Nucleon structure studies with CLAS12 at Jefferson Lab : timelike Compton scattering and the central neutron detector

Pierre Chatagnon

université
PARIS-SACLAY

Nucleon Structure studies with CLAS12 at Jefferson Lab: Timelike Compton Scattering and the Central Neutron Detector

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 576, Particules, Hadrons, Énergie et Noyau : Instrumentation, Image, Cosmos et Simulation (Pheniics)
Spécialité de doctorat: Physique hadronique
Unité de recherche: Université Paris-Saclay, CNRS, IJCLab, 91405, Orsay, France
Référent: Faculté des sciences d'Orsay

Thèse présentée et soutenue à Orsay, le 15 Octobre 2020, par

Pierre CHATAGNON

Composition du jury:

Michel Guidal Directeur de recherche, Université Paris-Saclay	Président
Stepan Stepanyan Professeur, Old Dominion University	Rapporteur & Examinateur
Marc Vanderhaeghen Professeur, Johannes Gutenberg-Universitaet Mainz	Rapporteur & Examinateur
Raffaella De Vita Professeur, INFN Genova	Examinatrice
Fabrizio Ferro Professeur, INFN Genova	Examinateur
Bernard Pire Directeur de recherche, École Polytechnique	Examinateur
Silvia Niccolai Directrice de recherche, Université Paris-Saclay	Directrice de thèse

Thèse de doctorat

NNT: 2020UPASP039

À Papa, Maman, Thomas et Julie

Acknowledgments

This manuscript is the result of three intense years of work, meetings, discussions, conferences, travels, moments of unforgettable joy and moments of heavy stress, and much much more. The journey started by a Skype call between Cambridge and Orsay in January 2017. Since then many people have made my PhD adventure possible. But before, I have to apologize for all these I will forget to include in this page: Thank you a lot !

This adventure was only possible thanks to Silvia Niccolai who have supervised my work, provided me with invaluable advice and constant support in all the good and bad times of these three years. Thank you for your quasi-24/7 availability, our countless meetings, mails and Whatsapp messages. I could always rely on your help no matter the time zone and no matter the issue. Thank you also for the trust you have placed in me, from the very start of the thesis; first to take part in the installation of "your" detector, and then to take in charge the important task of its calibration. Your early and total confidence in my work have been very rewarding and encouraging for the rest of the thesis. I am indebted to have worked with you, and very glad to consider you as a mentor and friend. *Grazie !*

The three years I have spent at IPN would not have been the same without my colleagues and friends. Thanks to Raphaël Dupré, Carlos Muños Camacho, Dominique Marchand, Eric Voutier and Gabriel Charles whose comments and discussions were a real source of improvement for my work. A special thanks to Mostafa Hoballah (aka Adam le Magicien) for your willingness to share your seemingly infinite knowledge on new data analysis techniques and coffee. I owe Michel Guidal for his help on the theory side of my analysis. Your inputs have been crucial. Especially, I am very grateful for your help in organizing my trip to Mainz, as well as being part of my jury as *président*. Finally I am thankful to all the postdocs and students for the delightful times spent with them: Victoria, Carolina, PoJu, Rong, Sami, Sizar Aziz, Chun Lu, and Lucien.

I also want to thank my dear office collaborators, Mathieu Ehrhart and *Docteur* Ho San Ko. Our work on ultra-relativistic rotating *flageolets* will provide, without a doubt, a *vérie goude* framework for future collaborations. I am grateful for our after work meetings, usually discussing about the deep implications of our theory around a glass of apple juice. I shall never forget these moments, although I am still not fully sure they are *haricots*. Of course, I shall not forget the late addition to our collaboration, in the person of Mylène Caudron. You have provided new insights and perspectives to the collaboration, and I bet it shall be recognized too. Especially, your knowledge on apple juice shall now be acknowledged even on the other side of the corridor.

Many thanks are also due to all the CLAS collaboration and JLab members that accompanied my work, their precious inputs made this work possible. I acknowledge the indispensable help of the *ee* analysis group: Valery Kubarovsky, Pawel Nadel-Turonski, Nathan Baltzell. Especially, I want to thank Stepan Stepanyan, who have greatly assisted me in the analysis part of this manuscript. I want to thanks you for your sharp remarks and your constant support. Your guidance was invaluable during these three years. Thanks also for being *rapporteur* of this work. I would also like to express my full recognition to Rafayel Paramuzyan for your precious advice and inputs on the TCS analysis. Also, I want to thank Raffaella De Vita, for your availability and our precious discussions about CLAS12, its software and the calibration suite. Many thanks for accepting to assess my work by being *examinatrice*.

Finally, I thank Marouen for our long discussions about physics (or not), Stuart for the pub quizzes at Tradition and for saving me (twice) from my canceled-flight curse, Joseph for your help on the analysis, Maxime Defurne, Maxime, Francesco Bossu, Guillaume, Marie and the rest of the French mafia, the Scottish mafia and especially Daria Sokhan and Paul, and all the other people who made my trips to NN very enjoyable moments.

This manuscript would not have been the same without the crucial discussions I had with Marc Vanderhaeghen in Mainz. This visit has been a pivotal point for my work, with the introduction to the concept of Forward/Backward asymmetry which turned to be a very clean and neat TCS observable. I am truly indebted to you for inviting me to Mainz. I also want to thank you for providing me with theoretical predictions to compare with my data points and for accepting to be *rapporteur* of this manuscript.

I am also very grateful to Bernard Pire, for our discussions on TCS and on TDAs which gave me great theoretical insights on my work but also beyond. I warmly thank you for accepting to be *examinateur* of my thesis. I am also thankful to Pawel Sznajder for the discussions on the PARTON software, and for providing me with model predictions for TCS.

Many thanks to Fabrizio Ferro for accepting to be *examinateur* of my thesis. I am delighted to work with you in the future.

Je veux aussi remercier toutes les personnes en dehors de l'IPN avec qui j'ai pu partager de bons moments ces trois dernières années. Merci Guillaume et la famille Marcelin. Merci Thomas pour nos Flam's trimestriels. Merci à Julien, Sébastien et Mathieu pour l'aventure arctique; les -20° resterons inoubliables ! Merci à Nicolas pour nos rituels orléanais resto indien/buzzer beater sur NBA 2K, qui m'ont permis de relâcher la pression de la troisième année.

Glòria, molts petons per tots els nostres moments viscuts junts durant aquests dos anys i mig, i pel teu amor malgrat la distància. Aprecio enormement la teva ajuda en la redacció del manuscrit i per la confecció de pastissets (i també panellets), i espero poder-te-la tornar l'any que ve (amb un pastís de xocolata) !

Enfin et surtout, Julie, Thomas, Papa et Maman: je vous aime !

Saint Jean de Touslas
22 octobre 2020

"Considérons les Anciens, les Grecs, les Indiens, les musulmans qui m'ont précédé, ils ont écrit abondamment dans toutes ces disciplines. Si je répète ce qu'ils ont dit, mon travail est superflu; si je les contredis, comme je suis constamment tenté de le faire, d'autres viendront après moi pour me contredire. Que restera-t-il demain des écrits des savants ? Seulement le mal qu'ils ont dit de ceux qui les ont précédés. On se souvient de ce qu'ils ont détruit dans la théorie des autres, mais ce qu'ils échafaudent eux-mêmes sera immanquablement détruit, ridiculisé même par ceux qui viendront après. Telle est la loi de la science; la poésie ne connaît pas pareille loi, elle ne nie jamais ce qui l'a précédée et n'est jamais niée par ce qui la suit, elle traverse les siècles en toute quiétude. C'est pour cela que j'écris mes robaïyat. Sais-tu ce qui me fascine dans les sciences ? C'est que j'y trouve la poésie suprême : dans les mathématiques, le grisant vertige des nombres; avec l'astronomie, l'énigmatique murmure de l'univers. Mais de grâce, qu'on ne me parle pas de vérité !"

— Amin Maalouf, *Samarcande*

Contents

List of acronyms

ADC	Analog to Digital Converter
BDT	Boosted Decision Tree
BH	Bethe-Heitler process
BSA	Beam Spin Asymmetry
CD	Central Detector
$CEBAF$	Continuous Electron Beam Accelerator Facility
CFD	Constant Fraction Discriminator
CFF	Compton Form Factor
$CLAS12$	CEBAF Large Acceptance Spectrometer 12
CND	Central Neutron Detector
COM	Center Of Mass
$CTOF$	Central Time Of Flight
CVT	Central Vertex Tracker
DAQ	Data Acquisition system
DC	Drift Chamber
DD	Double Distribution
DIS	Deep Inelastic Scattering
$DVCS$	Deeply Virtual Compton Scattering
$DVMP$	Deeply Virtual Meson Production
EB	Event Builder
EC	Electromagnetic Calorimeter
$ECIN$	Inner Calorimeter
$ECOUT$	Outer Calorimeter
$fADC$	Flash Analog to Digital Converter
FB	Forward-Backward
FD	Forward Detector
FF	Form Factor

FT	Forward Tagger
FTOF	Forward Time Of Flight
GEMC	Geant4 Monte Carlo
GPD	Generalized Parton Distribution
HTCC	High Threshold Cherenkov Counter
HV	High Voltage
LO	Leading Order
LTCC	Low Threshold Cherenkov Counter
MIP	Minimum Ionizing Particle
MLP	Multi Layer Perceptron
MVA	Multivariate analysis
MVT	Micromegas Vertex Tracker
NLO	Next to Leading Order
PCAL	Pre-shower Calorimeter
PDF	Parton Distribution Function
PID	Particle Identification
PMT	Photo-Multiplier Tube
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics
RF	Radio Frequency
RGA	Run Group A
RICH	Ring Imaging Cherenkov
ROC	Receiver Operating Characteristic
SF	Sampling Fraction
SLC	Sector Layer Component
SVT	Silicon Vertex Tracker
TCS	Timelike Compton Scattering
TDC	Time to Digital Converter
TOF	Time Of Flight

Introduction

The vast majority of the mass, more than 99%, of the visible universe is formed by atoms. The mass of the atom is itself almost entirely concentrated in its nucleus. The current model of the atom, a heavy nucleus surrounded by a cloud of electrons, was developed after the first alpha-particle scattering experiment performed by Rutherford at the beginning of the 20th century. His work led to the discovery of the composite nature of the atom. The later discoveries of the nucleons, first the proton (Rutherford, 1919) and then the neutron (Chadwick, 1932), opened the way to the exploration of the structure of the nucleus as a complex system composed of nucleons. The evidence that the nucleons themselves are complex composite objects was provided in 1969 by the Deep Inelastic Scattering results published by the SLAC collaboration. After this discovery, the Parton Model was developed and later confirmed by the discovery of the J/Ψ in 1973. These discoveries form the foundation of the field of hadronic physics, which aims at describing the interactions of the fundamental constituents of matter, the quarks and gluons, inside the hadrons and, in particular, the nucleons.

Quantum chromodynamics (QCD) is the theory describing the interaction of colored objects, the partons, which interact via the strong force, mediated by gluons. The theory of QCD, although perturbatively calculable at high energy, reveals its complexity at low energy (of the order of the nucleon mass). Its coupling constant increases with decreasing energy, at the extent that the perturbative approach cannot be applied at low energy. This behavior, called asymptotic freedom, is the core of hadronic physics. Indeed, understanding the structure of the nucleon ultimately enables us to fully understand non-perturbative QCD.

As explicit calculations of low-energy QCD are not achievable, the main tools to understand hadrons are ad-hoc structure functions, encoding the complex behavior of the partons inside the hadrons. The first set of structure functions describing the nucleon structure was introduced in the 1950s. The Form Factors (FF) and the Parton Distribution Functions (PDF) allow to explore, respectively, the spatial and the momentum-related structure of the nucleons. FFs are of great importance to determine the radius of the nucleon, while PDFs are used to parameterize the partonic content of the nucleons in terms of momentum, which is essential in high-energy proton collision experiments. Furthermore, their generalization in terms of Generalized Parton Distributions (GPD) provides even more information on the fundamental properties of the nucleon, such as its spin or its mechanical properties. The GPDs were introduced in the late 1990s. Their rich phenomenology, including strong links to FFs and PDFs, but also completely new interpretations, such as their direct link to the nucleon spin, have driven a large international effort, both experimental and theoretical, aiming to measure them.

The experimental program of the CLAS12 detector is largely dedicated to hadronic physics measurements, and in particular the extraction of GPD observables. This large acceptance detector is housed in the experimental Hall B at Jefferson Lab, in Virginia, USA. Its large coverage makes it ideal to measure the Deeply Virtual Compton Scattering process (DVCS, $ep \to e'p'\gamma$), the experimental reaction the most sensitive to GPDs. The 11-GeV beam provided to CLAS12 by the upgraded Continuous Electron Beam (CEBAF) also allows to measure the DVCS time-reversal process, Timelike Compton Scattering (TCS, $\gamma p \to e^+e^-p'$), in the resonance-free region. This process plays a crucial role in our understanding of GPDs. First, it is the simplest reaction, besides DVCS, that can be parameterized by GPDs. Its measurement and the comparison with DVCS results can provide evidence for the univer-

sality of the GPD theoretical framework. In addition, TCS has a singular sensitivity to the real parts of the Compton Form Factors (CFFs), which are GPD-based quantities accessible in DVCS and TCS measurements. The real parts of the CFFs, that contain integrals of GPDs over the internal quark momentum fraction, have not yet been constrained by existing DVCS data. The measurement of TCS is expected to provide a deeper insight into it.

The work reported in this manuscript focuses on the experimental approach to GPDs, using the CLAS12 detector. Two independent tasks were carried out, and they are described in two separate parts of this manuscript. The first task involves the installation, development of the calibration and reconstruction software, and commissioning of the Central Neutron Detector of CLAS12. This detector, designed and built at Institut de Physique Nucléaire d'Orsay and dedicated to the detection of the recoil neutron in the nDVCS reaction ($ed \rightarrow e'n'(p')\gamma$), is a key element in the measurement of DVCS observables on the neutron. These observables are an essential step in the determination of the contribution of the valence-quarks angular momentum to the total spin of the nucleon. The second task consisted in the analysis of the CLAS12 data in order to extract TCS observables.

These two projects are presented within six chapters:

- Chapter 1 introduces the concepts and the theoretical tools needed for this work. It is composed of three sections. In the first section, the FFs, the PDFs, and the polarized structure functions are presented. Then the GPDs and their links to FFs and PDFs, as well as their multiple interpretations are discussed. The final section is dedicated to the experimental path to GPDs, via DVCS and TCS. The interest for TCS is emphasized and the relevant observables are presented.

- Chapter 2 describes the experimental setup used during this thesis, focusing first on the CEBAF accelerator, then describing in details the CLAS12 detector, its subsystems and the associated software.

- Chapter 3 focuses on the Central Neutron Detector (CND). The motivations for the measurement of nDVCS are discussed. Then the work realized on the CND, from the development of the reconstruction and calibration algorithms and software to the assessment of its performances, is presented.

The second part of this manuscript deals with the CLAS12 data analysis, aiming to extract TCS observables. It is decomposed in 3 Chapters:

- Chapter 4 summarizes the particle identification algorithms applied on data. It also provides a complete explanation of the positron identification techniques developed for the TCS analysis, taking advantage of multivariate analysis tools such as neural networks. The development of momentum corrections and fiducial cuts applied to the data is also reported.

- Chapter 5 focuses on the work realized in order to extract the TCS observables. The simulation framework is presented, followed by the description of the exclusivity cuts applied on data. A phenomenological study of the Forward-Backward asymmetry, an observable not studied for TCS before this thesis, is also exposed in this chapter.

- Chapter 6 summarizes all the results obtained in this analysis. The results are compared with theoretical predictions. Finally physical interpretations and conclusions are drawn.

Chapter 1

Physics motivations

1.1 Nucleon structure studies with electromagnetic probes

Nucleons are composed of three valence quarks (*uud* for protons, *udd* for neutrons). The valence quarks are surrounded by a cloud of quark-antiquarks pairs and gluons called the *sea*. The interaction between the partons (quarks and gluons) is described by the theory of Quantum Chromodynamics (QCD). At energies comparable to the mass of the nucleon, QCD cannot be computed pertubatively. In this regime, structure functions have to be introduced to described the structure of the nucleon. QCD is introduced in Subsection 1.1.1. In the two following subsections, structure functions accessible in elastic (Subsection 1.1.2) and inelastic (Subsection 1.1.3) scattering experiments are presented.

1.1.1 Quantum Chromodynamics

The strong force describes the interaction of particles carrying a color charge (red, blue or green). Quarks, which carry color charge, interact by the exchange of massless bosons, the gluons. The gluons also carry color charge and interact with each other and with themselves. The theory of QCD is described by a Lagrangian which is gauge invariant under the SU(3) symmetry. This Lagrangian is invariant under the transformation $\psi(x) \rightarrow U(x)\psi(x)$, where U is a unitary 3x3 matrix of determinant one, acting on the color state of $\psi(x) = (\psi_R(x), \psi_G(x), \psi_B(x))$. The Lagrangian of QCD is written:

$$\mathcal{L} = -\frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu} + \sum_k \overline{\psi}_{ki} \left(i \not{D}_{ij} - m_k \right) \psi_{kj}, \tag{1.1}$$

where

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu \tag{1.2}$$

is the gluon field strength tensor,

$$(D_\mu)_{ij} = \partial_\mu \delta_{ij} - i g A^a_\mu T^a_{ij} \tag{1.3}$$

is the gauge derivative, ψ_k are the quark fields where k runs over quark flavors and i, j run over color charge, A^a_μ are the gluon fields where a runs from 1 to 8, T^a_{ij} are the 8 SU(3) generator matrices, f^{abc} are the structure constants of SU(3) and g is the strong coupling constant. By analogy to the fine-structure constant in quantum electrodynamics, it is convenient to define the strong fine-structure constant α_S as:

$$\alpha_S = g^2/4\pi. \tag{1.4}$$

The SU(3) structure of the Lagrangian is at the origin of two main properties of QCD, asymptotic freedom and confinement. Confinement refers to the fact that the only stable states allowed by QCD are color singlet states, which have net color charge. Experimentally quarks and gluons are never observed alone but in composite bound hadrons (baryons or mesons). Confinement is directly related to the QCD potential increasing with distance. If one tries to pull away two quarks, the energy stored in the potential increases and can become high enough to produce a pair of quark-antiquark, restoring the net zero color charge of the system.

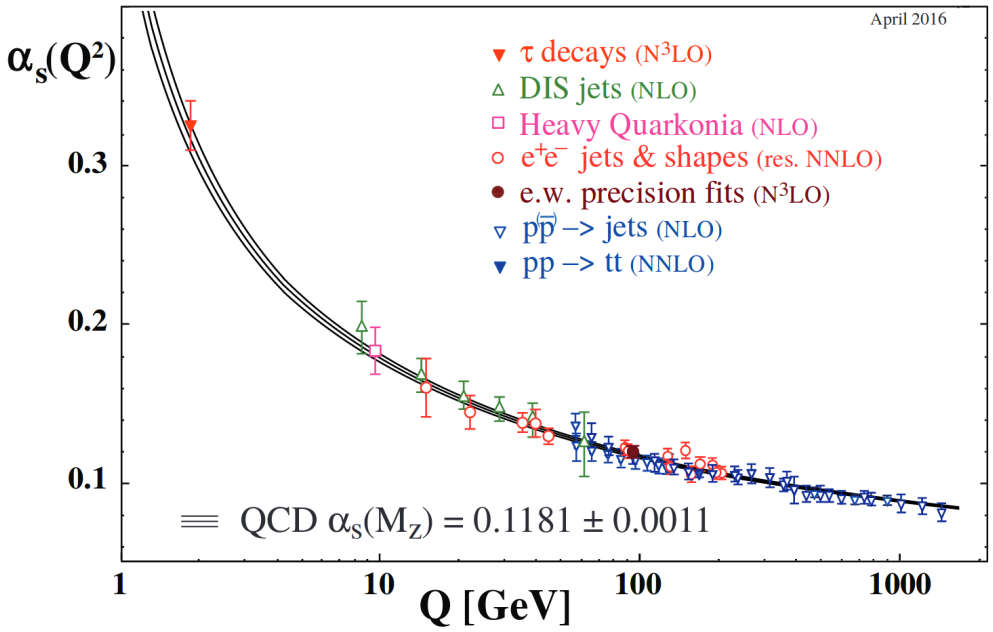


Figure 1.1: Summary of the measurement of α_S as a function of Q. Extracted from [1]

The second property, asymptotic freedom, arises from the renormalisation of QCD. As for QED, renormalization of QCD (ie. fixing an energy scale at which a process is observed) leads to a dependence of the strong constant α_S with respect to the renormalisation scale μ as:

$$\alpha_S(\mu^2) = \frac{12\pi}{(33 - 2n_f) \cdot ln\left(\frac{\mu^2}{\Lambda^2}\right)} \tag{1.5}$$

where μ is the renormalisation scale, $\Lambda \approx 250$ MeV, and n_f is the number of flavors. One can see that the strong coupling constant decreases with the scale, as shown in Figure 1.1. At high energy, α_S is small and calculations can be done pertubatively. At low energy, comparable to the nucleon mass, QCD cannot be calculated through a pertubative expansion in powers of the coupling constant α_S. In this regime, the interactions of partons inside the nucleons must be described by structure functions. Experiments, along with phenomenology and Lattice QCD, are currently the main ways to study the properties of these structure functions.

1.1.2 Elastic Scattering and Form Factors

Electron-proton elastic scattering experiments have historically played an important role in the discovery of the composite structure of the proton and the subsequent study of its structure. In the following, largely taken from [2] and [3], an historic approach of elastic scattering is presented.

The elastic scattering of a point-like particle on extended objects provides an efficient way to explore the structure of the latter. Rutherford was the first to use elastic scattering techniques to probe the structure of the gold atom in 1911. Using a beam of alpha particles, that can be considered as point-like compared to gold atoms, Rutherford showed that the mass of the atom is concentrated at its center. The Rutherford cross section formula, that applies for non-relativistic probes scattering on infinite mass point-like targets via the electromagnetic interaction, is given by:

$$\frac{d\sigma}{d\Omega} = \frac{\alpha_{em}^2}{16E^2 sin^4\left(\frac{\Theta}{2}\right)}, \tag{1.6}$$

where $\alpha_{em} = \frac{e^2}{4\pi} \simeq \frac{1}{137}$ is the electromagnetic fine structure constant, E the energy of the probe, Θ is the laboratory scattering angle and Ω is the solid angle where the probe is scattered.

Taking into account the relativistic effects and spins of the probe and the target, one can derive the Mott cross section formula. This formula describes the scattering of a massless particle (an electron) on a massive particle (a proton) via the exchange of a single photon as shown in Figure 1.2:

$$\frac{d\sigma}{d\Omega} = \frac{\alpha_{em}^2}{4E^2 sin^4\left(\frac{\Theta}{2}\right)} \frac{E'}{E} \left(cos^2\frac{\Theta}{2} + \frac{Q^2}{2m_p^2} sin^2\frac{\Theta}{2} \right), \tag{1.7}$$

where E is the energy of the incoming probe, E' is the energy of the outgoing probe, m_p is the mass of the target, and $Q^2 = -q^2 = -(k' - k)^2$ is the virtuality of the exchanged photon.

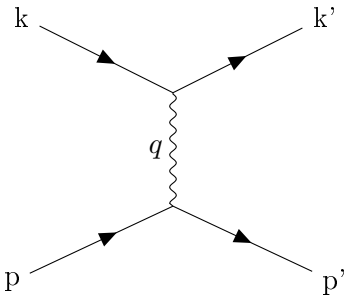


Figure 1.2: Point-like particle scattering

The Mott formula only applies to low energy processes ($\sqrt{Q^2} \ll m_p$). The distance (in GeV^{-1}) probed by a virtual photon with virtuality q is approximately $1/\sqrt{Q^2}$. In the regime where the Mott cross section applies the distance probed by the electromagnetic interaction is bigger than the size of the proton. In this case, the target and the probe are considered point-like particles. Therefore it is not possible to investigate the structure of the target in this energy regime.

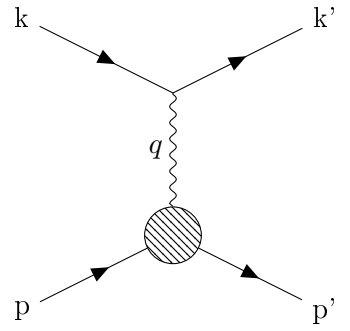


Figure 1.3: Feynman diagram for elastic scattering.

As the virtuality of the virtual photon increases, smaller distances can be probed. This was achieved experimentally with the increase of the energy of the available electron beams. When $\sqrt{Q^2} > m_p$ the proton size has to be accounted for. It is necessary to introduce structure functions, Form Factors (FFs), to account for the proton finite size. The Mott cross section can be re-written in terms of two FFs. The electron-proton elastic scattering is described by the Feynman diagram in Figure 1.3. The complete scattering formula was introduced by Rosenbluth in 1950 [4] and reads:

$$\frac{d\sigma}{d\Omega} = \frac{\alpha_{em}^2}{4E^2 sin^4\left(\frac{\Theta}{2}\right)} \frac{E'}{E} \left(\frac{G_E(Q^2)^2 + \tau G_M(Q^2)^2}{1+\tau} cos^2\frac{\Theta}{2} + 2\tau G_M(Q^2)^2 sin^2\frac{\Theta}{2} \right), \tag{1.8}$$

where τ is given by

$$\tau = \frac{Q^2}{4m_p^2}, \tag{1.9}$$

and $G_E(Q^2)$ and $G_M(Q^2)$ are the Sachs electric and magnetic FFs. It is convenient to introduce the Dirac and Pauli FFs, $F_1(Q^2)$ and $F_2(Q^2)$, respectively, as:

$$G_E = F_1 - \frac{Q^2}{4m_p^2}F_2, \tag{1.10}$$

$$G_M = F_1 + F_2. \tag{1.11}$$

The Form Factors can be directly related to the spatial distribution of the charge and the magnetic moment in the nucleon. This was shown by Hofstadter in 1956 [5] [6] and allowed for a first determination of the charge radius of the proton. In the Breit frame, which is, in the case of elastic scattering, the center-of-mass frame of the electron-nucleon system, the electric (resp. magnetic) FF can be interpreted as the Fourier transform of the transverse charge (resp. magnetization) distribution. In this frame, the exchanged photon carries no energy ($Q^2 = \vec{q}^2$), and the charge density can be written as:

$$\rho(\vec{r}) = \int G_E(\vec{q}^2)\frac{M}{E}e^{-i\vec{q}\cdot\vec{r}}\frac{d^3\vec{q}}{(2\pi)^3}. \tag{1.12}$$

The same equation applies to G_M and gives the magnetization density. From this interpretation, one can derive the charge and magnetic mean squared radii of the nucleon given by:

$$\langle r_E^2\rangle = -6\frac{dG_E(Q^2)}{dQ^2} \tag{1.13}$$

$$\langle r_M^2\rangle = -\frac{6}{G_M(0)}\frac{dG_M(Q^2)}{dQ^2}. \tag{1.14}$$

The current value of $\langle r_E^2\rangle$ derived from both elastic scattering measurements and hydrogen spectroscopy is 0.879 ± 0.008 fm. Proton radii extracted from muonic hydrogen spectroscopy have also been published [7]. The current value obtained by this technique, 0.84184 ± 0.00064 fm, is almost 8 sigmas away from the other value. Understanding this discrepancy has been a topic of great interest in the last years on both the experimental and theoretical point of view. New results published recently were obtained from an electron scattering experiment (PRAD) [8] and are in accordance with the muonic hydrogen radius. This result was achieved using a window-less hydrogen target and high resolution calorimeters for electron detection. This set-up was designed to achieve a more precise measurement than any previous scattering experiments.

FFs for both protons and neutrons have been measured. They have very different behaviors. For protons F_1 dominates for all values of Q^2. On the contrary, F_2 dominates in the low Q^2 region for neutrons, as shown in Figure 1.4.

Finally, one can also define axial and pseudo-scalar FFs [10], where one replaces the virtual photon by charged and neutral weak bosons. Axial and pseudo-scalar FFs G_A and G_P are measurable in neutrino scattering experiments, muon capture and pion electro-production processes.

1.1.3 Inelastic Scattering

Deeply Inelastic Scattering

For transferred momentum $Q^2 >> m_p^2$, the virtuality of the exchanged photon is sufficient to probe distances smaller than the size of the proton. If the photon energy is big enough, $\nu > m_p$, the proton is also very likely to break down. This regime is called Deep Inelastic Scattering (DIS). The DIS reaction is written $ep \to eX$ where X represents any possible hadronic final state produced by the breakup of the proton. In the single-photon-exchange approximation, DIS is described by the Feynman diagram in Figure 1.5a. The mass of the final state particles is given by $W = (p+q)^2$ with the $W \gg m_p$ condition. Contrary to elastic scattering, which is parameterized by only one variable (Q^2), DIS is described by two independent variables (Q^2 and ν).

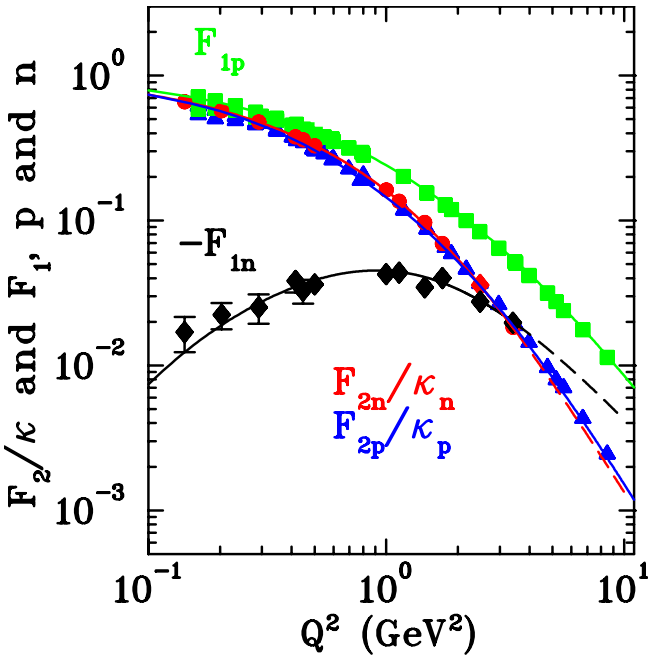
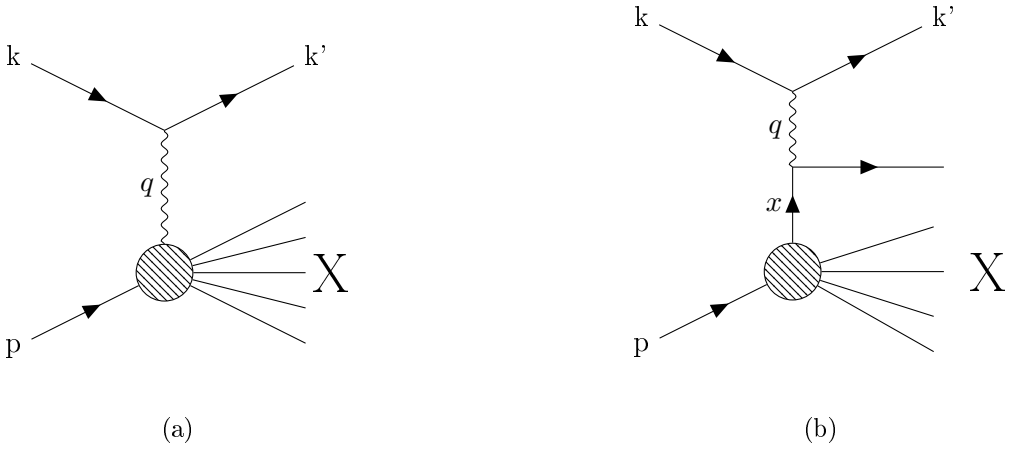


Figure 1.4: Proton and neutron FFs, taken from [9]



(a) (b)

Figure 1.5: (a) Feynman diagram for DIS, (b) DIS diagram where the virtual photon interacts with a single quark.

The cross section of DIS is given by:

$$\frac{d\sigma}{d\Omega dE'} = \frac{\alpha_{em}^2}{4E^2 sin^4\left(\frac{\Theta}{2}\right)}\left(W_2(\nu, Q^2)cos^2\frac{\Theta}{2} + 2W_1(\nu, Q^2)sin^2\frac{\Theta}{2}\right), \tag{1.15}$$

where the incoming and outgoing electron momenta are defined as in the previous section and $\nu = E - E'$. This formula is very similar to the Rosenbluth (Formula (1.8)), the main difference being the dependence of the structure functions W_1 and W_2 on two variables.

This formula is conventionally re-written as:

$$\frac{d\sigma}{d\Omega dE'} = \frac{\alpha_{em}^2}{4E^2 sin^4\left(\frac{\Theta}{2}\right)}\left(\frac{F_2^{DIS}(x_B, Q^2)}{\nu}cos^2\frac{\Theta}{2} + 2\frac{F_1^{DIS}(x_B, Q^2)}{m_p}sin^2\frac{\Theta}{2}\right), \tag{1.16}$$

where $F_1^{DIS}(x_B, Q^2)$ and $F_2^{DIS}(x_B, Q^2)$ are defined as:

$$F_1^{DIS}(x_B, Q^2) = m_p W_1(\nu, Q^2), \tag{1.17}$$

$$F_2^{DIS}(x_B, Q^2) = \nu W_2(\nu, Q^2) = \frac{p \cdot q}{m_p} W_2(\nu, Q^2), \tag{1.18}$$

and where we defined the Bjorken variable as:

$$x_B = \frac{Q^2}{2p \cdot q}. \tag{1.19}$$

In 1969, the SLAC group showed that $F_1^{DIS}(x_B, Q^2)$ and $F_2^{DIS}(x_B, Q^2)$ are independent of Q^2 [11]. This property is now referred as scaling. The Q^2-independence of the structure functions implies that the exchanged photon interacts with an object with no internal structure (ie. no dependence on the probed distance $1/\sqrt{Q^2}$). The scaling results published by the SLAC collaboration are shown in Figure 1.6. This was the first evidence of the existence of quarks. Indeed in the DIS regime, the virtuality of the photon is large enough to probe distances much smaller than the size of the proton and observe its fundamental components.

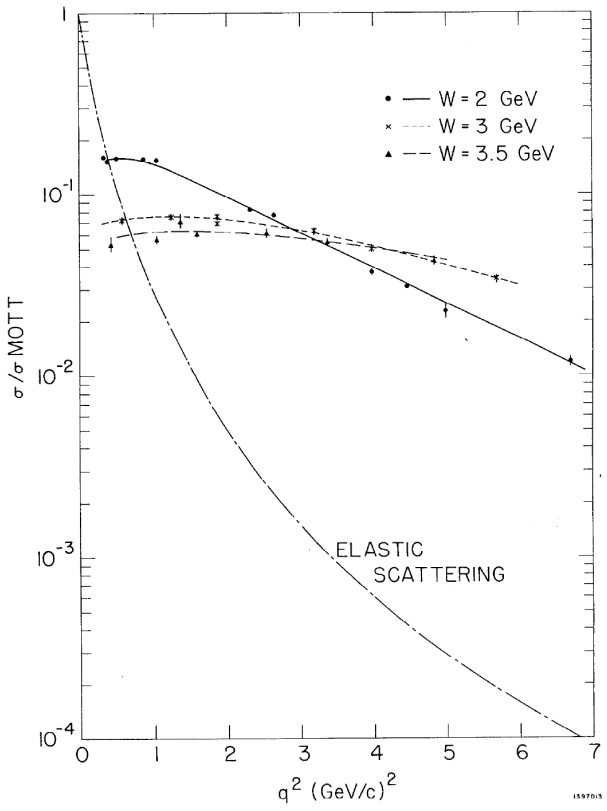


Figure 1.6: Ratio between the DIS cross section and the Mott cross section as a function of the squared of the momentum of the exchanged photon. These results were obtained by the SLAC collaboration in [11].

The parton model and Parton Distribution Functions

In order to better understand the SLAC results, the DIS cross section can be interpreted within the parton model [12]. In this model introduced by Feynman [13], the proton is composed of point-like particles, the partons. The parton model has a clear interpretation in the frame where the nucleon has an infinite momentum in the z-direction (the light-cone frame). In this infinite momentum frame, partons can be considered as non-interacting particles due to time dilatation. Thus the virtual photon interacts with a single quark as shown in Figure 1.5b. Defining x as the fraction of the nucleon

momentum carried by the quark before the interaction, $\tilde{p}$ the momentum of the struck quark after the interaction, and noting that the final state struck quark must be on-shell, one can write:

$$x \cdot p + q = \tilde{p}, \tag{1.20}$$

and after squaring and neglecting the quark mass:

$$x = \frac{Q^2}{2p \cdot q} = x_B. \tag{1.21}$$

The Bjorken variable x_B can be interpreted as the fraction of the proton momentum carried by the struck quark in the infinite momentum frame. The cross section can then be written as a sum of cross sections of electron scattering off a single point-like parton, weighted by the probability of finding such a parton with momentum fraction x. $F_1^{DIS}(x, Q^2)$ and $F_2^{DIS}(x, Q^2)$ can be re-written as:

$$F_1^{DIS}(x, Q^2) = \frac{1}{2} \sum_i e_i^2 \left(q_i(x) + \overline{q_i}(x)\right), \tag{1.22}$$

$$F_2^{DIS}(x, Q^2) = x \sum_i e_i^2 \left(q_i(x) + \overline{q_i}(x)\right), \tag{1.23}$$

where e_i is the charge of a quark of flavor i and q_i is the Parton Distribution Function (PDF) for the quarks of flavor i, with the bar denoting antiquarks. The PDFs are thus interpreted as the probability of interaction with a quark carrying a momentum fraction x. Furthermore, from Equations (1.22) and (1.23), $F_1^{DIS}(x)$ and $F_2^{DIS}(x)$ verify the relation:

$$F_2^{DIS}(x) = 2xF_1^{DIS}(x), \tag{1.24}$$

known as the Callan-Gross relation. This equality is a consequence of the fermionic nature of quarks. Its experimental verification was a further confirmation of the quark model validity.

The current fitted PDFs q_i are shown in Figure 1.7. The quark flavor separation of the PDFs is achieved by considering proton and neutron PDFs. The separation between quarks and antiquarks is done using neutrino scattering data. In Figure 1.7 one can also see that PDFs depend on Q^2, contradicting the simple parton model explained above. This behavior, called scaling violation, is understood by taking into account gluons radiated by the struck quark and is described by the DGLAP equations [14].

Polarized PDFs

One can also define polarized PDFs by considering the spin of the partons inside a polarized nucleon. There are two polarized structure functions g_1 and g_2 (see [16, 17]). The polarized structure function g_1 can be written as:

$$g_1(x, Q^2) = \sum_i e_i^2 \left(\Delta q_i(x) - \Delta\overline{q_i}(x)\right), \tag{1.25}$$

with

$$\Delta q_i(x) = q_i^{\uparrow} - q_i^{\downarrow}, \tag{1.26}$$

where $q_i^{\uparrow}$ (resp. $q_i^{\downarrow}$) represents the probability density of quarks with helicity parallel (resp. antiparallel) to the one of the nucleon. The g_1 structure function can be measured in longitudinally polarized target-longitudinally polarized beam DIS experiments. The second structure function g_2 is expected to vanish at leading twist (see definition of twist in Section 1.2), but has non-zero twist 2 and 3 contributions and is measurable in transversely polarized target experiments [18].

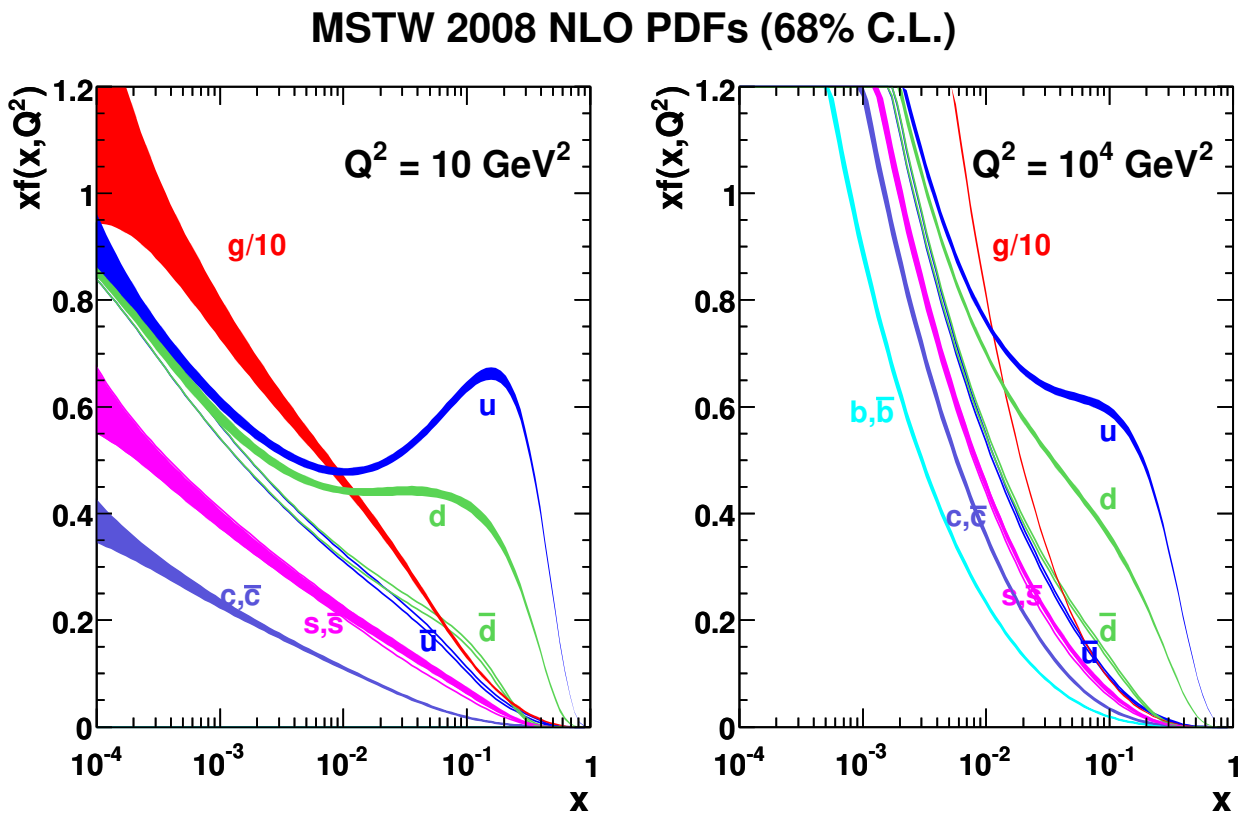


Figure 1.7: Parton distribution functions as a function of x for $Q^2 = 10$ GeV^2 and $Q^2 = 10^4$ GeV^2 determined by a fit over a large range of data [15].

1.2 Generalized Parton Distributions

FFs and PDFs introduced in the previous section describe the nucleon structure respectively in terms of spatial coordinates and momentum coordinates. However no correlations between the position and the momentum of the struck quark can be extracted from these functions. The concept of Generalized Parton Distributions (GPDs) was developed in the early 90's [19, 20, 21] and provided a framework to interpret the partonic structure of the nucleon in terms of traverse position, longitudinal momentum and their correlations.

The concept of GPDs is introduced in Subsection 1.2.1. Their link to the structure functions described in the previous section is highlighted. Their close relation to the nucleon spin puzzle and the pressure distribution inside nucleons is presented in Subsection 1.2.2. Finally, currently available models of GPDs are highlighted in Subsection 1.2.3 and generalizations of GPDs are presented in Subsection 1.2.4.

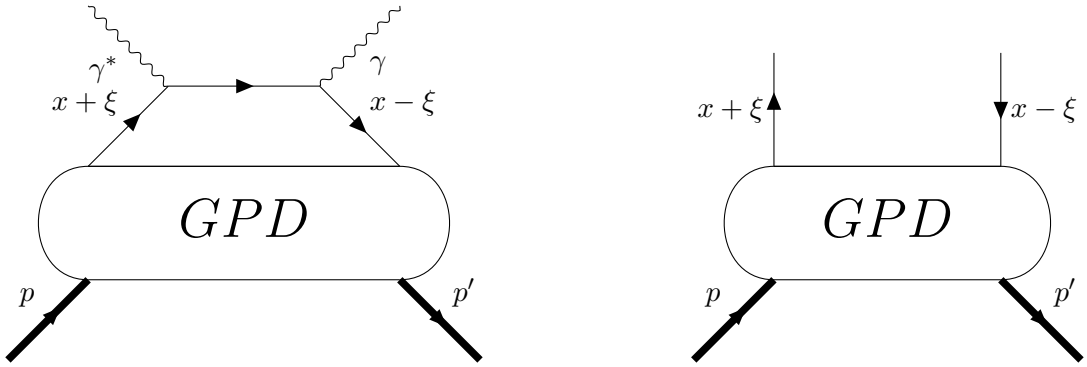
1.2.1 Phenomenology of Generalized Parton Distributions

Factorization

The concept of GPDs is based on QCD factorization. As shown in Figure 1.1 in the previous section, the strong interaction coupling constant α_S varies with the energy scale of the studied process. At high-energy scale α_S is small and one can apply power series expansion to calculate matrix elements. However at low energies, α_S becomes large and the power expansion is not possible. To calculate the amplitude of a high energy process involving a complex QCD object like a nucleon, one has to separate a point-like high-energy (i.e. hard) interaction, between an electron and a quark for example, from the long-range low-energy (i.e. soft) structure of the nucleon. The hard part can be described using

the Feynman diagram formalism while the soft part has to be parameterized using structure functions. PDFs, FFs and GPDs are such sets of functions representing this soft part.

Definitions



(a) (b)

Figure 1.8: The Feynman Diagram for DVCS at leading twist. An electron interacts with a quark from the nucleon via the exchange of a virtual photon; this quark then emits a real photon. (a) The full DVCS diagram, (b) the soft part of the diagram parameterized by GPDs.

Quark GPDs GPDs [22, 23, 24] were first introduced to describe the nucleon soft structure contribution to the exclusive production of a photon or a meson off a nucleon. Deeply Virtual Compton Scattering (DVCS) is the exclusive electro-production of a photon induced by a high virtuality (Q^2) initial photon (see Figure 1.8a) and is one of the processes parameterized by GPDs. DVCS is a key reaction for the GPD experimental program and will be further present in Section 1.3. The DVCS amplitude can be expanded as a series of operators, were each term is sorted according to its power in α_S (referred as its *order*) and to its *twist*, which is defined as its dimension minus its spin. Terms with high order are suppressed by increasing power of α_S, while high-twist terms are suppressed by increasing powers of $1/Q$. At leading order and leading twist (twist-2, which is assumed in the following), the amplitude of DVCS is given by the "handbag" diagram of Figure 1.8a. The top high-energy part of the graph is described by the usual perturbative Feynman rules, while the bottom part describing the soft structure of the nucleon (see Figure 1.8b) is parameterized by GPDs. The GPDs are universal functions and their properties should not depend on the process studied. Thus GPDs can be studied not only from DVCS but also from other processes, as presented in Section 1.3.

Formally, quark GPDs (gluon GPDs are defined in the next paragraph) are defined in the light-cone frame (introduced in the previous section), where the incoming and outgoing momenta of the nucleon (p and p') are collinear to the z-axis. In this frame, one can conveniently define plus/minus components of a 4-vector a as $a^{\pm} = a^0 + a^3/\sqrt{2}$. GPDs are related to the Fourier transform of the non-local non-diagonal matrix element $\langle p'|\bar{\psi}_q(0)\mathcal{O}\psi_q(y)|p\rangle$ represented by the diagram in Figure 1.8b. In this matrix element ψ_q is the flavor q quark field, $|p'\rangle$ and $|p\rangle$ are the quantum states of the incoming and outgoing nucleon with respective momenta p and p', and $\mathcal{O}$ is an operator appearing in the convolution with the hard part of the diagram (usually products of gamma matrices). This matrix element is non-local, as the quark field is taken at different space-time coordinates, and non-diagonal, as the nucleon states $|p'\rangle$ and $|p\rangle$ are different. In the case of processes conserving the quark helicity such as DVCS, the initial and final quark fields have the same helicity state. In the light-cone frame, the non-zero matrix elements appearing in the amplitude of these processes read $\langle p'|\bar{\psi}_q(0)\gamma^+\psi_q(y)|p\rangle$ and $\langle p'|\bar{\psi}_q(0)\gamma^+\gamma^5\psi_q(y)|p\rangle$. They are written in terms of four GPDs, H^q, $\tilde{H}^q$, E^q, and $\tilde{E}^q$, as:

$$\frac{P^+}{2\pi}\int dy^- e^{ixP^+y^-}\langle p'|\bar{\psi}_q(0)\gamma^+\psi_q(y)|p\rangle\Big|_{y^+=\vec{y}_\perp=0}$$
$$= H^q(x,\xi,t)\;\bar{N}(p')\gamma^+N(p)\;+\;E^q(x,\xi,t)\;\bar{N}(p')i\sigma^{+\nu}\frac{\Delta_\nu}{2m}N(p)\;,$$
$$\frac{P^+}{2\pi}\int dy^- e^{ixP^+y^-}\langle p'|\bar{\psi}_q(0)\gamma^+\gamma^5\psi_q(y)|p\rangle\Big|_{y^+=\vec{y}_\perp=0}$$
$$= \tilde{H}^q(x,\xi,t)\;\bar{N}(p')\gamma^+\gamma_5N(p)\;+\;\tilde{E}^q(x,\xi,t)\;\bar{N}(p')\gamma_5\frac{\Delta^+}{2m}N(p)\;, \tag{1.27}$$

where $P=(p+p')/2$ is the average nucleon momentum, $\Delta=(p'-p)$ is the nucleon transferred momentum, N and $\bar{N}$ are the initial and final nucleon spinors, and $\sigma^{\mu\nu}=\frac{i}{2}[\gamma^\mu,\gamma^\nu]$ are the gamma matrices commutators.

Each of the four helicity-conserving GPDs corresponds to a combination of the possible quark-nucleon helicity-spin flips (while conserving the quark helicity). The helicity-spin decomposition of the GPDs is illustrated in Figure 1.9. Some exclusive processes such as Deeply Virtual Meson Production (DVMP) ($ep\to e'p'm$, where m is a meson), are described by GPDs for which the quark flips its helicity and which are called *tranversity* GPDs. There are four transversity GPDs (H_T^q, $\tilde{H}_T^q$, E_T^q, and $\tilde{E}_T^q$) defined as the Fourier transforms of the matrix elements $\langle p'|\bar{\psi}_q(0)\sigma^{+\nu}\psi_q(y)|p\rangle$ and $\langle p'|\bar{\psi}_q(0)\sigma^{+\nu}\gamma^5\psi_q(y)|p\rangle$. The following parts of this manuscript focus on the properties of helicity-conserving quark GPDs. They are simply referred to as GPDs.

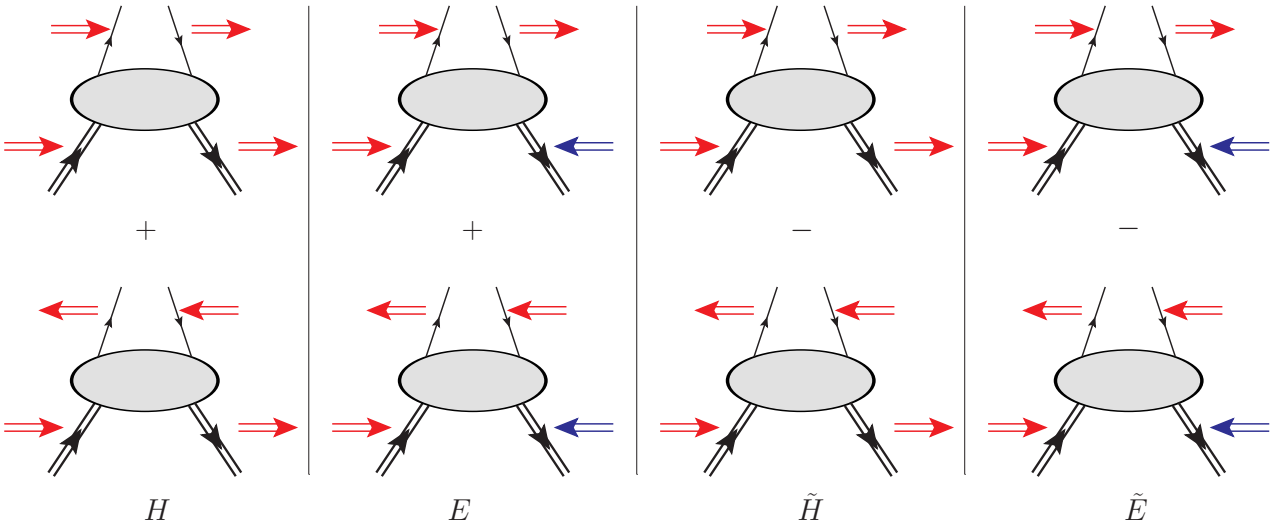


Figure 1.9: The four GPDs corresponding to the four different helicity-spin combinations of the quark-nucleon system.

Quark GPDs depend explicitly on three variables, x, the momentum imbalance ξ, and the Mandelstam variable t, where $x+\xi$ is the momentum fraction of the initial quark and $x-\xi$ the momentum fraction of the final quark. They read:

$$t=(p'-p)^2, \tag{1.28}$$

$$\xi=\frac{-\Delta^+}{2P^+}. \tag{1.29}$$

GPDs also depend on the hard scale (usually the virtuality Q^2 of the incoming virtual photon in DVCS and DVMP). Their evolution with Q^2 is well known and usually this dependence is omitted. The quark momentum fraction x varies between -1 and 1, while ξ varies from 0 to 1. In the region $\mid x \mid > \xi$, GPDs represent the probability amplitude of removing a quark with a momentum fraction $x+\xi$ and putting it back in the nucleon with a momentum fraction $x-\xi$. This region is called the DGLAP region in reference to the QCD evolution equation [14]. In the region $-\xi < x < \xi$, the GPDs can be interpreted as the probability amplitude of extracting a quark-antiquark pair from the nucleon. This region is referred as the ERBL region [25].

Gluon GPDs Gluon GPDs can also be defined. There are four gluon GPDs (H^g, $\tilde{H}^g$, E^g, and $\tilde{E}^g$). Gluon GPDs have a similar definition to quark GPDs (see [22]). They describe the probability of picking a gluon with momentum fraction $x+\xi$ and putting it back with the $x-\xi$ fraction as shown in Figure 1.10. In the valence region (for energies of the order of 1 GeV) and at leading twist, gluon GPDs only enter the DVCS amplitude at next-to-leading order in α_S, and are believed to have a small contribution to the DVCS cross section. However this might not be true in specific cases (see Subsection 1.3.4 for details). We do not consider gluon GPDs in the following discussion of GPDs properties and models.

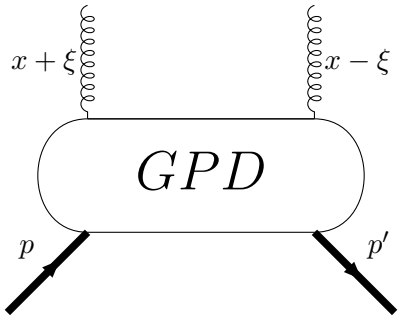


Figure 1.10: Diagram representing the gluon GPDs.

Forward limit

GPDs are related to PDFs via the following model-independent relations:

$$H^q(x,0,0) = \begin{cases} q(x), & x > 0 \\ -\bar{q}(-x), & x < 0 \end{cases} \tag{1.30}$$

$$\tilde{H}^q(x,0,0) = \begin{cases} \Delta q(x), & x > 0 \\ \Delta\bar{q}(-x), & x < 0 \end{cases} \tag{1.31}$$

where q, $\bar{q}$, Δq, and $\Delta\bar{q}$ are the PDFs and polarized PDFs (see Section 1.1) associated to the quark flavor q. This property is derived from the optical theorem illustrated in Figure 1.11, which links the DIS cross section to the imaginary component of the amplitude of the $\gamma^* N \to \gamma^* N$ process in the forward limit (at $t = 0$). These forward-limit properties provide strong constraints to GPDs as PDFs are relatively well known. There are no such relations for the E and $\tilde{E}$ GPDs, which make them more difficult to constrain.

GDP x-moments

Similarly to the PDFs in the forward limit, FFs can also be related to GPDs through their first x-moments as:

$$\int_{-1}^{1} dx H^q(x,\xi,t) = F_1^q(t), \tag{1.32}$$

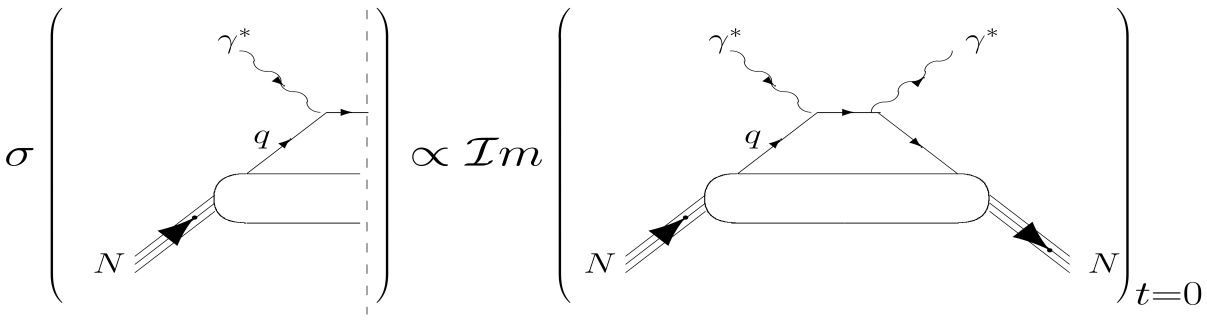


Figure 1.11: Schematic description of the optical theorem that links DIS cross section to the imaginary part of the $\gamma^* N \to \gamma^* N$ process. Figure from [22].

$$\int_{-1}^{1} dx E^q(x,\xi,t) = F_2^q(t), \tag{1.33}$$

$$\int_{-1}^{1} dx \tilde{H}^q(x,\xi,t) = G_A^q(t), \tag{1.34}$$

$$\int_{-1}^{1} dx \tilde{E}^q(x,\xi,t) = G_P^q(t). \tag{1.35}$$

where F_1^q, F_2^q,G_A^q and G_P^q are, respectively, the Pauli FFs, axial, and pseudo-scalar FFs defined in Section 1.1. Because of the restricted definition domain of ξ, these integrals only depend on t. Furthermore higher x-moments of GPDs verify a polynomiality condition. The n-th x-moment of the GPD H can be written as a polynomial in ξ, of order n for even n, and of order $(n+1)$ for odd n, as:

$$\begin{aligned} \text{if } n \text{ even} &: \int_{-1}^{1} dx\, x^n H(x,\xi,t) = a_0 + a_2\xi^2 + a_4\xi^4 + ... + a_n\xi^n, \\ \text{if } n \text{ odd} &: \int_{-1}^{1} dx\, x^n H(x,\xi,t) = a_0 + a_2\xi^2 + a_4\xi^4 + ... + a_{n+1}\xi^{n+1}, \end{aligned} \tag{1.36}$$

where the a_i coefficients only depend on t. For E the same conditions apply with the opposite sign for the coefficients a_i. For $\tilde{H}$ and $\tilde{E}$, the odd case is of maximum order $(n-1)$. The polynomiality condition is an important property as it provides strong constraints on GPDs models.

1.2.2 Physical interpretations

GPDs have a direct link to PDFs and FFs. Thus they encode both transverse spatial and longitudinal momentum information but also their correlations (as shown in Figure 1.12). Their interpretation provides a deeper insight on the structure of the nucleon than FFs and PDFs alone. In this subsection the three main physical interpretations of GPDs are presented. The tomography interpretation, the link between GPDs and the nucleon spin, and finally the relation with the internal forces and the pressure in the nucleon are introduced.

Tomography of the nucleon

It is possible to interpret GPDs as the probability distribution of finding a parton at transverse position $b_\perp$ with respect to the "center of momentum" of the nucleon, with longitudinal momentum fraction x, via a Fourier transform [26, 27, 28] :

$$H^q(x,b_\perp) = \int \frac{d^2\Delta_\perp}{(2\pi)^2} e^{-ib_\perp \Delta_\perp} H^q(x,0,-\Delta_\perp^2). \tag{1.37}$$

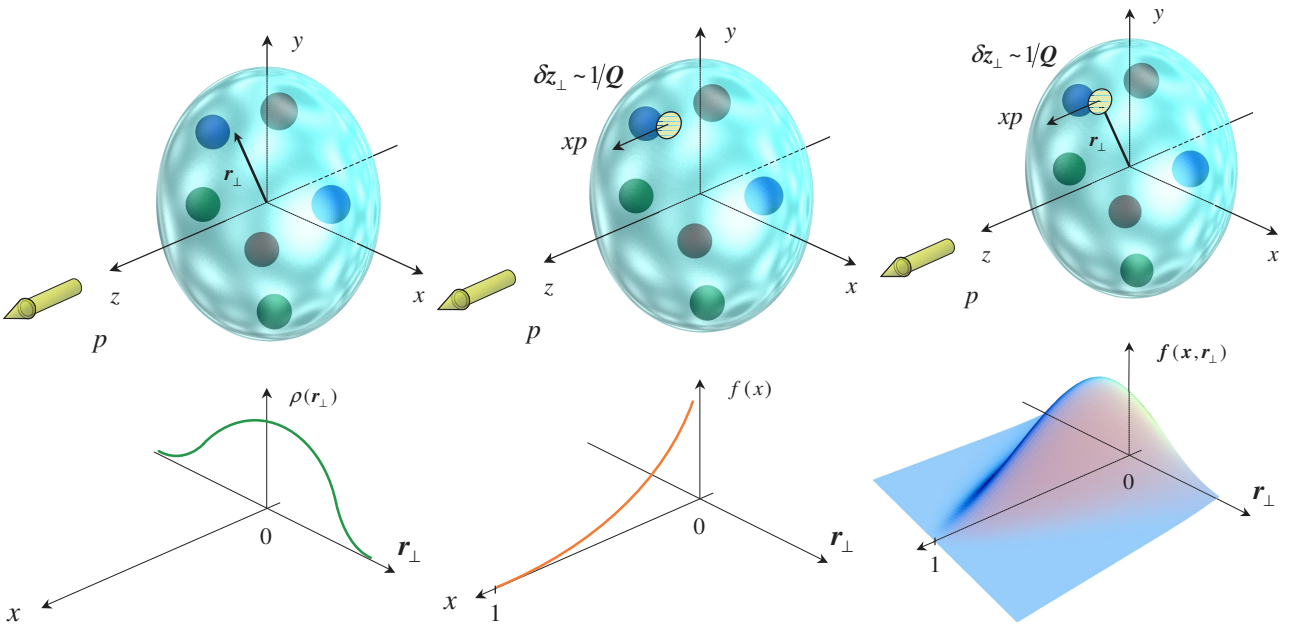


Figure 1.12: GPDs encode information from FFs and PDFs but also correlations not accessible with the previous functions alone. A tomography of the nucleon can be performed using GPDs, providing a picture of the internal structure of the nucleon in the transverse-spatial coordinates at different values of longitudinal momentum. Figure from [24].

Hence GPDs permit to perform a 2+1D (two transverse position dimensions and 1 momentum dimension) picture of the nucleon [29]. The interpretation of GPDs as a tomography of the nucleon is illustrated in Figure 1.12.

GPDs and the spin puzzle

GPDs are related to momentum and position of quarks in the nucleon. This relation implies that information about the angular momentum of quarks can be extracted from GPDs. This link was investigated by X. Ji in a 1997 publication [21] where the Ji's sum rule was first introduced. According to this sum rule, the spin of the nucleon can be decomposed as:

$$\frac{1}{2} = J_Q + J_G, \tag{1.38}$$

where J_Q and J_G are the total contributions of, respectively, quarks and gluons to the total spin of the nucleon. The contribution of the quarks can be further decomposed as:

$$J_Q = \frac{1}{2}\Delta\Sigma + L_Q, \tag{1.39}$$

where $\frac{1}{2}\Delta\Sigma$ and L_Q are the contributions from, respectively, intrinsic spin and angular momentum. It is not trivial that all of these contributions exactly add up to 1/2, and each contribution is a priori unknown. Polarized DIS experiments give access to the intrinsic spin component $\frac{1}{2}\Delta\Sigma$ as:

$$\Delta\Sigma = \sum_i \int_0^1 dx \Delta q_i(x) \tag{1.40}$$

where the sum runs over quark flavors, and the $\Delta q_i(x)$ are the polarized PDFs defined in Subsection 1.1.3. It was shown that this term contributes at most to 30% [30] of the total nucleon spin. The remaining contributions must account for the remainder. The angular momentum contribution from quarks can be accessed using GPDs. The Ji's sum rule relates the second moment of the sum over all quark flavors of the H^q and E^q GPDs at $t = 0$ to the total angular momentum of the quarks:

$$J_Q = \sum_q \frac{1}{2} \int_{-1}^1 dx\ x(H^q(x,\xi,0) + E^q(x,\xi,0)). \tag{1.41}$$

A complete measurement of H^q and E^q opens a way to extract the contribution to the nucleon's spin of the angular momentum of the quarks $L_Q = J_Q - \frac{1}{2}\Delta\Sigma$. The Ji's rule also apply to gluons. However the separation between the angular momentum contribution and the spin contribution as in Equation (1.39), is not applicable to gluons [31].

Forces inside the nucleon: the D-term

The first x-moment of H and E can be expressed as:

$$\int_{-1}^{1} dx\ xH^a(x,\xi,t) = A^a(t) + \xi^2 D^a(t) \tag{1.42}$$

$$\int_{-1}^{1} dx\ xE^a(x,\xi,t) = B^a(t) - \xi^2 D^a(t) \tag{1.43}$$

where the quantities A^a, B^a, and D^a are the Energy-Momentum Tensor (EMT) Form Factors and a denotes gluon or quark flavors. The EMT FFs are form factors appearing when one describes the interaction of the gravitational field with a nucleon [32, 33], as shown in Figure 1.13. From Equations

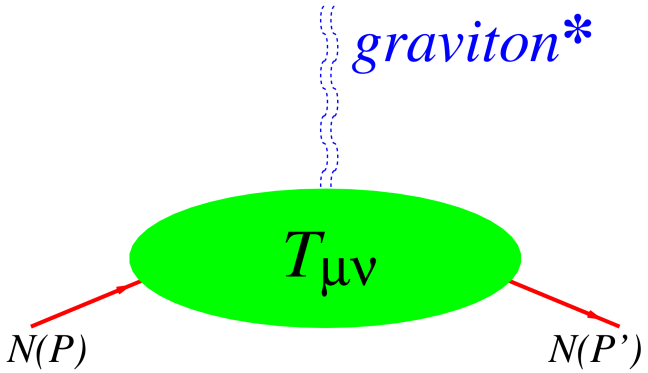


Figure 1.13: Diagram representing the interaction of a graviton on a nucleon, taken from [34]

(1.42) and (1.43), one can see that the EMT FFs A^a and B^a are related to the Ji's sum rule presented in the previous section. The D-term [34, 32], defined as $D(t) = \sum_a D^a(t)$, contains information about the force and the pressure distribution inside the nucleon. It can be related to the pressure inside the nucleon by the model independent formula [35]:

$$D(t) \propto \int d^3\mathbf{r}\ p(r) \frac{j_0(r\sqrt{-t})}{t}, \tag{1.44}$$

where j_0 is the first Bessel function. The D-term can also be related to Compton Form Factors [36, 37, 38], which are quantities measurable in DVCS experiments and which are presented in Subsection 1.3.1.

1.2.3 GPD Models

There are several parametrizations of GPDs. The main models are the double-distribution models (VGG [39], GK [40]), the dual parametrization [41] and the Mellin-Barnes model [42]. In the following section, the double-distribution model and its implementation in the VGG model are described.

Double Distribution parameterization

At fixed hard scale, GPDs depend on three variables (x, ξ and t). The double distribution (DD) model introduced in [43] and [44] is used to parameterize the x and ξ dependence of GPDs. The usual momentum notation for GPDs is shown in Figure 1.14a, where the incoming proton has momentum

P in the light-cone frame, and the initial and final quarks have momenta defined as fractions $(x+\xi)$ and $(x-\xi)$ of P. The main idea of the DDs parameterization is to perform a change of variable and define the initial quark momentum as $(\beta P - (1+\alpha)\frac{\Delta}{2})$, as in Figure 1.14b. Since ξ is given by:

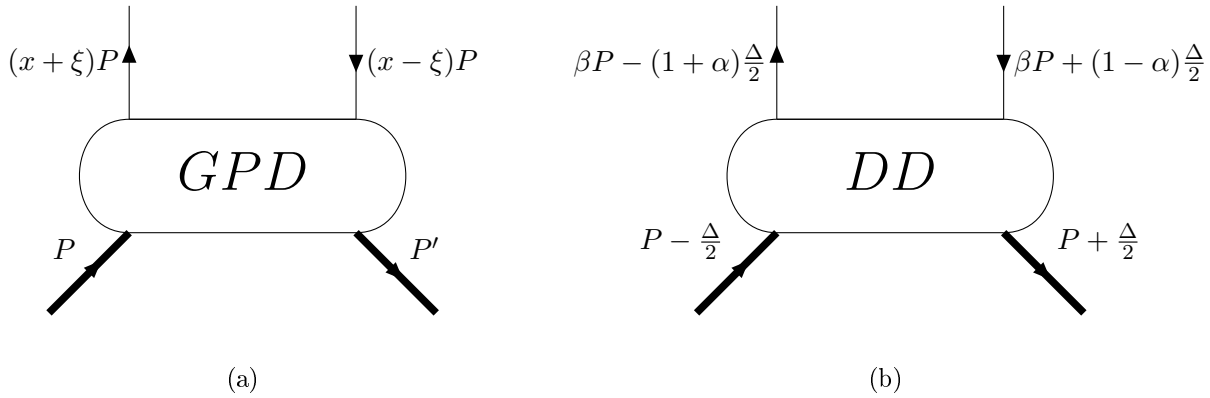
$$-2\xi = \frac{\Delta}{P}, \tag{1.45}$$

one can relate x, ξ, α and β using the relation:

$$x = \beta + \alpha\xi. \tag{1.46}$$

This relation authorizes to re-write each GPD as an integral of DDs as:

$$GPD^q(x,\xi) = \int_{-1}^{1} d\beta \int_{-1+|\beta|}^{1-|\beta|} d\alpha \delta(x-\beta-\xi\alpha) DD(\alpha,\beta). \tag{1.47}$$



(a) (b)

Figure 1.14: (a) Usual momentum parametrization of the GPDs, (b) Momentum parametrization used in the double distribution model.

This change of variables from GPDs to DDs has two main advantages. First, because of the δ function in the integral of Equation (1.47), the polynomiality conditions described in Subsection 1.2.1 are always satisfied, up to the n-th term. The second advantage of DDs is that their properties can be inferred from two limiting cases. The first case, when $\Delta = 0$, corresponds to the case where the incoming and the outgoing protons have equal momenta. In this case the DDs take the form of the PDFs measured in DIS (see Figure 1.11, relating GPDs at $t = 0$ and PDFs). The second limiting case is $P = 0$ and $\Delta \neq 0$. This corresponds to extracting a quark-antiquark pair with momenta $(1+\alpha)\frac{\Delta}{2}$ and $(1-\alpha)\frac{\Delta}{2}$ for the quark and the antiquark, respectively. In this limit, one can expect the DDs to have the form of Distribution Amplitudes (DAs) (see Figure 1.15). DAs describe the probability of finding a quark/antiquark pair with momentum fraction z and $(1-z)$ in a meson (or equivalently to form a meson from the quark-antiquark pair).

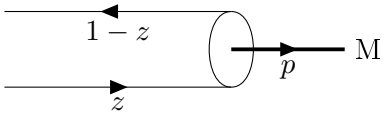


Figure 1.15: Feynman diagram of a DA. A quark with momentum fraction z interacts with an antiquark with momentum fraction $(1-z)$ to form a meson with momentum p and mass M.

Finally DDs are obtained by taking these two limiting cases and finding the best way to interpolate between the two limits. This was first done in [43]. In this paper the proposed DD takes the form of:

$$DD(\alpha,\beta) = h(\beta,\alpha)q(\beta), \tag{1.48}$$

$$h(\alpha,\beta) = \frac{\Gamma(2b+2)}{2^{2b+1}\Gamma^2(2b+1)} \frac{\left[(1-\mid\beta\mid)^2-\alpha^2\right]^b}{(1-\mid\beta\mid)^{2b+1}}, \tag{1.49}$$

where q is the PDF for the quark q, and where b is a parameter controlling the ξ dependence (note that there are two such parameters, for valence and sea quarks, b_{val} and b_{sea}). The DD formalism is the core of several GPDs parametrizations. In the following, the VGG model based on DD is presented in more details.

D-term The DD description ensures that the n-th moment in x of the GPDs are ξ-polynomials of order n. However the polynomiality constraints allow for an additional degree of freedom when n is odd for H and E. To ensure that the DD model is complete, one has to add an additional quantity. This quantity is closely connected to the quantity presented in Subsection 1.2.2. It was first introduced in [45], and is given by:

$$D(t,z) = (1-z^2)\sum_{n\ odd} d_n(t) C_n^{3/2}(z), \tag{1.50}$$

where $z = x/\xi$, and the $C_n^{3/2}$ is the n-th Gegenbauer polynomial of parameter 3/2. The coefficients $d_n(t)$ are free parameters.

The VGG model

In this subsection, the VGG model [46, 47, 39] for the GPD H is described. Other GPDs have different parametrizations but use the same concepts (see [22] for details). The (x,ξ) dependence is described using the DDs introduced previously. The t-dependence is factorized and reads:

$$F^q(\beta,\alpha,t) = DD(\alpha,\beta)\,\beta^{-\alpha'(1-\beta)t} \tag{1.51}$$

where $DD(\alpha,\beta)$ is given in Equation (1.48), and α' is taken from Regge trajectories and describes the t-dependence of the DD. Finally, the D-term is added, and the GPD H is given by:

$$H^q(x,\xi,t) = \int d\alpha d\beta\ \delta(x-\beta-\xi\alpha)F^q(\beta,\alpha,t) + \theta(\xi-|x|)\frac{1}{N_f}D\left(\frac{x}{\xi},t\right), \tag{1.52}$$

where the $\frac{1}{N_f}$ factor accounts for the identical contribution of each quark flavor to the D-term. The GPD E has a very similar parametrization with opposite sign D-term contribution. Note that $D(t,z)$ and the EMT FF introduced in Subsection 1.2.2 have the same name and are closely related. Indeed when performing the x-integration of the GPD H in the DD formalism, the only contribution to the second order term in ξ comes from the D-term and reads:

$$\int_{-1}^{1} dx\ x\ \theta(\xi-|x|)\frac{1}{N_f}D(\frac{x}{\xi},t) = \frac{1}{N_f}\xi^2\frac{4}{5}d_1(t), \tag{1.53}$$

where the orthogonality properties of the Gegenbauer polynomials are used to get the last relation (see Appendix A for more details on the calculation).

1.2.4 Beyond GPDs: GTMDs, TDAs

GPDs are part of a wide range of functions describing the nucleon, as shown in Figure 1.16. They can be related to 1-dimensional quantities such as PDFs and FFs by the relations described previously. In the same logic, GPDs can be extended. The Generalized Transverse Momentum Distributions (GTMDs) [48] are a generalization of GPDs. They describe the nucleon structure in terms of its constituent quarks transverse position, longitudinal and transverse momentum.

Another generalization of GPDs are the Transition Distribution Amplitudes (TDAs) [49, 50, 51]. These functions describe the transition from a baryon (the nucleon for example) to a meson. They are defined by a three-quarks operator similar to Equation (1.27), and can be accessed through backward-angle meson production [52].

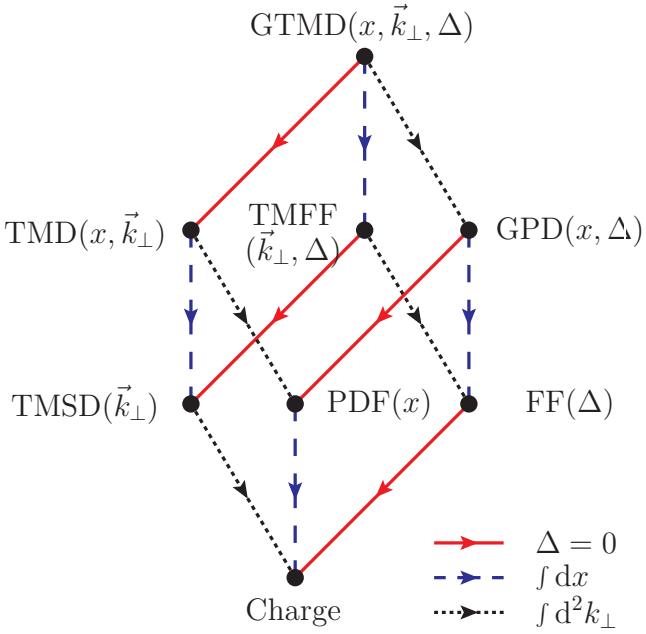


Figure 1.16: Structure functions scheme. GTMDs have an additional transverse momentum dependence compared to GPDs. TMDs describe the nucleon in terms of transverse momentum. TMFF refer to transverse-momentum form factors. TMSD are transverse-momentum spin densities. Figure taken from [48]

1.3 The experimental path to GPDs

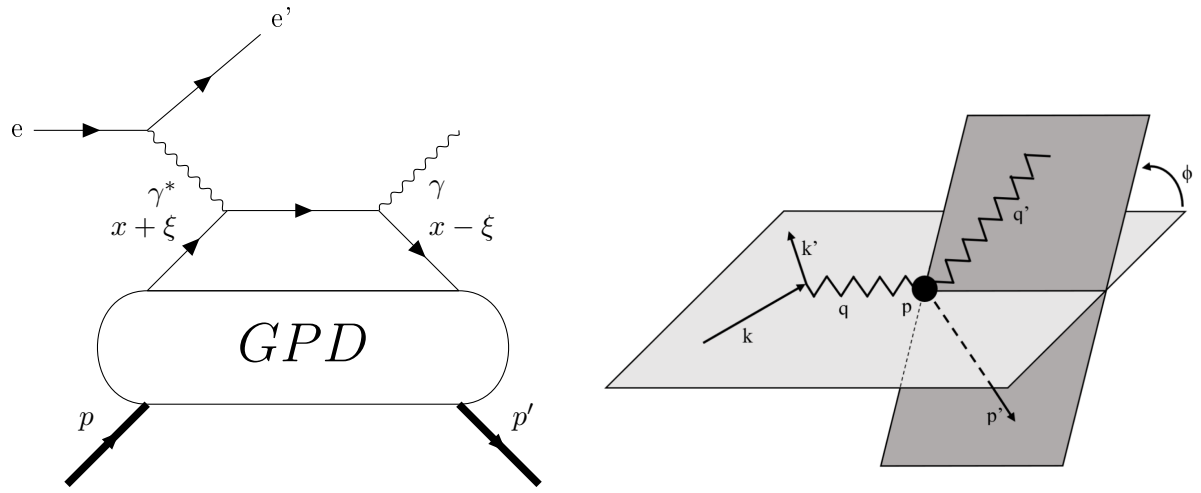
GPDs were first introduced to describe exclusive electro-production of photons and mesons. The cross sections of these processes can be fully described by these functions and the experimental measurement of these processes is therefore a direct way to study GPDs. In this section the experimental path towards the extraction of GPDs is introduced. First, Deeply Virtual Compton Scattering (DVCS) is presented in Subsection 1.3.1. Then one of its complementary process, Timelike Compton Scattering (TCS), is described in Subsection 1.3.2. Finally an overview of the current experimental status is provided.

1.3.1 Deeply Virtual Compton Scattering

As mentioned in the previous section, the simplest process to access GPDs is proton DVCS [53]: $ep \to e'\gamma^* p \to e'p'\gamma$. In the Bjorken regime (when the virtual photon γ^* has large virtuality $Q^2 = (k'-k)^2 \to \infty$ and large energy $\nu \to \infty$, where notations from Figure 1.17b are adopted), factorization can be applied. The virtual photon scatters off a single quark. At leading twist, leading order and in the Bjorken regime, the DVCS amplitude is a convolution of the hard scattering of a virtual photon off a single quark with the soft structure of the nucleon parameterized by GPDs. A diagram for the DVCS amplitude is shown in Figure 1.17a.

The DVCS cross section depends on four variables: $-t = (p'-p)^2$, Q^2, ϕ (the angle between the hadronic and leptonic planes defined in Figure 1.17b) and $x_B = \frac{Q^2}{2m\nu}$. In the Bjorken regime, the momentum skewness ξ is given by:

$$\xi = \frac{x_B}{2 - x_B}. \tag{1.54}$$



(a) (b)

Figure 1.17: (a) "Handbag" diagram for DVCS at leading twist and leading order. An electron interacts with a quark of the nucleon via the exchange of a virtual photon; this quark then emits a real photon. (b) The relevant kinematic variables for the $ep \to e'p'\gamma$ reaction.

The proton DVCS amplitude reads:

$$\mathcal{M} \propto \sum_q e_q^2 \Bigg\{ \int_{-1}^{1} dx\ x \left[\frac{1}{x-\xi+i\epsilon} + \frac{1}{x+\xi-i\epsilon} \right] \Big[A \cdot H^q(x,\xi,t) + B \cdot E^q(x,\xi,t) \Big] + \int_{-1}^{1} dx\ x \left[\frac{1}{x-\xi+i\epsilon} - \frac{1}{x+\xi-i\epsilon} \right] \Big[C \cdot \tilde{H}^q(x,\xi,t) + D \cdot \tilde{E}^q(x,\xi,t) \Big] \Bigg\}, \tag{1.55}$$

where the factors A,B,C and D are given in full details in [22, 54]. This amplitude depends on the proton GPDs defined as $H_p = \sum_q e_q^2 H^q$, where the sum runs over the quark flavors, and that we simply denote H in the following (the same definition applies for the other GPDs). One can also notice that the closed quark loop in the "handbag" diagram introduces an integral over the momentum fraction x. Therefore DVCS only accesses GPDs through such integrals, and the x-dependence of GPDs cannot be unfolded from DVCS observables. These integrals, depending only on ξ and t, are called Compton Form Factors (CFFs).

Compton Form Factors

Definition The CFFs are GPD integrals that appear naturally when integrating over the quark loop of the DVCS diagram. The CFFs depend on the process considered (see Subsection 1.3.2 for details on CFFs in the Timelike Compton Scattering channel). At leading order the DVCS CFFs are given by:

$$\{\mathcal{H}, \mathcal{E}\}(\xi,t) = \int_{-1}^{1} dx\ \{H, E\}(x,\xi,t) \left(\frac{1}{\xi - x - i\epsilon} - \frac{1}{\xi + x - i\epsilon} \right), \tag{1.56}$$

$$\left\{\tilde{\mathcal{H}}, \tilde{\mathcal{E}}\right\}(\xi,t) = \int_{-1}^{1} dx\ \left\{\tilde{H}, \tilde{E}\right\}(x,\xi,t) \left(\frac{1}{\xi - x - i\epsilon} + \frac{1}{\xi + x - i\epsilon} \right), \tag{1.57}$$

with the convention used in [55] (this convention includes an additional overall minus sign compared to the convention used in [22]).

CFFs are complex quantities. Consequently there are eight real functions to extract from DVCS experiments, the real and imaginary parts of the four CFFs. Figure 1.18 shows the area constrained by the real and imaginary part of CFFs for a given kinematic point. While the imaginary part is sensitive to the GPDs on the diagonals $x = \xi$ and $x = -\xi$, the real part provides information about the whole x range at given ξ.

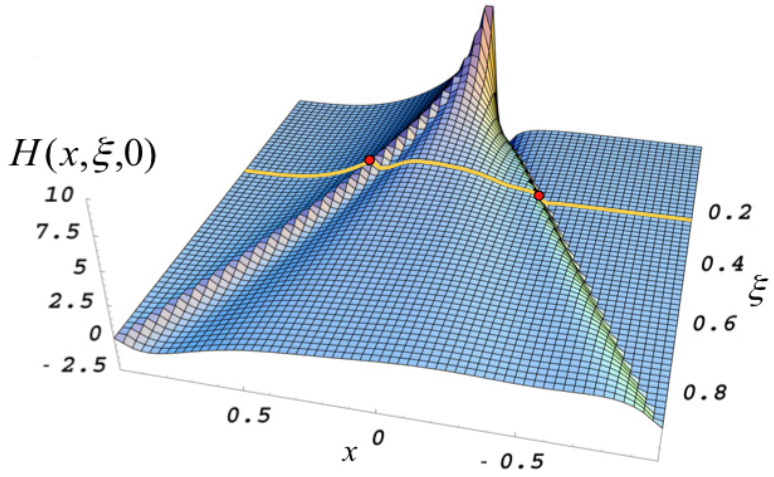


Figure 1.18: GPD H from the VGG model. The red points show regions which are obtained via the imaginary part of CFFs. The orange line displays the region constrained by the real part. Figure taken from [22]

Dispersion relation and D-term The real part of the $\mathcal{H}$ CFFs can be written using a dispersion relation (DR) involving the imaginary part and a subtraction term $\Delta(t)$ as:

$$\mathrm{Re}\mathcal{H}(\xi,t) = \int_{-1}^{1} \left(\frac{1}{\xi - x} - \frac{1}{\xi + x}\right) \mathrm{Im}\mathcal{H}(\xi,t)\ dx + \Delta(t). \tag{1.58}$$

The CFF $\mathcal{E}$ follows the same DR, with opposite sign subtraction term. $\tilde{\mathcal{H}}$ and $\tilde{\mathcal{E}}$ also follow similar DRs, with no subtraction term. Assuming a DD parametrization including the D-term of the GPD H, one can show that the subtraction term reads:

$$\Delta(t) = \frac{2}{N_f} \int_{-1}^{1} dz \frac{D(z,t)}{(1-z)} = \frac{4}{N_f} \sum_{n \text{ odd}} d_n(t). \tag{1.59}$$

This relation allows to directly relate $\Delta(t)$ to the EMT FF $D(t)$ introduced in Subsection 1.2.2, assuming that only $d_1(t)$ contributes in the sum:

$$\Delta(t) \propto d_1(t) \propto D^Q(t), \tag{1.60}$$

where D^Q denotes the quark contribution to the EMT FF $D(t)$. Measuring both the imaginary and the real part of the CFFs allows to use the DR as a tool to extract the substraction term $\Delta(t)$ and have access to the mechanical properties of the proton. The proton D-term has been extracted from data in [56], and the result is shown alongside other models and calculations in Figure 1.19, extracted from [34].

Bethe-Heitler process

The $ep \to e'p'\gamma$ reaction has not only contributions from DVCS but also from the Bethe-Heitler (BH) process. BH corresponds to the reaction where the initial (or final) state electron radiates a real photon that interacts with the proton as a whole. The two leading order diagrams of BH are shown in Figure 1.20. As the virtual photon interacts with the proton itself, BH is fully described using the usual FFs.

The DVCS and BH amplitudes add coherently when calculating the total cross section of the $ep \to e'p'\gamma$ process. The total cross section has contributions from both BH and DVCS and also from their interference and reads:

$$\sigma_{ep\to e'p'\gamma} = \sigma_{DVCS} + \sigma_{BH} + \sigma_{INT}. \tag{1.61}$$

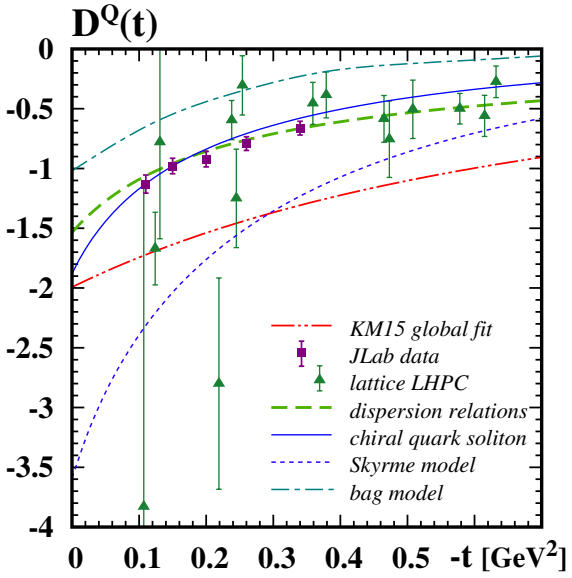


Figure 1.19: EMT FF $D(t)$ for the quarks as a function of the transferred momentum $-t$. The purple points are JLab data points [56] and the green dashed line is obtained using the CFF DR [37]. Figure taken from [34].

The relative importance of σ_{DVCS}, σ_{BH} and the interference term σ_{INT} depends strongly on the considered phase-space region. However the interference term between BH and DVCS can be used to extract CFFs conveniently, as they appear in σ_{INT} in a linear combination, whereas CFFs appear in bilinear combinations in σ_{DVCS}.

DVCS helicity-spin observables

Alongside the DVCS cross section and BH-DVCS interference terms, helicity-spin observables are an important way to extract CFFs. Experimentally one can change the helicity of the incoming electron, as well as the polarization of the proton target. It is then possible to measure cross-section asymmetries between different beam helicity-target spin configurations. Such asymmetries present two main advantages as they depend linearly on CFFs and are usually easier to extract than cross sections. Some DVCS asymmetries and their dependencies on the CFFs are given by:

$$\Delta\sigma_{LU} \propto \sin(\phi)\ \mathrm{Im}\left\{(F_1\mathcal{H} + \xi(F_1+F_2)\tilde{\mathcal{H}} - \frac{t}{4m_p^2}F_2\mathcal{E} + ...\right\}, \tag{1.62}$$

$$\Delta\sigma_{UL} \propto \sin(\phi)\ \mathrm{Im}\left\{(F_1\tilde{\mathcal{H}} + \xi(F_1+F_2)(\mathcal{H} + \frac{x_B}{2}\mathcal{E}) - \xi\left(\frac{x_B}{2}F_1 + \frac{t}{4m_p^2}F_2\right)\tilde{\mathcal{E}} + ...\right\}, \tag{1.63}$$

$$\Delta\sigma_{LL} \propto (A + B\cos(\phi))\ \mathrm{Re}\left\{(F_1\tilde{\mathcal{H}} + \xi(F_1+F_2)(\mathcal{H} + \frac{x_B}{2}\mathcal{E}) - \xi\left(\frac{x_B}{2}F_1 + \frac{t}{4m_p^2}F_2\right)\tilde{\mathcal{E}}...\right\}, \tag{1.64}$$

$$\Delta\sigma_{Ux} \propto \sin(\phi)\ \mathrm{Im}\left\{\frac{-t}{4m_p^2}(F_2\mathcal{H} - F_1\mathcal{E}) + ...\right\}, \tag{1.65}$$

where $\Delta\sigma$ denotes the difference of polarized cross sections. The first index refers to the beam polarization, the second one to the target spin. The index U stands for unpolarized, L for longitudinally polarized, x for transversely polarized. The dots stand for higher-twist terms. These asymmetries are mostly sensitive to the BH-DVCS interference cross section, thus they are sensitive to linear combinations of CFFs convoluted with FFs. These asymmetries also depend on different combinations of CFFs. A complete experimental measurement of these observables can thus provide a way to extract each contribution.

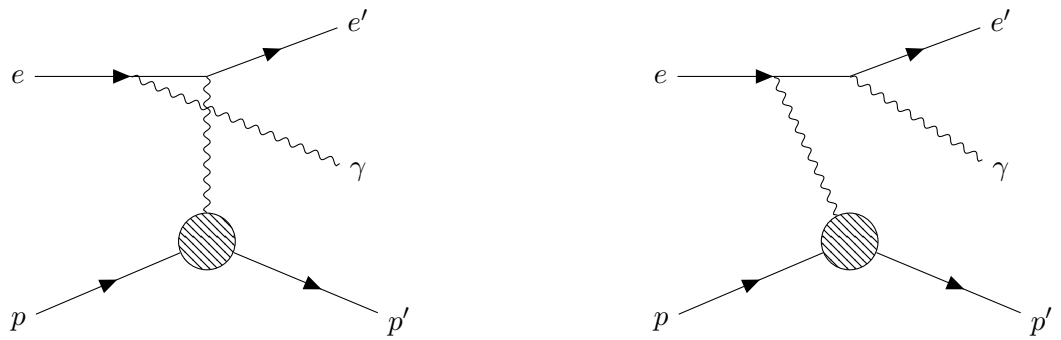


Figure 1.20: Bethe-Heitler contribution to the $ep \to e'p'\gamma$ reaction. In DVCS, the real photon is emitted by the proton. In BH, the real photon is emitted by either (a) the incoming or (b) the outgoing electron.

1.3.2 Timelike Compton Scattering

GPDs have been experimentally studied mainly through DVCS polarization obervables. Such observables are mainly sensitive to the imaginary part of the CFFs. To access the real part of CFFs, DVCS doubly polarized beam-target asymmetries (Equation (1.64)) or unpolarized cross sections are needed. It is also possible to access the real part of CFFs through the time-reversal symmetric process of the DVCS: the Timelike Compton Scattering (TCS)[57, 58, 59].

Measuring TCS observables also provides a test for the universality of GPDs. Indeed photon-polarization dependent cross section of TCS is sensitive to the imaginary part of CFFs. Comparing the results obtained in TCS and DVCS will help proving that GPDs are universal functions and are not only related to DVCS. In this subsection, the TCS process is presented. The associated BH process is described and the cross-section formulae are provided.

Phenomenology of TCS

TCS is the time-reversal symmetric process of DVCS. The reaction of interest is $\gamma p \to \gamma^* p'$, where the incoming photon is real ($Q^2 = 0$) and the outgoing photon is virtual. The virtual photon decays in a lepton pair which can be detected. The full reaction is therefore $\gamma p \to p' l^+ l^-$. Note that in the following we refer to the $\gamma p \to p' l^+ l^-$ reaction directly as the *TCS reaction*. Contrary to DVCS, where the large spacelike virtuality of the incoming photon gives a hard scale which ensures factorization, the TCS hard scale is given by the timelike virtuality Q'^2 of the outgoing photon. For large Q'^2 such that $\frac{t}{Q'^2} \ll 1$, factorization can be applied. The real incoming photon scatters off a single quark, which emits a virtual photon. The leading order, leading twist diagram for TCS is given in Figure 1.21.

The relevant variables to study the TCS reaction are shown in Figure 1.22. They are the virtuality of the outgoing photon $Q'^2 = (k + k')^2$, the transferred momentum to the nucleon $t = (p' - p)^2$, the γp Center-of-Mass (COM) energy $s = (p + q)^2$ (or equivalently the real photon energy E_γ), the azimuthal angle ϕ between the leptonic plane and the hadronic plane, and the angle of the outgoing electron in the lepton COM frame, θ. Finally, as for DVCS, the quark momentum imbalance can be defined as:

$$\xi = \frac{\tau}{2 - \tau}, \tag{1.66}$$

where $\tau = Q'^2/(s - m_p^2)$, which plays the symmetric role of x_B in DVCS (see Equation (1.54)).

As for DVCS, a Bethe-Heitler (BH) process also contributes to the $\gamma p \to p' l^+ l^-$ reaction and interferes with TCS. Figure 1.23 displays the two timelike BH diagrams, where the real photon decays in a lepton pair, from which one lepton then interacts with the proton.

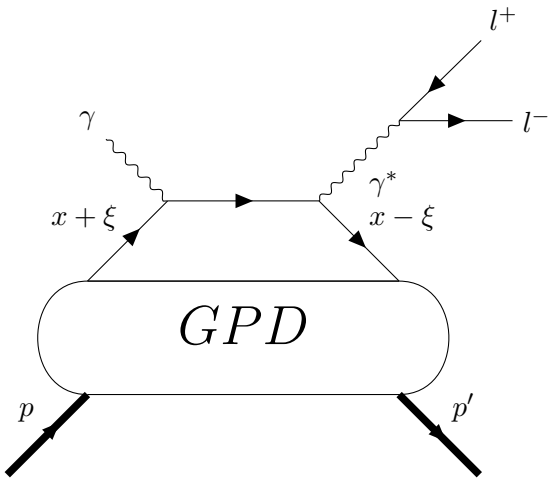


Figure 1.21: The Feynman diagram for TCS in the reaction $\gamma p \to p' l^+ l^-$

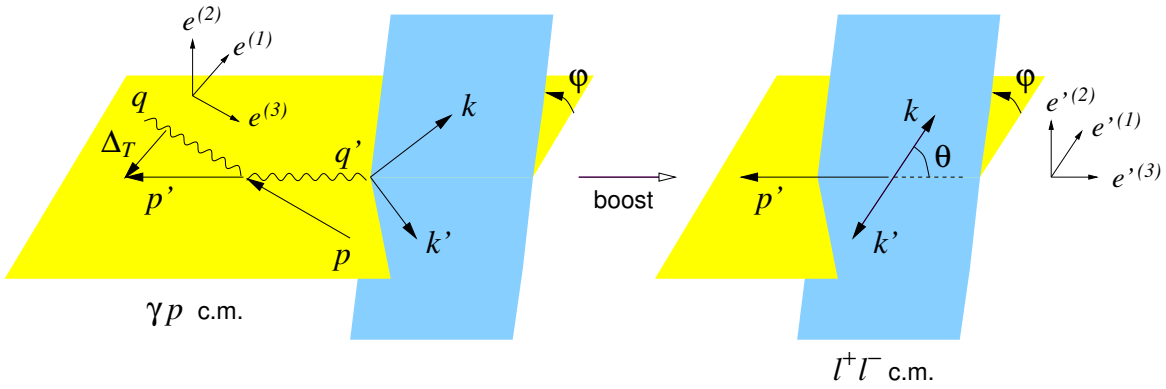


Figure 1.22: Frame definition and relevant variables for TCS, taken from [57]. The yellow plane containing the momenta of the target and recoil protons is called the hadronic plane. The blue plane containing the momenta of the two leptons is the leptonic plane. The angle between these planes is called ϕ. The angle θ is defined as the angle between the lepton with momentum k and the direction of the recoil proton momentum in the lepton pair COM frame.

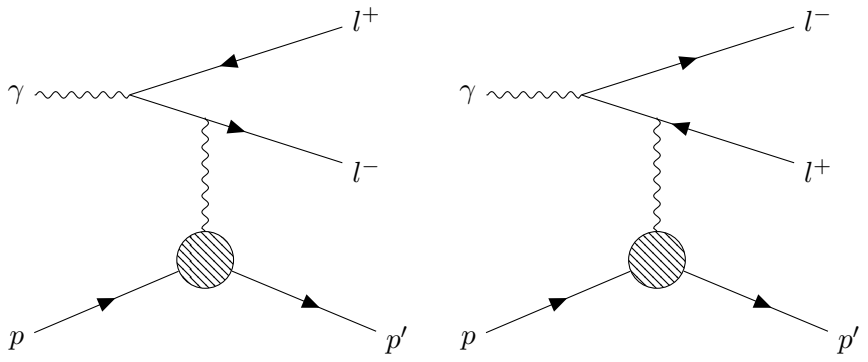


Figure 1.23: Feynman diagrams for the Bethe-Heitler processes that contribute to the $\gamma p \to p' l^+ l^-$ reaction.

The CFFs of TCS

Similarly to DVCS, the quark loop in the TCS diagram forces GPDs to appear inside CFFs in the amplitude. The TCS CFFs have similar forms to the DVCS CFFs defined in Equations (1.56) and

(1.57) up to a sign and a complex conjugation. The two kinds of CFFs are related by the following relations:

$$\begin{aligned} \mathcal{H}_{TCS} &= \mathcal{H}^*_{DVCS} \qquad \tilde{\mathcal{H}}_{TCS} = -\tilde{\mathcal{H}}^*_{DVCS} \\ \mathcal{E}_{TCS} &= \mathcal{E}^*_{DVCS} \qquad \tilde{\mathcal{E}}_{TCS} = -\tilde{\mathcal{E}}^*_{DVCS} \end{aligned} \tag{1.67}$$

In this section we use TCS CFFs unless specified otherwise.

TCS cross section

Unpolarized cross section The unpolarized cross section for $\gamma p \to p' l^+ l^-$ can be expanded as:

$$d^4\sigma(\gamma p \to p' e^+ e^-) = d^4\sigma_{BH} + d^4\sigma_{TCS} + d^4\sigma_{INT}. \tag{1.68}$$

Each term is written explicitly according to the formulas given in [57] in the following. The BH cross section is parametrized by FFs only. It reads:

$$\frac{d^4\sigma_{BH}}{dQ'^2 dt d(\cos\theta) d\phi} = \frac{\alpha_{em}^3}{4\pi(s-m_p^2)^2}\,\frac{\beta}{-tL}\left[(F_1^2 - \frac{t}{4m_p^2}F_2^2)\frac{A}{-t} + (F_1+F_2)^2\,\frac{B}{2}\right], \tag{1.69}$$

where

$$\begin{aligned} A &= (s-m_p^2)^2\Delta_T^2 - t\,a(a+b) - m_p^2 b^2 - t\,(4m_p^2 - t)Q'^2 \\ &\quad + \frac{m_l^2}{L}\left[\Big\{(Q'^2 - t)(a+b) - (s-m_p^2)\,b\Big\}^2 + t\,(4m_p^2 - t)(Q'^2 - t)^2\right] \\ B &= (Q'^2 + t)^2 + b^2 + 8m_l^2 Q'^2 - \frac{4m_l^2(t+2m_l^2)}{L}\,(Q'^2 - t)^2, \\ \beta &= \sqrt{1 - 4m_l^2/Q'^2}, \end{aligned} \tag{1.70}$$

with m_l the lepton mass and

$$a = 2(k-k')\cdot p', \tag{1.71}$$

$$b = 2(k-k')\cdot(p-p'), \tag{1.72}$$

$$L = \frac{(Q'^2 - t)^2 - b^2}{4}. \tag{1.73}$$

The BH cross section is plotted in Figure 1.24 for different θ and ϕ. One can see that the cross section is largely enhanced around $\phi = 0°$ for high values of θ, and around $\phi = 180°$ for low values of θ.

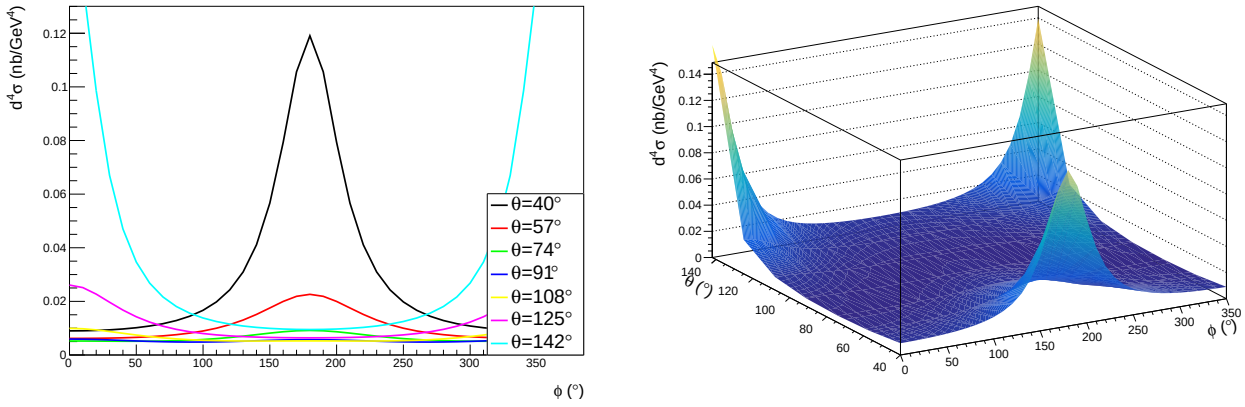
The TCS contribution reads:

$$\frac{d\sigma_{TCS}}{dQ'^2 dt d(\cos\theta) d\phi} \approx \frac{\alpha_{em}^3}{8\pi s^2}\,\frac{1}{Q'^2}\,\frac{1+\cos^2\theta}{4}\sum_{\lambda,\lambda'}|M^{\lambda' -,\lambda -}|^2, \tag{1.74}$$

where

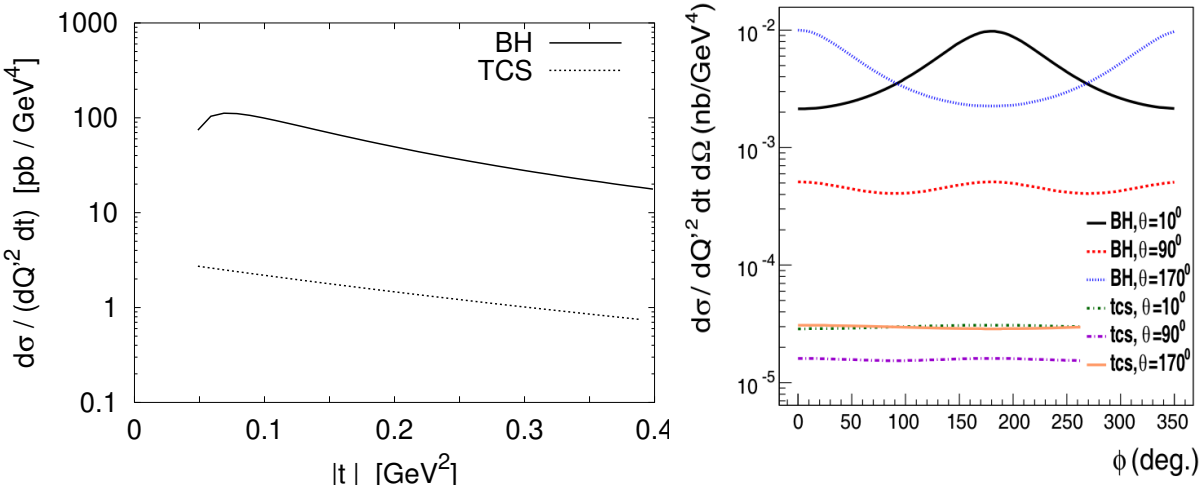
$$\begin{aligned} \frac{1}{2}\sum_{\lambda,\lambda'}|M^{\lambda' -,\lambda -}|^2 &= (1-\xi^2)\Big(|\mathcal{H}|^2 + |\tilde{\mathcal{H}}|^2\Big) - 2\xi^2\,\mathrm{Re}\Big(\mathcal{H}^*\,\mathcal{E} \;+ \tilde{\mathcal{H}}^*\,\tilde{\mathcal{E}}\;\Big) \\ &\quad - \Big(\xi^2 + \frac{t}{4M^2}\Big)\,|\mathcal{E}|^2 - \xi^2\frac{t}{4M^2}\,|\tilde{\mathcal{E}}|^2. \end{aligned} \tag{1.75}$$

Figures 1.25a and 1.25b show the comparison between the TCS cross section and the BH one. The TCS contribution to the total cross section is two orders of magnitude less important than the BH



(a) (b)

Figure 1.24: BH cross section at the kinematic point $Q'^2 = 4$ GeV^2, $\xi = 0.1$ and $-t = 0.2$ GeV^2, (a) for different values of θ as a function of ϕ, (b) as a function of both θ and ϕ.



(a) (b)

Figure 1.25: (a) BH (solid line) and TCS (dotted line) contributions to the total $\gamma p \to p' e^+ e^-$ cross section for $\sqrt{s} = 5$ GeV and $Q'^2 = 5$ GeV^2, integrated over ϕ in $[0, 2\pi]$ and θ over $[\pi/4, 3\pi/4]$. Figure extracted from [57]. (b) BH and TCS cross sections for different values of θ, as a function of ϕ. Figure extracted from [58].

one. Therefore measuring the TCS cross section is very challenging. One way to avoid this issue is to extract the BH-TCS interference term. This term reads:

$$\frac{d^4\sigma_{INT}}{dQ'^2 dt d\Omega} = -\frac{\alpha_{em}^3}{4\pi s^2}\frac{1}{-t}\frac{m_p}{Q'}\frac{1}{\tau\sqrt{1-\tau}}\frac{L_0}{L}[\cos(\phi)\frac{1+\cos^2(\theta)}{\sin(\theta)}\mathrm{Re}\tilde{M}^{--} - \cos(2\phi)\sqrt{2}\cos(\theta)\mathrm{Re}\tilde{M}^{0-} + \cos(3\phi)\sin(\theta)\mathrm{Re}\tilde{M}^{+-} + O(\frac{1}{Q'})], \quad (1.76)$$

where

$$L_0 = \frac{Q'^2\sin^2(\theta)}{4}, \quad (1.77)$$

and the $\tilde{M}$ terms are CFFs combinations defined in [60]. At leading order and leading twist, the only term contributing to the cross section is the one proportional to $\tilde{M}^{--}$. The CFF dependence of $\tilde{M}^{--}$

is given by:

$$\tilde{M}^{--} = \frac{2\sqrt{t_0 - t}}{m_p} \frac{1-\xi}{1+\xi} \left[F_1 \mathcal{H} - \xi(F_1 + F_2)\tilde{\mathcal{H}} - \frac{t}{4m_p^2} F_2 \mathcal{E} \right], \tag{1.78}$$

where

$$t_0 = \frac{4\xi^2 m_p^2}{(1-\xi^2)}. \tag{1.79}$$

Furthermore, the dominant term in Equation (1.78) at JLab kinematics is the one containing the CFF $\mathcal{H}$ ($\tilde{\mathcal{H}}$ is suppressed by a factor $\xi \approx 0.3$, $\mathcal{E}$ is suppressed by a factor $\frac{t}{4m_p^2} < 0.25$). The $\cos(\phi)$ modulation of the interference cross section thus gives direct access to the real part of the CFF $\mathcal{H}$.

Transversely polarized photon cross section As mentionned for DVCS, the helicity-spin observables are a powerful tool to extract the imaginary part of CFFs. The same reasoning can be applied to TCS. In the case of transversely polarized photons, the interference cross section is expressed as:

$$\frac{d\sigma_{INT}}{dQ'^2 dt d(\cos\theta) d\phi} = \frac{d\sigma_{INT}\,|_{\text{unpol.}}}{dQ'^2 dt d(\cos\theta) d\phi} - \nu \frac{\alpha_{em}^3}{4\pi s^2} \frac{1}{-t} \frac{M}{Q'} \frac{1}{\tau\sqrt{1-\tau}} \frac{L_0}{L} \left[\sin(\phi) \frac{1 + cos^2(\theta)}{sin(\theta)} Im\tilde{M}^{--} \right.$$
$$\left. - \sin(2\phi)\sqrt{2} cos\theta Im\tilde{M}^{0-} + \sin(3\phi) sin(\theta) Im\tilde{M}^{+-} + O\left(\frac{1}{Q'}\right) \right], \tag{1.80}$$

where ν is the circular polarization of the incoming real photon. The additional polarization term exhibits the same CFF content as the unpolarized cross section, except that it now depends on the imaginary parts via $\sin(n\phi)$ factors. Extracting the $\sin(\phi)$ component of the polarized cross section enables to access the imaginary part of $\mathcal{H}$. This is an important test of the universality of GPDs once compared with DVCS data.

1.3.3 TCS observables

In this subsection, three TCS observables, the R ratio and the Forward/Backward asymmetry sensitive to the real parts of the CFFs, and the photon polarization asymmetry sensitive to the imaginary parts of the CFFs, are presented.

R ratio

The R ratio, introduced in [57], is defined as:

$$R(\sqrt{s}, Q'^2, t) = \frac{\int_0^{2\pi} d\phi \cos(\phi) \frac{dS}{dQ'^2 dt d\phi}}{\int_0^{2\pi} d\phi \frac{dS}{dQ'^2 dt d\phi}}, \tag{1.81}$$

where

$$\frac{dS}{dQ'^2 dt d\phi} = \int_{\pi/4}^{3\pi/4} d\theta \frac{L}{L_0} \frac{d\sigma}{dQ'^2 dt d\phi d\theta}. \tag{1.82}$$

It has to be noted that the definition used in this manuscript differs by a factor 2 from the original definition. This was chosen to be consistent with the CLAS TCS analysis in [59].

The R ratio is directly sensitive to the real part of $\tilde{M}^{--}$. The integration domain is set to $[\pi/4, 3\pi/4]$ to avoid kinematic regions where TCS is too small compared to BH. Furthermore the enhancing effect of multiplying by $\frac{L}{L_0}$ before the integration can be seen in Figure 1.26. Figure 1.27 shows a projection for the R ratio as a function of $-t$. The importance of the D-term is highlighted by showing the amplitude of the observable for different D-term hypotheses.

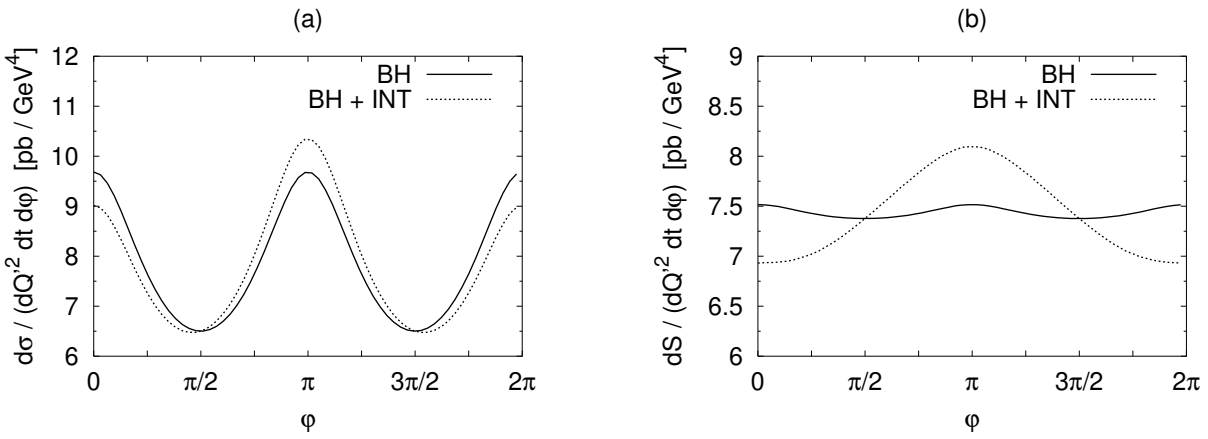


Figure 1.26: Contribution of BH (solid line) and BH+INT (dotted line) to the integrated cross section over θ in $[\pi/4, 3\pi/4]$ as a function of ϕ and for $\sqrt{s} = 5$ GeV, $Q'^2 = 5$ GeV^2 and $\mid t \mid = 0.2$ GeV^2. (a) is without the $\frac{L}{L_0}$ factor, (b) is including this factor. Figures extracted from [57].

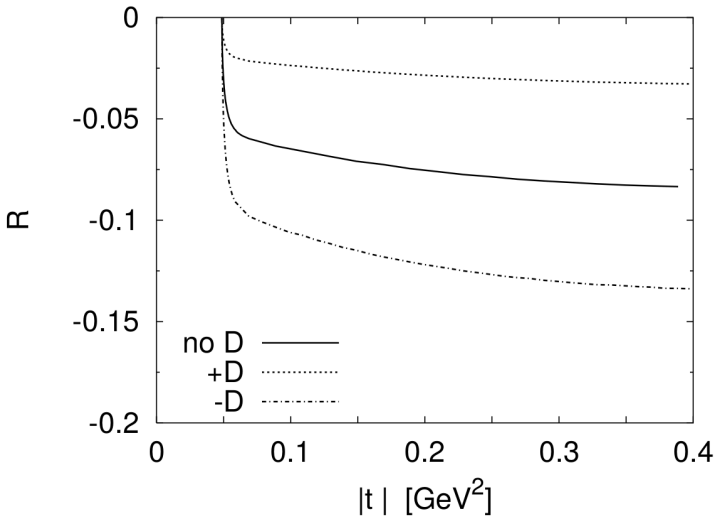


Figure 1.27: Projection for the R ratio for $\sqrt{(s)} = 5$ GeV and $Q'^2 = 5$ GeV^2. The dotted line represents the impact of adding the D-term in the GPD model used. The dotted-dashed line is obtained by subtracting the D-term. This figure shows the sensitivity of the R ratio to the D-term. Figure extracted from [57].

Forward-Backward asymmetry

The idea of Forward-Backward Asymmetry (A_{FB}) was initially proposed for J/Ψ threshold photoproduction studies in [61]. The A_{FB} is defined as:

$$A_{FB}(\theta, \phi) = \frac{d\sigma(\theta, \phi) - d\sigma(180^\circ - \theta, 180^\circ + \phi)}{d\sigma(\theta, \phi) + d\sigma(180^\circ - \theta, 180^\circ + \phi)} \tag{1.83}$$

where only the θ-ϕ dependence of the cross section is explicitly written.

The transformation ($\theta \to 180^\circ - \theta, \phi \to \phi + 180^\circ$) corresponds to inverting the vectors k and k' in the COM frame of the lepton pair. This transformation leaves both L (Equation (1.73)) and L_0 (Equation (1.77)) unchanged, as the term b appears squared in L. Moreover both $\frac{d^4\sigma_{BH}}{dQ'^2 dt d\Omega}$ and $\frac{d^4\sigma_{TCS}}{dQ'^2 dt d\Omega}$ remain unchanged under this transformation. Assuming leading order and leading twist, only the $\cos(\phi)\tilde{M}^{--}$ term contributes in Equation (1.76) and the interference cross section is transformed as:

$$\frac{d^4\sigma_{INT}}{dQ'^2 dt d\Omega} \xrightarrow{FB} -\frac{d^4\sigma_{INT}}{dQ'^2 dt d\Omega}. \tag{1.84}$$

Finally the A_{FB} can be explicitly written as:

$$A_{FB}(\theta_0,\phi_0) = \frac{-\frac{\alpha_{em}^3}{4\pi s^2}\frac{1}{-t}\frac{m_p}{Q'}\frac{1}{\tau\sqrt{1-\tau}}\frac{L_0}{L}\cos\phi_0\frac{(1+\cos^2\theta_0)}{\sin(\theta_0)}\mathrm{Re}\tilde{M}^{--}}{d\sigma_{BH}}, \tag{1.85}$$

where we neglect the TCS contribution in the denominator. This observable is sensitive to the same quantity as the R ratio. However, it is not integrated over a large phase space. It is therefore less sensitive to detector acceptance effects. First predictions for the TCS A_{FB}, realized with the VGG model, are presented in Section 5.9.

Photon polarization asymmetry

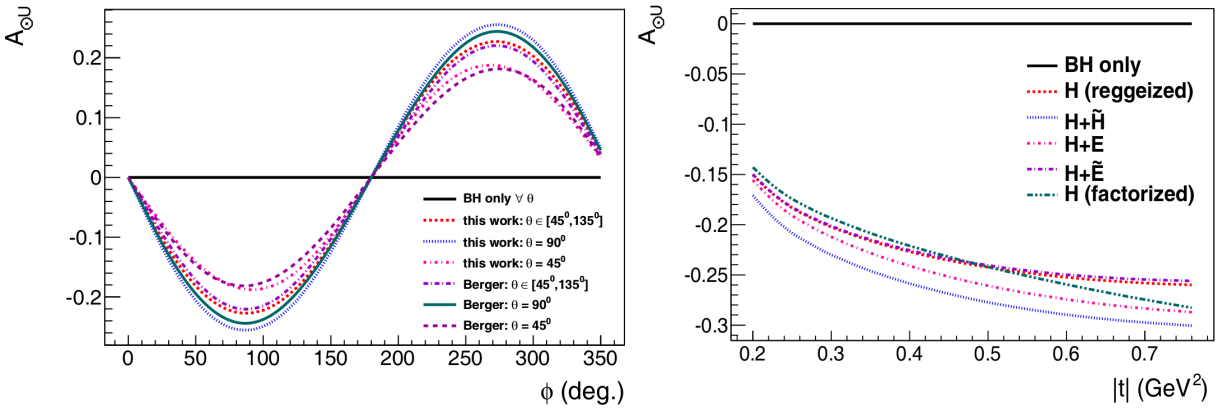
The photon polarization asymmetry $A_{\odot U}$ (also referred in the following as Beam Spin Asymmetry (BSA) because of its similarity with the DVCS BSA) is defined as:

$$A_{\odot U} = \frac{\sigma^+ - \sigma^-}{\sigma^+ + \sigma^-}, \tag{1.86}$$

where indexes $+(-)$ refer to the right(left)-handed circular polarization of the incoming real photon, and U to the unpolarized target. It can be written explicitly as:

$$A_{\odot U} = \frac{-\frac{\alpha_{em}^3}{4\pi s^2}\frac{1}{-t}\frac{m_p}{Q'}\frac{1}{\tau\sqrt{1-\tau}}\frac{L_0}{L}\sin\phi\frac{(1+\cos^2\theta)}{\sin(\theta)}\mathrm{Im}\tilde{M}^{--}}{d\sigma_{BH}}. \tag{1.87}$$

Projections for this observable have been made in [58]. The results obtained in this paper are shown in Figure 1.28. The asymmetry as a function of ϕ and as a function of $-t$ (only the amplitude at $\phi = 90°$ is then plotted) are shown in Figures 1.28a and 1.28b, respectively. Similar predictions, made during this thesis with the VGG model, are shown in Figure 1.29. The $-t$ and E_γ dependence are studied for different values of θ. Note that the sign change between the two predictions. This is due to a sign mistake in [58] which is discussed in more details in Section 5.9.

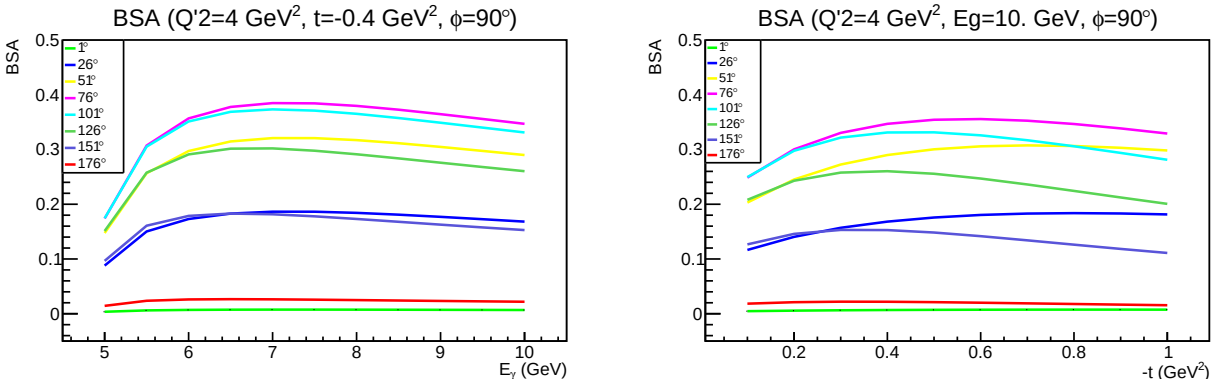


(a) (b)

Figure 1.28: (a) $A_{\odot U}$ as a function of ϕ for BH and BH+Int for different cases of integration over θ. (b) Amplitude of the $A_{\odot U}$ as a function of $-t$. In this case only the amplitude of $A_{\odot U}$ at $\phi = 90°$ is plotted. Figures from [58].

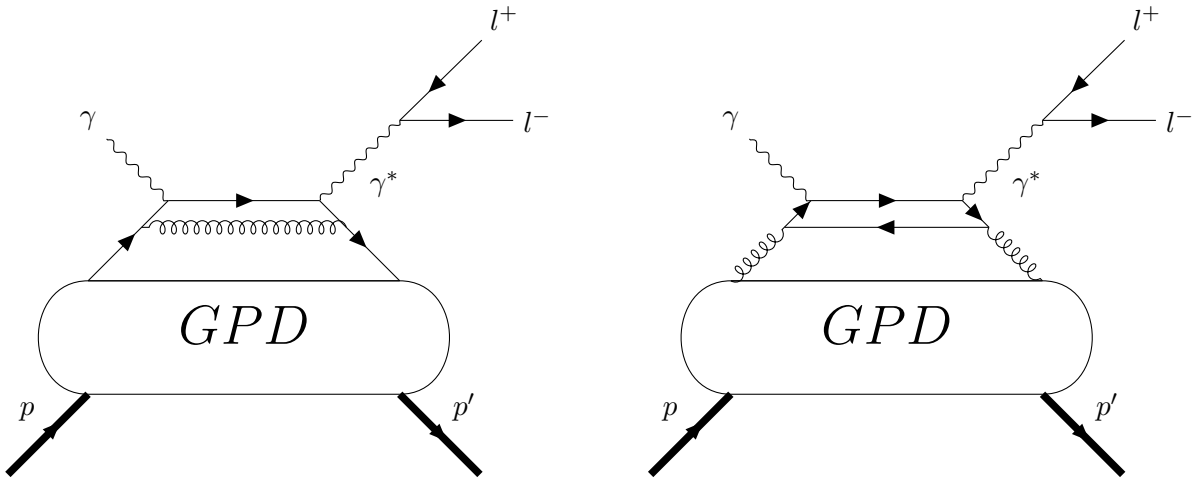
1.3.4 NLO corrections to the TCS amplitude

One last aspect to mention for the TCS process is its sensitivity to NLO corrections [62, 63, 64, 65]. NLO corrections in α_S can be represented by the diagrams in Figure 1.30. These corrections appear when one considers a hard gluon in the hard part of the diagram (Figure 1.30a) or including gluon GPDs (Figure 1.30b).



(a) (b)

Figure 1.29: (a) Amplitude of the $A_{\odot U}$ at $\phi = 90°$ as a function of E_γ. (b) Amplitude of the $A_{\odot U}$ as a function of $-t$, for different values of θ.



(a) (b)

Figure 1.30: Examples of two diagrams contributing at NLO in the TCS amplitude: (a) involving quark GPDs, (b) involving gluon GPDs.

To account for these corrections, the definition of the CFFs is generalized as:

$$\mathcal{H}(\xi, t, \mathcal{Q}^2) = \int_{-1}^{1} dx \sum_{i=u,d,\cdots,g} T^i(x,\xi) H^i(x,\xi,t,\mu^2), \tag{1.88}$$

where similar equations can be written for other GPDs, μ^2 is the factorization scale and $\mathcal{Q}$ denotes the hard scale, Q^2 for DVCS and Q'^2 for the TCS case. The T^i hard coefficients are, at leading order, defined in Equations (1.56) and (1.57). At NLO the hard coefficients for TCS and DVCS are different [62]. This difference leads to the NLO correspondence between TCS and DVCS CFFs:

$$\mathcal{H}_{TCS} \overset{\text{NLO}}{=} \mathcal{H}^*_{DVCS} - i\pi\, \mathcal{Q}^2 \frac{\partial}{\partial \mathcal{Q}^2} \mathcal{H}^*_{DVCS}\,, \tag{1.89}$$

$$\widetilde{\mathcal{H}}_{TCS} \overset{\text{NLO}}{=} -\widetilde{\mathcal{H}}^*_{DVCS} + i\pi\, \mathcal{Q}^2 \frac{\partial}{\partial \mathcal{Q}^2} \widetilde{\mathcal{H}}^*_{DVCS}\,, \tag{1.90}$$

where similar equations apply for $\mathcal{E}$ and $\tilde{\mathcal{E}}$. Because of the $i\pi$ factor in Equation (1.89), the real part of CFF $\mathcal{H}_{TCS}$ has a contribution from the imaginary part of $\mathcal{H}_{DVCS}$, which can be sizable. This fact is illustrated in Figure 1.31, where the LO and NLO real parts of $\mathcal{H}$ are plotted against ξ for the TCS and DVCS cases. A very large effect of the NLO correction is reported at low ξ (below 10^{-2}) where the sign of the CFF $\mathcal{H}$ for TCS is expected to flip.

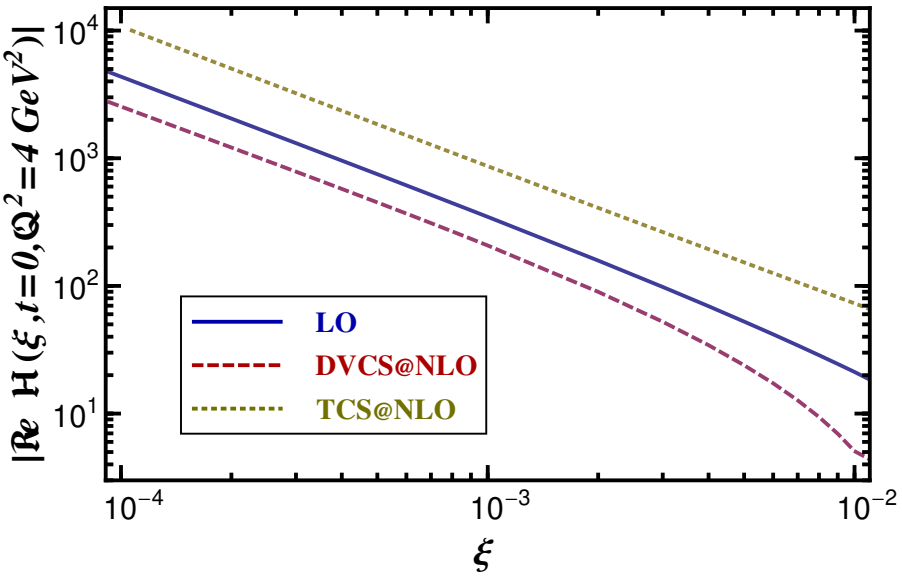


Figure 1.31: Real part of the $\mathcal{H}$ CFFs at LO and NLO in the DVCS and TCS cases. In the TCS case, the opposite sign value is plotted. Figure from [62].

1.3.5 TCS experimental status

There are currently no published data on TCS. However a complete analysis [59] was performed using the CLAS detector at Jefferson Lab. It used a 5.7 GeV electron beam impinging on a liquid-hydrogen target. Quasi-real photoproduction events where selected using the missing-mass technique. The results for the R ratio obtained in this analysis are shown in Figure 1.32. The ratio is calculated within the CLAS acceptance and compared to theoretical predictions. The DD parametrization of GPDs seems to underestimate the data, while the Dual parametrization seems to reproduce them well. However these results should be interpreted with care as the invariant mass range accessible by CLAS covered a region where vector meson production is not negligible.

Following the 12-GeV upgrade of the CEBAF accelerator, the newly built CLAS12 detector (see Chapter 2) reaches a wider range of lepton-pair masses. The results obtained by CLAS12 are presented in Chapter 6. Other experiments are planned in Hall A (with the SOLID detector) and Hall D at JLab. Finally there is a growing interest to measure TCS at higher energies in Ultra-peripheral collisions and possible measurements could be performed at CERN [66].

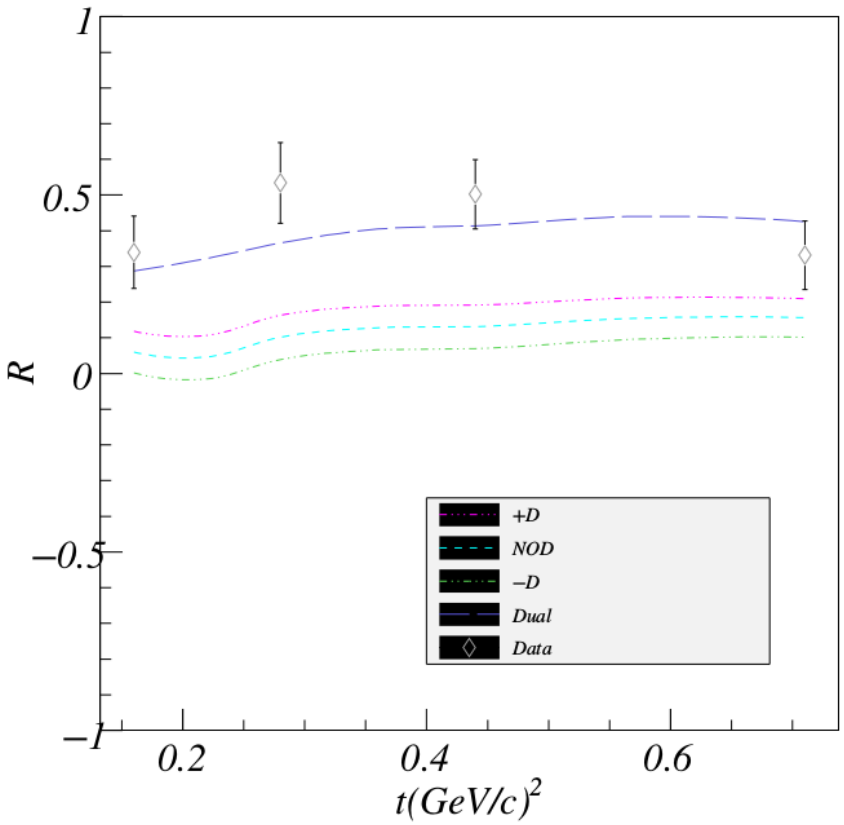


Figure 1.32: R ratio measured in the CLAS acceptance. Data points are compared with DD and Dual GPDs parametrizations.

Chapter 2

The experimental setup

The experimental activities described in this manuscript are performed at the Thomas Jefferson National Accelerator Facility. Its main asset is a nearly continuous polarized electron beam used to probe the structure of matter at the femto-meter scale. In this chapter the experimental setup is described. First an overview of the laboratory is given, followed by a detailed description of the CLAS12 detector (Sections 2.3, 2.4, 2.5 and 2.6) and associated softwares (Section 2.7). Finally the data set used in this manuscript is presented in Section 2.8.

2.1 Jefferson Lab

The Thomas Jefferson National Accelerator Facility (commonly referred as JLab) is a physics laboratory run by the U.S. Department of Energy, in Virginia, USA. It is located in Newport News, 300 kilometres south of Washington DC. More than 600 people work at JLab and more than 2000 scientists collaborate to run experiments in the laboratory. It is devoted to the study of the strong interaction since its foundation in 1984.

The Jefferson Lab physics program relies on the Continuous Electron Beam Accelerator Facility (CEBAF), an electron accelerator that provides a continuous electron beam to four experimental halls (Hall A to D). Each hall has a dedicated physics program and detectors. Hall A and C are equipped with narrow acceptance spectrometers and can handle large luminosities (of the order of $10^{38}\text{cm}^{-2}\text{s}^{-1}$). Hall D is dedicated to hadron spectroscopy and has a dedicated photon beamline. Finally, Hall B houses the CLAS12 (CEBAF Large Acceptance Spectrometer 12) detector described in the following. In 2009 the laboratory started an upgrade program that has led to the construction of the fourth experimental hall (Hall D), to the doubling of the beam energy from 6 GeV to 12 GeV and to the upgrade of the Hall B detector CLAS to the current CLAS12 detector. The schematic description of the CEBAF and JLab experimental halls is shown in Figure 2.1.

2.2 The Continuous Electron Beam Accelerator Facility

The core facility at Jefferson Lab is the CEBAF. This superconducting radio-frequency accelerator delivers an electron beam with energy up to 12 GeV for Hall D and 11 GeV for all three other halls. The beam electrons are produced at the injector where four circularly polarized lasers (one for each hall) illuminate a gallium arsenide cathode. The polarization of the lasers allows to extract polarized electrons from the cathode. The lasers operate at 250 MHz with an individual phase shift in order to distinguish each beam bunch. The power of each laser can be controlled individually allowing each hall to have its own beam current value. The beam is then injected to the accelerator. The accelerator is composed of two linacs with 25 cryomodules each. Each cryomodule is divided in 8 radio-frequency (RF) cavities which are synchronized at the total frequency of the accelerator ($250\times6=1500$ MHz). The two linacs are connected to each other by recirculation arcs on both ends. A different recirculating arc is used at each pass of the beam bunch in the accelerator. The maximum beam energy delivered in Hall A, B and C ($\sim$ 11 GeV) is obtained after a total of five passes in the whole accelerator. The beam is then split by two separators, working respectively at 750 MHz and 500 MHz (see [67] for more

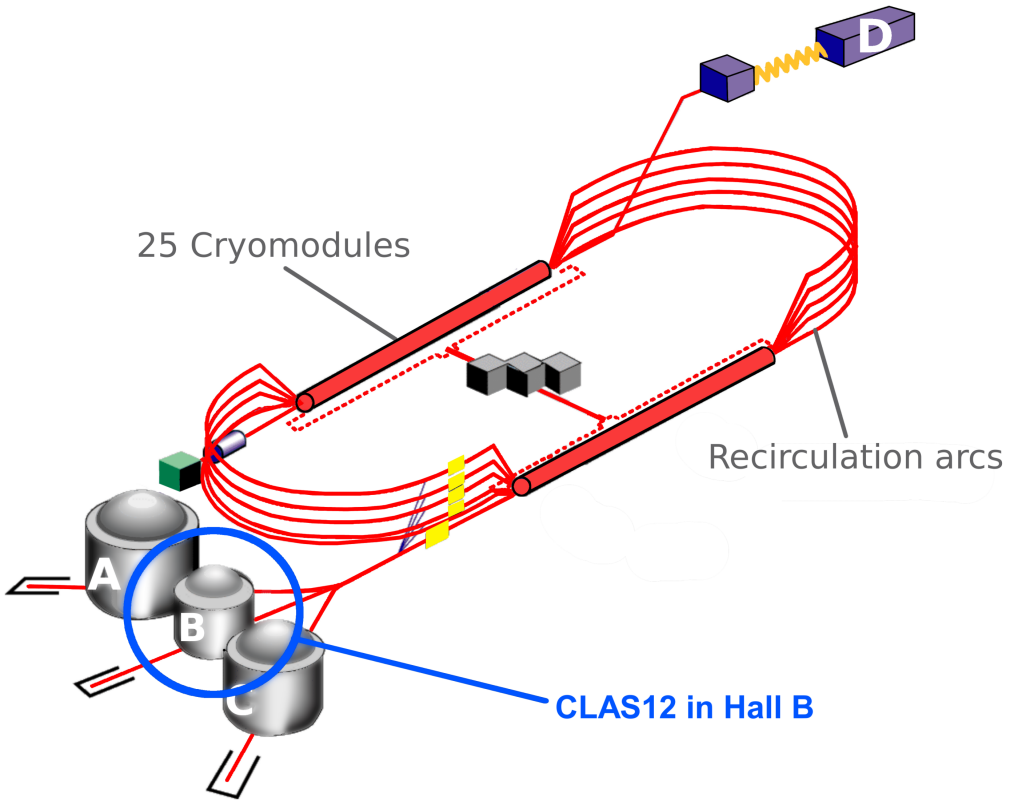


Figure 2.1: Schematic description of the CEBAF. The accelerator consists of two linacs connected to each other by magnetic recirculating arcs. The accelerator delivers electron bunches alternatively to each experimental hall. Hall D was built during the 12 GeV upgrade. Detectors in Hall A, B and C were upgraded and each linac was extended with five new cryomodules in order to increase the beam energy. The CLAS12 detector, described in Section 2.3, is located in Hall B.

details on the beam structure), providing each hall with a 250 MHz electron beam (one beam bunch every 4 ns). In Hall B, the maximum beam energy obtained during the data taking was 10.6 GeV and the nominal luminosity delivered by the CEBAF was $10^{35}\text{cm}^{-2}\text{s}^{-1}$.

2.3 CLAS12 general design

The CLAS12 detector is housed in Hall B [68]. It has a very large acceptance allowing to measure inclusive and exclusive processes over a large phase space. CLAS12 is the successor of the first detector of Hall B, the CLAS detector. Some detector systems of CLAS have been refurbished for CLAS12. CLAS12 is composed of two main subsets of detectors: the Central Detector (CD) and the Forward Detector (FD). The FD uses existing parts of the CLAS detector and newly built detectors. It detects particles with polar angles between 5° and 35° with respect to the beam direction. The CD is made of newly built sub-detectors and aims at measuring backward recoiling particles, with polar angles between 35° and 135°. In addition to the CD and the FD, the Forward Tagger (FT) covers very low polar angles and is dedicated to tagged photo-production measurements. It is situated close to the beam line downstream of the target. A solenoid magnet, with a central field of 5 T, and a torus magnet, with $\int Bdl = 0.5 - 2.7$ Tm, allow for charged-particle momentum measurements respectively in the CD and in the FD. The nominal luminosity of CLAS12 is $10^{35}\text{cm}^{-2}\text{s}^{-1}$, which is one order of magnitude greater than the luminosity of the previous CLAS detector. A schematic view of CLAS12 is shown in Figure 2.2. Table 2.1 extracted from [68] summarizes the nominal CLAS12 performances.

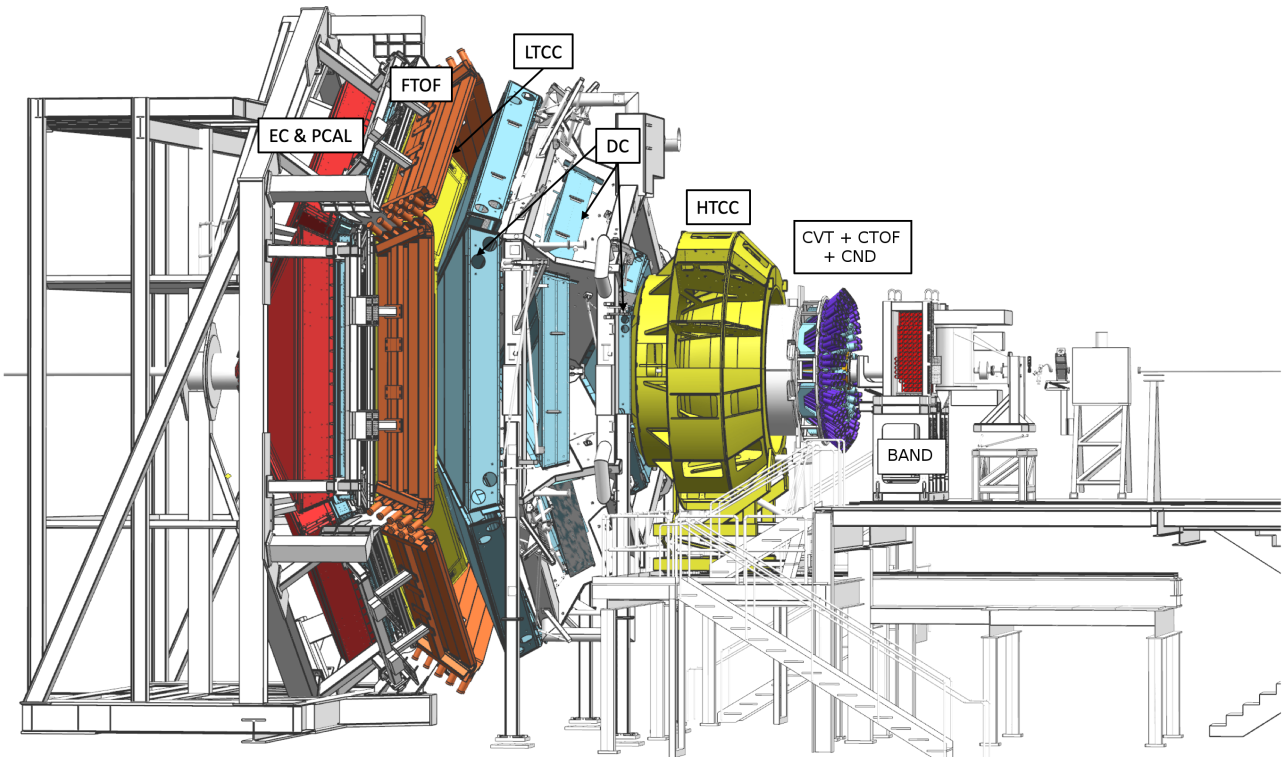


Figure 2.2: The CLAS12 detector in the Hall B. The electron beam provided by the CEBAF accelerator arrives from the right of the picture. It interacts with the target located at the center of the Central Detector (CD). Particles emitted at high polar angles will be detected in the CD which surrounds the target. The CD is composed of the Central Vertex Tracker (CVT), the Central Time Of Flight (CTOF) and the Central Neutron Detector (CND). Because of the fixed-target kinematics, the majority of particles are emitted at low polar angles. They are detected in the Forward Detector (FD) which is located downstream of the target. The FD is composed of three regions of Drift Chambers (DC), a time-of-flight (FTOF), a calorimeter (EC) and two Cherenkov Counters (HTCC and LTCC). Figure taken from [68].

2.4 The CLAS12 Central Detector

The Central Detector (CD) of CLAS12 is built between the target and the inner wall of the solenoid magnet. It is a barrel detector with an almost 2π azimuthal coverage. It is composed of a tracking system, the Central Vertex Tracker (CVT) (made of a silicon tracker (SVT) in the innermost region and surrounded by a micromegas tracker (MVT)), and two time-of-flight detectors, the Central Time Of Flight (CTOF) and the Central Neutron Detector (CND), dedicated to the identification of charged and neutral particles, respectively. Figure 2.3 shows a view from the CLAS12 Event Display (CED) of a reconstructed event in the CD, where all the detector layers are visible.

The CVT is used to measure the momenta of charged particles using the curvature of their helical trajectories in the longitudinal magnetic field produced by the solenoid magnet. The momentum p perpendicular to the direction of the magnetic field B is given by :

$$R = \frac{p}{qB}, \tag{2.1}$$

where R is the radius of curvature, and q the charge of the particle. The above formula only gives access to the transverse component of the momentum, the longitudinal component is deduced from the pitch of the helical track. In practice only the ratio $\frac{p}{q}$ is deduced from the radius of the track. The charge is deduced from the orientation of the track, as the tracks of particles with opposite charge curl in opposite directions in a magnetic field.

Capability	Quantity	Status
Coverage & Efficiency	Tracks (FD)	$5° < \theta < 35°$
	Tracks (CD)	$35° < \theta < 125°$
	Momentum (FD & CD)	$p > 0.2$ GeV
	Photon angle (FD)	$5° < \theta < 35°$
	Photon angle (FT)	$2.5° < \theta < 4.5°$
	Electron detection (HTCC)	$5° < \theta < 35°$, $0° < \phi < 360°$
	Efficiency	$\eta > 99\%$
	Neutron detection (FD)	$5° < \theta < 35°$
	Efficiency	$\leq 75\%$
	Neutron detection (CD)	$35° < \theta < 125°$
	Efficiency	10%
	Neutron Detection (BAND)	$155° < \theta < 175°$
	Efficiency	35%
Resolution	Momentum (FD)	$\sigma_p/p = 0.5 - 1.5\%$
	Momentum (CD)	$\sigma_p/p < 5\%$
	Pol. angles (FD)	$\sigma_\theta = 1 - 2$ mrad
	Pol. angles (CD)	$\sigma_\theta = 10 - 20$ mrad
	Azim. angles (FD)	$\sigma_\phi < 1$ mrad$/\sin\phi$
	Azim. angles (CD)	$\sigma_\phi < 1$ mrad
	Timing (FD)	$\sigma_T = 60 - 110$ ps
	Timing (CD)	$\sigma_T = 80 - 100$ ps
	Energy (σ_E/E) (FD)	$0.1/\sqrt{E\ (\text{GeV})}$
	Energy (σ_E/E) (FT)	$0.03/\sqrt{E\ (\text{GeV})}$
Operation	Luminosity	$L = 10^{35}$ cm^{-2}s^{-1}
DAQ	Data Rate	20 kHz, 800 MB/s., livetime. 95%
Magnetic Field	Solenoid	$B_0 = 5$ T
	Torus	$\int Bdl = 0.5 - 2.7$ Tm at $5° < \theta < 25°$

Table 2.1: Summary of CLAS12 performances, published in [68].

2.4.1 The solenoid magnet

A solenoid magnet [69] encloses the CD sub-detectors. It is made of four superconducting cylindrical coils. These coils produce a field primarily along the beam axis. A fifth coil is located outside of the four main coils and produces an opposite direction field and acts as an active shielding. At full operating current the solenoid magnet generates a 5 T magnetic field at its center. The solenoid magnetic field is firstly used for charge and momentum measurements of charged particles in the CD. It also provides a powerful shielding to Möller electrons. Möller electrons ($e^- + e^- \rightarrow e^- + e^-$) are produced when electrons from the beam scatter on atomic electrons in the target material. The majority of Möller electrons are curled along the longitudinal solenoid field and are collected in a tungsten shield placed downstream of the target, the Möller cone. The solenoid magnet has a cylindrical bore coaxial with the beam with a diameter of 78 cm where all sub-detectors of the CD are located.

2.4.2 Target

The CLAS12 experiments are regrouped in "run groups". The experiments of a given run group (RG) share the same experimental setup and the same target. The two first run groups of CLAS12 used the detector setup described below and unpolarized targets. The target system of CLAS12 (described in [70]) is the same as the one used in the CLAS experiment (see Figure 2.4). The cryogenic target cell consists in a 5 cm Kapton cone containing the target material. The first run group (RG-A) of CLAS12 uses liquid hydrogen (LH_2), while liquid deuterium (LD_2) was used during the second experiment (RG-B). The beam enters and leaves the target through 30-μm-thick aluminum windows. The target

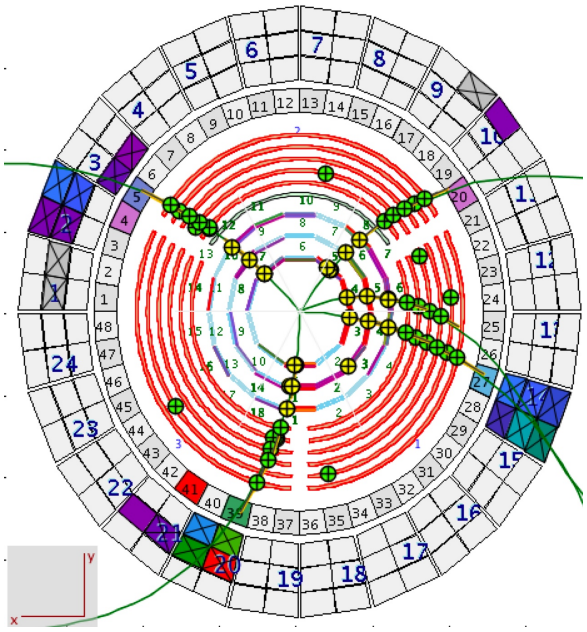


Figure 2.3: CLAS12 Event Display view of an event in the CD of CLAS12. The tracking system consists of 6 layers of silicon tracker followed by 3 micromegas layers. The Central Time Of Flight encloses the tracking system. Finally the Central Neutron Detector is the outermost detector in the central region.

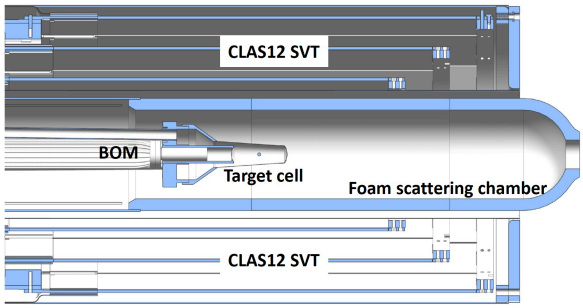


Figure 2.4: Target system of CLAS12. The target cell contains the target material (LH_2 or LD_2). The foam scattering chamber aims at reducing the material budget of particles emitted in the FD.

is located inside a 45-cm-long scattering chamber made of low density Rohacell XT110 foam that aims at reducing the material thickness crossed by particles emitted from the target. The beam position is monitored at the target location using a Beam Offset Monitor (BOM). This is achieved by placing a glass cylinder parallel to the beam. The Cherenkov light produced by the beam halo is read out with optical fibers to a multi-anode PMT and provides information on the offset of the beam at the target position.

2.4.3 The Central Vertex Tracker

The Central Vertex Tracker (CVT) of CLAS12 is the tracking system in the CD. It is composed of two subsystems using two distinct technologies: a three-layer Silicon Vertex Tracker (SVT) close to the target followed by a six-layer Micromegas Vertex Tracker (MVT). A schematic view of the complete CD tracking system is shown in Figure 2.5.

Silicon tracker

The SVT [71] is designed to provide a good momentum resolution in the CD. It is composed of three layers of double-sided silicon micro-strip modules. Each module is composed of two layers of 256 silicon micro-strip foils separated by a resistive wafer. Silicon strips have a small constant angle pitch (1/85°) with their nearest neighbor, allowing a 3D localization of the interaction point. Each layer is composed of respectively 10, 14 and 18 modules placed parallel to the beam line. This design is used to maximize the momentum resolution while minimizing the material budget in the inner region of the CD. A full description of the SVT is provided in [71].

Micromegas

The six outer layers of the CD tracking use the micromegas technology, and are referred as Micromegas Vertex Tracker (MVT). Micromegas are gaseous tracking detectors. A micromegas layer is composed of a drift region where free electrons are produced by the crossing of high-energy particles, which then drift toward a micro-mesh foil, delimiting the amplification gap. In the amplification gap, a large electric field accelerates the electrons, creating important electromagnetic showers. The signal is collected on parallel readout strips, located on the far side of the amplification gap. For a full description of the detector see [72]. Three layers of the MVT have their readout strips oriented longitudinally and the three others perpendicularly to the beam direction. This design increases the polar and azimuthal angular resolution of the tracking in the CD.

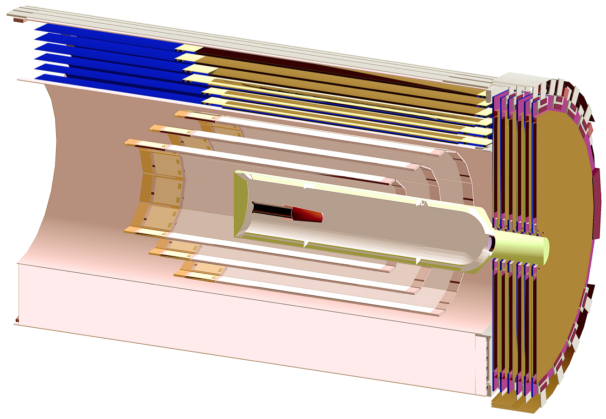


Figure 2.5: View of the CVT and the target system of CLAS12. The SVT surrounds the target, the MVT provides outer tracking.

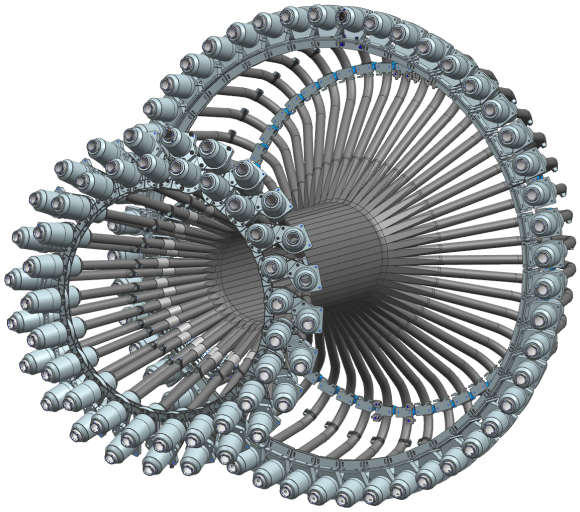


Figure 2.6: 3D view of the CTOF. It consists of 48 plastic scintillators arranged in a cylinder surrounding the CLAS12 central tracking system. The light produced by the interacting particles is guided outside the solenoid magnetic field via bent light guides.

2.4.4 Central Time Of Flight

The Central Time Of Flight of CLAS12 (CTOF) [73] consists of a barrel of plastic scintillators surrounding the central tracking system (see Figure 2.6). The scintillating light is collected on both sides of the scintillating paddles. It is then guided out of the magnetic field by light guides toward photomultipliers (PMTs). The downstream light guides have a bent design allowing readout on the upstream side of the CD. This design was forced by the presence of the High Threshold Cherenkov Counter (HTCC), close to the solenoid magnet volume. The CTOF provides a precise measurement of the time of flight of particles from the interaction point to their detection point. The particle identification is performed by comparing the momentum measured by the CVT to the expected one calculated from the measured time-of-flight, assuming various mass and charge hypotheses. The resolution of each paddle is estimated to be around 80 ps (see [73]), allowing for a good charged particle identification in the CD.

2.4.5 Central Neutron Detector

The Central Neutron Detector (CND) [74] is the outermost sub-detector of the CLAS12 CD. It consists of three radial layers of 48 plastic scintillator counters. It aims at increasing the neutron

detection efficiency in the central region. The CND is a single sided readout detector. Two adjacent paddles are optically coupled by a light guide at their upstream ends. This design was enforced by the bent-light-guide design of the CTOF preventing any readout from the downstream end of the CND. A full description of this detector is provided in Chapter 3 of this manuscript.

2.5 The CLAS12 Forward Detector

The Forward Detector (FD) of CLAS12 is located downstream of the target. It aims at detecting particles with polar angles comprised between 5° and 35°. It is divided into six sectors, each sector being equipped with the same set of sub-systems.

2.5.1 The torus magnet

The torus magnet of CLAS12 [69] produces a magnetic field perpendicular to the beam direction. Particles are deflected, depending on their charge, toward or outward of the beam direction. The curvature of the tracks allows for momentum reconstruction. The torus magnet of CLAS12 consists of six trapezoidal supra-conductor coils located between each drift chamber (see 2.7). The six coils are connected in series thus providing the same magnetic field in all the six sectors.

2.5.2 Drift Chambers

The forward tracking is achieved by three consecutive regions of Drift Chambers (DC) [75], each region being divided in a set of six sectors. The design of the DC is shown in Figure 2.7. The mechanical structure of the DC is provided by the torus magnet support. Each sector is separated by one of the six torus coils. Each region is composed of 12 layers of 112 sense wires. Six consecutive layers are grouped into a superlayer. Two superlayers of a given region have a ±6° stereo angle with respect to each other in order to increase the momentum resolution. The structure of a superlayer is shown in Figure 2.9. The charged particles crossing the layer ionize the gas mixture and the electrons produced then drift along the electric field created by field wires toward the sense wires. The recorded drift time permits to reconstruct the track of the particle through the chamber.

2.5.3 Forward Time Of Flight

The Forward Time Of Flight (FTOF) [76] is situated after the DCs and is used to measure the time of flight of charged and neutral particles emitted in the FD. It is divided in six sectors, each sector being composed of three panels of double-sided readout plastic scintillator paddles (panel-1b and panel-1a situated consecutively in the low polar angle regions, and panel-2 covering high polar angles). Figure 2.8 shows a schematic view of the FTOF and its mechanical support. The time resolution achieved by the FTOF ranges from 50 ps for the short paddles located close to the beam direction to 200 ps for the longer ones.

2.5.4 Electromagnetic Calorimeter

The CLAS12 Electromagnetic Calorimeter (EC) [77] consists of six independent calorimeters, one for each forward sector. Each sector is divided in two modules. The downstream module consists of two layers, the inner calorimeter (ECIN) and the outer calorimeter (ECOUT) which have been refurbished from the CLAS experiment. The Pre-shower Calorimeter (PCAL) is located in front of the former. The modules consist of lead sheets sandwiched with scintillator bars with single sided readout. Both modules use a triangular hodoscope geometry with stereo readout as described in Figure 2.10. The EC is used for particle identification through the measurement of the deposited energy and for kinematic reconstruction of neutral particles using time and position of the measured showers. The EC (ECIN and ECOUT) is composed of 39 layers of 10-mm-thick scintillator bars followed by 38 2.2-mm-thick lead sheets.

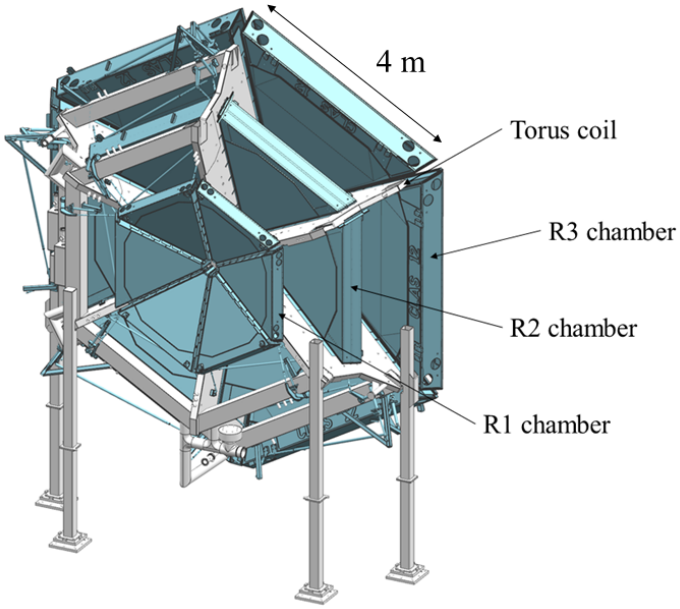


Figure 2.7: CLAS12 forward tracking system. It consists of three consecutive drift chamber regions, each of them covering one of the six forward sectors of CLAS12.

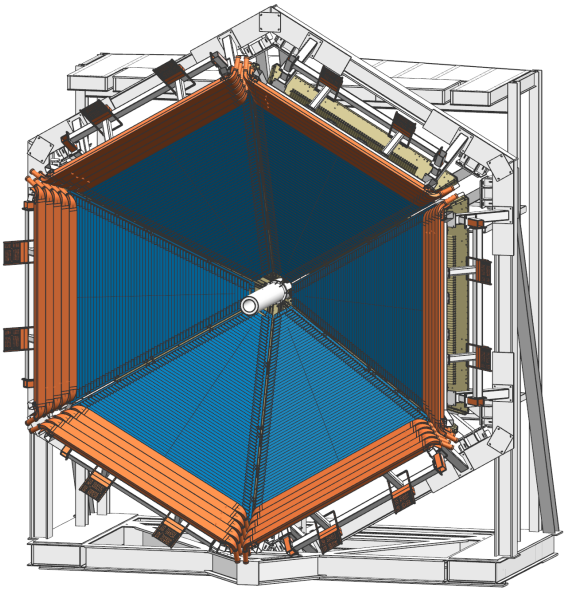


Figure 2.8: Forward Time Of Flight of CLAS12. It consists of three different panels with paddle timing resolutions ranging from 50 to 200 ps.

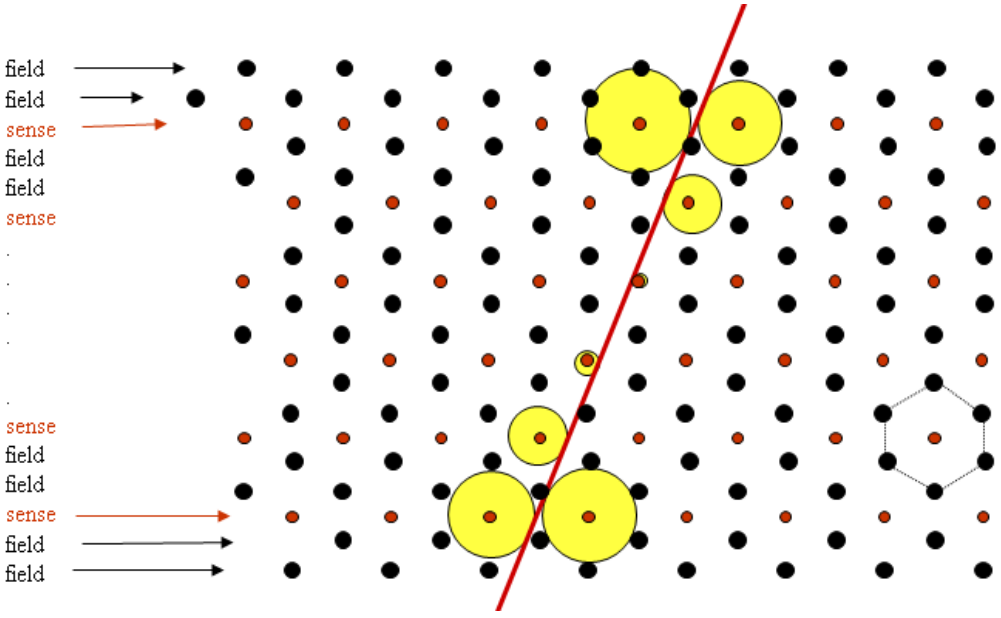


Figure 2.9: Wire layout for one superlayer of the CLAS12 drift chambers. One layer of sense wires is surrounded by two layers of field wires on each side, forming a "honeycomb" pattern. The red line represents the path of a charged particle being detected in the chamber. The drift distance (deduced from the drift time) from the track to the closest signal wire (represented by yellow circles) allows to retrieve the track characteristics.

Pre-shower Calorimeter

The Pre-shower Calorimeter (PCAL) has been designed and built for the CLAS12 upgrade. It is located just downstream of the FTOF. It aims at increasing the radiation length of the EC, hence

providing a better measurement of the total shower energy. It is also used in the identification of electrons as they tend to deposit most of their energy in a small shower volume while pions (that are minimum ionizing particles at CLAS12 kinematics) deposit energy in all EC layers. It is composed of 15 layers of 1-cm-thick scintillators bars sandwiched with 14 layers of 2.2-mm-thick lead sheets. The layers are composed of 84 strips in the U planes and 77 strips in the V and W planes. The PCAL has narrower strips (4.5 cm) than the EC (10cm). This feature provides a better determination of the position of the particle compared to the EC alone and allows to separate close clusters produced by photons in high energy π^0 decays. At large angles and for all layers, each pair of strips is combined into a single readout channel to optimize the number of readout channels.

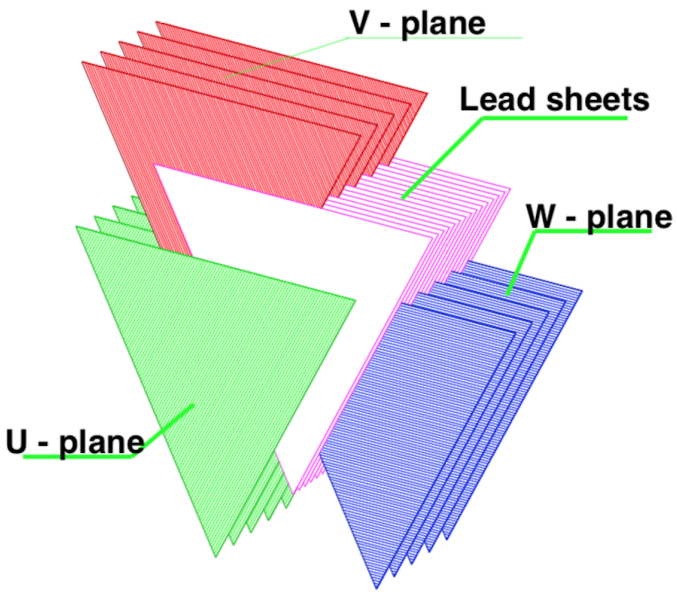


Figure 2.10: Layout of a CLAS12 EC module. Lead sheets are sandwiched with scintillator bars with alternating orientation (denoted U, V and W planes).

2.5.5 Cherenkov Counters

There are two main Cherenkov counters systems in CLAS12: the High Threshold Cherenkov Counter (HTCC) and the Low Threshold Cherenkov Counter (LTCC). These two sub-detectors are used for particle identification. Cherenkov counters are usually filled with a large volume of gas (CO_2 for HTCC and C_4F_6 for LTCC). A particle crossing the gas volume at a speed higher than the speed of light in the medium emits a cone of light. This phenomenon, known as the Cherenkov effect, allows to distinguish particles as the emission of light is possible only if the momentum of the particle is higher than a threshold momentum p_{th} given by:

$$p_{th} = \frac{mc}{\sqrt{n^2 - 1}}, \tag{2.2}$$

where n is the refraction index of the gas and m the mass of the particle. The Cherenkov light is then collected by mirrors located around the gas volume and sent to PMTs where it is amplified and read out.

High Threshold Cherenkov Counter

The High Threshold Cherenkov Counter (HTCC) is the most upstream detector of the FD. It is located just outside of the solenoid outer wall, covering all azimuthal angles (see Figure 2.11). All particles emitted at low polar angles (between 5° and 35°) in the FD will cross the HTCC. The light emitted is collected by 12 ellipsoidal mirrors located on the downstream wall of the gas volume. PMTs

are located in the outermost part of the detector. This design ensures that the quantity of matter crossed by particles is as low as possible. The HTCC is designed to accomplish a 99% electron/pion separation for momenta below 4.9 GeV.

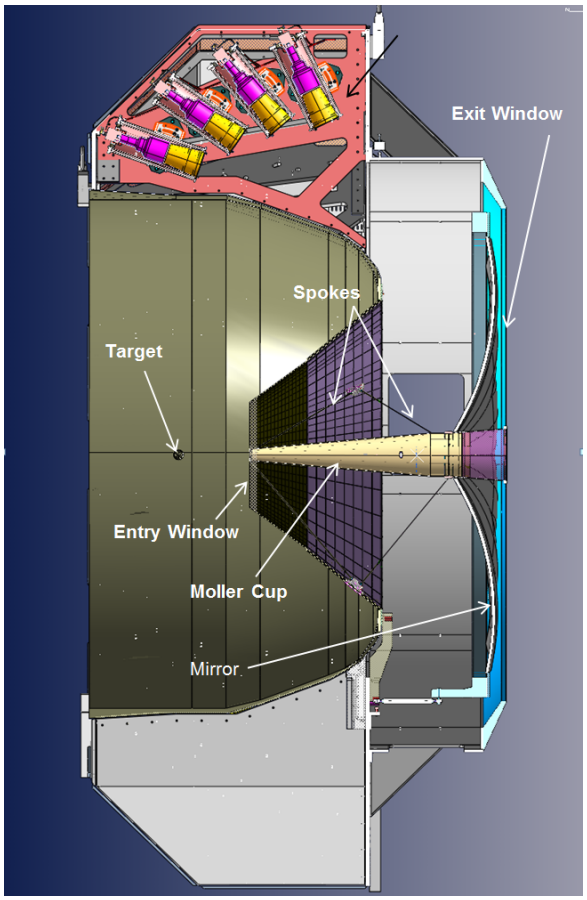


Figure 2.11: Schematic lateral view of the High Threshold Cherenkov Counter of CLAS12. Particles with a polar angle between 5° and 35° pass through the chamber of the HTCC, allowing for the separation between electrons and pions.

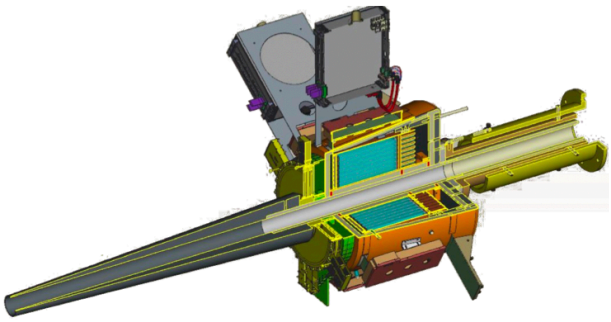


Figure 2.12: The Forward Tagger of CLAS12 is situated downstream of the target at very low polar angles ($2° < \theta < 5°$). It is used to detect electrons and photons scattered close to the beam pipe.

Low Threshold Cherenkov Counter

The Low Threshold Cherenkov Counter (LTCC) [78] is designed to separate pions and kaons in the 4 to 8 GeV region. It covers two sectors and it is made of the refurbished CLAS Cherenkov counters.

RICH detector

In addition to the two main Cherenkov systems, a Ring Imaging Cherenkov detector (RICH) [79] is installed in one of the sectors of the FD. This detector has the capability of measuring the angle θ at which the Cherenkov light is emitted, which is given by:

$$\cos(\theta_C) = \frac{1}{n\beta}. \tag{2.3}$$

This allows for kaon identification in the 3-8 GeV momentum range, by matching their β measured from the time of flight with the Cherenkov cone angle θ_C.

2.5.6 Forward Tagger

The Forward Tagger (FT) [80] is a composite detector made of a calorimeter, a scintillator hodoscope and two double-layers of micromegas trackers. It is located at very low polar angles (2° to 5°) and it aims at identifying electrons and photons emitted close to the beamline (see Figure 2.12). This detector is mainly used for tagged-photo reactions and DVCS measurements.

2.6 Other devices

2.6.1 Beamline

Before interacting with the target, the electron beam provided by the CEBAF is stirred and monitored (see Figure 2.13 extracted from [70]). The beam position monitoring is achieved upstream of the target by a series of Beam Position Monitoring (BPM) systems. The transverse profile of the beam is recorded by wire harp systems composed of tungsten wires that can be moved into the beam (this is an invasive measurement only performed before data taking). Finally halo counters (PMTs located at various angles and positions close to the beamline) monitor undesired beam halos and backgrounds.

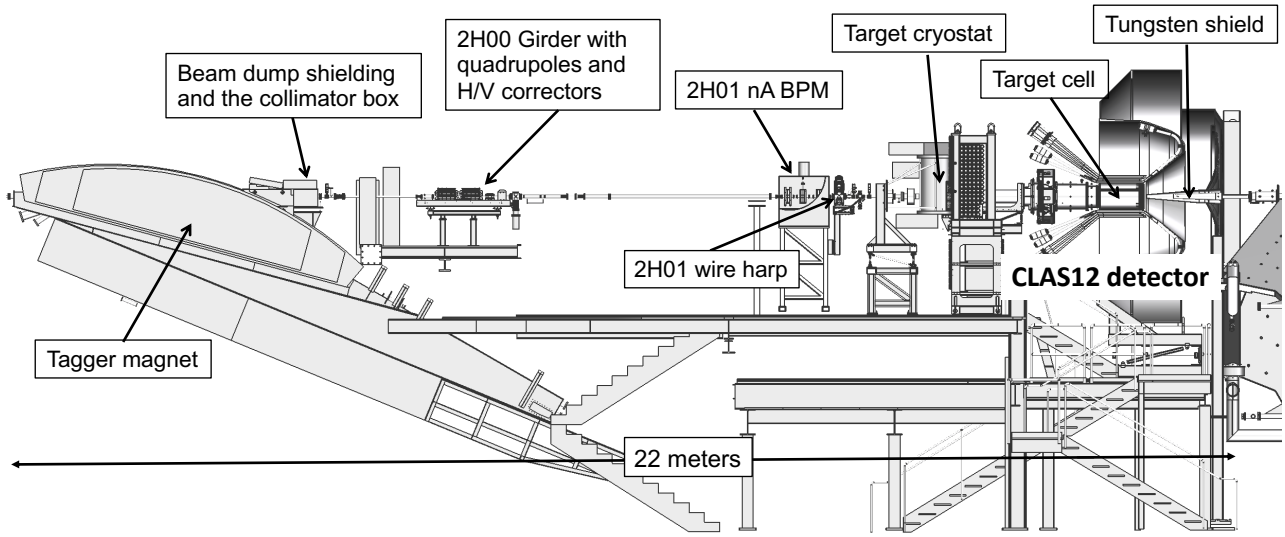


Figure 2.13: Schematic description of the CLAS12 beamline.

2.6.2 Möller polarimeter

The electrons accelerated by the CEBAF are polarized. The polarization efficiency is around 85%. In order to correct for this incomplete polarization, it is measured regularly during data taking. This measurement is performed using the Möller polarimeter located just before the beamline enters the experimental hall. Polarization measurements make use of the Möller scattering of longitudinally polarized electrons from the beam off atomic electrons from the longitudinally polarized Möller target ($\vec{e}+\vec{e} \rightarrow e+e$). The measurement is performed by putting a polarized target (made of Permendur, an iron-cobalt alloy) in the beam. Scattered electrons are detected in two scintillator detectors located at symmetric angles with respect to the beam. The scattered electrons are extracted from the beam by

quadrupole magnets. The difference of the coincidence rate for each beam helicity can be related to the polarization of the beam. The layout of the Möller polarimeter is shown in Figure 2.14.

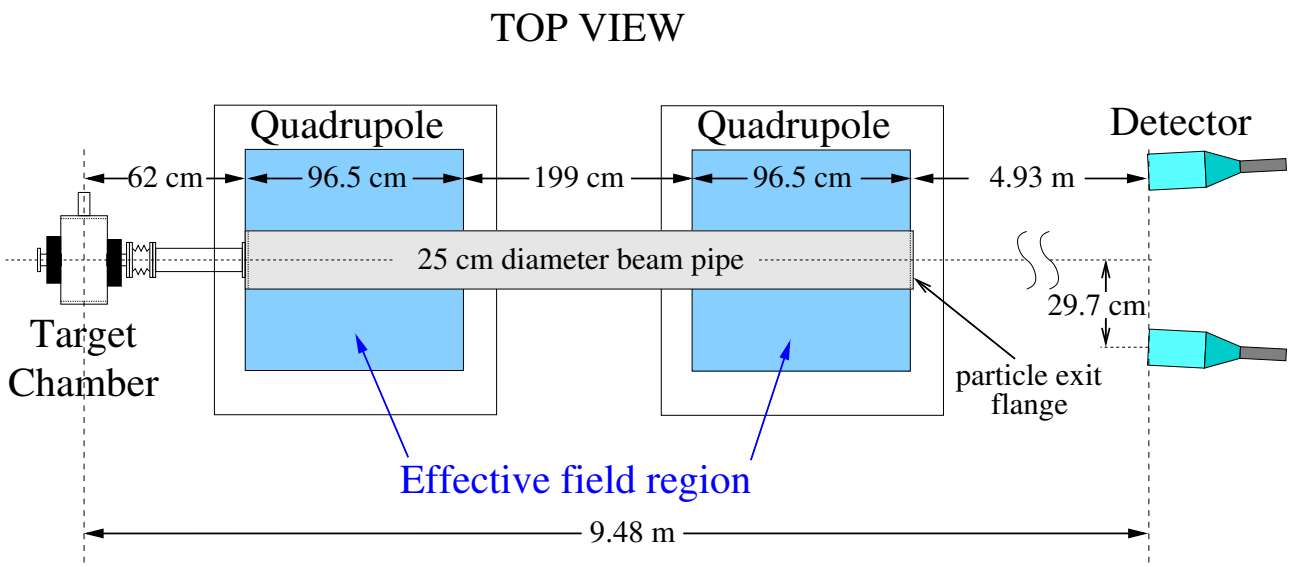


Figure 2.14: Möller polarimeter layout. It is composed of a polarized target, two quadrupole magnets and a pair of detectors.

2.7 Data acquisition and processing

2.7.1 Trigger system and Data Acquisition system

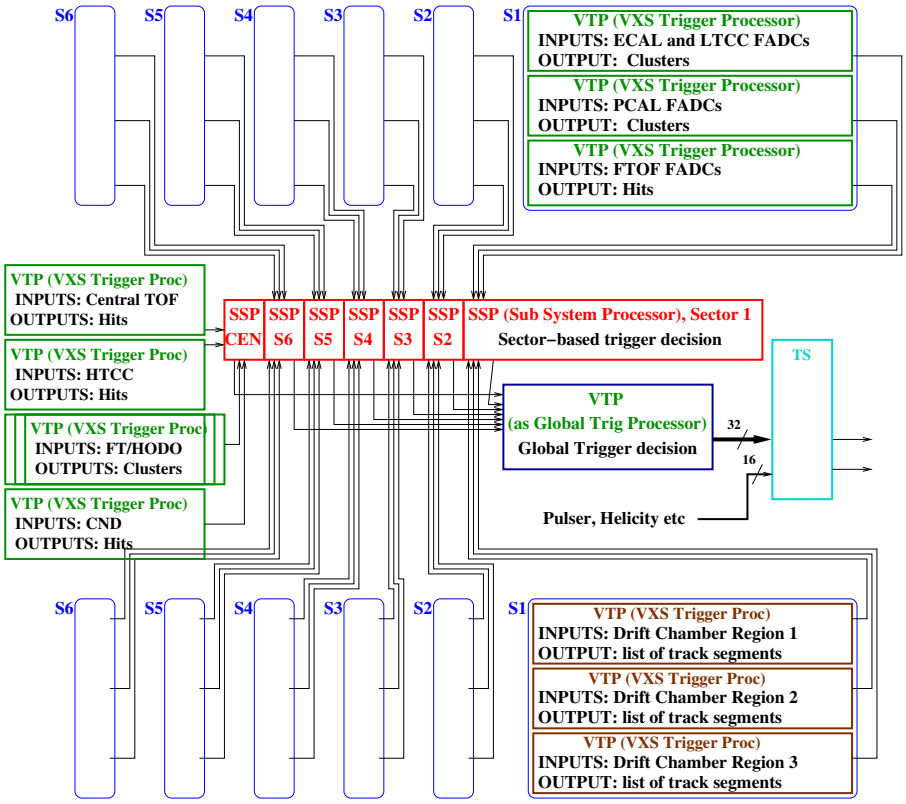


Figure 2.15: Diagram of the trigger system of CLAS12.

The trigger system [81] is the first step of data acquisition. It aims at providing the Data Acquisition system (DAQ) a trigger signal for data recording. The goal of the trigger is to keep only events where an electron effectively interacted with the target and reject non-interesting events such as cosmic rays for example. The trigger is composed of three stages that determine online if an event is worth recording for later offline analysis. The trigger system is represented on Figure 2.15. The first stage of the trigger is composed of VXS Trigger Processor boards (VTP) that collect the responses of the detectors (ADCs and TDCs, see Chapter 3) which are included in the trigger procedure and identify possible signals coming from particles. Each VTP provides a series of bits corresponding to the state of its associated detector. The bit series is then fed to a Sub System Processor (SSP). There are seven SSPs, one for each forward sector and one for the CD. SSPs combine information from each detector at the sector level (and CD) and send their response to a final VTP module that produces the final trigger bits. The total latency (time to process an event) of the trigger system is at most 8 μs. During RG-A data taking, the final trigger rate was lower than 20 kHz and the live time (the fraction of the time during which data were recorded on tapes) was around 95%. Triggered events are then read out by the DAQ system of CLAS12. The data-to-tape rate is reported to be up to 500 MB/s.

2.7.2 Event reconstruction and processing

Data are stored on large data tapes in the EVIO (Event Input/Output) format. The first step of the offline processing is referred as decoding. The decoding process takes EVIO files as input and produces output in the HIPO (High Performance Output) format. It involves the fitting of the signal waveforms and the translation from the DAQ electronic notation (crate/slot/channel) to the detector notation (sector/layer/component). This step is performed once, as it requires a large computing power. The output HIPO files contain detector related banks for each event. Calibration and reconstruction are then performed using these decoded files. The reconstruction of CLAS12 data is based on the CLARA framework [82]. CLARA is a multi-threaded framework that allows to run services according to a user-defined architecture (see the CLAS12 reconstruction architecture shown in Figure 2.16 for example). Each service receives input (I), processes it and produces output (O). I/O are organized in banks, each bank being associated to one event and containing custom information. This architecture was chosen for its versatility and the possibility to add custom services at a later stage of the experiment. As shown in Figure 2.16, each sub-detector of CLAS12 has its own reconstruction service. The output of each subsystem service is fed to the Event Builder (EB) of CLAS12. The EB associates all the sub-system responses and produces the particle information (particle identification, momenta and vertex).

2.7.3 Simulation

A full simulation of CLAS12 [83] is implemented in the Geant4 Monte Carlo (GEMC) package. Each sub-detector volume is implemented as a Geant4 volume, the response of the subsystems are given by specific "hit process" routines. Data taken with random trigger can be added to the simulated data to reproduce the background (see Section 5.5 for details).

2.8 Data set

The dataset used in the work of this manuscript was taken during the Fall 2018 run period, between October and December 2018. The run period is part of the CLAS12 run group A (RGA), which is dedicated to the study of the proton. The target used during RGA is LH2. The torus magnet was set to bend negative particle toward the beam pipe (i.e. inbending configuration). The accumulated charge taken during each shift and the total charge taken during this run period is shown in Figure 2.17. The total accumulated charge during this run period corresponds to approximately $200 fb^{-1}$.

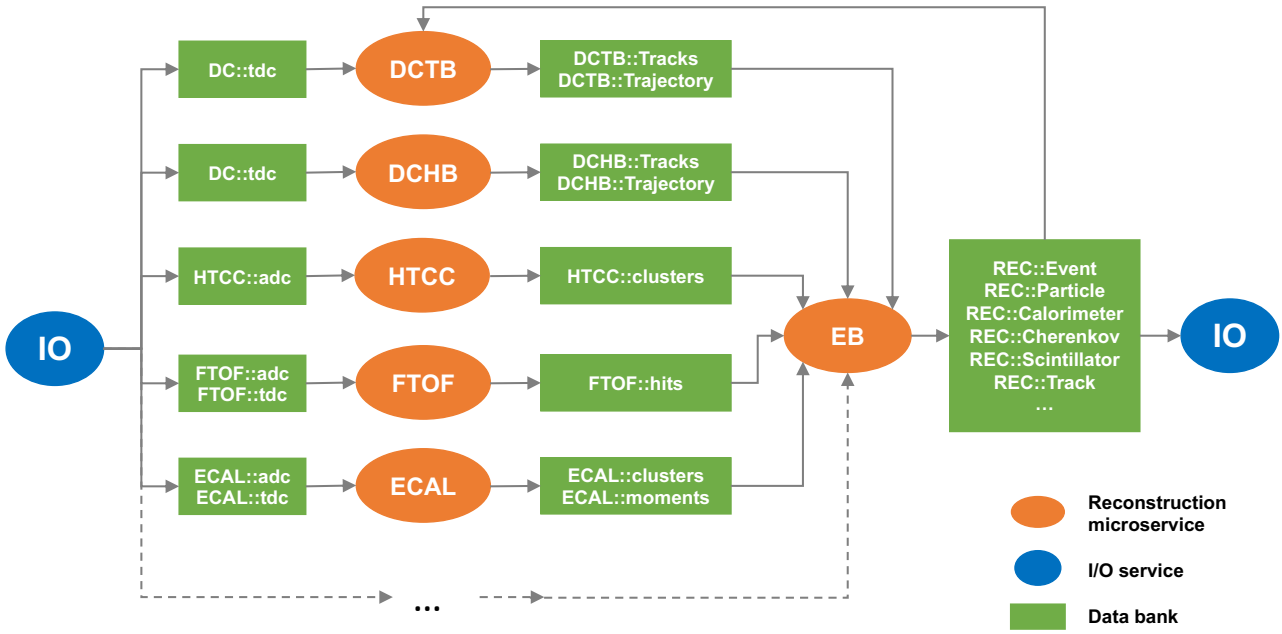


Figure 2.16: Reconstruction framework of CLAS12, implemented in a CLARA architecture.

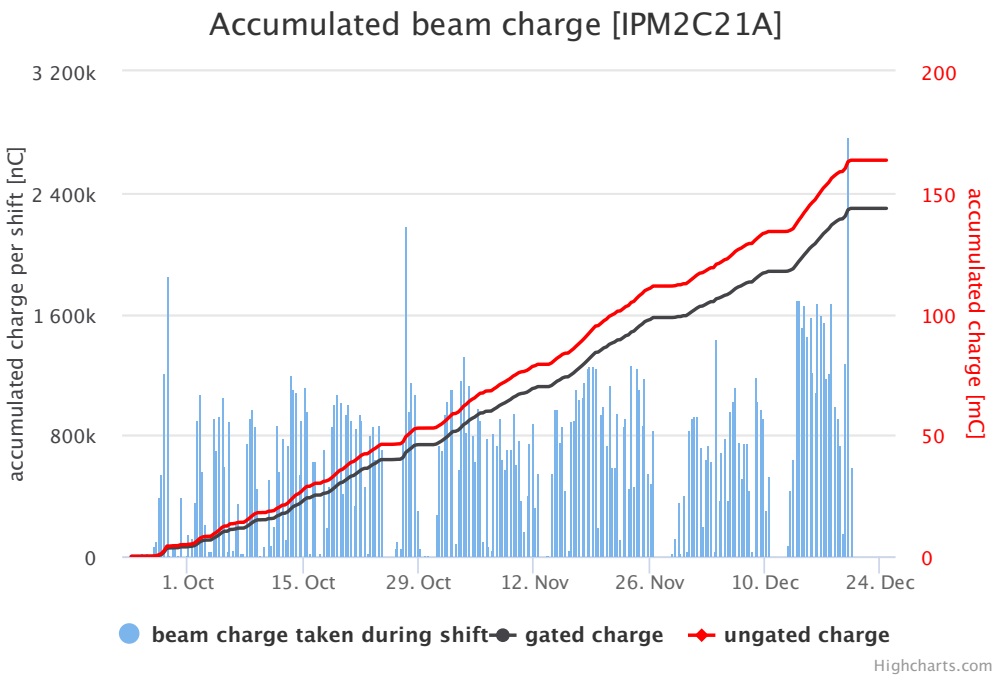


Figure 2.17: Accumulated charge per shifts and total accumulated charge. The gated charge corresponds to the ungated charge weighted by the live time of the CLAS12 DAQ. It corresponds to the actual amount of data recorded during this run period. Plot extracted from the CLAS12 monitoring system.

Chapter 3

The Central Neutron Detector

The Central Neutron Detector (CND) [74, 84] is the outermost of the subsystems composing the Central Detector of CLAS12. This detector was designed and built at the Institut de Physique Nucléaire d'Orsay (IPN, now part of IJCLab). It consists of a barrel of three layers of scintillators coupled at their downstream ends with U-turn light guides. The readout is performed at the upstream ends of the scintillators by photomultiplier tubes (PMTs) connected to the bars via 1-m-long bent light guides. The PMTs are positionned in the fringe-field region of the CLAS12 5-T superconducting solenoid. The CND was installed in the CLAS12 solenoid, and subsequently started its data taking, in the Fall of 2017. It is dedicated to the detection of neutrons with polar angles between 40° and 135°. The main purpose of the CND is the measurement of DVCS observables on the neutron (nDVCS) [85]. Measuring DVCS observables on neutron ($en \to en'\gamma$) is one of the necessary steps toward a full understanding of the structure of the nucleon in terms of GPDs. Quark-flavor separation of GPDs is possible by combining proton and neutron GPDs. Furthermore nDVCS plays a complementary role to the transversely polarized proton target DVCS experiments for its sensitivity to the GPD E. The GPD E plays an important role in understanding the spin structure of the nucleons as it enters directly the Ji's sum rule that describes the angular momentum contribution of the quarks to the spin of the nucleon. The CND will also be used in other nDVCS experiments [86] and whenever the detection of the recoil neutron may be required (N^* program, for instance, or for all the deeply-virtual meson production reactions on the neutron). The first beam data collected by CLAS12 on a proton target confirmed the design performance.

This chapter is divided in eleven sections. Section 3.1 presents the motivations for building the CND. In Section 3.2 the nDVCS experiment requirements and their consequences on the design of the CND are presented. Then in Sections 3.3 to 3.9, each aspect of the CND software is discussed. Section 3.10 presents the detection performance of the CND. Preliminary results for the nDVCS experiments are shown in Section 3.11.

3.1 Motivations

In Section 1.3, the Deeply Virtual Compton Scattering (DVCS) process is presented as one of the main physical processes described by GPDs. As of today, most DVCS data have been taken on proton target (pDVCS). Yet neutron DVCS (nDVCS) data are also of great interest [87, 88, 89].

One of the main motivations for the neutron DVCS is the flavor separation of GPDs. As mentioned in the PDFs case in Section 1.1, it is possible to decompose the valence-quark flavor dependence of proton and neutron GPDs using isospin symmetry as:

$$H^p = \frac{4}{9}H^u + \frac{1}{9}H^d, \tag{3.1}$$

$$H^n = \frac{4}{9}H^d + \frac{1}{9}H^u, \tag{3.2}$$

where similar equations apply to the other GPDs. By measuring both proton and neutron observables, an experimental extraction of quark GPDs can be performed.

Another point advocating for neutron DVCS is its complementary sensitivity to CFFs with respect to proton DVCS. As shown in Section 1.3, DVCS asymmetries are sensitive to CFFs weighted by FFs, but the proton and neutron FFs have different behaviors (see Figure 1.4). For example the beam-spin asymmetry (BSA) is moslty sensitive to the imaginary part of the CFF $\mathcal{H}$ in proton DVCS, while it is sensitive to the imaginary part of $\mathcal{E}$ in neutron DVCS (the $\widetilde{\mathcal{H}}$ contribution in neutron DVCS is expected to be small due to the cancellation between u and d quarks). Therefore measuring the BSA for nDVCS allows to constrain the E GPD, which is poorly constrained by current proton data (transversely polarized protons experiments are also sensitive to GPD E, but are also very challenging experimentally). The neutron DVCS will thus play a central role in the measurement of the Ji's sum rule.

However nDVCS is a much more challenging process to measure than pDVCS. As neutron targets are not stable, a deuterium target is needed. The use of a deuterium target implies that final-state-interaction corrections have to be applied to the extracted quantities, as neutrons in deuterium are bound. Furthermore neutrons are more difficult to detect than protons. They are electrically neutral and consequently cannot be detected in tracking detectors. In addition, as neutrons mostly interact with matter through elastic scattering, their detection in calorimeters and scintillator detectors has a lower efficiency than for protons. This issue can be dealt with by increasing the active volume of detection, in order to obtain a larger efficiency. This was the solution chosen for the CLAS12 nDVCS experiment. The dedicated neutron detector is presented in the next sections.

3.2 Design and hardware

3.2.1 Requirements

At CLAS12 kinematics, in the nDVCS channel, electrons and photons are mostly detected in the Forward Detector (FD) and in the Forward Tagger (FT), while neutrons are mainly emitted at large polar angles in the Central Detector (CD) ($\theta > 40°$). Thus high neutron detection efficiency, good momentum and angle resolution as well as high neutron-photon separation power are required in the CD in order to measure nDVCS observables. As described in Chapter 2, the CND is situated between the CTOF and the solenoid magnet as shown in Figure 3.2. The CND is providing most of the neutron detection efficiency in the CD.

Without the CND the neutron detection capability of the CD is given by the Central Time-Of-Flight (CTOF) of CLAS12. The CTOF is a barrel detector which consists of a single layer of 3-cm-thick plastic scintillators. It provides a relatively low neutron detection efficiency of around 3%. The CND has a radial thickness of 9 cm corresponding to the full available space between CTOF and the magnet. It is designed to have a neutron detection efficiency of around 10%.

Experimentally nDVCS is performed on a deuterium target and the spectator proton stays undetected ($ed \to e'n\gamma(p)$) because of its low recoil momentum. To ensure that the spectator particle is a proton, the missing mass of the final state spectator $e'n\gamma(X)$ is calculated and cut around the mass of the proton. In the computation of this missing mass the kinematic variables of the electron and the photon are given by the FD or FT while the neutron ones are given by the CND. It was shown in the early stage of the R&D process of the CND that the momentum resolution of the CND should not exceed 10% in order for the neutron kinematics to have low effect on the proton missing mass. This is shown in Figure 3.1 extracted from [84].

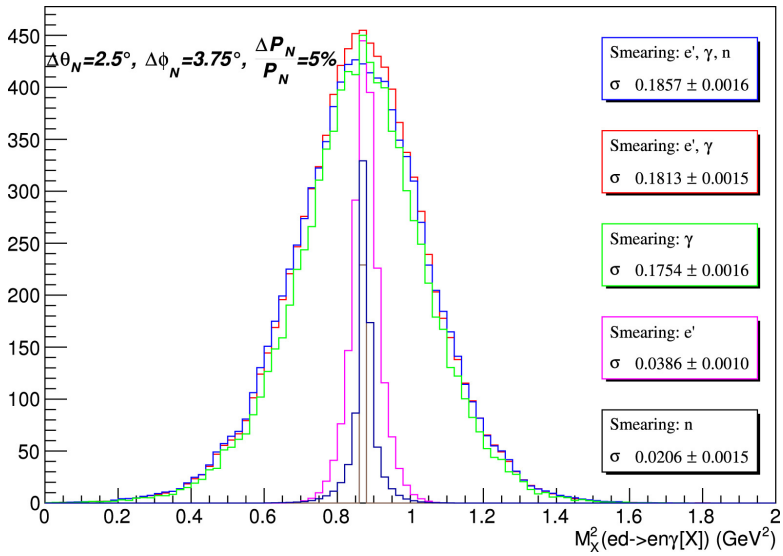


Figure 3.1: Missing mass squared of the $e'n\gamma(X)$ system, for the nDVCS channel. The different colors correspond to different combinations of particles detected with or without resolution effects. One can see that the width of the missing mass is driven by the resolution on the kinematics of the photon. The neutron contribution to the total width of the missing mass is small for a resolution on its momentum of 5%. Figure taken from [84].

The CND has to provide a good neutron-photon separation in the CD. The separation between neutrons and photons is achieved by measuring β for a given neutron candidate:

$$\beta = \frac{d_{\text{path}}}{TOF \times c} \tag{3.3}$$

where TOF is the time-of-flight of the particle, d_{path} is the path traveled by the particle from the target to the interaction point with the CND, and c is the speed of light. Both TOF and d_{path} are calculated using the CND (see Section 3.4 for more details). d_{path} is obtained as:

$$d_{\text{path}} = \sqrt{z^2 + r^2} \tag{3.4}$$

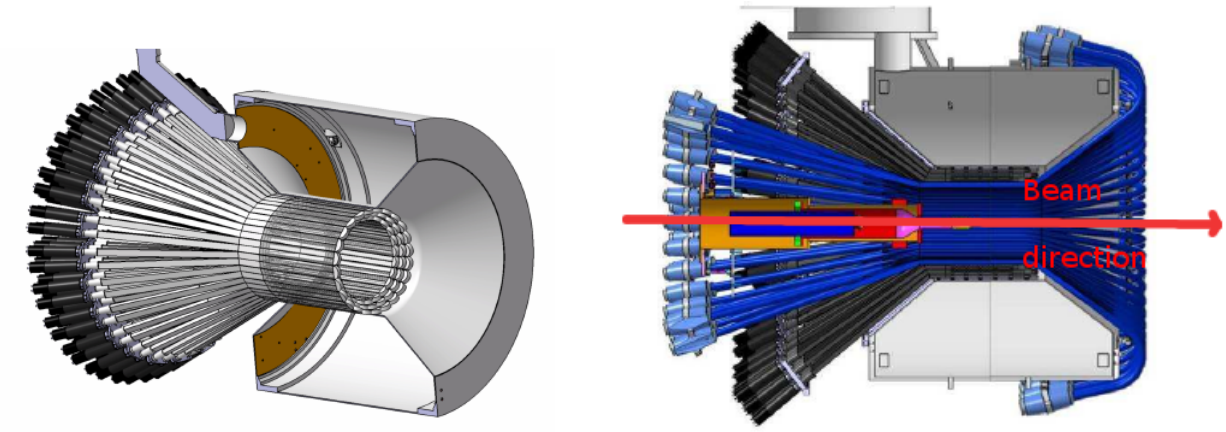
where z is the longitudinal position of the hit and r its radial position. A radial segmentation is needed in order to achieve a good measurement of r. The longitudinal position z is given by:

$$z \propto \Delta t \times v_{eff} \tag{3.5}$$

where Δt denotes the difference of the time measured at both ends of the hit scintillator and v_{eff} is the effective velocity of the light in the paddle. The TOF is proportional to the sum of the measured times at both ends of the paddle. Thus the β neutron-photon separation is only dependent on the time resolution of the CND. Geant4-based simulations show that to ensure a 3-σ separation up to 0.9 GeV the time resolution of the CND should be about 150 ps.

3.2.2 Design description

The CND uses the space left between the CTOF and the solenoid magnet. The downstream light-guides of the CTOF do not allow for downstream readout of the CND as shown in Figure 3.2b. Thus the CND was designed with light readout at the upstream end only. The CND consists of three radial layers of 48 plastic scintillator paddles with trapezoidal shape that are collinear to the direction of the beam. The scintillating material is EJ-200 (Polyvinyl Toluene, PVT) from Eljen Technology. The radial thickness of all the paddles is 30 mm while the other dimensions vary depending on the layer, as shown in Table 3.1.



(a) (b)

Figure 3.2: (a) Drawing of the CND inserted in the solenoid magnet. (b) Drawing of the CND (in black) inserted between the CLAS12 Central Time-Of-Flight (in blue) and the solenoid (in grey). The target system is also shown inside the CTOF.

Layer	Inner face width (mm)	Outer face width (mm)	Length (mm)
1	35.92	39.87	665.72
2	40.0	43.95	700.0
3	44.08	48.03	734.28

Table 3.1: Dimensions (mm) of the trapezoidal scintillator bars of the CND. The layer numbers go from the innermost (1) to the outermost (3). The thickness of all the bars is 30 mm.

Adjacent paddles are coupled two-by-two at their downstream ends with a u-turn lightguide. This allows for single-sided readout as one of the two paddles plays the role of lightguide for its neighbor, as shown in Figure 3.3. The signal extracted from the paddle where the hit occurred is called *direct* while the *indirect* signal is retrieved from the coupled paddle. Each bar is wrapped in an aluminum foil acting as reflector. Three pairs of coupled paddles are stacked up, one over the other, forming a sector or block. Overall there are 24 sectors of 6 paddles, for a total of 144 scintillators, numbered according to the scheme defined in Figure 3.4. The scintillation light produced by the interaction of particles is guided from the upstream ends of the scintillators to the PMTs through 1.5 m-long lightguides. The PMTs are located outside of the solenoid magnetic field surrounding the central detector of CLAS12 and are individually shielded by a 1-mm-thick mu-metal cylinder and a 5-mm-thick mild steel outer cylinder. The amplification of the light signal and the readout is done by Hamamatsu R10533 PMTs located at the end of the lightguides. This designed was tested before the installation. The timing resolution of each block was measured using cosmic rays as described in [74]. The results of these tests are shown in Figure 3.5: the timing resolution is compatible with the R&D specifications. The installation of the CND took place in Fall 2017. The support structure, designed alongside the detector, was fixed to the solenoid magnet. All the sectors were then installed one by one. The PMTs were coupled to the lightguides using optical grease. A picture of the detector after the installation is shown in Figure 3.4.

3.2.3 High voltage power supply, electronics and readout

The PMTs are energized by a high voltage (HV) of around 1500 V. The HV is supplied by CAEN SY527 power supplies on which CAEN A734N multi-channels boards (16 channels, 3kV max voltage and 3 mA max current) are set.

The signals from the PMTs are read by the electronic readout, situated a few meters away from the CND (see Figure 3.6). First the signal is sent to active splitters. The three splitter modules used for

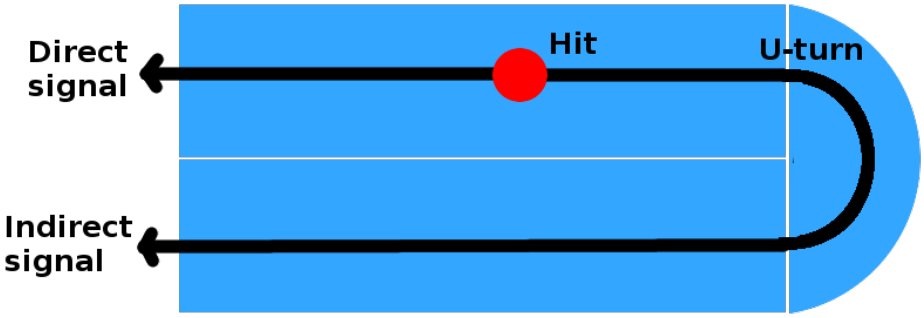


Figure 3.3: Concept of particle detection in the CND.

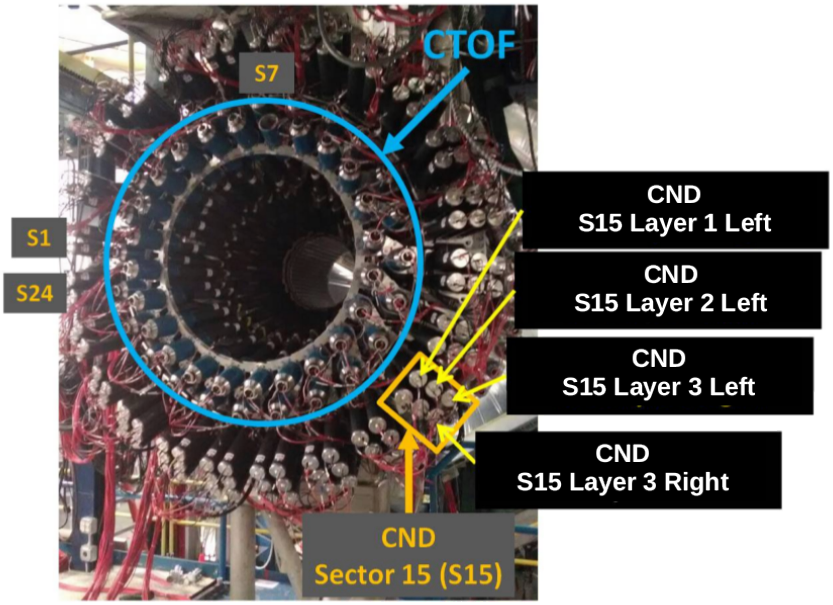


Figure 3.4: The channel numbering scheme of the CND. The CND is composed of 144 paddles, divided into 24 sectors of 3 layers of coupled paddles.

the CND were developed at IPN Orsay for the G0 experiment in Hall C at Jefferson Lab [90]. There are 64 input SMA channels on each module. For each input channel, two output signals are provided at the front end of the splitter modules: a signal for the charge measurement and a signal for the time measurement referred as time and charge signal respectively in the following. The gain is set to unity to ensure no loss of amplitude.

The charge signals are sent via LEMO cables to Flash Analogue to Digital Converter (fADC) modules (see Figure 3.6) that provide the total integrated charge of the signal (referred as ADC in the following). The time signals are sent to Constant Fraction Discriminators (CFD) GAN'ELEC FCC8 that have been designed for the TAPS collaboration [91]. The use of CFDs ensures that no time-walk correction is needed at the calibration stage. The time signal is then sent to Time to Digital Converters (TDC) which have a resolution of 25 ps per channel. The output digitized raw time is referred to as TDC in the following. From these two sets of numbers (ADCs and TDCs) the time and the deposited energy of the particle that produced the signal are extracted.

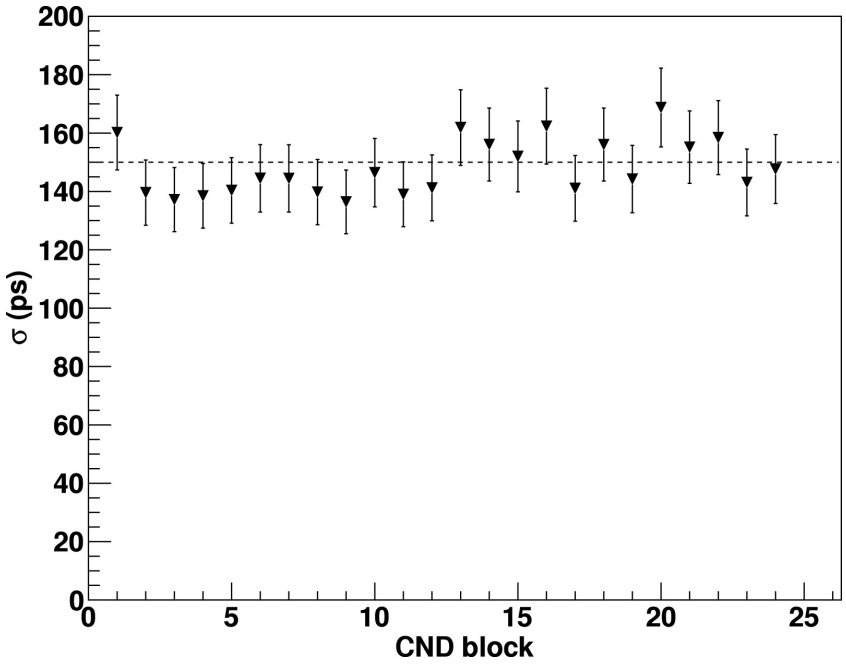
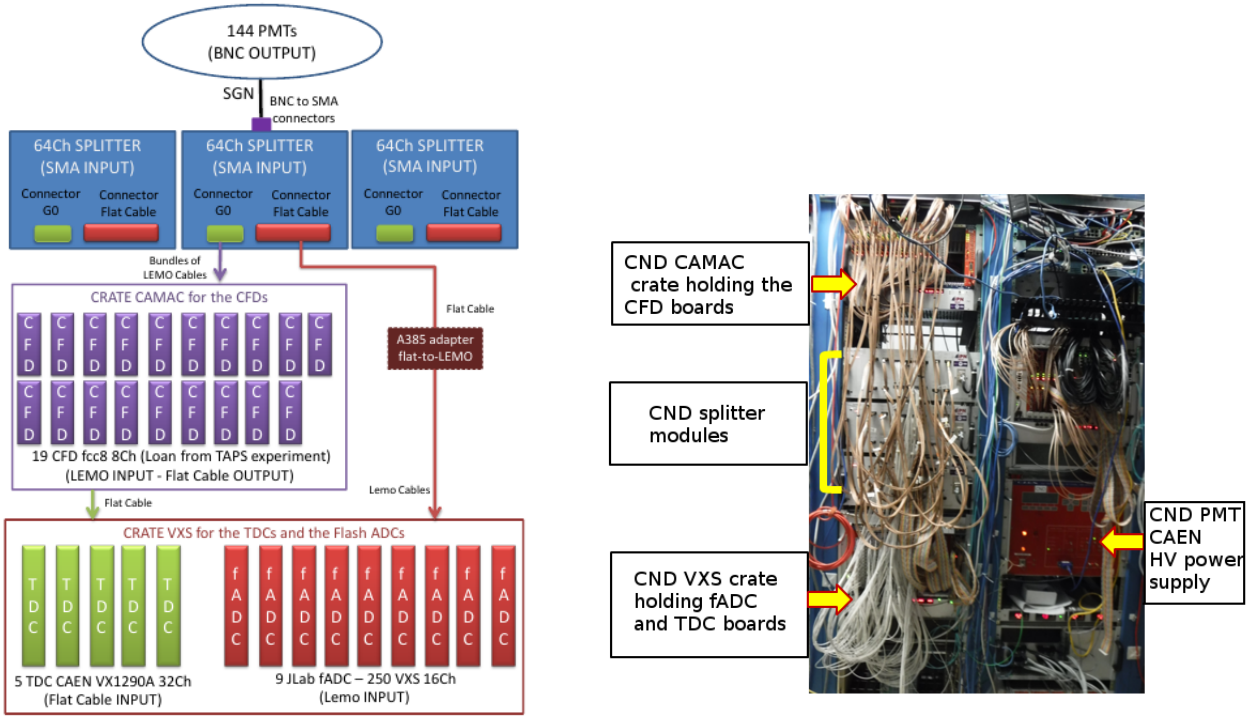


Figure 3.5: CND timing resolution measured using cosmic rays during the validation tests prior to the shipping to JLab. The average timing resolution is found to be 150 ps.



(a) (b)

Figure 3.6: (a) Schematic description of the readout electronics chain for the CND. The signals from the PMTs are splitted and transmitted to CFDs and TDCs, and fADC modules. (b) The CND electronics in Hall B.

3.2.4 High voltage setting

Each PMT was initially energized with a high voltage corresponding to a nominal gain that was determined during individual PMT tests prior to the installation. The high voltages were then adjusted

according to real data output. The value of the HV provided to each PMT determines its gain. It is necessary to have similar gains in each PMT/lightguide/scintillator system to have a constant detection efficiency for all the kinematics. The HV setting is performed with cosmic data taken with opposite sector trigger and for different value of HV. The opposite sector trigger ensures that cosmic rays cross paddles in straight tracks and deposit a constant energy. The position of the indirect ADC peak is used to equalize the gains of the PMTs. The indirect ADC is defined as the ADC associated with the indirect signal of a hit. For the HV setting, the indirect signals were defined as the signals with higher TDC (this provides a good approximation before the calibration is performed). The indirect ADC spectra was fitted with a Lorentzian and the maximum was extracted for different values of HV (see Figure 3.7). The ADC indirect peak versus HV distributions are then fitted with an exponential (see Figure 3.8) for each PMT. The HVs are then adjusted to yield the same ADC indirect peak position for every paddle. During the data taking, the gains of the PMTs were adjusted manually after long shutdowns using cosmic data, in order to have a similar indirect ADC peak for all the paddles.

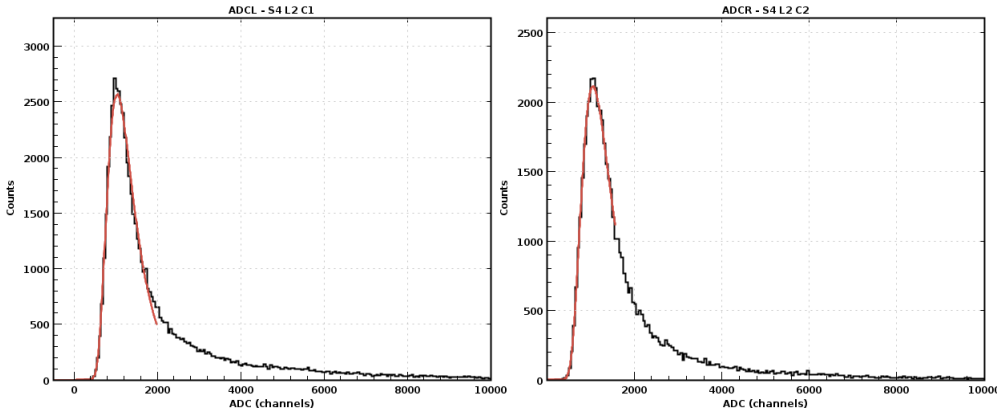


Figure 3.7: Indirect ADC spectra for cosmic rays fitted with a Lorentzian function. The position of the maximum of the fit defines the ADC indirect peak.

3.3 Decoding

During the data collection, the information recorded by the CND are the TDCs and the binned waveforms from the fADC modules. These data are associated to an electronic channel corresponding to the crate and to the board of the acquisition module. The raw data are then processed offline to reconstruct physics-related quantities.

The first step of the offline data processing is decoding. Decoding consists in two steps. First the charge signal is integrated and ADCs are calculated. The value of the pedestal is subtracted from the charge signal waveform. The pedestal corresponds to the constant value of the charge signal when no particle is detected. The charge signal waveform is binned in 4-ns bins. A threshold for pulse detection is applied to the binned waveform: the first bin with an amplitude higher than the threshold is recorded. The signals that fail the pulse detection are discarded. The integration of the charge signal is then performed in a window around the first bin above the threshold. During the CND data taking the ADC integration window was set to 12 bins before the threshold and 60 bins after. The procedure is shown in Figure 3.9, where in this case the integration window is set to a smaller width. The thresholds were set to 60 ADC units above the pedestal. Each electronics channel is then assigned a sector (1 to 24), a layer (1 to 3) and a component (left or right) according to a translation table stored in the CLAS12 database. The decoded files are written in the HIPO format and are used for the reconstruction and the calibration (see Chapter 2 for more details). The reconstruction and calibration steps are closely related and are performed alternatively at the early stage of the CND data processing. A first reconstruction is performed to match ADCs and TDCs. The file obtained is then used for the calibration before the final reconstruction happens. For clarity, the reconstruction

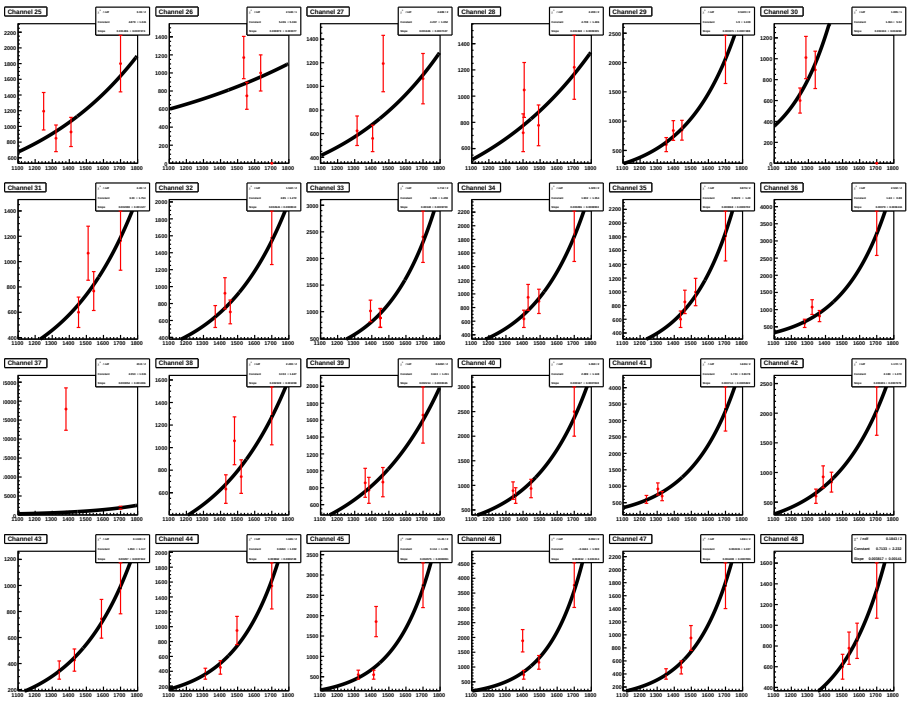


Figure 3.8: Indirect ADC peak versus HV for some paddles of the CND. The points are fitted with an exponential. The values of the HV are then chosen to yield the same indirect ADC peak in every paddle.

algorithm is presented below (Section 3.4) followed by the calibration algorithm (Section 3.5). The aim of the offline processing is to produce high-level quantities from ADCs and TDCs. In the following we call "hit" the interaction of a particle with the detector. Each hit is associated with a position, a time and an energy.

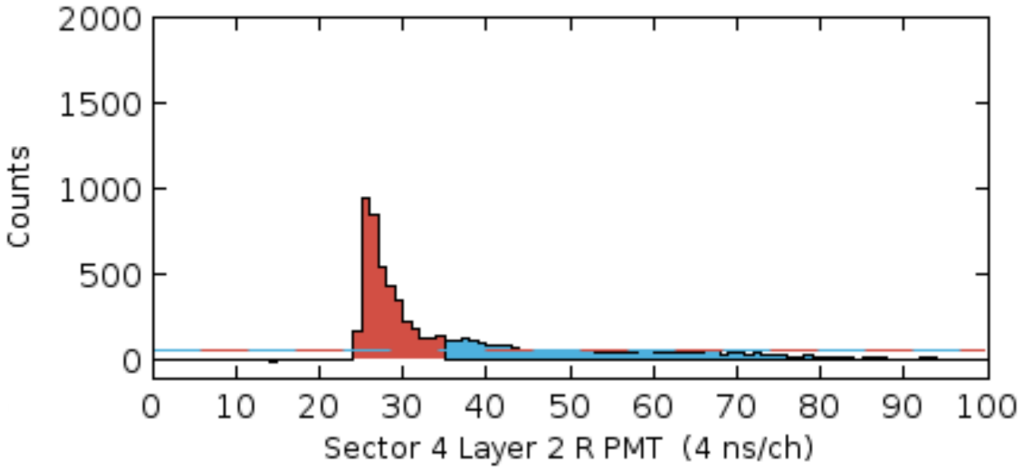


Figure 3.9: Raw signal obtained from a fADC module. The red area corresponds to the part of the signal used for the calculation of the charge integral.

3.4 Reconstruction

The reconstruction of the CND data is done in three steps. The first step is the reconstruction of the time and the position of the hit in the paddle. The second step is the reconstruction of the deposited

energy. The last step is the matching of CND hits with charged particle tracks. The reconstruction uses calibration constants that are determined during the calibration step. For the sake of clarity we suppose in this section that the calibration constants are already determined. During the CND offline data processing, the first reconstruction performed before the calibration uses either constants from previously calibrated runs or constants determined during tests prior to the installation. The constants used in the reconstruction are summarized in Table 3.2.

Constants	Name	Number of constants	Units
Effective velocity	v_{eff}	144	cm/ns
U-turn time loss	u_{t}	72	ns
Left-Right timing offset (adjusted)	t_{LRad}	72	ns
Global time offset	t_{off}	72	ns
Attenuation length	A	144	cm
Energy constants	MIP_{D}, MIP_{I}	144 each	no units

Table 3.2: Constants used in the CND reconstruction.

The effective velocity v_{eff} is the speed of light in the scintillating plastic of the paddle. The U-turn time loss term u_t accounts for the time necessary for the scintillating light to cross the u-turn light guide and reach the neighboring paddle. The two time offsets t_{LRad} and t_{off} are respectively the offset between two coupled paddles and the offset of a pair of paddles with respect to the CLAS12 common time reference (referred to as *start time* in the following). Finally the attenuation length A and the energy constants MIP_D and MIP_I are used in the energy reconstruction steps. They correspond respectively to the characteristic attenuation length of light in a paddle and to the ADC values for a direct (resp. indirect) signal of a Minimum Ionizing Particle (MIP) detected in a given paddle.

3.4.1 Associating TDCs and ADCs

The first step of the reconstruction consists in matching ADCs and TDCs associated with the same paddle. This is performed by looping through both the ADC and TDC lists and comparing sectors, layers and components (SLC). An ADC and a TDC with equal SLC are grouped together to form a half-hit. If more than one TDC or ADC are associated with one half-hit, the half-hit is discarded. Half-hits are then grouped in hits. Two half-hits form a hit if they have equal layer and sector and complementary component (a half-hit in each of the coupled paddles). The position, the time and the energy of each hit is then calculated.

3.4.2 Raw time

The raw hit times $t_{\text{L/R}}$ are obtained from the measured TDC channel using a slope constant of 0.0234 ns/channel for all the channels.

$$t_{\text{L/R}} = TDC_{\text{L/R}} \cdot 0.0234, \tag{3.6}$$

where the subscript L/R is the component index and refers to the left and right paddle of a pair, respectively.

One can then write the raw times in terms of the calibration constants defined in Table 3.2. We now denote the raw times as t_{XY} where X and Y can be R or L and refer respectively to the paddle from which the time is obtained and the paddle in which the hit occurred. For a hit in the left paddle the raw times read:

$$t_{\text{LL}} = t_{\text{off}} + t_{\text{tof}} + \frac{z}{v_{\text{eff}_{\text{L}}}} + t_{\text{S}} + \text{TDC}_{\text{j}}, \tag{3.7}$$

$$t_{\text{RL}} = t_{\text{off}} + t_{\text{tof}} - \frac{z}{v_{\text{eff}_{\text{L}}}} + \frac{L}{v_{\text{eff}_{\text{L}}}} + \frac{L}{v_{\text{eff}_{\text{R}}}} + u_{\text{t}} + t_{\text{S}} + t_{\text{LRad}} + \text{TDC}_{\text{j}}, \tag{3.8}$$

where t_{tof} is the time-of-flight of the detected particle from the vertex to the hit, z is the location of the hit along the beam direction from the upstream edge of the paddle, L is the length of the paddle, t_{S} is the event start time, TDC_{j} is the CLAS12 Radio Frequency clock jitter and the subindexes L/R refer to left or right paddles.

For a hit occurring in the right paddle the times read:

$$t_{\text{LR}} = t_{\text{off}} + t_{\text{tof}} - \frac{z}{v_{\text{eff}_{\text{R}}}} + \frac{L}{v_{\text{eff}_{\text{L}}}} + \frac{L}{v_{\text{eff}_{\text{R}}}} + u_{\text{t}} + t_{\text{S}} + \text{TDC}_{\text{j}}, \tag{3.9}$$

$$t_{\text{RR}} = t_{\text{off}} + t_{\text{tof}} + \frac{z}{v_{\text{eff}_{\text{R}}}} + t_{\text{S}} + + t_{\text{LRad}} + \text{TDC}_{\text{j}}. \tag{3.10}$$

3.4.3 Choice of the paddle where the hit occurs

The paddle in which the hit occurs is then determined prior to any further step. Defining Δ and Δ' as:

$$\Delta = \frac{L}{v_{\text{eff}_{\text{L}}}} - \frac{L}{v_{\text{eff}_{\text{R}}}}, \tag{3.11}$$

$$\Delta' = t_{\text{LX}} - t_{\text{RX}} + t_{\text{LRad}}, \tag{3.12}$$

where the index X can be R or L, one can compute $\Delta' - \Delta$ for both cases (hit in the left paddle or hit in the right paddle). If the hit is in the left paddle:

$$\Delta' - \Delta = \frac{2z}{v_{\text{eff}_{\text{L}}}} - \frac{2L}{v_{\text{eff}_{\text{L}}}} - u_{\text{t}} < 0. \tag{3.13}$$

If the hit is in the right paddle:

$$\Delta' - \Delta = \frac{2L}{v_{\text{eff}_{\text{R}}}} - \frac{2z}{v_{\text{eff}_{\text{R}}}} + u_{\text{t}} > 0. \tag{3.14}$$

Δ' and Δ are calculated for each hit. If $\Delta' < \Delta$, the paddle in which the hit happened is the left one, otherwise it is the right one.

3.4.4 Hit position and time reconstrution

The derivation of the time and the position of a hit is done in the case of a hit in the left paddle. In the case of a hit in the right paddle, the corresponding equations are obtained by switching L↔R indexes.

Subtracting the time offset known from calibration, the start time and the time jitter from Equations (3.7) and (3.8), we define t_{Lprop} and t_{Lprop} as:

$$t_{\text{Lprop}} = t_{\text{tof}} + \frac{z}{\text{v}_{\text{eff}_{\text{L}}}}, \tag{3.15}$$

$$t_{\text{Rprop}} = t_{\text{tof}} - \frac{z}{\text{v}_{\text{eff}_{\text{L}}}} + \frac{L}{\text{v}_{\text{eff}_{\text{L}}}} + \frac{L}{\text{v}_{\text{eff}_{\text{R}}}} + u_{\text{t}}. \tag{3.16}$$

The position of the hit is obtained from the difference of the above quantities as :

$$z = \frac{\text{v}_{\text{eff}_{\text{L}}}}{2}\left(t_{\text{Lprop}} - t_{\text{Rprop}} + L\left(\frac{1}{\text{v}_{\text{eff}_{\text{L}}}} + \frac{1}{\text{v}_{\text{eff}_{\text{R}}}}\right) + u_{\text{t}}\right). \tag{3.17}$$

The sum of t_{Lprop} and t_{Rprop} provides the time-of-flight of the particle that produces the hit:

$$t_{\text{tof}} = \frac{1}{2}\left(t_{\text{Lprop}} + t_{\text{Rprop}} - L\left(\frac{1}{v_{\text{eff}_{\text{L}}}} + \frac{1}{v_{\text{eff}_{\text{R}}}}\right) - u_{\text{t}}\right). \tag{3.18}$$

3.4.5 Energy reconstruction

For hits in the left paddle, the two associated ADCs can be written as:

$$ADC_{\mathrm{L}} = \frac{E_{\mathrm{L}}}{E_0}\, MIP_{\mathrm{D}}\, e^{\frac{-z}{A_{\mathrm{L}}}}, \tag{3.19}$$

$$ADC_{\mathrm{R}} = \frac{E_{\mathrm{R}}}{E_0}\, MIP_{\mathrm{I}}\, e^{\frac{-(L-z)}{A_{\mathrm{L}}}}, \tag{3.20}$$

where $E_{\mathrm{L/R}}$ are energies propagated in the left (resp. right) paddle. E_0 is half of the energy deposited by a MIP particle over the thickness of the scintillators. E_0 is given by:

$$E_0 = \frac{dE/dX\ h\rho}{2}\ \mathrm{MeV}, \tag{3.21}$$

where h is the thickness of each scintillator, ρ is the density of PVT equal to 1.023 g.cm^{-3} and dE/dX is the nominal energy deposited by MIPs in PVT and equal to 1.956 MeV.g^{-1}.cm^2. The above equations are valid for hits in left paddles, while for hits in the right paddles the corresponding equations are obtained by switching L↔R indexes. The energies measured in both paddles are then given by:

$$E_{\mathrm{L}} = \frac{ADC_{\mathrm{L}}\ E_0}{MIP_{\mathrm{D}}}\, e^{\frac{z}{A_{\mathrm{L}}}}, \tag{3.22}$$

$$E_{\mathrm{R}} = \frac{ADC_{\mathrm{R}}\ E_0}{MIP_{\mathrm{I}}}\, e^{\frac{L-z}{A_{\mathrm{L}}}}. \tag{3.23}$$

The total deposited energy is the sum of E_{L} and E_{R}:

$$E_{\mathrm{deposited}} = E_{\mathrm{L}} + E_{\mathrm{R}}. \tag{3.24}$$

3.4.6 Hit/Track matching

Tracks from charged particles crossing the CLAS12 Central Vertex Tracker (CVT) are associated to hits in the CND. This allows, for each CND hit matched with a CVT track, to calculate the position of the hit from the extrapolated track, the path-length between the track vertex and the CND hit, and the path traveled inside the hit paddle. This information is used in the calibration, as well as to veto charged particles when looking for neutrons in the CND. The CVT tracks are extrapolated to radii corresponding to the entry, middle and exit points of the track in the paddle as shown in Figure 3.10 These points are defined as the intersections between the helix of the track and the physical volume of the CND paddles. A CVT track and a CND hit are matched if the hit coordinates (x, y, and z) and the extrapolated coordinates at the middle of the counter (x_{m}, y_{m}, and z_{m}) verify the relations:

$$\mid x - x_{\mathrm{m}} \mid < \sigma_x, \tag{3.25}$$

$$\mid y - y_{\mathrm{m}} \mid < \sigma_y, \tag{3.26}$$

$$z_{\mathrm{m}} \in [-\sigma_z, L + \sigma_z], \tag{3.27}$$

where $\sigma_{\mathrm{z}} = 1.5$ cm, L is the length of a paddle, and σ_{x} and σ_{y} are given by:

$$\sigma_x = \sqrt{x^2 \frac{\sigma_{\mathrm{R}}^2}{R^2} + y^2 \sigma_\phi^2}, \tag{3.28}$$

$$\sigma_y = \sqrt{y^2 \frac{\sigma_{\mathrm{R}}^2}{R^2} + x^2 \sigma_\phi^2}, \tag{3.29}$$

where R is the radius of the hit, σ_{R} is half the thickness of a paddle (1.5 cm) and σ_ϕ is the azimuthal resolution of each paddle (3.75°). The path traveled by the particle in the paddle is approximated as the distance between the entry and exit points.

The path length between the vertex and the hit position is also calculated at this stage.

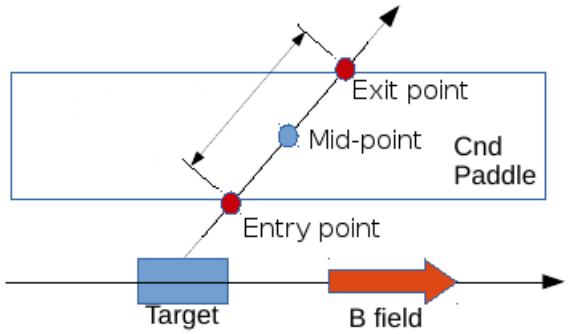


Figure 3.10: The path traveled by a particle in a CND paddle is defined as the distance between the entry and exit points of the track helix.

3.5 Calibration

The calibration of the CND is performed after a first reconstruction is performed on raw data to provide baseline data to work with. The calibration constants are obtained from these first pass reconstructed data using information from other detectors of CLAS12. Once the calibration is finished, the calibration constants are uploaded on the CLAS12 database and a second pass of reconstruction is performed with the correct constants. A schematic view of the calibration process is shown in Figure 3.11.

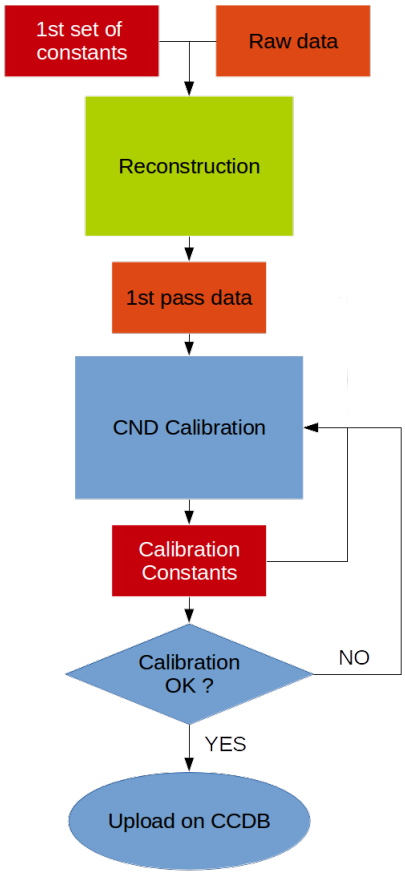


Figure 3.11: Flow chart of the CND calibration. A first pass of the reconstruction is done before the calibration in order to reconstruct hits and associate them to CVT tracks. The calibration constants can be uploaded in the calibration suite as some calibration steps require previously calculated constants.

3.5.1 Calibration constants for the CND

The calibration of the CND with beam data is done in two steps: the timing calibration, which allows us to obtain effective velocities and time offsets, which are, in turn, necessary to deduce timing and position information of the hits; and the energy calibration, in which attenuation lengths and energy conversion factors are extracted. Table 3.3 summarizes the calibration constants necessary to reconstruct CND hits and the order in which they are calculated during the calibration.

Step #	Constant name	Output	Number of constants	Units
1	Left-right timing offset	t_{LRoff}	72	ns
2	Effective velocity	v_{eff}	144	cm/ns
3	U-turn time loss	u_{t}	72	ns
4	Left-right timing offset (adjusted)	$t_{\mathrm{LR_{ad}}}$	72	ns
5	Global time offset	t_{off}	72	ns
6	Attenuation length	A	144	cm
7	Energy constants	MIP_D, MIP_I	144 each	no units

Table 3.3: The steps and the corresponding constants of the CND calibration.

3.5.2 Timing Calibration

There are five calibration constants that must be determined as part of the CND timing calibration: the two left-right time offsets (t_{LRoff} and $t_{\mathrm{LR_{ad}}}$), the effective velocity (v_{eff}), the propagation time in the U-turn (u_{t}), and the global time offset with respect to the event start time (t_{off}). The calibrations of these constants must be done in the following order: t_{LRoff}, v_{eff}, u_{t}, $t_{\mathrm{LR_{ad}}}$, and finally t_{off}. Each of these constants is determined using charged particles from beam interactions in the target.

At this stage TDCs have been converted to raw times and the hit paddle has been determined by the procedure described in Subsection 3.4.3.

3.5.3 Left-right timing offset

The left-right time offset refers to the time mis-alignement between two coupled paddles. The goal of this calibration step is to find this offset. This offset is determined in two steps. The first step relies on the u-turn structure of the CND to extract a first estimate of this offset t_{LRoff}. The second part of the algorithm adjusts this first value to the real value $t_{\mathrm{LR_{ad}}}$ by taking into account the effective velocities of both coupled paddles. There is one value of t_{LRoff} and $t_{\mathrm{LR_{ad}}}$ for each pair of paddles.

There are two different algorithms to find t_{LRoff} depending on whether the data were taken with or without magnetic field of the solenoid.

- If the solenoid field is off, the u-turn light guide coupling two adjacent CND paddles induces a gap in the time difference $t_{\mathrm{R}} - t_{\mathrm{L}}$ plots. The t_{LRoff} constant is defined as the time difference value at the center of the gap. This method was developed initially but requires special calibration runs as the effect of "double hits" described in the following was not anticipated at the early stage of software development.

- If the solenoid field is on, "double hits" occur. When the trajectory of a charged particle bents in the solenoid field crosses two adjacent coupled paddles, the two L and R signals have very similar TDC values (see Figure 3.12). Such hits induce a peak instead of a gap in the time difference plots (see Figure 3.13). In this case, t_{LRoff} is defined as the position of this peak.

Both cases are illustrated in Figure 3.13. Typical values of the offsets are below 5 ns. t_{LRoff} is not used in the reconstruction, but it is nonetheless necessary to remove double hits from the subsequent calibration steps. The constant $t_{\mathrm{LR_{ad}}}$, defined below, is used in the reconstruction.

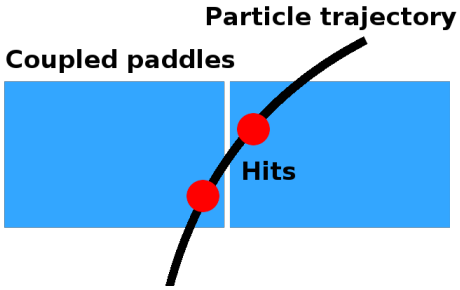


Figure 3.12: Double hits in the CND produced by the trajectory of a charged particle curved in a magnetic field. Both hits have similar TDCs resulting in a peak in the time difference distribution.

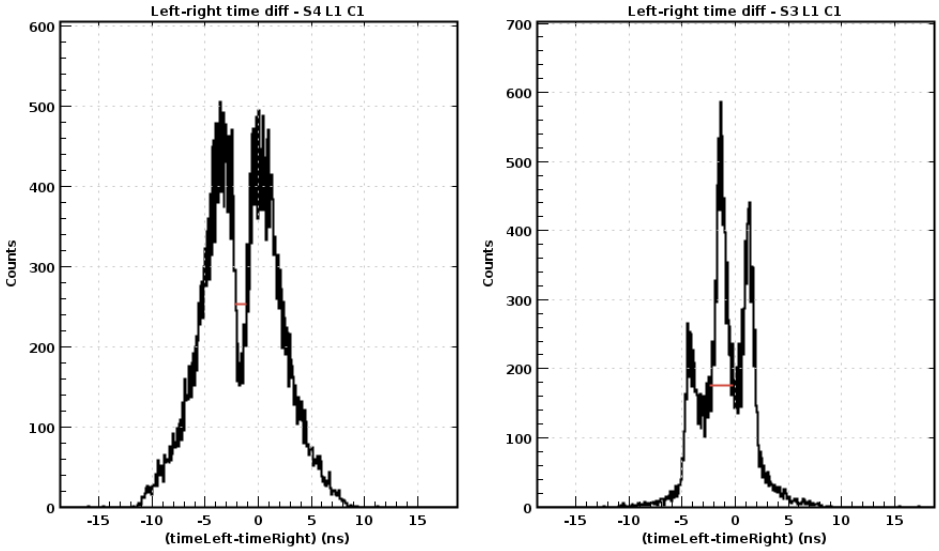


Figure 3.13: Time difference plots. The left plot corresponds to the case with no solenoid field. In this case the u-turn light guide induces a gap in the distribution. The right plot is for data with magnetic field. In this case double hits are possible (see Figure 3.12). These double hits have very similar values of both TDCs resulting in a peak instead of a gap.

Once the t_{LRoff} constants have been determined, they are corrected to account for the different effective velocities of the two coupled paddles.

For hits in the left paddle, the two associated TDCs can be expressed with Equations (3.9) and (3.10).

The offset $t_{\mathrm{LR_{ad}}}$ is defined as:

$$t_{\mathrm{LR_{ad}}} = t_{\mathrm{off_R}} - t_{\mathrm{off_L}}, \tag{3.30}$$

where $t_{\mathrm{off_R}}$ and $t_{\mathrm{off_L}}$ are time offsets associated with the left and right coupled paddles, respectively. From Equations (3.9) and (3.10), one can write:

$$\frac{t_{\mathrm{LL}} - t_{\mathrm{RL}}}{2} = \frac{z}{v_{\mathrm{eff_L}}} - C_{\mathrm{L}}, \tag{3.31}$$

where C_{L} is the opposite in sign of the intercept of $\frac{t_{\mathrm{L}} - t_{\mathrm{R}}}{2}$ vs z, and is given by:

$$C_{\mathrm{L}} = \frac{L}{2\ v_{\mathrm{eff_L}}} + \frac{L}{2\ v_{\mathrm{eff_R}}} + \frac{u_{\mathrm{t}}}{2} + \frac{t_{\mathrm{LR_{ad}}}}{2}. \tag{3.32}$$

For hits in the right paddle, the corresponding intercept C_{R} is given by:

$$C_{\mathrm{R}} = \frac{L}{2\ v_{\mathrm{eff_R}}} + \frac{L}{2\ v_{\mathrm{eff_L}}} + \frac{u_{\mathrm{t}}}{2} - \frac{t_{\mathrm{LR_{ad}}}}{2}. \tag{3.33}$$

Combining Equations (3.32) and (3.33), $t_{\mathrm{LR_{ad}}}$ is obtained as:

$$t_{\mathrm{LR_{ad}}} = C_{\mathrm{L}} - C_{\mathrm{R}}. \tag{3.34}$$

3.5.4 Effective velocity

The effective velocity v_{eff} is the speed of the light in the scintillators. There is one v_{eff} value for each coupled paddle. For a hit in the left paddle, the v_{eff} is obtained from the following equation:

$$z = (t_{\mathrm{LL}} - t_{\mathrm{RL}})\frac{v_{\mathrm{eff_L}}}{2} + c, \tag{3.35}$$

where z is the longitudinal position of the hit in the CND with respect to the upstream end of the CND paddles and c is a constant detailed in the previous section. For hits in the right paddles, the indexes of the time difference must be changed. The longitudinal position z is obtained independently from the CND, using the extrapolation of the particle track measured by the CVT. The v_{eff} is extracted by fitting the $\frac{t_{\mathrm{L}}-t_{\mathrm{R}}}{2}$ vs z distribution as shown in Figure 3.14. For each slice in z, the position of the maximum from a Gaussian fit is plotted against z. The slope of the obtained distribution gives v_{eff}. The expected values for v_{eff} are around 14-16 cm/ns.

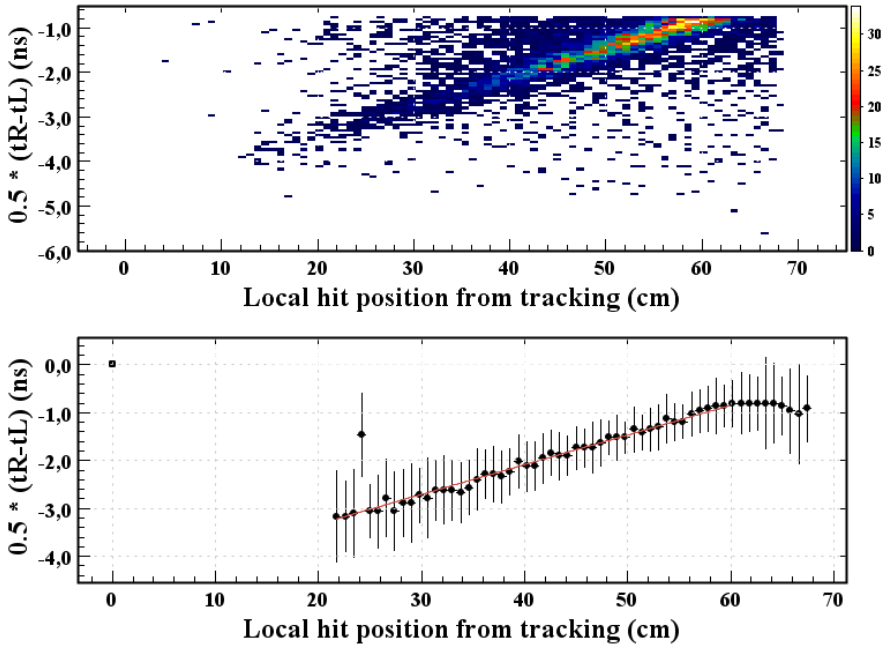


Figure 3.14: Plots used to determine the effective velocity for a right CND paddle. The top plot shows the raw $\frac{t_{\mathrm{R}}-t_{\mathrm{L}}}{2}$ vs z and the bottom plot is the distribution showing the linear fit.

3.5.5 U-turn time loss

The u-turn propagation time u_{t} is the time spent by the light to travel through the u-turn lightguide. It is used as a time offset on the indirect signal in the time and position reconstructions. There is one u_{t} value for each pair of paddles. The algorithm to extract u_{t} is very similar to the one used in the v_{eff} procedure: the intercept of the $\frac{t_{\mathrm{L}}-t_{\mathrm{R}}}{2}$ vs z distribution (see Figure 3.14) is extracted for both coupled paddles to determine u_{t}.

From the sum of the intercepts C_{L} in Equation (3.32) and C_{R} in Equation (3.33), u_t is obtained as :

$$u_{\mathrm{t}} = C_{\mathrm{R}} + C_{\mathrm{L}} - L\left(\frac{1}{v_{\mathrm{eff_R}}} + \frac{1}{v_{\mathrm{eff_L}}}\right). \tag{3.36}$$

The values for u_{t} are typically in the range from 0.5 ns to 1.5 ns, with values for the layer 1 around 0.6 ns, the layer 2 around 1 ns, and the layer 3 around 1.4 ns.

3.5.6 Global time offset

The global time offset t_{off} refers to the time difference between the start time value and the vertex time computed from the CND hit time and the CVT path length information. There is one t_{off} value for each pair of coupled paddles. It is given by:

$$t_{\text{off}} = \frac{t_{\text{L}} + t_{\text{R}}}{2} - t_{\text{S}} - t_{\text{tof}} - \frac{L}{2} \cdot \left(\frac{1}{v_{\text{eff}_{\text{R}}}} + \frac{1}{v_{\text{eff}_{\text{L}}}} \right) - \frac{u_{\text{t}}}{2} - \frac{t_{\text{LR}_{\text{ad}}}}{2} - \text{TDC}_{\text{j}}, \tag{3.37}$$

where t_{tof} is calculated using CVT information assuming the particles are pions. For this, only hits with negative charge tracks are used, as most of the negative particles in the CD are pions. The time-of-flight is then fully determined by the tracking information as:

$$t_{\text{tof}} = \frac{d_{\text{path}} \cdot \sqrt{p^2 + m^2}}{p \cdot c} \tag{3.38}$$

where d_{path} is the distance traveled by the pion from the vertex to the hit calculated at the hit-track matching stage of the reconstruction algorithm, p is the momentum of the track obtained from its curvature in the magnetic field, m is the mass of the pion and c the speed of light. The start time t_{S} is provided event by event and is calculated from the FTOF measured time and the RF bucket matching (see Chapter 4 for more details). The start time also accounts for the position of the vertex. The position of the peak in the distribution shown in Figure 3.15 gives t_{off}. This constant mainly depends on the start time t_{S}, which is calculated using the CLAS12 Forward Time-of-Flight System (FTOF). The CND global time offset can only be calibrated after the calibration of the FTOF. The variations of t_{off} between different pairs of paddles are typically below 10 ns.

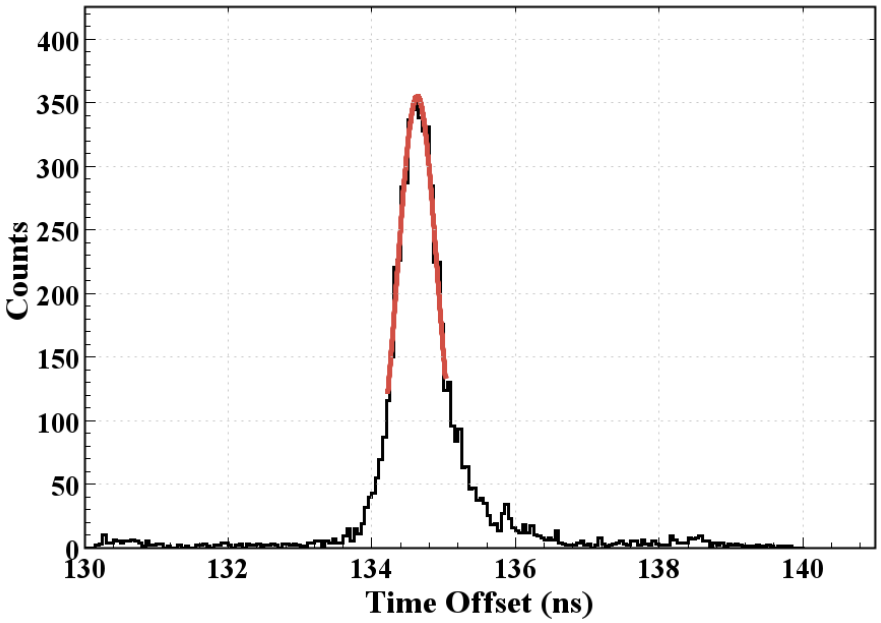


Figure 3.15: Plot used to determine the global time offset. The distribution is fit with a Gaussian and the position of the peak corresponds to t_{off}.

3.6 Energy calibration algorithms

There are three calibration constants for the energy determination in each paddle of the CND: the attenuation length (A), the ADC-to-energy constants for direct minimum-ionizing particles (MIPs) (MIP_{D}), and the ADC-to-energy constants for indirect MIPs (MIP_{I}). These three calibration steps can be performed almost independently from the timing calibration. However $t_{\text{LR}_{\text{ad}}}$ is needed to determine if an ADC signal is direct or indirect (i.e. the hit happened in the considered paddle or in its coupled partner).

3.6.1 Attenuation length

The attenuation length A accounts for the light attenuation along the length of the scintillators and light guides. There is an A value for each paddle. For hits in the left paddle, the two associated ADCs can be written as:

$$ADC_{\mathrm{L}} = \frac{E}{E_0} MIP_{\mathrm{D}}\, e^{\frac{-z}{A_{\mathrm{L}}}}, \tag{3.39}$$

$$ADC_{\mathrm{R}} = \frac{E}{E_0} MIP_{\mathrm{I}}\, e^{\frac{-(L-z)}{A_{\mathrm{L}}}}, \tag{3.40}$$

where MIP_{D} and MIP_{I} are constants defined in Subsection 3.6.2 below (MIP_{I} includes the effect of the light attenuation in the R paddle), E is half the energy deposited by the particle in the scintillator, and E_0 is half the energy deposited by a MIP in the scintillators. We have assumed here that the deposited energy is shared equally in the direct and indirect signals. E_0 is given by:

$$E_0 = \frac{2.001\ h}{2}\ \mathrm{MeV}, \tag{3.41}$$

where h is the thickness of each scintillator. All the above equations are valid for hits in the left paddles, while for hits in the right paddles the corresponding equations are obtained by switching the L↔R indexes. From Equations (3.39) and (3.40) the following relation is derived:

$$\ln(ADC_{\mathrm{L}}/ADC_{\mathrm{R}}) = c - \frac{2z}{A_{\mathrm{L}}}, \tag{3.42}$$

where c is a constant depending on MIP_{D}, MIP_{I} and L. A_{L} is given by the slope of the distribution in Equation (3.42) as shown in Figure 3.16. Values for A are typically around 150 cm.

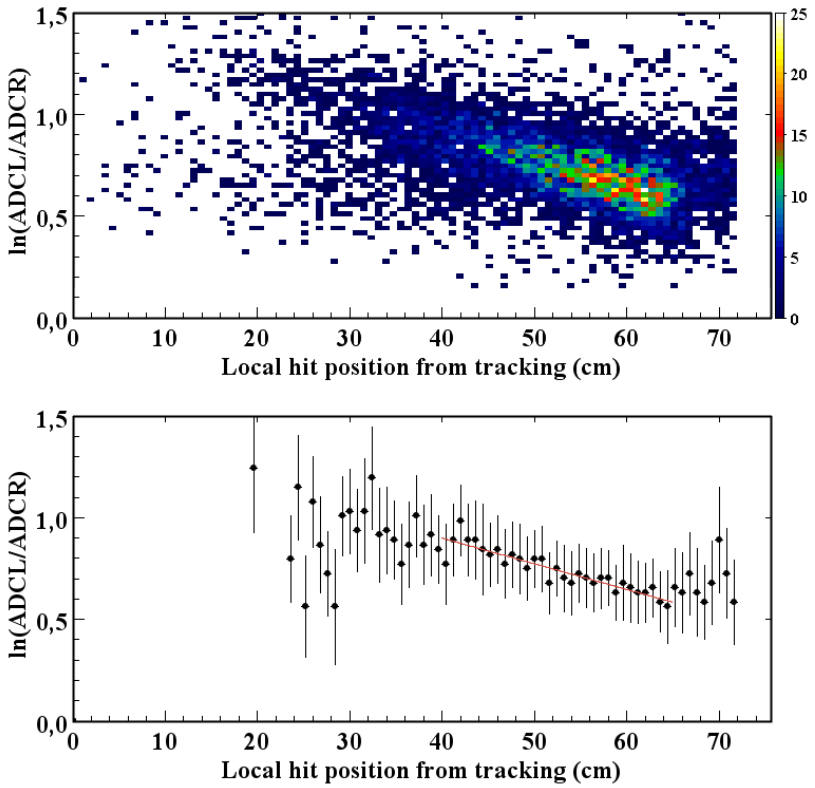


Figure 3.16: Plots used to determine A. The top plot shows the raw $\ln(ADC_{\mathrm{L}}/ADC_{\mathrm{R}})$ vs z distribution for one pair of paddles. Slices in z are fit with a Gaussian and the mean is plotted against z. The resulting distribution and its associated linear fit are shown in the bottom plot.

3.6.2 Energy calibration

The final step of the calibration of the CND is the determination of the energy conversion parameters MIP_{D} and MIP_{I}. There are two energy parameters for each paddle, thus there are four energy parameters for each pair of coupled paddles, denoted as MIP_{DL}, MIP_{IL}, MIP_{DR}, MIP_{IR}.

In the following, we only consider a hit in the left paddle. Equations for hits in right paddles are obtained by switching the L↔R indexes. For hits in the left paddle, only MIP_{DL} and MIP_{IL} can be obtained. In the following they are referred to as MIP_{D} and MIP_{I}. From Equations (3.39) and (3.40), one gets:

$$\ln\left(\frac{ADC_{\mathrm{L}}}{ADC_{\mathrm{R}}}\right) = \ln\left(\frac{MIP_{\mathrm{D}}}{MIP_{\mathrm{I}}}\right) + \frac{L}{A_{\mathrm{L}}} - \frac{2z}{A_{\mathrm{L}}} \tag{3.43}$$

$$\sqrt{ADC_{\mathrm{L}}\ ADC_{\mathrm{R}}} = \frac{E}{E_0}\sqrt{MIP_{\mathrm{D}}\ MIP_{\mathrm{I}}}\ e^{-\frac{L}{2A_{\mathrm{L}}}}. \tag{3.44}$$

From Equation (3.43) the intercept of the $\ln\left(\frac{ADC_{\mathrm{L}}}{ADC_{\mathrm{R}}}\right)$ vs z distribution (in Figure 3.16) gives the ratio $\frac{MIP_{\mathrm{D}}}{MIP_{\mathrm{I}}}$. The product $MIP_{\mathrm{D}} \times MIP_{\mathrm{I}}$ is obtained using Equation (3.44) after filtering MIPs and correcting for the path traveled by the MIP in the scintillators. Indeed for MIPs, E can be written as:

$$E = \frac{path}{h} \cdot E_0, \tag{3.45}$$

where *path* is the path length traveled by the MIP in the scintillator, which is obtained using the CVT tracking information by extrapolating the particle trajectory at the radius of the CND hit. Selecting MIPs and correcting for the path length removes the energy dependence from Equation (3.44), which becomes:

$$\sqrt{ADC_{\mathrm{L}}\ ADC_{\mathrm{R}}} = \frac{path}{h}\sqrt{MIP_{\mathrm{D}}\ MIP_{\mathrm{I}}}\ e^{-\frac{L}{2A_{\mathrm{L}}}}. \tag{3.46}$$

The distribution of $\sqrt{ADC_{\mathrm{L}}\ ADC_{\mathrm{R}}}\ \frac{h}{path}$ is fit with a Landau function and the position of the peak p is extracted as shown in Figure 3.17. Finally MIP_{D} and MIP_{I} are given by:

$$MIP_{\mathrm{D}} = p \times e^{\frac{i}{2}}, \tag{3.47}$$

$$MIP_{\mathrm{I}} = p \times e^{\frac{-i}{2}+\frac{L}{A}}, \tag{3.48}$$

where i and p are the intercept and the peak position defined above. The values of MIP_{D} and MIP_{I} are typically around 2000 and 500, respectively.

3.7 Calibration suite

The CND calibration algorithm is embedded in a calibration suite developed in Java. The framework of the suite is common to all the CLAS12 subsystems. Each calibration step is performed within a dedicated tab. The required previous constants for a given step can be uploaded using directly the output text files of the suits. Calibrators can check the goodness of each fit visually as well as numerically using the error on the extracted constants. Each plot used in a particular calibration step is shown for each paddle and can be exported for a later review. A screenshot of the suite is shown in Figure 3.18. Once the calibration is satisfactory, the output text files can be uploaded on the CLAS12 database (CCDB) for later use.

3.8 Clustering

Particles can deposit energy in several paddles of the CND. All the hits produced by a given single particle are grouped in a cluster. The clustering of CND hits is based on the geometrical space-time distance between them. The determination of the maximal distance for clustering two hits together takes into account the measured resolutions for the position and the timing of the CND counters.

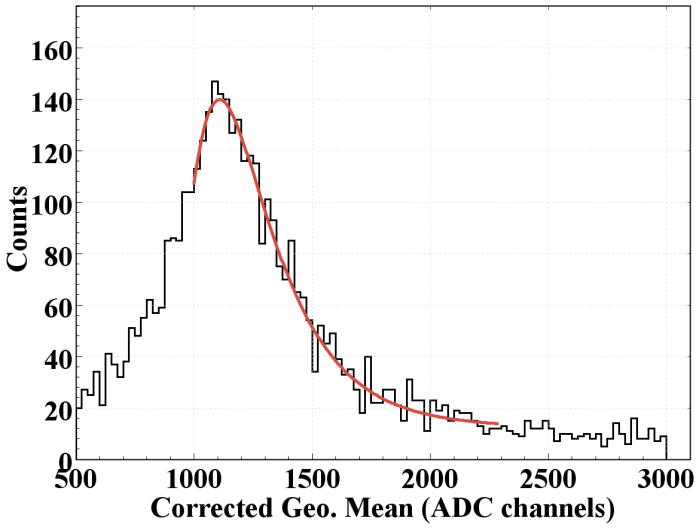
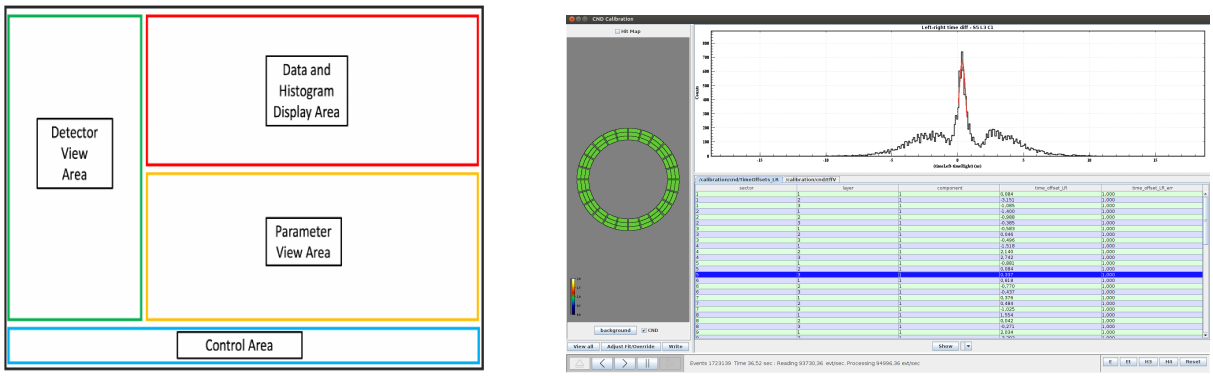


Figure 3.17: $\sqrt{ADC_{\mathrm{L}}\ ADC_{\mathrm{R}}}\ \frac{h}{path}$ distribution fit with a Landau function. The events in this plot are identified as MIPs by requiring a pion. The particle identification is performed requiring a negative charge, as most of the negatively charged particles in the Central Detector are pions.



(a) (b)

Figure 3.18: (a) General framework for the CLAS12 calibration suites. (b) CND calibration suite screen display. The plots used at each steps are displayed as well as the extracted constants for each component of the detector. The various calibration steps are performed within individual tabs.

The algorithm uses standard hierarchical clustering [92]. A scan of all the hits in an event is performed and only the hits with a deposited energy greater than 1 MeV are considered for clustering. The two closest hits are combined into a single hit with an associated energy defined as the sum of the energies of both hits. The position and the timing of the cluster hit are defined as those of the hit with the highest energy, i.e. the seed hit. The same algorithm is recursively run on the remaining hits. Finally, the leftover hits that are relatively far from each other are called clusters. The sector, layer, and component of each cluster are those of the seed hit.

3.9 Simulation geometry and hit process

The CND geometry is implemented in Geant4 based CLAS12 simulation framework (GEMC) using its geometry API. The paddles are Geant4 generic trapezoids (see Figure 3.19). The u-turn light guides are Geant4 "polycones" (volumes with cylindrical symmetry with varying radius along one axis). The paddles are assigned the scintillator material and are associated with the CND hit process routine described below.

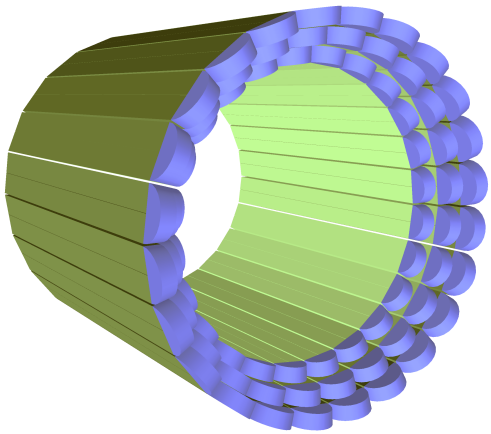


Figure 3.19: Overall view of the CND detector in GEMC. Pairs of scintillators are connected through a scintillator u-turn junction.

The energy deposited is scaled based on the position in the paddle using the calibrated light attenuation length. Two signals are then propagated, one to the PMT attached to the scintillator ("direct hit"), and one traveling through the u-turn lightguide into the coupled paddle and its PMT ("indirect hit"). Layer-dependent factors, applied to the two signals, account for the light loss in the u-turn and in the neighboring paddle. These factors were determined during R&D tests.

The corrected energy is converted into the theoretical number of photons N_{th} using the constant 1210 γ/MeV, which accounts for light propagation in the 1.4-m-long light guides, for losses at the junctions and for the quantum efficiency of the PMT. A Poissonian distribution is used to calculate the actual number of photons N_{actual} and the resulting "smeared" energy is then converted to ADC.

The absolute hit time is corrected using the calibration constants estimated from data (effective velocity and time offset). The Birks attenuation effect is also taken into account. The Birks factor, modifying the deposited energy depending on the particle type, enters in the timing calculation as follows: the direct and indirect times are smeared with a Gaussian function having a width directly proportional to an empirically determined Birks constant, and inversely proportional to the square root of the measured number of photons (which is, in turn, proportional to the attenuated energy). The time window of the CND is set to 400 ns: all of the Geant4 steps within the same paddle and time window are collected into one hit. The time is then digitized using the TCD module sampling time.

3.10 Performances

The performances of the CND are estimated in three distinct ways. First we estimate the time, position and energy resolutions using π^-, which are minimum ionizing particle in CLAS12 kinematics. Also we estimate the particle identification capabilities of the CND. Finally we measure the neutron efficiency of the CND using the exclusive reaction $ep \to en\pi^-$.

3.10.1 Time, position and energy resolutions

The timing performance for each of the three layers of the CND is illustrated in Figure 3.20b, showing the vertex time difference t_v for selected negative tracks, assumed to be all pions, integrated over all the sectors. It is defined as

$$t_v = t_{\mathrm{CND}} - (t_{\mathrm{S}} - vz_{corr}) - \frac{d_{\mathrm{path}}}{c\ \beta}, \tag{3.49}$$

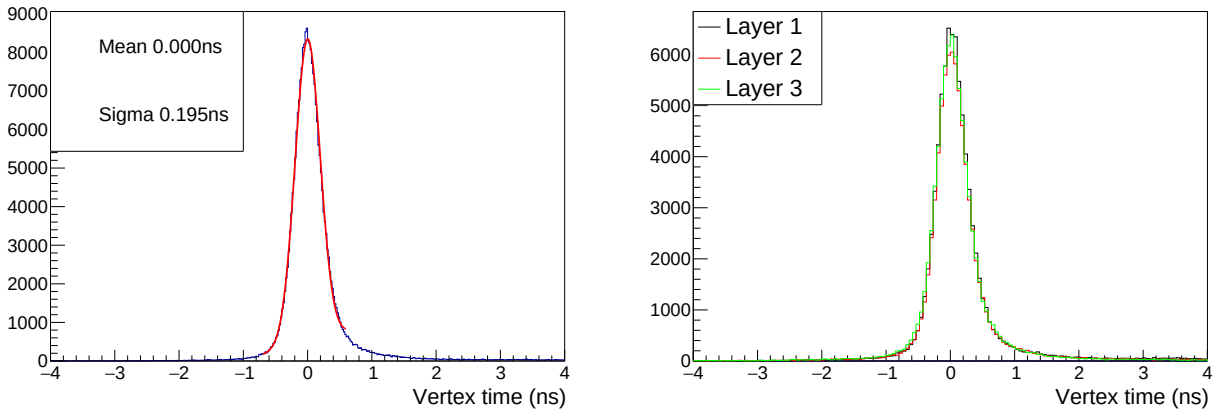
where t_{CND} is the mean time reconstructed from the L and R paddles, t_{S} is the event start time determined by the RF bucket matched with the FTOF vertex time, vz_{corr} accounts for the actual z position of the interaction vertex, d_{path} is the path length from the event vertex to the CND, and

$$\beta = \frac{p}{\sqrt{p^2 + m^2}}, \tag{3.50}$$

with p the momentum measured by the CVT, and m the pion mass. The distribution of t_v is centered at 0. From the width of the t_v distribution, the timing resolution of each CND paddle convoluted with the CVT resolution can be determined as:

$$\sigma_t = \sqrt{\sigma_{v_t}^2 - \sigma_{t_S}^2} = 185 \text{ ps}, \tag{3.51}$$

assuming that the resolution of the start time is σ_{t_S}=20 ps. This value is a higher limit of the start time resolution and thus the CND resolution is slightly overestimated. However using the lower estimate value of the start time resolution σ_{t_S}=2 ps, the CND resolution estimate varies by less than 1%. The timing resolution for the 144 individual CND counters is shown in Figure 3.21. Its average (indicated by the horizontal line) is around 185 ps, which is more than the average 148 ps intrinsic timing resolution measured with cosmic rays (see Section 3.5). This discrepancy is due to multiple factors, such as the current, not fully optimized, quality of the calibration and reconstruction of CLAS12, as well as the uncertainty of the path length and other non-CND contributions to Equation (3.49). The resolution integrated over all the paddles of the CND is shown in Figure 3.20a. The overall resolution (195 ps) is slightly higher than the average single paddle resolution due to remaining small mis-alignments.



(a) (b)

Figure 3.20: Difference between the vertex time of particles with an associated negative track, computed with CND and CVT information, and the start time computed by the FTOF: (a) integrated over all the CND paddles and (b) in each of the three layers of the CND and integrated over all paddles in each layer.

The position reconstruction performance of the CND is shown in Figure 3.22, which displays the difference between the z coordinate (along the beamline) computed by the CND and by the CVT for negative tracks in each of the three layers of the CND integrated over all the paddles. Its Gaussian width is ∼3 cm, which is consistent with the timing resolution shown in Figure 3.21 multiplied by typical effective light velocity values of 14 cm/ns. The individual position resolutions are shown for all the paddles in Figure 3.22b. Finally the extrapolated position of the hits measured by the CVT versus their position calculated by the CND for all the detected charged particles is shown in Figure 3.23. Hits are located on the diagonal, showing a good position calibration. The band at high z_{CND} corresponds to double-hits (see Figure 3.12) where the direct and indirect signals have similar timings mimicking hits close to the u-turn lightguide.

Figure 3.24 shows the energy deposited divided by the path length for selected MIPs. It peaks around the expected value of 2.001 MeV/cm for MIP particles.

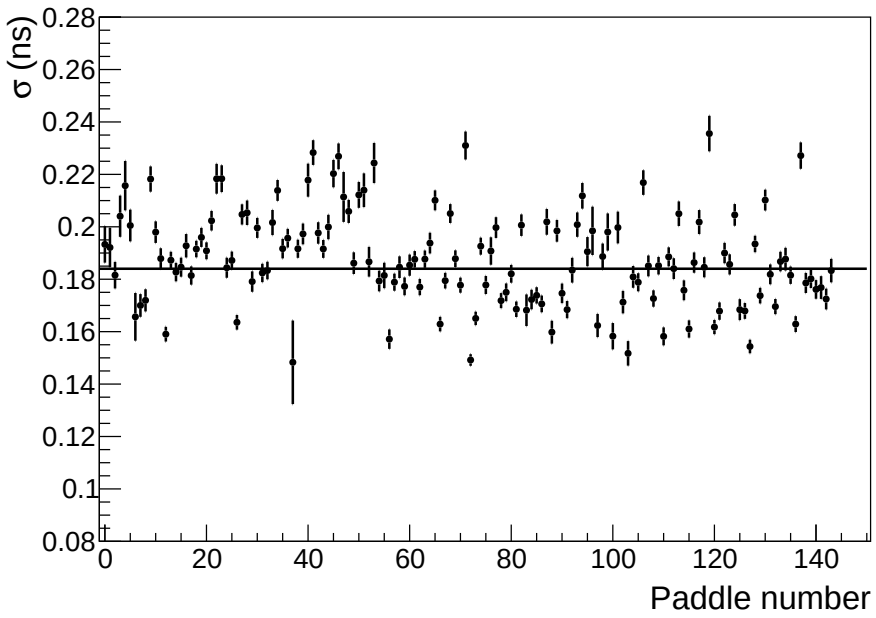
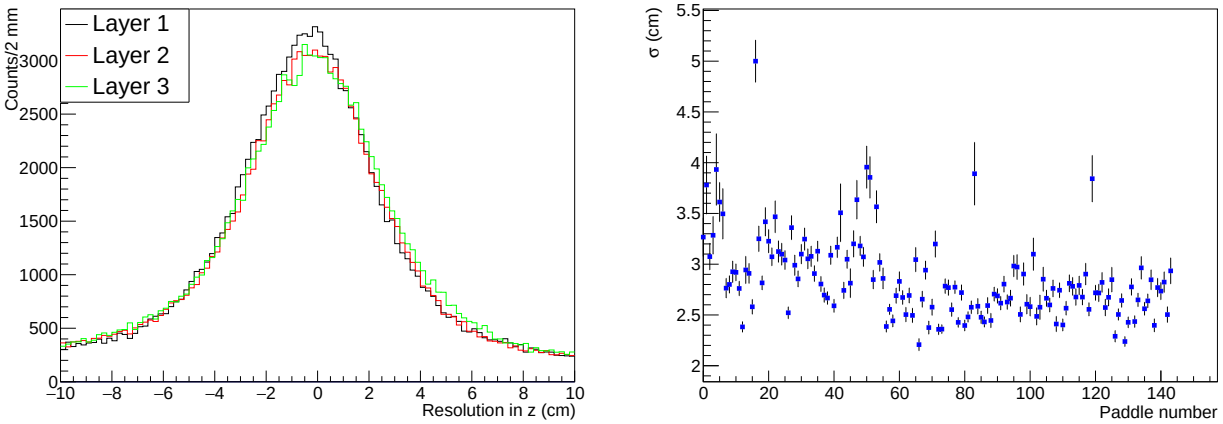


Figure 3.21: Timing resolution for each CND counter convoluted with the CVT resolution. The horizontal line indicates the average value.



(a) (b)

Figure 3.22: (a) Difference between the z coordinate (along the beamline) computed by the CND and the CVT for negative tracks in all the CND layers integrated over all the paddles. (b) Position resolution for each paddle of the CND. Points with large position resolution are located above CVT holes where the track reconstruction performance is reduced.

3.10.2 Particle identification performances

Charged particle identification

The CND was designed to detect neutrons. However it is also possible to use it to identify charged particles. We can assess the quality of charge particle identification by plotting β_{CND}, calculated from

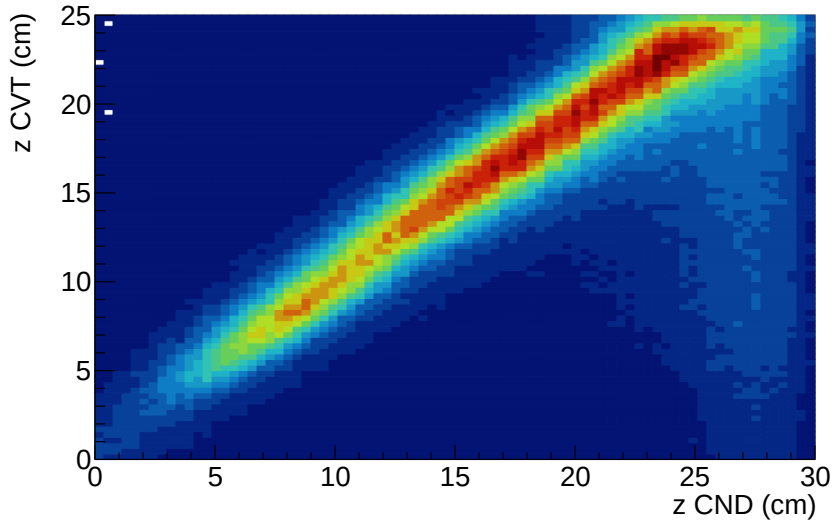


Figure 3.23: Position of the hits in the CND extrapolated from central tracks versus the measured position calculated by the CND. The vertical band at high z_{CND} is due to double-hits.

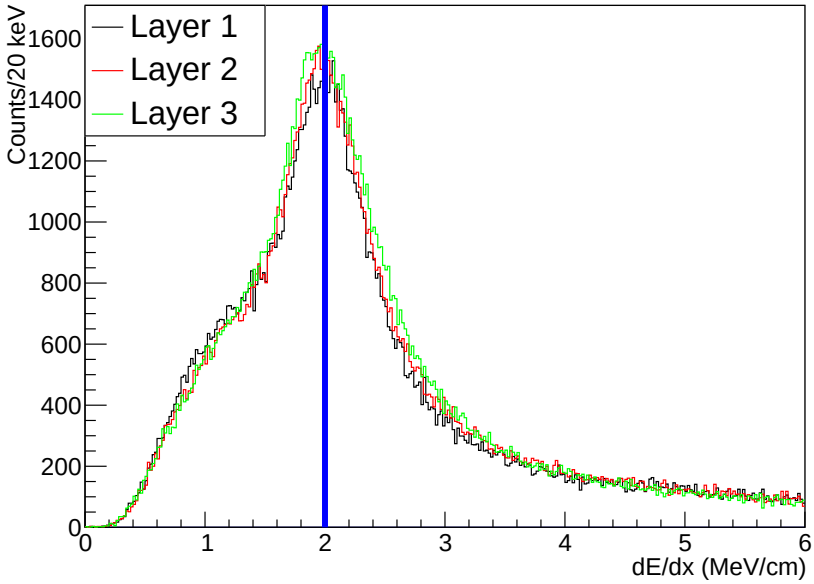


Figure 3.24: dE/dx for MIPs in the three layers of the CND integrated over all the sectors. The blue line indicates the nominal value for the expected energy deposit of a MIP in a centimeter of plastic scintillator.

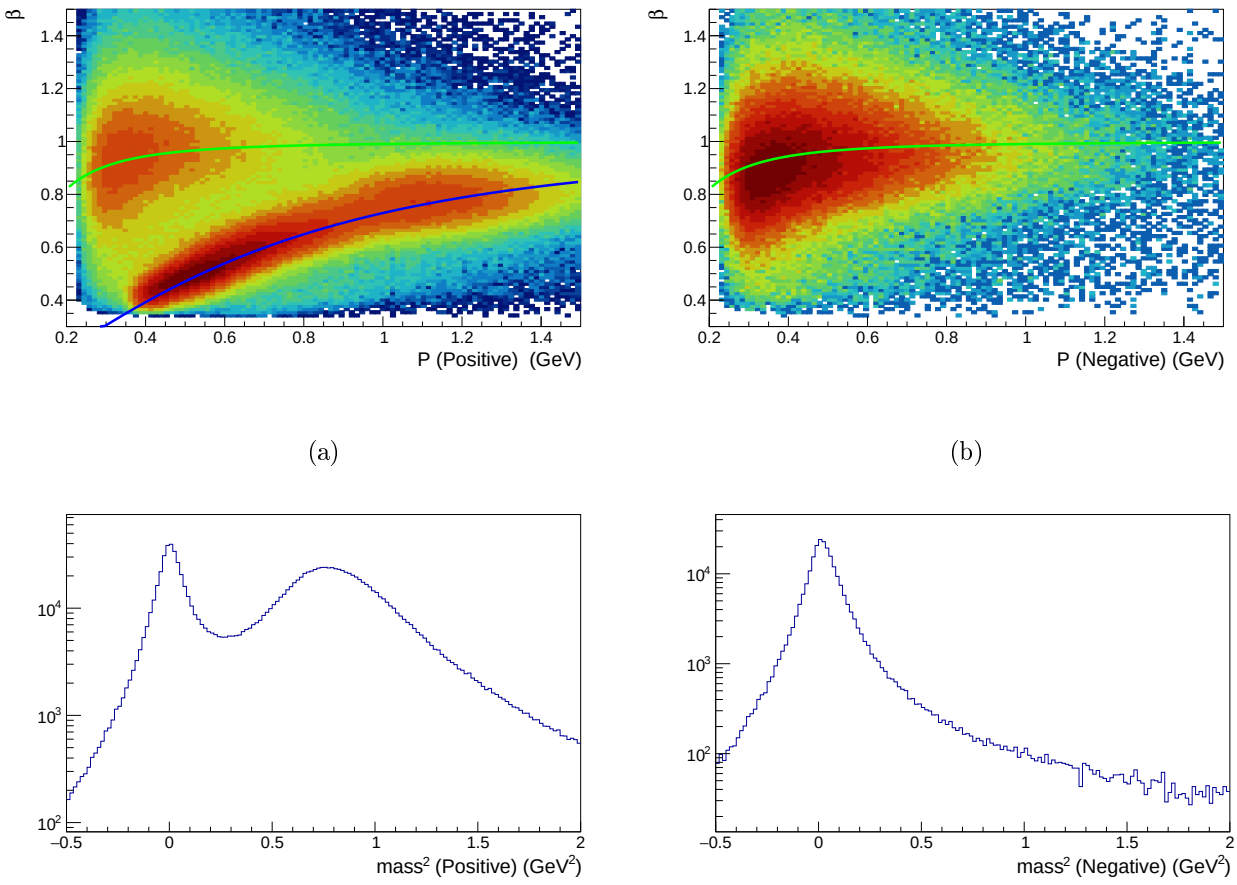
the CND time, against the momentum p measured from tracking, with β_{CND} defined as:

$$\beta_{CND} = \frac{d_{\text{path}}}{(t_{CND} - t_{\text{S}})\ c}. \tag{3.52}$$

Here d_{path} is the path traveled by the particle from the vertex to the detection point and t_{S} is the event start time. Figures 3.25a and 3.25b show β_{CND} versus momentum plots for both positive and negative central tracks. Protons and pions are clearly separated and follow the expected dependencies. The time-of-flight squared masses in both cases are also shown in Figures 3.25c and 3.25d. The time-of-flight

squared masses are given by:

$$m^2_{\text{tof}} = p^2 \left(\frac{1}{\beta^2_{CND}} - 1 \right). \tag{3.53}$$



(a) (b)

(c) (d)

Figure 3.25: Charged particles identification plots using the CND and the CVT. β_{CND} vs p is shown for (a) positive tracks and (b) negative tracks. The time-of-flight mass m^2_{TOF} is shown also for (c) positive and (d) negative tracks.

Neutron-photon separation

The timing resolution of the CND is designed to be around 150 ps in order to be able to separate neutrons from photons in the 0.2 to 1 GeV momentum region. The effective timing resolution of the CND in CLAS12 is estimated to be around 160 ps for real data after optimized calibration. To estimate the separation power achievable with such a resolution multiple methods are compared.

First, a simple error propagation calculation is performed. We assume that the timing resolution δt (set at 160 ps) and the position resolution δz are constant along a CND paddle, and that the position resolution is given by the timing resolution multiplied by the typical value of the light effective velocity (14 cm/ns). We also assume that the timing resolution for neutrals is similar to the one measured for MIPs. The neutrons and the photons are separated using a cut on the value of β calculated as:

$$\beta = \frac{d}{t\ c}, \tag{3.54}$$

where d is the distance between the vertex and the interaction point and t is the time of flight measured by the CND. The uncertainty of β is given by:

$$\delta\beta = \frac{\beta}{d}\sqrt{(\delta d)^2 + (\delta t)^2 \beta^2 c^2}. \tag{3.55}$$

The uncertainty on the path length is determined by:

$$\delta d = \frac{1}{2}\left(\sqrt{R^2+z^2+2z\delta z+\delta z^2}-\sqrt{R^2+z^2-2z\delta z+\delta z^2}\right)\approx\frac{1}{2}\frac{z\delta z}{d}, \tag{3.56}$$

where all the variables are defined in Figure 3.26. We do the calculation in the case of a hit in the first layer (R=30 cm, where δd is bigger). The values of z and d are given by the polar angle for which the calculation is made.

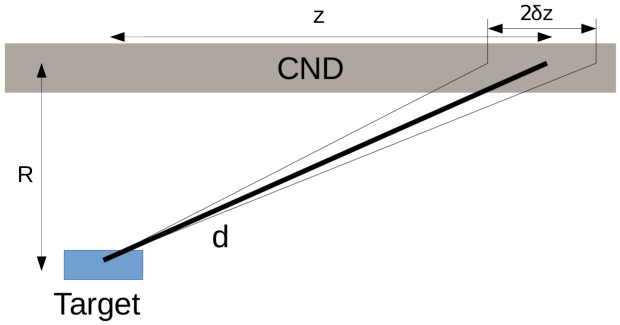


Figure 3.26: Schematic definition of the variables used in the determination of the neutral particle separation. The radius R is taken to be the radius at the center of the first layer. The vertex is located at the center of the target which is also taken as the origin of the z axis.

A second method to estimate the β resolution consists in calculating β from randomly picked values of t and d within the resolution specified before. A random path length value d_{R} is generated from a gaussian with mean d and standard deviation δd. The value of t_{R} is generated from a gaussian with mean $d/(t\beta)$ and standard deviation δt. The corresponding random value β_{R} is defined as $d_{\mathrm{R}}/(t_{\mathrm{R}}c)$. This process is iterated 10000 times and the resulting distribution of β_{R} is fitted with a gaussian. The standard deviation of the fitted function give the resolution. This method is referred as *fit* method in the following.

One can also estimate the β resolution using directly the CLAS12 data. In this case, the exclusive reaction $ep \to e'n\pi^+$ (see Subsection 3.10.3 for more details) is used. The distribution of the difference between the expected β from missing neutron events and the one measured by the CND is fitted with a gaussian for different slices in momentum. The fitted standard deviations give the resolution in each momentum bin. In the case of photons, there is no exclusive reaction that can be used to reliably extract the resolution. Thus only the β resolution for neutron is extracted from data.

Finally, the resolution is also extracted from simulations. This is done by simulating neutrons and photons in the CND using GEMC. The reconstructed β is then compared to the generated one. The resolution is obtained as for the real data case, using gaussian fits in each momentum bins.

Figure 3.27 visualizes the neutron/photon separation power of the CND, showing β as a function of momentum for photons (squares) and neutrons (dots). The error bars correspond to 2 σ, where σ is the β resolution for the two kinds of particles, evaluated with the different methods described above. The results obtained with the error propagation, fit, simulations and real data are in good agreement in the high-momentum region, which is the crucial one for photon/neutron separation. For the photons, we relied only on simulations (black), error propagation (green) and *fit* (red). We can conclude that neutrons and photons can be separated at a 2-sigma level for momenta up to 0.8 GeV. It must be noted that in [84] simulations showed a separating power up at a 3-sigma level for neutrons up to 0.9 GeV. The reason for the worsened performances is under investigation. The main change since [84] was in the clustering algorithm of the data reconstruction, passing from a energy-weighed average time for the cluster to a method based on the attribution of the time seed-hit (the most energetic one) to the whole cluster. The latter method was chosen to match the scheme adopted for the CTOF. However, this method may not be optimal for neutrals, and further studies are needed.

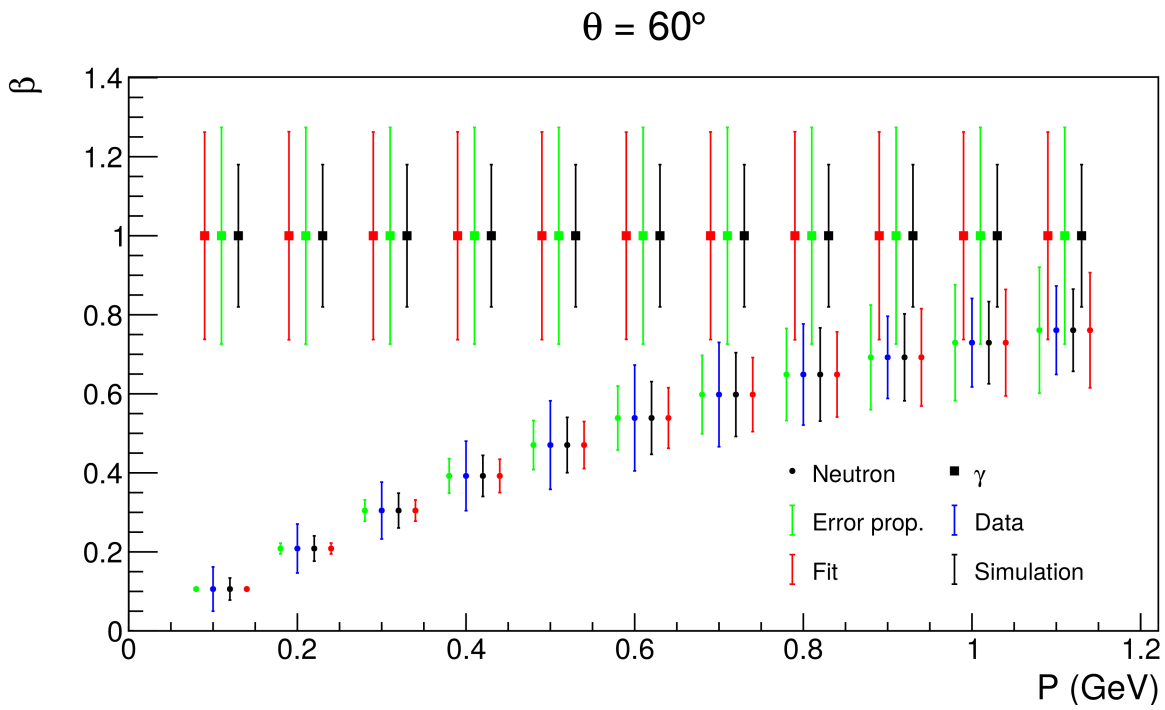


Figure 3.27: β versus momentum for the neutral particles emitted at polar angle 60° and detected in the CND. The error bars correspond to 2 σ. The photons are represented by the squares and the neutrons by the dots. The different colors correspond to the different methods presented in the text. The green error bars are obtained using the error propagation method, the red ones from the *fit* method, the blue ones from real data and the black ones from simulations.

3.10.3 Neutron detection efficiency

Plastic scintillators have a neutron detection efficiency of around 1% per cm of matter. The CND has a radial width of 9 cm. From these simple considerations we expect a detection efficiency of around 10% for neutrons. Early simulations studies showed that, for all detection angles, one can expect such an efficiency (see Figure 3.28).

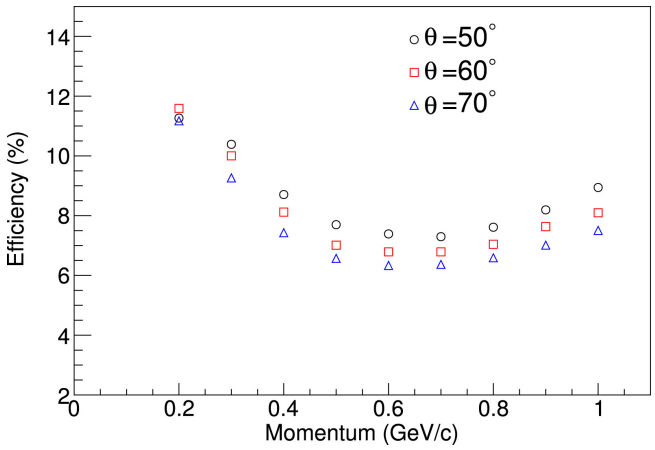


Figure 3.28: Estimation of the CND neutron efficiency performed on simulated events with a single neutron. The neutron efficiency is expected to be approximately 10%. Figure taken from [74].

The exclusive reaction $ep \rightarrow e'n\pi^+$ is analyzed to evaluate the neutron detection efficiency of the CND. The dataset used was taken with a 7.5-GeV electron beam incident on a liquid-hydrogen target. The events with an electron and a π^+ in the CLAS12 FD are selected. The missing mass of the $e'\pi^+X$

system is plotted versus β_X in Figure 3.29, and one can clearly identify a peak at the neutron mass. The missing particle is required to be in the CLAS12 Central Detector ($\theta > 40°$). The effect of this selection is shown in Figure 3.29, where the high-mass component of the spectrum is removed after the cut. We apply an additional cut on β of the missing neutron ($0.2 < \beta_X < 0.8$) and on the missing mass (0.7 GeV $< M_X < 1$ GeV) to ensure the exclusivity of the final state. From this set of $ep \to e'(n)\pi^+$ events, those with a neutron identified by the CND (i.e. with a CND cluster with $E_{dep} > 2.5$ MeV, no associated CVT tracks and $\beta < 0.8$) are selected. If multiple neutron candidates are detected by the CND, the neutron with the smallest azimuthal separation from the missing neutron is kept. A cut on $\beta > 0.1$ is applied to remove out-of-time hits that can be mistaken as neutrons. Finally, the detected neutron and the missing neutron azimuthal angle difference is constrained to be less than 20°.

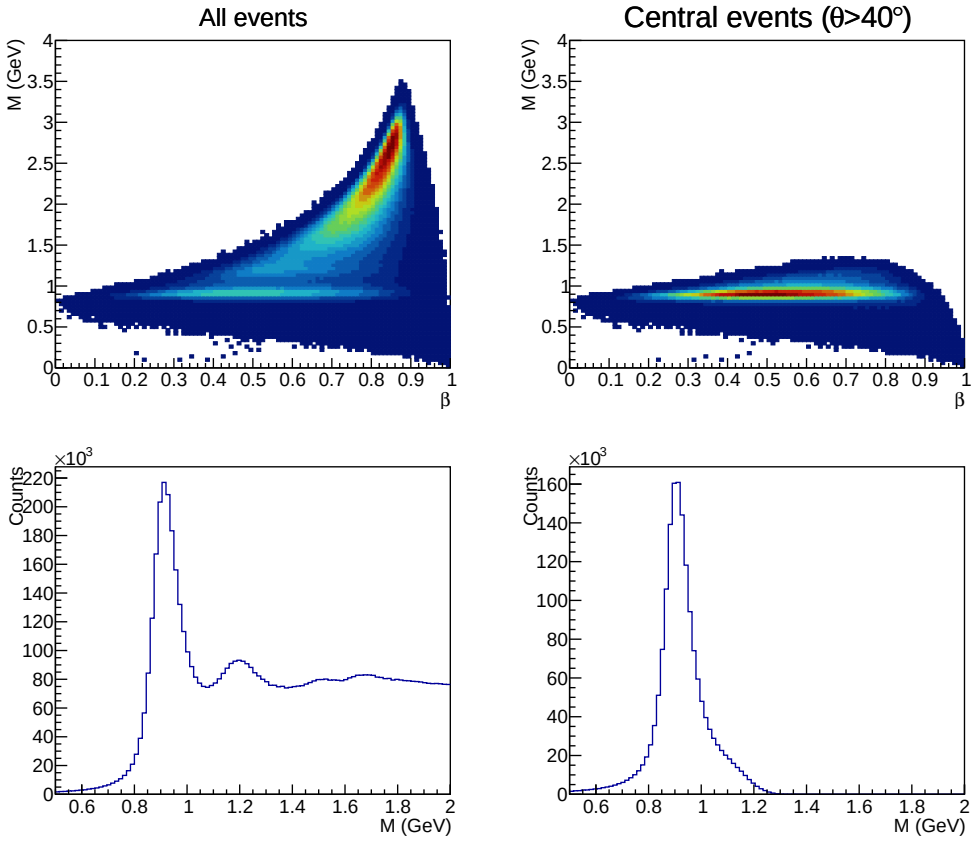
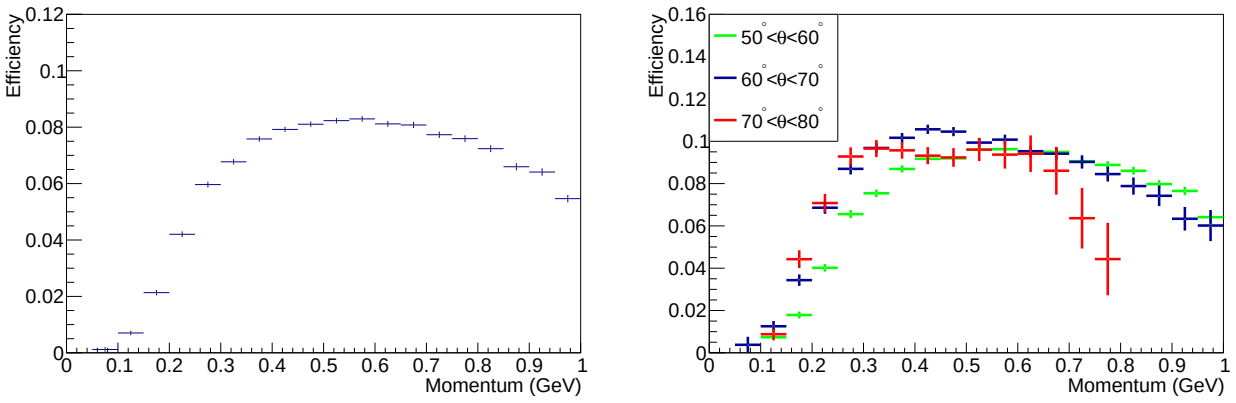


Figure 3.29: Selection of the exclusive neutron peak for the $ep \to e(n)\pi^-$ reaction. The top two plots show the missing mass versus β of the missing neutron for all the events (left) and for the events where the missing neutron is in the central detector ($\theta > 40°$) (right). The bottom two plots show the effect of the cut on the mass spectrum of the missing particle by comparing the spectrum before (left) and after (right) the cut. After applying the CD cut only the neutron peak remains.

The efficiency is measured in bins of the missing neutron polar angle and as a function of its missing momentum. For each bin in polar angle and momentum, the efficiency is defined as the ratio of events with a detected neutron to the number of missing neutron events. The result is shown as a function of the missing neutron momentum in Figure 3.30 for bins in polar angle and integrated over all azymuthal angles. The measured efficiency ($\sim 10\%$) is in agreement with the one expected during the R&D development of the CND. From this study, an estimate of the momentum resolution of the detected neutron is calculated. We find a resolution of $\frac{\Delta P}{P_{Mis.}} = 16\%$, where $\Delta P = P_{det.} - P_{Mis.}$ is the difference between the detected and the missing neutron momenta, as shown in Figure 3.31. This value is higher than the 10% value from the R&D specifications. However it includes effects coming from the detection of the electron and pion, especially FD-CD mis-alignments.

The same study is performed using Monte Carlo simulations. nDVCS events were processed in the full simulation and reconstruction chain of CLAS12. In this case, the efficiency is calculated as



(a) (b)

Figure 3.30: (a) Integrated neutron detection efficiency and (b) efficiency for three polar angle θ bins measured from data.

the number of reconstructed neutrons divided by the number of generated neutrons. The obtained efficiencies are compared in Figure 3.32 for different bins in polar angle and in Figure 3.33 integrated over all polar angles. The results for the efficiency from simulations and from data have consistent values. The discrepancies, observed at high momenta (above 0.8 GeV) and low momenta (below 0.3 GeV), are due to the two different final states used and the fact that the expected neutron kinematics are perfectly known in the case of the simulations.

3.11 Premilinary results for nDVCS based on neutron detection in the CND

To conclude this chapter, very preliminary results for the nDVCS reaction, obtained by M. Hoballah and S. Niccolai are presented. The data shown in the following were taken with an electron beam impinging on a deuterium target. Events with a photon in the FT and a neutron in the CD are selected. Considering that the neutron is quasi-free inside the deuterimum nucleus, the reaction selected from this topology is:

$$en \rightarrow e'n'\gamma X. \tag{3.57}$$

The missing mass M_X is shown in Figure 3.34. Exclusivity cuts are then applied to ensure that no other particle is produced in the reaction. The missing mass spectrum after exclusivity cuts is superimposed in Figure 3.34.

Once the reaction is identified, the raw beam spin asymmetry is calculated. The integrated BSA over the whole phase space available is shown as a function of Φ (see Section 1.3 for the definition). The BSA is fitted with the function $f(\Phi; a, b)$:

$$f(\Phi; a, b) = \frac{a\ \sin(\Phi)}{1 + b\ \cos(\Phi)}, \tag{3.58}$$

where the parameters a and b can then be related to CFFs. Figure 3.35 shows the raw beam spin asymmetry and the results of the fit. Note that the result shown here is not corrected for the main known background, coming from π^0 electro-production, where one photon stays undetected:

$$en \rightarrow e'n'\pi^0 \rightarrow e'n'\gamma(\gamma). \tag{3.59}$$

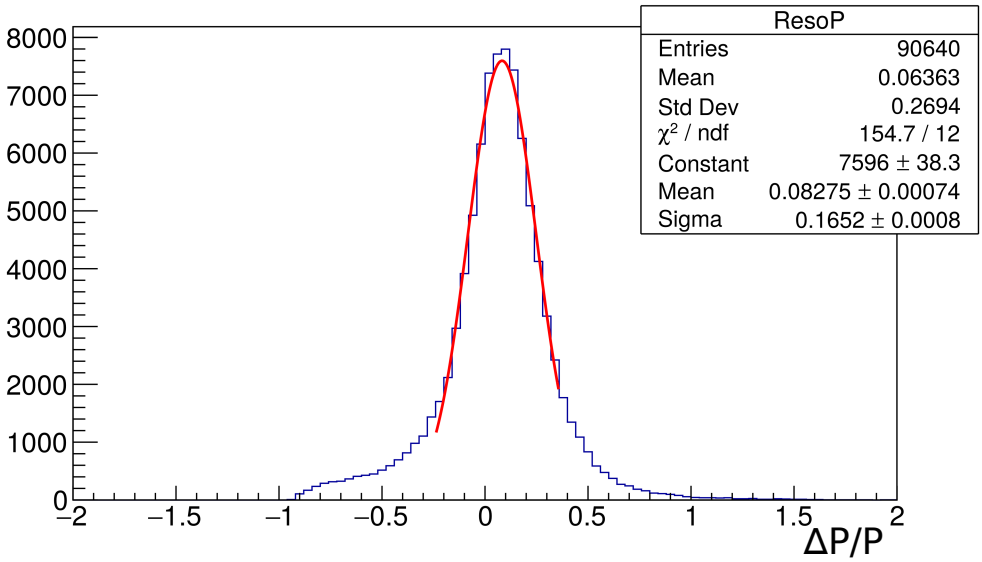


Figure 3.31: Neutron momentum resolution measured from data. The obtained value is higher than the design value. Discrepancies are explained by FD/CD misalignements inducing shifts in the missing neutron momentum.

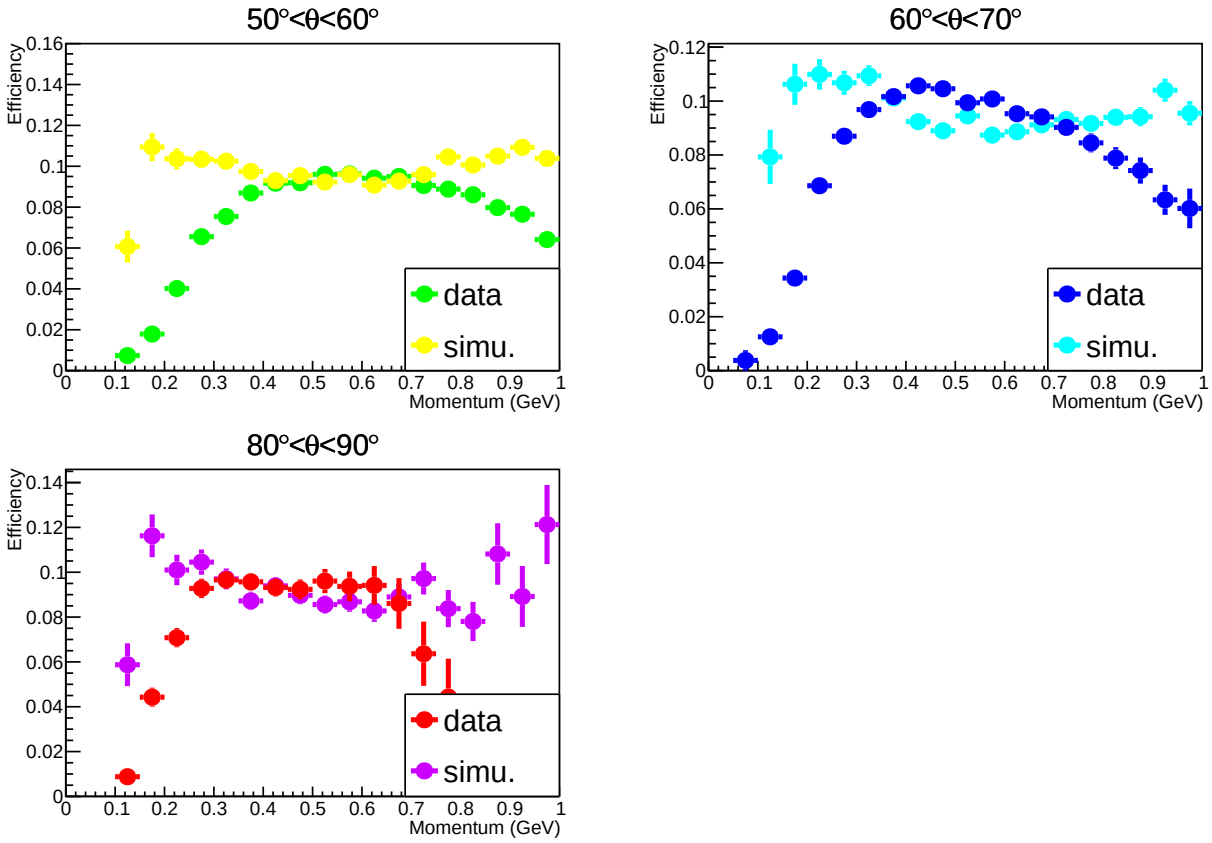


Figure 3.32: Comparison of the neutron detection efficiency for three polar angle θ bins measured from data and simulations.

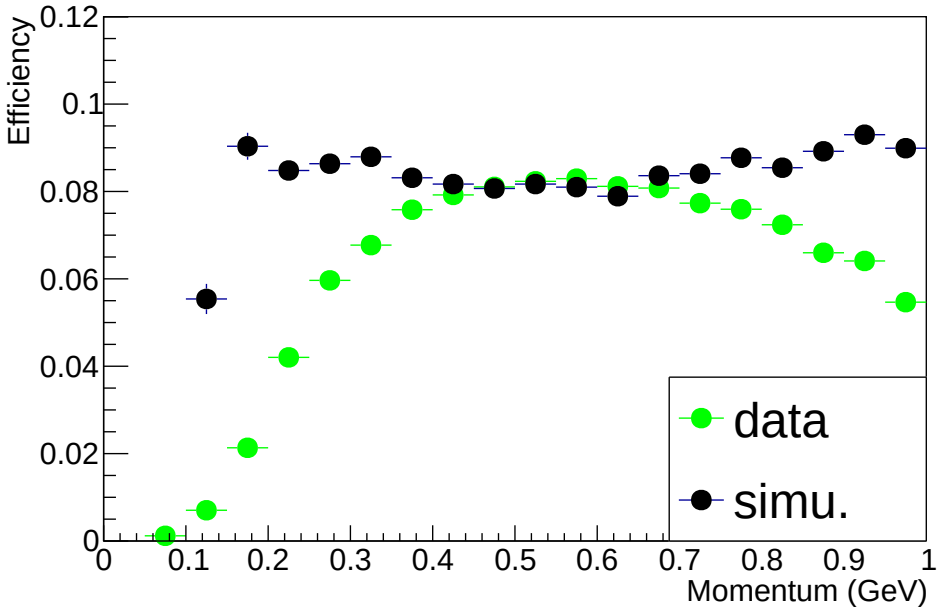


Figure 3.33: Comparison of the integrated neutron detection efficiency measured from data and simulations.

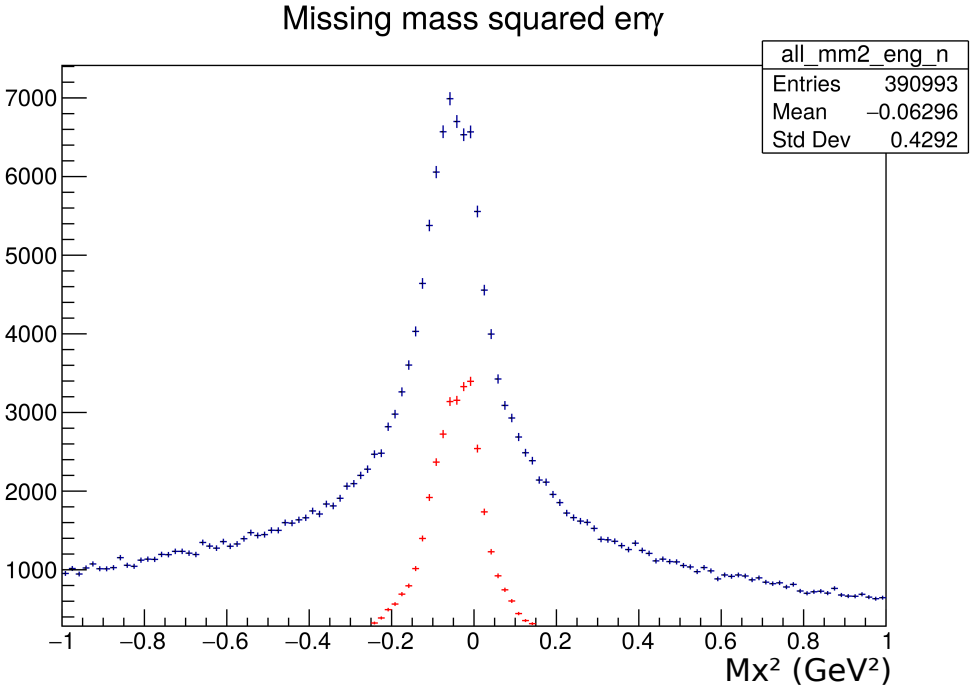


Figure 3.34: Missing mass of the $e'n'\gamma X$ before (blue) and after (red) exclusivity cuts are applied. Plot provided by M. Hoballah and S. Niccolai.

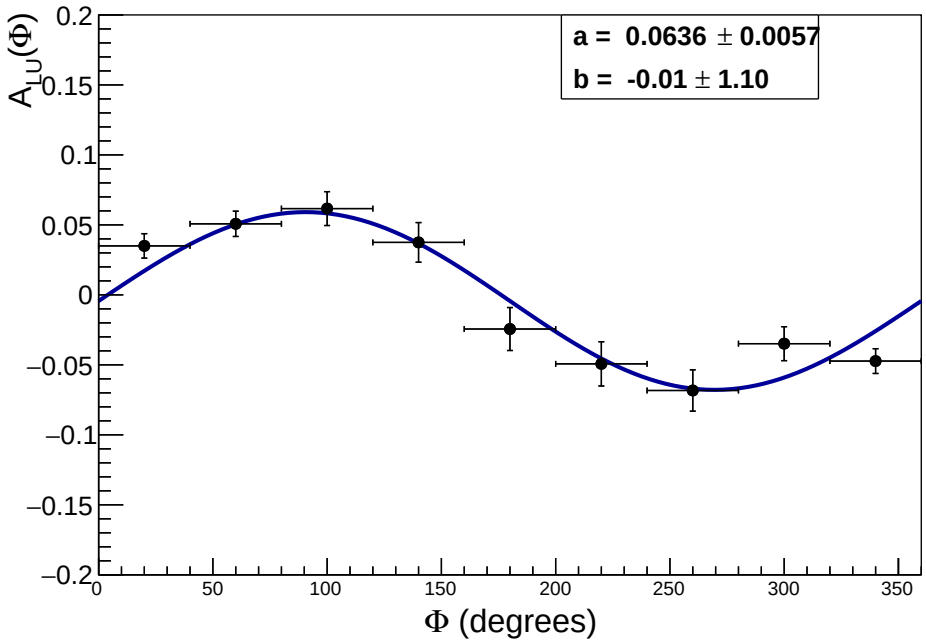


Figure 3.35: Fitted nDVCS BSA as a function of Φ. Plot provided by M. Hoballah and S. Niccolai.

Timelike Compton Scattering data analysis

This part of the manuscript groups three chapters. The work presented in the following is dedicated to the measurement of Timelike Compton Scattering observables from the CLAS12 RGA dataset.

The first step of this measurement is to retrieve events recorded by CLAS12 where the $\gamma p \to p' e^+ e^-$ reaction occurs. In order to select these events, the final-state particles must be identified. This step is presented in Chapter 4. The event builder algorithm of CLAS12 is presented. This algorithm associates the various responses of the CLAS12 sub-detectors for each event, to recover the momenta and the vertices of all the recorded particles. An enhanced positron identification algorithm, crucial in the TCS measurement, is also presented in this chapter.

The second step of the analysis consists in correcting the data. These corrections aim at matching the measured momenta of the detected particles with their actual momenta at the vertex. This step is also presented in Chapter 4.

Once the final state particles are well identified and their momenta are corrected, exclusivity cuts are applied in order to make sure that each event is a TCS event (i.e. from the $\gamma p \to p' e^+ e^-$ reaction). The observables are then computed with the kinematic variables of these good events. This step is performed in Chapter 5.

Finally in Chapter 6 the results of the full analysis of the CLAS12 data are displayed and discussed against theoretical predictions.

Chapter 4

Particle identification and momentum corrections

Particle Identification (PID) is the last step of the data reconstruction. It aims at producing a list of particles, their associated momentum and vertex coordinates, as well as the associated list of responses in CLAS12 detectors. This step is performed by the Event Builder (EB) described in [82]. Tracks are associated to detector responses to form particles. Each particle is then identified according to this list of detector response. In this chapter, the identification procedures for protons (Section 4.1) and leptons (Section 4.2) are presented first. In Section 4.3 an enhanced positron identification algorithm developed for the TCS analysis is discussed.

This chapter also describes momentum corrections developed to correct CLAS12 data and simulations. These corrections have been developed in the perspective of the TCS analysis but can be used in any CLAS12 analysis. A full set of momentum corrections for protons was put in place. Subsection 4.4.1 presents the Monte-Carlo based corrections, and Subsection 4.4.2 the data-driven corrections in the CD. Momentum corrections for leptons are also presented in Section 4.5. Finally, fiducial cuts for leptons are presented in Section 4.6.

4.1 Proton identification

Protons, and more generally heavy hadrons, are identified with their time of flight (tof) from the vertex to their interaction point with Time Of Flight (TOF) detectors. There are two TOF detectors in CLAS12, the CTOF in the CD and the FTOF in the FD. In both cases the proton identification procedure is similar. A precise time reference, the start time, is determined using a fast moving particle detected in the FD. Then the tof is calculated and matched to the expected one obtained from tracking.

Start time

The start time is determined using the FTOF response of the trigger particle of the event. The trigger particle is defined as the most energetic lepton in the EC or, if no lepton is detected, the pion with the higher momentum. The uncorrected vertex time t_v of the trigger particle is calculated as:

$$t_v = t - \frac{P_L}{c}, \tag{4.1}$$

where t is the measured time in the FTOF of the trigger particle and P_L is the path length from the vertex to the detection point in the FTOF. The vertex time is then corrected by the position of the vertex z_v (to account for the time that the beam bunch is taking to propagate to the vertex from the origin of CLAS12, defined as the center of the target z_0), and the RF time provided by the accelerator t_{RF}, as:

$$\Delta t_{RF} = t_v + \frac{(z_0 - z_v)}{c} - t_{RF} - \left(N + \frac{1}{2}\right)\frac{1}{f_{RF}}, \tag{4.2}$$

where f_{RF} is the frequency of the accelerator and N is a large integer (typically 800). This allows to find from which beam bunch the event originated and to precisely match the vertex time in the time window between two bunches, using:

$$\Delta t'_{RF} = \mathrm{mod}(\Delta t_{RF}, \frac{1}{f^{RF}}) - \frac{1}{2f_{RF}}. \tag{4.3}$$

Finally the start time is defined as:

$$t_S = t_v - \Delta t'_{RF}. \tag{4.4}$$

Momentum/tof matching

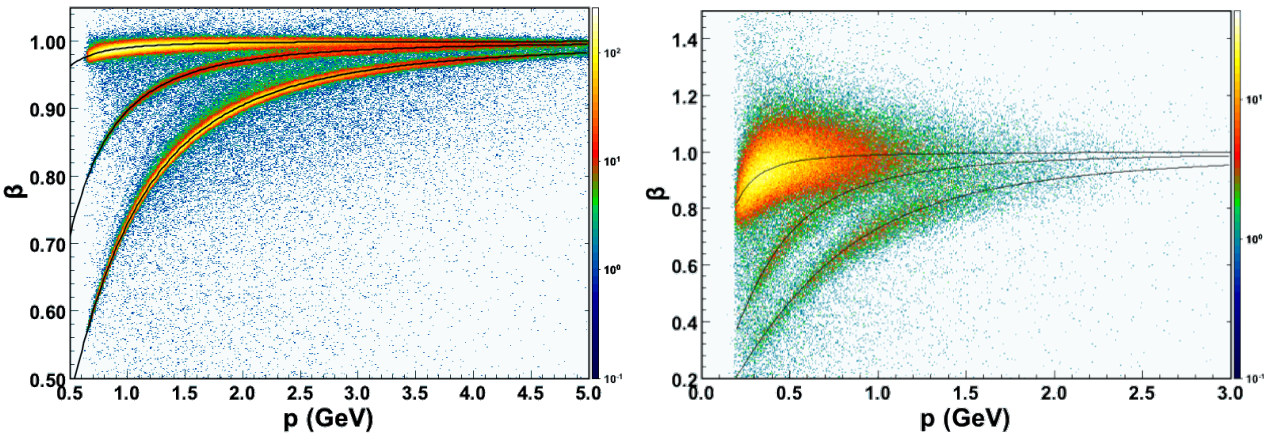
The tof t_{tof} of a hadron is given by:

$$t_{tof} = t_{TOF} - t_S, \tag{4.5}$$

where t_{TOF} is the time associated with the detector in which the hadron was detected. The expected tof from tracking t_{track} is given by:

$$t_{track} = \frac{P_L\sqrt{p^2+m^2}}{pc}, \tag{4.6}$$

where m is the mass of the particle and p its momentum from tracking. The mass hypothesis which minimizes the difference $t_{tof} - t_{track}$ is assigned to the particle. Figure 4.1 shows the velocity β versus momentum distributions in the FD and the CD.



(a) (b)

Figure 4.1: β versus momentum distribution for positively charged particles in the FD (a) (resp. in the CD (b)). The black lines correspond to three mass hypotheses (top: pion, middle: kaon, bottom: proton).

4.2 Lepton identification

Electrons and positrons have very low mass and will likely be detected in the FD of CLAS12. The timing resolution of the FTOF doesn't allow to separate leptons and pions at the CLAS12 kinematics. As a consequence, the pid of leptons is solely based on the Sampling Fraction (SF) of the EC and the number of photo-electrons in the HTCC. The SF is defined as:

$$SF = \frac{E_{dep}}{P}, \tag{4.7}$$

where E_{dep} is the total energy deposited in the EC, and P the momentum measured by the DCs. The EB of CLAS12 assigns the particle ID, in the Lund convention, ± 11 (i.e. electron or positron) for particles that fulfill all the following requirements:

- A track in the DCs and an associated EC shower (the curvature of the track in the torus magnetic field gives the charge of the particle)
- A minimum deposited energy in the PCAL: $E_{PCAL} > 60$ MeV

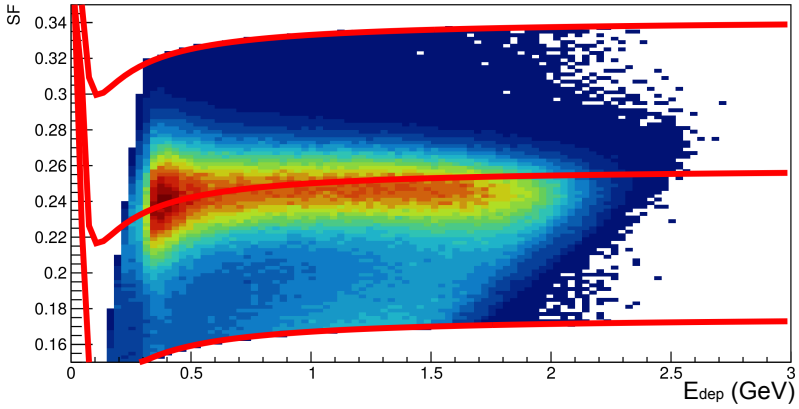


Figure 4.2: SF versus E_{dep} for electrons in the EC, from data. The three red curves represent the parametrized SF_M extracted from simulations, and its 5-σ limits. In the final data analysis, these three functions are calibrated for each sector.

- A total measured SF, $SF_M(E_{dep})$, within 5 σ of the parametrized SF, $SF_P(E_{dep})$: $\mid SF_M(E_{dep}) - SF_P(E_{dep}) \mid < 5\sigma_P$, as shown in Figure 4.2. The parametrized SF is defined by:

$$SF_P(E_{dep}) = a\left(b + \frac{c}{E_{dep}} + \frac{d}{E_{dep}^2}\right). \tag{4.8}$$

  The parameters are calibrated for each sector and depend on the run range.

- If $P < 4.9$ GeV, a minimum number of HTCC photo-electrons: $N_{PHE}(HTCC) > 2$. The distribution of the number of photo-electrons in the HTCC for detected electrons is shown in Figure 4.3.

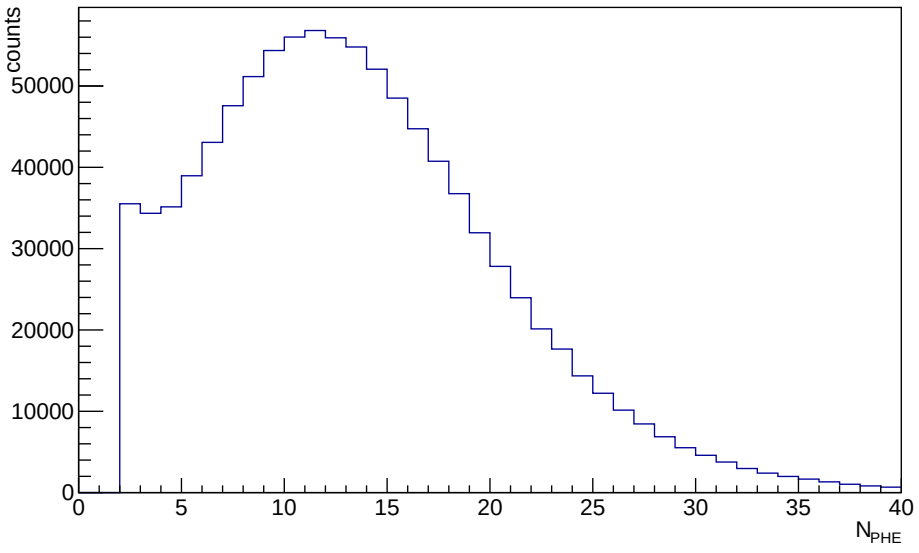


Figure 4.3: Number of photo-electrons for electrons, from RG-A data.

There are two distinct regions for lepton identification: below 4.9 GeV leptons are identified using the HTCC and the EC. Above this threshold, the HTCC produces a signal for pions and leptons and thus cannot help in separating them. In the high-momenta region, the lepton pid is then only based on the EC. An additional pid procedure to enhance the positron purity in this high momentum region is presented in Section 4.3.

4.3 Positron identification

Measurements of TCS observables require the detection of a pair of leptons from the decay of a virtual photon. For momenta lower than 4.9 GeV, the HTCC of CLAS12 provides good pion/lepton separation [93]. The HTCC electron detection efficiency is estimated to be higher than 99%. For momenta higher than the HTCC threshold (4.9 GeV) data and simulation show a large contamination of π^+ in positron samples. In this section, evidence for pion contamination is shown and a multivariate analysis is proposed to reduce this contamination. The results of this approach are shown and assessed. An estimate of the remaining pion contamination is given.

In the following we consider particles that have been assigned ID -11 and identified as positrons by the EB. The electron case is treated at the end of this section.

4.3.1 Evidence of π^+ contamination

In order to measure TCS observables, we aim at achieving a clean identification of leptons. The positron ID is crucial to reach this goal. However the pid cuts described in Section 4.2 are not sufficient to remove potential mis-identified π^+ from the positron sample. This contamination can be seen in both the data and simulations.

We use the expression "true-positron" for Generated positron-reconstructed positron and "mis-id. pion" for Generated pion-reconstructed positron in the simulation samples. For the data samples, analogous designations are used, replacing "Generated" by "Produced". Finally, we also refer to these categories as respectively "Signal" and "Background" when methods to distinguish both sets are described.

π^+ contamination from data

Positron momenta spectrum A first evidence of pion contamination is seen by investigating events with an electron, a proton, a positron and any other particles in CLAS12. One can see that there is an excess of positrons above the HTCC threshold. This is visible in Figure 4.4 where the polar angle of the positrons are plotted against their momenta.

Exclusive reaction A second evidence of pion contamination is seen when investigating the exclusive reaction $ep \to e\pi^+ n$. Events with at least a particle with ID -11 and momentum bigger than 4.4 GeV and an electron with momentum lower than 4.4 GeV were filtered from the CLAS12 dataset. Both particles are required to be detected in the FD. Cuts on the electron momentum ensure that it is a true electron. The particles with ID -11 are assigned the π mass. The missing mass of the system $ep \to ee^+_{m_\pi} X$ is then calculated. The notation $e^+_{m_\pi}$ is used to refer to the particle with ID -11 and mass equal to the π mass. The missing mass spectrum obtained is shown in Figure 4.5.

A clear peak at the neutron mass is visible. This peak is produced by the reaction $ep \to e\pi^+ n$ were the π^+ has been identified as a positron. These exclusive mis-identified pion events are a good way to quantify the pion contamination. Furthermore the momenta of the mis-id. π^+ cover most of the momentum range of interest, from 4.9 GeV to 10.6 GeV as shown in Figure 4.6. We will use these events later in Subsection 4.3.5 as a scale to quantify the pion background.

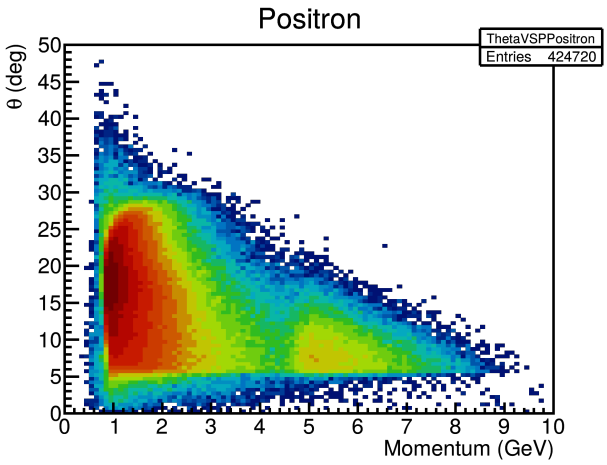


Figure 4.4: Polar angle θ versus momentum of positrons for events with one electron, one proton and one positron in CLAS12.

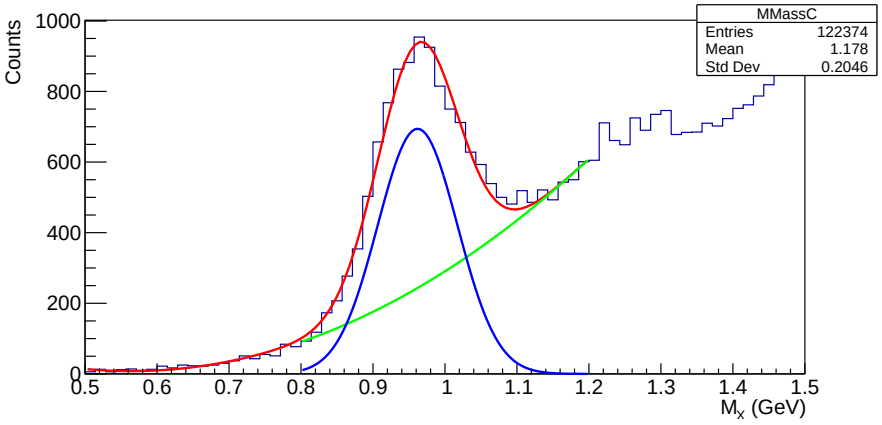


Figure 4.5: Missing mass spectrum of the reaction $ep \to ee^+_{m_\pi}X$ in the neutron mass region.

π^+ contamination from simulations

The π^+ contamination is also visible in simulations. Two test samples were generated, one sample of positrons and one of positively charged pions. They were passed to the GEMC CLAS12 simulation and reconstructed using the standard CLAS12 software. Particles were simulated within the CLAS12 acceptance and within the range of momenta of interest (4 GeV $< P <$ 10.6 GeV and transverse momentum within 0.5 GeV $< P_t <$ 2 GeV). The output of both samples were then skimmed identically: only particles with ID -11 were kept. The kinematics distributions for both true-positrons and mis-id. pions are shown in Figure 4.7.

The behavior observed in the data (excess of positrons due to contamination from π^+) is reproduced in the simulations.

4.3.2 1D and 2D cuts from the simulations

The results of the simulations described above were used to explore simple cuts to try to remove the pion contamination. In this section, various pid cuts based on these simulations are defined.

χ^2 cut

The CLAS12 EB gives the deviation of the total sampling fraction from a parametrized model. This quantity is referred as χ^2 in the CLAS12 software (although the name commonly used in literature is

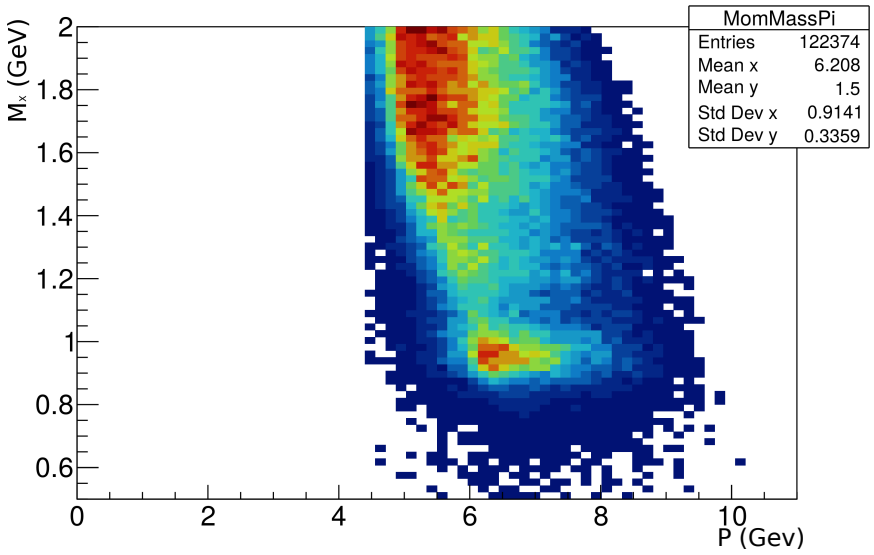


Figure 4.6: Mass of the missing particle versus momentum of the particle with ID -11 .

pull value) and it is defined as:

$$\chi^2 = \frac{SF_M(E_{dep}) - SF_P(E_{dep})}{\sigma_P}. \tag{4.9}$$

The EB requires that $-5 < \chi^2 < 5$ to identify a particle as a lepton. The χ^2 of true-positrons and mis-id. pions is shown in Figure 4.8. One can see that mis-id. pions mostly populate the low χ^2 region. Their sampling fraction is just big enough for them to be identified as positrons. From these distributions, two simple strategies can be tested: cutting on the absolute value of χ^2 ($\mid \chi^2 \mid < c$) or cutting on low values of χ^2 ($c < \chi^2$). These strategies are referred as Symmetric and Asymmetric χ^2 cuts in the following.

2D Sampling Fraction cuts

Another way to distinguish positrons from π^+ is to look at their partial SFs. The total sampling fraction can be decomposed according to the three layer of the EC (PCAL, ECIN and ECOUT):

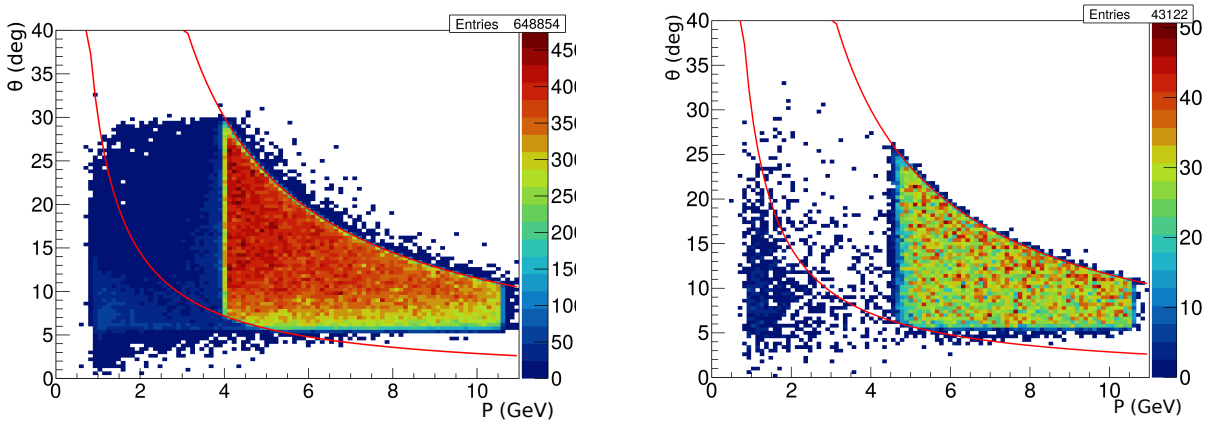
$$SF_{PCAL} = \frac{E_{dep}(PCAL)}{P}. \tag{4.10}$$

Analogous equations can be written for ECIN and ECOUT. The longitudinal segmentation of the EC proves useful to distinguish positrons. Positrons are more likely to deposit all their energy in the first layers of the EC (PCAL and ECIN). On the contrary π^+ are Minimum Ionizing Particles (MIPs) and are more likely to deposit energy in all the layers of the EC. Figure 4.9 shows the SF in the ECIN versus the SF in the PCAL for simulated true-positrons and mis-id. pions. One can see that a 2D cut along the anti-diagonal of the distribution can be applied to separate them. In the following this cut is referred as "SF cut".

4.3.3 Multivariate analysis approaches

In the previous section, simple approaches involving cuts on 1 or 2 quantities were shown. However these approaches do not allow to use the full amount of information provided by the EC. Additionally to the SFs of all three EC layers, one can access the shape of the electromagnetic shower in each layer. The square of the width of the shower is defined for each coordinate (U,V,W) and for each layer as:

$$M_2 = \frac{\sum_{\text{strip}}(x - D)^2 \ln(E)}{\sum_{\text{strip}} ln(E)}, \tag{4.11}$$



(a) (b)

Figure 4.7: (a) Polar angle θ versus momentum of the simulated true-positrons (b) Polar angle θ versus momentum of the simulated mis-id. pions. The red lines represent the transverse momentum limits (0.5 GeV $< P_t <$ 2 GeV) applied on the generated particles. The positrons detected at low momenta in Figure (a) are positrons which radiated a photon in the target material. The momenta of these positrons are corrected when the radiated photons are detected (see Section 4.5.2).

where D is the log-weighted mean position of the shower defined as:

$$D = \frac{\sum_{\text{strip}} x \ \ln(E)}{\sum_{\text{strip}} ln(E)}, \tag{4.12}$$

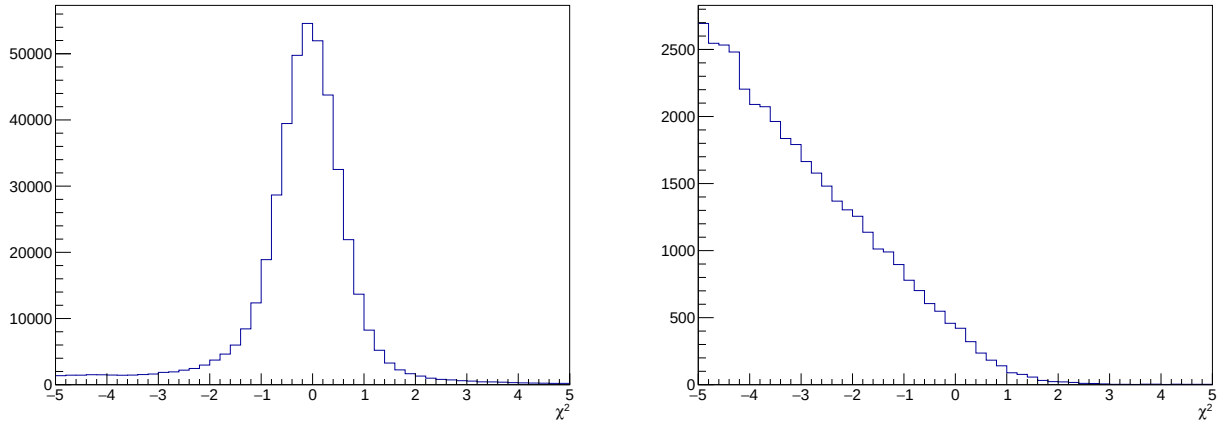
and where x is the position of the EC hit along the considered coordinate and E is the deposited energy associated to the hit. There are potentially at least 12 variables (3 sampling fractions and 9 shower widths) to investigate to help separating π^+ and e^+. A simple approach relying only on assessing correlations between each couple of variables one by one is not applicable here. A multivariate approach is needed. In this section some multivariate techniques are introduced. Their application to the π^+/e^+ separation problem is presented.

The TMVA Root package [94] was used for this analysis. This package has been developed to train, test and compare a large range of multivariate tools. Multivariate Analysis (MVA) classifiers take several quantities as inputs and produce a single output on which one can then apply a cut. The value of the cut that maximizes the background rejection and the signal efficiency is then found and applied to the output distribution. In the following we only focus on three of them: Fisher discriminant (see Appendix B for more details), Boosted Decision Tree (BDT) (in Appendix C) and Multilayer Perceptron (MLP) (in Appendix D).

The MVA classifiers presented in the following were trained on the simulation samples described in Subsection 4.3.1.

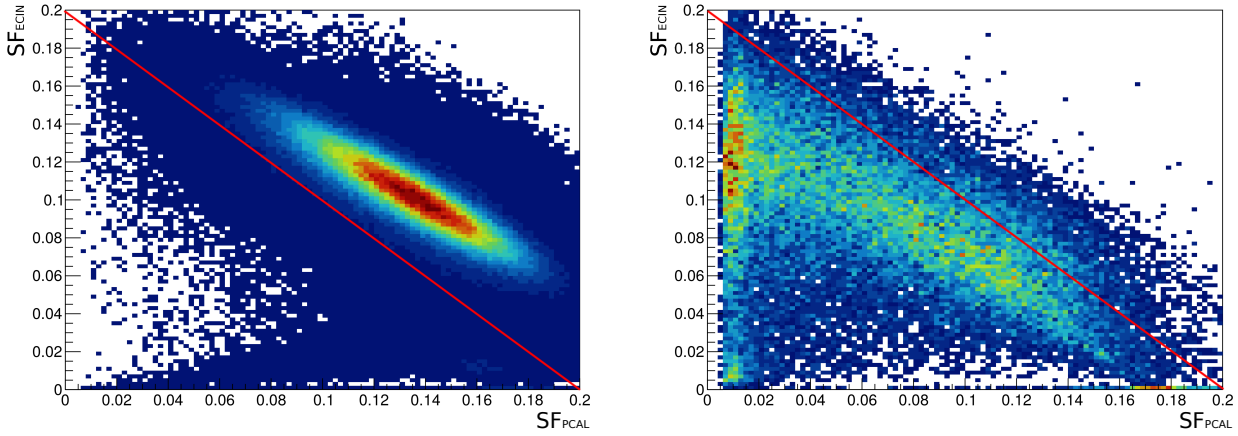
3D analysis

As a first step, the SFs of the three layers of the EC were considered as inputs. The distributions of these three input variables for true-positron (blue) and mis-id. pion (red) are shown in Figure 4.10. Three methods (Fisher, BDT, and MLP) were trained and tested on these three variables. This analysis is referred as 3D in the following. The structure of the MLP neuron network for this analysis is shown in Figure D.3 in the appendices.



(a) (b)

Figure 4.8: Sampling fraction χ^2 of (a) true-positrons and (b) mis-id. pions.



(a) (b)

Figure 4.9: SF in the ECIN versus SF in the PCAL for simulated (a) true-positrons and (b) mis-id. pions. A possible cut to remove mis-id. pions is to keep only particles above the anti-diagonal represented by the red lines. Most mis-id. pions are removed while most positrons are kept.

6D analysis

A 6D multivariate analysis was also studied. In this approach Fisher, BDT and MLP methods were applied to the three SFs and to the average width of the shower in each layer. The average of the square of the width in the PCAL is defined as:

$$M_{2/PCAL} = \frac{M_{2/U/PCAL} + M_{2/V/PCAL} + M_{2/W/PCAL}}{3} \tag{4.13}$$

where $M_{2/U/PCAL}$ is the square of the width of the shower along the U direction as defined in Equation (4.11). Similar equations apply for the V and W directions. The distribution of the input squared shower widths are shown in Figure 4.11. Figure D.4 in Appendix D shows the architecture of the neural network used for the 6D analysis.

4.3.4 Training, testing and comparison of MVA classifiers on simulations

Fisher discriminant, BDT and MLP are trained on the sets of variables presented in the previous two subsections using the TMVA package. A first series of tests is also performed on simulations. The

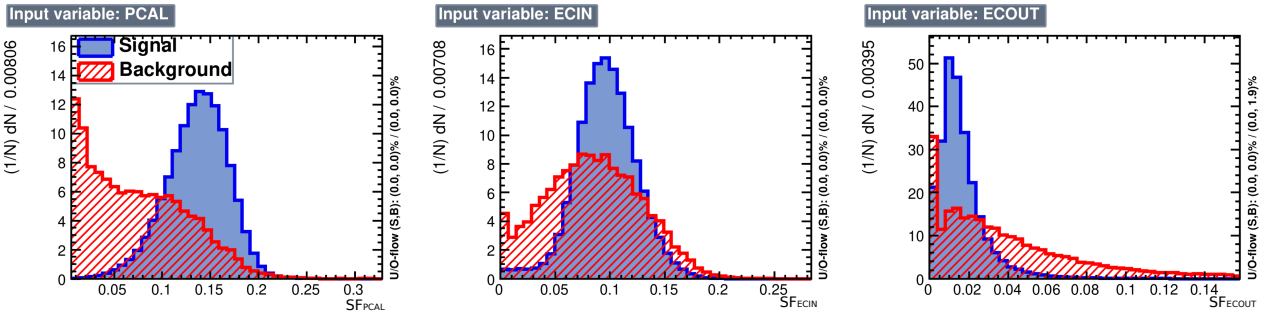


Figure 4.10: Input variables provided for the training of the multivariate classifiers. The leftmost plot shows the SF_{PCAL} distributions for the signal and the background. The middle plot SF_{ECIN}, the rightmost SF_{ECOUT}.

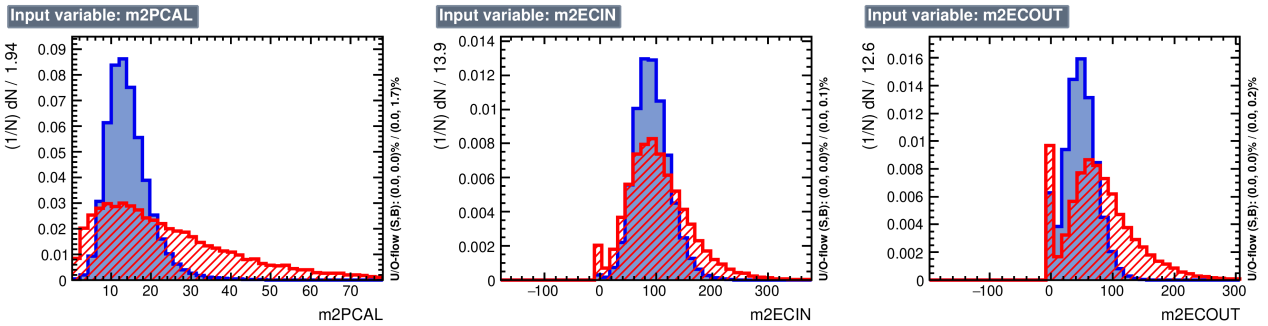


Figure 4.11: Average squared shower width for the PCAL (left), ECIN (middle), and ECOUT (right).

complete input sample is divided in two randomly selected subsets: a training and a testing sample. For each classifier, the training is performed on the training sample. The classifiers are then tested on the test sample. The distributions of classifiers outputs for the testing set are superimposed on the training distributions and presented in Appendix E. An example is shown in Figure 4.12. These checks indicate that it is possible to apply reliably these trained classifiers on simulations. Training and testing outputs show good agreement for both 3D and 6D approaches and for all the classifiers. One can also look at indicators specific to each technique in order to assess the quality of the classifiers. An example is shown in Figure 4.13 where the convergence of the MLP with three input variables is shown.

One can then compare the strength of different classifiers using Receiver Operating Characteristic (ROC) curves. ROC curves display the signal efficiency (fraction of signal kept) of a classifier versus its background rejection (fraction of background removed). The method which achieves higher signal efficiency and higher background rejection is preferred. The ROC curves for four 3D and three 6D classifiers are shown in Figure 4.14.

One can clearly see that 6D classifiers are more powerful than 3D ones. This behavior is expected as more information is taken into account by the classifiers. Indeed as shown in Appendix F, shower widths are not fully correlated with the deposited energy. All the three tested techniques seem to have similarly high efficiencies.

4.3.5 Test and comparison of MVA classifiers on data

Once the classifiers were trained and tested on simulations, they were used on CLAS12 real data and their performances were compared. In this section, the method used to assess the performance of the classifiers on CLAS12 data is presented.

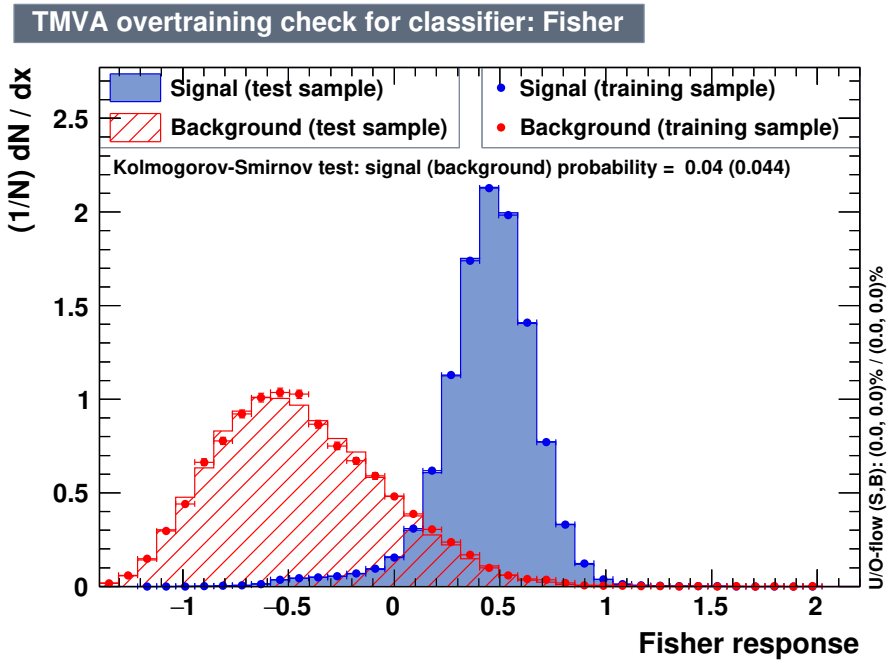


Figure 4.12: Training and testing outputs of the 3D Fisher classifier. Both distributions match very well for both the background and the signal.

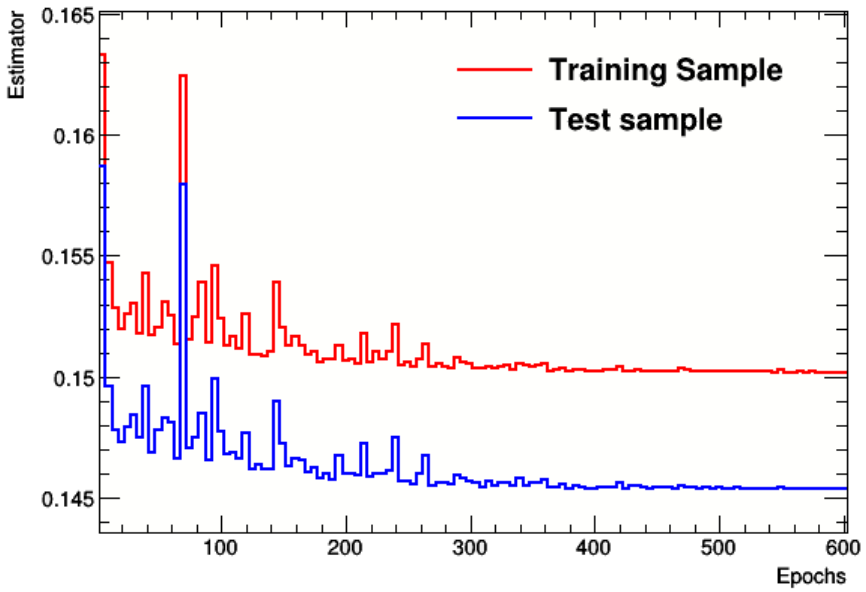
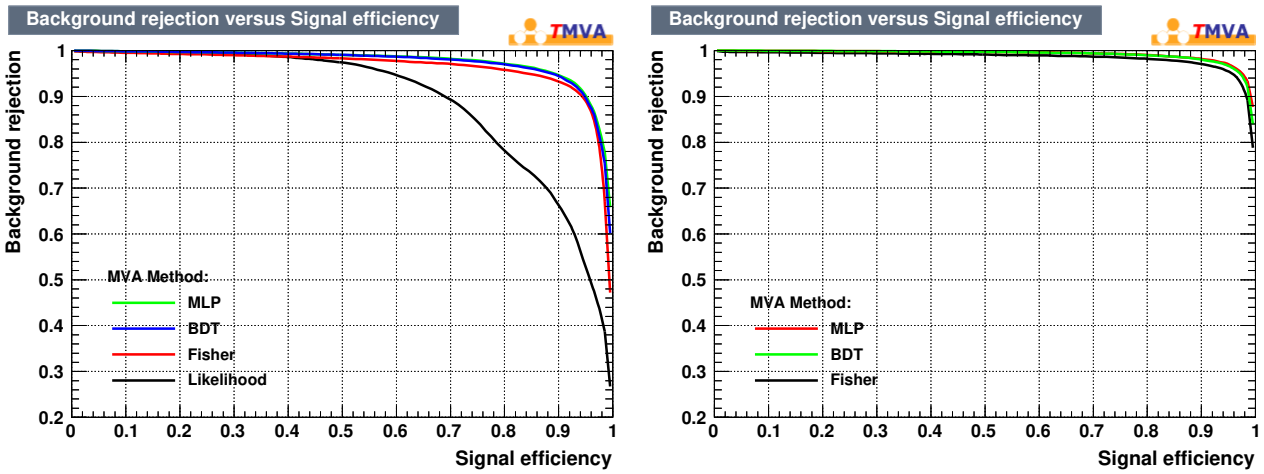


Figure 4.13: MLP 3D convergence test for the training and testing samples. The MLP error function defined in Appendix D and summed over all the events in the training (red) and testing (blue) sets, respectively, is shown as a function of the training iterations. One should check that the convergence is reached after a certain number of training cycles.

A benchmark to assess the π^+/e^+ separation on data

Clean signal and background samples are needed to evaluate the performance of classifiers. In simulation samples, this is easily achievable as the nature of each simulated particle is known. In data, a prior knowledge is necessary. The background sample (mis-id. pion) is described in Subsection 4.3.1. The neutron peak events presented in Figure 4.5 were used. As shown in this figure, the missing mass



(a) (b)

Figure 4.14: ROC curves for four classifiers (a) in the 3D case and (b) in the 6D case. In the 3D case, a Likelihood classifier was also tested, but not kept for later analysis due to its poor performance. The ROC curves of the 6D classifiers are consistently above 3D ones.

spectrum is fitted with a Gaussian for the peak and with a second order polynomial for the underlying background. The integral of the Gaussian gives a scale of the number of mis-id. pions in the sample. In the following this sample is denoted "neutron sample".

On the contrary, defining a clean signal sample (true-positron) from data is more difficult. TCS events (at least one electron, one positron and one proton; missing transverse momentum fraction $< 5\%$, and missing mass squared < 0.4 GeV^2, ensuring photo-production) with a positron with momentum bigger than 4 GeV were used and are referred as the "TCS sample" (see Chapter 5). This sample is not completely clean, it should a priori be a mixture of true-positrons and mis-id. pions. However the requirement of the e^-e^+p final state as well as the exclusivity cuts should enhance the true-positron fraction. This hypothesis is tested in the following section. To evaluate the "Signal+Background" in the TCS sample, the number of events is counted.

Once the two data test samples are defined, the different cuts presented in the previous sections are applied. For χ^2 and MVA approaches, one can vary the cut applied on the output variable and compare the number of remaining events in the neutron sample and in the TCS sample. This is realized by varying the cuts on the output variable in the specific output range and evaluating the integral of the neutron peak and the number of remaining TCS events. An example of the results of this procedure is shown in Figure 4.15, where the neutron peak integral and the number of TCS events are plotted against the value of the applied cut for the 3D BDT.

One can define a pseudo-ROC curve by associating these two curves. For each value of the cut, the number of TCS events is plotted against the corresponding number of neutrons. The curve obtained is not a ROC curve as the number of TCS events is a mixture of the signal and the background. However this pseudo-ROC curve still allows to compare the proposed π^+/e^+ separation strategies and evaluate their performances.

Strategy comparison and choice

The procedure presented in the previous section allows to draw for each identification technique a pseudo-ROC curve. All the pseudo-ROC curves are plotted in Figure 4.16. Both axes have been normalized to one. The blue circle denotes the case where no cuts are applied. The pink triangle represents the PCAL/ECIN SF anti-diagonal cut described in Subsection 4.3.2.

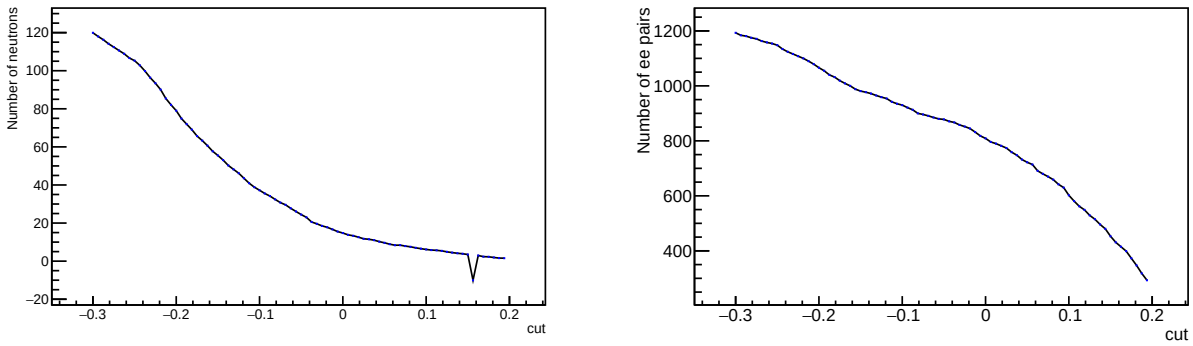


Figure 4.15: Number of neutrons as a function of the cut applied on the output of the 3D BDT (left). Number of TCS events as a function of the cut applied on the output of the 3D BDT (right).

The pseudo-ROC curves shown in Figure 4.16 (zoomed version in Appendix G) exhibit two regimes. At high background strength and large number of TCS events, the curves show a linear trend. This means that for a given number of mis-id. pions removed in the neutron sample, a proportional number of them is also removed from the TCS sample. On the contrary, when the background strength approaches zero, the number of TCS events decays dramatically. This happens because the cut applied on the output of the classifier is too hard and starts to remove true-positrons and mis-id. pions indistinctly. One should apply a cut on the classifier output that maximizes the background rejection while keeping most signal (at the limit between the two regimes). Figure 4.16 also permits to select the best approach to separate positrons from pions in CLAS12 data. One should use the method with the best pion rejection power while conserving most of the signal events. This is achieved by picking the method for which the pseudo-ROC curve gets closer to the (0,1) point. The MLP 6D classifier was chosen according to this criterion.

Adding skewness to the classifier

It is clear from Figure 4.16 that adding the width of the shower to the analysis increases the positron/pion separation efficiency. In order to test if adding more variables to the classifier training further increases its power, the 3rd moment of showers in the EC, the skewness μ, was added to train a MLP and a BDT. The skewness of a shower is defined as:

$$\mu = \frac{\sum_{\text{strip}} (x - D)^3 \ln(E)}{M_2^{3/2} \sum_{\text{strip}} \ln(E)}. \tag{4.14}$$

The pseudo-ROC curves for these two classifiers are compared to the 6D MLP and BDT in Figure 4.17. One can see that all curves superimpose. Adding the skewness to the inputs of the classifiers does not improve their separation power.

4.3.6 Remaining background estimation

The pseudo-ROC curves in Figure 4.16 show a linear behavior at high background strength. This behavior can be explained by considering that mis-id. pions in the TCS sample are removed at the same rate as mis-id. pions in the neutron sample when the cut is varied. True positrons might also be removed in the process and we can suppose this removal is small in the region where the background is important. If the classifier behave well, few signal events will be removed while most of the background will be cut away.

Let y be the variable describing the *normalized number of TCS events* and x describing the *normalized background strength* in Figure 4.16. In the region $0.1 < x < 1$ the pseudo-ROC curves in Figure

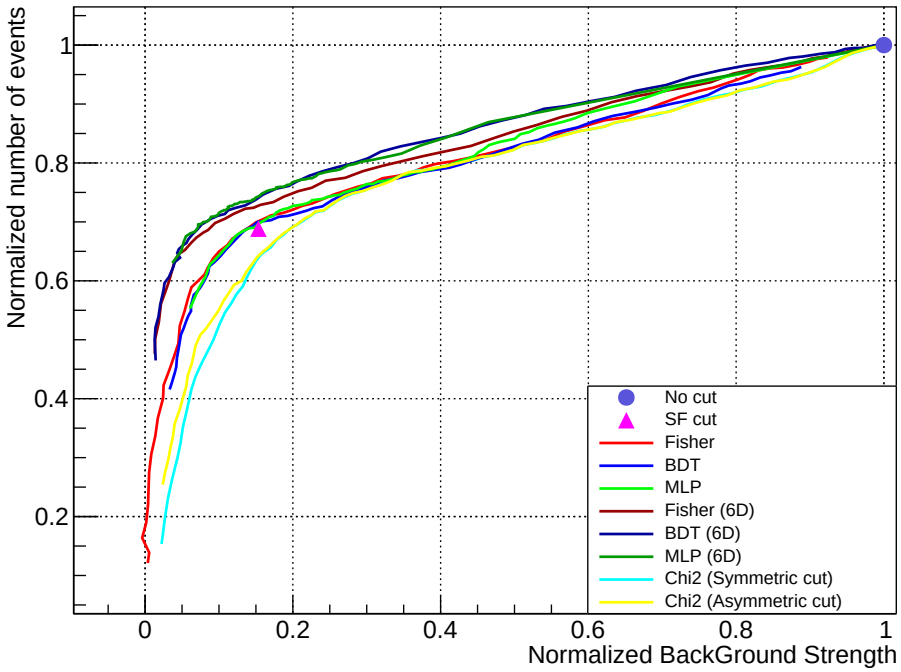


Figure 4.16: Pseudo-ROC curves obtained from data for different classifiers (3D classifiers are referred by their name only) and for the symmetric and asymmetric χ^2 cuts. The number of events in the TCS sample is plotted against the integral of the neutron peak in the neutron sample. Both axes are normalized to 1 when no cut is applied. The anti-diagonal SF cut is represented by the pink triangle. The blue dot represents that case were no additional cuts besides EB ones are applied. A zoomed version of this plot is displayed in Appendix G.

4.16 are linear. Let $B(x)$ and $S(x)$ respectively be the number of background (mis-id. pions) and signal events (true-positrons) in the TCS sample for the corresponding x background strength.

As stated before, the number of background events in the TCS sample is linear with the background strength x:

$$B(x) = \beta x, \tag{4.15}$$

where β is the number of background events in the TCS sample when no cut is applied. The function $S(x)$ is unknown but we assume it is increasing with x (when background is removed, signal events might also be removed by mistake), and does not vary much with x (signal events should not be removed by the classifier, the loss is estimated on simulations and is expected to be less than 1%, as shown in Figure 4.20).

The normalized number of TCS events can then be written in the linear region as:

$$y(x) = \frac{S(x) + B(x)}{S(1) + B(1)} = \frac{S(x) + x\beta}{S(1) + \beta}. \tag{4.16}$$

Although this formula is only applicable in the linear region, we can extrapolate it to $x = 0$:

$$y(0) = \frac{S(0)}{S(1) + \beta}. \tag{4.17}$$

At a given value x_0 of the normalized background strength achieved with the chosen cut, the corresponding normalized number of TCS events is:

$$y(x_0) = \frac{S(x_0) + x_0\beta}{S(1) + \beta}. \tag{4.18}$$

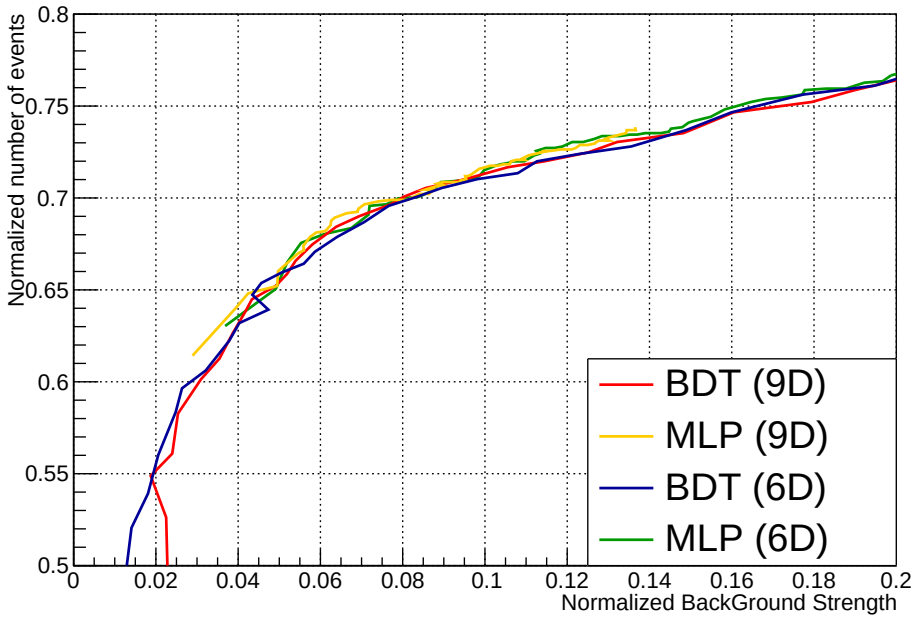


Figure 4.17: Pseudo-ROC curve obtained from data for 6D (SFs, shower widths) and 9D (SFs, shower widths and shower skewness) classifiers. The plot is zoomed in the region where the background is minimal while the signal is maximal. Adding the skewness information in the input of the classifiers does not improve their performances.

Solving Equation (4.18) for $B(x_0)$ and assuming that the signal is almost constant $S(x_0) = S(0)$, one can then obtain an estimate of the background/signal ratio at x_0:

$$\frac{B}{S}(x_0) \approx \frac{y(x_0)}{y(0)} - 1. \tag{4.19}$$

The full derivation is given in Appendix H. In order to get a simple reading of the background/signal ratio as a function of x, the pseudo-ROC curves of the 6D BDT and MLP are fitted with a linear function in the range $0.1 < x < 1$. The pseudo-ROC curves $y(x)$ are then transformed as:

$$\frac{B}{S}(x) = \frac{y(x)}{y(0)} - 1, \tag{4.20}$$

where $y(0)$ is the intercept of the fit. The obtained curves are shown in Figure 4.18, from which one can get an estimation of the background/signal ratio as a function of the normalized background strength in the region where the function is linear.

The B/S ratio is estimated to be close to 0.05 for a cut at 0.5 on the MLP output corresponding to a Background strength of 0.15 (see Figure 4.19 for values of the background as a function of the cut applied on the output of the 6D MLP classifier). Finally the pion contamination C_{π^+} above 4 GeV is estimated as:

$$C_{\pi^+} = \frac{B}{S+B} = \frac{1}{1+\frac{S}{B}} \approx 5\%. \tag{4.21}$$

4.3.7 Systematic checks on simulations

A second systematic check of the efficiency of the 6D MLP classifier was performed on simulations. A sample of BH-weighted events (see Chapter 5 for details on the TCS simulations) was used to test the classifier. This sample is completely uncorrelated from the training sample. It also assumes no hypothesis on momentum, polar angle and transverse momentum, apart from the ones arising from

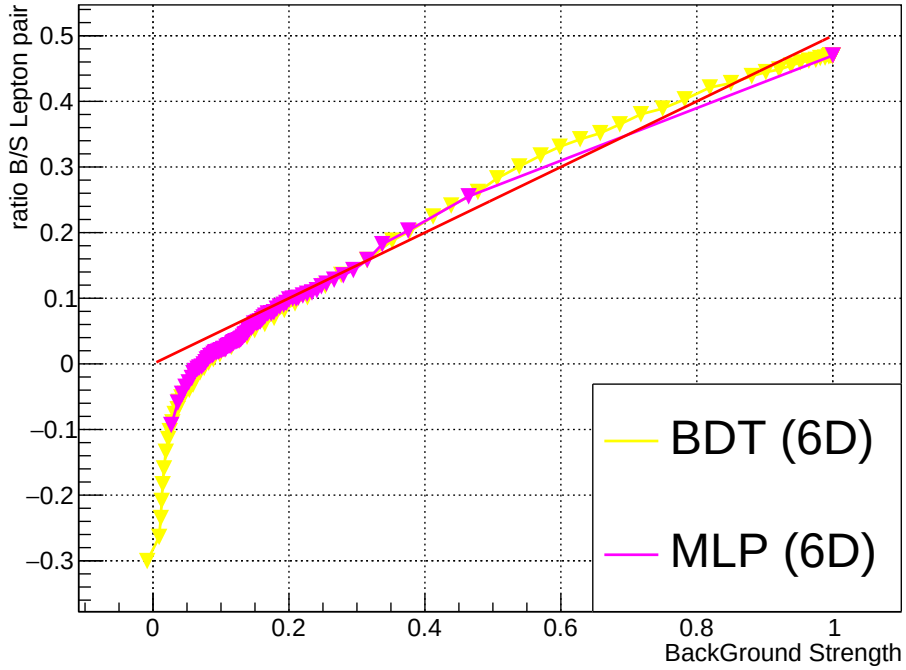


Figure 4.18: Background/Signal ratio in the TCS sample as a function of the background strength evaluated with the neutron sample. The red line corresponds to the linear fit of the BDT curve between 0.1 and 1.

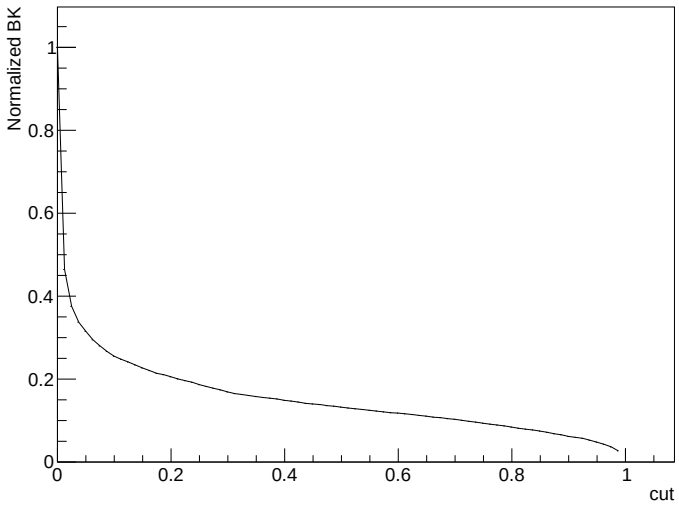


Figure 4.19: Normalized remaining background as a function of the cut applied to the output of the 6D MLP classifier.

the event generation. The positron signal efficiency is shown in Figure 4.20. This figure shows that the loss of true positrons in simulations when a cut is applied at 0.5 on the 6D MLP classifier output is less than 1%. Further tests on the kinematic distribution were also performed. Figures 4.21a, 4.21b and 4.22 show the distributions of momentum, polar angle and azimuthal angle of simulated positrons before and after the 0.5 cut was applied. No large systematic variation is seen, although some events do not fall in the kinematic limits of the training sample described in Subsection 4.3.1. This confirms

that this classifier can be applied to events in the kinematics region of TCS events without losing signal events. One can see in Figure 4.22 that most true positrons that are identified as mis-id. pions are located on the outside edge of sector 3 (ϕ between 100° and 150°) and 5 (between -150° and -100°) of the FD of CLAS12. These two sectors have LTCC modules located between the DCs and the FTOF/EC carriage. The showers that initiate in LTCC module walls could be a reason for these mis-identifications.

Finally a similar check is performed on pion simulations. A sample of π^+ with flat kinematics is simulated and the kinematic distribution of mis-id. pions before and after the 6D MLP 0.5 cut are compared. Figure 4.23 shows the momentum and θ distributions of π^+ before and after the cut. No large systematic efficiency shift is seen.

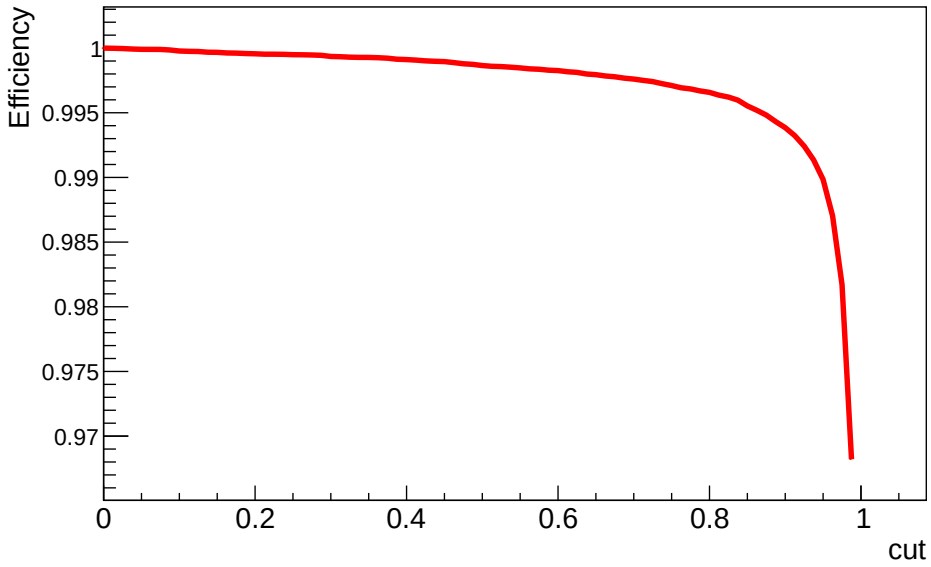


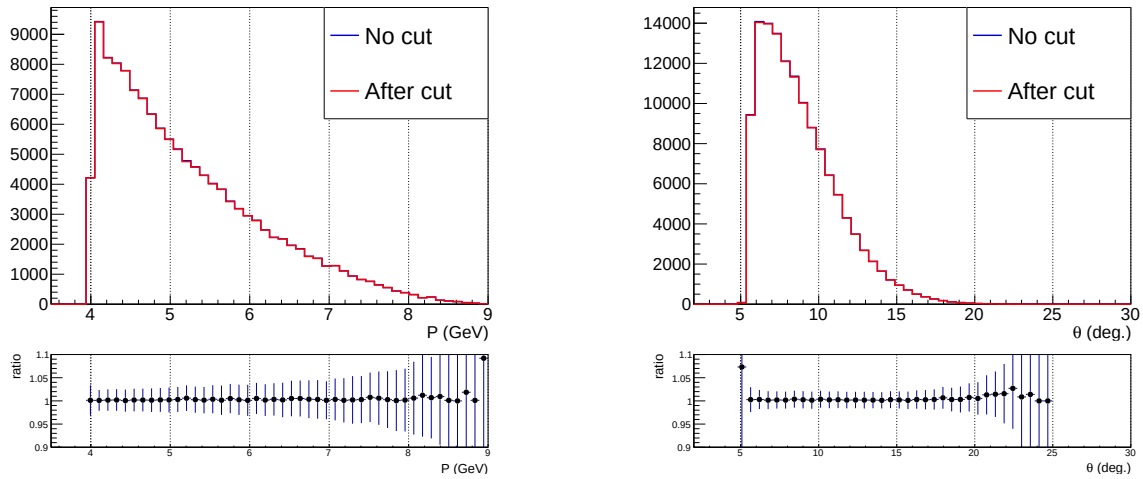
Figure 4.20: Signal efficiency for simulated TCS events (inbending electron) using the 6D MLP classifier as a function of the cut applied on the classifier output. At cut=0.5, the fraction of signal which is lost is less than a 1%.

4.3.8 Effect on data

The 6D MLP classifier was finally applied to the full available TCS data set. The output of the classifier and the signal efficiency on high-momentum positron candidates are shown in Figures 4.24 and 4.25, respectively. One can clearly see the mis-id. pions (around 0) and the true positrons (around 1) in the classifier output distribution. The effect of the classifier on the momentum distribution is assessed in Figure 4.26. One can clearly see an excess of positrons above the HTCC threshold in raw data. After applying the cut, the excess is removed.

4.3.9 What about electron PID?

So far only the case of positron identification was treated. Figure 4.27 shows the distribution of the momentum of electrons versus the momentum of positrons for TCS events without any other cuts that the ones from EB and the ones described in Subsection 4.3.5. The π^+ contamination, mostly coming from photo-production events $ep \to e'p'\pi^+(\pi^-)$, is visible in the positron high-momentum region. On the contrary no clear sign of π^- contamination in the electron high-momentum region is visible. This can be explained by the fact that high-momentum electrons $P_{e^-} > 4.9$ GeV are produced along with low momentum positrons $P_{e^+} < 4.9$ GeV. In this kinematics the positron is identified with a 99% efficiency by the HTCC. Lepton number conservation imposes that an electron is produced with the positron. Thus the simplest events that would contribute to the π^- contamination in the TCS data are



(a) (b)

Figure 4.21: (a) Momentum spectrum of positrons for simulated TCS events. (b) θ spectrum of positrons for simulated TCS events. In both cases the histogram before and after MVA pid cut, of value 0.5, superimpose. Ratios of both histograms are also shown. The number of cut events per bin is always lower than 3%.

the ones with final state $pe^+\pi^-(e^-e^-_{scat}\pi^+)$, where the electrons and π^+ would be undetected. Such events are suppressed by, at least, a factor of $\alpha_S^4 < 1\%$.

In addition to these considerations, the 6D MLP positron classifier was tested on electrons with momentum higher than 4.9 GeV from the TCS sample. This classifier is assumed to provide good results on electrons as the shower mechanism does not depend on the charge of the initiating particle at high energies. The signal efficiency as a function of the cut applied to the classifier output for simulation and data electrons is shown in Figure 4.28. One can see that simulated electrons are suppressed at most up to 3% for a 0.8 cut. The same behavior is seen for data electrons, showing that the background is less prominent than in the positron case. No further pid cuts for electrons than the EB ones are added in the analysis.

Finally the double pion contamination (where both leptons are mistaken for pions) is absent in the high-momentum region. The kinematic region were both leptons have momenta higher than 4.9 GeV is not accessible in CLAS12 kinematics.

4.4 Proton momentum corrections

The proton momentum corrections are split in two main contributions. The first corrections, presented in Subsection 4.4.1, are determined by comparing Monte-Carlo generated and reconstructed proton kinematics. The shifts observed in this case is attributed to the energy lost by the proton while propagating in the various detector materials. The determination of the parameters of this correction solely relies on simulations. The second contribution is a data-only based correction, which aims at correcting mis-alignments and inefficiencies of the actual detectors, not accounted for by the simulation. In particular, data-driven corrections are developed for the momenta of the protons in the CD in Subsection 4.4.2.

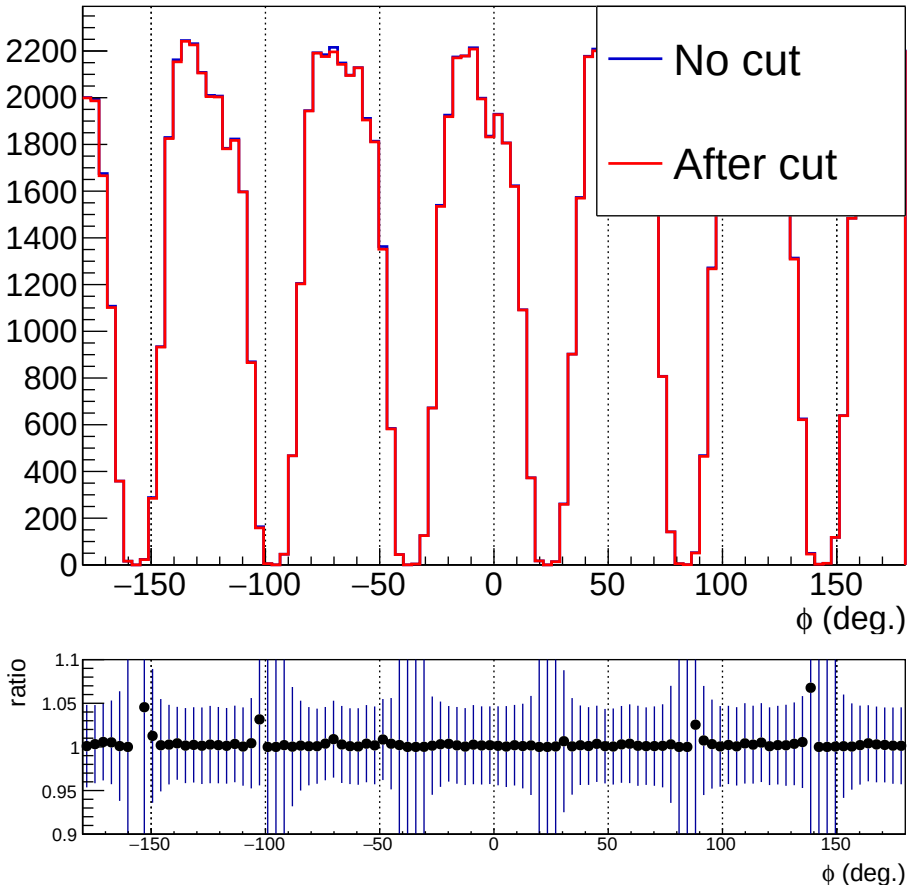
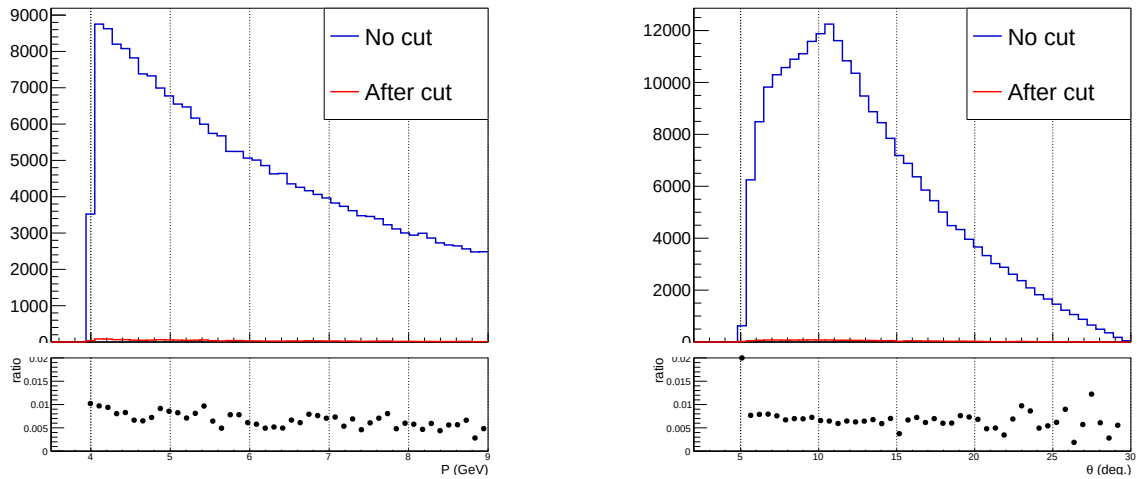


Figure 4.22: ϕ spectrum of positrons for simulated TCS events. The loss of positrons happens at very specific ϕ angles corresponding to the edges of sectors which accommodate an LTCC module.



(a) (b)

Figure 4.23: (a) Momentum spectrum of mis-id. pions. (b) θ spectrum of mis-id. pions. The histograms before and after the MLP cut at 0.5 are shown.

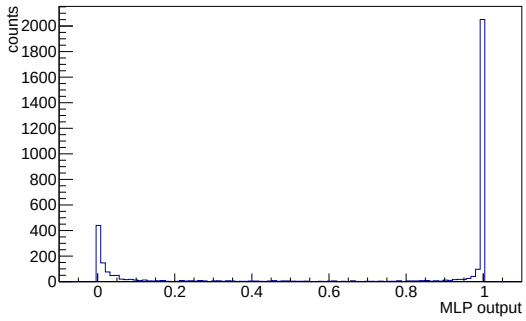


Figure 4.24: Output of the 6D MLP neural network on the TCS events used in the analysis.

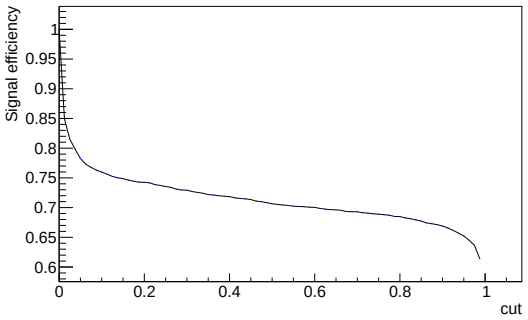


Figure 4.25: Signal efficiency of the cut applied on the MLP output. This corresponds to the proportion of remaining events for a given cut.

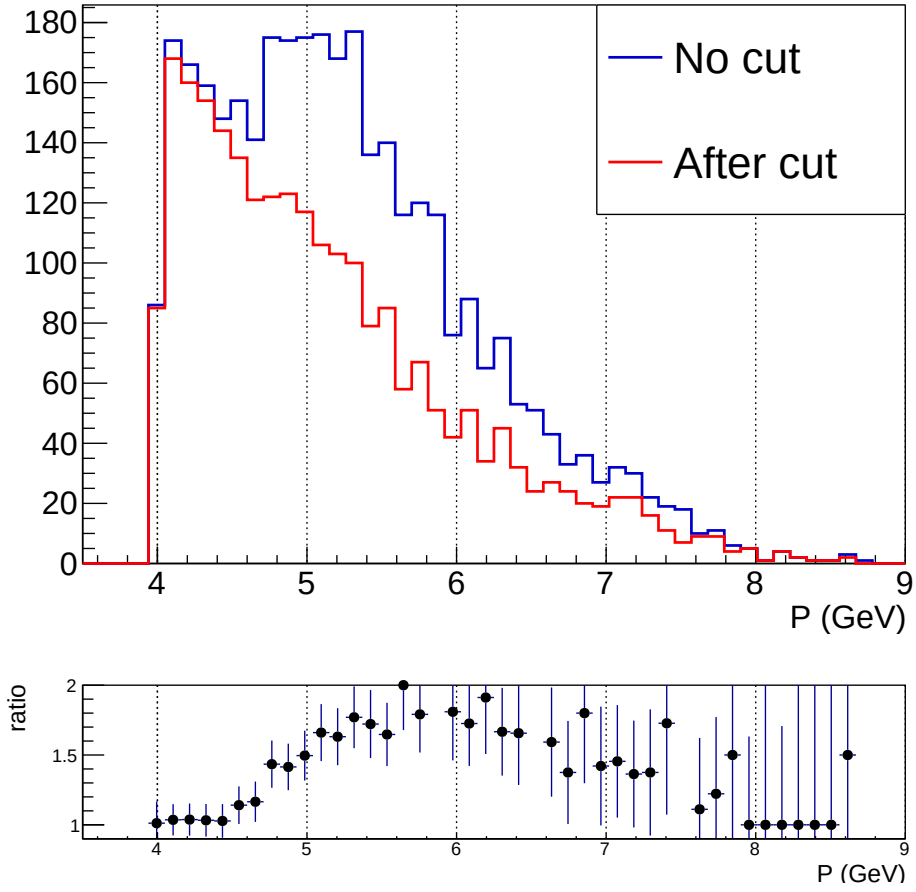


Figure 4.26: Momentum of positrons for TCS events. The blue histogram corresponds to all events. The red histogram corresponds to events that passed a 6D MLP output cut of 0.5.

4.4.1 Monte-Carlo corrections

The Monte-Carlo (MC) momentum corrections for the proton are derived using BH simulations (see Section 5.1 for details). The goal of these corrections is to match the momenta of the generated protons with the momenta of the reconstructed protons. These corrections account for the energy lost by the proton while crossing the various detector parts of CLAS12.

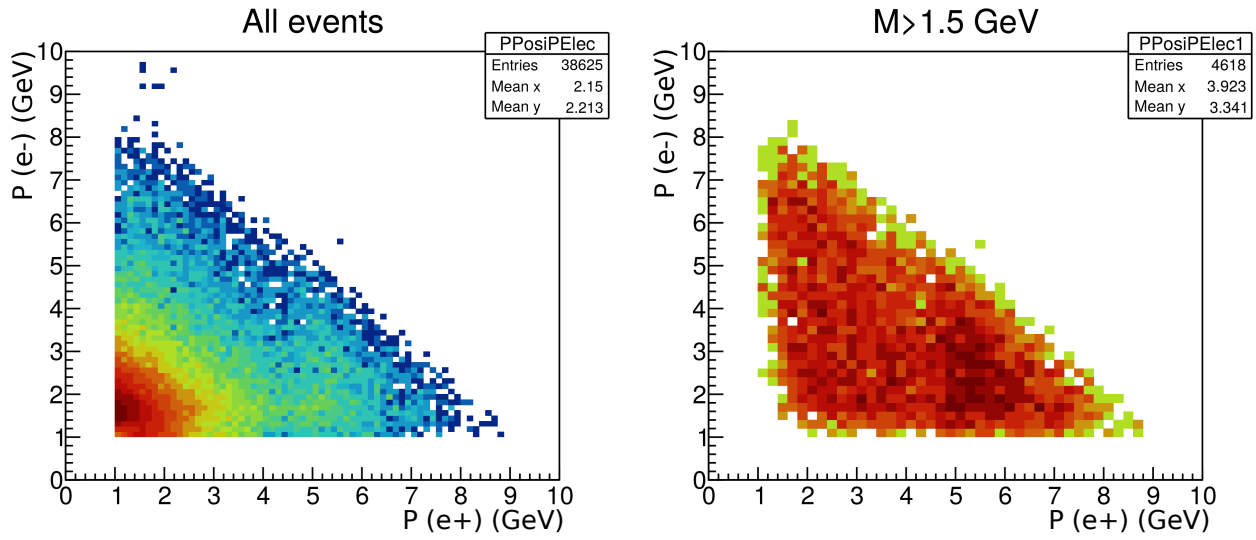


Figure 4.27: Momentum of the positron versus momentum of the electron in e^-e^+p events for all invariant masses (left) and only invariant masses higher than 1.5 GeV (right). No additional PID cuts than the EB ones are used in these two plots. The π^+ contamination is clearly visible in the $P(e^+) > 4.9$ GeV region in the left figure. On the contrary no π^- contamination is visible in the $P(e^-) > 4.9$ GeV.

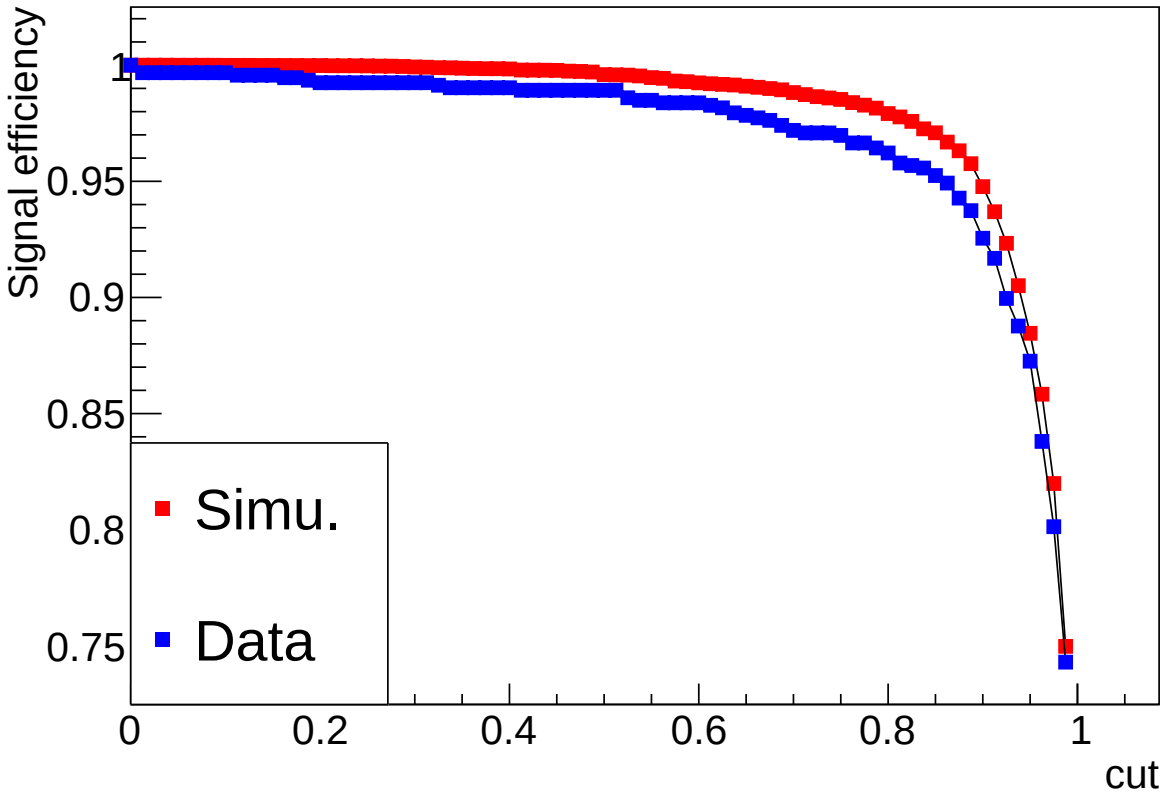


Figure 4.28: Signal efficiency of the MLP classifier for electrons with momentum bigger than 4 GeV, for simulated TCS events (red) and data (blue).

The difference between the generated and reconstructed momenta,

$$\Delta P = P_{Gen.} - P_{Rec.}, \tag{4.22}$$

is studied as a function of the polar angle of the proton, θ. The plot in Figure 4.29 shows the difference between the generated and reconstructed proton momenta as a function of θ for protons detected in the FD. The momenta difference shows different behaviors in two distinct regions. Below 27° protons cross little material before being detected and the momentum difference is small (below 20 MeV). Above 27° the material budget between the target and the DCs is larger, especially due to the HTCC and the forward CTOF light-guides. In this region the momentum resolution is degraded and the momentum difference can reach up to 80 MeV.

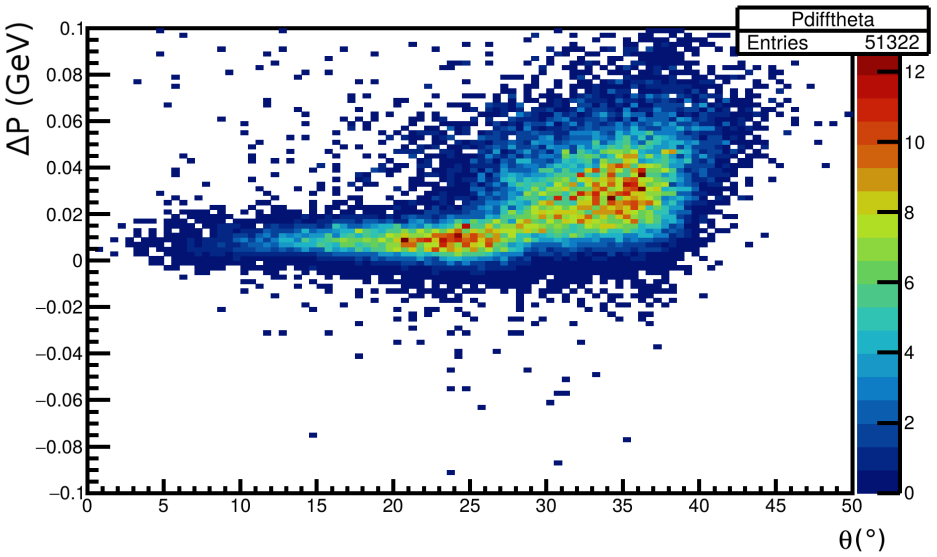


Figure 4.29: Difference between the generated and reconstructed momenta for protons detected in the FD of CLAS12. One can see two distinct regions, below and above $\theta = 27°$.

The MC corrections are derived in three different CLAS12 regions: the two regions in the FD described above and one region in the CD. In each region the momenta difference is parametrized as a function of the reconstructed momentum as shown in Figure 4.30. These corrections are at most of the order of 4% for low-momenta proton (~ 0.45 GeV) in the high-polar angle region of the FD. The corrections are applied to protons in both the simulations and the data.

4.4.2 Data-driven momentum corrections

Data-driven momentum corrections for the proton are motivated by the fact that the simulations depict an "ideal" detector, and therefore do not perfectly reproduce the data. Due to the detection inefficiencies of the CVT, the reconstructed momentum in the CD can be shifted from its actual value. To investigate this issue a method using exclusive two-pion production events was developed. This method relies on the exclusive measurement of the $ep \to e'p'\pi^+\pi^-$ reaction, where the scattered electron and the pions are detected in the FD. The kinematics of the scattered proton can then be studied in two different ways. In one case the proton can be detected by CLAS12, in the other case its kinematics can be inferred by calculating the missing 4-momentum of X in the $ep \to e'\pi^+\pi^-X$ reaction.

The missing mass spectrum obtained from the latter analysis is shown in Figure 4.31. One can see that the missing mass shows a clear peak at the proton mass. Furthermore, looking at the dependence of the missing mass as a function of the missing polar angle in Figure 4.31b, one can see that at high polar angles (above 35°) the high-mass component is suppressed and the missing mass spectrum has a contribution only from the scattered proton. This allows to compare directly the kinematics of the missing proton to the kinematics of the detected proton.

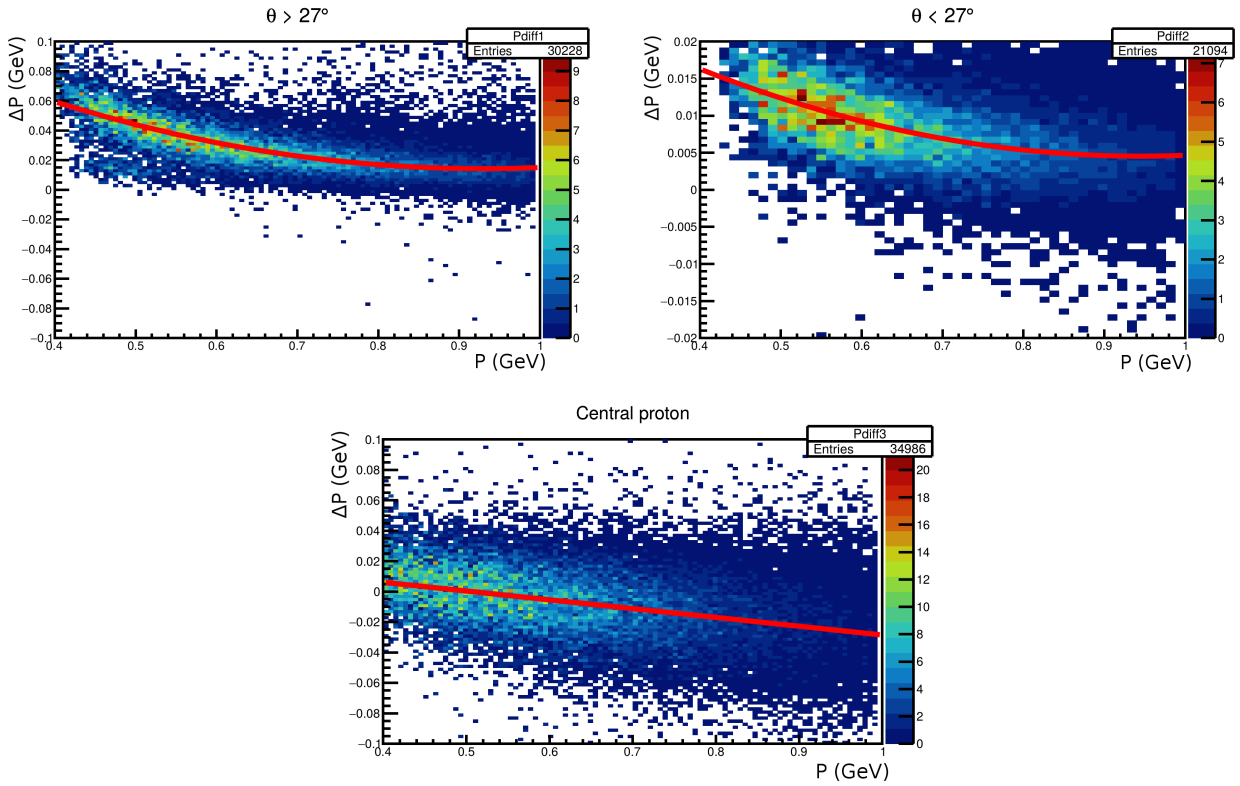
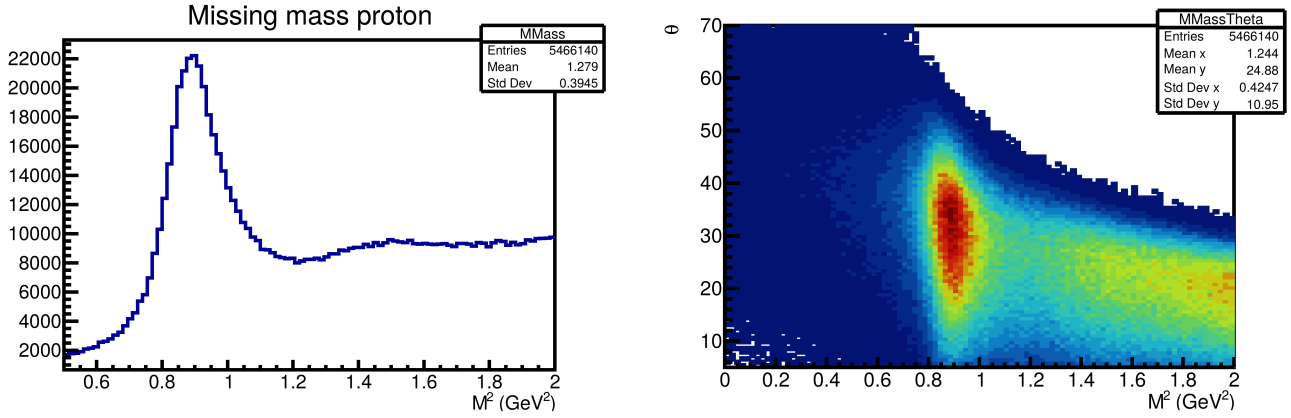


Figure 4.30: Top left: momenta difference (defined in Equation (4.22)) as a function of the reconstructed momentum for polar angles above 27° in the FD. The distribution of the mean of each momentum slice (obtained by a gaussian fit) is fitted with a 2nd order polynomial. The result of this fit is superimposed to the original distribution (red line). Top right: Corresponding figure for protons with polar angles below 27° in the FD. Bottom: Corresponding figure for protons in the CD.

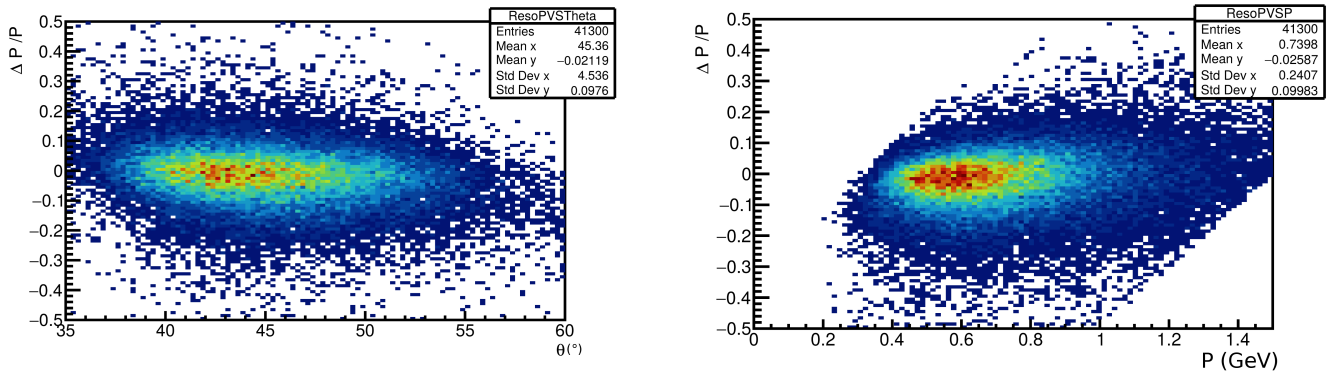


(a) (b)

Figure 4.31: (a) Squared missing mass of the $ep \to e'\pi^+\pi^- X$ reaction. One can see a clear peak at the proton mass and a higher-mass continuum. (b) Missing polar angle as a function of the squared missing mass for the same reaction. Once can see that the high-polar-angle region, corresponding to topologies where the missing proton goes in the CD, is free of high-mass background.

Figure 4.32 shows the momentum resolution:

$$\frac{\Delta P}{P} = \frac{P_{Missing} - P_{Rec.}}{P_{Rec.}} \tag{4.23}$$



(a) (b)

Figure 4.32: (a) Data-driven momentum resolution as a function of the polar angle for protons in the CD. (b) Momentum resolution as a function of the momentum.

as a function of the detected polar angles (Figure 4.32a) and the detected momenta (Figure 4.32b). No large dependencies are seen. The momentum resolution is also plotted as a function of the local azymuthal angle in the last layer of the CVT, ϕ_{CVT}, as shown in Figure 4.33.

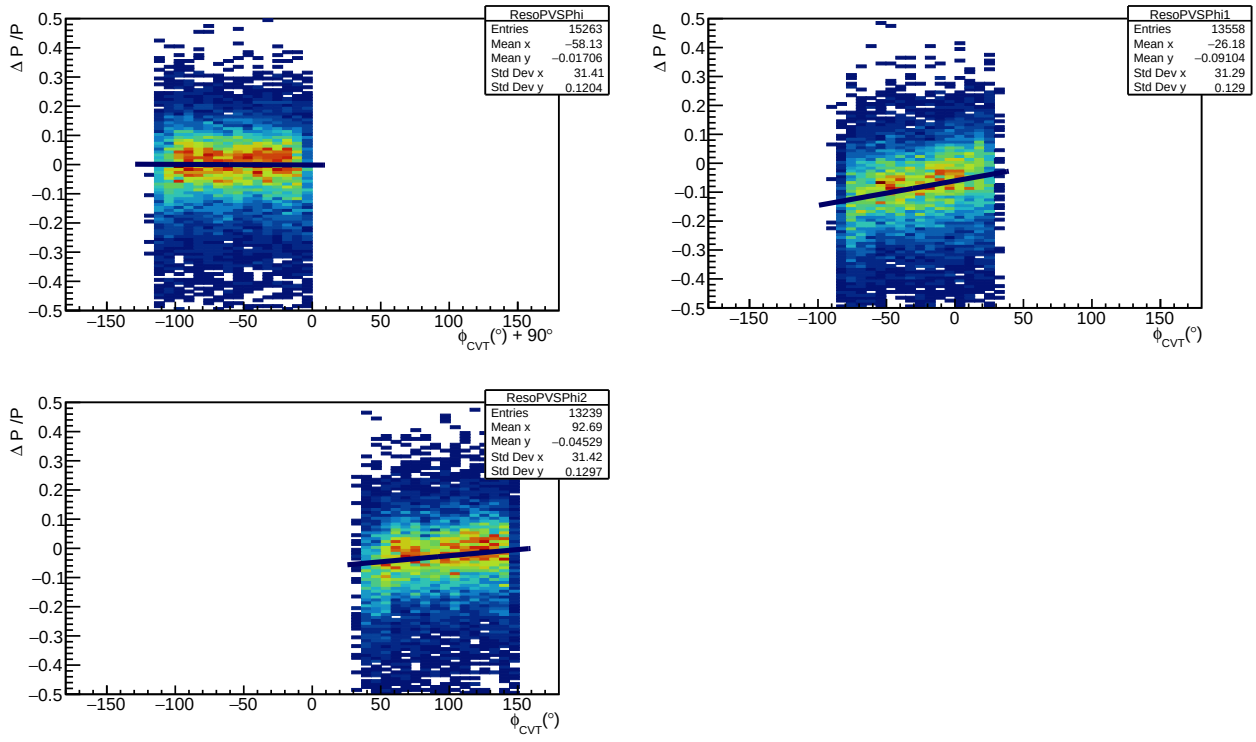


Figure 4.33: Momentum resolution as a function of the local azymuthal angle for protons in the CD, for the three regions of the CVT. The superimposed black line is the correction function for each region.

Each subplot corresponds to one of the three CVT regions. The distribution of the gaussian means of each ϕ_{CVT} slices is fitted with a linear function, for each subplot. The resulting resolution corrections are applied to the proton momenta in the data. The corrections range from almost zero for protons in the region $-210° < \phi_{CVT} < -90°$ to up to 14% at the lower edge of the $-90° < \phi_{CVT} < 30°$ region. These corrections are performed after the MC corrections presented in the previous subsection.

4.5 Lepton momentum corrections

In this section we present the corrections applied to leptons. First, in Subsection 4.5.1, the shift between the Monte-Carlo generated and the reconstructed kinematics is analyzed. Second, in Subsection 4.5.2, a specific detected-photon correction is presented.

4.5.1 Monte-Carlo corrections

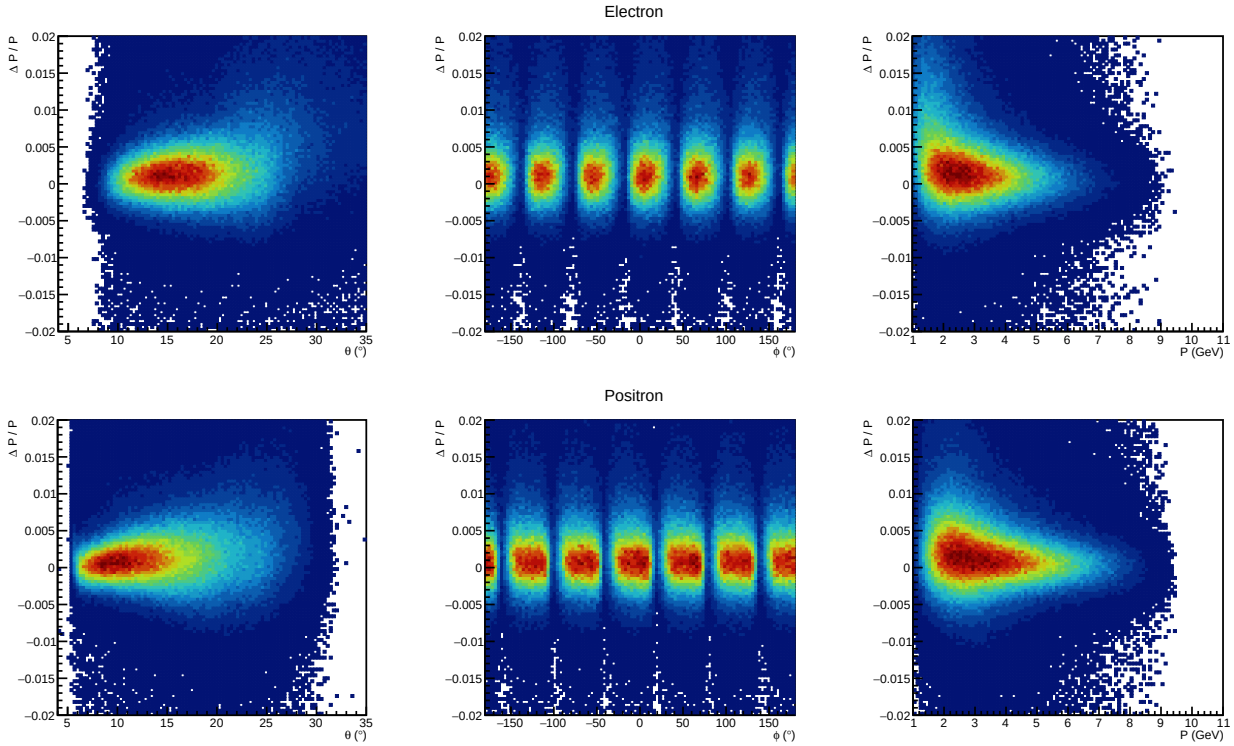


Figure 4.34: Reconstructed momentum resolution for electrons (top) and positrons (bottom), as a function of the lab angles θ, ϕ and the momentum.

As for proton, BH-simulation events are used to investigate the impact of the energy lost by the lepton in the material of CLAS12 before being detected. This energy loss ultimately affects the reconstruction of the kinematic variables, especially the momentum. Figure 4.34 shows, for the electrons and the positrons, the momentum resolution:

$$\frac{\Delta P}{P} = \frac{P_{Gen.} - P_{Rec.}}{P_{Rec.}}, \tag{4.24}$$

where $P_{Gen.}$ is the generated momentum and $P_{Rec.}$ the reconstructed one, as a function of their polar and azimuthal angles in the lab reference frame, θ and ϕ, and their momentum. These plots show that the shift in momentum due to the energy loss of leptons is always smaller than 1%. For this reason, we decided not to add MC-driven corrections to the momentum of leptons in the subsequent analysis.

4.5.2 Detected radiated photon correction

Leptons propagating through the target material can lose energy by radiating photons. These photons are emitted at low angles around the direction of the lepton. The process is represented by the diagram of Figure 4.35.

One way to recover the initial momentum of the lepton, before any radiative energy loss, is to use the detected photon in the CLAS12 EC. Figure 4.36a shows, in the CLAS12 data, the uncorrected momenta of electrons versus the difference in the polar angle at the vertex between electrons and detected

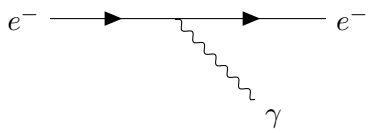
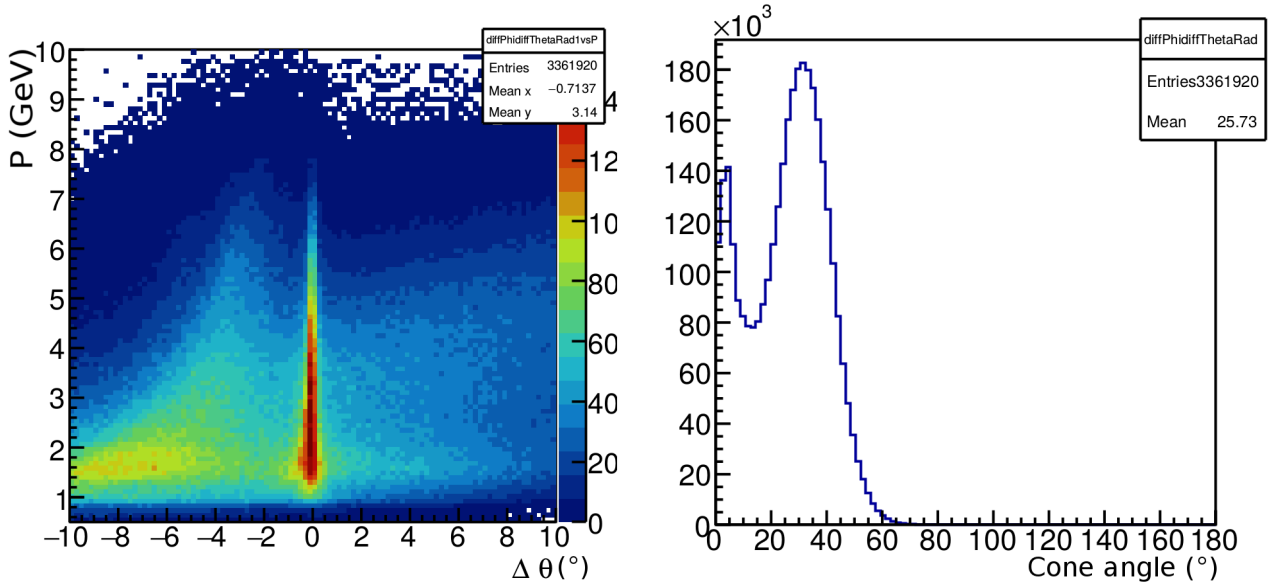


Figure 4.35: Diagram representing the radiation of a photon from an electron.

photons, $\Delta\theta_\gamma$, while Figure 4.36b shows the cone angle between electrons and detected photons. One can see that in both plots there is a significant number of photons detected in the close vicinity of the electron. The same behaviour is seen for positrons. Similar results are also obtained for simulations, as seen in Figure 4.37.



(a) (b)

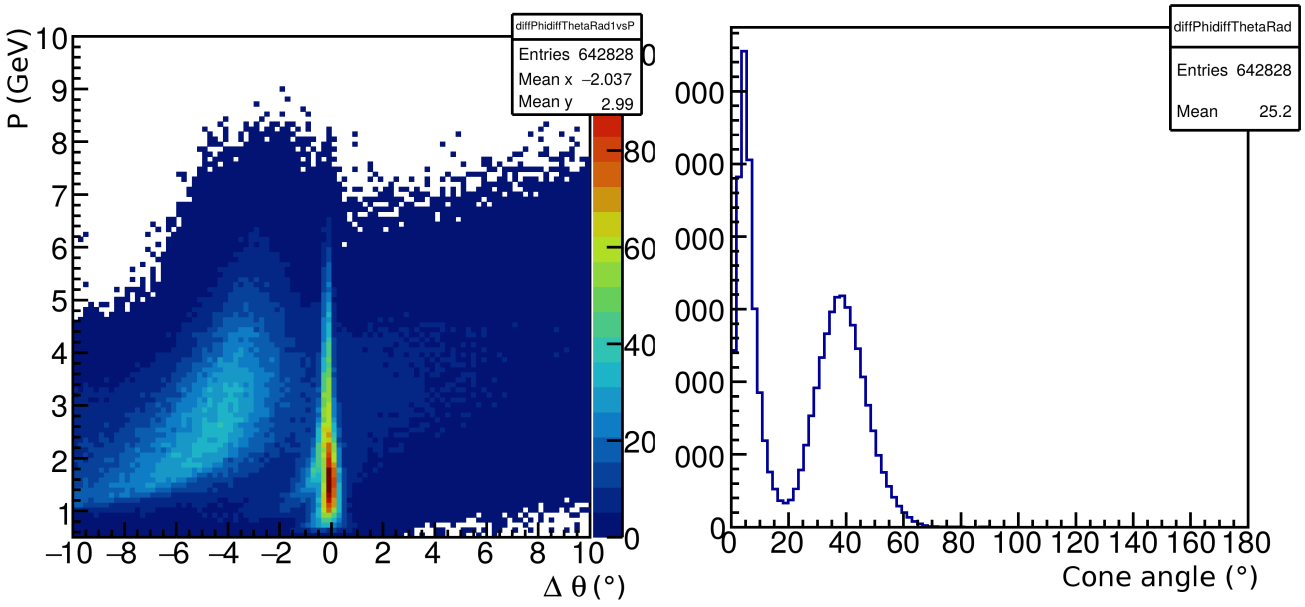
Figure 4.36: (a) Electron momentum as a function of the difference of polar angle between the electron and the photons detected in each event, (b) Cone angle between the electron and same event photons. These plots are produced with real CLAS12 data.

The 4-momenta of photons within $-1.5° < \Delta\theta_\gamma < 1.5°$ and with a cone angle below 10° are added to the reconstructed 4-momentum of the corresponding lepton. The effect of this correction can be seen in Appendix I. The improvements in the momentum and angular resolutions are shown for the electrons for which this correction is applied.

4.6 Fiducial cuts

Fiducial cuts consist in excluding regions where the detection efficiency is not well reproduced by the simulations. This mismatch between the data and the simulations mainly occurs on the edge of the detectors, where the detection efficiency varies fast. For calorimeters, this occurs in regions where the energy deposition is incomplete. In this analysis we develop a fiducial cut for the PCAL. This cut removes parts of the detector where electromagnetic showers are reconstructed too close to the edge of the active region. This ensures that most of the energy of an electromagnetic shower reconstructed in the PCAL is measured.

The PCAL fiducial cut developed in this section is based on the width of the measured electromagnetic shower. The definition of the square of the shower width in the PCAL is given in Formula (4.11). The mean shower size $\sqrt{M_2}$ is calculated for each of the three calorimeter coordinates (U, V and W).



(a) (b)

Figure 4.37: (a) Electron momentum as a function of the difference of polar angle between the electron and the photons detected in each event, (b) Cone angle between the electron and same event photons. These plots show simulated electrons.

Figures J.1 and J.2 in Appendix J show the distributions, for sectors 1 to 3, of the mean shower size as a function of respectively low and high values of U, V and W. The shower mean size is estimated away from the edge of the calorimeter (red lines). The value of the cut is then set at a distance equal to the mean shower size from the edge of the distribution (black lines). The cut values used in the TCS analysis are summarized in Table 4.1, where the maximum and minimum values of V and W are given for each sector of CLAS12. Only the cuts along V and W are used, as cutting on the U bars was found to be redundant.

sector	$V_{\min}$	$V_{\max}$	$W_{\min}$	$W_{\max}$
1	9.78924	402.06	9.47359	393.895
2	8.62768	402.389	8.57818	402.064
3	9.23112	403.875	8.23956	403.622
4	19.2814	403.021	8.26354	392.355
5	8.73336	402.915	9.28017	403.634
6	9.12088	403.681	8.13996	403.886

Table 4.1: Values of the fiducial cuts used. The minimum and maximum values for the U and V coordinates in the PCAL are given for each sector.

The effect of the PCAL fiducial cuts on the electron distribution is shown in Figure J.3 in Appendix J. The left plot shows the removed electrons in black. The plot on the right shows the regions kept for the analysis. The fraction of electrons lost by these fiducial cuts was estimated on inclusive events (at least one electron detected in CLAS12) from the inbending data set described in Section 2.8. It ranges from 5% at 1 GeV to up to 30% at 10 GeV.

Other approaches have been used to define fiducial cuts on the EC of CLAS12 [95]. The values of the cuts derived with these methods are similar to the ones presented here. For this reason, the TCS analysis presented in the next chapter adopts the values of the cuts defined in this section.

Chapter 5

Simulations and extraction of the TCS observables

In this chapter, the various steps towards the experimental measurement of the TCS observables are presented. The simulation software is presented in Section 5.1. Exclusivity cuts based on simulations and used for event selection are justified in Section 5.2. The comparison of the kinematic distributions of data and simulation is reported in Section 5.3. Acceptance and proton efficiency studies are displayed, respectively, in Sections 5.4 and 5.6. The background reactions for the TCS events are discussed in Section 5.7. The formulae used to calculate the TCS observables from data are detailed in Sections from 5.8 to 5.11. The binning used for data is displayed in Section 5.12. Finally the systematic errors are estimated in Section 5.13. Results from this analysis will be presented in the next chapter.

5.1 TCS event generator and simulations

Simulations of the $\gamma p \to e^- e^+ p'$ reaction are necessary for the analysis of TCS. They are used especially to determine relevant exclusivity cuts as well as estimating the acceptance of CLAS12 for this reaction. Two independent generators are used in this analysis. The first one, *TCSGen*, was developed during the exploratory TCS analysis of CLAS data [59]. A second generator developed for HERA, called *GRAPE* is also used to cross-check *TCSGen*, as well as to explore pair production background and interference between final-state electrons (see Subsection 5.7.1).

5.1.1 GRAPE

The *GRAPE* Generator [96] was developed for experiments running at HERA, at the DESY laboratory in Hamburg. It is a di-lepton generator to study the electromagnetic background mainly in J/Ψ and Υ measurements. It is a full generator including the kinematics of the beam electron producing the real photon. It also includes pair production from the incident electron as well as interferences between leptons in the final state. The diagrams which can be included in the generator are shown in Figure 5.10. It is a non-weighted event generator.

5.1.2 TCSGen

TCSGen is a generator developed by R.Paremuzyan for the first CLAS analysis of TCS, and completed and corrected for the purpose of this analysis. It generates weighted events, with the possibility to use BH-only weights or to include the TCS-BH interference cross section. Contrary to *GRAPE*, the initial electron from the beam is not included and no possible interferences are taken into account. Each event is assigned a weight w equal to:

$$w = psf \cdot \sigma \cdot flux_\gamma, \tag{5.1}$$

where psf is the phase space factor, σ is the cross section of the considered processes, and $flux_\gamma$ is the equivalent photon flux given in [97]. The phase space factor is the product of the ranges allowed

for each kinematic variable. First the energy of the incoming real photon is randomly picked between a minimum energy $E_{\gamma\ Min}$, defined by the user of the generator, and the maximum possible energy, equal to the electron beam energy E_b. The photon energy phase-space factor is $psf_{E_\gamma} = E_b - E_{\gamma\ Min}$. Second, the range allowed for the square of the transferred momentum to the proton $-t$ is fully defined by the value of E_γ previously picked. The limit t_{min} and t_{max} of the possible range for $-t$ are given by Equation (5.10) in page 121 of [98]. The associated phase space factor is $psf_t = t_{max} - t_{min}$. After randomly picking a value for $-t$, the kinematically accessible invariant mass is fully defined. The associated phase space is $psf_M = M_{Max} - M_{Min}$. The limits $M_{Min/Max}$ are also given in [98]. Finally the center-of-mass angles θ and ϕ are randomly chosen, in the range $[0 - \pi]$ and $[0 - 2\pi]$, respectively. The final-state particles are then boosted to the lab frame and a final azymuthal rotation is performed. The formula for the phase-space factor is:

$$psf = psf_{E\gamma} \cdot psf_t \cdot psf_M \cdot psf_{\theta\ \mathrm{CoM}} \cdot psf_{\phi\ \mathrm{CoM}} \cdot psf_{\phi\ \mathrm{Lab}}. \tag{5.2}$$

The cross section σ can be either the BH one or include the TCS interference term. The formulae used are those from [57].

The kinematic distributions of events generated with *TCSGen* and with *GRAPE* in the same kinematic region and with BH cross section only are compared in Appendix K. The two generators agree, up to a normalization constant. For the TCS analysis, the acceptance calculations were performed using *TCSGen* as it allows to cover the whole phase space with less iterations than *GRAPE*, as it is a weighted generator and no generated events are discarded.

5.2 Event selection

5.2.1 Final state selection

The initial step of the extraction of TCS observables from the complete RGA dataset is the event selection. First the final state of interest is selected. The complete dataset is skimmed to select events with exactly one proton, one electron and one positron. We allow any other particle in the event to avoid removing good events where accidental TOF/EC hits or false tracks are recorded. The particles are selected according to the pid given by the CLAS12 reconstruction software (see Chapter 4), and using the dedicated positron identification neural network described in Chapter 4. The cut applied on the output of the neural network is set at 0.5. The momenta of the detected particles are corrected and fiducial cuts are applied. Finally a cut on the lepton momentum is also applied (leptons with momenta greater than 1 GeV are kept). This cut is motivated by the poor momentum reconstruction for tracks with large curvature (low momentum) in the FD.

5.2.2 Exclusivity cuts

Once the events with the relevant final state are retrieved, exclusivity cuts are applied. The reaction of interest is :

$$\gamma p \to e^- e^+ p'. \tag{5.3}$$

However the RGA dataset was not taken with a beam of photons but with electrons impinging directly on the target. Therefore quasi-real photo-production events are used in this analysis. A quasi-real photon is emitted by an electron from the beam and it interacts with the proton, producing a e^+e^- pair in the final state, as:

$$ep \to (e')\gamma p \to (e')e^- e^+ p'. \tag{5.4}$$

The corresponding conservation of momentum equation is:

$$p_{\mathrm{beam}} + p_{\mathrm{beam}} = p_{\mathrm{scat.}} + p_\gamma + p_{\mathrm{target}} = p_{\mathrm{scat.}} + p_{e^+} + p_{e^-} + p_p. \tag{5.5}$$

The 4-momenta of the scattered electron and initial real photon are fully defined by the measurement of the 4-momenta of the final state particles:

$$p_\gamma = p_{e^+} + p_{e^-} + p_p - p_{\mathrm{target}}, \tag{5.6}$$

$$p_{\text{scat.}} = p_{\text{beam}} + p_{\text{target}} - (p_{e^+} + p_{e^-} + p_p). \tag{5.7}$$

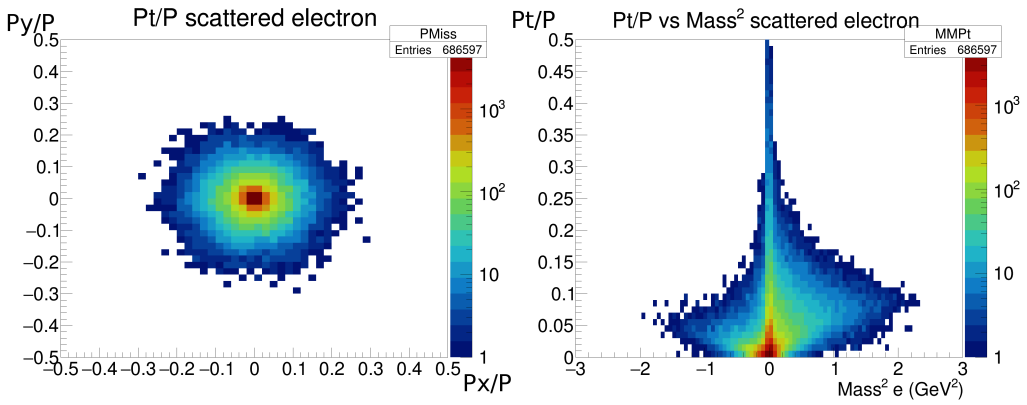
Two exclusivity cuts are applied on the missing particle (X) of the pe^+e^-X system. The mass and the transverse momentum fraction of the missing particle X are constrained to be close to zero. The mass cut ensures that the missing particle is an electron. The transverse momentum cut ensures the low virtuality of the photon. Indeed, the virtuality of the incoming photon can be written:

$$Q^2 = 2E_b E_X (1 - \cos(\theta_X)), \tag{5.8}$$

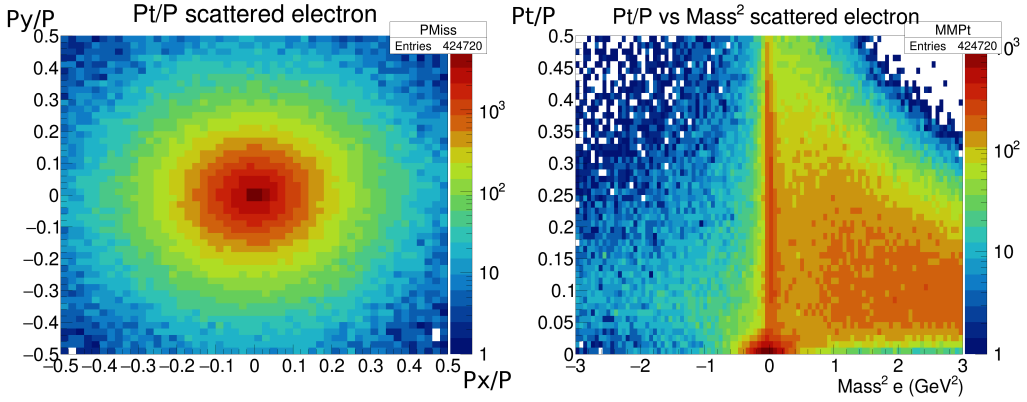
where E_b is the energy of the electron beam, E_X is the energy of the undetected scattered electron, θ_X is its scattering angle in the lab frame given by:

$$\tan(\theta_X) = \frac{Pt_X}{P_X}, \tag{5.9}$$

and P_X and Pt_X are, respectively, the momentum and transverse momentum of the missing particle. The values of these cuts are motivated by simulations. Figure 5.1a shows the distribution of simulated events in the transverse missing momentum fraction plane (left plot) and the transverse momentum fraction as a function of the missing mass (right plot). Photo-production events correspond to events at small transverse momentum fraction and small missing mass. The same distributions are shown for the data in Figure 5.1b. A similar behavior is seen, with the addition of high missing mass background.



(a)



(b)

Figure 5.1: Distributions of the exclusivity variables for (a) simulations and (b) the data.

The cuts used in the analysis are:

$$\left|\frac{Pt_X}{P_X}\right| < 0.05 \tag{5.10}$$

and

$$|M_X^2| < 0.4 \text{ GeV}^2. \tag{5.11}$$

The resulting distribution of incoming photon virtuality, calculated using Equation (5.8), is shown in Figure 5.2. Finally, the variation of the extracted quantities with different exclusivity cuts is investigated and accounted for in the systematic uncertainties (see Section 5.13).

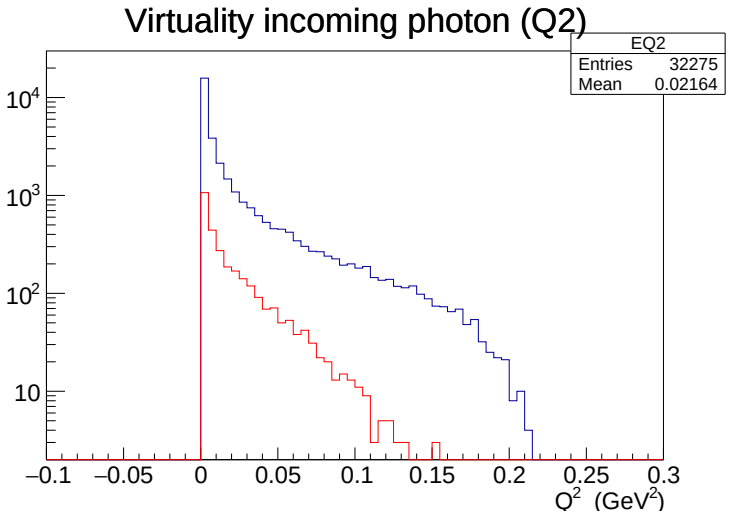


Figure 5.2: Virtuality of the incoming photon for all the events passing the exclusivity cuts (blue) and the events included in the analysis (red).

5.3 Phase space of interest and Simulations/Data comparison

After applying the exclusivity cuts detailed above, the invariant mass of the lepton pair is extracted. Figure 5.3 shows the invariant mass spectrum obtained for the full RGA Fall 2018 dataset. Meson resonances decaying into an electron-positron pair are visible.

The mass region between 1.5 GeV and the J/Ψ mass (3096 MeV) is selected for the measurement of TCS. The first reason for this choice is theoretical: the hard scale of TCS is given by the invariant mass of the lepton pair and has to verify $Q'^2 \gg m_p$ and $Q'^2 \gg -t$ in order to meet the conditions for the GPD factorization. As shown in the later analysis, a typical value of proton momentum transfer in this analysis is 0.3 GeV^2, which satisfy the previous condition. The second reason is that the mass range above 2 GeV is free from vector-meson resonances decaying into a di-lepton pair. The range between 1.5 GeV and 2 GeV has contributions from the wide vector mesons ρ (1450) and ρ (1700). In order to check the impact of these resonances, the mass spectrum obtained from the data is compared with BH-weighted simulations, between the Φ and the J/Ψ mass, in Figure 5.4. In this figure the simulated spectrum is normalized in order to have equal integral as the data spectrum, between 1.1 and 3 GeV. According to this plot there is no obvious meson contamination in the mass range of interest. The effect of this contamination on the observables is studied by extracting them in the two mass ranges ([1.5 GeV − 2 GeV] and [2 GeV − 3 GeV]).

Eventually, the phase space selected for the analysis is:

- 1.5 GeV $< M = \sqrt{Q'^2} <$ 3 GeV
- 0.15 GeV$^2 < -t <$ 0.8 GeV^2
- 4 GeV $< E_\gamma <$ 10.6 GeV.

Note that, unless specified otherwise, the observables are integrated over these kinematic ranges.

The distributions of the three variables $-t$, E_γ and M for data and simulations in the selected phase space are compared in Figure 5.5. The simulation distributions are normalized to have the same integral as data distributions. The simulations-data agreement for all three variable is good and no

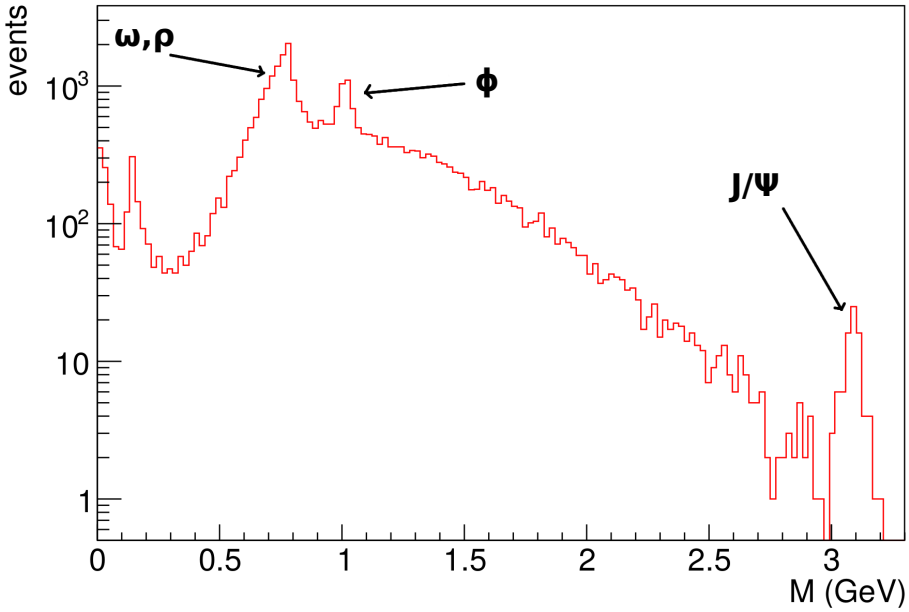


Figure 5.3: Lepton pair invariant mass spectrum after exclusivity cuts, extracted from data. Meson resonances (ω/ρ (770/782 MeV); Φ (1020 MeV) and J/Ψ (3096 MeV)) are visible. The peak at zero mass is due to photon conversion in the target material ($\gamma \to e^+e^-$) and π^0 Dalitz decay ($\pi^0 \to \gamma e^+e^-$). The peak at 0.1 GeV is an artifact induced by noise in the PMTs of the HTCC associated with two DC tracks in the same sector of CLAS12.

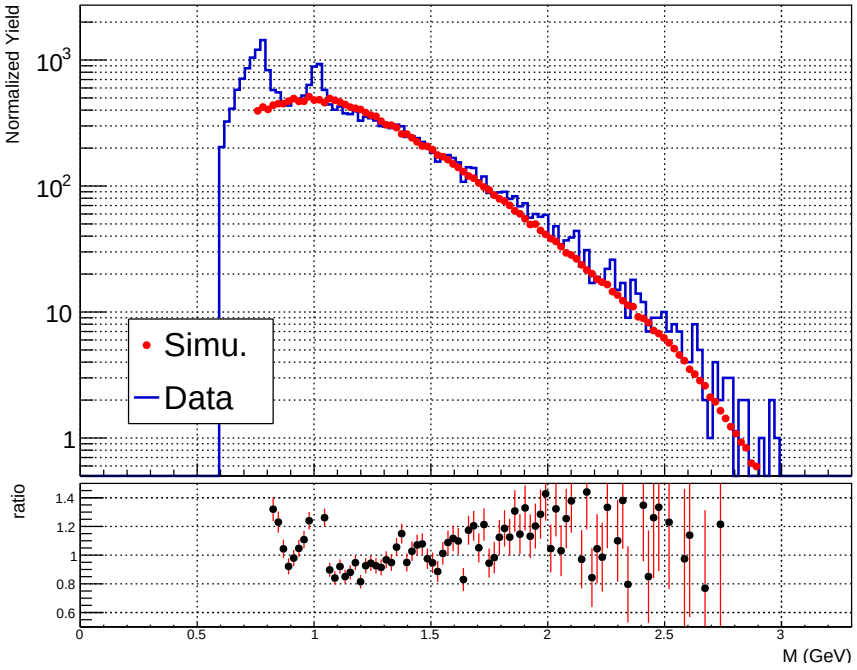


Figure 5.4: Comparison between data and simulations in the high-mass range. The ratio between the two plots is also shown. Cuts on the square of the transferred momentum (0.1 GeV^2 $< -t <$ 1 GeV^2) and the photon energy (3 GeV $< E_\gamma <$ 10 GeV) are applied.

large discrepancies are seen. The kinematics of the particles in the laboratory coordinates P and θ_{Lab} are displayed in Appendix L.

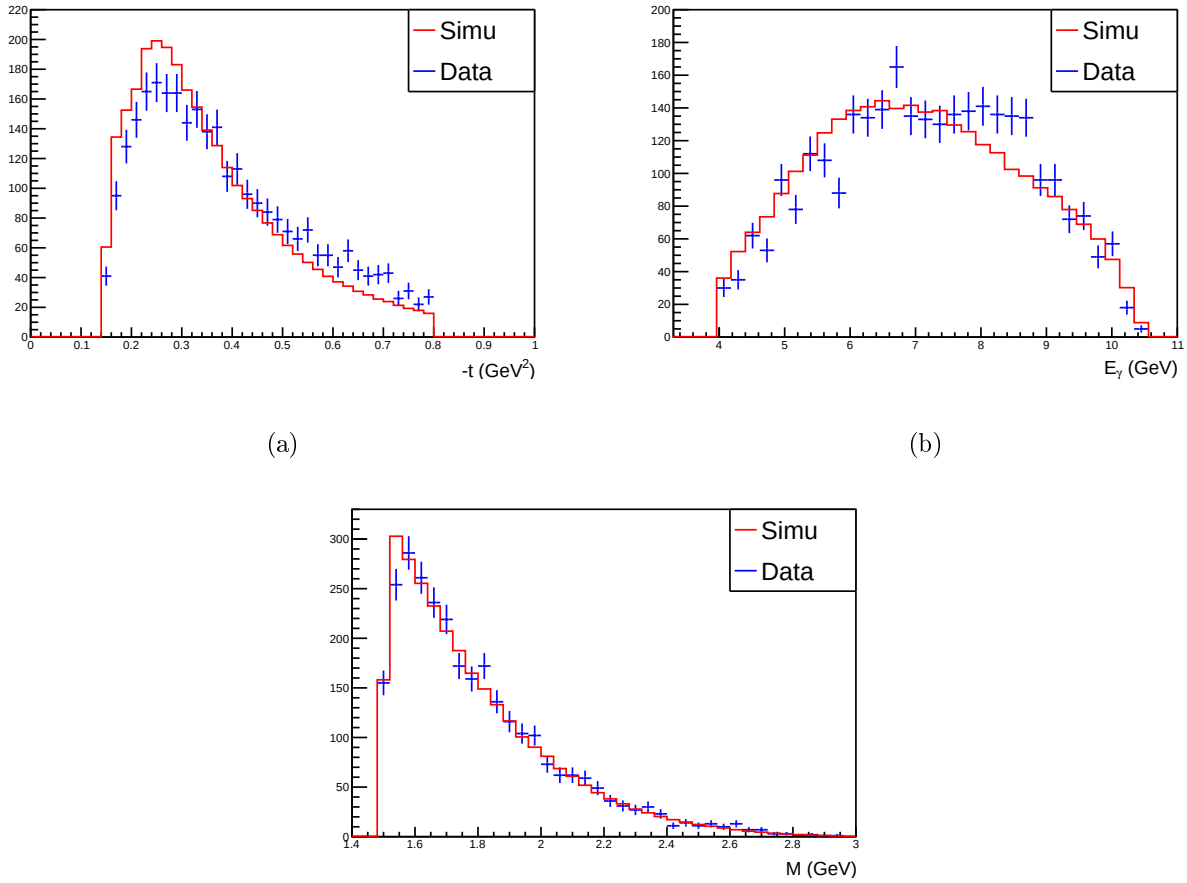


Figure 5.5: Simulations-data comparison for the distributions in (a) $-t$, (b) E_γ and (c) M. The simulation distributions are normalized to have equal integral as the data. Both the simulations and the data are in agreement, and no large discrepancies are seen.

5.4 Proton efficiency correction

In order to take into account the differences in the proton detection efficiency between real data and simulations, a proton efficiency correction to the GEMC simulation is implemented. This correction is derived using the same data sample as for the proton momentum corrections discussed in Subsection 4.4.2. The $ep \to e(p')\rho \to e(p')\pi^+\pi^-$ reaction is selected by applying a cut on the invariant mass of the two pions, 0.6 GeV $< M_{\pi^+\pi^-} <$ 1 GeV. The same reaction is generated using the *genev* event generator [99] and passed through the GEMC and the CLAS12 reconstruction softwares. The kinematics of the missing proton are assumed to be well reconstructed and are used to derive the correction. The proton efficiency is measured for data and simulations as:

$$Eff^{\text{Data/Simu.}}(\Omega_{Mis.}) = \frac{N_{Rec.}^{\text{Data/Simu.}}(\Omega_{Mis.})}{N_{Mis.}^{\text{Data/Simu.}}(\Omega_{Mis.})}, \tag{5.12}$$

where $N_{Mis.}^{\text{Data/Simu.}}(\Omega_{Mis.})$ is the number of events with a missing proton in the kinematic bin $\Omega_{Mis.} = P_{Mis.}; \theta_{Mis.}; \phi_{Mis.})$, $N_{Rec.}^{\text{Data/Simu.}}(\Omega_{Mis.})$ is the corresponding number of events with a detected proton.

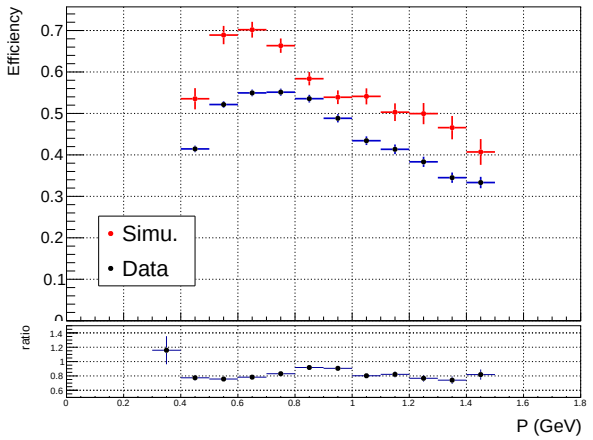
The proton efficiency correction is then encoded in the ratio:

$$Eff_{Corr} = \frac{Eff^{\text{Data}}}{Eff^{\text{Simu.}}}. \tag{5.13}$$

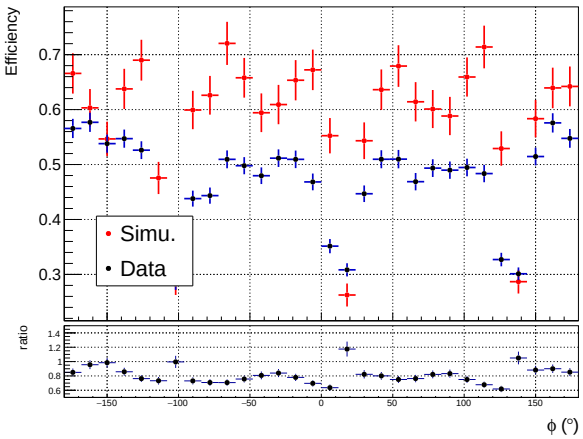
The correction is computed in the CD and in the FD independently, using similar procedures described in the next two subsections.

5.4.1 Efficiency correction in the central detector

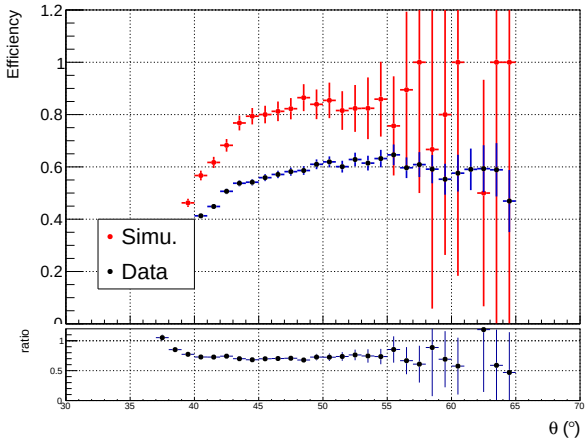
As shown in Figure 4.31b, there is very little background under the missing-proton mass peak in the high-polar-angle region. The number of events with a missing proton or a detected proton is then given by the number of events in each bin. The integrated efficiencies as a function of the momentum, the polar and the azimuthal angles of the missing proton are shown in Figure 5.6. The efficiency calculated in the simulations case is higher than for the data. The efficiency correction is calculated as a function of the three variables, with 2 bins in θ (from 37° to 45° and from 45° to 65°), 4 bins in momentum (spanning the 0.4 to 1.5 GeV range evenly) and 30 bins in ϕ (from −180° to 180°). The limits of the binning are driven by the variation of the correction as a function of each variable. The maximum value of the correction in the CD is 60%.



(a)



(b)



(c)

Figure 5.6: Proton efficiency in the CD, as a function of (a) the momentum, (b) the azimuthal angle and (c) the polar angle of the missing proton; for simulations in red and data in blue.

5.4.2 Efficiency correction in the forward detector

The proton efficiency is also derived for FD protons. Unlike in the case of the CD, where there is no background under the proton peak, there is a large high-mass background in the missing mass spectrum for protons at polar angles below 37°. In this region the number of events with a missing or reconstructed proton is calculated by fitting the missing proton peak with a gaussian plus a linear background as shown in Figure 5.7. The integral of the gaussian defines the number of events. The fit on the missing mass peak is done in both the missing-proton case and the reconstructed-proton case. In the case of the reconstructed proton, only the events with a reconstructed proton are kept in the missing mass spectrum.

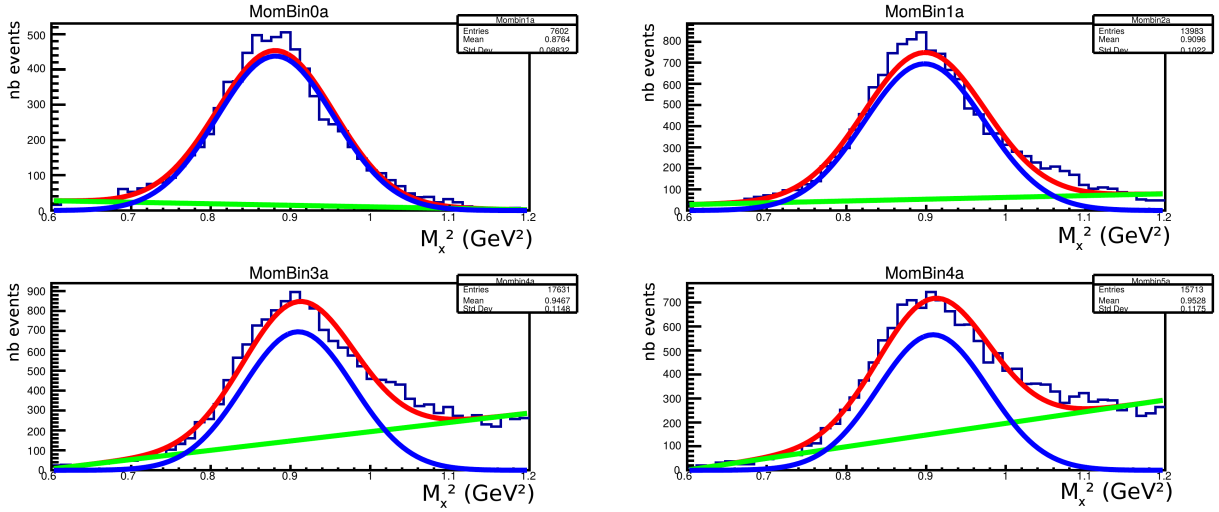


Figure 5.7: Examples of fits performed for the proton efficiency analysis in the FD. The missing mass peak is fitted with a gaussian and a linear background.

Figure 5.8 shows the efficiency as a function of the momentum, the azymuthal and the polar angles of the missing proton. As the ratios between simulations and data efficiencies are fairly constant in θ and ϕ, a single differential correction in momentum is applied according to the efficiencies shown in Figure 5.8a. The correction ranges from 80% at high momentum, around 1 GeV, and drops to 40% for low-momentum protons (around 0.4 GeV).

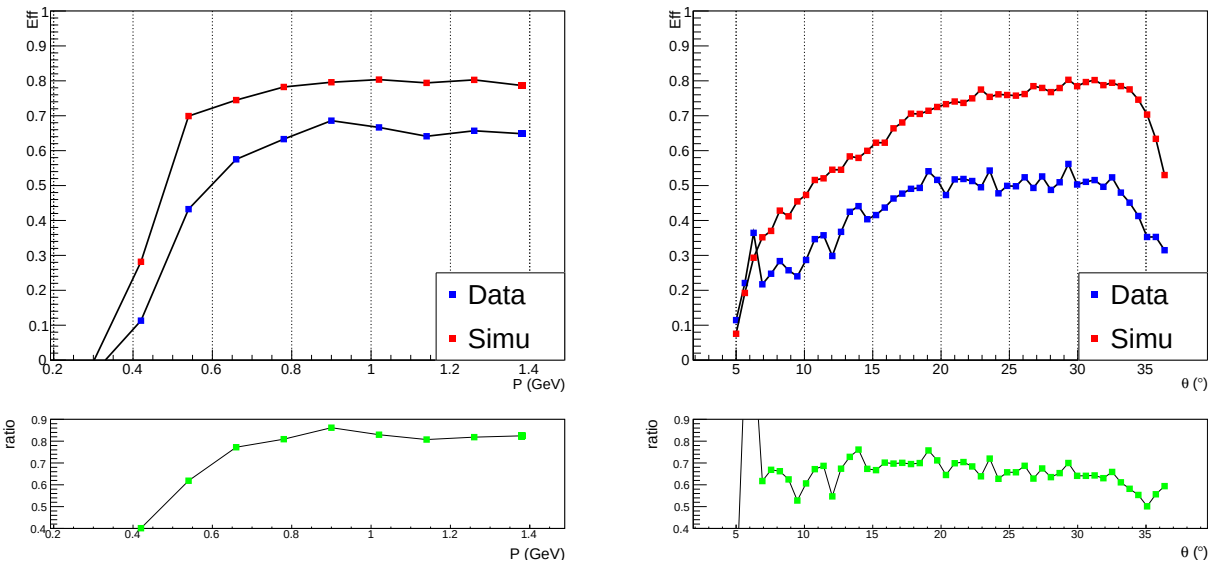
5.5 Background merging

The background merging [100] consists in mixing data events recorded with random trigger with simulation events. The random trigger events are recorded regularly (with a frequency of few hundreds hertz) during the data taking. For each simulation event, the ADCs and TDCs of the CTOF, FTOF, DCs, SVT, MVT, EC and HTCC from a random trigger event are added to the list of ADCs and TDCs obtained from the GEMC simulation. The merged events are then reconstructed with the standard reconstruction software. This background merging procedure aims at better reproducing the detection efficiencies in the simulation by mimicking the backgrounds present during the actual data taking.

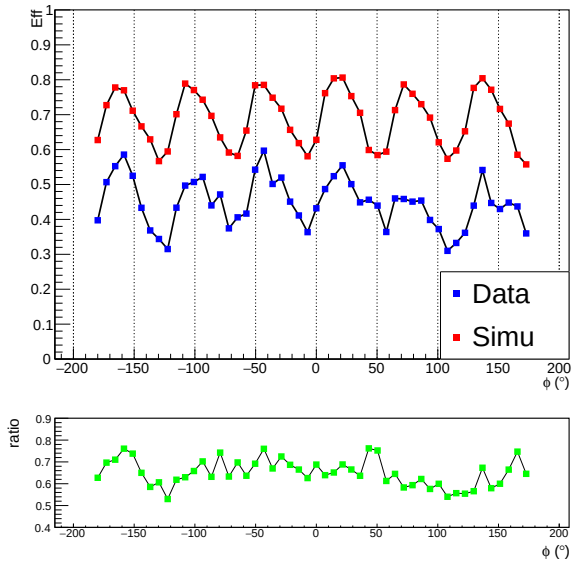
In the TCS analysis, the background merging procedure is used as an alternative to the previously presented proton efficiency correction. The acceptance (see Section 5.6) is calculated twice, first using the efficiency correction and without background merging and then using the background merging only. The difference between the results obtained with both methods is assigned as a systematic uncertainty, as explained in Section 5.13.

5.6 Acceptance estimation

The large angular coverage of CLAS12 permits to detect most of the particles produced when an electron interacts with the target. However, due to holes between detector subsystems, some particles



(a) (b)



(c)

Figure 5.8: Proton efficiency in the FD, as a function of (a) the momentum, (b) the azimuthal angle and (c) the polar angle of the missing proton; for simulations in red and data in blue.

may escape without being detected. Furthermore, because of detection inefficiencies, the reconstructed particle kinematics may differ from their initial values at the vertex. In order to correct for these effects, the acceptance of CLAS12 for the $\gamma p \to e^- e^+ p'$ reaction is estimated using simulations. The acceptance is calculated in 5-dimensional bins. In a given bin $\mathcal{B}$, the acceptance is defined as the number of events reconstructed in this bin devided by the number of events generated in this bin:

$$Acc_{\mathcal{B}} = \frac{N_{\mathcal{B}}^{REC}}{N_{\mathcal{B}}^{GEN}}. \tag{5.14}$$

The number of reconstructed events in $\mathcal{B}$ is:

$$N_{\mathcal{B}}^{REC} = \sum_{\{E_\gamma, -t, Q'^2, \theta, \phi\}_{REC} \in \mathcal{B}} Eff_{corr}\ w, \tag{5.15}$$

where the efficiency correction defined in Section 5.4 is applied, w is the weight of the event given by *TCSGen* (Equation (5.1)) and the sum is performed over all events with reconstructed kinematics inside the considered bin $\mathcal{B}$. The number of generated events in $\mathcal{B}$ is:

$$N_{\mathcal{B}}^{GEN} = \sum_{\{E_\gamma, -t, Q'^2, \theta, \phi\}_{GEN} \in \mathcal{B}} w, \tag{5.16}$$

with the same definition as before, except that the sum is now done over events with generated kinematics inside the bin $\mathcal{B}$. This definition encodes both acceptance and bin migration effects, provided that resolutions are well reproduced in simulations. Each event in the analysis is corrected by the acceptance factor corresponding to its reconstructed kinematics.

The binning used in the analysis is given in Table 5.1.

Variable	Bin limits
$-t$	0.15 - 0.25 - 0.34 - 0.48 - 0.8
E_γ	4.0 - 6.2 - 8.4 - 10.6
Q'^2	2.25 - 3.5 - 5.0 - 9.0
θ	from 30° to 160°, 10° bins
ϕ	from −180° to 180°, 10° bins

Table 5.1: Multi-dimensional binning used for the calculation of the acceptance.

For the results shown in Section 6.1, 20 million events were generated using *TCSGen*, passed through GEMC and reconstructed using the CLAS12 reconstruction software. In order to remove bins where there are too few events to perform the acceptance estimation, two additional cuts are used. The bins with acceptance bellow 5% are removed from the analysis. A cut on the relative error of the acceptance shown in Figure 5.9 is also applied:

$$\frac{\delta Acc}{Acc} < 50\%. \tag{5.17}$$

The resulting acceptance distributions are shown in Appendix M, as a function of θ and ϕ for each $\{E_\gamma, -t, Q'^2\}$ bin.

5.7 Background estimations

5.7.1 Electro-production of a lepton pair $ep \to pe^+(e^-)e_{scattered}$

In Section 5.2, where the exclusivity cuts are presented, the scattered electron is assumed to stay undetected at low angles. However it could be deflected at high angle and the electron from the pair could stay undetected at low angles, mimicking the TCS reaction. Also, because the final state of the reaction has effectively two electrons, interferences between them have to be investigated. These effects are included in the *Grape* generator which allows to investigate the effects of using quasi-real photons for the TCS measurement. Two samples of events were generated. The first sample contains events in which two electrons, the scattered one and the one from the pair, are generated. All possible pair-production channels are added: the BH channel (top plots in Figure 5.10), the pair production from radiated photons (bottom plots in Figure 5.10), as well as the interference between the two electrons involved in the reaction. In the second sample the primary electron is generated at very low angles and only the BH process is considered.

In the first sample including all possible effects, the two electrons can be distinguished by their transverse momentum. The electron with a high transverse momentum is likely to be detected and

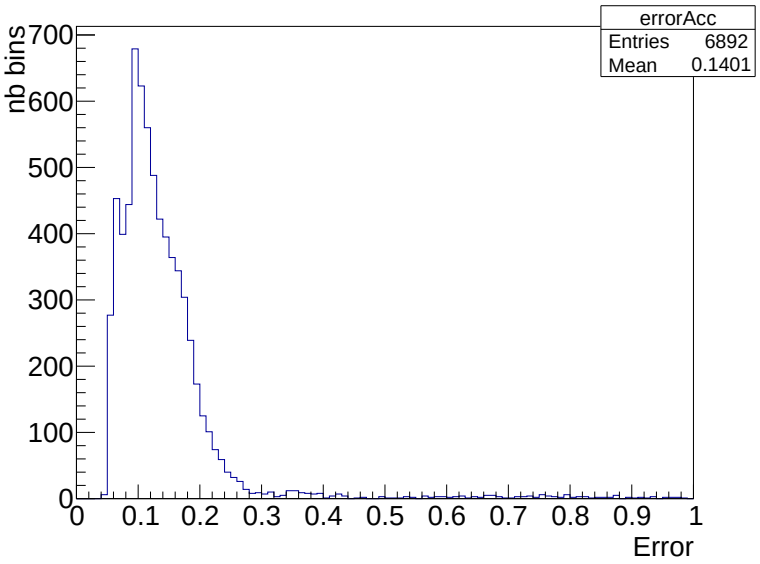
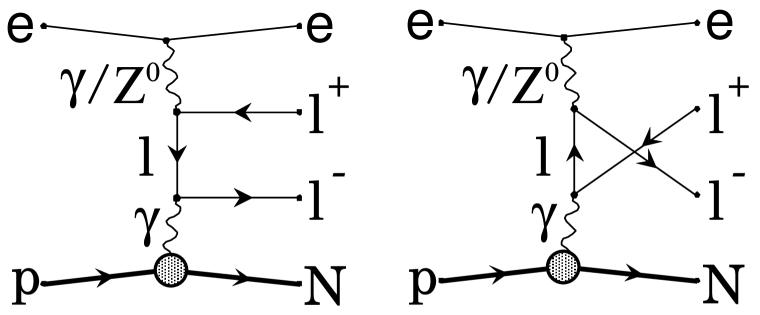
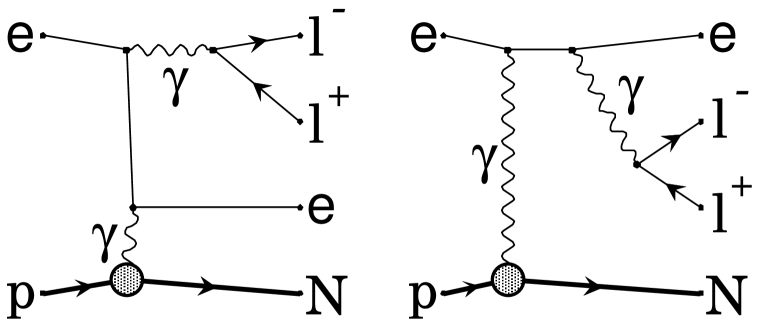


Figure 5.9: Relative error on the acceptance calculation, after bins with acceptance below 5% are removed. A cut is applied to remove bins with relative error higher than 50%.



(a) Bethe-Heitler type diagrams

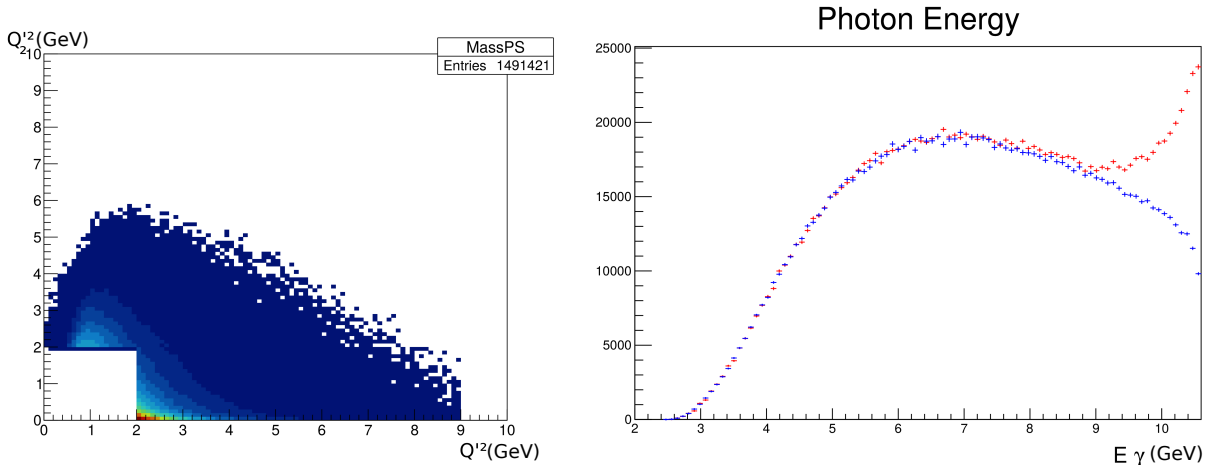


(b) QED-Compton type diagrams

Figure 5.10: Diagrams of the processes included in the *GRAPE* event generator. The top two plots (a) correspond to the BH processes associated with TCS where the incoming photon is radiated from an electron. The bottom two plots (b) represent the processes where a radiated photon emits a photon that then decays in a lepton pair.

identified as a "pair" electron, while the low transverse momentum electron is likely to stay undetected. Following this observation, two configurations for the measurement are possible (a positron and a low or high transverse momentum electron). Depending on which electron is detected, the reconstructed invariant mass of the lepton pair, as well as the calculated photon kinematics, are different. Figure

5.11a shows the distribution of generated events in the plane defined by the two possible invariant masses, Q'^2 calculated from the high transverse momentum electron and $Q_2'^2$ from the low transverse momentum electron. All the mass configurations that could lead to a reconstructed invariant mass in the range 2 GeV^2 – 9 GeV^2 are generated. Finally, Figure 5.11b shows the distribution of the photon energy calculated assuming that the high transverse momentum electron is the one from the lepton pair. The blue histogram corresponds to the BH-only case, the red one to the full simulation. Both simulations are normalized to have equal integrals between 2 GeV and 7 GeV.



(a) (b)

Figure 5.11: (a) Generated invariant mass phase space in the case where all possible processes leading to the final state $pe^+(e^-)e_{scattered}$ are considered. (b) Generated photon energy distributions for BH-only events (blue) and including all the possible processes (red).

The two generated samples are then passed through the CLAS12 simulation chain and exclusivity cuts are applied. The resulting measured distributions are shown in Figure 5.12. No large effect coming from the use of quasi-real photo-production is noticed.

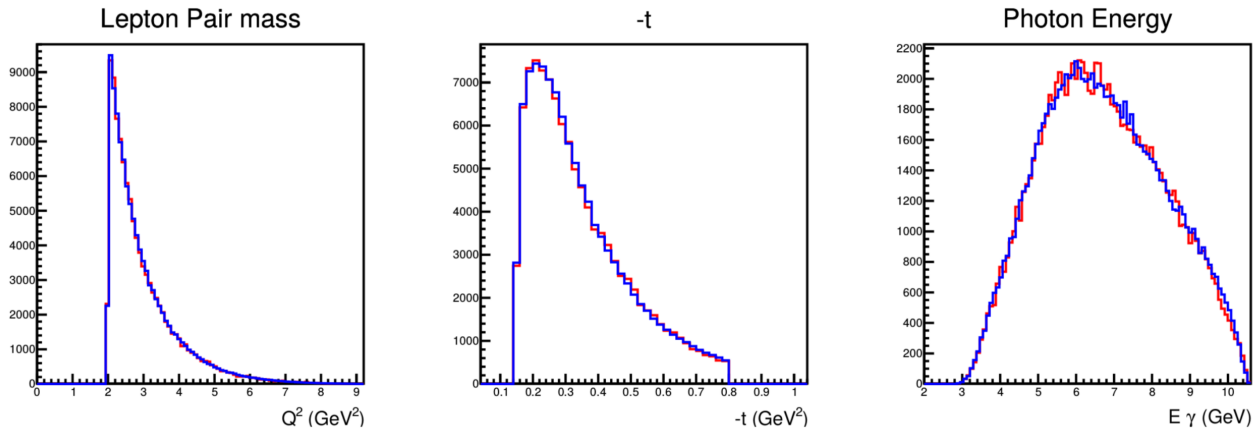


Figure 5.12: Measured invariant mass Q'^2 (left), proton transferred momentum $-t$ (middle) and photon energy E_γ (right) distributions, for the generated BH-only events (blue) and with all the processes included (red).

5.7.2 Pion contamination

Pions can be mistakenly identified as leptons, especially at high momenta (above $\sim$ 4.5 GeV) where the HTCC produces signals for leptons and pions. The pion background is addressed in Chapter 4. The neural network PID is implemented in the TCS analysis with a cut at 0.5. We show in Chapter

4 that this method reduces the background ratio by a factor 10, to less than 5%. To account for the remaining background and to estimate its impact on the measured observables, the background rejection cut is varied. This latter point is addressed in the Section 5.13.

5.8 Experimental cross-section ratio

The theoretical R ratio is calculated over the full angular phase space of TCS. However CLAS12 acceptance is limited and only covers a fraction of the angular phase space. The theoretical R ratio is thus inaccessible, but it can be calculated over the CLAS12 acceptance. Following the notations in [59], we call R' the ratio integrated over the CLAS12 acceptance. It is calculated as:

$$R' = \frac{\sum_\phi Y_\phi \ \cos(\phi)}{\sum_\phi Y_\phi}, \tag{5.18}$$

where the sum over ϕ is done in 10° bins and the $\cos(\phi)$ factor is calculated at the center of each bin. The Y_ϕ quantity is calculated for each ϕ bin as:

$$Y_\phi = \sum_{\text{events in } \phi \text{ bin}} \left(\frac{L}{L_0}\right)\left(\frac{1}{Acc}\right), \tag{5.19}$$

where the ratio $\frac{L}{L_0}$ is calculated event-by-event (the factors L and L_0 are given in Equations (1.73) and (1.77), respectively), and Acc is the acceptance in the kinematic bin of each event (given in Equation (5.14)). The statistical error of Y_ϕ is calculated as:

$$E^2(Y_\phi) = \sum_{\text{events in } \phi \text{ bin}} \left(\frac{L}{L_0}\right)^2\left(\frac{1}{Acc}\right)^2, \tag{5.20}$$

Statistical uncertainty using Monte Carlo method

As the numerator and the denominator used in the computation of R' are correlated, a Monte Carlo approach is used to compute the statistical uncertainty. It works as follows:

- Y_ϕ and $E(Y_\phi)$ are computed as described above,
- for each ϕ bin, a value Y_ϕ^R is randomly generated following a gaussian probability distribution of mean Y_ϕ and sigma $E(Y_\phi)$,
- the ratio is calculated using the randomly generated values: $R^R = \frac{\sum_\phi Y_\phi^R \ \cos(\phi)}{\sum_\phi Y_\phi^R}$.
- The previous two steps are repeated 10000 times and the resulting distribution of R^R is fitted with a gaussian,
- the σ of the fitted gaussian is defined as the statistical uncertainty of the measurement. An example of the fitted distribution is shown in Figure 5.13.

5.9 Phenomenological study of the TCS Forward-Backward asymmetry

As mentioned in Section 1.3 where the concept of Forward-Backward asymmetry (A_{FB}) is introduced, no predictions have yet been published for this observable in TCS. In this section the phenomenological work performed on the TCS A_{FB} is presented. The main goal of this analysis is to estimate the size of the asymmetry as well as to gain an insight on its kinematic dependencies in order to determine if CLAS12 can measure it.

This section is divided in three parts. In Subsection 5.9.1, the limitations of CLAS12 acceptance and the consequences on the measurement of the A_{FB} are presented. Initial results are compared to other predictions provided by independent groups and models. The results obtained according to the conclusion of the first section are then displayed in subsequent Subsections 5.9.2 and 5.9.3.

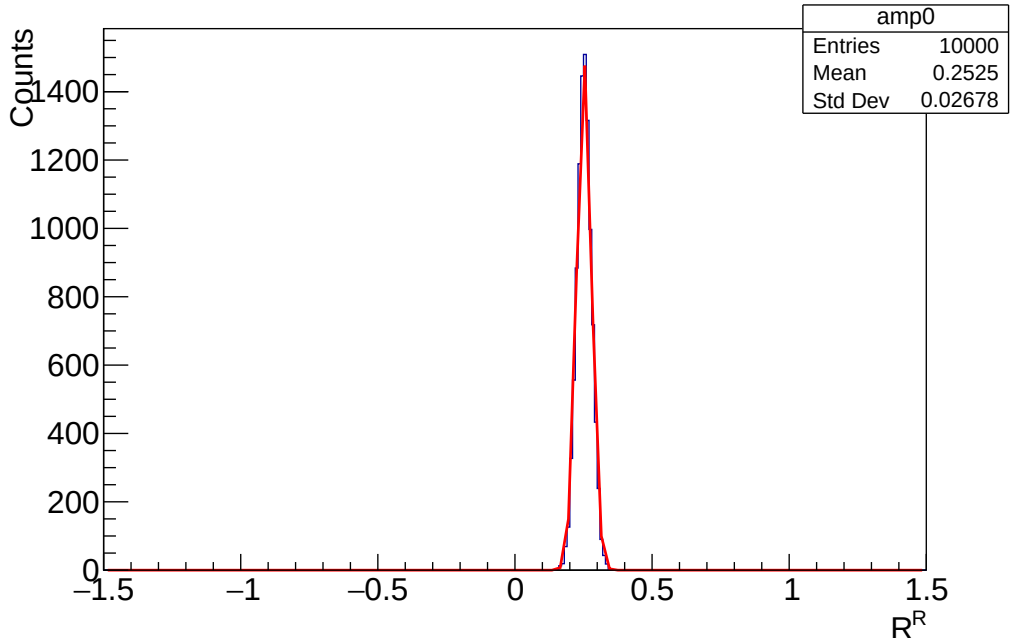


Figure 5.13: Distribution of R^R for the determination of the statistical error for first bin in ξ. The full results are shown in Figure 6.2. The gaussian fit is overlaid.

5.9.1 Early considerations and comparison with other models

To study the TCS A_{FB}, the TCS and BH processes and their interference cross sections have been calculated using the VGG code [39] provided by Michel Guidal. The formulas used in this code are described in [58]. The cross section is estimated in the forward and backward directions and the obtained results are then combined to produce the asymmetry curves shown in the following. The GPD model used is the VGG model described in [22] and in Section 1.2 of this manuscript. The D-term contribution to the GPD H is included (unless mentioned otherwise). The α' coefficient fixing the t-dependence of the GPDs is set to 1.098, the b_{val} and b_{sea} parameters fixing the ξ dependence are set to 1 (unless specified otherwise).

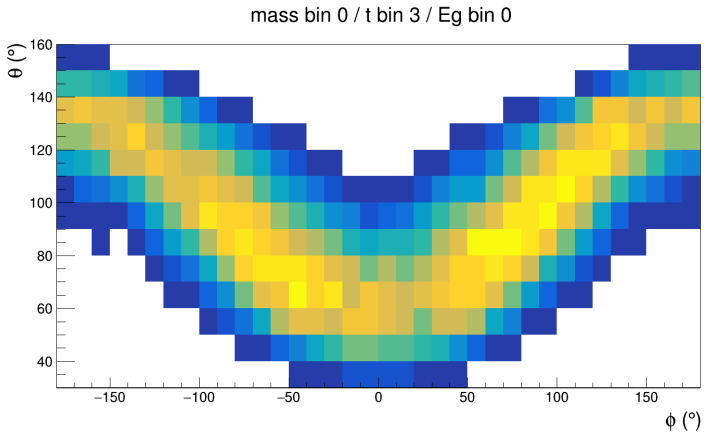
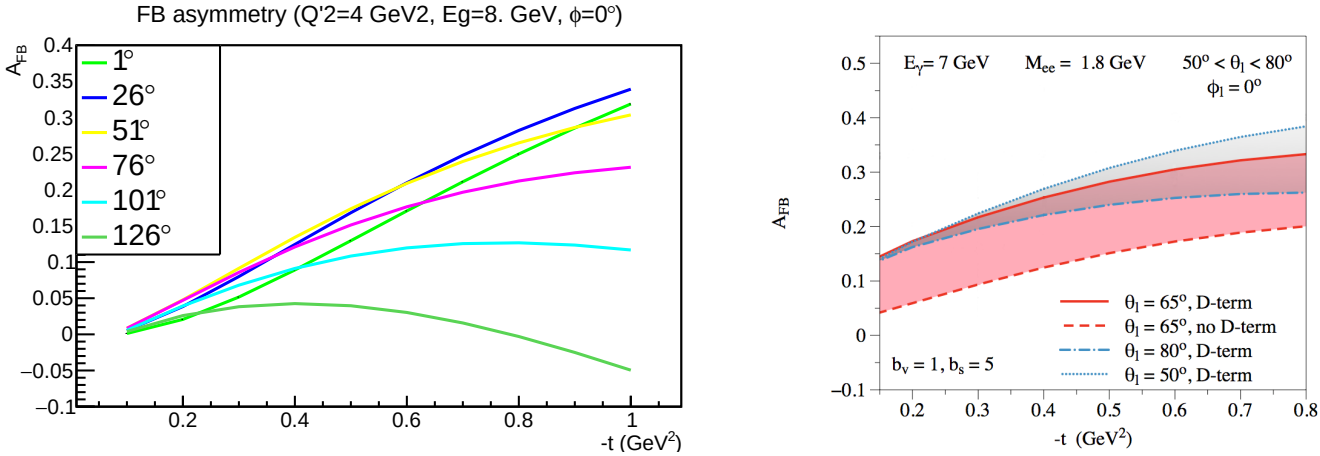


Figure 5.14: Acceptance of CLAS12 in the θ/ϕ plane. The region around $\phi = 0°$ and low polar angle, as well as $\phi = \pm 180°$ and high polar angle are well covered by CLAS12.

The CLAS12 acceptance does not cover the full ϕ - θ plane of TCS. An example of the acceptance coverage, obtained from the CLAS12 acceptance study presented in Section 5.6, is shown in Figure 5.14. The regions around $\phi = 0°$ and low θ, and around $\phi = \pm 180°$ and high θ are well covered by CLAS12. Following this observation, it is decided to study the FB asymmetry at $\phi_0 = 0°$.

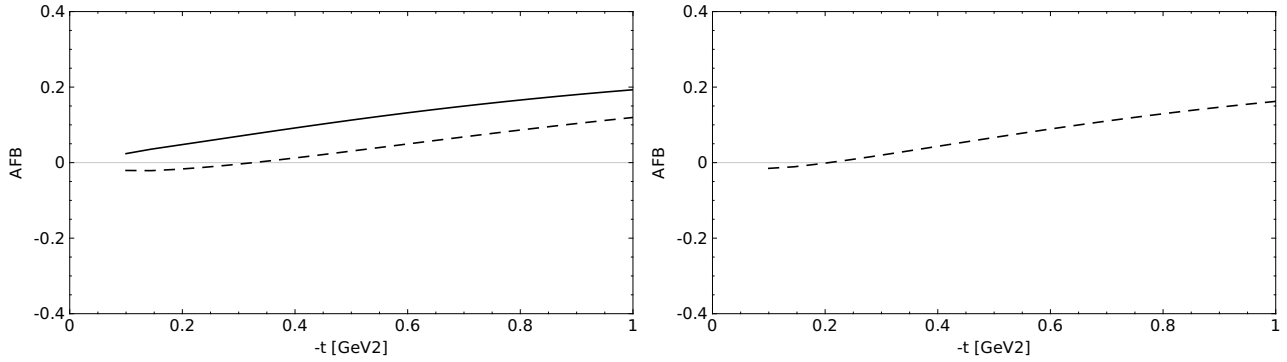
A prediction for the asymmetry is shown in Figure 5.15a, where the t-dependence of the asymmetry is plotted for various values of θ_0. These results were cross checked independently by M.Vanderhaeghen also using the VGG code. Figure 5.15b shows the results of this independent analysis. Both results are pointing toward an asymmetry with a positive value. Note that these results were obtained after correcting a missing minus sign in Equation (17) of [58]. The consequence of adding this minus sign is to flip the sign of the asymmetry.



(a) (b)

Figure 5.15: (a) FB asymmetry as a function of $-t$ at fixed $\phi_0 = 0°$ for different values of θ_0. (b) Asymmetry as a function of $-t$ at fixed $\phi_0 = 0°$ for different values of θ_0 and different models: the impact of the D-term on the asymmetry is shown. Figure courtesy of M. Vanderhaeghen.

Later predictions realized by Pawel Sznajder [101], using the *PARTONS* software [102], are shown in Figure 5.16. These predictions also point to a positive asymmetry, independently of the model used for the CFFs as well as high order corrections.



(a) (b)

Figure 5.16: FB asymmetry as a function of $-t$ at fixed $\phi_0 = 0°$, $\theta_0 = 60°$, $E_\gamma = 7$ GeV and $Q'^2 = 1.8$ GeV ^{2} (a) using the GK model at LO (dashed) and NLO (plain), (b) using the VGG model at LO. Figures courtesy of Pawel Sznajder.

Furthermore, the analytical formulae for TCS derived in [57] were used to cross-check the asymmetry in VGG. This was done after making sure the CFF conventions are consistent. The results obtained for the analytical formulae are shown in Figure 5.17. The sign of the asymmetry obtained is consistent with the *PARTONS* results and with the corrected VGG asymmetries.

A final consistency check was performed by calculating the asymmetry from the TCS only and BH only cross sections. The absence of asymmetry in both cases is well verified.

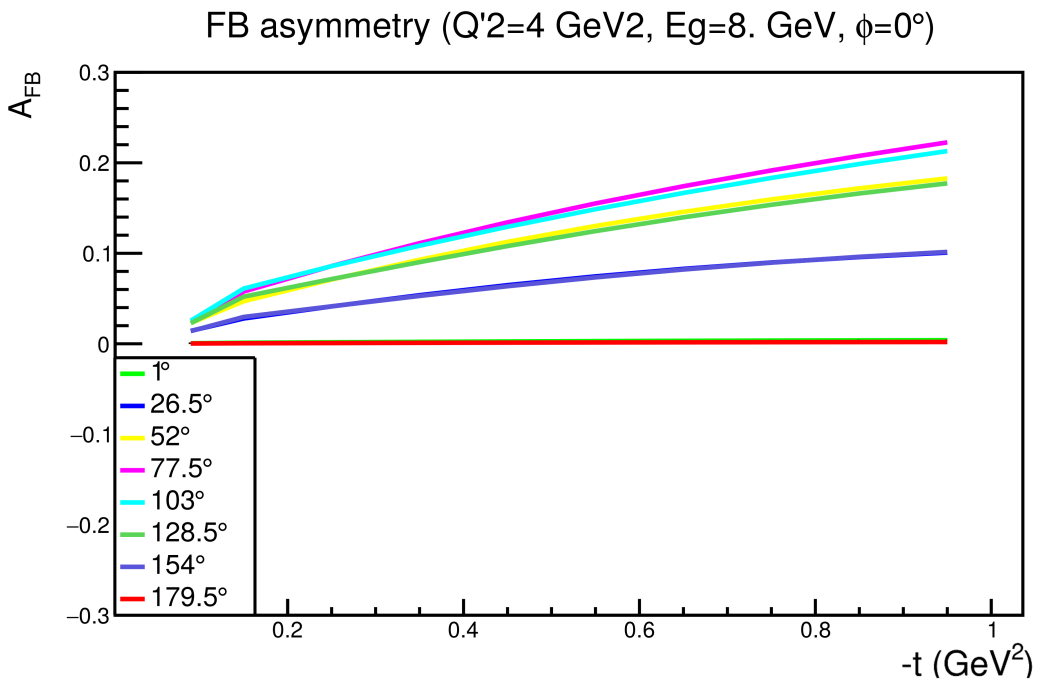


Figure 5.17: FB asymmetry as a function of $-t$ at fixed $\phi_0 = 0°$ for different values of θ_0 using formulas derived in [57].

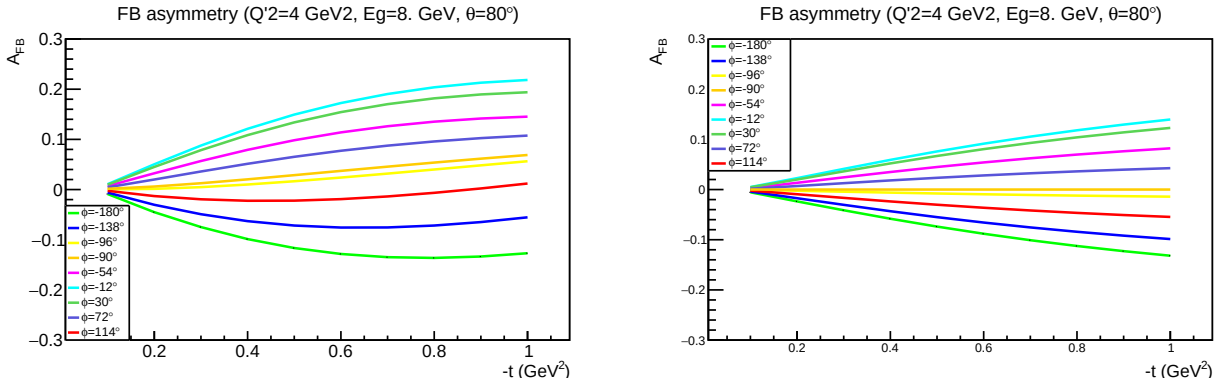
5.9.2 TCS A_{FB} kinematic dependencies

A main feature of the FB asymmetry is that it can be measured over a small portion of the TCS angular phase space. This prevents any large detector-induced false asymmetry caused by holes in the acceptance, which instead occurs in the case of the R ratio (see Section 5.8). On the contrary, the statistics dramatically falls if the angular phase space of the measurement is too narrow. The main goal of this analysis is to determine the angular range that can be used for the measurement in CLAS12, in order to maximize statistics. One has to make sure that the FB asymmetry conserves its sign over the angular integration domain to maximize the size of the measured asymmetry. Therefore, the angular dependence of the FB asymmetry is studied. From the explicit expression of the A_{FB} derived in Section 1.3 following [57]:

$$A_{FB}(\theta_0, \phi_0) = \frac{-\frac{\alpha_{em}^3}{4\pi s^2}\frac{1}{-t}\frac{m_p}{Q'}\frac{1}{\tau\sqrt{1-\tau}}\frac{L_0}{L}\cos\phi_0\frac{(1+\cos^2\theta_0)}{\sin(\theta_0)}\mathrm{Re}\tilde{M}^{--}}{d\sigma_{BH}}, \tag{5.21}$$

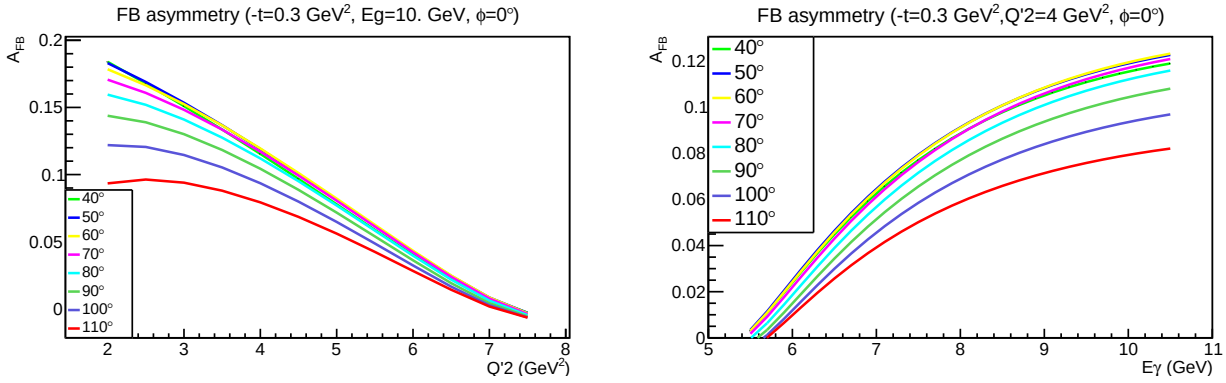
one can see that the ϕ_0 dependence is driven by the $\cos(\phi_0)$ factor. The ϕ_0 behavior is shown in Figure 5.18 for fixed $\theta = 80°$. The asymmetry changes sign at $\phi_0 \approx \pm 90°$. This behavior is reproduced by both the VGG model and the analytical model. The differences are attributed to the terms that are neglected in the analytical model (see [57] for full details). Following this investigation, one can see that the FB asymmetry in the region around $\phi_0 = 0°$ is maximal and that the sign is constant in its vicinity. In Figure 5.15a, showing the asymmetry as a function of $-t$ at $\phi_0 = 0°$ and for different values of θ_0, the asymmetry has constant sign over the θ_0 range accessible in the CLAS12 acceptance (from $\sim 40°$ to $\sim 100°$).

Finally, the asymmetry was studied as a function of the incoming photon energy E_γ and the square of the invariant mass of the lepton pair Q'^2, aiming at increasing the statistics available for the measurement. Figure 5.19 shows the Q'^2 and E_γ dependencies of the FB asymmetry at $\phi_0 = 0°$ and for various values of θ_0. These plots indicate that the sign of the asymmetry is constant over the kinematic range accessible by CLAS12. This implies that it is possible to integrate the measurement over a large portion of the phase space and still measure an asymmetry.



(a) (b)

Figure 5.18: FB asymmetry as a function of $-t$ at fixed $\theta_0 = 80°$ for different values of ϕ_0 computed (a) with the VGG code, (b) using the formulas derived in [57].



(a) (b)

Figure 5.19: FB asymmetry calculated using the VGG code (a) as a function of Q'^2 and (b) as a function of E_γ at fixed $\phi_0 = 0°$ and for different values of θ_0.

5.9.3 TCS A_{FB} model dependencies

After exploring the kinematic dependencies of the FB asymmetry, its dependencies on GPD model parameters are tested.

D-term dependence

The first model dependence investigated is the D-term. The FB asymmetry is sensitive to the real part of the CFFs and thus to the D-term. This makes this observable a good candidate to investigate its strength. The FB asymmetry is calculated with and without the D-term contribution. The D-term used is the one presented in [103]. The effect of adding the D-term to the GPDs parametrization is shown in Figure 5.20. One can see a sizable effect on the asymmetry induced by the D-term. It is also important to notice that the amplitude of the asymmetry increases with the addition of the D-term, and that the sign of the asymmetry does not change when adding it.

b_{sea}-dependence

The dependence of the asymmetry to the skewness parameter b_{sea} is also explored. Indeed in [28] it is suggested that DVCS data, sensitive to the imaginary part of the $\mathcal{H}$, are better described with skewness parameters $b_{val} = 1$ and $b_{sea} = 5$. Following this observation, the value of the sea parameter is varied in the asymmetry calculation, from 1 (strong skewness dependence) to 8 (low skewness dependence), to verify if the FB asymmetry can help in the determination of this parameter. The results are shown

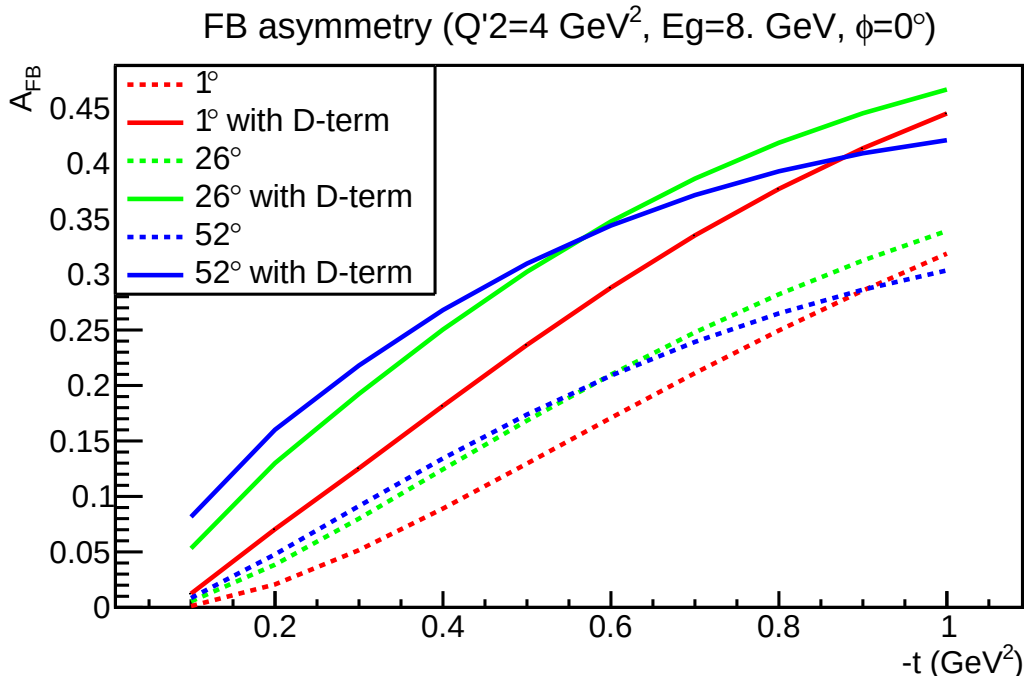
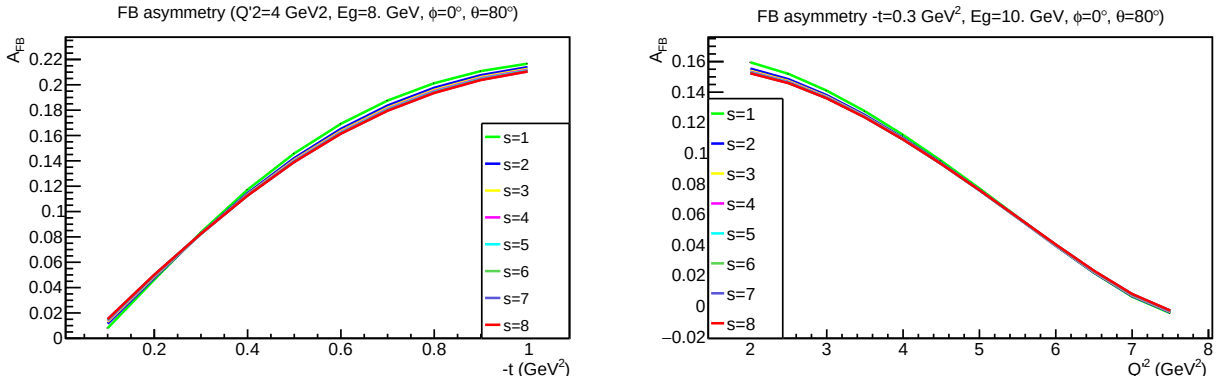


Figure 5.20: FB asymmetry as a function of $-t$ at $\phi_0 = 0°$ for different values of θ; and with (solid lines) or without (dashed lines) the D-term contributions to the GPD H.

in Figure 5.21, where the asymmetry dependencies on $-t$ and Q'^2 are plotted for various values of the sea parameter. No large effect is seen when varying the parameter b_{sea}.



(a) (b)

Figure 5.21: (a) $-t$ dependence of A_{FB} for different values of the b_{sea} parameter. (b) Q'^2 dependence of A_{FB} for different values of the b_{sea} parameter.

5.10 Experimental Forward-Backward asymmetry

According to the conclusions of the phenomenological study in Section 5.9, the FB asymmetry is calculated at $\phi_0 = 0°$. Furthermore, by looking at the angular coverage of CLAS12, it is decided to integrate over the following angular bin in the forward direction:

- $-40° < \phi_0 < 40°$
- $50° < \theta_0 < 80°$.

The corresponding backward bin limits are:

- $140° < 180° + \phi_0 < 220°$
- $100° < 180° - \theta_0 < 130°$.

Bin volume correction

Figure 5.22 shows the acceptance of CLAS12 in the θ - ϕ plane for a given $-t$, E_γ and Q'^2 bin. The limits of the forward and backward bins are shown in green and red, respectively. Although the limits of the angular bins have been chosen to be covered by the CLAS12 acceptance, some $\{E_\gamma, -t, Q'^2\}$ acceptance bins do not fully cover the forward or backward angular bins. The difference in coverage between the forward and the backward direction can be the source of fake asymmetries. To correct for this, a bin volume correction is applied during the calculation of the FB asymmetry. For each acceptance bin $\{E_\gamma, -t, Q'^2\}$, the fraction of the angular bins covered by the acceptance, $CorrVol_{F/B}$, is calculated. Each event detected in the forward (resp. backward) bin is then assigned a correction weight equal to the inverse of the fraction of the volume covered by the acceptance in this bin. This correction assumes that the cross section of the TCS reaction is relatively constant within the volume of the forward (resp. backward) bin and that it can be estimated only by measuring it in the volume covered by the acceptance of CLAS12. This hypothesis is verified by extracting the FB asymmetry with BH-weighted simulation events and the difference between the expected value for BH (null asymmetry) and the extracted value is assigned as a systematic uncertainty (see Section 5.13). Two sets of volume correction coefficients are obtained, one for the forward and one for the backward angular bins.

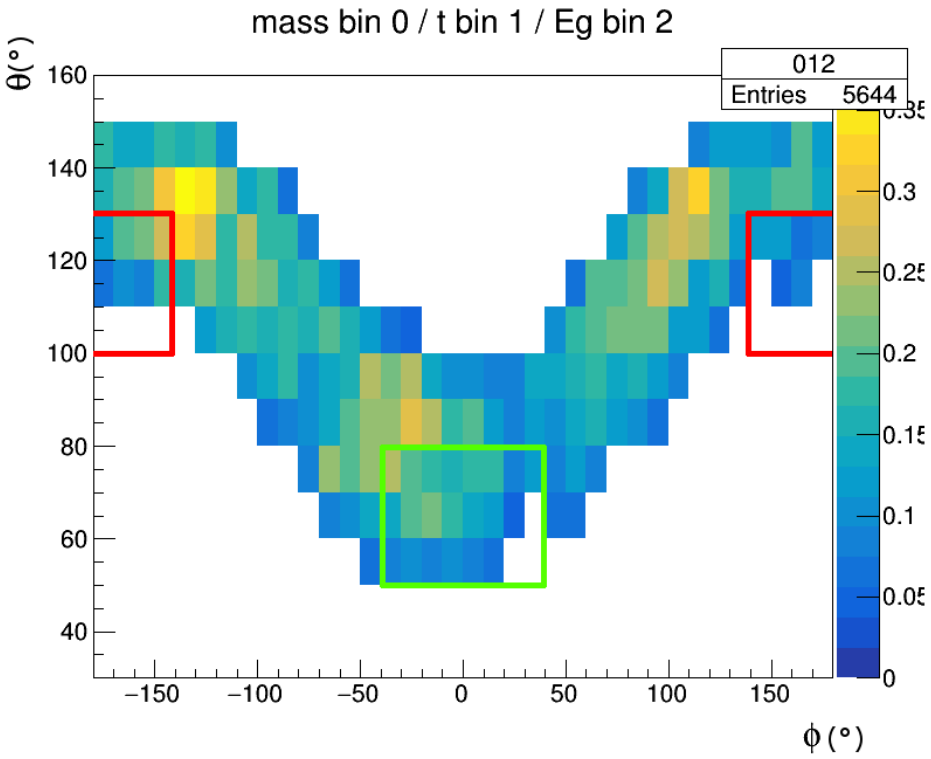


Figure 5.22: Volume correction for the A_{FB} calculation. The forward bin is represented by the green square. The red square delimits the backward bin. One can see that CLAS12 does not fully cover the angular bins for some $\{E_\gamma, -t, Q'^2\}$ bins.

For a given bin, the value of the FB asymmetry is calculated as the ratio:

$$A_{FB} = \frac{N_F - N_B}{N_F + N_B} \tag{5.22}$$

where $N_{F/B}$ is the number of events in the forward (resp. backward) angular bin, corrected by the acceptance and the bin volume, as:

$$N_{F/B} = \sum \frac{1}{Acc \times CorrVol_{F/B}}. \tag{5.23}$$

The reported statistical error bars are calculated by propagating the weighted error on $N_{F/B}$:

$$E^2(N_{F/B}) = \sum \left(\frac{1}{Acc \times CorrVol_{F/B}} \right)^2. \tag{5.24}$$

5.11 Experimental beam-spin asymmetry

As mentioned previously, this analysis is done on quasi-real photoproduction events, where a real photon is radiated by the initial electron beam. In this configuration, the circular polarization of the photon can be inferred from the initial longitudinal polarization of the electron beam. An electron polarized in the direction (resp. opposite) of the beam emits a right-(resp. left-) handed circularly polarized photon, with a polarization transfer $Pol_{transf.}$ fully calculable analytically in QED (see [104, 105] for details and Appendix N for the detailed formulas used to calculate the polarization transfer).

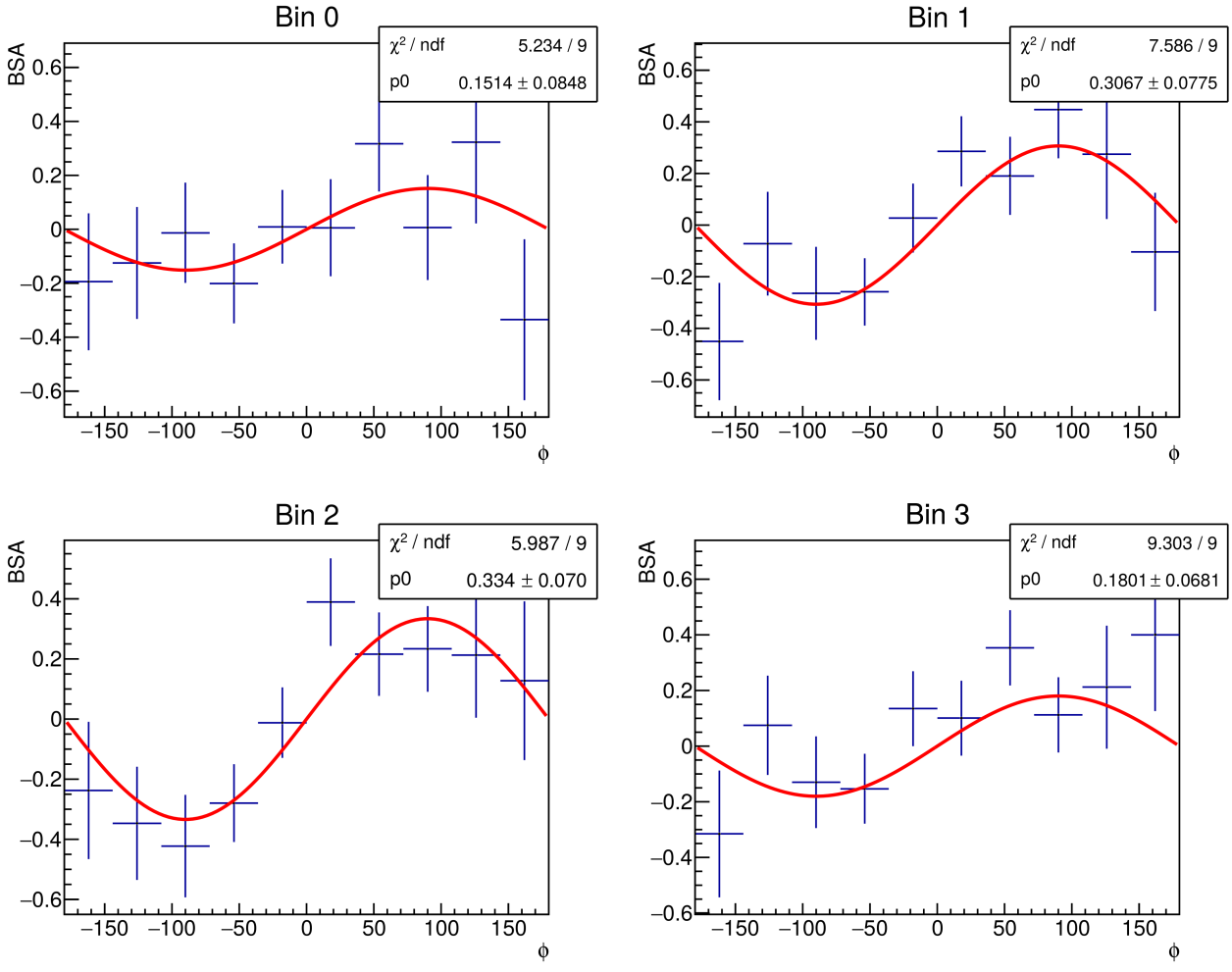


Figure 5.23: Distributions of the photon helicity asymmetry as a function of ϕ for four bins in $-t$ (as described in Section 5.12). The fit used to extract the asymmetry is overlaid. The fit amplitudes are displayed in Figure 6.11.

The photon polarization asymmetry is calculated as a function of the azymuthal angle ϕ as:

$$BSA(-t, E\gamma, M; \phi) = \frac{1}{Pol_{eff}} \frac{N^+ - N^-}{N^+ + N^-}, \tag{5.25}$$

where the number of events with reported positive N^+ (resp. negative N^-) electron helicity in each bin is corrected for the acceptance and the polarization transfer as:

$$N^{\pm} = \sum \frac{1}{Acc} Pol_{transf.}. \tag{5.26}$$

Pol_{eff} is the effective polarization of the CEBAF electron beam, which is estimated to be 85% on average. The ϕ-distribution is then fitted with a sinus function:

$$BSA(-t, E\gamma, M; \phi) = BSA(-t, E\gamma, M) \sin(\phi), \tag{5.27}$$

as shown in Figure 5.23. Following the definition of [58], the amplitude of the sinus function given by the fit (BSA at $\phi = \pi/2$) is extracted. The reported statistical error bars are given by the fit error on the amplitude of the function.

5.12 Binning of the data

The phase space used for the analysis is binned to have similar number of events in each bin for the t, E_γ and ξ variables. This aims at achieving similar error bars in each bins. There are four bins in $-t$, three in E_γ and ξ. For the lepton invariant mass, the ranges from 1.5 GeV to 2 GeV and from 2 GeV to 3 GeV are divided in two bins each, in order to study the possible effect of vector-meson contamination in the low-mass bin. Besides, the binning in $-t$ is different in the two mass ranges. The bin limits are summarized in Table 5.2 and superimposed on the kinematic distributions of the data in Figure 5.24.

Variable	Bin limits
$-t$ (1.5 GeV $< M <$ 2 GeV)	0.15 - 0.25 - 0.34 - 0.48 - 0.8
$-t$ (2 GeV $< M <$ 3 GeV)	0.15 - 0.35 - 0.45 - 0.55 - 0.8
E_γ	4.0 - 6.4 - 8.0 - 10.6
$M = \sqrt{Q'^2}$	1.5 - 1.7 - 2.0 - 2.5 - 3.0
ξ	0.0 - 0.12 - 0.15 - 0.4

Table 5.2: The binning grid used for the data in this analysis. Two different binnings for $-t$ are used, one for each invariant mass range.

5.13 Systematic errors estimation

In this section the methods used to estimate the systematic errors are described. Five different sources of systematic uncertainties have been studied. For each source of uncertainty, a value of systematic shift is calculated for every observable and for each bin. The errors are then added in quadrature.

Method

The method used to calculate the observable involves binned acceptance corrections and bin volume corrections for the FB asymmetry. To estimate the impact of these corrections on the extracted values, the method systematic error is computed using simulations. First the observables are calculated using a sample of generated BH-weighted events, without passing them through GEMC and with no corrections applied. This corresponds to the green point in Figure 5.25a. In the case of the R' ratio, only events within a non-zero acceptance bin are kept. In the case of the FB asymmetry, the observable is calculated

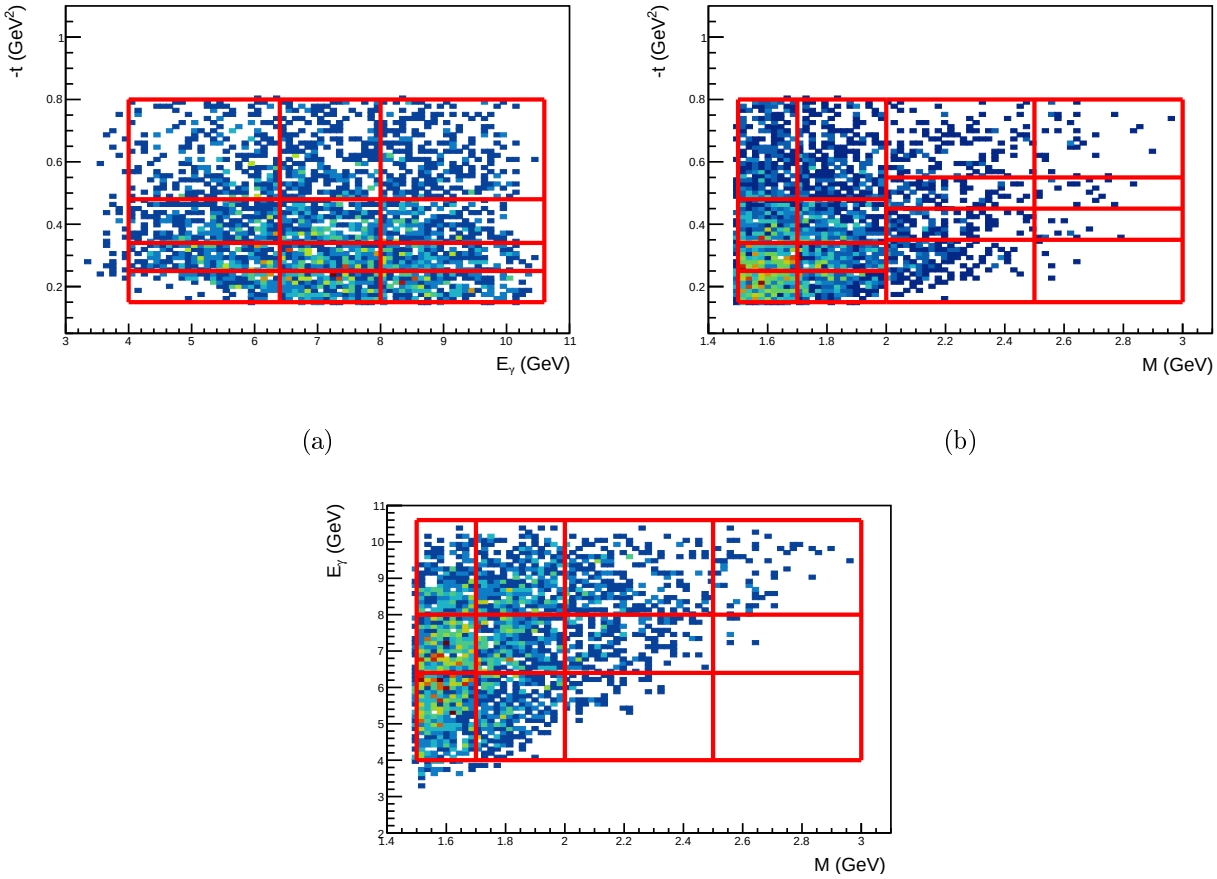


Figure 5.24: Binning grid for the TCS observables in (a) the E_γ - $(-t)$ plane (the binning in $-t$ used in this figure is the one used for the [1.5 GeV – 2 GeV] mass range), (b) the M - $(-t)$ plane and (c) the M - E_γ plane.

within the full angular bin defined in Section 5.10. Finally, in the case of the BSA, a random beam helicity is associated to each event. Then the observables are calculated a second time with the full-chain simulation events, applying all the corrections, except that the acceptance is not corrected for efficiency and data-driven momentum corrections are not applied. Two samples of simulation events are used, one using weighted events from *TCSGen* and one generated with *GRAPE* (blue points and red points on Figure 5.25a). The systematic uncertainty associated with the extraction method is the difference between the ideal case (no corrections) and the value extracted after the full analysis procedure. The systematic error is set between 0 and the difference between the ideal case and the extracted case for the two simulation samples. This procedure can thus result in an asymmetric error in order to reflect the shift induced by the measurement method. This systematics mostly affects observables binned in ξ (for example the R' ratio in Figure 6.2), as the acceptance is not binned in this variable. The maximum error induced on the extracted observables is at most 0.1 for the A_{FB} binned in ξ in Figure 6.7.

Efficiency / Background merging

A method to estimate the proton efficiency is described in Section 5.4. A second method, the background merging, used to reproduce the detection efficiencies is presented in Section 5.5. The systematic error associated with these corrections is estimated by the difference of the values of the observables obtained with both methods, Δ_{Eff}. The systematic error bar is then defined as $\pm\Delta_{Eff}/2$. Figure 5.25b illustrates the determination of the efficiency systematic error. As shown in Section 6.1,

the efficiency systematics is most of the time dominated by other sources of systematic errors for most of the observables. The maximum observed shift on the extracted observables is 0.1 for the A_{FB} as a function of $-t$ (Figure 6.9) and the BSA as a function of M (Figure 6.10).

Positron identification

In this analysis the positron identification algorithm plays an important role. To estimate the impact of the remaining pion contamination, the cut applied on the output of the neural network is varied around the chosen value, 0.5 ± 0.1. The acceptance is also recalculated accordingly. The difference between the observables extracted with the standard and the shifted cuts is assigned as the positron identification systematics. This systematics can be asymmetric as the variation of the extracted observables with the shifted positron cuts can be different in each case. As illustrated in Figure 5.25c, the variation is small for most of the observables. The induced shift is at most 0.1 for the A_{FB} as a function of $-t$ in the high-mass region in Figure 6.9, but usually limited to 0.03 for most of the observables.

Acceptance model

The dependence of the extracted results on the model used in the acceptance is also studied. The acceptance is calculated with BH-weigthed events and events without weights (i.e. phase-space generator). The difference between the two methods Δ_{Acc} is defined as the acceptance model systematics, and the associated error bar is set to $\pm\Delta_{Acc}/2$. The method is illustrated in Figure 5.25d. This source of systematics becomes large for observables with low statistics, such as the A_{FB}, for which the induced absolute shift can reach values up to 0.05 (as in Figure 6.4).

Exclusivity cuts

Finally, the last source of systematic uncertainty studied is the values chosen for the exclusivity cuts. To estimate this uncertainty, the analysis is performed with tighter cuts than those presented in Section 5.2:

$$\left|\frac{Pt_X}{P_X}\right| < 0.04 \tag{5.28}$$

and

$$|M_X^2| < 0.3 \text{ GeV}^2. \tag{5.29}$$

The difference between the two extractions Δ_{Exclu} is assigned as the exclusivity-cuts systematic error and the associated error bars are symmetric around zero $\pm\Delta_{Exclu}/2$, as shown in Figure 5.25e. This systematics dominates in bins where the statistics is limited, especially for the A_{FB} and the BSA at high mass in Figure 6.10. This shows that the exclusivity cut error is largely correlated with the statistical error in these bins. However, the observed systematic shift always remains within the statistical error bars.

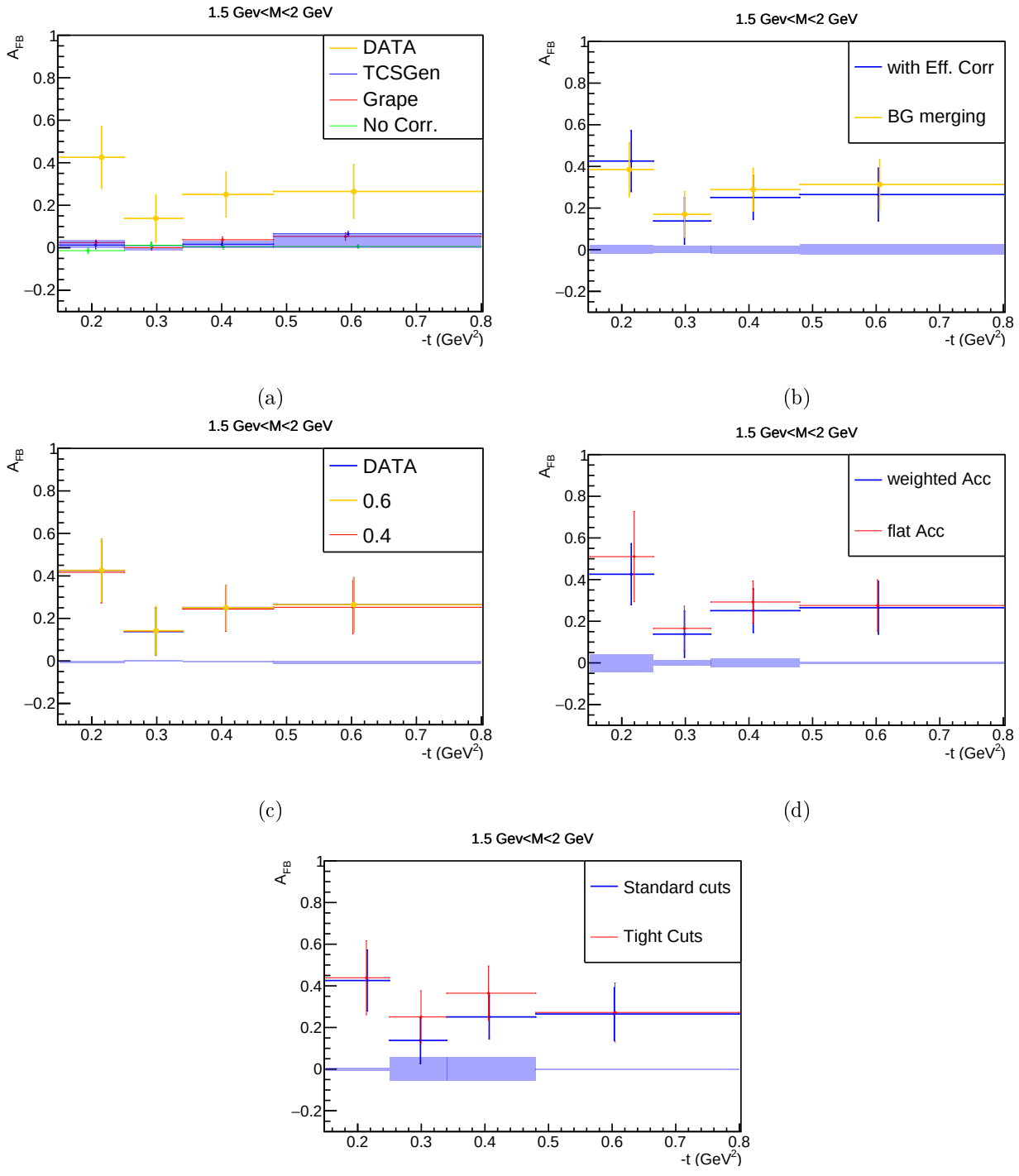


Figure 5.25: Systematic errors for the FB asymmetry in the 1.5 GeV $< M <$ 2 GeV mass range, as blue bands with respect to a reference value of 0. The vertical error bars correspond to the statistical errors and horizontal error bars expand along the bin size. The same study is done for all the observables extracted in this analysis. (a) Systematic error from the extraction method. The green points are calculated from generated events, the blue and the red points are obtained after the full chain simulation and analysis. The data points are also displayed in yellow in order to compare this systematics with the statistical error bars. (b) Efficiency systematic error obtained from the values of the observables with the proton efficiency correction (blue) and the background merging (yellow). (c) Positron ID systematic error. The cut on the output of the neural network is varied from 0.5 (blue) by ± 0.1 (red and yellow) to obtain the value of this error. (d) Acceptance model systematics obtained by calculating the observable with a flat acceptance (red). (e) The error associated to the exclusivity cuts is estimated by tightening the transverse momentum and the missing mass cuts (red) with respect to the standard cut (blue).

Chapter 6

Results and comparison with model predictions

This chapter presents the results obtained with the Fall 2018 CLAS12 dataset in the inbending configuration (see Section 2.8 for more details). The three TCS observables (R' ratio, A_{FB} and BSA) are extracted for different kinematic bins as a function of different variables. First, all the extracted values of the TCS observables and their systematic errors are shown in Section 6.1. A comparison with the CLAS data is shown in Section 6.2. Finally some results are compared with model predictions and discussed in Section 6.3.

6.1 Complete CLAS12 results for the TCS observables

In this section all the data points obtained in the analysis are displayed. For each observable and for each bin, the statistical error bar is shown as a vertical error bar. The horizontal error bar corresponds to the size of the bin. The horizontal position of the data points is the average value of the variable in the bin, corrected by the acceptance (and in the case of the A_{FB}, also by the bin volume correction). The grey bands show the total systematic uncertainty defined as the quadratic sum of all of the systematic contributions described in Section 5.13. The decomposition of the systematic uncertainty is shown under each plot. The red points reported on each plots correspond to the expected values of the observables in each bin for BH-only events. These points are calculated using BH-weighted simulation events and, within the acceptance of CLAS12 for the R' ratio and the BSA, and within the experimental forward and backward bin for the FB asymmetry. Finally, the mean values of the integrated kinematic variables, calculated using the same simulations, are given above each plot.

6.1.1 R' ratio

The R' ratio is extracted as a function of $-t$ (Figure 6.1) and ξ (Figure 6.2), with all the other variables integrated. The dependence on $-t$ is further explored by plotting the ratio as a function of $-t$ in the invariant mass bin [2 GeV − 3 GeV] (Figure 6.3).

The measured values of the R' ratio are always bigger than the values expected if only the BH process was contributing to the $\gamma p \to p' e^+ e^-$ cross section (red points in the following figures). This behavior is also seen in the high-mass region [2 GeV − 3 GeV] in Figure 6.3, where no vector-meson contamination is expected. This observation validates the fact that the CLAS12 data are sensitive to the BH-TCS interference cross section.

6.1.2 A_{FB}

The FB asymmetry is extracted as a function of E_γ (Figure 6.4), M (Figure 6.5), $-t$ (Figure 6.6) and ξ (Figure 6.7), with all the other variables integrated. In order to explore the invariant mass dependence further, and especially to investigate the effect of low-mass vector-meson resonances, it is also measured

as a function of $-t$ for the two additional mass bins, in the low-mass bin [1.5 GeV – 2 GeV] (Figure 6.8) and in the high-mass bin [2 GeV – 3 GeV] (Figure 6.9).

As in the case of the R' ratio discussed above, the extracted values of the A_{FB} shown in Figures 6.4, 6.5 and 6.6 are not compatible with zero, confirming that the BH process does not contributes alone to the $\gamma p \to p' e^+ e^-$ cross section. A none-zero asymmetry is seen in both the low-mass and high-mass bins. The signal seen in the high-mass bin, in Figure 6.9, can be attributed to the BH-TCS interference cross section, as there is no known vector-meson resonance in this mass range.

6.1.3 BSA

The BSA is extracted in CLAS12 acceptance, as a function of M (Figure 6.10), $-t$ (Figure 6.11) and ξ (Figure 6.12), with all the other variables integrated.

In each of these figures a clear photon polarization asymmetry is reported. This is a further confirmation that we observe the BH-TCS interference in the CLAS12 dataset, as the expected asymmetry for the BH process only is zero.

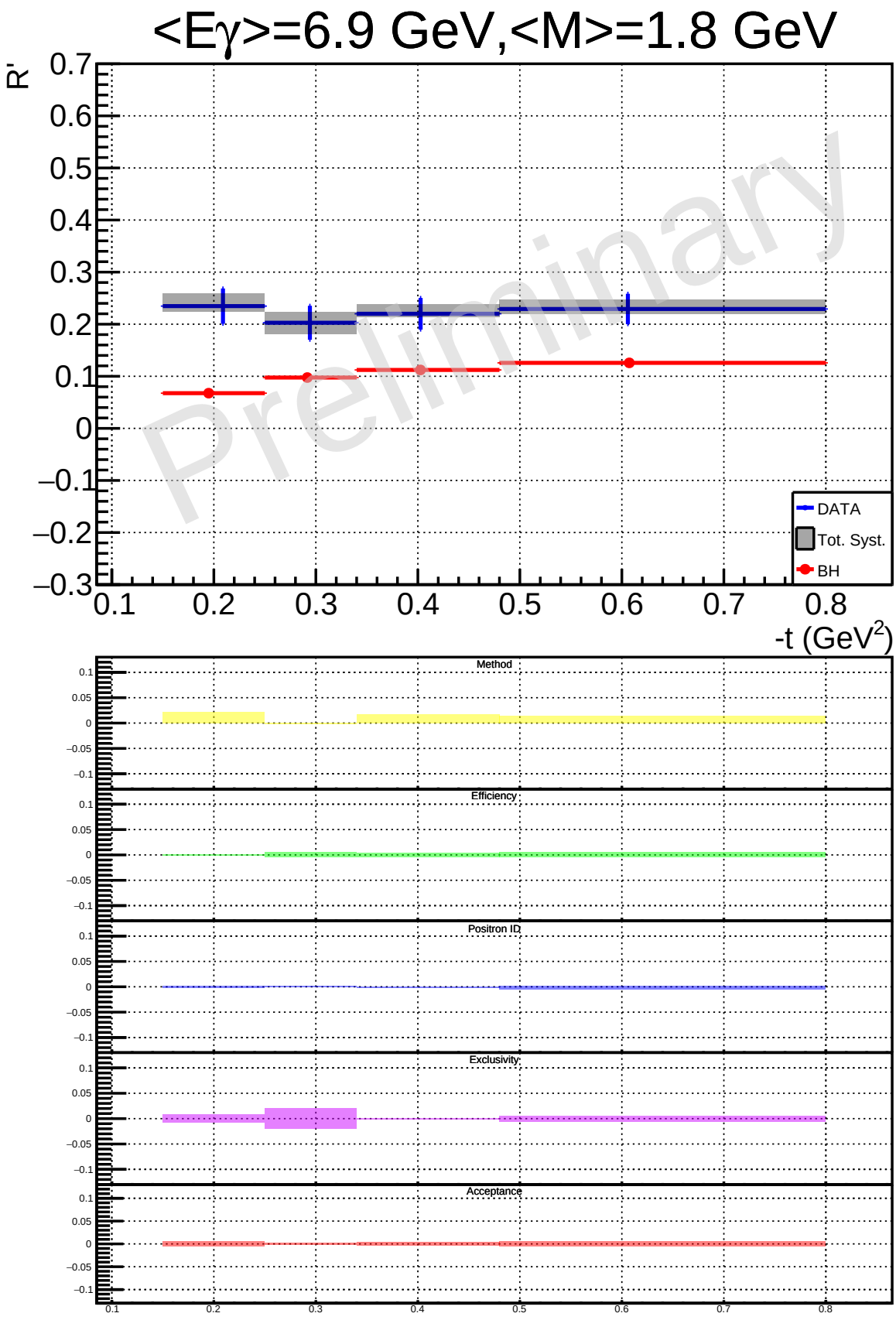


Figure 6.1: R' ratio as a function of $-t$, integrated over all the other variables. The horizontal error bars denote the bin size, the vertical error bars are statistical errors and the grey bands display the total systematic uncertainty. The red points are expected values for BH-only events, obtained with simulations. Tabulated values in Table O.1 in Appendix O.

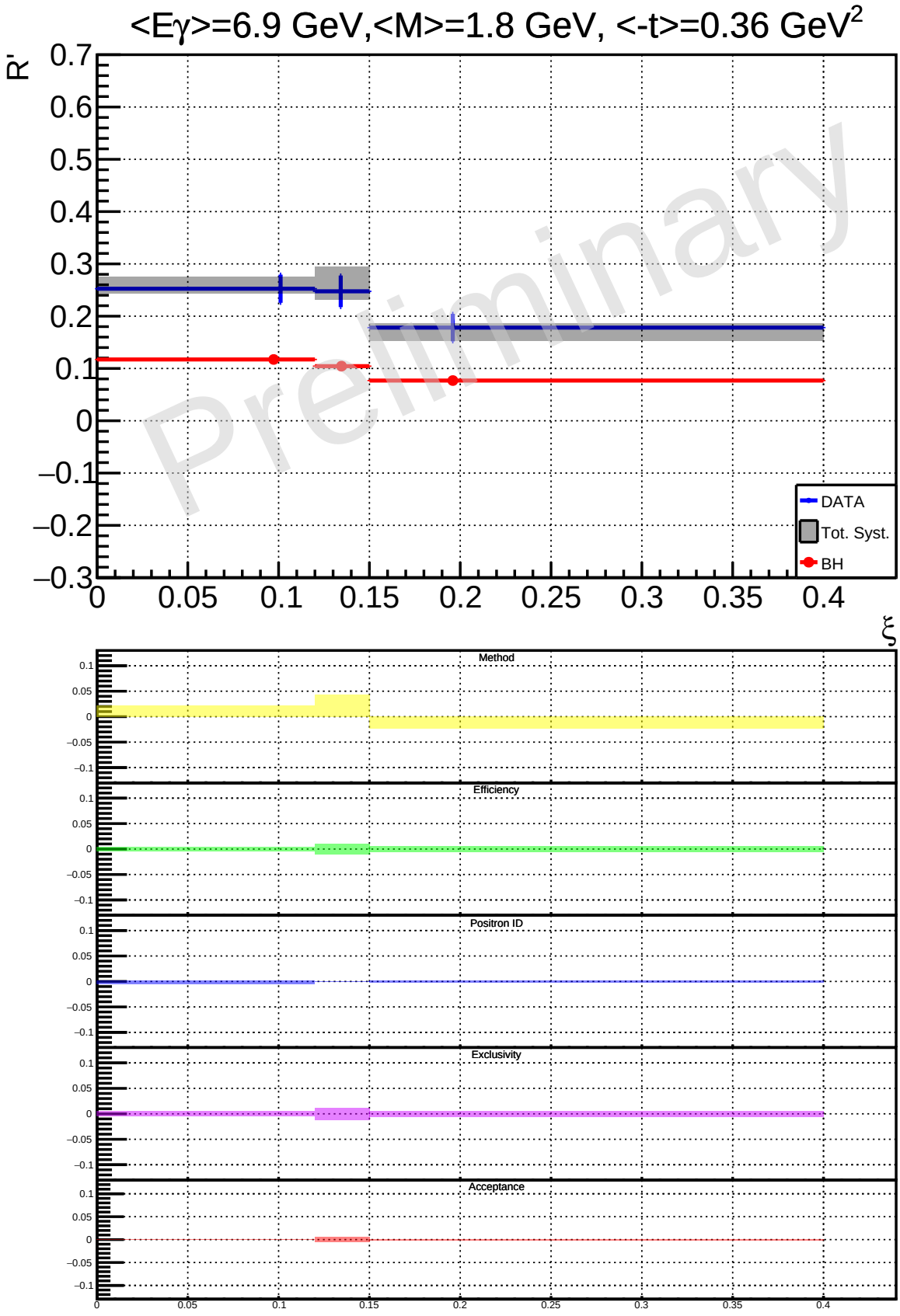


Figure 6.2: R' ratio as a function of ξ, integrated over all the other variables, using the same plotting conventions as in Figure 6.1. Tabulated values in Table O.2 in Appendix O.

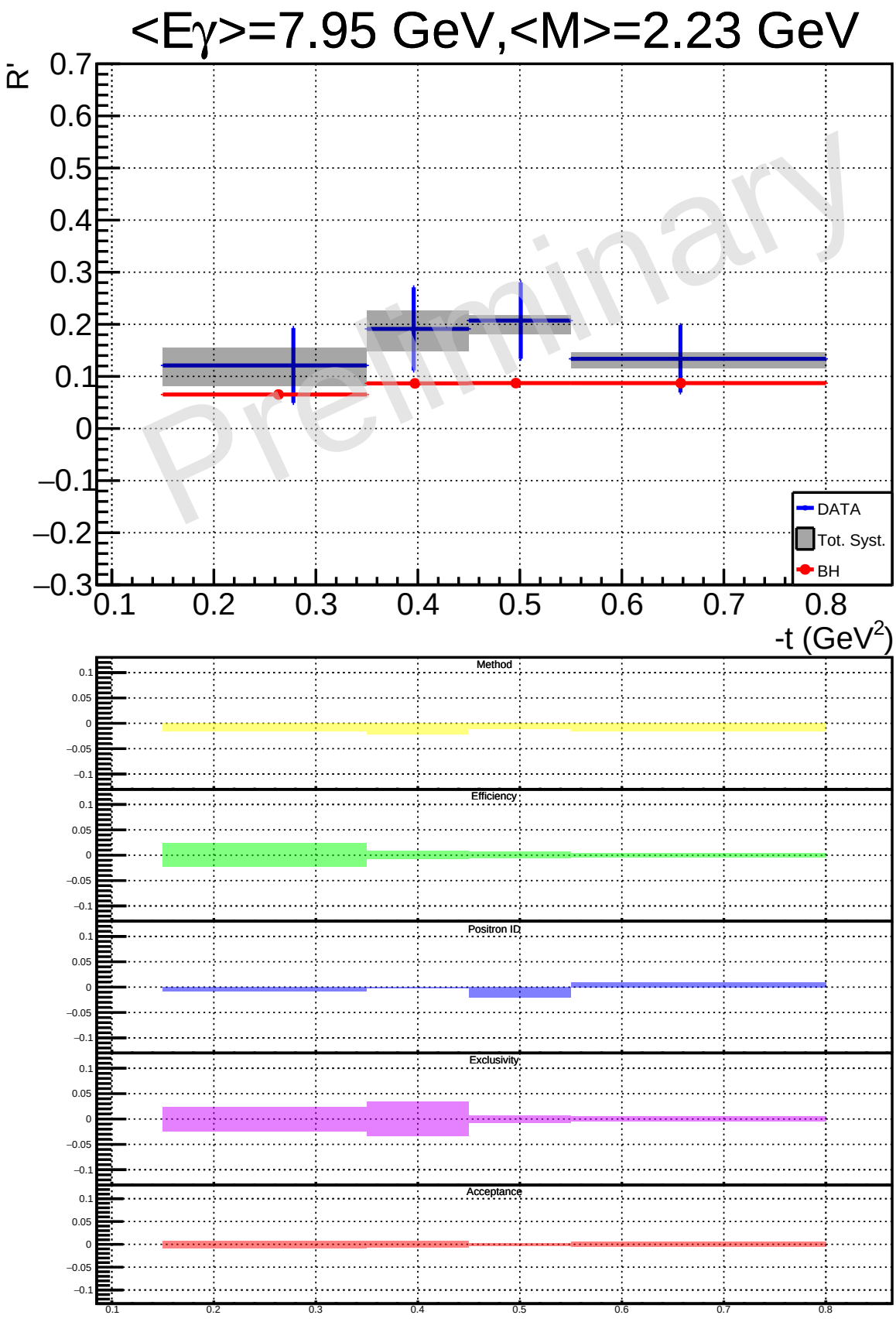


Figure 6.3: R' ratio as a function of $-t$, for the mass range [2 GeV – 3 GeV], integrated over all the other variables, using the same plotting conventions as in Figure 6.1. Tabulated values in Table O.3 in Appendix O.

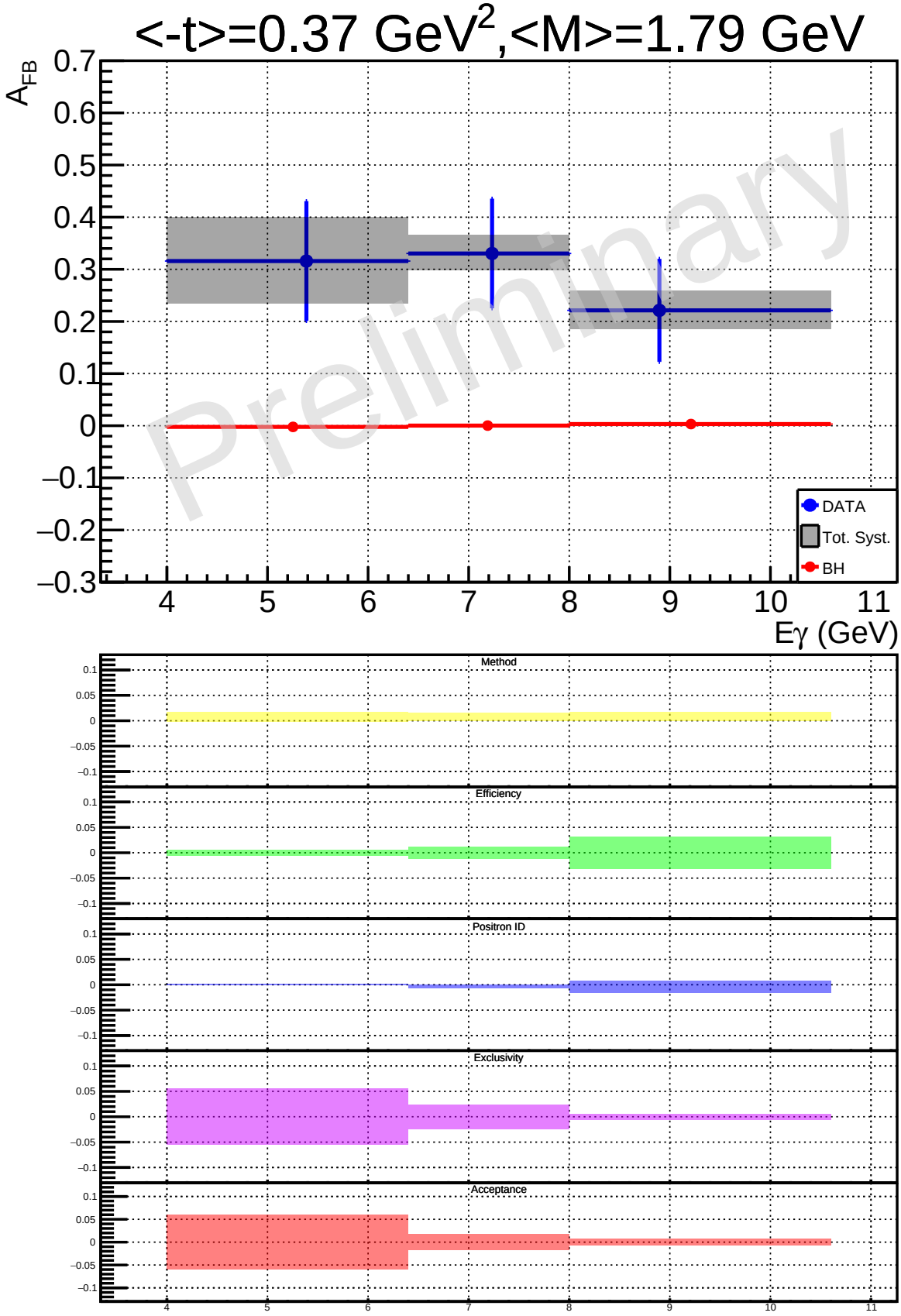


Figure 6.4: A_{FB} as a function of E_γ, integrated over all the other variables, using the same plotting conventions as in Figure 6.1. Tabulated values in Table O.4 in Appendix O.

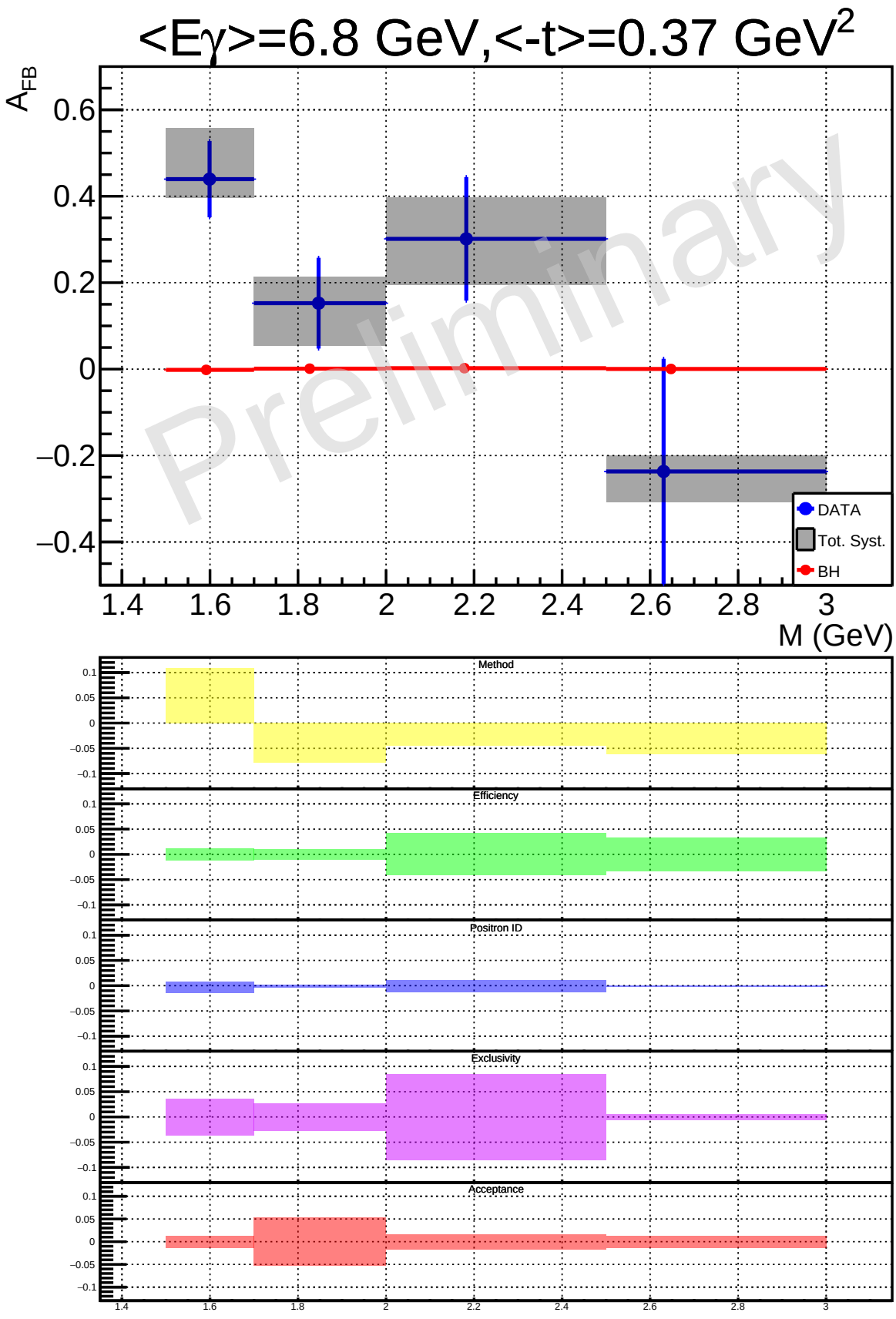


Figure 6.5: A_{FB} as a function of M, integrated over all the other variables, using the same plotting conventions as in Figure 6.1. Tabulated values in Table O.5 in Appendix O.

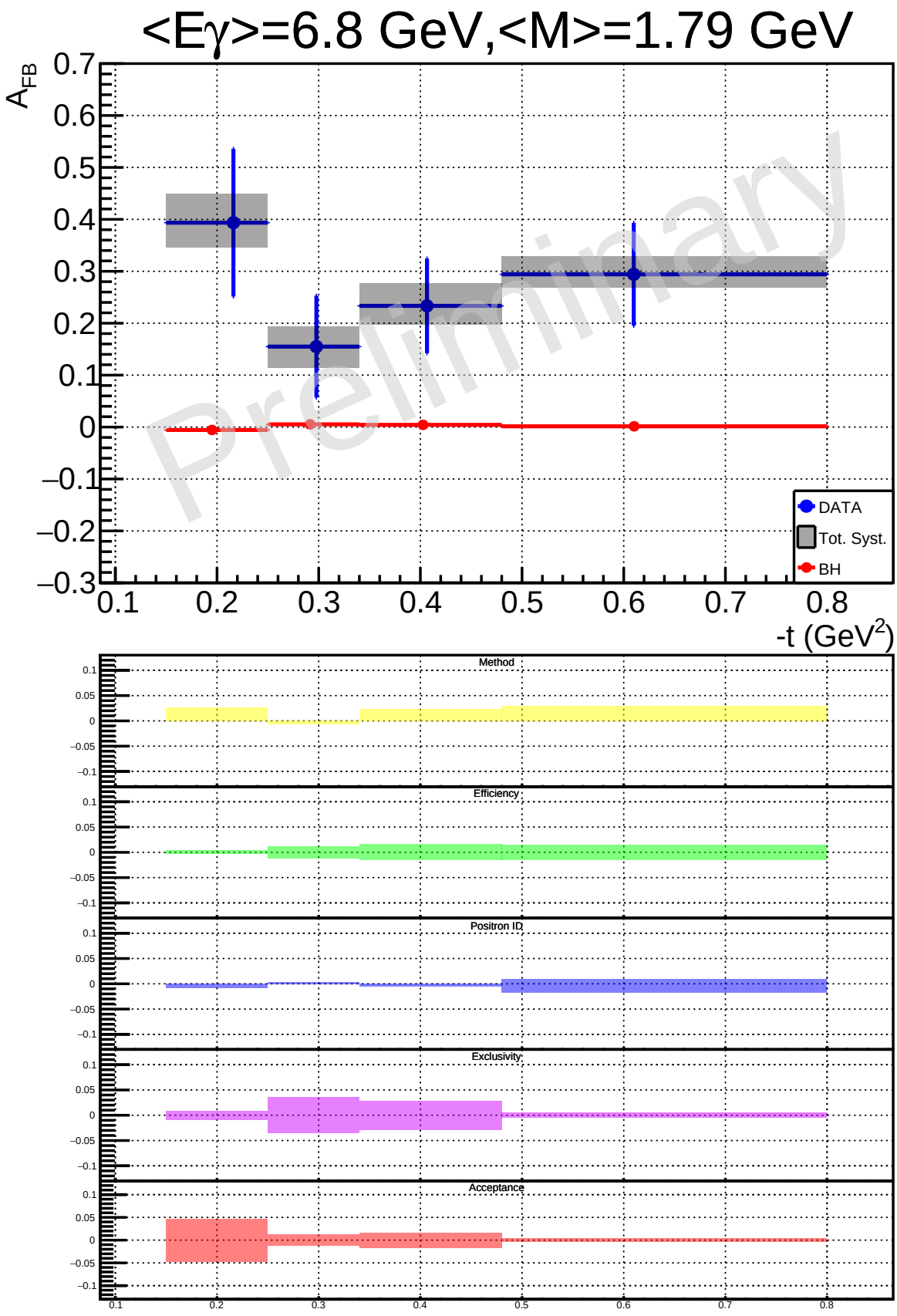


Figure 6.6: A_{FB} as a function of $-t$, integrated over all the other variables, using the same plotting conventions as in Figure 6.1. Tabulated values in Table O.6 in Appendix O.

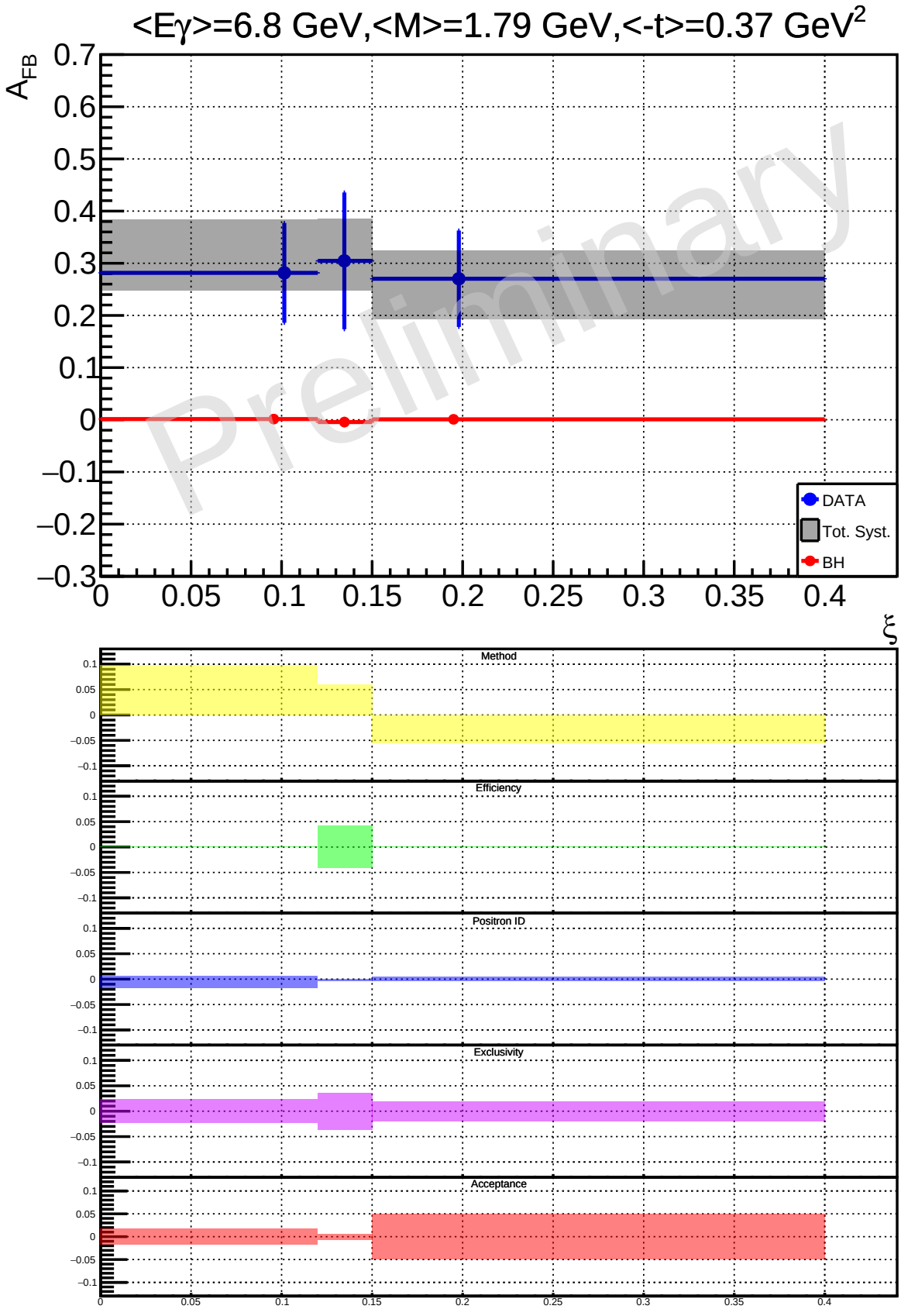


Figure 6.7: A_{FB} as a function of ξ, integrated over all the other variables, using the same plotting conventions as in Figure 6.1. Tabulated values in Table O.7 in Appendix O.

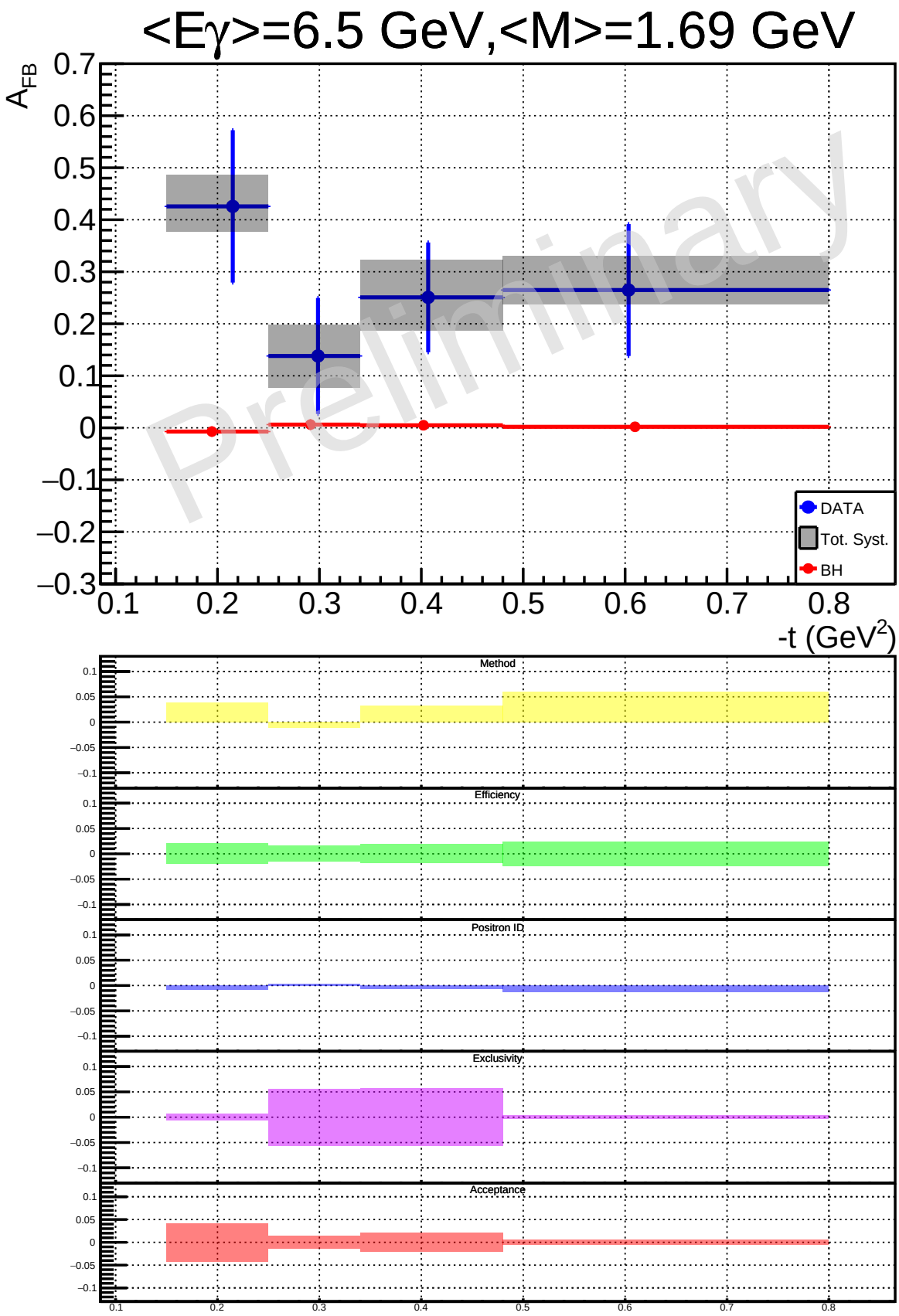


Figure 6.8: A_{FB} as a function of $-t$, in the [1.5 GeV – 2 GeV] mass range, integrated over all the other variables, using the same plotting conventions as in Figure 6.1. Tabulated values in Table O.8 in Appendix O.

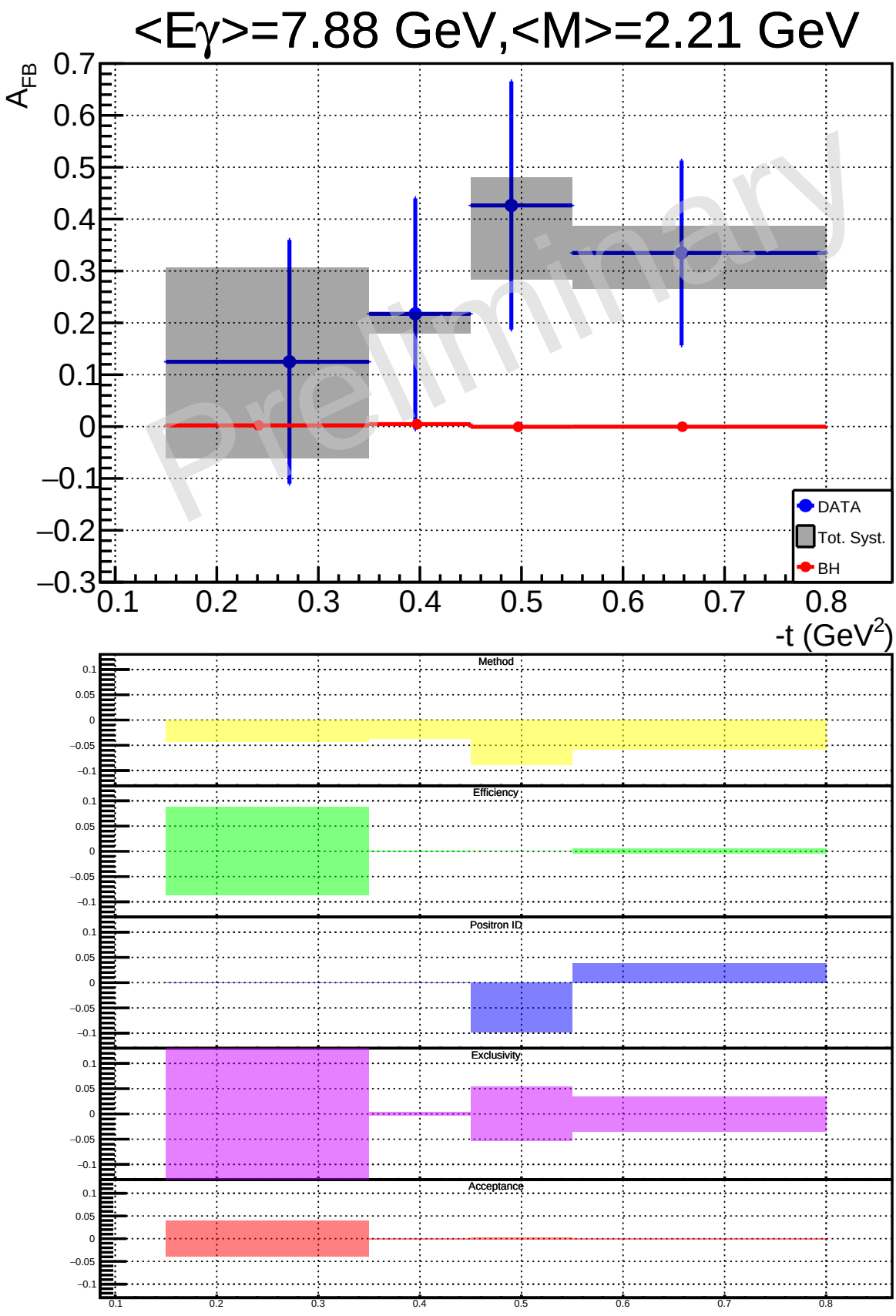


Figure 6.9: A_{FB} as a function of $-t$, in the [2 GeV − 3 GeV] mass range, integrated over all the other variables, using the same plotting conventions as in Figure 6.1. Tabulated values in Table O.9 in Appendix O.

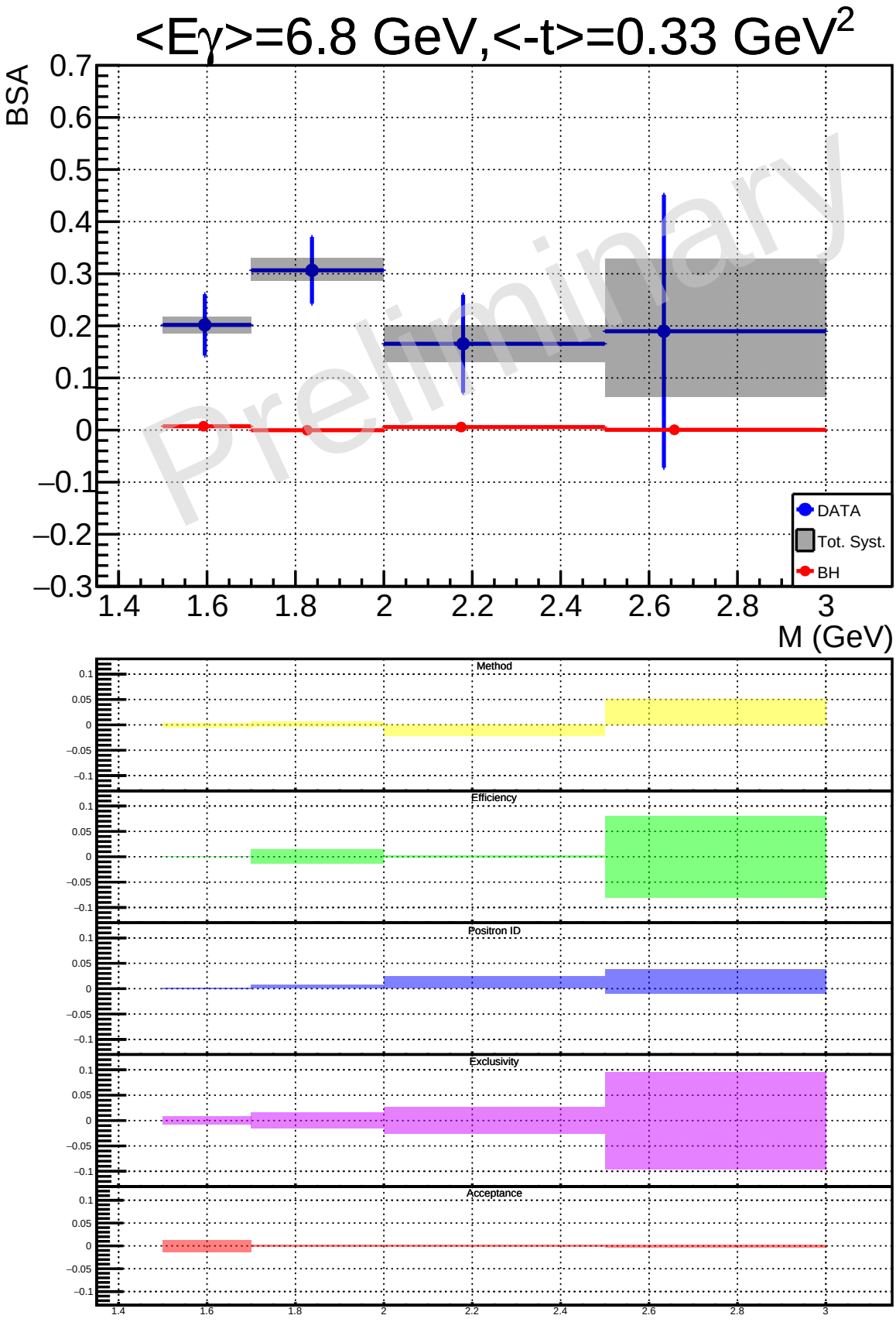


Figure 6.10: BSA as a function of M, integrated over all the other variables, using the same plotting conventions as in Figure 6.1. Tabulated values in Table O.10 in Appendix O.

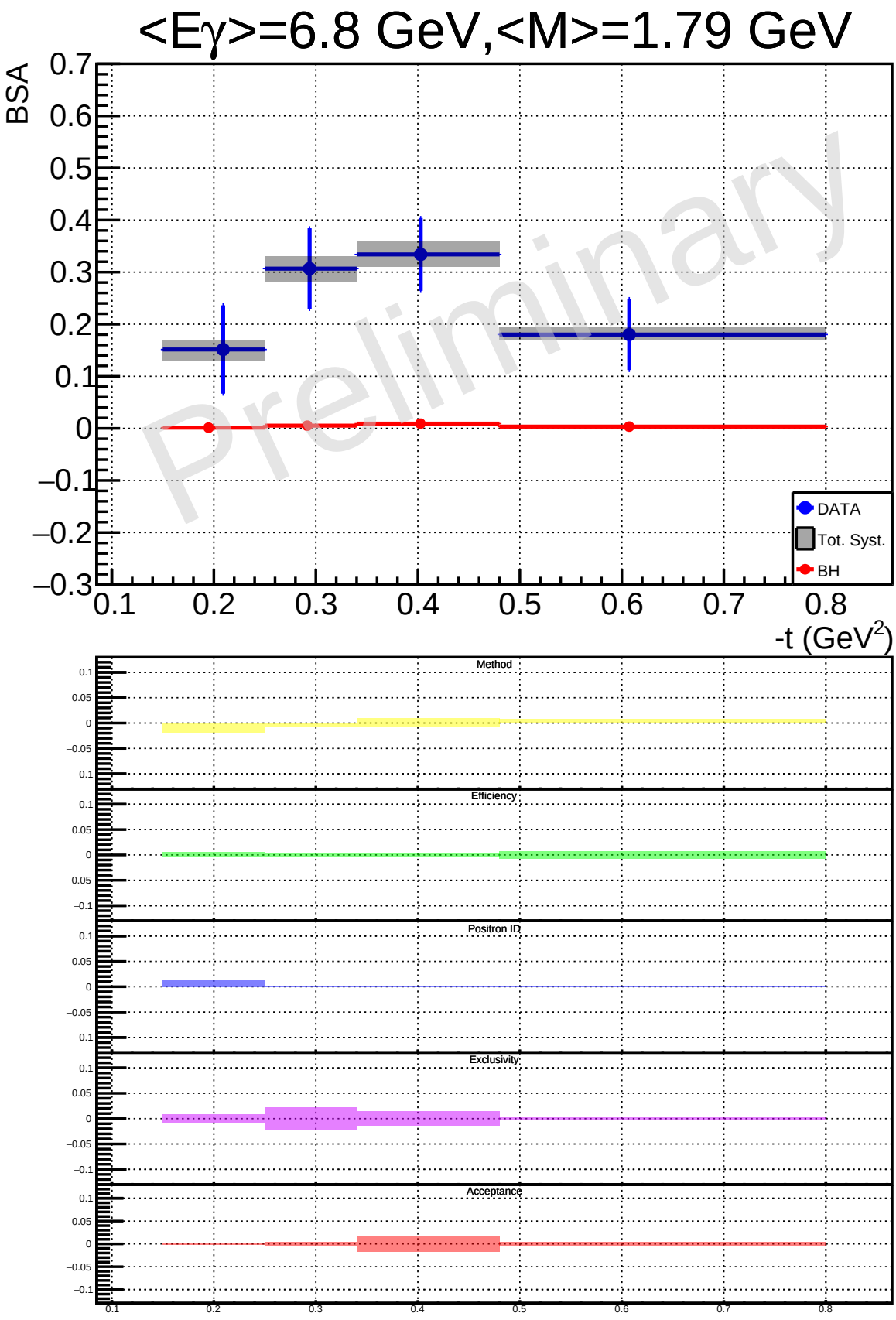


Figure 6.11: BSA as a function of $-t$, integrated over all the other variables, using the same plotting conventions as in Figure 6.1. Tabulated values in Table O.11 in Appendix O.

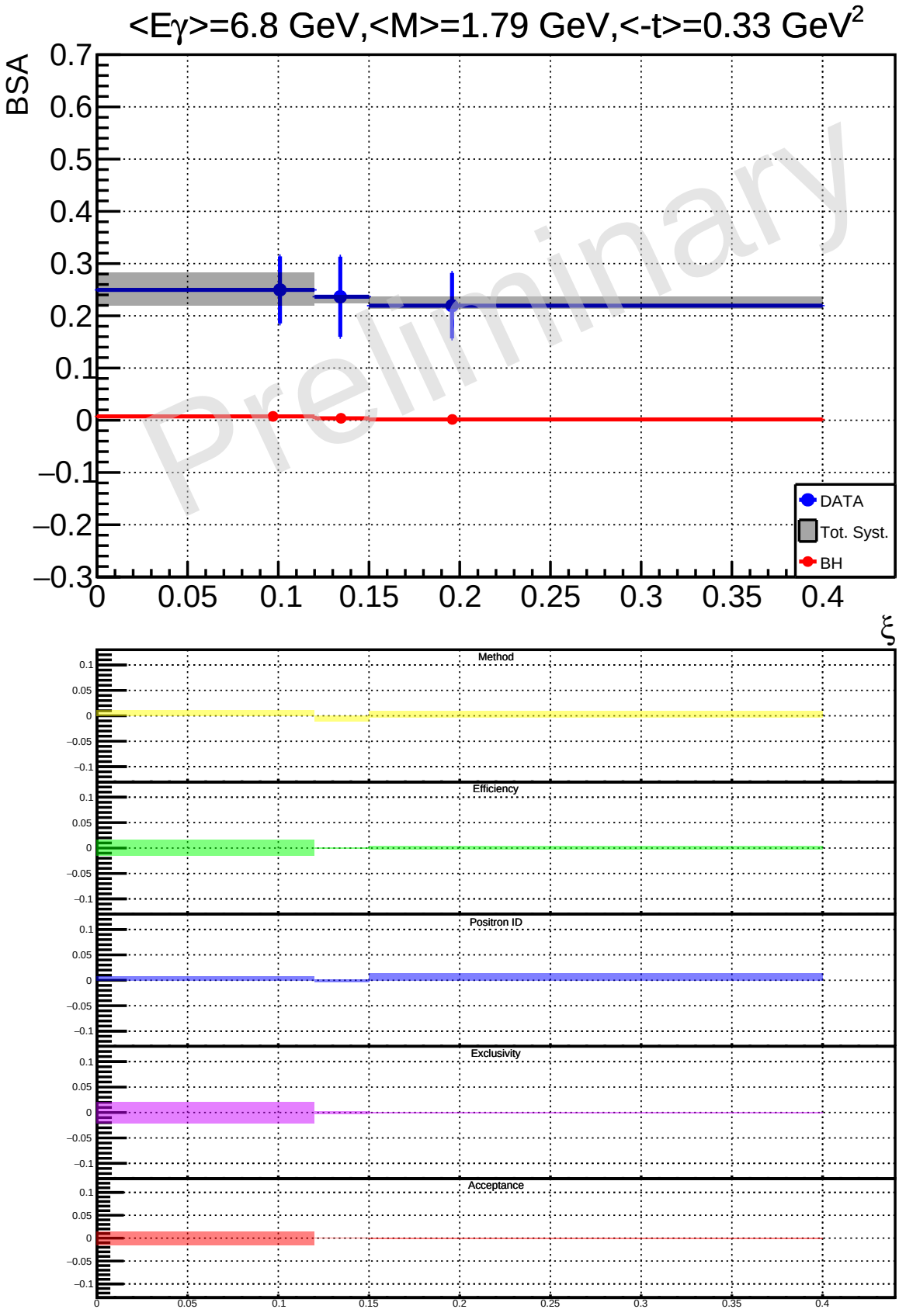


Figure 6.12: BSA as a function of ξ, integrated over all the other variables, using the same plotting conventions as in Figure 6.1. Tabulated values in Table O.12 in Appendix O.

6.2 Comparison with CLAS results

In the exploratory study performed on CLAS data, a first extraction of the cross-section ratio R' was performed. Because of the low energy of the electron beam delivered to CLAS (5.48 GeV), the lepton invariant mass range was limited between 0 and 2 GeV. The analysis was thus performed in the mass region above the $\phi(1020)$, and the ratio was extracted as a function of the squared proton transferred momentum [59].

For the comparison presented in this section, the same kinematic region as for the CLAS analysis is selected from the CLAS12 dataset:

- 2 GeV $< E_\gamma <$ 5 GeV
- 0.15 GeV$^2 < -t <$ 0.8 GeV^2
- 1.1 GeV $< M <$ 1.7 GeV

and the R' ratio was extracted and compared to the CLAS results. The CLAS12 dataset allows to have a thinner binning and to keep statistical error bars well below the CLAS ones. The ratio R' is calculated for eight $-t$ bins and the size of each bin is indicated by the horizontal error bars in the plot bellow. Note that the CLAS12 points are obtained without performing a complete acceptance calculation; the lowest-mass and lowest-energy bin of the acceptance presented in Section 5.6 is used.

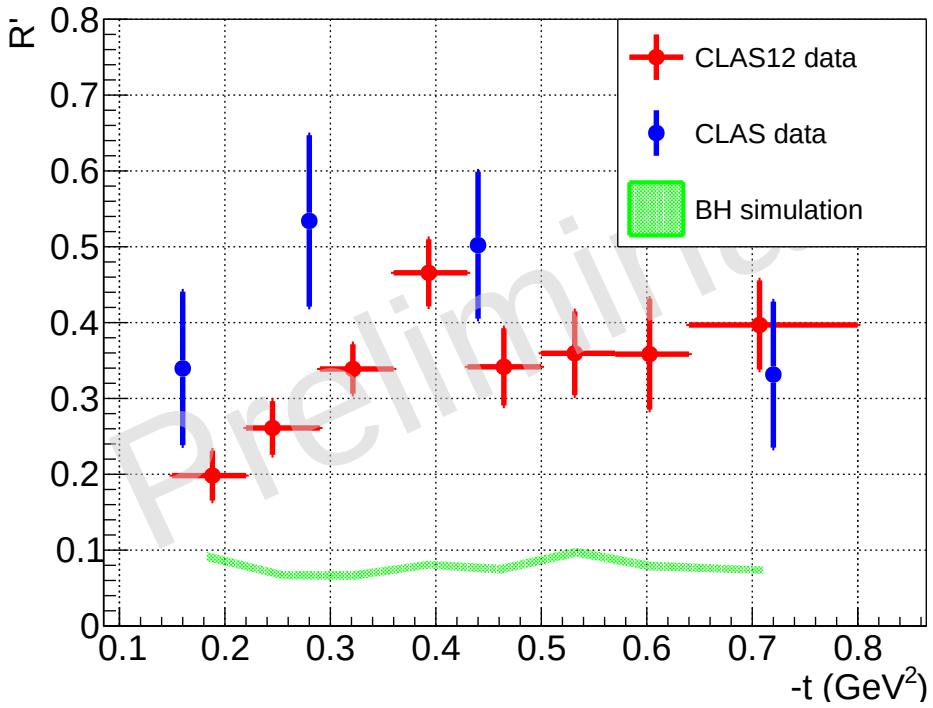


Figure 6.13: CLAS12 results for the R' ratio as a function of $-t$ in the kinematic region accessible by CLAS, together with the CLAS results obtained in [59] and BH-only simulations for comparison. The error bars correspond to the statistical error only.

Figure 6.13 shows the data point from CLAS (blue), CLAS12 (red) as well as the R' ratio calculated from BH-only simulation events with kinematics inside the CLAS12 acceptance (green). Both CLAS and CLAS12 datasets give roughly compatible results, given the error bars, indicating an asymmetry well above the one obtained from BH simulations. The difference between the CLAS and CLAS12 results can be explained by the fact that the ratio is calculated within the respective detector acceptance. Although in this mass region the energy scale might be too low to ensure factorization and vector-meson resonances might be too important to extract any information on TCS, both analyses have coherent results, indicating that the extraction method is under control.

6.3 Comparison Data/Models and physical interpretations

The TCS reaction is of great interest as it allows to measure the D-term via the the R' ratio and the A_{FB} asymmetry, both sensitive to the real part of the $\mathcal{H}$ CFF, as well to verify the universality of GPDs by extracting the photon polarization asymmetry (or BSA), which is sensitive to the imaginary part of $\mathcal{H}$. In this section, the data points obtained for the A_{FB} and the BSA are compared to model predictions provided by M.Vanderhaeghen using the VGG model [46, 47, 39] and by P.Sznajder using the GK model [40] in the *PARTONS* software [102].

BSA interpretation

Most of the data used to constrain GPDs have been measured in DVCS and Deeply Virtual Meson Production (DVMP). However, DVMP and DVCS data are difficult to compare directly as the former reaction involves Meson Distribution Amplitudes that must be measured using other reactions before being able to interpret DVMP in terms of GPDs. Contrary to DVMP, TCS does not involve distribution amplitudes and is only parameterized by GPDs, making it directly comparable to DVCS. The comparison between these two processes is an important test of the universality of the GPDs. This can be done using the TCS BSA, as it is directly sensitive to the imaginary part of the CFF $\mathcal{H}$ which is itself well constrained by DVCS data.

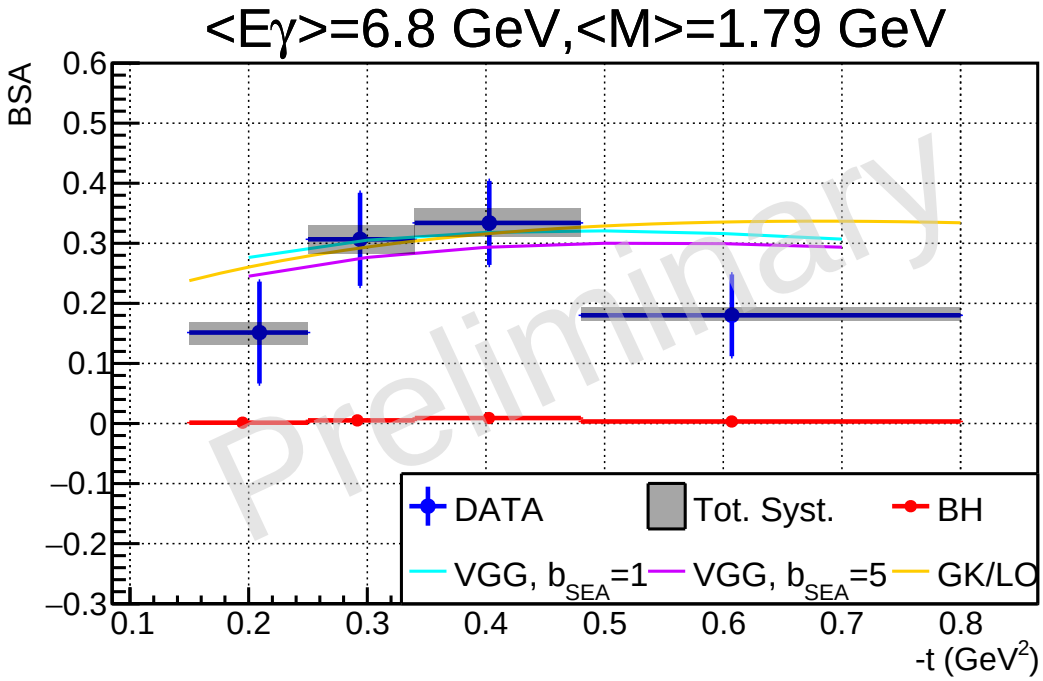


Figure 6.14: CLAS12 data points for the TCS BSA as a function of $-t$, evaluated at $\phi = 90°$, integrated over CLAS12 acceptance and over all the other variables. The vertical blue error bars are statistical uncertainties while the grey bands correspond to systematic uncertainties. Three model predictions, obtained using the VGG and GK models, are also displayed. The model predictions are calculated at the mean kinematic point given above the plot. The red points are the expected values of the BSA for BH-only events, obtained using BH-weighted simulations.

Figure 6.14 shows the TCS BSA extracted from the CLAS12 data as a function of $-t$, compared to three theoretical predictions. The two VGG curves (cyan and magenta) display the $-t$ dependence of the BSA evaluated at $\phi = 90°$ and for θ integrated from $\pi/4$ to $3\pi/4$. The other variables, E_γ and M, are set to 7 GeV and 1.8 GeV respectively. The BSA is calculated for two different values of the sea skewness parameter (see Subsection 1.2.3 for the definition). The hypothesis $b_{sea} = 1$ (cyan) is the default value of the VGG model, while the $b_{sea} = 5$ hypothesis (magenta) seems to be favored by the analysis of DVCS data in [29]. The GK prediction (orange) is evaluated at the mean kinematic point

of the plot, and the angular kinematics and integration are identical to those of VGG.

The values of the BSA extracted from the CLAS12 data are in agreement, within error bars, with the three theoretical predictions. This observation tends to validate the use of the GPD formalism to describe TCS data and is a hint for the universality of the GPDs, as the VGG model also describes well the DVCS data [29]. However, our data points do not strongly favor any of the VGG hypothesis. Further studies on the dependence of the TCS BSA with the b_{sea} parameter should be made in order to identify the kinematic regions where one could discern between both values.

Figure 6.15 shows the measured BSA as a function of the invariant mass of the lepton pair, M. The prediction obtained with the GK model, for the mean kinematic point specified above the plot and for angular dependencies identical to the ones in Figure 6.14, is superimposed (orange line). The GK curve is only displayed in the mass range between 1.5 GeV to 2.3 GeV, as values of the mass higher than 2.3 GeV are not kinematically allowed at the mean kinematic point of this plot. Nevertheless, the data points are in agreement, within error bars, with the theoretical prediction in the [1.5 GeV – 2.3 GeV] mass range. This is an indication that the GPD factorization seems to apply to TCS in a large part of the mass range accessible by CLAS12.

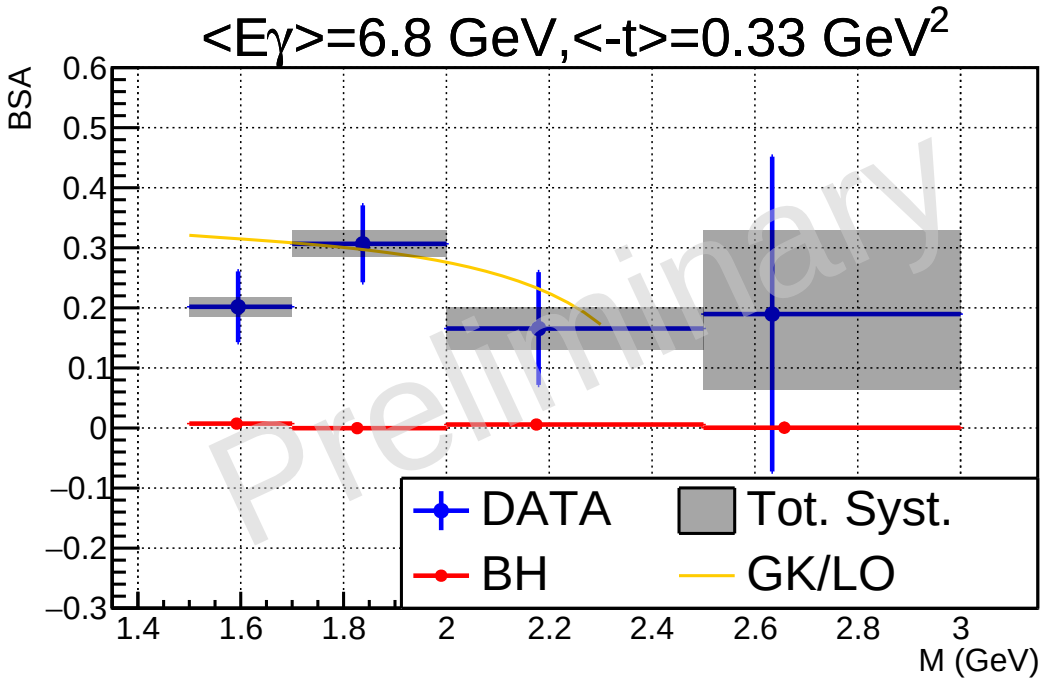


Figure 6.15: CLAS12 data points for the BSA as a function of M. The blue error bars are statistical uncertainties while the grey bands represent the systematic errors. The orange curve is the GK model prediction at the mean kinematic point of the data. The red points are the expected values for BH only (from simulations).

A_{FB} interpretation

The A_{FB} asymmetry has the advantage that it can be easily compared to theory, as it does not involve acceptance limits. Also, as shown in Section 5.9, it has a large sensitivity to the D-term, making it a valuable observable to extract this quantity.

The t-dependence of the A_{FB} extracted from data is compared to theoretical predictions in three cases. In all three cases the GK predictions are calculated for the average E_γ and the invariant mass of the lepton pair given above each plot. For the VGG predictions the mean kinematics are given in the corresponding figure captions.

In Figure 6.16 the A_{FB} asymmetry data points are plotted against $-t$, and all the other variables are integrated over the phase space detailed in Section 5.12. They are compared to predictions obtained

using the VGG model for $M = 1.8$ GeV (red lines) and for different values of θ_0 (at the lower edge of the angular bin $\theta_0 = 50°$ (dashed), at its center $\theta_0 = 65°$ (solid), and at its upper edge $\theta_0 = 80°$ (dash-double-dotted)). The effect of changing the average mass is also illustrated with the green curve calculated for $M = 1.5$ GeV. The cyan curve is calculated without the contribution of the D-term to the GPD H (the D-term contribution used for this calculation is the one described in [37]). Finally the GK prediction (orange) shown here is obtained by integrating the BH-TCS cross section in the forward and backward angular bin defined in Section 5.10. Also it has to be noticed that the GK prediction does not include the contribution of the D-term.

As already mentionned in Section 5.9, the VGG curves produced with and without the D-term clearly indicate that the D-term contribution to the GPD H has a large effect on the value of FB asymmetry. The data points are better described by the VGG model when the D-term is included, although error bars are still too large to completely rule out the case without the D-term. The GK model prediction seems to largely underestimate the asymmetry. This could be explained by the absence of D-term in this prediction. In order to ensure that this interpretation is valid for the whole mass range studied, the same comparison is done in the low-mass region [1.5 GeV – 2 GeV] and in the high-mass region [2 GeV – 3 GeV]. Indeed low mass events dominate when the asymmetry is integrated over the full mass range [1.5 GeV – 3 GeV], therefore low-mass vector-meson resonances (e.g. $\rho(1450)$) could be the origin of the observed FB asymmetry.

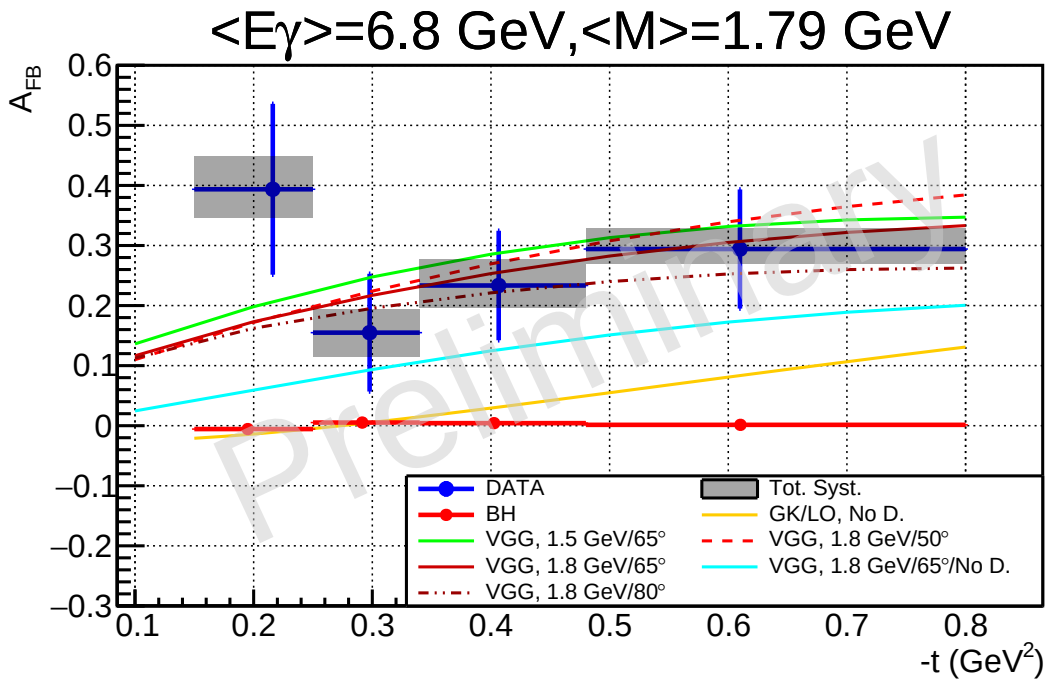


Figure 6.16: CLAS12 A_{FB} as a function of $-t$, integrated over all other variables. The data points are compared with theory predictions realized using the VGG and GK models. The cyan line is calculated with VGG and without the D-term contribution to H. The red lines correspond to three different θ_0 values for the forward direction, using the VGG model. The green line is the prediction of the VGG model for a lower invariant mass value (1.5 GeV). All the VGG curves are calculated at $E_\gamma = 7.0$ GeV, and at the invariant masses indicated in the legend. The orange line is the GK model prediction obtained with *PARTONS*, at the mean kinematic point of the plot, and integrated in the same angular range as the experimental A_{FB}. The red points are the expected values for BH only (from simulations).

Figure 6.17 shows the data points extracted in the [1.5 GeV – 2 GeV] low-mass region. They are compared with the same VGG predictions as in the full mass range case, as the average mass and the average photon energy do not change dramatically. The GK prediction is however recalculated at the mean kinematic point of this plot. The data points do not change substantially when the mass range is restricted to the [1.5 GeV – 2 GeV] range as the events included in this analysis mostly have a low

invariant mass, as seen in the mass distribution in Figure 5.5c. Therefore the conclusions drawn for Figure 6.16 also apply to Figure 6.17. Again the data points are better described when the D-term contribution is taken into account.

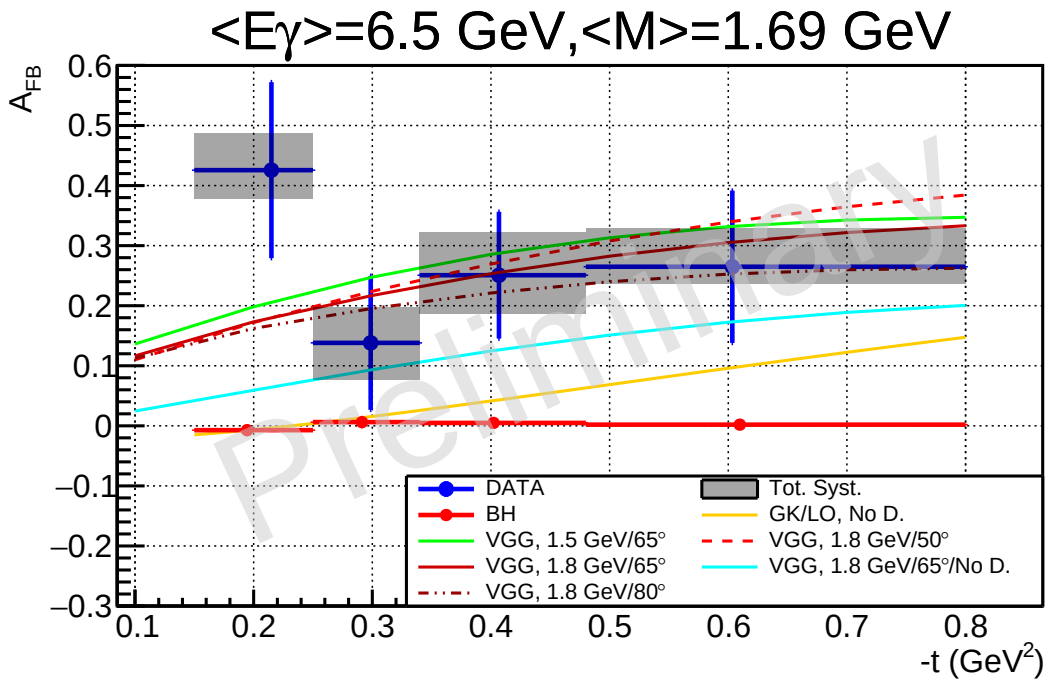


Figure 6.17: CLAS12 A_{FB} as a function of $-t$, integrated over all the other variables, with the invariant mass of the lepton pair integrated in the range [1.5 GeV − 2 GeV]. The VGG curves are the same as in Figure 6.16. The GK prediction is calculated at the mean kinematic point. The red points are the expected values for BH only (from simulations).

The A_{FB} measured in the high-mass region [2 GeV − 3 GeV] is shown in Figure 6.18. The data points are compared with predictions for the FB asymmetry calculated with the VGG model at the center of the angular bin ($\theta_0 = 65°$) and for two mass and photon-energy hypotheses. The prediction obtained when neglecting the D-term is displayed (cyan). The GK model prediction (orange) is also shown. While the effect of changing the kinematic point has little effect on the predicted asymmetry, the D-term plays again a very important role in the value of the asymmetry. As in the case of the full invariant mass integration, the data points tend to indicate that the D-term contribution to the asymmetry is necessary to explain its value. Indeed, both the VGG curve without the D-term and the GK curve underestimate the value of the asymmetry. The error bars shown in this plot do not allow for further conclusions. However, by increasing the available statistics, one could fit directly the D-term contribution and extract the pressure distribution inside the proton from this observable.

The mass and the photon-energy dependences of the extracted A_{FB} data points are also compared to GK model predictions. In both cases the GK model is evaluated at the mean kinematic point of the plots and integrated over the experimental forward and backward bins. Figure 6.19 shows the CLAS12 A_{FB} as a function of M. The GK prediction is only plotted in the [1.5 GeV − 2.3 GeV], as higher mass are kinematically forbidden at the mean kinematic point used to calculate the curve. A prediction for the GK model with a slightly different mean $-t$ ($-t = 0.33$ GeV^2) is also shown. The data points are not well reproduced by the GK model in the [1.5 GeV − 2.3 GeV] mass range. This discrepancy could originate from a possible vector-meson contamination, although the conclusions drawn from Figure 6.16 seem to indicate that the absence of the D-term in the model is the reason why the data and the GK model do not agree.

Figure 6.20 displays the A_{FB} data points as a function of the photon energy, E_γ. The GK prediction for this observable is also shown. As already observed in Figures 6.16 and 6.19, the GK predictions largely underestimate the measured FB asymmetry.

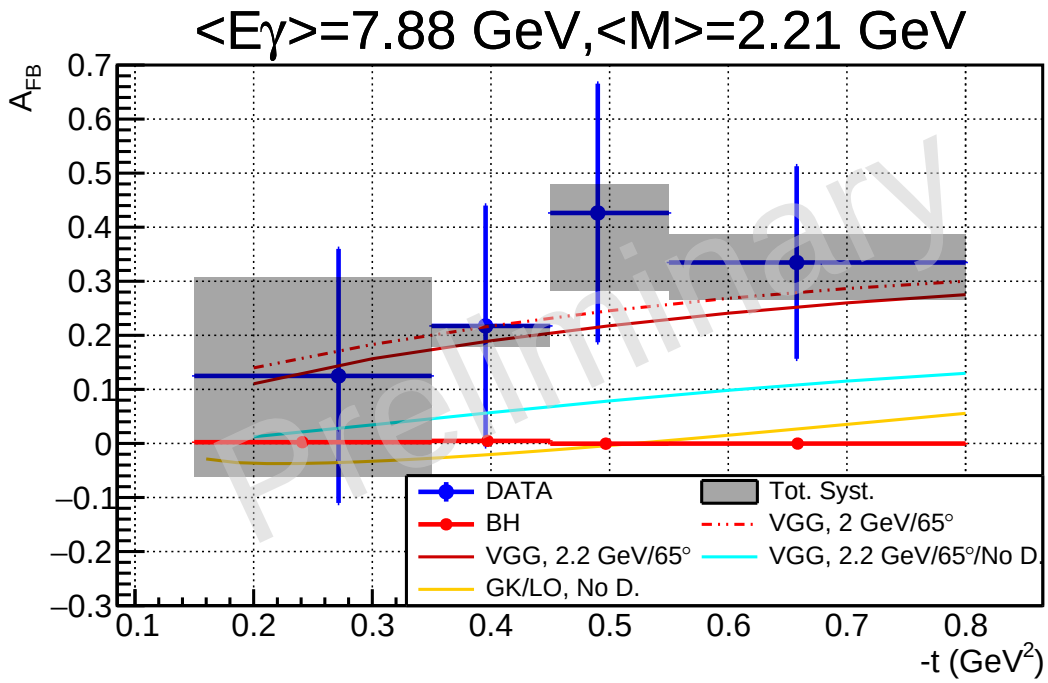


Figure 6.18: CLAS12 A_{FB} as a function of $-t$, integrated over all the other variables and with the invariant mass integrated in the range [2 GeV – 3 GeV]. The VGG model cyan curve does not take into account the D-term contribution. The plain red curve is calculated using VGG with the invariant mass set to 2.2 GeV and $E_\gamma = 7.88$ GeV while for the dot-dashed curve M is set to 2 GeV and $E_\gamma = 7.0$ GeV. The orange line is the GK model prediction at the mean kinematic point. The red points are the expected value for BH only (from simulations).

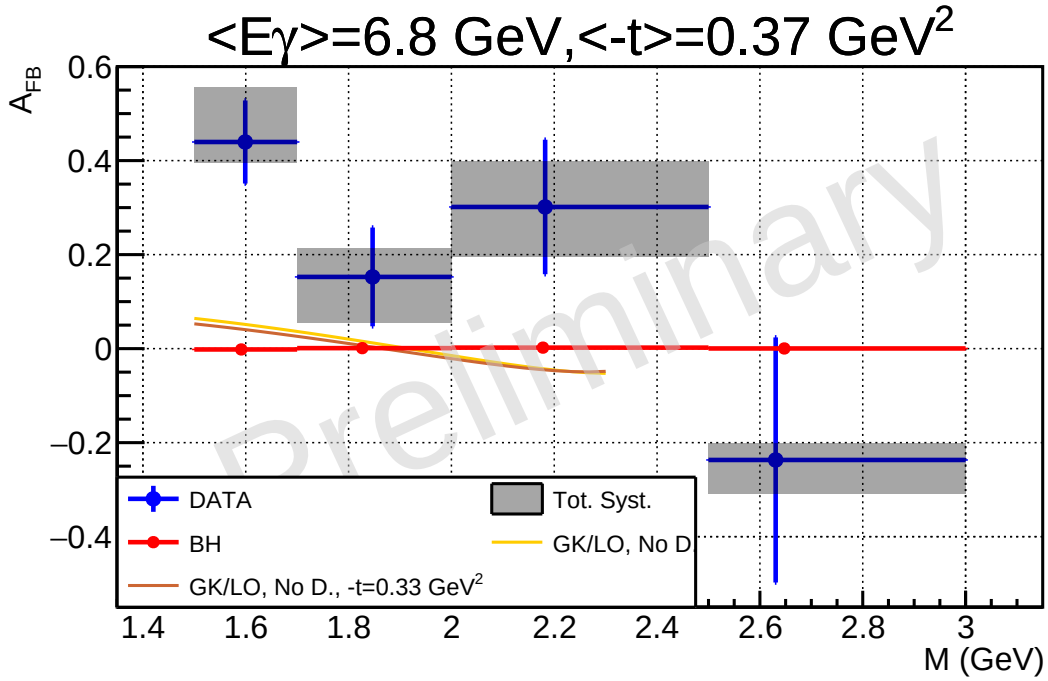


Figure 6.19: CLAS12 A_{FB} as a function of M, integrated over all the other variables. The vertical blue error bars are statistical errors and the grey bands are systematic uncertainties. The orange and brown curves are two predictions obtained with the GK model, for $-t = 0.37$ GeV^2 and $-t = 0.33$ GeV^2, respectively. The red points are the expected values for BH only (from simulations).

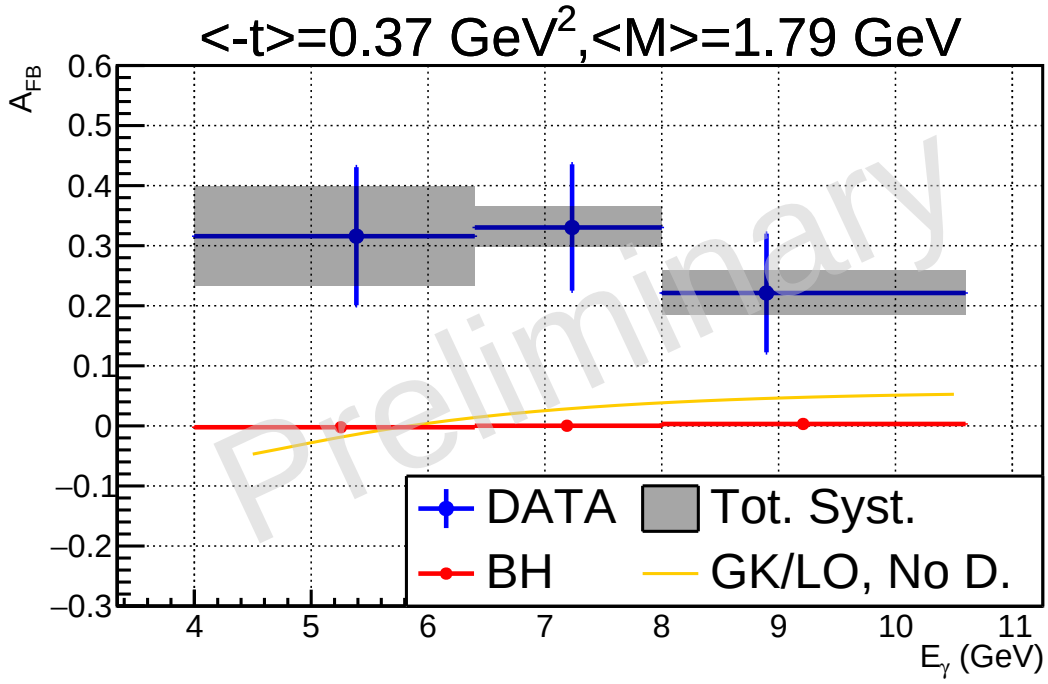


Figure 6.20: CLAS12 A_{FB} as a function of E_γ, integrated over all the other variables. The vertical blue error bars are statistical errors and the grey bands are systematic uncertainties. The orange curve is obtained with the GK model. The red points are the expected value for BH only (from simulations).

Conclusion

The description of the inner structure of the nucleon is a complex problem due to the nature of the interaction between its constituents, quarks and gluons. The theory of quantum chromodynamics describing their dynamics cannot be computed perturbatively at low energy. This phenomenon, called asymptotic freedom, imposes the use of ad hoc structure functions to encode the distribution of partons inside the nucleons. This manuscript is dedicated to the study of one set of structure functions, the Generalized Parton Distributions, particularly investigating two main topics: the commissioning of the Central Neutron Detector of CLAS12 and the measurement of the Timelike Compton Scattering process.

The first part of this manuscript is dedicated to the installation, calibration, reconstruction and performance validation of the Central Neutron Detector (CND). The CND allows to measure DVCS observables on the neutron by detecting the struck neutron in the $ed \rightarrow en'(p')\gamma$ reaction. This measurement plays a crucial role in the determination of the contribution from the angular momentum carried by valence quarks to the total spin of the nucleon. It also allows to extract GPDs in terms of quark flavors. The CND, located on the outer layer of the Central Detector of CLAS12, was installed at the beginning of this thesis. The cabling and part of the electronics were also installed during that time. The work described in this manuscript was performed after the installation.

The reconstruction and calibration softwares presented in this work were developed in parallel. The algorithms used in these sofwares are similar to the ones used in the other TOF detectors of CLAS12. However, due to the unconventional single-sided readout design of the CND, they were adapted to take into account features induced by this architecture. The CND calibration suite, a visual interface designed to provide an easy calibration procedure to the CND calibrators, was also partly developed during this thesis. Finally the performances of the CND were estimated from real data. The extracted timing resolution ($\sim$ 185ps) and position resolution ($\sim$ 3 cm) are slightly above the design values. However we showed in this manuscript that despite these resolutions, the CND still achieves to separate neutrons from photons for momenta up to 0.8 GeV. The neutron detection efficiency was also estimated from data using the $ep \rightarrow e'\pi^+(n)$ reaction. The efficiency value obtained ($\sim$ 10%) is in agreement with the design value. A comparison with the efficiency calculated from simulations was also done and it shows a good agreement over the whole range of momenta and polar angles accessible by the CND. Finally, very preliminary results obtained with real data taken by the CND were shown to illustrate its impact on the neutron DVCS measurement.

The second part of the manuscript focuses on the measurement of TCS observables using the RGA dataset recently taken by the CLAS12 detector. After presenting the experimental setup, the data analysis framework was explained in details. One of the main issue encountered during the analysis was the positron identification. Indeed two-pions electroproduction events where the negative pion stays undetected ($ep \rightarrow e'p'\pi^+(\pi^-)$) may mimic TCS events ($\gamma p \rightarrow p'e^+e^-$). In particular, when the momentum of the π^+ is higher than the HTCC threshold, the pion can be mistakenly identified as a positron. An algorithm based on neural networks was developed to solve this issue. The approach presented in this manuscript allows to reduce the π^+ contamination at high momenta by a factor ten, from 50% to 5%. The neural network was validated both on simulation events and on real data. Other tools were also developed for this analysis, such as proton efficiency corrections. These corrections were put in place in order to correct the momenta shifts observed in the data and which were not well reproduced in the simulations.

A phenomenological study on the Forward-Backward asymmetry (A_{FB}) for TCS was also performed during this thesis. Indeed, it was the first time that this observable was investigated for TCS, both theoretically and with real data. We proved that this observable is strongly sensitive to the real part of the CFFs. In particular it is shown in this thesis that the contribution of the D-term, a quantity which is linked to the pressure distribution inside the nulceon, has an important effect on the A_{FB}.

Finally the last part of this work summarizes our results on TCS. All the extracted quantities were compared to the values that are expected for them if only the BH process contributes. We observed that the data points always differ from the ones obtained with the BH-only contribution. This behavior was also reported in the resonance-free mass region, proving that we indeed extracted the BH-TCS interference contribution of the $\gamma p \to p' e^+ e^-$ reaction. These results were also compared with the predictions of two models, VGG and GK. The photon polarization asymmetry data points suggest that the GPD formalism describes well the CLAS12 TCS data. This observation advocates for the description of TCS in terms of GPDs, and thus for their universal nature. We also noted the fact that the FB asymmetry data points are better reproduced by the VGG model, which includes the D-term. This indicates the important role played by the D-term in the modeling of GPDs. However, because of the size of our error bars, no direct extraction of the D-term could be performed. Nevertheless, these results point out that a direct extraction of the D-term should be possible with the higher statistics provided by the complete CLAS12 dataset.

Appendices

Appendix A

Calculation of the ξ^2 term in the 1st x-moment of the GDP H

In this appendix the calculations of Equation (1.53) in Subsection 1.2.3 are detailed. The contribution to the ξ^2 term in the first x-moment of the GPD H comes from the D-term integral and reads:

$$\int_{-1}^{1} dx\ x\ \theta(\xi-|x|)\frac{1}{N_f}D(\frac{x}{\xi},t) = \frac{1}{N_f}\sum_{n\ \text{odd}} d_n(t)\int_{-1}^{1}\theta(\xi-|x|)dx\ x(1-z^2)C_n^{3/2}(z), \tag{A.1}$$

where $D(\frac{x}{\xi},t)$ is expended as a sum of Gegenbauer polynomials as:

$$D(t,z) = (1-z^2)\sum_{n\ odd} d_n(t)C_n^{3/2}(z), \tag{A.2}$$

and z is given by $z = x/\xi$, with the property:

$$|z| \leq 1. \tag{A.3}$$

The integral of Equation (A.1) can be written as:

$$\int_{-1}^{1}\theta(\xi-|x|)dx\ x(1-z^2)C_n^{3/2}(z) = \int_{-\xi}^{\xi} dx\ x(1-z^2)C_n^{3/2}(z) = \xi^2\int_{-1}^{1} dz\ z(1-z^2)C_n^{3/2}(z), \tag{A.4}$$

where in the first equality we have taken advantage of the step function and the fact that $\xi < 1$, and in the second equality the change of variable $x \to z\xi$ has been done.

Using the Gegenbauer polynomial orthogonality properties:

$$\int_{-1}^{1} dz\ (1-z^2)^{\alpha-\frac{1}{2}}C_n^{\alpha}(z)C_m^{\alpha}(z) \propto \delta_{n,m}, \tag{A.5}$$

and noting that $C_1^{\alpha}(z) = 2\alpha z$, one can re-write the right-hand side of Equation (A.4) as:

$$\xi^2\int_{-1}^{1} dz(1-z^2)\frac{1}{3}C_1^{3/2}(z)C_n^{3/2}(z) = \delta_{n,1}\ \xi^2\int_{-1}^{1} dz\ 3z^2(1-z^2) = \xi^2\delta_{n,1}\frac{4}{5}. \tag{A.6}$$

The ξ^2 contribution thus reads:

$$\int_{-1}^{1} dx\ x\ \theta(\xi-|x|)\frac{1}{N_f}D(\frac{x}{\xi},t) = \frac{1}{N_f}\xi^2\frac{4}{5}d_1(t). \tag{A.7}$$

Appendix B

Fisher discriminant

The Fisher discriminant method (also referred as linear discriminant analysis) consists in finding an axis in the hyperspace of the input variables such that, when projecting the signal and the background on this axis, the signal subset is as far as possible from the background subset, while all the signal (resp. background) events stay in the same neighborhood.

Let $x_{S(B),k}(i)$ be the value of the variable $k \in [1, N_{var}]$ for an event i which is in the signal subset (S) (resp. in the background subset (B)), let $\overline{x}_{U,k}$ be the mean of the variable k for events in the subset $U = \{S, B\}$.

The within-class matrix is defined as:

$$W_{k\ell} = \sum_{U=S,B} \langle (x_{U,k} - \overline{x}_{U,k})(x_{U,\ell} - \overline{x}_{U,\ell}) \rangle = C_{S,k\ell} + C_{B,k\ell} \,, \tag{B.1}$$

where $C_{S(B)}$ is the covariance matrix of the signal (background) sample.

The Fisher coefficients, F_k, are given by

$$F_k = \frac{\sqrt{N_S N_B}}{N_S + N_B} \sum_{\ell=1}^{N_{var}} W_{k\ell}^{-1} \left(\overline{x}_{S,\ell} - \overline{x}_{B,\ell} \right) \,, \tag{B.2}$$

where $N_{S(B)}$ is the number of signal (background) events in the training sample. The Fisher discriminant $y_{Fi}(i)$ of an event i is given by:

$$y_{Fi}(i) = F_0 + \sum_{k=1}^{Nvar} F_k x_k(i) \,. \tag{B.3}$$

where the offset F_0 centers the sample mean $\overline{y}_{Fi}(i)$ of all the events at zero.

Appendix C

Boosted decision tree

Decision tree

Decision trees classify signal and background events by applying successive cuts on the input variables. Let $\mathbf{x}_i$ be the set of input variables for the event i. A decision tree first finds a cut c_1 and a variable x_1 that results in the best separation of signal and background. Then the total signal+background sample is divided into two subsets according to the cut. These two subsets are called nodes. The final nodes with no sons are called leafs. For each node the purity of the subset is defined as:

$$P = \frac{\sum_S W_S}{\sum_S W_S + \sum_B W_B} \tag{C.1}$$

where the subscript S (B) denotes signal (background), and $W_{S(B)}$ is the weight of the event. So far all the weights are equal and set to $1/N$ where N is the total number of events. The Gini score G of a node is defined as:

$$G = \sum_{i=1}^{n} W_i \; P(1-P) \tag{C.2}$$

where the index i runs over the n events in the given node. The value of the cut c is optimized to minimize the total score of the two son nodes. The process is repeated until the algorithm reaches some predefined stop criterion (e.g. the maximum depth of the tree is reached or the number of events in a leaf is too small). The depth of a tree is defined by the number of nodes to cross to reach a leaf (e.g. the tree shown in Figure C.1 has a depth of 3). If a leaf has purity higher then 1/2 then it is called a signal leaf, otherwise a background leaf. Events that end up in signal leaves are classified as signal, the others as background. The schematic structure of a decision tree is shown in Figure C.1 and a real decision tree used in the MVA analysis of Chapter 4 is shown in Figure C.2. .

Boosting

A large number of decision trees are created sequentially for a single classifier. Boosting consists in changing the weights W_i of events i that have been mis-classified by the tree k for the training of the subsequent tree $k+1$. In the following the notation of [107] is used. Let:

- $y_i = 1$ if event i is signal, $y_i = -1$ otherwise
- $T_k(\mathbf{x}_i) = 1$ if the tree k classifies event i as signal, $T_k(\mathbf{x}_i) = -1$ otherwise
- $I_i = 0$ if the tree k classifies event i in the good subset (i.e. $y_i = T_k(\mathbf{x}_i)$), $I_i = 1$ otherwise.

In the following, the adaptative boosting algorithm is presented as it is used in Chapter 4. Other boosting algorithms are possible and are described in [106].

The classification error of the tree k is defined as:

$$err_k = \frac{\sum_{i=1}^{N} W_i I_i}{\sum_{i=1}^{N} W_i}. \tag{C.3}$$

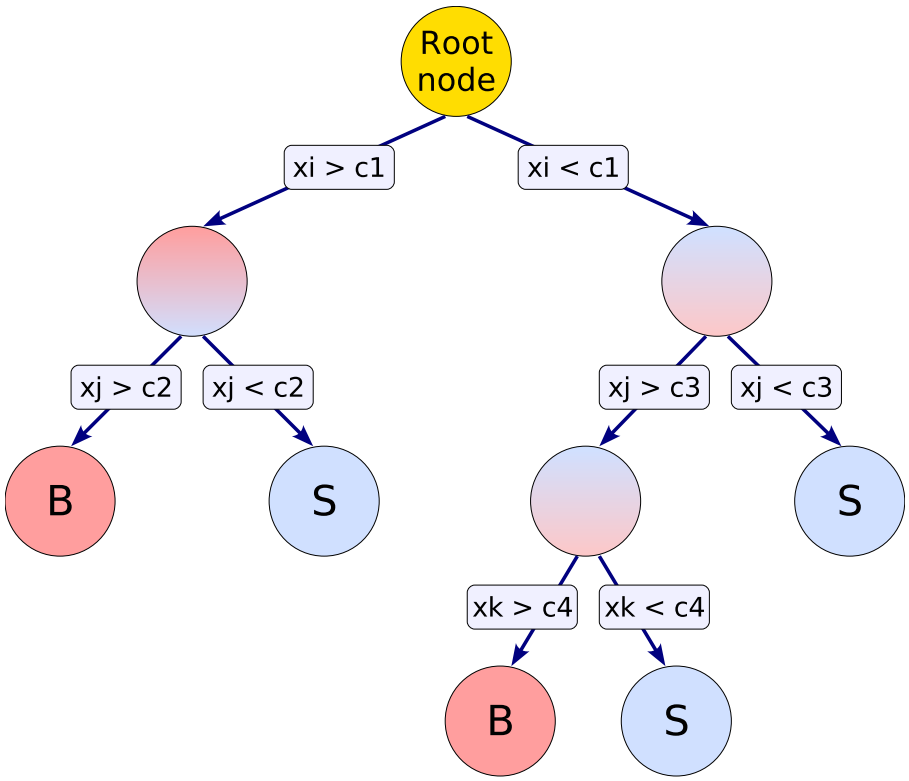


Figure C.1: Schematic structure of a decision tree. The event set at each node is split according to the best available cut to distinguish signal and background. Figure extracted from [106].

The boost weight of tree k is defined as:

$$\alpha_k = \beta \ln\left(\frac{1 - err_k}{err_k}\right), \tag{C.4}$$

where β is called the learning rate and is fixed before training. After the training of tree k, all weights are changed as:

$$W_i \to W_i \, e^{\alpha_k I_i}, \tag{C.5}$$

and normalized:

$$W_i \to \frac{W_i}{\sum_{i=1}^{N} W_i}. \tag{C.6}$$

The new weights are then used to build the tree $k+1$. The previous steps increase the weight of mis-classified events by the tree k. The next tree $k+1$ will thus look for better cuts to distinguish these specific high-weight events. In practice the first trees to be built will have a classification error lower than later trees. The output of the complete classifier for event i is defined as:

$$T(\mathbf{x}_i) = \sum_{k=1}^{N_{tree}} \alpha_k T_k(\mathbf{x}_i), \tag{C.7}$$

where the sum runs over all the trees included in the classifier. The adaptative boosting method is powerful on small individual decision trees (with depth lower or equal to 3). Although small individual decision trees have a low separation power, boosting allows the classifier to reach good performances [106]. Furthermore small decision trees are less subject to overtraining (i.e. classification based on features specific to the training sample). In the analysis presented in Chapter 4 we use trees with depth smaller or equal to 3. The learning rate β is set to 0.5 as it is reported to give good results [106, 107]. The number of decision trees trained in our analysis is 850. Finally we require that a leaf has to have at least 2.5% of the total number of training events.

Bagging

In addition to the boosting algorithm, one can apply the bagging procedure during the training of the classifier. Bagging refers to the process of picking with replacement a subset of events from the whole training set, and perform the training on this subset. In practice, at each iteration of the boosting procedure (ie: for each tree k), a subset of the training sample is randomly picked and the tree k is built using this subset. In the analysis of Chapter 4 half of the whole training set was used at each iteration.

Architecture

The architecture of the first 6D tree used in the analysis presented in Chapter 4 is shown in Figure C.2.

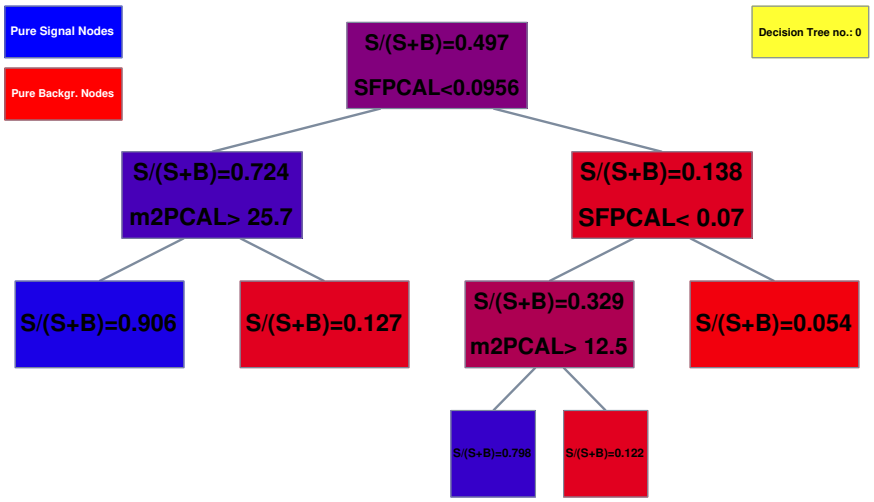


Figure C.2: First iteration of the BDT algorithm for the 6D analysis presented in Chapter 4. One can see that the cut made in the first node is on the sampling fraction of the PCAL, which is the most efficient variable to distinguish the true positrons from the pions.

Appendix D

Multilayer perceptron

Architecture

A Multilayer Perceptron (MLP) is a classifier formed of entities called neurons (see Figure D.1) linked together by connections. In a MLP neurons are arranged in layers and connections are only allowed between neurons from consecutive layers. A network starts with an input layer with one neuron for each input variable and ends with an output layer with a single neuron. There can be any number of layers and any number of neurons per layer between the input and output layers: such layers are called "hidden layers". A theoretically perfect classifier would be a MLP with only one layer with an infinite number of neurons. In the analysis presented in Chapter 4 an MLP with a single hidden layer is used. The number of neurons in the hidden layer is equal to the number of input variables plus five.

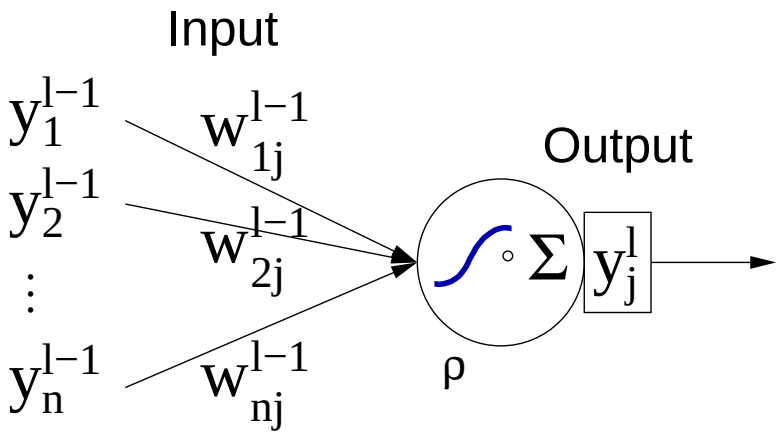


Figure D.1: Schematic description of a single neuron. The neuron is located in layer l, has n connections to the previous layer, and each connection has a weight w_{ij}^{l-1}. Figure taken from [106].

Neuron response function

In the following we use the notation of [106], illustrated in Figure D.1. Each neuron of a given layer is connected to all the neurons from the previous layer. A neuron j in a layer l transforms its input variables y_i^{l-1} provided by the neurons from the layer $l-1$ via n connections with weights w_{ij}^{l-1} into its output y_j^l. The function $(y_1^{l-1}, ..., y_n^{l-1}) \rightarrow y_j^l$ is called the neuron response function. It can be decomposed into the *synapse* function and the *activation* function. The synapse function combines

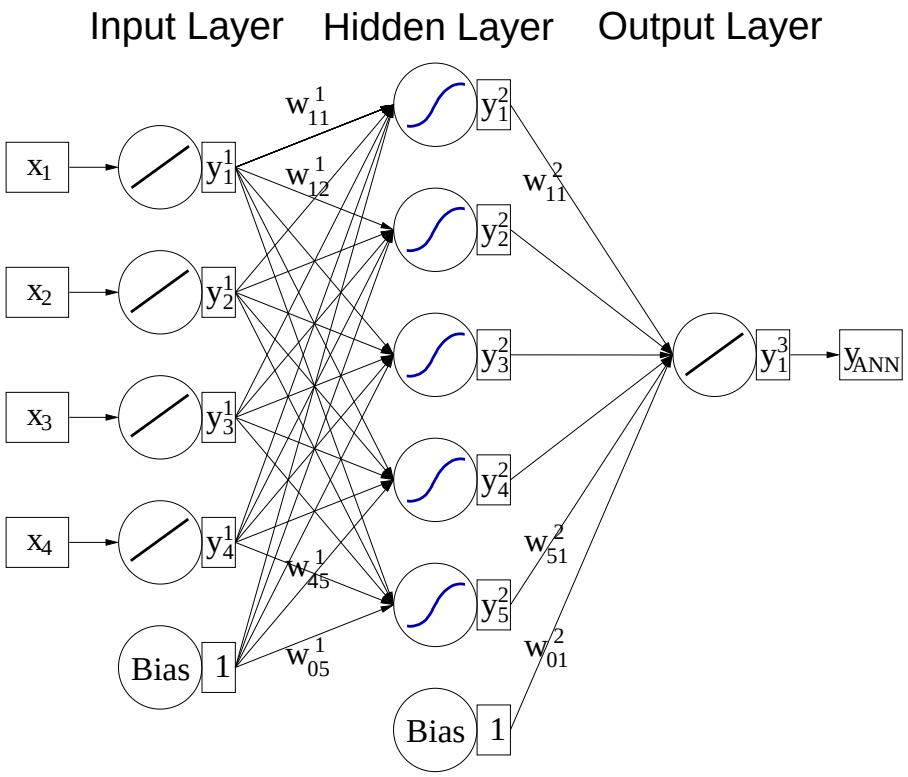


Figure D.2: An example of a multilayer perceptron with one hidden layer. Figure taken from [106].

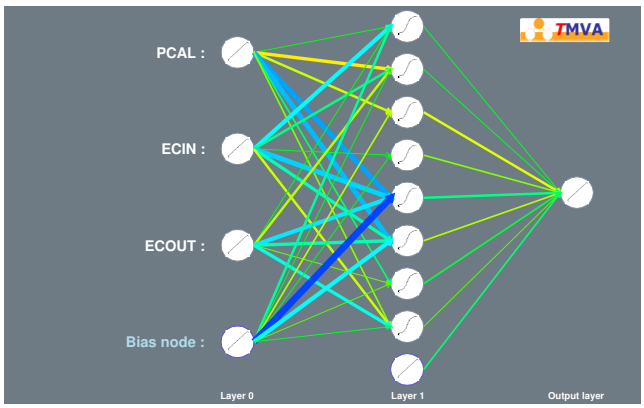


Figure D.3: MLP using three input variables developed for the positron identification algorithm described in Chapter 4.

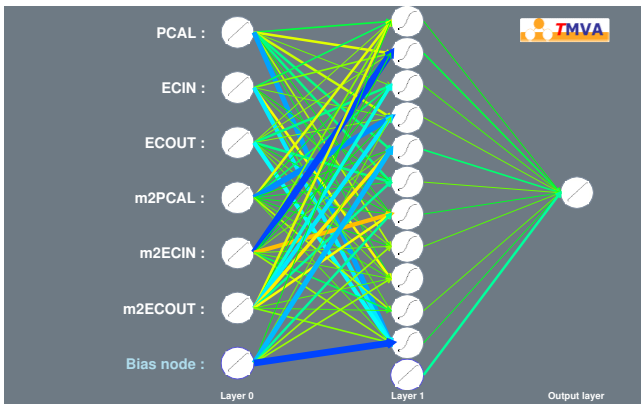


Figure D.4: MLP using six input variables developed for the positron identification algorithm described in Chapter 4.

the input variables and the connections weights in one of the following ways:

$$(y_1^{(\ell)},..,y_n^{(\ell)}|w_{0j}^{(\ell)},..,w_{nj}^{(\ell)}) \rightarrow \begin{cases} w_{0j}^{(\ell)} + \sum_{i=1}^{n} y_i^{(\ell)} w_{ij}^{(\ell)} & \text{Sum,} \\ w_{0j}^{(\ell)} + \sum_{i=1}^{n} \left(y_i^{(\ell)} w_{ij}^{(\ell)}\right)^2 & \text{Sum of squares,} \\ w_{0j}^{(\ell)} + \sum_{i=1}^{n} |y_i^{(\ell)} w_{ij}^{(\ell)}| & \text{Sum of absolutes.} \end{cases} \quad \text{(D.1)}$$

The activation function transforms the output of the synapse function to give the output of the neuron in the desired way:

$$x \to \begin{cases} x & \text{Linear,} \\ \frac{1}{1+e^{-kx}} & \text{Sigmoid,} \\ \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} & \text{Tanh,} \\ e^{-x^2/2} & \text{Radial.} \end{cases} \tag{D.2}$$

In this analysis we used the Sum synapse function and the Tanh activation function as recommended in the TMVA documentation [106].

Training of the neural network with Back-propagation

For a MLP with the previously described architecture (one hidden layer, Sum synapse function and the Tanh activation function), one can write the output of the network as:

$$y_{MLP} = \sum_{j=1}^{n_{\mathrm{h}}} y_j^{(2)} w_{j1}^{(2)} = \sum_{j=1}^{n_{\mathrm{h}}} \tanh\left(\sum_{i=1}^{Nvar} x_i w_{ij}^{(1)} \right) \cdot w_{j1}^{(2)} \,, \tag{D.3}$$

where the notations are the same as in Figure D.2 and N_{var} and n_{h} are the number of neurons in the input layer and in the hidden layer, respectively.

The aim of the training process is to adjust the set of weights w_{ij}^{l} of the network such that the output for all the training events obeys the following rule: if an event is background the output of the MLP is 0, otherwise it is 1. In the training sample the nature of the events (signal or background) is known a priori so one can compare the expected response with the actual response of the network and adjust the weights accordingly. The initial weights are randomly picked. Then the following algorithm is repeated until the weights converge. For an event with input variable vector $\mathbf{x}$ one can calculate the response of the network y_{MLP} and compare it to the expected output $\hat{y}$ (0 or 1). The comparison is performed using the error function E such that:

$$E(\mathbf{x}|\mathbf{w}) = \frac{1}{2} \left(y_{MLP} - \hat{y} \right)^2 \,, \tag{D.4}$$

where $\mathbf{w}$ is the set of weights.

The weights are updated according to the gradient descent method, that is to say in the direction in the $\mathbf{w}$-space where the error decreases the most. After a number of iterations ρ of this process, the weights $\mathbf{w}^{(\rho+1)}$ read:

$$\mathbf{w}^{(\rho+1)} = \mathbf{w}^{(\rho)} - \eta \boldsymbol{\nabla}_{\mathbf{w}} E \,, \tag{D.5}$$

where η is the called the learning rate and it is set to 0.02 in this analysis. Working out Equation (D.5) one can write the variation of weights for each layer. The weights connected with the output layer are updated by

$$\Delta w_{j1}^{(2)} = -\eta \frac{\partial E}{\partial w_{j1}^{(2)}} = -\eta \left(y_{MLP} - \hat{y} \right) y_j^{(2)} \,, \tag{D.6}$$

and the weights connected with the hidden layers are updated by

$$\Delta w_{ij}^{(1)} = -\eta \frac{\partial E}{\partial w_{ij}^{(1)}} = -\eta \left(y_{MLP} - \hat{y} \right) y_j^{(2)} (1 - y_j^{(2)}) w_{j1}^{(2)} x_i \,, \tag{D.7}$$

The weight iteration is done for each event. Thus this method requires to randomize the input events. The training sample is then used repeatedly on the network, each time with a new randomized events sequence.

Appendix E

Training tests from the TMVA package

This appendix shows the training tests performed on the different MVA classifiers used for the positron identification analysis. Each plot shows the output of the corresponding classifier obtained after the training, and the one obtained with the test sample. Both distributions should superimpose to validate the training, which is the case in both the 3D and the 6D analysis of the three classifiers (see Figures E.1, E.2, E.3, E.4, E.5 and E.6).

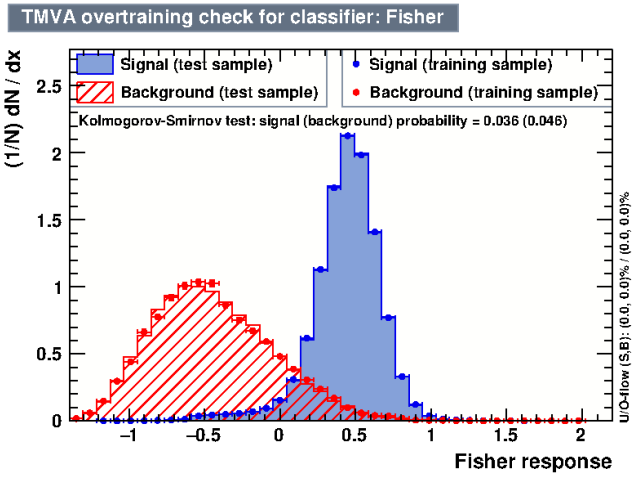


Figure E.1: 3D Fisher training histograms.

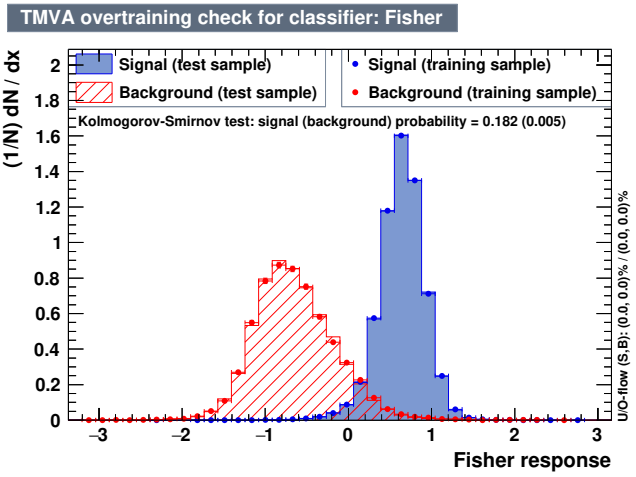


Figure E.2: 6D Fisher training histograms.

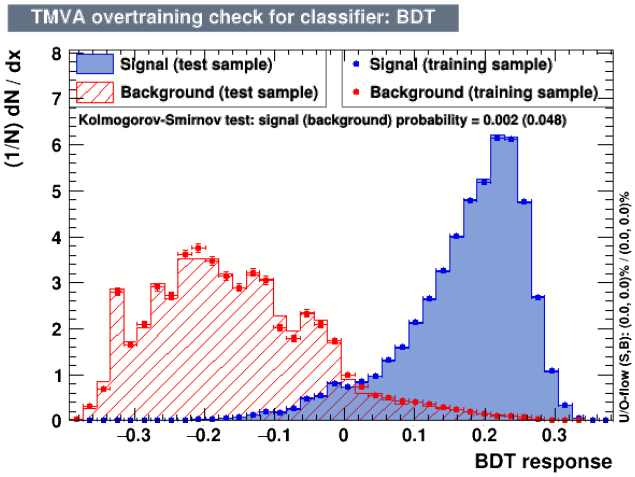


Figure E.3: 3D BDT training histograms.

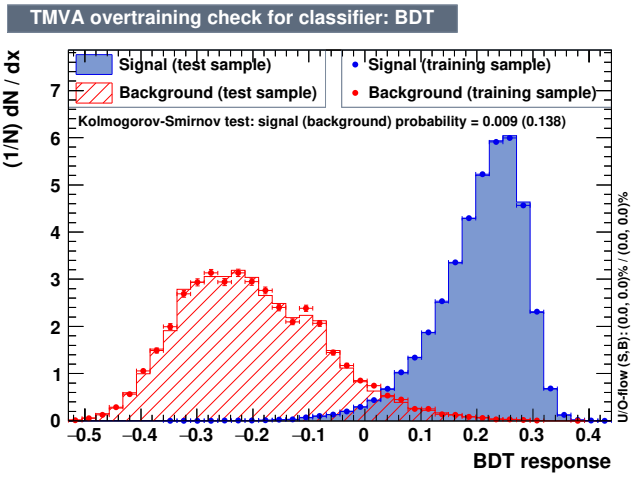


Figure E.4: 6D BDT training histograms.

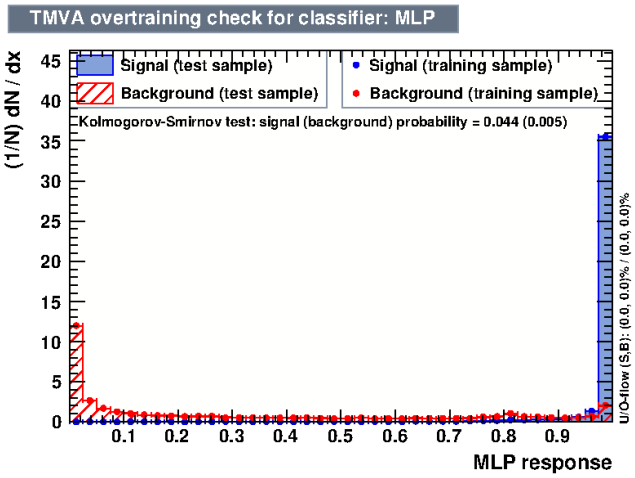


Figure E.5: 3D MLP training histograms.

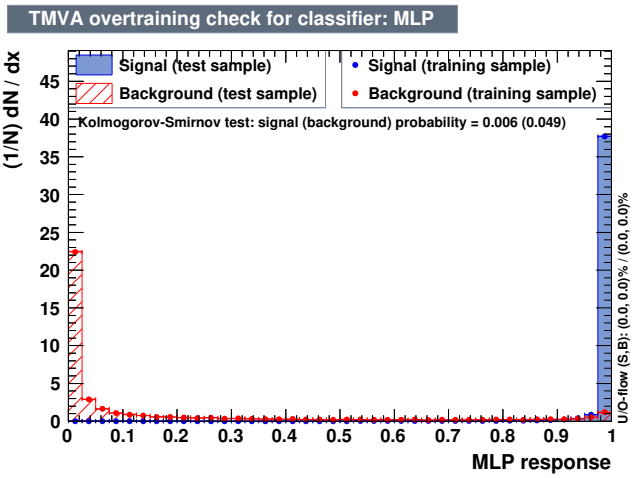


Figure E.6: 6D MLP training histograms.

Appendix F

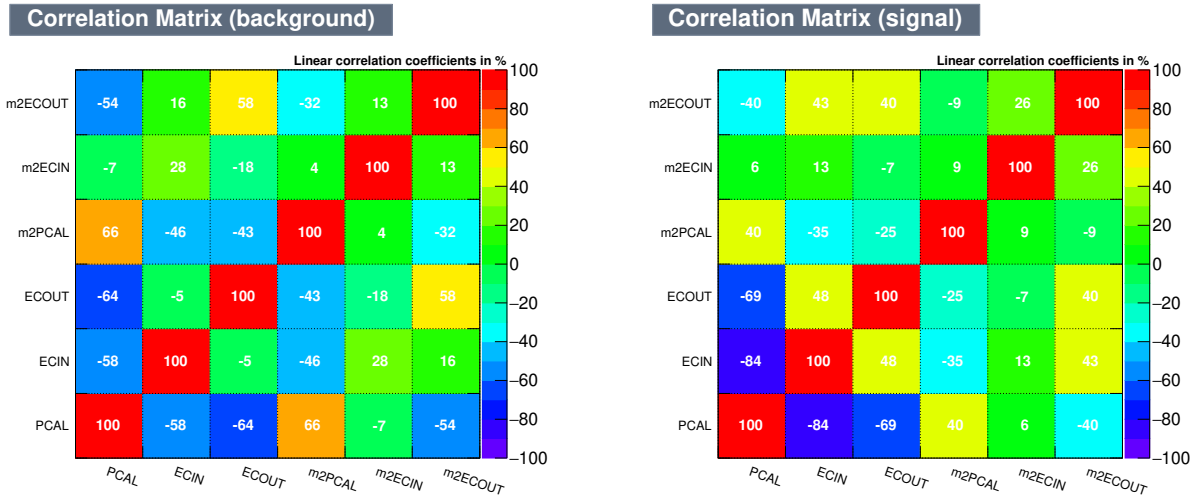
Correlations of the input variables

In this appendix we display the linear correlations between the variables used in the MVA analysis. The linear correlation of two random variables is defined as:

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - E[X])(Y - E[Y])]}{\sqrt{E[(X - E[X])^2]}\sqrt{E[(Y - E[Y])^2]}}, \tag{F.1}$$

where E stands for expected value.

Figure F.1a shows the correlations of the input variables for true positrons, while Figure F.1b is for π^+ that are mistakenly identified as positrons. One can see that the sampling fraction in the PCAL is strongly anti-correlated with the sampling fraction in the ECIN and in the ECOUT, in both cases. This means that if the sampling fraction is high in the PCAL, it will be low in the ECIN/ECOUT; or the opposite situation. This behavior is explained by the fact that positrons tend to deposit most of their energy in a single calorimeter layer. We also observe a large correlation between the sampling fraction in the PCAL and the shower $m2$ in this layer, depicting the relation between a high energy deposition and a large shower radial size.



(a) (b)

Figure F.1: Linear correlation matrix of the 6 input variables for (a) Background (mis-id. pions) and (b) Signal (true-positrons)

Appendix G

Pseudo-ROC curves from data

Figure G.1 corresponds to a zoomed version of the pseudo-ROC curves displayed in Figure 4.16 in Chapter 4. The region of low background and high number of TCS events is selected, in order to compare the performances of the different classifiers. This figure clearly shows the gradual improvements in the suppression of the background, from the χ^2 based approaches, to the 3D MVA techniques and finally the 6D approaches, which are the most efficient in selecting the signal while rejecting the most background.

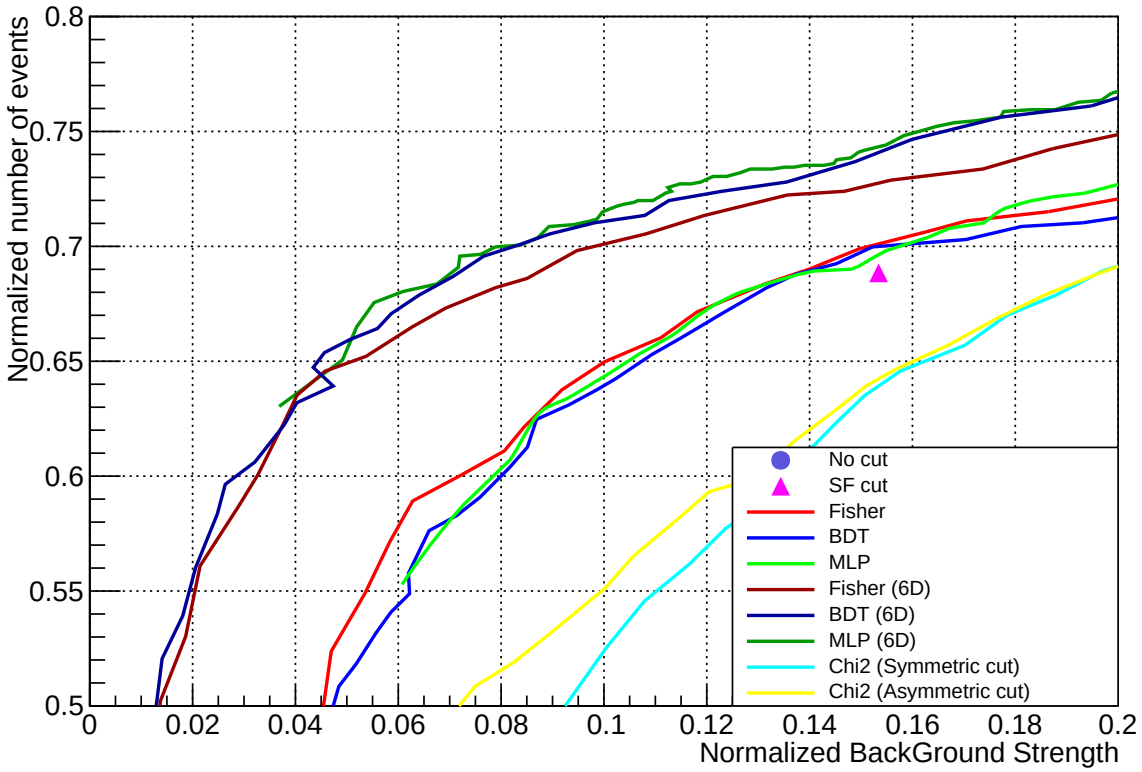


Figure G.1: Zoom on the high signal efficiency and high background rejection region of Figure 4.16.

Appendix H

Derivation of the background/signal ratio

Let $B(x)$ and $S(x)$ respectively be the number of background (mis-id. pions) and signal (true positron) events in the TCS sample for a background strength x ($x \in [0,1]$).

The number of background events in the TCS sample is linear with the background strength x:

$$B(x) = \beta x, \tag{H.1}$$

where β is the number of background events in the TCS sample when no cut is applied.

The function $S(x)$ is unknown but we assume it is increasing with x (when background is removed, signal events might also be removed by mistake), and does not vary much with x (signal events should not be removed by the classifier).

The normalized number of TCS events can then be described in the linear region with the following expression:

$$y(x) = \frac{S(x) + \beta x}{S(1) + \beta}. \tag{H.2}$$

Although this formula is only applicable in the linear region, we can extrapolate it to $x = 0$:

$$y(0) = \frac{S(0)}{S(1) + \beta}. \tag{H.3}$$

Assume the chosen cut yields to a normalized background strength x_0 in the linear region. The corresponding normalized number of TCS events is:

$$y(x_0) = \frac{S(x_0) + \beta x_0}{S(1) + \beta}. \tag{H.4}$$

We want to estimate the background/signal ratio for a given normalized background strength x_0, $\frac{B(x_0)}{S(x_0)}$. Solving $B(x_0)$ from Equations (H.1) and (H.4) one can write for the B/S ratio at x_0:

$$\frac{B(x_0)}{S(x_0)} = y(x_0)\frac{S(1) + \beta}{S(x_0)} - 1 = \frac{y(x_0)}{y(0)}\frac{S(0)}{S(x_0)} - 1\,. \tag{H.5}$$

We can estimate the ratio $\frac{S(x_0)}{S(0)}$ using simulations. It is most of the time very slightly bigger than 1 as seen in Figure 4.20 of Chapter 4. Therefore one finds:

$$\frac{y(x_0)}{y(0)} - 1 \simeq \frac{B(x_0)}{S(x_0)}, \tag{H.6}$$

and the quantity $\frac{y(x_0)}{y(0)} - 1$ gives a good estimate of the ratio $\frac{B(x_0)}{S(x_0)}$.

Appendix I

Lepton radiative corrections

In this appendix Figure I.1 shows the effect of the detected radiated photon correction applied to the simulated electrons, as described in Subsection 4.5.2 of Chapter 4. The plots shown contain only electrons corrected for this effect. The reconstructed kinematics are compared to the generated ones.

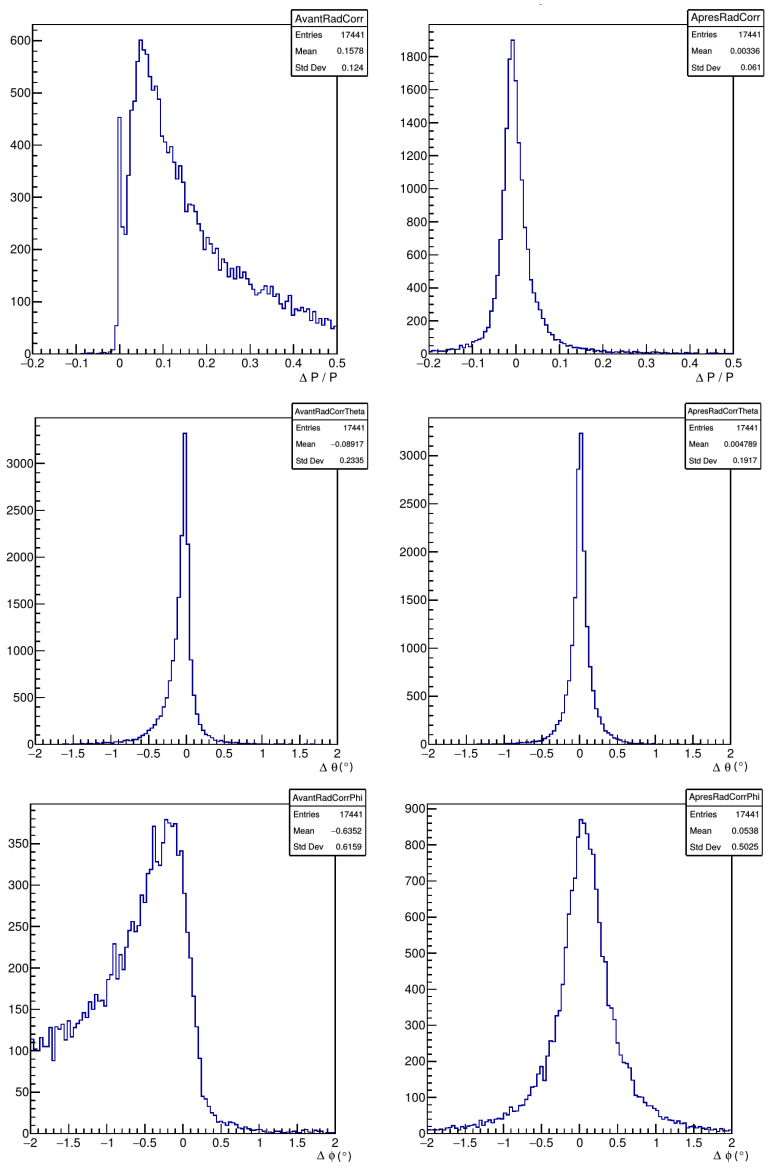


Figure I.1: Distributions of the momentum resolution (top), the θ difference (middle) and the ϕ difference (bottom) before and after applying the detected radiated photon correction to the simulated electrons (in the left and right plots, respectively).

Appendix J

Fiducial Cuts

The fiducial cuts described in Section 4.6 in Chapter 4 are shown in Figures J.1 and J.2 in this appendix, while Figure J.3 compares the electron distributions in the PCAL of CLAS12 data before and after the fiducial cuts are performed.

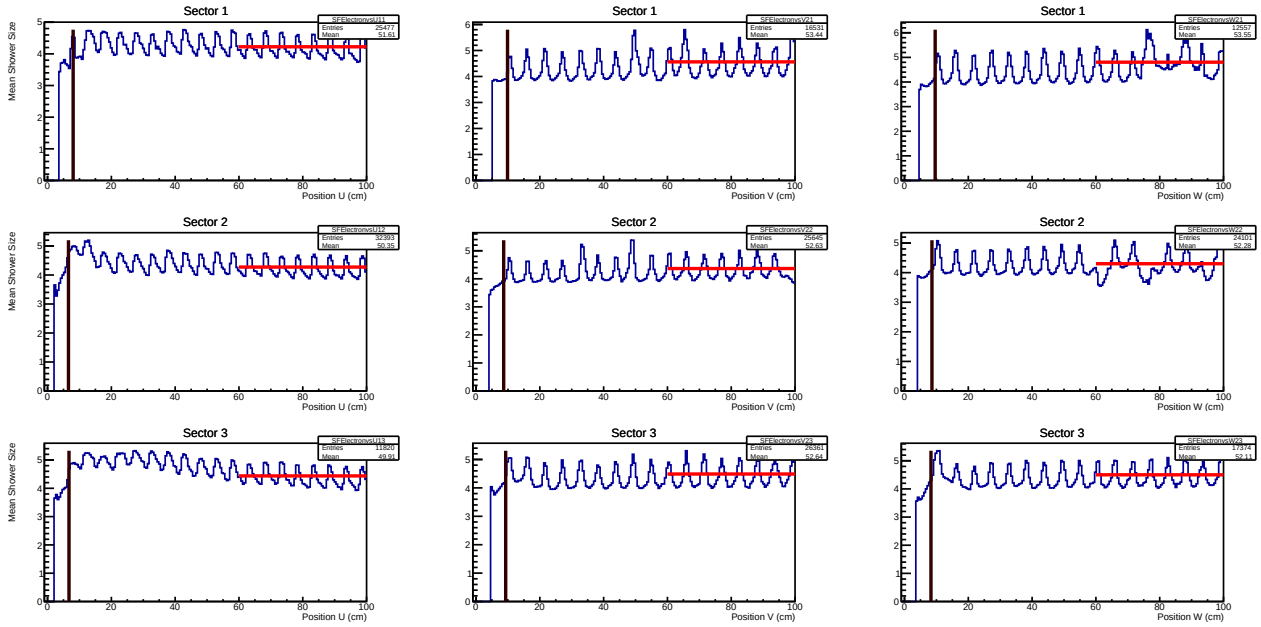


Figure J.1: Mean shower width as a function of U, V and W (low values) for three sectors of CLAS12. The red line corresponds to the mean shower size calculated away from the edge of the distribution. The black line corresponds to the fiducial cut.

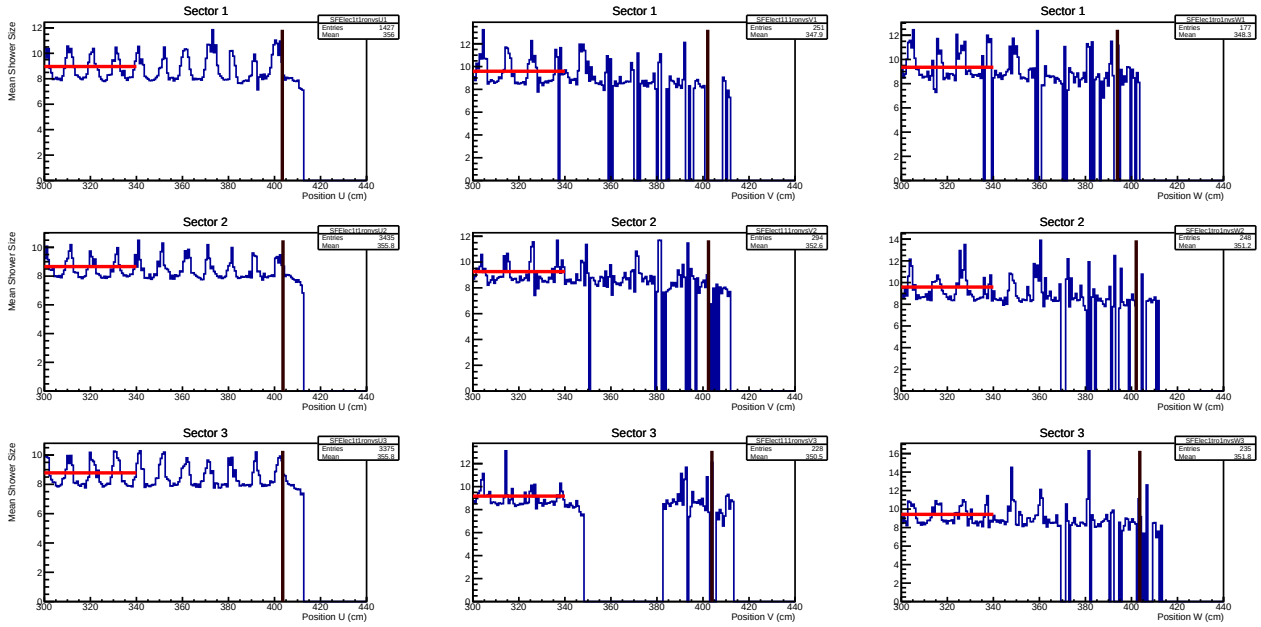


Figure J.2: Same plots as in Figure J.1 for the case of high values of U, V and W.

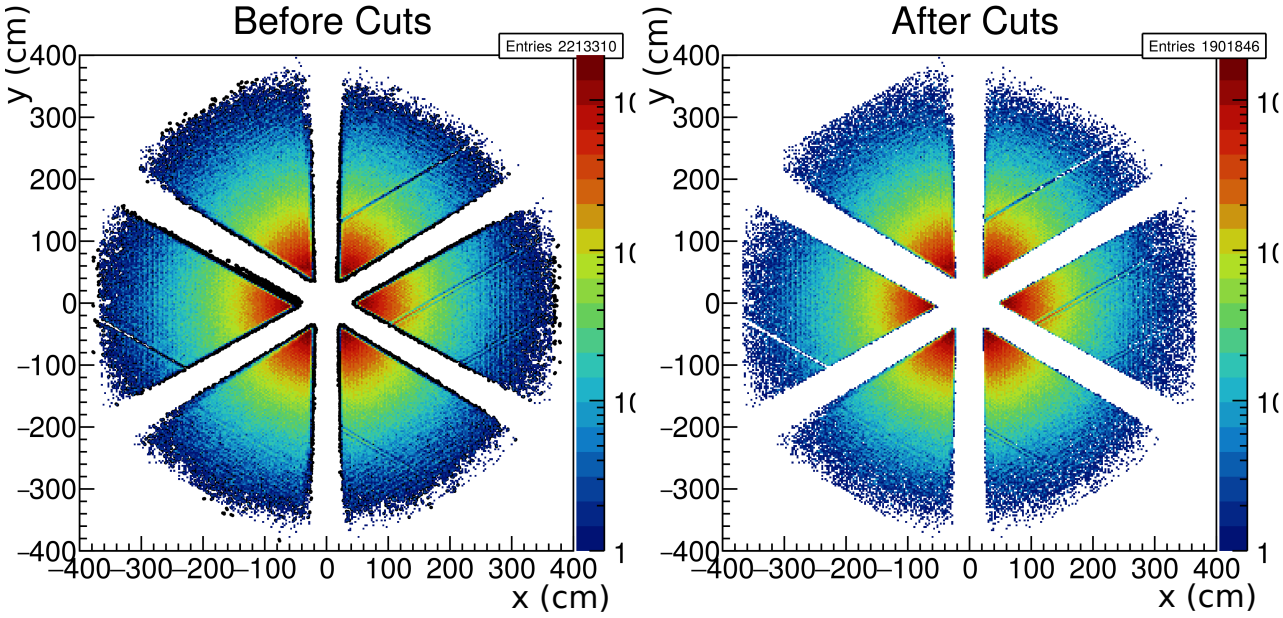


Figure J.3: Effect of the fiducial cuts on the electron distribution in the natural coordinate plane of the PCAL. The left plot shows the electrons which are removed in black. The plot on the right shows all the electrons that passed the PCAL fiducial cuts.

Appendix K

Generator checks: comparison between *GRAPE* and *TCSGen*

Bethe-Heitler events have been generated using *TCSGen* and *GRAPE* within the following phase space:

- Lepton momenta bigger than 1 GeV
- Lepton polar angle in the lab in CLAS12 acceptance $5° < \theta_{\text{Lab } e^{+/-}} < 45°$
- $0.01\ GeV^2 < -t < 1\ GeV^2$
- $1.7\ GeV < M < 3.08\ GeV$
- $2\ GeV < E_\gamma < 10.6\ GeV$
- $40° < \theta_{\text{COM}} < 140°$,

and the results of both generators have been compared.

The following plots in Figures K.1 to K.5 show the distributions of all five relevant TCS variables, obtained for both generators as well as their ratio. Each distribution is normalized by the integral of the E_γ distribution obtained for each generator.

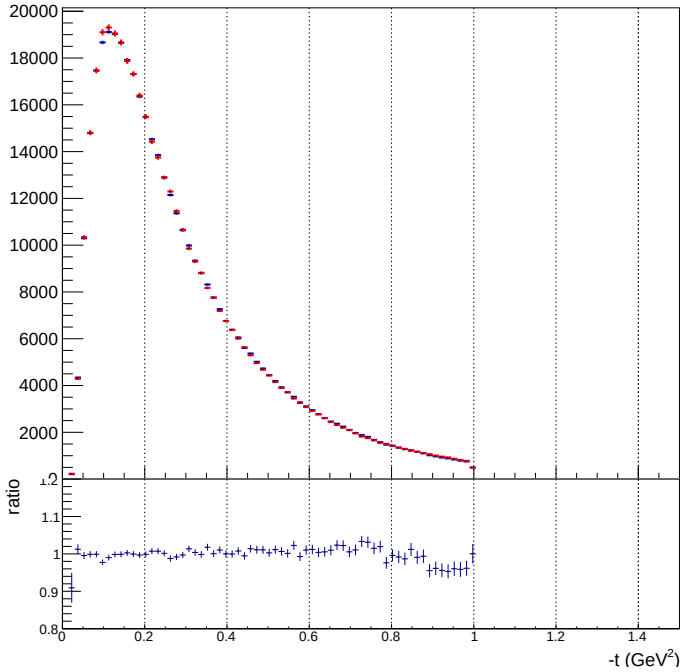


Figure K.1: Comparison of the generated proton momentum transfer squared $-t$ distributions obtained with *TCSGen* (in red) and *GRAPE* (in blue), as well as their ratio.

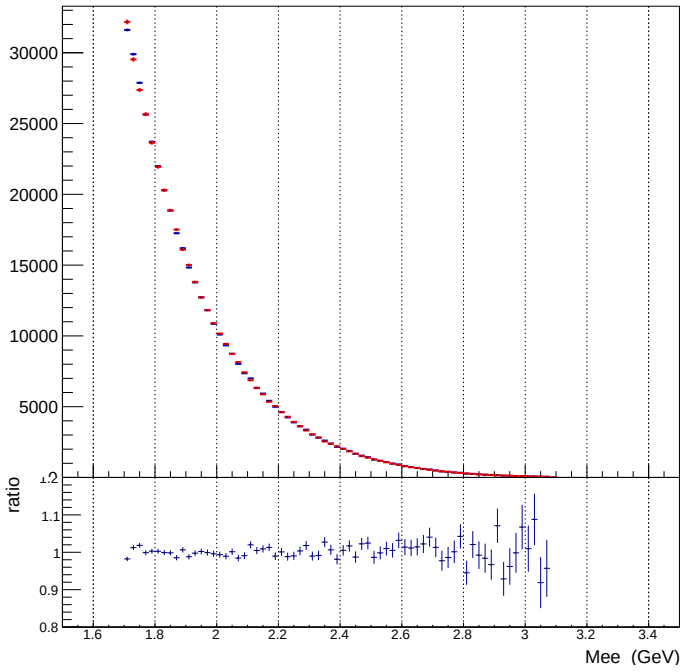


Figure K.2: Comparison of the generated invariant mass of the lepton pair M distributions obtained with *TCSGen* (in red) and *GRAPE* (in blue), as well as their ratio.

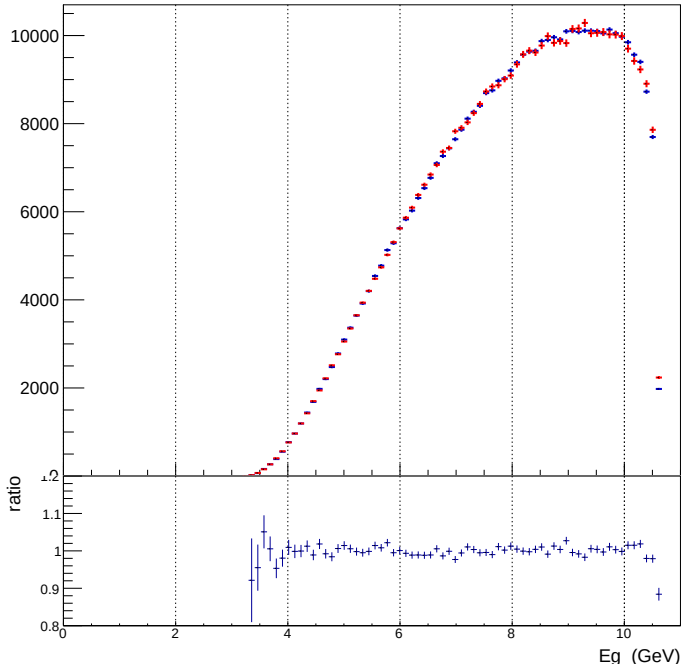


Figure K.3: Comparison of the generated photon energy E_γ distributions obtained with *TCSGen* (in red) and *GRAPE* (in blue), as well as their ratio.

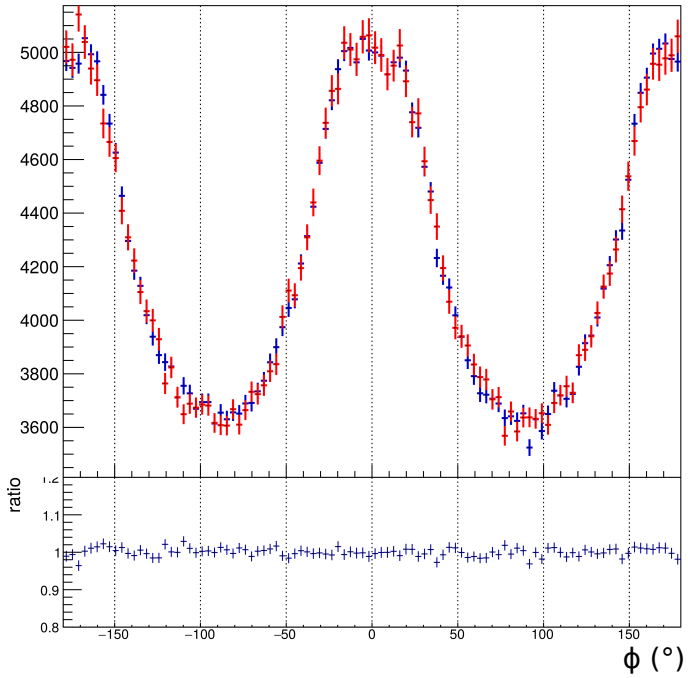


Figure K.4: Comparison of the generated COM azymuthal angle ϕ distributions obtained with *TCSGen* (in red) and *GRAPE* (in blue), as well as their ratio.

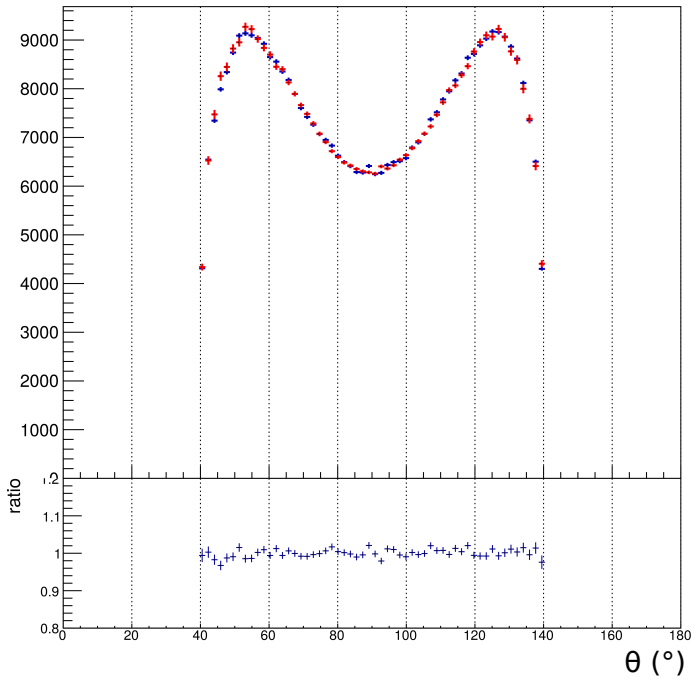


Figure K.5: Comparison of the generated COM polar angle θ distributions obtained with *TCSGen* (in red) and *GRAPE* (in blue), as well as their ratio.

Appendix L

Final state particle kinematics

Figures L.1 and L.2 in this appendix show the kinematics of the three final state particles of the TCS reaction in the CLAS12 data. The top three plots of Figure L.1 show these kinematics for all events in CLAS12 data in which a proton, an electron and a positron are detected. The three bottom plots in the same figure, show the kinematic distributions once the exclusivity cuts presented in Section 5.2 are applied. The plots displayed in Figure L.2 show the kinematics of the final state particles for the events selected in the analysis of this manuscript.

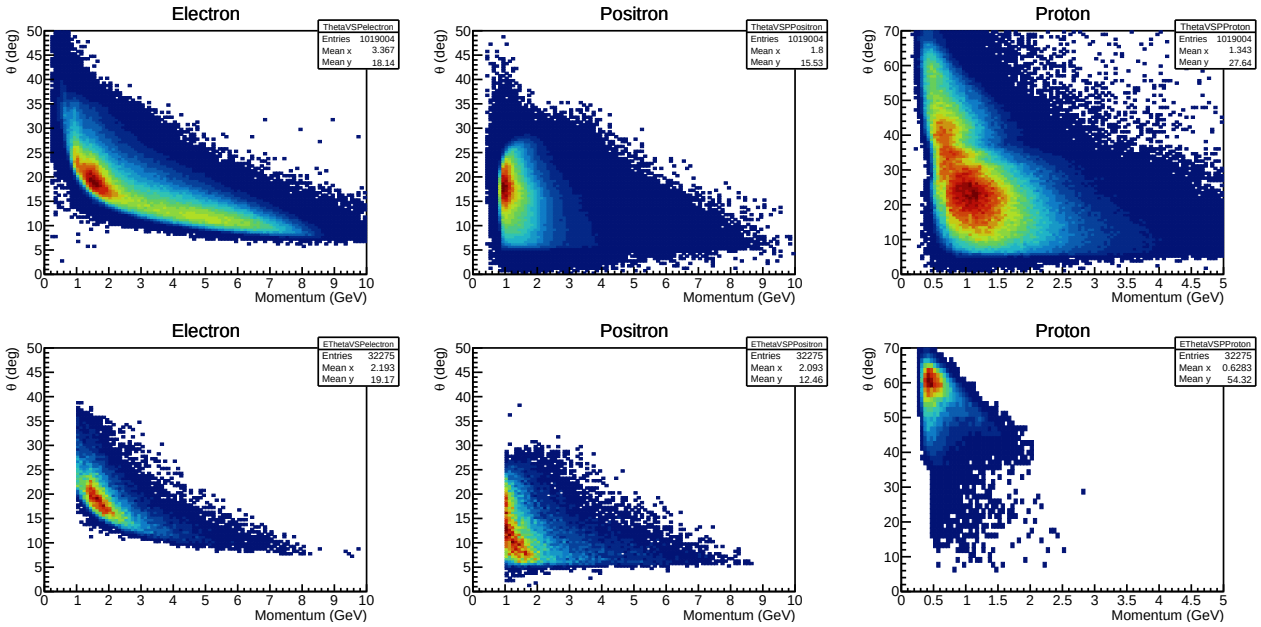


Figure L.1: Top: Kinematics of the TCS final state particles (polar angle in the laboratory frame vs momentum) for events with one proton and two opposite-sign leptons. Bottom: Same plots after the exclusivity cuts to select the photoproduction events are applied.

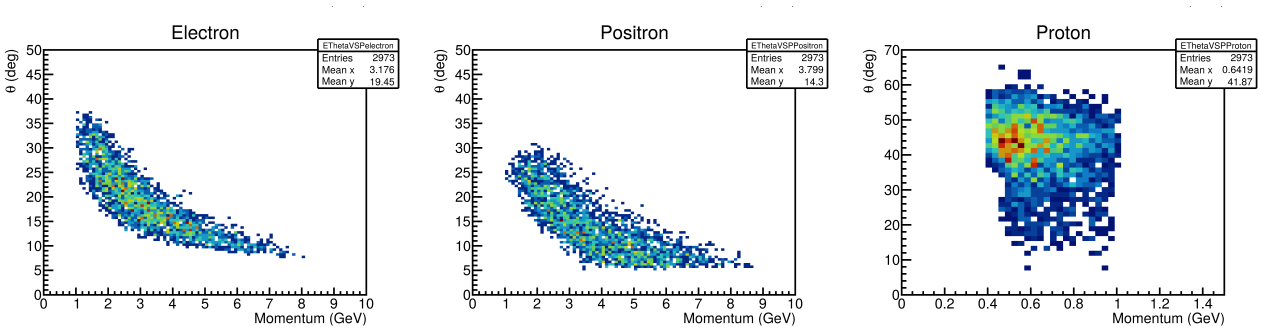


Figure L.2: Kinematics of the final state particles for TCS events with invariant mass between 1.5 and 3 GeV.

Appendix M

CLAS12 Acceptance for TCS

Figures M.1, M.2 and M.3 show the acceptance of CLAS12 for the TCS process. Each plot in the grids corresponds to a single $\{E_\gamma, -t, Q'^2\}$ bin, as defined in Table 5.1 in Chapter 5. The acceptance plots shown here are made prior to applying the acceptance quality cuts described in Section 5.6.

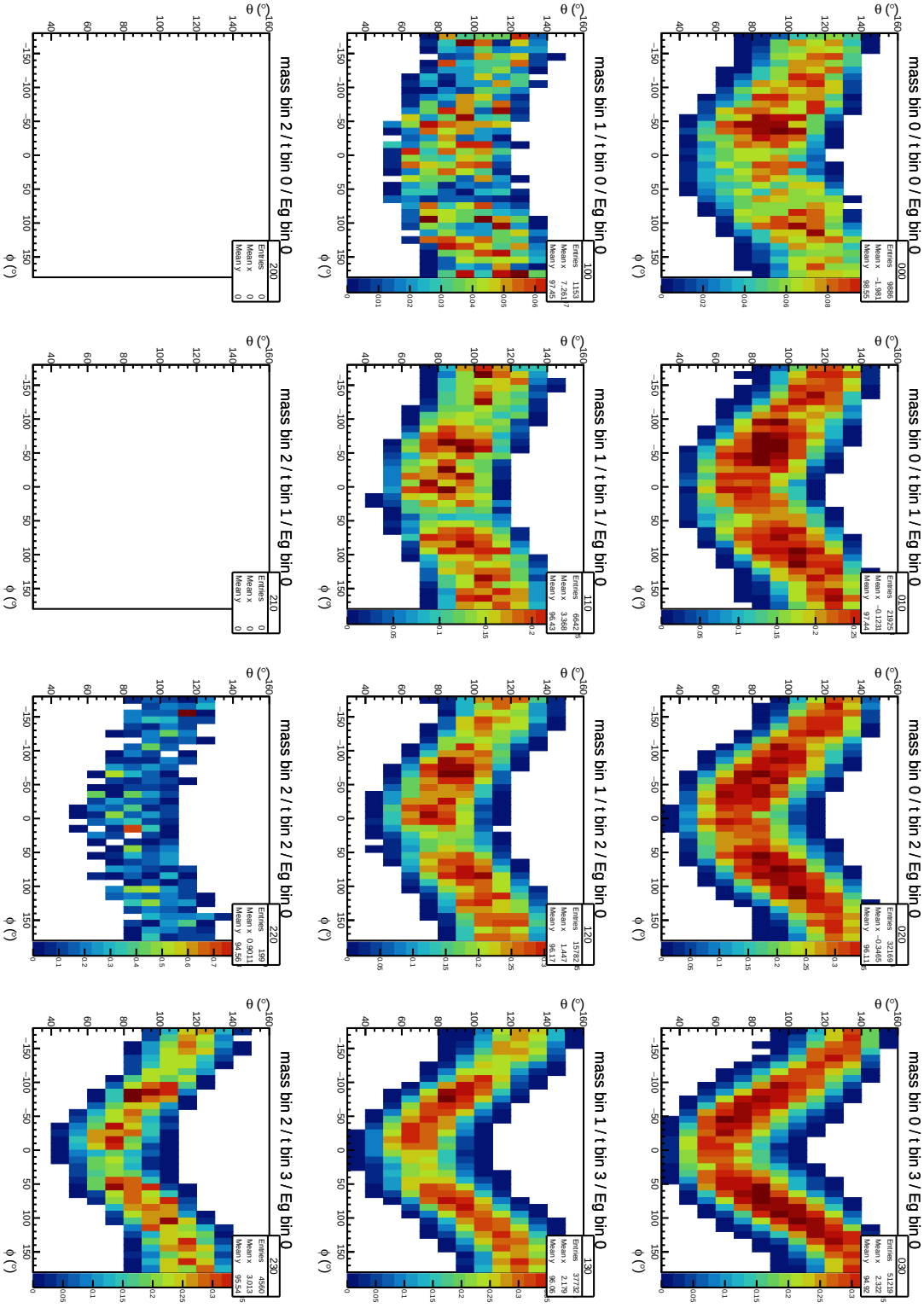


Figure M.1: CLAS12 acceptance for TCS in the ϕ/θ plane for the first bin in E_γ.

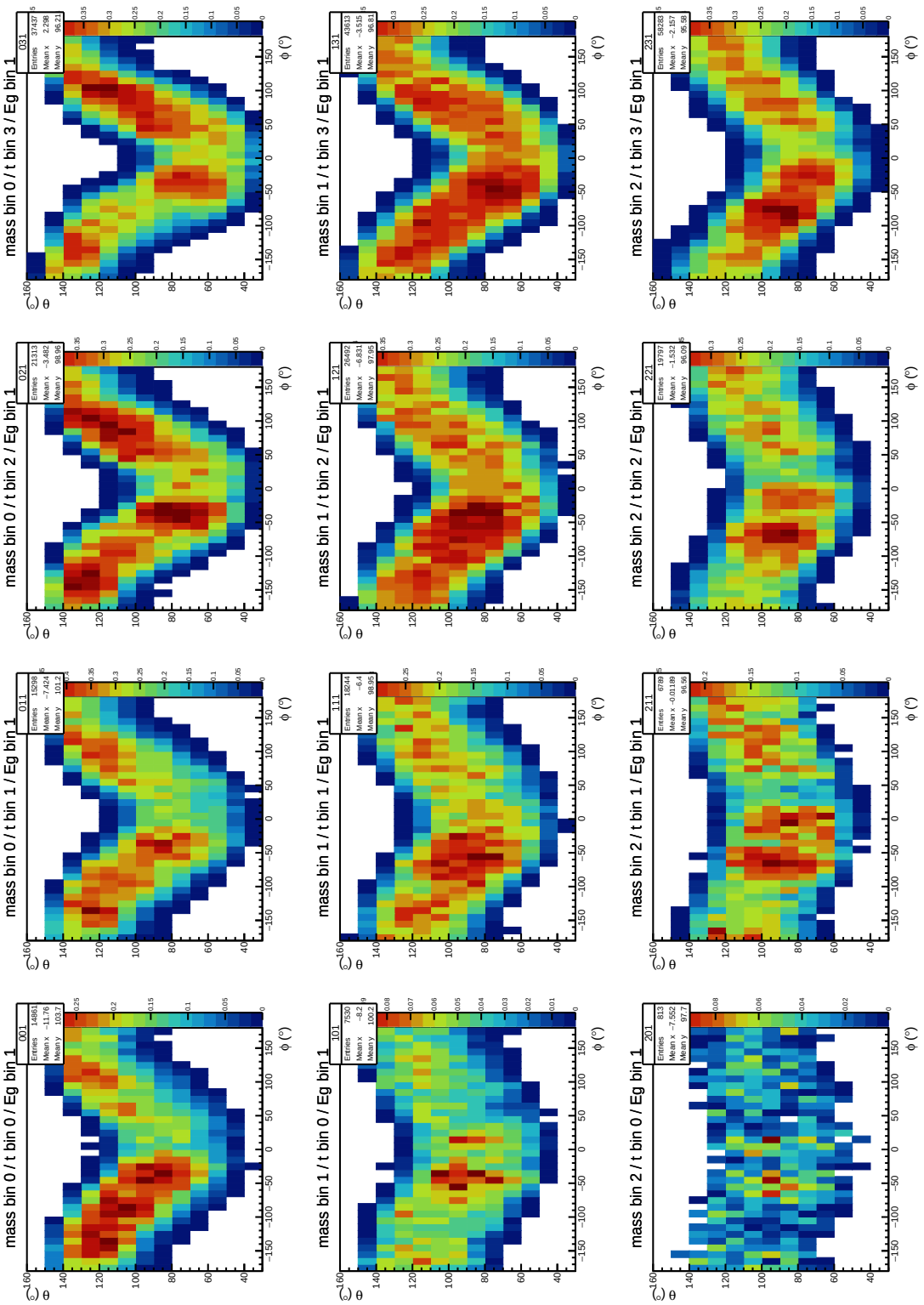


Figure M.2: CLAS12 acceptance for TCS in the ϕ/θ plane for the second bin in E_γ.

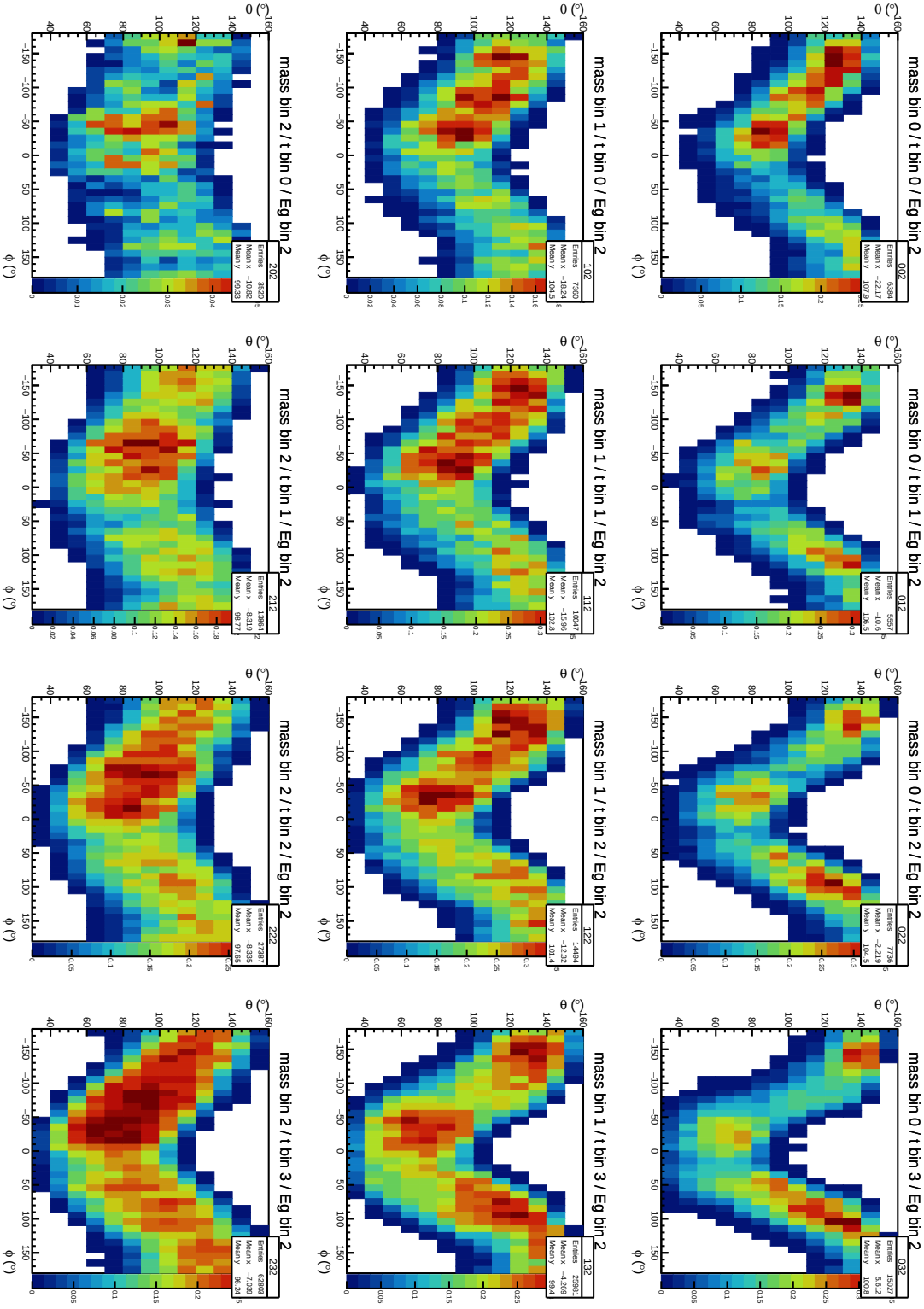


Figure M.3: CLAS12 acceptance for TCS in the ϕ/θ plane for the third bin in E_γ.

Appendix N

Photon polarization transfer

The circular polarization P of the incoming quasi-real photon in the $ep \to (e')\gamma p \to (e')e^-e^+p'$ reaction can be fully calculated in QED. It is given by the helicity S of the electron from the beam that emitted the photon multiplied by a polarization transfer coefficient L. If the electron has positive helicity the photon has right-handed polarization, while a negative helicity electron will emit a left-handed polarized photon. The following formulas are developed in [104], and the notation of [105] is used.

The photon circular polarization P is given as a function of the incoming electron beam helicity S as:

$$P = S\ L, \tag{N.1}$$

$$L = k\left[(E_1 + E_2)(3 + 2\Gamma) - 2E_2(1 + 4u^2\xi^2\Gamma)\right]/I_0, \tag{N.2}$$

$$I_0 = (E_1^2 + E_2^2)(3 + 2\Gamma) - 2E_1E_2(1 + 4u^2\xi^2\Gamma). \tag{N.3}$$

where L is the polarization transfer factor, E_1 and E_2 are respectively the energy of the incident and the scattered electrons and $k = (E_1 - E_2)$ is the energy of the photon. The Coulomb screening factor Γ is given by:

$$\Gamma = \mathcal{F}\left(\frac{\delta}{\xi}\right) - \ln(\delta) - 2 - f(Z), \tag{N.4}$$

where $\delta = k/2E_1E_2$. The factor ξ is calculated as $\xi = 1/(1 + u^2)$, with $u = E_1 \sin\theta_\gamma$, where θ_γ is the angle between the incoming electron and the radiated photon. The function $\mathcal{F}$ accounts for the screening effects. Tabulated values for $\mathcal{F}$ are given in Table N.1. The screening function is also plotted in Figure N.1. Finally the Coulomb correction function $f(Z)$ is given by:

$$f(a) = a^2 \sum_{n=1}^{\infty} \frac{1}{n(n^2 + a^2)}, \tag{N.5}$$

where $a = \alpha Z$, α is the electromagnetic coupling constant and Z is the atomic number of the target material where the radiation of the photon occurs.

Δ	0.5	1	2	4	8	15	20	25	30	35
$\mathcal{F}\left(\frac{\delta}{\xi}\right)$	0.0145	0.0490	0.14	0.3312	0.6758	1.126	1.367	1.564	1.731	1.875
Δ	40	45	50	60	70	80	90	100	120	
$\mathcal{F}\left(\frac{\delta}{\xi}\right)$	2.001	2.114	2.216	2.393	2.545	2.676	2.793	2.897	3.078	

Table N.1: Tabulated values of the screening function $\mathcal{F}$ as a function of $\Delta = (6Z^{-1/3}/121)(\xi/\delta)$.

The polarization transfer function L used for the TCS analysis is plotted as a function of the ratio between the energy of the photon and the energy of the beam in Figure N.2.

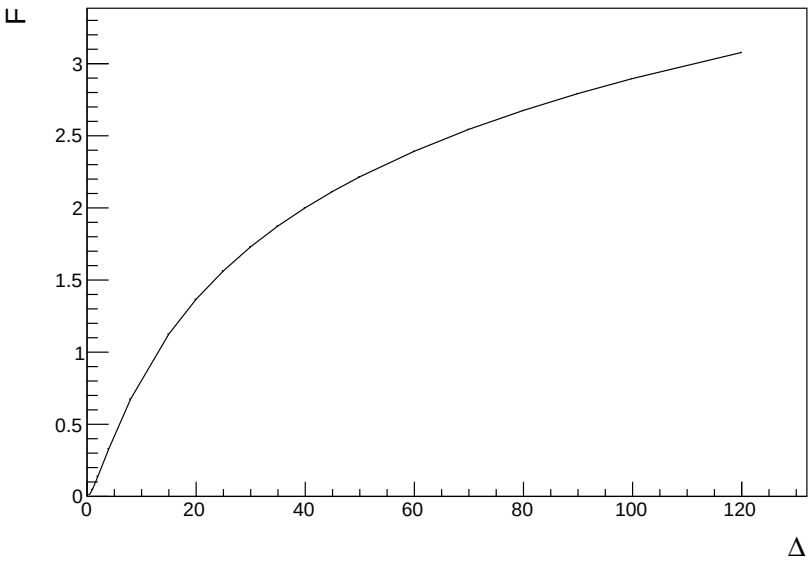


Figure N.1: Screening function $\mathcal{F}$ as a function of Δ.

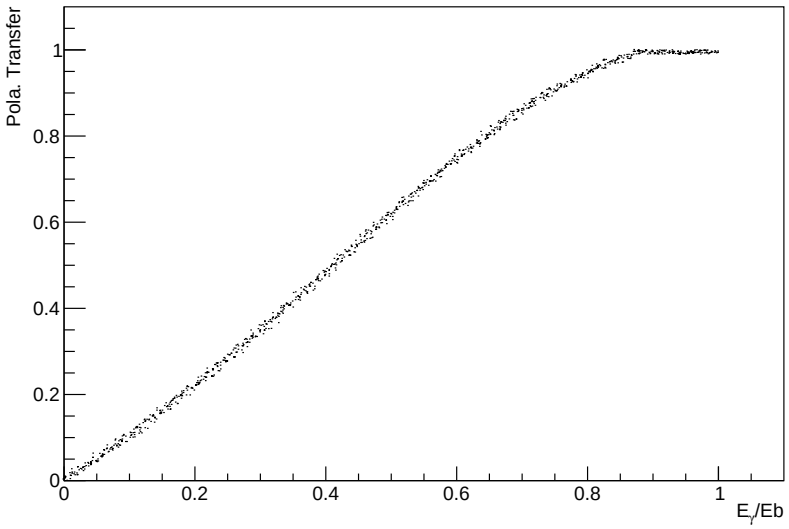


Figure N.2: Polarization transfer function used for the TCS analysis plotted as a function of the ratio between the energy of the radiated photon and the energy of the electron beam. At high energy (for $E_\gamma/E_b > 0.86$), the calculated polarization transfer is bigger than one. The polarization transfer is set to one when this happens.

Appendix O

Tabulated results

The results for the three TCS observables obtained from Fall 2018 CLAS12 dataset (see Chapter 6) are tabulated in this appendix, with the corresponding statistical and systematic uncertainties. Tables O.1, O.2 and O.3 contain results for the R' ratio, Tables O.4 to O.9 display the results for the FB asymmetry and Tables O.10, O.11 and O.12 show the BSA tabulated results.

$-t(GeV^2)$	R'	Stat. error	Low Syst. Error	High Syst. Error
0.209	0.235	0.0339	0.00992	0.0247
0.294	0.203	0.0326	0.0202	0.0202
0.403	0.22	0.0304	0.0042	0.0184
0.606	0.229	0.0292	0.00921	0.0169

Table O.1: R' ratio as a function of $-t$ (see Figure 6.1).

ξ	R'	Stat. error	Low Syst. Error	High Syst. Error
0.101	0.253	0.027	0.00731	0.0231
0.134	0.248	0.0302	0.0135	0.0461
0.196	0.178	0.0264	0.0241	0.00697

Table O.2: R' ratio as a function of ξ (see Figure 6.2).

$-t(GeV^2)$	R'	Stat. error	Low Syst. Error	High Syst. Error
0.278	0.121	0.072	0.0291	0.0229
0.396	0.191	0.0798	0.0428	0.0365
0.501	0.207	0.0734	0.0264	0.0117
0.657	0.134	0.0647	0.0205	0.0156

Table O.3: R' ratio as a function of $-t$ in the mass range [2 GeV $-$ 3 GeV] (see Figure 6.3).

$E\gamma(GeV)$	A_{FB}	Stat. error	Low Syst. Error	High Syst. Error
5.39	0.316	0.115	0.0822	0.0839
7.23	0.33	0.105	0.0322	0.0352
8.9	0.221	0.0988	0.0362	0.0378

Table O.4: A_{FB} as a function of E_γ (see 6.4).

$M(GeV)$	A_{FB}	Stat. error	Low Syst. Error	High Syst. Error
1.6	0.44	0.0884	0.0433	0.117
1.85	0.153	0.105	0.0987	0.0606
2.18	0.301	0.143	0.106	0.0965
2.63	-0.237	0.261	0.0711	0.0359

Table O.5: A_{FB} as a function of M (see 6.5).

$-t(GeV^2)$	A_{FB}	Stat. error	Low Syst. Error	High Syst. Error
0.216	0.394	0.142	0.0482	0.0548
0.298	0.155	0.098	0.0399	0.0394
0.407	0.233	0.0912	0.0367	0.0433
0.61	0.294	0.0992	0.0246	0.0352

Table O.6: A_{FB} as a function of $-t$ (see 6.6).

ξ	A_{FB}	Stat. error	Low Syst. Error	High Syst. Error
0.102	0.282	0.0957	0.034	0.102
0.135	0.305	0.131	0.0559	0.0812
0.198	0.27	0.0924	0.0774	0.0534

Table O.7: A_{FB} as a function of ξ (see Figure 6.7).

$-t(GeV^2)$	A_{FB}	Stat. error	Low Syst. Error	High Syst. Error
0.215	0.426	0.146	0.0482	0.061
0.299	0.138	0.112	0.0609	0.06
0.407	0.251	0.106	0.0639	0.0717
0.603	0.265	0.127	0.028	0.0647

Table O.8: A_{FB} as a function of $-t$ in the mass range [1.5 GeV $-$ 2 GeV] (see 6.8).

$-t(GeV^2)$	A_{FB}	Stat. error	Low Syst. Error	High Syst. Error
0.272	0.125	0.235	0.187	0.181
0.395	0.217	0.223	0.0378	0.00355
0.49	0.426	0.239	0.143	0.0539
0.658	0.335	0.178	0.0685	0.0519

Table O.9: A_{FB} as a function of $-t$ in the mass range [2 GeV $-$ 3 GeV] (see 6.9).

$M(GeV)$	BSA	Stat. error	Low Syst. Error	High Syst. Error
1.59	0.202	0.0589	0.016	0.016
1.84	0.307	0.064	0.0211	0.0234
2.18	0.166	0.0939	0.0346	0.0361
2.63	0.19	0.262	0.126	0.14

Table O.10: BSA as a function of M (see Figure 6.10).

$-t(GeV^2)$	BSA	Stat. error	Low Syst. Error	High Syst. Error
0.209	0.151	0.0848	0.0209	0.0165
0.294	0.307	0.0775	0.0245	0.0236
0.403	0.334	0.0699	0.0236	0.0244
0.607	0.18	0.0681	0.00951	0.0132

Table O.11: BSA as a function of $-t$ (see Figure 6.11).

ξ	BSA	Stat. error	Low Syst. Error	High Syst. Error
0.101	0.25	0.0642	0.0299	0.0329
0.134	0.237	0.0767	0.0117	0.00396
0.196	0.219	0.0627	0.00569	0.0176

Table O.12: BSA as a function of ξ (see Figure 6.12).

Bibliography

[1] Ian Hinchliffe and Aneesh V. Manohar. The qcd coupling constant. *Annual Review of Nuclear and Particle Science*, 50, 04 2000.

[2] M. Thomson. *Modern Particle Physics*, chapter 7 and 8. Cambridge University Press, 2013.

[3] A. W. Thomas and W. Weise. *The Structure of the Nucleon*. WILEY-VCH, Verlag Berlin GmbH, 2001.

[4] M. N. Rosenbluth. High energy elastic scattering of electrons on protons. *Phys. Rev.*, 79:615–619, Aug 1950.

[5] Robert Hofstadter. Electron scattering and nuclear structure. *Rev. Mod. Phys.*, 28:214–254, Jul 1956.

[6] R. Hofstadter, H. R. Fechter, and J. A. McIntyre. High-energy electron scattering and nuclear structure determinations. *Phys. Rev.*, 92:978–987, Nov 1953.

[7] Randolf Pohl et al. The size of the proton. *Nature*, 466:213, 2010.

[8] W. Xiong et al. A small proton charge radius from an electron–proton scattering experiment. *Nature*, 575:147, 2019.

[9] V. Punjabi et al. The structure of the nucleon: Elastic electromagnetic form factors. *The European Physical Journal A*, 51(7):79, Jul 2015.

[10] Véronique Bernard, Latifa Elouadrhiri, and Ulf-G Meißner. Axial structure of the nucleon. *Journal of Physics G: Nuclear and Particle Physics*, 28(1):R1–R35, nov 2001.

[11] M. Breidenbach et al. Observed behavior of highly inelastic electron-proton scattering. *Phys. Rev. Lett.*, 23:935–939, Oct 1969.

[12] J. D. Bjorken and E. A. Paschos. Inelastic electron-proton and γ-proton scattering and the structure of the nucleon. *Phys. Rev.*, 185:1975–1982, Sep 1969.

[13] Richard P. Feynman. Very high-energy collisions of hadrons. *Phys. Rev. Lett.*, 23:1415–1417, Dec 1969.

[14] Yuri L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics. *Sov. Phys. JETP*, 46:641–653, 1977.

[15] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the lhc. *The European Physical Journal C*, 63(2):189–285, Sep 2009.

[16] Y. Goto et al. Polarized parton distribution functions in the nucleon. *Phys. Rev. D*, 62:034017, Jul 2000.

[17] T. Gehrmann and W. J. Stirling. Polarized parton distributions in the nucleon. *Phys. Rev. D*, 53:6100–6109, Jun 1996.

[18] K. Abe et al. Measurements of the proton and deuteron spin structure function g_2 and asymmetry a_2. *Phys. Rev. Lett.*, 76:587–591, Jan 1996.

[19] A. V. Radyushkin. Nonforward parton distributions. *Phys. Rev. D*, 56:5524–5557, Nov 1997.

[20] D. Müller, D. Robaschik, B. Geyer, F.-M. Dittes, and J. Hořejši. Wave functions, evolution equations and evolution kernels from light-ray operators of qcd. *Fortschritte der Physik/Progress of Physics*, 42(2):101–141, 1994.

[21] Xiangdong Ji. Gauge-invariant decomposition of nucleon spin. *Phys. Rev. Lett.*, 78:610–613, Jan 1997.

[22] Michel Guidal, Hervé Moutarde, and Marc Vanderhaeghen. Generalized parton distributions in the valence region from deeply virtual compton scattering. *Reports on Progress in Physics*, 76(6):066202, 2013.

[23] M. Diehl. Generalized parton distributions. *Physics Reports*, 388(2):41 – 277, 2003.

[24] A.V. Belitsky and A.V. Radyushkin. Unraveling hadron structure with generalized parton distributions. *Physics Reports*, 418(1):1 – 387, 2005.

[25] G. Peter Lepage and Stanley J. Brodsky. Exclusive processes in perturbative quantum chromodynamics. *Phys. Rev. D*, 22:2157–2198, Nov 1980.

[26] Matthias Burkardt. Impact parameter space interpretation for generalized parton distributions. *arXiv:hep-ph/0207047*, 2002.

[27] A.V. Belitsky and D. Müller. Nucleon hologram with exclusive leptoproduction. *Nuclear Physics A*, 711(1):118 – 126, 2002. European Workshop on the QCD Structure of the Nucleon.

[28] Raphael Dupre, Michel Guidal, and Marc Vanderhaeghen. Tomographic image of the proton. *Physical Review D*, 95, 06 2016.

[29] R. Dupré, M. Guidal, S. Niccolai, and M. Vanderhaeghen. Analysis of deeply virtual compton scattering data at jefferson lab and proton tomography. *The European Physical Journal A*, 53(8):171, Aug 2017.

[30] Daniel de Florian, Rodolfo Sassot, Marco Stratmann, and Werner Vogelsang. Extraction of spin-dependent parton densities and their uncertainties. *Phys. Rev. D*, 80:034030, Aug 2009.

[31] Alexandre Deur, Stanley J Brodsky, and Guy F de Téramond. The spin structure of the nucleon. *Reports on Progress in Physics*, 82(7):076201, jun 2019.

[32] M.V. Polyakov. Generalized parton distributions and strong forces inside nucleons and nuclei. *Physics Letters B*, 555(1):57 – 62, 2003.

[33] Cédric Lorcé, Hervé Moutarde, and Arkadiusz P. Trawiński. Revisiting the mechanical properties of the nucleon. *The European Physical Journal C*, 79(1):89, Jan 2019.

[34] Maxim Polyakov and Peter Schweitzer. Forces inside hadrons: Pressure, surface tension, mechanical radius, and all that. *International Journal of Modern Physics A*, 33:1830025, 09 2018.

[35] Hyun-Chul Kim, Peter Schweitzer, and Ulugbek Yakhshiev. Energy–momentum tensor form factors of the nucleon in nuclear matter. *Physics Letters B*, 718(2):625 – 631, 2012.

[36] I. Anikin and O. Teryaev. Dispersion relations and subtractions in hard exclusive processes. *Physical Review D*, 76, 05 2007.

[37] B. Pasquini, M.V. Polyakov, and M. Vanderhaeghen. Dispersive evaluation of the d-term form factor in deeply virtual compton scattering. *Physics Letters B*, 739:133 – 138, 2014.

[38] M. Diehl and D. Ivanov. Dispersion representations for hard exclusive processes: Beyond the born approximation. *Eur. Phys. J. C 52(2007),919*, 52, 11 2007.

[39] M. Guidal, M. V. Polyakov, A. V. Radyushkin, and M. Vanderhaeghen. Nucleon form factors from generalized parton distributions. *Phys. Rev. D*, 72:054013, Sep 2005.

[40] S. V. Goloskokov and P. Kroll. An attempt to understand exclusive $\pi+$ electroproduction. *The European Physical Journal C*, 65(1):137, Nov 2009.

[41] M. Polyakov and A. Shuvaev. On "dual" parametrizations of generalized parton distributions. 08 2002.

[42] Dieter Muller and Kresimir Kumericki. Deeply virtual compton scaterring at small xbj. *Modern Physics Letters A*, 24(35n37):2838–2847, 2009.

[43] A. V. Radyushkin. Double distributions and evolution equations. *Phys. Rev. D*, 59:014030, Dec 1998.

[44] A.V. Radyushkin. Symmetries and structure of skewed and double distributions. *Physics Letters B*, 449(1):81 – 88, 1999.

[45] M. V. Polyakov and C. Weiss. Skewed and double distributions in the pion and the nucleon. *Phys. Rev. D*, 60:114017, Nov 1999.

[46] M. Vanderhaeghen, P. A. M. Guichon, and M. Guidal. Hard electroproduction of photons and mesons on the nucleon. *Phys. Rev. Lett.*, 80:5064–5067, Jun 1998.

[47] M. Vanderhaeghen, P. A. M. Guichon, and M. Guidal. Deeply virtual electroproduction of photons and mesons on the nucleon: Leading order amplitudes and power corrections. *Phys. Rev. D*, 60:094017, Oct 1999.

[48] C. Lorcé, B. Pasquini, and M. Vanderhaeghen. Unified framework for generalized and transverse-momentum dependent parton distributions within a 3q light-cone picture of the nucleon. *Journal of High Energy Physics*, 2011(5):41, May 2011.

[49] Bernard Pire. Generalized parton distributions and generalized distribution amplitudes: New tools for hadronic physics. *Annales Henri Poincaré*, 4(1):243–246, Dec 2003.

[50] J. P. Lansberg, B. Pire, and L. Szymanowski. *Transition Distribution Amplitudes*, pages 367–374.

[51] L. Szymanowski, B. Pire, and Kirill Semenov-Tian-Shansky. Transition distribution amplitudes : from jlab to eic. 10 2019.

[52] Kyung-Soon Park et al. Hard exclusive pion electroproduction at backward angles with clas. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, 780:340–345, 05 2018.

[53] Nicole d'Hose, Silvia Niccolai, and Armine Rostomyan. Experimental overview of deeply virtual compton scattering. *The European Physical Journal A*, 52(6):151, Jun 2016.

[54] Baptiste Guegan. *Study of Generalized Parton Distributions and Deeply Virtual Compton Scattering on the nucleon with the CLAS and CLAS12 detectors at the Jefferson Laboratory.* PhD thesis, Universite Paris-Sud, Rensselaer Polytechnic Institute, 2012.

[55] A.V. Belitsky, D. Müller, L. Niedermeier, and A. Schäfer. Leading twist asymmetries in deeply virtual compton scattering. *Nuclear Physics B*, 593(1):289 – 310, 2001.

[56] V. D. Burkert, L. Elouadrhiri, and F. X. Girod. The pressure distribution inside the proton. *Nature*, 557(7705):396–399, May 2018.

[57] E.R. Berger, M. Diehl, and B. Pire. Timelike compton scattering: exclusive photoproduction of lepton pairs. *The European Physical Journal C - Particles and Fields*, 23(4):675–689, Apr 2002.

[58] Boër, M., Guidal, M., and Vanderhaeghen, M. Timelike compton scattering off the proton and generalized parton distributions. *Eur. Phys. J. A*, 51(8):103, 2015.

[59] R. Paremuzyan. *Timelike Compton Scattering.* PhD thesis, Yerevan Physics Institute, 2010.

[60] Markus Diehl, Thierry Gousset, Bernard Pire, and John P. Ralston. Testing the handbag contribution to exclusive virtual compton scattering. *Physics Letters B*, 411(1):193 – 202, 1997.

[61] Oleksii Gryniuk and Marc Vanderhaeghen. Accessing the real part of the forward $j/\psi - p$ scattering amplitude from j/ψ photoproduction on protons around threshold. *Phys. Rev. D*, 94:074001, Oct 2016.

[62] D. Mueller, B. Pire, L. Szymanowski, and Jakub Wagner. Timelike and spacelike hard exclusive reactions. *Physical Review D*, 86, 03 2012.

[63] A.V. Belitsky, D. Müller, L. Niedermeier, and A. Schäfer. Deeply virtual compton scattering in next-to-leading order. *Physics Letters B*, 474(1):163 – 169, 2000.

[64] H. Moutarde, B. Pire, F. Sabatié, L. Szymanowski, and J. Wagner. Timelike and spacelike deeply virtual compton scattering at next-to-leading order. *Phys. Rev. D*, 87:054029, Mar 2013.

[65] B. Pire, L. Szymanowski, and J. Wagner. Next-to-leading order corrections to timelike, spacelike, and double deeply virtual compton scattering. *Phys. Rev. D*, 83:034009, Feb 2011.

[66] B. Pire, L. Szymanowski, and J. Wagner. Can one measure timelike compton scattering at lhc? *Phys. Rev. D*, 79:014010, Jan 2009.

[67] Reza Kazimi. Simultaneous Four-hall Operation for 12 GeV CEBAF. In *4th International Particle Accelerator Conference*, page THPFI091, 2013.

[68] V.D. Burkert et al. The clas12 spectrometer at jefferson laboratory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 959:163419, 2020.

[69] R. Fair et al. The clas12 superconducting magnets. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 962:163578, 2020.

[70] N. Baltzell et al. The clas12 beamline and its performance. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 959:163421, 2020.

[71] M.A. Antonioli et al. The clas12 silicon vertex tracker. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 962:163701, 2020.

[72] A. Acker et al. The clas12 micromegas vertex tracker. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 957:163423, 2020.

[73] D.S. Carman et al. The clas12 central time-of-flight system. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 960:163626, 2020.

[74] P. Chatagnon et al. The clas12 central neutron detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 959:163441, 2020.

[75] M.D. Mestayer et al. The clas12 drift chamber system. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 959:163518, 2020.

[76] D.S. Carman et al. The clas12 forward time-of-flight system. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 960:163629, 2020.

[77] G. Asryan et al. The clas12 forward electromagnetic calorimeter. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 959:163425, 2020.

[78] M. Ungaro et al. The clas12 low threshold cherenkov detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 957:163420, 2020.

[79] M. Contalbrigo et al. The clas12 ring imaging cherenkov detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, page 163791, 2020.

[80] A. Acker et al. The clas12 forward tagger. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 959:163475, 2020.

[81] B. Raydo et al. The clas12 trigger system. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 960:163529, 2020.

[82] V. Ziegler et al. The clas12 software framework and event reconstruction. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 959:163472, 2020.

[83] M. Ungaro et al. The clas12 geant4 simulation. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 959:163422, 2020.

[84] S. Niccolai et al. The central neutron detector for clas12. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 904:81 – 92, 2018.

[85] Deeply virtual compton scattering on the neutronwith clas12 at 11 gev. `https://www.jlab.org/exp_prog/proposals/11/PR12-11-003.pdf`. Accessed: 2020-04-01.

[86] Hall b clas12 run group b experiment summaries. `https://misportal.jlab.org/mis/physics/experiments/downloadOnePageSummary.cfm/summary.pdf?paperId=510`. Accessed: 2020-04-01.

[87] S. Niccolai and V. Kubarovsky. Deeply virtual compton scattering on the neutron with clas12 ar 11 gev for jlab pac38.

[88] M. Benali et al. Deeply virtual compton scattering off the neutron. *Nature Physics*, 16(2):191–198, Feb 2020.

[89] M. Hattawy et al. Neutron dvcs measurements with bonus12 in clas12.

[90] D. Androić et al. The g0 experiment: Apparatus for parity-violating electron scattering measurements at forward and backward angles. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 646(1):59 – 86, 2011.

[91] Gan'elec constant fraction discriminator fcc8. `https://wwwusers.ts.infn.it/~rui/univ/Acquisizione_Dati/Manuals/GANELEC%20CFD%20FCC8.pdf`. Accessed: 2020-04-01.

[92] William H. E. Day and Herbert Edelsbrunner. Efficient algorithms for agglomerative hierarchical clustering methods. *Journal of Classification*, 1, 1984.

[93] Y.G. Sharabian et al. The clas12 high threshold cherenkov counter. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, page 163824, 2020.

[94] Andreas Hoecker et al. Tmva: Toolkit for multivariate data analysis. *PoS*, ACAT:040, 2007.

[95] CLAS collaboration. Fiducial studies. `https://clasweb.jlab.org/wiki/index.php/Run_Group_A#tab=Fiducial_Studies`, 2020.

[96] Tetsuo Abe. Grape-dilepton (version 1.1): A generator for dilepton production in ep collisions. *Computer Physics Communications*, 136(1):126 – 147, 2001.

[97] Paul Kessler. Photon fluxes and the EPA. In *Two-Photon Physics at LEP and HERA - Status of Data and Models*, pages 183–190, 5 1994.

[98] Eero Byckling and K. Kajantie. *Particle Kinematics*. University of Jyvaskyla, Jyvaskyla, Finland, 1971.

[99] CLAS collaboration. Genev webpage. `https://clasweb.jlab.org/wiki/index.php/Event_generator`, 2014.

[100] S. Stepanya at al. Clas12 fd charge particle reconstruction efficiency and the beam background merging. private communication, to be released.

[101] Pawel Sznajder. private communication.

[102] B. Berthou, D. Binosi, N. Chouika, L. Colaneri, M. Guidal, C. Mezrag, H. Moutarde, J. Rodríguez-Quintero, F. Sabatié, P. Sznajder, and J. Wagner. Partons: Partonic tomography of nucleon software. *The European Physical Journal C*, 78(6):478, Jun 2018.

[103] K. Goeke, M.V. Polyakov, and M. Vanderhaeghen. Hard exclusive reactions and the structure of hadrons. *Progress in Particle and Nuclear Physics*, 47(2):401 – 515, 2001.

[104] Haakon Olsen and L. C. Maximon. Photon and electron polarization in high-energy bremsstrahlung and pair production with screening. *Phys. Rev.*, 114:887–904, May 1959.

[105] Jonathan Dumas. *Feasibility studies of a polarized positron source based on the bremsstrahlung of polarized electrons.* PhD thesis, Université Grenoble Alpes, 2011.

[106] A. Hoecker et al. Tmva - toolkit for multivariate data analysis, 2007.

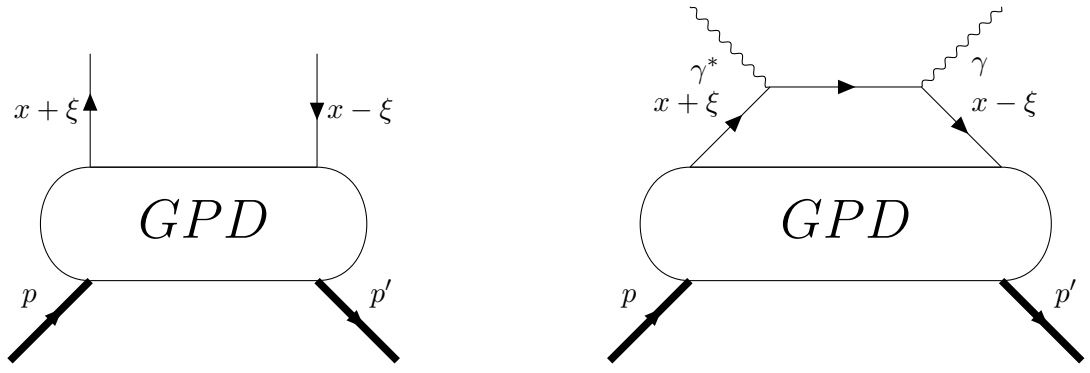
[107] Byron P. Roe et al. Boosted decision trees as an alternative to artificial neural networks for particle identification. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 543(2):577 – 584, 2005.

Résumé en Français

Introduction

La majorité de la matière visible de l'univers est formée d'atomes. Le noyau atomique est lui même composé de nucléons, les protons et les neutrons. Les découvertes successives du proton en 1919 par Rutherford, du neutron en 1932 par Chadwick, et plus tard la mise en évidence de la structure composite des nucléons par les expériences de diffusion profondément inélastique (DIS) en 1969 à SLAC, forment les bases de la physique hadronique. Ce domaine de la physique a pour but la compréhension des interactions entre quarks et gluons, les constituants fondamentales des nucléons.

La théorie qui décrit l'interaction forte entre les quarks et les gluons (aussi appelés partons) est la Chromodynamique quantique (QCD). Cette théorie possède un caractère différent suivant les énergies considérées. À hautes énergies, la QCD est calculable par l'approche perturbative car la constante de couplage de l'interaction forte est petite: c'est le phénomène de la liberté asymptotique. Cependant cette approche n'est plus possible à faibles énergies, par exemple pour des énergies similaires à la masse des nucléons. Dans ce cas, il est nécessaire d'introduire des fonctions de structure pour décrire les interactions entre partons. Historiquement, les premières fonctions de structure introduites sont les facteurs de forme (FF). Ces fonctions décrivent la distribution spatiale des partons dans le nucléon. Les FFs sont accessibles dans les expériences de diffusion élastique de leptons sur des protons par exemple. Un second ensemble de fonctions de structure sont les fonctions de distribution des partons (PDF). Ces fonctions décrivent la probabilité de trouver, à l'intérieur d'un nucléon, un quark ou un gluon possédant une certaine fraction x de l'impulsion du nucléon.



(a) (b)

Figure Résumé-Fr-1: (a) Représentation diagrammatique des GPDs. Les GPDs dépendent des variables x, ξ et $t = (p' - p)^2$ présentées dans ce diagramme. (b) Représentation diagrammatique de la réaction DVCS. Un photon virtuel interagit avec un unique quark, possédant une fraction $x + \xi$ de l'impulsion p du proton initial. Un photon réel est émis et le proton reste intact durant la réaction.

Dans cette thèse, je présente ensuite une troisième famille de fonctions de structure, les distributions généralisées de parton (GPD). Ces fonctions ont été introduites dans le courant des années 1990, dans le but d'étendre le concept des FFs et des PDFs. Il y a quatre GPDs conservant l'hélicité du quark, H, $\tilde{H}$, E et $\tilde{E}$. Ces GPDs peuvent être représentées par la Figure Résumé-Fr-1a. Elles peuvent

être reliées, de manière indépendante des modèles, aux FFs et PDFs. La réaction la plus simple et qui est principalement utilisée expérimentalement pour accéder aux GPDs est la diffusion Compton profondément virtuelle sur proton (DVCS, $ep \to e'\gamma^* p \to e'p'\gamma$). La Figure Résumé-Fr-1b représente le diagramme de Feynman associé à DVCS. Les GPDs apparaissent dans les formules de sections efficaces du DVCS au sein de quantités complexes, les facteurs de forme Compton (CFFs).

Cette thèse est ensuite composée de trois parties. Dans la première partie, je présente le détecteur CLAS12 situé à Jefferson Lab, aux États-Unis, et dont les données sont utilisées dans ce travail. La partie suivante présente le détecteur central de neutron de CLAS12. Enfin la dernière partie du manuscrit est consacrée à l'analyse des données de CLAS12, dans le but d'extraire des observables de la réaction de diffusion Compton de genre temps.

JLab et le détecteur CLAS12

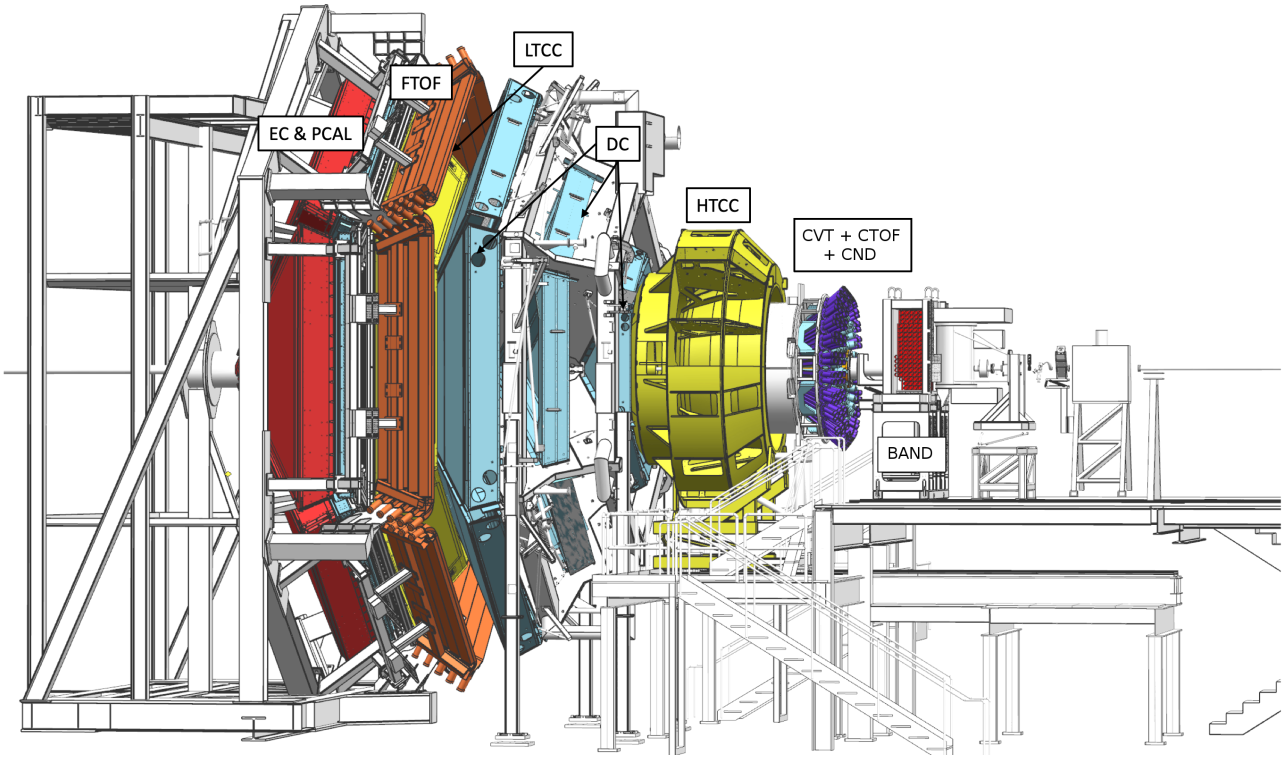


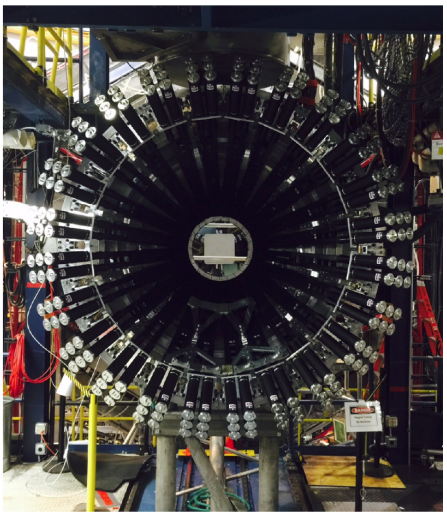
Figure Résumé-Fr-2: Le détecteur CLAS12 est situé dans le Hall B de JLab. Le faisceau d'électron délivré par le CEBAF arrive depuis la droite de l'image, et interagit avec la cible située dans le détecteur central. Les particules possédant un faible angle polaire sont détectées par le détecteur avant. Il est composé d'un aimant toroïdal, d'un système de chambres à dérive (DC) qui mesurent l'impulsion des particules, de détecteurs de temps de vol (FTOF), de calorimètres (EC et PCAL) et de compteurs Cherenkov (HTCC et LTCC) pour l'identification des particules. Les particules émises à grands angles polaires, sont détectées dans le détecteur central construit à l'intérieur d'un aimant solenoïdal. Un système de détection de trace (CVT) permet la mesure de l'impulsion. Le détecteur central de temps de vol (CTOF) permet l'identification des particules. Enfin le détecteur central de neutron (CND) est présenté dans une autre partie de la thèse.

Le détecteur CLAS12 (CEBAF Large Acceptance Spectrometer 12-GeV) est un détecteur de particules installé au laboratoire national américain, Thomas Jefferson National Accelerator Facility (JLab). Ce laboratoire, situé à Newport News sur la côte Est des États-Unis, étudie la physique hadronique

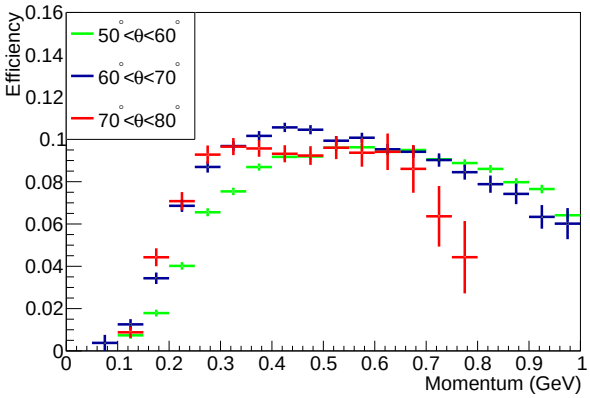
depuis sa création en 1984. JLab est construit autour du CEBAF (Continuous Electron Beam Accelerator Facility) qui est un accélérateur d'électrons polarisés pouvant atteindre une énergie de 12 GeV. Le détecteur CLAS12 (voir Figure Résumé-Fr-2) est installé dans le Hall B de JLab et reçoit une luminosité nominale de $10^{35}\text{cm}^{-2}\text{s}^{-1}$. Il est composé du détecteur à l'avant (FD), qui identifie les particules émises à bas angles polaires, et d'un détecteur central (CD) construit autour de la cible. Les données montrées dans cette thèse ont été prises par CLAS12 en automne 2018, avec une cible d'hydrogène liquide et un faisceau d'électrons avec une énergie de 10,6 GeV.

Le détecteur central de neutrons

Le détecteur central de neutrons de CLAS12 (CND) est un détecteur cylindrique de 10 cm d'épaisseur situé entre le CTOF et la paroi interne de l'aimant solenoïdal de CLAS12. Il a été conçu dans le but d'augmenter la capacité de détection des neutrons émis à grands angles polaires. Ce détecteur va ainsi permettre la mesure de la réaction DVCS sur le neutron, étape nécessaire dans la compréhension de la structure du spin du nucléon. Ce détecteur a été conçu et fabriqué à l'Institut de Physique Nucléaire d'Orsay et a été installé au début de ma thèse (voir la photographie de la Figure Résumé-Fr-3a). Dans cette thèse, je présente le travail réalisé sur la reconstruction et la calibration du CND. Enfin je montre les performances du détecteur. Plus particulièrement, l'efficacité de détection des neutrons est estimée à l'aide de la réaction exclusive de production de pion positif ($ep \to e\pi^+ n$). L'efficacité est calculée comme le ratio du nombre d'événements où le neutron est détecté dans le CND et du nombre totale de ces événements avec un neutron manquant susceptible d'être détecté par le CND. L'efficacité mesurée est de 10% (voir Figure Résumé-Fr-3b), en accord avec la valeur nominale, définie lors de la phase de R&D du CND.



(a)



(b)

Figure Résumé-Fr-3: (a) Photographie prise après l'installation du CND à Jefferson Lab en Octobre 2017. (b) Efficacité de détection des neutrons mesurée à l'aide de la réaction exclusive $ep \to e\pi^+ n$. Les performances de détection du CND sont proches des valeurs nominales.

Analyse de données: Diffusion Compton de genre temps

La dernière partie de la thèse présente le travail que j'ai réalisé pour l'analyse de la réaction de diffusion Compton de genre temps (TCS). Après avoir présenté les motivations pour la mesure de cette réaction, la stratégie d'analyse est détaillée. Les résultats obtenus pour trois observables sont présentés. Une étude des incertitudes systématiques est aussi décrites. Finalement l'interprétation des résultats obtenus à l'aide des prédictions de deux modèles théoriques est discutée.

Motivation

Le processus Compton de genre temps (TCS) est la réaction symétrique en temps de la réaction DVCS. Un photon réel interagit avec un quark, qui émet un photon virtuel, et décroît ensuite en une paire de leptons ($\gamma p \to \gamma^* p' \to e^+ e^- p'$). Le diagramme correspondant est présenté en Figure Résumé-Fr-4. Le processus de Bethe-Heitler (BH), purement électromagnétique, contribue aussi à la réaction de photo-production d'une paire de lepton.

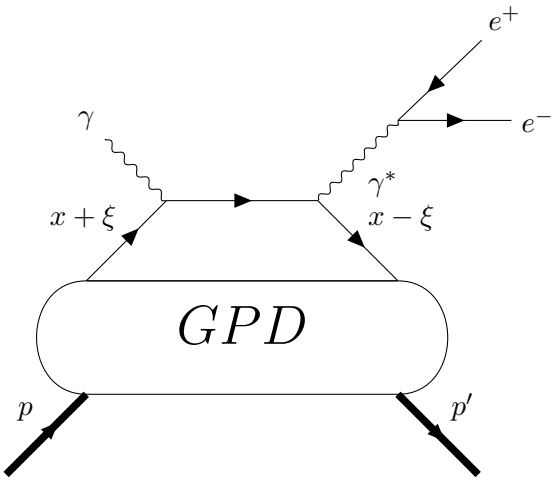


Figure Résumé-Fr-4: Le diagramme de Feynman de la réaction TCS $\gamma p \to p' e^+ e^-$

La mesure du TCS permet:

- d'accéder à la partie imaginaire du CFFs $\mathcal{H}$, et ainsi de pouvoir tester le caractère universel des GPDs,
- de mesurer la partie réelle de $\mathcal{H}$, une quantité encore mal connue, et qui peut être reliée aux propriétés mécaniques des nucléons.

Sélection de la réaction d'intérêt

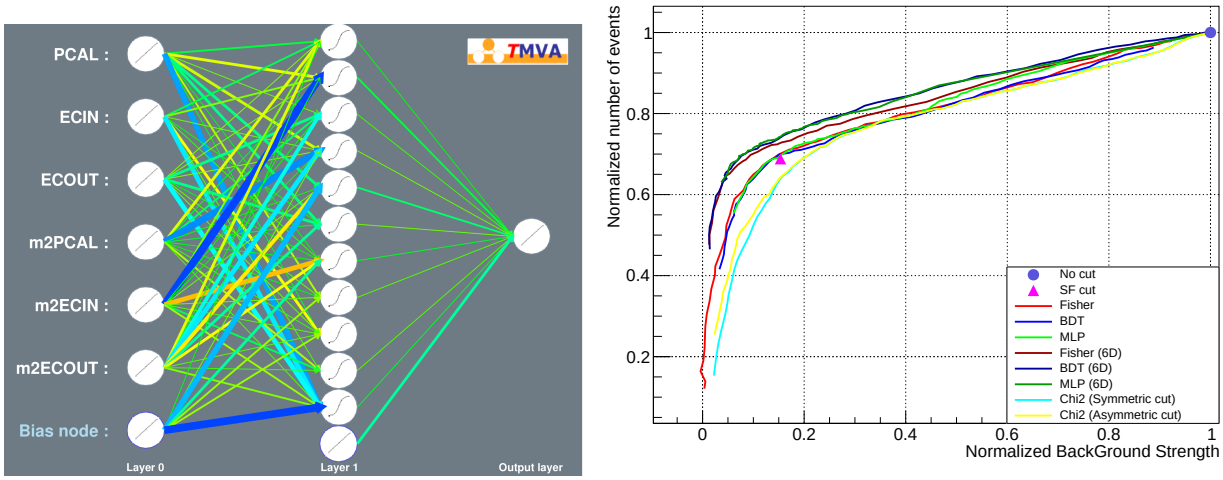
Pour sélectionner la réaction d'intérêt ($ep \to (e')\gamma p \to (e')p' e^+ e^-$), la stratégie suivante est adoptée:

- les particules de l'état final sont sélectionnées. Le proton est identifié par le technique du temps-de-vol. Les leptons (électrons et positrons) sont identifiés à l'aide du calorimètre et du compteur Cherenkov à haut seuil (HTCC). Un réseau de neurones est utilisé pour identifier les positrons à grandes impulsions ($P > 4.9$ GeV). L'architecture du réseau de neurones est montrée dans la Figure Résumé-Fr-5a. La courbe ROC caractérisant les performances de cette approche se trouve en Figure Résumé-Fr-5b,
- des coupures fiduciaires sont appliquées et les impulsions des particules sont corrigées pour prendre en compte les pertes d'énergies ainsi que les effets intrinsèques aux systèmes de détection de traces,
- des coupures exclusives sont appliquées pour s'assurer de l'exclusivité de la réaction mesurée.

Résultats

Trois observables du TCS sont mesurées dans cette thèse. La première observable est l'asymétrie de polarisation circulaire du photon réel (BSA), sensible à la partie imaginaire de $\mathcal{H}$. Elle est calculée comme dans l'équation suivante:

$$BSA = \frac{\sigma^+ - \sigma^-}{\sigma^+ + \sigma^-},$$



(a) (b)

Figure Résumé-Fr-5: (a) Réseau de neurones utilisé pour l'identification des positrons de grandes impulsions. (b) Courbes ROC de l'identification des positrons pour plusieurs méthodes testées durant cette thèse, en particulier pour le réseau de neurones présenté précédemment. Ces courbes sont calculées à l'aide de données réelles prises par CLAS12.

où σ^+ (resp. σ^-) est la section efficace du TCS lorsque le photon réel a une polarisation circulaire droite (resp. gauche). Le BSA en fonction du carré de l'impulsion transférée au proton t est présenté en Figure Résumé-Fr-6. Les points expérimentaux sont comparés avec des prédictions théoriques issues de deux modélisations de GPDs, le modèle VGG et le modèle GK. Les asymétries mesurées sont supérieures aux asymétries prévues en présence du processus de Bethe-Heitler seulement. Cela est une preuve de la mesure du processus TCS. L'amplitude de l'asymétrie est bien reproduite par les prédictions théoriques. Cela est un indice de l'universalité des GPDs car ces modèles reproduisent également les mesures expérimentales du DVCS.

La seconde observable mesurée est le ratio intégré de la section efficace pondérée par $\cos(\phi)$, et définie par:

$$R(\sqrt{s}, Q'^2, t) = \frac{\int_0^{2\pi} d\phi \cos(\phi) \frac{dS}{dQ'^2 dt d\phi}}{\int_0^{2\pi} d\phi \frac{dS}{dQ'^2 dt d\phi}}.$$

Néanmoins cette observable est sensible à l'acceptance angulaire de CLAS12, ce qui induit de fausses asymétries. Ce phénomène rend difficile la comparaison de ce ratio avec des prédictions théoriques.

Une troisième observable, l'asymétrie avant/arrière, est donc introduite et discutée dans cette thèse. L'asymétrie avant/arrière (A_{FB}), dont aucune prédiction n'a été faite pour la réaction du TCS avant ce travail, permet de s'affranchir des problèmes d'acceptance rencontrés pour le ratio de section efficace. L'asymétrie A_{FB} est définie comme:

$$A_{FB}(\theta_0, \phi_0) = \frac{d\sigma(\theta_0, \phi_0) - d\sigma(180° - \theta_0, 180° + \phi_0)}{d\sigma(\theta_0, \phi_0) + d\sigma(180° - \theta_0, 180° + \phi_0)},$$

où $d\sigma(\theta_0, \phi_0)$ (resp. $d\sigma(180° - \theta_0, 180° + \phi_0)$) correspond à la section efficace de la réaction lorsque l'électron final est émis à l'avant (resp à l'arrière), la direction de l'électron étant définie par un angle polaire θ_0 et un angle azimutal ϕ_0 dans le centre de masse de la paire de leptons. L'asymétrie A_{FB} permet d'accéder à la partie réelle de $\mathcal{H}$. Figure Résumé-Fr-7 représente l'asymétrie mesurée. Les points expérimentaux sont comparés aux prédictions théoriques. Comme pour le BSA, l'A_{FB} mesurée est non-nulle, ce qui traduit la mesure du processus TCS en plus du processus BH. De plus, la comparaison avec les modèles théoriques nous renseigne sur les propriétés du terme D, qui est un

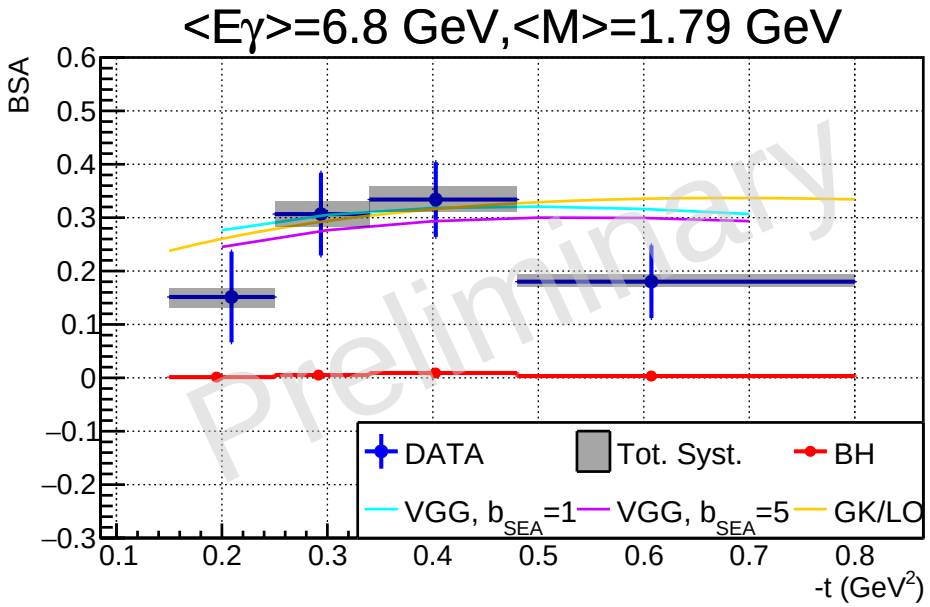


Figure Résumé-Fr-6: BSA en fonction de t. Les points expérimentaux sont en bleu. Les rectangles grisés représentes les incertitudes systématiques pour chaque intervalle. Les courbes théoriques sont calculées pour les cinématiques moyennes citées au dessus du graphique. Les points rouges représentent les valeurs attendues si seulement le processus BH contribuait.

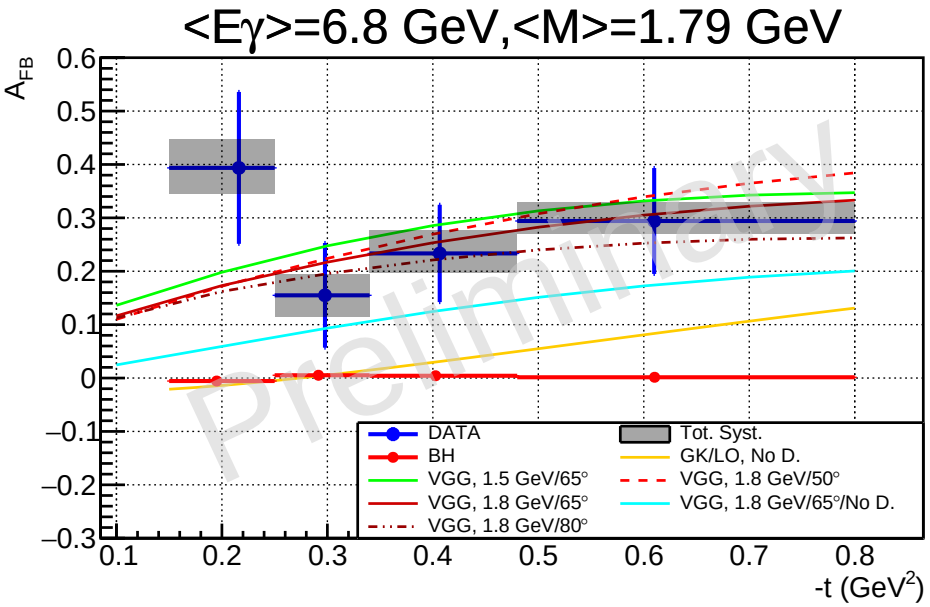


Figure Résumé-Fr-7: A_{FB} en fonction de t.

facteur important dans la modélisation de la partie réelle de $\mathcal{H}$. En effet les données sont mieux reproduites par le modèle VGG quand le terme D est inclus.

Conclusion

La description de la structure interne des nucléons est un problème complexe, aussi bien d'un point de vue théorique qu'expérimentale. En effet, il est nécessaire d'utiliser des fonctions de structure, telle que les GPDs. De plus, l'extraction expérimentale des GPDs nécessite de mesurer de nombreuses observables sur différentes réactions sensibles aux GPDs. Dans cette thèse, j'ai présenté deux aspects

de la mesure expérimentale des GPDs. Dans une première partie, le travail réalisé sur le détecteur central de neutron de CLAS12 est présenté. Ce détecteur permet la mesure de la réaction de DVCS sur le neutron, étape nécessaire dans la compréhension de la structure du spin du nucléon. La procédure de calibration mise en place durant ma thèse a permis d'atteindre des performances comparables à celles attendues. Cela permet déjà d'obtenir des résultats préliminaires encourageant pour la mesure du DVCS sur neutron. La seconde partie est dédiée à la mesure expérimentale du TCS, réaction complémentaire du DVCS dans l'extraction des GPDs. Les résultats obtenus sont comparés à des prédictions théoriques, qui permettent de tirer des conclusions sur l'universalité des GPDs, ainsi que sur la structure mécanique des nucléons.

Titre: Étude de la structure du nucléon avec CLAS12 à Jefferson Lab: Diffusion Compton de genre temps et le détecteur central de neutron.

Mots clés: GPDs, Gluons, Structure du nucléon, Réaction exclusive, Diffusion d'électron, Quarks

Résumé: Les protons et les neutrons sont les constituants principaux de la matière visible de l'univers. Leur structure, constituée de trois quarks de valence baignés dans un nuage de quarks de la mer et de gluons, est régis par la théorie de la chomodynamique quantique (QCD). Cependant, aux énergies comparables à la masse du nucléon, les propriétés de QCD ne peuvent pas être calculées par des méthodes perturbatives. Des fonctions de structure doivent être utilisées pour pouvoir décrire les nucléons. Les distributions de parton généralisées (GPD) sont un ensemble de fonctions de structure, introduites dans le courant des années 90. Elles modélisent la position transverse et le moment longitudinal des quarks et des gluons, les constituants élémentaires des nucléons. La phénoménologie de ces fonctions est très singulière. Elles sont en particulier étroitement liées à la structure de spin et aux propriétés mécaniques des nucléons. La mesure des GPDs est donc un élément déterminant dans la compréhension de la structure élémentaire de la matière. Le but de cette thèse est de fournir de nouvelle donnée pour l'étude des GPDs, en particulier avec la mesure inédite de la diffusion Compton de genre temps avec le détecteur CLAS12 à Jefferson Lab. Cette thèse est divisée en trois parties. Dans la première partie, la théorie des GPDs et leur modélisation est présentée. Le lien entre ces fonctions et des réactions mesurables est aussi établi, le concept de facteurs de forme Compton (CFF) est notamment introduit. De plus, les relations entre les GPDs et les différentes contributions des quarks au spin du nucléon, la correspondance entre la partie réelle des CFFs et les propriétés mécaniques du nucléons et enfin la possibilité de réaliser une image 3D du nucléon sont mises en lumière. La seconde partie du manuscrit est consacrée au travail que j'ai réalisé sur le détecteur central de neutrons de CLAS12 (CND). Le CND est un détecteur cylindrique formé par des scintillateurs en plastique. Il a été conçu pour augmenter les capacités de détection des neutrons dans la partie centrale de CLAS12. Après avoir présenté les motivations physiques de la construction du CND, le design, la procédure de calibration, de reconstruction et de simulation sont expliqués. Enfin, les performances du CND, mesurées à partir de données réelles sont comparées aux spécifications du design initial. Enfin, dans la dernière partie, la mesure expérimentale de la réaction de photo-production d'une pair de lepton sur le proton, ou diffusion Compton de genre temps (TCS) est exposée. Cette réaction permet d'accéder à des propriétés des GPDs encore mal connues, comme la partie réelle des CFFs. Le dispositif expérimental utilisé pour cette expérience est d'abord présenté. L'analyse de données est ensuite détaillée et les résultats obtenus sont présentés et discutés.

Title: Nucleon structure studies with CLAS12 at Jefferson Lab: Timelike Compton Scattering and the Central Neutron Detector.

Keywords: GPDs, Gluons, Nucleon Structure, Exclusive reactions, Electron Scattering, Quarks

Abstract: The nucleons, protons and neutrons, are the main constituents of visible matter in the universe. Their structure, three valence quarks surrounded by a cloud of sea quarks and gluons, is described by the theory of quantum chromodynamics (QCD). However, the properties of QCD cannot be computed perturbatively at energies comparable to the nucleon mass. Hence, structure functions were adopted to model the inner structure of nucleons. The Generalized Partons Distributions (GPD), were introduced in the 90's to provide a description of the nucleon in terms of both the transverse position and the longitudinal momentum of its quarks and gluons. These functions contain a large amount of information and are closely related to the nucleon spin and mechanical architecture. Their experimental measurement is a key element for the understanding of fundamental properties of matter. The main focus of this thesis is to provide new data for GPD studies, with a first-time measurement of Timelike Compton Scattering at Jefferson Lab with the CLAS12 detector. This thesis is divided in three parts. The first part presents the theory of GPDs, current models and their link with physical processes that can be experimentally measured. The relation between GPDs and experimental observables is discussed, and the concept of Compton Form Factors (CFF) is introduced. In addition, the link between the spin and the mechanical properties of the nucleon, as well as the possibility of performing a 3D imaging of the nucleon with GPDs is highlighted. The second part of the manuscript is dedicated to the work I performed on the Central Neutron Detector (CND). The CND is a plastic scintillator barrel built to increase the neutron detection capabilities of CLAS12 in its central region. After presenting the physical motivations for the building of this detector, its hardware implementation, calibration, reconstructions and simulation aspects are detailed. At the end of this part, the CND performances using real data are measured and compared to its design specifications. Finally, the third part covers the experimental measurement of the photo-production of a lepton pair off the proton, the Timelike Compton Scattering process (TCS). This reaction offers an insight on some properties of GPDs which are not well constrained by the reactions measured so far, in particular the real part of CFFs. The experimental setup used for data taking is described. The subsequent data processing and analysis is explained, and results for three different observables are shown.