Update on the TMT Adaptive Optics Real Time Controller

C. Boyer[a], L. Gilles[a], B. Ellerbroek[a], G. Herriot[b], J.P. Véran[b]

[a]TMT Observatory Corp., 2632 E. Washington Blvd, Pasadena, USA
[b]Herzberg Institute of Astrophysics, 5071 W. Saanich Road, Victoria, BC V9E 2E7, Canada

ABSTRACT

The Thirty Meter Telescope (TMT) will implement a first light facility Laser Guide Star Multi Conjugate Adaptive Optics System, NFIRAOS, which will feed three science instruments on the telescope Nasmyth platform. This system will include two deformable mirrors, six laser guide star wavefront sensors, and multiple tip/tilt/focus wavefront sensors located in the instruments. The Real Time Controller (RTC) is one of the most innovative and essential components of this first light AO system. In this paper, we provide an update on the NFIRAOS RTC overall requirements and challenges, and in particular, on the tomography and fitting wavefront reconstruction algorithms. Several implementations of a minimum variance reconstructor are presented together with their processing and memory requirements.

Keywords: Thirty Meter Telescope, Adaptive Optics, NFIRAOS, Real Time Controller, wavefront reconstruction algorithms

1. INTRODUCTION

The initial Adaptive Optics (AO) architecture for the TMT has been defined to provide near-diffraction-limited wavefront quality and high sky coverage in the near Infra-Red (IR) for the early light TMT science instruments IRIS and IRMS.[1–3] It is a Laser Guide Star (LGS) Multi Conjugate AO (MCAO) architecture consisting of:

- the Narrow Field Infra-Red Adaptive Optics System (NFIRAOS),[4] which provides turbulence compensation over a moderately large field of view (1-2 arcmin) in order to sharpen the images of the natural guide stars and improve the sky coverage,

- the Laser Guide Star Facility (LGSF),[5] which generates multiple LGS in the mesopheric sodium layer with the brightness, beam quality and asterism geometry required by both NFIRAOS and the second generation of TMT AO systems

- and the AO Sequencer (AOSQ),[6] which automatically coordinates the actions of NFIRAOS and the LGSF during science observations.

The initial TMT AO control architecture is presented in figure 1.

The NFIRAOS facility includes (i) two Deformable Mirrors (DM) conjugated at 0km (61x61 DM) and 12km (73x73 DM), (ii) one Tip/Tilt Platform (TTP) serving as a mount for the ground level DM, (iii) six 60x60 LGS Wavefront Sensors (WFS) observing an asterism of five LGS equally spaced on a radius of 35 arcsec and one additional on-axis LGS, (iv) up to three Tip/Tilt/Focus (TTF) WFS working in the near IR and located within each NFIRAOS instrument, (v) one 60x60 Natural Guide Star (NGS) WFS for observations without LGS, (vi) two Truth Wavefront Sensors (TWFS), which are used to calibrate for slow-varying wavefront biases due to variations in the sodium layer profile during LGS observations, and to calibrate for long term drifts in image quality for NGS observations, and (vii) a Real Time Controller (RTC),[6] which processes the inputs from the multiple WFS to compute the commands to the deformable mirrors and the tip/tilt platform.

The RTC is one of the most challenging systems of NFIRAOS. The requirements for the RTC represent a significant advance over the current generation of astronomical AO systems and require computationally efficient implementation of minimum variance wavefront reconstruction algorithms as well as innovative hardware architectures.

Section 2 of this paper provides an update on the NFIRAOS RTC requirements and functional block diagrams. Several modes of operations are defined and the control algorithms are detailed further. Section 3 focuses on the LGS wavefront

---

cboyer@tmt.org; phone: 1 626 395 1623; www.tmt.org

reconstruction algorithm, which is the most challenging in terms of processing requirements. Four possible candidates algorithms have been studied in great details and are described here. They are considered to provide equivalent AO performance but have different processing and memory requirements that may lead to different hardware implementations. Finally, section 4 summarizes the processing and memory requirements of the most demanding NFIRAOS RTC processes.
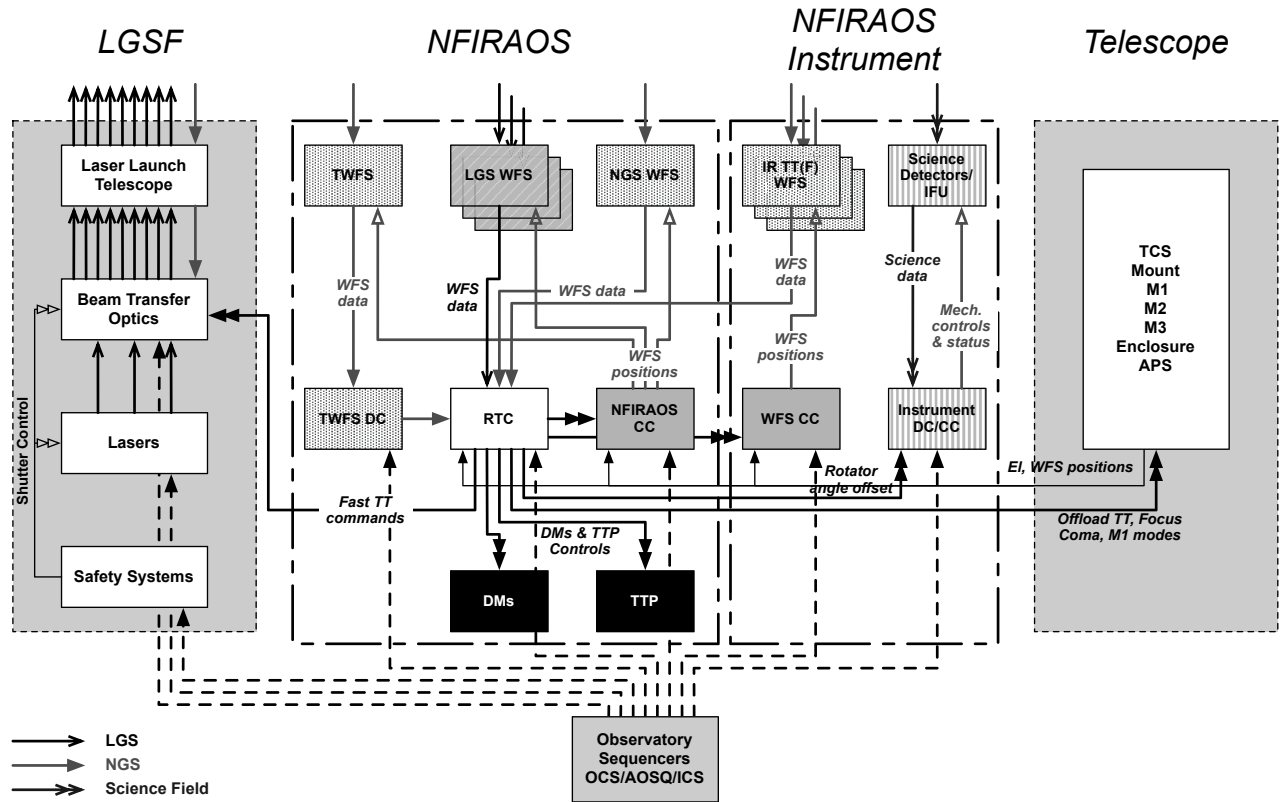

Figure 1. Initial TMT AO architecture

2. THE RTC REQUIREMENTS

2.1. Functional requirements

The functional requirements for the NFIRAOS RTC remain very similar to what was defined in a previous paper.[6] The NFIRAOS RTC will provide all the control functions that are required for calibration, test and real time atmospheric turbulence compensation. It will accept and process the inputs from several WFS and compute the commands to several wavefront correctors on the basis of these WFS measurements. It will operate at up to a 800Hz sampling rate with a 1000μs end-to-end latency (and a strong goal of 400μs). The RTC will update and optimize its control algorithms in real time as the observing parameters and the atmospheric conditions change. It will offload the persistent, low spatial frequency components of the wavefront corrector commands to the telescope sub-systems via the AO Sequencer. It will compute the commands to the LGSF fast steering mirrors to stabilize the pointing of the LGS on the sky. It will store up to one night of circular buffer data for diagnostic purposes or for PSF reconstruction in post processing. Finally, as illustrated in figure 1, the RTC will interface with several other AO systems and telescope sub-systems as necessary for real time atmospheric turbulence compensation.

2.2. RTC block diagrams

The RTC will implement 3 modes of operations:

- LGS AO mode. In this mode, the RTC uses the six LGS WFS and up to three TTF WFS to compute the commands for the two DMs and the TTP. The low order TTF WFS allow the measurement and estimation of the (i) tip/tilt for fast guiding, (ii) focus for LGS WFS calibration and (iii) tilt anisoplanatism for higher-order wavefront control. Roughly 100 low-order modes of the commands to the DMs and to the TTP are offloaded to the telescope sub-systems.

- NGS AO mode. In this mode, the RTC uses the high order NGS WFS and, if possible, one of the TTF WFS to compute the commands to the DM conjugated at 0km and to the TTP (the second DM being flattened). The TTF WFS measurements are used to improve the tip/tilt estimate (less anisoplanatic error and/or reduced measurement noise) and also to compensate for flexure within the instrument. As for the LGS case, roughly 100 low-order modes of the commands to the DM and to the TTP are offloaded to the telescope sub-systems.

- Seeing limited mode. There will be science cases for the IRMS instrument where NFIRAOS will be used solely as an active optics system. The DMs will be flattened and a combination of TTF WFS and TWFS measurements will be used to compute the telescope active optics commands.

The RTC will also implement a calibration mode to measure, WFS gains, biases, reference vectors, and noise statistics, DM-to-WFS influence matrices, pupil alignment measurements for each WFS, etc.

Figures 2 and 3 give the block diagrams of the TMT RTC for the LGS AO mode and the NGS AO mode. Block diagrams for the 3 operation modes have been developed, including the cases with and without an additional TTF WFS for the NGS AO mode. The block diagrams share identical modules, like the wavefront corrector control or the offload to the telescope modules. Some of the individual processes have been detailed, in particular the LGS or NGS wavefront reconstruction, the LGS WFS processing and the LGS WFS reference vector computation. Finally, some of the background processes are now executed by the Reconstructor Parameter Generator (RPG) system, which is responsible for computing and providing all of the matrices/vectors/parameters needed by the RTC to perform the wavefront reconstruction.
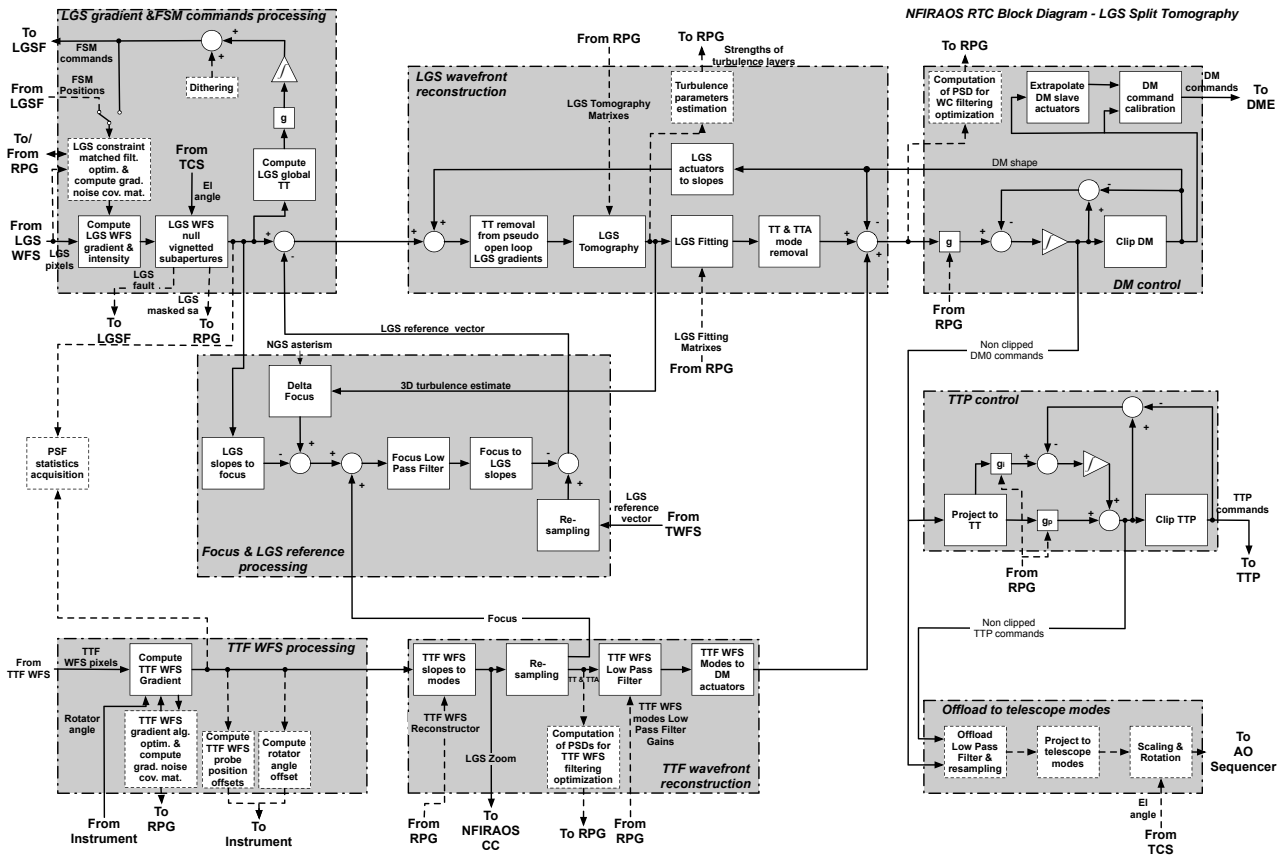

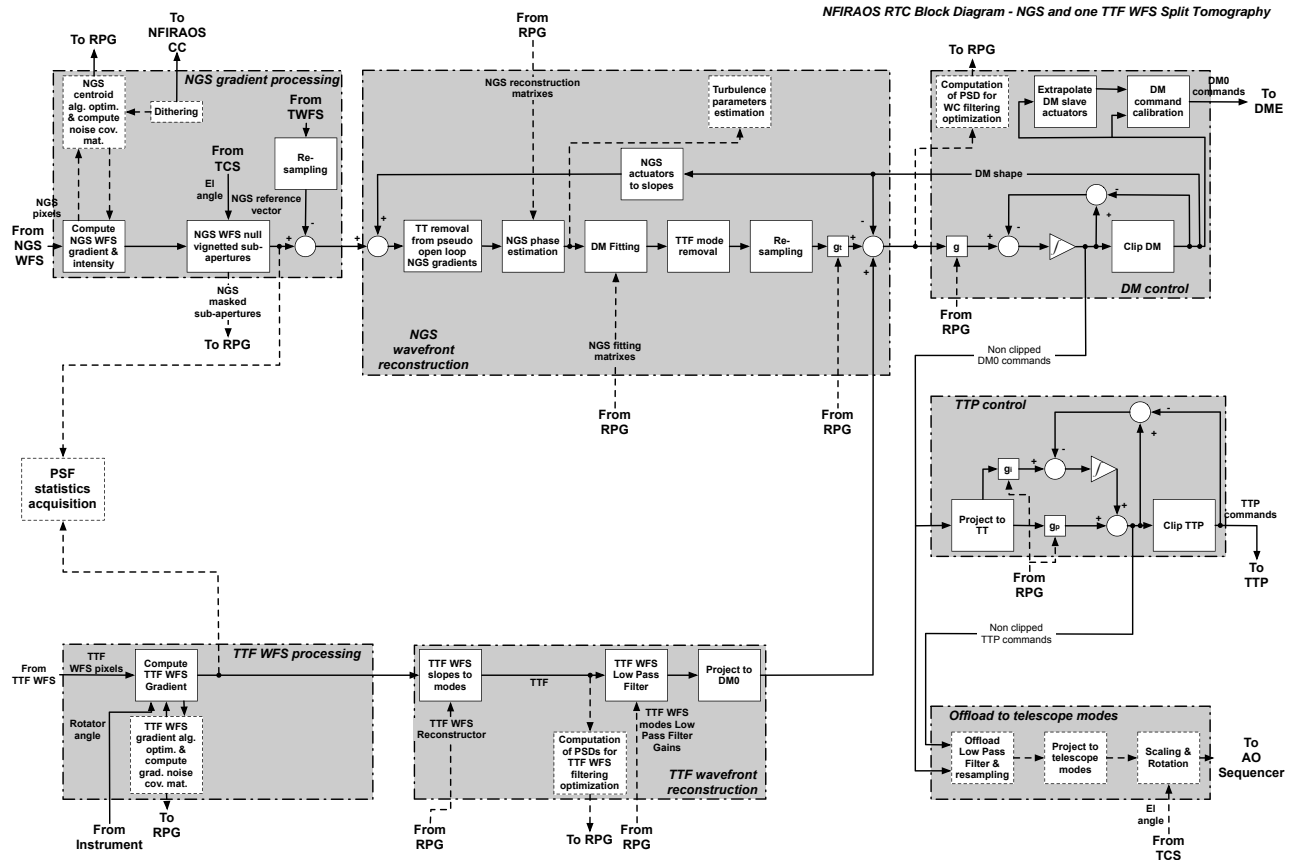**Figure 2.** RTC block diagram for the LGS AO mode

Figure 3. RTC block diagram for the NGS AO mode with a TTF WFS included

2.3. WFS processing

The 60x60 Shack Hartmann NFIRAOS LGS WFS will utilize the polar coordinate CCD array designed specifically for use with elongated laser guide stars.[7] The polar coordinate arrays are constituted of small rectangular islands of pixels oriented along lines radiating from the center of the array. The number of pixels per sub-aperture ranges from 6x6 to 6x15 and is proportional to the amount of the elongation. The total number of pixels per WFS to read and calibrate is ∼205 kpixels. From these measurements, the RTC computes, at a sampling rate of 800Hz, ∼5800 gradients per WFS using a constrained matched-filter processing algorithm.[8] The LGS WFS processing is performed synchronously with the digitization of the LGS WFS pixel intensities (500μs), with the final sub-aperture gradient computed within 10μs after the final pixel has been provided. The constrained matched filter is optimized in real time, at a sampling rate of 1Hz (with a goal of 10Hz) by dithering the LGS pointing with the LGSF fast steering mirrors. For each LGS WFS, the RTC computes the intensity level of the WFS sub-apertures at a rate of 800Hz and masks the unusable sub-apertures due to telescope obscuration, Rayleigh backscatter or any M3 alignment problem. Then, for each LGS WFS, a reference vector is subtracted from the LGS WFS gradients. The LGS reference vectors are computed at 800Hz, based upon (i) the TTF WFS defocus, (ii) the NGS-to-LGS delta focus derived from the 3D turbulence profile estimate, (iii) the LGS WFS defocus measurements and (iv) the TWFS measurements. The resulting gradient vectors are used as inputs to the LGS wavefront reconstruction process. The pointing commands to the LGSF fast steering mirrors are computed at a rate of 800Hz, based on the reference-subtracted gradients and the dithering signals needed for the optimization of the matched filter.

It is expected that the 60x60 NGS WFS will use a 240x240 CCD with sub-electron read noise, providing 2x2 pixels per sub-aperture and two guard rows/columns between the sub-apertures. The total number of pixels to read and calibrate is ∼12kpixels. From these calibrated pixels, the RTC computes ∼5800 gradients at a sampling rate varying between 10Hz and 800Hz as required by the NGS signal level. A standard quadrant detector algorithm is used to compute the gradients,

with the centroid gains updated* in real time at a rate of at least 0.1Hz, to account for variations in seeing and AO system performance. The RTC also computes the intensity level of each sub-aperture and masks the unusable sub-apertures due to telescope obscuration and M3 alignment problems. A reference vector computed from the TWFS measurements is subtracted to the NGS WFS gradients, and the resulting vector is used as input to the NGS wavefront reconstruction process. As for the LGS WFS processing, the NGS WFS processing is performed synchronously with the digitization of the NGS WFS pixel intensities (500μs), with the final sub-aperture gradient computed within 10μs after the final pixel has been digitized.

Finally, the TTF WFS are low order Shack Hartmann WFS with either 2x2 or 1x1 sub-apertures. The sampling rate at which the gradients are computed depends upon the signal level of the corresponding guide stars. The number of pixels to read and calibrate is small in comparison to the NGS WFS or the LGS WFS: $\sim$3kpixels (a maximum of 16x16 pixels per sub-aperture). The gradients are computed using a constrained matched filter, which is updated at a rate of at least 0.1Hz based upon variations in seeing and AO system performance. The gradients are de-rotated based on the instrument rotator angle and then used as inputs for the TTF wavefront reconstruction process.

2.4. Wavefront reconstruction and control of the wavefront correctors

As shown in figures 2 and 3, the low order modes of the wavefront reconstruction are controlled separately (split tomography) from the high order components of the LGS or NGS wavefront reconstruction.

The LGS wavefront reconstruction is described in great detail in section 3. The NGS wavefront reconstruction follows the same approach as the LGS wavefront reconstruction and the same set of algorithms is proposed. However, in the NGS case, all block structured operators/vectors reduce to standard operators/vectors, and the bilinear interpolation operators reduce to the identity.

A least squares modal reconstructor is used to compute the six low order modes (tip, tilt, focus and tilt anisoplanatism) in LGS AO mode or 3 low order modes (tip, tilt and focus) in NGS AO mode from the TTF WFS measurements. The low order modes are then temporally filtered and transformed onto DM commands at a rate of 800Hz in LGS AO mode, or at the TTF WFS sampling rate in NGS AO mode. The focus mode is used to control the LGS zoom corrector and the LGS WFS reference vectors in the LGS AO mode, and is not projected to the DM commands. The temporal filter parameters for the low order modes are updated in real time at a rate of 0.1Hz, consistently with the updates to the temporal filters for the wavefront corrector control process, to enhance control loop stability.

The number of DM actuators is given in Table 1:

	Active actuators	Edge actuators	Total actuators/DM
DM0	2821	304	3125
DM12	$\sim$3916	$\sim$421	$\sim$4337
Total	$\sim$6737	$\sim$725	$\sim$7462

Table 1. DM0 and DM12 number of actuators

The pair of DM error signals computed from the TTF WFS measurements and the high order wavefront reconstruction process are combined with the current DM commands applied after clipping. A simple integrator with an adjustable gain is applied to the DM error signals computed by the wavefront reconstruction processes, to enhance the control loop stability and balance the residual wavefront errors due to WFS measurement noise and time delay.

A woofer/tweeter algorithm is implemented for the control of the tip/tilt modes. The TTP is used as the woofer (low-bandwidth-high-stroke device) and the DM0 is used as the tweeter (high-bandwidth-low-stroke device). The TTP commands are obtained by applying an additional proportional-integrator filter to the tip/tilt component of the filtered DM commands.

The filtered DM and TTP commands are clipped to avoid saturation based upon the dynamic range and slew rate limits of the DMs and TTP. Integrator windup is prevented by subtracting such clipping adjustments from the inputs of the

---

*A dedicated tip/tilt mirror located in the NGS path may be dithered for this purpose.

integrator. Finally, the RTC computes the commands for the DM edge actuators to maintain a continuous facesheet figure at the boundary, and calibrates the active and edge actuator commands to account for variations in the actuator gains and offsets due to temperature. The clipped and calibrated commands for the DMs and TTP are then sent to the DM and TTP control electronics.

The temporal filter parameters for the wavefront corrector control process are updated at a rate of 0.1Hz (together with the modal temporal filter parameters for the TTF wavefront reconstruction) to enhance control loop stability. The RTC computes the Power Spectral Densities (PSD) of the actuator commands and the TTF wavefront modes, and transfers the PSDs to the RPG System. The actual optimization of the temporal filter parameters is then performed within the RPG.

At last, the ground level DM and TTP commands obtained after filtering are projected onto telescope modes (tip, tilt, focus, coma, and up to 100 primary mirror modes). The telescope modes are offloaded at a relatively slow rate to the telescope sub-systems via the AO Sequencer.

2.5. Other background processes

The RTC includes several additional background processes to monitor the AO performance of the NFIRAOS system. In particular, the RTC computes the following atmospheric parameters in the LGS AO mode at a rate of at least 0.1Hz: (i) the strengths of the turbulence layers, (ii) the Fried parameter r_0 and seeing, (iii) the atmospheric time constant τ_0, (iv) the isoplanatic angle θ_0 and the (v) the outer scale L_0. In NGS AO mode, only the parameters (ii), (iii) and (v) are computed. The Fried parameter r_0, the time constant τ_0, the outer scale L_0 and the strengths of the turbulence layers are computed from spatial and temporal structure functions of estimated turbulence phase screens. The isoplanatic angle θ_0 is computed from the estimated strengths and (a priori) ranges of the turbulence layers.

Another important background process is the acquisition of the telemetry data. The RTC will implement a telemetry storage system, allowing the acquistion and storage of up to one night of data (90TB and a data rate of 3.5GB/s are required). As a baseline, this system will also be used to acquire the PSF statistical data (WFS gradients and WFS gradient noise covariance matrices) needed to estimate the PSF in post-processing.

3. LGS WAVEFRONT RECONSTRUCTION ALGORITHMS

LGS wavefront reconstruction refers to the computation of wavefront values at DM actuator locations from closed loop LGS gradient measurements. The approach adopted is based on computationally efficient algorithms that provide iterative solutions to open loop minimum variance wavefront reconstruction with a sparse approximation of the regularization term.[9, 10] In this framework, wavefront values are first estimated on a set of phase screens (tomography step), and this estimate is then projected onto DM actuator locations (fitting step) to optimize performance (wavefront variance and Strehl ratio) averaged over a specified field of view (FoV). However, the proposed reconstruction architecture involves the following important modifications from conventional open loop minimum variance reconstruction:

(i) 5 modes, namely tip/tilt (TT) and 3 plate scale or tilt anisoplanatism (TA) modes, are controlled at a lower bandwidth by a second servo loop using a noise-weighted least-squares reconstruction matrix operating on closed loop TTF WFS gradients. The 5 NGS modes can be specified as follows: (i) TT on the ground level DM with no command on the upper DM; (ii) defocus and astigmatism on the ground level DM with a scaled version of these 3 quadratic modes on the upper DM, with a (negative) scale factor chosen to ensure that for point sources at the range of the LGS, the wavefronts propagated to the aperture-plane consist of pure TT, which is invisible to the TT-removed LGS WFSs.[11]

(ii) All modes orthogonal to these 5 TT/TA modes are controlled at high bandwidth based upon pseudo open loop LGS gradient measurements, using a computationally efficient algorithm which approximately solves the sparse tomography and fitting matrix systems.

(iii) TT is removed from the pseudo open loop LGS gradients on account of LGS position uncertainty, but tip/tilt removal is not performed inside the tomography matrix.

(iv) The 5 TT/TA modes are projected from the LGS wavefront reconstruction at the output of the fitting step to reduce the cross-coupling between the LGS and NGS loops.

(vi) The tomography matrix incorporates a sparse, scaled, Laplacian squared regularization matrix (high-pass spatial filter) which approximates the atmospheric turbulence inverse covariance matrix, summed with a scaled identity matrix to prevent the system from being singular.

(vii) Piston removal and piston regularization constraints are not imposed on the fitting system.

The main benefits of the split tomography approach are: a simpler formulation of the LGS tomography step with reduced computation, simpler, more flexible control of the TT and TA modes, and reduced coupling betweeen the LGS and TT/TA modes.

LGS wavefront reconstruction performs operations on 3 types of wavefront grids: a set of atmospheric phase screens, an aperture-plane grid, and two DM grids. In what follows, four algorithms will be discussed to perform atmospheric tomography. Three of them are conveniently implemented on square grids using a cone coordinate system for the atmospheric grids where mesh size is squeezed with range according to the cone compression factor $\Delta_k = \alpha_k \xi_k \Delta_0$. Here $\xi_k = 1 - h_k/h_{\text{lgs}} \leq 1$, $\alpha_k = 1$ or 2, and $\Delta_0 = d_{\text{sa}}/2 = 1/4$ m denotes the aperture-plane grid mesh size. Note that regardless of the choice of coordinate system, any LGS tomographic wavefront reconstruction algorithm needs to be updated slowly with time due to sodium layer range variability. In the cone coordinate system, this update is most conveniently accomplished by updating the atmospheric grids themselves, i.e. updating the above ξ_k parameters.

Four core operations are involved in the overall LGS wavefront reconstruction process:

- Bilinear interpolation. This operation computes wavefront values at the intercepts of rays traced through phase screens (atmospheric and DM grids) to/from the aperture plane, for sources at either finite or infinite range. Rays traced to/from the aperture plane correspond to forward/backward wavefront propagations, which are transposes of one another. The projected aperture-plane grid point coordinates on a plane at range h for a source at range h_{src} in direction θ are given by $x' = (1 - h/h_{\text{src}}) x + h \theta$, where x denotes aperture-plane grid point coordinates.

- Aperture-plane gradient and gradient transpose. The gradient operator computes each LGS WFS gradient measurement from wavefront values at up to 9 points of the aperture-plane grid on a 3×3 stencil covering the associated subaperture. Stencil weights are fully determined by the amplitude profile defining a subaperture. At least 80% of the subapertures in each WFS are expected to be fully illuminated, and the gradient operator associated with these subapertures has a regular stencil coupling only the 6 points that border opposite subaperture edges (3 points on each side with opposite weights). Similarly, the gradient transpose operator associated with these subapertures has a regular stencil coupling either 0, 2 or 4 subapertures into each grid point.

- Atmospheric wavefront curvature squared. This operation regularizes the tomography step. The curvature squared operator couples 13 points lying on a 5×5 stencil, with 3 points from each of the 4 corners removed. The 13 weights are identical for all atmospheric phase screens, and are identical throughout each screen, except near the boundary.

- Aperture weighting. This operation is performed in the DM fitting step to capture aperture edge effects and yield a lower fitting error by cross-coupling aperture-plane grid points with up to 8 nearest neighbors on a 3×3 stencil. Stencil weights are fully determined by the amplitude profile defining the aperture, and are identical throughout the aperture, except near the bounday. Fully interior stencils have 2-dimensional Simpson weights.

All four of these operations shall be described in greater detail below. Grid-based as opposed to sparse matrix-based operations are preferable since they have the highest potential to fully exploit the underlying structure of these operators, minimize storage requirements, and most importantly, exploit hardware parallelism.

3.1. LGS Pseudo Open Loop Gradient Computation

For stability reasons, the LGS loop must operate on pseudo open loop gradients. In the absence of measurement noise and systematic errors (WFS pupil distortion and misregistration/calibration errors), the poles of the LGS closed loop transfer matrix are then equal to those of a conventional minimum variance reconstructor with closed loop constraints. Pseudo open-loop gradients represent an estimate of the open loop gradients that would have been measured with ideal flat DMs and ideal, fully linear wavefront sensors. As a result, the pseudo open loop architecture eliminates the need for an invisible

mode removal step as in a conventional closed loop system. The pseudo open loop gradients associated to LGS WFS k are computed as follows:

$$[s_{\mathrm{dm}}]_k = \Gamma \left[\begin{array}{ccc} \left[\widetilde{H}_{\mathrm{dm}}\right]_{k1} & \cdots & \left[\widetilde{H}_{\mathrm{dm}}\right]_{kN_{\mathrm{dm}}} \end{array} \right] \left[\begin{array}{c} [\widehat{x}]_1 \\ \vdots \\ [\widehat{x}]_{N_{\mathrm{dm}}} \end{array} \right], \quad 1 \le k \le N_{\mathrm{lgs}}, \tag{1}$$

$$[s_{ol}]_k = [s]_k + [s_{\mathrm{dm}}]_k, \tag{2}$$

where $N_{\mathrm{dm}} = 2$ denotes the number of DMs, s_{ol} and s denote respectively the pseudo open loop and closed loop gradient vectors, $\widehat{x}$ is the concatenated DM actuator vector, Γ denotes the aperture-plane gradient operator, and $\left[\widetilde{H}_{\mathrm{dm}}\right]_{kl}$ represents the bilinear interpolation operator for DM grid l for rays traced from a finite range source (indicated by the tilde notation) in the direction of LGS number k to the aperture-plane grid. $\left[\widetilde{H}_{\mathrm{dm}}\right]_{kl}$ will generally have the following stencil:

$$w = \left[\begin{array}{c} \beta_y \\ 1 - \beta_y \end{array} \right] \left[\begin{array}{cc} 1 - \beta_x & \beta_x \end{array} \right] = \left[\begin{array}{cc} (1 - \beta_x)\,\beta_y & \beta_x\,\beta_y \\ (1 - \beta_x)\,(1 - \beta_y) & \beta_x\,(1 - \beta_y) \end{array} \right], \tag{3}$$

where (β_x, β_y) is the ray intercept offset relative to the DM grid mesh size from the lower left neighbor node. Note that this offset is not constant, since DM grids are not in the cone coordinate system and the intercepts are from finite range sources. The β_x/β_y offsets for each LGS can be stored separately in two one-dimensional arrays, since β_x is only a function of the x-intercept of a ray in the aperture plane (and similarly for β_y). Finally, note that the values of β_x/β_y need to be updated every few seconds as the range to the sodium layer varies.

3.2. LGS Pseudo Open Loop Gradient Tip/Tilt Removal

The first step of LGS tomography consists in removing global tip/tilt from the LGS pseudo open loop gradients. This step is trivially accomplished by projecting off the gradient tip/tilt modes from each LGS WFS pseudo open loop gradient vector, which can be represented as follows:

$$[c]_k = (T^T\,W_k\,T)^{-1}\,T^T\,W_k[s_{ol}]_k, \tag{4}$$

$$[\overline{s_{ol}}]_k = [s_{ol}]_k - T\,[c]_k, \tag{5}$$

where $[s_{ol}]_k$ denotes the pseudo open loop gradient vector of LGS WFS k, $[c]_k$ is the estimate of the tip/tilt coefficients for $[s_{ol}]_k$, W_k is the pre-computed thresholded LGS measurement noise inverse covariance matrix (sparse weighting matrix with 2 non-zero elements per row) for LGS WFS k, and T is a common 2-column, $\{0, 1\}$-valued matrix containing the gradient tip/tilt modes.

3.3. Computing the LGS Tomography System Right Hand Side Vector

The second step of LGS tomography consists in assembling the right hand side vector of the tomography matrix system from the tip/tilt removed LGS pseudo open loop gradients, which is accomplished by the following operations:

$$[y]_k = \Gamma^{\mathrm{T}}\,W_k\,[\overline{s_{ol}}]_k, \qquad 1 \le k \le N_{\mathrm{lgs}} \tag{6}$$

$$[b]_l = \left[\begin{array}{ccc} \left[\widetilde{H}_{\mathrm{atm}}\right]_{1l}^{\mathrm{T}} & \cdots & \left[\widetilde{H}_{\mathrm{atm}}\right]_{N_{\mathrm{lgs}}l}^{\mathrm{T}} \end{array} \right] \left[\begin{array}{c} [y]_1 \\ \vdots \\ [y]_{N_{\mathrm{lgs}}} \end{array} \right], \qquad 1 \le l \le N_{\mathrm{ps}} \tag{7}$$

where $[b]_l$ denotes the right hand side tomography sub-vector corresponding to phase screen l, and $\left[\widetilde{H}_{\mathrm{atm}}\right]_{kl}$ is the bilinear interpolation operator between the nodes of atmospheric layer l and intercepts of rays traced to the aperture-plane grid

from the direction of LGS WFS k. The tilde again indicates ray tracing from a finite range source. Note that the interpolation offset (β_x, β_y) is constant throughout the grid when grid nodes are defined in the cone coordinate system, which is slowly updated to reflect sodium layer range variability. This property is particularly well suited for a parallel grid-based implementation. The bilinear interpolation operators $\left[\widetilde{H}_{\text{atm}}\right]_{kl}$ are updated at a rate of 0.1 Hz as the range to the screens (i.e. zenith angle) and range to the LGS vary. The nodes of the aperture-plane grid can always be chosen to coincide with nodes of the ground-layer phase screen, such that no interpolation is needed for that screen.

3.4. Computing an approximate solution to the LGS Tomography System

The tomography matrix system takes the form $A\,x = b$, where x denotes the concatenated tomography vector of unknowns (atmospheric phase screens), A is the block structured tomography operator, and b is the tomography right hand side vector. All tomography solvers discussed below perform $[A]_{kl}\,[v]_l$ operations ($1 \leq k, l \leq N_{\text{ps}}$), which can be expressed as follows:

$$\psi = \text{Diag}(\Gamma^{\text{T}}\,W_1\,\Gamma, \cdots, \Gamma^{\text{T}}\,W_{N_{\text{lgs}}}\,\Gamma) \begin{bmatrix} \left[\widetilde{H}_{\text{atm}}\right]_{1l} \\ \vdots \\ \left[\widetilde{H}_{\text{atm}}\right]_{N_{\text{lgs}}l} \end{bmatrix} [v]_l \;, \tag{8}$$

$$[p]_k = \begin{bmatrix} \left[\widetilde{H}_{\text{atm}}\right]_{1k}^{\text{T}} & \cdots & \left[\widetilde{H}_{\text{atm}}\right]_{N_{\text{lgs}}k}^{\text{T}} \end{bmatrix} \psi \;, \tag{9}$$

$$[q]_k = (\gamma_k\,L^2 + \epsilon\,I)\,[v]_k \;, \tag{10}$$

$$[A]_{kl}\,[v]_l = [p]_k + [q]_k \;, \tag{11}$$

where L^2 is the curvature squared regularization operator coupling 13 grid points in a 5×5 stencil on each screen, γ_k is a scaling constant proportional to Δ_k^{-4} and to the 5/6 th power of the Fried parameter of layer k (computed by the RTC in real time), and ϵ is a small regularization parameter to prevent A from being singular. Note that eq.(8) does not incorporate a tip/tilt removal step after each multiplication by Γ even though global tip/tilt has been removed from the pseudo open loop gradients.

Four algorithms falling in two categories are proposed for solving the LGS tomography system $A\,x = b$: two block oriented algorithms operating on the blocks (i.e. layers) of the system one by one and two system-oriented algorithms operating on the whole system (i.e. all layers) at once. Detailed Monte Carlo simulations indicate that the closed loop performance of all four algorithms agree to within 15 nm RMS in quadrature.[12] These algorithms are respectively:

- BGS-CBS: Block Gauss-Seidel with Cholesky Back Substitutions. The core of this algorithm performs forward/back substitutions on N_{ps} smaller systems $[A]_{kk}\,[x]_k = [v]_k$, where the right hand side vector $[v]_k$ depends on the solution obtained for all diagonal blocks above or below k (the later direction depends on whether forward or backward BGS is implemented and forward BGS is assumed below without loss of generality). Memory requirements and operation counts are minimal when the algorithm is implemented on uncompressed non-square atmospheric grids which are matched in size to the metapupil associated with the LGS asterism and the science FoV. In condensed notation, starting with a warm initial guess $x^{(0)}$ the algorithm solves the following system:

$$(A_{\text{D}} + A_{\text{L}})\,x = y \;, \tag{12}$$

$$y = b - A_{\text{U}}\,x^{(0)} \;, \tag{13}$$

where A_{D}, A_{L} and A_{U} denote respectively the block diagonal, block lower and block upper parts of the tomography operator, and b is the tomography right hand side vector. The algorithm is organized in two main steps: assembling the transformed right hand side vector y, and block lower triangular system solution layer by layer. The transformed right hand side vector (13) can be computed as follows:

$$[y]_{N_{\mathrm{ps}}} = [b]_{N_{\mathrm{ps}}}; \psi = 0$$
$$\text{For } k = N_{\mathrm{ps}} - 1, \cdots, 1$$

$$\psi := \psi + \mathrm{Diag}(\Gamma^{\mathrm{T}} W_1 \Gamma, \cdots, \Gamma^{\mathrm{T}} W_{N_{\mathrm{lgs}}} \Gamma) \begin{bmatrix} \left[\widetilde{H}_{\mathrm{atm}} \right]_{1k+1} \\ \vdots \\ \left[\widetilde{H}_{\mathrm{atm}} \right]_{N_{\mathrm{lgs}}k+1} \end{bmatrix} \left[x^{(0)} \right]_{k+1}$$

$$[y]_k = [b]_k - \begin{bmatrix} \left[\widetilde{H}_{\mathrm{atm}} \right]_{1k}^{\mathrm{T}} & \cdots & \left[\widetilde{H}_{\mathrm{atm}} \right]_{N_{\mathrm{lgs}}k}^{\mathrm{T}} \end{bmatrix} \psi$$

End

whereas the block lower triangular system (12) is processed as follows:

$$[\widehat{x}]_1 = \mathrm{Solve}([A]_{11}, [y]_1)$$
$$\psi = 0$$
$$\text{For } k = 2, \cdots, N_{\mathrm{ps}}$$

$$\psi := \psi + \mathrm{Diag}(\Gamma^{\mathrm{T}} W_1 \Gamma, \cdots, \Gamma^{\mathrm{T}} W_{N_{\mathrm{lgs}}} \Gamma) \begin{bmatrix} \left[\widetilde{H}_{\mathrm{atm}} \right]_{1k-1} \\ \vdots \\ \left[\widetilde{H}_{\mathrm{atm}} \right]_{N_{\mathrm{lgs}}k-1} \end{bmatrix} [\widehat{x}]_{k-1}$$

$$[v]_k = [y]_k - \begin{bmatrix} \left[\widetilde{H}_{\mathrm{atm}} \right]_{1k}^{\mathrm{T}} & \cdots & \left[\widetilde{H}_{\mathrm{atm}} \right]_{N_{\mathrm{lgs}}k}^{\mathrm{T}} \end{bmatrix} \psi$$

$$[\widehat{x}]_k = \mathrm{Solve}([A]_{kk}, [v]_k)$$

End

where the function $x = \mathrm{Solve}(A, b)$ provides an exact solution of the generic system $A x = b$ with Cholesky forward/back substitutions (CBS).

Layer solution via CBS does not operate on A, but rather on the Cholesky factor of a re-ordered version of A. It is by far the most economical algorithm in terms of the number of operations on account of the high sparsity of the N_{ps} Cholesky factors. However, it requires to store these factors in cache, and partially sequential forward/back substitutions for each layer. The Cholesky factors are updated at a rate of 0.1 Hz as turbulence conditions and range to the LGS change.

- BGS-CG20: Block Gauss-Seidel with 20 iterations of Conjugate Gradient (CG) instead of the above Cholesky backsubstitutions implemented in the Solve(A, b) function. This approach has the advantage of being much more parallelizable. 20 CG iterations per layer provide optimal performance. The algorithm is well suited to operate on square grids in the cone coordinate system.

- CG30: 30 CG iterations operating on the whole tomography system. The algorithm is well suited to operate on square grids in the cone coordinate system. Simulation results indicate that around 30 iterations provide optimal performance, although 20 may still be acceptable with a $\sim$ 20 nm RMS performance degradation penalty in quadrature.

- FD3: 3 iterations of Conjugate Gradient with a Fourier Domain preconditioning matrix.[10] The FDPCG algorithm operates in the spatial domain with a sparse pre-computed Fourier Domain Hermitian preconditioning matrix $\widehat{M}$ updated at a rate of 0.1 Hz as the range to the layers, range to the LGS, SNR, and turbulence conditions change. Simulation results indicate that 3 iterations provide optimal performance. The preconditioning step is implemented in parallel via N_{ps} ffts, a Hermitian matrix vector multiplication and N_{ps} inverse ffts. The algorithm must operate on square grids in the cone coordinate system.

The following points are worth mentioning: simulation results indicate that an additional gradient tip/tilt removal step inside the tomography matrix provides only a modest (a few nm) performance improvement; warm restart (the solution from a given frame is used as the initial guess for the next frame) is used in all these iterative algorithms, and is crucial

since it reduces the number of iterations significantly; BGS-CG20 and CG30 have virtually memory-free implementations and are order N, where N denotes system size.

3.5. Propagating the LGS Tomography Estimate

The first step of LGS fitting consists in the geometrical propagation of the LGS tomography estimate (ray trace interpolation/accumulation) along N_{evl} evaluation directions from sources at infinite range, which can be represented as follows:

$$[y]_k = \left[\begin{array}{ccc} [H_{\text{atm}}]_{k1} & \cdots & [H_{\text{atm}}]_{kN_{\text{ps}}} \end{array} \right] \left[\begin{array}{c} [\widehat{x}]_1 \\ \vdots \\ [\widehat{x}]_{N_{\text{ps}}} \end{array} \right], \qquad 1 \le k \le N_{\text{evl}} \tag{14}$$

where $[\widehat{x}]_l$ denotes the tomography estimate for layer l, and $[H_{\text{atm}}]_{kl}$ is a sparse bilinear interpolation operator between the nodes of atmospheric grid l and intercepts of rays traced to the aperture-plane grid from the evaluation direction k. The absence of tilde on H indicates ray tracing from infinite range sources. Note that the interpolation offsets (β_x, β_y) are not constant on a given phase screen, since atmospheric phase screens are in the cone coordinate system and ray tracing is from infinite range sources. However, the offsets need to be stored for only 1 row/column of grid points for each infinite range source, and are updated at a rate of 0.1 Hz as the range to the LGS and the choice of evaluation directions vary. Typically, ray tracing is done in $N_{\text{evl}} = 9$ evaluation directions.

3.6. Computing the LGS Fitting System Right Hand Side Vector

The fitting right hand side vector is computed as follows:

$$[q]_k = \omega_k W [y]_k, \qquad 1 \le k \le N_{\text{evl}} \tag{15}$$

$$[b]_l = \left[\begin{array}{ccc} [H_{\text{dm}}]_{1l}^{\text{T}} & \cdots & [H_{\text{dm}}]_{N_{\text{evl}}l}^{\text{T}} \end{array} \right] \left[\begin{array}{c} [q]_1 \\ \vdots \\ [q]_{N_{\text{evl}}} \end{array} \right], \qquad 1 \le l \le N_{\text{dm}} = 2 \tag{16}$$

where $[b]_l$ is the right hand side sub-vector for DM grid l, ω_k is the scalar weight for evaluation direction k, W is a fixed sparse aperture-plane weighting operator coupling each grid point with up to 8 nearest neighbors on a 3×3 stencil (fully interior stencils have Simpson weights), and $[H_{\text{dm}}]_{kl}$ is a sparse bilinear interpolation operator between the actuators of DM grid l and intercepts of the rays traced to the aperture-plane grid from the evaluation direction k. The absence of tilde on H indicates projection from infinite range sources. The interpolation operator for DM number 1 at $h = 0\,\text{km}$ is fixed and independent of the evaluation direction; the interpolation operator for DM number 2 is a function of the evaluation direction, but the interpolation offsets are constant for all points in the aperture plane grid. These offsets (β_x, β_y) are updated as the choice of evaluation directions vary.

3.7. Computing an approximate solution to the LGS Fitting System

The LGS fitting matrix system to solve has the form $A\,x = b$, where x denotes the concatenated DM actuator vector of unknowns, b is the concatenated fitting right hand side vector, and A is the block-structured fitting operator. The proposed fitting solver consists of 4 CG iterations operating on the whole fitting system. The solver performs $[A]_{kl}\,[v]_l$ operations ($1 \le k, l \le N_{\text{dm}} = 2$), which can be expressed as follows:

$$\psi = \text{Diag}(\omega_1 W, \cdots, \omega_{N_{\text{evl}}} W) \left[\begin{array}{c} [H_{\text{dm}}]_{1l} \\ \vdots \\ [H_{\text{dm}}]_{N_{\text{evl}}l} \end{array} \right] [v]_l, \tag{17}$$

$$[A]_{kl}\,[v]_l = \left[\begin{array}{ccc} [H_{\text{dm}}]_{1k}^{\text{T}} & \cdots & [H_{\text{dm}}]_{N_{\text{evl}}k}^{\text{T}} \end{array} \right] \psi. \tag{18}$$

3.8. Tip/Tilt and Tilt Anisoplanatism Mode Removal

The final step of LGS wavefront reconstruction consists in removing the 5 TT/TA modes from the fitting estimate. Mode coefficients are estimated by performing an inner product with the modal matrix as follows:

$$c = \Sigma\, M^{\mathrm{T}}\, \widehat{x}\,, \tag{19}$$

where M is the modal matrix, Σ is the inverse of the 5×5 modal cross-coupling matrix $(M^T M)$, and $\widehat{x}$ is the fitting estimate. The mode-removed DM actuator vector is then computed as:

$$\overline{x} = \widehat{x} - M\, c\,. \tag{20}$$

The matrices M and Σ are a function of the range to the sodium layer and are updated at a rate of 0.1 Hz.

4. RTC COMPUTATION AND MEMORY REQUIREMENTS

This section describes the computation and memory requirements of the most demanding RTC processes. These are the LGS WFS gradient computation and the LGS wavefront reconstruction processes, which operate at 800Hz. In comparison, the other processes have very small requirements and are not described in this paper. Reconstructing the AO-corrected PSF in real time is not a RTC requirement as the moment, but is contemplated as a goal. The requirements of this process are very challenging and could easily drive the processing and memory requirements of the RTC.

	Memory (MB)	Number of operations (MMAC) per frame
LGS WFS processing	9.5	1.5
LGS wavefront reconstruction		
BGS-CBS	41.6	39.5
BGS-CG20	1.6	204.5
CG30	1.6	195.6
FD3	39.3	208.5

Table 2. Memory and processing requirements of the LGS WFS gradient computation and LGS wavefront reconstruction processes. Two bytes fixed point arithmetic has been used to estimate the memory for the real time LGS WFS gradient processing. Four bytes floating point arithmetic has been used to estimate the memory for the background tasks, which optimize the LGS WFS gradient computation. Four bytes fixed and floating point arithmetic has been used to estimate the memory for the LGS wavefront reconstruction process.

5. CONCLUSION

The NFIRAOS RTC requirements and the functional block diagrams have been updated. Four possible implementations of a minimum variance reconstruction algorithm have been defined together with their estimated processing and memory requirements. However, more work needs to be done to finalize our requirements and our block diagrams, in particular for the control of the LGS zoom corrector and the optimization of the NGS WFS gradient computation algorithm. Conceptual design studies will be conducted during the next 6 to 9 months. The purpose of these studies will be (i) to select a wavefront reconstruction algorithm, (ii) to define an hardware architecture optimally suited for the chosen algorithm, based upon existing processor technologies such as Digital Signal Processors (DSP) and Field Programmable Gate Arrays (FPGA) and (iii) to demonstrate that the combination of the chosen algorithm and the hardware architecture meet the end-to-end 1000μs latency requirement for the NFIRAOS RTC.

ACKNOWLEDGEMENTS

REFERENCES

1. G. Sanders and J. Nelson, "The status of the Thirty Meter Telescope project," in Ground-based and Airborne Telescopes, L. Stepp and R. Gilmozzi ed., SPIE 7012, 2008
2. B. Ellerbroek and al., "Progress towards developing the TMT adaptive optical systems and their components," in Adaptive Optics Systems, SPIE Proc 7015, 2008.
3. D. Crampton, L. Simard and D. Silva , "Early Light TMT Instrumentation," in Ground-based and Airbone instrumentation for Astronomy II, SPIE Proc 7014, 2008.
4. G. Herriot and al., "NFIRAOS - TMT's initial adaptive optics system," in Adaptive Optics Systems, SPIE, Proc 7015, 2008
5. C. Boyer and al., "The TMT Laser Guide Star Facility Design Update," in Adaptive Optics Systems, SPIE, Proc 7015, 2008
6. C. Boyer and al., "TMT Adaptive Optics Systems Control Architecture," in Advances in Adaptive Optics II, SPIE, Proc 6272, 2006
7. S. Adkins, O. Azucena and J. Nelson, "The design and optimization of detectors for adaptive optics wavefront sensing," in Advances in Adaptive Optics II, SPIE, Proc 6272, 2006
8. R. Conan and al., "Modeling of the Thirty Meter Telescope matched filter-based wavefront sensing," in Adaptive Optics Systems, SPIE, Proc 7015, 2008
9. B. L. Ellerbroek, L. Gilles and C. R. Vogel, "Numerical simulations of multi-conjugate adaptive optics wavefront reconstruction on giant telescopes," Appl. Opt. 42, 4811-4818 (2003).
10. C. R. Vogel and Q. Yang, "Fast optimal wavefront reconstruction for multi-conjugate adaptive optics using the Fourier domain preconditioned conjugate gradient algorithm," Opt. Exp. 14, 7487-7498 (2006).
11. B. L. Ellerbroek and F. Rigaut, "Methods for correcting tilt anisoplanatism in laser-guide-star-based multiconjugate adaptive optics," J. Opt. Soc. Am. A 18, 2539-2547 (2001).
12. L.Gilles and B.L.Ellerbroek, "Split atmospheric tomography using laser and natural guide stars," submitted to J. Opt. Soc. Am. (2008).