A prototype architecture for ELT-scale MCAO and LTAO based on many core CPU technologies (POSTER-WAV-REC-173)

David R. Jenkins[a,b], Alastair G. Basden[b], and Richard M. Myers[b]

[a]Gemini Observatory, 670 N. A'ohoku Place, Hilo, USA
[b]Durham University, Stockton Road, Durham, UK

ABSTRACT

The computational demands of AO RTC for next generation ELT-scale telescopes will increase massively compared to those of even the largest telescopes today. The effects of atmospheric seeing will have an even larger effect on the reduction of the effective diffraction limit of ELT-scale telescopes, making AO a necessary technology to ensure their scientific success. The AO RTC problem size scales to the fourth power of telescopes diameter requiring novel techniques and technologies to be investigated to provide the necessary capacity and performance. The ELT will use multiple laser guide stars to provide wide field AO capabilities either through the MCAO, LTAO or MOAO modes, increasing the requirements for the RTC even further. We present an update on a prototype hardware and software architecture for an ELT-scale RTC to process the MCAO and LTAO modes. The many-core CPU based architecture utilises an optimised and upgraded version of DARC and testing has been performed with simulated camera streams using an implementation of a UDP based ESO standard. We investigated the novel Intel Xeon Phi CPU architecture due to its high core count and high bandwidth memory as well as multi socket NUMA systems that can provide similar capabilities. The current performance capabilities of the RTC architecture will be presented for the SCAO, MCAO and LTAO modes based on specifications of first light ELT instruments. We have demonstrated SCAO on ELT-scales with an RTC latency of $< 450\mu s$. Similarly, for MCAO and LTAO we present our results with latencies of $< 1000\mu s$ for either system and with RMS jitter of $< 40\mu s$. Further optimisation and performance gains are expected. We have found that CPU based RTCs can deliver the performance, flexibility and maintainability necessary to ensure the success of next generation AO RTC systems.

Keywords: adaptive optics; real time control; simulation

1. INTRODUCTION

The next generation of Extremely Large Telescopes, including the Thirty Meter Telescope (TMT), ESO's Extremely Large Telescope (ELT) and the Giant Magellan Telescope (GMT), will all include adaptive optics as a fundamental aspect of telescope observations. This will be achieved by the use of facility wave-front sensors and deformable mirrors that come before the primary mirror in the telescopes optical path. The TMT and GMT will both have deformable secondary mirrors, with the GMTs secondary being segmented similar to its primary, and the ELT will have a deformable M4 mirror which comes before all science instruments. Due to the integrated nature of the deformable mirrors and the difficulty in maintaining the precise alignment of these enormous segmented telescopes, it is unlikely that any of the future ELTs will be able to provide a pure "seeing" limited mode with no adaptive correction of the wavefront. The efficient and seamless control of these AO systems is therefore paramount to the success of the ELT-scale observatories.

To maintain the WFS sampling density of current AO systems over the full primary aperture of the ELTs, the total number of WFS sub-apertures needs to scale with the second power of telescope diameter. This, coupled with the similar square scaling of actuator count over the wavefront correctors, gives an overall fourth power

Further author information: (Send correspondence to D.R.J.)
D.R.J.: E-mail: djenkins@gemini.edu, Telephone: 1 808 974 2569
A.G.B.: E-mail: a.g.basden@durham.ac.uk, Telephone: +44 (0) 191 33 42229

scaling of the reconstruction problem size to primary mirror diameter. For the next-generation ELTs with main telescope aperture diameters ranging from 24.5m to 39.3m, they will require up to 25x more WFS measurements over its pupil to achieve similar wavefront spatial sampling density as current state of the art observatories.

For an AO system to provide a desirable level of wavefront correction, the RTC needs to provide new DM commands in a timely and regular fashion. This means that both the latency and jitter of the RTC need to be as small as possible. The latency of an RTC is defined as the time delay due to the processing of wavefronts to DM command corrections and is measured from the time the last WFS pixel is received by the RTC and the time the DM command is made available to the mirror hardware. Therefore the RTC latency adds an overall delay to the AO correction loop and must be kept small to ensure that the corrections are applied while they are still valid for the observed atmospheric turbulence.

The jitter of the RTC can be defined as the deviation of the latency from the desired or target latency, this concept is shown in Figure 1. The RTC hardware is decided such that it can provide an RTC latency below a set requirement based on the degree of wavefront correction required of the AO system. However due to the non-determinant nature of computer systems, it is extremely difficult to ensure that this target latency is achieved on every frame. The jitter defines the distribution of latencies that are above the requirement and therefore have a negative impact on the wavefront correction of the AO system.

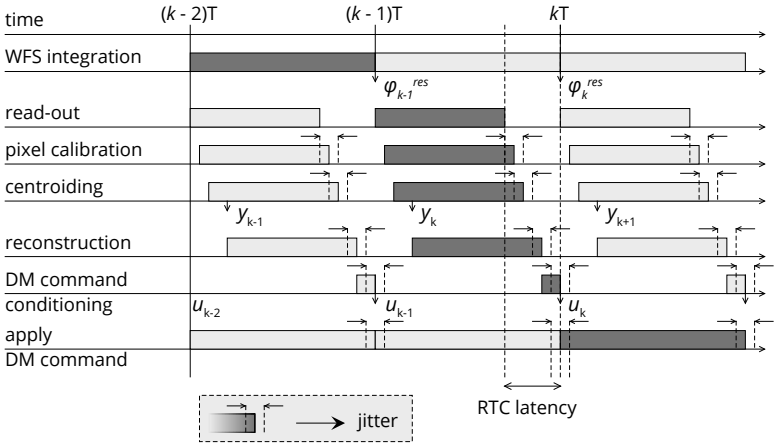

Figure 1. Two frame delay AO loop chronogram showing the overlapping computation times of the pipelined RTC operations. The jitter on the RTC latency is shown as the varying end times for each computation step which results in an overall jitter on the time the DM command is applied. The parameter k designates the time step, ϕ^{res} is the residual wavefront phase, y is the wavefront gradient vector, and u is the DM command vector.

Current AO systems use a range of different computing hardware to process the RTC, including CPUs, GPUs, FPGAs and DSPs. In the past FPGAs and DSPs have been used due to their high level of processing performance compared to contemporary CPUs/GPUs and due to their deterministic processing times which can help to reduce the overall loop jitter of the RTC. These devices however have several drawbacks which make them less than ideal for an AO RTC, especially now that that the processing performance of CPUs and GPUs has been massively improved over the last decade. The main drawback of an FPGA and DSP AO system is the difficulty of AO scientists and engineers to design and develop software for them as they typically don't use standard software languages or instruction sets. It is also difficult to find commercial off the shelf (COTS) devices that can meet the AO RTC requirements and have a long enough life time for an observatory AO RTC system.

The processing performance of standard x86-64 CPU systems has increased dramatically over the last decade with constant development due the ubiquity of the architecture in all computing related industries. This means that modern server CPU systems are more than adequate to perform the processing of current AO RTC systems and have also been demonstrated to be capable of ELT-scale operation,.[1] GPUs are also being actively investigated for their use in AO RTCs for ELT-scale telescopes, however due to GPUs requiring a host CPU, the

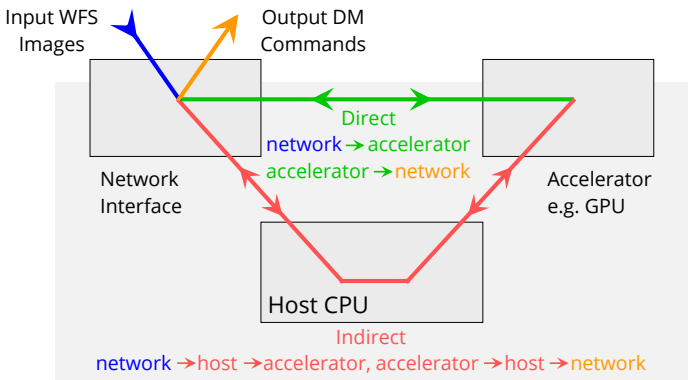

Figure 2. A comparison of the standard indirect data transfer through the host CPU and the more efficient direct to accelerator transfer scheme. A downside to the latter is that it currently involves non-portable software and the use of proprietary libraries.

standard method of data transfer to the GPUs involves an indirect transfer through the host CPU as shown in Figure 2.

There are novel methods to mitigate this, as demonstrated by Ref. 2, however the methods currently require custom hardware and software to directly transfer the data from the network interface to the GPU devices as shown by the "Direct transfer" route in Figure 2. Due to the non-standard nature of the direct transfer to GPU hardware, we instead continue to perform the RTC calculations on the host CPU to remove this added transfer latency.

2. METHODS

2.1 DARC

DARC is a fully modular on-sky tested AO RTC, which has been used extensively for the CANARY project. There are many publications available outlining its functionality, on-sky tested modes, and results.[3] It operates in a multi-threaded configuration for improved performance as AO RTC can be very parallel. It has also been optimised for Many Core and NUMA CPU systems.[1] An overview of the processes of an AO RTC such as DARC are shown in Figure 3.

2.2 MCAO and LTAO architecture

To achieve the required performance of the AO real-time control system for ELT-scale instruments, the SCAO architecture presented in Ref. 1 is scaled up with a single processing node used for each of the MCAO/LTAO LGS WFS as shown in Figure 4.

2.3 Multi-node SCAO

The multi-node architecture for MCAO and LTAO can be re-applied to the SCAO case to accelerate the ELT-scale problem by the use of multiple processing nodes per WFS camera, as shown in Figure 5. This improved the maximum achievable frame rate with the Xeon Phi processor and has the potential to reduce the RTC latency with modern many-core CPU systems.

2.4 Testing Configurations

The prototype architecture was tested using MAORY-like and HARMONI-like configurations, to demonstrate ELT-scale suitability with the most pertinent parameters being,

- 80x80 subaperture LGS WFS
- 5316 actuator M4 + 2 act M5

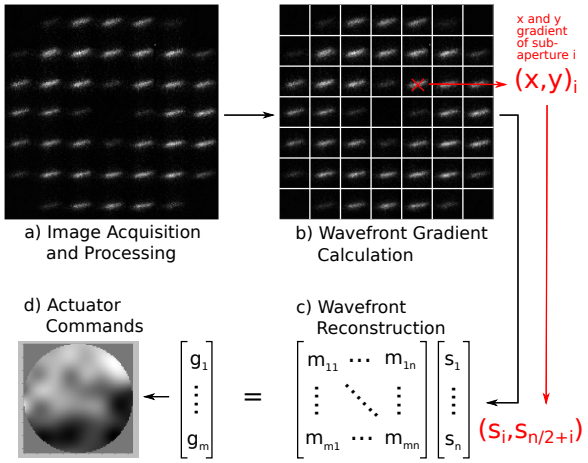

Figure 3. A figure showing the basic RTC operations, including a) image acquisition and processing (background subtraction, flat field application and threshold application), b) local wavefront gradient computation (using a centre of gravity algorithm), c) wavefront reconstruction (using a Matrix Vector Multiplication, MVM) and d) output of actuator commands. A thread will process a defined set of subapertures from beginning to end. For each subaperture, the local wavefront gradients are placed in a slope vector such that all the x gradients come first and then the y gradients ($(x,y)_i \longrightarrow (s_i, s_{n/2+i})$). The result of the MVM is a vector of actuator commands which can be reformatted to show the resulting shape of the correcting element.

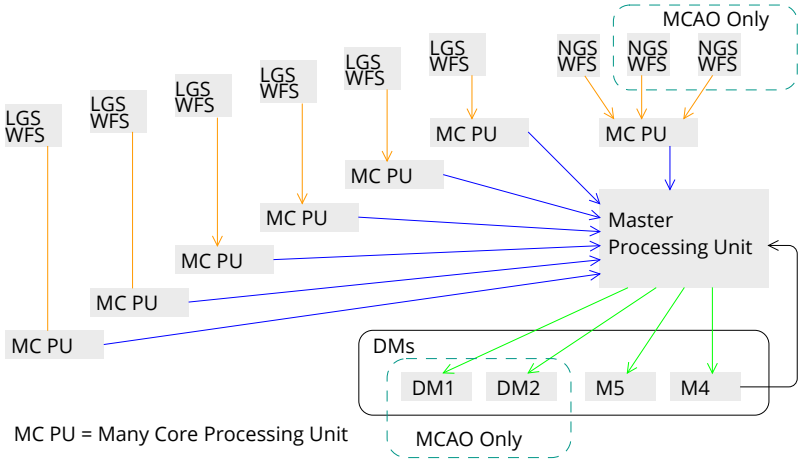

Figure 4. The proposed architecture for the MCAO and LTAO multi node RTC. The 6 LGS WFSs are each processed by a single many-core node. The NGS are all processed on the same node, three for MAORY and one for HARMONI. The master node sends out the final DM commands once it has finished summing and processing the partial vectors. The master node should also able to receive feedback from the ESO ELT M4 to integrate the actual M4 shape used in the next command.

- For MCAO + 2x 500 act DM.

We also investigated the use of the implicit POLC method as described in Ref. 4.

3. RESULTS

Results for the full MCAO and LTAO architecture described above are shown in Table 1 which are also presented in Ref. 5. Also included are results of varying the number of WFSs in the MCAO configuration.

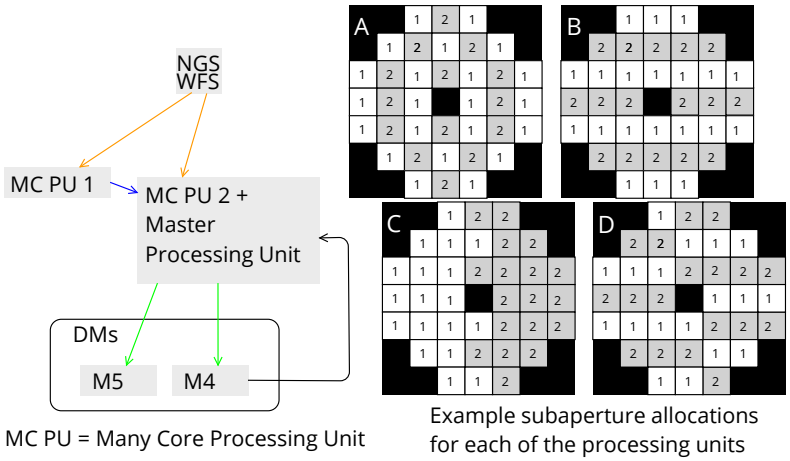

Figure 5. The architecture of the multi-node SCAO system setup as described in Section 2.3, this is an example for the ESO ELT. The WFS is multicast to the two processing units each of which process half of the total WFS subapertures, the master processing unit receives the partial DM vectors before combining them and delivering them to the DM.

Table 1. Latency, RMS jitter and largest outliers results for the full MCAO and LTAO architectures described in this paper. For all results the a) columns correspond to results from 1.5×10^5 continuous iterations at 500 Hz for a total time of 300 s for each test case. The b) columns correspond to results from a subset of no less than 2×10^4 continuous iterations chosen from the larger a) data sets for a total time of 40 s each. The b) column subsets were chosen to avoid any large outliers that result from simulated camera delays to give a better representation of the "steady" latency.

	Mean	RMS		Largest	
	Latency	Jitter (μs)		Outlier (μs)	
AO Mode	(μs)	a)	b)	a)	b)
Full MCAO	985	33	29	4465	1235
Full LTAO	894	29	28	4434	1174
MCAO telemetry	1085	32	30	2988	1466
MCAO POLC	1090	45	44	2880	1312
MCAO 6 LGS	992	47	46	3498	1143
MCAO 5 LGS	979	46	44	2870	1261
MCAO 4 LGS	969	45	45	3305	1285
MCAO 3 LGS	943	43	42	2817	1182
MCAO 2 LGS	951	42	43	2858	1119

3.1 Performance Comparison of Different CPUs

In Figure 6 we show the relative latency performance of different many-core CPU platforms that are suitable for ELT-scale AO RTC. It includes the single socket Xeon Phi processors alongside dual/quad socket AMD and Intel solutions.

4. CONCLUSIONS

We have demonstrated a many-core CPU-based AO RTC architecture for ELT-scale instruments. The results show that the Xeon Phi platform is able to obtain suitable performance for the operation of the MCAO and LTAO systems types. Other many-core CPU systems show potential for even greater performance. Future work would involve investigating an MOAO architecture and developing the MCAO/LTAO architecture for on-sky use with the ESO ELT instruments.

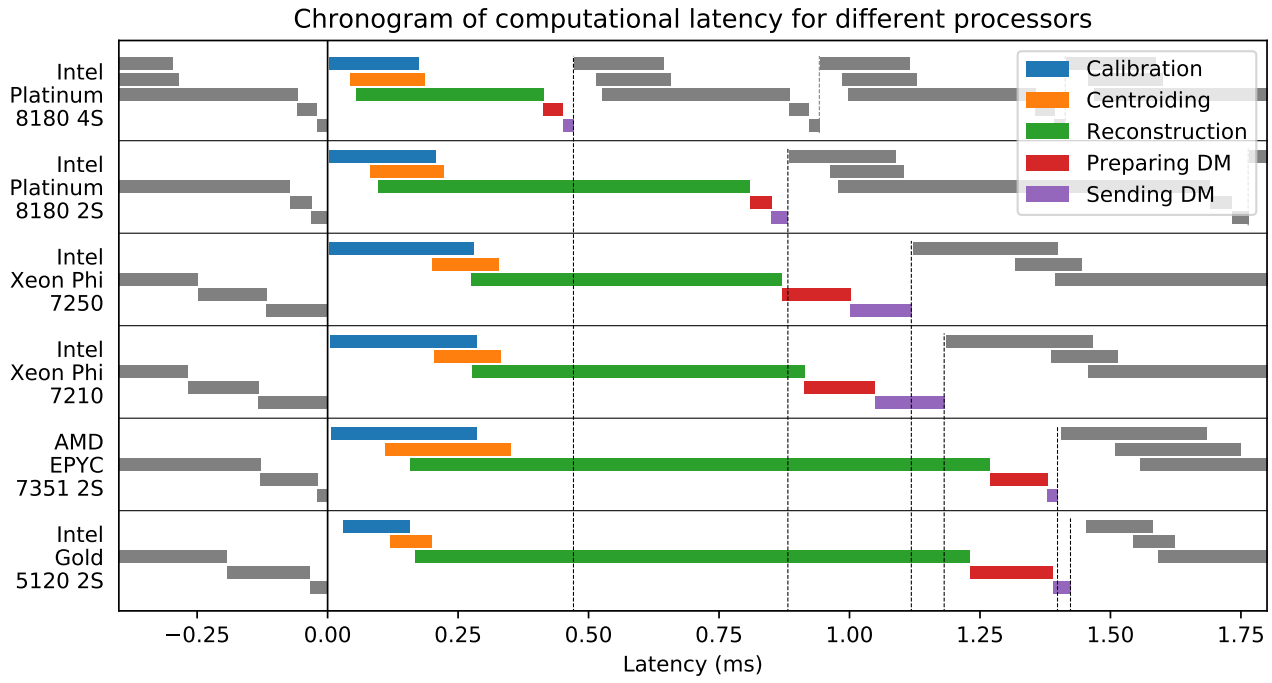

Figure 6. A comparison of the computational latency for different hardware platforms.

ACKNOWLEDGMENTS

REFERENCES

[1] Jenkins, D. R., Basden, A., and Myers, R. M., "Elt-scale adaptive optics real-time control with the intel xeon phi many integrated core architecture," *Monthly Notices of the Royal Astronomical Society* **478**(3), 3149–3158 (2018).

[2] Gratadour, D., "The cosmic rtc platform," (2018).

[3] Basden, A., Geng, D., Myers, R., and Younger, E., "Durham adaptive optics real-time controller.," *Applied optics.* **49**, 6354–6363 (November 2010).

[4] Basden, A. G., Jenkins, D., Morris, T. J., Osborn, J., and Townson, M. J., "Efficient implementation of pseudo open loop control for adaptive optics on extremely large telescopes," **486** (03 2019).

[5] Jenkins, D. R., Basden, A., and Myers, R. M., "A many-core cpu prototype of an mcao and ltao rtc for elt-scale instruments," *Monthly Notices of the Royal Astronomical Society* (2019).