Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **433,** 1718–1728 (2013)
Advance Access publication 2013 June 8

doi:10.1093/mnras/stt859

Phase errors in diffraction-limited imaging: contrast limits for sparse aperture masking

M. J. Ireland[1,2,3]★

[1]*Australian Astronomical Observatory, PO Box 296, Epping, NSW 1710, Australia*
[2]*Department of Physics and Astronomy, Macquarie University, NSW 2109, Australia*
[3]*Macquarie University Research Centre in Astronomy, Astrophysics & Astrophotonics, Macquarie University, NSW 2109, Australia*

Accepted 2013 May 12. Received 2013 May 12; in original form 2013 January 25

ABSTRACT

Bispectrum phase, closure phase and their generalization to kernel phase are all independent of pupil-plane phase errors to first order. This property, when used with sparse aperture masking behind adaptive optics, has been used recently in high-contrast observations at or inside the formal diffraction limit of large telescopes. Finding the limitations to these techniques requires an understanding of spatial and temporal third-order phase effects, as well as effects such as time-variable dispersion when coupled with the non-zero bandwidths in real observations. In this paper, formulae describing many of these errors are developed, so that a comparison can be made to fundamental noise processes of photon noise and background noise.

I show that the current generation of aperture-masking observations of young solar-type stars, taken carefully in excellent observing conditions, are consistent with being limited by temporal phase noise and photon noise. This has relevance for plans to combine pupil remapping with spatial filtering.

Finally, I describe calibration strategies for kernel phase, including the optimized calibrator weighting as used for LkCa15, and the restricted kernel phase POISE (phase observationally independent of systematic errors) technique that avoids explicit dependence on calibrators.

Key words: instrumentation: adaptive optics – instrumentation: high angular resolution – techniques: image processing – techniques: interferometric.

1 INTRODUCTION

The concepts of closure phase, bispectrum phase (e.g. Hofmann & Weigelt 1993), self-calibration and now kernel phase (Martinache 2010) are well known as techniques that cancel out many instrumental effects due to pupil-plane phase errors. Despite the very long history of aperture masking with a focus on fringe visibility amplitude (Fizeau 1868; Michelson 1891; Schwarzschild 1896), it was the use of closure phase that first enabled image reconstruction from this technique (Baldwin et al. 1986) as well as recent efforts in high-contrast imaging (e.g. Lloyd et al. 2006; Kraus & Ireland 2012).

A simple explanation of closure phase comes from a counting argument. From an interferometer with M (sub)-apertures, the complex visibilities can be independently measured on each of the $M(M-1)/2$ baselines consisting of each pair of (sub)-apertures. An optical aberration consisting of a piston on each of the (sub)-apertures amounts to $M-1$ degrees of freedom in the phase differences, leaving $(M-1)(M-2)/2$ additional measured quantities, which are the linearly independent set of closure phases. A set of observables which are independent of pupil-plane phase form an ideal starting point for precise model fitting and imaging at the diffraction limit. This argument applies to both redundant and non-redundant pupil geometries, as realized by Martinache (2010). But if phase errors on a pupil are large, a redundant pupil configuration is at a disadvantage, because the pairs of pupil locations that form any given Fourier component may add out of phase and destructively interfere. In the case of observations taken behind adaptive optics (AO), the choice of one technique over the other is not obvious.

In this paper, I will outline the causes of contrast limitations in the aperture-masking interferometry and kernel phase techniques, and methods to maximize contrast. In Section 2, the main causes of kernel phase errors will be outlined. In Section 3, I will describe why the statistical correlations between closure phases mean that kernel phases are preferred as a primary observable, and will compare the contrast limits achievable by different pupil geometries. In Section 4.1 I will describe standard closure phase calibration and its limitations, in Section 4.2 I will describe the calibration strategy as used in Kraus & Ireland (2012) to maximize contrast in aperture-masking interferometry observations, and in Section 4.3 I will describe the simpler phase observationally independent of systematic errors (POISE) calibration strategy. In Section 5, I will conclude and outline the key areas where further research is needed.

★E-mail: mireland@aao.gov.au

1.1 Kernel phase

The definition of kernel phase as used in this paper will be slightly simplified from the definition of Martinache (2010), as we will avoid the use of the 'redundancy' matrix **R**. To first order in pupil-plane phase (i.e. with a nearly flat wavefront), we can write the observed phase Φ_m in the Fourier transform of an image as

$$\Phi_m = \mathbf{A}\varphi + \Phi_o, \tag{1}$$

where φ is the pupil-plane phase and Φ_o is the phase of the Fourier transform of the object. These are represented as vectors where each vector element is one discrete point in the model pupil plane or the image discrete Fourier transform. The matrix **A** encodes the information about which parts of the pupil form each Fourier component. For example, a non-redundant baseline formed by two discrete pupil components only would have a $+1$ and -1 in that row of **A**, with all other elements taking the value 0. This matrix is described in detail in Martinache (2010). Using singular value decomposition, we then find a matrix **K**, the kernel of **A**, such that $\mathbf{KA} = 0$. By choosing **K** such that its number of non-zero rows is equal to its rank, this matrix enables us to project the Fourier phases on to a subspace, which we will call the *kernel phases* θ by $\theta = \mathbf{K}\Phi$. On this subspace, the observables are not affected by pupil-plane phase errors at first order:

$$\begin{aligned}\theta_m &= \mathbf{K}\Phi_m \\ &= (\mathbf{KA})\varphi + \mathbf{K}\Phi_o \\ &= \mathbf{K}\Phi_o. \end{aligned} \tag{2}$$

A model of the object can therefore be directly compared to the observed kernel phases by computing the Fourier transform and multiplying by the matrix **K**. For all reasonable two-dimensional pupils, the rank of **A** is at least half the length of Φ_o, meaning that at least half the object Fourier phase information is preserved when transforming from Fourier phase to kernel phase.

2 CAUSES OF KERNEL PHASE ERRORS

There are three broad classes of kernel phase errors: those that vary rapidly approximating white noise in a sequence of exposures (random errors), those that are static throughout an observing run and can therefore be calibrated by observation of unresolved calibrator stars (static errors) and those which vary from one target to another (calibration errors). Calibration errors include quasi-static errors with a time variability measured in minutes or hours, as well as errors that depend on e.g. the sky position or the spectrum of the source observed. The goal of any combination of observing technique and analysis strategy is both to minimize the random errors and to develop a calibration strategy where residual calibration errors are smaller than typical random errors. The following sections include error causes that could manifest themselves as one or several of these error classes.

2.1 General pupil-plane phase errors

We will examine first an abstract representation of pupil-plane phase errors that could cause random, calibration or static errors. We consider a closing triangle containing apertures A, B and C, as depicted in Fig. 1. Each aperture has the same size and shape, and each baseline $1 \equiv A \rightarrow B$, $2 \equiv B \rightarrow C$ and $3 \equiv C \rightarrow A$ has data taken at the same time. That is, there are equivalent coordinate systems describing apertures A, B and C, centred on each aperture. This means that

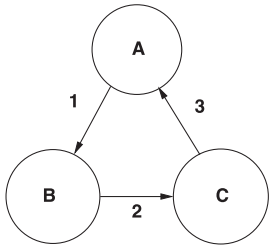

Figure 1. An abstract representation of closure phases formed by baselines 1, 2 and 3, in turn formed by congruent apertures A, B and C.

the visibility on each baseline is formed by the incoherent integral of visibilities arising from common spatio-temporal coordinates in sub-apertures A, B and C.

We will assign the symbols φ_A, φ_B and φ_C to the phase in sub-apertures A, B and C, the symbols Φ_1, Φ_2 and Φ_3 to the phase on baselines 1, 2 and 3, respectively, and will neglect amplitude variations (i.e. scintillation). The complex visibilities are then formed by

$$\begin{aligned} V_1 &= \overline{\exp \mathrm{i}(\varphi_B - \varphi_A)} \\ V_2 &= \overline{\exp \mathrm{i}(\varphi_C - \varphi_B)} \\ V_3 &= \overline{\exp \mathrm{i}(\varphi_A - \varphi_C)}, \end{aligned} \tag{3}$$

where the bar represents an average over the spatio-temporal coordinates corresponding to each aperture. This can be expanded to third order in phase to

$$V_1 \approx 1 + \mathrm{i}\overline{(\varphi_B - \varphi_A)} - \frac{1}{2}\overline{(\varphi_B - \varphi_A)^2} - \frac{\mathrm{i}}{6}\overline{(\varphi_B - \varphi_A)^3}, \tag{4}$$

with similar expressions for V_2 and V_3. The bispectrum is given by the product of these three visibilities, which can be again expanded to third order in phase:

$$b_{ABC} = V_1 V_2 V_3 \tag{5}$$

$$\Re(b_{ABC}) \approx 1 - \frac{1}{2}\left[\overline{(\varphi'_B - \varphi'_A)^2} + \overline{(\varphi'_C - \varphi'_B)^2} + \overline{(\varphi'_A - \varphi'_C)^2}\right] \tag{6}$$

$$\Im(b_{ABC}) \approx -\frac{1}{6}\left[\overline{(\varphi'_B - \varphi'_A)^3} + \overline{(\varphi'_C - \varphi'_B)^3} + \overline{(\varphi'_A - \varphi'_C)^3}\right], \tag{7}$$

where we have considerably simplified the expansion by introducing the piston-corrected phases:

$$\varphi'_A = \varphi_A - \overline{\varphi_A} \tag{8}$$

$$\varphi'_B = \varphi_B - \overline{\varphi_B} \tag{9}$$

$$\varphi'_B = \varphi_B - \overline{\varphi_B}. \tag{10}$$

A more complete derivation of this expansion is given in Appendix A. The closure phase $\phi_{\mathrm{cp}} = \Phi_1 + \Phi_2 + \Phi_3$ is then most simply approximated by taking the leading terms in the real (zeroth order) and imaginary (third order) components of the bispectrum, giving $\phi_{\mathrm{cp}} = \Im(b_{ABC})$.

It is also worthwhile briefly considering the effects of averaging the visibilities for baselines 1, 2 and 3 over different spaces. This could be caused by differing sub-aperture shapes in conventional aperture-masking interferometry (amounting to non-closing triangles), or by disjoint integration times as found in other forms of

interferometry. In this case, the leading terms in the closure phase errors become first order rather than third order in pupil-plane phase. Clearly, this is something to be avoided at considerable effort in the case of high-contrast aperture masking. The pupil 'shape' can also be thought of as the pupil-plane amplitude within each sub-aperture. Where amplitude errors are taken into account, these closure phase errors then become second order, i.e. first order in phase and first order in amplitude, and could plausibly be the leading term.

2.2 Temporal phase errors

Our first application of equation (7) to closure phase errors is rapid temporal effects, which cause a random kernel phase error. There are two key regimes that temporal errors operate in behind an AO system. Either exposure times are comparable to or shorter than the inverse of the AO system bandwidth (the short-exposure regime) or exposure times are significantly longer than these time-scales (the long-exposure regime). Given typical coherence times at $\sim$2.2 μm or shorter wavelengths of $<$50 ms, and typical AO system bandwidths in the range 10–100 Hz, exposure times longer than $\sim$100 ms in the near-infrared are in the long-exposure regime.

In the long-exposure regime, we can make the approximation that piston noise is white up to some cutoff frequency f_c. This is not very unrealistic, because in the frozen turbulence approximation, the atmosphere has an amplitude spectrum proportional to $f^{-5/6}$, while the error signal from a proportional–integral–differential controller in the mid-frequency range where the proportional term dominates gives residual errors proportional to the input signal amplitude multiplied by the frequency f. This gives a resultant error amplitude proportional to $f^{1/6}$, up to the servo loop cutoff. At this cutoff, three independent phenomena all tend to cut off the error spectrum rapidly: the $f^{-5/6}$ atmospheric amplitude spectrum, the rapidly lowering gain of the servo approaching its Nyquist sampling frequency and effects of spatial filtering.

We will now make a second set of approximations by assuming that the phase piston on each sub-aperture making a closing triangle is uncorrelated and has identical phase noise σ_φ. This may not be reasonable for some AO systems (e.g. if tip/tilt errors dominate due to tip/tilt mirror bandwidth) but as this depends on reconstructor and wavefront sensor details, it is a good first approximation.

An exposure of total time T can then be split into $f_c T$ sub-exposures, each of which has independent phase noise, so that in each exposure we have pupil-plane sub-aperture piston phases given by normal distributions:

$$\varphi_A \sim \mathcal{N}(0, \sigma_\varphi) \tag{11}$$

$$\varphi_B \sim \mathcal{N}(0, \sigma_\varphi) \tag{12}$$

$$\varphi_C \sim \mathcal{N}(0, \sigma_\varphi). \tag{13}$$

Applying equation (7) to this phase noise distribution for $f_c T \gg 1$ gives the standard deviation of closure phase (see Appendix B for a derivation):

$$\sigma(\phi_{\mathrm{cp,temporal}}) = \sigma_\varphi^3 \sqrt{3/f_c T}\,\mathrm{rad}. \tag{14}$$

In the short-exposure regime, we are dominated by atmospheric piston, as in the case with aperture-masking interferometry without AO (e.g. Tuthill et al. 2000). In this regime, for typical exposure times Δt less than $\sim$20 ms at a 2.2 μm wavelength, or $\sim$50 ms at 4 μm wavelength without AO or fringe tracking, we can still consider phase errors at third order with reasonable accuracy. By evaluating equation (7) numerically based on Kolmogorov turbulence, we arrive at

$$\sigma(\phi_{\mathrm{cp,temporal}}) = 0.0177 \left(\frac{\Delta t}{t_0}\right)^{15/6} \mathrm{rad}, \tag{15}$$

which is valid for $\Delta t \lesssim t_0$. This kind of relationship also has relevance to long-baseline interferometry in the case of measurements where visibilities are measured simultaneously. Examples of this are the Michigan InfraRed Combiner (Monnier et al. 2006) or the Precision Astronomical Visible Observations combiner (Ireland et al. 2008) at the Center for High Angular Resolution Astronomy array. This relationship does not apply to scanning beam combiners, where fringes can be recorded non-simultaneously depending on group delay tracking accuracy.

2.3 Spatial closure phase errors

In this section, we will examine how wavefront phase corrugations affect closure or kernel phases, occurring as random, calibration and static errors. Calibration errors occur when there are slowly time-variable spatial aberrations (often called *quasi-static* speckles). To most easily compare kernel phase to closure phase, we adopt a factor of $1/\sqrt{3}$ scaling to the closure phase, so that adding the three baseline phases is equivalent to multiplying by a unit vector (e.g. one of the orthonormal columns of the matrix $\mathbf{V}$ from Martinache 2010).

Fig. 2 shows a comparison between simulated sparse aperture masking and kernel phase data analysis for a variety of aberration spatial frequencies and aberration amplitudes. For each amplitude and spatial frequency, the position angle of a sinusoidal aberration was randomly varied and the overall rms kernel phase computed. It can be seen that although both kernel phase and closure phase appear equivalent to first order, they have quite different responses to high-order pupil-plane errors. The spatial filtering of an aperture mask means that it can be effectively used at much lower instantaneous Strehl ratios than unobstructed pupil kernel phase, but in a high-Strehl regime, kernel phase is in principle superior. For the 0.35 rad rms phase error case (right-hand figure), equation (7) predicts closure phases approximately two times lower than the simulation, possibly due to Fourier sampling and windowing effects in the sparse aperture-masking pipeline used, and possibly due to effects higher than third order in pupil-plane phase. For very high instantaneous Strehls, kernel phase in both geometries is expected to scale as the cube of the pupil-plane phase error, which is $(1 - S)^{3/2}$ in the Maréchal approximation.

A comparison between imaging with an unobstructed aperture and with sparse aperture masks is complicated somewhat by the ability to window data, which smoothes over high-spatial-frequency aberrations. This gives a further advantage in principle to an unobstructed aperture or a mask with large holes where the interferogram has a relatively small spatial extent. An example of a regime where fine spatial scale aberrations may dominate phase errors post-calibration is when aberrated pupil-plane elements or masks shift due to flexure effects.

2.4 Flat field errors

In sparse aperture masking, many pixels are used to record fringes from objects with intrinsically small spatial extents. If target and calibrator objects are not acquired on the same pixels, then the effect of flat-field errors is to add random phase errors across the Fourier plane. These random errors are only static if alignment is

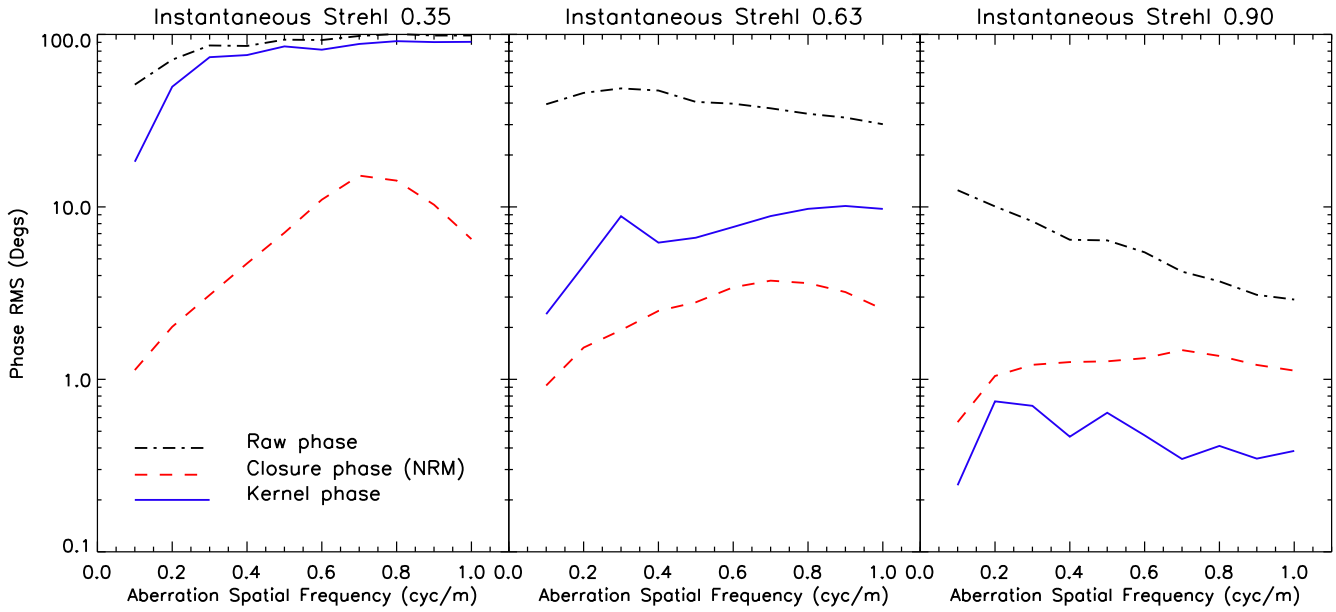

Figure 2. The effect of rms pupil-plane phase errors of 1 (left), 0.7 (centre) and 0.35 rad (right) on raw aperture-masking Fourier phase (black dot–dashed), full-pupil kernel phase (blue solid) and aperture-masking closure phase (red dashed) scaled by a factor of $1/\sqrt{3}$ as described in the text. The pupil geometries are the Keck non-redundant nine-hole mask and the full Keck pupil.

perfect between target and calibrator star observations – otherwise flat-field errors become a calibration error. A flat-field error can be modelled as multiplication in the image plane by a function that is 1.0 everywhere plus white noise with standard deviation σ_F. A typical value of σ_F is 10^{-3}, arising from a series of flat-field exposures with a total of 10^6 photoelectrons per pixel. Multiplication by this flat is equivalent to convolution in the Fourier domain, which spreads the power from the zero to near-zero spatial frequency components over the full Fourier plane. Clearly, phase errors will then be proportional to σ_F and inversely proportional to visibility. Numerical simulations give the following relationship for closure phase in sparse aperture-masking observations:

$$\sigma(\phi_{\mathrm{cp,photon}}) \lesssim 0.3\frac{\sigma_F}{V}\mathrm{rad}, \tag{16}$$

where V is the fringe visibility, referenced to a perfect Strehl interferogram of a point source. The constant of $\sim$0.3 varies between approximately 0.2 and 0.3 for different bandpass filters and aperture masks. To ensure that these errors are less than 10^{-3} rad with typical visibilities of 0.3, we need $\sigma_F < 10^{-3}$, meaning at least 10^6 photons per pixel recorded when taking flat-fields.

2.5 Bad pixels

The existence of bad pixels on an imaging array can often destroy sensitivity in traditional imaging over a small portion of the field of view. Like flat-field errors, incorrectly accounting for bad pixels can cause significant calibration errors. By spreading the information over many pixels, it may seem that at first glance bad pixels would always do significant harm to the information content in aperture-masking observations. However, the limited Fourier support of this kind of observation, as long as it is better than Nyquist sampled, means that bad pixels can be very effectively corrected. In simulations, the algorithm below has proved effective at contrasts beyond 10^6 for arrays far worse than those found at telescopes where aperture masks are installed, meaning that if properly corrected, bad pixels are not a cause of kernel phase errors.

The principle of this bad pixel correction algorithm is to assign the values to the bad pixels so that the power in the Fourier domain outside the region of support permitted by the pupil geometry is minimized. We will call this region of the Fourier plane the zero region Z. We can turn this problem into a linear one by realizing that the Fourier components corresponding to the set of bad pixel coordinates $\boldsymbol{x}_b$ form a subspace of Z, and we can find a vector of bad pixel offsets $\boldsymbol{b}$ to subtract so that the image Fourier transform on this subspace is identically zero.

The first step in this process is to create the matrix $\mathbf{B}_Z$ which maps the bad pixel values on to Z. The measured values f_Z in the Fourier-plane region Z are then modelled as

$$f_Z = \mathbf{B}_Z\boldsymbol{b} + \epsilon_Z, \tag{17}$$

with ϵ_Z being the remaining Fourier-plane noise. The bad pixel adjustments $\boldsymbol{b}$ are then found using the Moore–Penrose pseudo-inverse of $\mathbf{B}$:

$$\boldsymbol{b} = \mathbf{B}_Z^+ f_Z \tag{18}$$

$$= (\mathbf{B}_Z^*\mathbf{B}_Z)^{-1}\mathbf{B}_Z^* f_Z. \tag{19}$$

The Moore–Penrose pseudo-inverse can also be found by other methods such as singular value decomposition rather than direct computation of an inverse as in equation (19), but this method suffices for a relatively small number of bad pixels. Although this algorithm is very quick (the matrix $\mathbf{B}^+$ is pre-computed), the bad pixel correction (equation 18) has to be applied for every frame, with the computed values $\boldsymbol{b}$ subtracted off each frame. It can also be used to correct for saturated pixels at the core of a point spread function (PSF), pixels affected by transient events such as cosmic rays, or an acquisition error where a small portion of the interferogram is truncated by the detector edge.

2.6 Dispersion and wavelength-dependent phase errors

Kernel phase observations are often made in a broad-band filter where different wavelengths are affected by both the atmosphere

and optics in different ways. This causes a static kernel phase error, which can become a calibration error unless observing conditions and spectrum are matched between target and calibrator observations. A general analysis of these errors is particularly difficult and beyond the scope of this paper, because the definition of kernel phase is inherently monochromatic. However, we can put some limits on when this effect might become important and the order of magnitude of the effect. We write the air refractive index difference of Δn between the blue and red edges of a filter, and the spectral difference between a target and calibrator is ΔF covering a fraction f of the bandpass. Assume that both objects are observed at the same airmass. The image Fourier-plane phase error arising from this difference is

$$\Delta\varphi \approx 2\pi\Delta F f \Delta\alpha B_{\max}/\lambda_{\mathrm{mean}}, \tag{20}$$

where the change in angle on the sky between long- and short-wavelength part of the filter is

$$\Delta\alpha = \Delta n \tan(z). \tag{21}$$

Here z is the zenith distance angle, and this formula gives only the value for airmasses less than approximately 3. The kernel phase signature of this dispersion effect is very similar to that of a close companion of separation $\Delta\alpha$ and magnitude difference $f\Delta F$. For values of $\Delta\alpha$ greater than about $0.5\lambda_{\mathrm{mean}}/B_{\max}$, the kernel phase error $\Delta\theta$ is of the same magnitude as $\Delta\varphi$, and for smaller values of $\Delta\alpha$, the kernel phase error goes as $(\Delta\alpha B_{\max}/\lambda_{\mathrm{mean}})^3$ (e.g. see equation 5 of Le Bouquin & Absil 2012). As an example, observing in the full H band with a zenith angle of 45° from an altitude of 2600 m gives $\Delta\alpha = 31$ mas, which is larger than $0.5\lambda_{\mathrm{mean}}/B_{\max}$ for $B_{\max} = 8$ m. A 10 per cent difference in the spectrum over the long-wavelength 10 per cent of the H bandpass would then give $\Delta\theta \approx \Delta\varphi \approx 0.01$ rad.

The effect of observing at different airmasses is much more complex, because for flat spectra, dispersion does not give a non-zero kernel phase. In general, it may be a non-linear interaction between pupil-plane aberrations and dispersion that dominate the calibration errors.

2.7 Photon, background and readout noise

Finally, we consider the fundamental limitation of random errors caused by photon, background and readout noise. The closure phase error due to photon (shot) noise is

$$\sigma(\phi_{\mathrm{cp,photon}}) = \frac{N_{\mathrm{h}}}{N_{\mathrm{p}} V}\sqrt{1.5\left(N_{\mathrm{p}} + N_{\mathrm{b}} + n_{\mathrm{p}}\sigma_{\mathrm{ro}}^2\right)}, \tag{22}$$

where V is the fringe visibility, N_{p} is the total number of photons collected in an interferogram, N_{b} is the number of background photons and N_{h} is the number of holes in the aperture mask. The factor of $\sqrt{1.5}$ includes a factor of $\sqrt{3}$ due to photon noise from three independent baselines making up the closure phase, as well as a factor of $\sqrt{1/2}$ due to the shot noise power at any non-zero spatial frequency being split equally between the real and imaginary parts. The readout noise in photon units is σ_{ro} and the number of pixels is n_{p}. The effect of both readout and background noise is affected by the size of the window function used prior to making the Fourier transform to compute the visibilities, and this effect can be minimized if fringes are directly fitted to the data (e.g. the SAMP pipeline of Lacour et al. 2011).

2.8 Dominant error terms

The most common kinds of kernel phase data taken so far have been sparse aperture masking behind natural guide star AO, particularly at 1.5–2.4 μm wavelengths, so we will consider this regime first. We will also consider that adequate flat-fields have been taken and bad pixels properly corrected. The AO system only locks when there are at least ∼100 visible photons per Shack–Hartmann lenslet in ∼0.01 s, or $\sim 10^6$ photons in 100 s. With a similar near-infrared and visible photon rate, and a similar masking sub-aperture size to a Shack–Hartmann lenslet size, equation (22) would predict an ∼0.°4 photon-limited closure phase uncertainty for a 100 s integration and a nine-hole aperture mask.

We can use equation (14) to predict the effect of temporal phase errors: in particular good seeing, $\sigma(\varphi)$, could be as low as 0.3 rad (giving a temporal-phase-noise-limited Strehl of ∼0.9) and f_{c} could have a value of 10 Hz. This would give a temporal phase noise component to closure phase uncertainty of ∼0.°1. Perhaps not surprisingly, given how much light an aperture mask blocks, photon noise would dominate in this regime. However, for less than ideal seeing conditions and targets which are brighter in the infrared, the temporal phase noise dominates over photon noise. A characteristic 'typical seeing' predicted closure phase error for 0.5 rad rms pupil-plane phase error is 0.°5 for a 100 s integration.

The closure phase uncertainties predicted here are similar to the typical closure phase uncertainties computed from the standard error of the mean of individual observation sets in survey papers such as Kraus et al. (2008). However, it is certainly true that the residuals when subtracting closure phases from two point sources are not always statistically consistent with these standard errors. This kind of residual is often called a *calibration error*, where the non-zero closure phases described in Section 2.3 are not fully corrected by observations of a calibrator star. Typical uncalibrated closure phases from the Keck nine-hole aperture mask are 3.°5 in the H and K bands (CH4S and Kp filters), and 7° in the L band (Lp filter). These non-zero closure phases are consistent with having quasi-static spatial aberrations of ∼0.5 rad amplitude in the CH4S and Kp filters (e.g. Fig. 2) and atmospheric dispersion in the Lp filter (Section 4). A small change in the amplitude of these non-zero closure phases leads to miscalibrations that can be larger than the temporal (sub-aperture piston) phase and photon noise effects.

3 CLOSURE PHASE CORRELATIONS

One of the more confusing aspects of aperture-masking data analysis is knowing what to do with a linearly dependent set of closure phases. As described in Kulkarni (1989), these phases may be linearly independent in the case of very low signal-to-noise per exposure when the bispectrum is averaged, but in the high signal-to-noise limit considered here, with M non-redundant sub-apertures, there are $M(M-1)(M-2)/6$ closure phases but only $(M-1)(M-2)/2$ linearly independent closure phases. A redundant aperture has an even higher degree of correlation of the bispectrum phases.

Simply choosing an arbitrary independent set of closure phases for the purpose of modelling is not possible without a full consideration of the covariance matrix. If one considers only the simplest forms of closure phase errors, namely that due to readout noise, then the problem of modelling the covariance matrix is not difficult. However, there are many other kinds of errors that can cause correlations between closure phase errors.

Previous work has either gone to great lengths to diagonalize the measured covariance matrix of closure phase (e.g. Kraus et al.

2008) or has made an approximate scaling of fitting errors to account for the closure phase correlations (e.g. Hinkley et al. 2011). The difficulty in any approach based on real data is that the sample covariance matrix must be modelled and cannot in general be measured completely from the data. The reason for this is that where there are fewer data frames taken than independent closure phases, the sample covariance matrix is necessarily singular.

These difficulties are all avoided if rather than considering closure phases as a primary observable, the linear combinations that make the kernel phases are seen as the primary observables. This has added benefits of being able to extend the aperture-mask technique to considering baselines within each sub-aperture (consequently extending the usable field of view) and using the same language for all AO image analysis that is independent of pupil-plane phase to first order.

Of course, there are many different ways to form a set of kernel phases from a set of closure phases, or indeed a linearly independent set of kernel phases. Martinache (2010) suggested that kernel phases should be constructed so that only orthonormal linear combinations of Fourier phase are considered. However, this does not guarantee statistical independence. In the simplest case of a centrally concentrated image limited by photon noise, the spatial concentration of the image variance means that neighbouring Fourier components have highly correlated phase errors. This amounts to a contrast loss when considering n-sigma excursions of kernel phase, because just like aperture masking, the kernel phase technique, as described by Martinache (2010), has a nearly flat contrast limit curve beyond separations of $\sim\lambda/D$. However, standard imaging can have increasing contrasts as separations increase beyond the PSF centre. This apparent loss in sensitivity can be regained by properly considering the correlation between Fourier phases, as shown below.

3.1 Statistically independent kernel phase

Following from Section 1.1 we will define the matrix that transforms the Fourier phase vector Φ to the vector of kernel phases $\mathbf{K}_o$. This is an N_K by N_F matrix, where N_K is the number of kernel phases and N_F is the number of Fourier phases. The subscript o indicates that this matrix produces an orthonormal set of phase linear combinations. We can compute the sample covariance matrix of kernel phases $\mathbf{C}_K$ either directly or from the sample covariance matrix of Fourier phases $\mathbf{C}$. This matrix can be diagonalized by the finite-dimensional spectral theorem

$$\mathbf{S}^{\mathrm{T}}\mathbf{D}\mathbf{S} = \mathbf{C}_K = \mathbf{K}_o\mathbf{C}\mathbf{K}_o^{\mathrm{T}}. \tag{23}$$

The matrix $\mathbf{S}$ is then a unitary matrix which allows us to construct a set of statistically independent kernel phases based on a new kernel phase operator $\mathbf{K}_S$:

$$\theta_S = \mathbf{K}_S\Phi = \mathbf{S}\mathbf{K}_o\Phi. \tag{24}$$

As an example of the utility of this approach, I have simulated the effects of photon noise on kernel phase contrast limits, as shown in Fig. 3. The contrast standard deviation was estimated by first estimating the standard deviation of each kernel phase (i.e. neglecting covariances), forming a vector $\boldsymbol{\sigma}(\theta)$, then computing the contrast error using standard formulae for weighted averages:

$$\theta_m = \mathbf{K}\Phi_m \tag{25}$$

$$\sigma_c^2 = 1\Big/\sum_k \frac{\theta_{m,k}^2}{\sigma_k^2(\theta)}. \tag{26}$$

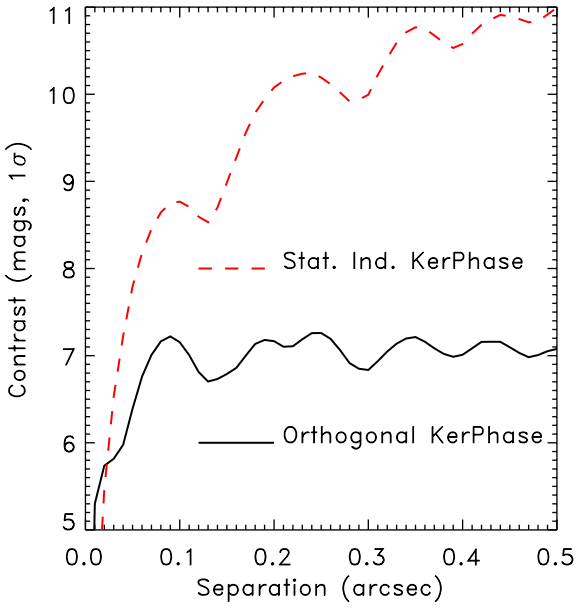

Figure 3. The effect of photon noise on kernel phase detections, based on a simulated photon-limited image with 10^6 photons taken with the unobstructed Keck telescope in the Lp filter. The decreased number of photons far from the PSF core means that kernel phases sensitive to these spatial locations have smaller errors, increasing the achievable contrast. Although the kernel phases in each situation are equivalent, the uncertainties are not equivalent and would require a full covariance matrix in the case of the orthogonal kernel phase.

Here Φ_m is the model phase divided by the contrast in the high-contrast limit, e.g. for a 100:1 brightness ratio companion, the phase would be approximated well by $0.01\Phi_m$. It is clear that the contrast achieved by considering statistically independent kernel phases defined by $\mathbf{K}_S$ is superior to the contrast achieved by orthonormal kernel phases defined by $\mathbf{K}_o$, for companions away from the PSF core.

4 CALIBRATION STRATEGIES

For the situation where phase errors are mostly random, calibration is not required. This has been the case for faint aperture-mask observations with a laser-guide star system, where obtaining calibration observations has a very significant observing time cost (e.g. Dupuy, Liu & Ireland 2009). When static phase errors dominate and random errors are larger than calibration errors, only a single suitable calibrator observation is required. A more typical situation in sparse aperture masking has been where random errors are small compared to calibration errors, and the choice and weight assigned to calibrator observations are critical in achieving the lowest possible model fit residuals and the highest contrasts. In this regime, there is an obvious danger – where calibrators are chosen to minimize the calibrated kernel phase, this biases the kernel phase away from a detection, and may result in deeper contrast limits being quoted for a non-detection than is justified by the data. This problem is also in common with the Locally Optimised Combination of Images (LOCI) algorithm (Lafrenière et al. 2007).

4.1 Nearest-neighbour calibration

The simplest calibration technique is to subtract the kernel phases from a calibrator observed closest to the target in time or space. A

Figure 4. An illustration of a situation where negative weighting of a calibrator may be optimal. Dispersion (illustrated by the lower red and upper blue circles) causes systematic kernel phases such that the kernel phases of Calibrator 2 (ϕ_{C2}) is the average of kernel phases of the Target (ϕ_T) and Calibrator 1 (ϕ_{C1}). The best estimate of the kernel phases caused by dispersion for the Target is then $2\phi_{C2} - \phi_{C1}$.

small extension to this technique (e.g. Evans et al. 2012) is to use the average of several calibrators observed nearby in time, rejecting outlier calibrator observations. Outliers are most easily rejected by looking for calibrators that when used to calibrate the target give spuriously large closure phases. For N_c calibrators, this amounts to calibrator weightings $\{a_k\}_{k=1}^{N_c}$, where each a_k is either 0 or $1/N_u$, with N_u the number of calibrators used. There are, however, several weaknesses to this technique.

(i) With small numbers of calibrator observations, it is difficult to avoid subjectivity in the choice to reject particular calibrators.
(ii) For particularly noisy calibrator observations and small systematic kernel phases, this process only adds noise.
(iii) All calibrators are weighted evenly, when the optimal weighting of individual calibrators may even be negative.
(iv) Any astrophysical structure in calibrators, e.g. undetected faint companions, contributes to any signal in final calibrated data.

The third point may not be obvious and is illustrated in Fig. 4. Whenever calibrators are all on one side of the calibrator in some space, then optimal calibration may extrapolate past the position of the calibrators to the target. This space may be real (such as zenith distance which produces non-zero kernel phases due to dispersion) or a one-dimensional parametrization of a hidden variable describing a time-variable aberration. This approach is similar to the potentially negative weighting of astrometric reference stars in precision astrometry (Lazorenko 2006).

4.2 Optimized calibrator weighting

We will now proceed to define a more optimal set of calibrator weightings $\{a_k\}_{k=1}^{N_c}$. This set of calibrator weightings must minimize the residual closure phases after fitting a model, without significantly biasing the model fit. In this section, we will describe this process as applied in Kraus & Ireland (2012), where the starting point is closure phases rather than kernel phases.

Following appendix A of Kraus et al. (2008), we begin by considering the closure phases only on a subspace spanned by the N_{ind} linearly independent set of closure phases. Furthermore, we construct a basis vector set on this subspace such that the closure phase covariance matrix is diagonal (or nearly so) when projected on to it. To see how this is done, first note how closure phases can be constructed linearly from phases:

$$\boldsymbol{\phi}_{cp} = \mathbf{T}\boldsymbol{\theta}_p. \tag{27}$$

The matrix $\mathbf{T}\mathbf{T}^t$ then projects any set of closure phases on to the set spanned by the linearly independent set of closure phases. This matrix can be diagonalized as $\mathbf{T}\mathbf{T}^t = \mathbf{U}_1^t\mathbf{D}_1\mathbf{U}_1$ by a diagonal matrix $\mathbf{D}_1$ and a unitary matrix $\mathbf{U}_1$. The eigenvalues on the diagonal of $\mathbf{D}_1$ are either 0 or 1. By considering only the non-zero eigenvectors of $\mathbf{D}_1$, we can write

$$\mathbf{T}\mathbf{T}^t = \mathbf{P}_1^t\mathbf{P}_1 \tag{28}$$

for an $N_{ind} \times N_{cp}$ projection matrix $\mathbf{P}_1$. $\mathbf{P}_1$ projects on to a subspace $\mathbb{S}$ spanned by an orthonormal set of linear combinations of closure phases.

Next, given a closure phase covariance matrix $\mathbf{C}_{cp}$, we can modify the projection matrix so that it projects on to a set of basis vectors for $\mathbb{S}$ with a diagonal covariance matrix. To accomplish this, we diagonalize the projection of $\mathbf{C}_{cp}$:

$$\mathbf{P}_1\mathbf{C}_{cp}\mathbf{P}_1^t = \mathbf{U}_2^t\mathbf{D}_2\mathbf{U}_2. \tag{29}$$

Then our new matrix $\mathbf{P}_2 = \mathbf{U}_2\mathbf{P}_1$ is a projection matrix on to $\mathbb{S}$ satisfying

$$\mathbf{P}_2\mathbf{C}_{cp}\mathbf{P}_2^t = \mathbf{D}_2. \tag{30}$$

Representing the data in this way enables, for example, the construction of χ^2 variables that can be computed by the sum over variance-normalized square deviates of a set of independent data, without the explicit use of covariance matrices. A potential problem with this approach is that the sample covariance matrix estimated from the data has a rank equal to $\min(N_{ind}, N_{fr} - 1)$, where N_{fr} is the number of data frames. Taken at face value, with $N_{fr} < N_{ind}$, this process unreasonably restricts the closure phases of a model of the target to lie on a very limited subspace in the space spanned by the observed departures from the mean closure phase. For this reason, we take $\mathbf{C}_{cp}$ above to be the weighted mean sample covariance matrix of all target and calibrator observations weighted by the inverse of the trace of each sample covariance matrix. We form the estimated errors of the target by

$$\mathbf{P}_2\mathbf{C}_t\mathbf{P}_2^t = \mathbf{D}_2'. \tag{31}$$

Our data and errors are then transformed to a set of kernel phases $\boldsymbol{x}$:

$$\boldsymbol{x} = \mathbf{P}_2\boldsymbol{\phi}_{cp} \tag{32}$$

$$\boldsymbol{\sigma}^2(\boldsymbol{x}) = \mathrm{diag}(\mathbf{D}_2') + \Delta^2. \tag{33}$$

The non-diagonal terms of $\mathbf{D}_2'$ are ignored, and any values on the diagonal less than the median are set to the median. This is a crude method to ensure that our statistics are reasonably robust, without resorting to studentizing a multidimensional distribution. An alternative to this approach might be a bootstrapping technique; however, in this case there is no obvious way to estimate the a_k variables below or to account for the error in their estimation. The additional uncertainty Δ^2 accounts for calibration errors, to be further defined below.

The next step is to find an optimal linear combination of weights $\{a_k\}_{k=1}^{N_c}$, where N_c is the number of possible calibrators. By *optimal*, we mean that we want to maximize the likelihood function for $\{a_k\}$ based on a null model for calibrated kernel phases $\boldsymbol{x}_c$:

$$\boldsymbol{x}_c = \boldsymbol{x}_t - \sum_{k=1}^{N_c} a_k\boldsymbol{x}_k \tag{34}$$

$$L(\{a_k\}) = \exp\left(-\sum_i \frac{x_{c,i}^2}{2\sigma_i^2(\boldsymbol{x}_t)}\right)\pi(\{a_k\}), \tag{35}$$

where we have explicitly subscripted $\boldsymbol{x}_c$ with i and where $\pi(\{a_k\})$ is a Bayesian prior distribution for $\{a_k\}$. The use of a restrictive prior as a regularizer is essential where there are many calibrators in use, because if $N_c > N_{ind}$ and there is a random error component, then there almost surely exists an $\{a_k\}$ such that $\boldsymbol{x}_c = 0$, subtracting any real astrophysical signal. The prior chosen in Kraus & Ireland (2012)[1] was

$$\pi(a_k) = \exp\left(-\frac{a_k^2}{2}\sum_i \frac{\sigma_i^2(\boldsymbol{x}_k)}{\sigma_i^2(\boldsymbol{x}_t)}\right), \tag{36}$$

where $\sigma_i^2(\boldsymbol{x})$ is the variance of the ith component of $\boldsymbol{x}$. This is certainly not the only choice of such a prior, but it does have the essential feature of preferring calibrator weights of zero and also of reducing the weighting of calibrators with large internal sample variances.

Once an optimal set of weights $\{a_k\}$ has been found by maximizing the likelihood function, the uncertainty on the calibrated kernel phases $\boldsymbol{x}_c$ is given by

$$\sigma_i^2(\boldsymbol{x}_c) = \sigma_i^2(\boldsymbol{x}_t) + \sum_k a_k^2\sigma_i^2(\boldsymbol{x}_k). \tag{37}$$

Note that this neglects any uncertainty in estimating the $\{a_k\}$.

Finally, the calibrator observations $\{\boldsymbol{x}_k\}$ do not necessarily span the space of the hidden parameters causing non-zero-point-source kernel phases. For this reason, the additional 'calibration error' term Δ^2 in equation (33) was iteratively added so that the reduced χ^2 for the null model was 1.0, i.e.

$$\chi_r^2 = \frac{1}{N_{ind}}\sum_i \frac{x_{c,i}^2}{\sigma_i^2(\boldsymbol{x}_c)} = 1.0. \tag{38}$$

In approximately half of the data sets tested in the work leading up to Kraus & Ireland (2012), no calibration error Δ^2 was needed. With values of the calibrated kernel phases $\boldsymbol{x}_c$ and their errors $\boldsymbol{\sigma}(\boldsymbol{x}_c)$ so computed, a model such as a bright star plus faint companion or a more complex image can be fitted using least squares. This is, however, a biased fit just like the LOCI technique (Lafrenière et al. 2007), because the process of computing the weights $\{a_k\}$ partly removes the binary signal, due to the null model for kernel phase in equation (35). For this reason, in Kraus & Ireland (2012), final values of model parameters were computed after re-computing the $\{a_k\}$ with the best-fitting model subtracted iteratively from the $\boldsymbol{x}_c$.

4.3 Restricted kernel phase (POISE)

An alternative to the complexity of the calibration strategy in the previous section is to ignore the kernel phases that require calibration, i.e. those kernel phases that are most affected by systematic errors. This is similar to choosing a prior in equation (35) in order that the calibrator is ignored for some kernel phases ($\pi(a_k) = \delta(0)$) and left uniform for other kernel phases, so that both calibration errors and astrophysical signal are subtracted. The difference between this and the technique described in this section is that only the restricted set of kernel phases where calibration is not required is used for subsequent analysis. We will call these restricted observables the POISE observables. This technique is very similar to the technique of ignoring dominant Karhunen–Loève eigenimages as a means of calibrating more wide-field PSFs (Soummer, Pueyo & Larkin 2012).

Following equation (28), we find a set of kernel phases $\boldsymbol{y}_k$ for each image k by a projection of the Fourier phases $\boldsymbol{\theta}_p$:

$$\boldsymbol{y}_k = \mathbf{S}_c\boldsymbol{\theta}_p \tag{39}$$

for the general kernel phase, remembering that

$$\boldsymbol{\theta}_p = \mathbf{P}_1\boldsymbol{\phi}_{cp} = \mathbf{P}_1\mathbf{T}\boldsymbol{\theta}_p \tag{40}$$

for aperture masking. The matrix $\mathbf{S}_c$ is formed in a similar way to equation (23), using the matrix $\mathbf{X} = \{x_k\}$ of calibrator observations, which is an (N_K by N_C) matrix, with N_C the total number of calibrator frames:

$$\mathbf{S}_c^{\mathrm{T}}\mathbf{D}\mathbf{S}_c = \mathbf{X}\mathbf{X}^{\mathrm{T}}. \tag{41}$$

This definition is almost the same as taking diagonalizing the covariance matrix, except that we do not subtract the mean kernel phases from the $\boldsymbol{x}_k$.

The calibrator kernel phases on this new subspace $\boldsymbol{y}_k$ with zero covariances are naturally subdivided into image sets C_j for each PSF calibrator observation j. Within each image set, uncertainties are dominated by random errors, but between image sets, there is a combination of random and calibration errors. We consider the sample variance for kernel phase i computed over all images k as *systematic* if

$$\delta_i^2 = s_i^2(\{\boldsymbol{y}_k \forall k\}) - s_i^2(\{\boldsymbol{y}_k : k \in C_j\}) > 0 \tag{42}$$

for all calibrator image sets j. In the POISE technique, we simply compute the systematic error components δ_i^2 for each kernel phase i, and

(i) ignore kernel phases y_i whenever

$$\delta_i^2 > \beta\langle s_i^2(\{y_k : k \in C_j\})\rangle_j. \tag{43}$$

A typical value for β is 1, which rejects approximately 1 to 3 out of 28 kernel phases for nine-hole Keck aperture-masking data.

(ii) Add δ_i^2 to each target observation's uncertainty estimate for the remaining kernel phases i.

This means that the process of calibration is completely independent of the target, which was not the case in Section 4.2, because in that technique calibrator weights were chosen to minimize the calibrated target kernel phases. The technique requires at least three calibrator image sets to differ significantly from simpler calibration techniques.

As an example of the use of this technique, we consider the data set used in the 2010 November K' sparse aperture-mask observations of the LkCa 15 system (Kraus & Ireland 2012). This data set consisted of 13 calibrator image sets of 12 images each, and 12 target image sets of 12 images each, all taken in good (0.6 arcsec) seeing. This is an ideal data set, especially given that all calibrators had previous sparse aperture-mask observations and were known to be single stars, and observations were continuous over a time period of 3.5 h, with target and calibrator observations interspersed. This is also the highest contrast detection published in the literature so far, which is the K-band detection of structure modelled as three compact sources around the star, with details reproduced in Table 1. Although much higher contrast is possible for brighter stars, especially when extreme AO may enable negligible piston phase errors,

[1] This equation, as presented in equation 1 of Kraus & Ireland (2012), was potentially confusing, because the division $\frac{(\cdot)}{(\cdot)}$ was element-by-element division, and the vector l^2-norm $|\cdot|$ was used without being explicitly described.

Table 1. A comparison between a fit with three additional point sources to 2010 November K' sparse aperture-mask data using linear combinations of calibrator observations (Kraus & Ireland 2012, KI12) and using the POISE observables. Parameters are separation (ρ), position angle (θ) and magnitude difference with respect to the primary (Δm). When adding uncertainties in quadrature, differences are always consistent within 2σ and in seven out of nine cases within 1σ.

Parameter	KI12	POISE
ρ_1 (mas)	67.0 ± 3.2	65.1 ± 3.1
θ_1 (deg)	12.3 ± 2.8	10.9 ± 2.9
Δm_1	7.40 ± 0.19	6.89 ± 0.18
ρ_2 (mas)	64.4 ± 1.5	62.6 ± 1.9
θ_2 (deg)	334.8 ± 1.5	333.4 ± 2.5
Δm_2	6.59 ± 0.09	6.36 ± 0.11
ρ_3 (mas)	82.5 ± 2.4	78.0 ± 4.1
θ_3 (deg)	302.3 ± 1.5	302.3 ± 2.8
Δm_3	7.06 ± 0.12	7.02 ± 0.18

at $V \sim 12$ this is roughly the brightest star of its class – no known $<$5 Myr solar-mass star is in any association closer than Taurus.

When applying the POISE algorithm to this data set with a β value of 1.0 in equation (43), only 1 of the 28 kernel phases are removed as 'systematic' by the calibrator observations, meaning that 96 per cent of the closure phase information is retained. A three-point-source fit to these restricted kernel phases had a reduced χ^2 value of 0.92, as shown in Table 1. With a reduction of β to 0.25, four kernel phases are removed as 'systematic', the reduced χ^2 becomes 1.00 but no fitted parameters change by even 1σ. In addition, the variance of the mean for 50 per cent of the image-set kernel phases is dominated by random errors, and not the δ_i^2 values from equation (42). This means that quasi-static spatial aberrations in this case do not significantly limit the signal-to-noise in the final image. For this kind of observation, spatially filtering the input wavefront (e.g. Huby et al. 2012; Jovanovic et al. 2012) could not significantly improve the achievable calibration-limited contrast. The random errors of $\sim$0.°5 in each 240 s image set are also consistent with temporal phase piston errors, which would not be improved by spatial filtering. This argument of course falls over for brighter targets (i.e. generally higher mass or closer and older targets) where existing AO systems perform much better, and extreme AO is possible. In these situations, σ_φ in equation (14) can be smaller than 0.3 rad, f_c can exceed 100 Hz and spatial filtering may become essential at the $\sim$10 mag contrast range enabled by this improved AO performance.

4.4 Imaging with POISE

For sufficiently complex sources, model fitting is replaced with imaging. In general, imaging from kernel phases alone is computationally intensive because of the non-linear relationship between the image plane and Fourier phase. However, in the high-contrast regime, where interferometric visibility amplitudes are unity within errors, we can approximate the Fourier transform $F(\boldsymbol{u})$ of an image $I(\boldsymbol{x})$ normalized to a total flux of unity as

$$F(\boldsymbol{u}) \approx 1 + \mathrm{i} \int \sin(2\pi \boldsymbol{u} \cdot \boldsymbol{x}) I(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}. \tag{44}$$

In turn, the phase Φ becomes

$$\Phi(\boldsymbol{u}) \approx \int \sin(2\pi \boldsymbol{u} \cdot \boldsymbol{x}) I(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}. \tag{45}$$

We can consider the image to be made of discrete pixel values arranged in a vector $\boldsymbol{p} = \{p_j\}$, so that the integral in equation (45) becomes a sum, and the values of Fourier phases $\boldsymbol{\phi}$ and kernel phases $\boldsymbol{\theta}$ are represented by matrix multiplication:

$$\boldsymbol{\Phi} \approx \boldsymbol{M}\boldsymbol{p} \tag{46}$$

$$\boldsymbol{\theta} \approx \mathbf{K}\boldsymbol{M}\boldsymbol{p}$$

$$\approx \mathbf{A}\boldsymbol{p}. \tag{47}$$

This linear approximation to imaging means that minimizing kernel phase χ^2 subject to a differentiable regularizer can be rapidly computed using a gradient descent method. An example of such a regularizer is the maximum entropy regularizer (e.g. Narayan & Nityananda 1986)

$$S = -\sum_j p_j \ln(p_j/q_j), \tag{48}$$

for some prior image $\boldsymbol{q}$, often taken to be a uniform image in some finite field of view and zero elsewhere. The problem of maximum entropy image construction is then simply a problem of minimizing the sum of the χ^2 value and the regularizer:

$$\boldsymbol{p}_{\mathrm{MaxEnt}} = \arg\min_p \left\{ \sum_i \frac{(\theta_i - \mathbf{A}\,\boldsymbol{p})^2}{\sigma_i^2} + \alpha \sum_j p_j \ln(p_j/q_j) \right\}. \tag{49}$$

The value of α is typically chosen so that the final image has a reduced χ^2 value of 1.0[2]. Then to see the result of this approach to imaging, we will again use the K' data set from Kraus & Ireland (2012). In that original paper, the optimized calibrator weighting scheme (see Section 4.2) enabled the MArkov Chain Imager (MACIM) algorithm (Ireland, Monnier & Thureau 2006) to be used to create images directly from the closure phases via an OIFITS input file. This approach ignored correlations between closure phases. The image created directly by fitting to kernel phases imaging with the maximum entropy regularizer can be seen in Fig. 5, where the resolved structures contain 1 per cent of the total system flux and the reduced χ^2 of the image is 1.0. Note that arbitrary point-symmetric flux could be added to this image and it would still fit the kernel phases. A weakness of imaging from kernel phases alone is that point-symmetric flux added to a bright central point source does not produce any phase information.

The image in Fig. 5 is cosmetically at least as good as that shown in Kraus & Ireland (2012), but comes with the significant benefit that the calibration process does not directly affect the image: the POISE observables are independent of the calibrator observations.

5 CONCLUSIONS

Aperture-mask interferometry has proven to be a powerful technique to recover high-contrast (up to $\sim$8 mag at 1σ), asymmetric information at the diffraction limit ($\sim$0.5–5λ/D) of large telescopes. The reason for this success is the ability for closure phase,

[2] Image reconstruction code in the PYTHON language using this regularizer can be found at http://code.google.com/p/pysco, the repository where all codes in this paper are intended to go after translation to PYTHON.

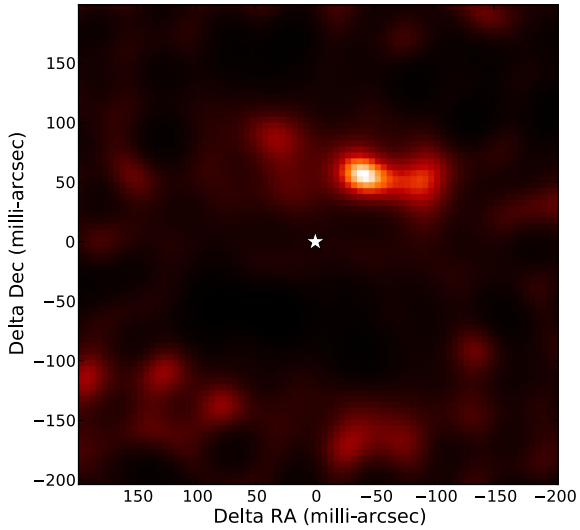

Figure 5. An imaging fit to the 2010 November data set of LkCa 15, originally published in Kraus & Ireland (2012). A uniform prior was used which had a total flux of 2 per cent of the image flux, and the final fit at a reduced χ^2 of 1.0 contained 1 per cent of the image flux, with the remaining 99 per cent contained within the point source star at the image centre.

a kind of kernel phase, to give an observable largely independent of time-variable aberrations. I have described many of the key sources of phase errors in this technique, as well as several strategies for mitigating them. Of note are the POISE observables, which are a subset of all possible linear combinations of closure phases. Observations of calibrator stars inform which linear combinations of phases constitute the POISE observables, but the analysis of the target observations is performed quite independently of the calibrator observations, leading to a more robust calibration method.

The generalization of the aperture-mask technique to full-pupil images shows great promise in the form of the full-pupil kernel phase observables. Simulations show that pupil-plane phase errors higher than third order affect full-pupil kernel phase more than aperture-mask kernel phase, meaning that full-pupil kernel phase will likely be restricted to moderately high Strehl observations.

The analysis presented here has implicitly involved only a monochromatic PSF from an imaging system. Although the effect of dispersion was discussed and the POISE calibration technique ameliorates the effects of dispersion, a mathematical framework to clearly predict the effects of dispersion on kernel phase was not developed. A future study of the effect of very broad bandwidths is needed. More importantly, an extension of this technique to work for the simultaneous wavelength-dispersed images formed by an integral field unit could be very powerful. The scaling of PSF with wavelength as a speckle suppression technique could be equally well applied to observables in the Fourier domain as it has been in image-plane analyses.

ACKNOWLEDGEMENTS

MI would like to acknowledge many helpful conversations with and encouragement from a large number of people over the past 10 years as these ideas developed and have been tested in various contexts, in particular Jean-Philippe Berger, Adam Kraus, Shri Kulkarni, Sylvestre Lacour, David Lafreniére, James Lloyd, Frantz Martinache, John Monnier, Laurent Puyeyo, J. Gordon Robertson, Anand Sivaramakrishnan and Peter Tuthill. The manuscript was also substantially improved following helpful comments from an anonymous referee.

REFERENCES

Baldwin J. E., Haniff C. A., Mackay C. D., Warner P. J., 1986, Nat, 320, 595
Dupuy T. J., Liu M. C., Ireland M. J., 2009, ApJ, 699, 168
Evans T. M. et al., 2012, ApJ, 744, 120
Fizeau H., 1868, C. R. Acad. Sci., 66, 932
Hinkley S., Carpenter J. M., Ireland M. J., Kraus A. L., 2011, ApJ, 730, L21
Hofmann K.-H., Weigelt G., 1993, A&A, 278, 328
Huby E. et al., 2012, A&A, 541, A55
Ireland M. J., Monnier J. D., Thureau N., 2006, in Monnier J. D., Schöller M., Danchi W. C., eds, Proc. SPIE Conf. Ser. Vol. 6268, Advances in Stellar Interferometry. SPIE, Bellingham, p. 58
Ireland M. J. et al., 2008, in Schöller M., Danchi W. C., Delplancke F., eds, Proc. SPIE Conf. Ser. Vol. 7013, SPIE, Optical and Infrared Interferometry. Bellingham, p. 63
Jovanovic N. et al., 2012, MNRAS, 427, 806
Kraus A. L., Ireland M. J., 2012, ApJ, 745, 5
Kraus A. L., Ireland M. J., Martinache F., Lloyd J. P., 2008, ApJ, 679, 762
Kulkarni S. R., 1989, AJ, 98, 1112
Lacour S., Tuthill P., Amico P., Ireland M., Ehrenreich D., Huelamo N., Lagrange A.-M., 2011, A&A, 532, A72
Lafrenière D., Marois C., Doyon R., Nadeau D., Artigau É., 2007, ApJ, 660, 770
Lazorenko P. F., 2006, A&A, 449, 1271
Le Bouquin J.-B., Absil O., 2012, A&A, 541, A89
Lloyd J. P., Martinache F., Ireland M. J., Monnier J. D., Pravdo S. H., Shaklan S. B., Tuthill P. G., 2006, ApJ, 650, L131
Martinache F., 2010, ApJ, 724, 464
Michelson A. A., 1891, Nat, 45, 160
Monnier J. D. et al., 2006, in Monnier J. D., Schöller M., Danchi W. C., eds, Proc. SPIE Conf. Ser. Vol. 6268, Advances in Stellar Interferometry. SPIE, Bellingham, p. 55
Narayan R., Nityananda R., 1986, AR&A, 24, 127
Schwarzschild K., 1896, Astron. Nachr., 139, 353
Soummer R., Pueyo L., Larkin J., 2012, ApJ, 755, L28
Tuthill P. G., Monnier J. D., Danchi W. C., Wishnow E. H., Haniff C. A., 2000, PASP, 112, 555

APPENDIX A: THIRD-ORDER BISPECTRUM EXPANSION

We will begin by writing the combination of equations (4) and (5) explicitly:

$$b_{ABC} = \left(1 + \mathrm{i}\overline{(\varphi_B - \varphi_A)} - \frac{1}{2}\overline{(\varphi_B - \varphi_A)^2} - \frac{\mathrm{i}}{6}\overline{(\varphi_B - \varphi_A)^3}\right) \times \left(1 + \mathrm{i}\overline{(\varphi_C - \varphi_B)} - \frac{1}{2}\overline{(\varphi_C - \varphi_B)^2} - \frac{\mathrm{i}}{6}\overline{(\varphi_C - \varphi_B)^3}\right) \times \left(1 + \mathrm{i}\overline{(\varphi_A - \varphi_C)} - \frac{1}{2}\overline{(\varphi_A - \varphi_C)^2} - \frac{\mathrm{i}}{6}\overline{(\varphi_A - \varphi_C)^3}\right). \quad \text{(A1)}$$

The zeroth-order terms in the $\Delta\phi$s are trivially collected as 1, and the first order terms clearly cancel to give 0. The second-order terms are

$$\Re(b_{ABC}) \approx -\frac{1}{2}\left[\overline{(\varphi_B - \varphi_A)^2} + \overline{(\varphi_C - \varphi_B)^2} + \overline{(\varphi_A - \varphi_C)^2}\right] - \left[\overline{(\varphi_B - \varphi_A)} \cdot \overline{(\varphi_C - \varphi_B)} + \overline{(\varphi_C - \varphi_B)} \cdot \overline{(\varphi_A - \varphi_C)}\right.$$

$$+\overline{(\varphi_A-\varphi_C)}\cdot\overline{(\varphi_B-\varphi_A)}\Big]. \tag{A2}$$

Moving from this equation to equation (6) requires the substitution of equations (8) through (10), as well as a recognition of the following classes of trivial identities:

$$\overline{(\varphi_B-\varphi_A)}=(\overline{\varphi_B}-\overline{\varphi_A}) \tag{A3}$$

$$\overline{(\varphi'_B-\varphi'_A)}=0. \tag{A4}$$

The third-order terms of equation (A1) are collected (after minor simplification of the coefficient 1/2 terms) as

$$\begin{aligned}\Im(b_{ABC})\approx&-\frac{1}{6}\Big[\overline{(\varphi_B-\varphi_A)^3}+\overline{(\varphi_C-\varphi_B)^3}+\overline{(\varphi_A-\varphi_C)^3}\Big]\\&+\frac{1}{2}\Big[\overline{(\varphi_B-\varphi_A)}\cdot\overline{(\varphi_B-\varphi_A)^2}\\&+\overline{(\varphi_C-\varphi_B)}\cdot\overline{(\varphi_C-\varphi_B)^2}+\overline{(\varphi_A-\varphi_C)}\cdot\overline{(\varphi_A-\varphi_C)^2}\Big]\\&-\overline{(\varphi_B-\varphi_A)}\cdot\overline{(\varphi_C-\varphi_B)}\cdot\overline{(\varphi_A-\varphi_C)}.\end{aligned} \tag{A5}$$

Again, equation (7) follows after substitution of equations (8) through (10) as well as applying trivial identities such as A3 and A4.

APPENDIX B: TEMPORAL PHASE ERRORS

In applying equation (7) to temporal phase errors, we write the instantaneous values of φ_A, φ_B and φ_C as random variables X_A, X_B and X_C, respectively, which take a new random value at N statistically independent time steps. We can then write

$$\mathrm{Var}(\phi_{\mathrm{cp}})=\frac{1}{36}\mathrm{Var}(\overline{(\varphi'_B-\varphi'_A)^3}+\overline{(\varphi'_C-\varphi'_B)^3}+\overline{(\varphi'_A-\varphi'_C)^3}) \tag{B1}$$

$$\approx\frac{1}{36\,N}\mathrm{Var}\Big((X_B-X_A)^3+(X_C-X_B)^3+(X_A-X_C)^3\Big) \tag{B2}$$

$$=\frac{1}{4\,N}\mathrm{Var}\Big(X_A^2X_B-X_AX_B^2+X_B^2X_C-X_BX_C^2+X_AX_C^2-X_A^2X_C\Big) \tag{B3}$$

$$=\frac{3\sigma_\varphi^6}{N}. \tag{B4}$$

Here Var represents the variance of a quantity, which in this special case of quantities of zero mean is simply the expectation of the square. The approximately equals sign ($\approx$) in equation (B2) is used because we are ignoring the piston subtraction, applicable only for $N\gg1$ (and with an error of the order of N^{-1}). All the variables X_A, X_B and X_C are independent Gaussian variables with mean 0 and standard deviation σ_φ, so their moments are standard results, and the expectation of a product of their moments is simply the product of the expectation of their respective moments. The variance on the right-hand side of equation (B3) can be thus be simply but tediously evaluated as the sum over 36 mutual covariances to give a value of $12\sigma_\varphi^6$. Finally, equation (14) follows directly from equation (B4), noting that the number of independent phase samples $N=f_cT$.

This paper has been typeset from a TEX/LATEX file prepared by the author.