

FPGA based microserver for high performance real-time computing in Adaptive Optics

C. Patauner^a, R. Biasi^a, M. Andrighettoni^a, G. Angerer^a, D. Pescoller^a, F. Porta^a,
D. Gratadour^b

^a Microgate Srl, via Stradivari, 4 – 39100 Bolzano-Bozen (BZ) – Italy

^b LESIA, Observatoire de Paris, 5 pl. Janssen – 92190 Meudon– France

ABSTRACT

The real-time data pipeline of ELT-oriented adaptive optics control systems requires very large communication bandwidth, flexible interfacing, high computational power and memory bandwidth, low latency and low jitter. A potential solution to all these challenging requirements has been developed by Microgate in the frame of the GreenFlash project and is based on two specifically developed FPGA boards with PCIe backplane interface with excellent energy efficiency. The μ XComp board is dedicated to high throughput computational tasks, with particular focus on the memory bandwidth, which is often the limiting factor for heavy matrix-vector multiplication, while still featuring flexible, high bandwidth interfaces. The second board, called μ XLink, is dedicated to the smart interfaces and features a system-on-chip ARM processor enabling the implementation of a stand-alone Microserver concept. In the Microserver, the μ XLink board interconnects and arbitrates the operation of various acceleration and interface boards, e.g. μ XComp boards or GPU accelerators, without requiring a host computer. Such Microserver concept can be adapted to satisfy the requirements of different AO configurations, being compatible with various wavefront cameras and deformable mirrors, and perform also other dedicated computation and control tasks within an AO system. We report the design, the performances and the tests performed on the hardware realized so far.

Keywords: E-ELT, adaptive optics, deformable mirror, μ XComp, μ XLink, FPGA based accelerator card, real-time computation, large memory bandwidth, energy efficiency, HPC

1. GREEN FLASH REAL-TIME DATA PIPELINE

High Performance Computing (HPC) has become a critical component as well in the field of adaptive optics for large scientific instruments such as the European Extremely Large Telescope (E-ELT), a 39m diameter telescope project. At the core of telescope operations is the adaptive optics (AO) module, used to compensate in real-time the effect of atmospheric turbulence on the wavefront to maintain the best image quality out of the telescope during the observations. For this new category of extreme scale scientific equipment, HPC is used to operate its complex AO sub-systems that require the real-time control, at the millisecond rate, of deformable optics with thousands of actuators. This major European project can be used as an example of the requirements of HPC computing in science and technology in the coming decade. In the course of a Horizon2020 funded project called **GreenFlash** [1] with academic and industrial partners the member **Microgate** is **developing, implementing and testing a Real-time HPC based on FPGAs**.

Within the GreenFlash project, a general breakdown structure is defined and shown in Figure 1 that applies to a typical adaptive optics Real-Time Control System for the E-ELT [2]. The real-time data pipeline highlighted in Figure 1 performs all calculations and control steps to pilot the actuators of a deformable mirror based on sensor data that derive from wavefront sensors (WFS).

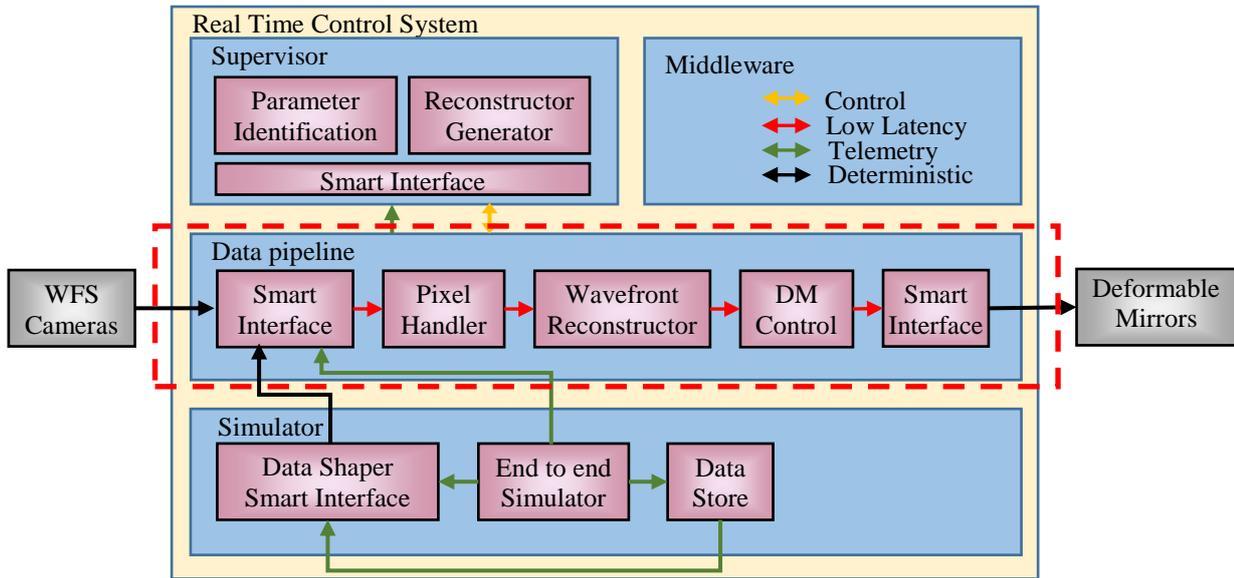


Figure 1: Green-Flash block diagram

The Microgate development of a hardware solution for implementing the real-time data pipeline will fulfill the GreenFlash overall specifications, which have been derived analyzing the various AO configurations of the E-ELT instruments. For each requirement, we are considering the most demanding case, excluding for the moment the EPICS case, which is not reasonably affordable with the today technology in terms of sustained computational throughput. Table 1 summarizes the requirements of the various instruments and highlights the ones that Microgate will tackle with its development of a Microserver that implements the real-time data pipeline.

Table 1: GreenFlash requirements

Instrument	Frame rate [Hz]	Latency [ms]	Jitter [μ s]	Pixel rate [Gb/s]	TMAC/s for MVM
HARMONI	800	2.5	125	139.3	0.2
MAORY/MICADO	500	4	200	138.2	0.46
METIS SCAO	1000	2	100	10.2	0.002
METIS LTAO	1000	2	100	194.6	0.37
MOSAIC	250	8	400	74.2	1.4
HIRES	500	4	200	138.2	0.28
EPICS	3000	.667	33	122.9	11

2. MICROSERVER

The implementation of the Microserver concept comprises a PCIe backplane, on which two different kind of FPGA-based boards are attached. The two FPGA boards are named μ XComp and μ XLink. Our primary microserver configuration consists of one μ XLink board and one or more μ XComp boards as can be seen in Figure 2. The μ XLink board is used to implement the PCIe root port and interfaces to the external world and to the user. Therefore, the μ XLink provides several different interfaces as well as a powerful microprocessor on which an operating system e.g. embedded Linux can be installed. The μ XComp boards are used to execute the heavy computational tasks of the real-time data pipeline. While the PCIe backplane allows as well to use other accelerator cards (e.g. GPUs) in addition to the μ XComp boards, the μ XLink board is essential for the microserver concept to be operated as a stand-alone system. Both FPGA boards can be used also in standard PCs or server machines as endpoint cards in order to provide computational power and/or flexible high bandwidth interfaces for other fields of application.

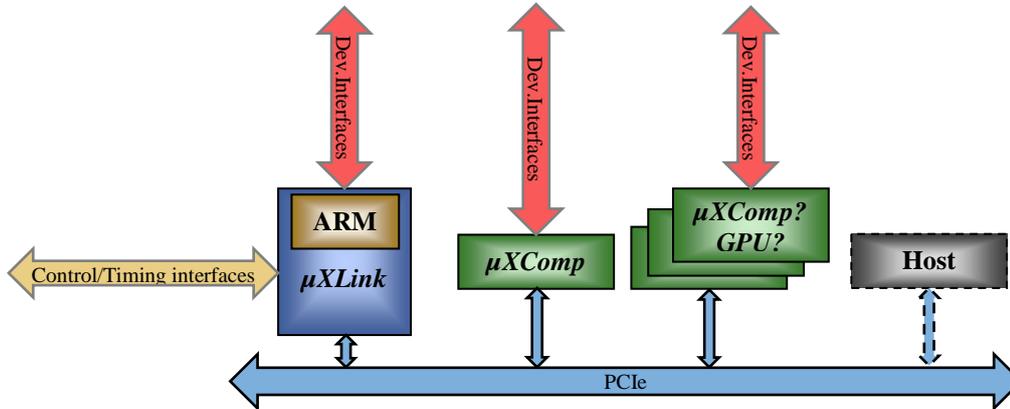


Figure 2: Microserver concept

With these FPGA boards the microserver achieves **high computational power, optimal energy efficiency, low data transfer latency, low jitter and efficient management of telemetry data**. The features of the microserver can be summarized as follows:

- The microserver shall allow stand-alone operation using SoC FPGA-CPU combination, while preserving compatibility with standard servers
- It will allow to insert different accelerator cards based on FPGAs, GPUs or CPUs
- The system will be scalable to computational throughputs up to some TFLOPs to adapt to the Real Time Reconstructor requirements of different AO instruments; in this frame, it will provide also different interfaces to wavefront cameras and deformable mirrors
- It will guarantee low latency and low jitter to fulfill the real-time requirement
- It will be energy efficient in comparison to other hardware solutions with similar performance

Microgate has selected for their boards state-of-the-art Intel FPGAs from the **ARRIA 10** family in combination with novel memory and communication devices. The details of the two boards are described hereafter.

2.1 μ XComp board

The μ XComp board is designed for **high throughput deterministic real-time computation**. It is based on the Intel FPGA **ARRIA 10 GX 1150**. In addition, a state-of-the-art memory device **Hyper Memory Cube (HMC)** with ultra-fast data transfer rate is selected to obtain a large memory bandwidth. Several real-time AO algorithms are based on matrix-vector multiplication (MVM); this processing is typically limited by the **memory bandwidth** required to feed the processing engine with the coefficients data matrix. A block diagram of the μ XComp board with its components and interfaces can be seen in Figure 3.

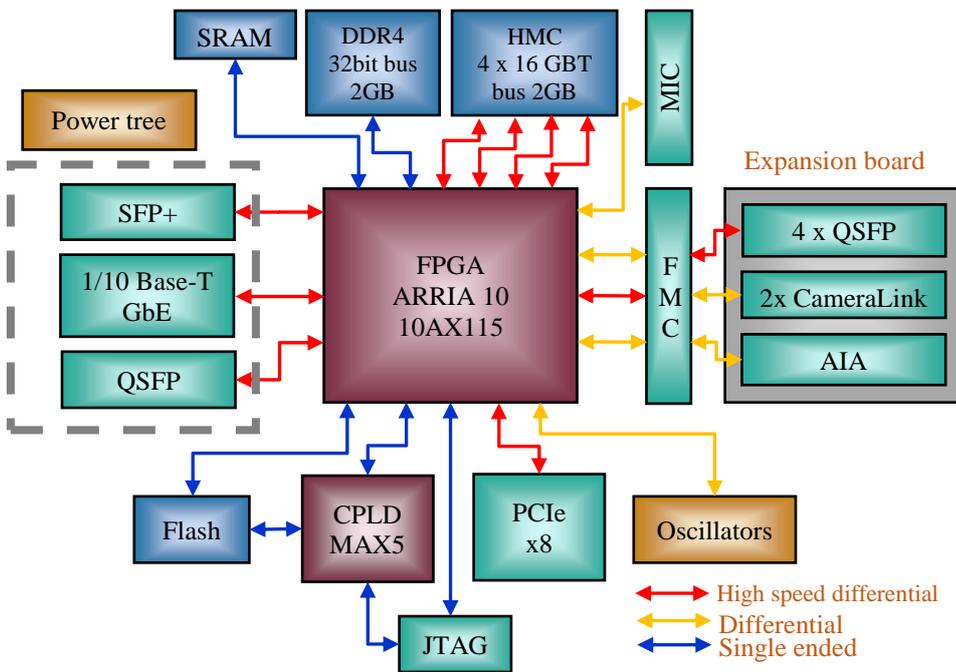


Figure 3: μ XComp board block diagram

The ARRIA 10 GX 1150 contains the highest number of high-speed transceivers from this family, which are used to interface the HMC with the FPGA. The HMC provides 4 links, each link containing 16 transceiver lanes. For the maximum memory bandwidth all 64 transceiver lanes of the HMC are connected to the FPGA. Each transceiver can be operated in parallel and full-duplex with three different transfer rates 10Gb/s, 12.5Gb/s and 15Gbps. This allows an aggregate maximum memory bandwidth of theoretically 120GB in each direction.

The FPGA contains more than 1500 DSP blocks. Each DSP block can perform a single precision floating-point multiply-accumulate operation at a frequency of about 200MHz and the DSP blocks can be instantiated to work in parallel. This gives a high computational power and together with the large memory bandwidth of the HMC this board provides a large computational performance especially for MVM operations. At the same time it has a notably small power consumption compared to other types of accelerator cards based e.g. on GPUs. The time deterministic nature of FPGA implementations with low latency and jitter makes this solution perfectly suitable for real-time computational applications.

A PCIe interface is provided containing 8 lanes and is compliant up to the generation 3 standard. This interface allows to combine the μ XComp board with the μ XLink board and other boards in the microserver and to be used as accelerator card in standard PCs or servers. The PCIe maximum theoretical communication bandwidth is 8GB/s over the PCIe backplane.

In addition to the PCIe interface, several other kinds of interfaces to communicate with the external world are implemented on the μ XComp board. A QSFP and SFP+ cage on the front-panel allow attaching optical fiber modules for 10Gb/s and 40Gb/s Ethernet or Infiniband. At the front-panel a RJ-45 connector is implemented using the newest Ethernet PHY chip from Marvell that allows to attach standard Ethernet copper cables with transfer rates from 10Mb/s up to 10Gb/s.

The back side of the board contains two additional connectors, the Microgate Interface Connector (MIC) and the FPGA Mezzanine Card (FMC) connector. These connectors allow to extend the functionality of the μ XComp using expansion boards to provide even more interfaces. The MIC connector contains 20 LVDS pairs that can be used as well as single-ended IOs and provides a lightweight solution to attach flat-cables with small interface boards or debugging and control signals. The much larger connector FMC contains 16 transceiver links, 32 LVDS pairs and some clock signals together with 3.3V and 12V power supply lines. Different off-the-shelf or custom expansion boards can be attached this FMC to provide interfaces as CameraLinks, additional QSFP or SFP+ links or AIA interfaces.

A programmable oscillator chip provides the different reference clock signals for the transceiver links and for the FPGA logic. Different communication protocols and speeds can be implemented by changing the Firmware of the FPGA and reprogramming the oscillator to provide the required reference frequencies.

All these flexible interface options make this board a highly flexible computational board with much wider field of application than only as accelerator cards in PCIe backplane server machines.

To optimize the implementation of a simple soft-core microprocessor as the NIOS II in the FPGA two additional memories are connected to the FPGA; a SRAM that can be used to store the microprocessor code and a DDR4 memory with a 32bit bus and 2GB capacity for the data memory. The DDR4 can also be used as an additional computational memory with a bandwidth of about 10GB/s or as diagnostic buffers. A CPLD chip MAXV acts as board controller and performs the FPGA programming, housekeeping and board safety.

2.2 μ XLink board

The μ XLink board is designed to feature very flexible interfaces, so to act as a Smart Interface from/to real time sensors and actuators and acts as **Microserver host and arbiter**. It is based on an **Intel System-on-Chip FPGA ARRIA 10 SX 660 with embedded ARM Cortex-A9 dual-core microprocessor** and has **PCIe root** capability. All the interfaces on the μ XComp board are present as well on the μ XLink board extended with additional types of interfaces to provide a high flexibility of the board. The different interfaces and components are summarized in the block diagram in Figure 4.

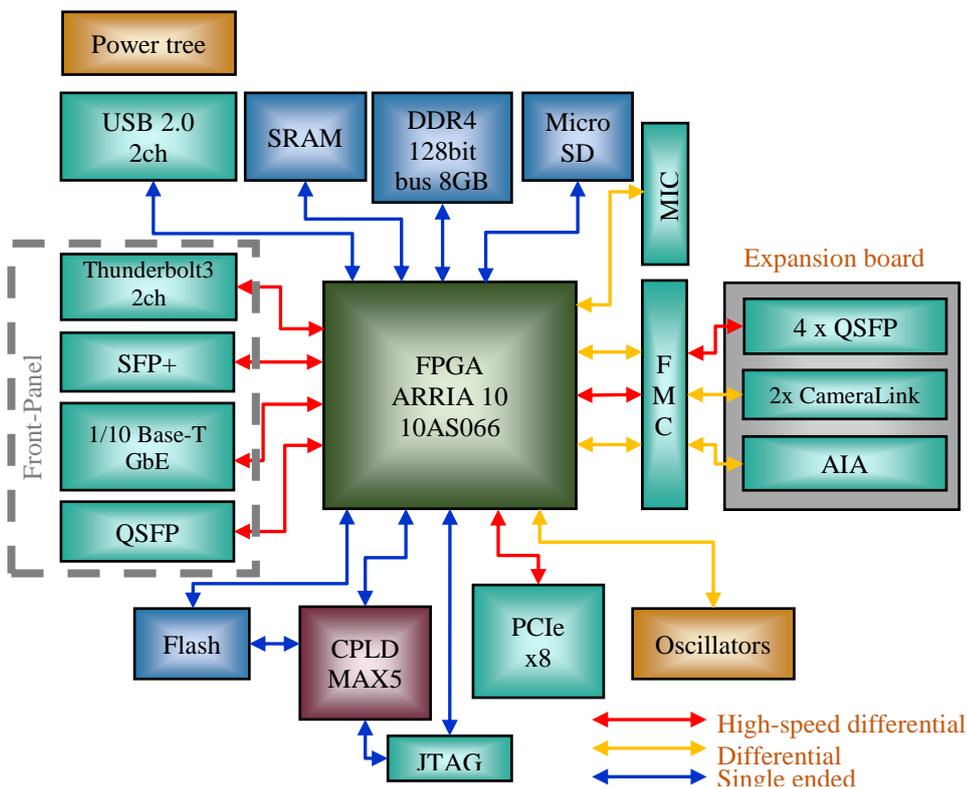


Figure 4: μ XLink block diagram

In contrary to the FPGA on the μ XComp board, this board contains a FPGA with a powerful hard-wired dual-core microprocessor on the same IC. This ARM processor allows to run an Operating System (OS) e.g. embedded Linux, with which the board does not require a Host PC or server machine but provides a stand-alone system, the microserver. In addition, the ARM microprocessor allows to control a PCIe root port by instantiating the PCIe controller in the ARRIA 10 accordingly. Electrically the PCIe interface of the μ XLink is equal to the one of the μ XComp board and therefore a PCIe

crossover adapter board is required to use it as root port. This root port allows the μ XLink to connect to several PCIe endpoint cards as the μ XComp board in order to perform a powerful stackable stand-alone HPC machine.

The ARRIA 10 SX 660 contains in total 48 transceiver and 1855 DSP blocks. The DSP blocks are equal to the ones in the μ XComp FPGA and perform a single precision floating-point multiply-accumulate operation at a frequency of about 200MHz. For the data storage several DDR4 chips are attached to the FPGA with total memory bandwidth of 40GB/s and a capacity of 8GB (upgradable to 16GB). With these components the μ XLink has as well a very **powerful computational engine in the FPGA and in the embedded processor** and can be used to implement the **telemetry data management of AO systems**.

Two Intel Thunderbolt 3 interfaces are foreseen on the front-panel, besides the above mentioned interfaces of SFP+, QSFP and RJ-45. The Thunderbolt 3 interfaces that are nowadays present in modern notebooks, allow transfer rates up to 40Gb/s per link and are compatible with the USB 3.1 standard and earlier. They can be used to attach monitors, keyboard, mouse and fast external hard drives to build together with the OS on the ARM a full operating system.

Some additional interfaces like USB 2.0, MicroSD Card slot and eventually a triple-speed Ethernet will be directly attached to the ARM processor dedicated pins to facilitate the communication, debugging and programming of the ARM.

Equally to the μ XComp board the μ XLink contains the FMC and MIC connector on the back side to attach expansion boards for extending the functionality and/or interfaces of the μ XLink. The position and pinout of the FMC is the same on both boards in order to allow using the same expansion cards.

3. CURRENT STATUS AND BOARDS AVAILABILITY

The design, prototyping and test of the μ XComp board has been fully completed. The 18 layer PCB board has a length of 200mm and a height of 111mm as can be seen in Figure 5. With these dimensions it fits in the PCIe full height and 3/4 length slots. An expansion card with a maximum length of 120mm can be attached on full length slots.

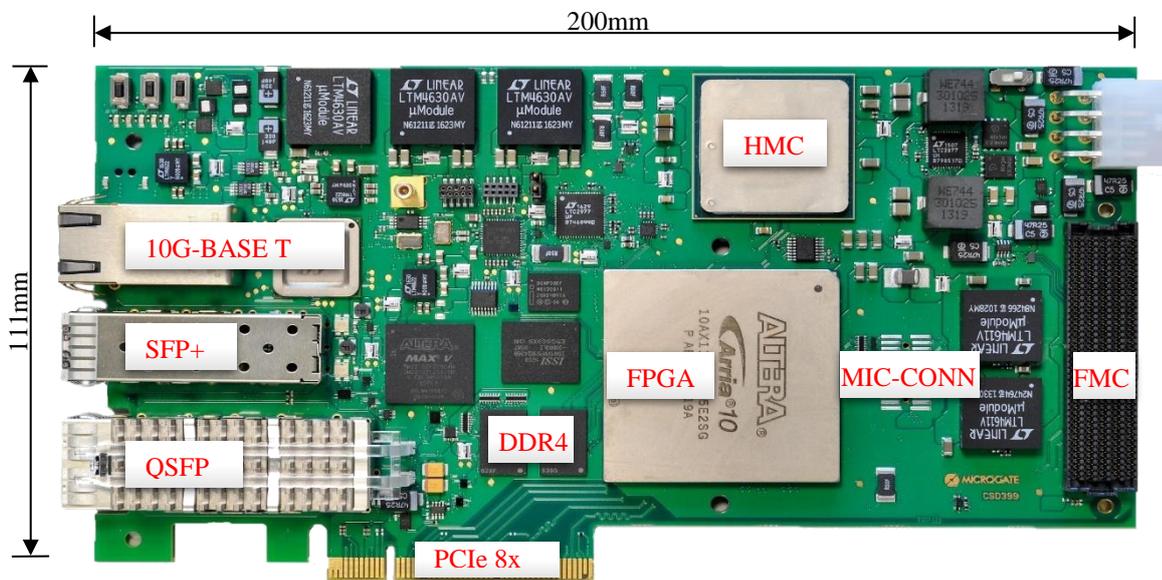


Figure 5: μ XComp board prototype

The deep hardware validation tests have proven that all internal and external interfaces are functional and perform as expected by design. The front-panel interfaces (SFP+, QSFP, RJ-45) were tested with 1G and 10G Ethernet and fulfill the standards. The PCIe is tested with Gen2 and Gen3 using 4 and 8 lanes and performs as expected achieving a maximum effective data rate of 6.2GB/s in each direction. All the transceiver links of the FMC were tested as well up to a 10G Ethernet standard. The power sequence of the different power rails is tested to guarantee a safe power-up and power-down cycle of the board.

One of the most demanding tests was the verification of all the 64 transceivers for the HMC memory. All 4 HMC links were tested with 10 Gb/s and 12.5Gb/s transition speed and showed stable functionality in read and write direction showing a total effective memory bandwidth of 71GB/s and **89GB/s** respectively. With an already planned future upgrade of the μ XComp using a higher core speed grade off the same ARRIA 10 device also the 15Gb/s transceiver speed can be supported for the HMC allowing a memory bandwidth up to 107GB/s.

We are currently implementing on this board the Saturation Management of the E-ELT M4 adaptive mirror: this is an excellent application test case, comprising the transformation from modal to zonal space of the commands for the 5316 actuators and the computation and smart clipping of their forces.

Moreover, one of the board prototype has been delivered to the GreenFlash partner PLDA to make it compatible with their QuickPlay software-defined FPGA development platform.

The power consumption of the μ XComp board depends on the implementation in the FPGA; which frequency the logic runs and how many transceivers are in use and at which transfer rates. Because the HMC uses the most of the transceiver channels the power consumption is largely dominated by the HMC usage. Table 2 shows the test results of the power consumption regarding the HMC operation.

Table 2: μ XComp power consumption and temperature regarding the HMC usage

# HMC links active	Power consumption	ARRIA 10 temperature	HMC temperature
One link only @ 10Gbps	27.8W	42°C	44°C
All 4 links @ 10Gbps	52.8W	60°C	54°C
One link only @ 12.5Gbps	31.9W	42°C	45°C
All 4 links @ 12.5Gbps	62.0W	62°C	54°C

A total power consumption of 80 W for the μ XComp board using all transceivers can be expected by extrapolating the data in Table 2. The power supply of the board is designed and tested up to 100W total power.

The μ XLink board is currently in final design. Prototyping and first tests are planned for the end of 2017.

4. CONCLUSIONS

Microgate is developing a complete solution for the implementation of the **real-time pipeline using state-of-the-art FPGA boards, fully developed in house**. In this way we will be able to support current AO projects along their whole lifetime. This is a crucial aspect, that Microgate has already demonstrated in the past with the previous generation telescopes (e.g. Keck, LBT, Magellan), where the AO pipeline is based on mid 2000s Microgate technology but still fully supported and expandable.

The first of the two types of FPGA boards is already fully tested and has proven to perform as expected by design. The μ XComp board is an optimal computation board for real-time applications, providing an energy efficient time deterministic computational power up to several FLOPS and 100GB/s on-board memory bandwidth.

The μ XLink is under development and initial integration of the Microserver is planned for late 2017/early 2018.

The Microserver concept is not only limited to the use of Microgate PCIe board, but will be further developed to interface seamlessly other acceleration boards, so to allow flexible tailoring to the target application. The Microgate boards can be used independently as well outside the microserver in standard PCs or servers for other fields of application.

The boards will be introduced to the market in 2018 and be available as COTS components.

REFERENCES

- [1] “Green FLASH: energy efficient real-time control for AO; Gratadour, D. et al; Proc. SPIE 9909, Adaptive Optics Systems V, 99094I (27 July 2016); doi: 10.1117/12.2232642
- [2] *Green Flash: Exploiting future and emerging computing technologies for AO RTC at ELT scale*; Gratadour, D. et al; in this conference