

UNIVERSITÀ DEGLI STUDI DI TORINO

Facoltà di Scienze Matematiche, Fisiche e Naturali

Dottorato di ricerca in Fisica XVII Ciclo

Spin asymmetries for the
reaction $\mu\bar{D} \rightarrow \mu'\Lambda X$ at
COMPASS: MonteCarlo
simulations using *GRID*

Candidato

Antonio Amoroso

Relatore

Prof.ssa Maria Pia Bussa

Coordinatore del ciclo

Prof. Ezio Menichetti

2002-2004



0.1 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 20 Institutions. Two research physics programs have been planned. One centered on muon physics and the other one on hadron physics. COMPASS is a fixed target experiment in the North Area site at CERN. It uses beams produced by the SPS accelerator. The experiment is placed in the EHN2 hall (Bldg. 888) in the Preveessin (F) site of CERN. The purpose of the experiment is the study of the structure and the spectroscopy of hadrons using different high intensity beams of lepton and hadrons with energies ranging from 100 to 200 GeV.

The experiment aims to collect large samples of charmed particles. From the measurement of the cross section asymmetry for open charm production in deep inelastic scattering of polarized muons on polarized nucleons, the gluon polarization ΔG will be determined and compared with the available theoretical predictions.

Hadron beams are used to study the semi-leptonic decays of charmed doubly charmed baryons. Both measurements will allow to study fundamental problems regarding the hadron structure and to test Heavy Quark Effective Theory (HQET) calculations. Moreover the production of exotic states, which are foreseen by the QCD but have not yet been established, will be investigated. Also the transverse spin distribution function, whose theoretical relevance has only recently been recognized, will be examined.

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 last 2004 shift, on November 10th 2004, it collected about 700 TB of data.

0.2 The importance of MonteCarlo simulations

In High Energy Physics experiments the production of MonteCarlo simulated events is an important step to investigate several aspects of the experimental apparatus. Simulated events are used mainly for the following reasons:

- to correct the acceptance of the experiment;
- to test the reconstruction software;
- to estimate the reconstruction efficiency;
- to study systematic effects.

By producing simulated events it is possible to minimize the instrumental error due to target and experimental apparatus geometry 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.

0.3 Why GRID?

The huge demand of computing power for the coming in 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 centres.

The COMPASS experiment has a lot of computing resource installed at CERN and Karlsruhe. It also owns a “farm” installed in the compute centre of INFN Torino. It comprises 30 hosts (one unit (1U) Rack Mountable machine with 2 1GHz CPU PIII and 1024MB RAM), one server for the queue manager with an external RAID5 storage (1,2TB) and another disk server with 2TB of RAID5 disk space. This farm is used by the COMPASS-Torino group to analyze

data collected during COMPASS experiment's runs. The main tasks of the farm are the downloading and the analysis of the mDST preprocessed at CERN. The COMPASS farm is also used to produce Montecarlo simulations.

The main MC task is to produce the Λ 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 MonteCarlo 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 MonteCarlo events are insufficient to fulfil the demand. I therefore suggested to use the new possibilities given by the GRID computing. Thanks to my experience as "site manager" of the Torino site (LCG2 and INFN Production GRID) I tested how COMPASS could access to GRID resources. Using GRID I was able to produced the simulated Λ events without using the COMPASS computing resouces. When COMPASS will officially join the GRID project it will be able to exploit an incredible number of CPUs.

The italian GRID currently comprises 23 sites with 860 CPU and 120 TB of disk space.

0.4 Outline of the thesis

This thesis is divided in two parts. In the first part I will give a general view of the COMPASS experiment, and of the GRID Project. In chapter 1 I will describe the main physics tasks of COMPASS collaboration, giving an overview of the muon and hadron physics programs.

I will describe the COMPASS experimental apparatus and the Data Acquisition (DAQ) system in chapter 2.

In chapter 3 I will give an overview of the physics problem that I have

studied.

Chapter 4 may be considered an “introduction for pedestrians to GRID computing”. I will give some definition together with a complete description of the INFN Production GRID and the LHC Computing GRIDs, with a particular stress on for the architecture and submission procedure.

In the second part I will describe the original work of this thesis aimed to enable the use of the GRID by the COMPASS Torino Group to generate simulated semi-inclusive events with transversal and longitudinal polarizations.

In chapter 5 I will describe the software packages used to perform MonteCarlo simulations, the events generator (LEPTO), the simulation program (COMGEANT) the reconstruction program (CORAL) and the analysis program (PHAST).

In chapter 6 I will describe the method used to install on the COMPASS MonteCarlo simulation software on the GRID resource.

In chapter 7 I will give a report of the MC production. I have submitted about 8500 jobs resources on GRID, producing a large sample of simulated events (about $2 \cdot 10^{10}$). I will show the results of my work during 2004 and I will describe the results from the GRID point of view.

In chapter 8 I will analyse the simulated data produce and compare the results with the analysis of real events.

0.5 Introduzione

COMPASS (COmmon Muon and Proton Apparatus for Structure and Spectroscopy) [1] e' un apparato sperimentale molto complesso progettato e costruito da una collaborazione di piu' di 20 Istituzioni scientifiche. Sono stati pianificati due distinti programmi scientifici: il primo riguardante la Fisica muonica ed il secondo la Fisica adronica. L'esperimento COMPASS e' installato nell'area nord del CERN di Ginevra nell'edificio 888 del sito di Preveessin (F). Utilizza fasci secondari prodotti dal SuperProtoSincrotrone (SPS). Lo scopo dell'esperimento e' lo studio della struttura degli adroni utilizzando diversi fasci ad alta densita'

di leptoni e adroni con energie comprese tra 100 e 200 GeV.

L'esperimento si propone di acquisire una grande quantità di dati arrivando a determinare la polarizzazione ΔG del gluone, confrontandola con le previsioni teoriche, ottenuta dalla misura dell'asimmetria nella sezione d'urto nello scattering profondamente anelastico di muoni polarizzati su nucleoni polarizzati.

I fasci di adroni sono utilizzati per lo studio del decadimento semi-leptonico di barioni charmati e doppiamente charmati. Entrambe le misure permetteranno lo studio di problemi fondamentali riguardanti la struttura degli adroni. Inoltre verrà studiata la produzione di stati esotici, la cui presenza è prevista dalla QCD ma non ancora provata sperimentalmente. In aggiunta verrà esaminata la funzione di distribuzione della spin trasverso.

Per permettere tutte queste misure, l'apparato sperimentale necessita di una buona risoluzione spaziale con un rateo di particelle di $2 \cdot 10^8$ particelle/spill. Un trigger e un'elettronica di read-out di nuova concezione completano le prestazioni dello spettrometro.

L'acquisizione dei dati sperimentali è iniziata nel 2002. Alla fine del periodo di presa dati del 2004 l'esperimento COMPASS ha acquisito circa 700 TB di dati.

0.6 L'importanza delle simulazioni MonteCarlo

In un esperimento di Fisica della alte energie, come COMPASS, la produzione di eventi MonteCarlo riveste importanti funzioni per la comprensione di diversi aspetti dell'esperimento. Produrre eventi simulati è importante, principalmente per le seguenti ragioni:

- correggere l'accettanza dell'apparato sperimentale;
- testare il software di ricostruzione degli eventi;
- stimare l'efficienza di ricostruzione del eventi;
- studiare l'effetto di errori sistematici

Producendo eventi simulati e' possibile minimizzare gli errori strumentali dovuti alla geometria del bersaglio e dell'apparato sperimentale. Inoltre e' possibile minimizzare gli errori dovuti ai tagli operati in fase di analisi. E' possibile inoltre testare gli algoritmi di ricostruzione dei vertici primari e secondari, la cui identificazione gioca un ruolo primario nello studio della Λ .

0.7 Perche' GRID

La grande richiesta di computing power proveniente dagli esperimenti di Fisica delle alte energie, come ad esempio ALICE, ATLAS, CMS e LHCb a LHC del CERN, non puo' essere soddisfatta dai tradizionali sistemi di calcolo. Inoltre la creazione di un sistema di calcolo distribuito e' richiesta per organizzare e condividere le risorse di calcolo eterogenee di un gran numero di istituzioni e centri di calcolo. L'esperimento COMPASS, oltre alle risorse di calcolo installate al CERN e a Karlsruhe, possiede una "farm" di calcolatori installata al centro di calcolo dell'INFN di Torino. La farm e' composta da 30 host ognuno con due processori PIII 1GHz e 1024MB di RAM, montati in un rack, un server per la gestione delle code di sottomissione, con storage esterno RAID5 (1,2TB); dispone inoltre di un server con ulteriori 2TB RAID5 di spazio disco.

Questa farm e' utilizzata dal gruppo COMPASS di Torino per l'analisi dei dati acquisiti dall'esperimento COMPASS. Questi dati, preprocessati al CERN vengono trasferiti sulla farm e quindi analizzati. La farm di COMPASS e' inoltre utilizzata per la produzione di eventi MonteCarlo.

L'obiettivo principale della produzione di eventi MC e' la produzione di Λ simulate ($2 \cdot 10^8$), richieste dall'analisi dei dati reali. Per generare una cosi' grande quantita' di eventi, utilizzando un singolo processore della farm, impiegheremmo circa $1.7 \cdot 10^5$ ore (CPU PIII 1GHz, 1024MB RAM). Potendo utilizzare a pieno regime i 60 processori che formano la farm di COMPASS sarebbero necessari circa 120 giorni. Durante questo periodo la farm dovrebbe essere utilizzata esclusivamente per la produzione di eventi MC. Questo scenario e' impraticabile poiche' la farm e' correntemente utilizzata per l'analisi dei dati

reali e per il test del software di ricostruzione e analisi.

E' quindi chiaro che le risorse che COMPASS puo' dedicare alla produzione di eventi MonteCarlo e' insufficiente a soddisfare la richiesta. Ho quindi suggerito l'utilizzo delle nuove possibilita' offerte dal "GRID computing". Grazie all'esperienza acquisita come "site manager" del sito di Torino della GRID di produzione dell'INFN e Lhc Computing GRID, ho testato come l'esperimento COMPASS poteva accedere alle risorse GRID.

Utilizzando GRID ho potuto produrre i set di dati MC richiesti dall'analisi dei dati dell'esperimento COMPASS senza sovraccaricarne le risorse di calcolo.

Quando COMPASS entrera' ufficialmente a far parte degli esperimenti abilitati all'utilizzo della GRID potra' disporre di un incredibile numero di CPU e di spazio disco.

Attualmente la GRID italiana e' composta da 23 siti per un totale di 860 CPU e 120 TB di spazio disco.

0.8 **Struttura della tesi**

Il mio lavoro di tesi e' suddiviso in due parti; nella prima descrivo l'apparato sperimentale di COMPASS e il progetto GRID. Nel primo capitolo descrivo i principali obiettivi della collaborazione COMPASS, fornendo una panoramica del programma muonico e del programma adronico.

L'apparato sperimentale di COMPASS e il sistema di acquisizione dati e' descritto nel capitolo 2.

Nel capitolo 3 descrivo in modo approfondito il problema fisica da me studiato.

Il capitolo 4 puo' essere considerato come una "introduzione per principianti" al "GRID computing". Descrivendo gli elementi che compongono una GRID fornisco una descrizione completa della GRID di produzione INFN (INFN Production GRID) e della LHC Computing GRID (LCG), dando particolare risalto all'architettura e alle procedure di sottomissione dei "job".

Nella seconda parte descrivo il lavoro da me condotto nell'ambito degli anni

del Dottorato per permettere al gruppo di Torino dell'esperimento COMPASS di utilizzare la GRID per generare eventi simulati di reazioni semi-inclusive con polarizzazione trasversa e longitudinale.

Nel capitolo 5 descrivo il software utilizzato per la produzioni delle simulazioni MC. Descrivo il generatore di eventi (LEPTO), il programma di simulazione (COMGEANT), il programma di ricostruzione (CORAL) e il programma di analisi (PHAST).

Nel capitolo 6 descrivo la procedura utilizzata per l'installazione del software di simulazione sulle risorse GRID.

Nel capitolo 7 descrivo i risultati, dal punto di vista GRID, della produzione di eventi MC effettuata durante il 2004. In questo periodo ho sottomesso 8500 job su risorse GRID italiane, producendo un gran numero di Λ simulate ($2 \cdot 10^{10}$).

Nel capitolo 8 ho analizzato i set di dati in configurazione trasversa degli anni 2002 e 2003 (3 periodi di presa dati).

Contents

0.1	Introduction	i
0.2	The importance of MonteCarlo simulations	ii
0.3	Why GRID?	ii
0.4	Outline of the thesis	iii
0.5	Introduzione	iv
0.6	L'importanza delle simulazioni MonteCarlo	v
0.7	Perche' GRID	vi
0.8	Struttura della tesi	vii
1	The COMPASS experiment	1
1.1	The Muon Physics program outline	1
1.1.1	Gluonic contribution to nucleon spin	1
1.1.2	Sea quark contribution to the nucleon spin	3
1.2	The hadron physics program outline	5
1.2.1	Charm hadrons	5
1.2.2	Exotic states: glueballs and hybrids	7
1.2.3	The Primakoff reaction	8
2	The experimental apparatus	9
2.1	The M2 beamline	10
2.2	The polarized target	14
2.3	Tracking system	18
2.3.1	Beam definition	18
2.3.2	Muon trigger	19

2.3.3	Calorimetry	22
2.3.4	Event reconstruction and particle identification	22
2.4	Data Acquisition System	32
3	Seminclusive Λ production from polarized targets	37
3.1	Lepton-nucleon scattering process	38
3.1.1	Kinematics	38
3.1.2	The inclusive DIS cross-section	39
3.1.3	Unpolarized lepton-nucleon scattering	41
3.1.4	Polarized lepton-nucleon scattering	42
3.1.5	Sum rules	45
3.1.5.1	The nucleon spin and the Ellis-Jaffe sum rules	46
3.2	EMC and the “proton spin crisis”	47
3.3	Gluon polarization $\Delta G/G$ and COMPASS	48
3.4	Polarization of strange sea quarks of the nucleon	50
3.4.1	Λ polarization in deep inelastic lepton scattering	51
3.4.2	Target fragmentation region	51
3.5	The transverse spin of the nucleon	52
3.6	Λ production with a transversely polarized target	54
4	The GRID	57
4.1	GRID Computing	57
4.2	LHC Computing Grid - LCG-2	57
4.3	INFN Production GRID	59
4.4	The Grid Architecture (LCG-2 and INFN Production Grid)	62
4.5	GRID Elements	63
4.5.1	User Interface	63
4.5.2	Computing Element	64
4.5.3	Storage Element	65
4.5.4	Information System	65
4.5.5	Data Management	66
4.5.6	Job Management	68

4.6	Service Interactions and Job Flow	69
4.6.1	Job Submission	69
4.6.2	Data Management	72
4.6.3	Information System	73
4.7	Entering the GRID	74
4.7.1	Obtainig a Certificate	74
4.7.2	Registering With LCG-2 - INFN Production Grid	75
4.7.3	VOs - Virtual Organizations	75
4.7.4	Setting up the User Account	76
4.8	JDL - Job Description Language	76
4.8.1	JDL example	78
4.8.2	Job Status Definition	80
5	The software packages	83
5.1	Introduction	83
5.2	The LEPTO event generator	83
5.3	The GEANT simulation	84
5.4	COMGEANT	85
5.5	The reconstruction: CORAL	89
5.6	PHAST	90
6	Software Installation for Lambda Transversal production	93
6.1	A typical farm	94
6.2	Software installation	96
6.3	Installation TEST	104
7	COMPASS@GRID	107
7.1	Roadmap	107
7.2	Longitudinal simulation (First phase)	109
7.2.1	Results	109
7.2.2	Submission technique	112
7.2.3	Disk space	113

7.3	Transverse simulation (Second phase)	113
7.3.1	Results	114
7.3.2	Submission technique	114
7.3.3	Disk space	116
7.4	CPU performance	116
7.5	DC Chronology	118
8	Transverse target spin data analysis	121
8.1	Transverse polarization	121
8.2	Useful symmetries	127
8.3	MonteCarlo simulations	127
8.4	Event selection	130
8.4.1	K_S^0 mesons identification	131
8.4.2	Primary and secondary vertex selection	131
8.4.3	Cuts on the decay product ($V^0 \rightarrow T^+T^-$)	132
8.4.4	Other selection criteria	132
8.5	Results of the analysis of MonteCarlo events	133
8.6	Results of transverse polarization analysis	135
A		141
A.1	GRID acronyms	141
A.2	Submit cron	143
A.3	JDL file for Λ transverse production	147
A.4	Simulation software installation script	147
	Bibliography	155

Chapter 1

The COMPASS experiment

1.1 The Muon Physics program outline

It is now firmly established that the spin content of the nucleon is not entirely due to the quark spins. The EMC result, of $\Delta\Sigma = 0.19 \pm 0.07$ [?], has been recently confirmed by experiments at CERN and SLAC. More than one theoretical explanation of such phenomena exist. In the gluon interpretation the polarized gluon ΔG lowers the quark's contribution to the nucleon spin. In an alternative model negatively polarized strange quarks are the responsables. Several are the ways to resolve these ambiguities.

1.1.1 Gluonic contribution to nucleon spin

The spin equation for the proton (or neutron) can be expressed as:

$$\frac{1}{2} = \frac{1}{2}\Delta\Sigma + \Delta G + \langle L_z \rangle \quad (1.1)$$

where $\Delta\Sigma = \Delta u + \Delta d + \Delta s$ is the contribution of the quark spin, ΔG is the contribution of the gluons and $\langle L_z \rangle$ is the possible contribution of the orbital angular momentum of the quarks and gluon. $\Delta u + \Delta d + \Delta s$ are the difference of distribution function of the quarks (anti-quarks) of a given flavour with spin parallel or anti-parallel to the nucleon spin. These functions can be expressed

as:

$$\Delta q_f = \int_0^1 dx [(q_f^{\uparrow\uparrow} + \bar{q}_f^{\uparrow\uparrow}) - (q_f^{\uparrow\downarrow} + \bar{q}_f^{\uparrow\downarrow})] \quad (1.2)$$

where x is the x-Bjorken variable, that is the fraction of momentum carried to the total momentum by the quark.

In the Parton Model the three valence quarks are in the S-state, therefore $\langle L_z \rangle = 0$ and $\Delta\Sigma = 1$. This result, is not confirmed by the experimental results.

If one measures the gluon polarization function via the photon-gluon fusion process leading to open charm production, one can access to ΔG , as is shown in Fig.1.1.

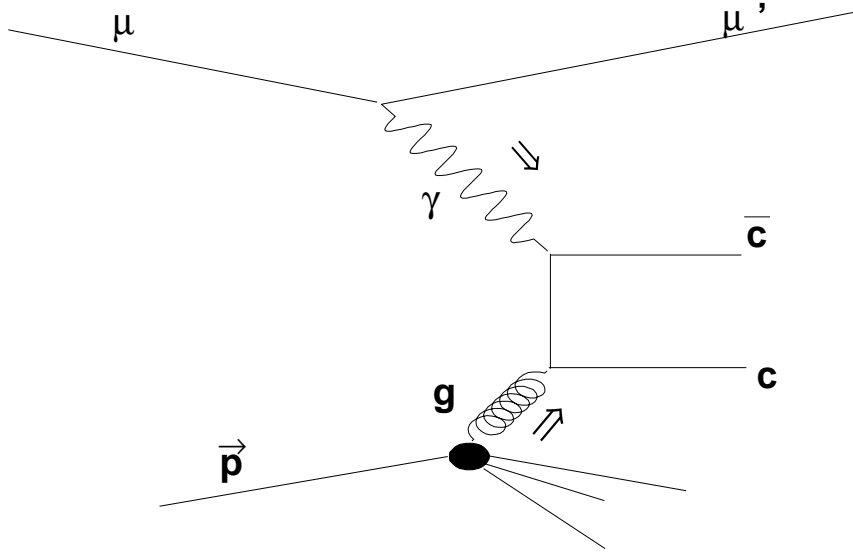


Figure 1.1: Photon-gluon fusion diagram for $c\bar{c}$ production

To reach this purpose longitudinal polarized muon scattering on polarized target can be used.

The quantity $\Delta G/G$ is determined from the measurement of the cross section asymmetry in the production of charmed particle in deep inelastic scattering (DIS) of longitudinally polarized muon on longitudinally polarized target. The measurement is based on the reconstruction of the D^0 meson from its hadronic

decay products ($D^0 \rightarrow \pi K$).

ΔG is also accessible by the measurements of the asymmetry in single or correlated production of large transverse momentum p_T particles in DIS. The measurement is performed looking for two correlated high- p_T hadrons with p_T larger than a certain threshold. A threshold of $(1.0 \div 1.5)$ GeV/c is sufficient to have a signature of the photon-gluon fusion process. The diagram illustrating this reaction is shown in Fig.1.2:

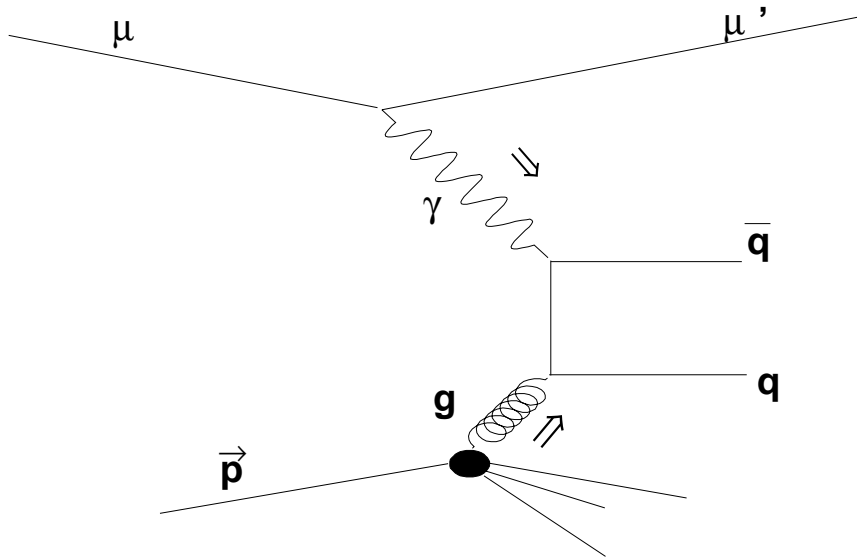


Figure 1.2: Photon-gluon fusion diagram for $q\bar{q}$ production

1.1.2 Sea quark contribution to the nucleon spin

The self-analyzing properties of the Λ make this particle particularly suited for the spin physics studies. The question is whether the spin of the strange sea-quarks is really anti-correlated with the proton spin, as proposed in some models, can be solved measuring the lambda polarization in the target fragmentation region. The use of polarized muon beam with unpolarized target selects reaction where quarks are oriented anti-parallel to the virtual photon helicity. For example, as shown in Fig.1.3, a positive polarized u quark with respect the

polarization on the nucleon, would leave behind a negative polarized $s\bar{s}$ pair (in the scope of the intrinsic strangeness model). This would lead a negative polarization of the Λ , when produced in the target fragmentation region. For the polarized gluon model, as shown in Fig.1.4, if a strange quark in the Λ is produced from a polarized gluon, assuming the gluon polarization is the same of the proton, a positive polarization of the Λ with respect to the nucleon would be observed. All polarization are considered negative or positive with respect to the nucleon polarization.

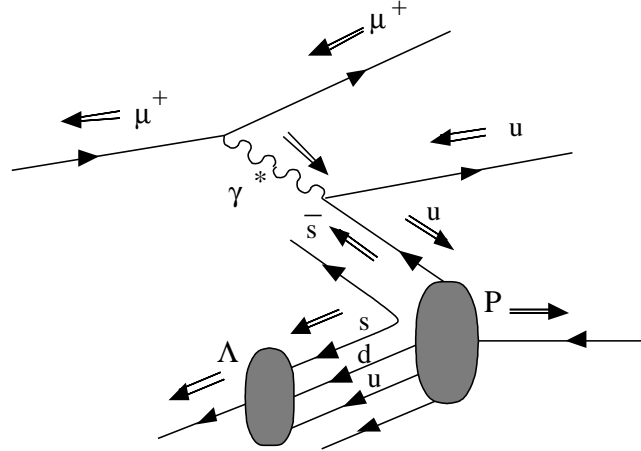


Figure 1.3: Λ polarization arising from polarized strange quarks, the \Rightarrow represents the helicity of the particle.

The observation of the Λ in the current fragmentation region addresses the problem of the spin dependent fragmentation functions. The measurement of the polarization of a forward produced Λ directly determines the spin dependent fragmentation function. The probability that a quark spin in transversely polarized nucleon is oriented parallel or anti-parallel to the nucleon spin can in principle be measured in DIS, requiring a transversely polarized target and the knowledge of the spin dependent fragmentation function for this case. This function can be measured selecting a particular hadron in the final state. The hadron is the Λ , to take advantage from its self-analyzing properties.

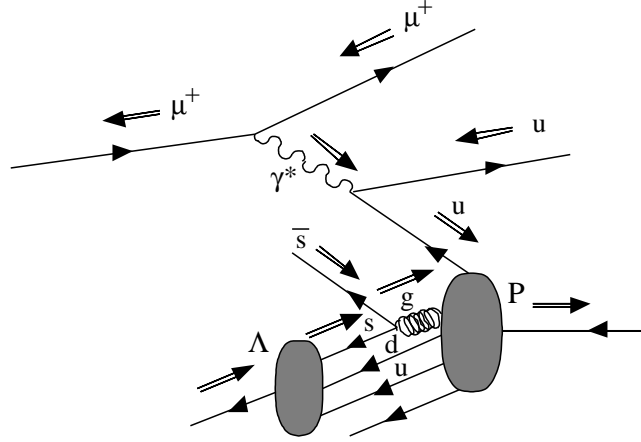


Figure 1.4: Λ polarization arising from polarized gluon, the \Rightarrow represent the helicity of the particle

1.2 The hadron physics program outline

COMPASS experiment with hadronic beam, p , π^+ , π^- , K^+ , K^- , aims to collect a sample of data many times greater than previously, in particular in the charmed hadrons sector for the spectroscopy of the light quark system and glueballs, and in the investigation of hadronic structure of unstable particles in selected channels like the Primakoff.

1.2.1 Charm hadrons

The knowledge of the semi-leptonic decay widths is one of the most relevant issues in the charm physics. Although many experiments were made in the last years, the understanding of the charmed particles is still lacking. Even if the life time of the states, decaying weakly is already measured the information obtainable from these data is limited by the accuracy of the measurement, not better than 50%.

The systematical studies of the semi-leptonic decay (inclusive and semi-inclusive) are the best way to verify the now days knowledge of the charm particles. Precise theoretical predictions on rates and form factor are available.

In particular the q^2 dependence of the form factor was recently calculated in the framework of the HQET. Although the charm quark mass is at the lower limit of applicability of the theory, corrections are believed to be on the order of only 20%. In the absence of a precise knowledge of charmed hadron production rates, semi leptonic decay provide a link between lifetime and hadronic branching ratio: the theoretical semi-leptonic decay width together with the experimental ratio of the semi-leptonic to hadronic branching fraction allow a determination of hadronic partial widths. The absolute hadronic branching ratio results from the multiplication with the measured lifetime.

$$\Gamma_{adr} = \left(\frac{BR_{adr}}{BR_{step}} \right) \cdot \Gamma_{step} \quad (1.3)$$

$$BR_{adr} = \Gamma_{adr} \cdot \tau \quad (1.4)$$

Almost all states of singly charmed baryons were observed, but nothing is known about doubly charmed baryons. Their structure is probably like a heavy meson: a heavy cc -diquark in the center surrounded by a light quark. The very low production cross section for these doubly charmed baryons requires an extremely high-rate experiment and beam of 10^7 protons per second.

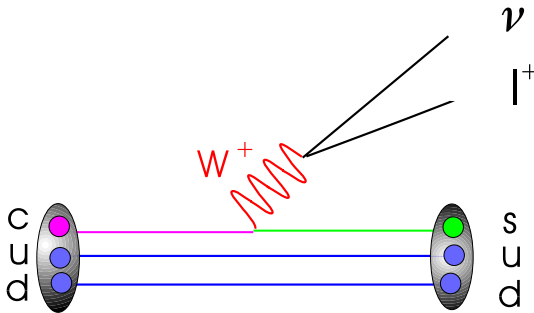


Figure 1.5: *Semi leptonic decay for charmed particle*

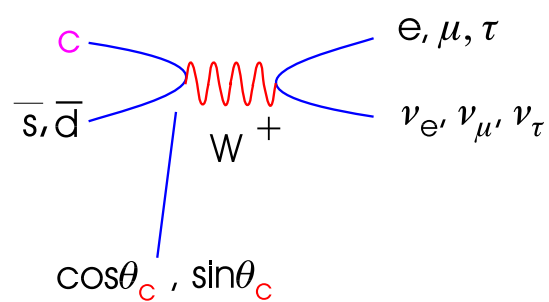


Figure 1.6: *Leptonic decay for charmed particle*

1.2.2 Exotic states: glueballs and hybrids

Quantum Chromo Dynamics (QCD), is a non-Abelian theory that predicts the existence of states containing valence gluons: glueballs (gg , ggg) and hybrids ($q\bar{q}g$). In these states the gluons contribute to the overall quantum numbers. QCD can not predict the spectrum of the non- $q\bar{q}$ states. Many models predict the low lying exotic states in the mass range from 1 GeV to 3 GeV, and from 1.5 GeV up to 1.8 GeV for the lightest glueball. In this mass range scalars and other mesons abound. A candidate for the scalar glueball ground state 0^{++} is the $f_0(1500 \text{ MeV})$ first observed at LEAR by the Crystal Barrel experiment. The COMPASS experiment studies light meson spectroscopy in the gluon rich reactions, i.e. the proton-proton central production and the diffractive pion and kaon scattering on nuclei. The identification of glueballs requires high statistics data sample, reconstruction of final states containing both neutral, especially η and η' particles, and charged particles, observation of the same meson in many different channels and the production of mesons in different reactions. The COMPASS experiment is suited to satisfy all of the above mentioned requirements.

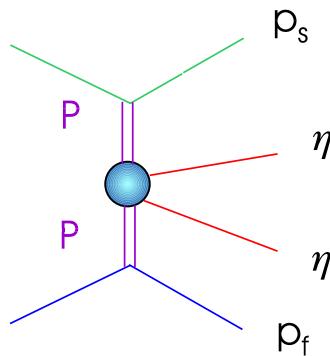


Figure 1.7: *Semi leptonic decay for charmed particle*

1.2.3 The Primakoff reaction

The Chiral Perturbative Theory χPT allows to make predictions for a number of quantities such as polarizabilities and cross section for the $\pi\gamma$ interactions. Currently such studies are almost addressed at low energy electron accelerators. High energy pion, kaon and hyperon beams allow complementary measurements using the Primakoff mechanism, which corresponds to a Compton scattering of an hadron on a virtual photon in the inverse kinematics. It is supposed that an hadron has an induced dipole moment associated to its quark structure due to the presence of a external electromagnetic field. For the pion a polarizabilities measurement already exists. For the kaon such measurement will be performed for the first time. Other χPT prediction are the width for the decay $\gamma \rightarrow 3\pi$ which can be tested by the COMPASS apparatus through the Primakoff mechanism. In general, the Primakoff experiment requires high Z target, low mass tracking detectors, good charged particle momentum resolution and high spatial and energy resolution electromagnetic calorimeter. It is also important to use a veto system around the target to reject events with an excitation and break-up of the nucleus.

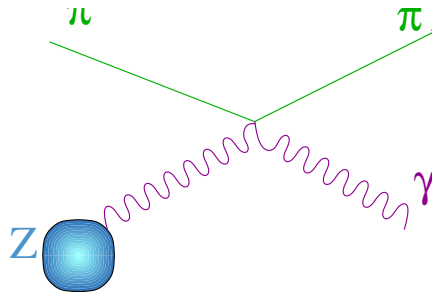


Figure 1.8: *Primakoff reaction*

Chapter 2

The experimental apparatus

The COMPASS apparatus is a double magnetic spectrometer with complementary kinematical ranges. The two spectrometers have a similar structure and are based on dipole magnets called SM1 and SM2 respectively. Each spectrometer is equipped with several tracking devices with different spatial and timing resolutions, and one RICH (Ring Image Cherenkov) detector for particle identification.

Tracking of charged particles is performed by means of a large variety of detectors. 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. This kind of detectors provide an active area of $\sim 30 \times 30 \text{ cm}^2$ and a spatial resolution of $\sim 70 \mu\text{m}$, and are complemented in the central region by scintillating fibers. Large area tracking is performed using drift chambers, straw detectors, MWPC chambers and plastic Jarrocci 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

LAS	SAS
SM1 aperture: $2.0 \times 1.6 \text{ m}^2$	SM2 aperture: $2.0 \times 1.0 \text{ m}^2$
SM1 field integral: 1 Tm	SM2 field integral: 4.4 Tm
Angular acceptance: $\theta > 30 \text{ mrad}$	Angular acceptance: $\theta < 30 \text{ mrad}$
Momentum range: $p < 60 \text{ GeV}/c$	Momentum range: $p > 10 \text{ GeV}/c$

Table 2.1: *Characteristics of the Large Angle Spectrometer (LAS) and of the Small Angle Spectrometer (SAS) of the COMPASS apparatus.*

additional track points for muons scattered at large angles. Particle identification is performed using a Ring Image Cherenkov (RICH) detector located between SM1 and SM2. Muon identification is performed by means of two thick Fe hadron absorbers, called μF1 and μF2 . 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. Tab.2.1 summarizes the characteristics of the two spectrometers with respect to the allowed kinematical ranges and angular acceptances.

A typical event with Λ candidates in the final state is shown in Fig.2.1; the scattered muon is tracked along the full apparatus and identified by the crossing of the second muon filter, while Λ s are reconstructed from the “ V^0 -like” decay vertexes in the first spectrometer.

2.1 The M2 beamline

The muon beam used in COMPASS is a tertiary beam of positive muons produced in the M2 beamline at the CERN SPS. The available kinematic range goes from 60 to 190 GeV/c with fluxes up to $2 \cdot 10^8$ muons per SPS cycle. It is a pulsed beam with a burst length of $\sim 4 \text{ s}$ and a period of $\sim 14 \text{ s}$, produced using a 450 GeV primary proton beam extracted from the SPS toward the North Experimental Area. A fraction of the proton beam is selected by two stages of

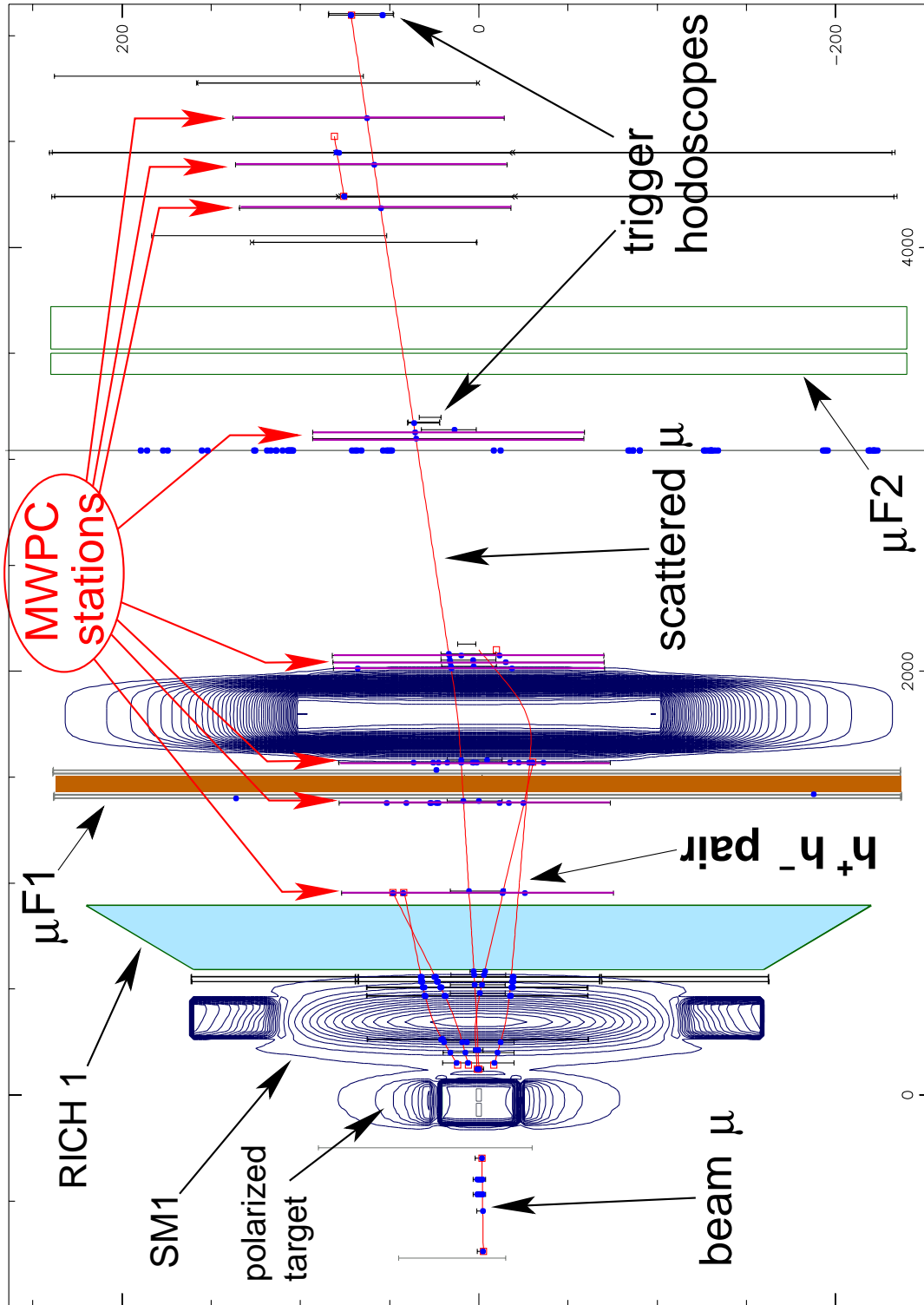


Figure 2.1: DIS event with two h^+/h^- pairs.

septum magnets and interacts with the primary T6 target, a 500 mm long and 3 mm thick beryllium plate. A secondary beam of positive or negative pions and kaons is derived from the target at zero production angle. The beam line can transport to the experiment either the primary π and k or the decayed μ or electron. In the latter case pions and kaons with a momentum spread of 5%, are selected by the first bending magnet and are transported through a 500 m long line of focusing and defocusing quadrupoles. (see Fig.2.2) In this section the tertiary muon beam is generated from pion and kaon decay and is transported together with the hadrons. At the end of the hadron section the beam is focused onto a 9.9 m long hadron absorber made of beryllium and located 700 m downstream of the production target. At this point the beam contains a muon fraction of about 6% for the standard pion energy of 208 GeV/c.

The absorber is located inside the bending magnet, which selects a muon momentum band of about 3%. In the following 300 m long muon section the beam is cleaned and focused onto the target of the experiment.

The muons produced from the parity violating pion decay $\pi \rightarrow \mu\nu$ (see Fig.2.3) are naturally longitudinally polarized. The polarization depends on the decay angle in the pion rest frame with respect to its direction of flight in the laboratory rest frame. Positive forward (backward) muons are polarized anti parallel (parallel) to their momentum in the laboratory frame. Highest polarization is obtained for a muon to pion energy ratio close to unity. For the SMC experiment an operating point of $E_\mu/E_\pi = 0.91$ was chosen as a compromise of muon flux and polarization, leading to a measured polarization of -0.8. COMPASS uses a muon beam with the same characteristics. The polarization value measured in SMC is considered valid also for the COMPASS beam.

The COMPASS setup of the M2 muon beam provides a spot size at the polarized target of 7.8×7.8 mm and a divergence of 0.4×0.8 mrad, with a flux of $\sim 2 \cdot 10^8$ using the 500×3 mm beryllium target. Different target lengths can be selected in order to reduce the beam flux and allow data taking with spectrometer magnets off for alignment purposes.

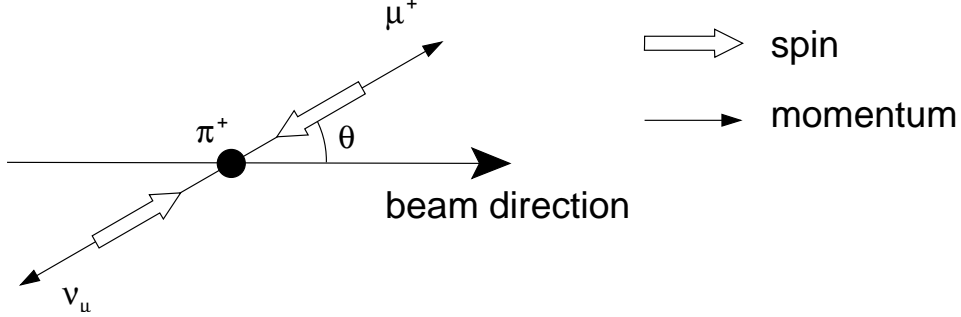


Figure 2.3: Weak π^+ decay in the pion rest frame. Outgoing muon polarization is obtained by selecting θ angles close to zero.

2.2 The polarized target

The COMPASS apparatus for the muon program uses a solid state polarized target in order to achieve high luminosities. The polarization of the target material is obtained using the technique of dynamic nuclear polarization which allows to greatly exceed the values given by the Curie law, which for spin- $\frac{1}{2}$ and spin-1 particles reads

$$P_{\frac{1}{2}} = \frac{N_{\frac{1}{2}} - N_{-\frac{1}{2}}}{N_{\frac{1}{2}} + N_{-\frac{1}{2}}} = \tanh\left(\frac{\hbar\omega}{2kT}\right) \quad (2.1)$$

and

$$P_1 = \frac{N_1 - N_{-1}}{N_1 + N_0 + N_{-1}} = \frac{4 \tanh\left(\frac{\hbar\omega}{2kT}\right)}{3 + \tanh^2\left(\frac{\hbar\omega}{2kT}\right)}. \quad (2.2)$$

Here $\omega = \mu B/\hbar$ is the Larmor frequency, μ the magnetic moment of the particle, k the Boltzmann constant and N_m the population of the magnetic sublevel m . The formula 2.1 gives for the electron a polarization value of 0.998 for a typical field of $B = 2.5$ T and a thermal equilibrium temperature of $T = 2.5$ K, while at the same conditions the result for proton and deuteron is 0.0051 and 0.0011 respectively.

The dynamic nuclear polarization technique allows to transfer part of the almost complete electron polarization of the electrons to the nucleons using microwave irradiation. At the temperature of 0.5 K basically only the two

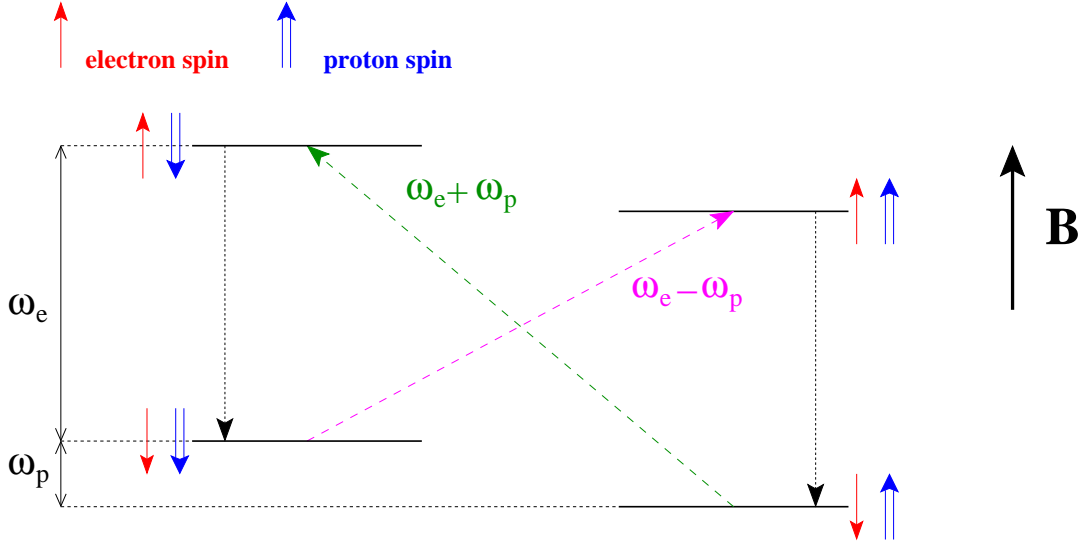


Figure 2.4: Energy levels of an electron-proton pair in a strong magnetic field, B . The NMR-induced transitions exploited by the DNP technique are also shown.

lower magnetic sublevel are populated, with the spin of the electron aligned anti-parallel to the direction of the magnetic field (electron's magnetic moment parallel to the field) and the spin of the nucleons either parallel or anti-parallel to the electron's spin (see Fig.2.4). The transitions $|\downarrow\uparrow\rangle \rightarrow |\uparrow\downarrow\rangle$ and $|\uparrow\downarrow\rangle \rightarrow |\downarrow\uparrow\rangle$ can be induced by microwaves of frequency $\omega_e + \omega_p$ and $\omega_e - \omega_p$ respectively. Since the relaxation time of the electron is of the order of 10^{-3} s while the relaxation time of the proton is 10^6 times bigger the microwave irradiation results in a net positive (negative) polarization of the protons for frequencies equal to $\omega_e + \omega_p$ ($\omega_e - \omega_p$).

Target cell polarization direction can be reversed applying a varying solenoidal and dipolar magnetic fields. The rotation procedure takes around 30 minutes and implies a minimal loss in the polarization of the target material. The systematic reversal, performed every 8 hours, allows to minimize the influence of the apparatus acceptance on asymmetry measurements, when data from two different target cells are compared.

If rotation is stopped halfway and a dipolar magnetic field of ~ 0.5 Tesla

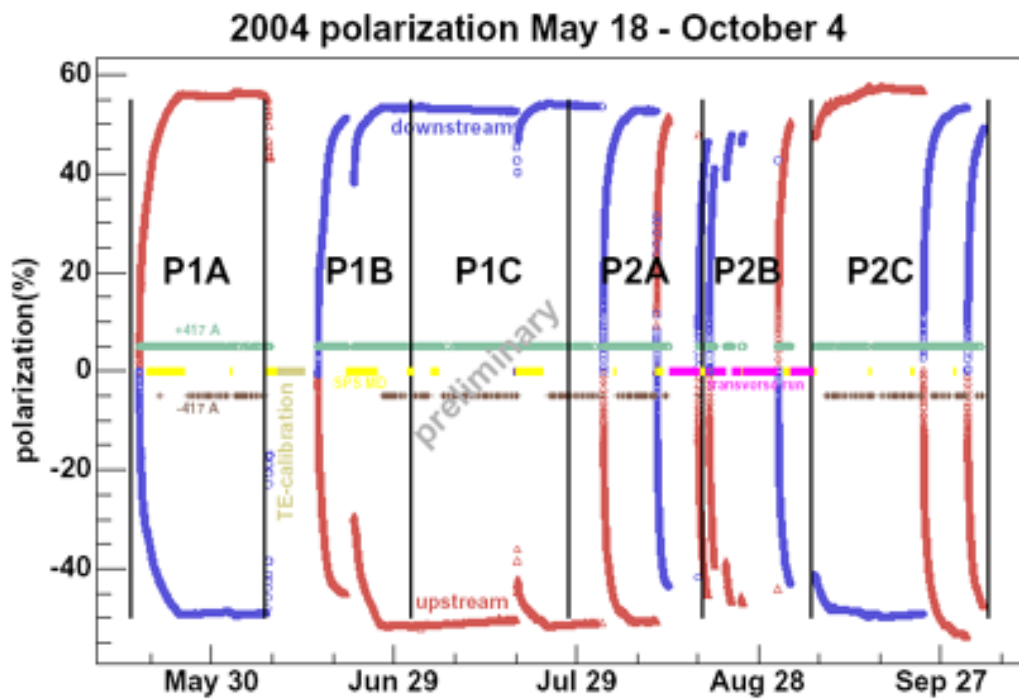


Figure 2.6: Measured upstream and downstream target cell polarizations during 2004 data taking period.

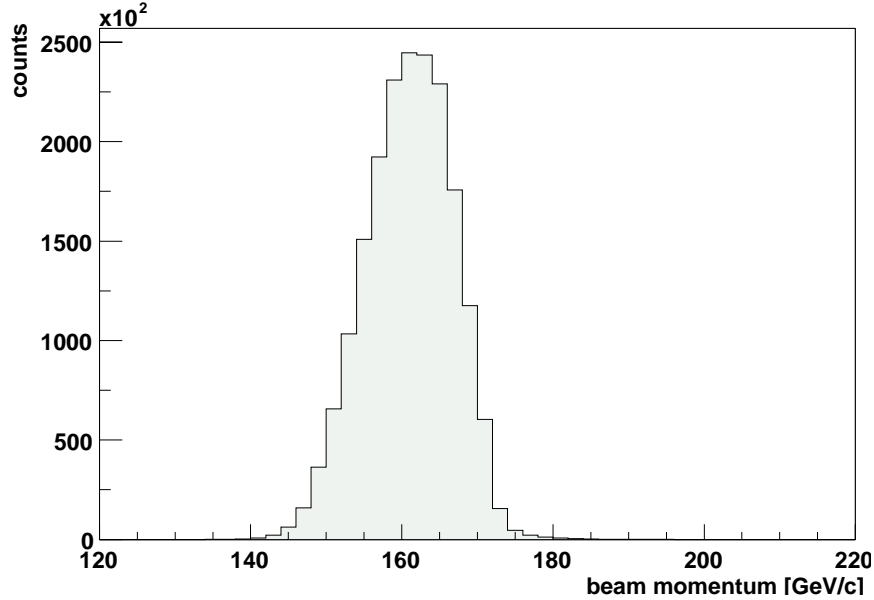


Figure 2.7: Muon beam momentum measured with the BMS system. The plot is obtained by requiring 4 fired BMS planes. Nominal beam momentum is ~ 160 GeV/c.

2.3 Tracking system

2.3.1 Beam definition

The momentum of each incoming muon is measured at the Beam Momentum Station (BMS) consisting of four scintillating fiber planes and the last dipole magnet bending the muon beam vertically into the experimental hall (see Fig.2.2). The scintillator planes form two telescopes upstream and downstream of the magnet and allow the measurement of the muon momentum. Only tracks with at least three fired BMS planes are reconstructed. Time correlation between BMS and beam hodoscopes is used to select good beam muon track candidates in the BMS. Fig. 2.7 shows the reconstructed beam momentum when 4 fired BMS planes are required.

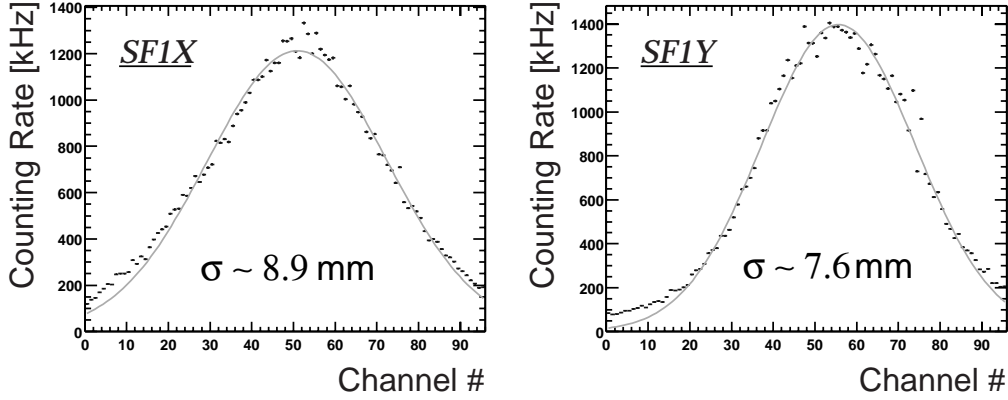


Figure 2.8: *Beam profiles in X and Y directions as recorded by the first SciFi station SF1.*

2.3.2 Muon trigger

The COMPASS experiment is equipped with a multi-trigger acquisition system, which allows to perform several measurements at the same time. Each event is tagged with the kind of trigger which activated the front-end equipments, allowing fast offline selection of different kinematic regions of the DIS process.

The trigger is built starting from the signal of several hodoscopes, made of scintillator plates read out via photomultipliers. The hodoscopes are used to detect the scattered muons at various angles, and are grouped according to the angular regions they cover. Depending on their scattered muon acceptance, they will correspond to different Q^2 and ν regions, where ν is the energy of the exchanged virtual photon. The available triggers, with their Q^2 acceptance, are:

- Inner (Inclusive) Trigger (IT), $Q^2 < 1 \text{ GeV}^2$, low ν
- Ladder Trigger (LT), $Q^2 < 3 \text{ GeV}^2$, high ν
- Middle Trigger (MT), $0.1 < Q^2 < 10 \text{ GeV}^2$
- Outer Trigger (OT), $Q^2 > 0.3 \text{ GeV}^2$

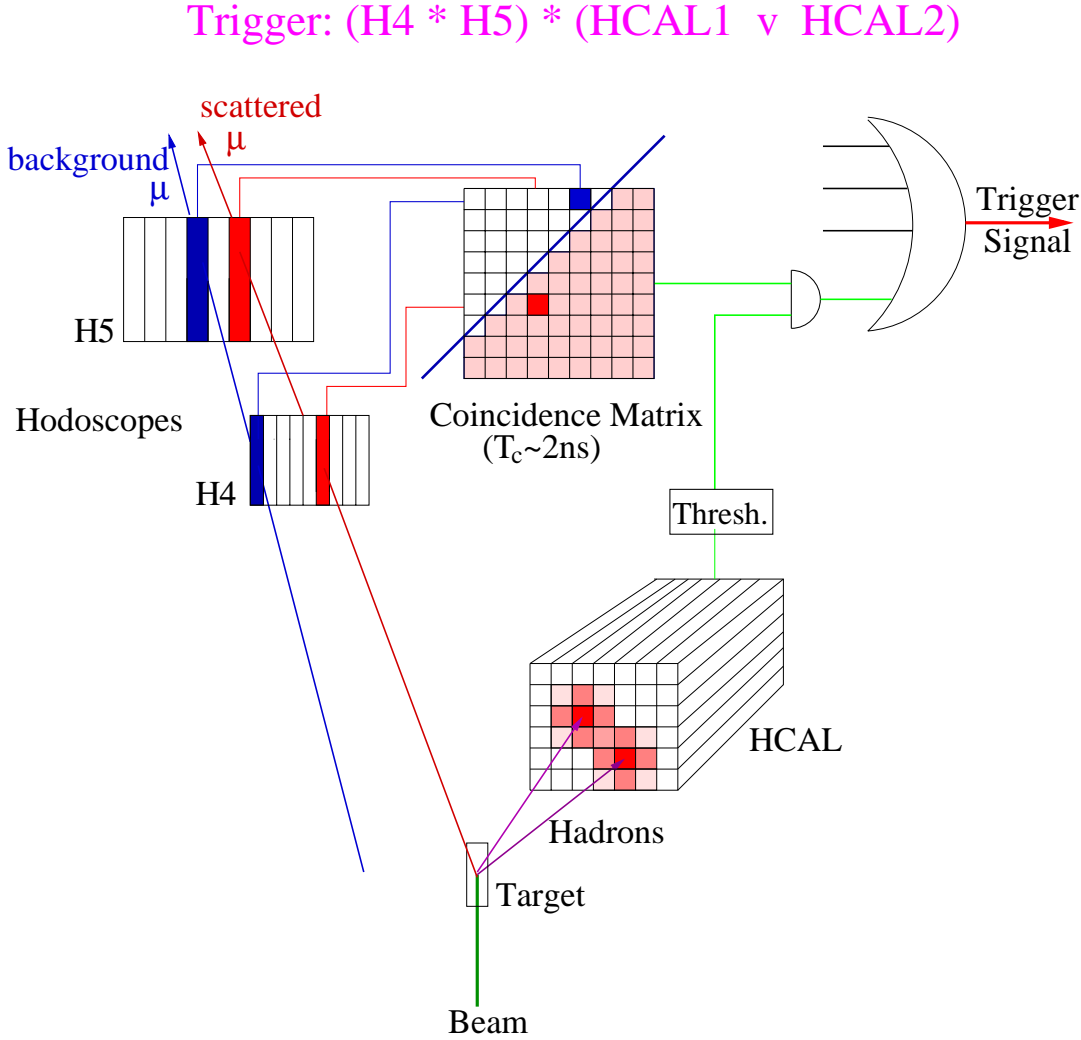


Figure 2.9: Schematic view of the trigger logic used in COMPASS. A coincidence matrix allows to separate scattered and background muons.

A beam trigger signal also exists, defined by the logical OR of all the channels of the scintillating fiber station placed in front of the polarized target. A veto system, composed of scintillator hodoscopes placed upstream the target and around the beam spot, is used to identify and reject halo muons.

The Inner and Ladder systems are used to trigger on quasireal photoabsorption events, among which photon-gluon fusion events are specially interesting for the ΔG measurement. The Middle and Outer systems are on the other hand

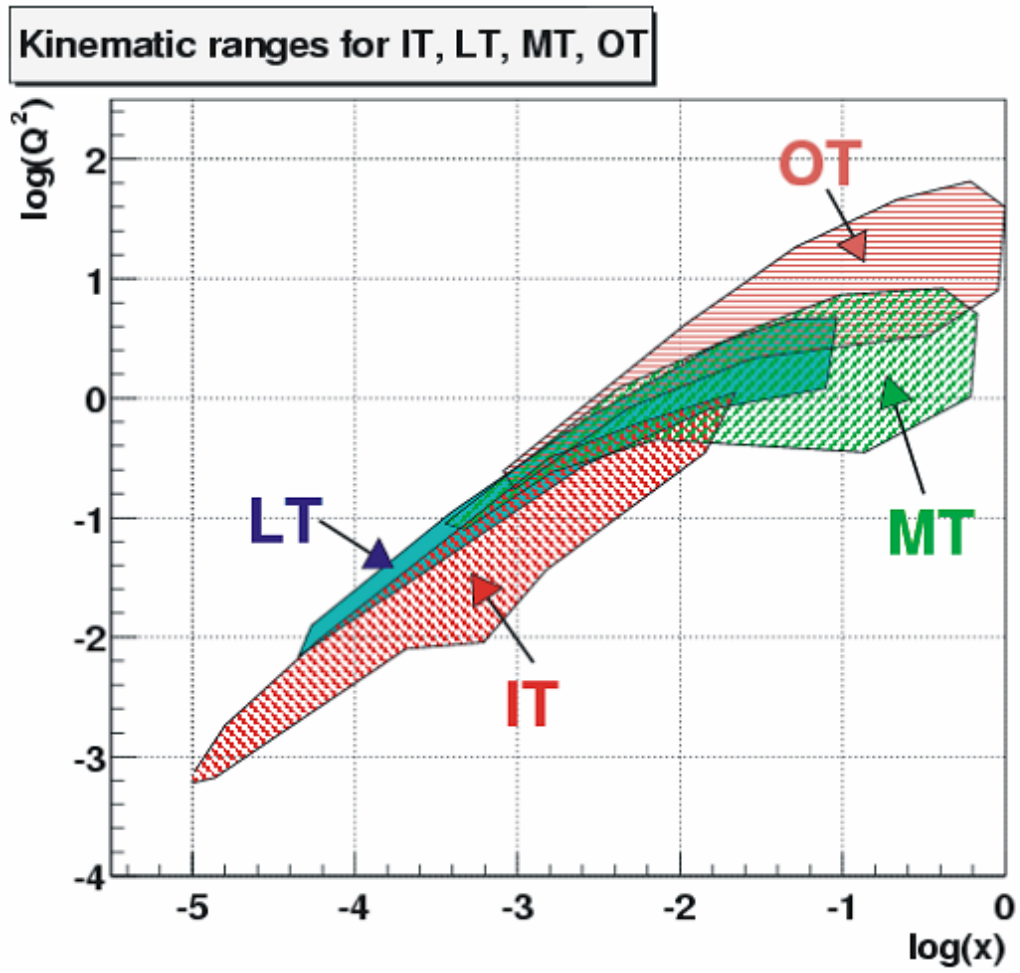


Figure 2.10: Q^2 and x_{Bj} acceptances of the various trigger types.

used to trigger on deep-inelastic events with $Q^2 > 1 \text{ GeV}^2$. The regions of the (Q^2, x_{Bj}) plane covered by the various trigger types are shown in Fig.2.10.

Each trigger system is composed of two scintillating detector stations placed at different distances from the target center. The digitized signal of the various scintillator plates in the two stations is fed to a coincidence matrix with adjustable delays for each row and column input, as schematically shown in Fig.2.9. The hit correlation between the two stations allows to discriminate between muons scattered from the target and halo muons, with an efficiency determined by the cut applied on the coincidence matrix. The wider is the cut, the higher is the efficiency, but the purity, i.e. the percentage of “good” muons among the accepted events, decreases. A coincidence matrix with optimal cuts is defined for each type of trigger. The outputs of the various coincidence matrices are then prescaled to have an approximate balance among all the available triggers and obtain an overall trigger frequency of $\sim 10 \text{ kHz}$.

The Inner and Ladder triggers are used for the detection of open charm production events, and require an additional information from the hadronic calorimeters, in order to reject radiative and μe scattering events as well as low energy halo tracks.

2.3.3 Calorimetry

The complete COMPASS setup is equipped with electromagnetic and hadronic calorimetry in both spectrometers. Tab. 2.2 summarizes the calorimetry performances.

2.3.4 Event reconstruction and particle identification

In the following the main characteristics of the various tracking detectors used in COMPASS are presented.

- **Scintillating Fibers**

Beam tracks are detected using scintillating fiber detectors with a size of $50 \times 50 \text{ mm}$, schematically shown in Fig.2.11. They provide good

LAS

Hadronic (HCAL1)	$\frac{\sigma}{E} = \frac{80\%}{\sqrt{E}} \oplus 8\%$
Electromagnetic (ECAL1)	$\frac{\sigma}{E} = \frac{2.8\%}{\sqrt{E}} \oplus 0.5\%$

SAS

Hadronic (HCAL2)	$\frac{\sigma}{E} = \frac{60\%}{\sqrt{E}} \oplus 6\%$
Electromagnetic (ECAL2)	$\frac{\sigma}{E} = \frac{5.5\%}{\sqrt{E}} \oplus 1.5\%$

Table 2.2: Characteristics for the COMPASS calorimeters.

spatial and time resolution with a very fast response which makes them also suitable as trigger detectors. Four stations are located around the polarized target, two upstream and two downstream. Time resolution of SciFi detectors is measured by looking at the time difference of hits in two correlated planes, as SF1X&SF2X or SF1Y&SF2Y. The result is shown in Fig.2.11-bottom for two different beam intensities and gives a measured time resolution of ~ 1.5 ns.

- **Silicon Detectors**

The tracking system upstream the polarized target comprises two stations of Silicon detectors with four measured coordinates each. Silicon detectors provide a better spatial resolution if compared to scintillating fibers and are used to improve the tracking precision of beam muons. The SciFi-Silicons beam telescope provides an overall beam reconstruction efficiency of the order of 40%.

- **Micromegas**

In this type of gaseous detectors the sensitive volume is divided into two

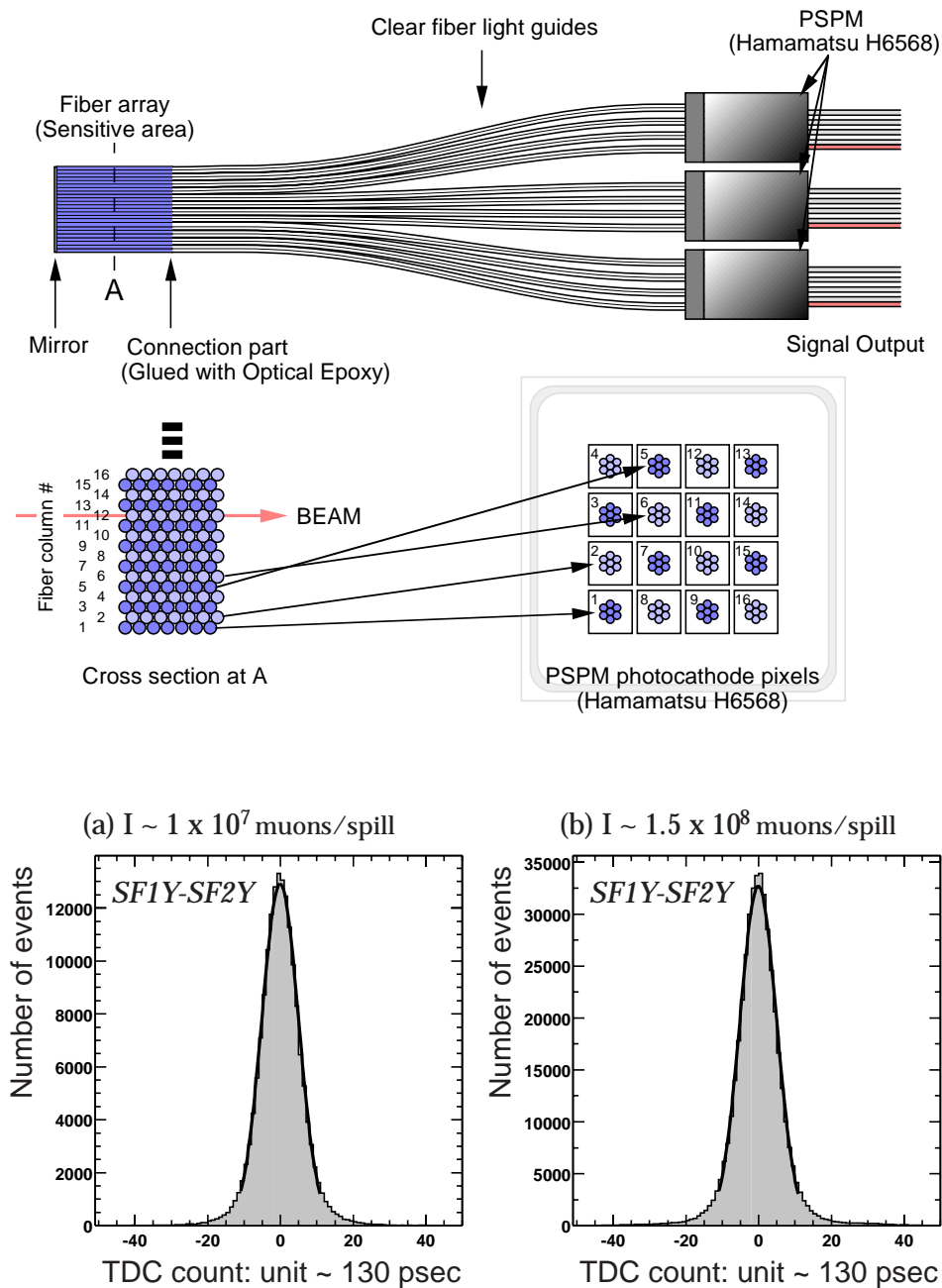


Figure 2.11: *SciFi detectors design.* Upper figure shows the schematic view of one *SciFi* plane; each channel is made of seven scintillating fibers read out by the same photomultiplier, in order to achieve full efficiency. Bottom plots show the time resolution measured from time difference of hits belonging to two correlated planes, for (a) low intensity beam and (b) high intensity beam.

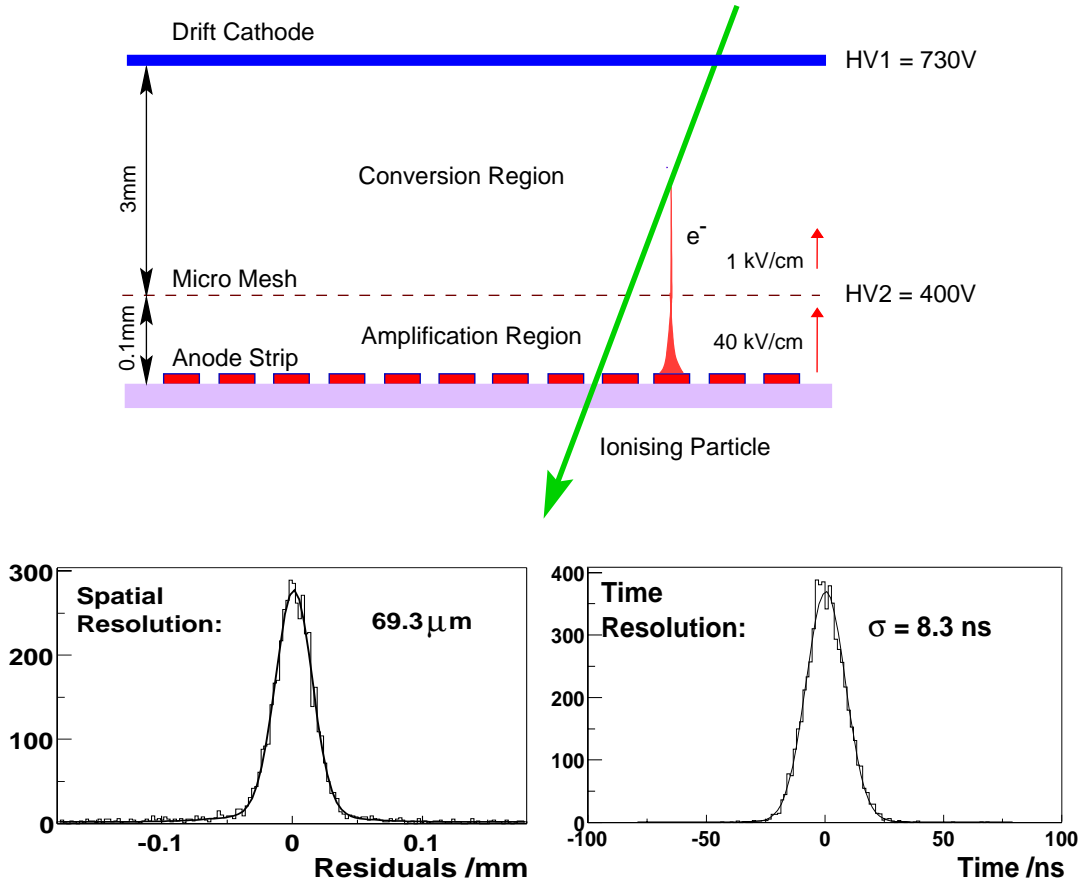


Figure 2.12: Principle of operation and performances of Micromegas gaseous detectors.

regions by a Micromesh foil (see Fig.2.12). The upper region in the picture (Conversion Region) occupies most of the sensitive volume and is characterized by a low electric field which makes the electrons produced in it by ionization drift toward the micromesh foil. This is a thin metallic grid which lets the electrons pass to the lower region, where the electric field is much more intense and multiplication occurs (Amplification Region). The electrons are collected by anode strips which are then read out and provide 1-D information on the localization of the charge avalanche.

- **GEM**

GEM technology uses a different approach to produce charge multipli-

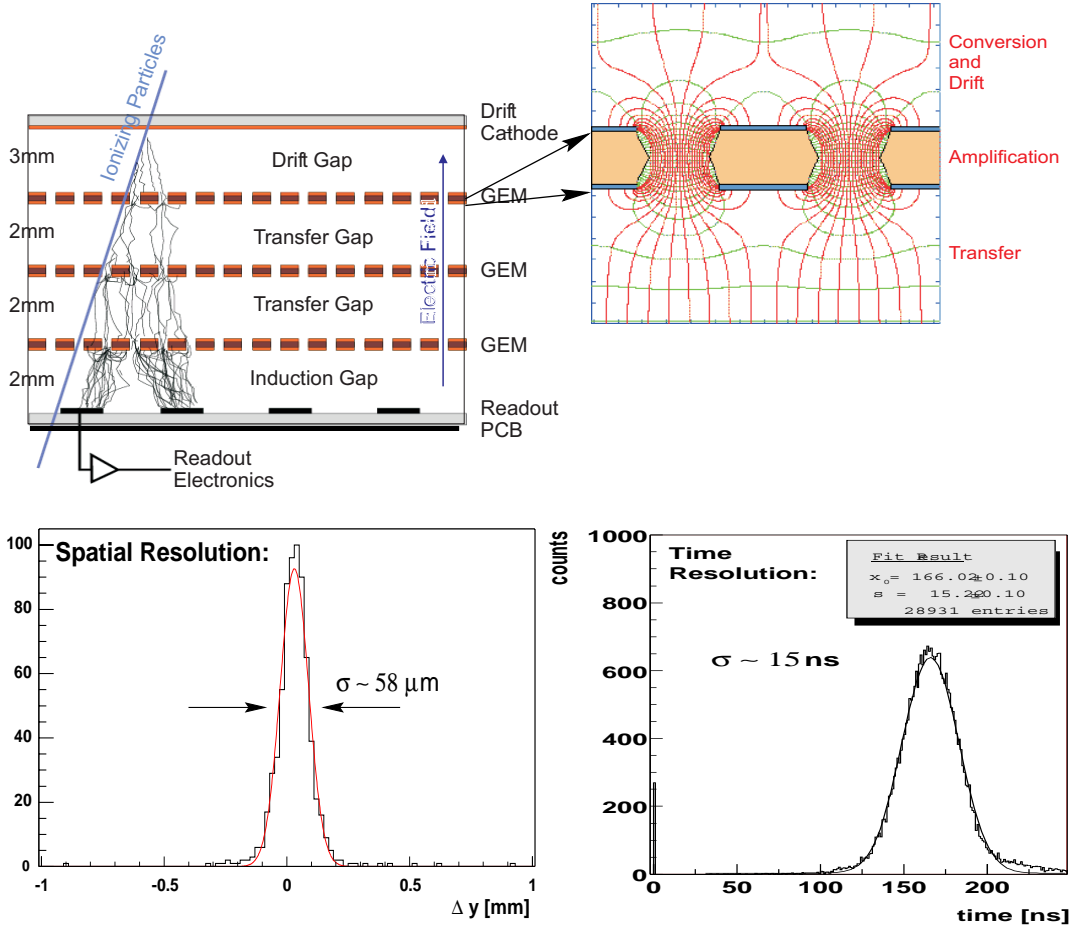


Figure 2.13: Principle of operation and performances of GEM detectors.

cation. In this case high electric field are generated inside small holes created in a thin insulating mylar foil. The foil is coated on both sides with aluminum and a potential difference is applied. Electrons created by ionization in the conversion region above the first GEM foil (see Fig.2.13) drift toward it and some of them enter the holes. Here the electric field is high enough to produce charge avalanches and multiplication occurs. The newly produced electrons then start drifting in the transfer region below the GEM foil, while the ions are immediately collected by the bottom aluminized side of the GEM foil. This prevents space-charge effect to distort the electric field, except for the small gas volume inside the holes. One of the advantages of the GEM technology is that several foils can be stacked

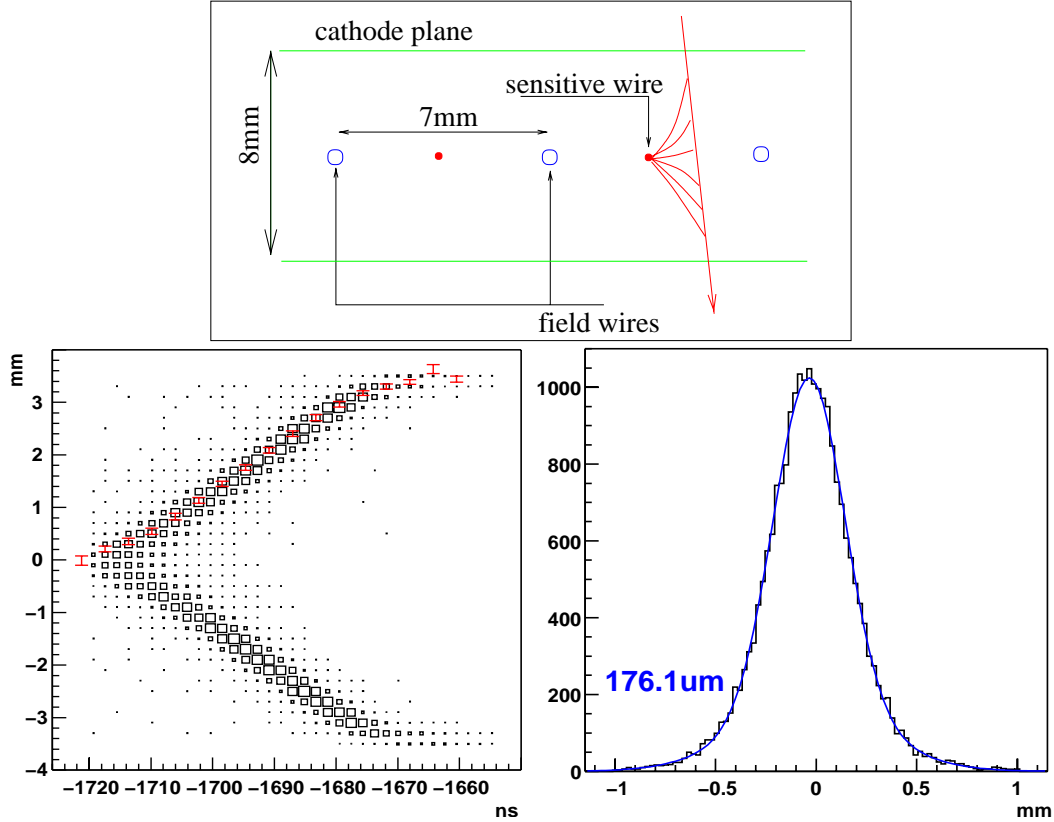


Figure 2.14: Structure of the COMPASS drift chambers cell (above) and spatial resolution (below).

one over the other with a small transfer gap in between with a constant electric field. This allows to keep the amplification factor of each GEM foil low and to adjust the voltage drop at each foil in order to optimize both gain and detector stability. In COMPASS triple GEM technology is used, with gas gain of the order of 10^4 .

- **Drift Chambers and Straws**

Large area tracking in the first spectrometer is performed using drift chambers and straw detectors. The COMPASS drift chambers are characterized by a 8 mm cathode/cathode gap and 7 mm cell size (see Fig.2.14). The typical spatial resolution is $\sim 180 \mu\text{m}$, and the sensitive area size is $1.5 \times 1.0 \text{ m}^2$.

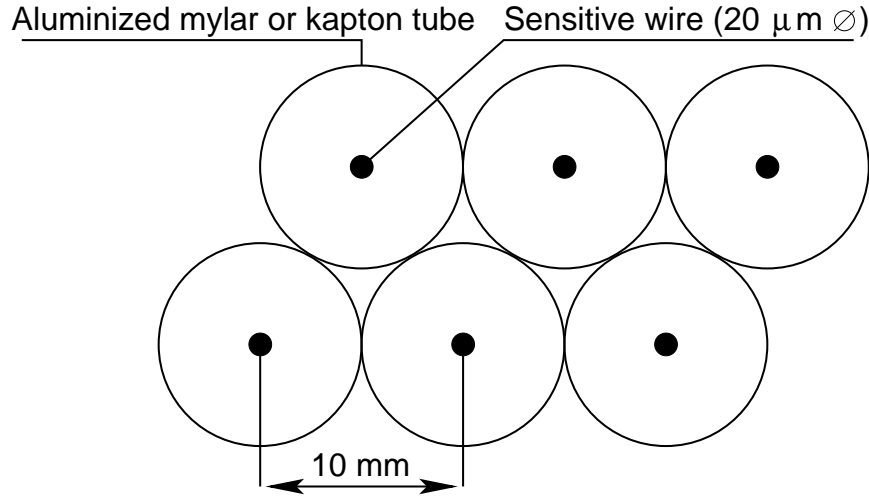


Figure 2.15: Structure of a straw double layer with 10 mm diameter tubes.

Straw detectors consist of thin drift tubes made of aluminized mylar or kapton with a sensitive wire stretched in the middle of the tube. Spacers placed inside the tube grant the accurate wire positioning along the whole tube length. The straw technology allows the construction of drift detectors with large surfaces without adding too much material on the particles flight path. COMPASS straw detectors have a size of $4 \times 5 \text{ m}^2$.

- **W4/5 drift chambers**

The tracking of scattered muons at large Q^2 is mainly based on the W4/5 drift detectors, located in front of the second hadron calorimeter. The W4/5 detectors are two large area drift chamber stations which have been previously used in the SMC experiment. The active area is $4 \times 5 \text{ m}^2$, the drift cell is $\sim 2 \text{ cm}$, and the spatial resolution is $\sim 300 \mu\text{m}$. Due to the big size of the drift cell this chambers cannot sustain high particle fluxes, and the central region is deactivated by a circular dead zone of 50 cm of diameter, which is covered by MWPC stations.

- **MultiWire Proportional Chambers**

MultiWire Proportional Chambers (MWPC) represent the backbone tracking system of the small angle spectrometer. They are characterized by an

	A-type	A*-type	B-type
available chambers:	7	1	6
external dimensions:	202x146 cm ²	202x146 cm ²	202x96 cm ²
active area:	178x120 cm ²	178x120 cm ²	178x80 cm ²
planes/chamber:	3	4	2
wire orientation ⁽¹⁾ :	1v, 2t	1v, 2t, 1h	1v, 1t
dead zone \varnothing :	16 \div 20 mm	16 mm	22 mm
wire pitch:	2 mm		
anode/cathode gap:	8 mm		
wires/plane ⁽¹⁾ :	v,t: 752, h: 512		

(1): v = vertical, t = 10.14° w.r.t. v, h = horizontal

Table 2.3: *Characteristics of Ω MWPC.*

active area of 1780×1200 mm², wire spacing of 2 mm and spatial resolution of ~ 0.7 mm. In COMPASS three types of chambers are used: A-type chambers have three planes of sensitive wires, one vertical and two tilted by ± 10.14 degrees; A*-type is a modified A-type chamber with an additional horizontal plane; B-type have only two planes, one vertical and one either right or left tilted by 10.14 degrees. The dimensions of B-type chambers are reduced to 1780×800 mm². A circular area with diameter of ~ 20 cm around the beam crossing point is deactivated in order to prevent detector damage due to excessive particle flux. All the main characteristics are summarized in Tab.2.3.

- **Muon Wall 1**

Muon identification in the first spectrometer is performed by means of the Muon Wall 1 detector. The detector consists of two large stations

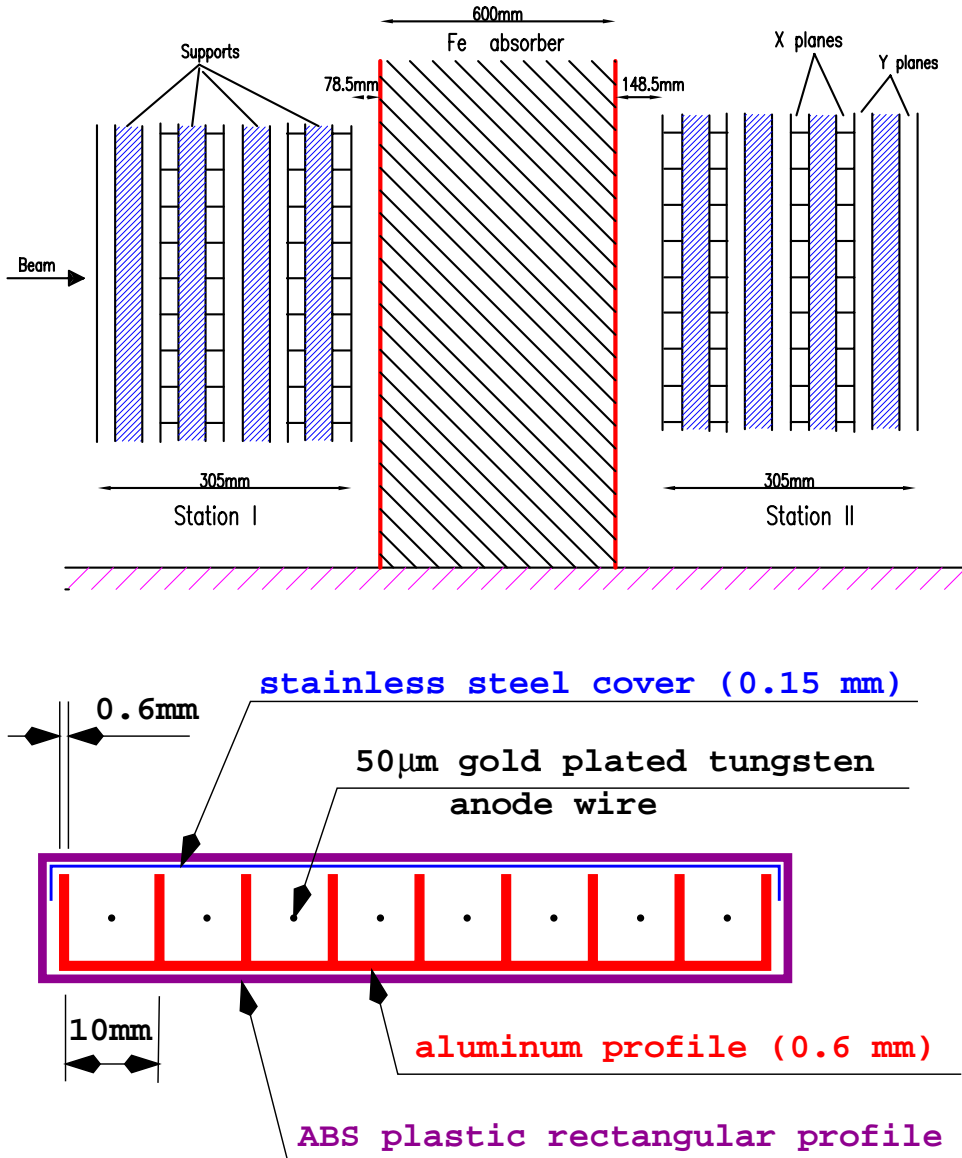


Figure 2.16: Top figure: placement of the two stations of the MW1 detector, before and behind the first hadron absorber. Bottom figure: schematic view of one MW1 module.

of Proportional Tubes, similar to the well-known plastic Iarocci tubes. The stations are placed on both sides of a 0.6 m iron absorber which stops the hadrons (see Fig.2.16). The detector is divided in submodules

composed of a $600\ \mu\text{m}$ thick aluminum profile with 8 cells, closed on the open side by a $150\ \mu\text{m}$ stainless steel cover. A $50\ \mu\text{m}$ gold plated anode wire is stretched in the center of each cell. The submodule is covered by a rectangular plastic profile which provides gas tightness.

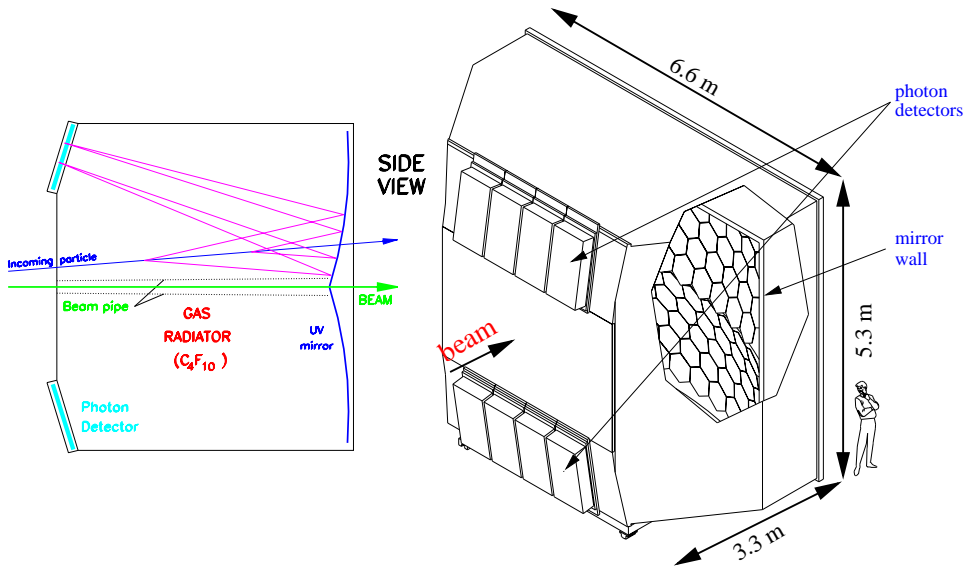


Figure 2.17: Schematic view of Rich 1 detector and of the photon focusing optics.

- **Rich 1**

Particle identification in the first spectrometer is based on a Ring Image Cherenkov detector (RICH 1) designed mainly to separate kaons, pions and protons. Cherenkov photons emitted by relativistic particles above threshold in the 3.3 m long gas radiator are reflected by a mirror wall on two matrices of photon detectors placed above and below the entrance window and outside the spectrometer acceptance. Particles crossing the detector with positive angles with respect to the horizontal plane emit photons which are focused on the upper photon detectors. Photons emitted along the particle path at fixed Cherenkov angle are focused by the mirror wall and form ring images on the photon detectors plane. The ring radius is proportional to the Cherenkov emission angle and allows to

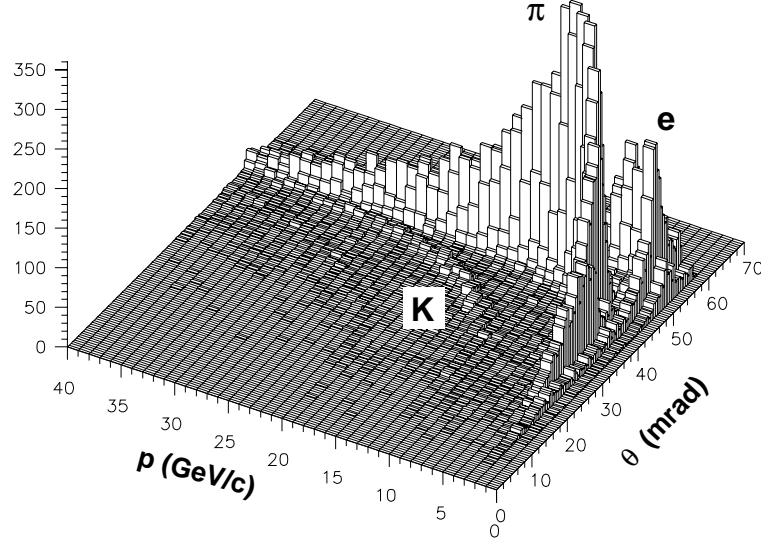


Figure 2.18: *RICH results: particle momentum versus measured Cherenkov angle.*

calculate the particle speed, once the reflective index of the gas is known, from the well known formula:

$$\cos \theta_c = \frac{\beta}{n} \quad (2.3)$$

Results on particle identification with RICH1 are shown in Fig.2.18.

2.4 Data Acquisition System

The COMPASS Data Acquisition System (DAQ) is designed to collect ~ 2 Terabyte of data per day, at a maximum trigger rate of 10^5 Hz. Such performances are achieved by performing data digitalization at the front-end level wherever possible and by using pipelines at various levels to reduce the readout dead time (see Fig.2.19). Frontend data digitalization is performed mainly using dedicated TDC chips ($\mathcal{F}1$ TDC)[12] which convert wire address and hit time to a formatted 24-bits word which is transmitted to CATCH modules through a 40 Mbits/s serial connection. Local event building is performed in hardware by CATCH

modules using FPGA technology and data are transmitted to the Global Event Builder machines through an high-speed fiber optics S-Link connection. Finally the full event is sent to the Central Data Recording at CERN IT division for permanent storage on tape.

The DAQ system allows to parasitically extract samples of the recorded events for monitoring purposes, without degrading the recording speed. This feature is used by the online monitoring software for controlling the overall performances of the various detectors through the graphical presentation of the characterizing detector parameters, as the hit profiles, hit time correlations, ADC spectra, etc. The graphical interface of the monitoring software is shown in Fig.2.20

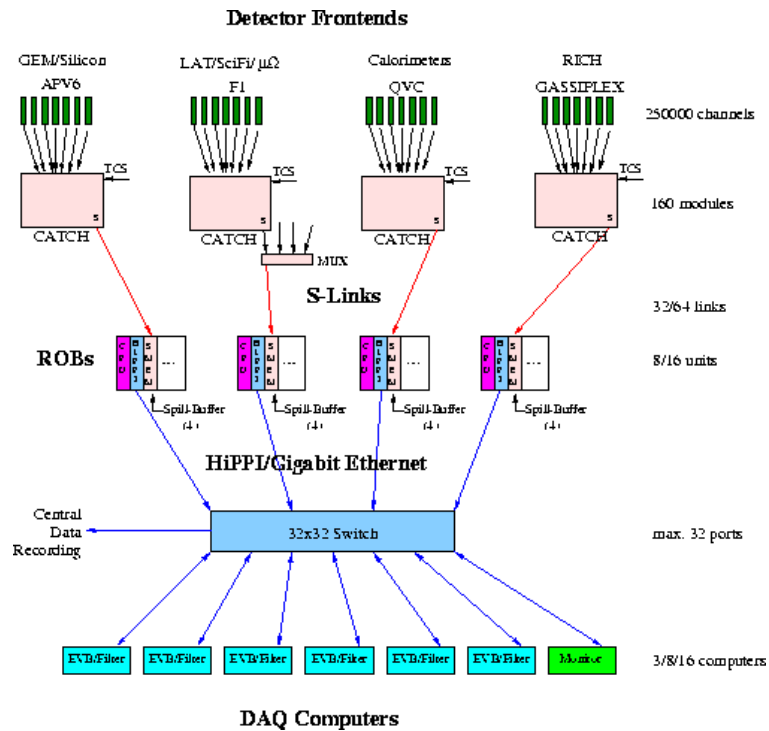


Figure 2.19: Schematic view of the COMPASS DAQ. The frontend data is collected by the CATCH modules, which implement in hardware the task of local event builders. The data is then transmitted to the global event builders (EVB) through readout buffer machines (ROB). The ROBs are equipped with large memories, enough to store the data corresponding to one complete spill.

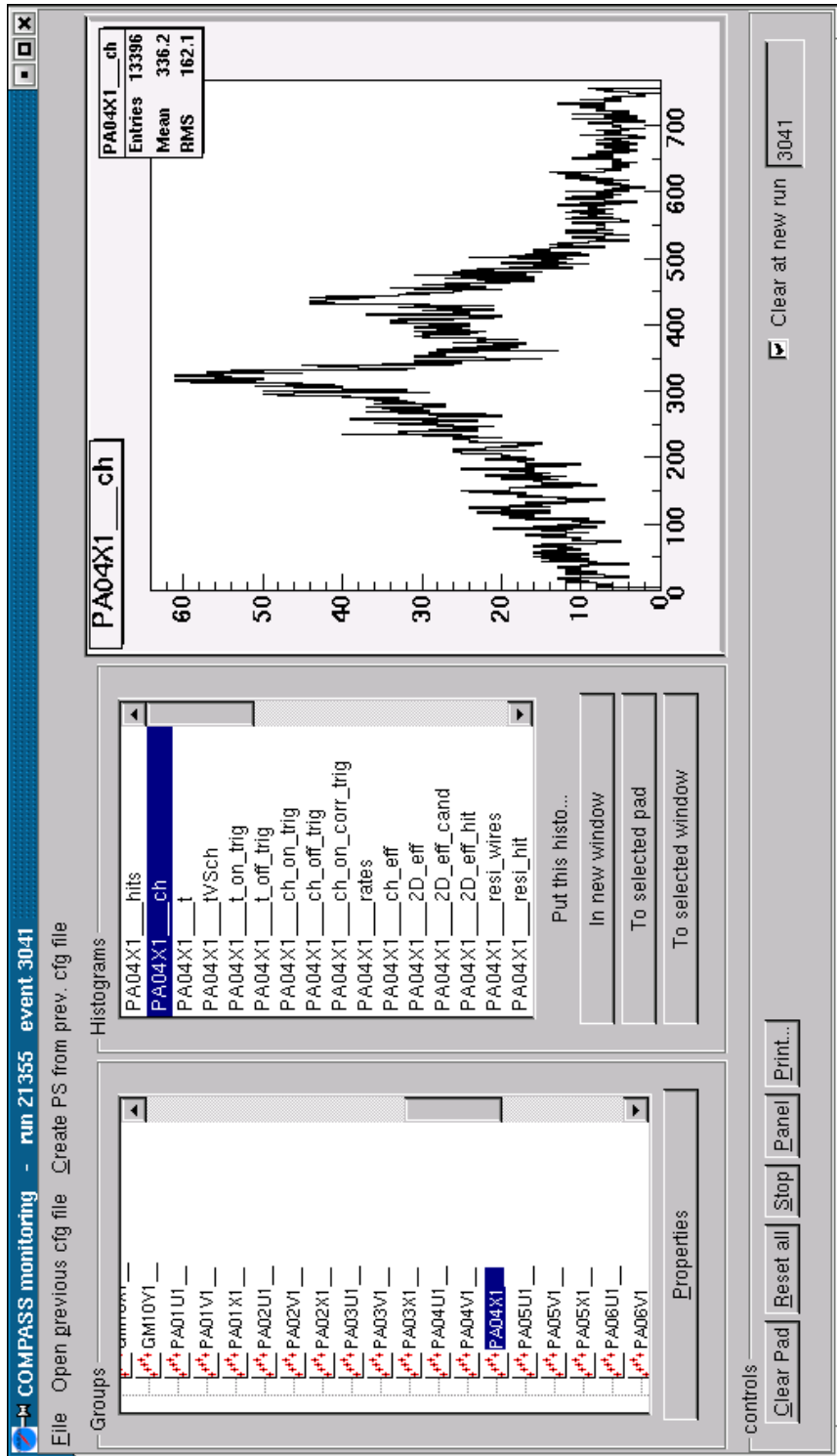


Figure 2.20: The COMPASS online monitoring program (COOL).

Chapter 3

Seminclusive Λ production from polarized targets

The scattering of polarized leptons on polarized nucleon targets has been used since 1974 by several experiments to probe the structure of the nucleons and to measure the partonic and gluonic contribution to the total nucleon spin.

The data from polarized lepton-nucleon deep-inelastic scattering (DIS) show that the spin structure of nucleons is much more complicated than what is described by the parton model. The experimental results suggest that contributions from sea-quarks, gluons and their angular momenta can be significant. In this context measurements of spin dependent structure functions for the individual gluon and strange quark contributions are needed. Moreover, the transversity distribution $h_1(x)$, which is the equivalent of the $g_1(x)$ for a transversely polarized target, is currently of great theoretical interest. The measurement of $h_1(x)$ involves the study of semi-inclusive hadron leptonproduction, and requires additional informations on the quark fragmentation functions available from polarized proton-proton scattering.

In this chapter I will describe the physics of polarized DIS and sketch the measurements that are foreseen in COMPASS, with a special highlight on semi-inclusive Λ leptonproduction and on the feasibility of the measurement of $h_1(x)$ in this particular channel.

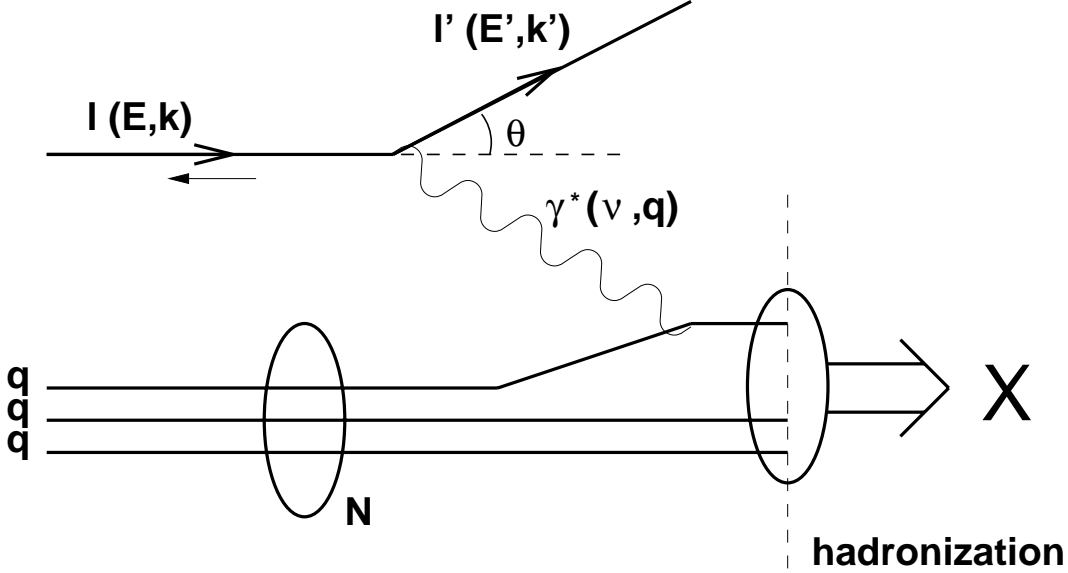


Figure 3.1: Deep-inelastic lepton-nucleon scattering process. The lepton emits a virtual photon which is absorbed by one of the quarks in the nucleon. The struck quark and the target remnants fragment into hadrons in the final state.

3.1 Lepton-nucleon scattering process

In the lepton-nucleon DIS, schematically described in Fig.3.1, an incoming lepton emits a virtual photon which is absorbed by a quark in the nucleon (struck quark). As a consequence, the nucleon breaks up and the struck quark and the target remnants fragment into hadrons in the final state. In the inclusive DIS measurement only the incoming and scattered leptons are detected. Semi-inclusive measurements require the additional detection of one of the hadrons in the final state, and provide additional informations on the nucleon structure and on the properties of the DIS process.

3.1.1 Kinematics

The lepton-nucleon DIS kinematics is usually described by the following Lorentz-invariant variables:

1. The squared four-momentum transfer,

$$Q^2 = -q^2 = -(k - k')^2 = 4EE' \sin^2(\theta/2); \quad (3.1)$$

2. the energy transfer to the hadronic system,

$$\nu = p \cdot q/M = E - E'; \quad (3.2)$$

3. the square of the invariant mass of the photon-nucleon system,

$$W^2 = M^2 + 2M\nu - Q^2; \quad (3.3)$$

4. the Bjorken scaling variable,

$$x = Q^2/2p \cdot q = Q^2/2M\nu; \quad (3.4)$$

5. the fraction of the incident energy transferred,

$$y = p \cdot q/p \cdot k = \nu/E. \quad (3.5)$$

In these equations, k , k' , p and q are the four-vectors of the incoming and scattered lepton, the target nucleon and the exchanged boson. M is the mass of the target nucleon, whereas the lepton mass is not considered. E , E' are the energies of the incident and scattered lepton, and θ the lepton scattering angle in the laboratory frame.

3.1.2 The inclusive DIS cross-section

The inclusive lepton-nucleon DIS cross-section can be expressed as

$$\frac{d^2\sigma}{d\Omega dE'} = \frac{\alpha^2}{2MQ^4} \frac{E'}{E} L_{\mu\nu} W^{\mu\nu} \quad (3.6)$$

where α is the fine structure constant and M is the mass of the target nucleon. $L_{\mu\nu}$ is a tensor which describes the lepton-photon vertex; it can be calculated exactly in the framework of Quantum Electro Dynamics (QED):

$$L_{\mu\nu} = \sum_{\lambda'} [\bar{u}(k', \lambda') \gamma_\mu u(k, \lambda)] [\bar{u}(k', \lambda') \gamma_\nu u(k, \lambda)]^*, \quad (3.7)$$

where k (k') and λ (λ') are the four-momentum and the helicity of the incoming (scattered) lepton, $u(k, \lambda)$ is the Dirac spinor, γ_μ are the Dirac matrices, and $\epsilon_{\mu\nu\alpha\beta}$ is the totally antisymmetric Levi-Civita tensor. In the expression for $L_{\mu\nu}$ an implicit summation is performed over repeated indexes. The $L_{\mu\nu}$ can be factorized in two terms, one symmetric (S) and one anti-symmetric (A) under the exchange of the μ and ν indexes:

$$L_{\mu\nu} = 2L_{\mu\nu}^S + 2iL_{\mu\nu}^A, \quad (3.8)$$

$$L_{\mu\nu}^S = k_\mu k'_\nu + k'_\mu k_\nu - g_{\mu\nu} k \cdot k', \quad (3.9)$$

$$L_{\mu\nu}^A = \lambda \epsilon_{\mu\nu\rho\beta} k^\rho k'^\beta. \quad (3.10)$$

The tensor $W^{\mu\nu}$ describes the absorption of the virtual photon by the nucleon; it contains the whole information about the unknown nucleon structure, and there is not an exact calculation for it. The form of the $W_{\mu\nu}$ tensor can be derived from symmetry considerations and conservation laws; as for $L_{\mu\nu}$, it can be splitted into symmetric and anti-symmetric terms [4]:

$$W_{\mu\nu} = W_{\mu\nu}^S + iW_{\mu\nu}^A, \quad (3.11)$$

$$W_{\mu\nu}^S = -g_{\mu\nu} F_1(\nu, Q^2) + \frac{p_\mu p_\nu}{\nu} F_2(\nu, Q^2), \quad (3.12)$$

$$W_{\mu\nu}^A = \frac{\epsilon_{\mu\nu\alpha\beta} q^\alpha s^\beta}{\nu} g_1(\nu, Q^2) + \frac{\epsilon_{\mu\nu\alpha\beta} q^\alpha (p \cdot q s^\beta - s \cdot q p^\beta)}{\nu^2} g_2(\nu, Q^2). \quad (3.13)$$

Taking into account the fact that the product of a symmetric tensor with and anti-symmetric one is null, the DIS cross-section can be written as

$$\frac{d^2\sigma}{d\Omega dE'} = \frac{\alpha^2}{2MQ^4} \frac{E'}{E} [L_{\mu\nu}^{(S)} W^{\mu\nu(S)} - L_{\mu\nu}^{(A)} W^{\mu\nu(A)}]. \quad (3.14)$$

The symmetric and the anti-symmetric terms in square brackets in Eq.3.14 are related to the unpolarized and to the polarized DIS processes respectively.

3.1.3 Unpolarized lepton-nucleon scattering

When the DIS cross-section is averaged over all the possible spin states of the incoming and outgoing lepton and over those of the target (unpolarized DIS), the differential cross-section is expressed by

$$\frac{d^2\sigma}{dQ^2 dx} = \frac{4\pi\alpha^2}{Q^2} \frac{1}{x} \left[xy^2 F_1(\nu, Q^2) + \left(1 - y - \frac{Mxy}{2E}\right) F_2(\nu, Q^2) \right] \quad (3.15)$$

In passing from Eq.3.14 to Eq.3.15 the azimuthal lepton emission angle Φ has been integrated over the 2π range and a transformation $dE' d\cos\theta \rightarrow dQ^2 dx$ has been performed.

When $(\nu, Q^2) \rightarrow \infty$ but the ratio ν/Q^2 , and therefore x , remains finite (Bjorken limit), the structure functions are no more dependent on both Q^2 and ν , but they are only functions of one variable $x = Q^2/2M\nu$ (Bjorken scaling). The dependence on a single kinematical variable is typical of the scattering off point-like particles. If the nucleon is considered in a reference frame where its momentum is infinite, the transverse momentum of the partons inside the nucleon can be neglected and the x variable is interpreted as the fraction of the nucleon momentum carried by the struck parton.

In the Quark Parton Model (QPM), the scaling phenomenon has been interpreted as due to the incoherent sum of scattering processes between the lepton and point-like spin- $\frac{1}{2}$ constituents of the nucleon, called quarks.

The structure functions $F_1(x)$ and $F_2(x)$ have a simple interpretation in the Quark Parton Model (QPM):

$$F_1(x) = \frac{1}{2} \sum_i e_i^2 [q_i(x) + \bar{q}_i(x)] \quad (3.16)$$

$$F_2(x) = \sum_i e_i^2 x [q_i(x) + \bar{q}_i(x)], \quad (3.17)$$

where $q_i(x)$ ($\bar{q}_i(x)$) is the probability to find a quark (anti-quark) of flavor i and carrying a fraction x of the momentum of the nucleon, and e_i the charge of the quark flavor i . In the hypothesis of spin- $\frac{1}{2}$ quarks, the two structure functions $F_1(x)$ and $F_2(x)$ obey the so-called Callan-Gross relation:

$$F_2(x) = 2xF_1(x). \quad (3.18)$$

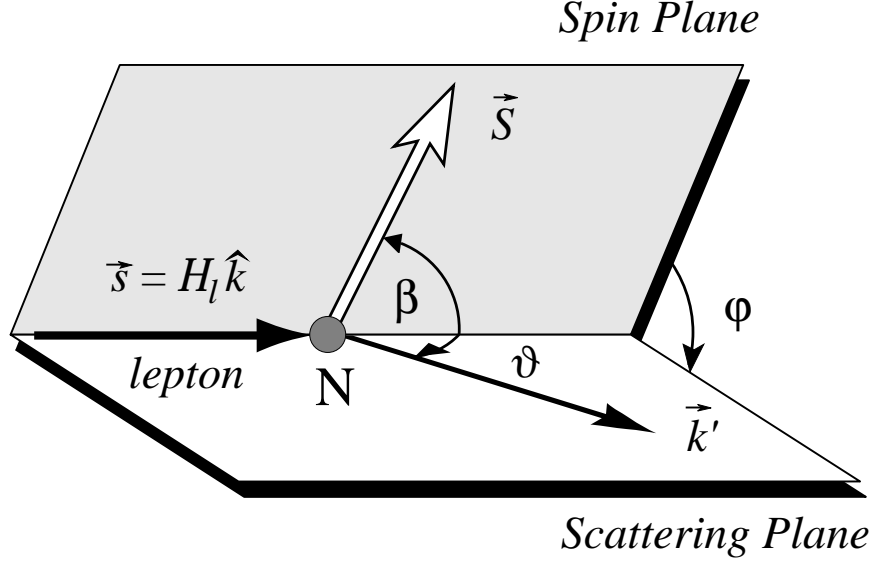


Figure 3.2: Kinematics of polarized deep-inelastic lepton-nucleon scattering

3.1.4 Polarized lepton-nucleon scattering

Let's now turn to the case of a longitudinally polarized lepton beam and a polarized target. The differential cross-section additionally depends on the relative direction of the target spin and of the incoming lepton momenta. It is useful to introduce two planes, the scattering plane and the spin plane, defined respectively by the directions of the incoming and scattered lepton, and of the incoming lepton and target spin axis (see Fig3.2). The corresponding cross-section can be decomposed into a spin-independent term σ_0 and a spin-dependent contribution $\Delta\sigma$ [2]:

$$\frac{d^3\sigma(\beta)}{dQ^2 dx d\phi} = \frac{d^3\sigma_0}{dQ^2 dx d\phi} - \frac{d^3\Delta\sigma(\beta)}{dQ^2 dx d\phi}, \quad (3.19)$$

where the β represents the angle between the target spin vector and the incoming lepton momentum vector, and ϕ is the angle between the scattering and the spin planes, as shown in Fig.3.2.

The spin-independent part $\frac{d^3\sigma_0}{dQ^2 dx d\phi}$ of the cross-section is given by Eq.3.15, while the spin-dependent part $\frac{d^3\Delta\sigma(\beta)}{dQ^2 dx d\phi}$ contains the structure functions $g_1(\nu, Q^2)$

and $g_2(\nu, Q^2)$, related to the anti-symmetric part of the hadronic tensor $W_{\mu\nu}$ (see Eq.3.13). It can be expressed as

$$\begin{aligned} \frac{d^3\Delta\sigma(\beta)}{dQ^2 dx d\phi} \propto & \cos(\beta) [a g_1(\nu, Q^2) + b g_2(\nu, Q^2)] \\ & - \cos(\phi) \sin(\beta) c \left[\frac{y}{2} g_1(\nu, Q^2) + g_2(\nu, Q^2) \right]. \end{aligned} \quad (3.20)$$

The parameters a , b and c are of the order $\mathcal{O}(1)$, $\mathcal{O}(1/Q^2)$ and $\mathcal{O}(1/Q)$ respectively when $Q^2 \rightarrow \infty$. Therefore, the distribution $g_1(x)$ dominates at large Q^2 in the first term of the right hand side of the above expression.

The functions $g_1(\nu, Q^2)$ and $g_2(\nu, Q^2)$ present the same scaling properties as $F_1(x)$ and $F_2(x)$. The $g_1(x)$ function has a simple interpretation in the QPM:

$$g_1(x) = \frac{1}{2} \sum_i e_i^2 [(q_i^+(x) + \bar{q}_i^+(x)) - (q_i^-(x) - \bar{q}_i^-(x))], \quad (3.21)$$

where $q_i(x)$ ($\bar{q}_i(x)$) is the probability of finding a quark (anti-quark) of flavor i , at a certain value of x . The sign $+$ ($-$) indicates that the spin of the quark is oriented in the same (opposite) direction as the spin of the nucleon. The same interpretation holds for $\bar{q}_i^+(x)$ ($\bar{q}_i^-(x)$) in the case of anti-quarks.

An intuitive picture of how $g_1(x)$ is sensitive to spin degrees of freedom and how it can be measured in polarized lepton-nucleon DIS is given in Fig.3.3. When a beam of longitudinally polarized leptons is scattered off a longitudinally polarized target, the lepton preferentially emits a circularly polarized virtual photon with spin projection pointing to the right of the picture[4]. Angular momentum conservation requires that the virtual photon is absorbed by a quark with spin oriented in **opposite** direction (pointing to the left of the picture), since the final state quark must have spin $1/2$. This corresponds to a value of the angle β appearing in Eq.3.20 of either 0° or 180° . Therefore, only the first term in the right hand side of that equation contributes. In this term the main contribution comes from the $g_1(x)$ distribution, since $g_2(x)$ is suppressed by the parameter b .

When the spin orientation of the nucleon is that of the struck quark (top picture in Fig.3.3), the cross-section is sensitive to the distribution $q^+(x)$. This

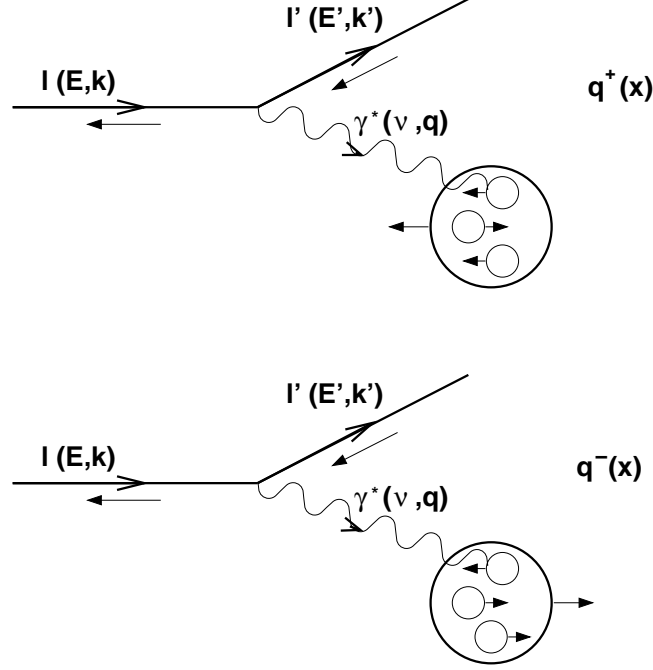


Figure 3.3: *Diagram of polarized deep inelastic scattering. A longitudinally polarized lepton emits a virtual photon that preferentially hits a quark with opposite spin orientation. The upper diagram is therefore sensitive to the $q^+(x)$ distribution, while the bottom one is sensitive to $q^-(x)$ (see text).*

cross-section will be denoted as $\sigma^{\uparrow\downarrow}$, where \uparrow is the direction of the spin of the incoming lepton and \downarrow is the direction of the nucleon spin. If the target spin is reversed (pointing to the left of the picture) the measured cross-section corresponds to parallel beam and target spins, and is denoted as $\sigma^{\uparrow\uparrow}$ (bottom picture in Fig.3.3). The elementary photon-quark interaction is however unchanged, because the spin of the struck quark must be oriented anti-parallel to the photon spin. Therefore, the $\sigma^{\uparrow\uparrow}$ cross-section is sensitive to the $q^-(x)$ distribution. The same rule holds when the virtual photon is absorbed by an anti-quark.

In inclusive DIS the flavor of the struck quark is not identified, and a summation over all possible quarks and anti-quarks is implicit in the measurement. From Eq.3.21 it is clear that $g_1(x)$ can be extracted from the double-spin asym-

metry

$$A_{\parallel} = \frac{\sigma^{\uparrow\downarrow} - \sigma^{\uparrow\uparrow}}{\sigma^{\uparrow\downarrow} + \sigma^{\uparrow\uparrow}}. \quad (3.22)$$

As already mentioned, the longitudinal spin asymmetry is dominated by the $g_1(x)$ structure function. To access $g_2(x)$, the target spin has to be oriented perpendicular to the beam direction ($\beta = 90^\circ$ or 270°), thus selecting the second term in Eq.3.20. However, the asymmetry is in this case suppressed by a factor $c \sim \mathcal{O}(1/Q)$, making the measurement more difficult.

3.1.5 Sum rules

The measurement of the parton distribution functions can give important informations on the validity of the models used to depict the internal structure of the nucleon. The experimental observation of the scaling properties of the $F_1(\nu, Q^2)$ and $F_2(\nu, Q^2)$ structure functions confirmed the hypothesis that the nucleon is made up of point-like, spin 1/2 constituents. The small deviations from the scaling observed have been explained by the Altarelli-Parisi equations in the perturbative QCD (PQCD) theory.

No theoretical predictions are currently available for the x -dependence of the spin-dependent structure functions. Nevertheless definite predictions do exist for the quantities

$$\Gamma_1^N = \int_0^1 g_1^N(x) dx, \quad (3.23)$$

where the superscript N denotes the nucleon type (p or n). A certain number of sum rules have been derived, which relate the Γ_1^N to the polarizations Δq_i of the quarks in the nucleon, defined as

$$\Delta q_i = \int_0^1 [(q_i^+(x) + \bar{q}_i^+(x)) - (q_i^-(x) + \bar{q}_i^-(x))] dx. \quad (3.24)$$

The most fundamental of these relations is the so-called **Bjorken sum rule**, which predicts the difference of the Γ_1 moments for the proton and the neutron:

$$\Gamma_1^p - \Gamma_1^n = \frac{1}{6}(\Delta u - \Delta d) = \frac{1}{6} \left| \frac{g_A}{g_V} \right|, \quad (3.25)$$

where g_A and g_V are the axial and vector weak coupling constants of the neutron β -decay. The Bjorken sum rule does not rely on a particular model of the internal structure of the nucleons, and is verified at a 10% level by several DIS experiments.

3.1.5.1 The nucleon spin and the Ellis-Jaffe sum rules

The most general expression for the spin of the proton (or the neutron) is given, in units of \hbar , by

$$\frac{1}{2} = \frac{1}{2}\Delta\Sigma + \Delta G + L. \quad (3.26)$$

In this expression $\Delta\Sigma = \sum_i \Delta q_i$ is the overall quark contribution to the nucleon spin, and L is the angular momentum contribution. The gluon contribution ΔG is given by

$$\Delta G = \int_0^1 [G^+(\eta) - G^-(\eta)] d\eta, \quad (3.27)$$

where $G^{+(-)}(\eta)$ is the probability of finding a gluon with the spin oriented parallel (anti-parallel) to the nucleon spin, and η is the fraction of the nucleon momentum carried by the gluon.

A prediction for the quantity $\Delta\Sigma$ can be derived in the hypothesis that the nucleon is composed only by three quarks flavours ($u(\bar{u})$, $d(\bar{d})$ and $s(\bar{s})$). Inserting in Eq.3.23 the expression 3.21 for $g_1(x)$ one gets

$$\Gamma_1^p = \frac{1}{2} \left(\frac{4}{9}\Delta u + \frac{1}{9}\Delta d + \frac{1}{9}\Delta s \right), \quad (3.28)$$

and, from isospin invariance,

$$\Gamma_1^n = \frac{1}{2} \left(\frac{1}{9}\Delta u + \frac{4}{9}\Delta d + \frac{1}{9}\Delta s \right), \quad (3.29)$$

Additional informations on the Δq_i can be extracted from the neutron and hyperon β decays. Linear combinations of the Δq_i moments are related to the

weak axial-vector constants a_0 , a_3 and a_8 by [2][3]

$$a_0 = \Delta u + \Delta d + \Delta s \equiv \Delta\Sigma \quad (3.30)$$

$$a_3 = \Delta u - \Delta d = \left| \frac{g_A}{g_V} \right| = F + D \quad (3.31)$$

$$a_8 = \Delta u + \Delta d - 2\Delta s = 3F - D, \quad (3.32)$$

where F and D are $SU(3)$ coupling constants. Then Eqs.3.28–3.29 can be rewritten as

$$\Gamma_1^{p(n)} = +(-)\frac{1}{12}a_3 + \frac{1}{36}a_8 + \frac{1}{9}a_0. \quad (3.33)$$

The equations 3.33 are the Ellis-Jaffe sum rules for the proton and the neutron.

Using the values of $F + D = 1.257 \pm 0.003$ and $F/D = 0.575 \pm 0.016$ derived from the neutron and hyperon β -decays, and with the hypothesis that the strange sea-quarks are unpolarized ($\Delta s = 0$, $a_0 = a_8$), one obtains $\Delta\Sigma = \Delta u + \Delta d \simeq 0.58$ [2]. The corresponding Ellis-Jaffe prediction for Γ_1^p is $\Gamma_1^p \simeq 0.19$.

3.2 EMC and the “proton spin crisis”

The Ellis-Jaffe sum rules can be verified by measuring the $g_1(x)$ distribution over a range of x as wide as possible, both with proton and neutron targets. The experimental value of Γ_1^p can also be used to derive, together with the a_3 and a_8 constants, the total quark contribution $\Delta\Sigma$ to the nucleon spin. The measurement has been performed by the EMC experiment at CERN (see Fig.3.4), with a surprising result: the calculated value of Γ_1^p implied that $\Delta\Sigma = 0.12 \pm 0.16$ [5], i.e. the quark contribution to the proton spin is compatible with zero.

The EMC result, confirmed later by the SMC and E154 experiments, raised new possible descriptions for the structure of the nucleon within the framework of QCD. However, despite the theoretical efforts a clear explanation of the polarized DIS results is still missing. Various models have been proposed which either claim for a large gluon contribution to the nucleon spin, or introduce

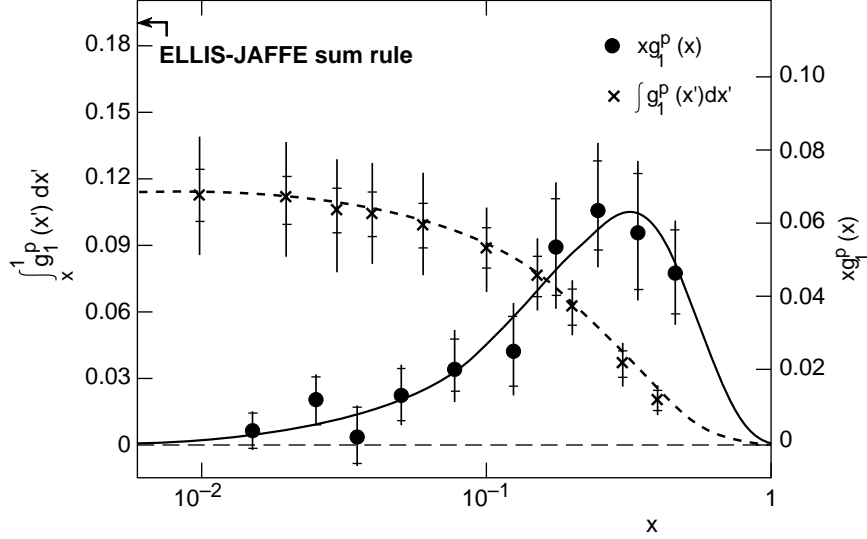


Figure 3.4: *The EMC result for the Ellis-Jaffe sum rule: solid line is the measured structure function $xg_1^p(x)$, dashed line is the integral $\int_x^1 g_1^p(x') dx'$. The measured $\Gamma_1^p = \int_0^1 g_1^p(x) dx$ is compared with the Ellis-Jaffe prediction. Curves are plotted as a function of x .*

a seizable negative strange sea-quark polarization which effectively “screens” the valence quark contribution. Angular momentum could contribute as well to the nucleon spin, but it is not clear if a direct measurement of the angular momentum contribution is possible [6]. In order to discriminate between the existing models, it is necessary to perform semi-inclusive polarized DIS measurements, which allow to independently measure the quark and gluon polarized distribution functions and verify the model predictions.

3.3 Gluon polarization $\Delta G/G$ and COMPASS

A possible solution of the “proton spin crisis” can be given by a large contribution of the gluon polarization. It is therefore mandatory to perform precise measurements of the polarized gluon distribution function $\Delta G(\eta)$.

The measurement of $\Delta G(\eta)/G(\eta)$ distribution, where $G(\eta)$ represents the unpolarized gluon distribution function, is one of the main programs of the

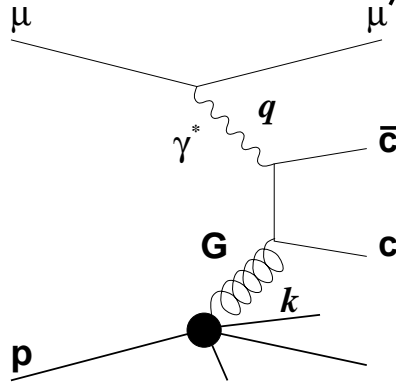


Figure 3.5: *The photon-gluon fusion diagram. The incoming muon emits a virtual photon which is absorbed by a gluon in the nucleon, through an intermediate state, that fragments into charmed hadrons in the final state.*

COMPASS experiment. Among the various suggestions to measure the gluon distribution function, the most promising for COMPASS is the longitudinal spin asymmetry of open charm leptonproduction. In the dominant process for this production, charm quarks are produced via the photon-gluon fusion (PGF) diagram shown in Fig.3.5. The $\Delta G(\eta)$ distribution is extracted from the measured asymmetry $A_{\mu N}^{c\bar{c}}$ for charm muon-production, with parallel and anti-parallel longitudinal spin orientations of the incoming muon and of the target nucleon. The experimentally measured asymmetry for charm production

$$(3.34)$$

where $N_{c\bar{c}}^{\uparrow\downarrow}$ ($N_{c\bar{c}}^{\uparrow\uparrow}$) is the number of charm events for anti-parallel (parallel) orientations of the spin of the incoming lepton and of the target nucleon, is proportional to the virtual photon asymmetry $A_{\gamma N}^{c\bar{c}}$. This quantity is in turn function of the ratio $\Delta G(\eta)/G(\eta)$.

The charm quarks hadronize into D^0 mesons approximately 60% of the time, and 20% in D^{*+} . The COMPASS apparatus has been specifically designed to detect, from their $D^0 \rightarrow K^- \pi^+$ decay, D^0 mesons produced in the hadronization of the $c\bar{c}$ pair. The $D^{*+} \rightarrow D^0 \pi^+ \rightarrow (K^- \pi^+) \pi^+$ decay can be tagged by requiring an additional soft pion in the final state. A specific trigger has also

been prepared to allow the reconstruction of the DIS kinematics in the quasi-real photon absorption region ($Q^2 \approx 0$), where the cross-section for the photon-gluon fusion process is expected to be maximal.

3.4 Polarization of strange sea quarks of the nucleon

The content of strange quarks in the nucleon is currently of great theoretical and experimental interest. This because the strange sea quarks could be negatively polarized with respect to the parent nucleon, and this fact could explain the small value of $\Delta\Sigma = \Delta u + \Delta d + \Delta s$ measured by the experiments.

A model has been proposed in [7] which is based on two major observations. First of all, the fact that the masses of pions and kaons are small at the typical hadronic scale can be attributed to the existence of a strong attraction between quarks and anti-quarks in the pseudoscalar channel $J^{PC} = 0^{-+}$. Therefore all the diquark pairs $q_i\bar{q}_j$ inside the nucleon are expected to have these quantum numbers, which imply spin singlet states. The second point is that in the vacuum the quantum numbers must be $J^{PC} = 0^{++}$, therefore only the 3P_0 state is allowed. Between the diquark pairs of the vacuum the strange content ($s\bar{s}$) is expected to have a density of the same order of that of light quarks.

Therefore a possible contribution to the proton spin can be attributed to an interaction of the valence quarks of the proton wave function with the strange anti-quarks of the vacuum, through an interaction that should select a spin singlet state. Therefore the spin of the strange anti-quarks must be opposite oriented to that of the valence quark. The requirement that the $s\bar{s}$ pairs of the vacuum are in a triplet state supports that also the s quark has a spin opposite to that of the valence quark, and hence to that of the nucleon.

Therefore the contribution to the nucleon spin of the strange content (Δs) is expected to be negative.

3.4.1 Λ polarization in deep inelastic lepton scattering

The Λ hyperons produced in polarized deep inelastic lepton scattering are of special interest for the measurement of the strange sea-quarks properties. In the simple parton model representation of hadrons, the Λ hyperons contain a ud pair in a singlet state, and an s quark carrying all the spin of the hyperon. The Λ hyperons undergo weak decay in $p\pi^-$ in 64% of the cases. Due to the parity-violating properties of weak interactions, the distribution of the decay products in the Λ rest frame is not isotropic, but the proton is emitted preferentially in the direction of the Λ spin vector. Thus if the produced Λ have a net polarization P_Λ along a certain direction \hat{n} , the decay proton distribution is given by

$$W(\theta_p^*) = \frac{1}{4\pi} [1 + \alpha P_\Lambda \cos(\theta_p^*)], \quad (3.35)$$

where θ_p^* is the polar angle between the proton direction and the polarization axis \hat{n} , measured in the Λ center of mass frame.

When discussing the production of Λ hyperons in polarized lepton-nucleon DIS, it is fundamental to distinguish between the *target fragmentation region* and the *current fragmentation region*, according to the sign of the x -Feynman (x_F) kinematical variable, defined as the ratio between the measured and the maximum allowed Λ longitudinal momenta, in the γ^*N center of mass frame:

$$x_F = \frac{p_L^\Lambda}{p_{L_{MAX}}^\Lambda} = \frac{2p_L^\Lambda}{\sqrt{s}}. \quad (3.36)$$

3.4.2 Target fragmentation region

The polarization of the strange sea quarks inside the nucleon can be measured by looking at the polarization of Λ s produced at $x_F < 0$ (target fragmentation region) in the deep-inelastic scattering of longitudinally polarized muon beams on longitudinally polarized nucleons. The argument is based on the fact that the cross-section for the electromagnetic interaction of longitudinally polarized leptons and nucleons depends on the longitudinal polarization states of the quarks in the nucleon[9]. The exchanged virtual photon is preferentially absorbed by a quark with the spin oriented anti-parallel to the spin of the beam muon. In this

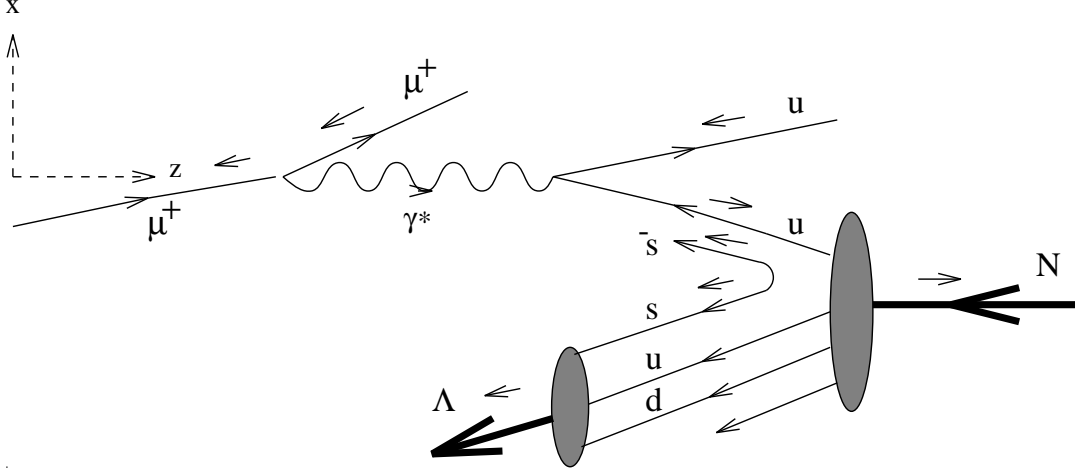


Figure 3.6: *Expected Λ polarization in deep-inelastic μN scattering on a valence u quark. The exchanged virtual photon preferentially selects a quark with the spin aligned parallel to the proton spin. The opposite orientation of the pair is reflected in the polarization of the Λ produced in the target fragmentation region.*

case the Λ hyperons produced in the target fragmentation region are expected to originate from a spin-triplet $s\bar{s}$ pair with their spins oriented anti-parallel to the spin of the struck quark. If s -quark polarization is transferred to the final state Λ , the observation of the distribution of the decay proton with respect to the direction of the virtual photon gives information on the polarization of the strange sea-quarks.

3.5 The transverse spin of the nucleon

The leading-order description of the internal structure of the nucleon requires, for each quark and anti-quark flavor, three different distribution functions. Two of them are the $q(x)$ and $\Delta q_i(x) = q_i^+(x) - q_i^-(x)$ distributions which have been already introduced when discussing the unpolarized DIS and the polarized DIS with longitudinally polarized targets respectively. The third one is introduced to describe the distribution of quarks in a transversely polarized nucleon, and

is denoted as $\Delta_T q_i(x) = q_i^\uparrow(x) - q_i^\downarrow(x)$. Here $q_i^\uparrow(x)$ ($q_i^\downarrow(x)$) represents the probability of finding a quark at a certain x and with its spin oriented parallel (anti-parallel) to the **transverse** spin of the nucleon. A similar definition applies for anti-quarks. We also introduce a new structure function, denoted with $h_1(x)$ and defined as:

$$h_1(x) = \sum_i \left[(q_i^\uparrow + \bar{q}_i^\uparrow) - (q_i^\downarrow + \bar{q}_i^\downarrow) \right], \quad (3.37)$$

which is the equivalent of $g_1(x)$ for a transversely polarized target.

Transversity is a pure relativistic effect: when the nucleon is at rest, there is no reason why the quark spin distributions should differ if the nucleon spin is rotated of 90 degrees. However the distribution functions are usually defined in the so-called infinite momentum frame, in which the transverse momentum of the quarks can be neglected. In this frame the longitudinal spin configuration is not equivalent to the transverse spin anymore, since Lorentz boosts and rotations do not commute. The difference between helicity and transversity distributions therefore derives from relativistic effects which can give important informations on the internal structure of nucleons.

The measurement of the transversity distribution functions is experimentally more difficult than the elicity-dependent distributions $\Delta q_i(x)$. The transverse spin of quarks cannot be probed in inclusive DIS with longitudinally polarized leptons. An intuitive argument to explain this fact is the following. The transverse spin states can be expressed in the helicity basis as:

$$|\uparrow\rangle = \frac{1}{2} [|+\rangle + |-\rangle], \quad |\downarrow\rangle = \frac{1}{2} [|+\rangle - |-\rangle]. \quad (3.38)$$

Therefore a transversely polarized quark can be found in both $+$ and $-$ helicity states with equal probability. For this reason, the elementary lepton-quark interaction, and therefore the inclusive DIS process, does not depend on the transverse spin state of the struck quark.

A more rigorous argument is based on the fact that the transversity distributions are chiral-odd objects, involving an helicity-flip of the struck quark. The elementary photon-quark scattering conserves chirality, and cannot reveal

a chiral-odd object like $h_1(x)$. The measurement of $h_1(x)$ is possible only if combined with another chiral-odd distribution. In polarized SIDIS this distribution is a chiral-odd quark fragmentation function $D_q^H(z)$, describing the probability for a quark q to fragment into a certain hadron H carrying a fraction z of the quark momentum. However, semi-inclusive measurements, involving the detection of a particular hadron H in addition to the scattered lepton, are experimentally more difficult than inclusive ones. In the next section the Λ production case will be considered.

3.6 Λ production with a transversely polarized target

The polarization of Λ hyperons produced in the scattering of an unpolarized lepton off a transversely polarized nucleon can be expressed as[10]:

$$P_{0\uparrow}^H = \hat{D}_{NN}(y) \frac{\sum_i e_{q_i}^2 \Delta_T q_i(x) \Delta_T D_{q_i}^H(z)}{\sum_i e_{q_i}^2 q_i(x) D_{q_i}^H(z)}. \quad (3.39)$$

The factor $\hat{D}_{NN}(y)$ takes into account the spin dependence of the elementary lepton-quark interaction:

$$\hat{D}_{NN}(y) = \frac{d\hat{\sigma}^{lq^\uparrow \rightarrow lq^\uparrow} - d\hat{\sigma}^{lq^\uparrow \rightarrow lq^\downarrow}}{d\hat{\sigma}^{lq^\uparrow \rightarrow lq^\uparrow} + d\hat{\sigma}^{lq^\uparrow \rightarrow lq^\downarrow}} = \frac{2(1-y)}{1+(1-y)^2}. \quad (3.40)$$

In Eq.3.39 we have introduced the transversity fragmentation functions $\Delta_T D_{q_i}^H(z)$ defined as

$$\Delta_T D_{q_i}^H(z) = D_{q_i^\uparrow}^{H^\uparrow}(z) - D_{q_i^\uparrow}^{H^\downarrow}(z), \quad (3.41)$$

where $D_{q_i^\uparrow}^{H^\uparrow(\downarrow)}(z)$, which represent the probability for a quark q_i of fragmenting into a Λ hyperon carrying a fraction z of the quark momentum and with the spin correlated (anti-correlated) with the transverse spin of the quark.

In the polarized SIDIS, the quantities $\Delta_T q(x)$ can only be measured together with the transversity fragmentation functions $\Delta_T D_q^H(z)$. For this reason, additional informations on the $\Delta_T D_q^H(z)$ functions are needed in order to isolate

the $\Delta_T q(x)$ distribution functions. Such information can be extracted from the data coming from the $pp^\uparrow \rightarrow \Lambda^\uparrow X$ reaction[11], and the combined measurements of pp^\uparrow and μN interactions can lead to the $\Delta_T q(x)$ measurement.

Chapter 4

The GRID

4.1 GRID Computing

Grid computing is a kind of distributed computing that involves coordinating and sharing computational power, applications, data, storage, or network resources across dynamic and heterogeneous organizations. Grid technologies promise to change the way organizations tackle complex computational problems. Since sharing such a huge amount of resources has never been done yet, this is an evolving area of computing, where standards and technologies are still being developed.

Upgrading and purchasing new hardware is a costly proposition, and with the rate of technology obsolescence, it is eventually a losing one.

The great innovation is to use the existing resources in a new way, never used before, being able to scale up to limits never reached. Before speaking about European and Italian GRID infrastructure it is better to give a table of acronyms that will be used in next sections, which can be found in appendix A.

4.2 LHC Computing Grid - LCG-2

The purpose of the LHC computing Grid Project (LCG) [19] is to deploy the computing infrastructure for the simulation, processing and analysis of data

for the LHC collaborations: ALICE, ATLAS, CMS and LHCb. This includes both the common infrastructure of libraries, tools and frameworks required to support the physics application software, and the development and deployment of the computing services needed to store and process data, providing batch and interactive facilities for the community of physics involved in LHC.

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. Since it is not considered feasible to fund all of the resources at one site, it has been agreed that the LCG computing service will be implemented as a geographically distributed Computational Data Grid. This means that the LCG computing service will use computing and storage resources, installed at a large number of Regional Computing Centres in many different countries, interconnected by fast networks. Particular software, referred to generically as Grid Middleware, will hide much of the complexity of this environment from the user, giving the impression that all of these resources are available in a coherent virtual computing centre.

In the first phase of the project, 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. LCG-2 is the current release of LCG. This version runs in 2004 with the main goal of providing a stable service. LCG-2 expands the services of LCG-1, with enough resources and functionality for the 2004 Computing Data Challenges. In addition, more Tier 1 and Tier 2 centres will join the project, following the Monarc model, as in the previous LCG-1 release.

The core sites deploying the new release are CERN, Karlsruhe, Barcelona, FNAL, CNAF, NIKHEF, Taipei and RAL. In Fig.4.1 is shown a map of LCG sites. LCG-2 is organized into Virtual Organizations: dynamic collections of individuals and institutions sharing resources in a flexible, secure and coordinated manner. In this environment, the user encounters unique authentication, authorization, resource access, resource discovery, and other features.



Figure 4.1: Map of LHC Computing Grid sites

The current installed software release is LCG2 2.2.0

4.3 INFN Production GRID

The INFN-Grid project [18] started in 1999 to develop the first Italian Grid. It is now a reality, integrated with various Grid infrastructures all over the world. The INFN Grid comprises more than 20 sites among 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 (bio-medicine, earth observation, etc.) and to in-

dustry. It is a successful example of collaboration between physicists, software engineers, computer professionals, computer scientists and Italian industries. With a grant received from the MIUR-FIRB funds (government funds for investment in fundamental research), for the project called Grid.it, INFN with other National Research Institutions start the development of a common grid infrastructure supporting the Italian Research Area (IRA).

In collaboration with CERN, various European countries and some industries, INFN has launched in 2001 the largest European grid project, DataGrid, a milestone towards an infrastructure supporting the common European Research Area (ERA). With the same partners it has furthermore promoted the DataTAG project, which has provided interoperability solutions for the development of a World Grid for Science with US and Asian-Pacific areas.

This work has been further extended by the CERN LHC Computing Grid Project (LCG), where INFN is one of the main contributing partners, whose goal is to deploy a worldwide computational grid service, integrating the capacity of scientific computing centres spread across Europe, America and Asia to enable the analysis of the large amount of data produced by the new accelerator starting in 2007.

With the CERN coordination and the collaboration of other institutes from almost all European countries, INFN has promoted the new EGEE (Enabling Grids for E-science in Europe) project that has been approved and has started on April 2004. The primary goal of this project is to create a seamless European Grid infrastructure for the support of the European Research Area.

In Fig.4.3 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 currently installed release is INFN Production Grid 2.2.0. This release and the previous one are superset of LCG ones.



Figure 4.2: Map of INFN Production Grid and Grid.it

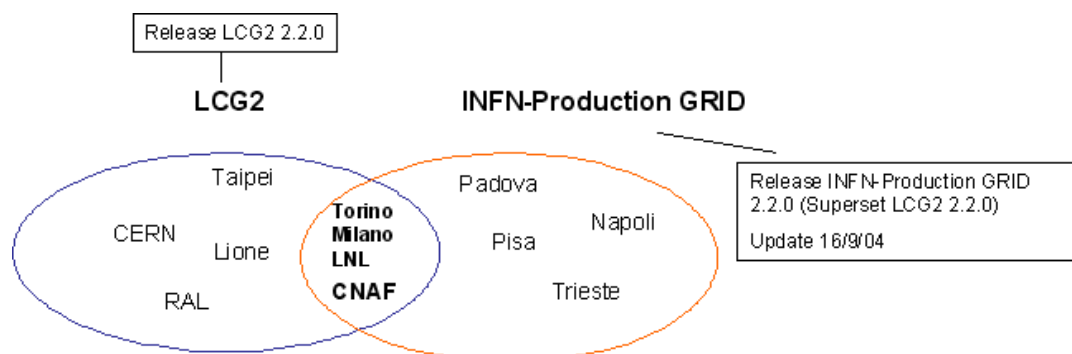


Figure 4.3: Relation between INFN Production Grid and LCG sites

4.4 The Grid Architecture (LCG-2 and INFN Production Grid)

The user, through a user interface (UI), gains access to computing resources (Computing Element (CE)) and data resources (Storage Element (SE)).

On the grid the user has access to different kinds of computing and data services which are connected via the Internet [17]. The user knows which services are offered and which services may be used, via Information Services (IS). A grid infrastructure needs to know where resources are and their states.

In the EDG DataGrid, information services for hardware resources (the EDG Information Service) and data resources (the EDG Replica Catalog) were provided. The EDG Information service also provides information about the grid services deployed. Currently, EDG provides only limited support for user service discovery. Individual users may however use their own services for this purpose. Another resource for hardware information are the information providers which are installed on every CE/SE and provide local hardware information to the EDG Information service. On a Computing Element this service is only installed on the front-end node and provides information about all the WNs within the computing element. The EDG Information Service collects both static (e.g. hardware configuration) and dynamic (e.g. current load or available memory) information about the system.

The information services (Information Providers, Replica Catalog, Info Service) are the low level infrastructure upon which higher level services are based. How to select among the resources and how to access them is the task of a set of higher level Grid services available on EDG.

The job submission service allows the user to specify requirements for the jobs to be run on the grid. Based on the information provided by the user and on the current situation of the grid provided by the EDG information services the Resource Broker (RB) schedules the job on appropriate resources.

The replica management service helps the user in managing his data in an efficient way by providing means for replicating (copying) data to several

storage resources. In particular, data may be replicated to locations close to the available computation resources thereby reducing data access costs. EDG currently provides two packages for replica management: GDMP and the edg-replica-manager, an improved version of the globus replica manager. Moreover, mass storage systems such as Castor or HPSS are supported.

The systems and services provided by EDG are logically clustered into 6 hosts types, each providing a particular service:

- 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 Information Service runs the EDG information services
- Replica Catalog hosts the data information service.
- The Storage Elements provide storage resources on the Grid
- The Computing Elements supply CPU power. CEs are typically clusters of Worker Nodes and a front-end node that manages the WNs via a Local Resource Management System like PBS, LFS, etc..
- The Resource Broker is the core of all activities on the Grid. Specifically, it schedules user jobs on appropriate CEs.

In a physical EDG Testbed the GRID Element needs to be mapped onto physical host. The information service, replica catalog, resource broker, and user interface needs to be set up on a host. If enough machines are available, each of these logical host may be mapped to a separate physical one as well.

4.5 GRID Elements

4.5.1 User Interface

The initial point of access to the Grid is the User Interface (UI), 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. This is the component that allows users to access the functionalities offered by the Information, Workload and Data management services. It provides a Command Line Interface (CLI) to perform some basic grid operations:

- 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.

4.5.2 Computing Element

A Computing Element (CE) is defined as a Grid batch queue and it is identified by a pair <hostname>/<batch queue name>. A Computing Element is a typical farm of computing nodes called Worker Nodes (WNs) and a node acting as a Grid Gate (GG) or front-end to the Grid. The GG runs a Globus gatekeeper, the Globus GRAM (Globus Resource Allocation Manager), the master server of a Local Resource Management System (LRMS), together with the EDG Logging and Bookkeeping services.

In LCG-2 and INFN Production Grid the types of LRMS supported are PBS, LSF and Condor. While a CE must have both out-bound and in-bound connectivities, the WNs can use private addressing with only out-bound connectivity. The GG provides a uniform interface to the computational resources it manages. On the WNs, all commands and Application Programming Interface (API) for performing actions on Grid resources and Grid data are available. Each GRID site runs at least one CE and a farm of WNs behind it.

4.5.3 Storage Element

A Storage Element (SE) provides uniform access and services to large storage spaces. The Storage Element may control large disk arrays, mass storage systems (MSS) and the like. The current LCG-2 and INFN GRID releases include a SE, which has a GridFTP server as the data transfer protocol to the storage resource. It is responsible for secure, fast and efficient file transfer to/from the Storage Elements. As MSS, LCG-2 will support disk pool (with GridFTP and rfiio as transfer protocols), tape archiving systems (with GridFTP and rfiio) and nstore (with GridFTP). Each site provides one or more SEs.

4.5.4 Information System

The resources described up to now constitute the computing and storage power of the Grid. Together with that infrastructure, additional services are provided 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. These are the Information System and the Data Management services.

The Information System (IS) provides information about the Grid resources and their status. In LCG-2 and INFN Production Grid, the Monitoring and Discovery Service (MDS) from Globus has been adopted as the provider of this service. Information is propagated in a hierarchy way.

Computing and storage resources at each site report (via the Grid Resource Information Servers, or GRISes) their static and dynamic status to the Site Grid Index Information Server (GIIS). The site GIISes register with one or more regional GIISes, for redundancy reasons and to nicely divide the administration domains. This structure is transparent to the user.

Due to dynamic nature of the GRID, the GIISes might not contain information about resources that are actually available on the Grid but that, for some reasons, are unable to publish updated information to the GIISes. Because of this, the Berkeley DB Information Index (BDII) was introduced. The BDII

queries the regional GIISes and acts as a cache storing information about the Grid status in its database. Every time a resource appears in one of the GIISes, its existence is registered in one of the BDII. There is one BDII running at each site where a Resource Broker (see later) is installed. Users and other Grid services (such as the RB) can query BDII to get information about the Grid status. Very up-to-date information can be found by directly interrogating the site GIISes or the local GRISes that run on the specific resources.

4.5.5 Data Management

The Data Management services are provided by the Replica Management System (RMS) of the European Data-Grid (EDG). In a Grid environment, the data files are replicated, possibly on a temporary basis, to many different sites depending on where the data is needed. The users or applications do not need to know where the data is located. They use logical names for the files and the Data Management services are responsible for locating and accessing the data. The files in the Grid are referenced by different names: Grid Unique Identifier (GUID), Logical File Name (LFN), Storage URL (SURL) and Transport URL (TURL). While the GUID or LFN refer to files and not replicas, and say nothing about locations, the SURLs and TURLs give information about where a physical replica is located.

A file can always be identified by its GUID; this is assigned at data registration time and is based on the UUID standard to guarantee unique IDs. A GUID is of the form: `guid:<unique string>`. All the replicas of a file will share the same GUID. In order to locate a Grid accessible file, the human user will normally use a LFN. LFNs are usually more intuitive, human-readable strings, since they are allocated by the user as GUID aliases. Their form is: `lfn;jany aliasj`. The SURL is used by the RMS to find where a replica is physically stored, and by the SE to locate it. Currently, the SURLs are of the form: `sfn:<SE hostname>/<local string>`, where `<local string>` is used internally by the SE to locate the file.

Finally, the TURL gives the necessary information to retrieve a physical

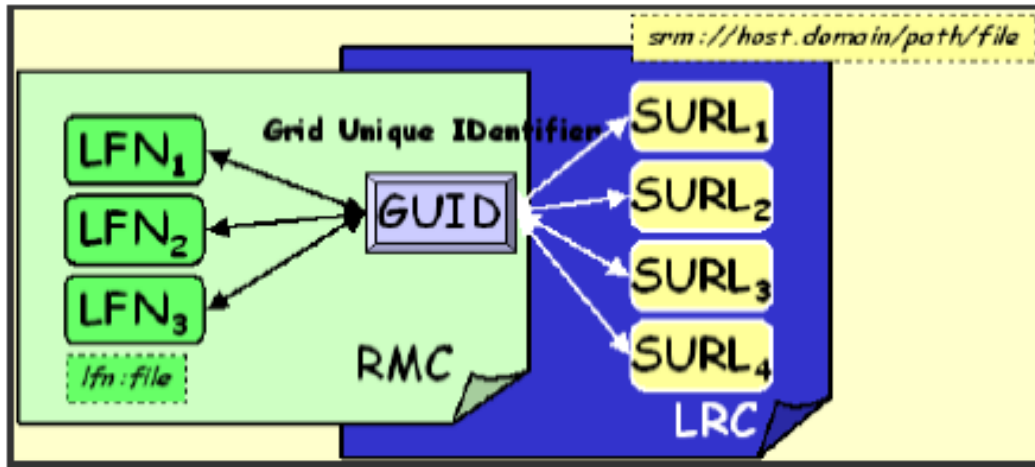


Figure 4.4: Relation between LFN and FFN

replica, including hostname, path, protocol and port (as any conventional URL); so that the application can open and retrieve it. Fig.4.4 shows the relation between the different file names.

The main services offered by the RMS are: the Replica Location Service (RLS) and the Replica Metadata Catalog (RMC). The RLS maintains information about the physical location of the replicas (mapping with the GUIDs). It is composed of Local Replica Catalogs (LRCs) which hold the information of replicas for a single VO. The RMC stores the mapping between GUIDs and the respective aliases (LFNs) associated with them, and maintains other metadata information (sizes, dates, ownerships) The last component of the Data Management framework is the Replica Manager.

The Replica Manager presents a single interface for the RMS to the user, and interacts with the other services. In the LCG-2, this interface is integrated with the User Interface described earlier.

When SRMs are already working, files stored there will use srm as the prefix for their SURLs, instead of sfn. This distinction will allow the RMS to distinguish which kind of storage the file is on.

For the moment these catalogues are centralized and there is one RLS (with

its LRC and RMC) per VO. In the first phase, all RLSs are run at CERN.

4.5.6 Job Management

The services of the Workload Management System (WMS) is responsible for the acceptance of job submits and the dispatching of those jobs to the appropriate CE, depending on the job requirements and the available resources. For that purpose, it must retrieve information from the BDII, and the RLS. The Resource Broker (RB) is the machine where the services of the WMS run. These services are:

- Network Server (NS), which accepts the incoming job requests from the UI, and provides support for the job control functionality.
- Workload Manager, which is the core component of the system.
- Match-Maker (also called Resource Broker), whose task is finding the best resource matching the requirements of a job (match-making process).
- Job Adapter, which prepares the environment for the job and its final description, before passing it to the Job Control Service.
- Job Control Service (JCS), which finally performs the actual job management operations (job submission, removal. . .)

In addition, the Logging and Bookkeeping service (LB) is usually also run on a RB machine. The LB logs all job management Grid events, which can then be retrieved by users or system administrators for monitoring or troubleshooting. Multiple RBs are available in LCG-2 and INFN Production Grid. Participating sites are free to install their own RBs.

The last component of the LCG-2 Grid described here is the Proxy Server (PS). When a user accesses the Grid, he/she is provided with a temporary certificate, called proxy, that has an expiration time. If the user proxy expires before the user job has finished, all subsequent requests for service will fail due to unauthorized access. In order to avoid this, the Workload Management

Service provided by EDG allows for proxy renewal before the expiration time has been reached if the job requires it. The PS is the component that allows such functionality.

4.6 Service Interactions and Job Flow

This section describes briefly what happens when a user submits a job to the LCG-2 Grid to process some data and how the different components interact, also give a description of the components of the Data Management system.

4.6.1 Job Submission

- a. After obtaining a digital certificate x.509 from one of the trusted Certification Authorities, registering with a Virtual Organization and obtaining an account on an User Interface (once only actions), the user is ready to use the Grid. He logs to the UI and creates a proxy certificate that authenticates him in every further secure interaction, and has a limited lifetime (typically 12h).
- b. The user submits the job from the UI to the WMS, where the job will be executed on a CE. The user can specify in the job description file one or more files to be copied from the UI to the CE; this set of files is called Input Sandbox. The event is logged in the LB and the status of the job is “SUBMITTED”.
- c. The WMS, and in particular the Match-Maker component, looks for the best available CE to execute the job. To do so, the Match-Maker queries the BDII to get the status of computational and storage resources and the RLS to find location of data. The event is logged in the LB and the status of the job is “WAIT”.
- d. The WMS Job Adapter prepares the job for submission creating a wrapper script that is passed, together with other parameters, to the JCS

for submission to the selected CE. The event is logged in the LB and the status of the job is “READY”.

- e. The Globus Gatekeeper on the CE receives the request and sends the Job for execution to the LRMS (e.g.PBS, LSF or Condor). The event is logged in the LB and the status of the job is “SCHEDULED”.
- f. The LRMS handles the job execution on the available local farm worker nodes. Users files are copied from the RB to the WN where the job is executed. The event is logged in the LB and the status of the job is RUNNING.
- g. While the job runs, Grid files can be accessed on the (close) SE using either the rfiio protocol or local access if the files are copied to the WN local filesystem. In order for the job to find out which is the close SE, or what is the result of the Match-Maker process, a file with this information is produced by the WMS and shipped together with the job to the WN. This is known as the BrokerInfo file. Information can be retrieved from this file using the BrokerInfo CLI or the API library.
- h. The job can produce new output data that can be uploaded to the Grid and made available for other Grid users to use. This can be achieved using the Data Management tools described later. Uploading a file to the Grid means copying it on a Storage Element and registering its location, metadata and attribute to the RMS. At the same time, during job execution or from the User Interface, data files can be replicated between two SEs using again the Data Management tools.
- i. If the job reaches the end without errors, the output (not large data files, but just small output files specified by the user in the so called Output Sandbox) is transferred back to the RB node. The event is logged in the LB and the status of the job is “DONE”.
- j. The user can retrieve the output of his job from the UI using the WMS CLI or API. The event is logged in the LB and the status of the job is

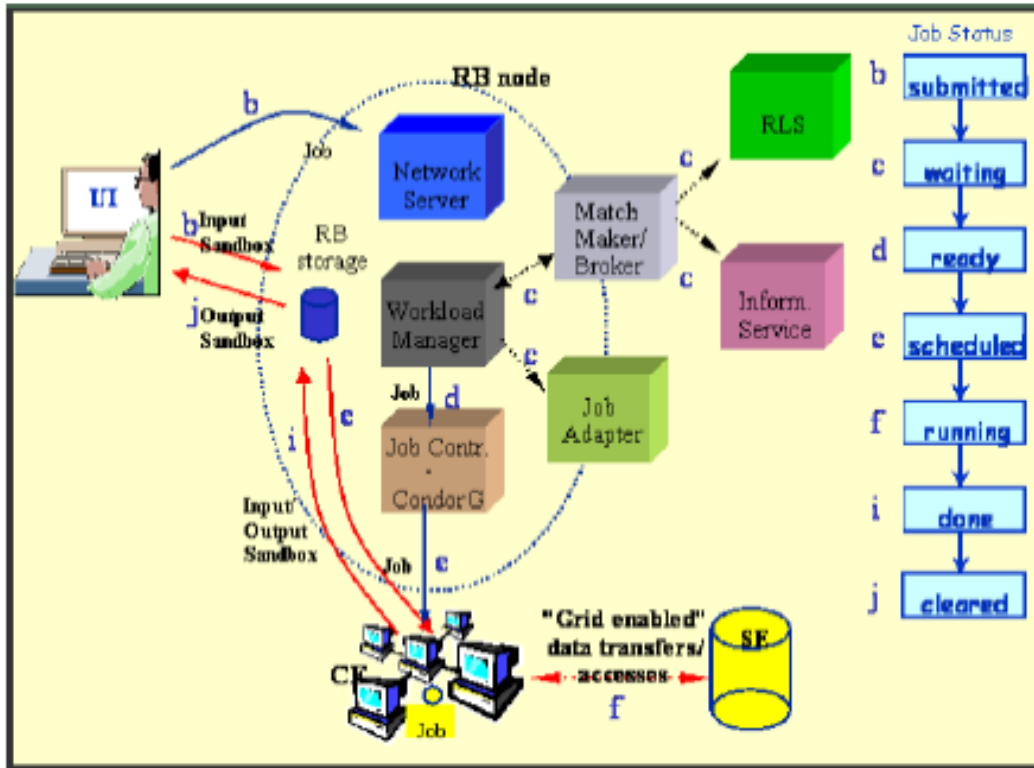


Figure 4.5: Graphical view of job status

“CLEARED”.

- k. Queries of the job status are addressed to the LB database from the UI machine. Also, from the UI is it possible to query the BDII for a status of the resources.
- l. If the site where the job is being run falls down, the job will be automatically resent to another CE that is similar to the previous one, with match the same requirements the user asked for. In the case that this new submission is disabled, the job will be marked as “ABORTED”. Users can get information about the scenario by simply questioning the LB service.

Fig.4.5 shows what has been described in previous steps.

- SUBMITTED - the user has submitted the job via UI
- WAITING - the RB has received the job
- READY - A CE, which matches job requirements, has been selected, and the job is transferred to the JSS
- SCHEDULED - the JSS has sent the job to the CE
- RUNNING - the job is running on the CE
- DONE - this state has different meanings:
 - DONE(ok) : the execution has terminated on the CE (WN) with success
 - DONE(failure) : the execution has terminated on the CE (WN) with some problems
 - DONE(cancelled) : the job has been cancelled with success
- OUTPUTREADY - the output sandbox is ready to be retrieved by the user The state OUTPUTREADY reflects the time difference between end of computation on CE and the moment RB got necessary notification via Condor and JSS. (job is terminated)
- CLEARED - the user has retrieved all output files successfully, and the job bookkeeping information is purged some time after the job enters in this state.
- ABORTED - the job has failed The job may fail for several reasons one of them is external to its execution (no resource found).

4.6.2 Data Management

The Input/Output Sandbox is a mechanism for transferring small data files needed to start the job or to check the final status over the Grid. Large data files are available on the Grid and known to other users only if they are stored

on SEs and registered in the RMS catalogues. In order to optimize data access and to introduce fault-tolerance and redundancy, data files can be replicated on the Grid.

The EDG Replica Manager, the Replica Location Service and the Replica Metadata Catalog are the tools available for performing these tasks. Only anonymous access to the data catalogues is supported: the user proxy certificate is not used to control the access to them. In the LCG-2, a file is identified uniquely by the GUID, but the user may refer to it using different aliases. Also, there will be probably several physical replicas of each file. The user should never interact with the RMC or the RLS catalogs directly. Instead, he should always use the EDG RM.

When a new file is produced, the file should be registered to the Grid to be known and usable by Grid services or other Grid users. This can be done using the EDG Replica Manager commands for copying and registering a file. Before running a job on the Grid, the user can ask the WMS to run the job on a CE close to an SE containing the data of interest, or, at run time, the job can ask the RMS to replicate a file on a SE close or even on the WN where the job is running. If a file is no longer needed, it can be deleted from the Grid and all its references removed from the data catalogues.

4.6.3 Information System

The architecture of the Information System in the LCG-2 Grid has been already described. Users can query the IS to retrieve static or dynamic information about the status of the LCG-2 infrastructure. In order to have an optimal answer, users are encouraged to query the BDIIs or the site GIISes, but not the regional GIISes. Also, the specific GRISes can be queried. The IS is based on OpenLDAP, an open source implementation of the Lightweight Directory Access Protocol (LDAP).

LDAP is a protocol that provides the access for a directory service. A directory service is a specialized database optimized for reading, browsing and searching information. No transaction or roll-back features are normally offered.

In particular in LCG-2 Grid, only anonymous access to the catalogue is offered. This means that all users can browse the catalogues and all services are allowed to enter information into it.

The LDAP information model is based on entries. An entry is a collection of attributes which form together a globally unique Distinguished Name (DN), a name that uniquely identifies the entries. Each of the entries attributes has a type and one or more values. The types are typically mnemonic strings, while the syntax of the values depends on the attribute type. An LDAP schema describes the attributes and the types of the attributes associated with entries.

The LCG-2 Grid deploys the GLUE (Grid Laboratory for a Uniform Environment) Schema for information description. The GLUE Schema activity aims to define a common conceptual data model to be used for grid resources monitoring and discovery. There are three main components of the GLUE Schema. They describe the attributes and value of Computing Elements, Storage Elements and binding information for Computing and Storage Elements.

4.7 Entering the GRID

This section describes the preliminary steps to gain access to the LCG-2 and INFN Production GRID.

4.7.1 Obtainig a Certificate

Before GRID resources can be used, a user is required to register some personal data and information about the Virtual Organization he belongs to with the GRID Registration Server, at CERN for LCG-2, and CNAF for INFN Production Grid.

The Grid Security Infrastructure (GSI) in LCG-2 enables secure authentication and communication over an open network. GSI is based on public key encryption, X.509 certificates, and the Secure Sockets Layer (SSL) communication protocol. The first requirement the user must fulfil is to have a valid X.509 certificate issued by a trusted Certification Authority (CA). The role of a CA is

to guarantee that a user is who he claims to be and is entitled to own his certificate. In general CAs are organized geographically and by research institute (e.g. INFNCA for INFN). Each CA has its own procedure to issue certificates. An important feature of a certificate is the subject, a string: containing information about the user, that identifies the owner.

4.7.2 Registering With LCG-2 - INFN Production Grid

Before using any GRID service, a registration of some personal data with the registration server(hosted at CERN for LCG-2 and at CNAF for INFN Production Grid) plus some additional steps are required.

4.7.3 VOs - Virtual Organizations

A second requirement for the user is to belong to at least a Virtual Organization (VO). A VO is an entity, which corresponds typically to a particular organization or group of people in the real world. The membership of a VO grants specific privileges to the user. For example, a user belonging to the ALICE VO will be able to read the ALICE files or to use resources reserved to the ALICE collaboration.

In LCG there are five possible Virtual Organizations (VOs) a user could be affiliated to: one for the DTeam (LCG Grid Deployment Group) for administration purpose and one for each of the four LCG HEP experiments.

In the INFN Production Grid there are much more VOs, the same of LCG plus the VOs for other INFN Physics Experiment (also COMPASS in the future) and also the VOs for the experiment of other italian research institution as INFN, INFM, INGV, ecc.) Entering the VO of an experiment usually requires being a member of the collaboration; the user must comply with the rules of the VO relevant to him to gain membership. Of course, it is also possible to be expelled from a VO when the user fails to comply with these rules.

A Virtual Organization Server maps users certificates to users data and lists the certificates as belonging to users that are part of a VO or HEP experiment.

The VO Server for the DTeam was and still is run at CERN, while the VO Servers for the four experiments are run at NIKHEF and CNAF for INFN Production Grid VOs. But LCG-2 can support many more VOs. Each site installing the new release is free to support any VOs. Users can be aware of the VOs supported at a given site by asking directly to that site.

A user is authorized to use Grid resources by means of the grid-mapfile mechanism. Each host part of the Grid has a local grid-mapfile which maps user certificates subject to local pool accounts. When a user request-for-service reaches a host, the certificate of the user is checked in the local grid-mapfile. If the user certificate subject is found there, then the local account to which the user certificate is mapped is used to serve the request.

It is not possible to access the Grid without being member of any VO. Every user is required to select his VO when registering with LCG-2 and INFN-GRID and the supplied information is forwarded to the VO administration and resource providers for validation before the registration process is completed. However, it is possible to belong to more than one VO at the same time. In that case, the user must choose VOs, at proxy creation time, and cannot exploit the advantage of being in two VOs at the same time.

4.7.4 Setting up the User Account

To access the Grid, the user must also have an account on a User Interface. To obtain such an account, a local system administrator must be contacted.

As an alternative, the user can install the UI software on his machine. Once the account has been created, the user certificate and key must be copied in `.globus` directory.

4.8 JDL - Job Description Language

In LCG-2 and INFN-GRID, job description files (`.jdl` files) are used to describe jobs for execution on Grid. These files are written using a Job Description Language (JDL).

VOs	LCG	INFN Prod. GRID and GRIDIT
Dteam (LCG Deployment Team)	X	only Tier 1-2 sites
ALICE (LHC Experiment)	X	only Tier 1-2 sites
ATLAS (LHC Experiment)	X	only Tier 1-2 sites
LHCb (LHC Experiment)	X	only Tier 1-2 sites
CMS (LHC Experiment)	X	only Tier 1-2 sites
INFN-GRID (INFN-GRID Project)		X
BABAR		X
BIO (Grid.it Biology Group)		X
CDF		X
compchem (Chemistry Dep. Univ. of Perugia)		X
ENEA		X
GRIDIT (General Grid.it Project VO)		X
INAF		X
INGV (Geophysics and Vulcanology)		X
Theophys (INFN Theoretical group)		X
virgo (INFN Virgo Experiment)		X
zeus		X

Table 4.1: List of supported VOs for LCG and INFN Prod. GRID - GRIDIT.

JDL is based upon Condor's Classified Advertisement Language (CLASSAD) which is a simple expression-based language to specify both, resources and requests. Classad facilitates a matching between resources and customers. Classad describes properties by means of attribute = value pairs.

The JDL adopted within the Grid is the Classified Advertisement (ClassAd) language defined by the Condor Project, which deals with the management of distributed computing environments, and whose central construct is the ClassAd, a record-like structure composed of a finite number of distinct attribute names mapped to expressions. A ClassAd is an highly flexible and extensible data model that can be used to represent arbitrary services and constraints on their allocation. The JDL is used to specify the desired job characteristics and constraints, which are used in by match-making process to select the resources that the job will use.

4.8.1 JDL example

Below is reported an example of a JDL file:

```
Executable = "gridTest";
StdError = "stderr.log";
StdOutput = "stdout.log";
InputSandbox = {"home/joda/test/gridTest"};
OutputSandbox = {"stderr.log", "stdout.log"};
InputData = "LF:testbed0-00019";
ReplicaCatalog = "ldap://sunlab2g.cnaf.infn.it:2010/lc=test,
                  rc=WP2 INFN Test, dc=infn, dc=it";
DataAccessProtocol = "gridftp";
Requirements = other.Architecture=="INTEL" && \
               other.OpSys=="LINUX" && other.FreeCpus >=4;
Rank = "other.MaxCpuTime";
```

To submit a job with this .jdl file the user will issue `edg-job-submit` command with this syntax:

```
edg-job-submit [-r <res_id>] [-c <config file>]
               [-o <output file>] <job.jdl>
```

- with `-r` the job is submitted by the RB directly to the computing element identified by `-res_id-`
- `-c` the configuration file `-config file-` is pointed by the UI instead of the standard configuration file
- `-o` the generated `dg_jobId` is written in the `-output file-`

Useful for other commands, the `-i` option (input file or `dg_jobId`) display the status information about `dg_jobId` contained in the input file.

WMS provides apart from the `dg` job submission command a set of command line tools that allow to "manage" jobs already submitted:

- `edg-job-list-match` command returns the list of resources fulfilling job requirements;
- `edg-job-cancel` command cancels one or more submitted jobs;
- `edg-job-status` command displays bookkeeping information about submitted jobs. Used without options, followed directly by `dg_jobId`, only displays bookkeeping information regarding that particular `jobId`;
- `edg-job-get-output` command requests the RB for the job output files (specified by the `OutputSandbox` attribute of the JDL) and stores them in the UI local disk;
- `edg-job-get-logging-info` displays logging information about submitted jobs (all the events pushed by the various components of the WMS), very usefull for debugging;
- `edg-job-id-info` command just parses the `dg_jobId` string and displays formatted information contained in the job identifier.

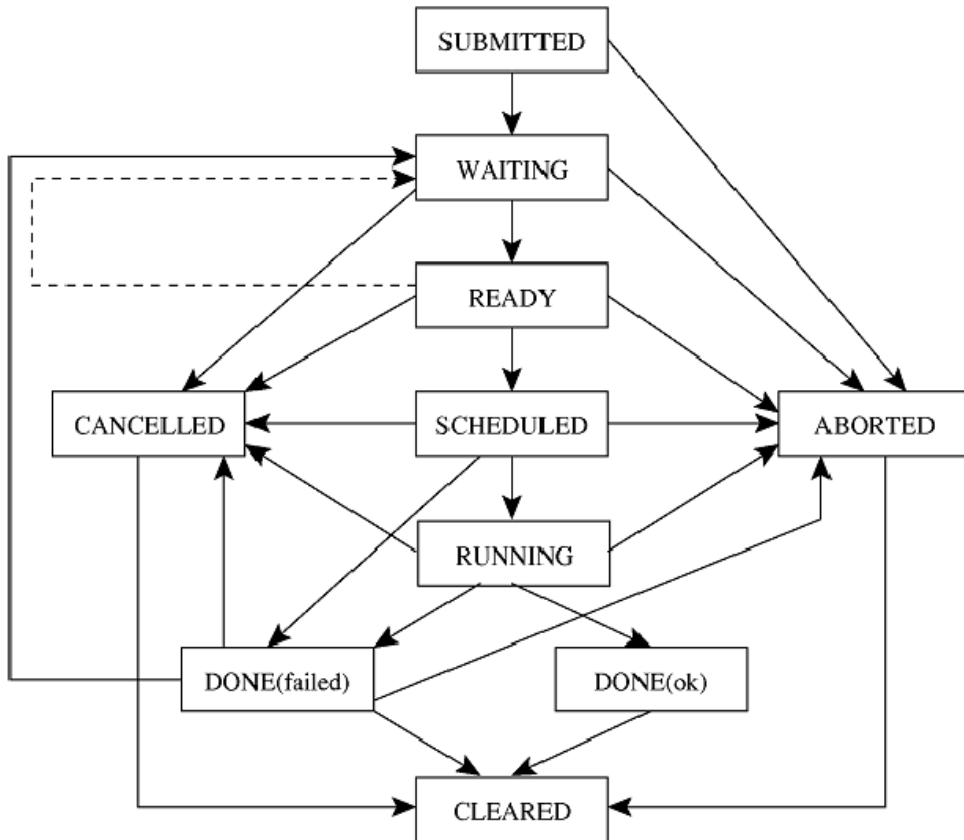


Figure 4.6: Possibility job states in the LCG2 and INFN Prod. GRID

4.8.2 Job Status Definition

As already mentioned a job can find itself in one of several possible states, the definition of which is given in Tab.4.2, in Fig.4.6 is shown a block diagram with all job status possibilities.

Status	Description of job status
SUBMITTED	submitted by the user via UI but not yet processed by the Network Server
WAITING	RB has received the job, accepted by the Network Server but not yet processed by the Workload Manager
READY	assigned to a Computing Element, which matches job requirements, but not yet transferred to it
SCHEDULED	waiting in the Computing Elements queue
RUNNING	running on WN
DONE	finished with different meanings
DONE(ok)	execution terminated on the CE (WN) with success
DONE(failure)	execution terminated on the CE (WN) with some problems
DONE(cancelled)	the job has been cancelled with success
OUTPUTREADY	output sandbox is ready to be retrieved by the user
ABORTED	aborted by the WMS for several reasons (e.g. it was too long, or the proxy certificated expired, no resource found, etc.)
CLEARED	Output Sandbox transferred to the User Interface

Table 4.2: *Job definitions*

Chapter 5

The software packages

5.1 Introduction

The software packages used to perform simulation and analysis are:

- The event generator, based on LEPTO or PHYTIA, which is integrated in COMGEANT
- The package COMGEANT based on GEANT
- The CORAL reconstruction program to perform the simulation
- The PHAST analysis program to filter the events and produce mDST suitable for further analysis

5.2 The LEPTO event generator

LEPTO is a general Monte Carlo (MC) [13] program to simulate complete lepton-nucleon scattering events and integrate cross sections. The parton interaction is based on the standard model electroweak cross sections, which are fully implemented in leading order for any lepton of arbitrary polarization, and different parametrizations of parton density functions can be used. First order

QCD matrix elements for gluon radiation and boson-gluon fusion are implemented and higher order QCD radiation is treated using parton showers. The code is written completely in standard FORTRAN77.

5.3 The GEANT simulation

The COMPASS simulation program is based on an interface to GEANT 3.21 [14], initially developed for the WA89 experiment, and later adapted to different fixed target experiments. The modified version for COMPASS was called COMGEANT.

GEANT is a software package for the simulation of particle passage through the matter. This software was initially developed at CERN for the high energy physics.

In the COMPASS experiment, GEANT is used mainly 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.

The GEANT package performs the following tasks:

- description of the experimental setup as a structure of geometrical volumes. An identification number is assigned to each volume (*medium number*); each material present in the volumes is described by a set of parameters useful to trace the particle in it, e.g the atomic number, the atomic weight, the density.
- receiving as input the events produced by the Monte-Carlo event generator
- tracking the particle in steps through the setup volumes, taking care of the geometrical dimension of the volumes, the magnetic field and the physical interaction with the material.
- storing the particle trajectory and the detector response in dedicated data bank.
- displaying interactively the tracks and the detectors.

Initialization of all parameters, checking of the event processing and preparation of a output to be analyzed are implemented with calls to standard GEANT subroutines written in Fortran. The data used during the event generation and tracking are stored in dynamic structures called “banks”, whose name starts with J: e.g. the bank JPART contains all particle characteristics, JVOLUM all information on the geometrical volumes, and so on. The data communication through the various segments is ensured by the data structure itself and by a definition of a global variable, contained in common block loaded during the initialization phase. The different subroutines communicate among themselves through local variables and the variable defined in the common block.

5.4 COMGEANT

The body of the simulation program, COMGEANT, is resident in a public area of the collaboration computers. Every COMPASS member can access these files and make a local copy of it, for modifications, in fact the setup is not predefined in the code, but it is defined by an external file, whose format is FFREAD (Format Free READ), which allows any modification made by a user to not affect the official version. Therefore when the program runs, automatically it reads the local configuration.

The choice of different event generators can be done in the COMGEANT “run card”, named myomg.ffr. COMGEANT contains calls to the event generators like JETSET and PYTHIA, which provide the production of charmed particles or LEPTO, that generate DIS events. It is also possible produce a file of events with an external event generator, and the produced file can be used as a COMGEANT input. All the events usually are generated at the target point, therefore the COMGEANT purpose is to trace them through the setup. The processed events are stored into an NTUPLE or a ZEBRA file, both readable from the reconstruction program. ZEBRA is a Fortran library that permits

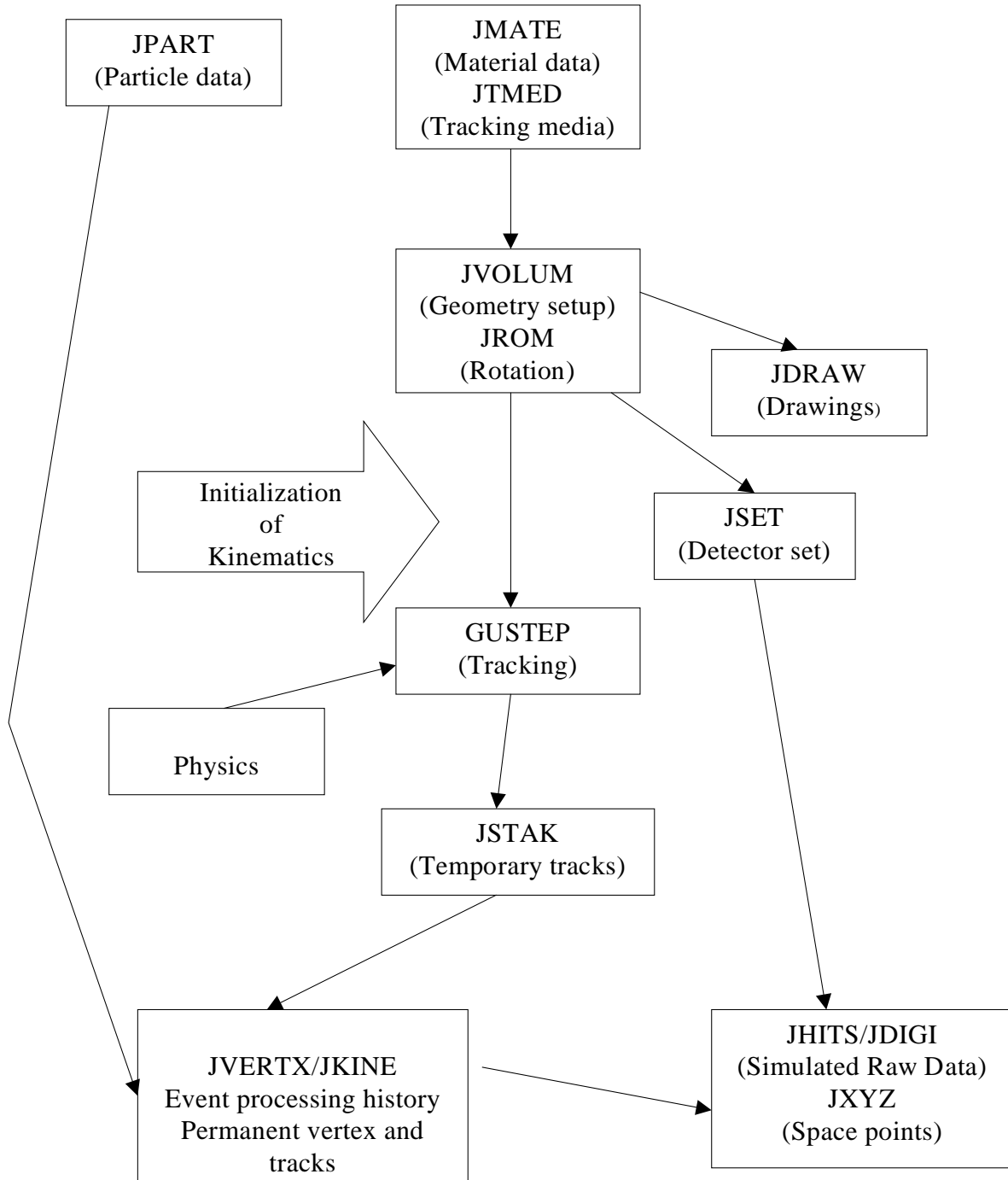


Figure 5.1: Schematic view of the GEANT structure

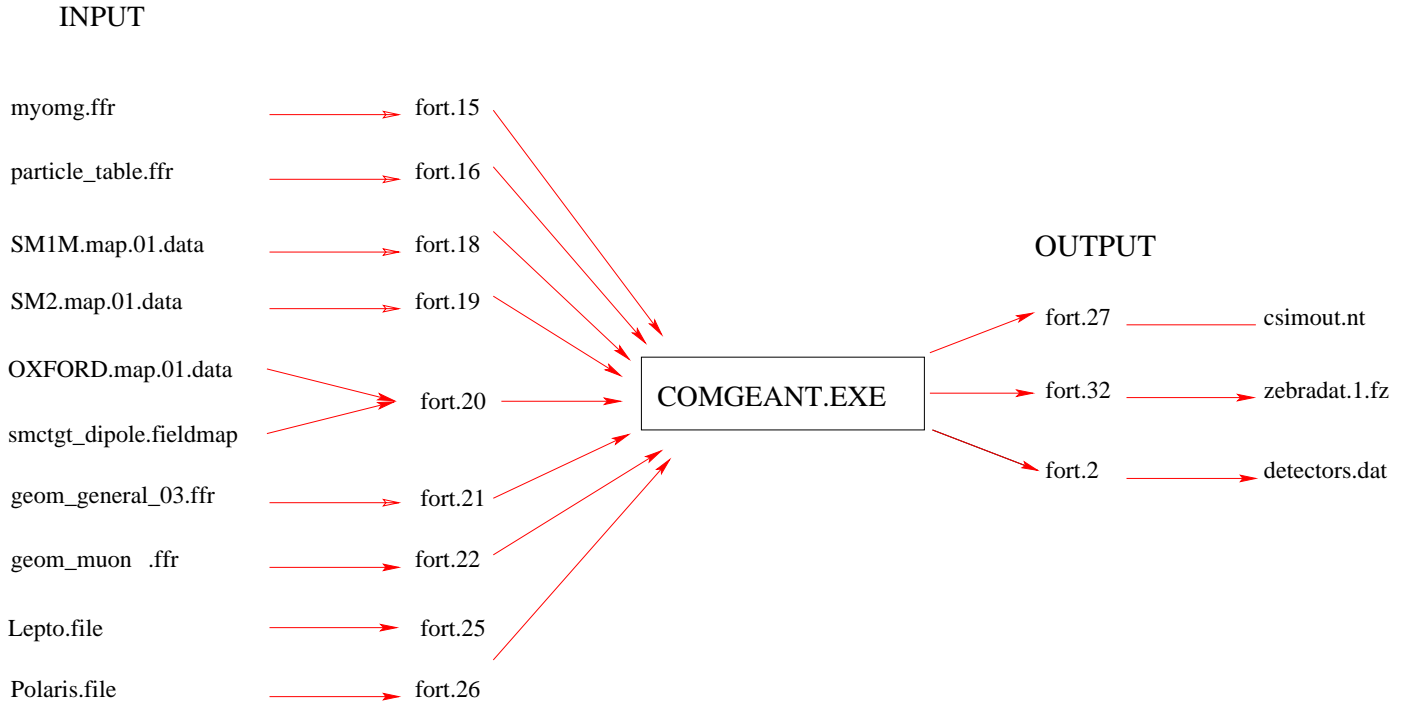


Figure 5.2: Schematic view of the COMGEANT input-output

saving and restoring the contents of a set of variables to and from a data file. Every connection to the geometrical setup description is defined by FFREAD cards.

COMGEANT received in input the files for the magnetic field description of SM1 and SM2 magnets, the file for the geometrical detectors description and the beam description. The user is allowed to modify the geometry files only. The program structure is illustrated in fig. 5.2.

All the input parameters are defined through the FFREAD cards and read in sequence as shown in figure 5.2. The main volume is called HALL and is defined in the main code. The HALL dimensions are read by a FFREAD file through the AREA definition, a COMGEANT entity. The HALL dimensions, id est the dimensions of the experimental hall in which the experiment takes place, are: $8000 \times 1000 \times 1000 \text{ cm}^3$ with the medium number equal to 1, corresponding to an experimental hall filled by air. The origin of the reference system of the

volume coincides with the center of the HALL, in which x-axis is parallel to the beam, y-axis goes from Saleve side to Jura side and z-axis along the direction down-up. Most of the geometrical definitions are independent from the choice of the axes direction, while this choice is important for the beam simulation. The choice of the reference system can be changed by an FFREAD card.

The materials, that fill the volumes, are identified in COMGEANT by different names. They can be simple or compounds. It is also possible to define a medium in which the particle propagates, e.g air or air with magnetic field.

To define a magnetic field it is necessary to previously define the volume occupied. The magnetic field maps are read from an external file. They contain the experimental measurements or calculations of the three components of the field on a grid. Therefore it is possible to interpolate on the value of the field any arbitrary point by a dedicated routine.

In COMGEANT two types of detectors are present: those who have an internal structure, for example a wire sensitive plane, and those who have no internal structure, which do not require additional description. COMGEANT identifies detectors by names, while GEANT identifies the detectors by meaning volumes.

A detector is described as a hierarchy of volumes that are filled with media. The physical properties of a volume are given by the definition of the medium. These properties include density, radiation length, transparency for magnetic field, sensitivity to the passage of ionizing particles, etc.

The coordinates of an hit, originated by the passage of a particle through the detector sensitive area, contribute to the form of one or two projections of the track itself. This measure is made in the detector reference system (DRF), in which the detector volume itself is defined.

The table of the particles interesting for the experiment is provided to COMGEANT by the file `particle_table.ffr`. The two and three body decays are simulated by GEANT routines, the four body decay is simulated by OMGEANT routine using a bidimensional random generator.

The track is produced at the vertex in the target point and propagated by

steps through the setup.

The output of COMGEANT is stored in a ZEBRA file which contains all the informations on the beam, on the GEANT tracks and the hits left on the sensitive detectors. Specific data on the detectors, e.g. the position in the setup, the wire direction, the pitch, the calorimeters structure etc, is provided by the file detectors.dat.

5.5 The reconstruction: CORAL

CORAL is the COMPASS Reconstruction and AnaLysis program [15] , it is written in C++ language using Object Oriented techniques and contains all the software necessary to reconstruct an event both from raw data and from simulated data.

As COMGEANT , the CORAL framework is piloted by an option file. This file contains the switches for the packages, the mapping file, the magnetic field maps and the parameters needed to select input. The output of the reconstruction is stored in a ROOT file. ROOT is a C++ Object Oriented library for data analysis developed at CERN in the context of the NA49 experiment. A ROOT file is organized as a UNIX file system to access a specific part of the object without reading all of the contents. This way provides a fast access to the interesting portion of data. The ROOT output, can contains histograms, showing distribution of some interesting variables and set of organized data; therefore sometimes it's used to store the informations event by event. This can be easily done in the ROOT tree, a flexible structure which can manage any kind of data such as object, vector, array or simpler data types. Both these features are used in the analysis.

An outstanding CORAL feature is the *modularity*. This allows for maximum flexibility in the choice of how an event is reconstructed, since the order of the modules can be changed easily or individual modules can be switched on and off, using an option file where a number of internal parameters can be set. The option file contains also the names of the input and output files, the output

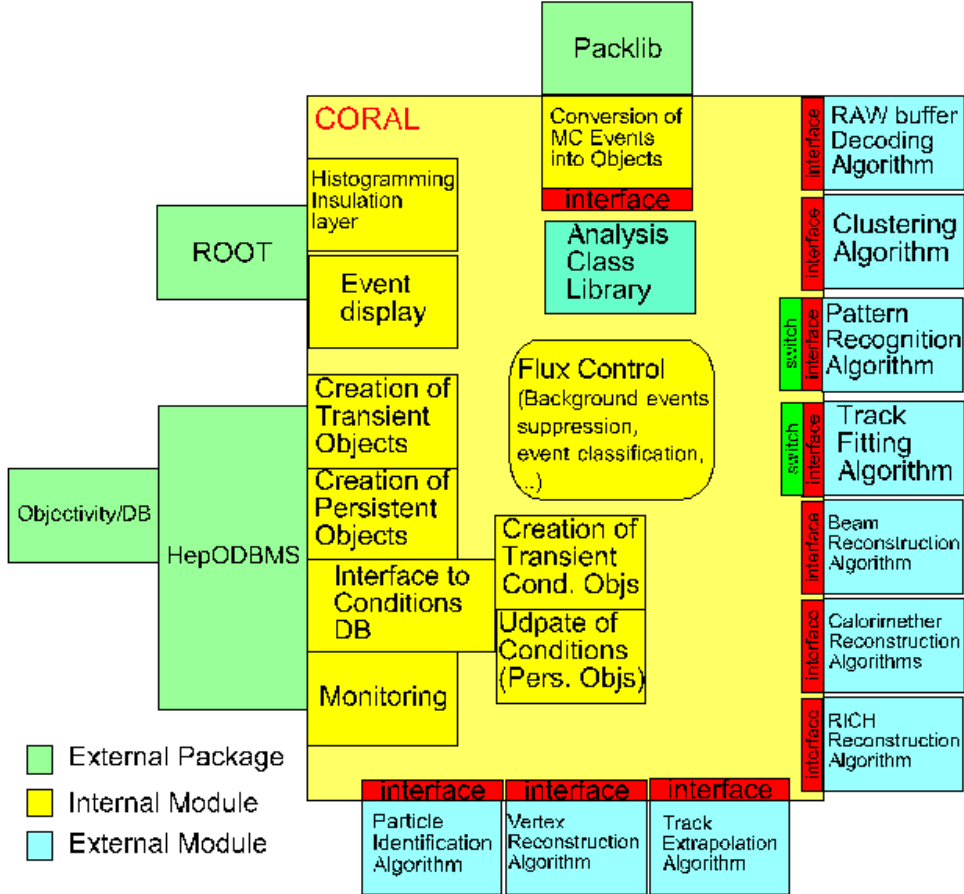


Figure 5.3: Schematic view of the CORAL structures

format and the address of the geometric description of the apparatus. This way to pass the geometrical description of the apparatus allow us to use the same code for different setups for example the muonic or hadronic one, both produced by COMGEANT.

5.6 PHAST

The software PHAST (PHysics Analysis Software Tools) [16] allows to filter the reconstructed events, and supplies the environment for the development of the analysis programs. In analysis PHAST is used twice. First PHAST

calls CORAL in order to reconstruct the traces discarding the events in which, for example, diffused traces of the muon is not reconstructed or the primary vertex is not defined. The output file is called miniDST (mDST) and it is reprocessed with PHAST, using a user defined code called User CARD. For example in the analysis of Λ seminclusive production by lepton scattering, the user card reconstruct the invariant mass of vertex V0, fills the Armenteros plot, calculates the value of Q transferred, and the values of y Biorken, x Feynman (x_f) and of the vertex azimuth angle ϕ . PHAST organizes the output information in multidimensional matrices. For Λ analysis PHAST produces two types of matrices: one for analyzing power A_n and one for the transverse polarization P_s .

Chapter 6

Software Installation for Lambda Transversal production

This section gives a short description of a typical GRID farm and a step by step description of the installation method used for COMPASS software on GRID environment. The listing of the script used to perform installation and test for COMPASS software is given in appendix A. The software used to perform the simulation for COMPASS experiment is based on three different programs as shown in chapter 4. Before using the software on the GRID it is necessary to install and test the software in a typical GRID environment. The software is installed on a server with a directory structure that reproduce faithfully the directory structure of a Computing Element. After the software compile and the test of the installation, I make the `.tar` file that will be downloaded during the installation procedure. This not gracefully method is necessary to get over the software compile difficulty, that is studied to be used in a “conventional” farm. The `.tar` file is stored on a web server (www.to.infn.it/grid-prod/software/). At this moment the COMPASS experiment has not, yet, a proper Virtual Organization (VO) so the software installation is performed using a particular ALICE VO user, mapped as `alicesgm`. The “SGM” (software grid manager) user is the only member of a VO that has write permission on a Computing Element (CE). The job submission is performed using a DTeam/Infngriid user account,

this user is mapped to “DTeam” on LCG-Tier2 resources and to “Infngrid” on INFN Production GRID, “GRIDIT only”, resources.

So it is possible to use a related run time environment variable:

```
GlueHostApplicationSoftwareRunTimeEnvironment: VO-alice-cmp-3.1.2
```

“alice” is for ALICE experiment, and “cmp” is for COMPASS. This variable will be published after installation over the GRID and will be required by the production job to select only the resources where the software is actually installed. The JDL file, (see par.1.8 chapter 4) is usually located on a UI and contains the system requirements and the scripts to be executed.

6.1 A typical farm

In this section a typical GRID farm is described taking as an example the one available in the INFN computing centre at the Institute of Physics of the Turin University. In Fig.6.1 is shown the GRID farm near to the COMPASS Experiment farm.

A minimal farm is composed by at least one of each GRID element:

- User Interface (UI)
- Storage Element (SE)
- Computing Element (CE)
- Worker Node (WN)

Tab.6.1 shows the relations between GRID elements and filesystem directories of the Turin site GRID farm. In the UI there are the user’s home directories where the single user can put submission and monitoring scripts and where the job output log file are stored; in the CE there is the experiment installation directory where SGM user can install the experiment software; in the SE there is the storage directory where job output can be saved (like the output simulation for COMPASS). Each VO has a different directory. The WN imports



Figure 6.1: *Typical GRID farm (INFN Turin Computing Centre)*

GRID Element	GRID Directory	Description
UI	/home	Users home directory
CE	/opt/exp_software/VO/	experiment installation software directory
SE	/flatfiles/SE00/VO	storage directory
WN	/opt/exp_software/VO/	imported from CE via nfs

Table 6.1: *GRID elements with exported and imported directories .*

the previously described directory via `nfs` from CE. The WN can save the jobs output on the SE via `gridftp` (GRID File Transfer Protocol).

The farm is composed by 20 one unit (1U) rack mountable servers, with two Xeon 2,4GHz CPUs, 1024MB RAM and 40GB of hard disk; the storage disk is a different 4U rack mountable server with two Xeon 2,4GHz CPUs, 2048MB of RAM, 40GB of system disk space and 2TB of storage disk space. The 2TB storage is composed by two 1TB raid 5 array joined via LVM (Large Volume Management).

6.2 Software installation

First an `edg-job-list-match` is issued to get the list of the available resources on the GRID. Using an appropriate `.conf` file it is possible to choose the VO and a particular Resource Broker to submit the job to e.g. `edt003.cnaf.infn.it` (gridit Resource Broker).

```
[inst-user@ui-alien test-bussola]# edg-job-list-match
--config-vo gridit.conf inst-lambda_trans.jdl
```

```
Selected Virtual Organisation name (from --config-vo option): alice
Connecting to host edt003.cnaf.infn.it, port 7772
```

COMPUTING ELEMENT IDs LIST

The following CE(s) matching your job requirements have been found:

CEId

```

grid0.fe.infn.it:2119/jobmanager-lcgpbs-infinite
grid0.fe.infn.it:2119/jobmanager-lcgpbs-long
grid0.fe.infn.it:2119/jobmanager-lcgpbs-short
grid002.ca.infn.it:2119/jobmanager-lcgpbs-infinite
grid002.ca.infn.it:2119/jobmanager-lcgpbs-long
grid002.ca.infn.it:2119/jobmanager-lcgpbs-short
grid008.to.infn.it:2119/jobmanager-lcgpbs-infinite
grid008.to.infn.it:2119/jobmanager-lcgpbs-lcg
grid008.to.infn.it:2119/jobmanager-lcgpbs-long
grid008.to.infn.it:2119/jobmanager-lcgpbs-short
gridba2.ba.infn.it:2119/jobmanager-lcgpbs-infinite
gridba2.ba.infn.it:2119/jobmanager-lcgpbs-long
griditce01.na.infn.it:2119/jobmanager-lcgpbs-infinite
griditce01.na.infn.it:2119/jobmanager-lcgpbs-long
lxn1181.cern.ch:2119/jobmanager-lcgpbs-infinite
lxn1181.cern.ch:2119/jobmanager-lcgpbs-long
lxn1184.cern.ch:2119/jobmanager-lcglsf-grid
t2-ce-01.lnl.infn.it:2119/jobmanager-lcglsf-grid_lcg1
atlasce.lnf.infn.it:2119/jobmanager-lcgpbs-infinite
grid20.bo.ingv.it:2119/jobmanager-lcgpbs-long
grid20.bo.ingv.it:2119/jobmanager-lcgpbs-short

```

The JDL file contains a shell script (`inst-lambda_trans.sh`) that downloads the software in one of the listed Computing Elements and installs it in the `/opt/exp_software/` directory. Each queue (short, long, infinite, lcg, ecc.) is listed as a single Computing Element.

In order to install the software on e.g. `grid002.ca.infn.it` (INFN Cagliari Computing Element) the job must be forced to this specific host. Since the installation can take several minutes because of the software download and its distribution on the WNs via `nfs`, a “long” queue should be chosen:

```
grid002.ca.infn.it:2119/jobmanager-lcgpbs-long.
```

I have used the submission command with `--resource` option to select Cagliari resources, and I selected the “long” queue that accepts jobs up to 10 hours of CPU time:

```
[inst-user@ui-alien test-bussola]# edg-job-submit --resource
grid002.ca.infn.it:2119/jobmanager-lcgpbs-long
--config-vo gridit.conf inst-lambda_trans.jdl
```

The job is thus submitted on the GRID. The RB returns a message with information about the VO, RB and Logging hosts used:

```
Selected Virtual Organisation name (from --config-vo option): alice
Connecting to host edt003.cnaf.infn.it, port 7772
Logging to host lxn1177.cern.ch, port 9002
*****
                        JOB SUBMIT OUTCOME
The job has been successfully submitted to the Network Server.
Use edg-job-status command to check job current status.
Your job identifier (edg_jobId) is:
- https://edt003.cnaf.infn.it:9000/UHBLs7IehltE2wUosuYqA
*****
```

The submission returns the job ID (`edg_jobId`), that univocally identifies a job on the GRID. The ID is automatically stored in a file. The status of a job should be periodically checked using the `edg-job-status` command with the job ID as argument. The status is at first “READY”:


```
[inst-user@ui-alien test-bussola]#
edg-job-status https://edt003.cnaf.infn.it:9000/UHLLS7IehltE2wUosuYqA
*****
BOOKKEEPING INFORMATION:
Status info for the Job :
https://edt003.cnaf.infn.it:9000/UHLLS7IehltE2wUosuYqA
Current Status:      Ready
Status Reason:       unavailable
Destination:         grid002.ca.infn.it:2119/jobmanager-lcgpbs-long
reached on:          Mon Jul  5 09:43:55 2004
*****
```

When the job reaches the LRMS the status becomes “SCHEDULED”:

```
[inst-user@ui-alien test-bussola]#
edg-job-status https://edt003.cnaf.infn.it:9000/UHLLS7IehltE2wUosuYqA
*****
BOOKKEEPING INFORMATION:
Status info for the Job :
https://edt003.cnaf.infn.it:9000/UHLLS7IehltE2wUosuYqA
Current Status:      Scheduled
Status Reason:       Job successfully submitted to Globus
Destination:         grid002.ca.infn.it:2119/jobmanager-lcgpbs-long
reached on:          Mon Jul  5 09:44:20 2004
*****
```

A further check can be performed: the job status in the queue on LRMS. The job is the 945th on the pbs server grid002.ca.infn.it. The user is mapped as alicesgm and the queue is the long one. The test is made from a different host (lcg-ui-01) and by a different user:

```
[prod-user@lcg-ui-01 lambda-trans]#
globus-job-run grid002.ca.infn.it/usr/bin/qstat
```

Job id	Name	User	Time Use	S	Queue
945.grid002	STDIN	alicesgm	0	R	long

```
[prod-user@lcg-ui-01 lambda-trans]#
globus-job-run grid002.ca.infn.it/usr/bin/qstat
```

Job id	Name	User	Time Use	S	Queue
945.grid002	STDIN	alicesgm	00:00:03	R	long

The status is now “RUNNING”:

```
[inst-user@ui-alien test-bussola]#
edg-job-status https://edt003.cnaf.infn.it:9000/UHBLs7IehltE2wUosuYqA
*****
BOOKKEEPING INFORMATION:
Status info for the Job :
https://edt003.cnaf.infn.it:9000/UHBLs7IehltE2wUosuYqA
Current Status:      Running
Status Reason:       Job successfully submitted to Globus
Destination:         grid002.ca.infn.it:2119/jobmanager-lcgpbs-long
reached on:          Mon Jul  5 09:46:43 2004
*****
```

The job is eventually completed. The status is “DONE with SUCCESS”:

```
[inst-user@ui-alien test-bussola]#
edg-job-status https://edt003.cnaf.infn.it:9000/UHBLs7IehltE2wUosuYqA
*****
BOOKKEEPING INFORMATION:
Status info for the Job :
```

```

https://edt003.cnaf.infn.it:9000/UHBLLs7IehltE2wUosuYqA
Current Status:      Done (Success)
Exit code:           0
Status Reason:      Job terminated successfully
Destination:        grid002.ca.infn.it:2119/jobmanager-lcgpbs-long
reached on:         Mon Jul  5 10:05:37 2004
*****

```

In order to check the job execution the standard output/error files should be retrieved and examined, using `edg-job-get-output` command:

```

[inst-user@ui-alien test-bussola]# edg-job-get-output
https://edt003.cnaf.infn.it:9000/UHBLLs7IehltE2wUosuYqA

```

```

Retrieving files from host: edt003.cnaf.infn.it (for
https://edt003.cnaf.infn.it:9000/UHBLLs7IehltE2wUosuYqA)
*****
                                JOB GET OUTPUT OUTCOME
Output sandbox files for the job:
- https://edt003.cnaf.infn.it:9000/UHBLLs7IehltE2wUosuYqA
have been successfully retrieved and stored in the directory:
/tmp/jobOutput/inst-user_UHBLLs7IehltE2wUosuYqA
*****

```

We can read output file in a temporary directory:

```

-rw-r--r-- inst-user tolcgi    153  Jul  5 12:33 inst-lambda_trans.out
-rw-r--r-- inst-user tolcgi      0  Jul  5 12:33 inst-lambda_trans.err

```

The error file is empty and the `.out` file, listed below, contains the software installation log:

- with the `hostname` command we ask on which WN our installation job is “RUNNING”:

```
-- Starting LAMBDA_TRANS installation on
grid014.ca.infn.it at Mon Jul  5 11:45:10 CEST 2004 --
```

- The installation script gets the name of the Computing Element where the job is to be install and check which scheduler is installed on the CE. On the INFN Production GRID are installed either PBS (Portable Batch System) or LSF (Load Sharing Facility):

```
-- Step 0: checking for /var/spool/pbs/server_name --
Oooh I've found a /var/spool/pbs/server_name file ... c'e' PBS
MY_CE = grid002.ca.infn.it
```

- The script checks for old .tgz files, the old archives will be removed.

```
-- Step 1: removing old lambda_trans.tgz --
```

- Then the disk space availability is checked. If the necessary disk space is not available, the installation will be stopped and an error is reported. The .tgz file, downloaded from the web server occupies about 1GB of disk space on the tmp directory. After uncompress the archive in the installation directory the COMPASS simulation software occupies about 2,5 GB of disk space:

```
-- Step 2: disk space availability --
Space Available on /tmp = 9231876 Byte
Space remaining = 8231876 Byte   /tmp OK
Space Available on /opt/exp_software = 47698528 Byte
Space remaining = 45698528 Byte   /opt/exp_software OK
```

- the .tgz archive is downloaded from a web site using the `wget` command:

```
-- Step 3: Downloading ARCHIVE lambda_trans.tgz
from http://www.to.infn.it/grid-prod/software/lambda_trans.tgz
cd /tmp
wget -N -q http://www.to.infn.it/grid-prod/software/lambda_trans.tgz
terminated lambda_trans.tgz download at Mon Jul  5 11:56:47 CEST 2004
```

- check the installation directory; if the right directory is not there it will be created. The .tgz archive is unpacked in the compass directory:

```
-- Step 4 : Unpack the ARCHIVE lambda_trans.tgz
cd /opt/exp_software/alice
Oooh I've found a /opt/exp_software/alice/compass dir
cd /opt/exp_software/alice/compass
tar -xzf /tmp/lambda_trans.tgz
terminated lambda_trans.tgz unpack at Mon Jul  5 12:01:05 CEST 2004
```

- after the installation procedure the script sets the environment variable, that will be published by the CE. When the user asks for available resources, the RB lists only the CEs that publish the right “envar”:

```
-- Step 5: setting envar --
lcg-ManageVOTag -vo alice -host grid002.ca.infn.it
--add -tag VO-alice-cmp-3.1.2
done setting tag at Mon Jul  5 12:01:08 CEST 2004
```

- the .tgz file is removed to clean up the CE disk:

```
-- Step 6: Removing ARCHIVE lambda_trans.tgz
rm -f /tmp/lambda_trans.tgz*
```

- the installation procedure tooks, in this specific case, about 16 minutes:

```
-- Terminated LAMBDA_TRANS installation at
   Mon Jul  5 12:01:09 CEST 2004 --
```

The installation time may vary in different site depending on network bandwidth, CPU speed of the CE, farm occupancy, etc.

6.3 Installation TEST

The next step is a small job submission performed through `run-20-lambda-trans.sh`. The submitting user (prod-user) is different from the installing one. An `edg-job-list-match` should be issued first to check the environment.

The list-match returns more than one CE. It should return all the CEs with the required version of software (VO-alice-cmp-3.1.2). Choosing the CE, where I have just instal the software, a short job is submitted.

As before the job progress can be monitored by using `edg-job-status` each one having as argument of every command is the job identifier (`edg_jobId`). The job change its status from “READY” to “SCHEDULED”:

The current status of the job is then “RUNNING”:

```
[prod-user@lcg-ui-01 lambda-trans]#
edg-job-status https://edt003.cnaf.infn.it:9000/uC-De3IiWgnolhKheSxMug
```

```
*****
BOOKKEEPING INFORMATION:
Status info for the Job :
https://edt003.cnaf.infn.it:9000/uC-De3IiWgnolhKheSxMug
Current Status:      Running
Status Reason:      Job successfully submitted to Globus
Destination:        grid002.ca.infn.it:2119/jobmanager-lcgpbs-infinite
reached on:         Mon Jul  5 14:46:15 2004
*****
```

with the `globus-job-run` command I can issue a `qstat` command to check the status of the job in the queue. The job is the 957th on the pbs server `grid002.ca.infn.it`. The user is mapped as `infngrid001` and the queue is the “infinite” one:

```
[prod-user@lcg-ui-01 lambda-trans]#
globus-job-run grid002.ca.infn.it/usr/bin/qstat
```

Job id	Name	User	Time Use	S	Queue
957.grid002	STDIN	infngrid001	00:14:17	R	infinite

I can check, again, the job status using the `edg-job-status` command. Now the job is finished. The status is “DONE with SUCCESS”:

```
[prod-user@lcg-ui-01 lambda-trans]#
edg-job-status https://edt003.cnaf.infn.it:9000/uC-De3IiWgnolhKheSxMug
```

```
*****
BOOKKEEPING INFORMATION:
Status info for the Job :
https://edt003.cnaf.infn.it:9000/uC-De3IiWgnolhKheSxMug
Current Status:      Done (Success)
Exit code:           0
Status Reason:       Job terminated successfully
Destination:         grid002.ca.infn.it:2119/jobmanager-lcgpbs-infinite
reached on:          Mon Jul  5 14:53:05 2004
*****
```

Using the `edg-job-get-output` command I can retrieve the output/error files.

I can then check the size of the file the output sandbox. I know that a standard simulation produces `.out` and `.err` files with a typical size of 1K for `.err` and 1MB for `.out`:

```
[prod-user@lcg-ui-01 lambda-trans]#  
ls -al /tmp/jobOutput/prod-user_uC-De3IiWgnolhKheSxMug/run-20-lambda_trans*  
-rw-r--r-- 1 prod-user  tolcgi  985 Jul 5 16:53  run-20-lambda_trans.err  
-rw-r--r-- 1 prod-user  tolcgi 893162 Jul 5 16:53  run-20-lambda_trans.out
```

In `run-20-lambda_trans.out` file is reported the result of the job submission.

Chapter 7

COMPASS@GRID

In this chapter I will report about the result of job submission for COMPASS simulation on the GRID; particular attention be devoted to the statistics of submitted, failed and correctly completed jobs. Also some statistics will be given about disk usage, transfer rate and cpu time. The preliminary tests on compatibility, installation and submission procedure are performed on LCG and INFN Production GRID installed in the INFN Torino site. For COMPASS Data Challenge I exploited the functionalities of the workload management system while I didn't use the data management part of the GRID middleware. All the output files are transferred to the COMPASS farm storage for the analysis procedure, and then deleted from the Torino SE.

7.1 Roadmap

The work is split into two separate phases that reflect two quite different sets of simulated events with longitudinal and transverse polarizations. In the first phase I simulated Λ decay with longitudinal polarization, with at least one Λ per event. In the second phase I started by simulation Λ decay with transversal polarization, with at least one Λ per event; in the future I plan to start a Data Challenge to analyze background Λ decay with transversal polarization. All the output are transfer directly to the Torino SE (grid009.to.infn.it) via gridftp

(GRID File Transfer Protocol) therefore at this moment I don't use the Data Management system. The output files are stored temporarily in a dedicated area on the SE. Periodically the output are transferred to the disk storage of the COMPASS farm for the analysis.

In 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

I planned to use more than one RB to prevent the COMPASS MonteCarlo Data Challenge from stopping altogether during the period of unavailability of a specific RB (upgrade to new release, power cuts, network problem, ecc.). To select a given RB it is sufficient to specify an appropriate configuration file in the argument of the submit command; wich Virtual Organization should be used is specified in the same file. Below a `.conf` file is reported that selects two RBs to submit the job. If the first RB is unavailable the second one will be used:

```
[
##VirtualOrganisation = "dteam";
##VirtualOrganisation = "alice";
VirtualOrganisation = "infngrid"; <----

## HLR location is optional. Uncomment and fill correctly for
## enabling accounting
#HLRLocation = "fake HLR Location"
## MyProxyServer is optional. Uncomment and fill correctly for
## enabling proxy renewal. This field should be set equal to
## MYPROXY_SERVER environment variable
```

```
##MyProxyServer = "fake My Proxy Server"

NSAddresses = {"edt003.cnaf.infn.it:7772","grid014.ct.infn.it:7772"}; <----
LBAddresses = {"edt003.cnaf.infn.it:9000","grid014.ct.infn.it:9000"}; <----

##HLRLocation = "fake HLR Location"
]
```

7.2 Longitudinal simulation (First phase)

Five sites were chosen for the initial test of phase one with longitudinal configuration. The choice criteria included the number of available CPUs for each site and the speed of the CPUs. I choose to install the software in a large site with at least 20 CPUs (Torino, Milano, Napoli and Roma1), or in sites with fast CPUs (Cagliari P4 2,8GHz), to test the different CPU performance .

7.2.1 Results

The submission report is shown in Tab.7.1 and in Fig.7.1. At the end of this DC 2851 jobs were submitted 2851 jobs. Among then, 305 test jobs, were submitted without using any GRID middleware on Torino resources, 2408 jobs submitted via the CNAF RB (edt003.cnaf.infn.it) and 445 via the Catania RB (grid014.ct.infn.it). Since the 305 test jobs were submitted manually (interactive mode) on the Torino WNs, is impossible to qualify these jobs as “ABORTED”. The report of jobs “ABORTED” because of GRID problem is shown in Tab.7.4; 5% (142) of the jobs are aborted for different reasons: network, power cuts, WNs disk full, RB problem, ecc. However from the COMPASS point of view there was a further 3% (85) of failed simulations (unreadable .root output file), even if in the GRID sense these jobs were succesfully completed. On the whole I simulated about 85 Million of events, submitting about 30 job/day for 75 day, with a peak of 72 job/day (18/05/2004).

SITE	CE	Submitted jobs	Aborted jobs
TORINO	grid008.to.infn.it (manual)	305	X
TORINO	grid008.to.infn.it (via RB)	1095	49
MILANO	t2-ce-01.mi.infn.it	495	0
NAPOLI	griditce01.na.infn.it	165	50
ROMA1	t2-de-01.roma1.infn.it	431	49
CAGLIARI	grid002.ca.infn.it	2044	5

Table 7.1: *Submitted and aborted jobs for longitudinal production*

Resource Broker	Submitted jobs	Aborted jobs
edt003.cnaf.infn.it	2408	103
grid014.ct.infn.it	445	8

Table 7.2: *Submitted and aborted jobs by Resource Broker for longitudinal production*

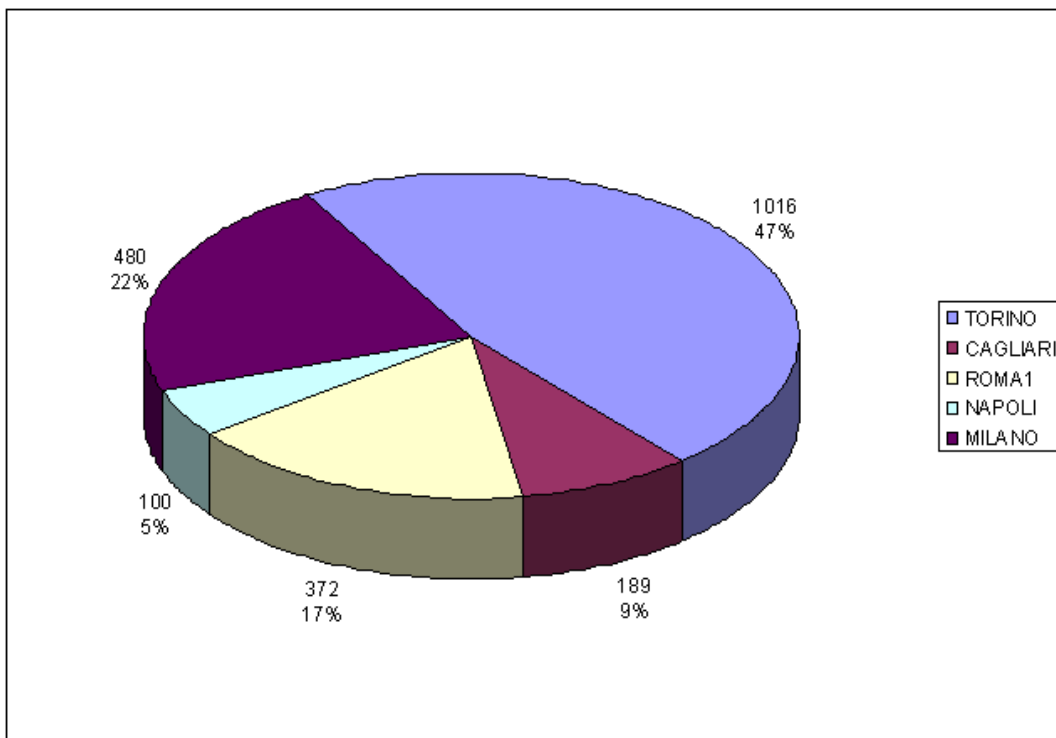


Figure 7.1: Jobs-site distribution for longitudinal production.

7.2.2 Submission technique

Since the CPU load generated by such a DC was relevant for a GRID resource, I planned to use a “non-invasive” submission method to not interfere with “official” VOs data challenges. I always used the `edg-job-submit` command with “`--resources`” option to select a particular CE. In the submission script I specified how many jobs should run on the site queue simultaneously; periodically via the UNIX cron facility the script checked the status of every job submitted in a CE. When a job was finished the script retrieved the output sandbox and afterwards it submitted automatically some fresh jobs up to the limit, that I decided before, for the site. I didn’t want to saturate a site with COMPASS jobs. The listing of the script used for the first phase of submission test is reported in appendix A. The output sandbox, retrieved automatically from RB, is saved in dedicated directories for “DONE with SUCCESS” jobs. The job ID of a FAILED job is saved in a special directory for further investigation. By saving `.log` files it is possible to check single job histories and know why a job failed or aborted. A job fails for problem due to network, RB, disk full on WNs, power cuts, ecc.

“DONE with SUCCESS” means that a job was processed correctly from the GRID point of view, so it is possible that a job aborted from the COMPASS point of view, e.g. for a bad configuration, is registered as “DONE” for GRID middleware. For the GRID completed jobs there is also a `.log` file saved in the SE. The `.log` file contains the name of the `.root` output, the hostname where the job ran and some information about CPU, memory etc.:

```
--Job Log mDST_mc_trans.lnl.infn.it.2004_11_24-07_58_54-06.log --
hostname : cmsfarm-03-03.lnl.infn.it
model name : Intel(R) Xeon(TM) CPU 2.80GHz
cache size : 512 KB
cpu MHz : 2799.302
bogomips : 5583.66
MemTotal:      2064804 kB
```

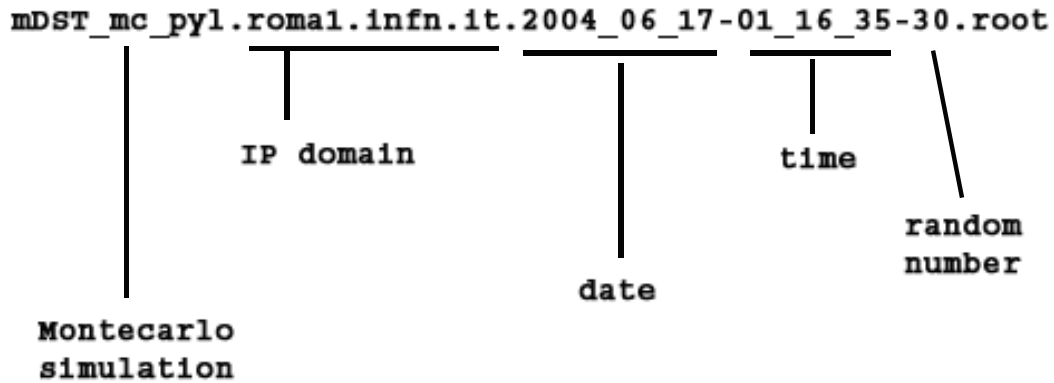


Figure 7.2: Output file name composition

```
run 24490
put 24
date 2004_11_24-14:47:28
-----
```

The name of the output file is automatically generated as shown in Fig.7.2. The name contains the IP domain where it is executed, the job start time and a random number to avoid that jobs started at the same time can get the same name.

7.2.3 Disk space

Every job produces an output simulation file (`.root`) that occupy about 65MB of disk space for 30K longitudinal polarization events. The 2462 output files (`mDST .root`) altogether use about 170GB of disk space.

7.3 Transverse simulation (Second phase)

For the transversal simulation DC I selected four sites: Torino, LNL, Cagliari, Ferrara. I did not include, in this second phase, Milano and Roma1 resources

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 7.3: *Submitted and aborted jobs for transverse production.*

because they were exclusively “reserved” for ATLAS DC since the end of June 2004.

7.3.1 Results

The submission report is shown in Tab.?? and in Fig.7.3. At the end of this DC I have submitted 5808 jobs, 4746 jobs via the CNAF RB (edt003.cnaf.infn.it), 563 jobs via the CATANIA RB (grid014.ct.infn.it) and 269 jobs via the EGEE RB (egee-rb-01.cnaf.infn.it). I simulated $1,07 * 10^8$ events. In Tab.?? is shown the report of jobs “ABORTED”; 3,7% (214) of the jobs are aborted because of GRID problem, mainly for RB problem. 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 contex. On the whole I submitted 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).

7.3.2 Submission technique

For the transversal simulation I used the “non-invasive” submission technique described in the previous section. However for a short period of two weeks I used a different technique. During this period I submitted my jobs without specifying the “--resources” option. The RB searches the available resources

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 7.4: Submitted and aborted jobs by Resource Broker for transverse production

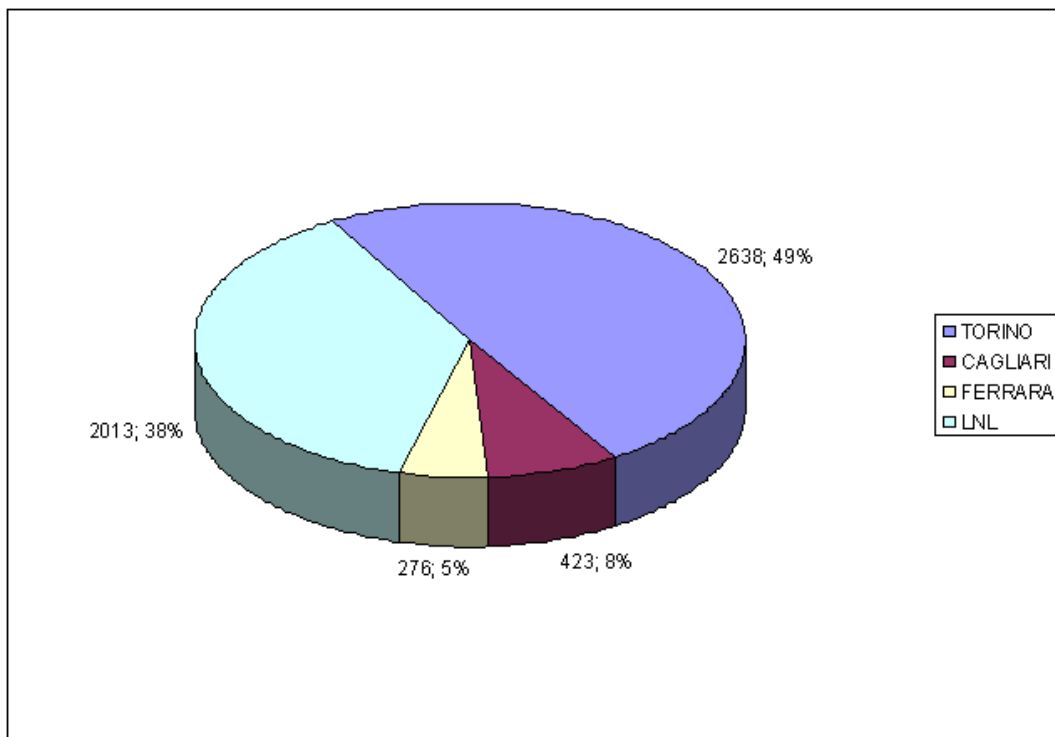


Figure 7.3: Jobs-site distribution for transverse production.

asking for the published environment variable. I preferred the “non-invasive” technique, because it is easier to control the job submission. Also for transversal production the name of the output file is generated automatically, e.g.:

```
mDST_mc_trans.ca.infn.it.2004_07_05-16_30_14-26.root
```

The output file and a small log file are stored in different directories on the SE:

```
cmp-trans/log  
cmp-tran/output
```

the output sandboxes are stored in the UI (lcg-ui-01.to.infn.it)

7.3.3 Disk space

Every job produces a simulation file (`.root`) that takes about 60MB disk space for 20K transverse polarization events. The 5356 output files (mDST `.root`) used about 320GB of disk space.

7.4 CPU performance

Thanks to the large amount of jobs submitted (7216 jobs) it is possible to test the real performance of different CPU. Reading `.log` file it is possible to extrapolate the CPU time utilisation for a simulation job. A plot of CPU time by CPU speed for Λ transverse simulations (5353 jobs) is shown in Fig.7.4. If the jobs uses CPUs with the same internal architecture (e.g. PIII with different clock frequencies), the decreasing of CPU time is proportional to the increasing of CPU speed. With a different kind of CPU architecture (e.g. P4 or Xeon) this proportionality is not verified. It is necessary to mark that a job do not use only the CPU, but also memory and disk space so in Fig.7.4 a spread in CPU time for the same CPU type can be observed. However the proportionality time/speed is obvious.

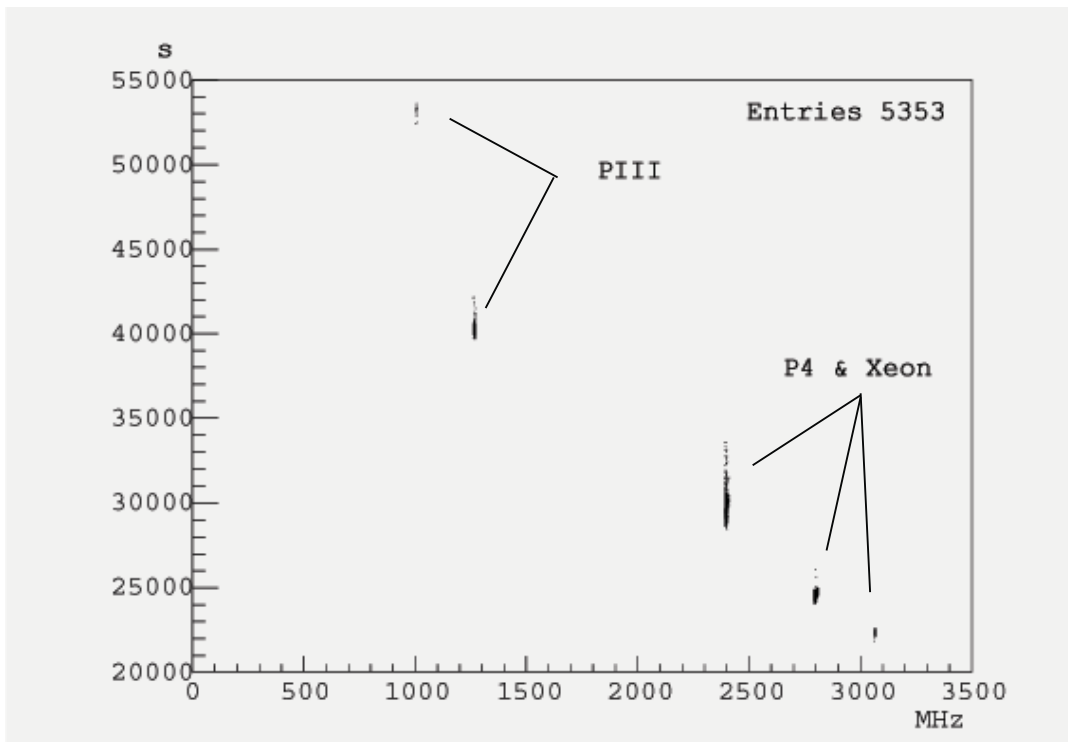


Figure 7.4: CPU time vs CPU speed for Λ transverse simulations

7.5 DC Chronology

In this section the history of the COMPASS@GRID project is reported. It started at the beginning of 2004 with a preliminary installation of COMPASS simulation package on a dedicated GRID host. This preliminary installation aimed at checking the right environment needed by COMPASS software to run and also to check the post-installation manual procedure to set several internal variables or file permissions. Another aim was to find out possible conflicts between COMPASS simulation software and GRID middleware. Below I report the history of this test:

- 20/03/2004 - First job submission of a 20 event longitudinal simulation as compatibility test between COMPASS software and GRID middleware software. These compatibility tests are submitted on Torino site resources.
- 08-20/04/2004 - Installation test on Torino Computing Element. I performed a preliminary installation procedure. No remote installation was performed; the software was manually installed directly on the software directory of the CE.
- 20-30/04/2004 - submission test on Torino CE. I start by submitting 2 jobs simultaneously
- 06/05/2004 - installation and first submission on the Cagliari CE
- 11/05/2004 - installation and first submission on the Roma1 CE
- 13/05/2004 - installation and first submission on the Milano CE
- 18/05/2004 - installation and first submission on the Napoli CE

The first phase of submission work stopped at 02/07/2004 with the submission of the last COMPASS longitudinal event simulation. Simultaneously, from 09/06 I started the preliminary installation test of a new software release for COMPASS transversal event simulation

- 19/06/2004 - installation and first submission on the Torino CE, using last version of installation procedure as shown in chapter 6
- 06/07/2004 - installation and first submission on the Cagliari CE
- 07/07/2004 - installation and first submission on the LNL CE
- 08/07/2004 - installation and first submission on the Ferrara CE

Chapter 8

Transverse target spin data analysis

During summer 2002 and 2003 three data taking periods have been devoted to the DIS with transverse target polarized in transverse mode. In the middle of each period the spin direction of the two target cells has been reversed. Sets of runs of comparable size have been recorded, three for each target spin configuration. The six data sets are denoted, following the COMPASS convention (SPS period convention), as **P2B**, **P2C**, **P2H(1-2)**, **P1G** and **P1H**. The target spin configuration for the different data sets are shown in Fig.8.1.

The conventions used in this Chapter for the orientation of the reference system are the following: the z axis coincides with the beam direction with origin at the center of the target, the y axis is directed vertically, from bottom to top, and the x axis direction forms a right-handed system with the other two axes.

8.1 Transverse polarization

The transverse polarization of Lambdas produced from a transversely polarized target allows to access the transversity quark distributions of the nucleons,

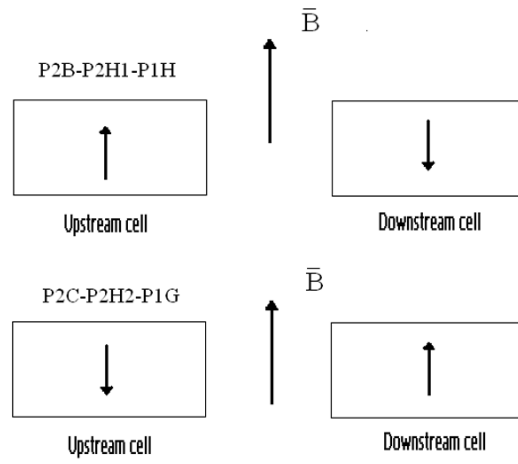


Figure 8.1: Configurations of the target cells polarizations for the $P2B-P2H1-P1H$ (top) and $P2C-P2H2-P1G$ (bottom) data sets. The \vec{B} arrow represents the target dipole magnetic field direction.

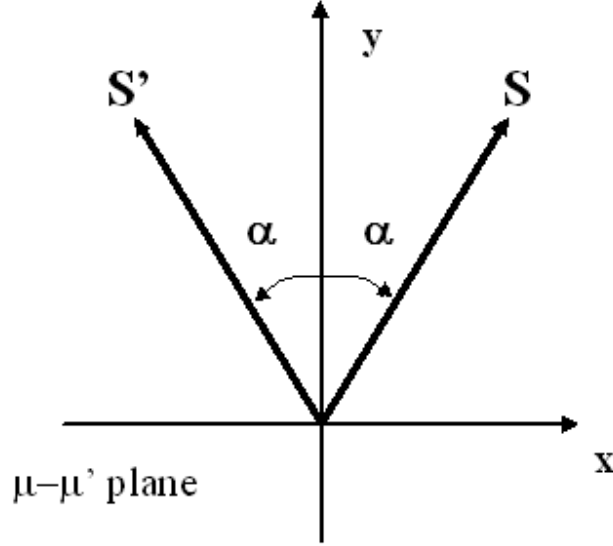


Figure 8.2: Definition of the S' axis used for the extraction of the Λ and $\bar{\Lambda}$ transverse spin.

through the following relation:

$$\begin{aligned}
 P_{\Lambda}^{S'} &= \frac{d\sigma^{lp^{\dagger} \rightarrow \Lambda^{\dagger} X} - d\sigma^{lp^{\dagger} \rightarrow \Lambda^{\dagger} X}}{d\sigma^{lp^{\dagger} \rightarrow \Lambda^{\dagger} X} + d\sigma^{lp^{\dagger} \rightarrow \Lambda^{\dagger} X}} = \\
 &= P_T' f D(y) \cdot \frac{\sum_q e_q^2 \Delta_T q(x) \Delta_T D_{\Lambda/q}(z)}{\sum_q e_q^2 q(x) D_{\Lambda/q}(z)} \quad (8.1)
 \end{aligned}$$

Here $P_{\Lambda}^{S'}$ is the Λ polarization along the S' vector, obtained by:

1. taking the component S of the target polarization vector perpendicular to the direction of the exchanged virtual photon,
2. defining S' as the reflection of S with respect to the normal to the scattering plane, as shown in Fig.8.2.

In general the transverse polarization $P_{\Lambda}^{S'}$ cannot be directly extracted from the experimental data, due to the distortion introduced by the apparatus acceptance. However this distortion can be corrected by exploiting the symmetries

of the apparatus, as it has been done in the case of the extraction of D_{NN} . In the transversity case the acceptance cancellation is slightly different, since the polarization axis changes event by event and therefore the up/down symmetry cannot be exploited.

Let's first of all introduce the expressions of the number of Lambdas emitting the proton in the angular bin $\theta_{S'}^* \pm \Delta$ ($\theta_{S'}^*$ is the angle formed by the proton momentum and the polarization axis, in the Lambda rest frame). The numbers are explicitly given for the two target spin orientations (+ or -) and the two data taking periods (1 or 2).

$$\begin{aligned}
 N_{1(2)}^+(\theta_{S'}^*) &= \Phi_{1(2)}^+ \left(\frac{d\sigma}{d\Omega} \right)^0 \left(1 + \alpha P_{\Lambda}^{S'} \cos \theta_{S'}^* \right) \cdot Acc_{1(2)}^+(\cos \theta_{S'}^*) \\
 N_{1(2)}^-(\theta_{S'}^*) &= \Phi_{1(2)}^- \left(\frac{d\sigma}{d\Omega} \right)^0 \left(1 - \alpha P_{\Lambda}^{S'} \cos \theta_{S'}^* \right) \cdot Acc_{1(2)}^-(\cos \theta_{S'}^*)
 \end{aligned} \tag{8.2}$$

The following equivalences of acceptance functions hold:

$$Acc_1^+(\theta_{S'}^*) = Acc_2^-(\theta_{S'}^*) \tag{8.3}$$

$$Acc_1^+(\pi - \theta_{S'}^*) = Acc_2^-(\pi - \theta_{S'}^*) \tag{8.4}$$

$$Acc_1^-(\theta_{S'}^*) = Acc_2^+(\theta_{S'}^*) \tag{8.5}$$

$$Acc_1^-(\pi - \theta_{S'}^*) = Acc_2^+(\pi - \theta_{S'}^*) \tag{8.6}$$

from which it follows

$$\begin{aligned}
 &\sqrt{Acc_1^+(\theta_{S'}^*) \cdot Acc_2^+(\theta_{S'}^*)} + \sqrt{Acc_1^-(\pi - \theta_{S'}^*) \cdot Acc_2^-(\pi - \theta_{S'}^*)} = \\
 &= \sqrt{Acc_1^+(\pi - \theta_{S'}^*) \cdot Acc_2^+(\pi - \theta_{S'}^*)} + \sqrt{Acc_1^-(\theta_{S'}^*) \cdot Acc_2^-(\theta_{S'}^*)}
 \end{aligned} \tag{8.7}$$

Let's now consider the following counting rate asymmetry:

$$\epsilon_T(\theta_{S'}^*) = \frac{\left[\sqrt{\frac{N_1^+(\theta_{S'}^*) \cdot N_2^+(\theta_{S'}^*)}{\Phi_1^+ \cdot \Phi_2^+}} + \sqrt{\frac{N_1^-(\pi-\theta_{S'}^*) \cdot N_2^-(\pi-\theta_{S'}^*)}{\Phi_1^- \cdot \Phi_2^-}} \right] - \left[\sqrt{\frac{N_1^+(\pi-\theta_{S'}^*) \cdot N_2^+(\pi-\theta_{S'}^*)}{\Phi_1^+ \cdot \Phi_2^+}} + \sqrt{\frac{N_1^-(\theta_{S'}^*) \cdot N_2^-(\theta_{S'}^*)}{\Phi_1^- \cdot \Phi_2^-}} \right]}{\left[\sqrt{\frac{N_1^+(\theta_{S'}^*) \cdot N_2^+(\theta_{S'}^*)}{\Phi_1^+ \cdot \Phi_2^+}} + \sqrt{\frac{N_1^-(\pi-\theta_{S'}^*) \cdot N_2^-(\pi-\theta_{S'}^*)}{\Phi_1^- \cdot \Phi_2^-}} \right] + \left[\sqrt{\frac{N_1^+(\pi-\theta_{S'}^*) \cdot N_2^+(\pi-\theta_{S'}^*)}{\Phi_1^+ \cdot \Phi_2^+}} + \sqrt{\frac{N_1^-(\theta_{S'}^*) \cdot N_2^-(\theta_{S'}^*)}{\Phi_1^- \cdot \Phi_2^-}} \right]} \quad (8.8)$$

Since the two target cells see in the same period the same muon flux ($\Phi_{1(2)}^+ = \Phi_{1(2)}^-$), one has:

$$\Phi_1^+ \cdot \Phi_2^+ = \Phi_1^- \cdot \Phi_2^- = \Phi_1 \cdot \Phi_2 \quad (8.9)$$

Therefore it is possible to demonstrate that

$$\epsilon_T(\theta_{S'}^*) = \alpha P_\Lambda^{S'} \cos \theta_{S'}^*, \quad (8.10)$$

and the Lambda polarization can be extracted from the slope of the $\epsilon_T(\theta_{S'}^*)$ distribution.

The formula above becomes simpler if only two bins in the proton decay angle distributions are considered. For that we need to introduce the expressions for the number of protons emitted in the direction of the polarization axis (U) and opposite to it (D). We will derive the expressions separately for the two target spin orientations (+ or -) and the two data taking periods (1 or 2).

$$\begin{aligned}
N_{1(2)}^+(D) &= \int_{-1}^0 \frac{N_{Tot}^{1(2),+}}{2} \left(1 + \alpha P_{\Lambda}^{S'} \cos \theta_T^*\right) \cdot Acc_{1(2)}^+(\cos \theta_T^*) d \cos \theta_T^* = \\
&= \frac{N_{Tot}^{1(2),+}}{2} \left(1 - \frac{\alpha P_{\Lambda}^{S'}}{2}\right) Acc_{1(2)}^+(D) \tag{8.11}
\end{aligned}$$

$$\begin{aligned}
N_{1(2)}^+(U) &= \int_0^1 \frac{N_{Tot}^{1(2),+}}{2} \left(1 + \alpha P_{\Lambda} \cos \theta_T^*\right) \cdot Acc_{1(2)}^+(\cos \theta_T^*) d \cos \theta_T^* = \\
&= \frac{N_{Tot}^{1(2),+}}{2} \left(1 + \frac{\alpha P_{\Lambda}}{2}\right) Acc_{1(2)}^+(U) \tag{8.12}
\end{aligned}$$

$$\begin{aligned}
N_{1(2)}^-(D) &= \int_{-1}^0 \frac{N_{Tot}^{1(2),-}}{2} \left(1 - \alpha P_{\Lambda} \cos \theta_T^*\right) \cdot Acc_{1(2)}^-(\cos \theta_T^*) d \cos \theta_T^* = \\
&= \frac{N_{Tot}^{1(2),-}}{2} \left(1 + \frac{\alpha P_{\Lambda}}{2}\right) Acc_{1(2)}^-(D) \tag{8.13}
\end{aligned}$$

$$\begin{aligned}
N_{1(2)}^-(U) &= \int_0^1 \frac{N_{Tot}^{1(2),-}}{2} \left(1 - \alpha P_{\Lambda} \cos \theta_T^*\right) \cdot Acc_{1(2)}^-(\cos \theta_T^*) d \cos \theta_T^* = \\
&= \frac{N_{Tot}^{1(2),-}}{2} \left(1 - \frac{\alpha P_{\Lambda}}{2}\right) Acc_{1(2)}^-(U) \tag{8.14}
\end{aligned}$$

The counting rate asymmetry is given in this simplified case by

$$\begin{aligned}
\epsilon_T &= \frac{\left[\sqrt{\frac{N_1^+(U) \cdot N_2^+(U)}{\Phi_1^+ \cdot \Phi_2^+}} + \sqrt{\frac{N_1^-(D) \cdot N_2^-(D)}{\Phi_1^- \cdot \Phi_2^-}} \right] - \left[\sqrt{\frac{N_1^+(D) \cdot N_2^+(D)}{\Phi_1^+ \cdot \Phi_2^+}} + \sqrt{\frac{N_1^-(U) \cdot N_2^-(U)}{\Phi_1^- \cdot \Phi_2^-}} \right]}{\left[\sqrt{\frac{N_1^+(U) \cdot N_2^+(U)}{\Phi_1^+ \cdot \Phi_2^+}} + \sqrt{\frac{N_1^-(D) \cdot N_2^-(D)}{\Phi_1^- \cdot \Phi_2^-}} \right] + \left[\sqrt{\frac{N_1^+(D) \cdot N_2^+(D)}{\Phi_1^+ \cdot \Phi_2^+}} + \sqrt{\frac{N_1^-(U) \cdot N_2^-(U)}{\Phi_1^- \cdot \Phi_2^-}} \right]} = \\
&= \frac{\alpha P_{\Lambda}^{S'}}{2} \tag{8.15}
\end{aligned}$$

The following equivalences of acceptance functions hold:

$$Acc_1^+(U) = Acc_2^-(U) \tag{8.16}$$

$$Acc_1^+(D) = Acc_2^-(D) \tag{8.17}$$

$$Acc_1^-(U) = Acc_2^+(U) \tag{8.18}$$

$$Acc_1^-(D) = Acc_2^+(D) \tag{8.19}$$

from which it follows

$$\begin{aligned} & Acc_1^+(U) \cdot Acc_1^-(D) + Acc_2^+(U) \cdot Acc_2^-(D) = \\ & = Acc_1^+(D) \cdot Acc_1^-(U) + Acc_2^+(D) \cdot Acc_2^-(U) \end{aligned} \quad (8.20)$$

8.2 Useful symmetries

In the COMPASS apparatus the angular acceptance of the polarized target is not equal for the two cells, as shown in Fig.8.3. A comparison of the production yields from the two target cells is strongly affected by false asymmetries arising from differences in the apparatus acceptance. Two measurements with opposite target cells spin configurations, may eliminate this false asymmetry

The rotation invariance properties of the SIDIS process can be used to correct for the apparatus acceptance even in the case of a single measurement. The Λ production process, in the hypothesis of photon exchange, involves electromagnetic and strong interactions which are invariant under a π rad rotation along the beam axis.

A Λ produced with momentum components p_x^Λ and p_y^Λ from a target cell polarized upward is equivalent to a Λ with momentum components $-p_x^\Lambda$ and $-p_y^\Lambda$ coming from a downward-polarized target cell (see Fig.8.4). Provided that the apparatus acceptance is symmetric with respect to the plane defined by the y and z axes, the left and right-produced samples play a role analogous to the opposite target cell orientations.

8.3 MonteCarlo simulations

As shown in previous chapter, I used GRID resources to produce a large amount of MonteCarlo simulated events, in order to investigate various aspects of the experimental apparatus and of the Λ reconstruction efficiency.

In the transverse target spin configuration a 0.5 Tesla dipole magnetic field is present in the polarized target. This field is directed along the y axis, per-

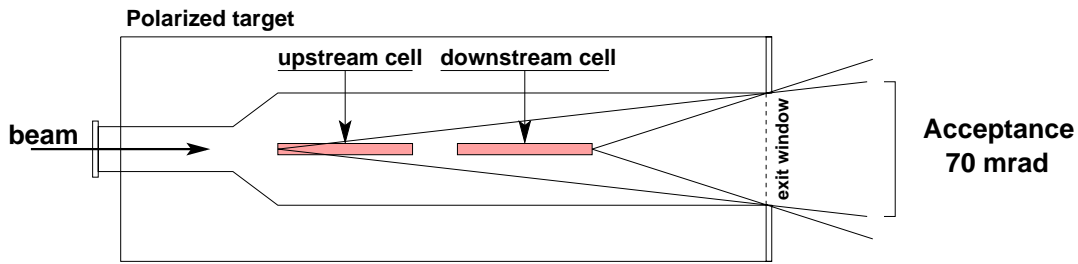


Figure 8.3: The apparatus acceptance for the two cells of the polarized target is constrained by the size of the target exit window. The minimum acceptance is 70 mrad.

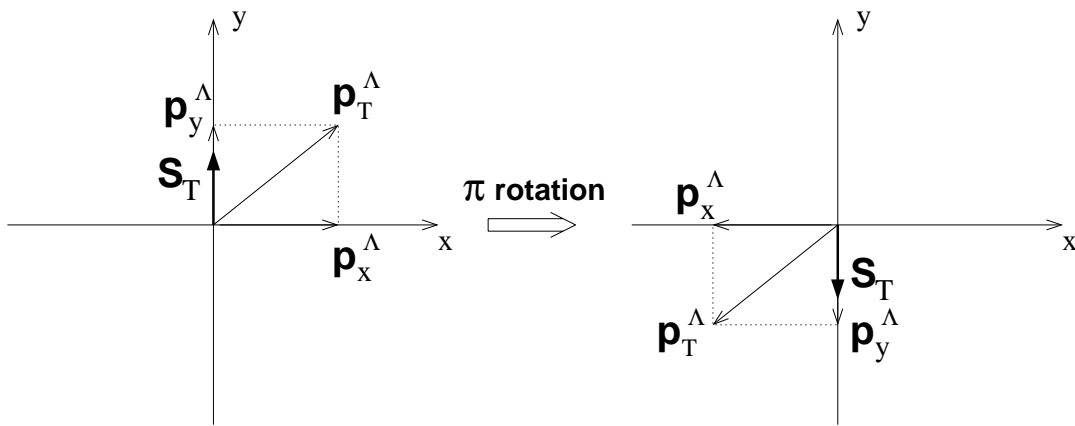


Figure 8.4: Invariance of the Λ production cross-section for a rotation of π radians around the beam axis. S_T is the direction of the target spin, p_T^Λ is the Λ transverse momentum, and p_x^Λ and p_y^Λ are the x and y components of the Λ momentum.

pendicular to the direction of the muon beam. The effect of the dipole field is to bend the beam by a mean angle of about 3 mrad on average. Therefore the beam line has to be adjusted accordingly; the beam enters the polarized target with an angle, such that on average the beam muons are directed along the z axis at the target center.

The dipole field deviates the scattered muons and the charged particles produced in the DIS as well. An asymmetry with respect to the beam axis is therefore expected, since the dipole field deviates the charged particles systematically in the same direction. The effect on the scattered muons is expected to be small, at least for small Q^2 values. On the contrary the effect can be significant for low-momenta hadrons, which can be stopped by the outer structure of the polarized target.

The Λ hyperons are detected from the $\Lambda \rightarrow p\pi^-$ decay products. The momentum of the decay pion strongly depends on the angular distribution of the decay process: pions emitted collinear to the Λ flight path and backward in the Λ center of mass frame can have very low momenta, and most of them do not pass through the target exit window. Moreover, π^- are systematically deviated on the left with respect to the beam axis, so that a left/right asymmetry on the Λ reconstruction efficiency is expected.

The situation is different in the case of longitudinal polarized target. The solenoid field, although intense (~ 2 Tesla), is with good approximation directed along the beam axis, and almost collinear with all the charged particles created in the interaction with the target. For this reason the deviation due to the solenoid field is small, and symmetric with respect to the beam axis.

In order to investigate the importance of these effects, I produced a big amount of MonteCarlo events both with the longitudinal polarization configuration and with the transverse one.

The Λ events have been produced using simulation software described in Chapter 5. To increase the statistics, the data set has been “artificially enriched” by keeping only events with at least one Λ hyperon in the final state. This includes both directly produced and secondary Λ . The simulated DIS pro-

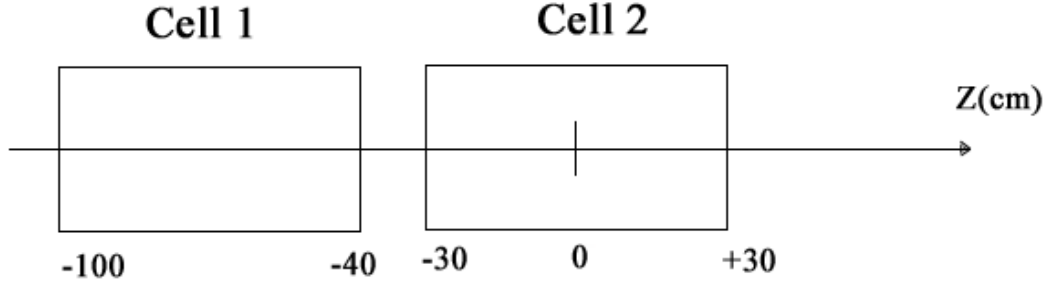


Figure 8.5: Position of the target cell along the z axis in the Main Reference System.

cess does not include polarization effects and allows the study of experimental apparatus contributions to the measured spin observables.

8.4 Event selection

In the analysis Λ particles are identified from their weak decay $\Lambda \rightarrow p\pi^-$, 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 deviated in opposite direction by the spectrometer magnets. The event selection is performed applying the following conditions to the data sets:

- the selection criteria of the Λ is mainly based on the requirement that the primary vertex is inside one of the two target cells.
- the primary vertex must be completely reconstructed (beam muon and scattered muon with associated momenta), plus at least one secondary V^0 vertex.

Such sample contains a small fraction of Λ vertices, and some additional cuts are required in order to reject the background sources and clean the Λ signal.

8.4.1 K_S^0 mesons identification

Also K_S^0 mesons are, as Λ , neutral particles decaying with a V^0 vertex. K_S^0 mesons have a larger production cross-section, and decay in $K_S^0 \rightarrow \pi^+\pi^-$ with a branching ratio of $\sim 68\%$. Thid particles if not indentified and rejected contribute to the background of the Λ invariant mass distribution.

The most effective way to discriminate between K_S^0 and Λ decays would be to identify the positive decay particle using RICH information on the particle mass. K_S^0 and Λ can also be separated using their different lifetimes. The $c\tau$ constant is 2.6762 cm for K_S^0 and 7.89 cm for Λ , and the distribution of the distance between the production and decay vertices is different in the two cases. In our analysis a cut of 35 cm minimum has been applied.

8.4.2 Primary and secondary vertex selection

Fig.8.5 shows the position of the two target cell, along the z axis. The two cells are cylinders with 60 cm length and 1.5 cm of radius. The requirements that the primary vertex must be inside the target cells can be written as:

$$-100 < z_{pri} < -40, -30 < z_{pri} < 30 \quad (8.21)$$

and its distance from the axis of the cell must be $< r_{cuts} = 1.4\text{cm}$:

$$\sqrt{x_{pri}^2 + y_{pri}^2} < r_{cuts} \quad (8.22)$$

The secondary vertex V^0 must be downstream of the primary vertex:

$$z_V^0 - z_{pri} > -2\sqrt{\sigma_{x_{pri}}^2 + \sigma_{y_{V^0}}^2} \quad (8.23)$$

and the z_{V^0} coordinate must be less than 140 cm ($-100 < z_{V^0} < 140$).

Fig.8.6 shows the primary and secondary vertex distribution in the analized periods after the applied cuts.

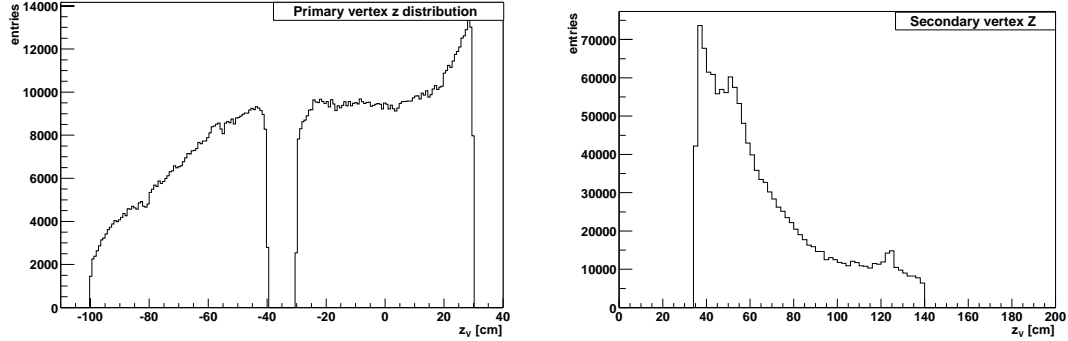


Figure 8.6: Primary (top) and secondary (bottom) vertex distribution.

8.4.3 Cuts on the decay product ($V^0 \rightarrow T^+T^-$)

- T^+T^- must completely pass through the magnet;
- 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, to not confuse with muons.

8.4.4 Other selection criteria

Additional selection criteria are the following:

- the path of the scattered muons must be 30 radiation lengths.
- Other V^0 vertices are originated by e^+e^- pairs from photon conversion. In this case the two electrons have a small transverse momentum, and the sum of their momenta has a small polar angle. 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 other vertices are rejected by applying a cut of $\theta > 0.01$ rad on the positive particle polar angle. These vertices, if not rejected, would lead to reconstructed Lambdas with too high momentum.

Another source of background is the presence of fake V^0 vertices. The vertex reconstruction package of the COMPASS reconstruction program does not apply a selection on fitted vertices: any combination of tracks which leads to fitted vertex with good χ^2 is maintained for further analysis. However, only some combinations correspond to real vertices, while many of them are only due to the combinatorial background. The rejection of fake vertices is mainly based on the request that the reconstructed neutral particle points to the primary vertex. This is done by calculating the angle between the particle direction and the line connecting the two vertices. Events are accepted only if the cosine of this angle is higher than 0.9999. All vertices surviving the cuts are then selected on a χ^2 basis.

The Λ vertex selection criteria have been checked in the MonteCarlo simulations to get numerical estimates of the rejection efficiencies. The MonteCarlo information allows to discriminate between “real” and “fake” secondary vertices, and to evaluate the contamination due to combinatorial background.

These criteria, validated with the MonteCarlo analysis, have been applied to real data to extract the transverse polarisation.

8.5 Results of the analysis of MonteCarlo events

In this section the results of the analysis of the MonteCarlo events are reported. I submitted on the GRID 5350 jobs for the transverse spin configuration. The number of generated events is about 10^8 . The result of the analysis of $5.2 \cdot 10^6$ events generated by 2600 GRID jobs is shown below. $1.3 \cdot 10^5$ events of the sample have been reconstructed correctly, with an overall efficiency of 0.25%.

The invariant mass of the sample passing all the selection cuts is shown in Fig.8.7 in the Λ hypothesis. To increase the Λ statistics the data set has been “artificially enriched” by keeping only events with at least one Λ hyperon in the final state. This includes both directly produced and secondary Λ . For this reason the combinatorial background is very low. The simulated DIS process does not include polarization effects and allows the study of experimental apparatus

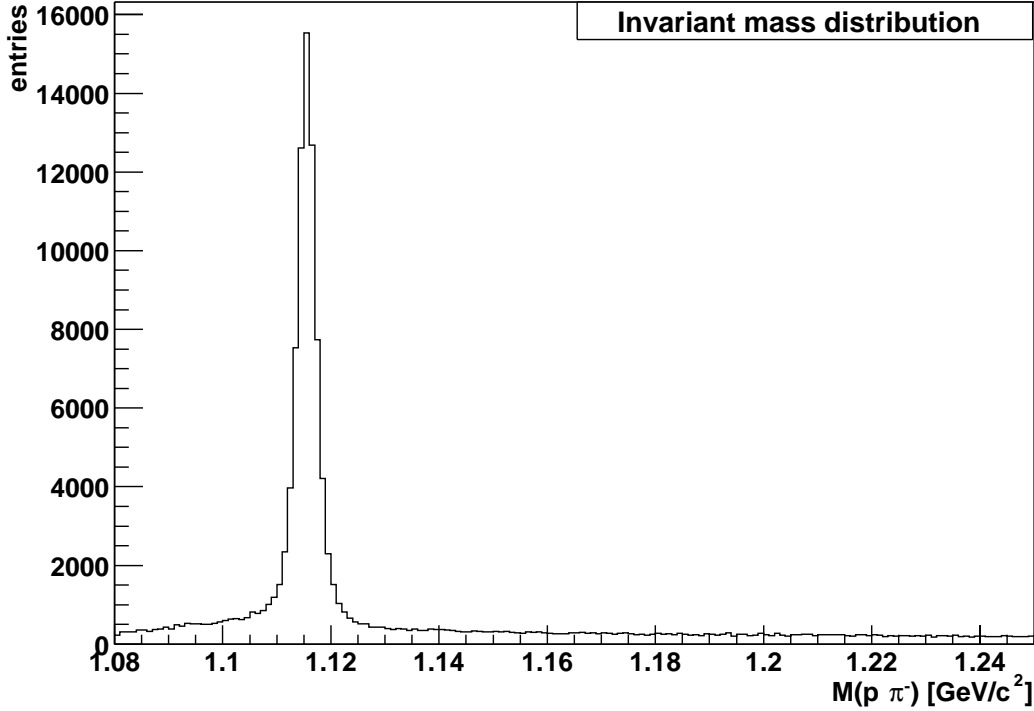


Figure 8.7: Λ invariant mass distribution after all the selection cuts for the MC sample.

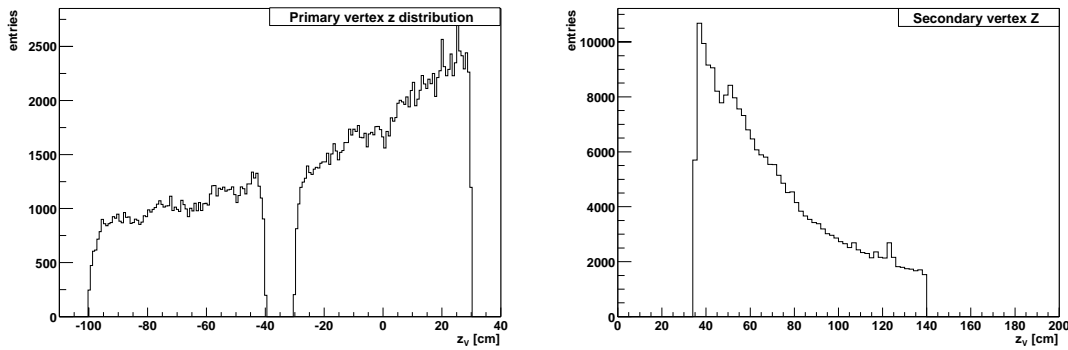


Figure 8.8: Primary (left) and secondary (right) vertex distribution

contributions to the measured spin observables.

Fig.8.8 shows the primary and secondary vertex distribution in the analyzed sample of data after the cuts. In Fig.8.9 the plot of Feynman x_F and the reconstructed Λ momentum p_Λ are given.

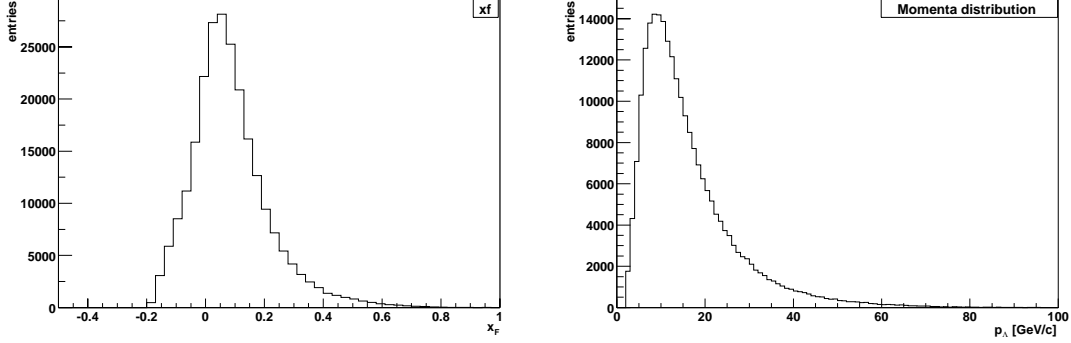


Figure 8.9: Feynman x_F (left) and reconstructed Λ momentum p_Λ (right) for MC sample.

The correction of the apparatus acceptance for the Λ polarization analysis has been verified with the MonteCarlo simulation], in which an unpolarized Λ sample has been generated. The obtained polarization is 0.006 ± 0.012 and is compatible with zero within error bars, with a statistical accuracy almost 10 times better than that of each of the data taking periods. Therefore I conclude that the effect of the non-ideal apparatus acceptance (see Fig.8.10) is well corrected by the algorithm uses to extract the polarization (see Eq.8.15).

8.6 Results of transverse polarization analysis

In this section the preliminary results for the transverse polarisation P_S of the Λ are reported.

The invariant mass of the sample passing all the selection cuts is shown in Fig.8.11 in the Λ hypothesis. The sample shown refers to one of the set included in the present analysis. The plot shows the difference between the reconstructed invariant mass of the $p\pi^-$ -system and the Λ PDG mass value. The peak is centered at zero as expected.

In Fig.8.12 the Armenteros plot for the same statistics is shown. The plot is a graphical representation of the neutral particle decay kinematics, and allows to distinguish Λ , $\bar{\Lambda}$ and K_0 particles according to the value of the Armenteros

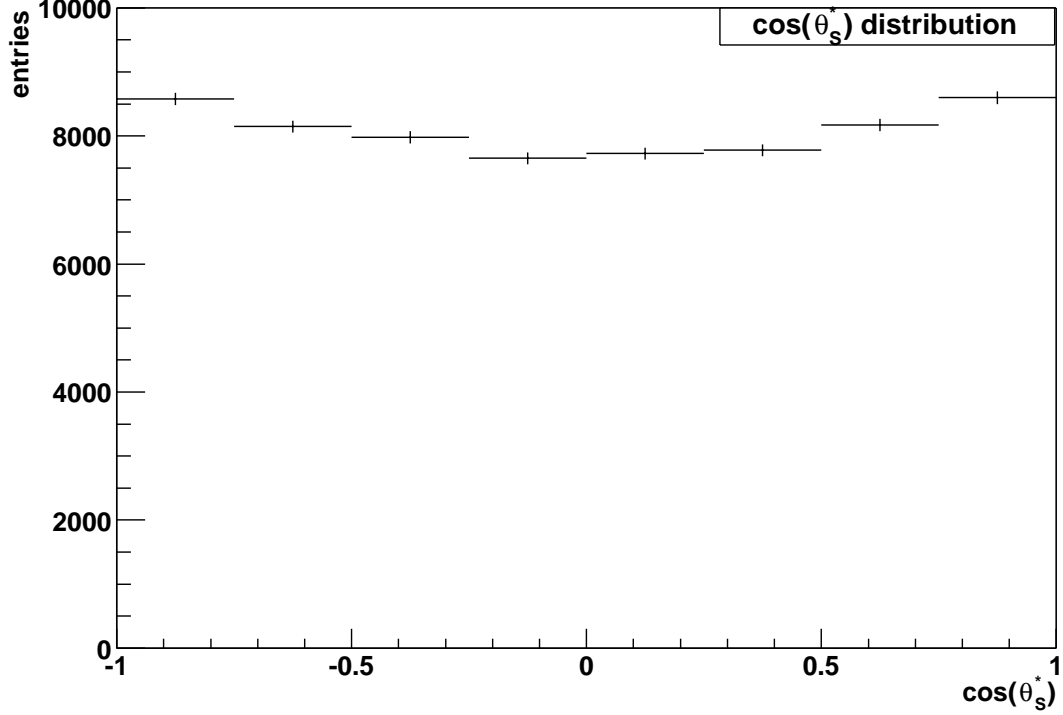


Figure 8.10: $\cos\theta$ distribution for MC sample

variable α , defined as

$$\alpha = \frac{p_L^{h^+} - p_L^{h^-}}{p_L^{h^+} + p_L^{h^-}}, \quad (8.24)$$

where $p_L^{h^{+(-)}}$ is the component of the positive (negative) decay particle momentum along the direction of the neutral decaying particle. When the transverse momentum of the positive decay hadron $p_T^{h^+}$ is plotted against α , the signature of neutral decaying particles are semi-ellipses in the $(\alpha, p_T^{h^+})$ plane.

The value of transverse polarisation for the three analyzed periods is shown in Fig.8.13.

The polarization has been extracted using two different methods to evaluate the number of reconstructed lambdas (background subtraction). The first method is based on a fit of the invariant mass distribution with a gaussian plus a polynomial function for the background parametrization. The second method

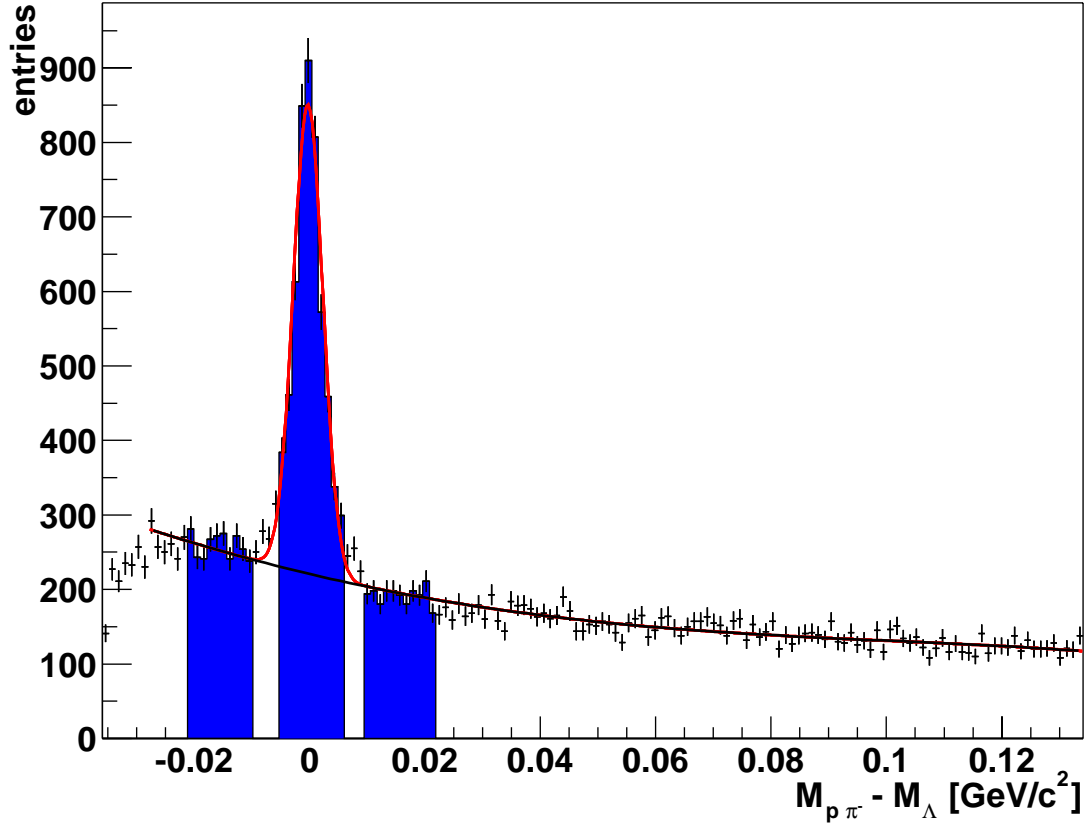


Figure 8.11: Λ invariant mass of the sample passing all the selection cuts.

(*sidebins subtraction method - ss*) uses the integral of the invariant mass in a window of 2σ (where σ is $2.6\text{MeV}/c^2$) on the left and right of the mass peak to estimate the amount of background below the peak. This second method can be safely applied only if the background is not polarized, and should give results consistent with the first one. The value of the Λ polarization measured with both methods and the background polarization measured with the second method are shown in Fig.8.13.

For the whole sample analyzed, the values for Λ and background are compatible with zero. They differ from zero by one σ (1.5σ for P2B.P2C period).

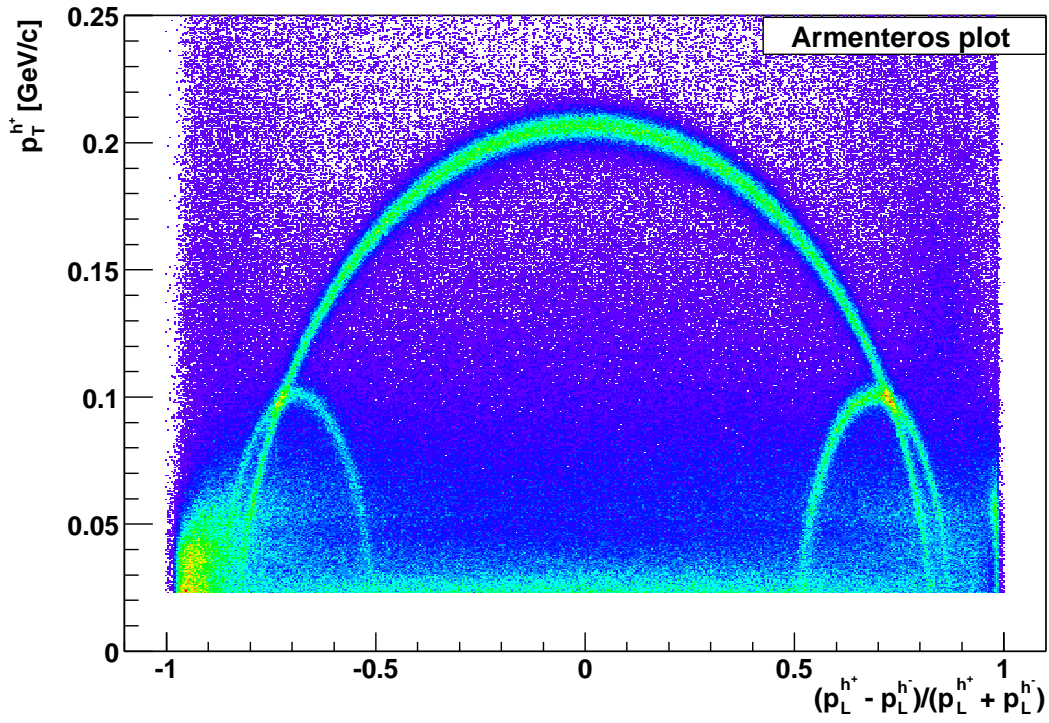


Figure 8.12: Armenteros plot. The K^0 , Λ and $\bar{\Lambda}$ decays are represented by semi-ellipses in the $(p_T^{h^+}, \alpha)$ plane, where $p_T^{h^+}$ is the transverse momentum of the positive decay product with respect to the decaying neutral particle direction, and α is the Armenteros variable.

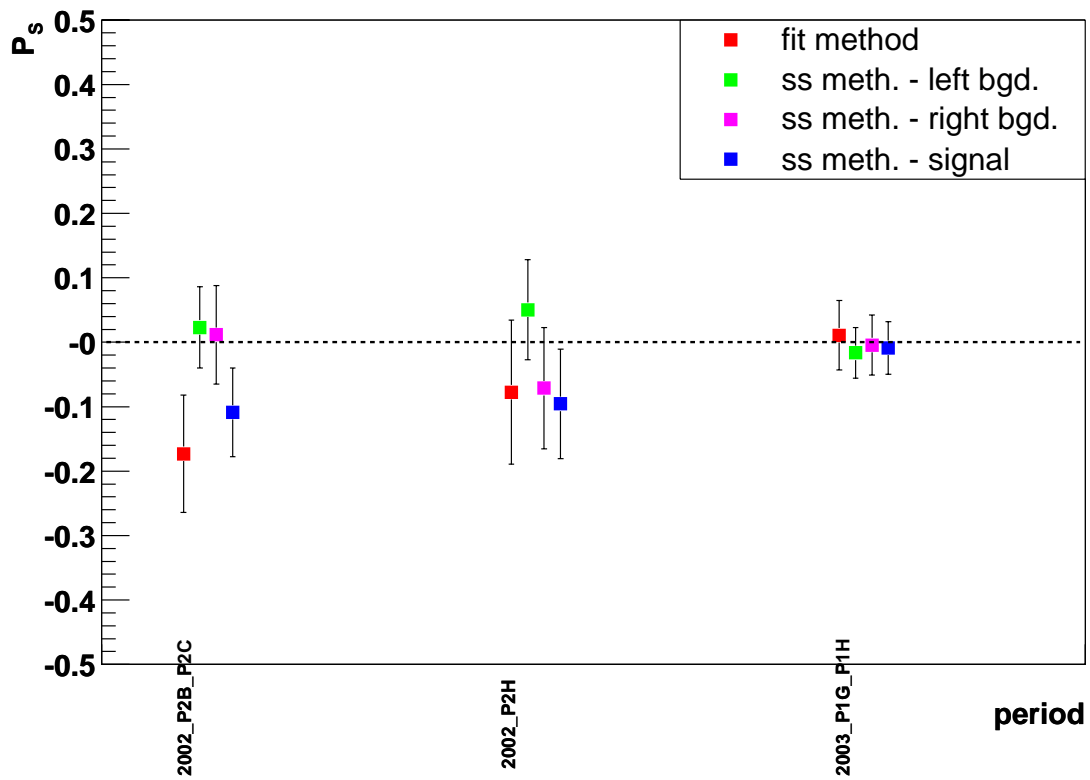


Figure 8.13: polarisation of the Λ for the period $P2B_P2C$, $P2H(1_2)$ and $P1H_P1G$.

Appendix A

A.1 GRID acronyms

Below is report the meant of the acronyms used in chapter 4:

API: Application Programming Interface

BDII: Berkeley Database Information Index

CE: Computing Element

CERN: European Laboratory for Particle Physics

ClassAd: Classified advertisement

CLI: Command Line Interface

CNAF: INFNs National Center for Telematics and Informatics

DIT: Directory Information Tree

DN: Distinguished Name (LDAPs)

EDG: European DataGrid

EDT: European DataTag

FNAL: Fermi National Accelerator Laboratory

GIIS: Grid Index Information Server

GLUE: Grid Laboratory for a Uniform Environment

GRAM: Globus Resource Allocation Manager

GRIS: Grid Resource Information Service

GSI: Grid Security Infrastructure

GUI: Graphical User Interface

GUID: Grid Unique ID

ID: Identifier

INFN: Istituto Nazionale di Fisica Nucleare

IS: Information Service

JCS: Job Control Service

JDL: Job Description Language

LB: Logging and Bookkeeping Service

LDAP: Lightweight Directory Access Protocol

LFN: Local File Name

LRC: Local Replica Catalog

LRMS: Local Resource Management System

LSF: Load Sharing Facility

MDS: Monitoring and Discovery Service

MPI: Message Passing Interface

MSS: Mass Storage System

NS: Network Server

OS: Operating System

PBS: Portable Batch System

PFN: Physical File name

PID: Process Identifier

POOL: Pool of Persistent Objects for LHC

RAL: Rutherford Appleton Laboratory

RB: Resource Broker

RLI: Replica Location Index

RLS: Replica Location Service

RM: Replica Manager

RMC: Replica Metadata Catalog

RMS: Replica Management System

ROS: Replica Optimization Service

SASL: Simple Authorization and Security Layer (LDAP)

SE: Storage Element

SMP: Symmetric Multi Processor

SRM: Storage Resource Manager
SURL: Storage URL
TURL: Transport URL
UI: User Interface
URI: Uniform Resource Identifier
URL: Universal Resource Locator
UUID: Universal Unique ID
VDT: Virtual Data Toolkit
VO: Virtual Organisation
WMS: Workload Management System
WN: Worker Node
WPn: Work Package number

A.2 Submit cron

Below the cron used to submit the MC simulation jobs on the GRID resources is reported.

```
#!/bin/bash

export PATH=/opt/lcg/bin:/opt/globus/bin:/opt/globus/sbin:/opt/edg/bin:
/usr/local/bin:/bin:/usr/bin:/usr/local/bin:/usr/bin/X11:/usr/X11R6/bin:
/opt/gpt/sbin:/usr/java/j2sdk1.4.1_01/bin:/opt/edg/bin:/opt/edg/sbin:
/opt/edg/bin:/opt/edg/sbin:/home/prod-user/bin
export EDG_WL_LOCATION=/opt/edg

TMP_DIR=/old-home/prod-user/lambda-trans/tmp          <--| directories
OUTPUT_DIR=/old-home/prod-user/lambda-trans/output/  <--| where are
OLD_DIR=/old-home/prod-user/lambda-trans/old         <--| stored JDL and
JDL_DIR=/old-home/prod-user/lambda-trans             <--| .log file
JOB_FILE=JOB.LNL.*
```

```

JFN=JOB.LNL
#CONF_FILE=gridit.conf          <----| .conf file to
#CONF_FILE=katane.conf          <----| select one or
CONF_FILE=more-dteam.conf       <----| more RB
##CONF_FILE=senza003-more-dteam.conf <----|
#CONF_FILE=egee.conf           <----|
#MAX_RUN=10
MAX_RUN=51  <-----| number of maximum jobs run
RUNNING=0
COM_STATUS=edg-job-status
VERBOSITY="-v 0"

echo "-----"
grid-proxy-info -e -h 24      <---| check of the proxy status.
if [ $? -gt 0 ]              <---| If the proxy expired the cron
then                          <---| stops the submitting procedure
    echo "no good proxy"     <---|
    exit                     <---|
fi

SubmitJob() {
NUM='date +%Y_%m_%d-%H_%M_%S'
NUM='date +%s'
NOW='/bin/date'
cd $JDL_DIR
/bin/echo "$NOW - sottometto $JFN.$NUM"
edg-job-submit -o tmp/$JFN.$NUM --config-vo $CONF_FILE
                run-20000-lambda_trans-LNL.jdl <-----| submission command
sleep 10
}

```

```

getOutput() {
echo "$JOBID GET OUTPUT from $FILE"
edg-job-get-output -dir $OUTPUT_DIR -i $FILE
mv $FILE $OLD_DIR
}

checkJob() {                                <-----| Start of the routine that
cd $TMP_DIR                                  <-----| checks the jobs status
for FILE in  $TMP_DIR/$JOB_FILE* <-----|
do
  NOW='date +%Y-%m-%d-%H-%M-%S'
#  echo "$COM_STATUS $VERBOSITY -i $FILE | grep Current |
      sed 's/Current Status:/' | sed 's/ //g'"
#  RET1='$COM_STATUS $VERBOSITY -i $FILE'
#  echo "RET1 = $RET1"
  RET='$COM_STATUS $VERBOSITY -i $FILE | grep Current |
      sed 's/Current Status:/' | sed 's/ //g'
#  echo "RET = $RET"
  echo -n $NOW $FILE $RET
  case $RET in
    Running)
      echo "  OK is running"
      ;;
    "Done(Success)")
      echo "  OK finished"
      JOBID='grep -h https $FILE'
      getOutput
      ;;
    Ready)
      echo "  OK Ready ---> do nothing"
      ;;
  esac
done
}

```

```

Scheduled)
    echo "  OK Scheduled ---> do nothing"
    ;;
*)
    echo BOH eheheh.....
    ;;
esac
done
}

checkJob          <----| This routine check the
                  <----| number of running jobs
cd $TMP_DIR       <----| and the number of
for file in $TMP_DIR/$JOB_FILE* <----| available slots
do
    if [ -f $file ]; then
        RUNNING='expr $RUNNING + 1'

    fi

done
echo "RUNNING=$RUNNING"
TO_RUN='expr $MAX_RUN - $RUNNING'
if [ $TO_RUN -gt 0 ]
then
    echo "Slot Available = $TO_RUN ... submitting " <---| If there are
else                                           <---| available slot
    echo "NO Slot Available ==> NO submitting " <---| go to submit
fi                                           <---| routine, if no
LOOP=0                                       <---| wait for
while [ $LOOP -lt $TO_RUN ]                 <---| available

```



```
do                                     <---| slots
  SubmitJob
  LOOP='expr $LOOP + 1'
done
```

A.3 JDL file for Λ transverse production

```
Executable = "run-20000-lambda_trans.sh"; <----| Shell script name.
                                                This .sh program
                                                starts the
                                                simulation software
```

```
StdOutput = "run-20000-lambda_trans.out"; <----| .out and .err file
StdError = "run-20000-lambda_trans.err"; <----| name. They will be
                                                saved in the UI in
                                                dedicated directory.
```

```
InputSandbox = { "run-20000-lambda_trans.sh"};
OutputSandbox = { "run-20000-lambda_trans.err" ,
                  "run-20000-lambda_trans.out"};
Requirements = Member("VO-alice-cmp-3.1.2", <-----|requirements
other.GlueHostApplicationSoftwareRunTimeEnvironment) && ( on RunTimeEnv
other.GlueCEPolicyMaxCPUTime > 1400 ) && Member("LNL", and CPU speed
other.GlueHostApplicationSoftwareRunTimeEnvironment);
```

A.4 Simulation software installation script

Below is reported the script used for the installation of the software for the longitudinal events simulation. The script for the install the software for transversal events production is similar. The difference is on the name of the file and on the directory name.

```
#!/bin/bash
###
RANGE=100

CAS=$RANDOM
let "CAS %= $RANGE"
NUM_CAS=$(printf "%02d" $CAS)
echo "NUM CAS = $NUM_CAS"

NUM='date +%Y_%m_%d-%H_%M_%S'-$NUM_CAS
echo "argomento = $NUM"

FQDN='hostname -f'
echo "Running on HOST $FQDN"
DOMAIN='hostname -d'
echo "DOMAIN= $DOMAIN"

#OUTDIR=/flatfiles/SE00/dteam
OUTDIR=$HOME/tmp/pythia_$NUM

SCRDIR=/opt/exp_software/alice/compass/lambda

echo "OUT_DIR $OUTDIR"

#####INIZIO##EXTGEN#####

GENDIR=${SCRDIR}/phast/coral/simulazione
echo "GEN_DIR for generation: $GENDIR"

LOCALDIR=$HOME/tmp/pythia_$NUM
echo "LOCALDIR for generation: $LOCALDIR"
```

```
echo "# creating dir $LOCALDIR"
mkdir -p $LOCALDIR
echo "# cwd $LOCALDIR"
cd $LOCALDIR
WD='pwd'
echo "# wd is now $WD"

ln -s $GENDIR/run.com run.com
echo "# creating sym link run.com ==> $GENDIR/run.com"
ln -s $GENDIR/pythia.exe pythia.exe
echo "# creating sym link pythia.exe ==> $GENDIR/pythia.exe"

ln -s $GENDIR/pythia_30000.opt pythia.opt
echo "# creating sym link pythia.opt ==> $GENDIR/pythia_30000.opt"

echo "# executing run.com pythia pythia.opt"
./run.com pythia pythia.opt

#####INIZIO##COMGEANT#####

HOMEDIR=${SCRDIR}/phast/coral/simulazione
echo "# HOMEDIR=$HOMEDIR"

echo "# creating sym link geant.exe ==> $HOMEDIR/geant.exe"
ln -s $HOMEDIR/geant.exe geant.exe

echo "# creating sym link fort.15 ==> $HOMEDIR/main_muon_071_30000.ffr"
ln -s $HOMEDIR/main_muon_071_30000.ffr fort.15
```

```
echo "# creating sym link fort.16 ==> $HOMEDIR/fort.16"
ln -s $HOMEDIR/fort.16 fort.16
echo "# creating sym link fort.18 ==> $HOMEDIR/fort.18"
ln -s $HOMEDIR/fort.18 fort.18
echo "# creating sym link fort.19 ==> $HOMEDIR/fort.19"
ln -s $HOMEDIR/fort.19 fort.19
echo "# creating sym link fort.20 ==> $HOMEDIR/fort.20"
ln -s $HOMEDIR/fort.20 fort.20
echo "# creating sym link fort.21 ==> $HOMEDIR/fort.21"
ln -s $HOMEDIR/fort.21 fort.21
echo "# creating sym link fort.22 ==> $HOMEDIR/fort.22"
ln -s $HOMEDIR/fort.22 fort.22
echo "# creating new empty file zebradat.1.fz"
touch zebradat.1.fz
echo "# creating sym link fort.32 zebradat.1.fz"
ln -s zebradat.1.fz fort.32

LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/usr/local.nis/lib:
        $SCRDIR/SHIFT_20021210/lib:
        $SCRDIR/prod-200303/coral/lib/Linux
echo "\$LD_LIBRARY_PATH = $LD_LIBRARY_PATH"

#export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/usr/local.nis/lib:
        $SCRDIR/SHIFT_20021210/lib

export LD_LIBRARY_PATH
echo $LD_LIBRARY_PATH

echo "# executing geant.exe"
./geant.exe
```

```
echo "# renaming csimout.nt PythiaL_${NUM}.nt"
mv csimout.nt PythiaL_${NUM}.nt
echo "# copy zebradat.1.fz PythiaL_${NUM}.fz"
cp zebradat.1.fz PythiaL_${NUM}.fz

#####
# start of PHAST
#####
CORALDIR=${SCRDIR}/coral/
echo "CORAL_DIR = $CORALDIR"

echo "# exporting env HOMEDIR_CORAL = $LOCALDIR"
export HOMEDIR_CORAL=$LOCALDIR

if [ -d cards ]
then
echo "# the directory cards exists"
else
echo "# the directory cards DOES NOT exists"
echo "# creating dir cards"
mkdir cards
fi

# entries to be analysed
# stored to disk
echo "# writing ./cards/eventtoread"
echo 'events to read 10000'>./cards/eventtoread
echo "# writing ./cards/eventtoskip"
echo 'events to skip 0'>./cards/eventtoskip
echo "detector table $LOCALDIR/detectors.dat">./cards/detfile
```

```
ls -l $LOCALDIR/detectors.dat
# zebra input
# stored to disk

echo "# writing ./cards/opt_zebra"
echo "Monte Carlo file ${LOCALDIR}/PythiaL_${NUM}.fz">./cards/opt_zebra

SAVEDIR=$LOCALDIR
# output files
OUTFILE=${SAVEDIR}/trafdic.mc_${NUM}.root
echo "# OUTFILE = $OUTFILE"
echo "# writing ./cards/opt_outfile"
echo "histograms home ${OUTFILE}" > ./cards/opt_outfile
echo "# adding ./cards/opt_outfile"
echo "mDST   file   mDST_mc_pyl.${NUM}.root" >>./cards/opt_outfile
echo "# adding ./cards/opt_outfile"
echo 'mDST   selection  0' >> ./cards/opt_outfile

#coral option file
CORAL_OPT=${SCRDIR}/phast/coral/trafdic.mc.opt
echo "# CORAL_OPT = $CORAL_OPT"
if [ -r $CORAL_OPT ]
    then
        echo "# the file $CORAL_OPT exists readable"
    else
        echo "# Cannot open $CORAL_OPT "
        echo "error"
    fi

echo "# creating sym link $LOCALDIR/coral.exe ==>
```

```
    $SCRDIR/phast/coral/coral.exe"
ln -s $SCRDIR/phast/coral/coral.exe $LOCALDIR/coral.exe
echo $?

echo "# ~/.bashrc"
. ~/.bashrc
echo "# unset CORAL"
unset CORAL
echo "# cd $CORALDIR"
cd $CORALDIR
echo "# . setup.sh"
. setup.sh
echo "exporting HOMEDIR_CORAL = $LOCALDIR"
export HOMEDIR_CORAL=$LOCALDIR
echo "# cwd $LOCALDIR"
cd $LOCALDIR
pwd
echo "# ls -altr"
ls -altr
echo "# setting envvar $localdir"
export localdir=$LOCALDIR
echo "localdir=$localdir"
echo "# setting envvar $num"
export num=$NUM
echo "num=$num"

echo "# time ./coral.exe $CORAL_OPT"

export localdir=$LOCALDIR; export num=$NUM;
    time ./coral.exe $CORAL_OPT
```

```
#####INIZIO##TRASFERIMENTO#####

# transferring back output results

# trasferimento mDST_mc_*.root

echo "# searching for mDST_mc_pyl.$NUM.root"
find ~ -name "mDST_mc_pyl*"

#echo "# copying ./mDST_mc_pyl.$NUM.root $OUTDIR"
#cp ./mDST_mc_pyl.$NUM.root $OUTDIR
echo "# transferring out-file via gridftp"

globus-url-copy file://`pwd`/mDST_mc_pyl.$NUM.root <--| The output
                gsiftp://grid009.to.infn.it/ <--| file is
                /flatfiles/SE00/dteam/long_prod <--| transferred
                /mDST_mc_pyl.$DOMAIN.$NUM.root <--| via gsiftp

echo "# cleaning $LOCALDIR : rm -Rf $LOCALDIR "
rm -Rf $LOCALDIR

exit 0
```

Bibliography

- [1] COMPASS, A Proposal for a COmmon Muon and Proton Apparatus for Structure and Spectroscopy, (CERN/SPSLC 96-14, SPSLC/P297, 1996).
- [2] E.W. Hughes and R. Voss, *Annu. Rev. Nucl. Part. Sci.* **49** (1999), 303
- [3] J. Ellis and R.L. Jaffe, *Phys. Rev.* **D 9** (1974), 1444; *Phys. Rev.* **D 10** (1974), 1669
- [4] M. Vetterli, **hep-ph/9812420**.
Prepared for “Lake Louise Winter Institute: Quantum Chromodynamics”,
Lake Louise, Alberta, Canada, 15-21 Feb 1998.
In “Lake Louise 1998, Quantum chromodynamics”, 49-98.
- [5] J. Ashman *et al.*, *Nucl. Phys.* **B328** (1989), 1
- [6] A. Bacchetta, PhD thesis.
- [7] M. Alberg *et al.*, *Phys. Lett.* **B356** (1995), 113
- [8] J. Ellis *et al.*, *Nucl. Phys* **A673** (2000), 256
- [9] J. Ellis, D. Kharzeev and A. Kotzinian, *Z. Phys.* **C69** (1996), 467
- [10] M. Anselmino *et al.*, *Phys. Lett.* **B 481** (2000), 253
- [11] D. de Florina *et al.*, *Phys. Lett.* **B439** (1998) no.1/2, 176
- [12] J. Fischer *et al.*: *Nucl. Instr. Meth. Phys. Res.* **A238** (1985), 248

[13] LEPTO Documentation page: <http://www3.tsl.uu.se/thep/lepto/>

[14] GEANT Documentation page: <http://wwwinfo.cern.ch/asd/geant/>

[15] CORAL Documentation page: <http://coral.cern.ch>

[16] PHAST Documentation page: <http://gs.home.cern.ch/ges/phast/>

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

[18] INFN GRID Project Home Page: <http://grid.infn.it/>

[19] LHC Computing Grid Project Home Page: <http://lcg.web.cern.ch/LCG/>