Monte Carlo simulations in a GRID environment for the COMPASS experiment

AMOROSO ANTONIO

Department of General Physics "A.Avogadro" and I.N.F.N. Torino

ON BEHALF OF THE COMPASS COLLABORATION

In HEP experiments, like COMPASS at CERN, the production of Monte Carlo simulations is an important step to investigate several aspects of the experimental apparatus, minimizing the instrumental errors due to target acceptance, geometry and the errors deriving from the analysis cuts. Such a study is particularly important for the measurement of spin observables, where the control of the instrumental asymmetries is needed and larger statistics is required to get small error bars. Producing such a large amount of data requires considerable computing power. A way to satisfy these requirements is the new GRID technology. Since the beginning of 2004 the first truly accessible GRID infrastructure, deployed by INFN, is available. I will report on the new tests and results obtained using the INFN GRID computing resources to simulate the Λ polarization in the semi-inclusive DIS of muons on a deuterium target polarized both longitudinally and transversely. These results will be compared with the available experimental results from COMPASS.

PACS: 80.89.20.Ff
Key words: GRID, Monte Carlo, Λ polarisation, COMPASS

1 Introduction

COMPASS (COmmon Muon and Proton Apparatus for Structure and Spectroscopy) [1] is a complex experimental apparatus assembled by an international collaboration of more than 30 institutions. COMPASS is a fixed target experiment in the North Area site at CERN, it uses beams produced by the SPS accelerator. The purpose of the experiment is the study of the structure and the spectroscopy of hadrons using different high intensity beams of muons and hadrons with energies ranging from 100 to 300 GeV. For the muon physics program a polarized muon beam of 160 GeV/c impinges on a ^{6}LiD polarized target. The use of the Λ weak decay as a polarimeter to study longitudinal and transverse parton distribution functions (PDF) inside the nucleon requires accurate measurements of the asymmetries in the polarization. To perform these measurements the spectrometer needs a good spatial and symmetrical particle identification at rates of about $2 \cdot 10^8$ particles/spill. Dedicated trigger and a new concept of read-out electronics complete the performance of the spectrometer. The experiment started running for physics in the year 2002. Up to the last 2004 shift it collected about 700 TB of data.

2 The experimental apparatus

The COMPASS apparatus is composed of two spectrometers, the first one is called Large Angle Spectrometer, while the second is the Small Angle Spectrometer. The two spectrometers have a similar structure and are based on dipole magnets

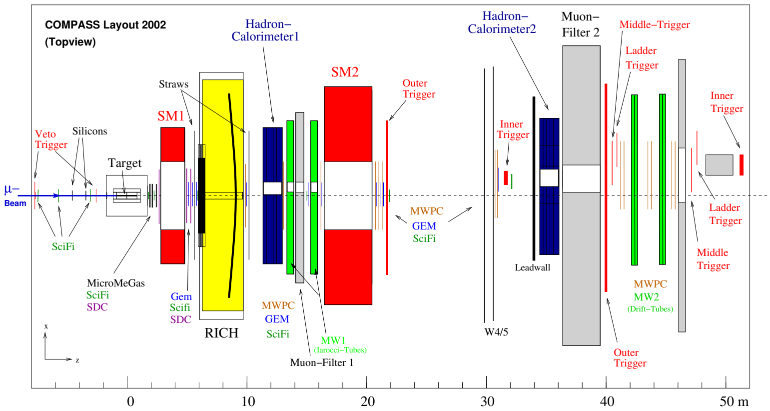

Fig. 1. Schematic view of the COMPASS setup

called SM1 and SM2 respectively (see Fig.1). Each spectrometer is equipped with several tracking devices with different spatial and timing resolutions. Beam particles are tracked using scintillating fiber and silicon detectors, which provide good spatial and time resolutions and are suitable for high particle fluxes. Outside the beam spot region gas multiplication detectors are used. Particles emitted with small angles are detected using Micromegas detectors in the region between the target and SM1, and Gas Electron Multiplier detectors in the rest of the apparatus. Large area tracking is performed using drift chambers, straw detectors, MWPC chambers and plastic Iarocci tube detectors. Drift chambers and straws provide the tracking of large angle particles before and behind the first spectrometer magnet SM1, while MWPCs represent the main tracking system of the second, small angle spectrometer. Two large area drift chambers, mounted in front of the second muon absorber, provide additional track points. Particle identification is performed using a RICH detector located between SM1 and SM2. Muon identification is performed by means of two thick hadron absorbers, called muon filter 1 and 2. Tracking stations before and behind the absorbers allow to separate muons and charged hadrons, since the latter are stopped and do not produce signals in the detectors downstream of the absorbers.
The target consists of two cells of 60cm length each and 3cm diameter, separated by 10 cm. The two cells are polarized in opposite direction, every 8 hours the spin directions in both cells are reversed by rotating the magnetic field direction. This is done to ensure that the variations of flux and acceptance cancel out in the

asymmetry calculations.

3 The importance of Monte Carlo simulations

In high energy physics experiments the production of Monte Carlo simulated events is an important step to investigate several aspects of the experimental apparatus. Asymmetries due either to acceptance, reconstruction efficiency and instrumental geometry can dilutes the signal for the study of the spin variables. Our purpose was the production of a very large sample (10^8) of Monte Carlo events, that is almost 10 times longer than the amount of the corresponding events collected in the data taking periods.
By producing simulated events it is possible to correct the instrumental error due to the geometry of the target and of the experimental apparatus and the errors deriving from the analysis cuts. We can also test the algorithms used to reconstruct the primary and secondary vertices, which play a fundamental role in the Λ identification.

4 The GRID Project

The huge demand of computing power required by the new high energy physics experiments - like ALICE, ATLAS, CMS and LHCb - at the Large Hadron Collider (LHC) at CERN cannot be satisfied by traditional computing systems and hence the development of a distributed computing environment is required to combine and share the heterogeneous resources of a large number of collaborating institutions and computing centers. The COMPASS experiment has a lot of computing resource installed at CERN and Karlsruhe. It also owns a "farm" installed in the computing center of INFN Torino. The COMPASS Torino group use this farm to analyze data collected during the COMPASS data taking. The main tasks of the farm are the downloading and the analysis of files (the mDST) preprocessed at CERN. The main Monte Carlo task is to produce the simulated events ($2 \cdot 10^8$ events) required by the analysis of experimental data. To generate this big amount of data a single CPU of the COMPASS farm would need for $1.7 \cdot 10^5$ hours (CPU PIII 1GHz, 512MB RAM) and with 60 CPUs available, the production would take about 120 days. In this period the COMPASS computing resources should be exclusively devoted to Monte Carlo simulations. This scenario is impossible because the farm is currently being used for the analysis of real data. It is also used to test the analysis software. It is thus clear that the COMPASS computing resources that can be devoted to the production of Monte Carlo events are insufficient to fulfill the demand.

4.1 The LHC Computing Grid

The purpose of the LHC computing Grid Project (LCG) is to deploy the computing infrastructure for the simulation, processing and analysis of data for the LHC collaborations. The requirements for LHC data handling are very large, in terms of computational power, data storage capacity, data access performance and the associated human resources for operation and support. From 2002 through 2005, LCG will develop and prototype the computing services and deploy a series of computing data challenges of increasing size and complexity to demonstrate the effectiveness of the software and computing models selected by the experiments. This version did run in 2004 with the main goal of providing a stable service. The current installed software release is LCG2 2.5.0

4.2 The INFN Production GRID

The INFN-Grid project started in 1999 to develop the first Italian Grid integrated with various Grid infrastructures all over the world. The INFN Grid comprises more than 20 sites (see Fig.2) including the most important Italian universities and, although primarily focused on the development of computing infrastructures for physics, it has been, from the beginning, open to other research subjects (biomedicine, earth observation, etc.) and to industry. It is a successful example of collaboration between physicists, software engineers, computer professionals, computer scientists and Italian industries. In Fig.3 the relation between INFN Production Grid and LCG site is shown. The two grid infrastructure are separated but they are sharing many sites (CNAF, Torino, Milano, LNL). The GRID has a three level (Tier) structure. The main site (Tier0) is CERN that provides the new release and shares the major part of computing resources and disk space; in every country there is one regional center (Tier1), CNAF for Italy, and many other site (Tier2). The currently installed release is INFN Production Grid 2.6.0. This release and the previous one are superset of the LCG ones. The INFN Production Grid shares up to 1000 CPUs and 2PB of disk space.

4.3 The GRID infrastructure

The systems and services are logically clustered into different hosts types, each providing a particular service. The core of the GRID is the Middleware (Information Service, Data Management services and Resource Broker). It provides all the services to locate and report on the status of Grid resources, to find the most appropriate resources to run a job requiring certain data access and to automatically perform data operations necessary before and after a job is run. The Resource Broker (RB) schedules user jobs on appropriate Computing Elements (CE). Every regional site provides all the Middleware services. In every GRID site is installed a computing farm with at least one User Interface, one Computing Element and one Storage Element, that constitute the computing and storage power of the Grid.

– The User Interface (UI) hosts all client programs allowing the user to interact with the Grid. This is the place from where a user retrieves information about the grid status and submits jobs. The UI is the initial point of access to the Grid, where users have a personal account and where the user certificates are stored. From the UI, a user can be authenticated and authorized to use the Grid resources. It provides a Command Line Interface to perform some basic grid operations:

Fig. 2. Map of INFN Production Grid and Grid.it

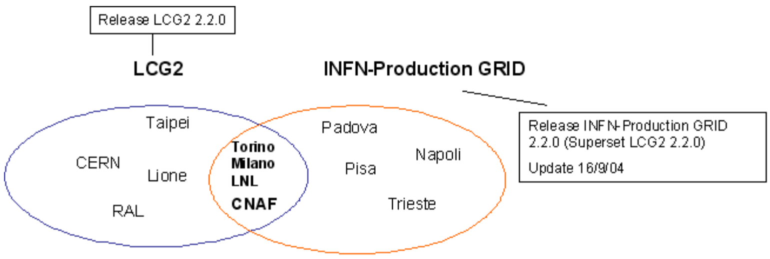

Fig. 3. Relation between INFN Production Grid and LCG sites

- ∗ submit a job for execution on a Computing Element;
- ∗ list all the resources suitable to execute a certain job;
- ∗ replicate and copy files;
- ∗ cancel one or more jobs;
- ∗ retrieve the output of one or more completed jobs;

* show the status of one or more submitted jobs.

One or more UIs are available at each GRID site.

- The Storage Elements provide storage resources on the Grid. A Storage Element (SE) provides uniform access and services to large storage spaces. The SE controls large disk arrays and mass storage systems. Each site provides one or more SEs.
- The Computing Elements (CE) supply CPU power. CEs are typically clusters of Worker Nodes (WN) and a front-end node that manages the WNs via a Local Resource Management System (LRMS). In LCG-2 and INFN Production Grid the types of LRMS supported are Open Portable Batch System (OpenPBS) and Load Share Facility (LSF). A Computing Element (CE) is defined as a Grid batch queue. Each GRID site runs at least one CE and a farm of WNs behind it.

In Fig.4 the structure of a GRID with different GRID Elements is shown.

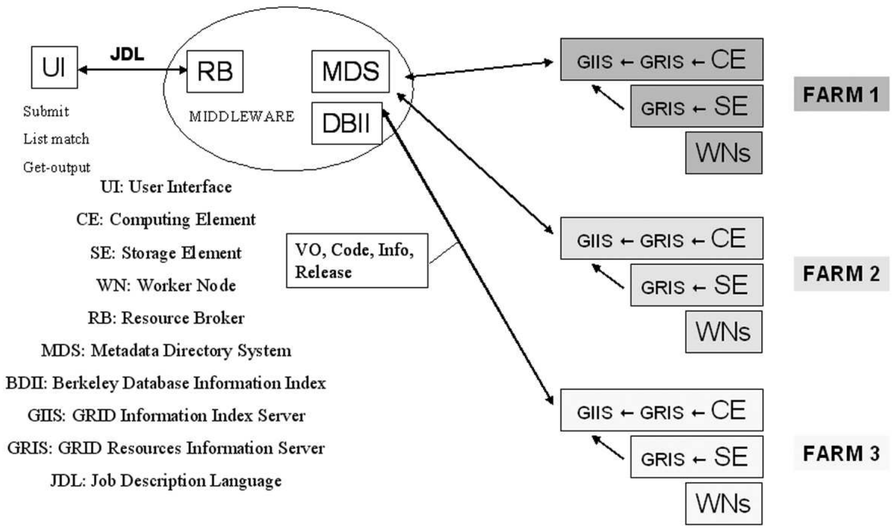

Fig. 4. Structure of a GRID

5 Simulation software

The COMPASS experiment uses different software packages to produce Monte Carlo simulated events. The simulation chain is structured in four steps: event generation, simulation and hit production, digitalization and reconstruction, event filtering and output. The event generator, (LEPTO) is a Monte Carlo [2] program to simulate complete lepton-nucleon scattering events and cross sections. The simulation program for COMPASS experiment is COMGEANT based on GEANT3. It is used to trace the particles through the setup, to simulate the detectors response and to have a graphic representation of both the setup and the particle trajectories. CORAL is the COMPASS Reconstruction and AnaLysis program [4], and contains all the software necessary to reconstruct an event both from raw data and from simulated data. For the simulated data it provides also the event digitalization. The software PHAST (PHysics Analysis Software Tools) [5] allows to filter the reconstructed events, and supplies the environment for the development of the analysis programs. It produces the mDST files suitable for further analysis.

6 COMPASS@GRID

In the INFN Production GRID (GRIDIT and EGEE) and LCG2 many Resource Brokers are available:

– INFN Prod. GRID - edt003.cnaf.infn.it - grid014.ct.infn.it - prod-rb-01.pd.infn.it

– LCG2 - wn-02-32-a.cr.cnaf.infn.it

– EGEE - egee-rb-01.cnaf.infn.it

During the submission period different RB are used, to prevent the COMPASS Monte Carlo data production from stopping altogether during the period of unavailability of a specific RB (upgrade to new release, power cuts, network problem, etc.). To select a given RB it is sufficient to specify an appropriate configuration file in the argument of the submit command. In Tab.1 the report of the submission

SITE	Resource Broker	Submitted jobs	Aborted jobs
CNAF	edt003.cnaf.infn.it	4746	124
CATANIA	grid014.ct.infn.it	563	91
EGEE (CNAF)	egee-rb-01.cnaf.infn.it	269	7

Table 1. Submitted and aborted jobs by Resource Broker

for different RB is shown. The CNAF RB was mainly used, the high percentual of aborted jobs for the Catania RB was due to unavailability of the chosen GRID resources during the period of submission via Catania RB.

6.1 Results

In Tab.2 the efficiency of the Monte Carlo production is shown. On the whole were submitted 5808 jobs, 4746 jobs via the CNAF RB (edt003.cnaf.infn.it), 563 jobs

SITE	CE	Submitted jobs	Aborted jobs
TORINO	grid008.to.infn.it	2744	92
LNL	t2-ce-01.lnl.infn.it	2093	115
FERRARA	gridrb.fe.infn.it	286	5
CAGLIARI	grid002.ca.infn.it	433	11

Table 2. Submitted and aborted jobs for transverse production.

via the CATANIA RB (grid014.ct.infn.it) and 269 jobs via the EGEE RB (egee-rb-01.cnaf.infn.it). The number of simulated events is $1,07 * 10^8$ as requested from the analysis. The 3,7% (214) of the jobs are aborted because of GRID problem, mainly for RB problems. However from the COMPASS point of view there was a further 4% (223) of unreadable `.root` output files, even if these jobs were "DONE with SUCCESS" in the GRID context. About 30 jobs/day for 162 days from 19/06 to 27/12 with a stop from 30/7 to 3/9 for summer holidays and for the upgrade to the new release (INFNGRID-2.1.0 to INFNGRID-2.2.0) with a peak of 126 jobs/day (16/07/2004) on the whole have been submitted.

7 Event selection

In the analysis, Λ particles are identified from their weak decay $\Lambda \to p\pi^-$, which results in a "V^0"-like vertex. The incoming neutral particle is not detected, and appears only when it decays in two particles with opposite charges, which are deflected in opposite directions by the spectrometer magnets. The selection criteria of the Λ is mainly based on the requirement that the secondary vertex is downstream of the primary vertex. The events of interest are selected by requiring the complete reconstruction of the main interaction vertex (beam and scattered muons, with associated momenta). The vertex must be contained inside the target cell volume. In addition, at least one secondary vertex reproducing the Λ decay signature must be present in the event, with a coordinate along the beam direction -35< z <140. Such sample still contains a background, due to other decay like Kaons into $\pi^+\pi^-$ pair and accidental track associations, and some additional cuts are required in order to reject the background sources and clean the Λ signal.

7.1 Other selection criteria

- The decay product $V^0 \to T^+T^-$ must be bridged through at least one of the spectrometer magnets to measure their momenta;
- the momentum of the decay particles is required to be > 1 GeV/c, to provide better mass resolution;
- the decay hadrons traverse less than 10 radiation lengths, in order to reduce the probability of associating a muon track with the candidate proton particle.

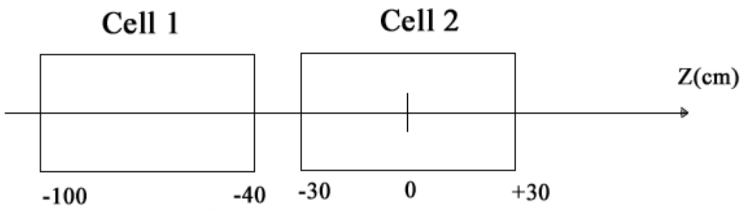

Fig. 5. Position of the target cell along the z axis in the Main reference system

– the candidate scattered muon must traverse at least 30 radiation lengths.

– A cut on the minimum transverse momentum $p_T > 0.023$ GeV/c of the positive decay particle with respect to the Λ direction is applied to reduce the background from e^+e^- pairs.

Some of the reconstructed V^0 vertices are generated by the accidental combination of uncorrelated tracks ("fake" vertices). Such combination are suppressed by requiring cos$\alpha < 0.9999$, where α is the angle between the particle direction and the line connecting the two vertices. The Λ vertex selection criteria have been checked in the Monte Carlo simulations to get numerical estimates of the rejection efficiencies. The Monte Carlo information allows to discriminate between "real" and "fake" secondary vertices, and to evaluate the contamination due to combinatorial background. These criteria, validated with the Monte Carlo analysis, have been applied to real data to extract the polarisation.

8 Results of the analysis of Monte Carlo events

In this section the results of the analysis of the Monte Carlo events are reported. A part of the sample used in this analysis comprises $5.2 \cdot 10^6$ events generated by 2600 GRID jobs is shown below. $1.3 \cdot 10^5$ events of the sample fulfilled all the selection criteria, with an overall efficiency of 0.25%.
The invariant mass derived from the analysis of Monte Carlo simulated events passing all the selection cuts is shown in Fig.6 in the Λ hypothesis. The statistical accuracy for the Monte Carlo sample is almost 10 times better than that of each of the data taking periods. To increase the Λ statistics the Monte Carlo data set has been "artificially enriched" by keeping only events with at least one Λ in the final state. This includes both directly produced and secondary Λ. The simulated DIS process does not include polarization effects and allows the study the contribution of the experimental acceptance to the measured spin observables. The effect of the non-ideal apparatus acceptance (see Fig.7) is well corrected by the algorithm used to extract the polarization (see Fig.8). That is well fitted with a constant with a value of χ^2 lower than 2.

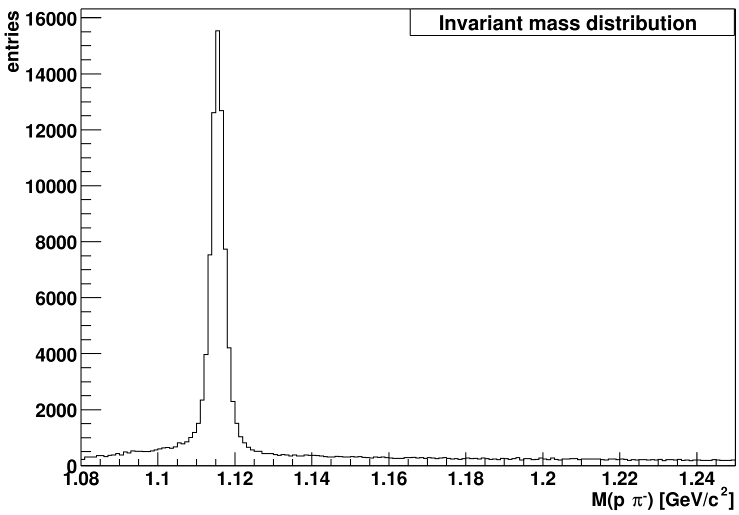

Fig. 6. Λ invariant mass passing all the selection cuts.

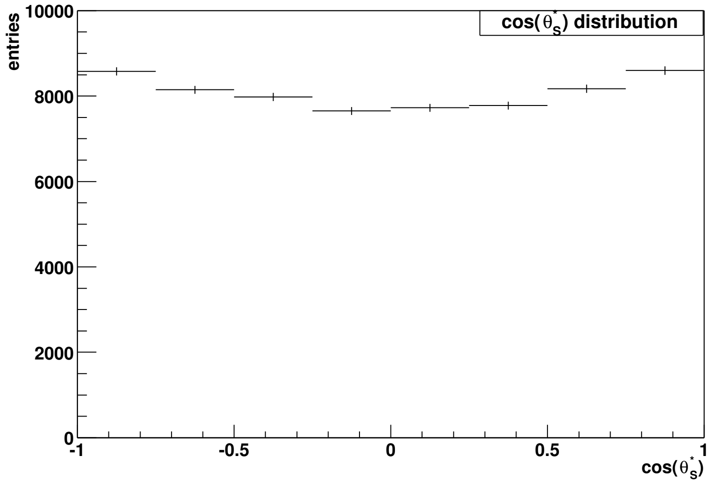

Fig. 7. $\cos\theta_s^*$ distribution for Monte Carlo sample.

9 Conclusion

Modern High Energy Physics experiments require a large amount of simulated events to reach the desired statistical accuracy. The use of GRID resources is a

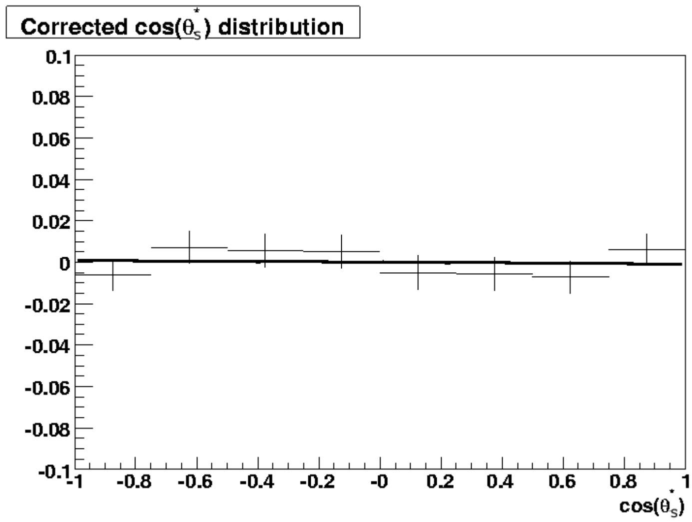

Fig. 8. Corrected $\cos\theta_s^*$ distribution for Monte Carlo sample.

powerful method to produce this large amount of Monte Carlo events. In this particular case we use this facility to produce $2 \cdot 10^8$ unpolarised SIDIS Λ events with transversely polarized target configurations. Such number of events is sufficient to calculate the effect of the apparatus acceptance for the measure of the Λ polarisation with an accuracy four times better than real data. The full data sample was produced during a period of 160 days using an average of 50 CPU. The framework described in this paper is the first step towards a complete implementation of the Monte Carlo simulation chain and of the data analysis for the COMPASS experiment on the GRID infrastructure.

This work has been done with the support of INFN and PRIN 2003029177_003 year 2003

References

[1] COMPASS, A Proposal for a COmmon Muon and Proton Apparatus for Structure and Spectroscopy, (CERN/SPSLC 96-14, SPSLC/P297, 1996).

[2] LEPTO Documentation page: *http://www3.tsl.uu.se/thep/lepto/*

[3] GEANT Documentation page: *http://wwwinfo.cern.ch/asd/geant/*

[4] CORAL Documentation page: *http://coral.cern.ch*

[5] PHAST Documentation page: *http://gs.home.cern.ch/ges/phast/*

[6] A. Delgado Peris et al. LCG-2 Userguide

[7] INFNGRID Project Home Page: *http://grid.infn.it/*

[8] LHC Computing Grid Project Home Page: *http://lcg.web.cern.ch/LCG/*