ON THE DETECTION OF THE FIRST STARS AND GALAXIES WITH THE 21-CM GLOBAL SIGNAL

Jack O. Burns1,9, T. Joseph W. Lazio2, Steven Furlanetto3, Jordan Mirocha1,3, Geraint Harker4, Abhirup Datta1, Judd Bowman3, Richard Bradley6, Dayton Jones2,11, Justin C. Kasper7, Abraham Loeb6, Michael Bicay9, Jonathan Pritchard10, Michael Seiffert8

1Center for Astrophysics and Space Astronomy, Department of Astrophysical and Planetary Science, University of Colorado, Boulder, CO 80309, USA
2Jet Propulsion Laboratory, California Institute of Technology, MS 138-308, 4800 Oak Grove Dr., Pasadena, CA 91109, USA
3Department of Physics and Astronomy, University of California at Los Angeles, Los Angeles, CA 90095, USA
4Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
5Arizona State University, School of Earth and Space Exploration, P.O. Box 876004, Tempe, AZ 85287, USA
6National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA
7Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI 48109, USA
8Center for Astrophysics, 60 Garden St., MS 51, Cambridge, MA 02138, USA
9NASA Ames Research Center, Moffet Field, CA 94035, USA
10Department of Physics, Blackett Laboratory, Imperial College, London SW7 2AZ, UK and
11Space Science Institute, 4750 Walnut Street, Suite 205, Boulder, CO 80301, USA

ABSTRACT

The redshifted 21-cm monopole is expected to be a powerful probe of the epoch of the first stars and galaxies (10 < z < 35), never before observed. This hyperfine transition of HI is sensitive to the thermal and ionization state of hydrogen gas and thus provides a tracer of sources of energetic photons – primarily hot stars and accreting black holes – which ionize and heat the high redshift intergalactic medium. Sky-averaged spectra of the redshifted 21-cm background are expected to exhibit broad absorption and emission features relative to the microwave background at low radio frequencies, ≈40 – 120 MHz. In a low altitude lunar orbit, such spectral observations avoid terrestrial radio frequency interference and ionospheric corruption of the signal while on the farside. Since the 21-cm spectrum has broad spectral features whereas the foregrounds do not, the signal can be extracted through Bayesian methods including Markov Chain Monte Carlo techniques. We illustrate how this signal can be detected in the presence of the foregrounds using the instrument parameters for the Dark Ages Radio Explorer (DARE) as an example. We further demonstrate that meaningful constraints on the Ly-α, Lyman continuum, and X-ray emissions of high redshift galaxies, along with the characteristic minimum mass of the first star-forming halos, can be extracted from the 21-cm global signal.

Keywords: cosmology: dark ages, reionization, first stars - cosmology: observations

1. INTRODUCTION

One of the last frontiers of observational cosmology is the time period stretching from the end of the Dark Ages through Cosmic Dawn (≈80–500 million years after the Big Bang). This is a virtually unobserved yet key epoch in the early Universe. During this interval, the first luminous objects including stars, galaxies, and active galactic nuclei (powered by accreting black holes) “turned on” (e.g., Loeb & Furlanetto 2013). Furthermore, this time period saw the birth of structural complexity in the Universe. At the beginning of the Dark Ages, corresponding to the Epoch of Recombination, the Universe was smooth to 1 part in 10^8 as evidenced by the Cosmic Microwave Background (e.g., Mather et al. 2013). Yet less than a billion years later, the Universe was teeming with complex structure spanning many orders of magnitude in scale. Thus, this transition time in the Universe is vital to understanding how the core components and structures of today’s Universe came to be.

The highly-redshifted 21-cm spectral line of neutral hydrogen, produced by a spin-flip hyperfine transition (Field 1958), provides an observable window into the early Universe’s intergalactic medium (IGM) before the Epoch of Reionization (EoR) was complete (Madau et al. 1997a). The heating and ionization caused by the “first objects to light up the Universe”1 serve as indirect probes of the nature of the first stars and galaxies. With an effective optical depth of ≈1%, and sensitivity to low temperatures, the resulting signal measured against the CMB permits us to investigate a large evolutionary range from the Dark Ages through the end of the EoR (e.g., Furlanetto et al. 2006; Morales & Wyithe 2010; Pritchard & Loeb 2012).

The 21-cm all-sky or global signal (Shaver et al. 1999; Pritchard & Loeb 2010) (Figure 1) is an attractive observational target for either a single antenna (e.g., Bowman & Rogers 2010; Burns et al. 2012; Patra et al. 2013; Voytek et al. 2014; Bernardi et al. 2015; Sokolowski et al. 2015a) or a small, compact array of antennas (e.g., Presley et al. 2015; Vedantham et al. 2015; Singh et al. 2015). Features in the spectrum may provide the first constraints on the birth and nature of the first luminous objects (e.g., Furlanetto 2006). Such an experiment for 21-cm cosmology is analogous to the COBE measurement of the CMB blackbody spectrum, which set the stage for

more detailed studies of spatial fluctuations by WMAP and Planck.

In this paper, we describe how the 21-cm global signal can be used to place the first constraints on the time of formation and the characteristics of the first stars and galaxies. We demonstrate how signal extraction using Bayesian statistical techniques, in the presence of strong galactic and extragalactic foregrounds, can measure spectral features (called “Turning Points”) and bound the spectra and masses of the first luminous objects.

The case is presented for performing these observations at low frequencies down to 40 MHz in an orbit around the Moon with data taken only above the lunar farside, which is free of distortion effects from the Earth’s ionosphere (Vedantham et al. 2014; Datta et al. 2014) and human-generated radio frequency interference (e.g., McKinley et al. 2013). In addition, the Moon shields the instrument (about half the time) from variable solar emission caused by flares and coronal activity (e.g., Mercier & Trotter 1997). Above the nighttime, pristine, radio-quiet lunar farside (as seen by RAE-2, Alexander & Kaiser 1976), such observations bypass the overwhelming challenges presented by the Earth and the Sun, and enables precision measurements of the global 21-cm signal.

The key insight permitting Cosmic Dawn to be detected with relatively simple observations is that once the Moon blocks solar and terrestrial RFI, all other foregrounds vary smoothly with frequency in contrast to the expected 21-cm spectral structure. The 21-cm signal strength is much weaker than the Galactic foreground, but the 21-cm signal is spatially uniform at angular scales $\gtrsim 10^\circ$ (e.g., Bittner & Loeb 2011; Liu et al. 2013); and has distinct spectral features whereas the foreground varies spatially and is spectrally featureless, permitting a clean separation. The 21-cm cosmological signal can then be extracted using algorithms similar to those employed for CMB observations implemented via a Markov Chain Monte Carlo framework (Harker et al. 2012, 2015).

This paper is organized as follows. In Section 2, an overview of the expected sky-averaged signal is presented. Section 3 describes the nature and brightness of foregrounds on Earth, in Earth orbit, and at the Moon which must be considered in efforts to measure the much weaker 21-cm Cosmic Dawn signal. Section 4 describes our approach for signal extraction using Bayesian statistical methods in the presence of extraterrestrial foregrounds. We use the instrument parameters and sensitivity for the Dark Ages Radio Explorer (DARE) (Burns et al. 2012) to illustrate the practical application of this statistical formalism. In Section 5, we describe the physical parameters (and their errors) associated with the first stars, black holes, and galaxies that are expected to be measured using the 21-cm all-sky spectrum. Section 6 presents a summary of the potential use of the 21-cm background to detect the first luminous objects in the early Universe.

2. OVERVIEW OF THE 21-CM GLOBAL SIGNAL

The expanding Universe began from a hot, nearly uniform mixture of ordinary matter (baryons) and dark matter. Once the temperature of the cosmic plasma dipped below 3000 K, protons and electrons combined to form hydrogen atoms, and the Universe became trans-parent to the blackbody radiation bath that filled it (the CMB) (Planck Collaboration et al. 2014, and references therein). Tiny fluctuations in the matter density at these early times, primarily on small physical scales, grew gravitationally and eventually led to the collapse of the first gas clouds $\sim 30$ Myr after the Big Bang (Abel et al. 2002; Bromm et al. 2002). These clumps seeded the first bursts of star formation, illuminating the Universe in the Cosmic Dawn (e.g., Loeb & Furlanetto 2013; Bromm & Yoshida 2011). Because the Cosmic Dawn of structure formation has not yet been investigated observationally, its direct detection is one of the most exciting frontiers in astrophysics.

X-ray and UV emission from these objects and their descendants heated and ionized the tenuous gas that lies between galaxies (the IGM), culminating in the Epoch of Reionization several hundred Myrs later — when high-energy photons from these sources ionized all of the IGM’s hydrogen atoms.

The 21-cm background can measure the effects of radiation from the first stars and black holes with the hyperfine line of the neutral hydrogen (HI) gas pervading the Universe. The expansion of the Universe redshifts these photons from earlier epochs in the Universe to lower observed frequencies, $\nu = 1420/(1+z)$ MHz (e.g., at $z = 30$, $\nu = 45$ MHz). Importantly, this phenomenon allows us to reconstruct the history of the Universe as a function of time (as each observed frequency corresponds to a different moment in the Universe’s history).

Figure 1 shows some example predictions (amongst the huge range currently allowed) for the 21-cm spectrum during the Dark Ages and Cosmic Dawn. The brightness temperature of this 21-cm signal is given by (e.g., Madau et al. 1997b; Furlanetto et al. 2006; Shaver et al. 1999),

$$\delta T_B \simeq 27 x_{\text{HI}} \left( \frac{T_s - T_{\text{CMB}}}{T_s} \right) \left( \frac{1 + z}{10} \right)^{1/2} \left( 1 + \delta_b \right) \left( \frac{\partial_v u_r}{(1+z)H(z)} \right)^{-1} \text{mK}$$

where $x_{\text{HI}}$ is the fraction of neutral gas, $T_s$ is the 21 cm spin temperature, $T_{\text{CMB}}$ is the CMB temperature, $\delta_b$ is the baryon overdensity (taken here to be $\delta_b \sim 0$), and $H(z)$ is the Hubble parameter. The last term in the above equation includes the effect of the peculiar velocities with line of sight velocity derivative $\partial_v u_r$. The effects of the last term in Equation 1 are negligible for observations of the sky-averaged global signal (e.g., Bharadwaj & Ali 2004; Barkana & Loeb 2005).

Several important physical processes drive the evolution of $T_B$ with redshift. These include: (1) UV radiation from the first stars, which “activates” the spin-flip signal through the Wouthuysen-Field mechanism (Wouthuysen 1952; Field 1958); (2) X-ray heating, likely generated by gas accretion onto the first black holes; and (3) ionizing photons from the first galaxies (which destroy the neutral

2 While X-ray heating is the most widely-accepted heating mechanism in the IGM, Kaurov & Gnedin (2015) have argued that shock heating may instead dominate. We focus here on the X-ray heating scenario because other studies have shown only minor effects from shock heating (Furlanetto & Loeb 2004; McQuinn & O’Leary 2012).
The 21-cm Global Signal

Figure 1. The evolution of a slice of the Universe is presented, from high redshifts at left to low redshifts at right (upper panel), as well as the corresponding sky-averaged 21-cm signal relative to the CMB (lower panel). The first stars form at $z \sim 30$ in the fiducial model simulation (labeled C) and “turn on” the 21-cm signal in absorption against the CMB (light blue in top panel). The fiducial model assumes the first sources are similar to those in nearby galaxies. At $z \sim 20$ (Turning Point C), X-rays from accreting black holes heat the gas, transforming the signal from absorption to emission ($z \sim 12$, Turning Point D, red in top panel). At later times, ionized bubbles (black, upper panel) grow and eventually fill the universe by $z \sim 6$. The timing of these transitions is highly model-dependent. The bottom panel illustrates a range of plausible models for the first stars and black holes. Blue curves vary the $\text{Ly}-\alpha$ production from the first stars by a factor of 10 above (dashed) and below (dotted) the fiducial value, while green curves vary the X-ray heating rate, again by factors of 10 above (dashed) and below (dotted) the input value. The red bars show the expected errors (at 95% confidence) for DARE (Burns et al. 2012), demonstrating how precision observations can distinguish between these models. Figure adapted from Pritchard & Loeb (2010) using the new reference model from Mirocha et al. (2015). Cross-hatched areas indicate the maximum allowed amplitude of the signal, which in absorption is set by the adiabatic cooling of the IGM, and in emission, by the “saturated limit” in which $T_S \gg T_{\text{CMB}}$ and $x_{\text{HI}} = 1$.

hydrogen).

The relative importance of these radiation fields determines the shape of the spin-flip spectrum. While the timing and amplitude of the features varies dramatically, Turning Points B, C, and D generally persist over large regions of parameter space. (Furlanetto et al. 2006; Pritchard & Loeb 2010; Mesinger et al. 2011) (Figure 1). Given this, we will use a scenario with three spectral features as a convenient test case for quantitative estimates.

In Figure 1, the absorption trough around Turning Point A ($z \sim 80, \nu \sim 18\text{ MHz}$) is produced solely by the cosmological conditions within the Dark Ages (i.e., adiabatic cooling of the IGM results in gas temperatures that are $< T_{\text{CMB}}$). The Cosmic Dawn window, which we focus on in this paper, begins at $z \sim 35$ when gas is nearly invisible (Turning Point B). When the first stars appeared, their radiation drove $T_s$ toward the cold temperatures characteristic of IGM gas (Madau et al. 1997a). This period, Turning Point B ($z \sim 30, \nu \sim 45\text{ MHz}$) in our fiducial model, marks a crucial transition in the Universe’s history. As more stars appear and the radiation field strengthens, the spin temperature continues to decline, which we can observe as a deep absorption trough (between Turning Points B and C). Shortly after these stars appear, black holes likely formed, e.g., as remnants of the first stars (e.g., Heger & Woosley 2002). These remnants, particularly those in binary systems, likely dominated the X-ray emission of galaxies in the early Universe, if high-$z$ galaxies are similar to nearby star-forming galaxies (Mineo et al. 2012a). Thermal bremsstrahlung from the hot interstellar medium (Mineo et al. 2012b) or inverse-Compton emission from supernova remnants (Oh 2001), however, may also be important. These X-rays are crucial because they can travel large distances through the intergalactic medium and deposit a fraction of their energy as heat (Shull & van Steenberg 1985; Furlanetto & Johnson Stoever 2010). In our scenario, the heating caused by this growing X-ray background caused the absorption trough to peak at Turning Point C ($z \sim 20, \nu \sim 68\text{ MHz}$) and then transform the spin-flip signal from absorption into emission as the gas becomes hotter than the CMB (between Turning Points C and D).

Finally, at Turning Point D in our model, photons from these stars and black holes start ionizing the IGM gas (when $x_{\text{HI}} \lesssim 90\%$, $z \sim 12, \nu \sim 108\text{ MHz}$). Then, the spin-flip signal declines and eventually disappears at $z \sim 6$.

Although many models follow this pattern, the unknown properties of early stars and black holes imply large uncertainties in the amplitudes of the relevant radiation fields, spanning at least several orders of magnitude (e.g., Pritchard & Loeb 2010; Mesinger et al. 2014; Mirocha 2014; Fialkov et al. 2014; Pacucci et al. 2014; Kaurov & Gnedin 2015). Figure 1 illustrates the effects of these uncertainties on the 21-cm signal. The depths and locations of the Turning Points vary dramatically between the different scenarios. Potential instruments, such as that proposed for DARE, can distinguish between the current extremely wide range of possible models for the first luminous objects (see also Sections 4 and 5).

3. FOREGROUNDS
The most challenging aspect of measuring the expected features in the 21-cm global spectrum shown in Fig. 1 is the presence of foregrounds which are orders of magnitude brighter than the cosmological signal (Figure 2). For any space or ground-based observations, important foregrounds include the diffuse Galactic emission and the extragalactic sources. In addition, ground-based observations are affected by two additional foreground-like contributions from Earth’s ionosphere and human-generated radio frequency interference (RFI). In Earth orbit or on the lunar nearside, RFI persists as a major source of signal corruption. In contrast, the lunar far-side offers an RFI-free environment. Above the farside, the thermal emission from the Moon forms an additional foreground, which is still orders of magnitude less than the foregrounds on Earth.

3.1. Earth’s Ionosphere

Earth’s ionosphere corrupts low frequency radio observations (Vedantham et al. 2014; Datta et al. 2014; Rogers et al. 2015; Vedantham & Koopmans 2015; Sokolowski et al. 2015b). Vedantham et al. (2014) modeled the effect of ionospheric refraction and absorption. They considered a static ionosphere using typical values of electron densities in the D and F layers of a mid-latitude ionosphere. Their results show that the residual brightness temperature due to the ionospheric effect is \( \sim 2 - 3 \times T \) orders of magnitude higher than the expected global 21-cm signal (Figure 2).

Earth’s ionosphere exhibits both spatial and temporal variations. The ionospheric variations are directly related to the forcing action by the solar UV and X-ray radiation (Davies 1990), as well as to changes in local plasma conditions from solar wind ram pressure plus B-field, and to variations in the circulation currents in geospace and the resulting variability of the geomagnetic field. The dynamic characteristics of the ionosphere reflect the same scales of solar temporal variations in solar input and the turbulent nature of the solar wind plasma (Liu et al. 2011). Solar activity, including solar radio bursts, exhibits flicker noise or 1/f noise (Polygannakis et al. 2003) which translates to similar non-Gaussian variations in the ionosphere electron density and temperature on timescales of at least 10 hours (Temerin & Kintner 1989; Sokolowski et al. 2015b; Datta et al. 2014). In Datta et al. (2014), a dynamic ionosphere was modeled using GPS measurements of the integrated line-of-sight ionospheric electron densities \((n_e)\) and backscatter radar-based knowledge of the electron temperatures \((T_e)\). This study included the effects of refraction, absorption and emission due to the dynamic night-time mid-latitude ionosphere around the last solar minimum. These dynamic ionospheric effects result in time-variable foregrounds with residual brightness temperatures that are \(10^2 - 10^5\) times higher than the expected global 21-cm signal features (Figure 2).

These ionospheric effects scale as \( \sim \nu^{-2} \), so they are much worse below 100 MHz. The “1/f” noise observed in the ionosphere over timescales of \(\approx 10\) hrs (Sokolowski et al. 2015b) would require hundreds of years of integration to achieve RMS levels of \(\approx 1\) mK needed to fully characterize the expected Turning Points in the global 21-cm spectrum (Datta et al. 2014).

Another recent study by Rogers et al. (2015) used real observations from the EDGES experiment at frequencies between 80 and 185 MHz to derive the ionospheric electron density and temperature. The resultant ionospheric parameter values agree with those used for modeling the effects of the dynamic ionosphere in Datta et al. (2014).

As a result of the severe ionospheric corruption of low-frequency observations on Earth, Datta et al. (2014) concluded that detection of the all-sky 21-cm signal from the Dark Ages and Cosmic Dawn down to \(\nu \approx 10\) MHz would best be done from the farside of the Moon. NASA’s Astrophysics Roadmap\(^3\) effectively came to a similar conclusion in describing a small mission “mapping the Universe’s hydrogen clouds using 21-cm radio wavelengths via a lunar orbiter observing from the farside of the Moon.”

3.2. Radio Frequency Interference

The VHF frequency range 40–120 MHz is heavily used by civil and military transmitters, producing an intense RFI environment on Earth (e.g., McKinley et al. 2013). FM radio transmissions (87.5-110 MHz) are the strongest source of RFI in this band\(^4\). In addition, reflection of human-generated RFI off space debris, aircraft, and ionized meteor trails back to radiometers on the ground presents a potential challenge (e.g., Tingay et al. 2013) from which there is no shielding even in remote regions of the Earth’s surface. Vedantham et al. (2015) calculated that radio Earth-shine from such spacecraft debris may pose limitations to ground-based Cosmic Dawn experiments.

In low Earth orbit, the effective antenna temperature from a single typical FM transmitter (with a broadcast power of 100 kW in a 100 kHz bandwidth) in direct line-of-sight could exceed \(10^9\) K in a 1 MHz spectral channel. In geosynchronous orbit, the effective antenna temperature would be reduced only to \(10^8\) K. Since an Earth-orbiting spacecraft has an entire Earth hemisphere in its potential field of view, the collective effects of all the FM transmitters could make the RFI environment in orbit worse than on the ground.

When observed at the distance of the Moon, a typical FM transmitter has an RFI brightness temperature \(T_{RFI} \sim 10^6\) K; without lunar occultation, a spacecraft would have to be > 4 au from Earth to reduce terrestrial RFI below the desired sensitivity goal for 21-cm experiments (RMS < 5 mK, Figure 2). At the lunar L2 Lagrange point, \(\approx 60,000\) km behind the Moon, the lunar disk does not sufficiently occult the Earth because of diffraction of radio waves around the lunar limb.

In contrast, a lunar-orbiting spacecraft will be shielded from Earth-based RFI while above the lunar farside (Genova et al. 2015). The radio-quiet cone above the farside (including diffraction effects) must be defined by a -90 dB attenuation of the RFI signals from Earth. Inside the radio-quiet cone, the RFI brightness temperature is reduced to < 1 mK. This was verified by both RAE-2 and by Apollo.

---


The “leakage” of emissions from geosynchronous (GEO) satellites observed at the Moon is unimportant because they contribute $T_{RFI} < 10^{-5}$ K at the Moon; reflections of terrestrial RFI from GEO spacecraft with size $\sim 10$ meters produce $T_{RFI} \sim 0.1$ mK. Emissions from lunar orbiting satellites such as LRO and ARTEMIS could be important, although it is unknown if these or other spacecraft will be operating beyond this decade. For spacecraft adhering to the MIL-STD-461F standard, they must be more than 5000 km distant in order to maintain $T_{RFI} \leq 1$ mK. Active orbital phasing of spacecraft with LRO and added orbital time to account for RFI from ARTEMIS (or other lunar orbiting satellites) can be used to maintain the required science integration time free from interference. In this regard, it is worth noting that LRO does not transmit when on the lunar farside, and frequency-dependent RFI from both spacecraft were measured before launch.

3.3. Galaxy/Extragalactic Foregrounds

The extraterrestrial foregrounds have several components (Figure 2). The most important arises from our Galaxy (Shaver et al. 1999) and a sea of extragalactic sources which, at low angular resolution, appear as another diffuse, spectrally-featureless, nonthermal foreground. Electrons at a single energy yield a smooth spectrum; this single-energy spectrum is convolved with the broad distribution of electron energies to yield an even smoother total spectrum (e.g., Bernardi et al. 2015; Petrovic & Oh 2011). The latter work predicts that, for interferometric measurements of the 21-cm power spectrum, the foreground is sufficiently well approximated by a 3rd order log-polynomial.

Current measurements of the foregrounds are consistent with these theoretical predictions. From 100–200 MHz, EDGES does not find foreground spectral structure at levels $> 12$ mK over 10 MHz spectral scales (Rogers 2015). Jelić et al. (2008) also found that a third-order polynomial was sufficient to fit a detailed model of the foregrounds for LOFAR.

Higher orders can easily be accommodated in the fitting with only a slight increase in the uncertainty on the turning point positions. The necessity to increase the complexity of the foreground model may be detected by comparing the Bayesian evidence for the different models, for example using a nested sampling approach, as discussed by Hurker (2015).

The Galactic foreground is also polarized whereas the Cosmic Dawn signal is not. Polarization mismatch and frequency dependencies of the power patterns of the linearly polarized dipole antennas could result in partially linearly polarized Galactic emission producing spectral features. One potential solution is to slowly rotate the 21-cm instrument with respect to the sky to address the spectral effects of Faraday rotation. Combined with the broad antenna power pattern, this approach produces an

Figure 2. Potential low radio frequency foregrounds, compared to the 21-cm signal (note the logarithmic scale, which changes the signal’s appearance compared to Fig. 1). The Galactic spectrum is shown for a typical region away from the Galactic Center, and is split into contributions from synchrotron and free-free radiation. We also show a typical spectrum from averaging over many extragalactic point sources, with a low angular resolution instrument. For a lunar-orbiting instrument, these diffuse foregrounds also reflect off the Moon (Evans 1969) into the backlobe of the antenna; their effective value after attenuation by the ground plane is shown as the dotted gray line. The Moon’s thermal emission at these frequencies arises from cold, uniform subsurface layers, and is similarly attenuated. As an example, the DARE sensitivity curve is shown to illustrate the potential baseline performance after 1000 hours of integration. Lunar dust impacting the surface of a lunar-orbiting spacecraft produces puffs of radio plasma, with a broadband emission spectrum; this is, however, very weak, as shown. For observations from the Earth, the ionosphere corrupts the incident Galactic radiation. The yellow curve shows the residual after subtracting the original galactic spectrum from this corrupted spectrum. The complex spectral residuals due to ionosphere is similar to and degenerate with the 21-cm spectrum, making it more difficult to be detected from the ground.
azimuth-dependent effect from which residual polarized (and lunar) pickup can be identified and removed.

Finally, as seen in Figure 3, the foregrounds are spatially variable. The featureless spectrum and spatial variability of the foregrounds contrast with the Turning Points and spatial uniformity of the 21-cm spectra, making them separable (see Section 4).

### 3.4. Other Foregrounds

For lunar-based observations, foreground emission from the Moon via the instrument’s backlobe will be present. The lunar spectrum comprises (1) thermal emission from a $\approx100$ km subsurface layer (electrical skin depth) that is isothermal and characterized by a single temperature at a given frequency (and lunar depth) (Salisbury & Fernald 1971; Keihm & Langseth 1975) and (2) reflected Galactic emission, requiring a parameter to describe the Moon’s reflectivity (Davis & Rohlf 1964; Vedantham et al. 2015). This background is included in our model results (see Section 4).

The Sun is also a potential and variable foreground. Solar emission at low radio frequencies consists of optically thin thermal bremsstrahlung and coherent plasma emission from flares and coronal mass ejections. The thermal bremsstrahlung is a complex integral over the entire solar corona and will vary over the course of a 27-day rotation of the Sun and the solar cycle, but it is expected to have little frequency structure. Plasma emission from solar bursts is intense and highly structured in frequency and time. The inclusion of daytime observations by focusing on periods with low levels of solar activity and templates for the thermal bremsstrahlung will need further investigation. For the modeling in the next two sections, we will assume observations in lunar orbit when both the Earth and the Sun are eclipsed by the Moon above the farside.

Other processes have a minor effect on the spectrum. Hyperkinetic impacts of dust particles from the interplanetary medium and the lunar exosphere on spacecraft surfaces have been shown to generate radio transients due to ejected plasma (e.g., Meyer-Vernet 1985). But the dust distribution around the Moon (e.g., Stubbs et al. 2010), the capacitance of the spacecraft, and solar wind plasma conditions near the Moon are such that this emission occurs at frequencies well below those of interest for 21-cm Cosmic Dawn experiments (Le Chat et al. 2014).

While transient emission from Jupiter and interactions between Jupiter and Io are among the strongest sources of radio emission in the solar system, this emission occurs below 40 MHz (Panchenko et al. 2013; Cecconi et al. 2012). In the absence of transient emission, Jupiter is a radio source with power $\sim10^{-4}$ that of the Sun ($\approx10$ mK) at 40–120 MHz (Zarka 2004)). It will be modeled as a point source in the signal extraction.

Finally, carbon atoms in cold, diffuse gas produce radio recombination lines (RRRLs; Morabito et al. 2014). These lines are sharp ($\sim10$ kHz wide), but spaced at known intervals of $\sim1$ MHz. So, with $\approx50$ kHz spectral resolution for 21-cm Cosmic Dawn observations, the RRRLs are expected to constitute a negligible fraction of the data and may be discarded.

4. EXTRACTING THE COSMIC 21-CM FEATURES

For the purpose of demonstrating the extraction of the 21-cm signal from the foregrounds, we will assume in Sections 4 and 5 the properties of the DARE science instrument described by Burns et al. (2012) with data collection above the lunar farside. We have continued to update the design of the radiometer to achieve high sensitivity while maintaining a featureless response with minimal systematic biases. The instrument consists of (a) an antenna composed of dual polarization biconical dipoles above a ground plane resulting in a wide bandwidth, spectrally featureless reflection coefficient; (b) a receiver that is thermally controlled with a pseudo-correlation architecture optimized to maintain a spectrally featureless response over a bandwidth of 40 to 120 MHz (corresponding to $35 < z < 11$); and (c) a wide-bandwidth, high fidelity digital spectrometer. The frequency response of this radiometer is shown in Figure 2 with an RMS sensitivity of $\approx5$ mK at 65 MHz.

Observations and modeling demonstrate that the 21-cm Turning Points can be extracted in the presence of these foregrounds. Our MCMC approach has proven heritage in cosmology, including in measurements of the CMB. An earlier version of our signal extraction pipeline was described by Harker et al. (2012), but it has since been extended in power and flexibility (Harker et al. 2015), as we summarize in this section.

#### 4.1. Inputs to the parameter estimation

The endpoint of the instrument calibration process is a series of estimates of the calibrated antenna temperature, $T_{\text{ant}}(\nu)$, which can also be denoted as the result of convolving the antenna power pattern, $B(\theta, \phi, \nu)$, by the sky brightness temperature distribution, $T_{\text{sky}}$,.

$$T_{\text{ant}} = \int d\Omega T_{\text{sky}}(\theta, \phi, \nu) B(\theta - \theta_0, \phi - \phi_0, \nu), \quad (2)$$

where $\{\theta, \phi\}$ are spherical coordinates and $\{\theta_0, \phi_0\}$ label a specific pointing direction and instrumental effects are taken out due to calibration. Following DARE, we will assume that spectra will be collected in a small number of specific pointing directions, yielding a small number of estimates of $T_{\text{ant}}$ representing different areas of sky. The number of independent directions is set by the angular extent of the antenna power pattern depending on frequency. For the analysis that follows, we will assume that we have only four independent measurements of $T_{\text{ant}}(\nu)$, having avoided the brightest regions associated with the galactic center and disk.

The estimates of $T_{\text{ant}}$ which enter the modeling are noisy, so that the final quantity we must model is given by $T_{\text{ant}}(\nu) + T_{\text{N}}(\nu)$, where $T_{\text{N}}(\nu)$ is noise, which is Gaussian under the assumption that there are no systematic errors left over after calibration. The RMS value of $T_{\text{N}}(\nu)$ reduces with time according to the radiometer equation, $T_{\text{RMS}}(\nu) = T_{\text{sys}}(\nu)/\sqrt{\Delta t \Delta \nu}$, where we introduce a ‘system temperature’ $T_{\text{sys}}$ which has contributions from the sky, antenna and receiver. $\Delta t$ is the integration time and $\Delta \nu$ is the width of the frequency channel.

#### 4.2. Parameterizing the model

$T_{\text{sky}}$ is the sum of contributions from the foregrounds and the 21-cm signal. To apply our fitting procedure, we must have parameterized models for the foregrounds and
the signal, so that given a set of inputs we can generate a model $T_{\text{sky}}$, fold this through the instrument response, and compare the result to the data.

Whereas previous studies parameterized the signal using the frequencies and brightness temperatures of the three turning points (Harker et al. 2015), we have found that using a series of $\tanh$ functions to model the Ly-\alpha background intensity history, $J_\alpha(z)$, the IGM thermal history, $T_K(z)$, and ionization history, $x_{\text{HI}}(z)$, to be a more powerful and physically-motivated approach. It recovers the turning points despite the fact that they are not explicit parameters of the model (Harker et al. 2015). A generic $\tanh$ model takes the form

$$A(z) = \frac{A_{\text{ref}}}{2} \{1 + \tanh[(z_0 - z)/\Delta z]\},$$

such that its free parameters are an amplitude, $A_{\text{ref}}$, a redshift, $z_0$, and a duration, $\Delta z$. For physical histories, though, we can fix the $A_{\text{ref}}$ parameter for the ionization history to unity. Similarly, $A_{\text{ref}}$ for the thermal history is fixed at $10^3$ K, though our results are insensitive to this choice.

For the diffuse (Galactic and extragalactic) foregrounds, we use the coefficients $\{T_0^i, a_1^i, a_2^i, \ldots\}$ of a polynomial in $\log(\nu)$–$\log(T)$, i.e. of $\log T_\text{FG} = \log T_0 + \sum_{j=1}^n a_j^i [\log(\nu/\nu_0)]^j$, where $n$ is the order of the polynomial and $i = 1, \ldots, N$ where $i$ labels the sky region and $N$ is the number of sky regions. $\nu_0$ is an arbitrary pivot frequency which we take to be 80 MHz. Meanwhile, the Moon is characterized by a temperature (for its thermal emission) and a reflectivity. However, we are easily able to incorporate alternative models for any of these components (for example, models of the signal based on physical parameters rather than the shape of the signal), or additional foregrounds, if these are required to model the data accurately. Currently, we are assuming that a time-dependent calibration procedure will be applied to the raw-data before being separated in different sky-regions, time-averaged and fed into the signal extraction procedure. However, our current implementation of the signal extraction procedure is fully capable of modeling some residual instrumental effects if needed.

4.3. Constraining the parameters

Given a set of (observed or simulated) noisy spectra, we infer the posterior probability distribution function (PDF) for all the parameters (and therefore, in addition, their best-fitting values and confidence regions) using an MCMC code. This explores parameter space, sampling more densely those regions where the modeled spectra best match the data. In this way, the features of the 21-cm signal, which are broad-band and occupy the entire frequency range, are found by fitting the shape of the spectrum at all frequencies (and across all sky regions) simultaneously.

The sampler we use for this parameter exploration is emcee (Foreman-Mackey et al. 2013), a free, open-source, MIT-licensed Python implementation of the affine-invariant Goodman & Weare (2010) sampler. It also allows for the computation of the Bayesian evidence (via a version of the sampler that implements parallel tempering) in order to select between competing models, as well as fitting the parameters of those models, though at present an accurate estimate of the evidence in this mode is computationally intractable.

Figure 4 illustrates an example of expected results from an end-to-end signal extraction pipeline for 1000 total hours of integration (corresponding to 2 years in lunar orbit shielded from both the Earth and Sun) divided among 4 sky regions. It uses the foreground model sky convolved with the frequency-dependent antenna power pattern and the instrument sensitivity/systematics described in Section 6. The figure shows the fully marginalized posterior probability distributions for the frequencies and brightness temperatures of the turning points along the diagonal of the plot. It also shows the pairwise joint probability distributions for these parameters (contour plots in the interior panels). The resulting constraints on the positions of the turning points will provide physical constraints on the first stars and black holes (Figure 1).

Smoothly varying components of the signal are degenerate with the smooth foregrounds, i.e. a smooth spectrum added to the signal would be treated as part of the foreground. The signal extraction must therefore exploit frequency structure in the signal (though, as noted in Section 3, we also gain some leverage from the fact that the signal is known to be spatially constant at low
Figure 4. The end-to-end extraction process results in measurements of the Turning Points shown above. The interior panels show 2-D PDFs with contours for 68% (blue) and 95% (green) confidence, while 1-D marginalized PDFs appear along the diagonal. Dotted lines denote the input values. The brightness temperature of Turning Point B is excluded from this plot as it is tightly coupled to the frequency of Turning Point B, as required by our parameterization of the global 21-cm signal.

angular resolution, while the foregrounds have spatial structure). Our technique fully captures these degeneracies between parameters, including those relating to the signal, the foregrounds and (potentially) the instrument (see, e.g. Harker et al. 2012), allowing us to thoroughly explore and characterize the data.

5. PHYSICAL PARAMETER ESTIMATION

In the modeling described in this Section, we assume an instrument with an RMS sensitivity of $\approx 5$ mK at 65 MHz. Such a detection will be an important result in and of itself as was the case of COBE measuring a blackbody spectrum for the CMB. Such measurements will provide the first constraints on the nature of the first luminous objects. Additional measurements from the CMB, the 21-cm power spectrum at $> 100$ MHz, and JWST when combined with global signal results will break degeneracies between physical parameters of the first star-forming halos (e.g. Liu & Parsons 2015).

The products of our signal extraction pipeline are measurements of the turning point positions and their associated uncertainties. In translating these constraints to the properties of the IGM, one must consider that (1) it is computationally expensive to generate physical models for the global 21-cm signal, which makes a thorough exploration of parameter space challenging, and (2) degeneracies exist between model parameters of interest, which makes them difficult to constrain independently of one another. In principle, the IGM properties at the redshifts corresponding to the turning points can be constrained in a fairly model-independent (Mirocha et al. 2013) and parameterization-independent (Harker et al. 2015) way, though constraints on the properties of the Universe's first galaxies will of course be model-dependent.

Sophisticated MCMC fits that accurately compute the Bayesian evidence will be able to provide evidence for one parameterization over another, though such exercises are very computationally expensive and beyond the scope of this paper (Harker 2015). For simplicity, we will follow Mirocha et al. (2015) and model the global 21-cm signal using a four-parameter model, which ties the metagalactic Lyman-α, hydrogen-ionizing, and X-ray backgrounds to the growth rate of dark matter halos (e.g. Liu & Parsons 2015).

The free parameters are the efficiency of Ly-α ($\xi_{\text{Ly},\alpha}$), ionizing ($\xi_{\text{ion}}$), and X-ray ($\xi_X$) photon production per unit star formation, and the minimum virial temperature of star-forming halos ($T_{\text{min}}$). Cosmological parameters are held fixed, as is the mass function of dark matter halos,
The 21-cm Global Signal

Figure 5. Constraints on the parameters of a simple galaxy formation model. From left to right, these include the efficiency of X-ray production ($\xi_X$), the efficiency of photon production in the Lyman-Werner band ($\xi_{LW}$), ionizing photon production ($\xi_{ion}$), and the minimum virial temperature of star-forming halos ($T_{min}$). Blue regions in the interior panels correspond to 68% confidence regions, while green regions denote 95% confidence regions. Dotted lines mark the input values.

which we assume is the Sheth & Tormen (1999) form.

Fitting the above model to the Turning Point constraints illustrated in Figure 4 yields the constraints shown in Figure 5. All parameters of our physical model are constrained to $\sim 0.1 - 0.2$ dex about their input values. Thus, the precision of these observations will place significant and meaningful bounds on the characteristics (i.e., spectra and mass) of the first stars and galaxies.

6. CONCLUDING REMARKS ON THE POTENTIAL OF 21-CM GLOBAL SIGNAL OBSERVATIONS

To achieve the science potential of 21-cm global spectral observations in the presence of multiple sources of foregrounds, careful consideration must be given to the local environment, the instrument, and methods for signal extraction. In this paper, we described how a lunar-orbiting 21-cm science instrument eliminates the brightest foregrounds present on the Earth’s surface, namely the ionosphere and human-generated radio frequency interference. Such observations have the best chance of success to detect the theoretically-predicted broad features in the 21-cm monopole in the presence of extraterrestrial foregrounds, primarily from our Galaxy, the “haze” of extragalactic sources, and thermal emission from the Moon.

Signal extraction in the presence of bright foregrounds is the greatest challenge for all observations of the 21-cm cosmological spectrum. Utilizing a Markov Chain Monte Carlo numerical inference technique, we showed that it is possible to accurately recover the frequencies and brightness temperatures of the the Turning Points in the spectrum in the presence of bright foregrounds with the instrument characteristics of the Dark Ages Radio Explorer (DARE) for $\approx 1000$ hrs of integration. This is in part because of the improved $tanh$ parameterization of the global 21-cm signal, which can capture the shape of physical models of the global 21-cm signal more accurately than previous approaches.

We also simulated the global 21-cm spectrum using a four-parameter model consisting of backgrounds from Ly-$\alpha$, Ly continuum, and X-rays along with the growth rate of star-forming halos. Again using an MCMC approach applied to a simulated DARE spectrum, we are able to constrain these parameters to 0.1-0.2 dex relative to their input values. Thus, these models further demonstrate the potential capability of the 21-cm global signal to quantitatively characterize the properties of the first generation of stars, black holes, and galaxies in the early
Universe.

Accurate parameter estimation is a core capability required for 21-cm global signal observations and interpretations. Bayesian methods have proven successful for similar CMB observations which also suffer from foregrounds. MCMC methods have been shown to accurately recover spectral and physical parameters from noisy spectra. The next step for analyses of the global 21-cm spectrum is to compute the Bayesian evidence that will help us to distinguish between different levels of structure in the beam-convolved foreground and also to distinguish between different physical models of the first halos. This is a highly computationally intensive process. However, recent developments of Nested Sampling algorithms for high dimensional parameter space which operate in massively parallel computer architectures (Handley et al. 2015) have great potential for 21-cm cosmology applications.

In conclusion, measurements of spectral features in the 21-cm spectrum will answer key science questions from the NRC Astrophysics Decadal Survey: “What were the first objects to light up the Universe and when did they do it?” With a clean measurement within the radio-quiet environs of the Moon’s farside and proven technology, the 21-cm global signal will open a new window of discovery into the early Universe.

This research was supported by the NASA Ames Research Center via Cooperative Agreements NNA09DB30A and NN1X5AD20A to JB, and by NASA ATP grant NN1X5AK80G to SF and JB. Part of this research was conducted at the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration. J.M. acknowledges support through the NASA Earth and Space Science Fellowship program, grant number NN1X4AN79H. G.J.A.H is supported through the People Program (Marie Curie Actions) of the European Union’s Seventh Framework Program (FP7/2007-2013) under REA grant agreement n. 327999. SF was partially supported by a Simons Fellowship in Theoretical Physics; SF also thanks the Observatories of the Carnegie Institute of Washington for hospitality while much of this work was completed. We acknowledge Ball Aerospace and Technologies Corporation for their partnership in this research. We thank the following colleagues for their important contributions including Jill Bauman, Butler Hine, Robert Hanel, Stephanie Morse, Bill Purcell, William Deininger, John Jonaitis, Karen Lee, Damon Russell, Sharon Kedar, Bang Nhan, Ken Galal, Andres Donoperez, Fan Yang, Anthony Genova, Jay Bookbinder, Scott Mitchell, David Ruppel, Pete Worden, and Eugene Tu. We also benefited from helpful conversations with John Mather, Harry Partridge, Sandy Weinreb, Aaron Parsons, Jonathan Pober, Jim Condon, Greg Taylor, Adam Lidz, Greg Schmidt, and Robert Jarnot.

REFERENCES
Boorman, J. D., & Rogers, A. E. E. 2010, Nature, 468, 796
Davies, K. 1990, Ionospheric radio No. 31 (IET)
Field, G. B. 1958, Proceedings of the IRE, 46, 240
Goodman, J., & Weare, J. 2010, Communications in Applied Mathematics and Computational Science, 5, 65
Keliih, S. J., & Langseth, M. G. 1975, Icarus, 24, 211
Meyer-Vernet, N. 1985, Advances in Space Research, 5, 37
Patra, N., Subrahmanyan, R., Raghunathan, A., & Udaya Shankar, N. 2013, Experimental Astronomy, 36, 319
The 21-cm Global Signal


Sokolowski, M., Tremblay, S. E., Wayth, R. B., et al. 2015a, PASA, 32, 4


Temerin, M., & Kintner, P. M. 1989, Washington DC American Geophysical Union Geophysical Monograph Series, 53, 65


