Nucleon Structure Knowledge and Analysis Toolkit (NuSKAT)

A letter of Intent for HORIZON-INFRA-2025

Coordinator: P. Chatagnon (IRFU, CEA)

1. Research objectives

Precision in the description of the structure of the nucleon is becoming a central pursuit of the hadronic physics community, building on more than four decades of theoretical and experimental progress. The Generalized Parton Distributions (GPDs) and Compton Form Factors (CFFs) are structure functions directly related to fundamental properties of the nucleons including the origin of their spin and the generation of their mass. Major facilities such as Jefferson Lab (JLab), HERMES at DESY, and COMPASS at CERN have been instrumental in providing an abundance of experimental data aimed at addressing these fundamental questions. Looking ahead, new experimental programs, notably the ePIC experiment at the future Electron-Ion Collider (EIC) in Brookhaven National Laboratory (BNL), are being developed to further expand our knowledge on GPDs and more generally on the dynamics of partons inside nucleons. To extract GPDs from these data — and thus achieve a multidimensional picture of the nucleon — it is essential to analyze a wide variety of exclusive reactions. These analyses often involve sophisticated and resource-intensive workflows. With the advent of large-acceptance and high-luminosity experiments, the need for standardized, efficient, flexible and broadly accessible analysis tools has become increasingly pressing.

The Nucleon Structure Knowledge and Analysis Toolkit (NuSKAT) project is a direct response to these challenges. It aims at supporting and accelerating the development of innovative and shareable analysis tools for the experimental hadronic physics community, with a strong emphasis on Artificial Intelligence (AI). These AI-driven solutions will address critical tasks such as fast Monte Carlo generation, detector effect unfolding, particle identification, and efficient event selection. A key objective of the project is to centralize and openly disseminate these tools to maximize their visibility, reusability, and impact across the community. Additionally, several workshops and tutorials will be organised to promote the effective use of these tools. This is particularly important for the AI-based tools, which are increasingly essential in the skillset of early-career scientists.

NuSKAT aims to provide and centralize the following services:

1. Generic Public Tools – Usable directly or as a foundation for more tailored applications:
   - AI-based particle identification algorithm (e.g. muon identification for ePIC and CLAS12, calorimeter-based PID for leptons and pions). Such PID based improvements will be crucial to design a second detector at the EIC (CEA Saclay, INFN Ferrara, UAH),
   - AI-based unfolding methods focused on exclusive reactions (Glasgow),
   - Fast Monte-Carlo based on AI, with strong links to theory tools to test both the impact of new measurements on CFFs and GPDs extraction, and the model dependance introduced by using a specific model to determine the detection efficiency for a given measurement (IJCLab, CEA Saclay),
2. Specialized tools for exclusive reactions and GPD extraction:
   - Compton Form Factors extraction tool to test the impact of future experiments or the reprocessing of existing data with new techniques such as the ones mentioned above (IJCLab),

- AI-based event selection, aiming at fully exploiting data produced at Jlab, BNL and in ultra-peripheral collisions at the LHC. Tools exploiting event topologies with missing particles and for which Machine Learning AI tools efficiently capture the non-linear decision boundaries in the multivariate space will be developed (CEA Saclay, IJCLab, INFN Ferrara, Glasgow),
- Extension of the framework 3DPARTONS and the EpIC generator to include the study of exclusive processes in ultra-peripheral collisions from RHIC and the LHC, expanding the kinematic reach down to very low Bjorken-x ($\sim 10^{-6}$). This work will have strong synergies with the initiative on the theory of GPDs. (UAH)

Summary of the deliverables:

1. **Development of AI-based tools** for key tasks such as particle identification, event selection, unfolding, and fast simulation.
2. **Creation of a centralized website** to host the developed tools, including source code, reference datasets, documentation, tutorials, and publications.
3. **Community engagement and training**, through the organization of yearly workshops, a dedicated summer school, and targeted tutorial sessions to promote collaboration and ensure broad adoption of the tools.

2. Connection to Transnational Access infrastructures and Virtual Access projects

NuSKAT will empower the experimental hadronic physics community to fully exploit data currently being collected and data that will be produced at pre-selected infrastructures such as JLab, the LHC and BNL. By integrating state-of-the-art analysis techniques, the project will enhance the interpretation of existing datasets and support the design and optimization of future experiments, particularly those planned at the EIC.

Given NuSKAT's strong focus on the experimental study of GPDs, close collaboration with the theoretical community is essential. The project will therefore build on established infrastructures such as 3DPARTONS (specifically for the tools related to CFFs projections), and incorporate future developments anticipated within the initiative on the theory of GPDs (GPDPortal). The NuSKAT proposal targets specific scientific objectives and methodological specificities related to the measurement of exclusive reactions and the experimental extraction of GPDs. Nonetheless, synergies with other initiatives, such as the Artificial Intelligence for Hadron Spectroscopy and Interactions (AI4HSI) project, could be pursued to mutualize tools when relevant and maximize the impact of AI-driven methodologies across related efforts.

3. Estimated budget request:

760k **(**6 years of postdocs: 630k; Travels: 80k; Workshops, schools, and tutorials: 50k)

4. Participating and partner institutions

- IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France,
- IJCLab, CNRS/IN2P3, Université Paris-Saclay, Orsay, France,
- INFN (Sezione di Cosenza & Università della Calabria, Ferrara, Genova), Italy,
- Universidad de Alcalá, Alcalá de Henares, Spain,
- University of Glasgow, Glasgow, United Kingdom.