Fundamentals of Radio Interferometry

Rick Perley, NRAO/Socorro

Fourteenth NRAO Synthesis Imaging Summer School
Socorro, NM
Topics

• Why Interferometry?
• The Single Dish ... as an interferometer
• The Basic Interferometer
  – Response to a Point Source
  – Response to an Extended Source
  – The Complex Correlator
  – The Visibility and its relation to the Intensity
  – Picturing the Visibility
Why Interferometry?

• Because of Diffraction: For an aperture of diameter D, and at wavelength λ, the image resolution is

\[ \theta_{rad} \approx \frac{\lambda}{D} \]

• In ‘practical’ units:

\[ \theta_{arcsec} \approx 2 \frac{\lambda_{cm}}{D_{km}} \]

• To obtain 1 arcsecond resolution at a wavelength of 21 cm, we require an aperture of ~42 km!

• The (currently) largest single, fully-steerable apertures are the 100 meter antennas near Bonn, and at Green Bank.

• So we must develop a method of synthesizing an equivalent aperture.

• The methodology of synthesizing a continuous aperture through summations of separated pairs of antennas is called ‘aperture synthesis’.
The Single Dish – as in Interferometer

• A parabolic reflector has a power response (vs angle) roughly as shown below.
• The formation of this response follows the same laws of physics as an interferometer.
• A basic understanding of the origin of the focal response will aid in understanding how an interferometer works.

Illustrated here is the approximate power response of a 25-meter antenna, at $\nu = 1$ GHz.
The Parabolic Reflector

- Key Point: Distance from incoming phase front to focal point is the same for all rays.
- The E-fields will thus all be in phase at the focus – the place for the receiver.
• An antenna’s response is a result of coherent vector summation of the electric field at the focus.
• First null will occur at the angle where one extra wavelength of path is added across the full width of the aperture:

$$\theta \sim \frac{\lambda}{D}$$

(Why?)
Specifics: First Null, and First Sidelobe

- When the phase differential across the aperture is 1, 2, 3, … wavelengths, we get a null in the total received power.
  - The nulls appear at (approximately): \( \theta = \lambda/D, 2\lambda/D, 3\lambda/D, \ldots \) radians.
- When the phase differential across the full aperture is \( \sim 1.5, 2.5, 3.5, \ldots \) wavelengths, we get a maximum in total received power.
  - These are the ‘sidelobes’ of the antenna response.
  - But, each successive maximum is weaker than the last.
  - These maxima appear at (approximately): \( \theta = 3\lambda/2D, 5\lambda/2D, 7\lambda/2D, \ldots \) radians.

![Antenna Power Response at 1 GHz](image)
Interferometry – Basic Concept

- We don’t need a single parabolic structure.
- We can consider a series of small antennas, whose individual signals are summed in a network.
- This is the basic concept of interferometry.
- Aperture Synthesis is an extension of this concept.
Quasi-Monochromatic Radiation

- Analysis is simplest if the fields are perfectly monochromatic.
- This is not possible – a perfectly monochromatic electric field would both have no power ($\Delta \nu = 0$), and would last forever.
- So we consider instead ‘quasi-monochromatic’ radiation, where the bandwidth $\delta \nu$ is very small.
- For a time $dt \sim 1/d\nu$, the electric fields will be sinusoidal.
- Consider then the electric fields from a small solid angle $d\Omega$ about some direction $\mathbf{s}$, within some small bandwidth $d\nu$, at frequency $\nu$.
- We can write the temporal dependence of this field as:

$$E_\nu(t) = A \cos(2\pi \nu t + \phi)$$

- The amplitude and phase remains unchanged to a time duration of order $dt \sim 1/d\nu$, after which new values of $A$ and $\phi$ are needed.
Simplifying Assumptions

- We now consider the most basic interferometer, and seek a relation between the characteristics of the product of the voltages from two separated antennas and the distribution of the brightness of the originating source emission.
- To establish the basic relations, the following simplifications are introduced:
  - Fixed in space – no rotation or motion
  - Quasi-monochromatic (signals are sinusoidal)
  - No frequency conversions (an ‘RF interferometer’)
  - Single polarization
  - No propagation distortions (no ionosphere, atmosphere …)
  - Idealized electronics (perfectly linear, no amplitude or phase perturbations, perfectly identical for both elements, no added noise, …)
Symbols Used, and their Meanings

• We consider:
  – two identical sensors, separated by vector distance $b$
  – receiving signals from vector direction $s$
  – at frequency $\nu$ (angular frequency $\omega = 2\pi\nu$)

• From these, the key quantity

$$\tau_g = \frac{b \cdot s}{c}$$

is formed. This is the ‘geometric time delay’ – the extra time taken for the signal to reach the more distant sensor.

• Finally, the phase corresponding to this extra distance is defined:

$$\Theta = \omega \tau_g = \frac{2\pi b \cdot s}{\lambda}$$
The Stationary, Quasi-Monochromatic Radio-Frequency Interferometer

Geometric Time Delay

\[ \tau_g = \mathbf{b} \cdot \mathbf{s} / c \]

\[ V_1 = E \cos[\omega(t - \tau_g)] \]

\[ V_2 = E \cos(\omega t) \]

\[ R_C = P \cos(\omega \tau_g) \]

\[ A^2 \left[ \cos(\omega \tau_g) + \cos(2\omega t - \omega \tau_g) \right] \]

The path lengths from sensors to multiplier are assumed equal!

Rapidly varying,

Unchanging
Pictorial Example: Signals In Phase

2 GHz Frequency, with voltages in phase:
\[ b.s = n\lambda, \text{ or } \tau_g = n/\nu \]

- Antenna 1 Voltage
- Antenna 2 Voltage
- Product Voltage
- Average

Time (ps)
Pictorial Example: Signals in Quad Phase

2 GHz Frequency, with voltages in quadrature phase:
\[ b.s = (n +/\ 1/4) \lambda, \ \tau_g = (4n +/ 1)/4v \]

- **Antenna 1 Voltage**
- **Antenna 2 Voltage**
- **Product Voltage**
- **Average**
Pictorial Example: Signals out of Phase

2 GHz Frequency, with voltages out of phase:
\[ b.s = (n +/\ - \frac{1}{2})\lambda \quad \tau_g = \frac{(2n +/\ - 1)}{2\nu} \]

- Antenna 1 Voltage
- Antenna 2 Voltage
- Product Voltage
- Average
Some General Comments

• The averaged product $R_C$ is dependent on the received power, $P = \frac{E^2}{2}$ and geometric delay, $\tau_g$, and hence on the baseline orientation and source direction:

$$R_C = P \cos(\omega \tau_g) = P \cos\left(2\pi \frac{b \cdot s}{\lambda}\right)$$

• Note that $R_C$ is not a function of:
  – The time of the observation -- provided the source itself is not variable.
  – The location of the baseline -- provided the emission is in the far-field.
  – The actual phase of the incoming signal – the distance of the source does not matter, provided it is in the far-field.

• The strength of the product is dependent on the antenna collecting areas and electronic gains – but these factors can be calibrated for.
To illustrate the response, expand the dot product in one dimension:

\[
\frac{\mathbf{b} \cdot \mathbf{s}}{\lambda} = u \cos \alpha = u \sin \theta = ul
\]

Here, \( u = \frac{\mathbf{b}}{\lambda} \) is the baseline length in wavelengths, and \( \theta \) is the angle w.r.t. the plane perpendicular to the baseline.

\( l = \cos \alpha = \sin \theta \) is the direction cosine.

Consider the response \( R_c \), as a function of angle, for two different baselines with \( u = 10 \), and \( u = 25 \) wavelengths:

\[
R_c = \cos(20 \pi l)
\]
Whole-Sky Response

• Top: \( u = 10 \)
  \[ R_c = \cos(20 \pi l) \]
  There are 20 whole fringes over the hemisphere.
  Peak separation 1/10 radians

• Bottom: \( u = 25 \)
  \[ R_c = \cos(50 \pi l) \]
  There are 50 whole fringes over the hemisphere.
  Peak separation 1/25 radians.
From an Angular Perspective

Top Panel:
The absolute value of the response for $u = 10$, as a function of angle.
The ‘lobes’ of the response pattern alternate in sign.

Bottom Panel:
The same, but for $u = 25$.
Angular separation between lobes (of the same sign) is
$$\delta \theta \sim \frac{1}{u} = \frac{\lambda}{b} \text{ radians}.$$
Hemispheric Pattern

- The preceding plot is a meridional cut through the hemisphere, oriented along the baseline vector.
- In the two-dimensional space, the fringe pattern consists of a series of coaxial cones, oriented along the baseline vector.
- The figure is a two-dimensional representation when \( u = 4 \).
- As viewed along the baseline vector, the fringes show a ‘bulls-eye’ pattern – concentric circles.
The Effect of the Sensor

• The patterns shown presume the sensor (antenna) has isotropic response.
• This is a convenient assumption, but doesn’t represent reality.
• Real sensors impose their own patterns, which modulate the amplitude and phase, of the output.
• Large antennas have very high directivity -- very useful for some applications.
The Effect of Sensor Patterns

- Sensors (or antennas) are not isotropic, and have their own responses.

- **Top Panel:** The interferometer pattern with a \( \cos(\theta) \)-like sensor response.

- **Bottom Panel:** A multiple-wavelength aperture antenna has a narrow beam, but also sidelobes.
The Response from an Extended Source

- The response from an extended source is obtained by summing the responses at each antenna to all the emission over the sky, multiplying the two, and averaging:

\[ R_C = \left\langle \iiint V_1 d\Omega_1 \times \iiint V_2 d\Omega_2 \right\rangle \]

- The averaging and integrals can be interchanged and, providing the emission is spatially incoherent, we get

\[ R_C = \iiint I_v(s) \cos(2\pi \nu \mathbf{b} \cdot \mathbf{s}/c) d\Omega \]

- This expression links what we want – the source brightness on the sky, \( I_v(s) \), – to something we can measure - \( R_C \), the interferometer response.

- Can we recover \( I_v(s) \) from observations of \( R_C \)?
A Schematic Illustration in 2-D

- The correlator can be thought of ‘casting’ a cosinusoidal coherence pattern, of angular scale $\sim \lambda/b$ radians, onto the sky.

- The correlator multiplies the source brightness by this coherence pattern, and integrates (sums) the result over the sky.

- Orientation set by baseline geometry.
- Fringe separation set by (projected) baseline length and wavelength.
  - Long baseline gives close-packed fringes
  - Short baseline gives widely-separated fringes
- Physical location of baseline unimportant, provided source is in the far field.
A Short Mathematics Digression – Odd and Even Functions

• Any real function, $I(x,y)$, can be expressed as the sum of two real functions which have specific symmetries:

$$I(x, y) = I_E(x, y) + I_O(x, y)$$

An even part:

$$I_E(x, y) = \frac{I(x, y) + I(-x, -y)}{2} = I_E(-x, -y)$$

An odd part:

$$I_O(x, y) = \frac{I(x, y) - I(-x, -y)}{2} = -I_O(-x, -y)$$
Why One Correlator is Not Enough

- The correlator response, $R_c$:

$$R_c = \iint I_\nu(s) \cos(2\pi \nu b \cdot s/c) d\Omega$$

is not enough to recover the correct brightness. Why?

- Only the even part of the distribution is seen.

- Suppose that the source of emission has a component with odd symmetry:

$$I_o(s) = -I_o(-s)$$

- Since the cosine fringe pattern is even, the response of our interferometer to the odd brightness distribution is 0.

$$R_c = \iint I_o(s) \cos(2\pi \nu b \cdot s/c) d\Omega = 0$$

- Hence, we need more information if we are to completely recover the source brightness.
Why Two Correlations are Needed

- The integration of the cosine response, $R_c$, over the source brightness is sensitive to only the even part of the brightness:

$$R_c = \iiint I(s) \cos(2\pi \nu \mathbf{b} \cdot \mathbf{s} / c) \, d\Omega = \iiint I_e(s) \cos(2\pi \nu \mathbf{b} \cdot \mathbf{s} / c) \, d\Omega$$

since the integral of an odd function ($I_o$) with an even function ($\cos x$) is zero.

- To recover the ‘odd’ part of the brightness, $I_o$, we need an ‘odd’ fringe pattern. Let us replace the ‘cos’ with ‘sin’ in the integral

$$R_s = \iiint I(s) \sin(2\pi \nu \mathbf{b} \cdot \mathbf{s} / c) \, d\Omega = \iiint I_o(s) \sin (2\pi \nu \mathbf{b} \cdot \mathbf{s} / c) \, d\Omega$$

since the integral of an even times an odd function is zero.

- To obtain this necessary component, we must make a ‘sine’ pattern. How?
Making a SIN Correlator

• We generate the ‘sine’ pattern by inserting a 90 degree phase shift in one of the signal paths.

\[ \tau_g = b \cdot \frac{s}{c} \]

\[ V = E \cos[\omega (t - \tau_g)] \]

\[ V = E \cos(\omega t) \]

\[ P[\sin(\omega \tau_g) + \sin(2\omega t - \omega \tau_g)] \]

\[ R_s = P \sin(\omega \tau_g) \]
Define the Complex Visibility

• We now define a complex function, the complex visibility, \( V \), from the two independent (real) correlator outputs \( R_C \) and \( R_S \):

\[
V = R_C - iR_S = Ae^{-i\phi}
\]

where

\[
A = \sqrt{R_C^2 + R_S^2}
\]

\[
\phi = \tan^{-1}\left(\frac{R_S}{R_C}\right)
\]

• This gives us a beautiful and useful relationship between the source brightness, and the response of an interferometer:

\[
V_v(b) = R_C - iR_S = \int \int I_v(s)e^{-2\pi ib \cdot s/c} \, d\Omega
\]

• With the right geometry, this is a 2-D Fourier transform, giving us a well established way to recover \( I(s) \) from \( V(b) \).
The Complex Correlator and Complex Notation

• A correlator which produces both ‘Real’ and ‘Imaginary’ parts – or the Cosine and Sine fringes, is called a ‘Complex Correlator’
  – For a complex correlator, think of two independent sets of projected sinusoids, 90 degrees apart on the sky.
  – In our scenario, both components are necessary, because we have assumed there is no motion – the ‘fringes’ are fixed on the source emission, which is itself stationary.

• The complex output of the complex correlator also means we can use complex analysis throughout: Let:

$$V_1 = A \cos(\omega t) = \text{Re} \left( A e^{-i\omega t} \right)$$

$$V_2 = A \cos[\omega (t - b \cdot s / c)] = \text{Re} \left( A e^{-i\omega (t - b \cdot s / c)} \right)$$

• Then:

$$P_{corr} = \left\langle V_1 V_2^* \right\rangle = A^2 e^{-i\omega b \cdot s / c}$$
Wideband Phase Shifters – Hilbert Transform

• For a quasi-monochromatic signal, forming a the 90 degree phase shift to the signal path is easy --- add a piece of cable \( \lambda/4 \) wavelengths long.

• For a wideband system, this obviously won’t work.

• In general, a wideband device which phase shifts each spectral component by 90 degrees, while leaving the amplitude intact, is a Hilbert Transform.

• For real interferometers, such an operation can be performed by analog devices.

• Far more commonly, this is done using digital techniques.

• The complex function formed by a real function and its Hilbert transform is termed the ‘analytic signal’.
Picturing the Visibility

- The source brightness is Gaussian, shown in black.
- The interferometer ‘fringes’ are in red.
- The visibility is the integral of the product – the net dark green area.

Long Baseline

Short Baseline
Examples of 1-Dimensional Visibilities

- Simple pictures are easy to make illustrating 1-dimensional visibilities.

- Unresolved Doubles

- Uniform

- Central Peaked
More Examples

- Simple pictures are easy to make illustrating 1-dimensional visibilities.

- Resolved Double

- Resolved Double

- Central Peaked Double
Another Way to Conceptualize …

• For those of you adept in thinking in terms of complex functions, another way to picture the effect of the interferometer may be attractive …

• The interferometer casts a *phase slope* across the (real) brightness distribution.
  – The phase slope becomes steeper for longer baselines, or higher frequencies, and is zero for zero baseline.
  – The phase is zero at the phase origin.
  – The amplitude response is unity (ignoring the primary beam) throughout.

• The Visibility is the complex integral of the brightness times the phase ramp.
The Complex Integral

Amplitude

Phase

\[ 2\pi \]

\[ \frac{1}{u} \]

\[ \ell = \sin \theta \]
Basic Characteristics of the Visibility

• For a zero-spacing interferometer, we get the ‘single-dish’ (total-power) response.
• As the baseline gets longer, the visibility amplitude will in general decline.
• When the visibility is close to zero, the source is said to be ‘resolved out’.
• Interchanging antennas in a baseline causes the phase to be negated – the visibility of the ‘reversed baseline’ is the complex conjugate of the original. (Why?)
• Mathematically, the visibility is Hermitian. \( V(u) = V^*(-u) \).
Some Comments on Visibilities

• The Visibility is a unique function of the source brightness.
• The two functions are related through a Fourier transform.
  \[ V_{\nu}(u, v) \Leftrightarrow I(l, m) \]
• An interferometer, at any one time, makes one measure of the visibility, at baseline coordinate \((u, v)\).
• ‘Sufficient knowledge’ of the visibility function (as derived from an interferometer) will provide us a ‘reasonable estimate’ of the source brightness.
• How many is ‘sufficient’, and how good is ‘reasonable’?
• These simple questions do not have easy answers…