ISBN: 978-3-95450-193-9 doi:10.18429/JACoW-ICALEPCS2017-FRAPL05

# HARDWARE ARCHITECTURE OF THE ELI BEAMLINES CONTROL AND DAQ SYSTEM

P. Bastl†, P. Pivonka, B. Plötzeneder, O. Janda, J. Trdlicka, V. Gaman, J. Sys
ELI Beamlines/Institute of Physics of the ASCR, Prague, Czech Republic

*Abstract*

The ELI Beamlines facility is a Petawatt laser facility in the final construction and commissioning phase in Prague, Czech Republic. End 2017, a first experiment will be performed. In the end, four lasers will be used to control beamlines in six experimental halls. The central control system connects and controls more than 40 complex subsystems (lasers, beam transport, beamlines, experiments, facility systems, safety systems), with high demands on network, synchronisation, data acquisition, and data processing. It relies on a network based on more than 15.000 fibres, which is used for standard technology control (PowerLink over fibre and standard Ethernet), timing (WhiteRabbit) and dedicated high-throughput data acquisition. Technology control is implemented on standard industrial platforms (B&R) in combination with uTCA for more demanding applications. The data acquisition system is interconnected via Infiniband, with an option to integrate OmniPath. Most control hardware installations are completed, and many subsystems are already successfully in operation. An overview and status will be given.

## INTRODUCTION

ELI Beamlines [1] is an emerging high-energy, high-repetition rate laser facility located in Prague, Czech Republic. Four laser beamlines (ranging from the inhouse developed L1 with <20fs pulses exceeding 100mJ at 1kHz based on DPSS technology to the 10PW-L4, developed by National Energetics) will supply six experimental halls which provide various secondary sources to users. Facility commissioning, and installation work of lasers and experiments is progressing, and first user experiments are expected in 2018.

The central control system connects, supervises and controls all technical installations used for the operation of this facility, which are more than 40 complex subsystems (lasers, beam transport, beamlines, experiments, plant systems (HVAC, vacuum), safety systems) with high demands on network, synchronisation, data acquisition, processing, and storage.

This paper describes the hardware architecture of this control and data acquisitions system, and addresses the challenges to be faced in the upcoming years.

## APPROACH

There are three factors that make the development of the ELIs' control system challenging:

First, ELI has been designed to be multifunctional, and to provide a highly diverse selection of lasers and secondary sources to researchers. In practise this means that we have to integrate a **multitude of very diverse subsystems** developed by internal and external suppliers; leading to an initially very inhomogeneous technical landscape, and complex system interfaces.

At the same time, ELI is **building groundbreaking technology** and its demands on synchronisation and data acquisition are pushing the boundaries on what is possible with current technology. Demands are especially high on safe operation, synchronization, and data acquisition.

Third, in ELI **commissioning and operational phases overlap**. While one part of the facility is still under development, others are being installed, and again others will be already serving early users (whose experiments need to be supported technically, and for whom laser and beam transport operation, safety, timing and data acquisition services must be provided).

We are using three approaches in our hardware architecture to deal with these challenges:

- **Standardization of hardware interfaces,** for example for camera interfaces [2], but also more complex interfaces like our lasers. This reduces complexity and software development effort, and allows us to integrate new systems with less or at least known effort.
- Use of **common hardware based on open standards**, which allows common just-in-time procurement (taking advantage of high-volume pricing), gives us full control and documentation, and flexibility for future updates and maintenance.
- **Implementation of a test-bed infrastructure**, which provides a representative system with all technologies used within the central control system for testing of new equipment, development of hard- and software and the opportunity to integrate subsystems in a controlled, offline environment.

Combined with model-based and standardized software development [3], we see promising early successes with this very streamlined and industrial approach.

† pavel.bastl@eli-beams.eu

ISBN: 978-3-95450-193-9 doi:10.18429/JACoW-ICALEPCS2017-FRAPL05

## STRUCTURAL OVERVIEW

Figure 1 shows a structural overview of the control system, which is divided into top level control (upper part) and local level control (lower part).

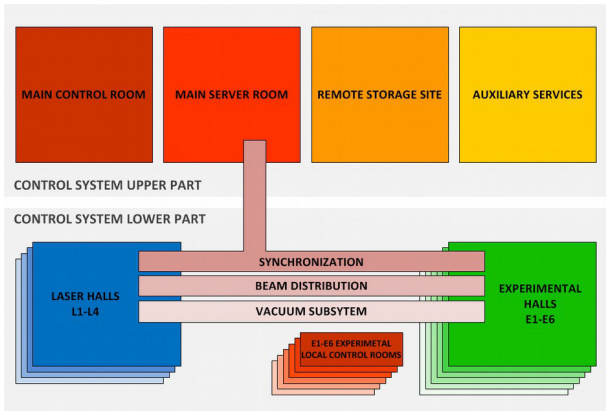


Figure 1: Control system structure.

The **top level control system** is responsible for integration from a functional and access point of view. The core components are located in the main server room (Figure 2) which directly connected to the main control room with its 21 seats (Figure 3). Auxiliary systems are installed in dedicated rooms: The timing system infrastructure (GPS antenna / receiver) under the roof of the building to reduce distances, and hardware for the central vacuum systems inside of the plant rooms.

Figure 2: View of the main server room.

Figure 3: Main Control Room, during installation (09/2017). The room is operational, large screens will be delivered 10/2017

The main server room also houses control racks for the (independently developed) laser control systems [4] and the *Eclipse* HPC [5] that is currently used for simulations.

The **local level control system** components are physically distributed over experimental and laser halls and form a distributed control system.

At this level, local control and local data acquisition ares implemented, and infrastructure for secondary sources and user experiments is provided, including local control rooms and standardized service hubs with access to network, data acquisition and timing systems as well as power and technical gasses.

All these systems are interconnected with a **network** based on more than 15.000 single-mode optical fibres. Mainly for safety reasons, but also to be able to optimize for performance, we actually divided it into three physically separated networks, for control, synchronization and data acquisition; following the *scientific DMZ network architecture* [6]

Finally, all network, control and DAQ systems are represented within our **test-bed** (Figure 4), which is also connected to a vacuum test system and a fully-equipped optical laboratory for integration tests.

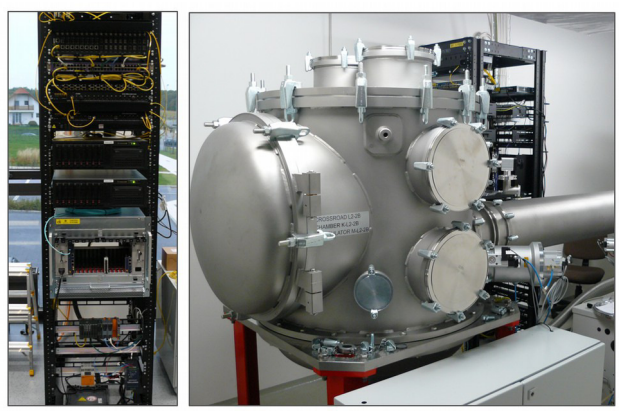

Figure 4: Test-bed Control/DAQ rack, vacuum test system.

ISBN: 978-3-95450-193-9 doi:10.18429/JACoW-ICALEPCS2017-FRAPL05



## LOGICAL OVERVIEW

As already introduced in [7][8], the ELI Beamlines control hardware can be divided into three separate logical units following the physical network division:

- Control System
- Data Acquisition System
- Synchronization System

### *Control System Architecture*

#### Network Overview

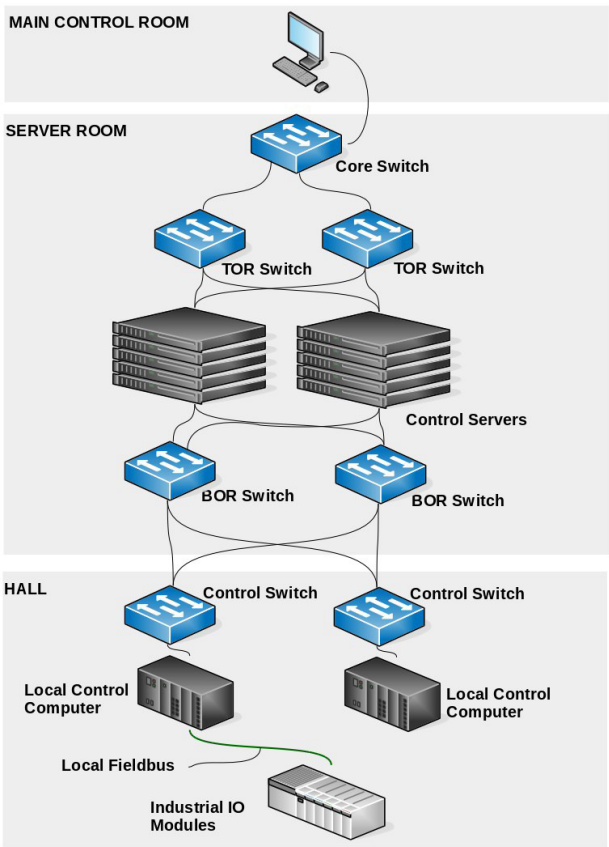


Figure 5: Control System Network.

Figure 5 shows a top-down overview of the control network, showing the communication from the user/operator in the main control room down to the equipment inside the experimental or laser hall. The entire network is a standard Ethernet network which is configured for high throughput, with no guarantee of synchronicity (real-time behaviour) in all but the case of the local fieldbus network, which will be discussed later. All components are connected to the system redundantly to increase reliability.

The main control room is connected to the server room via the core switch (*Cisco Nexus 7700*), which acts as a high-throughput, top-level aggregation and access switch. In the control room, control servers are connected to the upper layer using Top-Of-Rack (TOR)-switches (*Cisco Nexus 5672*) and to the lower layer using Bottom-Of-Rack switches (*Cisco Nexus 56128*).

The lower level is connected to this system using control switches (*Cisco Catalyst 2960X*) integrated into local service racks. On the field level, different types of DIN-rail mounted industrial switches are used.

#### Control Servers

Ten standard 2U servers (Figure 6) are used to implement the top level control servers (*Lenovo System x3650m5*), each has24 cores in total, 256GB of RAM memory and contains two NICs with two 10Gb/s SFP+ interfaces connected in redundantly as shown in Figure 5.

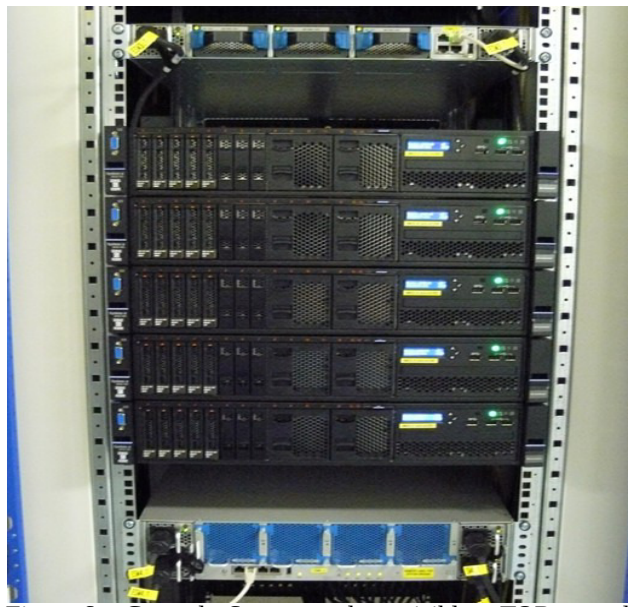

Figure 6: Control Servers, also visible: TOR- and BOR-Switches.

#### Local control hardware

In ELI, we distinguish between two classes of local control hardware:

- **Industrial control hardware** is used for undemanding applications (for example vacuum or motion control) with standard industrial interfaces, analog and digital I/O, motor controls and similar requirements. Real-time in the range of 100us may be required, as well as the possibility of optical-fibre communication due to long distances and EMP.

- **Advanced control hardware** is used for challenging applications, with high demands on data rates, response times or complex processing needs.

ISBN: 978-3-95450-193-9 doi:10.18429/JACoW-ICALEPCS2017-FRAPL05

### Industrial Control

The market offers a wide range of comparable industrial control systems and PLCs that fulfil the basic requirements as described above. Because of the large volumes in ELI, price is a significant deciding factor – and so is interoperability.

Two major systems stand out by being based on open, accessible and consortium driven fieldbusses on top of Ethernet: *PowerLink* [9], mainly represented by *B&R* [10] and *EtherCAT* [11], mainly represented by *Beckhoff* [12]. Both are capable of real-time operation and can be used in a ring-topology for redundancy.

While a classical PLC has to operate locally with the software provided by the vendor, the open standards allow us to use the interface cards alternatively

- directly with our own C++ based stack (without interfaces or PLC code); deterministic if used on a system with RT patch
- as field-nodes that are remote-controlled from the server room, with the fieldbus routed through our standard network hardware (required: VLANs and appropriate QoS configuration)

Both *EtherCAT* and *PowerLink* were successfully evaluated, the final decision for *PowerLink* was made based on cost and a more attractive range of available interface cards, as well as the openness of the IP core of PowerLink.

At this point, ca 40 IPCs (*PC910*) and PLCs and hundreds of interface cards are deployed, a number that is expected to grow rapidly over the next years of commissioning.

### Advanced Control

For more complex control demands, the physics-driven standard *MTCA.4* (*Micro Telecommunications Computing Architecture*) [13] was chosen – because of its high flexibility and modularity, redundant key components, agnostic backplane and advanced management.

A basic MTCA system starts with a crate / chassis which is managed by a *MTCA carrier hub* (MCH). This unit takes care of powering (often via redundant power supplies), cooling and switching / timing signals in the backplane, which can be configured to be PCIe, 10Gb Ethernet or Serial Rapid I/O. The crate can hold up to 12 *Advanced Mezzanine Cards* (AMCs) which communicate with each other over the backplane.

In ELI, we use two kinds of AMCs:

- CPU Boards (*AM90x/41x, Concurrent Technology*); within one crate, there can be multiple boards, allowing us to have independent systems.
- FMC Carrier Boards with *Artix XC7A200T* and *Kintex XC7K325T (Creotech)* These boards carry *FPGA Mezzanine Cards (FMCs)*, which can be flexibly chosen and exchanged. We currently have FMCs for timing, DI/O and analog acquisition.

At the moment, 22 MTCA crates are used in ELI, with 32 CPU boards, a number which is expected to grow as more demanding control applications arise.

## *Data Acquisition Architecture*

### Network Overview

Figure 7 shows the DAQ network in ELI [7][8]. Following the data flow, we start with two different types of **local DAQ hardware** in the experimental and laser halls.

We connect them using NICs (SFP+/QSFP) to FPGA cards installed in the **DAQ server** in the server room. At a later point, we plan direct interfaces from local FPGA cards in order to bypass PCIe limitations and reduce CPU load on the local hardware.

From the DAQ server, the data either goes into a multi-tier storage or to further processing using a a low-latency network, in our case Infiniband (which might be changed in the future).

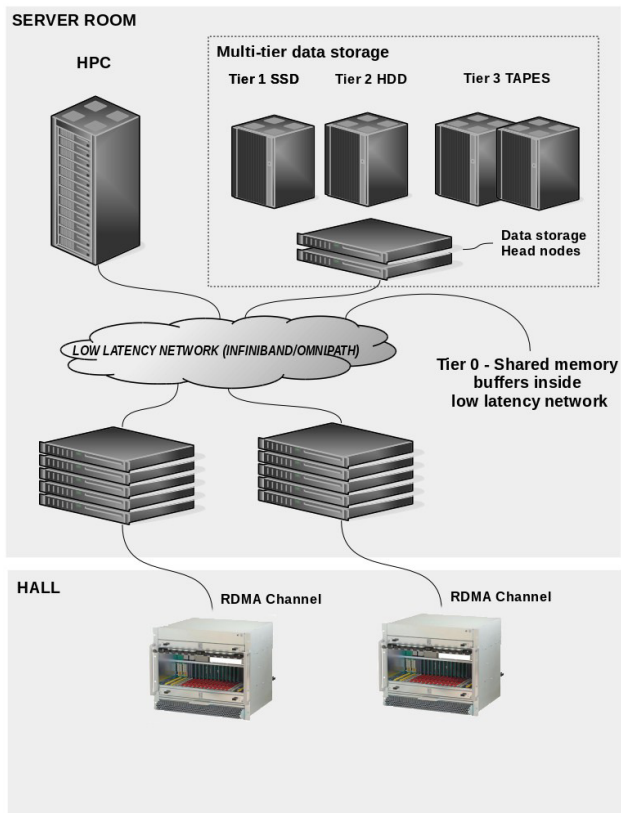


Figure 7: Data Acquisition Network.

ISBN: 978-3-95450-193-9 doi:10.18429/JACoW-ICALEPCS2017-FRAPL05



**Top Level Data Acquisition System**

In the server room, data is acquired, processed and stored using three main components

- a blade server for DAQ which contains a 2x Infiniband FDR switch in the rear, a 2x Ethernet switch 10/40GBASE-X in the rear and currently 14 blades, each blade has 24 cores and 768GB of memory for the buffer pool.
- the Infiniband [14] network (network interface cards, cabling and switches) acting as a low-latency interconnection inside the blade server and to the data storage
- the multi-tier data storage where the tier-1 is based on flash drives, tier-2 is based on standard hard drives and tier-3 is based on tape library. (Implemented in stages, initially 6PB)
  The RAM of the DAQ server, which acts as a memory buffer, is considered tier-0.

The DAQ servers are therefore used for three purposes

- **Data aggregation** from the DAQ hardware installed in the hall using either pure NICs (*Mellanox MCX4121A-XCAT*: 2x 10GBASE-X, RDMA support) or FPGA cards (*Alpha-Data ADM-PCIE-KU3*: FPGA XCKU060, 2x QSFP, SDAcell support, 8GB DDR; *Mellanox Innova Flex LX-4 :*FPGA XCKU060, 1x QSFP, RDMA support, 2GB DDR*)*
- as a **memory buffer pool** (tier-0) that is safe from EMP and can be shared across processing units using Infiniband, giving also access to associated systems such as our HPC
- as a host for **online data processing** using both computre cores and the above described FPGA cards for acceleration. We do prefer FPGA as core accellerator (directly incoming data, no bottleneck from PCIe), but provide some Xeon Phi / GPU for users.

**Local Data Acquisition Hardware**

In ELI Beamlines, there are two main reasons for high data rates:

- Our scientists are often working on very short phenomenons (femtosecond laser pulses), which require very high sampling rates
- Some lasers are operating with high repetition rates (1kHz) and need to be imaged (cameras and other 2D-detectors)

At the moment, we see applications with 10GS/s and more from digitizers (for example: *ADQ7-DC-F10-MTCA*) and are preparing for 2D-detectors with this and higher rates in the upcoming year.

We have two types of local DAQ hardware:

- standard **PCIe-based DAQ systems** (*Supermicro*) with 128GB of RAM, 24 cores and 10 PICex8 slots. These servers are low-cost, can provide large memory buffers (terrabytes) and there is a wide variety of PCIe-cards for different applications

- **MTCA-based DAQ systems**, which have the advantage of clock support (timing system, see next section) in the backplane and allow card-to-card-connection without involving the CPU. They have one limit: Due to the card size, the AMCs can only have 16GB of buffer.

However, when we want to use the advantages of both (clock support, card-to-card-connections, large memory buffers), our MCH (*NAT-MCH-PHYS80)* allows to connect its internal PCIe switch through optical cable to PCIe based local DAQ server. The connection setup (shown in Figure 8) provides PCIe x16 interface and can be implemented using PCIe on MTCAs' agnostic backplane and its fat pipes.

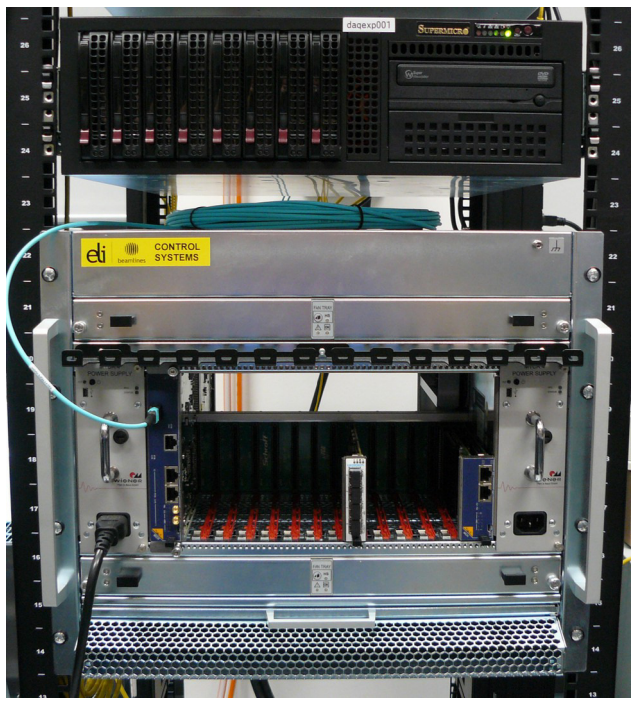


Figure 8: Standard PCIe DAQ system (top) connected with MTCA DAQ system (bottom) via optical fibre / PCIe bus

We use low-latency networks not only inside the server-room (in the form of Infiniband / OmniPath [14][15] – which is limited by distance), but also for the connection between local acquisition hardware and DAQ server in two forms: When NICs support it, we use Remote Direct Memory Access (RDMA); with normal Ethernet infrastructure we use RDMA over Converged Ethernet (RoCE).

ISBN: 978-3-95450-193-9 doi:10.18429/JACoW-ICALEPCS2017-FRAPL05

### *Synchronization Architecture*

The synchronization network, as shown in Figure 9, is used for the Electronic Timing System and based on WhiteRabbit [16].

In such a network, one switch is configured as the "GrandMaster", which distributes absolute timing information and an external clock to the entire network of connected switches using Synchronous Ethernet and PTP.

In ELI, we actually have the option to switch between two sources of precise time: we get time via CESNET [17], which uses a Cesium clock and a White Rabbit Grandmaster in Prague and roots the signal over ca 40km optical fibre to the ELI facility; and via a GPS receiver that is installed on top of the roof and acts as a backup (Figure 10).

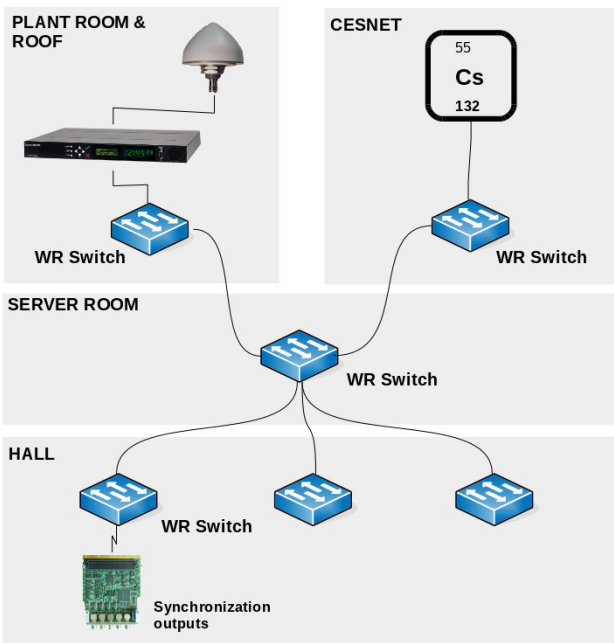


Figure 9: Synchronization network.

We currently installed 15 White Rabbit switches from Creotech with different types of FMCs (for both PCIe and MTCA carriers). In 2015, we tested the performance of our timing hardware and managed to reach sub-nanosecond accuracy and ca 9ps jitter on a still uncalibrated system [17].

In the near future, we will synchronize with the laser timing systems. This is planned in three phases:

First, simple triggers will be directly transmitted over spare optical fibres to the experimental halls – we have developed custom TX/RX modules with ns-accuracy for this and used them in several field-tests. This acts as ad-hoc solution for installation / comissioning.

As a next step, we transmit triggers (directly obtained from the lasers, ideally based on seed / injection signals) via WhiteRabbit. Control Systems software interfaces will be needed to compensate for varying beam travel time (for example: number of passes through an amplifier depending on laser power; varying distances from injector to experimental chamber).

Since the "heartbeat" of the laser facility is an optical oscillator (seed laser), that is disciplined by an 80MHz electronic oscillator, our final goal is to use this frequency as a basis for our timing system.

Figure 10: A White Rabbit "Grandmaster" switch is connected to an antenna under ther roof.

## OUTLOOK

In the upcoming years, lasers and secondary sources will be gradually installed and put into operation, which will be certainly challenging for the control systems team.

Specifically in 2018, we are expecting to work on the interfaces to two lasers (L1 / L3) and to control a number of secondary sources in their early stage of operation (for example ELIMAIA , HHG, an Ellipsometer, TEREZA, PXS and the MAC chamber with mostly vision/mottion control and detectors. The L3 beam transport will also provide another challenge with a very high number of controllable devices. The hardware is certainly ready for that, the challenge will be the efficient workload management of the local control implementation.

The data acquisition needs of the users are still quite modest, but expected to rapidly grow once the secondary sources are producing signals for the detectors. The way the systems and network were designed allows for easy scaling, we are first planning to increase storage capacities up to 12 PB, and maybe later on increase local transceiver modules from 10GB/s to 40GB/s or even 100GB/s.

Another topic are getting for is network load management. Like most facilities, we see high fluctuations in network load and anticipate these to worsen with increasing data loads. We are planning to use our top level DAQ server hardware as a **buffer for network load balancing.** Many network component vendor (including CISCO, which is currently our main vendor) are starting to offer technologies for automatic balancing by directing the traffic on switch level, which we believe to be foolish: Such systems can only work well in homogeneous environments, and lead to vendor lock-in. Instead we are working a pilot project and evaluating the capability of Infiniband (Mellanox) [14], OmniPath (Intel) [15], and CAPI (OpenPOWER foundation) [18][19] together with our Tier-0 storage as a network RAM buffer for load flattening.

ISBN: 978-3-95450-193-9 doi:10.18429/JACoW-ICALEPCS2017-FRAPL05



## CONCLUSION

This paper gave an overview of the hardware architecture of ELI Beamlines, as it exists in the commissioning phase of 2017. Our baseline is standing and operational, and we are now anticipating early operation to see how it holds up to reality.

## REFERENCES

[1] ELI Beamlines, https://www.eli-beams.eu

[2] B. Plötzeneder, V. Gaman, O. Janda, P. Pivonka, P. Bastl, "Cameras in ELI Beamlines: A standardized approach", in *Proc. ICALEPCS'2017*, Barcelona, Spain, paper THMPL06 (this conference)

[3] P. Bastl, O. Janda, A. Kruchenko, P. Pivonka, B. Plötzeneder, S. Saldulkar, J. Trdlicka, "Control System Software Environment in ELI Beamlines" in *Proc. ICALEPCS'2017*, Barcelona, Spain, paper THPHA171 (this conference)

[4] J. Naylon et al, "Control System Architecture for the L1 Laser" in *Proc. ICALEPCS'2015*, Melbourne, Australia, paper TUD3O02

[5] Computer Cluster Eclipse https://www.eli-beams.eu/en/facility/computing-simulations/computer-cluster

[6] E. Dart, L. Rotman, B. Tierney, M. Hester, J. Zurawski, "The Science DMZ: A Network Design Pattern for Data-Intensive Science" in *Proc. SC'13: The International Conference for High Performance Computing, Networking, Storage and Analysis*, Denver, CO, USA

[7] P. Bastl, "Control system architecture: HW architecture", Internal Design Report, 08/2014

[8] P. Bastl, "ELI Tango Workshop", Szeged, Hungary,02/2015 http://www.eli-alps.hu/sites/default/files/tangows/20150224-1450-ELI-Beamlines-DAQ-PavelBastl.pdf

[9] OpenPOWERLINK Consortium, http://openpowerlink.sourceforge.net/web/

[10] B&R Automation, https://www.br-automation.com

[11] EtherCAT Technology Group, https://www.ethercat.org/

[12] Beckhoff Automation, https://www.beckhoff.com/

[13] H. Schlarb, T. Walter, K. Rehlich, F. Ludwig, "Novel crate standard MTCA.4 for industry and research" in *Proc. IPAC'2014, paper THPWA003,* Dresden, Germany

[14] Mellanox Infiniband, http://www.mellanox.com/

[15] Intel OmniPath, https://en.wikipedia.org/wiki/Omni-Path

[16] The White Rabbit Project, https://www.ohwr.org/projects/white-rabbit

[17] V. Gaman, P. Bastl, White Rabbit Timing System Evaluation, in *ELI Beamlines Scientific Challenges 2015*, Prague, Czech Republic

[18] OpenPOWER Consortium, https://openpowerfoundation.org/

[19] OpenCAPI Consortium, http://opencapi.org/