10-30-2013

Hadronization Studies via π^0 Electroproduction off D, C, Fe, and Pb

Taisiya Mineeva

University of Connecticut - Storrs, mineeva@jlab.org

Follow this and additional works at: <https://opencommons.uconn.edu/dissertations>

Recommended Citation

Mineeva, Taisiya, "Hadronization Studies via π^0 Electroproduction off D, C, Fe, and Pb" (2013). *Doctoral Dissertations*. 266.
<https://opencommons.uconn.edu/dissertations/266>

Hadronization Studies via π^0 Electroproduction off D, C, Fe, and Pb

Taisiya Mineeva, Ph.D.

University of Connecticut, 2013

Propagation of partons and formation of hadrons is a topic of interest to multiple communities. New data available from Drell-Yan measurements at FermiLab, heavy ion collisions in RHIC and LHC, SIDIS measurements from HERMES at DESY and Jefferson Lab, all bring different types of information on short distance processes. DIS data obtained in the well understood nuclear medium provide direct information on hadron formation, essential to lay the groundwork for testing theoretical tools. A series of semi-inclusive DIS measurements were performed on D, C, Fe, Pb nuclei. The data were collected during the EG2 run period using the CLAS at Jefferson Lab. A double-target system consisting of liquid deuterium and one of the solid targets was exposed to a 5.014 GeV electron beam. The goal of the experiment is to extract hadronic multiplicity ratios (R_A^h) off nuclei of varying size. These are believed to have sensitivity to the parton fragmentation as well as in-medium hadronization.

Hadronization Studies via π^0 Electroproduction off D, C, Fe, and Pb

Taisiya Mineeva

B.S. Physics, Moscow State University, Moscow, Russia, 2005

M.Sc. Physics, Moscow State University, Moscow, Russia, 2007

M.Sc. Physics, University of Connecticut, Storrs, 2008

A Dissertation

Submitted in Partial Fulfilment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2013

Copyright by

Taisiya Mineeva

2013

APPROVAL PAGE

Doctor of Philosophy Dissertation

Hadronization Studies via π^0 Electroproduction off D, C, Fe, and Pb

Presented by

Taisiya Mineeva,

Major Advisor

Kyungseon Joo

Associate Advisor

William K.Brooks

Associate Advisor

Richard Jones

University of Connecticut

2013

ACKNOWLEDGEMENTS

The years I have spent on my research were a fascinating experience in science and life. I have encountered a pleiad of truly amazing people, invoking both professional and personal admiration. Many of them I know will stay, taking a special place in my mind and in my heart. I complete this journey a very different person than when I started, feeling strong and enduring yet with a tender heart, filled with the deepest gratitude to all the people who helped me to achieve the accomplishment that this work represents.

To begin with, I would like to thank those who had the most direct impact on this work, my adviser Prof. Kyungseon Joo and my supervisor Dr. Will Brooks. With his ethics of very hard work and persistence in problem solving, Kyungseon has helped me to improve my own abilities in those areas. He gave me the right amount of freedom to be independent in approaching the challenges of research, and yet, when I would find myself torn in between solving too many problems, his guidance was strong and precise. I am grateful to Kyungseon for funding my bread, and my attendance to many conferences. Except for the first year of graduate school, I have been permanently stationed at Jefferson Lab, in an environment that was perfect to stimulate my scientific progress. This would not have been possible without Kyungseon relentless support for which I owe him a

big debt of gratitude. I was blessed to have Will as my scientific adviser. I have met few physicists as competent and astute, in both experiment and theory, as Will. He would always make time to provide me with valuable insights in regards to my analysis and to give me meticulous comments about my presentations and my write ups. I have learned a great deal from Will, and I always wished I would learn more. He has always been a great source of encouragement, patience and diplomacy.

Thanks are due to EG2 working group, to Hayk Hakobyan, Aji Daniel, Raphael Dupre, Steve Manly, Lamiaa el Fassi and Kawtar Hafidi. Will Brooks, the leader of the group, has put much effort in organizing our meetings and keeping us on track of each other's work. Profound thanks to Hayk for our friendship and fruitful work discussions.

A good part of this thesis is due to the great environment that Jefferson Lab is. I am thankful to Maurizio Ungaro who had kindly supervised me for some time at JLab. Being both a very knowledgeable and at the same time very nice person would make a great adviser of Mauri if he ever decides to have his own students. I owe a great deal of gratitude to the entire staff of Hall-B and CLAS collaboration. In particular, to Volker Burkert, Valery Kubarovsky, Victor Mokeev, Eugene Pasyuk and Stepan Stepanyan who would inevitably find a moment to share their expertise to help me navigate through the caveats of CLAS analysis. Outside the JLab, I much enjoyed the refreshing bike rides on the weekends with Eugene (and

Valery), who knows all the best routes in the vicinity for any season, and who would always kindly wait for me to catch up with him. I would like to thank Victor for sharing his funniest russian jokes that seemed to be endless in his store. I had a chance to briefly work with Peter Bosted on the radiative corrections code, it was a valuable and fascinating experience for which I am very grateful for. I would like to thank Franz Klein for the opportunities he offered to me, and the wisdom that came along with it. Many thanks to Latifa Elouadrhiri, Patrizia Rossi and Raffaella De Vita for their most interesting and warm company. For theoretical assistance, I owe many thanks to Alberto Accardi for his expansive answers to my question related hadronization studies as well as to Alexei Prokudin for helping me to navigate through QCD.

I would like to further thank my friends I have met over the years in JLab: Nick Zachariou, Daria Sokhan, Paul Mattione, Sucheta Jawalkar, Erin Seder, Wes Gohn, Igor Senderovich, Brian Vernarsky, Jason Bono and Or Hen. Being now scattered around the globe, it was always a celebration when we had a chance to get together on the occasion of one of the collaboration meetings, and have great time socializing and catching up. Those unforgettable times made my stay in Jlab much happier. When Daria visited JLab, we would enjoy our long conversations and make little outdoor trips, which felt like a gulp of fresh air. Debussy's masterpieces often remind me of Brian, and his beautiful performances on our concert piano at JLab. I especially would like to thank Erin for her help

on the distance in navigating through the administrative aspects of the graduate school at UConn. Many kind thanks are due to Nick Zachariou, Balsa and Denis Terzic for their heartfelt hospitality, and the many warm cozy evenings we used to spend at their house. I am grateful to have met Jenia and Alena through their husbands working in JLab, the two amazing women I have much enjoyed the company of, and very talented photographers who gifted me with their attention. The russian-armenia community, when gathered for a celebration was much joy, for which I would like to thank Nick and Natalya, Sergei and Jenia, Andrey and Olga, Vania and Masha, Arthur, Gleb and Iliya. JLab environment was truly a wonderful place to meet good people, it gifted me with dear friends. Deep thanks to Natalie Waldorf and Nick Markov for being very special friends. Nick and his wife Natasha were the first people I have encountered when I came to the lab 6 years ago. From the beginning, I knew I could rely on them to come to rescue in any circumstances. Sharing the same office walls with Nick, he had heard many of my questions, complains, desperations, victories, jokes and news - no matter what it was, he would always remain a respectful listener, with a great sense of humor and a spree of calmness. I owe Nick many thanks for helping me occasionally with my work and for always making my days in the office calmer and brighter. With a wide smile and deep gratitude, I would like to thank Natalie. Though sharing different views on politics and having slightly different cultural background, it did not prevent us to become so close we could easily joke about our differences and

give each other a truly good company. She has generously provided me with her help on many occasions, both personal and professional. In storm and under the sun, she would always be there, with her support, encouragements, kind attention and life-celebrating spirit. Without her ability and efforts to help me vent, at the end of this journey I would be but a pressure cooker. I feel very lucky to have found a heartfelt friend like Natalie.

I can not speak of my time in United States without mentioning my old good friends from Russia. It has been many years I was away from home, and we did not have a chance to meet. When Sasha flew to DC to visit me, and Kostia came to Newport News to stay for couple days as a part of his american trip, I could not have had been happier. My dear Katia, a friend of almost two decades, has never let me to disappear in studies for too long, she would always find me on the phone and we would have conversations as no time have elapsed. I further thank Misha, Ira, Daria, Anton, and Jenia for staying in touch.

It would not be complete without a very special thanks to Andy and Olimpia Stein. They generously open the doors of their beautiful house every Friday night to host milongas of our (largely scientific) tango community. Thanks to Andy and Olimpia, the Friday nights are the ever-lasting faery of the week for those infected with the joy of argentine tango. Yet, there is substantially more to milongas than dancing (and eating!), it is the most wonderful people that comprise our community, people whom I was looking forward to see in the context of tango and

outside of it. Friday nights spent at La Tangueria in the atmosphere of friendliness feel like home, a soothing place to recharge and be merry. Andy and Olimpia, Daphne, Gary, Jessica and Andy, Jamie, John, Lena, Matt, Nikos, Onnie, Skip and Paula, Sharon, Troy and all of you, thank you for making our community such a cherished place to be. A special thanks to Nikos for providing me with an access to a perfect place in the library where I could productively work during the weekends. Each of us being at the end of the PhD, going through similar roller costers, we shared a great deal of support that helped us to get to our goals.

Finally and most of all, I owe the deepest thanks of all to those in my life who helped to get where I am now, to my family back in Russia and to my husband Francois. As much of a challenge as this journey has been for me, I know it has been harder for them. My family has been my fortress, with their kind understanding of my prolonged stay away from home, care and 24/7 support. My mother's love and trust in me empowered me to step on the scientific path, it nourished me throughout, and celebrated my accomplishment with the utmost joy. Francois has been my intellectual and spiritual lodestar all these years. His love, support and expertise kept me going like none else. With my deepest gratitude, I now have a privilege to dedicate this work to him.

TABLE OF CONTENTS

1. Physics Motivation	5
1.1 Deep inelastic scattering	5
1.1.1 DIS cross section	8
1.1.2 Quark-parton model	10
1.1.3 Quark-parton model and pQCD	13
1.1.4 Factorization	16
1.2 Hadronization	17
1.2.1 Hadronization in vacuum	19
1.2.2 Hadronization in nuclear environment	22
1.2.3 Nuclear effects in SIDIS	24
1.2.4 Observables	26
1.2.5 Overview of existing data	27
1.2.6 Hadronization in nuclear environment: models	31
1.3 Connection to neutrino experiments	50
1.4 Motivation	57
2. Experiment	61
2.1 CEBAF	61
2.2 CLAS	64
2.2.1 Drift chambers	66

2.2.2	Čerenkov counters	69
2.2.3	Time-of-flight Counters	71
2.2.4	Electromagnetic calorimeter	72
2.3	EG2 experiment	75
2.4	Triggering and data acquisition	76
2.5	Data Processing	78
3.	Data analysis	80
3.1	Particle ID scheme for electrons	80
3.1.1	Momentum	81
3.1.2	Čerenkov signal	81
3.1.3	Sampling fraction	85
3.1.4	Inner energy	89
3.1.5	EC and SC coincidence time	90
3.1.6	EC fiducial cut	92
3.1.7	DC fiducial cuts	93
3.1.8	Vertex correction	94
3.1.9	Kinematics	98
3.2	Particle ID scheme for photons	99
3.2.1	Recapitulation of cuts, and available statistics	103
3.3	Identification	105
3.3.1	Photon energy correction	107

3.3.2	Kinematical constraints on the photon	115
3.3.3	Event mixing method for π^0 background subtraction	118
3.4	Binning	130
3.5	p_T^2 cutoff	131
3.6	Exclusive contribution	133
4.	Monte Carlo Simulation	138
4.1	Adaptation of LEPTO	139
4.1.1	Fermi momentum	142
4.2	Simulations and reconstruction	145
4.2.1	GSIM	146
4.2.2	GPP	148
4.2.3	RECSIS	149
4.3	Particle identification	150
4.3.1	Electron identification	150
4.3.2	Photon identification	153
4.3.3	Photon energy correction	155
4.4	Phase space and shape of MC events	160
4.4.1	Resolution	170
4.4.2	Purity	174
4.5	Acceptance	177
4.5.1	Electron acceptance	180

4.5.2	$e\pi^0$ acceptance	184
5.	Radiative corrections	189
5.1	Inclusive radiative corrections	190
5.2	SIDIS radiative corrections with HAPRAD.	191
5.2.1	Nuclear structure functions	193
5.2.2	Input cross sections: SIDIS	200
5.2.3	Input cross sections: exclusive	202
5.2.4	Comparison of input cross sections	206
5.2.5	Exclusive contribution at low W	208
5.2.6	Calculation grid	214
5.2.7	Model dependence of RC	217
6.	Systematic uncertainties	221
6.1	Overview	221
6.2	Electron ID	223
6.2.1	Target vertex cut	223
6.2.2	Target leakage	225
6.2.3	Sampling fraction cut	227
6.3	Photon ID	229
6.3.1	EC time (beta) cut	229
6.4	π^0 identification	232

6.4.1	Systematic error on the signal shape	236
6.4.2	Acceptance extraction method	237
6.5	Recapitulation of all the cuts	240
7.	Results and Discussion	243
7.1	Multiplicity ratios results	244
7.2	Multiplicity ratios discussion	248
7.2.1	z Dependence	249
7.2.2	p_T^2 Dependence	250
7.2.3	Q^2 Dependence	252
7.2.4	ν Dependence	253
	Bibliography	262

LIST OF FIGURES

1.1	Schematic drawing of the deep inelastic scattering process. Incoming lepton l scatters off a nucleon h exchanging a virtual boson V	6
1.2	Illustration of the SIDIS process $l(k) + h(p) \rightarrow l'(k') + h(p) + X$. Lepton momenta $\vec{k}$ and $\vec{k}'$ define the scattering plane, while the hadron plane is defined by the direction of virtual photon $\vec{q}$ and that of the hadron $\vec{p}_h$. The angle between leptonic and hadronic plane is ϕ_h	8
1.3	SLAC measurement of the Callan-Gross ratio $\frac{2xF_2}{F_1}$ plotted versus $x = \frac{Q^2}{2M\nu}$. The plot was taken from [2].	12
1.4	The proton structure function F_2^p given at two values of Q^2 . The factors shown in brackets represent normalization of various datasets. The plot was taken from [4].	14
1.5	Strong coupling $\alpha_s(Q^2)/\pi$ obtained from JLab spin structure function data and the world data [6].	15
1.6	A classical picture of leading hadron production taken from [14]. The parton, following hard scattering, regenerates its color field via gluon emission during the production time. Once the color is neutralized, a formation time begins until the full hadronic wave function is formed. . . .	20

1.7	Illustration of $e^+e^- \rightarrow (Z/\gamma) \rightarrow H$ process. Here, D_j^h is the FF given by a probability that the parton of flavor j fragments into observed hadron h with fractional momentum z . As e^+e^- process does not proceed through hadronization stage, it allows direct extraction of the FF's.	21
1.8	Schematic illustration of the parton interaction as it travels inside the nuclear medium. Produced hadron can be formed outside (left) or inside (right) the nucleus, or somewhere in between. The cartoon is taken from [45].	24
1.9	The EMC effect on carbon relative to deuterium. The solid curve is the fit from SLAC to the carbon ratio. The plot is taken from [18].	25
1.10	EMC measurements of the multiplicity ratios as a function of p_T^2 in two ν bins [41].	29
1.11	HERMES one-dimensional π^0 multiplicity ratios in ν , z , Q^2 , and p_T^2 bins. Kinematical coverage is: $1 < Q^2 < 10 \text{ GeV}^2$, $7 < \nu < 23 \text{ GeV}$, $z > 0.2$ and $2 < p_h < 15 \text{ GeV}$. Results taken from [43].	31
1.12	Formation time τ_f in one-time scale absorption model [21].	35
1.13	Schematic illustration of the Lund string model. The vertical axes represents time, and the horizontal - direction. The black points denote the breaking points of the string due $q\bar{q}$ production and formation of the prehadron. Hadronization proceeds through the evolution of the string pieces (blue and green). P_i are the formation points of the hadrons. . .	37

1.14	HERMES data on charged hadrons fitted with the model predictions based on PYTHIA and FRITOF, and BUU transport model. Good agreement with data is reached when hadron formation length $l_y = 0.5$ fm and the prehadron interaction cross section $\sigma_h = \frac{\sigma}{3}$. Adapted from [25]. . . .	38
1.15	Predicted nuclear modification of FF by Wang and Wang [27] compared to HERMES data.	41
1.16	Comparison of HERMES attenuation data with the calculations carried out by Majumder. The partonic energy loss was evaluated in single scattering and single gluon emission (red) as well as multiple emission (blue) scenarios. In addition to partonic energy loss, hadronic energy loss were incorporated in the model (green).	43
1.17	Charged hadrons multiplicity ratios from the EMC on Cu (left) and HERMES data on Kr (right) fitted with model predictions by Accardi based on: pure rescaling (dashed), rescaling plus absorption according to Bialas-Chmaj (solid), rescaling plus absorption according to Bialas-Gyulassy (dotted). Adapted from [30].	45
1.18	Model fits to HERMES data for charged hadrons: energy loss models (blue) [27] and the absorption models (red and green) [25]. Taken from [36] .	46
1.19	Color-dipole model predictions for z -dependence of the multiplicity ratios on N and Kr for pions with inclusion of induced radiation (solid) and without (dashed). Taken from [32].	49

1.20	Color-dipole model predictions for p_T^2 -dependence binned in z of the multiplicity ratios on N (left) and Kr(right). Adapted from [32].	49
1.21	Probability distribution of the leading pre-hadron production time $t = t_p$ calculated on average HERMES kinematics. Adapted from [32]. . . .	50
1.22	Quark propagation inside a target nucleus ("cold QCD matter") illustrated for the deep-inelastic scattering in lepton-nucleus reaction (left) and Drell-Yan process in hadron-nucleus collisions (center). Right: Hard scattered parton traveling through the "hot QCD matter" produced in nucleus-nucleus collisions. The diagram is taken from [8].	58
2.1	The Continuous Electron Beam Accelerator Facility (CEBAF). Electron beam starts at the injector, passes two linear accelerators and the bending recirculating arcs, then terminates at the experimental halls (Hall A, B, C).	63
2.2	A schematic top view of the CLAS detector cut along the beam line. Typical photon, electron, and proton trajectories (from top to bottom) from an interaction in the target are superimposed on the figure.	66
2.3	Contours of constant absolute magnetic field for the CLAS toroid in the midplane between two coils.	67
2.4	Hexagonal cells of drift chambers with a typical track indicated by shaded areas (left). Vertical cut of drift chambers indicates geometry of the regions (right).	68

2.5	Diagram of one optical module of the CLAS Čerenkov counter illustrating the Čerenkov light from the electrons reflected twice from the mirrors into a Winstone cone which collects light on the surface of PMT. . . .	70
2.6	Schematic view of the scintillator strips for one sector with respect to the beam direction.	72
2.7	Schematic view of one of the six CLAS electromagnetic calorimeter modules.	74
2.8	The double target system from the EG2 run showing one solid target (in blue square) inserted in the beam line, and five targets retracted. The solid targets are held by mechanical arms allowing to change them remotely. The liquid (in red square) is in the back enveloped in thermally insulated aluminum foils.	77
3.1	10·Nphe distribution for all electron candidates with negative charge and time-based tracking.	82
3.2	θ_p vs mirror number distributions. The plot on the left contains electron candidates that satisfy all the electron ID cuts but the Nphe and CC match (left), while the plot on the right has the same electron sample but after the CC matching cut is applied.	82
3.3	10·Nphe distributions in each sector before the CC matching cut (blue) is applied and after (red). Vertical lines correspond to the 10·Nphe cut for each sector as defined in equation 3.1.4.	84

3.4	E_{tot}/P vs $\sigma(p)$ distribution for all electron candidates for which there is a signal in the DC, SC, CC and EC.	86
3.5	Carbon target set up E_{tot}/P vs $\sigma(p)$ distribution for all electron candidates for each CLAS sector. Black solid line is the second order polynomial fit to the mean value $\mu(p)$ of E_{tot}/P in each momentum bin. Red lines are the cuts on the sampling fraction corresponding to $\pm 2.5\sigma(p)$	88
3.6	Iron target E_{tot}/P vs $\sigma(p)$ distribution for all electron candidates for each CLAS sector. Black solid line is the second order polynomial fit to the mean value $\mu(p)$ of E_{tot}/P in each momentum bin. Red lines are the cuts on the sampling fraction corresponding to $\pm 2.5\sigma(p)$	89
3.7	E_{out} vs E_{in} distributions. Left - electron candidates satisfying negative charge and hit-based tracking; right - electrons satisfying all the identification cuts but E_{in} . Red dotted lines correspond to the cut off $E_{in} > 60$ MeV.	90
3.8	Distribution of $t_{EC} - (t_{SC} - \frac{l_{EC} - l_{SC}}{c})$ for: a) electron candidates satisfying negative charge and hit-based tracking (left); b) electrons satisfying all electron ID cuts except the EC and the SC coincidence time cut (right). Distribution inside red dotted lines corresponds to the remainder after the cut.	91
3.9	EC fiducial cuts. U, V, W distribution for the electrons containing all the ID cuts except the EC fiducial.	92

3.10	Y vs X distribution before (black) and after (color) imposed U, V, W cuts for: left - all electron candidates, right - electrons satisfying all the ID cuts except the EC fiducial one.	93
3.11	ϕ vs θ distribution before (color) and after (black) applying DC fiducial cuts: left - all electron candidates, right - electrons satisfying all the ID cuts except the DC fiducial one.	94
3.12	Electron vertex distribution on the example of deuterium and carbon target system plotted for each sector: left plot is regular z vertex distributions, right plot corresponds to z vertex corrected for the beam offset. The small peak in between the targets corresponds to the aluminum foil. . .	95
3.13	z vertex distribution for: D+C, D+Fe, and D+Pb. Dashed lines correspond to the vertex cut applied to D (similar for all three cases), and the solid lines illustrate vertex cuts applied to each solid target.	96
3.14	z vs ϕ for electrons. Left plot illustrates distributions before beam offset correction of the vertex, right plot - after. The red dashed lines indicate the cutoffs for the solid (top) and deuterium (bottom) targets.	97
3.15	y vertex of the electrons plotted in each sector, vertical lines correspond to the applied y -cut.	97
3.16	Q^2 vs $x_{Bjorken}$ plotted on the left for the full kinematic range in Q^2 and W , and on the right - in the kinematic range for SIDIS: $Q^2 > 1 \text{ GeV}^2$ and $W > 2 \text{ GeV}$	98

3.17	$t_{EC} - t_{start} - l_{EC}/30$ plotted for all neutral tracks with the hit in the EC.	
	Lines in red correspond to the photon coincidence time cuts.	100
3.18	Beta distribution. Left plot is β of all neutral tracks with a hit in the EC, right - after the time coincidence cut was imposed. The red lines indicate a possible cutoff on the velocity, which becomes unnecessary if the coincidence time cut is applied. The particles with bad status were assigned $\beta = 2$	101
3.19	U, V, W distributions for the photon satisfying all the cuts but the EC fiducial. Red lines correspond to each cut for U, V, W.	101
3.20	X vs Y global coordinates distribution for: all the neutral particles with a hit in the EC (left-hand plot), and photons satisfying the minimum deposited energy cutoff as well as the coincidence time (right-hand plot). Black distribution illustrates regions that have been removed by applying the U, V, W cuts.	102
3.21	π^0 invariant mass distribution for the liquid deuterium (left) and solid carbon (right) targets. This distribution was fitted with a Gaussian plus polynomial background function, where background is plotted in blue and the π^0 peak with background subtracted is plotted in red. Events have been distinguished between the two targets (liquid and solid) based on the electron vertex cuts according to equations 3.1.11.	107

3.22	$\theta_{\gamma_1\gamma_2}$ vs $E_{\pi^0} = E_{\gamma_1} + E_{\gamma_2}$ for all π^0 events (left-hand plot) and the events that satisfy the mass range cut [3.3.3] (right-hand plot). Black dashed curve corresponds to the theoretical value of the θ_{min} between the two photons.	108
3.23	Step 1 of photon energy correction. π^0 invariant mass distribution, binned in E_{γ_1} , is constructed of the photons with $E_{\gamma_1} \approx E_{\gamma_2}$ (example of carbon).	110
3.24	Step 2 where π^0 invariant mass was constructed with either E_{γ_1} or E_{γ_2} being in the range where correction of step 1 is valid, i.e. $0.3 < E_{\gamma}^{uncorr} < 1.1$ GeV, while the other photon was restricted in the energy range $0.35 < E_{\gamma}^{uncorr} < 2.5$ GeV.	112
3.25	$m_{\gamma_1\gamma_2}/0.135$ as a function of photon energy for the C (left) and Fe (right) targets. The black curve corresponds to the correction obtained from step 1, colored - from step 2, and magenta corresponds to the final correction applied.	114
3.26	$m_{\gamma_1\gamma_2}/0.135$ vs E_{γ} . Here the points in black correspond to the corrected values of π^0 , whereas the points in green have not been corrected. The black line is the linear fit to the black points with a constant value of parameter p0, indicating a flat behavior around the unity. Left plot corresponds to the C target, the right one - to the Fe target.	115

3.27	The distribution of E_{tot}/P vs P for the electrons in the example of C (left) and Fe (right) targets. The black curve represents a mean value of the Gaussian fit to the electron sampling fraction per momentum bin. The superimposed red and blue curves are the corrected sampling fraction of the photon $f_s \cdot corr(E_\gamma)$	116
3.28	Photon energy correction factor plotted for two targets: Fe plotted according to equation 3.3.10 in blue, and C (or Pb) plotted according to equation 3.3.9 in red. Green dashed line corresponds to the minimum photon energy cutoff imposed in this analysis to be above 0.3 GeV.	117
3.29	Angular distribution $\alpha_{e\gamma}$ between the scattered electron and one of the two photons, γ_1 and γ_2 , of a π^0 candidate. To avoid bremsstrahlung contamination of the photons, we impose a cutoff, indicated by the green line, on the $\alpha_{e\gamma} > 12^\circ$	118
3.30	The invariant mass of tagged π^0 on the outcome of LEPTO (left), with added smearing on the energy and position of the photon given realistic resolution in CLAS detector (right).	121

3.31	General event mixing method on the example of the lowest z-bin ($0.3 < z < 0.4$).	
	Left plot is the mixed background distribution (yellow) fitted with fourth order polynomial function (black). Right plot illustrates 2γ invariant mass distribution fitted with Gaussian (red) plus mixed background (black), the latter is normalized to the number of events in the peak. The number of events calculated from the fit (N_{gaus}) is $4.7 \pm 0.1\%$ smaller than the actual number of generated π^0 's.	122
3.32	Illustration of the correction procedure in general mixing method. In order to preserve kinematical correlations in the pair of mixes uncorrelated photons, a photon from one event (blue) is rotated to the photon in another (green) by the difference in their azimuthal angles.	125
3.33	Combinatorial background based on the improved mixing method: mixed events originated from the same (Q^2, ν, z, p_T^2) bins, with kinematics of one event rotated by the $\delta\phi$ to coincide with the other event, and mixed based on the photon energy ordering. Left plot is the mixed background distribution (yellow). Right plot is 2γ invariant mass distribution fitted with Gaussian (red) plus mixed background (black). The accuracy of π^0 extraction is at the percent level.	126

3.34	Invariant mass of the two photons from lead target in the slice of bins $0.3 < z < 0.4$, $0 < p_T^2 < 0.1 \text{ GeV}^2$ and from left to right bins in ν : $2.2 < \nu < 3.2$, $3.2 < \nu < 3.73$ and $3.73 < \nu < 4.25 \text{ GeV}$. Top panel is fitted with combinatorial background based on event mixing (red line), bottom panel is the second order polynomial fit (blue). The fit is done once on the wide range to obtain approximate width σ , the second fit is then a refit on the range $\pm 5\sigma$, indicated by the color of the background.	128
3.35	Sample of reconstructed MC events fitted on $\pm 5\sigma$ range in the same slice of bins as described in Fig.3.34. Top panel corresponds to combinatorial background from mixing reconstructed events employing the same method as developed for data, the bottom panel is the second order polynomial fit. The curvature of the polynomial background varies depending on the fit limits (black compared to blue), while combinatorial background is stable regardless.	129
3.36	Electron phase-space is divided into 3 bins each in Q^2 and ν (left). π^0 is divide into 5 and 6 bins correspondingly in z and p_T^2 (right).	130

3.37	p_T^2 distribution (blue) in the slice of (Q^2, ν, z) bins for the case of the lead target (minimum statistics and maximum fluctuations compared with other targets). The black lines correspond to the Gaussian fit. The green vertical line corresponds to the drop of the fit function down to one event per bin, and the red line is the corresponding p_T^2 cutoff. The number of events is plotted per 3 MeV ² in p_T^2	132
3.38	Missing mass distribution W_x . The black curve is W_x in the full SIDIS kinematics, the blue is constrained on $0.8 < z < 1.0$, where missing mass approaches exclusive kinematics: $W_x \rightarrow M_p$	134
3.39	p_T^2 vs z distribution for π^0 events generated using LEPTO without (left) and with (right) Fermi motion smearing of the target nucleon from deuterium. The W_x curves illustrate two cases: detected π^0 originating from exclusive reaction (red) or semi-inclusive (green).	136
3.40	p_T^2 vs z distribution of π^0 candidates from data. The W_x curves illustrate limits on the exclusive π^0 production at $W_x = M_p$, and SIDIS at increasing values of W_x	137
4.1	Generated z and p_T^2 distributions for the D, C, Fe, and Pb targets using LEPTO switch for the mass and charge of the nuclear target. This switch is responsible for the ratio of valence quarks u and d in the initial state. Its effect appears negligible on the characteristics of π^0 produced in the final state.	143

4.2	Probability n_k per GeV of Fermi momentum P_F distribution from Ciofi-Simula parametrization for D(green), C(red), Fe(blue) and Pb(black).	144
4.3	From left to right are Q^2 , ν , z and p_T^2 distribution for the two cases: target nucleon is at rest (black), and target nucleon carries Fermi momentum randomized according to Ciofi-Simula parametrization (blue).	144
4.4	π^0 multiplicity ratios of events that were generated using Ciofi-Simula parametrization for target nucleon. Ratios are extracted by combining statistics on for C, Fe, Pb targets normalized to D. The deviation from 1 demonstrates possible contributions from Fermi momentum to the multiplicities. From left to right multiplicities are plotted as a function of Q^2 , ν , z and p_T^2 respectively.	145
4.5	The double-target assembly for the EG2 run as implemented in the production version of GSIM. It consists of: a deuterium cryotarget with the cell and solid target with its support structures.	147

4.6	The sampling fraction $f_s=(E_{dep}/P)$ vs electron momentum (P) in each sector (1 to 6 from top left to bottom right, respectively). The black crosses are the mean values per momentum bin of the Gaussian fit to f_s . The black solid curves are the 2nd order polynomial fits to the mean values in each sector; the black dashed curve is the same fit but to the mean values integrated over all sectors. Since the difference between the solid and the dashed curve is small, we use the latter to define the cutoff for f_s , same for all sectors.	151
4.7	Width of the Gaussian fit to the sampling fraction plotted as a function of electron momentum (P) in each of the six CLAS sectors (black crosses). The red crosses are the mean values of fit to the sampling fraction integrated over all sectors. The latter is fitted with the inverse square of electron momenta, illustrated by the red line. The upper and lower limits of the cut off on sampling fraction are determined from the deviation of mean values in the range 2.5σ , where σ is given by the red fit. . . .	152
4.8	The top panel corresponds to the z-position, the bottom panel is the-y position of the target vertex. For both panels distributions, on the left hand side are the vertexes of generated events, distributions on the right - are reconstructed events with vertex cuts indicated by dashed curves. The values of vertex cuts are the same as those in data. Green color indicates the deuterium target, black is the iron target.	153

4.9	Distribution of photon velocities obtained using different start times: as taken directly from the EVNT bank (red), as calculated using time and path information from the SC (blue) or EC (green).	154
4.10	Angular distribution between the generated and reconstructed photons. The cutoff $\alpha_{\gamma(rec)\gamma(gen)} > 3^\circ$ determines the angular matching criteria for a given reconstructed photon come from a given generated photon. . .	156
4.11	Step 2 of reconstructed photon energy correction. $M_{\gamma\gamma}$ is binned in $E_\gamma = 0.15$ - 3.0 GeV with width varying from 0.025 GeV at low and medium photon energies to 0.2 GeV at high energies. π^0 candidates, fitted with a Gaussian function in red, are selected by matching an angle of reconstructed and generated photons which leads to a decreased background under the $M_{\gamma\gamma}$ peak. One of the photons in $M_{\gamma\gamma}$ is in the range 0.15 - 1.1 GeV corrected by step 1 function, while the second photon is uncorrected in the range 0.15 - 3 GeV.	157

4.12	Correction factors for the reconstructed photon energies vs the energy. Yellow curve (left) indicates correction obtained from the step 1. Green curve (middle) is the result of step 2 of photon energy correction. Since the green and yellow corrections do not converge due to the finite bin size, we make an iterative guess for the correction in between the two, and obtain the red curve. Right plot illustrates corrected two photon invariant mass normalized to $m_{\pi^0} = 0.135$ as a function of photon energy.	159
4.13	Q^2 vs ν distributions for electrons for generated MC events (left), reconstructed MC events (middle) with all selection criteria applied and data (right). Phase space considered in this analysis is enclosed within the box. For generated events phase space allowed for generation is larger in order to include bin migration effects.	161
4.14	z vs p_T^2 distributions for π^0 generated events (left), and π^0 candidate for reconstructed MC and data (middle, right) with their kinematical limits indicated in black solid lines.	161

4.15	Q^2 , ν and θ distributions on the example of the iron target. Top panel compares generated (blue) and reconstructed electrons (red), bottom panel compares electrons from reconstructed (red) events and data (black). axes, indicated by blue, red and black colors correspond to the full statistics available on iron target for generated, reconstructed MC events and data, respectively.	163
4.16	z , p_T^2 and $\phi_{(l,h)}$ distributions of π^0 's on the example of the iron target. Generated π^0 's (blue) have been selected by their MC PID, while reconstructed π^0 candidates from MC (red) and data (black), were required to be in the range $0. < M_{\gamma\gamma} < 0.3$ GeV. The colors and axes are the same as in Fig.4.15.	164
4.17	Illustration of weighting procedure. Plotted are: (1a) generated according to $e^{-5 \cdot x}$ function (an ansatz for the physical distribution further used as weight); (2a) generated flat (an ansatz for outcome of MC); (3a) events from (2a) weighted by (1a) ('corrected' generated MC); (1b) same as (1a), but smeared with a resolution $dx/x = 0.5$ ('reconstructed' physical events); (2b) same as (2a) with after smearing ('reconstructed' MC events); (3b) events from (2b) weighted by (1a) ('corrected reconstructed' MC); events (3a,b) are scaled to the sum of all weights for comparison. In reality, distribution (1a) is not known, hence the weight must be used on the reconstructed level.	165

4.18	Illustration of weighting method performed at the level of reconstructed with resolution $dx/x = 0.25$ events. Left plot are the events generated exponentially and smeared by the resolution (weight 1). Middle plot are the events generated flat and smeared by the same resolution ('reconstructed' MC). Plot on the right contains two iterations of weighting 'reconstructed' MC events: (1) red and blue lines are the result of the first iteration (2) black line is the result of the second iteration, where weight 2=weight 1/(reconstructed MC, weight 1). Comparison of the distributions on the left and the right bears slight deviations, which could be further corrected by the next iteration(s) leading to a convergent result.	166
4.19	p_T^2 distributions for measured data (black), reconstructed MC (red) and weighted MC in different levels of iteration (blue scale). After the 4th iteration of weighting reconstructed p_T^2 using parametrization in a power law $\alpha=1.5$ result still does not converge.	167

4.20	p_T^2 distributions for measured data (black), weighted generated MC (blue) and reconstructed MC (red). Left plot corresponds to the sample of π^0 candidates for which p_T^2 is reconstructed with resolution of 3σ or less from the mean value, while the right plot corresponds to the sample outside of this resolution. After the first iteration the sample on the left almost converges with data, whereas the sample on the right does not. This is the main reason we can not apply event weighting procedure to fix the shape of our π^0 candidates.	168
4.21	Fe/D ratios for reconstructed MC (purple) and data (black), with data points not corrected for acceptance. Normalization factor of MC ratios is the number of events in the data sample. Top panel ratios are plotted in Q^2 , ν , and θ ; bottom panel are z , p_T^2 , and $\phi_{(l,h)}$ for π^0 candidates in the range $0 < M_{\gamma\gamma} < 0.3$ GeV.	169
4.22	p_T^2 distributions for MC generated events (left) and reconstructed events (right) for D(black) and Fe(blue).	170
4.23	Resolution defined as $res(x) = x_i^{gen} - x_i^{rec}$ is plotted in each bin in z and fitted with a Gaussian function within 1σ	172

4.24	Top panel is resolution in z , bottom in p_T^2 . From left to right, the y-values are respectively: the width of resolution σ extracted from a Gaussian fit; relative resolution σ normalized to the mean value of z or p_T^2 in each bin; ratio of the bin width in z or p_T^2 normalized to the resolution in a given bin.	173
4.25	Illustration of 4D purities (Q^2, ν, z, p_T^2) plotted as a function of p_T^2 for the deuterium target. Each cell corresponds to (Q^2, ν) bins with z bins indicated in color code in the top right of the first cell.	175
4.26	Illustration of 3D Purities in (ν, z, p_T^2) bins plotted as a function p_T^2 in bins of incrementing ν (each histogram) and z (color code) on the example of the deuterium target. The purity is calculated by requiring that in a given set of 3D bins reconstructed π^0 kinematics match kinematics of the generated π^0 . The bin for which <i>purity</i> < 30 % ($0.3 < z < 0.4$, $2.2 < \nu < 3.2$, $0.75 < p_T^2 < 0.9$) is further excluded from the analysis. . .	176
4.27	3D Purities in (Q^2, ν, z) bins plotted as a function of z in bins of incrementing Q^2 (each histogram) and ν (color code) on the example of the deuterium target.	176
4.28	Electron acceptance in $(Q^2, \nu)=(6,6)$ bins (left) and associated relative error (right) on the example of a solid target.	182

4.29	The ratio of number of the electrons produced on a liquid target to that on deuterium: black points correspond to the uncorrected yields, while red ones were corrected using acceptance in $(Q^2, \nu)=(6,6)$ bins. From left to right are the ratios plotted for carbon, iron and lead targets. Note, that for the ease of comparing the acceptance factors between the three targets, the ratio of electron yields are normalized by the factor of deuterium to solid targets thickness (section 2.3).	183
4.30	Relative statistical error on the number of electrons before (left) and after (right) acceptance correction. The contribution from acceptance to the total statistical error is minimal.	183
4.31	$e\pi^0$ acceptance vs z in bins of ν (indicated by color) and Q^2 (each histogram) illustrated for the example of iron. Limits of integration over p_T^2 are the same as in data where maximum values reaches to $p_T^2=0.9 \text{ GeV}^2$. . .	185
4.32	Slice in the reconstructed invariant mass in (ν, z, p_T^2) bins on the example of a solid target. Fit to the invariant mass is comprised from the Gaussian plus polynomial functions, where the shape of the latter is defined using a mixed background technique based upon reconstructed photons. . .	186
4.33	$e\pi^0$ acceptance plotted as a function of p_T^2 in bins of z (indicated by color) and ν (vertical binning) integrated over Q^2 . Based on our studies of purities, we exclude one bin for which <i>purity</i> < 30 % for either target: $0.3 < z < 0.4$, $2.2 < \nu < 3.2$, and $0.75 < p_T^2 < 0.9$	187

4.34	Statistical errors in the ratio of a solid target to deuterium associated with acceptance in (Q^2, ν, z) bins (left) and in (ν, z, p_T^2) bins (right). . . .	188
5.1	Radiative correction factor (top) and the ratios (bottom) of solid targets to deuterium for the inclusive part of multiplicity ratio calculated in (Q^2, ν) bins.	191
5.2	Feynman diagrams contributing to the Born cross section (a) and the lowest order QED effects (b-e).	193
5.3	Three-dimensional distribution of π^0 candidates in (x, z, p_T) bins on the example of the iron target (data).	196
5.4	A slice of acceptance corrected p_T distributions (red points) in (x, z) bins of π^0 candidates (iron target). The p_T distribution is in the range 0-1.5 GeV. The blue lines illustrate the goodness of the Gaussian fit of p_T obtained as a simultaneous function of (x, z) . This parametrization is used as an input in nuclear structures functions, $\mathcal{H}_1$ and $\mathcal{H}_2$, in radiative correction code.	198

5.5	A slice of ϕ distributions in the range $(0, 180)^\circ$ of π^0 candidates without (black) and with (red) acceptance correction in (x, Q^2, z) bins (iron target). In each bin, ϕ distributions were initially fitted with $A+B\cos(\phi)+C\cos(2\phi)$ function. The fit showed that A, B and C depend mainly on z , thus, to simplify extraction of the function that fits cosine moments over all kinematical range, the fit was rearranged as: $f'(\phi)=A (1 + B'\cos(\phi) + C'\cos(2\phi))$ where $B'=B'(z)$ and C' is constant. Red fit is $f'(\phi)$ fitted with one parameter fixed (B'), the green fit has both parameters (B' and C') fixed.	199
5.6	The Born cross sections $\sigma_{SIDIS}^{xyz}=\sigma_T + \epsilon\sigma_L$ at fixed (Q^2, W) for deuterium calculated in two approaches: theoretical (black) and numerical (red). The black curve, integrated over $p_\perp^2$ and normalized to 2π , obtained following the formalism of equation 5.2.6. The red curve corresponds to the cross section integrated post full differential HAPRAD calculation. Comparison verifies that full HAPRAD calculations, in which nuclear structure functions enter with their own normalizations, are compatible with straightforward calculations depending only on $\mathcal{H}_1$ up to a leptonic kinematical factor.	201

5.7	Exclusive cross sections $\sigma_T + \epsilon\sigma_L$ on the proton as a function of W . Left and right plots are $Q^2=1.5 \text{ GeV}^2$ and $Q^2=3.0 \text{ GeV}^2$, respectively, both at $\theta_{\pi^0}=45^\circ$. The red curve corresponds to the HAPRAD output which employs MAID cross sections at $W<2 \text{ GeV}$ and Cornell parametrization at $W>2 \text{ GeV}$. The blue and the green cross sections are superimposed result of another code calculating structure functions on entire range: the blue curve is MAID result plotted for reference, green is VK parametrization extended for illustration to $W=1.5 \text{ GeV}$. The transition at the threshold $W=2 \text{ GeV}$, while acceptable at narrow kinematical range, is predominantly not smooth over the entire range.	205
5.8	Input cross sections $(\sigma_T + \epsilon\sigma_L)$ before radiation plotted as a function of z at fixed kinematics with M_x cut imposed. From left to right: σ_{SIDIS} , plotted in red, is compared to σ_{excl}^{VK} in green and to σ_{excl}^{MAID} in blue, respectively.	208
5.9	The top plots are the generated distribution of photon radiation according to equation 5.2.17, in regular (left) and logarithmic (right) scales. The bottom plot is the cumulative distribution of radiative factor based on equation 5.2.16 (left), and its derivative corresponding to the photon radiation before and after interaction (right). .	211

5.10	Distribution of the radiated W spectrum providing measured values are: $W = 2$ GeV and $x_B = 0.25$. Using Mo and Tsai formalism, we quantify that only 10% of events come from the range $W <$ 1.8 GeV.	213
5.11	Total radiative correction factors (semi-inclusive together with exclusive contribution) for Fe and D targets and their ratios in (Q^2, ν, z) bins. .	219
5.12	Total radiative correction factors (semi-inclusive plus exclusive contribu- tion) for Fe and D targets and their ratios in (ν, z, p_T^2) bins.	220
6.1	z vertex distribution. Original iron distribution is plotted in blue and deu- terium in red. The green dashed and solid lines illustrate stringent and loose cuts, respectively, around each of the two targets, determined based on the width of the aluminum target positioned in between them.	224
6.2	z vertex distribution of the aluminum target placed in between deuterium and iron. Its width $\sigma = 0.19$ cm, comparable with that of solid targets, is used in systematic studies as illustrated in Fig.6.1.	225
6.3	z vertex distribution of the double target system with test regions indicated by the vertical red (region 2) and blue (region 1) lines. The distance between the two targets and the distance between the test regions and the target target is the same (yellow).	226

6.4	Sampling fraction E_{tot}/P versus electron momentum $\sigma(p)$. The red lines indicate default cutoff, corresponding to $ E_{tot}/P - \mu(p) < 2.5 \cdot \sigma(p)$. The black curves illustrate tighter cutoff, which corresponds to $ E_{tot}/P - \mu(p) < 2 \cdot \sigma(p)$. The green curves are loosed cutoff, $ E_{tot}/P - \mu(p) < 3 \cdot \sigma(p)$	228
6.5	The timing cutoff (left) and respective distribution of photon β (right). The red are the default values, the green correspond to a tighter cutoff (equation 6.3.1) and the blue distributions indicate a loose cutoff (equation 6.3.2).	230
6.6	Systematic uncertainty due to photon energy cutoff in the example of iron target. RMS per multiplicity ratios are plotted together for all the bins in (ν, z, p_T^2) on the left and (Q^2, ν, z) on the right. From those histograms we extract average RMS per set of the two bins.	231
6.7	Illustration of invariant mass fits at fixed kinematics ($2.2 < \nu < 3.2$, $0.3 < z < 0.4$ and $0. < p_T^2 < 0.1$). Left and right plots illustrate fits employed for the systematical studies. Left plot contains fit in the range $(-5.5, 3.5)\sigma$ in blue, and $(-4.5, 3.5)\sigma$ in green. Right plot contains fit in the range $(-3.5, 4.5)\sigma$ in green and $(-3.5, 5.5)\sigma$ in blue. Plot in the middle illustrates the default fit in the range $(-5, 5)\sigma$ in red.	234

6.8 Systematical uncertainties in (ν, z, p_T^2) bins associated with background tilt.

Here, δR_{tilt} is the RMS of acceptance corrected multiplicities calculated within $\pm 0.5\sigma$ range on the left and right from the default value R_{π^0} 235

7.1 π^0 multiplicity ratios in (ν, z, p_T^2) set plotted as a function of p_T^2 in bins of ν (indicated at the bottom by the horizontal ν axis) and z (indicated by the color code in the first box). The bottom panel is carbon multiplicities, the middle - iron, and the top panel is lead multiplicities. The inner error bar (color) reflects statistical error in a given bin, while the outer error bar (black) corresponds to the total error, for which statistical and systematic uncertainties were added in quadratures. R_{π^0} corresponding to the $0.5 < z < 0.6$ bin are plotted at the center of p_T^2 bin, while in all other bins of z the values R_{π^0} are horizontally to the left and to the right from the p_T^2 center for visualization. 245

7.2	π^0 multiplicity ratios in (Q^2, ν, z) set plotted as a function of z in bins of Q^2 (indicated by the horizontal Q^2 axis) and ν (indicated by the color). The red color gradient corresponds to the carbon target, the blue - to the iron, and the black - to the lead target. The inner error bar (color) reflects statistical error in a given bin, while the outer error bar (black) corresponds to the total error. R_{π^0} corresponding to the $3.2 < \nu < 3.73$ bin are plotted at the center of z bin, while in other two bins in ν R_{π^0} are shifted to the left and to the right from the center of z -bin for visualization.	246
7.3	One-dimensional π^0 multiplicity ratios in Q^2 , ν , z , and p_T^2 bins. The error bars reflect statistical errors only.	247
A.1	MC distribution of the number of events assumed to be the invariant mass of the two photons. By construction, the peak is purely a Gaussian and the background is the first order polynomial. The statistical fluctuation on the number of events in the peak, relevant to the fitting method, are given by the normalization of the Gaussian fit (± 170 events). The fluctuations stemming from the sideband subtraction method are defined by the statistics under the peak in the full 6σ (± 200 events). As anticipated, the errors of two methods happen to agree, moderately.	258

A.2 The distributions of number of counts, generated 1k times, filled with: the difference between number of events calculated using two methods (top left - Gaussian fit, top right -sideband subtraction) relative to the generated number of events; the mean values (bottom left) of the two methods, and their absolute values (bottom right). The RMS and mean values of the top two histogram agree well with the predicted, by construction, fluctuations of the method themselves. The mean value calculated by taking the difference of the two methods, unlike its absolute value, proves to be a relevant estimation of the systematic error. . . . 259

LIST OF TABLES

3.1	Carbon target coefficients for the sampling fraction.	87
3.2	Iron target coefficients for the sampling fraction.	87
3.3	Lead target coefficients for the sampling fraction.	87
3.4	Fraction of the events rejected for each electron cut.	104
3.5	Fraction of the events rejected for each photon cut.	104
3.6	Number of bins and corresponding bin limits in Q^2 , ν , z , and p_T^2	130
4.1	Set of bins in $(Q^2, \nu) = (6, 6)$ used to calculated acceptance factors and correct electron yields.	181
5.1	Calculation grid for exclusive cross sections.	214
6.1	Summary of averaged systematic uncertainties in (Q^2, ν, z) bins for C, Fe and Pb multiplicities. Values are given in percentage from the values of multiplicity ratio.	241
6.2	Summary of averaged systematic uncertainties in (ν, z, p_T^2) bins for C, Fe and Pb multiplicities. Values are given in percentage from the values of multiplicity ratio.	242

Introduction

The hypothesis of confinement which stems from the observation that only color singlets propagate as free asymptotic states is still one of the deepest puzzles of the fundamental theory of the strong interaction, Quantum Chromo-Dynamics(QCD). Our modern understanding of the strong interaction was conceived in the late 1960's with development of the parton model, and flourished a decade later with the appearance of asymptotic freedom and confinement. So far, no free colored objects was observed in experiments, including quarks and gluons, for the potential energy between two colored objects in QCD is weak at small relative distances (asymptotic freedom), while at large distances it grows strongly with separation (confinement). Intimately related to confinement is the processes of hadronization. In this process, a struck parton, briefly liberated and traveling as a 'free' colored particle, evolves in space-time. Respecting the color charge neutrality, it subsequently emerges as a colorless hadron. In hard collisions, the parton is produced with a large virtuality which is then reduced by a sequence of partonic emissions. The development of the partonic shower can be calculated using perturbative methods of QCD, namely DGLAP equations, as long as the virtuality at any interaction vertex is large compared to the QCD scale (Λ_{QCD}^2). Once the virtuality drops, the strength of the interaction α_s becomes such that the perturbative techniques are no longer applicable. A feature of high energy interactions,

known as factorization, allows one to separate hadronization process into two parts. The first part involves hard partonic interactions, calculable using pQCD. The second part requires a non-perturbative approach to address the question of how hadrons are built out of quarks and gluons. Therefore, as hadronization is intrinsically a non-perturbative process, our understanding of it relies heavily on phenomenological models, to verify which experimental datasets are of a valuable importance.

In deep-inelastic scattering, the extent of the perturbative regime depends upon the initial energy the electron transferred to the quark, the fraction carried away by the detected hadron and the size of the medium itself. One way to study hadronization is to perturb the nuclear environment surrounding it, which is a sensitive probe to study the space-time dynamics of hadronization by means of comparing properties of final states produced on nuclei of varying sizes. Nuclear modifications of hadron production, normalized to the corresponding processes on the proton, have been extensively studied in deep-inelastic lepton nucleus scattering nDIS ($l^\pm A$), in hadron-nucleus (hA) and heavy-ion collisions (AA). The advantage of nDIS is the known properties of the environment, while leptoproduction has a virtue that the energy and momentum of struck quark are well determined by tagging the scattered lepton. By using nuclei of increasing size, the time development of hadronization can be inferred: if hadrons are produced at small distances compared to nuclear size, the relevant interactions in the medium

involve well-known hadronic cross section; if, on the other hand, hadronization takes place at large distances, the interaction involves partonic mechanisms accompanied by the emission of gluons and production of quark-antiquark pairs. The two mechanisms, or most likely their combination, lead to different predictions of production yields of observed hadrons, known as attenuation, on the nuclei as compared to that on a nucleon. The hadron attenuation data has been explained in terms of two different processes: interaction of color singlet prehadron with the surrounding medium (partonic energy loss type of models) and modification of in-medium fragmentation functions (hadron absorption type of models). During the last decade, the HERMES collaboration has published pioneering results on a wide range of hadrons, nevertheless, it has thus far not been possible to determine the relative magnitude of these two processes from available data. The CLAS data offers two orders of magnitude more integrated luminosity, lower energies and a wide range of target masses. Such experimental conditions provide an access to uniquely determine the relative contributions of the two mechanisms. Much theoretical work is, however, required for definitive results. It is the purpose of this thesis to present new experimental data on neutral pion attenuation using semi-inclusive deep inelastic scattering (SIDIS).

The thesis is organized as follows. In the first part of Chapter 1, the semi-inclusive deep inelastic scattering in the framework of quark parton model is discussed. The second part of Chapter 1 focuses on hadronization in nuclear medium

and reviews theoretical models along with currently available experimental data. Chapter 2 provides a description of the CLAS apparatus and experimental set up. In Chapter 3, the particle identification of electron and π^0 is detailed. The MC simulations, extraction of acceptance correction factors and purities are presented in Chapter 4. In Chapter 5 we discuss procedures, and present results on the radiative corrections for the inclusive and semi-inclusive processes. Systematic uncertainties are presented in Chapter 6, while final results and discussions are summarized in the Chapter 7.

Chapter 1

Physics Motivation

1.1 Deep inelastic scattering

Deep inelastic scattering (DIS) is one of the main tools to study hadron structure. This process is generically of the form: $l(k) + h(p) \rightarrow l'(k') + X$, where $l(k)$ and $l(k')$ represents initial and scattered lepton of momentum k^μ and k'^μ , h is a target hadron carrying momentum p^μ , and X is an arbitrary hadronic state. The process is initiated by the exchange of a vector boson V . Generally, the incoming lepton may be an electron, muon or neutrino, and the exchanged boson photon, $W^\pm$, or Z . Normally, h will be a nucleon or a nucleus. The initial and the final momentum of the scattered lepton are k^μ and k'^μ , respectively. Here we will only treat pure electromagnetic interaction with the incoming electron and single exchanged virtual photon γ^* . At CLAS kinematics, the energy of lepton beam (5 GeV) is well below M_Z , therefore weak interaction can be neglected. The momentum transferred between lepton and hadron is this of the virtual photon:

$$q = k - k' \tag{1.1.1}$$

The virtuality of the probe (Q^2) is:

$$Q^2 \approx 4E_k E_{k'} \sin^2 \left(\frac{\theta}{2} \right) \quad (1.1.2)$$

$$-q^2 = Q^2$$

where E_k and $E_{k'}$ are the energy of incoming and scattered lepton, respectively, and θ is the polar angle of the lepton. The 'deep' refers to the high virtuality of

Fig. 1.1: Schematic drawing of the deep inelastic scattering process. Incoming lepton l scatters off a nucleon h exchanging a virtual boson V .

the photon which gives an ability to probe distances that are smaller compared to the size of the target hadron. The range of virtualities to resolve a parton is not precisely known but it is assumed to be $Q^2 > 1 \text{ GeV}^2$ which corresponds to spatial resolution of the order of 0.2 fm. The energy transferred from the lepton to the target is defined as:

$$\nu = \frac{p \cdot q}{M} (\nu = E_k - E_{k'} \text{ in the target rest frame}) \quad (1.1.3)$$

In DIS the hadronic final state X invariant mass is much larger than that of a nucleon. The squared mass of the total hadronic system is ($M_x^2 \equiv W^2$):

$$W^2 = (p + q)^2 = M^2 + 2M\nu - Q^2 \quad (1.1.4)$$

Here M is the mass of the target. W^2 also indicates the inelasticity of the reaction.

If the scattering is elastic then $W^2=M^2$ and:

$$2M\nu - Q^2 = 0 \quad (1.1.5)$$

The Bjorken scaling variable:

$$x = \frac{Q^2}{2M\nu} \quad (1.1.6)$$

For an inelastic process $0 < x < 1$ and for the elastic process x is 1.

Another dimensionless variable y is related to ν , and denotes the fractional energy of the virtual photon:

$$y = \frac{p \cdot q}{p \cdot k} \left(y = \frac{\nu}{E_k}, \text{ in the target rest frame} \right) \quad (1.1.7)$$

DIS can be parametrized with all listed variables because the initial and the final electrons are observed. To separate DIS from other processes, for instance, resonance production or the diffractive regime, scattering is required to occur on the constituents of the nucleon ($Q^2 > 1 \text{ GeV}^2$), and the mass of the final states must exceed hadronic resonances ($W > 2 \text{ GeV}$). In case of inclusive reactions only the scattered lepton is detected, while in the semi-inclusive (SIDIS) process, additionally one final-state hadron is detected. In SIDIS the four-momentum of the

detected hadron p_h can provide information on the kinematics of the struck quark if the hadron was detected at high z (for low z it usually will not). The variable z is defined as fraction of the virtual photon energy carried by the produced hadron and is given by:

$$z = \frac{p \cdot p_h}{p \cdot q} \left(z = \frac{E_h}{\nu} \text{ in the target rest frame} \right) \quad (1.1.8)$$

Here, E_h is the energy of the final-state hadron. The transverse momentum of the hadron p_T is defined with respect to the direction of the virtual photon. The angle between leptonic and hadronic plane is ϕ_h as illustrated in Fig.1.2.

Fig. 1.2: Illustration of the SIDIS process $l(k) + h(p) \rightarrow l'(k') + h(p) + X$. Lepton momenta $\vec{k}$ and $\vec{k}'$ define the scattering plane, while the hadron plane is defined by the direction of virtual photon $\vec{q}$ and that of the hadron $\vec{p}_h$. The angle between leptonic and hadronic plane is ϕ_h .

1.1.1 DIS cross section

The differential DIS cross section for a scattered lepton in the one-photon exchange approximation inside solid angle $d\Omega$ in the laboratory frame with energy

between E' and $E+dE'$ is given by:

$$\frac{d^2\sigma}{dE'd\Omega} = \frac{\alpha^2}{2MQ^4} \cdot \frac{E'}{E} \cdot L_{\mu\nu} W^{\mu\nu} \quad (1.1.9)$$

where $\alpha^2 = \frac{e^2}{4\pi}$ is the electromagnetic coupling, $L_{\mu\nu}$ is the leptonic tensor describing leptonic interactions ($l \rightarrow l' + \gamma^*$), and $W^{\mu\nu}$ is the hadronic tensor which carries information about the structure of the target nucleon ($p + \gamma^* \rightarrow X$). The spin-averaged leptonic tensor for unpolarized electrons can be calculated using QED. The hadronic tensor is parametrized by functions describing the hadron in terms of parton structures and can be studied through the measurement of the cross section. For the unpolarized photon exchange reaction, the parametrization of the hadronic tensor is accomplished by introducing two structure functions, W_1 and W_2 which depend on x and Q^2 [1]. Contracting the hadronic with the leptonic tensor yields the following DIS cross section:

$$\frac{d^2\sigma}{dE'd\Omega} = \frac{\alpha^2}{4E^2\sin^4(\theta/2)} \left(W_2(x, Q^2)\cos^2\left(\frac{\theta}{2}\right) + 2W_1(x, Q^2)\sin^2\left(\frac{\theta}{2}\right) \right) \quad (1.1.10)$$

This result is similar to the Rosenbluth formula describing elastic scattering of an electron off a Dirac particle. At large Q^2 , inelastic electron-proton scattering can be considered as elastic scattering of the electron on a 'free' quark inside the proton. It should be possible therefore to obtain the Rosenbluth formula from equation 1.1.10 by the appropriate choice of W 's. The functions W_1 and W_2 are usually replaced, for the purposes of presenting data, by alternate and equivalent

structure functions F_1 and F_2 , which will turn out to be particularly simple in the parton model.

1.1.2 Quark-parton model

In the parton model we assume that hadrons are extended objects made of constituents which are held together by their mutual interaction. The underlying theory of hadron structure in QCD has so-called asymptotic freedom, i.e. at short distances partons (quarks) behave as if they were essentially free particles. At larger distance the strong coupling increases giving rise to confinement. The latter explains why quarks are strongly bound together in color neutral objects. The parton model is applicable to any hadronic cross section involving a large momentum transfer. At low energies the photon 'sees' the proton as a point-like particle - this gives the Mott cross section. At intermediate energies of elastic scattering the photon probes the proton structure - this gives the Rosenbluth formula and elastic form factors. At high energies and short wavelengths, the photon interacts with a single, essentially free, quark - this gives Bjorken scaling and Callan-Gross relation.

In the late 1960's, Bjorken predicted [3] that in the deep inelastic scattering regime the dependence of structure functions on Q^2 fades away and only depen-

dence on x remains providing Q^2 and ν are sufficiently large and x is fixed:

$$MW_1(Q^2, x) \equiv F_1(Q^2, x) \rightarrow F_1(x) \quad (1.1.11)$$

$$\frac{Q^2}{2Mx} W_2(Q^2, x) \equiv F_2(Q^2, x) \rightarrow F_2(x) \quad (1.1.12)$$

where M is the mass of the proton, and x is the fraction of nucleon momentum carried by the parton (coincides with Bjorken x) in the infinite momentum frame. The first experimental evidence of this behavior, known as 'scaling', was observed at Stanford Linear Accelerator Center (SLAC) and became the confirmation of the parton model.

The interpretation of DIS in terms of parton distributions is possible in a reference of infinite momentum frame (IMF). In this frame the nucleon of infinite momentum moves along the z -axis. Then the rest masses and transverse momenta of the proton constituents are much smaller than the longitudinal component, and therefore can be neglected. Contrary to the center-of-mass frame, in the IMF interaction of the virtual photon with vacuum fluctuations are suppressed thus allowing direct access to the structure of the proton. Two important things happen to a hadron moving in the IMF: it is Lorentz contracted in the direction of the collision and processes connected to its internal structure are time-dilated. The latter leads to the increase of lifetime of the partonic state relative to the time it takes for the electron to interact with the hadron. When the time of virtual photon interactions is short, the hadron can be considered as a state characterized by a definite number of partons. The rate at which partons interact with one

another is slowed down and can be neglected. Then each parton may be thought of as carrying a fraction x of the total hadron momentum. In such a scenario, the electron interacts with a single, essentially free parton.

Fig. 1.3: SLAC measurement of the Callan-Gross ratio $\frac{2xF_2}{F_1}$ plotted versus $x = \frac{Q^2}{2Mv}$.

The plot was taken from [2].

With these assumptions, the cross section for hadron scattering can be computed by combining probabilities rather than amplitudes. Up to kinematical factors, the scattering is directly proportional to the density of partons 'frozen' over a short scattering time-scale:

$$F_2(x) = x \sum_i e_i^2 q_i(x) \quad (1.1.13)$$

where e_i is the quark charge, $q_i(x)$ is the parton distribution function (PDF) which represents the probability that the electron will encounter a parton of species i with the fraction x of hadron momentum, and the sum runs over all flavors of

quarks and antiquarks. Callan and Gross suggested that in the high Q^2 limit, F_2 is related to F_1 as: $F_2(x)=2xF_1(x)$. This relation was confirmed at SLAC and elsewhere, implying the fact that proton consists of point-like spin one-half particles. A measurement of the ratio $\frac{2xF_2}{F_1}$ is illustrated in Fig. 1.3.

1.1.3 Quark-parton model and pQCD

The scaling prediction by Bjorken is related to the assumption that the average transverse momentum of the partons in the IMF is small compared to the longitudinal momentum. However, the radiation of gluons from the quarks violates this assumption. As seen in Fig. 1.4, the F_2 structure function is basically independent of Q^2 in the intermediate range of x . Yet, at low and high x , the structure functions were observed to depend on Q^2 , the phenomenon known as Bjorken scaling violation. At high Q^2 it manifests as an increased probability to find a parton carrying small x and a decreased chance of finding one at high x . The behavior is related to the fact that high momentum quarks lose their energy by gluon radiation. Quarks interact with each other by exchanges of gluons, which are the carriers of force in QCD. Unlike photons, gluons carry charge (color) and therefore interact with one another as well as with quarks. The dependence of the coupling on the scale of the strong interaction in the first order QCD is given by:

$$\alpha_s(\mu^2) = \frac{12\pi}{(33 - 2N_f \ln(\mu^2/\Lambda^2))} \quad (1.1.14)$$

Fig. 1.4: The proton structure function F_2^p given at two values of Q^2 . The factors shown in brackets represent normalization of various datasets. The plot was taken from [4].

where N_f is the number of quark flavors, μ is renormalization scale, and Λ is the characteristic scale of applicability of pQCD. The renormalization scale helps to solve the problem of divergences at high virtualities of radiated gluons. In DIS this scale is set to Q . The application of the QCD expansion is only valid for $\alpha_s \ll 1$, which is satisfied if $\mu^2 \gg \Lambda^2$. It is hard to determine Λ from experimental data but it appears to be in the range: $100 \text{ MeV} < \Lambda < 500 \text{ MeV}$ [5].

The gluon radiation produces the Q^2 evolution of parton densities. The evolution of PDF in QCD is described by Dokshitzer-Gribov-Lipatov-Altarelli-

Fig. 1.5: Strong coupling $\alpha_s(Q^2)/\pi$ obtained from JLab spin structure function data and the world data [6].

Parisi (DGLAP) equations:

$$\begin{aligned} \frac{\delta q_f(x, Q^2)}{\delta \ln Q^2} &= \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{dx'}{x'} \left[q_f(x', Q^2) \cdot P_{qq}\left(\frac{x}{x'}\right) + g(x', Q^2) \cdot P_{qg}\left(\frac{x}{x'}\right) \right] \\ \frac{\delta g(x, Q^2)}{\delta \ln Q^2} &= \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{x}{x'} \left[g(x', Q^2) \cdot P_{gg}\left(\frac{dx'}{x'}\right) + \sum_f q_f(x', Q^2) \cdot P_{qg}\left(\frac{x}{x'}\right) \right] \end{aligned} \quad (1.1.15)$$

where $q_f(x', Q^2)$ and $g(x', Q^2)$ are the parton and gluon densities respectively, and $P_{ij} \frac{dx'}{x'}$ are the so-called splitting functions. $P_{ij}(\frac{x}{x'})$ gives the probability that the parton (gluon) i with momentum fraction x was radiated from a parton j with momentum fraction x' . DGLAP equations cannot predict *a priori* the dependence of PDF on x . In order to resolve the above differential equation, a parametrization of initial distribution function at a given Q_0^2 is necessary. Thus, PDF are parametrized at a starting point, Q_0^2 , and only then the predictions of QCD are

compared to the data at other scales. Typically they are extracted from 'global QCD fits' to inclusive hadron electroproduction in lepton-nucleon DIS.

1.1.4 Factorization

Factorization theorems allow to separate the DIS cross section by convoluting the three parts, depending on the distance scale: (a) the hard scattering cross section of a lepton off parton, which can be calculated exactly from QED; (b) the parton distribution function (PDF) $q_f(x, Q^2)$ which at leading order is a probability to find a parton of fractional momentum x ; and (c) the fragmentation function $D_f^h(z, Q^2)$ (FF) which represents the probability of a parton f to fragment into observed hadron h of a fractional energy z . Typically, PDF are extracted from the fits to global data of inclusive hadron production in lepton-nucleon DIS, for which FF are such that $D_f^h(z, Q^2) \rightarrow 1$. The FF for large hadron fractional momentum ($z > 0.1$) can be calculated from the DGLAP evolution equations. In such calculations, the value for the initial scale Q_0^2 is obtained via the measurements of electron-positron annihilation into hadrons.

In the quark parton model, the hadron production cross section in the process $eN \rightarrow ehX$, integrated over transverse momentum of the hadron, is given by:

$$\frac{d^3\sigma_{SIDIS}}{dx dQ^2 dz} = \frac{\sum_f e_f^2 q_f(x, Q^2) \cdot D_f^h(z, Q^2)}{\sum_f e_f^2 q_f(x, Q^2)} \cdot \frac{d^2\sigma_{DIS}}{dx dQ^2} \quad (1.1.16)$$

Assuming the above factorized form of SIDIS cross section, hadronic multiplicity per DIS event is:

$$N_h(Q^2, z) = \frac{\sum_f e_f^2 q_f(x, Q^2) \cdot D_f^h(z, Q^2)}{\sum_f e_f^2 q_f(x, Q^2)} \quad (1.1.17)$$

The factorization theorem has been proven for the leading twist of the structure functions and to all orders of α_s [7], however, not proven for higher twist contributions. An important consequence of the factorization theorem is the universality, or process-independence, of PDF and FF. However, hadron production in nuclear systems were experimentally observed to break down the universality: PDF and FF in nucleus are different from those in a free nucleon at the probing scale Q_0 (the EMC effect outlined in section 1.2.3).

1.2 Hadronization

The process of fragmentation of quarks and gluons into observed hadrons is a fundamental process of QCD, commonly referred to as hadronization. Due to the property of confinement, an energetic parton cannot propagate in the vacuum as a free particle, instead it hadronizes sharing its energy with produced jets of hadrons, thus fulfilling the requirement of color neutralization. The colored objects, partons and gluons, cannot be observed directly, yet their properties may be inferred through indirect means, such a measurements of collision products.

The processes of hadronization can be conditionally divided into two parts. The first part, related to parton propagation followed by gluon emission, can

be described by employing perturbative methods. In a broad sense, perturbative means describing the strong interaction domain in the language of quarks and gluons (short distance, small coupling) [13]. The second part addresses the question of how hadrons are built out of quarks and gluons, and requires a non-perturbative approach (large distance effects). Due to the difficulty in carrying out calculations in the non-perturbative domain, a number of approximate theoretical models were built. Recent developments of lattice QCD [11] study the behavior of heavy quarks at large distances based on the linearly rising potential. Though providing first strong theoretical input on the quark-quark forces, lattice calculations are limited to static potential, which has no known connection to the real-world QCD. Light-quark confinement is a dynamical phenomenon, which cannot be expressed via static forces but can be connected with the analytic properties of QCD Dyson-Schwinger equations (DSE) [12]. DSE is a differential equation relating dressed-propagators which require ad-hoc assumptions for the constituent quark mass. Given the complexity of the first principle calculations, experimental guidance is crucial to support theoretical developments. Atomic nuclei of varying sizes, employed for those studies, act as a femtodetectors: by varying its size at medium energies, the space-time dynamics of hadronization is varied, enforcing it to occur inside or outside the nuclear medium. Particles, produced in the collisions on the nuclear targets and measured at macroscopic distances, carry important information on the characteristics of time-distance scales of hadronization. To in-

fer this information in the case of the DIS, one of the derived quantities would be the hadronic multiplicity R_A^h , an observable that measures the ratio of number of hadrons produced in the heavy relative to the light nuclei. At leading order, this multiplicity corresponds to the ratio of FF in cold nuclear medium (nucleus A) to those in 'vacuum' (deuterium). One can also use hadronization process to investigate the properties of the medium itself, for example, the dense medium produced at heavy-ion collisions. This, however, requires that hadronization dynamics are reliably understood.

1.2.1 Hadronization in vacuum

The virtual photon from an incident lepton is absorbed by the quark (strictly speaking, for this picture to be valid, $x_B > 0.1$ to avoid quark-pair production). Following the hard scattering, the quark regenerates its color field and loses energy via gluon emission during the time, referred to as production time. At the moment when the gluon emission ceases, the quark picks up a sea antiquark, thus forming a colorless dipole, or a 'prehadron'. The later will emerge as final-state hadron after the time it takes to form full hadronic wave function, i.e - formation time. Note, that production and formation times are introduced as phenomenological rather than the well defined quantities in order to separate two processes: the propagation of asymptotically free parton treated via the pCQD methods and formation of colorless dipole addressed by the non-perturbative approaches. The

Fig. 1.6: A classical picture of leading hadron production taken from [14]. The parton, following hard scattering, regenerates its color field via gluon emission during the production time. Once the color is neutralized, a formation time begins until the full hadronic wave function is formed.

physical picture of the leading hadron production in the DIS process is illustrated in Fig. 1.6.

The basic process of quark fragmentation in a vacuum was first described using the recollection of the data from the single-inclusive hadron production in e^+e^- annihilation (SIA) [9]. In this process, illustrated in Fig. 1.7, an e^+e^- pair annihilates into Z/γ , which in turn decays into $q\bar{q}$ pair. The propagating $q\bar{q}$ pair expands like a string, emitting gluon radiation. When the string tension saturates, the $q\bar{q}$ pair fragments into a number of colorless sprays of hadrons, collimated into two back-to-back jets. However, such probabilistic parton evolution picture, which became an inspiration for the first Monte Carlo generators, is as approximate as it is limited [19]. Combined analysis of the hadron spectra from a large set of e^+e^- data, performed by a number of theoretical groups, has lead to an extraction of quark fragmentation functions for pions and kaons. The most simple functional

Fig. 1.7: Illustration of $e^+e^- \rightarrow (Z/\gamma) \rightarrow H$ process. Here, D_j^h is the FF given by a probability that the parton of flavor j fragments into observed hadron h with fractional momentum z . As e^+e^- process does not proceed through hadronization stage, it allows direct extraction of the FF's.

form, based on the Lund string model, was chosen to be: $D_j^h \sim z^\alpha(1-z)^\beta$, where α and β are the free parameters, and D_j^h is calculated at some initial scale μ_0 for the Q^2 -evolution. The fits confirm the scale dependence of the fragmentation functions D_j^h as predicted by the DGLAP evolution equation. The SIA analysis, nevertheless, bears a serious limitation to disentangle between the favored and unfavored FF. The latter, together with universality and an explicit check of factorization, was established in the global analysis that compared the SIA results with those extracted in SIDIS using the lepton-proton and proton-proton reactions.

1.2.2 Hadronization in nuclear environment

Why a nuclear target?

One way to study parton fragmentation and hadronization is to perturb the surrounding environment by introducing a nuclear medium. The influence of the nuclear medium on the evolution of hadronization will have measurable consequences. They are signaled by the modification the final hadron yield distributions on the nuclei as compared to those produced in 'vacuum'. By measuring those modifications, one can access important information both on the early stages of hadronization that take place at the femtometer distances from the origin of hard interaction, as well as on the quark-gluon system as it propagates through the nuclear medium and interacts with it. Such an approach works extremely well in the window of medium energies DIS where a nuclei with well-defined properties is employed ("cold QCD matter"). There, the hadronization distances are of the order of the nuclear diameter; the kinematics is known provided the beam energy and the momenta of reaction products; the multiplicity of the final states is relatively low, allowing precise measurements. The unique ability of the DIS reaction to have a good handle on the kinematical variables is an important advantage that allows to isolate events containing initial fast moving quark (leading hadrons) struck by the virtual photon.

Physical picture in nuclear medium

As the parton travels in a vacuum, it 'dresses' in a color field of bound partons, which then evolve into the observed hadrons. The same process of dynamical parton dressing takes place as the parton transverses the nuclear medium, however, now it is modified by medium interactions. The process of in-medium hadronization is illustrated in Fig. 1.8. As the quark transverses the color field of the nuclei, it undergoes multiple interactions with the medium, which could be viewed as quark-quark interactions. The exact cross section of such processes is not known, although analysis of experimental data (HERMES, Drell-Yan) indicates that the interaction cross section is small, of the order of mb. The interpretation of low interactions probability is related to the cancellation of induced inelastic radiation emitted by a quark due to the destructive interference of gluons radiated at the two scattering centers (formation zone phenomena of Landau, Pomeranchuk, Migdal) [15]. This in turn manifests as a very small probability of quark being absorpted within the nuclear medium. The quark passage through the medium is instead signaled by multiple collisions which result in additional gluon bremsstrahlung (collisional energy losses). The in-medium scattering leads to an increase, on average, of the quark's primordial transverse momenta relative the initial virtual photon direction. The formation time for the propagating quark to neutralize its color depends on its virtuality and energy when it was produced. A colorless dipole, or prehadron, can be formed entirely inside the nuclear

Fig. 1.8: Schematic illustration of the parton interaction as it travels inside the nuclear medium. Produced hadron can be formed outside (left) or inside (right) the nucleus, or somewhere in between. The cartoon is taken from [45].

medium, or outside, or somewhere in between due the interference of the two amplitudes. The prehadron is likely to develop an inelastic cross section comparable with the ordinary hadronic cross section of the order of $O(40 \text{ mb})$ [8]. Therefore, if the prehadron is formed within the nuclei, its interaction with the surrounding medium will be dominantly inelastic, leading to the hadron absorption. Thus, to gain insight on prehadron evolution and interaction, one can either employ a nuclei of large size ensuring that prehadron formation occurs inside the medium, or, at small x_B , one can study the diffractive processes in which the $q\bar{q}$ pair is produced directly from the virtual photon [16].

1.2.3 Nuclear effects in SIDIS

In order to examine the effects pertaining to the nuclear medium, one compares measurements carried out on the nuclear target relative to those on the deuterium target. A heavy target contains a combination of both protons and neutrons,

affected by nuclear binding and Fermi motion. For that reason, a deuterium target is preferred instead of a hydrogen one. Using a deuterium target as a reference helps to minimize the effects caused by the difference of the structure function $F_2(x)$ on the proton and the neutron. In the scenario that the nuclear target is asymmetric in the number of protons and neutrons, isoscalar corrections become relevant.

Fig. 1.9: The EMC effect on carbon relative to deuterium. The solid curve is the fit from SLAC to the carbon ratio. The plot is taken from [18].

In 1983, the European Muon Collaboration (EMC) discovered that per nucleon deep-inelastic structure function $F_2(x)$ on iron differs from that on deuterium. This phenomenon, called the EMC effect, manifests as clear suppression of high momentum quarks on the range $0.3 < x < 0.8$ in the nucleon belonging to the iron nuclei relative to that of deuterium. This observation has been further confirmed by the measurements at SLAC [17] and JLab [18]. While principal features of

this effect have been explain, and modeling over all A and x has been extensively studied, there is still no consensus on a unique description.

1.2.4 Observables

The commonly used experimental observables are presented in terms of the hadronic multiplicity ratio R_A^h and transverse momentum broadening Δp_T^2 . In order to isolate the effects due to nuclear modification, a reference nucleus is used, which could be a proton, deuterium or even carbon. Access to the detailed multivariate dependencies of R_A^h is crucial in such types of studies, and thus requires high luminosity experiments.

When a parton propagates through the medium, it experiences multiple collisions and performs Brownian motion in transverse direction. The rate at which a traveling parton emits gluons due to multiple parton scattering is expected to be slightly greater than the rate of radiation which occurs in a vacuum. Those effects contribute to an increase of the width of the transverse momentum distribution of the final state hadron. The transverse momentum broadening for the observed final state hadron with respect to the direction of the virtual photon is defined as [8]: $\Delta\langle p_T^2 \rangle = \langle p_T^2 \rangle_A - \langle p_T^2 \rangle_D$, where $\langle p_T^2 \rangle_A$ is an average hadron momentum squared produced on a nuclear target A , and $\langle p_T^2 \rangle_D$ is the same quantity for a deuterium target. Since the hadron $\langle p_T^2 \rangle$ is mainly accumulated from the elastic scattering at the stage of quark propagation and gluon emission, it is supposed to

be a sensitive probe to the quark lifetime, i.e. production time, as well as medium-stimulated energy losses. This observable is a topic for a separate discussion, not covered in this work.

The hadron multiplicity ratio or attenuation ratio represents the ratio of the number of hadrons of type h produced per DIS event on a nuclear target of mass A to that from a deuterium target:

$$R_A^h(\nu, Q^2, z, p_T, \phi) = \frac{\left. \frac{N_h(\nu, Q^2, z, p_T, \phi)}{N_e(\nu, Q^2)|_{\text{DIS}}} \right|_A}{\left. \frac{N_h(\nu, Q^2, z, p_T, \phi)}{N_e(\nu, Q^2)|_{\text{DIS}}} \right|_D} \quad (1.2.1)$$

where N_h is the yield of semi-inclusive hadrons in a (ν, Q^2, z, p_T, ϕ) bin and N_e is the number of inclusive DIS events in the (ν, Q^2) bin. Normalizing hadron yields to the DIS events allows one to practically cancel initial state effects such as nuclear modification of PDF. The ratio R_A^h can be analyzed within a model to determine the formation lengths of hadrons as a function of the relevant variables, typically Q^2 , p_T^2 , z , and ϕ_h .

Recent HERMES measurements observed the dependence of attenuation as a function of atomic number and the hadron type. At large z this ratio is below unity indicating the breakdown of the universality of FF.

1.2.5 Overview of existing data

A wealth of data has been collected in the past four decades on hadron collisions and leptonproduction experiments. The latter we briefly discuss in this section. Electroproduction of the hadrons from various nuclear targets was first

investigated by Osborne *et.al* in the earlier 1970's at SLAC [38]. They utilized a 20.5 GeV electron beam incident on number of targets: ^2D , ^9Be , ^{12}C , ^{64}Cu , and ^{119}Tl . Due to detection limitations, the measurements of hadrons were summed up based on the charge. The ratio of single inclusive hadrons per target density (g per cm^2) to the same number for deuterium was measured as a function of z and p_T . The attenuation R_A^h of hadrons was observed for the first time in the semi-inclusive cross section ratio, clearly showing evidence to increase with the size of target nucleus. In addition, the nuclear absorption increases for the forward hadrons (higher z). The caveat of this measurement is the fact that unlike multiplicity ratio defined by equation 1.2.1, this observable is sensitive to the initial-state effects as it does not account for the nuclear modification of PDF (EMC effect). One of the first pioneering measurements with ultra-high energy muon beam was conducted in FNAL [39], and further studied, at higher luminosities at CERN, by the European Muon Collaboration [41] (well known for the EMC effect, discussed in section 1.2.3). Nuclear targets (^{12}C , ^{63}Cu , ^{119}Sn) and ^2D were measured for the first time simultaneously which canceled most of the systematic uncertainties in the multiplicity ratio. R_A^h was measured in the range of $20 < \nu < 220$ GeV in two Q^2 bins, and as a function of p_T in two ν bins. For large nuclei (Cu, Sn), a distinct reduction in the multiplicity ratio was observed for the fast hadrons (high z), whereas for carbon it was consistent with unity over an entire range in z . A slight 10% decrease of multiplicities below unity was

observed for energies $\nu < 60$ GeV. The p_T dependence of the ratio was observed to rise above unity at high p_T 1.10, which is analogous to the earlier reported Cronin effect [42] in hadron-nucleus collisions.

To summarize, the earlier measurements, from SLAC with electron beams

Fig. 1.10: EMC measurements of the multiplicity ratios as a function of p_T^2 in two ν bins [41].

and CERN with muons beam, have shown that multiplicity ratio mainly depends on the variables ν and z . This was confirmed later in 1990's by the Fermilab E665 experiment [40] carried out with 490 GeV muon beam incident on ^2D and ^{132}Xe targets. The performed measurements have also shown that the optimal transferred energy ν for studying nuclear effects on the multiplicity ratio ranges from a few GeV to a few tens of GeV [8].

The latest generation of experiments conducted in HERMES at DESY and CLAS at Jefferson Lab stimulated a new wave of interest in the extensive studies

of hadronization. HERMES conducted a series of SIDIS measurements with a 27.6 GeV positron beam incident on ^2D , ^4He , ^{14}N , ^{83}Kr and ^{131}Xe targets. Multiplicity ratios were presented for the various hadrons ($\pi^\pm, \pi^0, K^\pm, p, \bar{p}$) as a function of ν , z , Q^2 and p_T . A two dimensional analysis in z , ν and p_T bins (individually) was presented [43]. The data supported earlier results on the decrease of R_A^h with increase of the nuclear size. R_A^h is almost independent of p_T , except for high values of p_T (similar to the “Cronin” effect). For the first time, data showed more attenuation for higher values of z and less attenuation with increase of ν (averaged over all other kinematic variables). This behavior may be explained as being due to an increase of the formation length at high ν , resulting in a larger fraction of hadronization occurring outside of the nucleus. The average R_h for $\pi^\pm, \pi^0, K^+$ are similar, however, for K^-, p , and $\bar{p}$ significant difference is observed. A clear $A^{\frac{1}{3}}$ dependence of the Δp_T is observed for the first time. As a result of HERMES measurements, an extensive set of data to guide hadronization modeling has been collected. The results presented for π^0 are, however, limited to one-dimensional dependencies as illustrated in Fig. 1.11.

Fig. 1.11: HERMES one-dimensional π^0 multiplicity ratios in ν , z , Q^2 , and p_T^2 bins.

Kinematical coverage is: $1 < Q^2 < 10 \text{ GeV}^2$, $7 < \nu < 23 \text{ GeV}$, $z > 0.2$ and $2 < p_h < 15 \text{ GeV}$. Results taken from [43].

1.2.6 Hadronization in nuclear environment: models

How to let quarks know that they should not take away color (and fractional electric charges by the way)?! [19].

Decades of work on understanding hadronization data have resulted in many sophisticated models. Due to its non-perturbative nature, hadronization cannot be described from first principles, instead, it relies heavily on gaining insight from the experimental data. While common ground, i.e. hadronization in vacuum, may be similar between the models, many differences arise once the nuclear medium is introduced. The caveat in connecting model predictions with data lies in the fact that experimental data can be equally and relatively well described by a good number of models making it virtually impossible to pinpoint the leading mech-

anisms. The multidimensionality of data provides a stronger ground to address those challenges. Theoretical efforts aimed at describing the space-time evolution of hadronization have been expanding in the past three decades since the first EMC data became available. The wealth of published theoretical work spans both the hot dense matter of heavy-ion collisions, as well as processes in cold nuclear matter. Some of the models are successful to address both processes, while others are limited in the validity of their assumptions. In this section we will briefly discuss the models that describe with a certain degree of success the DIS in cold nuclei.

Estimates of hadronization times

Though hadronization is a complex process, it is possible to obtain back-of-the-envelope estimates of the hadronization time scales based on general grounds. Assuming that the quark has absorbed all the energy of the virtual photon, its initial energy would be ν (neglecting the quark mass). Then, if the hadron has carried away energy $E_h = z / \nu$, the conservation of energy requires that string retains energy $(\nu - E_h)$, or equivalently $\nu(1 - z)$. In reality, due to the quark energy losses, associated with gluon bremsstrahlung, the retained energy will be less by a factor $\frac{dE}{dx} \approx k$. Here, k is the string tension such that $k \approx \text{GeV/fm}$ as obtained from string models. Therefore, an estimate of the distance over which the quark

propagates as quasifree is approximately:

$$\tau_p \approx \frac{1}{c} \frac{\nu(1-z)}{k} \quad (1.2.2)$$

For example, if a hadron emerged from a quark of initial energy ν , and carried away a fraction of its energy $z = 0.6$, the production time scale will be that of $\tau_p \approx 2 \text{ fm}/c$. Given the size of the carbon nuclei, $R_C = 2.7 \text{ fm}$, this estimate indicates that production phase takes place within the nuclei.

The regeneration time of the color field of a propagating quark can be obtained from classical considerations [19]. The charge moving along the z axis will be surrounded by a disk of Lorenz contracted EM field of radius R . In order for the field from a point charge to spread to a distance R , it requires time a $t \geq R$ in the reference frame of the charge. Translating it to the laboratory frame, this time will be slowed by a factor $\gamma = E/m$. Therefore, the field at distance R will appear not sooner than:

$$\tau_f \approx \frac{R}{c} \frac{E}{m} \quad (1.2.3)$$

where R is the interquark distance equal to the hadronic size, and m is the mass of its constituents, which varies from the minimum mass of the two bare quark, to the maximum mass of the fully formed hadron. These two parameters, R and m , are closely related when considering the light quark system: $m_q \sim \sqrt{\langle k_\perp^2 \rangle} \sim R^{-1}$, where $k_\perp$ is an average transverse momentum of the quark. The consequences of such classical considerations indicate that the bare quark following interaction will be able to hadronize only after time τ_f . The above estimate of τ_f can be equally

obtained from a quantum-mechanical approach. There, the regeneration time of a gluonic field confining the quarks can be viewed as the static field, which in turn is well understood in QED: $\tau_f \sim \frac{k_{\parallel}}{k_{\perp}^2}$. In the hadron rest frame, the components of this field are such that: $k_{\parallel} \sim k_{\perp}^2 \sim k \sim R^{-1}$. In the lab frame, they will be boosted: $k_{\perp}^2 = k = R^{-1}$ and $k_{\parallel} = \gamma k_{\parallel} = \frac{E}{m} R$. Therefore, the same expression for the formation time can be obtained from both quantum-mechanical and classical considerations. Considering a 4 GeV pion of radius 0.66 fm gives the formation time of $\tau_f \approx 20 \text{ fm}/c$.

Early string models

Modeling of hadronization began in 1980's with string models, which lay the foundation for a large number of sophisticated theoretical models as well as complex Monte Carlo generators emerging a decade later. One of the first string models was a one-step fragmentation model developed by Bialas [20]. It is based on the assumption that the nucleon is transparent to the propagating quark, and therefore the hadron is produced directly. The time scale of hadron production would then be simply $\tau \sim \nu$. The one-time scale model, further improved by Bialas and Chmaj (BC) [22], described the HERMES data well on nitrogen. The BC model introduced two probabilities: the probability that the quarks interacts with medium, given by $P_q(z) = \exp(-z/\tau)$, and the probability that prehadron or a hadron interact, given by $P_h(z) = (1 - P_q(z))$. The probability that the

propagating object does not interact with the nucleon depends on $P_q(z)$, $\sigma(q)$, $P_h(z)$, σ_h and nuclear density ρ_A . Using this technique, the HERMES analysis found that production time in a form $\tau = 1.4 \nu (1 - z)$ provides good agreement with the data, from which they found assuming $\sigma_q=0$ mb that $\sigma_h=25$ mb. Later, the model was expanded by Bialas and Gyulassy (BG) [21] and included the nuclear absorption factor $N_A(z, \nu)$, which represents the probability that neither prehadron nor the hadron interacted with a nucleon. Formation time of the hadron in BG model is computed as illustrated in Fig. 1.12.

Fig. 1.12: Formation time τ_f in one-time scale absorption model [21].

The effect of the nuclear medium on the string breaking process was considered in the context of "string-flip" model [26] and independent string fragmentation model [23] both based on the Lund model, described in the following section. In string-flip model the struck quark interacts with the nucleon via a color-exchange cross section of the order of the hadronic one. The mean free path of the quark is assumed to be larger than the size of the nucleus, thus the hadron does not suffer

absorption. This model has successfully described the EMC data for the range of energies $\nu > 20$ GeV. The independent string fragmentation model is built on the relative relevance of parton to prehadron (constituent) length l_c versus the prehadron to fully formed hadron (yo-yo) length l_y . The key point is that at large z , $l_c \propto (1 - z) \nu$, while $l_y \propto z \nu$ (neglecting elastic scattering). Model fits to EMC data indicate that in high energy interactions, the constituent length is the dominant mechanism. This suggests that the color strings behave as hadron-like objects with $\sigma_q \approx \sigma_{preh}$.

Lund string model

The Lund string model [24] is a successful non-perturbative model of hadronization which is based on the idea that in QCD for large charge separation, the color field lines are compressed in tubelike regions forming a string. The string, stretched between the struck q and $\bar{q}$, contains a constant amount of energy stored per unit length, which corresponds to a linearly rising potential $V(r) = kr$ (omitting Coulomb correction). The maximum length between the quarks is $L = 2 m_q / k$, where $2m_q$ is the mass of $q\bar{q}$ system. At this length, the string breaks symmetrically on the left and on the right, governed by the fragmentation function $f(z) \propto z^\alpha(1 - z)^\beta$ with the probability modeled by tunneling: $P = \exp \frac{\pi(p_\perp^2 + m_q^2)}{k}$. The Gaussian spectrum in $p_\perp$ reflects the Fermi motion; production of heavy quarks is suppressed; and diquarks are treated like an antiquark

Fig. 1.13: Schematic illustration of the Lund string model. The vertical axes represents time, and the horizontal - direction. The black points denote the breaking points of the string due $q\bar{q}$ production and formation of the pre-hadron. Hadronization proceeds through the evolution of the string pieces (blue and green). P_i are the formation points of the hadrons.

for baryon production. The fragmentation concept in the Lund string model is schemed in Fig. 1.13. Here, the prehadron formation point is denoted by black dot which breaks the string in smaller pieces. Hadronization is further modeled to proceed through the evolution of string pieces (blue and green). The hadron produced from the struck quark is denoted by a rank 1 (P_1), counted from the right of the figure. The Lund model consequently distinguishes two-step dynamical scales: constituent formation length l_c , defined as the length between the interaction and the formation point of the first quark of final hadron P_1 , and the yo-yo length l_y , defined as distance at which the hadron is formed. At large $z \rightarrow 1$, the string

breaks early, meaning that $l_c \rightarrow 0$, to leave all the energy to the hadron $E_h \rightarrow \nu$. At small $z \rightarrow 0$, the hadron is created at high rank after the string underwent multiple breaking. The production and the formation time scales are computable:

$$\begin{aligned} \langle \tau_p \rangle &= f(z)(1-z) \frac{z\nu}{k} \\ \langle \tau_h \rangle &= \langle \tau_p \rangle + \frac{z\nu}{k} \end{aligned} \quad (1.2.4)$$

For a $\nu=14$ GeV pion at HERMES, $\langle \tau_p \rangle \sim 5$ fm which is of the order of nuclear size, while $\langle \tau_h \rangle \sim 10$ fm.

Fig. 1.14: HERMES data on charged hadrons fitted with the model predictions based on PYTHIA and FRITOF, and BUU transport model. Good agreement with data is reached when hadron formation length $l_y = 0.5$ fm and the prehadron interaction cross section $\sigma_h = \frac{\sigma}{3}$. Adapted from [25].

The Lund string model provides a base for the PYTHIA/JETSET Monte Carlo even-by-event generator. It produces hadrons guided by the fragmentation pro-

cesses in vacuum. The GiBUU event generator [25] also relies on PYTHIA (Lund fragmentation) for the hard interaction and fragmentation, while treating the pre-hadron interactions using a semi-classical BUU (Boltzmann-Uehling-Uhlenbeck) transport model; unlike PYTHIA, it includes a number of nuclear effects (Fermi motion, Pauli blocking, nuclear shadowing). The BUU model approach treats final states with high accuracy using classical transport equations. The model gives good description of HERMES data as illustrated in Fig. 1.14 if production length $l_c=0$ fm and formation length $l_y = 0.5$ fm, with prehadron interaction cross section $\sigma_h = \frac{\sigma}{3}$.

Pure energy loss models

In this class of models the dominant mechanism of nuclear attenuation is attributed to quark energy losses, comprised of vacuum and medium induced gluon radiation. Typically, large pion formation time is used in order to justify assumptions that neglect interactions of forming color field with the medium. Cold QCD matter is an ideal testing ground to compare different energy loss formalisms since the properties of the medium (density and geometry) are well known. Yet, it is important to separate energy losses due to the initial-state effects, related to propagating quark, and final-state effects, relevant at hadron formation stage. For that reason, the Drell-Yan mechanism is particularly suited to access the quark energy losses because the lepton pair, created following $q\bar{q}$ annihilation, is not

subject to strong interactions with the medium, while propagating quarks are.

The quantity, which is both relevant in DIS and DY, is the transport coefficient $\hat{q} = \langle \mu \rangle / \lambda$, where $\langle \mu \rangle$ is an average momentum transfer to the quark in a collision, or $k_\perp$, and $\lambda = [\sigma\rho]^{-1}$ is the mean free path. Transport coefficient measures the 'scattering power' of the medium and thus is proportional to the number of scattering centers, i.e. density of the medium. It can be calculated perturbatively as prescribed by Baier *et.al* (BDMPS):

$$\hat{q} = \frac{4\pi^2\alpha_s C_R}{N_c^2 - 1} \rho x G(x, Q^2) \quad (1.2.5)$$

where G is the gluon density, $\rho=0.17 \text{ fm}^{-3}$ is the nuclear density, C_R is color charge which is equal to N_c for gluons and $(N_c^2-1)/2N_c$ for partons. In cold medium the transport coefficient is roughly estimated to be $\hat{q}^{cold} \approx 0.05 \text{ GeV}^2/\text{fm}$, while in hot medium at RHIC temperatures it is significantly larger: $\hat{q}^{hot} \approx 2.2 \text{ GeV}^2/\text{fm}$ [8]. A simple relation between transport coefficient and the quark energy loss was derived in [35]:

$$-\frac{dE}{dz} = \frac{\alpha_s N_c}{4} \hat{q} L \quad (1.2.6)$$

The transport coefficient is an important quantity that enters in the calculations of the nuclear dependencies of FF. The effect of the quark energy loss reduces quark energy at the moment of fragmentation, therefore modifying in-medium fragmentation function $D(z, Q^2, A)$ in a way that if the quark loses an energy ϵ , it leads to a shift in z : $z^* \rightarrow z/(1 - \epsilon/\nu)$. Based on the BDMPS formalism calculations, carried out by Arleo [37] suggest that in-medium multiple scattering

of produced quark may be responsible for the attenuation R_A with ν and z . Assuming that the transport coefficient $\hat{q} = 0.75 \text{ GeV}^2/\text{fm}$, the average energy loss would be $-dE/dz \approx 0.62 \text{ GeV}/\text{fm}$ in a large ($L \approx 5 \text{ fm}$) nucleus. This result is close to the calculations by Wang and Wang [27] which found the $dE/dz \approx 0.5 \text{ GeV}/\text{fm}$ for a 10 GeV quark in Au.

Fig. 1.15: Predicted nuclear modification of FF by Wang and Wang [27] compared to HERMES data.

The FF in Wang and Wang are based on pQCD calculations of leading twist and twist-4 contributions from double scattering processes. They contain only one free parameter, $\tilde{C}$. This parameter is related to the nuclear broadening of transverse momentum and can be obtained from fitting the data. The predicted shape of the z and ν dependence agrees well with the experimental data, as illustrated in Fig. 1.15, as well as with BDMPS model. A remarkable feature of predictions is the fact that nuclear modification of FF has a quadratic $\propto A^{2/3}$ nuclear size dependence. The calculations were further extended to heavy-ion

collisions by including expansion of hot nuclear medium, and demonstrated a qualitative agreement with π^0 data from PHENIX.

Rescaling models

Given the factorized form of DIS cross section (equation 1.1.10), the hard parton scattering, followed by the gluon radiation, and leading to the non-perturbative fragmentation, would be contained within the definition of fragmentation function. The vacuum FF has a perturbatively calculable dependence on the scale Q^2 , which is given by the DGLAP evolution equations 1.1.15. The nuclear FF require, however, a special treatment. The approach, developed by Majumder [28], generalizes the cascade process of partonic radiation, where each radiation leads to a drop in virtuality, by including a dependence on the distance scale which the parton has traveled in the medium. The calculations were carried out with consideration of one gluon emission as well as inclusion of multiple emissions. As may be seen from the comparison with data in Fig. 1.16, in case of neon, the single scattering and single gluon emission describe data adequately. Proceeding to the larger nuclei, such as xenon, this description worsens. The reason may lay in the fact that as one transits to the larger nuclei, the possibility of multiple scattering and multiple emission increases, as marked by the calculations which include multiple emission.

Fig. 1.16: Comparison of HERMES attenuation data with the calculations carried out by Majumder. The partonic energy loss was evaluated in single scattering and single gluon emission (red) as well as multiple emission (blue) scenarios. In addition to partonic energy loss, hadronic energy loss were incorporated in the model (green).

Rescaling model: energy loss and prehadron absorption

In this class of models, color neutralization typically takes place inside the nuclear medium, consequently, prehadron formation undergoes nuclear scattering. The model developed by Accardi *et.al.* [30] is a modern version of string-based absorption models which incorporates modification of fragmentation functions. The time-scales of prehadron and hadron formation are adapted from the Lund

string model, while the modification of FF is based on the deconfinement models. There, it is assumed that the deconfinement scale λ_A in nuclei is larger as compared to λ_0 on a free nucleon. The larger confinement scale translates in a smaller string tension, indicating that the hadronization in the nuclei starts earlier. In such a scenario, given that $Q^2 \sim 1/\lambda$, quark fragmentation functions are rescaled by factor $(\frac{\lambda_A}{\lambda_0})^2$. As the quark, confined on a scale λ_0 (or λ_A) carries momentum Q_0 (or Q_A), the running of the strong coupling must be accounted for in the construction of the overall rescaling factor $\xi_A(Q^2)$. For consistency, it is necessary that partial deconfinement in nuclei not only modifies PDF, but also FF. Based on the above considerations, the nuclear fragmentation function can then be written as:

$$D_f^{h|A}(z, Q^2) = D_f^h(z, \xi_A(Q^2)Q^2) \quad (1.2.7)$$

The effect of rescaling alone can explain high energy hadronization data produced by the EMC with an average energy transferred $\nu = 64$ GeV. Yet, it fails to reproduce lower energy data from HERMES where average $\nu = 12$ GeV, as illustrated in Fig. 1.17.

At HERMES kinematics, in particular for heavier nuclei, nuclear absorption becomes the dominant effect with a tendency to mask the rescaling. In order to incorporate nuclear absorption, rescaling of FF was considered in combination with Bialas-Chmaj(BC) or Bialas-Guylassy (BG) approaches which carry practical calculations of the nuclear absorption factor $N_A(z, \nu)$. There are two free parameters in those calculations: prehadron cross section and the production time. The BC

Fig. 1.17: Charged hadrons multiplicity ratios from the EMC on Cu (left) and HERMES data on Kr (right) fitted with model predictions by Accardi based on: pure rescaling (dashed), rescaling plus absorption according to Bialas-Chmaj (solid), rescaling plus absorption according to Bialas-Gyulassy (dotted). Adapted from [30].

predictions carried with prehadron cross section $\sigma_q = 0.5\sigma_q$ slightly overestimate but reproduce the correct shape of the data in the region $0.2 \leq z \leq 0.9$. The model allows for flavor separation, giving good agreement with data for the three pion states and charged kaons.

A step further was proposed by Accardi [36] to conjecture that the multiplicity ratios R_A should scale as a combination of ν and z : $R_A = R_A [\tau (\nu, z)]$. The scaling variable τ is related to production time in a form: $\tau = C z^\lambda (1 - z) \nu$, where λ is a scaling component extracted from fitting the data. The sign of λ suggests which of the mechanisms dominate: the absorption ($\lambda \geq 0$) or the energy loss ($\lambda \leq 0$). Based on fits to HERMES data on charged hadrons, the values for λ

with $C = 1.4 \text{ GeV/fm}$ were found to be $\lambda \geq 0.4$, which is a clear indication that hadronization process starts inside the nucleus on a scale of few fm.

Fig. 1.18: Model fits to HERMES data for charged hadrons: energy loss models (blue) [27] and the absorption models (red and green) [25]. Taken from [36]

Gluon bremsstrahlung (or color dipole) model

The gluon bremsstrahlung, or as often referred to - the color dipole model, developed by Kopeliovich *et.al* [32], is a semi-classical model which employs a probabilistic description of hadronization development while treating parton evolution quantum-mechanically. The model considers production of the leading hadrons, containing the struck quark, with $z \geq 0.5$. The mechanism of leading pion production in the Born approximation can be thought of such that the struck quark q_1 radiates a gluon which then splits into $\bar{q}_2 q_3$. Next, the struck quark recombines with an antiquark to form a colorless dipole $q_1 \bar{q}_2$ which afterward is then projected onto the pion wave function. In this model, hadronization is a two-scale process as illustrated in Fig. 1.6. The main contributor to the quark energy losses are the

vacuum energy losses which may have either (or both) nonperturbative origin due to string tension $\frac{dE}{dz} \approx 1 \text{ GeV/fm}$, or perturbative color field regeneration [33]:

$$\Delta E_{vac} = \frac{2\alpha_s}{3\pi} Q^2 \quad (1.2.8)$$

Up to a certain distance scale, the initial stage at which a quark radiates gluons has a constant rate of energy loss. Following it, the time development of energy losses is introduced by observing that the radiated gluon can be physically distinguished from the propagating quark only after a coherence time $t_c = \frac{2E\alpha(1-\alpha)}{k_T^2}$, where E the energy of the quark before it radiates, k_T (or as in earlier notation, $k_\perp$) and $\alpha = E_g/\nu$ are the transverse momentum and energy fraction taken by the emitted gluon [32]. There exists a limit on the transverse momentum $k_T > Q$ when gluons are considered a part of the quark. If the leading hadron is produced at large z and contains a struck quark, gluons could be only radiated with $\alpha < 1-z$. This results in the energy losses that have a $1/t$ dependence. This effect is called the Sudakov suppression. As a quark propagates in the medium, it experiences additional source of energy loss caused by the interactions with the medium. A quark, undergoing multiple collisions in the medium, increases its mean transverse momentum squared linearly with the path length L as it undergoes Brownian motion in the transverse plane. The induced energy loss could then be calculated as derived by Baier *et al.*: $\Delta E_{ind} = \frac{3}{8} \alpha_s \Delta < p_T^2 > L$. The total quark energy loss is then: $E = E_{vac} + E_{ind}$. The induced energy losses are, however, vanishingly small compared to those in a vacuum. While they lead to an additional

hadron suppression, their magnitude rules out the energy loss scenario as dominant mechanism for hadron attenuation, as observed in HERMES, and illustrated in Fig. 1.19. Meanwhile, the energy conservation imposes important restrictions on the color neutralization time:

$$l_p \leq \frac{\nu}{dE/dz}(1 - z) \quad (1.2.9)$$

which must vanish as $z \rightarrow 1$. Combined with equation 1.2.8, color neutralization time is then controlled by transferred energy to the quark, virtuality and fractional energy of the hadron:

$$l_p \propto \frac{\nu}{Q^2}(1 - z) \quad (1.2.10)$$

Physically, it is interpreted as the fact that a quark which is struck by a photon of high virtuality radiates more intensely, and therefore travels shorter distances and hadronizes sooner.

The model of nuclear absorption treats the space-time development of hadronization probabilistically, by computing the probability distribution $W(t_p, z, Q^2, \nu)$ that the prehadron is formed at time t_p following the hard interaction. Numerical results, presented in Fig. 1.21, indicate that at high z , production time goes to zero. Such behavior is expected based on the energy conservation: if the detected hadron carried away all the energy of the struck quark, then the prehadron must be formed immediately, otherwise it loses its energy by gluon bremsstrahlung. The formation time t_f , dilated in the laboratory frame, is similar to equation 1.2.3 and written as $t_f \propto E_h \cdot \Lambda_{QCD}$. The prehadron is considered a colorless dipole $q\bar{q}$

Fig. 1.19: Color-dipole model predictions for z -dependence of the multiplicity ratios on N and Kr for pions with inclusion of induced radiation (solid) and without (dashed). Taken from [32].

Fig. 1.20: Color-dipole model predictions for p_T^2 -dependence binned in z of the multiplicity ratios on N (left) and Kr(right). Adapted from [32].

formed with a transverse size $r_T \sim 1/Q^2$. As the dipole propagates through the medium, it fluctuates in size and, according to color transparency, attenuates with an absorptive cross section that is controlled by the dipole size. The propagation

Fig. 1.21: Probability distribution of the leading pre-hadron production time $t = t_p$ calculated on average HERMES kinematics. Adapted from [32].

of the dipole is described by summing up all possible paths of q and $\bar{q}$. Finally, the effective in-medium fragmentation function $D_A^h(z, Q^2, \nu)$ is calculated by convoluting the nuclear transparency, or the probability that the hadron is absorbed, with the probability distribution $W(t_p, z, Q^2, \nu)$. The dipole formalism, while bearing some caveats, describes well, in a parameter free way, data on the multiplicities and momentum broadening from HERMES. It could further be extended to the description of relativistic heavy ion collisions.

1.3 Connection to neutrino experiments

The discovery and study of neutrino oscillations in the late 1990's by Super-Kamiokande has renewed interest in neutrino-nucleus interactions. Accelerator based neutrino oscillation experiments operate in the medium-energy regime (sev-

eral GeV). At these energies, the prevalent scale of interaction is hadronic with an average interaction length of 1 fm. Due to the use of nuclear targets, employed both in production of neutrino beams and for enhancement of the neutrino detection rate, the role of nuclear effects becomes increasingly important in understanding the topology and the total energy of measured hadronic final states. To account for a wide range of nuclear effects, experiments rely heavily on MC simulations. The discrepancy between MC predictions and experimental data is known to contribute the largest sources of systematic errors, affecting the precision of the measurements of the neutrino mixing matrix which relates neutrino mass differences squared Δm^2 and the mixing angles of flavor eigenstates. Below we briefly review the sources of uncertainties related to beam profile and efficiency of neutrino detection.

Conventional accelerator-based neutrino beams are created by directing an intense proton beam onto a nuclear target (typically beryllium or carbon). The produced charged pions and kaons are focused by magnetic devices and directed into long beam lines where they decay into neutrinos. The uncertainties related to hadron beam focusing and geometry are small and readily calculable. However, the uncertainties on the hadron spectrum, produced when primary protons strike the nuclear target, are large, and result in the uncertainties on the neutrino flux. In a particular example, a MINOS experiment was set up to measure the neutrino Δm^2 and θ_{13} under the assumption that the observed deficit of ν_μ in the far detec-

tor (734 km) compared to the near detector (1 km) is caused by oscillations. The dominant contribution to the uncertainty in the neutrino flux, in which 87% of ν_μ are produced by π^+ decay and 13% by K^+ decay, is caused by the uncertainty in the yield of hadrons off the carbon target as a function of p_z and p_T (components of hadron momentum along and transverse to the beam-line). This affects the neutrino energy distribution at the near and far detectors since the focusing system cannot focus all transverse momenta equally. The only possibility is to constrain parameters on the functional prediction $f(p_z, p_T)$ of hadron yields by comparing MC and experimental data using the range of several proton beam energies. The quoted uncertainty on neutrino flux is $\pm 9\%$ for $E_\nu < 6$ GeV or $\pm 6\%$ above [47]. In another example from the T2K experiment, which aimed at measuring $\bar{\nu}_\mu \leftrightarrow \bar{\nu}_e$ oscillations, it is important to constrain electron neutrino backgrounds, which requires the knowledge of produced pion to kaon ratios. Part of the MC predictions on hadron yields were constrained using pion production data from NA61 experiment which used a T2K replica carbon target and 31 GeV/c proton beam. Outside the phase-space of the NA61 experiment, the production of hadrons was modeled resulting in systematic uncertainties on the production rate of pions of 50%, while that on kaons varied from 15 to 100%. This factor subsequently contributed 7.3/4.8% (with and w/o oscillations) in the overall uncertainty on the neutrino flux [49].

Neutrino-nucleus quasi-elastic (QE) scattering is the simplest process used

in neutrino identifications, influenced by both the target material and detector technology. In the bubble chamber era, experimental selection of ν_μ was robust based on the identification of three final-state tracks: muon, proton and spectator proton. Thus the selected QE events bore samples of impressive purity, reaching up to 99%. Modern era experiments are oriented towards neutrino oscillation searches where statistics are at a premium. This has driven experimental designs towards employment of heavy targets and mainly two types of detectors, tracking and Čerenkov, aimed at identification of one or two-track samples. While augmenting statistics by orders of magnitude, such methods produce a bias in the measurement due to the presence of background events misidentified as QE. In particular, this affects measurement of the incoming neutrino energy calculated by summing the energy of muons and that of hadronic final states.

Nuclear effects related to the rescattering inside the nuclear medium of hadrons produced from the initial neutrino interaction before they exit the target nucleus [50] are one of the largest sources of background. Those effects are associated with final-state interactions (FSI), and have significant impact on the visible energy of hadronic final states, accounting, for example, for part of the missing energy. The evaluation of systematic uncertainties in calorimetric response related to the uncertainty of the hadronic shower scale relies on the modeling of two processes: hadronization, which determines the set of particles produced from a particular DIS event using formation time estimates, and internuclear rescatter-

ing, which determines how the produced set of hadrons is altered by FSI while exiting the target nucleus. Reaction types, considered in those models, depend on the incoming neutrino energy and include: charge exchange, elastic and inelastic scattering, secondary pion production, and absorption [51]. Considering again an example from the MINOS experiment, the magnitude of the uncertainty in the simulation of neutrino production and detection in iron is 8.2% [47]. It is the second largest error on Δm^2 measurements. This error is composed of a number of uncertainties in the model assumptions. The first key assumption in the rescattering model is related to the modification of the free hadron cross section at low energies. Since this parameter is unknown, the modification of absorption cross section is carried out by increasing the nuclear size of the hadron. The second assumption is related to the amount of missing energy in the pion/nucleon absorption reaction dominated by the production of low momentum nucleons (usually 2-4) which do not register in the detector. To treat this issue, cascade processes, which accounts for missing energy in the reaction, is simulated. The uncertainty on the hadron formation time in the hadronization model is 50% and is currently under improvement by incorporating a more sophisticated model based on the recent data on hadron attenuation.

A recent work compared the MC predictions based on GENIE generator with experimental data obtained from JLab EG2 experiment showed a significant progress in tuning final-state interactions models [48]. Taking another example

from the T2K experiment, contribution to the uncertainty on the total number of predicted events in the Super-Kamiokande detector from the uncertainties due to FSI is smaller: 3.2%/5.9% (with and without oscillations).

While misreconstruction of neutrino energy due to FSI is a major source of background, another important source is the neutral current (NC) π^0 production relevant in $\nu_\mu \leftrightarrow \nu_e$ ($\bar{\nu}_\mu \leftrightarrow \bar{\nu}_e$) oscillations searches. In the particular example of the MiniBooNE experiment, running in antineutrino low-energy mode, neutral current π^0 's are produced via resonant ($\bar{\nu}N \rightarrow \bar{\nu}\Delta \rightarrow \bar{\nu}\pi^0 N$) or coherent ($\bar{\nu}A \rightarrow \bar{\nu}A\pi^0$) mechanisms. A π^0 , promptly decaying into two photons, can be identified in the MiniBooNE Čerenkov detector as two-electron rings. If only one track is resolved, this event can be misidentified as a single-electron ring, which is the event signature of the $\bar{\nu}_e$ (ν_e) charged current (CC) interaction (specifically here, $\nu_e + {}^{12}\text{C} \leftrightarrow e^- + \text{X}$) [52].

Events containing π^0 are the dominant physics background to the ν_e appearance signal at Super-Kamiokande. To achieve the required precision on the ν_e appearance measurement in the T2K experiment, observed via $\nu_e + n \rightarrow e^- + p$, the NC π^0 rate ($\nu_\mu + N \rightarrow \nu_\mu + N + \pi^0 + \text{X}$) must be measured. For that purpose, a dedicated π^0 detector (Pi-Zero detector), located at the same off-axis angle as Super-Kamiokande, was designed. It provides precise measurements of the neutral current processes on a water target using two electromagnetic calorimeters for detection of π^0 decay products. Hadronization models, that can predict the

probability of interacting quark to hadronize into a π^0 , which would then propagate through the nuclei and emerge as final state, are of potential importance to MC simulations of the above processes.

Neutrinos and electrons scatter directly from partons or nucleons throughout the nuclear target because of their weak interaction probability. At large invariant masses ($W > 2 \text{ GeV}/c^2$) hadronic systems produced in neutrino scattering are similar to those produced in charged lepton and hadron scattering. Thus, by combining results from the neutrino and electron scattering experiments, encompassing the weak (exchange of Z^0 , $W^\pm$) and electromagnetic interaction (exchange of γ), one has at hand various methods to extract information on the parton production and propagation. In order to reach the desired precision, neutrino experiments must minimize systematic errors correlated with nuclear models and FSI of hadrons. Experiments on electron scattering offer an opportunity to study those processes in an environment similar to neutrino-nucleus interactions. The viability of existing MC schemes used for predictions of background contamination and uncertainties on the flux in neutrino scattering experiments can be currently tested on available experimental data on electron-nucleus scattering. Neutrino experiments can therefore tune their MC from the improved understanding of parton dynamics and hadronization mechanism constrained by multidimensional data on electroproduction.

1.4 Motivation

In DIS, propagation of the parton is subject to a sequence of parton and hadro-production processes resulting in the observed hadron. The leptonic part of the interaction is described by QED (virtual photon emission, radiative corrections). The hard scattering and parton evolution can be accessed via perturbative QCD. The regime at which hadronization and formation of final state hadrons takes place cannot be described using perturbative methods, rather only by phenomenological models which rely on experimental guidance. The experiments at Fermilab, DESY, and Jefferson Lab through the Drell-Yan and SIDIS have uncovered new information on partonic processes in the *cold* nuclear medium. Experiments in RHIC have generated a strong interest in the same partonic processes in the *hot* medium, which are currently being investigated at the LHC.

The diagram on the left in Fig. 1.22, illustrates DIS where a virtual photon from the incident lepton is absorbed by the quark, which propagates over some distance, hadronizes inside or outside the nucleus, and emerges as a hadron. Inclusive DIS on nuclei is used to study initial state effects (EMC), while the semi-inclusive and exclusive reactions provide information on final state effects. The center diagram depicts the Drell-Yan process where a quark from an incident hadron annihilates with an antiquark from the target, forming a photon which eventually emerges as a dilepton pair. This pair carries information about the quark's passage through the nuclear medium. The diagram on the right illustrates a heavy-ion collision

Fig. 1.22: Quark propagation inside a target nucleus ("cold QCD matter") illustrated for the deep-inelastic scattering in lepton-nucleus reaction (left) and Drell-Yan process in hadron-nucleus collisions (center). Right: Hard scattered parton traveling through the "hot QCD matter" produced in nucleus-nucleus collisions. The diagram is taken from [8].

in which a scattered parton propagates over some distance within the hot dense medium, and hadronizes at a later times. Initial and final state effects are entangled in heavy-ion collisions, and enter mainly as background for the jet search algorithms. All three pictures contain the fundamental process of quarks interacting strongly with the medium via gluon exchange, and two contain information on hadron formation following the color singlet stage which fulfills requirements of confinement. To test hadronization mechanisms and color confinement dynamics, a precise knowledge on parton propagation and hadronization can be obtained from nuclear DIS and Drell-Yan reactions. To explicate the fundamental prop-

erties of QCD is the primary goal of the measurements in the cold medium, in particular in the space-time domain. The knowledge obtained from nDIS experiments can be used in Drell-Yan production to factor out final state effects from initial state effects. As a secondary benefit, the insights gained on *cold* nuclear matter would be beneficial to refine theoretical tools and understanding of relativistic heavy ion collisions. Much theoretical work is need to actually accomplish this, however, there are already attempts to describe both hot and cold matter within a unified picture. Finally, correction of nuclear effects represents a large source of systematical uncertainty for neutrino scattering experiments which use nuclear targets aimed at enhancing experimental rates.

Jefferson Lab experiments promise to uniquely measure space-time properties of hadronization inaccessible to any other experiments with sufficient luminosity and kinematic reach. A new wave of data will be delivered following the 12 GeV upgrade at Jefferson Lab from experiment E12-06-117 "Quark Propagation and Hadron Formation" [65]. An 11 GeV electron beam will probe a number of nuclear targets with three orders of magnitude more integrated luminosity than the HEMRES program. The dependency of observables and thus derived quantities (such as production and formation times, transport coefficients, in-medium cross sections, *etc.*) on mass, flavor, and number of valence quarks will be explored. The 12 GeV program will thoroughly investigate the low-energy phenomena associated with formation of light baryons and mesons, will provide an access to

the low rate events, such as, for instance, $\bar{p}$ and ϕ . The program of studies will expand further with proposed experiments at the the Electron Ion Collider [45] accessing unique energy range and good luminosity for eA reactions. In Jefferson lab design project - MEIC, the two collider rings can store up to 20 to 100 GeV protons or up to 40 GeV per nucleon for ions up to Pb, and 3 to 11 GeV electrons. This regime provides as access to higher ν values ($10 < \nu < 1600$ GeV), where one expects a quasifree parton to have a longer life time, and thus being produced outside the nucleus. Clear separation of partonic phase from the hadronic one allows to study pure partonic energy loss, which are not currently accessible in CLAS and HERMES. At small x_B , parton densities are expected to saturate. The scale of saturation is directly proportional to the transverse momentum broadening acquired at partonic stage. The direct measurement of the saturation scale and determination of quark energy loss are exiting topics important both for their fundamental nature, as well as for their relevance to explorations at high energies [46].

Chapter 2

Experiment

2.1 CEBAF

The Jefferson Lab Continuous Electron Beam Accelerator Facility (CEBAF) delivers GeV range electron beams in three experimental halls. One of the distinguished features of the CEBAF accelerator is its high duty factor ($\approx 100\%$). The accelerator, schematically illustrated in Fig. 2.1, is basically a linear accelerator built 8 m below the Earth's surface consisting of a pair of superconducting RF linear accelerators (linac) 1.4 km in length. The north and south linacs are connected to each other by two magnetic arc sections. The radius of the recirculating arcs is large enough so that synchrotron radiation by traveling electrons is negligible, therefore because electrons are ultrarelativistic traveling at the same speed, the multiple linacs along each side were combined into a single linac thus saving resources. From bottom to top the recirculating arcs on each side correspond to increasing electron energy, counting 9 arcs total this allows 5 complete passes. Each of the two linacs is comprised of 20 cryomodules accelerating the

electron beam up to 600 MeV, hence an electron beam making up five successive orbits will reach a maximum energy of nearly 6 GeV. Eight niobium superconducting cavities form one cryomodule, each of which is connected to its neighbor by a section containing beam vacuum pipe and magnetic elements, quadrupoles and steering dipoles to focus and guide the beam [56]. The liquid helium is used to cool the niobium approximately to 2 K, minimizing electrical resistance thus allowing the most efficient transfer of energy to an electron. Located in between the two linacs, liquid helium is produced in the Central Helium Liquifier (CHL). The multi-user beam distribution system has three key elements: the injector, the availability of separated beams of different energies inside recirculation lines, and the use of RF-separators. The electron beam is produced in the injector by a laser hitting a photocathode (electron gun) at a certain radiofrequency. The injector enclosure has been designed to accommodate two electron guns to provide either unpolarized or polarized electron beams (currently achieved maximum polarization is $\approx 88\%$). There is one laser per each hall which supplies electrons at a frequency of 499 MHz shifted by a phase. The electrons are accelerated in the cavities at a frequency of 1497 MHz ($2/3$ ns) per bucket with a bunch length of 0.3 ps [55]; divided between the three halls the beam is delivered every 2 ns into one of the experimental halls. The initial electron beam is bunched and then accelerated in two five-cell cavities, and then in the two full-sized cryomodules to the required energy of 45 MeV before injection into the north linac. At the end of each

Fig. 2.1: The Continuous Electron Beam Accelerator Facility (CEBAF). Electron beam starts at the injector, passes two linear accelerators and the bending recirculating arcs, then terminates at the experimental halls (Hall A, B, C).

linac the beam is sent vertically to a different recirculation arc using a separator which splits electron bunches depending on their momentum. The RF-separator deflects the beam, while the septum magnets amplify this separation and extract the beam from the machine. The beam recirculation is terminated at the end of south linac and transported to the experimental halls using the same separating principle. The beam energy can be of the five-pass energies, thus delivering a beam from the lowest operating energy up to the highest one. The construction of the beamline makes it possible to deliver simultaneously beams not only with different energies but importantly with very different currents. For example, detectors in

Hall B due to their large acceptance require low current on target (1-100 nA), while spectrometers in Hall A have small acceptance and require high current (up to 200 μA).

The CEBAF accelerator delivers beam currents sufficient to achieve luminosities of several times $10^{38}\text{cm}^{-2}\text{s}^{-1}$ to Halls A and C. The maximum luminosity achievable in Hall B is limited by drift chamber occupancies due to low energy Moller electrons (beam electron scattering on atomic electron); it reaches up to $1\cdot 10^{34}\text{cm}^{-2}\text{s}^{-1}$ for proton, and double that for a large nucleus.

2.2 CLAS

Hall B houses the CEBAF Large Acceptance Spectrometer (CLAS) which was originally designed to study excited states of the nucleons predicted by, for example, the constituent quark model. The current physics program of CLAS comprises diverse experiments to investigate spin structure of the nucleon, spectroscopy of excited states and transition form factors, as well as modification of meson propagation and short-range correlations in the nuclear medium. The construction of high duty cycle and high current accelerator allows to study reactions of interest with high statistical sensitivity. To achieve high detection efficiency for multi-particle final states a large acceptance detector is required. The CLAS apparatus [57] is an assembly of four types of detectors arranged in an onion-like pattern around the beamline and covering 3π with a diameter of almost 6

m. The design of CLAS is based on a toroidal magnetic field generated by six superconducting coils made of niobium titanium alloy (NbTi) kept at 4.5 K by a recirculating helium flow. The direction of the toroidal field points along ϕ such that the charged particles conserve their azimuthal angle along their trajectory, except near the coils, which lies in the plane containing the beam axis. The kidney shape of the magnet was designed to provide a strong field gradient for the forward going particles carrying high momentum and a lower field gradient for particles emitted at larger angles (Fig. 2.3). At the maximum current of 3860 A the integral magnetic field reaches 2.5 T·m in the forward direction, dropping to 0.6 T·m at a scattering angle of 90° . Another magnet used in CLAS is mini torus. Placed between the target and the first region of drift chambers, it reduces the background produced by scattered Moller electrons.

The six coils separate CLAS naturally into six independent tracking areas (or sectors). The particle leaving the target crosses (Fig. 2.2) three regions of drift chambers (DC) which provide a measurement of a charged particle trajectory in the toroidal field, Čerenkov counters (CC) provide identification and separation of particles carrying the same charge, scintillator counters which measure the time of flight (TOF), and finally the electromagnetic calorimeters (EC) measures energy and enables detection of neutral particles.

Fig. 2.2: A schematic top view of the CLAS detector cut along the beam line. Typical photon, electron, and proton trajectories (from top to bottom) from an interaction in the target are superimposed on the figure.

2.2.1 Drift chambers

The drift chambers (DC) [58] are the first detectors encountered by the particles as they scatter from the target. Covering polar angles $8^\circ < \theta < 142^\circ$, they were designed to determine the trajectory of charged particles of momentum above 200 MeV with polar angle resolution of 1 mrad and momentum resolution of 0.5%. The wire chamber relies on the detection of large a fraction of the charge as the charged particle transverses a volume filled with appropriate gas and produces

Fig. 2.3: Contours of constant absolute magnetic field for the CLAS toroid in the midplane between two coils.

electron-ion pairs along its path. The drift chambers are organized in three radial layers (Fig. 2.4), referred to as 'Regions', in each of the six sectors making it a total of 18 separate drift chambers. Region 1 (R1) is the innermost section of the drift chambers located closest to the target in an almost field-free volume. Region 2 (R2) is situated between the magnetic coils, and thus is characterized by the highest magnetic field which is used to determine the particle's curvature. Located outside the volume of magnetic field, Region 3 (R3) is aimed at global track reconstruction in connection with other CLAS detectors. Each DC region covers the same angular range, leading to an increase in size of each region as the distance from the target increases. The CLAS toroidal magnet bends charged particles towards or away from the beam axis depending on the particle charge and

orientation of the magnetic field. The reconstruction of the particle's momentum p of charge q is based on the direction and curvature r of the track and the strength of the magnetic field B . Each region of drift chambers consists of two superlayers of wires. The first layer comprises axial wires, strung approximately parallel to the direction of the magnetic field. The second is a layer of stereo wires, tilted at a 6° angle with respect to the axial wires. Such a superposition of wires provides an increased pattern recognition of the track in the azimuthal direction.

Fig. 2.4: Hexagonal cells of drift chambers with a typical track indicated by shaded areas (left). Vertical cut of drift chambers indicates geometry of the regions (right).

Each super layer consists of six layers of hexagonal wire cells (Fig. 2.4) (except for Region 1 which has only four layers of stereo wires due to space limitation). The hexagon represents a drift cell where the sense wire, where a signal is detected,

is positioned in the middle, and the field wires are located in the vertexes. A high voltage system maintains the sense wires at positive potential, while the field wires are maintained at a negative potential 50% lower than the positive value. The high difference of potentials creates an avalanche of the electrons induced by the ionizing particle. The hexagonal shape of the cell mimics a circular geometry cell in which the drift time to drift distance is independent of entrance angle.

The volume of each region of drift chamber is filled with a high purity gas mixture composed of 90% argon and 10% CO₂. Such a choice provides high drift velocity (0.04 m/ μ sec) and fast collection time which in turn improves momentum resolution. The tracking efficiency does not exceed 95% for the chamber hit occupancies up to 4%.

2.2.2 Čerenkov counters

The primary purpose of Čerenkov counters (CC) [59] is to differentiate between electrons and negative pions with momenta below 2.5 GeV after they have passed through the drift chambers. The CLAS detector is equipped with a CC in the forward region covering polar angles up to 45° in each sector.

When the charged particle traverses a medium with the velocity greater than the speed of light in that medium ($v > c/n$, where n is refractive index), it emits electromagnetic radiation known as Čerenkov radiation. The active volume of CC is filled with C₄F₁₀ gas, which was chosen for its high index of refraction

Fig. 2.5: Diagram of one optical module of the CLAS Čerenkov counter illustrating the Čerenkov light from the electrons reflected twice from the mirrors into a Winstone cone which collects light on the surface of PMT.

($n = 1.00153$) producing a high yield of photons. Each sector of the detector holds approximately six cubic meters of gas. The electrons start to emit Čerenkov radiation at energies of ≈ 0.9 GeV while the threshold for pions is above ≈ 2.5 GeV making π/e separation less efficient in this region. The light-collecting optics consists of 216 modules: each sector of CC is divided into two sub-sectors each containing 18 modules. The optical element of CC module comprises an assembly of one elliptical and one hyperbolic mirror providing primary light focusing into a light collecting cone, a cylindrical mirror used to compensate for imperfections in the focusing, and a photomultiplier used to count the number of photons in the light cone. To prevent the degradation of energy resolution, light-collecting cones and the PMT's are placed in the regions of ϕ already obscured by the torus coils.

The single-photoelectron response of the photomultiplier is used to equalize the gain and preserve detection sensitivity. The path of a typical electron through the CC and its light collection in one module is shown in Fig. 2.5.

2.2.3 Time-of-flight Counters

The time-of-flight system (TOF), or as it is called otherwise scintillator counters (SC), surrounds the drift chambers of CLAS covering the same polar angles $8^\circ < \theta < 142^\circ$. It is located radially outside the tracking system and Čerenkov counters in front of the electromagnetic calorimeters. Each TOF counter is made of organic plastic scintillator (Bicron BC-408) with a PMT at each end [60]. The TOF system was designed to measure the velocity of charged hadrons, namely, π , K and p , utilizing the ionization principle. A charged particle traversing the matter leaves behind a number of excited molecules which release a fraction of their energy as optical photons. A high light yield and fast response time allow the possibility of subnanosecond timing resolution. The time of flight of the particles is reconstructed by taking the difference between the event start time at the target (RF time) and the time measured by the scintillators, t_{SC} . The particle velocity is then $\beta = l_{SC}/(t_{SC} \cdot c)$, where l_{SC} is the length of the trajectory, t_{SC} is the time of flight and $c = 29.97 \text{ cm/nsec}$ is the speed of light. With momentum defined from the drift chambers, the particle mass can be reconstructed as: $m = p\sqrt{(1 - \beta^2)}/\beta$. Each scintillator plane is positioned such that it is perpendicular to the average

local particle trajectory (Fig. 2.6). The scintillator thickness is 5.08 cm, the width of the forward counters is 15 cm and the large-angle counters are 22 cm wide. The timing resolution for a single counter varies with the length and the width of the strip from about 60 ps for the shorter scintillator paddles to up to 120 ps for the longer paddles. The system is capable of separating charged pions and kaons up to 2 GeV/c.

Fig. 2.6: Schematic view of the scintillator strips for one sector with respect to the beam direction.

2.2.4 Electromagnetic calorimeter

The sampling electromagnetic calorimeter (EC) consists of an active medium (scintillator strips) which produces signal output and a passive medium which functions as an absorber aimed at inducing a shower (lead sheets). High energy electrons (and positrons) lose energy in the EC dominantly by bremsstrahlung, while photons of the GeV energies lose their energy by $e^+ e^-$ pair production. The

total light collected is proportional to the initial energy of the electron (incoming photon). Heavier particles, such as charged pions, behave like minimum ionizing particles depositing small constant fraction of energy in the scintillator; they also lose their energy through a showering process, however, instead of bremsstrahlung, the fundamental process is nuclear interactions. The difference in the mechanisms of energy deposition helps to separate electrons and charged pions. The CLAS EC serves three main goals: detection and triggering of electrons of energies above 500 MeV, detection of photons of energies above 200 MeV (leading to reconstruction of π^0 and η via the measurement of their 2γ decay) and detection of neutrons. Due to EC detection efficiencies, the discrimination between photons and neutrons for momenta up to 2.5 GeV is done using time-of-flight measurements. Meanwhile, the separation of π/e is optimal at higher momenta ≈ 2.5 GeV where the pion rejection reaches its threshold in CC.

The electromagnetic calorimeter is installed in the forward region and covers the same angles $8^\circ < \theta < 45^\circ$ as those of the Čerenkov counters. The detector is constructed from alternating layers of scintillators sandwiched between lead sheets with a total thickness equal to 16 radiation lengths. A lead to scintillator ratio of 0.24 was used [57]. In such a configuration roughly 1/3 of the total energy of showering particles is deposited in the scintillator. Each EC module is a sandwich of 39 layers of 10 mm thick plastic scintillator and 2.2 mm thick lead sheets (Fig. 2.7). The modules have a shape of approximately equilateral

triangles. Each layer is composed of 36 scintillator strips parallel to one of the sides of triangle such that the orientation of the strips is rotated by approximately 120° in each successive layer. Thus, there are three views, labeled U, V, and W, each containing 13 layers which are further subdivided into inner and outer stacks. Such a configuration provides stereo information on the hit location; the time and the energy are then calculated by accounting for the path length from the hit to the readout edge.

Fig. 2.7: Schematic view of one of the six CLAS electromagnetic calorimeter modules.

The energy resolution of the EC is $\frac{\sigma}{E} < \frac{10\%}{\sqrt{E(\text{GeV})}}$, and a position resolution $\delta r \approx 2.3 \text{ cm}$ at 1 GeV [70]. Efficiency of neutron detection reaches up to 60 % above 1.6 GeV.

2.3 EG2 experiment

The studies on hadron propagation E04-104 [62] and the search for color transparency E02-110 [63] were run in parallel during the EG2 experiment. The experiment was divided into three runs, labeled a, b, c depending on the incoming electron energy: 4.0 GeV, 4.5 GeV and 5.014 GeV, respectively. The analysis presented in this thesis was performed on the third run, EG2c, providing the largest statistics amongst the three. Commissioning of the experiment took place in the summer of 2003, and ran through March of 2004 (with interruption due to hurricane Isabel). Since both of the proposed experiments use normalization to a deuterium target, the complete target system [64] was designed to have two targets exposed simultaneously to the incoming beam: the first target exposed was the deuterium target, referred further to as the liquid target, and the target was the solid target (Fig. 2.8). The latter can be chosen to be a carbon (C), aluminum (Al), iron (Fe), tin (Tn), or a lead (Pb) target. The largest dataset was accumulated for C, Fe and Pb targets from which pion production will be analyzed. The advantage of having a double target system is that several systematic effects related to beam and detector properties will cancel in the nuclear ratio. While one of the solid targets was exposed to the beam, the other solid targets were held out using a special system of 'fingers' positioned in the shadow of the CLAS coils. The targets were interchanged remotely via an EPICS (Experimental Physics and Industrial Control System) interface. The length of the liquid target was 2 cm, while the

solid targets had a form of a circular disk with a radius of 0.15 cm and variable in thickness: 0.014 cm, 0.04 cm and 0.17 cm for Pb, Fe and C correspondingly. The dimensions of the targets were chosen to have approximately the same number of nuclei along the target length. The separation distance between the two targets is 4 cm. The solid target was placed at $z=-25$ cm away from the CLAS center ($z=0$), while the center of the liquid target was at $z=-30$ cm. Positioning the target system backwards with respect to CLAS center was motivated by increasing acceptance for the negatively charged particles: the orientation of CLAS torus polarity was such that negative particles were inbending, thus moving the target away from the center avoiding the forward angles already limited by detector acceptance. The double target system with all the support structures was implemented to GSIM - a GEANT simulation package for the CLAS detector.

2.4 Triggering and data acquisition

To record the events of interest and minimize the dead time of detector components, a double layer triggering system of CLAS was developed. The Level 1 trigger is a fast response system which processes all designated raw signals to a preliminary definition of a possible event candidate. In the case of electroproduction runs, the Level 1 trigger selects events with a possible electron using information from the pre-trigger boards of the EC, CC, and the TOF to form a coincidence signal. The Level 1 trigger ensures that sufficient energy was deposited in the EC,

Fig. 2.8: The double target system from the EG2 run showing one solid target (in blue square) inserted in the beam line, and five targets retracted. The solid targets are held by mechanical arms allowing to change them remotely. The liquid (in red square) is in the back enveloped in thermally insulated aluminum foils.

and that the signals in the EC, CC and TOF are in coincidence. The information for this event candidate is then digitized to provide a readout for to the Level 2 trigger. The Level 2 trigger reads information from the DC to find possible track candidates. If no track candidates are found, the Level 2 trigger issues a clear signal, and more triggers are accepted. When a track candidate is found, the event is sent to the Event Builder (EB), which forms a complete event by putting together responses from different parts of detector in the form of individual tables which are prefixed by headers to form 'banks'. The Event Recorder (ER) picks up data for permanent storage from the Event Transfer (ET) system which uses

shared memory of CLAS computers. First the data are transferred to the redundant array of disks (RAID). From RAID it is then transferred to magnetic tape (silo) for permanent storage [57].

2.5 Data Processing

The DAQ (data acquisition system) captures the raw data from every subsystem of the CLAS detector. It continuously operates at event rates between 3 and 4 kHz. The data collected during experimental run are saved in separate files in RAW format, and contain roughly 10 M events corresponding to a file size of 2 GB. For the purpose of analysis the RAW format should be further processed by the offline reconstruction system RECSIS. This process is referred to as 'cooking'. The 'cooking' procedure is divided into three stages. At the first stage the calibration of the signals from all detector subsystems is carried out. The responses from detector electronics are comprised of ADCs (analogous-to-digital converter) which is used to determine the energy deposition in each detector, and TDCs (time-to-digital converters) which provide timing information for the hits. In the second stage of 'cooking', detector stand alone packages use calibration constants to convert ADC and TDC signals into the hits and record this information in the BOS format. The final stage the processes above BOS file through the event reconstruction software (RECSIS) which provides information about detector response directly in terms of variables characterizing the event - tracking, particle

charge, momentum, hit coordinates and timing. The final processed data is stored in various formats including BOS banks, PAW or ROOT ntuples.

The 'cooking' and calibration procedures were performed in 2005 by Lamiaa El Fassi and Lorenzo Zana. However, later in 2009 the EG2 dataset was completely recooked by the author of this thesis. Since 2005 the CLAS analysis software has undergone changes both in the tracking procedure as well as in the reconstruction of single detector outputs. These changes lead to the following statistical improvements: $\approx 44\%$ more electrons with refined ID (described in the next chapter) and $\approx 86\%$ more of two photon pairs were reconstructed compared to the number of electrons and photon pairs reconstructed from 2005 cooking. The consequence of recooking the full data set is almost doubled statistics of neutral pions.

Chapter 3

Data analysis

Below we describe our identification scheme for each of the particles as determined using standard CLAS instrumentation, followed by the discussion of the binning scheme employed for analysis.

3.1 Particle ID scheme for electrons

During the data “cooking” process SEB (Simple Event Builder) selects possible electron candidates based on the sign of the charge (from the DC magnetic field) such that the track matches in position and time with both the CC hit and isolated energy cluster in the EC. SEB marks those candidates with $\text{id} = 11$ and stores this information in the EVNT bank. The resulting particle ID is, however, very approximate. What follows further illustrates the refined criteria of the electron selection. The electron candidate, satisfying a “time-based” tracking criteria at the level of DC is defined by a number of cuts, each of which is discussed below.

3.1.1 Momentum

The electromagnetic calorimeter is usually set up to trigger on electrons at energies above 0.5 GeV. However, a low momentum cut P_{min} depends on the EC low total threshold of the trigger discriminator [?]:

$$P_{min}(\text{MeV}) = 214 + 2.47 \cdot EC_{threshold}(\text{mV}) \quad (3.1.1)$$

The low threshold for the EG2 experiment was 172 mV. Therefore, a minimum momentum cut $P_{min}=0.64$ GeV has to be applied in order to avoid inefficiencies for the deposited energies close to the EC threshold.

3.1.2 Čerenkov signal

The CC detector triggers electrons and separates them from low momentum negative pions based on the magnitude of the output signal which is defined as the number of photoelectrons (Nphe). The distribution of $10 \cdot \text{Nphe}$, where the factor of 10 is employed for the mere convenience of reading, is illustrated in Fig.3.1.

The distribution of electrons in each single PMT represents a Poisson distribution centered around 8-10 photoelectrons. The peak located at 1-2 photoelectrons corresponds to low energy pions that produce a knock-out electron in the Čerenkov gas (δ -rays). Hence, an effective cut is necessary in order to decrease negative pion contamination.

Fig. 3.1: $10 \cdot \text{Nphe}$ distribution for all electron candidates with negative charge and time-based tracking.

Fig. 3.2: θ_p vs mirror number distributions. The plot on the left contains electron candidates that satisfy all the electron ID cuts but the Nphe and CC match (left), while the plot on the right has the same electron sample but after the CC matching cut is applied.

For that purpose, we define a CC projectile plane which represents an imaginary plane behind the CC where the Čerenkov radiation would arrive if the track is extrapolated straight to the PMT before it was reflected [67]. The precision for hit determination from the hit segment number is significantly higher in the projected variables (θ_p, ϕ_p) as compared to the θ, ϕ determined from the DC. The distribution of the θ_p as a function of mirror number is shown in Fig. 3.2. Mismatch of the mirror number and θ_p corresponds to the pion contamination, when the knocked out electron gives a signal in a neighboring mirror as compared to the pion initial direction θ_p [68]. Excess of events at mirror number corresponding to the multiples of 10 are those events that have a 'bad' status, since no cut on the CC status word is applied.

The distribution of θ_p vs mirror number has been fitted with a second order polynomial, and the matching area has been selected such that:

$$|\theta_p - (a + b * MirrorNumber + c * MirrorNumber^2)| < 2.5 \cdot \sigma \quad (3.1.2)$$

$$MirrorNumber = 10 * mirrorN + CCstatus \quad (3.1.3)$$

where $0 \leq CCstatus \leq 10$, $0 \leq mirrorN \leq 18$, and coefficients $a = 7.30$, $b = 0.14$, $c = 3.7 \cdot 10^{-4}$. The CC matching cut is the only CC cut used in this analysis. Nevertheless, there exists another alternative approach that can replace or even be combined with the CC matching cut. A threshold cut $10 \cdot N_{phe}$ may be applied

Fig. 3.3: 10·Nphe distributions in each sector before the CC matching cut (blue) is applied and after (red). Vertical lines correspond to the 10·Nphe cut for each sector as defined in equation 3.1.4.

to the CC signal, however, this cut must be sector dependent:

$$Nphe > 25 \text{ for sector} = 1,2 \quad (3.1.4)$$

$$Nphe > 26 \text{ for sector} = 3$$

$$Nphe > 21 \text{ for sector} = 4$$

$$Nphe > 28 \text{ for sector} = 5,6$$

Illustrated in Fig. 3.3 is the 10·Nphe distribution together with Nphe cut in each sector before the CC matching cut (red) is applied and after (blue), providing the

electron satisfies all its identification cuts. The distribution in blue corresponds to the clean sample of electrons; evidently, it discriminates from pion contamination which peaks at the low count of $10 \cdot N_{phe}$ according to equation 3.1.4. The reason the $10 \cdot N_{phe}$ cut was not applied in addition to the CC matching cut is due to the fact that it cuts out good electrons (3%), leaving, however, the level of negative pion contamination essentially unchanged.

3.1.3 Sampling fraction

The energy deposited by electrons in the active area of the EC is a fraction of their total energy E_{tot} , that is proportional to their momentum P for energies above few hundred MeV. In order to distinguish electrons from high momentum π^- that loses amount of the energy per scintillator independent of its momentum, a momentum dependent sampling fraction cut (E_{tot}/P) must be applied. Fig. 3.4 demonstrates the total energy deposited in the electromagnetic calorimeter (E_{tot}/P) versus P for all electron candidates.

Ideally, the ratio of total energy deposited and momentum, measured from DC tracking information, should be a constant equal to 0.271 (based on GSIM simulations 4.2.1). Yet, due to the electromagnetic calorimeter sampling of the shower and attenuation of the light collected, this slope is not constant and varies with sector number. The electron identification cut is calculated by fitting the mean value $\mu(p)$ and the width $\sigma(p)$ of the E_{tot}/P distribution in each bin of P

Fig. 3.4: E_{tot}/P vs $\sigma(p)$ distribution for all electron candidates for which there is a signal in the DC, SC, CC and EC.

with a second order polynomial. The resulting functions, $\mu(p)$ and $\sigma(p)$, let us construct a cut such that:

$$|E_{tot}/P - \mu(p)| < 2.5 \cdot \sigma(p) \quad (3.1.5)$$

$$\mu(p) = a + b \cdot p + c \cdot p^2 \quad (3.1.6)$$

$$\sigma(p) = \sqrt{d^2 + \frac{f^2}{\sqrt{p}}} \quad (3.1.7)$$

where a , b , c , d , and f are constants, depending on the sector number and target type. The distribution of the sampling fraction is the same for the solid target and its deuterium, thus, we combined the two into deriving the above dependence. The constants are indicated in Table. 3.1, 3.2, and 3.3 respectively for carbon, iron, and lead target set up together with their deuterium.

	sector 1	sector 2	sector 3	sector 4	sector 5	sector 6
a	2.52e-01	2.78e-01	2.62e-01	2.51e-01	2.63e-01	2.55e-01
b	1.22e-02	1.87e-02	2.31e-02	2.01e-02	9.55e-02	2.32e-02
c	-7.94e-03	-2.38e-03	-3.54e-03	-3.32e-03	-1.02e-03	-3.05e-03
d	9.55e-03	1.399e-02	9.32e-03	8.21e-03	2.25e-02	1.17e-02
f	3.41e-02	3.75e-02	2.90e-02	2.99e-02	3.06e-02	3.64e-02

Table 3.1: Carbon target coefficients for the sampling fraction.

	sector 1	sector 2	sector 3	sector 4	sector 5	sector 6
a	2.22e-01	2.34e-01	2.52e-01	2.51e-01	2.72e-01	2.52e-01
b	2.23e-02	1.95e-02	2.42e-02	2.08e-02	1.18e-02	2.28e-02
c	-2.41e-03	-2.08e-03	-3.39e-03	-3.27e-03	-1.87e-03	-3.11e-03
d	9.23e-03	8.66e-03	1.08e-02	7.22e-03	1.84e-02	4.11e-03
f	2.98e-02	3.09e-02	2.64e-02	2.98e-02	3.48e-02	3.55e-02

Table 3.2: Iron target coefficients for the sampling fraction.

	sector 1	sector 2	sector 3	sector 4	sector 5	sector 6
a	2.53e-01	2.49e-01	2.54e-01	2.55e-01	2.76e-01	2.62e-01
b	1.38e-02	1.47e-02	2.26e-02	1.90e-02	1.11e-02	1.92e-02
c	-1.40e-03	-1.49e-03	-3.05e-03	-3.05e-03	-1.76e-03	-2.62e-03
d	7.67e-03	7.53e-03	8.13e-03	7.20e-03	1.81e-02	1.99e-03
f	3.54e-02	3.38e-02	2.77e-02	3.04e-02	3.53e-02	3.76e-02

Table 3.3: Lead target coefficients for the sampling fraction.

Fig. 3.5: Carbon target set up E_{tot}/P vs $\sigma(p)$ distribution for all electron candidates for each CLAS sector. Black solid line is the second order polynomial fit to the mean value $\mu(p)$ of E_{tot}/P in each momentum bin. Red lines are the cuts on the sampling fraction corresponding to $\pm 2.5\sigma(p)$.

Fig. 3.5-3.6 illustrates E_{tot}/P versus P distribution for electrons that satisfy all the ID cuts except the deep-inelastic kinematics and sampling fraction, for the example of carbon and iron targets together with their deuterium. While in certain sectors the fit to the mean value of sampling fraction exhibits different behavior such as the decrease of the sampling fraction at high electron momentum. It should be noted that this behavior does not affect the electrons of interest. After imposing DIS cuts along with the requirement $y < 0.85$, the effective range of electron momenta becomes: $0.75 < P < 2.8$ GeV. Within this range the sampling fraction fit in all the sectors and for all the targets is not decreasing with the increase of the electron momentum.

Fig. 3.6: Iron target E_{tot}/P vs $\sigma(p)$ distribution for all electron candidates for each CLAS sector. Black solid line is the second order polynomial fit to the mean value $\mu(p)$ of E_{tot}/P in each momentum bin. Red lines are the cuts on the sampling fraction corresponding to $\pm 2.5\sigma(p)$.

3.1.4 Inner energy

Each strip of the calorimeter is made out of 13 layers. The total collected energy per strip is composed of two parts: inner energy (E_{in}), which is deposited in the first 5 super layers, made of 15 cm thick scintillators and 2.3 mm thick lead shifts; and outer energy(E_{out}), deposited in the last 8 superlayers, made of 24 cm thick scintillator and lead shifts. Electrons below 500 GeV are known to deposit more energy into the inner stack of the calorimeter compared to the outer stack, however, at higher electron energies it is the reverse. Negative pions are minimum ionizing particles that deposit a constant amount of energy in each super layer,

equal to $\sim 2 \text{ MeV/g}\cdot\text{cm}^2$. This amounts to $E_{in} \simeq 30 \text{ (MeV)}$ and $E_{out} \simeq 50 \text{ (MeV)}$. The long tail signal in the outer stack is due to strong interaction of negative pions which produces forward going hadronic debris. This can be seen in Fig. 3.7 where E_{in} is plotted versus E_{out} . In order to suppress π^- contamination, E_{out} must be positively defined, and a cut on E_{in} is imposed, such that:

$$E_{in} > 60 \text{ (MeV)} \quad (3.1.8)$$

Fig. 3.7: E_{out} vs E_{in} distributions. Left - electron candidates satisfying negative charge and hit-based tracking; right - electrons satisfying all the identification cuts but E_{in} . Red dotted lines correspond to the cut off $E_{in} > 60 \text{ MeV}$.

3.1.5 EC and SC coincidence time

The scintillators are located radially outside the tracking system but in front of the calorimeter. An electron gives a signal in the EC, following the SC. If an

accidental particle was misidentified as an electron, it may produce a signal in either the EC or the SC. Those events can be suppressed by applying scintillator and calorimeter coincidence time cut, as illustrated in Fig. 3.8. The EC, SC

Fig. 3.8: Distribution of $t_{EC} - (t_{SC} - \frac{l_{EC}-l_{SC}}{c})$ for: a) electron candidates satisfying negative charge and hit-based tracking (left); b) electrons satisfying all electron ID cuts except the EC and the SC coincidence time cut (right). Distribution inside red dotted lines corresponds to the remainder after the cut.

coincidence time cut is of the form:

$$|t_{EC} - (t_{SC} - \frac{l_{EC} - l_{SC}}{c})| < 5 * \sigma \text{ (ns)} \quad (3.1.9)$$

where $\frac{l_{EC}-l_{SC}}{c} \approx 0.7$ ns and $\sigma = 0.35$ ns is the width of the Gaussian fit to the distribution. This cut is based on the sample where electrons satisfy all the identification cuts except the coincidence time cut.

3.1.6 EC fiducial cut

The coordinate perpendicular to the borders of EC are traditionally denoted as U, V, W. When a particle interacts at the edges of the calorimeter, the electromagnetic shower is not fully contained in the detection volume. Hence, the fiducial region (uniform efficiency) of the calorimeter must be defined. It excludes events that do not satisfy the following geometric cuts:

$$40 < U < 400, V < 360, W < 390 \text{ cm} \quad (3.1.10)$$

The distributions of U, V, W are plotted in Fig. 3.9. The distribution of X, Y of the calorimeter along with the imposed cuts are plotted in Fig. 3.10.

Fig. 3.9: EC fiducial cuts. U, V, W distribution for the electrons containing all the ID cuts except the EC fiducial.

Fig. 3.10: Y vs X distribution before (black) and after (color) imposed U, V, W cuts
for: left - all electron candidates, right - electrons satisfying all the ID cuts
except the EC fiducial one.

3.1.7 DC fiducial cuts

The CLAS toroidal magnetic field bends charged particles toward or away from the beam axis, changing the polar angle θ , however leaving the azimuthal angle ϕ essentially unchanged. A procedure has been developed to determine DC regions with low efficiencies depending on the type of the particle as a function of (P, θ, ϕ) . For the purpose of defining flat acceptance regions, the whole data set was divided into small bins: 0.1 GeV momentum bins, 1° wide in θ and 6 bins, corresponding to six sectors, in ϕ [69]. The effect of these cuts is illustrated in Fig.3.11.

Fig. 3.11: ϕ vs θ distribution before (color) and after (black) applying DC fiducial cuts: left - all electron candidates, right - electrons satisfying all the ID cuts except the DC fiducial one.

3.1.8 Vertex correction

During the EG2 running period, two targets were in the beam simultaneously: 2 cm liquid deuterium at $z=-30$ cm and a solid target at $z=-25$ cm, where $z=0$ corresponds to the CLAS center. The tracking reconstruction of vertex position assumes that the beam position is at $(x,y)=(0,0)$. The vertex is then defined by the intersection of the track with the plane containing the ideal beam position and perpendicular to the sector mid-plane. Therefore, the vertex reconstruction should be corrected by taking into account the real beam position offset. The beam offset (x,y) was determined using the proton elastic scattering reaction and was measured to be: $x=-0.043$ cm and $y=0.33$ cm. Due to this offset the distance between the perpendicular to the sector mid-plane going through the CLAS center

Fig. 3.12: Electron vertex distribution on the example of deuterium and carbon target system plotted for each sector: left plot is regular z vertex distributions, right plot corresponds to z vertex corrected for the beam offset. The small peak in between the targets corresponds to the aluminum foil.

and the mid-plane containing the real beam is different for each sector. Hence, the vertex position of the reconstructed track is measured as sector dependent. In order to eliminate this dependence, the track was extrapolated linearly along the momentum to the plane containing the real beam position. The resulting z vertex correction as a function of ϕ angle is presented on the Fig. 3.12 and Fig. 3.14. The z position after correction is aligned and independent of the sector number. A cut on the corrected values of the z vertex position has been applied for each target C, Fe, Pb, and D as follows:

$$-25.33 < z^C < -24.10 \text{ (cm)} \quad (3.1.11)$$

$$-25.65 < z^{Fe} < -24.26 \text{ (cm)}$$

$$-25.54 < z^{Pb} < -24.36 \text{ (cm)}$$

$$-31.80 < z^D < -28.40 \text{ (cm)}$$

The values of the vertex cut for the solid targets were chosen from a Gaussian fit in the range of $\langle z \rangle \pm 3 \cdot \sigma$ (Fig.3.13).

Fig. 3.13: z vertex distribution for: D+C, D+Fe, and D+Pb. Dashed lines correspond to the vertex cut applied to D (similar for all three cases), and the solid lines illustrate vertex cuts applied to each solid target.

The reconstruction of the y position of the target is usually worse than that of z due to the positioning of the stereo wires. In order to decrease the background under each of the targets, a cut on the corrected value of the y , identical for solid

Fig. 3.14: z vs ϕ for electrons. Left plot illustrates distributions before beam offset correction of the vertex, right plot - after. The red dashed lines indicate the cutoffs for the solid (top) and deuterium (bottom) targets.

and deuterium targets, was applied (Fig. 3.15):

$$-2.2 < y_{corr} < 2.0 \text{ cm} \quad (3.1.12)$$

Fig. 3.15: y vertex of the electrons plotted in each sector, vertical lines correspond to the applied y -cut.

3.1.9 Kinematics

The kinematical constraints imposed on the scattered electrons were identical for all analyzed data: $Q^2 > 1 \text{ GeV}^2$ (parton probe) and $W > 2 \text{ GeV}$ (excludes nucleon resonances). It should be mentioned that the applied electron momentum cut is slightly higher than the minimum momentum cut permitted by the trigger threshold (equation 3.1.1). The reason for that is to avoid large contribution from radiative processes, to reduce which we impose a cut $y = \frac{\nu}{E_k} < 0.85$. At our energies this would be equivalent to: $P > 0.75 \text{ GeV}$.

Fig. 3.16: Q^2 vs x_{Bjorken} plotted on the left for the full kinematic range in Q^2 and W , and on the right - in the kinematic range for SIDIS: $Q^2 > 1 \text{ GeV}^2$ and $W > 2 \text{ GeV}$.

3.2 Particle ID scheme for photons

In order to ensure the presence of the electron, only the events where the first particle was an electron, satisfying all of the above ID cuts, were accepted. In those events, we consider all the neutral hits in the EC. Next, we require that the minimum deposited energy of the photon candidate in the calorimeter is above 300 MeV. One of the challenges in photon reconstruction is to separate them from the neutrons. During cooking, the neutral particles are pre-identified as neutrons if their velocity $\beta < 0.95$ and as photons if $\beta > 0.95$. To distinguish photons from the neutrons, as well as to clean up the sample from accidental photon candidates, a cut based on the matching of the EC time with the path which the photon has traveled in the EC was used. Assuming $\beta=1$ and the speed of light is 30 cm/ns, the following cut, illustrated in Fig. 3.17, is applied:

$$-2.2 < t_{EC} - t_{start} - l_{EC}/30 < 1.3 \quad (3.2.1)$$

where t is in (ns) and l is in (cm).

One can surely cut on the particle velocity β , calculated as a path length of the particle over time, directly as $0.95 < \beta < 1.95$. The upper limit of this cut off will take care of the particles with bad status which were assigned $\beta = 2$ (Fig. 3.18). Meanwhile, the coincidence time cut if imposed alone, without a direct cut on β , takes care not only of unwanted neutrons but also includes the resolution on β , which consists of the time and spatial resolution of the EC.

Fig. 3.17: $t_{EC} - t_{start} - l_{EC}/30$ plotted for all neutral tracks with the hit in the EC.

Lines in red correspond to the photon coincidence time cuts.

The fiducial cuts in EC for photons are similar to, but not the same as the ones of the electrons. Radiation lengths, as well as the showering mechanism for electrons and photons in the EC are quite alike. Yet, due to constraints from the effective volume of the CC detector and the fact that electron is bending in the magnetic field of the DC, the electron UVW cuts hence come out tighter as compared to the photons. The resulting UVW cuts are illustrated in Fig. 3.19 and Fig. 3.20 and are of the form:

$$40 < U < 410, V < 370, W < 410 \text{ (cm)} \quad (3.2.2)$$

Fig. 3.18: Beta distribution. Left plot is β of all neutral tracks with a hit in the EC, right - after the time coincidence cut was imposed. The red lines indicate a possible cutoff on the velocity, which becomes unnecessary if the coincidence time cut is applied. The particles with bad status were assigned $\beta = 2$.

Fig. 3.19: U, V, W distributions for the photon satisfying all the cuts but the EC fiducial. Red lines correspond to each cut for U, V, W.

Fig. 3.20: X vs Y global coordinates distribution for: all the neutral particles with a hit in the EC (left-hand plot), and photons satisfying the minimum deposited energy cutoff as well as the coincidence time (right-hand plot). Black distribution illustrates regions that have been removed by applying the U, V, W cuts.

3.2.1 Recapitulation of cuts, and available statistics

To conclude particle identification, we estimate the impact of all the cuts employed for electron and photon identification. The fraction of the events that would be accepted providing all the cuts but the cut under consideration are applied, except for the DIS kinematics (Q^2 , W) and invariant mass cut for the photon case, is given in Table 3.2.1 and Table 6.5. The kinematical cut aimed to select DIS events reduces the clean electron sample by 38%. The total number of electrons for C and D double target are 13.5 M and 15.6 M, for Fe and D are 24.4 M and 26.0 M, and finally for Pb and D are 9.6 M and 22.1 M respectively. Approximately 40% of all two photon combinations with their invariant mass being in the range $0. < m_{\gamma_1\gamma_2} < 0.27$ GeV are accepted to be π^0 . Number of π^0 per double target system for D and C, D and Fe, D and Pb are: 267.5 K and 194.3 K; 434.5 K and 255.6 K; 378.7 K and 78.0 K respectively.

Electron cut	Fraction of rejected events (%)
DC momentum	2.3
CC mirror-angle matching	3.2
CC Nphe	3
EC sampling fraction	3
EC inner energy	4.3
EC-SC coincidence time	2
EC UVW coordinates	1.6
DC fiducial cut	12.2

Table 3.4: Fraction of the events rejected for each electron cut.

Photon cut	Fraction of rejected events (%)
EC momentum	11
EC time	26 ¹
TOF β	4.8
EC UVW	8

Table 3.5: Fraction of the events rejected for each photon cut.

¹ effect of EC time cut is estimated without TOF β applied.

3.3 Identification

The reconstruction of π^0 was performed by its two photon invariant mass. Candidate events corresponding to $\pi^0 \rightarrow \gamma\gamma$ decays were required to have at least two neutral hits in any sector of the EC. Since a real photon has no mass, its $E_\gamma = P_\gamma$, with photon momentum taken from the ECPB bank as $P_\gamma = \text{Max}(E_{tot}, E_{in} + E_{out})/0.273$. Hence, one can calculate invariant mass of π^0 as follows:

$$q_1 = (E, P_{x_1}, P_{y_1}, P_{z_1}) \quad (3.3.1)$$

$$q_2 = (E, P_{x_2}, P_{y_2}, P_{z_2})$$

$$m_{\pi^0}^2 = (q_1 + q_2)^2 = 4E_1 E_2 \sin^2\left(\frac{\theta_{\gamma_1 \gamma_2}}{2}\right)$$

$$\cos(\theta_{\gamma_1 \gamma_2}) = \frac{\vec{P}_1 \cdot \vec{P}_2}{E_{\gamma_1} E_{\gamma_2}}$$

where $\theta_{\gamma_1 \gamma_2}$ is the opening angle between the two photons. The source of the restriction on the possible values of the opening angle comes from the finite angular resolution of the CLAS detector as well as the kinematical constraints on the available energy of the photon. Providing polar angular resolution of the EC is $\sigma_\theta \approx 1$ mrad (0.573°) [57], the minimum angular distance for the two clustering photons to be resolved is $2 \cdot 3\sigma_\theta \approx 3.45^\circ$. Meanwhile, the minimum opening angle θ_{min} between the two photons in the lab frame depends on their energy as:

$$\tan \frac{\theta_{min}}{2} = \frac{m_{\pi^0}}{E_{\gamma_1} + E_{\gamma_2}} \quad (3.3.2)$$

Since the maximum available energy of the final hadron is 4.25 GeV (for the scattered electron $P_e > 0.75$ GeV) and the minimum available energy is 0.6 GeV

(for the photon momenta $P_\gamma > 0.3$ GeV), the possible values of the opening angle are constrained within $4.5^\circ < \theta_{min} < 48.4^\circ$ respectively. Therefore, the kinematical limits of the available π^0 energy lead to more stringent opening angles rather than the angular resolution of the EC. In order to suppress possible accidental events in the π^0 distribution, one can think of applying an energy dependent cutoff on the opening angle between the two photons or, correlated to it, a π^0 mass cut. The latter one appears to be more appropriate for the purpose of our study. The distribution of the two photon invariant mass calculated according to equation 3.3.1, where electrons and photons satisfy all previously mentioned conditions, is illustrated in Fig. 3.21. Based on the Gaussian plus background fit to the invariant mass, we determine a mass range of the π^0 candidate such as:

$$|\mu - m_{\pi^0}| < 3 \cdot \sigma \quad (3.3.3)$$

where μ is the mean position of the peak (given by the coefficient p_4), and σ is the width of the peak (give by coefficient p_5). Fig. 3.22 illustrates the $\theta_{\gamma_1\gamma_2}$ dependence of the $E_{\pi^0} = E_{\gamma_1} + E_{\gamma_2}$. The left-hand plot contains all the π^0 candidate events, whereas the right-hand plot contains only those events which satisfy equation 3.3.3. The plot clearly illustrates that the cut off on the invariant mass range is sufficient in order to select events above minimum angular resolution of the detector. The theoretical curve of the minimum opening angle of the photons, θ_{min} , is always above the detector limitations, even for the highest energy π^0 .

Fig. 3.21: π^0 invariant mass distribution for the liquid deuterium (left) and solid carbon (right) targets. This distribution was fitted with a Gaussian plus polynomial background function, where background is plotted in blue and the π^0 peak with background subtracted is plotted in red. Events have been distinguished between the two targets (liquid and solid) based on the electron vertex cuts according to equations 3.1.11.

3.3.1 Photon energy correction

The only information available on the photon momentum comes from the energy deposited in the EC, which is in fact only a fraction of the initial photon energy. It is expected that the sampling fraction $f_s = \frac{E_{tot}}{P}$ is similar for both the electrons and the photons. The electron sampling fraction, defined as the ratio of reconstructed energy in the EC to the momentum determined from the DC, is of the form $f_s = \text{const} \cdot (E - E_0)/P$, where $f_s \rightarrow \text{const}$ at electron energies above 3 GeV/c [70]. Meanwhile, the momentum dependence of f_s for the photon is not

Fig. 3.22: $\theta_{\gamma_1\gamma_2}$ vs $E_{\pi^0} = E_{\gamma_1} + E_{\gamma_2}$ for all π^0 events (left-hand plot) and the events that satisfy the mass range cut [3.3,3.3] (right-hand plot). Black dashed curve corresponds to the theoretical value of the θ_{min} between the two photons.

so well understood. If the constant choice of $f_s = 0.273$ is employed for photons, the π^0 invariant mass in the EC can be in principle reconstructed with an accuracy of $\pm 5\%$ [70] from its PDG value. Our data set requires a photon energy correction also in order to account for the systematic uncertainties between the targets. For example, in the case of the C or Pb targets (and their corresponding D target), the mean value of the two photon invariant mass $M_{\gamma_1\gamma_2}$ integrated over the entire kinematics is reconstructed 1% below its nominal value. This deviation will be different for the case of the Fe target, which is reconstructed below M_{π^0} by 4%. If one considers a limited range in $z > 0.4$, the situation will be the opposite: the reconstructed $M_{\gamma_1\gamma_2}$ for the C and Pb is above M_{π^0} by 5%, whereas for the Fe

target by only 1%. Larger deviations occur when our dataset is finely binned.

The correction function for the photon energy can be obtained directly from the π^0 events following the procedure described in [71]. For the invariant mass of π^0 calculated according to equation 3.3.1 the correction function is defined by:

$$E_{corr} = \frac{E_\gamma}{corr(E_\gamma)} \quad (3.3.4)$$

where $E_\gamma = Max(E_{in} + E_{out}, E_{tot})/0.273$. The corrected two photon invariant mass would then become $m_{\gamma_1\gamma_2}^{corr} = 0.135$ GeV :

$$m_{\gamma_1\gamma_2}^{corr} = \sqrt{\frac{E_{\gamma_1}}{corr(E_{\gamma_1})}} \cdot \sqrt{\frac{E_{\gamma_2}}{corr(E_{\gamma_2})}} \cdot \sin(\frac{\theta_{\gamma_1\gamma_2}}{2}) \quad (3.3.5)$$

The photon energy correction was performed in two steps. In the first step, two photons of almost equal energies were selected, such that $E_{\gamma_1} - E_{\gamma_2} < 0.2 \cdot \text{bin width}$ (photons have been ordered by their decreasing energy). Providing that $E_{\gamma_1} \approx E_{\gamma_2}$, the ratio of (3.3.1) to (3.3.5) becomes:

$$\frac{m_{\gamma_1\gamma_2}}{0.135} = corr(E_{\gamma_1}). \quad (3.3.6)$$

The photons of step 1 belong to the range $0.3 < E_{\gamma_1} < 1.1$ GeV. The entire data sample has been divided in 9 bins of E_{γ_1} , where the width of the bins is of variable size (from 50 to 100 and 250 MeV at highest energy). The invariant mass of the two photons was fitted with a Gaussian and a second order polynomial in each bin, providing the mean value of $m_{\gamma_1\gamma_2}$ as a parameter $p4$ in the fit box illustrated in Fig. 3.23. The mean value from this fit normalized to the actual π^0 mass

$m_{\gamma_1\gamma_2}/0.135$ and plotted as a function of E_{γ_1} (Fig. 3.25) gives the first estimation of the photon energy correction. This function has the following dependence on the photon energy:

$$\text{corr}(E_{\gamma}) = a + \frac{b}{E_{\gamma}} + \frac{c}{E_{\gamma}^2} \quad (3.3.7)$$

Fig. 3.23: Step 1 of photon energy correction. π^0 invariant mass distribution, binned in E_{γ_1} , is constructed of the photons with $E_{\gamma_1} \approx E_{\gamma_2}$ (example of carbon).

In the second step, a π^0 candidate is constructed with either E_{γ_1} or E_{γ_2} being in the range where the correction of step 1 is valid (0.3-1.1 GeV). The photon from this range has its energy corrected using equation 3.3.7 obtained from step 1. Therefore, we can now calculate a correction for the other photon:

$$\frac{m_{\gamma_1\gamma_2}}{0.135} = \sqrt{\text{corr}(E_\gamma)}. \quad (3.3.8)$$

The uncorrected photon was restricted to the energy range $0.35 < E_\gamma^{\text{uncorr}} < 2.5$ GeV. The two photon invariant mass, divided into 14 bins by the energy of its uncorrected photon, has been fitted with a Gaussian plus background function. The background shape has been accessed via a number of functions as illustrated in Fig. 3.24: the red curve corresponds to the second order polynomial $a+b\cdot x+c\cdot x^2$, the blue one to $a + b/x + c/x^2$ and the green one to $a/x + b \cdot x + c \cdot x^2$. The mean value of the invariant mass obtained from the fit and normalized to the π^0 mass was fitted again as a function of E_γ (Fig. 3.25). The color scheme of each fit, or corrections obtained from step 2, corresponds to one of the shapes of the background function under the π^0 peak. The functional dependence of photon energy correction in step 2 is the same as defined in step 1 (3.3.7).

The choice of the coefficients a , b , c in the final correction function, which summarizes corrections obtained following step 1 and step 2, is the result of an iterative process. At first, the best guess for the correction from step 1 is fed back into the program to be applied in step 2.

Fig. 3.24: Step 2 where π^0 invariant mass was constructed with either E_{γ_1} or E_{γ_2} being in the range where correction of step 1 is valid, i.e. $0.3 < E_{\gamma}^{uncorr} < 1.1$ GeV, while the other photon was restricted in the energy range $0.35 < E_{\gamma}^{uncorr} < 2.5$ GeV.

Next, the correction obtained from step 2 (which is more sensitive to the higher energy photons) is reapplied back in order to improve the initial guess of correction function from step 1. This procedure was repeated until the fits corresponding to step 1 (black curve) and step 2 (red, blue or green) overlay in the maximum proximity, meaning that the results of many iterations have converged (Fig. 3.25). In reality, due to finite bin size along with the spread in the photon energy when the two photons were required to carry the same momentum (step 1), we are forced to guess the average correction function in between those obtained from step 1 and step 2. Independently of the choice of the background functions in step 2, the final correction function (corresponds to the thin magenta curve in Fig. 3.25) was found to be the same for the two double target systems, namely C and Pb (along with their D):

$$corr(E_\gamma) = 1.129 - \frac{0.05793}{E_\gamma} + \frac{1.07e - 12}{E_\gamma^2} \quad (3.3.9)$$

Following the exact same procedure as illustrated in the example of the C target, the coefficients for the final correction fit (equation 3.3.7) were found to be different for the Fe target:

$$corr(E_\gamma) = 1.116 - \frac{0.09213}{E_\gamma} + \frac{0.01007}{E_\gamma^2} \quad (3.3.10)$$

Finally, to check the validity of our corrections, we plot the uncorrected two photon invariant mass normalized to 0.135 as a function of photon energy E_γ , where $E_\gamma = (E_{\gamma_1} + E_{\gamma_2})/2$, and compare it with the same quantity after the

Fig. 3.25: $m_{\gamma_1\gamma_2}/0.135$ as a function of photon energy for the C (left) and Fe (right) targets. The black curve corresponds to the correction obtained from step 1, colored - from step 2, and magenta corresponds to the final correction applied.

photon energy correction was applied. In this check we require that the energy of each photon in the pair, E_{γ_1} and E_{γ_2} , belong to the same energy bin. Fig. 3.26 illustrates the mean values of the Gaussian fit to invariant masses before (green points) and after (black points) correction. The corrected values were fitted with a constant function which is consistent with 1 within 1% for all the targets.

It should be mentioned that the photon energy correction is in fact a correction to the photon sampling fraction f_s . A relative comparison between sampling fraction for electrons and the one for corrected photons reveals a systematic difference between the two to be of the order of 10% (Fig. 3.27). The electron sampling fraction comparison between Fe and other targets (C, Pb) confirms our results for

the target dependent photon energy correction, pointing to variations in the EC energy calibration within the entire data set.

Fig. 3.26: $m_{\gamma_1\gamma_2}/0.135$ vs E_γ . Here the points in black correspond to the corrected values of π^0 , whereas the points in green have not been corrected. The black line is the linear fit to the black points with a constant value of parameter p_0 , indicating a flat behavior around the unity. Left plot corresponds to the C target, the right one - to the Fe target.

3.3.2 Kinematical constraints on the photon

Apart from the identification cuts discussed earlier, two additional constraints on the photon kinematics were imposed. The first one is related to the minimum energy deposited in the EC. The design of CLAS calorimeter allows detection of the photons at energies above 0.2 GeV [57]. By examining the magnitude of photon energy correction factors for the Fe and both C and Pb targets, illustrated in Fig. 3.28, we impose more stringent cutoff on the minimum photon energy. At

Fig. 3.27: The distribution of E_{tot}/P vs P for the electrons in the example of C (left) and Fe (right) targets. The black curve represents a mean value of the Gaussian fit to the electron sampling fraction per momentum bin. The superimposed red and blue curves are the corrected sampling fraction of the photon $f_s \cdot corr(E_\gamma)$.

energies $E_\gamma = E_\gamma/0.273=0.2$ GeV, the correction factor for the Fe target is rather large, 20%. One has to be absolutely certain in the procedure to apply such large correction. In our case of limited statistics, it would be wise to choose a cut off avoiding such large corrections. At higher energies $E_\gamma=0.3$ GeV, the correction factor for all target systems is below 10%. Therefore, for each photon in the pair, before applying photon energy correction, we require that the deposited energy E_{dep} is:

$$\frac{E_{dep}}{0.273} > 0.3 \text{ (GeV)} \quad (3.3.11)$$

To purify our photon sample pertaining to a π^0 decay from the bremsstrahlung

Fig. 3.28: Photon energy correction factor plotted for two targets: Fe plotted according to equation 3.3.10 in blue, and C (or Pb) plotted according to equation 3.3.9 in red. Green dashed line corresponds to the minimum photon energy cutoff imposed in this analysis to be above 0.3 GeV.

photons radiated by propagating electron, we impose a cut off on the angle between the scattered electron and the photon under consideration. The angle peaking approximation, first proposed for (e,e') experiments by Schiff in 1952 [72], states that most of the bremsstrahlung photons from the electron are emitted either in the direction of incoming or scattered electron. Defining $\alpha_{e\gamma}$ as an angle between scattered electron and detected photon, the following cut, illustrated in Fig. 3.29, was imposed:

$$\alpha_{e\gamma} > 12 (^{\circ}) \quad (3.3.12)$$

Fig. 3.29: Angular distribution $\alpha_{e\gamma}$ between the scattered electron and one of the two photons, γ_1 and γ_2 , of a π^0 candidate. To avoid bremsstrahlung contamination of the photons, we impose a cutoff, indicated by the green line, on the $\alpha_{e\gamma} > 12^\circ$.

3.3.3 Event mixing method for π^0 background subtraction

Due to the shortness of the π^0 life time, its direct identification is impossible, therefore an invariant-mass analysis is the only procedure available when considering the SIDIS reaction. Construction of all possible combinations of two photons within one event bears a Gaussian-type peak distributed around the $m_{\gamma\gamma}$ invariant mass on top of a combinatorial background. In such type of analysis it is essential to separate the peak containing π^0 events from its combinatorial background. A straightforward approach to fit the peak on top of the background is to employ a fitting function consisting of Gaussian plus a polynomial. In a large statistical

sample of π^0 candidates a free-parameter polynomial treatment of background would usually produce a good fit to the data. However, applied to the sample of scarce statics and asymmetric background, it introduces large fluctuations which then lead to large errors on the extracted number of events. Since our data is finely binned with some of the bins poorly populated, *a priori* knowledge of the background shape is essential for reliable fitting.

A general guideline for evaluating combinatorial backgrounds is to use an event mixing procedure in which a background distribution is constructed from two or more particles originating from uncorrelated events. In particular, in order to access the π^0 background, a photon originating from one event would be combined with a photon from a different event. The mixing of uncorrelated events is carried out on the restricted kinematics - an improvement to the general approach, accounting for the difference of global characteristics of the two events, discussed further in this section.

Method

A general approach of constructing combinatorial background was tested on a sample of Monte Carlo events simulated via the LEPTO [77] event generator which models π^0 spectra in deep-inelastic kinematics at the energies of our experiment (described in the following chapter). In order to perform a test of the event mixing method we implemented realistic energy and position resolution to mimic

the actual calorimeter response of CLAS. The photon energy is smeared using the following parametrization:

$$\frac{\sigma}{E} = \frac{10.3\%}{\sqrt{(E)(GeV)}}$$

the angular smearing is calculated using position resolution $\sigma_R = 2.3$ cm and distance from the target to the CLAS center $R \approx 530$ cm. By smearing each photon originating from a π^0 decay, the resulting mass distribution of the generated π^0 will no longer be a δ -function but a Gaussian-like distribution bearing a width similar to the one observed in real data (Fig. 3.30). Note, that due to the default generator rounding (usually, three significant digits), the invariant mass constructed from the two photons would come out smeared around $m_{\pi^0} = 0.135$ GeV/c². To avoid the digital smearing, the precision of the generator output was changed to 6 digits. The testing of the mixing methods relies on the specific kinematical case which is constrained by the current data analysis. The sample of events is chosen such that the electron satisfies all the DIS cuts as those used in data, and the photons are selected within the range: $8^\circ < \theta_\gamma < 45^\circ$, $E_\gamma > 0.3$ GeV, $\alpha_{e,\gamma} > 12^\circ$ where $\alpha_{e,\gamma}$ is the angle between the scattered electron and the photon. These conditions were imposed based on the the detector design and kinematical constraints discussed earlier in section 3.3.2. It is important to consider the kinematic range employed in data analysis for the reason that if the photon kinematics is expanded over the full range of energies and angles, the tests of the general approach prove themselves viable while limited kinematical case is more complex. This is related

Fig. 3.30: The invariant mass of tagged π^0 on the outcome of LEPTO (left), with added smearing on the energy and position of the photon given realistic resolution in CLAS detector (right).

to the fact that the deviation of mixed shape when fitted to the data is reduced as it is averaged over the fluctuations in energy of the mixed photon pairs. By restricting photons phase space we conduct a more stringent test of the method by pinpointing the effects otherwise weakly pronounced. Before proceeding to the combinatorial methods of extracting the background shape, it should be noted that the second order polynomial fit describes well the generated data [73], however, if it is used to describe the real data, it's curvature is not sufficient. When a higher order polynomial fit is used to fit low statistic data, it introduces large uncertainties in the parameter fits and thus becomes unreliable. For that reason, it is important to establish a shape of the background that would work well on a

Fig. 3.31: General event mixing method on the example of the lowest z -bin ($0.3 < z < 0.4$). Left plot is the mixed background distribution (yellow) fitted with fourth order polynomial function (black). Right plot illustrates 2γ invariant mass distribution fitted with Gaussian (red) plus mixed background (black), the latter is normalized to the number of events in the peak. The number of events calculated from the fit (N_{gaus}) is $4.7 \pm 0.1\%$ smaller than the actual number of generated π^0 's.

small available statistics.

The general approach to event mixing consists of combining each single photon from one event with each single photon from another event into a pair, thus obtaining a mixed invariant mass distributions. Next, in order to extract the background shape per kinematical bin, the mixed background is divided into (Q^2, ν, z) and (ν, z, p_T^2) set of bins, the number and width of which correspond to the binning scheme used in data, and then fitted with appropriate function. Normalized to

the actual number of events, the shapes of the mixed background together with the Gaussian are used to fit the real data. Given the fact that the normalization is the only free parameter in the background fit to data, the error on the extracted N_{π^0} is reduced compared to the background subtraction methods with more free parameters. The mixed shape along with its fit to the invariant mass is illustrated in the example in the first z bin in Fig. 3.31. Other bins of 1D distributions in p_T^2 , Q^2 , ν bins exhibit a similar behavior in terms of the goodness of the fit. Adjusting the values of normalization and/or adding additional terms to the background fit, or choosing another non-Gaussian function to fit the peak does not improve the degree of deviation of fit from the actual distribution. Using a known number of generated π^0 's, the number of π^0 's extracted with the mixed background is smaller by $4.7 \pm 0.1\%$ in the first z -bin. Therefore, the general event mixing method has limited accuracy to correctly reproduce the actual combinatorial background, in particular, at low z as well as low p_T^2 bins.

The main reason for which the general approach to event mixing fails lies in the difference in the characteristics of the events used to construct a mixed pair. Unlike mixing photons from the same event, when two photons from different events are mixed, modifications of the global characteristics of an event must be accounted for. Consider two uncorrelated photons: the first one originates from a certain γ^* and a π^0 , the second photon comes from a different event with a different parent π^0 and a different γ^* . Combining those two photons into a pair

results in an event that carries neither direction nor kinematics of the first or the second π^0 . In principle, a mixed photon pair does not violate kinematics with respect to the single-event a photon pair, however, on a large statistical sample this recombination will not respect the tails and shapes of the actual 2γ distributions. In order to resolve this, we restrict the phase-space available for picking up two uncorrelated photons (say, the "old" and the "new" one) by imposing the following condition: the scattered electron and π^0 candidate from the "old" event must belong to the same (Q^2, ν, z, p_T^2) bin as those from the "new" event. If only the above condition is satisfied, a photon from the "old" π^0 candidate will be combined with a photon from the "new" π^0 candidate. Next, all the photons in an event are ordered by decreasing energy, so that when the pair is constructed a more energetic photon from the "new" event will be combined with a less energetic one from the "old" event, and vice versa. This helps to avoid double counting of pairs mainly constructed by combining the two highest or only the two lowest energy photons. Finally, photons originating from an "old" event are rotated around the beam direction on the angle $\delta\phi = \phi_{new}^{e^-} - \phi_{old}^{e^-}$ which brings "old" and "new" photons closer to each other as if they have originated from the same γ^* (Fig. 3.32). Adding rotation also on $\delta\theta$ would bring photons from different events on top of each other, however, it would also shift the actual beam position in the rotated event. Therefore, only the $\delta\phi$ rotation is performed. To further refine the procedure, the following options were considered. One of them

Fig. 3.32: Illustration of the correction procedure in general mixing method. In order to preserve kinematical correlations in the pair of mixes uncorrelated photons, a photon from one event (blue) is rotated to the photon in another (green) by the difference in their azimuthal angles.

consists of imposing an additional condition where a π^0 resulting from mixing of two uncorrelated photons would be accepted only if it belongs to the same bin in p_T^2 and z as both π^0 's, the ones from which the "old" photon and the "new" photon originated. Another option would be to select events for mixing such that their initial kinematics is alike. This could be done by imposing a restriction on the momentum transfer by the lepton requiring, for example, that the difference in the momentum of two virtual photons does not exceed 50 GeV. While such refinement of mixing procedure demonstrated ever better fit to the

Fig. 3.33: Combinatorial background based on the improved mixing method: mixed events originated from the same (Q^2, ν, z, p_T^2) bins, with kinematics of one event rotated by the $\delta\phi$ to coincide with the other event, and mixed based on the photon energy ordering. Left plot is the mixed background distribution (yellow). Right plot is 2γ invariant mass distribution fitted with Gaussian (red) plus mixed background (black). The accuracy of π^0 extraction is at the percent level.

generated distribution [73], [74], it was not adopted due to the fact that either of the additional conditions significantly decrease statistics available for mixing events from the experimental data sample.

To summarize, the combinatorial background was extracted by mixing events originated from the same (Q^2, ν, z, p_T^2) bins, with kinematics of one event rotated by the difference in azimuthal angle to coincide with the other event, and mixed

based on the photon energy ordering. The result is illustrated in Fig. 3.33 on the example of a single bin in z . The developed mixing scheme, as applied on the sample of MC generated events, leads to a percent level accuracy on the extraction of the number of π^0 events from 2γ invariant mass combinations.

The improvement of general event mixing proved accurate on the sample of MC generated events. In case of its application to the experimental data and reconstructed from CLAS MC events, the improved method generally performs well, yet, slight deviation in certain kinematics were observed. The latter is mainly related to the energy differentiation in EC of the two neighboring photon clusters. When the two photons (partially) overlap, the energy of the clusters is divided between them, however, the correlation that exists within one event, will no longer be preserved when mixing two uncorrelated events. This gives rise to the background deviations, in particular, at low photon deposited energies. Nevertheless, in order to fit data and reconstructed MC events, a priori knowledge of the background shape is crucial as it provides a more accurate fit in those bins where background shape is strongly non-linear (as exemplified in Fig. 3.34, and 3.35), or those where overall statistics is low.

Fig. 3.34: Invariant mass of the two photons from lead target in the slice of bins

$0.3 < z < 0.4$, $0 < p_T^2 < 0.1 \text{ GeV}^2$ and from left to right bins in ν : $2.2 < \nu < 3.2$, $3.2 < \nu < 3.73$ and $3.73 < \nu < 4.25 \text{ GeV}$. Top panel is fitted with combinatorial background based on event mixing (red line), bottom panel is the second order polynomial fit (blue). The fit is done once on the wide range to obtain approximate width σ , the second fit is then a refit on the range $\pm 5\sigma$, indicated by the color of the background.

Fig. 3.35: Sample of reconstructed MC events fitted on $\pm 5\sigma$ range in the same slice of bins as described in Fig.3.34. Top panel corresponds to combinatorial background from mixing reconstructed events employing the same method as developed for data, the bottom panel is the second order polynomial fit. The curvature of the polynomial background varies depending on the fit limits (black compared to blue), while combinatorial background is stable regardless.

3.4 Binning

The available π^0 statistics does not allow us to perform four-fold binning in the variables under interest. Hence, we divide and analyze our data in two sets of three-fold bins: (Q^2, ν, z) and (ν, z, p_T^2) . The number and width of the bins in a given variable are the same for each of the two sets, except for p_T^2 which we will discuss in the following section. Fig. 3.36 illustrates phase space of electrons divided into (Q^2, ν) bins and phase space of π^0 's divided into z and p_T^2 .

Variable	N_{bins}	Bin edges
Q^2	3	(1.0, 1.33, 1.76, 4.1)
ν	3	(2.2, 3.2, 3.73, 4.25)
z	5	(0.3, 0.4, 0.5, 0.6, 0.7, 0.8)
p_T^2	6	(0., 0.1, 0.25, 0.4, 0.55, 0.75, 0.9)

Table 3.6: Number of bins and corresponding bin limits in Q^2 , ν , z , and p_T^2 .

Fig. 3.36: Electron phase-space is divided into 3 bins each in Q^2 and ν (left). π^0 is divide into 5 and 6 bins correspondingly in z and p_T^2 (right).

3.5 p_T^2 cutoff

While the binning scheme in z is guided by the physics interest and is confined inside a larger kinematical space, the upper limit of p_T^2 distribution have to be determined based on the statistical significance. The tail of the p_T^2 spectrum is likely to be contaminated by bin migration and statistical fluctuations, hence, a careful upper limit on p_T^2 values must be determined.

First we consider the set (ν, z, p_T^2) binned in p_T^2 up to 1.5 GeV^2 . The amount of events discarded above this cut relative to the total number of events in the bin varies up to 2 % as a function of increasing (ν, z) bin. At low values of ν and z , the distributions in p_T^2 fall rapidly below the cutoff. There are no fluctuations at the tail, mainly due to the large statistical sample available in those bins. At higher values of ν and z , the tail of the p_T^2 distribution is smeared above the cutoff, however, statistics in those bins drops by a factor ten, thus giving rise to the fluctuations at the end of the spectrum. The cutoff becomes significant at higher values of ν and z , where the tail is smeared and statistics are dropped by an order of magnitude.

Next, we look at (Q^2, ν, z) set integrated over p_T^2 . Here the limits of integration must exclude tails related to the statistical fluctuations. Since statistics largely differ between the targets, for each target type we investigate an individual cutoff based on the fluctuations at the end of p_T^2 spectrum. A handful of approaches have been investigated [75]. The chosen criteria of determining a cutoff consists

Fig. 3.37: p_T^2 distribution (blue) in the slice of (Q^2, ν, z) bins for the case of the lead target (minimum statistics and maximum fluctuations compared with other targets). The black lines correspond to the Gaussian fit. The green vertical line corresponds to the drop of the fit function down to one event per bin, and the red line is the corresponding p_T^2 cutoff. The number of events is plotted per 3 MeV^2 in p_T^2 .

of fitting the p_T^2 distributions in each (Q^2, ν, z) bin with a Gaussian function defined on a limited range of p_T^2 . We found that such a choice of a function describes the p_T^2 distributions the most effectively before the statistics drops at the tails. At the values of p_T^2 , for which our function contains less than one event per bin, we define a cutoff. The amount of rejected events will depend on (Q^2, ν, z) bins and the target type with the tightest cut due to statistics for the lead target, as illustrated in Fig. 3.37. So calculated limits on maximum p_T^2 confirm that the actual cutoff at $p_T^2 = 0.9 \text{ (GeV}^2\text{)}$ that we employ in the analysis is safe to use for the integration in each single bin of (Q^2, ν, z) .

3.6 Exclusive contribution

In certain kinematical regime, a π^0 , identified as semi-inclusive, may in fact originate from an exclusive reaction on a proton or a neutron: $eN \rightarrow e'N\pi^0$. The exclusive processes dominates the SIDIS process at high momentum transferred Q^2 and fixed x_B , providing the detected pion carries a large fraction of transferred energy z and large transverse momentum p_T (Fig. 3.38). At these kinematical limits, exclusive process would be a substantial contribution to the measurement of the SIDIS process. Below those limits, in the kinematical domain where SIDIS prevails, exclusive reactions can still occur when, for instance, outgoing pion scatters backwards relative to the initial virtual photon direction. The fraction of carried away energy, z , in such scenario would be small, $z \rightarrow 0$. The cross section

Fig. 3.38: Missing mass distribution W_x . The black curve is W_x in the full SIDIS kinematics, the blue is constrained on $0.8 < z < 1.0$, where missing mass approaches exclusive kinematics: $W_x \rightarrow M_p$.

of such a process is small as well; it is accounted for in the total SIDIS cross section at the level of radiative corrections described in section 5.2.5. The goal in investigating the exclusive processes is to select and exclude a kinematical range where exclusive process is likely to dominate, while correcting for it everywhere else by means of radiative corrections where its contribution is small.

Given the scattering off the elementary target, the missing mass W_x of the recoil system in an exclusive reaction would be that of a target nucleon. We further use the notion of W_x (3.6.2) in connection to z and p_T of the detected pion:

$$\begin{aligned}
 W_x^2 &= (P_{\gamma^*} + P_p - P_\pi)^2 \\
 &= W^2 + m_\pi^2 - 2\nu^2 z + 2P_\pi \sqrt{Q^2 + \nu^2} \cos\theta - 2M_p \nu z
 \end{aligned}
 \tag{3.6.1}$$

Such correlation is used to determine the kinematical limits which are dominated

by SIDIS processes while minimizing the contamination from the exclusive processes. For a detected π^0 to be produced in a SIDIS reaction, a minimum requirement is a production of an additional pion in the process: $eN \rightarrow e'N\pi^0\pi^0$. Thus, the low limit on the missing energy, separating exclusive and semi-inclusive pion production, is such that $W_x > (M_p + M_\pi^0)$. Based on this knowledge, the limits of z and p_T can be deduced correspondingly.

The case of nuclear targets is complicated by Fermi motion as well as intra-nuclear interactions. Those effects lead to the smearing of the missing mass spectrum as illustrated in Fig. 3.39. This in turn introduces a bias in determining the low limit on the M_x cut which could have been a viable quantity in separating exclusive and SIDIS events. For that reason, instead of imposing arbitrary cut on the missing mass, we exclude regions in z and p_T^2 where exclusive contribution is potentially large. Clearly, as seen in Fig. 3.40, at both values of high z ($0.8 < z < 1.0$) and high p_T^2 ($p_T^2 > 0.9 - 1.0 \text{ GeV}^2$), the phase space covers dominantly exclusively allowed kinematics with W_x around that of the mass of the proton. The convergence of nuclear smearing along with radiative effects, as well as detector resolution represents a challenge in evaluating the relative contribution of the semi-inclusive π^0 populating kinematically exclusive domain. For that reason, we exclude high z and p_T^2 imposing an upper boundary to be that of $z < 0.8$ and $p_T^2 < 0.9 \text{ GeV}^2$. A detailed study of the described issue, using both LEPTO and binned data, can be found in [76].

Fig. 3.39: p_T^2 vs z distribution for π^0 events generated using LEPTO without (left) and with (right) Fermi motion smearing of the target nucleon from deuterium. The W_x curves illustrate two cases: detected π^0 originating from exclusive reaction (red) or semi-inclusive (green).

Fig. 3.40: p_T^2 vs z distribution of π^0 candidates from data. The W_x curves illustrate limits on the exclusive π^0 production at $W_x=M_p$, and SIDIS at increasing values of W_x .

Chapter 4

Monte Carlo Simulation

The large coverage of the CLAS detector has many advantages for studying electromagnetic interactions with nucleons and nuclei. Along with those benefits come associated difficulties in the data analysis due to complexity of the detector and high particle yields, both of which result in data reduction. Therefore, in order to extract detector independent results as well as estimate related uncertainties, measurements performed with the CLAS detector must be corrected for acceptances and adjusted for resolution effects. The task associated with simulation can be virtually divided into two parts: generation of physical events and their reconstruction in detector components. To accomplish the first, a Monte Carlo (MC) event generator LEPTO, based on realistic physical processes, is used. LEPTO generates complete DIS events, employing the DGLAP evolution to the simulation of partonic processes, and LUND string model for hadronization in order to produce hadronic final states. The second step consists of feeding LEPTO generated events into routines which simulate the response of the CLAS detector. Its geometry is built in the Geant SIMulation (GSIM) program which allows one

to model the passage of various particles through the detector. Next, Gsim Post Processor (GPP) eliminates signals from the dead wires and smears signals in the detector components so to match the momentum distributions in real data. The output is then processed with the standard CLAS reconstruction software used on the experimental data. By performing a multidimensional comparison of the number of events generated on the output of LEPTO, and the number of events reconstructed in CLAS, we calculate the acceptance correction factors to our data.

4.1 Adaptation of LEPTO

The analysis presented here uses Monte Carlo event generator LEPTO 6.5.1 [77] in order to simulate complete events in the deep-inelastic lepton-nucleon scattering in the kinematics of our experiment. The simulation of physics events in LEPTO is a complex multi-stage processes with a large number of tunable parameters and options. In general, following a hard interaction, LEPTO incorporates perturbative QCD calculations of the partonic final states and non-perturbative hadronization models describing hadronic final states. QCD processes are calculated exactly to the first order matrix elements (ME) that contain the following processes: leading order parton level process $\gamma^*q \rightarrow q$, first order gluon radiation $\gamma^*q \rightarrow qg$, the boson-gluon fusion $\gamma^*g \rightarrow q\bar{q}$. Higher order QCD effects are accounted for by a bremsstrahlung-like parton shower (PS) approach in the leading $\log Q^2$ approximation of the DGLAP evolution equation. While ME and PS are

default approaches to hard and soft emission, an alternative realization of parton showers is provided by the color dipole model (CDM) called via the QCD cascade generator ARIADNE. The CDM differs from QCD cascades by describing gluon bremsstrahlung in terms of radiation from color dipoles between partons instead of treating them as independent emitters. Fragmentation into final state hadrons is performed with the Lund string model [24] based on the effective string picture of QCD at strong coupling. It is implemented in LEPTO via the JETSET package [78].

The default values of LEPTO parameters are set to describe the physics at HERA energies. The adaptation of LEPTO to JLab energies and the physics of interest includes a number of modifications listed below with corresponding switches indicated in the brackets:

- energy of incoming electron beam is set to 5.014 GeV (plz=5.014 GeV)
- simulation of electromagnetic process (lst(23)=1);
- choice of independent variables of integration to be x and Q^2 (lst(31)=1);
- simulation of QCD effects is set by ARIANDE (lst(8)= 9);
- choice of parton distribution functions are set according to CTEQ2D (lst(15)=10);
- cutoffs against divergences in QCD ME elements (parl(8)=0.02, parl(9)=3.5);
- width of the primarily Gaussian transverse momentum in fragmentation is narrowed (PYTHIA switch parj(21)= 0.3 GeV);
- tail and width of the secondary Gaussian distributions were widened (PYTHIA

switch parj(23)=0.02, parj(24) = 3.5 GeV) ;

- parameters of Lund fragmentation functions were adjusted (PYTHIA switch parj(41) = 0.6, parj(42) = 0.1);

- energy of initial kinematics is allowed to vary from event to event in order to account for nucleon Fermi motion (lst(17)=1);

- effective kinematic range calculated from the specified cuts on W , x , Q^2 :

$$0.09 < x < 0.7, 0.4 < y < 1.0, 0.8 < Q^2 < 6.3, 4.0 < W^2 < 9.4, 1.0 < \nu < 5.0;$$

- precision of the output has been changed from 3 significant digits to 6 by modification inside the LEPTO source code.

The above choice of parameters used to generate DIS events in our kinematical range is not unique. Most of the switches related to parton cascades and fragmentation can be set to default values without significant change to the final state distributions. The list of modified parameters is an artifact of many tunings aimed at finding the right configuration for which continuous generation of many events is possible, in particular for the heavier targets. LEPTO is known for containing a switch which enables to specify the number of nucleons A and protons Z in the target nucleus. This switch, however, merely accounts for the relative content of u to d quarks without taking into account nuclear modifications of fragmentation functions. As a result, kinematical characteristics of final state π^0 's are nearly identical for various combinations of A and Z as illustrated in Fig. 4.1. Finally, random seed used for single event generation was improved

compared to the default value by setting it to the multiplication of the current date and time.

A good agreement of reconstructed electrons with electrons from experimental data is observed. A disagreement, however, arises in the shape of hadron kinematics, most prominently at the high- p_T^2 tail. The deficit of high p_T^2 events stems from the cut off embedded in the generator which is proportional to the scale defined by Q^2 . Hence, altering switches and tuning parameters, whilst leaving initial kinematics unchanged, does not solve the problem of the high- p_T^2 shape. This problem can be addressed internally within LEPTO source code. One can introduce a parametrization on the p_T^2 distribution at the generation level, or modify kinematical constraints on the phase space of produced hadrons as a function of initial kinematics. Both of these options were considered, however, given the time investment of iterative introduction of parametrization as well as complexity of source code and its connection to other packages, implementing either one is rather non-trivial. Later we will describe a possible external method to adjust p_T^2 after reconstruction.

4.1.1 Fermi momentum

Fermi momentum of the target nucleon has been incorporated in LEPTO source code by calling an external C++ function which returns projections of momentum randomized according to theoretical parametrization. The main objective to

Fig. 4.1: Generated z and p_T^2 distributions for the D, C, Fe, and Pb targets using LEPTO switch for the mass and charge of the nuclear target. This switch is responsible for the ratio of valence quarks u and d in the initial state. Its effect appears negligible on the characteristics of π^0 produced in the final state.

include nucleon momentum smearing was to compensate for the low probability of generated final state hadrons to carry transverse momentum above $p_T^2=1.5 \text{ GeV}^2$. The smearing of nucleon momentum for the D, C, Fe, and Pb nuclei have been done using Ciofi-Simula parametrization [86] which represents a probability to find in a nucleus a nucleon with momentum P_F as depicted in Fig. 4.2. Projections of this momentum are calculated in a randomly generated whole solid angle.

Adding momentum to the target nucleon leads to the modification of initial kinematics as compared to the one when the nucleon is at rest. Therefore, characteristics of events in the final state are also modified as illustrated in Fig. 4.3 on

Fig. 4.2: Probability n_k per GeV of Fermi momentum P_F distribution from Ciofi-Simula parametrization for D(green), C(red), Fe(blue) and Pb(black).

Fig. 4.3: From left to right are Q^2 , ν , z and p_T^2 distribution for the two cases: target nucleon is at rest (black), and target nucleon carries Fermi momentum randomized according to Ciofi-Simula parametrization (blue).

the example of electrons and π^0 's. Fermi momentum distributions are similar for all heavy targets, hence, the difference in their relative effect on the kinematics of final state hadrons will be small. To evaluate possible contribution of Fermi motion on the π^0 multiplicity ratios, we have combined generated statistics from

three nuclear targets (C, Fe, Pb) and divided it by the number of yields produced on D. Fig. 4.4 illustrates that the contribution to the ratios from Fermi motion is less than 1% as a function of Q^2 and ν but it reaches up to 5% in certain z and p_T^2 bins. When compared to statistical uncertainties in π^0 yields, this effect is not likely to be pronounced.

Fig. 4.4: π^0 multiplicity ratios of events that were generated using Ciofi-Simula parametrization for target nucleon. Ratios are extracted by combining statistics on for C, Fe, Pb targets normalized to D. The deviation from 1 demonstrates possible contributions from Fermi momentum to the multiplicities. From left to right multiplicities are plotted as a function of Q^2 , ν , z and p_T^2 respectively.

4.2 Simulations and reconstruction

The main motivation of MC simulations is to extract acceptance correction factors which we discuss in the following section. The architecture of the simulation and reconstruction processes is straightforward. The text output of the generated events from LEPTO is converted into a BOS (Bank Organization System) file to

be fed into the program that contains idealistic response of the CLAS detector as well as target geometry. The latter produces another BOS file which further proceeds through a map of detector efficiencies. Finally, the physical events are reconstructed using a realistic detector package.

4.2.1 GSIM

The GSIM program produces a detector response on the passage of the particle. This response incorporates geometry and resolution of the detector along with physical processes happening within. The smearing of DC resolution is calculated as a difference of the two distances between the track and the sensible wire: one calculated utilizing the drift time of the wire and another one by using tracking procedure based on the response from all the cells. TOF and EC responses were smeared using known time resolution and energy or position resolutions correspondingly. It is, however, not possible to fully simulate the complexity of all detector components, for example, energy deposited by the electromagnetic cascade in the EC, especially near the edges, or accurate geometry of Čerenkov mirrors and thus the distribution of the number of the photoelectrons. The precision of the simulation is specified in the configuration file by selecting a threshold energy (1 MeV in our case). Once a particle reaches this threshold, GSIM considers that all the particle's energy be deposited in the given geometric volume. Lowering the threshold energy considerably increases the time of the simulation. The largest

Fig. 4.5: The double-target assembly for the EG2 run as implemented in the production version of GSIM. It consists of: a deuterium cryotarget with the cell and solid target with its support structures.

contributors to the acceptance of charged particles, apart from the physical geometry of the detector, is toroidal magnetic field which causes particles to bend in the polar angle with the strength dependent on their momenta. Geometry of the detector and toroid coils are contained in the standard GSIM and can be modified through configuration file depending on the characteristics of the experiment. Meanwhile, target geometry for each run must to be implemented separately.

The double target system has been added by H.Hakobyan [79] in the production version of GSIM in the CLAS software repository. It consists of: a deuterium cryotarget together with the cell which includes a cone, standoff, entrance and exit foils; solid target (material varies depending on the target choice) and its support structures which include carbon fiber loop and the rod, the aluminum

frame holder together with the steal rod and pivoting arm (Fig. 4.5). The foam scattering chamber was implemented with exact geometries and densities. To run simulations for the eg2 target set up, one has to merely specify target type keys [80]. The broader definition of acceptance correction includes physical processes which take place at the level of reconstruction. They are accounted for through their implementation in GSIM [82] and include: energy loss, multiple scattering and bremsstrahlung, all occurring as a result of an interaction as the particle traverses the medium; particle decay of the short-lived final state which takes place between the event vertex and detector elements; and emission of the secondary particles while the original particle transverses different parts of the detector.

4.2.2 GPP

In the exit file from GSIM the detector response to the passage of the particle is given in the assumption that all detector components are fully functional while in reality each experiment has its own set of dead channels and dead photomultiplier tubes. GSIM Post Processor (GPP) carries out the adjustments to the reconstructed GSIM data in order to better match experimental data. It removes events for which detector response falls in one of the channels with low efficiency. A map for the Drift Chambers Inefficiencies has been created by L.Zana [83] for each of the target configurations.

4.2.3 RECSIS

The output of GPP has been processed with standard CLAS reconstruction software RECSIS (REConstruction and AnalySIS package) written mostly in FORTRAN. It was built using the exact same tracking packages, contained in the `user_ana`, as the ones utilized for data reconstruction. Run control parameters for RECSIS are provided through a tcl script. Most of the default values can be used, but some (torus current, target position, etc) have to be set according to the specific experiment. The architecture of RECSIS is based on the encapsulation of each detector sub-system, which has its own shared libraries at the level of compilation and execution. Each event can be assigned to one of three types: header, scalar or physical event. In the case of the physical event, RECSIS acquires accelerator RF information which determines the time of the primary vertex. Next, for the electron runs each event is treated in the following order: reconstruction in the EC; determination of the hit-based track in the DC in each of the three regions which are further matched by the position of the wires in which the signal was detected; reconstruction of the time-of-flight from the scintillators; calculation of the number of photoelectrons from the Čerenkov; crude identification of the particles associated with different detector subsystem; secondary revision of the DC tracks based on the use of the time-of-flight (time-based hit); complete reconstruction of the event (Event Builder). Finally, reconstructed data is written out in the BOS banks and can be checked using monitoring histograms.

4.3 Particle identification

The GSIM package, as discussed above, contains idealized geometrical description and response of the detector components to the passage of the particle. Therefore, our criteria for particle identification, as established for the experimental data set, has to be revised for reconstructed MC events. Due to the complexity of shower simulation in the EC, we recalculate our electron and photon corrections on the deposited energy using the same methods as developed for real data.

4.3.1 Electron identification

To select an electron sample, similarly to the selection determined in data, we first require a negative track in the DC to have an associated hit in the EC and SC. Secondly, we apply the exact same kinematics cuts which define the DIS boundaries in data along with geometrical cuts, which excludes low efficiency regions of the DC and EC as well as a mismatch between angle and mirror number in the CC. Deuterium and solids targets are distinguished using the same vertex cuts as applied in data. These cuts are illustrated in the example of generated and reconstructed electrons in Fig. 4.8. Finally, using an electron sample that satisfies the above conditions, we revise the momentum dependent cut on the electron sampling fraction in the EC. The mean value and the width of the Gaussian fit to the sampling fraction as a function of electron momenta to a good approximation are constant as a function of the CLAS sector (in contrast to data). Therefore,

Fig. 4.6: The sampling fraction $f_s = (E_{dep}/P)$ vs electron momentum (P) in each sector (1 to 6 from top left to bottom right, respectively). The black crosses are the mean values per momentum bin of the Gaussian fit to f_s . The black solid curves are the 2nd order polynomial fits to the mean values in each sector; the black dashed curve is the same fit but to the mean values integrated over all sectors. Since the difference between the solid and the dashed curve is small, we use the latter to define the cutoff for f_s , same for all sectors.

one cut for all sectors has been employed:

$$\left| \frac{E_{dep}}{P} - (0.2623 + 0.0089 \cdot P - 0.0019 \cdot P^2) \right| < 2.5 \cdot \sqrt{0.0057^2 + \frac{0.0305^2}{P}} \quad (4.3.1)$$

where E_{dep} is the maximum energy between the total energy deposited in the EC and the sum of energies deposited in the inner and outer layers; P is the electron momentum as calculated from tracking information in the DC. The above

parametrization has been obtained from the mean μ and the width σ of the Gaussian fit to the sampling fraction E_{dep}/P in each P bin as presented in Fig. 4.6 and Fig. 4.7.

Fig. 4.7: Width of the Gaussian fit to the sampling fraction plotted as a function of electron momentum (P) in each of the six CLAS sectors (black crosses). The red crosses are the mean values of fit to the sampling fraction integrated over all sectors. The latter is fitted with the inverse square of electron momenta, illustrated by the red line. The upper and lower limits of the cut off on sampling fraction are determined from the deviation of mean values in the range 2.5σ , where σ is given by the red fit.

Fig. 4.8: The top panel corresponds to the z-position, the bottom panel is the y position of the target vertex. For both panels distributions, on the left hand side are the vertexes of generated events, distributions on the right - are reconstructed events with vertex cuts indicated by dashed curves. The values of vertex cuts are the same as those in data. Green color indicates the deuterium target, black is the iron target.

4.3.2 Photon identification

The cut off on the minimum energy deposited by a single photon in the EC is imposed at 300 MeV (before energy corrections). Photon identification as development from experimental data requires to remove fiducial regions at the edges of the calorimeter and include a cut on the matching between the time of the photon measured in the EC and the path which it has traveled. We applied the same EC fiducial cuts on the reconstructed sample, however, timing cut were revised due to the trigger time inconsistencies. There are at least three ways to extract trigger time: directly from the EVNT bank, or by constructing the

difference between the time measured in detector components relative to the event start time and the actual path traveled by the electron is determined either from the time-of-flight counters, or from the EC. To check which method of calculating the trigger time gives us the most consistent result, we examine corresponding distributions of photon velocities (Fig.4.9):

$$\beta = \frac{l^{EC}}{(t^{EC} - t_{trig}) \cdot c} \quad (4.3.2)$$

where t_{trig} is calculated either as $t_{trig}=t^{SC} - l^{SC}/c$, or $t_{trig}=t^{EC} - l^{EC}/c$, or is taken directly from the EVNT bank.

Fig. 4.9: Distribution of photon velocities obtained using different start times: as taken directly from the EVNT bank (red), as calculated using time and path information from the SC (blue) or EC (green).

Unlike measured data, distribution of photon velocities has large displacement from unity and an additional structure at $\beta=0.95$ if calculated using the start time from the EVNT bank. Meanwhile, velocities obtained using information from the SC or EC (blue and green curves in Fig. 4.9) are in better agreement with the data. Since trigger time calculated from the time-of-flight counters produces the smearing of velocities closer to the ones observed in data, we employ it in our cut off. Therefore, the following cut off the coincidence time is applied:

$$-2.2 < t^{EC} - \frac{l^{EC}}{30} - t_{trig}^{SC} < 1.3 \quad (4.3.3)$$

Further, in a number of studies, for example, photon energy correction and bin purity, which benefit from minimizing the background under the invariant mass peak, we will be using an angular matching criteria. This criteria is based on the idea that if a given reconstructed photon comes from a given generated photon, there momentum vectors would be in the vicinity of one another. We can extend this idea to find matching π^0 's: to ensure that given reconstructed π^0 comes from a given generated π^0 , we require that both photons in reconstructed π^0 are matched by the angle with the photons from generated π^0 . This is done by requiring that the angle $\alpha_{\gamma(rec)\gamma(gen)} > 3^\circ$ as illustrated in Fig. 4.10.

4.3.3 Photon energy correction

The next step, following photon identification, consists in correcting the energy that the photon deposited in the EC. A straight forward way to do it would be

Fig. 4.10: Angular distribution between the generated and reconstructed photons.

The cutoff $\alpha_{\gamma(rec)\gamma(gen)} > 3^\circ$ determines the angular matching criteria for a given reconstructed photon come from a given generated photon.

to correct a sampling fraction E_{dep}/E_{gen} as a function of E_{dep} . Here, E_{dep} refers to the energy deposited in the EC by the reconstructed photon and normalized to a constant sampling fraction equal to 0.2806. The reconstructed photon is matched by the angle to the generated photon which carries energy E_{gen} . By fitting the sampling fraction in each slice of E_{dep} , we obtain momentum-dependent corrections. This correction, however, when applied to the deposited energy of the reconstructed photon, does not provide a proper correction to the invariant mass of π^0 , underestimating it at lower and medium energies. The reason is that the sampling fraction distribution is not strictly a Gaussian in the vertical slices of E_{gen} , it is comprised of the photons in the entire range of energies, and consequently - resolutions.

Fig. 4.11: Step 2 of reconstructed photon energy correction. M_γ is binned in $E_\gamma = 0.15\text{-}3.0$ GeV with width varying from 0.025 GeV at low and medium photon energies to 0.2 GeV at high energies. π^0 candidates, fitted with a Gaussian function in red, are selected by matching an angle of reconstructed and generated photons which leads to a decreased background under the M_γ peak. One of the photons in M_γ is in the range 0.15 - 1.1 GeV corrected by step 1 function, while the second photon is uncorrected in the range 0.15 - 3 GeV.

In the second step, we correct one of the two photons by the factor defined in the first step, provided this photon lays in the same energy range as required in the first step, while the other photon can carry arbitrary energy in the range 0.15 - 3 GeV. By the same means as in the first step, we obtain improved functions for the photon energy correction. Ideally, iteration of the above steps must lead to a divergence of corrected invariant mass to its true value, and correction obtained in step 1 should be the same as the correction. For that reason, we will correct photon energy using the same method as employed in data. Due to larger available statistics in our simulations, the bin sizes in which corrections are performed, are adjusted to be much finer. In order to minimize the background, we select π^0 's for which reconstructed photons are matched by the angle with generated photons.

The first step consists of selecting photons in the narrow energy range, which was set to vary within 1/10 of the energy of the bin size. The width of the bin increases from 50 MeV at low energies and increments up to 300 MeV at high energies. The fit to the invariant mass of a pair of photons stemming from the same bin provides a mean value of invariant mass. Normalized to the actual π^0 mass, this ratio as a function of photon energy leads to the first guess for the energy correction (Fig. 4.12). However, due to the finite bin size and relatively wide range in the photon coincidence energy, such an outcome is not likely. Hence, we make an iterative guess for the final correction function to be somewhere in

between that of step 1 and 2:

$$E_{dep}^{corr} = 1.0 + 0.005/E_{dep} - 0.0052/(E_{dep} * E_{dep}) \quad (4.3.4)$$

The resulting photon energy correction factor leads to a good agreement of the mean value of the two photon invariant mass with an actual mass of π^0 at the low and medium photon energies but overestimates it at high energy by a few percent.

Fig. 4.12: Correction factors for the reconstructed photon energies vs the energy.

Yellow curve (left) indicates correction obtained from the step 1. Green curve (middle) is the result of step 2 of photon energy correction. Since the green and yellow corrections do not converge due to the finite bin size, we make an iterative guess for the correction in between the two, and obtain the red curve. Right plot illustrates corrected two photon invariant mass normalized to $m_{\pi^0} = 0.135$ as a function of photon energy.

4.4 Phase space and shape of MC events

In order to compute acceptance corrections, we first define a phase space where event selection takes place. The phase space of electrons is illustrated in Fig. 4.13 as a Q^2 versus ν distribution for generated and reconstructed MC as well as experimental data. The phase space allowed for the MC generation is larger than the one available in data in order to account for the possible bin migration effects. Following complete event generation, a kinematical restriction is applied on the generated set by imposing DIS cuts as defined in data: $1.0 < Q^2 < 4.1 \text{ GeV}^2$, $2.2 < \nu < 4.25 \text{ GeV}$, $W > 2 \text{ GeV}$ and $15 < \theta_e < 52^\circ$. Phase space of π^0 as defined in generated, reconstructed MC and data is illustrated in Fig. 4.14 on the example of z vs p_T^2 distributions. To visualize the order of magnitude of CLAS acceptance and the effect of detector resolution, we plotted generated and reconstructed distributions and compared them with experimental data using full statistical sample for electrons and π^0 .

While inclusive distribution of reconstructed MC events and real data match relatively well, π^0 distributions at high values of z and p_T^2 are poorly reproduced. A couple of approaches have been considered to fix this. The first one is to vary parameters of MC generation by experimenting with LEPTO, Pythia and JetSet switches, and study their effect on the shape of (z, p_T^2) distributions. To mention a few, relevant parameters to modify would be: choice of parton distributions functions (LST(15)); choice of fragmentation process (MSTJ(11) in Pythia); string

Fig. 4.13: Q^2 vs ν distributions for electrons for generated MC events (left), reconstructed MC events (middle) with all selection criteria applied and data (right). Phase space considered in this analysis is enclosed within the box. For generated events phase space allowed for generation is larger in order to include bin migration effects.

Fig. 4.14: z vs p_T^2 distributions for π^0 generated events (left), and π^0 candidate for reconstructed MC and data (middle, right) with their kinematical limits indicated in black solid lines.

tension (PARJ(33) from Pythia); the width of a Gaussian distribution for the primordial transverse momentum k_T of partons (PARL(3)); fragmentation width of the two-dimensional Gaussian p_x and p_y transverse momentum distribution for primary hadrons (PARJ(21) in Pythia), etc. While a large set of tunable parameters was tested, in general, neither of them or their combination lead to a significant changes in the shape of both z and p_T^2 distributions, it has merely affected the tails of p_T^2 distribution, which were eventually smeared by adding Fermi motion inside the LEPTO code.

The second approach would be to modify the shape of distributions post generation by applying a certain weight. To begin with, we will describe this procedure in general terms using an idealized example. At first, we randomly generate two sets of events for comparison: one according to the exponential distribution with slope equal to 5, and the other one just flat (plots 1a and 2a of Fig. 4.17). We further assume that the first set is a known cross section of the process, or a weight, while the second set is generated MC distribution for this process. If the cross section is known, flat MC distribution can utilize it as a weight, thus converging to a physical distribution (plot 3a of Fig. 4.17). Next, taking into account detector resolution on a given variable to be $dx/x = 0.5$, we smear generated distributions accordingly, so that they become 'reconstructed' events containing some effect of bin migration (plots 1b and 2b of Fig. 4.17). In that scenario, if we weight reconstructed MC events by, again, the known cross section, we will

Fig. 4.15: Q^2 , ν and θ distributions on the example of the iron target. Top panel compares generated (blue) and reconstructed electrons (red), bottom panel compares electrons from reconstructed (red) events and data (black). axes, indicated by blue, red and black colors correspond to the full statistics available on iron target for generated, reconstructed MC events and data, respectively.

obtain a new MC distribution which would be mathematically identical to the physical one after smearing (compare plots 3b with 1b of Fig. 4.17). In principle, the procedure works exactly and accounts properly for the bin migration effects provided the weight is calculated at the generated kinematics. In practice, the cross section of our process has not been yet measured, so we can not form the

Fig. 4.16: z , p_T^2 and $\phi_{l,h}$ distributions of π^0 's on the example of the iron target.

Generated π^0 's (blue) have been selected by their MC PID, while reconstructed π^0 candidates from MC (red) and data (black), were required to be in the range $0. < M_{\gamma\gamma} < 0.3$ GeV. The colors and axes are the same as in Fig.4.15.

ratio of generated MC distributions to generated distributions as they appear in nature. The approximate weight we can use would be that of the physical process after its passage through the detector (or measured data). The resulting function should be close enough to the true one such that a few iterations suffice for convergence (Fig. 4.18). Once the method passed analytical testing, we apply it on the actual MC distribution in order to correct reconstructed p_T^2 distribution.

Fig. 4.17: Illustration of weighting procedure. Plotted are: (1a) generated according to $e^{-5 \cdot x}$ function (an ansatz for the physical distribution further used as weight); (2a) generated flat (an ansatz for outcome of MC); (3a) events from (2a) weighted by (1a) ('corrected' generated MC); (1b) same as (1a), but smeared with a resolution $dx/x = 0.5$ ('reconstructed' physical events); (2b) same as (2a) with after smearing ('reconstructed' MC events); (3b) events from (2b) weighted by (1a) ('corrected reconstructed' MC); events (3a,b) are scaled to the sum of all weights for comparison. In reality, distribution (1a) is not known, hence the weight must be used on the reconstructed level.

Fig. 4.18: Illustration of weighting method performed at the level of reconstructed with resolution $dx/x = 0.25$ events. Left plot are the events generated exponentially and smeared by the resolution (weight 1). Middle plot are the events generated flat and smeared by the same resolution ('reconstructed' MC). Plot on the right contains two iterations of weighting 'reconstructed' MC events: (1) red and blue lines are the result of the first iteration (2) black line is the result of the second iteration, where weight 2=weight 1/(reconstructed MC, weight 1). Comparison of the distributions on the left and the right bears slight deviations, which could be further corrected by the next iteration(s) leading to a convergent result.

We calculate and parametrize the weight function, determined as a ratio of measured data to reconstructed MC evaluated at the generated p_T^2 , at each level of iteration. After multiple iterations, the resulting distribution, however, neither converges with distribution in data, nor changes in shape (Fig. 4.19). The underlying reason is most likely related to a poor EC resolution in reconstructing π^0 's p_T^2 . This statement was verified by conducting the following test using MC

Fig. 4.19: p_T^2 distributions for measured data (black), reconstructed MC (red) and weighted MC in different levels of iteration (blue scale). After the 4th iteration of weighting reconstructed p_T^2 using parametrization in a power law $\alpha=1.5$ result still does not converge.

reconstructed events. There, we have selected a sample of π^0 candidates for which p_T^2 is reconstructed with resolution of 3σ or less from the mean value, and compared it with the sample outside of this resolution. As illustrated in Fig. 4.20, following the first iteration, weighted p_T^2 distribution, corresponding to a 'good' resolution, converges fast to the shape of distribution in data. This positive trend is contrary to the poor convergence of the p_T^2 distribution for which the range of resolution is outside 3σ . Therefore, we refrain from attempting to use this method for correcting the shape of MC events. One can certainly think of other alternatives, for example, convoluting p_T^2 distribution from data with detector acceptance and resolution in order to extract p_T^2 shape as occurs in nature

Fig. 4.20: p_T^2 distributions for measured data (black), weighted generated MC (blue) and reconstructed MC (red). Left plot corresponds to the sample of π^0 candidates for which p_T^2 is reconstructed with resolution of 3σ or less from the mean value, while the right plot corresponds to the sample outside of this resolution. After the first iteration the sample on the left almost converges with data, whereas the sample on the right does not. This is the main reason we can not apply event weighting procedure to fix the shape of our π^0 candidates.

before the detector smearing. However, those procedures are complex and require significant time investment. It is not in the scope of this work to extract cross sections or transverse momentum broadening, but ratio of yields from two targets. For that reason, we have plotted iron to deuterium ratios of the reconstructed MC distributions and compared them to the same quantity in data (Fig. 4.21). Ratios in p_T^2 are close to unity for MC events, however, in data they show a Cronin-like effect (enhancement at high p_T^2). The effect of Fermi motion in deuterium and iron

Fig. 4.21: Fe/D ratios for reconstructed MC (purple) and data (black), with data points not corrected for acceptance. Normalization factor of MC ratios is the number of events in the data sample. Top panel ratios are plotted in Q^2 , ν , and θ ; bottom panel are z , p_T^2 , and $\phi_{(l,h)}$ for π^0 candidates in the range $0 < M_{\gamma\gamma} < 0.3$ GeV.

nuclei contributes to the tail of p_T^2 at the high end, after reconstruction procedure this effect is smeared out (Fig. 4.22). MC ratios of z distributions are close to unity and, evidently, do not exhibit attenuation properties as the ones observed in data. Therefore, since we observe no difference in the shapes of p_T^2 and z ratios in MC, Given the variance in the shape of MC events and data, we will further study and attribute a systematic error on our result, reflecting this discrepancy.

Fig. 4.22: p_T^2 distributions for MC generated events (left) and reconstructed events (right) for D(black) and Fe(blue).

4.4.1 Resolution

Particle resolution is the deviation of reconstructed momentum and angle from their true values at the interaction vertex. Tracking uncertainties for the charged particles stem from scattering in the detector material, misalignment of detector components and lack of precise knowledge of the magnetic field, as well as single wire resolution. These effects must be considered as they may introduce systematic fluctuations on the measured quantity, in particular, if it is measured in bins whose width is fine compared with the resolution on a given variable. Below studies of the detector resolution are aimed at verifying that the binning scheme used for our analysis is viable.

Electron momentum resolution σ_p/p varies from 0.5% for $\theta < 30^\circ$ to 1 – 2% otherwise; polar angle resolution $\sigma_\theta \approx 1(\text{mrad})$ [57]. Below we estimate Q^2

resolution for the two realistic cases. In the first instance, we consider the low Q^2 region with high momenta electron $P_e=2.6$ GeV scattered at a low angle $\theta_e=18(^{\circ})$. This yields $\delta Q^2/Q^2 = 0.008/1.272 \approx 0.6\%$. Second, for the high Q^2 region we choose low momenta electron $P_e=1(\text{GeV})$ scattered at a large angle $\theta_e=40(^{\circ})$, which in turn leads to the resolution $\delta Q^2/Q^2 = 0.047/2.339 \approx 2\%$. Meanwhile, resolution on ν depends only on the scattered electron momentum, and hence, is within the 2% range from its value. Therefore, our choice of Q^2 and ν bin widths is significantly larger than their resolution.

Resolution of the neutral particle detection in the EC depends on the total energy deposited in the scintillator, which is expressed as a fraction of the incident particle energy (sampling fraction), as well as on the thickness of the material, measured in radiation length. The energy resolution for the electrons can be parametrized as follows:

$$\frac{\sigma}{E} = \frac{10.3\%}{\sqrt{(E)(\text{GeV})}} \quad (4.4.1)$$

While single particle resolution improves with increasing the amount of deposited energy, the main challenge in resolving two neighboring photon clusters from a high energy decay of π^0 is to divide them into the energy strips. The current EC cluster reconstruction algorithm has an efficiency up to 90% for the $P_{\pi^0} > 2.5$ GeV [81]. Nevertheless, since distributions in z and p_T^2 involve a combination of two photons, whose individual energy can vary from the lowest to the highest available, detailed understanding of π^0 resolution becomes impor-

tant. Using MC simulation, we compare the choice of the bin width in z and p_T^2 versus their detector resolution, which we define as a difference between their reconstructed value and generated one on the event by event basis:

$$res(x) = x_i^{gen} - x_i^{rec} \quad (4.4.2)$$

where x indicates the bin in either z or p_T^2 variable and x_i is z or p_T^2 value within the bin range. In order to construct x_i^{rec} , we select reconstructed π^0 candidates such that each of its photons are matched to the photons from generated π^0 by fixing an angle: $\alpha(\gamma_{rec}, \gamma_{gen}) < 3^\circ$. The distribution of $res(x)$ is then fitted with

Fig. 4.23: Resolution defined as $res(x) = x_i^{gen} - x_i^{rec}$ is plotted in each bin in z and fitted with a Gaussian function within 1σ .

Fig. 4.24: Top panel is resolution in z , bottom in p_T^2 . From left to right, the y-values are respectively: the width of resolution σ extracted from a Gaussian fit; relative resolution σ normalized to the mean value of z or p_T^2 in each bin; ratio of the bin width in z or p_T^2 normalized to the resolution in a given bin.

a Gaussian function as presented in Fig. 4.23. In each bin $res(x)$, we extract the mean value of the resolution and compare it with the width of the bin (Fig. 4.24). From these plots we observe that the width in some of the bins is at the limit of the resolution. However, overall, the width is never less than 1.8σ , therefore, we conclude that the choice of the bin width is acceptable.

The above examination of detector resolution are rather simplified estimates. In reality, electron and π^0 resolutions have the level of entanglement which depends on the dimensions of the analysis. To understand better the effect of the finite

bin size and its correlation with the resolution in the multidimensional variables we further look at the bin purity.

4.4.2 Purity

Purity is one of the measures of the bin migration effect, arising dominantly from the resolution. Bin migration refers to a situation when an event is generated in one bin and reconstructed in another. Purity of the bin is studied using MC events, and is typically defined as:

$$Purity = \frac{N_{gen}^{rec}}{N^{rec}} \quad (4.4.3)$$

where N^{rec} is the total number of events reconstructed in a given bin, and N_{gen}^{rec} is the number of the events that were both reconstructed and generated in that bin. In practice, we construct purity using a transfer matrix of the number of generated and reconstructed events in each bin of the (Q^2, ν, z, p_T^2) grid. To calculate purity in the above 4D bins, we require that in the numerator the generated bin is equal to the reconstructed bin of the denominator in all four variables simultaneously (resulting 4D purities are illustrated in Fig. 4.25). The reconstructed π^0 candidates are selected using an angular matching criteria with generated events. Starting with the 4D grid of generated and reconstructs events per bin, there are two ways to go down in the number of dimensions in which purity is measured.

Considering an example of 3D binning in (Q^2, ν, z) , one can either sum already calculated 4D purities in (Q^2, ν, z, p_T^2) over p_T^2 , or construct a new purity

Fig. 4.25: Illustration of 4D purities (Q^2, ν, z, p_T^2) plotted as a function of p_T^2 for the deuterium target. Each cell corresponds to (Q^2, ν) bins with z bins indicated in color code in the top right of the first cell.

in (Q^2, ν, z) grid requiring that the reconstructed π^0 kinematics match generated kinematics in only (Q^2, ν, z) bins. The latter condition omits the requirement on the purity in p_T^2 bin by simply integrating over this variable. Integration of multidimensional purity down to (Q^2, ν, z) gives values of purity that are less by 10% on average per bin as compared to a purity calculated directly in 3D bins by integrating over all p_T^2 values. This difference is caused by the poor resolution in p_T^2 . If instead we choose a different variable of integration, for example Q^2 , the difference between the two methods would be negligible due to a fine resolu-

Fig. 4.26: Illustration of 3D Purities in (ν, z, p_T^2) bins plotted as a function p_T^2 in bins of incrementing ν (each histogram) and z (color code) on the example of the deuterium target. The purity is calculated by requiring that in a given set of 3D bins reconstructed π^0 kinematics match kinematics of the generated π^0 . The bin for which *purity* < 30 % ($0.3 < z < 0.4$, $2.2 < \nu < 3.2$, $0.75 < p_T^2 < 0.9$) is further excluded from the analysis.

Fig. 4.27: 3D Purities in (Q^2, ν, z) bins plotted as a function of z in bins of incrementing Q^2 (each histogram) and ν (color code) on the example of the deuterium target.

tion in Q^2 . For illustration of the two methods refer to [87]. Since the analysis, both count extraction and acceptance calculations, is carried out on the 3D grid with one variable integrated, we rely on the purity exclusion based on the second method.

Purity is not used directly but rather as a 'control distribution'. Assuming a flat cross section with perfect efficiency in 1D bin, which has a minimum resolution of 1σ , would yield $purity=68\%$. From this assumption we can deduct a rule of thumb asking for $purity > (0.68)^n$ for a n-dimensional binning [85]. Therefore, for n=4 dimensional case bins with $purity < 20\%$ should be excluded, whereas for n=3 case we exclude bins with $purity < 30\%$ as illustrated in Fig. 4.26-4.27. Purities of each bin in (Q^2, ν, z) exceed 50% and show little dependence on the target type. Meanwhile, overall purities in (ν, z, p_T^2) are significantly lower, in particular for the deuterium target. Yet, only one bin with $purity < 30\%$ exists at $0.3 < z < 0.4$, $2.2 < \nu < 3.2$, $0.75 < p_T^2 < 0.9$, and thus it will be excluded from our final results.

4.5 Acceptance

The CLAS detector is a complex system, and its acceptance correction encompasses the geometrical acceptance, detection and track reconstruction efficiencies. The CLAS polar and azimuthal angular coverage is less than 4π and the distance to the CLAS center $z=0$ from the solid target is 5 times more than the separation distance of the two targets. The above conditions are sufficient to create an imbalance in the particle acceptance of one target with respect to the other, especially when the scattered particles emerge on the edges of CLAS acceptance. Moreover, a small fraction of events, in particular from the downstream deuterium target, may pass through the support structures of the upstream solid target, scatter and

still emerge within CLAS acceptance with modified values of energy and angle. For that reason, despite the fact that the two targets were exposed in the beam simultaneously and were separated only by 4 cm along the beam axis, acceptance factors do not cancel out in the ratio of number of events emerging from solid target normalized to that in deuterium.

Considering that the difference in the geometry of target holders and maps of DC efficiencies between the three solid targets is small, only one of them, namely iron, has been simulated. Putting available computational resources on simulating one target is a compromise in favor of extracting acceptance correction factors in multidimensional binning as opposed to the acceptance calculations performed in a smaller number of bins. Approximately 355 M generated events for each D and Fe targets, have been input to GSIM. Each of these events contains at least one π^0 or η meson. Since the phase space available for the MC generation has been chosen larger than the phase space selected in data, the actual number of the 'good' generated triggers that satisfied our DIS cuts came down to 206M (D) and 234M (Fe). The number of 'good' reconstructed triggers, selected as described in section 4.3.1, is 66 M (D) and 77 M (Fe), which is almost 3 times greater compared to the same numbers from data. The total generated number of π^0 's, given a 'good' generated trigger, for D and Fe targets was 290 M and 258 M, respectively. The number of reconstructed π^0 candidates for which $0. < M_{\gamma\gamma} < 0.3$ GeV, provided a 'good' reconstructed electron has been detected, is 10.9 M (D) and 9.2 M (Fe).

This exceeds the number of π^0 candidates in the data by almost 13 and 19 times for the D and Fe targets, correspondingly. Despite the fact that the difference in the number of reconstructed triggers in data and MC is relatively small, a significant difference in the number of π^0 candidates is due to the fact that every single generated event in MC was selected with at least one π^0 . Simulation were carried out for the deuterium and iron targets only, omitting the carbon and lead. The reason is related to the fact that all four solid targets were positions at the same distance with respect to the CLAS center, and the maps of the DC inefficiencies were very similar. Thus, instead of running simulations for 4 solid targets, it was possible to accumulate large statistics for one.

As a general notation, let us assume that the data under consideration is divided in multidimensional bins $\vec{x}$, and define the acceptance correction factor as follows:

$$A(\vec{x}) = \frac{N_{rec}(\vec{x})}{N_{gen}(\vec{x})} \quad (4.5.1)$$

where $N_{gen}(\vec{x})$ and $N_{rec}(\vec{x})$ are the number of events generated and reconstructed in bin $\vec{x}$, respectively. Since MC simulations are performed with a finite number of events, one has to account for the statistical fluctuations in the correction factor. Calculation of statistical uncertainty strongly depends on the relative amount of the bin migration effects. In a scenario where the size of the bin is much larger compared to detector resolution, a good estimation of statistical uncertainties would be given by a binomial distribution. If, however, data is divided into smaller

bins such that the fraction of events generated in one bin and reconstructed in another is large, the probability to find a reconstructed event will have more than 2 outcomes. Therefore, a Poisson distribution will give a better estimate of the uncertainty [84].

$$\delta A(\vec{x}) = \sqrt{\frac{A(\vec{x})(1 + A(\vec{x}))}{N_{gen}}} \quad (4.5.2)$$

Expanding the above formula will give a straightforward way to calculate the acceptance error using the number of generated N_{gen} and reconstructed N_{rec} events:

$$\frac{\delta A(\vec{x})^2}{A(\vec{x})^2} = \frac{1}{A(\vec{x})} \frac{(1 + A(\vec{x}))}{N_{gen}} = \left(\frac{\delta N_{rec}(\vec{x})}{N_{rec}(\vec{x})} \right)^2 + \left(\frac{\delta N_{gen}(\vec{x})}{N_{gen}(\vec{x})} \right)^2 \quad (4.5.3)$$

Finally, the calculation of the statistical error on the number of acceptance corrected events includes statistical error on the number of events before correction (δN) and statistical error on the acceptance itself (δA):

$$\delta N_{corr} = N_{corr} \cdot \sqrt{\left(\frac{\delta N}{N} \right)^2 + \left(\frac{\delta A}{A} \right)^2} \quad (4.5.4)$$

4.5.1 Electron acceptance

The acceptance correction factor $A(\vec{x})$ for the electron arm is evaluated in $\vec{x} = \{Q^2, \nu\}$ bins and is applied on a bin-by-bin basis. Given the large set of MC events, the number of bins in each of the inclusive variables is not limited by statistics, but rather by detector resolution. Therefore, the number and the width of the bins in which correction is calculated can be optimized to describe

structures in acceptance. For that purpose, three sets of bins in $(Q^2, \nu)=(3, 3)$, $(6, 6)$, $(9, 9)$ with variable width were considered.

(Q^2, ν)	(6,6)
Q^2	(1.0, 1.17, 1.33, 1.55, 1.76, 2.7, 4.1)
ν	(2.2, 2.9, 3.2, 3.46, 3.73, 3.99, 4.25)

Table 4.1: Set of bins in $(Q^2, \nu) = (6, 6)$ used to calculate acceptance factors and correct electron yields.

Before applying acceptance corrections, it is important to consider and treat, if necessary, the bins which may feature the following conditions:

- large statistical error $\delta A/A$
- zero acceptance

In each set of bins, we have compared acceptances and their relative errors as well as fluctuations from bin to bin of acceptance corrected electron yields in the ratio of D to Fe targets [88]. Given the three sets of acceptance factors, the effect of the bin size in which electrons are corrected was found to be less than 1% for the ratio of D to Fe yields. Finally, the set in $(Q^2, \nu)=(6,6)$ bins was chosen to correct electrons in the measured data.

Electron acceptance (Fig. 4.28) varies from 6 % in the lowest (Q^2, ν) bins up to almost 60 % at the highest Q^2 and medium ν . As follows from the example of the iron target, the sizes of the bins were well tuned to avoid large statistical

Fig. 4.28: Electron acceptance in $(Q^2, \nu)=(6,6)$ bins (left) and associated relative error (right) on the example of a solid target.

fluctuations: average statistical error on electron acceptance factor is on average 0.1 % with one exception, at the edge of phase-space reaching up to 0.6 %. There is only one bin with zero acceptance, in which no event was generated and none were reconstructed. The content of the same bin was checked in the measured data, and confirmed that zero acceptance bin corresponds to the same empty bin in data. For the case when electrons are corrected in $(Q^2, \nu)=(6,6)$ bins and summed down to $(Q^2, \nu)=(3,3)$, as binned in a multiplicity ratio, the overall magnitude of electron correction on the ratio of liquid to solid targets is the largest for lowest bin in (Q^2, ν) and reaches up to 6% (Fig. 4.29).

Statistical error on the number of acceptance corrected electrons, plotted in Fig. 4.30 in $(Q^2, \nu)=(3,3)$ bins, was obtained by adding in quadratures statistical errors calculated in $(Q^2, \nu)=(6, 6)$ bins according to equation 4.5.4.

Fig. 4.29: The ratio of number of the electrons produced on a liquid target to that on deuterium: black points correspond to the uncorrected yields, while red ones were corrected using acceptance in $(Q^2, \nu)=(6,6)$ bins. From left to right are the ratios plotted for carbon, iron and lead targets. Note, that for the ease of comparing the acceptance factors between the three targets, the ratio of electron yields are normalized by the factor of deuterium to solid targets thickness (section 2.3).

Fig. 4.30: Relative statistical error on the number of electrons before (left) and after (right) acceptance correction. The contribution from acceptance to the total statistical error is minimal.

4.5.2 $e\pi^0$ acceptance

Acceptance correction factors for semi-inclusive π^0 were evaluated provided an event contains 'good' generated and reconstructed electron. The equivalent two sets of bins, in which the multiplicities are extracted, were used for calculations of acceptance factors: $(Q^2, \nu, z)=(3, 3, 5)$ and $(\nu, z, p_T^2)=(3, 5, 6)$. Generated numbers of π^0 were defined using a MC particle ID. Reconstructed π^0 events were determined using the same method as applied in data by fitting invariant mass of the two photons with a Gaussian plus a background function. Mixed background shape per each bin was calculated based on the reconstructed photons from uncorrelated events combined in pairs as described in section 3.3.3. Each of the generated and reconstructed photons from π^0 was constrained by the same conditions as those imposed on a photon from a π^0 candidate in data (energy and photon angle with electron).

First, consider results in a set in (Q^2, ν, z) bins integrated over p_T^2 . Each of the 54 bins, filled with reconstructed invariant mass, has large available statistics and is never empty, therefore there is no concern about zero acceptance bins or large relative errors on the acceptance. As illustrated in Fig. 4.31, the magnitude of acceptance varies from a fraction of percent up to 9%, and the relative statistical error on the acceptance itself is between 0.4% and 1.2%. The overall correction factors for the multiplicity ratios binned in (Q^2, ν, z) due to acceptance correction of hadronic arm are within 15%. Fits to the reconstructed invariant masses in

Fig. 4.31: $e\pi^0$ acceptance vs z in bins of ν (indicated by color) and Q^2 (each histogram) illustrated for the example of iron. Limits of integration over p_T^2 are the same as in data where maximum values reaches to $p_T^2=0.9 \text{ GeV}^2$.

(Q^2, ν, z) bins and comparison of acceptances for each target are reflected in [89].

Second, we consider acceptances calculated in (ν, z, p_T^2) bins by integration over Q^2 . Each of the 108 bins contains sufficient statistics to perform invariant mass fits to reconstructed events as reflected in Fig. 4.32. Acceptance factors, plotted in Fig. 4.33, vary from a fraction of percent up to 11 % on average with statistical errors varying from 0.4 % to 1.5 % on average reaching up to 3 % in a single bin. In case of (ν, z, p_T^2) bins, the overall correction factor is within 20 %, however, a number of fluctuations are present.

Fig. 4.32: Slice in the reconstructed invariant mass in (ν, z, p_T^2) bins on the example of a solid target. Fit to the invariant mass is comprised from the Gaussian plus polynomial functions, where the shape of the latter is defined using a mixed background technique based upon reconstructed photons.

Fig. 4.33: $e\pi^0$ acceptance plotted as a function of p_T^2 in bins of z (indicated by color) and ν (vertical binning) integrated over Q^2 . Based on our studies of purities, we exclude one bin for which *purity* < 30 % for either target: $0.3 < z < 0.4$, $2.2 < \nu < 3.2$, and $0.75 < p_T^2 < 0.9$.

Fig. 4.34: Statistical errors in the ratio of a solid target to deuterium associated with acceptance in (Q^2, ν, z) bins (left) and in (ν, z, p_T^2) bins (right).

Chapter 5

Radiative corrections

Electromagnetic photon radiation affects incident or scattered electron such that its energy and scattering angle at the vertex can be different than those measured in the detector. The purpose of radiative corrections (RC) is to account for processes other than the one-photon exchange (Born cross section) which lead to the modification of the kinematical variables due to photon emission. This modification affects both leptonic and hadronic components in the multiplicity ratio. Due to nuclear effects, different in solid and deuterium targets, as well as the difference in radiation lengths of each target, radiative effects will not cancel in the ratio. Radiative effects in the inclusive channel were calculated using KeppelRad code. The ratio of radiative corrections on deuterium to solid targets due to lepto radiation were found to reach 3%. The SIDIS radiative corrections generally consist of internal and external radiation. The energy loss due to external radiation, i.e emission of bremsstrahlung photon after interaction in the target chambers and walls as well as in subsequent components of detector, were taken into account by GEANT simulations, and enter in the acceptance

correction factors. The semi-inclusive internal radiation, i.e radiation of a hard photon in the field of nucleus, was calculated using HAPRAD code, which also accounts for the effects related to the exclusive π^0 contribution. The ratio of solid to deuterium targets of total radiative corrections (RC) to semi-inclusive π^0 were found to vary from unity up to 20% in selected bins.

5.1 Inclusive radiative corrections

The radiative correction for the inclusive part has been calculated using Keppel-Rad code. The radiation is treated via the Mo and Tsai approach [99]. An exact formula for computing external photon radiation (straggling effect) is used, while internal bremsstrahlung radiation is calculated via equivalent radiator approximation given the known nuclear target thicknesses in radiative length (0.0026, 0.009, 0.023, 0.025 r.l. correspondingly for D, C, Fe, and Pb).

The model for the inclusive cross section was developed on the range $1 < W < 2$ GeV, $0.5 < Q^2 < 10.0$ GeV² and links smoothly to a reliable global fit to all SLAC deep inelastic data in the range $0.1 < x < 0.9$ and $0.6 < Q^2 < 20$ GeV². The radiated and Born cross sections were calculated in the range of input kinematics in $1. < Q^2 < 4.1$ GeV² and $0.75 < E' < 2.9$ GeV. The radiative correction (RC) factors were obtained by taking the ratio in the bins of interest: $(Q^2, \nu) = (3, 3)$. The magnitude of radiative corrections varies from 8% at lowest Q^2 and highest ν up to 27% at high virtualities and low energy transferred. The ratio of RC on

Fig. 5.1: Radiative correction factor (top) and the ratios (bottom) of solid targets to deuterium for the inclusive part of multiplicity ratio calculated in (Q^2, ν) bins.

deuterium to solid target, which enters as inverse factor in the multiplicity ratios, varies at maximum from 3% above unity to 3% below unity (Fig.5.1).

5.2 SIDIS radiative corrections with HAPRAD.

In handling radiative correction two codes employing different approaches have been considered. One of the approaches is based on exact calculations of electroproduction processes obtained in terms of convolution of known leptonic tensors (with and without an additional radiated photon) and hadronic tensors (Bardin and Shumeiko approach) which relies on the model for structure functions [100]. Another approach is based on MC simulations of external and internal

radiations using both angle-peaking and the equivalent radiator approximations (Mo and Tsai approach) [101]. Associated code, SEMIRC, has been developed by Peter Bosted [102]. To correct our data, we use HAPRAD_cpp code [103], which is an adaptation to C++ of the original HAPRAD code written in FORTRAN; its ideology is based on the first approach employing exact calculations of internal radiation only.

Here, only lowest order QED correction are considered: real photon emission from the lepton leg as well as the additional virtual photon between the initial and final leptons and the correction due to virtual photon self-energy (Fig. 5.2). The total radiative correction at the lowest order is obtained as the sum of the above contributions and effects of vacuum polarization all normalized to Born cross section: $RC = \frac{\sigma_{obs}}{\sigma_{Born}}$. The original HAPRAD code is designed to calculate radiative corrections to unpolarized cross section of semi-inclusive electroproduction off a proton target; the next version, HAPRAD2 [104] includes additional contribution of π^+ exclusive radiative tail. The version of HAPRAD2 code tailored to extract radiative corrections for π^0 includes modification of structure functions due to nuclear effects, as well as exclusive π^0 contribution to radiative tails. To cross check that the two versions, the original FORTRAN code HAPRAD2 and its C++ version, are consistent, the RC calculated with default structure functions were compare, and found to agree within 2% [105].

5.2.1 Nuclear structure functions

Since the original version of HAPRAD code is aimed at extracting radiative corrections on the proton, a modification must be introduced in order to account for the nuclear effects on the targets under consideration: D, C, Fe, and Pb. The SIDIS cross section within the leading twist formalism depends on four hadronic structure functions: $\sigma_{SIDIS} = \sigma_{SIDIS}(\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3, \mathcal{H}_4)$ (exact expression is given in 5.2.5). The structure functions enter in σ_{SIDIS} each with their kinematical factors calculated based on leptonic kinematics. The nuclear effects are accounted for by replacing default proton structure functions for the nuclear structure functions in the expression for SIDIS. Within the code notation, the $\mathcal{H}_3$ and $\mathcal{H}_4$ correspond

Fig. 5.2: Feynman diagrams contributing to the Born cross section (a) and the lowest order QED effects (b-e).

to the cosine moments of cross section, while $\mathcal{H}_1$ and $\mathcal{H}_2$ are interrelated by the hadron kinematical factor. The model for structure functions is constructed based on the convolution of parton distribution function (f_q), fragmentation function

(D_q) , and transverse momentum distribution ($\mathcal{G}$) as prescribed in [106].

$$\mathcal{H}_1 = \sum_q e^2 f_q D_q \mathcal{G} \quad (5.2.1a)$$

$$\mathcal{H}_2 \approx \mathcal{H}_1 \quad (5.2.1b)$$

$$\mathcal{H}_3 = f(x, Q^2, z)|_{\cos(\phi)} \sum_q e^2 f_q D_q \mathcal{G} \quad (5.2.1c)$$

$$\mathcal{H}_4 = f(x, Q^2, z)|_{\cos(2\phi)} \sum_q e^2 f_q D_q \mathcal{G} \quad (5.2.1d)$$

where $f(x, Q^2, z)$ within code notation corresponds to cosine moments of cross section fitted with $A + B \cdot \cos(\phi) + C \cdot \cos(2\phi)$ function and $\mathcal{G}$ is transverse momentum distribution:

$$\mathcal{G} = \frac{1}{2\pi\sigma} \cdot \exp - \frac{(p_T - \mu)^2}{2\sigma^2} \quad (5.2.2)$$

The expression for $\mathcal{G}$ is obtained through a multidimensional fit to data. Here the μ and σ are the mean and the width of the Gaussian distribution in p_T which was fitted simultaneously in (x, z) bins. The fit parameters bear the following dependence on (z, x) :

$$\mu = p_0 + p_1 \cdot z + p_2 \cdot z^2 + p_3 \cdot x + p_4 \cdot x^2 \quad (5.2.3a)$$

$$\sigma = A + B \cdot z + C \cdot z^2 + D \cdot x + E \cdot x^2 \quad (5.2.3b)$$

Each set of parameters, $(p_0, p_1, p_2, p_3, p_4)$ and (A, B, C, D, E) , was obtained from the data individually for each target. The entire data set was divided and acceptance corrected in $(x, z, p_T) = (5, 5, 16)$ bins. The fitting procedure on the

three-dimensional set of bins, omitting those with few statistics, was performed in three steps. First, three-dimensional distributions in (x, z, p_T) (Fig. 5.3) were projected onto one-dimensional distributions in p_T in each bin in z and x , and then fitted with a Gaussian function. Second, the values of μ and σ of the Gaussian fit from the first step were plotted for each value of x and fitted with a polynomial function in order to obtain the x -dependent expressions for $\mu(x)$ and $\sigma(x)$. Similarly, dependencies of Gaussian parameters μ and σ were obtained as a function of z . Finally, given the above ansatz on the shape of two-dimensional distributions $p_T(x)$ and $p_T(z)$, the three-dimensional distributions were fitted simultaneously as $p_T(x, z)$. The three-step fitting routine is performed iteratively and separately for each target. The quality of the resulting three-dimensional fit is illustrated using p_T distribution in the slice of (x, z) bins in Fig. 5.4 on the example of the iron target.

Next, we examine ϕ distributions in the data which will be used to construct $\mathcal{H}_3$ and $\mathcal{H}_4$ structure functions. For that purpose, we start by dividing and correcting for acceptance our data set in $(x, Q^2, z, \phi) = (2, 2, 3, 24)$ bins. The initial guess of the fitting function to ϕ distribution in each of (x, Q^2, z) bins was a convectional expression: $f(\phi) = A + B\cos(\phi) + C\cos(2\phi)$. The fit parameters to ϕ distributions have shown to have a small dependence on x and Q^2 , thus it became possible to avoid a multistep fitting procedure by rearranging coefficients such that they only depend on z . The next guess is to fit ϕ dependence with

Fig. 5.3: Three-dimensional distribution of π^0 candidates in (x, z, p_T) bins on the example of the iron target (data).

$f(\phi) = A (1 + B'(z)\cos(\phi) + C'\cos(2\phi))$, where $B'(z)$ and C' have the following parametrization over the entire kinematics:

$$B'(z) = a + b \cdot z + c \cdot z^2 \quad (5.2.4a)$$

$$C' = \text{const} \quad (5.2.4b)$$

Here, (a, b, c, const) are the numerical coefficients from the fit individually for each target. Consequently, $B'(z)$ and C' enter in place of the $f(x, Q^2, z)|_{\cos(\phi)}$ and $f(x, Q^2, z)|_{\cos(2\phi)}$ in the $\mathcal{H}_3$ and $\mathcal{H}_4$ structure functions (5.2.1d). The overall normalization factor A up to a kinematical constant is in fact proportional to the sum of $\mathcal{H}_1 + \mathcal{H}_2$. In addition, each of the four SIDIS structure functions has a common factor corresponding to the pion threshold. Its purpose is to constrain the

available phase-space in order to exclude possible contributions from the exclusive range.

In the following sections we illustrate the input cross sections in the HAPRAD code before radiation, and compare them relative to one another. Note that the architecture of the HAPRAD code is such that cross section calculations are done at a fixed kinematical point for which input values in $(Q^2, \nu, z, p_\perp, \phi)$ are required (equivalent up to a beam energy and proton mass to the actual code calculation in $(x, y, z, p_\perp, \phi)$). Thus, no *a priori* integration over one of the variables could be done.

Fig. 5.4: A slice of acceptance corrected p_T distributions (red points) in (x, z) bins of π^0 candidates (iron target). The p_T distribution is in the range 0-1.5 GeV. The blue lines illustrate the goodness of the Gaussian fit of p_T obtained as a simultaneous function of (x, z) . This parametrization is used as an input in nuclear structures functions, $\mathcal{H}_1$ and $\mathcal{H}_2$, in radiative correction code.

Fig. 5.5: A slice of ϕ distributions in the range $(0, 180)^\circ$ of π^0 candidates without (black) and with (red) acceptance correction in (x, Q^2, z) bins (iron target). In each bin, ϕ distributions were initially fitted with $A+B\cos(\phi)+C\cos(2\phi)$ function. The fit showed that A, B and C depend mainly on z , thus, to simplify extraction of the function that fits cosine moments over all kinematical range, the fit was rearranged as: $f(\phi)=A (1 + B'\cos(\phi) + C'\cos(2\phi))$ where $B'=B(z)$ and C' is constant. Red fit is $f'(\phi)$ fitted with one parameter fixed (B'), the green fit has both parameters (B' and C') fixed.

5.2.2 Input cross sections: SIDIS

The Born cross section as implemented in HAPRAD code, is:

$$\sigma_{SIDIS} = \frac{d\sigma}{dx dy dz dp_{\perp}^2 d\phi} = \frac{N}{Q^4} (A + A^c \cos\phi + A^{cc} \cos 2\phi) \quad (5.2.5)$$

where N is the factor related to lepton kinematics, and A depends on four structure functions ($\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3, \mathcal{H}_4$). The unit of the SIDIS cross section on the exit of HAPRAD is [nbarn].

In this section we compare Born cross sections integrated over ϕ and $p_{\perp}^2$ for the case of nuclear targets, using two approaches: explicit integration of the HAPRAD output versus theoretical integration. This comparison allows us to verify that full HAPRAD calculations, in which nuclear structure functions enter in the expression of cross section with their own normalizations, are compatible with straightforward calculations depending only on $\mathcal{H}_1$ up to a leptonic kinematical factor. When integrated over the kinematical variables ϕ and $p_{\perp}^2$, the cross section, according to equation 5.2.5, coincides with well know formula for unpolarized SIDIS cross section ([106]) calculated within the Quark Parton Model:

$$\sigma_{SIDIS}^{xyz} = \frac{d\sigma}{dx dy dz} = \frac{2\pi\alpha^2}{yQ^2} (y^2 + 2 - 2y) \sum_q e^2 f_q D_q \quad (5.2.6)$$

Next, consider numerical integration of fully differential cross section at the exit of the HAPRAD code. Integrating it over ϕ , according to equation 5.2.5, resets the ϕ -dependent terms leaving only $\sigma_{SIDIS} = A(\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3, \mathcal{H}_4)$. Providing

Fig. 5.6: The Born cross sections $\sigma_{SIDIS}^{xyz} = \sigma_T + \epsilon \sigma_L$ at fixed (Q^2, W) for deuterium calculated in two approaches: theoretical (black) and numerical (red). The black curve, integrated over $p_\perp^2$ and normalized to 2π , obtained following the formalism of equation 5.2.6. The red curve corresponds to the cross section integrated post full differential HAPRAD calculation. Comparison verifies that full HAPRAD calculations, in which nuclear structure functions enter with their own normalizations, are compatible with straightforward calculations depending only on $\mathcal{H}_1$ up to a leptonic kinematical factor.

that $H_n \approx e^{-p_\perp^2}$, the $p_\perp$ dependence of the cross section can be factorized out as $\sigma_{SIDIS} = f(x, y, z) \cdot e^{-p_\perp^2}$. This allows for numerical calculation of the integral over $p_\perp$. Since cross section calculations in HAPRAD code require for one of

the inputs a fixed value in $p_{\perp}=p_{\perp}^0$, the numerical integration over $p_{\perp}$ is done post calculation: the cross section evaluated at $p_{\perp}^0$ is multiplied by the value of integral, calculated on the range $p_{\perp}=0-10$ GeV, and normalized to the value of that same integral calculated at the point $p_{\perp}^0$. Thus, the resulting SIDIS cross section will be reduced to $\sigma_{SIDIS}^{xyz}=\frac{d\sigma}{dx dy dz}$, and could be compared with equation 5.2.6.

Comparison between the two methods of integrating cross section down to σ_{SIDIS}^{xyz} is given at fixed Q^2 and W in Fig. 5.6. The two curves, with the black one based on (5.2.6) and the red one based on numerical integration, agree to a good degree starting at $z>0.2$. As the fragmentation function $D(z) \approx \frac{1}{z}(1-z)^n$, at low z it is expected to rise steeply. For high values of z , at given kinematics between $0.88 < z < 0.94$, we observe a gap related to the transition from the exclusive regime ($z=1$) down to the semi-inclusive in which at least one additional pion is produced. The part of the SIDIS cross section at exclusive range does not in fact enter into the calculations of radiative corrections. When the radiated SIDIS cross section is evaluated, a cut on the missing mass $M_x > M_p + M_{\pi}$ is applied by default in HAPRAD calculations. This cut ensured that only the SIDIS contribution enters in SIDIS radiative tails as illustrated further in Fig. 5.8.

5.2.3 Input cross sections: exclusive

The exclusive cross section, before radiation, in the CM-system is a differential over the energy of scattered lepton (dE'), as well as its solid angle ($d\Omega$) along with

that of the hadron in the CM ($d\Omega_h^*$) [107]:

$$\sigma_{EXCL} = \frac{d\sigma}{dE'd\Omega d\Omega_h^*} = \Gamma \frac{d\sigma}{d\Omega_h^*} \quad (5.2.7)$$

$$\Gamma = \frac{\alpha}{2\pi^2} \frac{E'}{E} \frac{k_\gamma}{Q^2} \frac{1}{1 - \epsilon}$$

where Γ is the flux of the virtual photon field in the Hand convention [108]; $k_\gamma = \frac{W^2 - M^2}{2M}$ is the 'photon equivalent energy' to excite the hadronic system with energy W ; and polarization $\epsilon = 1/(1 + 2 * (1 + \nu^2/Q^2 \tan^2 \frac{\theta}{2}))$. Integration over hadronic and leptonic ϕ yields following the differential form for exclusive cross section: $\frac{d\sigma}{dE d\cos\theta_e d\cos\theta_{CM}}$. Extraction of this quantity from the HAPRAD code is done by adding contributions from σ_T and σ_L , and calculating ϵ such that $\sigma_{EXCL} \approx \sigma_T + \epsilon\sigma_L$. The default units of the exclusive cross section in the code are mbarn/sr per nucleon.

The cross sections for exclusive π^0 production on proton and neutron in the resonance region are available through MAID2007. The MAID model is constructed by means of a multipole analysis based on the recollection of photoproduction data in the ranges corresponding to the electroproduction kinematics of $1.1 < W < 2$ (GeV) and $0. < Q^2 < 5$ (GeV²). Outside of this kinematics, above $W = 2$ GeV, HAPRAD employs extrapolation of MAID cross sections. The extrapolation is based the parton-interchange model that fits Cornell [109] π^+ electroproduction dataset in the range of $1.2 < W < 3.0$ GeV and $1.2 < Q^2 < 4.4$ GeV². As implemented in the code, the exclusive cross section for $W > 2$ GeV is but a normalization of MAID cross section at $W = 2$ GeV by a

factor of $(2/W)^{const \cdot \theta_\pi^2}$. The transition at $W = 2$ GeV is by construction smooth, however, Cornell parametrization is not a reliable source of determining π^0 cross section at high W .

The recently published data from Hall B on exclusive π^0 electroproduction [110] measured in the range of $1. < Q^2 < 4.6$ GeV² and $2 < W < 3$ GeV were investigated on its relevance to our calculations of exclusive part of radiative corrections (section 5.2.5). Parametrization of this dataset by V.Kubarovsky [111] (further referred to as VK parametrization) is based on the GPD inspired phenomenological model which agrees well with the measurement. The physical components of the parametrization have the following features: the dependence of structure functions on Q^2 is that of dipole type $(1/(Q^2 + M_p^2))$; the behavior of σ_{LL} and σ_{LT} is correctly reproduced at threshold where the structure functions go to zero at a kinematically forbidden domain ($t < t_{min}$); the t -dependence of the structure functions is taken as the one commonly used in the GPD models ($\sim \ln x_B$). Such parametrization, based on phenomenological description of experimental data, ideally suits our goals. Extending it down to at least $W = 1.8$ GeV is a reliable alternative of describing the exclusive reaction below $W = 2$ GeV [112].

To have exclusive π^0 cross sections on the wide range of W , one could consider results from MAID below $W = 2$ GeV, while at $W > 2$ GeV one can employ VK parametrization. (Note, exclusive π^0 cross sections were also measured in the medium low range at $1.1 < W < 1.4$ GeV [66], however, parametrization of struc-

Fig. 5.7: Exclusive cross sections $\sigma_T + \epsilon\sigma_L$ on the proton as a function of W . Left and right plots are $Q^2=1.5 \text{ GeV}^2$ and $Q^2=3.0 \text{ GeV}^2$, respectively, both at $\theta_{\pi^0}=45^\circ$. The red curve corresponds to the HAPRAD output which employs MAID cross sections at $W < 2 \text{ GeV}$ and Cornell parametrization at $W > 2 \text{ GeV}$. The blue and the green cross sections are superimposed result of another code calculating structure functions on entire range: the blue curve is MAID result plotted for reference, green is VK parametrization extended for illustration to $W=1.5 \text{ GeV}$. The transition at the threshold $W=2 \text{ GeV}$, while acceptable at narrow kinematical range, is predominantly not smooth over the entire range.

ture functions is currently not available.) The complication of using both MAID and VK arises at $W = 2 \text{ (GeV)}$: the transition at threshold, while acceptable at narrow kinematical range, is predominantly not smooth over the entire range, requiring a thoughtful parametrization. In the following section we investigate the relevance of the covered kinematical range in either model.

5.2.4 Comparison of input cross sections

In this section we intercompare the behavior of input SIDIS and the exclusive cross sections as a function of z . In order to do that, first, we must derive the expression for $z = E_\pi/\nu$ in the lab frame from an exclusive set of variables, i.e $(Q^2, W, \cos \theta_{\text{CM}})$. The pion energy E_π^* for the reaction $\gamma^*p \rightarrow \pi p$ is derived below in CM frame.

$$s = (E_{1\text{CM}} + E_{2\text{CM}})^2$$

$$s = (P_1 + P_2)^2 = m_1^2 + m_2^2 + 2E_{1\text{CM}}E_{2\text{CM}} + 2P_{\text{CM}}^2$$

In our case, we denote P_1 as corresponding to a π^0 with mass m , and P_2 to a proton with mass M . Given that $P_{\text{CM}}^2 = E_1^2 - m_1^2$ and the initial invariant energy of the reaction $W^2 = s$, we obtain:

$$E_\pi^* = \frac{W^2 + m^2 - M^2}{2W} \quad (5.2.8)$$

To find corresponding E_π in the lab frame, we use the Lorentz transformation to the virtual photon-nucleon CM-system :

$$\begin{aligned} \begin{pmatrix} E_\pi \\ p_\pi \end{pmatrix}_{\text{lab}} &= \begin{pmatrix} \gamma_{\gamma^*M} & \gamma_{\gamma^*M}\beta_{\gamma^*M} \\ \gamma_{\gamma^*M}\beta_{\gamma^*M} & \gamma_{\gamma^*M} \end{pmatrix} \begin{pmatrix} E_\pi^* \\ p_\pi^* \end{pmatrix}_{\text{CM}} \\ \beta_{\gamma^*M} &= \frac{P_{\gamma^*M}}{E_{\gamma^*M}} = \frac{\sqrt{Q^2 + \nu^2}}{\nu + M} \\ \gamma_{\gamma^*M} &= \frac{E_{\gamma^*M}}{M_{\gamma^*M}} = \frac{\nu + M}{W} \end{aligned} \quad (5.2.9)$$

Thus, following the Lorentz transformation, the pion energy in the lab frame is:

$$\begin{aligned}
 E_\pi = z\nu &= \frac{\nu + M}{W} \frac{W^2 + m^2 - M^2}{2W} + \frac{\sqrt{Q^2 + \nu^2}}{W} \sqrt{\left(\frac{W^2 + m^2 - M^2}{2W}\right)^2 - m^2 \cos^2 \theta_{\text{CM}}} \\
 z(\text{in the limit } m^2 \rightarrow 0) &\approx \frac{W^2 - M^2}{2W^2} \left[1 + \frac{M}{\nu} + \sqrt{1 + \frac{Q^2}{\nu^2} \cos^2 \theta_{\text{CM}}} \right] \\
 p_\perp^2(\text{in the limit } m^2 \rightarrow 0) &\approx \left(\frac{W^2 - M^2}{2W}\right)^2 (1 - \cos^2 \theta_{\text{CM}}) \quad (5.2.10)
 \end{aligned}$$

Next, comparison of the amplitudes of the two cross sections requires a common basis. The SIDIS cross-section σ_{SIDIS} depends on five variables corresponding to a set of $(x, y, p_\perp^2, z, \phi)$. The exclusive cross section depends only on four variables, $\sigma_{\text{EXCL}}(Q^2, W, \cos \theta_{\text{CM}}, \phi)$. In order to compare the two cross sections, the following must be accounted for :

- the SIDIS cross-section must be integrated over either z or $p_\perp$;
- the Jacobian of transformation from SIDIS cross section variables to the exclusive cross section ones must be computed.

Integration of σ_{SIDIS} over $p_\perp^2$ is carried out as described earlier in section 5.2.2.

To compare the SIDIS cross section reduced to $\frac{d\sigma}{dx dy dz}$ with exclusive cross section reduced to $\frac{d\sigma}{dE d\cos\theta_e \cos\theta_{\text{CM}}}$, we employ formula 5.2.8 together with the following

Jacobian transformations:

$$\begin{aligned}
 \frac{d\sigma}{dx dy dz} &= \frac{\partial(E', \cos \theta_e)}{\partial(x, y)} \frac{\partial \cos \theta_{\text{CM}}}{\partial z} \Big|_{\text{fixed } Q^2, W, \phi} \cdot \Gamma \frac{d\sigma}{d \cos \theta_{\text{CM}}} \\
 \frac{\partial(E', \cos \theta_e)}{\partial(x, y)} &= \frac{M\nu}{E'}
 \end{aligned}$$

$$\left. \frac{\partial \cos \theta_{\text{CM}}}{\partial z} \right|_{\text{fixed } Q^2, W, \phi} \approx \frac{2W^2}{W^2 - M^2} \sqrt{\frac{\nu^2}{Q^2 + \nu^2}} \quad (5.2.11)$$

Fig. 5.8: Input cross sections ($\sigma_T + \epsilon\sigma_L$) before radiation plotted as a function of z at fixed kinematics with M_x cut imposed. From left to right: σ_{SIDIS} , plotted in red, is compared to $\sigma_{\text{excl}}^{\text{VK}}$ in green and to $\sigma_{\text{excl}}^{\text{MAID}}$ in blue, respectively.

5.2.5 Exclusive contribution at low W

As mentioned earlier, the transition between MAID and VK at $W = 2 \text{ GeV}$ is not smooth. Thus, a compromise between choosing either model or finding a proper parametrization at the threshold should be done. In order to understand the relevance of kinematical ranges covered by MAID and VK parametrization within our experimental range, we estimate the amount of contribution stemming from lower W which, due to electromagnetic radiation, ends up being measured at a

higher values of W . For simplicity, we assume that the cross section is independent of W , and radiation is produced according to Mo and Tsai formalism, as described below. The relation between the measured elastic cross-section and the lowest order Born cross-section in the factorized form is given at lowest order by:

$$\left. \frac{d\sigma}{d\Omega} \right|_{\text{meas}} = (1 + \delta_{rad}) \left. \frac{d\sigma}{d\Omega} \right|_{\text{Born}} \quad (5.2.12)$$

where δ_{rad} is the sum over several contributions, including virtual diagrams δ_{virt} and infrared-divergent part δ_{inf} . The expression for measured cross section is infrared-divergent with respect to the maximum radiated energy ΔE :

$$\begin{aligned} \delta_{inf} &= -\frac{2\alpha}{\pi} \left[\ln \frac{Q^2}{m^2} - 1 \right] \ln \frac{E}{\Delta E} \\ \delta &= \frac{2\alpha}{\pi} \left[\ln \frac{Q^2}{m^2} - 1 \right] \\ \delta_{inf} &= \ln \left(\frac{\Delta E}{E} \right)^\delta \end{aligned} \quad (5.2.13)$$

where E is the energy of incident or scattered electrons and m is its mass. As observed by Schwinger [113], one cannot physically distinguish a single electron from an electron propagating with “zero energy photons”. Therefore, given that the probability to detect an electron with no additional radiation is zero, the measured cross-section, due to multiple photon emission, should vanish when $\Delta E \rightarrow 0$.

Resummation to all orders is accomplished with exponentiation of the infrared divergent part :

$$\begin{aligned}\left.\frac{d\sigma}{d\Omega}\right|_{\text{meas}} &= \left.\frac{d\sigma}{d\Omega}\right|_{\text{Born}} (1 + \delta_{\text{virt}}) e^{\delta_{\text{inf}}} \\ &= \left.\frac{d\sigma}{d\Omega}\right|_{\text{Born}} (1 + \delta_{\text{virt}}) \left(\frac{\Delta E}{E}\right)^{\delta}\end{aligned}\quad (5.2.14)$$

In order to generate a realistic spectrum of photons radiated before and after interactions, we do not need virtual correction diagrams δ_{virt} to radiative processes (such as loops in the virtual photon propagator or at the electron vertex), instead we consider contribution from the infrared part only. The original Mo and Tsai formalism is then modified with a square-root of energy products :

$$\delta_{\text{inf}} = -\frac{2\alpha}{\pi} \left[\ln \frac{Q^2}{m^2} - 1 \right] \ln \frac{\sqrt{E_{\text{before}} E_{\text{after}}}}{\Delta E} \quad (5.2.15)$$

This modification cancels by a factor two in the expression for δ , leading to a decomposition of δ_{inf} onto a radiation before and after, which is consistent with collinear approximation and resummation to all orders by exponentiation.

$$\begin{aligned}\delta &= \frac{\alpha}{\pi} \left[\ln \frac{Q^2}{m^2} - 1 \right] \\ \delta_{\text{inf}} &= \ln \left(\frac{\Delta E}{E_{\text{before}}} \right)^{\delta} \left(\frac{\Delta E}{E_{\text{after}}} \right)^{\delta} \\ \left.\frac{d\sigma}{d\Omega}\right|_{\text{meas}} &= \left.\frac{d\sigma}{d\Omega}\right|_{\text{Born}} (1 + \delta_{\text{virt}}) \left(\frac{\Delta E}{E_{\text{before}}} \right)^{\delta} \left(\frac{\Delta E}{E_{\text{after}}} \right)^{\delta}\end{aligned}\quad (5.2.16)$$

Fig. 5.9: The top plots are the generated distribution of photon radiation according to equation 5.2.17, in regular (left) and logarithmic (right) scales. The bottom plot is the cumulative distribution of radiative factor based on equation 5.2.16 (left), and its derivative corresponding to the photon radiation before and after interaction (right).

In order to generate a spectrum of photons, ΔE , satisfying (5.2.16), we use the inverse transform method [114]. The cumulative distribution u which gives the probability to radiate a photon of energy fraction between 0 and $\Delta E/E$, is

uniform between (0, 1). Then ΔE can be then generated as :

$$\begin{aligned} u &= (\Delta E/E)^\delta \\ \Delta E &= u^{1/\delta} E \end{aligned} \tag{5.2.17}$$

where E is replaced at the generation level by the incoming beam energy or by scattered electron energy. If radiation takes place before interactions, then $\Delta E < E_{beam}$, or if radiation happens after interaction, then $\Delta E < E'$. Thereby generated distribution of photon radiation are illustrated in Fig. 5.9 at fixed kinematics on a plain and logarithmic scale.

To verify the validity of the generated distributions, they were compared with the functional form of the ΔE derived from equation 5.2.16. The radiative factor $R(\Delta E/E)$ represents an integral on the range (0, ΔE) of the probability density to emit a photon of energy ΔE . Therefore, in order to obtain the energy distribution, we differentiate the integral over ΔE leading to: $f(\Delta E) = \frac{\partial R(\Delta E/E)}{\partial \Delta E} = \frac{\delta}{\Delta E} \left(\frac{\Delta E}{E} \right)^\delta$. The radiative factor of (5.2.16) is illustrated in Fig. 5.9 along with the distribution of radiated energy.

Finally, to quantify the contamination of exclusive events coming from $W < 2 \text{ GeV}$ in our SIDIS sample, we consider an example in the extreme kinematics with measured $W = 2 \text{ GeV}$, and at $x_B = 0.25$. Following generation of photon radiation, the radiated spectrum of W is then calculated using 'new' values of incoming and outgoing electron energies, now reduced relatively to the measured ones. Given that the electron is losing energy before and after interaction, the total energy

Fig. 5.10: Distribution of the radiated W spectrum providing measured values are: $W = 2$ GeV and $x_B = 0.25$. Using Mo and Tsai formalism, we quantify that only 10% of events come from the range $W < 1.8$ GeV.

of the recoiling system would be less than the one measured as illustrated in Fig. 5.10. At extremely low kinematics, we find that only 10% of events come from $W < 1.8$ GeV. The exclusive contribution drops further down at another end of extreme high kinematics: at $W = 2.8$ GeV exclusive contribution reduces to merely 1% [115]. Based on the above results, we conclude that on our kinematical range using VK parametrization, reliably reproducing exclusive cross section down to at least $W = 1.8$ GeV, is significantly more relevant than using the MAID plus Cornell parametrization.

5.2.6 Calculation grid

Since input cross sections, σ_{SIDIS} and σ_{excl} , are differentials over a different set of variables, there are two separate grids onto which they are evaluated. Exclusive cross sections, using VK parametrization, are calculated in (Q^2, W, θ) set of bins, equidistant in each variable as indicated in Table 5.2.6.

Variable	Range	Bin width
Q^2	(0-5) GeV ²	0.3 ¹
W	(1.08-3.04) GeV	0.2
θ	(0-180) °	3

Table 5.1: Calculation grid for exclusive cross sections.

Exclusive π^0 production takes place on both the proton and neutron. The VK parametrization is available on the proton only. We further assume that exclusive π^0 production on the neutron can happen with a probability equal to the half of that on proton: $\sigma_{excl}^n = \frac{1}{2} \cdot \sigma_{excl}^p$. Therefore, due to the nucleon content for each nuclear target of mass number A and proton number Z , the exclusive cross section is normalized by the factor $\frac{A+Z}{2A}$.

Radiated and Born cross sections are calculated at the center of each bin on the 5-dimensional grid in $(\nu, Q^2, z, p_T^2, \phi)$. The choice of the binning in each variable must be done consistently with the binning scheme of our experimental data (see

¹ With exception for the last bin where the width is 0.2 GeV²

section 3.4) for which $(\nu, Q^2, z)=(3, 3, 5)$ number of bins and $(\nu, z, p_T^2)=(3, 5, 6)$, respectively. There are two caveats associated with calculating radiative correction on the widely spaced bins, as those used in our data. The first one is possible model dependence, for which the geometrical center of the bin, where cross sections are calculated, might differ from the actual center of the distribution. The second effect is related to the edges of kinematical coverage: the center of the last, wide bin may be kinematically inaccessible, while finer division (and consequent integration) enables the calculation in the centers which are kinematically allowed, leading to a non-zero cross section at a given bin. Taking into account the computational time, the grid used for the radiative correction calculations was chosen to be $(\nu, Q^2, z, p_T^2, \phi)=(30,30,30,30,40)$ grid. Each of the experimental bins in Q^2 and ν was divided equidistantly by another 10 bins, each of the z and p_T^2 bins were similarly divided by another 5, while all of the 40 bins in ϕ are equally spaced in the range $(0,180)^\circ$.

To transit from the 5-dimensional grid to the 3-dimensional set of bins used in our data, two methods of integration were considered: rectangular and trapezoid. Providing that the bins are non-equidistant and the grid is coarse, summation by the trapezoid method provides smaller deviation from the true integral, while in case of fine grid the two methods are compatible (methods and results are illustrated in [116]). We adopt a trapezoid method of integration in all of our summation schemes. Integrating out one of the variables by trapezoid, for in-

stance, z , is done as following:

$$\sigma(Q^2, \nu, p_T^2, \phi) = \frac{1}{2} \sum_i (\sigma(Q^2, \nu, z_{i+1}, p_T^2, \phi) - \sigma(Q^2, \nu, z_i, p_T^2, \phi)) \cdot (z_{i+1}^{cent} - z_i^{cent})$$

where z_i^{cent} is the center value of the bin i , and $\sigma(Q^2, \nu, z_i, p_T^2, \phi)$ is the value of the cross section in a corresponding bin. Averaging over the number of bins, for instance in the case of z , is done by summing over $n=5$ equidistant bin widths in order to reduce from 30 to 5 bins, is carried as:

$$\sigma(Q^2, \nu, z, p_T^2, \phi) = \frac{1}{2(n-1)} (\sigma(n \cdot z_i) + 2\sigma(n \cdot z_{i+1}) + 2\sigma(n \cdot z_{i+2}) + 2\sigma(n \cdot z_{i+3}) + \sigma(n \cdot z_{i+4}))$$

where $n=5$ is the number of bins averaged over, $i=(0,4)$ is the final total number of bins in z , and $\sigma(n \cdot z_i) = \sigma(Q^2, \nu, z_{n \cdot i}, p_T^2, \phi)$.

To conclude, radiative correction factors, applied on the bin-by-bin basis on our experimental data, are calculated, first, by averaging and integrating the five-dimensional cross sections to the three-dimensional ones, and, finally, the radiative factors are obtained by taking the ratio of the reduced radiated cross section to the Born cross section. It was observed that radiative factors are similar for all three solid targets within a fraction of a percent. Thus, to correct for radiation we employ the factors calculated from the iron target for all solid targets, so that the fit to data, which is an input in structure functions, is based on a slightly larger statistical set. We examine separately the RC that originated from SIDIS tails versus the full corrections which consist of SIDIS plus exclusive contribution. In the case of pure SIDIS contribution for the set of (Q^2, ν, z)

bins, absolute RC per nuclear target vary from unity at lowest z , down to 20-30% at highest z . Meanwhile the ratio of RC factors varies in range from 12% below unity at lowest Q^2 and highest ν up to 18% above unity at highest Q^2 and lowest ν [117]. An addition of exclusive contribution mainly affects the highest values in $z > 0.8$, which are beyond the scope of this work; otherwise in the range of z under consideration, exclusive contribution decreases the ratio of RC by 2% at maximum. Considering RC factors in the case of (ν, z, p_T^2) set of bins, contribution from SIDIS per nuclear target varies from unity at lowest bins in z down to 24% at highest z . The ratio of SIDIS RC factors on solid target to deuterium is consistent with unity uniformly, with an exception of the few lowest bins in p_T^2 , where it goes down to 5-8%. Exclusive radiative tail mainly affects the lowest p_T^2 bins in the highest bins of z , adding additional decreasing to the ratio of SIDIS RC factor on solid to D target by a couple percent only (Fig. 5.12).

5.2.7 Model dependence of RC

A brief remark to the model dependence in the context of the choice of the grid used in the radiative correction calculations. The RC factors were calculated on both the coarse grid with $(Q^2, \nu, z, p_T^2, \phi) = (3, 3, 6, 6, 8)$, and the fine $(30, 30, 30, 30, 40)$ grid. The latter was integrated down to extract RC factors in the same set of bins as the first binning scheme. The comparison showed that the largest variation between the two grids for the ratio of SIDIS radiative factors

on Fe to that on D in the set of (ν, z, p_T^2) reaches up to 3% in few bins where p_T^2 values are low, yet, elsewhere remaining on average within 1%. In case of (Q^2, ν, z) this difference reaches up to 4% at the highest Q^2 and the lowest ν , but not exceeding 1% in all other bins [118].

Fig. 5.11: Total radiative correction factors (semi-inclusive together with exclusive contribution) for Fe and D targets and their ratios in (Q^2, ν, z) bins.

Fig. 5.12: Total radiative correction factors (semi-inclusive plus exclusive contribution) for Fe and D targets and their ratios in (ν, z, p_T^2) bins.

Chapter 6

Systematic uncertainties

6.1 Overview

The target system of the EG2 experiment was designed such that the systematic errors are minimized. Since the two targets, the liquid and solid, were positioned in the beam simultaneously, the time-varying effects, such as fluctuations of beam current and occurrence of dead channels, are the same for both targets. The choice of observable in a form of a super-ratio helps to reduce, or even cancel at first order, some systematic uncertainties, such as, for example, luminosity. Other uncertainties, which stem from particle identification, reflect the fact that at times it is not possible to determine an ideal position of a specific cut, thus variations around it must be examined. To do that, we vary the cut within reasonable limits, making it more stringent or loose, and recalculate the final multiplicity with it. The systematic error is then determined on the bin-by-bin basis as the RMS of the deviation of the modified multiplicity ratio Δ_n from the original one: $\Delta_{RMS} = \sqrt{\sum \Delta_n^2} / \sqrt{n}$, where n corresponds to the number of the cuts

used for a given systematic study. In this section we describe identified sources of systematic uncertainties pertaining to the detection of electron, photon, and π^0 .

The following sources of uncertainties were considered:

Electron ID

- Target vertex cut
- Target leakage
- Sampling fraction cut

Photon ID

- Photon energy cutoff
- EC time (beta) cut

π^0 reconstruction

- Background shape
- Signal shape
- Acceptance extraction method
- Acceptance in finite bin width

6.2 Electron ID

6.2.1 Target vertex cut

The target vertex cut is used to identify the liquid and solid targets in a double target system. Due to detector resolution, the positions of the targets are smeared. For that reason, each solid target was fitted with a Gaussian function and identified within $\pm 3\sigma$ of its width. The deuterium target, being 2 cm wide, was cut at the tails. To examine systematic uncertainties, we vary the cuts on the deuterium and solid targets simultaneously and by the same quantity, as illustrated in Fig. 6.1. To choose the range for our systematic studies, we examine the aluminum target which was positioned in between the deuterium and liquid targets, and make a careful choice of the cut so as not to introduce contamination into the deuterium or solid targets. A Gaussian fit to the aluminum target (Fig. 6.2) gives a width $\sigma_{Al} = 0.19$ cm which we use in order to move the original cuts on deuterium and solid targets inside and outside by $\pm 1\sigma_{Al}$. The stringent cuts would correspond to the combination:

$$\begin{aligned} -25.46 < z_{Fe}^{in} < -24.45 \\ -31.6 < z_D^{in} < -28.59 \end{aligned} \tag{6.2.1}$$

The loose cuts meantime are as following:

$$\begin{aligned} -25.84 < z_{Fe}^{out} < -24.07 \\ -31.99 < z_D^{out} < -28.21 \end{aligned} \tag{6.2.2}$$

Fig. 6.1: z vertex distribution. Original iron distribution is plotted in blue and deuterium in red. The green dashed and solid lines illustrate stringent and loose cuts, respectively, around each of the two targets, determined based on the width of the aluminum target positioned in between them.

Next, in each bin of (Q^2, ν, z) and (ν, z, p_T^2) , we constructed RMS such that $\text{RMS} = \sqrt{(R - R(\text{in})^2) + (R - R(\text{out})^2)} / \text{sqrt}2$, where $R(\text{in})$ and $R(\text{out})$ are multiplicities corresponding to targets defined as in equation 6.2.1 and 6.2.2. The resulting systematic uncertainty would be the average RMS per bin set, which in case of (Q^2, ν, z) is merely 0.3%, and in case of (ν, z, p_T^2) is 0.5%.

Fig. 6.2: z vertex distribution of the aluminum target placed in between deuterium and iron. Its width $\sigma = 0.19$ cm, comparable with that of solid targets, is used in systematic studies as illustrated in Fig.6.1.

6.2.2 Target leakage

In order to evaluate the contamination from the liquid target to the solid target and vice versa, we select regions upstream and downstream from the targets, where, presumably, no events should be detected. Let us assume that the width of the targets, as defined by our cuts in section 3.1.8, is Δz_D for deuterium and Δz_{Fe} for iron, and that the distance between the two targets is Δ_{Fe-D} . To examine the contamination in deuterium target from the iron target, we select a region downstream, Δ_{Fe-D} away from the lower border of deuterium, of width Δz_{Fe} (region 1). The contamination in the iron target from the deuterium target would be consequently defined in the region Δ_{Fe-D} away from the upper border of the iron target of width Δz_D (region 2).

The test regions are illustrated in Fig. 6.3 by the vertical lines.

$$\begin{aligned}
 -35.94 < \text{region 1} < -34.55 \\
 -21.51 < \text{region 2} < -18.11
 \end{aligned}
 \tag{6.2.3}$$

We find that the integrated number of electrons in region 1 is 0.3% from the total number of electrons in the iron target, while in region 2 the number of electrons is 0.2% from total number in deuterium target. The contamination in the sample of π^0 candidates is 0.5% in solid target, and 0.7% in iron. Combining the uncertainties on the number of electrons and π^0 candidates, the uncertainty on the multiplicity is 0.9%.

Fig. 6.3: z vertex distribution of the double target system with test regions indicated by the vertical red (region 2) and blue (region 1) lines. The distance between the two targets and the distance between the test regions and the target target is the same (yellow).

6.2.3 Sampling fraction cut

The purpose of the sampling fraction cut, discussed in section 3.1.3, is to clean up the electron sample from high energy pion contamination. The default cutoff is momentum and sector dependent, and is of following form: $|E_{tot}/P - \mu(p)| < 2.5 \cdot \sigma(p)$. Here, $\mu(p)$ and $\sigma(p)$ are the mean and the width of the Gaussian of E_{tot}/P in each bin of P , which are then fitted with second order polynomial. To study associated systematic uncertainty, we vary the width of the fit by $\pm 0.5\sigma$, as illustrated in Fig. 6.4. The average RMS associated with sampling fraction systematic uncertainty is 0.5% in case of (Q^2, ν, z) set is 0.5% and 0.4% for (ν, z, p_T^2) set of bins.

Fig. 6.4: Sampling fraction E_{tot}/P versus electron momentum $\sigma(p)$. The red lines indicate default cutoff, corresponding to $|E_{tot}/P - \mu(p)| < 2.5 \cdot \sigma(p)$. The black curves illustrate tighter cutoff, which corresponds to $|E_{tot}/P - \mu(p)| < 2 \cdot \sigma(p)$. The green curves are looser cutoff, $|E_{tot}/P - \mu(p)| < 3 \cdot \sigma(p)$.

6.3 Photon ID

6.3.1 EC time (beta) cut

In order to separate photons from the neutrons, we use a cutoff which matches the time and the path that the photon transversed in the EC. This cutoff, discussed in section 3.2, is akin a cutoff on β . It is applied both in data (equation 3.2.1) and MC (equation 4.3.3). To evaluate the systematic uncertainty, we consider the sample of events from data. The default cutoff is: $-2.2 < t_{EC} - t_{start} - l_{EC}/30 < 1.3$, where t_{EC} (ns) and l_{EC} (cm) are time and path measured in EC. This condition in turn corresponds to $\beta > 0.93$. To identify limits for systematic studies, we chose the limits on the the time to path cutoff based on the corresponding values of β as illustrated in Fig. 6.5. In case of a stringent cut, the timing cutoff matches a cut on $\beta > 0.92$, such that:

$$-2.45 < t_{EC} - t_{start} - l_{EC}/30 < 1.55 \quad (6.3.1)$$

A loose cutoff, corresponding to often used in CLAS cut $\beta > 0.95$ to select photons, is as follows:

$$-1.87 < t_{EC} - t_{start} - l_{EC}/30 < 0.97 \quad (6.3.2)$$

The average uncertainty due to this cut for the set of (Q^2, ν, z) bins is 0.8% and for the set of (ν, z, p_T^2) is 0.6%.

Fig. 6.5: The timing cutoff (left) and respective distribution of photon β (right). The red are the default values, the green correspond to a tighter cutoff (equation 6.3.1) and the blue distributions indicate a loose cutoff (equation 6.3.2).

Minimum photon energy cutoff

If we assume that the efficiency of the photon reconstruction is 100%, the ratio of $e\pi^0$ acceptance correction factor must be independent of the photon energy. In reality, the efficiency of π^0 reconstruction changes with energy. To study this effect, we vary the default minimum cutoff on the deposited photon's energy (obtained before photon energy correction) from the original 300 MeV value to 250, 400, and 500 MeV. The photon energy cutoffs were applied both to MC events in order to determine corresponding acceptance correction factors, as well as to the data. To fit two-photon invariant masses in the case of MC reconstructed and data events (in order to extract acceptances and multiplicities), combinatorial background was recalculated for each energy cutoff, separately for MC and data. Comparison of acceptance correction factors and distributions of acceptance

corrected π^0 yields and ratios can be found here [122], and [123]. We calculate the RMS in each bin of the two sets of bins, (Q^2, ν, z) and (ν, z, p_T^2) such that we construct the difference between the default multiplicity with only one of the three, that which gives the largest deviation from the default. Next, we combine all the bins in each set 6.6 and find the average RMS of the distribution to be 1.2% and 2.2% for (Q^2, ν, z) and (ν, z, p_T^2) in case of Fe(C) target, and 2.7% and 2.2% for Pb.

Fig. 6.6: Systematic uncertainty due to photon energy cutoff in the example of iron target. RMS per multiplicity ratios are plotted together for all the bins in (ν, z, p_T^2) on the left and (Q^2, ν, z) on the right. From those histograms we extract average RMS per set of the two bins.

6.4 π^0 identification

The main source of systematic uncertainty on the extraction of the number of π^0 's stems from the fit to the invariant mass distribution. The procedure, to which we will further refer to as the default, consists in fitting the binned invariant mass in the range $(-5\sigma, 5\sigma)$ using a Gaussian distribution with a polynomial background. The number of π^0 events are calculated from the height of the normalized Gaussian. The shape of background is defined via mixing of two uncorrelated photons. To study the systematics associated with this fitting procedure, we begin by considering a number of possible approaches to π^0 extraction, the results of which are presented here [92]. From a large number, it is important to carefully select those methods which must be included in our systematic studies while rejecting those less viable. For example, we discard the sideband subtraction method which will produce a priori large systematic deviations due to non-linearity of our background, in particular at low z and p_T^2 bins. We also discard the method of background subtraction using the second, or higher, order polynomial fit which, despite fixing the range of parameters, fails to fit the invariant peak in a large number of bins. Consequently, only two methods will be employed for the estimation of systematical uncertainties on π^0 extraction. The first method consists of fitting the background shape on the left $(\pm 0.5\sigma - 5\sigma, 5\sigma)$ and on the right $(-5\sigma, \pm 0.5\sigma + 5\sigma)$ from the Gaussian peak. The result of this procedure will indicate a mismatch between an actual background shape used to

fit the invariant mass and the left/right tilt of the background beyond the signal tails. It will be therefore be attributed as our systematical uncertainty on the background shape. The second method is sensitive to the deviation of the signal distribution from the Gaussian. In contrast to default extraction, in the second method we calculate number of π^0 's by taking the difference between the number of events in the default $(-5\sigma, 5\sigma)$ range calculated from the Gaussian fit with the total number of events in the same range minus the number of events under the background fit.

Systematic error on the background shape

The background shape has been constructed based on the event mixing technique. While the technique, tested on the sample of MC events, is viable, it fails to take into account the detector resolution effects (overlapping photon clusters in the EC) within one event. Therefore, the true background may have a systematically different line shape from the one we constructed using the event mixing technique. This results in slight inconsistencies in fitting the left and the right background component of the signal peak, for which the χ^2 fit finds compromise. The default range of the fit to the 2-photon invariant mass was carefully chosen to be within 5σ width from the mean of the peak. To evaluate the effect due to the left/right tilt, we fit the background separately on the left and on the right while correspondingly varying the lower and upper limits of the fit by 0.5σ from

the default 5σ value, meanwhile fixing the range on the other end at default value $\pm 5\sigma$. Thus, the ranges for our systematical studies are: $R_1=(-5.5, 5)\sigma$ and $R_2=(-4.5, 5)\sigma$ on the left, and $R_3=(-5, 4.5)\sigma$ and $R_4=(5, 5.5)\sigma$ on the right as illustrated on Fig 6.8. The RMS (equation 6.4.1), constructed from the deviation of $R_{1,...,4}$

Fig. 6.7: Illustration of invariant mass fits at fixed kinematics ($2.2 < \nu < 3.2$, $0.3 < z < 0.4$ and $0. < p_T^2 < 0.1$). Left and right plots illustrate fits employed for the systematical studies. Left plot contains fit in the range $(-5.5, 3.5)\sigma$ in blue, and $(-4.5, 3.5)\sigma$ in green. Right plot contains fit in the range $(-3.5, 4.5)\sigma$ in green and $(-3.5, 5.5)\sigma$ in blue. Plot in the middle illustrates the default fit in the range $(-5, 5)\sigma$ in red.

from the default value R , provides an estimate of the systematical uncertainties related to the linear component and curvature of the background. A test carried out using a set of MC events has validated this approach (see appendix 7.2.4). The study of the background tilt using varying range is as well a sensitive measure of how the parameters of fit depend on the range in which the distribution is fitted.

$$\delta R = \sqrt{\frac{(R - R_1)^2 + (R - R_2)^2 + (R - R_3)^2 + (R - R_4)^2}{4}} \quad (6.4.1)$$

Fig. 6.8: Systematical uncertainties in (ν, z, p_T^2) bins associated with background tilt. Here, δR_{tilt} is the RMS of acceptance corrected multiplicities calculated within $\pm 0.5\sigma$ range on the left and right from the default value R_{π^0} .

where R and $R_{1,...,4}$ denote multiplicity ratios for each target, corrected for acceptance and radiation, such that number of π^0 's is extracted from the Gaussian distribution which fits the invariant mass on the specified range of number of σ 's.

The systematical uncertainties were extracted on a bin-by-bin basis in each set of (Q^2, ν, z) and (ν, z, p_T^2) bins individually for each target (due to the difference in statistics and associated goodness of the fit). The average RMS in (Q^2, ν, z) bins, calculated according to equation 6.4.1, and summed over all bins is 1.2%, 0.8% and 0.8% for C, Fe and Pb targets, respectively. Meanwhile, the average variation in all the bins does not exceed 2% with the exception of a couple bins, where the error is higher, reaching up to 8% [120]. Considering the (ν, z, p_T^2) bins, the averaged over all bins RMS for C, Fe and Pb is almost identical, being 2.3%, 2.3% and 2.4% correspondingly. The per bin values vary from a fraction of percent up to 1.5-2% as illustrated in Fig. 6.8.

6.4.1 Systematic error on the signal shape

For a fixed photon energy, the calorimeter resolution is fixed, therefore, the two-photon invariant mass peak would be a true Gaussian. However, in bins where the energy of the two photons varies significantly from one another, the mass peak becomes a convolution of a Gaussian with the resolution that depends on the photon energy spectrum. To estimate the systematic uncertainty, we compare our default multiplicity ratios R , where the number of events were calculated based

on the parameters of the Gaussian fit, with the multiplicities R_{back} obtained by subtracting number of events under background from the total number of events in the histogram in the 5σ band. The average error integrated over all (Q^2, ν, z) bins is 0.88% for both C and Fe and 2.65% for Pb; integration over (ν, z, p_T^2) bins yields 2.1% for both C and Fe and 4.5% for Pb. Per bin error spread in (ν, z, p_T^2) set is rather pronounced [120] varying within 6 % for most of the bins, and reaching up to 10 % in a few bins. Meanwhile, in (Q^2, ν, z) the uncertainty is within 3 % for the C and Fe bins, going up to 6 % for some bins in Pb. For the case of the Pb multiplicities, where statistics is scarce, the overall systematic uncertainty tends to be larger than in case of C and Fe multiplicities that bear a larger statistical sample.

6.4.2 Acceptance extraction method

In order to extract acceptance correction factors, the number of generated events is calculated exactly using MC ID, while the number of reconstructed events is calculated based on a Gaussian plus a combinatorial background fit to invariant mass (with background shape calculated from reconstructed events based on method developed in section 3.3.3). Alternative way to determine acceptance correction factors is to employ angular matching criteria between reconstructed and generated photons (described in section 4.3.2) in selecting reconstructed π^0 events. Such method results in a clean π^0 sample bearing a minimal background.

Therefore, a rather straightforward estimation of the number of events in the peak is possible. Whether calculated by taking the total number of π^0 candidates in the histogram in the range 0-0.3 GeV, or by fitting the invariant mass distribution with a Gaussian function and extracting number of π^0 events from the fit, the ratio of acceptance correction factors on the solid target to deuterium would remain the same. The average discrepancy between the two methods of acceptance extraction studied in (Q^2, ν, z) and (ν, z, p_T^2) bins is 0.5 % and 0.8 %, respectively, from the value of the multiplicity ratio, corrected for the specified acceptance and radiation. While there is no direct necessity to estimate the difference between alternative and default acceptance extraction method, as the later method employs the same procedure of extracting reconstructed π^0 events as in data, the reason behind this estimate is related to the systematic uncertainty estimation in the following subsection.

Acceptance in finite bin width

As discussed in section 4.4, the electron kinematics is well reproduced by our MC simulations in comparison with data. Yet, the π^0 distributions were observed to deviate in MC reconstructed shapes of z and p_T^2 from those in data. This can ultimately introduce model dependence of the $e\pi^0$ acceptance correction factor. The width of the bins in which acceptances are calculated becomes relevant when the distribution inside the bin is non-linear. To exclude model dependence

of acceptance extraction, one would ideally divide the space into the infinitely small bins, inside which the distribution would be virtually linear. The electron acceptance correction was tested on a set of finer grids than the one used for the multiplicity ratios. The ratio of acceptances on solid target to deuterium were found to vary within a fraction of percent relative to the default grid (see section 4.5.1). To avoid an additional, though small source of systematic uncertainty, electron acceptances were, nevertheless, applied on a finer set of bins. In the case of $e\pi^0$ acceptance, the investigation of bin width dependance is, however, limited by the statistics. When divided into the finer bins, the statistics of π^0 candidates becomes insufficient in order to apply the same methods of extraction as employed in data. As quantified in section 6.4.2, there are two comparable methods of acceptance calculations. Thus, for these studies, in order to avoid an introduction of additional uncertainties associated with the fit to the background, we use an alternative (angular matching) method in order to calculate the fine grid acceptances.

The bin width in the default binning scheme is divided by a half for each z and p_T^2 bin. This leads to a twice finer grid in the set of (Q^2, ν, z) and four times finer binning in case of (ν, z, p_T^2) . Number of reconstructed events were calculated from the total number of counts in a histogram, and then summed to the default binning grid. The resulting systematic uncertainty on the multiplicity ratio, averaged over all bins, is 3 % and 1.5 % for (Q^2, ν, z) and (ν, z, p_T^2)

correspondingly. The systematic uncertainty associated with the bin width is applied on a bin-by-bin basis, and varies within 3 % for (Q^2, ν, z) set and within 6 % (with an exception of one bin where it reaches up to 10 %) for the (ν, z, p_T^2) set.

6.5 Recapitulation of all the cuts

The systematic uncertainties associated with the shape of the π^0 peak and the shape of the background under the peak are accounted for in the multiplicity ratios on a bin-by-bin and target-by-target basis due to their large relative variation. The uncertainties, due to the width of the bins in which acceptances are calculated, are independent on the target multiplicities, and are applied on the bin-by-bin basis. All other systematic uncertainties are the same for all the bins and targets, and therefore enter as normalization scale uncertainty to the multiplicity ratio. Below we summarize the total systematic uncertainty for the identified sources.

Systematic uncertainty	$\Delta_{RMS}^C(\%)$	$\Delta_{RMS}^{Fe}(\%)$	$\Delta_{RMS}^{Pb}(\%)$
Target vertex cut	0.3	0.3	0.3
Target leakage	0.9	0.9	0.9
Sampling fraction cut	0.5	0.5	0.5
Photon energy cutoff	1.2	1.2	1.2
EC time (beta) cut	0.8	0.8	0.8
Background shape	1.2	0.8	0.8
Signal shape	0.9	0.9	2.6
Acceptance extraction method	0.5	0.5	0.5
Acceptance in finite bin width	1.5	1.5	1.5
Total	2.9	2.7	3.6

Table 6.1: Summary of averaged systematic uncertainties in (Q^2, ν, z) bins for C, Fe and Pb multiplicities. Values are given in percentage from the values of multiplicity ratio.

Systematic uncertainty	$\Delta_{RMS}^C(\%)$	$\Delta_{RMS}^{Fe}(\%)$	$\Delta_{RMS}^{Pb}(\%)$
Target vertex cut	0.5	0.5	0.5
Target leakage	0.9	0.9	0.9
Sampling fraction cut	0.4	0.4	0.4
Photon energy cutoff	2.1	2.1	2.1
EC time (beta) cut	0.6	0.6	0.6
Background shape	2.3	2.3	2.4
Signal shape	2.1	2.1	4.5
Acceptance extraction method	0.8	0.8	0.8
Acceptance in finite bin width	3.0	3.0	3.0
Total	5.1	5.1	6.4

Table 6.2: Summary of averaged systematic uncertainties in (ν, z, p_T^2) bins for C, Fe and Pb multiplicities. Values are given in percentage from the values of multiplicity ratio.

Chapter 7

Results and Discussion

In this section we present the experimental results on the π^0 multiplicity ratios in the two sets of 3-fold bins, $R_{\pi^0}(Q^2, \nu, z)$ and $R_{\pi^0}(\nu, z, p_T^2)$, as well as integrated one-dimensional ratios as functions of each kinematical variable, $R_{\pi^0}(Q^2)$, $R_{\pi^0}(\nu)$, $R_{\pi^0}(z)$ and $R_{\pi^0}(p_T^2)$. Presented results are corrected for acceptances and radiative effects, and include both systematic and statistical errors. The DIS region was selected based on the following constraints: $Q^2 > 1 \text{ GeV}^2$, $W > 2 \text{ GeV}$ and $2.2 < \nu < 4.25 \text{ GeV}$ (the upper limit corresponds to $y = \nu/E < 0.85$). The cut on W is applied in order to exclude the events originating from resonance regions, while the cut on y is aimed at reducing the magnitude of radiative corrections. The kinematical constraints imposed on selected π^0 's were: $0.3 < z < 0.8$ and $0 < p_T^2 < 0.9 \text{ GeV}^2$. The upper bound on z and p_T^2 are stipulated to avoid contributions from exclusive kinematics. The variable z is binned such that the conventional region of current fragmentation ($0.4 < z < 0.7$) can be localized.

7.1 Multiplicity ratios results

Hadronic multiplicity ratios were measured as three-fold dependencies in the two sets of bins: (Q^2, ν, z) and (ν, z, p_T^2) . π^0 was reconstructed from the two-photon invariant mass fitted in each bin using a combination of two functions: a Gaussian, which fits the signal peak, and a fourth order polynomial, which fits the background. The shape of the polynomial function was determined using the event mixing technique. Next, acceptance correction factors and radiative corrections for both leptonic and hadronic legs were applied on a bin-by-bin basis as a normalization of the multiplicity ratio. The final results, corrected for acceptances and radiation, are presented in the geometrical center of the bin. as illustrated in Fig. 7.2, and 7.1. For better visualization of dependencies and associated uncertainties, the values of the multiplicity ratios were shifted in the horizontal axes around the bin center as illustrated in Fig. 7.2, and 7.1. Statistical errors correspond to the inner errors on the multiplicity ratios. The total error, which is the sum in quadratures of the statistical and systematic errors, correspond to the outer error bars. Integrated results reduced to one-dimensional dependencies are shown in Fig. 7.3.

Fig. 7.1: π^0 multiplicity ratios in (ν, z, p_T^2) set plotted as a function of p_T^2 in bins of ν (indicated at the bottom by the horizontal ν axis) and z (indicated by the color code in the first box). The bottom panel is carbon multiplicities, the middle - iron, and the top panel is lead multiplicities. The inner error bar (color) reflects statistical error in a given bin, while the outer error bar (black) corresponds to the total error, for which statistical and systematic uncertainties were added in quadratures. R_{π^0} corresponding to the $0.5 < z < 0.6$ bin are plotted at the center of p_T^2 bin, while in all other bins of z the values R_{π^0} are horizontally to the left and to the right from the p_T^2 center for visualization.

Fig. 7.2: π^0 multiplicity ratios in (Q^2, ν, z) set plotted as a function of z in bins of Q^2 (indicated by the horizontal Q^2 axis)

and ν (indicated by the color). The red color gradient corresponds to the carbon target, the blue - to the iron, and the black - to the lead target. The inner error bar (color) reflects statistical error in a given bin, while the outer error bar (black) corresponds to the total error. R_{π^0} corresponding to the $3.2 < \nu < 3.73$ bin are plotted at the center of z bin, while in other two bins in ν R_{π^0} are shifted to the left and to the right from the center of z -bin for visualization.

Fig. 7.3: One-dimensional π^0 multiplicity ratios in Q^2 , ν , z , and p_T^2 bins. The error bars reflect statistical errors only.

7.2 Multiplicity ratios discussion

The dependencies of the multiplicity ratios $R_{\pi^0}(Q^2, \nu, z)$ and $R_{\pi^0}(\nu, z, p_T^2)$ on kinematical variables and their correlations are discussed in detail in the following subsections. The global features of the data are the increase of attenuation (decrease of R_{π^0} below unity) with increasing value of mass number A and with increasing values of z . The value of R_{π^0} moderately rises with increasing p_T^2 , except at the largest values of p_T^2 , where an enhancement above unity is observed. The p_T^2 dependence of multiplicity ratios shows a significant dependence on z such that at small z there is a pronounced enhancement at highest p_T^2 , while at large z it flattens out. The dependencies on Q^2 and ν are rather weak. The discussion of R_{π^0} is lead in the context of comparing it with existing one-dimensional measurement from HERMES as illustrated on Fig. 1.11. It is important to point that extraction of the differential behavior of R_{π^0} provides crucial information since the dependencies of R_{π^0} on Q^2 , ν , z , p_T^2 do not generally factorize. Therefore, integration over a wide range of kinematics and variables may introduce false dependencies unless otherwise investigated by means of multidimensional studies. In our case, this is mainly relevant in the context of multiplicity depending on both p_T^2 and z .

7.2.1 z Dependence

The R_{π^0} dependencies on z can be visualized in both sets of two-dimensional bins. As illustrated in Fig. 7.2 and Fig. 7.3, in the case of carbon, R_{π^0} is largely a constant, however, for iron and even more so for lead target there is an observable decrease with increasing z . A behavior, similar to CLAS carbon multiplicities, was observed by HERMES for π^0 multiplicity on the lightest nuclei - helium and neon. Yet, HERMES result on heavier nuclei, krypton and xenon, showed a stronger gradient of attenuation as a function of z . Importantly, the values at the largest CLAS bin in z ($0.7 < z < 0.8$) for the lead target quantitatively agrees with HERMES measurement in the same range in z for the heaviest targets. The controversy rises at low z , where HERMES observes less attenuation. Based the differential behavior on ν and Q^2 of our data, we also observe that the z -dependence of R_{π^0} is essentially independent on the lepton kinematics within the uncertainties of the measurement. The decrease of the multiplicity ratios with z and weak dependence on ν and Q^2 was equally reported in CLAS data for the charged pion states [79], [124].

The dependence on z is in the heart of hadronization: hard parton scattering cross section is contained within the definition of fragmentation functions, which have known evolution with Q^2 , and depend primarily on z . In parton energy loss models, or rescaling models, such dependence of R_{π^0} on z is interpreted in terms of decrease of fragmentation functions at large z due to a shift in z towards the lower

values which results from parton energy loss [27]. In modern version of energy loss models [28], the overall attenuation as a function of z and the nuclear size is related to the assumption that as one transits from the small to the larger nuclei, the possibility for multiple gluon emission and for parton scattering increases. The absorption types of models with a combination of partonic energy losses, for example, the color dipole model discussed in section 1.2.6, interprets the overall decrease of the multiplicity ratio with increasing z due to the vanishing production length. As the prehadron carried away large fraction of the initial parton's energy, the parton in turn must have undergone small energy losses which is possible in the scenario that the time of its propagation was short. Vanishing production length leads to a larger path available for the prehadron absorption.

7.2.2 p_T^2 Dependence

The p_T^2 dependence of R_{π^0} is illustrated in Fig. 7.1, and Fig. 7.3. Both the light and heavy nuclei show a gradual rise of R_{π^0} from low p_T^2 values transitioning to a strong enhancement in the highest p_T^2 bin. Comparing to one-dimensional HERMES measurement, which observed small dependence for light nuclei and rise, by 40%, for the heaviest nuclei, enhancement at CLAS kinematics is more uniform as a function of atomic number, as well as more moderate: integrated over z and ν , p_T^2 -dependence of R_{π^0} shows enhancement by 20%. Such difference between HERMES data and CLAS may be related to the fact, that HERMES

integrates over lower z values as well as over a larger range in ν . It may be speculated that at this regime parton rescattering mechanisms and energy losses are significant, leading to larger broadening of transverse momentum.

The effect, characterized by a large increase of the multiplicity at large transverse momenta ($p_T^2 \geq 1 \text{ GeV}^2$), resembles Cronin-like effect. It was first observed in the measurements by EMC (shown in Fig. 1.10), latter by FNAL in hadron-nucleus collisions [42], and further confirmed in lepton-nucleus reactions by the HERMES [44] and CLAS data on charged pions. Theoretically, such an effect has been explained in terms of initial and final state multiple parton scattering prior to its fragmentation. As illustrated in Fig. 7.1, the Cronin effect is largely independent of ν , however, it vanishes for the largest z . The latter is consistent with the idea that the Cronin effect is of partonic nature (such as multiple quark re-scattering in the nuclear target). In the limit $z \rightarrow 1$, as the quark is not allowed to lose any energy, the production time of the propagating quark is vanishing, therefore, leaving no room for re-scattering. Meanwhile, such a scenario indicates that the attenuation in the limit $z \rightarrow 1$ is purely due to a hadron absorption mechanism. The z -dependence of the Cronin effect in our data quantitatively agrees with predictions from the color-dipole model as illustrated in Fig. 1.20.

7.2.3 Q^2 Dependence

As illustrated in Fig. 7.2, the dependence of $R_{\pi^0}(Q^2, \nu, z)$ on Q^2 is weak, and maybe be assumed constant within the uncertainties of the measurement. However, following integration over ν and z , as presented in Fig. 7.3, we observe a slight rise within 10% of $R_{\pi^0}(Q^2)$ with scale Q^2 . A moderate increase of multiplicities with Q^2 has been also demonstrated for positive charged pions in CLAS data [79]. HERMES results, covering a larger range in Q^2 up to 10 GeV, showed a small increase with Q^2 [43] for the three pion states, which is slightly stronger for heavier nuclei. They concluded that attenuation is not very sensitive to Q^2 , and, therefore, integrating multiplicities over Q^2 would not introduce false dependencies. The Q^2 coverage in our current measurement is modest, yet it will be extended to much higher values reaching up to 9 GeV² at the future CLAS12 program [65].

The dependence of multiplicity ratios on Q^2 is an important measurement being sensitive to the underlying dynamics of the hadronization process. In the context of early string models, there was no indication that the hadron attenuation will be affected by the Q^2 scale. In the modern string models, on the contrary, a different kind of Q^2 dependence is explored. The deconfinement model, discussed in section 1.2.6, predicts a slight decrease of multiplicity with Q^2 , which is not supported by data. Considering the prehadron as a color dipole, the dependence of the prehadron cross section is sensitive to the Q^2 scale due to color transparency

as $\sigma_h \propto 1/Q^2$. Since the initial size of the hadron shrinks with Q^2 , the survival probability of the prehadron increases. The color-dipole model calculated Q^2 dependence from two counteracting processes (color transparency and contraction of production length). Their calculations resulted in very small dependence of multiplicity rising with both Q^2 and ν scales. When describing attenuation in terms of a modification of the fragmentation functions only [37], a slight increase of multiplicities is also predicted with Q^2 .

7.2.4 ν Dependence

The multidimensional set of $R_{\pi^0}(Q^2, \nu, z)$ and $R_{\pi^0}(\nu, z, p_T^2)$ does not reveal a flat ν -dependence, however, within our uncertainties it is very small. A slight rise of multiplicities with increasing ν is more pronounced when integrated over Q^2 and z , as shown in Fig. 7.3. A slight increase with ν has been previously observed in CLAS kinematics in the case of charged pions. One of the first measurements by EMC in the range $10 < \nu < 220$ GeV indicated a slight increase of the multiplicity with increasing value of ν . The HERMES data, which covers the range $7 < \nu < 23$ GeV, showed a similar trend of increasing R_π with ν ; while the increase is small for light nuclei, it becomes prominent for the heavier nuclei. The trend of larger nuclear suppression at low energies and vanishing suppression with increasing energy is well reproduced by the color dipole model. At higher energies, production time rises leaving less room for absorption, yet, formation time due to

Lorentz dilatation rises too, slowing down the prehadron evolution leading to a lesser attenuation due to color transparency.

Conclusions

The microscopic information on space-time characteristics of hadronization can now be accessed via DIS by employing nuclear medium as a spacial analyzer. The space-time evolution of hadronization is influenced by the nuclear medium and has measurable consequences which manifest themselves in the modification of the hadron yields produced in nuclear matter compared to those in vacuum.

The 3-fold kinematic dependencies of π^0 multiplicity ratios, measured in semi-inclusive deep-inelastic scattering of electrons on the three nuclear targets relative to deuterium, have been measured in two sets of bins: $R_{\pi^0}(Q^2, \nu, z)$ and $R_{\pi^0}(\nu, z, p_T^2)$. The differential multiplicities were then integrated over two of the three variables in order to extract $R_{\pi^0}(Q^2)$, $R_{\pi^0}(\nu)$, $R_{\pi^0}(z)$, and $R_{\pi^0}(p_T^2)$ dependencies and compare them with existing 1-fold HERMES measurements. Subsequently, HERMES published the two-fold differential measurements for charged particles, but not for the neutral pion. The neutral pion is substantially more difficult to measure than the charged pions. This is largely, but not exclusively, related to the fact that a combinatorial background of the two photons is always present under the π^0 peak; it must be estimated and removed in all the bins, in both data and simulation, using a controlled procedure. Thus, the measurement

carried out in the present work, is a world's first in that it is a 3-fold differential measurement for the neutral pion, going beyond the 1-fold distributions of HERMES.

The most prominent features of the data were found to be in qualitative agreement between HERMES and CLAS measurements on three pion states: systematic increase of attenuation (decrease of R_{π^0} below unity) with increasing value of atomic number A ; attenuation of forward hadrons carrying a large fraction of initial energy z , in particularly pronounced for large A ; moderate increase of R_{π^0} with p_T^2 at highest z , and an enhancement at large p_T^2 at lower values of z . We also observe a very small rise of R_{π^0} with increasing ν , and a very weak dependence on Q^2 . In total, an extensive dataset on R_{π^0} has been obtained to guide modeling of hadronization in nuclear matter. The 3-fold dependencies, once explored in the framework of existing theoretical models, will provide detailed information, previously inaccessible.

In future, based on the current dataset, it is planned to continue the exploration to extract the transverse momentum broadening of π^0 as a function of atomic number A in 1-fold dependencies on ν or z . Measurements of $\pi^\pm$ broadening in a series of nuclei were published by HERMES, and extracted from CLAS data. However, p_T^2 broadening for the neutral pion has never been published, due to the difficult nature of the background subtraction and broader calorimeter resolutions. Another measurement would be related to the extraction of multiplicity ratios of

η mesons. A visible signal with a useful statistical sample was observed from the current dataset, making it feasible to compare the suppression in DIS of η to π^0 which has never been previously accessed in cold QCD matter. Both hadrons have similar up and down quark content, however, unlike π^0 , the wavefunction of the η meson contains an $s\bar{s}$ component. Thus, given different masses and hadron-nucleon cross sections of the two mesons, these studies would be a novel and important tool to shed light on the universality of light hadron suppression and the relative contribution of the in-medium parton quenching versus pre-hadron absorption.

The future program with CLAS12 (E12-06-117) will provide by far the best experimental access to medium-stimulated parton energy loss and enable extraction of 4D multiplicities for a large spectrum of hadrons. Proposed experiments in the EIC offer high energy eA collisions, further extending the studies of hadronization in the luminosity and energy frontiers.

Appendix A

MC example for the calculation of systematic uncertainties

Before discussing the results of the studies of systematic uncertainties, first we consider a simple example illustrating a number of approaches that can be taken towards the calculation of errors themselves. The reason behind this illustration is the fact that no *a priori* prescribed method of calculation exists, thus one must proceed with care to investigate the main possible culprits in every specific experiment. In this example we extract the number of events which are assumed to be the invariant mass of two the photons in a neutral pion decay. To keep the example simple and uncontroversial, the peak is purely a Gaussian and the background is linear by construction. To estimate the systematical uncertainties, we will compare two approaches in obtaining the number of counts in the peak. We generate a histogram randomly filled with 100k events according to a linear background between 0 and 300 MeV/c². With 10% probability, we add events in the Gaussian peak centered at the π^0 mass with the width of 10 MeV. The two methods of extracting the number of events in the peak will be further considered:

- a χ^2 fit using the Gaussian plus linear background. By construction, this method is unbiased and does not have systematics associated with it.
- a sideband subtraction in the range of 3σ . Since the background is linear, its estimation is unbiased. However, given the finite range of the sidebands,

there is on average 0.27% of the signal events outside of the peak. Those events are not counted in the signal but are subtracted as a background. Thus there is a systematic deficit of 0.54% on average in the estimation of the number of signal events using this method.

Fig. A.1: MC distribution of the number of events assumed to be the invariant mass of the two photons. By construction, the peak is purely a Gaussian and the background is the first order polynomial. The statistical fluctuation on the number of events in the peak, relevant to the fitting method, are given by the normalization of the Gaussian fit (± 170 events). The fluctuations stemming from the sideband subtraction method are defined by the statistics under the peak in the full 6σ (± 200 events). As anticipated, the errors of two methods happen to agree, moderately.

Fig. A.2: The distributions of number of counts, generated 1k times, filled with: the difference between number of events calculated using two methods (top left - Gaussian fit, top right -sideband subtraction) relative to the generated number of events; the mean values (bottom left) of the two methods, and their absolute values (bottom right). The RMS and mean values of the top two histogram agree well with the predicted, by construction, fluctuations of the method themselves. The mean value calculated by taking the difference of the two methods, unlike its absolute value, proves to be a relevant estimation of the systematic error.

The two methods are illustrated in Fig. A.1. There is a natural statistical fluctuation of ± 100 events in the signal peak. Extraction of signal events via fitting method increases statistical fluctuation up to ± 170 events which accommodates the normalization error of the Gaussian fit (error on the parameter p_0 times the bin width found on Fig. A.1). The statistical fluctuation relevant for sideband subtraction method consists of the total number of events under the peak in the full 6σ ($N_{peak+back}+N_{side}$); by construction it is ± 200 events.

We repeat generation of this example 1k times, and construct four histograms containing: the difference between result from each method relative the true number of events, the difference between the two methods, and the absolute value of the difference between the two methods. The results are illustrated in Fig. A.2. As anticipated, the distribution for the number of signal events extracted using the fitting methods is a Gaussian (top left). Its width, or RMS, is ≈ 141 counts which compares well with the expected statistical fluctuations stemming from the normalization error of the Gaussian fit; its mean value is within the 3% accuracy as anticipated from repeating the procedure 1k times. The distribution of signal events from the sideband subtraction (top right) is also a Gaussian with its width ≈ 186 comparable to the expected statistical fluctuation of the number of events due to the subtraction method. The mean value of this distribution, centered around 57 counts, reflects the systematical error which happens to agree very well with the expected number of 54 counts due the accuracy of the measurement

within the 6σ range. In practice, however, one would not know the true number of events, for which reason we construct the histogram of the difference between the two methods given in the bottom left of Fig. A.1. Since both methods fit histograms with the similar fluctuations, the resulting RMS of the distribution is less than each RMS of the two method. The mean value of the difference between the two methods provides an excellent estimate of the systematic error due to sideband subtraction. However, if one would choose to use the histogram of the absolute value of the difference (bottom right), one would introduce a systematic bias by overestimating the uncertainty because the absolute value is always positive. Therefore, in our studies we generally construct the mean value of the distribution in calculating the systematical error.

Bibliography

- [1] The CTEQ Collaboration, *Handbook of perturbative QCD* Version 1.0, Rev.Mod.Phys 67, pages 157-248, (1995).
- [2] F.Halzen, A.Martin, *Quarks and leptons: an introductory course in modern particle physics*, ISBN 0-471-88741-2, (1984).
- [3] J.D.Bjorken, *Asymptotic sum rules at infinite momentum*, Phys.Rev. 179 (1968).
- [4] PDG, J.Phys.G: Nucl.Part.Phys (2012).
- [5] D.Griffiths, *Introduction to elementary particles*, ISBN 0-471-60386-4 (1987).
- [6] A. Deur, V. Burkert, J.P. Chen, W. Korsch, *Determination of the effective strong coupling constant $\alpha_{s,g1}(Q^2)$ from CLAS spin structure function data*, Phys.Lett.B665, pages 349-352 (2008).
- [7] J.C.Collins, D.E.Soper, G.Sterman, *Factorization of hard processes in QCD*, Adv. Ser. Direct, High Energy Phys.5 (1988).
- [8] A.Accardi, F.Arleo, W.K.Brooks, D.D'Enterria and V.Muccifora, *Parton Propagation and Fragmentation in QCD matter*, Riv.Nuovo Cim., vol.32, pages 439-553 (2010).
- [9] G.Lafferty, P.Reeves and M.Whalley, *A compilation of inclusive particle production data in $e+e-$ annihilation*, J.Phys.G, vol. G21, (1995).
- [10] D. de Florian, R.Sassot and M.Stratmann, *Global analysis of fragmentation functions for pions and kaons and their uncertainties*, Phys.Rev., vol. D75, 114010 (2007).
- [11] A.Nakamura, T.Saito, *Color Confinement and Coulomb Gauge QCD*, Prog.Theor.Phys 115 (2006).
- [12] C.Roberts, *Looking into the matter of light-quark hadrons*, arXiv:1109.6325

- [13] Y.L.Dokshitzer, *QCD phenomenology*, Lectures at the CERN-Dubna school, Pylos (2002).
- [14] B.Z.Kopeliovich, J.Nemchik, I.Schmidt, *Time evolution of jets and perturbative color neutralization*, Nucl.Phys. A782, pages 224-233 (2007).
- [15] S.J.Brodsky, *Novel QCD effects in nuclear collisions*, Nucl.Phys. A544, pages 233-236 (1992).
- [16] V.Del Duca, S.J.Brodsky, *Space-time structure of deep-inelastic lepton-hadron scattering*, Phys.Rev.D46, pages 931-943 (1992).
- [17] J.Gomez *et al.*, *Measurements of the A dependence of deep-inelastic electron scattering*, Phys.Rev.D49, 4348-4373 (1994).
- [18] J.Seely, A.Daniel, D.Gaskell, *New Measurements of the european muon collaboration effect in very light nuclei*, PRL 103, 202301 (2009).
- [19] Y.L.Dokshitzer, V.A.Khoze, A.H.Mueller and S.I.Troian, *Basics of perturbative QCD*, Gif-sur-Yvette, France: Ed. Frontieres (1991).
- [20] A.Bialas, *Attenuation of High energy particles leptonproduced in nuclear matter*, Acta Phys.Pol B11 (1980).
- [21] A.Bialas, M.Gyulassy, *Lund model and an outside-inside aspect of the inside-outside cascade*, Nucl.Phys. B291, pages 793-812 (1987).
- [22] A.Bialas, T.Chmaj, *Leptonproduction of hadrons from nuclear targets and fragmentation of quarks into hadrons*, Phys.Lett. B133, pages 241-244 (1983).
- [23] A. Bialas, J. Czyzewski, *Attenuation of coloured strings in nuclear matter*, Phys.Lett. B222, pages 132-134 (1989).
- [24] B.Andersson, *The Lund model*, Cambridge University Press (1998).
- [25] T.Falter, W.Cassing, K.Gallmeister, U.Mosel, *Hadron attenuation in deep-inelastic lepton-nucleus scattering*, Phys.Rev. C70 (2004).
- [26] M.Gyulassy, M.Plumer, *Jet quenching in lepton nucleus scattering*, Nucl.Phys. B346, pages 1-16 (1990).
- [27] E.Wang, X.N.Wang, Phys.Rev.Lett. 89, 163301 (2002).
- [28] A.Majumder, *The In-medium scale evolution in jet modification*, arXiv:0901.4516

- [29] A. Accardi, D. Grunewald, V. Muccifora and H. J. Pirner, Nucl.Phys. A761, 67 (2005).
- [30] A. Accardi, V. Muccifora, H.J.Pirner *Hadron production in deep inelastic lepton-nucleus scattering*, Nucl.Phys A720, pages 131-156 (2003).
- [31] A. Accardi, *Space-time evolution of hadronization*, Eur.Phys.J. C49, pages 347-353 (2007).
- [32] B.Z.Kopeliovich, J.Nemchik, E.Predazzi, A.Hayashigaki *Nuclear Hadronization: Within or Without?* Nucl.Phys A740, pages 211-245 (2003).
- [33] B.Z.Kopeliovich, J.Nemchik, I.Schmidt *Time Evolution of Jets and Perturbative Color Neutralization* Nucl.Phys. A782, pages 224-233 (2007).
- [34] R.Baier, Yu.L.Dokshitzer, A.H.Mueller, S.Peigne, D.Schiff *Radiative energy loss and $p_{\perp}$ -broadening of high energy partons in nuclei* Nucl.Phys. B484, pages 265-282 (1997).
- [35] R.Baier, Yu.L.Dokshitzer, A.H.Mueller, D.Schiff *Medium-induced radiative energy loss; equivalence between the BDMPS and Zakharov formalisms* Nucl.Phys. B531, pages 403-425 (1998).
- [36] A.Accardi, *Can we distinguish energy loss from hadron absorption?*, Acta Phys.Hung. A27, pages 189-192 (2006).
- [37] F.Arleo, *Tomography of cold and hot QCD matter: tools and diagnosis* JHEP 11, 044 (2002).
- [38] L.S. Osborne *et al.*, Phys. Rev.Lett. 40, 1624 (1978).
- [39] L. Hand *et al.*, Z. Phys. C1, pages 139-142 (1979).
- [40] E665 Collaboration, M. R. Adams *et al.*, Z.Phys. C65, pages 225-244 (1995).
- [41] European Muon Collaboration J. Ashman *et al.*, Z. Phys. C52 (1991).
- [42] J.W Cronin *et al.*, Phys.Rev. D11, 3105 (1975).
- [43] A. Airapetian *et al*, Nucl.Phys. B780, 1 (2007).
- [44] A. Airapetian *et al*, *Multidimensional study of hadronization in nuclei*, Eur.Phys.J A47 (2011).
- [45] A.Accardi, J.L. Albacete, M. Anselmino, N. Armesto, E.C. Aschenauer, A. Bacchetta, D. Boer, W. Brooks *et al.*, *Electron Ion Collider: The Next QCD Frontier - Understanding the glue that binds us all*, arXiv:1212.1701

- [46] W.K.Brooks *Physics with Nuclei at an Electron Ion Collider*, arXiv:1008.0131
- [47] P.Adamson *et al.*, *A study of muon neutrino disappearance using the Fermilab Main Injector Neutrino beam* arXiv:0711.0769
- [48] S.Manly, H.Lee, *Measurements of pion production in eA with the CLAS detector*, Proceedings of NUINT 2012.
- [49] K.Abe *et al.*, *First muon-neutrino disappearance study with an off-Axis beam*, arXiv:1201.1386
- [50] H.Gallagher, G.Garvey, G.Zeller, *Neutrino-Nucleus Interactions*, Annu.Rev.Nucl.Part. Sci. 2011.61:355-78.
- [51] S.Dytman, H.Gallagher, M.Kordorsky, *Hadronic Shower Energy Scale Uncertainty in the MINOS Experiment*, arXiv:0806.2119.
- [52] V.T.Nguyen for MiniBooNe collaboration, *NC π^0 production in the Mini-BooNe Antineutrino*, arXiv:0801.0628.
- [53] S.Assylbekov *et al.*, *The T2K ND280 Off-axis Pi-Zero detector*, arXiv:1111.5030
- [54] A.Ereditato, *The appearance of tau neutrinos* Scienza in primo piano vol26, no3-4 (2010).
- [55] C.Leeman, D.Douglas, G.Krafft, *The Continuous Electron beam Accelerator Facility: Cebaf at the Jefferson Laboratory*, Annu.Rev.Nucl.Part Sci 2001.51:413-450 (2001).
- [56] Southern University Research Association (SURA), *Conceptual Design Report, CEBAF Basic Experimental Equipment*, CEBAF, Newport News, Virginia (1993).
- [57] N. A. Mecking *et al.*, *The CEBAF large acceptance spectrometer (CLAS)*, Nucl. Inst. and Meth. A 503, 513 (2003).
- [58] M.Mestayer *et al.*, *The CLAS drift chamber system*, Nucl.Instrum.Meth., vol.A449, (2000).
- [59] G.Adams *et al.*, *The Clas Čerenkov detector*, Nucl.Instrum.Meth., vol.A465 (2001).
- [60] E.Smith *et al.*, *The time-of-flight system for CLAS*, Nucl.Instrum.Meth., vol.A432 (1999).

- [61] M.Amarian *et al.*, *The CLAS forward electromagnetic calorimeter*, Nucl.Instrum.Meth., vol.A460 (2001).
- [62] W.K.Brooks *et al.*, *Quark propagation through cold QCD matter*, Proposal to PAC22 (2002).
- [63] K.Hafidi *et al.*, *Q^2 dependence of nuclear transparencies for incoherent ρ^0 electroproduction*, Proposal to PAC22 (2002).
- [64] H.Hakobyan, W.K.Brooks *et al.*, *A double-target system for precision measurements of nuclear medium effects*, Nucl. Inst. and Meth. A592, pages 218-223 (2008).
- [65] K.Hafidi, W.K.Brooks *et al.*, *Quark propagation and Hadron formation*, Proposal to PAC30 (2006).
- [66] M. Ungaro, *π^0 electroproduction from $\Delta(1232)$ at high momentum transferred with CLAS.*, PhD thesis (2005).
- [67] A. Vlassov, <http://www.jlab.org/vlassov/cc/>
- [68] M. Osipenko and A. Vlassov, *CLAS Note 2006-09*.
- [69] L. Zana, http://nuclear.unh.edu/lzana/Fiducial_cut/
- [70] M. Amarian *et al.*, *The CLAS forward electromagnetic calorimeter*, Nucl. Inst. and Meth. A460, pages 239-265 (2001).
- [71] R. De Masi *et al.*, *CLAS Note 2006-015*.
- [72] L. I. Schiff, *Radiative Correction to the Angular Distribution of Nuclear Recoils from Electron Scattering*, Phys. Rev. 87, pages 750-752 (1952).
- [73] T.Mineeva https://userweb.jlab.org/mineeva/mcmix_corr/mcmix_corr.html
- [74] T.Mineeva <https://userweb.jlab.org/mineeva/mixcompare/mixcompare.html>
- [75] T.Mineeva http://clasweb.jlab.org/wiki/index.php/PT2_CutOff
- [76] T.Mineeva https://clasweb.jlab.org/wiki/index.php/Mx_studies
- [77] G. Ingelman, A. Edin and J. Rathsman, *LEPTO 6.5 - a Monte Carlo generator for deep inelastic lepton-nucleon scattering*, Comp. Phys. Comm. 101, 108 (1997).
- [78] T. Sjostrand, *The Lund Monte Carlo for jet fragmentation and e^+e^- physics -JETSET version 6.2*, Comp. Phys. Comm. 39, 347 (1986).

- [79] H.Hakobyan, *Observation of quark propagation and pattern in nuclear medium*, PhD thesis (2008).
- [80] H. Hakobyan, *CLAS NOTE 2008-002* (2008).
- [81] N.Dashyan, S.Stepanyan, *CLAS NOTE 2006-016* (2006).
- [82] CLAS Acceptance Document Review Commitee, *CLAS Acceptance Calculations: Basic Principles, Recommended Procedures*, (2001).
- [83] L.Zana, <http://www.jlab.org/Hall-B/secure/eg2/Efficiency>
- [84] M. Osipenko, *CLAS NOTE 2009-006* (2009).
- [85] E.Rizvi, *Methods and Technics in Experimental Particle Physics, series of lectures in the Queen Mary University of London* (2010).
- [86] C.Cioffi degli Atti, S.Simula, *Realistic model of the nucleon spectral function in few- and many-nucleon systems*, Phys.Rev. C53,4 (1996).
- [87] T.Mineeva http://clasweb.jlab.org/wiki/index.php/3D_Purity
- [88] T.Mineeva <http://clasweb.jlab.org/wiki/index.php/Taya#Acceptance>
- [89] T.Mineeva http://clasweb.jlab.org/wiki/index.php/Epi0Acceptance_qvz_3D
- [90] T.Mineeva http://clasweb.jlab.org/wiki/index.php/Compare_BeforeAfter
- [91] T.Mineeva http://clasweb.jlab.org/wiki/index.php/Results_pi0Background
- [92] T.Mineeva http://clasweb.jlab.org/wiki/index.php/Pi0_Background
- [93] T.Mineeva http://clasweb.jlab.org/wiki/index.php/Results_pi0Peak
- [94] M. D. Mestayer et al., *Nucl. Inst. and Meth. A* 449, 81 (2000).
- [95] G. Adams et al., *Nucl. Inst. and Meth. A* 465, 414 (2001).
- [96] E. S. Smith et al., *Nucl. Inst. and Meth. A* 432, 265 (1999).
- [97] M. Osipenko et al., *CLAS NOTE 2004-020*, 2004.
- [98] F. X. Girod, *PhD thesis, CEA-Scalay, France, 2006*
- [99] L.W.Mo, Y.S.Tsai, *Radiative corrections to elastic and inelastic ep and μp scattering* Reviews of Modern Physics Volume 41, 1 (1969).
- [100] I.Akushevich, N.Shumeiko, A.Soroko, *Radiative effect in the processes of hadron electroproduction*, Eur. Phys. J C10 (1999).

- [101] P.Bosted, *Radiative Correction for polarized SIDIS*, EG1-DVCS Technical Note #24 (2010).
- [102] P.Bosted, *SEMIRC code*, <http://www.jlab.org/Hall-B/secure/eg1-dvcs/technotes/radcor/>
- [103] H.Hakobyan, S.Mancilla, *Migration of HAPRAD2 to C++*, https://github.com/usm-data-analysis/HAPRAD_cpp
- [104] I.Akushevich, A.Ilyichev, M.Osipenko, *Lowest order QED radiative corrections to five-fold differential cross section of hadron leptonproduction*, arXiv:0711.4789.
- [105] T.Mineeva, <https://clasweb.jlab.org/wiki/index.php/Hapradcpp.Haprad2>
- [106] P.J.Mulders, R.D.Tangerman, *The complete tree level results up to order $1/Q$ for polarized deep inelastic leptonproduction*, Nucl.Phys. B461 (1996).
- [107] D.Drechsel, L.Tiator, *Threshold pion photoproduction on nucleons*, J.Phys.G: Nucl Part Phys 18 (1992).
- [108] L.N.Hand *Experimental investigations of pion electroproduction*, Phys.Rev.129, 1834.1 (1963)
- [109] A. Browman et al., *Electroproduction of single pions with large transverse momenta* Phys. Rev. Lett. 35, 1313 (1975).
- [110] I.Bedlinsky et al., *Measurement of exclusive π^0 electroproduction structure functions and their relationship to transversity GPD* Phys.Rev.Lett 109, 112001 (2012).
- [111] V.Kubarovsky https://clasweb.jlab.org/rungroups/e1-dvcs/wiki/index.php/MC_Generator_for_pi0
- [112] V.Kubarovsky, private communication.
- [113] J. Schwinger, Phys. Rev. 76 (1949).
- [114] J. Beringer et al.(PDG), *Monte-Carlo techniques* PR D86, 010001 (2012) (<http://pdg.lbl.gov>)
- [115] T.Mineeva http://clasweb.jlab.org/wiki/index.php/Exclusive_Range
- [116] T.Mineeva <https://clasweb.jlab.org/wiki/index.php/Integration>
- [117] T.Mineeva https://clasweb.jlab.org/wiki/index.php/HAPRAD_cpp_VK

- [118] T.Mineeva https://clasweb.jlab.org/wiki/index.php/Calculation_Grid
- [119] T.Mineeva https://clasweb.jlab.org/wiki/index.php/Iteration_Haprad
- [120] T.Mineeva <https://clasweb.jlab.org/wiki/index.php/Systematics>
- [121] T.Mineeva https://clasweb.jlab.org/wiki/index.php/Result_AccMethods
- [122] T.Mineeva https://clasweb.jlab.org/wiki/index.php/Epi0_AccPhotEnergy
- [123] T.Mineeva https://clasweb.jlab.org/wiki/index.php/Npi0Acc_Egamma
- [124] R.Dupre, *Quark fragmentation and hadron formation in nuclear matter*, PhD thesis (2011).