Coherent Wavefront Sensor for Optical Communications through Atmospheric Turbulence in Strong Background Environments

Victor A. Vilnrotter,* Lewis C. Roberts Jr.,† and J. Christopher Shelton†

ABSTRACT. — Deep space optical communications is an important component of the NASA roadmap, with the potential for greater data-rates and greater reach for both robotic and human missions in the future. The baselined signaling format for optical deep space communications is pulse-position modulation (PPM), detected directly with a large aperture ground-based optical receiver. While this design is perfectly adequate for low background reception, as encountered during nighttime operation, photon-counting receiver performance degrades significantly during the day when turbulence distorts the optical phase-front causing the receiver to open up its field-of-view (FOV) to collect the available signal energy, thus letting in more background photons that interfere with photon-counting detection. A plausible approach to improving daytime optical reception is to use Adaptive Optics (AO) to reconstruct the optical fields, enabling operation with much narrower FOV and thus reducing background interference. A key component of the AO system is the Shack-Hartmann Wavefront Sensor, initially developed for astronomical imaging applications, which however is very sensitive to daytime background interference. Here we propose and evaluate a Coherent Wavefront Sensor concept that virtually eliminates background interference, and in addition improves detection performance in high background environments, thus enabling efficient daytime optical communications over the photon-starved deep-space optical channel.

I. Introduction

Deep-space optical communication over interplanetary distances that often exceed many astronomical units (AU) will require sensitive detection techniques and efficient algorithms to close the communications link. The main goal of planetary exploration via manned or robotic spacecraft is discovery, either in the form of scientific measurements or close-up images, hence it is important to relay the information back to Earth accurately and in a timely manner. This implies the ability to transmit high-rate telemetry over great distances, which in turn implies high-power transmitters and large optical telescopes, both

* Communications Architectures and Research Section.

† Optics Section.

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

of which are rare commodities on a spacecraft. Optical communications holds the promise of greater reach than current microwave technology can provide, for future deep-space exploration. In terms of parameters relevant to optical communications, the requirements generally translate to high-power lasers and narrow optical beams, suitably modulated to enable rapid delivery of large amounts of data to an earth-based receiver.

Atmospheric turbulence over the terrestrial receiver distorts the phase-front of the received optical fields, effectively generating multiple spatial modes arriving from different directions [1], and causing the optical receiver to increase its field-of-view (FOV) in order to collect the available signal energy. But increasing the FOV admits more background radiation into the receiver, degrading communications performance.

In strong background and turbulence characteristic of daytime operation, the performance degradation of the receiver may be so severe that the optical link cannot be maintained. A promising approach to mitigating turbulence is to employ Adaptive Optics (AO) for optical communications. Some of the received signal energy is diverted to a Shack-Hartmann Wavefront Sensor, which employs a lenslet array and a quad-cell for each lenslet that analyzes the tilt of the phase-front over a large number of points in the pupil-plane of the optical receiver (an image of the receiving aperture), and controls a deformable mirror (DM) in a closed-loop configuration to reconstruct the distorted optical phase-front. In extreme cases, as much as 50% of the received signal energy may have to be diverted to the wavefront sensor in order to achieve the desired detection performance.

The Coherent Wavefront Sensor (C-WFS) described in this paper replaces each lenslet quad-cell with a single PIN diode, thus reducing each array dimension by a factor of 2, or the total number of PIN diode elements by a factor of 4. With the coherent detection approach, the C-WFS is flooded with a strong collimated local laser beam of essentially the same optical wavelength, combining with the signal field on the detector's surface and generating a cross-term at the detector output proportional to the signal amplitude, but with a large magnification factor that elevates the signal above the shot-noise. In strong background, the bandwidth can be limited to the signal modulation bandwidth, typically on the order of tens to hundreds of MHz. On the other hand, optical pre-detection filter bandwidths at 1064 and 1550 nm are on the order of 10–100 GHz wide, hence the direct-detection Shack-Hartmann WFS electronics observe much more background interference than the Coherent Wavefront Sensor electronics, where the background power is greatly reduced by filtering to the signal bandwidth.

In the signal-processing section of the Shack-Hartmann WFS electronics the direct-detected quad-cell outputs are summed pairwise, then differenced to obtain the position estimate in each direction, resulting in further performance degradation via the increase in error variance. By contrast, with coherent detection the PIN-diode array-phase is determined directly from the averaged in-phase (I) and quadrature (Q) components via a single calculation, resulting in significantly better performance in strong background, since coherent detection virtually eliminates background noise via electrical filtering.

In the rest of this paper, we derive the structure of the coherent wavefront sensor, and develop the signal processing algorithms for the C-WFS and the coherent optical array detection system. A detailed MATLAB simulation is developed employing Kolmogorov

phase-screens to simulate the turbulent optical fields, and input to the C-WFS to evaluate its performance. Next, the Cramer-Rao bound on array phase estimator performance is derived, and used to validate the simulation results. Finally, the performance of photon-counting and coherently detected optical PPM signals are evaluated and compared in a strong background environment characteristic of daytime reception from a Mars orbiting spacecraft.

II. Coherent Array Processing for Turbulent Optical Fields

Properly designed PIN-diode detector arrays can extract array phase information directly from the optical fields, to drive a deformable mirror or combine for coherent detection, thus reducing receiver complexity. The conceptual block diagram shown in Figure 1 refers to the telescope's pupil plane for ease of implementation, but equivalent performance can also be obtained with a properly designed focal-plane array. Figure 1 highlights the key components of a pupil-plane array coherent processing system using free-space optical components.

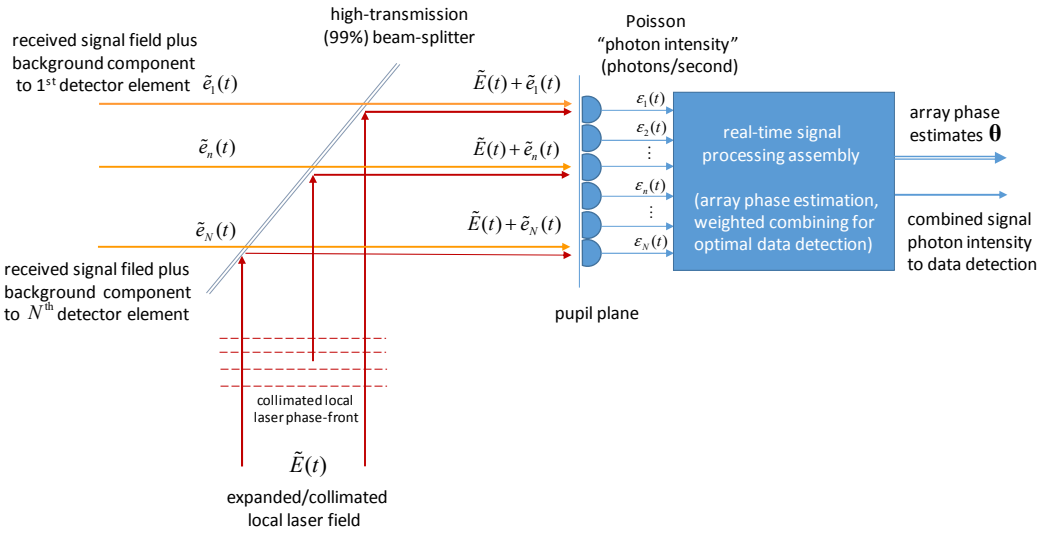

Figure 1. Conceptual block diagram of a coherent array optical receiver, designed to process turbulence-distorted optical fields for real-time array phase estimates and coherent signal detection.

Coherent detection signal model. The turbulence-distorted received signal field over each detector element is represented by the optical field components $\tilde{e}_n(t), 1 < n \le N$, in Figure 1. A locally generated optical field $\tilde{E}(t)$ with constant amplitude, frequency and phase is added to the signal components at the active surface of each detector element, resulting in the sum-field $\tilde{E}(t)+\tilde{e}_n(t)$. The *n*-th detector element responds to the optical intensity of the sum-field generating an electric current $\varepsilon_n(t)$ as the observable output.

The signaling format of interest for future deep space optical communication is pulse position modulation, or PPM, where a single optical place is placed in one of *M* consecutive slots to convey $\log_2(M)$ bits of information. It is tacitly assumed that

successive pulses from the laser transmitter are phase-coherent, which can be achieved in practice by controlling the quenching of the pump laser. The temporal evolution of the local and received signal field components can be represented as complex sinusoids,

$$\tilde{E}(t) = E\exp(-j\omega_{LO}\,t), \quad \tilde{e}_n(t) = s_n(t)\exp[\,j(\omega_0\,t + \theta_n)] \tag{1}$$

where E is the amplitude of the local field, ω_{LO} and ω_0 are the optical frequencies of the local and received fields, respectively, $s_n(t)$ is the real information-bearing modulation of the received field with amplitude s_n, and θ_n is the phase with respect to the receiver clock. For example, if the signal is intensity modulated PPM of order M with unit pulse amplitude s_n, the modulation consists of a "1" in the signal-slot and "0" in the remaining (M – 1) slots.

Assuming the local oscillator laser is set to the nominal optical frequency plus an intermediate frequency (IF), such that $\omega_0 - \omega_{LO} = \omega_{IF}$ is much greater than the modulation bandwidth, the optical photon-count process can be represented as follows [1]:

$$\begin{aligned}
\varepsilon_n(t) &= \tfrac{\eta}{h\nu}\int_{A_n} |\,\tilde{E}(t) + \tilde{e}_n(t)\,|^2\,dxdy = \tfrac{\eta A_n}{h\nu}\{|\,\tilde{E}(t)\,|^2 + |\,\tilde{e}_n(t)\,|^2 + 2\,\mathrm{Re}[\tilde{E}(t)\tilde{e}_n(t)]\} \\
&= \tfrac{\eta A}{h\nu}\Big(|\,\tilde{E}(t)\,|^2 + |\,\tilde{e}_n(t)\,|^2 + 2\,\mathrm{Re}\{E\exp(-j\omega_{LO}\,t)s_n(t)\exp[\,j(\omega_0\,t+\theta_n)]\}\Big) \\
&\cong \tfrac{\eta A}{h\nu}\{E^2 + 2E\,s_n(t)\cos(\omega_{IF}\,t + \theta_n)\}
\end{aligned}$$

According to the Poisson model, the function derived in Equation (1) generates a Poisson process with photon-count rate $\varepsilon_n(t)$. The continuous IF waveform is converted to samples for digital processing via short-term integration, which is equivalent to low-pass filtering, yielding: ${}_{IF}\chi_n(t_i) = \frac{1}{\Delta t}\int_{(i-1/2)\Delta t}^{(i+1/2)\Delta t} \varepsilon_n(t)\,dt, \quad t_i = i\Delta t$. The samples are Δt seconds apart, at a rate greater than the IF frequency hence justifying the approximation ${}_{IF}\chi_n(t_i) \cong \varepsilon_n(t_i)$. It is clear that with the Δt^{-1} scaling the samples become scaled Poisson-distributed random variables with conditional short-term mean $\overline{\chi_n(t_i)} = \varepsilon_n(t_i)$ and variance $\sigma^2(t_i) = \Delta t^{-2}\varepsilon_n(t_i)\Delta t = \varepsilon_n(t_i)/\Delta t$. Since the local laser field amplitude is much greater than the received signal amplitude, $E >> s_n$, it follows that the variance of the i-th sample is dominated by the local laser intensity, hence the approximation $\sigma^2(t_i) = \varepsilon_n(t_i)/\Delta t \cong \frac{\eta A}{h\nu}E^2/\Delta t$ is valid. The conditional sample mean is due to both the high-intensity local laser and the modulation: since the local laser intensity is constant, the mean count due to the local laser can be subtracted out, leaving the conditional short-term mean due to the signal modulation. Since high-intensity Poisson shot noise is approximately Gaussian, the "signal plus additive Gaussian noise" model can now be invoked to represent the IF samples:

$${}_{IF}r_n(t_i) = \overline{{}_{IF}\chi_n(t_i)} + n(t_i) = 2\tfrac{\eta A}{h\nu}E\,s_n(t_i)\cos(\omega_{IF}\,t + \theta_n) + n(t_i) \tag{2}$$

where $\overline{{}_{IF}\chi_n(t_i)} \triangleq \overline{\chi_n(t_i)} - \frac{\eta A}{h\nu} E^2$. The variance of the IF noise samples is $\sigma^2 = \mathcal{E}[n^2(t_i)] \cong \frac{\eta A}{h\nu} E^2 / \Delta t$, where $\mathcal{E}$ is the expectation operator.

The next step is downconversion to complex baseband. This is accomplished by multiplying the IF samples by sinusoidal samples at the IF frequency, namely $2\cos(\omega_{IF} t_i)$ and $2\sin(\omega_{IF} t_i)$. Invoking the in-phase and quadrature representation of bandpass noise, the filtered shot-noise can be represented as $n(t_i) = n_I(t_i)\cos(\omega_{IF} t_i) - n_Q(t_i)\sin(\omega_{IF} t_i)$, where $\mathcal{E}[n_I^2(t_i)] = \mathcal{E}[n_Q^2(t_i)] \cong \frac{\eta A}{h\nu} E^2 = \sigma^2$. Carrying out the downconversion operation and low-pass filtering (LPF) to eliminate the double-frequency terms yields the following baseband samples $r_{n,I}(t)$ and $r_{n,Q}(t)$:

$$\begin{aligned} r_{n,I}(t_i) &= 2\tfrac{\eta A}{h\nu} E s_n(t_i)\cos(\omega_{IF} t_i + \theta_n) 2\cos(\omega_{IF} t_i) + n(t_i) 2\cos(\omega_{IF} t_i)\,|_{LPF} \\ &= 2\tfrac{\eta A}{h\nu} E s_n(t_i)\cos(\theta_n) + n_I(t_i) \end{aligned}$$

$$\begin{aligned} r_{n,Q}(t_i) &= 2\tfrac{\eta A}{h\nu} E s_n(t_i)\cos(\omega_{IF} t_i + \theta_n) 2\sin(\omega_{IF} t_i) + n(t_i) 2\sin(\omega_{IF} t_i)\,|_{LPF} \\ &= 2\tfrac{\eta A}{h\nu} E s_n(t_i)\sin(\theta_n) + n_Q(t_i) \end{aligned}$$

The complex baseband sample $\tilde{r}_n(t_i)$ can now be expressed as $\tilde{r}_n(t_i) = r_{n,I}(t_i) + j r_{n,Q}(t_i)$, or in terms of signal parameters $\tilde{r}_n(t_i) = 2\tfrac{\eta}{h\nu} A E s_n(t_i)\Delta t[\cos(\theta_n) + j\sin(\theta_n)] \triangleq \alpha_{n,i}\exp(j\theta_n)$. Similarly, the complex noise samples are defined as $\tilde{n}(t_i) = n_I(t_i) + j n_Q(t_i)$, where the in-phase and quadrature noise components each have variance σ^2 , hence the variance of the complex noise sample is $\mathcal{E}\,|\tilde{n}(t_i)|^2 = \mathcal{E}\,|n_I(t_i) + j n_Q(t_i)|^2 = \mathcal{E}\,n_I^2(t_i) + \mathcal{E}\,n_Q^2(t_i) = 2\sigma^2 \triangleq \sigma_n^2$.

Note that multiplication of a complex noise sample by an arbitrary phasor such as $\exp(-j\psi)$, rotates the phase of the signal component, but leaves the variance of the complex noise unchanged:

$$\tilde{r}_n(t_i)\exp(-j\psi_n) = \alpha_{n,i}\exp[j(\theta_n - \psi)] \text{ where } \mathcal{E}\,|\tilde{n}(t_i)\exp(-j\psi)|^2 = \sigma_n^2$$

This observation enables compensating the phase of the received signal, θ_n, using an estimate of the phase, $\hat{\theta}_n$. Hence, if an accurate estimate of the sample phase is available, such that $\hat{\theta}_n \cong \theta_n$, then multiplying the received sample by $\exp(-j\hat{\theta}_n)$ yields a real sample with effectively zero phase, without changing the variance of the noise term:

$\tilde{r}_n(t_i)\exp(-j\hat{\theta}_n) = \alpha_{n,i}\exp[j(\theta_n - \hat{\theta}_n)] + \tilde{n}(t_i)\exp(-j\hat{\theta}_n) \underset{\theta_n=\hat{\theta}_n}{=} \alpha_{n,i} + \tilde{n}'(t_i)$, where $\mathcal{E}\,|\tilde{n}(t_i)\exp(-j\theta_n)|^2 = \sigma_n^2$.

Sample-SNR with Coherent Detection: The sample signal-to-noise ratio (sample-SNR) is a key parameter that determines communications performance with coherently detected turbulent fields, as well as the performance of phase estimation algorithms. Here we examine the signal-to-noise ratio of phase-compensated samples, which is critical in determining coherent-detected communications performance.

For the case of phase-compensated samples, the sample signal-to-noise ratio is defined as the ratio of the square of the sample mean (the signal term in the coherent model) and the sample noise variance. In the absence of background radiation, coherent signal processing yields the following expression for the sample signal-to-noise ratio, or sample-SNR of the *i*-th sample from the *n*-th detector element:

$$SNR_n(t_i) = \frac{\left(2\frac{\eta}{h\nu}AEs_n(t_i)\right)^2}{\frac{\eta}{h\nu}AE^2/\Delta t} = 4\frac{\eta}{h\nu}As_n^2(t_i)\Delta t = 4K_s(t_i)/N \tag{3}$$

where *N* is the number of PIN-diode detectors in the array, $K_s(t_i)$ is the average signal photon-count per sample over the entire receiver aperture, and hence $K_s(t_i)/N = \frac{\eta}{h\nu}As_n^2(t_i)\Delta t$ is the average signal count over each detector element of the *N*-dimensional array, assuming that signal energy is divided equally among the detector elements in the pupil-plane. The SNR is critical for determining the performance of systems that employ the Gaussian additive noise model.

Consider the case where turbulence distorts the received optical fields, and in addition strong background fields impinge on the detector array. Similar to the signal fields, the background field components reaching the detector elements in the pupil-plane are also multiplied by the local laser field, and hence the detector output can be expressed as $\tilde{\mathcal{O}}_n(t_i) = 2\frac{\eta}{h\nu}AE\tilde{b}_n(t_i)$. The variance of each component is the expected value of the squared magnitude of the background field components: $\mathcal{E}\,|\frac{\eta}{h\nu}AE\tilde{b}_n(t_i)|^2\,\Delta t \equiv K_b/N$. With background, the SNR includes both the shot-noise and the amplified background (thermal noise can generally be neglected compared to shot-noise), yielding:

$$SNR_{b,n}(t_i) = \frac{\left(2\frac{\eta}{h\nu}AEs_n(t_i)\right)^2}{\frac{\eta}{h\nu}AE^2/\Delta t + \xi\mathcal{E}\left|2\frac{\eta}{h\nu}AE\tilde{b}_n(t_i)\right|^2} = \frac{4\frac{\eta}{h\nu}As_n^2(t_i)\Delta t}{1+4\xi\frac{\eta}{h\nu}A\mathcal{E}\,|\tilde{b}_n(t_i)|^2\,\Delta t} = \frac{4K_s(t_i)/N}{1+4\xi K_b/N} \tag{4}$$

where ξ is the bandwidth-suppression factor defined as the ratio of electrical sample bandwidth (inversely proportional to Δt) to optical filter bandwidth, and $K_b/N = \frac{\eta}{h\nu}AE\mathcal{E}\,|\tilde{b}_n(t_i)|^2\,\Delta t$ and K_b is the average number of background photons per sample entering the receiver.

Background suppression example: The signal parameters in the following example were taken from [2], in order to demonstrate the impact of filtering the background noise to the signal bandwidth. The electrical bandwidth of a 0.17 nm optical filter at 1550 nm can be calculated by differentiating the expression $f = c / \lambda$ with respect to λ and multiplying through by $d\lambda$, yielding $df = -c\, d\lambda / \lambda^2$. Replacing the differentials by discrete increments yields the electrical bandwidth, Δf, of the 0.17 nm or $\Delta\lambda = 1.7\times10^{-10}$ m optical filter at $\lambda = 1550$ nm as follows:
$\Delta f = c\,\Delta\lambda / \lambda^2 = (3\times10^8)(1.7\times10^{-10}) / (2.4\times10^{-12}) = 2.125\times10^{10}$ Hz or 21.25 GHz.

Consider PPM signaling at 2 MBPS data-rate from 2 AU as suggested in [2]. With rate ½ forward error correction coding, the slot-width becomes $\tau_{FER} = \varsigma \log_2(M) / RM$. With a coding rate of $\varsigma = 0.5$ and $M = 16$ we have $\tau_{FER} = \varsigma \log_2(M) / RM = 0.5\times4 / (32\times10^6) = 6.25\times10^{-8}$ seconds. Using the time-bandwidth product to estimate the downconverted signal bandwidth, that is letting $\tau_{FER}\Delta f = 1$, this corresponds to a signal bandwidth of approximately $\Delta f_{FER} = 1 / \tau_{FER} = 16\times10^6$ Hz or 16 MHz. For an electrical bandwidth of 16 MHz, the ratio of signal to optical filter bandwidth is $\xi = (16\times10^6) / (2.125\times10^{10}) = 7.53\times10^{-4}$, implying that a coherent receiver with lowpass filter matched to the signal bandwidth observes a factor of $\xi = 7.53\times10^{-4}$ fewer background photons per slot, than a direct detection photon-counting receiver. For example, the high-background case with $K_b = 10$ photons per slot, yields $SNR_b(t_i) = \dfrac{4K_s(t_i)}{1+4(7.53\times10^{-4})(10)} = \dfrac{4K_s(t_i)}{1.03}$, which is only 0.1 dB loss from the SNR of the shot-noise limited sample-SNR, due to strong background. Even in bright daylight with Mars in the field-of-view when the background is maximized, when the background slot-count can be as great as $K_b = 31$ photons/slot, the loss with coherent detection remains at a relatively small 0.4 dB, which does not seriously impact detection performance. Even this simple "order of magnitude" example illustrates that coherent detection is very effective in suppressing strong background interference, when the ratio of electrical signal bandwidth to optical pre-detection filter bandwidth is small.

III. Maximum Likelihood Phase Estimator for Known Waveforms

The objective of coherent signal processing is to enable the detection of high data-rate optical PPM symbols with the required fidelity. With coherent processing, a detector array composed of high quantum-efficiency PIN diodes can be used to detect the sum of the signal and local optical fields. In this section, we develop the structure of the maximum likelihood array phase estimator, and evaluate its performance via Cramer-Rao bounds on the variance of phase estimation error.

The phase distribution of the turbulent optical signal field over the receiving aperture with diameter D can be transferred to the telescope's pupil plane with diameter d determined by the magnification factor M of the optical train as $d = D/M$. For example, generating a 1-cm pupil-plane for a 5-m diameter receiving telescope requires a magnification of $M = 500$, which is well within the capabilities of free-space optical components.

A conceptual block diagram of the coherent processing system for real-time estimation of the array phase is shown in Figure 2. An array of N PIN-diode detector elements placed in the pupil-plane is flooded with a collimated local laser field, and the detector outputs downconverted to complex baseband for processing. The sum of signal plus local laser fields over each detector element generates an intensity process at a convenient IF frequency that contains the signal modulation in the cross-term. The signal plus shot-noise are band-pass processes filtered to the signal bandwidth around the IF frequency, resulting in bandlimited signals that are further downconverted to baseband, yielding the following complex baseband samples at the output of the n-th detector element:

$$\tilde{r}_n(t_i) = \tilde{x}_n(t_i) + \tilde{n}_n(t_i) = \alpha_{n,i} \exp(j\theta_n) + \tilde{n}_n(t_i) \tag{6}$$

where $\alpha_{n,i} = 2\frac{\eta}{h\nu} AE\, s_n(t_i)$, $s_n(t_i)$ is a real modulation function that enables the baseband representation of PPM and other amplitude modulations, and $\tilde{n}_n(t_i)$ is circular Gaussian noise with variance $\sigma_n^2 = 2\sigma^2$, assumed to be the same for each detector element. The phase θ_n is slowly-varying with respect to the modulation but in general correlated across the array, hence it is essentially constant over the estimator's update rate, selected to be short compared to the coherence-time of the phase process due to turbulence.

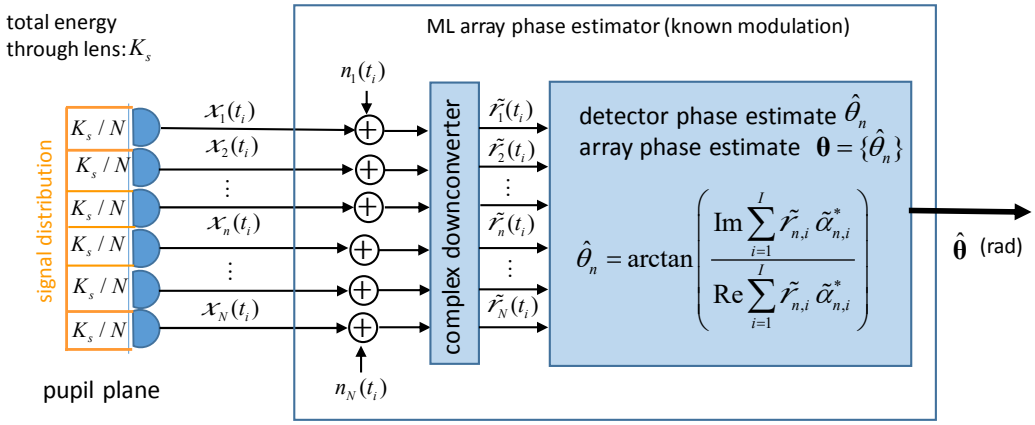

Figure 2. Conceptual block diagram of maximum likelihood phase estimator for an array of PIN diodes, with coherent detection.

Under the assumption that the array phases θ_n are unknown but constant during the observation interval, noise samples from different (spatially disjoint) detectors are statistically independent, $\mathcal{E}(\tilde{n}_{n,i}\tilde{n}^*_{m,k}) = 0 \quad n \neq m, i \neq k$, with variance $\mathcal{E}(|\tilde{n}_{n,i}|^2) = \sigma_n^2$. Hence, the joint probability density of the noise-sample vector is the product of individual densities, due to statistical independence:

$$p(\tilde{\mathbf{n}}) = \left(\pi\sigma_n^2\right)^{-N\times I} \prod_{n=1}^{N}\prod_{i=1}^{I} \exp(-|\tilde{n}_{n,i}|^2/\sigma_n^2) \tag{7a}$$

Recalling that the received samples are the sum of the signal plus noise samples, we can write the noise in terms of the signal as $\tilde{n}_{n,i} = \tilde{r}_{n,i} - \alpha_{n,i}\exp(j\theta_n)$, and rewrite the joint conditional density of received samples as

$$p(\tilde{\boldsymbol{r}}\,|\,\boldsymbol{\theta}) = \left(\pi\sigma_n^2\right)^{-N\times I} \prod_{n=1}^{N}\prod_{i=1}^{I} \exp(-|\tilde{r}_{n,i} - \alpha_{n,i}\exp(j\theta_n)|^2/\sigma_n^2) \tag{7b}$$

The maximum likelihood (ML) estimates of the parameters are those values that simultaneously maximize the conditional joint probability density, or equivalently its natural logarithm, known as the "conditional log-likelihood function" $\Lambda(\tilde{\boldsymbol{r}}\,|\,\boldsymbol{\theta})$:

$$\Lambda(\tilde{\boldsymbol{r}}\,|\,\boldsymbol{\theta}) \equiv \ln[\,p(\tilde{\boldsymbol{r}}\,|\,\boldsymbol{\theta})] = \frac{2}{\sigma_n^2}\mathrm{Re}\left(\sum_{n=1}^{N}\sum_{i=1}^{I}\tilde{r}_{n,i}\,\alpha_{n,i}\exp(-j\theta_n)\right)$$
$$-\frac{1}{\sigma_n^2}\sum_{n=1}^{N}\sum_{i=1}^{I}|\alpha_{n,i}|^2 - \frac{1}{\sigma_n^2}\sum_{n=1}^{N}\sum_{i=1}^{I}|\tilde{r}_{n,i}|^2 - (N\times I)\ln(\pi\sigma_n^2)$$

Therefore the ML estimate of the vector $\boldsymbol{\theta}$ is that vector, $\hat{\boldsymbol{\theta}}$, that maximizes the reduced log-likelihood function $\Lambda_0(\tilde{\boldsymbol{r}}\,|\,\boldsymbol{\theta})$, where irrelevant terms that do not contribute to the maximization have been removed:

$$\hat{\boldsymbol{\theta}} = \max_{\boldsymbol{\theta}} \Lambda_0(\tilde{\boldsymbol{r}}\,|\,\boldsymbol{\theta}) = \max_{\boldsymbol{\theta}} \mathrm{Re}\left(\sum_{n=1}^{N}\sum_{i=1}^{I}\tilde{r}_{n,i}\,\alpha_{n,i}\exp(-j\theta_n)\right) \tag{8}$$

Expanding the double-sum and rearranging terms yields the following expression:

$$\hat{\boldsymbol{\theta}} = \max_{\boldsymbol{\theta}} \Lambda_0(\tilde{\boldsymbol{r}}\,|\,\boldsymbol{\theta}) = \max_{\boldsymbol{\theta}} \mathrm{Re}\left(\sum_{n=1}^{N}\sum_{i=1}^{I}\tilde{r}_{n,i}\,\alpha_{n,i}\exp(-j\theta_n)\right)$$
$$= \max_{\boldsymbol{\theta}} \mathrm{Re}\left(\exp(-j\theta_1)\sum_{i=1}^{I}\tilde{r}_{1,i}\,\alpha_{1,i}\right) + \cdots + \max_{\boldsymbol{\theta}} \mathrm{Re}\left(\exp(-j\theta_N)\sum_{i=1}^{I}\tilde{r}_{N,i}\,\alpha_{N,i}\right)$$

Maximizing with respect to the vector $\boldsymbol{\theta}$ is equivalent to maximizing over each of its components, θ_n individually. Let $\tilde{z}_n = \sum_{i=1}^{I} \tilde{r}_{n,i}\,\alpha_{n,i}$. The expression $\mathrm{Re}\{\tilde{z}_n \exp(-j\theta_n)\}$ is maximized with respect to θ_n when $\theta_n = \arg(\tilde{z}_n)$, taking on its maximum value $|\tilde{z}_n|$. Therefore, the maximum likelihood estimate of θ_n is the argument of $\tilde{z}_n$, yielding

$$\hat{\theta}_n = \arctan\left(\frac{\mathrm{Im}\sum_{i=1}^{I} \tilde{r}_{n,i}\,\alpha_{n,i}}{\mathrm{Re}\sum_{i=1}^{I} \tilde{r}_{n,i}\,\alpha_{n,i}}\right) \tag{9}$$

Finally, the ML estimate of array phase is the full array composed of its components: $\hat{\boldsymbol{\theta}} = \left(\hat{\theta}_1 \quad \hat{\theta}_2 \quad \cdots \quad \hat{\theta}_n \quad \cdots \quad \hat{\theta}_N\right)$. Note that this form of the ML phase estimator needs to have a copy of the modulation as contained in $\alpha_{n,i}$, in order to perform the correlation with the received signal in Equation (9). This implies that known frame-markers, headers and other suitable segments of the high-rate data-stream are best suited for phase estimation in this application.

IV. Cramer-Rao Bound on Array Phase Estimator Performance

The Cramer-Rao Bound (CRB) is the lower bound on the variance of any unbiased estimator, hence it is useful in determining achievable estimator performance. As shown by the background suppression example above, even strong background power passing through the optical filter is reduced to negligible levels when filtered to nominal signal bandwidths, hence the dominant noise source with coherent detection is shot noise generated by the local laser when the bandwidth ratio ξ is on the order of 10^{-3} or less. Therefore, in the following we consider the shot-noise limited case characterized by the signal-to-noise ratio of Equation (3).

The CRB can be determined from the log-likelihood function, including the noise-variance scaling in this derivation since performance depends on the variance of the additive noise: $\Lambda_0(\tilde{\boldsymbol{r}} \mid \boldsymbol{\theta}) = \frac{2}{\sigma_n^2}\mathrm{Re}\left(\sum_{n=1}^{N}\sum_{i=1}^{I} \tilde{r}_{n,i}\,\alpha_{n,i}\exp(-j\theta_n)\right)$. The CRB is the expected value of the negative of the inverse of the Fisher Information Matrix (FIM) $\boldsymbol{\mathcal{I}}$, expressed mathematically as $CRB \equiv \min \mathrm{var}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) = -\boldsymbol{\mathcal{I}}^{-1}$ where $\boldsymbol{\mathcal{I}} = \begin{bmatrix} \mathcal{I}_{11} & \cdots & \mathcal{I}_{1N} \\ \vdots & \ddots & \vdots \\ \mathcal{I}_{N1} & \cdots & \mathcal{I}_{NN} \end{bmatrix}$ and $\mathcal{I}_{m,n} = -\mathcal{E}\dfrac{\partial^2 \Lambda(\tilde{\mathbf{r}} \mid \boldsymbol{\theta})}{\partial\theta_m\,\partial\theta_n} \quad m,n = 1,\ldots,N$. Consider the general diagonal terms $J_{n,n}$ and off-diagonal terms $J_{m,n}$ of the FIM, before taking the negative of the expectation:

$J_{n,n} = \dfrac{\partial^2 \Lambda(\tilde{\mathbf{r}} \mid \boldsymbol{\theta})}{\partial \theta_n^2}$ and $J_{m,n} = \dfrac{\partial^2 \Lambda(\tilde{\mathbf{r}} \mid \boldsymbol{\theta})}{\partial \theta_m \partial \theta_n}$. Carrying out the differentiation for the diagonal elements yields:

$$J_{nn} = \frac{\partial_n^2 \Lambda(\tilde{\mathbf{r}} \mid \boldsymbol{\theta})}{\partial \theta_n^2} = \frac{\partial}{\partial \theta_n}\left(\frac{\partial}{\partial \theta_n} \mathrm{Re}\left(\exp(-j\theta_1) \sum_{i=1}^{I} \tilde{r}_{1,i}\, \alpha_{1,i} \right) + \ldots + \frac{\partial}{\partial \theta_n} \mathrm{Re}\left(\exp(-j\theta_N) \sum_{i=1}^{I} \tilde{r}_{N,i}\, \alpha_{N,i} \right) \right)$$

$$= \frac{\partial}{\partial \theta_n} \mathrm{Re}\left(-j \exp(-j\theta_n) \sum_{i=1}^{I} \tilde{r}_{n,i}\, \alpha_{n,i} \right) = -\mathrm{Re}\left(\exp(-j\theta_n) \sum_{i=1}^{I} \tilde{r}_{n,i}\, \alpha_{n,i} \right)$$

because $\dfrac{\partial}{\partial \theta_n} f(\theta_m) = 0 \quad m \neq n$. Similarly, carrying out the differentiation for the off-diagonal elements shows that the off-diagonal elements are all zero:

$$J_{nm} = \frac{\partial^2 \Lambda(\tilde{\mathbf{r}} \mid \boldsymbol{\theta})}{\partial \theta_n \partial \theta_m} = \frac{\partial}{\partial \theta_m}\left(\frac{\partial}{\partial \theta_n} \mathrm{Re}\left(\exp(-j\theta_1) \sum_{i=1}^{I} \tilde{r}_{1,i}\, \alpha_{1,i} \right) + \cdots + \frac{\partial}{\partial \theta_n} \mathrm{Re}\left(\exp(-j\theta_N) \sum_{i=1}^{I} \tilde{r}_{N,i}\, \alpha_{N,i} \right) \right)$$

$$= -j \frac{\partial}{\partial \theta_m} \mathrm{Re}\left(\exp(-j\theta_n) \sum_{i=1}^{I} \tilde{r}_{n,i}\, \alpha_{n,i} \right) = 0 \,.$$

Hence, the FIM is an $N \times N$ diagonal matrix,

$$\boldsymbol{I} = -\mathcal{E} \underset{N \times N}{\mathrm{diag}}(J_{nn}) = \mathcal{E} \underset{N \times N}{\mathrm{diag}} \left[\mathrm{Re}\left(\exp(-j\theta_n) \sum_{i=1}^{I} \tilde{r}_{n,i}\, \alpha_{n,i} \right) \right].$$

Finally, substituting the expression for the received samples $\tilde{r}_n(t_i)$ and carrying out the expectation of the general term yields:

$$\mathcal{E}\, \mathrm{Re}\left(\frac{2}{\sigma_n^2} \sum_{i=1}^{I} \tilde{r}_{n,i}\, \alpha_{n,i} \exp(-j\theta_n) \right)$$

$$= \mathrm{Re}\left(\frac{2}{\sigma_n^2} \sum_{i=1}^{I} \mathcal{E}[\alpha_{n,i} \exp(j\theta_n) + \tilde{n}_n] \times \alpha_{n,i} \exp(-j\theta_n) \right)$$

$$= \frac{2}{\sigma_n^2} \sum_{i=1}^{I} \alpha_{n,i}^2 = \frac{1}{\sigma^2} \sum_{i=1}^{I} \left| 2 \frac{\eta}{h\nu} AE\, s_n(t_i) \right|^2 = \frac{4I}{\frac{\eta}{h\nu} AE^2 / \Delta t} \left(\frac{\eta}{h\nu} AE\, s_n(t_i) \right)^2$$

$$= 4I \frac{(\frac{\eta}{h\nu})^2 A^2 E^2 s_n^2(t_i)}{\frac{\eta}{h\nu} AE^2 / \Delta t} = \frac{4IK_s(t_i)}{N} \quad \text{where } K_s = \frac{\eta}{h\nu} A s_n^2(t_i) \Delta t.$$

Therefore the CRB for pupil-plane phase estimation reduces to an $N \times N$ diagonal matrix, $CRB = -\mathbf{I}^{-1} = CRB_n \times \mathbf{I}_N$, with components given by the expression

$$CRB_n = \frac{N}{4IK_s} = \frac{1}{I \times SNR_n} \qquad 1 \le n \le N \tag{10}$$

where $\mathbf{I}_N$ is the N-dimensional identity matrix, I is the number of samples used in computing the phase estimate, K_s is the average number of signal photons per sample over the entire pupil, and $SNR_n = 4K_s / N$. Therefore, the CRB for each detector element is directly proportional to the number of matched detector elements N, and inversely proportional to the total number of complex samples I, as well as the total number of photons per sample over the entire receiver aperture.

V. Numerical Results

The maximum likelihood phase estimator described above was simulated in MATLAB, along with realistic phase processes generated via Kolmogorov phase-screens, in order to evaluate the performance of the array phase estimator. An example of a Kolmogorov phase-screen is shown in Figure 3, for a 5-meter optical receiver aperture with average seeing characterized by Fried parameter $r_0 = 6$ cm. The phase trajectory examples for a 5-meter diameter receiver were generated by selecting pixels near the left edge of the screen spanning a 5-meter aperture (or roughly 83 r_0 units), and sliding it across the phase-screen horizontally to simulate the motion of the turbulent optical phase across the aperture.

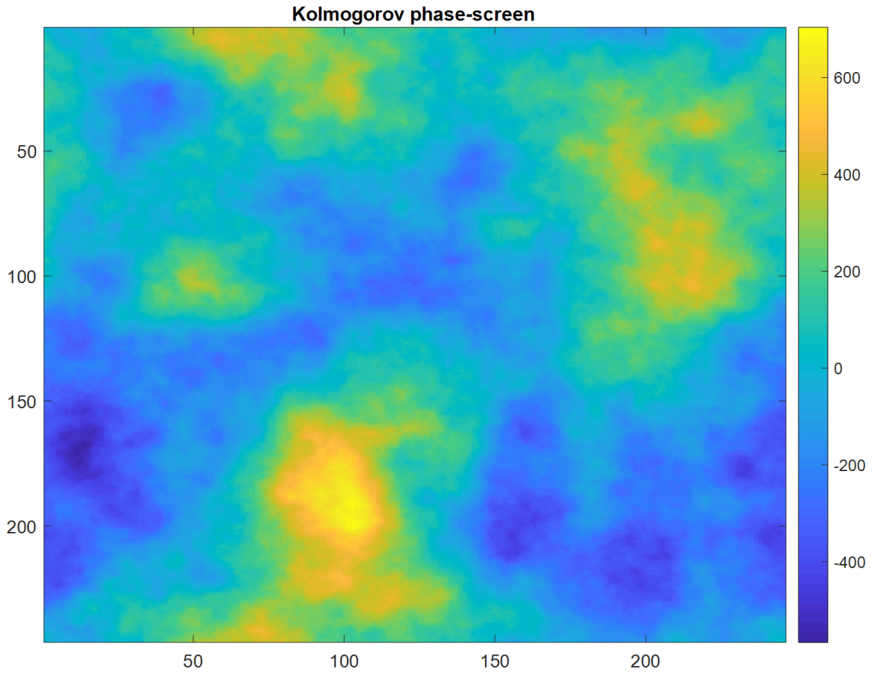

Figure 3. Kolmogorov phase-screen in units of Fried parameter r_0, for $r_0 = 6$ cm. Intensity scale represents phase, in radians.

An example of the resulting phase trajectories are shown in Figure 4 as the black dashed lines, representing the time-evolution of the phase-process at four selected points across the 5-m aperture. The phase-trajectories were unwrapped and initialized modulo 2π, which does not impact signal-combining but helps to visualize the evolution of the phase estimates from a common point in Figure 4.

The received signals were simulated by adding shot-noise to the received complex samples, and applying the noise-corrupted samples to the ML phase estimator structure defined in Equation (9). For PPM, correlation of the received PPM sequence with a known noise-free reference waveform (such as frame-marker, header, or idle sequences) observes only the signal-slots and nulls the noise-slots, hence the relevant SNR is that of the signal-slot. In the following, it is assumed that temporal synchronization has been established in the receiver, and that phase estimates are obtained at the matched signal-slots, with one sample per slot by setting $I = 1$ in Equation (10). Note that for the purposes of phase estimation, higher-SNR estimation samples can always be created if needed by adding together multiple high-rate PPM signal-slot samples to create improved estimation samples, provided the summation interval remains much smaller than the turbulence coherence-time (typically on the order of 10^{-3} seconds).

When generating Figure 4, the signal amplitude over each detector element in the pupil-plane was set to 1 unit, with the standard deviation of the samples set to 0.3 units. This implies a signal-to-noise ratio of roughly 10 dB for each estimation sample consisting of one or more PPM signal-slots, which is reasonable for high-rate data communication. The ML phase estimator acquires the phase on the first estimate, and continues to track the phase trajectory accurately, as can be seen in Figure 4. Note that when N channels are combined coherently in the pupil-plane, as would be the case for signal detection, the SNR of the combined output signal is N times greater than that of the individual channels.

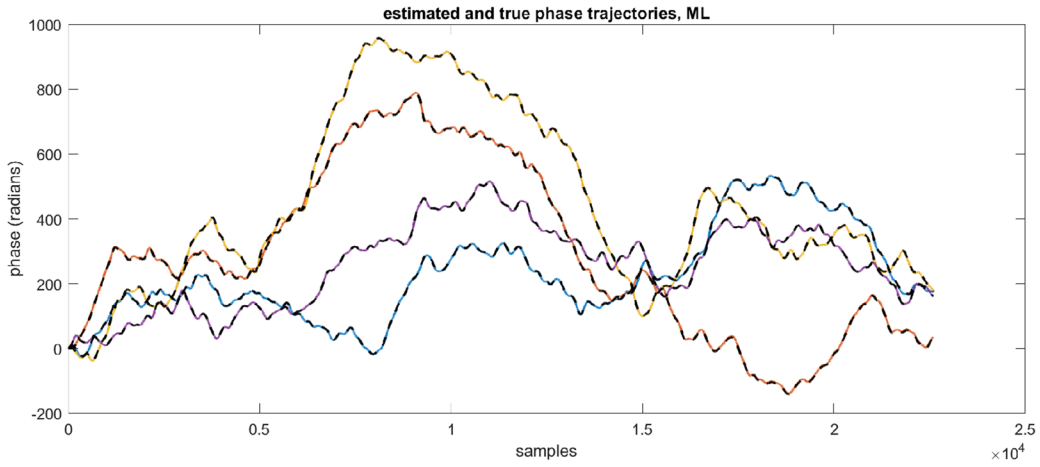

Figure 4. Example of 4 representative phase trajectories, generated by sliding pixels across the Kolmogorov phase-screen. Simulated phase trajectories are shown as dashed black lines; ML phase estimates are solid colors.

The standard deviation of the phase estimation error is a key parameter characterizing wavefront sensor and signal-combiner performance. This quantity can be estimated by

forming the difference between the true phase trajectory and the estimate, but at least one sample later to account for the inevitable delay between starting the measurement and applying the estimate. The results of the simulation are shown in Figure 5, where the red circles represent the simulation points and the dashed blue line is the CRB on phase estimation error, as defined in Equation (10). The standard deviation of the phase estimation error is shown as a function of the underlying signal-to-noise ratio, or SNR, in dB.

In Figure 5, the phase estimation error increases dramatically in the threshold region below 10 dB sample-SNR, mainly due to the way estimation error is defined: it is the difference between the true phase and the estimated phase. In the region below threshold a slight increase in noise begins to cause cycle-slips, or instantaneous changes of 2π radians, as shown in Figure 6a. This type of error may not impact some receiver functions that depend on phase only modulo 2π, such as coherent symbol detection.

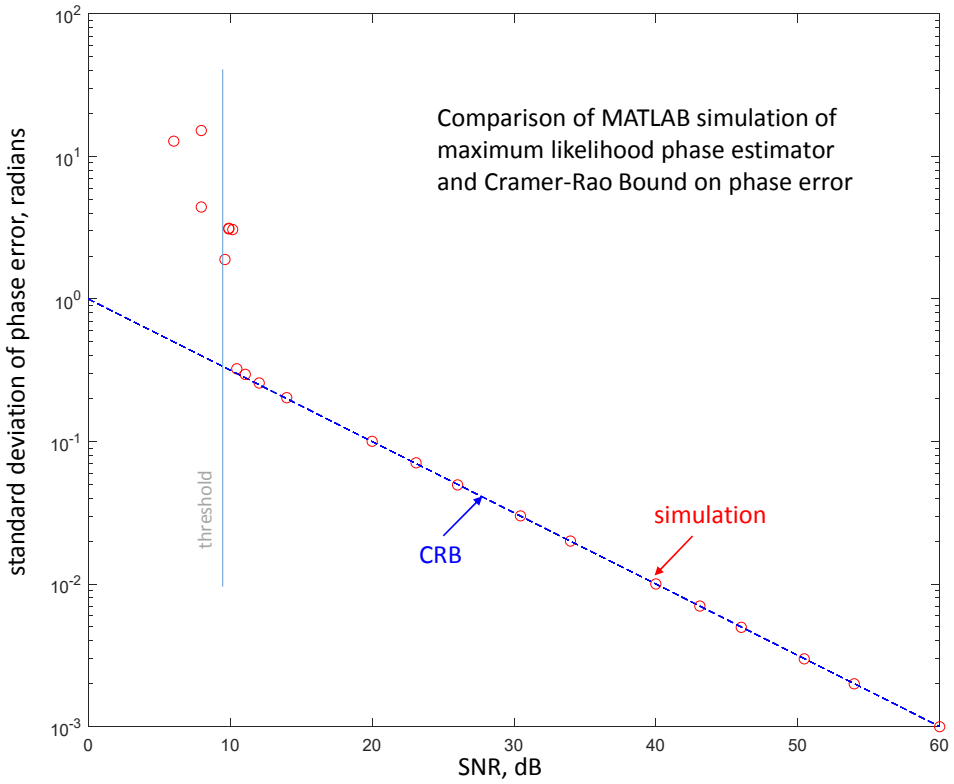

Figure 5. Comparison of MATLAB simulation and CRB for the ML phase estimator derived in Equation (10), over a wide range of sample-SNRs.

However, in applications where phase is first converted to delay or distance before applying it to a device such as the deformable mirror, cycle-slips may cause catastrophic errors. Figure 6b is an example of continuous 2π phase-jumps, or cycle-slips, due to increased noise below the indicated threshold of 10 dB.

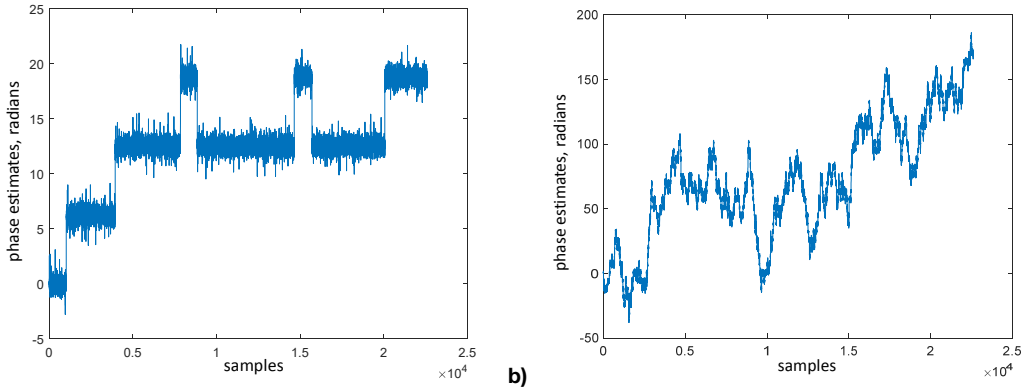

Figure 6. a) Example of occasional 2π cycle-slips near threshold, at SNR = 8 dB; b) Continuous cycle-slips at SNR = 0 dB.

The impact of phase estimation errors on the combined output can be seen in Figure 7, where the simulated output of a 7-channel combiner, selected from detector elements located across the 5-m simulated aperture with 6 cm Fried parameter, by sliding the Kolmogorov phase-screen across the aperture as described above. It can be seen that the combined output begins to degrade below 10 dB sample-SNR, consistent with the threshold effect shown in Figure 5. However, the degradation in combined SNR is graceful, and not instantly catastrophic as indicated in Figure 5. The reason for this effect is that the rms phase error simulated in Figure 5 includes occasional 2π phase-jumps as shown in Figures 6a and 6b, which however do not impact combining performance since adding 2π to any phase does not change the modulo-2π phase. The combined SNR degrades below threshold due to continuous cycle-slips, and not due to occasional 2π phase-jumps. Note that above 20 dB sample-SNR the combined SNR begins to saturate, due to the 1-sample delay between the time the estimate is made, and applied.

The combining gain measures the deviation of the simulated combined channel SNR from the ideally combined SNR (dashed blue line), in dB. It can be seen that between 10 and 20 dB the combining gain is roughly 8 dB, close to the ideal combining gain of 8.45 dB (or a factor of 7 for a 7-channel combiner), but degrades significantly below 10 dB. However, above 20 dB the combining gain also degrades due to delay-error that dominates the phase error in the high-SNR region.

The combined signal power P_Σ for a large number of channels can be estimated via the following equation, which assumes small and equal residual phase error variance σ_θ^2 for each channel: $P_\Sigma = N + N(N-1)\exp(-\sigma_\theta^2)$. As shown in [4], this expression is exact for all *N*, and saturates as $P_\Sigma \underset{N>>2}{\cong} N^2 \exp(-\sigma_\theta^2)$ for large *N*. Note that for any *N*, the combining loss depends only on the individual channel phase error variance σ_θ^2. The combined power can be extrapolated to any number of array channels using this expression, since the phase error variance is the same in every channel of the pupil-plane system model described in this paper.

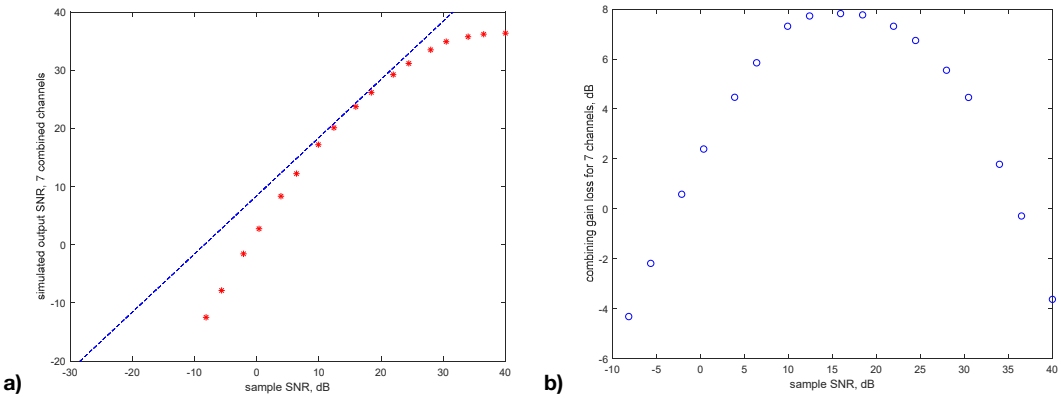

Figure 7. a) Simulated output SNR as a function of sample-SNR, in dB: ideally combined 7-channel output (dashed blue line); simulated combined output with 1-sample delay (red asterisks); b) Gain-loss in dB, as a function of sample-SNR.

VI. Coherent Detection Performance in Strong Background

Comparison of photon-counting with coherently combined PPM Modulation. Here it is assumed that phase compensation is ideal even in strong background, due to the background suppression property of the C-WFS and coherent detection, as shown in Figure 7 at sample-SNRs of 10–20 dB and small processing delay. The baseline modulation format for future Deep-Space Optical Communications is pulse-position modulation, or PPM. With this modulation format the binary information bit-stream is grouped into words consisting of $\log_2(M)$ bits, and the slot corresponding to the numerical value of the binary word is selected as the signal-slot. For example, with $M = 64$ slots, a binary word consisting of the bits (100010) translates to the number 34, hence this information word corresponds to the PPM symbol with a single laser pulse in the 34th slot. At the receiver, correct identification of the signal-slot decodes the binary word into (100010), thus recovering 6 bits of information with a single PPM symbol. Communication at a bit-rate of R bits/s requires decoding $R/\log_2(M)$ symbols per second, which means that M slots must be received every $\log_2(M)/R$ seconds. Therefore each slot is of duration $\tau = \log_2(M)/RM$ seconds. If forward error correction (FER) coding is employed with a code rate of ς , then the slot-width decreases by this same factor to accommodate the spectral overhead of FER coding, yielding $\tau_{\text{FER}} = \varsigma\log_2(M)/RM$.

With PPM signaling and background interference, the signal slot contains an average of $K_s + K_b$ photons, and the remaining $(M - 1)$ slots contain K_b photons each. With photon-counting statistics the counts are Poisson distributed, where the probability of k counts is given by $p(k) = K^k \exp(-K)/k!$ and where K is the average number of photons (signal, or background, or a combination of the two). The maximum likelihood receiver chooses the slot with the greatest photon counts as the signal slot. Ignoring equalities in maxima among the slots (a negligibly low-probability event in strong background), the probability of correct detection is equal to the probability that every non-signal slot has lower photon count than the signal slot. This must be true for every possible photon count from $k = 1$ to $k = \infty$ in the signal-slot, for correct detection to occur. With independent Poisson count

statistics in each slot, the formula for correctly detecting the PPM symbol, $P_s(C)$, becomes [3]:

$$P_s(C) = \sum_{k=1}^{\infty} (K_s + K_b)^k \exp[-(K_s + K_b)] / k! \left(\sum_{i=0}^{k} K_b^i \exp[-K_b] / i! \right)^{M-1} \quad (11)$$

The probability of symbol-error is $P_s(E) = 1 - P_s(C)$. Converting symbol-error probability to bit-error probability (BER) yields $P_b(E) = \frac{M}{2}[1 - P_s(C)] / (M - 1)$.

Computation of the BER for coherent detection employs the same approach as photon counting, except that Poisson statistics are replaced by Gaussian statistics. This yields an effective signal-to-noise ratio of $4K_s / (1 + 4\xi K_b)$ when background fields are present, and can be interpreted as corresponding to a signal mean of $\mu = \sqrt{4K_s}$ in the signal-slot, $\mu = 0$ in the noise-slots, and noise variance of $\sigma^2 = (1 + 4\xi K_b)$ in every slot. The coherently detected signal model leads to the continuous Gaussian probability density: $p(x) = \sqrt{2\pi\sigma^2} \exp[-(x - \mu)^2 / 2\sigma^2$. The maximum likelihood detection strategy is the same as for photon-counting detection, except that here we are dealing with continuous voltages modeled as Gaussian random variables: the slot with the greatest voltage is selected as the signal-slot. Correct detection occurs when the voltage in the signal-slot exceeds every other voltage in the remaining (M – 1) slots, yielding the following expression for correct detection:

$$P_s(C) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\{-[x - \mu]^2 / 2\sigma^2\} dx \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\{-(y)^2 / 2\sigma^2\} dy \right\}^{(M-1)} \quad (12)$$

The BER curves for coherent and photon-counting detection are shown in Figure 8 over a link distance of 2 AU, at a bit-rate of 2 MBPS and FER coding rate of 0.5, for PPM modulation with M = 4, 16, and 64.

This example was selected because it represents the maximum range discussed in [2], and incorporates the maximum background contributed both by atmospheric scattering and a bright planet (Mars) in the receiver's FOV. It can be seen in Figure 8 that coherent detection is 4.4 dB better than photon-counting detection with M = 64, and 6.6 dB better for M = 16 in the high background regime, due primarily to the dramatic reduction in electrical bandwidth matched to the signal bandwidth, compared to the much greater electrical bandwidth of the optical pre-detection filter.

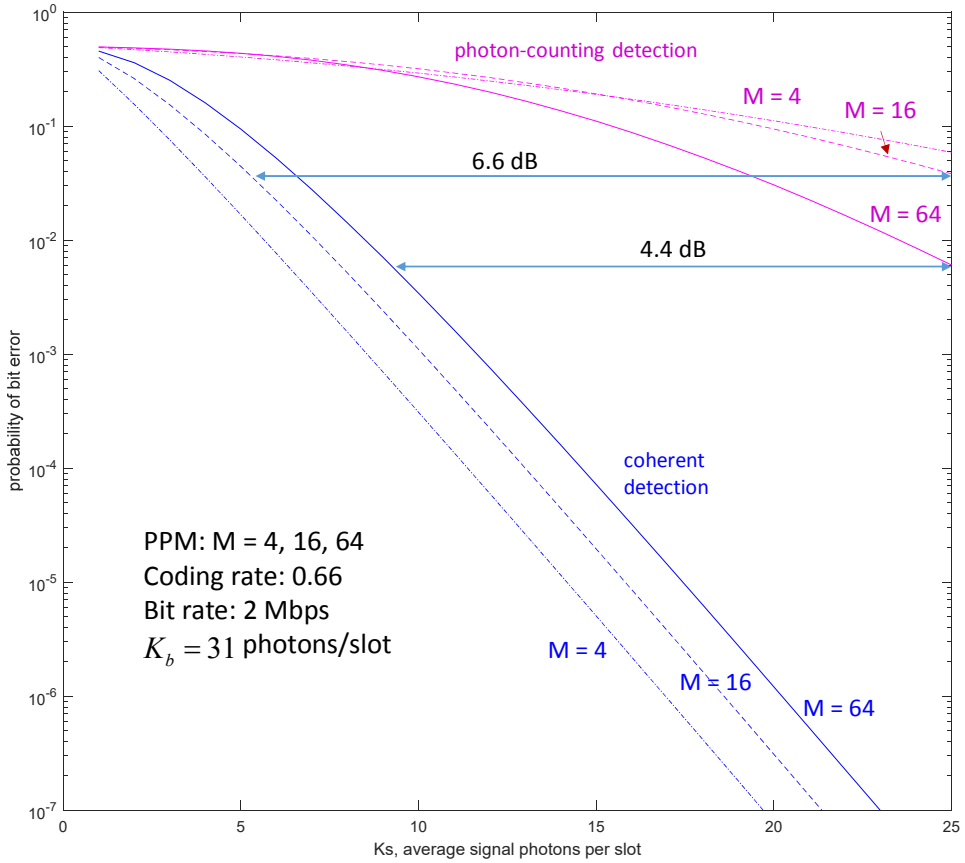

Figure 8. Comparison of coherent and photon-counting receiver BER performance for PPM signals, in the presence of strong background.

VII. Summary and Conclusions

In this article we considered the potential use of a novel adaptive optics wavefront sensor utilizing coherent optical detection. It was shown that coherent optical detection is virtually insensitive to background radiation, provided that the signal bandwidth is orders of magnitude smaller than the electrical bandwidth of the optical pre-detection filter, which is generally valid for nano-meter optical and mega-Hz modulation bandwidths.

It was shown that with ideal phase compensation, coherent detection reduces background radiation by several orders of magnitudes, yielding detection performance comparable to shot-noise limited error probabilities.

The Cramer-Rao bound on the variance of an array of phase estimates was derived, and used to validate the simulation results at high sample-SNRs. The structure of the coherent wavefront sensor was developed and its performance evaluated via simulation, using Kolmogorov phase-screens to generate realistic optical phase trajectories and taking into account processing delays. Excellent agreement with the CRB was demonstrated above a threshold of approximately 10 dB sample-SNR, validating the simulation results. It was shown that nearly ideal combining gain can be achieved above threshold even with processing delay, however the combining gain saturates above 20 dB demonstrating the

importance of minimizing processing delay in a practical implementation. This improvement is significant for Adaptive Optics applications, where the performance of conventional Shack-Hartmann wavefront sensors is known to degrade significantly in high background environments.

Finally, the BER performance of the coherent array combiner was compared to conventional photon-counting detection under high background conditions. It was shown that coherent array detection outperforms conventional photon-counting detection by as much as 6.6 dB with $M = 16$ PPM and 2 MBPS data-rate, applicable to an Earth–Mars link at maximum distance. We conclude that coherent optical array processing offers significant improvements in high-background environments, whether the coherent wavefront sensor is employed in an adaptive optics configuration, or used directly to combine the compensated optical field for deep-space communications applications.

References

[1] R. Gagliardi and S. Karp, Optical Communications, Second Edition, Wiley, NY, 1995.

[2] A. Biswas, M. Srinivasan, S. Piazzolla, D. Hoppe, "Deep Space Optical Communications," *Proc. SPIE,* 2018.

[3] V. Vilnrotter, M. Srinivasan, "Adaptive detector arrays for optical communications receivers," *IEEE Transactions on Communications,* 2002, vol. 50, issue 7, pp. 1091–1097.

[4] V. Vilnrotter, "Power Spectrum of Uplink Array Signals with Random Phase and Delay Errors," *Proceedings of IEEE Aerospace Conference,* Big Sky, MN, 2012.