PROCEEDINGS OF SCIENCE

Fundamentals of Radio Interferometry

Richard Porcas*
Max-Planck-Institut für Radioastronomie, Bonn
E-mail: porcas@mpifr-bonn.mpg.de

In this lecture I attempt to explain the fundamentals of radio interferometry and the use of interferometer data for imaging radio sources. The text follows closely the presentation, even to the point of including the original viewgraphs, which are augmented by text in a "one page per step" format. The topics covered in the various steps are:

1-7: a review of the nature of radio astronomy signals
8-18: what a radio interferometer measures
19-23: imaging and the CLEAN algorithm
24-28: calibration and phase self-calibration
29-32: dealing with imperfect path compensation – fringe-fitting

2nd MCCT-SKADS Training School Radio Astronomy: fundamentals and the new instruments
August 26- September 4, 2008
Sigüenza, Spain

*Speaker.

PoS(2nd MCCT-SKADS)013

RADIO NOISE AND RADIO WAVES

RADIO SOURCE EMISSION

RECEIVED NOISE POWER

UNIT OF FLUX DENSITY, S jansky Jy

$= 10^{-26}\ W\ Hz^{-1}\ m^{-2}$

($\equiv$ optical magnitude)

SURFACE BRIGHTNESS (of extended source)

FLUX DENSITY / UNIT SURFACE AREA b

UNIT Jy / beam area (in radio maps)

PHYSICAL UNIT Brightness temperature T_b

The temperature at which a black body would give the same radio emission per unit area

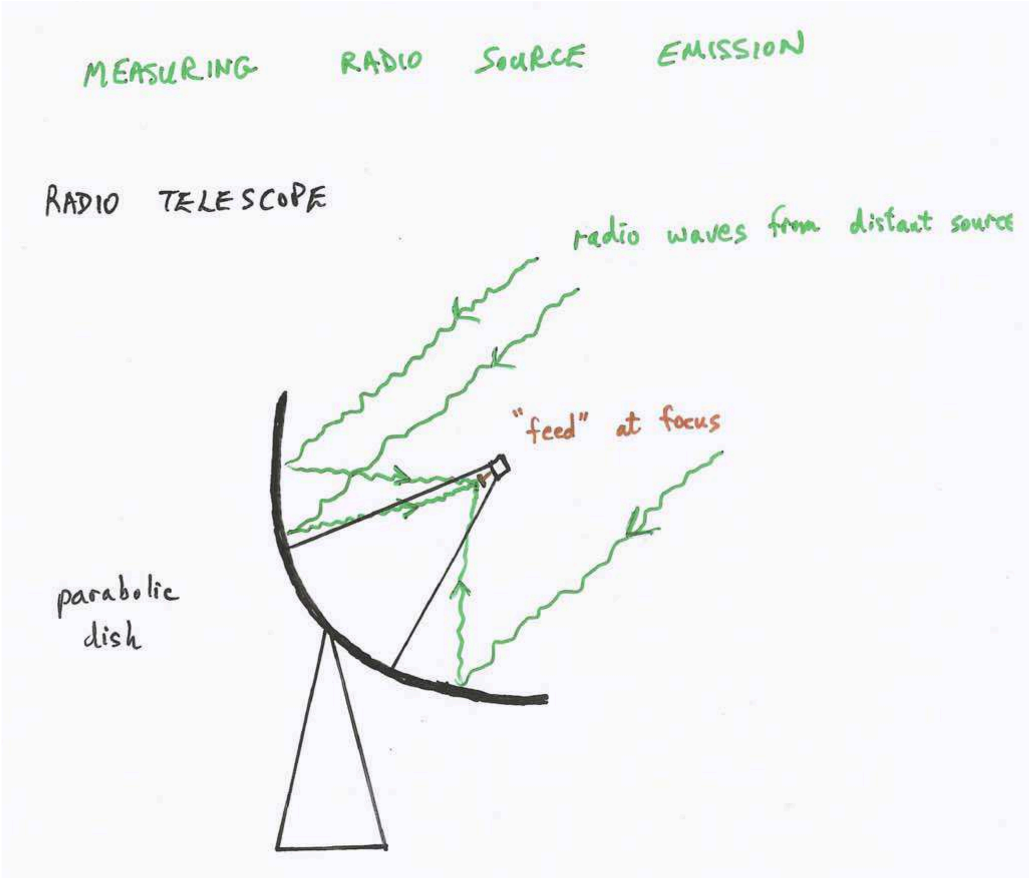
$$T_b = \frac{S \lambda^2}{2 k \Omega}$$

k = Boltzmann's constant
Ω = observed solid angle

SURFACE BRIGHTNESS IS INTRINSIC (DISTANCE INDEPENDANT)

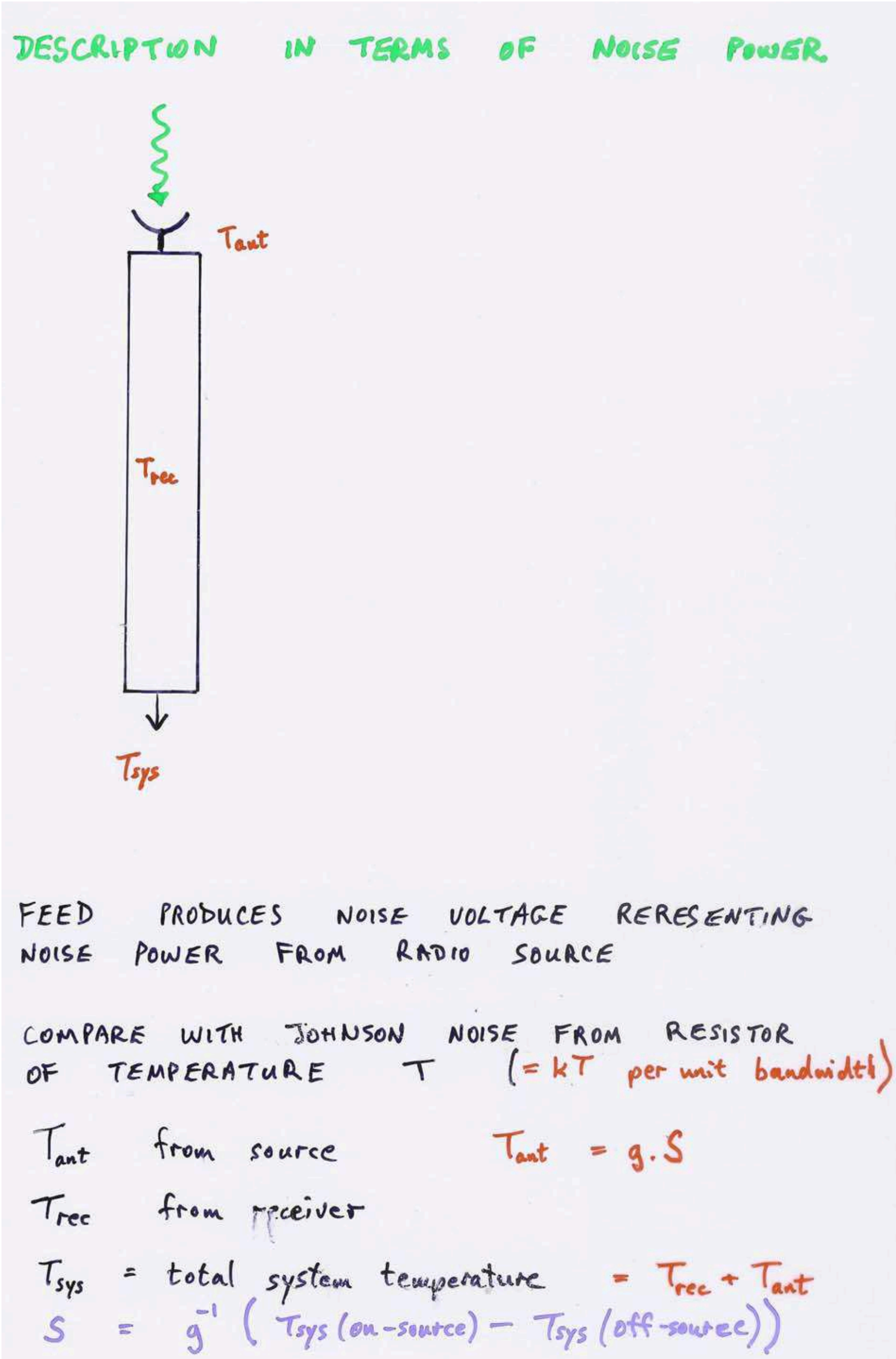
1. Review of radio-astronomical quantities

This is a reminder of the concepts used in measuring the "noise" power of radio sources; flux-density, S, surface brightness, b, and brightness temperature, T_b. "Noise" is just a radio engineer's term for what radio astronomers might normally call the "signal" !



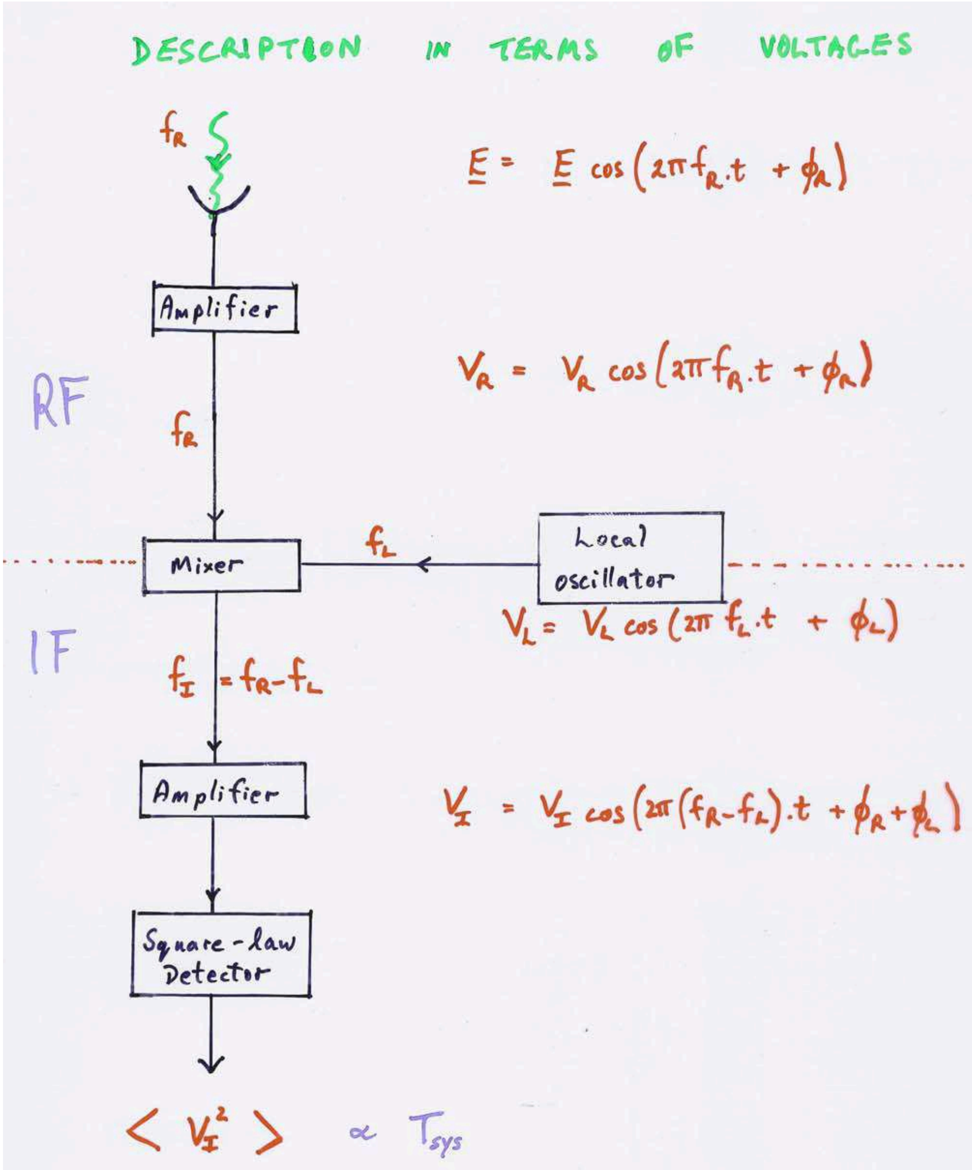
2. Review of radio waves and noise power

To understand radio interferometry we need to consider the wave (amplitude and phase) aspects of radiation in addition to the power. A radio telescope focuses radiation and converts the wave amplitude to a voltage, whose (time-averaged) square represents the power received. "Noise power" recognises the fact that in general this is indistinguishable from the Johnson noise from a resistor, which is proportional to temperature, T. Thus it is customary to measure noise power in K. If we use a single radio telescope to measure radio emission we need to note that the total noise power output from the radio detector ("system temperature", T_{sys}) is the sum of a number of contributions, in particular T_{ant} from the telescope ("antenna") aperture and T_{rec} from the radio receiver electronics.



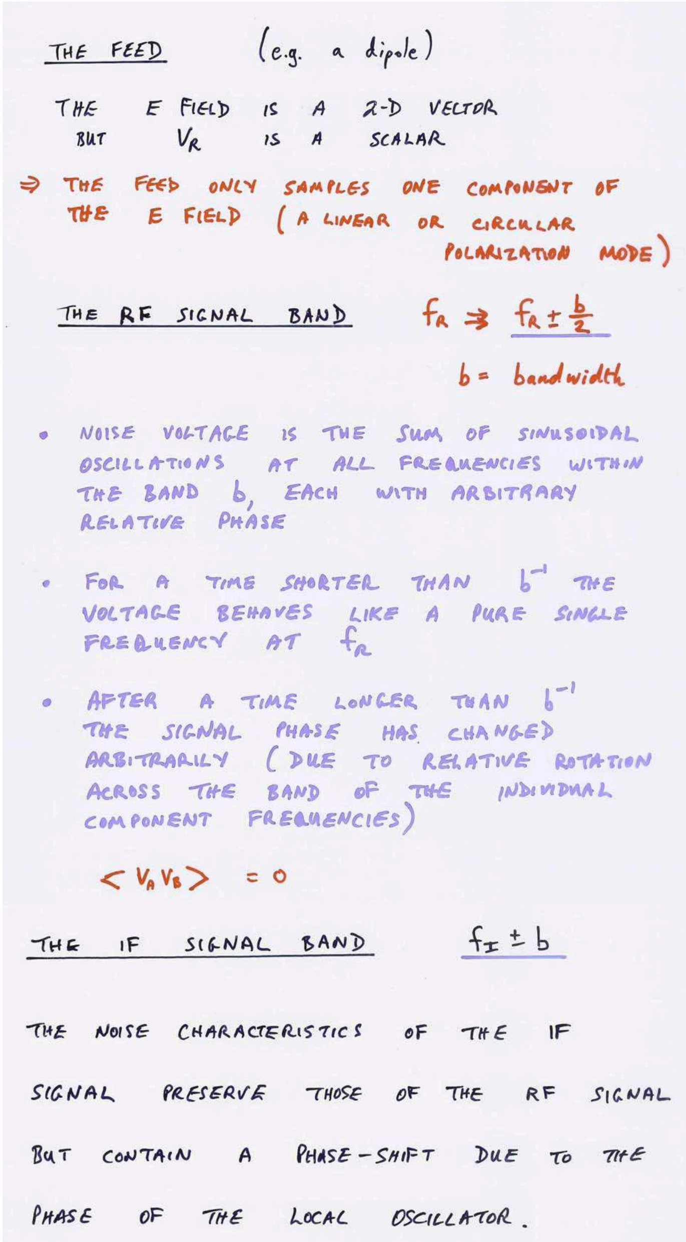
3. Noise power received by a single radio telescope

The flux density, S, of a source is related to T_{ant} by a telescope gain factor, g measured in K/Jy. It can be obtained from the difference between an on-source and off-source T_{sys} measurement.



4. Radio telescope noise power described with signal voltages

The oscillating electric field vector, **E**, of the wave is converted to a voltage, V_R of the same frequency, f_R ("RF" = radio frequency) and amplified. For most radio frequencies, heterodyne receivers are used. These mix the RF signal with a pure, locally-generated, single frequency (f_L) tone (the "LO" = local oscillator) producing a signal at a lower frequency, f_I ("IF" = intermediate frequency) which is easier to analyse. Of course, we in fact observe a band, b, of frequencies covering a range $f_R \pm$ b/2, which converts down to $f_I \pm$ b/2. These wideband signals act like a single frequency for a signal "train" shorter than a time 1/b but acquire a random phase for times longer than 1/b.



THE FEED (e.g. a dipole)

THE E FIELD IS A 2-D VECTOR
BUT V_R IS A SCALAR

⇒ THE FEED ONLY SAMPLES ONE COMPONENT OF THE E FIELD (A LINEAR OR CIRCULAR POLARIZATION MODE)

THE RF SIGNAL BAND $f_R \Rightarrow f_R \pm \frac{b}{2}$

b = bandwidth

- NOISE VOLTAGE IS THE SUM OF SINUSOIDAL OSCILLATIONS AT ALL FREQUENCIES WITHIN THE BAND b, EACH WITH ARBITRARY RELATIVE PHASE
- FOR A TIME SHORTER THAN b^{-1} THE VOLTAGE BEHAVES LIKE A PURE SINGLE FREQUENCY AT f_R
- AFTER A TIME LONGER THAN b^{-1} THE SIGNAL PHASE HAS CHANGED ARBITRARILY (DUE TO RELATIVE ROTATION ACROSS THE BAND OF THE INDIVIDUAL COMPONENT FREQUENCIES)

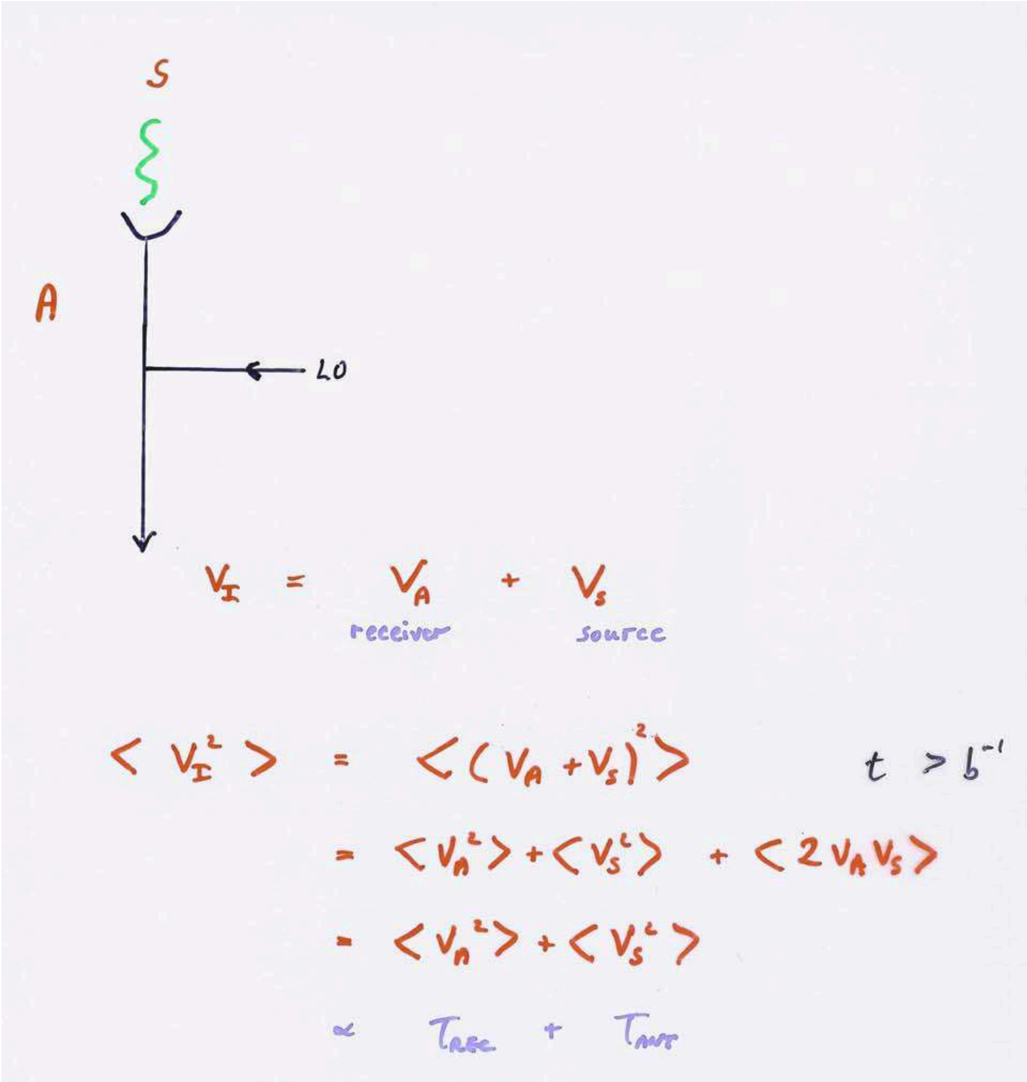
$\langle V_A V_B \rangle = 0$

THE IF SIGNAL BAND $f_I \pm b$

THE NOISE CHARACTERISTICS OF THE IF SIGNAL PRESERVE THOSE OF THE RF SIGNAL BUT CONTAIN A PHASE-SHIFT DUE TO THE PHASE OF THE LOCAL OSCILLATOR.

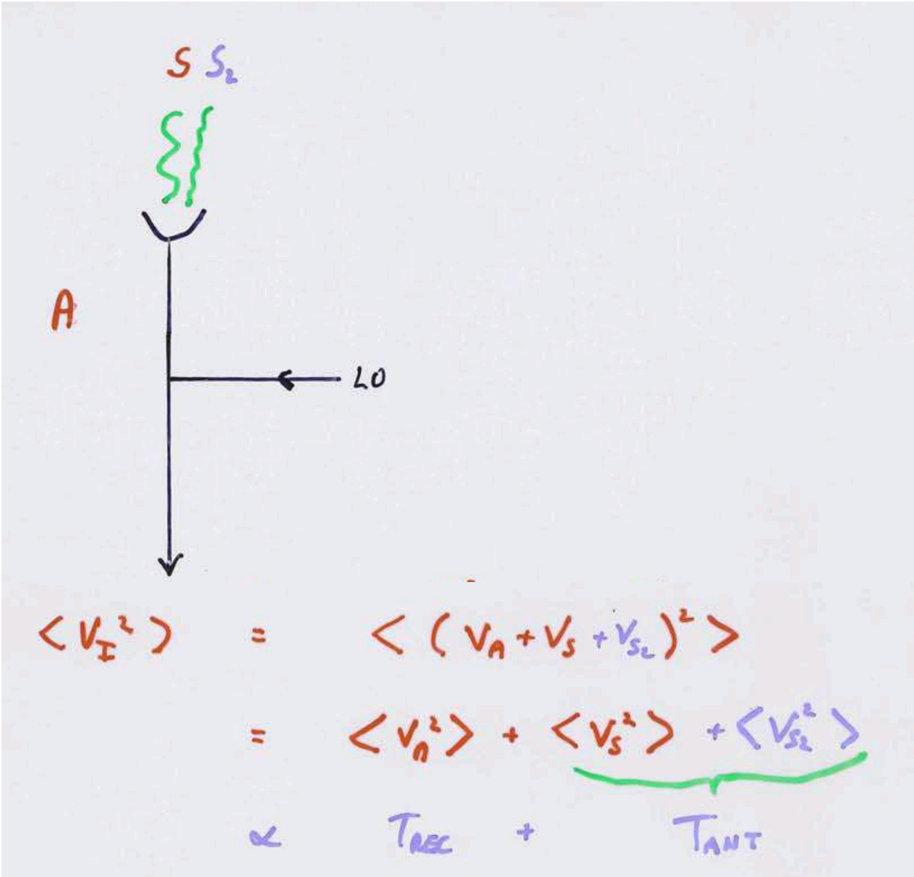
5. A few further points to note regarding the signal

(1) As the voltages V_R and V_I are scalars, they represent only a single component of the (2-D) electric field vector $\mathbf{E}_R$, i.e. a single polarization component. Dual-channel receivers are necessary to detect both polarization components.
(2) The average product of the voltages V_A and V_B at 2 points of the wave-train separated by a time greater than 1/b is zero.
(3) The coherence characteristics of the IF signal preserve those of the RF signal.



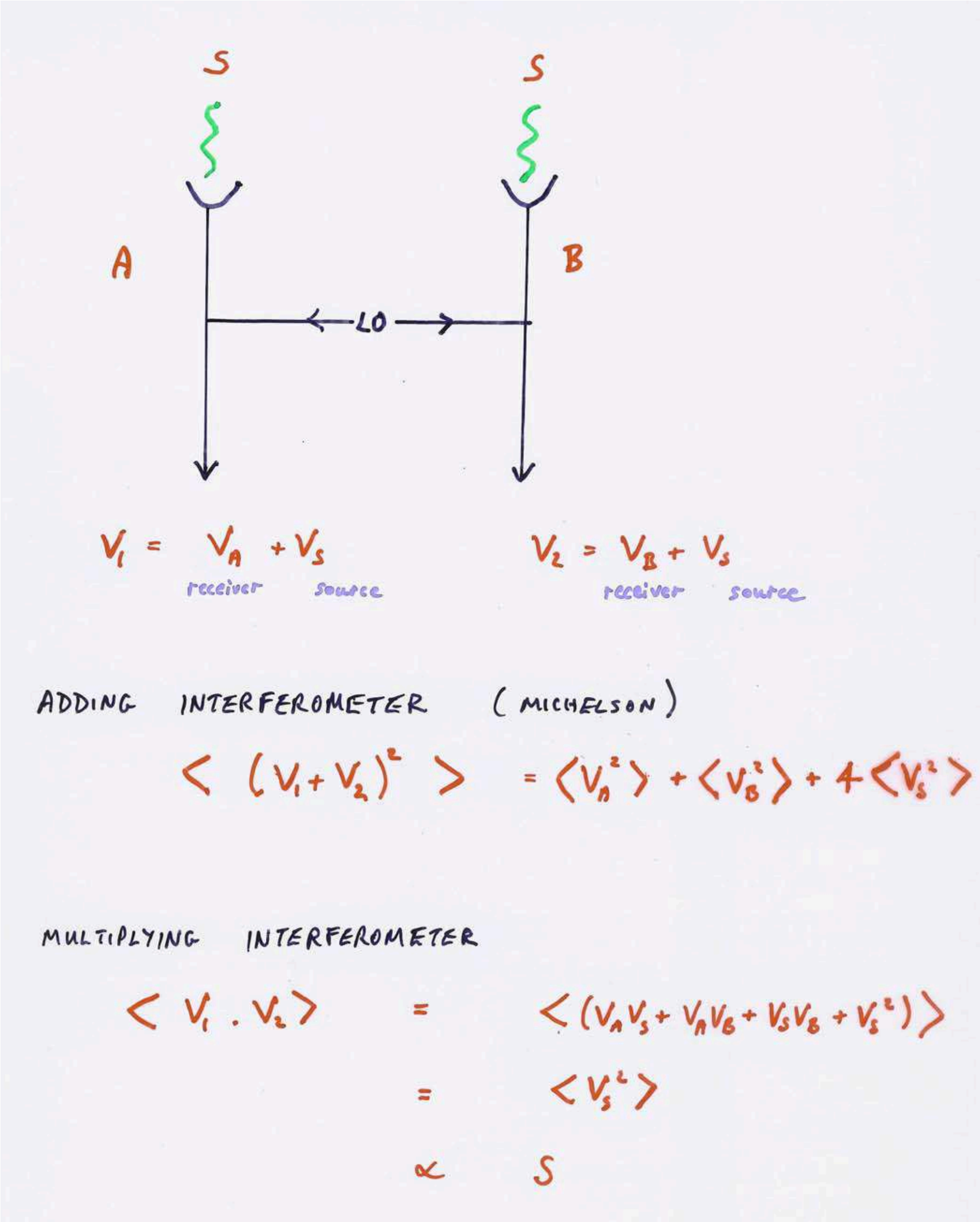
6. Addition of independent sources of noise

We consider the IF voltage, V_I, as having contributions from two components. Whereas "instantaneous" voltages add, the long-term ($t > 1/b$) time-averaged square of the sum of voltages from independent sources (here the receiver noise, V_A and the source noise V_S) is simply the sum of the individual squared voltages since $< V_A V_S > = 0$.
Thus $< V_I^2 >$ is proportional to $T_{sys} = T_{rec} + T_{ant}$.



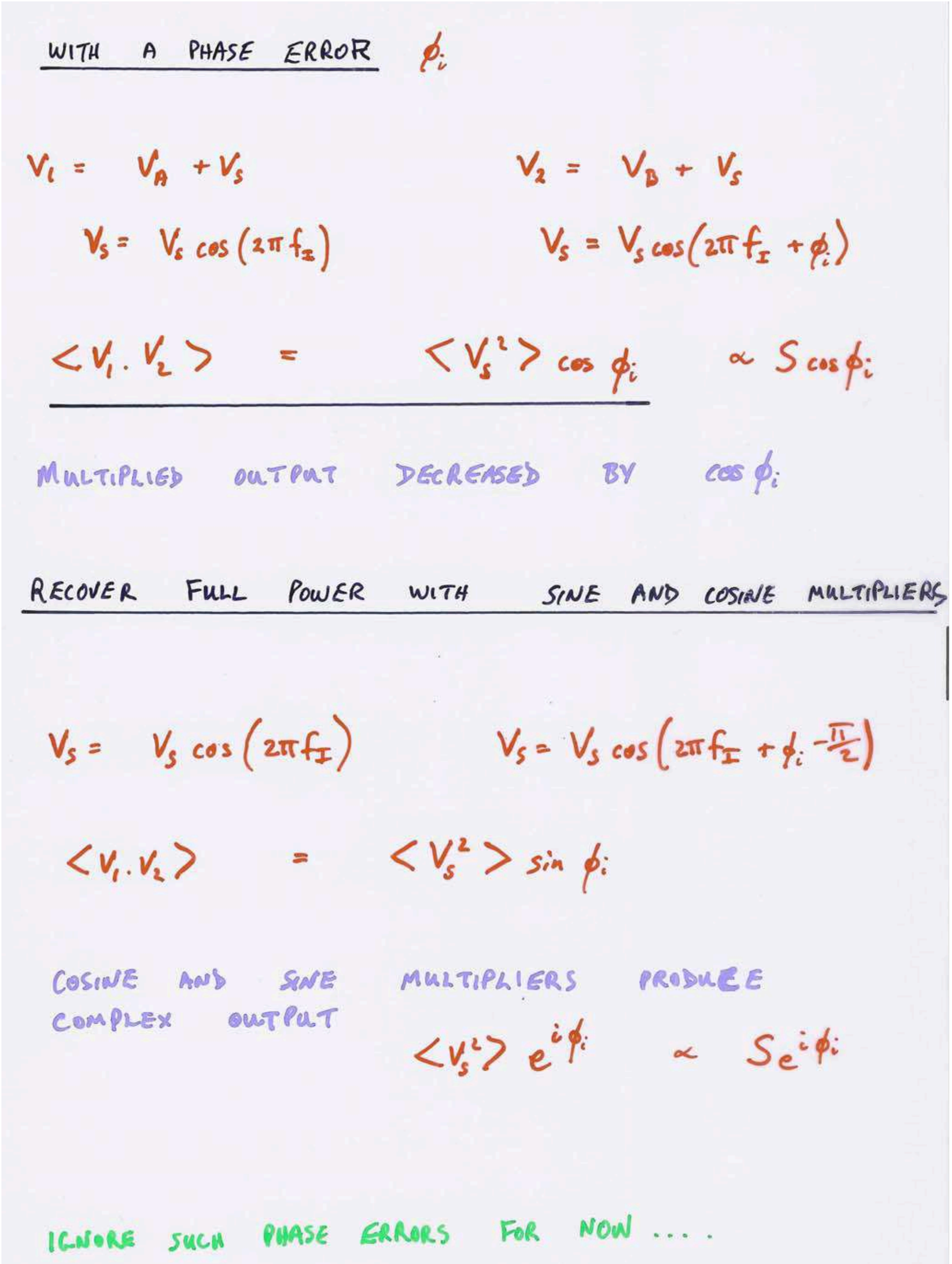
7. And the same for two cosmic radio sources

We observe two sources S and S_2 at the same time. T_{ant} is just the sum of what one would obtain for each source individually.



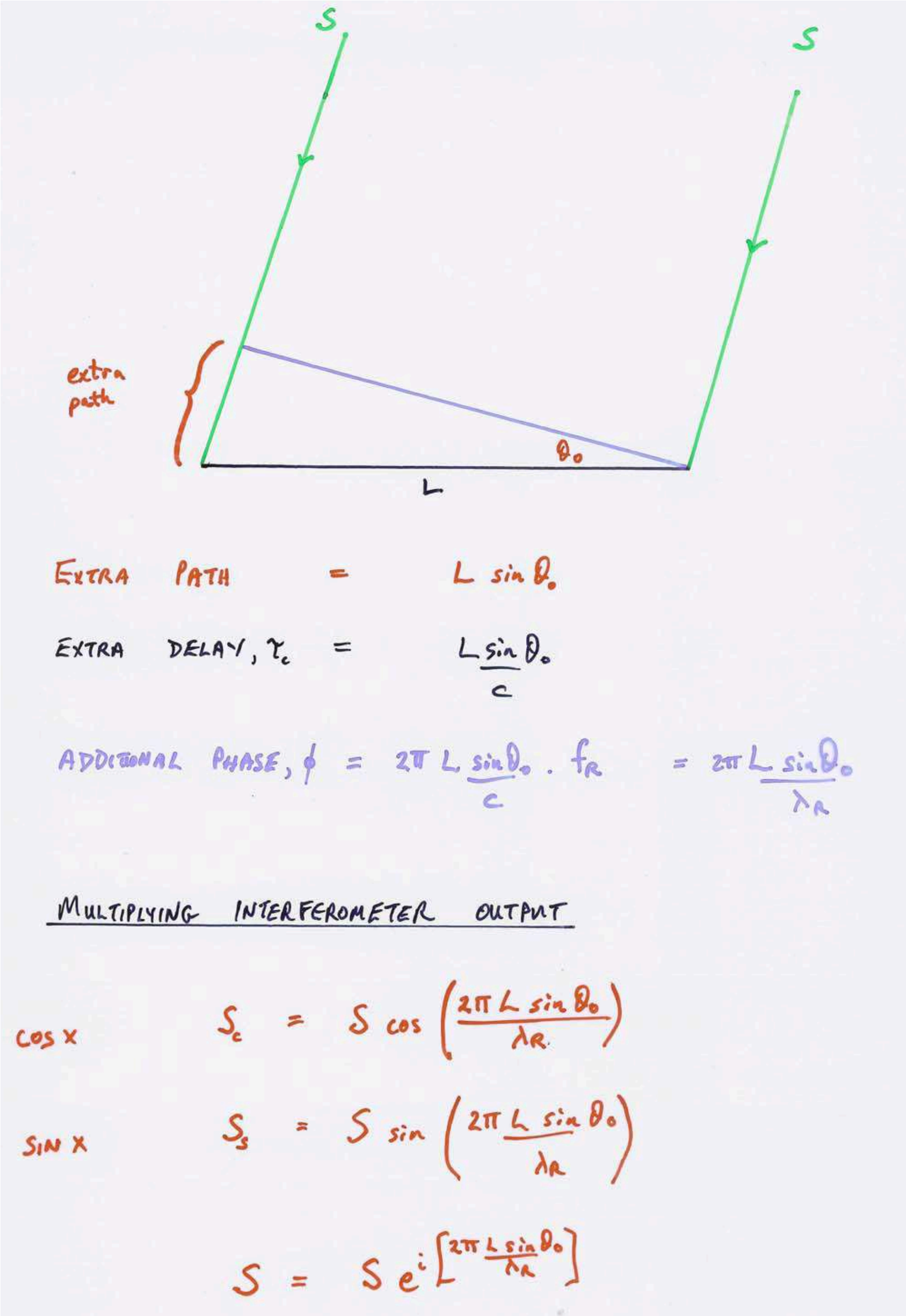
8. Interferometry: combining signals from two radio telescopes

For simplicity we will ignore here the RF/IF distinction and combine the voltages from two telescopes A and B, whose physical separation is zero ! In optical (Michelson) interferometry we ADD the signals. It is usual in radio interferometry to MULTIPLY the signals. Whereas the source voltage contribution is common to the 2 telescopes, the receiver contributions are independent. Hence the time average product (also referred to as the CORRELATION) represents only the source noise power.



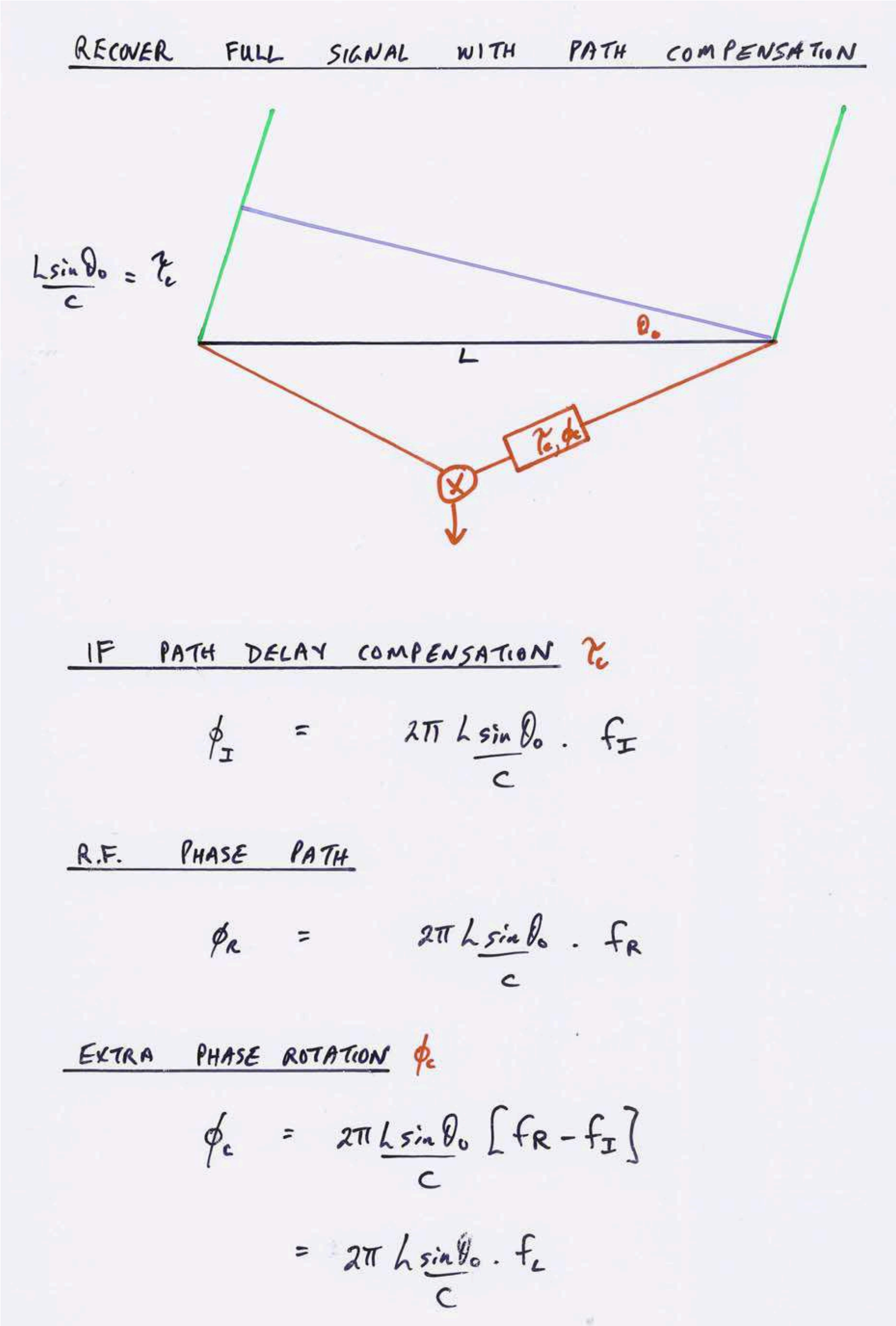
9. Effect of a phase error, ϕ_i

If there is a small phase difference between the source signals in A and B, the voltage product is reduced by a factor $cos\,\phi_i$. The full product can be recovered by making both "cosine" and "sine" (one signal offset by 90 degrees) multiplications. Hereafter we use the complex notation to describe the result of multiplication.



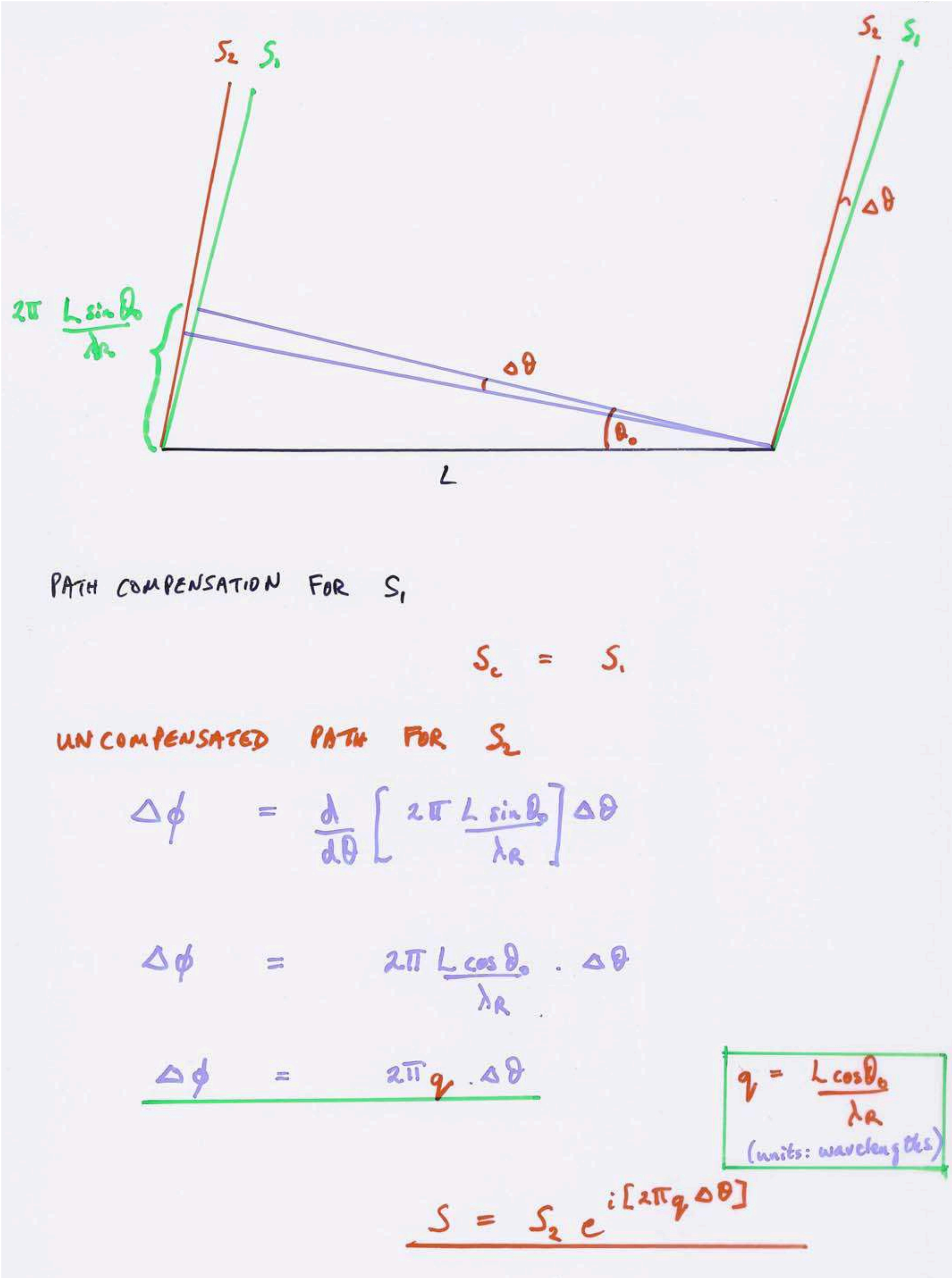
10. The geometrical aspects of radio interferometry

We consider two telescopes separated by distance L, and ignore, for now, Earth rotation. The diagram represents the plane containing the baseline and the direction to a source, S. The unequal paths produce a delay, τ_c, between the source signals at the two telescopes, and hence also a phase difference, introducing a phase term in the complex interferometer output.



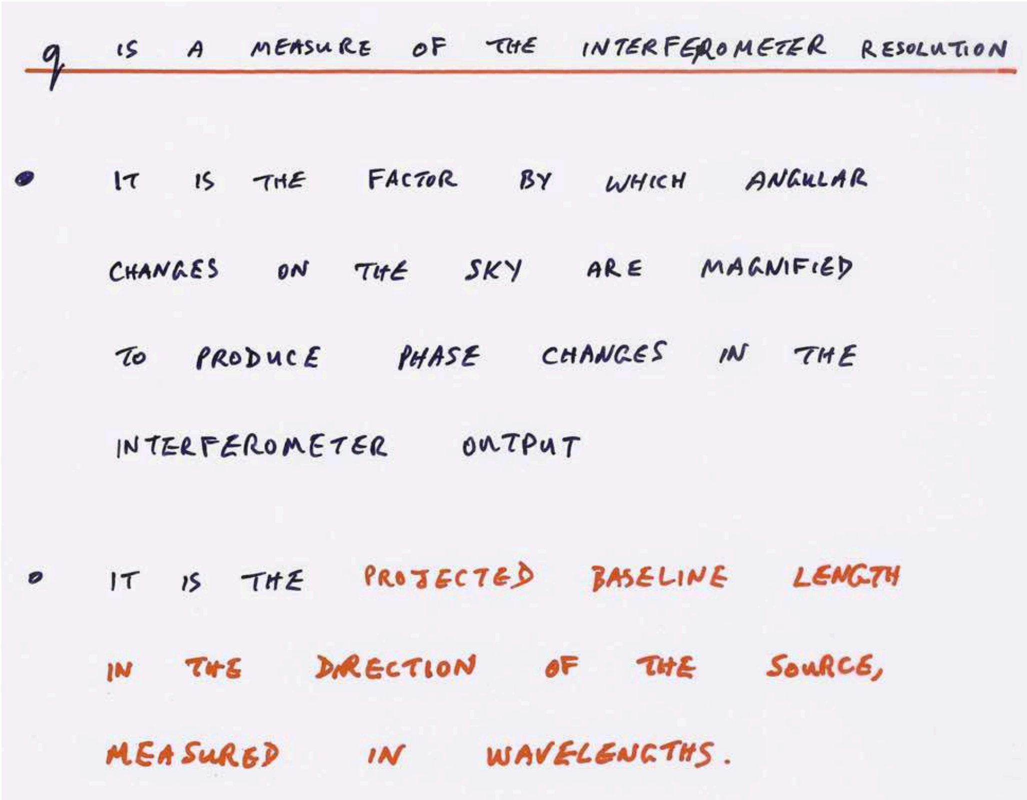
11. Signal path compensation

If τ_c is $> 1/b$ the source signals will be incoherent. A compensating electronic delay, calculated knowing the source position, is therefore introduced in one arm of the interferometer to recover the interferometer output. This is done at IF, so an additional phase rotation is also necessary.



12. Interferometer response to an offset source

Suppose we compensate for a nominal source position S_1 but in fact observe a source S_2 offset in position by $\Delta\theta$. Then the complex interferometer output has a phase term, $2\pi q \Delta\theta$.



13. Interferometer resolution

The author introduced the letter q to denote the 1-D equivalent of the more conventional u and v designations for the resolution factors in the Right Ascension (-x) and Declination (y) directions (see later).

We are now in a position to understand the final basic step of interferometry - the relationship between the radio powers of a distribution of sources on the sky and the resulting output of an interferometer. Remember that the powers are modified by a phase term dependent on the offset from a nominal interferometer "pointing" position, and that the powers of independent sources add.

TOTAL INTERFEROMETER RESPONSE TO S_1 AND S_2

$$S_{TOT} = S_1 + S_2 e^{i2\pi q \Delta\theta}$$

FOR MANY POINT SOURCES

$$S_{TOT} = S_1 + S_2(\Delta\theta_2) e^{i2\pi q \Delta\theta_2} \dots + S_i(\Delta\theta_i) e^{i2\pi q \Delta\theta_i}$$

FOR CONTINUOUS DISTRIBUTION

$$dS = b(\Delta\theta)\, d\theta$$

$$S_{TOT} = \int b(\Delta\theta)\, e^{i2\pi q \Delta\theta} \cdot d\theta$$

$$\Theta = \theta - \theta_0$$

$$\boxed{S_{TOT} = \int b(\Theta)\, e^{i2\pi q \Theta}\, d\Theta}$$

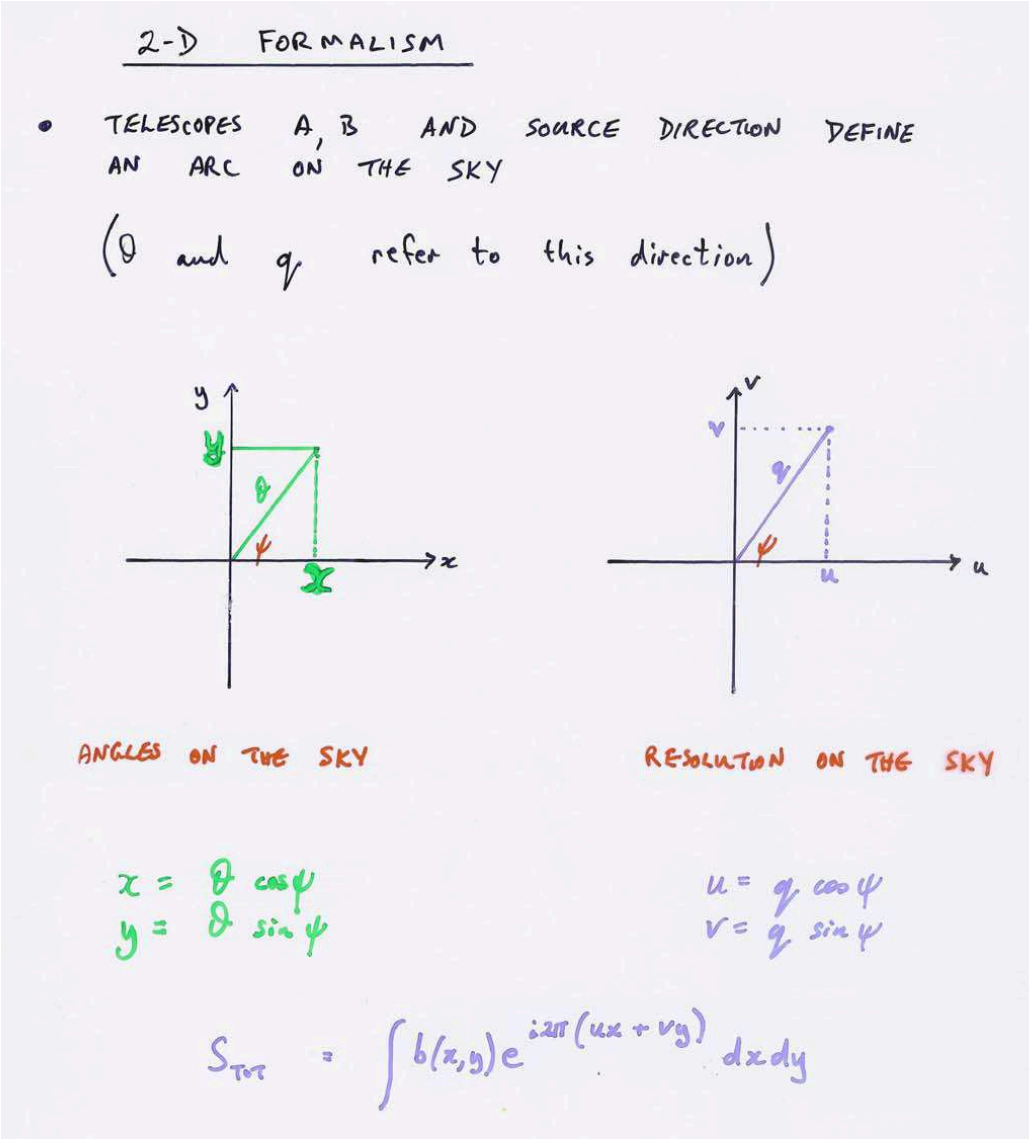
INTERFEROMETER FLUX IS ONE COMPONENT OF

THE FOURIER TRANSFORM OF $b(\Theta)$

SPATIAL FREQUENCY q

14. Interferometer response to a distribution of sources

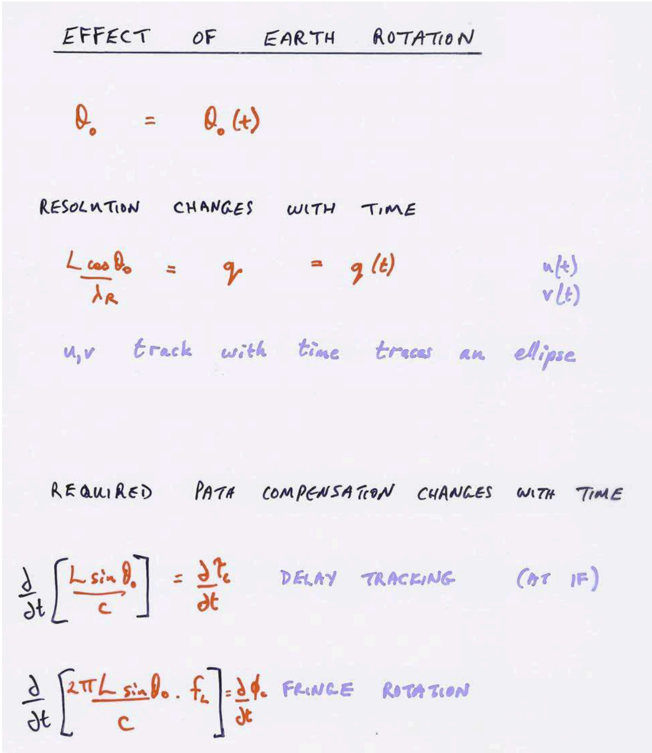
The total interferometer output it simply the sum of all the flux densities of the sources visible to the two telescopes, but each modified by a phase term dependent on its offset in the resolution direction, ψ, defined by the angle of the intersection with the sky of the plane containing the baseline and the source direction. Noting that source extension in the direction perpendicular to the resolution direction has no effect on this phase, we can generalize and instead form the integral of the “brightness distribution”. It is apparent that S_{tot} is a component of the Fourier Transform of the brightness distribution, with “spatial frequency” q. S_{tot} is referred to as the **source visibility**.

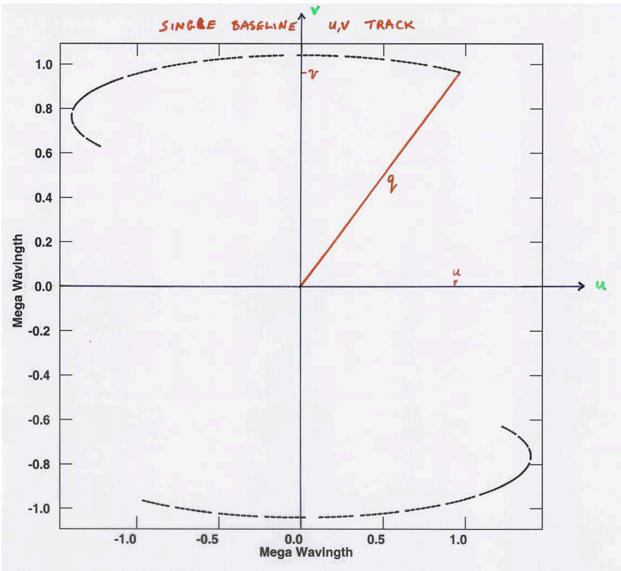


15. Resolution in 2 dimensions

Now we just need to tidy up a little. The sky brightness distribution is clearly 2-dimensional - we use x and y to denote directions parallel to astronomical RA and Dec. Its Fourier Transform is, similarly, 2-dimensional with spatial frequency (resolution) axes u and v. The resolution factor, q, and direction, ψ, correspond to some particular values of u and v which reflect the specific geometry of the angle of the interferometer baseline and the source direction.

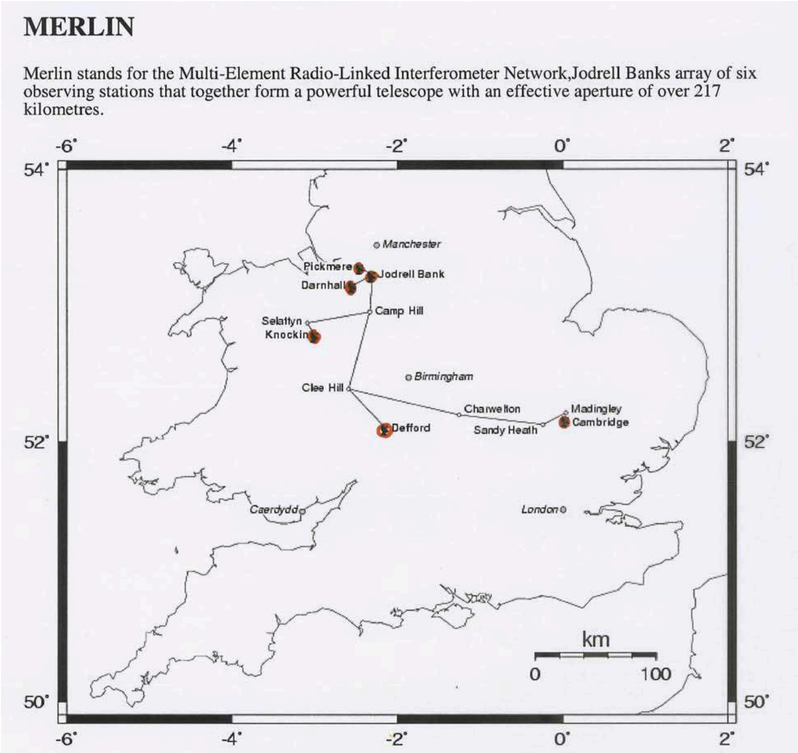
We may also note that the sign of the visibility phase depends on which of the telescope signals we measure the delay with respect to. If we change the "reference" telescope we are sampling the point (-u,-v); thus both points are sampled by a single measurement. This reflects the fact that the sky brightness distribution is a real function, and hence its Fourier Transform is Hermitian.

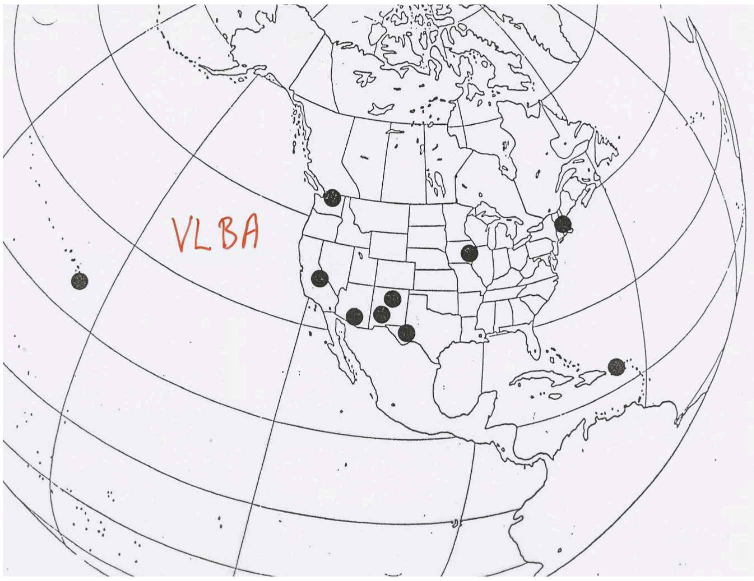




16. Effect of Earth rotation

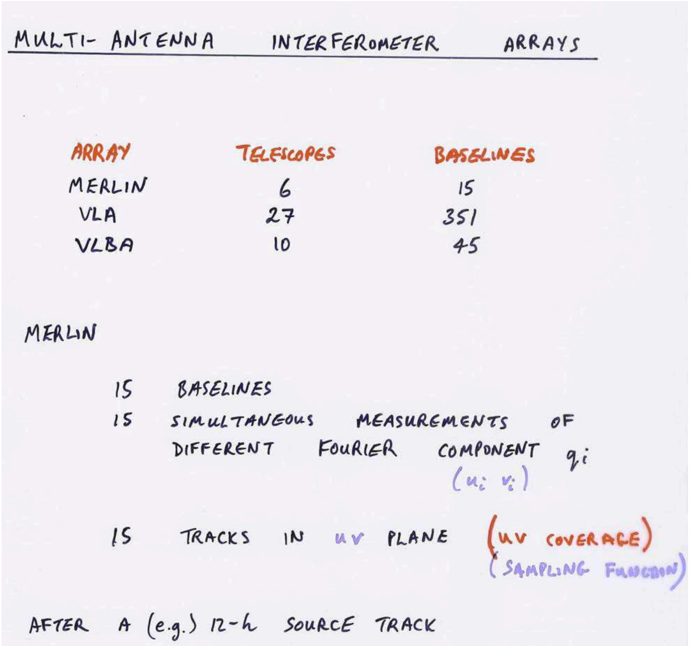
Of course, as the Earth rotates, the source angle with respect to the baseline changes continuously, allowing visibility measurements along a locus (track) of points in the u,v plane to be made. For telescopes fixed to the Earth the locus is an ellipse; for the special case of a baseline along an E-W line, the centre of the ellipse is at the centre of the u,v plane.

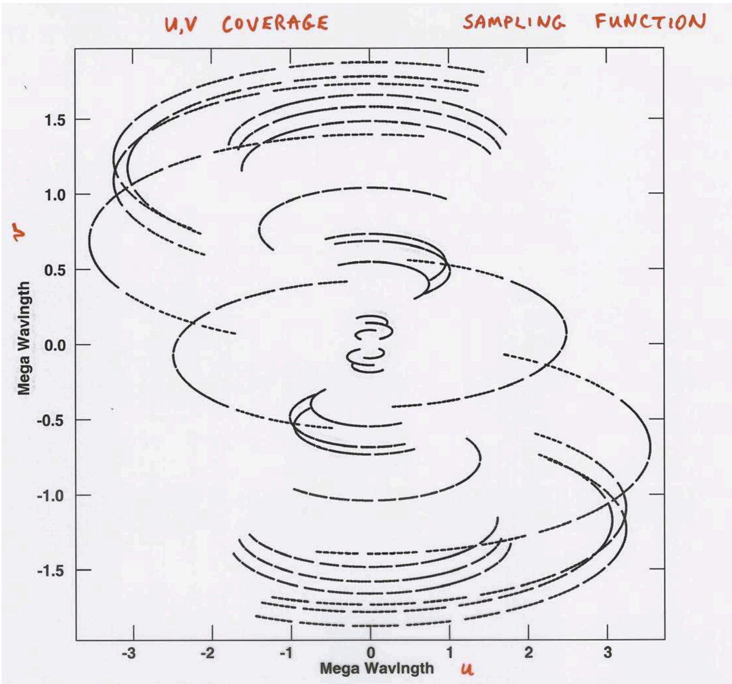




17. Observing with an array of N telescopes

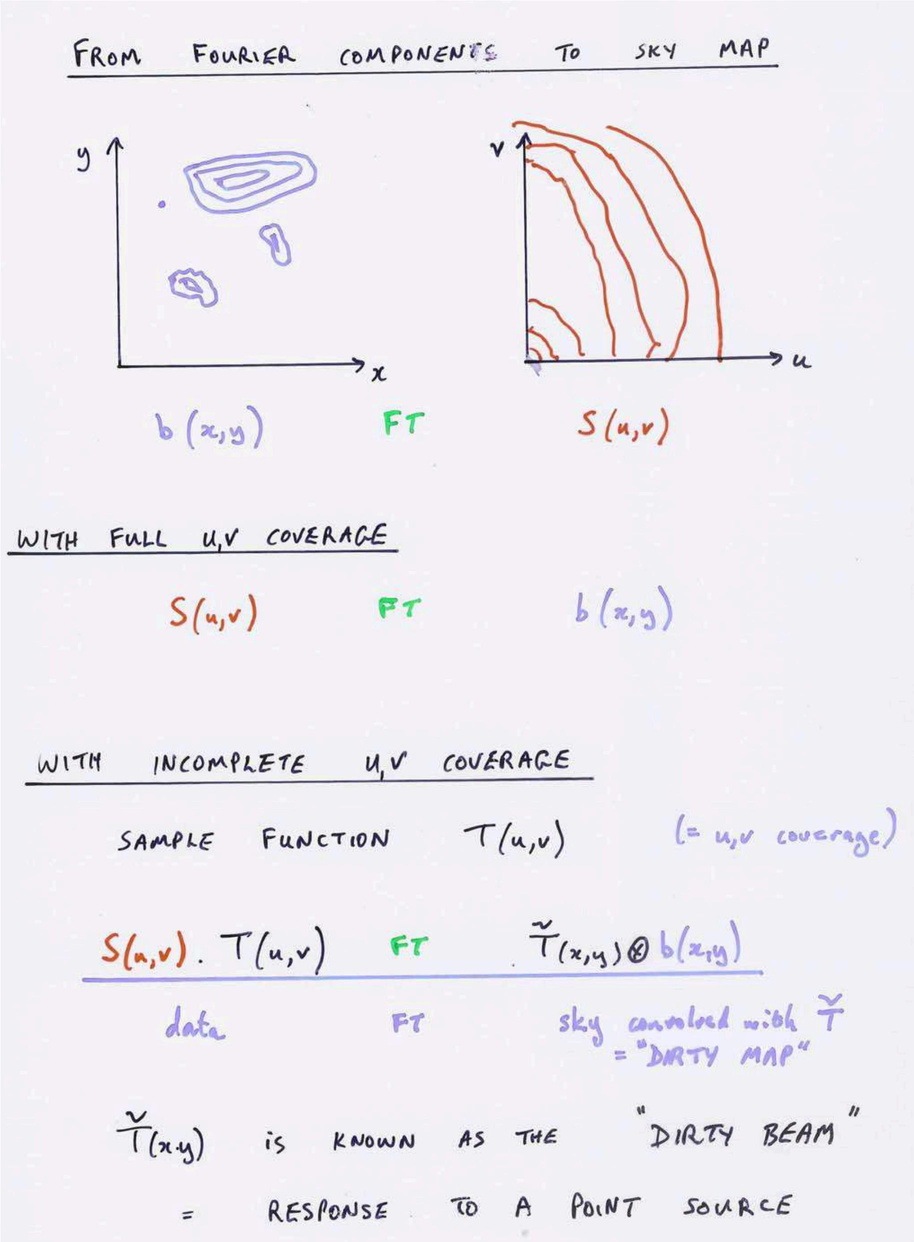
A single interferometer gives only very sparse information about the sky-plane structure. Better "u,v coverage" is obtained by observing with an array of N telescopes which provide N(N-1)/2 simultaneous interferometer baselines.





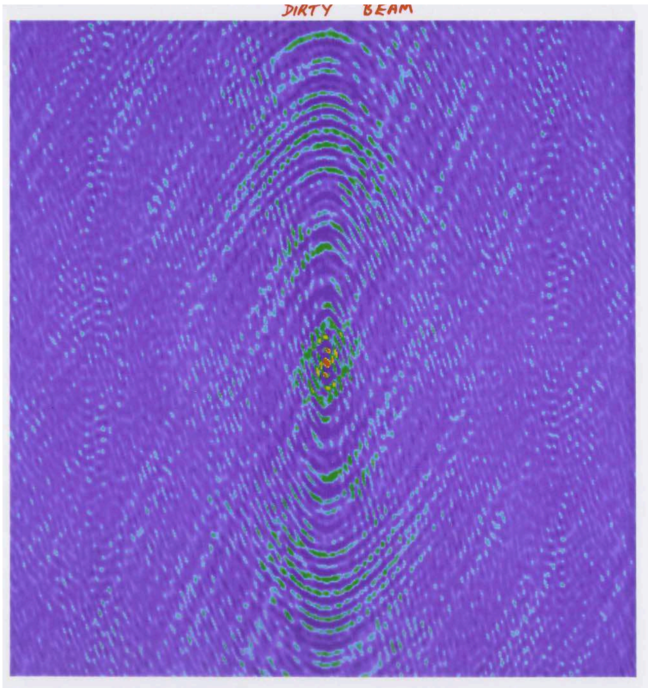
18. The u,v coverage of an observation

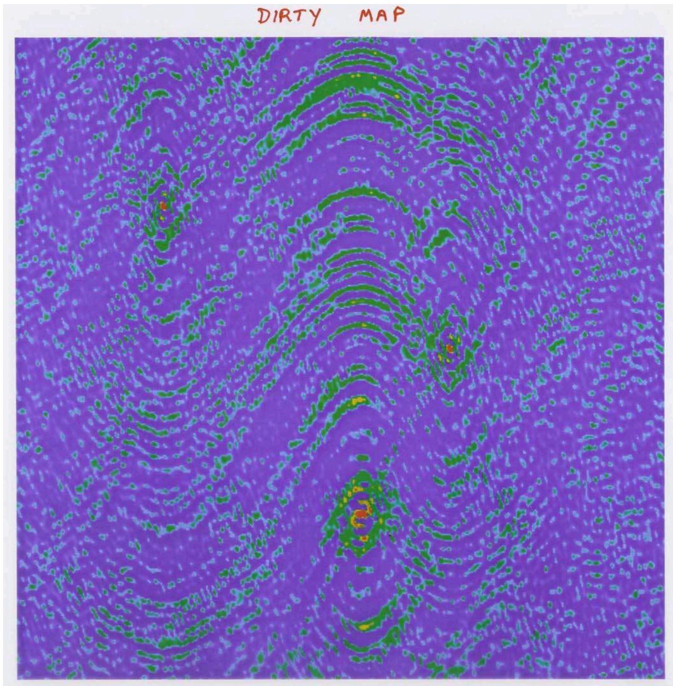
The lower plot shows the u,v coverage (the "sampling function") for a 12h observation with the MERLIN array. There are 15 (u,v) tracks (and 15 counterparts reflected through the origin). Note that none of the baselines is oriented E-W.



19. Imaging: the "dirty map" and "dirty beam"

We could recover the sky brightness distribution, b(x,y) by Fourier Transformation of the visibility measurements in the (u,v) plane if we had full coverage (out to some maximum resolution u_{max}, v_{max}). With only partial coverage (represented by the sampling function), Fourier Transformation yields a distorted picture of the sky - the so-called "dirty map". The dirty map is the convolution of the true brightness distribution with the Fourier Transform of the sampling function; the latter is referred to as the "dirty beam".





20. The dirty beam and dirty map resulting from a MERLIN observation

The source is the gravitational lens MG 2016+112

DECONVOLUTION ALGORITHM "CLEAN"

- COMPUTE DIRTY BEAM $T(u,v)$ FT $\tilde{\Upsilon}(x,y)$
- COMPUTE DIRTY MAP $S(u,v).T(u,v)$ FT $\tilde{\Upsilon}(x,y) * b(x,y)$

(•) TRY TO RECOGNIZE DIRTY BEAM RESPONSE IN DIRTY MAP

- LOCATE POSITION OF BRIGHTEST POINT IN DIRTY MAP
- SUBTRACT SCALED VERSION OF DIRTY BEAM CENTRED AT THAT POINT
  - ~~SCALED BY DIRTY MAP VALUE~~
  - SCALED BY L × (DIRTY MAP VALUE)
    L = "LOOP GAIN" ~ 0.1
- STORE THE VALUE AND LOCATION OF THE POINT-SOURCE RESPONSE REMOVED = CLEAN COMPONENT $cc(x,y)$

- REPEAT LAST 3 STEPS UNTIL THE BRIGHTEST POINT IN THE DIRTY MAP IS BELOW THE MAP NOISE LEVEL

- WE ARE LEFT WITH A DENUDED DIRTY MAP (= "RESIDUAL MAP") AND A LIST OF CLEAN COMPONENTS.
  RESIDUAL MAP IS HOPEFULLY JUST NOISE.

21. CLEANing the dirty map

The CLEAN algorithm is based on the convolution relation between the true sky brightness distribution, b, the dirty map and the dirty beam. It attempts to model b as a distribution of delta-functions located at discrete positions - the pixels of the dirty map. It thus seeks to recognise incarnations of the dirty beam response (to a delta-function) in the dirty map, and removes them in decreasing order of brightness. They are referred to as "clean components".

- THE $CC(x,y)$ REPRESENT A MULTI-POINT-SOURCE MODEL OF $b(x,y)$ WHICH, WHEN CONVOLVED WITH THE DIRTY BEAM, WOULD REPRODUCE THE DIRTY MAP.

- IN GENERAL A POOR REPRESENTATION OF A CONTINUOUS BRIGHTNESS DISTRIBUTION $b(x,y)$

- INSTEAD WE CONVOLVE THE $CC(xy)$ WITH AN IDEALIZED BEAM FUNCTION – THE CLEAN BEAM

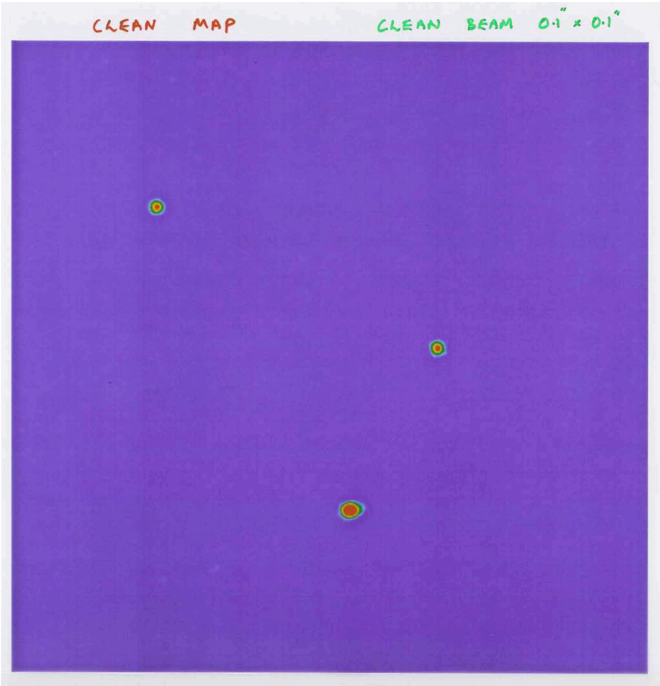
- CLEAN BEAM SIZE CHOSEN TO CORRESPOND TO THE DIRTY BEAM RESOLUTION (CENTRAL PEAK REGION OF DIRTY BEAM)

- ADD TO RESIDUAL MAP TO INDICATE NOISE

$$\underbrace{CM(x,y)}_{\text{clean map}} = \underbrace{CB(x,y)}_{\text{clean beam}} \circledast \underbrace{CC(x,y)}_{\text{clean components}} + \underbrace{RM(x,y)}_{\text{residual map}}$$

22. The CLEAN map: "restoring" the clean components

If the algorithm has run successfully, the dirty map could simply be recovered by convolving the clean components with the dirty beam. The "clean" map is produced by convolving with an idealised beam (the "clean" or "restoring" beam) which is chosen to have a resolution similar to that of the dirty beam but no "sidelobes". It is typically a 2-D elliptical Gaussian function.



23. CLEANed MERLIN maps

These show the source MG 2016+112, derived from the dirty map using CLEAN. Two different restoring beams are used.

INSTRUMENTAL PHASE ERRORS

$$S_{ij} = \int b(\theta) . e^{i2\pi q_{ij}.\theta} . d\theta$$

$$S_{ij} = a_{ij} \; , \; \phi_{ij}$$

$$S_{ij} = a_{ij} \; , \; \phi'_{ij} \qquad \phi'_{ij} = \phi_{ij} + \phi^{e}_{ij}$$

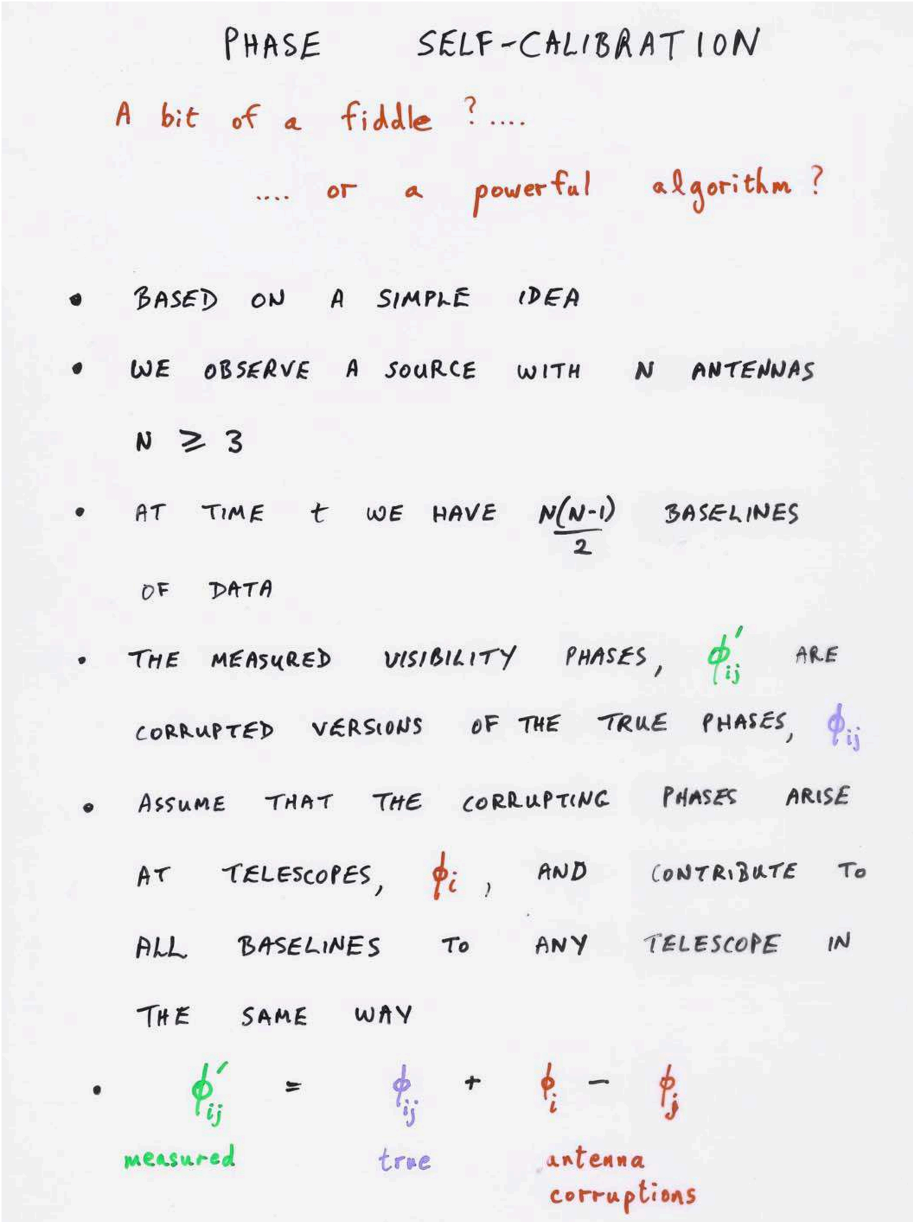
CALIBRATION ON POINT SOURCE OF KNOWN POSITION

$$S^{CAL}_{ij} = a^{CAL}_{ij} \; , \; \underset{=0}{\phi^{CAL}_{ij}}$$

INTERFEROMETER PHASE MEASURED ON CALIBRATOR IS THE ERROR PHASE ϕ^{e}_{ij}

24. Interferometer calibration

A practical implementation of radio interferometry requires, of course, calibration. Both the amplitude and the phase of the visibility function can rarely be calibrated *ab initio*; they depend on the telescope gains and system temperatures and numerous places where additional signal path-lengths arise, not least in the troposphere and ionosphere. And, of course, they are strongly time-dependent - the phase errors especially so. In the end, interferometers are usually calibrated by frequently observing point-like radio sources of known strength and position, whose phase, in the absence of errors, would be zero. The phase error on an interferometer baseline can be calibrated out by observing such a point-like source.



25. Phase self-calibration

This concept is widely used for imaging (= mapping) radio sources with interferometer arrays. It is based on the simple notion that, if one knew in advance the source structure contribution to the visibility phase on each baseline, one could deduce the phase error on each baseline from the visibility measurements themselves. The key idea is to assume that these phase errors result from path length errors at or above the individual telescopes, and hence a baseline phase error is the difference of two "telescope" phase errors.

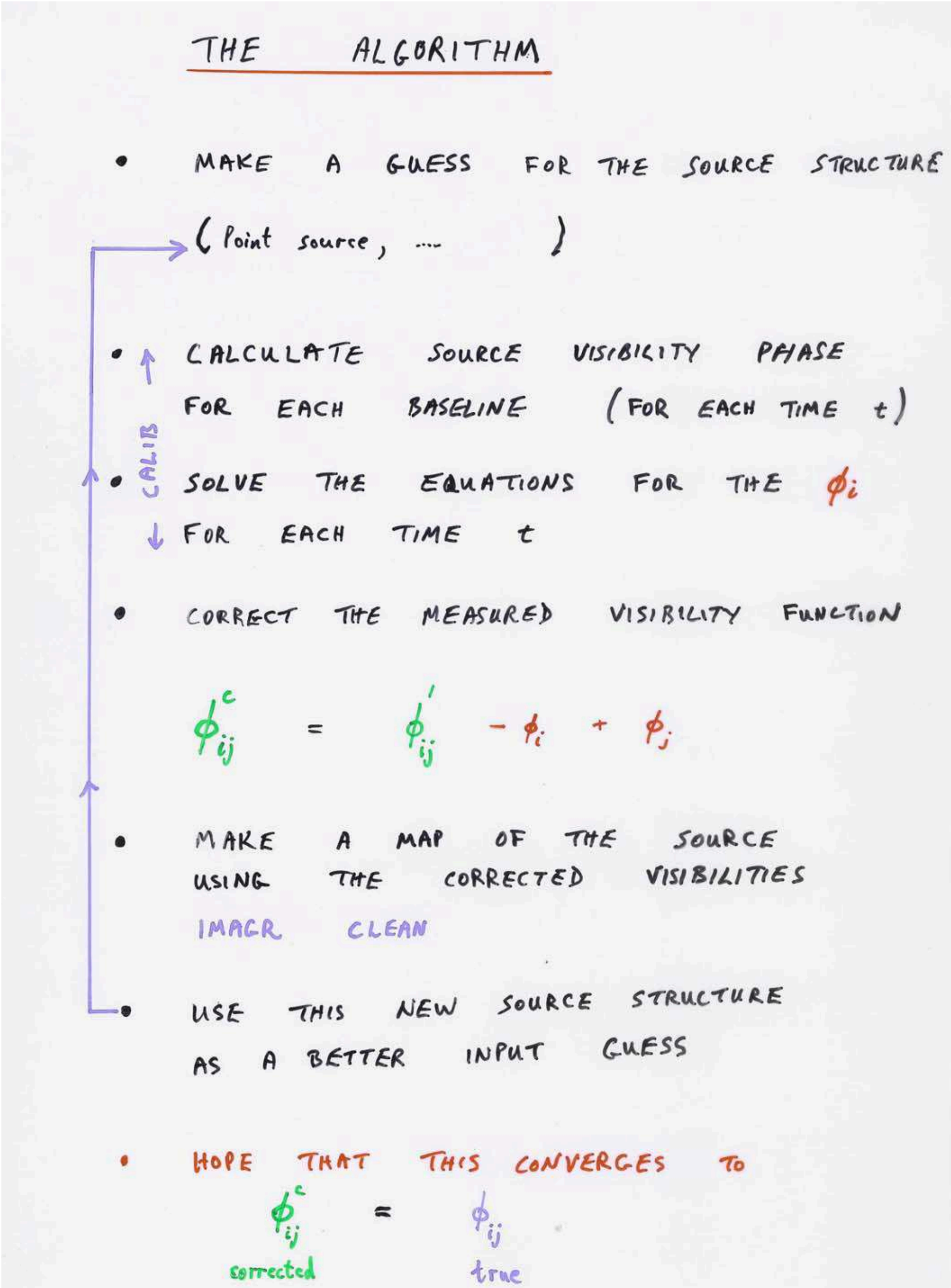
- NOW THE TRICK ......
- SUPPOSE WE KNOW THE TRUE VISIBILITY PHASES, ϕ_{ij} (because we know the source structure and can thus compute them for each baseline for time t)
- THEN WE CAN SOLVE THE $\frac{N(N-1)}{2}$ SIMULTANEOUS EQUATIONS TO DETERMINE THE CORRUPTING PHASES, ϕ_i, BECAUSE THERE ARE MORE EQUATIONS THAN UNKNOWNS
- $$\begin{aligned} \phi'_{12} &= \phi_{12} + \phi_1 - \phi_2 \\ \phi'_{13} &= \phi_{13} + \phi_1 - \phi_3 \\ \phi'_{14} &= \phi_{14} + \phi_1 - \phi_4 \\ \phi'_{23} &= \phi_{23} + \phi_2 - \phi_3 \\ &\vdots \end{aligned}$$

  measured calculated unknowns (N)

  $\frac{N(N-1)}{2}$ equations
- $\frac{N(N-1)}{2} \geq N$ SO SYSTEM IS OVER-DETERMINED : USE LEAST SQUARES MINIMIZATION
- BUT ...... WE DO NOT KNOW THE SOURCE STRUCTURE YET .... SO ....

26. Over-determination leads to source structural information

For an array of N telescopes there are N(N-1)/2 baselines but are only N pathlengths and hence only (N-1) independent pathlength differences. The system is thus over-determined, and the extra information allows one to extract corrections to the assumed source structure phase contributions.



27. The hybrid mapping algorithm

In this algorithm (task CALIB in AIPS) phase self-calibration is used in an interative scheme together with CLEAN imaging to tease out source structure corrections to an initial guess.

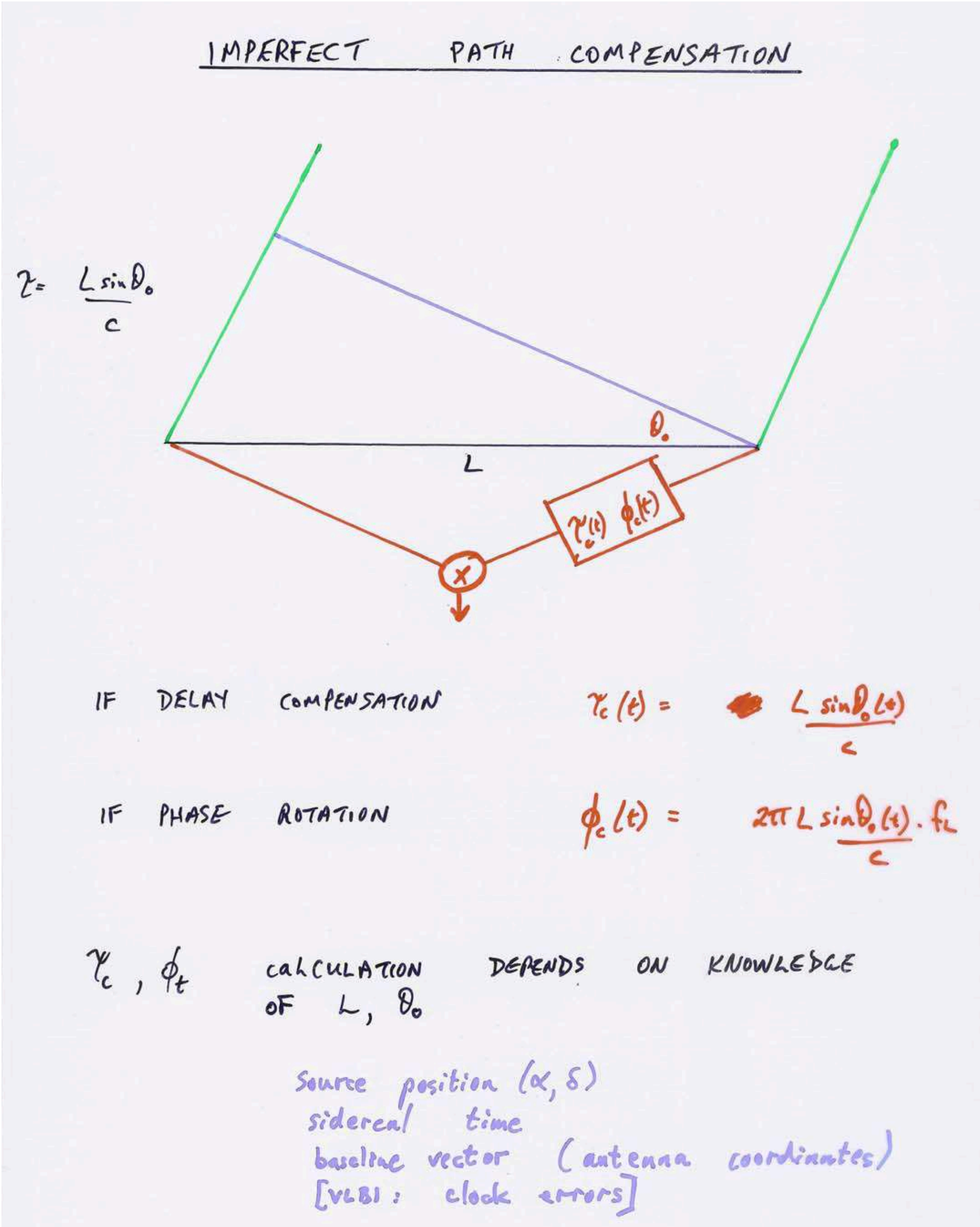
- THE ALGORITHM IS KNOW AS "HYBRID MAPPING"
- IT IS BASED ON PHASE SELF-CALIBRATION
- IT ONLY WORKS IF THERE IS GOOD SNR FOR THE VISIBILITIES
- IT DESTROYS THAT PART OF THE PHASE INFORMATION WHICH GIVES THE POSITION OF THE SOURCE
- THE POSITION OF A SOURCE IN A HYBRID MAP SIMPLY REFLECTS THE POSITION OF THE INITIAL MODEL GUESS (e.g. POINT AT THE MAP ORIGIN)
- ONLY N-1 VALUES OF ϕ_i CAN BE DETERMINED; WE SET THAT OF A REFERENCE ANTENNA TO ZERO ($\phi_{ref} = 0$) SINCE FOR ANY SOLUTION SET ϕ_i, THE VALUES $\phi_i + k$ ARE ALSO A SOLUTION SET

28. Some properties of hybrid mapping

The most important points to note are

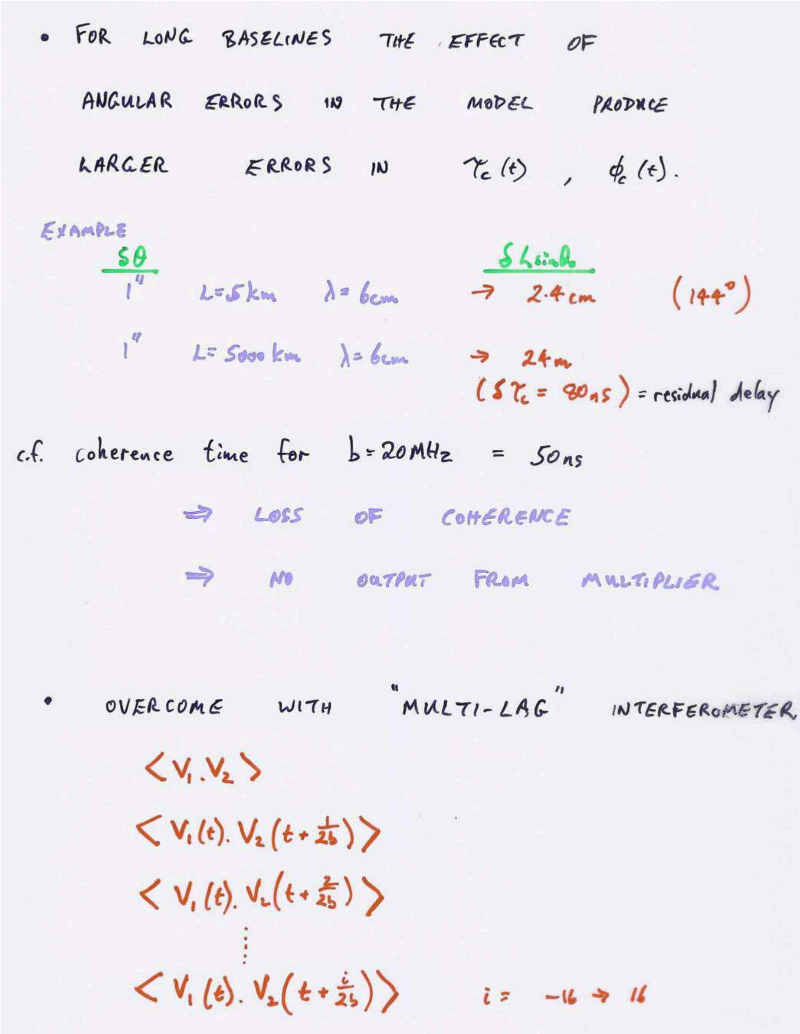
1) The source must be sufficiently strong to be "detected" within the chosen "solution interval" i.e. the time interval chosen between successive phase solutions. In practice this means the source S/N-ratio on a typical baseline must be $> 5/\sqrt{(N-1)}$ where N is the number of telescopes in the array.

2) The position of the source in the hybrid map is determined by the *coordinates of the input model* and not the interferometer phases. For a "point source at the origin" model, the source structure will be centred at the map origin, regardless of where it really is with respect to the sky coordinates.



29. Imperfect path compensation

Recall from **11** that an electronic delay τ_c(t) is used to compensate the path difference in an interferometer for an assumed source position. Small errors in $\theta_\circ$ and L produce phase errors in the visibility function. Large errors may result in errors in $\tau_c > 1$/b (see **4** & **5**) and hence produce a loss of coherence. Similarly, if errors in the compensating IF phase rotation, θ_c(t), change by $> \pi/2$ within a visibility averaging time, coherence is also lost.



30. Multi-delay ("multi-lag") Correlation

Very Long Baseline Interferometry (VLBI) is especially prone to path length errors due to the magification of small angular errors by the long baselines. In addition, use of independent clocks can produce constant delay errors. One solution is to perform the correlation at many different values of compensating delay, separated by delay steps $\Delta\tau = 1/2$b, over a range spanning the delay uncertainty, τ_{range}. This is referred to as "multi-lag" correlation.

Alternatively, one can perform correlation in "spectral line" mode whereby the IF signal bands of each telescope are first Fourier transformed to frequency space, and then multiplied, frequency channel by frequency channel. This is referred to as "FX" (Fourier transform, then multiply) correlation. The frequency resolution is chosen to be $<< \tau_{range}{}^{-1}$.

It is also common, in either multi-lag or spectral line correlation, to divide the observing band into smaller sub-bands, and perform multi-band correlation.

- NEED TO CHOOSE WHICH VALUE OF i HAS THE SIGNAL
- SIGNAL MUST BE STRONG ENOUGH TO RECOGNIZE (e.g. S/N-ratio > 5)
- WE MAY NEED TO INTEGRATE THE SIGNAL TO ACHIEVE SUFFICIENT S/N-ratio.
- BUT ERRORS IN MODEL PRODUCE ERRORS IN FRINGE ROTATION PHASE, ϕ_c, WHICH CHANGE WITH TIME AND PREVENT (VECTOR) INTEGRATION

EXAMPLE

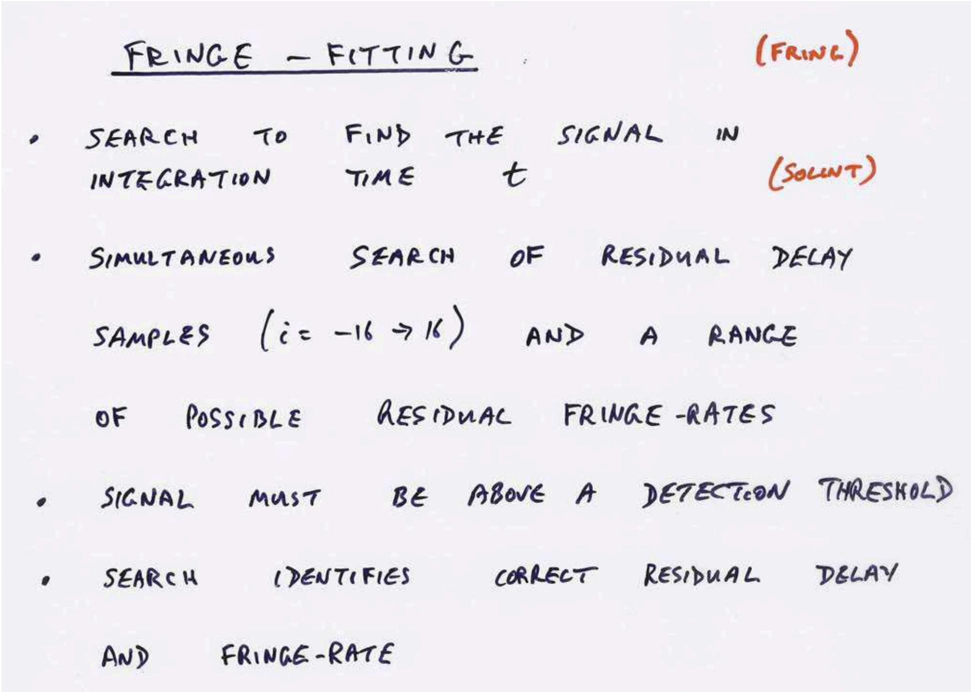
$\delta\theta$ $1''$ $L = 5000$ km $\lambda = 6$cm $\delta\dot{\phi}_c$ 33 mHz (turn in 30^s)

= residual fringe rate

- NEED TO TRY DIFFERENT RESIDUAL FRINGE RATES TO FIND WHICH IS THE CORRECT ONE!

31. The Multi-lag Visibility Function

Multi-lag correlation results in multiple visibility functions, only one of which may contain the source response. The source signal must be sufficiently strong to be recognized above the noise level. The noise can be reduced by averaging, but at the expense of requiring an additional search in “residual fringe-rate” space, formed by a Fourier transform of the visibility time samples over the averaging interval.



32. Fringe fitting

The search for the signal in "residual delay" – "residual fringe-rate" space is referred to as fringe fitting. This can be done baseline by baseline, or by searching all baselines simultaneously to find station-based residual delays and rates, using the same considerations as in phase self-calibration (see **27**). A key consideration to note, as for phase self-calibration, is that the S/N-ratio must be above a certain threshold (e.g. >5) within the solution interval for fringe-fitting to work. The AIPS task FRING performs "global fringe-fitting", i.e. solves for station-based residual delays and rates; it also simultaneously does a station-based phase self-calibration.

"Spectral line" mode correlation naturally results in multiple visibility functions in observed frequency space; AIPS stores visibilities in this way. Multi-lag visibility functions can be converted to multi-frequency ones by Fourier transformation. As a final twist, one should note that the fringe search for residual delay involves a Fourier transform of the multi-frequency visibility function back to delay space.

With multi-band correlation, fringe-fitting may be performed separately in each individual band, if S/N-ratio considerations permit. The maxmimum S/N-ratio can be obtained by simultaneously fitting all bands; this involves also searching for the "multiband delay", which manifests itself as a linear phase gradient with frequency between the visibilities in each band.

Acknowledgments

I wish to thank Gabi Breuer for scanning my original viewgraphs, Helge Rottmann for a careful and critical reading of the draft, and the Editors for their understanding and patience.