Differentiable, End-to-End Forward Modeling for 21 cm Cosmology: Robust Systematics Error Budgeting and More

Nicholas Kern*
MIT Kavli Institute for Astrophysics and Space Research
Cambridge, MA
nkern@mit.edu

Abstract

A new generation of radio telescopes are being built to map the growth of cosmological structure throughout the majority of the observable universe, giving us access to new cosmological information that will shed light on outstanding questions in astrophysics and cosmology. These telescopes use 21 cm emission from neutral hydrogen as a tracer of structure, but at the low radio frequencies that they operate face a daunting systematics suppression challenge. These systematics are wide ranging, and are generally considerably brighter than the underlying cosmological signal of interest, setting up a delicate signal separation problem that has yet to be overcome by the field. We present the first differentiable, end-to-end forward model for 21 cm cosmology that will allow us to better subtract low-level systematics and compute more robust errorbars. This framework is made possible by high-performance machine learning frameworks that use automatic differentiation to quickly compute exact posterior gradients that are then fed to gradient-aware optimization and posterior sampling routines. In this work we give an overview of the problem statement, demonstrate a proof of concept for the framework, and discuss near-term improvements for future works.

1 Introduction

Neutral hydrogen is ubiquitous throughout the universe, making it an ideal probe of the growth of cosmological structure over the near entirety of cosmic time. In particular, the hyperfine splitting of the neutral hydrogen ground state emits or absorbs a photon at a wavelength of 21 cm, which has been widely used to study the structure of our own Milky Way galaxy and of nearby galaxies. However, this signal can also be used to map out large scale structures at cosmological distances [1], putting its measured redshifted emission in the low-frequency part of the radio spectrum ($\sim$ 100 MHz).

A new generation of radio telescopes are being built to capitalize on the vast quantity of cosmological information traced by this spectral line, which will shed light on an array of science topics, ranging from the formation of the first stars and galaxies at Cosmic Dawn, the nature of dark energy, and the nature of inflation [2]. However, these telescopes face the challenge of bright contaminating emission from non-thermal synchrotron emission at these low radio frequencies, emanating not only from our own galaxy but from the numerous population of radio galaxies spanning the sky. Theoretical models tell us that the background cosmological signals of interest traced by the redshifted 21 cm emission is roughly 10^5 times weaker than these foreground contaminants [3], setting up a delicate, high dynamic range signal separation problem that has yet to be solved by the field.

*NASA Hubble Fellow

Machine Learning and the Physical Sciences Workshop, NeurIPS 2023.

While 21 cm analysis pipelines are becoming increasingly sophisticated, the field still struggles in being able to propagate uncertainty from the various components in our complex, high-dimensional and non-linear data model down to our final data products. In other words, we have yet to build a framework capable of exploring the joint posterior distribution between systematics, foregrounds, and the cosmological signal. This is part because of the immense scope of the data model for 21 cm interferometers, which include multiplicative responses in our sky and instrumental model (thus resulting in a set of model parameters that have non-linear relationships), in addition to the many Ns of our raw dataset: every night these telescopes produce datasets with $\mathcal{O}(10^3)$ frequency channels, $\mathcal{O}(> 10^3)$ time integrations, $\mathcal{O}(> 10^3)$ antenna-pair correlations (referred to as "baselines"), all coming from a sky brightness distribution that is generally represented with $\mathcal{O}(> 10^5)$ pixels.

In this work, we present the first end-to-end, differentiable forward model for redshifted 21 cm telescopes, which can jointly model sky signals along with the telescope's instrumental response at the various stages of its measurement. Fundamentally, the forward model is a computational representation of the radio interferometric measurement equation, which describes how the raw telescope observable (referred to as the *visibility*) is related to the incoming sky radiation as its measured by the instrument [4; 5]. The visibility, V_{jk}, formed between two distinct antennas j and k can be written as

$$V_{jk}(\nu) = G_j G_k^*(\nu) \int A(\hat{s}, \nu) e^{2\pi i \vec{b}_{jk} \cdot \hat{s} \nu / c} B(\hat{s}, \nu) d\Omega, \tag{1}$$

where $B(\hat{s}, \nu)$ is the specific intensity of electromagnetic radiation incident on the telescope from angular direction $\hat{s}$ and at observing frequency ν, $\vec{b}$ is the baseline vector formed between antennas j and k, $A(\hat{s}, \nu)$ is the antenna efficiency along the direction $\hat{s}$ and at frequency ν, also referred to as its *primary beam*, $d\Omega$ is a solid angle differential where the integral is taken over the full 4π steradians of the sky, and finally G_j, G_k are the complex-valued, front-end gains of the analog and electronic of the telescope for antennas j and k. Note that many more kinds of effects can in principle be considered (e.g. the ionosphere, radio frequency interference) but we will use this formulation for now as it is the most common. Also note that all terms Equation 1 are also in principle time-dependent (i.e. they can vary over the course of observations), although in this work we will assume they are time-independent for simplicity. The multiplicative relationship between the various components on the RHS of Equation 1 make this a non-linear optimization problem when trying to solve for them jointly. As a consequence, current approaches generally leave one component as a free parameter while assuming we know the other components, even though we generally do not know any of these components to better than a few percent.

2 End-to-End Forward Modeling

Today, end-to-end forward modeling approaches are commonplace in the physical sciences, particularly for analyzing datasets in astrophysics and cosmology [e.g. 6; 7; 8]. These have been made possible by the advent of large compute clusters equipped with high-memory graphics processing units (GPUs), as well as new machine learning frameworks that leverage a technique known as automatic differentiation [9] to compute gradients of a loss function very quickly. Utilizing GPUs can dramatically accelerate the forward model runtime, while having access to gradient information allows us to use high-performing optimization techniques such as quasi-Newton schemes, as well as efficient posterior Markov Chain Monte Carlo (MCMC) sampling techniques such as Hamiltonian Monte Carlo [10]. Together, this can bring some joint optimization and sampling inverse problems that previously were untractable into the realm of feasibility.

Leveraging these same tools, we present the first end-to-end, differentiable forward model for the 21 cm measurement equation (Equation 1), where by "end-to-end" we mean a model that simulates a sky signal all the way through the instrumental response down to the raw visibility data, and by "differentiable" we mean a model that is easily and *exactly* differentiable (e.g. through a technique like automatic differentiation). A visualization of the forward model is shown in Figure 1, demonstrating how our model explicitly parameterizes various kinds of foregrounds and instrumental responses, while seamlessly incorporating any prior knowledge we have on them (e.g. from lab measurements or other telescope observations) into our posterior distribution, which is then differentiated and fed to an optimization or MCMC sampling routine. Our model is implemented in PyTorch [11].

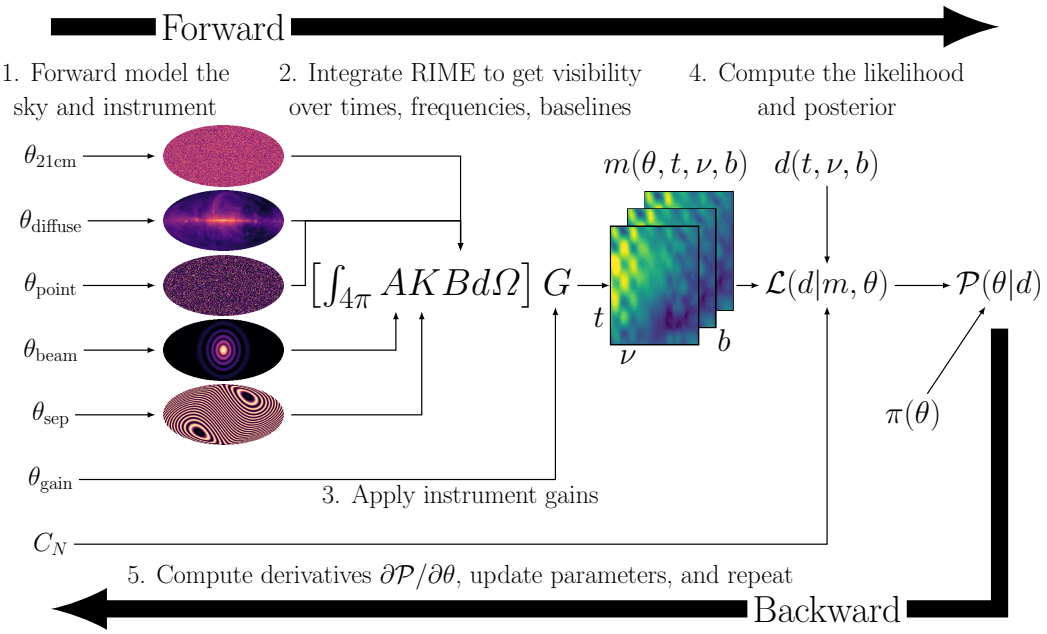

Figure 1: A visualization of an end-to-end forward model for 21 cm radio interferometric measurement equation (RIME), which produces the telscope observable known as the *visibility*. We parameterize sky signals such as the cosmological signal ($\theta_{21\mathrm{cm}}$) and foregrounds ($\theta_{\mathrm{diffuse}}$, θ_{point}), as well as the various components of the instrumental response (θ_{beam}, θ_{sep}, θ_{gain}). Model visibilities (m) are computed with a forward pass of the model, which are compared against raw visibility data to form a likelihood given a noise model (C_N), while our prior incorporates any previous knowledge we may have on those parameters, say, from lab measurements or from other telescope observations. Implementing this in a deep learning framework like `PyTorch` enables us to rapidly compute exact posterior gradients, which are used for parameter optimization and posterior sampling, as well as seamlessly push our forward model onto a (multi)GPU node for further speedups.

3 Proof of Concept: FGs, Instrumental Response, and 21 cm Signal

Here we will demonstrate our framework on a simplified problem setup. For each component of our model, we take their fiducial parameter values and perturb them to simulate a "truth" forward model output, which we then seek to reconstruct via a large-scale unconstrained optimization. The goal of this exercise is to 1. recover a residual signal that is above the thermal noise uncertainty, and 2. compute the propagated uncertainty from the foregrounds, instrumental response and the 21 cm signal on this residual.

Our telescope model will be a scaled-down version of the Hydrogen Epoch of Reionization Array (HERA) experiment [12], with 91 antennas packed in a hexagonal configuration. For simplicity, we only use baselines with a total length of 40 meters or less. Our simulated frequency range spans 120 – 130 MHz, and our simulated observations span 5 hours in drift-scan mode. All components of the model specified below have Gaussian priors with widths that are $\sim 10\%$ of the parameters' fiducial values, with the exception of the 21 cm model which has a flat prior. Other components of the data model include:

Diffuse galactic foregrounds – We use the (full-sky) Global Sky Model [13] as our fiducial model, and use a complex spherical harmonic decomposition as a parameterization with $\ell_{\mathrm{max}} = 80$. We perturb their initial values by roughly 5% to simulate a realistic level of uncertainty. We parameterize the frequency response with a second-order orthogonal Legendre polynomial.

Point source foregrounds – We use the GLEAM low-frequency radio point source catalogue [14] and take all sources with a perceived flux density (i.e. after accounting for primary beam attenuation) above 5 Jy, resulting in a few hundred sources. We perturb their amplitudes by 3%, and assume we know their positions while leaving their amplitudes as free parameters.

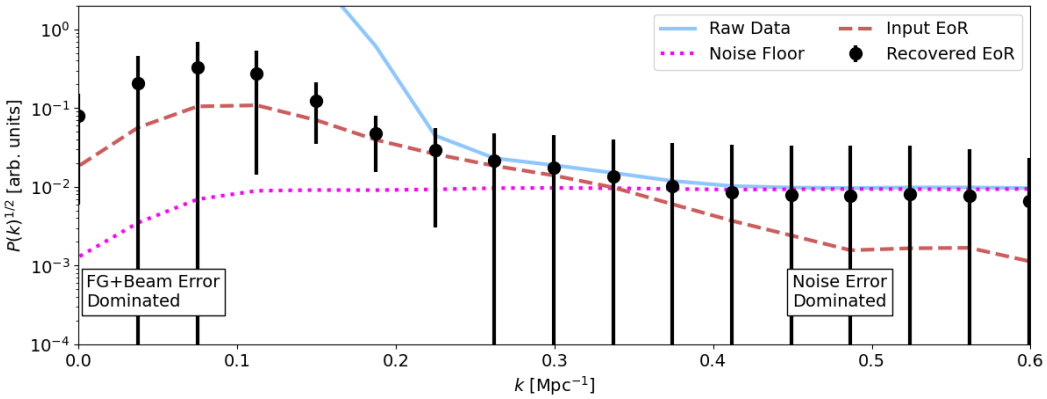

Figure 2: Power spectrum recovery after optimization and then posterior expansion. The raw data (FG+EoR+Noise) is plotted in blue, the intrinsic noise floor in magenta, and the fiducial EoR model in red. We plot the recovered EoR signal after optimization (black points) along with its marginal posterior standard deviations (68% credible region errorbars), showing non-detections at low-k (FG+Beam error dominated) and high-k (noise dominated), but detections at intermediate k modes. The key advancement is the ability to forward model "Systematics" errorbars which tend to peak at low-k.

Primary beam – We use an airy disk power pattern for the direction and frequency dependent antenna primary response, decomposed as spherical harmonics with $\sim$10% perturbations. We enforce that the beam response is unity at its zenith pointing throughout the optimization. We parameterize the frequency response with a fifth-order orthogonal Legendre polynomial.

21 cm Signal – Our signal model resembles our diffuse foreground model but with additional frequency terms (30 in total), and the angular extent is confined to a 10 degree x 60 degree patch on the sky. The fiducial model is the default output from the 21 cmFAST code [15].

We omit the G_j gain terms for simplicity, as the above model is sufficient to demonstrate the capability of the framework. In total, the model has roughly 50,000 parameters, We also add white thermal noise at a low level such that a 21 cm detection could be possible (e.g. with $\sim$1000 hours of HERA observations). Note that we apply a high-pass filtering to the visibilities across frequency to eliminate some of the foreground modes that are strongly degenerate with the 21 cm signal modes [16].

We first run a joint optimization of our data model against our noisy, truth simulation, using the LBFGS solver [17]. Then we expand the posterior distribution about its maximum a posteriori (MAP) value, AD to compute the Hessian and then inverting it (i.e. a Laplace approximation). The results are shown in Figure 2, showing the recovered EoR component (black points) with their marginal standard deviations. We see that, relative to the raw data (blue), our recovered points make EoR signal detections where previously the data were contaminated ($k \sim 0.15$ Mpc^{-1}). The other key advancement in this work is being able to forward model a "systematics" error budget (in this case a FG + beam error budget), which tends to peak at low Fourier k wavevectors. We see that the recovered spectra are noise dominated at high k and systematics dominated at low k, but we see fiducial detections at intermediate k modes. This allows us to be more confident when we begin to make real signal detections (at intermediate k) that we are accounting for the full error budget of the system (rather than just a thermal noise errorbar, which is generally the standard).

Future work will incorporate more types of known systematics, improve runtime efficiency to scale-up to high angular resolutions, and use HMC to directly sample the posterior instead of expanding it via a Laplace approximation.

4 Broader Impact

While the science drivers for 21 cm are diverse, the various kinds of systematics faced by 21 cm telescopes operating at low and high frequencies can be quite similar, thus a breakthrough by one experiment will help others as well. Lastly, 21 cm science is highly complementary to other

cosmological surveys (e.g. DESI, Euclid) as well as astrophysical telescopes (e.g. JWST, SphereX), and a robust 21 cm analysis pipeline would unlock a multi-pronged approach to inferring astrophysical and cosmological parameters.

Acknowledgments and Disclosure of Funding

NK acknowledges support from NASA through the NASA Hubble Fellowship grant #HST-HF2-51533.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555.

References

[1] S. R. Furlanetto, S. P. Oh, and F. H. Briggs, Physics Reports **433**, 181 (2006), arXiv:astro-ph/0608032 [astro-ph] .

[2] A. Liu and J. R. Shaw, PASP **132**, 062001 (2020), arXiv:1907.08211 [astro-ph.IM] .

[3] A. Mesinger, B. Greig, and E. Sobacchi, MNRAS **459**, 2342 (2016), arXiv:1602.07711 [astro-ph.CO] .

[4] J. P. Hamaker, J. D. Bregman, and R. J. Sault, A&AS **117**, 137 (1996).

[5] O. M. Smirnov, A&A **527**, A106 (2011), arXiv:1101.1764 [astro-ph.IM] .

[6] B. Popovic, D. Brout, R. Kessler, and D. Scolnic, arXiv e-prints , arXiv:2112.04456 (2021), arXiv:2112.04456 [astro-ph.CO] .

[7] C. Hahn, M. Eickenberg, S. Ho, J. Hou, P. Lemos, E. Massara, C. Modi, A. Moradinezhad Dizgah, B. Régaldo-Saint Blancard, and M. M. Abidi, arXiv e-prints , arXiv:2211.00723 (2022), arXiv:2211.00723 [astro-ph.CO] .

[8] BeyondPlanck Collaboration, K. J. Andersen, R. Aurlien, R. Banerji, A. Basyrov, M. Bersanelli, S. Bertocco, M. Brilenkov, M. Carbone, L. P. L. Colombo, H. K. Eriksen, J. R. Eskilt, M. K. Foss, C. Franceschet, U. Fuskeland, S. Galeotta, M. Galloway, S. Gerakakis, E. Gjerløw, B. Hensley, D. Herman, M. Iacobellis, M. Ieronymaki, H. T. Ihle, J. B. Jewell, A. Karakci, E. Keihänen, R. Keskitalo, J. G. S. Lunde, G. Maggio, D. Maino, M. Maris, A. Mennella, S. Paradiso, B. Partridge, M. Reinecke, M. San, N. O. Stutzer, A. S. Suur-Uski, T. L. Svalheim, D. Tavagnacco, H. Thommesen, D. J. Watts, I. K. Wehus, and A. Zacchei, A&A **675**, A1 (2023), arXiv:2011.05609 [astro-ph.CO] .

[9] A. Gunes Baydin, B. A. Pearlmutter, A. Andreyevich Radul, and J. M. Siskind, arXiv e-prints , arXiv:1502.05767 (2015), arXiv:1502.05767 [cs.SC] .

[10] R. Neal, in *Handbook of Markov Chain Monte Carlo* (2011) pp. 113–162.

[11] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, arXiv e-prints , arXiv:1912.01703 (2019), arXiv:1912.01703 [cs.LG] .

[12] D. R. DeBoer, A. R. Parsons, J. E. Aguirre, P. Alexander, Z. S. Ali, A. P. Beardsley, G. Bernardi, J. D. Bowman, R. F. Bradley, C. L. Carilli, C. Cheng, E. de Lera Acedo, J. S. Dillon, A. Ewall-Wice, G. Fadana, N. Fagnoni, R. Fritz, S. R. Furlanetto, B. Glendenning, B. Greig, J. Grobbelaar, B. J. Hazelton, J. N. Hewitt, J. Hickish, D. C. Jacobs, A. Julius, M. Kariseb, S. A. Kohn, T. Lekalake, A. Liu, A. Loots, D. MacMahon, L. Malan, C. Malgas, M. Maree, Z. Martinot, N. Mathison, E. Matsetela, A. Mesinger, M. F. Morales, A. R. Neben, N. Patra, S. Pieterse, J. C. Pober, N. Razavi-Ghods, J. Ringuette, J. Robnett, K. Rosie, R. Sell, C. Smith, A. Syce, M. Tegmark, N. Thyagarajan, P. K. G. Williams, and H. Zheng, PASP **129**, 045001 (2017), arXiv:1606.07473 [astro-ph.IM] .

[13] H. Zheng, M. Tegmark, J. S. Dillon, D. A. Kim, A. Liu, A. R. Neben, J. Jonas, P. Reich, and W. Reich, MNRAS **464**, 3486 (2017), arXiv:1605.04920 [astro-ph.CO] .

[14] N. Hurley-Walker, J. R. Callingham, P. J. Hancock, T. M. O. Franzen, L. Hindson, A. D. Kapińska, J. Morgan, A. R. Offringa, R. B. Wayth, C. Wu, Q. Zheng, T. Murphy, M. E. Bell, K. S. Dwarakanath, B. For, B. M. Gaensler, M. Johnston-Hollitt, E. Lenc, P. Procopio, L. Staveley-Smith, R. Ekers, J. D. Bowman, F. Briggs, R. J. Cappallo, A. A. Deshpande, L. Greenhill, B. J. Hazelton, D. L. Kaplan, C. J. Lonsdale, S. R. McWhirter, D. A. Mitchell, M. F. Morales, E. Morgan, D. Oberoi, S. M. Ord, T. Prabu, N. U. Shankar, K. S. Srivani, R. Subrahmanyan, S. J. Tingay, R. L. Webster, A. Williams, and C. L. Williams, MNRAS **464**, 1146 (2017), arXiv:1610.08318 [astro-ph.GA] .

[15] A. Mesinger, S. Furlanetto, and R. Cen, MNRAS **411**, 955 (2011), arXiv:1003.3878 [astro-ph.CO] .

[16] N. S. Kern and A. Liu, MNRAS **501**, 1463 (2021), arXiv:2010.15892 [astro-ph.CO] .

[17] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. (Springer, New York, NY, USA, 2006).