PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

# The Real-Time controller (RTC) for the Narrow Field Infrared Adaptive Optics System (NFIRAOS) for TMT final design

Jennifer Dunn, Dan Kerley, Malcolm Smith, Ed Chapin, Glen Herriot, et al.

Jennifer Dunn, Dan Kerley, Malcolm Smith, Ed Chapin, Glen Herriot, Jean-Pierre Veran, Corinne Boyer, Lianqi Wang, Luc Gilles, Gelys Trancho, "The Real-Time controller (RTC) for the Narrow Field Infrared Adaptive Optics System (NFIRAOS) for TMT final design," Proc. SPIE 10703, Adaptive Optics Systems VI, 1070317 (10 July 2018); doi: 10.1117/12.2314226

SPIE.

Event: SPIE Astronomical Telescopes + Instrumentation, 2018, Austin, Texas, United States

# The Real-time Controller (RTC) for the Narrow Field Infrared Adaptive Optics System (NFIRAOS) for TMT Final Design

Jennifer Dunn*[a], Dan Kerley[a], Malcolm Smith[a], Ed Chapin[a], Glen Herriot[a], Jean Pierre. Veran[a], Corinne Boyer[b], Lianqi Wang[b], Luc Gilles[b], Gelys Trancho[b]

[a]National Research Council of Canada - Herzberg, 5071 West Saanich Rd, Victoria, Canada
[b]Thirty Meter Telescope, Suite 300 - 100 West Walnut Street, Pasadena, California

ABSTRACT

The Real-Time Controller (RTC) for the Thirty Meter Telescope (TMT) Narrow Field Infrared Adaptive Optics System (NFIRAOS) is the software and server hardware that calculates wavefront corrector commands from wavefront sensor measurements. The RTC is fed data from up to six Shack-Hartmann Laser Guide Star wavefront sensors (LGS WFS), one high-order Natural Guide Star Pyramid Wavefront Sensor (PWFS), up to three Shack-Hartmann On-Instrument wavefront sensors (OIWFS) that are located in the client science instruments, and up to four on-detector guide windows (ODGW) also in the client instruments. The RTC supports laser guide star multi-conjugate adaptive optics (MCAO) and natural guide star adaptive optics (NGS AO) observing modes. The RTC controls two deformable mirrors, one conjugated to 0km (DM0, mounted on a tip/tilt stage) and another to 11.8km (DM11). During the final design phase we used extensive prototyping to demonstrate that off-the-shelf servers using general purpose CPUs are able to support the maximum 800 Hz loop-rate at which the RTC is required to operate, with acceptable latency and jitter. Prototyping was also performed to compare the flexibility of CPU- and Xeon Phi-based architectures. This paper discusses evolution in the RTC from the Preliminary to Final Design Phases, emphasizing its modular architecture, and ability to accommodate a wide range of observing conditions.

Keywords: real-time control, adaptive optics, wavefront sensors, real-time, deformable mirrors

## 1. INTRODUCTION

The Thirty Meter Telescope's (TMT) facility Adaptive Optics (AO) system, which is called the Narrow Field Infrared Adaptive Optics System (NFIRAOS), feeds up to three science instruments on the telescope Nasmyth platform operating at TMT's diffraction limit. It is a multi-conjugate adaptive optics (MCAO) system that has two different modes: laser guide star adaptive optics (LGS AO), and natural guide star adaptive optics (NGS AO). Wavefront sensor measurements are provided by up to six Shack-Hartmann Laser Guide Star wavefront sensors (LGS WFS) and one high-order Natural Guide Star Pyramid Wavefront Sensor (PWFS) that can also act as a truth WFS when in LGS AO mode. To sharpen the image at the science focal plane there are up to three Shack-Hartmann On-Instrument wavefront sensors (OIWFS), and up to four on-detector guide windows (ODGW) located within the client science instruments. The RTC controls two deformable mirrors: one conjugated to 0km (DM0) and another to

Adaptive Optics Systems VI, edited by Laird M. Close, Laura Schreiber, Dirk Schmidt, Proc. of SPIE Vol. 10703, 1070317 · © 2018 SPIE
CCC code: 0277-786X/18/$18 · doi: 10.1117/12.2314226

11.8km (DM11). In order to accommodate slower, but larger-amplitude tip/tilt errors, DM0 is mounted on a tip/tilt stage (TTS) that acts as a "woofer", handling frequencies up to ~20 Hz, while the DM itself acts as a "tweeter".

Building upon the RTC Preliminary Design that was described in [1][2][3], this paper summarizes updated architectural details and prototyping results that were presented as part of the RTC's Final Design, which successfully passed its external review in December 2017. This involved external reviewers from key international organizations. Everything was reviewed from interfaces, requirements, design, prototyping, along with programmatic readiness in the form of quality assurance, reliability, failure mode and risk analysis, full cost and schedule. The primary performance driver of the RTC design is a matrix vector multiplication (MVM) that is used to derive wavefront corrector commands (primarily 3125 DM0 and 4548 DM11 actuator values) from wavefront gradients (5792 from each of the six LGS WFS in MCAO mode, and 14776 from the PWFS in NGS AO mode), at a rate of up to 800 Hz. We have shown, with confidence, that by spreading the ~905 MB (~7000 controlled actuators x 32400 illuminated subaperture gradients) control matrix across several computers, the MVM can be performed using existing commercial off-the-shelf (COTS) CPU-based servers and networking hardware, with acceptable latency and jitter. In addition, the RTC design team, in conjunction with the TMT AO group, has considered a wide range of guide star acquisition and observing scenarios, to ensure that the RTC is sufficiently flexible to undertake and synchronize a number of activities in parallel with the MVM. Some of these activities include: (i) tiered assignment of guide stars of varying brightness to different NGS WFS, to constrain different modes in the AO system (the "Low Order", or LO path); (ii) measurement and use of matched filters to obtain WFS gradients; (iii) collection of statistics, including WFS subaperture power spectral densities (PSD); (iv) projection of wavefront errors into DM shape space, and vice versa (e.g., to remove unobservable modes); and (v) calculation and offloading of modes [e.g. to the TTS, and Telescope Control System (TCS)].

Section 2 provides historical context for the final RTC design, including the evolution from earlier concepts based on FPGAs and GPUs. Section 3 gives an overview of the final RTC software design, including modular aspects that allow it to partition the work across multiple physical servers. Section 4 describes the RTC hardware, including prototyping work that demonstrated that either CPU-based or Xeon Phi-based servers could be used with minimal changes to the code. Section 5 summarizes the main benchmarks that demonstrate that the RTC meets its timing requirements.

## 2. RTC EVOLUTION

Throughout the development of the RTC, several different hardware platforms have been investigated, including field-programmable gate array (FPGA) based systems and commercial off the shelf (COTS) CPU-based servers, the latter both with and without graphics processing units (GPU). A previous trade study ("Results of the NFIRAOS RTC trade study" [3]), with updated analysis also presented at the December 2016 Paris RTC4AO4 workshop concluded that a purely CPU-based solution was within the capabilities of existing servers at the time, and would be the least complex, and result in reduced risks (e.g. schedule and obsolescence), though with an increase in hardware cost and power consumption over the other approaches.

Since the trade study, CPUs have continued to improve in performance (more than doubled), with increasing core counts and improved memory bandwidth. For a given level of performance, modern CPUs have substantially lower costs and power consumptions than those that were available several years ago, and used by our group for prototyping.

Based on these studies, the NFIRAOS RTC final design uses high performance COTS servers running standard Linux distributions (currently CentOS 7.x) with the real-time patch. The use of the real-time patch is required in order to reduce timing jitter to an acceptable level, and eliminates the need for a more specialized real-time operating system, such as VxWorks.

Most of the work since the preliminary design has been focused on hardware testing, which resulted in numerous refinements and validation of the final RTC design. One substantial change was to the way real-time data (e.g., from WFS) are stored for short-term analysis. Whereas the preliminary design stored the data locally, spread among the various servers involved in the real-time calculations (using spinning hard disks), the final design now incorporates a separate (and simpler), dedicated Real-time Telemetry Storage (RTS) system, enabled by Redundant Arrays of Independent Disks (RAID) composed of Solid State Drives (SSDs), since SSDs have rapidly increased in size and dropped in price in recent years. This system is currently split into two physical servers (to improve bandwidth), and receives data from the rest of the RTC servers over 40 Gb Ethernet and stores the data directly to the SSD arrays. The RTS has sufficient capacity to store all of the real-time data from a full night of AO operations. In addition, the RTS packages and serves data used by the Point Spread Function Reconstructor (PSFR, a separate system from the RTC), which generates estimated PSF maps throughout the night that are later used in post-processing.

Another substantial change was further re-factoring of the RTC algorithms (both the "critical path" algorithms running at up to 800 Hz to handle high-order (HO) and low-order (LO) wavefront corrections, and the slower background tasks, such as matched filter optimization and truth WFS calculations) so that they are more efficiently split across physical servers. For example, the design was updated so that PWFS pixels are always sent to the same place, where they are calibrated and converted into gradients, rather than two different locations depending on the mode (LGS vs. NGS) as they were in the preliminary design. In NGS mode, the PWFS gradients are sent on to the High Order Processing (HOP) Servers that perform the HO MVM.

## 3. RTC

### 3.1 Overview

The RTC is composed of the real-time software and hardware that will process inputs from all of the wavefront sensors, compute the commands for the wavefront correctors, and offload telescope modes to the TCS. Running at a rate of up to 800 Hz, NFIRAOS will provide wavefront corrected light to one of three mounted science instruments.

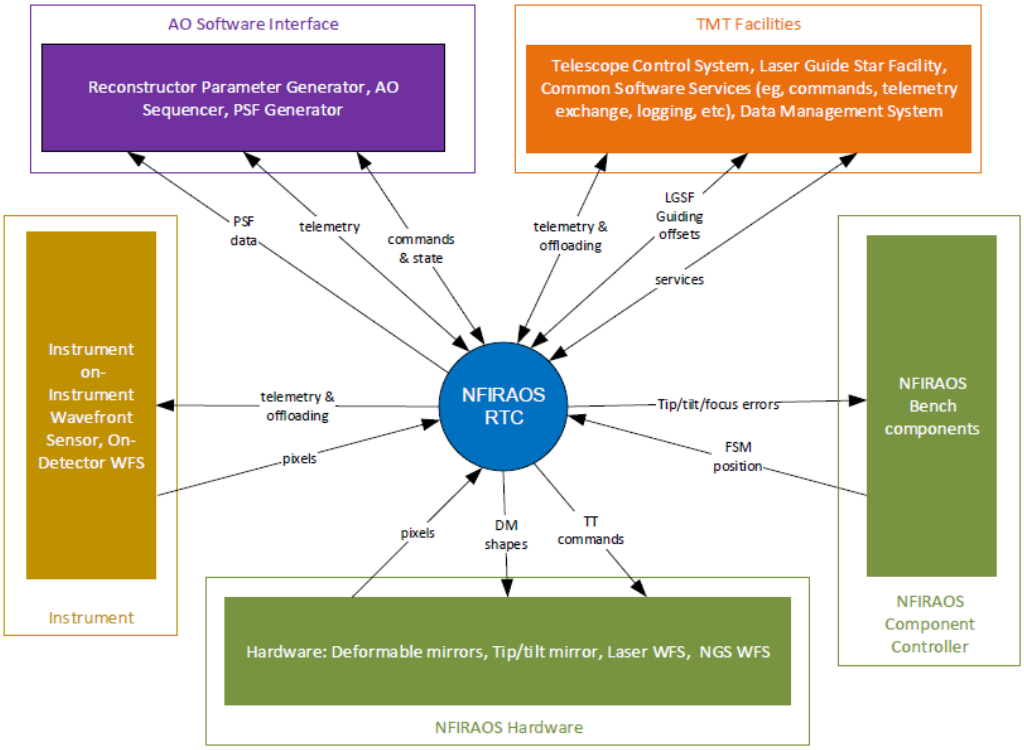


*Figure 1 - RTC Context Diagram*

Figure 1 shows an RTC in context with other NFIRAOS and TMT subsystems. The control of the RTC is hierarchical, with the TMT Adaptive Optics Sequencer (AOSQ) orchestrating the commands and configuration required to start the RTC and other subsystems involved in AO. Experience has taught us that having a highly-configurable AO system, particularly including the ability to change the sequence of events that lead to closing the AO loops, is essential when commissioning a new instrument. To this end, the RTC has been designed as a "tool box" with many different low-level commands, and therefore highly fine-grained control, made available to the AOSQ.

Complementing the real-time activities of the RTC is the Reconstructor Parameter Generator (RPG). This separate subsystem uses statistical data collected by the RTC to infer the atmospheric turbulence profile, and to periodically update a number of control parameters, the most important being the control matrix used by the HO MVM, so that the RTC can optimally derive wavefront corrector demands given the current observing conditions. This critical duplex communication is enabled by a dedicated high-bandwidth, low-latency network and protocol connecting the two subsystems directly. A third subsystem, the PSFR, mentioned in the previous section, combines turbulence models with data collected by the RTC to reconstruct the effective PSFs over the field of view throughout the night.

The TMT Common Software (CSW) services and framework are used to implement most of the interfaces with other subsystems, the most notable exception being the dedicated high-speed connection mentioned above. CSW is primarily a software communications backbone that allows the sharing of telemetry information, the transmission of commands, and raising alarms. The RTC uses CSW

to transfer lower-rate data (e.g., 20 Hz), such as the "offload telescope modes" (including tip-tilt, focus, magnification primary mirror scalloping, and other telescope modes) which the RTC publishes, via the CSW Event Service. The TCS will subscribe to these telemetry data, and relay the appropriate modes to the dedicated telescope subsystems (primary, secondary, tertiary and mount).

Operations of the slower NFIRAOS mechanisms on the NFIRAOS bench are handled by a separate software package called the NFIRAOS Components Controller (NCC). The RTC uses the CSW Event Service to provide the NCC with offloaded measurements that are used to adjust some of the slow real-time components, including focus errors for the LGS trombone mechanism, and tip/tilt/focus (TTF) errors for the PWFS Star Selection Mechanism (SSM), which are used to direct NGS light on to the PWFS.

The RTC has direct hardware interfaces, via a dedicated high-bandwidth Ethernet switch, with the two deformable mirrors (ground layer DM0 and high altitude DM11), six LGS WFSs, the PWFS, and the tip/tilt stage (TTS) to which DM0 is mounted. The TTS compensates for slow but high-amplitude tip/tilt errors that would saturate the DM. The RTC also has a dedicated high-speed connection to the TMT's laser guide star facility (LGSF), which generates a fixed asterism of six artificial guide stars by illuminating the sodium layer. In LGS mode, pixels from the LGS WFSs are used to perform tomographic reconstruction of the atmospheric turbulence in the volume above the telescope, while in NGS mode, pixels from the PWFS are used instead.

There are up to three science instruments connected to NFIRAOS. The RTC can accept input from up to three On-Instrument WFSs (OIWFSs) located in NFIRAOS client instruments; and up to four On-Detector Guide Windows (ODGWs), also located in NFIRAOS client instruments. These pixels are sent using the high-speed network and protocol to the RTC for incorporation into the low order wavefront correction. NFIRAOS and the connected client instruments reside on the Nasmyth platform, with the science instruments including rotators to perform de-rotation of the sky. The RTC detects and feeds-back rotational errors to those rotators. The RTC can also sense errors in OIWFS probe positions, and provide offsets to their positioning mechanisms (depending on the operational mode). These slow corrections are sent using the CSW Event Service.

### 3.2 RTC Server Roles

The RTC design described above is realized by spreading processes across different machines. The primary division of labor is to accommodate the HO MVM. With existing technology, it is practical to compute the full component of the DM commands corresponding to a single LGS WFS on a single quad-socket server, therefore a total of six servers are needed in the case of NFIRAOS. However, modularity in the design makes it straightforward to either increase or decrease the number of HOP servers depending on the size of the MVM. The partitioned HO MVM, as well as numerous other smaller processes are then grouped into server "roles", which would run concurrently on a single server with a particular hardware configuration. In the final design, the roles and the processes that they host are summarized in the following bullets:

- High Order Processing (HOP) [x6] : accepts LGS WFS pixel streams at up to 800Hz (or PWFS gradients in NGS mode) and performs high order wavefront reconstruction

- Wavefront Corrector Controller (WCC): performs PWFS pixel processing; low order reconstruction; high order wavefront summation; filtering, clipping and output of DM and TTS commands; generation of other low-order offloading outputs.
- Telemetry and Engineering Display (TED): interface with TMT services (commands, events), along with web-based interface to the Real-time Telemetry Storage.
- Real-time Telemetry Storage (RTS) [x2]: store real-time data streamed from the other servers (up to one night raw LGS WFS pixels and telemetry data, 2 GB/s or ~70 GB/night per server), fault tolerant, also provided data for real-time displays and PSFR data, and finally provides ability to "re-play" stored data for other RTC servers.
- Test server: provide interfaces (pixel sources, command sinks, external interfaces) to mimic the timing behavior of external systems.
- Calibration Software: early integration calibration
- Telemetry Analysis Computer (TAC): provide engineering interface for diagnostic data processing

### 3.3 Flexible Design

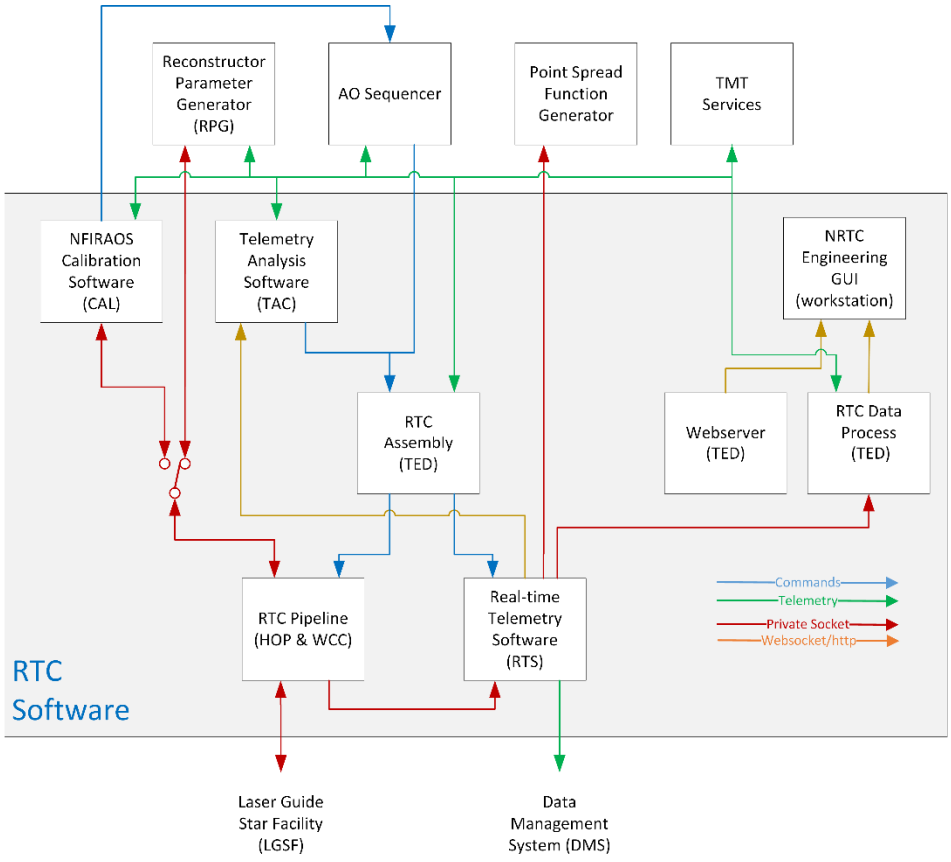


*Figure 2 - RTC Software Layout*

Figure 2 shows the layout of the RTC software. The grey box shows

- The “Calibration Software” (CAL) is used extensively throughout integration and commissioning, for purposes such as driving mechanisms and measuring coefficients for pointing models. CAL software is initially developed as part of the RTC (and will run on one of the RTC servers), but will ultimately be folded into the AOSQ software for future operational use.
- No AO system is complete without telemetry analysis software or calibration software. “Telemetry Analysis Software” (TAC) is a platform to execute commands and provide interfaces to assist with engineering tasks such as system debugging and performance analysis.
- The “RTC Assembly” is the interface between the RTC software and the AO Sequencer. It provides a TMT CSW command interface, and is also used to publish and subscribe to telemetry via the CSW Event Service.
- The RTC will have its own “Engineering Graphical User Interfaces” (GUI’s) and “Webserver” that will be essential during build and integration for controlling and monitoring the system. The decision was made early that the GUIs would be browser-based, and only the Webserver would be hosted on an RTC server.
- “RTC Data process” provides real-time data for the GUIs.
- The “RTC Pipeline” provides the RTC control systems and runs on the HOP and WCC servers

Spare servers capable of hosting each role type are included in the RTC rack, and role assignment can be performed purely through the use of software commands (i.e., there is no need to physically change wires etc.). In this way, recovery from server hardware failures can be accomplished quickly, ensuring that the RTC will meet its ambitious uptime goal. This is important for TMT as the combination of NFIRAOS and the Infrared Imaging Spectrograph (IRIS)[4] will be the only operating science instrument at first light.

The RTC architecture uses a pipeline that has spigots that can be turned on and off based on the configuration selected and commands initiated. The pipeline is active once commanded and can sit and wait to receive pixels and perform any processing. Because the RTC utilizes multiple servers, there can be multiple converging data streams, and a “synchronization manager” ensures that these data are received within a certain window of time, and deals with expired synchronization points. There are four major synchronization points:

- vector assembly upon completion of high-order processing
- vector assembly upon completion of low-order processing
- vector assembly upon completion of low-order truth processing
- wavefront reconstruction completion (high and low-order)

Associated with each of those synchronization points are different (configurable) rates, as shown by the colored lines in Figure 3.

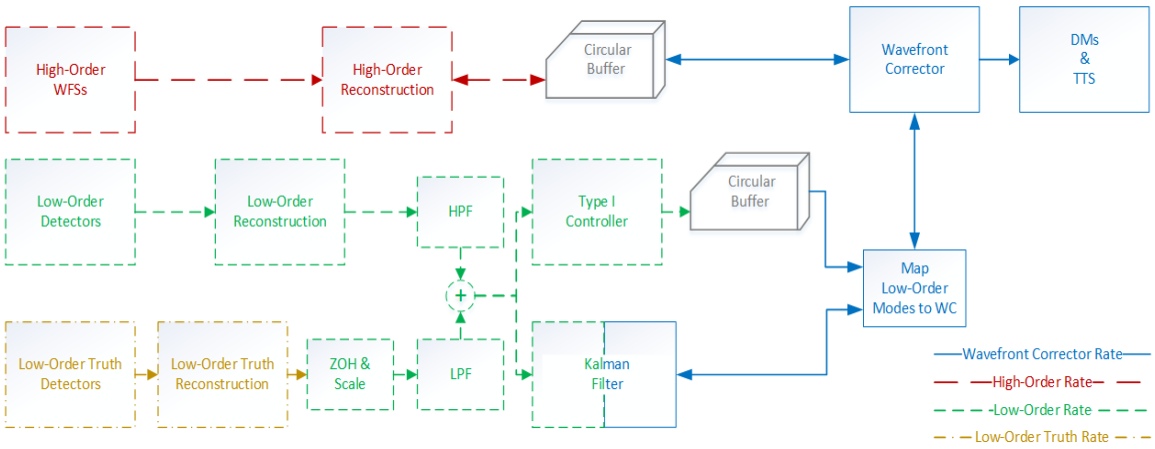


*Figure 3 - Showing different rates for the different loops*

The following sections provide further details of each of the two critical server roles, the HOP and WCC. Note that all servers also host a performance monitor to report their status.

### 3.4 High Order Processor (HOP) Functionality

The HOP servers perform the critical task of receiving LGS WFS pixels in LGS mode, processing those pixels into subaperture gradients, adding them to the current deformable mirror shape (converted into LGS gradient space) and feeding them into the High Order MVM reconstructor, which produces commands for the DMs and TTS.

These calculations are split across six HOP servers (one per LGS WFS) due to the sheer scale of the MVM. The control matrix has a size of ~905 MB (~863 MiB) resulting from ~7,000 controlled DM actuators, and ~32400 illuminated subaperture gradients. For comparison, the Gemini Multi-Conjugate Adaptive Optics System (GEMS), used on the 8-m Gemini South telescope has ~684 controlled actuators and ~2040 gradients, which results in a ~7.5MB control matrix (~7.1MiB or 5.3 MiB controlled), which could easily be accommodated on a single CPU.

When operating in NGS mode, the main difference is that the PWFS pixels are processed on a separate server (WCC), though the gradients are sent to four HOP servers (one per quadrant) to again be fed into the High Order MVM reconstructor.

In LGS mode, each HOP server also computes the full aperture tip/tilt for a single laser of the Laser Guide Star Facility (LGSF). These values are processed taking into account sensed positions already fed in, and converted to fast steering mirror commands for the LGSF. Also in LGS mode, the PWFS is used

as a truth wavefront sensor (TWFS) and sends modes to each HOP server to be converted into equivalent LGS WFS gradients.

Each HOP server sends a portion of the computed DM error vector to the WCC for command finalization. In LGS mode, the final DM shape is sent back from the WCC Server to each of the HOP Servers so that it can be converted into LGS WFS gradient space, and then added to the measured LGS WFS gradients to compute pseudo open-loop gradients.

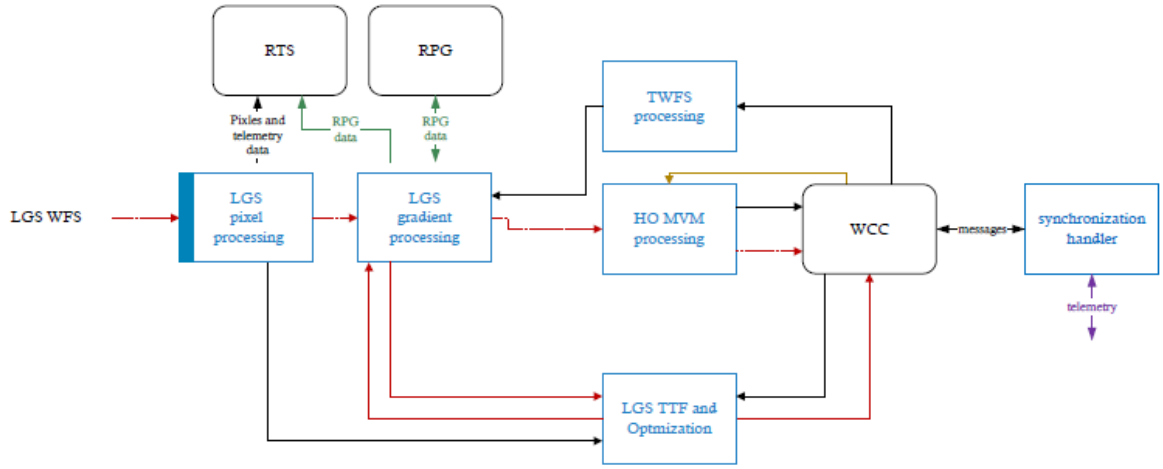


*Figure 4 - HOP processing, LGS mode*

### 3.5 Wavefront Corrector Controller (WCC) Functionality

The WCC is responsible for the final computation of the wavefront correction, processing PWFS pixels, and computing the LO mode corrections.

The finalization of the wavefront correction consists of:

- summing the high order HO DM error vectors from each of the HOP servers,
- adjusting the DM error vector summation based on the low-order modes,
- integrating the DM error vector,
- clipping the DM and TTS commands, if needed , and finally,
- sending the DM command vectors to the DM electronics and the tip/tilt command to the TTS.

The WCC also:

- processes pixel data from the OIWFSs and ODGWs and LO PWFS modes to perform the LO mode reconstruction and filtering,
- computes high-order PSDs from the computed DM errors after the removal of LO modes which are then used to optimize the loop gain, and
- computes the telescope modes for offloading to the TCS.

Regardless of whether the RTC is running in LGS or NGS mode, the WCC will process the PWFS pixels to compute gradients. These gradients are used to:

- compute truth wavefront sensor (TWFS) modes for the HOP servers in LGS mode (e.g., to compensate errors in the LGS system due to long-timescale variations in the structure of the sodium layer),
- compute HO POL gradients for the HOP servers in LGS mode, HOP (non-POL) gradients in NGS mode and,
- compute low-order modes in both LGS and NGS modes.

### 3.6 Configuration of the RTC

As mentioned above, it is important to have the ability to reconfigure the system quickly to facilitate integration and commissioning activities (or even future upgrades). For the core RTC Pipeline (as indicated by a box in Figure 2) the following components can be activated and configured:

- pipeline pixel processing,
- selection of the detectors that provide high- or low-order measurements, along with TT or TTF measurements,
- enabling high-order correction, and setting the loop rates,
- pseudo open-loop feedback,
- calibration of backgrounds,
- various controls loops, e.g., LGS focus, truth WFS matched filter, dithering of LGS or PWFS,
- selected gradient computation algorithms for LGS, PWFS, OIWFS/ODGW,
- offloading to the TCS,
- reset individual loop parameters,
- flatten DMs and TT Stage,
- manage RTS data storage

## 4. Hardware

The RTC contains a total of twelve quad socket Xeon servers. Each of the servers use Xeon Gold Scalable Processors and fall into one of three classes of machine: base server, network enhanced server or storage enhanced server. The base servers are used as the hardware for the HOP role and contain high core count CPUs. The network enhanced servers contain additional network connectivity and higher clock speed but lower core count CPUs compared to the base servers. The network enhanced servers are used for the WCC and Test roles. The storage enhanced servers contain additional SSD storage and RAID hardware and are used for the RTS role.

We also investigated servers based on Intel's Knights Landing (KNL) Xeon Phi processor, which contain 64-72 physical cores as an alternative to the quad socket base servers used for the HOP role. These processors include 16 GiB of integrated multi-channel DRAM which provides very high bandwidth (~450 GB/s) compared to motherboard based DRAM (~100 GB/s). This high bandwidth memory and the 512 bit vector units makes them very effective for the large MVM performed during the high order reconstruction. A server containing a single KNL Xeon Phi processor is able to perform the MVM within the required time. To reduce power consumption clock speeds and core complexity are reduced, compared to a regular Xeon, so sequential performance is reduced. The use of KNL based servers would significantly reduce the cost and power consumption of the HOP servers but the KNL Xeon Phi processor was not chosen as the baseline hardware because it increased system latency compared to the (more expensive) quad socket Xeon server. In addition, Intel has cancelled the successor to the Knights Landing processor, which puts the long-term future of the product line in question.

The server hardware specified at final design has not been purchased and the selection of the final hardware will be revised as available hardware improves. It is likely that improvements in hardware in the next year or two will allow the use of dual socket servers rather than quad socket servers. The use of dual socket servers is likely possible using current hardware for the network enhanced and storage enhanced servers. Using dual socket servers would significantly reduce cost and power consumption. The baseline design did not select dual socket servers since we have not yet benchmarked current generation (e.g. Xeon Gold and AMD Epyc) servers and specifying quad socket servers will definitely provide enough performance to meet the RTC requirements. In order to meet performance requirements while reducing cost and power consumption, careful evaluation of hardware options will be performed prior to purchasing the final hardware.

For AO systems with fewer DM actuators and WFS subapertures, dual socket servers would be more appropriate, with the number of servers being adjusted as necessary. The RTC software is written with the assumption that there are multiple NUMA (non-uniform memory architecture) regions per server so replacing a quad socket server with a dual socket server simply means updating a hardware configuration file which informs the server of core assignments and memory partitioning. A significant loss of performance can result if NUMA effects are ignored.

Internal communication between RTC servers is via multiple links per machine which all feed into a single private high speed (e.g. 40Gb Ethernet) switch. A 10Gb Ethernet switch provides connectivity to external TMT systems along with fiber connections to a third 10Gb Ethernet switch on the Nasmyth platform which is connected to the wavefront sensors and correctors.

Using Ethernet for all RTC inputs and outputs allows the Test server to act as the source of pixel streams and the sink for wavefront corrector commands, which allows exercising the RTC without requiring the use of the physical WFS controllers or DM electronics. This makes testing and benchmarking the RTC much simpler.

## 5. Prototyping and Performance

The most computationally demanding task of the RTC is the LGS mode large high-order MVM which must be performed each 800 Hz frame. In order to reduce the amount of time required to perform the MVM, the portion of the MVM corresponding to each LGS WFS is computed independently on a separate HOP server. Further, the LGS WFS pixel stream for each quadrant is sent as a separate stream and each quadrant is processed by one of the four CPUs within the HOP server. This allows each CPU and subsequently each HOP server to complete its portion of the MVM shortly after the last subapertures are converted into gradients.

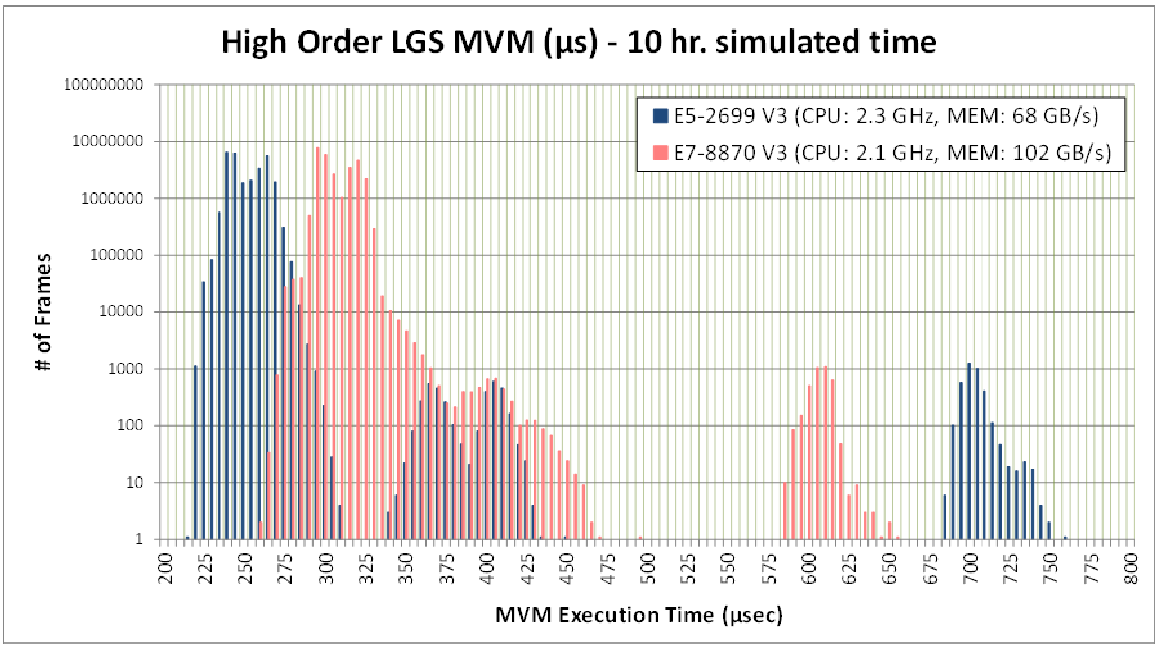


*Figure 5 - High Order LGS MVM from "Benchmarking hardware architecture candidates for the NFIRAOS real-time controller" SPIE 2014 [2]*

The figure above (Figure 5) shows the execution time histograms for the high order MVM of one quadrant using a Xeon E5-2699 V3 or a Xeon E7-8870 V3. At the right of the figure, the two peaks correspond to fetching a new control matrix; the E7 CPU with the greater memory bandwidth is ~100 µsec faster than the E5 CPU when the new control matrix is applied. When the control matrix resides within the CPU cache, the higher clock speed of the E5 CPU results in greater performance. It should be noted that the MVM is being performed column-wise as the gradients from the pixel processing become available. This allows the pixel reading and processing to run in parallel with the high order MVM. Since the pixel readout requires 500 µsec, the cached MVM performance is greater than required, but performing the MVM using a new (uncached) control matrix requires more time than the detector readout time. The benchmark represented in the figure does not include any prefetching of coefficients when the control matrix is updated so the peaks at 600 µsec and 700 µsec are pessimistic. After final hardware has been purchased, it will be determined if software prefetching is beneficial. Since the loop rate corresponds

to a period of 1250 µsec, and the MVM requires less than 500 µsec, there is ample time to prefetch control matrix coefficients.

Both CPUs used in the 'MVM only' benchmark (Figure 5) are two generations old and are significantly slower than a modern Xeon processor. As seen in the table below, a comparable Xeon Gold CPU has a higher clock speed and significantly higher memory bandwidth. The new Xeon CPUs also include 512 bit vector units compared to the 256 bit vector units of the older models. These CPU improvements will result in a faster MVM for both cached and uncached control matrices. At this time, the final RTC hardware has not been purchased and we have therefore not been able to benchmark the performance of the Xeon Gold CPUs.

*Table 1 - Clock speed and memory bandwidth for different CPU's used in prototyping*

| CPU | Clock Speed | Memory Bandwidth |
|---|---|---|
| E5-2699 V3 | 2.3 GHz | 68 GB/s |
| E7-8870 V3 | 2.1 GHz | 102 GB/s |
| Xeon Gold 6148 | 2.4 GHz | 128 GB/s |

The benchmark illustrated in Figure 5 only includes the high order MVM; it does not include reading pixels, performing pixel processing or sending results to a second server. Figure 6 below, is a more complete benchmark and includes the time required for the following steps:

- A server sends two quadrants of LGS WFS pixels to a dual processor HOP server.
- Pixels are sent as multiple datagrams spread over the 500µsec readout time using a single 10Gb Ethernet link.
- This was limited to two quadrants due to only having two CPUs to perform the MVM.
- HOP server reads pixels, calibrates pixels and computes gradients for the two quadrants.
- HOP server performs independent MVM for each quadrant and sends result to the test server. Rather than summing the two intermediate results on the HOP server, the HOP server sent both intermediate error vectors to the test machine for summation to increase the amount of Ethernet data transmission. A 10Gb Ethernet link was used to transfer the intermediate results.
- The test server reads computed DM error vectors (MVM output), sums the vectors and calculates the round trip time for each frame.

The simplified round trip benchmark was performed on dual socket versions of the first three generations of Xeon E5-2600 CPUs. Better performance is expected using a more modern CPU.

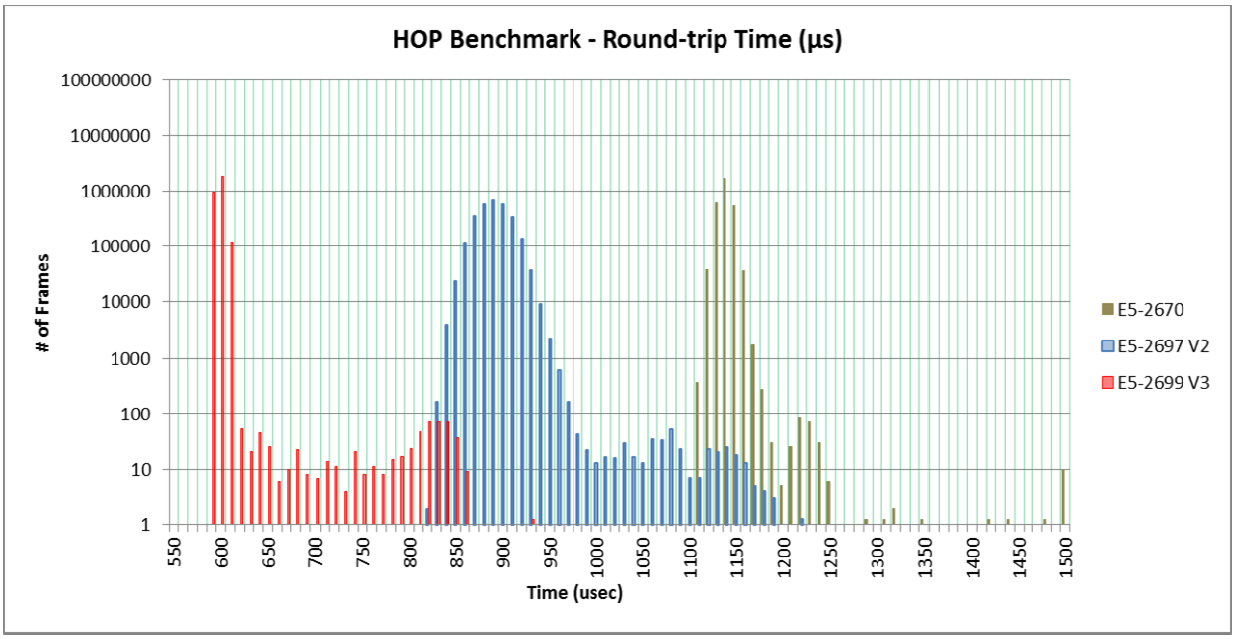


*Figure 6 – HOP MVM Benchmark showing round trip time from "Benchmarking hardware architecture candidates for the NFIRAOS real-time controller" SPIE 2014 [2]*

Figure 7 below shows the results of a Monte Carlo simulation of conservatively estimated execution times for the multiple quad socket Xeon Gold servers in the critical path of the RTC. The six HOP servers read pixels during the first 500 µsec of the frame (marked by the green line) with the pixel processing and MVM completing shortly after the last pixel arrives. The grey timings for the HOP servers represent the estimated worst case times and correspond to swapping in a new control matrix. The cached MVM times are sampled from a histogram produced using the E5-2699 V3 CPU round trip times. Notice that the Monte Carlo simulation results in the HOP servers transmitting their intermediate results to the WCC approximately 600 µsec after the start of the pixel readout which is consistent with the measured round trip times of approximately 600 µsec. For uncached control matrices, we used an MVM histogram corresponding to 100 µsec less than the round trip for the E5-2699 V3 CPU. We believe this is pessimistic since the Xeon Gold CPU has a higher bandwidth than the E7-8870 V3 CPU, which corresponds to the 100 µsec faster uncached MVM time.

Although execution time distribution of Figure 7 shows a noticeable tail corresponding to fetching the control matrices, it should be remembered that a new control matrix is only received once every ten seconds so there is one *slow* frame out of every 8000 and the impact on the AO performance is negligible. A logarithmic scale is used to show the details of the distribution; when a linear scale is used these rare events become invisible.

Additional benchmarking has also been performed, as listed below. These timings are included in Monte Carlo simulation when appropriate.

- Pixel processing including all meta-information (implementing code to parse the datagrams delivered by the VCAM WFS according to the published interface).

- CRC checksum calculation of each datagram is less than 1 μsec, resulting in less than 1 μsec additional latency for the system.
- Disk based telemetry storage has been replaced with SSD based storage. A system with three SSD based RAID arrays was tested using SATA SSDs. Each array sustained writes at a bandwidth exceeding 2.5GB/s. The telemetry storage requirements correspond to ~4GB/s total bandwidth, which will easily be met by the two RTS servers with two RAID arrays each.
- Sending intermediate DM error vectors from HOP servers to WCC was benchmarked using 40Gb Ethernet, which is the minimum speed of interconnect that the RTC will use for internal communication. The mean time to read six DM vectors was 75.6 μsec with a standard deviation of 1.9 μsec.
- Many transformation steps corresponding to smaller matrix vector multiplications (e.g. DM vector to low order modes, low order modes to DM vector, extrapolation of edge actuators) were benchmarked using simple single core dense and/or sparse routines. These timings are used in the extrapolation and DM pre-processing steps of the Monte Carlo simulation.
- An iterative method to perform DM actuator clipping and ensure that actuator stroke limits and inter-actuator stroke limits are not violated was also benchmarked. The DM clipping process was benchmarked using a single core (per DM) of an E5-2643 V4 system and takes an average of ~50 μsec during poor seeing.

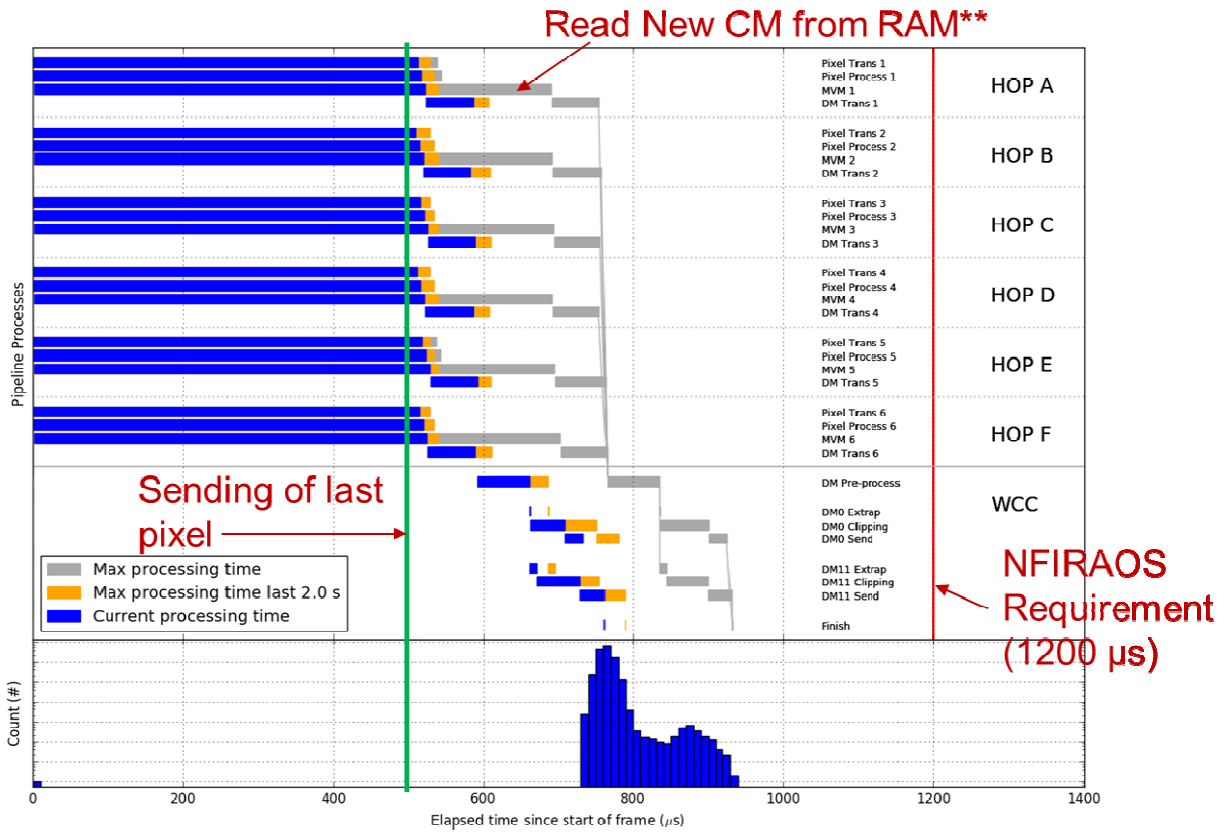


*Figure 7 - Monte Carlo Simulation. The green line shows the sending of the last pixels. Between the green and red line all processing needs to complete, and the histogram at the bottom shows that it completes well before the red line at the right which is the NFIRAOS requirement*

## 6. Conclusion

The RTC design has matured and many details have been added, all interfaces have been defined, benchmarking has been performed for the most critical requirements, verification plans have been defined, and much more. HAA successfully passed its Final Design Review in December of this year. Due to the hardware prototyping there is high confidence that a COTS CPU using a matrix vector multiplication (MVM) to derive wavefront corrector commands will be successful. The design of the control system is very flexible, which gives it the ability to reconfigure the system quickly to facilitate integration and commissioning activities. The hardware costs and power consumption have decreased over time and will likely decrease again, which means we will wait to buy the hardware until the last possible moment.

## 7. References


[1] Dan Kerley, Malcolm Smith, Jennifer Dunn, Glen Herriot, Jean-Pierre Véran, Corinne Boyer, Brent Ellerbroek, Luc Gilles, Lianqi Wang, "Thirty Meter Telescope (TMT) Narrow Field Infrared Adaptive Optics System (NFIRAOS) real-time controller preliminary architecture" in SPIE Proceedings Volume 9913: Software and Cyberinfrastructure for Astronomy IV, September 2014

[2] Malcolm Smith, Dan Kerley, Glen Herriot, Jean-Pierre Véran, "Benchmarking hardware architecture candidates for the NFIRAOS real-time controller" in SPIE Proceedings Volume 9148: Adaptive Optics Systems IV, September 2014

[3] Jean-Pierre Véran, Corinne Boyer, Brent L. Ellerbroek, Luc Gilles, Glen Herriot, Daniel A. Kerley, Zoran Ljusic, Eric A. McVeigh, Robert Prior, Malcolm Smith, Lianqi Wang, "Results of the NFIRAOS RTC trade study" in SPIE Proceedings Volume 9148: Adaptive Optics Systems IV, September 2014

[4] Edward L. Chapin, Jennifer Dunn, David Andersen, Glen Herriot, Dan Kerley, Takashi Nakamoto, Jimmy Johnson, Lianqi Wang, Gelys Trancho, Eric Chisholm, Brent Ellerbroek, Kim Gillies, Yutaka Hayano, James Larkin, Luc Simard, Mark Sirota, Ryuji Suzuki, Robert Weber, Shelley Wright, Kai Zhang, "The Infrared Imaging Spectrograph (IRIS) for TMT: closed-loop adaptive optics while dithering" in SPIE Proceedings Software and Cyberinfrastructure for Astronomy, June 2018