

National Research Council Canada Conseil national de recherches Canada



Wave-front Sensing & Reconstruction

CfAO Summer School, UC Santa Cruz Tuesday, August 17th 2021

> Garima Singh NRC Postdoctoral fellow

Herzberg Astronomy & Astrophysics Research Center

Victoria, BC, Canada

garima.singh@nrc-cnrc.gc.ca / guiding.honu@gmail.com

Table of content

- Brief introduction to the elements of wave front sensors (WFS)
- Classical WFS widely used in AO
 - Shack-Hartmann
 - Curvature
 - Pyramid
- Wavefront reconstruction
- Why we need Extreme-AO and what are its elements?

Acknowledgment

- Most of the introductory notes on Shack-Hartmann is gathered from two books: "Adaptive optics in Astronomy" by Francois Roddier and "Adaptive Optics for Astronomical Telescopes" by John W. Hardy
- Some slides are reused from talks given by the previous lecturers of the AO summer school. Credits to Don Gavel, Marcos Van Dam, Richard Clare, Lisa Poyneer and Olivier Guyon.

Reminder

- Wavefront (WF) aberrations are random.
- Waves are described as a complex number $\Psi = Ae^{i\phi}$, A and ϕ are real numbers representing the amplitude and phase of the fluctuating field.
- WF reference sources emit incoherent radiation, thus absolute measurements of optical phase is not possible.
- Phase information is translated into intensity signals.

A WFS is composed of three main components:

Optical device

Transforms WF aberrations into intensity variations.

Light detector

Transforms intensity into electrical signal.

Wave-front reconstruction

Convert signals into phase aberrations / WF sensor measurements into DM commands.

Requirements of a Wavefront Sensor (WFS):

- Large linear range WFS response should be a linear function of the input phase. Helps in faster reconstruction which means less computing time.
- High dynamic range Should be able to measure large WF errors.
 Allows AO loop to close in bad seeing.
- High sensitivity Should make efficient use of photons.
 Allows higher performance on faint sources. Requires detectors with high
 quantum efficiency and low noise.
- Spatial resolution Number of modes measured should ideally be as many as can be corrected by DM.
- Ability to work on extended sources.
- Ability to operate over a wide range of wavelengths.

It is difficult to get all properties simultaneously!

Wavefront Sensors



Wavefront Sensors

- A random WF error over a 2D aperture can be specified as Zonal or Modal.
 - **Zonal:** Aperture is divided into an array of independent subapertures or zones. WF is expressed in terms of the Optical Path Difference (OPD) over these spatial zones.



WF gradient/curvature is measured within the array of zones covering the telescope aperture.

• **Modal:** When the WF is expressed in terms of coefficients of the modes of a polynomial expansion over the entire pupil, for example, Zernike polynomials for a circular aperture.



WF is decomposed into specific surface shapes for example overall tilt, astigmatism and defocus etc

Sensing local gradients/slopes



- Divide aperture into zones/subapertures.
- Image the reference source within each such zone.
- Average WF slope is simultaneously measured over each zone. Note that the relative phase or piston component of each subaperture is lost.
- Technically, the WF slopes in each zone can be corrected individually, however, would result in incoherent superposition of the images formed by each zone. Thus, loss in angular resolution.

Individual slopes are fitted together and reconstructed into a continuous surface.

Shack Hartmann WFS and shearing interferometer are based on this principle.

Shack-Hartmann WFS

Shack-Hartmann: Measures the spatial first derivative i.e. the gradient of the wavefront (a Zonal WFS).



- A lenslet array is placed in a conjugate pupil plane to sample the incoming WF.
- Each lenslet create a map of local WF slope.
- The displacement of the spot is proportional to the WF slope.
- Need to calibrate the focus positions of the lenslet array.

Figure reference: "Adaptive optics in Astronomy" by Francois Roddier

Shack-Hartmann WFS

Oblique Astigmatism





SHWFS on 3.6 m telescope, uses an InGaAs focal plane array, operates at $1.6\mu m$



9 Slide credit: Michael Hart Spot size depends on the subaperture dimension d, the angular diameter of the source θ , the turbulence strength r_0 , and the sensor wavelength λ

- $d \approx$ the actuator pitch (often exactly the actuator pitch– Fried geometry, r_0)
- r_0 >d, angular size of the spot is $\sim \lambda/d$.
- $r_0 < d$, image of an unresolved source is determined by $r_0 (\sim \lambda / r_0)$
- Smaller aperture means more accurate WF measurements. If subapertures are too small, spot size
 increases due to diffraction, which degrades spot centroid estimate. Subaperture must be large enough
 to resolve the isoplanatic patch.
- Bigger subapertures means more light and better SNR in centroid measurement. However, poorer fit to WF.
- WF variations < subaperture size can not be measured. Slope sensor acts as a low-pass filter!

How to measure the position of spots?

A CCD to record all the images simultaneously



4 x 4 pixels or more

Estimate center of gravity position (c_x, c_y)

$$c_x = \frac{\sum_{i,j} x_{i,j} I_{i,j}}{\sum_{i,j} I_{i,j}} \text{ and } c_y = \frac{\sum_{i,j} y_{i,j} I_{i,j}}{\sum_{i,j} I_{i,j}}$$

 $I_{i,j}$ and $(x_{i,j}, y_{i,j})$ are the signals and the position coordinates of the CCD pixel (i,j).

A four quadrant detector (quad-cell) for each subaperture



Measured angle of arrival could be estimated as

$$\alpha_x = \frac{\theta_b}{2} \frac{I_1 + I_2 - I_3 - I_4}{I_1 + I_2 + I_3 + I_4}$$

Spot size (angular size on sky)

Sources of noise

Bias errors due to misalignment of the optics.

Unavoidable photon noise and read out noise. Measurement noise which depends on the size of the spot as seen in a sub-aperture ϑ_b , wavelength λ , sub-aperture size d, and signal-to-noise ratio SNR.

$$\sigma_{S-H} = \frac{\pi^2}{2\sqrt{2}} \frac{1}{SNR} \left[\left(\frac{3}{2} \right)^2 + \left(\frac{\vartheta_b d}{\lambda} \right)^2 \right]^{1/2} \text{ rad for } r_0 > d$$

$$\sigma_{S-H} = \frac{\pi^2}{2\sqrt{2}} \frac{1}{SNR} \left[\left(\frac{3d}{2r_0} \right)^2 + \left(\frac{\vartheta_b d}{\lambda} \right)^2 \right]^{1/2} \text{ rad for } r_0 < d$$
Hardy equation 5.16

SNR in a sub-aperture for fast CCD cameras is **dominated by read noise**. *S* is flux of detected photoelectrons per sub-aperture, n_{pix} is the number of detector pixels per sub-aperture and *R* is the read noise in electrons per pixel

$$SNR \approx \frac{St_{\text{int}}}{\left(n_{pix}R^2/t_{\text{int}}\right)^{1/2}} = \frac{S\sqrt{t_{\text{int}}}}{\sqrt{n_{pix}R^2}}$$

See McLean "Electronic Imaging in Astronomy", Wiley

How to measure the position of spots?

A CCD to record all the images simultaneously.

- Better linearity with small pixels size
- Dynamic range may be increased by using pixels of large size. There is no crossover between the subapertures, however, it lets in more sky background.
- More pixels on a subaperture means noisier estimation of the centroid due to read noise/dark current (SNR decreases). Noise can be reduced by windowing the centroid. Windowing is used to eliminate the background around the central core of the

image.



A four quadrant detector (quadcell) for each subaperture

- To convert position measurement into an angle the image size must be known!
 If images are diffraction-limited, θ_b = λ/d
 If seeing-limited, θ_b ~ λ/r₀, spot size depends on seeing conditions and is unknown.
- Quad-cell response can not be pre-calibrated for seeing-limited and extended sources!
- Estimated centroid position is linear with displacement over a small region (limited dynamic range). Could defocus the source image to a known spot size but at the cost of SNR
- Faster to read and compute centroid, less sensitive to detector noise

Advantages of a SHWFS

- Achromatic, because the OPD in turbulence is achromatic, hence works with broadband white light.
- Can operate with extended sources if the FOV is adapted to the source size.
- Linear with large dynamic range (when using a CCD).



 Aliasing errors: High frequency errors incorrectly measured as low-frequency errors! A spatial filter (a field stop of width λ/d in focal plane) can be used to overcome it.

It's a widely used sensor in Astronomy (Hale Telescope, Gemini South, Very Large Telescope), Ophthalmic applications, improving retinal imaging and mapping the aberrations of the eye (Dreher et al. 1989).

A challenge for the SHWFS

Laser guide elongation Shack-Hartmann sub-apertures see a line not a spot! Na layer 30 20 h Z -20 -30 θ -30 -20 -10 0 10 20 30 arcsec laser Laser projected subaperture from right ground $\leftarrow s \rightarrow$



A challenge for the SHWFS

Laser guide elongation Shack-Hartmann sub-apertures see a line not a spot!





 Radial format CCD.

•

Arrange pixels to be at same angle as spots

Measuring local curvatures

Curvature WFS (CS): Measures the second derivative of the phase i.e. its Laplacian (a Modal WFS!)



- Records the irradiance distribution at distance ±1 from the focus
- A local WF curvature in the pupil produces an access of illumination in one plane and a lack of illumination in the other!

Figure reference: "Adaptive optics in Astronomy" by Francois Roddier



Normalized difference between the irradiance distributions measured in defocused pupil planes P₁ and P₂

WF radial first derivative at the edge of the beam (radial tilts)

I) [∂φ (

 $\lambda f(f)$

measurement of the local WF curvature inside the beam

(5.10)

Choice of *l* is very important in curvature sensing!

The blur produced at defocused pupil planes ($(f - l)\theta_b$) should be < size of the WF fluctuations to be measured (to avoid smearing of the intensity variations),

The size of the WF fluctuations is the subaperture size!

Equation reference: "Adaptive optics in Astronomy" by Francois Roddier

Measuring local curvatures

- Sensitivity of the sensor is inversely proportional to the defocusing and dynamic range!
- Increasing the distance increases spatial resolution on the WF measurement but decreases sensitivity.
- A smaller distance yields a higher sensitivity to low-order aberrations. Smaller distance reduces both the aliasing error and the measurement noise.
- CS works well with incoherent white light and with faint targets.
- Curvature is a scalar field and requires one value per point! Thus intensity can be measured with a photon counting avalanche photodiodes without readout noise. Its cost effective!
- The photon error in a single subaperture of a CS is similar to SHWFS. However the error propagation in the reconstruction process is > SHWFS.
- Subaru Telescope AO188 system has a CS + Bimorph DM.



An irregular mirror tested with knife-edge test

Knife edge test for perfect lens (top), and one with spherical aberration (bottom). At right are observer views of pupil in each case.

- Simultaneous implementation of four Foucault knife-edge measurements.
- PyWFS splits focal plane in four quadrants, which are imaged by relay optics onto the pupil plane, producing four images of the pupil.



If the system is unaberrated (effects of diffraction are ignored), then the 4 pupil images should be identical.



Highly sensitive to small WF aberrations, limited dynamic range

Increased dynamic range at the cost of sensitivity.

Modulation is a trade-off between dynamic range and sensitivity

Fig. 1 (a) Illustration of a PyWFS. The prism angle is exaggerated here for illustrative purposes, but typically have values ~1 deg. (b) Focal plane (left) and pupil plane (right) images for a non-modulated PyWFS. (c) Focal plane (left) and pupil plane (right) images for a $4\frac{\lambda}{D}$ modulated PyWFS.



The slope-like signal for each subaperture is given by

 $S_x(x,y) = [(I_1(x,y) + I_2(x,y)) - (I_3(x,y) + I_4(x,y))]/I_0,$ $S_y(x,y) = [(I_1(x,y) + I_4(x,y)) - (I_2(x,y) + I_3(x,y))]/I_0,$

- $I_i(x, y)$: intensity in the subaperture located at (x, y)in the quadrant i. Intensities are integrated during a modulation cycle. I_0 is the average intensity per subaperture.
- As wavefront slopes become small, the PyWFS becomes a direct phase measuring device.
- Measurements can be more precise if the edge of pyramid is sharper.

When the aberrations are large (e.g. defocus below), all the incoming light will fall only on one facet of the pyramid producing highly non-linear response (reaches saturation).



SHWFS and PyWFS visualization



Demonstration of tip-tilt effects for:

- a star flux of 1000 photons per integration time,
- 100 sub-apertures projected on a 12 x 12 grid over an obstructed circular aperture.

PyWFS over SHWFS is:

- High sensitivity to WF errors when used with unresolved sources.
- Easily tuned to seeing conditions

Credit: Astronomical Observatory of Padova, https://www.youtube.com/watch?v=qXqkbPZZdcs

Choosing a WFS

Linear, large dynamical range, poor sensitivity at low spatial frequencies

- Shack-Hartmann
- Curvature
- Modulated Pyramid

Linear, large dynamical range, high sensitivity

- Fixed Pyramid
- Zernike phase contrast mask
- Pupil plane Mach-Zehnder interferometer



Wavefront Reconstruction

- AO WFS does not measure the WF directly but its gradient/Laplacian.
- To determine WF phase from a map of its gradient, we reconstruct the WF from the measurement by performing matrix multiplication.
- Reconstruction must achieve the most accurate, lowest noise estimate possible in the least amount of computation.
- Given a measurement vector **S** (*m* elements of x- and y-slopes), determine the unknown vector of Φ (*n* commands/actuators/phase values).
- How to estimate $\boldsymbol{\Phi}$ from **S**?
 - First, go from actuators to centroid space by building a system matrix (interaction/response matrix, H). S = $H \Phi$
 - Then invert **H** to go from sensor measurements to actuator commands. Let's call H^{-1} as **R**. **R** is the reconstruction (command) matrix. $\Phi = R S$

How to compute the system matrix *H*?

H describes how the deformations Φ , affects the centroids *S*. $S = H \Phi + noise$ Can be calculated theoretically or preferably experimentally (also known as calibrating the system).

Zonal method

Modal methods

Basis set is the actuators Use an orthogonal modal basis set to represent the actuators (Zernike/Fourier modes)



SHWFS as an example

Ensure that the registration is right by adding waffle to the DM









Waffle is a repeated astigmatic pattern at the period of the SHWFS sampling over the entire aperture

Wavefront Reconstruction: Compute H

To link measurements **S** to the incoming phase Φ , apply unitary voltages to each actuators (keeping all the others to 0) and record the centroids.

$$\mathbf{S}_{i,j}^{x} = [(\phi_{i+1,j+1} + \phi_{i+1,j}) - (\phi_{i,j} + \phi_{i,j+1})]/2d$$

$$\mathbf{S}_{i,j}^{\gamma} = [(\phi_{i+1,j+1} + \phi_{i,j+1}) - (\phi_{i,j} + \phi_{i+1,j})]/2d,$$

 $\boldsymbol{\Phi}_{i,j}$ are the phase values at the four corners of a sub-aperture.



slope or y-× an to Each row corresponds



Matrix H (n x m elements)

.

٠

- The columns of *H* are the measurement vectors associated with each actuator
- Includes system defects/imperfections for e.g., alignment errors, defective sensor element(s), defective actuator(s) and crosstalk.

Wavefront Reconstruction: Invert H

- We have the interaction matrix **H**.
- Compute the reconstruction matrix R, by inverting H to solve for Φ

 $H \Phi = S$ $H^{T} H \Phi = H^{T} S$ $\Phi = (H^{T} H)^{-1} H^{T} S$ R

- Usually, there are more measurements than unknowns, m > n. It is an overdetermined system.
- *H* is a rectangular matrix and is not invertible. Also, H is singular because a few modes such as piston and waffle (in Fried geometry) can not be sensed (called blind or null modes). Noise propagates through them.
 - Piston is the constant retardation or advancement of the phase over the entire beam.
 - In the presence of waffle mode, the WF has zero average slope over every sub-aperture and it can not be sensed by the SHWFS.

 $S = H \Phi + noise$

 $\Phi = R S$

Wavefront Reconstruction: Invert H

(H^T H)⁻¹ H^T is a least square reconstructor. (H^T H)⁻¹ is not invertible. Invisible modes piston (left) and waffle (right.)



- Solution is to take the pseudoinverse of *H* using the singular value decomposition (SVD).
- Decompose **H** into $U\Sigma V^T$

• Decompose **H** into $U\Sigma V^T$

U is unitary matrix (m x n elements) Σ is a diagonal matrix (eigenvalues σ_i , n x n matrix that has non-zero elements on the diagonal) V^{τ} is a unitary matrix (n x n square matrix)

- Pseudoinverse of H using SVD is computed straightforwardly: $H^{-1} = V \Sigma^{-1} U^{T}$ and simply, R = SVD(H)
- $\Sigma^{-1} = 1/\sigma_i$ if $\sigma_i > 0$, and 0 if $\sigma_i = 0$

 $\Sigma = \left\{ \begin{array}{ccc} \sigma_{0} & 0 \\ \sigma_{1} & \\ & \sigma_{2} & \\ & & \\ 0 & & \sigma_{n} \end{array} \right\} \qquad \Sigma^{-1} = \left\{ \begin{array}{ccc} 1/\sigma_{0} & 0 \\ 1/\sigma_{1} & \\ & 1/\sigma_{2} & \\ & & \\ 0 & & 1/\sigma_{n} \end{array} \right\}$

SVD sorts the eigenvalues in descending order of mode sensitivity.

Wavefront Reconstruction: SVD truncation



If some diagonal elements (*σ_i*) are zero or close to zero (the matrix *H* is singular), the corresponding mode (centroid) measurement are the noisiest (for e.g., piston and waffle). It means that some modes of the DM are not detected by the WFS, they should be filtered out/discarded using the SVD truncation. Otherwise, the noise will be amplified.

<u>Σ</u>⁻¹ Example from Keck Observatory



Set these values to zero before taking a dot product $H^{-1} = V \Sigma^{-1} U^{T}$

Check out the following reference below to learn how you can penalize piston and local waffle and regularize your SVD before inverting *H*: D. T. Gavel, "Suppressing anomalous localized waffle behavior in least squares wavefront reconstruction," Proc. SPIE 4839, pp. 972–980 (2002).

Wavefront Reconstruction: Modal

You can choose to reconstruct certain modes and avoid reconstructing unwanted modes (e.g., waffle)

Zernike modes



Noise propagation, $(H^T H)^{-1}$ Here, tip-tilt are the nosiest modes (mode 0 is tip actually).

🕕 🍣 🥌 🌒



Let's look at the response of the PyWFS to Zernike modes.



Defocus

Spherical

70

10

Turbide et al. 2013

Wavefront Reconstruction: Modal

Fourier modes: Apply a sine/cosine wave on the DM and record the response of the sensor (build system matrix **H**). Fourier transform of a sine wave generates two identical PSFs on either side of the central PSF.



Non-modulated case (Guyon 2005)







Response of the PyWFS for one modulation cycle for a sine wave applied on the DM.

AO to Extreme-AO

Application: Direct imaging of Exoplanets

WFS from AO to Extreme-AO



WFS from AO to Extreme-AO

AO

SR < 90% (current best systems like Keck/LBT/VLT) in near infrared (NIR), under good seeing



Simulated single short-exposure image (PSF is not scaled with the image on right, it is just an example) Simulated short exposure image under post-AO residuals seen by the SPHERE instrument at VLT for a circular pupil

WFS from AO to Extreme-AO

AO

SR < 90% (current best systems like Keck/LBT/VLT) in near infrared (NIR), under good seeing

ExAO/ High-contrast imaging of Exoplanets, Required Strehl ratio ~ 99% in NIR



Simulated single short-exposure image (PSF is not scaled with the image on right, it is just an example) Simulated short exposure image under post-AO residuals seen by the SPHERE instrument at VLT for a circular pupil

Long-exposure simulated normalized image under SPHERE/VLT post-AO residuals

Low-order aberrations

Causes: Temperature variations, thermal distortions, optical/mechanical vibrations, alignment errors due to telescope motors and chromatic errors.

Effects: Starlight leak around a coronagraphic mask, prevent detection at small angles.

Non-common path aberrations (NCPA)

Cause: different AO sensing and science imaging channels. Effects: Evolving quasi-static stellar speckles, which not only mask faint exoplanet signals but also create false positive signals.

Low-wind effect

Cause: Spider arms of the secondary can cool below the ambient air temperature due to radiative losses. Change in air index from one side of the spider to the other.

Effects: Each quarter of the pupil shows different piston and sometimes tip-tilt phase errors.

Wind driven halo

Cause: High wind speeds at the upper level of turbulence across the pupil moving faster then the speed of AO loop correction.

Effects: A typical butterfly-shaped structure in the focal plane image along the wind direction.

WFS from AO to Extreme-AO: a few references

Extreme-AO instruments

GPI (Gemini Planet Imager): Macintosh et al. 2014 SPHERE (Spectro-Polarimetric High-contrast Exoplanet REsearch): Beuzit et al. 2019 SCExAO(Subaru Coronagraphic Extreme Adaptive Optics project): Jovanovich et al. 2015

Focal-plane WFSs

Malbet, Yu & Shao (1995), Guyon et al. 2005, Give'On et al. 2006, 2007; Borde & Traud 2006, Baudoz, P., et al 2006, Sauvage et al. 2012, Paul et al. 2013, Martinache 2013, Galicher et al. 2010, Mazoyer, J., et al. 2014 Gerard et al. 2018, 2019 Potier et al. 2019, Singh et. Al. 2019

Low-order WFS

Guyon, O., Matsuo, T., & Angel, R. 2009, ApJ, 693, 75 Vogt, F. P. A., Martinache, F., Guyon, O., et al. 2011, PASP, 123, 1434 Wallace et al. 2011 Petit et al., 2014 Singh, G., 2014, 2015, 2017