High Fidelity Imaging

or,

Getting the right, (or at least the best), image you can

- •How Imaging Errors Affect your Map
- •Origins of Imaging Errors
- •An Appeal for Help!

1 What is High Fidelity Imaging?

•High Fidelity Imaging means getting the correct answer.

•An incorrect image can be caused by many different problems, such as:
- •Errors in your data
- •Errors in approximations used in the imaging process
- •Errors in your methodology
- •Insufficient information

•You'll never get a perfectly correct image, so the problem is to minimize the sources of error.

•The purpose of this lecture is to review some known, or suspecte d sources of error.

2 The Effects of Visibility Errors on Image Dynamic Range

•The most common, and simplest source of error is an error in the measures of the visibility (spatial coherence function).
•What is the effect of such errors on the output image?

•Consider a point source of unit flux density at the phase center, observed by a telescope array of N antennas.
•A visibility measurement is:

$$V(u) = (1+\varepsilon)\delta(u-u_0)e^{-i\phi}$$

where ε = the (additive) error in the visibility amplitude
ϕ = the error in the visibility phase.
and u_0 is the interferometer baseline.

•Now suppose that all but one of the N(N-1)/2 visibility measurements are perfect, (i.e., amplitude = 1, phase = 0) and the only error on that bad interferometer is a phase error, ϕ. What is the effect in the map?

The output image is calculated as:

$$I(l) = \int V(u) e^{i2\pi ul} du$$

Because we have N(N-1)/2 discrete measures, the integral becomes a sum.
For each 'good' baseline, the contribution to the output image is:

$$2\cos(2\pi u_k l)$$

(the factor of two arises because each measure is counted twice – once in its 'correct' location, $u = u_k$, and once, with its complex conjugate, at $u = -u_k$).
But for the 'bad' baseline, the contribution becomes:

$$2\cos(2\pi u_0 l - \phi) \approx 2[\cos(2\pi u_0 l) + \phi\sin(2\pi u_0 l)]$$

Where the approximation is valid for a small (ϕ << 1) error.
Adding them all up, we get:

$$I(l) = 2\phi\sin(2\pi u_0 l) + 2\sum_{k=1}^{N(N-1)/2}\cos(2\pi u_k l)$$

The 'perfect' image (a.k.a. the 'beam', or point-spread function), is given by:

$$B(l) = 2\sum_{k=1}^{N(N-1)/2}\cos(2\pi u_k l)$$

Deconvolution is accomplished by subtracting the 'beam' from the image, then replacing the 'dirty' beam with a 'clean' one, which is more pleasing to the eye.
In our simplified case, this process results in a residual given by:

$$R(l) = I(l) - B(l) = 2\phi\sin(2\pi u_0 l)$$

The 'error' image consists of a sinusoid, with amplitude 2ϕ, and period $1/u_0$.
The rms of this error is $\sqrt{2}\phi$, so that the 'dynamic range' is:

$$D = \frac{Peak}{Noise} = \frac{N(N-1)}{\sqrt{2}\phi} \approx \frac{N^2}{\sqrt{2}\phi} \qquad \text{(1 baseline)}$$

Where the approximation is valid for N >> 1.

To get a feel for the meaning of this, consider a VLA 'snapshot', for which N = 27. A typical phase error (say, due to an atmospheric perturbation) might be 0.1 radian (6 degrees). In this case, the residual on the 'clea ned' image will be about: D ~ 5100.

For an all-day integration, the effect of this small single error will become completely negligible – about one part in 100 million!

But errors rarely occur on a single baseline, at a single time. We can easily extend the simple argument to cover some typical situations.

Random errors on all baselines. If each error on each baseline is independent, then the image errors rise with $\sqrt{N_{baselines}}$, or,

$$D = \frac{N(N-1)}{\sqrt{2}\phi}\frac{\sqrt{2}}{\sqrt{N(N-1)}} = \frac{\sqrt{N(N-1)}}{\phi} \approx \frac{N}{\phi} \qquad \text{(All baselines)}$$

More realistically, errors are antenna-based – they affect all the baselines connected to a single antenna. So, rather than having single error in the image from this one error, we have N-1.
Assuming the effect on the image from these N-1 errors is independent (which is a decent approximation, since they all have different baseline lengths and orientations in an array like the VLA), the dynamic range becomes:

$$D = \frac{N(N-1)}{\sqrt{2}f}\frac{1}{\sqrt{N-1}} = \frac{N\sqrt{N-1}}{\sqrt{2}f} \approx \sqrt{\frac{N}{2}}\frac{N}{f} \qquad \text{(1 antenna)}$$

Most commonly, each of the N antennas has its own small error (caused, for example, by an atmospheric phase fluctuation). In this case, the dynamic range is lowered by another factor of $\sqrt{N}$:

$$D = \frac{1}{f}\sqrt{\frac{N(N-1)}{2}} \approx \frac{N}{\sqrt{2}f} \qquad \text{(All antennas)}$$

These approximate results apply as long as neither the errors for each antenna or correlator nor the locations of the errors (i.e., the interferometer baselines) change.
But of course, errors and baselines do change – which is a good thing here, since these reduce the effects of the errors.

The most straight-forward way to consider the effect of this is to imagine a long observation broken into M individual short ones, each of which has an independent set of errors (either due to change of geometry, or due to changes in atmospheric or instrumental conditions). The image dynamic range is improved by this averaging effect, by a factor roughly $\sqrt{M}$. Thus, we find:

$$D \sim \frac{\sqrt{M}N}{f} \qquad \begin{array}{l}\text{(M observations)}\\ \text{(all baselines)}\end{array}$$

(The factor 'M' for a full synthesis, can range from a few to 1000s.

So far, we have considered only phase errors. One can repeat the simple analysis for amplitude errors, and recover the same results with the substitution: $f \rightarrow e$.

The units of phase error are radians, while for amplitude it is a fraction of the correct value.

Hence, a 10% error in amplitude will have the same bad effect on an image as a 1/10 radian (6 degree) error in phase.

This is an important conclusion. Modern interferometers easily provide estimates of the visibility amplitude with stability often better than 1%. But the atmosphere (neutral and ionized) will very rarely give 1/100 radian phase stability. Phase errors are the dominant cause of poor imaging.

But self-calibration can change all of this!

We now apply these simple concepts to 'typical' observations.
Imagine we are being limited by the atmosphere, so we can ignore amplitude errors, and that the typical phase error at one time or place is 10 degrees. We find the 'dynamic range' is limited to:

1500:1	if the error is on a single baseline,
700:1	if the error is on a single antenna, and
100:1	if equal errors are on all the antennas.

These apply for a single 'snapshot'. If, however, the observations are extended over many hours, there will be many independent errors – say, 100. Then the resultant dynamic ranges will be a factor ~10 better for each of the cases given above.

If self-calibration can be employed, the residual errors might be reduced by a factor of ~100, giving images better by that factor.

For VLA 'continuum' data, the resulting dynamic ranges after self-calibration are typically a few tens of thousands. For strong sources, the remaining errors are definitely not due to thermal noise, so other error sources are responsible.

Experience shows that often the culprit is 'non-closing' errors – baseline-based errors which cannot be removed through antenna-based calibration techniques.

In some circumstances, these errors can be calculated and removed, resulting in images with dynamic ranges exceeding a few hundred thousand. Note that residual errors less than 0.1% (1/20 degree of phase) are needed to reach this level of accuracy.

3 Closing and Non-Closing Errors

A 'closing' error is one which can be identified with an antenna. Its effect thus occurs equally on all baselines which use that antenna. A 'non-closing' error cannot be separated into a pair of antenna-based errors – it is identified with a particular baseline.

Formally, we write:

$$\tilde{V}_{ij}(t) = g_i(t)\, g_j^*(t) G_{ij}(t) V_{ij(t)} + e_{ij}(t) + d_{ij}(t)$$

Here, the term on the LHS is the measured estimate of the visibility while V_{ij} is the true visibility. The g_i are the antenna-based (closing) gain errors, while G_{ij} is the baseline-based (non-closing) error which cannot be factored into a product of two antenna-based gains. The additive errors, ε and δ are baseline-based errors, representing an offset, and thermal noise, respectively. All quantities are considered complex.

Closing errors can be identified and removed through the well-established procedure of self-calibration. This process works well for two key reasons:

- The error is 'seen' identically on N - 1 baselines at the same time – improving the SNR by a factor ~ $\sqrt{N-1}$.
- The N – 1 baselines are of very different lengths and orientations, so the effects of errors in the model are randomized amongst the baselines, improving robustness.

Non-closing errors can also be calibrated out – but here the process is much less robust! The error is on a single baseline, so not only is the SNR poorer, but there is no tolerance to model errors. The data will be adjusted to precisely match the model you put in!

Some (small) safety will be obtained if the non-closing error is constant in time – the solution will then average over the model error, with improved SNR.

4 Origins of Residual Errors

The list of potential sources of errors which limit the accuracy of synthesis imaging is very long! Here I list a few of them that we have thought of, and which might be important. There are undoubtedly others that we haven't thought of, and which are important!

4.1 Thermal Noise.

This is the ultimate source of error. Because it is due to very fast fluctuations within the electronics which cannot be resolved by the correlator, it is a non-closing error – independent on each baseline. From the 'noise' lecture, we find:

$$dI = \frac{C}{\sqrt{N(N-1)\Delta Bt}}$$

Where C is a constant depending on the antenna size and efficiency, the system noise, and type of correlator.

System noise will affect gain solutions. The error in the estim ated gain is:

$$s_G \sim \frac{s_V}{S\sqrt{N}}$$

In this expression, the numerator is the rms of the noise on one baseline, in the time over which a solution is to be calculated, S is the calibrator flux density (in the same units as the rms noise), and N is the number of antennas.

An example: A 10-second solution on a 1 Jy object with the VLA will give an error in the estimate of the gain of each antenna of about 0.4% (or 0.2 degrees), which will limit the dynamic range to a few tens of thousands.

Improving the accuracy by increasing the solution time will eventually fail – when the change in the gain exceeds the errror of the estimate.

4.2 Atmospheric and System Phase and Amplitude Errors

This is the most common type of error in a modern radio telescope. These are 'closing' errors, provided their variations are temporally resolved by the correlator. If so, and if the object being observed is strong enough and small enough, these errors can be removed through the process of 'self-calibration'.

How strong and how small?

- Strong enough that the baselines to any given antenna have a signal a few times larger than the thermal noise on that baseline, in a time short enough to resolve the variability accurately enough to reach the desired dynamic range.

- Small enough that there remains sufficient signal on enough baselines to every antenna for a solution (within the time desired, etc. etc.)

How do you know if your object is small enough and strong enough? A few basic estimates are essential – but the 'bottom line' is to try the self-calibration procedure, and see if the image improves.

4.3 Temporally Unresolved Phase Winds.

If the atmosphere, or electronic, phases are changing on a timescale shorter than the correlator integration time, then the estimate of the visibility will be in error.

Suppose the phase is changing linearly with time. We can write the instantaneous complex visibility as:

$$V = Ae^{i(f_0 + ft)}$$

Thus, over a time integration of length Δt, the result is:

$$R = \frac{Ae^{i2pf_0}}{\Delta t}\int_{-\Delta t/2}^{\Delta t/2} e^{i2pft}dt = Ae^{i2pf_0}\frac{\sin(pf\Delta t)}{pf\Delta t} = Ae^{i2pf_0}\mathrm{sinc}(f\Delta t)$$

where f = the frequency in Hz of the phase wind, and the sinc function is defined as: sinc(x) = sin(πx)/πx. The term $f\Delta$t is then the number of turns of phase wind within the averaging time.

Note that a uniform phase wind does not affect the measurement of the visibility phase, but causes a loss of amplitude. This is a non-closing error, since the magnitude of the loss depends on the pair of antennas concerned (even through the phase wind is itself antenna based)! A non-closing phase error is a 2nd order effect.

A ten degree phase wind will cause a loss in amplitude of ~0.1%, sufficient to limit dynamic ranges to a few hundred thousand.

4.4 Phase or Amplitude Bandpass Errors

There is no essential difference between a phase wind which is unresolved in time, or a phase wind unresolved in frequency.

If the system phase changes with frequency, and the correlator averages over a frequency width Δν, the effect on the measured visibility is the identical to the expression derived in 4.3, with the term *f* replaced with $t = \partial f / \partial n$, and time, t, replaced with frequency, ν.

Thus, the loss in amplitude is: sinc(tΔν), where Δν is the frequency width over which the integration occurs.

It is worth noting that a phase slope over frequency is formally identical to a delay error. The term t (above) is the delay.

As in the temporal winding case, a linear wind does not affect the measured phase. Phase errors will be caused by a non-linear wind.

4.5 Correlator Quadrature Errors.

In the first lecture, I spoke on the complex correlator – a device which actually consists of two multipliers in quadrature – a 'COS' and a 'SIN' multiplier pair.

Suppose the inserted phase shift is not 90 degrees, but is actually, say, 90+φ degrees. It is then easy to show that the calculated visibility phase will be in error by ~φ degrees, thus limiting the dynamic range to levels roughly given in Section 2. A similar error will occur if the multipliers are not balanced in amplitude.

This type of error, which is clearly 'non-closing', can be estimated by an observation of a strong source of known structure. For the VLA's continuum correlator, the phase offset is typically one or two degrees, and the amplitude imbalance one or two percent.

4.6 Quantization Correction Errors

Nowadays, digital correlators (and digital electronics in general) are much preferred, due to their flexibility and precision. But they are not perfect! Replacement of a smoothly varying voltage with a discretely changing voltage results in (amongst other things) an error in the estimate of the complex visibility.

This error is non-linear (i.e. it is not proportional to the visibility magnitude) and acts independently on the COS and SIN correlators The error rapidly diminishes with multi-bit sampling, and can be generally be ignored with (say) 4-bit (16-level) sampling, or better.

The error can be corrected for at the correlator level – this is done with the VLBA correlator, but is not (properly) done on the VLA.

For the VLA, this error reaches about 0.1% for a source of ~50Jy .

4.7 Polarization Leakage

Regrettably, antennas and electronics are not perfect. One of the unhappy consequences of imperfection is that the correlator products labelled (for example), RR, RL, LR, and LL (in a circular polarization representation) or HH, VV, HV, and VH (for a linear polarization representation), are not all that they claim to be! In general, an antenna whose output voltage is labelled, say, V_r, for RCP, is actually a combination of both polarizations:

$$V_r' = V_r + D_r V_l$$

Where D_l measures the amplitude and phase of the 'leakage'.

We form the visibilities corresponding to the Stokes' parameters I, Q, U, and V through linear combination of the four possible complex correlations. But since each of these is contaminated with 'leakage' signal, so too will the I, Q, U, or V estimates.

The upshot of this is that when one is forming the 'I' polarization, one is actually getting a complex combination of all polarization states. For example,

$$V_I' = V_I + D_1 V_Q + D_2 V_U + D_3 V_V$$

Where the D's are sums over various cross-polarization leakages, and the V_i are the complex visibilities corresponding to the four Stokes' parameters.

In normal imaging, these leakages can be calibrated, and their residuals ignored, as the D's are typically a few percent, and the Q, U and V visibilities are also a few percent.

But in high-dynamic range imaging, since the polarized visibilities are not the same as V_I, the residuals will create non-closing errors which will damage your image.

4.8 Far-Out Effects.

Many of the assumptions used in generating those beautiful Fourier relationships shown in Lecture 1 break down at larger angles. Here is a short list:

4.8.1 Non-Coplanar Baselines. As covered in Lecture 1, many real interferometers (including the VLA) measure the visibility in a three-dimensional volume, while most imaging software employs a two-dimensional grid, after a phase adjustment which is valid for a single direction – and is incorrect for every other direction.

If the field of view roughly exceeds:

$$q_{FOV} \geq \sqrt{q_{res}}$$

then notable imaging errors can be expected.

This geometry-based error can be overcome through '3-dimensional' imaging techniques – to be covered in a later lecture.

4.8.2 Antenna Sidelobes and Other Nasty Things.

The technique of aperture synthesis requires the apparent source structure, position, and strength, to remain unchanged during the course of the observation.

Simple offsets in position and changes in strength (caused by electronic gain changes, or atmospheric phase screens) can be effectively removed by self-calibration.

But 'apparent' changes in the source structure, caused by spatially variable antenna gains (amplitude or phase, or both), are much more troublesome.

The most common effect is antenna pointing errors. Others include
- Gravitation warping – changes the primary beam shape.
- Strong sources in the antenna sidelobes – these are not circularly symmetric, and will change in response to elevation, temperature, wind, etc.

In principle, all of these can be handled in computing – but not cheaply!

4.8.3. Varying Phase Screens (The Isoplanatic Patch Problem).

If the phase of the atmosphere or ionosphere changes over the field of view of the object of interest – we're in trouble!

Most imaging algorithms don't know about this, so if object 'A' on one side of the primary beam is being seen through a different atmospheric screen than object 'B' on the other side, we won't be able to get a good image of both at the same time. Standard self-calibration won't help here – it assumes only a single solution, per field. One gets an 'average' solution, which will ruin both images.

This problem is especially severe at low frequencies, where the primary beam size is very large (>10 degrees at 74 MHz), and the ionospheric isoplanatic scale can be very small (~1 degree).

Once again, this problem can be handled in software, if there is enough signal to permit simultaneous, spatially-variant self-calibrations. This is an area of active research and development – the current capabilities will be described in the low-frequency lecture.

4.8.4. Antenna Beam Polarization

Similar to the spatially-variant antenna gain problem is the spatially-variant antenna polarization problem.

All real antennas mix polarization states – e.g. the output labelled 'R' is really a combination of 'R' and 'L'. The 'D' terms quantify this 'mixing', or leakage. They can be estimated, and their effects removed, with reasonable success.

Unfortunately, real antennas also have polarization characteristics which vary with angle – the 'D-terms' are spatially variant. (They are also probably time, elevation, and frequency variant too …).

Precise wide-field polarimetry will require correction for these effects. The variable D-terms can be measured using strong isolated sources, and the data corrected. The principles are understood – but no demonstrations have yet been attempted (to my knowledge), other than for snapshots (NVSS).

4.8.5 Baseline Errors

An error of δu (in wavelengths) in a baseline coordinate gives an error of :

$$\delta\phi = 2\pi\delta u\,\theta \quad \text{radians}$$

in the phase of the visibility. This means a sinusoidal component of the wrong spatial frequency and/or orientation is being placed on your image. The phase error increases with angle, so this problem affects wide-field imaging. How badly?

Suppose we set a tolerable limit of 1 degree in phase. We then find that the offset at which the phase error reaches this is:

$$\theta_{1\text{deg}} \sim \frac{6}{\delta u} \quad \text{arcmin}$$

In the D-configuration, the VLA's baselines are accurate to ~0.2 mm. The one degree error is then reached at an angle of 1.7 degrees at 1420 MHz, or 3.6 arcminutes at 43 GHz. This is good for high-fidelity mosaicing!

But in the A-configuration, the errors are 10X worse, and the fields of view thus ten times smaller – accurate mosaicing will be difficult.

4.9 Computational Problems.

Related to the problems using digital correlators are problems stemming from our use of digital computers, and the regularly sampled grids used in the FFT. Some problems include:
- Sparse sampling in coarse u-v grids. The visibility data don't lie at the centers of the (u,v) cells – but pass nearby. The effect is the same as a baseline error, but is much reduced if there are many data points per cell. Alternatively, you can just make a bigger map (which means the u-v cell size is smaller) – or even use a (slow) DFT.
- Aliasing of sources outside the image. This is caused by the regular grid employed by the FFT. It can be reduced greatly by clever convolution algorithms (but these need a lot of data to work well). Or you can consider a DFT. Note that the *real* sidelobes of an outside image cannot be reduced by convolution or DFTs. You have to map the offending source (either by making a bigger single image, or placing a small map on the source) to remove these.
- Computational Round-Off. The old days of 16-bit integer computations limited dynamic range to ~65000:1. This is no longer a problem.

4.10 Deconvolution Problems

Even if the measured visibility data are perfect (other than noise), important errors can occur in the imaging/deconvolution/self-calibration stages.

Consider observations of a unit point-source. The visibilities all have amplitudes of 1, and phases of 0 (plus/minus some small noise).

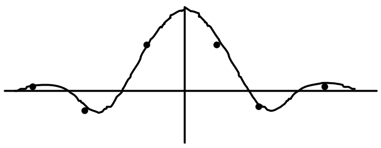
An image of this source, with the object at the center of a cell, will give the expected perfect answer, with the dynamic range ~ 1 billion.

But an image of this source, with the object placed in between two cells, returns an image with dynamic range of tens of thousands. What went wrong?

The problem lies in the deconvolution process and its use of a regularly sampled grid.

The 'dirty map' of our point source extends over many cells in the image. If the object is at the center of the cell, the image and the beam are identical, and a single component is sufficient.

But when the object is between cells, many components are needed.



The black dots are where the object lies on the grid. The 'CLEAN' components are constrained to lie on these positions.
If the CLEAN algorithm can 'find' only the two largest components, our point source has been turned into a double. Wrong answer!

If only the inner four components are found – still the wrong answer, (but better than only two).

In fact, an near-infinite number of components need to be found (i.e., ever gridded value) before the 'right' answer is obtained.

The 'CLEAN' algorithm, in fact, can find only the inner 6 or 8 components, after which it goes wandering around. This results in a deconvolution residual which limits the dynamic range.

This problem is exacerbated when these incomplete sets of components are used in the self-calibration algorithm. (Wrong input -> wrong answer!)

Any bright, bounded resolved object will suffer this problem– especially objects with sharp, unresolved boundaries.

This is in area desperately needing research and development.

Coverage Errors

Finally – one last source of errors to worry about.

Observations of a very extended object with the VLA's 'A'-configuration will result in incomplete sampling of the visibility function, with the most notable effect being that the total flux will be seriously underestimated. In simple terms, the short spacing visibilities (which are by far the largest in magnitude) will be missed, with an obvious 'bowl' being the visible manifestation.

Missing information can, in some cases, be guessed or interpolated in by clever algorithms. But the best remedy is to get the missing information from a smaller 'configuration', or array, or a 'single-dish'.

5 Conclusion (of sorts)

The purpose of this lecture is not to instill depression, or to convince you to change fields.

The principles of synthesis imaging are well established, and the process works beautifully!

Users must understand the limitations of the methodologies, in order to make the best use of it.

The major sources of error are well-understood, and we have good methods for correction.

Most minor sources of error are understood (we think!), and correction methods are under development (or should be!). The next generation of radio arrays will need to make these corrections.

Help in development of these algorithms and methods is needed!