The Measurement Process with Fully-Filled Apertures

C.J. Salter, NAIC, Arecibo Observatory, HC3 Box 53995, Arecibo, PR 00612.

1 Introduction

Perhaps the most remarkable thing about radio astronomical measurement is its peculiar mixture of the highly accurate, and the rather inaccurate. For example, we can measure the repetition period of the radio pulses from pulsars with an accuracy approaching 1 part in 10^{12}, and determine the redshift of a galaxy via its neutral hydrogen (HI) spectrum to perhaps 1 part in 10^4, yet still only estimate the flux density of a radio source to perhaps $\pm 10\%$, and its polarization parameters to even lower precision.

2 Basic Definitions

2.1 Intensity or Surface Brightness

At the risk of repeating what Paul said in the previous talk, the fundamental observable in radio astronomy is the INTENSITY of the radio waves arriving at Earth from a given direction. Neglecting the polarization properties of the waves, something Carl Heiles will address in considerable depth rather soon, intensity (or as it is often known, SURFACE BRIGHTNESS or even just plain BRIGHTNESS, B), is a function of four variables;

I(θ, ϕ, ν, t), where,

θ and ϕ are the spherical coordinates of a given direction on the two-dimensional sphere,

ν is the frequency,

and, t is time, i.e. the emission may be time variable.

Suppose that we just consider the energy arriving from a small solid angle, dΩ, in the sky, this being located in the direction, (θ, ϕ), and crossing an area, dA, normal to that direction. Then, if from that tiny cone an energy, dE, is received on dA in time, dt, and within a band of frequencies, dν, centered at frequency ν,

$$I(\theta,\, \phi,\, \nu,\, t) = B(\theta,\, \phi,\, \nu,\, t) = \lim_{dt,\, dA,\, d\nu,\, d\Omega \to \infty} \frac{dE(\theta,\, \phi,\, \nu,\, t)}{dt\ dA\ d\nu\ d\Omega} \tag{1}$$

Note that dE/dt is nothing but the power received from the small solid angle dΩ for a bandwidth dν by the area dA, so that I (= B) is the power received per unit area, per unit solid angle, per unit bandwidth from the direction (θ, ϕ). Its standard units are $\mathrm{W\,m^{-2}Hz^{-1}ster^{-1}}$.

2.2 Brightness Temperature

Rather than using these standard units, it is often more convenient to express the surface brightness as a BRIGHTNESS TEMPERATURE, $\mathrm{T_B}$. The physical significance of this is that were we able to replace our small element of sky, $d\Omega$, by a black body of temperature, $\mathrm{T_B}$ K, then at our observing frequency ν (and probably only at this frequency) we would measure the same intensity from this black body as from the real sky.

Let me stress that unless the piece of sky that we are observing is actually radiating like a black body, (and the Moon, the planets, and the 2.7 K cosmic microwave background are good approximations to this), then $\mathrm{T_B}$ is a function of frequency, and related to the Intensity, $\mathrm{I}(\theta, \phi, \nu, \mathrm{t})$, by the Rayleigh-Jeans approximation (unless the frequency is very high, and the temperature very low, when you will have to use the full Planck formula). Thus,

$$I(\theta,\,\phi,\,\nu,\,t) = \frac{2\,k\,T_B(\theta,\,\phi,\,\nu,\,t)}{\lambda^2} = \frac{2\,k\,\nu^2\,T_B(\theta,\,\phi,\,\nu,\,t)}{c^2} \tag{2}$$

where, λ is the wavelength, k is Boltzmann's constant, and c is the velocity of light.

2.3 Flux Density

Typically, we use our radio telescope to measure the distribution of brightness over a chosen area, and then use the data we record to produce a contour map or "radio photograph" of the radio emission there. From such images we find that the radio sky contains a multitude of discrete radio sources, each having well-defined boundaries, albeit impressed on a distributed background of emission. Clearly, for any frequency of measurement, ν, we would like to define a global parameter that characterizes the strength of the emission from any such discrete source. A suitable choice is the total power received by unit area per unit bandwidth from the whole source. Now to obtain this total power received, we need to integrate the intensity (i.e. surface brightness) over the total solid angle of the discrete source. This we call the FLUX DENSITY, $\mathrm{S}(\nu, \mathrm{t})$, of the source at frequency ν, and it is given by,

$$S(\nu,\,t) = \iint_{source} I(\theta,\,\phi,\,\nu,\,t)\;d\Omega \tag{3}$$

Hence, if we consider our tiny cone of sky as a discrete source,

$$\begin{aligned} S(\nu,\,t) &= I(\theta,\,\phi,\,\nu,\,t)\;d\Omega && (4)\\ &= \frac{dE(\theta,\,\phi,\,\nu,\,t)}{dt\;dA\;d\nu\;d\Omega} \times d\Omega && (5)\\ &= \lim_{dt,\,dA,\,d\nu\to\infty} \frac{dE(\theta,\,\phi,\,\nu,\,t)}{dt\;dA\;d\nu} && (6) \end{aligned}$$

The standard units of flux density are $\mathrm{W\,m^{-2}\,Hz^{-1}}$. However, the flux densities of all celestial radio sources are so small that a practical unit, the JANSKY, has been adopted, where, $1\,\mathrm{Jy} = 10^{-26}\,\mathrm{W\,m^{-2}Hz^{-1}}$. With the most sensitive telescopes available today, the typical flux densities measured are more conveniently expressed in units of milliJy ($1\,\mathrm{mJy} = 10^{-3}\,\mathrm{Jy}$) or even microJy ($1\,\mu\mathrm{Jy} = 10^{-6}\,\mathrm{Jy}$).

2.4 Distance Dependencies

If our small radio source is situated at a distance r, then by the inverse square law, the energy we receive from it is $\mathrm{dE} \propto r^{-2}$. However, the solid angle that it subtends is $d\Omega \propto r^{-2}$. Hence, $\mathrm{I} = \mathrm{dE}/(\mathrm{dt\,dA\,d}\nu\,\mathrm{d}\Omega)$ is distance independent.

In contrast, the flux density, $\mathrm{S} = \mathrm{dE}/(\mathrm{dt\,dA\,d}\nu) \propto r^{-2}$, and hence the flux density of a "standard candle" falls as the inverse square of its distance.

3 The Basic Radio Receiver

As everyone in the audience doubtless knows, the radio emissions from celestial bodies can be subdivided into two main spectral categories, either;

1. Broadband emission, known as continuum radiation, where the emission is present right across the radio spectrum. Typical examples of continuum emission would be that from planetary surfaces, HII regions, the Cosmic Microwave Background, and those Galactic and extragalactic sources that emit via the synchrotron process.
2. Line emission due to low-energy transitions within atoms and molecules in space. Here, the emission is localized to very specific frequencies in the radio spectrum.

The receivers for line observing are somewhat more complicated than are continuum receivers. Apart from Doppler correction to allow for the motion of the Earth, high frequency resolution is required to reveal the detailed shape of the line profile from a particular direction. Consequently, in this introductory talk I will restrict myself to talking about the basic total-power continuum receiver, and leave it to others to fill in the specific details of line systems later in the week.

3.1 The Super-Heterodyne Receiver

Given that the majority of sources we are likely to want to study are likely to only provide a power level of only $10^{-26} - 10^{-30}\ \mathrm{W\,m^{-2}\,Hz^{-1}}$, the signals that are collected by our antenna need to be amplified greatly before being recorded. However, we wish to achieve this amplification while adding as little extra noise as possible to the celestial signals, and avoiding significant instabilities.

In many ways, most radio-astronomy receivers resemble the super-heterodyne receiver found in your transistor radio. The very simplest of all is the total-power receiver.

Suppose that we are interested in measuring a band of frequencies centered around a frequency, ν_{RF} (where RF stands for Radio Frequency). The power arriving from our continuum source is, of course, very broadband, but the combination of our receiver horn (or feed), preamplifier, and perhaps the addition of a carefully designed electronic filter will select out for us the band in which we are interested.

The preamplifier is a very, very important component of our receiver as for many receivers it is this which provides the majority of the "artificial" noise against which we try to detect our celestial target source. Crudely put, the reason for this is that the noise produced by the the preamp is subsequently amplified by every other stage of the receiver!

Following the preamplifier comes the mixer whose job it is to change the central (or carrier) frequency of our signals to a (usually) lower frequency. This it must do without distorting the signal envelope which carries the astronomical information in which we are interested. This change of frequency is achieved by multiplying together the signals amplified by the preamp with a pure, strong, sinusoidal signal at a frequency ν_{LO} (where LO stands for Local Oscillator). The multiplication is achieved via a strongly non-linear device such as a Si crystal or a Schottky diode. The multiplication process generates sum and difference signals, of which only the difference signals are selected to be passed, this being the so-called INTERMEDIATE FREQUENCY (IF) signal. As the difference signal is selected;

$$\nu_{\mathrm{IF}} = \mid \nu_{\mathrm{RF}} - \nu_{\mathrm{LO}} \mid \tag{7}$$

You will notice that for any combination of values, ν_{IF} and ν_{LO}, there are two values of ν_{RF}, equi-spaced on either side of ν_{LO}, that will produce an output at a frequency of ν_{IF}. These are known as the upper and lower sidebands. Usually, the frequencies of the unwanted one of this "sideband pair" have already been "killed" by the frequency selective feed, preamp, filter combination ahead of the mixer. This is known as SINGLE SIDEBAND operation, and is usually the case at longer wavelengths, where the second sideband is likely to be filled with undesirable radio frequency interference. However, let me just add that double sideband operation may occasionally be preferred in the millimeter-wave bands.

You may well be wondering why this complicated operation of frequency changing is undertaken rather than merely amplifying the signal at ν_{RF} right down the receiver chain? Three of the most important reasons are as follows;

1. Traditionally, long lengths of cable to transit the signals from the telescope to the control room/electronics lab are less lossy for lower signal frequencies. Today, this is often of less importance than formerly due to optical fibers being used to transmit the signal.

2. The complete receiver chain will probably boost the received power by something like a factor of $10^6 - 10^7$. Should a little of the radiation from down the amplifier chain escape and be reabsorbed by the telescope feed, then if the signal has not had its frequency changed

early on in the signal path, a positive feedback loop could be established causing the receiver to oscillate.

3. If the LO frequency, ν_{LO}, can be varied over a wide range, and its value is chosen correctly, it is possible to use a single standard IF chain to service a wide range of receiver front-ends.

Following the chain of IF amplifiers, the signal is passed to the DETECTOR. This removes the carrier wave altogether, bringing the frequency of the band center down to DC (0 Hz). Usually, a square-law detector is employed, for which output voltage $\propto$ (input voltage)^2 $\propto$ input power. This operation is appropriate as it is the power of the astronomical signal collected by the antenna in which we are interested. Alternative strategies to using a square-law detector following the IF chain will be described later this week.

Next, the detected signal goes to an integrator which sums up the detector output voltage for a specified time in order to smooth out noise fluctuations, as I will detail in a short while. The simplest possible integrator is an R-C circuit.

Finally, the signals are recorded. Many years ago, this would have been done by a pen-on-paper X-Y chart recorder. Nowadays, it is much more likely that the integrated signals will be fed to an analog-to-digital converter (A->D converter), and the digitized signals recorded on computer disk for subsequent data processing.

4 Signal-to-Noise Considerations

4.1 System Noise

A number of independent factors contribute to the total level of noise power against which we try to study our target radio source. For example, at decimeter and longer wavelengths our own Milky Way is a source of considerable noise power against which our target is observed. The strength of the Milky Way emissions varies strongly both with direction in the sky and with frequency. In addition, all electronic devices generate noise internally and contribute to the unwelcome background noise.

At least conceptually, we can perform a very simple experiment to quantify the amount of noise that is generated inside our receiver itself. Suppose that we are able inject a variable amount of noise power into the input port of our preamplifier, and then measure the output power of the system as the input power is varied. The slope of the resulting plot of input power against output power is called the electronic GAIN of the system, and represents the power amplification factor of the system, often expressed in decibels (dB's), where,

$$\mathrm{Gain(dB)} = 10\ \log_{10}(\mathrm{dP_{out}}/\mathrm{dP_{in}}) \tag{8}$$

One further thing is striking from the input power against output power plot. For zero input noise power, a finite level of output power is still recorded! It is as if $\mathrm{P_R}$ amount of power is being generated within the receiver. This excess power can be quantified by expressing it as the temperature, T_R, of a matched resistor connected to the receiver input port, where,

$$P_R = k\, T_R\, \Delta\nu \tag{9}$$

and $\Delta\nu$ is the receiver bandwidth, and k is Boltzmann's constant.

A good present-day receiver at 1.4 GHz, with its front end cooled cryogenically to a very low temperature, would typically have a receiver noise temperature of $T_R < 10\,K$.

4.2 Cascaded Amplifiers

I stressed just now the importance of having a good quality preamp if we wish to achieve the lowest possible T_R. Why should this be? Well, our receiver chain consists of N amplifiers, each with its own gain, g_i, and noise temperature, T_i. Clearly, the gains are multiplicative, so that the total system gain, g, is given by,

$$g = g_1 \times g_2 \times g_3 \times \ldots \times g_N \tag{10}$$

Further, the contribution to the total receiver temperature from each succeeding amplifier down the receiver chain is reduced by the cumulative gain of all the amplifiers ahead of it. Hence,

$$T_R = T_1 + \frac{T_2}{g_1} + \frac{T_3}{g_1 g_2} + \frac{T_4}{g_1 g_2 g_3} + \ldots \tag{11}$$

As each $g_i \gg 1$, clearly, $T_R \approx T_1$, and the first-stage amplifier contributes much more to the total receiver noise temperature than any other stage, and needs to be of the finest quality.

4.3 Lossy Components

An associated situation situation is where a lossy component is introduced into the receiver system. For now, I'll just consider the case where there is loss in the feed system before the preamp, say a mediocre feeder cable or loss in a device such as a directional coupler, a filter, a circulator, or a hybrid ring. If the power loss between the horn and the preamp is by a factor of $(1 - \eta)$, where η is the feeder efficiency, and if the receiver noise temperature at the preamp input is T_R, then the effective noise temperature of the complete receiver system (referred to the horn) becomes,

$$T'_R = (\frac{1}{\eta} - 1)\, T_\circ + \frac{T_R}{\eta} \tag{12}$$

where, $T_\circ$ is the ambient temperature of the feeder components. As $\eta < 1$, $T'_R > T_R$. Clearly, one desires as little loss as possible before the preamp. Indeed, there should be good reasons for everything that is placed ahead of the preamp, (and often there is.) Also to be noted is that cooling the lossy components before the preamp helps, as $T_\circ$ is lowered.

An example of a component that does need to be placed as near the front of the receiver chain as possible is the directional coupler through which a known and constant amount of noise power can be injected into the system for calibration of the temperature scale and system temperature of an observation. Typically, this power is generated by a carefully calibrated noise diode that can be switched on and off at will. If the effective temperature of the noise step impressed on the recorded power level by firing the noise diode is known, then clearly calibration in K can be easily achieved. At millimeter wavelengths, an effective calibrated noise step can be obtained by placing an absorber at room temperature over the horn, followed by an absorber cooled (say) in a liquid nitrogen bath. Essentially one is performing thermometry.

4.4 The System Temperature

Going a step further, it is often not enough to just take the receiver noise, plus a correction for feeder loss, as being the total noise against which we observe a radio source. Usually there are extra contributions to the total noise power from such things as;

a) emission from the Earth's neutral atmosphere.

b) radiation from the ground picked up in the far (back) sidelobes.

c) celestial noise contributions from, (*i*) the background radiation from our own Milky Way galaxy, (*ii*) the 2.7 K radiation of the universal Cosmic Microwave Background (CMB) left over from the birth of the Universe, and (*iii*) perhaps a contribution from the noise power of the source that we are observing itself!

The sum of all these noise contributions (remembering that noise powers add) is known as the SYSTEM TEMPERATURE, T_{sys}, and can be expressed as,

$$T_{sys} = T_R + \underbrace{T_{atm} + T_{ground} + T_{gal} + T_{CMB} + T_{source}}_{T_A} \tag{13}$$

where sum of all contributions from outside the observing system can be called the ANTENNA TEMPERATURE, T_A.

In respect of the celestial noise contributions, the interstellar contribution from the Milky Way galaxy plus the CMB changes greatly with frequency. For example,

- at 38 MHz, $T_{gal} \geq 10^4 - 10^5\,K$
- at 327 MHz, $T_{gal} \sim 20 - 2,000\,K$
- at 5 GHz and above, T_{gal} is negligible, and is dwarfed by $T_{CMB} = 2.7\,K$

4.5 The Radiometer Equation

Finally, we are in a position to answer the question as to how weak a source we can detect with a given Rx. Firstly, I will quote (but not here derive) the single most important relationship when it comes to planning a radio-astronomical observation. This is the so-called RADIOMETER EQUATION.

The radiometer equation states that for our simple total-power receiver, the rms noise ripple against which we are trying to detect our target source can be expressed in units of antenna temperature as,

$$\mathrm{T_{rms}} = \frac{\mathrm{T_{sys}}}{\sqrt{\Delta\nu\ \tau}} \tag{14}$$

where, $\Delta\nu$ is the receiver bandwidth (in Hz), and τ is the integration time (in sec)

Now we are able to define the signal-to-noise ratio which we will measure when we move our telescope beam on and off a source which gives a rise in antenna temperature of $\Delta\mathrm{T_A}$, integrating for a time $\tau/2$ in each phase. This is,

$$\frac{\Delta\mathrm{T_A}}{\mathrm{T_{rms}}} = \frac{\Delta\mathrm{T_A}\sqrt{\Delta\nu\ \tau}}{2 \times \mathrm{T_{sys}}} \tag{15}$$

The factor of 2 arises as, while we are observing for a time τ, we are only recording the signals from the source for half the time, but we are recording noise for the whole time. (Note this assumes that the source does not contribute significantly to $\mathrm{T_{sys}}$.) Usually a signal-to-noise ratio of $\sim$5 is considered necessary for a significant detection.

Often, however, we are interested in a point source's flux density rather than its antenna temperature. Now Paul just showed that the power received by a telescope of effective collecting area, $\mathrm{A_{eff}}$, from a point source of flux density, S, in a single polarization is $1/2\,\mathrm{S\,A_{eff}}\,\Delta\nu$. The same power would be received were the feed to be replaced by a matched resistor of temperature, $\Delta\mathrm{T_A}$. In other words,

$$\frac{1}{2}\,\mathrm{S\,A_{eff}}\Delta\nu = \mathrm{k}\Delta\mathrm{T_A}\Delta\nu \tag{16}$$

or,

$$\Delta\mathrm{T_A} = \frac{\mathrm{SA_{eff}}}{2k} \tag{17}$$

Further, it often simpler when considering sensitivity in terms of flux density to express $\mathrm{T_{sys}}$ in terms of the point-source flux density that would exactly double the system temperature. This is called the System Equivalent Flux Density (SEFD), and is given by

$$\mathrm{SEFD} = \frac{2\,\mathrm{k}\,\mathrm{T_{sys}}}{\mathrm{A_{eff}}} \tag{18}$$

Using this, the radiometer equation can be written as,

$$\mathrm{S_{rms}} = \frac{\mathrm{SEFD}}{\sqrt{\Delta\nu\ \tau}} \tag{19}$$

4.6 Confusion

From the previous discussion you might think that if we were to keep on making receivers with lower $\mathrm{T_{sys}}$'s, broadening our bandwidth, and lengthening our integration times, we then would be able to detect incredibly weak sources. Sadly, it is often the case that there is an ultimate limit imposed by the size of our telescope beam. Suppose that we integrate our continuum image of a celestial field deeper and deeper. Eventually we might pass the sensitivity where we have (say) one source for each 100 telescope beam areas. If we continue to integrate down, at some point we would reach a sensitivity that should give one source per 10 beam areas, and eventually we would reach a sensitivity where we would expect an average of one source per beam area. However, at that point your sky might look a little as follows.

Clearly, all is not well here. Many apparent sources would really be blends of many individual objects, and as we continue to integrate to get below this level, the ripple on our scans is no longer set by the system noise, but by the distribution of sources in the sky, and our own beam pattern. In fact, continued integration no longer changes the appearance of the scan significantly. The observation is now deep in the regime where it is referred to as being CONFUSION LIMITED (as opposed to being sensitivity limited, where the noise on the scan is set mostly by the system). Jim Condon will be saying more in his talk on continuum observing as to how confusion can set a lower flux density limit for continuum surveys. In fact, confusion is especially a problem for single-dish continuum observing at low frequencies where sources are bright, and beams are broad.

5 Single-Dish Imaging and Spatial Sampling

Very often, we wish to image the brightness distribution of a region of sky with our single dish. The popular "On-the-Fly" method of mapping makes a raster of parallel scans by moving the telescope beam across the field of interest, separating adjacent scans by a fixed interval in celestial coordinates. We can now pose the question as to how we should space our scans (and also as to how often we should sample spatially along a scan) such as to lose no information which our telescope is capable of receiving? We can answer this by considering the one dimensional case.

Paul demonstrated that when a telescope of normalized beam pattern, $\mathrm{P}(\theta)$, observes a distribution of brightness temperature, $\mathrm{T_B}(\theta)$, the antenna temperature recorded is given by,

$$\mathrm{T_A}(\theta') = \frac{1}{\Omega_\mathrm{A}} \int \mathrm{T_B}(\theta)\, P(\theta' - \theta)\, d\theta \tag{20}$$

Now, as Paul noted, this is a convolution, and the Fourier Transform of a convolution is equal to the multiplication of the Fourier Transforms of the two functions being convolved. If the sky brightness distribution and beam pattern are expressed in radians, then their Fourier transforms are distributions in so-called spatial frequencies, given in units of cycles per radian. The spatial frequencies in the Fourier transform of the beam pattern of a telescope of size D, observing at a wavelength λ, is always zero beyond a value of $|\mathrm{D}/\lambda\,|$) cycles/rad. In other words, the beam of a single-dish telescope acts as a "low-pass Fourier filter" of the sky brightness distribution.

Returning to our problem of how often to sample our scans, sampling at discrete points is equivalent to multiplying the recorded antenna temperatures, $\mathrm{T_A}(\theta')$, by a set of delta functions separated by the sampling interval, s radians. This is equivalent to convolving the Fourier transform of $\mathrm{T_A}(\theta')$, (which remember is zero outside $\pm$ D/λ cycles/rad), by a series of delta functions spaced by $1/s$ cycles/radian. This produces a series of "islands" that are not overlapping, i.e. we retain full information, only if,

$$\frac{1}{s} > \frac{2\,\mathrm{D}}{\lambda} \tag{21}$$

Or in other words,

$$s < \frac{\lambda}{2\mathrm{D}} \tag{22}$$

If we sample more sparsely than this limit, then the "islands" overlap in the spatial-frequency plane, producing "aliasing", and losing information. Thus, in preparing your mapping proposal for a single dish, you should sample along your scans, and space those scans, at the critical sampling interval or finer, or be very aware of the risks that you are taking by not doing so!