Radio Astronomical Data Acquisition

John Ball[1]

Preface

These are introductory notes intended for students who know a little about physics and perhaps electrical engineering and who want to learn something about data-acquisition techniques in radio astronomy. There is no attempt at either rigor or completeness; most of these topics deserve at least a chapter in a radio-astronomy book.

Characteristics of the signals

Electromagnetic radiation: Rays, waves, or photons?

Wavelength and frequency ranges; what's λ, what's ν?

The radio-frequency range is arbitrarily defined to be from frequencies (ν) beginning at the super-sonic, meaning just above the range of sound waves that humans can hear, 20 kHz or so, up to about 600 GHz, where the far infrared begins. The corresponding wavelength (λ) range is from many kilometers down to about 0.5 mm or 500 microns. Frequency and wavelength are related, as for any conventional wave phenomena, by $\nu = c/\lambda$, where c is the speed of light, 299792.5 km/s.

Over this frequency range and at moderate equivalent temperatures, we can usually ignore photons; that is, we need not quantize the electromagnetic field. In most cases, this results in both simplification and insight that is lacking if one thinks photons instead of waves. (What's ν? Well, c/λ but not usually E/h.)

Many formulas used in radio astronomy (and electrical engineering) are based on the Rayleigh-Jeans approximation to the Planck black-body law. If, furthermore, the antenna structures are large compared to a wavelength, as is usually the case at centimeter and millimeter wavelengths, then we can often ignore waves also and use just ray-trace optics.

---

[1] MIT Haystack Observatory. Revised: 2005 May 23.

Law of large distances

Another helpful simplification involves the large distances and small angular sized of most astronomical objects. Angles in radians are, then, just the linear sizes divided by the distances.

Noise-like

Law of large numbers

Most astronomical sources are large in physical size, even though small in angular size, and radiation is emitted by a large number of statistically independent sources—atoms, molecules, or electrons. The resulting signals are noise-like, that is, the electric fields are Gaussian random variables with spectra that depend on the details of the emission mechanisms. Laboratory laser and maser oscillators are usually coherent because of the cavity in which they oscillate. Astronomical masers have never been found to have non-Gaussian statistics.

For similar reasons, many radio-astronomical sources are unpolarized; that is, signals in one polarization are statistically independent of signals in the orthogonal polarization. But some sources, especially those involving magnetic fields that extend over a significant part of the spatial extent of the source, can be polarized, and studying such polarization sometimes leads to significant insights.

Law of large sizes

As a rough rule with some exceptions, an incoherent source that is not resolved in angle can be seen to vary only on time scales long compared with the light travel time through the length of the emitting region. A black sphere with cyclic temperature variations, for example, will have these variations smoothed over as seen from a large distance if the cycle time is comparable to or less than the light travel time across a radius of the sphere. Some astronomical sources vary, and the time scales of the variations can sometimes be used to infer maximum sizes.

Significance for measurement techniques

Some of the techniques used in radio astronomy depend on these characteristics of the sources. One-bit autocorrelation spectroscopy, for example, depends on the signal voltage being a Gaussian random variable. And many-hour-long interferometry depends on the source being stable over that time.

Spectra

Continuum sources, black body or ν^n?

A black body at temperature T emits a brightness or specific intensity $B_\nu(T)$ (power per unit area and per unit frequency interval) given by the Planck black-body law,

$$B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/kT}-1}$$

where h is Planck's constant, k is Boltzmann's constant, k = 1380 Jy m^2/K, a Jansky, Jy, is 10^{-26}w/(m^{2}Hz), ν is frequency, and c is the speed of light. The Rayleigh-Jeans approximation for $h\nu/kT \ll 1$, which is almost always true in radio astronomy, is

$$B_\nu(T) = \frac{2\nu^2}{c^2}kT = \frac{2kT}{\lambda^2}$$

where λ is wavelength. (B_ν is sometimes written I_ν.) The thermal noise power per unit frequency interval from a resistor, also for $h\nu/kT \ll 1$, is

$$P_\nu = kT$$

This proportionality to temperature allows us to talk about intensities and powers in temperature units, °K or Kelvins. If a source really were thermal emission from a black or gray body, then this radiation temperature would be independent of wavelength. More typically, however, sources are "colored," that is, radiation temperatures vary with wavelength and are not necessarily related to physical temperatures of the sources. A synchrotron-emission source, for example, has a spectral index n, usually defined by $I_\nu \propto \nu^{-n}$ where I_ν is the specific intensity (per unit frequency interval) and where n is related to the energy distribution of the emitting electrons but is typically $0.5 < n < 1$. Thermal emission from a black or gray body would have $n = -2$ (from $B_\nu(T)$ above). Regardless of the emission mechanism or spectrum, we can use the

Rayleigh-Jeans equation above to *define* a brightness temperature, T_b, proportional to intensity, but then T_b will be, in general, a function of frequency. Even in the high-frequency low-temperature range where Rayleigh-Jeans is no longer useful, we can use this equation to define a convenient fake brightness temperature. Figure 1 is a cartoon example of such a spectrum.

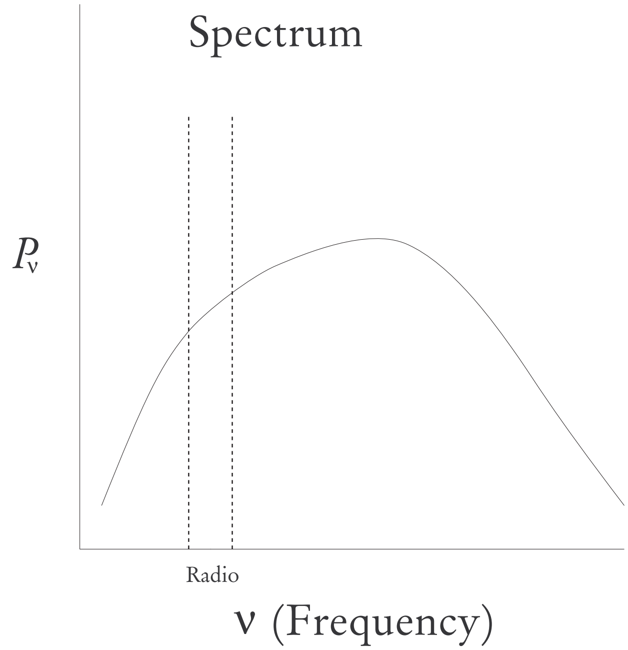

Figure 1

Electrical engineers implicitly use the Rayleigh-Jeans approximation also. The noise power per unit bandwidth from a warm resistor into a matched load is just kT provided that the frequency and temperature are in the range for which Rayleigh-Jeans is accurate. To see this, imagine a resistor coupled to a feed looking at a black body. Even in the range where Rayleigh-Jeans is no longer useful, in thermal equilibrium, equal power must flow each way, and the power from the resistor must follow Planck rather than Rayleigh-Jeans.

Atoms and molecules—emission and absorption lines

Emission and absorption lines in radio astronomy usually originate from atoms and small molecules or molecular ions in gaseous form, and molecular transitions at radio wavelengths are usually rotational. Emission lines result from warm gas overlying a colder background so that the intensity (or flux or radiation temperature) at the line frequency is sharply higher compared to nearby wavelengths. Absorption lines result from cool gas overlying a hotter background source so that the intensity at the line frequency is sharply lower compared to nearby wavelengths. If a gas cloud with temperature T_g and opacity τ (nepers) covers an opaque continuum source with temperature T_c then

$$T_a = T_c e^{-\tau} + T_g (1 - e^{-\tau})$$

where T_a is the brightness temperature seen at the antenna. Figure 2 is a cartoon of this situation. The opacity of the gas, τ, is nearly zero except near the line frequency. If $T_g > T_c$, we see emission at the line frequency, if $T_g < T_c$, we see absorption. If such a gas cloud is optically thick (i.e., $\tau > 1$ on the line) or opaque, then the specific intensity or brightness at the line frequency is given by the state temperature of the corresponding transition, which, at thermodynamic equilibrium, would be just the temperature. Thermodynamic equilibrium is, however, not very common in celestial sources.

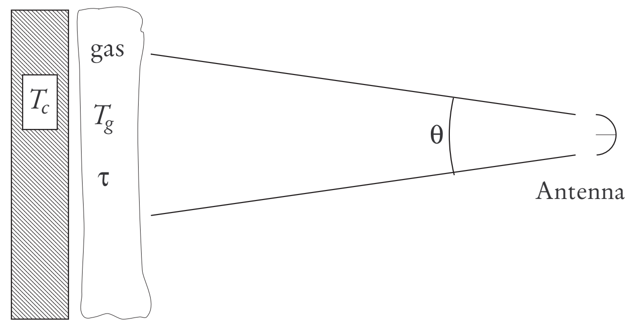

Figure 2

Figure 3 is a cartoon of such spectra. The upper spectrum in this figure is for $T_c = T_{c1} > T_g$ (absorption), and the lower spectrum is for $T_c = T_{c2} < T_g$ (emission).

Doppler shifts and kinematics

Doppler shifts are very important in spectral-line radio astronomy. The non-relativistic form is usually written as

$$\frac{\Delta \nu}{\nu} = -\frac{v}{c}$$

where $\Delta \nu$ is the change in frequency ν due to the Doppler velocity v, defined as the rate of change of distance from source to observer (hence the minus sign), and c is the speed of light. (This is an unfortunate choice of symbols since ν and v look so much alike especially in Times font.) Even when speeds are relativistic, this non-relativistic formula is sometimes still used to *define* a convenient fake velocity.

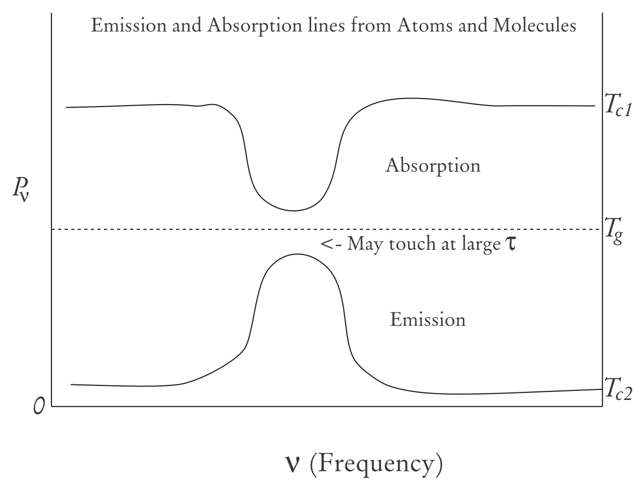

Figure 3

We observe from a moving location—the antenna on Earth—but since we know about Earth's motions (e.g., rotation and revolution), we can subtract them from the observed spectra to leave Doppler shifts characteristic of the sources.

Line widths and line shapes are influenced by several line-broadening mechanisms including *a*) natural line widths related to the lifetimes of the states involved in the transition, *b*) kinematic temperatures characterizing the small-scale random motions of the atoms or molecules, *c*) turbulence or larger-scale random motions, and *d*) kinematics, by which we mean large-scale ordered motions such as expansions, contractions, or rotations. A useful exercise is to estimate and sketch the spectrum that would be seen from some simple kinematic models such as a circumstellar shell or sphere that is expanding, contracting (infalling), or rotating but is not resolved in angle. In some cases the central star ionizes nearby gas, which makes a central continuum source. Such an object can show both emission and absorption features separated in velocity or width.

Antennas at radio wavelengths

Parabolic—why?

Radio-astronomical sources are far away, so incoming signals often look like plane waves from a specific direction (*point sources*), and the first goal of a radio telescope is to catch as much energy as possible from such a wave and avoid as much as possible any other signals, especially local interference. The signal at the antenna in this case can be characterized by a flux density in Janskys (1 Jy = 10^{-26} w/m^2/Hz), so the bigger the antenna, the more watts (well, pico-pico watts) we collect. A parabolic antenna (i.e., a parabola of revolution), which puts all this energy into a small spot where a *feed* can be placed, is usually an engineering optimum for centimeter and millimeter wavelengths. The small antenna at the focus of a larger parabolic antenna is traditionally called a feed by analogy with a radar antenna where the feed "feeds" radiated power to the main antenna. Many of the properties of an antenna system used for transmitting (e.g., beamwidth, gain) are the same as for receiving—this is the reciprocity theorem.

Aperture efficiency and K/Jy

Consider a black sphere with diameter d and temperature T at a distance r from a circular receiving antenna with diameter D. Assume that r is much larger than either d or D. Figure 4 is a cartoon of this situation. Then the power density (power per unit frequency interval) received by the antenna, P, can be calculated as its collecting area times the flux from the source at the antenna or as the specific intensity of the source times the solid angle of the antenna as seen from the source. Either way gives the same formula, namely,

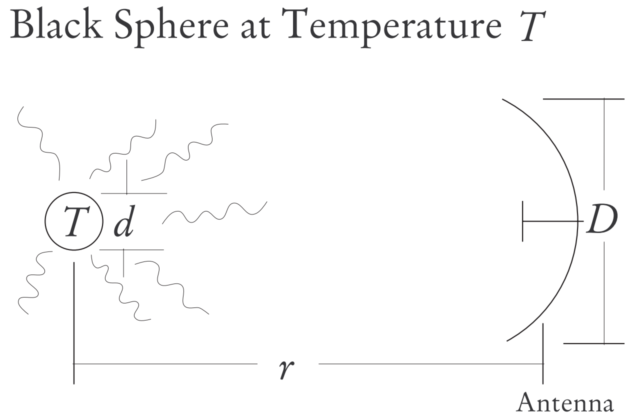

Figure 4

$$P = I \times \frac{\pi d^2}{4} \times \frac{1}{r^2} \times \frac{\pi D^2}{4}$$

The first three terms on the right combine to make up the source flux density, F, and $A = \pi D^2/4$ is the antenna's area, except that the effective collecting area is smaller than its physical area due to various losses. If we characterize P in temperature units, $P = 2kT_R$, as usual, then

$$\frac{T_R}{F} = \frac{A}{2k}$$

is an antenna's figure of merit, sometimes called *sensitivity*, typically in Kelvins per Jansky. That extra 2 is because the flux density refers to the total in both polarizations, but a single receiver can receive only one polarization. The ratio of the antenna's collecting area from this formula to its physical area is called *aperture efficiency*, usually expressed as a percentage, and usually 60% or less.

Beam efficiency and beam dilution

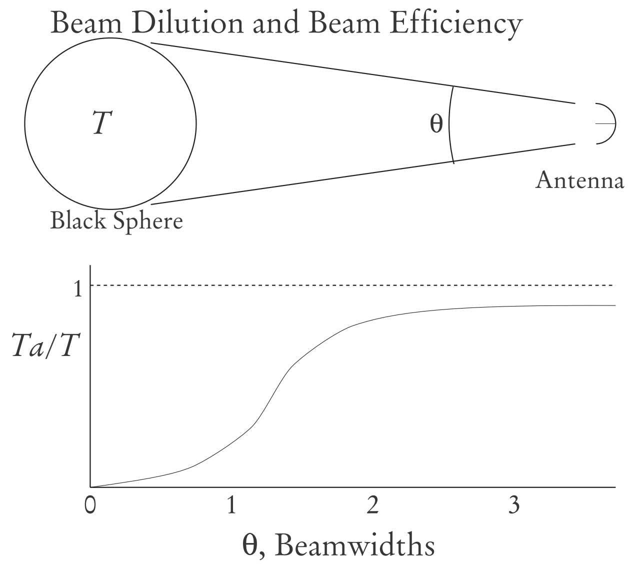

Figure 5

Another figure of merit, appropriate for sources extended in angle, is the *beam efficiency*, crudely defined as the ratio of T_R to the brightness temperature of the source. (There are more precise but less useful definitions.) Beam efficiency by this definition is, however, a function of the source angular size and shape, alas. A more useful parameter is the *beam dilution*, defined in the same way but for an assumed circular source with a specified diameter in beamwidths and as a function of this diameter. Planets with known brightness temperatures and angular sizes are candidate calibration sources for measuring aperture efficiency when they are small in angle compared with a beamwidth or points on the beam-dilution curve when they are larger in angle. Figure 5 is a cartoon of this situation.

Cassegrain—why?

A Cassegrain antenna comprises a parabolical main reflector and a concave hyperbolical *subreflector* near the *prime focus* of the main reflector to reflect incoming signals back to a spot near the center of the main reflector, the *secondary focus*, where the feed is placed. This feed is usually mounted on the front of a receiver box that fits through a hole in the center of the main reflector. This arrangement trades a little additional aperture blockage (the subreflector is larger than a prime-focus feed would be), for the ability to place additional equipment, such as a cryogenic refrigerator, near the secondary focus.

Beamwidth: λ/D

The *beamwidth* (full width to half power) of the radiation pattern of a circular antenna is approximately $1.2\lambda/D$ in radians, where λ is the wavelength and D is the diameter of the antenna in, of course, the same units. The 1.2 factor and the details of the beam shape (angular resolution function) depend somewhat on the feed illumination pattern, that is, on the pattern of the feed as seen from the main reflector. A circular antenna with circular illumination gives a circular beam.

When a finite antenna is used to map extended sources over a range of angles, one can show that the resulting maps are *band limited* in that they contain no angular frequencies above a maximum, called the Nyquist limit, that depends only on the wavelength and the antenna's diameter. Band-limited maps are, then, smooth continuous functions of two angles on the sky, but they can be specified or measured at a finite grid of evenly spaced points (finite number of pixels) provided that these points are no farther apart than a Nyquist step, which is $\lambda/(2D)$ in radians. A Nyquist is typically a little less than half a beamwidth. Smooth maps can be obtained from such finite grids of points by convolving by Bessel function $J_0(r)$.

Requirements for surface precision

The effective collecting area of an antenna is less than its physical area because of various losses, one of which is due to the departure of the surface from an ideal parabolic shape. An antenna with a surface that is rough on the scale of a wavelength will be almost useless because of low aperture efficiency and also susceptibility to interference scattered into the feed. An ideal antenna would have at least so-called 20^{th}-wave optics, meaning that the surface is within a 20^{th} of a wavelength of perfect. We must sometimes make do with antennas less than ideal; all antennas have some short-wavelength limit based on this criterion.

Interferometers

Why?

Resolution: λ/D

Two or more antennas looking at the same source at the same time can have their signals combined into an interferometer to give some of the information that would be obtained from a single antenna with a diameter equal to the spacing between the interferometer antennas. The resolution of a two-antenna interferometer is approximately λ/D (no 1.2), where D is now the spacing between antennas, but this resolution is only in the direction parallel to a line connecting the antennas and only for sources in a plane perpendicular to this line.

Data from many antennas or data taken with many different locations of one or several of the antennas can produce an array of fringe data whose Fourier transform is a map of the intensity distribution of a source on the sky. This is called *aperture synthesis* because we have synthesized an equivalent antenna aperture whose diameter is comparable to the baseline spacings among interferometer antennas.

Interferometers can be used to help solve two separate classes of problems: If we know the locations of our antennas on Earth and the various motions of Earth (e.g., rotation and revolution), then we can use interferometers to learn about angular locations, angular sizes, and angular structures of astronomical objects—this is astronomy. If, instead, we already know about such objects on the sky, then we can use interferometers to learn about the locations of our antennas on Earth and about Earth's motions—this is geodesy.

Imagine that you are at the location of the source on the sky and looking back toward a pair of interferometer antennas on Earth. Imagine a plane perpendicular to your line of sight on which you plot the projected baseline with the convention that one end of the baseline is always plotted at the origin of coordinates. At each moment of time, plot two points corresponding to having each end of the apparent baseline at the origin. These two points will be, of course, just opposite each other with the origin between. The distance from the origin to each point is the foreshortened baseline length, and the vector angle is that of the projected baseline in source-oriented coordinates, typically right ascension (*RA*) and declination (*Dec*). Figure 6 is a cartoon of such a situation.

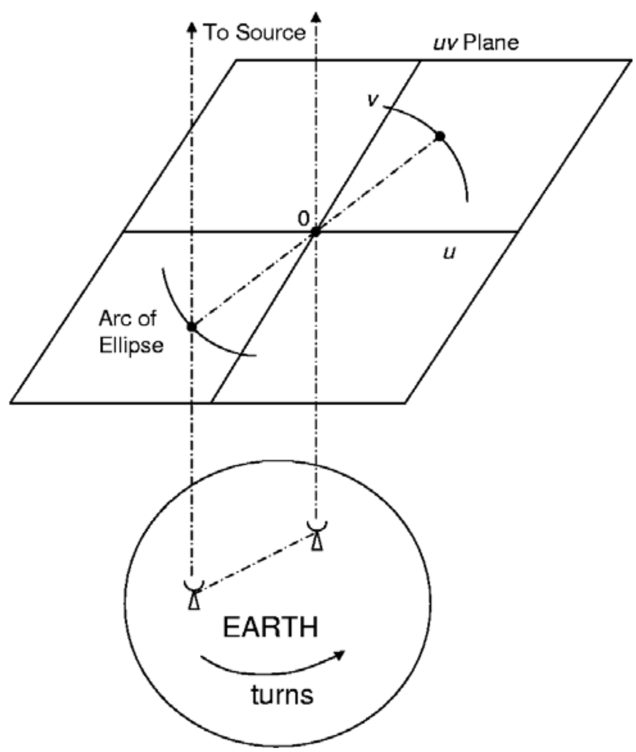

Figure 6

As time passes, Earth turns, and these points move and trace out two ellipses—either whole ellipses in 24 hours if the source is circumpolar (i.e., never sets) or arcs of ellipses that start when the source rises and end when the source sets as seen from the antennas. These ellipses depend on the locations of the antennas (especially the baseline as a vector) on Earth and also depend on the location of the source (especially its declination) on the sky. With axes labeled in wavelengths (rather than, say, meters), this is the famous *uv* plane, much discussed by radio astronomers.

A single measurement with a single pair of antennas gives a single datum to be placed at two points in the *uv* plane. The data along an ellipse obtained over a day can be combined provided that the source is unchanging over the time interval (typically hours) of the observations. The number of baselines and the number of these elliptic-arc pairs is $n(n-1)/2$, where n is the number of antennas. More antennas give more elliptic arcs, often crossing and so providing valuable redundant information, but unless the antennas actually touch or overlap, there will always be gaps or holes with no data in the *uv* plane.

Under most conditions, we can produce the data that would be obtained from a multi-antenna interferometer by moving one of a pair of antennas to various locations on Earth and observing for, say, a day at each location. This works provided that the source is unchanging over the time interval (days) of these observations. Many multi-antenna radio interferometers have moveable antennas on, for example, railroad tracks, in order to obtain more complete coverage of the *uv* plane than could otherwise be obtained without building more antennas.

The *uv* plane is roughly comparable to the surface of an equivalent single antenna, and the inevitable gaps in the *uv*-plane coverage correspond to an antenna or mirror with holes. These produce equivalent beam patterns with non-symmetric shapes and ugly sidelobes, that is responses to sources outside the main beam.

The raw data from a radio-interferometer measurement are fringe amplitudes and phases (fringes as complex numbers) or a generalization thereof called the cross-spectral function for each measured point in the *uv* plane. Data in this form can be obtained from the radio receivers in several ways. In one method, cross correlations of the signal-voltage waveforms are first calculated:

$$R_{12}(\tau) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} V_1(t)\, V_2(t-\tau)\, dt$$

where V_1 and V_2 are the voltages, t is time, and τ has units of time and is called delay or lag. This limit and integral are normally replaced by a finite sum since only a finite number of voltages can be sampled. Then the cross-spectral function is the Fourier transform of the cross correlation,

$$S_{12}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_{12}(\tau)e^{-i\omega\tau}d\tau$$

(again usually as a finite sum), where $\omega = 2\pi\nu$ represents signal frequency usually shifted down to video. $S_{12}(\omega)$ contains all the interferometer information and also the power spectrum of the source(s) within the bandpass.

There are simpler schemes that do not involve cross correlation. A multiplying interferometer is a special case of a cross-correlating interferometer in which the cross correlation is performed at only one value of lag. If, for example, we multiply the two voltages and average over a time long compared with a cycle (e.g., a second), then, as Earth turns, we can see fringes in time with periods typically of some seconds. Estimating fringe amplitudes and phases from such data plotted on, say, a chart recorder, can be done by hand. An even simpler scheme involves adding the two voltages and then square-law detecting. This produces the product together with some constant terms to be ignored. That is

$$(V_1 + V_2)^2 = V_1^2 + V_2^2 + 2V_1V_2$$

where V_1V_2 is the desired product. The two squared terms are proportional to power and presumed constant except for noise. A simple adding interferometer combining two Small Radio Telescopes (SRTs) looking at the Sun is described in http://web.haystack.edu/SRT/SRTI.pdf.

An actual or equivalent cross correlation of the signal voltages is usually done in special-purpose computer hardware but, for modest data rates, can also be calculated by a program running in a general purpose computer such as a PC. A cross correlation produces a lower-noise estimation of fringe amplitudes and phases compared to the simpler schemes, which are wasteful in that some information is not used.

After the data are obtained and gridded onto the *uv* plane, then a two-dimensional Fourier transform yields the intensity distribution (map or contours) of the source on the sky as seen by the equivalent antenna pattern. Because of the holes or gaps with no data in the *uv* plane, this initial intensity distribution, called a *dirty map*, is usually pretty ugly. There are helpful data-reduction techniques (such as Clean or MEM) that amount to model fitting or to interpolating or extrapolating across these gaps. But beware: These are tricks based on assuming that we know something about the source that was not actually measured.

Another outstanding advantage of multi-antenna interferometers is that the data are usually processed so as to produce multi-pixel maps—so-called aperture synthesis—the equivalent of a multi-feed array on the equivalent single antenna. An interferometer map can, in principle, show everything inside the beam of the individual antennas that make up the interferometer.

Connected vs VLBI

Interferometers with antennas spaced up to, say, a few kilometers usually have cables or fibers connecting the antennas to a central site where cross correlations are calculated. Such interferometers are called *connected* because they require at least one cable connecting the telescopes to the correlator. Fringes are created and can be seen in real time. By contrast, very-long-baseline interferometry (VLBI, http://web.haystack.edu/vlbi/vlbi.html), sometimes called tape-recorder interferometry, records signals together with precise (sub-microsecond) timing at each of two or several radio telescopes. The necessary time and frequency standard at each telescope typically comprises a hydrogen-maser atomic clock. Since they are not actually connected to the correlator, VLBI telescopes can be arbitrarily far apart—hence *very-long* in the name.

After recording, the VLBI magnetic tapes or disks are shipped to a central processing station where they are played, and cross correlations are calculated, typically in a hardware correlator, at a time that may be days or weeks later. This delay is a disadvantage, especially for troubleshooting, but the baseline lengths among telescopes and the resulting angular resolutions are limited only by the diameter of the Earth, and there is almost no other way to do large data rates on megameter baselines with their milliarcsecond resolutions.

Recent developments in VLBI include recording on disks rather than tapes (Mark5, http://web.haystack.edu/mark5/) and sending data to correlators using Internet or high-speed data links—so-called eVLBI for *electronic* VLBI (http://web.haystack.edu/eVLBI/)—which can produce fringes in almost real time. But this is not the same as a connected interferometer because each station needs its own time and frequency standard.

Figure 7 is an example *uv*-plane coverage for a realistic VLBI. Each elliptic arc is plotted with a letter corresponding to a single baseline, that is a single pair of antennas. This figure represents 21 baselines among 7 antennas. Each arc ends when the source sets at one of the two antennas.

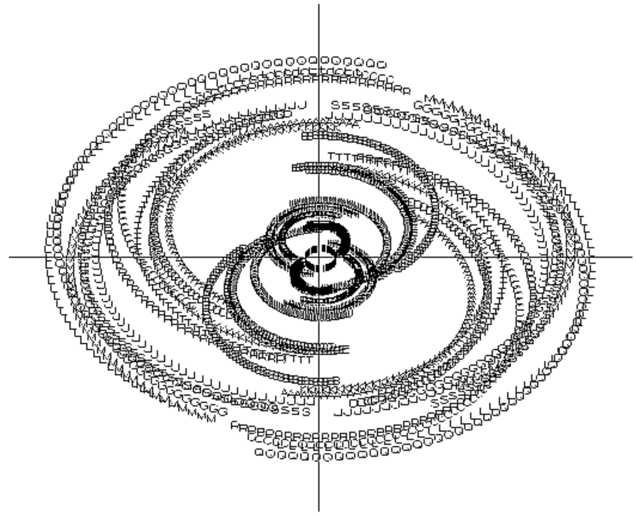

Figure 7

Receivers

Characterization—temperature units

Figure 8 is a schematic of a minimalist receiver for continuum radio astronomy. After the antenna, the first stage of the receiver, the low-noise amplifier (LNA), is probably the most important component of a radio telescope. Since the signals are so weak, the noise performance of the receiver is crucial, and this leads to extraordinary efforts, such as cryogenic cooling, to reduce noise in the LNA. The noise performance of radio-astronomy receivers is usually characterized by an equivalent system temperature, T_{sys} (in Kelvins), referred to the feed or even to outside Earth's atmosphere. Using temperature units for the system allows direct comparison with source temperatures. Typical system temperatures are ten to a hundred K for centimeter wavelengths or up to several hundred K for millimeter and submillimeter wavelengths. These numbers should drop as technological progress is made.

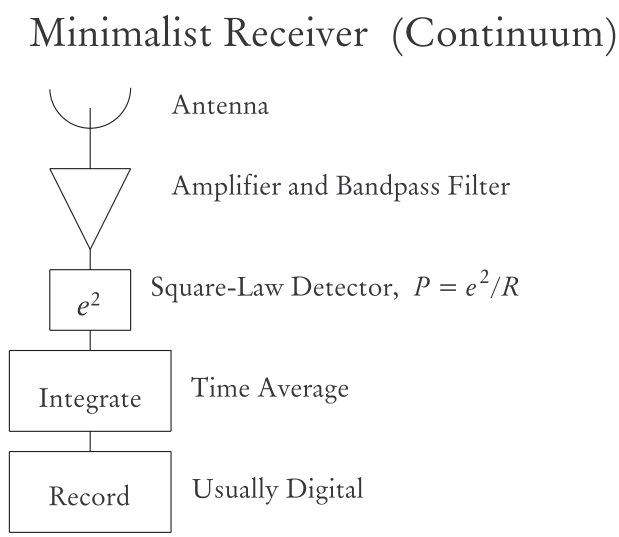

Figure 8

The noise equations

The usual rms noise calculations in radio astronomy are based on

$$\Delta T = \frac{\alpha \gamma T_{sys}}{\sqrt{\beta t}}$$

where T_{sys} is the system temperature in Kelvins, β is the noise bandwidth, which is approximately equal to the resolution in spectroscopy or the total usable bandwidth in continuum, and t is the total integration time (on plus off) for normal switched observing. Set α to 2 for ordinary single switching: That 2 is the product of two $\sqrt{2}$, one for spending half the time on source, another for differencing two equally noisy measurements. The correlation quantization correction, γ, is approximately 1.16 for Haystack's spectrometer with its modified 3×3 multiplication table. Set γ to 1 for continuum. Then ΔT is the rms fluctuation in the corresponding measurement. The βt in the denominator of this equation is, in effect, the number of samples averaged to make the measurement.

Why heterodyne?

Most receivers used in radio astronomy (almost all receivers used for spectroscopy) employ so-called superheterodyne schemes. The goal is to transform the frequency of the signal (SF) down to a lower frequency, called the intermediate frequency (IF) that is easier to process (e.g., filter) but without losing any of the information to be measured. This is accomplished by mixing the SF from the LNA with a local oscillator (LO) and filtering out any unwanted sidebands in the IF. A bonus is that the SF can be shifted around in the IF, or alternatively, the IF for a given SF can be shifted around by shifting the LO.

Why square-law detectors?

Inside radio-astronomy receivers, the signal begins as a voltage proportional to the electric field as collected by the antenna. But we normally want to measure power or power density. So, at least for continuum measurements and for calibration, we need a device that produces an output proportional to the square of the voltage, a so-called square-law detector, and also averages over at least a few cycles of the waveform.

Spectrometers and spectroscopy

Scanning filter

Filter shape, resolution

To measure spectral-line emission or absorption from molecules or atoms, we need a device to measure power spectra—a spectrometer. An intuitive method to measure power spectra is to scan a narrow tunable bandpass filter across the frequencies to be measured and record its power output as a function of frequency. A variant of this scheme, actually used in some spectrum analyzers, has a fixed filter in the IF that is, in effect, scanned by scanning an LO in a heterodyne configuration. Figure 9 shows an example of such a configuration. Variations in the power spectrum narrower than the width of the scanning filter are smoothed over and lost. The width and shape of this filter characterize the spectrometer's resolution. This scheme works but is wasteful because all the information outside the instantaneous position of the filter is ignored.

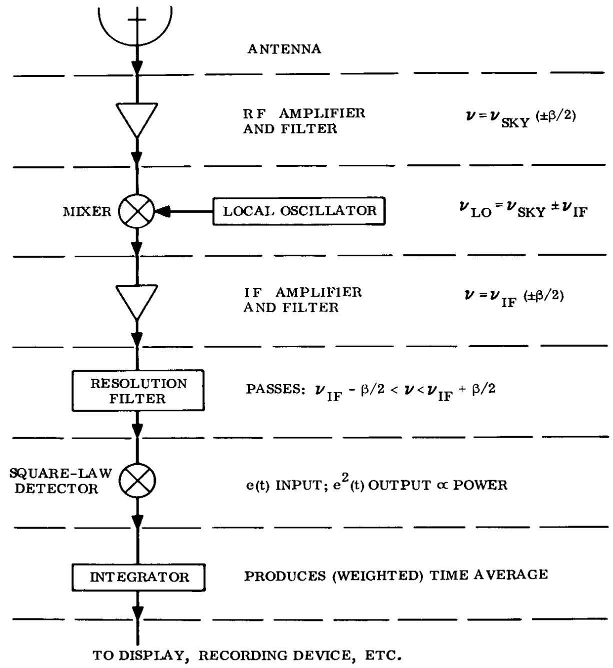

Figure 9: Block diagram of a scanning-filter receiver.

Comb of filters—filter bank

A significant improvement in observing efficiency results from having a comb or bank of bandpass filters placed side-by-side in frequency and recording all their outputs simultaneously. Figure 10 shows an example of such a filter bank. Choosing filter shapes and spacings for

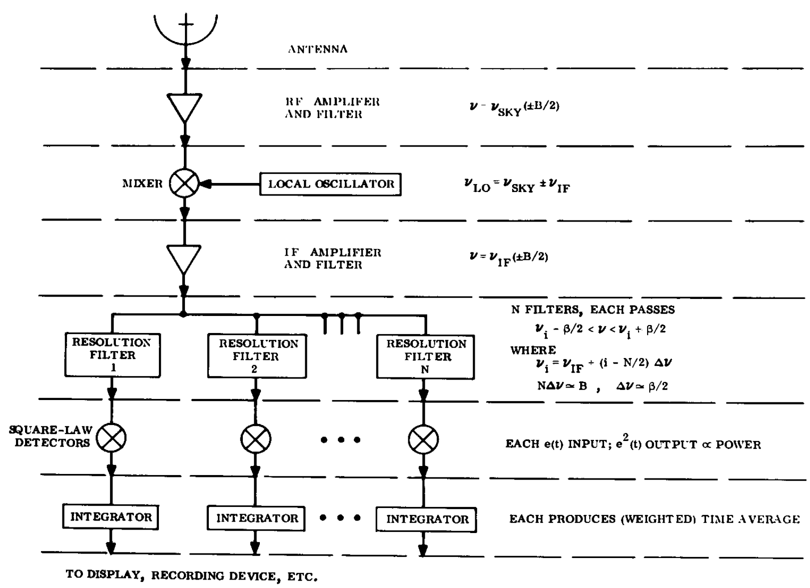

Figure 10: Block diagram of a multichannel filter receiver.

such a spectrometer is, however, not intuitive. The popular almost-square filters placed just touching, for example, give spectra that are difficult to interpret whenever spectral features are comparable to the filter widths. With today's technology, filter banks are expensive and troublesome compared with various digital alternatives.

Autocorrelations and Fourier transforms
Why?

Some authors define the power spectrum to be the Fourier transform of the autocorrelation of the voltage and then show that this definition accords, at least approximately, with the intuitive scanning filter. If autocorrelations are done for lags, τ, up to some maximum, τ_{max}, and Fourier transforms are done with no weighting, then, except for noise considerations, the resulting power spectra will be the same as would be obtained with a scanning filter whose shape is $\sin(2\pi\nu\tau_{max})/(2\pi\nu\tau_{max})$. If the spectra to be measured are band limited, perhaps by a preceding lowpass filter, then, by the Nyquist theorem, autocorrelations can be done at uniform finite lag steps, $\tau_s = 1/(2\nu_{max})$, where ν_{max} is the maximum frequency in this band. This sampling corresponds to two data per cycle of ν_{max}. The resulting spectra are also band limited because they contain no lags above τ_{max}, which is the Nyquist limit. The band-limited spectra from an autocorrelation spectrometer are, then, smooth continuous functions of frequency, and they can be specified at a finite set of evenly spaced points provided that these points are no farther apart than a Nyquist step, $\nu_{step} = 1/(2\tau_{max})$. The number of these frequency steps is the same as the number of lag steps, namely $2\nu_{max}\tau_{max}$, which sometimes leads to confusion. Smooth continuous spectra can be obtained from finite sets of points by convolution by the same $\sin(2\pi\nu\tau_{max})/(2\pi\nu\tau_{max})$.

The point of using autocorrelations to get power spectra and of quantizing both autocorrelations and spectra is to allow these operations to be done digitally. The autocorrelations are usually done in hardware, the Fourier transforms in software. This is usually a significant simplification compared to a filter bank.

Yet another alternative involves doing direct Fourier transforms (direct FFTs) on finite-length samples of the voltage. With modern high-speed computers, this is sometimes feasible even in software.

Multibit vs one bit

A further simplification is possible because of the nature of the signals to be measured. For Gaussian random noise, autocorrelations and power spectra can be computed from one-bit (just sign) samples of the voltage. Figure 11 shows an example of such a configuration. The price to pay for this simplification is about 31% additional noise and a little more computer arithmetic to correct the one-bit autocorrelations before the Fourier transform. Haystack now uses 1.5-bit (i.e., three-level) sampling, which adds about 16% additional noise (γ in the noise equation above) and is only a little more complex than one-bit.

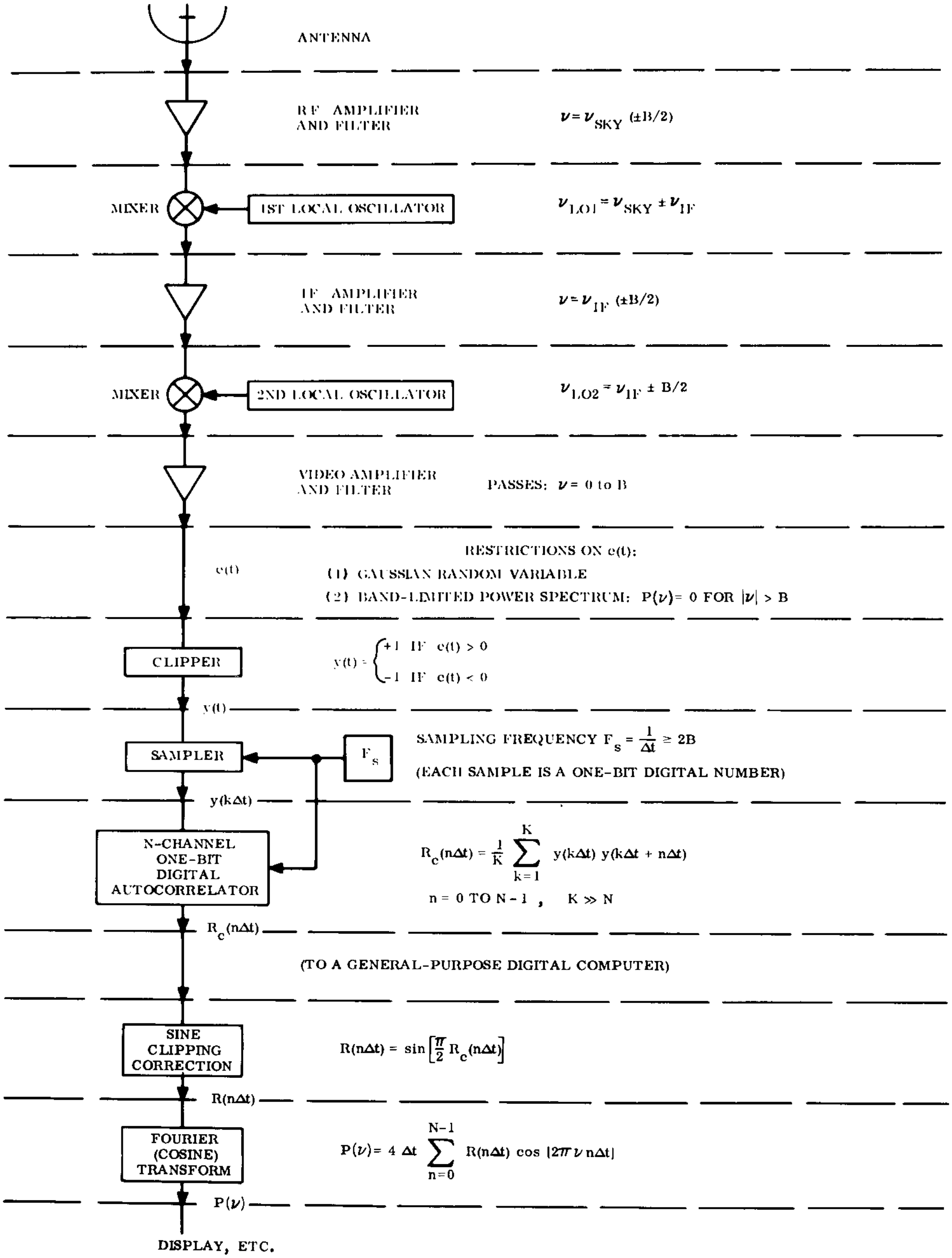

Figure 11: Block diagram of a one-bit digital autocorrelation receiver.

Switching schemes and baselines

Why switch?

Essentially all measurements in radio astronomy are made by subtracting the data between two situations: source present and source absent. Taking data with the antenna pointing toward the source and toward nearby empty sky is an example. The purpose is to reduce instrumental effects that otherwise overwhelm the very weak signals that we are trying to measure.

One of the most persistent and difficult problems in spectral measurements in radio astronomy involves the difficulty of obtaining good flat baselines—the parts of spectra with no signal. The corresponding baseline problem with continuum measurements involves stable data off source (e.g., cold sky) to subtract from on-source data. There are numerous instrumental effects that contribute to bumpy baselines, and many of the effects are a substantial percentage of the system temperature and are typically much larger than the signals to be measured. To reduce the severity of such instrumental effects, almost all radio-astronomy measurements are made using one of several possible switching schemes, usually called Dicke switching after R.H. Dicke, who first used this method. The ideal switching scheme would have the source itself turn off and on synchronously with a prescribed periodicity and with nothing else changing. Then the difference between signal and comparison would be precisely the desired measurement. Provided that T_s does not change between on and off source, one can show that noise is minimized by spending half the time on source, half off.

Among practical switching schemes, we can move the antenna pointing on and off the source either by actually moving the antenna or by offsetting the effective pointing by moving the feed or its image. But this works only if the source is confined in angle. Or we can move the source in and out of the passband by moving the LO frequency. But this works only if the source is confined in frequency. Or we can switch the input to the receiver alternately from the feed to an absorber or load. This last scheme is less desirable because many of the instrumental effects to be ameliorated are in the feed and beyond.

Switching schemes are almost always needed for either continuum or spectral measurements. The price to be paid for switching is increased noise for a given integration time. There are two contributions to the added noise: The receiver spends only half the time looking at the signal, and the answer is the difference between two equally noisy measurements. The result is twice the noise compared with not switching for the same integration time (this is α in the noise equation above), or four times the integration time to achieve the same noise.

Calibration

Why do radio astronomers spend so much time calibrating, and why is it such a messy complicated procedure? If you need to measure a voltage, just pick up a voltmeter, connect the probes to the right place, and see the answer on the display. Or, how about measuring a weight? Just put it on the scale and read the answer. Measurements in radio astronomy are much more complicated because the powers to be measured are minuscule—down in the noise, as we say—and because the measurements must be made through Earth's atmosphere, which is a variable attenuation. Radio-astronomy antennas are usually pushing the state of the art in mechanical design, and they are subject to distortions due, for example, to bending under their own weight as they point around the sky. And the receivers are as advanced in low-noise performance as effort and money can provide, which sometimes means unstable in gain or frequency response.

Optical astronomers have very similar problems, and their traditional solution is to compare a star or other celestial object to be measured with another nearby calibration star with known characteristics. This is a bootstrapping operation, of course, since someone somehow must have measured that calibration star, and it must be stable in time.

This scheme works for radio astronomy also, but calibration sources are fewer and hardly ever in the same field of view, hence calibration in radio astronomy usually must be a separate operation.

Using a small-angular-diameter source on the sky that has a known flux—a calibration source—allows one to measure directly the system-equivalent flux density (SEFD), which is the (only) factor needed to convert raw measured numbers (counts or volts) toward an unknown source to a flux density for that source:

$$\mathrm{SEFD} = \frac{2kT_{sys}}{A}$$

where k is Boltzmann's constant, T_{sys} is the system temperature, and A is the effective collecting area of the antenna or the aperture efficiency times the physical area. Since $2k/A$ is the factor to convert from Kelvins to flux units, SEFD is the system temperature expressed in flux units. Calculate SEFD by dividing a calibration source's flux by the fractional change in signal level (volts or counts from a square-law detector) from off to on source. Then, on the target source, multiply SEFD times the fractional change in signal to get this source's flux. But SEFD is a function of pointing elevation (because of attenuation in Earth's atmosphere and also antenna characteristics), weather, various receiver parameters that can vary with time, and, of course, wavelength.

Pointing back and forth between the target source and a calibration source at the same elevation is a good, if sometimes wasteful, calibration technique that allows one to ignore all the engineering variables that might be causing the SEFD to vary and so concentrate on doing astronomy. Measuring SEFD does not give T_{sys} or A separately, only their ratio.

The equivalent parameter for sources that are extended in angle is the system-equivalent brightness temperature (SEBT):

$$\mathrm{SEBT} = \frac{T_{sys}}{\eta_b}$$

where η_b is the beam dilution as discussed above. The SEBT is, however, a function of all the same variables as SEFD plus also the angular size of the source. Given a calibration source with a known brightness temperature and the same angular size as the target source (not too likely), one can use the same comparison as with a small-diameter source.

Because of such difficulties in using direct methods, most calibrations in radio astronomy are done using a two-part method that involves measuring antenna sensitivity (K/Jy) or aperture efficiency (for small-diameter sources) or antenna beam dilution (for extended sources) as a function of elevation and also a system temperature, T_{sys}, which is a characteristic of the receiver including re-emission from losses in the antenna feed and the atmosphere.

Sometimes one can justify assuming that the antenna sensitivity (or gain or aperture efficiency) as a function of elevation and wavelength is a fixed curve (the so-called gain curve, actually a surface), measure it once, and use it for subsequent observations. Then calibrating involves measuring system temperatures, a much quicker and easier task, and, depending on which system temperature is measured, estimating also the atmospheric attenuation from weather parameters. Some of these techniques are discussed below.

The raw output of a radio-astronomy radiometer is usually a voltage or counts representing a noise power as processed by a square-law detector. When the antenna is pointing away from any sources on the sky, this output represents the system temperature, T_{sys}, but in strange units rather than Kelvins. Calibration, in the simplest case, just involves measuring the conversion factor between these raw units, voltage or counts, to Kelvin degrees.

Some radio-astronomy receivers provide a noise-calibration source comprising a noise diode or a noise tube connected through a directional coupler (or equivalent) into the signal path usually between the feed and the first amplifier. The amount of noise added by turning on this noise source needs to be measured first by comparing with a standard noise source, usually a matched resistor with a variable physical temperature. This measurement is usually done by receiver engineers (they simultaneously measure their amplifier's noise figure), and it needs to be constant over the time since these primary noise measurements. This measured number, assumed constant, is called T_{cal}.

If v_{sys} is the voltage or counts from the square-law detector when pointing toward empty sky and v_{cal} is the additional voltage when this calibration noise is switched on (so that the total voltage is $v_{sys} + v_{cal}$), then the conversion from volts to Kelvins is T_{cal}/v_{cal} and the system temperature is

$$T_{sys} = \frac{T_{cal}}{v_{cal}} v_{sys}$$

Figure 12 is a cartoon of an imaginary chart recorder showing such a calibration sequence.

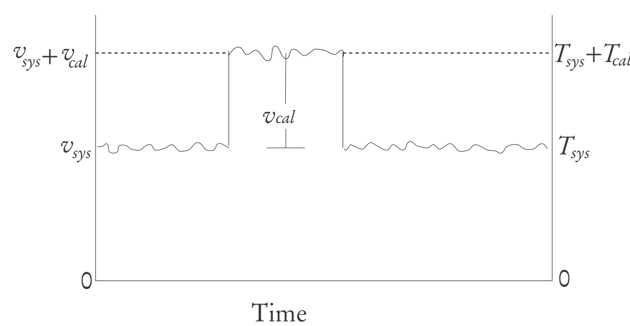

Figure 12

An alternative scheme, which actually measures a different system temperature called T_{sys}^*, uses an absorber, called a vane or chopper wheel, that covers the feed and supplies ambient-temperature noise. With this vane in place, the voltage from the square-law detector corresponds to an input noise temperature of $T_{vane} + T_{amp}$, where T_{vane} is the physical temperature of the vane (usually ambient temperature) and T_{amp} is that *part* of T_{sys} attributed to the receiver amplifier. So if we wanted to calculate T_{sys} using the vane, we would need a different formula. If, however, we use the same formula as above (with *vane* for *cal*), and if we assume that the atmosphere (and other losses, such as in the feed and radome) are at this same temperature, T_{vane}, then we get

$$T_{sys}^{*} = \frac{T_{vane}}{v_{vane} - v_{sys}} v_{sys}$$

which, with these assumptions, can be shown to equal the ordinary T_{sys} multiplied by the atmospheric loss (plus the warm part of other losses), so, T_{sys}^{*} is, at least approximately, the value to use to refer our measurements to outside Earth's atmosphere. There are small additional corrections because of the difference in temperature between the vane and the average temperature of the atmosphere at a scale height of the attenuation, but, when using the vane calibration in this way, do not use the normal atmospheric-attenuation correction. Figure 13 is a cartoon of an imaginary chart recorder showing such a calibration.

Calibrating Spectra

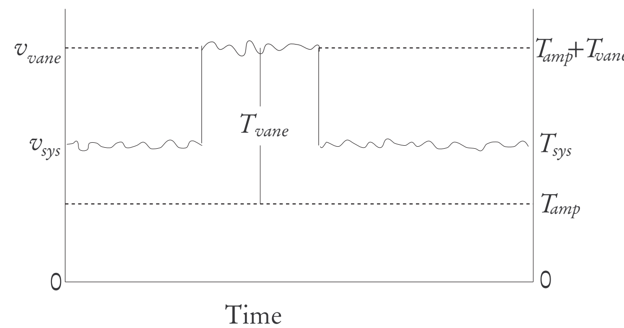

Figure 13

The result of measuring a spectrum in radio astronomy is usually a plot with apparent flux density or apparent brightness temperature on the ordinate and frequency or Doppler velocity on the abscissa. There are two aspects of properly calculating and labeling the ordinate of such a plot. The first aspect is just the same as calibrating a continuum measurement; we need to know the system temperature in flux or brightness-temperature units. With a filter-bank spectrometer, we could just do the continuum calibration scheme above for each filter separately. With an autocorrelation spectrometer, we need an additional calibration step. Imagine a spectral feature whose strength is constant but whose frequency moves about within the passband of the spectrometer. A proper calibration procedure should produce the correct amplitude for this feature regardless of where in the bandpass it falls. In addition to any gain variations with frequency in the spectrometer itself, the portion of the receiver prior to the spectrometer contains filters that shape the bandpass. The calibration procedure, then, must compensate for these spectral variations.

There are two useable calibration procedures, referred to as $(S-R)/R$ and $(S-R)/C$, based on two different assumptions about the behavior of the receiver. The $S-R$ in these pseudoformulas refers to signal minus reference in the switching scheme as discussed above. The denominator is intended to be a normalized gain function that represents the gain variation described above. To use the R formula, we assume that the system temperature referred to the front of the receiver is not a function of frequency across the spectrometer's passband, therefore the shape of this passband (R) is proportional to the gain variations that we want to compensate for. To use, instead, the C formula, we measure this gain by assuming that the calibration noise is not a function of frequency across the spectrometer's passband, but this is not simple for typical autocorrelators because they are deliberately insensitive to level changes. The R formula is more popular than the C formula because R involves fewer steps, and the assumption about system temperature is usually quite reliable. Using the C formula takes more time away from actual observing.

Thanks for using the facilities of MIT Haystack Observatory.

References and recommended reading

An Introduction to Radio Astronomy, B.F. Burke and F. Graham-Smith, Oxford University Press, Oxford, UK, 1997

http://www.haystack.mit.edu/

Interferometry and Synthesis in Radio Astronomy, 2nd edition, A. Richard Thompson, James M. Moran, and George W. Swenson, Jr., Wiley-Interscience, New York, 2001.

Methods of Experimental Physics, Volume 12, *Astrophysics*, Part B, *Radio Telescopes*, and Part C, *Radio Observations*, M.L. Meeks, ed., Academic Press, New York, 1976

Methods in Computational Physics, Volume 14, *Radio Astronomy*, B. Adler, S. Fernback, and M. Rotenberg, eds., Academic Press, New York, 1975

Radio Astronomy, 2nd edition, John D. Kraus, Cygnus-Quasar Books, Powell, Ohio, 1986

Tools of Radio Astronomy, 2nd edition, K. Rohlfs and T.L. Wilson, Springer-Verlag, Berlin, 1996

Problems

1) Calculate the beamwidth of the Haystack antenna at the 86.24-GHz line of SiO. Compare this with the diameter of the full moon.

2) Calculate the angular resolution of a VLBI from Haystack to Effelsberg, Germany at the same frequency. (Hint: The distance you need is not the great-circle distance but rather the straight line through Earth.)

3) From Haystack's web pages, find the spectrometer's resolution for the standard setup with one video converter and a bandwidth of 160 MHz and also one-third and one-ninth of this bandwidth. Calculate the bandwidth/resolution ratio for the three cases. Why isn't this number constant?

4) Calculate the rms noise (in Kelvins) expected at Haystack for a continuum measurement at 86 GHz with a bandwidth of 160 MHz and one minute of effective integration. (Hint: Get an approximate system temperature from Haystack's Web pages.)

5) Calculate the rms noise for the same conditions but using the spectrometer as in 3) above. Try to find and use the proper spectrometer clipping correction.

Problems revised: 2004 May 26, JAB

Lecture 3 Problems

John Ball

1) Calculate the beamwidth of the Haystack antenna at the 86.24-GHz line of SiO. Compare this with the diameter of the full moon.

The appropriate formula is 1.2 λ/D, where 1.2 is an approximation for a typical feed, λ is the wavelength, D is the diameter of the antenna, 120 feet, and the answer is in radians. Calculate λ from $\lambda = c/\nu$, where c is the speed of light, 299792.5 km/s and ν is the frequency, 86.24 GHz. Get

$$\lambda = \frac{299792.5\times10^3\,\mathrm{m/s}}{86.24\times10^9\,\mathrm{Hz}} = 3.48\times10^{-3}\,\mathrm{m} = 3.48\,\mathrm{mm}$$

Then

$$\frac{1.2\lambda}{D} = \frac{1.2\times3.48\times10^{-3}\mathrm{m}}{120\,\mathrm{ft}}\times\frac{\mathrm{ft}}{12\,\mathrm{inch}}\times\frac{\mathrm{inch}}{2.54\,\mathrm{cm}}\times\frac{\mathrm{centi}}{10^{-2}}$$

or

$$\frac{1.2\lambda}{D} = 1.14\times10^{-4}\mathrm{radian} = 6.53\times10^{-3}\mathrm{degree} = 23.5\mathrm{arcseconds}$$

The diameter of the full moon is approximately 0.5 degree, so this beamwidth is about a 77th of a full moon.

2) Calculate the angular resolution of a VLBI from Haystack to Effelsberg, Germany at the same frequency. (Hint: The distance you need is not the great-circle distance but rather the straight line through the earth.)

From ftp://dopey.haystack.edu/dist/pc-sched/antenna.sch (pointed to from pc-sched, etc.), the geocentric XYZ coordinates of the two sites in meters are

EFLSBERG	4033949	486989	4900431
HAYSTACK	1492407	-4457267	4296882

Then the straight-line distance is $D = \sqrt{(\Delta X^2 + \Delta Y^2 + \Delta Z^2)} = 5591903$ m and $\lambda/D = 6.217 \times 10^{-10}$ radians $= 3.56 \times 10^{-8}$ degrees $= 1.28 \times 10^{-4}$ arcseconds (using λ from problem 1 above). There is, of course, no 1.2. This resolution applies only when the source is just perpendicular to this line joining the sites.

3) From Haystack's web pages, find the spectrometer's resolution for the standard setup with one video converter and a bandwidth of 160 MHz and also one-third and one-ninth of this. Calculate the bandwidth/resolution ratio for the three cases. Why isn't this number constant?

From http://web.haystack.edu/37m/files/Introduc.pdf, section C.3, table 2, or from http://web.haystack.edu/Haystack.info/R.A.Observing.pdf, page 5,

Bandwidth, MHz	Resolution, kHz	Ratio
160	377.1	424
53.3	62.85	849
17.8	5.237	3395

Because of speed limitations in the hardware, the number of lags that the spectrometer can calculate varies depending on the sampling frequency, which is just twice the bandwidth (Nyquist sampling), and the bandwidth/resolution ratio is proportional to the number of lags, which is noted in the table cited.

4) Calculate the rms noise (in Kelvins) expected at Haystack for a continuum measurement at 86 GHz with a bandwidth of 160 MHz and one minute of effective integration. (Hint: Get an approximate system temperature from Haystack's Web pages.)

From http://web.haystack.edu/Haystack.info/R.A.Observing.pdf, page 6, the appropriate formula is

$$\frac{\alpha \gamma T_{sys}}{\sqrt{\beta \tau}}$$

where $\alpha = 2$, $\gamma = 1$ for continuum or 1.16 for spectroscopy, and T_{sys} is approximately 300°K from http://web.haystack.edu/37m/files/Introduc.pdf, section B, table 1, $\beta = 160$ MHz, and $\tau = 1$ minute. Under the radical, β is the noise bandwidth and τ is the effective integration time. For this case, get

$$\frac{2\times300°\mathrm{K}}{\sqrt{160\times10^{6}\,\mathrm{Hz}\times60\,\mathrm{s}}} = 0.0061°\mathrm{K}$$

5) Calculate the rms noise for the same conditions but using the spectrometer. Try to find and use the proper spectrometer clipping correction.

From the spectrometer table cited above, the resolution with 160-MHz bandwidth is 377.1 kHz, and this is an approximate noise bandwidth β. So get

$$\frac{2\times1.16\times300°\mathrm{K}}{\sqrt{377.1\times10^{3}\,\mathrm{Hz}\times60\,\mathrm{s}}} = 0.146°\mathrm{K}$$

but, of course, that is for each point on the spectrum.

Answers revised: 2004 May 26