

Required Instrumentation Information for Interferometrists and Radio Astronomers

Griffin Foster

February 9, 2015

When thinking about engineering and the development of an instrument from a practical point of view the key concept to get across is that *any measurement is a loss in information*. In the case of astronomy, there is the potential to measure infinite frequency bandwidth, frequency resolution, time bandwidth, and time resolution across a 4π area of the sky. In reality, the instruments that we can build have limits on all these parameters, so we must be selective about which information we retain based on the scientific goals and engineering limitations.

1 Classic (dish-based) Interferometry

We start with dish-based interferometry, which I am calling ‘classic’ interferometry. This is a collection of antennas which have some sort of curved optics, are mechanically pointed, and correlated together to form visibilities. Examples of this type of array are the Very Large Array (VLA), Karoo Array Telescope (KAT-7), Atacama Large Millimeter Array (ALMA), and Westerbork Synthesis Radio Telescope (WSRT). Using this type of interferometric array, we will follow the path of electromagnetic waves from the interface with the dish optics up to the final correlated visibilities.

1.1 Optics and EM Interactions

Given a dish diameter D , and an observing wavelength λ we think of two different interaction domains. When $D \gg \lambda$ light can be treated as a particle, this is the case for optical astronomy. For example, the ratio of D/λ for a small optical telescope with a 0.5 meter mirror, observing visible light (500 nm) is 1000000. When $D \sim \lambda$ (with in a few orders of magnitude at least) then light needs to be treated as a wave, with complex interactions, this is the case for radio astronomy. For example, a VLA dish is 25 meters in diameter, observing at L-band (21 cm), then $D/\lambda = 120$, in the limit where the light is wave like. Due to the scale of radio waves (10 m - 1 cm), and the inherent scale of humans and our devices, we are in this limit. Later we will see that one can even observe in the domain where $\lambda > D$ with out dish elements.

1.1.1 Dish Optics

To first approximation we can treat light as a particle and then look at the optics of different dish designs. In this simple case, for a monochromatic signal, ray-tracing is used to determine the focus of a dish. The *primary focus* optics (e.g. KAT-7) uses a paraboloid shape to focus light to a single point above the dish. This simple design has a single mirror with the receiver above the dish. A *cassegrain* design (e.g. VLA) also uses a paraboloid primary mirror design, but also has a convex secondary mirror which reflects the light down again to a focus at or below the primary beam dish. Both of these designs have an issue of occultation from the structures that hold the feed or secondary above the primary mirror, this leads to complex beam structures due to scattering (an issue of the wave-like nature of light). To avoid this blockage a *offset Gregorian* (e.g. MeerKAT) optics design can be used, this design uses a subset of a paraboloid to reflect the light away from the direction of incident to a secondary which does not block the primary. Extending

beyond a monochromatic signal, the focus of a wideband signal vary with frequency. This can be controlled with modifications to the mirror shapes, and the design of the receiver feed. This complex focus pattern is called a *caustic*, a common example of this is looking at the response of visible light through a glass of water.

1.1.2 Efficiencies

The mirror surface accuracy, type of material, and design of the dish structure all lead to light signal loss, this loss is measured as an *efficiency* denoted as η . Typically a surface needs to have structure variations on the scale of $\sim \lambda/10$ of the observing wavelength. For L-band (1.421 GHz) this is around 2 cm, for C-band (5 GHz) this is 6 mm, and for VHF (~ 100 MHz) this is 30 cm. As the surface variation increases relative to the observing frequency the surface efficiency decreases. The choice of mirror material is also important, there is some loss in any reflective material, additionally there is pickup from the ground which leaks into the material. The scattering of light by the structure above the dish also leads to a reduction in efficiency. Any efficiency is less than 1 and is a multiplicative effect on the received signal. The overall efficiency is called the *aperture efficiency*, and is of the form

$$\eta = \eta_{\text{surface}} \eta_{\text{blockage}} \eta_{\text{spillover}} \eta_{\text{taper}} \dots \quad (1)$$

where η_{surface} is the efficiency of the surface accuracy and material, η_{blockage} is the efficiency due to structure above the mirror blocking light, $\eta_{\text{spillover}}$ is the efficiency of the illumination of the feed at the focus, and η_{taper} is the efficiency due to the dish shape. Other efficiencies can be included to compute the aperture efficiency.

1.2 The Primary Beam

If the dish efficiencies describe the overall effect on the light measured at the receiver feed then the *primary beam* is a more complex and detailed effect which takes into account the direction of a source, and the frequency response of the dish design. The primary beam describes the complex gain for a given frequency and direction. For this discussion the dish structure and antenna feed are coupled together into the primary beam. The antenna feed, like the dish structure, has a complex time, frequency, and direction response based on the design. The antenna converts EM waves to a current, optimizing for parameters such as gain, bandwidth, polarization leakage, impedance matching, and the smoothness lead to different feed designs. A $\lambda/4$ (*quarter-wavelength*) *dipole*, a classical design, is a simple dipole made up of two monopoles of length $\lambda/4$ connected to a *balun*, which is an electronic device which converts an EM signal to a current. This design is very broad spatially, close to isotropic, with a peak sensitivity at λ with a smooth roll-off to other frequencies, but is not broadband or directive. Other feeds, such as log-periodic dipoles, Vivaldi, horns, etc. are optimized for different parameters depending on their use. Many books and lifetimes have been spent on designing antenna feeds.

1.2.1 Polarization

Further discussion of the primary beam first requires a digression into polarization. Light waves have some orientation of oscillation. A plane wave, which for our case we can approximate a light wave to be, can be described with two values based on orthogonal axes, either linear or circular. Figure 1 shows different types of possible orientation.

Using vector notation a EM wave is described by a 2-dimensional vector

$$\mathbf{e} = \begin{pmatrix} e_a \\ e_b \end{pmatrix} \quad (2)$$

where a and b are orthogonal modes of polarization. A full polarization antenna feed is made up of two feeds, either linear or circular, which measure orthogonal modes of polarization to fully measure the polarization

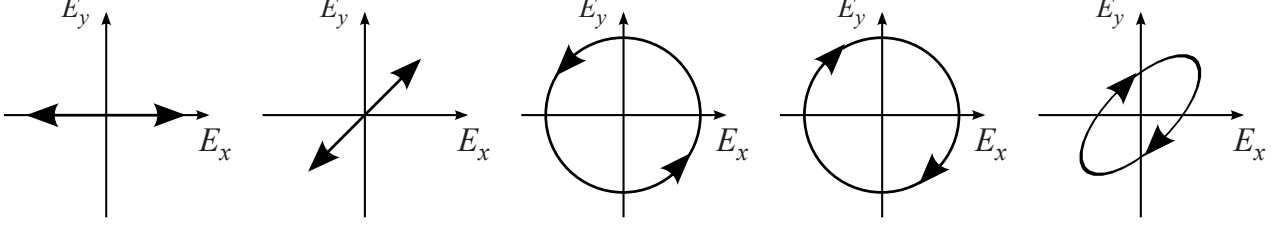


Figure 1: Modes of polarization: full linear, mixed linear, left circular, right circular, elliptical which is a combination of linear and circular. (wikimedia)

components of the EM wave. Then the primary beam is a frequency, position, and time dependent Jones \mathbf{J} matrix which transforms \mathbf{e} into a complex voltage \mathbf{v} which will be amplified by the analogue system.

$$\mathbf{v} = \begin{pmatrix} v_a \\ v_b \end{pmatrix} = \mathbf{J}\mathbf{e} = \begin{pmatrix} j_{a \rightarrow a}(\nu, t, l, m) & j_{a \rightarrow b}(\nu, t, l, m) \\ j_{b \rightarrow a}(\nu, t, l, m) & j_{b \rightarrow b}(\nu, t, l, m) \end{pmatrix} \begin{pmatrix} e_a \\ e_b \end{pmatrix} \quad (3)$$

In \mathbf{J} the diagonal elements represents the gain from converting the EM wave to complex voltage, where as the off-diagonal terms represent the *leakage* from between one orthongal mode feed into the other. Ideally the off-diagonal terms are close to zero.

Since we are not measuring individual waves, and we are making a stochastic measurement, we want to think about the polarization of a source as the average polarization over many wave. Most radio sources are ‘incoherent’, or unpolarized sources, meaning if only one polarization is used then only half of the signal from the source can be measured. Sources such as pulsars are ‘coherent’ and are highly polarized. There are many effects which can introduce and washout polarization to a source, but that is for the science experts to explain.

1.2.2 Back to the Primary Beam

Broadly looking at a primary beam there are three types of components: the primary lobe, side lobe, and back lobe. The primary lobe is the main lobe, which has the forward gain in which the antenna or dish was designed to maximize. The centre of this lobe is generally what is pointed towards the centre of the science field and is also called *boresight*. Side lobes are the spillover structures that are created from the design. They lead to pick up of signals from outside the primary lobe, which then leak into the receiver. During design, one optimization is to minimize or smooth the side lobes. The back lobe is another side lobe structure, but is directly opposite the primary lobe. This lobe can pickup signals from the dish and antenna electronics and the ground. Antenna design is a balancing between required primary lobe structure and the trade-offs with the side lobes and back lobes.

A typical metric to describe a beam in a single number is the *gain* term, this is a combination of the *directivity* which is a metric for the size of the primary lobe, and the *electrical efficiency* which is the efficiency of the conversion from electromagnetic waves to power. The gain term here is different from the complex gain Jones matrix which is commonly used in the RIME, discussed in another lecture.

The primary beam shape introduces a number of effects on radio sources. Taking a simple unpolarized point source, continuum radio source as a typical example we can see how the primary beam effects the source. If our source is not at the phase centre then there will be some distance-dependent attenuation in the total intensity flux. If the two orthogonal polarization feeds are out of alignment, or projected then the source will appear partially polarized. The structure of the primary beam as a function of frequency will introduce frequency-dependent structure into the source known as *spectrum colorization*. The primary beam is different on every dish and antenna because of manufacturing imperfections and the environment. This leads to variations between elements, and a time-dependent effect such as observing in windy conditions which will cause dishes to sway at different rates. As will be seen in other lectures, moving beyond the

simply observation cases, the primary beam introduces significant effects which must be taken into account when modeling an interferometric system.

1.2.3 Stokes Parameters and the IXR

The sources and skies we measure after forming an interferometric image are real valued powers, how do we go from complex voltages to powers? These real-valued powers are called *Stokes parameters* \mathbf{e}_S , and a full-Stokes measurement for a source is represented by four values: I, Q, U, V . Where I is the measure of the total intensity, the total flux measured, this is a positive real value. Parameters Q and U describe the linear polarization of the source, and V is the direction and intensity of the circular polarization, these values can be any real value. Given two complex signals \mathbf{e}_A and \mathbf{e}_B a *coherency vector* \mathbf{e}_J is formed by $\mathbf{e}_J = \mathbf{e}_A \otimes \mathbf{e}_B^*$, where \otimes is the *outer product*. Conversion from \mathbf{e}_J to \mathbf{e}_S is

$$\mathbf{e}_S = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix} = \mathbf{S}^{-1} \mathbf{e}_J = \mathbf{S}^{-1} \begin{pmatrix} e_{Ax} e_{Bx}^* \\ e_{Ax} e_{By}^* \\ e_{Ay} e_{Bx}^* \\ e_{Ay} e_{By}^* \end{pmatrix} \quad (4)$$

\mathbf{S} and \mathbf{S}^{-1} are the conversion matrices to transform between Stokes parameters and the brightness coherency vector. For reference they are:

$$\mathbf{S} = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & i \\ 0 & 0 & 1 & -i \\ 1 & -1 & 0 & 0 \end{pmatrix} \quad (5)$$

$$\mathbf{S}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & -i & i & 0 \end{pmatrix} \quad (6)$$

In the imaging domain it is useful to use the Stokes parameters and Mueller matrices which represent powers and transformations on the flux. In the gain chain and calibration domain it is useful to use the Jones matrix representations because they are operations acting on a complex signal before correlation. This is all discussed in more detail in the calibration and RIME lectures, but for now it is a necessary evil in introducing polarization concepts.

A final concept before we leave polarization and the primary beam is the metric we use for leakage between element feeds called the *intrinsic cross-polarization ratio* or IXR. This qualifies the coupling between two orthogonal feeds. Going back to the Jones formalism for the primary beam the matrix can be decomposed via *singular value decomposition* (SVD) into three matrices

$$\mathbf{J} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\dagger = \mathbf{U} \begin{pmatrix} \sigma_{\max} & 0 \\ 0 & \sigma_{\min} \end{pmatrix} \mathbf{V}^\dagger \quad (7)$$

The degree of information loss due to the diagonal components is described as the *matrix condition number* $\text{cond}_2(\mathbf{J}) \equiv \kappa(\mathbf{J}) = \frac{\sigma_{\max}}{\sigma_{\min}}$, where σ_{\max} and σ_{\min} are the maximum and minimum singular values. The IXR in Jones formalism is defined as

$$\text{IXR}_J = \left(\frac{\kappa(\mathbf{J}) + 1}{\kappa(\mathbf{J}) - 1} \right)^2 \quad (8)$$

In the Stokes/Mueller formalism it is

$$\text{IXR}_M = \frac{\kappa(\mathbf{M}) + 1}{\kappa(\mathbf{M}) - 1} = \frac{1 + \text{IXR}_J}{2\sqrt{\text{IXR}_J}} \quad (9)$$

When the condition number is near 1, then the matrix is well conditioned and the IXR will be very large, if the matrix is poorly conditioned then the IXR will approach 1.

1.3 Analogue Receiver Front-end

EM waves are converted to a current via the antenna feed (using a balun or waveguide), from there the signal can be treated as a voltage in an electronic circuit. But, the signal is very weak, it looks like weak Gaussian noise. That noise is what we would like to measure, but there are other types of noise, which we don't want, leaking into the system. At the earliest possible point in the analogue receiver chain the signal must be amplified using a *low-noise amplifier*. To understand why this amplification is necessary we must start with the *system temperature*

1.3.1 System Temperature

The sky signal, if we consider it to be a voltage in an electrical circuit, is a weak Johnson-Nyquist noise source, also known as a 'cold' source. Johnson-Nyquist noise quantifies the power of a voltage signal in units of Kelvin.

$$P = k_B T \Delta\nu \quad (10)$$

Where P is the power of the voltage signal, k_B is the Boltzmann constant, T is temperature in Kelvin, and $\Delta\nu$ is the bandwidth of the signal. At L-band (1.4 GHz) the sky temperature is around 10 Kelvin, 2.73 K of which are from the CMB. The temperature of the sky varies depending on frequency. At low frequency, say 100 MHz the sky is close to 1000 K. The earth, the ground that is around 300 K. We can see that if even a small percentage of the ground is picked up in the receiver then the sky signal can be buried underneath. The design of the early stages of the analogue receiver must pay important attention to which components are used, as each one adds to the system temperature. There is additional noise from ground pickup, the atmosphere (depending on observing frequency), the antenna feed, filters, amplifier, and passive electronic components.

A low-noise amplifier (LNA) typically will boost a very weak signal by a factor of 1000 (30 dB) or 10000 (40 dB) while introducing minimal noise. An LNA is introduced as earlier as possible into the analogue chain to have the effect of boosting the sky signal, so as it proceeds through the rest of the system, it remains the dominant signal. The system temperature is of the form

$$T_{\text{sys}} = T_{\text{sky}} + T_{\text{atmosphere}} + T_{\text{spillover}} + T_{\text{rx}} + \dots \quad (11)$$

where T_{rx} is the receiver noise of the form

$$T_{\text{rx}} = T_{\text{feed}} + \frac{T_{\text{passive}}}{G_{\text{feed}}} + \frac{T_{\text{LNA}}}{G_{\text{feed}} G_{\text{passive}}} + \frac{T_{\text{amp}}}{G_{\text{feed}} G_{\text{passive}} G_{\text{LNA}}} + \dots \quad (12)$$

where the T terms are the additional temperature introduced with each component, and G is the 'gain' term which can be less than 1 (attenuation) or greater than 1 (amplification). The gain from the feed and passive components will be less than 1, and are similar to an efficiency, these will cause the temperature from these components to increase in the system temperature. By having the LNA as early as possible we see that the weak sky signal is amplified, and noisier components such as filters and second stage amplifiers can be introduced without introducing significant noise compared to the sky signal. This LNA and the front-end analogue components are often cooled using a heat pump or a *cryostat*, a seal vacuum container which can be cooled down to ~ 10 Kelvin. For low-frequency systems the sky is so 'hot' compared to room temperature that additional cooling offer minimal improvements in the system temperature.

1.3.2 System Temperature

Why would we like our system to be as 'cold' as possible? The classic equation to describe the sensitivity of a radio antenna is the *radiometer equation*

$$\sigma_T = \frac{T_{\text{sys}}}{\sqrt{\Delta\nu\tau}} \quad (13)$$

which describes the necessary bandwidth ($\Delta\nu$) and integration time (τ) required to reach a desired noise level (σ_T) for a system with a given system temperature (T_{sys}) when observing a broadband source. A

example is useful here. If there are two systems, with the same bandwidth $\Delta\nu$, in which one has a system temperature twice that of the other, i.e. $T_{\text{sys},0} = 2T_{\text{sys},1}$, then $T_{\text{sys},0}$ will take 4 times longer to reach the same sensitivity as $T_{\text{sys},1}$.

1.3.3 The Bandpass

The bandpass response is the frequency-dependent structure of the signal. The frequency-dependent structure of the sky is modified by the optics, receiver, analogue components, and the digital signal processing. Generally, the dish element will have a frequency-dependent efficiency which have the effect of decreasing gain as the frequency increases. The feed will be tuned for a specific response, which can even introduce sharp features into the bandpass. Like the feed, the LNA will be tuned for a specific frequency response. Importantly in *heterodyne* system, which take advantage of Nyquist aliasing, bandpass filters are used to create sharp cut-offs on the edges of the band. Often it is necessary to introduce notch filters to reduce the effect of *radio frequency interference* (RFI).

RFI is a class of human-made radio signals which take on a number of different forms, but all of which cause a degradation of the recorded sky signal. Many governments allocate the radio spectrum, and use it as a common utility or sell off sections of the spectrum to companies. The use of these spectrum allocations introduce RFI into the environment. Typical RFI sources are FM radio (89-110 MHz) and GSM mobile (around 900 MHz and 1.8 GHz). RFI is characterized as being ‘narrow’-band where the RFI is on the scale of a few frequency channels (frequency channelization is discussed in the next section), and ‘wide’-band where the RFI dominates most of the observed spectrum. An RFI source will also have a *duty cycle*, which is the time-averaged ratio of time in which that RFI signal is on or detectable to the observing time. If a signal is always detectable then it will have a duty cycle of 1. An RFI signal with a duty cycle of 1 is manageable if that signal is narrow band, and we are not interested in the sky signal at that frequency. The signal can be managed by either introducing a notch filter to attenuate the signal in the analogue path, or digitize the signal and ‘flag’ the frequency channels as not usable. If an RFI signal is wide band with a duty cycle near or at 1, then little can be done to suppress that signal and pick out the sky signal. But, if the duty cycle of the wide band signal is low then the entire band can be flagged for that short time period with a minimal loss to the sensitivity. A typical wide band signal is aircraft RADAR, which is very strong, but moves quickly through the beam.

1.3.4 Digitization

After signal conditioning in the analogue front-end, we are ready to capture the signal into a digital form so that we can perform the correlations. The term digitization is worth thinking about for a moment, it is the process of taking a continuous in time analogue signal (which is described by real numbers), and converting it to a discrete in time and voltage signal (which are ‘digits’ described by integers). The last analogue component is the *analogue to digital converter* (ADC). A simple form of this is a cascading resistor ladder which measures discrete steps in an analogue signal, each output of the ladder triggers a bit value of either 0 or 1, thus a digital signal is created. A typical radio astronomy ADC might have 8 bits, that is the signal can take on 256 values (−128 to 127). There is a resolution to these steps, which introduces digitization noise to the signal as the continuous signal is truncated to the closest step. You can imagine that more bits will reduce this noise, as there will be more steps and the resolution will be smaller. Or, more bits can be used to increase the dynamic range of the ADC. Say there is a strong RFI source with a weak sky signal, we want to be able to sample the sky signal at the enough resolution, but also we need to fully sample the RFI signal. If the maximum ADC value is smaller than the RFI signal then the RFI will *saturate* the ADC, causing the output of the ADC to sit at the highest or lowest value and no sky signal can be captured. This is another reason why strong RFI is often managed in the analogue chain. The choice of number of bits and step resolution depend on the RFI environment, the amplification of the sky signal, and also the bandwidth of the ADC.

1.4 Digital Back-End

A typical digital back-end for a radio astronomy dish is a spectrometer, beamformer, pulsar timing engine, base-band recorder, or a correlator. These instruments do fairly simple operations, but these operations need to be performed very quickly. A *spectrometer* transform a time-domain voltage signal into a frequency-domain spectrum using a *fast Fourier Transform* (FFT), a power converters, and an accumulator. A *beamformer* adds together multiple antenna signals into one combined signal. The simplest instrument is the *base band recorder* which simply records an ADC output signal to disk, as in the case of VLBI. A *correlator* is the important instrument for interferometry. Two signals are correlated together by use of an FFT, multiplier, and accumulator. We will focus on the correlator in this discussion.

1.4.1 The Nyquist Sampling and Aliasing

Digitization samples a continuous signal at regular intervals, this introduces a frequency ambiguity due to the f_s sampling rate. With a sampling rate of f_s , the maximum frequency which can be resolved (the Nyquist frequency) is $f_s/2$. Frequencies above the Nyquist frequency are said to ‘alias’. If we are only sampling at a rate of f_s then any sine wave oscillating at a higher frequency can go through many cycles between sampling points. There exists another sine wave, going through fewer cycles at below the Nyquist frequency which will match in amplitude at the same sampling points. Thus these two sine wave can not be differentiated, in fact there are not just two but an infinite set of such waves due to the periodic nature of sine wave. Knowing this, if we would like to Nyquist sample a 100 MHz band, then we need an ADC which samples at 200 MHz, with appropriate bandpass filters.. This is a difficult topic and there are many books which will provide a thorough explanation of the topic. I want to stress this: it worth understanding aliasing and Nyquist sampling as they will come up in everything to do with Fourier Transforms and interferometry. I recommend Richard Lyons’ *Understanding Digital Signal Processing*.

1.4.2 The Convolution Theorem

The convolution theorem has been brought up in other lectures, but important topics are worth review. The theorem states that for two signals f and g

$$\mathcal{F}\{f * g\} = \mathcal{F}\{f\}\mathcal{F}\{g\} \quad (14)$$

where $\mathcal{F}\{\}$ is the Fourier transform of the signal and $*$ is the convolution operator. And, the convolution of two signals $h(z)$ is

$$h(z) = f * g = \int f(x)g(z - x)dx \quad (15)$$

Correlation and convolution are very similar, the difference lies in the details of using complex signals and using the conjugate of one of the signals. The importance in these equations is that by multiplying the Fourier transform of two signals we get the Fourier transform of the two functions convolved. A convolution is a computationally expensive task, where as two FFTs and a multiplication is fairly simple. By convolving sky signals from two dishes we produce the *interferometer visibility*, and when the visibilities from all interferometer pairs are combined, and Fourier transformed an image of the sky can be created. Convolution will come up again and again in imaging and calibration.

1.4.3 Correlators

In a practical way how are the visibilities computed with a correlator? There are two ways to go about it, the FX and XF architectures, but both ways produce the same result in the end. In the FX architecture a Fourier Transform is applied to a window of samples from two signals, this is the F component. Then the two Fourier transformed signals are multiplied together (one of the signals is conjugated to make this a correlation instead of a convolution), and the resulting signal is then accumulated, this is the X component. As you might guess, the XF architecture does the two operations in a reverse order. There are a few reasons to choose

one architectures over the other but, suffice to say the FX is generally favoured for computational reasons. The correlation of all antenna pairs is computationally intensive operation and is commonly implemented on field programmable gate arrays (FPGAs) or graphics processing units (GPUs).

1.4.4 FFTs and PFBs

The bandpass of the analogue system and digitization process select out a range of frequencies to be correlated between antennas. The full band from each antenna can be correlated against another antenna but this would only give information about the entire band. If there is strong narrowband RFI then any correlation will be dominated by that signal, and the sky signal will be lost. Or, if we are only interested in a weak narrow signal, such as a spectral line, that line structure will be hidden in the band correlation. In order to gain access to subsets of the frequency range, known as *subbands* or *channels*, a Fourier Transform must be employed. In this context a Fourier Transform will transform a wideband time-domain signal into a set of frequency-domain signals (for a given ‘window’ of samples). The 1-D Fourier Transform of a signal $x(n)$ to $X(k)$ is defined as

$$X(k) = \int_{-\infty}^{\infty} x(n)e^{-2\pi ink} dn \quad (16)$$

A simple way to think about this is that a signal is decomposed into a set of sine and cosine waves. As we are working on discrete samples, regularly sampled in time we will use the *Discrete Fourier Transform* (DFT) which is defined as

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-\frac{i2\pi}{N}nk} \quad (17)$$

where N is the number of samples in the data window. Using this form the naive computational cost of this transform is $\mathcal{O}(N^2)$, which can grow to be very computationally expensive for relatively small sizes of N . The solution to this computational problem is the *Fast Fourier Transform* (FFT) which is a class of algorithms which use the regularity of the sampling to approach a run time of $\mathcal{O}(N \log N)$. The Cooley-Tukey Radix-2 algorithm is the classic example of a FFT algorithm. An important note to make right here, you will often see FFTs of size 2^d , i.e. 256, 512, 1024, etc., being used, this is because when a size of 2^d is used the most efficient FFT can be used. Other FFT sizes, even sizes $< 2^d$ will take longer to compute. So, the point I am trying to make is *any time you are doing an FFT try to pick a size which is of the form 2^d* .

Back to our discretely-sampled time-domain signal, depending on how many samples we choose to include in our window for applying the FFT affects the subband resolution. For N real-valued time-domain signals $N/2$ complex frequency subbands are created. The apparent factor of 2 ‘loss’ in information due to the FFT is because the output of the FFT is complex and the input is real, so there are effectively 2 values per frequency subband. The more samples included in the window, the higher the resolution of subbands output. In the extreme case of an infinite window the DFT approaches the continuous Fourier Transform output. On the other side, a window of 1 sample will produce the same output as input. It is often useful to think of the FFT generating a series of narrow-band time domain signals, this is called a *filterbank* in DSP.

I have been using the term windows for blocks of samples which an FFT is applied to. If you take a continuous signal and select out a block of samples then you have altered the signal, you have effectively applied a *top-hat function* (rectangle) to the signal. This might not seem like an issue because none of the samples have been modified, but the continuity of the signal has been altered. When an FFT is applied to the window, from the convolution theorem, we are convolving the Fourier transform of the data with the Fourier transform of a top-hat function. The Fourier transform of the top-hat function is known as a *sinc function* which is defined as $\text{sinc}(x) = \frac{\sin x}{x}$. The top-hat function is known as a *windowing function*, and it turns out to be a generally poor one. The primary issue with the top-hat function is that there are high ‘sidelobes’ with a ‘slow’ roll-off. Many other windowing functions have been developed: Hamming, Hann, Blackman, Gaussian, etc. These windowing functions have a trade off in the subband width, sidelobe levels, sidelobe roll-off, and sharpness. A simple example as to why you would want to apply a windowing

function to samples is that if there is strong RFI in one subband, and the windowing function has high sidelobes (like a top-hat function), then power will *leak* from the RFI subband to clean subband, corrupting it.

A more advanced form of windowing is often applied before the FFT in what is known as a *finite impulse response* (FIR) filter which is a what of overlapping windowed samples to have a larger winding function. The FIR together with the FFT is known as a *poly-phase filterbank* (PFB), and all modern digital back-ends use this instead of the simple FFT.

1.4.5 Low-bit Correlation and Visibilities

After the bandpass signal has been channelized into subbands then the subbands are ready to be multiplied and accumulated. That is, for each antenna pair A_i, A_j , subband k from antenna A_i is multiplied with (the complex conjugate of) subband k of Antenna A_j . Only subbands with the same frequency are correlated. For an array of N_{antennas} antenna elements, the number of baseline correlations is

$$N_{\text{corr}} = N_{\text{auto corr}} + N_{\text{cross corr}} = N_{\text{antennas}} + \frac{N_{\text{antennas}}(N_{\text{antennas}} - 1)}{2} \quad (18)$$

where $N_{\text{auto corr}}$ is the number of antenna self-correlations, or *auto-correlations* and $N_{\text{cross corr}}$ is the number of antenna pair *cross-correlations*. In most cases when using an interferometric array, the auto-correlations are ignored due to issues of antenna self-noise, but that is a discussion for another time.

Before the multiplication and accumulation is performed that is a stage known as quantization. This stage reduces down the number of bits used to represent the signal in each subband, often down to 1, 2, or 4 bits per value. Quantization is performed to save in computational cost, and it turns out to have minimal impact on a sky signal. For example, quantizing a subband signal down to 4 bits for the real component and 4 bit for the imaginary component, there is a *quantization efficiency* of 0.989. That is, as opposed to an unquantized signal, there is only a 1.1% loss in correlation. Going down to 1 bit (2 levels) the quantization efficiency is 0.637, not too bad for a single bit.

After quantization the multiplication are performed, and accumulated to increase the signal to noise and reduce the output data rate. The final visibilities are collected into data cubes with dimensions of time, frequency, and polarization for each baseline.

This digital back-end overview has skipped some additional features included in these systems, such as timing keeping, fine and coarse delay correction, equalization, and fringe-stopping. All these features are important in the practical design on an interferometric array, and are worth reading up on.

2 Transiting Arrays, Aperture Arrays, and Phased-Array Feeds

Up to this point we have been focusing on using dishes with feeds for interferometric observations. But there are other types of ‘receiving elements’ which can perform that function, each with their own advantages and disadvantages. The *phased-array feed* (PAF) is a feed design for a dish which generates multiple feeds instead of the single ‘pixel’ feed we have been discussing. We can also generalize interferometry to not just dish-based receiving elements in the form of *transiting arrays* and *aperture arrays* (AAs). Many of these designs use a digital or analogue, or hybrid beamformer.

2.1 Beamforming

Beamforming is the process of adding together multiple signals, with different complex weights, from different receiving elements to form a new primary beam. The voltage signal from multiple elements, which can have wide and complex primary beams, are summed together to form a narrower, and smoother primary beam. For example, a dish is a form of mechanical beamforming, where the light is summed with different weights based on the point of interaction and reflection on the dish. Often beamforming is done with analogue components, such as the LOFAR HBA elements, or after the signal is digitized, such as in the LOFAR stations. There are

many techniques to perform beamforming, you would be surprised how complex it is to just add together numbers in a meaningful way.

Beamforming algorithms can also be broken down into *coherent* and *incoherent* methods, where in coherent beamforming there are phase weights added to each voltage signal where as incoherent beamforming is simply a weighted sum of the output powers and can be applied long after an observation. Coherent beamforming allows a narrowing of the overall beam to create higher forward gain or lower sidelobes. A common use is when observing a known pulsar which is low in flux, beamforming increase the sensitivity in a narrow field of view. Incoherent beamforming does not affect the field of view or sidelobes, but increases the overall sensitivity. This form of beamforming can be used to search for rare events where the location of occurrence is not know, so sensitivity across a large field of view is desirable.

2.2 Phased-Array Feeds

We can think of the classic dish with a single feed as being a single pixel camera, where the pixel covers the size of the primary beam. But, there is no geometric reason we can not construct a feed with multiple pixels that each cover a subset of the primary beam. Then each pixel could be correlated with a pixel in another dish which covers the same area of the sky. This is the idea behind the phased-array feed. It is a relatively new technology in radio astronomy and undergoing much development at the moment. It takes a form of a grid of active receiving elements, which are beamformed together to form multiple beams across a wide field of view.

The main advantage to a PAF-based dish is there is a greater sky coverage compared to a single pixel feed. This makes the design ideal for surveys and large-scale sources. This increased sky coverage does not come for free, each receiving element requires its own, isolated analogue front-end and digital back-end, making the feed more expensive. As PAF technology is new, it is undergoing development. At the moment, PAFs have a higher system temperature and more limited analogue bandwidth compared to a single pixel feed. Both these issues can potentially be overcome with further advances. Additionally, the calibration of PAFs are a new challenge, and the software and techniques are in an ongoing state of development.

As this is a new technology, many of the telescope arrays which use PAFs still being built and designed. Example of PAF-based arrays are the APERTIF Westerbork upgrade, ASKAP, and the SKA-MID survey.

2.3 Transiting Arrays

Instead of having a dish which can be used to track a particular point in the sky over many hours, one can build a transiting array which have elements that do not move (or have limited mobility) and the sky ‘transits’ through the primary beam of the elements. These are some of the simplest array designs, and some of the first built. A feed will have an effective primary beam, which will be large compared to that of a dish. As a source enters into that beam, its apparent flux starts small, gradually increases decreases as it moves across the beam sidlobes, then peaks when the source is at the local meridian, and then decreases as it sets to the horizon. Examples of these type of arrays are PAPER and the Medicina Northern Cross.

There are obvious cost advantages to building an array with no moving parts. But also, such arrays have wide fields of view, and can have the individual elements places close together. This allows for large-scale structure experiments such as EoR and BAO studies. The disadvantages compared to dishes create a new challenge for calibration an imaging. The individual elements are less sensitive compared to a dish, so more elements are needs, requiring larger correlator systems. Also, as the sky transits the apparent flux of sources changes, so the primary beam must be well know in order to get back to the *intrinsic* flux of the sky. Depending on the scale of the primary beam, a transiting array has set amount of time per day in which a section of the sky can be observed. This means, a deep integration of a region of sky is not possible, without observing for many days.

2.4 Aperture Arrays

Using a bit of technology we can turn a transiting array into a digital dish which can be pointed, also known as an aperture array. In fact, an aperture array can be pointed in many directions simultaneously. The idea with an aperture array is that by beamforming many extremely wide field of view, low-sensitivity elements the beamformed signal has a high sensitivity and can be pointed in any direction in the sky. By updating the beamforming weights, the beam of the aperture array can track a region of the sky. The aperture array takes advantages from both dishes and transiting arrays. The main cost is the analogue and digital electronics to build such an array. For this reason, the aperture array is mainly used for low-frequency science, such as LOFAR and the future SKA-low, as the components are cheaper. But, as technology progresses the price of higher frequency components will make it possible to increase the observable frequency. A second issue with aperture arrays is that the primary beam changes depending on pointing location and frequency. As the beam is a weighted sum of all the individual elements there is limited precision to the beam shape. There is also the issue of *sparse* and *dense* aperture arrays. When the elements of the aperture array are placed closer than $\lambda/2$ in observing wavelength the array is ‘dense’, the array is fully sampling the wavefront and there are no beam artefacts such as ‘grating lobes’(a type of sidelobe) which introduces significant structure into the beam. If the elements are further apart than $\lambda/2$ then these sidelobe structures appear and limit the sensitivity and field of view. Now, for a single observing frequency, designing an aperture array is simple as all the elements are spaced at $\lambda/2$. But, for wide-band arrays, if the elements are placed at $\lambda_i/2$ for a wavelength λ_i , then for any wavelength $< \lambda_i$ the array configuration under-samples that observing wavelength and introduces large grating lobes. So, a balance between observing bandwidth, cost, and dense versus sparse trade-off must be made during the array design.