A vision for the SWIFT-HEP2 project

This document provides a vision for what the second phase of the SWIFT-HEP project would look like. We discuss the plans and the ambitions we have after the completion of the current phase, from 2025 to the end of the decade. A detailed description of the workplan and of the positions sought is beyond the scope of this document. These will be provided as part of the PPRP submission once the budget envelope is defined.

SWIFT-HEP Phase 1 achievements

The SoftWare InFrasTructure for High Energy Physics (SWIFT-HEP) project was funded by STFC in 2021, initially for 3 years and then extended, at a constant effort, for a further 18 months to September 2025. We currently fund 4.5 FTE of effort at 8 institutions across the UK, plus travel and training to facilitate community discussions and workshops. The model is to fund part-posts at institutes to be matched by other experimental funding. This proved to be successful as it allowed developers from different experiments and theorists to work together on common projects.

The goals and deliverables set in the initial proposal document were largely achieved, full details can be found in the yearly reports submitted to STFC in June 2022 and July 2023. To mention a few achievements:

- **Optimization of Event Generator Code:** We successfully developed code that significantly improves the efficiency of generating standard model backgrounds at the LHC (V+jets, top+jets), reducing required CPU time by more than a factor of 10. We have similarly modernised generator codes needed for simulation of heavy hadron decays, enabling the generators to operate in multithreaded CPU environments, and reducing consumed CPU time;
- **Simulation of Electromagnetic Showers on GPU**: We are developing code to simulate electromagnetic showers on GPU and the ray tracing of optical photons, with the goal to reduce the cost of this processing task that currently uses a significant fraction of all grid resources. Initial results suggest an increased throughput of a factor of 2;
- **Multi-Architecture Tracking Algorithms:** We are leading the development of tracking algorithms that run on multiple architectures (GPU, FPGA) and can be used both offline and online, and are also developed to be used at future experiments;
- **Prototype for Analysis Workflows on GridPP:** We developed a prototype to run analysis workflows using industry standard tools (Dask) on GridPP resources. This reduces the barrier of entry to LHC researchers, as well as users from smaller experiments. The prototype led to the development of new features in the workflow management tools (DIRAC) and the definition of a data lake based on quality of service as a main parameter;
- **Leadership within HEP Software and Computing Community:** Among others, the UK currently provides 3 conveners to the 8 HEP Software Foundation (HSF) working group, with UK members having convened 6 out of 8 groups since 2020. In addition the UK provides the software liaison to the WLCG, a member of the Geant4 steering board and has a representation on the Geant4 oversight board. This underscores our role in guiding the community's software and computing strategies.

SWIFT-HEP Phase 2 plans

The second Phase of SWIFT-HEP starting in October 2025 for either 3 or 4 years should build on the model used during Phase 1. In Phase 1 we demonstrated the ability to grow the project, engage with the international community and deliver on the scope that was initially set. Phase 2 should also focus on the deployment by the experiments of the software developed to capitalise on the investments made.

This grant will enable us to retain researchers with key expertise, by providing them with some security and career progression prospects. It is vital to have a clear financial commitment ahead of time so people have job security and career development prospects and research software engineers (RSEs) can work on the project with continuity.

The current cost of SWIFT-HEP is around £500k/year, plus we had some additional effort as part of the ExCALIBUR programme (ending in December 2024). The initial SoI submitted to STFC in 2020 devised a project with a cost of about £2M/year, and in phase 2 we aim to further grow the project taking on further deliverables in agreement with the experiments and our international collaborators.

Phase 2 work packages

The work package structure should broadly reflect the organisation of the HSF to have clear points of interaction with our international colleagues. The Phase-1 structure was mostly efficient, although some restructuring is needed (merge WP1 and WP5), and additional work packages should be added to reflect the development in the software and computing landscape. The proposed organisation is:

- **WP0 Management and training**. To include the project leader and deputy, the WP leaders, a project manager, a project administrator and a training and outreach coordinator.
- **WP1 Distributed systems and analysis**. Combines the Phase-1 WPs on workflow and data management, with the work on analysis tools to ensure that different workflows access data in an optimal way depending on requirements (e.g. analysis jobs should get fast storage access, long production jobs can schedule recall from tape, etc). An interface for scheduling analysis jobs and smaller productions will be deployed following the "grand analysis challenge" developed by IRIS-HEP in the US. This will ease the scale out of user jobs on grid infrastructure.
- **WP2 Event generators.** Provides the software engineering needed to make optimal use of theoretical advances in the generation of physics processes and ensure that the use of higher order calculations is not limited by the compute power available. Continued software engineering is also needed to exploit optimally modern compute architectures, such as GPUs on HPCs and on the grid, and hence minimise the resource usage.
An important element of this WP will be the development and optimisation of interface tools (HepMC, LHAPDF, Rivet) needed for sharing simulated events between specialised packages (e.g. those providing precise simulation of final state radiation, particle decays, parton distribution functions, or the particle interaction codes of WP3), and for enabling validation of new simulation results.
- **WP3 Simulation of particle interactions** which contributes to the largest fraction (in some cases more than 50%) of compute resource usage for the LHC experiments. This will include the development of electromagnetic showering models and optical photon propagation to run on GPUs. We should also increase our participation in the Geant4 collaboration to a level consistent with that of other funding agencies with work that will

also impact other communities using Geant4 such as space and health. Further developments of fast simulation models, based on AI techniques, will also be possible in conjunction with WP7.

- **WP4 Reconstruction algorithms** to match the UK expertise in designing and developing particle physics detectors. Cross-experimental reconstruction tools are developed both for current and future experiments and face the challenge to reconstruct objects in high density environments with high accuracy within the time and power constraints of online triggering systems, as well as multiple architectures (GPUs, FPGAs, …).
- **WP5 Benchmarking and sustainability** is a new WP and will work on deploying software test beds on different compute infrastructure to measure the compute power, energy requirements and CO_2 footprints of our workflows. In collaboration with GridPP this WP will also develop the tools needed to monitor these footprints both on local and distributed systems, and will contribute to the STFC net zero drive.
- **WP6 Software frameworks for future experiments,** Phase 2 will see increasing interest in future experiments at the energy and intensity frontiers. New software frameworks (e.g. Key4HEP) are being developed to enable efficient optimisation of future experiments design. It is crucial to ensure the smooth integration of other SWIFT-HEP software developments in collaboration with the DRD programme.
- **WP7 AI/ML tools and emerging technologies**. Artificial intelligence is transformative in several disciplines and is driving the development of new hardware. To continue to benefit from AI developments, we need to develop the tools to optimise and deploy the ML algorithms used in HEP to AI-optimised architectures, including the latest HPC being deployed by UKRI. It's essential that our computing workflows evolve to efficiently utilise these emerging resources, following the FAIR principle.
  Another emerging technology is quantum computing. Given the quantum mechanical nature of many problems in HEP, such as scattering matrix element calculations, quantum computing and quantum machine learning offer exciting new technological frontiers.

Risk and opportunities

The primary risk for the project continues to be recruiting and retaining skilled staff. This can be mitigated with a clear financial plan that offers staff a stable view of their future work. While competing with industry salaries remains challenging, fostering a supportive and stimulating work environment is key to retaining talent and ensuring a beneficial turnover rate.

The other major risk is connected with the non-adoption of some of the software developments by experiments. This is a systemic risk in R&D projects and we plan to mitigate this by maintaining close interactions with the experiments, and the infrastructure providers, understanding their needs, and tailoring our developments accordingly.

With risks come also opportunities, as emerging technologies can change the paradigm of what we do. The maintenance of a risk and opportunity register, as well as good project management will capture these opportunities and review deliverables throughout the course of the project.