ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOST050

LIVERPOOL CENTRE FOR DOCTORAL TRAINING FOR INNOVATION IN DATA INTENSIVE SCIENCE*

C. P. Welsch†, University of Liverpool/Cockcroft Institute, UK

Abstract

The Liverpool center for doctoral training for innovation in data intensive science (LIV.INNO) is an inclusive hub for training three cohorts of students in data intensive science. Starting in October 2022, each year will train about 12 PhD students in applying data skills to address cutting edge research challenges across astrophysics, nuclear, theoretical and particle physics, as well as accelerator science. This framework is expected to provide an ideal basis for driving science and innovation, as well as boosting the employability of the LIV.INNO PhD students.

This paper gives examples of the accelerator science R&D projects in the center. It includes details about research into the optimization of 3D imaging techniques and the characterization of photocathodes for accelerator applications.

INTRODUCTION

Analysis and manipulation of Big Data has long been a specialty of researchers in STFC-funded research areas. The new Center for Doctoral Training (CDT) LIV.INNO will provide training for three cohorts of PhD students through an interdisciplinary approach [1]. At the University of Liverpool (ULIV) and Liverpool John Moores University (LJMU), we have an outstanding track record in the design, construction and operation of scientific instruments and the science they enable. This had led to strong links with many research laboratories and industry partners. The CDT will capitalize on the wider research and training activities at both partner universities where data science is an identified priority research theme. Amongst others, the CDT will benefit from a range of existing initiatives, such as the Virtual Engineering Centre (VEC), the strategic partnership between UoL and STFC Daresbury, as well as the new £12M Digital Innovation Facility (DIF) .

Liverpool researchers have an excellent track record in delivering science for STFC: At UoL the accelerator science group based at the Cockcroft Institute, particle physics and nuclear physics groups have the highest rate of STFC funding per academic in the UK, attracting more than £70M of funding from STFC over the past decade. Our Physics department has designed, built and contributed detectors to many particle and nuclear physics experiments, including ATLAS, LHCb and ALICE at the LHC, g-2, T2K, SNO+, proto-DUNE, LZ, AGATA, NuSTAR, AWAKE, HLLHC and R3B. The data analysis on these detectors and many running experiments and those that have recently completed data taking is aided by more than 100 PhD students. The very large data volumes from these experiments has given staff and PhD students expertise in data handling, data reduction techniques and the management of systematic uncertainties in these very large-scale analysis tasks. We have also developed code to run on non-traditional processor architectures, the integrated data acquisition, trigger and data compression required for readout systems and simulation of all of the expected and some of the possible physics processes the detectors will be sensitive to. Research will be complemented by the strong involvement of our Mathematics and Computer Science departments which contribute additional skills and important links with external partners. The 14-strong STFC-supported Theoretical Physics group in MS access national high-performance computing (HPC) facilities such as DiRAC to perform lattice field theory simulations for particle, nuclear and condensed matter physics, and employ machine learning techniques for complex problems such as the classification of string theory vacua.

We are very experienced as coordinator of large-scale postgraduate training initiatives, having coordinated training networks worth more than 20M€ for almost 100 Marie Curie Fellows across physics, engineering and life sciences. We have also hosted THE very successful STFC CDT LIV.DAT with LJMU between 2017 and 2020 with around 40 PhD students trained in recent years [2]. These initiatives have driven inclusivity and equality of opportunity: In our OMA network [3], 47% of the Fellows were female, in our CDT LIV.DAT, our students acted as ambassadors for the national IOP Bell Burnell fund to improve diversity in physics [4], and the LIV.INNO Director was one out of only five (and the only male) invited representatives within the 2021 UK country delegation at the International Conference on Women in Physics (ICWIP) where he presented Liverpool EDI initiatives [5] and an update on progress made in the UK [6].

The Astrophysics Research Institute (ARI) of LJMU consists of 55 academic, technical, admin and research staff and more than 40 research students. ARI carries out fundamental research across a wide area of the STFC priority astronomy areas: galaxy evolution, numerical simulations of galaxy formation, star formation and stellar evolution, time domain astrophysics and astronomical instrumentation. ARI research connects well to Big Data, e.g. execution and analysis of massively-parallel numerical simulations of cosmic structure formation, and in time-domain astrophysics covering gamma-ray bursts, super luminous supernovae, and potential electromagnetic counterparts to gravitational wave sources. Central to the ARI's time domain activities is the Liverpool Telescope, the world's largest fully robotic telescope, located on La Palma in the Canary Islands, owned and operated by LJMU and nationally funded by STFC.

* This project has received funding from STFC.
† c.p.welsch@liverpool.ac.uk

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOST050

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. LIV.INNO provides a comprehensive training programme to its three student cohorts to address this problem. The focus of the centre is on addressing the data challenges presented by research in astronomy, nuclear, theoretical, particle and accelerator physics. R&D is structured across the following 3 main work packages: Monte Carlo methods and High Performance Computing, Artificial Intelligence and Machine Learning, and Data Analysis. Detailed R&D plans, including student placements were agreed at the project start and supervisory teams identified so that student can be recruited right after project start.

- Monte Carlo (MC) methods and HPC are powerful tools for everything from modelling the birth and evolution of the universe to performing the numerical integrals needed to calculate cross sections for particle interactions;
- 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; our researchers are at the forefront of this research;
- Data Analysis. Projects will have students working on both STFC core physics topics and to expand the skills into wider applications. This will connect their work to the R&D basis generated in WP2 and expand it to the problem of combining data from different sources and providing an efficient analysis in view of the research goals.

All work packages are highly relevant for accelerator science. 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. In the following, examples of two specific accelerator R&D projects are given.

Optimization of 3D X-ray Imaging Systems

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 and company Adaptix Ltd has demonstrated stationary 3D imaging with this technology in dentistry, orthopaedics, veterinary medicine and non-destructive testing applications.

One project in LIV.INNO that will start later in 2022 will focus on upgrading this cutting edge technology to specifications suited for chest DT. This will include computer aided design studies into a system with a number of source arrays, as illustrated in Fig. 1.

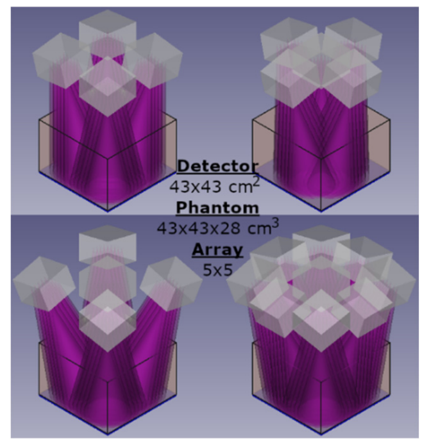

Figure 1: Multi-FPS DT systems designed by Ipioni and drawn in FreeCAD. White boxes: FPSs, purple cones: X-ray beams, transparent box: phantom, blue box: FPD.

Chest DT systems commonly operate at 120 kV, but lowering the accelerating voltage to around 90 kV could produce similar image quality at a fixed effective dose or even fixed emitted photon flux [7, 8]. Therefore, upgrading Adaptix' portable, compact DT systems to 90 kV could bring cheaper, lower dose chest 3D X-ray imaging to smaller clinics and patients' bedside.

Initial Monte Carlo (MC) simulations of this upgrade have already been carried out, involving the development of Ipioni, a software for computer-aided designs (CAD) of multi-FPS DT systems necessary for covering the larger volume of the chest, see Fig. 1. FLUKA [9] simulations were provided the necessary background information on a single monochromatic electron pencil beam of various energies incident perpendicularly on an infinitely wide, 20 μm thick X-ray target of undisclosed material. This revealed optimized combinations of accelerating voltage and target thickness with high photon yield [10].

A LIV.INNO student will further optimize the imager technology through Monte Carlo studies and experimental campaigns involving both, movable X-Ray sources and cold cathode arrays.

Characterization of Photocathodes

The search for high-performance photocathodes is a priority in the field of particle accelerators. Characterization measurements for four Caesium Telluride photocathodes synthesized at CERN were recently carried out and are presented in detail elsewhere at this conference [11]. The

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOST050

photocathodes were transported under ultra-high vacuum and analyzed at STFC Daresbury Laboratory, using ASTeC's Multiprobe for surface characterization via XPS and STM, and the Transverse Energy Spread Spectrometer (TESS) [12] for Mean Transverse Energy measurements.

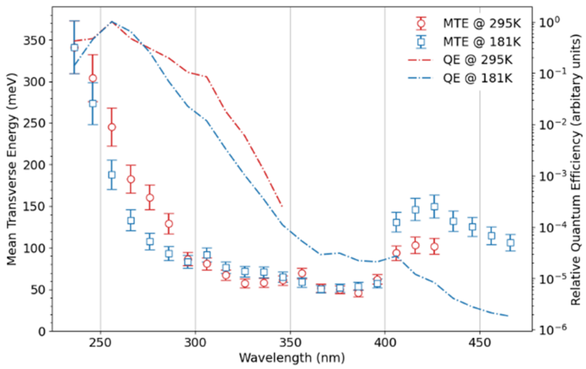

Figure 2: Comparison between extracted MTE values and estimated relative QE values based on known gain parameters and optical power for the CsTe P6 cathode at illumination wavelengths of 236 nm - 436 nm at both room and cryogenic temperatures.

Figure 2 shows the calculated relative Quantum Efficiency for one of the photocathodes which generally declines as the illumination wavelength increases. Future studies will focus on the detailed modelling and experimental verification of the emission characteristics of various cathode materials. The intention is to identify the optimum materials and preparation techniques to achieve the highest levels of electron beam brightness, with the lowest intrinsic emittance and the longest operational lifetime. The project will also continue the development, optimization and exploitation of TESS for the characterization of novel photocathode electron source materials.

TRAINING

The projects within LIV.INNO are very demanding in their own right - in combination within each student cohort and across all three cohorts, they will allow our students to tackle some of the big challenges in using Data Science to advance STFC science and drive innovation with identified partners. The training is designed to provide skills that boost the employability of the center's students:

- **Focused research skills and techniques** through cutting edge R&D projects;
- **Broader scientific experience** via realization of placements and participation in workshops and Schools, including interdisciplinary exposure;
- **Research project management, communication and presentation skills** via a complementary skills School in the first year, supported by Fistral. Outreach activities will be completed by the students and communicated internationally using the established communication channels of ULIV and LJMU.
- **Understanding of the benefits of an inclusive environment**, by integrating Equality, Diversity and Inclusion (EDI) in all aspects of the CDT.
- **Interpersonal skills, accepting responsibility, working in teams, networking** will be developed through project work, a student-organized charity event, and involvement in events open also to external delegates, such as the CUWIP 2023 which will be hosted in Liverpool and an LIV.INNO outreach Symposium at the ACC in 2024. This will encourage networking and ongoing development of relations to complement conference participation encouraged in the wider community.
- **Awareness of private sector constraints and knowledge of Intellectual Property (IP) rights and management** via industrial mentors and secondments, close collaboration with and coaching by the Business Development Manager (BDM) based in UoL Physics, complemented by training through IPR helpdesk webinars.
- **Career management, CV writing, interview techniques, proposal preparation and entrepreneurship** will be covered in an advanced researcher skills School in the final year and through commercialization mentoring by supervisory teams.

LIV.INNO's training program has been developed to create future leaders in data intensive science, with a focus on three scientific areas. The CDT's students will undertake training on a cohort basis, with the opportunity to personalize their program. They will complete a four-year program which includes the award of a PGCert in Big Data after completion of 120 credits in year 1. The program will be delivered by academic staff from both partner universities and supported by external partners, potential employers and users in Data Science. In years 2-4, students will 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 will include 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. The kickoff meeting of the new center [13] attracted more than 100 experts in May 2022 and also saw the official launch of the LIV.INNO video [14].

SUMMARY AND OUTLOOK

The LIV.INNO CDT will start in October 2022 with a first cohort of around a dozen PhD students. They will carry out interdisciplinary research across three scientific work packages, all related to Data Intensive Science. Research plans were illustrated on the example of two projects focusing on accelerator science R&D and a brief overview of the training program was also given. The center benefits from outstanding research partnerships in the UK and abroad and has attracted high interest from research organizations and businesses from its launch.

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOST050

REFERENCES

[1] LIV.INNO homepage, http://www.livinno.org

[2] C.P. Welsch, "Using Data Intensive Science for Accelerator Optimization", presented at IPAC'22, Bangkok, Thailand, June 2022, paper TUPOST051, this conference.

[3] C.P. Welsch, "Advances in the Optimization of Medical Accelerators", presented at IPAC'22, Bangkok, Thailand, June 2022, paper THPOMS032, this conference.

[4] Women in STEM: Improving diversity in physics, https://bit.ly/3GORVbB

[5] C.P. Welsch, *et al.*, "Liverpool Women in Physics: Initiatives and Progress", Proc. ICWIP 2021, *submitted*

[6] S. Jordan, *et al.*, "Women in Physics in the UK: Successes, Challenges and Wider Diversity", Proc ICWIP 2021, *submitted*

[7] C. Söderman *et al.*, "Image quality dependency on system configuration and tube voltage in chest tomosynthesis-A visual grading study using an anthropomorphic chest phantom", *Med. Phys.,* vol. 42, p. 1200–1212, February 2015. doi:10.1118/1.4907963

[8] 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

[9] A. Fassò, A. Ferrari, J. Ranft, and P. R. Sala, *FLUKA: A multi-particle transport code (program version 2005),* CERN, 2005.

[10] T. Primidis, *et al.*, "Design and Optimisation of a Stationary Chest Tomosynthesis System With Multiple Flat Panel Field Emitter Arrays: Monte Carlo Simulations and Computer Aided Designs", presented at IPAC'22, Bangkok, Thailand, June 2022, paper THPOMS033, this conference.

[11] L. Soomary, *et al.*, "Performance Characterisation at Daresbury Laboratory of Cs-Te Photocathodes Grown at CERN", presented at IPAC'22, Bangkok, Thailand, June 2022, paper THPOPT033, this conference.

[12] L.B. Jones *et al.*, "The Commissioning of Tess: An Experimental Facility for Measuring the Electron Energy Distribution From Photocathodes", in Proc. 35th Int. Free Electron Laser Conf. (FEL'13), New York, NY, USA, Aug. 2013, paper TUPSO33, pp. 290-293.

[13] LIV.INNO Kickoff meeting event page, indico.ph.liv.ac.uk/event/593/

[14] LIV.INNO – diving innovation through data science short film, https://youtu.be/ZzVnNcPyNfA