KAON 790

SPIE Astronomical Telescopes and Instrumentation
June 27-July 2, 2008, San Diego, CA

Wavefront control algorithms for the Keck next-generation adaptive optics system

Donald Gavel and Marc Reinig

University of California Observatories, UC Santa Cruz, 1156 High Street, Santa Cruz, CA, USA 95064

ABSTRACT

The Keck Next Generation Adaptive Optics (KNGAO) system promises to yield high-Strehl observations over a wide range of science wavelengths from the optical through the infrared. We describe the algorithms proposed for a Real-Time Controller (RTC) implemented in a massive parallel processor environment. These algorithms take advantage of the Fourier domain to speed up processing and ensure minimum variance control that incorporates prior as well as current data. We present the unique approach to the design that enables such a complex tomography processor to scale more favorably with telescope aperture size than the more traditional RTC approaches.

Keywords: Real-Time Control, Tomography

1. INTRODUCTION

The real-time controller (RTC) has the overall task of taking raw camera data from multiple wavefront sensors and determining the commands to place on multiple wavefront correctors. The processing of data and creation of commands is to happen within time-slices determined by the wavefront sensor frame sample period t_s with the wavefront sensors acting as the master reference clock on the cycle. For the Keck Next Generation Adaptive Optics (NGAO) system, we have designed the system for operation on a period of $t_s = 500$ μs.

The RTC algorithm must perform the following activities:

- Process the raw camera data so that its values are proportional to photocounts
- Perform a reconstruction process that estimates the wavefront correction needed for the science path correction and for correction of the tip/tilt stars.
- Allocate the corrections to woofer and tweeter DMs.
- Invoke a dynamic compensator, such as integral feedback control, to stabilize the closed-loop path (involving the woofer DM)
- Invoke a non-linearity compensation for all of the DMs, to convert wavefront phase to command voltages.
- Process low-order (tip/tilt/focus/astigmatism) wavefront data to determine tip/tilt for the science path and for the tip/tilt paths.

The NGAO laser guide star constellation is composed of 4 fixed laser guidestars dedicated to tomography and 3 roving laser guidestars assigned to correcting the wavefront of tip/tilt reference natural stars. The tomography guidestars address a ~40 arcsec field of view (FoV) volume above the on-axis science path, while the remaining 3 roving LGS are pointable over a 120 arcsecond field. Each LGS has an associated high-order wavefront sensor (HOWFS) whether used for tomography or for correcting tip/tilt stars. The tip/tilt/focus/astigmagtism measurements are determined in the low-order wavefront sensors (LOWFS) using the NGS light. The HOWFS data from the fixed 4 guidestars are sent to the RTC tomography engine to compute the wavefront correction to be placed on the science path deformable mirror. All the LOWFS data are sent to the RTC for processing in order to determine the tip/tilt command to send to the science path fast-steering stage.

The algorithms implementing the real-time controller are run on a massively parallel processing system whose architecture is designed to take advantage of the algorithms' inherent parallelism. Operations in parallel include the independent point-and-shoot wavefront corrections, the independent camera data processing from multiple wavefront sensors and the multiple DM nonlinearity compensation. In addition, a great deal of parallelism is achieved in the inverse volume tomography algorithm by implementing it in the Fourier domain. The Fourier

domain method invokes an approximate inverse, called a preconditioner, which must be iterated to convergence. Thus we have had to assure that time spent iterating don't defeat the parallelism benefit of the Fourier domain implementation. Our analyses show that wavefront accuracy requirements for NGAO are achieved with as few as three iterations per wavefront measurement cycle, providing a clear advantage to the Fourier domain processing.

This paper presents the mathematical description of key algorithmic elements in the NGAO RTC processing. Although the algorithms are tuned to be implemented in a fast, massively parallel machine architecture, there are no approximations in the implementation of minimum variance control except for spatial and temporal sampling, and digitization. The wavefront reconstruction and tomography algorithms produce the unique minimum variance solutions taking in to account a-priori statistics and measurement noise statistics. To double-check the correctness of calculations, an RTC simulator has been coded so that results can be compared against a "traditional" matrix-multiply invocation of the minimum-variance computation. The RTC simulator will also later serve as a validation tool for use during RTC hardware/software integration and test.

The spatial and temporal sampling and the digitization (finite length machine word) are design choices. We have analyzed these and established their design values so as to meet the overall error budget.

2. ALGORITHMS (MATHEMATICAL DESCRIPTIONS)

2.1 Definition of terms

We start with a general description of the wavefront formation from guidestars and turbulent atmosphere. We then explain the nonlinearities involved with the wavefront sensing and wavefront control. Starting from these basic models and definitions, the subsequent sections will describe each portion of the RTC's algorithm set.

The common accepted practice in today's astronomical AO systems is to assume that the wavefront formation process is linear to first order, i.e. that there is a linear relationship between delta index perturbations in the volume of atmosphere above the telescope and the wavefront phase as it enters the telescope aperture. The assumption is valid for weak atmospheric turbulence, where we can assume that the optical path distance (OPD) of a ray traversing through the atmosphere departs from that of a vacuum path according to a line-integral of the delta-index of the air along that straight path. The weak turbulence assumption is valid over scales that are larger than the Fresnel zone, a transverse length scale associated with ray crossing at the ground due to higher altitude phase aberrations or, equivalently, valid for spatial frequency amplitude variations where the associated Talbot length is much longer that the atmospheric path. Hardy [1] discusses these quantities for Earth's atmosphere. For the NGAO case, where subaperture sizes, mapped to the primary, are greater than or equal to 15 cm and we are using the baseline C_n^2 profile for design purposes, the weak conditions apply with high accuracy. This is not to say that scintillation in the Hartmann wavefront sensing is negligible. This scintillation is an important contributor to the error budget, but is tolerable; hence the baseline RTC algorithms are not designed to utilize scintillation information to aid in wavefront reconstruction.

Given that the ray-traces are line integrals, we approximate them as a Riemann sum:

$$\begin{aligned}\phi_k(\vec{\mathbf{u}}) &= \left(\frac{2\pi}{\lambda}\right)\int_0^{\bar{z}} \delta\eta\left(\alpha_k(z)\vec{\mathbf{u}} + \vec{\theta}_k z, z\right)dz \\ &\cong \sum_{l=0}^{L-1} \delta\phi\left(\alpha_{kl}\vec{\mathbf{u}} + \vec{\theta}_k z_l, z_l\right)\end{aligned} \tag{1}$$

where $\phi_k(\vec{\mathbf{u}})$ are the wavefront phases at the ground, at transverse positions on the telescope aperture $\vec{\mathbf{u}}$ from a guidestar in direction $\vec{\theta}_k$ and at distance z_k (sodium LGS are located at 90 km x sec(ζ) where ζ is the zenith pointing angle of the telescope, and natural stars are at $z_k = \infty$), $\delta\eta\left(\alpha_k(z)\vec{\mathbf{u}} - \vec{\theta}_k z, z\right)$ are the index of refraction changes due to atmospheric turbulence at altitude *z* along the line $\vec{\theta}_k$, $\alpha_k(z) = 1 - z/z_k$ is the scale factor associated with a cone-beam projection, λ is the wavelength of light for which the phase is defined, and L is the number of discrete layers, indexed by *l*, in the layered approximation.

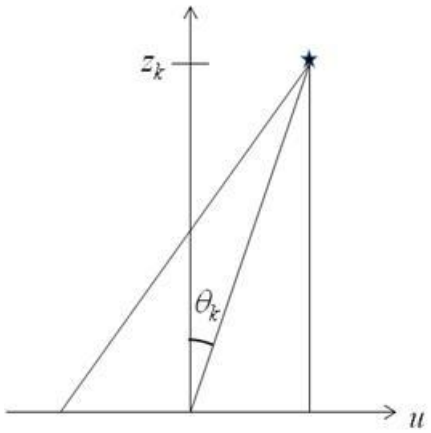

Figure 1. Geometry of ray projection from the k'th guide star.

Line up all the phase points (sampling in $\vec{\mathbf{u}}$-space) for all the wavefront sensors (indexed by *k*) and put them in a single vector, **y**. The number of elements in y is equal to the grand total count of all the subapertures in the system. Similarly, line up all the delta index-of-refractions over all the volume into one single vector, **x**. Note, the later are sampled in both $\vec{\mathbf{u}}$ (transverse) and *z* (longitudinal) space. With these definitions, the Riemann sum integral (1) can be written as a linear matrix equation

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{2}$$

where **A** is a matrix that maps every volume element to every sample on every wavefront sensor. The **A** matrix can be thought of as consisting of 1's and 0's (with a wavelength dependent factor in front to convert OPD to phase), however, with the discrete transverse sampling at each layer the lines do not necessarily pierce the layers at discrete index points, so the entries in **A** are the appropriate interpolation coefficients.

Hartmann wavefront sensors measure subaperture-averaged slopes of wavefronts. Again, we can use a Riemann sum to generate a discrete approximation to the averaged gradient operator.

$$\begin{aligned} \vec{s}_k\left(\vec{\mathbf{u}}_j\right) &= \frac{1}{A_j}\int_{\text{subap } j} \vec{\nabla}\phi_k\left(\vec{\mathbf{u}}\right) d\vec{\mathbf{u}} \\ &\cong \sum_{j' \in \text{subap } j} g_{j,j'}\phi_k\left(\vec{\mathbf{u}}_{j'}\right) \end{aligned} \tag{3}$$

where $\vec{\mathbf{u}}_j$ represents each discrete position (indexed by j) of subapertures in the aperture, and A_j is the area of that subaperture. Note that the Riemann sum allows for sub-sampling of the phase within a subaperture. Each Hartmann slope measurement is a 2-vector for each location, one for the x slope and one for the y slope.

Line up all the phase points (again, sampled in $\vec{\mathbf{u}}$-space) for all the wavefront sensors into one vector, **y**. Line up all the discrete slope measurements into one vector, **s**. Then the Hartmann wavefront sensing operation is denoted as

$$\mathbf{s} = \mathbf{G}\mathbf{y} \tag{4}$$

That is a linear operation (**s** is linearly proportional to **y**).

A note on the storage in the vector **s**: It is arbitrary how the sensor data are arranged in s (so long as the rows of **G** are correspondingly arranged). We adopt the following convention. Each wavefront sensor has a contiguous block of elements in **s**. Within each block assigned to a wavefront sensor, the x slopes are arranged in raster-scan order, then the y slopes.

We should also distinguish whether the elements of **s** and **y** include every grid location, or just those enclosed in the aperture and illuminated by the guidestar. This is an important distinction for a reconstruction algorithm that utilizes Fourier domain techniques, since Fourier transforms need points to be on a rectangular grid and on a rectangular

shaped domain. A similar consideration might apply for the elements of the **x** vector, whether or not it includes only volume elements pierced by guidestar rays or fills a rectanguloid shaped volume.

We adopt the following convention: **G** represents the matrix operation in (4) on a larger, aperture-enclosing square grid of subapertures. $\mathbf{G}_P$ represents a matrix with reduced row dimension that includes only illuminated subapertures within the telescope aperture. The pupil operator **P** is a dimension-reducing matrix that picks data (or matrix rows) that are inside the pupil. Thus

$$\mathbf{s}_P = \mathbf{P}\mathbf{s} \text{ and } \mathbf{G}_P = \mathbf{P}\mathbf{G} \tag{5}$$

The transpose of the pupil operator, $\mathbf{P}^T$ is a dimension-increasing operation that takes a vector consisting of data only from inside the pupil and produces a data vector with the data inside the pupil unchanged and entries for positions outside the pupil set to zero on an aperture-enclosing square grid. $\mathbf{P}^T\mathbf{P}$ is a masking operation that modifies data on a square grid by setting all the elements outside the pupil to zero.

The matrix **A** and vector **x** will not need to have this distinction. They will always represent the full 3-d rectanguloid volume.

For the Fourier domain algorithm discriptions, it is convenient to clearly designate whether a certain operation occurs in the spatial domain or in the spatial frequency (Fourier) domain. To do this, we introduce the Fourier transform operator **F**. The elements of **F** are:

$$F_{fj} = \exp\left\{-i2\pi\vec{\mathbf{u}}_j \cdot \vec{\mathbf{f}}_f\right\} \tag{6}$$

where $\vec{\mathbf{f}}_f$ is the spatial frequency, indexed by *f*. Note that

$$\mathbf{F}^{-1} = \frac{1}{N}\mathbf{F}^* \tag{7}$$

where *N* is the total number of sample points in the square grid. The Fourier transform of a vector of raster-scanned data is

$$\tilde{\mathbf{y}} = \mathbf{F}\mathbf{y} \tag{8}$$

We can also define the Fourier transform of a matrix so that it is consistent with matrix operations on vectors of data:

$$\tilde{\mathbf{G}} = \mathbf{F}\mathbf{G}\mathbf{F}^{-1} \tag{9}$$

The over-tilde designates the Fourier transform of an item. The Fourier transform can only be applied over aperture-enclosing square grids, thus it cannot be applied to *P* subscripted items directly (which consist only of data inside the aperture), but it can be applied to items that are inserted in to a square grid using the $\mathbf{P}^T$ operator. So we define

$$\tilde{\mathbf{s}}_P = \mathbf{F}\mathbf{P}^T\mathbf{s}_P \tag{10}$$

For tip/tilt sensors, the sensing process involves propagating to the far-field and determining the center of mass of the resulting far-field point spread function (PSF). Again, assuming the weak turbulence hypothesis, the center of mass shift is linearly proportional to the average wavefront tilt. This is not to say that there aren't significant errors in this assumption, i.e. the difference between "Z" tilt and "G" tilt is well documented [2]. In particular, an aberrated wavefront will have non-symmetric Zernike aberrations other than tip and tilt, such as coma, that will cause a further shift of the far-field PSF center of mass. In the NGAO case, this is considerably mitigated by the fact that the tip/tilt stars' wavefronts are partially corrected through the use of dedicated deformable mirrors, and by the woofer mirror, in the baseline point-and-shoot architecture. The resulting differences between Z and G tilt are then considered tolerable, and included in the error budget.

Each wavefront sensor and wavefront corrector has associated with it certain nonlinearities as well as grid and alignment distortions. These effects are very significant and so are accommodated by the RTC. Correction for grid distortions are implemented with a parallel linear multi-point interpolation method on the systolic array architecture.

A deformable mirror has, in general, complete non-linear cross-dependence of the surface response of the mirror to the voltages given to its actuators. We adopt the conventions presented in Morzinski and Gavel [3] where we presume shift-invariance, a linear plate-equation model of the deflection of the surface in response to forces, and arbitrary dependence of a given actuator's force as a function of voltage, but with no cross-coupling of actuators except through forces exerted by the top plate. The resulting formulation is:

$$\begin{aligned} f_p(\vec{\mathbf{u}}) &= \sum_j \left[f_e(V_j, \xi_j) + f_s(\xi_j) \right] \delta(\vec{\mathbf{u}} - \vec{\mathbf{u}}_j) \\ \nabla^4 \xi(\vec{\mathbf{u}}) &= f_p(\vec{\mathbf{u}}) \end{aligned} \tag{11}$$

where $f_p(\vec{\mathbf{u}})$ is the plate restorative force, $f_e(V_j, \xi_j)$ is the electrostatic deflection force of the *j*'th actuator, $f_s(\xi_j)$ is the spring return force of the *j*'th actuator, ξ_j is the deflection of the mirror surface at the actuator post location, $\xi(\vec{\mathbf{u}})$ is the deflection of the mirror surface over the contiguous face sheet, V_j is the voltage applied to the *j*'th actuator and $\vec{\mathbf{u}}_j$ is the location of the *j*'th actuator's post attachment to the face sheet. This model is not perfect, but it has been proven accurate to approximately 15 nm surface peak-to-valley on a Boston Micromachines MEMS deformable mirror. The error is incorporated in the error budget when applied to open-loop controlled DMs. It is not a critical issue for the woofer DM, which is controlled closed loop. The baseline woofer DM is a Cilas mirror that uses the latest technology piezo actuators having very little hysteresis. For the woofer, we may adopt the plate-equation and nonlinear actuator model (6), or a simpler linear superposition model:

$$\xi(\vec{\mathbf{u}}) = \sum_i d_j(V_j) r_j(\vec{\mathbf{u}} - \vec{\mathbf{u}}_j) \tag{12}$$

where $r_j(\vec{\mathbf{u}} - \vec{\mathbf{u}}_j)$ is a mirror displacement influence function and $d_j(V_j)$ is a non-linear voltage to displacement relation of the *j*'th actuator.

The grid and alignment distortion will, by convention, be referenced to a common coordinate system, which we arbitrarily attach to the 0 km conjugate optical plane at the location of the woofer DM. Even the woofer however will need a distortion grid map because the incident angle causes a foreshortening of one axis with respect to the other. We use the centered, symmetric Keck pupil to set the scale in x and y so that, nominally, $\Delta x = \Delta y$ in SI units (meters) on the Keck primary. Each of the deformable mirror actuator grids and each of the wavefront sensor subaperture grids must be mapped to this common grid. The RTC must use distortion, rotation, and displacement coefficients to map wavefronts internally represented in the common coordinate system to subaperture and actuator locations, as they are truly located in the optical system. Transformations will be of the form

$$\vec{\mathbf{u}}'_k = R_k(\vec{\mathbf{u}}_0) \tag{13}$$

where $\vec{\mathbf{u}}'_k$ is the coordinate system set by the respective DM actuator pitch or WFS subaperture pitch (DM or WFS #*k*) and $\vec{\mathbf{u}}_0$ is the common coordinate system. The distortion mapping is anticipated to be a relatively small perturbation that can be implemented in the RTC through localized interpolations. The algorithm is described in the Centroiding and DM Command Generation sections.

2.2 Wave Front Reconstruction

Wavefront reconstruction is the process of converting surface slopes to surface heights, under the assumption that the surface is continuous. The general idea is to take the numerical divergence of the slopes, which is equal to the Laplacian of the phases, then numerically inverting the Laplacian operator. In a matrix formulation, this process is equivalent to taking the pseudo-inverse of the G matrix in (4): $\mathbf{y} = \left(\mathbf{G}^T \mathbf{G}\right)^{-1} \mathbf{G}^T \mathbf{x}$. In fact, although this is the basic idea, there must be considerable modification because of several difficulties:

- The noise statistics need to be taken into account to form a minimum variance estimate
- The measurement data has a boundary, the edge of the aperture. Solutions must take in to account the discontinuity at this boundary to prevent severe error.

- The matrix G is not full rank. Piston, and, for some geometries, waffle mode, are in the null space. Thus the pseudo-inverse cannot be formulated as shown.

The approach in the NGAO reconstructor is to solve for the conditional mean phase given the slope data confined to the aperture.

$$\left\langle \mathbf{y} \middle| \mathbf{s}_P \right\rangle = \left\langle \mathbf{y}\mathbf{s}_P^T \right\rangle \left\langle \mathbf{s}_P \mathbf{s}_P^T \right\rangle^{-1} \mathbf{s}_P \tag{14}$$

where $\langle \bullet \rangle$ is the expectation operator, which when applied to the outer product of vectors is the cross-covariance of the vectors. In matrix form

$$\begin{aligned} \hat{\mathbf{y}} &= \mathbf{S}_y \mathbf{G}_P^T \left(\mathbf{G}_P \mathbf{S}_y \mathbf{G}_P^T + \mathbf{S}_n \right)^{-1} \mathbf{s}_p \\ &= \left(\sigma_n^2 \mathbf{S}_y^{-1} + \mathbf{G}_P^T \mathbf{G}_P \right)^{-1} \mathbf{G}_P^T \mathbf{s}_p \end{aligned} \tag{15}$$

where $\hat{\mathbf{y}} = \left\langle \mathbf{y} \middle| \mathbf{s}_P \right\rangle$, $\mathbf{S}_y = \left\langle \mathbf{y}\mathbf{y}^T \right\rangle$, $\mathbf{S}_n = \left\langle \mathbf{n}\mathbf{n}^T \right\rangle = \sigma_n^2 \mathbf{I}$, $\mathbf{n}$ is the measurement noise, and σ_n is the standard deviation of the measurement noise, which is assumed to be independent identically distributed Gaussian white noise for each channel of the Hartmann sensor. The second line of (15) is a result of applying the Matrix Inversion Lemma, also known as the Sherman-Morrison Theorem. The covariance matrices are computed off-line based on the statistical steady-state values. These are easily computed in the Fourier domain as each spatial Fourier component is assumed statistically independent. The covariance matrices may be changed on a slow time scale by the supervisory controller to reflect the changes in seeing conditions.

Equation (15) (first or second form) must be implemented by the wavefront reconstructor section of the RTC. We now show how we plan to implement it using massively parallel computer architecture.

Using the second form of (15), transform to the Fourier domain:

$$\begin{aligned} \tilde{\hat{\mathbf{y}}} &= \mathbf{F} \left(\sigma_n^2 \mathbf{S}_y^{-1} + \mathbf{G}_P^T \mathbf{G}_P \right)^{-1} \mathbf{G}_P^T \mathbf{s}_p \\ &= \left[\mathbf{F} \left(\sigma_n^2 \mathbf{S}_y^{-1} + \mathbf{G}_P^T \mathbf{G}_P \right) \mathbf{F}^{-1} \right]^{-1} \mathbf{F} \mathbf{G}_P^T \mathbf{s}_P \\ &= \left(\sigma_n^2 \tilde{\mathbf{S}}_y^{-1} + \tilde{\mathbf{G}}^H \mathbf{F} \mathbf{P}^T \mathbf{P} \mathbf{F}^{-1} \tilde{\mathbf{G}} \right)^{-1} \tilde{\mathbf{G}}^H \tilde{\mathbf{s}}_p \end{aligned} \tag{16}$$

where the superscript *H* denotes the Hermetian transpose (complex-conjugate of the transpose). Recall from our earlier definitions (equation (10)) that $\tilde{\mathbf{s}}_p$ is the slope data inserted in the aperture-enclosing square grid with zeros outside the pupil, then Fourier-transformed.

The advantage of using the Fourier domain is that matrices $\tilde{\mathbf{S}}_y^{-1}$, $\tilde{\mathbf{G}}$, and $\tilde{\mathbf{G}}^H$ are diagonal with

$$\left[\tilde{\mathbf{G}} \right]_{ff} = \begin{bmatrix} -i2\pi f_x \\ -i2\pi f_y \end{bmatrix} \quad \text{and} \quad \left[\tilde{\mathbf{G}}^H \right]_{ff} = \begin{bmatrix} i2\pi f_x & i2\pi f_y \end{bmatrix} \tag{17}$$

($\tilde{\mathbf{S}}_y^{-1}$ is defined in the section on parameters, below).

If it weren't for the nested pupil-mask inside the second term of the inverted matrix, this matrix too would be diagonal. Then equation (16) would become *N* independent scalar equations, amenable to an *N*-fold parallelization of its calculation.

Instead we use the iterative algorithm:

$$\tilde{\mathbf{c}} = \tilde{\mathbf{G}}^H \tilde{\mathbf{s}}_P \tag{18}$$

$$\text{iterate}\left\{\begin{array}{l}\tilde{\tilde{\varepsilon}}^{k} = \tilde{\tilde{\mathbf{c}}} - \left(\sigma_n^2 \tilde{\tilde{\mathbf{S}}}_y^{-1} + \tilde{\tilde{\mathbf{G}}}^H \mathbf{F}\mathbf{P}^T \mathbf{P}\mathbf{F}^{-1} \tilde{\tilde{\mathbf{G}}}\right)\tilde{\tilde{\mathbf{y}}}^{k} \\ \tilde{\tilde{\mathbf{y}}}^{k+1} = \tilde{\tilde{\mathbf{y}}}^{k} + \gamma \tilde{\tilde{\mathbf{C}}} \tilde{\tilde{\varepsilon}}^{k}\end{array}\right\} \text{until} \tilde{\tilde{\varepsilon}}^{k} \sim 0 \tag{19}$$

where

$$\tilde{\tilde{\mathbf{C}}} = \left(\sigma_n^2 \tilde{\tilde{\mathbf{S}}}_y^{-1} + \tilde{\tilde{\mathbf{G}}}^H \tilde{\tilde{\mathbf{G}}}\right)^{-1} \tag{20}$$

and γ is a positive gain-iteration factor (generally $\gamma \approx 0.5$). $\tilde{\tilde{\mathbf{C}}}$ is a diagonal approximation of the needed matrix inverse and is referred to as the Fourier-preconditioner. The elements on the diagonal are

$$\tilde{C}(\mathbf{f}) = 1 \Big/ \left(\sigma_n^2 \tilde{S}_y^{-1}(\mathbf{f}) + 4\pi^2 |\mathbf{f}|^2\right) \tag{21}$$

Algorithm (19) operates in parallel, one loop for each frequency domain component, and independently for each wavefront sensor. The operation requires a transform into the spatial domain to apply the aperture mask, **P**, then transforming back to the Fourier domain, each time around the iteration loop.

The iterative algorithm (19) is proven to be stably convergent [4] and our simulation experience has shown that about 30 iterations are necessary to converge to under 10 nm rms in r_0 = 15 cm seeing conditions, starting with $\tilde{\tilde{\mathbf{y}}}^0 = 0$. We advocate the use of a "warm-restart" in the RTC, where the phase estimate from the prior data time step is the starting point for the current time step, i.e.

$$\tilde{\tilde{\mathbf{y}}}^0(t) = \tilde{\tilde{\mathbf{y}}}^K(t-1) \tag{22}$$

where K is the last iteration at time step t-1. In the moderate baseline case wind conditions for NGAO, 10 m/sec, and 1 kHz sample rates, our simulations using warm restart and K=3 (one iteration per time step) typically converge to the 10 nm level in less than 100 ms.

2.3 Tomography

In tomography, the key operations are the forward propagation **A** and the back-propagation, $\mathbf{A}^T$.

The **A** matrix is of dimension $NM \times NL$ where N is the number of grid points in the aperture-enclosing square grid of data points, M is the number of wavefront sensors, and L is the number of atmospheric layers. For later convenience, we arrange **A** as an $N \times N$ array of $M \times L$ blocks. The elements of **A** are (conceptually, ignoring the issue of interpolation coefficients when **u** does not land on a sample grid point):

$$A_{lk}(\mathbf{u}_0, \mathbf{u}_l) = \delta\left(\mathbf{u}_0, (\mathbf{u}_l - \boldsymbol{\theta}_k z_l)/\alpha_{kl}\right) \tag{23}$$

where $\delta(\cdot,\cdot)$ is the Kroneker delta: = 1 if the two index arguments are equal, 0 otherwise, and a is the spatial stretching factor $\alpha_{kl} = 1 - z_l / z_k$ where z_k is the altitude of the laser guidestar.

In the Fourier domain the elements of $\tilde{\mathbf{A}}$ are:

$$\tilde{A}_{lk}(\mathbf{f}_0, \mathbf{f}_l) = \exp\{-i2\pi \mathbf{f}_l \cdot \boldsymbol{\theta}_k z_l\} \delta(\mathbf{f}_0, \mathbf{f}_l / \alpha_{kl}) \tag{24}$$

In either domain, the output domain map is stretched to conform on the cone beam. In the spatial domain, the points at an altitude are stretched apart when mapped to the ground. In the frequency domain, the spatial frequencies at an altitude are shrunk to lower frequencies on the ground.

The back propagation operation $\mathbf{A}^T$ in the spatial domain is

$$A_{lk}(\mathbf{u}_l, \mathbf{u}_0) = \delta(\mathbf{u}_l, \alpha_{kl}\mathbf{u}_0 + \boldsymbol{\theta}_k z_l) \tag{25}$$

and, in the frequency domain is

$$\tilde{A}_{lk}\left(\mathbf{f}_l,\mathbf{f}_0\right)=\exp\left\{i2\pi\mathbf{f}_l\cdot\boldsymbol{\theta}_k z_l\right\}\delta\left(\mathbf{f}_l,\alpha_{kl}\mathbf{f}_0\right) \tag{26}$$

Thus, in a similar manner similar to the forward propagator, the back propagation operator's output map is stretched to conform to the cone beam, but now in the opposite sense.

To compute a minimum-variance tomography solution, we utilize the chain rule of conditional expected values [4] to connect the reconstructed phase estimates from algorithm (19) for each wavefront sensor to a volume estimate of turbulence:

$$\left\langle\mathbf{x}(t)\middle|\mathbf{s}_P\right\rangle=\left\langle\mathbf{x}(t-1)\right\rangle+\left\langle\mathbf{x}\mathbf{e}_P^T\right\rangle\left\langle\mathbf{e}_P\mathbf{e}_P^T\right\rangle^{-1}\left(\left\langle\mathbf{y}_P\middle|\mathbf{s}_P\right\rangle+\varphi_W-\mathbf{A}_P\hat{\mathbf{x}}(t-1)\right) \tag{27}$$

where $\left\langle\mathbf{x}(t-1)\right\rangle$ is the prior estimate of the volume and φ_W is the phase on the woofer deformable mirror. In the NGAO hybrid closed-loop open-loop architecture, the woofer DM is upstream of the wavefront sensors, subtracting its phase optically, thus we need to add it back in numerically.

In matrix form the estimation formula is:

$$\begin{aligned}\hat{\mathbf{x}}(t)&=\hat{\mathbf{x}}(t-1)+\mathbf{K}\mathbf{P}\left(\hat{\mathbf{y}}(t)+\varphi_W-\mathbf{A}\hat{\mathbf{x}}(t-1)\right)\\ \mathbf{K}&=\mathbf{S}_x\mathbf{A}_P^T\left(\mathbf{A}_P\mathbf{S}_x\mathbf{A}_P^T+\mathbf{S}_{n_y}\right)^{-1}\end{aligned} \tag{28}$$

where $\hat{\mathbf{x}}=\left\langle\mathbf{x}\middle|\mathbf{s}_P\right\rangle$, $\mathbf{S}_x=\left\langle\left(\mathbf{x}-\hat{\mathbf{x}}(t-1)\right)\left(\mathbf{x}-\hat{\mathbf{x}}(t-1)\right)^T\right\rangle$ the covariance of the error in the current volume turbulence estimate and $\mathbf{S}_{n_y}=\left\langle\left(\mathbf{y}_P-\left\langle\mathbf{y}_P\middle|\mathbf{s}_P\right\rangle\right)\left(\mathbf{y}_P-\left\langle\mathbf{y}_P\middle|\mathbf{s}_P\right\rangle\right)^T\right\rangle$ the covariance of the error in the reconstructed phases in the wavefront sensors.

Assuming we can only approximate the matrix inverse by **Q**, the iterative reconstructor is

$$\begin{aligned}&\mathbf{e}_P(t)=\mathbf{P}\left(\hat{\mathbf{y}}+\varphi_W-\mathbf{A}\hat{\mathbf{x}}(t-1)\right)\\ &\text{iterate}\left\{\begin{aligned}\mathbf{r}_P^k&=\mathbf{e}_p(t)-\left(\mathbf{A}_P\mathbf{S}_x\mathbf{A}_P^T+\mathbf{S}_{n_y}\right)\mathbf{w}_P^k\\ \mathbf{w}_P^{k+1}&=\mathbf{w}_P^k+\gamma\mathbf{Q}\,\mathbf{r}_P^k\end{aligned}\right\}\text{until }\mathbf{r}^k\sim 0\\ &\hat{\mathbf{x}}(t)=\hat{\mathbf{x}}(t-1)+\mathbf{S}_x\mathbf{A}_P^T\mathbf{w}_P^K\end{aligned} \tag{29}$$

Transforming this to the Fourier domain

$$\begin{aligned}&\tilde{\mathbf{e}}(t)=\tilde{\hat{\mathbf{y}}}+\tilde{\varphi}_W-\tilde{\mathbf{A}}\tilde{\hat{\mathbf{x}}}(t-1)\\ &\text{iterate}\left\{\begin{aligned}\tilde{\mathbf{r}}^k&=\tilde{\mathbf{e}}(t)-\left(\tilde{\mathbf{A}}\tilde{\mathbf{S}}_x\tilde{\mathbf{A}}^H+\tilde{\mathbf{S}}_{n_y}\right)\tilde{\mathbf{w}}_P^k\\ \tilde{\mathbf{w}}_P^{k+1}&=\tilde{\mathbf{w}}_P^k+\gamma\mathbf{F}\mathbf{P}^T\mathbf{P}\mathbf{F}^{-1}\tilde{\mathbf{Q}}\mathbf{F}\mathbf{P}^T\mathbf{P}\mathbf{F}^{-1}\tilde{\mathbf{r}}^k\end{aligned}\right\}\text{until }\tilde{\mathbf{r}}^k\sim 0\\ &\hat{\mathbf{x}}(t)=\hat{\mathbf{x}}(t-1)+\tilde{\mathbf{S}}_x\tilde{\mathbf{A}}^H\tilde{\mathbf{w}}_P^K\end{aligned} \tag{30}$$

Where

$$\tilde{\mathbf{Q}}=\left(\tilde{\mathbf{A}}\tilde{\mathbf{S}}_x\tilde{\mathbf{A}}^H+\tilde{\mathbf{S}}_{n_y}\right)^{-1} \tag{31}$$

$\tilde{\mathbf{Q}}$ is a block diagonal matrix, with one block of size $M\times M$ for each spatial frequency. These blocks represent the spatial filtering that must be done to combine the *M* wavefront sensor phases so that they add constructively when they are back-propagated into the volume.

Accordingly, we call the quantity $\mathbf{w}_P^K$ the *preconditioned* wavefront error. When it is back propagated (and then post-conditioned by the covariance matrix $\mathbf{S}_x$) it will produce the minimum variance update to the volume estimate $\hat{\mathbf{x}}(t)$. Again, a "warm-restart" will be used in the RTC, where the preconditioned wavefront estimate from the prior time step is the starting point for the current time step. I.e.

$$\tilde{\mathbf{w}}_P^0(t) = \tilde{\mathbf{w}}_P^K(t-1) \tag{32}$$

where K is the last iteration at time step t-1. Simulation studies have shown that about 3 iterations per time step are sufficient under NGAO nominal conditions in order to maintain tomographic reconstruction error within error budget tolerance

The inverse-tomography reconstructor of equation (30) is implemented in the RTC as shown in Figure 2. Most of the operations in the loop are multiply-accumulates of a single data number by a constant, one per spatial frequency element. This is done with a massive array of compute elements, each dedicated to a single spatial frequency and all operating in parallel. Two Fourier transform pairs are required in algorithm (30). The Fourier transform is implemented as a DFT sum at each compute element, which accumulates the terms in the sum as data is shifted laterally across the array (shown as u_x and u_y in the figure). Interpolation of spatial frequencies is required during the forward and back propagation steps because of the cone beams. This is not a small grid distortion (metapupils contract ~20%) so the sample points contract about a 6 subaps in the most extreme cases. The interpolation step also requires lateral shifting across the array as depicted in the figure.

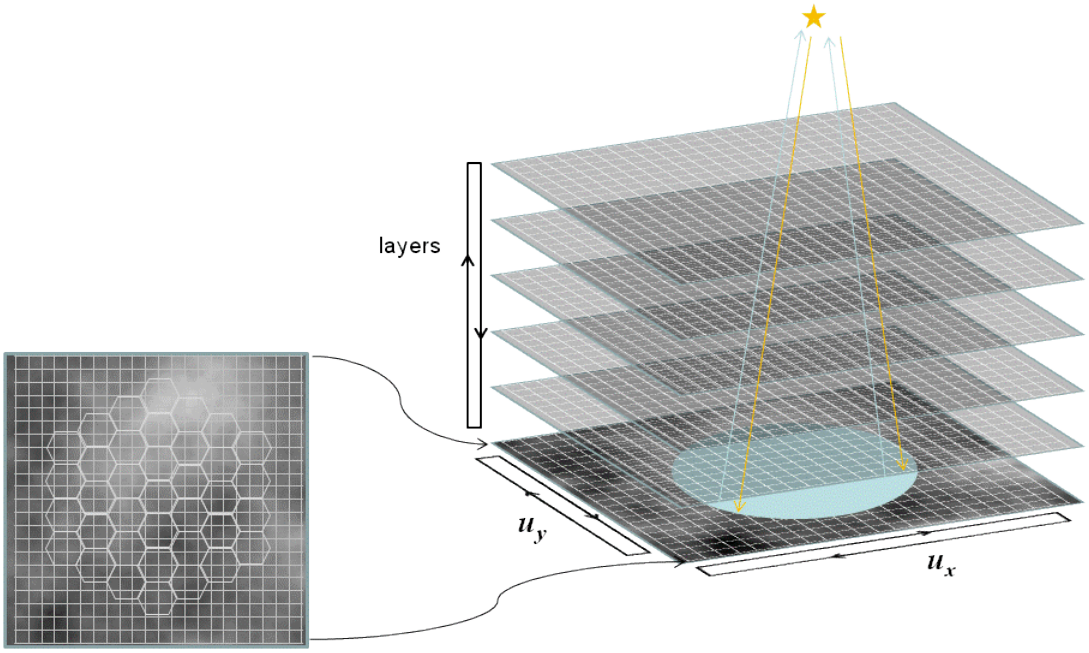

Figure 2. Mapping of the tomography problem to a massively parallel processor in systolic array architecture. Each white box depicts a separate compute element working in parallel. The compute elements communicate only with their neighbors. Some operations (DFT, interpolation) require broad distribution of data, implemented with lateral shifts (u_x, u_y) across the array.

2.4 DM Command Generation

DM command generation consists of the following steps:

- Given the volume estimate of delta-index variations $\hat{\mathbf{x}}(t)$ produced by the tomography engine, forward propagate in the science direction (on-axis) to the ground layer.
- Split the commands into woofer and tweeter components.
- Do a least-squares fit of the woofer actuator influence functions to the desired woofer shape to produce woofer actuator commands

- Apply the open-loop nonlinear model to the desired tweeter shape to produce tweeter actuator commands.

The forward propagation, since it is on-axis, involves simply adding all the layers in the volume:

$$\tilde{\phi}(\mathbf{f}) = \sum_{l=0}^{L} \tilde{x}(\mathbf{f}, z_l) = \tilde{\mathbf{A}}_s \tilde{x} \tag{33}$$

To determine which portion goes on the woofer, we make a least squares fit to the portion of the wavefront that has Frequency components less than the woofer's Nyquist frequency, f_w.

$$\tilde{\phi}_w(\mathbf{f}) = \begin{cases} \tilde{\phi}(\mathbf{f}); & f_x < f_w \text{ and } f_y < f_w \\ 0; & f_x \geq f_w \text{ or } f_y \geq f_w \end{cases} \tag{34}$$

We assume that the woofer's response is a linear superposition of actuator influence functions, described by the spatial filter function $\tilde{W}(\mathbf{f})$.

$$\tilde{\varphi}_w(\mathbf{f}) = \tilde{W}(\mathbf{f})\tilde{a}_w(\mathbf{f}) \tag{35}$$

We ignore the nonlinearities because this DM is driven in closed loop. An approximation $\tilde{B}(\mathbf{f})$ to the inverse of $\tilde{W}(\mathbf{f})$ is used to provide the least-squares fit of actuator commands to the desired low-order shape.

$$\tilde{a}_w(\mathbf{f}) = \tilde{B}(\mathbf{f})\tilde{\phi}_w(\mathbf{f}); \quad \mathbf{f} < \mathbf{f}_w \tag{36}$$

Where we've used the shorthand notation $\mathbf{f} < \mathbf{f}_w$ to represent inside the Nyquist box $f_x < f_w \text{ and } f_y < f_w$. $\tilde{B}(\mathbf{f})$ is actually a parameter that can be tailored by the AO operator. Since DMs tend to act like low-pass filters, one might be tempted to "boost" the high frequency components in order to compensate. The penalty is that $\tilde{W}(\mathbf{f})$ can have (depending on the DM) significant spatial frequency components beyond Nyquist, and these will be aliased and amplified by $\tilde{B}(\mathbf{f})$. It might be a better trade-off to allow some spectral energy near the woofer's Nyquist be corrected instead by the tweeter in order to limit the amount of re-correction of aliased woofer commands. The exact choice of $\tilde{B}(\mathbf{f})$ will have to be done experimentally once $\tilde{W}(\mathbf{f})$ is known so as to optimize the use of stroke on both deformable mirrors.

Outside the woofer-Nyquist box, the spectrum of actuator commands $\tilde{a}_w(\mathbf{f})$ must be mirror-replicated to account for aliasing

$$\begin{aligned} \tilde{a}_w(f_x, f_y) &= \tilde{a}_w(f_w - f_x, f_y); f_w < f_x < 2f_w \\ \tilde{a}_w(f_x, f_y) &= \tilde{a}_w(f_x - 2f_w, f_y); 2f_w < f_x < 3f_w \\ &\cdots \end{aligned} \tag{37}$$

until the entire spatial frequency grid associated with the high-order wavefront sensor (HOWFS) is filled out. Then, the shape of the woofer is

$$\tilde{\varphi}_w(\mathbf{f}) = \tilde{W}(\mathbf{f})\tilde{a}_w(f) \tag{38}$$

The desired shape on the tweeter is then

$$\tilde{\phi}_{tw}(\mathbf{f}) = \tilde{\phi}(\mathbf{f}) - \tilde{\varphi}_w(\mathbf{f}) \tag{39}$$

Tweeter command generation follows a more complicated set of operations because of the need to control these DMs open loop to high accuracy. The method outlined below is considerably more accurate than assuming a linear superposition fit [3].

The voltages that need to be applied to MEMS actuators depend on both the top plate displacement and the force that needs to be exerted on the top plate by the actuator. To close approximation, the relationship between the displacements and forces is linear (equation 12), albeit the forces are cross-coupled to neighboring actuators by the partial-differential plate equation.

The first step is to calculate these forces using the Fourier domain equivalent to the plate-equation's bi-Laplacian operator:

$$\tilde{f}_p(\mathbf{f}) = \tilde{\Phi}(\mathbf{f})\tilde{\phi}_{tw}(\mathbf{f}) \tag{40}$$

The second step is to transform both $\tilde{\phi}_{tw}(\mathbf{f})$ and $\tilde{f}_p(\mathbf{f})$ to the spatial domain

$$\begin{aligned} \phi_{tw} &= \mathbf{F}^{-1}\tilde{\phi}_{tw} \\ \mathbf{f}_p &= \mathbf{F}^{-1}\tilde{\mathbf{f}}_p \end{aligned} \tag{41}$$

then, for each actuator individually, use a nonlinear function (implemented as a lookup table or with spline fits) to determine the actuator voltages:

$$V_j = v\left(f_p(\mathbf{u}_j), \phi_{tw}(\mathbf{u}_j)\right) \tag{42}$$

3. SUMMARY

In this paper we have presented the baseline concept for the Keck Next Generation Adaptive Optics Real-Time Control system. This system has at its heart a tomography reconstructor which combines information from all of the wavefront sensors and results in an estimate of the 3-dimensional turbulence variations in the volume above the telescope from which wavefront phase correction in any given direction through that volume can be derived.

We also described the method of distributing the control to a woofer-tweeter combination of deformable mirrors, the methods for accounting for nonlinearities of deformable mirrors and of wavefront sensors, and our approach to accounting for pupil distortions amongst all of the relayed pupils within the system.

Convergence properties and accuracy of the presented algorithms have been validated with simulations. A preliminary design of a custom FPGA based hardware for the tomography engine is now under investigation.

4. ACKNOWLEDGEMENTS

This material is based upon work supported AURA through the National Science Foundation under Scientific Program Order No. 5 as issued for support of the Telescope Systems Instrumentation Program (TSIP), in accordance with Proposal No. AST-0335461 submitted by AURA.

The W. M. Keck Observatory is operated as a scientific partnership among the California Institute of Technology, the University of California, and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W. M. Keck Foundation.

REFERENCES

[1] Hardy JW, *Adaptive Optics for Astronomical Telescopes*. New York: Oxford University Press, 1998.

[2] Tyler GA, "Bandwidth considerations for tracking through turbulence," JOSA A, 11, 1, 358-367, (1994).

[3] Gavel DT and Morzinski KA, "The open-loop control of MEMS: modeling and experimental results," in *Proceedings of the SPIE 6467*, Orlando FL, 2007.

[4] Gavel D, "Stability of Closed Loop Tomography," UC Santa Cruz, Laboratory for Adaptive Optics, Internal Report 2004.