ANNUAL REPORT

2020

01

INTRO

Background

CERN openlab is a public-private partnership, through which CERN collaborates with leading technology companies and other research institutions on R&D projects related to scientific computing. This report provides a summary of all R&D activities carried out through this partnership in 2020.

The number of projects carried out through CERN openlab has grown rapidly in recent years: 16 in 2018, 25 in 2019, 28 in 2020. They involve teams spread across our Organization and the LHC experiments.

2020 was the final year in CERN openlab's sixth three-year phase, addressing ICT challenges related to data-centre technologies and infrastructures, computing performance and software, and machine learning and data analytics.

CERN openlab's sixth phase also saw the launch of investigations into quantum computing. These investigations have now grown into an important thread of CERN's new Quantum Technologies Initiative — launched in 2020 and led by Alberto Di Meglio, the head of CERN openlab.

Another important initiative launched in 2020 was CERN's new collaboration on high-performance computing (HPC) with SKAO, the organisation leading the development of the Square Kilometre Array radio-telescope; GÉANT, the pan-European network and services provider for research and education; and PRACE, the Partnership for Advanced Computing in Europe. This collaboration — led at CERN by Maria Girone, CERN openlab CTO — will help us to unlock the full potential of the next generation of HPC technologies.

On top of this, investigations continued into ways in which we can serve the scientific community beyond high-energy-physics. For example, the CERN openlab team contributed to the CERN against COVID-19 initiative, particularly through the BioDynaMo project, in collaboration with CERN's Knowledge Transfer Group.

In 2020, the COVID-19 pandemic also led to the cancellation of the CERN openlab Summer Student Programme. Nevertheless, the CERN openlab team was able to organise remote summer projects for 14 students. They also organised a series of new online lectures and workshops. On top of this, the team reworked the annual CERN Webfest hackathon as an online event; over 400 people signed up from 75 countries across the globe.

02

ABOUT

The concept

CERN openlab is a unique public-private partnership that accelerates the development of cutting-edge ICT solutions for the worldwide LHC community and wider scientific research. Through CERN openlab, CERN collaborates with leading ICT companies and research institutes.

Within the CERN openlab framework, CERN provides access to its advanced ICT infrastructure and its engineering experience – in some cases even extended to collaborating institutes worldwide. Testing in CERN's demanding environment provides the collaborating companies with valuable feedback on their products, while enabling CERN to assess the merits of new technologies in their early stages of development for possible future use. This framework also offers a neutral ground for carrying out advanced R&D with more than one company.

Industry collaboration can be at the associate, contributor, or partner level. Each status represents a different level of investment, with projects lasting typically between one and three years. The collaborating companies engage in a combination of cash and in-kind contributions, with the cash being used to hire young ICT specialists dedicated to the projects. The associate status formalises a collaboration based on independent and autonomous projects that do not require a presence on the CERN site. The contributor status is a collaboration based on tactical projects, which includes a contribution to hire an early-career ICT specialist supervised by CERN staff to work on the common project, in addition to the hardware and software products needed by the projects. The partners commit to a longer-term, strategic programme of work and provide three kinds of resources: funding for early-career researchers, products and services, and engineering capacity. The partners receive the full range of benefits of membership in CERN openlab, including extensive support for communications activities and access to dedicated events

CERN openlab was established in 2001, and has been organised into successive three-year phases. In the first phase (2003–2005), the focus was on the development of an advanced computing-cluster prototype called the "opencluster". The second phase (2006–2008) addressed a wider range of domains. The combined knowledge and dedication of the engineers from CERN and the collaborating companies produced exceptional results, leading to significant innovation in areas such as energy-efficient computing, grid interoperability, and network security. CERN openlab's third phase (2009-2011) capitalised and extended upon the successful work carried out in the second phase. New projects were added focusing on virtualisation of industrial-control systems and investigation of the then-emerging 64-bit computing architectures. The fourth phase (2012-2014) addressed new topics crucial to the CERN scientific programme, such as cloud computing, business analytics, next-generation hardware, and security for the ever-growing number of networked devices. The fifth phase (2015-2017) tackled ambitious challenges covering the most critical needs of ICT infrastructures in domains such as data acquisition, computing platforms, data-storage architectures, compute provisioning and management, networks and communication, and data analytics. It also saw other research institutes join CERN openlab for the first time.

Participants in the CERN openlab Technical Workshop 2020.

This annual report covers the final year of CERN openlab's sixth phase (2018-2020). The ICT challenges to be tackled in this phase were set out in a white paper (openlab.cern/whitepaper), published at the end of 2017. Challenges related to data-centre technologies and infrastructures, computing performance and software, machine learning and data analytics, and quantum technologies are being addressed.

At CERN openlab's annual technical workshops, representatives of the collaborating companies and research institutes meet with the teams, who provide in-depth updates on technical status. Collaborating companies and research institutes also elect representatives for CERN openlab's annual Collaboration Board meeting, which is an opportunity to discuss the progress made by the project teams and to exchange views on the collaboration's plans.

The CERN openlab team consists of three complementary groups of people: young engineers hired by CERN and funded by our collaborators, technical experts from the companies involved in the projects, and CERN management and technical experts working partly or fully on the joint activities.

The names of the people working on each project can be found in the results section of this report.

The members of CERN openlab's management team are shown in the right-hand column.

Alberto Di Meglio
Head of CERN openlab

Maria Girone
Chief Technology Officer

Fons Rademakers
Chief Research Officer

Sofia Vallecorsa
Quantum and AI Reasearch Lead

Kristina Gunne
Chief Finance and Administration Officer

Andrew Purcell
Chief Communications Officer

Miguel Marquina
Educational Programmes Advisor

Anastasiia Lazuka
Junior Communications Officer

Christina Bolanou
Junior Communications Officer

ORACLE

intel

SIEMENS

Micron

Google

beyys research

IBM

E4 COMPUTER ENGINEERING

CQC Cambridge Quantum Computing

360 Comtrade Three-Sixty

open systems

INFN Istituto Nazionale di Fisica Nucleare

Scimpulse foundation

KING'S College LONDON

ODTÜ METU

gluoNNet knowledge exchange for smart decisions

TU/e Technische Universiteit Eindhoven University of Technology

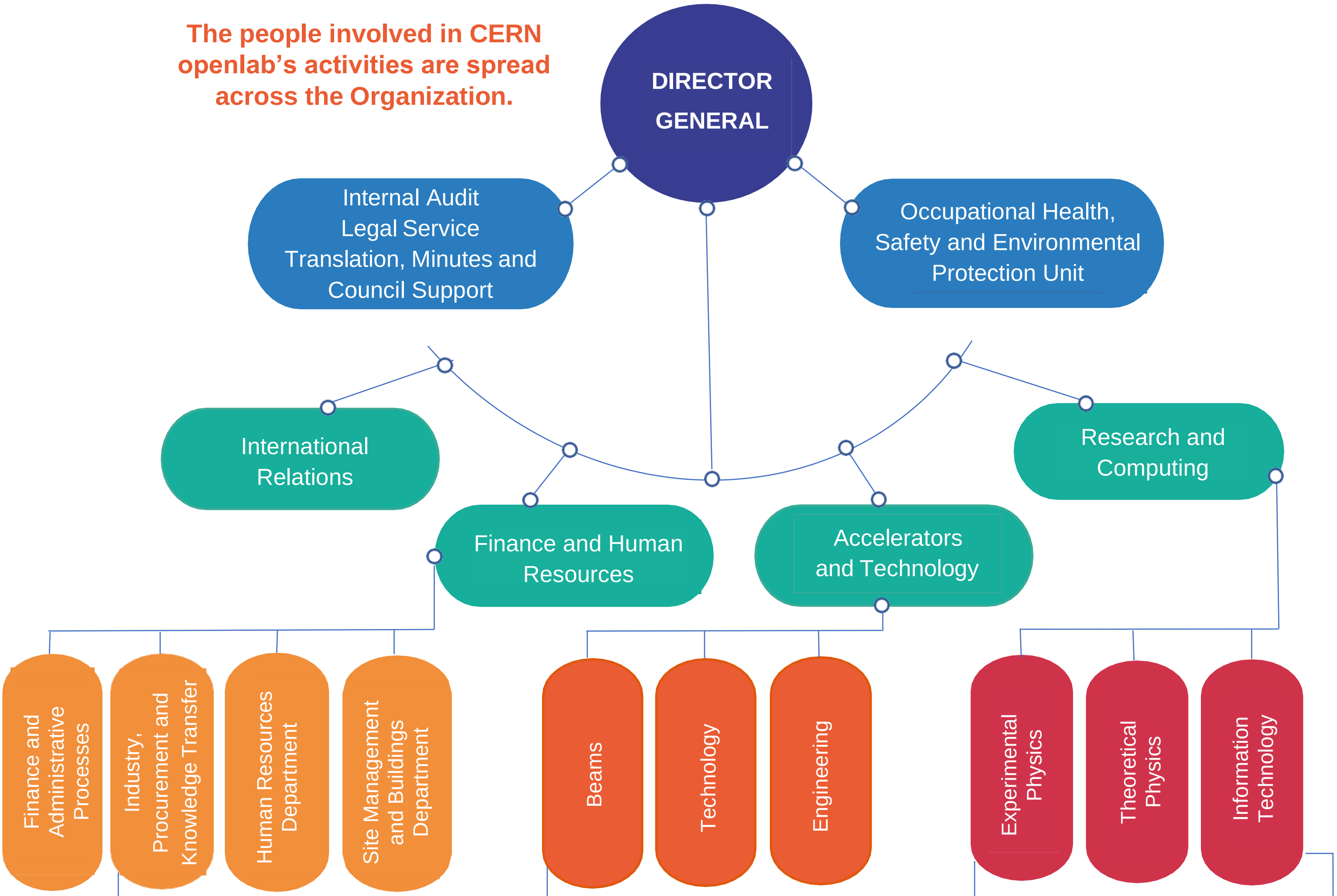
SAMARA UNIVERSITY

Fermilab

Newcastle University

POSITIONING CERN OPENLAB'S ACTIVITIES AT CERN

The people involved in CERN openlab's activities are spread across the Organization.

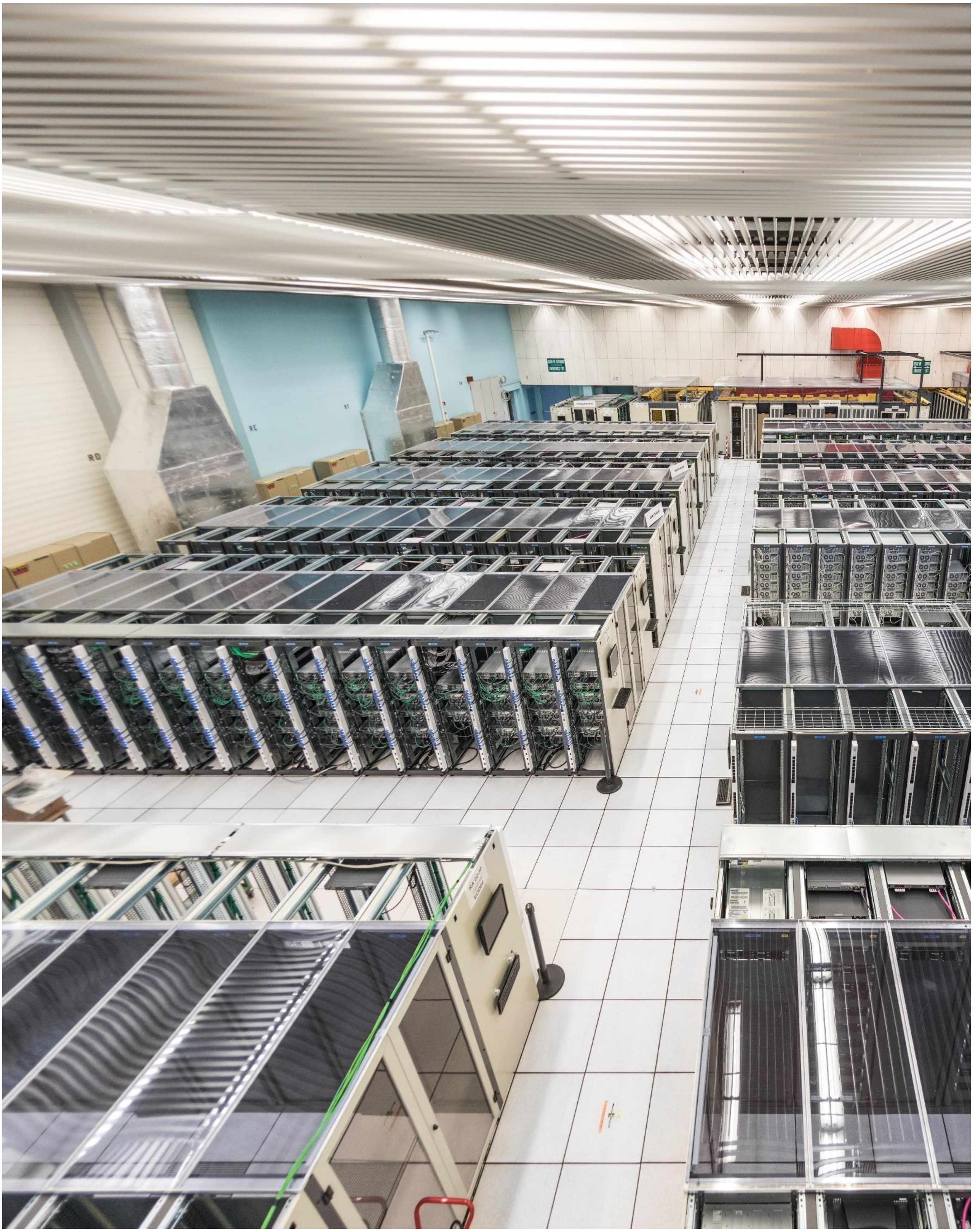


- Procurement & Industrial Services
- **Knowledge Transfer**
- European Union Support
- Office of the Department Head

- **Accelerators and Beams Physics**
- Administration, Safety & Resources
- Beam Instrumentation
- Controls
- Industrial Controls & Safety Systems
- Operation
- Radio Frequency

- Administration and General Services
- **ALICE Experiment**
- **ATLAS Experiment**
- **CMS Experiment**
- Office of the Department Leader
- **Detector Technologies**
- Electronic Systems for Experiments
- Linear Collider Detector
- **LHCb Experiment**
- **Neutrino**
- **Software Design for Experiments**
- Small and Medium-Sized Experiments

- Collaboration, Devices & Applications
- **Compute & Monitoring**
- **Databases**
- **Storage**
- **Computing Facilities**
- **Communications Systems**
- **Departmental Infrastructure**



03

RESULTS

Our projects

Information about each of the 28 technical projects that ran in 2020 can be found in this section. These projects are organised into five overarching R&D topics.

R&D topic 1:
Data-centre technologies and infrastructures

Designing and operating distributed data infrastructures and computing centres poses challenges in areas such as networking, architecture, storage, databases, and cloud. These challenges are amplified and added to when operating at the extremely large scales required by major scientific endeavours.

CERN is evaluating different models for increasing computing and data-storage capacity, in order to accommodate the growing needs of the LHC experiments over the next decade. All models present different technological challenges. In addition to increasing the capacity of the systems used for traditional types of data processing and storage, explorations are being carried out into a number of alternative architectures and specialised capabilities. These will add heterogeneity and flexibility to the data centres, and should enable advances in resource optimisation.

R&D topic 2:
Computing performance and software

Modernising code plays a vital role in preparing for future upgrades to the LHC and the experiments. It is essential that software performance is continually increased by making use of modern coding techniques and tools, such as software-optimising compilers. It is also important to ensure that software fully exploits the features offered by modern hardware architectures, such as many-core GPU platforms, acceleration coprocessors, and hybrid combinations of CPUs and FPGAs. At the same time, it is paramount that physics performance is not compromised in drives to maximise efficiency.

R&D topic 3:
Machine learning and data analytics

Members of CERN's research community expend significant efforts to understand how they can get the most value out of the data produced by the LHC experiments. They seek to maximise the potential for discovery and employ new techniques to help ensure that nothing is missed. At the same time, it is important to optimise resource usage (tape, disk, and CPU), both in the online and offline environments. Modern machine-learning technologies — in particular, deep-learning solutions applied to raw data — offer a promising research path to achieving these goals.

Deep-learning techniques offer the LHC experiments the potential to improve performance in each of the following areas: particle detection, identification of interesting collision events, modelling detector response in simulations, monitoring experimental apparatus during data taking, and managing computing resources.

R&D topic 4:
Quantum technologies

The HL-LHC is likely to require computing and storage capacity 2-3 times greater than today. Even corrected for the evolution of technology, this will result in a shortage of computing resources. Keeping with its mandate, CERN openlab is exploring new and innovative solutions to help physicists to bridge this resource gap which may put in danger the achievement of the goals of the HL-LHC experimental programme.

Following a successful workshop on quantum computing held at CERN in 2018, CERN openlab has started a number of projects in quantum computing that are at different stages of realisation.

R&D topic 5:
Applications in other disciplines

The fifth R&D topic is different to the others in this report, as it focuses on applications in other disciplines. By working with communities beyond high-energy physics, we are able to ensure maximum relevancy for CERN openlab's work, as well as learning and sharing both tools and best practices across scientific fields. Today, more and more research fields are driven by large quantities of data, and thus experience ICT challenges comparable to those at CERN.

CERN openlab's mission rests on three pillars: technological investigation, education, and dissemination. Collaborating with research communities and laboratories outside high-energy physics brings together all these aspects.

HYBRID DISASTER-RECOVERY SOLUTION USING PUBLIC CLOUD

R&D TOPIC 1:
DATA-CENTRE TECHNOLOGIES AND INFRASTRUCTURES

Project coordinator:
Viktor Kozlovzky

Technical team:
Aimilios Tsouvelekakis, Alina Andreea Grigore, Andrei Dumitru, Antonio Nappi, Arash Khodabandeh, Artur Wiecek, Borja Aparicio Cotarelo, Edoardo Martelli, Ignacio Coterillo Coz, Sebastian Lopienski

Collaborator liaisons:
Cris Pedregal, Alexandre Reigada, David Ebert, Vincent Leocorbo, Dimitry Dolgušin

Project goal

In 2020, the Database Services group in the CERN IT department launched this project, in collaboration with Oracle, to explore how integration of commercial cloud platforms with CERN on-premises systems might improve the resilience of services. The aim of this project is to understand the benefits, limitations and cost of the cloud environment for application servers and databases.

Background

Today, high availability is a key requirement for most platforms or services. The Database Services group maintains critical services for CERN that are used daily by the majority of users. This project enables us to review the current system, assess its scalability and explore further capabilities by integrating them with new technologies.

Progress in 2020

For the integration exercise, the group is using the Oracle Cloud Infrastructure (OCI).

The project team successfully replicated its process for creating virtual machines on OCI. This involved registering public-cloud virtual machines on the CERN main network, as well as integrating them with CERN's central configuration management system.

The ability to deploy machines on OCI within the CERN network enabled us to use OCI running Oracle databases like on-premises Oracle databases. Moreover, we automated the procedure for creating standby databases for our on-premises primary databases, and we configured a Data Guard broker for data synchronisation between them. We performed tests with different data sets to evaluate the performance on the network and storage sides.

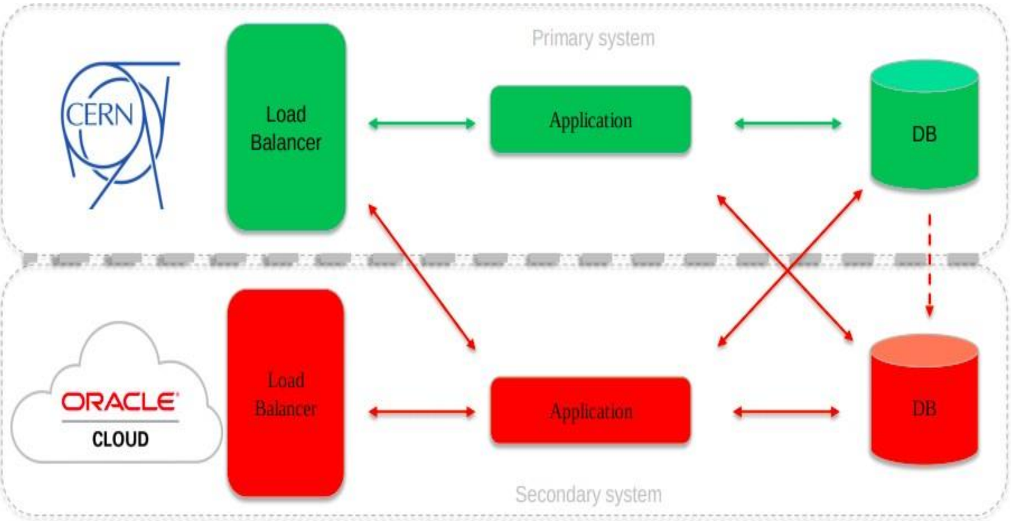
The team maintaining Kubernetes applications investigated the container engine provided by Oracle for the Kubernetes (OKE) platform. We extended the application deployment process, which is now capable of deploying applications to the OCI running OKE platform.

Furthermore, using an Oracle REST Data Services (ORDS) application, we performed a complex integration test; for forwarding network traffic, we used a proxy server. The tests showed vulnerabilities and integration limitations for the applications.

Next steps

We will continue to evaluate the performance of storage and network for databases and see what the best fits would be for our use cases. We will also investigate the cost of running the databases on OCI compared to running them in CERN's data centre.

The members of the project would like to thank the support teams for Oracle Cloud Infrastructure and Oracle Terraform for their valuable assistance.



The graphic above provides an overview of the Database Services group's hybrid disaster-recovery project in 2020.

ORACLE WEBLOGIC ON KUBERNETES

R&D TOPIC 1:
DATA-CENTRE TECHNOLOGIES AND INFRASTRUCTURES

Project coordinator:
Antonio Nappi

Technical team:
Lukas Gedvilas

Collaborator liaisons:
Monica Riccelli, Will Lyons, Maciej Gruzka, Cris Pedregal, David Ebert, Dimitry Dolgušin

Project goal

The aim of this project is to improve the deployment of Oracle WebLogic infrastructure in large-scale deployments, profiting from new technologies such as Kubernetes and Docker Containers. These technologies will help to make the infrastructure deployment process portable, repeatable, and faster, enabling CERN service managers to be more efficient in their daily work.

Background

The Oracle WebLogic service has been active at CERN for many years, offering a very stable way to run applications core to the laboratory. However, we would like to reduce the amount of time we spend on maintenance tasks or creating new environments for our users. Therefore, we started to explore solutions that could help us to improve how we deploy Oracle WebLogic. Kubernetes has now made our deployment much faster, reducing the time spent on operational tasks and enabling us to focus more on developers' needs.

Progress in 2020

In 2020, we focused on the structure of monitoring and logging systems; these dramatically improved the way we work.

Using Prometheus for the monitoring part, we were able to get much more information on both the infrastructure itself and the application layer. We can now easily see resource use level in Kubernetes, as well as how containers are behaving. Thanks to this, it will be much easier to build an efficient alert system.

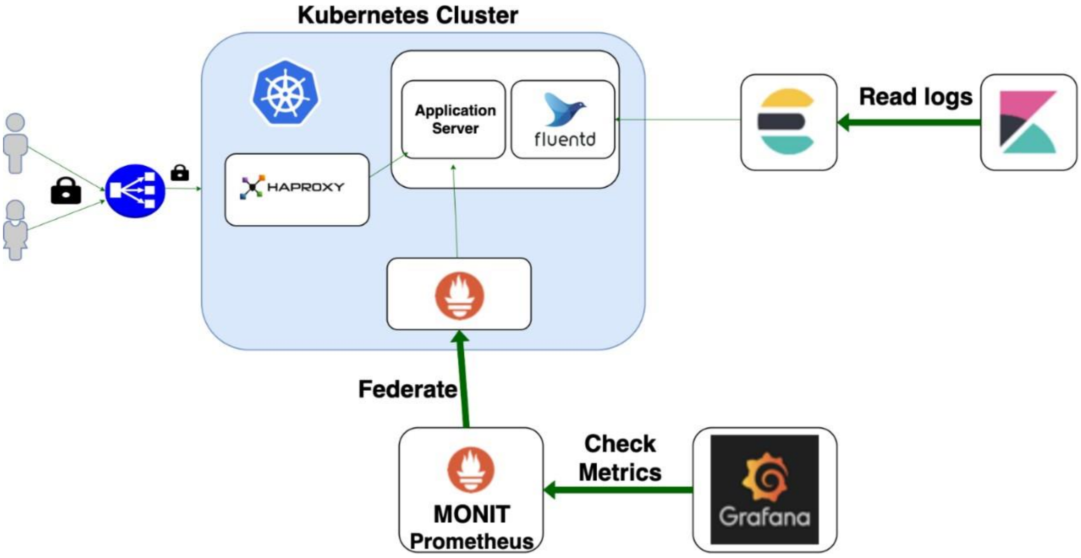
We decided to introduce Fluentd as our logging component. This helped us to reduce the number of hosts writing to elasticsearch and to standardise the logs produced by our system. Meanwhile, we also managed to migrate more applications to Kubernetes: 70-80% of our current production is now running on K8s.

Next steps

We aim to complete the migration of production applications by mid-2021. We will also focus on the integration of Prometheus into our new alert system. In addition, we would also like to start investigating ArgoCD and Flux2 as solutions for making application and infrastructure deployments more oriented towards GitOps.

Presentations

- W. Coekaerts, A. Nappi, Cloud Platform and Middleware Strategy and Roadmap (14-15 January). Presented at the Oracle OpenWorld Middle West, Dubai, 2020. *http://cern.ch/go/Qd8c*
- M. Gruszka, W. Lyons, A. Nappi, Deploying Oracle WebLogic Server on Kubernetes and Oracle Cloud (14-15 January). Presented at the Oracle OpenWorld Middle West, Dubai, 2020. *http://cern.ch/go/q7jk*
- W. Coekaerts, M. McMaster, R. Hussain, A. Nappi, Cloud Platform and Middleware Strategy and Roadmap (12-13 February). Presented at the Oracle Openworld, London, 2020. *http://cern.ch/go/9sQP*
- M. Gruszka, W. Lyons, A. Nappi, Deploying Oracle WebLogic Server on Kubernetes and Oracle Cloud (12-13 February). Presented at the Oracle Openworld, London, 2020. *http://cern.ch/go/69Dc*



Oracle WebLogic on Kubernetes architecture.

HETEROGENEOUS I/O FOR SCALE

R&D TOPIC 1:
DATA-CENTRE TECHNOLOGIES AND INFRASTRUCTURES

Project coordinator:
Maria Girone, Viktor Khristenko

Collaborator liaisons:
Ulrich Bruening (University of Heidelberg), Mondrian Nuessle (Extoll GmbH)

Project goal

We are working to develop a proof of concept for an FPGA-based I/O intermediary. The potential impact would be to change the way data ingestion happens when using remote storage locations. In view of the enormous amounts of data to be employed in the future for data analytics, it is crucial to efficiently manage the flow in order to harness the computational power provided by high-performance computing (HPC) facilities.

Background

One of the common aspects of all data-intensive applications is the streaming of recorded data from remote storage locations. This often imposes constraints on the network and forces a compute node to introduce complex logic to perform aggressive caching in order to remove latency. Moreover, this substantially increases the memory footprint of the running application on a compute node. This project, abbreviated to 'HIOS', aims to provide a scalable solution for such data-intensive workloads by introducing heterogeneous I/O units directly on the compute clusters. This makes it possible to offload the aggressive caching functionality onto these heterogeneous units. By removing this complicated logic from compute nodes, the memory footprint decreases for data-intensive applications. Furthermore, the project will investigate the possibility of including additional logic, coding/decoding, serialisation, I/O specifics, directly onto such units.

An integral part of the project will be the ability to integrate the units developed directly with current HPC facilities. One of the main outcomes of the project will be the reduced time required to extract insights from large quantities of acquired information, which, in turn, directly impacts society and scientific discoveries.

HIOS is one of the 170 breakthrough projects receiving funding through the ATTRACT initiative. ATTRACT, which is part of the European Union's Horizon 2020 programme, is financing breakthrough ideas in the fields of detection and imaging.

Progress in 2020

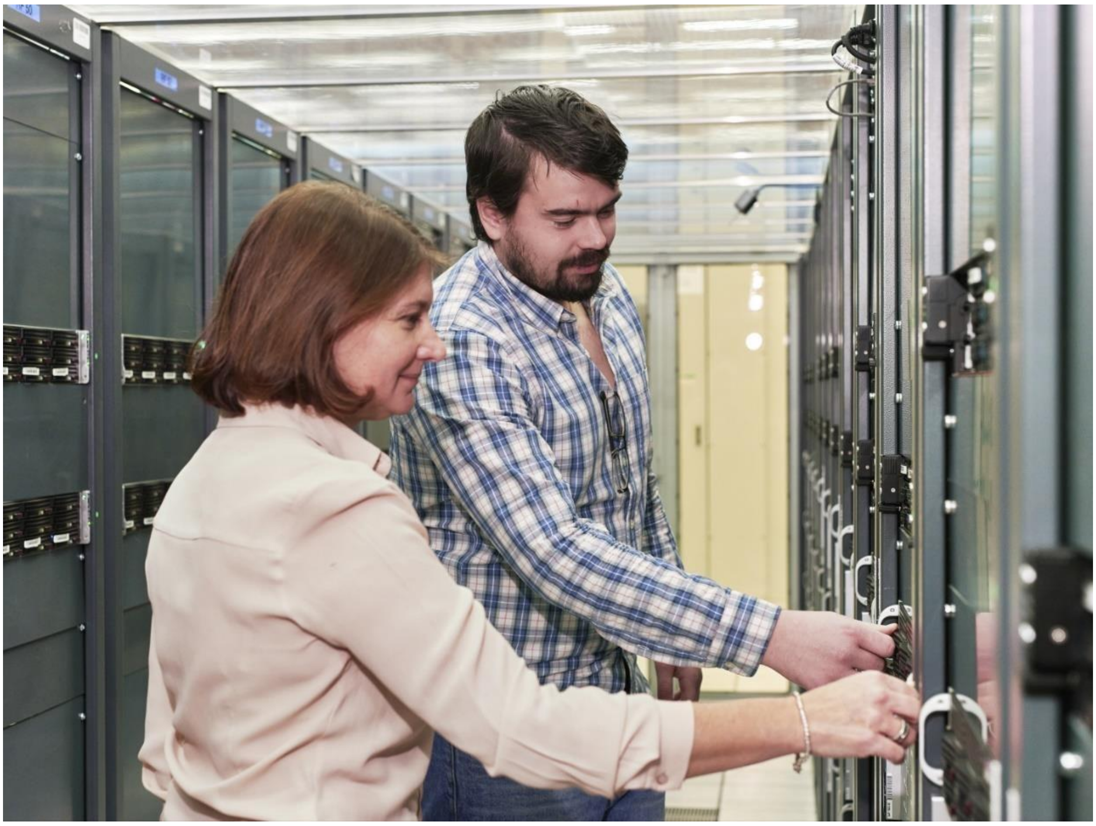
During the past year, the first phase of the ATTRACT project, we performed testing and validation of the TCP/IP network stack. This targeted an FPGA and could be used to integrate with Extoll's existing solution. At the same time, a range of compression/decompression algorithms (e.g. LZMA, LZ4, zstd, Deflate) were explored in order to incorporate not only network-related functionality, but also data-processing primitives.

Next steps

The first phase of the project was concluded at the end of 2020. Discussions are currently underway regarding how to take this work forwards.

Publications

- V. Khristenko, M. Girone, M. Nuessle, D Frey, U. Bruening, J. Schonbohm, HIOS: Heterogenous I/O for Scale. Published on ATTRACT, 2020. http://cern.ch/go/qH7N



Maria Girone and Viktor Khristenko, CERN members of the project team, pictured in the CERN data centre.

DYNAMICAL EXASCALE ENTRY PLATFORM – EXTREME SCALE TECHNOLOGIES (DEEP-EST)

R&D TOPIC 1:
DATA-CENTRE TECHNOLOGIES AND INFRASTRUCTURES

DEEP-EST collaboration

Project coordinator:
Maria Girone (for DEEP-EST task 1.7)

Technical team:
Viktor Khristenko (for DEEP-EST task 1.7)

Project goal

The main focus of the project is to build a new kind of a system that makes it possible to run efficiently a wide range of applications — with differing requirements — on new types of high-performance computing (HPC) resources. From machine-learning applications to traditional HPC applications, the goal is to build an environment which is capable of accommodating workloads that pose completely different challenges for the system.

Background

DEEP-EST is an EC-funded project that launched in 2017, following on from the successful DEEP and DEEP-ER projects. The project involves 27 partners in more than 10 countries and is coordinated from Jülich Supercomputing Centre at Forschungszentrum Jülich in Germany.

Overall, the goal is to create a modular supercomputer that best fits the requirements of diverse, increasingly complex, and newly emerging applications. The innovative modular supercomputer architecture creates a unique HPC system by coupling various compute modules according to the building-block principle: each module is tailored to the needs of a specific group of applications, with all modules together behaving as a single machine.

Specifically, the prototype consists of three compute modules: the cluster module (CM), the extreme scale booster (ESB), and the data-analytics module (DAM).

Applications requiring high single-thread

performance are targeted to run on the CM nodes, where Intel® Xeon® Scalable Processor (Skylake) provide general-purpose performance and energy efficiency.
The architecture of ESB nodes was tailored for highly scalable HPC software stacks capable of extracting the enormous parallelism provided by Nvidia V100 GPUs.
Flexibility, large memory capacities using Intel® Optane™ DC persistent memory and different acceleration capabilities (provided by Intel Stratix 10 and Nvidia V100 GPU on each node) are key features of the DAM; they make it an ideal platform for data-intensive and machine-learning applications.

CERN, in particular the CMS Experiment, participates by providing one of the applications that are used to evaluate this new supercomputing architecture.

Progress in 2020

During the past year, the focus was on testing the functionality developed during the previous year using the full DEEP-EST prototype system and optimising for the needs of the CMS Experiment. In particular, the tests we ran included more than 1000 cores and just under 100 Nvidia GPUs simultaneously, close to 100% of the available computing resources for our prototype. Furthermore, tests included not only traditional CMS Experiment reconstruction, but also a complex deep-learning pipeline. The workflow was used to benchmark the usage of Nvidia GPUs with MPI (over InfiniBand) for the purpose of distributed training of a model based on Graph Neural Networks.

Next steps

The work is ongoing to integrate the usage of Message Passing Interface (MPI) to add the ability to use not only GPUs available on the node that is performing data processing, but also on the remotely connected nodes. This will allow more efficient utilisation of heterogeneous resources, providing an overall speed up (i.e. higher throughput) of the CMS reconstruction application. The DEEP-EST project came to a successful conclusion during the first quarter of 2021.

Publications

- The following deliverable was submitted to the European Commission: Deliverable 1.5: Final report on application experience.

Presentations

- V. Khristenko, Exploiting Modular HPC in the context of DEEP-EST and ATTRACT projects (22 January). Presented at the CERN openlab Technical Workshop, Geneva, 2020. *cern.ch/go/rjC7*

The DEEP-EST prototype.

EOS PRODUCTISATION

R&D TOPIC 1:
DATA-CENTRE TECHNOLOGIES AND INFRASTRUCTURES

COMTRADE

Project coordinator:
Luca Mascetti

Technical team:
Fabio Luchetti, Elvin Sindrilaru

Collaborator liaisons:
Gregor Molan, Branko Blagojević, Ivan Arizanović, Svetlana Milenković

Project goal

This project is focused on the evolution of CERN's EOS large-scale storage system. The goal is to simplify the usage, installation, and maintenance of the system. In addition, the project aims to add native support for new client platforms, expand documentation, and implement new features/integration with other software packages.

Background

Within the CERN IT department, a dedicated group is responsible for the operation and development of the storage infrastructure. This infrastructure is used to store the physics data generated by the experiments at CERN, as well as the files of all members of personnel.

EOS is a disk-based, low-latency storage service developed at CERN. It is tailored to handle large data rates from the experiments, while also running concurrent complex production workloads. This high- performance system now provides more than 350 petabytes of raw disks.

EOS is also the key storage component behind CERNBox, CERN's cloud-storage service. This makes it possible to sync and share files on all major mobile and desktop platforms (Linux, Windows,
macOS, Android, iOS), with the aim of providing offline availability to any data stored in the EOS infrastructure.

Progress in 2020

In 2020, Comtrade's team continued to improve and update the EOS technical documentation. Due to the COVID pandemic and the travel restrictions during the year, the hosting of dedicated hardware resources at CERN, to support the prototyping of an

EOS-based appliance, was postponed.

The EOS-Comtrade team therefore focused their efforts on the development of a native way to interact from the Windows environment to the distributed storage system. Currently, Windows users access EOS only via dedicated gateways, where the system is exposed with SAMBA shares.

In order to provide more performant and seamless access to EOS for Windows users, Comtrade started developing a dedicated Windows client, codenamed EOS-wnc.

This new, dedicated, command-line Windows client was developed against the EOS server APIs. It now covers all EOS commands from the standard user listing and copy commands to the administrative ones.
On top of this, some EOS commands on Windows have additional functionalities, like improved tab-completion and history features.

Next steps

Together, we will focus on improving the current EOS Windows native client, adding new functionalities, and integrating it in the Windows user interface, developing a dedicated driver.

It is still planned to host dedicated hardware resources at CERN to support prototyping of an EOS-based appliance. This will enable Comtrade to create a first version of a full storage solution and to offer it to potential customers in the future.

Ivan Arizanović, Luca Mascetti and Branko Blagojević pictured at the CERN Data Centre.

Presentations

- L. Mascetti, EOS Comtrade Project (23 January). Presented at CERN openlab Technical workshop, Geneva, 2020. http://cern.ch/go/l9gc
- G. Molan, Preparing EOS for Enterprise Users (27 January). Presented at Cloud Storage Services for Synchronization and Sharing, Copenhagen, 2020. http://cern.ch/go/tQ7d
- L. Mascetti, CERN Disk Storage Services (3 February). Presented at CERN EOS workshop, Geneva, 2020. http://cern.ch/go/pF97
- G. Molan, EOS Documentation for Enterprise Users (3 February). Presented at CERN EOS workshop, Geneva, 2020. http://cern.ch/go/swX8
- G. Molan, EOS Windows Native Client (3 February). Presented at CERN EOS workshop, Geneva, 2020. http://cern.ch/go/P7DX
- G. Molan, EOS Storage Appliance Prototype (5 February). Presented at CERN EOS workshop, Geneva, 2020. http://cern.ch/go/q8gh
- G.Molan, EOS-wnc demo (30 October 2020). Presented at IT-ST-PDSComtrade workshop, Geneva, 2020.

KUBERNETES AND GOOGLE CLOUD

R&D TOPIC 1:
DATA-CENTRE TECHNOLOGIES AND INFRASTRUCTURES

Project coordinator:
Ricardo Manuel Brito da Rocha

Collaborator liaisons:
Grazia Frontoso, Karan Bhatia, Kevin Kissell

Project goal

The aim of this project is to demonstrate the scalability and performance of Kubernetes and Google Cloud, validating this set-up for future computing models. The focus is on taking both existing and new high-energy physics (HEP) use cases and exploring the best suited and most cost-effective set-up for each of them.

Background

Looking towards the next-generation tools and infrastructure that will serve HEP use cases, exploring external cloud resources opens up a wide range of new possibilities for improved workload performance. It can also help us to improve efficiency in a cost-effective way.

The project relies on well-established APIs and tools supported by most public cloud providers – particularly Kubernetes and other Cloud Native Computing Foundation (CNCF) projects in its ecosystem – to expand available on-premises resources to Google Cloud. Special focus is put on use cases with spiky usage patterns, as well as those that can benefit from scaling out to large numbers of GPUs and other dedicated accelerators like TPUs.

Both traditional physics analysis and newer computing models based on machine learning are considered by this project.

Progress in 2020

The year started with consolidation of the work from the previous year. New runs of more traditional physics analysis helped with the validation of Google Cloud as a viable and performant solution for scaling our HEP workloads. A major milestone for this work was the publication of a CERN story on the Google Cloud Platform website (see publications).

The main focus in 2020 though was on evaluating next-generation workloads at scale. We targeted machine learning in particular, as this places significant requirements on GPUs and other types of accelerators, like TPUs.

In addition to demonstrating that HEP machine-learning workloads can scale out linearly to hundreds of GPUs in parallel, we also demonstrated that public cloud resources have the potential to offer HEP users a way to speed up their workloads by at least an order of magnitude in a cost-effective way.

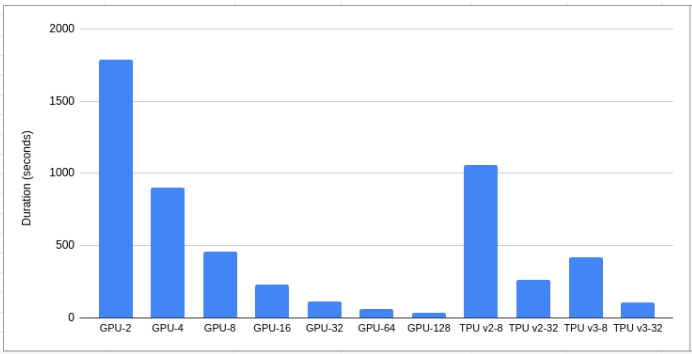
Next steps

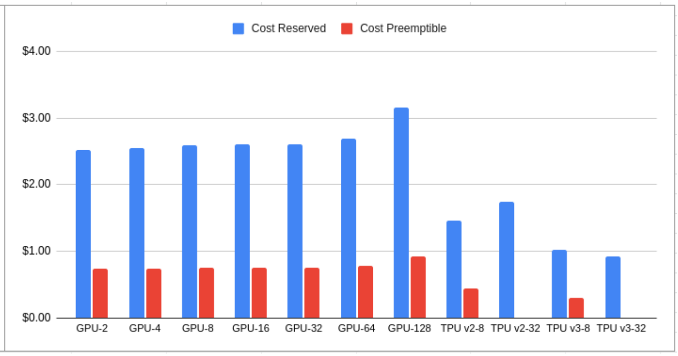
Work is now continuing in the following areas:

- Further expanding the existing setup to burst on-premises services to public cloud resources, building on cloud-native technologies.
- Onboarding more use cases where on-demand resource expansion is beneficial, with a special focus on workloads requiring accelerators (such as those for machine learning).
- Expanding the cost analysis for each of these workloads and improving the feedback to end users on the efficiency of their workloads.

Publications

- R. Rocha, Helping researchers at CERN to analyse powerful data and uncover the secrets of our universe. Published at the Google Cloud Platform CERN, 2020. http://cern.ch/go/Q7Tn





Results from one of the sample workloads taken for evaluation on Google Cloud Platform. Specifically, it shows the distributed training of a 3DGAN for fast simulation. The left-hand graph shows the raw results with up to 128 parallel GPUs (achieving linear scaling), as well as with TPUs (up to 32 cores). The right-hand graph shows the overall cost of each run. The GPU runs show a flat overall cost, even with the larger number of GPUs providing a speedup of over 52x. The TPU runs show that this type of resource is not only efficient, but it is also the most cost effective for this particular workload.

HIGH-PERFORMANCE DISTRIBUTED CACHING TECHNOLOGIES (DAQDB)

R&D TOPIC 2:
COMPUTING PERFORMANCE AND SOFTWARE

Project coordinator:
Giovanna Lehmann Miotto

Technical team:
Adam Abed Abud, Fabrice Le Goff, Lola Stankovic

Collaborator liaisons:
Claudio Bellini, Grzegorz Jereczek, Jan Lisowiec, Andrea Luiselli, Maciej Maciejewski, Jakub Radtke, Malgorzata Szychowska, Norbert Szulc, Johann Lombardi, Andrea Luiselli

Project goal

We are exploring the suitability of a new infrastructure for key-value storage in the data-acquisition systems of particle-physics experiments. DAQDB (Data Acquisition Database) is a scalable and distributed key-value store that provides low-latency queries. It exploits Intel® Optane™ DC persistent memory, a cutting-edge non-volatile memory technology that could make it possible to decouple real-time data acquisition from asynchronous event selection.

Background

Upgrades to the LHC mean that the data rates coming from the detectors will dramatically increase. Data will need to be buffered while waiting for systems to select interesting collision events for analysis. However, the current buffers at the readout nodes can only store a few seconds of data due to capacity constraints and the high cost of DRAM. It is therefore important to explore new, cost-effective solutions — capable of handling large amounts of data — that capitalise on emerging technologies.

Progress in 2020

The idea of DAQDB has proven to be very interesting. Nevertheless, the effort required to develop a mature product of this complexity was recognised to be too large for the available resources. Work thus continued in two parallel strands:

1) An evaluation of Distributed Asynchronous Object Storage (DAOS) was started, with the aim of assessing whether the future DAQ

needs of large experiments, such as ATLAS and CMS, may be addressed through this underlying technology, combined with a custom software layer.

2) Performance measurements of Intel® Optane™ DC persistent memory devices continued, in the context of data-acquisition systems, achieving very promising results.

Next steps

The DAQDB project concluded in 2020. A possible follow-up activity is being discussed between Intel and experiment representatives; this would focus on the further development and use of DAOS.

Team members Adam and Lola stood in front of the rack hosting the servers used for benchmarking DAQDB and Intel® Optane™ persistent memory.

Publications

- P. Czarnul, G. Gołaszewski, G. Jereczek, M. Maciejewski, Development and benchmarking a parallel Data AcQuisition framework using MPI with hash and hash+tree structures in a cluster environment. Published at the 19th International Symposium on Parallel and Distributed Computing, 2020. http://cern.ch/go/lK78
- A. A. Abud, D. Cicalese, G. Jereczek, F. L. Goff, G. L. Miotto, J. Love, M. Maciejewski, R. K. Mommsen, J. Radtke, J. Schmiegel, M. Szychowska, Let's get our hands dirty: a comprehensive evaluation of DAQDB, key-value store for petascale hot storage. Published at the 24th International Conference on Computing in High Energy and Nuclear Physics, 2020. http://cern.ch/go/7JGF

Presentations

- P. Czarnul, G. Gołaszewski, G. Jereczek, M. Maciejewski, Development and benchmarking a parallel Data AcQuisition framework using MPI with hash and hash+tree structures in a cluster environment (5-8 July). Presented at the 19th International Symposium on Parallel and Distributed Computing, Warsaw, 2020. http://cern.ch/go/W9zL
- A. A. Abud, Experience and performance of persistent memory for the DUNE data acquisition system (12-24 October). Presented at the 22nd virtual IEEE Realtime Conference, 2020. http://cern.ch/go/P8sD

TESTBED FOR GPU-ACCELERATED APPLICATIONS

R&D TOPIC 2:
COMPUTING PERFORMANCE AND SOFTWARE

Project coordinator: *Maria Girone*

Technical coordinators:
Andrea Bocci, Felice Pantaleo, Maurizio Pierini, Federico Carminati, Vincenzo Innocente, Marco Rovere, Jean-Roch Vlimant, Vladimir Gligorov, Daniel Campora, Riccardo De Maria, Adrian Oeftiger, Lotta Mether, Ian Fisk, Lorenzo Moneta, Jan Kieseler, and Sofia Vallecorsa.

Technical team:
Mary Touranakou, Thong Nguyen, Javier Duarte, Olmo Cerri, Roel Aaij, Dorothea Vom Bruch, Blaise Raheem Delaney, Ifan Williams, Niko Neufeld, Viktor Khristenko, Florian Reiss, Guillermo Izquierdo Moreno, Luca Atzori, Miguel Fontes Medeiros

Collaborator liaisons:
Cosimo Gianfreda (E4), Daniele Gregori (E4), Agnese Reina (E4), Piero Altoé (NVIDIA), Andreas Hehn (NVIDIA), Tom Gibbs (NVIDIA).

Project goal
The goal of this project is to adapt computing models and software to exploit fully the potential of GPUs. The project, which began in late 2018, consists of ten individual use cases.

Background
Heterogeneous computing architectures will play an important role in helping CERN address the computing demands of the HL-LHC.

Progress in 2020
This CERN openlab project supports several use cases at CERN. This section outlines the progress made in the two main use cases that were worked on in 2020.

Allen: a high-level trigger on GPUs for LHCb
Allen' is an initiative to develop a complete high-level trigger (the first step of the data-filtering process following particle collisions) on GPUs for the LHCb experiment. It has benefitted from support through CERN openlab, including consultation from engineers at NVIDIA.

The new system processes 40 Tb/s, using around 350 of the latest generation NVIDIA GPU cards. Allen matches — from a physics point of view — the reconstruction performance for charged particles achieved on traditional CPUs. It has also been shown that Allen will not be I/O or memory limited. Plus, not only can it be used to perform reconstruction, but it can also take decisions about whether to keep or reject events.

A diverse range of algorithms have been implemented efficiently on Allen. This demonstrates the potential for GPUs not only to be used as accelerators, but also as complete and standalone data-processing solutions.

In May 2020, Allen was adopted by the LHCb collaboration as the new baseline first-level trigger for Run 3. The Technical Design Report for the system was approved in June. From the start, Allen has been designed to be a framework that can be used in a general manner for high-throughput GPU computing. A workshop was held with core Gaudi developers and members of the CMS and ALICE experiments to discuss how best to integrate Allen into the wider software ecosystem beyond the LHCb experiment. The LHCb team working on Allen is currently focusing on commissioning the system for data-taking in 2022 (delayed from the original 2021 start date due to the COVID-19 pandemic). A readiness review is taking place in the first half of 2021.

End-to-end multi-particle reconstruction for the HGCal based on machine learning. The CMS High-Granularity Calorimeter (HGCal) will replace the end-cap calorimeters of the CMS detector for the operation of the High-Luminosity LHC. With about 2 million sensors and high lateral and transversal granularity, it provides huge potential for new physics discoveries. We aim to exploit this using end-to-end optimisable graph neural networks.

Profiting from new machine-learning concepts developed within the group and the CERN openlab collaboration with the Flatiron Institute in New York, US, we were able to develop and train a first prototype for directly reconstructing incident particle properties from raw detector hits. Through our direct contact with NVIDIA, we were able to implement custom tensorflow GPU kernels. Together with the dedicated neural network structure, these enabled us to process the hits of an entire particle-collision event in one go on the GPU.

Next steps

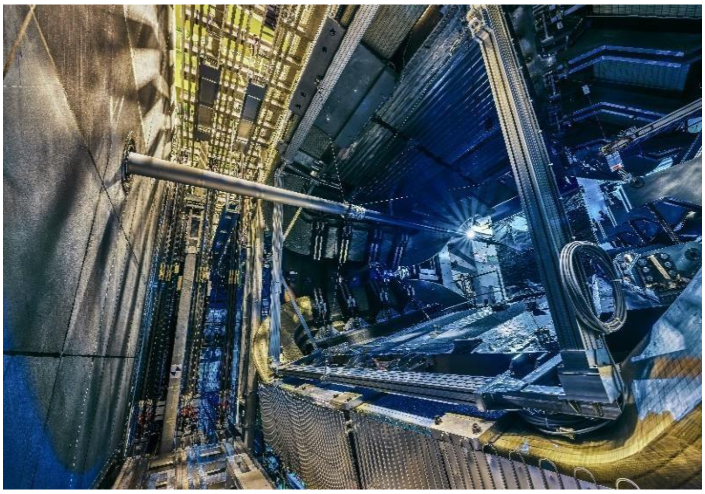
Work related to each of the project's use cases will continue in 2021.

Publications

- G. Passaleva, R. Linder, LHCb Upgrade GPU High Level Trigger Technical Design Report. Published at the CERN Document Server, 2020. http://cern.ch/go/FCV9
- R. Aaij, J. Albrecht, M. Belous, P. Billoir, T. Boettcher, A. Brea Rodríguez, D. vom Bruch, D. H. Cámpora Pérez, A. Casais Vidal, D. C. Craik, P. Fernandez Declara, L. Funke, V. V. Gligorov, B. Jashal, N. Kazeev, D. Martínez Santos, F. Pisani, D. Pliushchenko, S. Popov, R. Quagliani, M. Rangel, F. Reiss, C. Sánchez Mayordomo, R. Schwemmer, M. Sokoloff, H. Stevens, A. Ustyuzhanin, X. Vilasís Cardona, M. Williams, Allen: A high level trigger on GPUs for LHCb. Published at the arXiv.org, 2020. http://cern.ch/go/mX7g
- J. Kieseler, Object condensation: one-stage grid-free multi-object reconstruction in physics detectors, graph and image data. Published at the arxiv.org, 2020. http://cern.ch/go/w8SR

Presentations

- J. Kieseler, Application and development of advanced deep neural networks for high-granularity calorimeters (29 January 2020). Presented at the EP-IT Data science seminars, Geneva, 2020. http://cern.ch/go/6CBh



Allen has been adopted by the LHCb collaboration as the new baseline first-level trigger for the LHC's next run.

TPUs FOR DEEP LEARNING

R&D TOPIC 3:
MACHINE LEARNING
AND DATA ANALYTICS

Google

Project coordinator:
Sofia Vallecorsa

Collaborator liaisons:
Renato Cardoso

Project goal

CERN is running pilot projects to investigate the potential of hardware accelerators, using a set of LHC workloads. In the context of these investigations, this project focuses on the testing and optimisation of machine-learning and deep-learning algorithms on Google TPUs. In particular, we are focusing on generative adversarial networks (GANs).

Background

The high-energy physics (HEP) community has a long tradition of using neural networks and machine-learning methods (random forests, boosted decision trees, multi-layer perceptrons) to solve specific tasks. In particular, they are used to improve the efficiency with which interesting particle-collision events can be selected from the background. In the recent years, several studies have demonstrated the benefits of using deep learning (DL) to solve typical tasks related to data taking and analysis. Building on these examples, many HEP experiments are now working to integrate deep learning into their workflows for many different applications (examples include data-quality assurance, simulation, data analysis, and real-time selection of interesting collision events). For example, generative models, from GANs to variational auto-encoders, are being tested as fast alternatives to simulation based on Monte Carlo methods. Anomaly-detection algorithms are being explored to improve data-quality monitoring, to design searches for rare new-physics processes, and to analyse and prevent faults in complex systems (such as those use for controlling accelerators and detectors).

The training of models such as these has been made tractable with the improvement of optimisation methods and the availability of dedicated hardware that is well adapted to tackling the highly parallelisable task of training neural networks. Storage and HPC technologies are often required by these kinds of projects, together with the availability of HPC multi-architecture frameworks (from large multi-core systems to hardware accelerators like Google TPUs).

Progress in 2020

Machine learning has been used in a wide range of areas. Nevertheless, the need to make it faster while still maintaining accuracy (and thus the validity of results) is a growing problem for data scientists. Our work is exploring the Tensorflow distributed parallel strategy approach, in order to efficiently run a GAN model in a parallel environment. This includes benchmarking different types of hardware for such an approach.

Specifically, we parallelised a 3D convolutional GAN training process on multiple GPUs and multiple Google TPU cores. This involved two main approaches to the Tensorflow mirrored strategy: The first approach uses the default implementation and the built-in logic from the Tensorflow strategy deployment model, with training on several GPUs.The second approach uses a custom training loop that we optimised in order to increase control over the training process, as well as adding further elements to each GPU's work, increasing the overall speedup.

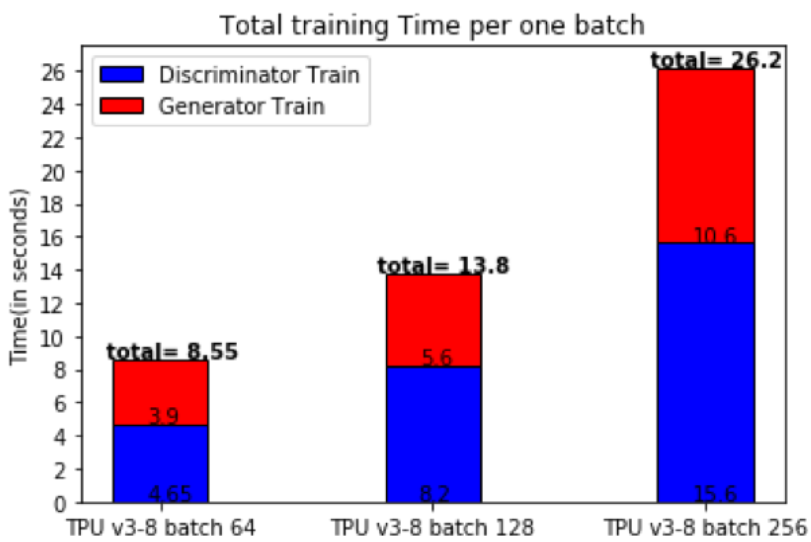
For the TPUs, we used the TPU distributed strategy present in Tensorflow, applying the same approaches as described for the mirrored strategy. This was validated by comparing with the results obtained with the original 3DGAN model, as well as the Monte Carlo simulated data. Additionally, we tested scalability over multiple GPU nodes by deploying the training process on different public cloud providers using Kubernetes.

Next steps

Our work in 2021 will be devoted to the extension of this approach to more efficient parallelisation strategies. We will also work to optimise deep-learning architectures; these are more demanding in terms of computing resources, meaning that they stand to benefit greatly from TPU architectures.

Presentations

- R. Cardoso, Accelerating GAN training using distributed tensorflow and highly parallel hardware (22 October). Presented at the 4th IML workshop, Geneva, 2020. http://cern.ch/go/H6xt



Total training time for the 3D convolutional GAN model measured on different TPU setups.

DATA ANALYTICS IN THE CLOUD

R&D TOPIC 3:

MACHINE LEARNING AND DATA ANALYTICS

ORACLE®

Project coordinator:
Eva Dafonte Perez, Eric Grancher

Technical team:
Luca Canali, Riccardo Castellotti

Collaborator liaisons:
Vincent Leocorbo, Cristobal Pedregal-Martin, David Ebert, Dmitrij Dolgušin

Project goal

This project is evaluating solutions that combine data engineering, machine-learning and deep-learning tools. They are being run using cloud resources — from Oracle Cloud Infrastructure (OCI) — and address a number of use cases of interest to CERN's community. This activity will enable us to compare performance, maturity, and stability of solutions deployed on CERN's infrastructure with the ones in OCI.

Background

Big-data tools — particularly related to data engineering and machine learning — are evolving rapidly. As these tools reach maturity and are adopted more broadly, new opportunities are arising for extracting value out of large data sets.

Recent years have seen growing interest from the physics community in machine learning and deep learning. One important activity in this area has been the development of pipelines for real-time classification of particle-collision events recorded by the detectors of the LHC experiments. Filtering events using so-called "trigger" systems is set to become increasingly complex as upgrades to the LHC increase the rate of particle collisions.

Progress in 2020

SWAN is a platform for performing interactive data analysis in the cloud. It was developed at CERN and integrates software, compute, and storage resources used by CERN physicists and data scientists. In 2020, we deployed a proof-of-concept version of SWAN on OCI resources.

As part of this work, we developed a custom Kubernetes deployment on OCI resources, in order to take advantage of GPU resources. This proved that it is possible to run interactive analytics workflows and ML in OCI while accessing datasets from CERN's storage systems.

We also performed a distributed machine-learning training exercise, with a recurrent neural network using over 250 GB of data. For this, we utilised 500 CPU cores and 10 GPUs using Kubernetes on OCI. Using OCI showed us that public clouds are particularly convenient for use cases that need a large number of resources for a short amount of time.

Next steps

The focus of the project in 2021 will include work to integrate CERN's analytics platform with OCI, enabling users to run their workloads on remote cloud resources using a common interface.

Over the longer term, we are planning to add new features to the analytics platform, with a focus on improving the full lifecycle of the development of machine-learning use cases.

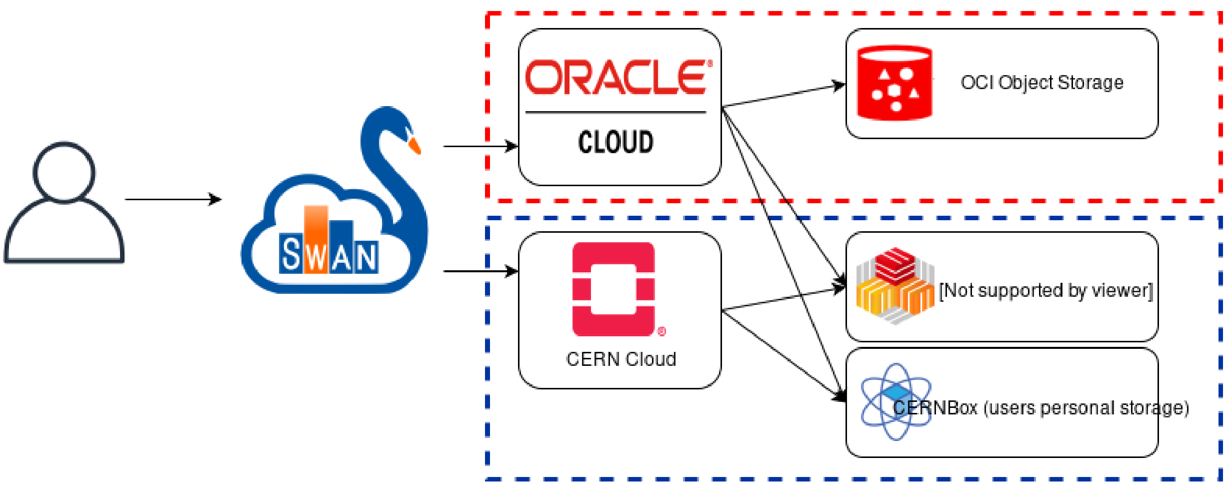
Publications

- M. Migliorini, R. Castellotti, L. Canali, M. Zanetti, Machine Learning Pipelines with Modern Big Data Tools for High Energy Physics. Published on Computing and Software for Big Science, 2020. *http://cern.ch/go/S8wV*

- R.Castellotti, L. Canali, Distributed Deep Learning for Physics with TensorFlow and Kubernetes. Published on CERN DB blog, 2020. *http://cern.ch/go/X6Xk*

Presentations

- R. Castellotti, L. Canali, P. Kothuri, SWAN: Powering CERN's Data Analytics and Machine Learning Use cases (22 October). Presented at 4th Inter-experiment Machine Learning Workshop, CERN, 2020. http://cern.ch/go/6jjX



SWAN (Service for Web based Analysis) is a platform for performing interactive data analysis in the cloud. The integration of SWAN with OCI means that users can run their workloads on OCI or at CERN regardless of where their data is stored.

DATA ANALYTICS FOR INDUSTRIAL CONTROLS AND MONITORING

R&D TOPIC 3:
MACHINE LEARNING AND DATA ANALYTICS

SIEMENS

Project coordinator:
Rafal Kulaga

Technical team:
Filip Široký, Marc Bengulescu, Fernando Varela Rodriguez, Rafal Kulaga, Piotr Golonka, Peter Sollander

Collaborator liaisons:
Thomas Hahn, Juergen Kazmeier, Alexey Fishkin, Tatiana Mangels, Elisabeth Bakany, Ewald Sperrer

Project goal

The main goal of the project is to render the industrial control systems used for the LHC more efficient and more intelligent. The focus is to develop a data-analytics platform that capitalises on the latest advances in artificial intelligence (AI), cloud and edge-computing technologies. The ultimate goal is to make use of analytics solutions provided by Siemens to provide non-expert end users with a turnkey data-analytics service.

Background

The HL-LHC project aims to increase the integrated luminosity — and hence the rate of particle collisions — by a factor of ten beyond the LHC's design value. Monitoring and control systems will therefore become increasingly complex, with unprecedented data throughputs. Consequently, it is vital to further improve the performance of these systems, and to make use of data-analytics algorithms to detect anomalies and to anticipate future behaviour. Achieving this involves a number of related lines of work. This particular project focuses on the development of a data-analytics platform that combines the benefits of cloud and edge computing.

Progress in 2020

One of the main achievements in 2020 was the experimental deployment of the Siemens Distributed Complex Event Processing (DCEP) technology to enable advanced data analytics and predictive maintenance for the oxygen-deficiency sensors in the LHC tunnel. This was initially done by deploying a suite of microservices on a pool of virtual machines in the cloud. In the latest phase, the cloud-edge gap was bridged by also adding a

Siemens IoT 2050 box as a worker to the computing pool.

Progress was also made on the optimisation of the ion-beam source for the LINAC3 accelerator at CERN. Multiple machine-learning and spectral techniques were employed and extensively tested, under the guidance of LINAC3 experts. We expect to continue this work in 2021.

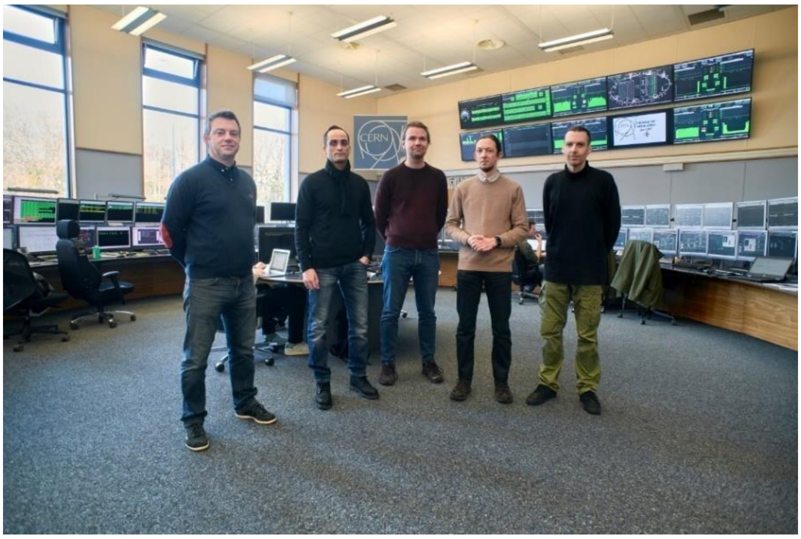
A prototype platform was developed to provide collection service for generic AI algorithms that could easily be employed by people who are not data scientists, helping them to perform advanced analytics on controls data. This is an ongoing effort together with colleagues from various groups at CERN, including the Cryogenics group in the Technology department and the Cooling and Ventilation group in the Engineering department.

Next steps

Depending on the priorities agreed with the company, the focus of the collaboration can shift to new areas, such as device management. A CERN-SIEMENS joint workshop has been scheduled in order to define the use cases for the following year.

Presentations

- F. Široký, Data Analytics and IoT for Industrial Control Systems (22 January). Presented at CERN openlab technical workshop, Geneva, 2020. http://cern.ch/go/F6NM
- M. Bengulescu, F. Široký, Distributed Complex Event Processing at the Large Hadron Collider (2 July). Presented at IoT Siemens Conference, 2020.
- F. Široký, M. Bengulescu, Predicting LINAC3 current with LSTM and wavelet spectral analysis (22 September). Presented at Siemens Corporate Technology Seminar, 2020.
- M. Bengulescu, F. Široký, Predicting LINAC3 current using RNNs and spectral decomposition (19 June). Presented at CERN ML Coffee, Geneva, 2020.
- F. Široký, Bayesian ML – a practical approach (6 May). Presented at CERN ML Coffee, Geneva, 2020.
- M. Bengulescu, F. Široký, Data Analytics for Industrial Controls (31 January). Presented at CERN ML Coffee, Geneva, 2020.



The team is focusing on the development of a data-analytics platform that combines the benefits of cloud and edge computing.

EXPLORING ACCELERATED MACHINE LEARNING FOR EXPERIMENT DATA ANALYTICS

R&D TOPIC 3: MACHINE LEARNING AND DATA ANALYTICS

Project coordinator:
Emilio Meschi, Paola Sala, Maria Girone

Technical team:
Thomas Owen James, Dejan Golubovic, Maurizio Pierini, Manuel Jesus Rodriguez, Saul Alonso-Monsalve, Ema Puljak, Lorenzo Uboldi

Collaborator liaisons:
Mark Hur, Stuart Grime, Michael Glapa, Eugenio Culurciello, Andre Chang, Marko Vitez, Dustin Werran, Aliasger Zaidy, Abhishek Chaurasia, Patrick Estep, Jason Adlard, Steve Pawlowski

Project goal

The project has two threads, each investigating a unique use case for the Micron Deep Learning Accelerator (a modular FPGA-based architecture). The first thread relates to the development of a real-time streaming machine inference engine prototype for the level-1 trigger of the CMS experiment.

The second thread focuses on prototyping a particle-identification system based on deep learning for the DUNE experiment. DUNE is a leading-edge, international experiment for neutrino science and proton-decay studies. It will be built in the US and is scheduled to begin operation towards the end of this decade.

Background

The level-1 trigger of the CMS experiment selects relevant particle-collision events for further study, while rejecting 99.75% of collisions. This decision must be made with a fixed latency of a few microseconds. Machine-learning inference in FPGAs may be used to improve the capabilities of this system.

The DUNE experiment will consist of large arrays of sensors exposed to high-intensity neutrino beams. The use of convolutional neural networks has been shown to substantially boost particle-identification performance for such detectors. For DUNE, an FPGA solution is advantageous for processing ~ 5 TB/s of data.

Progress in 2020

The CMS team primarily focussed on preparing a level-1 scouting system using the Micron SB-852 FPGA processing boards to capture and process trigger-level data at 40 MHz. We developed and optimised neural networks to improve analysis performance using level-1 scouting objects. In addition, we developed a system for level-1 anomaly detection using a variational auto-encoder approach. We implemented this using the Micron deep-learning framework.

The DUNE team organised a test on the protoDUNE Single Phase detector to analyse data from cosmic rays. This was the last chance to test it before protoDUNE's second run in 2022. The test aimed to examine the incoming triggered data using a triple AC-511 Micron FPGA as a hardware accelerator. The hardware ran an image-segmentation neural network designed to identify regions of interest. This setup was able to analyse data at a rate of ~1.42 Gb/s. The capability of the network to identify low-energy events was tested in data taken by protoDUNE with a neutron generator.

Next steps

The CMS team will focus on completing the installation and validation of the 40 MHz scouting system for LHC Run 3. This will require the integration and debugging of several software, FPGA-firmware, and hardware layers. The performance of a deep-learning-driven anomaly-detection algorithm will also be evaluated for use in LHC Run 3.

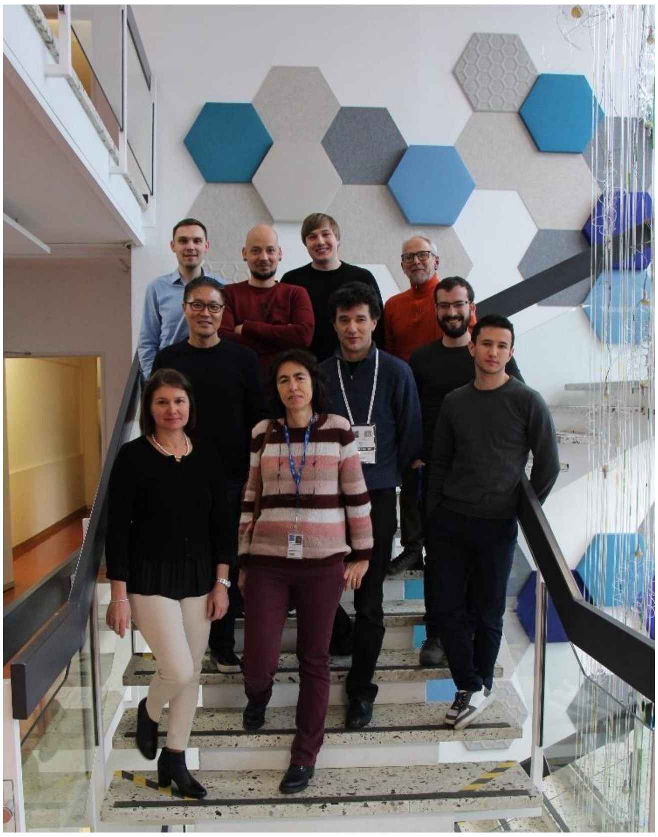
The DUNE team plans to continue analysing the neutron generator's data, as it demonstrates huge potential for the reconstruction techniques at DUNE. The goal is to perform studies in real time to understand the quality of the data taken using Micron's state-of-the-art hardware accelerators for the data-acquisition system of protoDUNE.

Publications

- D. Golubovic, 40 MHz scouting with deep learning in CMS. Published on Zenodo, 2020. *http://cern.ch/go/vJD9*
- M. Popa, Deep learning for 40 MHz scouting with level-1 trigger muons for CMS at LHC run-3. Published on Zenodo, 2020. *http://cern.ch/go/99St*

Presentations

- M. Rodríguez, S. A. Monsalve, P. Sala, Prototyping of a DL-based Particle Identification System for the Dune Neutrino Detector (22 January). Presented at CERN openlab Technical Workshop, Geneva, 2020. *cern.ch/go/zH8W*
- D. Golubovic, T. James, E. Meschi, 40 MHz Scouting with Deep Learning in CMS (22-24 April). Presented at Connecting the Dots Workshop, New Jersey, 2020. *http://cern.ch/go/C96N*



EVALUATION OF POWER CPU ARCHITECTURE FOR DEEP LEARNING

R&D TOPIC 3:
MACHINE LEARNING AND DATA ANALYTICS

Project coordinator:
Maria Girone and Sofia Vallecorsa

Technical team
Marco Rossi

Collaborator liaisons:
Eric Aquaronne, Oliver Bethmann

Project goal

We are investigating the performance of distributed training and inference of different deep-learning models on a cluster consisting of IBM Power8 CPUs (with NVIDIA V100 GPUs) installed at CERN. A series of deep neural networks is being developed to reproduce the initial steps in the data-processing chain of the DUNE experiment. More specifically, a combination of convolutional neural networks and graph neural networks are being designed to reduce noise and select specific portions of the data to focus on during the reconstruction step (region selector).

Background

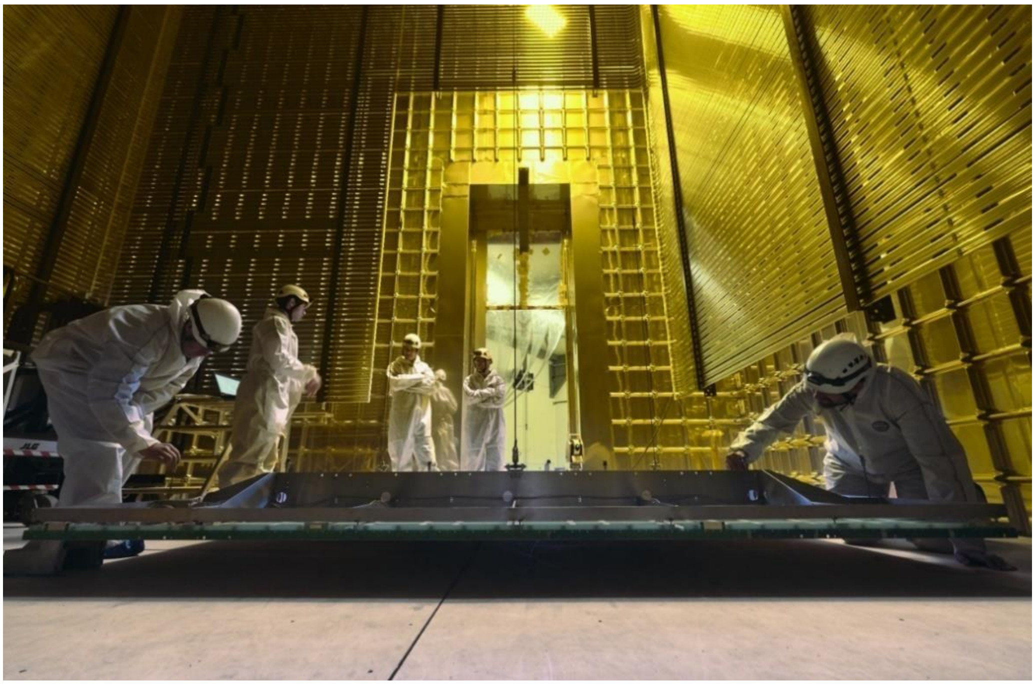
Neutrinos are elusive particles: they have a very low probability of interacting with other matter. In order to maximise the likelihood of detection, neutrino detectors are built as large, sensitive volumes. Such detectors produce very large data sets. Although large in size, these data sets are usually very sparse, meaning dedicated techniques are needed to process them efficiently. Deep-learning methods are being investigated by the community with great success.

Progress in 2020

We have developed a deep neural network architecture based on a combination of two-dimensional convolutional layers and graphs. These networks can analyse both real and simulated data from protoDUNE and perform the region selection and de-noising tasks, which are usually applied to the raw detector data before any other processing is run.

Both of these methods improve on the classical approaches currently integrated in the experiment software stack. In order to reduce training time and set up hyper-parameter scans, the training process for the networks is parallelised and has been benchmarked on the IBM Minsky cluster.

In accordance with the concept of data-parallel distributed learning, we trained our models on a total of twelve GPUs, distributed over the three nodes that comprise the test Power cluster. Each GPU ingests a unique part of the physics dataset for training the model.



Scientists working on the ProtoDUNE experiment.

NEXT GENERATION ARCHIVER FOR WINCC OA

R&D TOPIC 3:
MACHINE LEARNING
AND DATA ANALYTICS

SIEMENS

Project coordinator:
Rafal Kulaga

Technical team:
Anthony Hennessey, Rafal Kulaga, Mariusz Suder, Piotr Golonka, Peter Sollander, Fernando Varela, Marc Bengulescu, Filip Siroky

Collaborator liaisons:
Thomas Hahn, Juergen Kazmeier, Alexey Fishkin Tatiana Mangels, Mikhail Kalinkin, Elisabeth Bakany, Ewald Sperrer

Project goal

Our aim is to make control systems used for the LHC more efficient and smarter. We are working to enhance the functionality of WinCC OA (a SCADA tool used widely at CERN) and to apply data analytics techniques to the recorded monitoring data, in order to detect anomalies and systematic issues that may impact upon system operation and maintenance.

Background

The HL-LHC programme aims to increase the integrated luminosity — and hence the rate of particle collisions — by a factor of ten beyond the LHC's design value. Monitoring and control systems will therefore become increasingly complex, with unprecedented data throughputs. Consequently, it is vital to further improve the performance of these systems, and to make use of data-analytics algorithms to detect anomalies and anticipate future behaviour. Achieving this involves a number of related lines of work. This project focuses on the development of a modular and future-proof archiving system (NextGen Archiver) that supports different SQL and NOSQL technologies to enable data analytics. It is important that this can be scaled up to meet our requirements beyond 2020.

Progress in 2020

The most important milestone for the NextGen Archiver (NGA) project in 2020 was the preparation of the first stable version with Oracle and InfluxDB backends for all users at CERN. Despite challenges, it was successfully released in July, receiving positive feedback.

The release is currently being deployed in the ALICE systems, where the NGA will be used with a custom backend to stream data from the control systems to the new physics data readout and analysis architecture after Long Shutdown 2. This represents a crucial validation step for the massive deployment of the new archiver in all CERN systems, planned for Long Shutdown 3.

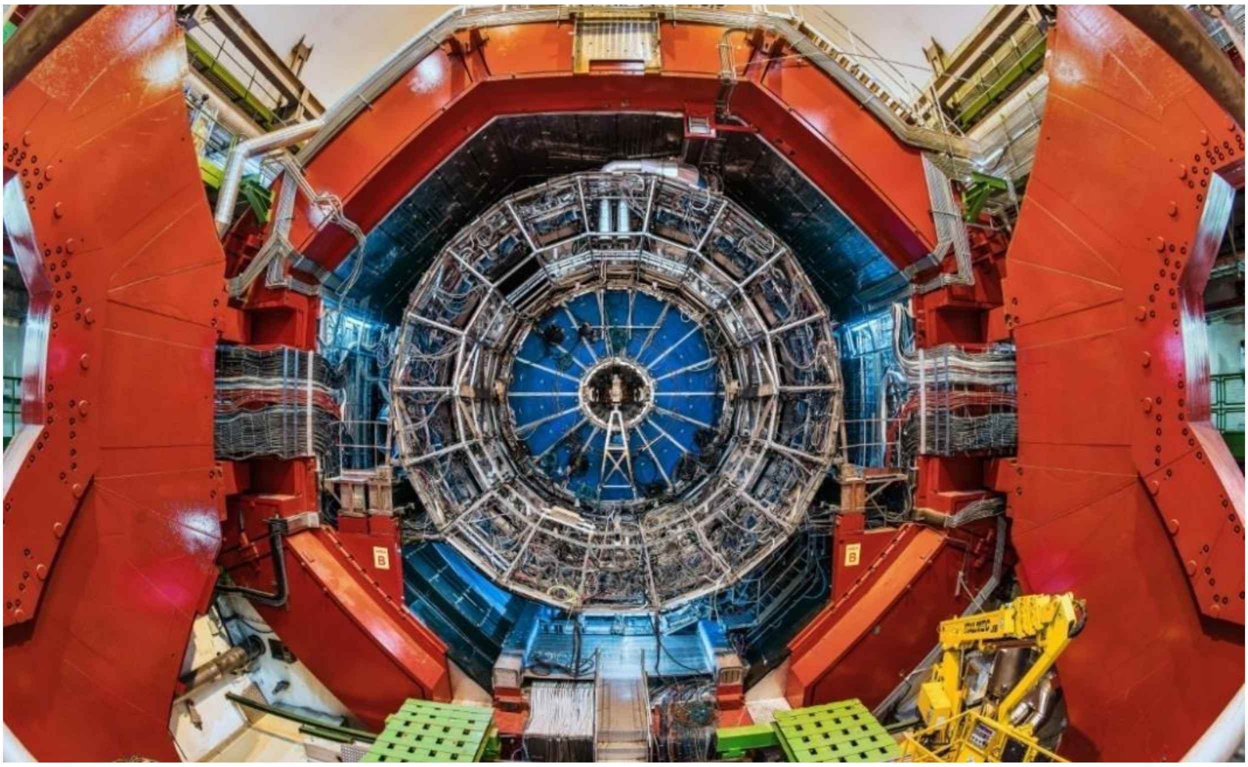
After the release, the team focused on developing several features and improving reliability. These upgrades will be included in the subsequent versions. One notable upgrade is the ability to send queries to selected backends, with the option to specify different time ranges for each of them. This functionality will be indispensable in systems where parallel archiving into multiple databases is used to improve performance; it will enable new use cases.

Next steps

The work on the project will continue on several fronts. The NGA will be deployed in all systems in the ALICE experiment. A further increase of test coverage is also one of the priorities, with particular attention to performance, handling of failure scenarios, and redundancy. The work on multiple features of the archiver will continue, including extensions to the query mechanisms and improvements in all the backends.

Presentations

- A. Hennessey, P. Golonka, R. Kulaga, F. V.arela, WinCC Open Architecture – Next Generation Archiver (23 January). Presented at CERN openlab Technical Workshop, Geneva, 2020. http://cern.ch/go/8Kq7



The ALICE experiment will pioneer the deployment of the NGA. It will play a critical role in a project known as O2, which will provide a new facility for handling the increased data produced by the ALICE experiment in LHC Runs 3 and 4.

ORACLE CLOUD TECHNOLOGIES FOR DATA ANALYTICS ON INDUSTRIAL CONTROL SYSTEMS

R&D TOPIC 3
MACHINE LEARNING
AND DATA ANALYTICS

Project coordinator:
Manuel Martin Marquez, Sébastien Masson

Technical team:
Manuel Martin Marquez, Sébastien Masson, Aimilios Tsouvelekakis

Collaborator liaisons:
Çetin Özbütün, Reiner Zimmermann, Michael Connaughton, Cristobal Pedregal-Martin, Engin Senel, Cemil Alper, Giuseppe Calabrese, David Ebert, Dmitrij Dolgušin

Project goal

CERN's control systems acquire more than 250 TB Of data per day from LHC and its experiments. Managing these extremely complex "Industrial Internet of Things" (IIoT) systems raises important challenges in terms of data management, retrieval, and analytics.

The main objective is to explore the capabilities of Oracle Autonomous Data Warehouse and Oracle Analytics Cloud for integrating heterogeneous control IIoT data, while improving performance and efficiency for the most challenging analytics requirements.

Background

Keeping the LHC and the rest of the accelerator complex at CERN running efficiently requires state-of-the-art control systems. A complex IIOT system has been developed to persist this data, making it possible for engineers to gain insights about temperatures, magnetic-field strengths, and more. This plays a vital role in ensuring the highest levels of operational efficiency.

The current system for persisting, accessing and analysing this data is based on Oracle Database. Today, significant effort is dedicated to improving performance and coping with increasing demand — in terms of data volume, analysis and exploration of bigger data sets.

Progress in 2020

During 2020, the team focused on three main aspects:

(i) Assessing extended capabilities for supporting custom data types when importing and dealing with complex data schemas on parquet files. The results showed that complex parquet schemas can now be handled and automatically ingested by Oracle Data Warehouse technology. This enabled us to work with data coming from control-system devices based on CERN's custom controls middleware, CMW.

(ii) Deploying a hybrid solution based on standard and external table partitions, with the objective of improving performance for data retrieval. This improved the performance compared to the solution based on external partitions and is thus a good fit for the most demanding real-time applications consuming data from the IIoT control system.

(iii) Increasing data retrieval/analytics complexity using real-life scenarios to explore Oracle Analytics technologies as a front-end solution for control engineers and equipment experts.

Keeping the LHC and the rest of the accelerator complex at CERN running efficiently requires state-of-the-art control systems.

We presented the outcomes of this work at events across Asia Pacific, Europe, Middle East, Africa and North America. Through our participation in these events, CERN openlab could share insights obtained with representatives of large companies from various industries, ranging from banking and telecommunications to research and educational institutions.

Next steps

The team will increase the data volume and complexity to assess the capabilities of the new Autonomous Database for drag-and-drop data loading and transformation, as well as for automatic insight discovery. This will be done with the goal of going one step further in terms of automating processes to improve operational efficiency.

Presentations

- M. Marquez, Managing 1 PB of Object Storage in the Oracle Cloud (9 July). Presented at the Oracle Cloud Platform Virtual Summit: The Modern Data Warehouse, North America, 2020.
- M. Marquez, Managing 1 PB of Object Storage in the Oracle Cloud (6 August). Presented at the Oracle Cloud Platform Virtual Summit: The Modern Data Warehouse, EMEA, 2020. *cern.ch/go/r7cX*
- M. Marquez, Managing 1 PB of Object Storage in the Oracle Cloud (15 August). Presented at the Oracle Cloud Platform Virtual Summit: The Modern Data Warehouse, JAPAC, 2020. *cern.ch/go/t6JQ*
- M. Marquez, CERN Industrial IoT data with Oracle Autonomous Data Warehouse (2 June). Presented at the Oracle Global Leaders Summer Meeting, Miami, 2020. *cern.ch/go/CH9d*
- S. Masson, La gestion des données en gros volume au quotidien (19 May). Presented at the Oracle Technology Day 2020: Data, innovation et continuité d'activité, 2020. *http://cern.ch/go/8xkP*

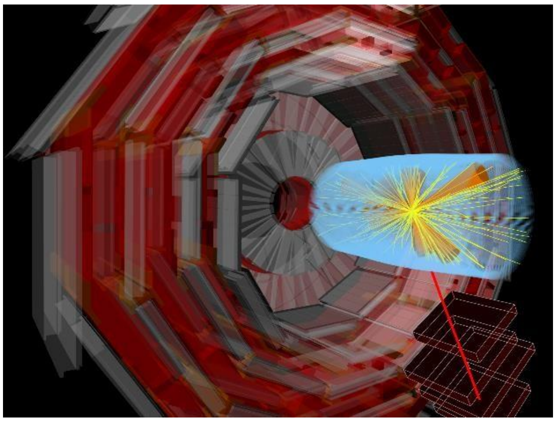
FAST DETECTOR SIMULATION

R&D TOPIC 3:
MACHINE LEARNING
AND DATA ANALYTICS

Project coordinator:
Sofia Vallecorsa

Technical team:
Florian Rhem, Gurlukh Khattak, Krisitna Jaruskova

Collaborator liaisons:
Intel: Claudio Bellini, Marie-Christine Sawley, Andrea Luiselli, Saletore Vikram, Hans Pabst, Adel Chaibi, Eric Petit. SURFsara BV: Valeriu Codreanu, Maxwell Cai, Damian Podareanu. Barcelona Suepercomputing Center: John Osorio Rios, Adrià Armejach Marc Casas



Simulating the response of detectors to particle collisions is an important step on the path to new physics discoveries.

Project goal

We are using artificial intelligence (AI) techniques to simulate the response of the particle detectors to collision events. Specifically, we are developing deep neural networks — in particular, generative adversarial networks (GANs) — to do this. Such tools will play a significant role in helping the research community cope with the vastly increased computing demands of the High Luminosity LHC (HL-LHC).

Once properly trained and optimised, generative models can simulate a variety of particles, energies, and detectors in just a fraction of the time required by classical simulation, which is based on detailed Monte Carlo methods. Our objective is to tune and integrate these new tools in the experiments' existing simulation frameworks.

Background

Simulating the response of detectors to particle collisions — under a variety of conditions — is an important step on the path to new physics discoveries. However, this work is very computationally expensive. Over half of the computing workload of the Worldwide LHC Computing Grid (WLCG) is the result of this single activity.

We are exploring an alternative approach, referred to as 'fast simulation', which trades some level of accuracy for speed. Fast-simulation strategies have been developed in the past, using different techniques (e.g. look-up tables or parametrised approaches).

However, the latest developments in machine learning (particularly in relation to deep neural networks) make it possible to develop fast-simulation tools that are both more flexible and more accurate than those developed in the past.

Progress in 2020

Most of the work in 2019 focused on the acceleration of the training process using a data-parallel approach. In 2020 we turned our attention to the optimisation and acceleration of the inference process. Industry is developing new hardware platforms that promise large acceleration factors for the training and inference processes related to deep neural networks (e.g. Intel XE). In most cases, low-precision data representation (e.g. half-precision floating points or half-precision integers) is one of the key strategies for achieving significant acceleration. Given this, we have carefully studied the effect of low-precision data representation on the 3DGAN model. We obtained a 1.8x speedup by running inference using a half-precision integer representation, compared to using single-precision float points. We verified that the precision of physics results is conserved with this approach. We also verified that using a mixed-precision approach for training (dynamically switching between single-precision and half-precision floating points) converges to stable results.

Next Steps

The work done so far on 3DGAN can be considered as an initial R&D phase. Our focus now is on moving from the prototyping stage to production, deployment and integration within the simulation software. To achieve this goal, it is essential to optimise resources, stabilising the training process and improving model convergence. At the same time, it is also important to perform systematic studies on model generalisation and robustness, as well as on results interpretability and reproducibility.

Validating the performance of a generative model is not an easy task. In particular, evaluating the number of missing modes (as well as their properties) is critical for ensuring that the simulated data are a good representation of the underlying theoretical models, thus meaning that they can be safely used to evaluate detector performance and model their response.

Building on the work done to optimise the 3DGAN discriminator network, the plan is to design a convolutional neural network (CNN) able to analyse the GAN-generated images and to act as a feature extractor. The CNN output can then be analysed by an XGBoost-based analyser, solving the final classification or regression problem.

Motivated by the issue of missing modes, we also intend to develop and optimise a boosting approach to improve the convergence of the 3DGAN model.

Publications

- J. Rios, A. Armejach, G. Khattak, E. Petit, S. Vallecorsa, M. Casas, Mixed-Precision Arithmetic for 3DGAN to Simulate High Energy Physics Detectors. Published at the ICMLA, 2020.

Presentations

- J. Rios, A. Armejach, G. Khattak, E. Petit, S. Vallecorsa, M. Casas, Mixed-Precision Arithmetic for 3DGAN to Simulate High Energy Physics Detectors (13 October). Presented at the 2020 IXPUG Annual Meeting, US, 2020. http://cern.ch/go/7PXj

- F. Rhem, Reduced Precision Strategies for Deep Learning: A GAN use case from High Energy Physics (19 October). Presented at the FS 2020 Accelerated Artificial Intelligence for Big-Data Experiments Conference, Urbana-Champaign, 2020. http://cern.ch/go/T6wj

- F. Rhem, S. Vallecorsa, Reduced Precision Strategies for Deep Learning: 3DGAN Use Case (21 October). Presented at the 4th IML Machine Learning Workshop CERN, Geneva, 2020. http://cern.ch/go/Tnf8

QUANTUM GRAPH NEURAL NETWORKS

R&D TOPIC 4:
QUANTUM TECHNOLOGIES

Project coordinator:
Sofia Vallecorsa

Team members:
Fabio Fracas, Cenk Tüysüz, Jean-Roch Vlimant, Bilge Demirkoz, Kristiane Novotny, Carla Rieger

Collaborator liaisons :
Daniel Dobos, Karolos Potamianos.

Project goal

The goal of this project is to explore the feasibility of using quantum algorithms to help track the particles produced by collisions in the LHC more efficiently. This is particularly important as the rate of collisions is set to increase dramatically in the coming years.

Background

The LHC at CERN is producing collisions at unprecedented collider energy. The hundreds of particles created during the collisions are recorded by large detectors composed of several sub-detectors. At the centre of these detectors there is usually a tracker detector, precisely recording the signal of the passage of charged particles through thin layers of active material. The trajectories of particles are bent by a magnetic field to allow the measurement of the particle momentum. There is an expected ten-fold increase in the number of tracks produced per bunch crossing after the high-luminosity upgrade of LHC.

Classical algorithms perform the reconstruction of the trajectory of charged particles; these make use of Kalman filter formalism. Although they are quite accurate, they scale worse than quadratically with the number of tracks. Several ways are explored to mitigate the increase in the computing needs, such as new detector layout, deep-learning and code parallelisation.

Quantum computing has been shown to provide speed ups for certain problems and different R&D initiatives are exploring how quantum tracking algorithms could exploit such capabilities. We are developing a quantum-based track-finding algorithm aimed at reducing the combinatorial background during the initial seeding stage for the Kalman filters. We are using the publicly available dataset designed for the Kaggle 'TrackML' challenge for this work.

Progress in 2020

We have developed a prototype quantum graph neural network (QGNN) algorithm for tracking the particles produced by collision events. The model uses a graph interpretation for trajectory reconstruction by representing detector hits with nodes in a graph and segments among hits as graph connections. The quantum model is based on cascade hierarchical classifiers, identifying nodes and segments belonging to particle tracks.

Several architectures have been investigated, ranging from tree tensor networks to multi-scale entanglement renormalization ansatz (MERA) graphs, and the results were compared against classical graph neural networks (GNNs). We studied the way in which both quantum effects, such as entanglement, and the expressibility of quantum circuits impacts the final performance. In addition, a student project during the summer focused on the design and optimisation of a hybrid approach to data embedding, combining a classical multi-layer perceptron to a quantum circuit. The results were presented at several international conferences and workshops, including Connecting The Dots 2020 and IOP Quantum 2020.

In parallel to the high-energy physics use case, we also started investigating the application of this methodology to the collision-avoidance systems used for commercial air traffic.

Next Steps

This work represents a first complete look at particle track reconstruction using a QGNNs. Our results show that the prototype can perform at a similar level to classical approaches. We have also learned about how scaling the network depth can improve its performance.

However, it is known that there are many obstacles to the efficient training of extended QGNNs. Therefore, we are currently maintaining a conservative attitude towards the advantages offered by such networks.

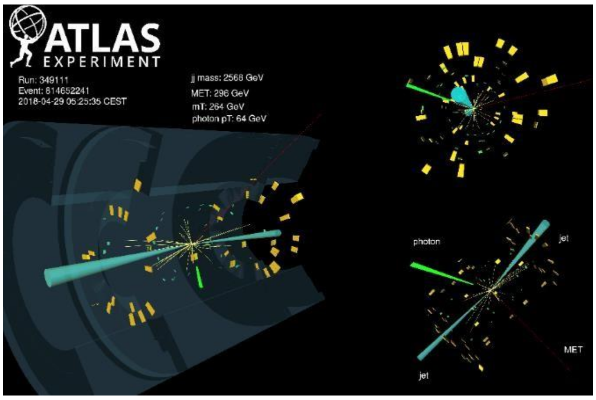
Our plan is to improve the model by introducing better strategies for encoding information. We will also continue the analysis of the properties of the quantum circuits, in order to avoid the problems typically encountered when training hybrid models, such as the effect of vanishing gradients.

Publications

- C. Tüysüz, F. Carminati, B. Demirköz, D. Dobos, F. Fracas, K. Novotny, K. Potamianos, S. Vallecorsa, J. R. Vlimant, A Quantum Graph Neural Network Approach to Particle Track Reconstruction. Published at the arXiv.org, 2020. http://cern.ch/go/z7zx

Presentations

- C. Tuysuz, A Quantum Graph Neural Network Approach to Particle Track Reconstruction (20 April). Presented at Connecting the Dots Workshop, New Jersey, 2020. http://cern.ch/go/N88n
- K. Novotny, Exploring (Quantum) Track Reconstruction Algorithms for non-HEP applications (21April). Presented at Connecting the Dots Workshop, New Jersey, 2020. http://cern.ch/go/8lX9
- C. Rieger (October). Presented at the IOP Quantum 2020 conference, 2020.



This project is exploring the feasibility of using quantum algorithms to help track the particles produced by collisions in the LHC more efficiently.

QUANTUM MACHINE LEARNING for SUPERSYMMETRY SEARCHES

R&D TOPIC 4:
QUANTUM TECHNOLOGIES

Project coordinator:
Koji Terashi

Team members:
Michiru Kaneda, Tomoe Kishimoto, Masahiko Saito, Ryu Sawada, Junici Tanaka, Federico Carminati, Sofia Vallecorsa, Fabio Fracas

Project goal

The goal of this project is to develop quantum machine-learning algorithms for the analysis of particle collision data from the LHC experiments. The particular example chosen is the identification and classification of supersymmetry signals from the Standard Model background.

Background

The analysis of LHC data for the detection of effects beyond the Standard Model requires increasing levels of precision. Various machine-learning techniques are now part of the standard analysis toolbox for high-energy physics. Deep-learning algorithms are increasingly demonstrating their usefulness in various areas of analysis, thanks to their suitability for exploring a much larger dimensional space.

This seems close to an ideal area of application for quantum computing, which offers a parameter space that is potentially enormous, as well as a correspondingly large level of computational parallelism. Moreover, the quasi-optimal Gibbs sampling features of quantum computers may enhance the training of the deep-learning networks.

Progress in 2020

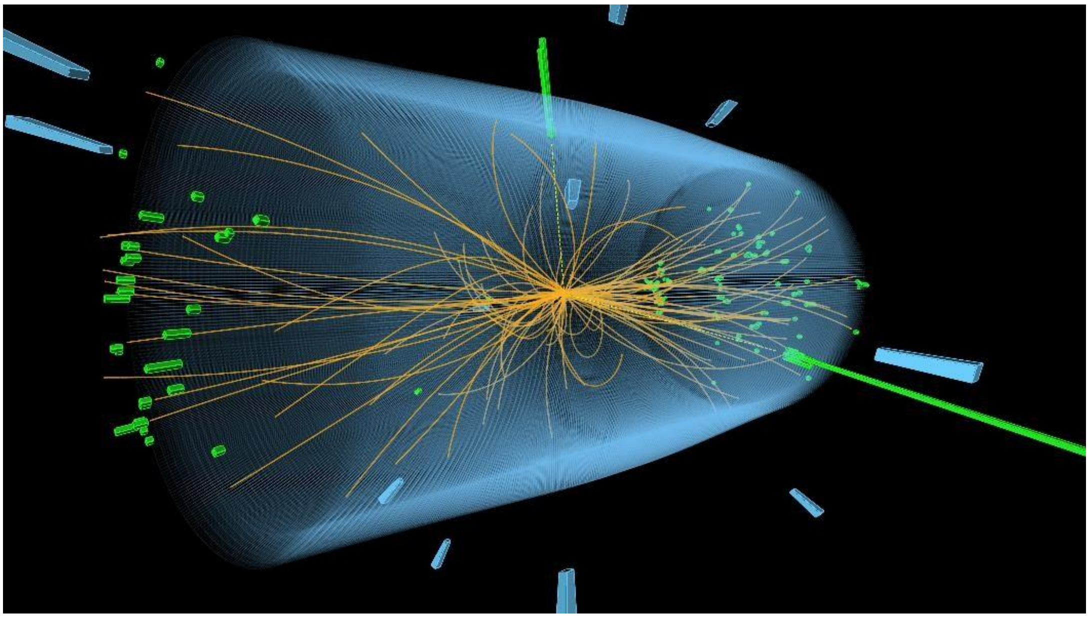
In 2020, the team in Tokyo pursued the study and characterisation of several quantum algorithms, in order to understand their applicability at different points in the data processing chain. One of the problems being addressed is the acceleration of the event-generation process. Event generators (in particular their integration step) are expected to be extremely demanding in terms of computation time, especially for the generation of multi-jet events at the experiments running in future on the High-Luminosity LHC.

In another line of work, we studied different quantum machine-learning models, such as quantum convolutional neural networks, as applied to both classical and quantum data.

With the limited qubit counts, connectivity and coherence times of present quantum computers, quantum circuit optimisation is crucial for making the best use of these devices. In addition to the algorithmic studies outlined above, we have studied a novel circuit-optimisation protocol, composed of two techniques: pattern recognition of repeated sets of gates and reduction of circuit complexity by identifying computational basis states. The developed optimisation protocol demonstrates a significant gate reduction for a quantum algorithm used to simulate parton shower processes.

Next steps

Current results are very promising. However, they have been obtained using simulations that assume ideal quantum circuits, without taking into account effects such as noise, measurement errors, or limited coherence time. All of these issues can affect quantum hardware significantly, particularly noisy intermediate-scale quantum devices. Our next steps will include a detailed study of strategies for error mitigation.



The analysis of LHC data for the detection of effects beyond the Standard Model requires increasing levels of precision.

QUANTUM COMPUTING for SIMULATION: INVESTIGATING QUANTUM GENERATIVE ADVERSIAL NETWORKS and QUANTUM RANDOM NUMBER GENERATORS

R&D TOPIC 4:
QUANTUM TECHNOLOGIES

Project coordinator:
Sofia Vallecorsa

Team members:
CERN: Su Yeon Chang
University of Oviedo: Elias Combarro
CQC: Simon McAdams, Ross Duncan, Mattia Fiorentini, Steven Herbert, Alec Edgington, Cameron Foreman, Florian Curchod, Chad Edwards, Kimberley Worral

Project goal

The collaboration with Cambridge Quantum Computing (CQC) is investigating the advantages and challenges related to the integration of quantum computing into simulation workloads. This work is split into two main areas of R&D: (i) developing quantum generative adversarial networks (GANs) and (ii) testing the performance of quantum random number generators. This second area involves testing the performance of such generators with respect to modern pseudo-random number generators in the context of simulation software used in high-energy physics.

Background

Research in high-energy physics, as in many other scientific domains, makes extensive use of Monte Carlo calculations — both in theory and in experiments.

GANs are among the most interesting models in classical machine learning. GANs are an example of generative models (i.e. models that learn a hidden distribution from a training dataset) and can sample new synthetic data. We have been investigating their use as an alternative to Monte Carlo simulation, obtaining remarkable results. Much faster than standard Monte Carlo algorithms, GANs can generate realistic synthetic data, while retaining a high level of accuracy (see our fast simulation project). Quantum GANs could have more representational power than classical GANs, making them better able to learn more complex distributions from smaller training datasets.

We are now training a quantum GAN to generate images of a few pixels.

For this we are investigating two possible approaches:

- A hybrid schema with a quantum generator learning the target PDF, using either a classical network or a variational quantum circuit as a discriminator (variational quantum generator).
- A full quantum adversarial implementation (quGAN).

Monte Carlo methods make extensive use of random numbers. To ensure simulated data is unbiased, it is essential to understand the quality of the random number generators, the question of true randomness, and other related issues. Modern pseudo-random number generators exhibit optimal performance; we are studying the effect of replacing those with real quantum numbers in different Monte Carlo applications.

Progress in 2020

GANs are systems composed of two networks (a generator and a discriminator) that are trained one against the other in an adversarial fashion. We developed both full quantum GAN models and hybrid classical-quantum GAN models. In the latter kind, the generator is implemented as a variational quantum system and the discriminator is a classical neural network.

We tested different configurations (in terms of number of qubits and circuit depth) in order to reproduce the energy pattern deposited along the longitudinal direction of a calorimeter.

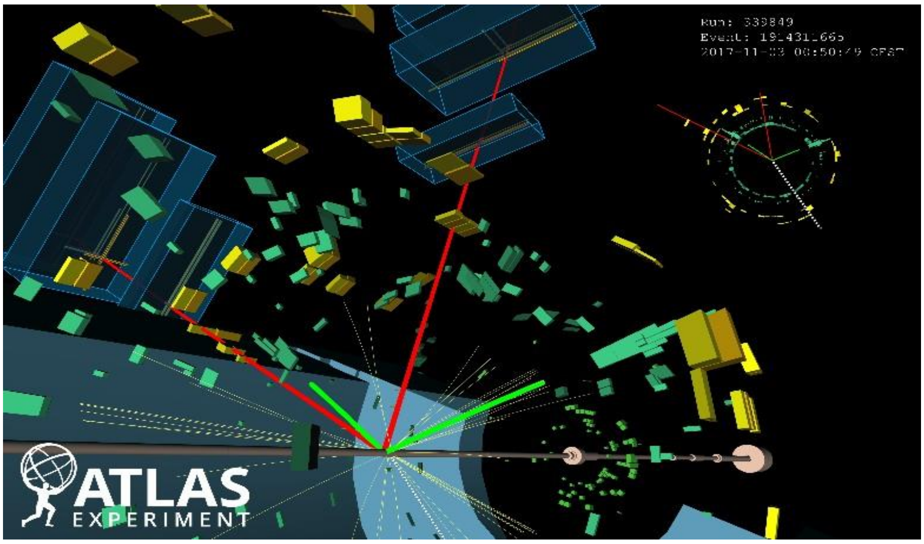
In terms of testing quantum random generators, we investigated various Ising models and machine-learning applications. The results obtained did not demonstrate sufficient sensitivity to show relevant differences between pseudo and real quantum random numbers.

Next steps

The results we obtained in 2020 are extremely promising. In 2021, we will explore techniques to stabilise GAN training and increase the size of the simulated output. We will also explore the effect of noise- and error-mitigation strategies on the quantum and hybrid GAN prototypes.

Presentations

- S. Chang, Quantum 2020 review (19-22 October). Presented at IOP Quantum, 2020.



This project is investigating the integration of quantum computing into simulation workloads.

QUANTUM OPTIMISATION FOR GRID COMPUTING

R&D TOPIC 4:
QUANTUM TECHNOLOGIES

Project coordinator:
Sofia Vallecorsa

Team members:
CERN: Fabio Fracas, Costin Grigoras, Latchezar Betev
Polytechnic Institute of Grenoble: Jaques Demongeot
Polytechnic University of Bucharest: Mircea-Marian Popa, Mihai Carabas, George Pantelimon Popescu

Project goal

The goal of this project is to develop quantum algorithms to help optimise how data is distributed for storage in the Worldwide LHC Computing Grid (WLCG), which consists of 167 computing centres, spread across 42 countries. Initial work focuses on the specific case of the ALICE experiment. We are trying to determine the optimal storage, movement, and access patterns for the data produced by this experiment in quasi-real-time. This would improve resource allocation and usage, thus leading to increased efficiency in the broader data-handling workflow.

Background

The WLCG has been essential to the success of the LHC's scientific programme. It is used to store and analyse the data produced by the LHC experiments. Optimal usage of the grid's resources is a major challenge: with the foreseen increase in the data produced by the LHC experiments, workflow optimisation — particularly for the data-placement strategy — becomes extremely important.

Simulating this complex and highly non-linear environment is very difficult; the complexity of the task goes beyond the capability of the computing hardware available today. Quantum computing could offer the possibility to address this. Our project, a collaboration with the Polytechnic Institute of Grenoble and the Polytechnic University of Bucharest, will develop quantum algorithms to optimise the storage distribution.

Progress in 2020

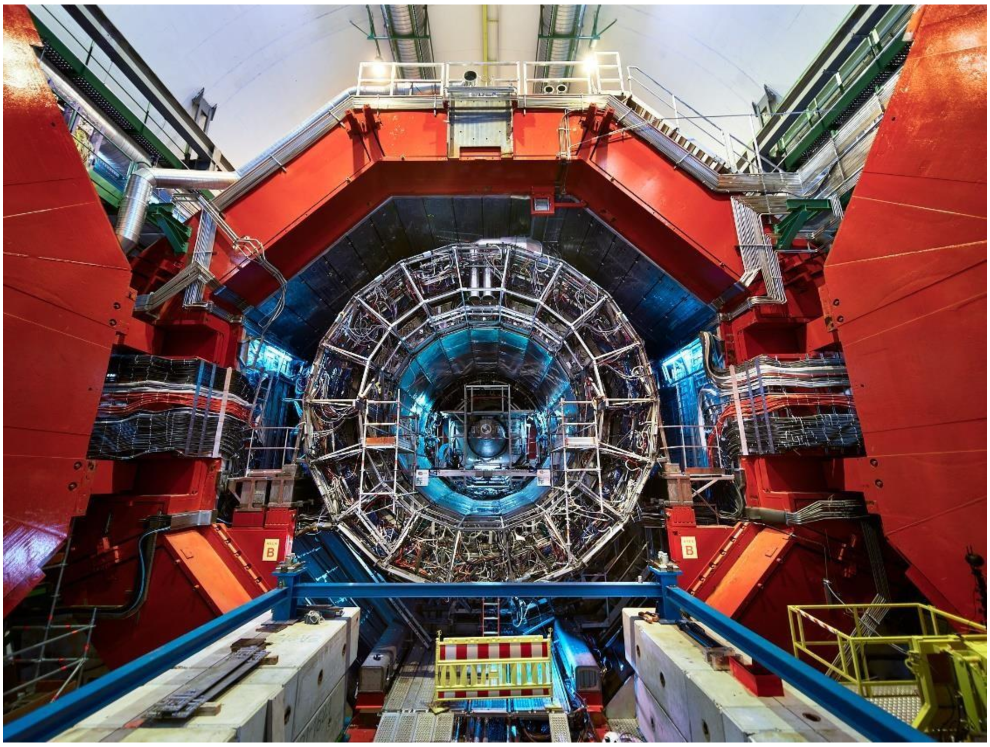
In May 2019, this project was awarded one-year funding under the European Union's ATTRACT initiative. This project, which has the full title of 'Quantum optimisation of Worldwide LHC Computing Grid data placement' is one of 19 ATTRACT projects in which CERN is involved. One of the major challenges faced by this project is the difficulty of defining a suitable description of the data set extracted from monALISA, the monitoring and scheduling tool used by the ALICE experiment for grid operations. We have developed a hybrid quantum reinforcement-learning strategy based on the implementation of a quantum Boltzmann machine using the D-wave quantum annealer simulator available in the D-Wave Ocean software suite. We tested our approach with a simplified use case. In addition, we developed a LSTM (log short-term memory) neural network architecture; this is capable of simulating the monALISA I/O throughput with a high level of accuracy.

Next steps

In 2021, we plan to extend this approach to the optimisation of a linear accelerator beam configuration problem in collaboration with the CERN Beams and Accelerators department. We will compare the performance of our quantum model to a classical reinforcement learning approach.

Presentations

- M. Popa, F. Carminati, S. Vallercosa, F. Fracas, C. Grigoraş, L. BetevM. Carabaş, G. P. Popescu, Quantum Optimization of Worldwide LHC Computing Grid data placement (22 January). Presented at CERN openlab Technical Workshop, Geneva, 2020. *http://cern.ch/go/g7Mk*



Initial work carried out through this project focuses on a specific use case related to the ALICE experiment.

QUANTUM SUPPORT VECTOR MACHINES FOR HIGGS BOSON CLASSIFICATION

R&D TOPIC 4:
QUANTUM TECHNOLOGIES

Project coordinator:
Wisconsin University: Sau-Lan Wu
IBM Zurich: Ivano Tavernelli
CERN: Sofia Vallecorsa, Alberto Di Meglio

Team members:
University of Wisconsin: Chen Zhou, Shaujun San, Wen Guan
IBM Zurich: Panagiotis Barkoutsos, Jennifer Glick

Project goal

This project is investigating the use of quantum support vector machines (QSVMs) for the classification of particle collision events that produce a certain type of decay for the Higgs boson. Specifically, such machines are being used to identify instances where a Higgs boson fluctuates for a very short time into a top quark and a top anti-quark, before decaying into two photons. Understanding this process — known by physicists as ttH production — is challenging, because it is rare: only 1% of Higgs bosons are produced in association with two top quarks and, in addition, the Higgs and the top quarks decay into other particles in many complex ways, or modes.

Background

QSVMs are among the most promising machine-learning algorithms for quantum computers. Initial quantum implementations have already shown performances comparable to their classical counterparts. QSVMs are considered suitable algorithms for early adoption on noisy, near-term quantum-computing devices. Several initiatives are studying and optimising input data representation and training strategies.

We are testing IBM's QSVM algorithm at the ATLAS experiment. Today, identifying ttH-production events relies on classical support vector machines, as well as another machine-learning technique known as 'boosted decision trees'. Classically, these methods are used to improve event selection and background rejection by analysing 47 high-level characteristic features.

Progress in 2020

Different quantum classifiers have been investigated, including QSVM and quantum kernel methods. We compared their performance to classical models in terms of classification accuracy, training dataset size and the number of input features. We studied different types of noise and how they affect the final performance. We also compared the results obtained on IBM quantum hardware for two different Higgs decay modes.

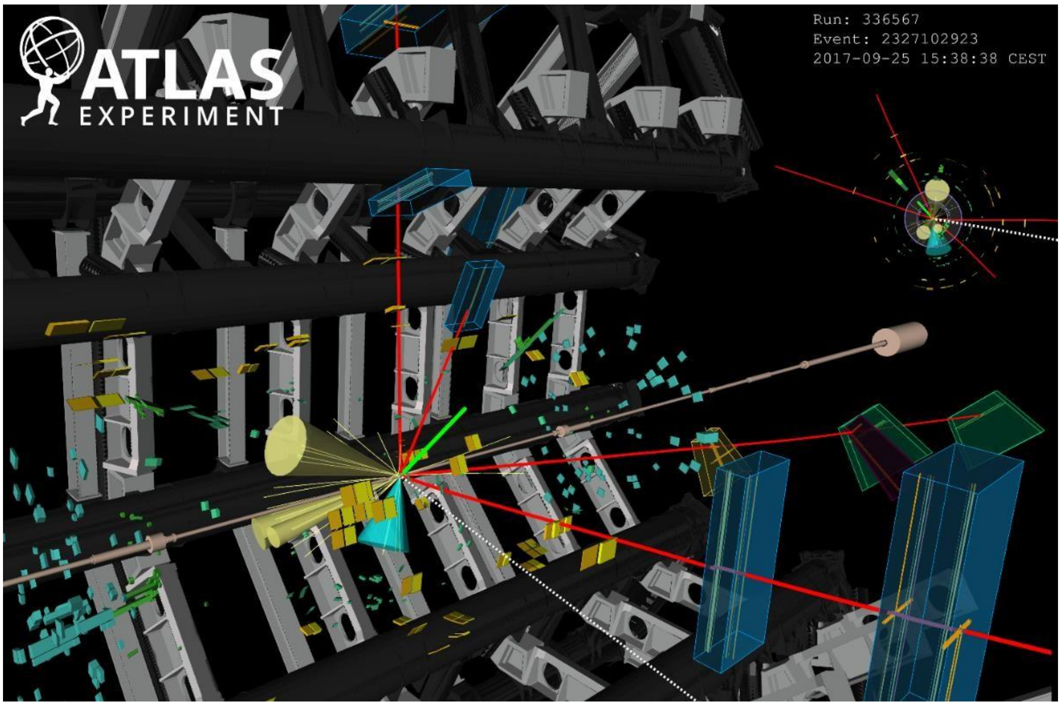
Preliminary results, obtained using the quantum simulator, show that the QSVM can achieve performance that is comparable to its classical counterpart in terms of accuracy, while also being much faster. In addition, a quantum neural network was also developed; its performance was compared with that of the other quantum methodologies.

Next steps

We will continue with the optimisation of the models developed so far. The goal is to improve their performance and increase the size of the problem in terms of number of qubits and training datasets. At the same time, we aim to reduce the classical computing resources needed for the simulation of quantum circuits, which today represent a bottleneck for most studies of this kind. This is a necessary step for creating quantum solutions suitable for more realistic problems.

Presentations

- S. Wu, Application of Quantum Machine Learning to High Energy Physics Analysis at LHC Using Quantum Computer Simulators and Quantum Computer Hardware (4 November). Published at the QuantHEP Seminar, 2020. http://cern.ch/go/wZW7



The project team is testing IBM's QSVM algorithm at the ATLAS experiment.

BIODYNAMO

R&D TOPIC 5:
APPLICATIONS
IN OTHER DISCIPLINES

Project coordinator:
Roman Bauer, Fons Rademakers

Technical team:
Lukas Breitwieser, Jean de Montigny, Ahmad Hesam

Collaborator liaisons:
Uri Nevo, Marco Durante, Vasilis Vavourakis, Klaus-Dieter Oertel

Project goal

We are aiming to create a platform through which life scientists can easily create, run and visualise three-dimensional biological simulations. Built on top of the latest computing technologies, the BioDynaMo platform will enable users to perform simulations of previously unachievable scale and complexity, making it possible to tackle challenging scientific research questions.

Background

Within the life-sciences community, computer simulation is being used more and more to model increasingly complex biological systems. Although many specialised software tools exist, establishing a high-performance, general-purpose platform would be a major step forward. CERN is therefore contributing its deep knowledge in large-scale computing to this collaboration, supported by Intel. Together, we are working to develop a unique platform. This project is co-financed by the CERN budget for knowledge transfer to medical applications.

Progress in 2020

In 2020, we generalised the simulation engine to make it applicable to other application areas beyond cell-based simulation. We used these improvements to implement a simple agent-based SIR model that simulates the spread of infectious diseases. Based on the results of this work, CERN and the University of Geneva started collaborating on a more detailed model for investigating realistic viral spread between individuals. For this purpose, BioDynaMo has been coupled with a fluid-mechanic simulation framework to simulate the precise spread of aerosols and droplets, in addition to agents' behaviour.

Furthermore, we continued our work to improve the performance and usability of BioDynaMo. We completely redesigned the visualisation component and achieved a speedup of up to two orders of magnitude. CERN technology (the ROOT framework) played a central role in this improvement.

Lastly, as part of the CERN openlab summer-student programme, we launched BioDynaMo Notebooks: an interactive web application for rapidly prototyping simulations. BioDynaMo Notebooks are powered by ROOT's fast C++ interpreter and enables researchers to use BioDynaMo through the well-known Jupyter interface.

Next steps

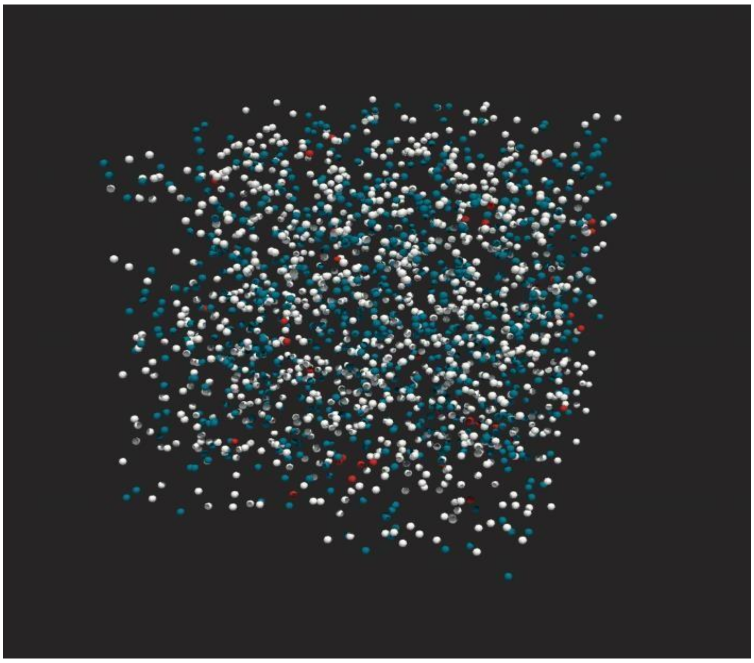
BioDynaMo is currently able to simulate millions of cells on one server. To improve the performance further, we will focus on two aspects.

First, we will continue development on the distributed runtime, to combine the computational resources of many servers. Second, we will improve hardware acceleration to fully utilise (multiple) GPUs in a system. This will not only reduce runtime on high-end systems, but will also benefit users that work on a standard desktop or laptop.

In terms of epidemiological simulation, several virus-spreading scenarios will be investigated in 2021. This will help to determine which conditions are best for avoiding virus build-up and reducing infection. We will add the possibility to optimise model parameters based on various fitting functions, to match realistic scenarios and to extrapolate new findings. This automated optimisation engine will be executable on distributed computing platforms.

Presentations:

- *L. Breitwieser, A. Hesam, The BioDynaMo Project (23 January). Presented at CERN openlab Technical Workshop, Geneva, 2020.*



Large-scale simulation of neural development using BioDynaMo.

CIRCULAR HEALTH

R&D TOPIC 5:
APPLICATIONS
IN OTHER DISCIPLINES

Project coordinator:
Alberto Di Meglio

Technical team:
Anna Ferrari

Collaborator liaisons:
Ilaria Capua, Luca Mantegazza, Elio Borgonovi, Claudio Bellariva

Project goal
The Circular Health project involves a collaboration of research institutes, universities, and not-for-profit organisation led by the One Health Center of Excellence at the University of Florida, US. Together, we are working on the definition, design and implementation of a large-scale open-access data platform to support novel paths to collaborative research for global challenges. Circular Health aims to contribute with both efficient technical tools and methodologies for governance.

Background
Global crises, like the COVID-19 pandemic, have shown the importance of accelerating multi-disciplinary research and easing barriers to locating, aggregating, processing and sharing data and results. Doing so is necessary for addressing complex challenges that involve medical, social, or economic data.

CERN has years of experience in designing large-scale, collaborative data platforms and has developed efficient tools like Zenodo, SWAN and REANA to facilitate sharing, reproducibility and collaboration. Through Circular Health, CERN can contribute to solving critical issues and can support scientific research beyond high-energy physics.

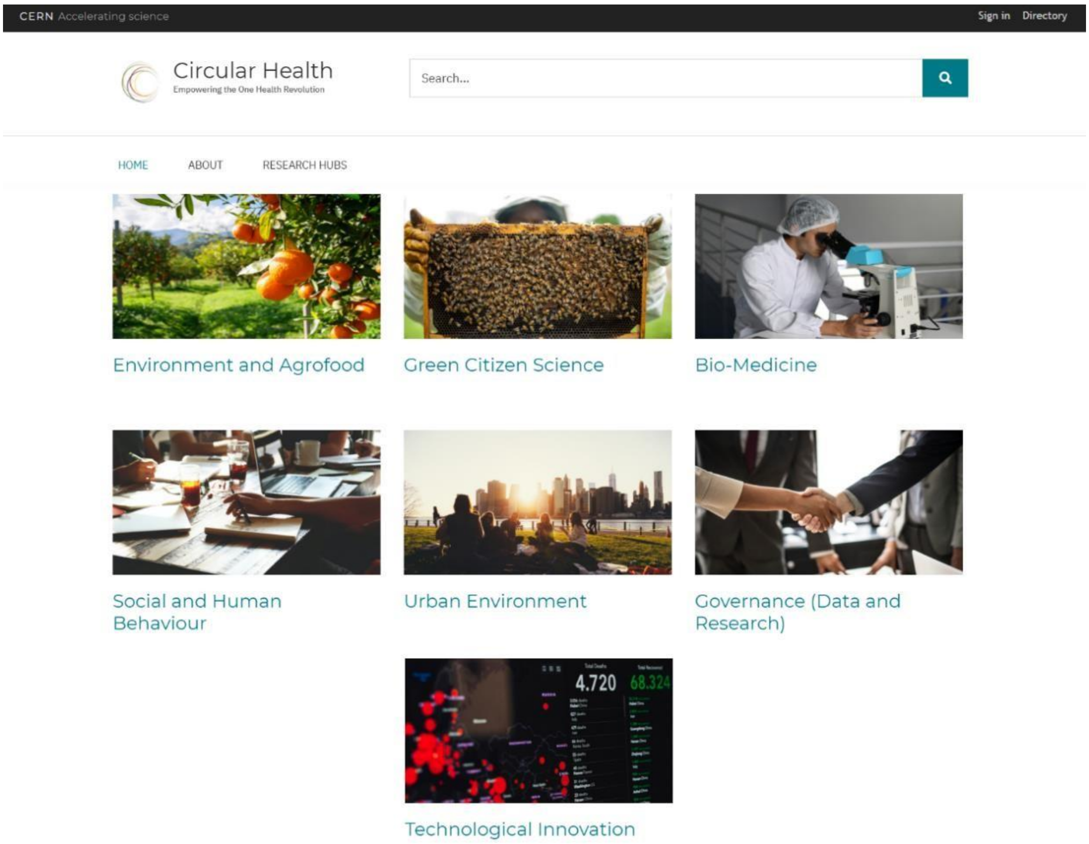
This project is being carried out in the context of CERN's strategy for knowledge transfer to medical applications, led by CERN's Knowledge Transfer group.

Progress in 2020

CERN started its collaboration with Circular Health as part of the CERN Against COVID-19 Task Force, which was established by CERN's management in March 2020. During 2020, experts from CERN openlab and the CERN IT department have collaborated with researchers at the One Health Center of Excellence in Florida, US, as well as at both Bocconi University and Milano-Bicocca University in Milan. Together, we defined the initial requirements for the data and computing infrastructure. We then deployed a first project focused on assessing excess mortality linked to COVID-19 through comparison with data from previous years.

Next steps

At the end of 2020, following our initial investigations, Fabiola Gianotti, the CERN Director-General, supported the creation of a new dedicated project called CERN Science for Open Data (CS4OD). Its mission is to integrate CERN tools and expertise into a platform for supporting international projects. It will use open-access data and will support the CERN-linked UN Sustainable Development Goals.



The research hubs of the Circular Health project.

CERN LIVING LAB

R&D TOPIC 5: APPLICATIONS IN OTHER DISCIPLINES

Project coordinator:
Alberto Di Meglio

Technical team:
Jose Cabrero, Anna Ferrari, Sofia Vallecorsa

Collaborator liaisons:
David Manset (be-studys),
Marco Manca (SCImPULSE)

Project goal

The project goal is to develop a big-data analytics platform and tools for large-scale studies of data under special constraints, such as information that is privacy-sensitive, or that has a varying level of quality, associated provenance information, or signal-to-noise ratio. Ethical considerations are also considered when necessary. This will serve as a proof-of-concept for federating and analysing heterogeneous data from diverse sources, in particular for medical and biological research, using ideas and expertise coming from CERN and the broader high-energy physics community.

Background

CERN is a living laboratory, with several thousand people coming to work at its main campuses every day. For operational purposes, CERN collects data related to health, safety, the environment, and other aspects of daily life at the lab. Creating a platform to collate and enable intelligent management and use of this data — while respecting privacy and other ethical and legal obligations — offers the potential to improve life at the lab. At the same time, such a platform provides an ideal testbed for exploring new data analytics technologies, algorithms and tools, including machine-learning (ML)/deep-learning (DL) methods, encryption schemes, or block-chain-based ledgers. It also provides a natural bridge to collaborate with other scientific research domains, such as medical research and biology.

This project is being carried out in the context of CERN's strategy for knowledge transfer to medical applications, led by CERN's Knowledge Transfer group.

Progress 2020

In 2020, the project activities focused mainly on the investigation of privacy-preserving techniques for data analysis, particularly in cases where machine-learning or deep-learning models are used. A systematisation of the state-of-the-art was conducted looking at different methodologies, such as homomorphic encryption, secure multi-party computation and federated learning. The existing implementations and their capabilities were assessed against reference use cases, including the extraction of features from brain MRI scans and aggregated data classification for epidemiological research. In 2020, two new collaborators joined the initiative: the University of Madrid, Spain, and the Seoul National University Bundang Hospital (SNUBH), South Korea, sharing expertise in security and the analysis of medical data.

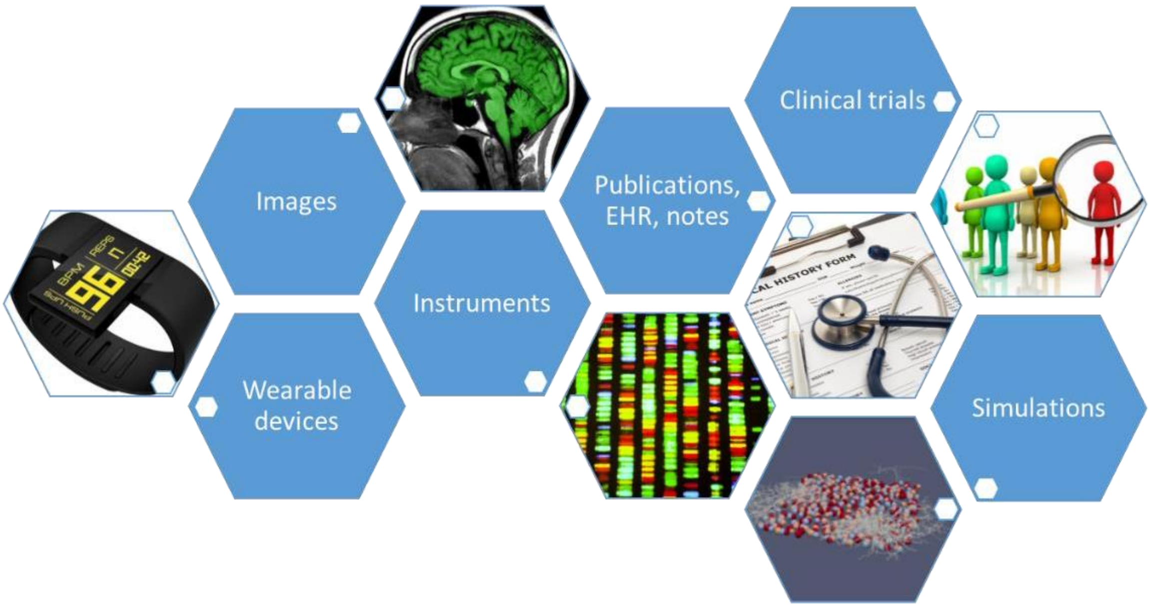
Next steps

After our initial systematisation of knowledge related to privacy-preserving methods, we will begin work to develop one or more methods, integrating them into the ML/DL inference algorithms of the reference use cases. Extension to the full model training process will then be addressed.

In November 2020, CERN openlab entered into a collaboration with the OpenQKD project to assess the use of distribution infrastructures for the quantum keys used for secure analysis of data. The integration of QKD in the data analysis process will be investigated as an additional layer for protecting transactions.

Presentations

- A. Di Meglio, The Living Lab Project (23 January). Presented at CERN openlab Technical Workshop, CERN, Geneva, 2020. http://cern.ch/go/Cf7R



Large-scale medical data requires a new approach to analysis and security.

HUMANITARIAN AI APPLICATIONS FOR SATELLITE IMAGERY

R&D TOPIC 5:
APPLICATIONS
IN OTHER DISCIPLINES

Project coordinator:
Sofia Vallecorsa

Technical team:
Suren Thapa

Collaborator liaisons:
Lars Bromley, Edoardo Nemni

Project goal

This project is making use of expertise in artificial intelligence (AI) technologies at CERN to support a UN agency. Specifically, we are working on AI approaches to help improve object recognition in the satellite imagery created to support humanitarian interventions. Such satellite imagery plays a vital role in helping humanitarian organisations plan and coordinate responses to natural disasters, population migrations, and conflicts.

Background

Since 2002, CERN has hosted UNOSAT, the Operational Satellite Applications Programme of UNITAR (The United Nations Institute for Training and Research) on the laboratory's premises. UNOSAT acquires and processes satellite data to produce and deliver information, analysis, and observations to be used by the UN or national entities for emergency response, to assess the impact of a disaster or a conflict, or to plan sustainable development in the face of climate change.

At the heart of this project lies the idea of developing machine-learning techniques that can help speed up analysis of satellite imagery. For example, predicting and understanding the movement of displaced persons by identifying refugee shelters can be a long, labour-intensive task. This project is working to develop machine-learning techniques that could greatly reduce the amount of time needed to complete such tasks.

Progress in 2020

In 2020, we focused on the challenge of simulating synthetic high-resolution satellite images. High-resolution satellite imagery is often licensed in a way that makes it difficult to share it across UN partners and academic organisations. This reduces the amount of image data available for training deep-learning models, thus hampering research in this area. We have developed a generative adversarial network (GAN) that is capable of generating realistic satellite images of refugee camps. Our tool was initially based on a progressive GAN approach developed by NVIDIA. We have now developed this further, such that it can combine multiple simulated images into a cohesive larger image of roughly 5 million pixels. The new model is built on a multi-network architecture combining several auto-encoders; their output is used to condition the image-generation step and to ensure each new image is consistent with previous ones. This method was tested on satellite images of a flooded area in Myanmar.

Next steps

Next year will be dedicated to the optimisation of the progressive GAN model. In particular, we will implement a distributed approach for training our network in parallel across multiple nodes. This should help us to reduce training time for the model and increase the maximum image size.

This project is working to develop machine-learning techniques to aid understanding and predicting of the movement of displaced persons.

SMARTANOMALY

R&D TOPIC 5:
APPLICATIONS
IN OTHER DISCIPLINES

Project coordinator:
Alberto Di Meglio

Technical team:
Yann Donon

Project goal

The SmartANOMALY project is an evolution and broadening of the SmartLINAC project, which launched in June 2019. The main goal of the original project was to create a platform for anomaly detection and maintenance planning for linear accelerators, which are used widely in medicine and high-energy physics research.

Background

Technologies related to artificial intelligence (AI) are opening up new possibilities for anomaly detection. Given the array of large particle accelerators at CERN, the Organization has significant expertise in detecting anomalies in highly complex systems. This expertise has the potential to be applied to a range of scientific and industrial activities — including (but not limited to) other fields where particle accelerators are used, such as medicine. This project has been supported by CERN's Knowledge Transfer group.

Progress in 2020

After more than a year of development, promising results were achieved, demonstrating the potential of our innovative algorithms for detecting anomalies — as well as perhaps even *predicting* their effects to some extent. Today, the project's primary focus is on medical accelerators. However, we see potential in training our solution on more sources, such as on compressor engines or complex industrial processes.

It is common practice to use alternative data sets when training anomaly-detection systems. Therefore, the distinguishing aspect of our research is that several approaches, based on statistics and neural-network technologies, are being combined in order to offer a system that can be adapted to different sources.

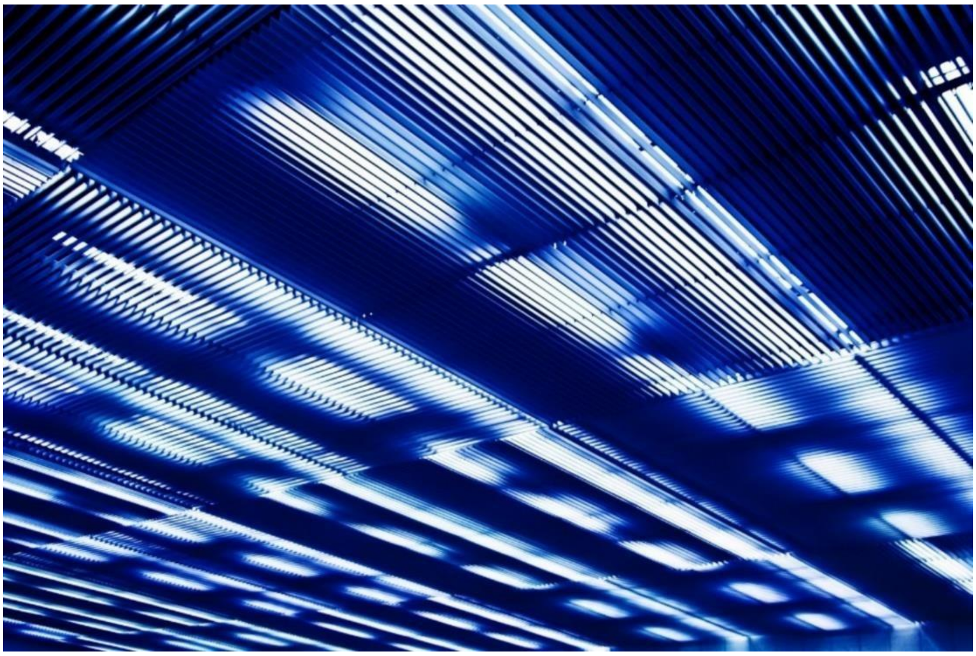
Given that demand for such tools is growing rapidly, we believe the time is right to formally enlarge the scope of the research started through SmartLINAC. Thus, we have created our new, broader SmartANOMALY project.

Next steps

This new project incorporates our existing investigations with linear accelerators, and will also allow new actors to take part in the development of the anomaly detection tool for complex systems. We are currently discussing possible applications in the automotive and food-processing industries. Given that our research is now entering a new phase, we encourage actors from industry and academia to get in touch with us. We are keen both to develop the existing activities within this project and to explore new opportunities for enlarging its scope.

Presentations

- Y. Donon, Anomaly detection in noised time series: the challenge of CERN's LINAC4 (24 January). Presented at The Open Data science meetup #3, Samara, 2020. cern.ch/go/9PZD



The SmartANOMALY project is an evolution and broadening of the SmartLINAC project.

SMARTANOMALY SPIKEFALL

R&D TOPIC 5:
APPLICATIONS
IN OTHER DISCIPLINES

Project coordinator:
Sofia Vallecorsa

Technical team:
Yann Donon

Collaborator liaisons:
Sauro Succi, Walter Rocchia, Nicola Scafuri

Project goal

This project was launched in the context of the fight against the COVID-19 pandemic, as a CERN openlab collaboration with the Italian Institute of Technology and CompBioMed. Through this project, we aim to propose a model based on machine learning for simulating the enhanced molecular dynamics of SARS-CoV-2's spike glycoprotein. Proteins' qualities can be described by their secondary structures, three-dimensional forms made of atoms groups under which protein residues (individual amino acid) can exist. The goal of SmartANOMALY Spikefall is to predict this change in structure by analysing just a few moments, or 'frames'.

Background

The energy cost for simulating proteins is high. As such, a new approach for predicting the qualities of proteins would be a very powerful tool for the research community. Such an approach could be a particular boon to the fight against the COVID-19 pandemic, with researchers seeking to accelerate their investigations.

This project is being carried out in the context of CERN's strategy for knowledge transfer to medical applications, led by CERN's Knowledge Transfer group.

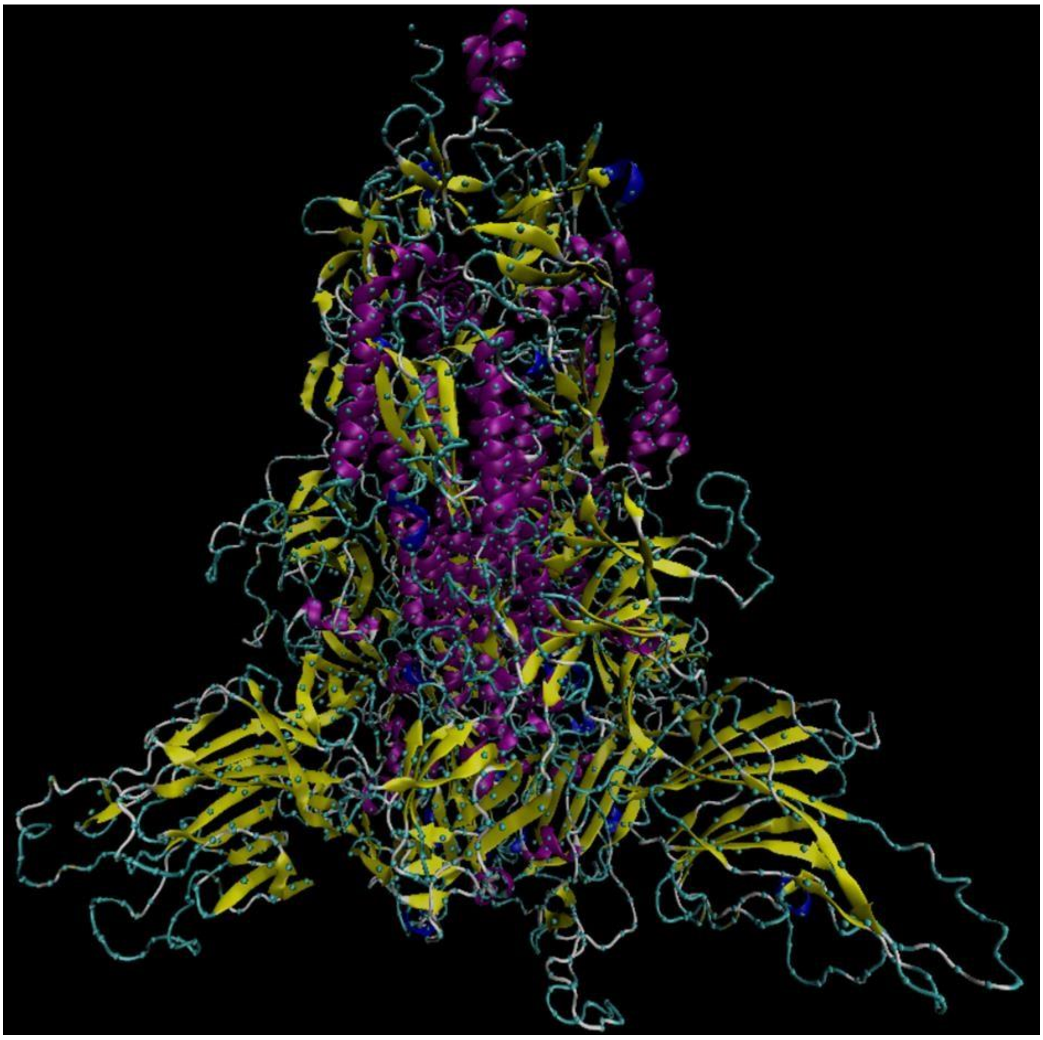
Progress in 2020

The project started in the fourth quarter of 2020, with several approaches for classification and behaviour prediction tested immediately. Currently, we are studying the behaviour of single atoms — in particular, alpha carbons — over a short period of time. Our first observations have led us to believe that understanding the instability of an atom can provide us with insight into the long-term behaviour of a proteins' secondary structure.

Next steps

Our first results are promising, encouraging us to investigate further. Rather than continuing to look at single atoms, we will now studying the behaviour of proteins' secondary structures in full, increasing the complexity and precision of results.

Following our first steps in 2020, we will now work to expand the project collaboration, with a view to helping us achieve the full potential of this work.



SARS-CoV-2 spike glycoprotein representation with its different secondary structures, and alpha

EARLY DETECTION of PARKINSONs DISEASE

R&D TOPIC 5:
APPLICATIONS
IN OTHER DISCIPLINES

UNIVERSITA' DEGLI STUDI DI MILANO
BICOCCA

Project coordinator:
Alberto Di Meglio, Sofia Vallecorsa

Technical team:
Anna Ferrari

Collaborator liaisons:
Daniela Micucci, Paolo Napoletano

Project goal

Our work will be organised into three main areas:

1) Identification of one or more suitable datasets from existing public data providers (e.g. the Michel J. Fox Foundation).
2) Implementation of a supervised learning strategy to analyse labelled data and classify patients affected by Parkinson's disease.
3) Implementation of an unsupervised learning strategy to detect and diagnose potential Parkinson's disease symptoms using anomaly-detection algorithms or other suitable approaches. This will then correlate relevant features (e.g. duration and intensity of the symptoms) with medical treatments and other factors.

Background

CERN openlab is currently running a project called CERN LivingLab to set up a distributed data-analysis platform providing specialised features to process data with sensitive content, such as personal or medical information. The platform is intended to be a technology demonstrator and a testbed for state-of-the-art functionalities, including advanced machine learning and deep-learning tools and algorithms, secure data transmission and storage, and encryption techniques.

This project is being carried out in the context of CERN's strategy for knowledge transfer to medical applications, led by CERN's Knowledge Transfer group.

Progress in 2020

Two public datasets released by the Michael J. Fox Foundation have been used for analysis. Both presented inertial data recorded from one smart-phone and one smart watch. Data was recorded during everyday life and was manually labelled by the user. The availability of labelled data enabled the use of supervised deep-learning and machine-learning techniques. Different kinds of strategies were selected and implemented: convolutional neural networks were applied to time series and extracted images.

The dearth of data led to insufficient performance in terms of accuracy. Furthermore, the high variability of the data meant an ad-hoc pre-processing phase was required, followed by an additional feature-extraction procedure. Both were implemented and used to feed traditional machine-learning algorithms, which then performed better than the analysis based on deep learning.

Next steps

The pre-processing and feature-extraction procedures are crucial steps for the reliability and sustainability of the results. We will work to improve these two procedures in 2021, enhancing the performance of the machine-learning algorithm.

Presentations:

- A. Ferrari, Deep Learning Analysis on Wearable Devices (23 January). Presented at CERN openlab Technical Workshop, Geneva, 2020. http://cern.ch/go/bKR6

CERN LivingLab is working to set up a distributed data-analysis platform for processing personal or medical information, as well as other kinds of sensitive data.

04
KNOWLEDGE

Education, training and outreach

CERN openlab is designed to create and share knowledge through a wide range of activities and programmes.

CERN openlab is a knowledge factory. We work to disseminate this knowledge through both outreach activities and educational programmes. As well as promoting our technical work among a variety of stakeholders, we are working to train the next generation of ICT specialists. Thus, CERN openlab provides a means for its collaboration members to share a joint vision of the future of scientific computing. This vision is communicated to a wide audience, including partner clients, policy makers, members of the press, and the general public. Together, we can shape the future of scientific computing for the benefit of both research and wider society.

Events move online due to pandemic

The COVID-19 pandemic resulted in many training and outreach activities moving online in 2020. It was also not possible to hold the usual CERN openlab summer-student programme. Nevertheless, online mentorship was provided for 16 of the students whose planned projects were deemed feasible for running remotely.

Recordings of previous summer-student lectures were also made available via a new page on the CERN openlab website: *https://openlab.cern/online-learning*. These lectures address topics such as machine learning, quantum computing, high-performance computing, and much more. Plus, four new online tutorials were organised, covering computer security, monitoring, parallel programming, and high-performance computing on GPUs.

In addition, two live Q&A sessions were streamed in July. These provided the students with the opportunity to ask their questions to leading ICT experts at CERN. Both of these sessions were recorded and are also now available via the above webpage.

Intel oneAPI

Intel sponsored six students to work on projects related to their oneAPI technology in 2020. In October and November, the students blogged about these projects via an Intel platform. Short videos were also created in which the students outline how they're investigating Intel OneAPI for several possible use cases across the laboratory. All materials can be found here: *http://cern.ch/go/tx7M.*

Following on from this, Intel launched a large, public competition around oneAPI in late November. Entrants were asked to demonstrate new and interesting use cases for this technology. Eugenio Marinelli, a PhD student from EURECOM in France was selected as the overall winner. He has been sponsored by Intel to participate in an internship with CERN openlab in 2021. Intel has also paid for flights and accommodation for three other competition winners to come to CERN for special guided tours. Should these physical visits not be possible due to COVID-19 restrictions, alternative prizes will be offered.

Reaching out around the world

Throughout 2020, we made a concerted effort to connect with people in their home countries, sharing our work — and our passion — with them.

Online workshops, training sessions and other meetings were organised online throughout the year. One of the highlights of 2020, was the CERN Webfest, our annual hackathon, which became an online event for the first time ever. Our report on this event is republished in full across the next two pages.

On 24 September, the interns who had worked remotely on projects throughout the summer gave short presentations summarising their investigations.

CERN Webfest goes online... and global!

The 'CERN Webfest' — CERN's annual hackathon based on open web technologies — was held on the weekend of 27-28 June. Due to the COVID-19 pandemic, the event was held online for the first time. Over 400 people signed up for the event from 75 countries across the globe.

Held each year since 2012, the Webfest brings together bright minds to work on creative projects. Participants work in small teams, often designing web and mobile applications that help people engage with CERN's research, physics, or even science in general.

The theme for this year's event was 'working together apart: accelerating collaboration'. Given the global COVID-19 crisis, the organisers were particularly keen to see projects that address the evolving ways in which we work together. Building on CERN's strong history of international collaboration, the Webfest provided an excellent opportunity to create tools to support the changing ways in which we do science.

Examples of projects developed over the course of the weekend include the following: an online detective-themed science show for school children, a web library of LaTeX equations, an app to assist with urban planning at large research centres, a platform for sharing remote access to lab equipment, a learning-management system, and a machine-learning program to help prevent social-media content from exacerbating depression. The projects listed above are just the six shortlisted projects from over 30 submitted. "The CERN Webfest was the best experience of my life," says Ashraful Alam Khan, one of the two final-year high-school students from Bangladesh behind this last project. "It was fantastic to be selected as one of the finalists".

Over 400 people signed up for the event from 75 countries.

Naturally, many of these projects are still at an early stage of development. However, guidance was offered to the participants about how they could take their projects further over the coming weeks and months. The Webfest organisers will be following all of the projects with interest.

A panel of 12 judges, representing a range of organisations, selected an overall winner from the event. They picked a team that is developing a platform to digitise electrocardiograms of COVID-19 patients, with a view to helping medical researchers better analyse this data. "The Webfest provided us with an opportunity to inspire others and to be inspired ourselves; the event helped us to believe that — through our hard work and dedication — we can work together to change the world for the better," says Sina Khezri, project leader and a medical student in Iran. His teammate Arjeta Semani, from Kosovo, agrees, saying: "It was a great experience that will last in our memories for a long, long time." The other members of this team were Ayush Parhi from India and Ali Fele Paranj, also from Iran.

"Selecting the best projects was not an easy task for us judges, as the amount of creativity, innovation and originality offered by the Webfest 2020 participants was fantastic," says Jean-Pierre Reymond, Chargé de Mission, Head of Innovation Partnerships at the Permanent Mission of Switzerland to the United Nations, Geneva. "The event was a great demonstration of how collective intelligence can be productive and cooperative."

"It was a privilege being part of the jury for this year's Webfest," says Charlotte Warakaulle, CERN Director for International Relations. "The quality of the projects was impressive. The participants demonstrated great creativity and understanding of the broader societal challenges that we are trying to address together, all of them showing how much we can really achieve when we work across boundaries for common goals. Any good idea needs to be nourished through exchange to be turned into truly transformative action."

While the projects worked on formed the core of the Webfest, there was so much more to this year's event. Six workshops were held across the weekend, during which experts from CERN shared important skills and insights with the participants. Highlights included a talk from CERN's head of computer security, Stefan Lüders, on staying safe online, and a talk from Manuella Cirilli of CERN's Knowledge Transfer group on how to give good presentations. Members of CERN's yoga and fitness clubs also provided online exercise sessions, and an online DJ set was organised for the close of the event.

Altogether, this made for a rich programme spanning the entire weekend. "I really enjoyed taking part in the Webfest," says Sarah Catherine Johnston, from the University of St Andrews in Scotland, UK. "I had never taken part in a hackathon before, and it wasn't quite like I had imagined, but only in the sense that it was a lot more fun and flexible than I thought it would be!"

The CERN Webfest was organised by members of CERN openlab and gluoNNet, an evidence-based analysis provider founded by physicists from CERN. Additional support was provided by members of CERN's HR department; IR sector; THE Port, which organises an annual humanitarian hackathon at CERN's IdeaSquare; and Remotely Green, a start-up company created by CERN users that specialises in virtual networking.

The hackathon was managed using a custom-built application; this was used to select projects, form teams, and communicate with the participants throughout the event. It was very important to connect this app — originally developed for the recent VersusVirus hackathon — to existing tools used at CERN, such as Indico and Mattermost, explain Daniel Dobos and Karolos Potamianos of gluoNNet.

"It was a memorable time for all the participants, including the organisers and judges," says Ben Segal, Webfest co-founder, who also helped to organise this year's event. "I saw lots of talent, learning, energy and engagement displayed from all corners of the globe. And in the true Webfest spirit, everyone had fun!"

A screenshot of the event on Zoom. The organisers worked remotely, as well as the participants.

ALBERTO DI MEGLIO
Head of CERN openlab
Alberto.Di.Meglio@cern.ch

MARIA GIRONE
Chief technology officer
Maria.Girone@cern.ch

FONS RADEMAKERS
Chief research officer
Fons.Rademakers@cern.ch

SOFIA VALLECORSA
Quantum and AI Research Lead
Sofia.vallecorsa@cern.ch

ANDREW PURCELL
Communications officer
Andrew.Purcell@cern.ch

KRISTINA GUNNE
Administrative / finance officer
Kristina.Gunne@cern.ch

CERN openlab 2020 Annual Report

Editor: Andrew Purcell

Additional subediting: Anastasiia Lazuka and Christina Bolanou

All images provided by the CERN Audio-Visual Production Service except:
Veronika McQuade (page 2); CERN openlab communications office (pages 14, 23, 37, 69, 71, 73); DEEP Projects (page 21); CNES (page 63).

ISBN 978-92-9083-621-6
Published by CERN

CERN openlab