ISBN: 978-3-95450-231-8 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2023-WEPA108

DATA INTENSIVE SCIENCE AND PARTICLE ACCELERATORS: DRIVING SCIENCE AND INNOVATION*

C. P. Welsch†, University of Liverpool, Liverpool, UK

Abstract

Particle accelerators and light sources are some of the largest, most data intensive, and most complex scientific systems. The connections and relations between machine subsystems are complicated and often nonlinear with system dynamics involving large parameter spaces that evolve over multiple relevant time scales and accelerator systems.

Data Intensive Science offers exciting prospects for accelerator design and operation. This includes the optimization of machine design and the reconstruction of transverse beam distributions using machine learning, as well as data analysis in high data rate monitors. This paper presents the new Liverpool Center for Doctoral Training for Innovation in Data Intensive Science (LIV.INNO) and its exciting research and training program.

INTRODUCTION

The Liverpool Center for Doctoral Training (CDT) for Innovation in Data Intensive Science (LIV.INNO) was established in 2022 to provide training for three cohorts of PhD students through interdisciplinary research projects [1]. The center offers PhD projects across nuclear and particle physics, accelerator science, astrophysics, mathematics and computer sciences. It hence covers the entire breath of STFC-funded research. The connecting element between all research projects is that they have demanding data science-related challenges in their core. The CDT is hosted between the University of Liverpool and Liverpool John Mores University in the North-West of England and offers a structured academic training program to around a dozen students every year, complemented by secondments to national and international research partners with strong industry contributions.

LIV.INNO is directly based on the partners' previous highly successful STFC-funded CDT LIV.DAT which has trained around 40 PhD students since 2017 [2]. More generally, the approach to postgraduate training is based on the concepts developed through the coordination of a large number of Marie Skłodowska-Curie Actions training networks in accelerator science coordinated by the University of Liverpool since 2008.

More than 100 experts joined the LIV.INNO kick-off meeting on Tuesday, 24 May 2022, see Fig. 1 [3]. The event included presentations on all aspects of the center, talks by industry and research partners, and a poster session showcasing research outcomes from the previous center, LIV.DAT.

* This work was supported by the Science and Technology Facilities Council (STFC) under grant agreement ST/W006766/1.
† c.p.welsch@liverpool.ac.uk

Figure 1: Participants at the LIV.INNO kick-off meeting in the Spine, Liverpool, UK on 24 May 2022.

The kick-off meeting discussed the aims of the centre and its ambitious R&D plans in detail. It also gave existing and prospective partners a platform to present their data science capabilities and research interests. The event was an ideal opportunity to discuss collaboration and future joint projects. A special treat for all participants was the premiere of the film "LIV.INNO – driving innovation through data science" [4].

RESEARCH

Managing, analyzing and interpreting large, complex datasets and high rates of data flow is a growing challenge for many areas of science and industry. Recent years have witnessed a dramatic increase of data in many fields of science and engineering, due to the advancement of sensors, mobile devices, biotechnology, digital communication and internet applications. Very little targeted training is provided internationally, and in the UK in particular, to address a growing skills gap in this area.

R&D in LIV.INNO is structured across the following three main scientific work packages: Monte Carlo methods and High Performance Computing, Artificial Intelligence and Machine Learning, and Data Analysis.

- **Work package 1: Monte Carlo (MC) methods and High Performance Computing (HPC)** are powerful tools for everything from modelling the birth and evolution of the universe to performing numerical integrals for calculating cross sections for particle interactions;
- **Work package 2: Artificial Intelligence (AI) and Machine Learning (ML).** Optimization techniques that can exploit deep structures and race-tuned implementation of Deep Learning on GPUs have resulted in pervasive and successful application of ML across the Big Data arena; Liverpool researchers are at the forefront of this research;

ISBN: 978-3-95450-231-8 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2023-WEPA108

- **Work package 3: Data Analysis.** Projects have students working on both, STFC core physics topics and wider applications. This connects their work to the R&D basis generated in work package 2 and expands it to the problem of combining data from different sources and providing an efficient analysis.

All three work packages are very relevant for accelerator science R&D. Beam control and manipulation, beam dynamics studies and detailed analysis of beam diagnostics output data all directly benefit performance enhancements. This is true for linear and circular machines, as well as for various particle species and beam energies, i.e. the results within LIV.INNO are highly relevant for a large number of projects within accelerator science. The following examples briefly describe several specific accelerator R&D projects in the center.

Reconstruction of Transverse Beam Distribution using Machine Learning

The beam transverse distribution in CERN's high-radiation environment is measured by imaging the light generated by the particle beam hitting a scintillating screen, using cameras produced in-house based on radiation hard tubes. Due to the cessation of radiation-hard tube production worldwide, CERN is investigating the transport of the beam image to low radiation areas using radiation tolerant optical fibers coupled to normal CMOS cameras [5].

In this framework, pioneering work to reconstruct the beam's transverse distribution using a single large-core multimode optical fiber began in 2020. The method takes advantage of advances in generative modeling using deep learning methods, such as convolutional neural networks, and attempts to apply them to beam diagnostics [6]. Specifically, conditional adversarial networks for image-to-image [7] translation have been trained to translate the output plane of the fiber, imaged on a CMOS camera, into the image of the beam on the scintillating screen, optically relayed to the input plane of the fiber.

To date, this proof-of-principle work has evolved in two steps. On the theoretical side, a simulated dataset has been created from the image propagation in a simplified model of an optical fiber using a commercial ray-tracing software; it has been used to train the networks to study the feasibility of the technique. In parallel, an experimental setup was installed at CERN's CLEAR facility and, benefiting from dedicated machine time, allowed to validate the technique by evaluating its potential and establishing a roadmap for improvements.

A PhD research project starting later in 2023 will optimize the fiber optic parameters and perform a wider survey of available radiation-resistant fibers. The student will complete a screening of available networks to be used for image translation, in particular the convolutional U-Net [8], widely used for biomedical image segmentation and the already used GANs. On this basis, they will build a machine learning model using simulated datasets and evaluate its performance. An experimental setup will then be established to validate the simulations, before measurements will be done at CERN's CHARM and CLEAR facilities to validate the results.

Optimization of 3D X-ray Imaging Systems

X-rays are a key imaging modality used in healthcare, and an important application of accelerator technology. Digital tomosynthesis (DT) allows 3D imaging by using a ~30° range of projections instead of a full circle as in computed tomography (CT). The technique promises patient doses ~10 times lower than CT and similar to 2D radiography, with a diagnostic ability that is significantly better than 2D radiography and similar to that of CT. In addition, cold-cathode field emission technology allows the integration of tens of X-ray sources into source arrays that are smaller and lighter than conventional X-ray tubes. The distributed source positions avoid the need for source movements. Adaptix Ltd. [9] has developed a product with the potential to change how DT is used in everyday clinical settings. They have created a 'desktop' DT device, which uses a 2D array of electron emitters etched upon a silicon layer [10] to fire electrons on to an x-ray target which, upon filtration and collimation, creates x-rays that are incident on a flat-panel detector that allows 3D reconstruction of the object imaged. This is currently on the market for voxel and small animal imaging, and work is undergoing to scale this up to a multi-panel system for chest DT scans.

The PhD project of Lauryn Eley investigates this scaling-up. A primary consideration here is the x-ray beam generation - namely, the part of the device that includes the x-ray target and filtration. Monte-Carlo simulations in Geant4 have proved an effective method of undertaking such investigations. An example is shown in Fig. 2.

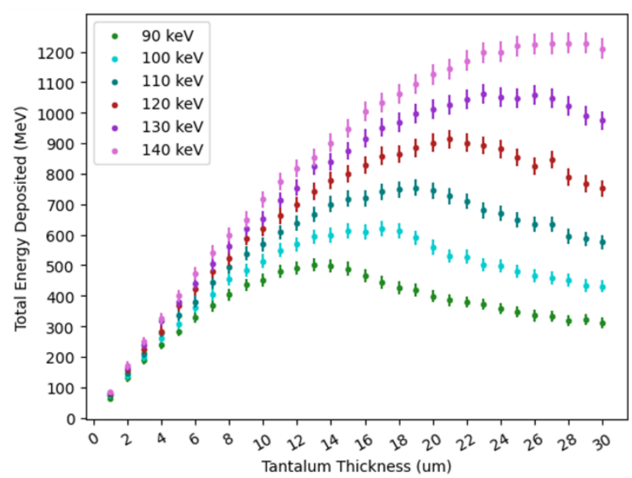

Figure 2: Geant4 simulation outputs showing how the total energy deposited on to a virtual detector is affected by the x-ray target thickness at different incident electron beam energies.

The plot shows results of a Geant4 simulation in which 10^7 electrons are incident on an x-ray target made of tantalum, and the total energy deposited on the target is found as a method of modelling patient dosage. In this simulation, the only objects created are the tantalum target, and a virtual detector, to allow the effect of the target geometry to be

ISBN: 978-3-95450-231-8 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2023-WEPA108

investigated in isolation. The incident electron beam is set to be Gaussian in shape, to imitate the shape of electrons emitted from the semiconductor surface in Adaptix' product. The electron energies were altered over a range realistic for clinical use. It can be seen in the plot that the increased number of atoms at higher thickness initially causes higher x-ray production, due to increased bremsstrahlung and atomic deexcitation interactions [11]. However, the increased target thickness also causes attenuation of the photons themselves in the target, causing the plots to take the shape seen in the figure. Additional Monte Carlo investigations will look into how other electron beam parameters affect the x-rays produced, such as the beam shape and size to help optimize this part of the x-ray device for chest scan applications.

Longitudinal Density Monitor for the Large Hadron Collider

At the Large Hadron Collider (LHC), ion bunches circulate in two counter-rotating beams and are brought into collision. Each bunch is confined within a bucket by the longitudinal focusing effect of the radio frequency (RF) cavities. The RF period is 2.5 ns, while the minimum bunch spacing is 25 ns. 9 out of every 10 buckets should be empty, as well as additional gaps to allow for the rise-time of injection and dump kickers. In practice, however, small numbers of particles can occupy these supposedly empty buckets, causing problems for machine protection and for the absolute calibration of the LHC's luminosity.

The Longitudinal Density Monitor (LDM) is a key diagnostic in the LHC for particle physics experiments and used to measure the profile with very good resolution [12]. This monitor produces a vast amount of data used for luminosity calibration and machine control. An example measurement is shown in Fig. 3.

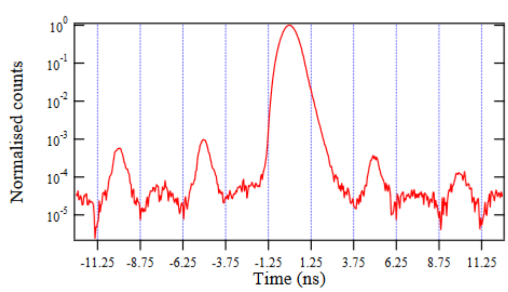

Figure 3: Example measurements with the LDM.

The PhD project of Alex Jury focuses on the development of machine learning-based approaches to analyze this data and develop a novel tool for precise real-time luminosity measurements. This includes techniques to correct for bias associated with different monitor types, beam distributions and beam-beam effects. Moreover, the impact of pileup effects is being assessed and mitigation strategies are being developed.

TRAINING

LIV.INNO provides targeted training to meet a growing skills gap in the area of data science. The training program has been developed to create future leaders in data intensive science. The CDT's students undertake training on a cohort basis, with the opportunity to personalize their program. This is primarily training through research, in addition to a bespoke program consisting of structured lectures, seminars, workshops, and schools.

The students complete 120 credits worth of training in data science during their first year. This is delivered by academic staff from both partner universities and supported by external partners, potential employers and users in data science. In years 2-4, the students continue to develop skills and expertise in response to their Development Needs Analysis (DNA), whilst engaging with the various international events organized by the CDT. This includes placements at identified businesses and research institutions in areas that fall outside the core of their PhD to broaden their skills and enhance their employability, and to apply STFC data skills to help address real-world challenges.

In terms of wider careers skills, the LIV.INNO training program is designed to address important areas such as project management, networking, communication and presentation skills, with the aim to provide all students with the skills set required for a successful career in either academia or industry. To this end, an introductory skills school for the current first year LIV.INNO students as well as for the other first year PhD students in the Department of Physics was run from 3-6 April 2023 in Liverpool, UK. The school was based on successful training developed over the past decade for pan-European training programs that have demonstrated that they boosted the skills of course participants. It included sessions on presentation skills, scientific outreach, intellectual property rights, project management, peer review and science writing. The outreach ideas developed during the week will be used during Daresbury Open Week which is a large-scale public engagement event that will run from 10 -15 July 2023. This initial school was important to establish close links within the cohort and considered very useful by participants. It will next be offered again in November 2023 to the next cohort of LIV.INNO students, year 1 PhD students in the Department of Physics, as well as to the Fellows within the new EuPRAXIA Doctoral Network [13].

SUMMARY AND OUTLOOK

The LIV.INNO Center for Doctoral Training started in October 2022 with a first cohort of around a dozen PhD students. It provides its students with an excellent framework to address cutting-edge research challenges across different disciplines, and a comprehensive training in data science and wider careers skills. The network will offer a number of international events that will be open to the wider scientific community and thus help address a growing shortage of experts in this important area.

ISBN: 978-3-95450-231-8 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2023-WEPA108

REFERENCES

[1] LIV.INNO home page, `http://www.livinno.org`

[2] C. P. Welsch, "Using Data Intensive Science for Accelerator Optimization", in *Proc. IPAC'22*, Bangkok, Thailand, Jun. 2022, pp. 980-982. `doi:10.18429/JACoW-IPAC2022-TUPOST051`

[3] LIV.INNO Kickoff meeting event page, `https://indico.ph.liv.ac.uk/event/593/`

[4] LIV.INNO – Driving Innovation through Data Science, `https://youtu.be/ZzVnNcPyNfA`

[5] D. Celeste, E. Bravin, S. Burger, and F. Roncarolo, "Development and Test of a Beam Imaging System Based on Radiation Tolerant Optical Fiber Bundles", in *Proc. IPAC'19*, Melbourne, Australia, May 2019, pp. 2658-2660. `doi:10.18429/JACoW-IPAC2019-WEPGW078`

[6] E. Kakkava, *et al.*, "Imaging through multimode fibers using deep learning: The effects of intensity versus holographic recording of the speckle pattern", *Optical Fiber Technology* 52 (2019) 101985. `doi:10.1016/j.yofte.2019.101985`

[7] P. Isola *et al.*, "Image-to-Image Translation with Conditional Adversarial Networks" (2018). `arXiv: 1611.07004`

[8] O. Ronneberger *et al.*, "U-Net:Convolutional Networks for Biomedical Image Segmentation", *Springer Lecture Series* (2015). `doi: 10.1007/978-3-319-24574-4_28`

[9] Adaptix home page, `http://www.adaptix.com`

[10] T. G. Primidis *et al.*, "3D chest tomosynthesis using a stationary flat panel source array and a stationary detector: a Monte Carlo proof of concept", *Biomed. Phys. Eng. Express* 8 (2021). `doi:10.1088/2057-1976/ac3880`

[11] M. M. Nasseri, "Determination of tungsten target parameters for transmission X-ray tube: A simulation study using Geant4", *Nuclear Engineering and Technology*, 48(3):795–798, 2016. `doi:10.1016/j.net.2016.01.006`

[12] A. Jeff, *et al.*, "Longitudinal density monitor for the LHC", *Phys. Rev. ST Accel. Beams* 15, 032803 (2012). `doi: 10.1103/PhysRevSTAB.15.032803`

[13] C. Welsch, "EuPRAXIA Doctoral Network", presented at the IPAC'23, Venice, Italy, May 2023, paper TUPA109, this conference.