1st AO4ELT conference, 07006 (2010)
DOI:10.1051/ao4elt/201007006

An FPGA Based Computing Platform for Adaptive Optics Control

G.J. Hovey[1,a], R. Conan[3], F. Gamache[2], G. Herriot[1], Z. Ljusic[1], D. Quinn[2], M. Smith[1], J.P. Veran[1], and H. Zhang[1]

[1] Herzberg Institute of Astrophysics, National Research Council, British Columbia, Canada
[2] Lyrtech Signal Processing, Quebec City, Quebec, Canada
[3] University of Victoria, Victoria, British Columbia, Canada

Abstract. We outline a concept design for an FPGA based real-time adaptive optics controller that meets the requirements of the Thirty Meter Telescope (TMT) first light Narrow Field Infrared Adaptive Optics system (NFIRAOS). The data and processing rates for NFIRAOS are very demanding. It has pixel input rates of about 5 GB/s and processing requirements of order several hundred GMACs. The hardware proposed is programmable and flexible, and can be used in other AO systems. It is based on a custom FPGA compute blade housed with COTS interface and computer boards in an ATCA chassis. The system accepts data from eight wave front sensors at 10 Gbps each via thirty-two serial front panel data ports (sFPDPs). Nine FPGA blades provide about 2.7 TMAC/s of processing power and are used to perform pixel processing, tomography and deformable mirror fitting. Two general-purpose computer boards are used to perform background processing and control, each with three Intel L7400 dual core processors. The system is also capable of storing 90 TB of telemetry data at rate of 3.5 GB/s.

1 Introduction

Turbulent atmosphere distorts the wavefront and limits the resolution of large ground-based optical telescopes, such as the proposed Thirty Metre Telescope (TMT). Such distortions can largely be compensated by adaptive optic controllers that re-construct the distortions along the light path and adjust deformable mirrors to improve the resolution of the telescope. Re-construction is an inverse problem that by nature is compute intensive. The problem is exacerbated by a large aperture and the time variability of the atmosphere. In the case of the TMT, megabytes of data need to be processed within a few hundred microseconds with low jitter and latency. Such a computing load is of order 10^{11} operations per second and cannot be handled practically using general purpose computing approaches. Such loads, however, are commonly handled by FPGA (Field Programmable Gate Array) based digital signal processing systems. What is atypical is using them to solve inverse type problems. In this paper we present our investigations on the feasibility of such an approach. We start by reviewing the requirements of such a system and compare the trade-offs of various alternative approaches. We then outline our concept design and the results from our benchmark measurements. In the final section we provide conclusions and a technology outlook for future real-time controllers.

2 AO Computing Requirements

First generation Adaptive Optics (AO) controllers were used on meter class ground-based telescopes. These had relatively low processing requirements and were implemented using CPUs. Such CPUs were capable of providing 10 to 100 mega multiplier accumulate (MMAC) operations per second and to achieve the required performance multiple processors were used. As apertures increased more powerful AO controllers were needed to realize the resolution potential of large telescopes. Second generation AO systems made use of digital signal processors (DSPs) that were capable of GMAC performance. The latest multi-core DSPs are capable of 1 to 20 GMACs. Very large telescopes such as

[a] e-mail: Gary.Hovey@nrc-cnrc.gc.ca

the TMT require 100s of GMACs of performance which presently exceeds the capability of standard CPUs and DSPs. Field Programmable Gate Array (FPGAs) offer about a factor of 10 in MAC performance over DSPs and are widely used in high-performance digital systems. FPGAs have the following key advantages:

- They can be programmed at the gate level permitting virtually any architecture to be implemented.
- They are flexible and efficient, as both the data word length (i.e., number of bits in a data element) and the architecture are programmable.
- CPU instruction overheads such as fetch, decode, and store cycles are not required.
- Data-flow architectures can be implemented that are pipelined or parallelized to optimize algorithm performance.
- Unlike CPUs and DSPs, there are no hidden cache or pipeline stall issues.
- Custom memory interfaces can be implemented that yield more bandwidth and performance than CPUs and DSPs.
- They have highly efficient I/O interfaces that permit fast data transfers directly into the execution pipeline.
- They generally consume less power than CPUs and DSPs for a given level of performance.

There are some drawbacks to FPGAs. One is they are generally more difficult to program; however, this is mitigated to some extent by high-level programming tools. They also have limited on-chip memory and like CPU and DSPs they need external memories to store large coefficient data sets. Like many DSPs they use fixed-point arithmetic and careful dynamic-range analysis is needed to avoid unacceptable numerical errors.

2.1 Overview of TMT Real-Time Controller Key Requirements

The TMT first-light AO system, NIFAROS [1], is a multi-conjugate AO system, NFIRAOS has the following wavefront sensors: a) six laser guide-star (LGS) wavefront sensors (WFS), b) up to three low-order natural guide-star (NGS) WFS and c) a high order NGS WFS. The real-time AO controller, or RTC, operates under the control of the TMT adaptive-optic sequencer and uses the WFS sensor data to control two deformable mirrors and a tip/tilt platform. Requirements and algorithms for the RTC are given in [2] and [3], respectively. The system operates in one of several AO modes. The mode that has the most demanding requirements is the LGS mode. A block diagram of this mode is given in figure 1[1]. The major real time blocks shown in the figure are as follows:

- The pixel processing block, which receives LGS-WFS data, calibrates it, computes statistics and estimates the gradient for each sub-aperture.
- The tomography processing block, which computes atmospheric phase-screen estimates based on sub-aperture gradients.
- The DM fitting block, which projects the phase-screen estimates onto deformable-mirror (DM) actuator locations.

The system operates at an 800 Hz update rate. Every 1.25 ms, roughly 2.5 MB of data is output from six wavefront sensors in 500 μs, resulting in a peak data rate of about 5 GB/s. Gradients are estimated for 5792 WFS sub-apertures within 10 μs after the last WFS pixel is readout, which requires about 17 GMAC/s and 80 GB/s of memory bandwidth. The TMT AO requirements permit the tomography block to be implemented using one of several solvers. We have selected the Block Gauss-Seidel (BGS) with Cholesky Back Substitution (CBS) solver because it requires only one iteration per frame. This results in the fewest operations and therefore has the lowest risk of numerical round errors. The requirements for each block is summarized in table 1.

[1] Figure 1 reflects the operations studied. Version 4 of the TMT requirements supersedes this diagram.

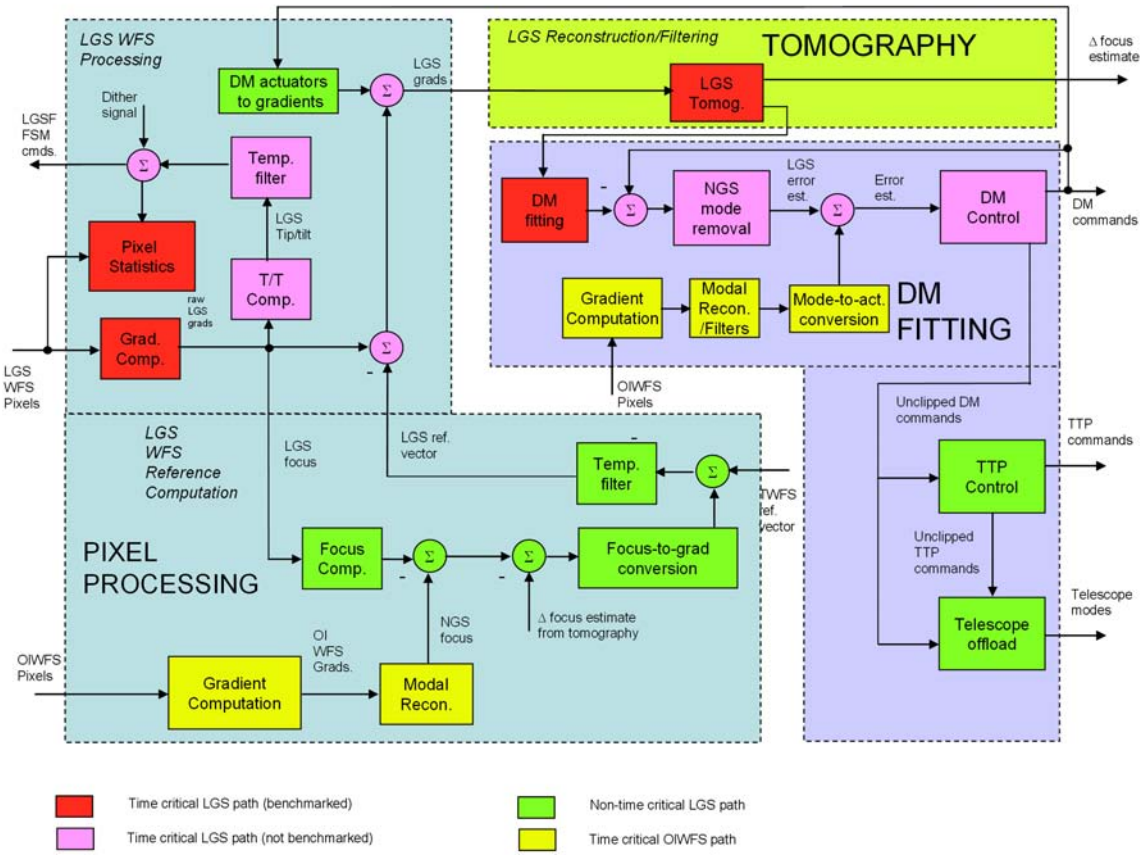

Fig. 1. A block diagram of the TMT LGS AO controller mode studied [1]

3 Implementation Overview

Because of the high performance and reliability needed, the concept system is based on Advanced Telecommunications Computing Architecture (ATCA) [4]. This is a mature well defined open architecture for building high-performance high-availability 19-inch rack mountable equipment. A block diagram of the concept system is given in Figure 2, and shows the following hardware:

- Nine identical FPGA compute-boards, each with six FPGAs.
- Two interface boards, with thirty-two 2.5 Gbps sFPDP full-duplex links.
- Two general purpose computer (GPC) boards, each with three Intel dual core L7400 processors.
- A high-speed data recorder capable of sustained write rates of 4.2 GBps.
- A display server capable of serving multiple real-time display client workstations.
- An ATCA chassis and backplane fabric, which houses the boards and provides standard Ethernet links for command and control as well as 120 2.5-Gbps high-speed point-to-point communication links between boards.

The system is highly modular and all but the FPGA boards are commercially available. A custom FPGA-board is used to perform high throughput functions, such as WFS processing and reconstruction. Nine FPGA-boards are used to implement the system and additional boards can be added to incrementally expand the system's capability. Multiple chassis can also be utilized to further expand the system. Alternatively, the FPGA board can be updated over time with new devices to improve performance. This is feasible largely because the backplane and architecture supporting the hardware exceeds the requirements "future proofing" the system by a large measure.

Table 1. TMT Real-Time Controller Requirements for Key Processing Blocks

Item	Requirement
Overall Requirements	
Frame rate	800 Hz (period 1.25 ms)
Latency from last Gradient to last DM command	< 1 ms (< 400 μs goal)
Residual Computation Error	< 20-nm RMS
Telemetry Storage Data Rate	3.5 GB/s
LGS WFS Processing	
Number of LGS WFS	6
Total WFS data per 1.25 ms frame	~2.5 GB
WFS data transfer period	500 μs
Peak WFS data rate	5 GB/s
Total sub-aperture gradients	5792 gradients (x and y)
Gradient computation time	<10 μs after WFS transfer period
Estimated processing rate per frame	~17 GMAC/s peak
Estimated memory bandwidth	80 GB/s peak
LGS Tomography Processing	
Solver	Block Gauss-Seidel with Cholesky Back Substitution
Estimated processing operations per frame	32.6 MMACs
Processing period	625 μs
Estimated processing rate	52 GMACs/s
Estimated memory bandwidth	241 GB/s peak
LGS DM Fitting Processing	
Solver	Conjugate Gradient
Number of DM actuators	Two DMs totaling 7500 actuators
Estimated processing operations per frame	46.4 MMACs
Processing period	375 μs
Estimated processing rate	123 GMACs/s
Estimated memory bandwidth	100 GB/s peak

3.1 The Real Time Processing Flow

Starting on the left side of figure 2, the high-speed WFS data are received over 2.5 Gbps fibre links by the RTC interface boards. These boards forward their data over the ATCA fabric to a pair of FPGA boards which perform the pixel processing operations on a quadrant basis. The fabric consists of multiple point-to-point links each with 2.5 Gbps throughput which can be aggregated to give the ~1 GBps peak performance needed to transfer high-speed WFS data. The gradients are then output from these boards over the fabric to the tomography engine. The engine is made up of two FPGA boards which mainly compute the Block Gauss-Seidel Solver and three FPGA boards which perform the forward and backward-substitution computation. Phase estimates from the tomography engine are output to the DM fitting engines which are implemented on a pair of FPGA boards. The DM commands are output over the ATCA fabric to the interface boards and on to the DM controller over sFPDP channels. The DM commands are output to the pixel-processing engine, where they are converted to slopes.

A key requirement is the storage of telemetry data. The driving requirement is the 3.5 GBps data rate which mainly comes from raw WFS pixel-data. The interface board is capable of re-directing data received out to the recorder without going into the system. Telemetry data can also originate from any of the other boards in the system (not shown). While the recorder is guaranteed to store data in real-time at 4.2 GBps, data can be retrieved on a best effort basis, so that real-time displays can be implemented without compromising data collection. A display server provides client display access to the recorder.

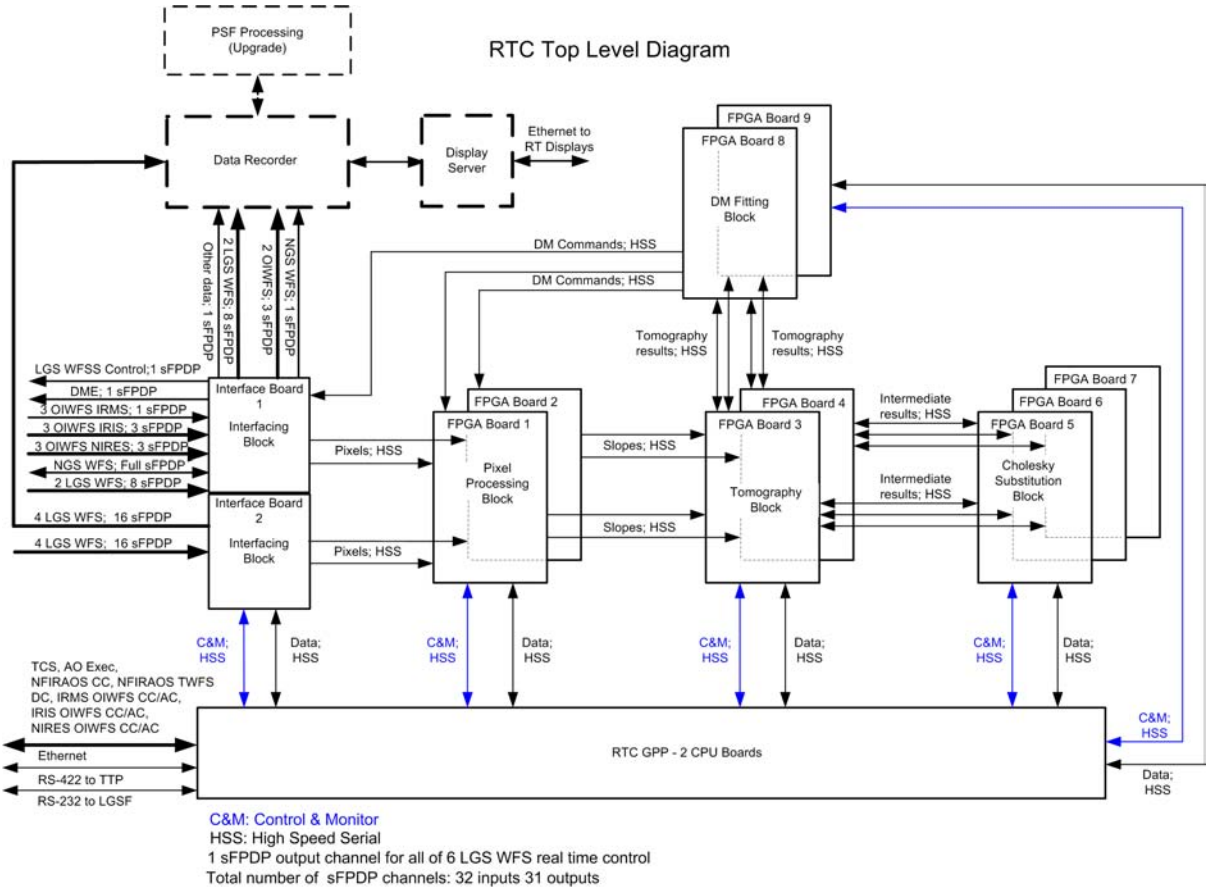

Fig. 2. A block diagram of the FPGA based real time controller

4 Results

To ensure the system met performance requirements the AO matrix operations were optimized for implementation on FPGAs. These were then implemented in fixed-point arithmetic using MATLAB. Scaled down versions of the operations were then implemented and run in Xilinx FPGAs to validate execution time and hardware requirements. The MATLAB fixed-point results were compared to a floating-point model to validate numerical-correctness and to determine the optimum bit-widths for the fixed-point operations. Hardware performance and correctness were validated by comparing the results of the MATLAB fixed-point model to the FPGA implementation. It was found that bit widths of 18, 24, 25 and 48 bits could be used as appropriate in the processing to achieve less than 11-nm residual error, easily meeting the 20-nm requirement. Pixel processing is performed in parallel with WFS readout and does not affect the critical processing-time. Gradients are available soon after the WFS data are received permitting initialization of some variables of the tomography algorithm. Within 100 ns from the end of the WFS readout, the last gradient is computed and the forward substitution of the BGS-CBS algorithm for the first atmospheric layer starts. It completes in 62 μs. The tomography and the first stage of the DM fitting algorithm then run in parallel. Together they are executed in 471 μs, which is set by the Tomography processing rate. Once the first stage of the DM fitting is complete, the conjugate-gradient algorithm computes the DM shapes in 191 μs. The overall computing time is 724 μs, which meets the 1 ms computing time requirement.

5 Conclusions and Technology Outlook

Real-time AO controllers for extremely large telescopes, such as the TMT, can be implemented using exsiting FPGA hardware. The system presented makes extensive use of pipeling and parallel execution. Higher frame-rates and throughput could be achieved by speeding up readout, reducing the number of algorithm operations and using higher FPGA clock-speeds. On-chip memory would also improve memory bandwidth performance and help achieve higher system clock-rates. Faster external and inter-chip communications would permit faster readout and update speeds.

Next generation FPGAs such as Xilinx Virtex-6 devices have I/O speeds that are a factor of 4 faster. The memory interface bandwidth is a factor of 2 higher. Clock speeds improvements however are marginal and only 10% faster. Memory available on-chip is also too little for AO ELT applications.

The roadmap for CMOS semiconductor performance is given in [5]. Looking ahead to the year 2017 we can expect the following:

- Factor of 10 improvement in flash and DRAM capacity. Memory-based disk drives may be the norm at altitude.
- Five times faster off-chip I/O speeds to 42 Gbps.
- An improvement of 5 to 8 times in cost-performance of transistors per die.
- A factor of 8 to 10 in processor engines per die.
- A factor of 1.5 in multi-drop lines – limiting conventional memory bandwidth.
- Only a factor of 3 reduction in dissipation as static power will be a significant.
- Slightly higher ratio of memory to processors.
- An 80% improvement in clock speeds.

Overall there may be a 7 to 8 times improvement in operations per second. It will result mainly from more transistors running in parallel not due to higher clock rates. Faster AO controllers will not be able to rely on higher semi-conductor clock speeds for performance, but rather better algorithms that have fewer operations and permit more parallelism.

References

1. B. Ellerbroek, ”Adaptive Optics Systems for the Thirty Meter Telescope”, Proceedings: Adaptive Optics for Extremely Large Telscopes, Paris, June, 2009 (in print)
2. ”Requirements Document for NFIRAOS RTC”, TMT.AOS.DRD.08.001.REL04, Thirty Meter Telescope Corporation, Pasadena, December, 2008
3. ”NFIRAOS RTC Algorithm Description”, TMT.AO.DRD.08.002.REL04, Thirty Meter Telescope Corporation, Pasadena, December, 2008
4. ”ATCA Advanced Telecommunications Computing Architecture”, http://www.intel.com/technology/atca
5. ”The International Technology Roadmap for Semiconductors: 2008 Update Overview”, ITRS, http://www.itrs.net/Links/2008ITRS/Update/2008_Update.pdf