



OPEN ACCESS

EDITED BY

Claudio Zunino,
CNR Institute of Electronics, Computer and
Telecommunication Engineering, Department
of Engineering, ICT and Technology for Energy
and Transport, Italy

REVIEWED BY

Isaias González,
CNR Institute of Electronics, Computer and
Telecommunication Engineering, Department
of Engineering, ICT and Technology for Energy
and Transport, Italy
Luca Durante,
Ionized Gas Institute, National Research
Council (CNR), Italy

*CORRESPONDENCE

Gabriele Manduchi,
✉ gabriele.manduchi@istp.cnr.it

RECEIVED 22 November 2023

ACCEPTED 24 July 2024

PUBLISHED 07 August 2024

CITATION

Manduchi G (2024), Overlaps between
industrial informatics and control, data
acquisition and management in Big Science.
Front. Ind. Eng. 2:1342734.
doi: 10.3389/fieng.2024.1342734

COPYRIGHT

© 2024 Manduchi. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Overlaps between industrial informatics and control, data acquisition and management in Big Science

Gabriele Manduchi^{1,2*}

¹Consorzio RFX (CNR, ENEA, Università di Padova, INFN, Acc. Venete), Padova, Italy, ²Istituto per la Scienza e la Tecnologia dei Plasmi, CNR, Padova, Italy

Big Science applications require very large infrastructures and often involve different countries in order to achieve important scientific results or to find solutions to the major problems of mankind, such as finding a clean and endless source of energy. Big Science applications represent not only a scientific challenge, but also large engineering applications involving a wide range of technologies shared with other industrial applications. As a consequence there is a significant overlap in technologies and approaches between Big Science and Industry. In this paper, the overlap between Big Science and industrial applications will be presented in more detail under the control perspective, that is, by highlighting the common aspects between industrial informatics and the control, data acquisition and data management in large scientific applications.

KEYWORDS

industrial informatics, data acquisition, control systems, scientific applications, open source, PLC

1 Introduction

The term “Big Science” describes the current trend in scientific research toward large scale experiments. While in the past relevant scientific results could be obtained by small groups in university or other research laboratories, major scientific achievements nowadays require a much larger infrastructure often involving different countries. To better illustrate what Big Science means, let’s consider two recent major results in physics research: the evidence of the Higgs Boson achieved in the CERN laboratories in 2012 (CERN, 2023) and the first observation of gravitational waves in 2015 detected at the twin LIGO laboratories (LIGO, 2023). The discovery of the Higgs Boson, almost 50 years after it was first proposed, is the result of the statistical analysis performed over a huge amount of measurements derived by collision events between accelerated protons in the CERN Large Hadron Collider (LHC) (The CERN LHC, 2023). LHC consists in a 27 Km ring where superconducting magnets boost the energy of two proton beams flowing in opposite directions up to a value that is of the same order as the kinetic energy of an Airbus 380 flying at a speed of 720 km/h. The two proton beams flow in separate paths except for four points in the ring where they are deflected in order to let them collide. The quarks and gluons inside the colliding protons interact to form a wide array of low-energy, ordinary particles. Occasionally, heavier particles are produced as well as energetic particles paired with their antiparticles. Every sector hosts different detector types so four separate experiments are hosted along the LHC. ATLAS and CMS use general-purpose detectors to investigate the largest range of physics

possible. ALICE and LHCb have detectors specialized for focusing on specific phenomena. The CERN laboratories hosting the LHC employ approx. 2,500 people on a permanent basis and more than 10,000 visiting researchers from all around the world. The successful exploitation of scientific research involves not only physicists but also a large number of engineers and technicians who are involved in the development and maintenance of the experiment components such as:

- *high vacuum* system for the pipes along the 27 km ring;
- *cryogenic* systems for maintaining ultralow temperature for the superconducting magnets bending the proton beams;
- *power supplies* driving magnet current and transferring energy to the beam;
- *support electronics* for the particle detectors;
- *precise timing* system to timestamp events and validate real events against noise;
- *data system* for storing and processing petabytes of data per day;
- *grid infrastructure* to let scientists all over the world access and analyze experimental data.

On Sept. 14, 2015, the gravitational waves originated from the collision of two black holes that occurred 1.3 billion years ago were observed for the first time in both the LIGO twin interferometers, thus confirming a major prediction of Albert Einstein's 1915 general theory of relativity. Each LIGO experiment consists of two 4 km long interferometers arranged in the shape of an L. The interferometers are locked in such a way that the light waveform of a laser generator destructively interferes, i.e., the generated and reflected waveforms elide each other producing no light. When a gravitational wave arrives, the length of the two arms in the L change so that the Laser interference changes and light is generated. The twin interferometers, located in Louisiana and Washington state, respectively, allow detecting the origin in the universe of the gravitational wave based on the difference in time at which the event has been detected. Despite the simplicity of the detection ground principles, the LIGO project represents a great engineering challenge because the change in the 4 km long interferometer arms, due to the interaction with the gravitational wave, is in the order of $1\text{E}-19$ m, i.e. 1/1000th the size of a proton, and the detection system must therefore compensate all the possible sources of vibration, such as traffic in the nearby roads and earthquakes, that are orders of magnitude larger. This is achieved via sophisticated active and passive vibration-damping systems. Ultra-high vacuum for a volume surpassed only by CERN LHC in the world, extremely accurate optic systems for the laser mirrors, and highly pure laser light generation represent some of the engineering challenges in the LIGO project. The amount of data required to discriminate real events from all the other noisy sources is several terabytes per day and cloud data availability is provided in order to share data with the other gravitational antenna experiments around the world to synchronize and validate detected events.

Big Science is not only involved in pure physics research, but it addresses also other important achievements of mankind, such as finding a clean and endless source of energy. In this context, nuclear fusion represents currently the main challenge. Reproducing on earth the nuclear fusion phenomena occurring on the sun would

provide a carbon-neutral source of energy avoiding those problems that affect nuclear fission reactors, i.e. (1) scarcity of fuel for the nuclear reaction, (2) risk of uncontrolled nuclear reactions and (3) nuclear waste management. Nuclear fusion is based on the reaction between deuterium and tritium, two isotopes of hydrogen with 2 and 3 neutrons, respectively, and producing an atom of helium and a fast neutron. An unlimited source of deuterium is available in the earth's oceans while tritium is produced in the reactor itself as a side product of the interaction of the generated neutrons with a lithium blanket. Intrinsic safeness, i.e., the impossibility that a failure in the reactor could produce uncontrolled fusion is guaranteed by the nature of the nuclear reaction itself. Finally, nuclear waste produced by the activation of the material during the fusion reaction has a decay time that is orders of magnitude shorter than that of materials that are activated by a fission reaction. The major technological challenge in fusion reactors is due to the fact that the mixture of deuterium and tritium must be kept at a temperature of 10 Million degrees in order to achieve nuclear fusion and the only way of keeping such a mixture of ionized gas (called plasma) in a container is to make it levitate by means of strong electromagnetic fields generated by coil currents so high that they can only be achieved by keeping the coils in the superconducting state. The inner wall of such a container (doughnut-shaped and called Tokamak) must sustain a heat flux in the order of 10 MW per square meter and at the same time, the whole tokamak and the surrounding coils must be immersed in a large cryostat at 4.5 K temperature. As with several other Big Science projects, also fusion research is the result of worldwide cooperation. ITER, currently the largest fusion device in the world ([The ITER project 2023](#)) is under construction in France and it is the result of the collaboration among Europe, the United States, China, Korea, Japan, India, and Russia.

From the above examples, it can be seen how Big Science applications represent not only a scientific challenge but also large engineering applications involving a wide range of technologies shared with other industrial applications. The overlap between Big Science and industrial applications will be presented in more detail under the control perspective, i.e. by highlighting the common aspects between industrial informatics and the control, data acquisition, and data management in large scientific applications. In the following, [Section 2](#) analyzes in detail the similarities and differences between industrial informatics and control and data acquisition systems in Big Science. These are presented in the context of different fields of application in order to cover most use cases that can be found in industry and research (plant control, real-time applications, FPGA applications, timing systems, data acquisition and storage, cloud and grid computing, and machine learning). The presented concepts are then summarized and discussed in [Section 3](#). Some general conclusions are finally drawn in [Section 4](#).

2 Application specific similarities and differences

Industrial informatics refers to the infrastructure that provides the development and deployment of real-world applications. Such an infrastructure consists of a collection of techniques and practices

that use information analysis and tools to achieve higher efficacy, effectiveness, reliability, and security within the industrial environment. These techniques represent the heart of the Industry 4.0 concept that encompasses control, the Internet of Things, artificial intelligence, robotics, and automation. The above definitions can be rephrased by replacing “industrial” with “research” to obtain the definition of the objectives of Control, Data Acquisition, and Management in Big Science. Of course, reality is more complex and requires a more detailed discussion based on more specific application fields, as shown below.

2.1 Plant control

Common components in Big Science are large vacuum and cooling systems. For CERN LHC, an ultra-high vacuum is necessary for the beam pipes. In LIGO, an ultra-high vacuum is required for the interferometer arms. In ITER, the 840 m³ volume where the plasma is formed must be kept in vacuum and the heat generated by nuclear fusion must be fully removed by cooling systems. Cryogenic systems are also required to achieve high vacuum via cryopumps and to cool the superconducting coils required to achieve extremely high electromagnetic fields to bend the beam in particle accelerators or to confine the plasma in fusion devices. Control for these plants is carried out by Programmable Logic Controllers (PLCs) and industrial solutions are often adopted. Indeed, these plants are often commissioned to industry rather than internally developed. Due to the large dimension of such plants, different components, possibly developed in the context of different contracts, must be integrated, relying in most cases on common practices in industrial informatics. As an example, the control of the large helium cryogenic system in the CERN LHC is compliant with the standard automation pyramidal organization of IEC-62264 and defines (1) an instrumentation layer that integrates a large number of industrial sensors and actuators via field buses, (2) a control layer based on standard industrial components (PLCs) and (3) a supervision layer implemented by a cluster of Linux Data Servers (Pezzetti, 2021).

Industrial control applications for vacuum, cooling, and cryogenics in Big Science have a large overlap in requirements with other applications outside research. Even if building the controlled equipment often represents a challenge with respect to other industrial plants (e.g., for their dimension, for the heat flow, and for the required vacuum level) the requirements in control functions and their dynamics do not differ significantly with respect to those of other industrial plants. Considering also the number of I/O signals in the plants involved in Big Science, this can be considered of the same order of magnitude with respect to the number of signals involved in a large industrial facility or a transportation system. As a consequence, the challenges in this context are similar for Big Science and large industrial appliances. Here the main challenge is the effective management of systems that are composed of a huge number of components that, not being complex *per se*, require strategies for handling overall system complexity. An important strategy in this context is automatic code generation targeting the difficulty in developing safe and efficient program code in large systems. For this reason, methods and tools for the automatic generation of PLC code based on high-

level system descriptions, including I/O signal lists and transformation specifications, are a common practice in large industrial applications. However, a general approach for automated code generation is missing (Koziolek et al., 2020). This is true also in Big Science where site-specific solutions for PLC code generation have been developed. For example, CERN developed UNICOS (*UNICOS Framework portal 2017*), a control system framework for designing and implementing control systems applications. It provides a methodology, an object library, and a set of tools to generate the control code for the target applications. A different solution addressing similar requirements has been developed at ITER in order to generate the code to coordinate roughly 200 plant systems and to integrate them into the control infrastructure (Stepanov et al., 2011). In this case, the high-level description, supported by a specialized editor, includes not only industrial control components but extends also to other services such as data archiving and network configuration.

A peculiar aspect of industrial control in Big Science applications is the widespread use of open-source SCADA tools in the control of the industrial parts of the experimental facility. The most widespread open-source SCADA solution is EPICS (*Experimental Physics and Industrial Control System 2023*). EPICS is an open-source software tool collaboratively developed and used in several experiments around the world, such as particle accelerators and telescopes, and adopted also in ITER, whose complexity is comparable to CERN LHC, for the integration of its plant control systems (Leone et al., 2023). The main concepts of EPICS are not different from those of other modern supervisory tools, defining a set of Process Variables (PVs) and exporting the value of physics quantities over the network. Above the core level of EPICS, handling the management and distribution of PVs, the Control System Studio provides a collection of tools and applications to monitor and operate large-scale control systems, such as the ones in the accelerator community (*Control System Studio, 2024*).

A notable exception in the accelerator community is CERN which uses Siemens WinCC-OA Supervisory Control and Data Acquisition (SCADA) System. It is worth noting that Siemens WinCC-OA derives from the former PVSS SCADA system, adopted by CERN in the year 2000 after a 3-year evaluation phase (SIMATIC, 2023). The industrial product WINCC-OA can be considered a spillover of CERN and represents a successful example of the synergy that can be achieved between industry and research.

The diffusion of EPICS and other open-source solutions in the scientific environment is also a consequence of the collaborative attitude of research people, often moving from one laboratory to another and bringing ideas and solutions. It is worth noting that what is considered a plus in the research environment, i.e., sharing ideas and solutions, can be a drawback in the industrial world where nondisclosure is often a requirement. Moreover, despite the advantages offered by open-source SCADA solutions, there are also drawbacks that need to be taken into consideration in design choices. This is in particular true in large experimental facilities where many plant systems are not developed within the scientific community but commissioned to industry. Typically, the (associations of) industries that are involved in plant systems development bring their internal expertise, including the support

tools for the development of plant control systems. In this case, rather than dictating the technology to be used, it is necessary to define the interface between the specific plant and the rest of the system. This interface should be defined via methods and protocols that are familiar in the industrial world, and in this case, adopting open-source components that are well known within the scientific community but not outside can bring additional costs. In any case, open-source solutions are gaining interest also in the industry. The most notable example, with implications also in Big Science, is the OPC Unified Architecture (UA) which is a cross-platform open-source IEC 62541 standard for data exchange (OPC, 2024). OPC UA has been adopted as the standard for Industry 4.0 to support digital transfer. This standard is implemented both in commercial SCADA systems and in open-source libraries and currently represents the preferred choice in the integration of plant systems in many Big Science projects.

2.2 Real-time applications

In industry, real-time processing is crucial for businesses that require continuous improvement in safety, efficiency, and reliability. Applications of real-time systems include process control systems, machine vision, robotics, manufacturing, and healthcare. Real-time applications are more and more involved in edge computing, where data processing occurs closer to where it is being generated, e.g., in Industrial Internet of Thing (IIOT) system, in order to improve response time and save bandwidth in the overall system communication (Musaddiq et al., 2018). Several Real-Time Operating Systems (RTOS) have been adopted, whose evolution towards more and more distributed applications followed the evolution in hardware towards multicore architectures. Nowadays many industrial applications are moving towards implementations based on Real-Time Linux. Originally not fit for real-time, the evolution of the Linux OS and in particular of its real-time extensions such as Messaging Real-time Grid (MRG), combined with more powerful computing hardware and multicore architectures, provide real-time performance parameters such as response time and jitter that are quite acceptable for a large number of industrial applications. Moreover, real-time performance integrated into a general-purpose operating system such as Linux allows prioritizing, managing, and executing real-time workloads over non real-time workloads, unifying the software and hardware solutions for complex systems. It is worth noting that Linux and its real-time extensions are not the unique emerging RTOS in both industrial and scientific environments. FreeRTOS, combining proven robustness, tiny footprint, and wide device support, is one emerging RTOS that is gaining more and more widespread usage in IIOT as well as in research applications (Guan et al., 2016). In general, for larger distributed applications involving many different actors with different requirements in hardware and software, a combination of RTOS may be required. For example, in an IIOT framework, FreeRTOS may represent the best choice for the smaller devices close to the sensors, while Linux may be best suited for the higher-level servers collecting overall data.

When considering overlaps in real-time requirements and solutions between industry and Big Science it is convenient to identify two broad categories under the common denomination

of real-time systems. The first category refers in general to distributed information processing systems with hardware and software components that can respond to events with predictable time constraints. This is the typical context of IIOT, involving a network of interconnected devices and sensors in an industrial setting. Examples of use cases in this category are industrial automation and energy distribution monitoring. The second broad category refers to real-time control of a given equipment such as an electrical power supply or a jet engine. In this case, timing constraints are typically stricter, being the controlled entity a physical component rather than a distributed infrastructure. This last category represents historically the starting point of the concepts and the solutions in real-time systems and in Big Science the focus on real-time is more oriented to this last category being often the experiment itself the target of real-time control. Depending on the nature of the experiment, computer-based real-time control may address the experiment core or be required for accessory, but crucial, aspects. In the LIGO gravitational antennas, for example, real-time systems are used in the active vibration damping where the vibration originated from human activity (e.g., traffic) and environmental (e.g., earthquakes) close to the detector site must be canceled. Vibration sensors send their signals to a computer that combines all the vibration signals and generates a net counteraction to cancel all external vibrations simultaneously (Matichard et al., 2015). In CERN LHC there is no need for closed-loop real-time control systems for the main experiment (albeit a large number of accessory instrumentation require active real-time controls) but real-time systems are nevertheless required in order to provide online analysis for the huge amount of data coming from the detectors in correspondence to proton-to-proton collisions. In the most recent implementation of the ALICE detector system, more than 3.5 terabytes per second are produced by the continuous readout of 12 billion detector pixels. After a first-level data processing carried out by FPGA, a stream of up to 600 gigabytes per second is produced and analyzed online on a high-performance computer farm, implementing 250 nodes, each equipped with eight GPUs and two 32-core CPUs. This allows us to further reduce the rate to a maximum of 100 gigabytes per second before writing the data to the disk (Nowakowski et al., 2024).

Among Big Science applications, computer-based active real-time control of the core phenomena is most important in nuclear fusion experiments. In fusion devices, real-time control of the plasma within its container is essential in order to ensure confinement and avoid any contact between the 10 M degree plasma and the tokamak inner wall that would otherwise melt immediately. To achieve plasma confinement, the system acquires real-time information about the magnetic fields via several hundreds of electromagnetic sensors as well as other plasma parameters derived by other plasma diagnostics and provides a sub-millisecond system reply in terms of reference currents for the power supplies. The power supplies then feed the superconducting coils in order to generate a response electromagnetic fields and compensate plasma instabilities (Perek et al., 2023). The dynamics of the involved physical phenomena in the plasma confinement require a reaction time for its control in the order of some hundreds of microseconds that is compatible with state-of-the-art technology in computer hardware and software. Due to the large amount of requested computation, mostly required to

derive plasma parameters from the input sensors measurements, and due also to the physical location of the sensors and of the actuators in a large experimental plant, the plasma control in large fusion experiments is implemented by a network of computers connected via a real-time network. Such a network was implemented in the past with ad-hoc solutions, but now the mainstream 10 Gigabit Ethernet allows reaching the required performance in network latency.

An additional requirement in real-time applications derives from their experimental nature and the research environment, requiring frequent updates in the control algorithms or even in the control system layout, depending on the availability of new plasma diagnostics, or on new ideas in control algorithms. This requirement in flexibility is in contrast with other requirements in reliability that are mandatory in a system controlling a billions-worth experimental plant. Moreover, algorithms and technical solutions are often shared in the scientific community thus requiring an effective way of integrating new solutions still retaining the reliability of the system. The solution for incorporating all these requirements is represented by real-time frameworks, i.e., collections of collaborating classes that provide a set of services for a given domain. The framework is then customized to a particular application by subclassing and composing instances of the framework classes. In this way new components can be integrated in the system without the need of re-writing code, but just adding a new component in the customized framework instance. Reliability can be maintained by reusing well-tested components and concentrating only on the newly added functionality, such as the implementation of a new control algorithm. For the above reasons, open-source real-time frameworks have been developed by the fusion research community targeting the need for complex and performing real-time control. For example, MARTE2, a framework for real-time applications developed by the European ITER agency, combines the flexibility required to integrate new components with strict software quality requirements (Avon et al., 2021). Even if some research is ongoing (Delgado et al., 2023), open-source real-time frameworks for plant control are not common in the industry in respect of nonopen-source ones. Perhaps the most widespread nonopen source solution in this context is the Object Execution Framework (OXF) that is used in the code generated within the IBM Engineering System Design Rhapsody (IBM Engineering Systems, 2024).

A further step towards fast and reliable component integration in real-time systems is the integration of code generated by the Simulink coder within the framework. MATLAB Simulink represents the 'lingua franca' in the control community both in research and industry. The Simulink coder tool provides automatically generated C code corresponding to the selected Simulink component, removing the need for the demanding and error-prone manual translation of the component into C or C++. However automatic code generation of Simulink components alone is not enough to ensure a fast and reliable component integration in a framework if the interface adapter must be implemented manually. This is a common requirement both in industry and Big Science and indeed both the frameworks cited above provide a wrapper interface that is able to integrate the Simulink component directly into the framework without any manual code development, based on the introspection capability provided in the Simulink generated code.

If we turn our attention to the first cited category of real-time systems, i.e., distributed information processing systems and IIOT,

open source framework solutions are much more common. This is not a surprise, as the number of applications in this case is much larger in respect of applications referring to the real-time control of a given physical system. It is worth noting that the real-time management of streams of information carried out by open-source frameworks such as Apache Kafka (Apache Kafka Portal, 2024) or Redis (Redis 2024) is also needed in Big Science when considering the related infrastructure. Indeed the requirements in managing in due time the flow of information in a Big Science facility, such as sharing real-time experimental results along a worldwide community, are not different from other high-performance, low-latency information systems in the industry. For example, Apache Kafka is used at CERN for radiation supervision and environmental protection (Ledeul et al, 2019) and Redis is used in the Fusion Energy Sciences Network (ESnet) to deliver highly reliable data transport capability in data-intensive sciences (ESnet final report 2021).

2.3 FPGA applications

FPGA applications are used when required throughput and latency cannot be achieved with computers. Real-time computer applications can achieve at most a control cycle time down to 20–50 μ s in closed loop applications, therefore shorter periods require different solutions such as FPGAs. Now prevalent in high-performance computing, FPGAs offer the benefit of task-specific customization of a generic computing architecture. Examples of FPGA usage in the industry are smart energy applications for efficient, reliable, and intelligent energy systems, robotics for low latency, deterministic computing, and connectivity, machine vision for direct ingest of data and pipelined processing, and motor and motion control. Big science applications have a large overlap with industry when considering FPGA technology, sharing hardware solutions and software tools, but differ significantly when considering applications. This is in particular true in particle accelerator experiments where FPGA systems are extensively used as the first processing of the huge amount of raw data coming from detectors. The first level of data filtering is required to reduce the data flow to an amount that can be stored and processed on more traditional CPU-based processing units. In this context, FPGAs are used to run algorithms looking for particle signatures in the raw detector signals (Harris, 2021). FPGA programming normally involves simple computation expressed in a Hardware Description Language such as Verilog or VHDL. However, in the accelerator community, and in particular at CERN, there has been a strong push towards more complex FPGA applications. Including more complex algorithms, such as Kalman filters, in the first-level processing, greatly improves the overall detector efficiency, and it represents the key to the CERN achievements in particle physics experimentation. This has been made possible thanks to the High-Level Synthesis (HLS) tools that allow the use of a high-level language for generating FPGA firmware. Major FPGA development tools such as XILINX Vivado or Intel Quartus provide HLS compilers for high-level specifications such as C++, dramatically lowering the development time to the extent that students and physicists can contribute to large parts of the development of detector electronics systems (CERN courier

2024). Integration of FPGA applications via HLS has been pushed further at CERN to implement in FPGA complex deep learning algorithms. Deep learning solutions are being considered for replacing former algorithms to extract particle signatures for raw detector data (Blago, 2023) and their implementation in FPGA has a potentially huge impact on detector efficiency. The *hls4ml* project (CERN hls4ml 2024) aims at providing automated translation to HLS starting from traditional open-source machine learning packages such as PyTorch and Keras.

Even if FPGA applications are pushed to the extreme in accelerator experiments, industrial FPGA solutions and tools are adopted also in other Big Science applications. In particular, Systems on Chips (SoC), leveraging CPU performance in FPGA applications, are increasingly used in nuclear fusion experiments. As stated before, nuclear fusion experiments rely on complex real-time computer systems to achieve the confinement of the plasma in its container. Even if most controlled phenomena have dynamics that allow their control via a computer-based real-time system, other phenomena related to fast plasma instabilities require a faster response in control that can be achieved only via FPGA control. In this context, SoC solutions provide an easy integration of FPGA functions within the plasma control system, where FPGA computation is orchestrated by the embedded Linux controller. Other FPGA applications are involved in plasma diagnostic systems and provide intelligent data processing prior to real-time control. For example, in ITER data from several thousands of electromagnetic probes requiring filtering and integration will be processed by a cluster of Ultra Scale Zynq SoC systems (Batista et al., 2017).

2.4 Timing systems

Correct time synchronization in industrial systems represents an important requirement especially when different systems with independent clocks interact with each other (Balakrishnan et al., 2023). Examples of industrial applications where accurate time synchronization plays a crucial role are cloud robotics, smart grids, and drone-based sensors in industries such as mining oil and gas. When considering Big Science applications, accurate time synchronization is in general needed to correlate a possibly large set of sensor measurements in order to derive useful information about the physical phenomena being investigated. Time synchronization requirements can be divided into two broad categories:

- Relative synchronization, typical in industrial plants where the entities in a plant such as field devices, controllers, and computers are synchronized with each other. Relative synchronization represents also the required synchronization in all Big Science experiments. For example, a plasma instability in a nuclear fusion experiment triggers signal evolution over hundreds of sensors. If the acquired signals were not synchronized with each other to a precision extent that is compatible with the dynamics of the instability it would not be possible to derive useful information from the measurements.
- Absolute synchronization, necessary to achieve relative synchronization when a direct connection is not feasible between synchronizing and synchronized components. The

most popular way of achieving absolute synchronization is using a Global Positioning System (GPS) obtaining in principle an accuracy in synchronization of tens of nanoseconds. However, the unavailability of the GPS signal due to radio jamming and building structures as well as high installation/maintenance costs hinder extensive GPS usage for synchronization both in industry and in Big Science.

Absolute synchronization can be also achieved by means of the distribution of time over a network, such as Ethernet. The network evolution towards packet switching has led to increased interest in time synchronization using packet-based methods, such as Network Time Protocol (NTP), ubiquitous in PLC-based applications for industrial control. The NTP intrinsic time precision is in the range of 10–100 ms and it is enough for most PLC-based control applications in industry as well as in many plants in Big Science applications, such as vacuum and cooling systems. There are however applications requiring a tighter, sub-millisecond, synchronization in the telecom industry, finance segment, and smart grids. In this context, the IEEE 1588 Precision Time Protocol (PTP) has been increasingly adopted both in industry and research (Girela-López et al., 2020). The PTP protocol uses standard network lines to offer hardware-level time synchronization accuracy in nanoseconds range, provided all the involved network components, such as switches and routers, are PTP enabled. PTP-based synchronization is being adopted in several Big Science projects. As an example, ITER implements PTP-based synchronization for the synchronization in control and data acquisition for all the diagnostic systems of the nuclear fusion experiment (Liu et al., 2018). In this context, being data from physical measurements collected at possibly high frequency (>10 MHz), exact data sample timestamping is essential to validate data coming from different sensors and acquired by different systems, possibly reflecting the occurrence of a fast physical phenomenon in the experiment. In other Big Science applications, such as CERN, timing requirements are even more stringent with respect to what even a properly calibrated PTP synchronization network can provide. In accelerators such as LHC the detection of particle interaction decay products results in data coming from a potentially high number of different detectors. Very accurate, sub-nanosecond timing precision is required to discriminate meaningful data from noise via temporal coincidence analysis. This requirement triggered the White Rabbit Project in 2012 (The White Rabbit Project, 2023). White Rabbit provides sub-nanosecond accuracy in synchronization for large distributed systems in which devices are interconnected in a network. It combines PTP and Synchronous Ethernet (SyncE) with clock loopback and phase adjustment. SyncE uses the data carrier frequency of the physical layer interface and a digitally implemented Phase Locked Loop (PLL) removes the jitter generated by the clock recovery circuitry. The cleaned clock is sent back to the originating master that measures the phase difference between its reference clock and the received one. The measured phase difference is finally sent to the slave node so that it can correct its phase offset such that the delay variations in the physical link are compensated (Jansweijer et al., 2013). Fully implemented in open-source technology WhiteRabbit has expanded outside the field of particle physics. In 2020, it was included in the PTP industry standard, governed by the Institute of Electrical and Electronics Engineers.

2.5 Data acquisition and storage

Data acquisition and storage systems are ubiquitous in industrial applications. The importance of data in industry is, among others, a consequence of the Cyber-Physical System approach in modern industrial applications, where cyber components like cloud servers compute results and make decisions according to data collected or generated by physical components such as sensors. The data oriented paradigm has proven fundamental for the technological transformation process that characterizes Industry 4.0. The same paradigm always represented a pillar in Big Science and more in general in all scientific experiments. Indeed data, i.e., scientific knowledge, is the ultimate result of scientific research, and therefore data storage and database technology represent an important component of the control systems in scientific experiments.

Restricting our attention to Big Science, data requirements are further stressed by the dimension of the experiment and by the dynamics and complexity of the underlying phenomena being investigated. For example, in CERN LHC up to one billion particle collisions can take place at every second inside the LHC experiment's detectors. A trigger system is therefore used to filter the data and select those events that are potentially interesting for further analysis. However, even after this drastic reduction, 1 petabyte of data per day must be placed and stored. An even larger amount of data per day, currently estimated at 2 petabytes, will be produced by the ITER nuclear fusion experiment (Abla et al., 2014). In this case, the biggest generators of raw data are camera-based diagnostics taking redundant measurements that are required to reduce uncertainties and increase confidence in values that cannot be measured directly and must be inferred from other measurements.

When considering data acquisition, a current trend in scientific experiments is to use commercial data acquisition modules (DAQs) whenever solutions for specific requirements, such as data sampling rate, are available on the market, limiting as far as possible in house development. This is particularly true for plant control, where PLC-based data acquisition solutions are ubiquitous both in industry and research. There is moreover an increasing interest from industry, research, and academia towards emerging open-source SoC platforms for data acquisition such as RedPitaya DAQ that may provide cheap and flexible alternatives to bulky and expensive instruments.

When considering requirements and solutions in database technology there are similarities and differences between industry and Big Science. To better understand them, it is necessary to identify what data represent in the two contexts. Borrowing terminology from Industry 4.0, an asset is defined as a "physical or logical object owned or held in custody by an organization, having a perceived real value to the organization" (De Oliveira et al., 2021). An asset in the industry context can be something physical (equipment, materials, products) or not (electronic documents, computer programs) or represent some sort of metadata, such as location, time, state of an asset, and relationships with other assets. In the research context, an asset will be most likely represented by a measurement and the associated metadata describing how such measurement has been derived. NoSQL data model is beneficial here because of the heterogeneity of data types and aggregates

required to describe assets both in industry and in research. An added requirement is the need for storing time series, i.e., samples over time for acquired plant signals. Even if time series may need to be stored both in industry and in scientific experiments, the requirements for the latter are typically much more stringent. Solutions such as InfluxDB and TimescaleDB, widely used in industry for time series databases, are not feasible in Big Science for all signals, but typically only for signals produced by the industrial components of the experimental application. For several Big Science applications, the intrinsic organization of the experimental time-dependent results is best described by a hierarchical database such as HDF5 (The HDF Group 2023). HDF5 is widely used in scientific simulations producing a very large amount of data and this database has been chosen in ITER to store the data produced by the experiments.

In addition to more stringent timing requirements (i.e., data sampling rate) that cannot be withstand by current industrial solutions, two additional requirements are more specific to scientific data, namely:

- Data dependency, experimental results are very often not represented by a direct measurement (such as a temperature or pressure measurement in an industrial plant), but they represent the result of an acquisition and computation chain involving parameters and raw data acquired by a possibly large set of sensors. In order to provide self-descriptive data, both raw and computed data must be stored in the database. Keeping raw data in the database after they have been used to derive scientific results is necessary because of the experimental nature of the application. For example, a new derivation of scientific results is required when a new, better algorithm developed by the scientific community is available, or when some parameters in the data acquisition chain must be corrected (we are dealing with an experiment, after all). A solution that allows at the same time to optimize database dimension and to maintain data consistency is to store in place of computed data only the raw measurements and a complete description of the computation needed to derive a meaningful physical measurement. The computation specified in the database is then activated on the fly whenever the signal is retrieved. This concept is implemented, for example, in the ROOT data system (CERN ROOT 2023) used at CERN and in the MDSplus (MD Splus, 2023) data system used in the fusion research framework.
- Consistency of scientific data among different facilities, even when defining self-describing data for experimental results, deriving first principles for a given research domain requires accessing experimental results from different Big Science applications. This is, for example, of interest in Nuclear Fusion research where the general principles of the underlying physics are deduced from experimental results of large experiments located in the US, Europe, China, Korea, and Japan. To help researchers prepare tools for scientific investigation, a common, inter-experiment representation of data is under development at the ITER organization so that data can be used across all the fusion machines in the world. This data representation model, called

Integrated Modelling and Analysis Suite (IMAS) is built based on a common data dictionary and allows researchers all around the world to prepare simulation and analysis tools for fusion research (Romanelli et al., 2020).

2.6 Cloud and grid computing

Industrial cloud computing refers to a vast concept including a suite of infrastructures, systems, and solutions. Several definitions of cloud computing, involving a different number of layers, can be found in literature, but in essence cloud computing provides data storage, computational power and software applications through the internet on demand. This approach offers several advantages in an industrial environment, such as pooling resources from different servers and allocating them on an as-needed basis, thus reducing inefficiencies and increasing elasticity and flexibility in response to fluctuating demand. In an industrial organization cloud computing eliminates the user's responsibility for software and hardware installations and maintenance and allows a better organization of resources in the production and in the supply chain. Considering the Big Science framework, the computational resources can be divided into two broad categories: (1) resources for operating the scientific experiment, such as control and data acquisition systems, and (2) resources needed for the storage of the experimental results and offline data analysis. While resources belonging to the first category cannot clearly be exported to remote sites (the same holds considering computing resources involved in an industrial plant), distributed architectures for the management of a large amount of data and analysis computation are adopted to keep pace with the ever-growing demand of data space and computation in Big Science applications. However, rather than a cloud organization with a central management of resources, the more flexible grid organization is the best fit in this context. One of the largest grid organizations is the Worldwide LHC Computing Grid (WLCG) (Worldwide LHC, 2023) which combines about 1.4 million computer cores and 1.5 exabytes of storage from over 170 sites in 42 countries. Such an amount of computing resources would be unaffordable even for a large international scientific organization such as CERN and indeed laboratories all around the world actively contribute to sharing resources (CERN provides about 20% of the overall resources of WLCG).

Related to the management of the Computing Grid infrastructure, and more in general to any infrastructure devoted to the supervision of the experiment operation, is the need for data analytics tools that allow the collection, transformation, and organization of cloud data. We are not referring here to the management of the mainstream experiment, but rather to its support infrastructure. In this context, the requirements are the same both in Big Science and large industrial applications. For example, the requirements in the supervision of a computing center or of the building infrastructure are the same, regardless they refer to a large industrial plant or a Big Science experiment. As a consequence, solutions now extensively used in the industry such as Kafka, Hadoop, and related data management tools such as Elasticsearch and InfluxDB are also adopted in large scientific

experiments. In particular, Grafana, an open-source and Web-based data visualization and monitoring tool used to create interactive and customizable dashboards is being increasingly used in scientific experiments (Hasmani et al., 2023). The main reason for its widespread usage is the possibility of integrating customized, application-specific data sources in the Grafana framework, in addition to the large set of available interfaces for SQL, NoSQL, and time series databases.

2.7 Machine learning

Machine Learning (ML) is a rapidly growing field with potentially endless applications. Considering industrial applications, ML can be used for quality control, automation, and customization in production lines and for data analysis to help make better decisions about inventories and prices. Not surprisingly, ML applications can be found in Big Science for solving common problems, such as feature recognition in camera images, as well as more specific ones, such as deriving physical parameters from a large set of experimental data. Convolutional Neural Networks (CNN) are largely adopted in industrial automation to detect defects in products and more in general to retrieve information from images acquired by camera devices. There is a great interest in the application of CNNs also in fusion research because camera-based diagnostics, placed at a safe distance from the reactor, will be used for plasma control in the next-generation of fusion devices. Indeed any other equipment for measuring physical plasma parameters would be soon destroyed by the high temperature and the neutron flux in these reactors. Another important application of ML in Fusion experiments is the prediction of plasma disruptions (Vega et al., 2022). A plasma disruption occurs upon sudden failure of plasma control, losing plasma confinement. In this case, the large energy stored in the plasma is transferred in a very short time to the walls and the structure of the container, with big damage to the experimental facility itself. It is therefore of paramount importance to detect in advance a disruption so that defense actions such as a fast, but controlled, termination of the plasma can be achieved.

Within the large number of ML applications in Big Science, it is possible to identify two classes of applications of particular interest with potential implications in industrial applications, namely FPGA implementation and Physics Informed Neural Networks. In particle accelerators a huge data volume needs to be efficiently analyzed in real-time to reconstruct and filter nuclear events of interest, requiring FPGA processing. In many cases, the algorithms for the derivation of parameters of interest from raw detector data, previously implemented with *ad hoc* fitting procedures, are now efficiently implemented by means of ML algorithms such as CNNs. Two factors however hinder straight implementation of deep learning algorithms in very low latency (sub-microsecond) FPGA applications. The first factor is the long development time that is required to translate physics-motivated data processing into the firmware, as engineering is a scarce and valuable resource. We have already seen that this problem has been tackled at CERN in the development of the *hls4ml* tool that allows physicists to rapidly prototype ML algorithms without extensive Verilog/VHDL experience, greatly reducing the "time to physics." The second

factor is the challenge in creating an optimal FPGA implementation to balance the FPGA resources needed to achieve the latency and throughput goals of the target algorithm. The adopted techniques for this purpose are (Duarte et al., 2019):

- *Compression*, attempting to reduce the number of synapses and neurons without suffering performance loss;
- *Quantization* that can reduce the required resources for computation (sums and multiplications) retaining an acceptable loss in precision
- *Parallelization* tuning the degree of reuse of the multiplication units for a given layer computation allowing a tradeoff between FPGA resource utilization and latency in response.

In the CERN framework, these techniques are implemented in the combined usage of the high-level ML framework (e.g., TensorFlow and PyTorch) and of the proper parametrization of the hls4ml tool. These techniques can boost processing efficiency also in other domains, beyond high energy physics, from energy efficiency gains in data centers to cell screening in medical applications. The research activity in this field triggered also a collaboration between CERN and an autonomous driving software company aiming at using the techniques and software developed at CERN for deploying deep learning on FPGAs for autonomous driving. Instead of particle-physics data, the FPGAs will be used to interpret huge quantities of data generated by normal driving conditions, using readouts from car sensors to identify pedestrians and vehicles.

Physics Informed Neural Networks (PINNs) represent an advancement in the use of Neural Networks (NNs) to solve linear and nonlinear partial derivative equations. In PINNs, the loss function that is minimized at every iteration in learning takes into account also the underlying physical law governing the transfer function to be learned. PINNs are of particular interest in nuclear fusion for diagnosis and control. For example, controlling the shape and position of the plasma requires knowledge in real-time of the magnetic configuration inside and outside the plasma column. This configuration is derived by solving an inverse problem based on the Grad Shafranov equation, a nonlinear partial derivative equation describing the equilibrium in a magnetized plasma. This task requires an interactive procedure to adjust the equilibrium in order to match the experimental measurements and the computation can be speeded by orders of magnitude in time by means of PINNs (Bonotto et al., 2024). More in general PINNs are being increasingly used in several application within scientific experiments where parameters of interest are derived from raw data combining in this way the plasticity of ML algorithms with the information brought by the knowledge of the underlying physical phenomena. This concept can be extended to industrial applications where the decision-making basis of the adopted ML models may be difficult to understand being based on a black box model. PINNs based solutions have already been recently proposed in industry, such as a method for investigating crack propagation in industrial applications (Tu et al., 2023).

Finally, it is worth mentioning recent results in the usage of reinforcement learning achieved in nuclear fusion. Reinforcement learning is already adopted in the industry for control-related tasks such as self-driving cars and robot applications. A completely new approach for plasma control based on reinforcement learning has

been developed at Google DeepMind and validated for the first time in a fusion experiment (Degraeve et al., 2022).

3 Discussion

The key element in the implementation of Industry 4.0 practice is the concept of Cyber-Physical System which refers to complex engineered systems that leverage embedded computing, sensing, and network communication to monitor, coordinate control, and integrate physical devices or processes (Zhang et al., 2023). These concepts were pioneered in the nineties in several physics experiments, even before the term Big Science was invented because they are intrinsic in the nature of the application--the experiment--requiring acquiring and saving recorded measurements and interacting both with the real experiment and its data image.

The Industry 4.0 concepts gradually expanded towards smart components incorporating intelligent robotics, machine learning technology and mobile computers. Similar concepts pervaded at the same time the technology involved in more recent Big Science projects. In particular, an important concept in the Industry 4.0 movement is the Digital Twin, which is the virtual replica of processes, production lines factories, and supply chains. The Digital Twin concept is also a key concept in Big Science for even more compelling reasons. Indeed, especially for large and expensive Big Science experiments, the digital replica of the experiment, that is its digital model, is mandatory in order to reduce risk and optimize the experimental sessions. As an example, in the 20 Billion Euro ITER experiment, simulated runs of a detailed model of the experiment itself are expected to represent a routine operation to be performed before every experimental session in order to validate the experimental setup and to reduce the risk that a wrong configuration may cause dramatic events, such as plasma disruptions, that may even destroy the experimental apparatus. In other words, if in industrial applications digital twins help increase productivity, improve workflows, and minimize downtime, in several Big Science applications, a digital twin is the only viable option to safely operate experiments.

It has been shown in the previous sections that, depending on the specific field of application, common solutions have been often adopted in industry and Big Science. At the same time, peculiar aspects of Big Science applications have triggered new developments from which industry eventually benefited. For example, considering the more traditional plant systems such as vacuum and cooling, the need for SCADA solutions able to integrate a large number of components at CERN (a typical use case in large experiments) triggered the development of WinCC-OA that is now also a widespread solution in large industrial applications. Also considering real-time applications, a variety of solutions, often open source, are shared between industry and Big Science, especially in the field of distributed information processing and IIOT, such as Kafka and Grafana. More specific solutions are instead adopted when referring to real-time control of critical equipment. Even if the adopted strategies in the industry and Big Science do not basically differ, the collaborative nature of scientific institutions led to open-source solutions that are however not common in the industry. Open source frameworks such as MARTE2 for real-time control or EPICS for plant supervision have been developed and are used in research, but they may well represent interesting solutions also in industry. There is, in any case, increasing consensus towards

using open source technologies in industrial applications for a variety of reasons, among which: (1) Cost effectiveness, (2) Flexibility due to the ease of modification whereas proprietary software is often hardly customizable, (3) innovation, as open source products are constantly being developed and improved by a large community of contributors (Ebert, 2008). There are however some disadvantages that must be taken into account especially when deploying open-source solutions. For example, some open source applications may be tricky to set up and use and, even if a community is normally present on the web helping in finding solutions, this fact has to be considered when time to production is an issue.

Originally evolved in the collaborative research environment, open source, and more recently open hardware, concepts, and approaches proved successful in the industry, leading to several open standards that allow for solutions that are interoperable, modular and vendor independent, such as OPC-UA for industry automation and MQTT for IIOT applications, open Modbus/TCP for equipment supervision and control, PROFINET, a high level network for industrial applications and EtherCAT for high speed and low latency plant communication.

Among Big Science applications, CERN is, without doubt, the most active scientific organization promoting the technological transfer of open-source solutions toward industry. In the previous sections the hls4ml package for machine learning inference in FPGA and the White Rabbit project for high precision timing have been presented, but there are many other cooperation projects for technology transfer between CERN and industrial partners (CERN, 2023).

4 Conclusion

In the paper, it has been shown how Big Science and industry took advantage of synergy in several fields of application. It is true that several concepts of Industry 4.0 have been pioneered in the scientific research environment, however, several practices that evolved in industry have been later imported successfully into

Big Science. Among those, software quality addressing characteristics like reliability, usability, performance, and security represents perhaps the most important contribution of industry in the research world. Disciplined methods and best practices that are likely to result in a higher quality product contrast with the programming anarchy that used to be common in research and university. This new approach required a change in mentality in many software development teams when they moved from small pioneering experiments to Big Science applications, closing a loop in which new concepts are generated in the research environment, but good practices are taken from the industrial experience.

Author contributions

GM: Writing—original draft, Writing—review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abla, G., Heber, G., Schissel, D. P., Robinson, D., Abadie, L., Wallander, A., et al. (2014). ITERDB—the data archiving system for ITER. *Fusion Eng. Des.* 89, 536–541. doi:10.1016/j.fusengdes.2014.02.025
- Apache Kafka Portal (2024). Apache Kafka portal. Available at: <https://kafka.apache.org/> (Accessed May, 2024).
- Avon, G., Buscarino, A., Neto, A., and Sartori, F. (2021). “MARTe2 embedded signal processing unit for the ITER magnetics diagnostics,” in IECON 2021 – 47th Annual Conference of the IEEE Industrial Electronics Society, Toronto, ON, Canada, 13–16 Oct. 2021.
- Balakrishnan, K., Dhanalakshmi, R., Sinha, B. B., and Gopalakrishnan, R. (2023). Clock synchronization in industrial Internet of Things and potential works in precision time protocol: review, challenges and future directions. *Int. J. Cognitive Comput. Eng.* 4, 205–219. doi:10.1016/j.ijcce.2023.06.001
- Batista, A. J., Capellà, L., Neto, A., Hall, S., Naylor, G., Stephen, A., et al. (2017). F4E prototype of a chopper digital integrator for the ITER magnetics. *Fusion Eng. Des.* 123, 1025–1028. doi:10.1016/j.fusengdes.2017.02.024
- Blago, M. P. (2023). Deep learning particle identification in LHCb RICH. *J. Phys. Conf. Ser.* 2438, 012076. doi:10.1088/1742-6596/2438/1/012076
- Bonotto, M., Abate, D., and Pigatto, L. (2024). Reconstruction of plasma equilibrium and separatrix using convolutional physics-informed neural operator. *Fusion Eng. Des.* 200, 114193. doi:10.1016/j.fusengdes.2024.114193
- CERN (2022). CERN technological transfer report. Available at: <https://report2023-kt.web.cern.ch/> (Accessed May, 2024).
- CERN (2023). The Higgs Boson. *Dyn. Stand. Model.* 434–457. Available at: doi:10.1017/9781009291033.016
- CERN (2024). CERN hls4ml project. Available at: <https://fastmachinelearning.org/hls4ml/index.html> (Accessed May, 2024).
- CERN ROOT (2023). CERN ROOT. Available at: <https://root.cern/> (Accessed November 14, 2023).
- Control System Studio (2024). Control system Studio. Available at: <https://www.controlsystemstudio.org/> (Accessed May, 2024).
- Degrave, J., Felici, F., Buchli, J., Neunert, M., Tracey, B., Carpanese, F., et al. (2022). Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature* 602, 414–419. doi:10.1038/s41586-021-04301-9
- Delgado, R., Jo, Y. H., and Choi, B. W. (2023). RT-AIDE: a RTOS-agnostic and interoperable development environment for real-time systems. *IEEE Trans. Ind. Inf.* 19, 2772–2781. doi:10.1109/tii.2022.3182790
- De Oliveira, V., Pessoa, M., Junqueira, F., and Miyagi, P. (2021). *SQL and NoSQL databases in the context of industry 4.0 machines 10010020*.
- Duarte, J., et al. (2019). “Fast inference of deep neural networks for real-time particle,” in Physics Applications Proceedings of the 2019 ACM/SIGDA

- International Symposium on Field-Programmable Gate Arrays, USA, February-24-26, 2019.
- Ebert, C. (2008). Open source software in industry. *IEEE Softw.* 25, 52–53. doi:10.1109/ms.2008.67
- ESnet (2024). ESnet portal. Available at: <https://www.es.net/> (Accessed May, 2024).
- Experimental Physics and Industrial Control System (2023). Experimental physics and industrial control system. Available at: <https://epics.anl.gov/> (Accessed November 13, 2023).
- FPGAs that speak your language (2024). CERN courier. Available at: <https://cerncourier.com/a/fpgas-that-speak-your-language/> (Accessed May, 2024).
- Girela-López, F., López-Jiménez, J., Jiménez-López, M., Rodríguez, R., Ros, E., and Díaz, J. (2020). IEEE 1588 high accuracy default profile: applications and challenges. *IEEE Access* 8, 45211–45220. doi:10.1109/access.2020.2978337
- Guan, F., Peng, L., Perneel, L., and Timmerman, M. (2016). Open source FreeRTOS as a case study in real-time operating system evolution. *J. Syst. Softw.* 118, 19–35. doi:10.1016/j.jss.2016.04.063
- Harris, P. (2021) Scientific applications of FPGAs at the LHC, proceeding of the 2021 ACM/SIGDA international symposium on field-programmable gate arrays
- Hashmani, R., Konyushikhin, M., Shan, B., Cai, X., and Demirköz, M. (2023). New monitoring interface for the AMS experiment. *Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrom. Detect. Assoc. Equip.* 1046, 167704. doi:10.1016/j.nima.2022.167704
- IBM Engineering Systems (2024). Design Rhapsody. Available at: <https://www.ibm.com/products/systems-design-rhapsody> (Accessed May, 2024).
- Jansweijer, P. P. M., Peek, H. Z., and De Wolf, E. (2013). White Rabbit: sub-nanosecond timing over Ethernet. *Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrom. Detect. Assoc. Equip.* 725, 187–190. doi:10.1016/j.nima.2012.12.096
- Koziolek, H., Burger, A., Platenius-Mohr, M., and Jetley, R. (2020). A classification framework for automated control code generation in industrial automation. *J. Syst. Softw.* 166, 110575. doi:10.1016/j.jss.2020.110575
- Ledeul, A., Savulescu, A., Segura Millan, G., and Bartłomiej, S., 2019 Data streaming with Apache Kafka for CERN supervision,” in Control and Data Acquisition System for Radiation and Environmental Protection 17th Biennial International Conference on Accelerator and Large Experimental Physics Control Systems (ICALPECS), New York, United States, 5 - 11 Oct 2019.
- Leone, D., Carrubba, V., Mazzaro, S., Nobili, M., Cucè, V., and Hamilton, D. (2023). EPICS application for ITER RH supervisory control system. *Fus. Eng. Des.* 112429.
- LIGO The LIGO gravitational antenna (2023) Available at: <https://www.ligo.caltech.edu/page/ligo-gw-interferometer> [Accessed November 3, 2023]
- Liu, G., Makijarvi, P., and Pons, N. (2018). The ITER CODAC network design. *Fusion Eng. Des.* 130, 6–10. doi:10.1016/j.fusengdes.2018.02.072
- Matichard, F., Lantz, B., Mittleman, R., Mason, K., Kissel, J., Abbott, B., et al. (2015). Seismic isolation of Advanced LIGO: review of strategy, instrumentation and performance. *Class. Quantum Gravity* 32, 185003. doi:10.1088/0264-9381/32/18/185003
- MD Splus (2023). MDSplus home page. Available at: <https://www.mdsplus.org/index.php/Introduction> (Accessed November 14, 2023).
- Musaddiq, A., Zikria, Y. B., Hahm, O., Yu, H., Bashir, A. K., and Kim, S. W. (2018). A survey on resource management in IoT operating systems. *IEEE Access* 6, 8459–8482. doi:10.1109/access.2018.2808324
- Nowakowski, P., Rokita, P., and Graczykowski, L. (2024). GPU propagation and visualisation of particle collisions with accurate model of ALICE detector magnetic field. *Comp. Phys. Comm.* 108912.
- OPC (2024). OPC foundation. Available at: <https://opcfoundation.org/> (Accessed May, 2024).
- Perek, P., Makowski, D., Kadziela, M., Lee, W., Zagar, A., Simrock, S., et al. (2023). Evaluation of ITER Real-Time Framework in plasma diagnostics applications. *Fus. Eng. Des.* 113623.
- Pezzetti, M. (2021). Control of large helium cryogenic systems: a case study on CERN LHC. *EPJ Tech. Instrum.* 8, 6. doi:10.1140/epjti/s40485-021-00063-w
- Redis (2024). Redis portal. Available at: <https://redis.io/> (Accessed May, 2024).
- Romanelli, M., Coelho, R., Coster, D., Ferreira, J., Fleury, L., Henderson, S., et al. (2020). Code integration, data verification, and models validation using the ITER integrated modeling and analysis system (IMAS) in EUROfusion. *Fusion Sci. Technol.* 76, 894–900. doi:10.1080/15361055.2020.1819751
- SIMATIC (2023). SIMATIC WinCC open architecture portal. Available at: <https://www.wincco.com/company.html> (Accessed November 13, 2023).
- Stepanov, D., Abadie, L., Bertin, J., Bourguignon, G., Darcour, G., and Liotard, O. (2011). *Integrated approach to the development of the ITER control system configuration data. Proceedings of ICALEPCS2011*. France: Grenoble.
- The CERN LHC (2023). The CERN LHC. Available at: <https://home.cern/science/accelerators/large-hadron-collider> (Accessed November 3, 2023).
- The HDF Group (2023). The HDF group. Available at: <https://www.hdfgroup.org/solutions/hdf5/> (Accessed November 14, 2023).
- The ITER project (2023). The ITER project. Available at: <https://www.iter.org/> (Accessed November 3, 2023).
- The White Rabbit Project (2023). The white Rabbit project. Available at: <https://white-rabbit.web.cern.ch/> (Accessed November 14, 2023).
- Tu, J., Liu, C., and Qi, P. (2023). Physics-Informed neural network integrating PointNet-based adaptive refinement for investigating crack propagation in industrial applications. *IEEE Trans. Ind. Inf.* 19, 2210–2218. doi:10.1109/tii.2022.3201985
- UNICOS (2024). UNICOS framework portal. Available at: <https://unicos.web.cern.ch/> (Accessed May 9, 2024).
- Vega, J., Murari, A., Dormido-Canto, S., Rattá, G. A., Gelfusa, M., Mailloux, J., et al. (2022). Disruption prediction with artificial intelligence techniques in tokamak plasmas. *Nat. Phys.* 18, 741–750. doi:10.1038/s41567-022-01602-2
- Worldwide LHC (2023). Worldwide LHC computing grid. Available at: <https://wlcg.web.cern.ch/> (Accessed November 14, 2023).
- Zhang, K., Shi, Y., Karnouskos, S., Sauter, T., Fang, H., and Colombo, A. W. (2023). Advancements in industrial cyber-physical systems: an overview and perspectives. *IEEE Trans. Ind. Inf.* 19, 716–729. doi:10.1109/tii.2022.3199481