

# Chapter 1

## Instrumentation for Zero-mode Signal Detection and Experiments

### 1.1 Introduction to zero-mode experiments–(LJG)

- names, locations, citations
- very concise / statement of what above citations focus on

#### **Radiometer basics**

Karl Jansky discovered in August 1931 that any sensor of the electromagnetic field placed in an open field will provide a sample of cosmic radio noise at its terminals, along with other local and terrestrial radio frequency interference. Electrical engineers call electromagnetic sensors antennas, and the simplest form of a system that may detect the zero mode of the radio sky is a single element antenna placed in an open field and followed by a detector of the received power. Such a basic radio telescope is often referred to as a radiometer, and usually has an antenna element followed by a low-noise amplifier, filters that define the bandwidth, and a detector that provides a measure of the power received by the antenna. Of course, the received power in an Earth-bound radiometer will have unwanted additives from terrestrial man-made interference, radio emission from the ground in the vicinity of the antenna, additive radio power from elements of the radiometer signal path, which are all detected along with the cosmic radio power and the accuracy in the detection of the sky radio emission will depend on the accuracy of the component separation and also the gain calibration that provides meaningful units to the received power in terms of the sky radio brightness.

*itemizeTheElectromagneticsensororantenna*

The sensor in any radiometer is in its simplest form usually a pair of conducting elements ending in a pair of terminals. When placed above ground, and considered as a receiving antenna, the incident electromagnetic waves from the sky above as well as from the ground beneath induce currents in the pair of conducting elements, which result in a voltage across the terminals. Across frequencies in the radio window there is a changing intensity in the sky brightness and also of the ground emission and hence the currents and voltage at the antenna terminals would have a spectral characteristic that represents the spectra of the incident fields

and also the transfer function of the antenna. The spatial arrangement of the conducting elements are usually purpose designed to tailor the transfer function and efficiency of the coupling to sky and also to make the coupling to ground inefficient. Considered as a radiator, voltages impressed upon the antenna terminals drive currents in the conducting elements, which result in the energy propagating away as electromagnetic waves. Alternately, the antenna may be viewed as a transformer that attempts to provide a matching between the impedance of free space on one side to the impedance of the transmission line connected to the antenna terminals.

A pair of quarter wavelength conducting elements adjacent to each other and in line forms what is commonly referred to as a dipole antenna; this is tuned to most efficiently transduce electromagnetic waves with wavelength twice the total length of the dipole. Linear dipoles receive a single linear polarisation mode oriented along the dipole. Helical conducting elements of helical antennas efficiently transduce electromagnetic waves with wavelength the circumference of the helix, and these receive a single circular polarisation mode. Any antenna can indeed receive or radiate in only a single polarisation mode, linear, circular or elliptic, represented by a point on the Poincare sphere. Thus sensing of the total incident electromagnetic field, which may be decomposed into a pair of orthogonal polarisation modes, would necessarily require a pair of radiometers with antennas designed to transduce orthogonal polarisations.

Linear dipoles and helical antennas are tuned to specific frequencies and are hence intrinsically narrow band sensors with performance degrading substantially away from the design frequency. CD/EoR signals are, however, wideband signals, requiring at least an octave bandwidth sensor to capture spectral structure that is sufficiently complex relative to the relatively smoother foregrounds, so that appropriate signal processing algorithms may be able to separate the components mixed in at the antenna terminals. Thus antenna design for CD/EoR arranges the conducting elements so that the antenna performs efficiently and usefully over at least an octave band. Dipole antenna elements are often made in 2D flat shapes and even 3D fat dipoles of shapes such as conical elements or more complex structures. Circularly polarised antenna elements are often planar spirals. These design evolutions from simple linear dipoles and helices attempt to encompass a range of wavelengths within the structures thus widening the structural bandwidth and hence including a significant part of the total global CD/EoR spectrum in its terminal voltage signal.

Arrangement of the conducting elements in a horizontal plane above ground, either as planar dipoles or planar spirals, even if the elements are fat, yields antenna designs that receive equally efficiently the radiation incident from the sky above and also from the ground below. The conductors are often designed to deviate from such symmetry, by drooping the arms of the dipole, or by extending the spiral structures into conical spirals, with a design goal of reducing efficiency to ground while enhancing that to the sky. Of course, in such antenna placed on or above real ground, the ground radiation is indeed incident on the elements along with the sky radiation and the preference is attained by designing the element shapes to have a variation in distance to ground so that the net ground induced voltage at the antenna terminals is preferentially canceled and simultaneously the sky induced signal is constructively enhanced. Such structural deformations enhance directivity of the antenna towards sky or, in other words, limit the field of view of the antenna to the sky.

Sensitivity to ground emission may be substantially reduced by covering the earth be-

neath the antenna by a conducting plane. Such a ground screen would then act as a reflector and hence the antenna would "see" the sky directly and also reflected on the ground screen. Thus every sky point would have two paths to the antenna and the net response towards any sky direction would be the vector sum of the fields induced via the two paths. Ground screens may be soldered or welded wire meshes, with mesh size a small fraction of the wavelengths of interest, so that the transparency of the ground screen and hence leakage of ground radiation through the ground screen would be small. If the antenna is of single linear polarisation, it is sufficient for the ground screen to be made of wires with orientation along the antenna polarisation as projected on the ground. Ground screens can be large but usually of finite size, and then there may be additional paths between antenna and sky via scattering off the edges of the ground screen. With the aim of randomising this scattered component, the edges of the ground screen are often constructed to be serrated, with the dimensions of the serrations significant compared to the wavelengths of operation.

As aspect of antennas is that there is a space around that is referred to as the "near field" region, where the antenna stores reactive energy when viewed as a radiator, and beyond which the field is radiative "far field".

Just as antennas are frequency selective, so also are antenna selective in their direction sensitivity. I know of no practical antenna that is isotropic. The relative sensitivity of an antenna across sky directions is referred to as the far-field antenna radiation pattern or beam pattern. Any antenna is directional and is selective is receiving radiation. A quarter wave dipole antenna is insensitive to radiation incident along its length: after all, radiation incident along that direction would have fields orthogonal to the dipole arms and hence cannot possibly induce currents. The quarter wave dipole antenna efficiently transduces signals in directions orthogonal to its orientation: quarter wave dipole antennas have toroidal beam patterns; dipole antennas are omni-directional. The larger an antenna is in its electrical size, the more directional is its response to sky radiation. Antennas with electrical size exceeding unity would naturally have beam patterns that have a main lobe that defines a sky solid angle where sensitivity is maximum, and sidelobes of lower sensitivity. Electrically small antennas would have a single main lobe without sidelobes.

Antennas are usually passive systems, hence reciprocal, and may be viewed as radiators or sensors, transforming incident power to signal power at its terminals or transforming power fed at the terminals to free-space radiation. The antenna as a transformer provides a transformed free-space impedance at its terminals, which is often referred to as antenna impedance. Viewed as a radiator, power fed to the antenna via a transmission line connected to the antenna terminals is only partly available for radiation, because depending on the impedance mis-match between the transmission line and the antenna a part of the power fed to the antenna is reflected back. Of the power available for radiation, a part of the power flows to sky via the main and side lobes of the radiation pattern, a part to ground via what might be considered to be back lobes and a part might be lost as ohmic loss that heats the antenna, if the conducting antenna elements are resistive. Reflection efficiency of the antenna defines the fractional power available, and radiation efficiency defines what fraction of the available power is radiated into free space. The antenna views the sky multiplied by the radiation pattern; the weighted average brightness of the sky, with a weighting defined by the radiation pattern, is the sky noise temperature in so far as the antenna is concerned. The radiation efficiency times this temperature is the sky noise temperature component at the

antenna terminals, and the reflection efficiency times that terminal noise temperature is the sky noise power propagating down the transmission line.

In the context of radiation efficiency, it may also be noted here that if the antenna elements are resistive and there is loss of cosmic signal as ohmic loss in the antenna, not only is there a loss of desired signal but the resistive elements also add their own thermal radio noise to the sky signal.

#### Analog Signal Processor or Receiver

A radio astronomy antenna receiving incident sky radiation provides a voltage waveform at its terminals that is a selective sampling of the electromagnetic field incident at the location of the antenna. The field arises from a multitude of independent atomic processes or radiation mechanisms in the sky and on the ground. The electric field is not a constant in time, it is time varying, the field is a random variable with zero mean, Central Limit theorem argues for the field having a normal (Gaussian) probability distribution. Man-made radio communication and collateral damage to the radio frequency environment from inadequate shielding of electrical and electronic equipment most often result in non-Gaussian signals; however, sensitive radio astronomy receivers are usually deployed in extremely radio quiet sites where the total radio power at the antenna terminals from such unwanted components is sub-dominant to the total radio power received from the celestial sphere.

The fluctuating electric field results in fluctuating currents in the antenna elements and hence a fluctuating voltage at its terminals, which propagates down the transmission line connected to the antenna. The voltage in the transmission line, due to cosmic radio noise power, is extremely small and hence needs to be amplified before any detector or sampler may make a useful measurement of the spectral power. Any analog amplifier inevitably adds its own noise to the incoming signal, apart from amplifying the signal. The degradation in the signal to noise, due to this additional noise originating in amplifiers, will obviously be less if the signal levels are larger. Therefore, most of the degradation in signal to noise occurs in the first amplifier encountered by the cosmic signal, and it is desirable that this first amplifier in the signal path be a low-noise amplifier: an amplifier specially designed to add less noise.

Just as resistive elements in the antenna add their own thermal noise to the signal, so also do all resistive elements that appear in the signal path in the receiver chain add their thermal noise contribution to the signal, thus degrading the signal to noise. The equivalent noise temperature of the signal progressively increases down the signal path with amplification and the equivalent noise temperature of the thermal noise added by resistive attenuations is a fraction of its physical temperature; therefore, the degradation in signal to noise is of less significance as the signal is amplified.

In the analog receiver chain the cosmic signal, which is essentially a voltage of Gaussian random noise form, is amplified. The amplification most often is done in stages. The total gain is selected so that the amplitude at the end of the chain is sufficient for the analog to digital converter that samples the signal. If the equivalent noise temperature of the cosmic noise at the terminals of the antenna is less than the equivalent noise added by the first stage low-noise amplifier, then the signal to noise in the voltage waveform in the signal path will be less than unity. In this case the power in the voltage waveform as it arrives at the end of the amplification stages will be dominated by noise power added within the receiver. Advances in technology and developments towards making low-noise amplifiers that add

lower and lower equivalent noise aim towards ensuring that in radiometers the cosmic radio noise dominates at all stages within the signal path.

It may be also noted here that amplifier noise as also the thermal noise from attenuations in the signal path propagate downstream and also upstream. The noise wave components propagating in the two directions are partially correlated. These components, as also the cosmic noise, are partially reflected and partially transmitted at every impedance discontinuity they encounter. The design of analog receiver chains attempts to minimise impedance discontinuities and to attenuate or cancel any signals that suffer back reflection and provide isolation and prevent back propagation. Attenuators are passive and reciprocal and provide equal attenuation to signals propagating both ways. Amplifiers provide voltage gain for signals propagating in the forward direction; for signals arriving at their outputs, they provide a gain loss or reverse isolation and a diminished version of the voltage waveforms at the amplifier output appear at the input and propagating in the reverse direction. Amplifier reverse isolation are almost always greater than the forward gain, so that any reverse propagating signal always suffers a net loss on going across the amplifier backwards and forwards again. Any chain of amplifiers and attenuators in a radiometer will have multi-path propagation of the input cosmic noise and noise added in amplifier and attenuator devices between their source and the final output of the receiver chain. Vector addition of wideband coherent signal components that arrive at the output of the receiver chain, with multi-path propagation within the chain, will result in net voltages that vary with frequency since the path differences between adding components would cause different frequency components to suffer different phase shifts.

An important function of the analog receiver, apart from voltage amplification, is to limit the band. There is inevitably a limiting of the band by the antenna, which usually has a design frequency, and is optimised in structure and in matching of its impedance to free space and to the connecting transmission line over a bandwidth. Amplifiers are also designed to have voltage gain over a certain bandwidth, and their gains fall off on either side thus limiting the band in the signal path. Apart from these, radiometers usually have filters in their analog signal paths. Most spectral radiometers have digital spectrometers to measure the spectrum of the cosmic noise, and digital spectrometers act on digital signals, hence requiring an analog to digital converter. Digital signals are a sequence of samples, and for sampling without loss of information it is essential that the samples are at Nyquist rate, which is twice the maximum frequency present in the signal presented to the sampler. This requires that the analog signal is band limited and that signal components beyond the upper limit of the low-pass or band-pass filter and adequately rejected, depending on the science requirement. This is done by the analog filters. The cosmic noise is often also high-pass filtered, so that the total band is defined in a bandpass for the science. The signals may be band limited by bandpass filters. Filter designs are usually either of Butterworth approximations, which have a maximally flat pass band shape, or of Chebyshev type, which provide sharper transitions from pass band to stop band by allowing for ripples in the pass or stop bands, or of Elliptic function approximations that allow for ripples in the pass and stop bands and provide the performance meeting a design goal of pass band ripple and stop band rejection with an approximation of lowest order. Often a combination of filters is used - one to provide excellent rejection in the stop band but with relatively slower rolloff in response, and others with sharp transitions at the design edges of the pass band.

### Digital Spectrometer

Digital signal processing has replaced analog filter banks and analog signal processing methods in most spectral radiometers.

Digital signal processing requires that the continuous or analog signal voltage waveform be converted into a digital signal, which is essentially a stream of numbers. The analog signal is sampled at a sampling rate, and converted to numbers in a quantiser. Improving technology provides devices with higher sampling rates and with ability to represent the samples with greater precision using larger word lengths. Often there is a trade off in that the fastest samplers available are those with fewer bits representing the samples. In other words, devices with the highest sampling rates have greater quantisation errors.

Quantisation with finite word length results in a non-linear distortion of the signal and hence a limit to the dynamic range of the measured spectrum. Additionally, quantisation results in addition of what is called quantisation noise with amplitude depending on the magnitude of the least significant bit. Analog to digital converters with larger numbers of bits are always preferred when higher dynamic range is required; in particular, when cosmic noise is sampled along with man-made radio frequency interference that is almost always non-Gaussian. However, quantisers also suffer from subtle errors arising from inaccuracies in its transfer function; therefore, simply having larger word lengths is not sufficient, limitations also come from inaccuracies in the representation.

The analog band is usually sampled at the Nyquist rate, which is twice the analog bandwidth. Higher sampling rates are sometimes adopted, referred to as oversampling, in an attempt to improve dynamic range. To some extent, oversampling compensates for finite word length in representation of sampled data.

Digital spectrometers in spectral radiometers are implemented following essentially two possible paths. The power spectrum of a signal may be computed by first Fourier transforming the time sequence and then squaring the complex coefficients of the discrete Fourier transform. Alternately, an auto-correlation of the time sequence may be computed for a range of time delays and then the discrete auto-correlation estimates may be Fourier transformed to get the power spectrum. The first method is often called the FX approach and the second is called the XF approach.

In the FX approach, the length of time sequence that is discrete Fourier transformed decides the spacing between Fourier components in the transform domain. If we assume that the analog signal has bandwidth  $B$ , and is sampled at Nyquist rate  $2B$ , and that blocks of  $N$  samples are discrete Fourier transformed, then in the transform domain there will be  $N/2$  samples of the spectrum over the band with a spacing  $2B/N$ . Similarly, in the XF approach, if  $N$  samples of the auto-correlation function are evaluated and these are discrete Fourier transformed, once again there will be  $N/2$  samples of the spectrum over the band with a spacing  $2B/N$ .

Let the total length of the time-domain sequence be  $T$ , where  $T = N/(2B)$ . If all of the  $N$  time domain measurements are used with equal or uniform weights, then in the transform domain the point spread function that defines the resolution in the spectrum is defined as the discrete Fourier transform of a rectangular function with time-domain extent  $T$ . This point spread function has a main lobe with width twice the channel spacing. Additionally, since the transforms of rectangular functions have large sidelobes, there will be significant leakage of spectral power across spectral channels. Windowing or multiplying the time domain by

an appropriate window function or, in other words, apodising the time domain sequence, is a method of reducing the sidelobes of the frequency-domain point spread function. However, this results in a widening of the main lobe and hence the spectral resolution.

In FX type digital spectrometers that employ windowed discrete Fourier transforms for channellisation, the design usually provides for significant overlap between successive time segments so that the sensitivity loss due to windowing is compensated. Loss in spectral resolution due to windowing is compensated by corresponding increase in the length of the time sequence processed in each transform. Another method that is being adopted in digital spectrometers is adoption of what is called poly-phase filter banks; this is an algorithm that processes longer time sequences in its internal memory to obtain point spread functions with low sidelobes.

Most FX digital spectrometers, which were a few decades ago implemented using discrete high-speed digital integrated circuits, are implemented today in Field programmable gate arrays, or FPGAs. The algorithms are now implemented in FPGAs as firmware, which allows for reconfiguration. With recent substantial increase in processing speeds of graphics processing units, or GPUs, there are implementations in which radio astronomy signals are routed into GPUs for digital signal processing and channellisation. Often the processing is today split between an FPGA followed by GPUs, with divisions optimised for processing accuracy and power consumption.

At the highest bandwidths that are beyond the sampling speeds of available analog to digital converters, signal processing for channellisation is even today done in analog domain. Filter banks may be implemented as an array of analog filters. Alternately the discrete auto-correlation function may be computed in real time over a range of time lags using a delay line with multiple taps and the correlations computed in Gilbert cell multipliers followed by analog power measurement. In such an implementation, as in the case of the XF spectrometer, the discrete Fourier transform is usually computed offline in a general purpose machine.

#### **Design considerations for special purpose radiometers for the zero-mode 21-cm signal**

Following cosmological recombination of the primeval plasma, the primordial gas is almost completely neutral and a significant fraction is hydrogen. Hydrogen has a hyperfine spin flip transition with rest frequency 1420.4 MHz and rest wavelength 21.106 cm. The ratio of populations in the upper triplet state and lower singlet state of this transition is described by the hydrogen spin temperature. The cosmic microwave background (CMB) streams freely through the gas following recombination because the free electron density is extremely small; thus we observe the CMB as it propagates to us through the neutral gas. If the spin temperature of the neutral hydrogen at any redshift exceeds the CMB temperature at that redshift, then the 21-cm emission from the gas adds to the CMB and will be observed at the present epoch at the corresponding redshifted frequency to have a positive distortion. On the other hand, if the spin temperature of the neutral hydrogen at any redshift is lower than the CMB temperature at that redshift, then 21-cm absorption by the gas decrements the CMB and will be observed at the present epoch at the corresponding redshifted frequency to have a negative distortion.

The Dark Ages signal in redshifted 21-cm zero-mode in standard cosmology is well defined by the cosmological parameters and the matter power spectrum to be a broad absorption

feature between about redshift of 30 to 300, corresponding to observing frequencies of about 5 to 45 MHz. This absorption is expected to be at peak about 40 mK. Cosmic Dawn and the light from the first stars competes in coupling the spin temperature to the cold gas kinetic temperature and also in heating the kinetic temperature to be raised above the ambient CMB temperature. The uncertain thermal history at Cosmic Dawn leads to a family of possible 21-cm distortion signals: efficient coupling to cold gas might result in absorption as deep as a few hundred mK, and efficient heating might result in a positive CMB distortion of magnitude at most about 30 mK. The 21-cm CMB distortions would almost certainly be absent at redshifts below 6, corresponding to observing frequencies above 200 MHz, since the intergalactic gas is almost certainly reionised by that epoch.

Thus a radiometer for detecting interesting physics during the formation of the first stars might have a focus on redshifted 21-cm during Cosmic Dawn and subsequent Epoch of Reionization that result in CMB distortions over the observing frequency band of 40 to 200 MHz, which are relatively uncertain. Having said that, the 21-cm absorption of the CMB during the Dark Ages is also a signal worth pursuing, at least to confirm the standard models and rule out several exotic physics that might cause the thermal evolution to deviate from standard expectations.

Detection of the 21-cm zero-mode CMB distortions from Dark Ages, Cosmic Dawn and Reionization requires a radiometer that ideally covers the entire band of 5 to 200 MHz with a single sensor and associated radiometer receiver. A single wide-band spectral observation of the all-sky averaged cosmic radio noise, calibrated for the transfer function of the antenna and receiver electronics, would provide the best separation between foregrounds and any cosmological 21-cm signal. As the radiometer band is reduced, it becomes increasingly difficult to separate between foregrounds and at least the 21-cm signal predictions from standard cosmology that does not admit exotic physics. The confusion results in reduced sensitivity to the 21-cm signals; therefore, radiometers with reduced bandwidths would require observing the sky spectrum with lower measurement noise and hence their associated calibration tolerance is also correspondingly tighter.

The following subsections examine in detail the design considerations specific to the various components of the radiometer.

- Antenna effective area for detection of the zero-mode 21-cm signal

This collecting area in radio telescopes is often increased by arraying sensor elements and combining their terminal voltages in phase, or by using shaped reflectors to concentrate the incident field in a small area where the sensor elements might be placed to transduce the coherently integrated cosmic noise into a voltage waveform. The antenna beam has a main lobe whose solid angle is inversely proportional to the effective collecting area and increasing the area in an attempt to enhance the received faint cosmic noise reduces the main lobe solid angle. This strategy that is appropriately adopted for detection of weak celestial sources is pertinent for sources that are unresolved by the telescope beam.

Viewed as a transformer, the antenna presents at its terminals an equivalent noise power corresponding to the mean brightness temperature of the celestial radio sky weighted by the telescope beam. Once the source extent on the sky exceeds the telescope beam, the antenna temperature (or the power in the cosmic noise as measured at the antenna



terminals) will simply saturate to the brightness temperature of the source. Since the zero-mode of the 21-cm signal is of extent  $4\pi$  steradians, which will always exceed the telescope beam, the antenna terminals will always present equivalent noise corresponding to the zero-mode regardless of the telescope beam and hence the effective collecting area. In other words, the antenna for the zero-mode may be of any arbitrary collecting area and increasing this area does not facilitate detection of the zero-mode.

- Mode coupling of spatial structure to spectral structure

Where as the zero-mode of the cosmological 21-cm signal is a uniform signal component over the sky, it may be noted here that the radio sky has spatial structures on various angular scales. The largest features on the sky are of the Galaxy, the Galactic plane and loops and spurs. H-ii regions and supernova remnants form intermediate scale structures. Compact sources in the Galaxy are more numerous in the Galactic plane. Extragalactic sources have an isotropic distribution, with some clustering, and with distributions in angular scales, in flux density and spectral indices. If the zero-mode telescope antenna beam is achromatic, and has identical beam patterns including main lobe profile and all sidelobe structure across all frequencies in the observing band, then the same celestial sources are averaged with same weighting at all frequencies. In this case the frequency spectrum of the antenna temperature will be the spectrum of the beam weighted sky and sources therein. Of course, as the achromatic beam scans the sky, as would happen if the sky were to drift across the beam, this spectrum of the antenna temperature would vary as different parts of the sky move in and out of the beam. After all, different sky regions and different sources in the sky do have somewhat different spectra.

A key design goal for the antenna is avoidance of what is termed "mode coupling". Mode coupling arises when the telescope beam is chromatic and varies across the observing band. Then different sky structures and sources are "visible" to the telescope at different frequencies and then the spectrum of the antenna temperature will not only be a result of a weighted averaging over sky sources but also a result of sky structure. With a chromatic beam, the antenna temperature will have a frequency structure even if the sky were of constant spectral index; in this case the frequency structure will reflect the sky spatial structure and the chromaticity of the beam pattern. Thus a chromatic beam results in a coupling of sky spatial structure into frequency spectral structure and hence this effect is termed mode coupling.

A radiometer for detection of the zero-mode 21-cm signal that views the celestial sky via its telescope beam will inevitably detect the Galactic and extragalactic foregrounds along with any 21-cm signal. Algorithms for detection of the 21-cm signal in the observed spectra will necessarily have to perform a separation of the foregrounds, and have a model for the expected spectral behaviour of the foregrounds, which will involve relevant radiation processes and source properties and statistics. Mode coupling introduces an added avoidable complexity into the analysis, which inevitably compromises detection of the 21-cm signal. Therefore, a key design goal for the zero-mode antenna is to make it achromatic within the observing band. The class of antennas that provide achromatic performance characteristics are what are called frequency in-

dependent antennas.

Aperture arrays have a fixed physical size and hence are usually highly chromatic in their beams; aperture arrays are then unsuitable for radiometer detection of the zero-mode 21 cm. Antennas with reflector optics may be made achromatic by deploying feeds that illuminate the reflectors over areas that are constant in wavelength units. Hence the effective areas are a constant in wavelength units and the physical effective area scales linearly with frequency, being larger at larger wavelengths.

- Dipole antennas

Most radiometer designs for the zero-mode 21 cm deploy single element antennas without reflector optics. The simplest forms are wideband dipole antennas, which may have arms that are planar 2D structures or 3D fat dipoles. Dipole antennas have omnidirectional beam patterns that may be toroidal at their design frequency; however, the patterns usually deviate somewhat away from a central design frequency. The beam patterns of the dipole antenna may be made achromatic in a band if the antenna is made electrically small and the arms of the dipole are significantly less than quarter wavelength at the highest frequency of operation. In that case the beam pattern tends towards being toroidal and with a single lobe of cosine square form cross-section. Dipoles are usually placed above a conductive and hence reflective ground plane, which covers the real earth below and in the vicinity of a horizontal dipole. A dipole placed above a reflector is inevitably achromatic, since the ground is a fixed physical distance below and cosmic radiation incident from any sky direction would have two paths to the antenna, a direct ray and another reflected off the ground screen, and the two arrive with a path difference and hence a phase difference that linearly varies with frequency. There are also additional paths via scattering off the edges of the ground screen. The more distant the edge the lesser is this effect; however, the greater distance to the edge results in greater phase variations across the band for the multi-path propagation. The beam patterns of dipoles above ground screens are usually designed to somewhat avoid these effects by shaping the dipole arms off the horizontal plane as, for example, in the case of "droopy" dipoles. Another design strategy is to serrate the edges of the ground screens and hence randomise the scattering. Nevertheless, considering how faint the zero-mode 21-cm signal is relative to the Galactic and extragalactic foreground, mode coupling has the potential to confuse the cosmological signal and hence dipoles over ground screens require corrections for mode coupling via modelling this unwanted spectral structure and correcting data. The modelling requires excellent knowledge of the beam patterns, the chromaticity in the beam and a model for the sky structure including its frequency dependence.

An alternate design strategy is to deploy an absorbing ferrite tile ground plane below the dipole antenna, which avoids reflections and hence any multi-path propagation between sky and antenna.

- Monopole antennas

An alternate design for the elemental antenna for zero-mode 21-cm signal is a monopole antenna. This structure has a vertical element, which is often shaped, and is above a

ground plane. As in the case of dipoles, the monopole beam is achromatic if it is electrically small and the vertical element is significantly smaller than quarter wavelength at the highest frequency of operation. The beam pattern of a short monopole above ground plane depends on the extent of the ground plane. If finite, then the omnidirectional pattern has a null towards zenith, null towards horizon, and a maximum at an elevation angle that depends on the extent of the ground plane. If the ground plane is infinite, the maximum is towards horizon, and as the extent is reduced the maximum lifts in elevation to a maximum of about 30 degrees. Therefore an electrically short monopole with an electrically short ground plane has a frequency independent beam pattern whose peak is at some angle about 30 degrees above the horizon.

- Antenna polarization

Dipoles that are oriented horizontally receive horizontal polarisation. A crossed pair of horizontal dipoles would be sensitive to orthogonal polarisations for a wave incident from zenith, where the dipole beam is a maximum. Since the Galactic and extragalactic foregrounds are composed of sources that often have significant fractions of linearly polarised emission, the foreground spectrum component detected in zero-mode radiometers will be polarisation dependent. Complications arise because of Faraday rotation of the polarisation orientation on the sky plane, either during propagation in the interstellar medium or during passage through the Earth's ionosphere. Faraday rotation is wavelength dependent and hence for a linearly polarised source, the source intensity received in any linearly polarised antenna will be frequency dependent. Thus Faraday rotation results in spectral structure that may potentially confuse detection of zero-mode 21-cm spectral structure. For this reason it is desirable and a design goal for zero-mode radiometers to be dual polarised pair of radiometers, with full polarisation calibration that allows for the Stokes I component of the sky cosmic noise to be derived.

- Antenna Reflection Efficiency

The antenna reflection efficiency  $(1 - \Gamma^2)$ , which is related to the voltage reflection coefficient  $\Gamma$  at the antenna terminals, determines what fraction of cosmic noise received by the antenna propagates into the receiver chain. In this consideration, it is the impedance of the antenna at its terminals, which is effectively the free space impedance transformed by the antenna to its terminals, as compared to the impedance of the first low noise amplifier encountered by the cosmic noise as transformed by the interconnecting transmission line to the antenna terminals, that decides the reflection coefficient  $\Gamma$ .

A design goal for zero-mode 21 cm is an antenna that has high reflection efficiency over the full observing band. However, since the foreground Galactic sky has a brightness temperature that is significantly greater than the noise temperatures of modern low noise amplifiers operating in the 10-200 MHz band, it is sufficient that the total efficiency of the antenna provide an antenna temperature that well exceeds the receiver noise. In that case, the system temperature and hence the measurement noise for any integration time would be independent of the receiver noise and improving the total

efficiency would not improve detection sensitivity or reduce the required observing time.

What is probably of greater importance is that the reflection efficiency be a smooth function of low order so that the product of the relatively bright foreground sky with the reflection efficiency, to give the dominant unwanted component of the observed spectrum, does not confuse the desired zero-mode 21 cm signal, and is separable from the 21-cm signal. If the antenna structure is electrically long, as would be the case, for example, in frequency independent spiral antennas with large structural bandwidth, the reflection efficiency would have fine structure in frequency. Therefore, from the viewpoint of designing the antenna element to have reflection efficiency that is exclusively of low order, it is advantageous to have electrically small antennas.

If the antenna does not have resistive elements, and the radiation efficiency is unity, then a measurement of the antenna reflection efficiency  $\Gamma$  would be a useful method for correcting the data for antenna efficiency and translating the measured spectrum to a sky spectrum. In this case, it is desirable and useful to provide a switch at the antenna terminals, which might allow a 1-port network analyser to access the antenna terminals and make an accurate measurement of  $\Gamma$ . This is best done at the observing site, where the antenna environment is the same as for the zero-mode observing. Deriving the reflection efficiency and total efficiency requires also a measurement of  $\Gamma$  for the low-noise amplifier, but that may be done in the laboratory provided that the amplifier temperature and operating conditions are the same.

- Antenna Radiation Efficiency, the ground and environment

If the zero-mode 21-cm antenna is lossless, has no resistive elements, then when viewed as a transmitter all of the power fed to the antenna, and not reflected back along the transmission line at the antenna terminals, will emerge as radiation. However, for an antenna on the ground part of the radiated power may be absorbed by the ground and only a part will go to the sky. For antennas that are on ground covered by a conducting ground plane, all of the radiated power goes to sky either directly or on reflection off the ground plane. Ground planes are ideally continuous metal planes. In practice, the ground plane may be a welded mesh, which would allow for some transmission and hence loss to ground. The ground plane may also be in the form of sheets joined along edges, where there may be leakage across the plane via the slot gaps. Passive reciprocal antennas may be viewed conversely as receivers where the loss in resistive elements of the antenna and loss in ground results in limited antenna radiation efficiency, and also emission from these resistive elements adding to the cosmic noise.

The additive component of ground emission has an imprint of the antenna radiation efficiency in a complex manner, making it difficult to separate from zero-mode 21-cm signals unless the radiation efficiency itself is designed to have characteristics orthogonal to the expectations for zero-mode signals. It is preferred to avoid the ground, and also to design the radiation efficiency to be a smooth function, as is also the design goal for the reflection efficiency.

Thus a design goal is to avoid resistive elements and also ground loss. The ground loss does depend on the ground characteristics, conductivity and dielectric constant, which

depend on soil characteristics and moisture content. Structure in the soil beneath the antenna; for example, a rock bed some distance beneath the soil surface, may also result in reflections at impedance discontinuities and hence multi-path propagation. Long wavelength electromagnetic waves in the frequency range of interest here does penetrate soil to substantial depths, which may be several metres in dry soil conditions typical of remote sites, and the soil may not be homogeneous at these depths.

The antenna efficiency is also influenced by the environment of the antenna, not only the ground beneath but also feature above like, for example, trees and other man-made structures. Conducting cables that may supply power to the radiometer and conduct signals to receivers located some distance away might also influence the total efficiency. In measurements of the reflection efficiency as a transmitting antenna, power transmitted by the antenna reflect and scatter off trees and structures in the environment and return to the antenna, as in a radar. Thus measurements of  $\Gamma$  sample the environment out to several tens of metres and beyond. Conversely, these environmental features will influence the receipt of cosmic radiation in reverse, which will also scatter off these objects and appear with spectral structure that is an imprint of the environment. Thus is it essential to have a clear space above ground and clear homogeneous soil below, to the extent at which the influence reduces below the zero-mode signal strength.

- Internal reflections of receiver and cosmic noise

Impedance discontinuities in the receiver path cause internal reflections of the system noise, which includes the cosmic noise component and also the receiver noise. Reflections of wideband noise over a physical length  $l$  result in interference that cause spectral structure with frequency scale  $v/2l$ , where  $v$  is the propagation speed of electromagnetic waves in the physical medium. For example, internal reflection within a cable of length 2 metres between the antenna and low-noise amplifier, with velocity factor 0.7, will result in spectral ripples with period 52.5 MHz. These structures would be a modulation of the receiver gain, and hence calibrated out, if the calibration includes these sections in their entirety. However, if the receiver calibration is internal, then reflections from the antenna terminal and also reflections of receiver noise from environment are omitted from the calibration. Ideally, receiver gain calibration using celestial sources, such as the passage of the Galactic plane across the radiometer beam, would best calibrate the signal path. In any case, it is desirable that the signal path have isolators that prevent back flow of signal path components to the antenna. Improved isolation also comes from use of amplifiers in which the forward gain is substantially greater than the reverse isolation, so that the net loss on propagation back and forth is substantial.

- Dynamic range

The cosmic signal is expected to be of peak amplitude in the range a few tens to a few hundred mK, and the foreground is expected to between a few hundred to a few thousand Kelvin brightness temperature. This requires a dynamic range of at least  $10^4$ , clean signal detection requires aiming for dynamic range of  $10^5$ . Because the algorithms for components separation, which depend on orthogonality between the

zero-mode 21-cm signal and other unwanted additives and foreground, are usually limited and the models for the unwanted components would subsume a significant part of the 21-cm signal; therefore, the design goal for the 21-cm radiometers is typically to achieve a spurious free spectrum of about 1 mK sensitivity, which is about  $10^6$  below the dominant foreground.

Thus the analog system components are necessarily required to be chosen to have operating points wherein inter-modulation distortion products between signal components are 60 dB below the total spectral power.

- Analog to digital conversion

A key component in the signal processing in a zero-mode 21-cm radiometer is the conversion from continuous to discrete data. Random and systematic errors in the representation of the analog signal leads to performance limitations. An important design consideration is the number of bits in the analog to digital converter. Larger effective number of bits is essential for greater spurious free dynamic range. A design goal of  $10^6$  for dynamic range requires 10 effective bits, which implies analog to digital converters of 12 bits or greater. The quality in the transfer function of the converter is also required to be accurate so that the spurious free dynamic range is  $10^6$ .

- Radio Frequency Interference (RFI)

Radio frequency interference (RFI), if present at the observing site, would naturally require improved performance of the radiometer to avoid spurious spectral structure that might confuse the zero-mode signal. In good sites, it is expected that the total power in RFI would be well below the total band power from cosmic noise. This is facilitated by adopting small antenna designs that have low gain, low effective collecting area, so that the cosmic noise is received without compromise but response to RFI is reduced.

Presence of radio frequency interference may drive the design to lower the power in the cosmic noise going to the analog to digital converter, so that the signal is not clipped in the sampling, which would lead to non-linear products spread across the band. Lowering the power in the cosmic noise, to provide greater headroom for the RFI components when they might occur, reduces the effective number of bits operating on the cosmic noise, thus requiring greater performance of the analog to digital converter.

### Experimental Challenges

CMB anisotropy and its spectrum has been measured with incredible precision; both measurements have been sufficient to explore theoretical models for the power spectrum and distortions with fractional accuracy comparable to or even exceeding that required for detection of the zero-mode 21-cm signal.

The key difference is that the wavelengths at which zero-mode 21-cm spectral distortions are expected are orders of magnitude larger, metres instead of millimetres. Scaling of physical sensors by factor 1000 is impractical: the COBE-FIRAS sensor was 2 metres long and scaling this to the redshifted 21-cm window would require a 200 metre long sensor. The new sensor designs required for radiometers at the longer wavelengths has also required new methods of calibration. Additionally, at the significantly longer wavelength, the radio sky is qualitatively different and that poses new challenges. We discuss these aspects below.

- Foregrounds

At millimetre wavelengths where the CMB peaks the sky is dominated by the CMB and the Galactic emission as well as extragalactic background is sub-dominant. Thus, in the case of mm-wavelength CMB spectral distortions, the dominant foreground spectrum is the CMB itself and the attempt is to detect distortions from Planck form in this dominant sky emission.

However, at metre wavelengths the Galaxy dominates the radio sky. The Galactic plane is the most prominent feature and structures like loops and spurs form the dominant structures. The sky brightness owing to extragalactic sources also completely dominates the CMB. Therefore, in the attempt to detect zero-mode 21-cm distortions, the dominant foreground is the Galactic and extragalactic sources, whose spectral form and intensity is only known with precision of 1-10% in the redshifted 21-cm band. To detect distortions that are expected to be a fraction  $10^{-5}$  of the foreground, when the foreground is only known with precision  $10^{-2}$  or worse, appears to be an impossible task! Moreover, the foreground varies across the sky.

- Ionosphere

The ionosphere has a time varying electron density and may be characterised by a time varying total electron content (TEC) along any line of sight. The ionosphere modifies the cosmic radio spectrum as seen from Earth in several ways. The ionosphere refracts rays, bending rays so that sky sources appear at higher elevations. The ionosphere partially absorbs the cosmic radio signal and also adds an ionospheric emission component to the sky spectrum. These effects of the ionosphere are strongly wavelength dependent and predominantly modifies the relatively longer wavelength radiation.

The TEC is measured and monitored with limited accuracy; however, the accuracy is inadequate for attempting a time-varying correction of measured sky spectra. TEC data is primarily of use in deciding the relative severity and rough magnitude with which data might be modified by the prevailing ionosphere.

The analysis of measurements from spectral radiometers must, therefore include models for ionosphere effects, whose parameters would require marginalisation while solving for the zero-mode 21-cm signal. Space missions that operate beyond the ionosphere avoid this problem.

- Radio Frequency Interference (RFI)

A significant part of the radio spectrum in which the zero-mode 21-cm signal is expected is allocated to terrestrial and space communications and transmitters at these wavelengths exist in most parts of the populated world. Radio quiet remote sites exist on Earth; however, transmissions do propagate across the world in normal as well as anomalous episodic modes and these have the potential to corrupt even data acquired in radio quiet sites and result in weak artefacts, which may be erroneously attributed to zero-mode 21-cm.

RFI is expected to be present even in Earth orbit and even at Earth Moon distances. Avoidance of terrestrial RFI may be best achieved by locating the radiometers for zero-mode 21-cm on the far side of the Moon, either in orbit or on the surface.

- Antenna design challenge

A critical component of zero-mode 21-cm radiometers is the antenna, which requires extremely careful design and development. Most available electromagnetic simulation packages are incapable of providing solutions with the precision necessary to qualify the antenna element; hence the design path necessarily requires iterations to achieve the exacting performance tolerances. There are no standard antennas in the communications engineering world that may be directly adopted for this science goal; detection of the zero-mode 21-cm requires purpose design.

For antennas located on ground, modelling the soil to the depths needed to achieve accuracy in solutions, which may be validated via measurements of the antenna beam patterns and reflection efficiencies, is an engineering challenge, particularly since several metres of soil beneath the antenna could have significant influence on the performance. In the case of antennas covered with ground screens, effects of finite screen, scattering at the edges of the screen, effect of ground beyond the screen, effect of propagation through the screen and through slot gaps between screen segments are all issues that limit usefulness of the modelling.

For antennas in space, where the ground is avoided, the antenna model needs to be inclusive of the satellite bus structure and all its conductive and dielectric materials.

Since mode-coupling may be a fundamental limitation to detection, because correcting measurement data for mode coupling requires accurate knowledge of the frequency dependent foregrounds and also a frequency dependent beam pattern, designing the antenna to be frequency-independent is essential. It is then necessary to be able to calibrate for the total efficiency, which includes reflection and radiation efficiencies, and calibrate for additives from resistive losses in antenna elements.

- Receiver design challenge

The receiver modifies the cosmic radio spectrum in its bandpass, and adds receiver noise; these require bandpass calibration and correction of the measurement data for additives. Most receiver designs adopt methods for cancellation of additives, via switching schemes. However, there may be additives that suffer multi-path propagation that includes reflection off the antenna impedance, which may not be canceled. In such cases, receiver designs often incorporate isolators and also minimise physical lengths of multi-path propagation, so that the unsubtracted additives are restricted to be of low order and hence largely orthogonal to the 21-cm signals.

### **Architectures and methods for zero-mode signal detection**

- Single-element sensor based radiometer

The simplest form of a radio telescope that would be appropriate for detecting the zero-mode 21-cm signal is a single elemental wideband antenna followed by a spectrometer. Such a radiometer would ideally have a frequency-independent antenna, a self-calibratable receiver that corrects for the bandpass, and switching schemes to cancel internal additives including receiver noise.



The considerations that drive the design of such a radiometer have been discussed above. Limitations and design challenges to the performance of such a single-element spectral radiometer are manifold. Therefore, there have been new concepts and design attempts to develop alternate schemes or configurations that might avoid some of the potential show stoppers.

- Outriggers to Fourier synthesis telescopes

A key challenge in zero-mode 21-cm radiometers is knowing the antenna beam pattern, its chromaticity, and the bandpass of the antenna element. Antenna measurements at long wavelengths are exceedingly difficult because of the parasitic effects of environment and the ground, which influence both the device under test and also the test and measurement antenna. Switched calibration, using broad band noise sources, may serve to calibrate the bandpass of the receiver chain; however, this leave the antenna bandpass, radiation efficiency, uncalibrated. This leads to a situation where the antenna characteristics may have to be derived from electromagnetic simulations, which may not have the accuracy needed for correction of the measurement data.

A solution to these issues is to deploy the single-element sensor based radiometer as an outrigger to an array of antennas, which operate in Fourier synthesis interferometer mode. The radiometer antenna and receiver chain then form another element of the array, which together observe the sky sources within the antenna primary beams. The measured spectral visibilities are then used to solve simultaneously for the sky model and also the instrument parameters, which include the bandpass and beam shape of the outrigger antenna. If all array elements are dual polarised, then full Stokes calibration is also possible, providing polarisation calibration solutions as well for the outrigger antenna.

- Interferometer methods

It is not often appreciated that interferometers are not totally blind to the zero-mode in the sky temperature distribution. If the sky were uniformly bright, then the electromagnetic fields at two points in space separated by less than a wavelength will show mutual coherence, which may be measured by an interferometer. A pair of half-wave dipoles placed adjacent to each other, in line, will respond to the zero-mode of the sky temperature distribution. If placed parallel to each other the interferometer response will contain the zero mode, and the response in this case will be greater. The interferometers may be thought of as sampling the zero-mode in the direction in which the projected spacing between the elements is zero. Of course, if the spacing between the dipoles increases, the response falls off progressively.

The coupling of the interferometer response to the zero-mode signal depends on the mutual coupling between the aperture fields in the pair of antennas. This is relatively high for closely spaced elemental antennas; however, a pair of aperture antennas placed adjacent to each other would have very little response to the zero-mode if their aperture illuminations have little overlap.

Interferometers have the advantage that when they deploy a pair of sensors of the electromagnetic field at two separated locations in space, and measure the mutual coher-

ence in the fields by cross correlating the received and amplified voltage waveforms, they are insensitive to the additive receiver noise in the two arms, which are uncorrelated.

Thus interferometer measurements are blind to internal receiver additives, thus scoring over single element radiometers on this count.

- Zero-spacing interferometer

If a pair of wideband antennas are placed adjacent to each other, with mutual coupling between the antennas, then the interferometer response includes the zero-mode signal. This response may be enhanced by placing a vertical beam splitter in between the antennas, so that incident field from any side propagates to the antenna on the far side through the screen and to the near side in a direct path and also after reflection off the screen. If the beam splitter is reactive, and lossless, then the response to zero-mode on the two sides of the screen cancel. However, if the beam splitter is resistive, then the interferometer responses add with the same sign. Optimally, it can be shown that the sensitivity is a maximum if the beam splitter has a sheet resistance equal to the impedance of free space.

Interferometers made from elemental antennas placed in a close packed configuration will have a telescope filter function, which defines the interferometer response to zero-mode 21-cm signals, that is highly frequency dependent and challenging to calibrate. The advantage of the zero-spacing interferometer made from frequency independent antennas is that the telescope filter function in this case is flat, at least over the frequency range in which the resistive screen is frequency independent.

- Moon block

An alternate and interesting approach to detection of the zero-mode 21-cm signal is via Fourier synthesis imaging of the Moon. The brightness measured towards the Moon, over frequency, would be a difference between the Moon brightness and the mean brightness of the radio sky. Thus if the Moon were assumed to be of flat spectrum, or if the temperature spectrum of the Moon were known, then the differential measurement may be used to infer the zero-mode 21-cm signal.

The measurement is not without difficulties. Firstly, it is not clear that the temperature spectrum of the Moon is flat or even smooth; the lunar regolith may have structure and layered in depth, which would give spectral structure in the emission brightness. Additionally, the Moon is reflective and hence the brightness of the Moon has a component that is a reflection of the radio sky. Finally, the synthetic beam of the Fourier synthesis telescope will have sidelobes that are chromatic; thus mode coupling will induce spectral structure whose removal from the data will be limited by the depth of deconvolution.

### **Calibration approaches (LJG)**

Bandpass calibration

Calibration of additives

some sub-units can be cal'd in the lab, others only in the field.

Ideal goal: flat spectral baseline to  $1:10^6$

discrimination btw. 21cm and foreground signals

Compromise to admit spectral baselines that are at least partially orthogonal to 21-cm signals

Switching

- Dicke switching
- ways in which to separate what is up and downstream from frontend switch
- inbuilt noise and calibration sources

correlation spectrometers – phase switching

when all else fails, fit a model to the spectral baseline

- physically motivated
- empirical
- dangers

### **LEDA, SARAS, and EDGES (LJG)**

- discussion of design considerations emphasized in each project
- status
- **First detection? EDGES. (LJG)**

### **Outlook (LJG)**

- Testing the claim
- Is there more to do after the first detection with reasonable S/N?