Theory of Phases & Image Reconstruction

in Optical Interferometry

John Young

University of Cambridge

Thanks to: Chris Haniff, John Monnier, J.-P. Berger

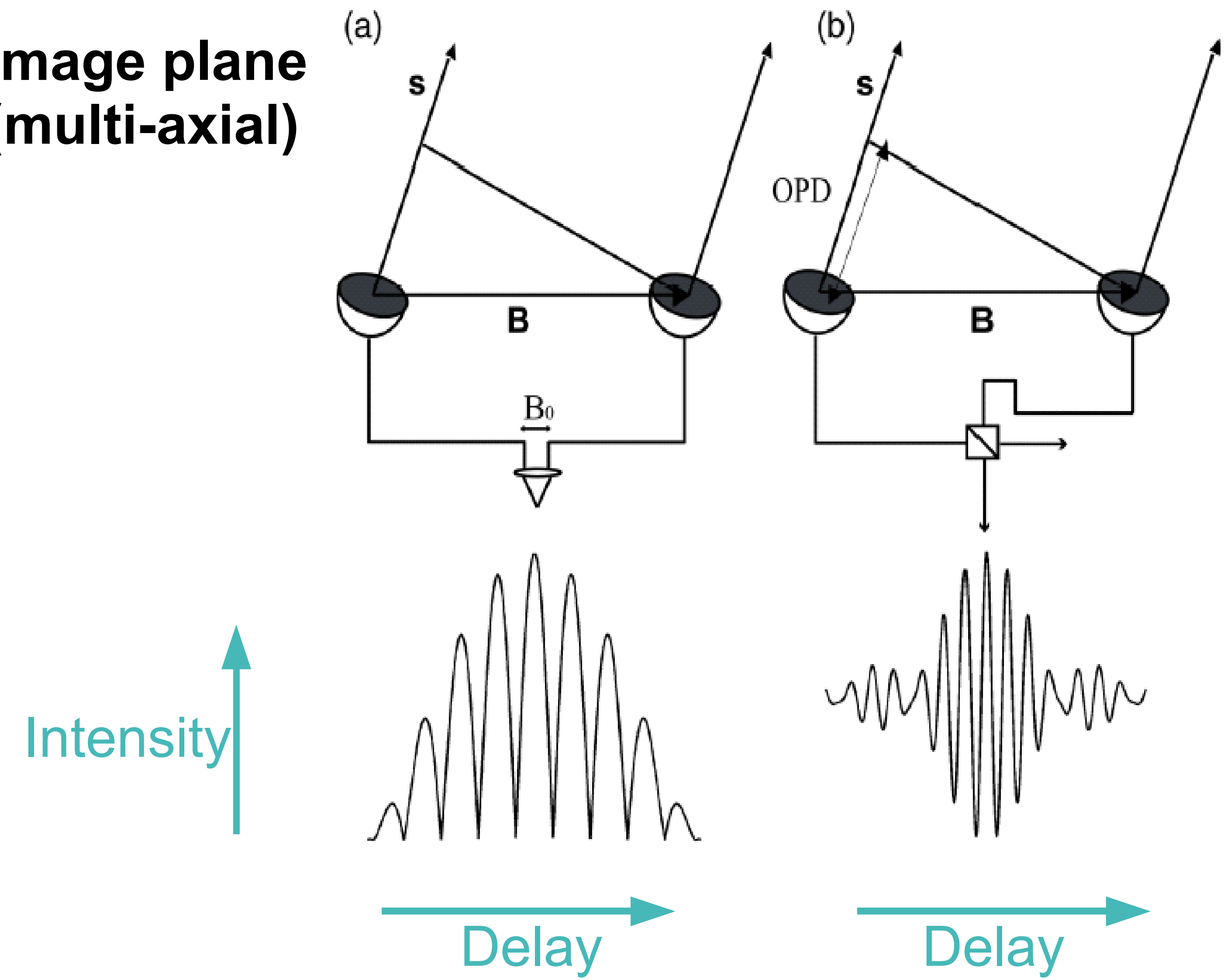
Outline of Talk

- Fringe patterns; amplitude and phase
- Impact of atmosphere
- Techniques to get phase information:
  - Phase Referencing
  - Differential Phase
  - Closure Phase
- Closure Phase
  - Definition and properties
  - Bayesian inference
  - Use of CP in model fitting
  - Use CP in mapping

Interferometers measure fringes (revision)

Image plane (multi-axial)

Pupil plane (co-axial)



Fringe amplitude and phase depend on FT of source brightness distribution (revision)

- Recall sinusoidal interferometer output (fringes):

  $P \propto 1 + \cos(kD), \qquad D = \mathbf{s}.\mathbf{B} + d_1 - d_2$

- This is related to the Fourier Transform of the source brightness distribution:

  $P(\mathbf{s}_0, \mathbf{B}, \delta) = I_{\text{total}} + \mathrm{Re}[V(\mathbf{B}) \exp(-ik\delta)],$

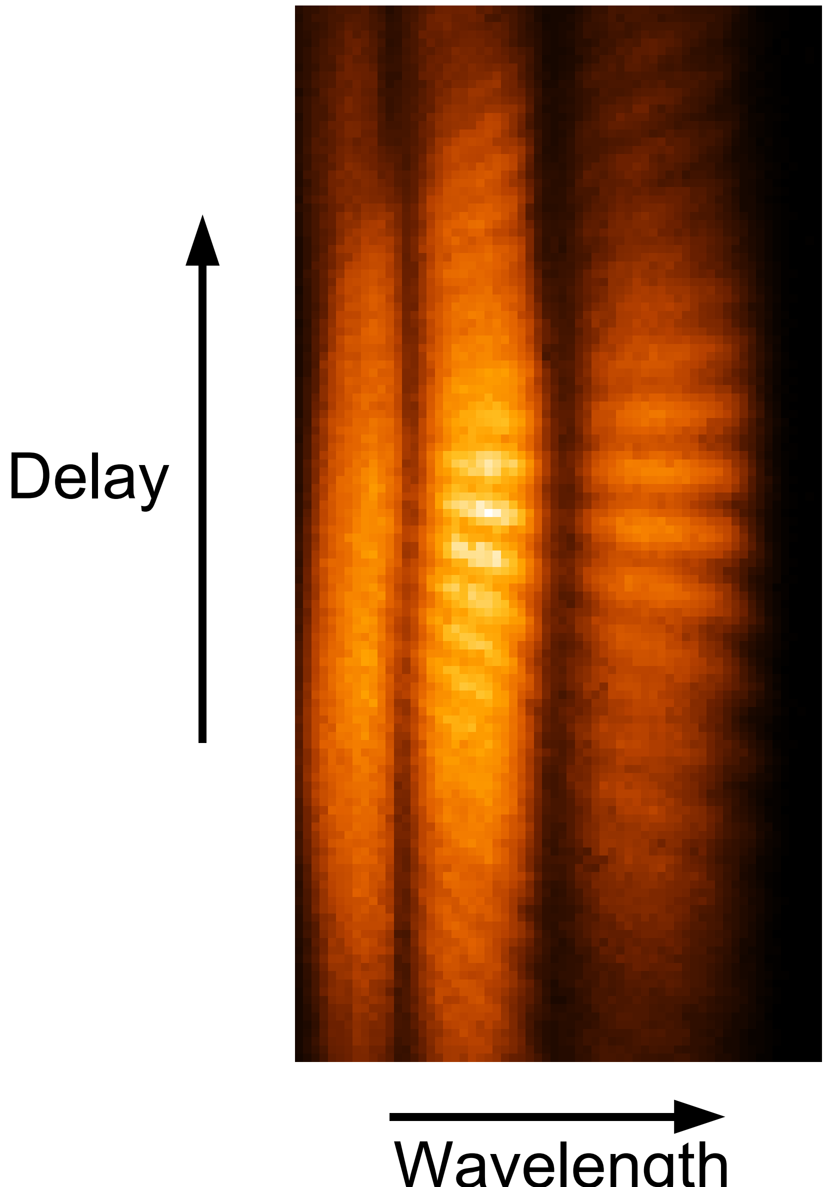
  $V(\mathbf{B}) = V(u,v) = \int I(\alpha, \beta) \exp(-2\pi i\,(\alpha u + \beta v))\, d\alpha\, d\beta$

- δ is offset from expected white-light fringe delay. Measure P at e.g. $\delta=0$ and $\delta=\pi/4$ to obtain complex V

BUT....

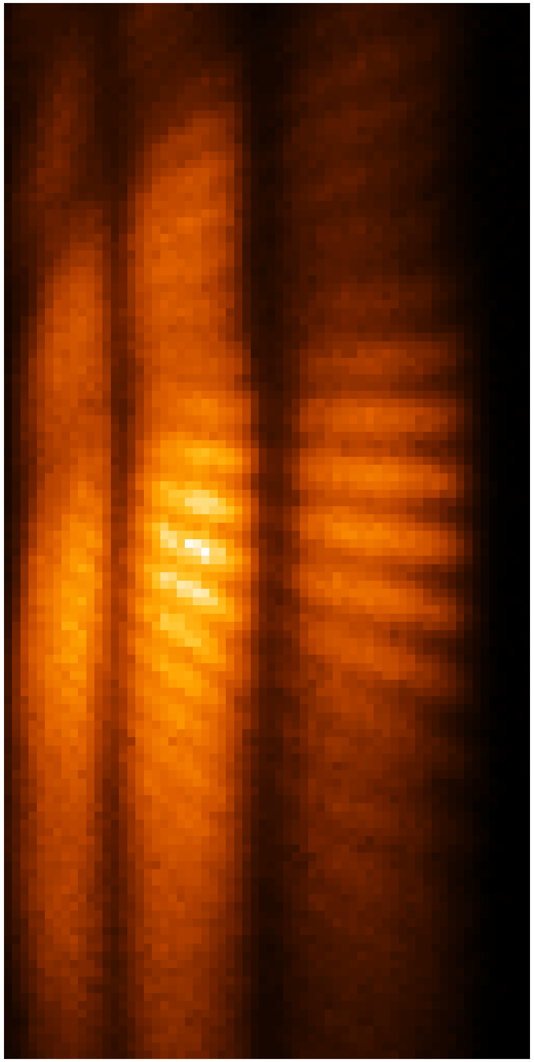
- When we observe through the Earth's atmosphere, the phase of the fringe pattern is perturbed
- We can still measure the phase, but it no longer tells us anything about the source

Actual (dispersed) fringe pattern



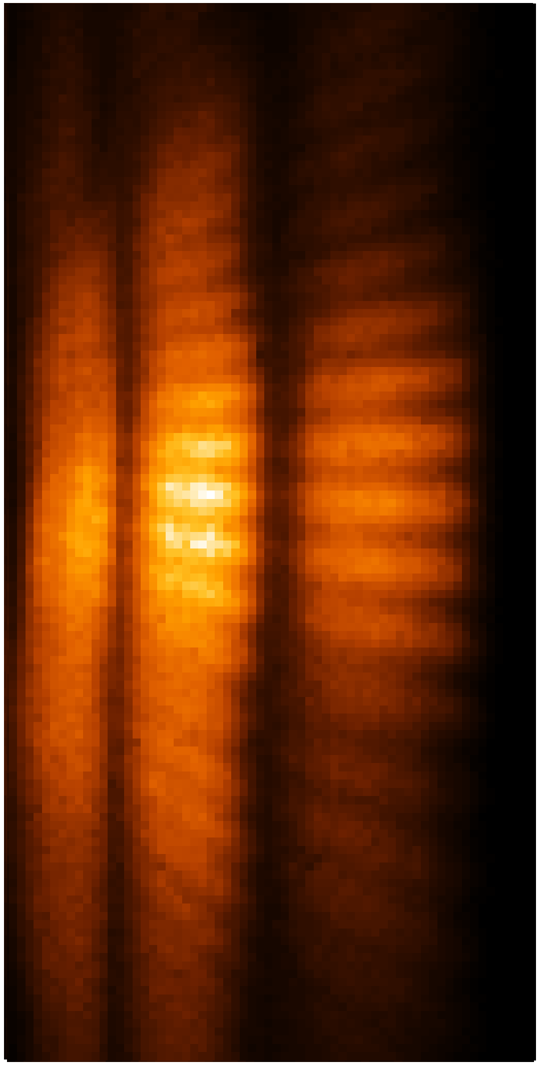
Fringe motion = phase fluctuations

- Note motions are > 1 fringe



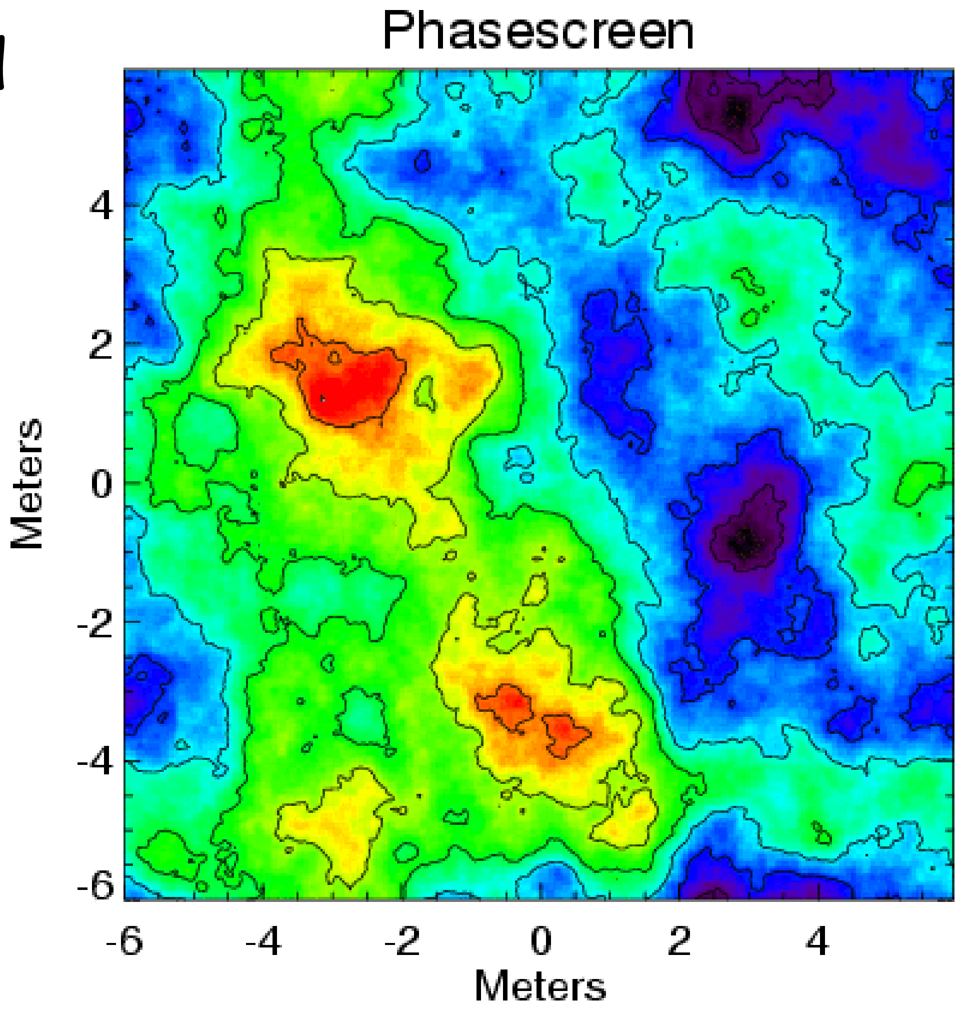
Fringe motion = phase fluctuations

- Note motions are > 1 fringe



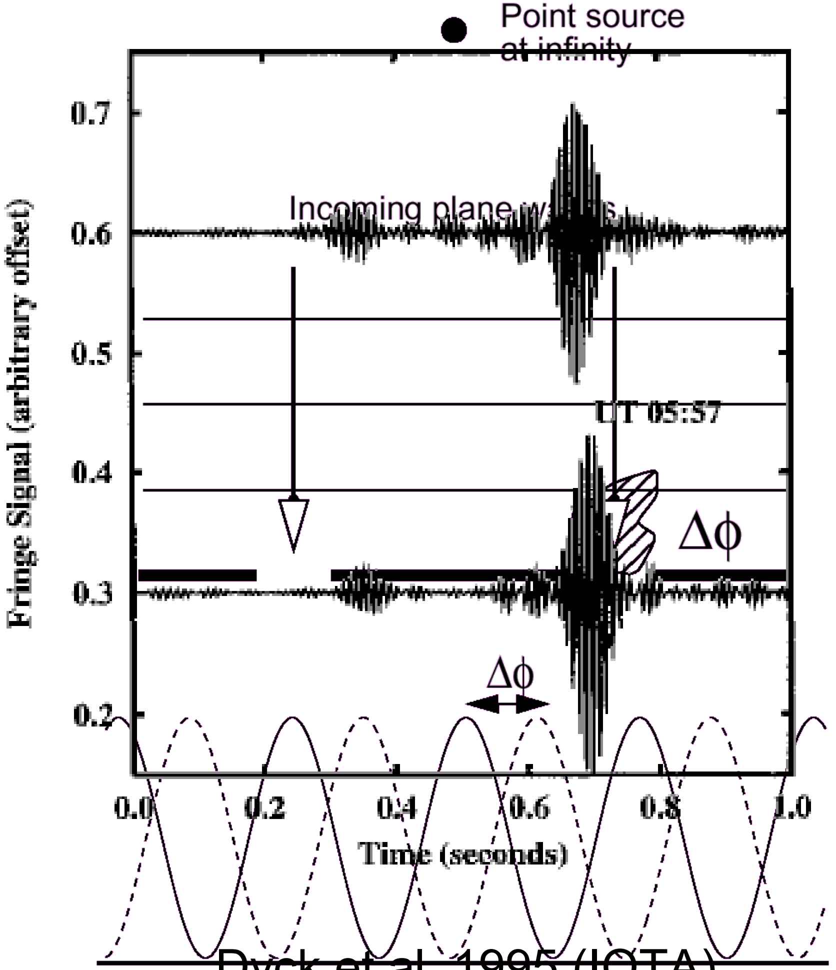
The atmosphere

- A 12m × 12m patch of good atmosphere
- Each contour represents one radian of phase delay for light at a wavelength of 2 microns



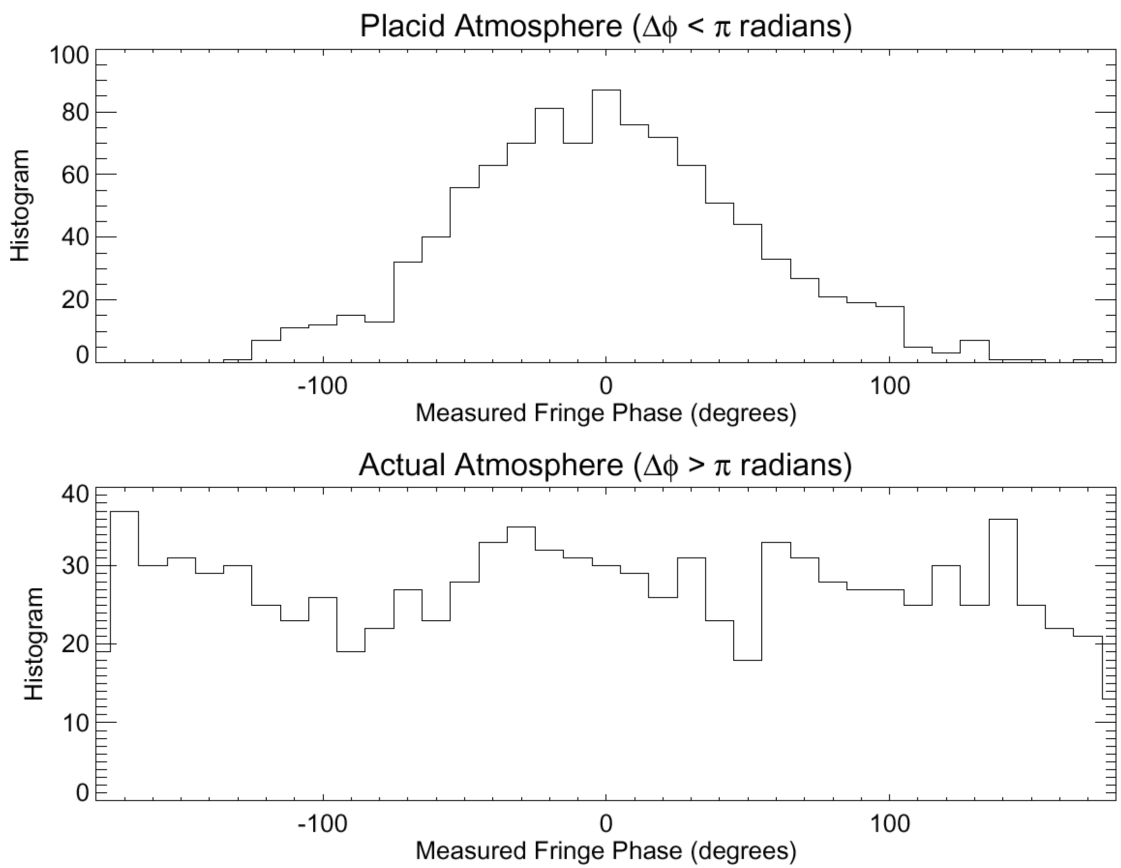
What this does

- Atmosphere introduces an unknown phase delay above each interferometric collector
- Example: suppose only one aperture affected
- Shifts output of interferometer away from expected white-light position



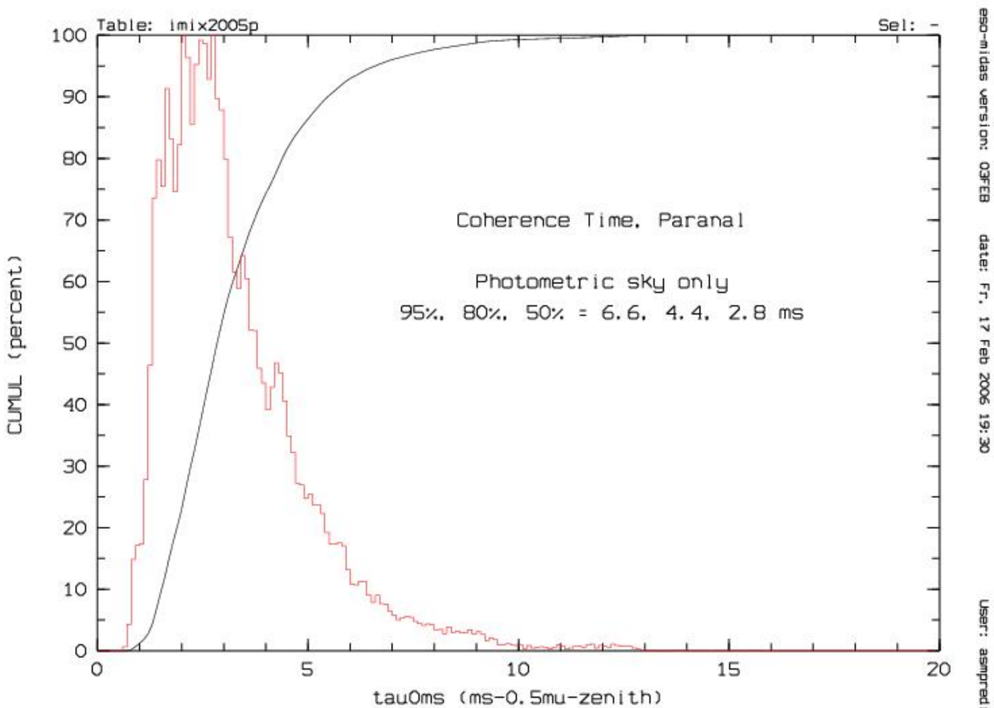
Dyck et al. 1995 (IOTA).

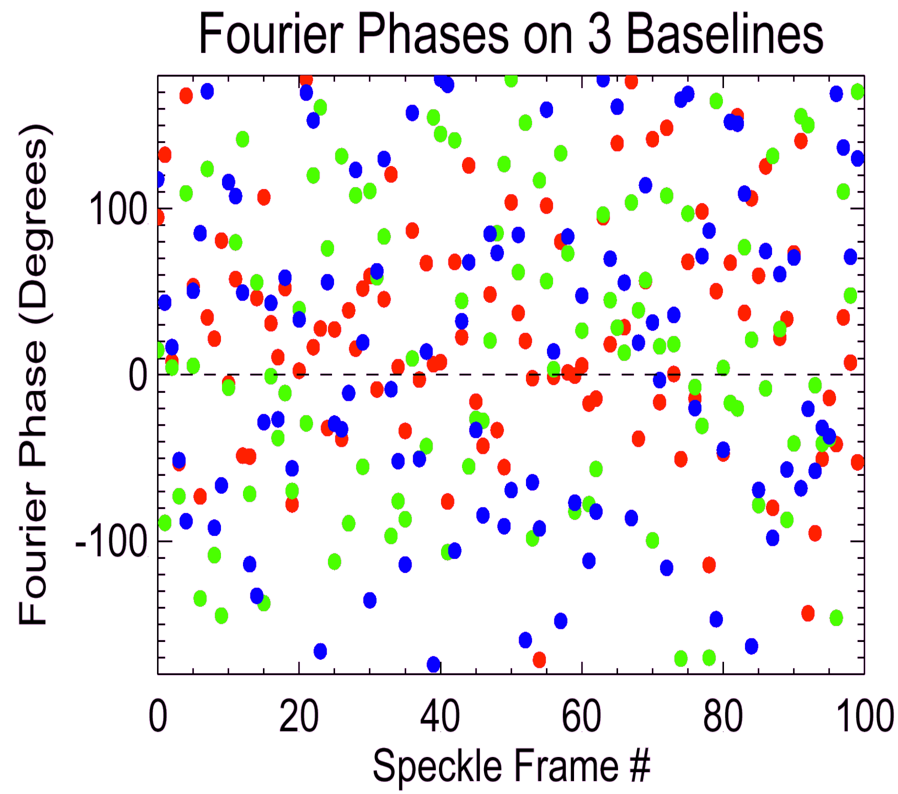
Why we can't average the phase



- Even on a good site in the near-infrared, phase excursions exceed π radians

Timescale for phase variations





- **Coherence time τ_0:**

  Interval over which RMS phase change is 1 rad

  - Scales as $\lambda^{6/5}$, so 3ms at 500nm corresponds to 18ms at 2.2μm
- Must use short exposures to avoid smeared fringes

Recap

- Interferometers measure fringe patterns:
  - Amplitude
  - Phase: location (in delay-space) of white-light fringe
- Fringe amplitude and phase are amplitude and phase of ***one Fourier component*** of the source brightness distribution
- Atmosphere perturbs measured phase by > π rad
- Timescale for phase perturbations is coherence time
  - Tens of milliseconds in NIR
  - Need short exposures even if only measuring amplitude

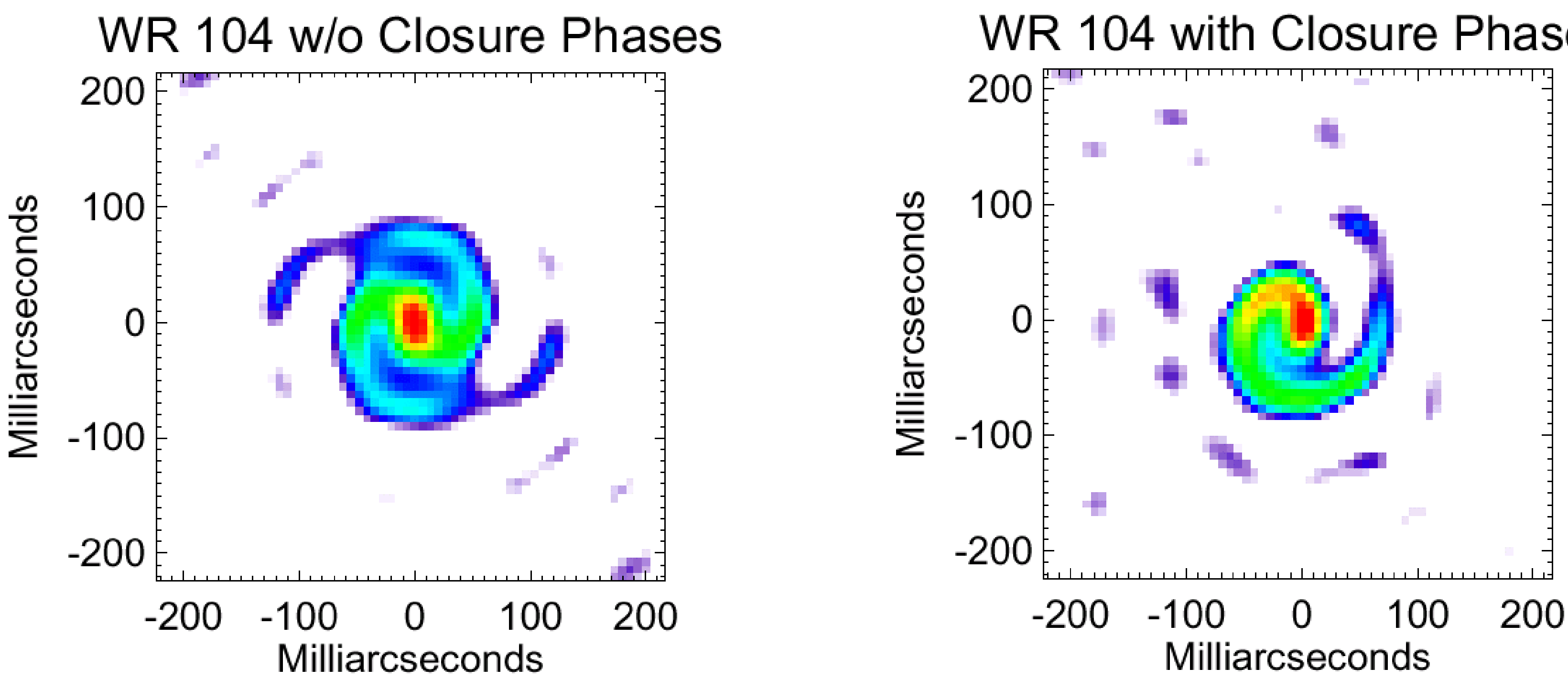
Questions?

?

Why do we care about phase?

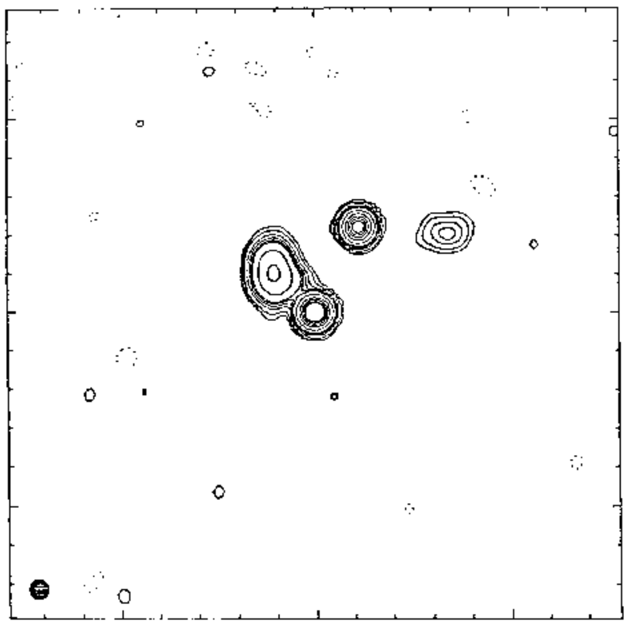
- Why don't we just measure the fringe amplitude then?
- Answer: Depends on what science you want to do. Sometimes just the amplitude is enough, often its not.
- Aside: atmosphere also affects measured visibility amplitude
- Mitigate by interleaving observations of science object and unresolved calibrator star

Most of information is in visibility phases (i)

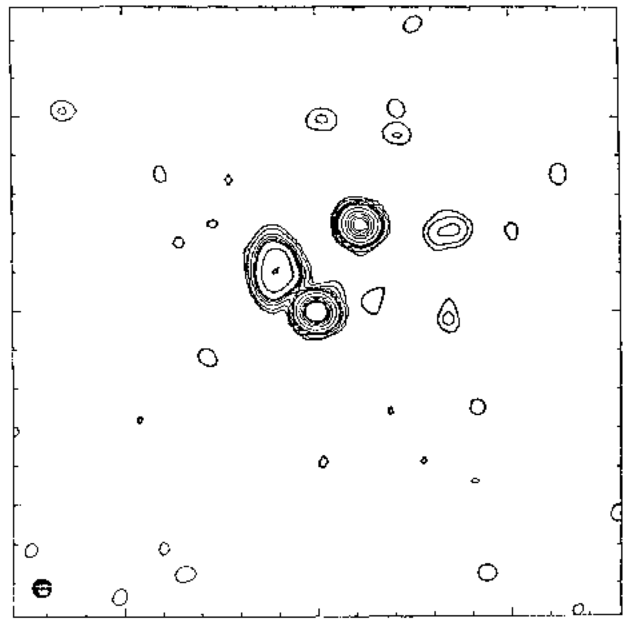


- Without closure phase data, map is necessarily centro-symmetric
- With phase information, correct source brightness distribution is discovered

Most of information is in visibility phases (ii)



Amplitudes and phases

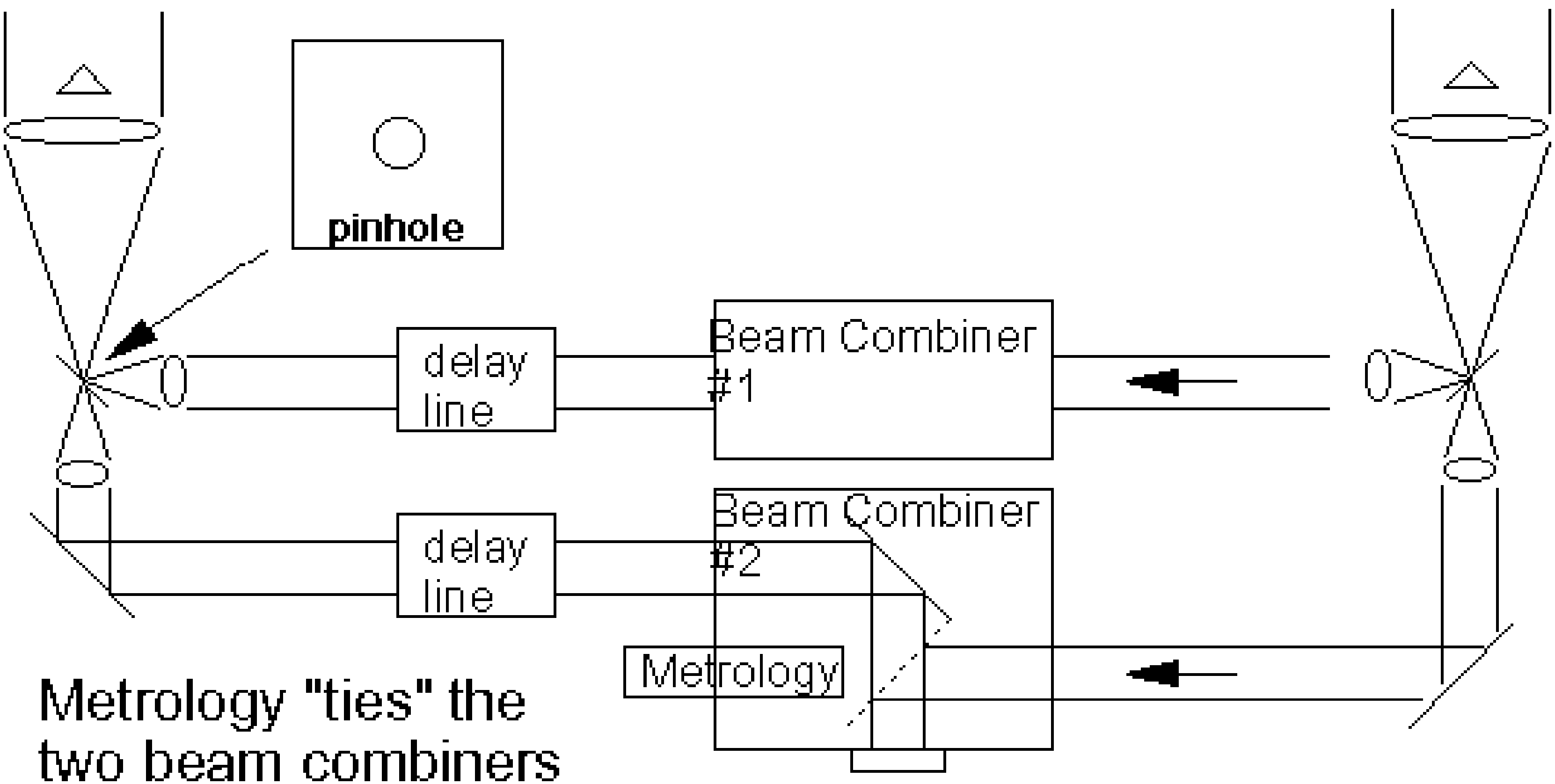


Phases only

Techniques to recover phase information

- Closure phase (most of this talk & practical session)
- Phase referencing
- Differential phase

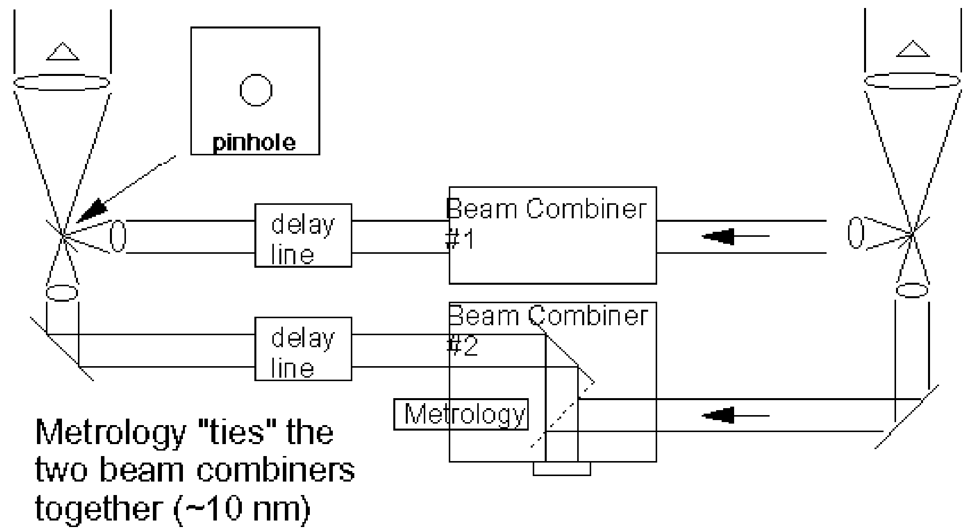
Phase Referencing
(Dual Star Interferometry)



Metrology "ties" the two beam combiners together (~10 nm)

Phase Referencing e.g. PRIMA

- Fringes on 2 sources simultaneously
- Track fringes on reference source using BC#1
- Measure amplitude and phase of science object fringes in BC#2
- Metrology system tells you phase zero-point for BC#2 – measured phase then equals true visibility phase

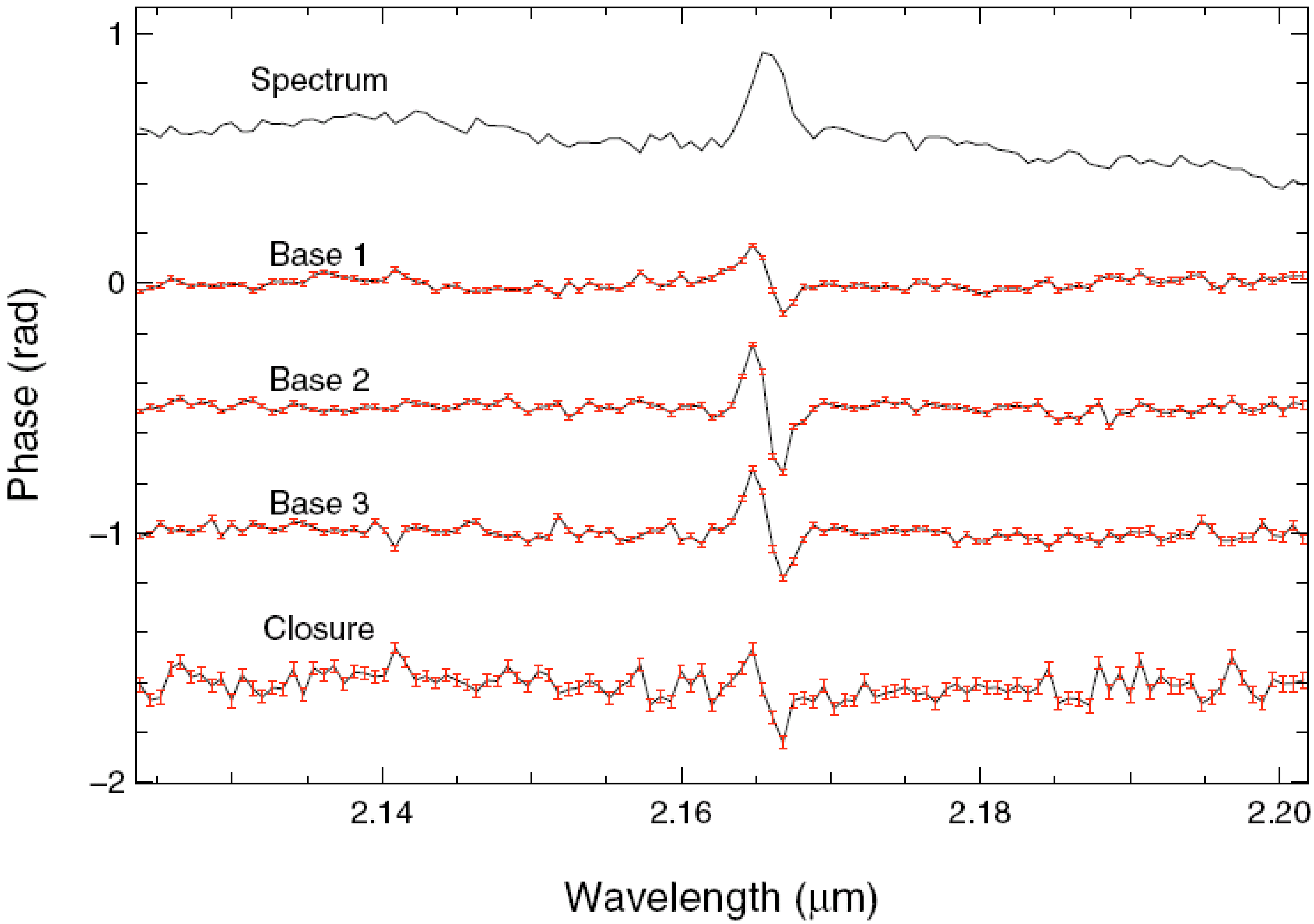


Differential phase e.g. AMBER/MIDI

- Extra hardware not required
- Nearby reference star not required

- Measure fringe phase as function of wavelength
- Model and remove atmospheric dispersion
  $\rightarrow \Phi_{DP} = \phi(\lambda) - \phi(\lambda_{ref})$
- Tells you photocentre shift w.r.t λ_{ref}
  (Fourier shift theorem)
- Need a model for the source to interpret further

Differential Phase Example



- From Meilland et al. (2007) *A&A* 464, 59

The Closure Phase

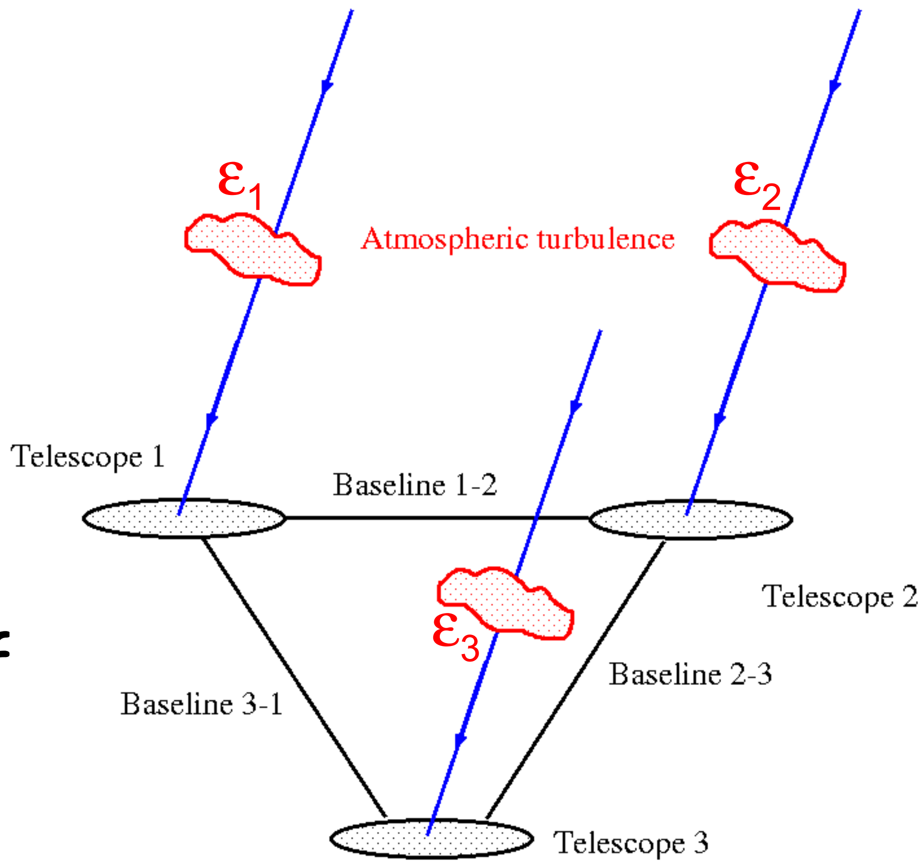
$$\Phi_{12} = \phi_{12} + \varepsilon_1 - \varepsilon_2$$

$$\Phi_{23} = \phi_{23} + \varepsilon_2 - \varepsilon_3$$

$$\Phi_{31} = \phi_{31} + \varepsilon_3 - \varepsilon_1$$

$$\Phi_{12}+\Phi_{23}+\Phi_{31} = \phi_{12} + \phi_{23} + \phi_{31}$$

- Sum of visibility phases around a closed triangle of baselines
- Telescope-dependent errors (e.g. atmosphere) cancel

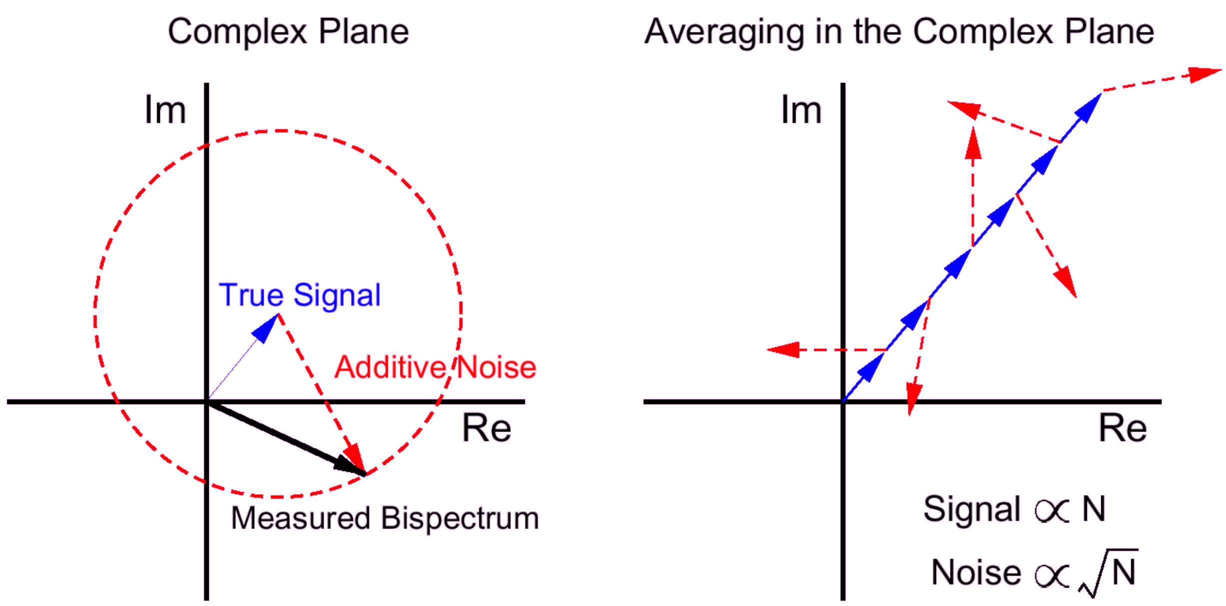


The Bispectrum

- Often more convenient to work with the **bispectrum** (a.ka. “Triple product”) = **product of complex visibilities** around a closed triangle of baselines
- Argument of bispectrum is the closure phase

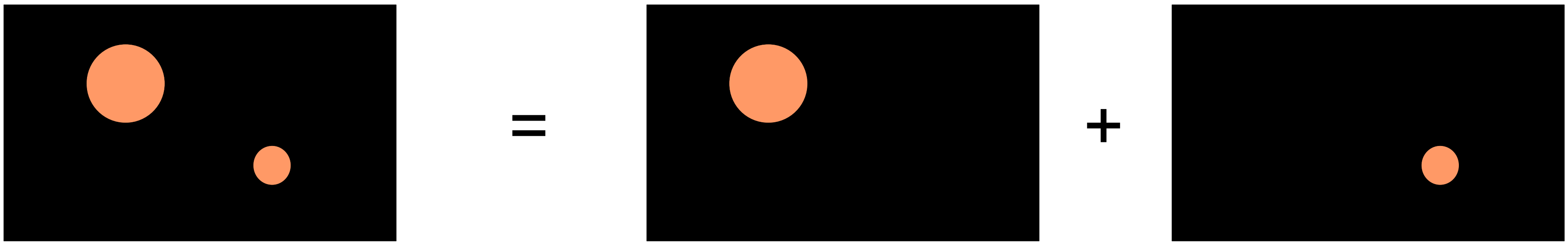
Bispectrum measurements can be averaged

- Successive bispectrum measurements can be averaged in the complex plane
- Average is useful even if SNR of individual measurements is low (even if << 1)



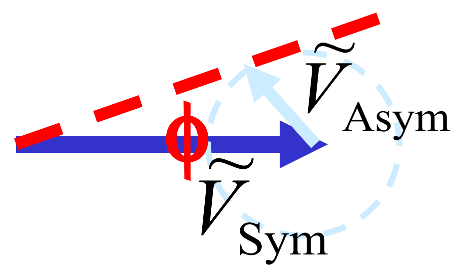
Recap on Visibility Functions

- FT is linear → V(component1 + component2) = V(component1) + V(component2)
  - Remember V is complex
  - Use this to predict bispectrum and hence closure phase for complicated sources



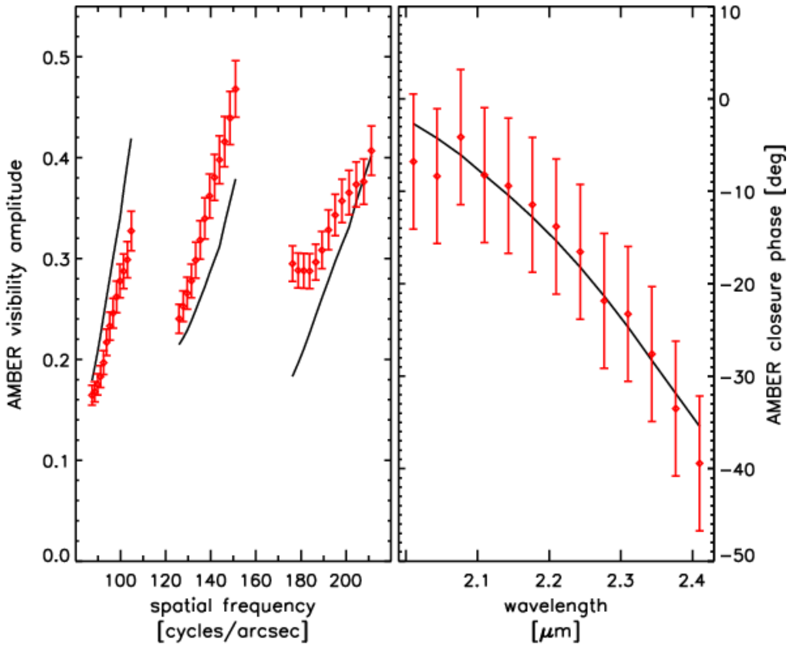
What does a closure phase measure?

- Insensitive to source position
  - Unlike visibility phase
- Point-symmetric sources have CP of 0 or 180°
  - Common examples: symmetric disc, equal binary
- CP measures fraction of asymmetric flux
  - On the angular scale to which you have resolved the source
  - $|\,CP/\mathrm{rad}\,| \approx F_{asymm} / F_{symm}$

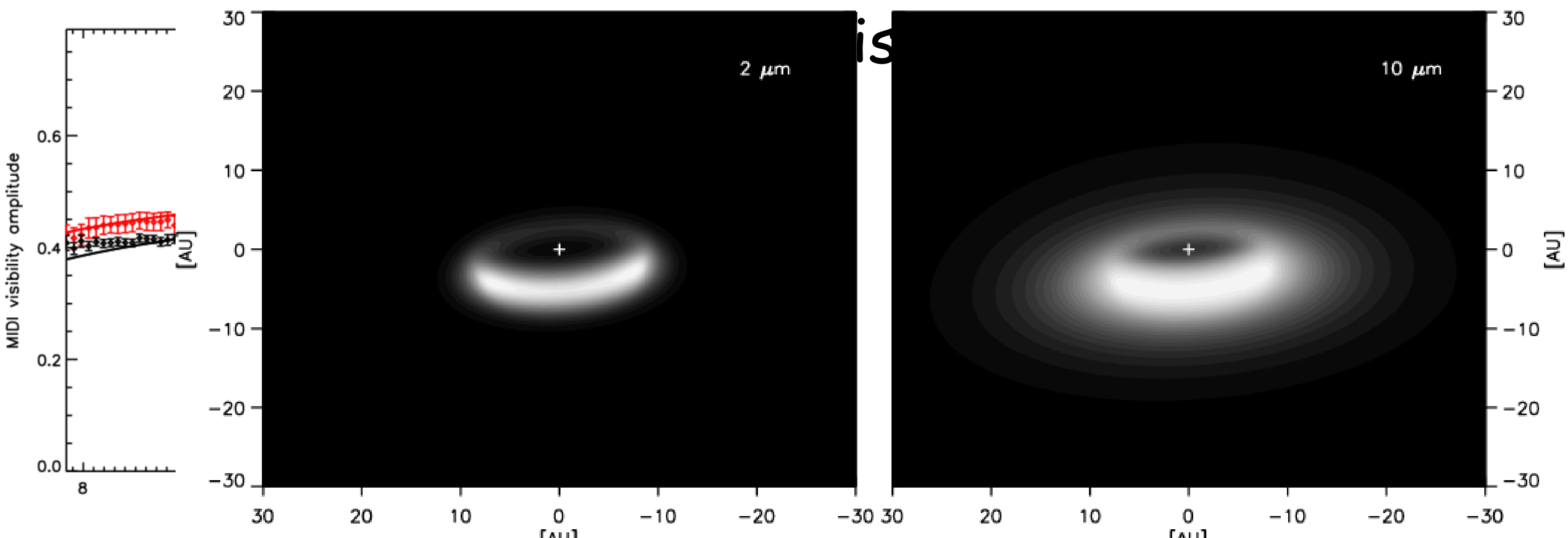


  - With enough closure phases, you can discover the nature of the asymmetry

Closure Phase Example



- Deroo et al. (2007) A&A 474, L45
- CP -> infer presence of inclined dust disk in post-AGB binary
- Note use of spectral
is



Recap: closure phase

- Bispectrum is product of complex visibilities around closed triangle of baselines
  - Closure phase is argument of bispectrum
- Bispectrum is “good observable” in presence of atmosphere
  - Can be averaged over many coherence times
- Closure phase measures fraction of asymmetric flux, on scale at which you have resolved the source

Questions?

?

Important Properties of Closure Phases

- More robust to calibration error than visibility amplitude
  - Atmospheric turbulence generally does not bias measurement
  - Reasonable hope of measurement error reducing as $\sqrt{N}$
  - There can be biases due to chromatic effects (same for visibility amplitude)
- Sensitive to asymmetries in brightness distribution
  - Bispectrum *real* for point-symmetry ($\Phi_{CP} = 0$ or $180°$)
  - Must resolve object to have significant signal
  - Critical for validating model fits to visibility amplitude data
  - Necessary for imaging (if no phase referencing)

How Much Phase Information?

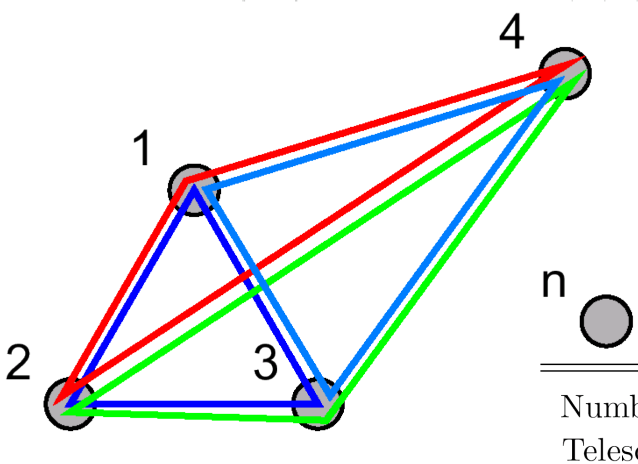
Closure Phases are not all independent from each other

Number of Closure Phases

$$\binom{N}{3} = \frac{(N)(N-1)(N-2)}{(3)(2)},$$

Number of Fourier Phases

$$\binom{N}{2} = \frac{(N)(N-1)}{2}$$



Number of Independent Closure Phases

$$\binom{N-1}{2} = \frac{(N-1)(N-2)}{2}$$

Number of Telescopes	Number of Fourier Phases	Number of Closing Triangles	Number of Independent Closure Phases	Percentage of Phase Information
3	3	1	1	33%
7	21	35	15	71%
21	210	1330	190	90%
27	351	2925	325	93%
50	1225	19600	1176	96%

Using Interferometric Observables

- i.e. statistical inference with squared visibilities, closure phases etc.
- Quantitative interpretation of sparse Fourier-plane data is an **inverse problem**:
  - What can we infer about the sky brightness distribution given the data we have measured?
- The most useful technique is Bayesian inference:
  - Model-fitting (a.k.a. parameter estimation, parametric imaging)
  - Mapping (a.k.a image reconstruction)
- Will also outline non-Bayesian mapping techniques

Classes of Inference Problems

- Given a hypothesis H, are some data D consistent with it?
  - Related problem is *Model Selection*: given two competing hypotheses H_1 and H_2, which is best supported by the data?
- *Parameter Estimation* can be thought of as hypothesis testing as well:
  - Given a family of hypotheses, identical apart from the value of some parameter, which is best supported by the data?

Interpretation of Probability (i)

- Consider hypothesis testing as introduced above
- We can ask "Given a hypothesis H, what is the probability that our noisy measurement process gave rise to the observed dataset?"
  - We can image a series of repeated trials of the measurements, each of which would give rise to a different realisation of the dataset
  - No intellectual leap from thinking about different outcomes of dice throwing or card drawing experiment

Interpretation of Probability (ii)

- But we only have one realisation of the dataset
- We are interested in the **Inverse Problem**: what can we say about the validity of the hypothesis given the data?
- A Bayesian would ask "Given the data, what is the probability that the hypothesis H is true?"
  - Requires a broad interpretation of the concept of probability

Bayes' Theorem (i)

- A Bayesian would ask "Given the data, what is the probability that the hypothesis H is true?"
  - Willing to interpret probability as reflecting their degree of belief in a particular hypothesis
- Accepting this, we can now use the sum and product rules for combining probabilities
- Hence derive *Bayes' Theorem* (after Rev. Thomas Bayes):

$$P(H|D, I) = P(D|H, I) \cdot P(H|I) \,/\, P(D|I)$$

Posterior = Likelihood · Prior / Evidence

- I is background information, assumptions
- Often, just proportionality is useful:

$$P(H|D, I) \propto P(D|H, I) \cdot P(H|I)$$

Bayes' Theorem (ii)

- *Posterior probability* is what we are interested in knowing = probability of H given measured dataset
- *Likelihood* is probability of obtaining our particular realisation of the dataset given H – we hopefully know enough about the measurement process to calculate this
- *Prior* encodes our *a priori* knowledge about H. Critics of Bayesian methods have particular problems with this:
  - Subjective
  - Hard to encode complete ignorance
  - But if changing prior alters your conclusions, you need more data!

Hypothesis Testing (Bayesian)

- Usually couched as a *Model Selection* problem: given two competing hypotheses H_1 and H_2, which has the highest posterior probability?
- Usually choose a simple uninformative (as possible) prior, then model selection boils down to calculating the likelihood = probability of obtaining measured data given hypothesis
- Slightly more complicated if competing hypotheses have different numbers of unknown parameters – but Bayes' theorem allows this to be handled objectively, by *marginalisation*

Assigning Probabilities

- Hypothesis testing requires calculation of the *likelihood* = prob. of obtaining measured data given hypothesis
- To make progress we must think about the *noise model* for our data, i.e. how to assign the probability of measuring certain values, given a hypothesis concerning e.g. parameter(s) we want to infer from the data

Example: Gaussian Noise

- Suppose N data x_i, measurements of a quantity whose true value is μ
- Gaussian noise, **known** standard deviation σ
- $P(x_i) = 1/(\sigma\sqrt{(2\pi)}) \exp[-(x_i-\mu)^2/(2\sigma^2)]$
- What is *likelihood* of dataset $\{x\}$ given particular value of μ?
- By product rule for combining probabilities:
  $P(\{x\}|\mu) = \prod P(x_i) = \prod (1/(\sigma\sqrt{(2\pi)}) \exp[-(x_i-\mu)^2/(2\sigma^2)]$
- Often more convenient to work with natural log:
  $\ln P = \text{const.} - \sum (x_i-\mu)^2/(2\sigma^2)$

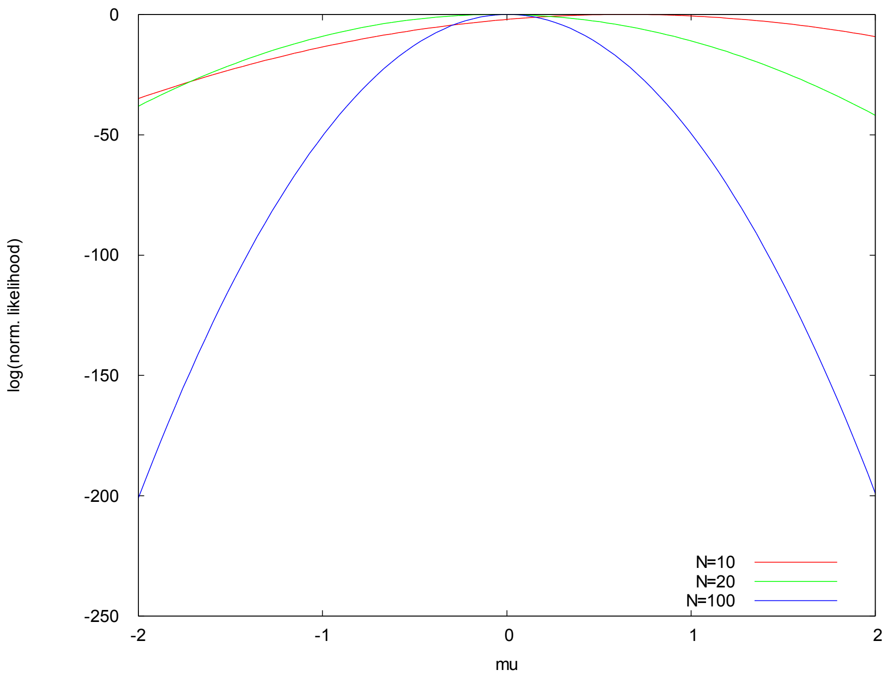
Gaussian Noise: Bayesian Approach

- From previous slide:
  $\ln P = \text{const.} - \sum (x_i - \mu)^2/(2\sigma^2)$
- Taking a Bayesian approach, for case of a *uniform prior* (over a certain range), posterior probability $P(\mu|\{x\})$ is proportional to likelihood – only likelihood depends on μ
- Best estimate of μ is given by maximum of the posterior pdf. This is a *Maximum Likelihood* method
- Unsurprisingly, P (and ln P) is maximised for $\mu_0 = (\sum x_i)/N$
  i.e. the mean of the samples $\{x\}$ is the best *estimator* of the population mean

Gaussian Noise: Posterior pdf

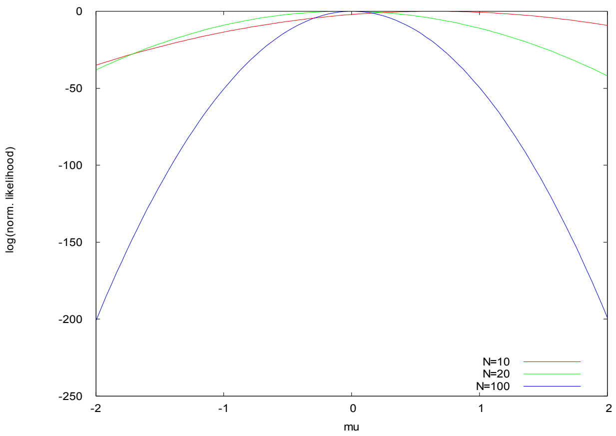
- $\ln[P(\mu|\{x\})]$ for true $\mu=0$, $\sigma=1$:
- As N increases
  - Pos'n of peak gets closer to true μ
  - Peak becomes sharper

i.e. mean of measurements gives better estimate of true μ as we acquire more data – this is generally true in parameter estimation provided you have designed a good experiment!



Posterior pdf

- In general, the (log) posterior pdf may be
  - Asymmetric
  - Multi-modal i.e. have multiple peaks
- In the region of a maximum X_0, we can always perform a Taylor expansion (note 1st derivative is zero):
  $\ln[P(X|\{x\})] = L(X|\{x\}) = L(X_0) + \frac{1}{2}\, d^2L/dX^2|_{X=X0}(X-X_0)^2 + \ldots$
  (now using X, rather than μ, as symbol for parameter under test)
- Taking the exponential again:
  $P(X|\{x\}) \approx A \exp[\frac{1}{2}\, d^2L/dX^2|_{X=X0}(X-X_0)^2]$



Posterior pdf

- So after Taylor expansion of ln P we have:
  $P(X|\{x\}) \approx A \exp[\frac{1}{2}\, d^2L/dX^2|_{X=X0}(X-X_0)^2]$
- We've approximated the posterior as a Gaussian pdf
  - With $\mu = X_0$, $\sigma_X = (-d^2L/dX^2|_{X=X0})^{-1/2}$
  - Hence interval $[X_0-\sigma_X, X_0+\sigma_X]$ contains **approximately** 68.3% of the posterior probability; $[X_0-2\sigma_X, X_0+2\sigma_X]$ contains 95% etc.
  - i.e. σ_X gives the "error bar" on the parameter estimate X_0
- In the Gaussian noise example, approximation is exact
  - $d^2L/dX^2|_{X=X0} = -N/\sigma^2$
  - Best estimate of mean of Gaussian process is $(\Sigma x_i)/N \pm \sigma/\sqrt{N}$

Chi-square Fitting

- If we have N data points (RVs) $y_i(x_i)$ with independent Gaussian errors σ_i, **maximum likelihood model-fitting corresponds to minimising chi-squared**. To see this:-
- $P(y_i|\{a\}) \propto \exp[-(y_i - y(x_i,\{a\}))^2/(2\sigma_i^2)]$,
  where $y(x_i,\{a\})$ is model-predicted value given params $\{a\}$
- Likelihood of dataset $\{y\}$ given particular $\{a\}$ is:
  $P(\{y\}|\{a\}) = \prod P(y_i) \propto \prod \exp[-(y_i - y(x_i,\{a\}))^2/(2\sigma_i^2)]$
- Maximising P is equivalent to maximising ln P, given by:
  $\ln P = \text{const.} - \sum(y_i - y(x_i,\{a\}))^2/(2\sigma_i^2) = \text{const.} - \chi^2/2$
- Obviously, maximising $-\chi^2/2$ is the same as minimising χ^2

Model-fitting in practice

- Lots of literature about how to minimise χ^2 (and other merit functions)
- Can choose to minimise χ^2 even if σ_i's are not Gaussian, but:
  - No longer a maximum-likelihood estimate
  - Cannot legitimately perform chi-square goodness-of-fit test
- Distinction between problems that are linear in {a} and those that are not – usually non-linear for OI
- Beware of:
  - May be several local minima in $\chi^2(\{a\})$
  - Unhelpful topology of χ^2 hyper-surface

Interpretation of best-fit χ^2

- Models with unlikely χ^2/ν values (say 3-5) are often described as "acceptable"
- Slightly high χ^2 may be wholly or partly due to underestimated or non-normal σ_i's
- May also indicate that there is some element of the physics that is unmodelled, but model may still be useful:
  - Good fit over some range(s) of *x*
  - Probably captures some physics, can form basis for more realistic model
  - Make useful/testable predictions
- Conversely χ^2<<1 suggests over-complex model is fitting noise and/or data/bins not independent

Model Fitting with Closure Phases

- Conventional techniques used to fit model for source brightness distribution to measured visibility amplitudes and closure phases/bispectra
  - Least-squares (only if equally-sized errors)
  - Bayesian: minimise negative log posterior probability

    $L = \text{Prior} + \Sigma\,(D_i - M(\boldsymbol{a}))^2/(2\sigma_i^2)$

    by varying vector ***a*** of model parameters
    - Here D_i is Squared Visibility or Closure Phase ($D_i - M$ calculated modulo 2π)
- Model is either:
  - Sum of uniform discs, elliptical Gaussians etc.
  - Output of radiative transfer code

Recap: Bayesian Model-fitting

- Recall *Bayes' Theorem*:

  P(H|D, I) = P(D|H, I) · P(H|I) / P(D|I)

  Posterior = Likelihood · Prior / Evidence

- If we have data with independent Gaussian errors, maximum likelihood model-fitting corresponds to minimising chi-squared.
- For sparse Fourier plane data, the chi-squared hyper-surface often has unhelpful topology
- We can derive error bars on fitted parameters from the shape of the peak in the posterior pdf

The Imaging Problem

- Suppose we wish to reconstruct a pixellated (model-independent) image instead

- Sampling of the (u,v) plane is necessarily incomplete
- In other words we have only measured some of the spatial frequencies (given by B/λ) in the sky brightness distribution

- Unless we do phase referencing, the phase of each measured visibility is unknown
- We only have linear combinations of the visibility phases = closure phases

Sampling

- We have sampled the visibility *V* at various points (u_i,v_i)
- In other words we have measured the product of FT(sky brightness) and a sampling function *S*(*u*,*v*) = sum of delta functions
- We can invert this to give a dirty image
- By the convolution theorem, the dirty image is the *convolution* of the true sky brightness with the dirty beam

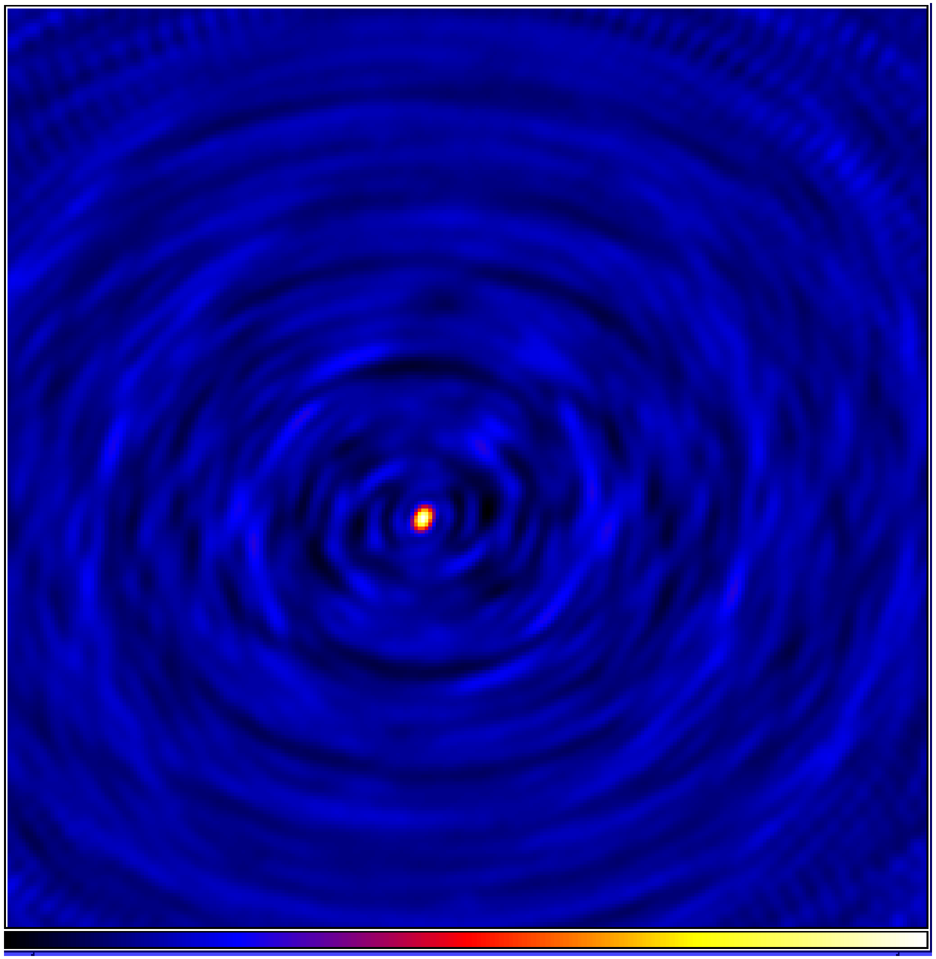
$$S(u,v)V(u,v) \overset{FT}{\Leftrightarrow} \underbrace{\Im S * \Im V}_{\text{Dirty image}} = \underbrace{\Im S}_{\text{Dirty beam}} * I$$

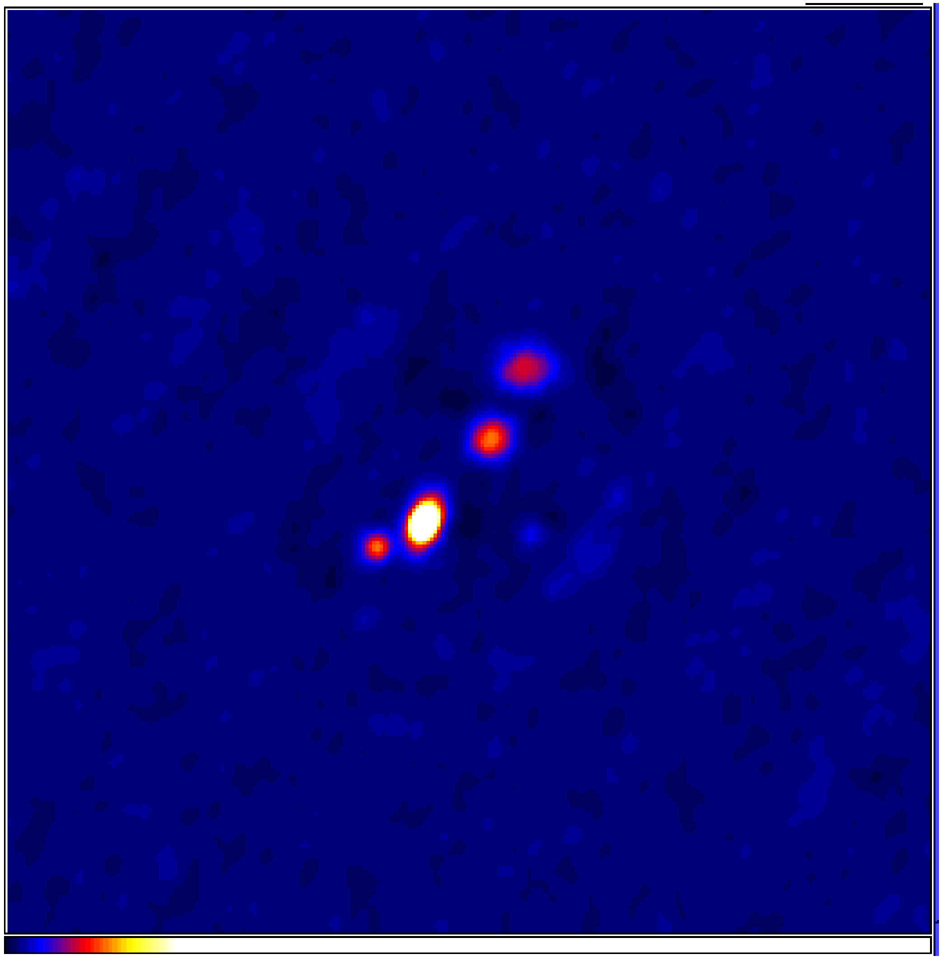
- The dirty beam is the Fourier Transform of the sampling function *S*(*u*,*v*)

Deconvolution in Principle

- There are an infinite number of maps that fit the data
- Most of these are physically unreasonable
- Need to use prior knowledge
- Universal:
  - Positivity
  - Finite extent (support)
- CLEAN-specific
  - Map consists of a relatively small number of point sources (convolved with a clean beam)
- MEM-specific
  - Map has a compressed range of pixel values
  - i.e. map is smooth in some sense

Deconvolution Does Work!





Mapping with Closure Phases: (a) Iterative Deconvolution

- Closure phases used as a constraint in an iterative scheme
  - Assign some phases consistent with the closure phases using a model
  - Fourier Invert→ dirty map
  - Deconvolve dirty map (e.g. CLEAN)→ new model
  - Start over with new model
  - Unless procedure has converged

Why do iterative schemes converge?

- In early iterations, models do not fit data perfectly
  - Generally over-simplified
- model = true sky + "error distribution"
- "Error distribution" produces errors in phases assigned to the data
- Hence get spurious features in dirty map. But these are
  - Weak
  - Spread over a large area
- Hence don't get incorporated into new model
- Deconvolution filters out the error distribution

Mapping with Closure Phases: (b) Fitting with Regularization

- Fit model consisting of pixel values to visibility amplitude and closure phases
- No unique solution, so constrain using prior knowledge:
  - Positivity
  - Limited Field of View
  - Regularization term to favour "simple" solutions
    - e.g. Maximum Entropy: compressed range of pixel values
    - e.g. local smoothness
  - Note that the iterative schemes outlined previously incorporate prior knowledge *implicitly*

Maximum Entropy (MEM/Maxent)

- Fit pixellated model (I_k) to data with the extra constraint that the "entropy" S is maximised

$$S = -\sum I_k \ln \frac{I_k}{M_k}$$

- Produces image with a compressed range of pixel values
  - *Hence image is "smooth" (but not locally)*
- M_k are pixel values for default image
  - *Allows specific prior knowledge to be incorporated*
- In practice constraints are maximum entropy and that χ^2 has its expected value, so maximise $\alpha S - \chi^2/2$

Entropic Prior

Monkeys throw a large number, *M*, of balls into *N* buckets (flux quanta into pixels).

Probability that they end up with configuration $\{n_i\}$ is

$$\mathrm{pr}(n_i \mid M, N) = N^{-M} \frac{M!}{n_1! n_2! \ldots}$$

Use Stirling's approximation, $\log(n!) \approx n \log(n) - n$

$$\log[\mathrm{pr}(n_i \mid M, N)] = -M \log N - \sum_{i=1}^{N} n_i \log\left[\frac{n_i}{M}\right]$$

Let the fraction of peanuts be, $f_i = n_i / M$

$$\log[\mathrm{pr}(f_i \mid M, N)] = -M \log N - M \sum_{i=1}^{N} f_i \log f_i$$

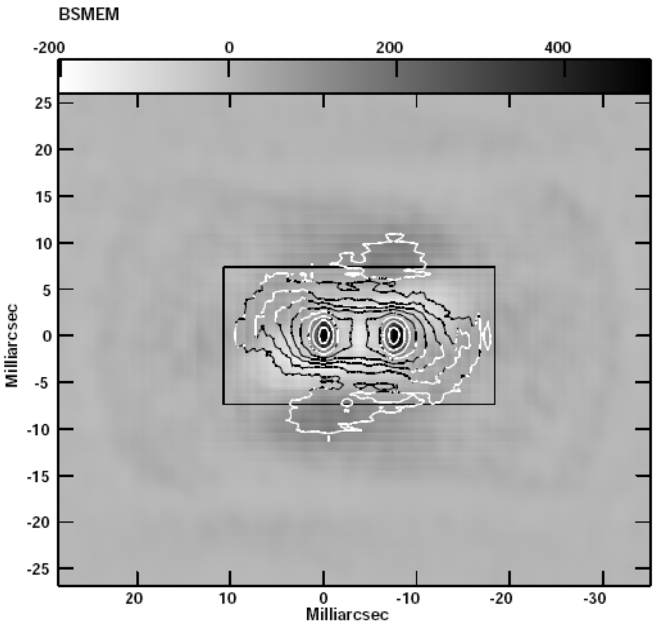
$$\mathrm{pr}(f_i \mid M, N) \propto \exp(\alpha S)$$ "entropic" prior pdf

$$S = -\sum_i f_i \log f_i$$ configurational entropy

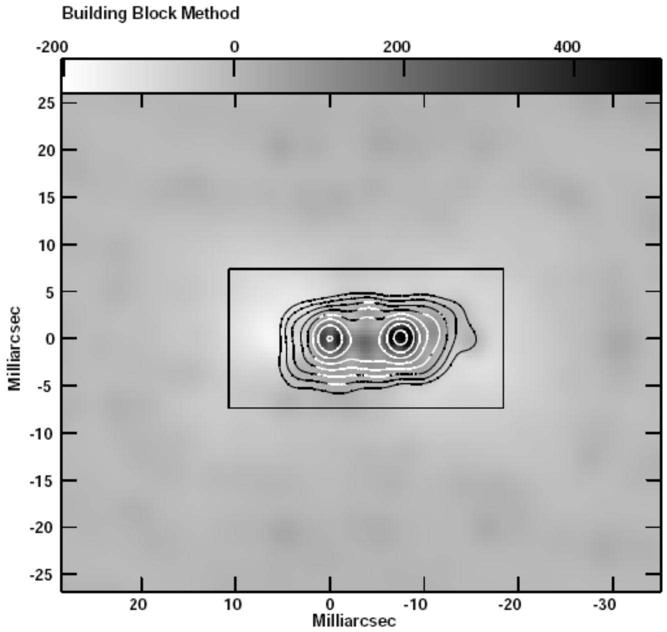
Fitting with regularization in practice

- Want to avoid doing inverse transform from data to model space (missing phase information)
- Start in model space, with initial (default) image
- Calculate merit function (S and χ^2 for MaxEnt)
- Determine improved model (gradient search, or model-space algorithm)
- Iterate

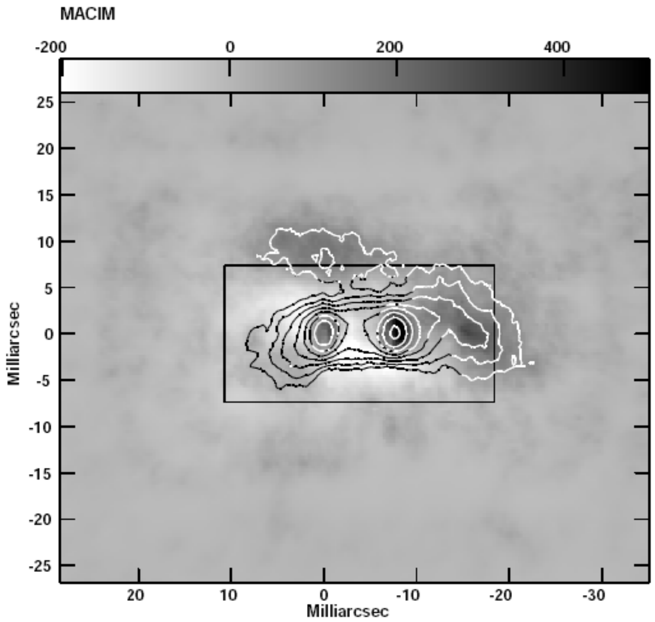
Available OI imaging codes



BSMEM (Cambridge)



Building Block Method (Bonn)



MACIM (Caltech)

- Also MIRA (Lyon;this school) and Wisard (ONERA)

Image Quality

- Dynamic Range: Ratio of peak brightness to faintest believable feature
  - Limited by errors on visibility data (random, systematic)
  - Few 100:1 typical for optical interferometry
- Image Fidelity: How close map is to true image
  - Hard to quantify!
  - Clearly dependent on (u,v) coverage

Rules of Thumb

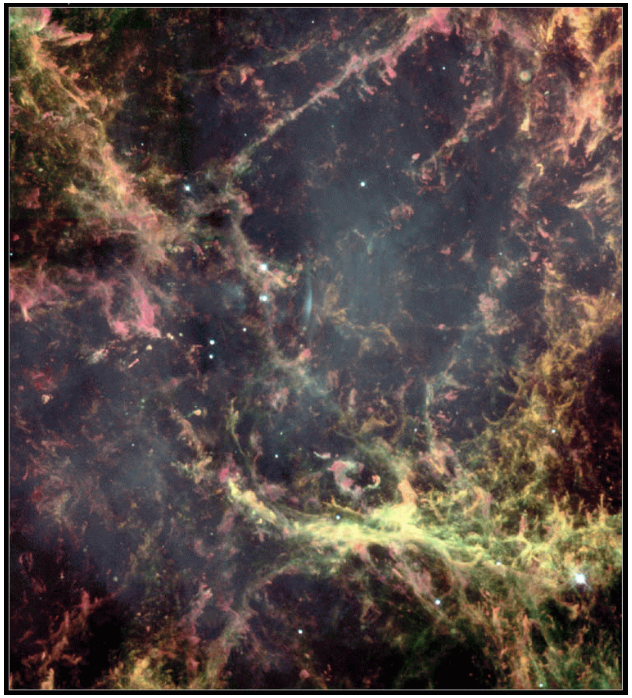
- The number of visibility data $\geq$ number of filled pixels in the recovered image
- The distribution of samples should be as uniform as possible
  - To aid the deconvolution process.
- The range of interferometer baselines, i.e. B_{max}/B_{min}, will govern the range of spatial scales in the map
- There is no need to sample the visibility function too finely
  - For a source of maximum extent θ_{max}, sampling very much finer than $\Delta u \approx 1/\theta_{max}$ is unnecessary.

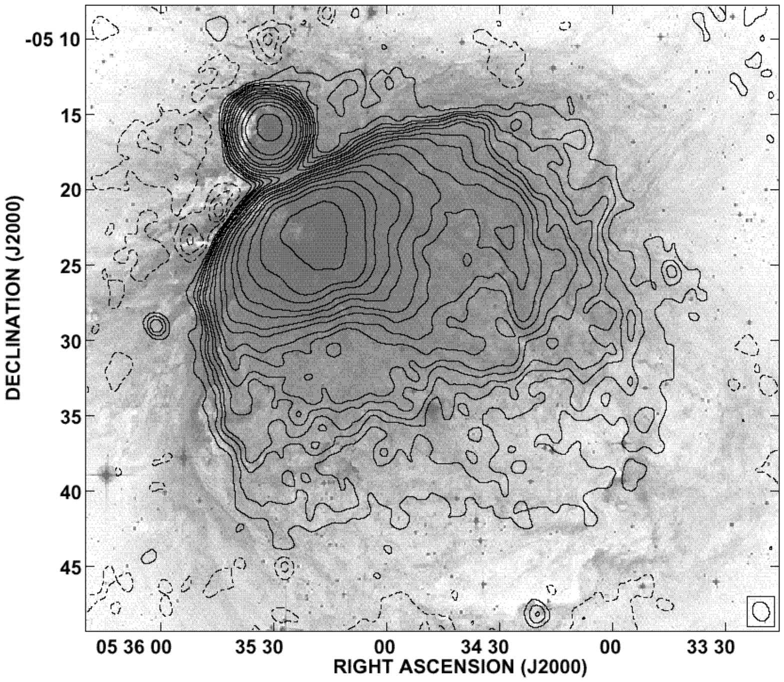
Field of View (revision)

The field of view will be limited by:

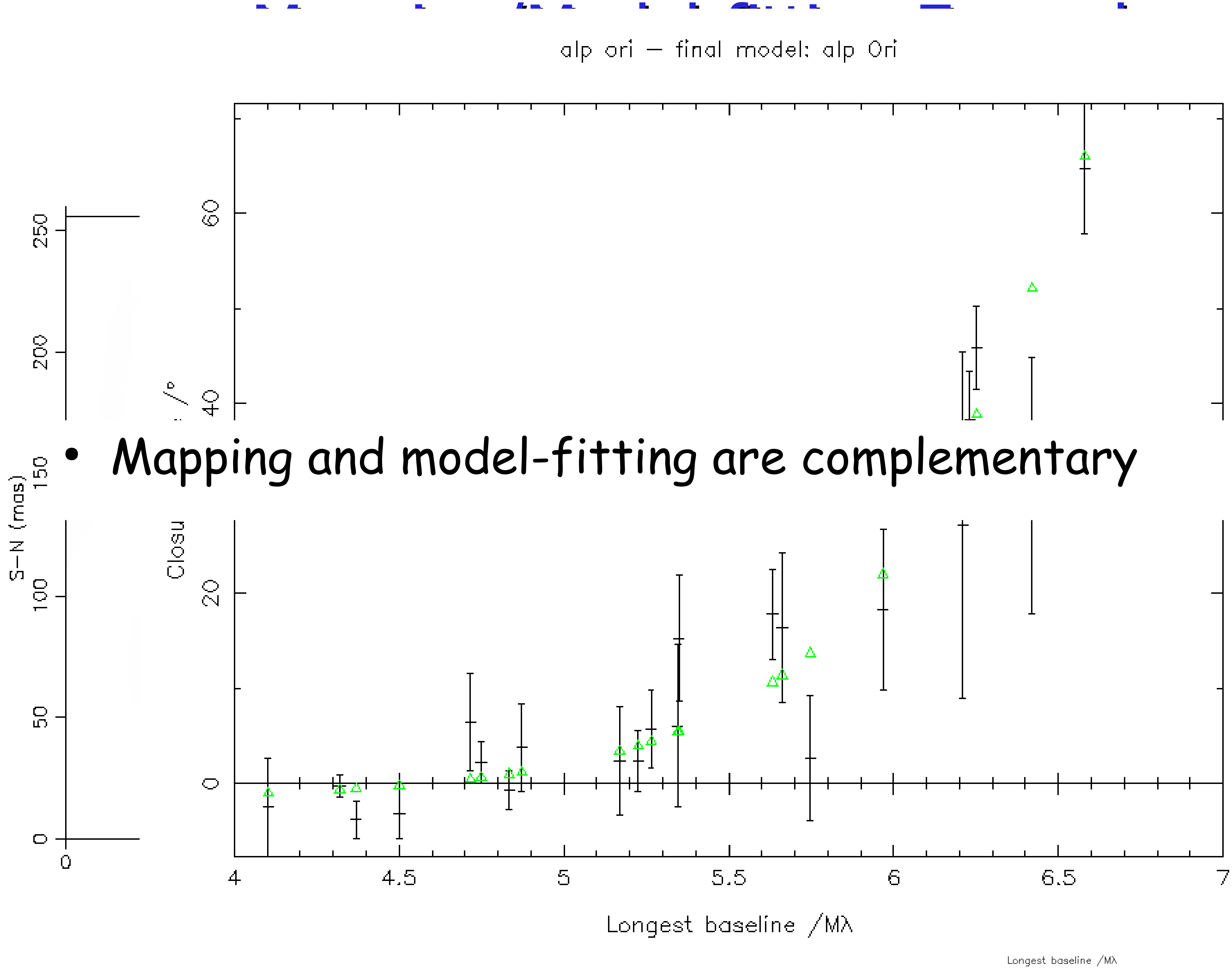
- Primary beam of the collectors
- Spectral resolution
  - OPD < $\lambda^2/\Delta\lambda$ must be satisfied for all field angles
  - Generally $\Rightarrow$ FOV $\leq [\lambda/B][\lambda/\Delta\lambda]$
    i.e. (spatial resolution)$\times$(spectral resolution)
- Shortest baseline in the array
  - Must sample low spatial frequencies i.e. large scales
- Chosen map size

Conventional vs. Interferometric Imaging





- Optical HST (left) and 330MHz VLA (right) images of the Crab Nebula and the Orion nebula. Note the differences in the:
  - Range of spatial scales in each image
  - The range of intensities
  - The complexity of each image
  - The field of view as measured in resolution elements



- Mapping and model-fitting are complementary

Recap: Bayesian Mapping

- No unique solution - need to use prior knowledge
- Implicit prior knowledge: positivity, finite map size
- Explicit prior knowledge: regularization term e.g. entropy
- Several public software packages – most do a good job
- Image quality is generally limited by poor uv coverage

Concluding Remarks

- Phase information important for unambiguous interpretation of Fourier plane data
- Several techniques for obtaining phase information in phase-unstable conditions
  - Closure phase
  - Differential phase
  - Phase referencing (see talk on PRIMA)
- Closure phases/differential phases can be used in mapping and model-fitting

Spare slides start here

Closure Amplitudes

- Combination of *amplitudes* that is unaffected by antenna-based errors

$$A_{1234} = \frac{|V_{12}^{\text{measured}}||V_{34}^{\text{measured}}|}{|V_{13}^{\text{measured}}||V_{24}^{\text{measured}}|}$$

$$= \frac{|G_1||G_2||V_{12}^{\text{true}}|\;|G_3||G_4||V_{34}^{\text{true}}|}{|G_1||G_3||V_{13}^{\text{true}}|\;|G_2||G_4||V_{24}^{\text{true}}|} = \frac{|V_{12}^{\text{true}}||V_{34}^{\text{true}}|}{|V_{13}^{\text{true}}||V_{24}^{\text{true}}|}$$

- Named by analogy with closure phase
- Need at least four elements
- Can use as an additional constraint in self-calibration, to correct the *amplitudes* of the telescope gains
  - Usually applied in later iterations only

Self-Calibration

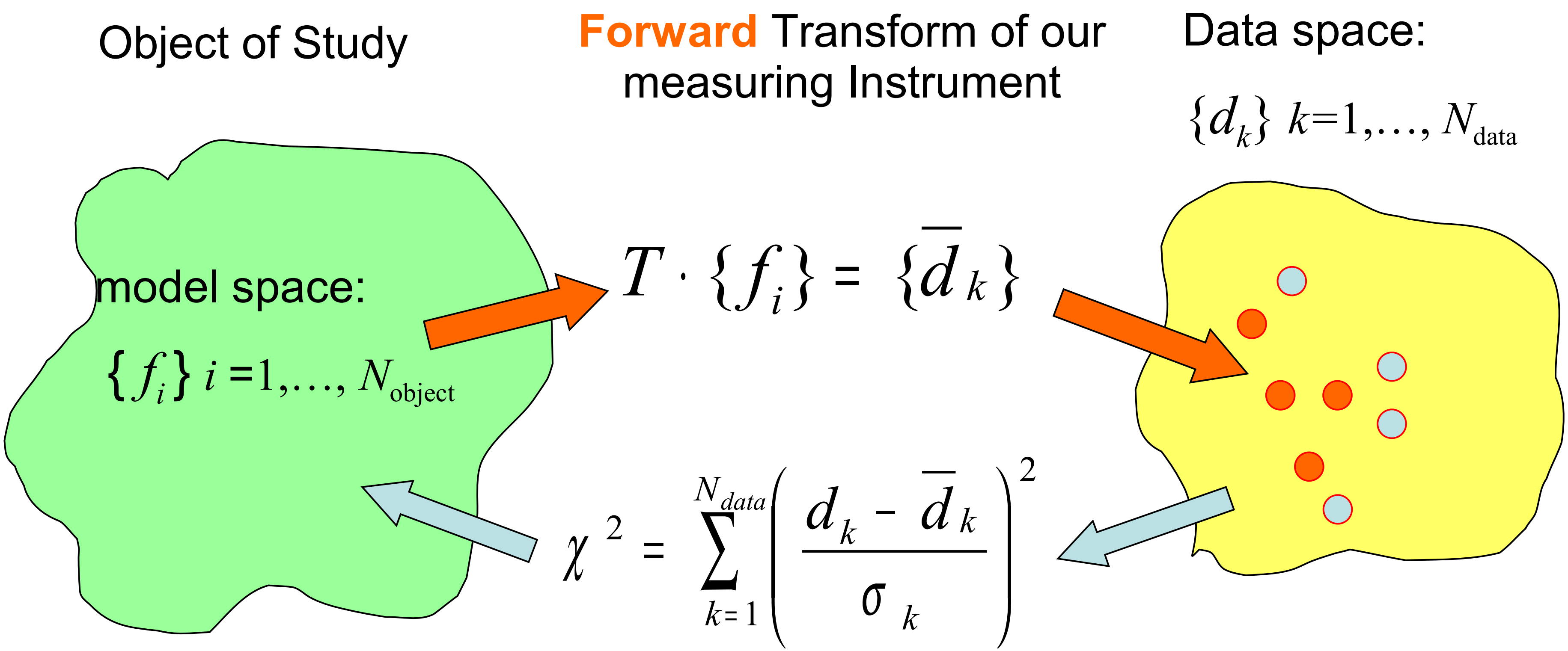
- Cornwell & Wilkinson (1981)
- Another specific example of above: closure phases used *implicitly*
- Explicitly solves for telescope phase errors ε_I: $\phi_{ij}^{\text{intrinsic}} = \phi_{ij}^{\text{measured}} - (\varepsilon_i - \varepsilon_j)$
- CALIB task in AIPS implements this

- Start with sky model: trial image
- Adjust telescope errors (corrections) so data is best fit by trial image
  - Take proper account of noise on measured amplitudes and phases
- Apply these corrections to data
- Invert corrected data and deconvolve
- If not converged, use resulting map as new trial image and start over

Features of Self-Calibration

- Uses a noise model in choosing optimum phases:
  - Allows for different SNRs on different baselines
  - Allows for different phase stability at different telescopes
  - Can specify timescale over which phase errors change
- This noise model is *wrong* for optical/infrared interferometry
- One telescope error is arbitrary
  - Assign zero error to a reference telescope
- CLEAN/MEM aids convergence in same way as other iterative schemes

MaxEnt data processing



Start from "flat" maximum entropy "object": $\{f_i\}$ all equal (α = ¥)

transform this to get "mock" data, and χ^2 degree-of-fit

Find new search directions (e.g. $\nabla\chi^2$, ∇S) and update $\{f_i\}$

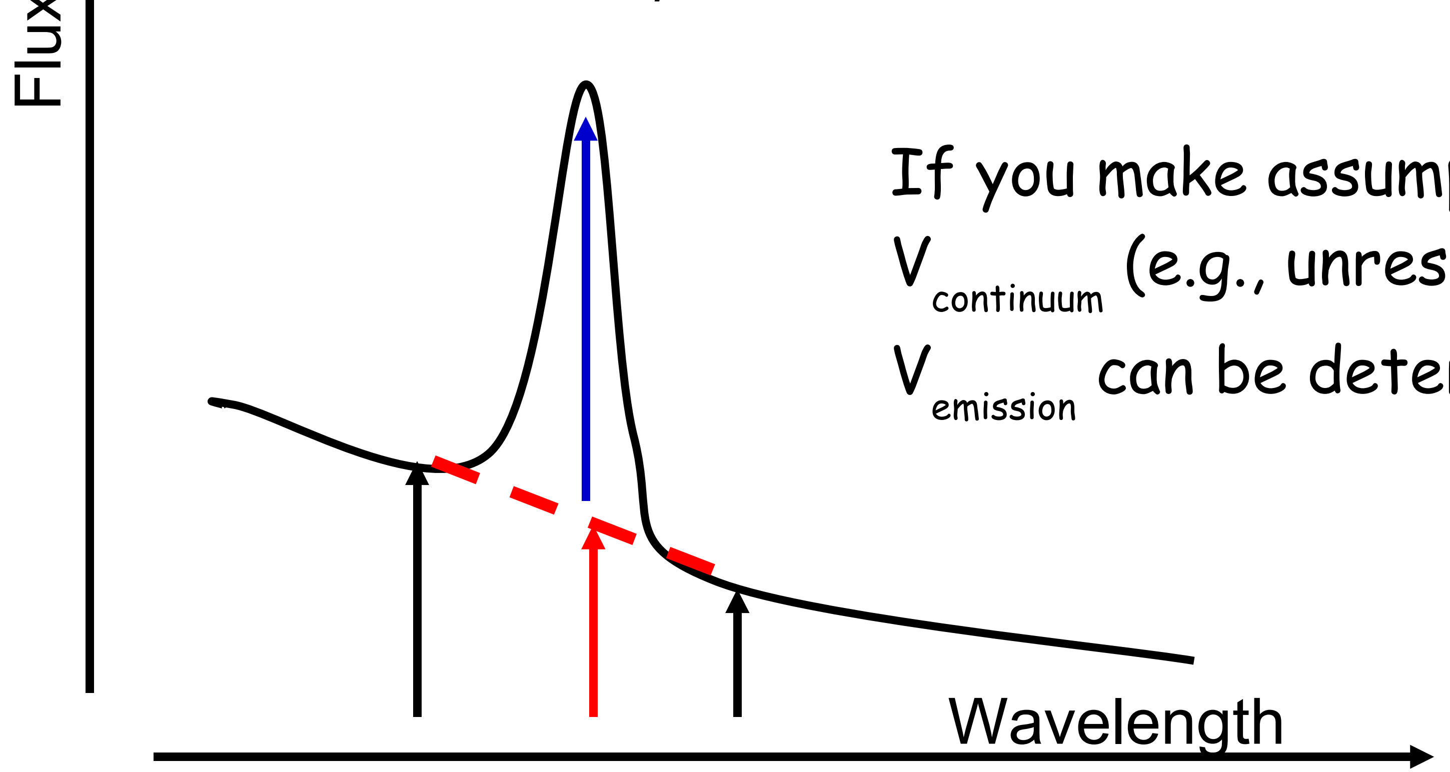
α relaxed from ¥to its most probable value

Differential Phase Example

$$\widetilde{V}'_{\text{line}} = \widetilde{V}'_{\text{continuum}} + \widetilde{V}'_{\text{emission only}}$$

Measured *Interpolated*

Prime indicates corrupted by atmospheric piston



If you make assumption about $V_{continuum}$ (e.g., unresolved), then $V_{emission}$ can be determined fully