

In system design, these various performance metrics describe the trade-space. One can pick high gain or low reflection, but have bad noise performance, or vice versa. As Friis tells us in the cascaded noise equation [5]

$$F_{\text{Total}} = F_1 + \frac{F_2 - 1}{G_{A_1}} + \frac{F_3 - 1}{G_{A_1} G_{A_2}} \cdots + \frac{F_N - 1}{\prod_{i=1}^{N-1} G_{A_i}} , \tag{2.8}$$

where G_{A_i} is the available gain of the ith amplifier, we must maximize the gain of the early stages to reduce the noise impact of noise contributors further down the signal chain. Again, part of the system design work is evaluating all these tradeoffs.

Traditionally, the exploration into these tradeoffs is rather brute-force. However, a more disciplined approach can be taken. We can formulate exact solutions for optimum noise under constraints, allowing us to explore theoretical limits given a device's intrinsic behavior and informing engineering decisions made in design.

Optimum Noise Under Constraints

Suppose we want to design an LNA with a given reflection coefficient or equivalent voltage standing wave ratio (VSWR). The amplifier we want will have minimum noise, with VSWR being less than or equal to some limit. Given that

$$\text{VSWR} = \frac{1 + |\Gamma_{\text{A}}|}{1 - |\Gamma_{\text{A}}|} \tag{2.9}$$

and

$$|\Gamma_{\text{A}}| = \left| \frac{\Gamma_{\text{In}} - \Gamma_{\text{S}}^*}{1 - \Gamma_{\text{In}} \Gamma_{\text{S}}} \right| , \tag{2.10}$$

where

$$\Gamma_{\mathrm{In}} = S_{11} + \frac{S_{12}S_{21}\Gamma_{\mathrm{L}}}{1 - S_{22}\Gamma_{\mathrm{L}}}, \tag{2.11}$$

we write this problem as a mathematical minimization equation via

$$\begin{aligned} \underset{\Gamma_{\mathrm{S}}}{\text{minimize}} \quad & F(\Gamma_{\mathrm{S}}) \\ \text{subject to} \quad & |\Gamma_{\mathrm{A}}| \leq |\Gamma_{\mathrm{A}}|_{\mathrm{Max}} . \end{aligned} \tag{2.12}$$

As written, this is a constrained, nonlinear optimization problem. These types of problems are characteristically difficult to solve and are usually non-convex. In this case, however, there is an exact solution.

In the previous section, we mentioned that constant noise induces circles in the Γ_{S} plane. The same is true for constant VSWR. For a given $|\Gamma_{\mathrm{A}}|$, we draw the circle with center

$$C_V = \frac{\Gamma_{\mathrm{In}}^*(1 - |\Gamma_{\mathrm{A}}|^2)}{1 - |\Gamma_{\mathrm{A}}\Gamma_{\mathrm{In}}|^2} \tag{2.13}$$

and radius

$$R_V = \frac{\left||\Gamma_{\mathrm{A}}||1 - |\Gamma_{\mathrm{In}}|^2\right|}{\left|1 - |\Gamma_{\mathrm{A}}\Gamma_{\mathrm{In}}|^2\right|} . \tag{2.14}$$

Points inside this circle have strictly lower reflection, so we know a solution to Equation 2.12 must lie on this circle or inside it. If we consider the case where Γ_{Opt} is not inside the constant VSWR circle, we fully constrain the optimum solution to lie on this circle. This is where we apply the noise circle formalism from before. As the radius of this circle scales with ΔF, we want to find the noise circle that just touches the constant VSWR circle. We know this will have a solution, as we can keep increasing the radius until it touches the VSWR circle. We also know that this will have exactly one solution, as any further intersections would arise from a larger circle with higher noise. This is the geometric intuition for the convexity of Equation 2.12.

Expressing this condition mathematically, we state

$$R + R_F = |C - C_F| , \tag{2.15}$$

or the sum of the two circles' radii equal the Euclidean distance between the centers of the two circles, given an arbitrary circle with center C and radius R and the noise circle stated above. Solving this relation for N (Equation 2.5) gives the quadratic result following [6]

$$N = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} , \tag{2.16}$$

where

$$
\begin{aligned}
c_1 &= |C|^2 - R^2 - 1 \\
c_2 &= |\Gamma_{\text{Opt}} - C|^2 - R^2 \\
A &= c_1^2 - 4R^2 \\
B &= 2c_1c_2 + 4R(|\Gamma_{\text{Opt}}|^2 - 1) \\
C &= c_2^2
\end{aligned}
$$

We then take this result and substitute it in to Equation 2.5 to solve for ΔF. To disambiguate between the two solutions, we apply the constraint that N must be positive, and if both solutions are positive, we take the smaller of the two. From here, we immediately compute a minimum theoretical noise under a VSWR constraint.

Worked Examples

To showcase these results, we investigate the 400 nm x 200 µm gate, four-finger (4F50) transistor from Diramics [7] at a bias of V_d=0.5 V, I_d=16 mA, using the vendor-supplied model. This transistor is an exceptionally high-performance device with very low noise, even at room temperature, and has been used in previous LNAs such as [8], [9]. We start by considering only the device's S and noise parameters, with no further circuitry. Some of these metrics are shown in Figure 2.2. This allows us to set the stage for input/output matching and feedback.

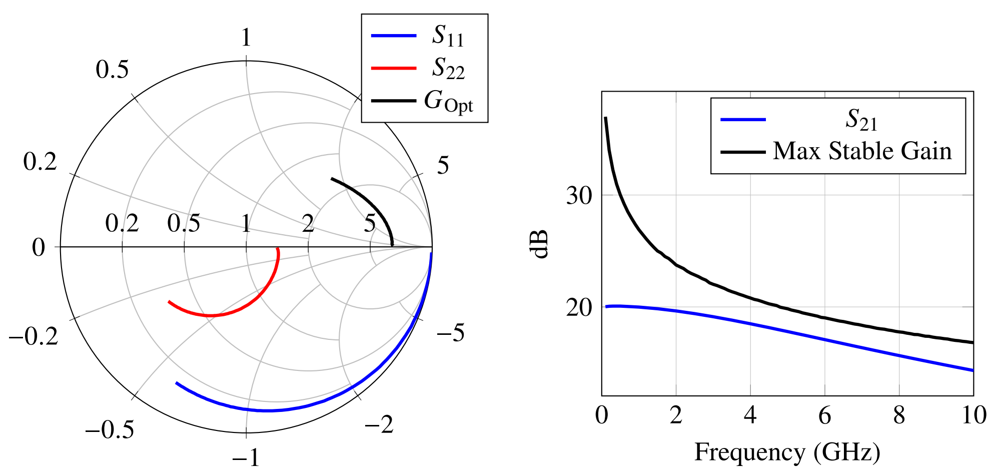


Figure 2.2: 4F50 S_{11}, S_{22}, S_{21}, and Γ_{Opt} from 0.1 GHz to 10 GHz.

We start by analyzing this device at a single frequency point, 1.5 GHz. If we wanted the optimum Γ_{S} for this device under the constraint that $|\Gamma_{\text{A}}| \leq 0.9$, we find ΔF of 0.012 or an F of 1.038 or 11.1 K of noise temperature. Using this result, we compute the center and radii of the two circles. Then, with both centers and either radius, we compute their osculation point using simple geometry

$$\Gamma_S = \begin{cases} C_F - R_F \frac{C-C_F}{|C-C_F|}, & \text{if } |C - C_F| \le R \text{ and } R_F \le R \\ C_F + R_F \frac{C-C_F}{|C-C_F|}, & \text{otherwise} \end{cases} . \tag{2.17}$$

Plotted on the Smith chart in Figure 2.3, we observe the result is as expected, with the intersection of the two circles giving the optimum Γ_S.

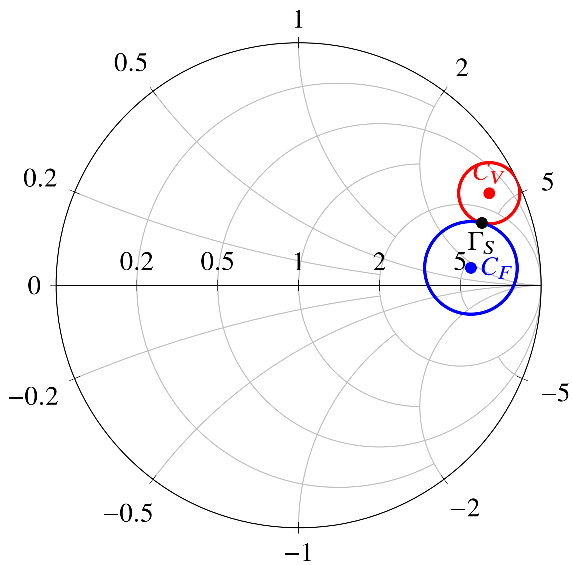


Figure 2.3: Diramics 4F50 $|\Gamma_A|$=0.9 optimum noise solution at 1.5 GHz.

We now begin to explore some useful results. First, we compute the achievable noise given a range of VSWR constraints to quickly analyze the tradeoff between the two. In Figure 2.4, we plot the achievable noise temperature in K under a reflection constraint sweeping $|\Gamma_A|$ from 0.1 to 0.95. A value for $|\Gamma_A|$ higher than 0.95 places Γ_{Opt} inside the VSWR circle, implying that for any constraint higher than 0.95 we would pick $\Gamma_S = \Gamma_{Opt}$ and our noise would be F_{Min}. This result alone tells us that we will have to make considerable sacrifices in reflection to achieve good noise performance.

It is important to note that the Equation 2.16 solution is agnostic to the constraining circle. Instead of VSWR, we could use a different circle such as constant gain instead. Specifically, we look at circles of constant available gain, as those are drawn on the Γ_S plane as is noise.

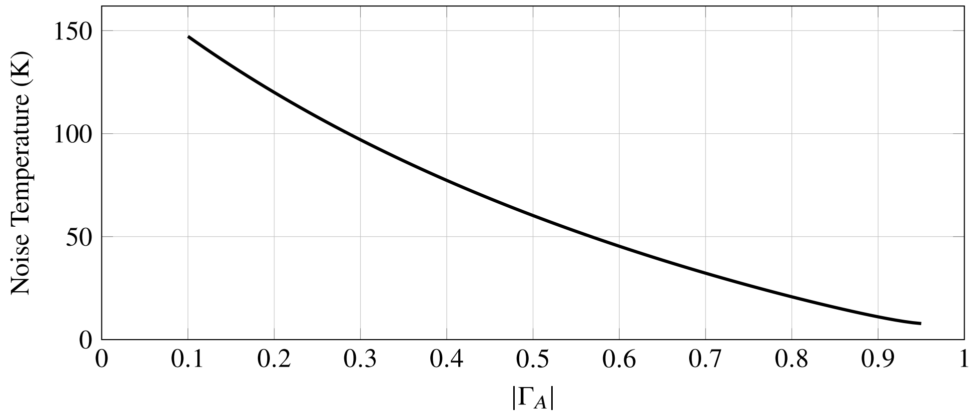


Figure 2.4: Optimum noise figure versus $|\Gamma_\mathrm{A}|$ at 1.5 GHz.

Available gain is defined as

$$G_A = \frac{P_\mathrm{AVN}}{P_\mathrm{AVS}} = \frac{\text{Power available from the network}}{\text{Power available from the source}}\,, \tag{2.18}$$

where

$$G_A = \frac{1-|\Gamma_\mathrm{S}|^2}{|1-S_{11}\Gamma_\mathrm{S}|^2}\frac{|S_{21}|^2}{1-|\Gamma_\mathrm{Out}|^2} \tag{2.19}$$

and

$$\Gamma_\mathrm{Out} = S_{22} + \frac{S_{12}S_{21}\Gamma_\mathrm{S}}{1-S_{11}\Gamma_\mathrm{S}}\,. \tag{2.20}$$

This function is only parameterized by Γ_S and the device characteristics, operating under the assumption that the output is conjugate-matched. In the same manner as Γ_A and noise, we solve for the values of Γ_S that result in the same available gain. If we define a normalized desired gain of

$$g_a = \frac{G_a}{|S_{21}|^2}\,, \tag{2.21}$$

the center and radius of the circles for constant g_a are

$$C_a = \frac{g_a\left(S_{11}^* - S_{22}\Delta^*\right)}{1+g_a\left(|S_{11}|^2-|\Delta|^2\right)} \tag{2.22}$$

and

$$R_a = \frac{\sqrt{1-g_a\left(1-|S_{11}|^2-|S_{22}|^2+|\Delta|^2\right)+g_a^2|S_{12}S_{21}|^2}}{1+g_a\left(|S_{11}|^2-|\Delta|^2\right)}\,, \tag{2.23}$$

where

$$\Delta = S_{11}S_{22} - S_{12}S_{21}\,.$$

Looking back at Figure 2.2, at 1.5 GHz, we achieve a maximum stable gain (MSG) of about 25 dB and an S_{21} (gain under 50 Ω conditions) of about 20 dB where

$$MSG = \frac{|S_{21}|}{|S_{12}|}\,. \tag{2.24}$$

If we draw the circle of constant available gain of 20 dB (which intersects the center of the Smith chart, as it is the value of S_{21}) and the intersecting minimum noise circle, we find the result shown in Figure 2.5. Notably, Γ_{Opt} is inside the constant gain circle, implying that Γ_{Opt} results in a higher gain than the chosen value. While usually we would want more gain, we will proceed with the G_A =20 dB solution for illustrative purposes.

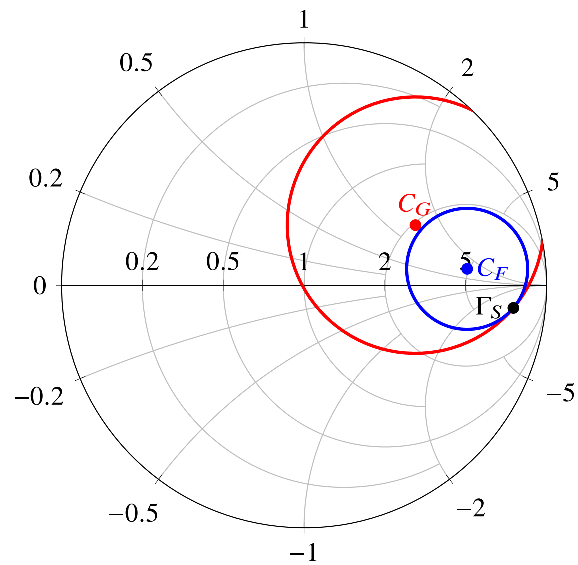


Figure 2.5: 20 dB optimum noise solution at 1.5 GHz.

Observing the relative positions of S_{11}^* and the VSWR circles from Figure 2.3, this Γ_S solution will have suboptimal VSWR. To be more precise, we can substitute this Γ_S into Equation 2.10 and find that the $|\Gamma_\text{A}|$ for this solution is 0.986 with a noise temperature of 13.2 K. From the other perspective, if we took the Γ_S solution from Figure 2.3 with a $|\Gamma_\text{A}| = 0.9$, we find that the available gain using Equation 2.19 has a high negative value, indicating this solution is unstable! If we use the unilateral approximation where $S_{12} = 0$, we find that the available gain for the Γ_S solution under the $|\Gamma_\text{A}| = 0.9$ constraint is 28 dB, 2 dB over the maximum stable gain. In fact, if we compute Γ_{Out} for this same Γ_S, we find a magnitude greater than unity, indicating oscillation.

Stability and Degeneration

Up to this point, we have failed to mention the impact of stability on the computed results. As we intend to design an amplifier and not an oscillator, it is imperative that the magnitudes of our reflection coefficients are less than unity. For a single gain element (transistor), we use Rowlett's [10] stability factor, K, to determine if the amplifier is stable.

$$K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}S_{21}|}\,. \tag{2.25}$$

When K is greater than unity and the magnitude of the determinant of the S matrix $|\Delta|$ is less than unity, the amplifier is unconditionally stable. This considers potential matching on both the source and load side of the device.

While stability is critical, we can be a bit more lax than "unconditional" stability. To expand on this, stability of every impedance over the load plane is not strictly required, as we will know precisely what the Γ_{L} value will be (assuming we are conjugate matching). On the input, we have a bit more to consider. As shown earlier, we have a range of impedances that represent various trade-offs. To make sure we are correctly comparing performance between options, those Γ_{S} solutions must be stable. Additionally, we have been operating under the assumption that this LNA will be attached to a $50\,\Omega$ source, which is not quite true. If the amplifier we are designing were a consumer product, we would have no idea what source impedance the customer might try to attach it to. For us in astronomy, the impedance of the feed antenna is typically designed to be close to $50\,\Omega$, but will degrade out of band. We need to ensure that over the frequency range that the amplifier has gain (up to tens of GHz for this high f_t Diramics device), the presented source impedance does not cause the amplifier to oscillate.

Instead of evaluating unconditional stability, we focus on the magnitude of the S parameters and the geometric stability factors [11]. Specifically, we compute μ', which represents the distance from the center of the Smith chart to the source-plane stability circle (the boundary of the impedance region that pushes the device into instability). If this quantity is greater than unity, no impedance in the source plane will cause the amplifier to oscillate.

$$\mu' = \frac{1 - |S_{22}|^2}{|S_{11} - S_{22}^{*}\Delta| + |S_{21}S_{12}|} \tag{2.26}$$

There are many ways one could stabilize a potentially unstable amplifier. Unfortunately, most of these methods degrade the amplifier's noise. An alternative strategy is to use the common technique of inductive degeneration [12]. For the common-source amplifier we are investigating, this degeneration is realized as an inductor in series with the source terminal. In implementation, this could be a transmission line, a bond wire, or a lumped component.

To begin evaluating the performance of the degenerated device, we compute the modified S parameters of the transistor under degeneration. The series-series feedback of the inductor attached to the source terminal is equivalent to the sum of the networks' Z parameters, shown in Figure 2.6.

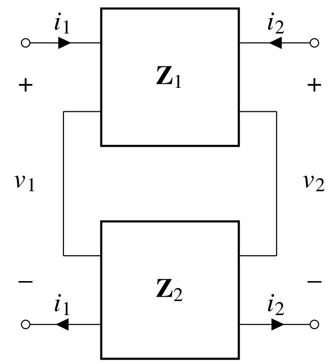


Figure 2.6: Series-Series feedback in Z parameters.

Converting the S parameters to Z parameters with

$$\mathbf{Z} = \frac{Z_0}{\Delta_s} \begin{bmatrix} (1+S_{11})(1-S_{22})+S_{12}S_{21} & 2S_{12} \\ 2S_{21} & (1-S_{11})(1+S_{22})+S_{12}S_{21} \end{bmatrix}, \quad (2.27)$$

where

$$\Delta_s = (1-S_{11})(1-S_{22}) - S_{12}S_{21}$$

and

$$\mathbf{S} = \frac{1}{\Delta_z} \begin{bmatrix} (Z_{11}-Z_0)(Z_{22}+Z_0)+Z_{12}Z_{21} & 2Z_0Z_{12} \\ 2Z_0Z_{21} & (Z_{11}+Z_0)(Z_{22}-Z_0)-Z_{12}Z_{21} \end{bmatrix}, \quad (2.28)$$

where

$$\Delta_z = (Z_{11}+Z_0)(Z_{22}+Z_0) - Z_{12}Z_{21}\,,$$

we take the transistor's S parameters, convert to Z parameters, add the Z parameters of the impedance of the feedback (ZJ_2 where J_2 is the 2x2 unit-matrix), and convert back to S parameters to evaluate the new S parameters and derived stability. The S parameters of this construction are

$$\mathbf{S}_{\text{Degen}} = \frac{1}{Z(S_{11} + S_{12} + S_{21} + S_{22}) - 2Z_0} \left(\begin{bmatrix} \Delta_A & \Delta_B \\ \Delta_B & \Delta_A \end{bmatrix} + 2Z_0\mathbf{S} \right) , \tag{2.29}$$

where

$$\begin{aligned} \Delta_A &= \Delta + S_{12} + S_{21} - 1 \\ \Delta_B &= -\Delta + S_{11} + S_{22} - 1 \, . \end{aligned}$$

We could attempt to substitute Equation 2.29 into Equation 2.26 and solve for the L that forces μ' to unity, but we are unfortunately not guaranteed a solution. An alternative approach is to plot μ' across a range of inductances to explore their relationship. However, the maximum gain of the modified amplifier also changes with degeneration, as expected for an amplifier under feedback. These two metrics are plotted versus the degeneration inductance in Figure 2.7.

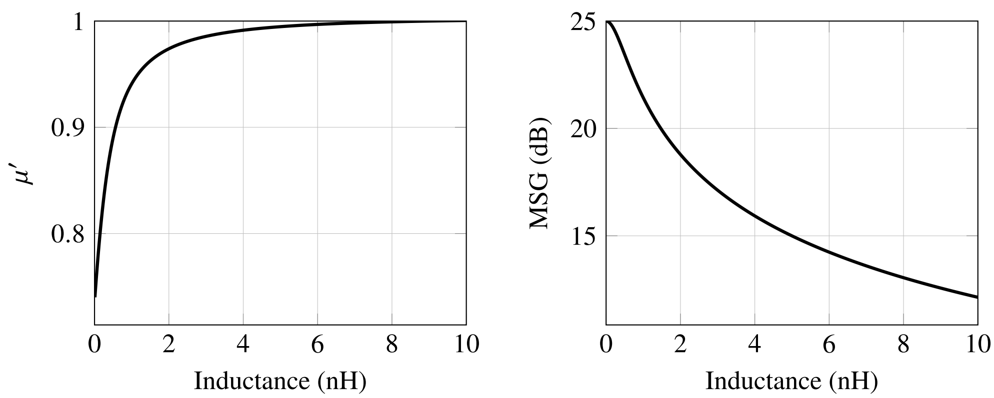


Figure 2.7: μ' and MSG versus inductive degeneration for the 4F50 at 1.5 GHz.

From these results, we determine that to achieve an MSG of 20 dB, we can apply at most 1.5 nH of degeneration. The μ' metric is the distance of the source stability circle from the center of the Smith chart, so we compare this value with the magnitude of the desired Γ_S to evaluate conditional stability. 1.5 nH is a reasonable value for the degeneration inductor, as we will be wire bonding the source pad of the die transistor, and the rule of thumb is about 1 nH of inductance per 1 mm of bond wire length.

Next, we must evaluate the effect of degeneration on the noise parameters. Starting from the Hartmann and Strutt result [13] which describes the changes in noise parameters under feedback, we compute the "noise parameter transformation matrix" as

$$n = \frac{1}{\Delta_n} \begin{bmatrix} \Delta_n & Z_0 Z(\Delta + S_{22} - S_{11} + 2S_{21} - 1) \\ 0 & 2S_{21} Z_0 \end{bmatrix}, \tag{2.30}$$

where

$$\Delta_n = Z(S_{11} - 1)(S_{22} - 1) + S_{21}(2Z_0 - S_{12} Z)$$

under the assumption of series feedback, as is the case for the degeneration inductance as shown earlier. From here, we use the relations

$$R'_n = R_n |n_{11} + n_{12} Y_{\text{Corr}}|^2 + G_n |n_{12}|^2 \tag{2.31}$$

$$G'_n = \frac{G_n R_n |n_{11} n_{22} - n_{12} n_{21}|^2}{R'_n} \tag{2.32}$$

$$Y'_{\text{Corr}} = \frac{R_n (n_{21} + n_{22} Y_{\text{Corr}})(n^*_{11} + n^*_{12} Y^*_{\text{Corr}}) + G_n n_{22} n^*_{12}}{R'_n}, \tag{2.33}$$

where R_n is the noise resistance, G_n is the noise conductance, and $Y_{\text{Corr}} = G_{\text{Corr}} + jB_{\text{Corr}}$ is the correlation admittance, with the noise figure equation here being expressed as

$$F = 1 + \frac{G_n}{G_s} + \frac{R_n}{G_s} |Y_S - Y_{\text{Corr}}|^2, \tag{2.34}$$

with $Y_S = G_S + jB_S$ being the source admittance. This equation for the noise figure uses different noise parameters than those in Equation 2.3, and we convert between the two using results from [14] with

$$Y_{\text{Opt}} = \sqrt{\frac{G_N}{R_N} + G^2_{\text{Corr}}} - jB_{\text{Corr}} \tag{2.35}$$

$$F_{\text{Min}} = 1 + 2R_N (G_{\text{Corr}} + G_{\text{Opt}}), \tag{2.36}$$

and

$$Y_{\text{Corr}} = \frac{F_{\text{Min}} - 1}{2R_N} - G_{\text{Opt}} - jB_{\text{Opt}} \tag{2.37}$$

$$G_N = R_N (G^2_{\text{Opt}} - G^2_{\text{Corr}}). \tag{2.38}$$

Using this result, we recompute the noise behavior of the degenerated device. Given a small degeneration inductance, the resulting noise parameters do not change much. [13] gives an approximation that under the conditions

$$L_{\text{Degen}} \ll \frac{1}{\omega} \left| \frac{2S_{21} Z_0}{(1 - S_{11})(1 - S_{22}) - S_{12} S_{21}} \right| \tag{2.39}$$

and

$$\mathrm{Re}\{n_{12}\} \ll \frac{R_n\,\mathrm{Re}\{Y_{\mathrm{Corr}}\}}{R_n|Y_{\mathrm{Corr}}|^2 + G_n}\,, \tag{2.40}$$

the noise parameters are unchanged. As we cannot make this assumption in general, we will always recompute the noise parameters in our analysis. The degenerated S parameters and Γ_{Opt} are shown in Figure 2.8 along with the optimum noise Γ_{S} under a 20 dB gain constraint, similar to Figure 2.5.

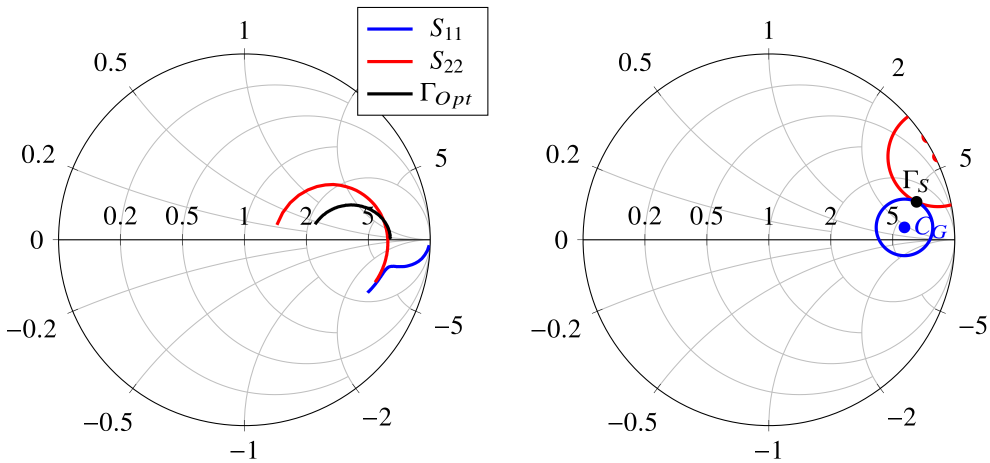


Figure 2.8: S_{11}, S_{22}, and Γ_{Opt} from 0.1 GHz to 10 GHz and noise optimum Γ_{S} at 1.5 GHz under 1.5 nH source degeneration.

While the Γ_S result appears similar to Figure 2.3, the degenerated solution achieves a noise temperature of 9.56 K with a $|\Gamma_{\mathrm{A}}|$ of only 0.195, or an $|S_{11}|^2$ of −14.2 dB, which is excellent for an LNA. Compare this with the Figure 2.3 solution, which had a noise of 11.1 K and $|S_{11}|^2$ of −0.45 dB, while also being unstable. The degenerated performance is an improvement as Γ_{Opt} and S_{11}^* have shifted closer together, at the expense of maximum stable gain.

If this were a single-frequency LNA, we could design a simple matching network to conjugate-match the output and provide the Γ_{S} solved here at the input. Much to the chagrin of the RF engineer, customers and scientists demand bandwidth. Next, we explore how these design techniques apply to wideband design.

2.3 Wideband Design Under Constraints

Using the techniques established thus far, we move to evaluate the wideband behavior of the transistor under investigation. To do so, we solve either the VSWR or gain constrained problem, and plot the resulting noise, gain, and S_{11} versus frequency.

For this section, we focus on the same device and bias point, over the 0.7 GHz to 2 GHz range, matching the target band for the DSA-2000.

We begin by choosing a desired gain constraint of G_A=20 dB. Given the previous analyses, this should be a good compromise between maximizing gain and minimizing noise. We then evaluate MSG over frequency and degeneration inductance to find the maximum degeneration value that allows for an MSG greater than our desired G_A across the band. We are maximizing degeneration to result in a maximally stable amplifier. The plot of MSG versus frequency and degeneration is shown in Figure 2.9. From this result, we observe that to achieve 20 dB of gain over our entire frequency range, we need an inductance of 0.8 nH, less than our 1.5 nH result from before.

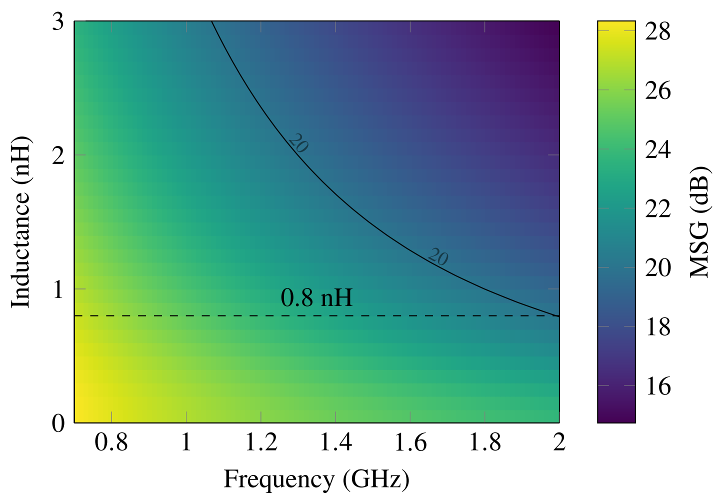


Figure 2.9: MSG vs Degeneration vs Freq and 20 dB contour.

With this degeneration value, we solve for the optimum noise Γ_S under a G_A=20 dB constraint, and compute the remaining parameters such as the input and output return loss and noise.

In Figure 2.10, we see that from 0.7 GHz to 1.75 GHz, we achieve higher gain than the requested value, with Γ_S taking on the value of Γ_{Opt}. After 1.75 GHz, however, we sacrifice noise to maintain our gain requirement. Additionally, we observe that the input return loss (IRL) and output return loss (ORL) are reasonable, notably positive (in loss) indicating this implementation is stable, as we expect from the MSG curve. However, at 2 GHz we are at the limit of stable gain under the gain matching condition where G_A = MSG. This may imply that small perturbations in the design will cause it to oscillate. To build a more robust amplifier, we will want additional margin. Furthermore, the noise gets significantly worse past 1.75 GHz.

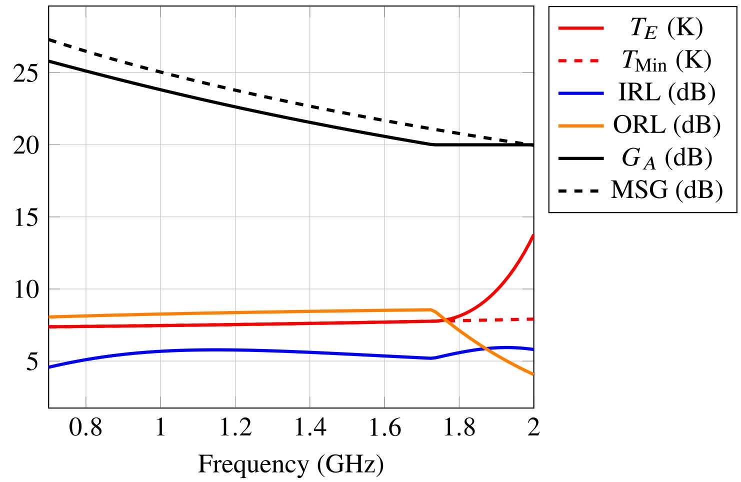


Figure 2.10: Performance under 0.8 nH degeneration with a 20 dB gain constraint.

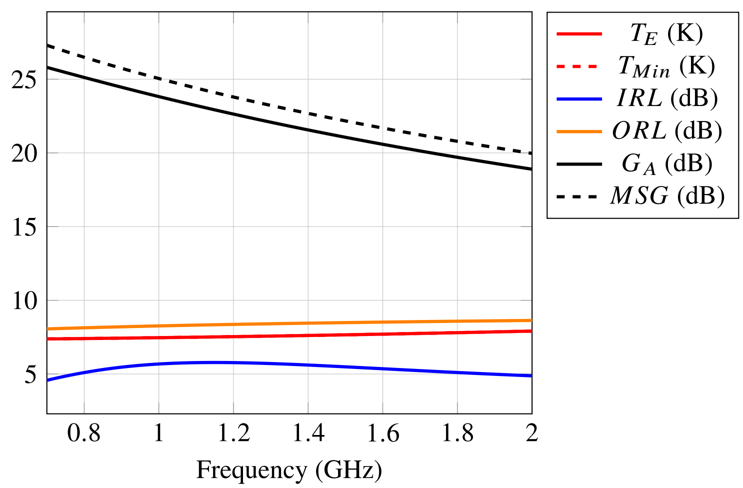


Figure 2.11: Performance under 0.8 nH degeneration with a 18 dB gain constraint.

With the same degeneration value, we then evaluate at a reduced gain constraint of 18 dB. Shown in Figure 2.11, we see that we are meeting our gain requirement across the whole frequency range, implying the optimum Γ_S is simply Γ_{Opt}.

For a radio telescope interferometer, we are usually not particularly interested in the return loss, rather gain and noise. Minimizing noise gives the telescope sensitivity, while gain reduces the impact of the noise generators after the LNA. Poor return loss would result in passband ripple, the effects of which are usually calibrated out. However, in some extreme cases such as 21 cm cosmology, these effects need to be carefully controlled [15]. We could evaluate the noise performance under return loss constraints, but we would see similar results to those shown under gain constraints.

As such, we move our attention to the implementation of the network that performs the match to Γ_{Opt}.

2.4 Wideband Nonuniform Line Matching

From the result in Figure 2.11, we know now we want to create a matching network directly for Γ_{Opt}, under 0.8 nH of inductive source degeneration. Unfortunately, there is not an analytic way to design a matching network for an arbitrary impedance response with frequency. There are many optimization approaches in the literature, perhaps the most popular being the "Real Frequency Technique" (RFT) [16], [17]. In RFT, coefficients for a Laplace-domain polynomial are optimized, minimizing an error function between desired and modelled behavior in the least-squares sense. Solving for these polynomials involves iterative root-finding, on top of the normal iterative gradient-descent optimization. After we have found the optimum Nth order polynomial, we still need to perform a synthesis step where we transform the result into lumped or distributed components. Unfortunately, there is no guarantee during optimization that the resulting implementation is physical. Additionally, the multistep iteration and finite-difference approximation of the gradient is slow and imprecise.

If we happened to know the topology of the matching network a priori, we could attempt to directly optimize it using the same least-squares approach. For the case of this amplifier we are attempting to design, there are very few matching topologies which would be acceptable. This primarily comes from the fact that F_{Min} is very small, so any discrete components will be too lossy and significantly impact our noise. So, we are left with transmission line matching, either as cascaded stubs, cascaded steps, or some combination of the two. One of the more interesting topologies to investigate is the nonuniform transmission line (NTL). These lines vary their impedance continuously as a function of position. NTLs have been widely studied in matching contexts, even for complex loads [18]. As NTLs are still just transmission lines, our matching network loss is driven by the attenuation constant of the implementation, which can be very low, resulting in an exceptionally low-loss match.

Here, we attempt to use NTLs to match to Γ_{Opt} over our wide bandwidth, formulated as a least-squares problem against Γ_{S}. The difficulty in this approach is parameterizing the NTL in a simple way that allows us to compute Γ_{S}. A common tactic is to discretize the line into many uniform sections and cascade their scattering behavior. With many sections, this yields a good approximation to the smooth transmission line.

In optimization, we use a gradient-descent approach, which will require the derivative of our cost function with respect to every input variable. We parameterize the NTL by N lines of characteristic impedance Z_i, each with constant length of δL across M distinct frequencies ν and source reflection coefficient (before the input matching network) of Γ_{S}, and write our cost function as

$$E = \frac{1}{M}\sum_{i=1}^{M}\left|\Gamma_{\mathrm{Out}}\left(\prod_{j=1}^{N} A(Z_j, \delta L, \nu_i), \Gamma_S\right) - \Gamma_{\mathrm{Opt}_i}\right|^2, \tag{2.41}$$

where the final ABCD parameters are converted to S parameters using the standard transformations [5]. As all the expressions thus far are analytic, the gradient for $\delta E/\delta Z_j$ has an analytic form. As N increases, however, this expression will have exponentially more terms. One could use the finite-difference method to compute this derivative, but this is inefficient and imprecise. Instead, we will make use of automatic differentiation to compute the exact gradients. Doing so has significant performance implications, which will be discussed in Chapter 3.

One important note on the cost function in Equation 2.41 is that we will minimize mean squared error to Γ_{Opt}, not noise. The distance between Γ_{S} and Γ_{Opt} is proportional to noise, but the true relationship follows from Equation 2.3, which contains the same geometric error term of $|\Gamma_{\mathrm{S}} - \Gamma_{\mathrm{Opt}}|^2$, but is divided by $1 - |\Gamma_{\mathrm{S}}|^2$. We could use the full noise equation in optimization, however this more simple geometric approximation yields a satisfactory result.

Given a set of Γ_{Opt} with frequency and a total length L, we compute the optimum Z_j versus linear distance profile. In the current formulation, the total length of the line L is not part of the optimization, so we must try several lengths. From [18], longer lines should perform better, although we will be at odds with insertion loss in actual implementation. We start our design with the length set to slightly less than $\lambda/4$ at 700 MHz, as NTLs have been used for miniaturization of quarter wave lines [19]. Assuming the transmission line is in free space, we start at a length of 80 mm. We also constrain the characteristic impedances between 45 Ω and 350 Ω as that is the range of what is practically realizable in suspended microstrip, an exceptionally low-loss topology. The resulting profile and performance is shown in Figure 2.12.

In Figure 2.12, several things are immediately apparent. One, the line is notably not smooth, rather oscillating between the high and low impedance limits. The

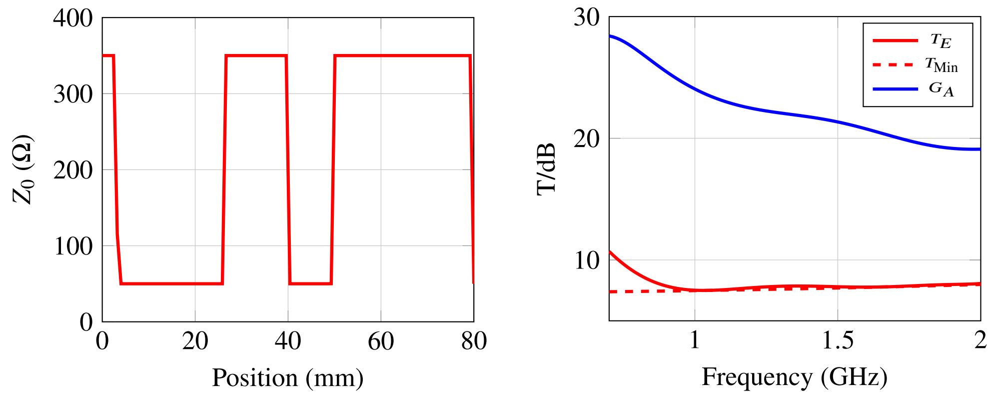


Figure 2.12: Optimal impedance profile, noise, and gain for an 80 mm NTL.

noise match, however, is reasonable, with a band-average of 7.99 K compared to the band-average of T_{Min} of 7.64 K. Unfortunately, the sharp discontinuities break our working assumption of smoothness in the nonuniform lines. These discontinuities create EM effects not captured by the cascaded uniform section model. For example, in microstrip, large step discontinuities such as these can drastically alter circuit behavior and must be accounted for [20]. The optimizer converging on a non-smooth result for an NTL has been shown in the literature [21], and in the limit approximates an R-transformer [22]. We have a few options to account for this behavior. One, we could add in a circuit model of the discontinuity, although this would rely on knowing the transmission line implementation. Second, we could add a constraint to the optimization to account for smoothness. We will choose the latter, as we want to make minimal assumptions about the transmission line topology.

In [23], one option to contend with the stepped behavior of NTL optimization is to add a constraint where every impedance step is no more than some preset limit. For a 100-section line, the implementation of the constraint adds a 99-element constraint vector, with an associated highly sparse constraint Jacobian. This adds a lot of complexity to the optimization problem and makes the performance significantly worse. We propose an alternative of regularizing for total smoothness by adding a Lagrange multiplier to the sum of impedance deltas. With this, the new cost function looks like

$$E = \frac{1}{M}\sum_{i=1}^{M}\left|\Gamma_{\text{Out}}\left(\prod_{j=1}^{N} A(Z_j, \delta L, \nu_i), \Gamma_S\right) - \Gamma_{\text{Opt}_i}\right|^2 + \frac{\lambda}{M}\sum_{i=2}^{M}|Z_i - Z_{i-1}|^2 , \quad (2.42)$$

where λ adjusts the strength of the regularization term.

For the same M=100, L=80 mm problem, but now with regularization for λ=0.005, we compute the result shown in Figure 2.13. While this result is smooth, the gain and noise profiles are worse. The average noise temperature is now 8.53 K, about half a Kelvin worse than the previous result and 1 K worse than the minimum noise. Lowering λ improves this, but of course, with sharper features in the impedance profile.

The performance of the smooth line allows us to conclude that we may be better served with a longer line. This agrees with [18], in that the longer line is similar to more cascaded transformers. Using the same λ=0.005 regularization, we re-run the optimization for a few different lengths shown in Figure 2.13, Figure 2.14, and Figure 2.15. The noise is decreasing with increased length, with the 120 mm design achieving 7.90 K, slightly improved over the 80 mm unregularized solution.

All the smooth solutions are realizable in suspended stripline/microstrip, but we have failed to account for the insertion loss of the matching network itself. As the line grows in length, so does its loss and therefore noise. As the resulting design achieves a noise of 10 K, even 0.15 dB of insertion loss will double the noise. Additionally, the loss of the line is roughly proportional to its impedance, so the more high impedance section line we have, the worse the noise due to the smaller cross-sectional area and associated resistance per unit length. The precise details of optimizing the line considering its loss is the subject of Chapter 3.

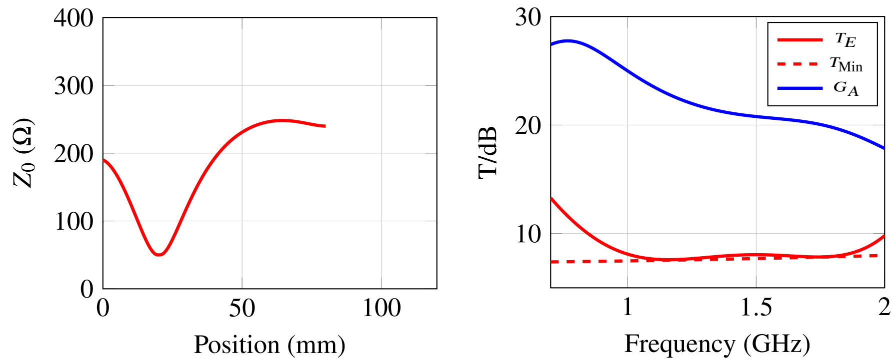


Figure 2.13: Impedance profile, noise, and gain for an 80 mm NTL, λ=0.005.

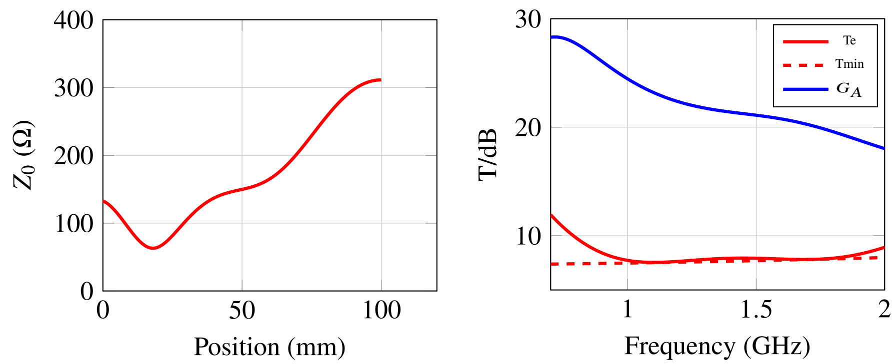


Figure 2.14: Impedance profile, noise, and gain for an 100 mm NTL, λ=0.005.

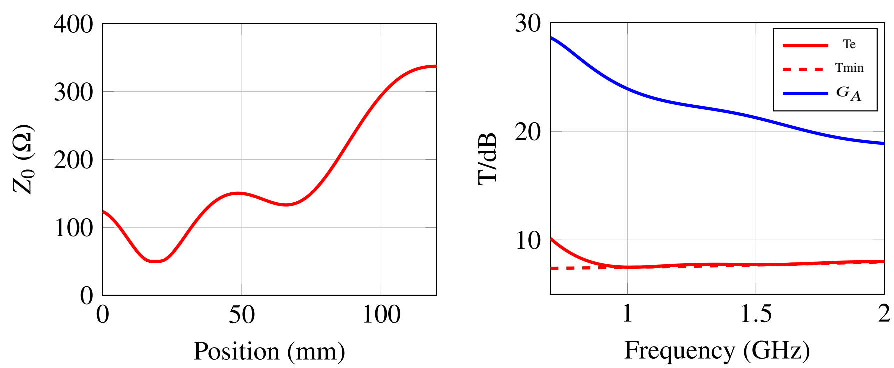


Figure 2.15: Impedance profile, noise, and gain for an 120 mm NTL, λ=0.005.

References

[1] H. Nyquist, “Thermal Agitation of Electric Charge in Conductors,” *Physical Review*, vol. 32, no. 1, pp. 110–113, Jul. 1928. DOI: 10.1103/PhysRev.32.110.

[2] W. Schottky, “Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern,” *Annalen der Physik*, vol. 362, no. 23, pp. 541–567, Jan. 1918. DOI: 10.1002/andp.19183622304.

[3] H. Rothe and W. Dahlke, “Theory of Noisy Fourpoles,” *Proceedings of the IRE*, vol. 44, no. 6, pp. 811–818, Jun. 1956. DOI: 10.1109/JRPROC.1956.274998.

[4] G. Gonzalez, *Microwave Transistor Amplifiers: Analysis and Design*, 2nd ed. Upper Saddle River, N.J: Prentice Hall, 1997, 506 pp.

[5] D. M. Pozar, *Microwave Engineering*, 4th ed. Hoboken, NJ: Wiley, 2012.

[6] G. Link and V. Rao Gudimetla, “Analytical expressions for simplifying the design of broadband low noise microwave transistor amplifiers,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 43, no. 10, pp. 2498–2501, Oct. 1995. DOI: 10.1109/22.466187.

[7] “Diramics pH-200 4F50.” (), [Online]. Available: https://diramics.com/wp-content/uploads/downloads/DIRAMICS-pH-100-4F50.pdf (visited on 02/04/2025).

[8] J. Shi and S. Weinreb, “Room-Temperature Low-Noise Amplifier With 11-K Average Noise From 0.6 to 2 GHz,” *IEEE Microwave and Wireless Technology Letters*, vol. 33, no. 11, pp. 1540–1543, Nov. 2023. DOI: 10.1109/LMWT.2023.3315269.

[9] S. Weinreb and J. Shi, “Low Noise Amplifier With 7-K Noise at 1.4 GHz and 25 °C,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 4, pp. 2345–2351, Apr. 2021. DOI: 10.1109/TMTT.2021.3061459.

[10] J. Rollett, “Stability and Power-Gain Invariants of Linear Twoports,” *IRE Transactions on Circuit Theory*, vol. 9, no. 1, pp. 29–32, Mar. 1962. DOI: 10.1109/TCT.1962.1086854.

[11] M. Edwards and J. Sinsky, “A new criterion for linear 2-port stability using a single geometrically derived parameter,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, no. 12, pp. 2303–2311, Dec. 1992. DOI: 10.1109/22.179894.

[12] P. Leroux and M. Steyaert, “Detailed Study of the Common-Source LNA with Inductive Degeneration,” in *LNA-ESD Co-Design for Fully Integrated CMOS Wireless Receivers*, Boston, MA: Springer US, 2005, pp. 73–110. DOI: 10.1007/1-4020-3191-2_4.

[13] K. Hartmann and M. Strutt, “Changes of the four noise parameters due to general changes of linear two-port circuits,” *IEEE Transactions on Electron Devices*, vol. 20, no. 10, pp. 874–877, Oct. 1973. DOI: 10.1109/T-ED.1973.17761.

[14] P. Heymann and M. Rudolph, “Noise of Linear Two-Ports,” in *A Guide to Noise in Microwave Circuits: Devices, Circuits and Measurement*, IEEE, 2022, pp. 101–108. DOI: 10.1002/9781119859390.ch7.

[15] J. D. Bowman, A. E. E. Rogers, R. A. Monsalve, T. J. Mozdzen, and N. Mahesh, “An absorption profile centred at 78 megahertz in the sky-averaged spectrum,” *Nature*, vol. 555, no. 7694, pp. 67–70, Mar. 2018. DOI: 10.1038/nature25792.

[16] B. Yarman and H. Carlin, “A Simplified "Real Frequency" Technique Applied to Broad-Band Multistage Microwave Amplifiers,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 30, no. 12, pp. 2216–2222, Dec. 1982. DOI: 10.1109/TMTT.1982.1131411.

[17] H. Carlin and P. Amstutz, “On optimum broad-band matching,” *IEEE Transactions on Circuits and Systems*, vol. 28, no. 5, pp. 401–405, May 1981. DOI: 10.1109/TCS.1981.1085001.

[18] G. Xiao and K. Yashiro, “Impedance matching for complex loads through nonuniform transmission lines,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 6, pp. 1520–1525, Jun. 2002. DOI: 10.1109/TMTT.2002.1006413.

[19] M. Khalaj-Amirhosseini, “Nonuniform Transmission Lines as Compact Uniform Transmission Lines,” *Progress In Electromagnetics Research C*, vol. 4, pp. 205–211, 2008. DOI: 10.2528/PIERC08082602.

[20] T. C. Edwards and M. B. Steer, “Discontinuities in Microstrip,” in *Foundations for Microstrip Circuit Design*, IEEE, 2016, pp. 227–267. DOI: 10.1002/9781118936160.ch9.

[21] P. Miazga, “Discrete shape optimization method of a non-uniform transmission line — advantages and drawbacks,” in *MIKON 2008 - 17th International Conference on Microwaves, Radar and Wireless Communications*, May 2008, pp. 1–4.

[22] S. Rosloniec, “Design of stepped transmission line matching circuits by optimization methods,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 42, no. 12, pp. 2255–2260, Dec. 1994. DOI: 10.1109/22.339750.

[23] P. Miazga, “Nonuniform transmission line matching circuits synthesis - Analytical versus optimization approach,” in *2014 20th International Conference on Microwaves, Radar and Wireless Communications (MIKON)*, Jun. 2014, pp. 1–4. DOI: 10.1109/MIKON.2014.6899913.

Chapter 3

COMPUTATIONALLY EFFICIENT DESIGN OF AN LNA INPUT MATCHING NETWORK USING AUTOMATIC DIFFERENTIATION

[1] K. A. Shila, “Computationally Efficient Design of an LNA Input Matching Network Using Automatic Differentiation,” *IEEE Journal of Microwaves*, May 7, 2025. DOI: `10.1109/JMW.2025.3568779`,

In this chapter, we present a method for the design of an LNA input matching network using automatic differentiation (AD), a technique made popular by machine learning. The input matching network consists of a non-uniform suspended stripline transformer, directly optimized with AD-provided gradients. Compared to the standard approach of finite-differences, AD provides orders of magnitude faster optimization time for gradient-based solvers. This dramatic speedup reduces the iteration time during design and enables the exploration of more complex geometries. The LNA designed with this approach improves over a previous two-section uniform-line design, achieving an average noise temperature of (11.53 ± 0.42) K over the frequency range of 0.7 GHz to 2 GHz at room temperature. We optimized the geometry in under 5 s, 40x faster than optimizing with finite-differences.

3.1 Introduction

Radio telescope receivers are typically cooled to give high sensitivity, but the cost of cooling is prohibitive for arrays of thousands of telescopes. Receivers with ambient-temperature LNAs are becoming competitive with cooled designs at low frequencies, resulting in dramatically decreased cost that enables the development of enormous arrays such as the DSA-2000 [1]. A key metric for these arrays is the speed at which they create images of the sky. The observation time required for some fixed noise in an image is $\propto T^2$, where T is the system noise temperature. At decimeter wavelengths, LNA noise typically dominates the system noise [2], so even 1 K improvement in the LNA significantly increases the survey speed and subsequent science return. The key LNA development has been the use of a low-noise transistor from Diramics [3] with a low-loss suspended stripline input matching network [4], [5], consisting of uniform transmission line (UTL) transformers. We have improved the LNA noise temperature by changing the input matching network to a non-uniform transmission line (NTL).

It is well known that NTLs are useful in wideband matching and are easily implemented in suspended stripline. For our LNA, an NTL gives a better trade-off between insertion loss and noise match, but designing such a line is nontrivial. The geometry must be directly optimized in situ, as no closed-form expression for the desired behavior of the NTL exists.

To optimize the geometry of the NTL, we follow the standard practice of gradient descent. In most electronic design automation (EDA) packages, the gradient is approximated via finite-differencing, but finite-differencing is imprecise and inefficient for large-dimensional optimization, as is the case for an NTL problem. While direct optimization is a common technique in designing NTLs [6]–[8], the performance overhead of computing the gradient is a limiting factor.

Instead of computing the gradient with finite-differencing, we used automatic differentiation (AD) [9], an algorithmic tool made popular by machine learning. AD is an accelerated way to compute the exact derivatives of an arbitrary computer program. As AD is an exact method, it is more precise than finite-differencing and typically higher performance. Gradient-descent optimization of NTLs with AD dramatically reduces optimization time, which enables more complex designs.

In this chapter, we describe the design, implementation, and experimental verification of an LNA whose input matching network is an NTL designed via direct optimization using AD. The NTL design achieves lower average noise temperature than the UTL design. Optimization of the NTL took under 5 s, 40x faster than solving the equivalent finite-differenced gradient-based problem in a commercial EDA package.

In Section 3.2, we discuss this mathematical formulation in detail with an overview of the LNA design problem posed as constrained nonlinear optimization, the AD method, and NTL design. In Section 3.3, we discuss the implementation details of applying this technique to our LNA design. Finally, in Section 3.4, we present experimental results to demonstrate the performance of the amplifier.

3.2 Mathematical Formulation

Low-Noise Amplifier Optimization

The priority for any LNA design is to minimize the amplifier's noise. Balancing this priority with other requirements such as input reflection, stability, gain, etc. is application-specific and is stated mathematically as

$$\begin{aligned} \underset{\Gamma_{\mathrm{S}}}{\text{minimize}} \quad & F(F_{\mathrm{Min}}, R_n, \Gamma_{\mathrm{Opt}}, \Gamma_{\mathrm{S}}) \\ \text{subject to} \quad & G(\mathbf{S}, \Gamma_{\mathrm{L}}, \Gamma_{\mathrm{S}}) \leq \xi \,, \end{aligned} \tag{3.1}$$

where F is the noise objective function, given here in the standard form of noise figure [10], driven by the device's noise parameters via

$$F = F_{\mathrm{Min}} + \frac{4R_n \left|\Gamma_{\mathrm{S}} - \Gamma_{\mathrm{Opt}}\right|^2}{Z_0 \left|1 + \Gamma_{\mathrm{Opt}}\right|^2 \left(1 - |\Gamma_{\mathrm{S}}|^2\right)} \,, \tag{3.2}$$

where F_{Min} is the minimum noise figure, R_n is the noise equivalent resistance, and Γ_{Opt} is the optimum source reflection coefficient that results in F_{Min}. G is the constraint, a function of the transistor's S parameters, $\mathbf{S}$, and input and output reflection coefficients, Γ_{S} and Γ_{L}, shown in Figure 3.1. The constraint function captures any of the non-noise requirements and is represented by an inequality with some scalar ξ. The solution to (3.1) is the Γ_{S} that minimizes the noise figure given a biased transistor's S and noise parameters under some constraint. Notably, the solution gives minimum noise without specifying any goal or target. It is then up to the designer to evaluate the result and determine if it is satisfactory, loosening the constraints as necessary. This is distinct from "goal-based" optimization, where the optimizer attempts to solve a potentially unfeasible problem and returns a goal-weighted least-squares solution.

If we were only interested in constraining input reflection, we would define ξ as a maximum allowed magnitude of Γ_{A} and write our constraint function as

$$G \leq |\Gamma_{\mathrm{A}}|_{\mathrm{Max}} \,. \tag{3.3}$$

Alternatively, we could constrain the gain to be greater than some value, either case being a trade-off between power and noise match. In this work, we will constrain $|\Gamma_{\mathrm{A}}|$.

The optimal Γ_{S} solution, however, is only valid at a single frequency and could be readily solved by hand. In practice, we sample the device's S and noise parameters across a set of N discrete frequency points, $\boldsymbol{\nu}$, and our goal is to design an amplifier that operates over a subset of that range. For this wideband LNA design, one formulation is to minimize the average noise

$$\begin{aligned} \underset{\Gamma_{\mathrm{S}}(\boldsymbol{\nu})}{\text{minimize}} \quad & \frac{1}{N}\sum_{i=1}^{N} F(\nu_i) \\ \text{subject to} \quad & G(\boldsymbol{\nu}) \leq |\Gamma_{\mathrm{A}}|_{\mathrm{Max}} \,. \end{aligned} \tag{3.4}$$

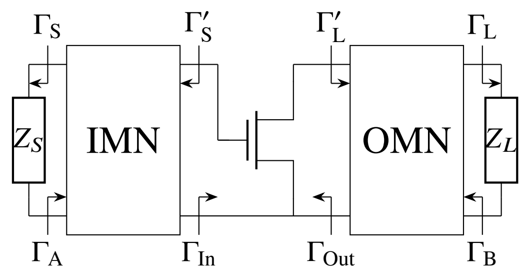


Figure 3.1: Reflection planes for a typical microwave transistor amplifier, its input matching network (IMN), and output matching network (OMN).

Here, the minimizer and constraints are vector-valued. In this form, (3.4) is a constrained, weighted, nonlinear least-squares problem, as minimizing the average noise figure over $\Gamma_S(\boldsymbol{\nu})$ is proportional to minimizing the R_n-weighted mean squared distance between $\Gamma_S(\boldsymbol{\nu})$ and $\Gamma_{Opt}(\boldsymbol{\nu})$, subject to the input reflection constraint.

In practice, the $\Gamma_S(\boldsymbol{\nu})$ solution is not directly useful, as ultimately, we aim to design the network that induces $\Gamma_S(\boldsymbol{\nu})$. Therefore, instead of minimizing over $\Gamma_S(\boldsymbol{\nu})$, we minimize over the input matching network directly. If, for example, this matching network were a series UTL with electrical length l, phase constant β, and characteristic impedance Z_L, the problem becomes

$$\begin{aligned} &\underset{l,\, Z_L}{\text{minimize}} && \frac{1}{N}\sum_{i=1}^{N} F(\Gamma_S(\nu_i)) \\ &\text{subject to} && G(\Gamma_S(\boldsymbol{\nu})) \leq |\Gamma_A|_{Max} \ , \end{aligned} \tag{3.5}$$

where

$$\Gamma_S(\nu) = Z_0 \left[\frac{Z_L + jZ_0 \tan(\beta(\nu)l)}{Z_0 + jZ_L \tan(\beta(\nu)l)} \right] . \tag{3.6}$$

Here, we reduce the dimensionality of the problem from N to two as we are solving for a single length and characteristic impedance. However, there is no guarantee that this topology achieves the same average noise figure as the optimum $\Gamma_S(\boldsymbol{\nu})$ solution. Solving this transformed problem is now akin to regression, as the structure of the matching network dictates the frequency response and the solver will try to "curve fit" that frequency response to the optimum $\Gamma_S(\boldsymbol{\nu})$ solution. Selecting an appropriate network topology is therefore crucial for achieving optimal results. In this work, we investigate NTLs as a general-purpose topology.

With a chosen a topology, we solve the minimization problem with gradient-based nonlinear optimization. The inclusion of an inequality in the constraint (3.3) implies the solver must contend with a system of Karush-Kuhn-Tucker (KKT) conditions [11] and therefore must be solved numerically. Gradient-based KKT solvers will evaluate the cost function, the constraints, and the gradient/Jacobian potentially thousands of times. Computing these values efficiently is essential to the performance of this design method.

Automatic Differentiation

Traditionally, there are three ways one can generate gradients for optimization. First, if the cost expression is simple, one could solve the gradient by hand by applying the rules of calculus. However, if the expression were large or if it were a black-box with unknown internals, deriving an expression for the gradients might be impossible. The second approach is to approximate the gradient via finite-differencing each term using the standard form,

$$\frac{df}{dx} \approx \frac{f(x+h) - f(x)}{h}\,. \tag{3.7}$$

This is a simple formulation which works on any well-behaved function and is the solution of choice for most EDA software. However, these approximations suffer from numeric instability and high sensitivity to the choice of the step size h [12].

The third option is to use a computer algebra system to perform “symbolic differentiation”. This method applies the rules of calculus to an expression tree and is a simple mechanistic process for computers. However, the complexity of computing a symbolic result grows exponentially with the number of terms, resulting in a problem known as “expression swell” [13]. For optimization, we are not interested in the symbolic form as we only need an accurate numeric gradient to know the direction of steepest descent, so this extra computation is not useful.

Automatic differentiation (AD) is a technique distinct from the aforementioned approaches, where a program algorithmically computes derivatives via the accumulation of values during execution. This method is not an approximation like (3.7), nor does it generate expressions like symbolic differentiation. While there are several implementation strategies, the general formulation yields accurate gradients to machine precision with a small, constant overhead. As AD operates on values rather than expressions, it can be applied to most programs, including those that utilize loops, branching statements, and recursion. Efficient implementations of AD have been transformational in the machine learning community, where large, complex models composed of arbitrary code can be differentiated and optimized [14]. From the AD perspective, a program that computes circuit behavior (noise, gain, etc.) is no different from a program that implements a machine learning model, as both are composed of simple, differentiable pieces.

The application of AD to circuits is not a new idea, but is surprisingly underutilized. As noted in [15], “[AD] can enable applications previously deemed impractical”. While AD has not had much exposure outside of machine learning, it has been recently used for parameter estimation of lumped-element circuit models [16] and sensitivity analysis of EM structures [17]. Some older works conclude the performance overhead may not justify use over finite-differences, but the explosion in machine learning research has accelerated the development of high-performance AD libraries.

AD operates by decomposing a complex function into a sequence of simple operations with known derivatives. As the function is evaluated, AD propagates these derivatives via the chain rule. The chain rule states that each subsequent derivative is multiplied by the previous. As multiplication is associative, we can compute the total derivative from input to output or vice versa. This leads to the two modes of AD: forward-mode and reverse-mode.

In forward-mode, derivatives are computed as the function is evaluated, starting with the input independent variables. Each operation in the function is evaluated along with its derivative with respect to the input, which propagates forward to compute the final derivative. This method is efficient when the function has few inputs and many outputs, as each input’s derivative must be propagated through the function one at a time, similar to finite-differences. In reverse-mode, the function is first evaluated normally, and then derivatives are propagated backward from the outputs to inputs, using the chain rule to accumulate derivatives with respect to each input. This is identical to the familiar “backpropagation” algorithm used in training neural networks. Reverse-mode is particularly efficient for functions with few outputs and many inputs, such as machine learning, where the goal is often to compute gradients of a scalar loss function with many parameters. The same is true for this work, where circuit optimization manifests as high dimensional optimization of a scalar cost function. In either mode, AD computes exact derivatives and is more computationally efficient compared to traditional numerical differentiation methods. A more complete mathematical treatment of AD is given in Appendix A.

Non-Uniform Transmission Lines

Non-uniform transmission lines (NTLs) are a class of transformer structures used in many microwave circuits. These transmission lines vary their impedance continuously as a function of position to achieve such goals as matching [18], filtering [19], or miniaturization [20]. For ambient-temperature LNAs designed around very low noise transistors, minimizing the loss of the input matching network is essential. As discussed in [4], suspended stripline matching structures provide the lowest loss. Implementing the suspended stripline as an NTL is a natural development of the stepped transformer in [5] and empirically results in a lower noise temperature. Other strategies such as lumped matching are simply too lossy and are not considered for this work.

Unfortunately, the scattering behavior for a general NTL has no analytic form. This arises from the fact that for a general transmission line, the voltage and current as a function of position are governed by a set of Riccati differential equations [21]. While closed-form solutions exist for special cases, such as exponential tapers [22], a standard approach for analyzing arbitrary NTLs is to discretize the line into many small lengths of constant impedance sections.

We represent the discretized approximation of an NTL as the multiplication of ABCD (cascade) matrices of the uniform sections [23]. For a uniform transmission line of length l with characteristic impedance Z_c and propagation constant γ, this matrix is

$$\mathbf{A}\left(Z_c, \gamma, l\right) = \begin{bmatrix} \cosh \gamma l & Z_c \sinh \gamma l \\ 1/Z_c \sinh \gamma l & \cosh \gamma l \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} . \tag{3.8}$$

For an NTL defined by a characteristic impedance and propagation constant profile with position, $Z_c(x), \gamma(x)$, and total length L, we pick a discretization of M points and compute

$$\mathbf{A}_{\mathrm{NTL}} \approx \prod_{i=1}^{M} \mathbf{A}\left(Z_c\left(iL/M\right), \gamma\left(iL/M\right), L/M\right) . \tag{3.9}$$

We convert this final ABCD matrix into scattering parameters via

$$\mathbf{S} = \frac{1}{\zeta} \begin{bmatrix} A + B/Z_0 - CZ_0 + D & 2(AD - BC) \\ 2 & -A + B/Z_0 - CZ_0 + D \end{bmatrix} , \tag{3.10}$$

where

$$\zeta = A + B/Z_0 + CZ_0 + D , \tag{3.11}$$

and Z_0 is the reference impedance for the S parameters and reflection coefficients in the design.

To design an NTL for some particular function, we can optimize the set of discrete impedances that approximate the NTL [6]. In this case, we are solving the “inverse

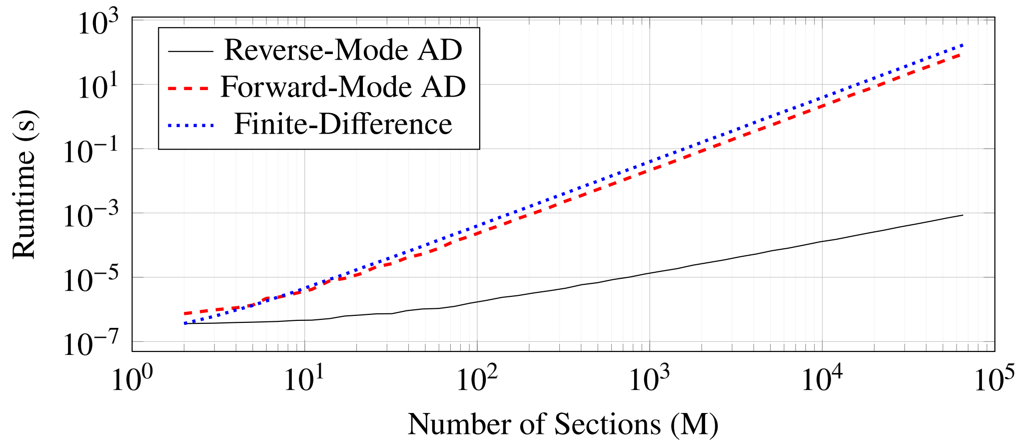


Figure 3.2: Experimental verification of runtime speed of gradient computation for various methods versus NTL discretization.

problem" of finding the discretized $Z_c(x)$ that yields some behavior. To perform this optimization, we take the set of impedances, compute the two-port scattering behavior via (3.9) and (3.10), and use this result to compute an error function against the desired behavior. For gradient-based optimization, we need the gradient of this error function with respect to the set of impedances.

For an M-section discretization of an NTL, a program needs to evaluate $M - 1$ matrix multiplications. Then, for the finite-difference method (or forward-mode AD), the program evaluates the sensitivity of the cost function with respect to every M section. These combined operations, performed for every iteration of an optimizer, have an algorithmic complexity of $O(M^2)$. This scaling dramatically increases the optimization time of designs with large M. However, large M is desirable as it reduces error due to discretization. With reverse-mode AD, which "back-propagates" the sensitivity of the cost function to the inputs, the complexity of computing the gradient is constant over M. This results in an optimization complexity of $O(M)$.

The massive speedup of computing the gradients using reverse-mode AD results in orders of magnitude faster NTL optimization. To demonstrate this, we computed the gradient of a simple scalar cost function ($|S_{11}|^2$) with respect to an M-discretized NTL of constant length. The runtime of this computation versus M is shown in Figure 3.2. For a typical M of 100 sections [24], reverse-mode AD yields gradients two orders of magnitude faster than forward-mode and without the numerical error of finite-differencing.

The utility of this direct-optimization method extends beyond solving the $Z_c(x)$

profile. To implement the NTL, one would use the $Z_c(x)$ result to synthesize physical dimensions. This problem is ill-posed, as realistic transmission lines have coupled Z_c and γ, both of which are functions of frequency. A simpler approach is instead to optimize the geometry directly [7], [8]. The geometry induces a Z_c and γ, which is then used in (3.9).

Recent work [25] has shown another strategy, using convolutional neural networks to capture the geometry-driven behavior of an NTL which is then used for optimization-based inverse design. However, this method is resource-intensive as it relies on many full-wave simulations of "example" lines as well as network training. While the simple cascade method that we use suffers from the M-discretization, it does not require training, or lengthy EM simulations, and is therefore much faster.

To perform the direct geometry optimization, we start with the per-unit-length equivalent resistance, inductance, conductance, and capacitance (RLGC) model of a transmission line with

$$\gamma = \sqrt{(R + j\omega L)\,(G + j\omega C)}\,, \tag{3.12}$$

and

$$Z_c = \sqrt{(R + j\omega L)\,/(G + j\omega C)}\,. \tag{3.13}$$

This model captures all the effects of loss and dispersion. While some transmission lines like coax have an analytic RLGC form, in general, we use curve-fit models. These models can be frequency-dependent, and are readily extracted from measured data or EM simulations [26]. For suspended stripline, we parameterize this model by widths of the center conductor $\mathbf{W}$ and total length L, as the mechanical design dictates fixed height, dielectric constant, etc. The model also enables constraints on geometry like maximum/minimum length or width. These constraints should be kept loose to maximize the likelihood of finding an optimal design.

We apply the NTL direct optimization method to our LNA problem via

$$\begin{aligned} \underset{L,\,\mathbf{W}}{\text{minimize}} \quad & \frac{1}{N}\sum_{i=1}^{N} F(\Gamma_{\mathrm{S}}(\nu_i, \mathbf{W}, L)) \\ \text{subject to} \quad & G(\Gamma_{\mathrm{S}}(\boldsymbol{\nu}, \mathbf{W}, L)) \le |\Gamma_{\mathrm{A}}|_{\mathrm{Max}}\ , \end{aligned} \tag{3.14}$$

where Γ_{S} follows (3.10). Here, the geometry of the NTL is the domain over which we solve the nonlinear least squares problem from before. This benefits our LNA design in that we simultaneously account for both the match to Γ_{Opt} and the insertion loss of the line itself. This proves to be critical to the design, as the insertion loss of the input matching network is close in magnitude to the minimum noise figure of the transistor. The direct optimization approach balances the two sources of noise to result in an optimal design.

3.3 Implementation

To design the improved amplifier, we started with the design from [5]. We make use of the same 400 nm x 200 µm gate, four-finger transistor from Diramics [3] at a slightly lower bias of V_d=0.5 V, I_d=16 mA to match the vendor's updated model. To improve linearity, we changed to a two-stage design versus the three-stage design of the prior work. We also replaced the second stage with a higher current transistor with near identical noise, the SAV-541+ from Mini-Circuits. As the first stage provides more than 20 dB of gain, the noise of the amplifier is practically independent of these modifications. These changes result in an amplifier with better linearity and slightly lower gain. Lastly, we borrow the general chassis geometry for the suspended stripline input matching network. Apart from these similarities, the designs are independent. The goal for this design was to improve the noise while maintaining similar gain and S_{11} over the same bandwidth of 0.7 GHz to 2 GHz.

For the suspended stripline input, we chose the dimensions to satisfy a few conditions. First, the total height of the completed amplifier must be ≤20 mm to satisfy a mechanical constraint for the telescope feed design. Second, the substrate must be close to the centerline of the amplifier to accommodate connectors. Finally, the distance between the substrate and the top and bottom ground planes should be close to equal to minimize the average current density and therefore insertion loss. The large substrate to ground spacing allows for the construction of very high impedance sections, as is required for this particular design due to the transistor's high Z_{Opt}. The resulting model cross-section geometry is shown in Figure 3.3. The substrate material is 0.508 mm-thick RT/duroid 5880 with 0.5 oz cladding. We chose a surface finish of electroless palladium / immersion gold (EPIG) over rolled copper to minimize surface roughness, to eliminate lossy nickel, and to provide a suitable surface for wire bonding.

We drew the stripline geometry in ANSYS's 2D Extractor, and solved for the RLGC parameters as a function of frequency from 0.1 GHz to 3 GHz in 51 uniform steps and as a function of trace width from 0.05 mm to 15 mm swept logarithmically with 111 steps per decade.

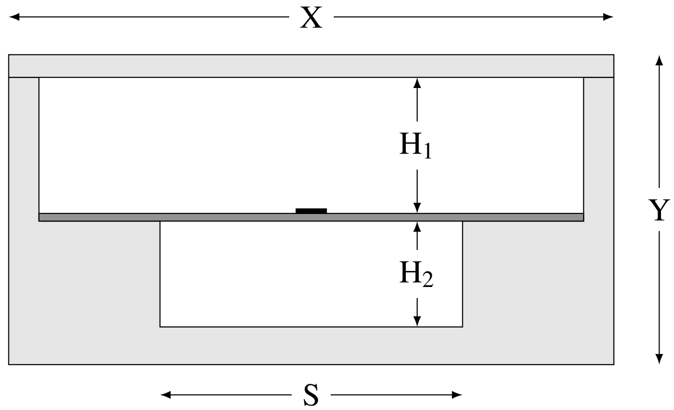


Figure 3.3: Cross-section of the suspended stripline with critical dimensions. X=33, Y=20, S=20, H_1=9.5, H_2=7, all values in mm.

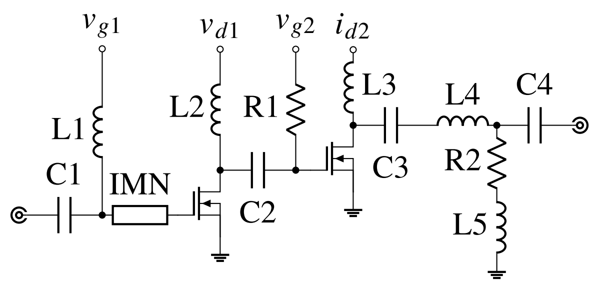


Figure 3.4: Schematic of the RF-portion of the amplifier. C1=3x47 pF, C2=2.5 pF, C3=2 pF, C4=20 pF, L1=100 nH, L2,L3=120 nH, L4=2 nH, L5=4.7 nH, R1=5.1kΩ, R2=41.2 Ω.

Next, we designed the section of the amplifier following the input matching network in Keysight's Advanced Design System (ADS). We used models of lumped components from Modelithics and the Diramics-provided model of the first stage transistor. This section of the design was unremarkable and used standard techniques. We used resistive matching to achieve wideband, high output return loss. We accomplished interstage matching with a single series capacitor. We designed the second stage active bias following the vendor's recommendation. Once we completed this design, we created a layout to perform a full wave EM co-simulation with the component circuit models, a crucial step in solving layout-related stability issues. Finally, we exported the partial amplifier S and noise parameters for the final optimization step of this work. The schematic for the amplifier is shown in Figure 3.4.

We wrote the NTL optimization code in the Julia programming language [27], an open-source, high-level, high-performance language for scientific computing. This language has an extensive ecosystem in both AD and in nonlinear optimization.

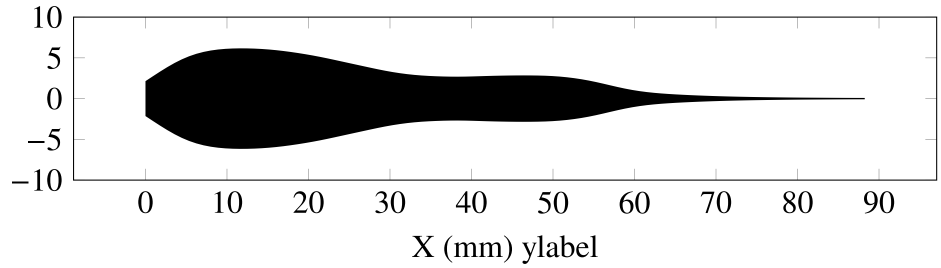


Figure 3.5: Optimized NTL input matching network.

Further, the AD implementations in Julia typically outperform implementations in other languages.

In Julia, we collected the RLGC data from the 2D EM simulation, as well as the S and noise parameters from the circuit simulation. We used the RLGC data to create a linear interpolation across width and frequency for each parameter. We constructed the optimization problem with *Optimization.jl*. This package exposes a structure for defining arbitrary, constrained optimization problems that provide their gradients using a pluggable AD backend. We used the state-of-the-art package *Enzyme.jl* [28] as the reverse-mode AD backend as-is with no modifications. We limited the optimization domain to widths between 0.1 mm and 15 mm and lengths between 50 mm and 120 mm. We set the constraints such that the input return loss was greater than ~9 dB for every point in frequency, representing a compromise between match and noise.

We solved the optimization problem for 100 sections, with an initial profile of a 100 mm-length linear taper from 15 mm to 0.1 mm using the *Ipopt* solver [29]. On an Intel Xeon Gold 6128 workstation, 50 iterations were completed in under 5 s. As gradient-based optimization only finds a local minimum, we ran the optimizer from different starting vectors to compare a few different results. The problem was robust against initial conditions, as solving with different starting profiles resulted in similar line shapes of equal performance. In the worst initial case of random widths, the solver took around 2x longer to converge. However, we know a priori the transformer should transition from 50 Ω to the high Z_{Opt}, so a linear profile is a reasonable initial condition. The resulting shape is shown in Figure 3.5.

Finally, we simulated the resulting geometry in a full-wave 3D EM solver to validate performance. Following this simulation, we removed a section of the line to place low-loss DC-blocking capacitors and added a high-Z stub for the DC gate bias. This

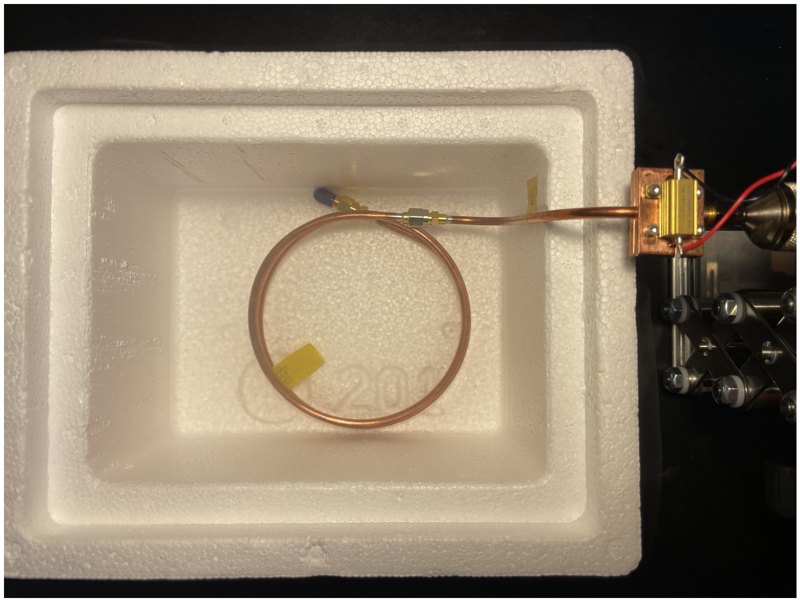


Figure 3.6: LN_2 cold termination measurement setup.

geometry was added at the lowest-Z section of the NTL to minimize the perturbation. We re-simulated the geometry to verify the performance with the DC block and bias structures.

To compare the presented AD method to a commercial EDA tool, we constructed the equivalent optimization problem in ADS, using the same interpolated RLGC data. To prevent ADS from re-simulating the circuit following the NTL, we optimized against the exported S and noise parameters. The timing results are shown in Table 3.1 for various finite-difference (FD) and gradient-free (GF) methods. All trials achieved the same noise temperature within 0.1 K, with the AD approach around 40x faster than the next best solver in ADS.

Table 3.1: Runtime Comparison of NTL Optimization Methods

Solver	Runtime (s)	Relative Time
Ipopt (Reverse-Mode AD)	**5**	**1.0**
Gradient Descent (FD)	193	38.6
Quasi-Newton (FD)	1718	343.6
Simulated Annealing (GF)	3026	605.2
Genetic (GF)	25200	5040.0

3.4 Experimental Results

Noise Measurement Procedure

Since the amplifier in this work is incredibly low noise, we measured the noise using the Y-factor method with a termination in liquid nitrogen (LN_2) and a separate room-temperature termination. If we know the atmospheric pressure, then we precisely know the boiling point of LN_2. Then, given the temperature of the cold termination, we need to consider the length of coax that exits the LN_2 bath, its frequency-dependent insertion loss, and its temperature gradient. As the thermal conductivity of most materials is dependent on their temperature, the temperature gradient along the length of the coax line is nonlinear and must be solved numerically, as described in [30].

Next, we must consider the thermal load on the cold termination. Even when submerged, heat from the warm side of the line will travel into the termination. To reduce this effect, we could use a stainless steel line, as it has low thermal conductivity. However, the temperature-dependent conductivity (and therefore insertion loss) of stainless steel is not well known. Instead, we follow the submerged termination with a long length of line before exiting the bath. In our case, we used an 8 cm section of RG402 coax connected to a submerged 1 m coil and termination. We used a heater, shown in Figure 3.6, to maintain the warm end of the coax at ambient temperature. Finally, we attached a previously characterized N to SMA adapter to the warm end of the cold load to mate with the N-type socket on the constructed LNA. This setup, shown in Figure 3.6, has a frequency-dependent output noise following the curve-fit result

$$T_{\text{Cold}}\,(\text{K}) = 78.0824 + 1.2425 f_{\text{GHz}} - 0.1166 f_{\text{GHz}}^2 \,. \tag{3.15}$$

The error on this result depends on uncertainty in the length of the coax between the surface of the LN_2 and the heater (nominally 8 cm), the warm-side temperature (nominally 297 K), and the atmospheric pressure (nominally 100 kPa). Assuming worst-case errors of ±1 cm of length, ±2 C in temperature, and ±0.5 kPa of pressure, we derive an error model in frequency,

$$\delta T_{\text{Cold}}\,(\text{K}) = 0.0776 + 0.0568 f_{\text{GHz}} - 0.0048 f_{\text{GHz}}^2 \,. \tag{3.16}$$

Finally, we must consider the errors due to changing source impedance between the hot and cold states. These errors are present in all Y-factor measurements but are exacerbated in our LN_2 approach as there are two distinct terminations. The impact

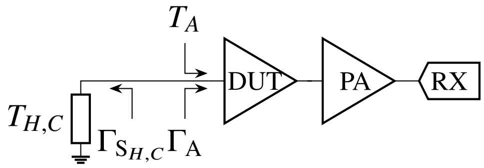


Figure 3.7: System schematic of the Y-factor measurement. The device under test (DUT) is presented a termination with reflection coefficients Γ_{S_H} and Γ_{S_C} at two temperatures T_H and T_C, respectively. The DUT is followed by a series of preamplifiers (PA) before being digitized by the receiver (RX). The total noise temperature referred to the input of the DUT is T_A. The reflection coefficient at the input of the device under test is Γ_A.

on the measured Y-factor stems from two effects: the shift in the amplifier's effective input noise due to its noise parameters (3.2) and changes in available gain.

First, to use the Y-factor method, we must assume that the amplifier's noise remains constant between the two states, implying that the impact of the noise parameters is negligible. Given that we measured both terminations to have return loss ≥25 dB and simulated $R_n \leq 2\,\Omega$ and $|\Gamma_{\text{Opt}}| \leq 0.6$, we expect no worse than 0.03 K error due to this assumption.

Next, we evaluate the impact due to changes in available gain. Given a termination at the input of a noisy receiver, shown in Figure 3.7, the output noise is given by

$$P = k_b B D_{\text{RX}} G_{\text{A}} \left(T_{H,C} + T_{\text{A}} \right) , \tag{3.17}$$

where k_b is Boltzmann's constant, B is the noise equivalent bandwidth, D_{RX} is a scaling term capturing other digital and analog gain factors in the receiver, $T_{H,C}$ is the hot or cold noise temperature at the amplifier input, $T_{\text{A}} = T_{\text{A}}(\Gamma_{\text{S}_{H,C}})$ is the equivalent input noise temperature, and $G_{\text{A}} = G_{\text{A}}(\Gamma_{\text{S}_{H,C}})$ is the available gain of the amplifier given by

$$G_{\text{A}} = \frac{|S_{21}|^2 \left(1 - |\Gamma_S|^2\right)}{|1 - \Gamma_S \Gamma_A|^2 \left(1 - |S_{22}|^2\right)} . \tag{3.18}$$

We write our expression for the measured Y-factor as a ratio of these hot and cold powers [31], but now without the assumption that G_{A} remains constant due to the impact of $\Gamma_{\text{S}_{H,C}}$,

$$Y = \frac{P_{\text{H}}}{P_{\text{C}}} = \frac{G_{\text{A,H}} \left(T_{\text{H}} + T_{\text{A}}\right)}{G_{\text{A,C}} \left(T_{\text{C}} + T_{\text{A}}\right)} = C \left(\frac{T_{\text{H}} + T_{\text{A}}}{T_{\text{C}} + T_{\text{A}}} \right) , \tag{3.19}$$

where the correction C due to mismatch error [32] is

$$C = \frac{1 - |\Gamma_{S_{\text{H}}}|^2}{1 - |\Gamma_{S_{\text{C}}}|^2} \frac{|1 - \Gamma_{S_{\text{C}}} \Gamma_A|^2}{|1 - \Gamma_{S_{\text{H}}} \Gamma_A|^2} . \tag{3.20}$$

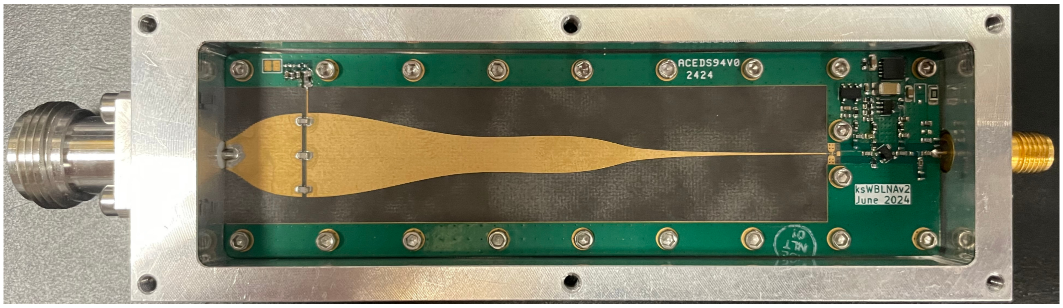


Figure 3.8: The assembled amplifier in a machined aluminum enclosure with terminations and the lid removed.

Solving (3.19) for T_A yields

$$T_A = \frac{T_H - T_C Y/C}{Y/C - 1} . \tag{3.21}$$

This equation for noise temperature is identical to the standard expression, but with a correction term applied to Y. Usually, this factor is close to unity, but as our amplifier is primarily designed for noise and not return loss, and the noise is exceptionally low, the interaction between Γ_A and Γ_S results in a nontrivial correction. For our amplifier and terminations, C is between 0.974 and 1.019, representing a change in the computed noise of about 2 K to 4 K, or a third of the total expected noise.

Additionally, we have uncertainty in the reflection coefficient measurements, which results in uncertainty in this correction factor. Following the datasheet for our network analyzer [33], we create a curve-fit function to describe the uncertainty in both phase and magnitude, as a function of magnitude. For our frequency range and IF bandwidth, these are

$$\delta|\Gamma| = 0.004 + 0.006|\Gamma| \left(1 + |\Gamma|\right) , \tag{3.22}$$

and

$$\delta \arg(\Gamma) = \frac{0.25^\circ}{|\Gamma|} + 0.5^\circ . \tag{3.23}$$

Finally, we need to remove the noise of the receiver. We do so by performing a measurement with a room temperature termination on the receiver's input and subtracting the linear power from the hot and cold powers of the device under test. We performed all these calculations also in the Julia language, making use of the *Measurements.jl* [34] package to propagate uncertainty.

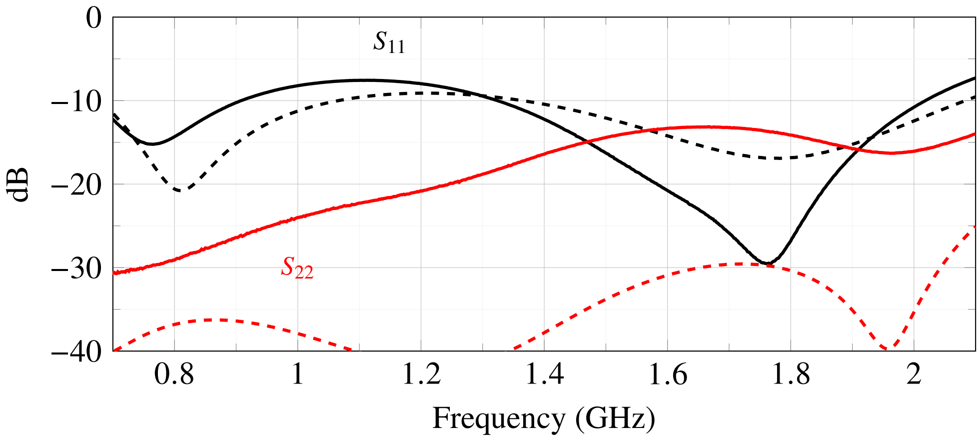


Figure 3.9: Measured (solid) versus simulated (dashed) S parameters.

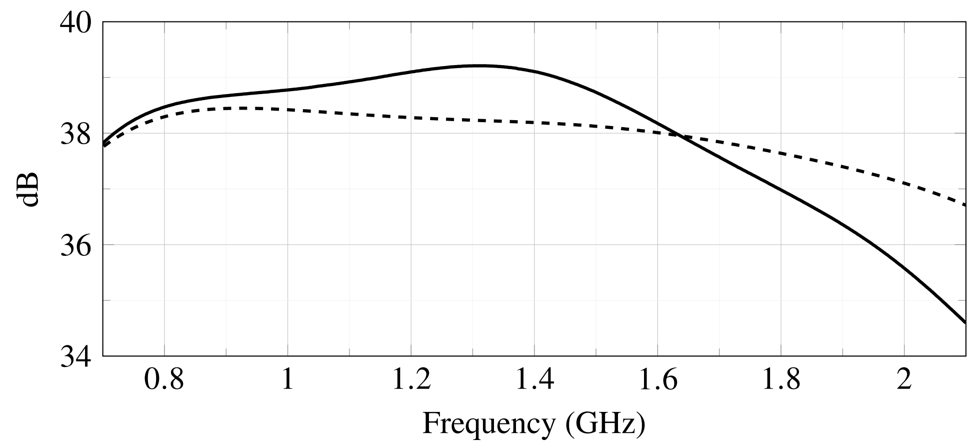


Figure 3.10: Measured (solid) versus simulated (dashed) S_{21}.

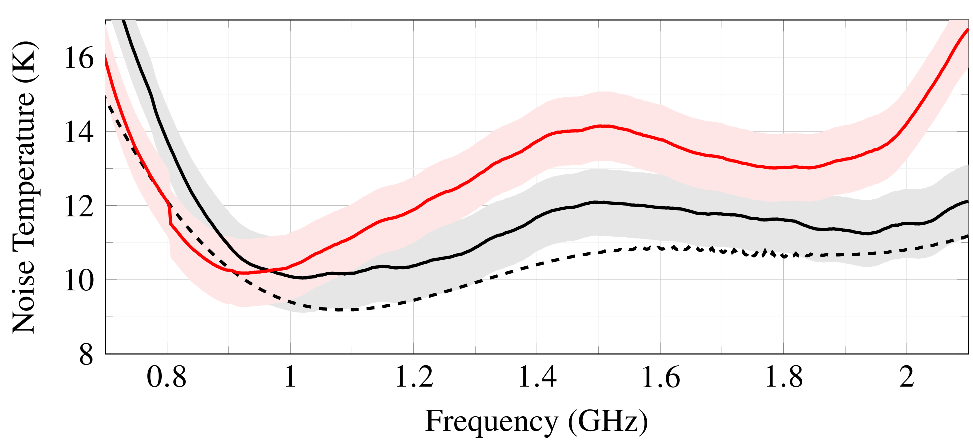


Figure 3.11: Measured (solid) and simulated (dashed) noise temperature of this work (lower trace), prior work [5] (higher trace), and 95% confidence regions.

Measured Results

We measured the S parameters of the assembled amplifier (shown in Figure 3.8) and the hot and cold terminations with a Keysight N5242A network analyzer with no averaging or smoothing. We used a Siglent SSA3032X as the noise receiver, preceded by a Mini-Circuits ZX60-P105LN+ amplifier and bias tee. We configured the spectrum analyzer to measure from 0.6 GHz to 2.1 GHz with a resolution bandwidth of 1 MHz, video bandwidth of 1 kHz, attenuation of 0 dB, preamplifier enabled, and log power averaging. After a 30 min warm-up time, data were saved for the noise of the receiver and the device under test with the hot and cold terminations, recording the physical temperatures. All measurements were performed at an ambient temperature of 297 K. We performed these power measurements five times each, saving the batch mean and standard deviation. In error analysis, we used these standard deviation data divided by $\sqrt{5}$ as the standard error of measurement.

We processed the measurements by computing the corrected Y-factor from (3.21), including uncertainty. Finally, these data were smoothed with a 150 MHz sliding window. The results are shown in Figure 3.11. The average noise temperature from 0.7 GHz to 2 GHz is (11.53 ± 0.42) K . The measured S parameters match the simulation well, as shown in Figure 3.9 and Figure 3.10, with disagreement in S_{22} due to an un-modeled output connector interface. S_{21} is reasonably flat with ≥35 dB across the band and has a maximum of 39 dB, and S_{11} is ≤ -8 dB. Differences between the simulated and measured S_{11} and S_{21} are due to imperfect wire bond geometry.

We performed these measurements on the amplifier from the prior work [5] using the same test setup, procedure, and physical temperature to ensure a fair comparison. The results of this new measurement are also shown in Figure 3.11. Our measurements of the average noise temperature for the prior amplifier is (12.48 ± 0.42) K . Comparing the two amplifiers across the same frequencies, the presented design shows a (0.95 ± 0.59) K mean improvement. As shown in Table 3.2, the presented amplifier outperforms not only the prior work in noise, but also other similar uncooled amplifiers intended for radio astronomy over a wide bandwidth.

3.5 Conclusion

In this chapter, we demonstrated the design of a non-uniform transmission line LNA input matching network using AD-accelerated gradient-based optimization. This amplifier achieves a frequency-averaged noise temperature of (11.53 ± 0.42) K from

0.7 GHz to 2 GHz at 297 K. The NTL design outperforms a previous uniform-line design by (0.95 ± 0.59) K . This design method is algorithmically more efficient than finite-difference optimization, and outperforms such solvers by more than 40x for this problem. AD is a general-purpose method for computing derivates and could be adopted for many other microwave problems such as parameter estimation and sensitivity analysis, as well as inverse design of structures besides NTLs.

Table 3.2: Comparison of Similar Radio Astronomy LNAs

Ref.	Freq. (GHz)	S_{21} (dB)	T_{50}[1](K)	Topo. / Tech.
This Work	0.7-2	36	11.5 ± 0.4	NTL / InP
[5]	0.7-2	40	12.5 ± 0.4[2]	UTL / InP
[35]	0.3-1.5	32	18 ± 6	Discrete / InP
[36]	0.7-1.4	17	20	CMOS

[1] Band-averaged noise temperature with a 50 Ω source impedance
[2] Our measurements of this amplifier

References

[1] G. Hallinan, V. Ravi, S. Weinreb, *et al.*, "The DSA-2000 — A Radio Survey Camera," *Bulletin of the AAS*, vol. 51, no. 7, Sep. 30, 2019.

[2] J. Flygare, "Wideband Low-Loss Feed Design for the DSA-2000 Ambient Temperature Array in Radio Astronomy," in *2024 IEEE International Symposium on Antennas and Propagation and INC/USNC-URSI Radio Science Meeting (AP-S/INC-USNC-URSI)*, Firenze, Italy: IEEE, Jul. 2024, pp. 5–6. DOI: 10.1109/AP-S/INC-USNC-URSI52054.2024.10686435.

[3] "Diramics pH-200 4F50." (), [Online]. Available: https://diramics.com/wp-content/uploads/downloads/DIRAMICS-pH-100-4F50.pdf (visited on 02/04/2025).

[4] S. Weinreb and J. Shi, "Low Noise Amplifier With 7-K Noise at 1.4 GHz and 25 °C," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 4, pp. 2345–2351, Apr. 2021. DOI: 10.1109/TMTT.2021.3061459.

[5] J. Shi and S. Weinreb, "Room-Temperature Low-Noise Amplifier With 11-K Average Noise From 0.6 to 2 GHz," *IEEE Microwave and Wireless Technology Letters*, vol. 33, no. 11, pp. 1540–1543, Nov. 2023. DOI: 10.1109/LMWT.2023.3315269.

[6] P. Miazga, "Nonuniform transmission line matching circuits synthesis - Analytical versus optimization approach," in *2014 20th International Conference on Microwaves, Radar and Wireless Communications (MIKON)*, Jun. 2014, pp. 1–4. DOI: 10.1109/MIKON.2014.6899913.

[7] P. Miazga, “Discrete shape optimization method of a non-uniform transmission line — advantages and drawbacks,” in *MIKON 2008 - 17th International Conference on Microwaves, Radar and Wireless Communications*, May 2008, pp. 1–4.

[8] J. Schaepperle and E. Luder, “Optimization of distributed parameter systems with a combined statistical-deterministic method,” in *Proceedings of IEEE International Symposium on Circuits and Systems - ISCAS ’94*, vol. 6, May 1994, 141–144 vol.6. DOI: 10.1109/ISCAS.1994.409546.

[9] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, “Automatic differentiation in machine learning: A survey,” *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 5595–5637, Jan. 1, 2017.

[10] G. Gonzalez, *Microwave Transistor Amplifiers: Analysis and Design*, 2nd ed. Upper Saddle River, N.J: Prentice Hall, 1997, 506 pp.

[11] H. W. Kuhn and A. W. Tucker, “Nonlinear programming,” in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, Univ. California Press, Berkeley-Los Angeles, Calif., 1951, pp. 481–492.

[12] M. E. Jerrell, “Automatic Differentiation and Interval Arithmetic for Estimation of Disequilibrium Models,” *Computational Economics*, vol. 10, no. 3, pp. 295–316, Aug. 1, 1997. DOI: 10.1023/A:1008633613243.

[13] G. F. Corliss, “Applications of Differentiation Arithmetic1,” in *Reliability in Computing*, R. E. Moore, Ed., Academic Press, Jan. 1, 1988, pp. 127–148. DOI: 10.1016/B978-0-12-505630-4.50013-4.

[14] A. Paszke *et al.*, “Automatic differentiation in pytorch,” in *NIPS 2017 Workshop on Autodiff*, Long Beach, California, USA, 2017.

[15] Feldmann, Melville, and Moinian, “Automatic differentiation in circuit simulation and device modeling,” in *1992 IEEE/ACM International Conference on Computer-Aided Design*, Nov. 1992, pp. 248–253. DOI: 10.1109/ICCAD.1992.279380.

[16] A. I. Mezza, R. Giampiccolo, and A. Bernardini, “Data-Driven Parameter Estimation of Lumped-Element Models via Automatic Differentiation,” *IEEE Access*, vol. 11, pp. 143 601–143 615, 2023. DOI: 10.1109/ACCESS.2023.3339890.

[17] J. I. Toivanen, R. A. E. Makinen, S. Jarvenpaa, P. Yla-Oijala, and J. Rahola, “Electromagnetic Sensitivity Analysis and Shape Optimization Using Method of Moments and Automatic Differentiation,” *IEEE Transactions on Antennas and Propagation*, vol. 57, no. 1, pp. 168–175, Jan. 2009. DOI: 10.1109/TAP.2008.2009657.

[18] G. Xiao and K. Yashiro, "Impedance matching for complex loads through nonuniform transmission lines," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 6, pp. 1520–1525, Jun. 2002. DOI: `10.1109/TMTT.2002.1006413`.

[19] T.-W. Pan, C.-W. Hsue, and J.-F. Huang, "Arbitrary filter design by using nonuniform transmission lines," *IEEE Microwave and Guided Wave Letters*, vol. 9, no. 2, pp. 60–62, Feb. 1999. DOI: `10.1109/75.755046`.

[20] M. Khalaj-Amirhosseini, "Nonuniform Transmission Lines as Compact Uniform Transmission Lines," *Progress In Electromagnetics Research C*, vol. 4, pp. 205–211, 2008. DOI: `10.2528/PIERC08082602`.

[21] I. Sugai, "Riccati's and Bernoulli's Equations for Nonuniform Transmission Lines," *IRE Transactions on Circuit Theory*, vol. 8, no. 3, pp. 359–360, Sep. 1961. DOI: `10.1109/TCT.1961.1086802`.

[22] H. Wheeler, "Transmission Lines with Exponential Taper," *Proceedings of the IRE*, vol. 27, no. 1, pp. 65–71, Jan. 1939. DOI: `10.1109/JRPROC.1939.228695`.

[23] J.-F. Mao and Z.-F. Li, "Analysis of the time response of nonuniform multiconductor transmission lines with a method of equivalent cascaded network chain," *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, no. 5, pp. 948–954, May 1992. DOI: `10.1109/22.137402`.

[24] K. Rae, V. Mahadevan, and S. Kosta, "Analysis of Straight Tapered Microstrip Transmission Lines - ASTMIC (Computer Program Descriptions)," *IEEE Transactions on Microwave Theory and Techniques*, vol. 25, no. 2, pp. 164–164, Feb. 1977. DOI: `10.1109/TMTT.1977.1129062`.

[25] P. G. Trémuel, E. Gavves, C. Würsch, K. Frick, and R. Vetsch, "Parameter-free Neural Field-based Optimal Design of Nonuniform Transmission Lines," in *2023 30th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, Dec. 2023, pp. 1–4. DOI: `10.1109/ICECS58634.2023.10382765`.

[26] J. Zhang, J. L. Drewniak, D. J. Pommerenke, *et al.*, "Causal RLGC(f) Models for Transmission Lines From Measured S -Parameters," *IEEE Transactions on Electromagnetic Compatibility*, vol. 52, no. 1, pp. 189–198, Feb. 2010. DOI: `10.1109/TEMC.2009.2035055`.

[27] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A Fresh Approach to Numerical Computing," *SIAM Review*, vol. 59, no. 1, pp. 65–98, Jan. 2017. DOI: `10.1137/141000671`.

[28] W. S. Moses and V. Churavy, "Instead of rewriting foreign code for machine learning, automatically synthesize fast gradients," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, ser. NIPS '20, Red Hook, NY, USA: Curran Associates Inc., Dec. 6, 2020, pp. 12 472–12 485.

[29] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical Programming*, vol. 106, no. 1, pp. 25–57, Mar. 1, 2006. DOI: 10.1007/s10107-004-0559-y.

[30] A. Soliman, A. Janzen, and S. Weinreb, "Thermal modelling of coaxial line for cryogenic noise measurements," in *2016 URSI Asia-Pacific Radio Science Conference (URSI AP-RASC)*, Aug. 2016, pp. 900–903. DOI: 10.1109/URSIAP-RASC.2016.7601381.

[31] H. Packard, "Fundamentals of RF and Microwave Noise Figure Measurements," Hewlett Packard, Tech. Rep. 57-1, Jul. 1983.

[32] P. Heymann and M. Rudolph, *A Guide to Noise in Microwave Circuits: Devices, Circuits, and Measurement*, 1st ed. Wiley, Dec. 21, 2021. DOI: 10.1002/9781119859390.

[33] K. Technologies, "Keysight 2-Port and 4-Port PNA-X Network Analyzer," Keysight Technologies, Tech. Rep. N5242-90007, Oct. 2018.

[34] M. Giordano, *Uncertainty propagation with functionally correlated quantities*, Oct. 2016. DOI: 10.48550/arXiv.1610.08716. arXiv: 1610.08716 [physics].

[35] M. Lai, V. Mackay, D. Wulf, P. Shmerko, and L. Belostotski, "0.3–1.5-GHz LNA With Wideband Noise and Power Matching for Radio Astronomy," *IEEE Microwave and Wireless Technology Letters*, vol. 33, no. 8, pp. 1163–1166, Aug. 2023. DOI: 10.1109/LMWT.2023.3272211.

[36] L. Belostotski and J. W. Haslett, "Sub-0.2 dB Noise Figure Wideband Room-Temperature CMOS LNA With Non-50 Ohm Signal-Source Impedance," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 11, pp. 2492–2502, Nov. 2007. DOI: 10.1109/JSSC.2007.907172.

Chapter 4

ANALOG SIGNAL PATH DESIGN FOR DSA-2000

The Analog Signal Path (ASP) subsystem in the DSA-2000 is responsible for all the signal processing before digitization. There is no frequency conversion, so the entire signal path consists of amplifiers, filters, attenuators, and most notably an RF over fiber optic (RFoF) link. While RFoF links are widely used in radio astronomy [1]–[3], none have been deployed at the scale of DSA-2000. This chapter discusses the ASP system as a whole, focusing on the design, implementation, and testing of the RFoF link. We also describe modifications to the LNA to improve the match to the feed and tests of an LNA with thermoelectric cooling of the first stage transistor, all to further improve the noise.

4.1 RF Over Fiber Link

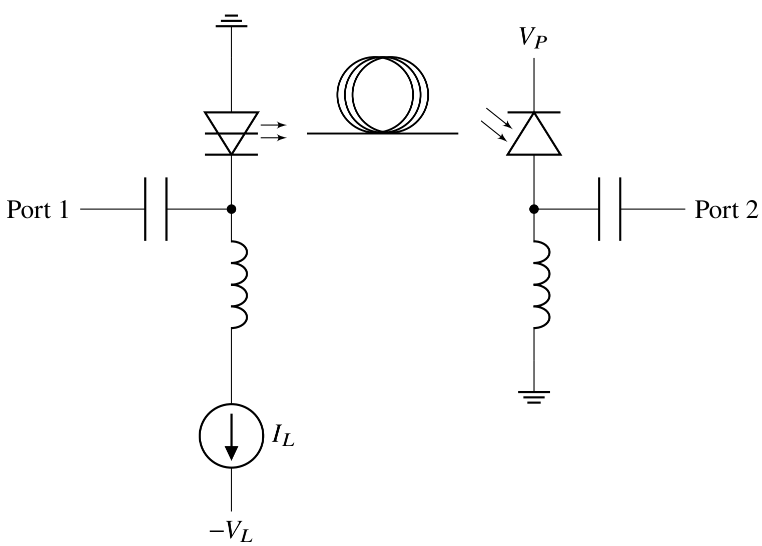


Figure 4.1: RFoF system diagram

To transport the RF signal from each antenna to the processing facility, we are using RF over fiber (RFoF) links. These links use RF power to directly modulate the current of a solid-state diode laser, resulting in intensity modulated light. After traveling through up to 20 km of single mode fiber, this optical signal is presented to a photodiode, where it is converted back into RF current. This style of analog RF transmission is used in cable television and commercial/military antenna remoting applications. It is important to note that direct modulation is similar, but architecturally very different from digital modulation, where 1s and 0s are achieved by pulsing the laser nearing 100% modulation depth. In RF or analog modulation, we operate in the small signal regime of the laser/photodiode transfer function and are subject to their gain, noise, and linearity characteristics. While this is a simple and inexpensive solution for long-haul analog transmission, the dynamic range is limited by the distortion in the laser and noise from the laser or photodiode, depending on the amount of optical loss in the link.

Noise

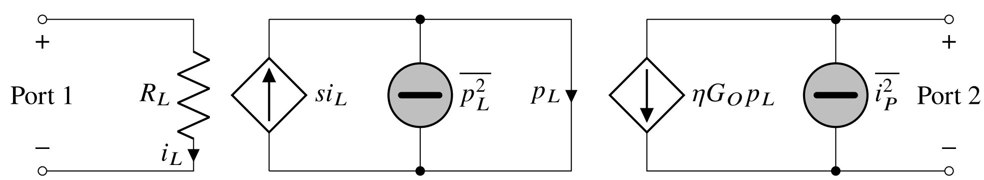


Figure 4.2: RFoF equivalent noise circuit

We begin by focusing on the noise figure of the intrinsic laser to photodiode link. Noise figure is defined as the degradation in signal-to-noise ratio (SNR) from input to output. The noise at the input is taken as thermal noise in the generator impedance at a reference temperature of T_0. There are several mechanisms by which the SNR is reduced in the link. There is added thermal noise due to various resistive components, relative intensity noise or RIN in the laser, optical loss in the fiber (whose noise figure is identical to that of a passive RF attenuator [4]), and shot noise in the receiving photodiode. In the simple, direct-modulated diode link we are studying, RIN and shot noise are the dominant terms, and as such we will ignore the various other sources of thermal noise [5].

The relative intensity noise of the laser (RIN) is the mean deviation in optical output power about the mean square output, given by

$$\text{RIN (dB/Hz)} = \frac{\overline{p_L^2}}{P^2}, \tag{4.1}$$

where $\overline{p_L^2}$ is the average fluctuation in optical power and P is the DC optical power. RIN is driven by fluctuations in the photon count in the laser gain cavity through various mechanisms. If the laser were biased well past its threshold current, RIN is dominated by shot noise. In this regime $\overline{p_L^2} \propto P$, so RIN $\propto 1/P$. At lower optical power, RIN is dominated by spontaneous emission [6] and is $\propto 1/P^3$. Unfortunately, this spectrum is not white, unlike pure shot noise. Near the laser's relaxation resonance, RIN is amplified [7], giving a slope to the power spectral density. Other device-dependent effects can also strongly shape the spectral and bias-dependent effects. As the behavior of RIN is strongly laser-dependent, we start with an experiment measuring our laser's RIN.

To measure the laser RIN, we employ the "subtraction method" [8]. This method sidesteps the need for an optical spectrum analyzer by measuring the induced photocurrent noise due to the laser's intensity noise on an RF spectrum analyzer. This experiment requires two measurements, one with the laser off and one with the laser on. In the off state (with the photodiode biased), we measure the power spectral density of the receiver with the dark current of the photodiode. Then, with the laser biased, we repeat the measurement, this time reading a power spectral density of the aforementioned noise sources superimposed with the laser's noise. With the laser biased, we note the DC photocurrent of the photodiode. A figure of the test setup is shown in Figure 4.3.

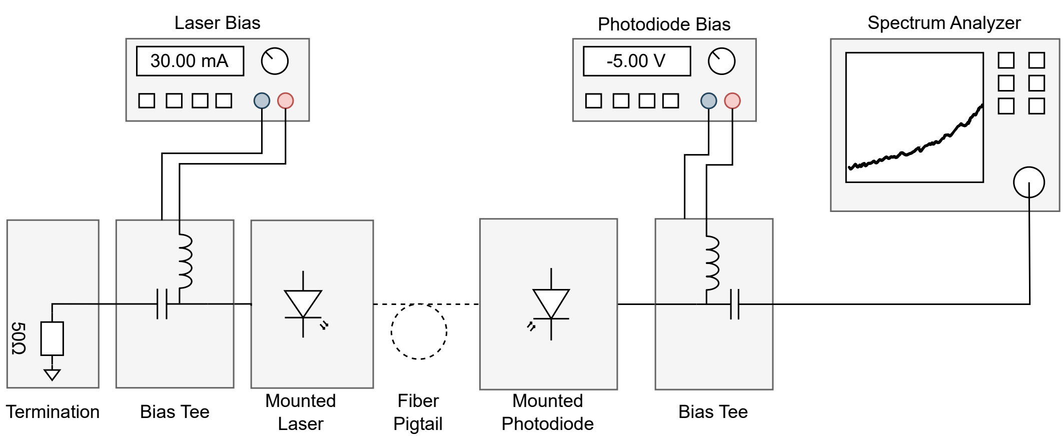


Figure 4.3: Diagram of RIN test setup

To compute the link noise alone, we subtract the "dark" noise from the measurement with the laser on. Then, we compute the noise power into the spectrum analyzer from the photodiode's DC shot noise, and subtract that as well to result in just the laser's noise power. Finally, as the noise power is proportional to $\overline{p_L^2}$, we divide by DC photocurrent's equivalent power to compute RIN, given by

$$\text{RIN (dB/Hz)} = 10 \log 10 \frac{S_T - S_{RX} - 2qI_{PD}Z_0}{I_{PD}^2 Z_0}, \tag{4.2}$$

where S_T is the total power spectral density, S_{RX} is the measured "dark" power spectral density, I_{PD} is the DC photocurrent, q is the electron charge, and Z_0 is the input impedance of the spectrum analyzer. We then repeat this measurement as a function of bias. The result of this experiment is shown in Figure 4.4. The ripple in this measurement is due to the mismatch between the high impedance photodiode and the length of transmission line and bias tee before the 50 Ω spectrum analyzer input. Missing data are due to the sensitivity limitations of the spectrum analyzer. The general shape of rising noise with frequency comes from the aforementioned laser relaxation resonance driving up RIN. The relaxation resonance for this laser is just above 3 GHz, past the top end of the measurement.

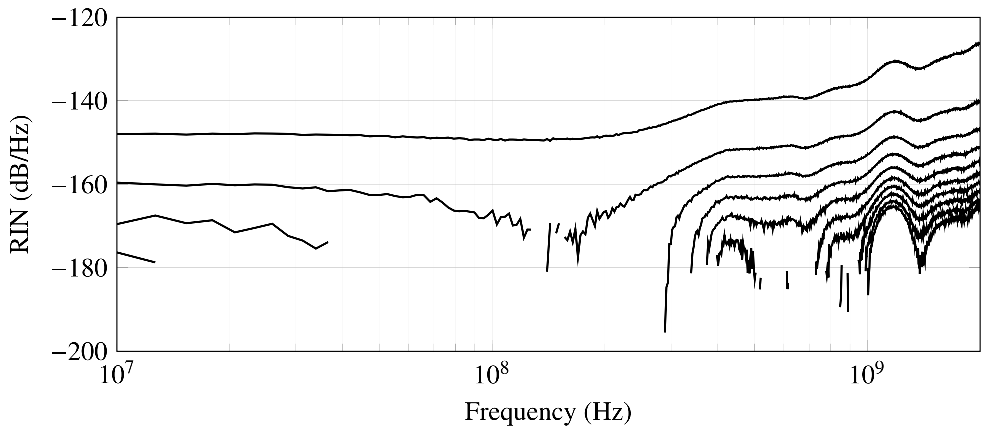


Figure 4.4: Measured RIN of an AGx laser. Top to bottom: 10 mA to 50 mA in steps of 4 mA.

To validate our assumptions about RIN, we look at the bias (optical power) response at several frequencies. To observe the dependence on laser power instead of laser bias current, however, we need to transform the laser current to power via the laser's threshold current I_{th} and slope efficiency s. As we also have measurements of the DC photocurrent, and know the photodiode's responsivity η and an approximate less-than-unity gain (loss) of the connector interface G_O, we can estimate the optical power from that direction as well. From the datasheet, I_{th}=8 mA, s=0.315 W/A, η=0.93 A/W. We estimate G_O to be 0.98.

Looking at the response in Figure 4.5, we observe that the datasheet values and our assumptions agree. This implies we can use either conversion factor to compute laser power instead of disconnecting the laser and attaching it to an optical power meter.

Figure 4.6 shows the RIN versus laser power at a few frequencies. We observe the steep increase in RIN at low laser power due to spontaneous emission, and the linear

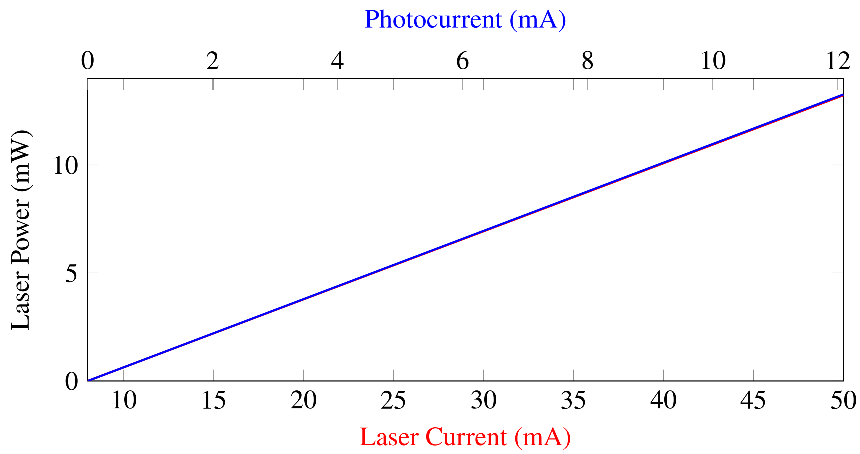


Figure 4.5: Laser output power vs. laser current and DC photocurrent.

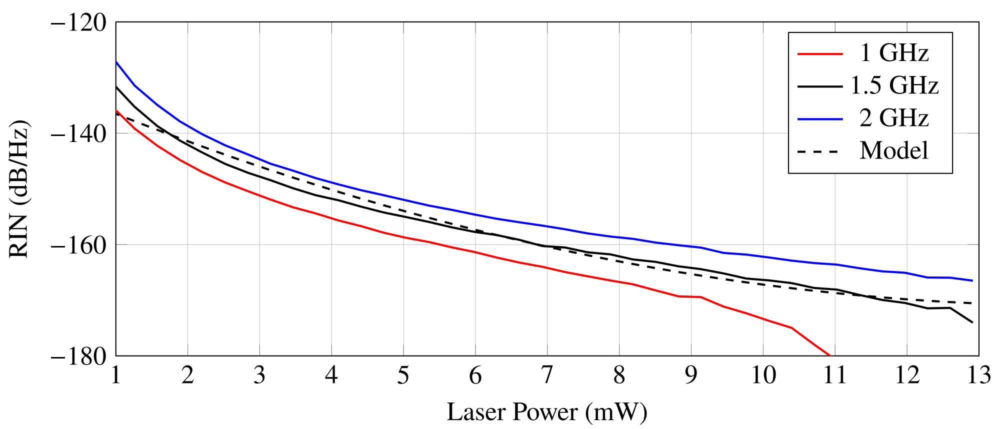


Figure 4.6: Measured RIN of an AGx laser vs laser output at three frequencies.

section due to shot noise at higher power, matching our prediction.

To aid in analysis, we ignore the frequency dependence for now and fit a simple quadratic model at 1.5 GHz

$$\mathrm{RIN}_{\mathrm{Model}} = -131.2 - 5.5P + 0.19P^2\,, \tag{4.3}$$

where P is the laser output power in mW, also plotted in Figure 4.6.

Now that we understand the bias and frequency response of RIN, we continue with our noise analysis. In the small signal case, the laser appears as a small resistor R_L, with deviations in current manifesting as deviations in the intensity of the light following the slope efficiency of the laser. The RIN is then superimposed over this modulated light, creating a mean square deviation over the DC optical power P

following

$$\overline{p_L^2} = \mathrm{RIN} \cdot P^2 B\,. \tag{4.4}$$

Finally, the light is attenuated by the less-than-unity gain of the fiber, G_O, and is converted back into current following the photodiode's responsivity. As the laser is biased to a fixed optical output from which it is then modulated, the DC photocurrent generates shot noise in the photodiode following

$$\overline{i_P^2} = 2qI_P B = 2qPG_O\eta B\,. \tag{4.5}$$

The equivalent noise circuit for this system is shown in Figure 4.2.

We simplify this circuit by combining the dependent sources and moving the noise currents forward towards the output, Port 2. The total current gain between the laser input and the photodiode is

$$g_i = sG_O\eta\,. \tag{4.6}$$

Noise in the optical power of the laser manifests as noise in the photocurrent, scaled by the square of the passive gain of the fiber and the photodiode responsivity, $(G_O\eta)^2$. As the shot noise and RIN are uncorrelated, the total noise current is their sum given by

$$\overline{i_{\mathrm{Out}}^2} = \overline{p_L^2}(G_O\eta)^2 + \overline{i_P^2} = PG_O\eta B(PG_O\eta \cdot \mathrm{RIN} + 2q)\,. \tag{4.7}$$

The simplified, output-referred circuit is shown in Figure 4.7.

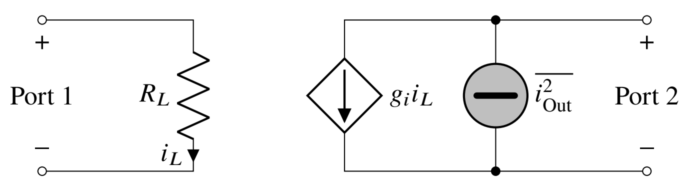


Figure 4.7: RFoF output-referred current noise

It is usual to work in input-referred noise and here we encounter the first quirk in noise analysis, where instead of simply referring $\overline{i_{\mathrm{Out}}^2}$ to the input by g_i^2, we must consider the impact of the source termination. Definitionally, noise figure is the degradation in SNR, where the “signal” originates from a power source, i.e., with a source impedance. This is emblematic of the difficulty in this type of analysis, as it would seem that moving the currents forward in the previous case is the same as moving them backwards towards the input. However, the subtle distinction here is that moving forward is asking, "how does current changing at the input branch manifest as current at the output", which depends solely on the gain, whereas moving backwards is asking, "what equivalent current at the input would manifest as current at the output", which depends on the circuitry attached to the input. We begin to solve this problem by working backwards, starting with the current we want at the input under the source termination condition and working towards the output.

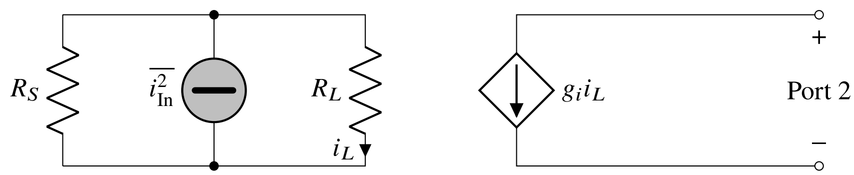


Figure 4.8: RFoF input-referred current noise

In this configuration, we use the current divider to find

$$i_L = i_{\text{In}} \frac{R_S}{R_L + R_S} \,. \tag{4.8}$$

Then, we use the current gain to move to the output

$$i_{\text{Out}} = g_i i_{\text{In}} \frac{R_S}{R_L + R_S} \,. \tag{4.9}$$

Finally, we solve for i_{In} in terms of i_{Out}, and take the time average of the square

$$\overline{i^2_{\text{In}}} = \frac{\overline{i^2_{\text{Out}}}}{g_i^2} \left(\frac{R_L}{R_S} + 1 \right)^2 \,. \tag{4.10}$$

To find the noise figure, we use the relation

$$F = \frac{\overline{i^2_{\text{Total}}}}{\overline{i^2_S}} = \frac{\overline{i^2_{\text{In}}} + \overline{i^2_S}}{\overline{i^2_S}} = \frac{\overline{i^2_{\text{In}}}}{\overline{i^2_S}} + 1 \,, \tag{4.11}$$

where $\overline{i^2_S}$ is the thermal noise in the generator, defined as

$$\overline{i^2_S} = 4 k_B T_0 G_S B \,, \tag{4.12}$$

where G_S is the source conductance and T_0 =290 K (IEEE standard). Combining these equations yield

$$F = 1 + \frac{\overline{i^2_{\text{Out}}} \left(\frac{R_L}{R_S} + 1 \right)^2}{4 k_B T_0 G_S B g_i^2} \,. \tag{4.13}$$

Finally, as we prefer to work in equivalent noise temperature, we use $T_e = T_0(F - 1)$, we find

$$T_e = \frac{P R_S}{4 k_B s^2 G_O \eta} \left(P G_O \eta \cdot \text{RIN} + 2q \right) \left(\frac{R_L}{R_S} + 1 \right)^2 \,. \tag{4.14}$$

This equivalent noise temperature result gives insight into the behavior of the link under different source impedances. If we are free to choose R_S through matching, we would want to know what R_S provides the best noise performance. We solve for this quantity by minimizing Equation 4.14, finding the R_S that solves $\delta T_e/\delta R_s = 0$. Taking the positive solution of the quadratic yields

$$R_{S,\mathrm{Opt}} = R_L \,. \tag{4.15}$$

This follows from the fact that all the noise sources are not electrically influenced by the source impedance, so by conjugate matching the laser, we maximize the link's gain and therefore minimize the input-referred noise. In this case, we can find the minimum noise temperature of

$$T_{\mathrm{Min}} = \frac{PR_L}{k_B s^2 G_O \eta} \left(PG_O\eta \cdot \mathrm{RIN} + 2q\right) \,, \tag{4.16}$$

which is a ratio of

$$\frac{T_e}{T_{\mathrm{Min}}} = \frac{(R_S + R_L)^2}{4R_S R_L} \,, \tag{4.17}$$

or 3.025 for R_S =50 Ω and R_L =5 Ω The plot for minimum noise alongside the 50 Ω noise is shown in Figure 4.9. This figure uses typical parameters for our laser[1], which has a RIN following our model in Equation 4.3, a slope efficiency s of 0.315 W/A, and a slope impedance of 5 Ω and for our photodiode[2], which has a slope efficiency η of 0.93 A/W. The conclusion from this plot is that for laser powers between 8 mW and 12 mW, the noise is about the same, within a factor of 2, and remains that way as we increase the loss.

With $T_e(R_S)$ and $R_{S,\mathrm{Opt}}$, we have essentially derived the noise parameters of the RFoF link. Notably missing from this analysis are the parasitics of the components, the series inductance of the leads of the packaged coaxial laser and photodiode, and the junction capacitance of the photodiode. To minimize noise, as we have seen, we will want to power match at the input of the network. As such, any inductance at the input will need to be tuned out in addition to matching the slope resistance.

S Parameters

In analyzing the gain and reflection coefficients of the RFoF link, it is more convenient to analyze S parameters instead of voltages and currents. However, determining the

[1] `https://www.agxtech.com/PDF/PLMR3+Rev+4.0.pdf`
[2] `https://www.agxtech.com/PDF/PPDA+R5.6.pdf`

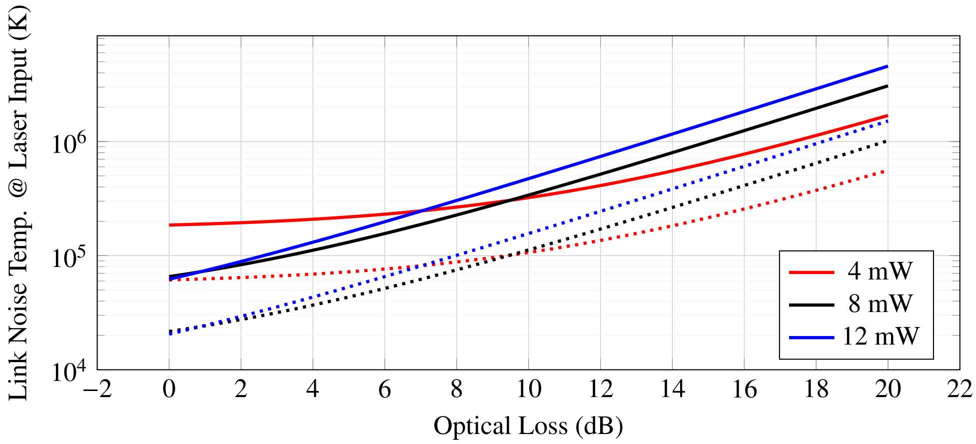


Figure 4.9: RFoF link noise ($R_S = 50\,\Omega$ (solid) and optimal (dotted)) versus optical attenuation.

S parameters directly from the Figure 4.2 model is nontrivial due to the dependent sources. Instead, it would be easier to consider the Y parameters and convert using the well-known formula. The Y parameters of Figure 4.2 are

$$Y_{\text{Link}} = \frac{1}{R_L}\begin{bmatrix} 1 & 0 \\ s\eta G_{\text{O}} & 0 \end{bmatrix}. \tag{4.18}$$

Converting to S parameters gives

$$S_{\text{Link}} = \begin{bmatrix} \dfrac{R_L - Z_0}{R_L + Z_0} & 0 \\ \\ \dfrac{2Z_0 s\eta G_{\text{O}}}{R_L + Z_0} & 1 \end{bmatrix}. \tag{4.19}$$

S_{11} of this result is as expected, it is simply the reflection coefficient of the laser's slope impedance. S_{21}, however, is more nuanced. If we substitute numbers into this equation such as Z_0=50 Ω, R_L=5 Ω and $s\eta G_{\text{Opt}}$=0.2, we find $|S_{21}|^2$=−8.79 dB. If we "match" the laser by setting the characteristic impedance equal to the laser's impedance $R_L = Z_0$, the equation reduces to $s\eta G_{\text{Opt}}$. However, using the same device parameters, we find $|0.2|^2$=−13.98 dB, unintuitively lower gain than before. The apparent drop in gain is because the equation for S_{21} assumes both the output and input ports have equivalent characteristic impedance, and changing the output impedance changes the power delivered from the photodiode. It would be more correct to analyze this circuit with an explicit matching network at the input.

For a real load as we are exploring, we can use an ideal transformer at the input as our matching network, which has an ABCD matrix of

$$A_{TF} = \begin{bmatrix} N & 0 \\ 0 & 1/N \end{bmatrix} . \tag{4.20}$$

We convert the Y matrix of the link to ABCD parameters using the standard equation to arrive at

$$A_{\text{Link}} = -\frac{1}{s\eta G_{\text{O}}} \begin{bmatrix} 0 & R_L \\ 0 & 1 \end{bmatrix} . \tag{4.21}$$

We then cascade to get

$$A_{\text{Combined}} = -\frac{1}{s\eta G_{\text{O}}} \begin{bmatrix} 0 & NR_L \\ 0 & 1/N \end{bmatrix} . \tag{4.22}$$

And then finally converting to S parameters, we find

$$S_{\text{Combined}} = -\frac{1}{Z_0 + N^2 R_L} \begin{bmatrix} Z_0 - N^2 R_L & 0 \\ 2NZ_0 s\eta G_{\text{O}} & 1 \end{bmatrix} . \tag{4.23}$$

If $N = \sqrt{Z_0/R_L}$, matching to the laser resistance, this reduces to

$$S_{\text{Matched}} = \begin{bmatrix} 0 & 0 \\ -\sqrt{Z_0} s\eta G_{\text{O}}/\sqrt{R_L} & 1 \end{bmatrix} . \tag{4.24}$$

Using the same parameters as before, the matched link now has an $|S_{21}|^2$ of -3.98 dB, or 4.81 dB higher gain than unmatched. A notable result from this is that positive power gain can be achieved, even though the intrinsic link exhibits a current gain of less than unity. For example, if $s\eta G_O$=0.2, Z_0=50 Ω, and R_L=2 Ω, the link's gain $|S_{21}|$ would be unity. While unintuitive, the transformer acts as a current gain device, for which a given turns ratio will counteract the current loss in the link. From both the noise and gain perspective, we want to power match the input of the laser. We can perform this match with a wideband transformer, although finding a device with such an extreme impedance ratio may prove difficult. Alternatively, transmission line matching or lumped matching could be employed.

Linearity

The final metric we evaluate for the RFoF link is the linearity. If radio frequency interference (RFI) is strong enough, nonlinearity in the signal chain would result in frequency products in addition to the center frequency of the RFI itself. The strength of these nth-order products depends on the strength of the incoming signal and the nth derivative of the transfer function of the system. One way to reduce the impact of this distortion is to simply limit the incoming signal strength. Doing so, however, is at odds with added noise, as we want to *increase* power before a noisy component to reduce that component's noise impact on the SNR. To compare RFoF links (bias settings, laser or photodiode choice, etc.), we use "spurious free dynamic range" or (SFDR) as a noise-normalized linearity metric. SFDR is defined as the maximum signal power at the output for which the power of the n-th order intermodulation product is equal to the noise power. This is defined following [9] as

$$\mathrm{SFDR}_n\ (\mathrm{dB}) = \frac{n-1}{n}(\mathrm{OIP}n - N_O)\,, \tag{4.25}$$

where OIP_n is the nth order output intermodulation intercept point and N_O is the output noise power, both defined in dBm. We only need to consider $n = 3$ for narrowband systems, but our radio telescope is more than an octave, so we must consider second order distortion as well, as distortion products from the lower half of the band will appear in the upper half.

Experimental Results

Noise

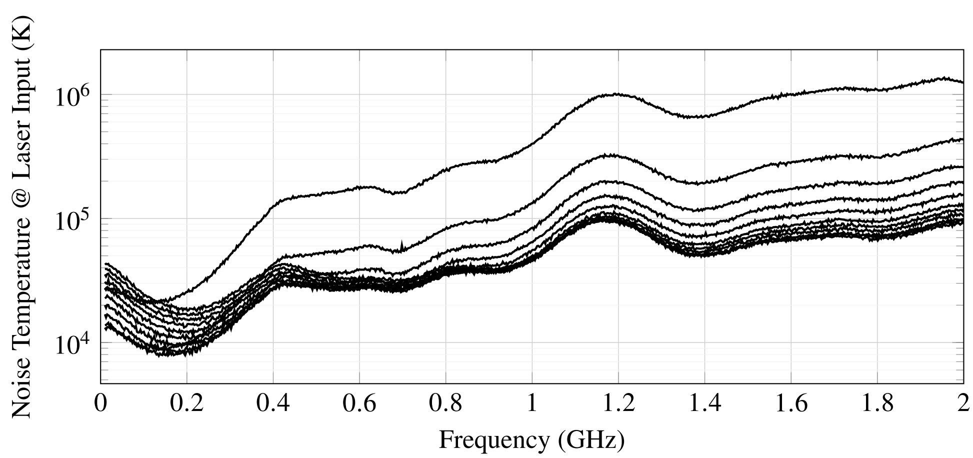


Figure 4.10: RFoF link input noise temperature from 10 mA to 50 mA (top to bottom) in steps of 4 mA.

To validate our noise derivation, we directly measured the link output noise and referred this to the input. The setup is identical to that used in the RIN characterization, shown in Figure 4.3. We used a spectrum analyzer as a power meter to measure the noise power, with the link input terminated with a room-temperature load. In addition to this "hot" measurement, we also measured the noise of the receiver, including the "dark current" of the photodiode. Finally, we subtracted the linear noise powers to leave us with just the noise of the link. Using measured power gain from a network analyzer, we referred the noise power back to the input.

We measured the noise power spectral density on a spectrum analyzer at the output of the photodiode, subtracting the receiver noise. The input-referred noise temperature in Figure 4.10 matches our expected result from Figure 4.9, with high laser currents resulting in an input noise temperature of about 60 000 K in the band of interest. While this is an excellent result, we need to consider both the noise and linearity to compute the SFDR.

Linearity

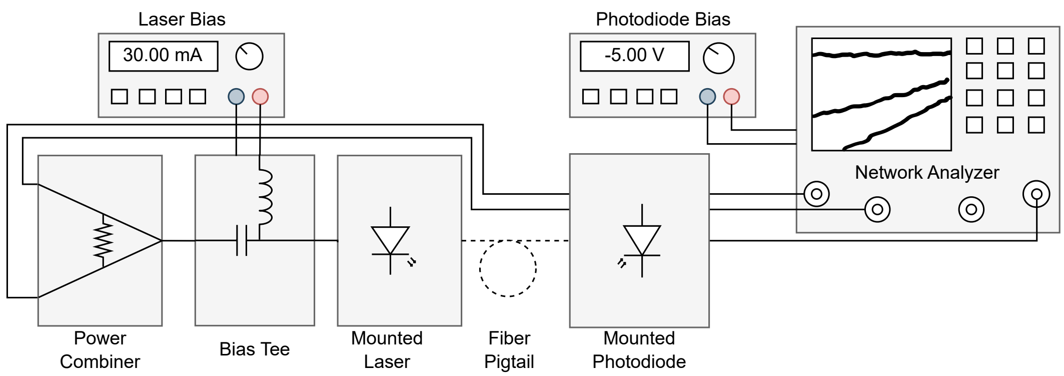


Figure 4.11: Diagram of linearity test setup

While it is simple to measure noise and gain to validate analytic expressions, linearity is more difficult as we do not have an expression for the actual transfer function of the laser and photodiode pair. The solution is to probe the individual derivatives of the transfer function by injecting a pair of tones and observing their intermodulation products, derived for reference in Appendix B. For the experiment, we inject two closely spaced tones, ω_L and ω_H (1 MHz apart) and observe five tones on the output. We observe the powers of the two fundamental tones to track the gain, P_L and P_H, $\omega_L + \omega_H$ for IM_2 to compute OIP2 via

$$\mathrm{OIP2\ (dBm)} = P_L + P_H - \mathrm{IM2}\,, \tag{4.26}$$

and $2\omega_H - \omega_L$ for $\mathrm{IM3}_H$, and $2\omega_L - \omega_H$ for $\mathrm{IM3}_L$ to compute OIP3 via

$$\mathrm{OIP3~(dBm)} = \frac{2P_L + P_H - \mathrm{IM3}_L}{2} \tag{4.27}$$

$$= \frac{P_L + 2P_H - \mathrm{IM3}_H}{2}. \tag{4.28}$$

As OIP3 has two unique solutions, we are typically interested in the worst case, which will be the minimum of the two OIP3 results. For most components, the two terms will be close to equal. Performing this experiment for the AGx laser and photodiode pair from before, we measured the results shown in Figure 4.12.

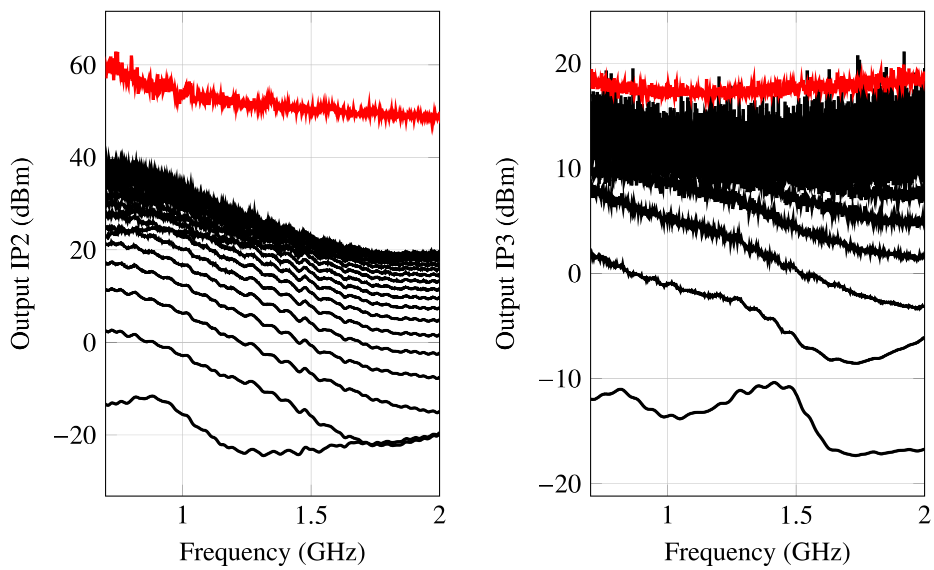


Figure 4.12: RFoF measured OIPn. Red trace is the receiver linearity, black traces bottom to top are laser currents from 10 mA to 50 mA in steps of 2 mA

This measurement was performed on a PNA-X network analyzer, utilizing the two independent sources as the two tones incident on the link under test, shown in Figure 4.11. We used a power combiner to join these two outputs, and calibrated out the loss of the combiner and bias tee to result in an incident power of −20 dBm for each tone. The analyzer was then configured to detune the receiver to follow the five tones of interest described earlier, measuring the power directly in dBm. This measurement includes a correction for the loss in the cable and internals following the photodiode.

Both the IP2 and IP3 results are excellent for this laser, being limited by the dynamic range of the PNA-X in the IP3 measurement. In the presented range, the gain only changes slightly, from an average of −5 dB to −4 dB across the band, with associated input values changing accordingly. The linearity improves through increased laser power, but then worsens as the device starts to saturate, near 50 mA.

Combining the noise and linearity measurements, we compute the spurious-free dynamic range as defined in Equation 4.25, shown in Figure 4.13. Here, we see

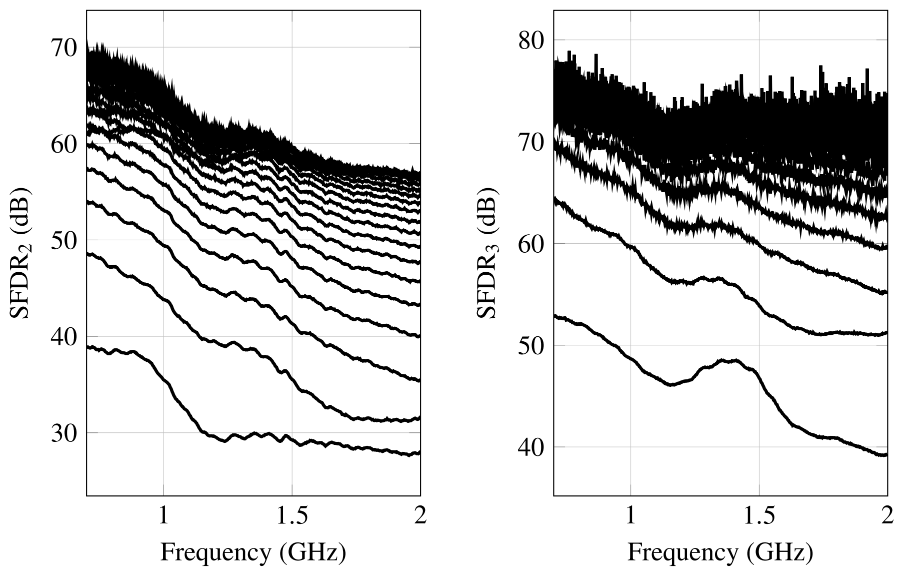


Figure 4.13: Spurious-free dynamic range for the intrinsic RFoF link. Traces bottom to top are laser currents from 10 mA to 50 mA in steps of 2 mA. Noise power is the measured intrinsic link noise under 100 kHz of noise equivalent bandwidth.

that we are limited primarily by the 2nd-order linearity, although it is important to note that after 1 GHz, the power in the second order tone (2 GHz) is out of band and will be filtered out). At high laser powers, in the band of 0.7 GHz to 2 GHz, we can expect SFDR of better than 60 dB for the intrinsic RFoF link.

As the SFDR falls off at high laser power while the noise improves, we find the optimum laser bias of around 38 mA shown in Figure 4.14.

Long-Haul RFoF

The final piece of the RFoF puzzle involves the consequences of the exceptionally long fiber that connects the most remote antennas in the telescope to the digitizers. Current estimates predict the longest fiber lengths will be about 20 km. As the lasers operate at 1310 nm, the effective loss in standard single-mode fiber is around 0.5 dB/km. Figure 4.9 indicates that the resulting 10 dB of optical loss (20 dB in RF power loss) will result in about a factor of four increase in noise. This is still an acceptable noise level, as there can be a significant amount of gain before the link.

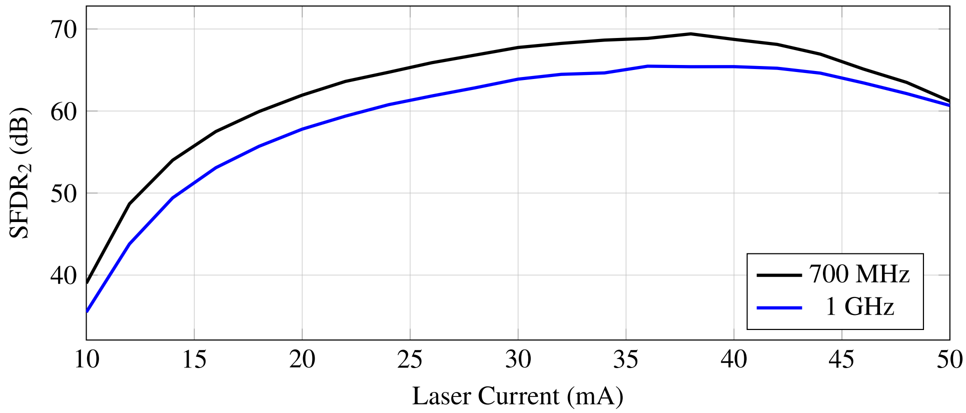


Figure 4.14: Spurious-free dynamic range for the RFoF link vs laser current

In initial experiments with a 20 km spool on the bench, linearity and noise both worsened significantly more than what was expected. We believe the origin of these effects are reflections off imperfect connectors and Rayleigh backscattering, coupled with poor isolation in the laser and laser chirp. While simple reflections are a well-known issue in RFoF links, the induced distortion due to the increased effect of Rayleigh backscattering in long fibers combined with the laser chirp is more nuanced. Only recently has this effect been discussed in the context of radio astronomy [1], [10].

Rayleigh backscattering is a phenomenon in optical fiber where imperfections in the refractive index results in scattering. While this would typically manifest as a linear effect (and is the dominant component of loss in fiber optic links [11]), the interaction with the chirp of the laser results in nonlinear effects. Chirp is the change in instantaneous optical frequency of the laser as it is directly modulated. As Rayleigh backscattering sends a reflected wave back into the laser, the time-dependence of chirp results in the different optical frequencies mixing and retransmitting, resulting in a distorted signal.

There are a few ways one could mitigate the distortion introduced from chirp and scattering. The most straightforward approach would be to use a lower chirp laser, or to use an external modulator where there would be no chirp. Both of these solutions, however, are typically expensive. The solution for one radio telescope (the SKA) [10] is to introduce a dithering tone. This low frequency modulation added on top of the RF modulation results in decoherence of the scattered light and the light emitted from the gain cavity of the laser. This solution, however, requires additional hardware and filtering. The most simple solution is to instead purchase lasers with more optical

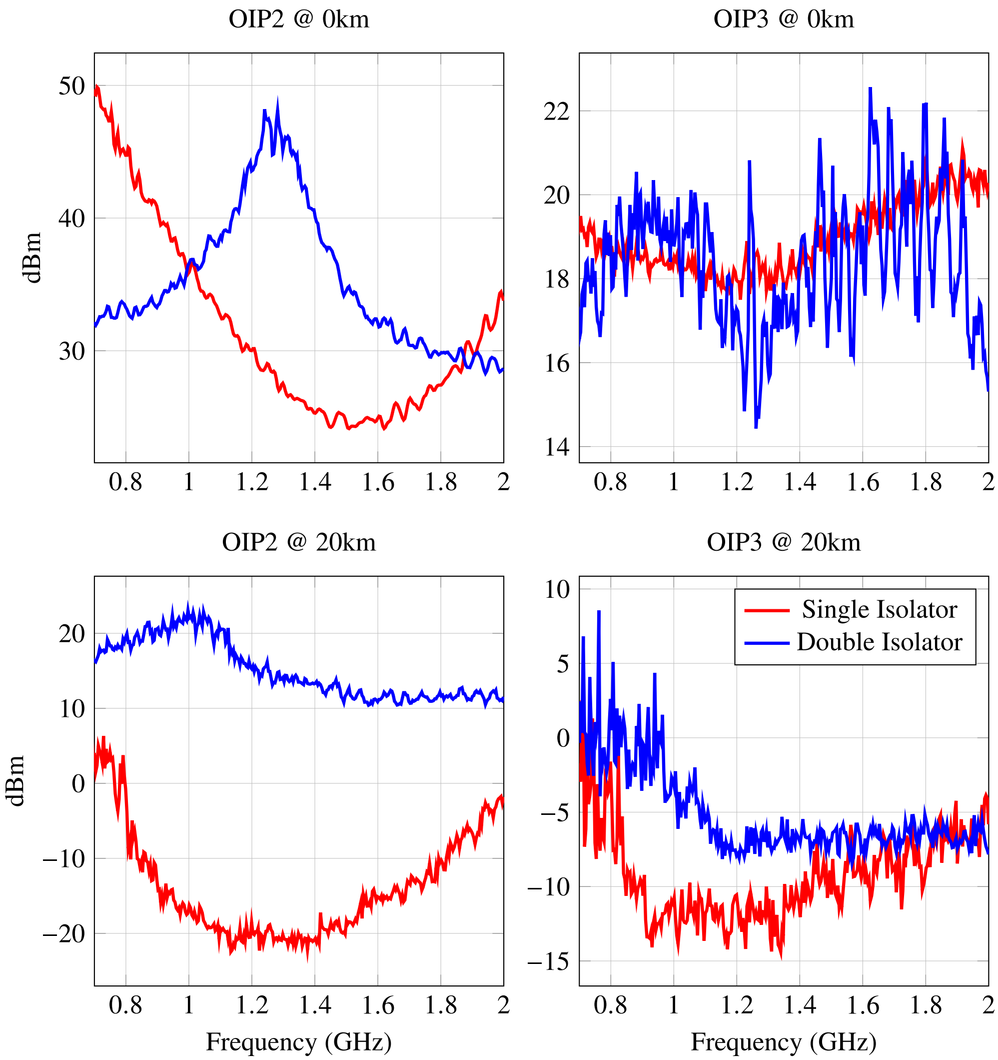


Figure 4.15: Linearity metrics for single- and double- isolated lasers at 0 and 20 km. Both lasers are operated at 40 mA.

isolation built in, or add in-line isolators in the path of the fiber near the laser. For most vendors, a double isolator (which increases the optical isolation by 20 dB) is on the order of a $10 increase in cost.

We find that adding additional isolators is an effective way to reduce the nonlinearities introduced by Rayleigh backscattering by repeating the earlier tests on the bench with a 20 km spool[3], shown in Figure 4.15. In this figure, the top row are tests with 0 km of fiber, with the laser and photodiode directly connected. The bottom row adds a 20 km spool of single-mode fiber between the two. The left column shows OIP2 and the right column shows OIP3. All four measurements are shown for the

[3]Thank you, Courtney!

single- and double-isolated lasers.

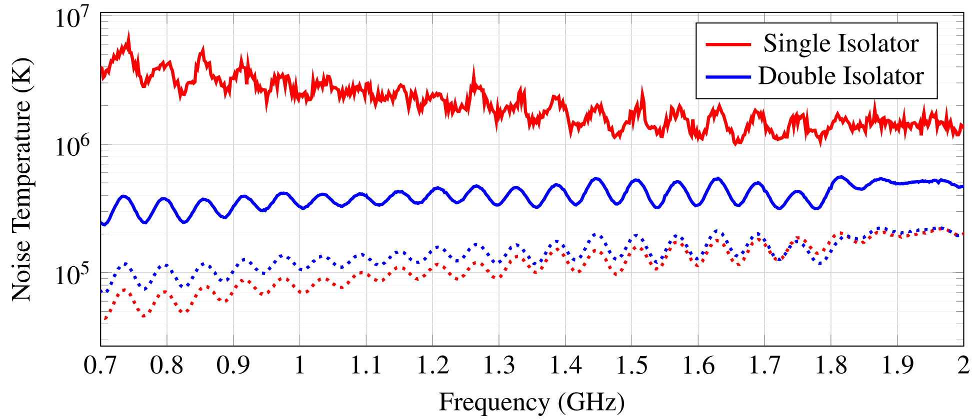


Figure 4.16: Noise temperature for single- and double- isolated lasers at 0 km (dotted) and 20 km (solid). Both lasers are operated at 40 mA.

From Figure 4.15, it is clear that at 20 km we observe Rayleigh scattering as the linearity for the single-isolated laser is significantly lower. Adding the double isolator improves the linearity for both second and third order distortion. It is important to note, however, that the overall output intercept points have dropped in the case of the long fiber. This is due to the fiber introducing 10 dB of loss, and the primary distortion mechanism is from the laser itself. If you add the 10 dB of loss back to the 20 km numbers, the double-isolated intercept points match the 0 km measurements, matching our expectations.

We can see similar behavior in noise by measuring the devices in the same manner as described earlier. The results are shown in Figure 4.16[4]. With the single-isolator laser, the noise increases by more than an order of magnitude when we add the 20 km of fiber, which is much more than the predicted results from Figure 4.9. With the double-isolated laser, however, the noise is almost exactly what we predict of around 2×10^5 K.

Adding a double isolator is a simple, low-cost solution that significantly improves the performance of the link in the long-haul limit. Additionally, doing so does not reduce performance in the short-fiber distances. The output power of the laser is lower in the double-isolated case due to the doubled insertion loss of the isolator, but this only mildly increases the noise.

[4]Ripples are due to uncalibrated cables that connect to the bare diodes, as this measurement was performed on a spectrum analyzer with a tracking generator and not a network analyzer

4.2 System Design

Link Budget

For the DSA-2000, we are presented with a few key science requirements that induce engineering requirements. For us, the most important metric is survey speed. Our ability to survey the sky to high sensitivity with a fast cadence is one reason this instrument will be impactful. Survey speed is a consequence of the radiometer equation, where we compute the RMS uncertainty in a noise measurement by

$$\sigma_T = \frac{T_{\text{Sys}}}{\sqrt{B\tau}}, \tag{4.29}$$

given system temperature T_{Sys}, bandwidth B and integration time τ. If our goal is to have high confidence in our results, say 5σ and the telescope has some fixed bandwidth, we have two "free" parameters: T_{Sys} and τ. These are free in the sense that T_{Sys} is an engineering requirement and τ is an operational requirement. If T_{Sys} were very low, we could integrate over much less time to achieve the same statistical result. For a large survey, we must perform this integration for every piece of the sky we want to observe. As such, driving down T_{Sys} is critical for the survey speed performance, quadratically so.

To not significantly perturb the survey speed, we chose to place a limit of no more than 1 K added to T_{Sys} from electronics following the LNA. As the LNA has ≈35 dB of gain, this implies a maximum tolerable noise of ≈3100 K at the output of the LNA.

Next, we focus on the total gain requirement. The amount of absolute power required to drive the digitizers is dependent on the exact digitizer we use. The current design uses the 12-bit AD9207[5] from Analog Devices. The full-scale peak-to-peak input voltage is 1.475 V across the differential 100 Ω input. We want to set the gain of the analog signal path such that the astronomical signal exercises about three bits [12]. A detectable signal must be above the device's quantization noise, reducing the effective bit-depth of the ADC to an "effective number of bits" (ENOB). Our ADC has an ENOB of about 9 at 2 GHz. This implies that three bits correspond to a peak-to-peak voltage of 23.05 mV, or −31.78 dBm into the 100 Ω differential input. At the input of the analog signal path, T_{Sys} induces a noise power of −93.48 dBm (25 K across our 1.3 GHz bandwidth). Therefore, we require at least 61.7 dB of gain. About 35 dB of gain is covered by the LNA, leaving only 26.7 dB needed for the remainder of the ASP.

[5] `https://www.analog.com/media/en/technical-documentation/data-sheets/ad9207.pdf`

Finally, we must consider linearity. The signal levels from astronomical sources will be incredibly faint, so their intermodulation products are unimportant. Unfortunately, radio frequency interference (RFI) can be terribly strong, wreaking havoc on our science data. The impact of RFI can be categorized into three domains. First, if the RFI is faint enough to generate intermodulation products beneath the power of T_{Sys}, all we must do is flag the carrier channels as contaminated. Second, the RFI power could be strong enough to generate intermodulation products that also contaminate the science band, implying we would need to flag more than the channel of the RFI. Finally, the RFI could be so strong that we compress the amplifiers or saturate the ADC, resulting in an unusable signal path. In the final case, there is not much that can be done in ASP design, depending on what saturates first. If everything were linear, this gives us ≈50 dB of headroom between the ADC noise and the brightest RFI we can tolerate before the ADCs saturate. This motivates using highly linear amplifiers after the LNA, such that we only need to worry about the ADC. In the first and second cases, we can derive a constraint in the second and third order intermodulation terms if we want the intermodulation power to be no more than some fraction of the noise power in a channel.

If the signal chain has 61.7 dB of gain and the full-scale power of the ADC is 4.34 dBm, RFI will saturate the ADC at $P_{\mathrm{RFI}} = -57.36$ dB into the LNA. If we constrain the linearity such that the ASP does not generate intermodulation products higher than the noise level of $P_N = -93.48$ dBm at the input, then we can compute lower limits on the input second-order intercept as $\mathrm{IIP2} = 2P_{\mathrm{RFI}} - P_N = -21.23$ dBm and third-order intercept as $\mathrm{IIP3} = (3P_{\mathrm{RFI}} - P_N)/2 = -39.29$ dBm. Referred to the output under 61.7 dB of gain, these bounds are an OIP2 and OIP3 of 40.46 dBm and 22.4 dBm, respectively.

Cascade Analysis

To optimize the analog signal path, we compute the input-referred noise and total cascaded linearity. Computing the total link noise follows from the Friis cascaded noise equation

$$T_{\mathrm{In}} = T_{\mathrm{LNA}} + \frac{T_1}{G_{\mathrm{LNA}}} + \frac{T_2}{G_{\mathrm{LNA}} G_2} \dots , \tag{4.30}$$

where each subsequent noise contributor is divided by the available gain of all the components before it. Linearity, however, is again a bit tricky. As the intermodulation powers are deterministic, they may be in-phase and add coherently. Following the derivation in [9], we find the total OIP3 of a cascaded system with coherent products as

$$\text{OIP3 (W)} = \left(\frac{1}{G_n \text{OIP3}_{n-1}} + \frac{1}{\text{OIP3}_n} \right)^{-1} , \tag{4.31}$$

where G_n is the gain of the nth component along with its OIP_{3_n}, $\text{OIP}_{3_{n-1}}$ is the cumulative OIP_3 through the previous $n - 1$ stages. All of these calculations are performed in linear units. This equation is a worst-case estimate, assuming every intermodulation product adds coherently. This will never be the case as there is a relatively random distribution of cable lengths, interconnects, and component phase delays. As such, we can approximate the phases between each stage to be random, and compute the associated cascaded intercept point as

$$\text{OIP3 (W)} = \left(\frac{1}{G_n^2 \text{OIP3}_{n-1}^2} + \frac{1}{\text{OIP3}_n^2} \right)^{-1/2} . \tag{4.32}$$

We can then input-refer these quantities by dividing by the accumulated gain as

$$\text{IIP}n = \text{OIP}n/G \, . \tag{4.33}$$

Stated another way, if we assume the intermodulation terms have random phase, then their powers would add. The derivation for 2nd order distortion and cascade analysis can be found in Appendix B.

Using these results, we compute the cascaded behavior for all the components along the analog signal path, shown in Table 4.1. Our noise only increases by ≈0.7 K, beneath our requirement of 1 K. The total gain is 74.7 dB, with adjustable attenuators allowing for control. We set the attenuators to match our gain requirement and to maximize dynamic range. Finally, the cascaded OIP2 and OIP3 are 45.83 dB and 33.57 dBm respectively, exceeding our requirements.

Table 4.1: Cascade analysis of analog signal path. Included are coherent (Coh.) and random (Rand.) output intercept points. Noise at the input LNA represents the entire system noise temperature. Laser noise is assuming a worst-case of 1×10^6 K. Laser linearity uses the double-isolator result from Figure 4.15. Passive components are taken to have infinite linearity.

		Component				Cascaded					
Description	Part Number	Gain (dB)	Noise (K)	OIP2 (dBm)	OIP3 (dBm)	Acc. Gain (dB)	Noise (K)	OIP3 Coh. (dBm)	OIP3 Rand. (dBm)	OIP2 Coh. (dBm)	OIP2 Rand. (dBm)
LNA	ksWBLNAv2	35	25	22	16	35	25	16	16	22	22
Cable	-	-3	288.63	–	–	32	25.09	13	13	18.99	19
Bias Tee	Lumped	-0.1	6.75	–	–	31.9	25.10	12.9	12.9	18.88	18.9
BPF	Lumped	-2	169.62	–	–	29.9	25.21	10.9	10.9	16.88	16.9
Amplifier	PGA-105	15	170	49	39	44.9	25.38	25.69	25.89	30.75	31.82
External Filter	-	0	0	–	–	44.9	25.38	25.69	25.89	30.72	31.82
Attenuator	F1958	-2	169.62	80	64	42.9	25.38	23.69	23.89	28.69	29.82
Amplifier	PGA-105	15	170	49	39	57.9	25.38	35.83	37.44	39.93	43.41
Attenuator	PAT-1220C	-3	288.63	–	–	54.9	25.39	32.83	34.44	36.87	40.41
Amplifier	PGA-105	15	170	49	39	69.9	25.39	38.47	38.98	44.29	48.11
BFP	Lumped	-1	75.09	–	–	68.9	25.39	37.47	37.98	43.17	47.1
Attenuator	PAT-1220C	-3	288.63	–	–	65.9	25.39	34.47	34.98	40.08	44.1
Bias Tee	Lumped	-0.1	6.75	–	–	65.8	25.39	34.37	34.88	39.89	44
RFoF Link	-	-25	1E6	20	0	40.8	25.66	-0.48	-0.02	11.06	16.46
Bias Tee	Lumped	-0.1	6.75	–	–	40.7	25.66	-0.58	-0.12	10.95	16.36
Attenuator	PAT-1220C	-3	288.63	–	–	37.7	25.68	-3.58	-3.12	7.95	13.36
Amplifier	PGA-105	15	170	49	39	52.7	25.71	11.42	11.88	22.53	28.32
Attenuator	F1958	-3	288.63	80	64	49.7	25.71	8.42	8.88	19.52	25.32
BPF	Lumped	-1	75.09	–	–	48.7	25.71	7.42	7.88	18.51	24.32
Amplifier	PGA-105	15	170	49	39	63.7	25.72	22.32	22.88	32.16	38.88
Attenuator	PAT-1220C	-3	288.63	–	–	60.7	25.72	19.32	19.88	29.14	35.88
Amplifier	PGA-105	15	170	49	39	75.7	25.72	33.05	34.57	40.21	46.83
AA Filter	Distributed	-1	75.09	–	–	74.7	25.72	32.05	33.57	39.13	45.83
ADC	AD9207										

FTX and FRX Boards

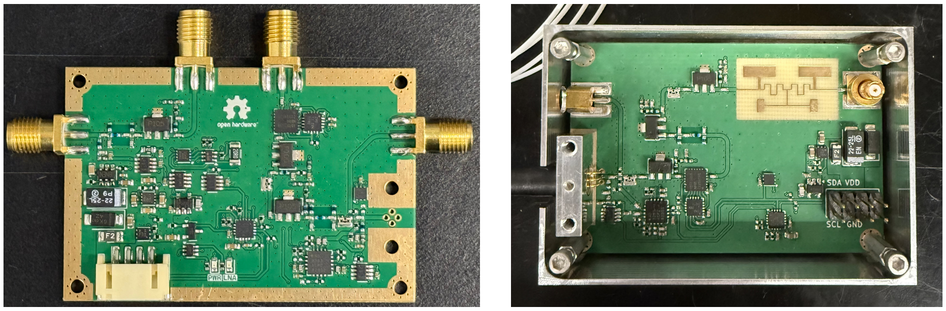


Figure 4.17: FTX and FRX PCBs

To implement the analog signal path, we designed two printed circuit boards, shown in Figure 4.17. These boards are the fiber transmitter (FTX) at the antenna station and the fiber receiver (FRX) at the central processing facility. Both of these boards include monitor and control capabilities, measuring board currents and temperatures, and controlling digital step attenuators and bias currents. This interface is achieved via an I^2C serial interface, whose digital lines are heavily filtered and will be disabled during observations.

FTX

The FTX board consists of three stages of amplification with a collection of filters and attenuators. The laser bias is provided by a controlled current source, whose set point is controlled by a digital potentiometer. The laser's anode is connected to the case, which implies we must drive the laser using a negative supply voltage. Controlled current sources that allow for this are uncommon, motivating a custom design. The design allows for a current selection between 0 mA and 50 mA, covering the full range of acceptable currents for the AGx laser.

In addition to the laser control, the FTX includes a connection point for an external filter. In the current design stage, it is unclear if we will require much filtering besides shaping the 1280-1530 MHz bandpass. In earlier stages of the project, strong interference causing significant distortion was a top concern. The current RFI environment of the DSA-2000 site seems to indicate that no additional filtering will be required. Nevertheless, the option to add the filters will remain if this situation changes.

The FTX board itself is housed inside an RFI-tight container, with feed-through connections for the RF, digital, and power signals. The switching power supply for the board is housed on the outside of the enclosure to prevent switching noise from leaking into the signal path. Each box contains electronics for both polarizations.

FRX

At the receiving side, the FRX module is a daughter board to the "radio camera frontend" or RCF. The RCF is a larger card that hosts the ADCs and FPGA and fits into a 10-unit Eurocard subrack. Two FRXes must then fit on a standard height card (100 mm), with some gap for edge interfaces. The RF signals out of the FRXes are connected to the digitizers on the RCF using a standard SMP board-to-board connector. The FRXes are shielded with an aluminum cover that encloses the modules and are electrically connected to a ground plane on the RCF PCB. The signals are routed on inner layers of the RCF PCB, so the FRX boards are surrounded by ground on all sides, except for a small slit that allows the fiber to exit. The fiber pigtail is connected to the front panel of the card with a standard fiber interface.

The design of the FRX card was straightforward, with a standard assortment of filters and amplifiers. Perhaps the most interesting component is the distributed low pass antialiasing filter at the output. This filter is an inverse-Chebyshov low pass filter that has been optimized in physical area. The center resonator was split in half and folded inwards to minimize space.

4.3 LNA Improvements

As the project progresses, the DSA-2000 team has been charged with pushing the performance even further, both in terms of what can be achieved before final design review and as avenues for future upgrades. As the goal is to continue maximizing survey speed, we want to exhaust all options to reduce system noise within our budget.

For the current frequency range, there is very little we can do to improve the noise from a circuit design perspective. The presented LNA from Chapter 3 is close to the theoretical minimum noise of the transistor, with typical losses of the matching network. While this design is optimal for room temperature operation, it could benefit from cooling, as is the case for any transistor. The entire premise of the ambient temperature LNA was to avoid cooling, but we can accomplish some cooling using Peltier microcoolers without resorting to expensive cryogenics. Additionally, the noise of the amplifier is dependent on the source impedance presented by the feed. As such, the design of the LNA input matching network could be modified to provide a better match to the feed instead of to 50 Ω. The following section discusses these two potential improvements.

Active Cooling

Thermoelectric cooling of LNAs is a well-known approach [13] for improving an amplifier's noise. However, most implementations of this strategy involve "bulk" cooling, where the entire amplifier assembly is cooled. The power required for bulk cooling to achieve reasonable noise improvements can be significant. Additionally, the hot-side of the cooler would need a large heat sink and fan for thermal management, adding size, weight, and power. An alternative for hybrid amplifiers is to cool only the first-stage transistor, as that has the highest impact on the LNA's noise. These transistors are very small and low power, requiring significantly less cooler power than what would be required to achieve the same noise improvement in bulk cooling. Modern advancements in miniaturization of Peltier coolers for optoelectronics[6] have dramatically lowered cost and improved performance. As such, we have designed an amplifier with an embedded microcooler for the first stage transistor to improve the noise without significant impact to cost or power consumption. Variants on this idea have been shown before [14], but have been limited to MMIC-based amplifiers, which require a larger cold-side area and cooling capacity.

The potential noise gains, however, need to be weighed against the complications that could arise due to creating a below-freezing point inside the amplifier. This presents a formidable design challenge, as the goal for the project is to have an exceptionally low mean-time-to-failure. If water vapor were to leak into the chassis, it would freeze over the transistor and result in failures. To prevent this, we need to design an LNA that is hermetically sealed, which adds complexity and cost. The improvement in survey speed, however, could be substantial depending on what is achievable and may outweigh the added engineering costs.

The current LNA design biases the transistor at a power of 8 mW (0.5 V at 16 mA of drain current). Models for this transistor at 2 GHz indicate a minimum noise temperature of 7 K at 300 K ambient and 0.6 K at 15 K ambient. This results in a noise slope of about 1 K per 45 K change in ambient temperature. To improve our amplifier by 1 K, we can integrate a small single-stage microcooler to generate the 45 K differential.

To fully estimate the heat load that we would need to cool, we must account for heat

[6]Typically used for lasers in high-speed digital communication

introduced from the short wire bonds. We can compute this via

$$Q_{\text{Wires}} = NK\frac{S}{L}\Delta T\,, \tag{4.34}$$

where N is the number of wires, K is the wire material thermal conductivity, S is the wire cross-sectional area, L is the wire length, and ΔT is the temperature differential. If we require $N = 5$, 20 µm diameter (S =2.54 × 10^{-10} m^2) wire bonds (three for the transistor and two for the thermistor) made of gold (K =314 $\frac{\text{W}}{\text{m K}}$) with an average length of L =500 µm across the ΔT of 45 K, we have an additional heat load of 36 mW. This implies that the bulk of the required cooling capacity comes from the wire bonds and not the transistor.

As the ΔT produced by a Peltier cooler behaves non-linearly under different heat loads, a bit of trial and error is required to estimate their performance properly. Estimations from the manufacturer[7] for their `1TC22-0044-25.H` model predicts a maximum ΔT of just beneath 30 K for 45 mW of heat. With this ΔT, however, the cooling requirement will be smaller as the heat from the wire bonds is less due to the lower temperature differential. In this scenario, we can compute a total wire bond head load of 32 mW, which would induce a new ΔT closer to 40 K. This all implies we can expect a measured ΔT of between 30 K and 40 K.

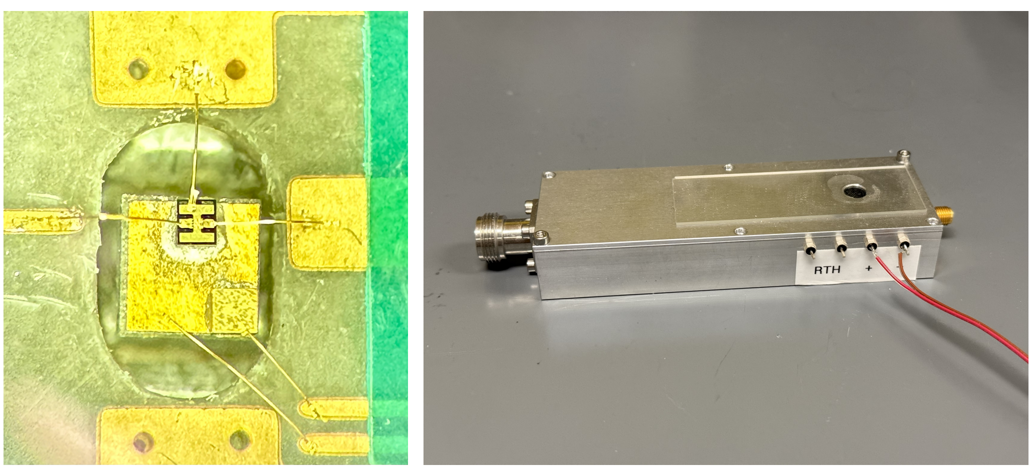


Figure 4.18: Bonded transistor on microcooler and constructed cooled amplifier

We constructed a copy of the amplifier described in Chapter 3, but with an embedded microcooler. While this amplifier was not hermetically sealed, two holes were added at either end, with dry nitrogen pumped through one side. Additionally, a hole and

[7] `https://www.tec-microsystems.com/`

microscope slide was added to the lid to observe the first stage transistor under cooling to watch for ice formation. The microcooler was powered via an independent supply, and a thermistor was mounted to the cold-stage, adjacent to the transistor to monitor the temperature. Finally, we performed the precision noise measurement described in Chapter 3. A close-up picture of the bonded transistor (center) and thermistor (bottom right) on the microcooler and assembled amplifier is shown in Figure 4.18. The transistor's wire bonds from left to right are the gate, source, and drain.

With zero power applied to the microcooler, we measured a thermistor temperature of 30 C. With the cooler powered to its maximum current, we measured a thermistor temperature of -10 C, for a ΔT of 40 C. The LNA noise temperature dropped by an average of about 0.75 K, shown in Figure 4.20, which is very close to the predicted 0.89 K and well within the error for this measurement.

While the experiment was a success, eventually ice did start to form on the first stage transistor, destroying the fragile air bridges. Work on this solid-state cooling approach needs further investigation in environmental control, especially as it pertains to mass-manufacture for the 4000 LNAs that would eventually need to be built. Additionally, our original estimations of the heat load which influenced the sizing of the microcooler did not include the heat load of the wire bonds. The manufacturer has suggested a new custom part, `1TC22-0040-21.H`, which should have a higher ΔT of about 50 K under our heat load of $\approx$40 mW. We could also investigate two-stage coolers for significantly more cooling capacity, but with a downside of doubling the cross-sectional area of the cold-stage. This would increase wire bond length which would reduce the heat load, but add significant inductance which will alter our matching network.

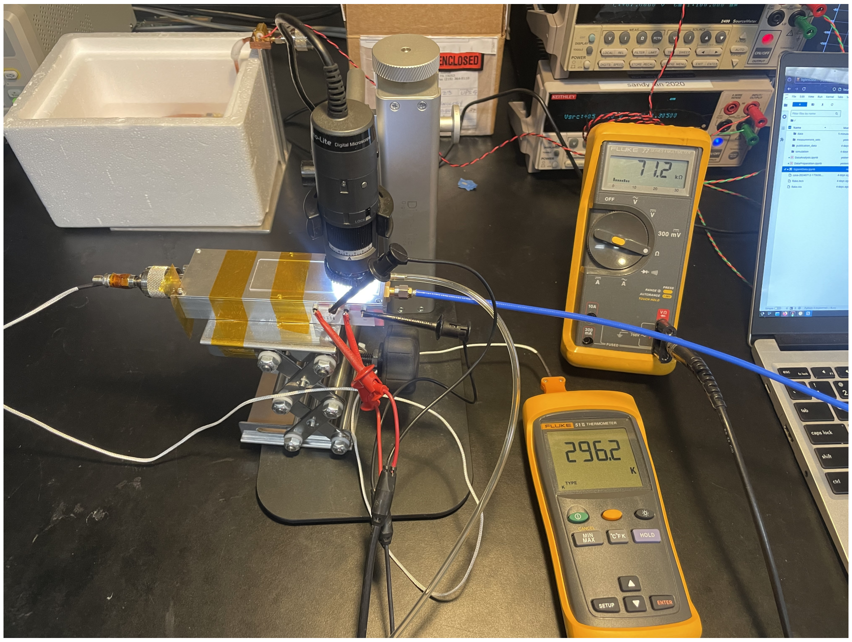


Figure 4.19: Test setup for cooled amplifier noise measurements

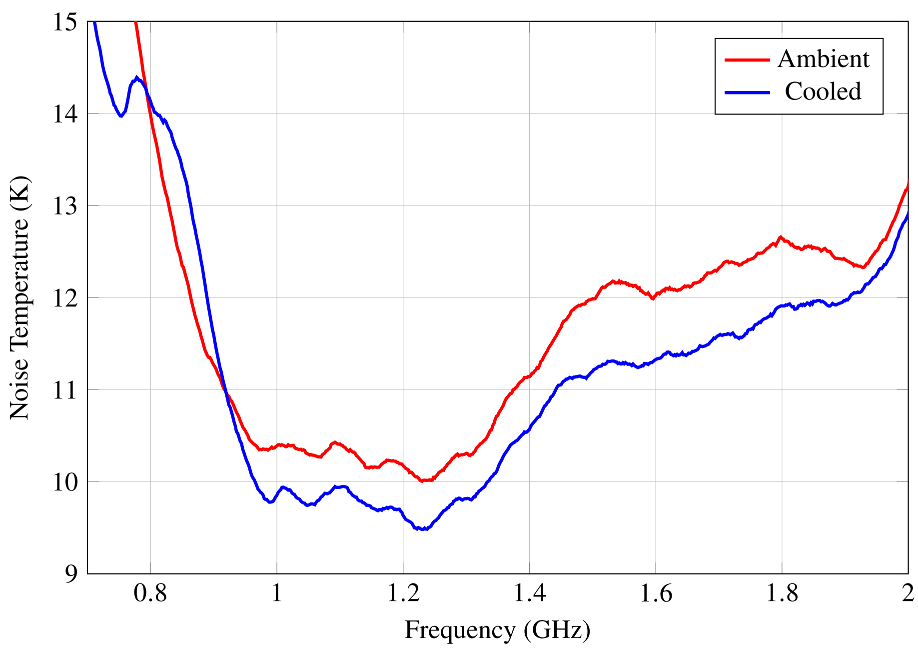


Figure 4.20: Measured noise temperature of cooled amplifier

Optimization of LNA Input Matching Network to Feed Impedance

All our amplifier designs up to this point are intended to operate in a $50\,\Omega$ system, but the impedance presented by the feed is not $50\,\Omega$. The feed has return loss greater than 10 dB over much of the band, but the mismatch at the band edges significantly degrades the noise.

To reduce the effect of the feed impedance "pulling" the noise, we can design a matching network that is optimized for the feed impedance. If the feed impedance were a large deviation from $50\,\Omega$, it would be difficult to validate performance in the lab, as all the noise experiments are done in a $50\,\Omega$ system. However, as the feed was intended to match to $50\,\Omega$, we only have to compensate for a small difference in impedance, leaving the $50\,\Omega$ bench test more than satisfactory for validating performance.

Performing the co-optimization follows the work from Chapter 3, with a small modification made to the definition of the noise goal in the cost expression. Before, we were simply using S_{22} of the nonuniform input matching network as the source reflection coefficient Γ_S needed to compute the noise via Equation 2.3. Now, we take measured S_{11} data of the feed and compute Γ_{Out} via

$$\Gamma_{\text{Out}} = S_{22} + \frac{S_{21}S_{12}\Gamma_S}{1 - S_{11}\Gamma_S}, \tag{4.35}$$

where Γ_S is the measured feed S_{11} (assuming everything is normalized to the same reference impedance).

The current feed design [15], [16], is a quad-ridge flared horn, following previous work at Caltech [17]. This feed includes a dielectric lens and "cake pan" choke rings to optimize the illumination of the DSA-2000 reflector. As the ridges form a balanced transmission line, there is a shorted-pin balun that transitions from coax. This transition is then brought to the edge of the feed structure to a standard microwave connector. In the current design, the LNA is attached directly to this connector, resulting in the feed and LNA as two separate line-replaceable units (LRUs).

While the separation of components is useful from a maintenance perspective, the long coax pin that attaches the connector to the throat of the feed introduces additional loss. An alternative design is to embed the LNA into the feed, using the ridge as the enclosure for the LNA PCB. This sidesteps the need for the long coax pin and allows us to start the matching network very close to the slotline transition. We can accomplish this while maintaining two distinct LRUs by designing an LNA chassis that slots into the ridge and connects to a blind-mate pin. The new LNA chassis has a silver-epoxied 50 Ω glass feed-through connector to connect to the internal blind-mate pin. A lid gasket and hermetic output connector seal the assembly and allow for solid-state cooling.

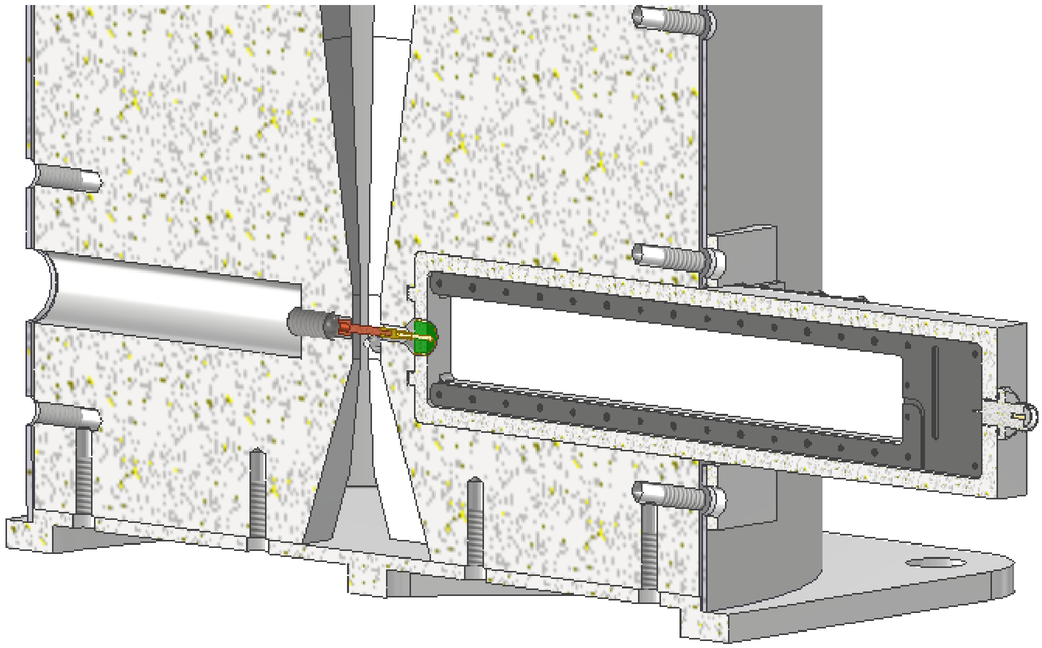


Figure 4.21: Cross-section of an LNA-embedded feed

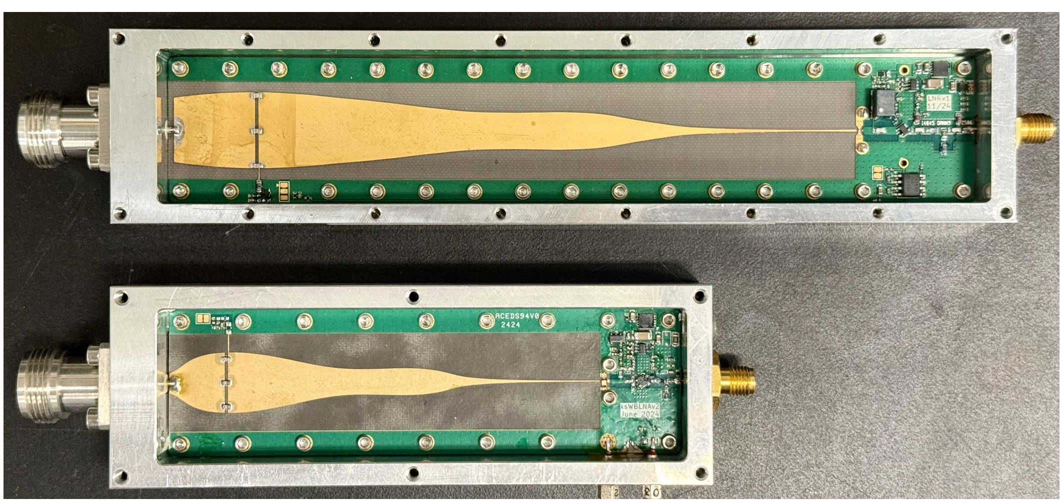


Figure 4.22: Comparison of the co-optimized LNA (in a test chassis) and the previous 50 Ω design

The results from the optimization were surprising, as the optimal nonuniform line is significantly longer than the previous design. This seems to indicate a shift in the optimum insertion loss versus noise match space. At its core, the nonuniform matching approach requires long lines to successfully match complex loads [18]. As discussed in Chapter 3, a well-matched, long line is at odds with the insertion loss of the line itself, implying that there exists some optimum compromise between the length of the line providing a quality noise match and the loss of the line. When co-optimizing with the DSA-2000 feed, this optimum has shifted to favor longer lines. In short, the design can tolerate the loss introduced from the longer line, as the noise otherwise introduced due to mismatch at the band edges is so great.

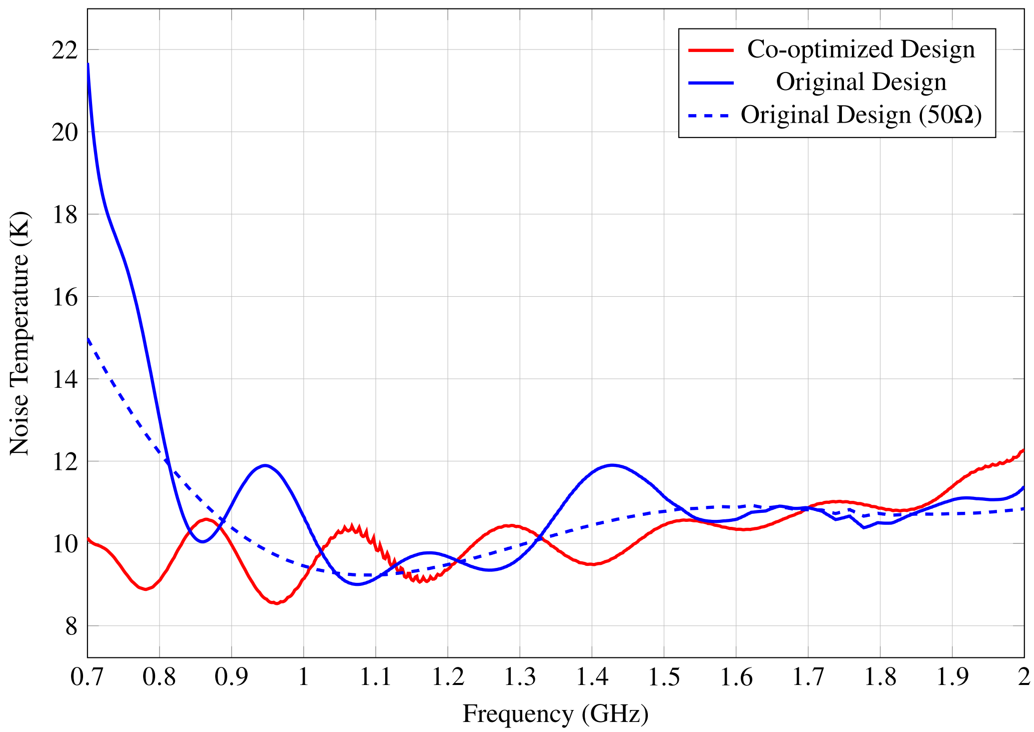


Figure 4.23: Simulated noise temperature of co-optimized and original LNA designs given a simulated feed source impedance

The simulated equivalent noise temperatures for the previous design and the new design are presented in Figure 4.23. Although the difference in mean noise temperature is small (≈1 K), the improvement in noise from 700 MHz to 800 MHz is massive. As survey speed is $\propto \lambda^2$, this improvement in the low frequency performance is huge.

References

[1] F. Perini, S. Rusticelli, M. Schiaffino, *et al.*, "Radio frequency over fiber technology for SKA-low receiver," *Journal of Astronomical Telescopes, Instruments, and Systems*, vol. 8, p. 011 016, Jan. 1, 2022. DOI: 10.1117/1.JATIS.8.1.011016.

[2] J. Mena, K. Bandura, J. .-. Cliche, M. Dobbs, A. Gilbert, and Q. Y. Tang, "A Radio-Frequency-over-Fiber link for large-array radio astronomy applications," *Journal of Instrumentation*, vol. 8, T10003, Oct. 1, 2013. DOI: 2017022022421000.

[3] S. Montebugnoli, M. Boschi, F. Perini, P. Faccin, G. Brunori, and E. Pirazzini, "Large antenna array remoting using radio-over-fiber techniques for radio astronomical application," *Microwave and Optical Technology Letters*, vol. 46, no. 1, pp. 48–54, 2005. DOI: 10.1002/mop.20898.

[4] C. Cox, E. Ackerman, and G. Betts, "Relationship between gain and noise figure of an optical analog link," in *1996 IEEE MTT-S International Microwave Symposium Digest*, vol. 3, Jun. 1996, 1551–1554 vol.3. DOI: 10.1109/MWSYM.1996.512232.

[5] C. Cox, E. Ackerman, G. Betts, and J. Prince, "Limits on the performance of RF-over-fiber links and their impact on device design," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 2, pp. 906–920, Feb. 2006. DOI: 10.1109/TMTT.2005.863818.

[6] H. Haug and H. Haken, "Theory of noise in semiconductor laser emission," *Zeitschrift für Physik A Hadrons and nuclei*, vol. 204, no. 3, pp. 262–275, Jun. 1967. DOI: 10.1007/BF01326200.

[7] R. Hui and M. O'sullivan, "Chapter 1 - Fundamentals of optical devices," in *Fiber-Optic Measurement Techniques (Second Edition)*, R. Hui and M. O'sullivan, Eds., Academic Press, Jan. 2023, pp. 1–135. DOI: 10.1016/B978-0-323-90957-0.00002-3.

[8] I. S. Kashevsky and V. S. Speransky, "Methods for Measuring Relative Noise Intensity (RIN) of Laser Radiation," *Synchroinfo Journal*, vol. 9, no. 4, pp. 15–20, 2023. DOI: 10.36724/2664-066X-2023-9-4-15-20.

[9] D. M. Pozar, *Microwave Engineering*, 4th ed. Hoboken, NJ: Wiley, 2012.

[10] J. Nanni, A. Giovannini, M. U. Hadi, *et al.*, "Controlling Rayleigh-Backscattering-Induced Distortion in Radio Over Fiber Systems for Radioastronomic Applications," *Journal of Lightwave Technology*, vol. 38, no. 19, pp. 5393–5405, Oct. 2020.

[11] H. Kanamori, "Transmission Loss of Optical Fibers; Achievements in Half a Century," *IEICE Transactions on Communications*, vol. E104.B, no. 8, pp. 922–933, Aug. 1, 2021. DOI: 10.1587/transcom.2020EBI0002.

[12] G. Bianchi, F. Perini, C. Bortolotti, J. Monari, S. Montebugnoli, and M. Roma, “ADC bit number and input power needed, in new radio-astronomical applications,” in *Proceeding of 13th Workshop on ADC Modelling and Testing*, 2008, pp. 1064–1069.

[13] H. S. Cao, R. H. Witvers, S. Vanapalli, H. J. Holland, and H. J. M. Ter Brake, “Cooling a low noise amplifier with a micromachined cryogenic cooler,” *Review of Scientific Instruments*, vol. 84, no. 10, p. 105 102, Oct. 1, 2013. DOI: 10.1063/1.4823528.

[14] F. Schreuder and J.-G. Bij de Vaate, “Localized LNA Vooling in Vacuum,” in *Proceedings of 12th International Workshop on Thermal Investigations of Ics*, Nice, France: TIMA Editions, Sep. 2006, pp. 175–179.

[15] J. Flygare and D. P. Woody, “Wideband Feed and Reflector Ground Shield for the Deep Synoptic Array (DSA-2000),” in *2023 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (USNC-URSI)*, Portland, OR, USA: IEEE, Jul. 2023, pp. 761–762. DOI: 10.1109/USNC-URSI52151.2023.10237516.

[16] J. Flygare, “Wideband Low-Loss Feed Design for the DSA-2000 Ambient Temperature Array in Radio Astronomy,” in *2024 IEEE International Symposium on Antennas and Propagation and INC/USNC-URSI Radio Science Meeting (AP-S/INC-USNC-URSI)*, Firenze, Italy: IEEE, Jul. 2024, pp. 5–6. DOI: 10.1109/AP-S/INC-USNC-URSI52054.2024.10686435.

[17] A. H. Akgiray, “New Technologies Driving Decade-Bandwidth Radio Astronomy: Quad-ridged Flared Horn and Compound-Semiconductor LNAs,” Ph.D. dissertation, California Institute of Technology, 2013. DOI: 10.7907/TYX5-2C48.

[18] G. Xiao and K. Yashiro, “Impedance matching for complex loads through nonuniform transmission lines,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 6, pp. 1520–1525, Jun. 2002. DOI: 10.1109/TMTT.2002.1006413.

Chapter 5

GREX: AN INSTRUMENT OVERVIEW AND NEW UPPER LIMITS ON THE GALACTIC FRB POPULATION

[1] K. A. Shila, S. Niedbalski, L. Connor, *et al.*, “GReX: An Instrument Overview and New Upper Limits on the Galactic FRB Population.,” *Publications of the Astronomical Society of the Pacific*, 2025, In Review. arXiv: `2504.18680 [astro-ph.HE]`,

We present the instrument design and initial results for the Galactic Radio Explorer (GReX), an all-sky monitor for exceptionally bright transients in the radio sky. This instrument builds on the success of STARE2 to search for fast radio bursts (FRBs) from the Milky Way and its satellites. This instrument has deployments across the globe, with wide sky coverage and searching down to 32 μs time resolution, enabling the discovery of new super giant pulses. Presented here are the details of the hardware and software design of the instrument, performance in sensitivity and other key metrics, and experience in building a global-scale, low-cost experiment. We follow this discussion with experimental results on validation of the sensitivity via hydrogen-line measurements. We then update the rate of Galactic FRBs based on non-detection in the time since FRB 200428. Our results suggest FRB-like events are even rarer than initially implied by the detection of a MJy burst from SGR J1935+2154 in April 2020.

5.1 Introduction

Fast radio bursts (FRBs) are short-duration ($\lesssim$ 100 ms), highly energetic ($\gtrsim 10^{40}$ erg s^{-1}) transients that have been detected between 100 MHz and 8 GHz [1], [2]. Prior to April 2020, all FRBs were extragalactic and most were at cosmological distances ($z \geq 0.1$). The discovery of SGR 1935+2154 by both STARE2 and the Canadian Hydrogen Intensity Mapping Experiment (CHIME/FRB) made a dramatic connection between extragalactic FRBs and Galactic magnetars [3], [4]. While FRB 200428 from SGR 1935+2154 was less luminous than all known extragalactic FRBs, its proximity made it the highest fluence of any exosolar radio burst ($\sim$ MJy ms at 1 GHz, compared to $\lesssim$ kJy ms at $\sim$1 GHz for high fluence extragalactic FRBs [5]) and enabled detection of the first prompt multi-wavelength emission [6].

In the years since the FRB-like emission from SGR 1935+2154, the link between young, energetic magnetars and the extragalactic FRB phenomenon has been muddied. FRBs have been found in globular clusters [7], at the outskirts of elliptical galaxies [8], [9], and may be underrepresented in low-mass star forming galaxies and the sites of recent core-collapse supernovae [10]. In other words, if all FRBs are magnetars, they are not all young remnants of CCSNe embedded in star-forming region, like SGR 1935+2154. These facts motivate a blind search survey, rather than monitoring known Galactic magnetars. The discovery of other new Galactic phenomena such as long-period radio transients [11]–[13] further motivates blind searches in new regions of parameter space.

The Galactic Radio Explorer is an international collaboration aimed at discovering exceptionally bright bursts in the radio sky. Building off the success of STARE2 [4], we are currently searching for FRBs emitted by Galactic sources as well as nearby galaxies. We search down to a high time resolution of 32.768 μs, saving raw 8.192 μs voltage data to enable more in-depth analysis for interesting single-pulse candidates. Additionally, the instrument has been designed to be easy to build and reproduce, allowing collaborators to quickly bring up a functioning station. The goal as mentioned in our original white paper [14] is to eventually have 4π steradian coverage with increased exposure to the galactic plane to improve sensitivity to galactic sources. As more sensitive aperture arrays for fast transient discovery come online, such as BURSTT [15] in Taiwan and similar efforts in the US, Australia, and Chile, the niche of GReX will be all-sky sensitivity to ultra-narrow, bright radio bursts at 1.4 GHz. Each GReX terminal is fully self-contained and low-cost, serving as a valuable pedagogical tool for the observatory or university operating it locally.

In this work, we describe the GReX instrument, and the current network of terminals around the world. We also place new upper-limits on the rate of ultra-bright FRBs based on non-detection.

5.2 Hardware Design

Analog Signal Path

The first major hardware improvement over STARE2 is the transition to a new, higher performance low noise amplifier (LNA) [16] developed at Caltech for the DSA-110 [17]. The new amplifier has a noise temperature of 7 K, versus the previous amplifier's noise temperature of $\approx$ 30 K, dramatically improving our system noise temperature and sensitivity. The LNA achieves record-breaking noise at ambient temperature via the use of an exceptional low noise transistor from Diramics and a low loss suspended stripline input matching network.

For the antenna, we use an improved version of the same "cake pan" antenna from STARE2, also designed for the DSA-110. This is a waveguide horn antenna with cake pans forming axial corrugations, enhancing the beam pattern. The antenna's full-width half-maximum (FWHM) is estimated to be $70° \pm 5°$ at 1.4 GHz with very little variation across our band of 1280 to 1530 MHz. We verify the estimation of the beam width in Section 5.5 by fitting HI intensity data. As this is an all-sky instrument, the large beam width trades sensitivity for field of view, matching our science goals.

Frontend Module

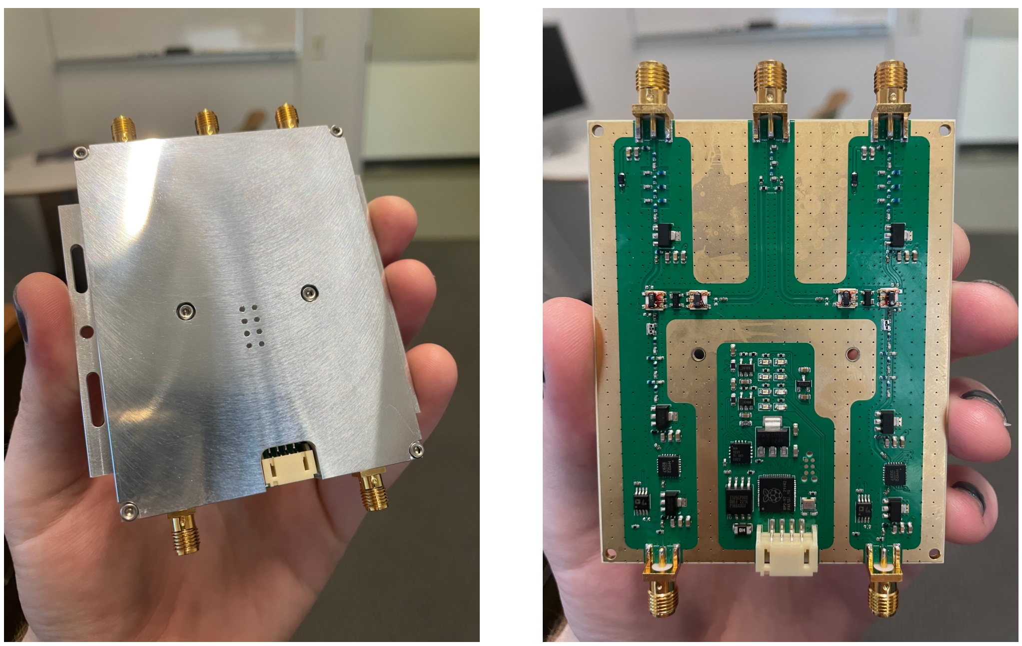


Figure 5.1: Completed FEMs

The heart of the analog signal processing in GReX is the frontend module (FEM). In STARE2, the analog processing was split between electronics at the antenna and in the server room, with an RF over fiber (RFoF) link between the two sides. Instead of RFoF, we digitize directly at the antenna, incorporating all the electronics into a single enclosure to which the antenna is bolted. As RFoF links are typically quite nonlinear, this section of the analog signal path could limit the linear dynamic range. To replace the RFoF circuitry, we designed the FEM to perform the frequency downconversion, filtering, and amplification needed to prepare the signal for digitization.

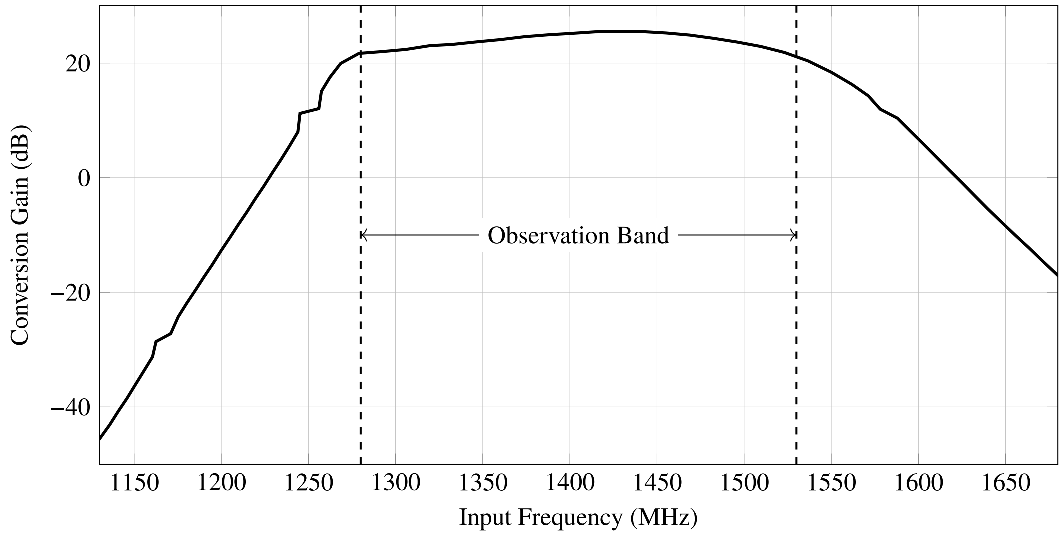


Figure 5.2: Measured conversion gain of the FEM. Total system gain adds 40 dB from the LNA and 15 dB from ADC preamplifiers

The FEM also contains digital control electronics using an RP2040 microcontroller. This device controls the power for the LNAs connected to the FEM inputs, sets the variable attenuators, and monitors temperatures and average RF power levels. We wrote the firmware in the Rust programming language using the RTIC [18] framework for real-time task-based concurrency. Monitoring and control is performed via a simple serial connection.

The FEM enclosure is custom machined with channels that isolate the two polarizations from the digital electronics. This enclosure also acts as the heat sink, as it is bolted directly to a large aluminum panel inside the box. The FEMs were fully factory-assembled, with the total per-unit cost under $200 USD. The completed FEM enclosure and PCB are shown in Figure 5.1. The conversion gain of the FEM is shown in Figure 5.2.

The Box

The constructed telescope is contained in a single weatherproof aluminum box 20″ tall, 16″ wide, and 6″ deep. This box contains a GPS timing system and oscillators, the FEM, the FPGA digitizer, power supplies, a 10G Ethernet switch, and a Raspberry Pi for monitor and control. As the box we chose had no inherent RFI shielding, we added an aftermarket weatherproof RFI gasket. The FEM, synthesizer, GPS receiver, and FPGA digitizer are bolted to a subpanel inside the box. The two polarization

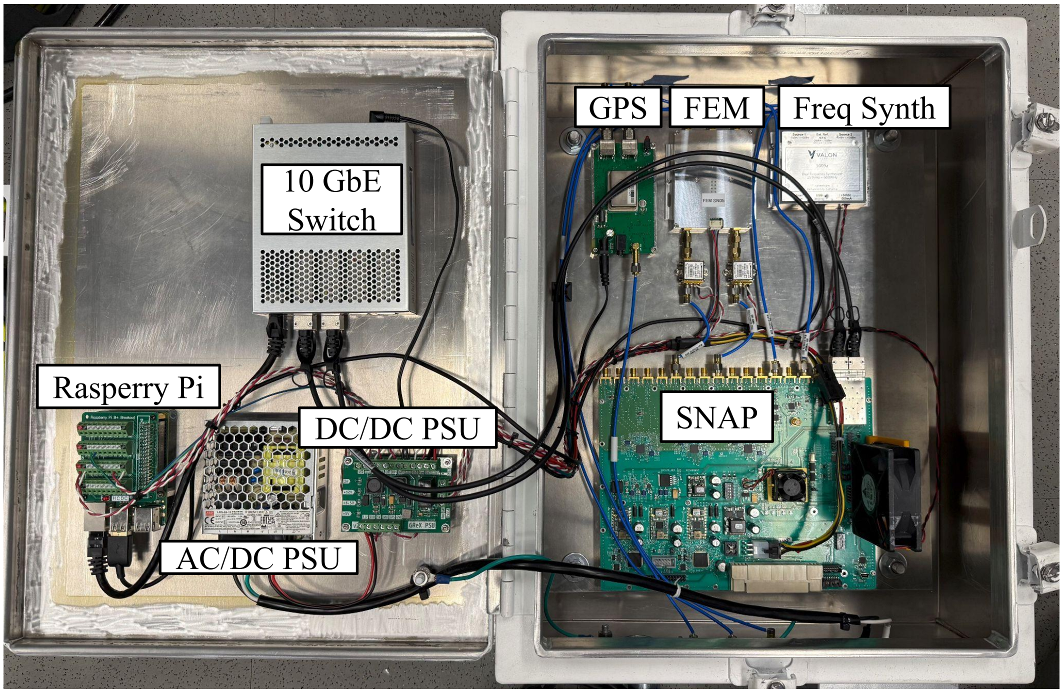


Figure 5.3: Inside a GReX box. The left panel contains the 10 GbE switch, two power supply units (PSU), and the Raspberry Pi for monitor and control. The right half contains the GPS timing receiver, frontend module (FEM), frequency synthesizer, and SNAP digitizer.

connections and GPS antenna are connected via feedthrough connectors on the bottom of the box. The optical fiber is routed through a waveguide cutoff pipe that is attached to the bottom of the box, where it is internally terminated to a 10G SFP+ connector, attached to the Ethernet switch. The box has mounting points on the outside for a stand or pipe and boltholes to mount the antenna. The antenna is bolted on the top of the box with weatherproofing gasket material to prevent any water leaks.

Installation of the box is indented to be simple, requiring two single-mode fibers for the Ethernet connection and mains AC power. Upon receiving a box, the operator needs to bolt the antenna to the top with the weatherproofing compound, then mount it to an appropriate place with a clear view of the sky. Ideally, the antenna should be covered with RF-transparent material for weatherproofing, but we found any piece of low-loss material such as a plastic trash can to be sufficient. A picture of an assembled box is shown in Figure 5.3.

5.3 Software Design

As this experiment is deployed globally, it is imperative that the software stack be easy to build, install, and use, all with very little assistance from the project maintainers. New members should be able to on-board their stations independently, just by following online documentation[1]. As such, we developed the software stack for this instrument with intense attention to these goals. As much as possible, software is tested in continuous integration and written in a way to reduce the likelihood of unexpected errors. Additionally, all the software for the project is freely accessible and under an open-source license.

Digitization and Initial Processing

Following the analog signal path, the high-frequency signals are digitized by the Smart Network ADC Processor (SNAP)[2] board. This platform contains three HMCAD1511 analog to digital converters (ADC) from Analog Devices and a Kintex-7 160T field-programmable gate array (FPGA) from Xilinx. This platform is supported by the CASPER [19] ecosystem, which we make use of here.

We wrote our gateware[3] in a combination of Simulink and SystemVerilog using the CASPER toolflow. To make maximum use of the available 10G connection, we stream 8+8 bit complex data for each polarization, channelized to 2048 frequency bins. As the FPGA core is clocked at 250 MHz, with the ADC clock running 500 MHz, the F-engine processes two ADC samples every clock cycle. The channelization is accomplished with a standard polyphase filterbank constructed via the combination of a finite impulse response (FIR) filter and fast Fourier transform (FFT), both making use of design blocks from the CASPER library. The FIR filter has 8 taps and a Hamming windowing function to reduce inter-channel leakage and scalloping loss [20]. To avoid overflows in the FFT, each stage is allowed to grow the number of bits representing each channel. The output of the FFT block is 18+18 bit, fixed-point complex numbers. If we attempted to stream these data directly, we would not have the bandwidth on the 10G link. As such, these data are requantized to a lower bit-depth of 8+8 bits to fit in a single 10G Ethernet payload. Finally, the resulting channelized voltage data is packetized for transmission. The block of data includes a 64-bit packet counter header, allowing the downstream software to detect data loss and re-ordering. These data were packed such that they match the layout of a C-struct,

[1] `https://grex-telescope.github.io`
[2] `https://casper.berkeley.edu/wiki/SNAP`
[3] `https://github.com/GReX-Telescope/gateware`

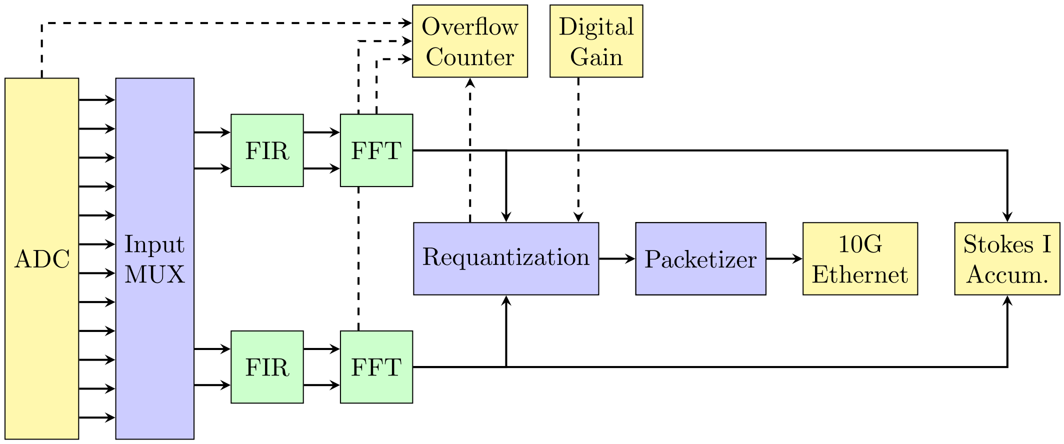


Figure 5.4: FPGA gateware internals: yellow blocks represent CASPER-provided interfaces to hardware, blue blocks are custom SystemVerilog, and green blocks are CASPER-provided DSP logic. Solid lines represent high-speed data, dashed lines represent slow monitor and control data. Reset lines and timing logic is not shown.

so no processing needs to be done to use the data downstream. The packetized data is transmitted over the optical 10G Ethernet link to the server, making use of the standard Ethernet block from CASPER. Every 2048 samples, a completed Ethernet payload is transmitted, including the header, totalling 8200 bytes. This occurs every 8.192 us, implying a total data rate of just over 1 GB/s or 8 GiB/s.

In addition to the primary functionality of the gateware, there are a few added utilities to assist in operation. First, there is an input multiplexer that allows the user to connect to any input on the SNAP without reprogramming. Second, there is a triggerable accumulator for Stokes I data that gives the user a snapshot view of the spectra without running high-speed packet capture. Additionally, this accumulation is performed on the raw output from the FFT block, providing insight into the required digital gain setting for requantization. Finally, there are various registers that monitor overflow conditions in both requatization, FFT bit growth, and raw ADC samples. This primarily gives us an indication that the total power level into the SNAP board is too high, or downstream digital gain is too high, either case being actionable via the adjustable RF attentuators in the FEM or in the digital gain settings. A figure of the data flow in the gateware is shown in Figure 5.4.

FPGA Software Interface

A critical piece of code in the project is the software interface to the running FPGA gateware. During telescope operation, status registers and accumulators are read, variables are set, etc. For this project, we chose to write a new interface to CASPER devices using the Rust programming language, *casperfpga_rs*[4]. The new library generates compile-time checked interfaces to the various components in the running FPGA design. As the generated code fully validates correctness at compile-time, runtime errors are all but eliminated.

This new library allows for much higher confidence in the deployment of CASPER designs. This library also contains unit tests against a mock interface, extensive documentation, and is written in a modular nature to encourage other members of the CASPER collaboration to add support for their hardware.

Server and First Stage Processing

After packets have been transmitted from the FPGA, they are captured and processed by a single high-performance server. This server contains a 24-core AMD Ryzen Threadripper Pro 5965WX with 128 GB of RAM. The server is also outfitted with an NVIDIA GeForce RTX 3090 Ti GPU, used for the brute-force dedispersion search. This system's kernel parameters are tuned to support the "jumbo Ethernet frames" emitted from the SNAP board. After the packets have been processed through the kernel, user-space programs perform the remaining processing.

The majority of the effort in software development for this project was in the first stage processing program, *T0*. This program performs the packet capture, computation of Stokes I, time downsampling, and management of ring buffers of voltage data to be written to disk on command. As this piece of software handles the raw voltage data, it is the most timing sensitive and performance-critical.

At the top level, *T0* consists of many independent threads, each performing a processing step called a subtask. One thread is dedicated to the packet capture subtask, one for time-downsampling, etc. We must use multithreading as the total processing time is longer than the incoming packet cadence, so parts of the processing need to be broken up and performed in parallel, i.e., pipelined. After each subtask completes computation, it passes the result to the subsequent subtask across a thread boundary.

Our implementation of this streaming processing pipeline is built off the inter-process communication model of channels with allocation-reusing ring buffers, similar in spirit to PSRDADA [21]. Unlike PSRDADA, the model is fully contained in a single executable, allowing for much stronger guarantees around the data being passed

[4]`https://github.com/kiranshila/casperfpga_rs`

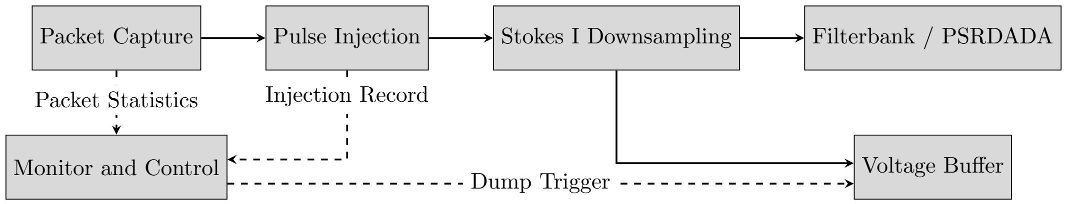


Figure 5.5: *T0* subtasks and inter-thread communication. Solid lines represent high-speed science data, dashed lines represent monitor and control.

around. Additionally, this model reduces the need for synchronization primitives such as mutexes, improving performance.

The complete architecture of *T0* is shown in Figure 5.5, with arrows showing the channels that move data across the various threads. After program start and timing synchronization (described in the Section 5.3), all the subtasks are spawned simultaneously. The first subtask again is packet capture, where incoming data is read from the network card. This subtask checks the value of the packet header against the previous to test for packet loss or reordering. Metadata about the packet statistics is transferred using a separate channel to the monitoring thread. The following subtask performs fake pulse injection. As we want to ensure our pipeline is working properly, we occasionally add synthetic data of various dispersion measures and fluence into the real-time data stream. We write information about the injection into a *SQLite* database, so downstream processing software can determine if a given candidate is synthetic. The next subtask performs time downsampling. Downstream processing, specifically brute-force dedispersion, has trouble handling exceptionally high time resolution data. As such, we need to perform downsampling in time, in addition to computing Stokes I as we will be performing incoherent dedispersion. Finally, depending on launch arguments, the downsampled Stokes I data is written to either a *SIGPROC*[5] filterbank file (using a high-performance Rust implementation[6] of the file format) or to a *PSRDADA*[7] ring buffer.

Running in parallel to the data processing pipeline, a thread is accumulating and distributing monitor information. This data includes occasional queries to the FPGA about overflow statistics, long-integration Stokes I spectra, and internal temperatures. Moreover, it collects log messages produced by the program and statistics about

[5] `https://sigproc.sourceforge.net/`
[6] `https://github.com/kiranshila/sigproc_filterbank`
[7] `https://psrdada.sourceforge.net/`

packet capture from the first subtask. The monitor subtask provides an HTTP API to query the monitor information for a Prometheus[8] time-series database. Log messages and traces are ingested by an *OpenTelemetry*[9] collector.

Data Loss

As the packets are transmitted over Ethernet with the UDP protocol, there is no guarantee that they are received. We have taken special care to ensure the server's operating system can efficiently process incoming UDP data without data loss and *T0* should be able to process incoming data in real time without issue. However, packet loss still does occur, albeit rarely. For the Owens Valley station, the average packet loss rate is $10^{-6}\%$ or one packet dropped per 10^8 packets processed. As our packet cadence is one packet per $8.192\,\mu$s, this works out to an average of one dropped packet per 1000 s. This amount of data loss is to be expected and is inconsequential to the performance of the telescope.

Timing and Synchronization

In the box, we include a GPS-discriminated 10 MHz oscillator. This device provides a stable reference and emits a pulse-per-second (PPS) square wave, connected to a general purpose digital input on the FPGA. The 10 MHz reference clock is used by a Valon 5009a frequency synthesizer to produce the 500 MHz clock used by the ADC and FPGA. Furthermore, this synthesizer is used to generate the 1030 MHz local oscillator signal used in the FEM for downconversion.

In the FPGA gateware, the PPS signal is used to timestamp data. In *T0* at program start, we query a network timeserver to get the current local time within tens of milliseconds. Then, the program waits until the next half second to "arm" the timing system on the FPGA. Once the next rising edge of the PPS signal arrives in the FPGA and starts the flow of data, we know that the very first voltage sample is coincident with the next whole second.

As each packet has a 64 bit packet counter, and we know the timestamp of the first packet and know the clock speed of the FPGA, we can work out the start time of every packet. Specifically, we know that packet N is exactly $N \cdot 8.192\,\mu$s past the time of the first packet. This information is used in the metadata stored in voltage dumps as well as in the candidate metadata.

[8]`https://prometheus.io/`
[9]`https://opentelemetry.io/`

RFI Cleaning

Following *T0*, data is transferred to an RFI cleaning program, *clean_rfi*[10] using *PSRDADA* ring buffers. This program implements a multistep cleaning approach following the implementation in CHIME/FRB [22].

For both *T0* and *clean_rfi*, we needed a Rust interface to the *PSRDADA* C library, as that is the only mechanism to stream data into the brute-force dedispersion program, *HEIMDALL* [23]. While Rust has a robust interface to C programs and libraries, they are all memory-unsafe by default, as there is no mechanism to guarantee Rust's invariants across the application binary interface (ABI) boundary. As such, we wrote a high-level Rust wrapper library called *psrdada_rs*[11]. Much like *casperfpga_rs*, this library adds a significant number of compile-time checks to guarantee correct usage before runtime.

clean_rfi implements simple masking, variance-cut, and detrending algorithms to iteratively remove RFI from a block of dynamic spectra. Specifically, we start with a static mask of bad channels for frequencies that are consistently contaminated. Then, we remove all samples that contain zeros, as the typical noise floor will be just above zero and pure-zeros represent dropped packets. Next, we remove bandpass variation by dividing each sample by the time-average frequency response. Finally, we perform variance cuts in both the frequency and time axes by removing data above some σ threshold. We perform these final variance cuts twice, with a higher σ in the second pass. Currently, the iterative variance cut process removes time samples / frequency channels that are greater than 3σ and then 5σ.

Real-Time Detection Pipeline

After the dynamic spectra data were cleaned, they are passed along to our fork[12] of HEIMDALL. We use HEIMDALL, and the associated *dedisp* library, as the implementation of the brute-force dedispersion search. Our fork removes the built-in RFI cleaning routines, removes candidate clustering, adds a structured logging library, and enables writing candidates to a network socket instead of a file. As we will be running the search in real time, the included RFI cleaning and clustering routines were too slow for our high time-resolution data.

HEIMDALL then writes lines of candidates over a local network socket to the candidate filtering task, *T2*. This program is a fork of the *T2* Python project written

[10] `https://github.com/GReX-Telescope/clean_rfi`
[11] `https://github.com/kiranshila/psrdada-rs`
[12] `https://github.com/GReX-Telescope/heimdall-astro`

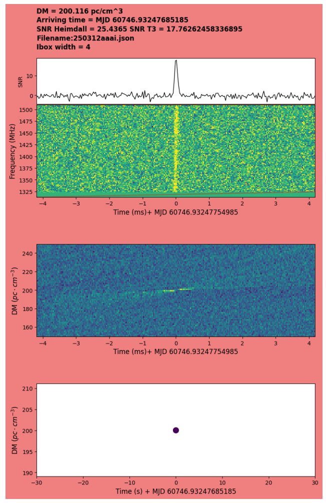


Figure 5.6: Detection candidate for an injected burst.

Figure 5.7: Grafana dashboard monitoring system performance.

for DSA-110, modified to work with our data formats. This program ingests the stream of candidates emitted by HEIMDALL, clusters them using HDBSCAN [24], filters the clustered results in dispersion measure, SNR, and boxcar width to produce candidates. Additionally, for every viable candidate, a message is sent to *T0* to dump the contents of the voltage ring buffer to disk.

T3 is the final stage in the pipeline, another fork from the DSA-110 project. This program watches the directory where candidate files are written and generates plots of the dedispersed, RFI-cleaned dynamic spectra. These candidate plots are then pushed to a Slack channel, where members of the collaboration get real-time notifications. An example of this notification is shown in Figure 5.6. Eventually, *T3* will perform machine learning-based candidate classification as well as inter-station communication for coincidencing and localization.

The whole pipeline is orchestrated through a single bash script. This script launches all the tasks sequentially using GNU Parallel [25], including the initial setup of the PSRDADA buffers. The script has several launch modes, allowing the operator to run the full processing pipeline, stream into a named PSRDADA buffer, stream into a filterbank file, etc. Once the detection pipeline has sent a trigger, *T0* writes a self-describing NetCDF binary file of voltage data to disk where the candidate can be processed offline.

Technosignature Searches

GReX units can also be used to carry out technosignature surveys. GReX's large field of view enables the constraining of upper limits on the prevalence of intelligent life in the local galaxy. Presently, two technosignature pipelines can be deployed. A high-spectral-resolution technosignature detection pipeline can be deployed to search for drifting narrowband technosignatures. In the case of GReX, coincidence rejection from two or more units can be used to mitigate non-terrestrial narrowband signals [26]. Additionally, artificially dispersed signals can be searched for, as described in [27], by deploying SPANDAK[13] on GReX. The minimum power of a transmitter that can be detected is given in terms of effective isotropic radiated power (EIRP), for a narrowband signal, this is expressed as

$$\mathrm{EIRP}_{\mathrm{min,narrow}}(f, l, b) = \sigma \cdot 4\pi d_\star^2 \frac{2k_b T_{\mathrm{sys}}(l, b)}{A_e(f)} \frac{1}{\sqrt{n_p t_{\mathrm{obs}} \delta\nu}}. \tag{5.1}$$

Here, σ is the required signal-to-noise ratio (SNR), $\delta\nu$ is the bandwidth of the received signal, t_{obs} is the observing integration time, $A_e(f)$ effective area of the telescope as function of frequency, $T_{sys}(l, b)$ system temperature as a function of sky position, n_p is the number of polarizations, and $d_\star$ is the distance between the transmitter and receiver, i.e., the distance to the star. A value of 1 Hz is assumed for $\delta\nu_t$. For narrowband signals considered in Doppler searches. Similarly, an artificially dispersed signal can be expressed as

$$\mathrm{EIRP}_{\mathrm{min,disp}}(f, l, b) = \sigma \cdot 4\pi d_\star^2 \frac{2k_b T_{\mathrm{sys}}(l, b)}{A_e(f)} \frac{1}{\sqrt{n_p\ \delta\nu\ \tau}}. \tag{5.2}$$

[13] `https://github.com/gajjarv/PulsarSearch/`

However, in this case τ represents the pulse duration of the dispersed burst where 1 ms is assumed. A technosignature emanated from Alpha Centauri (1.34 pc) would require 1.2×10^{12} W/Hz and 4.9×10^{12} W/Hz for a dispersed burst and a narrowband signal at an SNR of 10. Here the maximum effective area of GReX is used along with a T_{sys} of 50 K. The wide FoV of GReX places it as a useful tool in characterization of potential technosignatures enabling meaningful constraints in the observing band.

5.4 Deployment

Figure 5.8: Deployed GReX terminals from left to right: Hat Creek Radio Observatory (USA), Cornell University Space Sciences Building (USA), Owens Valley Radio Observatory (USA), and Rosse Observatory (Ireland), and the Smithsonian Astrophysical Observatory at Harvard University (USA).

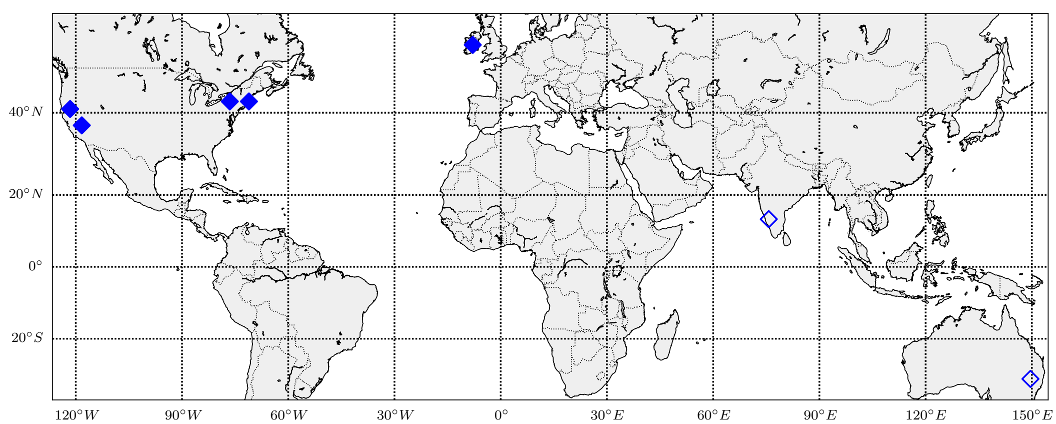


Figure 5.9: Mercator map with deployed and on-sky GReX stations (♦) along with stations at various stages of deployment (◊).

At the time of writing, five GReX units are currently operating on-sky (Figure 5.8). Their locations are shown in Figure 5.9, with two additional units planned for deployment at the Parkes Observatory in New South Wales, Australia and Gauribidanur, India. The first operational unit was deployed at Owens Valley, followed by Cornell, Hat Creek, Birr, and Harvard. Details of each deployment, including installation processes and site-specific considerations, are outlined below.

Owens Valley Radio Observatory

The deployment at the Owen's Valley Radio Observatory took the place of the previous STARE2 system in May 2024. We made use of the same location and fiber optic connections. Installing at the site involved the final assembly of the box, termination of fibers for the Ethernet connection, and testing. Once the box was powered, we performed Y factor measurements to ensure we achieved our desired sensitivity.

Cornell

The installment of the GReX terminal at Cornell University followed a three-stage plan: site testing and preparation, debugging GReX hardware and software, and site installation with first-light measurements. Full deployment of the system was complete with the start of burst injections in early July 2024. The terminal has remained on-sky at this site since this initial deployment.

Prior to receiving any hardware or software, we began a site-searching campaign to identify an optimal site to host the GReX terminal. This site needs to meet four primary requirements: (1) easy access to electricity and internet, (2) a weatherproof and secure location to host the GReX server, (3) an unimpeded view of the sky from the terminal, and (4) a minimal level of radio frequency interference (RFI) within the observing band. To facilitate this search, we developed a portable device for recording the RFI environment of a potential site. This device includes (i) a chargeable battery-pack for power, (ii) a Siglent Spectrum Analyzer for signal analysis, (iii) and a signal path consisting of a bias-T with 12-V power supply and a 1320-1580 MHz band pass filter[14] all of which is contained within a (iv) portable rack case. For consistency, we used the LNAs and cake pan antenna from the GReX system, with the initial goal of identifying any extreme continuous RFI that would immediately disqualify a potential site. We then ran follow-up measurements spanning at least one day of collecting time at sites with acceptable RFI levels to search for the presence of any intermittent RFI that was missed during the first survey. Following this second batch of measurements, we computed Y-factors and equivalent system temperatures (see Section 5.6 for details) for each of the remaining potential

[14]Mini-Circuits ZX75BP-1450-S+

host sites. All sites showed roughly equivalent baseline system temperatures across the observing band. The Space Sciences Building (SSB) roof on Cornell's campus is the lowest system temperature site of those surveyed that meets all four primary site requirements. This, combined with the ease of access from having the GReX terminal located in the middle of campus, has led to us selecting the roof of SSB as the permanent Cornell GReX terminal location.

The second stage of deployment is characterized by the arrival of hardware (the server and GReX terminal) on site and subsequent software setup and bug-fixes. We received the servers and terminal box at Cornell at the end of June 2023. As the second deployed GReX terminal, we spent a few months working closely with team-members at Caltech on hardware and software bug fixes. This effort helped develop the comprehensive GReX software guide[15] for "ground up" installation, from connecting the server to the Raspberry Pi and SNAP board in the terminal box to running the full GReX real-time data capture pipeline.

With this bug-fixing period complete, we moved to setting up the terminal box on our selected site Figure 5.8 (Center Left) and taking first-light data collection. This included an initial collection of ADC RMS values and observation of the spectra collected on site. To prevent overflowing or underflowing of the ADCs, we tuned the adjustable gain in the terminal FEM. Two easy astrophysical signals to detect on first light are the presence of the solar continuum (seen as an overall increase in the spectrum as the sun enters the beam) and the HI emission line from Galactic hydrogen centered near 1420.4 MHz. The absence of either of these signals indicates that the terminal is not properly observing the sky and requires further fine-tuning, usually either due to an inappropriate FEM gain or spectral leakage within the LNAs. Once these checks are passed, we move on to carry out tests of the terminal sensitivity (Section 5.6) which will give us estimates of the receiver gain, the system temperature, and the resulting system equivalent flux density (SEFD).

Hat Creek Radio Observatory

The GReX deployment at HCRO Figure 5.8 (Left) was completed in June 2024. Before installation, the box was first tested in a screened room using a spectrum analyzer connected to an omnidirectional antenna and external amplifier. Measurements were conducted with the GReX unit turned off for baseline measurements and then powered on with the box both open and closed with RF absorbers placed inside the

[15] `https://grex-telescope.github.io`

enclosure. Steel wool was added to the fiber cable opening to enhance shielding. These measurements (discussed in Section 5.4) show that the self-generated RFI is stronger than expected, and necessitated installation at a more remote location on-site to avoid interference with other experiments.

Rosse Observatory

The deployment of the GReX unit in Ireland was completed in December 2024 at the radio observatory of Trinity College Dublin. The Rosse Observatory is situated in Birr, in the relatively low population density Irish Midlands. The site already hosts the Irish LOw Frequency Array (LOFAR) station [28], [29] and other smaller experiments. As the GReX deployment was in close proximity to the LOFAR station, any unintended emission from the GReX unit in the LOFAR band was measured in the lab, prior to deployment, see Section 5.4. Due to the harsher weather conditions in Ireland, the GReX unit underwent further waterproofing with extra sealant on the box where all components are housed (see Figure 5.3). The most notable modification is the use of a radome housing, as it is radio transparent and often used on marine vessels to protect equipment from harsh outdoor elements Figure 5.8 (Right).

Self-Generated RFI

As GReX is intended to be hosted at radio observatories as well as universities, tests of the self-generated RFI are critical to maintain spectrum purity at sensitive sites. Lab tests of the box's emission were performed at HCRO and Birr, but not at OVRO due to lack of resources. Additionally, Stokes I data were taken using I-LOFAR at Birr while GReX was operational. The lab tests of the closed box's emission are shown in Figure 5.10. The results from the tests at HCRO in the 300 – 6000 MHz band indicate that GReX in its current form is not yet suitable for installation close to sensitive telescopes observing in this band. We believe the RFI retrofitting of the commercial enclosure used to house the electronics is primarily at fault. Further experiments with other RFI gaskets and conductive tape show a marked improvement in radiated power, but not totally acceptable for sensitive sites, especially at L-band. However, Stokes I data from I-LOFAR (shown in Figure 5.11) show no appreciable contamination. Additionally, the various ongoing experiments at OVRO such as the DSA-110 and LWA have also not noticed any increase in local RFI.

While some mitigations have substantially improved the unit's radiated power, the box remains a moderate RFI source. The self-generated RFI has not proven to be an issue for the FRB detection pipeline, nor has it interfered with experiments at operational radio observatories. However, the emission remains an open issue and prevents installation at some sites. Further work is needed to identify problematic components and improve shielding.

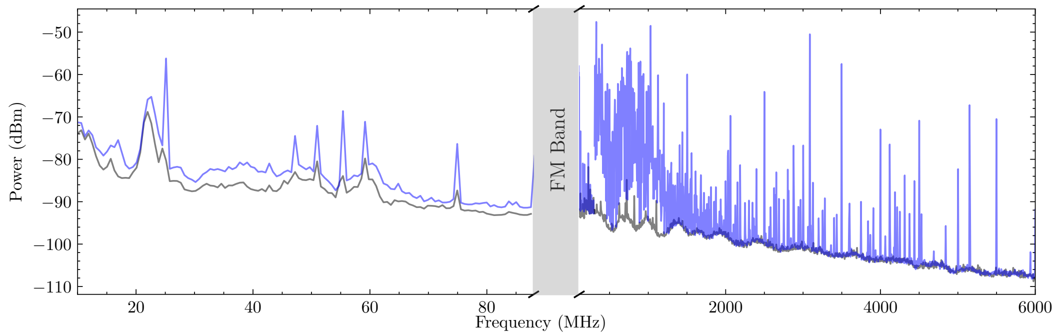


Figure 5.10: Measured power spectrum from 10 – 6000 MHz of a closed GReX box. The blue trace shows the spectrum when the unit is powered on, while the gray trace represents the baseline measurement. The frequency range of 10 – 300 MHz was measured at Birr in a lab and 300 – 6000 MHz was measured at HCRO in a screened room.

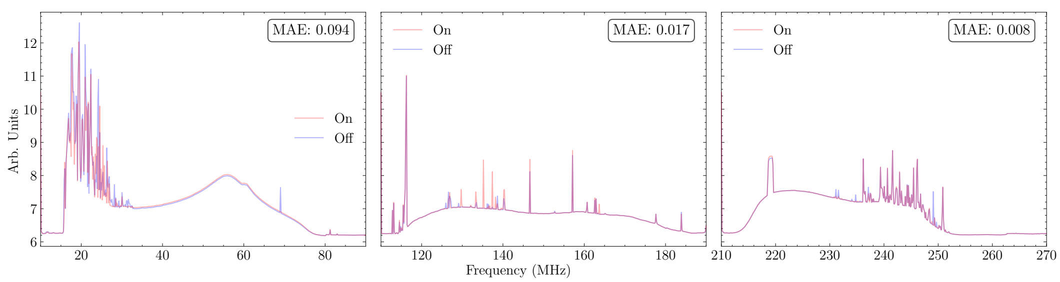


Figure 5.11: 15-minute Stokes I spectra measured using I-LOFAR with the GReX unit powered on and off, covering all observing modes from 10 – 270 MHz. The mean absolute error (MAE) was calculated for each measurement, comparing the powered-on spectrum to the baseline (powered-off) measurement.

5.5 Beam Modelling with Measurements of Galactic HI

The Hydrogen I (HI) emission line is an omnipresent and easily identifiable time-varying signal within the GReX observing band, making it an ideal metric for ensuring that a GReX terminal is observing the sky properly. Captured spectra in the *Prometheus* database are accessed by Grafana to display estimates of the HI SNR in real time at the top of all Grafana dashboard monitors. While the HI-line intensity provides a useful sanity check that a terminal is properly observing the sky, its usefulness does not end there. In this section, we present the results of our method for using the measured HI-line intensity at the OVRO and Cornell terminals to estimate the FWHM (and corresponding solid angle) of the GReX antenna beam-response.

Following the prescription detailed in Section C.2, we use the Leiden/Argentine/Bohn (LAB) all-sky survey of galactic neutral hydrogen [30] to simulate the expected antenna temperature as seen by the terminal over one sidereal day (sampled every 10 minutes) for fifty different beams with θ_{FWHM}s ranging from 50° to 100°. We query HI-line observations from the Cornell and OVRO *Prometheus* databases with time steps of ten minutes over a month. We only keep days where the data is continuous across the day to consistently align the data and simulations according to the local sidereal time (LST) before normalizing both such that the average temperature over one day is a constant between all simulations and observations. For each time sample across the sidereal day, we calculate the observed HI-line mean intensity and standard deviation. Any samples that exceed a 3σ deviation from the mean are considered to be impacted by RFI and are ignored when calculating the root-mean squared (RMS) residuals between observations and models. Any days with more than 5% of samples above this threshold are fully ignored. After RFI mitigation, there are 23 and 11 days remaining from the Cornell and OVRO measurements, respectively. With the data aligned, normalized, and cleaned of RFI, we calculate the RMS error between each day and each model. Figure 5.12 presents the aligned and normalized observed and simulated HI-line (left) alongside the RMS errors of each day against all modeled beam-shapes (right). While all RMS curves display consistent profiles which have minima that are clustered around the average FWHM, there are some curves in Figure 5.12 **(d)** that minimize at a higher RMS than is typical. We attribute this to the presence of RFI-affected samples that did not get removed by our thresholding process. The modeled θ_{FWHM}s corresponding to the RMS minimum for each day are taken as the population of best-fit parameters. We determine the general best fit model as the mean of this population, with a 1σ uncertainty in the FWHM given by the population standard deviation. Using this method, we determine that the Cornell terminal has a beam response at 1420.4 MHz that is well-characterized as a 2D Gaussian with $\theta^{\text{CU}}_{\text{FWHM}}(1420.4\,\text{MHz}) = 67.44° \pm 0.99°$. The OVRO terminal is similarly described by a $\theta^{\text{OVRO}}_{\text{FWHM}}(1420.4\,\text{MHz}) = 64.29° \pm 0.62°$.

Since all GReX terminals are outfitted with identical feeds, we consider the beam response of a typical feed to be a good description for all feeds. This typical beam

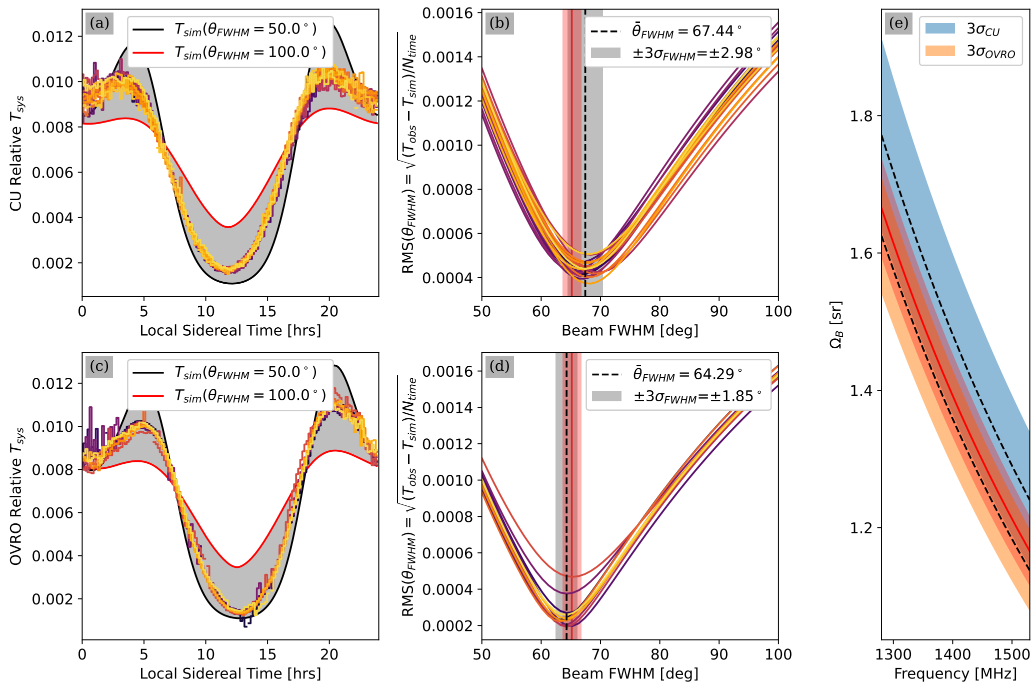


Figure 5.12: Observed data and corresponding simulations of the HI-line by the Cornell University ((**a**) and (**b**)) and OVRO ((**c**) and (**d**)) GReX terminals are used to fit the beam FWHM at 1420.4MHz. Simulations are rendered following the prescription in Section C.2. The observed (colored step-function lines) and simulated data (gray shaded region bounded by red and black curves) are aligned by local sidereal time and samples overly affected by RFI are excluded when calculating the RMS error. The population of FWHM values that minimize the RMS errors for each terminal and each day are then used to determine the mean (black dashed line, (**b**) and (**d**)) and standard deviation of each terminal FWHM. The characteristic FWHM of a GReX terminal (Equation 5.3) is presented in (**b**) and (**d**) as a solid red line with 3σ shading. We then scale the solid angle across the observing band according to the relationship $\Omega_B \propto f^{-2}$ in (**e**).

response at 1420.4 MHz is then characterized as

$$\theta_{\rm FWHM} = \frac{\frac{\theta_{\rm FWHM}^{\rm OVRO}}{\sigma_{\rm OVRO}^2} + \frac{\theta_{\rm FWHM}^{\rm CU}}{\sigma_{\rm CU}^2}}{\frac{1}{\sigma_{\rm OVRO}^2} + \frac{1}{\sigma_{\rm CU}^2}} \pm \sqrt{\frac{1}{\frac{1}{\sigma_{\rm OVRO}^2} + \frac{1}{\sigma_{\rm CU}^2}}} = 65.18^\circ \pm 0.53 \tag{5.3}$$

with corresponding solid angle at that frequency of

$$\Omega_B = \int_\Omega d\Omega \exp\left[-4\ln 2\left(\frac{\theta}{\theta_{\rm FWHM}}\right)^2\right] = 1.36\ \pm 0.02\ {\rm sr}. \tag{5.4}$$

We use $\Omega_B \propto f^{-2}$ to determine the solid angle as a function of frequency,

$$\Omega_B(f) = \frac{1931.7 \pm 28.4}{f^2}\,\mathrm{sr\,MHz}^2. \tag{5.5}$$

This functional relationship (and equivalent comparisons for the individual terminal fits) is presented in Figure 5.12 (**e**). We determine from this relation that the beam solid angle near our central frequency is $\Omega_B(1400\,\mathrm{MHz}) = 1.40 \pm 0.02\,\mathrm{sr}$. We also report the maximum effective area A_0 of the receiver across the band as $A_0 = 328 \pm 5\,\mathrm{cm}^2$ according to Equation C.6.

5.6 Sensitivity

Understanding the sensitivity of the GReX instrument to incoming signals is necessary for defining the strength of those signals and being able to place them in an astrophysical context.

The receiver gains and the system temperatures for the Cornell, OVRO, and Harvard stations are shown in Figure 5.13, computed using a Y-factor test (Equation C.12). Sky measurements are taken before daybreak or after sundown to ensure emission from the sun does not enter the beam. This allows us to assume a constant sky temperature of 5.5 K, with 2.7 K from the CMB, 1.9 K from atmospheric effects, and 0.9 K from galactic effects [31]. We placed a slab of ambient-temperature (nominally 290 K ± 5 K) radio-absorbent foam over the terminal antenna while taking hot data for the Y-factor test. Data were saved in the Stokes I filterbank format for the hot foam and cold sky states.

The system noise temperature for a GReX terminal is approximately 35 K, averaged across the band. Obstructions in the beam such as buildings and trees will worsen the noise, as well as the presence of strong RFI desensitizing the receiver. As such, optimal placement and configuration of attenuation/gains is critical to maximizing performance.

The second essential measure of system sensitivity is the forward gain (g_f) of the receiver. This quantity depends primarily on the integrated beam response function across the sky (Equation C.7, Equation C.6), since that is the value mediating between the true brightness temperature integrated across the telescope beam and the nominal effective temperature representing the observed temperature (Equation C.4, Equation C.5). Applying the effective area of the terminal found in Section 5.5 to Equation C.7, we compute the forward gain as $g_f = 84.2\,\mathrm{kJy\,K}^{-1} \pm 1.3\,\mathrm{kJy\,K}^{-1}$. Multiplying this value by the system temperature gives an SEFD of approximately

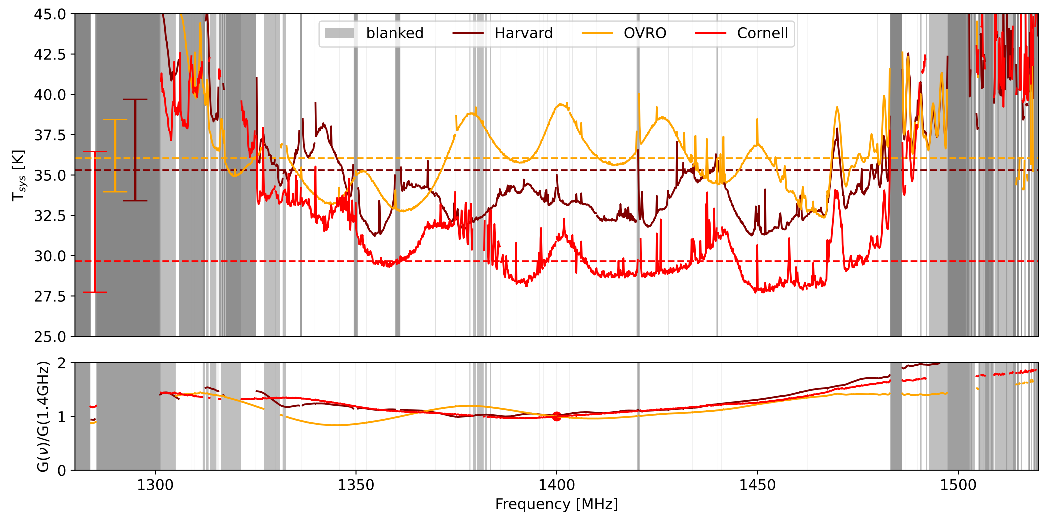


Figure 5.13: System temperature and relative receiver gain measured for the GReX terminals at Cornell, OVRO, and Harvard. Dashed lines show the median system temperature with associated 16^{th} and 84^{th} quantile error bars.

3 MJy. When determining our detection threshold, we consider a proto-typical burst that fills the band and lasts 1 ms. Our detectors are sensitive to an effective bandwidth of 188 MHz (75% of the total bandwidth) and the use of matched-filtering with a set of boxcars in the detection pipeline allows our integration time to match the 1 ms burst width. By averaging over these samples in frequency, time, and both polarizations, we find that for an event to attain an SNR of 10 (chosen as a fiducial value with a reasonable empirically-determined false-positive rate), it must have a corresponding flux density of at least ~ 50 kJy.

This is a minimum requirement, as the angular deviation of a source away from the center of the detector beam at the time of observation increases the required flux density necessary to detect the burst.

5.7 Upper Limit of Bright FRBs

We constrain the upper limit on FRBs at least as bright as our detection threshold by estimating the number of detections over some time as a Poissonian process given the single detection by STARE2 and, at present, zero detections by GReX. For this analysis, we will only consider the time on sky at the OVRO site. The Poissonian process has a probability distribution of

$$p(k|\lambda, \tau) = \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!}, \tag{5.6}$$

for k events given the event rate λ and observation time τ. To estimate the event rate λ with $k = 1$ after time on sky of τ_1, we solve for $p(\lambda|k = 1, \tau_1)$ via Bayes' theorem assuming a flat prior on λ where $p(\lambda) \propto 1$ via

$$p(\lambda|k = 1, \tau_1) = \frac{p(k = 1|\lambda, \tau_1)p(\lambda)}{p(k = 1)} = \frac{\lambda\tau_1 e^{-\lambda\tau_1}}{\int_0^\infty d\lambda\lambda\tau_1 e^{-\lambda\tau_1}} = \tau_1^2\lambda e^{-\lambda\tau_1}. \tag{5.7}$$

This result is a gamma distribution with a scale of 2 and a rate parameter of τ_1.

Given our center-frequency beam solid angle of ~ 1.5 sr from Equation 5.4, at any given time the station is sensitive to only 12% of the sky. At the time of the STARE2 detection of FRB 200428, it had been observing for 448 days. Assuming a 75% observational duty-cycle, we compute $\tau_1 = 1.23\,\mathrm{yr} \times 0.12\,\mathrm{sky} \times 0.75 = 0.11\,\mathrm{sky\ yr}$. STARE2 continued to observe after FRB 200428 until it was superseded by GReX in May 2024. In total, a station at OVRO has now been observing for 6.3 years. We then compute a new rate parameter of $\tau_1 = 6.3\,\mathrm{yr} \times 0.12\,\mathrm{sky} \times 0.75 = 0.567\,\mathrm{sky\ yr}$.

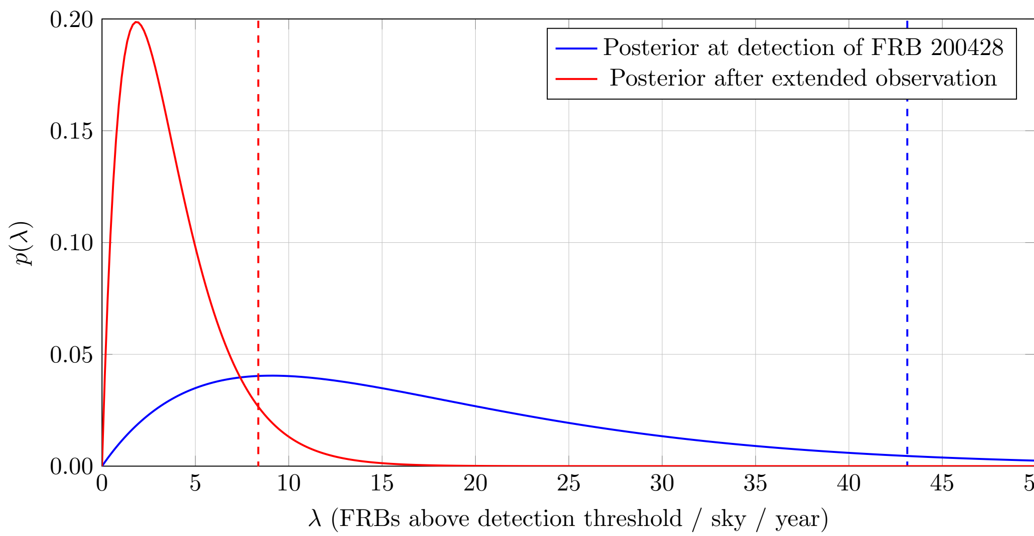


Figure 5.14: Posterior distribution of FRB event rate λ with associated dashed 95% confidence upper limits.

Shown in Figure 5.14 are the probability distribution functions for the event rate λ, implied by the observation time at the detection of FRB 200428 and the current total observation time. The STARE2 detection placed an upper limit on the rate parameter of $43.13\,\mathrm{sky}^{-1}\,\mathrm{yr}^{-1}$, with wide uncertainty due to the limited time on sky. Adding the subsequent non-detection time significantly reduces this estimate to $8.37\,\mathrm{sky}^{-1}\,\mathrm{yr}^{-1}$. A more thorough analysis of the rate including overlapping sky and time, luminosity, etc. will follow in a subsequent paper.

5.8 Conclusions and Outlook

We have described the Galactic Radio Explorer (GReX), a network of low-cost 1.4 GHz radio antennas searching for bright transients between 32.768 μs and 1.024 ms. Terminals have been deployed at multiple sites around the world, though most are currently in the US. Since the discovery of FRB 200428, GReX and an upgraded STARE2 have more than quadrupled the total square degree hours devoted to ultra-bright FRB searching. The improved system temperature of GReX LNAs over the original STARE2 hardware results in ~ twice the sensitivity, resulting in a deeper search. We also search for extremely narrow bursts $\ll$ 1 ms, which is not possible at lower frequencies or in systems with a large number of beams. With no detections, we update the all-sky rate of FRBs above ~ 50 kJy to be no larger than 8.37 sky^{-1} yr^{-1} at 95% confidence. This indicates that FRB-like emission from Galactic sources is rare, and FRB 200428 was even more unusual than initially thought.

The GReX boxes have proven to be replicable and stable. For example, the OVRO GReX terminal did not require on-site human intervention for more than a year of observation. Five sites are on-sky and continuously searching for bright FRBs. These locations are OVRO in California, the Cornell GReX in Ithaca, New York, Harvard GReX in Cambridge Massachusetts, Hat Creek GReX in California, and an Ireland station in Birr. A box has been shipped and tested in New South Wales, though not yet deployed. In the future, we hope to expand to multiple sites in the Southern Hemisphere to improve exposure to the Galactic plane, where most magnetars reside.

References

[1] J. M. Cordes and S. Chatterjee, "Fast Radio Bursts: An Extragalactic Enigma," *Annual Review of Astronomy and Astrophysics*, vol. 57, no. 1, pp. 417–465, Aug. 18, 2019. DOI: 10.1146/annurev-astro-091918-104501.

[2] E. Petroff, J. W. T. Hessels, and D. R. Lorimer, "Fast radio bursts," *The Astronomy and Astrophysics Review*, vol. 27, no. 1, p. 4, Dec. 2019. DOI: 10.1007/s00159-019-0116-6.

[3] The CHIME/FRB Collaboration, "A bright millisecond-duration radio burst from a Galactic magnetar," en, *Nature*, vol. 587, no. 7832, pp. 54–58, Nov. 2020, Publisher: Nature Publishing Group. DOI: 10.1038/s41586-020-2863-y.

[4] C. D. Bochenek, D. L. McKenna, K. V. Belov, *et al.*, "STARE2: Detecting Fast Radio Bursts in the Milky Way," *Publications of the Astronomical Society*

of the Pacific, vol. 132, no. 1009, p. 034 202, Jan. 2020. DOI: 10.1088/1538-3873/ab63b3.

[5] F. Kirsten, O. S. Ould-Boukattine, W. Herrmann, *et al.*, "A link between repeating and non-repeating fast radio bursts through their energy distributions," en, *Nature Astronomy*, vol. 8, no. 3, pp. 337–346, Mar. 2024, Publisher: Nature Publishing Group. DOI: 10.1038/s41550-023-02153-z.

[6] S. Mereghetti, V. Savchenko, C. Ferrigno, *et al.*, "INTEGRAL Discovery of a Burst with Associated Radio Emission from the Magnetar SGR 1935+2154," *The Astrophysical Journal Letters*, vol. 898, no. 2, p. L29, Aug. 1, 2020. DOI: 10.3847/2041-8213/aba2cf.

[7] F. Kirsten, B. Marcote, K. Nimmo, *et al.*, "A repeating fast radio burst source in a globular cluster," en, *Nature*, vol. 602, no. 7898, pp. 585–589, Feb. 2022. DOI: 10.1038/s41586-021-04354-w.

[8] T. Eftekhari, Y. Dong, W. Fong, *et al.*, "The Massive and Quiescent Elliptical Host Galaxy of the Repeating Fast Radio Burst FRB 20240209A," *The Astrophysical Journal Letters*, vol. 979, no. 2, p. L22, Feb. 1, 2025. DOI: 10.3847/2041-8213/ad9de2.

[9] V. Shah, K. Shin, C. Leung, *et al.*, "A Repeating Fast Radio Burst Source in the Outskirts of a Quiescent Galaxy," *The Astrophysical Journal Letters*, vol. 979, no. 2, p. L21, Feb. 1, 2025. DOI: 10.3847/2041-8213/ad9ddc.

[10] K. Sharma, V. Ravi, L. Connor, *et al.*, "Preferential occurrence of fast radio bursts in massive star-forming galaxies," *Nature*, vol. 635, no. 8037, pp. 61–66, Nov. 7, 2024. DOI: 10.1038/s41586-024-08074-9.

[11] N. Hurley-Walker, X. Zhang, A. Bahramian, *et al.*, "A radio transient with unusually slow periodic emission," *Nature*, vol. 601, no. 7894, pp. 526–530, Jan. 27, 2022. DOI: 10.1038/s41586-021-04272-x.

[12] M. Caleb, E. Lenc, D. L. Kaplan, *et al.*, "An emission-state-switching radio transient with a 54-minute period," *Nature Astronomy*, vol. 8, no. 9, pp. 1159–1168, Jun. 5, 2024. DOI: 10.1038/s41550-024-02277-w.

[13] A. C. Rodriguez, "Spectroscopic detection of a 2.9-hour orbit in a long-period radio transient," *Astronomy & Astrophysics*, vol. 695, p. L8, Mar. 2025. DOI: 10.1051/0004-6361/202553684.

[14] L. Connor, K. A. Shila, S. R. Kulkarni, *et al.*, "Galactic Radio Explorer: An All-sky Monitor for Bright Radio Bursts," *Publications of the Astronomical Society of the Pacific*, vol. 133, no. 1025, p. 075 001, Jul. 1, 2021. DOI: 10/gmb99r.

[15] H.-H. Lin, K.-y. Lin, C.-T. Li, *et al.*, "BURSTT: Bustling Universe Radio Survey Telescope in Taiwan," *Publications of the Astronomical Society of the Pacific*, vol. 134, no. 1039, p. 094 106, Sep. 1, 2022. DOI: 10.1088/1538-3873/ac8f71.

[16] S. Weinreb and J. Shi, "Low Noise Amplifier With 7-K Noise at 1.4 GHz and 25 °C," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 4, pp. 2345–2351, Apr. 2021. DOI: 10.1109/TMTT.2021.3061459.

[17] V. Ravi and D.-1. Collaboration, "The DSA-110: Overview and first results," *Bulletin of the AAS*, vol. 55, no. 2, Feb. 2023.

[18] J. Eriksson, F. Häggström, S. Aittamaa, A. Kruglyak, and P. Lindgren, "Real-time for the masses, step 1: Programming API and static priority SRP kernel primitives," in *2013 8th IEEE International Symposium on Industrial Embedded Systems (SIES)*, Jun. 2013, pp. 110–113. DOI: 10.1109/SIES.2013.6601482.

[19] J. Hickish, Z. Abdurashidova, Z. Ali, *et al.*, "A Decade of Developing Radio-Astronomy Instrumentation using CASPER Open-Source Technology," *Journal of Astronomical Instrumentation*, vol. 05, no. 04, p. 1 641 001, Dec. 2016. DOI: 10.1142/S2251171716410014.

[20] A. R. Thompson, J. M. Moran, and G. W. Swenson, "Digital signal processing," in *Interferometry and Synthesis in Radio Astronomy*. Cham: Springer International Publishing, 2017, pp. 309–390. DOI: 10.1007/978-3-319-44431-4_8.

[21] W. van Straten, A. Jameson, and S. Osłowski, "PSRDADA: Distributed Acquisition and Data Analysis for Radio Astronomy," *Astrophysics Source Code Library*, ascl:2110.003, Oct. 1, 2021.

[22] M. Rafiei-Ravandi and K. M. Smith, "Mitigating Radio Frequency Interference in CHIME/FRB Real-time Intensity Data," *The Astrophysical Journal Supplement Series*, vol. 265, no. 2, p. 62, Apr. 2023. DOI: 10.3847/1538-4365/acc252.

[23] B. R. Barsdell, M. Bailes, D. G. Barnes, and C. J. Fluke, "Accelerating incoherent dedispersion," *Monthly Notices of the Royal Astronomical Society*, vol. 422, no. 1, pp. 379–392, May 1, 2012. DOI: 10.1111/j.1365-2966.2012.20622.x.

[24] L. McInnes, J. Healy, and S. Astels, "Hdbscan: Hierarchical density based clustering," *Journal of Open Source Software*, vol. 2, no. 11, p. 205, 2017. DOI: 10.21105/joss.00205.

[25] O. Tange, *Gnu parallel 20250222 ('grete tange')*, Feb. 2025. DOI: 10.5281/zenodo.14911163.

[26] O. A. Johnson, V. Gajjar, E. F. Keane, *et al.*, "A Simultaneous Dual-site Technosignature Search Using International LOFAR Stations," en, *The Astronomical Journal*, vol. 166, no. 5, p. 193, Oct. 2023, Publisher: The American Astronomical Society. DOI: 10.3847/1538-3881/acf9f5.

[27] V. Gajjar, D. LeDuc, J. Chen, *et al.*, "Searching for Broadband Pulsed Beacons from 1883 Stars Using Neural Networks," *The Astrophysical Journal*, vol. 932, p. 81, Jun. 2022, Publisher: IOP ADS Bibcode: 2022ApJ...932...81G. DOI: 10.3847/1538-4357/ac6dd5.

[28] M. P. Van Haarlem, M. W. Wise, A. W. Gunst, *et al.*, "LOFAR: The LOw-Frequency ARray," *Astronomy & Astrophysics*, vol. 556, A2, Aug. 2013. DOI: 10.1051/0004-6361/201220873.

[29] P. C. Murphy, P. Callanan, J. McCauley, *et al.*, "First results from the REAL-time Transient Acquisition backend (REALTA) at the Irish LOFAR station," *Astronomy & Astrophysics*, vol. 655, A16, Nov. 2021. DOI: 10.1051/0004-6361/202140415.

[30] P. M. W. Kalberla, W. B. Burton, D. Hartmann, *et al.*, "The Leiden/Argentine/Bonn (LAB) Survey of Galactic HI: Final data release of the combined LDS and IAR surveys withimproved stray-radiation corrections," *Astronomy & Astrophysics*, vol. 440, no. 2, pp. 775–782, Sep. 2005. DOI: 10.1051/0004-6361:20041864.

[31] D. Le Vine and S. Abraham, "Galactic noise and passive microwave remote sensing from space at L-band," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 1, pp. 119–129, Jan. 2004. DOI: 10.1109/TGRS.2003.817977.

Appendix A

AUTOMATIC DIFFERENTIATION

At its core, AD is the recursive application of the chain rule:

$$\frac{d}{dx}\left[f\left(g\left(x\right)\right)\right] = f'\left(g\left(x\right)\right)g'\left(x\right)\ . \tag{A.1}$$

To apply the chain rule to computer programs, we must reinterpret the sequence of instructions that a program performs as function composition. For a complicated program with many steps, this is equivalent to nested composition $f(g(h(w(...(x)))))$. For deep nesting, it may be more useful to imagine the flow of values through the program as a computational graph. For the case of $f(g(x))$, there are three "primal" values: x, $g(x)$, and $f(g(x))$. Correspondingly, there are two "tangent" values, $g'(x)$ and $f'(g(x))$. This graph is shown in Figure A.1 with square nodes indicating start and stop values, circular nodes indicating computation, and edges showing the flow of values.

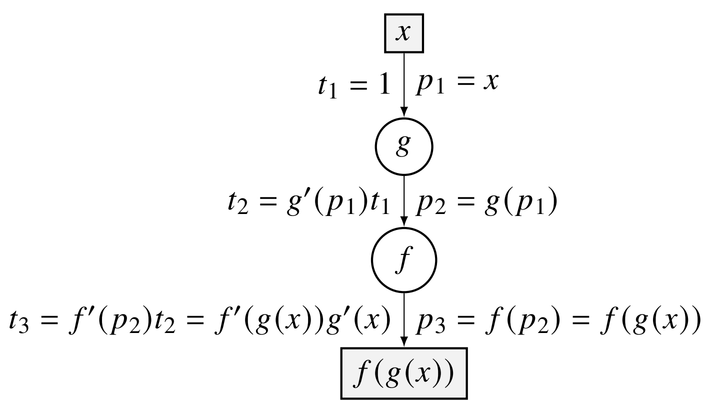


Figure A.1: Computational graph of forward-mode evaluation of $f(g(x))$.

The values flowing down the right side of the graph in Figure A.1 represent the normal evaluation of the program. The values on the left represent the accumulated tangents, with the initial value "seeded" as unity. Each primal computation relies on the values that immediately precede it. Similarly, the tangent components rely on the previous tangent and primal component. At the end of this graph, we have computed both the primal result and the tangent. This accumulation of tangent values alongside the primal values to build the resulting derivative is known as forward-mode AD.

To extend forward-mode AD to multivariate functions, we set the tangent "seed" value of the independent variable of interest to unity, with the remaining variables' tangents set to zero. Consider this more complicated computer program:

Algorithm 1 Example differentiable program

$f(a, b)$:
  $x \leftarrow a^2$
  $y \leftarrow \sin b$
  return $x * y$

Suppose we want to find $\partial f/\partial b$. The graph of all operations, primals, and tangents to do so is shown in Figure A.2.

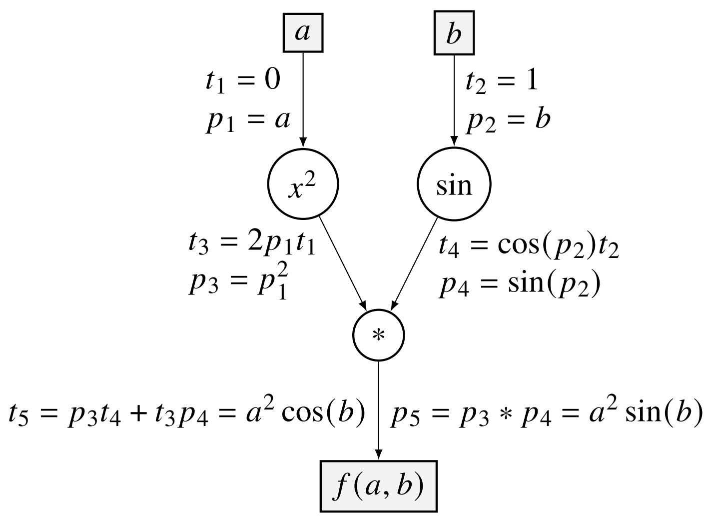


Figure A.2: Computational graph of forward-mode evaluation of $f(a, b)$, with the tangent seeded on b.

To compute $\partial f/\partial a$, we would set t_1 to unity and t_2 to zero and perform an additional sweep. This generalizes to an approach for evaluating gradients for functions of the form $f : \mathbb{R}^n \to \mathbb{R}^m$. Requiring one pass per independent variable implies that gradient evaluation in forward-mode has a complexity of $O(n)$ for an n-dimensional input. As such, forward-mode performs well for $f : \mathbb{R} \to \mathbb{R}^n$ but has no better complexity than finite-differencing for $f : \mathbb{R}^n \to \mathbb{R}$ (while maintaining machine-precision accuracy).

While forward-mode is simple and is straightforward to implement, the linear complexity poses a problem for high-dimensional functions such as large machine learning models or highly parameterized circuit optimizations. A solution is to change the manner in which we decompose the chain rule. Instead of computing the full derivative of each nested function recursively, we compute the derivate *with respect to* each nested function recursively and work backwards. In practice, this amounts to evaluating the primal trace of the function as before, but now collecting all primal values and their partial derivatives with respect to each parent value. Then, we seed the "adjoint" of the dependent variable (output) with unity and work backwards through the graph, applying the chain rule utilizing the precomputed partials. This is known as reverse-mode AD.

Looking again at our unary $f : \mathbb{R} \to \mathbb{R}$ function of $y = f(g(x))$, we draw the computational graph shown in Figure A.3.

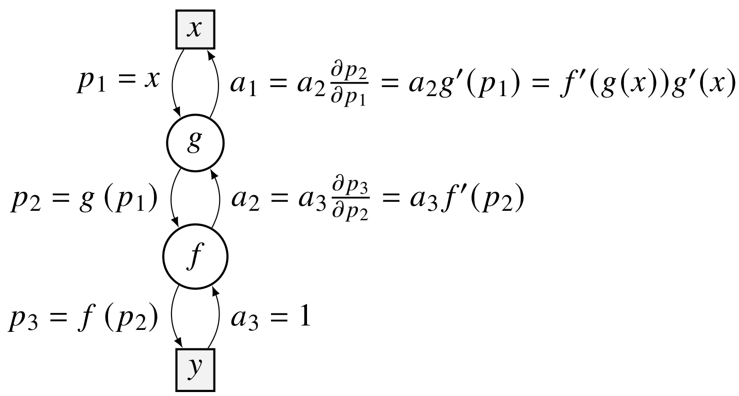


Figure A.3: Graph of reverse-mode evaluation of $f(g(x))$.

Here we again see the primal, normal program execution of the function on the left, and then the reverse pass accumulation of the adjoints on the right. The resulting adjoint, a_1, is identical to the resulting forward-mode tangent t_3 from Figure A.1. The difference is that in this case, all the intermediate values and the partial derivatives need to be stored, whereas in forward-mode, each step relies solely on the previous tangent and primal values in the graph.

The procedure becomes more complicated with less trivial, multi-arity functions. Consider the function example from Figure A.2, shown in Figure A.4 now in reverse-mode.

Again, in the forward sweep, we collect primal values as well as partial derivates. Then, in the reverse-mode, we collect adjoints down every branch by multiplying the previous adjoint by the appropriate precomputed partial derivates. This offers several advantages, the first of which is that in one pass, we have computed the full gradient. Additionally, we have done so with fewer operations than in the forward-mode. Often, this is a more efficient way of computing the gradient. However, the reverse

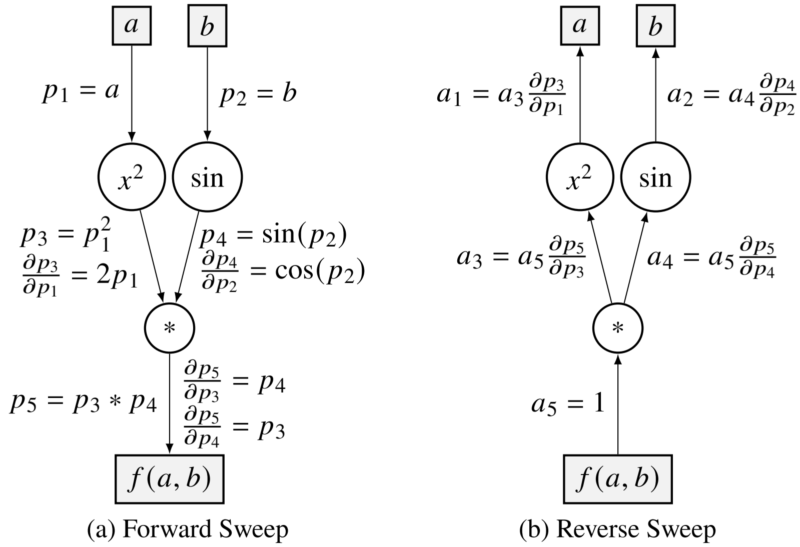


Figure A.4: Graph of reverse-mode evaluation of $f(a, b)$.

sweep requires us to build the computation graph in memory and store intermediate values and partials, while forward-mode only requires the previous tangents and primals. This is a classic example of a memory-time complexity tradeoff. In general, for functions of the form $f : \mathbb{R}^n \to \mathbb{R}^m$ where $m \gg n$, forward-mode is preferred. However, for general optimization problems, reverse-mode is typically higher performance, as it solves $f : \mathbb{R}^n \to \mathbb{R}$ for a scalar cost function. Benchmarking to choose the right approach is required when the input and output dimensions have similar orders of magnitude.

Efficient implementations of these algorithms have become commonplace due to their usage in training large machine learning models. For a more in-depth discussion on the field of AD and implementation details, see [1].

References

[1] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, “Automatic differentiation in machine learning: A survey,” *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 5595–5637, Jan. 1, 2017.

Appendix B

USEFUL RESULTS

B.1 Interferometer Survey Speed and Sensitivity

To derive the survey speed of a radio interferometer, we start with the radiometer equation,

$$\sigma_T = \frac{T_{\text{Sys}}}{\sqrt{\Delta\nu\tau}}. \tag{B.1}$$

For system noise temperature T_{Sys}, bandwidth $\Delta\nu$, and integration time τ, we compute the RMS error (in Kelvin). If we want to know the signal-to-noise ratio for an observation of a point source, we define the ratio of the source's induced antenna temperature T_a over this RMS noise level as

$$\text{SNR} = \frac{T_a}{\sigma_T} = \frac{T_a}{T_{\text{Sys}}}\sqrt{\Delta\nu\tau}. \tag{B.2}$$

Sources in the sky vary wildly in flux density, so to come up with a single metric, we can pick some arbitrary, minimum source flux density to set our SNR, S_ν. This point source flux density induces an antenna temperature via

$$T_a = \frac{S_\nu A_e}{2k_B}, \tag{B.3}$$

where A_e is the effective collecting area of the array and k_B is Boltzmann's constant. For reflector-based antennas, the effective aperture is equal to the reflector's physical area, scaled by some aperture efficiency, η_a, typically in the 0.7-range. We then multiply this by N antennas to compute the total effective collecting area for the entire array.

$$A_e = \frac{N\eta_a\pi D^2}{4}. \tag{B.4}$$

Now, we find the integration time required to achieve an SNR of unity as

$$\tau = \frac{1}{\Delta\nu}\left[\frac{8k_B T_{\text{Sys}}}{S_\nu N\eta_a\pi D^2}\right]^2. \tag{B.5}$$

As this is the time required to survey one portion of the sky, the last missing piece is how much of the sky we surveyed in this pointing. Again, for reflectors, we could use the full-width half-maximum in steradians, or the beam solid angle, Ω_B. For high-gain reflectors (as are used in radio astronomy), these are almost identical. For any antenna, the beam solid angle is defined as

$$\Omega_B = \frac{\lambda^2}{A'_e}\,, \tag{B.6}$$

where A'_e here is the effective aperture (of the primary beam, not the total array). So, we write

$$\Omega_B = \frac{4\lambda^2}{\pi\eta_a D^2}\,. \tag{B.7}$$

Finally, we can define survey speed as the time required to survey a given beam solid angle to some fixed minimum point source flux density as

$$\mathrm{SS\ (sr/s)} = \frac{\Omega_B}{\tau} = \eta_a\pi\Delta_\nu\left[\frac{\lambda S_\nu N D}{4k_B T_{\mathrm{Sys}}}\right]^2\,. \tag{B.8}$$

If we take S_ν to be in Jy (1×10^{-26} W/Hz/m^2) and convert to deg^2 per hour, we can approximate this as

$$\mathrm{SS\ (deg^2/hr)} \approx 1.217\eta_a\Delta_\nu\left[\frac{\lambda S_\nu N D}{T_{\mathrm{Sys}}}\right]^2\,. \tag{B.9}$$

The important implications of this relation are the quadratic terms in λ, N, D, and T_{Sys}, as these are all design parameters.

In addition to survey speed, we can recast Equation B.2 as a sensitivity metric. In other words, we find what flux density (S_ν) we can measure at a $1\sigma_T$-level in some fixed integration time τ via

$$S_\nu\ \mathrm{(W/m^2/Hz)} = \frac{8k_B T_{\mathrm{Sys}}}{\pi\eta_a D^2 N\sqrt{\Delta_\nu\tau}}\,. \tag{B.10}$$

We can again simply this if we assume τ is one hour, our answer is in μJy (1×10^{-32} W/Hz/m^2), and Δ_ν is in GHz, we can write

$$S_\nu\ (\mu\mathrm{Jy\ @\ 1\ hr}) \approx 1852\frac{T_{\mathrm{Sys}}}{\eta_a D^2 N\sqrt{\Delta_{\nu(\mathrm{GHz})}}}\,. \tag{B.11}$$

If we use some typical DSA-2000 numbers of $\Delta_\nu = 1.3$ GHz, $D = 5$ m, $N = 2\,000$, $\eta_a = 0.7$, $T_{\mathrm{Sys}} = 25$ K, at a center wavelength λ of 22 cm, we find

$$\mathrm{SS\ (deg^2/hr\ @\ 2\mu Jy)} = 31.3 \tag{B.12}$$

$$S_\nu\ (\mu\mathrm{Jy\ @\ 1\ Hr}) = 1.16\,. \tag{B.13}$$

B.2 Intermodulation Products

Given a voltage transfer function of some nonlinear RF component

$$v_{\text{Out}} = f(v_{\text{In}})\,, \tag{B.14}$$

we can find the Taylor (Maclaurin) series expansion of the nonlinear f to give us

$$v_{\text{Out}} = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} v_{\text{In}}^n = a_0 + a_1 v_{\text{In}} + a_2 v_{\text{In}}^2 + a_3 v_{\text{In}}^3 \ldots\,, \tag{B.15}$$

where the coefficients a_0, a_1, etc. are proportional to the n-th derivative of f. Consider an input v_{In} composed of a sum of two tones close in frequency,

$$v_{\text{In}} = V_0(\cos\omega_1 t + \cos\omega_2 t)\,. \tag{B.16}$$

Substituting into Equation B.15, we find

$$v_{\text{Out}} = a_0 + a_1 V_0(\cos\omega_1 t + \cos\omega_2 t) \tag{B.17}$$
$$+ a_2 V_0^2(\cos\omega_1 t + \cos\omega_2 t)^2 \tag{B.18}$$
$$+ a_3 V_0^3(\cos\omega_1 t + \cos\omega_2 t)^3 \ldots \tag{B.19}$$

Using trigonometric identities, we expand this to

$$\begin{aligned}
v_{\text{Out}} = a_0 &+ a_1 V_0(\cos\omega_1 t + \cos\omega_2 t) \\
&+ \frac{a_2 V_0^2}{2}(2\cos(\omega_1 + \omega_2)t + 2\cos(\omega_1 - \omega_2)t) \\
&+ \frac{a_2 V_0^2}{2}(\cos 2\omega_1 t + \cos 2\omega_2 t + 1) \\
&+ \frac{a_3 V_0^3}{4}(\cos 3\omega_1 t + \cos 3\omega_2 t + 9\cos\omega_1 t + 9\cos\omega_2 t) \\
&+ \frac{a_3 V_0^3}{4}(3\cos(\omega_1 - 2\omega_2)t + 3\cos(2\omega_1 - \omega_2)t) \\
&+ \frac{a_3 V_0^3}{4}(3\cos(\omega_1 + 2\omega_2)t + 3\cos(2\omega_1 + \omega_2)t) \ldots
\end{aligned} \tag{B.20}$$

Here, we see that increasing powers from the series expansion leads to sum and difference frequencies of the form $n\omega_1 \pm m\omega_2$ for positive integers m and n. As the input voltage in the two tones increases, so do these so called "intermodulation" products. If the input power is small, these products will be well beneath the noise floor. However, as the power increases, the output power of the fundamental ($\cos\omega_n t$ increases linearly while the power in the intermodulation powers grow with the input voltage squared for the second-order terms and with the cube for the third-order terms. If the amplifier did not saturate, eventually the power of these tones at the output would intersect. The hypothetical power at which these intermodulation tones intersect are the n-th order intercept points.

2nd Order Distortion

The total power delivered at either fundamental into a 1 Ω load is

$$P_{\omega_1} = \frac{1}{2}a_1^2 V_0^2\,, \tag{B.21}$$

recalling that these voltage phasors are defined with peak magnitudes, and so power is defined as $1/2Re\{VI^*\}$. The total power delivered at one of the 2nd order mixing terms ($\cos\omega_1 \pm \omega_2$) is

$$P_{\omega_1+\omega_2} = \frac{1}{2}a_2^2 V_0^4\,. \tag{B.22}$$

The definition of the intercept point is when the powers at the fundamental and the intermodulation product are equal, given by

$$\frac{1}{2}a_1^2 V_{IP2}^2 = \frac{1}{2}a_2^2 V_{IP2}^4\,. \tag{B.23}$$

Solving for V_{IP2} yields

$$V_{IP2} = \frac{a_1}{a_2}\,. \tag{B.24}$$

We define the output intercept point as the power at the fundamental which leads to this intersection

$$\text{OIP2} = P_{\omega_1}|_{V_0=V_{IP2}} = \frac{1}{2}a_1^2 V_{IP2}^2 = \frac{a_1^4}{2a_2^2}\,. \tag{B.25}$$

Finally, we can find that the power of the output intermodulation product can be rewritten in terms of the power of the fundamental and OIP2

$$P_{\omega_1+\omega_2} = \frac{1}{2}a_2^2 V_0^4 = \frac{\left(\frac{1}{2}a_1^2 V_0^2\right)^2}{a_1^4(2a_2^2)^{-1}} = \frac{P_{\omega_1}^2}{\text{OIP2}}\,, \tag{B.26}$$

or

$$\text{OIP2 }(W) = \frac{P_{\omega_1}^2}{P_{\omega_1+\omega_2}} \tag{B.27}$$

$$\text{OIP2 (dBm)} = 2P_{\omega_1} - P_{\omega_1+\omega_2}\,. \tag{B.28}$$

Cascaded Distortion

Given two components with power gains G_1 and G_2 with output nth-order output intercept points OIPn' and OIPn'', we want to find the effect of combining their distortions. As the tones are deterministic, the phase of the various intermodulation product impact the output. As such, we need to work out the cascade in voltages.

Using the result from Equation B.26, we can compute the output voltage of the 2nd order term as

$$V'_{\omega_1+\omega_2} = \sqrt{P'_{\omega_1+\omega_2} Z_0} = P'_{\omega_1}\sqrt{\frac{Z_0}{\mathrm{OIP2}'}}\,. \tag{B.29}$$

As the phase of these voltages is unknown, we can assume the worst-case of the distortion products adding in-phase. Therefore, the total 2nd-order distortion voltage at the output of the second component is

$$V''_{\omega_1+\omega_2} = P'_{\omega_1}\sqrt{\frac{G_2 Z_0}{\mathrm{OIP2}'}} + P''_{\omega_1}\sqrt{\frac{Z_0}{\mathrm{OIP2}''}}\,. \tag{B.30}$$

As $P''_{\omega_1} = G_2 P'_{\omega_1}$, we have

$$\begin{aligned} V''_{\omega_1+\omega_2} &= \left(\frac{\sqrt{G_2}}{G_2\sqrt{\mathrm{OIP2}'}} + \frac{1}{\sqrt{\mathrm{OIP2}''}}\right) P''_{\omega_1}\sqrt{Z_0} \\ &= \left(\frac{1}{\sqrt{G_2\mathrm{OIP2}'}} + \frac{1}{\sqrt{\mathrm{OIP2}''}}\right) P''_{\omega_1}\sqrt{Z_0}\,. \end{aligned} \tag{B.31}$$

The total output distortion power is

$$P''_{\omega_1+\omega_2} = \frac{\left(V''_{\omega_1+\omega_2}\right)^2}{Z_0} = \left(\frac{1}{\sqrt{G_2\mathrm{OIP2}'}} + \frac{1}{\sqrt{\mathrm{OIP2}''}}\right)^2 (P''_{\omega_1})^2 = \frac{(P''_{\omega_1})^2}{\mathrm{OIP2}}\,. \tag{B.32}$$

Therefore the 2nd order intercept of the cascaded system with coherent products is

$$\mathrm{OIP2} = \left(\frac{1}{\sqrt{G_2\mathrm{OIP2}'}} + \frac{1}{\sqrt{\mathrm{OIP2}''}}\right)^{-2}\,. \tag{B.33}$$

Although the worst case is useful for putting a bound on the distortion, in practical systems with many components, the phases between various stages may be randomly distributed. In that case, we treat the contributions as incoherent and directly sum the powers of the intermodulation products.

$$P''_{\omega_1+\omega_2} = G_2\frac{(P'_{\omega_1})^2}{\mathrm{OIP2}'} + \frac{(P''_{\omega_1})^2}{\mathrm{OIP2}''} = \left(\frac{1}{G_2\mathrm{OIP2}'} + \frac{1}{\mathrm{OIP2}''}\right)(P''_{\omega_1})^2\,. \tag{B.34}$$

Again solving for the total cascaded 2nd order intercept, incoherently combined gives

$$\mathrm{OIP2} = \left(\frac{1}{G_2\mathrm{OIP2}'} + \frac{1}{\mathrm{OIP2}''}\right)^{-1}\,. \tag{B.35}$$

The derivations for the 3rd order intercept can be found in [1].

References

[1] D. M. Pozar, *Microwave Engineering*, 4th ed. Hoboken, NJ: Wiley, 2012.

Appendix C

GREX ERRATA

C.1 Telescope Calibration

The factors of importance for determining the instrument sensitivity are three-fold: (1) the telescope response to an incoming signal as a function of angle, (2) the gain of the receiver, and (3) the system temperature of the instrument. Noise is added to an incoming signal during its propagation through the amplifiers and other electronics of the system. It is practical to begin by considering the specific intensity (I_ν) of a signal, since it is a conserved quantity through empty space. However, the quantity that describes the spectral power of the source received at the detector is the flux density,

$$S_\nu = \int_\Omega d\Omega \left[I_\nu(\theta, \phi) B(\theta) \cos\theta \right] . \tag{C.1}$$

$B(\theta)$ describes the response of the receiver to incoming radiation and is equivalent to the normalized effective area $A_e(\theta)/A_0$, where $A_0 = A_e(\theta = 0) = A_{e,max}$. In this notation, $\theta = 0$ points along the normal vector of the telescope receiver and ϕ describes the azimuthal angle about that normal vector. The $\cos\theta$ term in the integral describes the effect of projecting the incoming spectral power across the detector at different inclination angles. The beam size of a telescope is relatively small in most cases, so $B(\theta)$ falls to zero rapidly and $\cos\theta \approx 1$, but this approximation does not hold for the GReX instrument. Since the behavior of the $\cos\theta$ term describes the Flux Density seen at the detector, it is absorbed into the beam response function $B(\theta)$. It is standard practice to express the specific intensity (I_ν) of astrophysical signals in terms of their equivalent brightness temperature (T_b) using the Raleigh-Jeans approximation:

$$I_\nu \approx \frac{2 k_B T_b}{\lambda^2} . \tag{C.2}$$

For a given distribution of brightness temperature on the sky ($T_b(\theta, \phi)$), the detected flux density is

$$S_\nu = \frac{2 k_B}{\lambda^2} \int_\Omega d\Omega \left[T_b(\theta, \phi) B(\theta) \right] . \tag{C.3}$$

We can prescribe a constant effective temperature (T_{eff}) that gives the same flux density as for an arbitrary brightness distribution:

$$\int_\Omega d\Omega \left[T_b(\theta, \phi) B(\theta) \right] = T_{eff} \int_\Omega d\Omega \left[B(\theta) \right] = T_{eff} \Omega_B .$$

For a single-pixel instrument like GReX, this is equivalent to the antenna temperature contributed by the source. If the source does have a constant brightness temperature across the beam, then it is apparent that T_{eff} accurately describes the source temperature. However, for sources covering a small patch of sky ($T_b(\theta, \phi) = T_b\Delta'$, where$\Delta' \sim \delta(\theta - \theta', \phi - \phi')$) centered at (θ', ϕ') and with total solid angle Ω_Δ, the beam response and source brightness temperature are approximately constant and

$$T_{eff} \approx \frac{\int_\Omega T_b(\theta, \phi) B(\theta) \Delta' d\Omega}{\Omega_B} = T_b B(\theta') \frac{\Omega_\Delta}{\Omega_B}.$$

Since the flux density is related to the effective temperature like

$$S_\nu = \frac{2k_B\Omega_B}{\lambda^2} T_{eff},$$

a source at a distance $\delta\theta$ from the center of the beam will have its flux density at the receiver reduced by a factor $B(\delta\theta)$ than if it were centered in the beam:

$$S_\nu(\delta\theta) \approx B(\delta\theta) S_\nu(0).$$

While flux density is usually not used to describe extended sources with variable brightness temperature, we do use it for calibration of the instrument as described in Section 5.5 with measured and simulated values of the hydrogen I line brightness temperature from diffuse Galactic gas. As such, we include an expression for the antenna temperature for a well-known brightness distribution:

$$T_A = \int_\Omega d\Omega \left[T_b(\theta, \phi) \hat{B}(\theta) \right], \tag{C.4}$$

where $\hat{B}$ describes the normalized telescope response,

$$\hat{B}(\theta) = \frac{B(\theta)}{\Omega_B}. \tag{C.5}$$

The effective area of the telescope has the intrinsic property that

$$\langle A_e \rangle_\Omega = \frac{\int_\Omega}{4\pi} = \frac{\lambda^2}{4\pi}.$$

Following from this, and $B(\theta) = A_e(\theta)/A_0$, the maximum effective area is

$$A_0 = \frac{\lambda^2}{\Omega_B}. \tag{C.6}$$

The forward gain (g_f) relates the antenna temperature to the source flux density as

$$S_\nu = g_f T_{eff} \rightarrow g_f = \frac{2k_B\Omega}{\lambda^2},$$

equivalent with

$$g_f = \frac{2k_B}{A_0} \tag{C.7}$$

Since the forward gain relates the observed temperature of a source by the instrument to the flux density of that source, it is a critical component for defining the sensitivity of the instrument.

Due to the prevalence of describing observed signals in terms of T_A, it is convenient to describe all sources that contribute additive power to the signal within the system as a system temperature (T_{sys}), even though the noise is not necessarily thermal in nature. While there are many sources that contribute to T_{sys}, we focus on the thermal noise within the receiver electronics (T_{rec}) and the background temperature of the sky (T_{sky}). At any time, the total observed temperature by the system is

$$T_{obs} = T_{eff} + T_{sys} = T_{eff} + T_{rec} + T_{sky}. \tag{C.8}$$

However, GReX data is natively stored as linear intensities with arbitrary units (I_{arb}). The receiver gain (G, in K/arb) converts the linear intensities into Kelvin,

$$T_{obs} = GI_{arb}. \tag{C.9}$$

Since T_{sky} is well characterized to account for atmospheric effects and background radiation, our sensitivity analysis needs to account for G and T_{sys} so that the effective source temperature can be isolated as

$$T_{eff} = GI_{arb} - T_{sky} - T_{rec}. \tag{C.10}$$

We determine G and T_{sys} by performing a Y-factor test with two known values for T_{sky}. We first take measurements of the unobscured sky with a GReX terminal before collecting data again, this time with a slab of room-temperature radio-absorbent foam covering the receiver. Under the assumption that there are no significant astrophysical sources present in the data, we now have

$$I_1 = \frac{T_{sky} + T_{rec}}{G}, I_2 = \frac{T_{hot} + T_{rec}}{G}. \tag{C.11}$$

The Y-factor ($Y = I_2/I_1$) is then expanded according to the above equations and rearranged to give the following expression for the receiver temperature,

$$T_{rec} = \frac{T_{hot} - YT_{sky}}{Y - 1}. \tag{C.12}$$

From this value, the receiver gain is trivially

$$G = \frac{T_{hot} + T_{rec}}{I_2} = \frac{(T_{hot} - T_{sky})}{(I_2 - I_1)}. \tag{C.13}$$

Ideally, this receiver gain is roughly constant over observing epochs, but changes in temperature and degradation of the electronics can cause slow changes in this value. Of course, any changes to the programmable gain in the FEM of a GReX terminal will require remeasuring the receiver gain. The more volatile value is the system temperature, which can change depending on environmental effect such as prevalence of RFI and accumulation of water in the feed, among other potential issues. As such, it is useful to measure T_{sys} semi-regularly.

Once G and T_{sys} are well characterized for the terminal, we compare the measured T_{eff} of a source to its known flux density to determine the forward gain. We utilize all-sky maps of neutral galactic hydrogen to simulate the expected flux density as seen by the GReX terminal and compare this with the measured T_{eff} of the HI line by the instrument to calibrate g_f in Section 5.5. The conversion of temperature units into flux densities is commonly applied to the system temperature of an instrument to define its system equivalent flux density (SEFD),

$$\text{SEFD} = S_{\nu,sys} = g_f T_{sys}. \tag{C.14}$$

Since we are dealing with dynamic spectra that contain signals across multiple frequency channels, time samples, and polarizations, it is helpful to consider the power and temperature within the detector noise that limits the detectability of a signal. A signal spanning a bandwidth $\Delta\nu$ with total integration time τ that is seen in both polarizations of a dipole antenna will be present in $N = 2\Delta\nu\tau$ total samples. The detectability of such a signal depends on the signal-to-noise ratio

$$\text{SNR} = \frac{T_{src}}{T_{rms}}, \tag{C.15}$$

where $T_{rms} = \sigma_T/\sqrt{N}$ describes the root-mean-squared noise in the system temperature. Assuming that the thermal noise contributing to T_{sys} is exponentially distributed such that $T_{sys} = \mu_T = \sigma_T$, then

$$T_{rms} = \frac{T_{sys}}{\sqrt{N}}.$$

We can then rewrite the signal-to-noise ratio as

$$\text{SNR} = \frac{T_{src}}{T_{sys}}\sqrt{N}$$

The SNR is a useful measure since this ratio of temperatures is equivalent to the corresponding ratios of native system intensity units and the ratios of Flux Density.

Converting the above expression to flux densities gives

$$S_{\nu,src} = \mathrm{SNR}\frac{\mathrm{SEFD}}{\sqrt{N}}.$$

The flux density threshold for detection is then given by

$$S_{\nu,min} = \mathrm{SNR}_{min} \cdot \mathrm{NEFD},$$

where the NEFD (noise-equivalent flux density) is NEFD = $g_f T_{rms}$ = SEFD/$\sqrt{N}$.

C.2 HI Line Simulation

The LAB data is an all-sky map of the brightness temperatures of neutral galactic hydrogen binned into discrete velocity channels. These scalar velocities represent the component of the motion of the neutral galactic hydrogen along the line of sight from the solar system to the gas within the local standard of rest (LSR) frame. The original data format for the LAB survey had a velocity channel spacing of $\sim$ 1.031 km/s and a total velocity range of −450 km/s to +450 km/s, but this data was conglomerated into wider bins with a spacing of 10 km/s. This was done by averaging the brightness temperatures of the narrower bins into those respective larger bins. Thus, the LAB all-sky map is stored in *HEALPix* [1] file format with separate files for the difference wide velocity channels. Each *HEALPix* velocity-channel file has data stored in two fields: the first is a *TEMPERATURE* field, which gives the brightness temperature within the channel for each sky position pixel; the second is called the *SIMULATION* field, which gives the number of original $\sim$ 1.031 km/s velocity channels that were used in calculating that brightness temperature. From the *TEMPERATURE* field, we construct a map of brightness temperature, $T_{HI}(\ell_j, b_j, v_k)$, where (ℓ_j, b_j) describes the central position of the j^{th} pixel in galactic coordinates and v_k gives the k^{th} velocity channel in the LSR frame. To construct an expected spectrum of HI from this map for a specific terminal and observing time, we need to: (1) account for the velocity, $\vec{v}_{obs}$, and central pointing, $(\ell, b)_{obs}$, of the terminal in the LSR frame at that time; (2) remove the velocity of the observer in LSR along the LoS to each pixel from the neutral hydrogen velocity in LSR to get the gas velocity in the terminal's frame; (3) compute the angular deviation $(\delta\theta)$ of each pixel from the central pointing of the terminal feed and use that to construct a simulated 2D Gaussian beam response $B(\delta\theta)$; (4) convert hydrogen velocity in the observer frame to HI line frequency ($f_{HI,0}$ = 1420.40575 MHz) and bin frequencies into observing channels; and (5) for each frequency channel, integrate the brightness temperature map over the 4π steradians of sky modulated by the telescope beam response function $(B(\delta\theta, f))$ to get the expected spectrum of brightness temperature of HI as seen from a GReX unit.

To compute the velocity of neutral hydrogen in the reference frame of the GReX terminal along the LoS to each LAB data pixel, we need to start with an expression for the overall gas velocity in the observer frame:

$$\vec{v}_H, obs(t) = \vec{v}_{H,LSR} - \vec{v}_{obs,LSR}(t).$$

Then, the velocity of gas along the LoS to the j^{th} data pixel is

$$v_{H,obs,j}(t) = \hat{n}_j \cdot (\vec{v}_{H,LSR} - \vec{v}_{obs,LSR}(t)),$$

where $\hat{n}_j$ is the unit vector pointing along that LoS. The scalar velocity channels of the LAB data are already the gas velocity along the LoS, so our final expression is

$$v_{H,j,k,obs}(t) = v_{H,k,LSR} - v_{obs,j,LSR}(t),$$

where $v_{obs,j,LSR}(t) = \hat{n}_j \cdot \vec{v}_{obs,LSR}(t)$ is the component of observer velocity along the j^{th} LoS. *Astropy* handles the conversion of sky positions and velocities between different reference frames and coordinate systems to generate values for $v_{obs,j,LSR}$. We begin by defining an *EarthLocation* object for the terminal in geodetic coordinates ($x_{obs,geo}(\text{lat}, \text{lon})$), converting to a location in the barycentric celestial reference system at a specific time (t), which is then transformed directly into ICRS and then LSR coordinates ($x_{obs,LSR}(x_{obs,geo}, t)$). *Astropy* then computes the Cartesian differential of the *EarthLocation* object to determine the the GReX terminal velocity within the LSR frame ($\vec{v}_{obs,LSR}(t)$) in (x,y,z) coordinates. The LoS unit vector $\hat{n}_j = \hat{n}(\ell_j, b_j)$ that points from the observer towards the galactic coordinates ℓ_j and b_j is generated by converting an *Astropy SkyCoord* object at those galactic coordinates into (x,y,z) to remain consistent with $x_{obs,LSR}$ and $\vec{v}_{obs,LSR}(t)$. The terminal's velocity within the LSR frame along each LoS towards an individual LAB data pixel is then computed as the dot product

$$v_{obs,j,LSR}(t) = \hat{n}_j \cdot \vec{v}_{obs,LSR}(t).$$

The frequency of the HI line as seen by the GReX terminal along the j^{th} LoS and from the k^{th} LAB data velocity channel are computed from this scalar gas velocity according to

$$f_{HI,j,k} = \gamma_{j,k}(1 - \beta_{j,k}) f_{HI,0},$$

where $\beta_{j,k} = v_{obs,j,LSR}/c$, $\gamma_{j,k} = 1/\sqrt{1 - \beta_{j,k}^2}$, where $f_{HI,0}$ is the lab-frame frequency of emission for neutral hydrogen gas. After applying these transformations, our map of brightness temperature is formatted as $T_{HI}(\ell_j, b_j, f_{j,k})$. To standardize the data, we consider a hypothetical observing frequency channelization scheme labeled as f_i such that channel edges directly abut one another. The temperature map is then reconfigured such that

$$T_{HI}(\ell_j, b_j, f_i, t) = \sum_k T_{HI}(\ell_j, b_j, f_{j,k}(t))|f_{j,k}(t) \in f_i|,$$

and we now have an all-sky map of brightness temperature within each frequency channel at each observing epoch.

The actual brightness temperature seen by a GReX terminal is found by integrating the brightness temperature of your source over the normalized beam response function of the feed:

$$\widetilde{T}_{HI}(t, f) = \frac{\int_\Omega d\Omega \left[T_{HI}(\ell, b, f, t) B(\ell(t), b(t))\right]}{\int_\Omega d\Omega B(\ell(t), b(t))}.$$

In our case, we replace this integral with its numerical counterpart,

$$\widetilde{T}_{HI}(t, f_i) \approx \sum_j \Delta\Omega \left[T_{HI}(\ell_j, b_j, f_i, t) \hat{B}(\ell_j(t), b_j(t))\right],$$

which requires that we compute the normalized beam response function for a GReX terminal at each LAB pixel galactic coordinate as a function of time. We model the normalized beam response as a two-dimensional Gaussian,

$$\hat{B}(\theta, \theta_{\text{FWHM}}) = \Omega_B^{-1} e^{-4\ln 2[\theta/\theta_{\text{FWHM}}]^2},$$

with a full-width at half maximum of θ_{FWHM}. It is important to note that the $\Omega_{beam} \propto f^{-2}$ relationship according to the radiometer equation means that the beam needs to be simulated for every frequency channel. We numerically integrate over simulated beams across a range of different $\theta_{\text{FWHM}}$and interpolate between the resulting $\Omega_B(\theta_{\text{FWHM}})$. We consider a characteristic beam width $\Omega_B(f_{HI,0})$, which can be scaled across the band according to the inverse squared frequency dependence. We assume that $\Omega_B \approx \Omega_B(f_{HI,0})$ in the $\sim$ 2MHz of band used in this analysis. We parametrize the angle dependence of the beam response function into galactic coordinates such that $\theta \longrightarrow \delta\theta(\ell_j, b_j) = \delta\theta_j$, where $\delta\theta$ describes the off-axial deviation of the j^{th} data pixel coordinate from the beam central pointing $(\ell, b)_{obs}$. Thus,

$$\hat{B}(\theta, \theta_{\text{FWHM}}) \longrightarrow \hat{B}(\delta\theta_j, \theta_{\text{FWHM}}(f)) = \Omega_B^{-1} e^{-4\ln 2[\delta\theta_j/\theta_{\text{FWHM}}(f)]^2}$$

The normalizing factor Ω_B is found by numerically integrating the beam response over the full 4π steradians of sky at the *HEALPix* data coordinates,

$$\Omega_B(\theta_{\text{FWHM}}(f)) \approx \left[\Delta\Omega \sum_j e^{-4\ln 2\left(\frac{\delta\theta_j}{\theta_{\text{FWHM}}(f)}\right)^2}\right].$$

The central pointing of the GReX feed is needed to calculate $\delta\theta_j$. We first define an *AltAz* frame for the terminal at the appropriate *EarthLocation* and time, t which is then fed into a *SkyCoord* object with the corresponding altitude and azimuth values for the feed (which should be 90° and 0°, respectively). We then extract the galactic coordinates that correspond to this *SkyCoord* object, which acts as the central pointing of the feed, $(\ell, b)_{obs}(t)$. The difference in angle ($\delta\theta$) between the j^{th} data point and the terminal pointing in galactic coordinates is calculated according to

$$\cos\delta\theta_j(t) = \cos\left[90^\circ - b_{obs}(t)\right]\cos\left[90^\circ - b_j\right] + \sin\left[90^\circ - b_{obs}(t)\right]\sin\left[90^\circ - b_j\right]\cos\left[\ell_j - \ell_{obs}(t)\right].$$

We include a small-angle approximation for numerical stability, which gives the resulting angular distance as

$$\delta\theta_j(t) = \begin{cases} \sqrt{2[1 - \cos\delta\theta_j(t)]}, & |1 - |\cos\delta\theta_j(t)|| < 10^{-6} \\ \arccos\left[\cos\delta\theta_j(t)\right], & \text{else} \end{cases} .$$

Using this definition for $\delta\theta_j(t)$, we generate an all-sky map of the beam response at each *HEALPix* data coordinate for all observing times and frequency channels,

$$B_{\theta_{\mathrm{FWMH}, c}, f_c}(\ell_j, b_j, f_i, t) = \exp\left[-4\ln 2\left(\frac{\delta\theta_j(t)}{\theta_{\mathrm{FWHM}}(f_i)}\right)^2\right],$$

and compute the normalizing factor as $\Omega_B(t, f_i) = \Delta\Omega\sum_j B_{\theta_{\mathrm{FWHM}, c}, f_c}(\ell_j, b_j, f_i, t)$. Substituting the expressions for $\theta_{\mathrm{FWHM}}(f)$ and $\delta\theta_j(t)$ into the earlier expressions for B and Ω_B yields the normalized beam response ($\hat{B}$) as a sky map at the *HEALPix* galactic coordinates for each observing time and frequency channel.

We can now apply this set of simulated beam responses to the all-sky temperature map ($T_{HI}(\ell_j, b_j, f_i, t)$) and sum over the j data coordinates to generate the expected observed spectrum of neutral galactic hydrogen in temperature units as seen from the GReX terminal at each observing time:

$$\widetilde{T}(t, f_i) = \frac{\Delta\Omega}{\Omega_B(t, f_i)}\sum_j\left[T_{HI}(\ell_j, b_j, f_i, t)B_{\theta_{\mathrm{FWHM}, c}, f_c}(\ell_j, b_j, f_i, t)\right].$$

In Section 5.5, we use the method described above to generate the expected spectra of the HI-line for terminals at OVRO and Cornell University. We find that integrating across the spectrum at each observation time and comparing the total HI-line intensity was sufficient for determining the size of the terminal beams.

References

[1] K. M. Gorski, E. Hivon, A. J. Banday, *et al.*, "HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere," *The Astrophysical Journal*, vol. 622, no. 2, pp. 759–771, Apr. 2005. DOI: 10.1086/427976.