The problem of fringe acquisition in high precision space-based interferometers

Carlos E. Padilla, Valeri I. Karlov, Jun Li, Hon M. Chun

Moldyn, Inc., 1033 Massachusetts Ave.
Cambridge, MA 02138

John N. Tsitsiklis

Massachusetts Institute of Technology
77 Massachusetts Ave., Cambridge, MA 02139

Robert D. Reasenberg

Harvard-Smithsonian Astrophysical Observatory
60 Garden Street, Cambridge, MA 02138

ABSTRACT

This paper presents the results of an investigation of the problem of fringe acquisition (FA) as it applies to high-precision space-based optical interferometers. The POINTS (for Precision Optical INTerferometry in Space) instrument concept, being developed in a collaborative effort between the Smithsonian Astrophysical Observatory and the Jet Propulsion Laboratory (JPL) as a wide-band space-based optical interferometer dedicated to astrometry, is used as a test baseline. In this study we analyze three candidate FA algorithms: linear correlation techniques (CT); nonlinear least squares (NLS); and, a discrete Bayes approach (DBA). Analytical methods of evaluating the probability of detection for the NLS are developed which enable a multi-parametric study of the FA problem. Among other parameters, the study examines the effects of varying the magnitude of the electronic readout noise of the focal plane detector cells (assumed to be CCD's), of increasing the bolometric magnitude of the target stars, of allowing constant drifts of the optical path difference (OPD), and of co-adding CCD cells to reduce readout noise. A signal-to-noise measure is selected that exhibits a high correlation with the FA performance and a study of the a posteriori probability density space reveals insights into the nonlinear nature of the interferometric measurement function. The study concludes with selected Monte Carlo simulations to confirm the analytical predictions and to compare the performance and robustness of the CT and NLS to those attainable with the DBA.

KEYWORDS

Fringe Acquisition, Channelled Spectrum, Spaceborne Interferometry, Nonlinear Estimation, Loss of Lock

1. PROBLEM STATEMENT

The fringe acquisition (FA) problem can be succinctly defined as the problem of estimating the initial angle that the interferometer optical axis makes with a target star after the spacecraft attitude control system (and possibly a fine pointing system) has brought the telescopes to bear on the star with a pre-specified "coarse" precision. The goal of the fringe acquisition algorithms is to successfully estimate the true off-axis angle to within a specified "finer" accuracy. This estimate can then be handed over as the initial estimate for a tracking algorithm.

In order to better understand the nature of the problem and its fundamental limitations, we start by analyzing a simplified model wherein the spacecraft/instrument ensemble is assumed static (i.e., no drift or dynamic motion of the off-axis angle). Sensor noise, readout times, and stellar intensities are allowed to vary. Three "classes" of algorithms are analyzed and implemented: Nonlinear Least Squares method; Correlation or Matched Filter techniques; and, Discrete Bayesian estimation based on a Markov state approach.

0-8194-1830-7/95/$6.00

The goal of this work is to perform a multi-parametric study of the fringe acquisition performance. The following issues are to be explored:

- Cell grouping as a tradeoff between readout noise and resolution in optical frequency
- Dynamic changes in the off-axis angle (constant or linear drift)
- Stellar bolometric magnitude
- Connection between fringe acquisition and fringe tracking (how much should uncertainty be reduced after fringe acquisition in order to start stable tracking)
- Tradeoff between optimality and robustness in the estimation algorithms

This paper puts the emphasis on the analytical evaluation of the FA nonlinear estimation problem without involving time-consuming Monte Carlo simulations, although some amount of simulation is used to validate the analytical predictions.

2. INTERFEROMETER SYSTEMS UNDER STUDY

2.1 The channelled spectrum

The baseline configuration studied for the FA problem consists of an instrument "staring" at a target star with an off-axis angle δ, related to the Optical Path Difference (OPD) by OPD = $L \cdot \delta$, where L is the instrument baseline (see Refs. 1,2). Constant or linear drift in the OPD is considered in the study. There are nominally 512 CCD detector cells per beamsplitter output port at the focal plane (there are two ports, plus and minus, for a total of 1024 CCD detector cells). The CCD detectors in the focal plane are subject to electronic readout noise. For this study it is assumed normally-distributed with zero mean and nominal standard deviation of 3 electrons per readout. To reduce the effect of electronic noise, the CCD detector cells can be grouped in larger configurations at the expense of lower resolution in optical frequency. This tradeoff will be considered in this paper.

The measurement equation for POINTS (see Refs. 1,2) has the following form accounting for the effect of loss in visibility due to OPD drift:

$$\bar{Y}_{i\pm} = E[Y_{i\pm}] = \Delta t_{CCD} n(\nu_i) \Delta\nu_i \left[1 \pm \Lambda(\nu_i) \sin\left(\frac{2\pi L \nu_i}{c} \delta \right) \chi_i^{v} \chi_i^{t} \right] \tag{1}$$

$$\text{where} \quad \chi_i^{v} = \operatorname{sinc}\left(\frac{\pi L \Delta\nu_i}{c} \delta \right), \quad \chi_i^{t} = \operatorname{sinc}\left(\frac{\pi L \nu_i}{c} \dot{\delta} \frac{\Delta t_{CCD}}{2} \right)$$

Eq. (1) is written for each cell i, where $\pm$ stands for plus and minus port respectively; Δt_{CCD} is the readout time interval; $n(\nu)$ is the envelope of the spectrum and is dependent on the stellar bolometric magnitude and stellar temperature; $\Delta\nu_i$ is the optical frequency bandwidth of cell i; Λ is the fringe visibility; ν_i is the central optical frequency at cell i; L is the interferometer baseline; c is the speed of light; and, $L\delta$ is the optical path difference (OPD). In the dynamic case, the off-axis angle δ and its time derivative $\dot{\delta}$ are related to the middle of the CCD readout interval. The rate $\dot{\delta}$ is assumed to be constant over the CCD interval. It is assumed for this study that $|\dot{\delta}| \leq 5 \times 10^{-6}$ rad/sec which would correspond to a 1 arc sec/sec drift in the Attitude Control System (ACS). The assumptions used in the derivation of the measurement equation (Eq. (1)) are valid for relatively small CCD intervals (the nominal interval is 0.01 seconds). The fringe acquisition is performed from a sequence of CCD readouts. The number of these readouts is chosen to have a detection time within 1 second.

2.2 Drift model

Define t_0 as the initial time and t_k, k = 1, ..., K as the middle points of subsequent CCD readout intervals, the number of which is equal to K and the duration of each interval is equal to Δt_{CCD}. Then the following linear equation can be used

to model the OPD dynamics:

$$\theta(t_k) = \Phi(t_k - t_0)\theta(t_0) \tag{2}$$

where θ is the vector of parameters to be estimated, and Φ is the transition matrix. Approximating the dynamic motion of the spacecraft by a second-order polynomial model (linear drift), we have:

$$\theta = \begin{pmatrix} \delta \\ \dot{\delta} \\ \ddot{\delta} \end{pmatrix}, \quad \Phi(\Delta t) = \begin{pmatrix} 1 & \Delta t & \frac{\Delta t^2}{2} \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{pmatrix} \tag{3}$$

where the off-axis angle δ, its rate $\dot{\delta}$, and constant acceleration $\ddot{\delta}$ are to be estimated. In the case of a first-order model (constant drift) Eq. (3) is simplified by excluding the acceleration component.

The vector of initial conditions $\theta_0 = \theta(t_0)$ is chosen as a vector for direct estimation. The estimates of $\theta_k = \theta(t_k)$ can be obtained from the prediction equation (Eq. (3)). Combining Equations (1), (2) and (3) we have the nonlinear regression model for estimation, which exposes the relationship between measurements and parameters to be estimated:

$$\overline{Y_i(t_k)} = \overline{Y_i(t_k, t_o, \theta_o)}, \; i = 1,\ldots,m, \quad k = 1,\ldots,K \tag{4}$$

A dynamic model based on stochastic state space equations will be used in tracking. However, for the fringe acquisition problem, the regression model is justified due to the relatively short observation interval and lower accuracy required and because it provides a convenient framework for performance analysis.

2.3 Standing assumptions

For easy reference and to highlight the scope of this paper, the standing assumptions are listed:

- Stellar temperature of 7000 deg. Kelvin
- Stellar bolometric magnitude of 10 (nominal) - Range for study: 20.0 to 10.0
- Collection mirror of 35 cm diameter
- Visibility of 0.85
- Detection probability of 0.15 in the optics (efficiency in capturing photons)
- Transmission probability of 1.0
- Optical bandwidth of 0.25 to 0.9 μm
- Nonlinear dispersion
- 512 CCD detector cells per beamsplitter output port, 1024 total
- Cell grouping enabled
- CCD readout time of 0.01 sec (nominal) - Range for study: 0.001 to 0.01
- "Normally-distributed" readout noise
- Standard deviation of readout noise of 3 electrons (nominal) - Range for study: 0 to 4
- Maximum initial uncertainty in the OPD of 2.42×10^{-6} rad
- Initial uncertainty in the drift of 5×10^{-6} rad/sec (nominal)
- Initial uncertainty in the acceleration of 1×10^{-6} rad/sec (nominal)

Under the nominal assumptions, the expected number of photons per second per port is 69700. This results in 697 photons per port per nominal readout of 0.01 sec. It is important to realize that most high optical frequency cells only receive 0 or

1 photon which would then be "buried" by the electronic CCD readout noise. The maximum number of photons expected in any cell is 8 and it occurs at low frequencies (≈ 333 THz).

Assumptions on the maximum uncertainties in the drift of the OPD and its acceleration result in an OPD change of 5 × 10^{-8} rad per one readout (Δt_{CCD} = 0.01 sec) and 5 × 10^{-6} rad during the detection time (within 1 second). This indicates that the main contribution of the OPD dynamics to the degradation of the fringe acquisition performance comes from residual errors in the estimation of the drift and acceleration during the detection time rather than from loss in visibility due to dynamic effects at a given CCD interval.

3. CANDIDATE FRINGE ACQUISITION ALGORITHMS

Three classes of algorithms are studied. The first two assume continuous noise statistics. The Nonlinear Least Squares (NLS) method allows for the tuning of the filter to the expected intensity of the sensor noise. The Matched or Correlation filter (falling under the rubric of correlation techniques or CT) is a simplification of the NLS and only assumes the form of the measurement equations.

The third and most complex algorithm studied is a discrete Bayesian estimator based on a Markov chain approach (dubbed the Discrete Bayesian Approach or DBA). This algorithm results from Bayes' conditional probability density equations and its generality allows for the treatment of discrete and non-Gaussian noise sources in an exact manner. The Markov chain framework allows extension of the DBA to the case in which the off-axis angle varies in time and evolution equations must be included in the filter equations to extract maximum performance.

3.1 Nonlinear least squares method (NLS)

The NLS method is one of the most common techniques used to fit noisy data with parametric models (see Refs. 3, 4). For the FA problem, application of the NLS consists of minimizing the following cost function with respect to θ_o :

$$F(\theta_o) = \sum_{i=1}^{m} w_i \|\Delta_i(\theta_o)\|^2 = \sum_{i=1}^{m} w_i \|Y_i - \overline{Y_i(\theta_o)}\|^2 \tag{5}$$

where

$$\begin{aligned} Y_i &= [Y_i(t_1) \ \ldots \ Y_i(t_K)] \\ \bar{Y}_i(\theta_o) &= [\overline{Y_i(t_1,t_o,\theta_o)} \ \ldots \ \overline{Y_i(t_K,t_o,\theta_o)}] \\ \Delta_i(\theta_o) &= [\Delta_i(t_1,t_o,\theta_o) \ \ldots \ \Delta_i(t_K,t_o,\theta_o)] \end{aligned} \tag{6}$$

are 1 × K vectors with K components corresponding to the K readouts. $Y_i(t_k)$ is the number of detected photons over the k-th CCD readout interval in the *i*-th cell; $\overline{Y_i(t_k,t_o,\theta_o)}$ is the expected mean number of photons over the k-th CCD readout interval in the i-th cell for the parameter vector θ_o under a nominal model (see Eqs. 1-4); $\Delta_i(t_k, t_o, \theta_o)$ is the *i*-th residual $Y_i(t_k) - \overline{Y_i(t_k,t_o,\theta_o)}$ and m is the number of sensor cells.

Additionally, in Eq. (5) the notation $\|x\|^2$ stands for the 2- norm of the vector x, i.e. $\|x\|^2 = xx^T$; w_i is a weight associated with the cost of the i-th residual $\Delta_i(\theta_o)$ (assumed to be equal for all readouts). The following weighting function was chosen as a baseline:

$$w_i = (N_i + \sigma^2_{noise})^{-1} \tag{7}$$

where $N_i = \overline{Y_i(0)}$ is the Poisson variance (or spectrum in the i-th cell) when the OPD is zero, and σ^2_{noise} is the variance of the discrete Gaussian electronic readout noise. This weighting does not depend on the OPD to be estimated.

3.2 Correlation techniques (CT)

The CT can be considered a numerical simplification of the NLS. In the case under study the CT cost function is:

$$F(\theta_o) = \sum_{i=1}^{m} Y_i f_i^T(\theta_o) \tag{8}$$

where:

$$\begin{aligned} Y_i &= [Y_i(t_1), \ldots, Y_i(t_K)] \\ f_i(\theta_o) &= [f_i(t_1, t_o, \theta_o), \ldots, f_i(t_K, t_o, \theta_o)] \end{aligned} \tag{9}$$

are again 1 × K vectors each component of which corresponds to the k-th readout. $Y_i(t_k)$ is defined as above, and $f_i(t_k, t_o, \theta_o)$ is the correlation function for the i-th cell evaluated at the vector of initial conditions θ_o and corresponding to the k-th readout. The correlation function assumes the form of the measurement equation, i.e., $f(\cdot) = \bar{Y}(\cdot)$.

The maximization of the CT cost instead of the minimization of the NLS cost is justified when the quadratic term $\sum_i |\overline{Y_i(\theta_o)}|^2$ is approximately the same for all θ_o. This is true for large values of the first component of the vector θ_o, the off-axis angle δ_o. However, when the correlation function $f_i(\theta_o)$ is nearly linear in one of the components of θ_o (e.g., for the small values of the drift $\dot{\delta}_o$ and acceleration $\ddot{\delta}_o$, or in the case of small values for δ_o) neglecting the quadratic term leads to the degenerated problem of maximizing a linear cost. In other words, the CT estimator is effective in handling nonlinearities and loses its detection capabilities in the linear range. This property can be used to determine at what level of accuracy the fringe acquisition problem turns into a fringe tracking problem.

3.3 Discrete Bayes approach (DBA)

The Discrete Bayes Approach (DBA) provides the possibility of solving the simplified POINTS estimation problem under study without invoking any approximations. The DBA is capable of handling the nonlinearity of the measurement model and non-Gaussian statistics (i.e., Poisson distribution of photon arrivals and discrete Gaussian electronic readout noise). The resulting solution allows one to explore the limit in interferometric accuracy and assess how other approximation-based techniques compare with this ultimate limit. The main idea behind the DBA is to interpret the estimation problem in terms of numbers of photons detected rather than treating electron counts as continuous variables. The mathematical framework for the realization of this approach in the general case is the fundamental Bayes formula in its different forms (see Refs. 4-6) and the theory of discrete processes (Markov chains) (see Ref. 7). The estimation of θ_o is defined as the process of finding the $\hat{\theta}_o$ that is most likely given the measured data and the known statistical models for θ_o and the measurements. First, form hypotheses by discretizing the space of θ_o. Let hypothesis H_j represent the event that the vector of true parameters $\theta_0 = (\delta_o\ \dot{\delta}_o\ \ddot{\delta}_o)^T$ is inside a cube:

$$H_j = \{\theta_o : \theta_o^{(j-1)} \le \theta_o \le \theta_o^{(j)}\} \tag{10}$$

where the vectors $\theta_o^{(j-1)}$ and $\theta_o^{(j)}$ define the vertices of the cube. On the set H, define probabilities: $P = \{P_1, \ldots, P_n\}$. We will distinguish between *a priori* and *a posteriori* probabilities. The *a priori* probabilities $P^* = \{P_1^*, \ldots, P_n^*\}$, $P_j^* = P(H_j)$

are available before taking new measurements, e.g., from the previous readout. The DBA estimation algorithm transforms these probabilities into *a posteriori* ones $\hat{P} = \{ \hat{P}_1, \ldots, \hat{P}_n \}$, $\hat{P}_j = P(H_j/A)$, by taking into account the data (event A) and the problem's statistics.

The transformation is based on Bayes formula:

$$P(H_j/A) = \frac{P(H_j) \cdot P(A/H_j)}{\sum_{k=1}^{n} P(H_k) \cdot P(A/H_k)}, \quad j=1, \ldots, n \tag{11}$$

To evaluate Eq. (8), one has to compute the probabilities $P(A/H_j)$. The elaborate formulas for those probabilities are based on Poisson and Gaussian laws as well as on the nonlinear model of the interferometric measurement function (see Refs. 8, 9).

The implementation of the DBA is based on the genetic Markov chain: First, set a low-resolution net in the space of parameters and calculate the *a posteriori* probabilities of the hypotheses H_i (see Eq. (10)) using the DBA algorithm; once the search is localized, redefine a higher resolution net within the selected hypotheses. For bright stars the probability of detecting the true hypothesis will be close to one and it will be possible to localize further searches, i.e., the genetic chain has only one direction for evolution. In the case of dimmer stars, it is possible to continue the estimation process in a few directions (branches of the genetic tree). This is the so-called multiple-hypotheses tracking. At a certain level of the star dimness, the lack of information can be so significant that it will be of no use to increase the initial low resolution.

It is possible to precalculate the $P(A/H_j)$. This information can then be stored in the so-called detection matrix S for each readout: S_{kj} (where $k = (q - 1) \times m + i$; $q = 0, \ldots, N$; $j = 1, \ldots, n$; $L = (N + 1) \times m$) is a probability of detection of the event #q (corresponding to the number of photons) in the cell i when the parameter vector θ_o satisfies the hypothesis j. The detection matrix S directly accounts for the discrete nature of the interferometer measurement. After the (integer) numbers of photons are detected in each cell, it is possible to take the corresponding elements of S and use them in the Bayes formula (see Refs. 8, 9).

A formulation of the DBA that is robust to the uncertainties in model parameters and statistical characteristics can be developed by implementing a general philosophy in designing robust algorithms. It is based on simple averaging of the solution mechanism over the uncertainties in the problem's characteristics. In this case the detection matrix S should be averaged over the uncertain parameters. This can be done either analytically using perturbations techniques or numerically (through numerical integration for a small number of parameters or through Monte Carlo simulation). Physically, this averaging makes the detection matrix S "flatter" or "less tuned". In the case of complete uncertainty all elements of S will be equal, and the DBA will lose its detection capabilities. In practice it is necessary to find the best tradeoff between optimality and robustness, depending on how much uncertainty is actually present in the model.

Implementation of the DBA for processing measurement information from multiple readouts can be performed by using Bayes formula recursively. In this case, the *a posteriori* probabilities $\hat{P} = P(H_j/A)$ obtained at the previous readout are used as the *a priori* probabilities at the next readout.

4. ANALYSIS OF FRINGE ACQUISITION PERFORMANCE

4.1 Limitations of the Cramer-Rao bound

In the framework of the NLS it is possible to estimate the lower bound for the variance of the estimation error. This bound was derived by Cramer and Rao (see Ref. 5). In the present case of multi-parametric estimation, it can be written for the error covariance matrix:

$$P_\theta = E[(\theta_o - \hat{\theta}_o)(\theta_o - \hat{\theta}_o)^T]$$
$$\geq [S_\theta + \sum_{k=1}^{K} C_k^T(\hat{\theta}_o) \cdot (K_{\Delta_k}(\hat{\theta}_o))^{-1} \cdot C_k(\hat{\theta}_o)]^{-1} \quad (12)$$

where $\hat{\theta}_o$ is the estimate for the vector of initial conditions $\theta_0 = (\delta_o \dot{\delta}_o \ddot{\delta}_o)^T$, $C_k(\hat{\theta}_o)$ is the $m \times 3$ matrix of first-order sensitivities at the k-th readout:

$$C_k(\hat{\theta}_o) = \overline{\frac{\partial Y(t_k, t_o, \theta_o)}{\partial \theta_o}}\Big|_{\theta_o = \hat{\theta}_o} \quad (13)$$

$K_{\Delta_k}(\hat{\theta}_o)$ is the covariance matrix of measurement errors at the k-th readout:

$$K_{\Delta_k}(\hat{\theta}_o) = \text{diag}\{\overline{Y_i(t_k, t_o, \hat{\theta}_o)} + \sigma^2_{noise}\} \quad (14)$$

S_θ is the inverse *a priori* covariance matrix with zero off-diagonal elements and diagonal [0 D_2^{-1} D_3^{-1}] where D_2 is the *a priori* variance of the rate $\dot{\delta}$ and D_3 is the *a priori* variance of the acceleration $\ddot{\delta}$. The introduction of the matrix S_θ is needed as a regularization procedure to provide full rank to the matrix on the right-hand side of Eq. (10) starting from the first readout. The variances D_2 and D_3 should correspond to the expected level of uncertainty in the drift and acceleration.

The Cramer-Rao lower bound establishes the limit in accuracy which can be achieved by using the NLS under condition of Gaussian error statistics. Therefore, the art in designing a nonlinear filter will be in the ability to stay close to this accuracy limit. It is worth noting that an estimator which accounts directly for non-Gaussian statistics (Poisson distribution, discreteness of detected photons, etc.) can provide accuracy better than what the Cramer-Rao lower bound predicts. In this paper we will exploit the fact that the Cramer-Rao bound practically coincides with the actual covariance matrix P_θ when the estimation errors are localized in the vicinity of the estimate $\hat{\theta}$. In this case Eq. (12) is also called the Fisher information matrix (see Refs. 3-6).

4.2 Global nonlinear mechanism of loss of lock

In order to go beyond the Cramer-Rao bound in the analysis of FA performance, we first illustrate the global nonlinear mechanism of detecting the OPD. Figure 1 represents the cost of all considered techniques (NLS, CT, DBA). All three cost functions are linearly transformed into one-dimensional maximized functions to allow a qualitative comparison. In different realizations the cost functions are random (with larger variations for a shorter CCD interval). In Figure 1, for $\Delta t_{CCD} = 0.003$ [sec], we plot three cost functions averaged over 2000 realizations. Figure 1 reveals approximately the same locations of the primary and secondary peaks in the NLS, CT and DBA cost functions. These correspond to the most probable candidates for estimates of δ. The structure revealed is inherently nonlinear: the existence of two "mirror image" sets of similar spectra for the values of the off-axis angle under consideration. The first set groups symmetrically around the true $\delta = \delta^*$ with the maximum in the center and decaying peaks at both sides. The second set groups around the "mirror image" point, $\delta = -\delta^*$; but relative locations of the peaks are switched with the minimum points from the first set. Figure 2 demonstrates the physics involved by showing the information rate (spectra multiplied by squared optical frequency) for four δ values: true (1.12e-6), "mirror" (-1.12e-6), "left mirror" (-1.33e-6), and "right mirror" (-1.08e-6). The information rates of the true and "mirror" δ's significantly differ from each other. This facilitates discrimination between the corresponding δ values. On the other hand, the information rates of the "left mirror" and "right mirror" δ's very much resemble that of the true δ, particularly in the most informative low optical frequency sensor cells where the signal to noise ratio per detector cell is highest (maximum of information rate). This makes the detection process more difficult for low SNR situations.

4.3 Transformation of the NLS cost for analysis

Consider two NLS costs evaluated at the true parameter θ_o^* and a free parameter θ_o (see Eq. (5)):

$$F(\theta_o^*) = \sum_{i=1}^{m} w_i \| Y_i - \overline{Y_i(\theta_o^*)} \|^2$$
$$F(\theta_o) = \sum_{i=1}^{m} w_i \| Y_i - \overline{Y_i(\theta_o)} \|^2 \qquad (15)$$

For the *a priori* analysis, consider the ensemble of all possible realizations of the measurement vector Y_i:

$$Y_i = \overline{Y_i(\theta_o^*)} + \Delta_i \qquad (16)$$

where the residuals $1 \times K$ vectors Δ_i combine for form the random $m \times K$ matrix Δ (measurement error matrix for all cells and readouts). All elements of this matrix have zero mean, are statistically independent, and the variance of the [k, i]-th element is the following (see Eq. (14)):

$$\sigma^2_{\Delta_{ki}}(\theta_o^*) = \overline{Y_i(t_k, t_o, \theta_o^*)} + \sigma^2_{noise} \qquad (17)$$

Consider the difference of the two NLS costs (Eq. (14)):

$$\Delta F(\theta_o^*, \theta_o) = F(\theta_o^*) - F(\theta_o) \qquad (18)$$

Substituting Eq. (16) into the NLS costs and doing some algebra, we obtain the following form of Eq. (18):

$$\Delta F(\theta_o^*, \theta_o) = 2\sum_{i=1}^{m} w_i (\overline{Y_i(\theta_o)} - \overline{Y_i(\theta_o^*)}) \cdot \Delta_i^T - \sum_{i=1}^{m} w_i \| \overline{Y_i(\theta_o)} - \overline{Y_i(\theta_o^*)} \|^2 \qquad (19)$$

It is important to note for the analysis that follows that $\Delta F(\theta_o^*, \theta_o)$ is linear in the random factors Δ despite the quadratic cost of the NLS. Equation (19) can be easily generalized for the case when there are various mismatches between the model and the true system (e.g., in spectrum visibility, noise characteristics, etc.). All model uncertainties can be included now in the random factors Δ.

4.3.1 Calculation of the probability of detection

The NLS performance cost attains its minimum value at the vector of true parameters θ_o^*, so that the probability of detecting the true parameters θ_o^* value instead of some parameters θ_o, is:

$$p(\theta_o^*,\theta_o) = \Pr\{\Delta F(\theta_o^*,\theta_o) \le 0\} \tag{20}$$

In practice we are interested in evaluating the probability of detecting the true parameters θ_o^* within a given range of accuracy:

$$p_d(\theta_o^*) = \Pr\{|\hat{\theta}_o - \theta_o^*| \le \bar{c}\} \tag{21}$$

where $\hat{\theta}_o$ is the estimate for θ_o and $\bar{c}$ is the vector specifying the criterion of detection. It should be noted that in the fringe acquisition problem it is enough to consider only the first component of the vector θ_o, i.e. the off-axis angle δ_o, since the nonlinearities in this direction are the most significant.

It is clear that the event $D = \{\hat{\theta}_o: |\hat{\theta}_o - \theta_o^*| \le \bar{c}\}$ can be considered an intersection of the events R_q, q = 1, ..., N where:

$$R_q = \{\theta_o^{(q)}: \Delta F(\theta_o^*,\theta_o^{(q)}) \le 0, |\theta_o^{(q)} - \theta_o^*| > \bar{c}\} \tag{22}$$

and the number of all possible free parameters $\theta_o^{(q)}$ is sufficiently large. In other words, to detect the true parameter θ_o^* we must not mix it up with any free parameter outside the interval $\{\theta_o: |\theta_o - \theta_o^*| \le \bar{c}\}$.

The events R_q are not independent since the cost difference $\Delta F(\theta_o^*, \theta_o^{(q)})$ contains the same random factor Δ for different q. This significantly complicates the statistical analysis. Introduce for brevity the vector:

$$Z = \begin{pmatrix} \Delta F(\theta_o^*,\theta_o^{(1)}) \\ \vdots \\ \Delta F(\theta_o^*,\theta_o^{(N)}) \end{pmatrix} \tag{23}$$

The vector Z is Gaussian and its mean m_z and covariance matrix P_z can be easily calculated from Eqs. (17) and (19).

After the statistical characteristics of the vector Z are calculated, the detection probability $p_d(\theta_o^*)$ can be evaluated as the solution of the following probabilistic problem:

$$p_d(\theta_o^*) = \Pr\{Z \le 0\} = \Pr\{\max_q Z_q \le 0\},\ Z \varepsilon N\{m_Z, P_Z\} \tag{24}$$

A tractable formula for the detection probability $p_d(\theta_o^*)$ can be obtained by taking advantage of the interferometer physics. Combining "global" and "local" events (which are mutually dependent), the probability of detection can be determined as:

$$p_d(\theta_o^*) = P_z\{Z \le 0, |\Delta\hat{\delta}_o| \le \bar{c}_\delta\} \tag{25}$$

where $\Delta\hat{\delta}_o$ is the "local" estimation error, and $\bar{c}_\delta$ is the first element of the vector $\bar{c}$ in Eq. (21).

Now, if we introduce the extended vector

$$Z_\Sigma = \begin{pmatrix} Z \\ \Delta\hat{\delta}_o \end{pmatrix} \tag{26}$$

with the computed statistics and assume it has Gaussian characteristics (invoking the central limit theorem), the probabilistic problem can be solved by using the techniques from Ref. 10. Note that for the "global" effects, it is enough to consider 10 to 12 "remote points" (i.e., N = 10 or 12, see Ref. 9).

4.4 A signal to noise ratio measure for the channelled spectrum

The statistical analysis of the fringe acquisition performance is accompanied by a simpler Signal-to-Noise Ratio (SNR) analysis. After some numerical experiments, we defined the following SNR measure:

$$SNR = \frac{\overline{Y_s(k)}}{\sqrt{\overline{Y_s(k)} + k\sigma_{noise}^2}}, \quad s = \arg\max_i v_i^2 \overline{Y_i(k)} \tag{27}$$

where k is the current number of readouts, $\overline{Y_i(k)} = \sum_{j=1}^{k} \overline{Y_i(t_j)}$ is the number of expected photons in the i-th cell after k readout intervals, and σ_{noise} is the Standard Deviation (STD) of the electronic readout noise. The maximum operation in Eq. (27) emphasizes the most informative part of the spectrum ($v_i^2 \bar{Y}_i$ being defined as the information rate at cell i.)

The correlation of the simple SNR analysis with more elaborate statistical analysis and Monte Carlo simulation is of great interest. It turns out (see the simulation results, Section 5) that the proposed SNR exhibits a high level of correlation with the FA performance (probability of detection) when k = 1. For k > 1, however, the magnitude of the SNR is overly conservative, i.e., the given probability of detection actually can be reached with a lower SNR. The reason for this is the fact that estimation techniques utilize more information (by processing k independent measurements separately) than the SNR analysis which accounts only for the integral effect. It is possible to correct for this fact empirically by multiplying the SNR with a function of the number of readouts, $f(k)=k^\alpha$, where the index α depends on the detection criterion.

5. PARAMETRICAL STUDIES

All calculations of the performance characteristics are based on using analytical formulas with N = 10 (see Section 4). It should also be noted that the computational expense for the statistical analyses is reduced 2 to 3 orders of magnitude through the use of the NLS-based analytical predictive tool by eliminating the need for Monte Carlo runs.

Figure 3 presents the probability of detection as a function of m_{bol} and Δt_{CCD}. As can be seen, reliable detection from one readout is possible with Δt_{CCD} close to 0.01 *sec* and m_{bol} = 10-11. An increase in m_{bol} dramatically (in an exponential way) impedes the detection process.

Figure 4 illustrates the dependence of FA performance on the off-axis angle δ. First, it was encountered that in the situation when there are enough photons to localize the detection errors in the vicinity of the true OPD, the FA performance

(in terms of quantile† corresponding to the probability 0.97) is the worst in two points symmetric about the origin $\delta_1 = -0.6784 \times 10^{-7}$ rad and $\delta_2 = -0.6784 \times 10^{-7}$ rad (see Fig. 4). There is a simple physical explanation of this fact. In Figure 4 the measurement function (for a few cells in the negative port) and its first-order sensitivities are plotted. It can be seen that at the point δ_1 the sensitivities in all cells are close to 0; at the point δ_2 the situation is even worse since the measurement function itself is also close to 0 (for the positive port the points δ_1 and δ_2 will be switched). For other values of the off-axis angle the balance of low and high sensitivities is approximately equal for different cells. The low sensitivities in the presence of sufficiently intensive noise is a serious impediment for improving detection accuracy. Performance deteriorates with increasing readout noise intensity, as expected.

It is interesting to note that the sensitivities in Figure 4, which are exactly equal to 0, do not correspond to the cells with the lowest frequencies (where the average photon rate with $\delta = 0$ is maximal). There is a shift to higher frequencies with the exact "zero" point occurring around 500 THz. However, if we plot the information rate (measurement function weighted with ν_i^2 where ν_i is optical frequency), the zero sensitivities correspond exactly with the most informative photon rate at δ_1 and δ_2.

The situation where there is a lack of detected photons (e.g., for $m_{bol} = 11.2$) has provided an additional physical interpretation on the dependence of the FA performance on the OPD. Figure 5 shows the probability of detection within 1×10^{-8} and 1×10^{-7} rad. For the criterion of detection 1×10^{-8} rad the FA performance is again worst at the points δ_1 and δ_2. But, from analyzing the "global" and "local" probabilities (see Figure 5) it follows that the drop in the FA performance is caused by local effects (due to low sensitivities). At the same time, the "global" probabilities are the largest at the points δ_1 and δ_2. So, for the criterion of detection 1×10^{-7} rad, when global effects dominate, the FA performance is the best at the points δ_1 and δ_2. As we shall see, this result correlates well with the SNR analysis which accounts for the number of detected photons but not for the sensitivities with respect to OPD.

Another parameter for study is the number of cells grouped or co-added before readout. Cell grouping can be considered a technical measure to reduce the effect of readout noise. In effect, this is equivalent to having larger cells at the focal plane output ports. For simplicity, we consider the situation when the cells' configuration is represented by a single scalar parameter - the number of original cells grouped into one "enlarged" cell. Figure 6 takes a look at the FA performance (quantile with probability 0.97) for different OPD's and number of cells grouped. It is typical, for the given nominal parameters, that up to 10 cells can be grouped into one cell providing a significant noise reduction without involving noticeable loss in optical resolution. For remote OPDs the FA performance gets degraded already after grouping more than 10 cells while for OPDs close to the origin the cell grouping is more "flexible".

5.1 Correlation of FA performance with SNR analysis

Figure 7 summarizes all sets of parameters studied (including ones not mentioned in the paper) to reveal a close correlation between the FA performance (probability of detection for the criterion 1×10^{-7} rad) and the SNR. The parameter sets are split into four groups: multiple readouts (o), different STDs of electronic noise (+), cell grouping (*) and everything else (x). It was suggested in Section 4.4 that the SNR (Eq. (27)) should be multiplied by k^{α} where k is the number of readouts and $\alpha = -0.08$ (for criterion of detection 1×10^{-7} rad). This correction is already done in Figure 7. It is worth noting that for parameter sets with different cell grouping schemes there are "nonlinear features" in the correlation which result from the fact that the SNR analysis is incapable of predicting performance degradation due to a loss in optical resolution. The least correlated sets of runs combine the effects of the multiple readouts with uncertainties in OPD dynamics. For a detection criterion of 1×10^{-8} rad, effects due to sensitivities and estimation mechanisms destroy the nice correlation with SNR exhibited by the 1×10^{-7} rad criterion.

5.2 Connection Between Fringe Acquisition and Fringe Tracking

The results of the previous sections suggest that the criterion of detection within 1×10^{-7} rad has certain advantages over

† Quantile is defined as the rms accuracy achieved by a certain percentage of the samples, e.g., 0.97 quantile corresponds to about 2.17 σ in the Gaussian approximation.

the criterion within 1×10^{-8} rad. Detection within 1×10^{-7} rad not only shows good correlation with the SNR analysis for FA, but allows for the use of low resolution gridding in the FA search algorithm. This offers a convenient way to separate the FA and fringe tracking (FT) tasks. For relatively bright stars, and using this criterion of detection, it is not necessary to account for drifts, uncertainties in the spectrum, statistics etc. in the FA problem. Furthermore, a simple FA algorithm (e.g., CT) can be used involving only one-parameter gridding.

Fringe acquisition can be considered an initial step to "localize" nonlinearities and start a high-precision refinement of the estimates by a tracking algorithm. Note that for the case of very dim stars there is little distinction between the fringe acquisition and tracking problems. In this case the whole process can be considered as multiple-hypotheses tracking and the DBA can be used as a framework to design a fringe tracker (Ref. 11).

6. MONTE CARLO SIMULATIONS

In this memo the purpose of Monte Carlo simulation is three-fold: 1) to verify the results of the analytical predictions of the FA performance; 2) to compare different estimation techniques (NLS, CT, and DBA); and 3) to study robustness issues. We limit ourselves to the set of parameters when the electronic noise is present, its statistical characteristic σ_{noise} is uncertain (20% of the nominal) and there are uncertainties in the spectrum (10%) and visibility (20%) as well. The latter two uncertainties are modelled as Gaussian independent variables for each cell. Figure 8 demonstrates that the analytical prediction of the FA performance corresponds quite accurately to the actual FA performance of the NLS obtained through Monte Carlo simulation. The DBA can provide a significant improvement of the FA performance. For example, a star can be up to one bolometric magnitude dimmer, but the probability of its detection by the DBA is the same as by the NLS or CT. In its turn, the NLS outperforms the CT due to its robustness and excellent filtering properties (within Gaussian approximation).

7. CONCLUSIONS

In this paper three methods of solving the Fringe Acquisition (FA) problem have been studied: CT, NLS, and DBA. Selected Monte Carlo runs were performed to study the performance attainable using all three selected techniques as well as their robustness.

The DBA outperforms the other two methods because it directly accounts for the discrete Poisson distribution of photon arrivals. This results in its ability to observe dimmer (at about 1 bolometric magnitude) stars with the same performance as the NLS. In turn, the NLS outperforms the CT due to exploiting not only a "correlation" mechanism of estimation but also the "filtering" mechanism. As can be seen, the "filtering" mechanism provides high robustness properties for the NLS in the case when there is a mismatch in the statistical characteristics of the readout noise. The robustness is high both in terms of stability (i.e., not losing lock) and performance. The results presented also indicate that the DBA can be efficiently robustified to uncertainties in the statistical characteristics of the electronic readout noise.

A method of analytical prediction of the FA performance was developed within the framework of the NLS . This method accounts for the physical nonlinear mechanism of fringe detection which consists of both "global" and "local" effects. The method calculates the probability of detecting the "true" OPD within the specified interval by analyzing the multiple peaks of the *a posteriori* density of probabilities. The STD of the "local" error is computed through the conventional covariance relations (Cramer-Rao bond). The accuracy of the analytical predictions was verified by Monte Carlo simulations.

The multi-parametric studies performed using the NLS-based analytical tool made possible a number of principal observations. Two symmetrical regions near zero OPD were identified where loss of sensitivity in the cells with the highest information content leads to reduced fringe acquisition accuracy. At the same time, those regions can be detected with the highest probability (within 1×10^{-7} rad) due to their unique spectra. The multi-parametric studies also led to the conclusion that dynamical effects (e.g., drift in OPD) can be ignored for the bright star Fringe Acquisition task and can relegated to the Fringe Tracking stage. Finally, a signal-to-noise ratio (SNR) measure was devised which exhibits a high degree of correlation with the probability of fringe acquisition (within 1×10^{-7} rad).

8. ACKNOWLEDGMENTS

This work has been conducted under NASA/JPL Phase II SBIR contract #NAS7-1248, entitled Estimation Tools for Precision Interferometry.

9. REFERENCES

1) Reasenberg, R.D., Babcock, R.W., Noecker, M.C., Philips, J.D.
"POINTS: the First Small Step" (Invited Paper), Proceedings of the SPIE on Spaceborne Interferometry, Orlando, Florida, April, 1993.

2) Padilla, C.E., Chun, H.M., Matson, L., Reasenberg, R.D.
"Fringe Tracking Filters for Space-Based Interferometers", Proceedings of SPIE on Spaceborne Interferometry, Orlando, Florida, April 1993.

3) Gelb, A., Editor
Applied Optimal Estimation, The MIT Press, Cambridge, Massachusetts, 1989.

4) Cramer, H.
Mathematical Methods of Statistics, Princeton, Princeton University Press, 1946.

5) De Groot, M. H.
Optimal Statistical Decisions, McGraw-Hill Book Company, New York, 1970.

6) Walkup, J.F., Goodman, J.W.
"Limitations of Fringe-Parameter Estimation at Low Light Levels," Journal of the Optical Society of America, Vol. 63, no. 4, 1973.

7) Meyn, S.P., Tweedie, R.L.
Markov Chains and Stochastic Stability, Springer-Verlag, New York, 1993.

8) Padilla, C.E., Li, J.
"The Problem of Fringe Acquisition in High Precision Space-based Interferometers: Part I - Static Acquisition", Technical Memorandum, MOM394, ETPI TM194, September, 1994.

9) Padilla, C.E., Karlov, V.I.,
"The Problem of Fringe Acquisition in High Precision Space-based Interferometers: Phase II - An Exploration of the Parameter Space", Technical Memorandum, MOM1894, ETPI TM294, December, 1994.

10) Malyshev, V.V., Krasilçhikov, M.N., and Karlov, V.I.
Optimization of Observation and Control Processes, AIAA Education Series, 1992.

11) Padilla, C.E., Karlov, V.I., Tsitsiklis, J.N., Li, J., Reasenberg, B., Chun, H.M.
"A Study of fringe tracking for high precision space-based interferometers", Paper # 2477-07, SPIE Conference, Orlando, FL, April 1995.

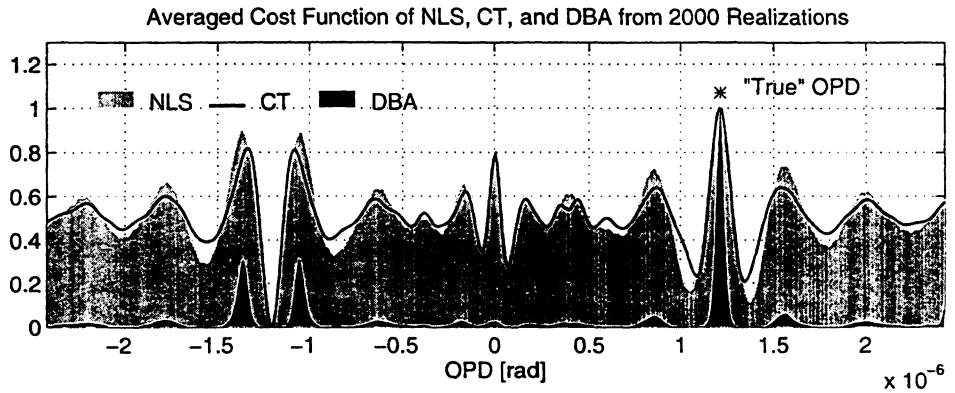

Figure 1 Global Nonlinear Mechanism of Detection

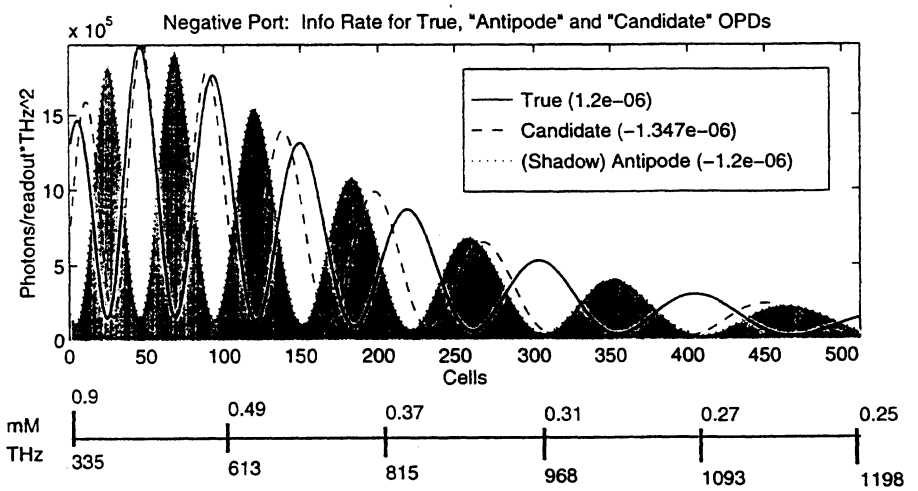

Figure 2 Information Rates at Different OPDS

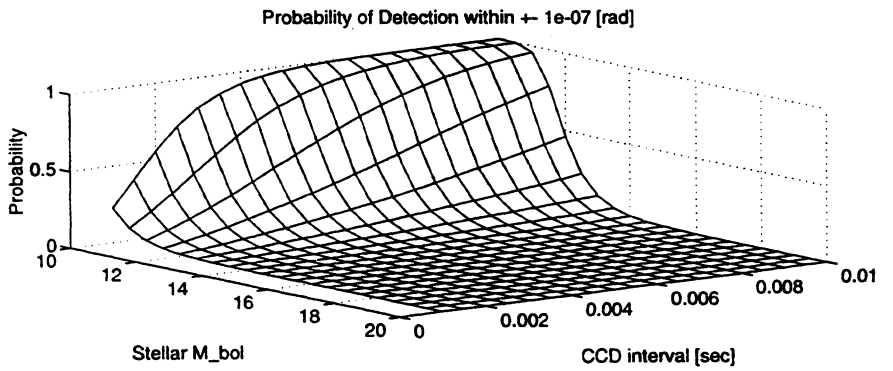

Figure 3 Dependence of Detection Probability on Stellar Magnitude and CCD Readout Time

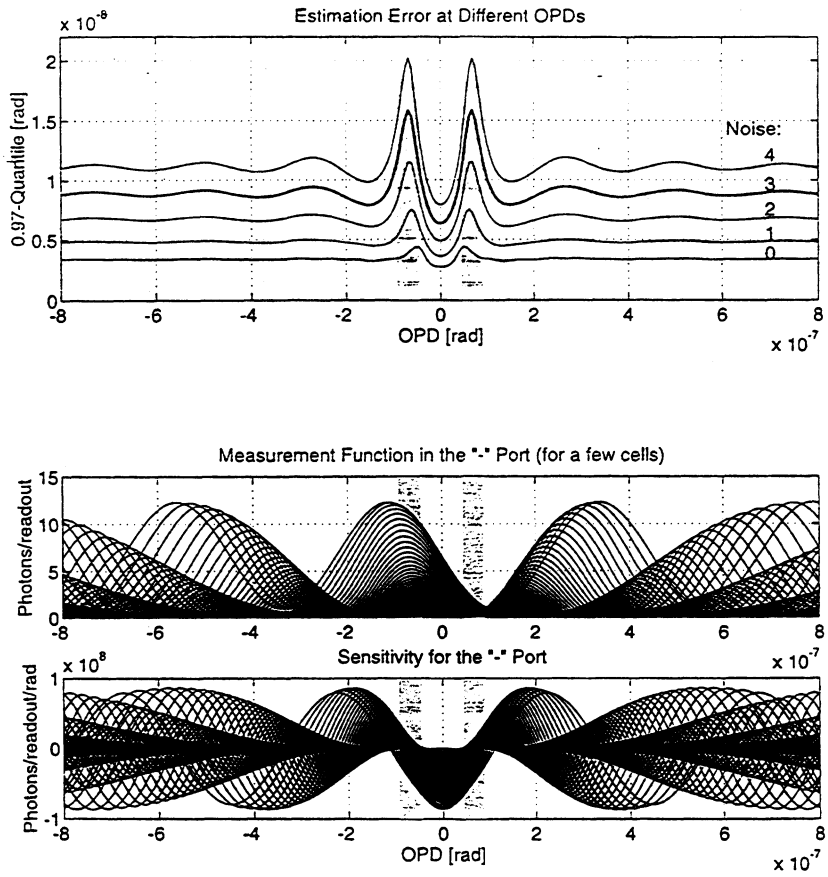

Figure 4 Dependence of Detection Accuracy on OPD

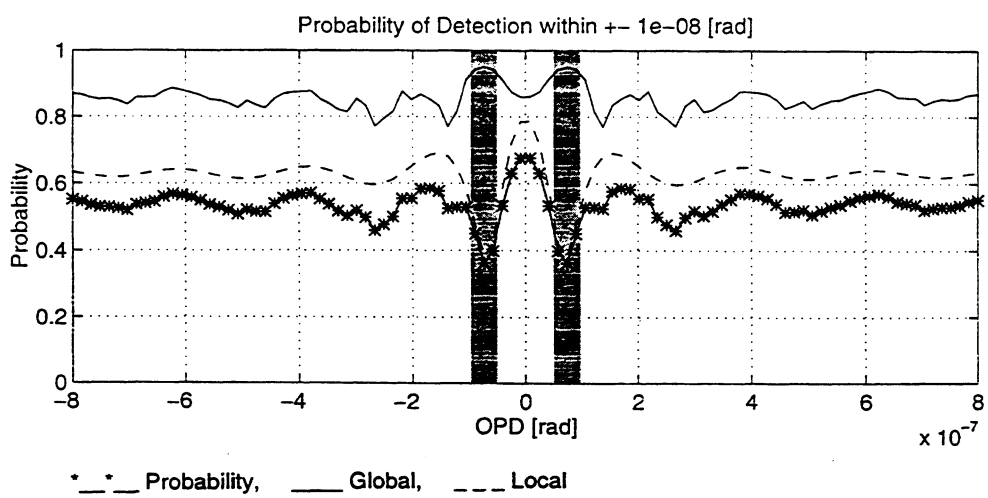

Figure 5 Dependence of Detection Probability on OPD

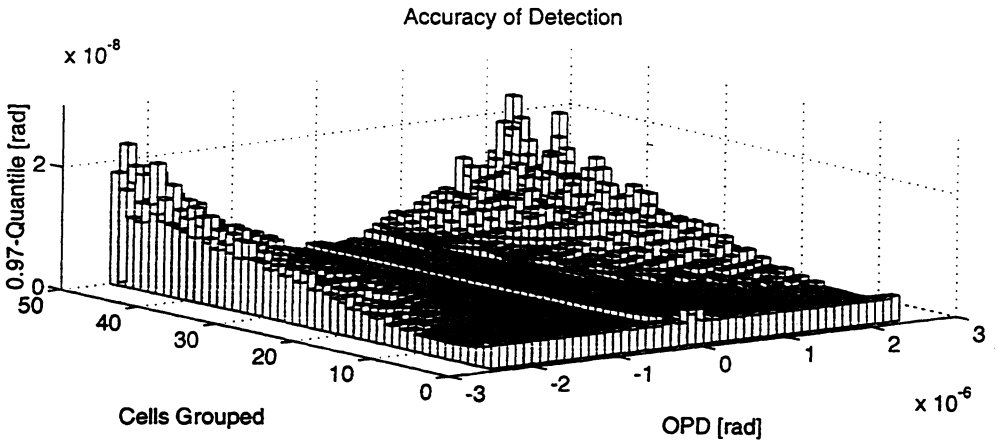

Figure 6 Dependence of Detection Probability on OPD and Cells Grouping

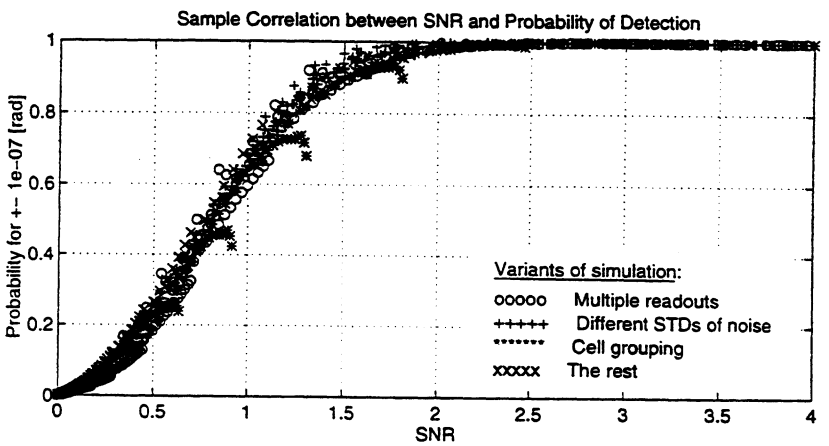

Figure 7 Correlation of Detection Probability and SNR

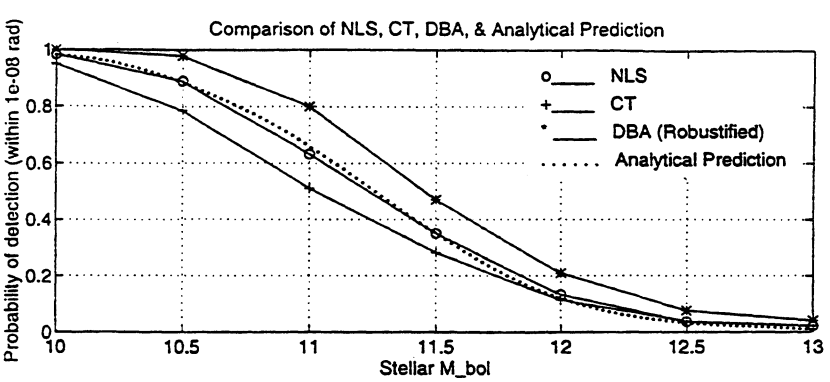

Figure 8 Monte Carlo Simulations and Analytical Predictions