The SMARTHEP European Training Network

James Andrew Gooding[1,*]

[1]Fakultät Physik, Technische Universität Dortmund, Dortmund, Germany

Abstract. **S**ynergies between **MA**chine learning, **R**eal-**T**ime analysis and **H**ybrid architectures for efficient **E**vent **P**rocessing and decision-making (SMARTHEP) is a European Training Network, training a new generation of Early Stage Researchers (ESRs) to advance real-time decision-making, driving data-collection and analysis towards synonymity.
SMARTHEP brings together scientists from major LHC collaborations at the frontiers of real-time analysis (RTA) and key specialists from computer science and industry. By solving concrete problems as a community, SMARTHEP will further the adoption of RTA techniques, enabling future HEP discoveries and generating impact in industry.
ESRs will contribute to European growth, leveraging their hands-on experience in machine learning and accelerators towards commercial deliverables in fields that can profit most from RTA, e.g. transport, manufacturing, and finance.
This contribution presents the training and outreach plan for the network, and is intended as an opportunity for further collaboration and feedback from the CHEP community.

1 Introduction

The **S**ynergies between **MA**chine learning, **R**eal-**T**ime analysis and **H**ybrid architectures for efficient **E**vent **P**rocessing and decision making (SMARTHEP) European Training Network is an EU Horizon-funded training network, with a focus on the development of expertise in real-time analysis (RTA) techniques through applications to high energy physics (HEP) research and industry. The network centres around the training of 12 Early Stage Researchers (ESRs) between September 2021 and September 2025.

2 SMARTHEP as a European Training Network

As a European Training Network (ETN), the primary aim of the network is in training ESRs, whilst deepening synergies between HEP and industry. The network takes a novel approach to building such synergies, structuring each ESR position (a 3 year period of doctoral study) around academic and industrial secondments. To achieve this, the network is formed of a series of partnerships between universities, research institutes and organisations in industry, as listed in Table 1.

The network is formally structured as 7 Work Packages (WPs), laid out in Figure 1. WP1 “Management”, covers the management of the network by the Project Manager/Project

*e-mail: Jamie.Gooding@cern.ch

Category	Partners
Universities	Lund University, Sorbonne University (LPNHE & LIP), Technische Universität Dortmund, University of Bologna, University of Geneva, University of Heidelberg, University of Helsinki, University of Manchester, Universidade Santiago de Compostela (IGFAE), Vrije Universiteit Amsterdam
Research institutes	CERN, CNRS, NIKHEF
Industry partners	IBM France, Lightbox, Point 6, Uni.versity of Manchester Institute for Data Science and AI, Verizon Connect, Ximantis

Table 1. The 18 organisations which form the SMARTHEP network.

Coordinator/Executive Board. WP2 "Training", sets out training and development of participants, by both network partners and external training providers. WP3 "Machine learning & advanced data analysis", develops machine learning (ML) techniques for use in real-time environments. WP4 "Hybrid architectures" focuses upon the deployment of non-CPU computing architectures in acceleration of data processing. WP5 "Decision-making in research and industry", applies WP3 and WP4 to develop RTA-based decision-making technologies. WP6 "Monitoring and discoveries", also applies WP3 and WP4 to RTA approaches in data analysis. WP7 "Dissemination and communication of results", publishes and propagates the results produced in WP6 and WP7.

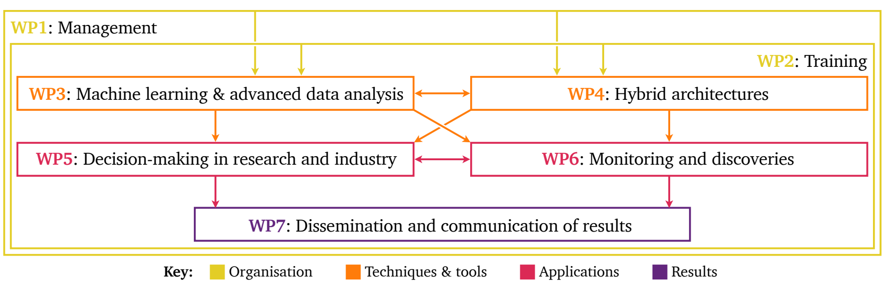

Figure 1. The structure of work packages within the SMARTHEP network. WP1 and WP2 define the organisation of the network; WP3 and WP4 introduce the techniques and tools of real-time analysis to the network; WP5 and WP6 use said techniques and tools to produce results for HEP and industry; WP7 makes these results available and promotes their wider use and adoption.

The network has a particular focus on physics at the Large Hadron Collider (LHC). Each ESR is thus affiliated to one of the four major experiments based at the LHC: ALICE, ATLAS, CMS and LHCb. The network duration coincides with Run 3 of the LHC (2022-2025), with many of the ESR projects framed in this context.

A unique feature of the network is the extensive cooperation between HEP and industry across all ESR positions. RTA approaches have seen significant adoption in industry in recent years, with many organisations turning to RTA as a means to handle "big data".

3 Real-time analysis

HEP and industry share a common challenge—the rapid processing of large amounts of data. [1] Recent advances in computing have enabled the possibility of processing data in real-time, i.e. as data is collected (also commonly referred to as "online" processing). [2] Real-time analysis (RTA) is thus an umbrella term for such approaches, encompassing a number of state-of-the-art data processing techniques. By processing data online, resources (computing power, storage space, energy, etc.) can be saved and new insights can be obtained from the data recorded, as shown in Figure 2.

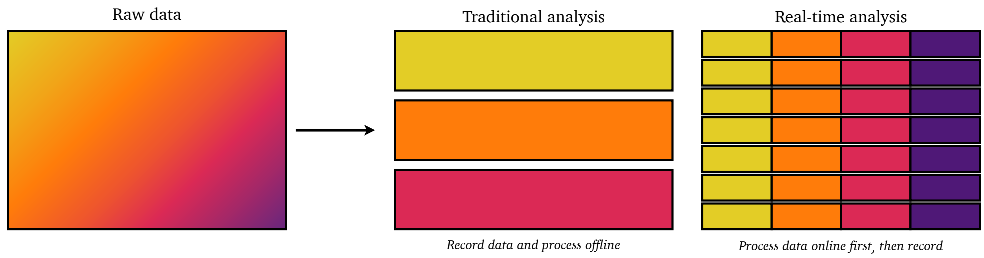

Figure 2. Traditional and real-time analysis approaches to data processing. Traditional approaches rely on recording all data and processing this offilne. Real-time analysis applies the reverse approach, processing data as it is produced, recording only the relevant processed information, enabling larger volumes of processed data to be stored.

RTA techniques have seen widespread adoption across HEP, in particular amongst trigger and data acquisition (TDAQ) systems. Since it is not possible to record a full detector readout and carry out event reconstruction at the LHC collision rate of 40 MHz, triggers must be devised to select only those events relevant to the physics goals of an experiment. Such triggers conventionally consist of a hardware system making coarse decisions from partial detector readout, and a staged software trigger, applying gradually more finely-grained selections with increasingly more detailed reconstructions of events.

In industry, the limitations of the scale of "big data" looms over many applications of computing. However,

Successful implementation of RTA approaches rests upon the . In particular, in the areas of machine learning and hybrid architectures, which are summarised in Sections 3.1 & 3.2.

3.1 Machine learning

ML, a catch-all term for a family of techniques and technologies wherein algorithms are trained to analyse data, enables rapid decision-making and pattern recognition across a broad range of use cases. [3] In HEP, the adoption of ML began with classifiers for offline physics analysis, since widening to a variety of classification and pattern recognition/anomaly detection techniques for use online and offline. [4]

ML classifiers are algorithms classed with classification, with Boosted Decision Trees (BDTs) and Neural Networks (NNs) amongst those most commonly employed in HEP. The use of BDTs in signal selection for offline physics analysis has become standard practice, e.g. suppresion of combinatorial background in the LHCb experiment measurement of the $B_s^0 - \bar{B}_s^0$ oscillation frequency. [5] Classifiers can also aid tasks such as event reconstruction, e.g. the CMS boosted event shape tagger, a NN trained to discriminate between possible t, $W^{\pm}$, Z^0 and H candidates within an event. [6] In industry, ML classifiers are used frequently in content deliver

ML techniques can also be applied to pattern recognition and anomaly detection—tasks which cannot be realistically carried out on the scales of data presently being analysed. In industry, this is applied intuitively to fraud detection, wherein subtle details in data such as the transactions of a customer. In HEP [7]

3.2 Hybrid architectures

Typical computational resources consist of Central Processing Units (CPUs), processors which are designed for general purpose computation (i.e. varied tasks with significant variety in computing/memory resources), with large on-board memory and often multiple processing cores. Alternative architectures, such as those shown in Figure 3, can be applied in conjunction or in place of CPU architectures to accelerate processing where tasks are not well-suited are to CPU architectures alones.

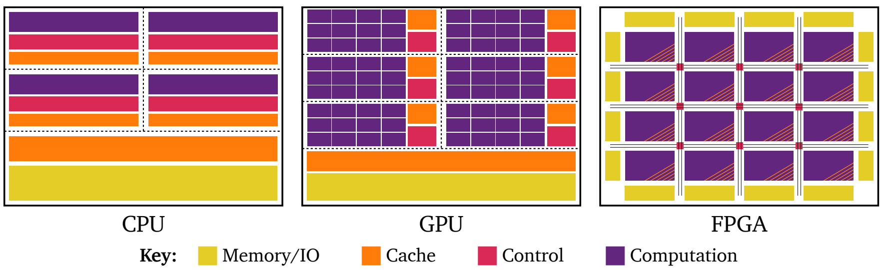

Figure 3. Comparison of CPU, GPU and FPGA architectures. A CPU is typically formed of several cores, each containing computational, control and cache resources, and a centralized cache, memory and input/output (IO) interface. GPUs are formed similarly, with centralized cache, memory and IO, and many multiprocessors; however, each multiprocessor contains a greater proportion of computational resources, which are themselves partitioned to perform tasks in parallel. FPGAs are structured in a radically different way, with memory and IO connected to many interlinked control blocks, each primarily formed of simpler logic gate arrangements and often accompanied by a small cache.

Field-programmable gate arrays (FPGAs) are integrated circuits which can be programmed by the user for specific tasks. [8] An FPGA consists of many simple logic circuits with a small amount of on-board memory, formed into logic blocks, each of which are connected via switch matrices. [9] FPGAs are thus well-suited to bespoke, computationally light but highly parallelisable tasks, such as low-level trigger decisions, wherein simple selections must be made rapidly. [10] In industry, FPGAs have seen adoption . However; a lack of on-board memory limits the deployment of FPGAs to more intensive tasks.

Graphical Processing Units (GPUs) can also be employed to . [11] Historically, the use of GPUs has been

4 Early stage researchers

12 ESRs form the core of the network, with the training and partnerships providing a scaffold for the completion of their respective outcomes. These ESRs work with HEP and industry partners throughout the 3 years of their doctoral study.

4.1 Structure of ESR positions

Each ESR is enrolled as a doctoral student at a partner university for 3 years. This university is typically also the primary institution of the ESR (2 industry-centred ESR positions are enrolled at a university near to the respective industry partner). At the end of the period of doctoral study, each ESR will receive a doctorate in particle physics from their university of enrollment.

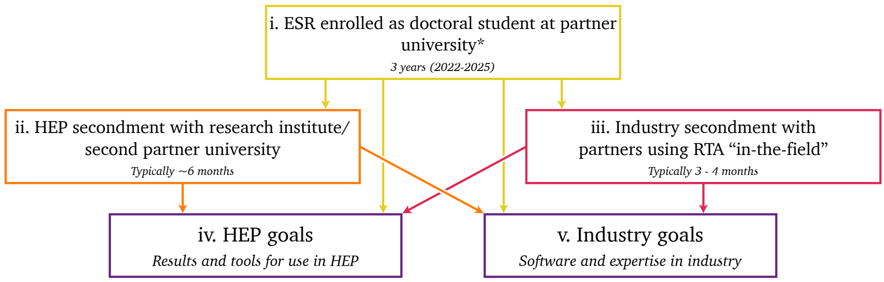

Figure 4. Diagram of the structure of a SMARTHEP ESR position. Each ESR is enrolled (i.), during which they will undertake secondments with network partners in HEP (ii.) and industry (iii.). Through the combination of primary and secondment work, each ESR will complete goals in HEP (iv.) and industry (v.), discussed in further detail in Section 5.1.

Each ESR will undertake a secondment in HEP during their studies, either at another partner university or a partner research institute. This is generally organised to last for 6 months, though this varies from project to project.

In addition to their time working in HEP, each ESR will also spend time working in industry, typically as a secondment of 3-4 months (though this is again flexible to the specific programme of each ESR position).

Laura Boggia

Carlos Cocha

Joachim Hansen

5 Network outcomes

The network is anchored by a set of concrete outcomes, guiding the progression of the network and its participants. These intended outcomes are summarised briefly below.

5.1 Goals of HEP and industry

The core outcomes of the network are a set of goals, completed on behalf of the network partners by the network participants. These goals include experiment commmissioning, HEP measurements and industrial results. These goals are typically defined with respect to a specific ESR position, e.g. the goal of "calibration of ALICE TPC for heavy-ion physics" associated to Joachim Hansen.

5.2 RTA whitepapers

Participants will write a series of three whitepapers reviewing the current RTA state-of-the-art, expected to be submmited in late 2023. These whitepapers will place a particular emphasis on the ongoing contributions of the network to their respective topics.

The first whitepaper reviews ML applications to HEP in RTA contexts and their corresponding best practices. Such applications range from data-taking, e.g. particle identification at ALICE, to offline analysis, e.g. anomaly detection for diject resonance searches at ATLAS. [12, 13] The use of hybrid architectures by LHC experiments is reviewed in the second whitepaper. Tasks such as selection and reconstruction were significantly accelerated during

Run 2 of the LHC (2015-2018). Run 3 deployments of hybrid architectures capitalise on this progress, e.g. in the use of FPGAs in the ALICE Central Trigger System and the use of GPUs in the ATLAS High Level Trigger farm. [14, 15] In the third whitepaper, TDAQ systems of LHC experiments are reviewed, with a focus upon best practices for both TDAQ hardware and software. As such, this whitepaper reviews topics ranging from upgrades to the ATLAS TDAQ system, to the Allen framework of the LHCb experiment enabling software trigger operation at a 30 MHz readout rate. [16, 17]

5.3 Training of ESRs

Funding is allocated within the network for participants to attend relevant training activities. In the case of ESRs, training and career devlopment is discussed in writing their respective PCDP, providing a clear overview of their intended and requested training activities. Examples of such activities include attendance of specialist industrial training sessions and academic schools. Training is also provided within the network, between participants, for example in the First SMARTHEP School on Collider Physics and Machine Learning discussed in Section 6.2.

5.4 Software and digital assets

The work of the ESRs will generate a number of digital assets (e.g. software packages, data processing tools), with many being applicable beyond narrow academic/industrial applications. The network is committed to producing making any produced digital assets **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable (FAIR). [18] Under these guiding principles our assets and results should be open and accessible to the wider community. To implement these commitments, a project on GitHub has been created to host such assets, https://github.com/SMARTHEP. Other assets and resources will be made available on the network website, https://smarthep.org.

6 Network events

Network events form the backbone of the SMARTHEP network, giving participants unique opportunities to meet, exchange ideas and develop. Whilst the events of the network are focused on an ESR audience, SMARTHEP has made many events available to additional interested early career scientists working/studying at SMARTHEP institutes.

6.1 SMARTHEP Kick-off Meeting

To commence ESR particpation in the network, a kick-off meeting was held at the University of Manchester from 21$^{\text{st}}$ to 25$^{\text{th}}$ November 2022, formally introducing the ESRs, and discussing objectives and organisation of the network. The meeting also served as an opportunity for participants from all sides of SMARTHEP to meet in person, network and discuss ideas. To facilitate this, a visit to the Jodrell Bank Centre for Astrophysics was held, and a review paper-writing course was provided by Scriptoria to train ESRs ahead of the writing of the 4 RTA whitepapers.

6.2 First SMARTHEP School on Collider Physics and Machine Learning

The First SMARTHEP School on Collider Physics and Machine Learning was hosted by the University of Geneva from 10th to 13th January 2023. The school provided a varied programme, including lectures on experimental physics at collider experiments by Anna Sfyrla and on theoretical physics & Monte Carlo event generators by Torbjörn Sjöstrand. Additionally, hands-on lessons in machine learning were led by Maurizio Pierini, with seminars also given on multimessenger astronomy by Teresa Montaruli and the CERN experimental programme by Jamie Boyd.

6.3 Upcoming events

At time of writing, the network still has two years ahead of it, in which further network events are planned. These will guide the governance of the network, develop ESR expertise and deepen collaborations with industry.

The network assembly sits as a cornerstone of network policy and governance.

Accelerator bootcamps in specific aspects of RTA . These are likely to take place in the summer of 2024. An industry applications school will round off the collaboration with industry, by which time it is expected that ESRs will have completed their industry secondments.

7 Conclusion

The network will provide valuable contributions in HEP, particularly to the commisioning and operation of LHC experiments throughout Run 3 of the LHC. Additionally, the network serves to further the adoption of RTA approaches in industry and do demonstrate the value of collaboration between HEP and industry.

Acknowledgements

We acknowledge funding from the European Union Horizon 2020 research and innovation programme, call H2020-MSCA- ITN-2020, under Grant Agreement n. 956086

References

[1] H. Hu, Y. Wen, T.S. Chua, X. Li, IEEE Access **2**, 652 (2014)

[2] W. Tian, Y. Zhao, in *Optimized Cloud Resource Management and Scheduling*, edited by W. Tian, Y. Zhao (Morgan Kaufmann, Boston, 2015), pp. 17–49, ISBN 978-0-12-801476-9

[3] J. Jovel, R. Greiner, Frontiers in Medicine **8** (2021)

[4] K. Albertsson, P. Altoe, D. Anderson, J. Anderson, M. Andrews, J.P.A. Espinosa, A. Aurisano, L. Basara, A. Bevan, W. Bhimji et al., *Machine learning in high energy physics community white paper* (2019), 1807.02876

[5] R. Aaij, C.A. Beteta, T. Ackernley, B. Adeva, M. Adinolfi, H. Afsharnia, C.A. Aidala, S. Aiola, Z. Ajaltouni, S. Akar et al., Nature Physics **18**, 1 (2022)

[6] A. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, T. Bergauer, M. Dragicevic, J. Erö, A.E.D. Valle, M. Flechl, R. Frühwirth et al., Journal of Instrumentation **15**, P06005 (2020)

[7] B. Nachman, *Anomaly Detection for Physics Analysis and Less Than Supervised Learning* (2022), chap. Chapter 4, pp. 85–112

[8] J. Serrano, *Introduction to FPGA design* (2008), https://cds.cern.ch/record/1100537

[9] U. Meyer-Baese, *Digital Signal Processing with Field Programmable Gate Arrays* (Springer Berlin Heidelberg, 2013), ISBN 9783662046135, https://books.google.co.uk/books?id=L6fvCAAAQBAJ

[10] J. Duarte, P. Harris, S. Hauck, B. Holzman, S.C. Hsu, S. Jindariani, S. Khan, B. Kreis, B. Lee, M. Liu et al., Computing and Software for Big Science **3** (2019)

[11] D. vom Bruch, Journal of Instrumentation **15**, C06010 (2020)

[12] Ł.K. Graczykowski, M. Jakubowska, K.R. Deja, M. Kabus, on behalf of the ALICE collaboration, Journal of Instrumentation **17**, C07016 (2022)

[13] G. Aad, B. Abbott, D.C. Abbott, A. Abed Abud, K. Abeling, D.K. Abhayasinghe, S.H. Abidi, O.S. AbouZeid, N.L. Abraham, H. Abramowicz et al. (ATLAS Collaboration), Phys. Rev. Lett. **125**, 131801 (2020)

[14] Kvapil, Jakub, Bhasin, Anju, Bombara, Marek, Evans, David, Jusko, Anton, Kluge, Alexander, Krivda, Marian, Kralik, Ivan, Lietava, Roman, Nayak, Sanket Kumar et al., EPJ Web Conf. **251**, 04022 (2021)

[15] A. Bocci, on behalf of the CMS Collaboration, Journal of Physics: Conference Series **2438**, 012016 (2023)

[16] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, O. Abdinov, R. Aben, B. Abi, O. AbouZeid, H. Abramowicz et al. (ATLAS), Tech. rep. (2013), final version presented to December 2013 LHCC., https://cds.cern.ch/record/1602235

[17] 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 et al., Computing and Software for Big Science **4**, 7 (2020)

[18] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.W. Boiten, L.B. da Silva Santos, P.E. Bourne et al., Scientific Data **3**, 160018 (2016)