Contribution ID: **18** Type: **Oral Presentation**

Distributing the Simulated Annealing workload for Quantum Unfolding in HEP

Tuesday, 18 March 2025 15:20 (20 minutes)

High-Energy Physics (HEP) experiments involve a unique detector signature - in terms of detector efficiency, geometric acceptance, and software reconstruction - that distort the original observable distributions with smearing and biasing stochastic terms. Unfolding is a statistical technique used to reconstruct these original distributions, bridging the gap between experimental data and theoretical predictions. The emerging technology of Quantum Computing offers potential improvements for unfolding, by addressing its computational complexity. To accomplish this task, a simple Python module named QUnfold has been developed, addressing such a challenge by means of the quantum annealing optimization process. In such a context, the regularized log-likelihood minimization formulation - required by the unfolding problem - is translated into a Quantum Unconstrained Binary Optimization (QUBO) model, which can be solved via quantum annealing systems.

Despite being a promising approach to tackle the unfolding problem, the scalability of Simulated Annealing poses hard challenges, especially with the increasing data volume expected during the high-luminosity phase of the LHC. To address this, the QUnfold library is being adapted to a distributed, high-throughput platform using tools like Jupyter, Dask, and HTCondor, offering users a more flexible and dynamic data access, as well as speeding up the overall execution time by distributing the workload. The approach is validated on Monte Carlo samples from the CMS Collaboration simulated at generator level (thus containing the parton level observables) and reconstructed with the full pipeline used in data taking conditions. A comparison between the current implementation of QUnfold - running serially on a local machine - and the distributed implementation using Dask is provided, highlighting the speedup in terms of the number of worker nodes used for the computation.

Primary authors: LORUSSO, Marco (Alma Mater Studiorum - University of Bologna); DIOTALEVI, Tommaso (INFN and University of Bologna); GASPERINI, Simone (University of Bologna & INFN)

Presenter: GASPERINI, Simone (University of Bologna & INFN)

Session Classification: Infrastructure Clouds & Virtualisation

Track Classification: Track 8: Infrastructure Clouds and Virtualizations