MAPPING PROTON QUARK STRUCTURE—LOOKING INSIDE THE PROTON: HOW DO QUARKS SPIN?

Allocation: NSF PRAC/9,440 Knh
PI: Caroline Riedl[1]
Co-PIs: Matthias Perdekamp[1], Naomi Makins[1]
Collaborators: Vincent Andrieux[1], Robert Heitz[1], Riccardo Longo[1], Angelo Maggiora[2], Artem Petrosyan[3]

[1]University of Illinois at Urbana-Champaign
[2]National Institute for Nuclear Physics (Torino, Italy)
[3]PNPI (Moscow, Russia)

EXECUTIVE SUMMARY

COMPASS (COmmon Muon Proton Apparatus for Structure and Spectroscopy) is a high-energy physics experiment that probes proton substructure by scattering high-energy pion and muon beams off of nuclear targets at CERN in Geneva, Switzerland. The experiment explores the momentum and coordinate phase space of quarks inside the proton. Observing correlations between proton spin and the intrinsic transverse momentum of quarks will shed light on the quark dynamics inside the proton, and will provide a critical test of fundamental predictions derived from Quantum Chromo Dynamics, the quantum field theory describing the nuclear force. The measurements will produce 10 petabytes of experimental and simulated data. Blue Waters' balance of processing capabilities and data storage and handling is well suited for the analysis of the large COMPASS data samples, as these require significant algorithmic processing per pion/muon-proton scattering event. In addition to raw data processing and physics-level analysis, Blue Waters allows for the detailed simulation of COMPASS detector properties and environmental effects.

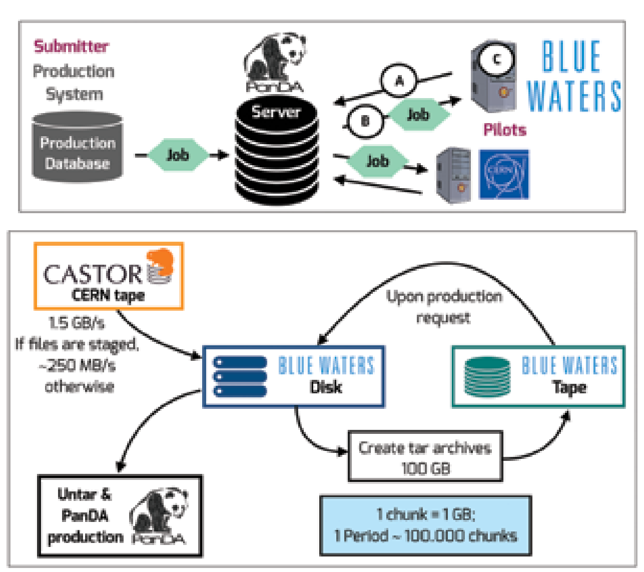

Figure 1: Top—work flow of COMPASS data production using PanDA. Bottom—management of raw experimental data on Blue Waters. Transfers between CERN and BW are handled by FTS3; transfers within BW by Globus Online.

RESEARCH CHALLENGE

Observation of the sign change of the Sivers quark distributions ("Sivers functions") in Drell–Yan scattering compared to existing measurements in semi-inclusive deep-inelastic scattering is one of the few Nuclear Science Advisory Committee [1] milestones for U.S. Department of Energy- and National Science Foundation-funded research in nuclear physics. (Sivers functions arise from correlations between proton spin and quark transverse momentum and thus appear connected to quark orbital motion inside the proton.) Such measurement requires polarization-dependent Drell–Yan data. The 2015 and 2018 Drell–Yan runs of the COMPASS experiment at CERN constitute the first measurements of this kind [2]: the negatively charged pion beam from the Super Proton Synchrotron was impinged on a target of transversely polarized protons.

With the 2016 and 2017 Generalized Parton Distributions (GPD) runs, COMPASS has added valuable observables in Deeply Virtual Compton Scattering (DVCS) to constrain spin-independent GPDs in the so-far unexplored kinematic domain between HERMES and the Jefferson Lab experiments on the one hand, and the HERA collider experiments on the other. The global community of model-fitters awaits the future COMPASS results in DVCS.

METHODS & CODES

For experimental data production, about 1 petabyte of raw COMPASS data collected at CERN has been transferred to Blue Waters. The average throughput speed is up to 1.5 GBs using the File Transfer System FTS3 [3], a bulk data mover created to distribute the Large Hadron Collider (LHC) data globally. More data—about 4 petabytes—will be transferred in 2018. The data are tared and stored on tape; upon production request, they are retrieved from tape (see Fig. 1, bottom).

For each triggered event in COMPASS, the information of the detectors is recorded by the Data AcQuisition system. The COMPASS Reconstruction Analysis Library (CORAL) software performs the transition from raw data information to physical quantities. CORAL's function is to reconstruct particle trajectories and momenta, as well as the position of vertices. The reconstructed information is stored in the form of Data Summary Trees, which are read and analyzed using the COMPASS PHysics Analysis Software Tools (PHAST).

The production of Monte Carlo data is performed in three steps: 1) The generation of signal and background events is carried out with event-generator packages. 2) For the simulation of the detector response to the physics event, a GEANT4 [4] toolkit is used based on the description of the COMPASS apparatus. 3) Simulated hits are subjected to the same CORAL and PHAST reconstruction codes as experimental data.

RESULTS & IMPACT

COMPASS accumulates a raw experimental data set of about 1 petabyte per year. A first step in the data analysis is the conversion of raw data into the physical properties of the fundamental particles created in a collision event. Approximately 6.5-million CPU hours and 50 days at the CERN computing cluster are needed for one annual production pass. Our team has adapted the COMPASS version of PanDA (Production ANd Distributed Analysis) [5], a data production and monitoring system developed for ATLAS–LHC, to Blue Waters (see Fig. 1, top). We expect to be able to process an annual COMPASS data set on Blue Waters within five days. Long-term, we are planning on four annual data sets, each with two passes.

Apart from the processed experimental data, simulated Monte Carlo (MC) data are an essential ingredient of the data analysis. Simulations of the detectors play a central role in understanding subtle detector effects and in removing background events from the data sample. With the available resources at CERN and collaborating institutions, the CPU-intensive part of the MC calculation—the simulation of the detector properties with GEANT—often cannot be afforded for extensive studies; for example, event pile-ups or time-dependent detector efficiencies. Two examples of MC activities are shown in Fig. 2: an environmental simulation that led to a modification in the experiment's shielding setup, and the simulation of the physics signal and its backgrounds for the Drell–Yan analysis.

This Blue Waters project involves students and young postdocs, and it will, in the future, attract more young physicists. It thus offers outstanding educational potential for a significant number of students and postdocs and is a step toward building a community capable of using petascale computing.

WHY BLUE WATERS

With the petascale resources of Blue Waters, COMPASS experimental and Monte Carlo data can be processed significantly faster (10 to 25 times faster compared to other computing resources available to COMPASS), which will allow the completion of publications and PhD theses in a timely manner. In the case of simulations, the data can also be generated in greater detail, delivering the high precision that keeps systematic uncertainties at the smallest possible levels. An example is the realistic pile-up of events in the particle collision. Blue Waters enables novel explorations; for example, detector resolutions in kinematic binnings and two-dimensional detector efficiency maps. In addition, Blue Waters staff provide essential guidance in terms of job flow, load distribution, data transfer, and BW-specific technical features.

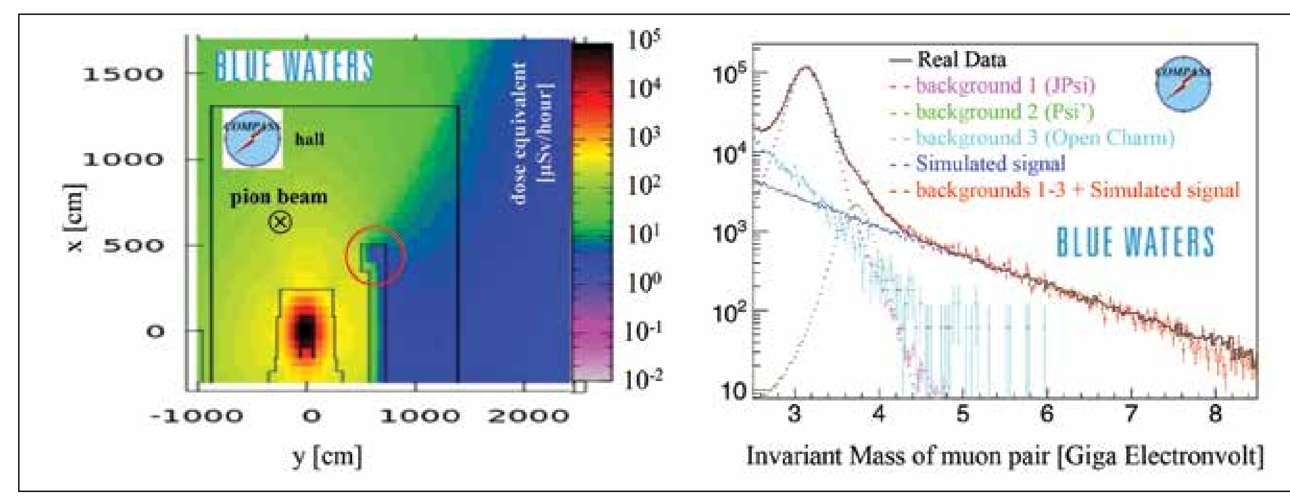

Figure 2: Left—BW FLUKA [6] simulation of hourly radiation dose in the COMPASS experimental hall. Additional concrete shielding blocks (red circle) were installed to satisfy CERN's radiation-protection requirements. Right—BW PYTHIA [7] simulation of muon pairs reconstructed with the COMPASS spectrometer in comparison with the experimental data.