DI

TN

MAPPING PROTON QUARK STRUCTURE USING PETABYTES OF COMPASS DATA

Allocation: Illinois/200 Knh
PI: Caroline Riedl[1]
Co-PIs: Matthias Grosse Perdekamp[1], Naomi Makins[1]
Collaborators: Vincent Andrieux[1], Robert Heitz[1], Marco Meyer[1]

[1]University of Illinois at Urbana-Champaign

EXECUTIVE SUMMARY

The COMPASS experiment at CERN probes proton substructure by scattering high-energy pion and muon-beams off of nuclear targets to measure 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 with 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 detailed simulation of COMPASS detector properties and environmental effects.

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 (SIDIS) is one of the few performance Nuclear Science Advisory Committee (NSAC) [1] milestones for Department of Energy- and National Science Foundation-funded research in nuclear physics. 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: The negatively charged pion beam from the Super Proton Synchrotron (SPS) was impinged on a target of transversely polarized protons. Sivers functions arise from correlations between proton spin and quark transverse momentum and thus appear connected to quark orbital motion inside the proton. The experimental measurement of these correlations will provide the first tomographic images of quarks in the proton in transverse momentum space (Fig. 1).

With the 2016 and 2017 runs, COMPASS will add valuable observables in Deeply Virtual Compton Scattering (DVCS) to constrain spin-independent Generalized Parton Distributions in the so far unexplored kinematic domain between HERMES and the Jefferson Lab experiments on the one side, and the HERA collider experiments on the other side.

METHODS & CODES

For experimental data production, the raw COMPASS data collected at CERN are transferred to Blue Waters at average throughput speed of about 1 GB/s using the File Transfer System FTS3 [3], a bulk data mover created to distribute globally the Large Hadron Collider (LHC) data. For each triggered event in COMPASS, the information of the detectors is recorded by the Data AcQuisition (DAQ) system. The COMPASS Reconstruction Analysis Library (CORAL) software performs the transition from raw data to physical quantities. CORAL's function is to reconstruct particle trajectories and momenta, and 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 reconstruction CORAL and PHAST codes as experimental data.

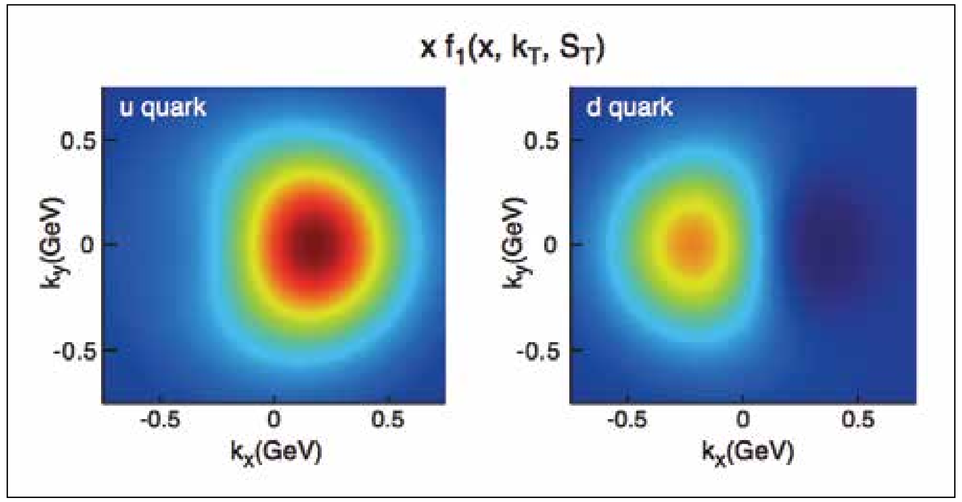

Figure 1: Quark densities in the transverse momentum plane for a transversely polarized proton (y-direction). Deep red (blue) regions indicate large negative (positive) values. Figure from [2], projecting the impact of future measurements on the knowledge of quark densities in transverse momentum space.

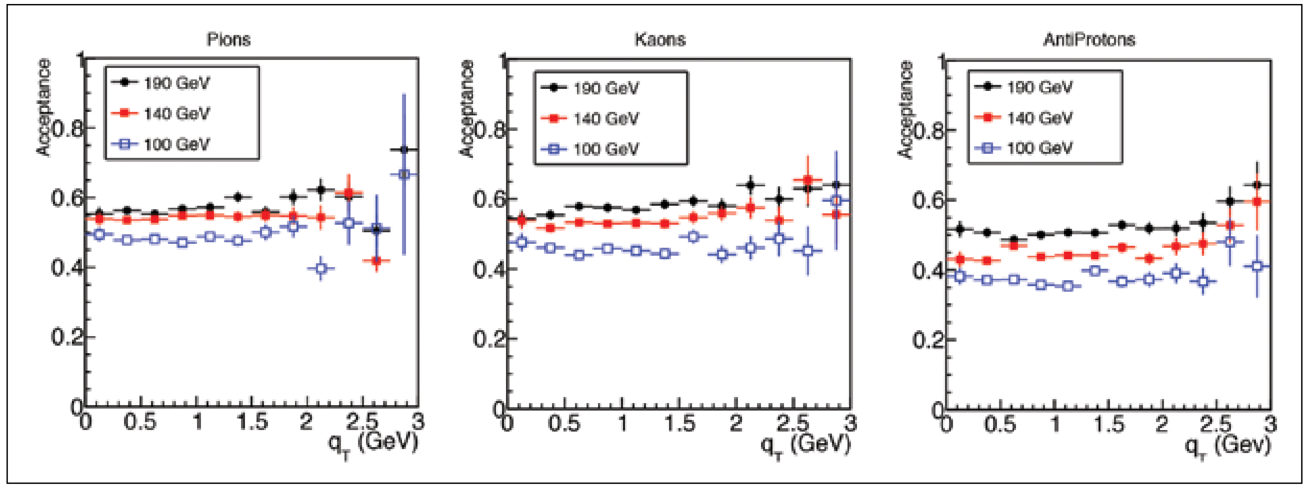

Figure 2: COMPASS acceptance for pion, kaon, and anti-proton beams simulated by the Monte Carlo generator PYTHIA [6] adapted to pion and kaon beams versus the transverse momentum of the Drell–Yan particle pair. Different beam energies are indicated by different colors.

RESULTS & IMPACT

COMPASS accumulates a raw experimental data set of about 0.75 petabytes 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. This data production is an iterative process that requires two or three passes. Approximately 6.5 million CPU hours are needed for one data production pass, which requires usually about three calendar months given the available resources at CERN. In an exploratory mass production in 2016, 5% of a yearly COMPASS data set was processed in 15 hours [5]. Since the task is scalable, a full year can be processed in about the same time if sufficient computing resources are available. This project has recently been granted millions of node hours. As of the middle of 2017, the launching of full-scale data production on Blue Waters was in preparation.

Apart from the processed experimental data, simulated Monte Carlo data are an essential ingredient to the data analysis. Simulations of the detectors play a central role in understanding subtle detector effects and removing background events from the data sample. With the available resources at CERN and collaborating institutions, the CPU-intensive part of the Monte Carlo—the simulation of the detector properties with GEANT—can often not be afforded for extensive studies, for example event pile-ups or detector efficiencies in a fine time binning. The production of simulated data for studies on a future fixed-target experiment at CERN using radio-frequency separated pion, kaon, and anti-proton beams has begun (Fig. 2).

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 towards building a community capable of using petascale computing.

WHY BLUE WATERS

Requiring both substantial Monte Carlo data production and up to three data campaigns in parallel, a timely analysis of COMPASS data appears difficult. A delay of several years between the end of data-taking in November 2018 and publication of all COMPASS results appears likely. Given the present computing resources at CERN and collaborating institutions, COMPASS might lose out to competing experiments scheduled to run in the future at FNAL, RHIC, and Jefferson Laboratory. The current projected schedule for data analysis also provides challenges to sustaining support from funding agencies for extended periods of times and for on-schedule completion of doctoral theses based on COMPASS data. With the petascale resources of Blue Waters, COMPASS experimental and Monte Carlo data can be processed significantly faster, and in the case of simulated data, also can be generated in greater detail.