

Diffraction Pupil Astrometry in the Toliman Space Telescope:

PSF as local 2D Ruler

Abstract

The concept of using an optical point spread function (PSF) as a precision two-dimensional ruler is explored. The point spread function of a telescope can be influenced by the design of a pupil plane mask. When we allow phase as well as amplitude masking the design possibilities become far reaching. We show how critical alignment properties of the PSF such as gradient energy (GE) and radial-weighted GE (RWGE) can be extended in range far beyond their usual diffraction levels. Other properties, such as peak intensity can be reduced dramatically and allow significantly increased utilisation of the image sensor dynamic range, sensor area and noise budget, not possible with amplitude-only masking.

Base System

All the following analyses assumes a base system as shown in the blue column below.

Nominal System Specs		
Two mirror Cassegrain Ritchey-Chrétien telescope (Hyperboloid primary, hyperboloid convex secondary; perfect on-axis), zero coma (aplanatic), non-zero field curvature and astigmatism.		
2kx2k CMOS image sensor, ~ 11 μ m sensor spacing (GSENSE)		
Pupil diameter 30cm		
Effective focal length	30m	15m
Focal-ratio	f-100	f-50
Airy pattern first zero ring radius ~ 0.45 λ	65.4 μ m	32.7 μ m
	5.9 pixels	3.0 pixels
Binary separation ~4 λ (Alpha Centauri A & B)	582 μ m	291 μ m
	52.9 pixels	26.5 pixels
Measurable change in separation 0.2 μ as = 9.70 pico-radians	291pico-metre	145pm
	3 micro-pixels	1.5 micro-pixels

The point spread function as an astrometric ruler

The key problem that overwhelmingly dominates astrometric error budgets is the stability of the image. When trying to reference stellar positions at micro-arcsecond scales, a host of small imperfections and minute mechanical drifts, warps and creep of optical surfaces create instabilities that can be orders of magnitude larger than the true signal. Rather than trying to directly contain all these errors, the Diffractive Pupil approach sidesteps them by creating a new ruler of patterned starlight against which to register positions in the image plane. The genius of this approach is that the diffractive grid of starlight suffers identical distortions and aberrations to the signal being measured. Drifts in the optical system therefore cause identical displacements of both the object and the ruler being used to measure it, and so the data are immune to a large class of errors that beset other precision astrometric experiments (paragraph text is from SPIE paper).

Diffractive PSF Ruler

We can broadly classify key PSF (intensity) properties in an ascending scale of complexity:

- Align-ability – namely total gradient energy
- Energy dispersion and dynamic range reduction – interrelated with the above
- Scale and spatial distortion sensitivity – radial weighted gradient energy
- Spectrographic calibration – needs large dispersion and high grating resolution

The lowest order property of the ruler is its absolute position measurement. The next level property is its relative position (or magnification) measurement. Both of these are 2D properties. Higher level properties corresponding to more general warps (non-linear distortion) can be defined.

Note that the diffractive pupil and the (coherent or amplitude) PSF are holographically duals of each other. The precision ruler formed by the PSF is dependent on the precision phase mask structure of the pupil.

Practical realisation of the diffractive PSF ruler concept is dependent on intensive computer modelling and image data processing to realise the ultimate micro-arc-second and micro-pixel precision theoretically allowed by photon noise processes.

The rest of this technical note is structured as follows:

Section A explores the alignability of star images and re-engineering PSFs

Section B explores the spatial self-calibrating possibilities of PSFs

A. Image Alignability and PSF Design

Fundamental limits: additive noise

In the past the task of astrometry was to estimate the position of a star based on a geometric definition of the star center or "centroid". The definition of centroid becomes problematic for star images with complex PSFs in noisy environments, and it becomes more natural to think in terms of maximum likelihood estimation. The task then becomes a registration problem between the image of a star and its idealization also known as the PSF.

In 2004 Robinson and Milanfar [1] established the Cramer-Rao lower bound (CRLB) for image registration under the assumption of additive white Gaussian noise (AWGN). Their measure of fit between two images (one of which maybe a real or an idealized image) is mean square error (MSE) which corresponds to Maximum Likelihood Estimation (MLE) with AWGN. The lower bound for the variance of the position error is σ_x , where:

$$\frac{1}{\sigma_x^2} = \frac{\sum_n \sum_m \left(\frac{\partial g(x_m, y_n)}{\partial x} \right)^2}{\sigma^2}$$

Here the intensity PSF is $g(x,y)$, the additive noise variance is σ^2 , and the summation is over the full spatial domain.

Clearly the greater the so-called gradient energy in the denominator, the smaller the CRLB. In most image registration tasks the gradient energy is predetermined by the image structure.

Robinson and Milanfar hit a limit of 2 milli-pixels. It turns out this is related to low order (quadratic) peak interpolation. In 2005 Pham [2] showed that the conventional (gradient-based shift) registration estimator (GBSE) exhibits a significant bias. The bias occurs even in noise-free systems but increases further with noisy images. The bias is essentially caused by low order interpolation errors and is thus related to the image Fourier spectrum and can be as large as about one quarter of the pixel separation – five orders of magnitude greater than our micro-pixel target. Pham shows how the bias can be reduced by an iterative algorithm. However, using accurate shift estimation the bias can be eliminated in one hit. The idea is based on perfect band-limited image shifting using Fourier interpolation. Note that the CRLB decreases as the image gradient Fourier spectrum increases; an opposite effect to bias.

Using Fourier shift and correlation it is possible to get down to the milli-pixel regime quite easily. However, moving three more orders of magnitude down to micro-pixels means that most previous assumption no longer hold. In particular:

- sensor element gain varies from element to element ("flat field"),
- sensor element gain centroid varies from element to element (pixel shift),

A number of systematic errors related to spatial sampling and discrete Fourier sampling become (potentially) significant in the micro-pixel domain. The most obvious effect is the interaction of

nearby (and not-so nearby) stars. Because the typical diffraction pattern from a point-like star is the Airy disc the intensity drop-off is the inverse of the radius cubed. As a rough guide, a star within 100 pixels of another may have a 10^{-6} influence on apparent intensity. The problem is entirely deterministic and can be resolved by accurate modelling.

Fundamental limits: photon noise

The matched filter position detection theory of Guyon [3] shows that the CRLB, in Poisson noise dominated imagery, is inversely proportional to the normalised gradient energy of the intensity PSF. Their equation can be written:

$$\frac{1}{\sigma_x^2} = \sum_n \sum_m \frac{\left(\frac{\partial g(x_m, y_n)}{\partial x} \right)^2}{g(x_m, y_n)}$$

We can interpret this equation loosely as the Robinson-Milanfar additive noise limit, except with the variance replaced by the signal itself. This result ties in with the recent analysis of Espinosa [4] which suggests that the CRLB can be reached in practice for Poisson noise dominated images if weighted least squares fitting replaces classical least squares. [Note that we should be aware of potential zero divisions in above formula!](#)

The above bound is rather different to the Robinson-Milanfar result for additive noise. In particular the simple gradient energy is re-weighted by the inverse of the PSF, which reduces the contribution of high image intensity regions and boosts the lower intensity regions.

For a telescope with a given light throughput (photons/sec) it is possible to increase the total gradient energy by broadening the PSF, dropping its peak value to just fill the dynamic range, but keeping the small fine structure size. This can be achieved by reducing the overall light flux of the telescope (i.e. partially occluding primary mirror) or introducing phase shifts in the pupil. The pure phase pupil approach is preferable because all photons are utilized. By suitable design of the pupil phase it is possible to significantly increase the total gradient energy in a captured image frame, as we shall see in the next section.

PSF Design

We have seen that astrometric precision is ultimately limited by the CRLB and in particular the (weighted) gradient energy of the star PSF. In a telescope we have a considerable degree of control over the system PSF by way of pupil masking (refs?).

It turns out that the pupil of a telescope can be phase modulated by sub-wavelength variations in the surface profile. Phase masking allows profound changes in PSF properties unlike amplitude masking where the central PSF peak is fixed. Phase allows total or partial cancellation in the final intensity PSF distribution. It is possible to change the local phase by π radians on a mirror by engineering a quarter wave step in surface profile (which induces a half wavelength step in OPL).

Diffractive Pupil Design Examples


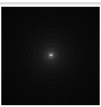

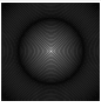
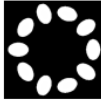
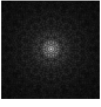



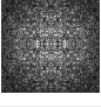




Pupil Design	Pupil	(Intensity) ^{0.25}	Relative Statistics	
P1 Full aperture			Normalised max	1.00
			Mean (#photon flux)	1.00
			Total signal energy	1.00
			Radial wtd radial gradient	0.68
			Total gradient energy	1.00
P2 Annular			Normalised max	1.00
			Mean (#photon flux)	5.60
			Total signal energy	0.94
			Radial wtd radial gradient	4.11
			Total gradient energy	1.33
P3 P. Tuthill 5-10 way symmetry			Normalised max	1.00
			Mean (#photon flux)	3.2
			Total signal energy	0.93
			Radial wtd radial gradient	1.71
			Total gradient energy	1.18
P4 Bipolar bullseye			Normalised max	1.00
			Mean (#photon flux)	9.30
			Total signal energy	2.92
			Radial wtd radial gradient	6.09
			Total gradient energy	2.82
P7 Semi-random symmetric bipolar			Normalised max	1.00
			Mean (#photon flux)	133
			Total signal energy	5.80
			Radial wtd radial gradient	76.78
			Total gradient energy	6.13
P8 Bipolar annular			Normalised max	1.00
			Mean (#photon flux)	100
			Total signal energy	5.8
			Radial wtd radial gradient	82.95
			Total gradient energy	6.32
P10 Bipolar version of P3			Normalised max	1.00
			Mean (#photon flux)	11.3
			Total signal energy	2.98
			Radial wtd radial gradient	7.00
			Total gradient energy	3.42

Fig 1, Initial table of possible diffractive pupil and corresponding PSF designs

Changes in height can be fabricated by masking during the reflector coating process, or by ion beam milling of the substrate itself.

The important PSF characteristics in the detailed design of the pupil pattern are

- reduction in peak height to just fill dynamic range
- spread of energy and gradient energy
- evenness of local peak heights
- resistance to spectral spreading

In a simple system the PSF is Fourier transform of the complex pupil function. In practical systems Fresnel propagation is required to compute the exact form of the PSF, but for our purposes here we use the Fourier approximation.

$$P(u, v) = \iint p(x, y) \exp(-2\pi i [ux + vy]) dx dy$$

$$p(x, y) = \iint P(u, v) \exp(+2\pi i [ux + vy]) du dv$$

Fourier duality of the pupil $P(u, v)$ and coherent PSF $p(x, y)$ is here represented by:

$$p(x, y) \longleftrightarrow P(u, v)$$

For brevity we have omitted the Fourier scaling factor with wavelength, focal length and system f-number. We are actually interested in the intensity PSF g as measured by the sensor. The intensity PSF or IPSF is then the transform of the pupil autocorrelation, where 2D correlation is written as cross inside a circle.

$$g(x, y) = |p(x, y)|^2 \longleftrightarrow P(u, v) \otimes P^*(u, v) = G(u, v)$$

Although we have considerable control of the coherent pupil, we have less control over the autocorrelation of this pupil.

The final characteristics have been tabulated below for various pupils. Each PSF has been normalised to have a peak intensity of 1, which means that the Toliman PSFs have a greater integrated intensity (\sim proportional to total number of photons per second) when each sensor image frame is assumed to have a filled dynamic range.

Peak intensity, $g_{peak}(\mathbf{r}) = \max \{g(\mathbf{r}), \mathbf{r} \in \mathbf{D}\}$

Total flux $\text{flux} = \sum_{\mathbf{D}} \sum g(x, y)$

Radial weighted gradient energy, $\text{RWGE} = \sum_{\mathbf{D}} \sum [\mathbf{r} \cdot \nabla g(x, y)]^2$

Mean intensity radius $r_{int}^2 = \sum_{\mathbf{D}} \sum [rg(x, y)]^2 / \sum_{\mathbf{D}} \sum [g(x, y)]^2$

Mean gradient radius $r_{grad}^2 = \sum_{\mathbf{D}} \sum [\mathbf{r} \cdot \nabla g(x, y)]^2 / \sum_{\mathbf{D}} \sum \|\nabla g(x, y)\|^2$

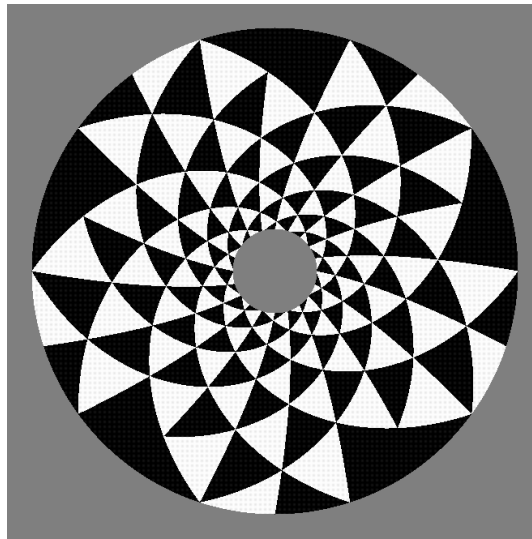


Fig 1, Early prototype Toliman telescope pupil design. White areas correspond to +1 transmission, black areas -1 transmission and the grey areas zero transmission (later design flipped some out zones)

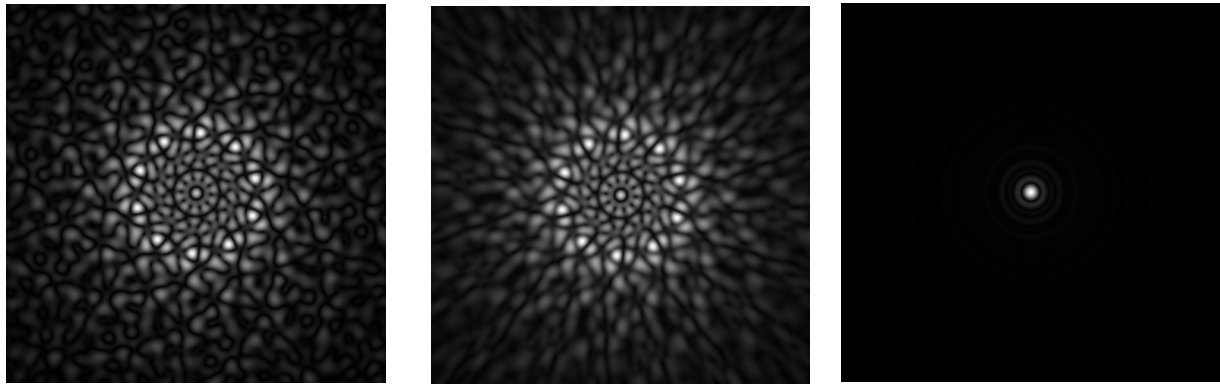


Fig 2 Toliman PSF 625 nm image on left, 600-650nm bandpass image middle, compared to annular PSF on right, all square root intensity grey scale.

Intensity Point Spread Function Characteristics (peak normalized)				
	Annular 625nm	Annular 600-650	Toliman 625nm	Toliman 600-650
Peak g_{peak}	1.0	1.0	1.0	1.0
Flux (photons)	1.0	1.001	71.77	80.18
Gradient Energy	1.0	1.0	5.295	5.036
Radial wtd GE	1.0	0.994	74.46	65.63
Intensity radius r_{int} (pixels)	1.215	1.213	26.97	25.98
Gradient radius r_{grad} (pixels)	1.781	1.772	25.04	23.21

Important points to to note are:

- The central peak reduction of the Toliman pupil allows about 70 times as many photons in an image frame without sensor saturation.
- A 50nm waveband causes only a small reduction in crucial **GE** and **RWGE** parameters.
- Second moment factors are very similar to the **GE**.
- For the annular pupil to achieve a 65 times improvement in magnification estimation (standard deviation) it would need a 65^2 increase in photons whereas the Toliman PSF only requires an 80 times increase in photons; a substantial 52 times advantage.
- The alignment improvement from a 79 times photon increase would be $\sqrt{79} \approx 9$ times, whereas the Toliman pupil achieves 5 times, albeit in a single frame exposure.

B. 2D PSF Ruler

Just as the alignment CRLB requires high gradient energy, so the 2D ruler needs de-localised gradient energy over a defined area.

Isotropic magnification correction

In the simplest case for using the PSF as a ruler we need to fit a model of the PSF to a single star image with unknown shift and unknown (but only slightly perturbed) magnification. If the gain is also unknown then a 4 parameter fit is required. Considering the simplified case where only the magnification is unknown allows us to see what parameters are important for the fit. Consider initially additive noise:

$$f(x, y) = g(x', y') + n(x, y) \quad \text{with} \quad x' = (1 + \varepsilon_0)x, \quad y' = (1 + \varepsilon_0)y$$

Then

$$s = \sum \sum n^2(x, y) = \sum \sum [f(x, y) - g(x + \varepsilon_0 x, y + \varepsilon_0 y)]^2$$

The LSE occurs at

$$\frac{\partial s}{\partial \varepsilon_0} = 0 = -2 \sum \sum [f(x, y) - g(x + \varepsilon_0 x, y + \varepsilon_0 y)] \frac{\partial g(x + \varepsilon_0 x, y + \varepsilon_0 y)}{\partial \varepsilon_0}$$

For small magnification errors a first order Taylor series suffices

$$g(x + \varepsilon_0 x, y + \varepsilon_0 y) \approx g(x, y) + \varepsilon_0 x \frac{\partial g(x, y)}{\partial x} + \varepsilon_0 y \frac{\partial g(x, y)}{\partial y} = g(x, y) + \varepsilon_0 \mathbf{r} \cdot \nabla g(x, y)$$

Hence

$$\sum \sum [f(x, y) - g(x, y)] \mathbf{r} \cdot \nabla g(x, y) = \varepsilon_0 \sum \sum [\mathbf{r} \cdot \nabla g(x, y)]^2$$

The result then being that magnification estimation depends upon a radial weighted gradient energy (RWGE)

$$\varepsilon_0 = \frac{\sum \sum [f(x, y) - g(x, y)] \mathbf{r} \cdot \nabla g(x, y)}{\sum \sum [\mathbf{r} \cdot \nabla g(x, y)]^2}$$

And we can expect this RWGE to replace the simpler GE in the CRLB for magnification estimation.

A suitable PSF ruler needs to optimise both GE and RWGE. The LRHF inspired Toliman pupil of Fig 1 has a greatly improved RWGE compared to a simple annular pupil with the same inner and outer diameters.

Anisotropic Magnification

We can repeat the above analysis with a more general anisotropic distortion of the PSF. A low order representation is given a 2D affine transformation with almost diagonal unity magnification:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} 1 + \varepsilon_a & \varepsilon_b \\ \varepsilon_c & 1 + \varepsilon_d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \mathbf{M}\mathbf{r} + \mathbf{r}_0$$

A similar affine equation applies in the Fourier domain, but in the Fourier domain it has the advantage that a positional shift \mathbf{r}_0 appears as a linear phase multiplier that can be detected separately and removed (see Bracewell [5]).

Anisotropic magnification is a distortion that is particularly important yet easy to analyse. In the simplest case it arises from a single tilt in the optical system – the sensor for example. In a direction parallel to the tilt axis the magnification is unchanged. However, perpendicular to this axis the classical cosine foreshortening occurs. Given that the nominal binary separation is 26 pixels on the sensor and we wish to resolve variations down to about 2 micro-pixels, then the tilt variation must be less than 8 microns over the width of the 20mm sensor, or 120 microns over primary mirror width of 300m. Long term mechanical stability of this order is difficult to maintain but we can use the PSF ruler to monitor slow (~hours) variations in the actual magnification.

General Warping/Distortion

Heuristically, for higher order warps we need to estimate magnification over many smaller domains, so it would seem that we need uniformly distributed PSF features and hence approximately uniform **GE** with a granularity on the scale of the warp variation. A uniform array of point-like intensities would be one such solution. The detailed maths has not yet been derived, but a **GE** metric with total **GE** and min-max **GE** over a defined region and granularity could be defined.

Sensor variation

Ignored in this technical note.

References

1. D. Robinson and P. Milanfar, "Fundamental performance limits in image registration," *Image Process. IEEE Trans. On* **13**, 1185–1199 (2004).
2. T. Q. Pham, M. Bezuijen, L. J. Van Vliet, K. Schutte, and C. L. L. Hendriks, "Performance of optimal registration estimators," in *Defense and Security* (International Society for Optics and Photonics, 2005), pp. 133–144.
3. O. Guyon, E. A. Bendek, J. A. Eisner, R. Angel, N. J. Woolf, T. D. Milster, S. M. Ammons, M. Shao, S. Shaklan, M. Levine, and others, "High-precision astrometry with a diffractive pupil telescope," *Astrophys. J. Suppl. Ser.* **200**, 11 (2012).
4. S. Espinosa, J. F. Silva, R. A. Mendez, R. Lobos, and M. Orchard, "Optimality of the Maximum Likelihood estimator in Astrometry," *ArXiv Prepr. ArXiv180503673* (2018).
5. R. N. Bracewell, *Two-Dimensional Imaging*, Signal Processing (Prentice Hall, 1995).