Mapping Proton Quark Structure in Momentum and Coordinate Phase Space using Petabytes of COMPASS Data

Project Report for Illinois Exploratory Allocation

Caroline Riedl[1] (PI and corresponding author) criedl@illinos.edu,
Vincent Andrieux[1] (co-PI) andrieux@illinois.edu,
Robert Heitz[1] (co-PI) rsheitz2@illinois.edu ,
Marco Meyer[1] (co-PI) meyerma@illinois.edu,
Matthias Grosse Perdekamp[1] (co-PI) mgp@illinois.edu

[1] University of Illinois at Urbana-Champaign

January 30, 2017

Summary

COMPASS probes proton substructure with high-energy negative pion and polarized muon beams at CERN. These measurements provide unique access to the momentum and coordinate phase space of quarks in the proton. Of particular interest are COMPASS' first measurements of observables in the Drell-Yan process, which shed light on the currently unknown orbital motion of quarks inside the proton.

Over four years, the measurement campaign at CERN will produce more than 10 petabytes of raw data, Monte Carlo data and reduced mini-Data Summary Trees. Performance tests were carried out between June and December 2016 in context of an exploratory Blue Waters phase. The tests included the transfer of ∼100 terabytes of COMPASS data from CERN to Blue Waters and pilot mass productions of experimental and simulated data.

With the presently available COMPASS computing resources it is estimated that the full analysis of COMPASS data sets will take through 2025, well beyond the funding contracts of many of the participating institutions. The unique petascale capabilities of Blue Waters to COMPASS data analysis will make it possible to finish the data analysis within three years after the data have been acquired, well within the reach of grant proposals and within acceptable durations for thesis research of graduate students in COMPASS.

1 Key Challenges

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 DOE- and NSF-funded research in nuclear physics.

Such a measurement requires polarization-dependent Drell-Yan data, which do not exist to date. The 2015 Drell-Yan run of the COMPASS experiment at CERN constitutes the worldwide first measurement of this kind: the negatively charged pion beam from the Super Proton Synchrotron was impinged on a target of transversely polarized protons. COMPASS has previously measured the Sivers asymmetry in SIDIS in the same phase space as for the Drell-Yan experiment [2] and is therefore in the unique situation to have access to the sign change in Drell Yan without the need to involve uncertainties related to so-called TMD evolution.

COMPASS runs 2016 and 2017 are dedicated to hard-exclusive physics and the measurement of deeply virtual Compton scattering. In 2018, a second year of Drell-Yan data taking will follow. In each annual data campaign (2015–2018), COMPASS collects experimental data from May through November resulting in a raw 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 so-called data production is an iterative process that requires 2 or 3 passes over the full data set. Approximately 6.5 million CPU hours are needed for one data production pass, which requires more than 3 calendar months given the available resources at CERN. The first mass production of 2015 COMPASS data was running at CERN between late July and late October 2016.

Monte-Carlo simulations of the detectors play a central role for understanding subtle detector effects. COMPASS uses Monte-Carlo data samples that contain more signal events than experimental data by orders of magnitude, avoiding that systematic uncertainties will be dominated by statistical uncertainties from the Monte-Carlo production. Simulated data are required for example to subtract events from background processes (for example, di-muon events in the J/ψ mass region); to tune the intrinsic transverse momentum of quarks in the proton ("k_T tuning") and other kinematic parameters to match the observed transverse momentum of the detected particles; or to study the effects of QCD radiation in hadron-hadron collisions.

With the present available resources at CERN and collaborating institutions, the CPU-intensive part of the Monte Carlo – the simulation of the detector properties with GEANT [3] – can often not be afforded for extensive studies: the standard COMPASS Monte-Carlo samples do not contain effects of pile-up (particles from more than one event illuminating detectors at the same time) and simulate only events of interest, i.e. are not "minimum-bias" data samples. Simulating these effects is very CPU-intensive and cannot be afforded with the computing resources presently available at COMPASS. The situation is similar for the determination of detector efficiencies from experimental data in a fine time binning.

2 Why it Matters

The nucleon is the fundamental bound state of the strong nuclear interaction. Protons and neutrons play an important role as a laboratory for studying quantum chromodynamics (QCD), the quantum field theory of the strong nuclear force. Understanding the proton's quark and gluon substructure, the creation of nucleon mass and the origin of the proton's spin in terms of its constituent spin and orbital angular momentum are highly important questions in particle and nuclear physics. A large number of experiments have been carried out over the past three decades at accelerator laboratories in Germany (DESY, ELSA, MAMI), the Netherlands (NIKHEF), Switzerland and France (CERN) and the United States (BNL, FNAL, Jefferson Lab, SLAC) exploring the quark and gluon structure of the proton.

With generous support from the NSF, Illinois faculty have played a strong leadership role in this field. Over the past twelve years, the Illinois Nuclear Physics Laboratory (NPL) faculty has focused on the question of intrinsic transverse-momentum dependent (TMD) degrees of freedom of quarks and gluons inside the proton at Jefferson Lab, at E866 (FNAL), at HERMES (DESY), at Belle (KEK) and at PHENIX (BNL). As a result of this experimental work and the simultaneous intense theoretical effort, a QCD-based quantitative description of TMDs for the nucleon has emerged for the first time. Objects of intense scrutiny are the so-called Sivers quark distributions. They 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 allow for first tomographic images of quarks in the proton in transverse momentum space (Fig. 1).

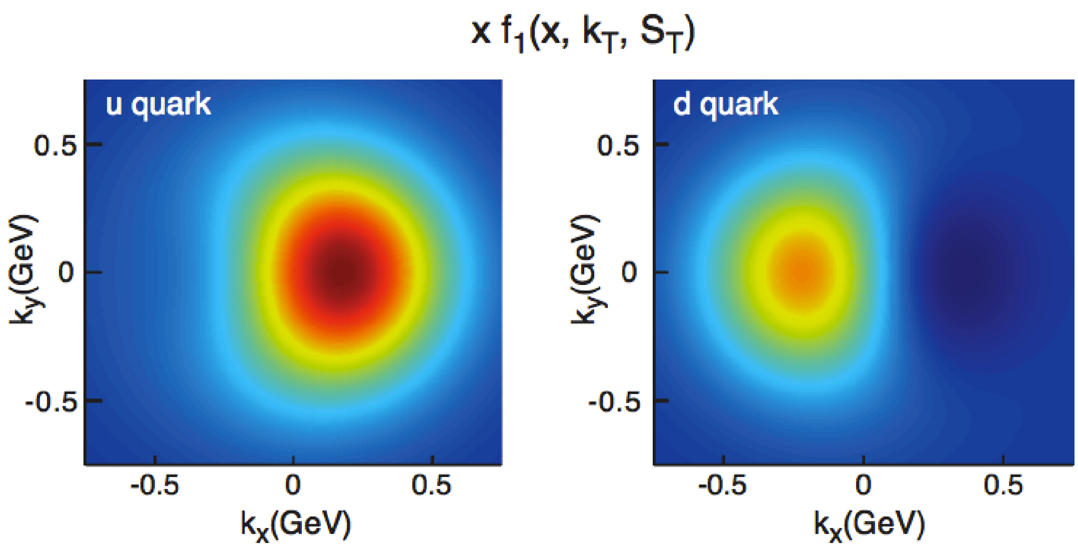

Figure 1: Quark densities in the transverse momentum plane for a transversely polarized proton (y-direction). Deep red (blue) regions indicate large negative (positive) values. Figure from Ref. [4], projecting the impact of future measurements on the knowledge of quark densities in transverse momentum space.

While the TMDs have the potential to describe the structure of the nucleon in momentum space, another set of objects of recent interest, Generalized Parton Distributions (GPDs), provide a tomography in coordinate space [5].

Very surprisingly the new QCD-TMD framework predicts that Sivers quark distributions are not universal, which is strikingly different from previously studied quark-momentum and quark-spin distributions in the nucleon [6] [7] [8]. For example, Sivers distributions are expected to change sign when comparing measurements in semi-inclusive deep inelastic lepton-proton scattering (SIDIS) with observables in polarized hadron-hadron Drell-Yan processes. Sivers functions also violate the well-known DGLAP evolution equations, which have been derived in QCD and which describe ab initio the dependence of quark and gluon distributions on the four momentum transferred by the probe particle in high-energy scattering processes testing proton substructure. The status of TMD-related experiments and theory has been summarized recently by Grosse Perdekamp and Yuan (LBL) [9].

Because COMPASS is particularly well suited to measure the expected Sivers sign change, co-PI Grosse Perdekamp and PI Riedl have joined the COMPASS collaboration at CERN. The Drell-Yan measurement in COMPASS is key to confirming the TMD-QCD framework that has emerged from our past experimental program and therefore has been generously supported by the NSF Physics Division. COMPASS has previously measured the Sivers asymmetry in SIDIS [2]. In 2015, COMPASS was measuring the polarization-dependent Drell-Yan process induced by a negatively charged pion beam on a transversely polarized proton target. The collaboration currently prepares the publication of the observables related to the Drell-Yan Sivers function. The phase space of this measurement overlaps with the previous SIDIS measurement at COMPASS. Therefore a direct comparison of the two measurements will allow to decide about the existence of the Sivers sign change without having to rely on model-dependent TMD evolution.

With the 2016 and 2017 Deeply Virtual Compton Scattering (DVCS) data, COMPASS will add valuable observables to constrain the spin-independent GPD H 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. The future COMPASS results in DVCS are awaited by the community of global fitters. Of particular interest is the magnitude of the $\cos\phi$ amplitude of the spin/charge-asymmetry at COMPASS since it has been measured to change sign between H1 (HERA) and HERMES kinematics. The DVCS cross section differential in the squared four-momentum transfer to the target will allow for sea-quark imaging of the proton.

An exploratory allocation on Blue Waters was granted in May 2016 to PI Riedl and her team. First studies on COMPASS data production on Blue Waters have recently been completed. The preliminary results were used to secure a campus Illinois allocation in 2017 to continue the commissioning of data production on Blue Waters. Work on this new allocation has recently started. The involvement of our group in Blue Waters will continue to educate students and young postdocs to work with supercomputing resources and environments.

3 Why Blue Waters?

With the need of substantial Monte-Carlo data production and the fact that 3 additional COMPASS data campaigns will follow for data taking years 2016, 2017, and 2018, a timely analysis of COMPASS data appears impossible. 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, COMPASS will not be able to deliver on the NSAC milestone in due time and might loose 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 the completion of doctoral theses based on COMPASS data on schedule.

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 generated in greater detail. The time period for publication of physics results from the 2015-2018 COMPASS data will be essentially decreased with Blue Waters.

To solve technical challenges during our exploratory allocation, it was extremely useful to be supported by the Blue Waters help team, as detailed in the next Section.

4 Accomplishments

In the half-year exploratory phase on Blue Waters that was granted with 50,000 node hours to our UIUC-COMPASS group [10], our UIUC project team has between June and December 2016 set up and tested a fully functional COMPASS data production and analysis environment. With support by CERN-IT, PI Riedl enabled and manages data transfers between CERN and Blue Waters and performed the first evaluation of COMPASS reconstruction code on Blue Waters. UIUC graduate student Marco Meyer installed COMPASS software and set up a user-friendly environment on Blue Waters. Meyer performed first bulk tests of COMPASS physics analysis software in comparison to the performance at CERN. In close collaboration with Blue-Waters IT, he found the solution to running the COMPASS calibration database on Blue Waters, which is needed for processing raw data, and has prepared the environment for the CORAL mass production of experimental data on Blue Waters. Andrieux runs the COMPASS simulation package TGeant on Blue Waters. UIUC graduate student Robert Heitz performs COMPASS Drell-Yan data analysis on Blue Waters using a PHAST program. Of the 50,000 available node hours, 22,000 were used for a CORAL pilot mass production, 8,000 for performance tests and physics analysis, and 20,000 for Monte-Carlo simulations.

Used Computational Codes. For each triggered event in COMPASS, the information of the detectors is recorded by the Data AcQuisition (DAQ) system. The transition from raw data information to physical quantities is performed by the COMPASS Reconstruction Analysis Library (CORAL) software. CORAL's function is to reconstruct particle trajectories and momenta, and the position of vertices. From a description of the position of the detectors (alignment file), magnetic field maps and calibration files, the raw data are converted into spatial hits in the detectors and energy deposits in the calorimeters. Specific algorithms merge adjacent hits into clusters, which are fitted into trajectories. For each interaction event, the converging trajectories (tracks) are matched into vertices.

The reconstructed information is stored in the form of Data Summary Trees (DSTs), which are used by COMPASS users for physics analyses. DSTs represent the data format after detector calibrations, alignment and particle tracking have been applied to the raw data. mDSTs (miniDSTs) are the most comprehensive data with all physical event information included. For a given physics analysis, users generate reduced information μDSTs (microDSTs) from mDSTs for minimizing the required processing and storage resources needed for data analysis. For Drell-Yan physics, a user might filter out events that contain at least two muon tracks. DSTs are read and analyzed using the COMPASS PHysics Analysis Software Tools (PHAST), which can be modified by each user according to their specific needs. A description of the software structure can be found in Ref. [11].

The production of Monte-Carlo data is performed in three steps. 1. The generation of physics events is carried out with event-generator packages. For Drell-Yan data, PYTHIA [12] is used for the generation of signal and background events. For hard-exclusive data, LEPTO [13] is used for the simulation of background events in lepton-proton scattering and HepGen [14] for the generation of exclusive single-photon or meson signal events. 2. For the simulation of the detector response to the physics event, a GEANT4 [3] toolkit is used based on the description of the COMPASS apparatus (TGeant or ComGeant). 3. Following GEANT4, simulated hits are subject to the same reconstruction CORAL and PHAST codes as experimental data.

Data Transfer Rates between CERN and Blue Waters (BW). In particle physics, the experimental data from particle collisions is collected and initially written to disk and later to tape on the site of the experiment. COMPASS produces large experimental data sets of the order of 1 petabyte in each year of operation at CERN. For the data analysis it is therefore essential to transfer the data efficiently to the computer system that will be used for data analysis. As a test bed for the data transfer to BW, 1 out of 9 periods of the 2015 raw COMPASS data was used: data period W12 (2 weeks of data taking). Period W12 consists of about 90,000 files. Each file is 1.1 GB in size. The files are referred to as "chunks" and the total volume of data period W12 is 95 TB. Also various mDST samples of 2015 data produced at CERN were transferred to BW, 20,000 files of each 3.5 GB, i.e. 67 TB in total. The data were transferred from CERN to BW

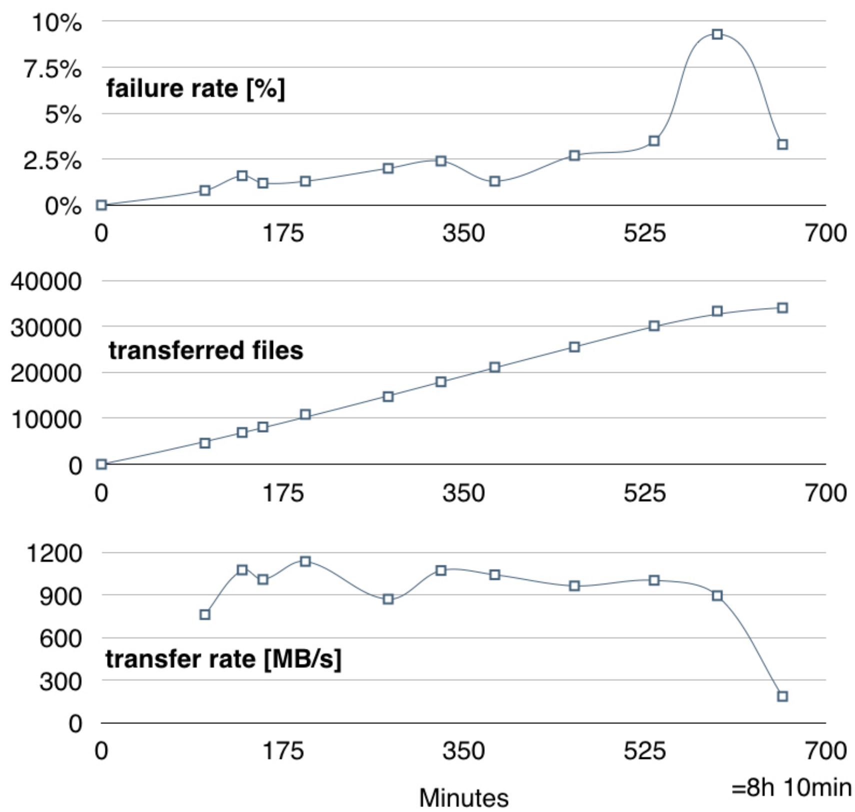

Figure 2: Data transfer example from CERN to Blue Waters: failure rate, files transferred, and transfer rate vs. time for part of COMPASS 2015 raw data period W12. One file is 1.1 GB.

using File Transfer System FTS3 [15], a bulk data mover created to distribute globally the multiple petabytes of LHC data. With FTS3, data transfers can be scheduled and monitored efficiently, maximizing the use of available network and storage resources. Most of the data at CERN are stored on tape and FTS3 includes the possibility of staging the files to disk before transferring them. We submitted up to 40,000 files for transfer at the same time, which allowed to transfer 250 to 400 files in parallel at good throughput conditions. The average observed transfer rates were between 890 and 1,200 MB/s. The failure rates were found to be below 10% and were mostly caused by temporary glitches or staging issues at the source. Most of the failed files could be successfully transferred at the second attempt. The bulk of period W12 (80 of 95 TB) was transferred within 24 hours, plus about 60 minutes for re-sending files from failed transfers. An illustration of the data transfer can be found in Fig. 2. In the example shown, 35,000 files were submitted for transfer at the same time. The average failure rate was 2.7% and the transfer took about 8 hours with an average transfer speed of 1.16 GB/s. Towards the end of the transfer, the failure rate peaks because of files that could not be retrieved from CERN-tape at first attempt. Also visible in the graph is a drop of transfer speed towards the end of the transfer because FTS3 works less efficient with too few files available for parallel transfer.

The observed data transfer rates are sufficiently high to transfer an annual COMPASS data set in 12 days from CERN to BW. These numbers were determined for transfers from CERN-tape to BW-disk (Lustre). When available, data are transferred directly from CERN-disk to BW-disk, thus avoiding the staging time from tape. We initially tried to transfer data directly to BW-tape. These transfers were found to be not as fast (about 250 MB/s) and not reliable with high failure rates. In our current computing model we therefore transfer large data sets to BW-disk (scratch) with FTS3 and then copy the data to BW-tape using the standard Globus Online tool. We have also transferred μDST data generated at BW to CERN for analysis by COMPASS collaborators. Using FTS3, we achieved a transfer rate of 470 MB/s for about 1,200 files of each about 1.5 MB of size. The total transferred data volume was 1.5 TB.

Software and Libraries Installation on Blue Waters (BW). The COMPASS software (CORAL, PHAST, ComGeant, TGeant) was found to be compatible with the BW computing facility and was set up in a common group directory, from where it can be sourced and used in different versions by all group members. Also several dependancies were installed, such as dedicated high-energy physics libraries. A master database containing calibration data was replicated from CERN on BW. This database is mainly needed by CORAL for data reconstruction.

CORAL and PHAST Performance on Blue Waters (BW). The performance of the COMPASS PHAST software on BW was evaluated by generating reduced μDSTs from CERN-produced mDSTs. In a typical example, the 4,242 CERN-produced mDSTs of COMPASS 2015 period W12 were processed by 133 jobs in parallel, with each job consuming about 13-18 node minutes on the XE nodes, see also Fig. 3. One complete data period of μDSTs could be processed in 30 minutes on BW, while it takes 1 day at CERN. This major gain in time arises from the independence of the nodes and the constant processing speed for each mDST. This is not the case at CERN, where CPUs are fair-shared between users according to priority and number of jobs already running, and are not fully dedicated to users.

Speed performances of PHAST μDST analysis were studied to determine the optimum number of μDST files to submit per node as well as a comparison between specific Drell-Yan PHAST analysis codes at BW and CERN. It was found that varying the number of submitted files per XE node with 32 processors from 28 to 32 resulted in only small overall time differences. Nine jobs

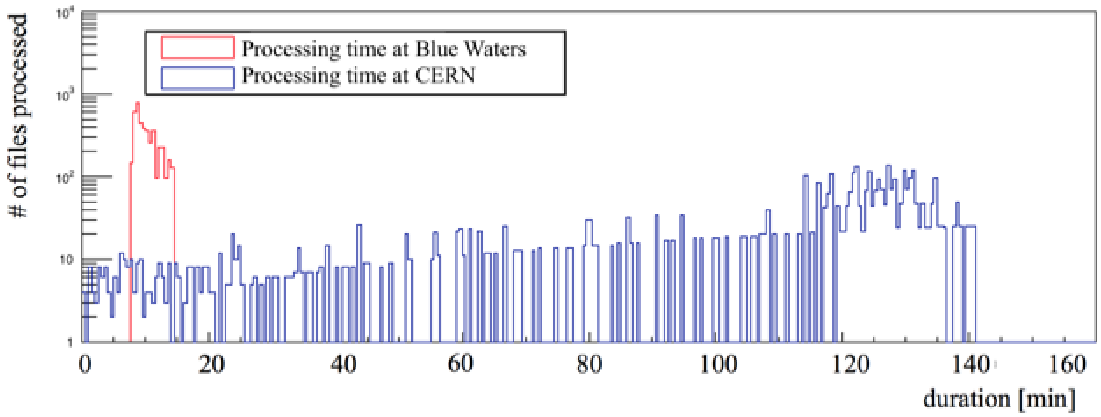

Figure 3: Performance of COMPASS analysis software PHAST for 4,242 files. Comparison between Blue Waters (red histogram) and CERN (blue histogram).

were submitted in parallel to analyze all W12 μDSTs. The CPU time per PHAST execution of one μDST file was found to be vary comparable between BW and CERN.

The full chain of experimental data analysis was carried out on 3 example chunks of COMPASS 2015 raw data by calculating COMPASS detector efficiencies. The raw data were reconstructed with CORAL and then processed with a version of PHAST that calculates the particle detection efficiency for a specified detector plane. The investigated detector was DC5, a NSF-funded large-area drift chamber built for COMPASS by our group in 2012-2015 [16]. The same 3 chunks were processed in the standard CERN computing environment with identical release versions of CORAL and PHAST. While CORAL needs 7-9 hours to process one chunk at CERN, it takes 5-10 hours on Blue Waters. PHAST runs only a few minutes per CORAL-processed chunk both at CERN and on BW. The results obtained on DC5 efficiency at CERN and on BW are compatible.

CORAL Mass Production of COMPASS Experimental Data on Blue Waters (BW). Our test bed was again period W12 of the 2015 COMPASS data with in total 90,000 chunks. The execution of a single raw chunk is not faster than at the standard computing environment at CERN. The advantage of BW over other computing facilities becomes more significant the more chunks are processed in parallel. CORAL processes data from different collision events sequentially. We have found that the best way to take advantage of the parallel architecture of BW is to send data chunks of about 1.1 GB to different processors. The BW XE (XK) nodes provide 32 (16) processors with each 2 GB memory. Since the COMPASS software is based on sequential CPU codes, we not expect to profit from using the the GPU-accelerated XK nodes.

Figure 4 shows our CORAL job organization for XE nodes: chunks are submitted to the BW grid in packages of 31 using one “aprun” command. The one processor not employed with CORAL execution is occupied with the parallel-command processor script “pcp” that we received from the BW support team and that is used to distribute jobs efficiently over the computing grid. 33 packages are submitted to the grid using one “qsub” command. As recommended by BW personnel, each of the qsub job scripts contains one aprun command at the very beginning to start the COMPASS calibration database on one of the grid nodes. There is one instance of the database running for 31×33=1,023 chunks and it is terminated after the last package of 31 chunks of that job has completed. In total, 44 jobs are submitted in parallel to process 45,000 chunks. This corresponds to 50% of period W12 (1 week of data taking). We can process 1,023 chunks using 500 node hours

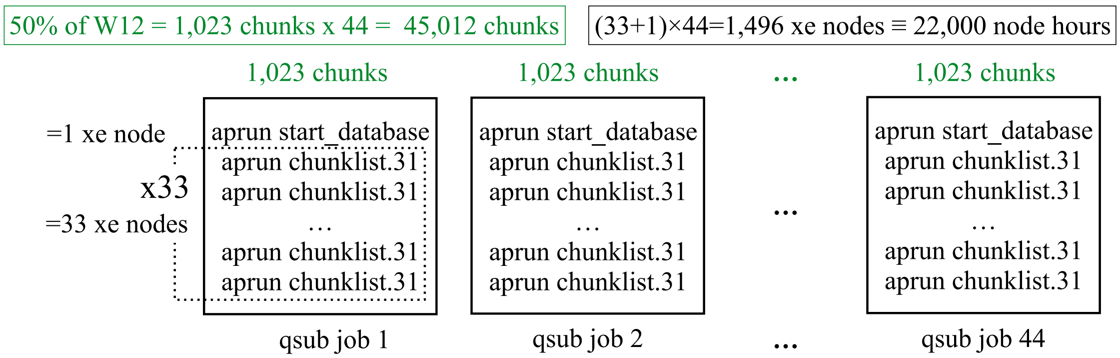

Figure 4: Scheme of job submission to Blue-Waters grid for mass production of COMPASS experimental data. The jobs run in parallel and take about 15 hours each.

(1 job) and 45,000 chunks using 22,000 node hours (44 jobs).

At CERN, COMPASS data period W12 was processed with up to 9,000 chunks in parallel. These 9,000 chunks were finished after 9 hours and processing the entire W12 period took about seven days at CERN. This results in an average of about 7 seconds per chunk. Using the resources from our exploratory phase on BW, we carried out a mass-production pilot run in that 50% of period W12 (45,000 chunks) were processed with about 1.2 seconds per chunk. This is about 6 times faster than at CERN. With more node hours available, the average time per chunk can be as low as 0.06 seconds if one entire year of data taking is processed in parallel. This is more than 100 times faster than at CERN. The scalability of COMPASS data production on BW is demonstrated in Fig. 5.

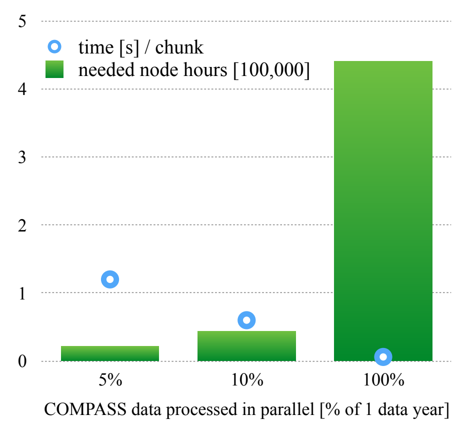

Figure 5: Scalability of COMPASS data mass production on Blue Waters. The case of 5% of was tested during an exploratory phase (50% of W12) .

The job submission to the BW grid will be further optimized towards COMPASS data mass production. In reality, the 44 jobs of our pilot mass production each consuming 34 nodes were not processed in parallel due to queuing times. We have found that the overall queuing time can be significantly decreased by submitting only one big job instead of 44 small jobs, which improves the priority coefficient of our jobs. We will continue to study the optimization of CPU distribution between different chunks. The execution (real) time per chunk increases by about 50% (from ∼10 to ∼15 hours) when $n = 31$ chunks are processed on one node instead of only one, $n = 1$. This increase is attributed to multi-threaded libraries that CORAL uses and that are not sufficiently fed when too many chunks run on one node, and/or to the memory available per chunk (2 GB for $n = 32$ and 64 GB for $n = 1$). We have tried to slightly lower the number of chunks on one node to $n = 28$, without significant change in execution time. Sending only one chunk per node results in a punishment in terms of node-hour charging. Therefore we decided to use $n = 31$ for the pilot mass production. The optimum solution might lie in between $n = 1$ and $n = 31$ and might also include splitting the chunks of each ∼1.1 GB into smaller portions to provide more CPU memory to CORAL.

We are planning to manage the mass productions with PanDA (Production ANd Distributed Analysis) [17], a data production and monitoring system developed for ATLAS at the LHC. COMPASS has started negotiations with the PanDA group and a first test production at CERN was successfully completed in July 2016. Two computing specialists from the PanDA team are presently supporting the UIUC team in setting up and testing the PanDA pilot on BW.

Monte-Carlo (MC) Data Production on Blue Waters (BW). In a pilot simulation campaign on BW, 5,000 events generated with PYTHIA8 were processed through a GEANT4 model of the COMPASS spectrometer (TGeant). While this job takes about 3-4 hours at the national computing center in Lyon (France), the exact same task takes 20 minutes on BW with charge of half a node hour. This corresponds to an improvement factor of 12. Of the three steps of MC simulation, GEANT4 is by far the most CPU-intensive task and is therefore usually the limiting factor in the volume of simulated data produced.

During the last month of the exploratory allocation, about 80 million PYTHIA8 events were generated and processed on BW. This simulated data sample serves as important input for the peer-reviewed publication of the 2015 COMPASS data, which is currently being prepared. The MC data are used to determine the so-called dilution factor of the transversely polarized COMPASS target, i.e. the fraction of polarizable protons in the target. The precision determination of the dilution factor involves the calculation of the Drell-Yan cross section at different kinematic points.

COMPASS data simulations will be significantly advanced by the usage of BW. The generation of larger MC data samples will allow us to keep systematic uncertainties at a minimum level. An example is the determination of the experimental acceptance, which is needed for example for Drell-Yan and DVCS cross-section measurements. Blue Waters will allow the exploration of the PYTHIA parameter space at much finer granularity than before. For a meaningful acceptance correction, extensive study and tuning of the event generator parameters has to be performed. Processing the full MC chain for many different combinations is usually not affordable due to the expensive CPU consumption of GEANT4. Usually this improvement of the tuning is assessed by a re-weighting method to avoid GEANT4.

Some aspects of the experimental conditions can only be simulated if sufficient CPU resources are available. Examples are the simulations of pile-up and minimum-bias events, both of which are not included in the COMPASS standard MCs. Pile-up occurs when particles from more than one event hit the detectors at roughly the same time. Minimum-bias simulation includes all possible event types, as opposed to only simulating events of interest (such as pure Drell-Yan events). We estimate the CPU consumption to increase by a factor of ∼15 and the file volume to increase by a factor of 16 when pile-up is included. Simulation of minimum-bias events is more CPU expensive by a factor of ∼25, increasing file sizes by about a factor of 8. The so-far generated events did not include pile-up and minimum bias.

References

[1] *The Frontiers of Nuclear Science: A Long-Range Plan*, The DOE/NSF Nuclear Science Advisory Committee Working Group, December, 2007, arXiv:0809.3137.

[2] COMPASS Collaboration, *Sivers asymmetry extracted in SIDIS at the hard scale of the Drell-Yan process at COMPASS*, arXiv:1609.07374 [hep-ex] and submitted to Physics Letters B.

[3] J. Allison, *Facilities and Methods: Geant4 - A Simulation Toolkit*, Nuclear Physics News, **17**:2, 20-24.

[4] A. Deshpande *et al.*, *Electron Ion Collider: The Next QCD Frontier*, BNL-98815-2012-JA, JLAB-PHY-12-1652, arXiv:1212.1701, December, 2014.

[5] M. Diehl, *Introduction to GPDs and TMDs*, Eur. Phys. J. A (2016) **52**: 149.

[6] COMPASS Collaboration, *The spin structure function* g_1^p *of the proton and a test of the Bjorken sum rule*, Physics Letters B **753** (2016) 18.

[7] COMPASS Collaboration, *Longitudinal double spin asymmetries in single hadron quasi-real photoproduction at high* pT, Physics Letters B **753** (2016) 573-579.

[8] COMPASS Collaboration, *Leading-order determination of the gluon polarisation using a novel method*, arXiv:1512.05053 [hep-ex] and submitted to Eur. Phys. J. C.

[9] F. Yuan and M. Grosse Perdekamp, *Transverse Spin Structure of the Nucleon*, Ann. Rev. Nucl. Part. Sci. **65** (2015) 429-456.

[10] M. Meyer, *Blue Waters, A Petascale Computer Facility for the reconstruction of CERN COMPASS-II data*, talk given at the 22nd International Spin Symposium, Urbana-Champaign, September 26-30, 2016.

[11] COMPASS Collaboration, *The Compass Experiment at CERN*, NIM A **577** (2007) 455-518.

[12] T. Sjoestranda et. al., *An Introduction to PYTHIA 8.2*, arXiv:1410.3012 [hep-ph].

[13] G. Ingelman, *LEPTO version 6.1: The Lund Monte Carlo for deep inelastic lepto - nucleon scattering*, Conf. Proc. C **911029** (1991) 1366.

[14] A. Sandacz and P. Sznajder, *HEPGEN generator for hard exclusive leptoproduction*, arXiv:1207.0333 [hep-ph].

[15] H. Riahi, M. Salichos, O. Keeble et al., *FTS3: Quantitative Monitoring*, J. Phys.: Conf. Ser. 664 (2015) 062051.

[16] R. Heitz, *A Large-Area Planar Drift Chamber for the COMPASS experiment at CERN*, talk given at the 22nd International Spin Symposium, Urbana-Champaign, September 26-30, 2016.

[17] A. Klimentov et al., *Beyond ATLAS: A Scalable Workload Management System For Data Intensive Science*, talk given at the International Symposium on Grid and Clouds Academia Sinica, Taipei (Taiwan), March 2014.