Report on the ESO Workshop

Real Time Control for Adaptive Optics 2012

held at ESO Garching, Germany, 4–5 December 2012

Enrico Fedrigo[1]

[1] ESO

The Real Time Control for Adaptive Optics workshop series was conceived to bring together international specialists in real-time control (RTC) for adaptive optics in order to share and exchange experience regarding the design and implementation of these systems. During two full days, the participants were presented with 28 talks divided into seven sessions, one panel discussion and two free-form open discussions. The major topics covered during this second RTC workshop are briefly reported.

The real-time control system is a crucial component for any astronomical adaptive optics (AO) system. The computational demands placed on the next generation RTCs for future extremely large telescopes (ELTs) are enormous, and even current systems require specialised skills to implement. The workshop series brings together international AO RTC specialists with the aim of sharing and exchanging experience in order to improve the design of new and proposed AO systems, increasing their performance and usability. Although the workshops are focused principally on astronomical AO, attendance of participants from non-astronomical areas, including industry, was welcomed, and indeed encouraged, to allow cross-disciplinary discussions to take place.

Figure 1. The QR code used to announce the website[2] of the RTC Workshop 2012.

Facts and figures

This was the second in a series of Real Time Control for AO workshops; the first was held in Durham in April 2011[1]. Sixty-six outside participants, complemented by a number of ESO staff members, attended the 2012 workshop: the 28 talks were divided into seven sessions and there was one panel and two open discussions (access the workshop programme via the QR code [Figure 1] or directly via the web page[2]). Most of the participants were European (see the breakdown in Figure 2), with about 10 % non-European participants (USA, Canada, east Russia), and almost 20 % of the participants were from industry, all former or current ESO partners. The workshop also invited participants to a social dinner in true Bavarian style at the Augustiner restaurant in central Munich, surrounded by the warm atmosphere of the Christkindl open-air market.

Technology

The major topics covered in the workshop were technology and algorithms, the former divided into three technological families: Central Processing Units (CPUs), Graphical Processing Units (GPUs) and Field Programmable Gate Arrays (FPGAs). Different groups took different approaches and none was really identified as the "silver bullet". FPGAs are components that can execute a program at hardware level and are therefore potentially very fast but, above all, extremely deterministic, since every phase of the process is under the control of the programmer. This is the approach taken by one group at the Thirty Meter Telescope (TMT) project [talk by Ljusic] who designed a custom board populated by a large number of FPGA chips. Something similar was made by Microgate [presented by Biasi], with custom boards made with both FPGA and DSPs (Digital Signal Processors), the latter a technology that does not seem to hold any prospects for the future. SPARTA[3] for the VLT also contains FPGAs and DSPs [Suarez], integrated in commercial boards.

While recognising the value of predictability and determinism, important for multi-year development projects, the panel discussion immediately pointed out the main problem with this technology, which is the long round-trip engineering cycle required to develop under FPGAs. This aspect was exacerbated in ESO's SPARTA project, as all FPGA development is outsourced and therefore any new requirement takes a long time to be completed, mostly due to contractual issues. With respect to this point, some talks [Dipper] identified the Open Computing Language (OpenCL) framework as a promising approach for enhancing FPGA design productivity that would, at the same time, unify the development between FPGAs, GPUs and the advanced mathematical units of the CPUs.

The second family of technology that was considered is the GPU, or rather General Purpose Graphical Processing Unit (GP-GPU), currently being looked at by many groups. When used in real time, this approach suffers from the computing model offered by the GPU cards presently on the market: no on-board input/output is featured and they must receive data from the central processing unit and its memory. Recent developments such as NVIDIA GPUDirect [talks by Gratadour and Dipper] promise to

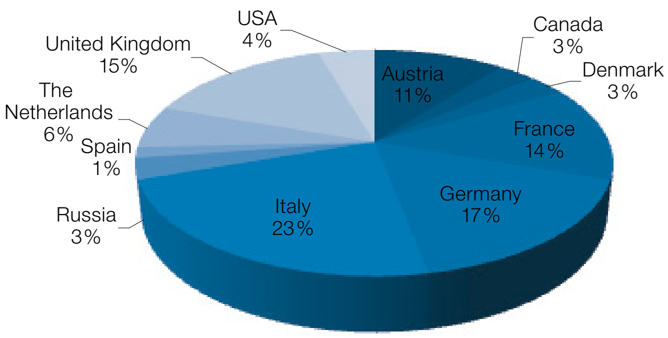

Figure 2. The breakdown of participants to the second Real Time Control for Adaptive Optics workshop.

alleviate some of the problems. Several prototypes and projects were presented by LESIA [Sevin, Gratadour], another group from TMT [Wang] and the Dutch Organization for Applied Scientific Research [TNO; Doelman], the latter projecting a possible GPU-based system for an instrument of the size of the E-ELT Exoplanet Imaging Camera and Spectrograph (EPICS; see Kasper & Beuzit, 2010).

However, even though of a different magnitude, GPUs suffer from a similar problem to the FPGAs: they need specialised tools and specialised knowledge to really harness the computing power available. GPUs are now very common; so many libraries are available to implement standard algorithms, without the need to learn a new technology. Also mentioned in several talks was the upcoming Intel Phi [Dipper, Gratadour], the first instance of the new Intel many-core architecture. It shares several aspects of a GPU since it sits on the same peripheral component interconnect (PCI) bus and acts as an accelerator without dedicated input/output. However inside it is a many-core system, so will be something potentially easier to program with standard tools and languages.

Especially relevant here was the invited talk delivered by a worldwide expert in the field of high performance computing, author of FFTW (Fastest Fourier Transform in the West) and Cilk (a computing language for multi-threaded parallel computing), Matteo Frigo. One of the issues clearly understood and mentioned also in several talks is the need to develop a strategy for managing obsolescence, since the main subject of the workshop is devoted to computer systems that will be operational for the first time in ten years from now and will run for 10 to 20 years. Frigo explained how he designed FFTW and Cilk in a fully portable way regardless of the number of available cores, their cache or interconnections, without penalising performance, hints that can be exploited in developing a CPU-based system, in particular with the Phi.

The last technology family, standard CPUs, was presented by several groups who are pursuing implementations of small or medium-sized AO systems on an all-CPU system, from solar astronomy at the Kiepenheuer Institute (Linux, ~ 500 subapertures; with plans to switch to the higher performance FPGA for the future 1600-subaperture system, [Bekerfeld]), to Force Technology (FreeBSD, with 2500 subapertures, [Kamp]), to the Durham DARC real-time controller (Linux, with ~ 200 subapertures, [Basden]), to the all-CPU version of ESO's SPARTA running on the Intel version of the real time operating system VxWorks (1300 subapertures, [Tischer]).

The clear advantage of all-CPU systems is the ease of programming and the short development time required to achieve a minimum functionality. This was clearly demonstrated by the Durham system where a number of strategies and algorithms were prototyped and tested [Basden]. What it takes to scale the system up is more debatable: how to tune it to the best performance possible and move the implementation from prototype to production. The major issue with this approach might not be the debate about whether performance can come by tweaking and hacking a freely available operating system or by purchasing an already optimised one together with its optimisation tools, but rather the people factor: any coding effort, including the “hacking” part, requires people to do it and perform long-term maintenance. It is this latter aspect that will play a major role in the decision process more than the technical solution.

The other problem with an all-CPU strategy that the audience identified is the unpredictability of any CPU-based implementation: until code is instantiated on a given platform you will never know how well it runs, making the cycle of design-on-paper-then-implement-on-silicon difficult to manage without proper prototyping. But it is clear that a system based on CPU with its inherent flexibility would be a powerful tool to test new strategies and algorithms as well as powering a laboratory system, where downtime is normally not the first issue to solve.

As a synthesis of all the presented technologies, the SPARTA team proposed a concept [Suarez] that encompasses many of the previous ideas in a flexible architecture originally presented at the SPIE 2010 meeting. The main idea is to maximise the reuse of the SPARTA architecture and software to achieve the highest cost-saving possible, while abandoning the VMEbus in favour of using server-class systems to host accelerators of various types, or even just using the available plain CPU to achieve lower hardware costs and better maintenance capability.

Algorithms

The other major topic of the workshop concerned which algorithm a real-time controller should implement. A typical measure of the complexity of an AO RTC is the product of the number of degrees of freedom at the input (the total number of gradients, or double the number of subapertures), the number of degrees of freedom at the output (the total number of mirror actuators) and the sampling rate. Without going into the details of latency requirements [as discussed by Fedrigo], this measure defines the minimum computing power required, assuming a standard least squares estimate reconstructor, which can be easily implemented as a matrix-vector multiplication (MVM). For certain systems like EPICS this measure can be prohibitively high, so this is where smart algorithms come into play.

Smart algorithms can be divided into two broad categories: ones that use a sparse, or sparsified, interaction matrix without inverting it; and others that use a model of the system, sometimes applying an interaction matrix to tune the model. In the former case the solution is found by using a conjugate gradient descent or a variation of it. In the latter case it depends on which model is used. Several talks in the algorithms session came from the Austrian Adaptive Optics group who are contributing to ESO with in-depth research on novel reconstruction algorithms. Some of their proposed algorithms have already been tested on the laboratory bench and even on sky (as reported at the workshop, [Bitenc]) and seem very promising. The portfolio of proposed algorithms spans a broad range running from a single conjugate AO (SCAO)/extreme AO (XAO) system (the CuRE family of algorithms, [Rosensteiner] and [Shatokhina])

Figure 3. One of the Real Time Control for Adaptive Optics workshop sessions in the auditorium of ESO Headquarters.

to multi-conjugate (MCAO)/laser tomography (LTAO) AO (Kaczmarz and wavelets, [Ramlau] and [Yudytskiy]) with multi-object AO in the works.

Further contributions from Lyon [Bechet] presented an optimisation of the already mature, at least in the simulation world, Frim algorithm, that has versions available for SCAO and MCAO. This algorithm is essentially a pre-conditioned conjugate gradient (PCG) method, with an arrangement that exploits the closed-loop nature of the process and thus reduces the iterations of the PCG to only one, pushing the others outside the latency cycle, a smart arrangement that is rather generic. A contribution from TU-Delft [Verhaegen] uses splines on triangular partitions to reconstruct the wavefront in a highly parallelisable architecture that, in a similar way to the CuReD (a fast wavefront reconstructor), could address the E-ELT planet-finder EPICS directly. Finally a method that only uses the images in the science detector was proposed [Koriakoski], however it is still in an early stage and suffers from several limitations. Overall, an impressive array of methods and algorithms are now available to replace traditional matrix-vector multiplication methods and the next step is to try them out.

Conclusions

Two trends are developing and converging: a push to make the real-time controller hardware faster and computationally more powerful, and a push to reduce the required computational power by means of approximated methods. The feeling of the audience was that the convergence point has already passed, such that smart algorithms can be implemented into relatively small and cheap systems, while traditional MVMs can now reach very large dimensions. It might now be a matter of choice: a relatively cheap system based on an approximated smart algorithm or a more expensive, bigger system based on the well-known and feature-rich MVM. Smart algorithms will certainly be more complex to implement than a plain MVM where extreme optimisations can be obtained; therefore smart algorithms might be best coupled with all-CPU systems or with CPU-based accelerators, while highly optimised MVM-based algorithms can be better served by GPU- or FPGA-based or accelerated systems.

Amongst the major drivers that need careful examination, the workshop identified on the one hand that the future RTC for AO will likely be a heterogeneous computing environment [Dipper, Suarez] and, on the other, that software development costs are going to be the important factor [Dipper]. Therefore the need to share developments across instruments arises. Amongst real systems, SPARTA for the VLT has already achieved that [Suarez], serving 20 AO instruments of various sizes and characteristics. Moreover it features an almost complete supervisor software that accounts for the majority of the coding effort. But other initiatives aim at achieving similar results.

A "killer" feature of the future platforms will be the capability to integrate a quasi real-time AO simulation system, similar to the end-to-end AO simulators used to predict system performance [Gratadour]. This will be desperately needed since the number of ELT-size AO benches and their availability will be extremely limited in the future and therefore the majority of the development and testing of the RTCs for the ELT AO-based instruments will be done without a bench and only with the aid of a simulation tool. Yet another challenge for the community.

Given the success of this second workshop the community agreed to convene again in about 18 months for a third one. Exact date and place will be published both on the workshop mini-site[2] and on the AO RTC collaborative web hub[4]. Meanwhile all of the presentations given at the workshop are available online[2].

Acknowledgements

The workshop was only possible thanks to the dedication of the members of the Scientific Organising Committee (Alastair Basden, Dolores Bello, Corinne Boyer, Enrico Fedrigo [chair], Glenn Herriot, Arnaud Sevin and Marcos Suarez [co-chair]), the Local Organising Committee (Samantha Milligan, Marcos Suarez, Enrico Fedrigo), the ESO IT Helpdesk, in particular the video-conferencing team, and many others, who provided logistical support. We are particularly grateful to Samantha Milligan and her passion for detail, that everything in the organisation ran smoothly, and the very positive spirit of all attendees made all tasks very pleasant. I thank all the organisers and participants.

References

Kasper, M. & Beuzit, J. L. 2010, The Messenger, 140, 24

Links

[1] The 2011 RTC for AO workshop pages: http://www.dur.ac.uk/cfai/ adaptiveoptics/rtc2011/
[2] The 2012 RTC for AO workshop pages: http://www.eso.org/sci/meetings/ 2012/RTCWorkshop.html
[3] SPARTA platform for AO RTC development for the VLT: http://www.eso.org/sci/facilities/develop/ao/tecno/sparta.html
[4] The AO RTC hub: http://aortc.ast.cam.ac.uk/