Title
Reducing adaptive optics latency using many-core processors

Authors
Barr, David
Basden, Alastair
Dipper, Nigel
et al.

Reducing adaptive optics latency using many-core processors

David Barr[a,c], Alastair Basden[b], Nigel Dipper[b], and Noah Schwartz[a]

[a]UK Astronomy centre, Royal Observatory, Edinburgh, UK
[b]Durham University, South Road, Durham, UK
[c]Heriot-Watt University, Edinburgh, UK

ABSTRACT

The high control frequency required of planned E-ELT adaptive optics modules (from several hundreds of Hz to thousands and typically in the region of 500 Hz for first light E-ELT instruments such as HARMONI), along with the high number of actuators, leads to very demanding computational needs for the real-time AO control (RTC) systems. The number of actuators, proportional the total area of the telescope primary mirror, will range from several thousands to tens of thousands (typically 4-6 thousand for first light E-ELT instruments). Traditional RTC architectures based on CPU only technologies are typically unable to achieve the required performance in a cost-effective and maintainable manner. Alternative hardware such as many-core hardware accelerators need to be considered to deliver the computational power. These many-core processors offering a highly parallel environment, have the potential of coping with the high computational load and of accelerating parts of the AO control loop. AO systems for the E-ELT however, also put heavy constrains on acceptable levels of jitter and latency.

In this paper, we investigate novel hardware that has the potential to accelerate wavefront reconstruction and wavefront pre-processing, respectively the Intel Xeon Phi and the TILERA TILE-Gx processors. We present a detailed performance analysis putting an emphasis on execution time, jitter (i.e. variation in execution time) and outliers (i.e. results significantly apart from mean). Results are explored both for specific first light E-ELT instruments and, to stay as general as possible and fully appreciate scalability issues, for a much wider range of AO system sizes. The paper also addresses anticipated near future hardware developments and examines their suitability for the E-ELT.

Keywords: Adaptive Optics, Real-time computing, wavefront processing, wavefront reconstructon, Tilera, Xeon Phi

1. INTRODUCTION

As we move into the era of Extremely Large Telescopes (ELTs), real-time control systems (RTC) for Adaptive Optics (AO) are becoming more complex and computationally demanding. A selection of E-ELT first light instrument specifications are shown in table 1. These instruments were selected as they are first light candidates and give a representative overview of the specifications for all the ELTs using similar instruments. The Integral Field Spectrograph (ELT-IFS) will use both an SCAO and LTAO module, whereas the Multi-Object Spectrograph (ELT-MOS) will rely mainly on an MOAO system.[1]

As a illustration, figure 1 shows the processing power needed for various Very Large Telescope (VLT) and E-ELT instruments. It clearly shows the required increase needed, even over a complex instrument such as the

Table 1. Proposed specifications for E-ELT showing generic and simplified characteristics of the proposed instruments.

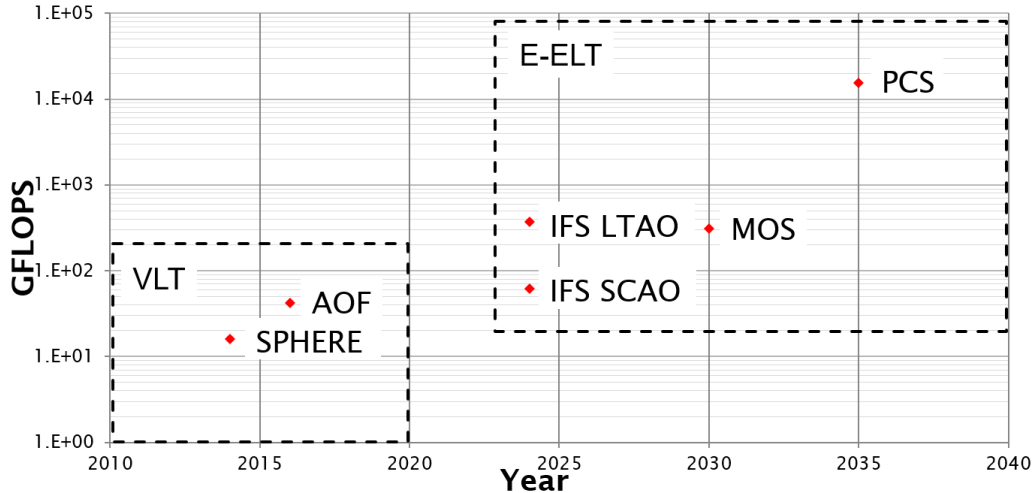| System	Telescope	Type	Channels	WFS Sub-Aps	Frequency (Hz)
IFS	E-ELT	SCAO	1	74×74	500
		LTAO	6	74×74	500
MOS	E-ELT	MOAO	10	74×74	250

Figure 1. Approximate computing power required for a selection of VLT and E-ELT instruments.

Adaptive Optics Facility (AOF), to be able to run first light instruments. The E-ELT Planetary Camera and Spectrograph (PCS) is shown to illustrate the long term computing challenges still ahead.

Traditional RTC architectures based on CPU only technologies are typically unable to achieve the required performance in a cost-effective and maintainable way. Investigating possible hardware that can deliver the desired performance, both in terms of computational power but also in terms of acceptable levels of jitter and latency, is essential. Many-core processors offering a highly parallel environment have the potential of coping with the high computational load and of accelerating parts of the AO control loop. In an effort of down selecting suitable computational technologies for the E-ELT AO modules, we investigate in this paper novel hardware that has the potential of accelerating the wavefront reconstruction and wavefront pre-processing, respectively the Intel Xeon Phi and the TILERA TILE-Gx processors.

The rest of the paper is divided into 4 sections. Section 2 describes the general RTC architecture and assumptions used in this paper for benchmarking the hardware. In section 3, we introduce and give performance estimates of the Tilera processor for the wavefront image calibration and processing part of the AO loop. In section 4, we focus on reducing the AO RTC latency of the wavefront reconstruction calculation using the Xeon Phi. Finally, we conclude the paper in section 5, where we discuss the anticipated evolution of these technologies and their place in the AO RTC control loop.

2. GENERAL RTC ARCHITECTURE

The generic function of the RTC control loop is to take the pixel data streamed from the wavefront camera and convert it to deformable mirror commands. This can be visualised in figure 2 which shows a simplified block diagram of an AO RTC control loop. This diagram only shows the main blocks of the processing chain. In this paper we will focus on the core parts of the AO loop (i.e. the time critical ones): processing the pixels from the wavefront camera and calculating the DM commands. Once read out, the pixel data has to be calibrated and processed to determine the slope vector. The slope vector is then used in the control calculation to calculate the deformable mirror commands. This is typically done through a matrix-vector multiplication, although other algorithms are being developed.

In order to minimise the time between the first pixel read out from the WFS camera and the last element of the DM command vector sent out, the processing blocks seen in figure 2 can be performed during the integration time of the WFS camera. Figure 3 shows the different processes overlapping in time to reduce the AO RTC latency as much as possible.

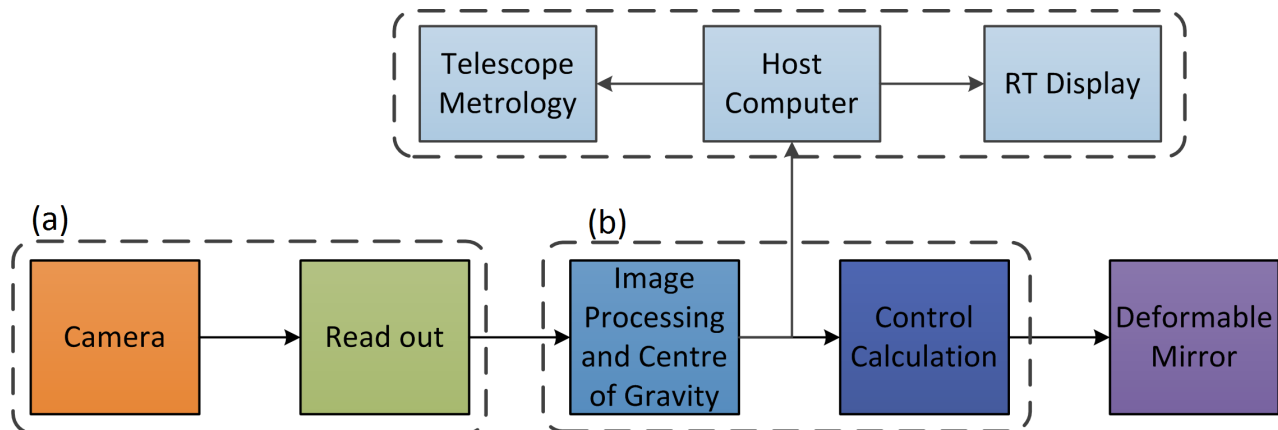Corresponding author David Barr: david.barr@stfc.ac.uk

Figure 2. Simplified block diagram of an AO RTC loop processing chain and their association in the AO RTC, (a) Wavefront Camera and (b) Real-time computer.
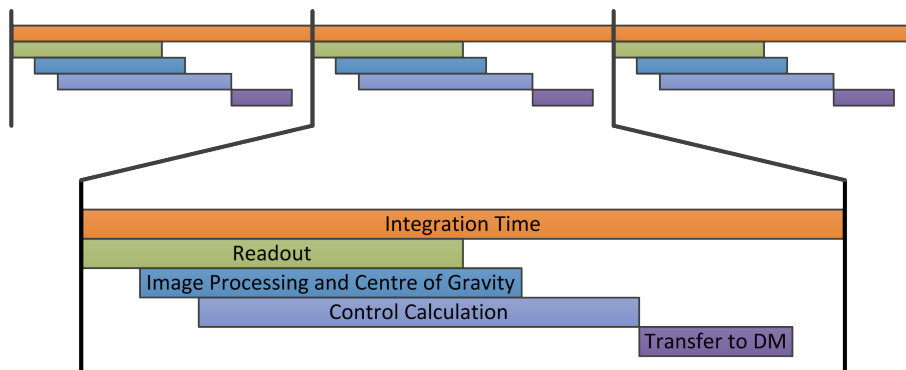

Figure 3. RTC timing diagram, showing the overlapping processing units.

One of the main factors driving the complexity of an AO system is the number of DM actuators or equivalently the number of sub-apertures of the WFS. The number of sub-apertures needed for the E-ELT NGS and LGS WFSs are anticipated to be 74×74 (see table 1) for first light instruments. However, this number does not refer to the number of valid sub-apertures. Due to the outer shape of the pupil, spiders and central obscuration, the true number of valid sub-apertures is going to be smaller. In reference to the ELT-IFS, the actual number of valid sub-apertures was taken in this paper to be in the region of 4000 (to be compared with $74 \times 74 = 5476$). For the rest of the paper, a similar scaling factor was used even for larger or smaller AO systems. When the number of sub-apertures is expressed as a single value, this always represents the total number of valid sub-apertures. However, when the number is expressed as N×N this refers to the total number of sub-apertures and not only the valid ones (to get the number of valid sub-apertures, a scaling factor needs to be applied).

3. WAVEFRONT IMAGE CALIBRATION AND PROCESSING

Figure 4 shows a typical wavefront camera processing chain for a Shack-Hartmann. Once the raw pixels have been read, they need to be calibrated with the dark map subtraction, flat fielding and background map subtraction. This step is common to both Pyramid and Shack-Hartmann WFSs. The second step differs slightly between the Pyramid and the Shack-Hartmann (SH), but are very similar in terms of computational load making general performance estimates applicable to both WFSs. For the SH, the processing involves a centre of gravity calculation and the subtraction o f the reference slopes. Here, we have used a simple centre of gravity algorithm with thresholding.

This step in the RTC chain is very important because it reduces the amount of data the RTC needs to transfer and/or process by at least an order of magnitude. The majority of work in reducing AO latency so far has focused on reducing the latency of the control calculation (typically an MVM). This has been studied
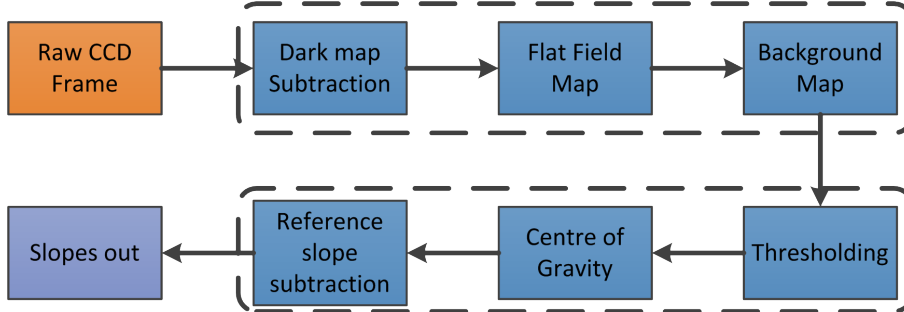
3

Figure 4. Typical processing chain for a Shack-Hartmann wavefront sensor. The calibration steps is in the upper dotted area and processing steps in the lower dotted area.

by many groups with technologies such as GPUs[2–4] and Xeon Phis.[5,6] These technologies are very efficient at accelerating the MVM due to there large number of cores and high memory bandwidths. They do not however offer very large I/O capabilities. Some systems are being developed to stream pixels directly into the hardware accelerators,[7] while others use techniques such as GPUdirect[8] or use FPGAs to allow for GPUdirect.[9] These techniques either involves a secondary copy by the CPU or developing specific hardware to allow direct memory access to the hardware accelerator. In other terms, these options either add unwanted delays to the AO loop or are expensive both in terms of development cost and time.

Off-the-self hardware have a number of advantages over ad-hoc hardware specifically developed for niche applications such as AO. They do not require any development time and generally increase maintainability and upgradability. They can however, suffer from performance issues and therefore need to be tested in suitable conditions. In this paper, we investigate an off-the-shelf solution (the Tilera TILE-GX36), a many-core processing card that offers excellent I/O capabilities by having multiple 10 Gbps ports directly on the card. This allows efficient movement of data to a parallel environment for processing. The Tilera does not offer a high memory bandwidth, so is not a suitable processor for the MVM calculation. What the Tilera can offer is a off-the-shelf platform that can receive pixel streams from a WF camera and perform the wavefront processing and calculate the wavefront slopes. This can reduce the subsequent amount of data needed to be sent to the hardware accelerators.

3.1 Tilera

The Tilera Tile GX series (shown figure 5) offers a variety of card, from 9 to 72 cores and from one to ten 10 Gbps ports. In this paper, we are testing the Tile GX36 which has 36 cores and four 10 Gbps Ethernet ports. Two modes of operation were tested: full frame and pipelined. In the first mode the processor waits for all pixels in a frame to arrive before starting to process the data. This is not necessarily representative in a real system but allows us to very accurately measure performance and investigate where the actual sources of time delay are present.

In the second mode, the processing takes place as soon as the data is available. A full frame from a camera is likely to be distributed over many data packets. Standard Ethernet packet payload allows for 1500 Bytes of data and jumbo allows for packet sizes of 9000 Bytes. A single frame from a $40{\times}40$ system with $2{\times}2$ pixels per sub-aperture would fill this jumbo frame. Systems closer to ELT scales with $74{\times}74$ would fill multiple packets per frame. As soon as a row of sub-apertures has arrived on the Tilera, the processing begins. This allows for overlap in time between the read-out and the frame processing, as can be seen in figure 3.

In normal operation, the Tilera runs a non real-time micro Linux kernel each core will be interrupted by the Linux kernel system calls. One of the more interesting modes of operation of the Tilera for real-time pixel processing is called the Zero Overhead Linux (ZOL) mode. In the ZOL mode, Tilera offers the posibility for the user to specify a subset of tiles (i.e. cores). Each of these cores will run a single, user-space task, without incurring any Linux system overheads.[11] The specified core is free of all Linux system overheads and interrupts. The ZOL mode allows for near real-time performance from the Tilera and is capable of removing major variations in execution time.
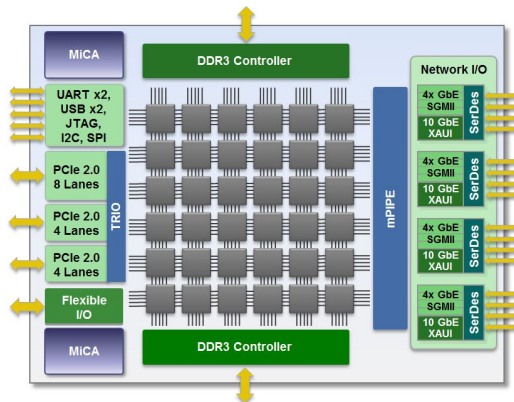
Figure 5. TILE-Gx8036 Processor Block Diagram.[10]

3.2 Full frame testing

In this configuration, the Tilera waits for a whole frame to arrive before processing the data. This mode allows quick testing of many different size detectors combined with accurate performance measurements. In order to disentangle between the different sources of time delay and actual limitations from a real camera, we have decided to emulate the WFS camera using a PC which streams pixels over Ethernet.

Figure 6 shows the mean calculation time (i.e. from pixel data to slope vector) as a function of the linear detector size N (the full detector size being N×N). Highlighted are two sizes of detectors of interest (i.e. 500×500 and 800×800). These different points represent different size detectors that are being considered for ELT first light instruments. A larger detector with 1200×1200 is also highlighted, showing the clear scalability to the timing estimates. The execution time seen is independent of the number of sub-apertures and is only dependent on the size of detector used. For example, one can estimate the approximate mean execution time for a given size detector using the data from figure 6.
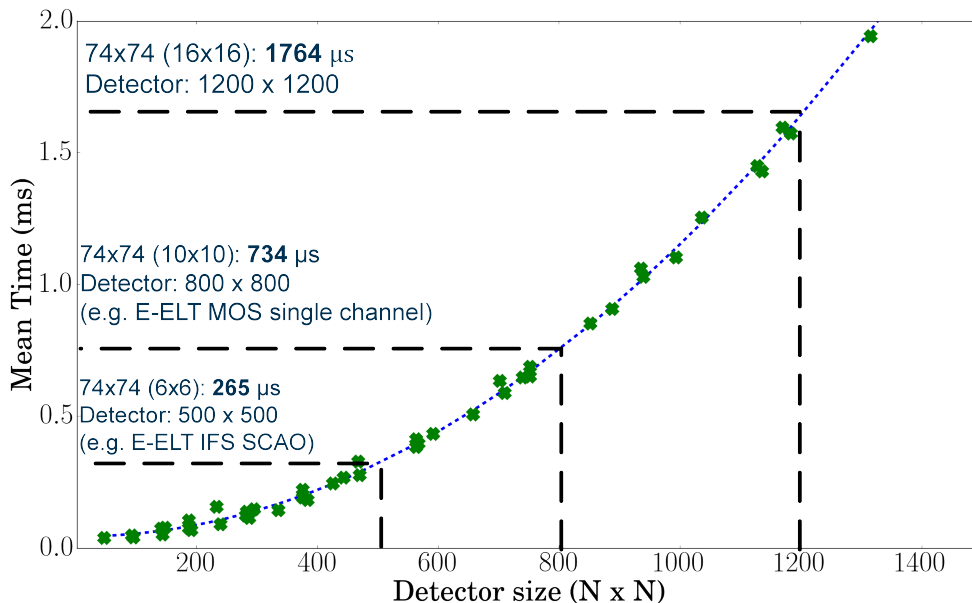

Figure 6. Mean WFS processing time as a function of linear detector size N. The green dots represent different combinations of number of sub-apertures and number of pixels per sub-aperture. The blue dotted curve is a N^2 fit.

5

As stated previously, the mean performance is not the only important factor for an RTC. The variation (or jitter) in this time is also important. We show the overall distribution of execution times for a single detector size in figure 7. We see that for over a million (10^6) frames tested, the system is very stable, with a distribution that is very close to a Gaussian. Data was obtained for 74×74 sub-apertures and for a detector size of ≈ 500×500. For this size we see a range of 13 μs and a standard deviation of $\sigma = 1.28$ s. It is important to note that approximately the same distribution shape, range and standard deviation is obtained for all detector sizes, only the mean position is modified.

The Tilera is capable of providing the performance required to process the current proposed detector sizes needed for the ELT first light instruments. We have made the assumption that the readout time of future wavefront cameras will be similar to current versions. This, combined with the low variation, suggests that the Tilera might be a strong contender for WFS pixel processing, it has similar performance to FPGAs and a lower development time and cost.
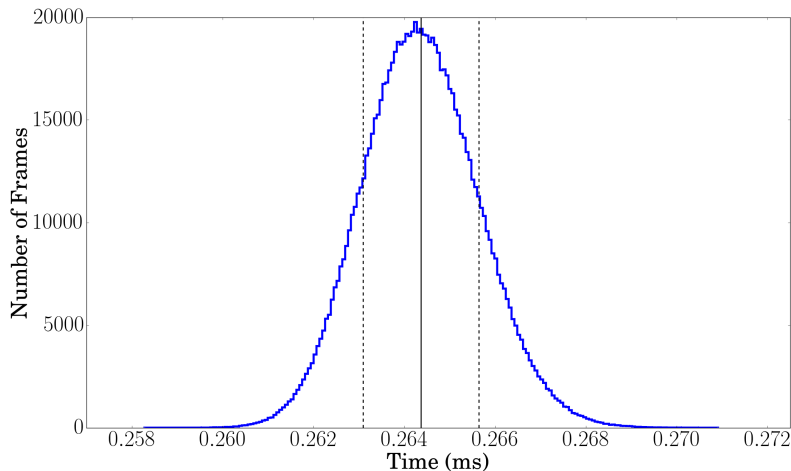

Figure 7. Variation in execution time for a 74×74 sub-apertures using a 500×500 detector. Distribution calculated for a million events. The black vertical line represents the mean and the dashed vertical lines the standard deviation.

3.3 Pipelining

The full-frame measurements, where the system needs to wait for all pixels to be received before any processing is started is far from optimal. A more efficient configuration, as depicted in figure 3, can be obtain by starting processing as soon as a sufficient number of pixels has been received (generally a full row of sub-apertures). Preliminary testing of the Tilera whereby processing starts as soon as the first row of sub-apertures has arrived show very encouraging results.

For a 74×74 sub-aperture system with a detector of approximately 500×500 pixels, the Tilera is able to reduce the time between the last pixel arriving from the camera to the last slope sent out to ≈ 50 μs. More testing is required to fully characterise performance, especially for large system sizes, but these initial results show that the Tilera can keep up with the output of a high-frame rate camera and only add minimal latency to the overall system.

4. WAVEFRONT RECONSTRUCTION

Wavefront reconstruction, translating measured wavefronts (slopes) into new DM commands, is by far the most computationally intensive algorithm that an ELT-scale RTC is required to perform. The most common wavefront reconstruction algorithm used is the Matrix-Vector Multiplication (MVM). The DM commands d are related to the slopes s through a linear equation $d = G^{-1}s$, where G^{-1} is the control matrix. The computational complexity for an MVM grows as $\mathcal{O}(M^2)$, where M is the number of degrees of freedom of the AO system.

Table 2. Comparison of advertised and achievable memory bandwidths for a CPU, GPU[15] and Xeon Phi of the same generation.

Test	Dual Xeon E5-2650	NVidia K40	Xeon Phi
Advertised Max. GBs^{-1}	2×51.20	288	320
STREAM GBs^{-1}	63.7	229	166.5
Percentage	62.2%	79.5%	52.03

In this next section, we investigate the performance of the Intel Xeon Phi for wavefront reconstruction using the MVM algorithm. The Xeon Phi uses x86 instruction set microprocessors (same as conventional CPUs), which may help in lowering the barriers to entry compared with GPUs or FPGAs, i.e. no specialised code base or API is required (unlike GPUs which require knowledge of CUDA or OpenCL). The implemented code can easily be modified and upgraded should a more performant hardware be released. The Xeon Phi also offers high memory bandwidth to accelerate memory-bound parallel algorithms. It is however, designed for high-performance computing where the requirements are more focused on the mean execution time rather than on the determinism of execution time. A detailed analysis of performance in a realistic AO environment is therefore essential. Previous investigations were limited and focused on non-real-time (Non-RT) Linux systems[12] or on a very specific AO system[6] making the generalisation to other systems difficult. In addition, a detailed analysis of the timings is crucial to fully understand the limitations of the hardware and extrapolate to future hardware developments. Different science cases will have different tolerances on the acceptable jitter (variation in execution time) or outliers (results significantly apart from the mean) for example, which may or may not impact science results significantly.

In this paper, we test the Xeon Phi 5110p with 60 cores and an advertised memory bandwidth of 320 GB s^{-1}. A more detailed analysis of the Xeon Phi performance can be found in.[5]

4.1 Achievable memory bandwidth

When calculating the MVM, the slope vector is updated at every iteration, while the control matrix remains constant for periods of time between tens of seconds to several hours. However, for large AO systems the matrix is too large to be stored in cache, and must be read from memory at each iteration. Memory bandwidth becomes a performance limiting factor. We first tested the Xeon Phi using the STREAM[13,14] test, a benchmark that give an accurate representation of the achievable memory bandwidth in multiple scenarios. Results are presented in table 4.1 and are compared to a GPU and CPU of the same generation.

A typical CPU has a relatively low memory bandwidth. We have tested a dual CPU configuration of the Xeon E5 which has an advertised combined memory bandwidth of 102.4 GBs^{-1}, and can only achieve approx. 62%. The GPU performs well, achieving $\approx 80\%$ of the advertised memory bandwidth, whereas the Xeon Phi is able to achieve $\approx 50\%$ of its advertised memory bandwidth. This means that even though the Xeon Phi has the highest advertised memory bandwidth, the achievable value is lower than that of the equivalent GPU. We believe that this is an important number to kept in mind when comparing or designing RTC systems, even if this test doesn't take into account the difficulty of achieving this value in a practical implementation.

4.2 Matrix-Vector Multiplication

Figure 8 shows a comparison of the mean MVM computation time between the Xeon Phi, a system using 2 Xeon Phis and a same generation Xeon E5 processor as a function of the number of valid sub-apertures. The measured times represent the time from the first data sent out from the host computer to the Xeon Phi(s) to the last data received back on the host computer. This includes data transfer (i.e. slope and DM command vectors) and MVM calculation on the Xeon Phi(s).

We see that for small systems with ≤ 2000 actuators (that is approx. 40×40 total sub-apertures) the CPU is faster at performing the calculation. This is due to the time required to transfer the required data between the host and the Xeon Phi over the PCIe bus. The transfer time between the host and the Xeon Phis makes the CPU a better option for small systems. When using multiple Xeon Phis, the same observation can be made. This conclusion still holds even when only half the data needs to be transferred to each respective Xeon Phi, because it is limited by the PCIe overheads. This effect is seen on GPUs as well, as they also require the transfer of data
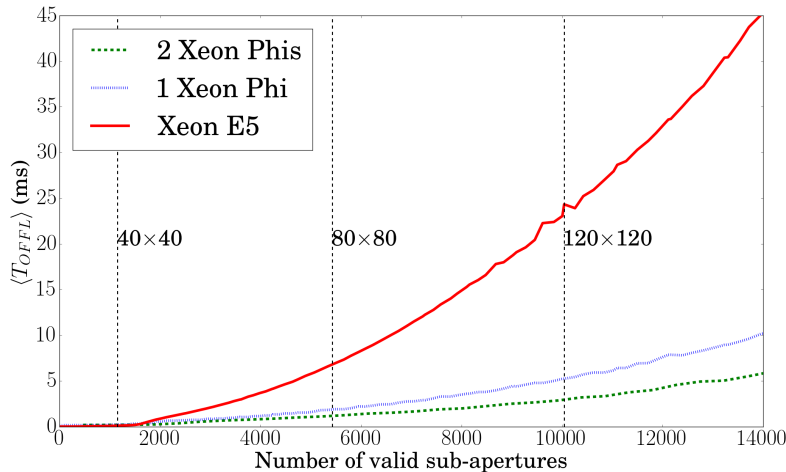
Figure 8. Comparison of the mean MVM computation time between the Xeon Phi (blue), a system using 2 Xeon Phis (green) and a same generation Xeon E5 processor (red) as a function of the number of valid sub-apertures performing the MVM control calculation. Dashed vertical lines represent approximate position of systems with 40×40, 80×80 and 120×120 total sub-apertures.

over PCIe. As the system size grows, the complexity of the MVM increases quickly (increasing as $\mathcal{O}(n^2)$), and the MVM computation itself becomes the dominating term. The Xeon Phi then out performs the CPU because of the large number of cores and the high memory bandwidth allowing for the acceleration of the calculation. For example, a 74×74 system would take in average 3.7 ms on a E5 CPU, 1.14 ms on a Xeon Phi and 0.85 ms on 2 Xeon Phis which could be acceptable for some first light ELT instruments.

4.3 Variation in excecution time

Figure 9 shows histograms of execution time (i.e. calculation and transfer of data) for three different configurations: 1 Xeon Phi used with a non-real-time Linux host, and 1 and 2 Xeon Phi used with a real-time Linux host. For each configuration, 10^6 measurements were taken. Without a real-time Linux kernel, large amounts of outliers are seen, with values up to $\approx 12\ ms$. This is far too large a variation to be acceptable for use in the RTC. A host running a RT Linux kernel greatly improves the stability of the system. For example, the standard deviation σ of the execution time is approximately 66 μs for the Xeon Phi and non-RT host, 39 μs for the Xeon Phi and RT host and 42 μs for the 2 Xeon Phis and RT host.

Similar tests were done by only measuring the calculation time on the Xeon Phi(s) therefore ignoring data transfer (data not shown). The extreme outliers visible in figure 9 are no longer present. This result indicates that the variations are caused by the host system and/or the data transfer. Even when the host system was reconfigured to support real-time Linux and priorities optimised to improve execution stability, multiple frames of delay were still present in the total offload time. This indicates that the major variations are due to the transfer of data and how the PCIe bus is handled by the Xeon Phi and not the host or the Xeon Phi itself.

5. EXPECTED TECHNOLOGY EVOLUTION AND THEIR PERFORMANCE

5.1 Wavefront processing unit: Tilera

The initial tests presented in section 3 show that the current generation of Tilera processors could offer the overall performance and real-time stability required for first light ELT instruments. The Tilera processors are a strong contender and more tests are required to estimate their performance in a real system. Future systems will have the ability to further improve over the current performance. For example, more memory bandwidth would naturally improve execution time of the image calibration step (which requires data to be read from memory regularly). Increasing the number of cores and clock speed will also improve the overall performance by allowing to have more margin in calculation times or allowing more complex centre of gravity algorithms to be implemented without penalty to performance.
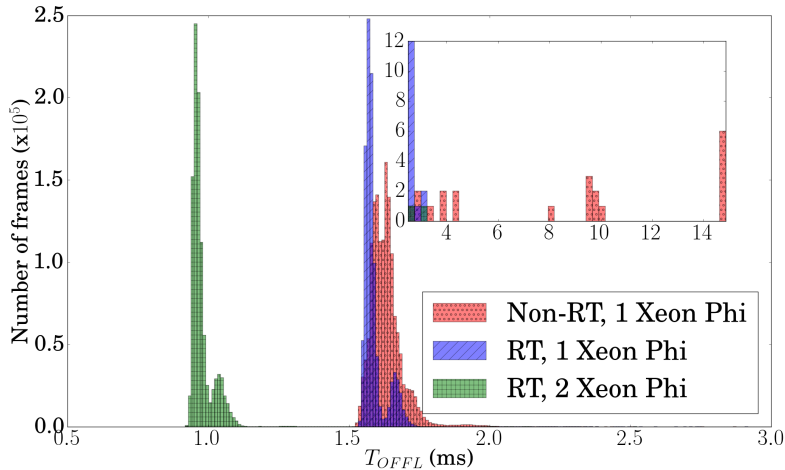
8

Figure 9. Histograms comparing the offload time for an 80×80 sub-aperture system calculated using 10^6 samples. The offload time encapsulates both MVM calculation and transfer time. Blue: Single Xeon Phi on a real-time Linux host; Red: single Xeon Phi on non-real-time Linux; Green: Two Xeon Phi on real-time Linux. Inset shows data from 2.5 ms to 15 ms with the number of frames ranging from 0-12 (showing outliers more clearly).

Table 3. A selection of mean execution time for the wavefront reconstruction measured using the current Xeon Phi and expected for the next generation (Knights Landing).

Total sub-apertures	Current Generation	Knights Landing
40×40	$350 \; \mu s$	$80 \; \mu s$
74×74	$1140 \; \mu s$	$500 \; \mu s$
80×80	$1865 \; \mu s$	$949 \; \mu s$

The next generation of Tilera, the Tile-MX series has been announced to be released in 2016.[16] This new series of chip will no longer use their in-house tiles (i.e. cores) but will rely on ARM processors. This version will offer higher memory bandwidth moving from DDR3 to DDR4 RAM. It will support programming in both C/C++ and Java and continue its impressive Zero Overhead Linux mode allowing real-time performance and no kernel interruptions of specified cores. The new generation is also advertised to have up to 100 cores. Even though these boards are not designed for AO RT applications, but rather for use in data centres, it is hoped that these changes should allow the Tile MX series to perform even better than the current GX series.

5.2 Wavefront reconstruction: Xeon Phi

The current Xeon Phi generation offers the average performance required for first light instruments but the variation in computation time is too large. We believe that it is due data transfer between the host and the Xeon Phi (through PCie) as this variation is not seen on processors themselves (i.e. where the actual MVM calculation is performed). The next generation of the Xeon Phi has been announced for early 2016.[17] In this next generation, code named Knights Landing, there are two versions of the processor available: the first as a co-processor similar to the version we have tested in this paper and the second as a standalone CPU. The standalone CPU version of the Xeon Phi can potentially solve the problem of data transfer as we will no longer need to offload the calculations over PCIe.

Figure 10 shows the predicted performance of the next generation of Xeon Phi. We have made the assumption that we can achieve half of the advertised memory bandwidth, as it was seen in section 4.1 and that there is no data transfer time. To highlight the potential performance increase, the 40×40, 74×74 and 80×80 cases have been extracted and can be see in table 5.2. We see that removing the data transfer and increasing the memory bandwidth should increase the overall performance of the Xeon Phi in terms of mean execution time. Interestingly, the Knights Landing version not only outperforms the current version for large AO systems, but also for small systems due to the lack of need of data transfers.
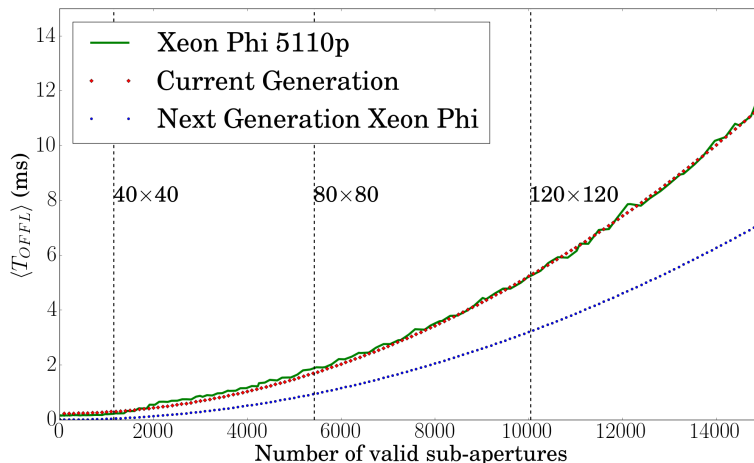
Figure 10. Comparison of mean MVM execution time between the current Xeon Phi (5110P) and the predicted next generation. These projections are based on the specifications found in.[17] Green: Current version; Red: current version based on predictive model; Blue: next generation estimation based on same model.

The main issue with the current Xeon Phi generation is the variability in execution time. This will be further reduced with the next generation standalone CPU (no PCIe data transfer needed). Running a real-time Linux kernel on the Xeon Phi should also reduce this variation further. These results are obviously only projections and will need to be tested when the Knights Landing processors are released early 2016.

6. CONCLUSIONS

In an effort of down selecting suitable computational technologies for the E-ELT AO modules, we have investigated in this paper novel hardware that has the potential of accelerating the wavefront reconstruction (by means of an MVM) and wavefront pre-processing; respectively the Intel Xeon Phi and the TILERA TILE-Gx36 processors. These technologies are off-the-shelf computing hardware that have the potential to reduce latency of the AO control loop for large systems.

We have showed that the current Tilera Tile-GX36 version has the capacity of offering the performance needed for E-ELT scale systems. It enables a large reduction in mean execution time for typical wavefront processing algorithms and excellent stability. It is hoped that future hardware generations will further improve this performance and will ameliorate upgradability and maintainability.

In this paper, we have shown that the Xeon Phi has acceptable performance in terms of mean execution time. However, variations are too large for typical real-time AO systems. We have showed that the major outliers and system instabilities come from the transfer of data over PCIe. The next version of the Xeon Phi, the Knights Landing, will be released in the next year as a standalone CPU (i.e. not a co-processor). It is expected the new version will eliminate most of the calculation jitter and make it a suitable candidate for future ELT AO systems.

ACKNOWLEDGMENTS

REFERENCES

[1] Benoit Neichel, Thierry Fusco, Jean-Francois Sauvage, Carlos Correia, Kjetil Dohlen, Niranjan Thatte, Fraser Clarke, Morgan Gray, Noah Schwartz, Hermine Schnetler, Ian Bryson, Matthias Tecza, and Joel Vernet. The ao modes for harmoni: from classical to laser-assisted tomographic ao systems. volume 1, 2015.

[2] A. G. Basden and R. M. Myers. The durham adaptive optics real-time controller: capability and extremely large telescope suitability. Monthly Notices of the Royal Astronomical Society, 424(2):1483–1494, 2012.

[3] Damien Gratadour, Arnaud Sevin, Julien Brulé, Eric Gendron, and Gérard Rousset. Gpus for adaptive optics: simulations and real-time control. volume 8447, page 84475R, 2012.

[4] Lianqi Wang and Brent Ellerbroek. Computer simulations and real-time control of elt ao systems using graphical processing units. volume 8447, page 844723, 2012.

[5] David Barr, Alastair Basden, Nigel Dipper, and Noah Schwartz. Reducing adaptive optics latency using xeon phi many-core processors. Monthly Notices of the Royal Astronomical Society, 453(3):3222–3233, 2015.

[6] Malcolm Smith, Dan Kerley, Glen Herriot, and Jean-Pierre Véran. Benchmarking hardware architecture candidates for the nfiraos real-time controller. volume 9148, page 91484K, 2014.

[7] Urban Bitenc, Alastair Basden, Nazim Bharmal, Nigel Dipper, Richard Myers, Andrew Reeves, Fabrice Vidal, Damien Gratadour, Gerard Rousset, Tim Morris, and Eric Gendron. Laboratory and on-sky test of cured and hwr fast wavefront algorithms. volume 1, 2015.

[8] Tuan N. Truong, Antonin H. Bouchez, Rick S. Burruss, Richard G. Dekany, Stephen R. Guiwits, Jennifer E. Roberts, Jean C. Shelton, and Mitchell Troy. Design and implementation of the palm-3000 real-time control system. volume 8447, pages 84472F–84472F-9, 2012.

[9] Denis Perret, Maxime Laine, Damien Gratadour, Arnaud Sevin, and Bertr Le Ruyet. A low-latency link between the nodes of a heterogeneous real-time computing platform for adaptive optics. volume 1, 2015.

[10] EZChip. The tile-gx36 processor product brief. Brochure, 2015.

[11] TILERA. Multicore development environment optimization guide. User Guide, 2011.

[12] David Barr, Alastair Basden, Nigel Dipper, Noah Schwartz, Andy Vick, and Hermine Schnetler. Evaluation of the xeon phi processor as a technology for the acceleration of real-time control in high-order adaptive optics systems. In SPIE Astronomical Telescopes+ Instrumentation, pages 91484B–91484B. International Society for Optics and Photonics, 2014.

[13] John D. McCalpin. Memory bandwidth and machine balance in current high performance computers. IEEE Computer Society Technical Committee on Computer Architecture (TCCA) Newsletter, pages 19–25, December 1995.

[14] John D. McCalpin. Stream: Sustainable memory bandwidth in high performance computers. Technical report, University of Virginia, 1991-2007. http://www.cs.virginia.edu/stream/.

[15] I. Z. Reguly, E. László, G. R. Mudalige, and M. B. Giles. Vectorizing unstructured mesh computations for many-core architectures. In Proceedings of Programming Models and Applications on Multicores and Manycores, PMAM'14, pages 39:39–39:50, New York, NY, USA, 2014. ACM.

[16] EZChip. Tile-mx multicore processor. Brochure, 2015.

[17] Intel and Eric Gardner. What public disclosures has intel made about knights landing?, 2015, (Accessed: 2015-12-07).