# THE STRUCTURE FUNCTION OF THE FREE NEUTRON AT HIGH $X$-BJORKEN

by

Madhusudhan Pokhrel
B. Sc. December 2013, Tribhuvan University, Nepal
M. Sc. December 2016, Tribhuvan University, Nepal
M. S. May 2019, Old Dominion University, USA

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

PHYSICS

OLD DOMINION UNIVERSITY
May 2025

Approved by:

Stephen Bueltmann (Director)

Sebastian Kuhn (Member)

Raul Briceno (Member)

Yuan Zhang (Member)

Sylvain Marsillac (Member)

# ABSTRACT

## THE STRUCTURE FUNCTION OF THE FREE NEUTRON AT HIGH $X$-BJORKEN

Madhusudhan Pokhrel
Old Dominion University, 2025
Director: Dr. Stephen Bueltmann

Understanding the internal structure of nucleons is one of the primary goal of nuclear physicists. As protons and neutrons are only the bound state solution of the QCD lagrangian (at least inside atomic nuclei), studying protons and neutrons helps uncover nuclear structure. Due to its easy availability, many studies on protons have been done on a wide range of kinematics. However, free neutron targets are not readily achievable. So, any information on neutrons has to be extracted from neutron-rich nuclei, and some nuclear models have to be used to subtract the contributions from other nucleons to extract the information on neutrons. So, the Barely Off-shell Nucleon Structure (BONuS12) experiment at Jefferson Lab was conducted to overcome these challenges by using spectator tagging. The experiment effectively created a quasi-free neutron target by scattering electrons off a deuterium target and detecting low-momentum, backward-moving protons using a custom-built Radial Time Projection Chamber (RTPC). Selecting the low momentum and backward-moving spectators would enable us to minimize the model-dependent effects due to final state interactions and target fragmentation. The RTPC was a 40 cm-long cylindrical detector that works on the principle of gaseous ionization. It had three layers of Gas Electron Multipliers (GEMs) for charge amplification and a surrounding readout pad. The scattered electrons were measured using the CLAS12 detector, and data were collected using a 10.4 GeV electron beam during Spring and Summer 2020. Using spectator tagging, we extracted the structure function ratio $F_2^n$ of the quasi-free neutron in the deep inelastic scattering at high $x$, upto $x \sim 0.8$. The result was extracted in the region with the invariant mass $W > 1.8$ GeV/$c^2$, and $Q^2$ in the range 1.3 to 11 GeV$^2$. This dissertation presents the methodology, event selection criteria and refinements, estimation and subtraction of backgrounds, and complete analysis of extraction of $F_2^n/F_2^p$ in a model-independent way. Also, systematic uncertainties in our final analysis will be discussed in detail.

To my parents, Tekraj Pokhrel and Maiya Devi Pokhrel, beloved wife Anupama, and my princess daughter Avni.

# ACKNOWLEDGMENTS

I want to thank all the people who have been directly and indirectly involved with me in completing my PhD journey. To begin with, I would like to thank my thesis advisor, Dr. Stephen Bueltmann, who believed in me and accepted me as his thesis student. Along with being a great Physicist, he is also a lovely human. His support and belief in me helped me reach my goal and complete my project.

I also want to thank Dr. Sebastian Kuhn. His guidance was crucial in successfully completing the BONuS12 experiments and the analysis. Over the past seven years, I have learned a lot of analysis techniques and ideas from him, thanks to those fruitful discussions we used to have in his office. Next, I would like to thank Dr. Mohammad Hattawy for his support and guidance over these years. His comments and ideas in every analysis meeting helped me learn much about data analysis.

Next, I would like to thank Jiwan Poudel, who helped me understand the RTPC detector and its DAQ system during my early days. Because of his help, I could easily catch up on my complex research project. I also want to thank my BONuS12 colleague Yu Chun Hung for the fruitful analysis discussions and ideas we shared during the time we shared as BONuS12 graduate students. Similarly, I would like to thank David Payette and Nathan Dzbenski, who helped me in my early research days. Additionally, I would like to thank Eric Christy for the ideas he provided during our analysis meetings. Furthermore, I would like to thank my thesis committee, Dr. Raul Briceno, Dr. Sylvain Marsillac, and Dr. Yuan Zhang. I would also like to thank my professors at ODU, who taught me different courses over the first two years of my studies at ODU. Thanks to Dr. Mark Havey, now only in our memory, for accepting my graduate application to study at ODU. I would also like to remember my classmates at ODU; working with you was terrific.

Thanks to my wife, Anupama, for continuous support in this life journey. Your smile never makes me feel low and always motivates me to work harder. Now, our daughter, Avni, further motivates me to move ahead for success. I thank my parents, Tekraj Pokhrel and Maiya Devi Pokhrel, for their continuous support and belief. Today, I have finally achieved my mother's dream. Also, I would like to thank my brother Rahul, my sister Jenny, and my brother-in-law Sameer. Lastly, I want to thank my friends back in my country who always believed I would be a good physicist one day. Though I haven't reached it yet, I have succeeded in taking the first step.

# TABLE OF CONTENTS

Page

Page

# LIST OF TABLES

Table Page

# LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

# CHAPTER 1

# INTRODUCTION

From the earliest days of civilization, humans have always been curious about what makes the matter around us. This question can be dated back to thousands of years. Ancient Greeks used to think that all the matter around us was made up of fire, water, air, and earth. They used to believe that everything is made up of specific proportions of them. However, the somewhat modern and semi-scientific viewpoint on constituents of matter could be dated back to Democritus (∼460 BCE). His ideas were based on simple observation. For example, if we cut an apple into smaller pieces and keep cutting it further, we will reach a certain point where we will have smaller apples, which are further indivisible. According to him, they are the smallest units of apples. He called them 'atomos' and assumed they were fundamental constituents of matter, and each material had unique, indivisible components. The philosophers of the time did not support it. So, this concept of matter being composed of tiny indivisible particles vanished until it was rediscovered 1000s of years later.

In the 17th and 18th centuries, Chemistry, as a field of science, was continuously growing. Several new developments were happening in the field. Concepts of chemical reactions were developed. It showed that through chemical reactions, one particle could transform into another. These new experiments indicated matter as things composed of further tiny particles. These observations led John Dalton to postulate his first modern "atomic theory" [1], according to which matter is composed of tiny indivisible particles called atoms, and all atoms of the same element are identical. Then, after J.J Thom*p*son discovered negatively charged electrons using cathode ray experiments, subatomic constituents of an atom were verified. After this, significant development in understanding the basic constituent of matter came through physicist Ernest Rutherford when he discovered atomic nuclei and protons through his famous alpha particle scattering experiment [2]. This experiment demonstrated the atom as a positively charged nuclei surrounded by electrons. Later, after the discovery of Neutrons by Chadwick [3], it was known that atomic nuclei are composed of neutral particles (neutrons) alongside positively charged particles (protons).

During the 1930s, Cockroft and Walton invented particle accelerators [4], which led to a new field in physics. With the help of these particle accelerators, which could produce

highly energetic particles, a significant breakthrough came when Robert Hofstadter, through elastic ep scattering [5] at Stanford University, showed that nucleons are not homogeneous as they were supposed to be but have internal structures and have charge and magnetic moment distributions. With similar other experiments, various sub-atomic particles were also discovered.

Later, during the 1970s, very high-energy accelerators were developed, and similar scattering experiments were performed at very high beam energy at SLAC and other experiment facilities. As very high energies were involved, the final state would produce a lot of different particles. This type of scatttering is called "Deep Inelastic Scattering (DIS)." So, the early results were not explainable by the available theories at that time. Hence, Feynman and Bjorken introduced the "Parton Model" to explain such results. The scattering results could be easily explained if we assumed that electrons scattered from tiny point-like particles inside the nucleons called partons. These could explain the DIS results with a very high degree of precision and accuracy. Glenn-Mann [6] and Zweigg [7] had independently proposed the quark model in the early 1960s, which was later identified to be equivalent to the partons that the parton model describes. Further experiments have verified that quarks are point-like and do not possess a substructure. The interaction between the quarks is explained by Quantum Chromodynamics, which is the theory of strong force that binds quarks.

So, from Rutherford's first alpha particle scattering experiment to the present experiments at different medium-to-high-energy labs around the world, scattering experiments have been the most powerful tool for revealing the nuclear constituent of a nucleon matter.

# CHAPTER 2

## PHYSICS MOTIVATION AND THEORETICAL BACKGROUND

At present, the Standard Model (SM) is the best theory for describing the universe's building blocks in the current nuclear and particle physics scenario. It describes the fundamental particles in the universe in two broad categories: quarks and leptons. It describes the interaction between these two groups of particles in terms of another group of particles called bosons.

The SM encompasses three fundamental forces of nature: Electromagnetic interaction, Strong interaction, and Weak interaction. In the SM, the particles involved in the interactions are fermions, whereas the mediators of the interactions of those fermions are bosons. For instance, Electromagnetism, an interaction between electric charges, is mediated by photons involving electric and magnetic fields. On the other hand, the strong force binds atomic nuclei together and is mediated by gluons, while the weak force is mediated by W and Z bosons. Still, the SM needs to incorporate gravitational interaction, which is still one of the unsolved problems in physics.

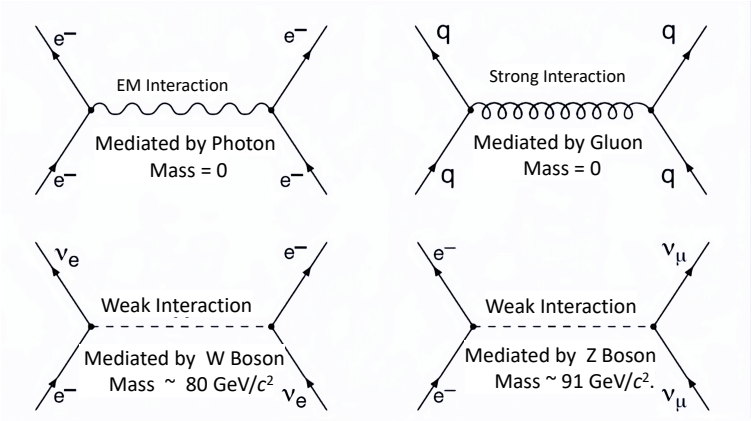



FIG. 1. Fenyman Diagrams for interaction between fundamental particles through exchange of vector bosons.

The fermions are a group of particles that are a combination of quarks and leptons. They follow Fermi - Dirac statistics and obey the Pauli exclusion principle. The fermions have half-odd integer spin. The quark is a fundamental constituent of matter. Quarks combine to form baryons and mesons, collectively called hadrons. The combination of three quarks is a baryon, whereas a quark-antiquark pair is called a meson. The quarks have a non-integral electric charge. Altogether, there are six quarks known as flavors: up, down, strange, charm, bottom, and top. Similarly, a lepton is an elementary particle of half-integer spin. There are two main classes of leptons: charged leptons (also known as the electron-like leptons or muons), including the electron, muon, and tauon, and neutral leptons, better known as neutrinos. The main difference between quarks and leptons is that they are unlike quarks. However, leptons are not subject to strong interaction.

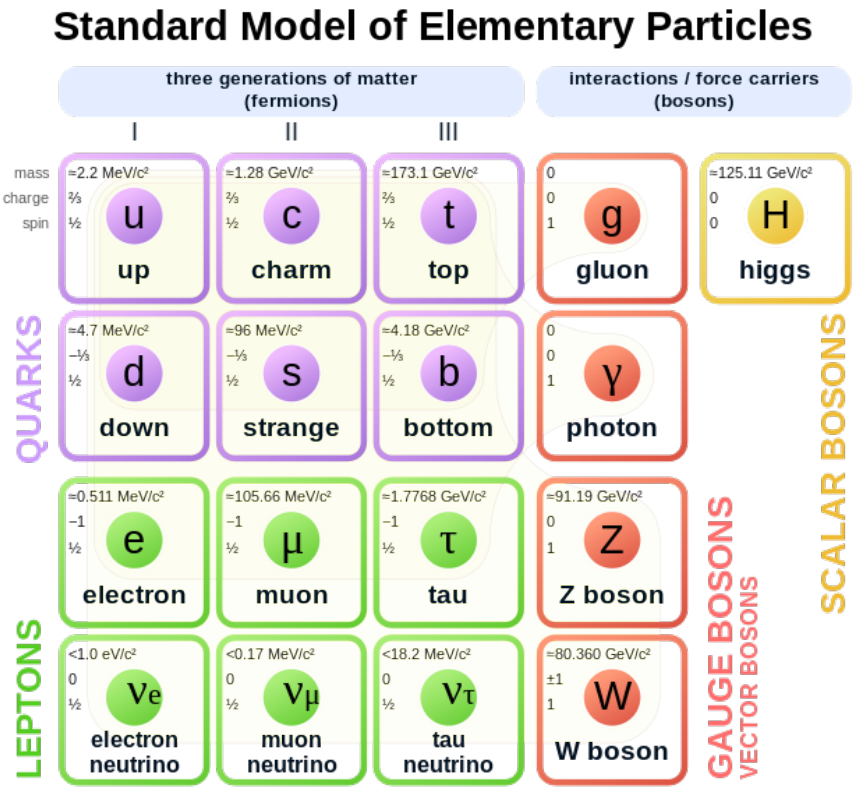



FIG. 2. Standard Model of Particle Physics [8].

Though quarks are the most fundamental division of hadrons, free quarks can never be observed due to color confinement. The "color" of quarks refers to the charge in QCD, which is similar to the electric charge in QED. This is explained by Quantum Chromodynamics (QCD), the theory of strong interaction. According to QCD, as quarks are pulled apart, the gluon field energy between them increases, unlike gravitational or electromagnetic forces, which follow an inverse distance square law. When the energy in the gluon field is large enough, it breaks into a quark-antiquark pair, which binds to the original quarks. This process prevents the isolation of quarks. Since it is impossible to observe an isolated quark, we infer about it through the bound states of quarks and gluons, mainly protons and neutrons. Because of this, it is challenging to understand QCD in the confinement region. Scattering experiments have been incredibly successful in observing these things. Inclusive electron-nucleon scattering is one of the significant scattering processes that has helped us understand the dynamics of quarks and gluons.

Scattering experiments have been widely helpful in understanding fundamental particles like protons and neutrons that are bound states of quarks and gluons. From the discovery of protons to the latest discovery of the Higgs boson [9], scattering experiments have helped understand fundamental interactions in nature.

## 2.1 SCATTERING EXPERIMENTS

Scattering experiments help understand the internal structure of nucleons or the interaction between different particles. In high-energy physics experiments, where we are interested in studying particles of very small sizes (e.g., the size of protons), a very high-energy beam of particles is needed. Electron scattering is very useful in understanding or investigating these things. Since electrons are point-like objects without any substructure, electron scattering experiments are instrumental in studying small things. The interaction between electrons and nucleons would be purely electromagnetic with a coupling constant, $\alpha = \frac{1}{137}$. This interaction is understood and calculated with high accuracy up to several decimal digits. Hence, a very small is played by higher-order effects.

### 2.1.1 KINEMATICS OF SCATTERING

Let us consider an electron beam four-momentum $p = (E, \vec{k})$ scattering off a proton with a rest mass $m_p$. The electron scatters with scattering angle $\theta$ in the lab frame, and after scattering, its energy changes to $E'$ with an exchange of a virtual photon $q$. The new four-momentum of the electron after scattering will be $p' = (E', \vec{k'})$. Under the one

photon exchange approximation, the scattering will be mediated by a virtual photon of four-momentum $q = (E - E', \vec{q})$, with $\vec{q} = \vec{k} - \vec{k'}$. Using this information, we can construct several Lorentz invariant quantities that help in explaining the scattering process,

$$\nu = E - E' = \frac{p \cdot q}{M_p}, \tag{1}$$

$$y = \frac{p \cdot q}{p \cdot k} = \left(\frac{\nu}{E}\right)_{\text{lab}}, \tag{2}$$

$$Q^2 = -q^2 \approx 4EE'(1 - \cos\theta), \tag{3}$$

$$x = \frac{Q^2}{2p \cdot q} = \frac{Q^2}{2M_p \cdot \nu}, \tag{4}$$

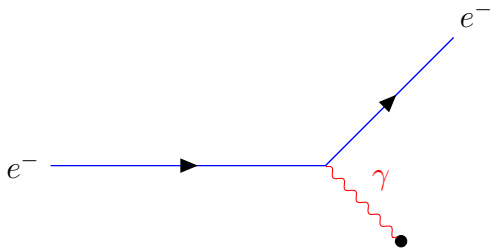
$$W^2 = (p + q)^2 = M_p^2 + Q^2 \left(\frac{1}{x} - 1\right). \tag{5}$$

So, if we try to probe a nucleon with an electron beam, it interacts with the nucleon by exchanging a virtual photon with wavelength $\lambda$. The wavelength is related to the transferred momentum by,

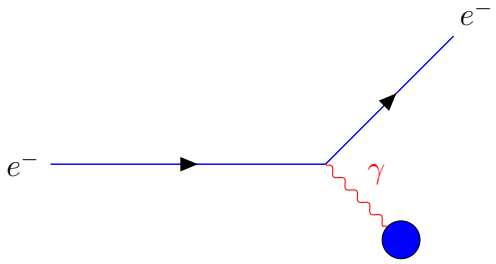
$$\lambda \sim \frac{1}{\sqrt{Q^2}}. \tag{6}$$

Therefore, larger $Q^2$ corresponds to higher resolution.

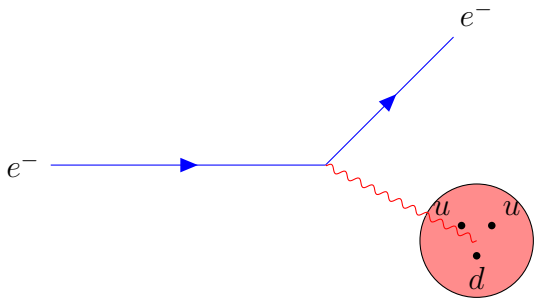
- At very low energies where $\lambda >> r_p$, it is equivalent to scattering from spin-less point-like objects.



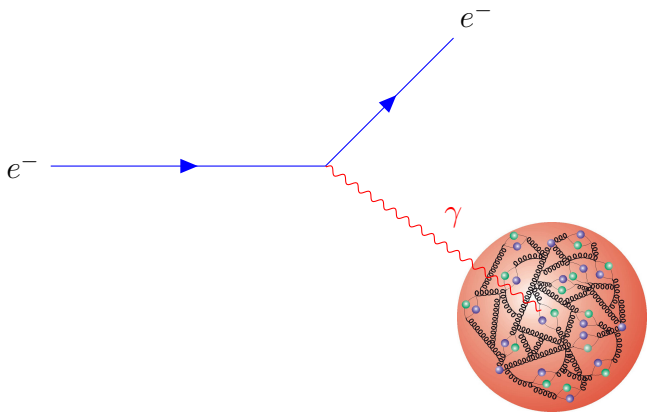

- At low energies $\lambda \sim r_p$, it is equivalent to scattering from extended charged objects; we can see charge distributions through form factors.



- At high energies $\lambda < r_p$, it is equivalent to scattering from individual constituents inside protons, which are quarks. The resolution increases, and we start observing valence quarks.




- At very high energies $\lambda <<< r_p$, our resolution of electron probe increases such that we start seeing protons as a sea of virtual quarks and gluons.




For elastic scattering, the invariant mass $W$ is equal to the mass of the scattering target (proton), and therefore we obtain,

$$2m_p\nu - Q^2 = 0 \implies x = 1 \qquad \text{(Elastic scattering)}, \tag{7}$$

$$2m_p\nu - Q^2 > 0 \implies 0 < x < 1 \qquad \text{(Inelastic scattering)}. \tag{8}$$

Whereas if,

$$Q^2 = -q^2 \gg m_p^2 \sim 1 \text{ GeV}^2 \qquad \textbf{Deep}, \tag{9}$$

$$W^2 = (p+q)^2 > m_p^2 \qquad \textbf{Inelastic}. \tag{10}$$

## 2.2 ELASTIC SCATTERING

At low energies, the dominant scattering process is the elastic scattering ($e^- + p \to e^- + p$), where the proton remains intact after the scattering. One of the earliest experiments, The

Rutherford Gold Foil Experiment, is the best example of this type of scattering, which proved the existence of a dense atomic nucleus in the center of the atom surrounded by a negative charge. The cross-section for such a process can be calculated by considering the scattering of non-relativistic electrons from the point-like target ($\beta_e\gamma_e << 1$). The scattering cross-section can be written as,

$$\left(\frac{d\sigma}{d\Omega}\right)_{\text{Rutherford}} = \frac{\alpha^2}{4E^2\sin^4\left(\frac{\theta}{2}\right)}, \tag{11}$$

where $\alpha$ is the fine structure constant in QED, $E$ is the energy of incident electron and $\theta$ is the scattering angle. This Rutherford cross-section could be obtained by considering the scattering of non-relativistic particles in the static columb potential of the nucleus. We do not need to take into account the intrinsic magnetic moments of electrons or protons. So, we can conclude that in the non-relativistic limit, only the charge of interacting particles matter.

For the Rutherford scattering, we assumed that the target nuclei don't recoil and the incoming electron is non-relativistic ($E_k < m_e$). For the relativistic electron electrons and still neglecting the target recoil, the new scattering cross section known as Mott's cross-section is given by,

$$\left(\frac{d\sigma}{d\Omega}\right)_{\text{Mott}} = \frac{\alpha^2}{4E^2\sin^4\left(\frac{\theta}{2}\right)}\cos^2\left(\frac{\theta}{2}\right). \tag{12}$$

We see a disagreement between Rutherford cross section (Eq. (11)) and Motts cross section(Eq. (12)). There is an extra factor of $\cos^2\theta/2$ in the Mott cross section, which comes from the relativistic spinor structure of the electron. If we take into account non-point-like charge distributions, the Mott cross section will be given by Eq. (13), where $|F(q^2)|^2$ gives information on the proton's charge distribution.

$$\left(\frac{d\sigma}{d\Omega}\right)_{\text{Mott}} = \frac{\alpha^2}{4E^2\sin^4\left(\frac{\theta}{2}\right)}\cos^2\left(\frac{\theta}{2}\right)|F(q^2)|^2. \tag{13}$$

For electron scattering of spin 1/2 protons, two form factors exist; one enables the proton's spin state to remain the same at all times, and the other allows the spin state to flip during the scattering process. These are Dirac form factor ($F_1(Q^2)$) and Pauli form factor ($F_2(Q^2)$). We can determine $F_1(Q^2)$ and $F_2(Q^2)$ by determining following matrix element in an electromagnetic current,

$$\langle p'|j^\mu(0)|p\rangle = \bar{N}(p')\left[F_1(Q^2)\gamma^\mu + F_2(Q^2)\frac{i\sigma^{\mu\nu}q_\nu}{2M_N}\right]N(p). \tag{14}$$

The expression $\langle p'|j^\mu(0)|p\rangle$ describes how an initial nucleon (like a proton or neutron) with momentum $p$ interacts with an electromagnetic field to transition to a final nucleon state with momentum $p'$. In this expression, $\langle p'|$ and $|p\rangle$ are the initial and final states of the nucleon, represented as spinors, with $p$ and $p'$ being the momenta of the initial and final nucleon states, respectively. The term $j^\mu(0)$ is the electromagnetic current operator, which couples to the nucleon in the interaction, where the index $\mu$ represents the four components of the current and the argument 0 denotes the interaction at time $t = 0$. The term $F_1(Q^2)\gamma^\mu$ involves the electric form factor $F_1(Q^2)$, which describes the distribution of electric charge inside the nucleon, and $\gamma^\mu$, the gamma matrix, which couples the nucleon to the photon in the vector current. The second term, $F_2(Q^2)\frac{i\sigma^{\mu\nu}q_\nu}{2M_N}$, includes the magnetic form factor $F_2(Q^2)$, which describes the nucleon's magnetic moment, and $\sigma^{\mu\nu}$, the spin tensor that encodes spin-dependent effects. Here, $q_\nu$ is the four-momentum transfer, defined as $q = p' - p$, which describes the momentum of the virtual photon in the interaction. The denominator $2M_N$ normalizes the magnetic term by the nucleon mass $M_N$. Both form factors, $F_1(Q^2)$ and $F_2(Q^2)$, depend on the momentum transfer $Q^2$, which represents the difference between the initial and final nucleon momenta. The electric form factor $F_1(Q^2)$ describes how the nucleon's internal charge responds to the electromagnetic field, while the magnetic form factor $F_2(Q^2)$ accounts for the nucleon's magnetic response and spin effects. In the limiting case, $Q^2 \to 0$, this represents zero virtual photons exchanged between electron and proton. This means we are probing the proton's static magnetic and electric charge using a very high $\lambda$ beam or large distance scale. Hence, in this limiting case, the form factors normalize to their electric charge and magnetic moments. We have the following relations in the limiting case,

$$F_{1p} = 1 \quad (1 \text{ unit of electric charge}), F_{2p} = \kappa_p, \tag{15}$$

$$F_{1n} = 0 \quad (\text{no electric charge}), F_{2n} = \kappa_n, \tag{16}$$

$\kappa_p = \mu_p - 1$ and $\kappa_n = \mu_n$, are the anomalous magnetic moments of proton and neutron, respectively, with values $\mu_p = 2.7928\mu_N$ and $\mu_n = -1.9130\mu_N$, where $\mu_N$ is given by,

$$\mu_N = \frac{e\hbar}{2m_p} = 3.1525 \times 10^{-14}\ \text{MeV} \cdot \text{T}^{-1}. \tag{17}$$

We can also express a linear combination of these form factors to get a new electric and

magnetic Sachs form factor [10],

$$G_E(Q^2) = F_1(Q^2) - \frac{Q^2}{(2M_N)^2} F_2(Q^2), \tag{18}$$

$$G_M(Q^2) = F_1(Q^2) + F_2(Q^2). \tag{19}$$

These equations are valid for protons and neutrons with their respective form factors. Now, these form factors represent the electric and magnetic charge distribution of protons and neutrons. In the limit of $Q^2 \to 0$, the Sachs form factors are,

$$G_{E_p}(0) = 1, G_{M_p}(0) = 1, G_{M_n}(0) = 1, G_{M_p}(0) = 1. \tag{20}$$

The differential scattering cross section for elastic scattering of electrons off the protons is given by the Rosenbluth formula,

$$\frac{d\sigma}{d\Omega} = \left(\frac{\alpha^2}{4E^2 \sin^4\left(\frac{\theta}{2}\right)}\right)\left(\frac{E'}{E}\right)\left[\frac{G_E^2 + \tau G_M^2}{1+\tau}\cos^2\left(\frac{\theta}{2}\right) + 2\tau G_M^2 \sin^2\left(\frac{\theta}{2}\right)\right], \tag{21}$$

with $\tau = \frac{Q^2}{4M^2}$. If we introduce the degree of polarization of the virtual photon $\epsilon = (1+2(1+\tau)\tan^2(\theta/2))^{-1}$, we can simplify this cross-section equation,

$$\frac{d\sigma}{d\Omega} = \left(\frac{d\sigma}{d\Omega}\right)_{\text{Mott}} \cdot \left[\frac{\epsilon G_E^2 + \tau G_M^2}{\epsilon(1+\tau)}\right]. \tag{22}$$

From Eq. (22), it is clear that Sachs form factors eliminate mix terms from the cross section and only depend on the square of $G_E$ and $G_M$. To extract the form factors from measured cross-sections, the Rosenbluth separation is usually used [10]. Fig. 3 shows the world data for $G_{E_p}$ and $G_{M_p}$ obtained by the Rosenbluth separation method. These were calculated in 1970s as the ratios of $G_{E_p}/G_D$ and $G_{M_p}/\mu_p G_D$, where $G_D$ is the dipole form factor and given by the Eq. (23).

$$G_D(Q^2) = \frac{1}{\left(1 + Q^2/0.71\ \text{GeV}^2\right)^2}. \tag{23}$$

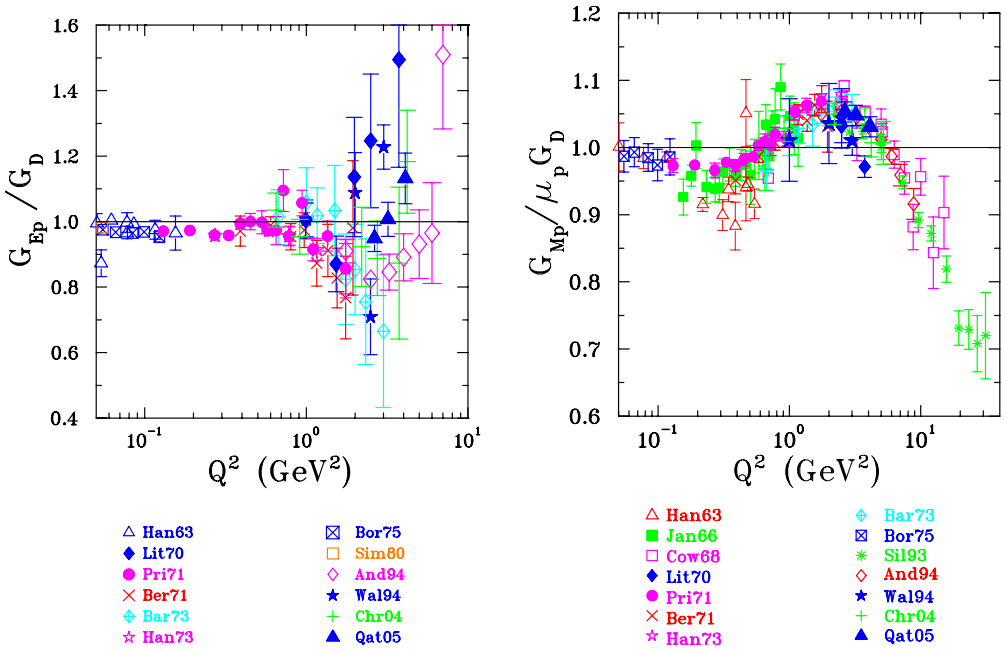



FIG. 3. World data points for $G_E$ and $G_M$ obtained by the Rosenbluth separation [11].

### 2.3 DEEP INELASTIC SCATTERING (DIS)

During inelastic scattering, the proton can break up into its constituent quarks, which then form a hadronic jet. At high $q^2$, this is known as deep inelastic scattering (DIS). The invariant mass, $M_x$, of the final state hadronic jet is given by Eq. (10). The system X will be a baryon. The proton is the lightest baryon, therefore $W > M_p$. Since $M_x = M_p$, $q^2$ and $\nu$ are two independent variables in DIS, and it is necessary to measure the beam energy $E$, scattered electron energy $E'$ and scattering angle $\theta$ in the lab frame to determine the full kinematics. Hence, the scattering cross section in DIS will be a double differential. Assuming one photon exchanged, the inclusive differential cross section for DIS can be written as,

$$\frac{d^2\sigma}{d\Omega\, d\nu} = \frac{1}{2M} \cdot \frac{E'}{E} \cdot |\mathcal{M}|^2, \tag{24}$$

where,

$$|\mathcal{M}|^2 = \frac{\alpha^2}{q^4} L^{\mu\nu} T_{\mu\nu}. \tag{25}$$

Here $L_{\mu\nu}$ is the leptonic tensor and $T^{\mu\nu}$ is the hadronic tensor. The leptonic tensor describes the coupling between the virtual photon and the electron. It can be written in terms of Dirac spinors ($u$) and the gamma matrices ($\gamma^\mu$) as,

$$L^{\mu\nu}(k, s_l; k') = \sum_{s'} \left[\bar{u}(k', s'_l)\gamma u(k, s_l)\right] \left[\bar{u}(k', s'_l)\gamma u(k, s_l)\right], \tag{26}$$

Summing over all possible spin states $s'$ of the final state electron, the leptonic tensor can be split into a symmetric part ($S$) and an antisymmetric part ($A$), under the exchange of Lorentz indices $\mu$ and $\nu$. The symmetric part ($S$) is spin independent. It is given by

$$L^{(S)}_{\mu\nu}(k; k') = 2\left(k_\mu k'_\nu + k_\nu k'_\mu + g_{\mu\nu}(m^2 - k_\sigma k'^\sigma)\right). \tag{27}$$

The spin dependent anti-symmetric part ($A$) is given by

$$L^{(A)}_{\mu\nu}(k, s; k') = 2m\,\epsilon_{\mu\nu\alpha\beta}s^\alpha(k^\beta - k'^\beta), \tag{28}$$

where $\epsilon_{\mu\nu\alpha\beta}$ is the totally antisymmetric Levi-Civita tensor with $\epsilon_{0123} = +1$, and $m$ is the mass of an electron. The hadronic tensor can also be divided into symmetric and antisymmetric parts, the former being spin-independent and the latter being dependent on the spin of the target nucleon ($S$). It can be written as,

$$W_{\mu\nu}(q; P, S) = W^{(S)}_{\mu\nu}(q; P) + iW^{(A)}_{\mu\nu}(q; P, S). \tag{29}$$

The symmetric part of the hadronic tensor can be parameterized in terms of two spin-independent inelastic form factors, $W_1$ and $W_2$, and the asymmetric one in terms of two spin-dependent form factors, $G_1$ and $G_2$,

$$\begin{aligned}\frac{1}{2M}W^{(S)}_{\mu\nu}(q; P) = &\left(-g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2}\right) W_1(P\cdot q, q^2) \\ &+ \frac{1}{M^2}\left(P_\mu - \frac{P\cdot q}{q^2}q_\mu\right)\left(P_\nu - \frac{P\cdot q}{q^2}q_\nu\right) W_2(P\cdot q, q^2),\end{aligned} \tag{30}$$

$$\begin{aligned}\frac{1}{2M}W^{(A)}_{\mu\nu}(q; P, S) = &\epsilon_{\mu\nu\alpha\beta}q^\alpha \left[\frac{MS^\beta}{M^2}G_1(P\cdot q, q^2)\right. \\ &\left.+\frac{1}{M^2}\left((P\cdot q)S^\beta - (S\cdot q)P^\beta\right) G_2(P\cdot q, q^2)\right].\end{aligned} \tag{31}$$

These inelastic form factors are Lorentz-invariant and generally expressed in dimensionless

quantities that depend on two DIS variables, $x$ (Eq (4)) and $Q^2$ (Eq. (3)).

$$F_1(x,Q^2) = MW_1(P\cdot q, q^2), \tag{32}$$

$$F_2(x,Q^2) = \nu W_2(P\cdot q, q^2), \tag{33}$$

$$g_1(x,Q^2) = \frac{(P\cdot q)^2}{\nu} G_1(P\cdot q, q^2), \tag{34}$$

$$g_2(x,Q^2) = \nu(P\cdot q) G_2(P\cdot q, q^2). \tag{35}$$

The functions $F_1$ and $F_2$ are independent of the spin, so they are usually referred to as **unpolarised structure functions**, whereas the spin-dependent functions $g_1$ and $g_2$ are known as **polarised structure functions**. In the pQCD framework, the structure functions cannot be evaluated analytically using the first principles. Hence, they can only be determined experimentally. Now, writing the symmetric and antisymmetric part of the hadronic tensor $T_{\mu\nu}$ in Eq. (30) and Eq. (31), in terms of these new structure functions, we get,

$$\begin{aligned} W^{(S)}_{\mu\nu}(q;P) = 2&\left(-g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2}\right) F_1(x,Q^2) \\ &+ \frac{2}{P\cdot q}\left(P_\mu - \frac{P\cdot q}{q^2}q_\mu\right)\left(P_\nu - \frac{P\cdot q}{q^2}q_\nu\right) F_2(x,Q^2), \end{aligned} \tag{36}$$

$$W^{(A)}_{\mu\nu}(q;P,S) = \frac{\epsilon_{\mu\nu\alpha\beta}q^\alpha}{2MP\cdot q}\left[S^\beta g_1(x,Q^2) + \left(S^\beta - \frac{S\cdot q}{P\cdot q}P^\beta\right) g_2(x,Q^2)\right]. \tag{37}$$

The scattering of an electron from a nucleon is an electromagnetic (EM) process and conserves parity. Hence, the terms with the same symmetry can only contribute to the cross-section in Eq. (25). Using this fact, the cross section in terms of hadronic and leptonic tensors reduces to,

$$\frac{dE'}{d\Omega} = \frac{\alpha^2}{2MQ^4}\frac{E'}{E}\left[L^{(S)}_{\mu\nu}W^{\mu\nu}_{(S)} - L^{(A)}_{\mu\nu}W^{\mu\nu}_{(A)}\right]. \tag{38}$$

Finally, if we average over all the spin in the initial state and sum over all the spin states in the final state, then only the spin-independent part contributes to the cross-section, which is the symmetric part of leptonic and hadronic tensors. Hence, the unpolarised cross-section in terms of the spin-independent part of leptonic and hadronic tensor is,

$$\begin{aligned} \frac{d^2\sigma_{\text{unpol}}}{dE'd\Omega} &= \frac{\alpha^2}{2MQ^4}\frac{E'}{E}L^{(S)}_{\mu\nu}W^{\mu\nu}_{(S)} \\ &= \left(\frac{d^2\sigma}{dE'd\Omega}\right)_{\text{Mott}}\cdot\left[\frac{2M}{\nu}F_1(x,Q^2)\tan^2\left(\frac{\theta}{2}\right) + \frac{1}{\nu}F_2(x,Q^2)\right]. \end{aligned} \tag{39}$$

Alternatively, we can write the cross-section in terms of inclusive variables $x$ and $Q^2$ as,

$$\frac{d^2\sigma}{dx\,dQ^2} = \frac{4\pi\alpha^2}{xQ^4}\left[y^2 \cdot 2xF_1(x,Q^2) + (1-y)F_2(x,Q^2)\right]. \tag{40}$$

## 2.4 QUARK PARTON MODEL

Until now, we tried to explain deep-inelastic scattering of electrons as a kinematic process where the target proton breaks up with $W >>> m_p$, and much hadronic debris is produced. This interpretation could be simplified if we explain this process in another Lorentz frame, called the "Breit Frame". Since this is an arbitrarily chosen frame, the underlying physics process does not affect our chosen frame. So, in this frame, the target proton is moving very fast, such that it has only momenta along the longitudinal direction, i.e., the transverse momenta and rest mass of the proton can be neglected. This is the underlying principle of the quark-parton model introduced by Fenyman and Bjorken. In this model, the constituents of protons are called partons, which were later identified as quarks and gluons. So, in this partonic frame, we can decompose a nucleon into independent partons, and the total interaction is the incoherent sum of all the individual interactions. However, one thing to note is that these approximations are valid as long as the interaction time between the virtual photon and the partons is so short that the interactions between the partons can be safely neglected. This is called **impulse approximation**.

In the DIS, the spatial resolution is given by the wavelength of the exchanged virtual photon, . This is not a Lorentz invariant, but it depends on the reference frame. In the Lab frame, where $q_0 = \nu/c$ is,

$$\lambda = \frac{\hbar}{|\vec{q}|} = \frac{\hbar c}{\sqrt{\nu^2 + Q^2c^2}} \approx \frac{\hbar c}{\nu} = \frac{2Mx\hbar c}{Q^2}. \tag{41}$$

In the Breit Frame, this equation will simplify to,

$$\lambda = \frac{\hbar}{|\vec{q}|} = \frac{\hbar c}{\sqrt{Q^2}}. \tag{42}$$

Hence, in the Breit Frame, the quantity $Q^2$ corresponds to the spatial resolution with which the nucleon structure can be resolved.

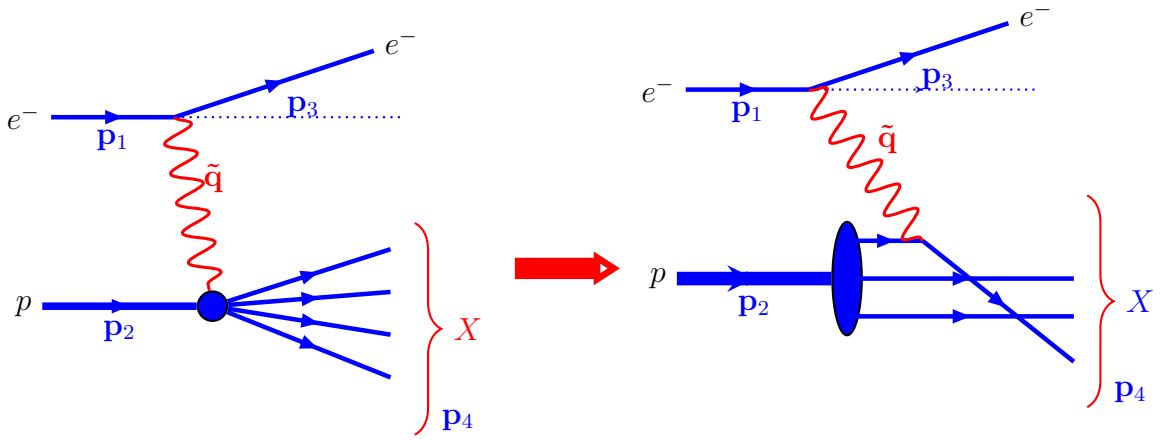



FIG. 4. Left: Scattering from proton with structure functions. Right: Scattering from individual quarks within proton.

In the quark-parton model, the interaction is the elastic scattering of spin half quark inside the proton. In the Infinite momentum frame, for proton with four momenta, $p_2 = (E_2, 0, 0, E_2)$, consider a quark (parton) carrying $\zeta$ fraction of the proton's four-momentum. Hence its four momentum will be $\zeta p_2 = (\zeta E_2, 0, 0, \zeta E_2)$.

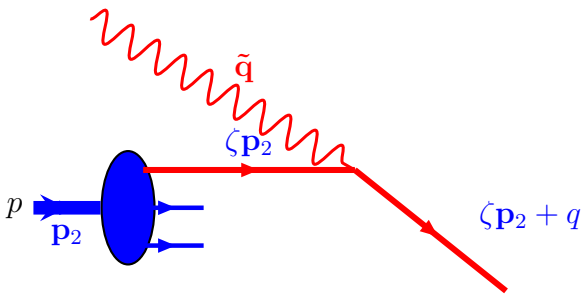



FIG. 5. ep scattering at partonic level.

After the interaction between the virtual photon and struck quark, the four-momenta of

the struck quark will be $\zeta p_2 + q$.

$$(\zeta p_2 + q)^2 = m_q^2, \tag{43}$$

$$\implies \zeta^2 p_2^2 + q^2 + 2\zeta p_2 \cdot q = m_q^2. \tag{44}$$

For any particle with four-momentum $p^\mu$, we have $p^\mu p_\mu = p^2 = m^2$. Using this fact for the quark with mass $m_q$ and four momenta $\zeta p_2$, we get $\zeta^2 p_2^2 \approx m_q^2 = 0$ (In the Breit frame, mass and longitudinal momenta of quarks can be neglected.) Using this fact in Eq. (44) we get,

$$q^2 + 2\zeta p_2 \cdot q = 0 \implies \zeta = -\frac{q^2}{2p_2 \cdot q} = \frac{Q^2}{2p_2 \cdot q} = x, \quad \text{using (3) and (4).} \tag{45}$$

Hence, in the Breit frame, Bjorken-x, $x$, is the fraction of momentum carried by the struck quark. Now, in terms of the proton's momentum, using Fig. 4,

$$s = (p_1 + p_2)^2 \cong 2p_1 \cdot p_2, \quad \text{with} \quad y = \frac{p_2 \cdot q}{p_2 \cdot q_1}. \tag{46}$$

For the interaction of underlying quarks,

$$s^q = (p_1 + xp_2)^2 = 2xp_1 \cdot p_2 = xs, \tag{47}$$

$$y_q = \frac{p_q \cdot q}{p_q \cdot q_1} = \frac{xp_2 \cdot q}{xp_2 \cdot q_1} = y, \tag{48}$$

$$x_q = 1 \quad \text{(Elastic Scattering from individual quarks).} \tag{49}$$

## 2.5 PARTONIC INTERPRETATION OF STRUCTURE FUNCTIONS

We have shown that the deep inelastic scattering of an electron from a proton is an elastic scattering of electrons from individual quarks within the proton. Now, for $eq \to eq$ (electron quark scattering), the Lorentz invariant elastic scattering cross section in the relativistic limit is given by,

$$\frac{d\sigma}{dq^2} = \frac{2\pi\alpha^2 e_q^2}{q^4}\left[1 + \left(1 + \frac{q^2}{s^q}\right)^2\right] \quad \text{where } e_q \text{ is charge of quark,} \tag{50}$$

$$\text{Using} \quad -q^2 = Q^2 = (s^q - m_q^2)x_q y_q \implies \frac{q^2}{s^q} = -y_q = -y, \tag{51}$$

$$\frac{d\sigma}{dQ^2} = \frac{2\pi\alpha^2 e_q^2}{Q^4}\left[1 + (1-y)^2\right], \tag{52}$$

$$\frac{d\sigma}{dQ^2} = \frac{4\pi\alpha^2 e_q^2}{Q^4}\left[(1-y) + \frac{y^2}{2}\right], \tag{53}$$

This is the expression for the differential cross-section for elastic electron scattering of a quark carrying a fraction $x$ of the proton's momentum. The introduction of the parton distribution function $q^p$ such that $q^p(x)dx$ represents the total number of quarks with momentum fraction between $x$ and $x + \mathrm{d}x$. The differential cross section for any particular quark within the proton with a momentum fraction between $x$ and $x + \mathrm{d}x$ is

$$\frac{d\sigma}{dQ^2} = \frac{4\pi\alpha^2 e_q^2}{Q^4}\left[(1-y) + \frac{y^2}{2}\right] \cdot e_q^2 q^p(x)dx. \tag{54}$$

To calculate the contributions from all quarks that are inside the proton, we sum this expression for all types of quarks.

$$\frac{d\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[(1-y) + \frac{y^2}{2}\right] \sum e_q^2 q^p(x)dx. \tag{55}$$

Comparing this expression with the electron-proton scattering cross-section in terms of structure-functions in Eq. (40)

$$\frac{d\sigma}{dxdQ^2} = \frac{4\pi\alpha^2 e_q^2}{Q^4}\left[(1-y)\frac{F_2(x,Q^2)}{x} + y^2 F_1(x,Q^2)\right]. \tag{56}$$

Comparing Eq. (55) and Eq. (56), we get the prediction of the structure function in terms of the partonic model,

$$F_2^p(x,Q^2) = 2F_1^p(x,Q^2) = 2\sum e_q^2 q^p(x)\,dx. \tag{57}$$

At fixed values of $x$, the structure function weakly depends on $Q^2$. This is the scaling property of the structure functions, i.e., $F_i(x,Q^2) \to F_i(x)$ [12] in the Bjorken limit, where $Q^2$ and $\nu \to \infty$ with $x$ fixed.

## 2.6 SCALING VIOLATIONS OF THE STRUCTURE FUNCTIONS

Until now, we discussed the partonic model of ep scattering. So, in the partonic model, the scattering is described as scattering from the point-like constituents within the proton. We also discussed that the structure function $F_2$ solely depends on the Bjorken scaling variable $x$. However, the high-precision measurements revealed a small $Q^2$ dependence. The Fig. 8 shows the measurements of $F_2^p$ as a function of $Q^2$ for different values of $x$. The data set spans over large kinematics in $x$ and $Q^2$. We observe that at very small- $x$, the structure function $F_2^p$ increases with increasing $Q^2$, at medium $x$ we see almost no dependence on $Q^2$, whereas, for high $x$, we have weak dependence where it decreases with increasing $Q^2$. This behavior is called scaling violation. Hence, we see that the structure function $F_2$ is not independent of $Q^2$, but is a function of $x$ and $Q^2$ together.

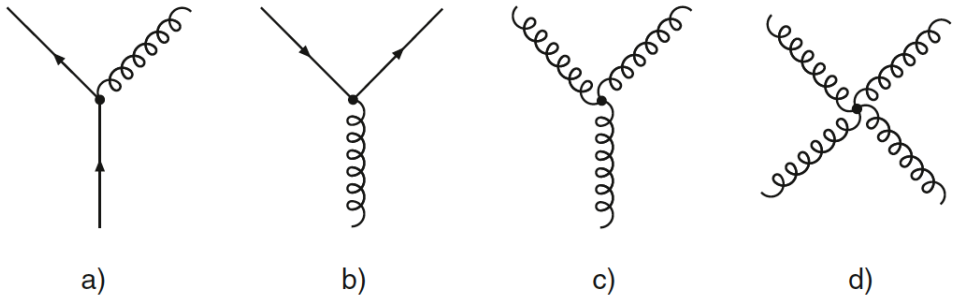



FIG. 6. Fundamental interactions in QCD: from left to right, gluon emission from a quark, gluon decay into a quark-antiquark pair, and gluon self-coupling.

The partonic interpretation of DIS scattering says that the scattering is from point-like constituents inside the nucleons. Also, in the equation (42), we showed that the $Q^2$ can be directly related to the spatial resolution inside the nucleon. It might lead us to think that the quarks have a finite size. However, this is not true. The finite size of quarks does not cause scaling violations. In the QCD framework, which describes the complete interaction of strong force, this can be understood as resulting from constant interactions between the particles inside the nucleon (Fig. 6). Since quarks can emit or absorb gluons, and gluons can either produce quark-antiquark pairs or emit additional gluons, the momentum distribution among the nucleon's constituents constantly changes. A virtual photon can resolve the dimension of the order of $\frac{\hbar c}{\sqrt{Q^2}}$. At small values of $Q^2$, specifically at a starting scale $Q_0^2$, the resolution of the probe is low, and quarks—as well as any potentially emitted gluons—cannot be individually resolved. Consequently, a quark distribution $q(x, Q_0^2)$ is measured, which effectively includes all unresolved radiation. However, the resolution is higher at larger values of $Q^2$; we can see smaller sub-processes like gluon splitting into quark-antiquark and Parton splitting. Hence, it causes a change in the distributions of partons and correspondingly structure functions. So, the effect of these must be taken into account when calculating PDFs. Fig. 6 shows that multiple fundamental interactions could lead to the origin of quarks with momentum fraction $x$ if the resolution is sufficiently high. A quark with a more significant momentum fraction could emit a gluon and decay to a quark with a smaller momentum fraction; a gluon could decay into a quark-antiquark. Each process has some associated probability functions, which are proportional to the strong coupling

constant $\alpha_s(Q^2)$ and splitting function unique to each process. These splitting functions can be calculated perturbatively within the Quantum Chromodynamics (QCD) framework.

This evolution of quark and gluon distributions with $Q^2$ is generally described by a system of coupled differential equations. These equations are described by DGLAP evolution named after their authors, Dokshitzer, Gribov, Lipatov, Altarelli, and Parisi [13–15]. So, by adding the higher order processes in QCD described by DGLAP evolutions to add corrections to the naive Quark Parton model, pQCD gives a rigorous approach at high $Q^2$. The coupling constant $\alpha_s$ is itself $Q^2$ dependent. For the pQCD expansion, the coupling constant must be small. Fig. 7 shows the running coupling constant, $\alpha_s$, with the energy scale in the x-axis. We see that at higher energy, $\alpha_s << 1$, where the perturbation theory is applicable. So, the perturbative approach fails at a lower scale $Q^2 <\sim 1$. The significance of QCD evolution is that once PDFs are known at an energy scale where pQCD is applicable, they can be calculated at any other level which is predicted via DGLAP evolution. Alternatively, the $\alpha_s(Q^2)$ and the gluon PDFs can be extracted using observed scaling violation of $F_2(x, Q^2)$.

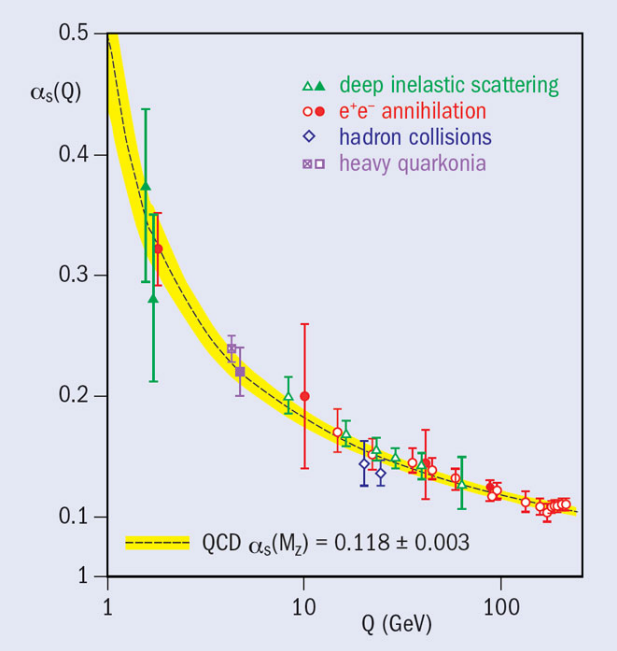



FIG. 7. The value of strong coupling constant $\alpha_s$ at different energy scale.

## 2.7 STRUCTURE FUNCTIONS OF NUCLEONS AND THEIR RATIO

For electron scattering off the proton, ignoring the contribution from strange quarks, we have

$$F_2^p(x) = \sum e_q^2 q^p(x) = x\left(\frac{4}{9}u^p(x) + \frac{1}{9}d^p(x) + \frac{4}{9}\bar{u}^p(x) + \frac{1}{9}\bar{d}^p(x)\right), \tag{58}$$

Similarly, for an electron scattering off a neutron, we have

$$F_2^n(x) = \sum e_q^2 q^n(x) = x\left(\frac{4}{9}u^n(x) + \frac{1}{9}d^n(x) + \frac{4}{9}\bar{u}^n(x) + \frac{1}{9}\bar{d}^n(x)\right). \tag{59}$$

The proton and neutron are "isospin symmetric", i.e., the $u$ quarks in the proton are the same as the $d$ quarks in the neutron and vice versa, i.e.

$$d^n(x) = u^p(x) \quad \text{and,} \quad u^n(x) = d^p(x). \tag{60}$$

So, defining the distributions in terms of the proton,

$$u(x) \equiv u^p(x) = d^n(x); \quad d(x) \equiv d^p(x) = u^n(x), \tag{61}$$

$$\bar{u}(x) \equiv \bar{u}^p(x) = \bar{d}^n(x); \quad \bar{d}(x) \equiv \bar{d}^p(x) = \bar{u}^n(x), \tag{62}$$

gives,

$$F_2^p(x) = x\left(\frac{4}{9}u(x) + \frac{1}{9}d(x) + \frac{4}{9}\bar{u}(x) + \frac{1}{9}\bar{d}(x)\right), \tag{63}$$

$$F_2^n(x) = x\left(\frac{4}{9}d(x) + \frac{1}{9}u(x) + \frac{4}{9}\bar{d}(x) + \frac{1}{9}\bar{u}(x)\right). \tag{64}$$

Integrating the above two equations,

$$\int_0^1 F_2^p(x)\,dx = \int_0^1 x\left(\frac{4}{9}u(x) + \frac{1}{9}d(x) + \frac{4}{9}\bar{u}(x) + \frac{1}{9}\bar{d}(x)\right)\,dx = \frac{4}{9}g_u + \frac{1}{9}g_d, \tag{65}$$

$$\int_0^1 F_2^n(x)\,dx = \int_0^1 x\left(\frac{4}{9}d(x) + \frac{1}{9}u(x) + \frac{4}{9}\bar{d}(x) + \frac{1}{9}\bar{u}(x)\right)\,dx = \frac{4}{9}g_d + \frac{1}{9}g_u, \tag{66}$$

Where,

$$g_u = \int_0^1 \left[xu(x) + x\bar{u}(x)\right]dx, \tag{67}$$

$$g_d = \int_0^1 \left[xd(x) + x\bar{d}(x)\right]dx. \tag{68}$$

Hence, $g_u$ gives the momentum fraction of the proton carried by $u$ and $\bar{u}$ quarks, whereas $g_d$ gives the momentum fraction of the proton carried by $d$ and $\bar{d}$ quarks. Experimentally,

$$\int_0^1 F_2^p(x)\,dx \approx 0.18, \int_0^1 F_2^n(x)\,dx \approx 0.12, \tag{69}$$

$$\implies g_u \approx 0.36 \quad \text{and} \quad g_d \approx 0.12. \tag{70}$$

As expected, the quarks carry half of the proton's momentum, whereas gluons carry the remaining half. Since gluons are charge-less, we cannot probe them through electromagnetic scattering. As we have already explained, DIS scattering is the region with high $Q^2$, and high $Q^2$ corresponds to high resolutions inside the nucleon during scattering. So, we start seeing protons as a complex distribution of partons. The Parton distribution functions $u^p = u(x)$ include a contribution from both valence quarks and sea quarks. So, defining the neutron distributions in terms of the proton,

$$u(x) = u_v(x) + u_s(x), \tag{71}$$

$$d(x) = d_v(x) + d_s(x), \tag{72}$$

$$\bar{u}(x) = \bar{u}_s(x), \tag{73}$$

$$\bar{d}(x) = \bar{d}_s(x), \tag{74}$$

The proton has two up valence quarks and one down valence quark.

$$\int_0^1 u_v(x)\,dx = 2, \quad \int_0^1 d_v(x)\,dx = 1. \tag{75}$$

However, we cannot estimate sea quarks. The sea quarks correspond to a virtual quark-antiquark pair formed by gluon radiation. Eqs. (63) and (64) become

$$F_2^p(x) = x\left(\frac{4}{9}u_v(x) + \frac{1}{9}d_v(x) + S(x)\right), \tag{76}$$

$$F_2^n(x) = x\left(\frac{4}{9}d_v(x) + \frac{1}{9}u_v(x) + S(x)\right), \tag{77}$$

where $S(x)$ refers to the contributions from sea quarks. The suffix "v" from the quark distribution functions will be dropped to simplify and make it more straightforward for readers. So, throughout the text, the quark distribution functions, whenever mentioned, will be referring to valence quarks unless mentioned otherwise. At low $x$, sea quarks dominate. Hence,

$$\frac{F_2^n(x)}{F_2^p(x)} \to 1 \quad \text{as} \quad x \to 0 \quad \text{(experimentally observed)}, \tag{78}$$

At high $x$, the valence quarks dominate. The contribution of sea quarks is negligible. This occurs in the region $x \gtrsim 0.4$. So, to the leading order

$$\frac{F_2^n(x)}{F_2^p(x)} \equiv \frac{(4d(x) + u(x))}{(4u(x) + d(x))} \quad x \to 1. \tag{79}$$

Arranging this equation, we can get

$$\frac{d}{u} \approx \frac{\frac{4F_2^n(x)}{F_2^p(x)} - 1}{4 - \frac{F_2^n(x)}{F_2^p(x)}}. \tag{80}$$

From the definitions of structure functions, $F_2^p$ and $F_2^n$, (see Eqs. (76) and (77)), we see that due to $e_q^2$ weighting, measurements on protons provide constraints on the $u$ quark distribution, whereas precise measurements of the $d$ quark distribution require measurements on neutrons. Scattering experiments with proton targets are more easily achievable because the simplest atom in the universe is hydrogen, and the nucleus of a hydrogen atom is a single proton. Therefore, hydrogen can be used as a target for measuring protons. With more than five decades of experiments at various high-energy physics labs worldwide, we have collected significant measurements on protons over a wide range of kinematics in $x$ and $Q^2$ with comparable precision. Fig. 8 shows the structure function of proton, $F_2^p$ measured at different values of $x$ as a function of $Q^2$.

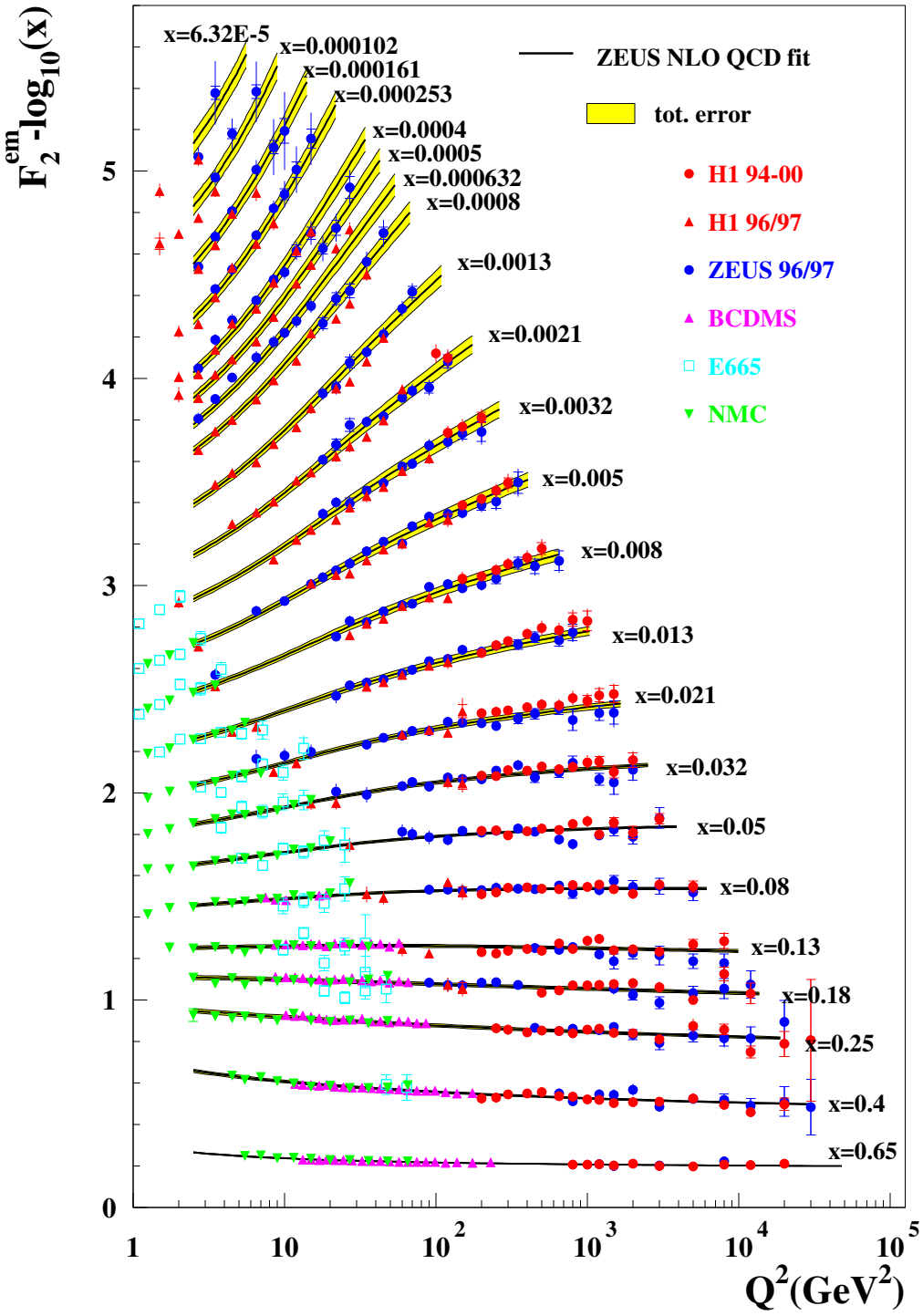



FIG. 8. The structure function of proton, $F_2^p$ measured at different values of $x$ as a function of $Q^2$ [16].

However, obtaining a free neutron target is a complicated task. Free neutrons decay in approximately 15 minutes [17]. Additionally, neutrons are charge-less, making it challenging to use confinement to form dense neutron targets. To collect information on neutrons, we use the simplest neutron-rich nuclei, such as $D_2$, $^3$H, $^3$He, and extract information on neutrons from measurements on these nuclei. However, these are not simple nuclei; they have protons and neutrons in their nuclei. So, to extract information on neutrons from the measurements on these nuclei, we have to use a different physics model that decouples the interaction between the nucleons and subtracts non-neutron backgrounds. Each model has predictions based on how it models the interaction between the protons and neutrons. However, no model is perfect. Each model inherits uncertainties in the model caused by Fermi motion, target fragmentation, off-shell effects of the nucleons, and final state interactions. With these effects, there will be ambiguities in the extracted structure function of neutrons, especially at high $x$, clearly seen in Fig. 9, where we can see the big uncertainty band in the $d/u$ ratio as we go high in $x$.

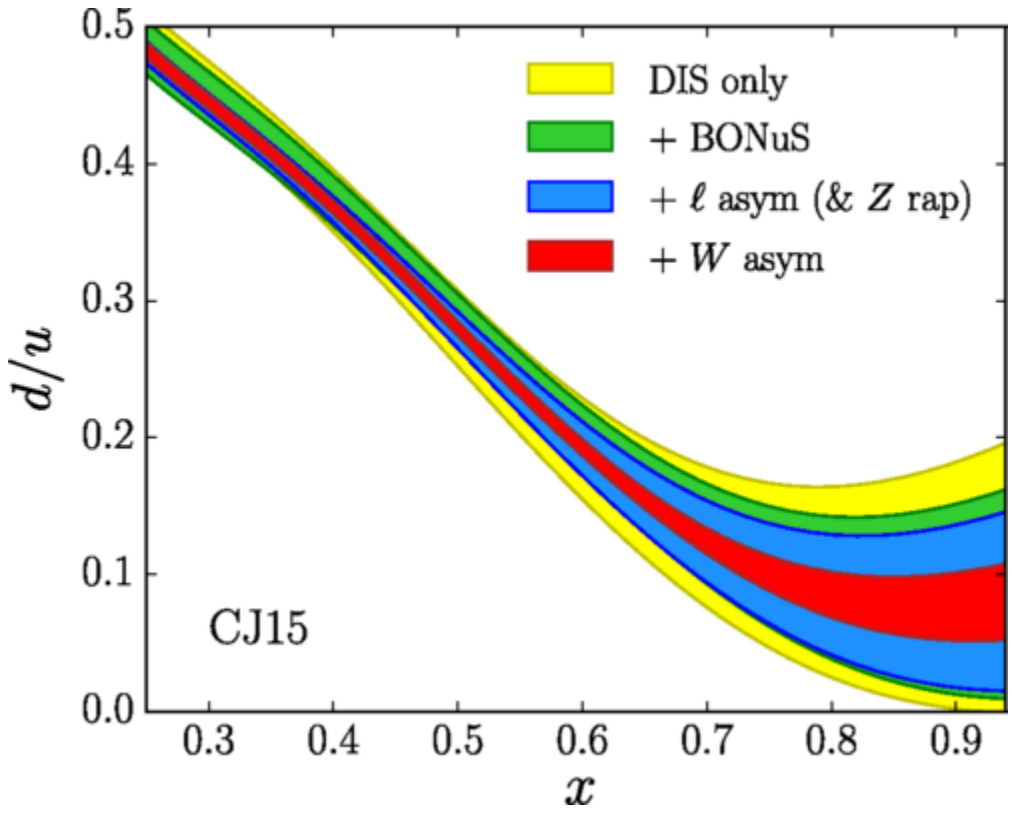



FIG. 9. $d/u$ ratio uncertainties from different experimental datasets [18].

## 2.8 SPECTATOR TAGGING AND THE BONUS12 EXPERIMENT

The main goal of the BONuS12 experiment is to extract the model-independent structure function $F_2^n$ of the free neutron at high $x$. For this, we take measurements of $D_2$, and using the method of spectator tagging, we only select events where we tag a backward-moving spectator proton in coincidence with the scattered electron in the forward direction. In the reaction,

$$eD \rightarrow e' p_s X, \tag{81}$$

we measure the spectator proton $p_s$, and the scattered electron $e'$. However, detecting the spectator proton does not ensure that the scattering of electrons takes place from a free neutron. So, we have to limit the kinematical phase space of the detected spectator to ensure this. Choosing the backward-moving spectator relative to the virtual photon exchanged ensures that we are largely free from effects like Final State Interactions (FSI) and target fragmentation. If, after scattering, the spectator re-scatters on DIS remnants of neutrons, this is called FSI. In this case, the measured structure function of neutrons would not be of free neutrons. The effects of FSI has been studied in [19]. It shows that if we choose our spectator kinematics as $p_s < 0.15$ GeV/$c$ and $\theta_{pq} > 100°$, then the effect of FSI is highly minimized in these regions as shown in Fig. 10. Similarly, the hadronic debris of struck neutrons could also produce low-momentum spectator protons. However, several theoretical calculations have shown that these fragmentation effects are highly suppressed in the backward hemisphere of scattering. Also, as the spectator momenta decreases, the contribution of protons coming through target fragmentation decreases, see Fig. 11.

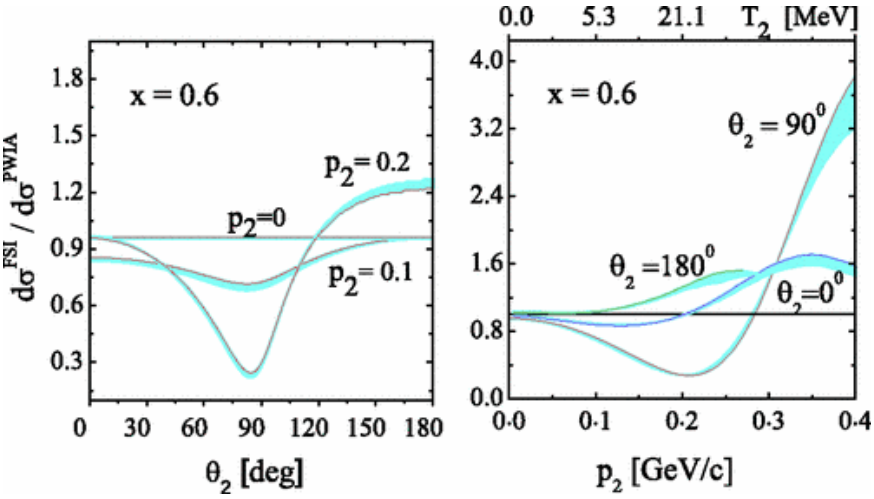



FIG. 10. The ratios of differential cross sections between FSI and PWIA, plotted as a function spectator angle $\theta_{pq}$ (left) and spectator momentum ($p_s$) (right), we see that as $p_s$ $< 0.15$ GeV/$c$ and $\theta_{pq} > 100°$ the ratio approaches unity [19].

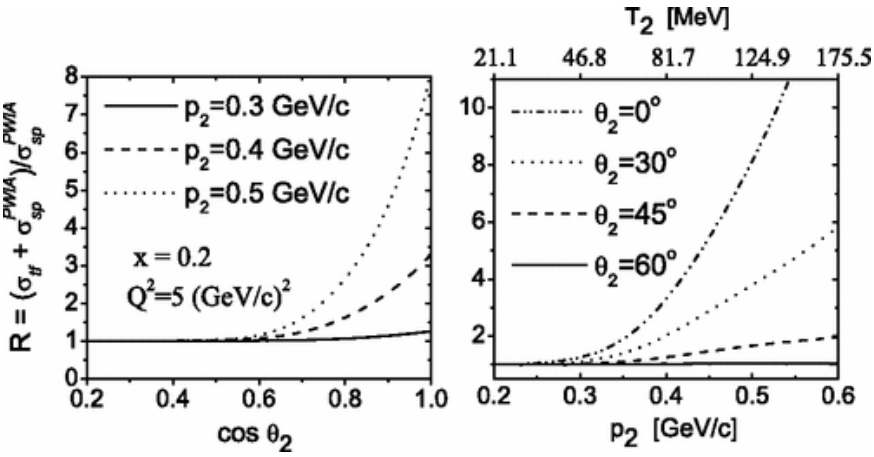



FIG. 11. Contribution of target fragmentation to nucleon emission in the process $D(e, e'p_s)X$. The ratio of the sum of cross sections $R = \frac{d\sigma_{tf}+d\sigma_{PWIA}}{d\sigma_{PWIA}}$, plotted versus $\cos\theta_2$ (left panel) and versus $|\vec{p}_2| \equiv p^2$ (right panel) [19].

Finally, off-shell effects are minimized for low-momentum spectators. The Off-Shell-ness can be calculated in our kinematics as,

$$\begin{aligned} m_n - m^* &= m_n - \sqrt{E^{*2} - \vec{p}_s^{\,2}} \\ &= m_n - \sqrt{(M_d - E_s)^2 - \vec{p}_s^{\,2}} \\ &= m_n - \sqrt{(M_d - \sqrt{m_p^2 + \vec{p}_s^{\,2}})^2 - \vec{p}_s^{\,2}} \\ &\approx M - \sqrt{(2M - M - \frac{\vec{p}_s^{\,2}}{2M})^2 - \vec{p}_s^{\,2}} \\ &\cong M - \sqrt{M^2 - 2\vec{p}_s^{\,2}}, \end{aligned} \tag{82}$$

**Where:**

- $M$ = average mass of nucleon,
- $m_n$ = neutron mass,
- $m_p$ = proton mass,
- $M_d$ = deuteron mass,
- $\vec{p}_s$ = momentum of spectator proton,
- $E_s = \sqrt{m_p^2 + \vec{p}_s^{\,2}}$ = energy of spectator.

**In BONuS12 kinematics**

$$\text{At } p_s = 70 \text{ MeV}/c \Rightarrow \text{Off-shellness} \lesssim 6 \text{ MeV}, \tag{83}$$

$$\text{At } p_s = 100 \text{ MeV}/c \Rightarrow \text{Off-shellness} \lesssim 11 \text{ MeV}. \tag{84}$$

Hence, in our kinematic measurement region, Off-Shell-ness is $\leq 2\%$ of neutron mass.

The calculation of the kinematic variables for the experiment is carried out in a covariant way, taking into account the four-momentum vector of the on-shell spectator proton for the calculation of the electron-neutron scattering. The Bjorken scaling variable for the interacting neutron then reads as,

$$x^\star = \frac{Q^2}{2\, p^\mu\, q^\mu} = \frac{Q^2}{2\left((M_d - E_s)\,\nu + \vec{q}\cdot\vec{p}_s\right)}, \tag{85}$$

where $q^\mu = (\nu; \vec{q})$, is the momentum transfer 4-vector carried by the virtual photon, and $p^\mu = (M_d - E_s, -\vec{p}_s)$ is the momentum 4-vector of the off-shell neutron, and $M_d$ is the mass of the deuterium nucleus. In the covariant description, the struck nucleon is on its energy shell but off its mass shell. The mass of the free nucleon $M$ is therefore replaced by the off-shell mass (or virtuality) of the bound nucleon:

$$M^{\star 2} = p_\mu p^\mu = (M_d - E_s)^2 - \vec{p}_s^{\,2}. \tag{86}$$

The invariant mass squared of the final hadronic state in $d(e, e'p_s)X$ can then also be written as,

$$\begin{aligned} W^{*2} &= (p^\mu + q^\mu)^2 \\ &= (M_d - E_s + \nu)^2 - (\vec{q} - \vec{p_s})^2. \end{aligned} \tag{87}$$

# CHAPTER 3

# THE BONUS12 EXPERIMENT

The Barely Off Shell Nucleon Structure experiment, known as the BONuS12 experiment, was conducted at Jefferson Lab in the Spring and Summer of 2020. This is an upgraded version of the BONuS experiment conducted at Jefferson Lab in 2006 [20], [21]. During the 2006 run, the experiment took data at a beam energy of 6 GeV, whereas after the energy upgrade of JLab to 12 GeV, it was conducted again with upgraded beam energy. The main principle behind this experiment is that we scatter the electron beam of the deuterium target and measure the scattered electron in the Forward detector. In contrast, the backward-moving recoil (spectators) is detected in the newly built Radial Time Projection Chamber (RTPC). Carefully choosing the kinematics of spectator particles to be in the low momentum range and scattering backward, we can extract information on DIS on the neutron in the reaction $D(e, e', p_s)X$, which helps us to calculate $F_2^n/F_2^p$.

## 3.1 CONTINUOUS ELECTRON BEAM ACCELERATOR FACILITY

The Thomas Jefferson National Accelerator Facility, abbreviated as Jefferson Lab, was built in 1984 in Newport News, Virginia. When it was first established, it was called Continuous Electron Beam Accelerator Facility (CEBAF), and in 1996, the name was changed to the new one. The main accelerator of this lab is the CEBAF accelerator [22]. It consists of a polarized electron source and injector. Further, it consists of superconducting Radio-Frequency accelerators. Initially, the accelerators could supply up to 4 GeV of electron beam, which increased to 6 GeV after design enhancement in the 2000s. After further energy upgrades during the 2010s, it could deliver up to 12 GeV electron beam energy. Jefferson Lab consists of four experimental halls, generally known as Halls A to D. Each hall has a separate physics program. Halls A, B, and C work with electron beams, whereas Hall D works with photon beams. The BONuS12 experiment took data in Hall B at a beam energy of 10.4 GeV. Below, we will discuss the detector system in Hall B and the RTPC detector installed in Hall B, especially for this experiment.

### 3.2 CLAS12 DETECTOR

The CEBAF Large Acceptance Spectrometer, generally known as the CLAS12 detector [23], is a combined name of a collection of individual detectors that are in Hall B. Initially, the detector was called CLAS detector, but after the energy upgrade to 12 GeV some detectors were upgraded, and some new detectors were installed. Hence, the name of this detector collection was changed to CLAS12. When the accelerated electrons reach Hall B, they are scattered inside the target, which is different depending on the experimental needs of that physics program. For the BONuS12 experiment, the target was deuterium gas. After scattering, multiple particles are produced in the final state. Many detectors in Hall B help track these charged/neutral particles over a wide range of solid angles. Each detector has its purpose. We can divide these detector groups into two divisions. The Forward Detectors (FD) detect and help to identify particles that are scattered in a forward direction at an angle of $5 - 35°$ in the lab frame, whereas the Central Detectors (CD) detect those in the central region in the angular range of $35 - 125°$. The Central Vertex Tracker (CVT) of the Central Detector was replaced by the RTPC detector. We will be discussing each detector in the CLAS12 detector system used for the BONuS12 experiment; see Fig. 12.

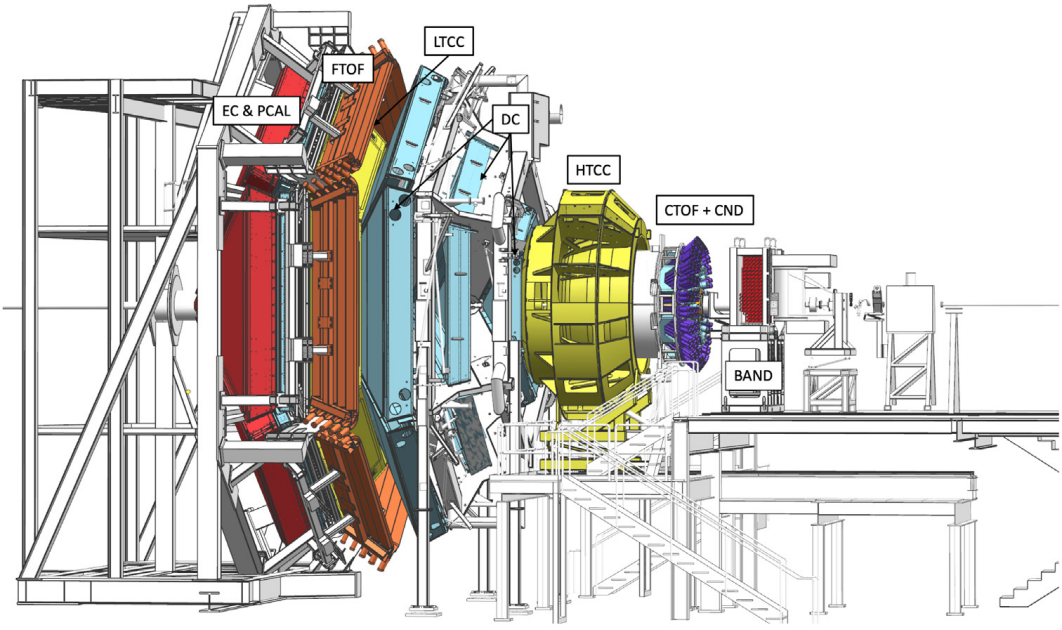



FIG. 12. CLAS12 Detector in Hall-B beamline [24].

### 3.2.1 TORUS AND SOLENOID MAGNETS

The CLAS12 detector system consists of a torus magnet that covers a polar angle up to 35° with a solenoid magnet system that covers a polar range from 35° to 125° and a nominal magnetic field of maximum 5 T. These magnet systems help track charged particles at high momentum and resolution at forward angles with the high luminosity that Hall B offers. This magnet system also helps effectively shield the Moller electrons produced by the beam electrons' scattering off atomic electrons of the target.

CLAS12 contains a torus magnet in the form of six superconducting coils arranged in a symmetrical configuration around the beamline and provides angular coverage of 5° to 35°. They operate at a nominal current of 3770 A and a peak field of 3.58 T at the inner turn of the torus. The inductance of this torus coil is 2.0 H, and the energy stored in the coil is 14.2 MJ.

The solenoid magnet of the CLAS12 detector is a superconducting magnet that provides a magnetic field aligned with the beamline. This magnet provides:

- Magnetic field for tracking charged particles in central detectors up to 125°.
- Shields Moller electrons (beam electrons scattered from atomic electrons from targets).
- Provides a uniform magnetic field for operating polarized targets in Hall B.

The superconducting coils could be powered at a maximum current of 2416 A, generating a 5 T field at the center and storing 20 MJ of energy [25].

### 3.2.2 DRIFT CHAMBERS (DC)

The CLAS12 detector system consists of drift chambers (DC), which are physically supported by the six superconducting coils of the torus magnet system. The three independent drift chambers are arranged in three regions (R1, R2, and R3). Each DC region consists of six sectors, each of which has 12 layers with 112 sense wires, which accounts for 36 layers per region. These three regions of DCs are arranged in such a way (see Fig. 13) that R1 DCs are located upstream (entrance of torus magnet region), R2 DCs are located inside the magnet where there is the maximum magnetic field, and R3 DCs are towards the downstream side of torus magnet where there is a minimum magnetic field. This arrangement of three DCs in a torus magnet system provides independent tracking of charged particles in torus sectors [26].

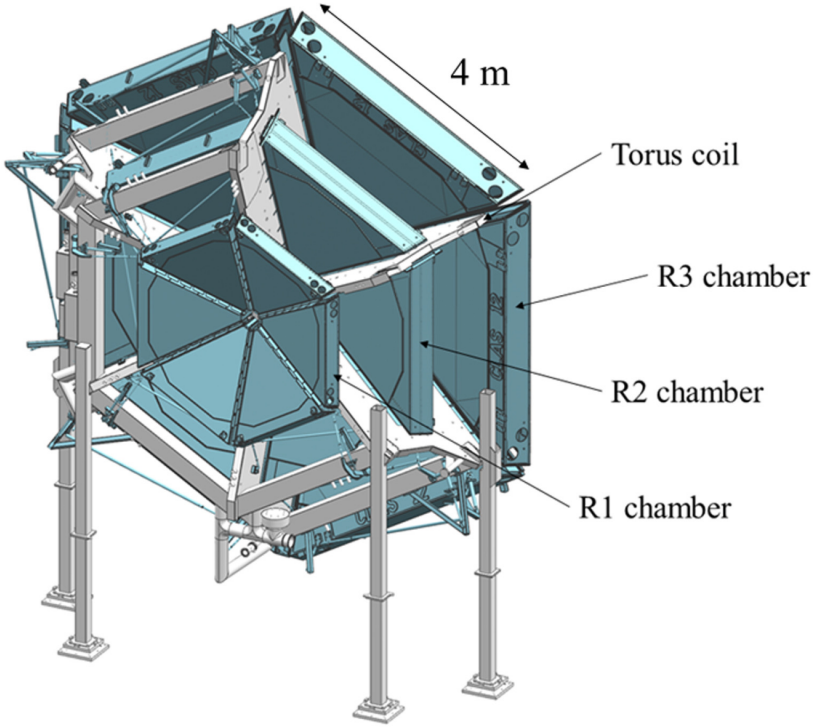



FIG. 13. A model drawing of torus magnet (light gray) with the Drift Chambers (light blue). Figure adapted from [26].

### 3.2.3 CHERENKOV COUNTERS (CC)

Whenever a charged particle travels through a medium with a speed greater than the speed of light in that medium, a type of radiation is emitted called Cherenkov radiation. This radiation is emitted at a particular angle relative to the direction of the charged particle, and hence, the wavefront of Cherenkov radiation forms a cone. The angle of the cone is given by, $\cos\theta_c = 1/\beta n$ . The Cherenkov light is only produced by particles above a certain speed threshold, $\beta > 1/n$. One of the benefits of Cherenkov counters is that they can differentiate particles with the same momenta but different masses. The CLAS12 detector system contains three different Cherenkov detectors: low-threshold Cherenkov Counters (LTCC) and high-threshold Cherenkov Counters (HTCC), as well as the Ring Imaging Cherenkov Detector (RICH). The HTCC is the main detector in the CLAS12 system that separates

electrons/positrons below momenta of 4.9 GeV/$c$ from pions, kaons, and protons. This detector is placed between the target and the R1 region of the DC. It is filled with dry $CO_2$ gas at 1 atmosphere, with a refractive index of 1.00041. The LTCC is used to detect $\pi^+/\pi^-$ at momenta greater than 3.5 GeV/$c$. It comprises a $C_4F_{10}$ radiator with a refractive index of 1.00134. Previously, the Cherenkov counters were used to detect low-energy electrons during the CLAS era. For CLAS12, the LTCC has been refurbished to increase $\pi^+/\pi^-$ detection efficiency. It can also be used to discriminate between pions/kaons. One of the sectors combined a RICH module instead of LTCC to improve pion/kaon separation [27].

### 3.2.4 TIME OF FLIGHT COUNTERS (TOF)

It is known that two particles having the same momentum but different masses will have different velocities. Hence, they will take different amounts of time to travel the same distance. Considering this fact, time of flight (TOF) detectors were integrated into the CLAS12 system. The CLAS12 detector system contains two different TOF detectors. One is the Forward Time of Flight detector (FTOF), and the next is the Central Time of Flight detector (CTOF). The FTOF consists of six sectors of plastic scintillators with readout on both sides. Each sector has three arrays of counters - panel 1a, panel 1b, and panel 1c with corresponding timing resolutions of 125 ps, 85 ps, and 155 ps, respectively. It measures the time of flight of charged particles that come out of the target. The FTOF covers the polar angle from 5° to 45°. The CTOF detector is used to detect charged particles emerging from the target in the momentum range of 0.3 to 1.25 GeV/$c$. It comprises 48 plastic scintillators with a photomultiplier tube (PMT) readout on both sides via focusing light guides. The timing resolution of the CTOF is 80 ps [28] and [29].

### 3.2.5 ELECTROMAGNETIC CALORIEMETERS (ECAL)

The CLAS12 detector consists of an electromagnetic calorimeter (EC) that stemmed from the CLAS era and a pre-shower calorimeter (PCAL) installed in front of the EC, called ECAL in combination. These calorimeters are used to detect charged and neutral particles. These are sampling calorimeters, which have six modules. The EC is divided into EC-inner (ECin) and EC-outer (ECout). They provide a longitudinal sampling of electromagnetic showers and also of hadronic interactions. Each module of these ECALs has a triangular shape with 15 layers for PCAL and ECin and 24 layers for ECout. Each layer consists of 1 cm thick scintillators segmented into wide strips of 4.5 cm for PCAL and 10 cm for EC sandwiched between 2.2 mm thick lead sheets. The total thickness of these strips corresponds to 20.5

radiation lengths [30].

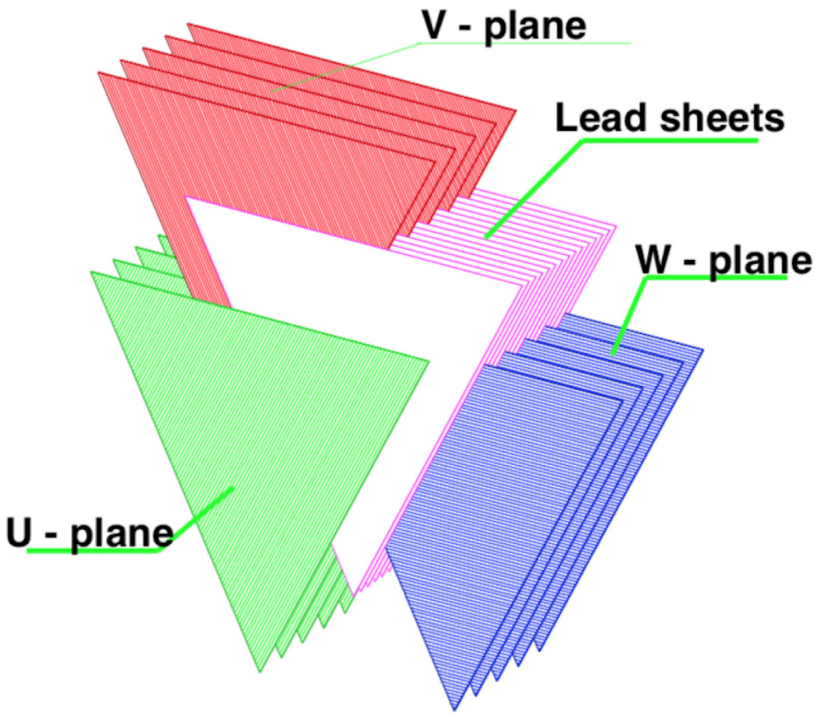



FIG. 14. A schematic showing how scintillator layers are interleaved with lead sheets. The PCAL consists of five layers of scintillator strips that define the U, V, and W planes. Figure adapted from [30].

### 3.2.6 CENTRAL NEUTRON DETECTOR (CND)

The CND is placed radially outside of the CTOF before the solenoid magnet. It comprises three layers of scintillator paddles, of which there are 48 paddles in each layer. These paddles are coupled by a semicircular light guide at the downstream end, whereas at the upstream readout end, PMTs are placed outside the high magnetic field region of the solenoid magnet. It can detect neutrons in the momentum range of 0.2 to 1 GeV/$c$ based on their time of flight in the TOF detector and energy deposited in scintillators. The CND has an angular coverage of 40° - 125° in the polar region and has a neutron detection efficiency of around

10% [31].

The performance of CLAS12 detector can be summarised as shown in Table 1,

| **Capability** | **Quantity** | **Status** |
|---|---|---|
| Coverage & Efficiency | Tracks (FD) | $5^\circ < \theta < 35^\circ$ |
| | Tracks (CD) | $35^\circ < \theta < 125^\circ$ |
| | Momentum (FD & CD) | $p > 0.2$ GeV |
| | Photon angle (FD) | $5^\circ < \theta < 35^\circ$ |
| | Photon angle (FT) | $25^\circ < \theta < 4.5^\circ$ |
| | Electron detection (HTCC) | $5^\circ < \theta < 35^\circ$, $0^\circ < \theta < 360^\circ$ |
| | Efficiency | $n > 99\%$ |
| | Neutron detection (FD) | $5^\circ < \theta < 35^\circ$ |
| | Efficiency | $\leq 75\%$ |
| | Neutron detection (CD) | $35^\circ < \theta < 125^\circ$ |
| | Efficiency | 10% |
| | Neutron Detection (BAND) | $155^\circ < \theta < 175^\circ$ |
| | Efficiency | 28% |
| Resolution | Momentum (FD) | $\sigma_f/\rho = 0.5 - 1.5\%$ |
| | Momentum (CD) | $\sigma_g/\rho < 5\%$ |
| | Pol. angles (FD) | $\sigma_g = 1 - 2$ mrad |
| | Pol. angles (CD) | $\sigma_g = 3 - 5$ mrad |
| | Azim. angles (FD) | $\sigma_g < 2$ mrad/sin $\delta$ |
| | Azim. angles (CD) | $\sigma_g = 3 - 15$ mrad |
| | Timing (FD) | $\sigma_r = 60 - 110$ ps |
| | Timing (CD) | $\sigma_r = 80 - 100$ ps |
| | Energy ($\sigma_E/E$) (FD) | $0.1/\sqrt{E}$ (GeV) |
| | Energy ($\sigma_E/E$) (FT) | $0.03/\sqrt{E}$ (GeV) |
| Operation | Luminosity | $L = 10^{18}$ cm$^{-2}$s$^{-1}$ |
| DAO | Data Rate | 20 kHz, 800 MB$_e$, LT. 95% |
| Magnetic Field | Solenoid | $I_R = 5$ T |
| | Trans | $J_{Ref} = 0.5 - 2.7$ Tm at $5^\circ < \theta < 25^\circ$ |

TABLE 1. CLAS12 performance parameters based on the current state of the reconstruction, subsystem calibrations, knowledge of the detector misalignments, and the understanding of the torus and solenoid magnetic fields.

## 3.3 RADIAL TIME PROJECTION CHAMBER

David R. Nygren first introduced the concept of Time Projection Chambers (TPCs) at Lawrence Berkeley Laboratory in the late 1970s [32]. These detectors work on a principle of gas ionization, which contains a filled detection volume in an electric field with a mechanism to collect electrons sensitive to the ionization position. A 2D array of the arrival position of ionized electrons in the readout system, along with the drift time, allows the reconstruction of the initial state ionization point and, hence, the reconstruction of the entire track. Over the years, several improvements have been made in such detector systems, and such detectors have been used in many high-energy experiments worldwide.

**Gas Electron Multipliers (GEM) [33]:** A GEM foil consists of thin Kapton foil clad in copper on both sides. It contains a regular and uniform high density of doubly conical-shaped holes in a hexagonal pattern. Typically, the Kapton foil has a thickness of 50 $\mu$m, and cladded copper has a thickness of 5 $\mu$m. The holes have a diameter of 40-140 $\mu$m, and the number of holes is $50 - 100$ mm$^{-2}$. The GEM foils play a crucial role in the amplification of ionization signals. When a charged particle passes through an active volume of the ionization detector, electron-ion pairs are formed. These electrons drift toward GEM foils if an external electric field is applied across the drift volume. Due to the biconical structure of the holes, an intense electric field is formed in the holes. As drifting electrons move to the surface of GEMs, the strong electric field of holes accelerates the electrons. Inside the holes, the accelerated electrons collide with gas molecules and form additional electron-ion pairs, producing secondary electrons. This process is known as the avalanche effect, which produces 100s of secondary electrons per primary electron. This amplification causes a signal that is large enough to be detected by the readout system. Hence, each hole acts as a proportional amplifier. So, because multiple secondary electrons come from tiny holes in GEM foils, it allows high gain, low noise, and spatial resolution. Also, multiple GEM foils can be cascaded to multiply the amplification for detection and localization of charged particles to improve detection efficiency and resolution.

### 3.3.1 BONUS12 RTPC

The BONuS12 experiment used a 10.4 GeV electron beam scattered from the $D_2$ target. The scattered electrons were detected in CLAS12, whereas the recoil protons were detected in the RTPC detector, as explained in previous sections. To measure events in which the

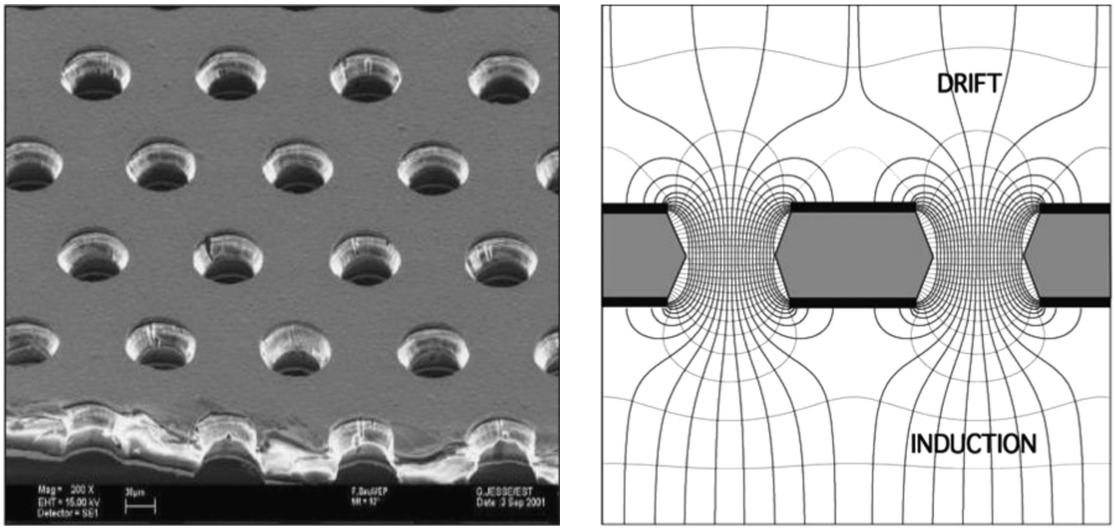



FIG. 15. Microscopic view of GEM foil(Left) and holes in GEM foils with electric field lines drawn (Right) [33].

electrons scatter off the neutrons, we need to detect protons inside the RTPC. In principle, if scattering occurred from neutrons, which are not bound to protons, then neutrons would be on-shell. In such a case, the proton would not leave the target, leading to no spectators in our detector; however, if scattering took place from a neutron bound with a proton, protons would-be spectators. The lower the momentum of the spectator proton, the more loose the bond between neutrons and protons in $D_2$ nuclei. So, we will be tagging protons with as low a momentum as possible. From the Bethe-Bloch energy loss formula, we know that a particle with smaller momentum traveling through a medium loses more energy in that medium. So, as we wanted to detect a low-energy recoil, it would lose much of its energy before reaching any detector in CLAS12. Hence, we needed to build a detector that is much closer to the target and contains very few materials so that the spectator proton would not lose much energy while traveling through the active detection region of the detector. To solve this issue, using the principle of GEM amplification, a Radial Time Projection Chamber (RTPC) was built for the original BONuS experiment at Jlab in 2006 [34]. The experiment

was conducted successfully, showing the use of GEM-based detectors for spectator tagging. Due to much improvement in GEM production quality, large-area GEMs became available, making it easy to construct cylindrical GEM layers without having to join multiple GEMs to form a cylinder, as in BONuS RTPC. Using these improvements, BONuS12 RTPC was built. It has three layers of GEM amplification. Hence, this BONuS12 RTPC has better $\theta$ and $\phi$ resolution, vertex reconstruction, and momentum reconstruction compared to its previous version and has a momentum threshold of 70 MeV/$c$.

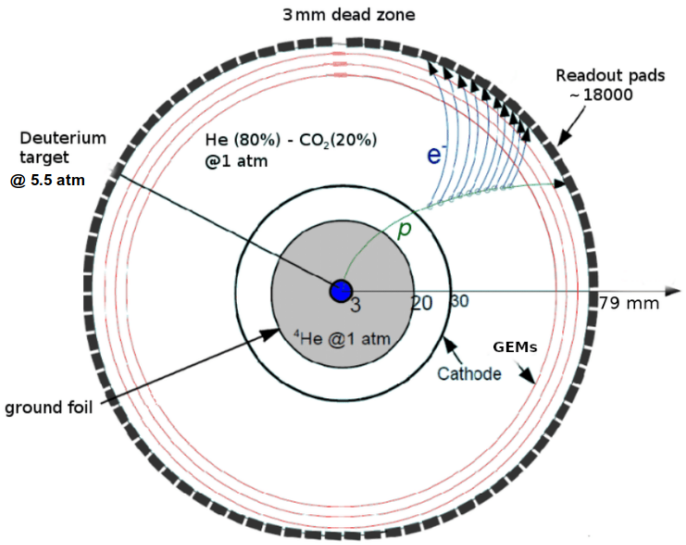



FIG. 16. Cross sectional view of RTPC seen from the beam axis.

### 3.3.2 DESIGN AND WORKING PRINCIPLE

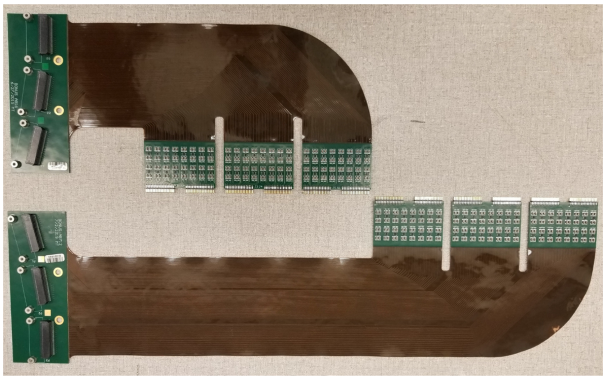
The BONuS12 RTPC is a cylindrical detector of an active length of 40 cm placed along the beam axis in the center of the CLAS12 solenoid. It has a 50 cm long target straw along the central axis, made up of aluminized polyimide (Kapton) of 63 $\mu$m thickness and 6 mm in diameter. During data taking, the target straw was maintained at a pressure of 5.5 atmospheres of the target gas. The outer region of the target is the buffer region, which contains $^4$He gas to minimize the effect of Moller electrons. Outside the buffer region is ground foil, made from 6 $\mu$m thick aluminized mylar. It prevents the target from the field

generated by the cathode. Then, the cathode is held at negative potential at a 30 mm radial distance from the central axis. The electric field is maintained radially outward. Radially outside the cathode is the drift region, which contains a mixture of $^4$He and $CO_2$ in the ratio of 80:20 by volume, where $CO_2$ is used as a quencher. This region extends up to 70 mm, where we have our first GEM layer. Further, at 73 mm and 76 mm, we have two more GEM layers, making three GEM layers in the RTPC detector altogether. Then, the outermost layer is the readout pads, which collect charges from GEM amplification and transfer information into front-end electronics through readout channels.

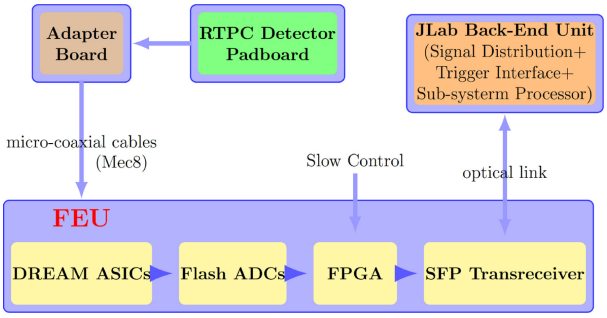
The RTPC works on the principle of gaseous ionization. Whenever a spectator particle leaves the target and enters the drift region, it ionizes the drift gas along its path. Since the cathode is held at a negative potential, only ionization electrons travel toward the GEMs placed radially outward. There, the ionization electrons are amplified through three GEM layers, producing a total amplification of around 1000 times. There is an electric field across each successive GEM for forward accelerating of charge. On the outermost surface, the amplified charges are collected at the padboard. The readout padboard consists of 17,280 conductive pads for charge collection. Between each pad, there is isolation. Each pad covers 4 mm in the z-direction and 2° in $\phi$.

The signals produced by amplified charges that are deposited on the readout pads are transferred to the DAQ system through adapter boards, as shown in Fig. 17a. On one side, the adapter boards are connected to the padboard, whereas on the other side, a Hitachi micro-coaxial cable is connected to it. The adapter board contains current limiting resistors that protect the electronics in case of a spark in the detector. The Hitachi cables transfer signal from readout pads to front-end units (FEUs). Each FEU is composed of eight MEC8 input connectors, eight DREAM (Dead-timeless Readout Electronics ASIC for Micromegas) chips, and an 8-channel flash ADC. Each DREAM chip contains 64 channels, with each channel equipped with an amplifier, shaper, analog buffer, and discriminator. The DREAM chip performs in a dead-timeless readout mode of up to 20 MHz for a trigger rate of up to 20 kHz. The FEUs used in the BONuS12 experiment used the same electronics as the Micromegas Vertex Tracker, which was removed and replaced with BONuS12 RTPC. After processing, the signals from FEUs are sent to the back-end unit (BEUs) of the CLAS12 DAQ system for signal distribution, trigger interface, and further processing.

**Official Start of Experiment** The BONuS12 experiment officially started on February 12, 2020. After the start of the experiment, the RTPC started showing some issues. The gain was unstable, and leakage current started to appear. By the middle of March, it was



(a) Adapter boards used to transfer signals from the RTPC padboard to the front-end electronics units (FEUs).




(b) Schematic of the RTPC data acquisition electronics, illustrating the signal flow from the detector to the back-end storage system.

FIG. 17. Overview of the RTPC signal transmission and data acquisition electronics.

inoperable. So, we replaced this RTPC (called RTPC-1) with a new RTPC (called RTPC-3). After installing RTPC-3 in Hall B, the experiment started again on March 20, 2020. However, due to the COVID outbreak, Jefferson Lab halted all experiments and closed the lab. After a couple of months, when the lab reopened, the experiment resumed data-taking on August 2, 2020. This part of the run is called "Summer Run," and the data-taking run with RTPC-1 is called "Spring Run." The experiment was officially completed on September 21, 2020. Since the spring run had problems with the stability of the RTPC detector, those data have not yet been analyzed. So, the analysis presented in this dissertation will only be based on experimental data from the Summer of 2020.

TABLE 2.

| Beam Energy (GeV) | Spring 2020 (02/12 - 03/24) | Summer 2020 (08/02 - 09/21) | Target | Total Triggers M |
|---|---|---|---|---|
| | 1 | 22 | Empty | 23 |
| 2.1 | 81 | 185 | $H_2$ | 266 |
| | 37 | 145 | $D_2$ | 82 |
| | 19 | 44 | $^4He$ | 63 |
| Total | 138 | 296 | | 434 |
| | 21 | 45 | Empty | 66 |
| 10.4 | 151 | 266 | $H_2$ | 417 |
| | 2275 | 2355 | $D_2$ | 4630 |
| | 77 | 51 | $^4He$ | 128 |
| Total | 2524 | 2717 | | 5241 |

TABLE 3. Total statistics collected during RGF.

# CHAPTER 4

# ADVANCED DATA ANALYSIS

## 4.1 CLAS12 DAQ, RECONSTRUCTION AND EVENT BUILDER

The CLAS12 Data Acquisition (DAQ) system, a pivotal component in the data collection and analysis, is entirely network-based. It comprises various hardware and software components that meet specific compatibility standards for CLAS12 operations. These components are interconnected through parallel optical fibers, which distribute synchronization resets and trigger signals to each component while collecting busy signals from the electronics. VME/VXS crate data are transferred to the Event Builder (EB) using 1 Gb or 10 Gb Ethernet links over TCP connections. After the event builder processes the events, they are written onto a disk for offline analysis by the users.

The CLAS12 detector system is huge and complex. During each experiment, large data volumes are generated. Due to this, the CLAS12 software team has developed a common analysis framework called CLARA (CLAS12 Reconstruction and Analysis framework). CLARA, which incorporates essential software libraries, reconstruction tools, and applications for analyzing physics data, is based on data-stream processing. It plays a crucial role in managing the complexity and volume of data, enabling service-oriented applications composed of micro-services connected by data-stream pipes. Each micro-service receives an input data stream, processes it, and sends the output in a CLAS12 bank structure to the next service in the workflow. CLARA supports Java, C++, and Python, although all reconstruction services within CLARA currently use Java. Further details on CLARA and its tools are available in Reference [35]. Within CLARA, data collected by the DAQ (containing detector identifiers and digitized ADC and TDC data) are initially decoded and organized into a bank structure for each event. These banks contain all necessary components (e.g., hits, time, clusters, and tracks) for event reconstruction in both the Central Detector (CD) and Forward Detector (FD) systems. Track reconstruction involves algorithms for both pattern recognition and track fitting. Charged particle tracks are reconstructed in both tracking systems through hit-based tracking, matching recorded hit positions across various detectors. Simultaneously, track reconstruction using hit timing and energy is performed in other

detector systems, with matching conducted by the Event Builder software. Unmatched hits and tracks, such as photons and neutrons, are reserved as neutral candidates. After matching, start time reconstruction occurs, enabling time-based tracking and factoring in the drift time in the Drift Chambers (DCs). The improved time-based tracks serve as inputs to the Event Builder, leading to the final event reconstruction. The CLAS12 DAQ stores raw data in EVIO (Event Input-Output) format, developed by the Jefferson Lab Data Acquisition Group. During decoding, EVIO is converted to the CLAS12 data format known as HIPO (High-Performance Output). This data structure minimizes disk space requirements and optimizes access during reconstruction and further analysis. HIPO supports both Java and C++ libraries.

The CLAS12 Event Builder (EB) [35], [36] is a crucial part of the reconstruction package that gathers and organizes information for each physics event reconstructed by CLAS12 detectors. It aggregates global event information through data banks from upstream services, combining all details for a single event. The EB also implements particle identification schemes and generates final event data in a standard bank structure. Various CLAS12 subsystems support particle identification, helping to differentiate particles in the final event-building process. See Fig. 18. The CLAS12 EB assigns Particle Identification (PID) numbers,

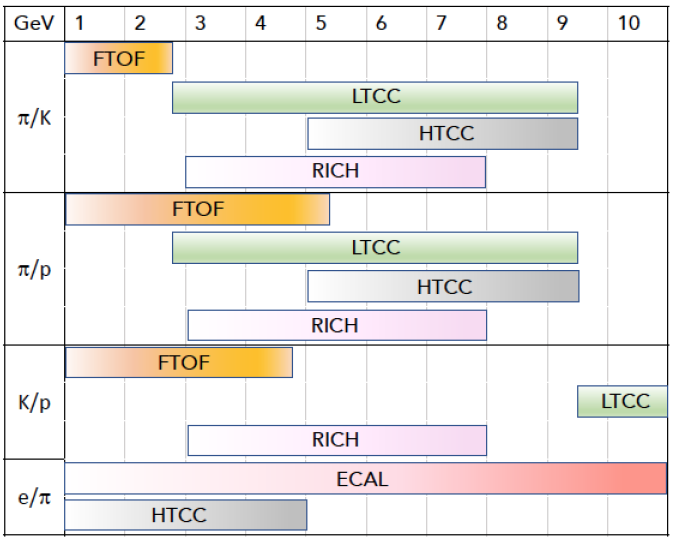



FIG. 18. A summary of the various detector subsystems in the CLAS12 Forward Detector (FD) used for particle identification as a function of momentum, where higher color intensity represents greater sensitivity to detection of that particle.

such as 11 for electrons and -11 for positrons, by matching charged particle tracks in the DC with clusters in the ECAL and associated hits in the FTOF, meeting specific HTCC and EC criteria. In summary, at the end of the CLAS12 offline software, a set of routines collectively called Event Builder does the following:

- Collect and analyze global event information, e.g., RF, helicity, scalers.
- Collect and organize reconstructed responses from the various CLAS12 detector subsystems.
- Associate those responses into particles and execute particle identification schemes.
- Output all necessary information for physics analyses into dedicated data structures known as data summary tapes (DST) in the form of data structures called "banks."

For each charged particle in the forward detectors, the Event Builder assigns $e^-$ (PID = 11) or $e^+$ (PID = -11) (depending on the bending direction of the reconstructed track in the DC) if a particle satisfies all following HTCC and ECAL requirements:

- 2.0 photoelectrons in HTCC.
- 60 MeV energy deposit in PCAL.
- 5-sigma cut on a parameterized momentum-dependent sampling fraction where "sampling fraction" is ECAL visible energy deposition (PCAL+Inner+Outer) divided by DC momentum.

The slow-recoiling final state particles are reconstructed and detected in the BONuS12 RTPC (momentum per charge less than 300 MeV/$c$), but they are not yet included in the Event Builder. However, we apply custom criteria to identify recoils.

For neutrals in the forward CLAS12 detector, only photons (PID = 22) and neutrons (PID = 2112) are considered, and particle identification is assigned based on a simple beta cut at 0.9. Currently, only one timing response is used for this, and for ECAL, the prioritization is PCAL, EC Inner, and EC Outer. The momentum direction is assigned based on the neutral's ECAL cluster position and the vertex of the charged particle used to determine the start time. For photons, the energy (magnitude of the momentum) is calculated from ECAL visible energy and momentum-dependent sampling fraction, while for neutrons, energy is calculated from beta. The CND is treated similarly to ECAL for the central detector, except only neutron PID is assigned based on beta, and the cut is at 0.8. This experiment's two

reactions of interest and selection are the inclusive scattering $D(e,e')X$ and neutron-tagged DIS scattering $D(e,e'p_s)X$. Therefore, our final state particles that need to be identified are the electrons in the forward CLAS12 detector and the slow-recoiling protons in the BONuS12 RTPC. The following subsections detail our criteria and selection cuts for these two final state particles.

## 4.2 ELECTRON IDENTIFICATION REFINEMENTS

### 4.2.1 VERTEX CUTS

The BONuS12 target has an active length of 40 cm, centered at the center of the CLAS12 solenoid. The target straw contains the target gas used in the experiment. Whenever an electron scatters off the target gas, the position along the target it scatters from is also reconstructed and saved in the database. Fig. 19 shows the reconstructed $z$-vertex for the collected electrons. One can see two sharp peaks corresponding to the two 27$\mu$m thick aluminum windows at the ends of the target. These particles must be rejected as they originate from outside the target's active region. As explained above, we have two reactions of interest: inclusive scattering and neutron-tagged DIS scattering. Hence, for the Summer 2020 data, we set two different $v_z$ cuts for electrons as: -20.35 cm $< v_z <$ 18.65 cm for inclusive analysis whereas, for the tagged analysis, it was -23.35 cm $< v_z <$ 20.05 cm. For the tagged case, we require an additional proton detected in RTPC in coincidence with electron in CLAS12. The optimized vertex cut for proton selection was -21 cm $< vz_p <$ 18 cm. It is described in detail in Section 5.2. Similarly, the $\Delta v_z(vz_{\rm el} - vz_p)$ cut for our final analysis has a width of approximately $\sim$13 mm (see Section 5.2.1) around the mean. For the tagged events selection we want acceptance of proton to guide the selection cuts rather that the electron's acceptance. So, because of these reasons we decided the electron selection cut for the tagged analysis to be wider than the inclusive analysis.

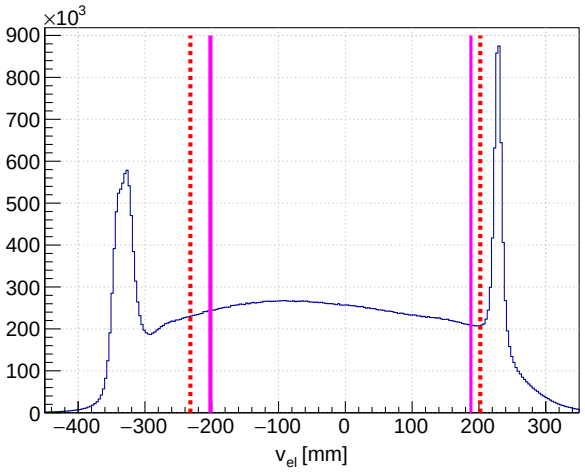



FIG. 19. $v_z$ distribution for the scattered electrons from the $D_2$ target using 10 GeV beam energy. The red vertical lines at $v_z$ = -23.35 cm and at 20.05 cm represent the cuts we apply to select the electrons that have a tagged event in RTPC, whereas the magenta vertical lines at $v_z$ = -20.35 cm and at 18.65 cm represent the cut to select the inclusive electrons irrespective of any tracks in RTPC.

Additionally, to clean the electron sample from the electrons that go through substantial additional material in the forward direction, we apply the following cuts on the correlation between the vertex ($v_z$) and the polar angle ($\theta_{el}$), see Fig. 20.

- (345.3 - ($v_z$ - $z_{center}$ ))tan($\theta_{el}$) > 25.0 [mm] and $\theta_{el} > 6.5°$: to remove the electrons that hit the FMT inner ring.
- (345.3 - ($v_z$ - $z_{center}$ ))tan($\theta_{el}$) < 248.5 [mm] and $\theta_{el} < 39$ °: to remove the electrons that hit the downstream side of the solenoid magnet.

Where $v_z$ is the electron's vertex along the RTPC longitudinal position, $z_{center}$ is the position of the center of the target with respect to the center of the CLAS12 coordinate system (= -20.5 mm).

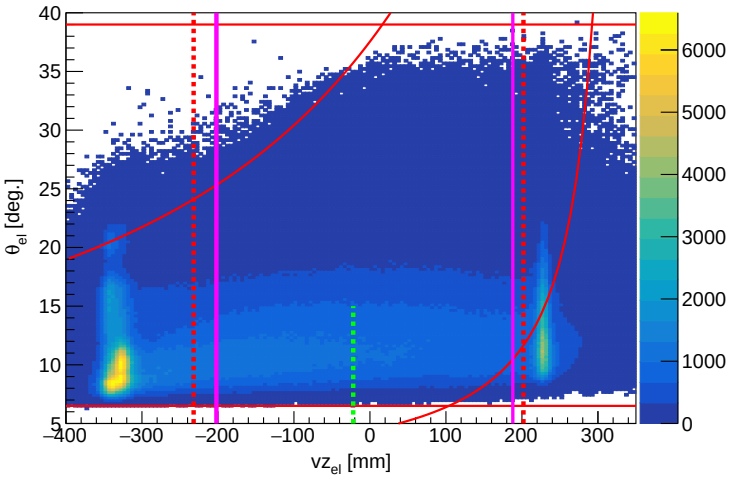



FIG. 20. 2D distribution of $v_z$ vs. $\theta$ for electrons. The red horizontal lines and curves represent the cuts on the correlation between $v_z$ and $\theta$ to clean the electrons sample from the ones that experience larger radiations. The green dashed lines represent the line showing the geometrical center of the target, and the vertical dotted red line and magenta lines represent the electron selection cuts for inclusive and tagged events as explained above.

### 4.2.2 CUT ON ENERGY OF SCATTERED ELECTRON

The low $E'$ region corresponds to the trigger threshold region. Also, for low $E'$ region, energy loss due to beam Bremsthrlaung is high ($dE \propto E'$). In this region, we have high pion contamination in the electron sample and pair symmetric contamination. So, to avoid these effects, we put a cut on minimum $E'$ to be greater than 2 GeV, see Fig. 21.

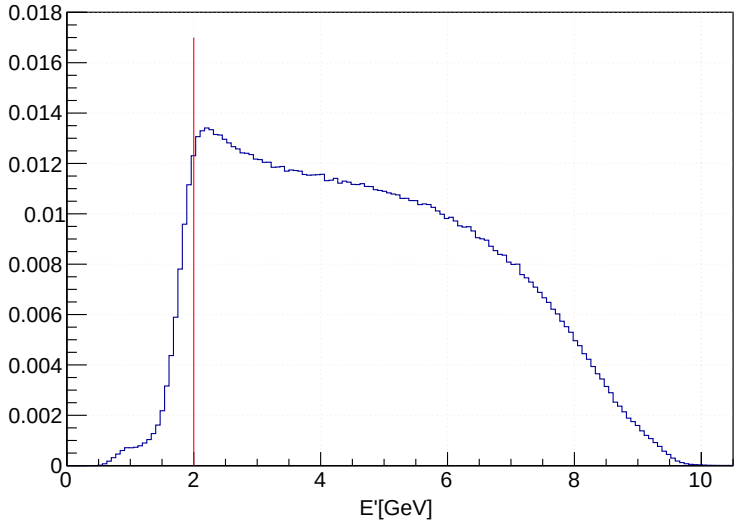



FIG. 21. The distributions of energy of scattered electrons. The red vertical line represents the cut at 2 GeV.

### 4.2.3 PCAL AND EC ENERGY CUTS

The CLAS12 calorimeters are sampling calorimeters that consist of high-density material (lead) sandwiched between scintillator bars. When electrons and photons interact with the calorimeter, they produce electromagnetic showers (production of secondary particles), primarily through Bremsstrahlung (for electrons) and Pair production (for photons). Most of the particles contaminating the sample in our raw electron sample are pions. The electrons and photons deposit energies in calorimeters, which is proportional to their momentum. The pions deposit almost a small constant energy in the calorimeters proportional to the thickness of the calorimeters, primarily through ionization. Hence, they are also called Minimum Ionizing particles (MIPs). Fig. 22 illustrates the importance of the different sub-detectors in the forward CLAS12 detector. It shows the deposited energies in the calorimeters for the negative reconstructed tracks at two stages in the identification: just being a negative particle and then associating a negative particle with the number of photoelectrons (*nphe*) in the HTCC and the sampling fraction. One can see the low energy band around 30 MeV deposited energy in PCAL that represents the MIPs. The CLAS12 Event Builder assigns a 60 MeV minimum visible energy deposit cut in the PCAL to separate electrons from MIPs.

Upon a closer investigation to remove any MIP tail in the PCAL energy spectrum and to obtain a clean electron sample, we increased the value of the cut to 100 MeV. Additionally, we noticed that some electrons do not make it to the EC. Therefore, to ensure that our electrons have energy deposited in at least two regions of the calorimeter system, we added a cut of a minimum $EC_{in}$ deposited energy of 10 MeV.

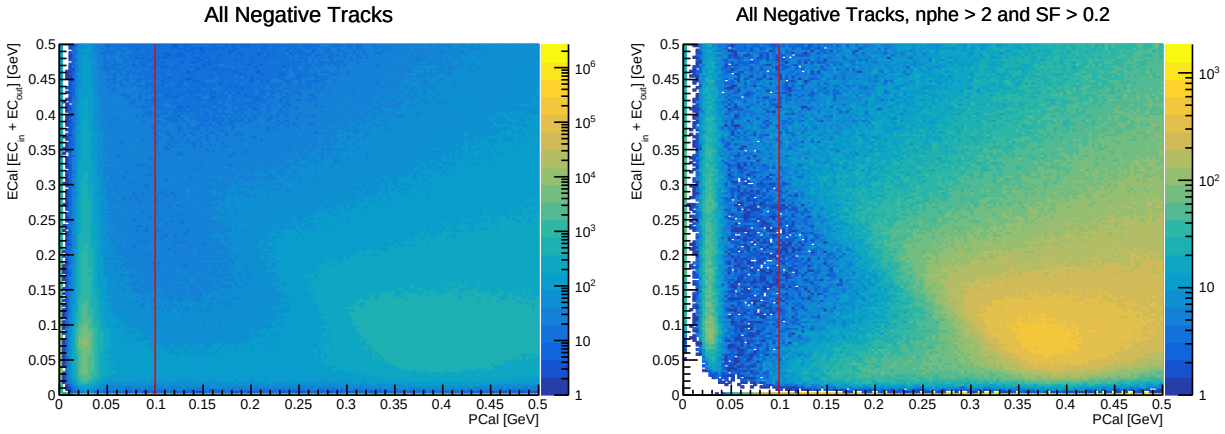



FIG. 22. ECAL ($EC_{in} + EC_{out}$) versus PCAL deposited energy distributions for the negative reconstructed particles in the CLAS12 forward detector (left), and for the negative particles after adding the requirements that the *nphe* in HTCC is greater than 2 and the sampling fraction is greater than 0.2 (right).

### 4.2.4 DC FIDUCIAL CUT

We developed the DC fiducial cut following the established procedures from CLAS12 Run Group A (a previous experiment at Hall-B, whose analysis procedures have already been validated and approved by the CLAS collaboration), where the cuts are defined based on the distribution of track quality ($\chi^2/NDF$) of the collected electrons. In the following procedures, we use the electrons originating from the active region of the target and have momenta of greater than 2 GeV/$c$. Using the hit position in the local coordinate system, we calculate the local angles $\theta$ and $\phi$ in region 1 of DC for the collected electrons. Then, we perform the following steps:

- For each DC sector, we segment the $\theta_{local}$ range, from 5.2° to 30.4°, into 15 slices.
- For each slice in $\theta_{local}$, we monitor the $\phi$ distribution of the average $\chi^2/NDF$.
- Then, the $\phi$ distributions are fitted with a constant function around the central value (0° for sector 1, 60° for sector 2, and so on for the remaining sectors).
- Then, the $\phi$ boundaries are determined where the value of average $\chi^2/NDF$ increases to 40% compared to the fitted central value in each slice. (see Figs. 23, 24, 25, 26, 27 and 28.
- After determining $\phi_{local}$ boundaries for each of the slices in $\theta_{local}$, we plot these $\phi$ boundaries versus $\theta_{local}$ (see Figs. 29, 30, 31, 32, 33 and 34). An additional point of $\phi_{local} = \pm$ 27.5° (around the center of the sector) at $\theta_{local} = 40°$ is added to reduce the fluctuations at high $\theta_{local}$. Finally, the points are fitted with the function $\phi = p_0 + p_1 \ln\theta + p_2\ \theta + p_3\ \theta^2$.

Fig. 35 shows the $\phi$ versus $\theta$ for the collected electrons before and after applying the DC fiducial cuts.

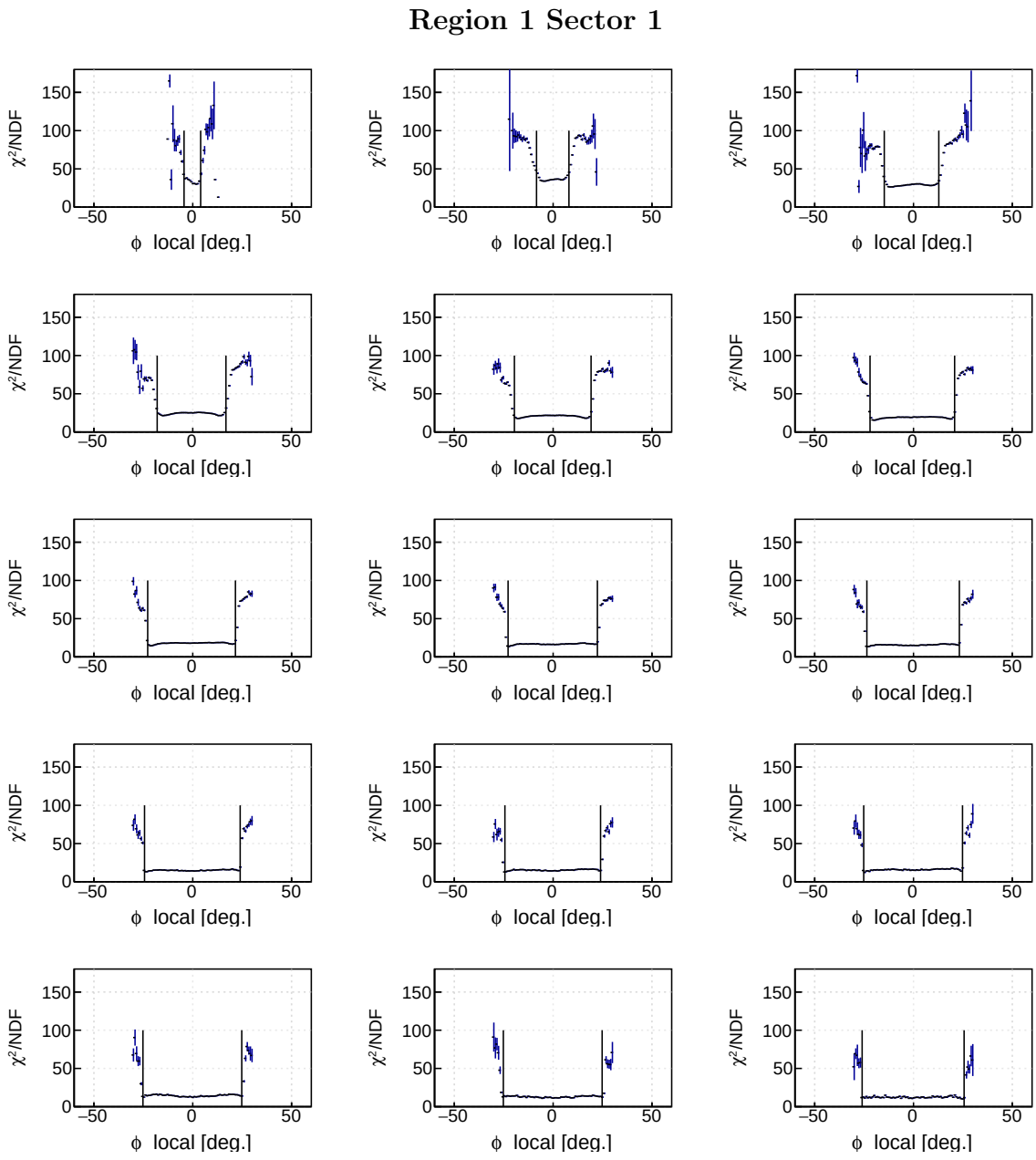



FIG. 23. Electron tracking $\chi 2$/NDF as a function $\phi$ in $\theta$ bins in Region 1 Sector 1.

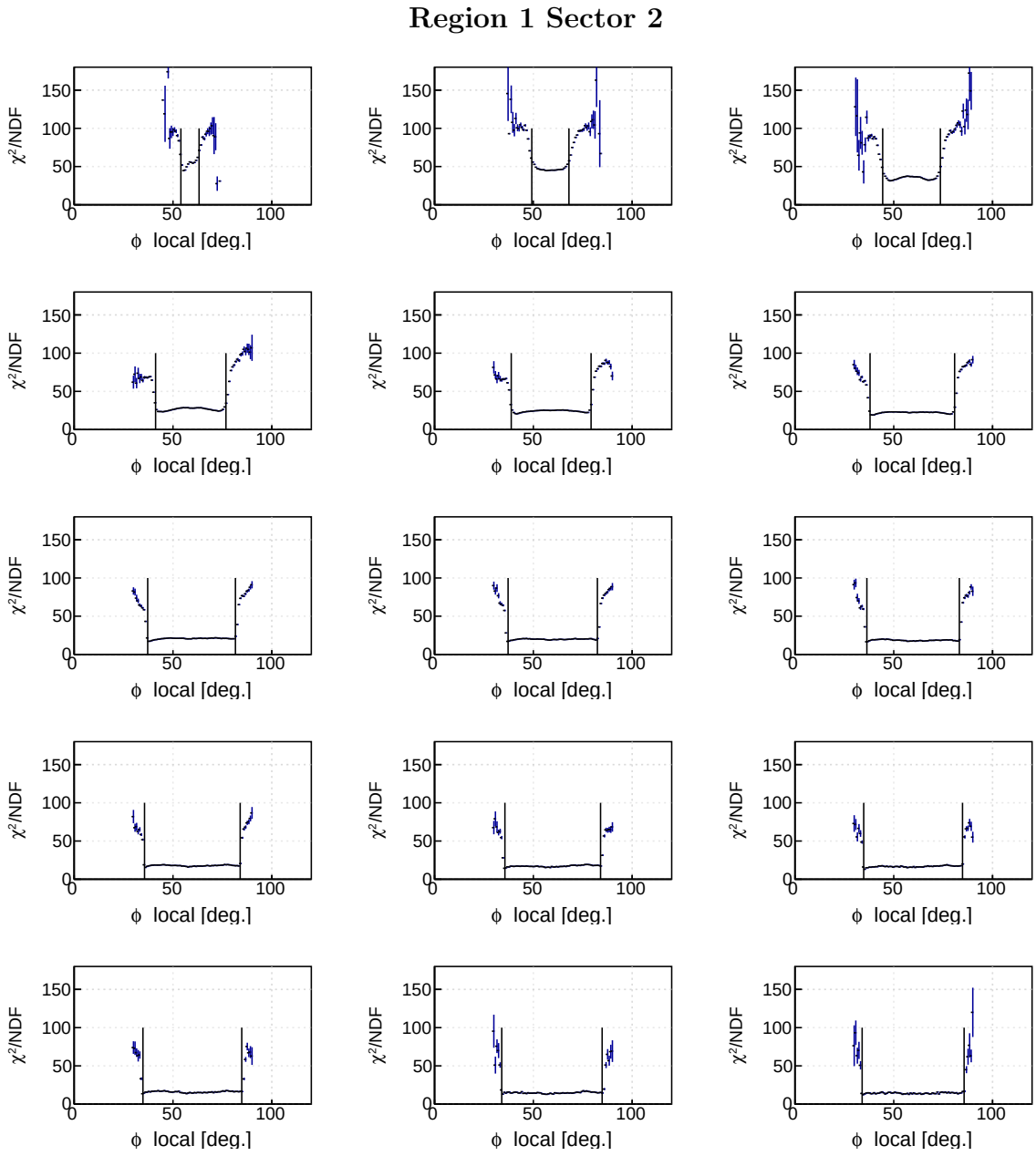



FIG. 24. Electron tracking $\chi 2$/NDF as a function $\phi$ in $\theta$ bins in Region 1 Sector 2.

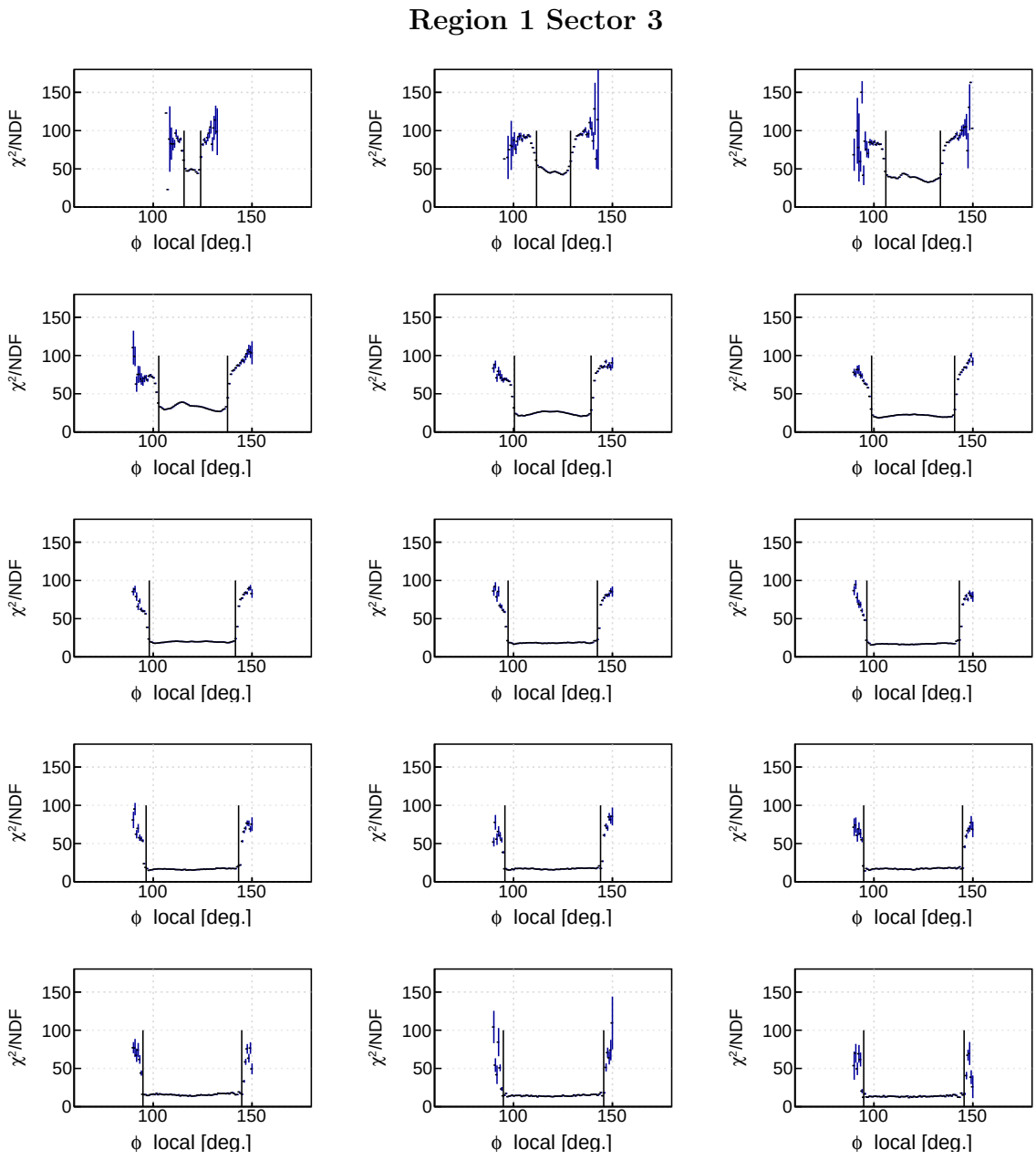



FIG. 25. Electron tracking $\chi 2$/NDF as a function $\phi$ in $\theta$ bins in Region 1 Sector 3.

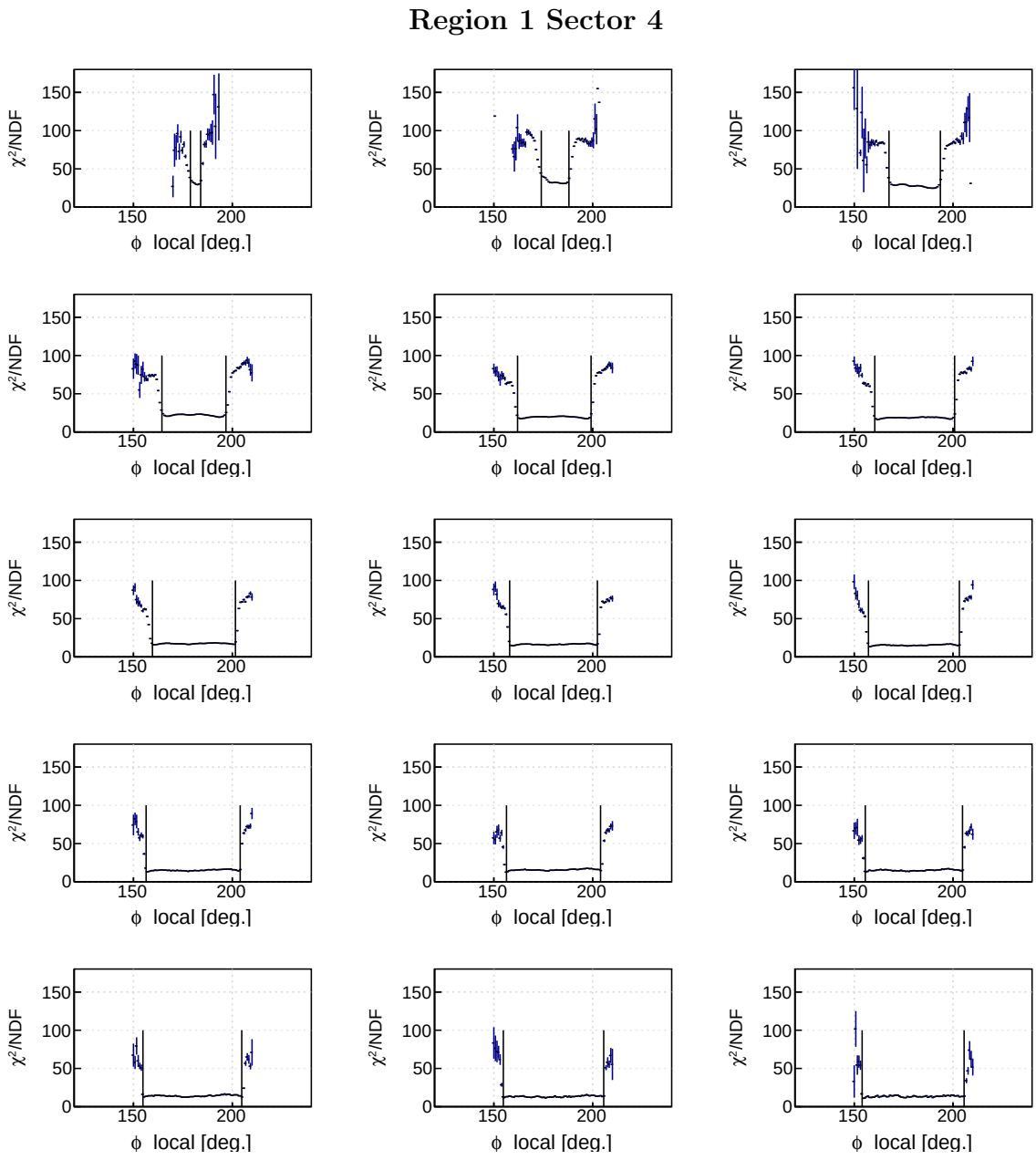



FIG. 26. Electron tracking $\chi 2$/NDF as a function $\phi$ in $\theta$ bins in Region 1 Sector 4.

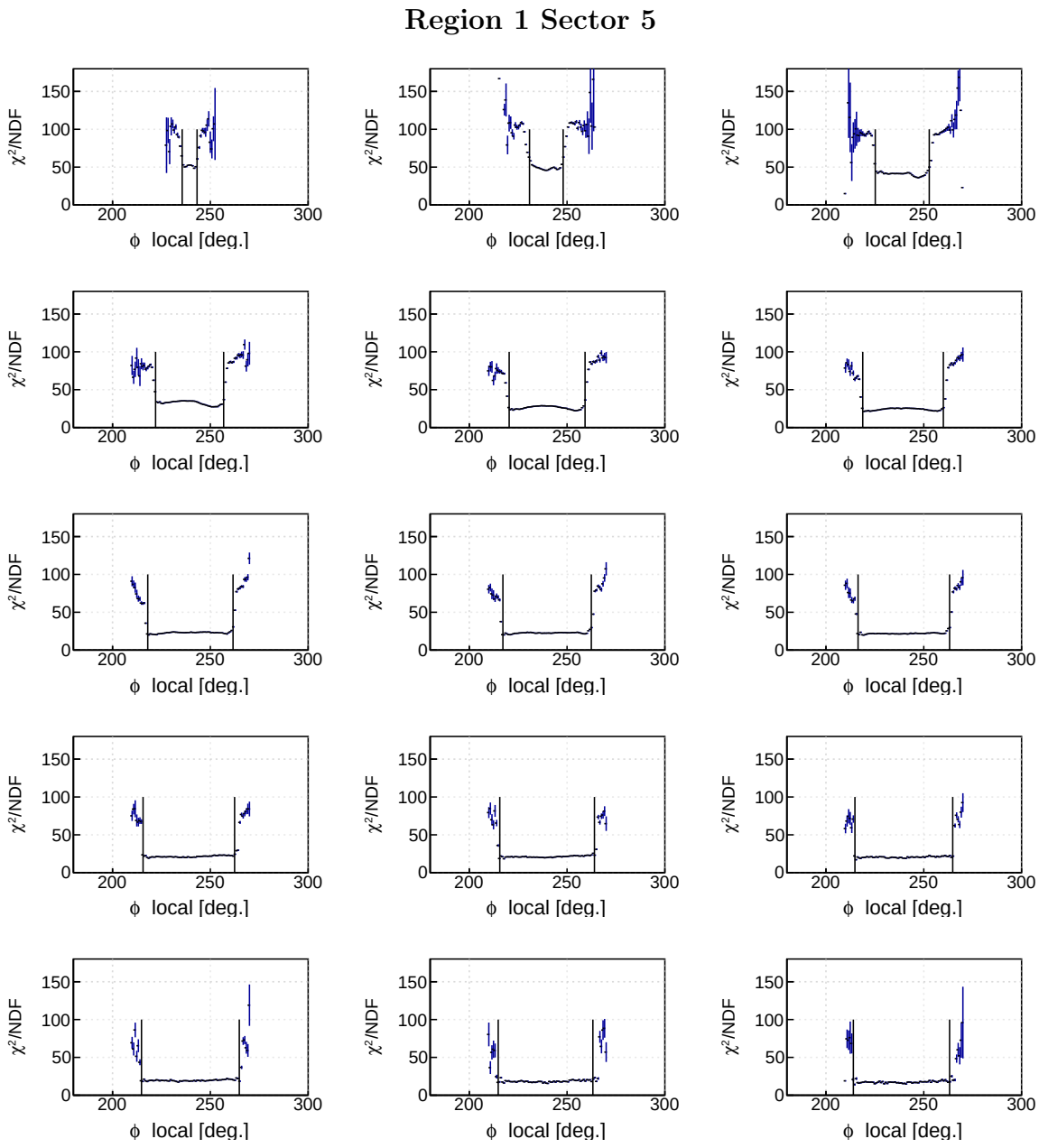



FIG. 27. Electron tracking $\chi 2$/NDF as a function $\phi$ in $\theta$ bins in Region 1 Sector 5.

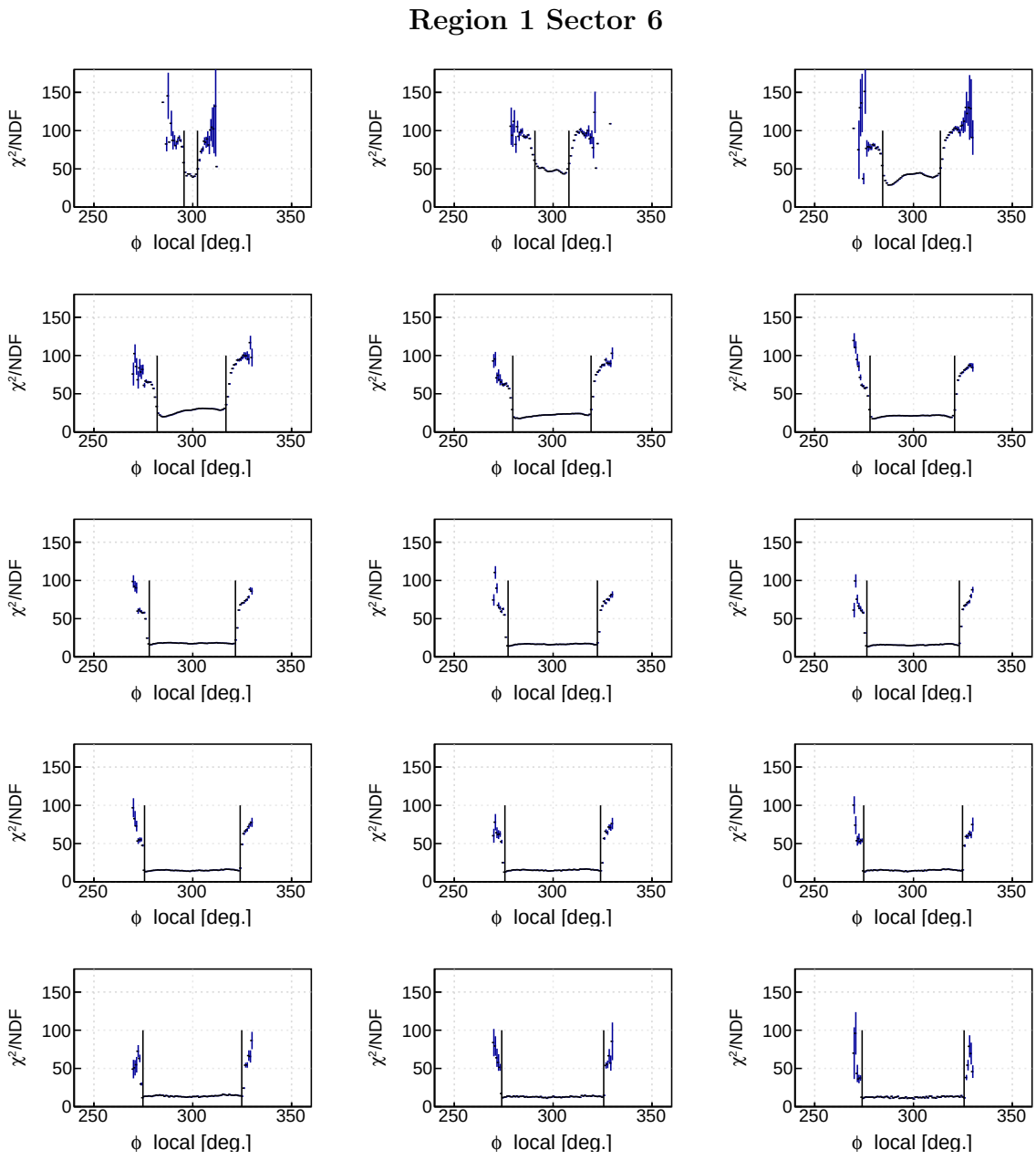



FIG. 28. Electron tracking $\chi 2$/NDF as a function $\phi$ in $\theta$ bins in Region 1 Sector 6.

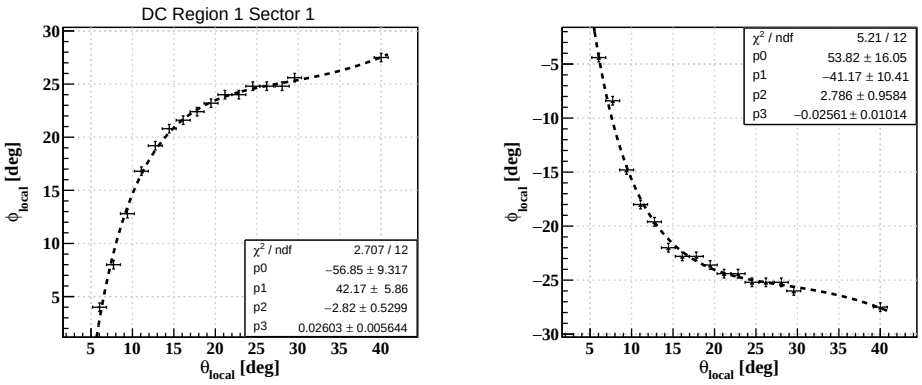



FIG. 29. Fitting for the cut border in DC Region 1 Sector 1 for electrons.

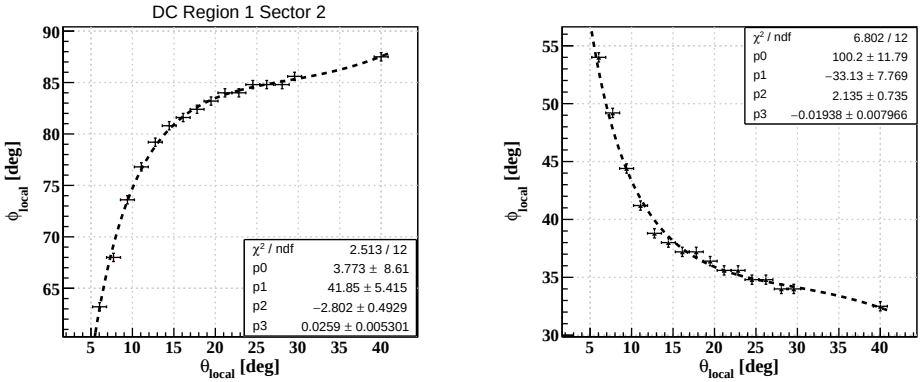



FIG. 30. Fitting for the cut border in DC Region 1 Sector 2 for electrons.

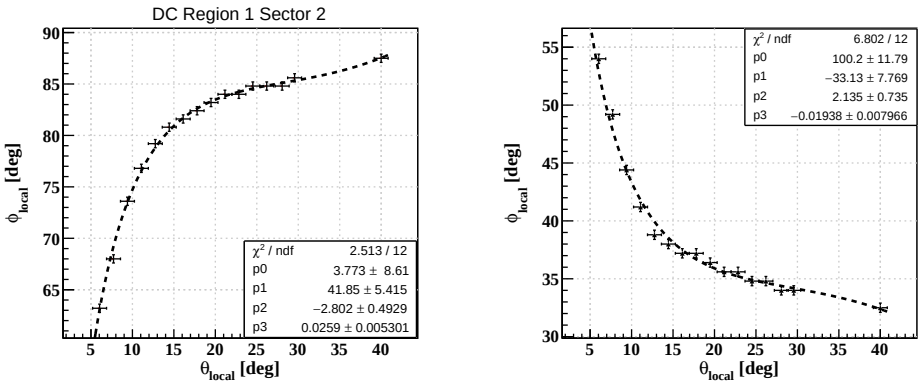



FIG. 31. Fitting for the cut border in DC Region 1 Sector 3 for electrons.

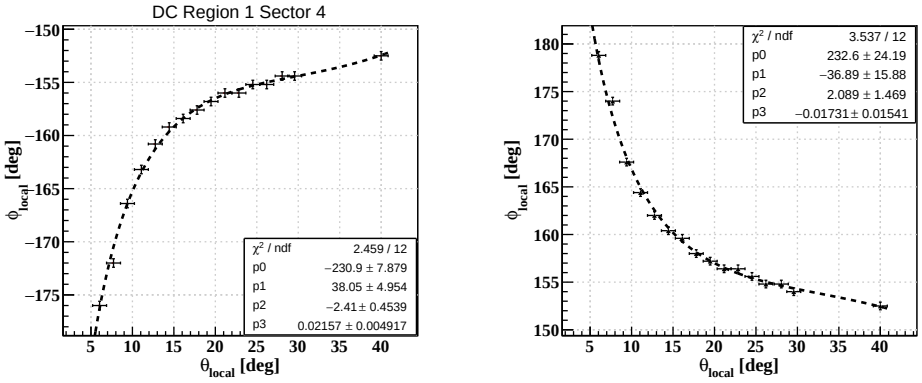



FIG. 32. Fitting for the cut border in DC Region 1 Sector 4 for electrons.

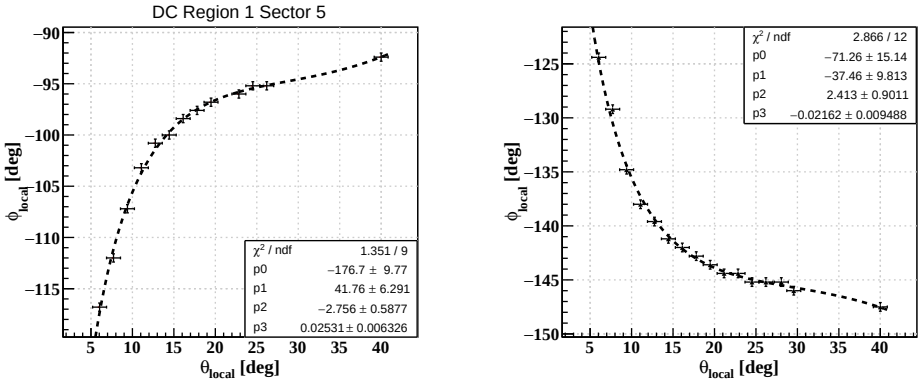



FIG. 33. Fitting for the cut border in DC Region 1 Sector 5 for electrons.

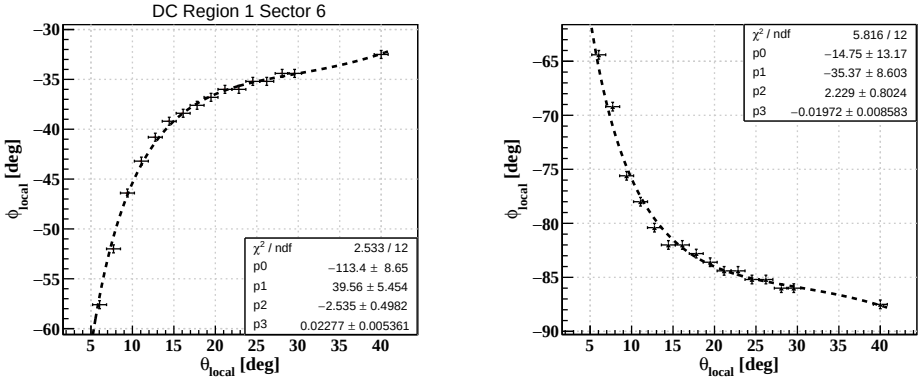



FIG. 34. Fitting for the cut border in DC Region 1 Sector 5 for electrons.

### Result of DC Fiducial Cuts

The following figure demonstrates its application after the development of fiducial cuts for the DC.

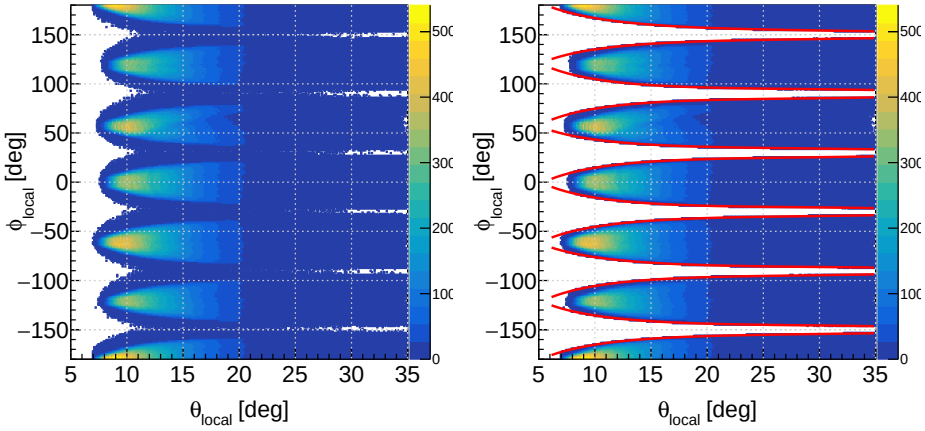



FIG. 35. $\theta_{local}$ vs. $\phi_{local}$ distributions (left) and Same distributions with DC fiducial cuts applied (right).

### 4.2.5 PCAL FIDUCIAL CUT

When electrons interact with the caloriemeter they produce electromagnetic showers. If the electron hits the calorimeter near its edges, a portion of the shower may fall outside the detector, resulting in incomplete energy collection. This leads to an underestimation of the reconstructed energy, which can affect the accuracy of the sampling fraction. Additionally, it becomes difficult to measure the cluster size accurately in these cases. Both factors reduce the calorimeter's ability to identify electrons correctly. Since the electron's momentum and angle are precisely measured using the drift chambers (DC), a perfectly accurate sampling fraction (SF) calculation is not essential. However, to match the agreement between the data and Monte Carlo, we still applied loose fiducial cuts in the PCal to remove the events at the edge of the detector.

During the investigation, we observed the strong dependence of the sampling fraction

distribution on PCal sectors. We decided to develop the fiducial cuts on PCal for each sector separately. Also, cut borders are defined independently for all calorimeter coordinates, namely, U, V, and W.

To determine fiducial cuts on PCal, we followed the following procedures:

- We calculated the sampling fraction for each sector as a function of calorimeter coordinates, namely, U, V, and W.
- For each distribution, we fitted the distribution of SF vs. U (V and W) with a constant function in the plateau region, where the distribution is flat (40 - 100 cm).
- After this, we calculated the coordinate for the distribution, where the average SF drops to 90% of the fit.
- Finally, that point was selected as a minimum cut for that sector's U, V, or W bar.
- In CLAS12, we have 6 PCal sectors and 3 PCal coordinates, so altogether, we had 18 cut points for PCal fiducial cuts. Figs. 36, 37 and 38 collectively illustrate the SF vs. the U, V and W coordinates for sector 6 in PCal and extraction procedures.

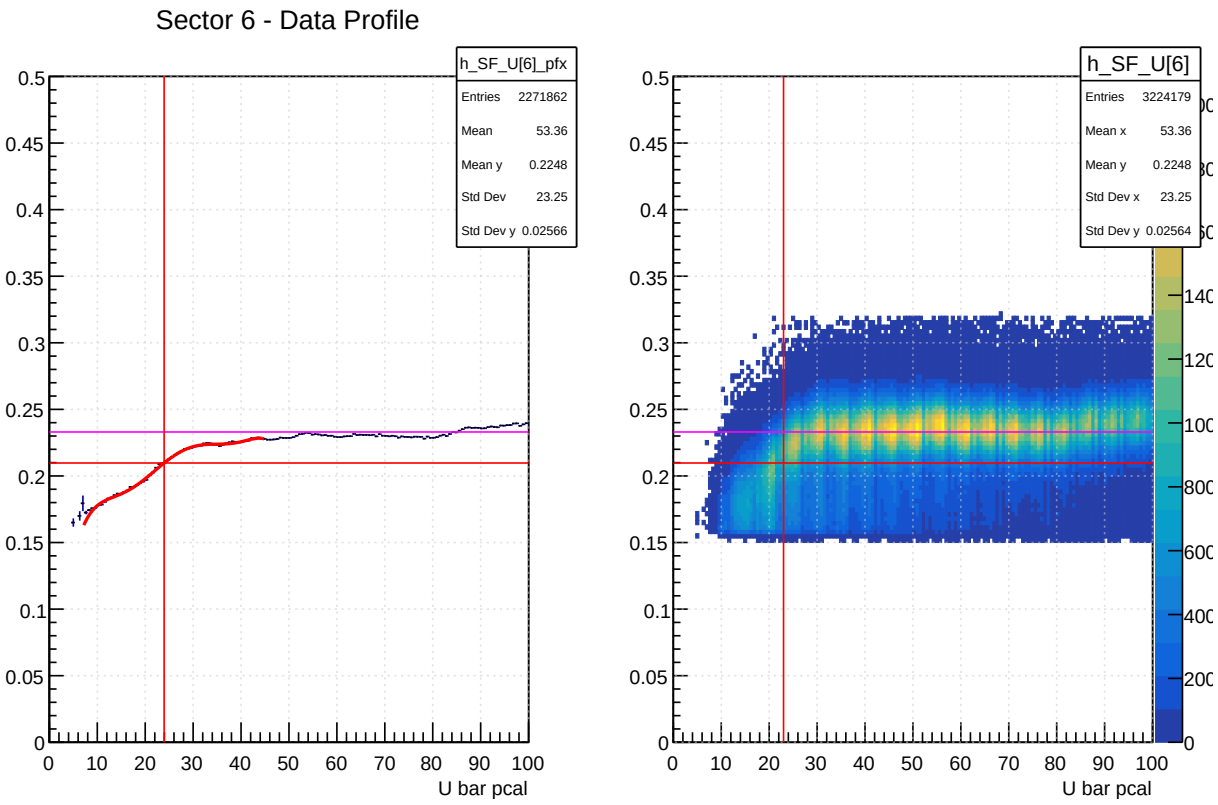



FIG. 36. SF vs U for sector 6 in PCal. On the left, the top horizontal line (in magenta) represents a constant fit (40 - 100 cm), and the bottom horizontal line (in red) represents the SF value corresponding to 90% of the fit. The vertical red line represents the corresponding x-coordinate where the average sampling fraction drops to 90% of the fit. On the right, the same fit lines and 90% lines drawn on top of the SF vs. U plot for sector 6.

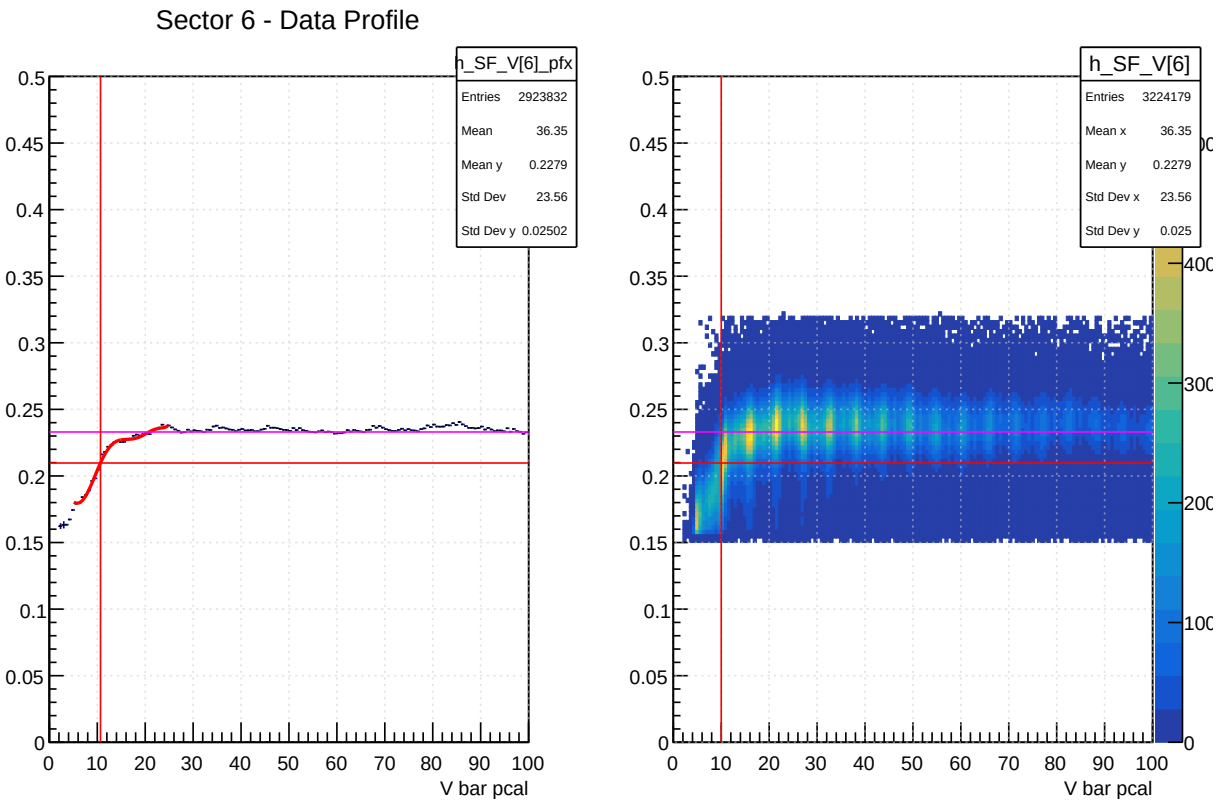



FIG. 37. SF vs V for sector 6 in PCal. On the left, the top horizontal line (in magenta) represents a constant fit (40 - 100 cm), and the bottom horizontal line (in red) represents the SF value corresponding to 90% of the fit. The vertical red line represents the corresponding x-coordinate where the average sampling fraction drops to 90% of the fit. On the right, the same fit lines and 90% lines drawn on top of the SF vs. V plot for sector 6.

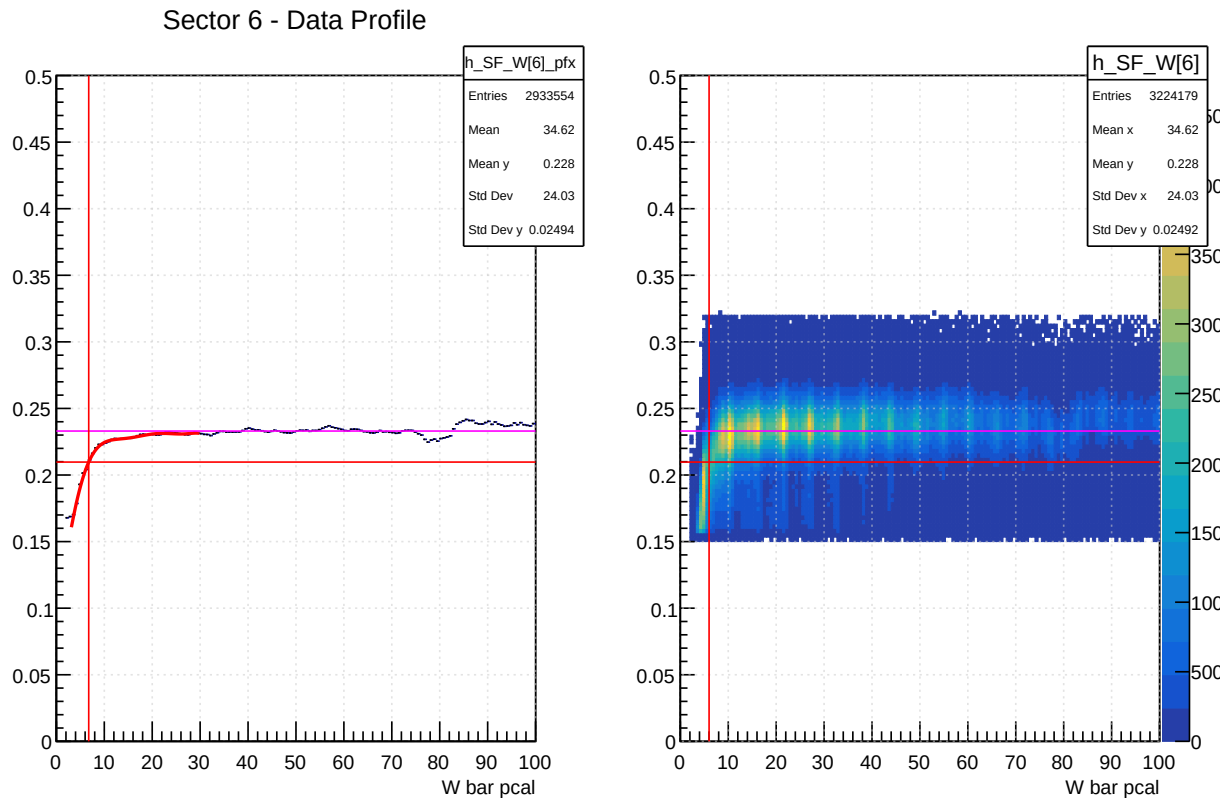



FIG. 38. SF vs W for sector 6 in PCal. On the left, the top horizontal line (in magenta) represents a constant fit (40 - 100 cm), and the bottom horizontal line (in red) represents the SF value corresponding to 90% of the fit. The vertical red line represents the corresponding x-coordinate where the average sampling fraction drops to 90% of the fit. On the right, the same fit lines and 90% lines drawn on top of the SF vs. W plot for sector 6.

We also observed that in the PCal W coordinate for sector 2, gains in some PCal bars were mis-calibrated. Fig. 39 demonstrated this effect. So, we decided to apply a cut to remove these bars in the analysis. This was done for both the data and the simulation.

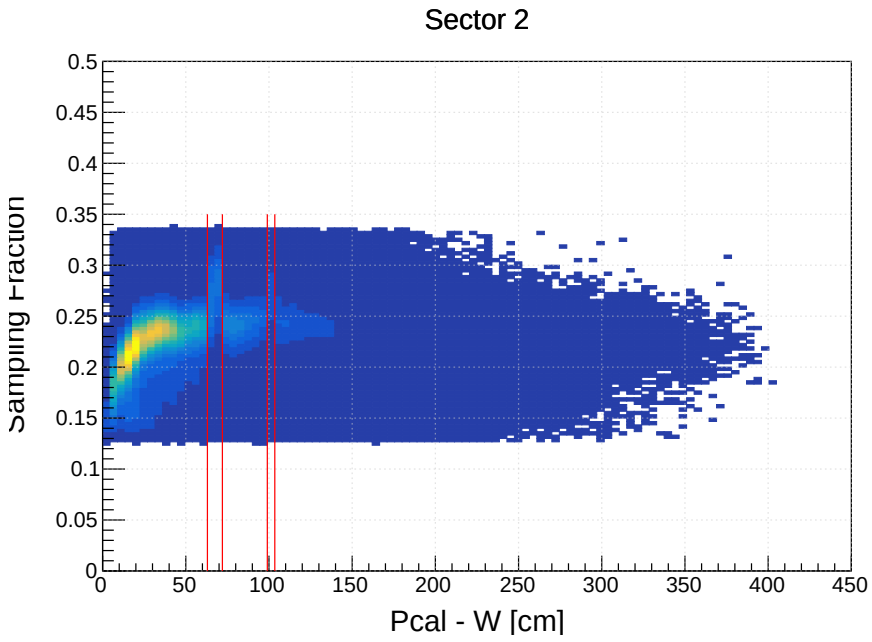



FIG. 39. SF vs. W coordinate for the sector 2. We clearly see a jump in SF for a couple of bars.

### 4.2.6 SAMPLING FRACTION CUT

The sampling fraction is defined as the total deposited energy in the calorimeter divided by the momentum of the charged particle,

$$SF = \frac{E_{PCAL} + EC_{in} + EC_{out}}{p}, \tag{88}$$

The $SF$ for electrons is expected to be a constant value of around 0.25. At the same time, for pions, it decreases with increasing momentum since pions deposit a constant amount of energy proportional to the fixed thickness of the calorimeter. In section 4.2.5, we described the development of fiducial cuts for PCal U, V, and W coordinates for electrons. Fig. 40 shows the distributions of sampling fraction vs. momentum for electrons before and after applying the PCal fiducial cuts. After applying these cuts, the sampling fraction versus momentum distribution for electrons appears much cleaner.

Now, as the next step, we slice this SF vs distributions along the x-axis in different

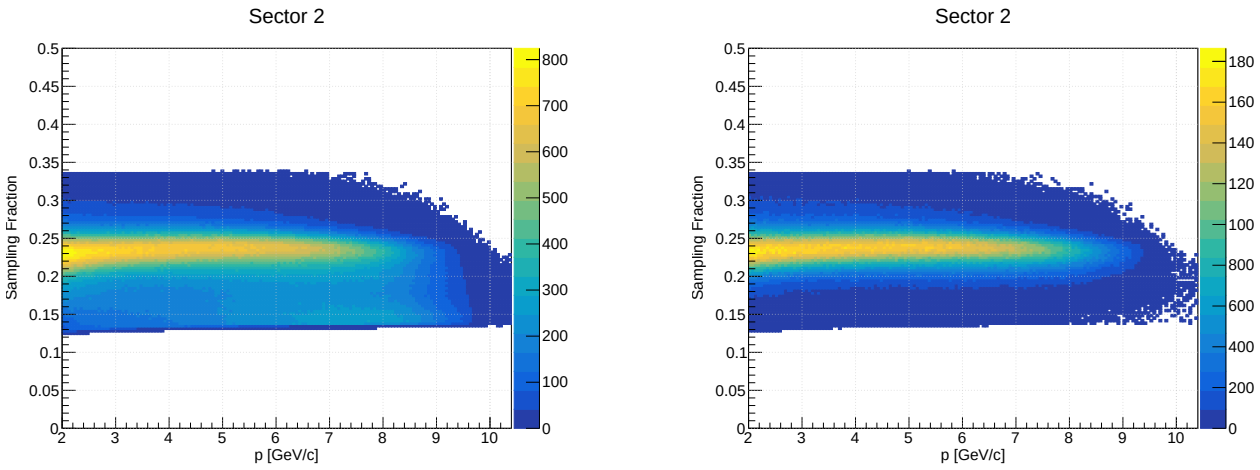



FIG. 40. $SF$ versus momentum for identified electrons (PID = 11) before [left] and after[right] the PCal fiducial cut, for sector 2 of PCal.

momentum bins. Then, for each slice in the momentum bin, we fit this slice with a one-dimensional Gaussian distribution. We extract the mean and sigma of the Gaussian distribution. Finally, we apply a cut of 3-sigma around the mean position for each momentum bin. This is shown in Fig. 41, and the curve obtained after extracting mean and sigma for all momentum slices is shown in Fig. 42.

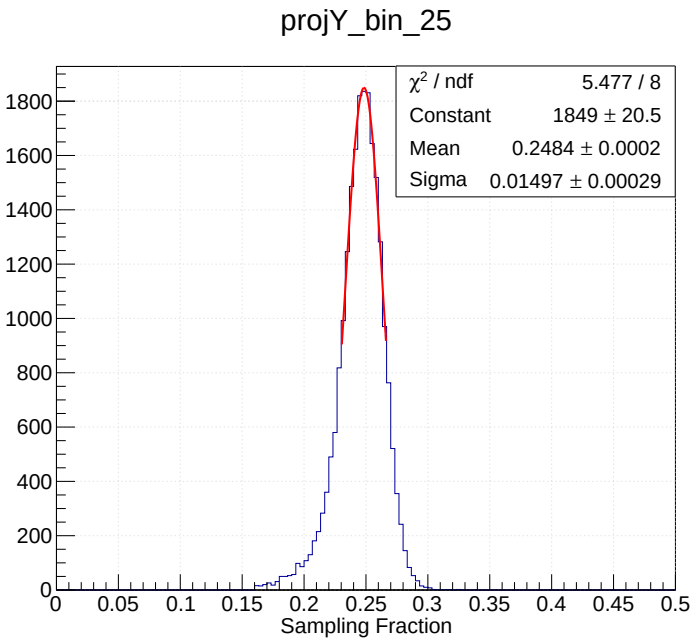



FIG. 41. Projection of a slice of $SF$ vs momentum distribution fitted with a Gaussian.

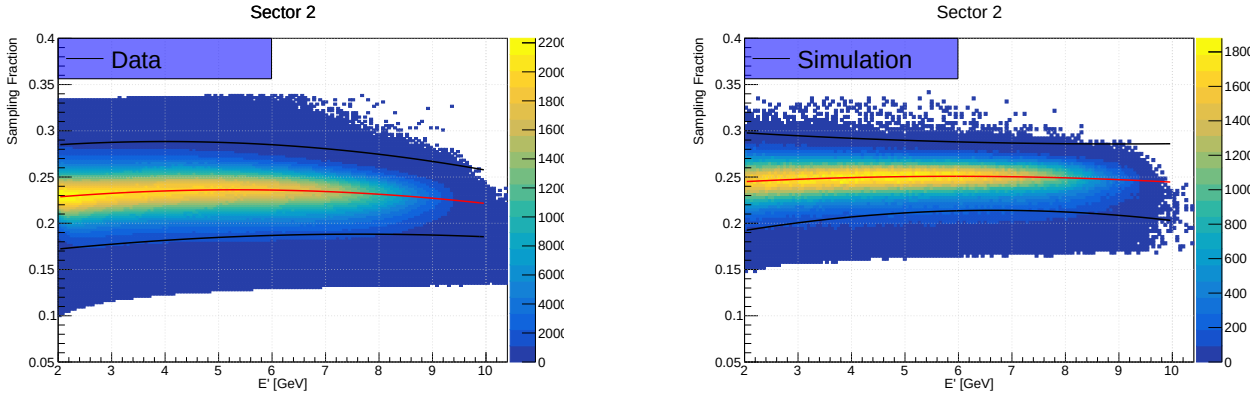



FIG. 42. $SF$ versus momentum plot for identified electrons with black curves showing three sigmas around the mean of projection fit of each slice, for data (left) and simulation (right).

# CHAPTER 5

# RTPC DATA PROCESSING

The BONuS12 RTPC was built to detect slow-moving spectator protons. In this chapter, first, I will give an introduction to RTPC reconstruction software in a nutshell, Then I will explain in detail the quality requirements for the reconstructed tracks inside the RTPC, the correspondence criteria between the final-state scattered electron and the RTPC recoils, the selection cuts for the DIS events, and our sample of VIP DIS sample of interest.

## 5.1 RTPC RECONSTRUCTION SOFTWARE

During the experiment, the CLAS12 Data Acquisition System was used. Whenever a scattered electron is detected in one of the ECal detector sections, a data readout trigger initiates the recording of the event readout in the forward and central detector of CLAS12. It also sends back trigger initiation information to the RTPC data acquisition system using optical fiber links. The RTPC used the fiber readout mode of two dedicated Back End Units (BEUs). Both BEUs distributed the global system *clock*, trigger, and synchronous commands to 36 Front End Units (FEUs), which needed to read out all RTPC channels. For each trigger electron, the BEU performed local event-building tasks by gathering data packets from all FEUs belonging to the same event (matching time stamps and event IDs). Data transfer rates of 200 MBytes/s were routinely achieved. The data recorded by the RTPC DAQ system contains information on the readout pad number (pad ID), timing information, and ADC information of each hit. The decoder software translates this information to the physical location in the detector (pad information) and time relative to the trigger time. This collective information is written to a file for track reconstruction for each triggered event.

The track-finding algorithm in the reconstruction software sorts individual hits, along with the information of ADC, timing, and pad ID, into possible track candidates for each event. Initially, hits are sorted by time bin. When chains of hits within the same time slice are geometrically close, they form "proto-tracks", and each hit belonging to a proto-track is assigned a unique track ID. If there are hits that are within the same time slice but are not close enough to any existing proto-track candidates, they serve as seeds for new

proto-tracks. The criteria for forming a proto-track are as follows: adjacent hits must be within a maximum distance of 8 degrees or four pad widths in azimuthal angle ($\phi$), 19 mm (approximately five pad lengths) along the detector length, and within a time difference of 360 ns, or 3 DREAM chip time slices. If two different proto-tracks share a hit, both track IDs are assigned to the track. Eventually, all proto-tracks that share the same hits and track IDs are merged to form hit-level track candidates. However, there may be tracks that cross each other closely in time and space but are not from the same physical track. These crossing tracks need to be disentangled, a process handled by a disentangling algorithm in the RTPC reconstruction software. For more details on the disentangling process, refer to [37]. After this step, the measured timing of each hit is converted to its ionization position from the central axis of the detector. Also, some tracks might be out of synchronization with the trigger time. Hence, we introduce a timing offset based on the fact that every track of a particle from the target must cross the cathode, which is located at the maximum drift time. This maximum drift time is a known parameter for our detector. Several simulation studies were done using Garfield++ to calculate radial drift velocity and Lorentz Drift correction. This Lorentz Drift is the additional kick experienced by the electrons due to the magnetic field in addition to moving radially in the radial electric field. After all these corrections, we have a unique position in $r$, $\phi$, and $z$ for each point on a track. Finally, all these points are fitted by a helix fitter, assuming the charged particles follow a helical path. Using the radius of curvature of the helix and magnetic field of the CLAS12 solenoid, we can calculate the momentum of the charged particle. Similarly, the length of the tracks of the particle, along with the ADC deposited by a track, helps in PID determination for different particles in the RTPC detector.

So, we apply several quality criteria to select good tracks in RTPC. The quality criteria are based on the performance of the helix fitter.

## 5.2 RTPC TRACK SELECTION CRITERIA

A good positively charged RTPC track passes the following requirements:

- **Radius of curvature**: The RTPC detects positively charged spectators. The radius of curvature from the helix fitter of the track must be negative for positively charged particles. Fig. 43 shows the radius of curvature distribution for the collected RTPC tracks.

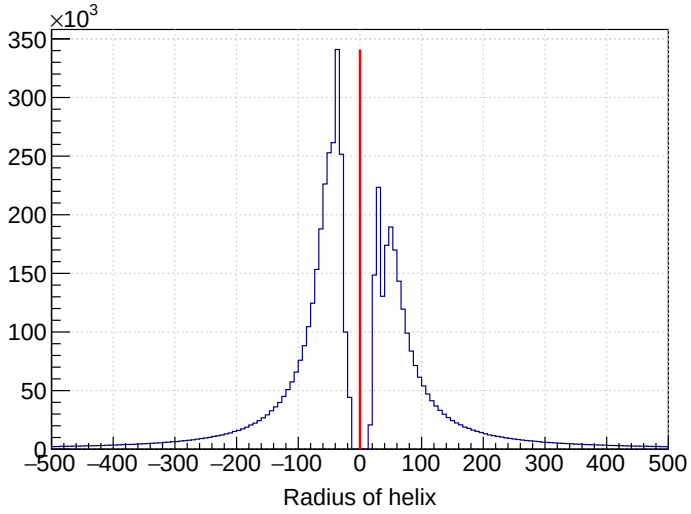



FIG. 43. The radius of curvature (in mm) of the reconstructed tracks in the RTPC. Negative radii are associated with positively charged particles.

- **Quality of the helix fit**: We applied a cut on the $\chi^2$/ndf value for each helix fit to be less than 5. Fig. 44 shows the $\chi^2$ distribution for the positive tracks originating within the RTPC.

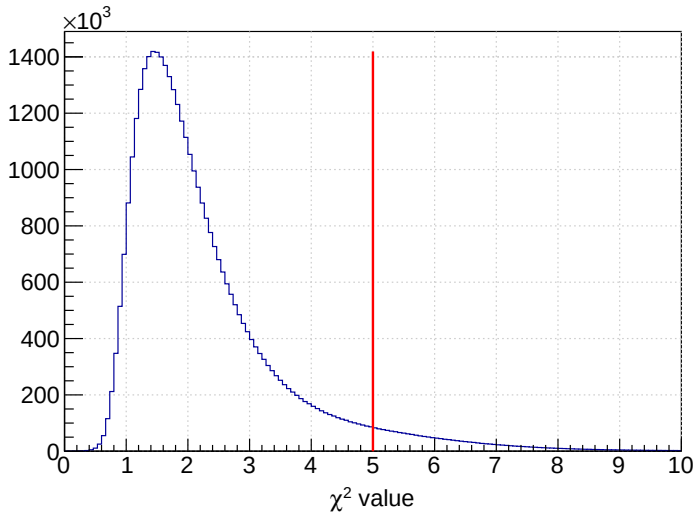



FIG. 44. $\chi^2$ distributions for reconstructed tracks in RTPC.

- **Maximum radius ($rmax$) cut**: $rmax$ is the distance between the beamline and the farthest hit, i.e., the closest to the readout anode. We require that $rmax > 67$ mm and $rmax < 72$ mm for a good track in order to ensure that the track has crossed the whole drift region up to the anode. See Fig. 45 for the relevant distribution. ’

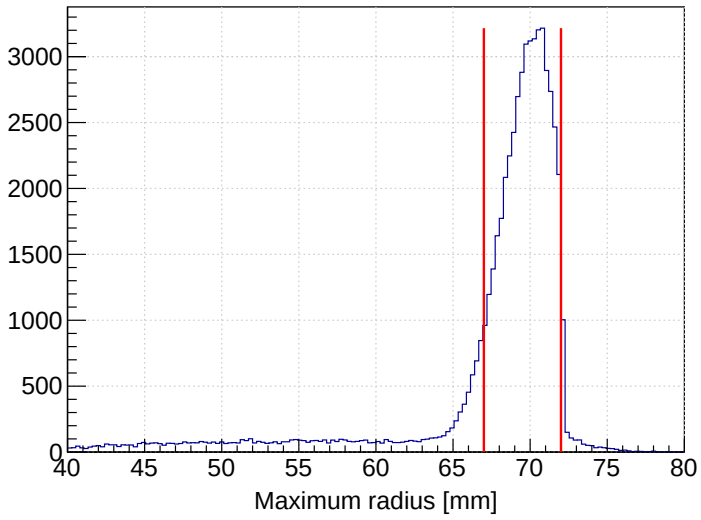



FIG. 45. The maximum radius for the tracks reconstructed in RTPC.

- **Number of hits in a track**: For a good track track candidate, we put a criterion that it has made at least 10 visible reconstructed hits in the drift region of the RTPC. See Fig. 46 for the relevant distribution.

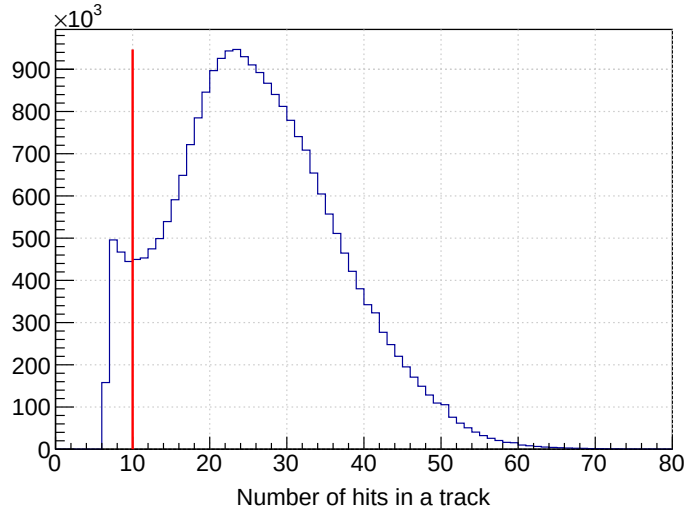



FIG. 46. The number of hits of reconstructed tracks in RTPC.

- **Timing cut ($t_{diff}$):** The timing parameter($t_{diff}$) is the time difference between the expected on-time track and the measured timing of the track in the RTPC. In other words, it is the time shift that is applied to the chain of the hits in a track to have the first hit being produced right after the cathode. It was observed that this sensitive track quality parameter changes even within the same run. Fig. 47 shows the $t_{diff}$ distribution for the collected RTPC tracks in a single file from a 10 GeV $D_2$ target run (12736), the mean and the sigma of the fitted $t_{diff}$ variable as a function of the file number in the same run.

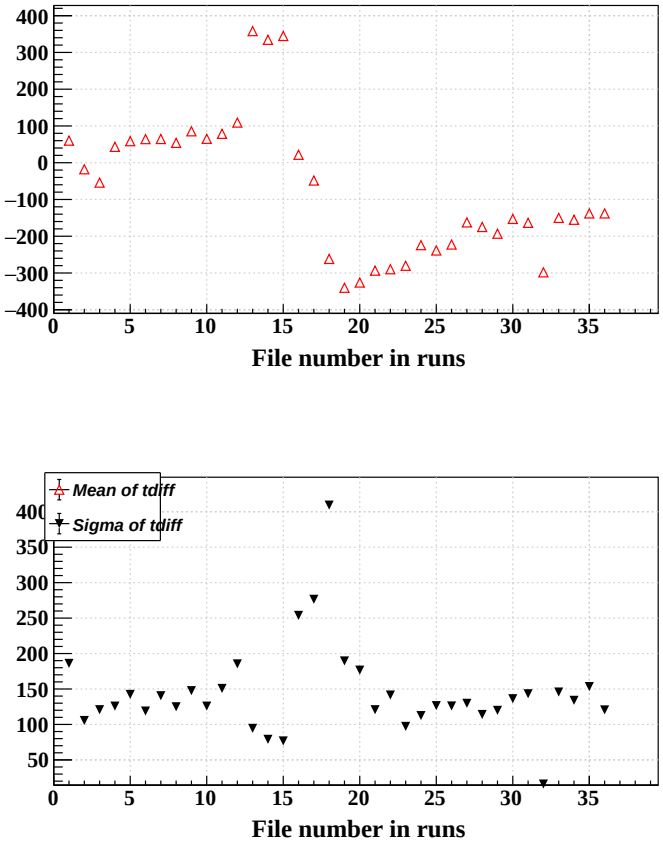



FIG. 47. Mean (top) and Sigma (bottom) for $t_{diff}$ distributions as a function of the data file number in run 12736.

In addition to the $t_{diff}$ file dependence, we observed that $t_{diff}$ depends on the maximum radius ($rmax$). This means that the time shift of the chain of the hits for a track is strongly correlated with the geometrical location of the farthest point in a track, which should not be. Fig. 48 shows this correlation between $rmax$ and $t_{diff}$ in run 12736. In the latter figure, the strips stand for the changing $t_{diff}$ as a function of the file number in the run 12736. Additionally, one sees the linear dependence between the two variables, $rmax$ and $t_{diff}$. This linear dependence has been parameterized and corrected for by defining a new corrected timing variable as $t'_{diff} = t_{diff} + 120(rmax - 70.0)$. Fig. 49 shows the dependence between $rmax$ and the corrected timing variable $t'_{diff}$ after the

correction is applied. One can clearly see that the dependence has been removed.

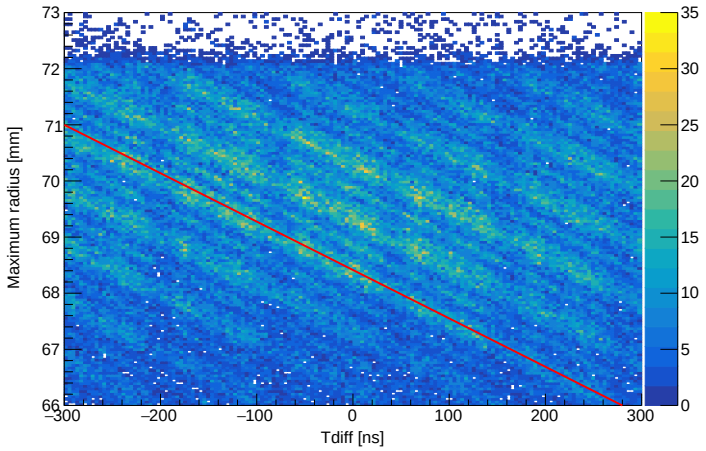



FIG. 48. $r_{\max}$ vs. $t_{\text{diff}}$ for run 12736. The red line shows the fit to the dependence between the two variables.

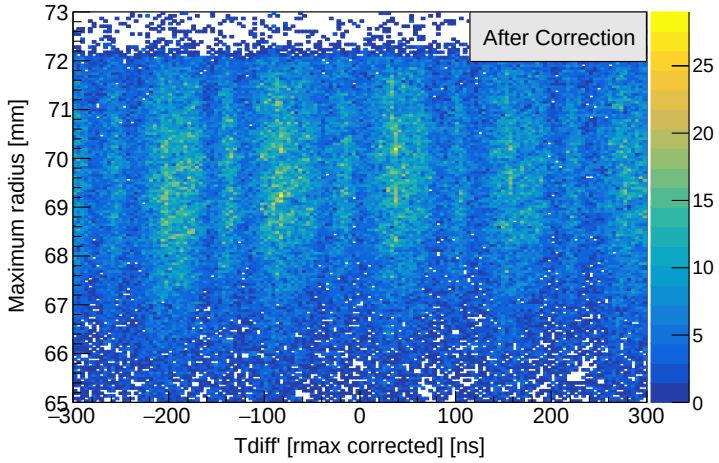



FIG. 49. $r_{\max}$ vs. $t'_{\text{diff}}$ [corrected $t_{\text{diff}}$] for run 12736.

After correcting $t_{diff}$ based on $rmax$, we again extracted the mean and sigma of corrected $t_{diff}$ for each run and data file and updated the table with these corrected values of mean and sigma to be used as a track selection cut. Fig. 50 shows the 1D distributions for the mean and sigma of the raw $t_{diff}$, after a first correction iteration, and the final distributions after the second iteration. One can clearly see that the final "magenta" distributions of the means and the widths of $t_{diff}$ for the data files are more coherent compared to the raw and the first iteration results.

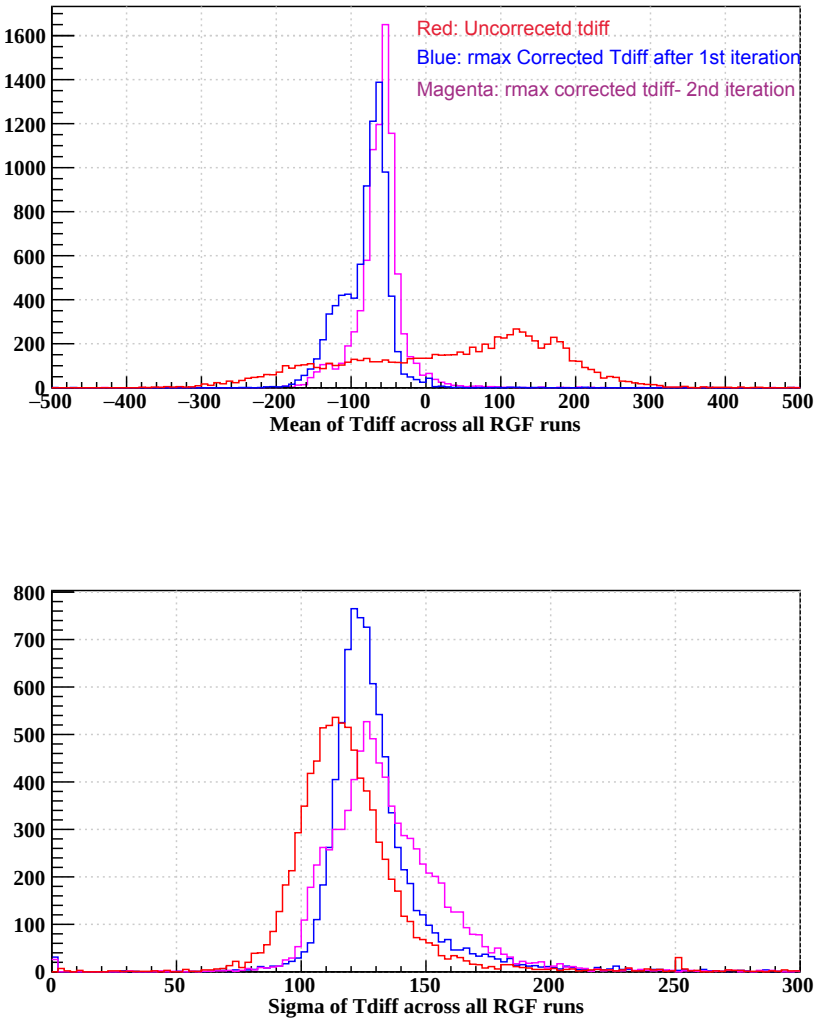



FIG. 50. The mean (top) and sigma (bottom) of $t_{diff}$ distributions for all RG-F Summer 2020 data. The red distribution is for old $t_{diff}$, blue is after a first correction based on $rmax$, and magenta is after the second final iteration of the correction performed by taking fit values from first iteration as initial parameters for the second iteration.

- **RTPC vertex cut**: To selects protons only coming from the target and to optimize the geometrical vertex cut of the reconstructed tracks along the RTPC, we have monitored

the $v_z$ of the reconstructed tracks after the following successive selection steps (see Figs. 51, 52 and 53). The definition of the vertex correspondence, $dE/dx$ PID cuts, VIP and nDIS cuts are discussed in following sections.

- After applying all the selection cuts on the scattered electron except its vertex cut, and, for protons, applying cuts on the quality of tracks (see legend).
- After applying additional coincidence cuts between the scattered electron and the recoil proton in the final state.
- After applying additional VIP and spectator cuts. These coincident cuts and spectator cuts will be discussed in the next section.

After these studies, we safely decided that our target fiducial cut for proton selection should be [-21, +18] cm.

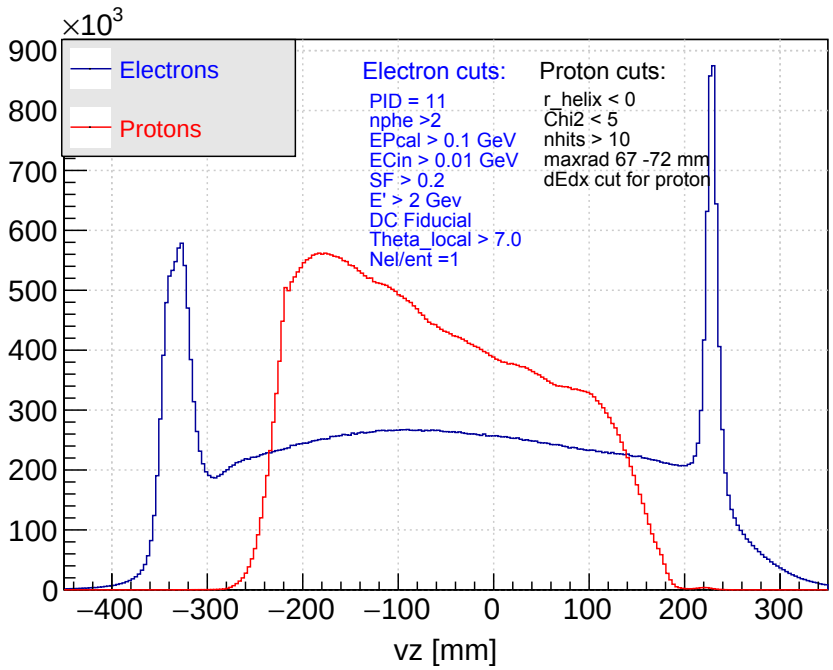



FIG. 51. $v_z$ distributions for electrons and RTPC tracks after the listed (in legend) quality cuts.

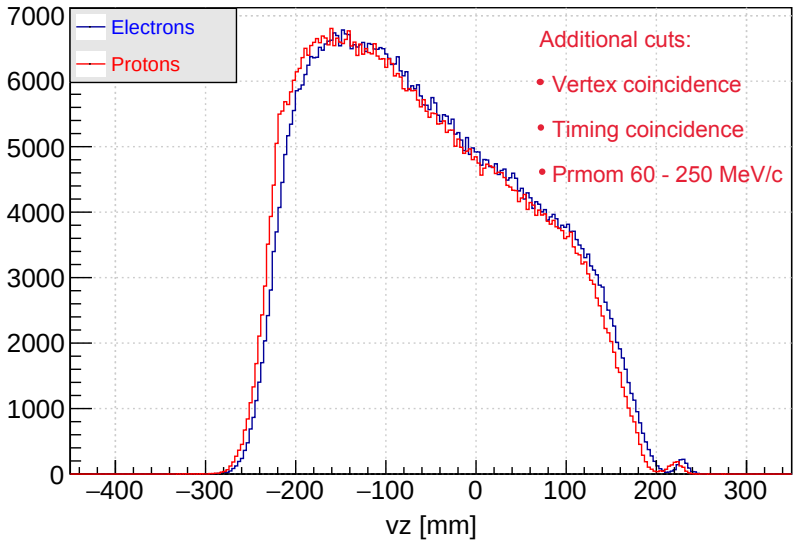



FIG. 52. $v_z$ distributions after the listed additional cuts in addition to the top quality cuts.

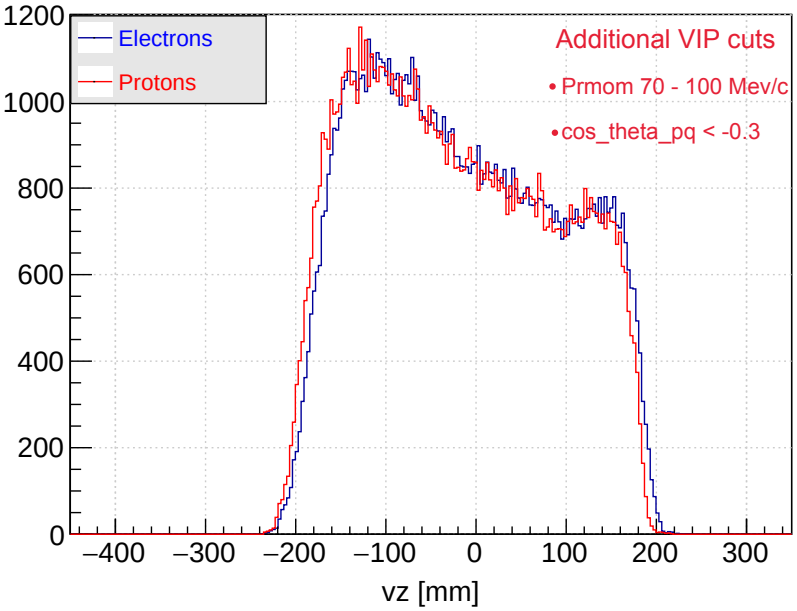



FIG. 53. $v_z$ distributions after all the cuts, including the nDIS VIP cuts.

### 5.2.1 VERTEX CORRESPONDENCE CUT

To ensure that the detected track in the RTPC and the electrons in CLAS12 originate from the same scattering event, we apply an additional vertex correspondence cut ($\Delta vz$). Fig. 54 shows the vertex difference between the electrons and the RTPC tracks from 10 GeV beam data. After sampling the accidental backgrounds that lie under signal peak and subtracting them [see chapter 6], the vertex difference for six different sectors in CLAS12 is shown in Fig. 55. Similarly, Fig. 56 shows the mean and standard deviation of the Gaussian fit. We clearly see the sector dependence in the $\Delta vz$ distribution.

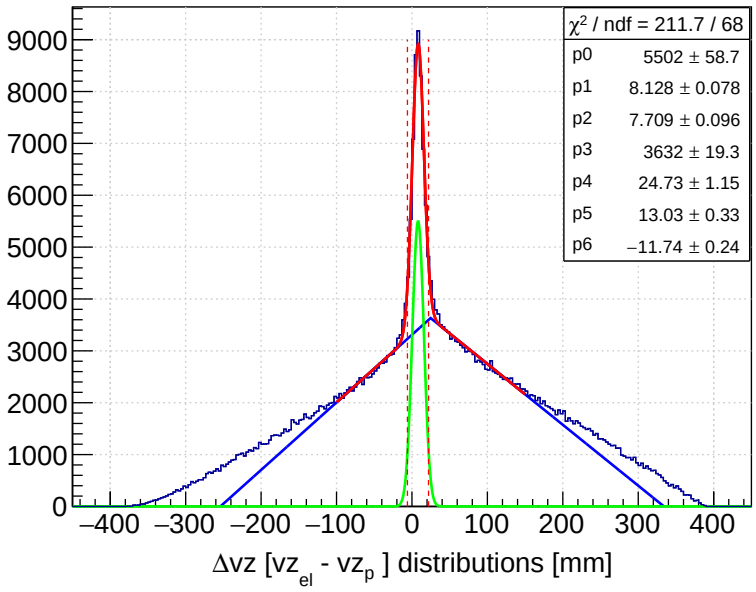



FIG. 54. Vertex difference between the scattered electron and the RTPC reconstructed tracks fitted with Gaussian signal over triangular background.

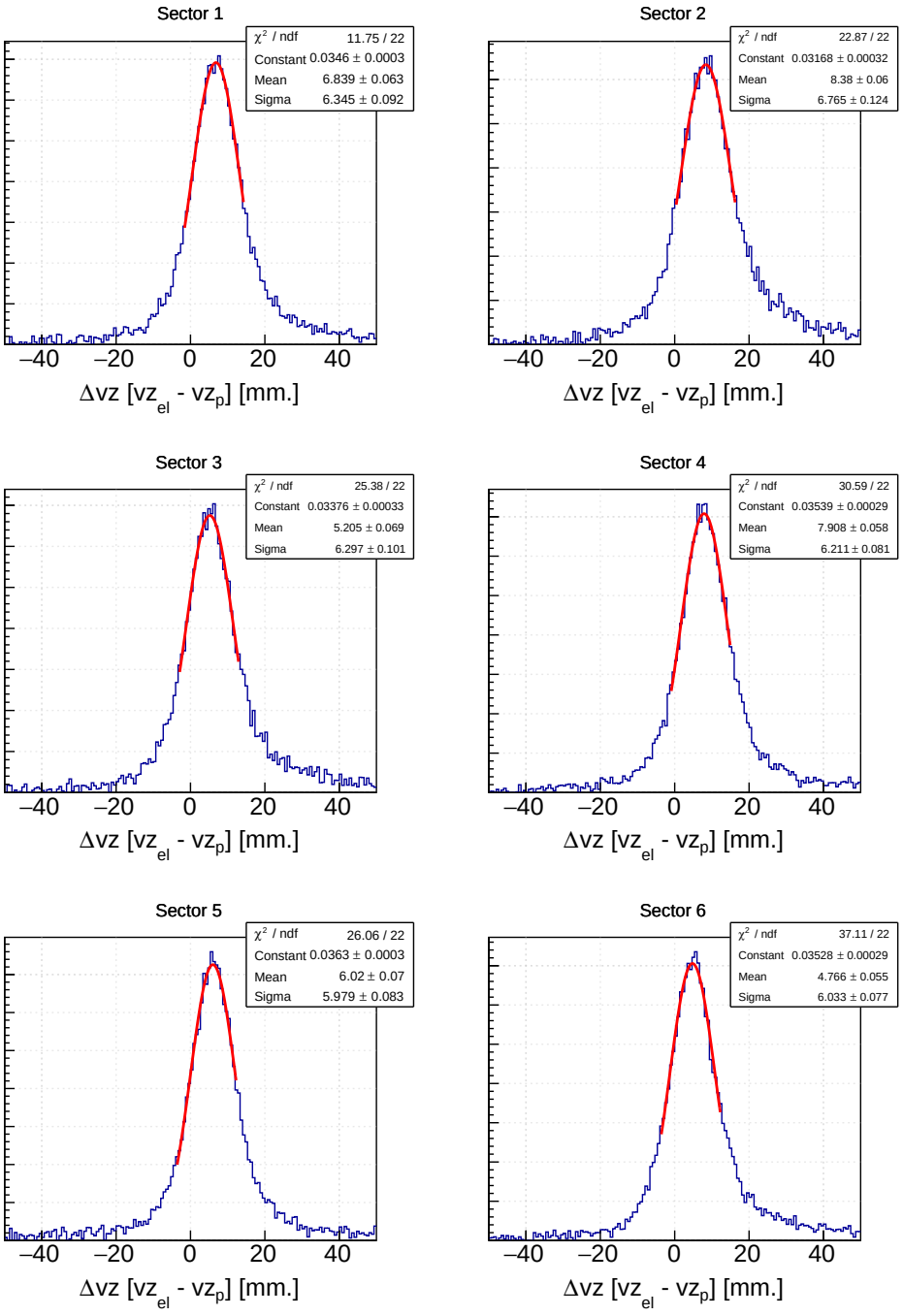



FIG. 55. Vertex difference between the scattered electron and the RTPC reconstructed tracks (with all DIS cuts) fitted with Gaussian fit for six different sectors of CLAS12.

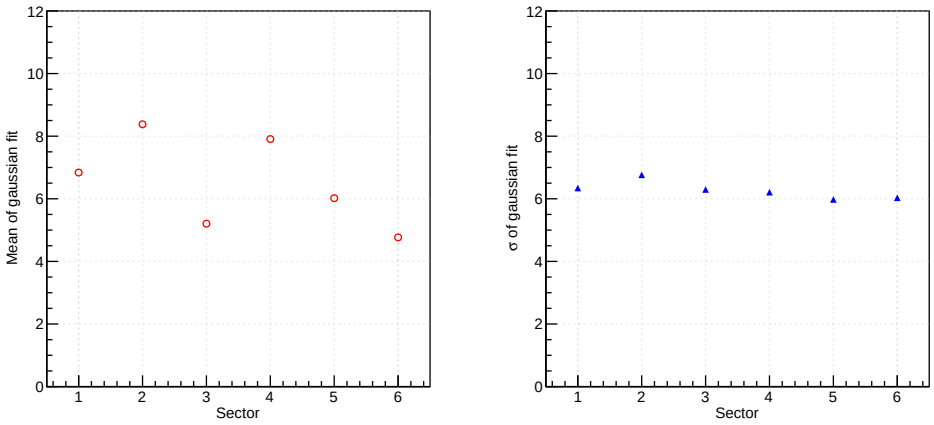



FIG. 56. The Mean and Standard Deviation of the Gaussian fit of deltavz distributions.

After calculating the mean and standard deviation of the signal region for each sector, to decide on how many sigmas around the mean we should take our cut, we calculated a **Signal Significance (SS)**. If the total number of event counts for any cut is given by $N$, and $S$ and $B$ represent the total signal and background counts within the cuts, then,

$$N = S + B,$$

$$S = N - B,$$

The standard deviation of the signal significance is given by

$$\sigma_S = \sqrt{\sigma_N^2 + \sigma_B^2},$$

Since these counts follow the Poisson distribution, $\sigma_N^2 = N$ and $\sigma_B^2 = B$, so we define Signal Significance (SS) as:

$$SS = \frac{S}{\sigma_S} = \frac{S}{\sqrt{N+B}} = \frac{S}{\sqrt{S+2B}}.$$

We want to maximize this quantity for a better signal that is free from backgrounds. We calculated this quantity for different values of the width of sigma around the mean. Fig. 57 shows the calculation SS for six different sectors of CLAS12 for different sigmas ranging from 1 to 3. After this, we calculate the sigma value at maximum SS. Fig. 58 shows this, and we see the fit shows 1.98 $\sigma$ as fit to six sectors. Hence, we decided on,

$$\Delta vz^{\text{cut}} = \langle \Delta vz \rangle_i \pm 2\sigma_i.$$

where $i$ represents the sector number of CLAS12 (1 to 6) $\langle \Delta vz \rangle_i$ and $\sigma_i$ are the mean and standard deviation of a Gaussian fit for each sector

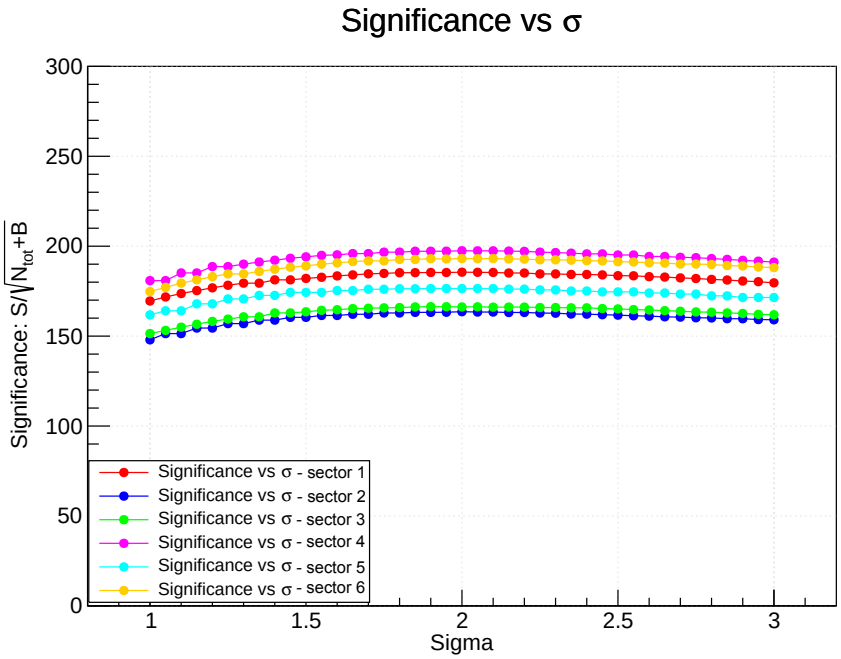



FIG. 57. Signal Significance for different widths of cuts around the mean, separately for each CLAS12 sector.

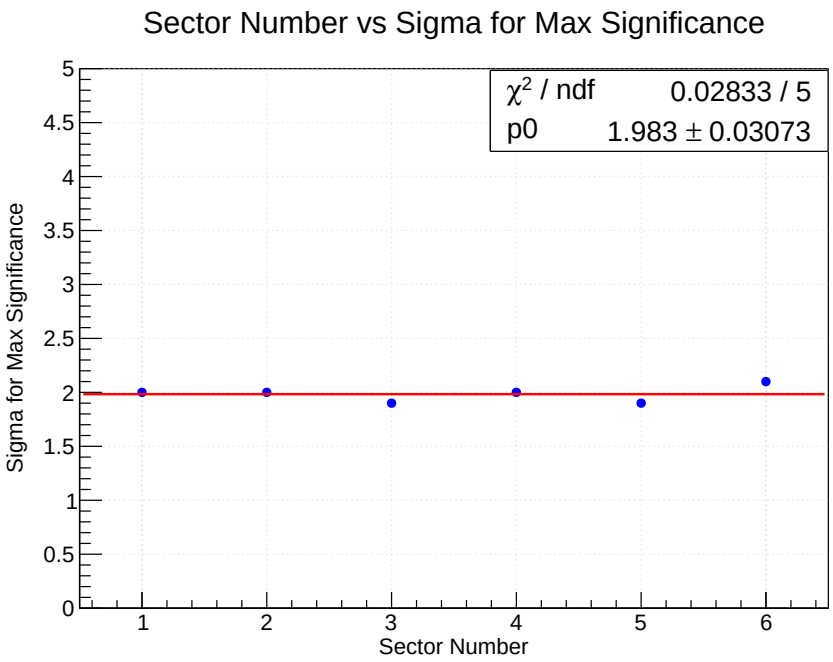



FIG. 58. Sigma value for maximum Signal Significance for each CLAS12 sector.

### 5.2.2 PID CUTS

Any ionizing particle that travels through a medium will lose energy due to interaction [mainly ionizing collisions] of the particle with the medium. The Bethe-Bloch formula gives the mean loss of energy of heavy particles traveling through a medium.

$$-\langle\frac{dE}{dx}\rangle = Kz^2\frac{Z}{A}\frac{1}{\beta^2}\left[\frac{1}{2}\ln\left(\frac{2m_ec^2\beta^2}{I}\right) - \beta^2 - \frac{\delta(\beta\gamma)}{2}\right], \tag{89}$$

where:

$\frac{dE}{dx}$ is the energy loss per unit path length,

$K$ is a constant,

$z$ is the charge of the particle,

$Z$ is the atomic number of the medium,

$A$ is the atomic mass of the medium,

$\beta$ is the velocity of the particle as a fraction of the speed of light,

$\gamma$ is the Lorentz factor,

$m_e$ is the electron mass,

$c$ is the speed of light,

$I$ is the mean excitation potential of the medium,

$\delta(\beta\gamma)$ is a density correction term.

From the Eq. (89), it is evident that different particles, for e.g., proton, $D_2$, $^3$He, $^4$He will have different amounts of energy loss. In a time projection chamber like our RTPC, simultaneous measurement of momentum and particle's energy loss per unit length in the medium will help provide particle identification. The procedure of converting energy deposited by any particle as it travels in the drift region of the RTPC will be explained in sec 5.4. The $dE/dx$ vs $p/q$ for the 2 GeV helium target run is shown below:

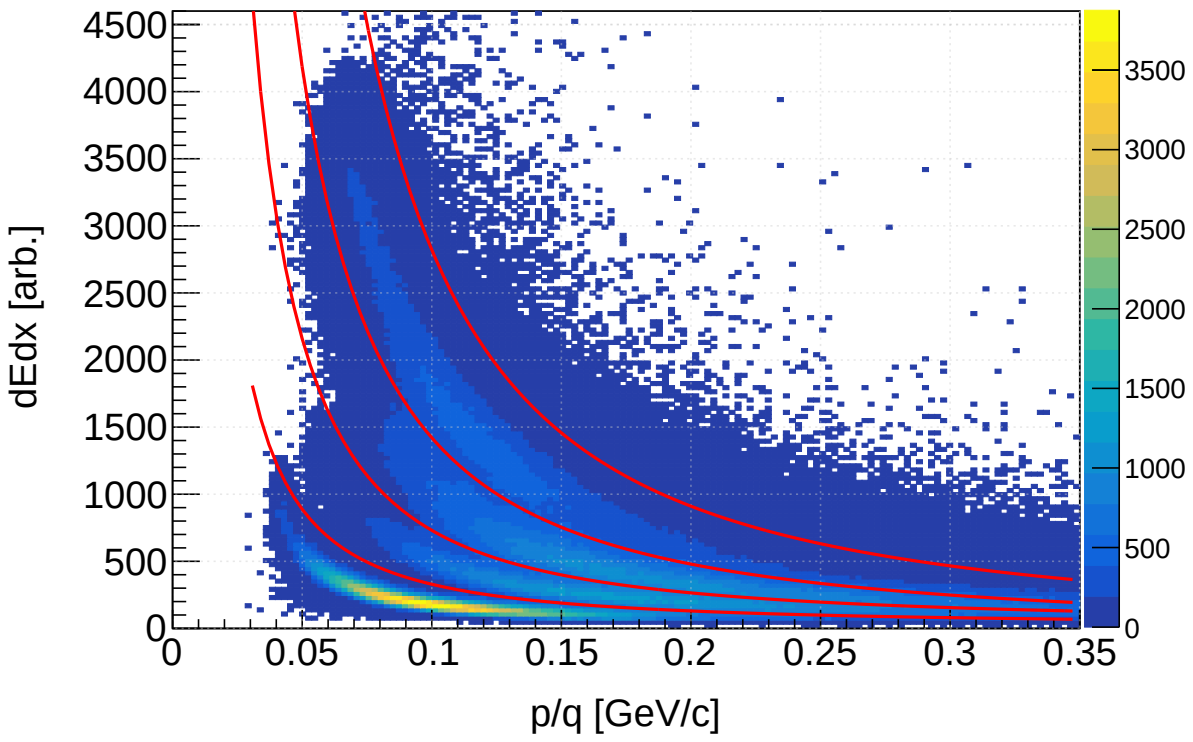



FIG. 59. $dE/dx$ vs $p/q$ plot for a helium run - It clearly shows different bands of proton,D,$^3$H/$^3$He and $^4$He.

So, the main goal is to draw curves between these bands so that they can be separated and selected during our analysis of the BONuS12 data. We want to reduce cross-band contamination while optimizing the statistics. First of all, we made an ad-hoc selection of each band by drawing an arbitrary curve that selects each band separately. Then, we made a profile plot of both the proton band and deuteron band in the $dE/dx$ vs $p/q$ plot.

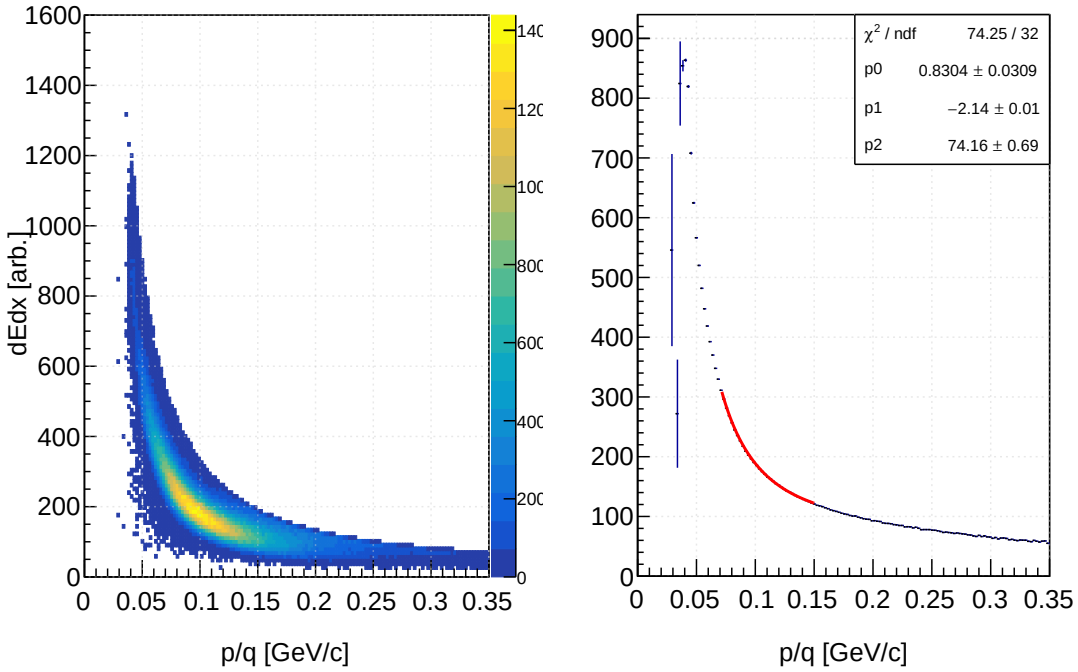



FIG. 60. Proton band in $dE/dx$ vs $p/q$ spectrum and its profile on right.

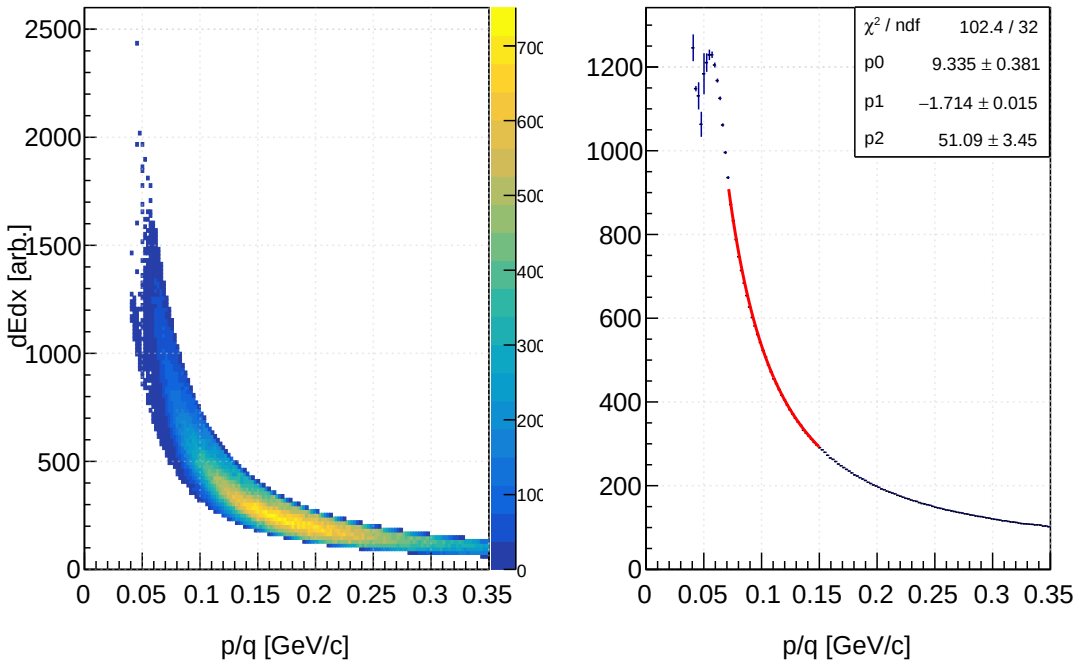



FIG. 61. Deuteron band in $dE/dx$ vs $p/q$ spectrum and its profile on right.

After drawing profiles of each band, we fitted those profiles. As shown in the plots above, a 3-parameter fit was used to fit the curve. The fitting curve was of the form, $dE/dx$ = a.$p^{\mathrm{b}}$ + c, where $p$ is momentum reconstructed in GeV/$c$ and a, b and c are parameters to be fitted.

We fitted this $dE/dx$ vs. $p$ profile for both the proton and deuteron bands. After fitting those profiles, we made another curve that was the average of the two profiles. It was done by taking the average of $dE/dx$ entries for each momentum bin in two profile curves of proton and deuteron,
$dE/dx_{\mathrm{i}} = (dE/dx_{\mathrm{i}}[\mathrm{p}] + dE/dx_{\mathrm{i}}[\mathrm{D}])/2$,
where p is for the proton profile, D is for the deuteron profile, and the indices are the bin centers of the momenta bin of the profile curve in either the proton or deuteron curve. After taking the average of the profile, the curve was again fitted with the same 3-parameter function. And the fit function is, $(dE/dx$ - 60.15$)p^{1.773} = 5.081$.
To check how well this looks, we plotted this expression vs $p$.

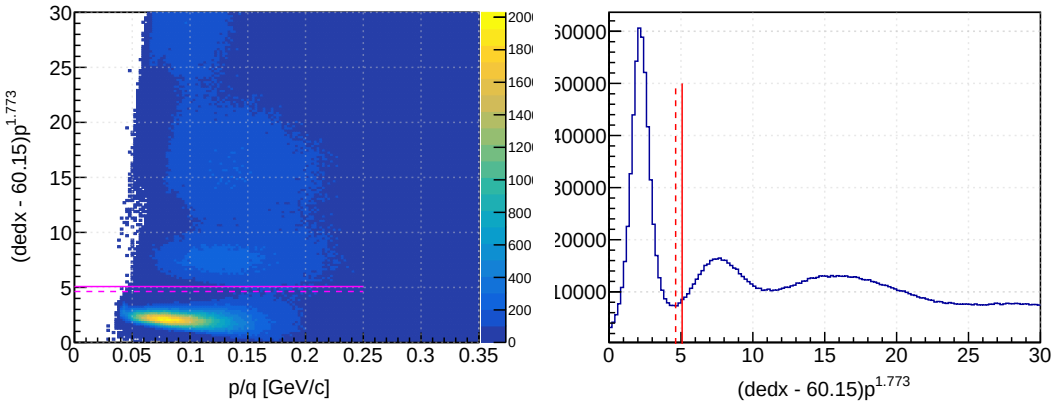



FIG. 62. On left 2D histogram of $(dE/dx$ - 60.15$)p^{1.773}$ vs $p$ and right 1D distribution of $(dE/dx$ - 60.15$)p^{1.773}$.

The vertical bold line on the right represents the value of 5.081 from the fit, and the dotted vertical line represents the line corresponding to the minimum $[dE/dx$ - 60.15$]p^{1.773}$ value on the x-axis which is 4.63. Hence, our final $dE/dx$ cut for proton selection is, $(dE/dx$

- 60.15)$p^{1.773} < 4.63$.

### 5.2.3 VIP SELECTION CUTS

As has been detailed in the Section 2.8, the detection of a slow recoiling spectator proton in the backward hemisphere will significantly reduce the ambiguities introduced by the nuclear model dependence in extracting the neutron's structure functions from the DIS reaction $D(e, e'p_s)X$. Therefore, we imply the detection of slow-recoiling backward-moving spectator protons in the RTPC in coincidence with the scattered detected electron in the forward CLAS12 detector. Fig. 63 shows the momentum distribution of the reconstructed protons in the RTPC. We apply a momentum cut of 70-100 MeV/$c$ on the spectator recoiling proton to select nDIS events where the initial neutron inside the deuteron is close to a quasi-free neutron. Moreover, Fig. 64 shows the $\cos\theta_{pq}$ distribution, where $\theta_{pq}$ is the angle between the spectator proton and virtual photon exchanged. A cut of $\cos\theta_{pq} <$ - 0.3 is implemented to select the backward recoiling protons, which have minimum final-state interaction effects on the structure of the knocked initial neutron.

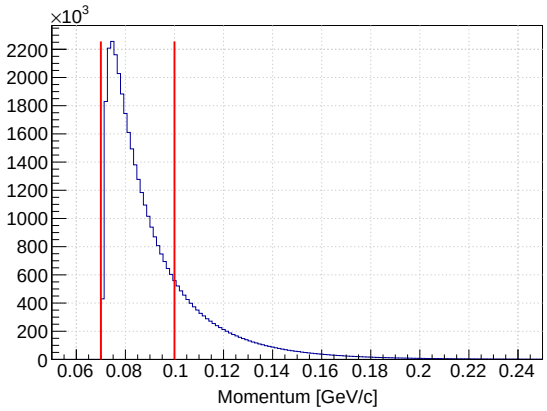



FIG. 63. The red vertical lines represent the VIP proton's momentum selection cuts.

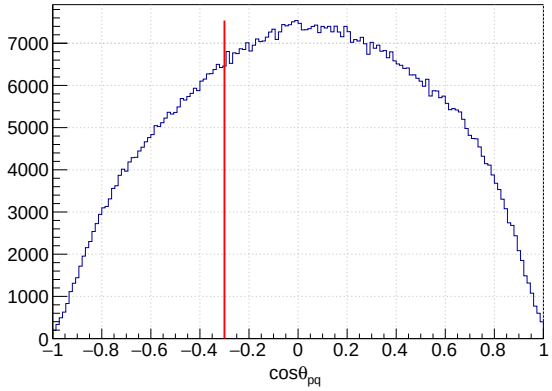



FIG. 64. The distribution of the cosine of the angle between spectator proton and virtual photon exchanged. The Red vertical line represents the boundary line for the applied cut.

### 5.2.4 DIS SELECTION CUTS

At high $Q^2$, the nucleon has a higher probability of breaking up, creating a complicated hadronic system. The invariant mass, $W > 1.8$ GeV/$c^2$ and $Q^2 > 1$ GeV$^2/c^2$ are typical values for the DIS regime. In fact, a virtual photon with $Q^2 > 1$ GeV$^2/c^2$ has a wavelength ($\lambda \propto 1/\sqrt{Q^2}$) smaller than the typical size of the nucleon (≈1 fm). Therefore, the virtual photon sees the partons of the nucleon. To ensure that the collected events are from the DIS regime, the condition $W > 1.8$ GeV/$c^2$ is applied to avoid the region where the nucleon is excited to prominent resonant baryon states. For inclusive events, $W$ is used, while $W^*$ is used for the tagged events, where $W^*$ is the corrected neutron invariant mass using the kinematics of the detected spectator proton in the RTPC. Figs. 65 and 66 present $W$ and $W^*$ distributions for the inclusive and tagged events, respectively.

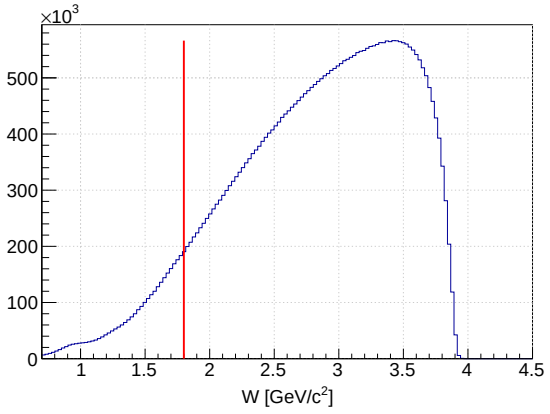



FIG. 65. The invariant mass ($W$) distribution for the inclusive event ($D(e,e')X$) with the red vertical line representing the applied cut.

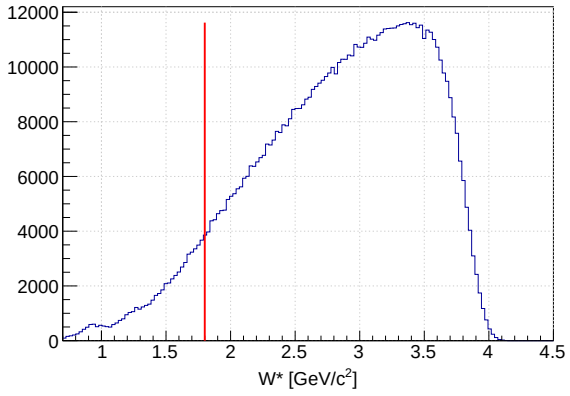



FIG. 66. The invariant mass ($W^*$) distribution for the tagged events $D(e,e'p_s)X$ with the red vertical line representing the applied cut.

## 5.3 CORRECTIONS

### 5.3.1 MOMENTUM CORRECTIONS

After the spectator particle leaves the deuterium nucleus it travels a certain distance through target gas, the target wall, buffer gas and a ground and cathode foil before it enters the drift region, and looses energy. Hence, we have to develop an empirical momentum correction to account for these energy losses for the momentum calibration of the RTPC detector. For this purpose we chose simulated data to extract the momentum corrections and the correction parameters are,

$$a = a_0 + a_1 \sin\theta_p + a_2 \sin^2\theta_p, \tag{90}$$

$$b = \frac{b_0}{\sin\theta_p} + b_1 + b_2 \sin\theta_p, \tag{91}$$

$$c = c_0 + c_1 \sin\theta_p + c_2 \sin^2\theta_p, \tag{92}$$

$$\text{temp} = \left(\frac{1}{\frac{1}{p_rec} + c}\right)^a, \tag{93}$$

$$p_{\text{corr}} = (\text{temp} + b)^{\frac{1}{a}}. \tag{94}$$

with

$$\begin{aligned} a_0, a_1, a_2 &= 2.18953, 1.41982, -0.0887623. \\ b_0, b_1, b_2 &= 0.000663994, -0.000847083, 0.000268637. \\ c_0, c_1, c_2 &= -0.237476, 0.81319, -0.304124. \end{aligned} \tag{95}$$

where, $p_{rec}$ and $\theta_p$ are the particle's reconstructed momentum (in GeV/$c$), and polar angle (in radians), resulting from the helix fit, and $p_{corr}$ is the reconstructed corrected momentum of the particle at the vertex.

#### Validation of momentum corrections for real data

To validate the momentum corrections that were extracted using simulated data, we chose radiative elastic scattering of electrons from protons using 2.1 GeV beam energy data on an $H_2$ target.

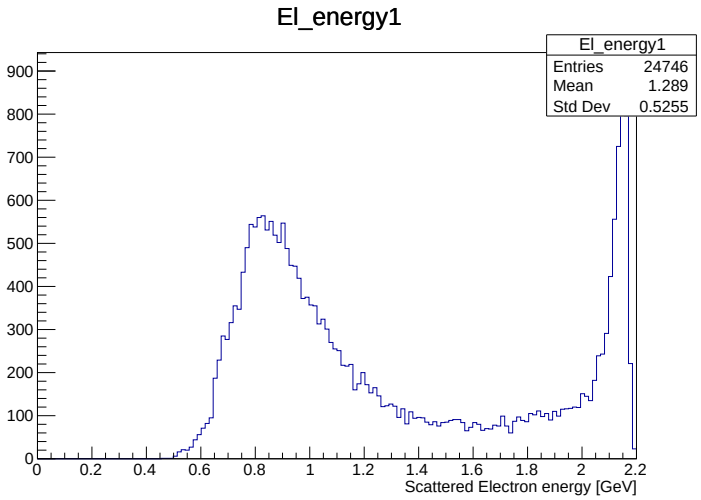



FIG. 67. $E'$ distributions for electrons scattering off $H_2$ target.

In the above 1D distribution of $E'$, we clearly see two peaks. One is around 2.1 GeV, which is due to elastic scattering, and another is around 0.9 GeV and a falloff. This is a clear signature of the electron beam Bremsstrahlung, where the electron loses energy in the form of radiation. Since the cross-section is dominated by pre-scattering radiation, most of these electrons would be radiating before scattering. Hence, we selected these events. Along with this cut on the $E'$ distribution we applied several other cuts on vertex coincidence, timing coincidence, azimuthal angle coincidence for elastic scattering ($\Delta\phi = 180°$), polar angle cut in expected region, and other tracks quality cuts.

Then we corrected the beam energy of electrons, assuming that they scattered elastically off protons in $H_2$ nuclei. After correcting the beam energy, we calculated the expected momentum and polar angle of protons in the RTPC based on the initial kinematics of the electrons and the target hydrogen gas. We compared this expected momentum with the momentum reconstructed by the helix fitter, incorporating the necessary momentum corrections. The result is shown in Figure 68. We reached the following conclusions:

- For the VIP protons (70 - 100 MeV/$c$), the $\sigma$ of $\Delta p/p$ is around 8%, which is well within our acceptable range.
- For the first two bins in momentum, the mean of $\Delta p/p$ tends to show around a 10% deviation. This is due to statistical limitations in the number of events within that

momentum range, combined with the RTPC detector's momentum threshold of around 70 MeV/$c$, causing this observed deviation.

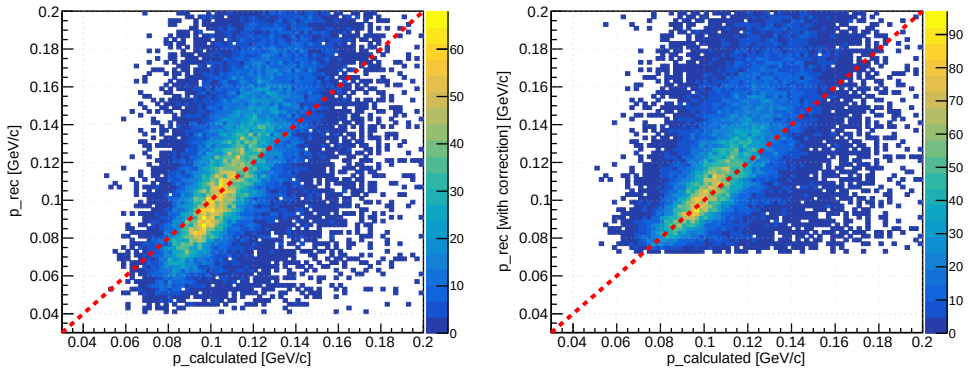



FIG. 68. 2D plot showing a correlation between reconstructed momentum in RTPC (before and after applying momentum correction) and calculated momentum using the kinematics of scattered electron in forward detector.

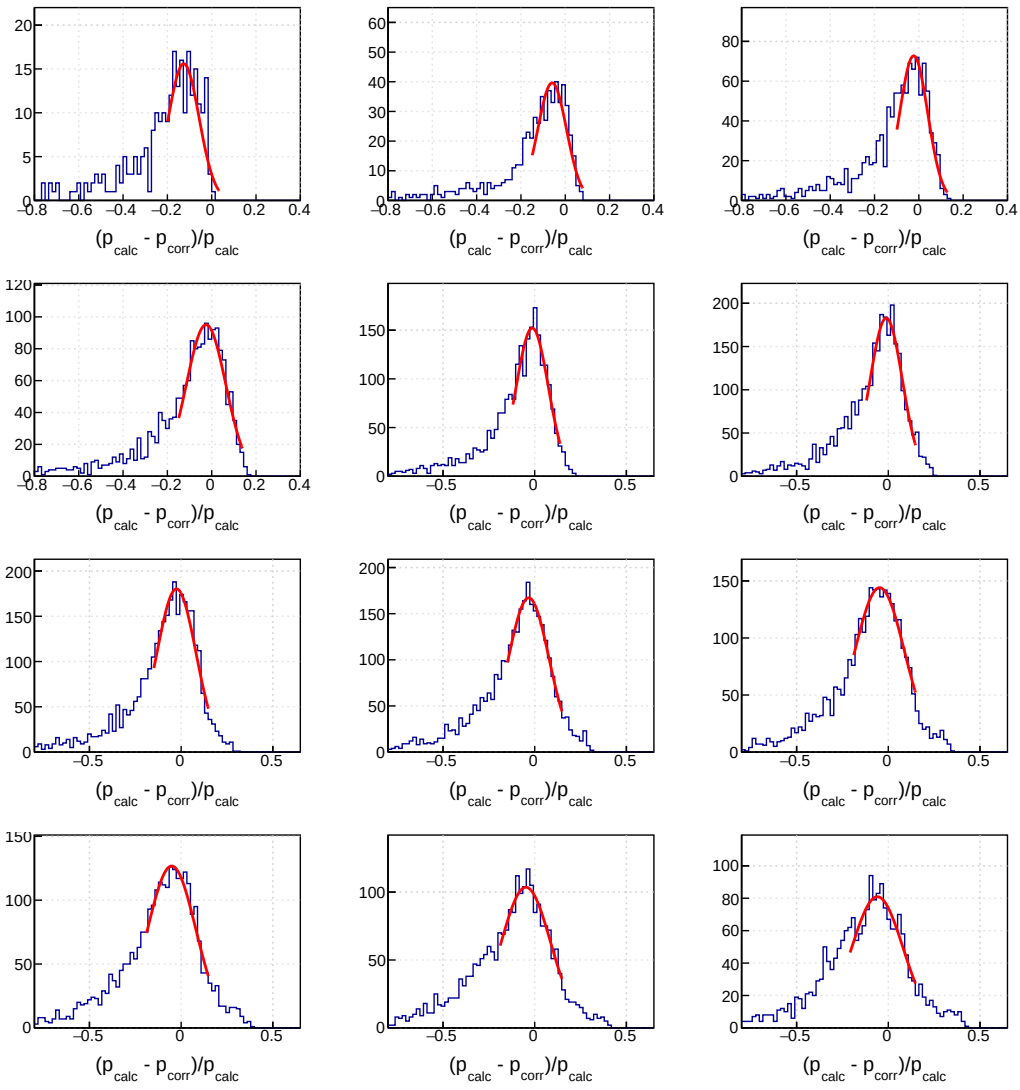



FIG. 69. Demonstration of the effect of momentum correction in different bins of calculated momentum, calculated using elastic kinematics.

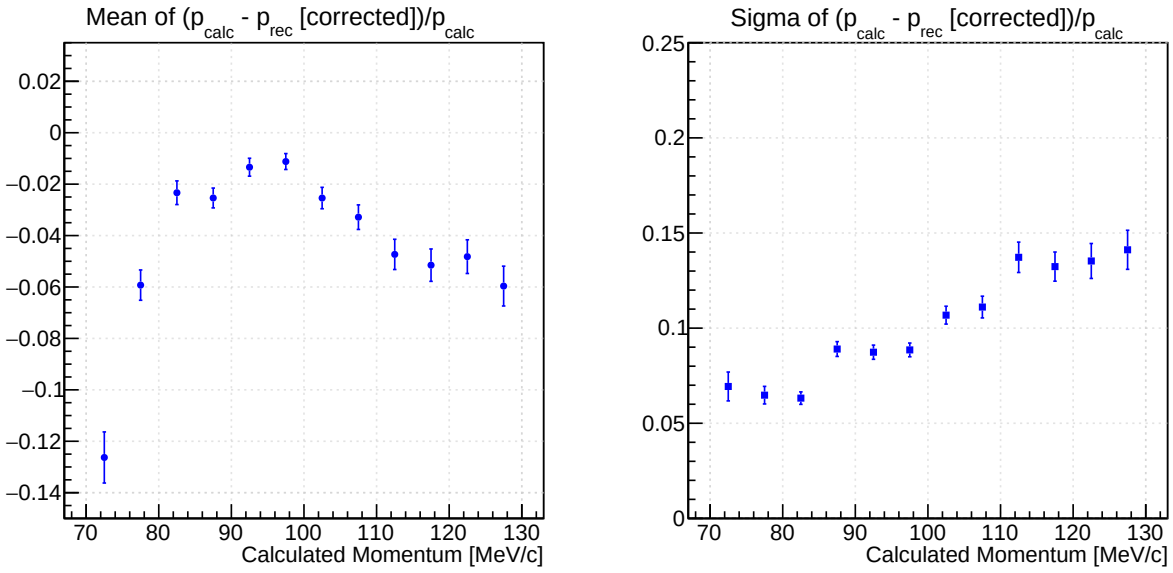



FIG. 70. The parameters of the Gaussian fit of the resolution of momentum reconstructed (after correction) as a function of calculated momentum.

## 5.4 RTPC GAIN CALLIBRATION

Whenever a spectator particle traverses the drift region of the RTPC, it loses a certain amount of energy in collisions with the drift gas. This information can be used to provide PID calibration for different species of spectator particles, as different particles deposit different amounts of energy based on their mass, charge, and momentum.
Experimentally, this energy loss can be computed as,

$$\langle \frac{dE}{dx} \rangle = \frac{\sum_i \frac{ADC_i}{G_i}}{vtl}.$$

where $\mathrm{ADC_i}$ is the energy deposited in each pad in ADC units, $G_i$ is the gain of each pad, $vtl$ is visible track length, and the summation runs over all pads that got hit in a track.

The RTPC had a total of 17,280 pads. The gain for each pad is a function of all hardware components that convert the electrical signal recorded in a pad into ADC values recorded in the acquisition system for the same pad and the local amplification of the 3 GEM foils. As a first step towards gain calibration, we took pad occupancy information for all pads in our detector, collecting sufficient statistics. For each pad, we divide the ADC signal recorded by

the total number of hits recorded on that pad. This number was our first order gain for each pad. So, we used this gain value to calculate $dE/dx$.
Mathematically,

$$G_{\rm i} = \frac{\sum_i ADC_i}{frq.\,of\,hits},$$

where summation in $i$ runs over all pads in the detector. However, during the analysis, we noticed that our detector has shown a dependence along the longitudinal position of the target ($v_z$). We noticed that the calculated $dE/dx$ was dependent on the position in the detector. Fig. 71 demonstrates this effect.

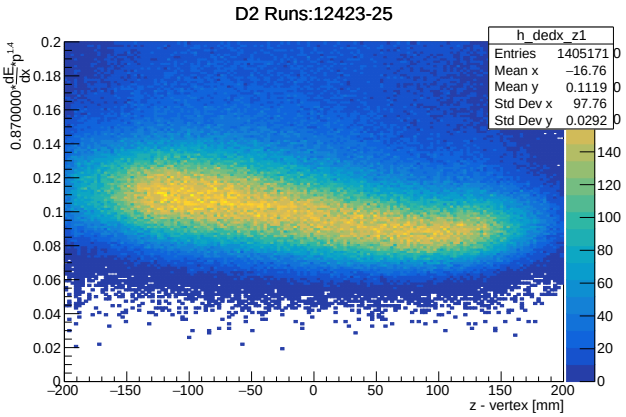



FIG. 71. $dE/dx$ (multiplied by the momentum ($p^{1.4}$) vs. $v_z$ for $D_2$ target runs ,12423-12425, where we can see the dependence of $dE/dx$ along the longitudinal position of the target ($vz$).

Additionally, this $dE/dx$ dependence on $v_z$ changes with time (run number). Fig. 72 shows the extraction procedures for this $dE/dx$ dependence on $v_z$, using the collected protons in the momentum band between 70 MeV/$c$ and 120 MeV/$c$. From the linear fit to the one-dimensional profile, we get the following variables:

- $dE/dx$ amplitude (the constant of the fit): which is related to the magnetic field and the HV setting on the GEM foils inside the RTPC. Fig. 73 shows the three observed plateaus in the amplitude of $dE/dx$ and the extracted factors to flatten $dE/dx$ as a function of the run number over the Summer 2020 data-taking period.

- $dE/dx$ slope versus $v_z$: Which represent the strength of the $v_z$ dependence.
- $dE/dx$ percentage ratio: Which is defined as $(targetLength * slope)/amplitude$ and represents our observable to the relative $dE/dx$ change over the length of the target.

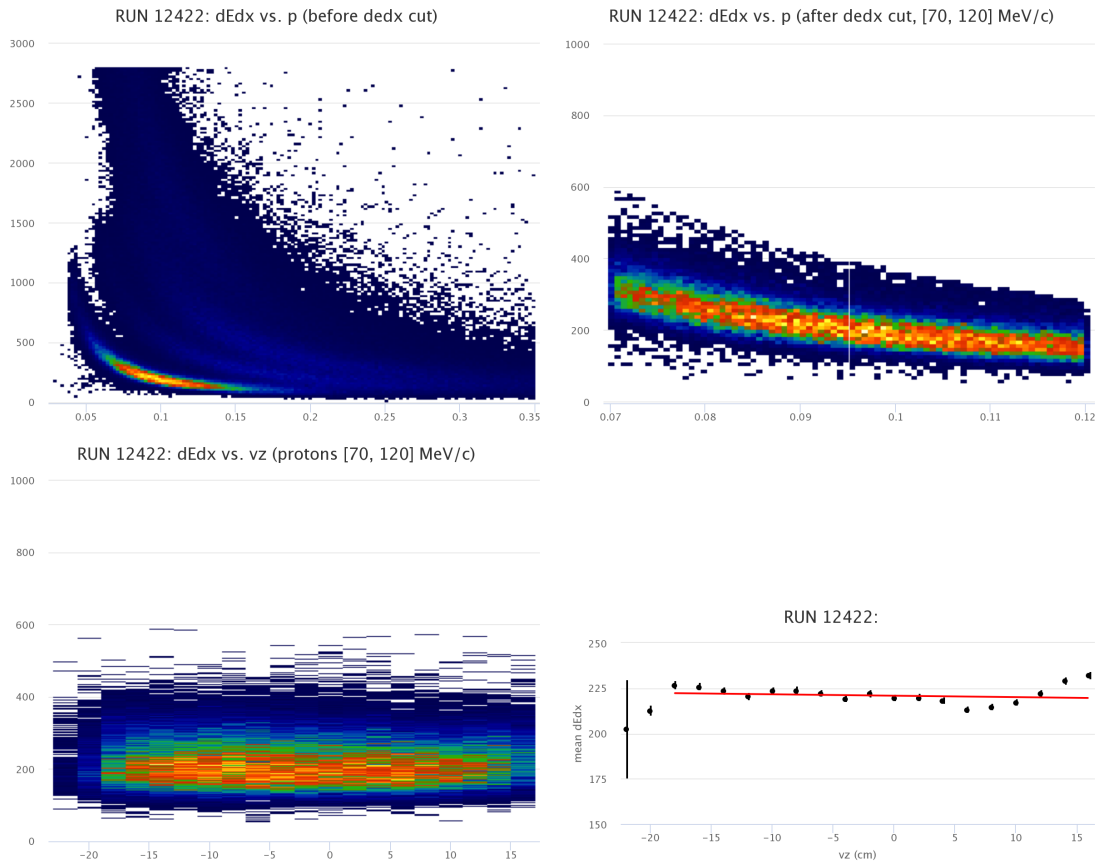



FIG. 72. Top-left: $dE/dx$ versus $p/q$ for the collected tracks in one run (12422), top-right: $dE/dx$ versus $p/q$ for selected protons between 70 MeV/$c$ to 120 MeV/$c$ momentum, bottom-left: The $dE/dx$ (multiplied by the momentum $(p^{1.4})$) profile versus $v_z$ for the selected protons and bottom-right: the 1D profile of $dE/dx$ versus $v_z$ with the red line represents a first-degree polynomial fit.

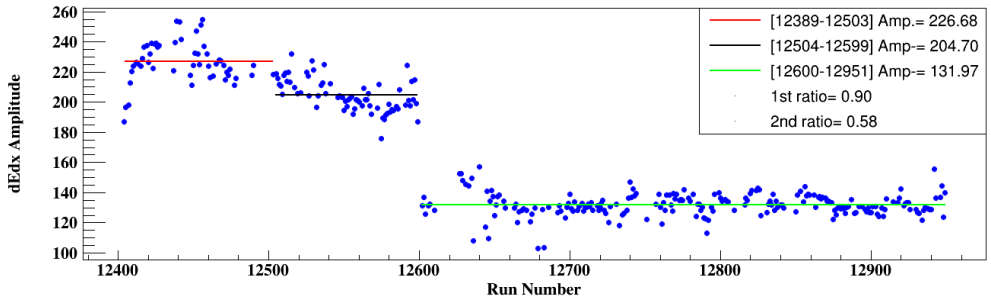



FIG. 73. $dE/dx$ amplitude as a function of the run number for the Summer 2020 period. Data before run 12600 were collected using 385 V across the GEM foils in the RTPC, while it was 375V after 12600.

So, to correct this effect, we did a second iteration of gain calibration. This time, we looked on-time reconstructed proton tracks that span over the full drift region, and we applied some quality cuts on those tracks to be sure they are good tracks in the RTPC. For each of those tracks, we had an RTPC::Hits bank (event database) associated with each event. This bank contains all the hit information on that track. For example, it contains information on all the pads that got fired in that track, the timing of the hit for each pad, ADC for each hit in the track, and several other useful information. So, using this information for each track, we removed the top 5% of maximum ADCs in a track. Then, for the remaining 95% of ADCs, we normalize them by the median of remaining hits. Hence, for each hit within a track:

$$\mathrm{ADC}_{\mathrm{new}}[i] = \frac{\mathrm{ADC}_{\mathrm{meas}}[i]}{\text{Median ADC of the track}}, \tag{96}$$

$$\tag{97}$$

$$G_i = \frac{\sum_i \mathrm{ADC}_{\mathrm{new}}[i]}{\text{Frequency of hits}[i]}. \tag{98}$$

Fig. 74 shows the effect of gain calibration on our data, which removed the $dE/dx$ dependence along the longitudinal position of the target.

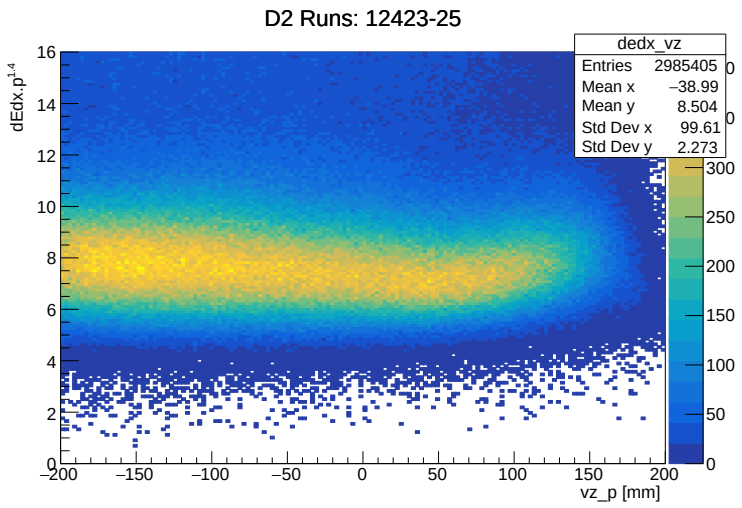



FIG. 74. $dE/dx$ (multiplied by the momentum ($p^{1.4}$) vs vz for $D_2$ Runs 12423-12425 after gain calibration. Here, we can see that introducing new gains has removed the dependence of $dE/dx$ along the longitudinal position of the target ($v_z$).

## 5.5 DEPENDENCE OF RTPC VARIABLES ALONG THE $\phi$ DIRECTION

We observed an unexpected behavior in the RTPC performance. The RTPC has a full $2\pi$ angular acceptance, meaning that all measured variables should be symmetric in azimuthal angle($\phi$). However, contrary to expectations, the reconstructed variables in the RTPC exhibit a clear $\phi$ dependence, resulting in a non-uniform $\phi$ distribution, as shown in Fig. 75. Since simulations do not reproduce this effect, we can rule out the possibility that it arises from geometry or underlying physics.

Although cylindrical TPCs have been successfully used in various experiments worldwide, this type of behavior has not been documented before, making it an interesting topic for further study. Fig. 76 presents the $\phi$ distribution in the RTPC with each additional good track selection cuts applied in every plot. The persistence of the structure, independent of these cuts, indicates that the effect is not a result of selection biases.

To investigate whether this $\phi$ dependence is correlated with the trigger electron's $\phi$, which is detected in the forward detector, we analyzed the $\phi$ distribution in the RTPC separately for each electron sector. The CLAS12 forward detector consists of six sectors, and we plotted six

different $\phi$ distributions, each corresponding to the sector that detected the trigger electron. Fig. 77 demonstrates that the $\phi$ distribution in the RTPC remains unchanged across all sectors, ruling out any correlation with the trigger electron and confirming that this effect originates purely from the RTPC detector itself.

Although our final experimental result will be integrated over $\phi$, meaning this effect will not immediately impact the measurement, it is still an important issue to understand. This anomaly could be an inherent feature of GEM-based TPCs. Given that our RTPC is a triple-GEM detector, where each GEM layer consists of 16 independent sectors, one possible hypothesis is that the boundaries between GEM sectors are less efficient than the central regions. However, despite extensive studies, a definitive explanation for this effect has yet to be found.

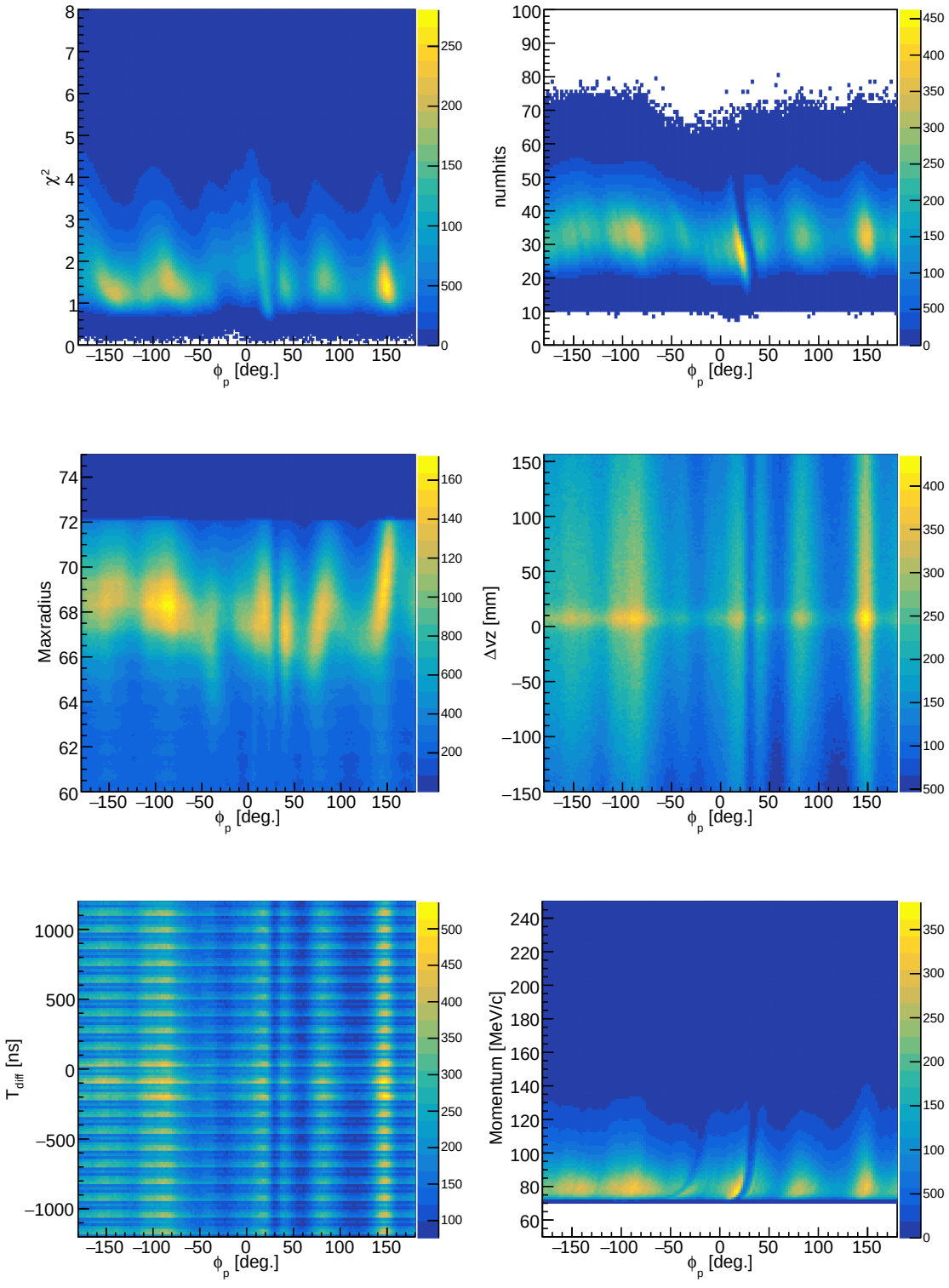



FIG. 75. The 2D histogram of the reconstructed variable vs. $\phi_{RTPC}$.

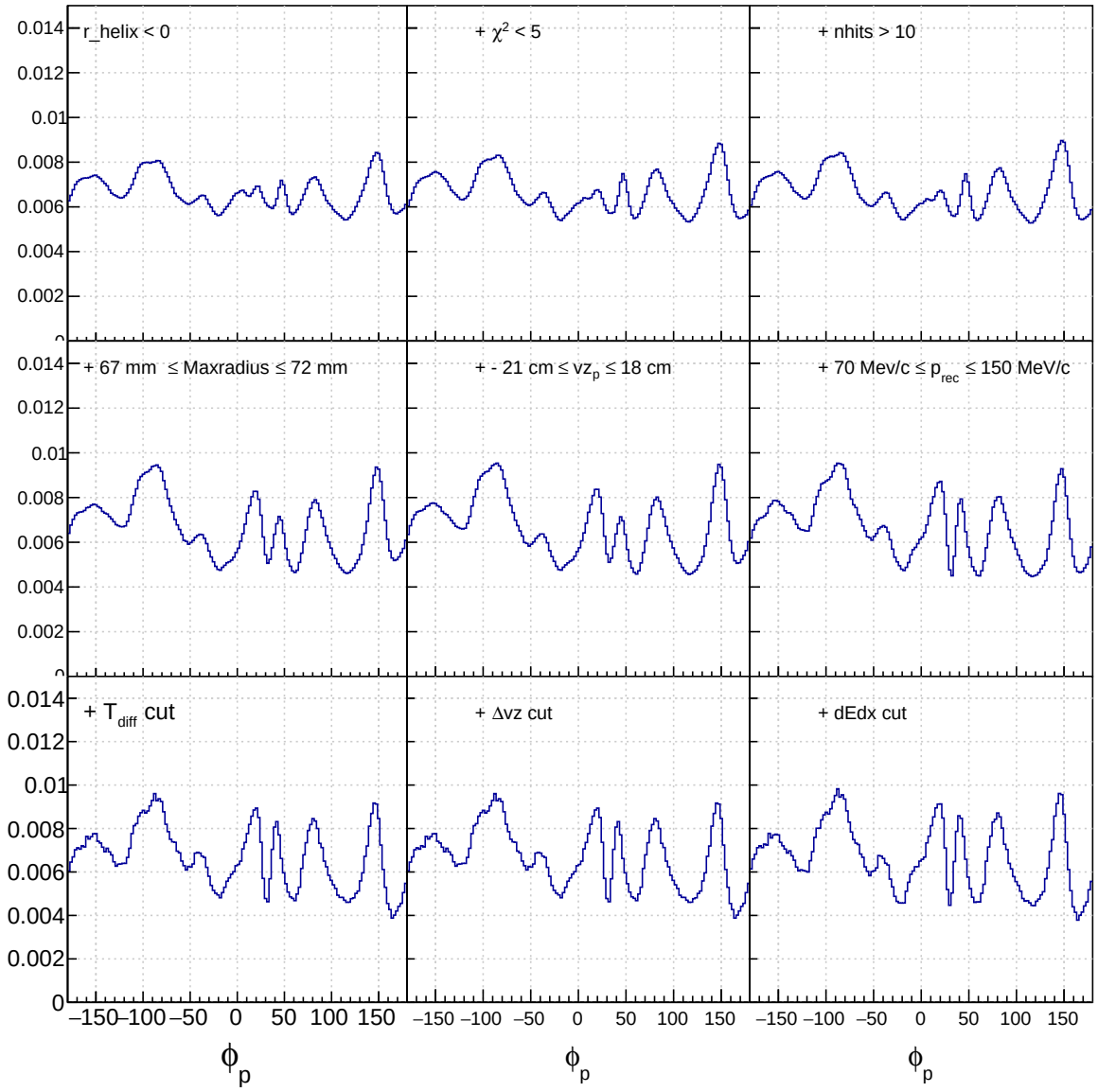



FIG. 76. The $\phi$ distribution in RTPC with each track selection cuts applied in order.

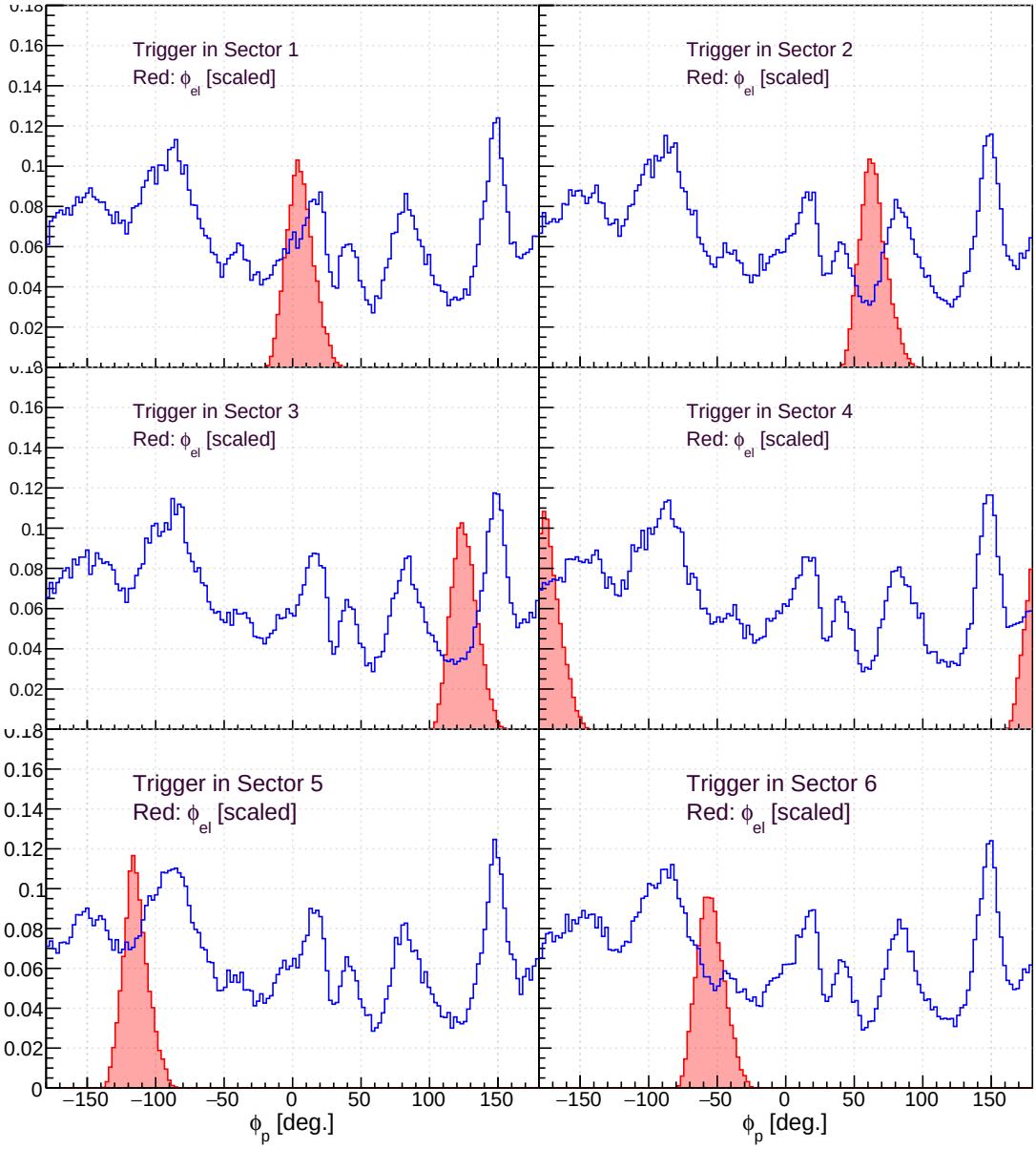



FIG. 77. The $\phi$ distribution in RTPC for different sectors of trigger electrons. The red histogram is the $\phi$ for the trigger electron.

# CHAPTER 6

# BACKGROUND SUBTRACTION

The BONuS12 experiment uses the RTPC detector to tag recoiling spectator protons in electron scattering events on the bound neutron inside deuterium. By identifying the spectator proton in coincidence with the scattered electron, one can reconstruct the initial kinematics of the struck neutron and suppress scattering events from tightly bound protons in deuterium. Additionally, by applying cuts on the spectator proton ("VIP") in momentum and angle, one can emphasize events where the struck neutron is not far from being off-shell and minimize the effects of Final State Interactions (FSI). However, the final-state electrons and protons do not always come from neutron deep-inelastic scattering events. Therefore, one needs to estimate the fraction of background events and subtract them from the data. In the following sections, we introduce the main background sources contaminating the sample of identified electron-neutron DIS events.

## 6.1 BONUS12 PROTON CONTAMINATION

### 6.1.1 ACCIDENTAL COINCIDENCES

Inside the RTPC, the ionization electrons drift relatively slowly, about 3.2 $\mu$s from the cathode to the first GEM foil. Therefore, hits belonging to a proton track have to be selected in coincidence with the trigger electron over a wide range in time ($\pm$1800 ns) around the coincidence peak. The resolution of the RTPC timing parameter $t_{diff}$ is around 200 ns; hence, a significant number of accidental coincidences can be found within any given timing cut.

After the previously presented selection cuts on the electrons (in CLAS12) and on the protons (in the RTPC, but without timing and vertex cuts), one expects to find roughly one proton per trigger electron in the final state. Most of these are random coincidences, but we estimate roughly 0.022 expected true coincidence protons per electron. The latter number is important as it leads to an overestimation of the random background. Since our cuts on vertex coincidence ($\Delta v_z$) are two sigmas wide and the timing coincidence cuts ($t_{diff}$) are 1.9

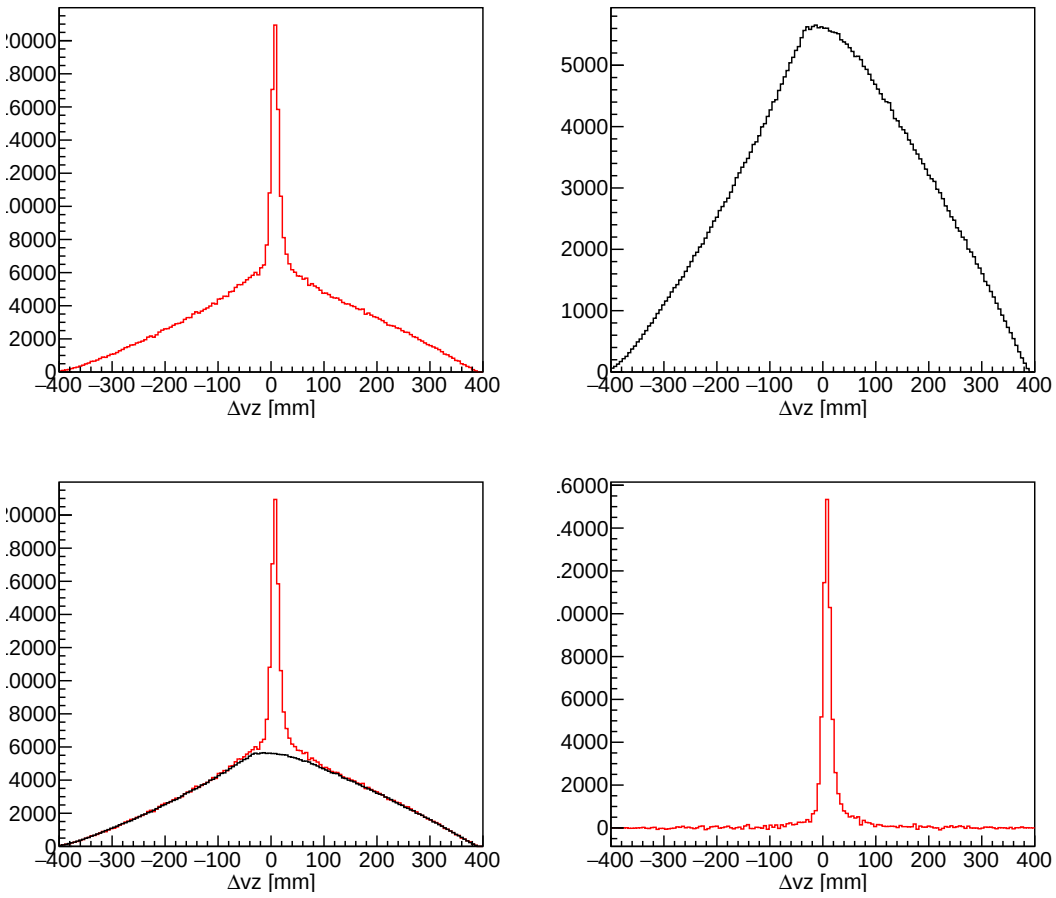



FIG. 78. Background subtraction demonstrated in the $\Delta v_z$ spectra. Top left: The experimentally observed $\Delta v_z$ distribution after all the PID cuts on the final state electron and recoil proton in addition to $t_{diff}$ and VIP cuts. Top right: same for the case where protons are matched with up to 14 electrons from different events, and the cuts are evaluated using those electron kinematics. One sees the expected triangular shape of the background. Bottom left: Overlay of both spectra after normalization. Bottom right: True coincident spectra after subtracting the mixed-event background.

sigmas wide, we have to develop a method to account for accidental background events that occur inside our cuts.

The "real" coincidence peak could be estimated by extrapolating the "side wings" outside

the peak into the coincidence window. In practice, this has several drawbacks:

- The wings in the $\Delta v_z$ distribution are not flat but rather triangular. This is a consequence of the fact that for a given $\Delta v_z$, contributions can only come from some fraction of the length of the target for both the electron and the proton unless $\Delta v_z = 0$. This makes an extrapolation more uncertain and lowers the background statistics.

- The $t_{diff}$ distribution is also not completely flat. Furthermore, due to how tracks are reconstructed in the RTPC, the range over which $t_{diff}$ is a reliable estimate for the actual time difference between the tracks is limited.

Instead, we match electrons and protons from different events to get better statistics on the accidental background and subtract those mixed events within our coincidence cuts from the signal. The following procedure is applied: We read in 15 consecutive events with trigger electrons that pass all electron cuts. On average, each of these events will have one random proton and 0.022 coincident protons. In any case, we will loop over all the protons. Then, we define the following:

- **Diagonal Events**: We define a diagonal event where the proton and the electron come from the same event, including true coincidence and accidental background that we need to estimate and subtract. Then, we apply all coincidence cuts to select only coincident protons and protons that form the remaining accidental background within our cuts.

- **Off-Diagonal Events**: For those same 15 electron-proton events, we loop over the remaining electrons from all other events for every proton. We call them off-diagonal events. We apply the same *vertex* coincidence cuts for each off-diagonal pair, but for the *timing* coincidence cut, we offset the time window by 1100 ns (see below). All surviving off-diagonal events are tabulated, and further cuts (VIP, $W^*$, etc.) are applied the same way as for the coincident diagonal events. Since one would expect 14 times as many random events (due to the combination with 14 different electrons) than in the diagonal sample, the resulting number is divided by 14 and subtracted (within each kinematic bin) from the corresponding diagonal event number, as shown in Fig. 78.

We apply this procedure for both $D_2$ as well as $^4$He data; the latter are used to obtain the coincident background from scattering events due to other target components than deuterons, as explained in the next section.

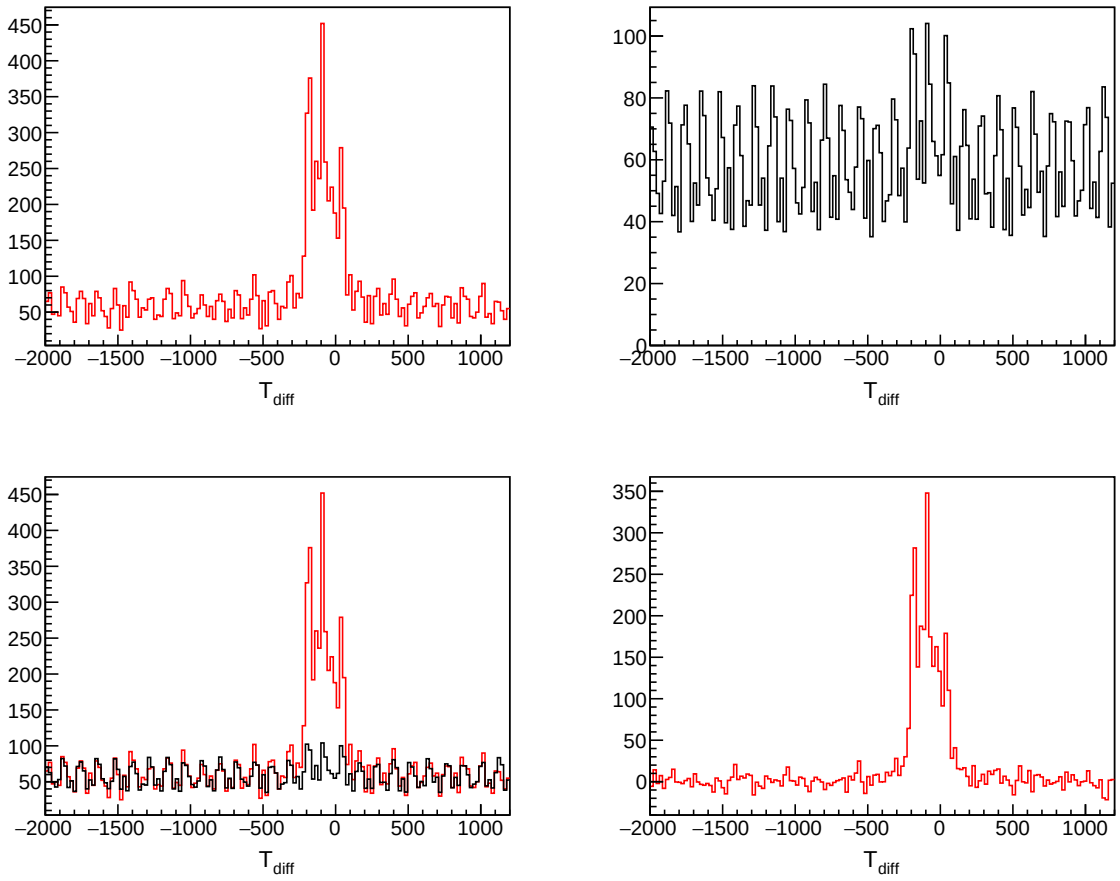



FIG. 79. Background subtraction demonstrated for $t_{diff}$ spectra. Top left: Number of events within $\Delta v_z$ and VIP cuts from the same event. Top right: same for the case where protons are matched with up to 14 electrons from different events, and the cuts are evaluated using those electron kinematics. Bottom left: Overlay of both spectra. Bottom right: "True coincident" spectra after subtracting the mixed-event background.

Figure 79 shows that our matching algorithm works as anticipated in the "wings" of the $t_{diff}$ distribution. However, despite being matched with electrons from different events, the Fig. 79 top right panel shows that there is an excess (about 10%) of mixed events in the central $t_{diff}$ region within our $t_{diff}$ cuts. This excess comes from those protons that belong to true coincidences: While they would end up at arbitrary $\Delta v_z$ when paired with an electron

from a different event, their $t_{diff}$ is determined by the trigger electron from the event in which they occur, independent of the electron they are matched with, and hence within the coincidence cut in $t_{diff}$. Given that we estimate about 0.33 such "true coincidence" electrons in our 15-event sample (0.022 coincidences per trigger), we expect an enhancement of about 10% within our $t_{diff}$ cut from these electrons since they are, by definition, concentrated in a region in $t_{diff}$ corresponding to about 20% of the full range. To solve this problem, we shifted the $t_{diff}$ window for mixed event selection by 1100 ns towards the left (more negative), where the $t_{diff}$ distribution is flat and coincides with that for the diagonal events (bottom left panel of Fig. 79).

In summary, Fig. 80 shows how we sample the accidental backgrounds under the signal peaks. In the figure, we have taken the example of 5 electron proton pairs for easy demonstration. So, we pair every electron we detect with the proton in the same event. We might have more than one proton per event, and this method also works for such cases. However, the figure explains it in the case of one proton per event. After every fifth, we make all possible combinatorial pairs among them, which are 25 electron proton pairs. Mathematically possible pairs are,

$$\text{pair}_{ij} = e_i p_j, \quad \text{where } i, j = 1, 2, 3, 4, 5.$$

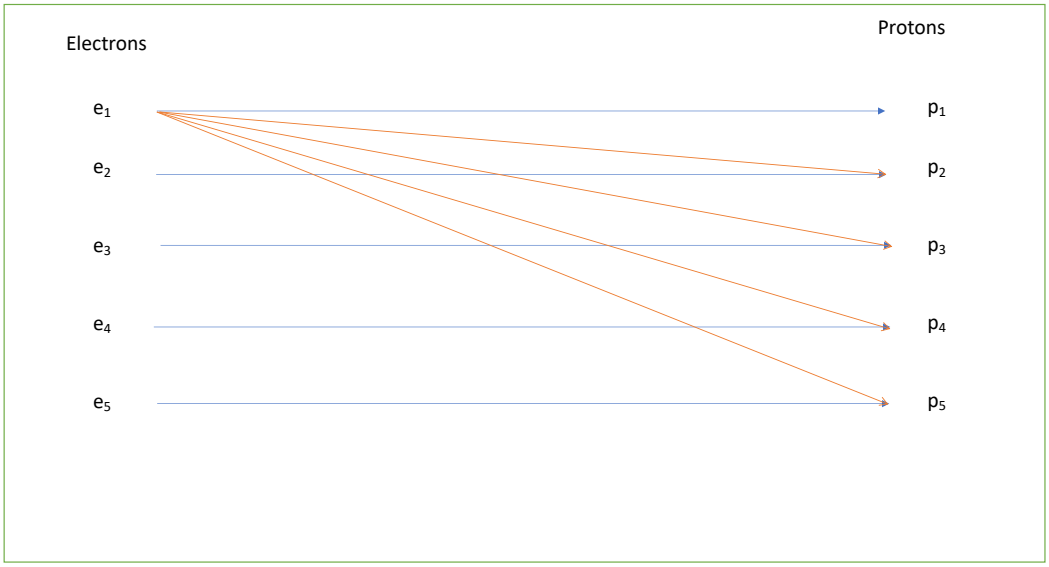



FIG. 80. Representative diagram showing event mixing of electrons and protons to sample the combinatorial backgrounds using five electron-proton pairs.

In matrix representation,

$$\text{pair}_{ij} = \begin{bmatrix} e_1p_1 & e_1p_2 & e_1p_3 & e_1p_4 & e_1p_5 \\ e_2p_1 & e_2p_2 & e_2p_3 & e_2p_4 & e_2p_5 \\ e_3p_1 & e_3p_2 & e_3p_3 & e_3p_4 & e_3p_5 \\ e_4p_1 & e_4p_2 & e_4p_3 & e_4p_4 & e_4p_5 \\ e_5p_1 & e_5p_2 & e_5p_3 & e_5p_4 & e_5p_5 \end{bmatrix},$$

The matrix shows all electron-proton pairs. $(e_ip_i)$ correspond to "signal pairs" (same event) where the off-diagonal elements $(e_ip_j,\ i \neq j)$ correspond to "combinatorial background" pairs (mixed events). This can be represented in a single equation as,

$$\text{pair}_{ij} = \delta_{ij} \cdot \text{signal} + (1 - \delta_{ij}) \cdot \text{background}.$$

Also, in our example, for every one electron pair, we have four times as many background pairs. So, to calculate background counts, we have to scale the background by four.

### 6.1.2 NON DEUTERON TARGET CONTAMINATION

The BONuS12 experiment used $D_2$ as the production target gas. We also collected data on $^4$He, $H_2$, and empty targets for calibration and background estimation purposes. Once we changed our target gas from $D_2$ to other targets and took data on those targets, we would flush the target multiple times with $D_2$ gas before re-starting production data taking. Fig. 83 represents a purge cycle of the target gas, where we fill and empty the target three times during each purge. This procedure would be repeated every day or even multiple times a day whenever beam delivery was delayed. This procedure ensured that we minimized remnants of the previously filled gas, and so minimized the contamination. In reality, the target straw was not perfectly leak-tight, and an outward leak rate of about 5% per hour was observed at 68 psig pressure. Additionally, $^4$He from the detector could diffuse into the target straw from the surrounding buffer gas region. One way to measure the presence of non-deuterium gas inside the target was to monitor the different particles reconstructed in the RTPC detector.

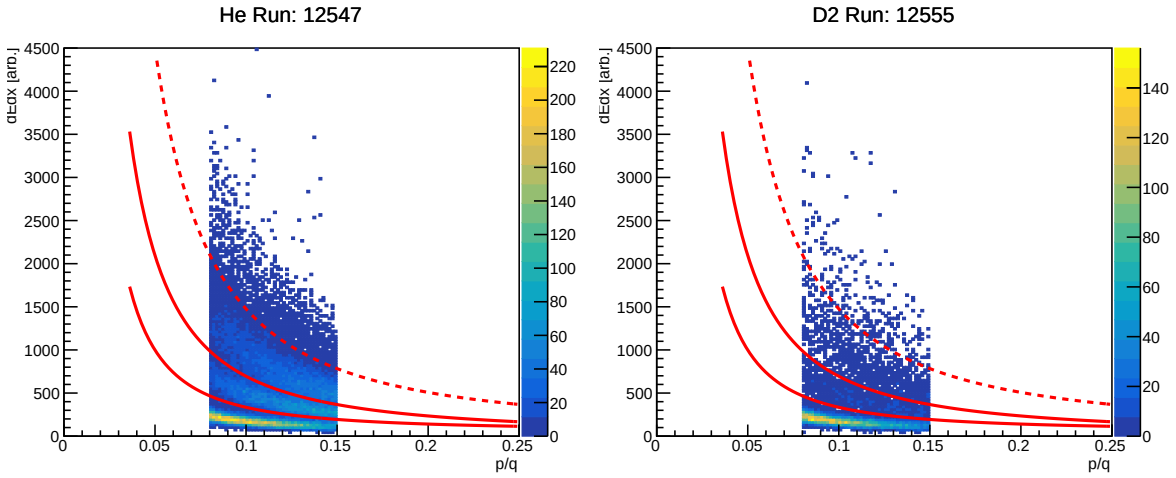



FIG. 81. From left to right: $dEdx$ versus $p/q$ for $^4$He and $D_2$ runs. The red curves present the PID selection cut for protons, $D_2$,$^3$H/$^3$He, and $^4$He from bottom to top.

Fig. 81 shows the distribution of $dEdx$ as a function of the measured momentum per charge ($p/q$) for the reconstructed tracks inside the RTPC from $^4$He and $D_2$ target gases.

For a pure deuterium target, one would not expect any recoil particles other than protons and deuterons. However, the data show heavier particles above the deuterium band, indicating the presence of non-deuterium target gas inside the target, mostly $^4$He. The following procedures have been carried out in order to estimate the $^4$He contamination in the deuterium data:

- Count the number of particles in the $^3$H/$^3$He band in $dE/dx$ vs $p$ spectrum in each run.
- Normalize the counts to the total measured beam charge in each run.
- Cross normalize those beam-charge normalised $^3$H/$^3$He counts to same quantity in the $^4$He runs. This number would then represent $^4$He contamination in each run.

Fig. 82 present the $^4$He contamination in each $D_2$ target that are used to subtract the nDIS events that have originated from $^4$He gas rather than $D_2$. For subtracting the $^4$He contributions from the final dataset, we define a contamination factor $\gamma$ as the ratio of total number of events in $^3$He/$^3$H band for $D_2$ runs to $^4$He runs, that we had taken during whole RGF experimental run period. Mathematically,

$$\gamma = N_{\mathrm{He_3}}^{D_2\,\mathrm{runs}} / N_{\mathrm{He_3}}^{^4\mathrm{He}\,\mathrm{runs}} \tag{99}$$

$$\beta = \left(N_{\mathrm{He_3}}^{D_2\,\mathrm{runs}} / \mathrm{FC}_{D_2}\right) / \left(N_{\mathrm{He_3}}^{^4\mathrm{He}\,\mathrm{runs}} / \mathrm{FC}_{^4\mathrm{He}}\right) \tag{100}$$

The true number of $D(e, e'p_s)X$ events in each bin after subtracting He contamination will be,

$$N_{true}^{tag} = [N_{sig}^{D} - N_{bkg}^{D}] - \gamma[N_{sig}^{He} - N_{bkg}^{He}]. \tag{101}$$

Where,

$N_{sig}^{D}$ is the number of the on-time tagged events from $D_2$ data,
$N_{bkg}^{D}$ is the number of the tagged events from the accidental background in the $D_2$ data,
$N_{sig}^{He}$ is the number of the on-time tagged events from the $^4$He data,
$N_{bkg}^{He}$ is the number of the tagged events from the accidental background in the $^4$He data,
$\gamma$ is the $^4$He contamination factor in the $D_2$ data.
$\beta$ is the total He contamination in our target (expressed as percentage).

This will be further defined in 9.2.1

| Target | Run Period | Count of $^3$He | FC Beamcharge (nc) |
|---|---|---|---|
| $D_2$ runs | Before 12600 | 93,544 | $4.86623 \times 10^7$ |
| | After 12600 | 247,263 | $1.52082 \times 10^8$ |
| $^4$He runs | Before 12600 | 18,082 | $1.05874 \times 10^6$ |
| | After 12600 | 41,626 | $4.08072 \times 10^6$ |
| **Contamination Ratios** | | **Before 12600** | **After 12600** |
| $\gamma$ | | 5.17 | 5.94 |
| $\beta$ | | 0.11 | 0.16 |

TABLE 4. Summary table for contamination calculation with ratios

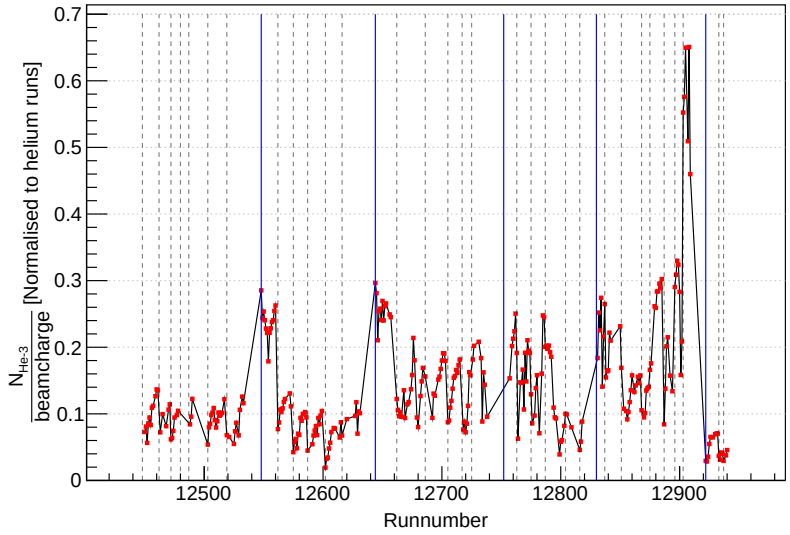



FIG. 82. Normalised $^3$H/$^3$He counts in $D_2$ runs cross normalized to same quantity in $^4$He runs as a function of run number. The number on the y-axis represents contamination. The vertical black line in the x-axis represents the run during which we flushed our target and filled it with fresh $D_2$ target gas. Hence, contamination will be minimal at those runs. Similarly, the vertical blue line represents the runs after the helium target runs.

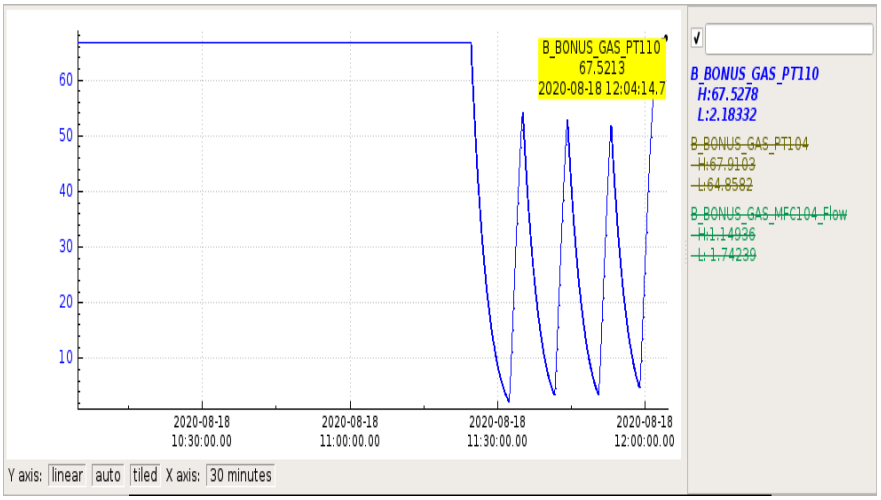



FIG. 83. The purging cycle of gas inside the target straw.

## 6.2 ELECTRON CONTAMINATION

The CLAS12 data acquisition system is set up to accept the first electron as the trigger particle if it satisfies all trigger criteria, regardless of whether it is the scattered electron or a secondary electron from radiation or the annihilation of other particles. In this analysis, we need to find this portion of secondary electrons that contaminate our sample of nDIS scattering events and correct it. The primary source of the contaminating secondary electrons are,

- Dalitz decay: $\pi^0 \rightarrow e^+e^-\gamma$ [1.2 % branching ratio]
- $\pi^0 \rightarrow \gamma\ \gamma \rightarrow e^+e^-\ \gamma$

These pair symmetric electrons are experimentally indistinguishable from electrons originating from actual nDIS physics events. To estimate the contribution from pair symmetric electrons, we monitor positrons, as there should be for every electron in the $e^+e^-$ pair a corresponding positron. After measuring the number of positrons, one can take the ratio of positrons to electrons in each kinematic bin of interest, giving the contamination ratio. However, the CLAS12 detector contains the solenoid and the torus magnets, which are used for tracking of charged particles. Since positrons and electrons are oppositely charged, they behave differently in these fields. For an inbending torus configuration, the electrons bend towards the beamline, whereas positrons move away from the beamline. Hence, the acceptance will not be the same for positrons and electrons for the same torus configuration, as can be seen in Fig. 84. One way to account for this is by comparing inbending electrons with outbending positrons and counting electrons and positrons with the same cuts that we apply for identification. These procedures have to be carried out for both analysis channels, i.e., inclusive and tagged analysis, as presented in the following subsections.

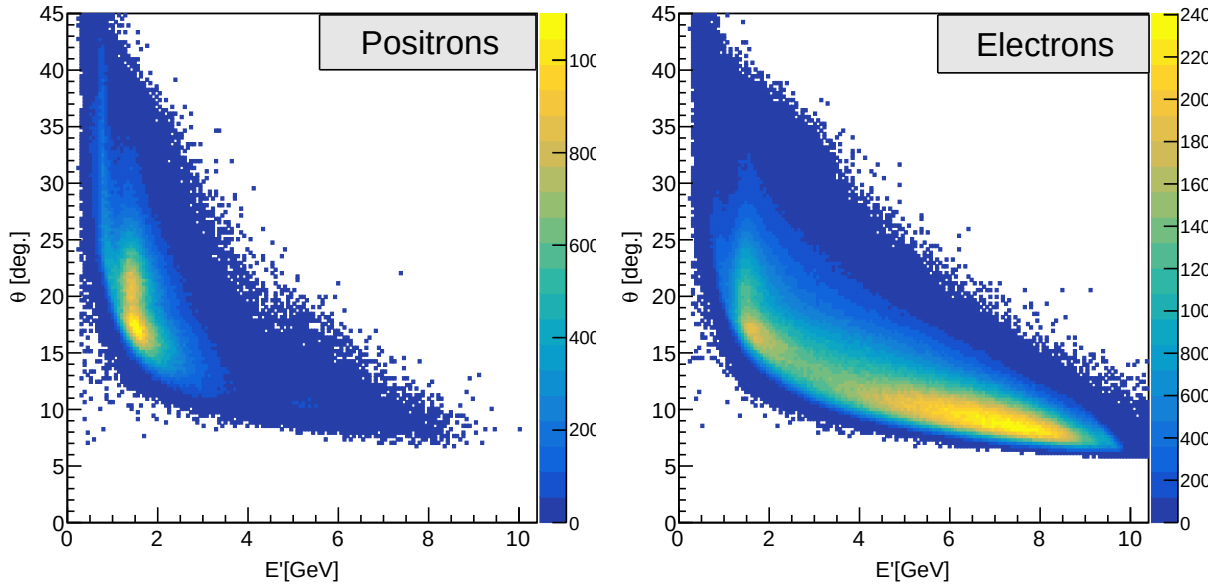



FIG. 84. 2D plot of $\theta$ versus energy for inbending positrons (left) and inbending electrons (right).

### 6.2.1 ANALYSIS OF INCLUSIVE DATA

As mentioned previously, in order to access a similar kinematic phase-space for electrons and positrons, one needs to compare the inbending electrons to the outbending positrons. BONuS12 did not collect outbending production data during the Summer 2020 period, but it did during the Spring 2020 run, which is used and presented here for this particular study. The same cuts are applied to the selection of electrons or positrons. The number of photo-electrons cut was changed to 7 for both electrons and positrons to assure the removal of the $\pi^+$ ($\pi^-$) contamination in the positron (electron) sample. After all analysis cuts, the collected positrons and electrons were binned in 8 $\theta$ bins. In each $\theta$ bin, the ratio of the number of positrons to those of electrons is calculated and plotted as a function of the energy as shown in Fig. 85. Then, the ratios were fitted with the function:

$$R = e^{p_0(\theta)E' + p_1(\theta)}, \tag{102}$$

where $p_0(\theta)$ and $p_1(\theta)$ are fit parameters that depend on polar angle $\theta$, and $E'$ is the energy.

The resulting fitting parameters $a_0(\theta)$ and $a_1(\theta)$ are plotted as a function of the polar angle $\theta$ and fitted to obtain a global fit for the fitting parameters as shown in Fig. 86. With global fitting for the parameters $a_0(\theta)$ and $a_1(\theta)$, one can calculate the $e^+/e^-$ ratio for each bin in $x$ and $Q^2$. Additionally, for each bin in $x$ and $Q^2$, the average value of the measured $E$‘ and $\theta$ distributions are used for this contamination subtraction.

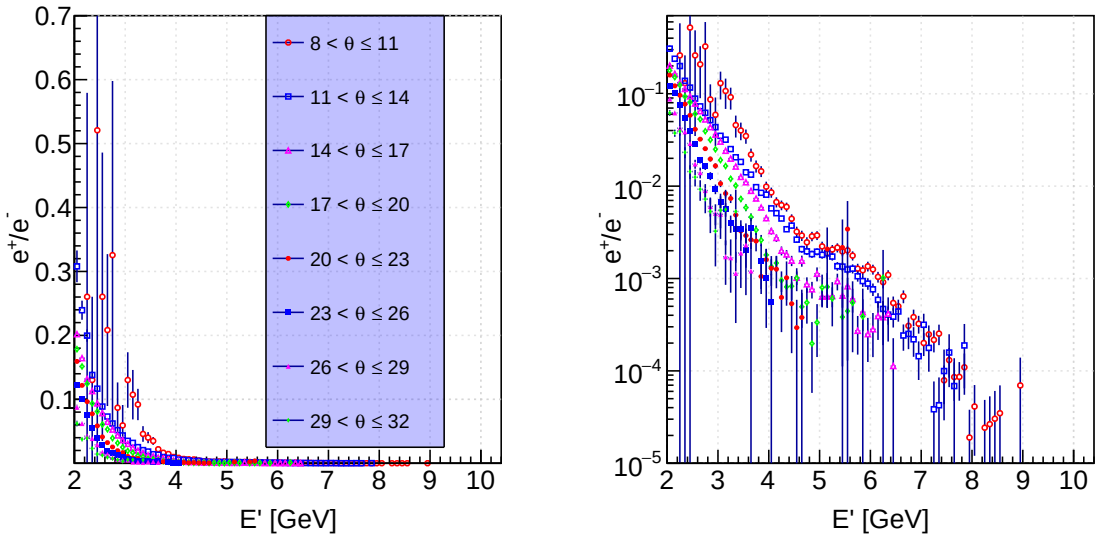



FIG. 85. [The $e^+/e^-$ ratio in the different $\theta$ bins.]The $e^+/e^-$ ratio in the different $\theta$ bins (left) as a function of their energy (left) and same ratio in logarithmic scale (right).

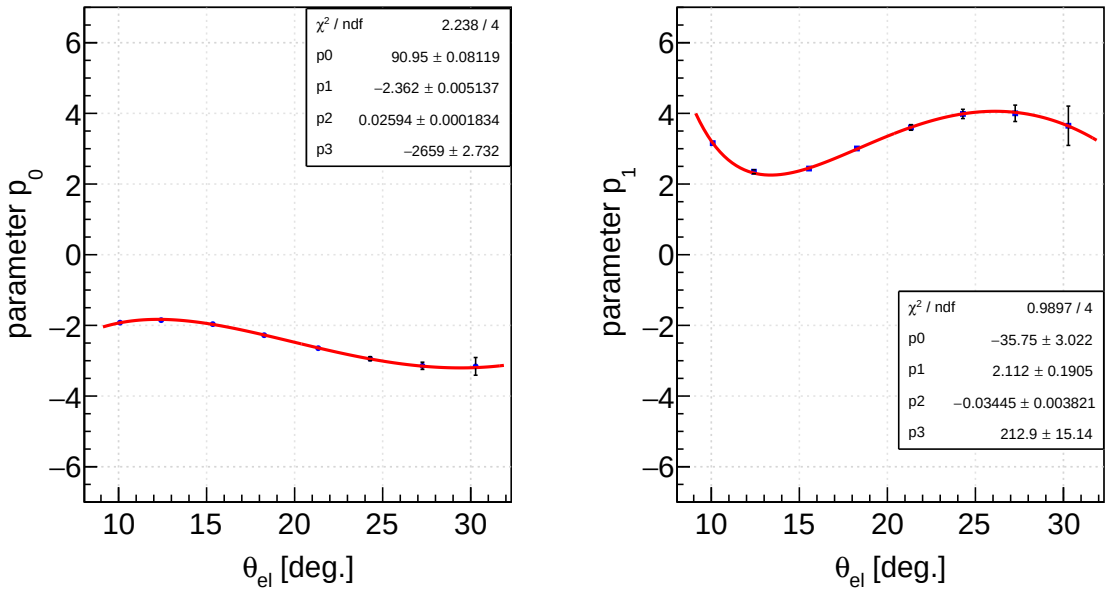



FIG. 86. Global fit for the $e^+/e^-$ ratio fitting parameters $p_0(\theta)$ (left) and $p_1(\theta)$ (right).

The global fit functions are:

- **for parameter $p_0$:**

$$p_0 = a_0 + a_1\theta + a_2\theta^2 + \frac{a_3}{\theta + 27}, \tag{103}$$

where:

$$a_0 = 90.94 \tag{104}$$
$$a_1 = -2.361 \tag{105}$$
$$a_2 = 0.02594 \tag{106}$$
$$a_3 = -2659. \tag{107}$$
$$\tag{108}$$

- **for parameter $p_1$:**

$$p_1 = b_0 + b_1\theta + b_2\theta^2 + \frac{b_3}{\theta}, \tag{109}$$

where:

$$b_0 = -35.75$$
$$b_1 = 2.112$$
$$b_2 = -0.03445$$
$$b_3 = 212.9.$$

### 6.2.2 ANALYSIS OF TAGGED DATA

To extract the $e^+/e^-$ ratio for events where a recoil proton is tagged and detected in the BONuS12 RTPC, we used only the summer 2020 data since the RTPC was not functioning correctly during the spring 2020 run. This ratio was then corrected using the inclusive $e^+/e^-$ ratios previously obtained from the spring and summer 2020 datasets. Fig. 87 shows the positron-to-electron ratios across different theta bins for the summer 2020 data. During the Summer 2020 run, data were collected with only one torus field configuration. As discussed earlier, due to the opposite bending of electrons and positrons in the magnetic field, proper estimation of electron contamination in a specific kinematic bin ideally requires positron data from the opposite field configuration. Since this was unavailable for Summer 2020, we applied correction factors derived from the Spring 2020 inclusive data.

$$\begin{aligned} \mathrm{R_{tag}}(\text{summer}, \text{true}) = &\left[\frac{\mathrm{R_{tag}}(\text{summer}, \text{inbend } e^+, \text{inbend } e^-)}{\mathrm{R_{inc}}(\text{summer}, \text{inbend } e^+, \text{inbend } e^-)}\right] \\ &\cdot \mathrm{R_{inc}}(\text{spring}, \text{inbend } e^-, \text{outbend } e^+) \end{aligned} \tag{110}$$

- $\mathrm{R_{tag}}(\text{summer}, \text{true})$: The true ratio of positrons to electrons in any kinematic bin.
- $\mathrm{R_{tag}}(\text{summer}, \text{inbend } e^+, \text{inbend } e^-)$: The ratio of positrons to electrons in a bin for the tagged summer data for the runs where we used inbending runs to count the number of positrons and electrons.
- $\mathrm{R_{inc}}(\text{summer}, \text{inbend } e^+, \text{inbend } e^-)$: The ratio of positrons to electrons in a bin for the inclusive summer data for the runs where we used inbending runs to count the number of positrons and electrons.
- $\mathrm{R_{inc}}(\text{spring}, \text{inbend } e^-, \text{outbend } e^+)$: The ratio of positrons to electrons in a bin for the tagged summer data for the runs where we used inbending runs to count the number of electrons whereas to count number of positrons we used outbending runs.

Fig. 87 displays the positron-to-electron ratios for the tagged case, where a spectator proton is detected in the RTPC in coincidence with the scattered electron. The final corrected $e^+/e^-$ ratios for the tagged Summer 2020 data, incorporating the Spring 2020 correction factors, are presented in Fig. 89.

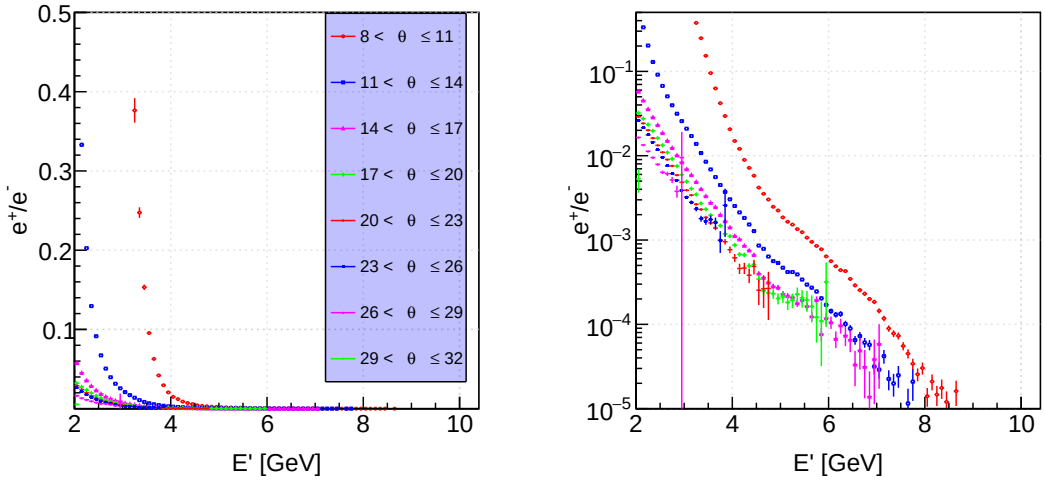



FIG. 87. The ratio of the number of $e^+/e^-$ for different $\theta$ bins (left) and the same ratio in logarithmic scale (right) from the inclusive Summer 2020 data.

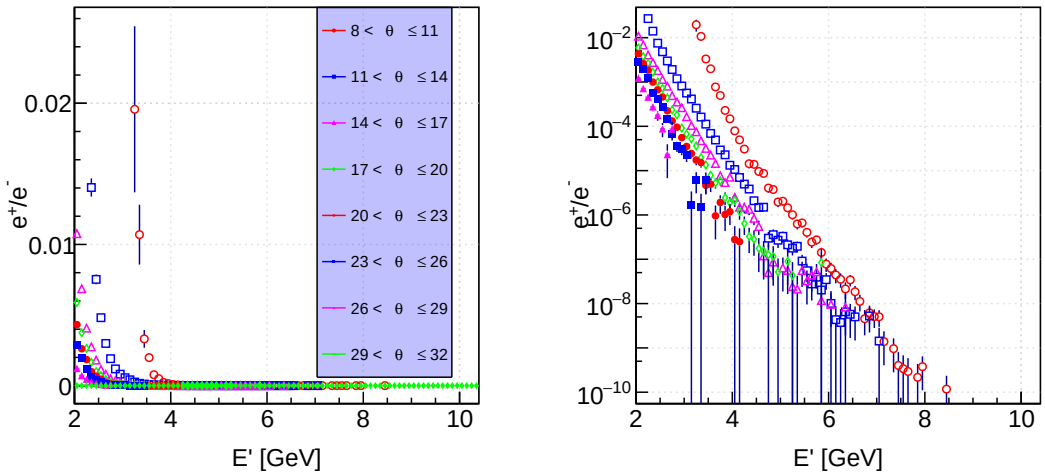



FIG. 88. The ratio of the number of $e^+/e^-$ for different $\theta$ bins (left) and same ratio in logarithmic scale (right) from the tagged Summer 2020 data.

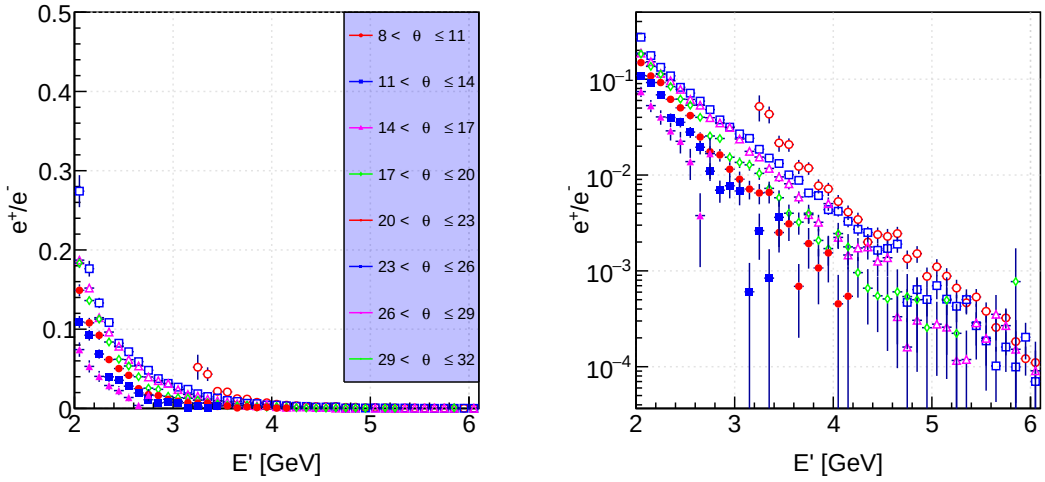



FIG. 89. The ratio of the number of $e^+$ to $e^-$ for different $\theta$ bins (left) and the same ratio in logarithmic scale (right) from the tagged Summer 2020 data after correcting for acceptance using spring parameters.

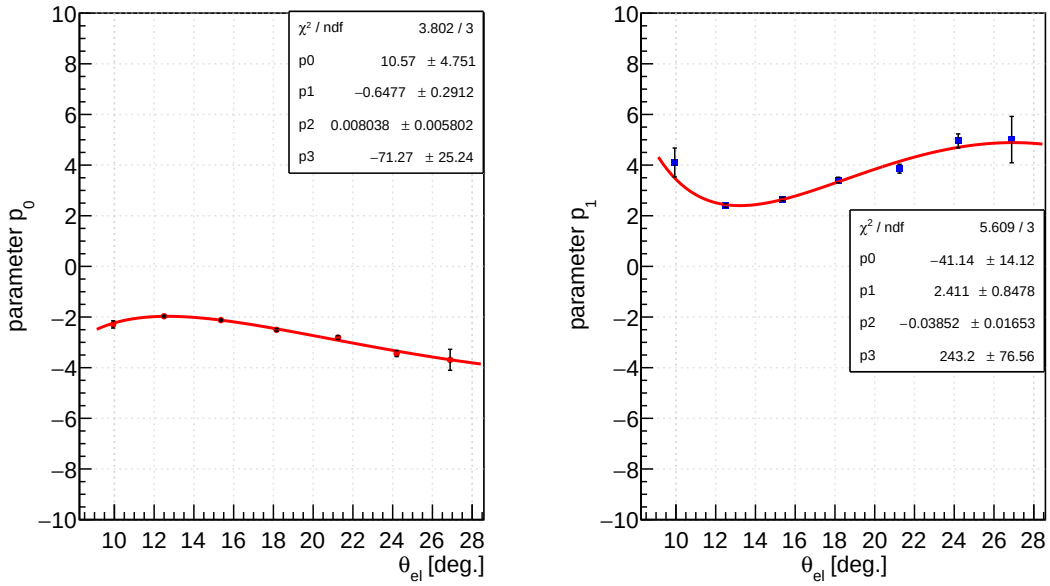



FIG. 90. Global fit for $e^+/e^-$ ratio fitting parameters, $p_0(\theta)$ (top) and $p_1(\theta)$ (bottom), for the tagged data from the Summer 2020 period.

Fig. 89 shows the final corrected $e^+/e^-$ ratio for tagged Summer 2020 data as a function of the energy in bins of the polar angle ($\theta$). These ratios are then fitted using the formula,

$$R = e^{c_0(\theta)E' + c_1(\theta)}, \tag{111}$$

where $c_0(\theta)$ and $c_1(\theta)$ are fit parameters that depend on the polar angle $\theta$, and the energy $E'$.

We then obtain a global fit for these fit parameters. The global fit functions are:

- **for parameter $\mathbf{p_0}$:**

$$c_0 = a_0 + a_1\theta + a_2\theta^2 + \frac{a_3}{\theta}, \tag{112}$$

where:

$$a_0 = 10.6, \tag{113}$$
$$a_1 = -0.6492, \tag{114}$$
$$a_2 = 0.008059, \tag{115}$$
$$a_3 = -71.48 \tag{116}$$
$$\tag{117}$$

- **for parameter $\mathbf{p_1}$:**

$$c_1 = b_0 + b_1\theta + b_2\theta^2 + \frac{b_3}{\theta}, \tag{118}$$

where:

$$b_0 = -41.29,$$
$$b_1 = 2.42,$$
$$b_2 = -0.03866,$$
$$b_3 = 244.1.$$

### 6.2.3 FINAL IMPLEMENTATION

After extracting parameters for both inclusive and tagged data, the pair symmetric background was corrected. For the corrected yield of electron counts in each bin in $x$ and $Q^2$, for either the inclusive or the tagged analysis, we obtained

$$N_{e^-measured} = N_{e^-scattered} + N_{e^+e^-pair}$$
$$N_{e^-measured} = N_{e^-scattered} + N_{e^-measured}(\frac{N_{e^+}}{N_{e^-}})$$
$$N_{e^-scattered} = N_{e^-measured}[1 - \frac{N_{e^+}}{N_{e^-}}]$$
$$N_{e^-scattered} = N_{e^-measured}[1 - R].$$

where,

$N_{e^- measured}$ is the experimentally measured total number of scattered electrons,

$N_{e^- scattered}$ is the true number of scattered electrons originating from scattering events,

$N_{e^+e^-}$ is the number of electrons stemming from pair production,

$R$ is the corresponding extracted $e^+/e^-$ ratio for the inclusive or the tagged analysis channels.

For each bin in $x$ and $Q^2$, one can then calculate the average value of $\theta$ and $E'$ for that bin. Using the value of $\theta$ and $E'$, one can then calculate the contamination ratio, $R$, using the above equation, which will result in the correction factor $(CF) = 1 - R$. For every bin, we can multiply the raw count with the correction factor to get the corrected count of scattered electrons.

# CHAPTER 7

# DATA QUALITY MONITORING AND FILE SELECTION

## 7.1 ELECTRON YIELD

Whenever an experiment starts in Experimental Hall B of Jefferson Lab, or when the DAQ system begins recording responses from individual detectors—generally referred to as the start of a run—the DAQ system assigns a unique identifier to the run, known as the run number. The analysis presented in this thesis is based on data from run 12448 to 12951. During each scattering event, the DAQ system collects responses from individual detectors. Additional details regarding this process have already been explained in chapter 5. Since an enormous amount of data is collected during each event, managing data size for each run is a crucial step in offline reconstruction. For our BONuS12 experiment, we implemented a constraint: whenever the total number of collected triggers in a run reached 10 million, we would terminate the run and start a new one. After completing each experimental run, during offline reconstruction, the data from that run was saved in multiple, smaller-sized files called Data Summary Tape (DST) files. For example, a file named **rec_clas_012736.evio.00125-00130.hipo** represents a reconstructed data file from run number 12736, containing reconstructed event data from file numbers 125 to 130 combined into a single HIPO file.

In principle, with constant beam current, target density, and pressure, the event rate has to be constant over the experimental data taking time. However, this rate changes due to changes in the experimental conditions, such as changing a trigger in a detector, a slight shift in the beam position, or a system failure somewhere. We minimize the effects of these changes on the reconstructed events by selecting good runs. To achieve this goal, we monitor the normalized electron yield to the beam charge and the ratio of the well-reconstructed protons in the RTPC to the number of detected good electrons as a function of experimental time (in other words, as a function of run number). So, for each DST file, we would count the total number of electrons in each file and monitor this quantity as a function of time. The main aim of the data quality monitoring procedure is to count the electron yield, denoted by $N_{el}$, normalized by the Faraday Cup charge (FC) or beamcharge.

The electron yield per beamcharge must be calculated for each file in each run. This ratio should be constant for different run periods with steady experimental conditions, the same DAQ conditions, the same trigger conditions, and configurations. However, changes in beam conditions, detector trips, target gas type, background conditions, trigger issues, and DAQ issues could cause outliers. So, the main aim of this study is to find such outliers and identify files with problems.

### 7.1.1 DETERMINING THE NORMALIZED ELECTRON YIELD

First, we must count the number of electrons satisfying all good electron selection criteria. We applied the following cuts for electron selection:

- Particle Identification (PID) = 11.
- -21 cm $< vz_{el} <$ 18 cm.
- Number of Photoelectrons in HTCC $> 2$.
- Sampling Fraction cut [see Chapter 4.]
- EPcal $> 0.1$ GeV.
- Only one good electron per event, and the selected electron is the first particle in an event.

We have applied the loose cuts on the electron selection for this procedure. After applying the cuts for electron selection, the total number of electrons that satisfy all of the above selection cuts was counted for all individual reconstructed files and documented in tabular form. These cuts have already been described in chapter 4.

#### Determining the Faraday-Cup Information for Each DST File

For each Data Summary Tape (DST) file, the total accumulated charge in the Faraday Cup must be calculated. This information is available in the RUN::Scaler bank, which is present in each DST file. The structure of this bank would look like the following:

```
{
    "name": "RUN::scaler",
    "group": 10000,
    "item": 14,
    "info": "Run-integrated, analyzed scalers.",
    "entries": [
        {"name":"fcupgated",  "type":"F", "info":"Beam charge, integrated from beginning of run, DAQ-gated (nano-Coulomb)"},
        {"name":"fcup",       "type":"F", "info":"Beam charge, integrated from beginning of run, ungated (nano-Coulomb)"},
        {"name":"livetime",   "type":"F", "info":"DAQ livetime during one scaler period"}
    ]
},
```


FIG. 91. Structure of RUN::Scaler bank in CLAS12 database.

We would use the fcupgated beam charge, which contains the Faraday cup measured beamcharge gated by DAQ lifetime, integrated from the beginning of each run. Similarly, another bank HEL::Scaler is also present, with the same and additional scaler information.


```
{
    "name": "HEL::scaler",
    "group": 10000,
    "item": 16,
    "info": "Helicity-gated, analyzed scalers.",
    "entries": [
        {"name":"fcupgated",  "type":"F", "info":"Beam charge from Faraday cup, DAQ-gated (nano-Coulomb)"},
        {"name":"fcup",       "type":"F", "info":"Beam charge from Faraday cup, (nano-Coulomb)"},
        {"name":"slmgated",   "type":"F", "info":"Beam charge from SLM, DAQ-gated (nano-Coulomb)"},
        {"name":"slm",        "type":"F", "info":"Beam charge from SLM (nano-Coulomb)"},
        {"name":"clockgated", "type":"F", "info":"Clock, DAQ-gated"},
        {"name":"clock",      "type":"F", "info":"Clock"},
        {"name":"helicity",   "type":"B", "info":"Helicity state, HWP-corrected"},
        {"name":"helicityRaw","type":"B", "info":"Helicity state"}
    ]
},
```


FIG. 92. Structure of HEL::Scaler bank in CLAS12 database.

The one difference to note in the fcupgated beamcharge between these two banks is:

- RUN::Scaler - beam charge from FC integrated from the beginning of the run.
- HEL::Scaler - beam charge from FC for different helicity of electron beam.

## 7.2 MISSING FC INFORMATION FOR GROUPS OF RUNS

Using the information from the Scaler banks, while we were trying to extract beam charge information for each reconstructed DST file, we noticed that there was no scaler information for runs 12857 to 12877. We tried to look for alternative scalers that could help us extract the beamcharge from these scalers' information, but none of the scalers below were available for the above runs.

- No FC gated [ RUN::Scaler and HEL::Scaler Bank].
- No FC ungated [RUN::Scaler and HEL::Scaler Bank].
- No Synchrotron Light Monitors (SLM) gated.
- No Synchrotron Light Monitors (SLM) ungated.

We had to find a way to calculate any measurable quantity from available scalers or any combination of scalers that could have a one-to-one correspondence with beamcharge.

### 7.2.1 BEAMCHARGE EXTRACTION USING BEAM CURRENT

The Hall B beamline has three beam position monitors (BPMs) - namely 2H01, 2C21, and 2C24. These monitors also measure beam current as the beam passes through them. This current information and several other details on EPICS Process Variables (PVs) are stored in the RAW::epics bank within the CLAS12 data stream. Using information from this bank, we tried to interpret beam charge as:

$$\text{Beamcharge} = \sum_{\text{events}} (\mathit{clock} \times \text{current}) \,,$$

where,

- **clock** is the DAQ-gated clock information (read from HEL::Scaler or RUN::Scaler bank).
- **Current** is read from RAW::epics (measured at 2C21, 2H01, and 2C24).

Among the three BPMs, 2C21 was selected because its current measurements showed the highest consistency with the overall experiment. Before proceeding with this procedure, the beam charge for the runs with existing information was recalculated using the new method and compared against the pre-existing information. If both methods produced comparable results, we could proceed with the new method.

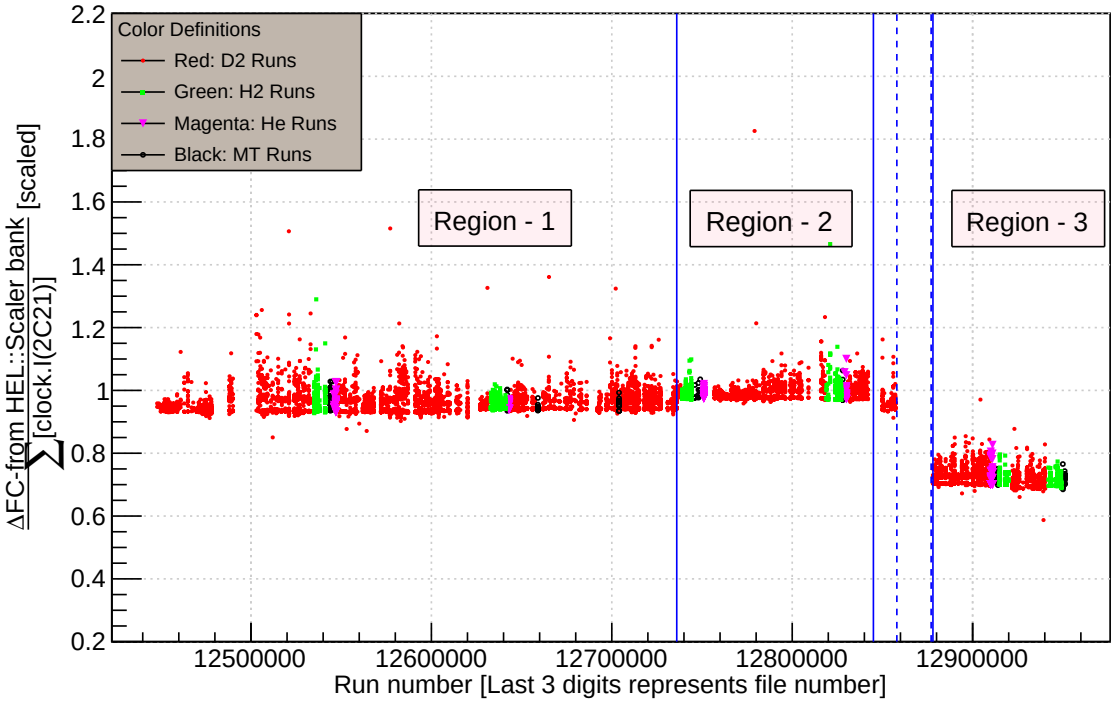



FIG. 93. The ratio of beamcharge to $clock.I_{2C21}$ is plotted against Run-File numbers.

Fig. 93 shows three regions in our run periods. The first region corresponds to runnumber $<$ 12736 or runnumber between 12845 and 12877. The second region corresponds to the runnumber between 12736 and 12845, and the third corresponds to the runnumber above 12877. The drop in Region 3 could be explained by the fact that the Beam Charge Monitor(BCM) was recalibrated before the beginning of those runs. Note: The gap between the two vertical dotted lines refers to runs that have scaler information missing in the RUN::Scaler and HEL::Scaler banks. We see consistency in ratios between beamcharge from the HEL::Scaler bank and $clock.I_{2C21}$. Although the ratio is consistent, we observed two different types of readings. The first type is a regular reading in the FC from the HEL::Scaler bank and $clock.I_{2C21}$. For example, for run number 12448 and file number 000, the beamcharge reading from the HEL::Scaler bank is 44240.2 units, whereas $clock.I_{2C21}$ is 46685.6 units, and their ratio is 0.95. The second type involves lower readings in the beamcharge from the HEL::Scaler bank and $clock.I_{2C21}$. By low reading, we mean lower relative to the

corresponding reading in the RUN::Scaler bank. For example, for run number 12449 and file number 005, the beamcharge reading from HEL::Scaler is 346.543 units, whereas $clock.I_{2C21}$ is 361.02 units, and their ratio is 0.96. Thus, the observation using this fact is that although we have two different types of readings, the ratios between them remained consistent. This behavior was observed for all data files, as shown in Fig. 93.

Fig. 94 shows the relationship between the $clock.I_{2C21}$ and FC from HEL::Scaler bank for all three regions.

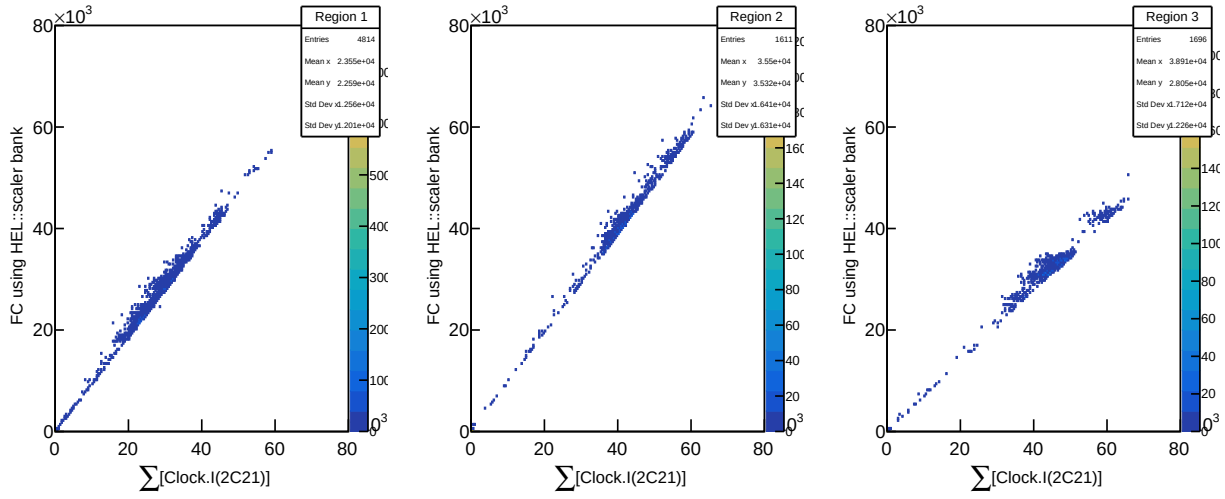



FIG. 94. Histogram, from left to right goes region 1 to region 3, of FC beamcharge, measured from HEL::Scaler bank in Y-axis vs. $clock.I_{2C21}$ in X-axis.

**Region 1**

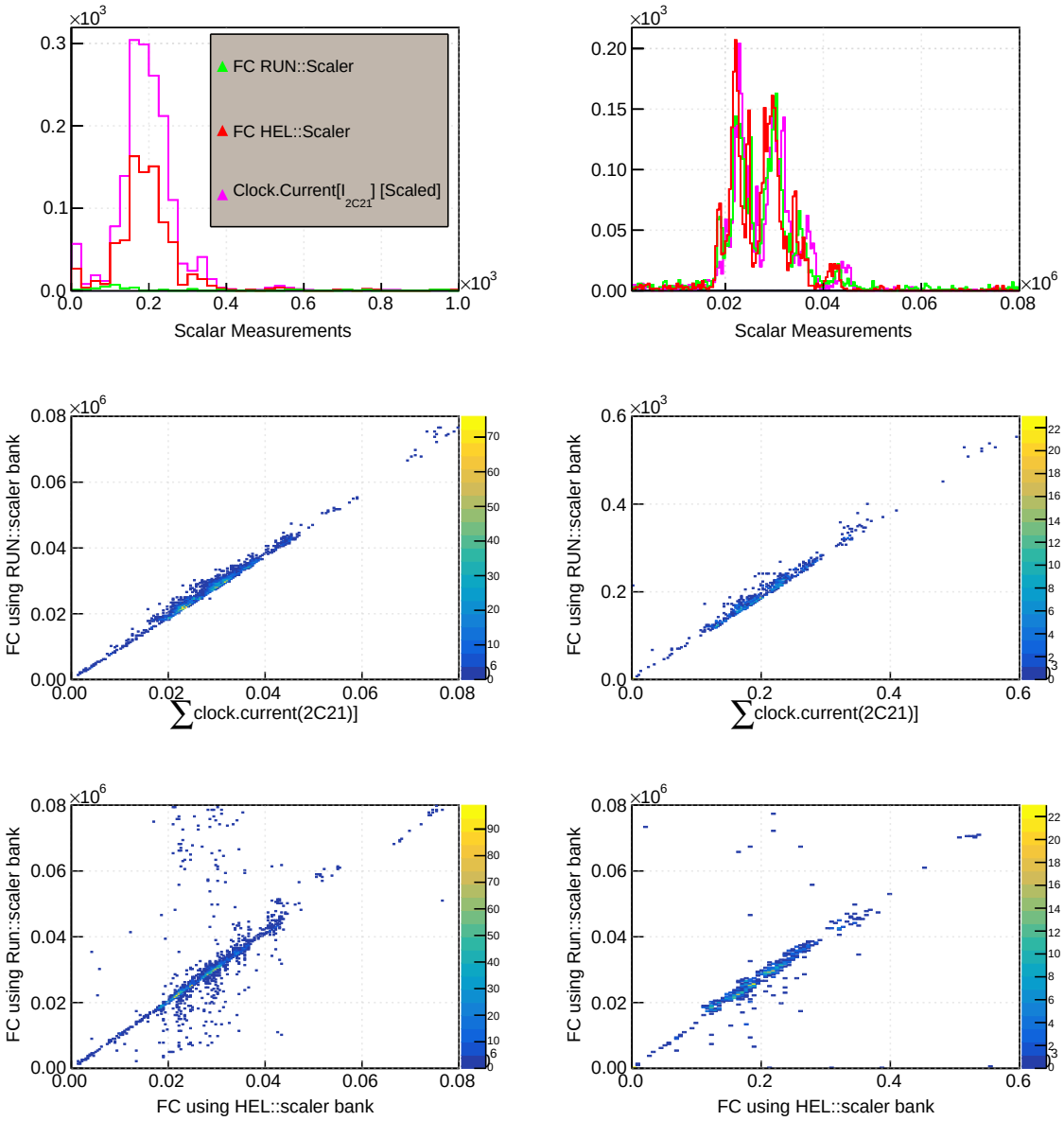



FIG. 95. Top left and right: 1D Histogram of scaler measurements (beamcharge from RUN::scaler bank, HEL::scaler bank, and $clock.I_{2C21}$, middle left and right: 2D Histogram of the beamcharge from HEL::scaler bank in the y-axis vs. the $clock.I_{2C21}$ in x-axis, bottom left and right: Histogram of FC measurements using RUN::scaler bank in y-axis vs. of FC measurements using HEL::scaler bank in x-axis.

**Region 2**

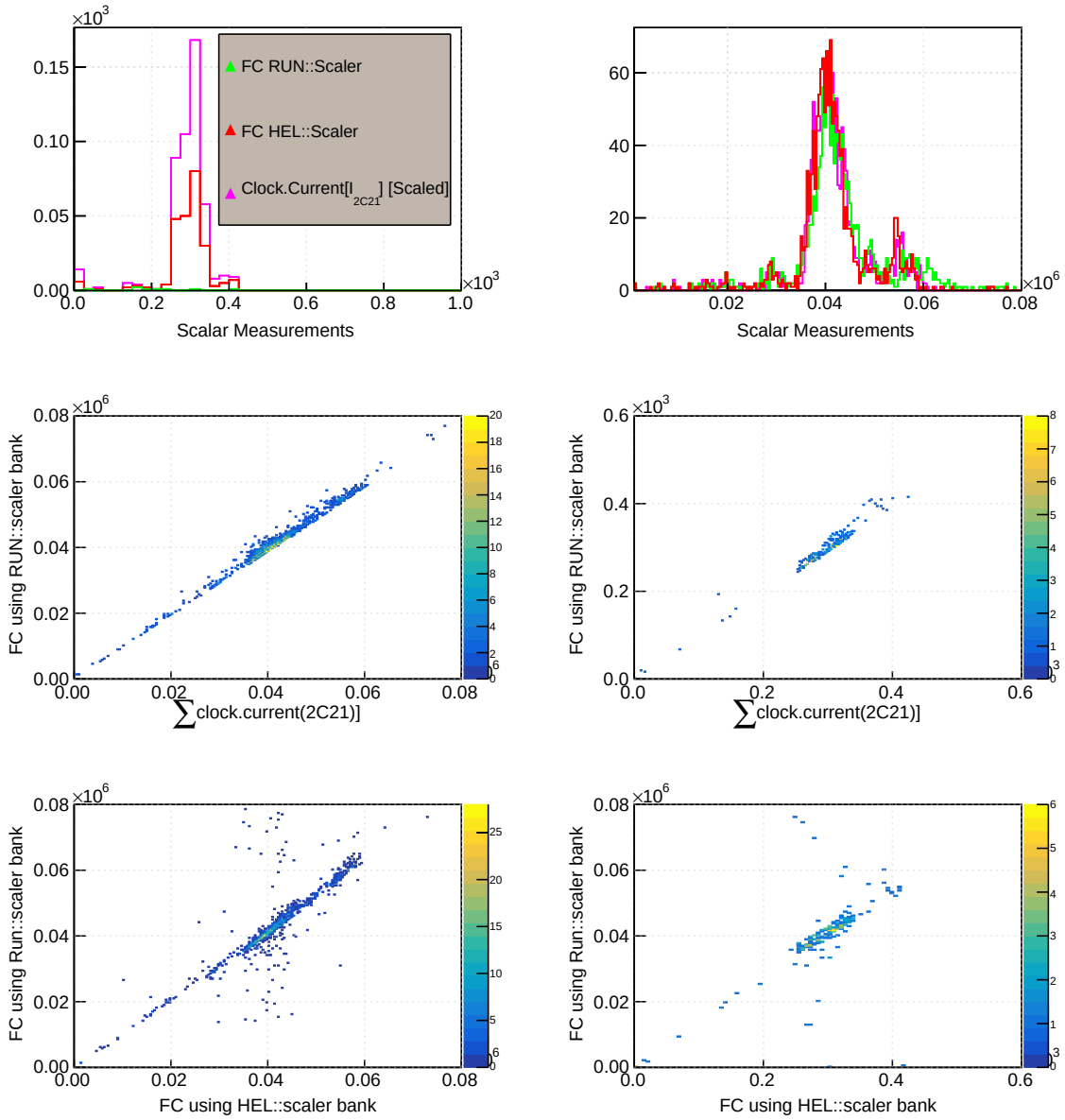



FIG. 96. Top left and right: 1D Histogram of scaler measurements (beamcharge from RUN::scaler bank, HEL::scaler bank, and $clock.I_{2C21}$, middle left and right: 2D Histogram of the beamcharge from HEL::scaler bank in the y-axis vs. the $clock.I_{2C21}$ in x-axis, bottom left and right: Histogram of FC measurements using RUN::scaler bank in y-axis vs. of FC measurements using HEL::scaler bank in x-axis.

**Region 3**

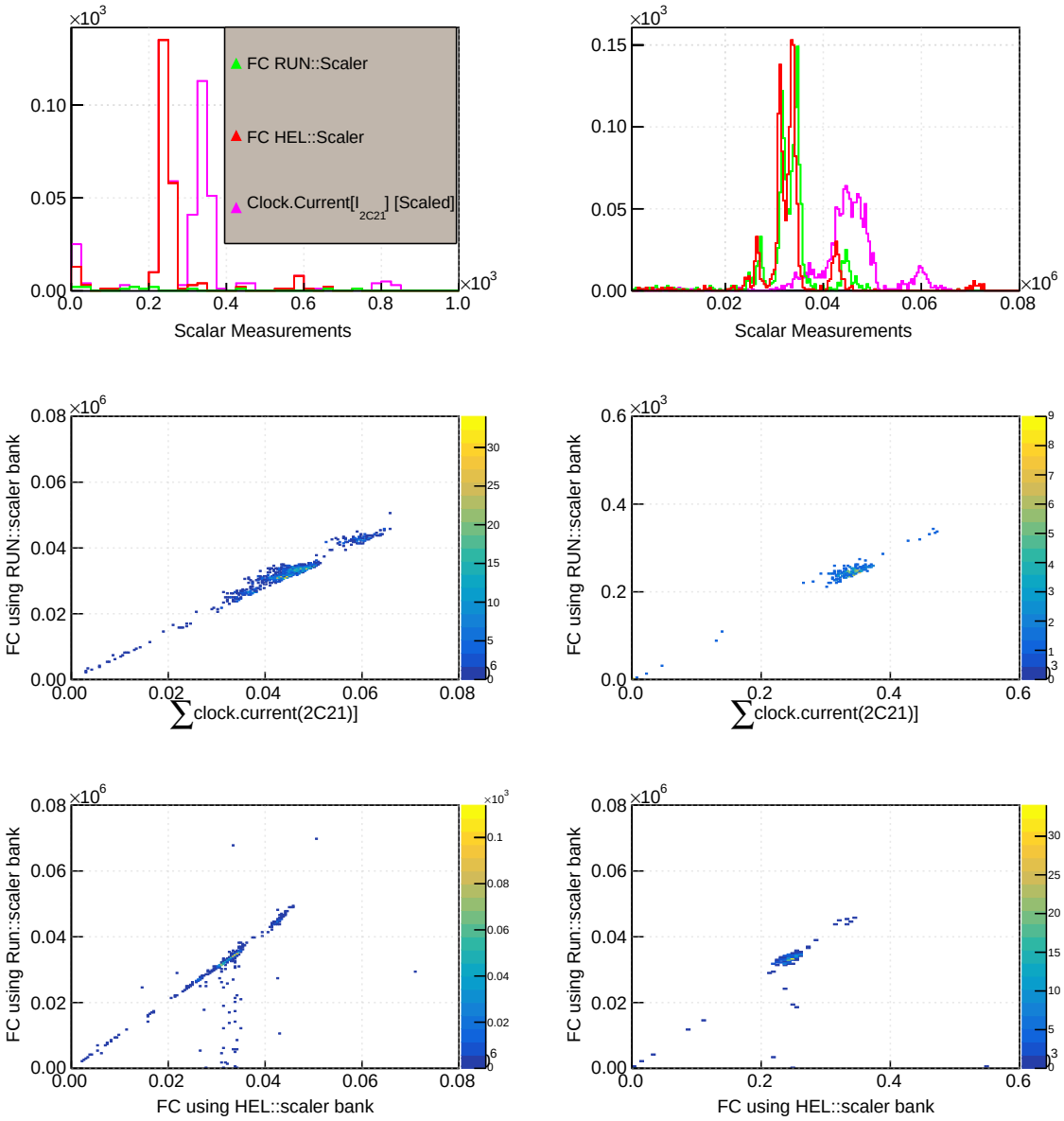



FIG. 97. Top left and right: 1D Histogram of scaler measurements (beamcharge from RUN::scaler bank, HEL::scaler bank, and $clock.I_{2C21}$, middle left and right: 2D Histogram of the beamcharge from HEL::scaler bank in the Y-axis vs. the $clock.I_{2C21}$ in X-axis, bottom left and light: Histogram of FC measurements using RUN::scaler bank in Y-axis vs. of FC measurements using HEL::scaler bank in X-axis.

After studying the above plots and the observations, we noticed that some files exhibit very low readings in HEL::scaler bank, which did not correspond to beamcharge in those cases; hence, we decided to use the RUN::Scaler bank and $clock.I_{2C21}$ to extract beam charge for the runs that have missing information. We followed the beamcharge extraction procedure for all three regions separately, as outlined below.

### 7.2.2 REGION 1 - BEAMCHARGE EXTRACTION

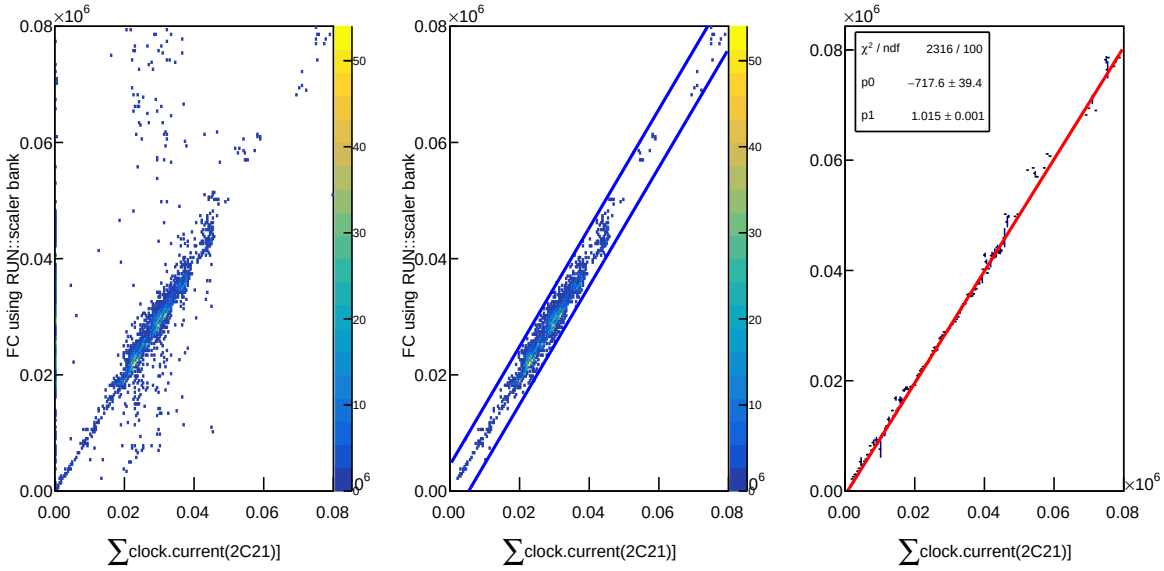



FIG. 98. Left: 2D Histogram of beamcharge from RUN::Scaler bank in y-axis vs. the $clock.I_{2C21}$ in the x-axis, middle: Same 2D Histogram as in the left Plot but only focusing on the region between 2 diagonal lines, right: a profile histogram of the central plot and a 1D profile fit.

After this:

- If $clock.I_{2C21}$ (summed for each event) was less than 1000, we used reading in RUN::Scaler bank as beamcharge.
- If the reading/data lies in the region as shown in the middle plot above, we use the

reading in the RUN::Scaler bank as beamcharge.

- If the reading lies outside the region described by the middle Plot above, we used fit parameters from the fit above right Plot to extract beamcharge from if $clock.I_{2C21}$.

The equation of the top line is, $y = 1.02x + 4500$, and the bottom line is, $y = 1.02x - 4500$. The profile fit on the right is of form $y = p_0 + p_1 x$ with the parameters: $p_0 = -717.6$ and $p_1 = 1.015$.

### 7.2.3 REGION 2 - BEAMCHARGE EXTRACTION

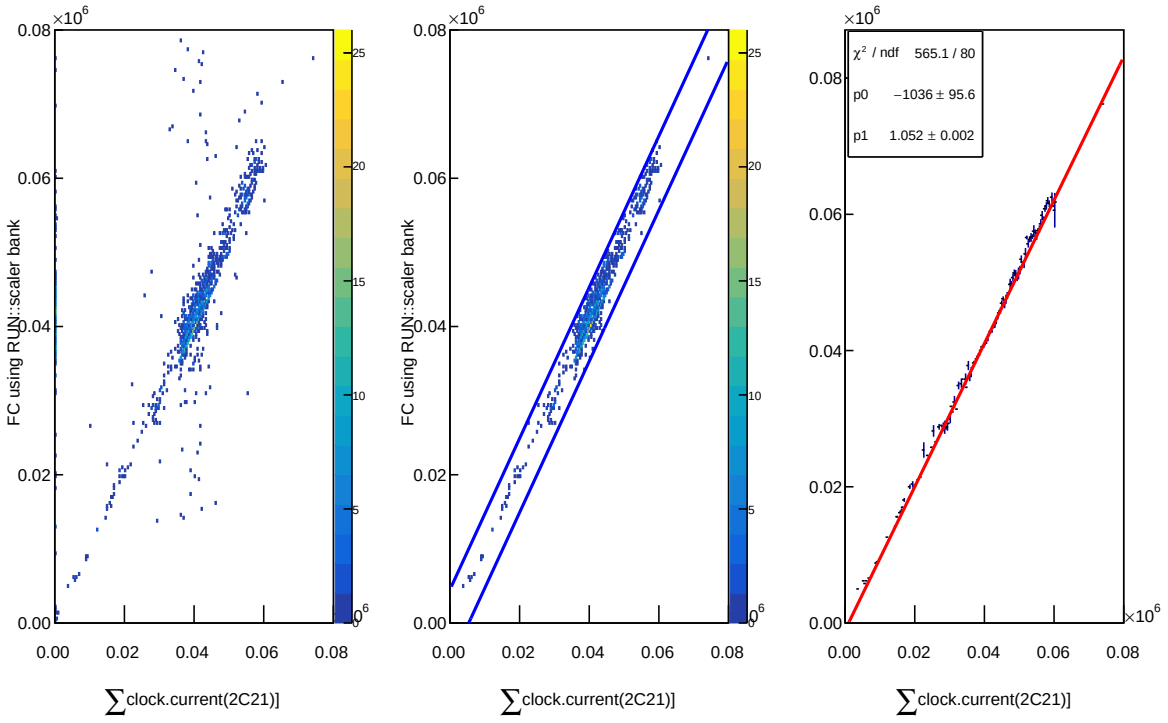



FIG. 99. Left: 2D Histogram of beamcharge from RUN::Scaler bank in y-axis vs. the $clock.I_{2C21}$ in the x-axis, middle: Same 2D Histogram as in the left Plot but only focusing on the region between 2 diagonal lines, and right: a profile histogram of the central plot and a 1D profile fit.

The equation of the top line is, $y = 1.02x + 4500$, and the bottom line is, $y = 1.02x - 5500$. The profile fit on the right is of form $y = p_0 + p_1 x$ with the parameters: $p_0 = -1036$ and $p_1 = 1.052$.

### 7.2.4 REGION 3 - BEAMCHARGE EXTRACTION

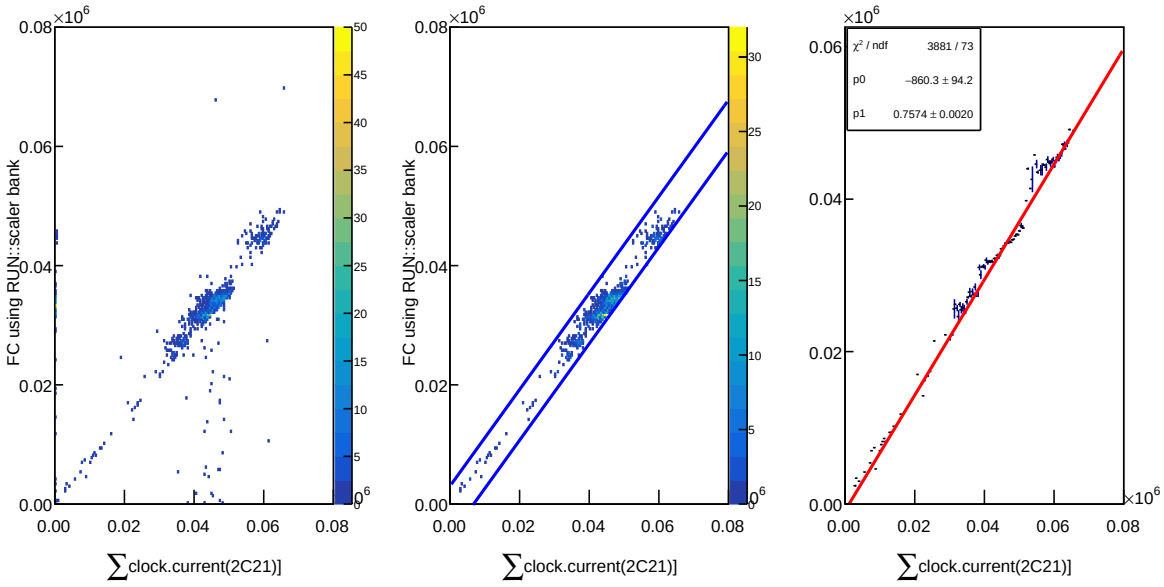



FIG. 100. Left: 2D Histogram of beamcharge from RUN::Scaler bank in y-axis vs. the $clock.I_{2C21}$ in the x-axis, middle: Same 2D Histogram as in the left Plot but only focusing on the region between 2 diagonal lines, and right: A profile histogram of the central plot and a 1D profile fit.

The equation of the top line is, $y = 0.81x + 3000$, and the bottom line is, $y = 0.81x - 5500$. The profile fit on the right is of form $y = p_0 + p_1 x$ with the parameters: $p_0 = -860.3$ and $p_1 = 0.756$. After obtaining these fit parameters for each region, we use these parameters to extract beamcharge for the runs that did not have beamcharge recorded in the CLAS12 database. Hence, we now have FC information for all RGF runs.

## 7.3 GOOD RUN/FILE SELECTION

### 7.3.1 NUMBER OF ELECTRON YIELDS NORMALIZED PER BEAMCHARGE

After all procedures of extracting beamcharge for the runs with missing Faraday cup (FC) information, for each data file in each experimental run, the total number of electrons that satisfy all particle identification cuts were calculated. Then, the number of electrons in each reconstructed DST file was normalized to the total Faraday cup beamcharge recorded for the file. Fig. 101 shows this ratio as a function of the file and run number for the summer 2020 period. Three regions can be identified in Fig. 101 over the experiment, which are separated by the vertical lines. A new trigger was implemented between the first and the second periods (starting by run 12579) [Hall-B logbook entry 3829393]. At the beginning of the third period, run 12878, a new calibration for the Beam Charge Monitor (BCM) was carried out [Hall-B logbook entry 3847392]. In the end, a file selection cut based on this quantity was developed to remove the outliers data files in the three periods.

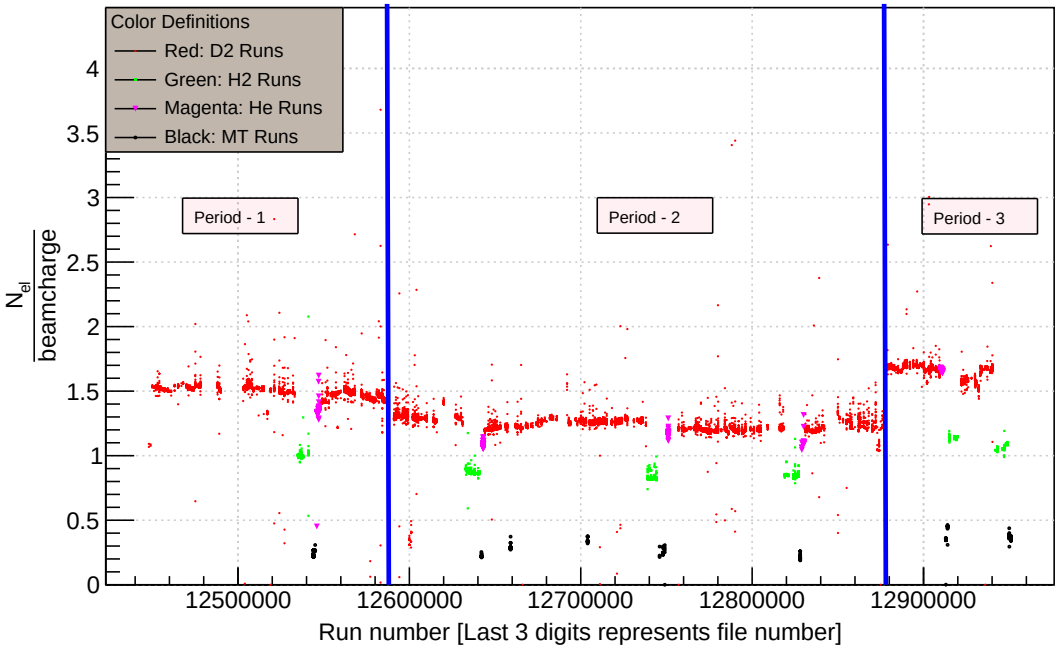



FIG. 101. Graph of normalized electron yield per beamcharge for all RGF files.

For the three periods, we observed in the above plots, Figs. 102, 103, and 104 show the 1D distribution of the number of electrons normalized by beam charge for each period. Using this 1D distribution for normalized yield per beamcharge for each target for each region, we put a cut of $2.5\sigma$ to select good files for final analysis.

**Period 1**

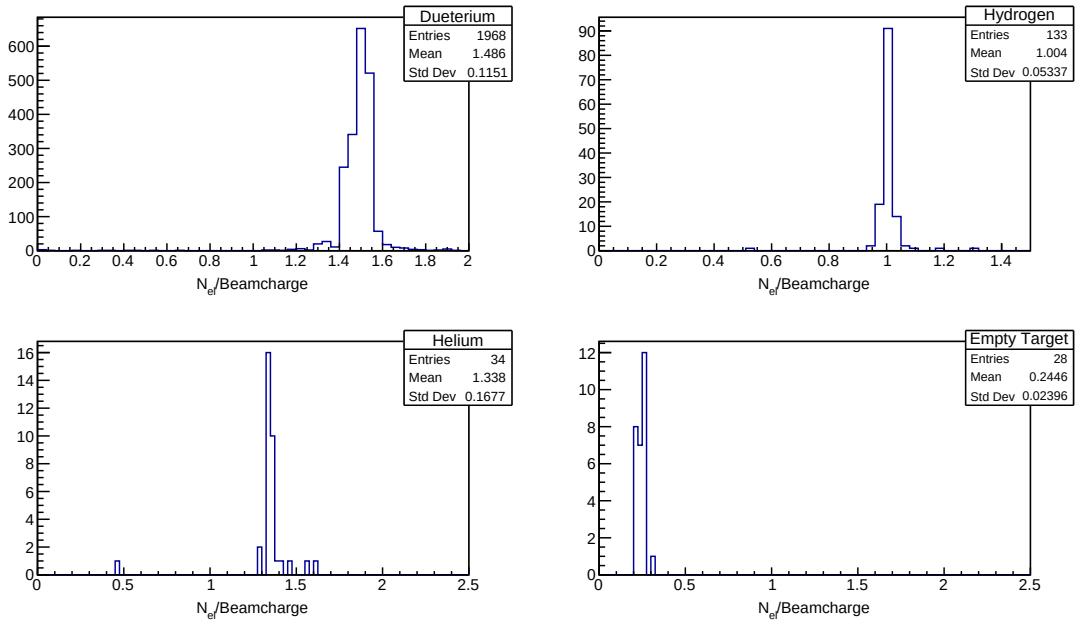



FIG. 102. Plot of the normalised electron yield per beamcharge for period-1 for each target type.

## Period 2

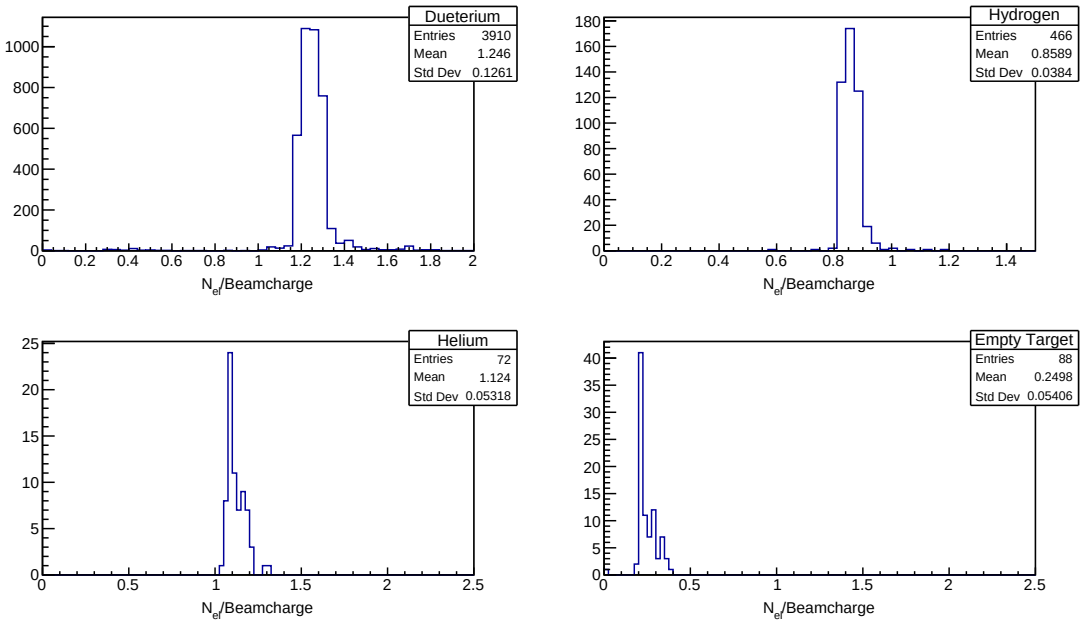



FIG. 103. Plot of the normalised electron yield per beamcharge for period-1 for each target type.

**Period 3**

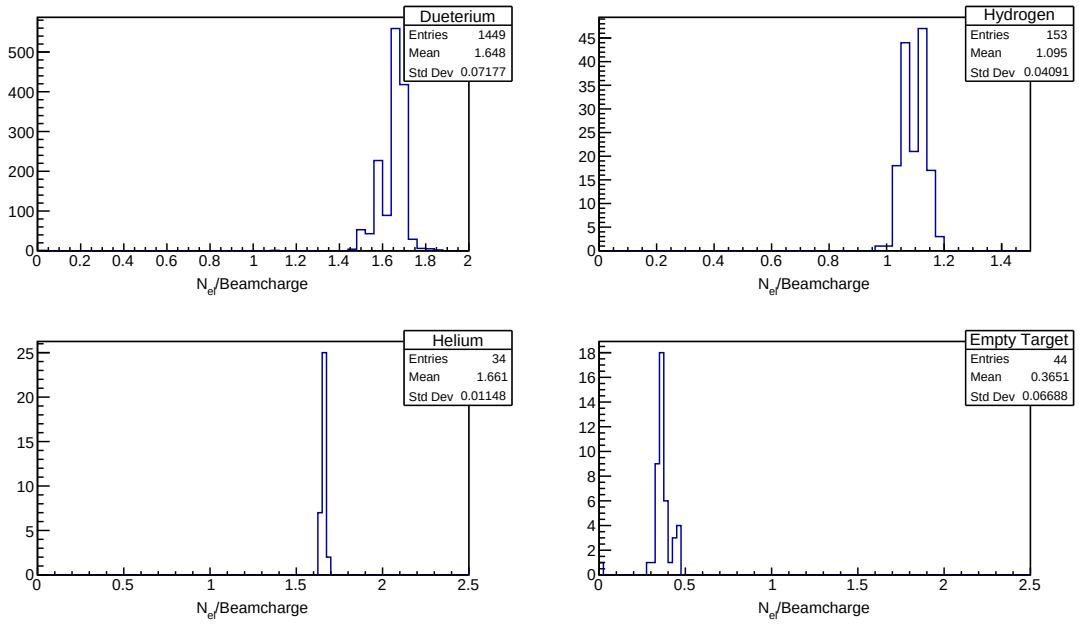



FIG. 104. Plot of the normalised electron yield per beamcharge for period-1 for each target type.

### 7.3.2 NUMBER OF PROTONS IN RTPC PER GOOD ELECTRON

The main goal of the BONuS12 experiment is to use RTPC to identify spectator protons from the reaction $D(e, e'p_s)X$. Therefore, we added file selection criteria based on the number of protons reconstructed in the RTPC detector per good electrons as a function of the file and run numbers. Fig. 105 shows the number of protons per electron rate as a function of experimental time (file and run numbers).

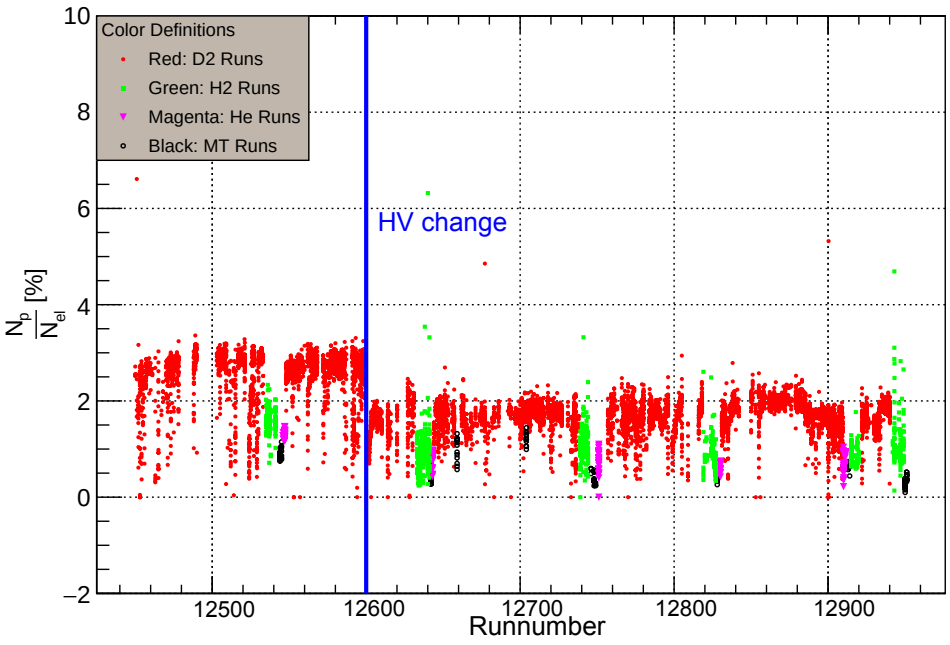



FIG. 105. The integrated number of proton tracks per electron rate as the function of file and run numbers. The blue line represents when the RTPC HV setting was changed on the GEM foils from 385 V to 375 V

Fig. 105 shows two distinct regions, before and after runs 12600. This separation is because of the HV change across the GEM foils inside the RTPC detector after run 12600, which decreased the proton yield. Fig. 106 and 107 show the distribution of number of proton tracks per good electron. Based on these distributions, we selected the optimum cut for selecting good runs based on the number of proton tracks to electron tracks.

**RTPC period 1 [Run:12448 to Run:12600]**

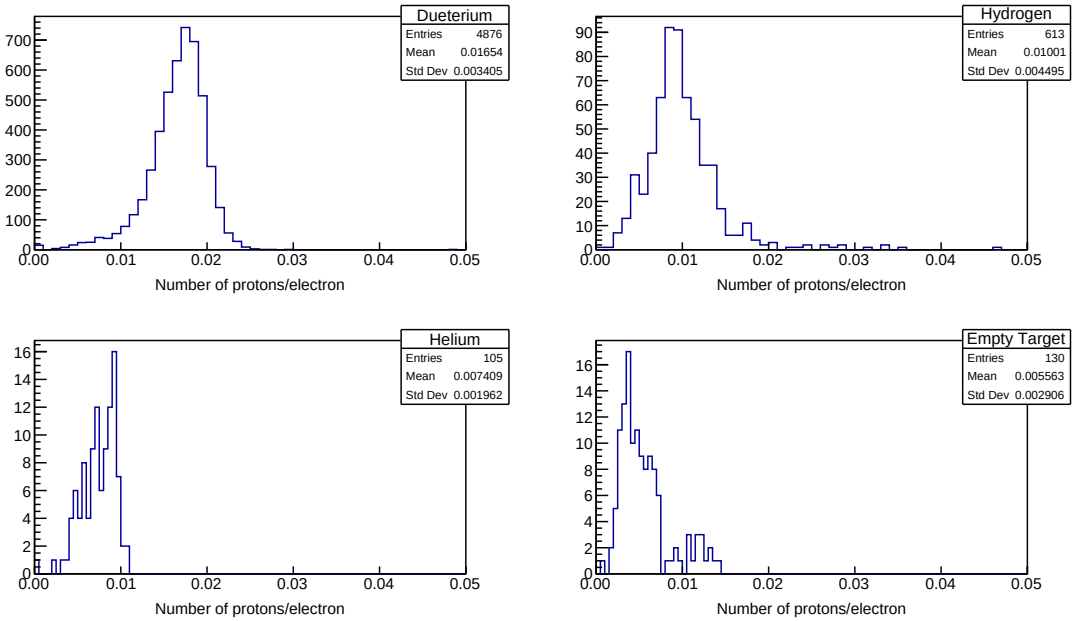



FIG. 106. Plot of the ratio between the number of proton tracks to a number of electrons for RTPC run period 1.

**RTPC period 2 [Run:12601 to Run:12951]**

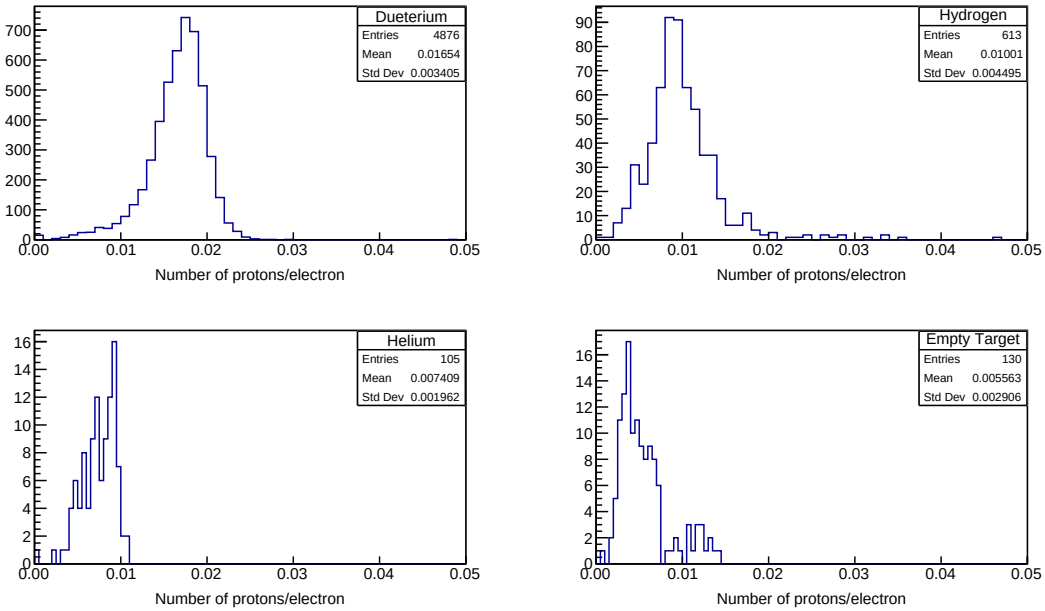



FIG. 107. Plot of the ratio between the number of proton tracks to the number of electrons for RTPC run period 2.

# CHAPTER 8

# PRESENTATION OF DATASET

In the previous chapters, we presented the particle identification criteria for proton and electron selection (chapters 3 and 4), focusing on the two reaction channels of interest: inclusive ($D(e,e')X$) and proton-tagged ($D(e,e'p_s)X$) deep inelastic scattering (DIS) events, aimed at extracting $F_2^n$ and the $d/u$ ratio. Similarly, in chapter 7, we discussed the selection of good data files for the final analysis. Additionally, background subtractions were addressed in chapter 6. After applying all particle and file selection criteria and subtracting backgrounds, this chapter presents the final data set in terms of its kinematic coverage.

The summary of our electron and proton selection cuts are as follows,

## ELECTRON SELECTION CUTS

- PID = 11.
- $-20.35$ cm $< vz_{el} <$ 18.65 cm [Electron selection for Inclusive analysis].
- $-23.35$ cm $< vz_{el} <$ 20.05 cm [Electron selection for Tagged analysis].
- Number of photoelectrons in HTCC $> 2$.
- $(345.3 - (v_z - z_{\text{center}})) \cdot \tan(\theta_{\text{el}}) > 25.0$ mm and $\theta_{\text{el}} > 9^\circ$ , $z_{\text{center}} = 6.5$ mm.
- $(345.3 - (v_z - z_{\text{center}})) \cdot \tan(\theta_{\text{el}}) < 248.5$ mm and $\theta_{\text{el}} < 39^\circ$.
- $E_{\text{pcal}} > 100$ MeV.
- $EC_{\text{in}} > 10$ MeV.
- Fiducial cuts on DC and PCal.
- Sampling fraction cut.
- $E' > 2$ GeV.
- Only one good electron triggering the event.
- **DIS cuts** $Q^2 > 1.3$ GeV$^2$ and $W > 1.8$ GeV.

**PROTON SELECTION CUTS**

- Radius of helix of track $< 0$.
- $\chi^2 < 5$.
- Minimum 10 hits in a track.
- Maximum radius between 67 and 72 mm.
- $-21$ cm $< vz_p < 18$ cm.
- Vertex coincidence cuts.
- Timing coincidence.
- Energy loss ($dE/dx$) within proton band.
- **VIP cut** Momentum of proton (70 - 100 MeV/c).
- **Spectator cut** $\cos(\theta_{pq}) < -0.3$.
- **DIS cuts** $Q^2 > 1.3$ GeV$^2$ and $W^* > 1.8$ GeV.

## 8.1 INCLUSIVE DISTRIBUTIONS

Figs. 108 and 109 shows the distributions of inclusive electrons. Inclusive distributions refer to measurements of electrons irrespective of whether other particles are being measured or not.

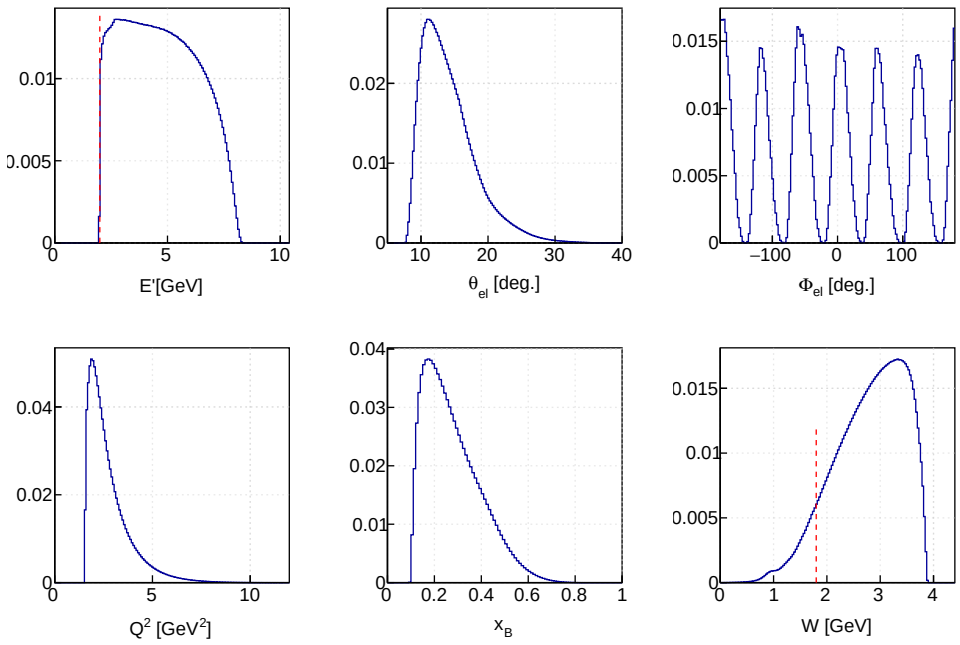



FIG. 108. The distribution of inclusive electron kinematics. The plots from left to right and from top to bottom are: energy ($E'$),polar-angle ($\theta$), azimuthal-angle ($\phi$), $Q^2$, $x$, and rest-nucleon invariant-mass ($W$).

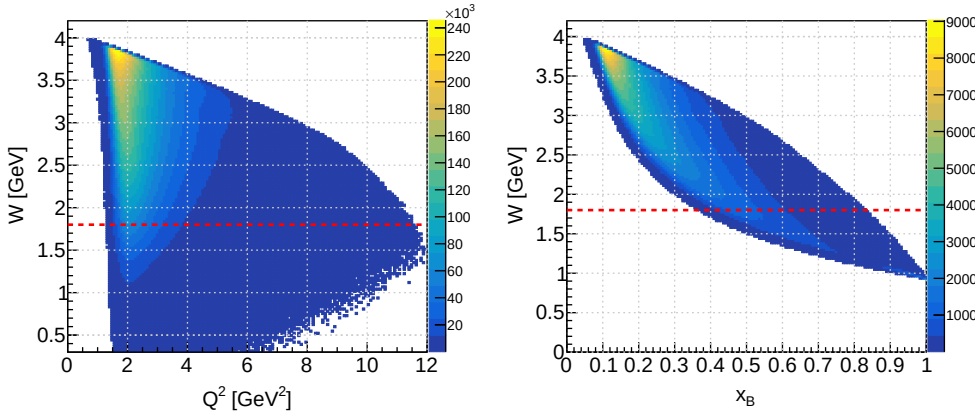



FIG. 109. 2D distribution of $W$ vs. $Q^2$ (left) and $W$ vs. $x$ (right) for inclusive electron kinematics.

## 8.2 TAGGED DISTRIBUTIONS

Fig. 110 shows the distributions of electrons and, 111 shows the distributions of protons, for the tagged case. Tagged distributions refer to measurements whenever an event is reconstructed in RTPC.

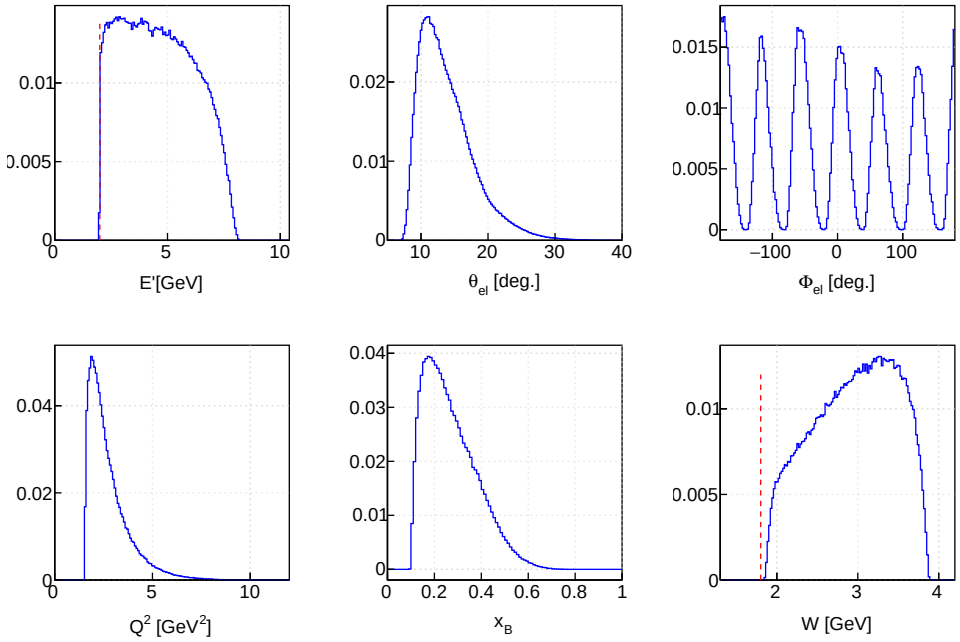



FIG. 110. The distribution of electron kinematics for the tagged case. The plots from left to right and from top to bottom are: energy ($E'$),polar-angle ($\theta$), azimuthal-angle ($\phi$), $Q^2$, $x$, and rest-nucleon invariant-mass ($W$).

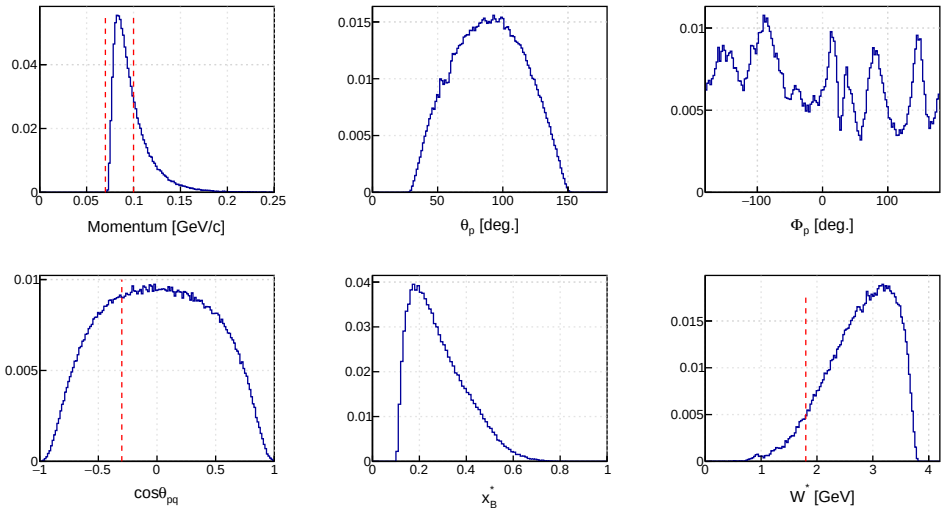



FIG. 111. The distribution of proton kinematics for tagged data. The plots from left to right and from top to bottom are: energy ($E'$), polar-angle ($\theta$), azimuthal-angle ($\phi$), cos$\theta_{pq}$, $x$, and invariant-mass ($W^*$) of the struct neutron.

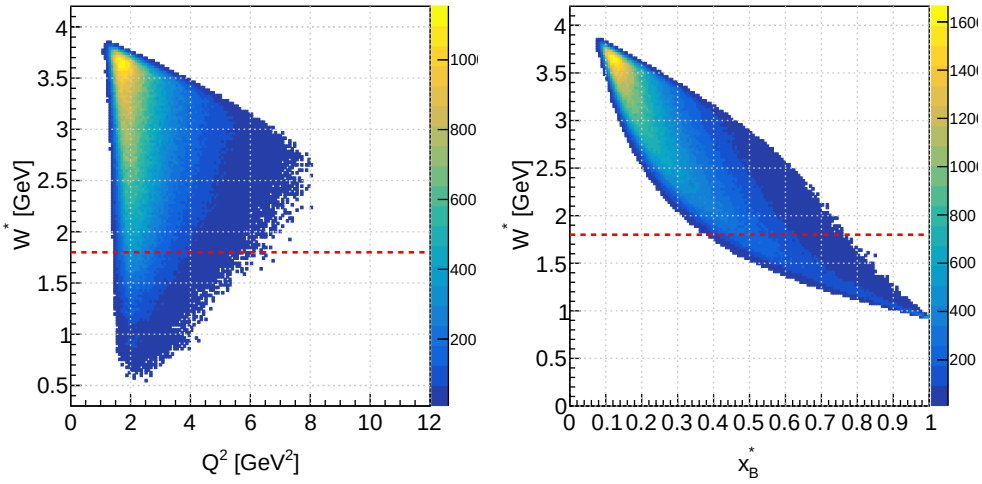



FIG. 112. 2D distribution of $W^*$ vs. $Q^2$ (left) and $W^*$ vs. $x^*$ (right) for tagged proton kinematics.

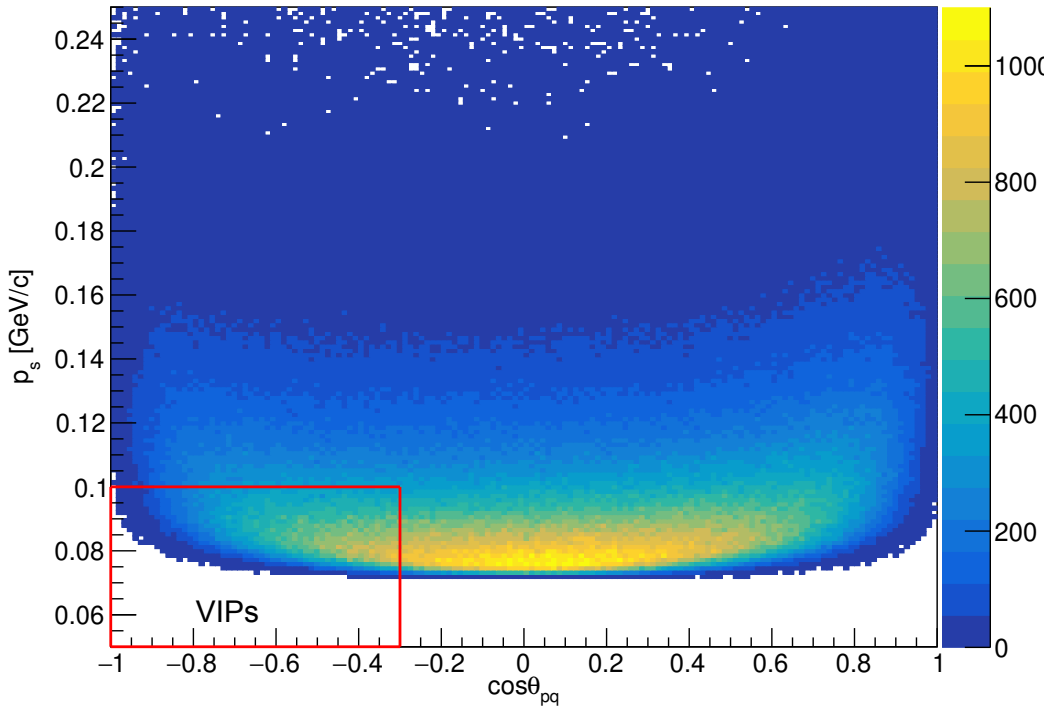



FIG. 113. 2D distribution of momentum reconstructed in RTPC vs. the cosine of the angle between spectator proton and the virtual photon exchanged.

# CHAPTER 9

# RESULTS AND CONCLUSIONS

With the previously presented particles' identification, the final state particles of the two channels of interest, i.e., inclusive ($D(e,e')X$) and proton-tagged ($D(e,e'p_s)X$) DIS events, are being identified and selected for further analysis of extracting $F_2^n$ and the $d/u$ ratio. This chapter presents our relatively model-independent extraction for $F_2^n/F_2^p$ with the associated statistical and systematic uncertainties.

The proposed spectator-proton tagging technique does not only tag events to ensure that the scattering has occurred on the bound neutron of interest, but the detection of the spectator proton gives complete information about the initial state of the explored bound neutron in addition to fine-tuning of the selection of the events where the bound neutron has a relatively small Fermi momentum and is close to being a quasi-free neutron. Fig. 114 demonstrates the effectiveness of the tagging technique using the calculated invariant mass distribution with the 2 GeV incident electron beam on our $D_2$ gaseous target. One can see that the calculated invariant mass distribution for the tagged data results in a better resolution for the (quasi)-elastic peak and in the resonance region compared to the inclusive data set.

## 9.1 STRUCTURE FUNCTION RATIO EXTRACTION

The 10.389 GeV production data on the $D_2$ gaseous target have been analyzed to identify the events of both physics channels of interest, i.e., inclusive ($D(e,e')X$) and proton-tagged ($D(e,e'p_s)X$) DIS events. All the cuts and conditions that we apply for final DIS events have been described in chapters 4 and 5, and summarised in chapter 8.

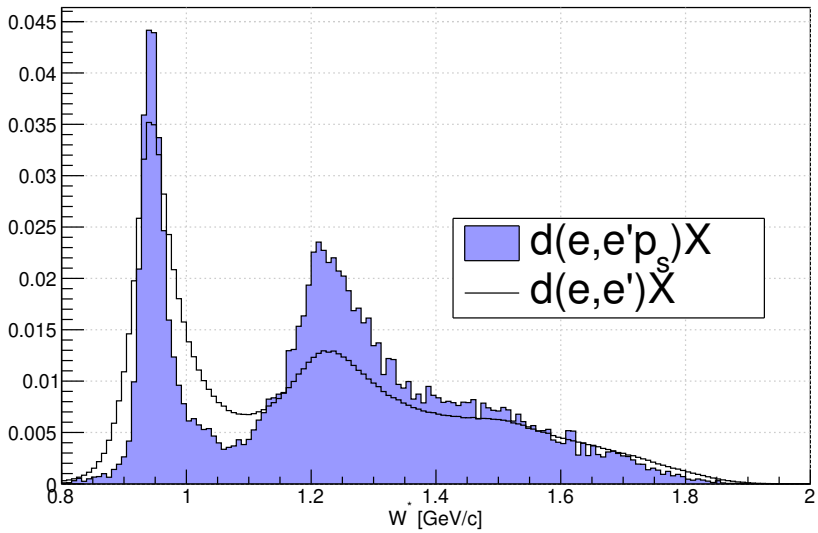



FIG. 114. The normalized neutron invariant mass distributions are calculated using the on-shell mass of the neutron (black line) through the inclusive data and the calculated off-shell mass of the initial neutron (blue shade) through the proton-tagged events.

Once the inclusive and the tagged events have been identified, these events are binned into 2D bins in $x$ and $Q^2$ for the inclusive data and $x'$ and $Q^2$ for the VIP proton-tagged data, following the binning, as can be seen in Fig. 115:

- $x/x^*$ bins: 50 equidistant bins from 0 to 1.
- $Q^2$ bins: 16 bins with bin boundaries (0.92, 1, 1.3, 1.56, 1.87, 2.23, 2.66, 3.17, 3.79, 4.52, 5.40, 6.45, 7.70, 9.19, 10.97, 13.09, 15.63).

After binning our data in those 2D bins, the following procedures will convert the yield in each bin into the structure function ratio of $F_2^n/F_2^p$ by analyzing both data sets as following.

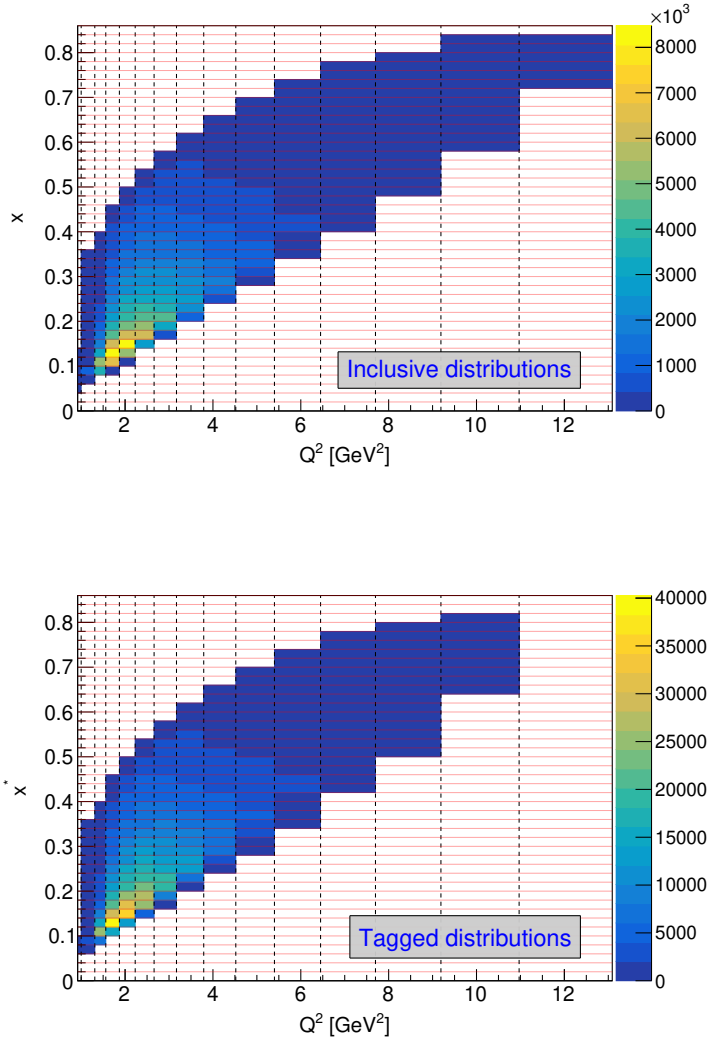



FIG. 115. (Top) $x$ versus $Q^2$ for the collected inclusive $(D(e,e')X)$ DIS events,(Bottom) $x'$ versus $Q^2$ for the collected VIP proton-tagged $(D(e,e'p_s)X)$ DIS events.

### 9.1.1 GEMC SIMULATION

My work presented in this thesis is based on experimental data. But to extract the final result $F_2^n/F_2^p$, a comparison to simulated data is required. Although I wasn't directly involved in generating these simulations, I've included a brief introduction to the GEMC

simulation framework to provide context for the simulated data used in this study.

To accurately model the response of CLAS12 and RTPC detector system and to assess its impact on the kinematics of detected events, detector acceptance effects, resolution effects, and the efficiency of the measurements of final state particles, we used the GEMC (Geant4 Monte Carlo) simulation package developed by the CLAS Collaboration at Jefferson Lab. It is a C++ based framework that utilizes Geant4's modeling of matter interactions along with the Monte Carlo method of randomized sampling to obtain the particle's response and trajectory in various detectors.

The first step in the simulation chain is to generate scattering events based on the input physics model. In our simulation chain, we adopted an approach in which the number of events generated in any kinematical bin is proportional to the integrated cross-section over that bin. These generated events were saved in LUND-based data format and later used as input files for GEMC simulation. At this point, it is already evident that this analysis is composed of two parts: inclusive analysis and tagged analysis. So, we generate separate LUND files for inclusive and tagged events. The event generator we used for the BONuS12 simulation is an updated version of the event generator used in the BONuS6 experiment [20]. The output of the generator contains a certain number of events (fixed during the running of the generator code), which contains the kinematics of scattered electrons for inclusive cases. In contrast, the tagged case contains scattered electrons and a spectator proton, plus any radiated photons from internal "after scattering" higher-order QED processes and a selectable number of background protons. However, for this analysis, the background protons were not generated. For tagged DIS, we modeled the deuteron as composed of a moving neutron and a spectator proton, with the spectator momentum distribution governed by the Argonne V18 deuteron wave function [38]. To enhance efficiency and avoid generating undetectable protons in real experimental counterparts, we impose a lower cutoff on spectator proton momenta at 60 MeV/c. We model the scattering using fully relativistic kinematics on the neutron rest frame. To do this, we use cross-section extracted from the nucleons' structure function parametrization by Christy and Bosted [39], which describes inclusive cross-sections over a wide range. The exact normalization of the cross-section is not super important due to its cancellation. However, since we extract the ratio $F_2^n/F_2^d$, it remains essential to model the shapes of all relevant distributions accurately. This ensures that effects such as detector smearing, bin migration, and radiative tails are correctly considered. Similarly, the effect of radiations (internal and external) is taken into account using the equivalent radiator method developed by Mo and Tsai. In the case of pre-scattering radiation, photons emitted before

scattering are assumed to follow the scattered electron's direction (following the peaking approximation) and are added to the LUND output. Final-state external radiation, along with energy losses in the detector materials, is handled by the GEMC simulation itself during particle propagation.

Inclusive DIS events are generated in a similar manner to tagged events, with the primary difference being that scattering off both moving neutrons and protons in the deuteron is included. The generator samples the entire available momentum phase space, and each event includes the scattered electron, any radiated photons, and, for tagged events, the spectator proton. Once the events are generated and saved in the LUND file, these events are passed through the GEMC framework, which simulates the whole CLAS12 detector geometry and response. The RTPC is also incorporated into the CLAS12 GEMC framework. GEMC tracks each particle through the detector, modeling all relevant interactions with detector materials and converting the outcomes into digitized signals. These simulated outputs are formatted identically to real data files, enabling the use of the same reconstruction software and analysis techniques used in processing experimental data. By comparing simulated events to real measurements, we can extract the structure function ratio $F_2^n/F_2^d$, corrected for detector acceptance, resolution, and kinematic smearing effects. This approach lets us cleanly separate physics-driven behavior from detector-induced distortions and estimate systematic uncertainties with high confidence.

### 9.1.2 INCLUSIVE YIELD RATIO

The inclusive yield ratio ($R_{\text{inc}}(x, Q^2)$) is defined as:

$$R_{\text{inc}}(x, Q^2) = \frac{Y_{\text{inc}}^{\text{Data}}}{Y_{\text{inc}}^{\text{MC}}}, \tag{119}$$

where,

- $Y_{\text{inc}}^{\text{Data}}(x, Q^2)$ is the total number of inclusive $D(e, e')X$ events reconstructed in a given $x$ and $Q^2$ bin after all corrections for backgrounds and all cuts.
- $Y_{\text{inc}}^{\text{MC}}(x, Q^2)$ is the total number of simulated inclusive $D(e, e')X$ events in the same bin after the full chain event generation, simulation, reconstruction, and analysis with equivalent cuts to real data.

These data ($Y_{\text{inc}}^{\text{Data}}(x, Q^2)$) and Monte-Carlo ($Y_{\text{inc}}^{\text{MC}}(x, Q^2)$) yields are defined as:

$$Y_{\text{inc}}^{\text{Data}}(x, Q^2) \sim \mathcal{L} \cdot \text{A}(x, Q^2) \cdot \eta(x, Q^2) \cdot \Delta\sigma_{\text{inc}}(x, Q^2), \tag{120}$$

$$Y_{\text{inc}}^{\text{MC}}(x,Q^2) \sim \mathcal{L}_{\text{LUND}} \cdot \text{A}(x,Q^2) \cdot \eta(x,Q^2) \cdot \Delta\sigma_{\text{inc}}^{\text{Sim}}(x,Q^2). \tag{121}$$

where,

- $\mathcal{L}$ is the total integrated luminosity for the duration of the entire data set, in beam-electrons times the aerial density of the target.
- $\mathcal{L}_{\text{LUND}}$[1] the equivalent luminosity from the event generator, which is essentially given by how many events we create in the LUND file divided by the cross-section per created event.
- $A \cdot \eta$ is the acceptance (fraction of the total solid angle that maps onto the bin in $x$ and $Q^2$ reconstructed by CLAS12) times the efficiency of all detectors and cuts required to record an event. The crucial assumption here is that GEMC is a high-fidelity model of CLAS12, and hence, those two are the same for data and simulation.
- $\Delta\sigma_{\text{inc}}$ is the "true" differential cross section integrated over the bin in $x$ and $Q^2$. For the data, of course, it is unknown, but we assume it is directly proportional to $F_{2d}$, with resolution smearing and radiation as well as multiple scattering. We assume that our generator is realistic enough. Hence, the effects of this folding are the same on the simulated and the real data.
- $\Delta\sigma_{\text{inc}}^{\text{Sim}}(x,Q^2)$ is the "smeared" and radiated cross-section from the simulation, but without the acceptance and efficiency that comes from the full simulation chain.

With the assumption that $\Delta\sigma \propto F_{2d}$ for both data and MC simulation, one can re-write the inclusive yield ratio as:

$$R_{\text{inc}}(x,Q^2) = \frac{Y_{\text{inc}}^{\text{Data}}}{Y_{\text{inc}}^{\text{MC}}} \propto \frac{F_{2d}^{\text{true}}(x,Q^2)}{F_{2d}^{\text{Gen}}(x,Q^2)}. \tag{122}$$

### 9.1.3 PROTON-TAGGED NDIS YIELD RATIO

Similar to the inclusive case, we define the proton-tagged nDIS ratio as:

$$R_{\text{tag}}(x^*,Q^2) = \frac{Y_{\text{tag}}^{\text{Data}}}{Y_{\text{tag}}^{\text{MC}}}, \tag{123}$$

[1]LUND format of data is a text-based file format which is highly used in High Energy Physics Simulation

where, the proton-tagged nDIS yield for data ($Y_{\rm tag}^{\rm Data}(x^*, Q^2)$) and simulation ($Y_{\rm tag}^{\rm MC}(x^*, Q^2)$) are defined as:

$$Y_{\rm tag}^{\rm Data}(x^*, Q^2) \sim \mathcal{L}^{\rm tag}\Bigg[\int_{\rm VIP} {\rm A_{el}} \cdot {\rm A_p} \cdot \eta_{\rm el}(x^*, Q^2) \cdot \eta_{\rm p}(x^*, Q^2) \cdot S(\alpha_s, p_s^T, Q^2) \cdot d{\rm p_s}\, d\alpha_{\rm s}\Bigg] \cdot \Delta\sigma_{\rm n}(x^*, Q^2), \tag{124}$$

$$Y_{\rm tag}^{\rm MC}(x^*, Q^2) \sim \mathcal{L}_{\rm LUND}^{\rm tag}\Bigg[\int_{\rm VIP} {\rm A_{el}} \cdot {\rm A_p^{Gen}} \cdot \eta_{\rm el}(x^*, Q^2) \cdot \eta_{\rm p}^{\rm Gen}(x^*, Q^2) \cdot S^{\rm Gen}(\alpha_s, p_s^T, Q^2) \cdot d{\rm p_s}\, d\alpha_{\rm s}\Bigg] \cdot \Delta\sigma_{\rm n}^{\rm sim}(x^*, Q^2). \tag{125}$$

with

- $\mathcal{L}^{\rm tag}$ is the total integrated luminosity for the duration of the entire data set for the events where tagged protons are detected in RTPC with all cuts applied.
- $\mathcal{L}_{\rm LUND}^{\rm tag}$ is the equivalent luminosity for the generator, which is essentially given by how many events we create in the LUND file with a spectator proton in RTPC divided by the cross-section per created event.
- ${\rm A_{el}} \cdot \eta_{\rm el}$ is the acceptance (fraction of the total solid angle that maps onto the bin in $x'$ and $Q^2$ reconstructed by CLAS12) times the efficiency of all detectors and cuts required to register an event for the events with tagged spectator in RTPC with all cuts applied.
- ${\rm A_p} \cdot \eta_{\rm p}$ is the acceptance (fraction of the total solid angle that maps onto the bin in $x'$ and $Q^2$ reconstructed by RTPC) times the efficiency of RTPC and cuts required to record an event.
- $\Delta\sigma_{\rm n}$ is the "true" differential cross section integrated over the bin in $x'$ and $Q^2$. For the data, of course, it is unknown, but we assume it is directly proportional to $F_{2n}$, the structure-function of the neutron.
- $\Delta\sigma_{\rm tag}^{\rm Sim}(x^*, Q^2)$ is the "smeared" and radiated cross-section from the simulation, but without the acceptance and efficiency that comes from the full simulation chain.

- $S(\alpha_s, p_s, Q^2)$ is the spectral function, which is the probability of detecting a VIP spectator with given kinematics with all cuts applied. The suffix "Gen" refers to the generator, and "Data" refers to experimental data.

similarly, with the assumption that $\Delta\sigma_{\rm n}$ is directly proportional to $F_{2n}$ for both data and simulation, one can re-write this proton-tagged nDIS yield ratio as:

$$R_{\rm tag}(x^*, Q^2) = \frac{Y_{\rm tag}^{\rm Data}}{Y_{\rm tag}^{\rm MC}} \propto \frac{F_{2n}^{\rm true}(x^*, Q^2)}{F_{2n}^{\rm Gen}(x^*, Q^2)}. \tag{126}$$

### 9.1.4 SUPER RATIO AND STRUCTURE FUNCTIONS

In our context, a Super Ratio ($SR$) is defined as:

$$\begin{aligned} SR &= \frac{R_{\rm tag}(x^*, Q^2)}{R_{\rm inc}(x, Q^2)} = \frac{\left(Y_{\rm tag}^{\rm Data}/Y_{\rm tag}^{\rm MC}\right)}{\left(Y_{\rm inc}^{\rm Data}/Y_{\rm inc}^{\rm MC}\right)} = \text{Constant} \cdot \frac{\left(F_{2n}^{\rm true}/F_{2n}^{\rm Gen}\right)}{\left(F_{2d}^{\rm true}/F_{2d}^{\rm Gen}\right)}, \\ \Rightarrow SR &= \frac{\left(Y_{\rm tag}^{\rm Data}/Y_{\rm inc}^{\rm Data}\right)}{\left(Y_{\rm tag}^{\rm MC}/Y_{\rm inc}^{\rm MC}\right)} = \text{Constant} \cdot \frac{\left(\dfrac{F_{2n}}{F_{2d}}\right)^{\rm true}}{\left(\dfrac{F_{2n}}{F_{2d}}\right)^{\rm Gen}}. \end{aligned} \tag{127}$$

Hence, the experimentally measured structure function ratio $\left(\dfrac{F_{2n}}{F_{2d}}\right)^{\rm true}$ is expressed as:

$$\left(\frac{F_2^n}{F_2^d}\right)^{\rm true} = \text{Constant} \cdot \left(\frac{F_2^n}{F_2^d}\right)^{\rm Gen} * \frac{\left(Y_{\rm tag}^{\rm Data}/Y_{\rm inc}^{\rm Data}\right)}{\left(Y_{\rm tag}^{\rm MC}/Y_{\rm inc}^{\rm MC}\right)}. \tag{128}$$

Where $Y_{\rm tag}^{\rm Data}/Y_{\rm inc}^{\rm Data}$ and $Y_{\rm tag}^{\rm MC}/Y_{\rm inc}^{\rm MC}$ are measured ratios calculated by counting the number of events in each bin in $x/x^*$ and $Q^2$. The $F_2^{n\,({\rm Gen})}$ and $F_2^{d\,({\rm Gen})}$ can be calculated using the input model of the generator.

Note: The inclusive data are binned in $x$ and $Q^2$, while the tagged measurements are binned in $x^*$ and $Q^2$, where $x^*$ is the spectator-corrected $x$. All the definitions are similar to the inclusive case but correspond to the case where the spectator proton is measured in the RTPC.

## 9.2 YIELD AND STATISTICAL UNCERTAINTY CALCULATIONS

Following the proposed $Q^2 - x$ two-dimensional binning of the data, the true number of events in each bin and the statistical uncertainties are defined as follows, depending on the channel of analysis.

### 9.2.1 PROTON-TAGGED NDIS EVENTS

The true number of $D(e,e'p_s)X$ DIS events in each bin is defined as:

$$N^{tag}_{true} = [N^{D}_{sig} - N^{D}_{bkg}] - \gamma[N^{He}_{sig} - N^{He}_{bkg}], \tag{129}$$

Where,

$N^{D}_{sig}$ is the number of the on-time tagged events from $D_2$ data,
$N^{D}_{bkg}$ is the number of the tagged events from the accidental background in the $D_2$ data,
$N^{He}_{sig}$ is the number of the on-time tagged events from the $^4$He data,
$N^{He}_{bkg}$ is the number of the tagged events from the accidental background in the $^4$He data,
$\gamma$ is the $^4$He contamination factor in the $D_2$ data.

In (129), all the count numbers are corrected for the positron contamination. Meaning that, if $(N^{D}_{sig})_{raw}$ is the experimentally collected on-time tagged events from $D_2$ data, then the positron-corrected number $(N^{D}_{sig})$ is

$$N^{D}_{sig} = (N^{D}_{sig})_{raw} \ \cdot \ CF(E',\theta). \tag{130}$$

with $CF$ being the positron correction factor, which will be calculated based on average $\theta$ and $E'$ for each kinematical bin. Section 6.2.3 demonstrates its application. Regarding the statistical uncertainties, each term is calculated as following:

- For on-time events: $\sigma^2_N = (CF)^2 \cdot N$.
- For accidental backgrounds: $\sigma^2_N = \frac{1}{14^2} \cdot (CF)^2 \cdot N$.

Where factor 14 comes from the fact that we use 15 by 15 *ep* event pairs to sample accidental background, out of which one electron-proton pair forms a signal and 14 are combinatorial backgrounds. Finally, the total statistical uncertainty in each bin can be calculated as:

$$(\sigma^{tag}_{\rm true})^2 = (\sigma^{D}_{\rm sig})^2 + (\sigma^{D}_{\rm bkg})^2 + \gamma^2 \left((\sigma^{\rm He}_{\rm sig})^2 + (\sigma^{\rm He}_{\rm bkg})^2\right). \tag{131}$$

### 9.2.2 INCLUSIVE EVENTS

Similar procedures have been followed to calculate the yields and the associated statistical uncertainties. Despite the fact that there are no accidental backgrounds in the case of the

inclusive analysis, one still needs to correct for the positron contamination and subtract the $^4$He contribution. The true number of $D(e,e')X$ DIS events in each bin is defined as:

$$N_{true}^{inc} = N_{el}^{D} - \gamma \cdot N_{el}^{He}, \tag{132}$$

Where,

$N_{el}^{D}$ is the number of the positron-corrected inclusive event from $D_2$ data,

$N_{el}^{He}$ is the number of the positron-corrected inclusive events from $^4$He data,

$\gamma$ is the $^4$He contamination correction factor.

And the statistical uncertainty as follows:

$$(\sigma_{\text{true}}^{inc})^2 = (\sigma_{\text{el}}^{D})^2 + \gamma^2(\sigma_{\text{el}}^{\text{He}})^2. \tag{133}$$

### 9.2.3 HE CONTAMINATION SUBTRACTION: ALTERNATIVE APPROACH

As explained in (129) and (132), we would correct the He contamination in our target during the experiment by subtracting the counts from the He target runs and using the He contamination factor $\gamma$. For He subtraction, similar to $D_2$ runs, we would bin $^4$He$(e,e'p_s)X$ DIS events in each kinematic bin, assuming it as $D_2$ runs, and then for each bin, we would subtract the corresponding measurement in $^4$He runs from $D_2$ runs. So, rather than subtracting $^4$He data bin by bin, we calculated accidental backgrounds and positron contamination corrected $^4$He counts in each $x^*$ bin for tagged data and $x$ bin for inclusive data. Then we extracted the ratio of $^4$He to $D_2$ in each bin. Finally, we fit this ratio and use the fit parameters to extract $^4$He contamination and subtract it. This would give much more precise contamination calculation across all bins, reducing the effect due to statistical fluctuation of events in each bin. Figs. 116 and 117 demonstrate the fitting procedure. With this approach, the He contamination subtracted counts will be,

$$N_{true}^{tag}(x^*,Q^2) = [N_{sig}^{D} - N_{bkg}^{D}] \cdot \text{CF}_{\text{He}}(x^*,Q^2) \qquad \text{Tagged,}$$

$$N_{true}^{inc}(x^*,Q^2) = N_{el}^{D} \cdot \text{CF}_{\text{He}}(x,Q^2) \qquad \text{Inclusive.}$$

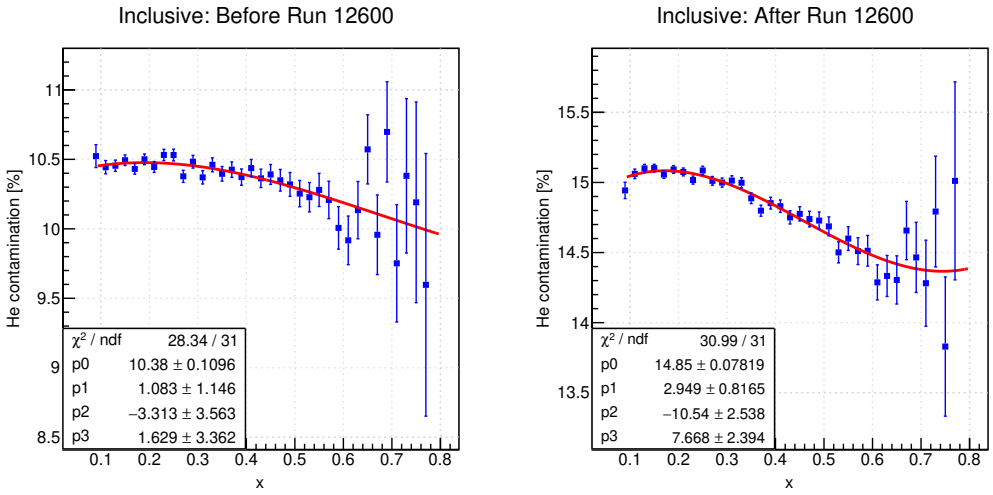



FIG. 116. He contamination percentage in the deuterium target for inclusive DIS data set for runs before 12600 (left) and for runs after 12600 (right).

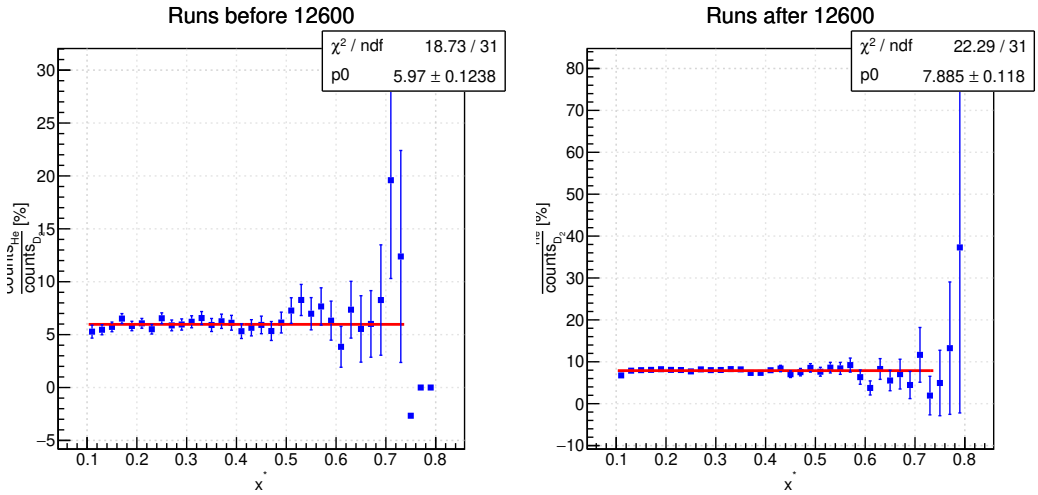



FIG. 117. He contamination percentage in the deuterium target for proton tagged DIS data set for runs before 12600 (left) and for runs after 12600 (right).

### 9.2.4 RATIO OF TAGGED TO INCLUSIVE EVENTS

For each bin in $x$ and $Q^2$, the yield ratio of tagged to inclusive events, both for experimental and simulated data, reads as:

$$R = \frac{N_{true}^{tag}}{N_{true}^{inc}}, \tag{134}$$

with its associated statistical uncertainty as:

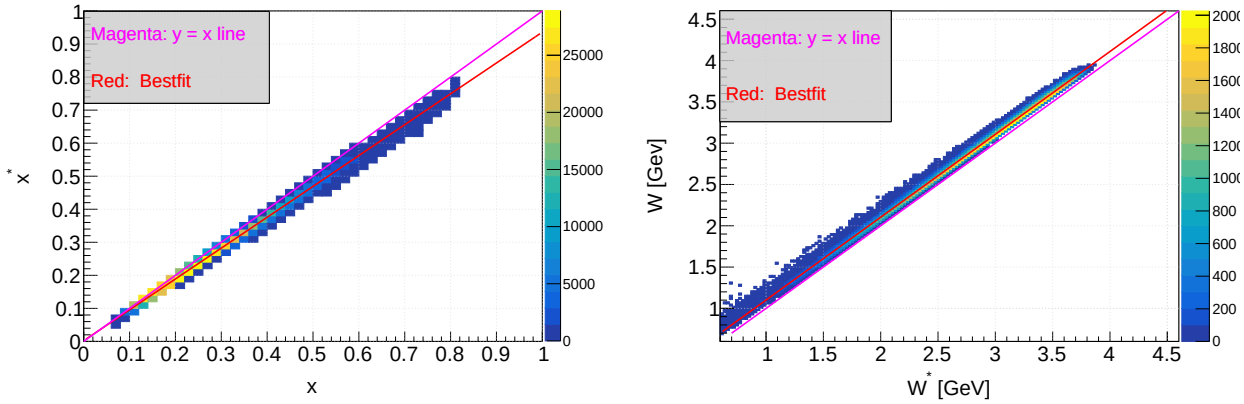
$$\sigma_R^2 = R^2 \left( \left( \frac{\sigma_{true}^{tag}}{N_{true}^{tag}} \right)^2 + \left( \frac{\sigma_{true}^{inc}}{N_{true}^{inc}} \right)^2 \right). \tag{135}$$

which is used to propagate the uncertainty in the final results for both, real and simulated data. Similarly, the same formula can be used to calculate the statistical uncertainty in the final Super Ratio (SR).

## 9.3 CLAS12 ACCEPTANCE MATCHING AND BIN FILTRATION

So far, we have explained that we will be binning our proton-tagged data in a 2D bin of $x^*$ and $Q^2$, and for the inclusive data, we will be binning in a 2D bin of $x$ and $Q^2$. We also applied a $W$ cut for inclusive events for the DIS event selection, whereas we modified the $W^*$ variable for the tagged events. The starred variables are calculated based on the kinematics of the spectator protons. They are calculated using the spectator's momentum. If the spectator proton has 4-momentum $P^\mu = (p_x, p_y, p_z, E_s)$, then, using the conservation law of 4-vectors, the 4-momentum of the initial struck nucleon is $n^\mu = (n_x, n_y, n_z, M_d - E_s)$.

$$\begin{aligned} x &= \frac{Q^2}{2m_p\nu}, \quad \text{where} \quad \nu = E - E', \\ x^* &= \frac{Q^2}{2n_\mu q^\mu}, \quad \text{where} \quad q^\mu = e^\mu - e'^\mu, \\ W &= \sqrt{m_p^2 + 2m_p\nu - Q^2}, \\ W^* &= \sqrt{n_\mu n^\mu + 2n_\mu q^\mu - Q^2}. \end{aligned} \tag{136}$$




(a) 2D distribution of $x^*$ vs. $x$.

(b) 2D distribution of $W$ vs. $W^*$.

FIG. 118. Comparison of kinematic variables between tagged and inclusive events.

From Eq. 136, it is evident that $x$ and $x^*$, as well as $W$ and $W^*$, are not identical. They differ from each other, as shown in Fig. 118a and 118b. These variables do not correspond to the same phase-space region or the same part of the CLAS12 detector for electrons. Consequently, if events occur near the detector's edges—where efficiency drops—acceptance and efficiency effects may come into play. The CLAS12 acceptance at the kinematical edges of the detector might, therefore, differ.

To mitigate potential discrepancies caused by this effect, we analyzed the tagged-to-inclusive event ratios for each kinematical bin and established two criteria for bin selection in the final analysis. These criteria were determined using simulated data. A bin is excluded from the final analysis if either of the following conditions below is met. See Figs. 119, 120, 122 and 122. After the filtration process

- The tagged-to-inclusive ratio is less than 0.4 [Green circle].
- The tagged-to-inclusive ratio deviates by more than 70 from the average value in that bin (after filtration using step 1) [Black Circle].
- The statistical uncertainty in the tagged-to-inclusive ratio exceeds 2.5 times the average value in that bin (after filtration using step 1) [Red circle].

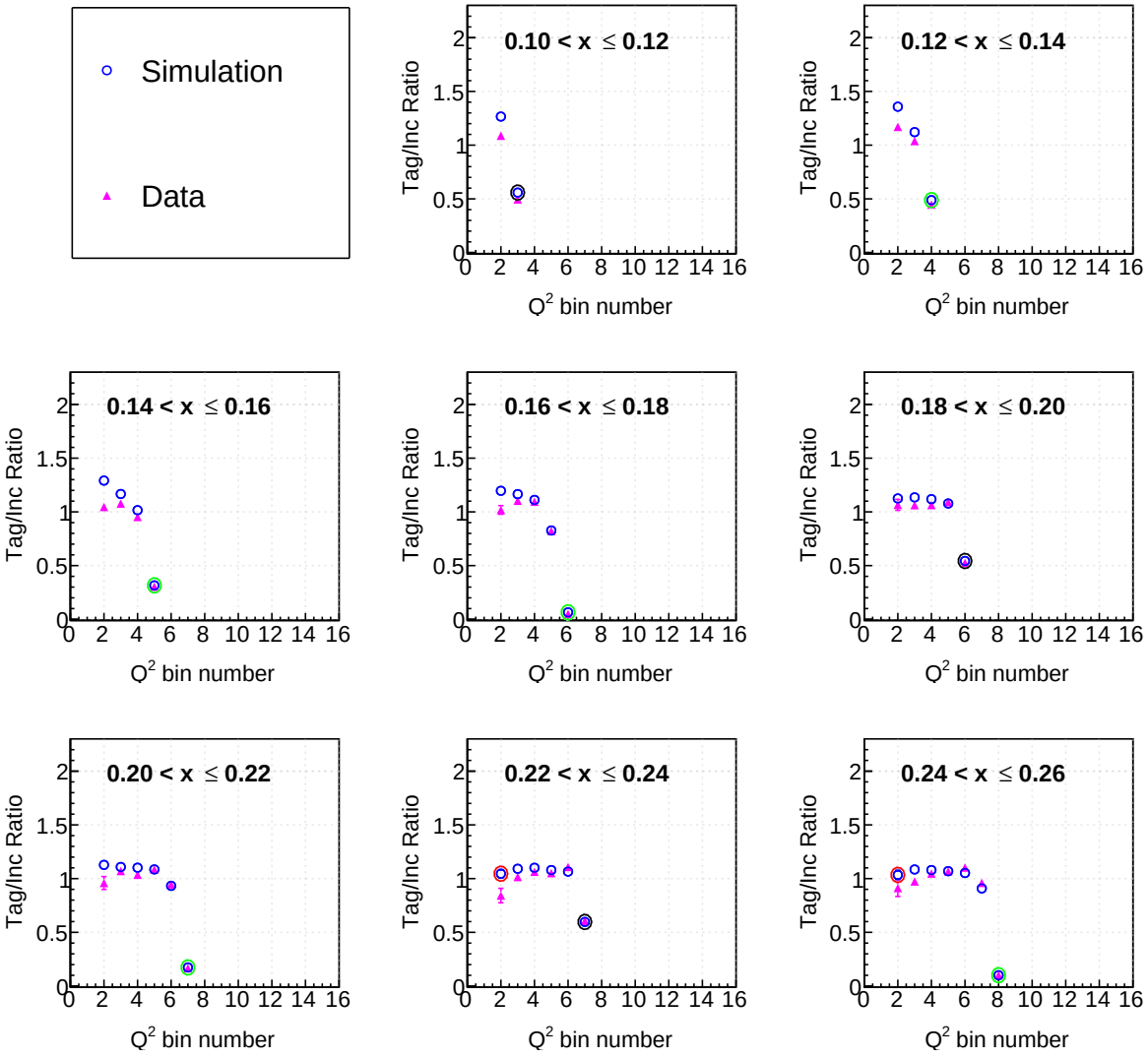



FIG. 119. Tag to inclusive ratio for both data and simulation (see legend) showing filtered bins.

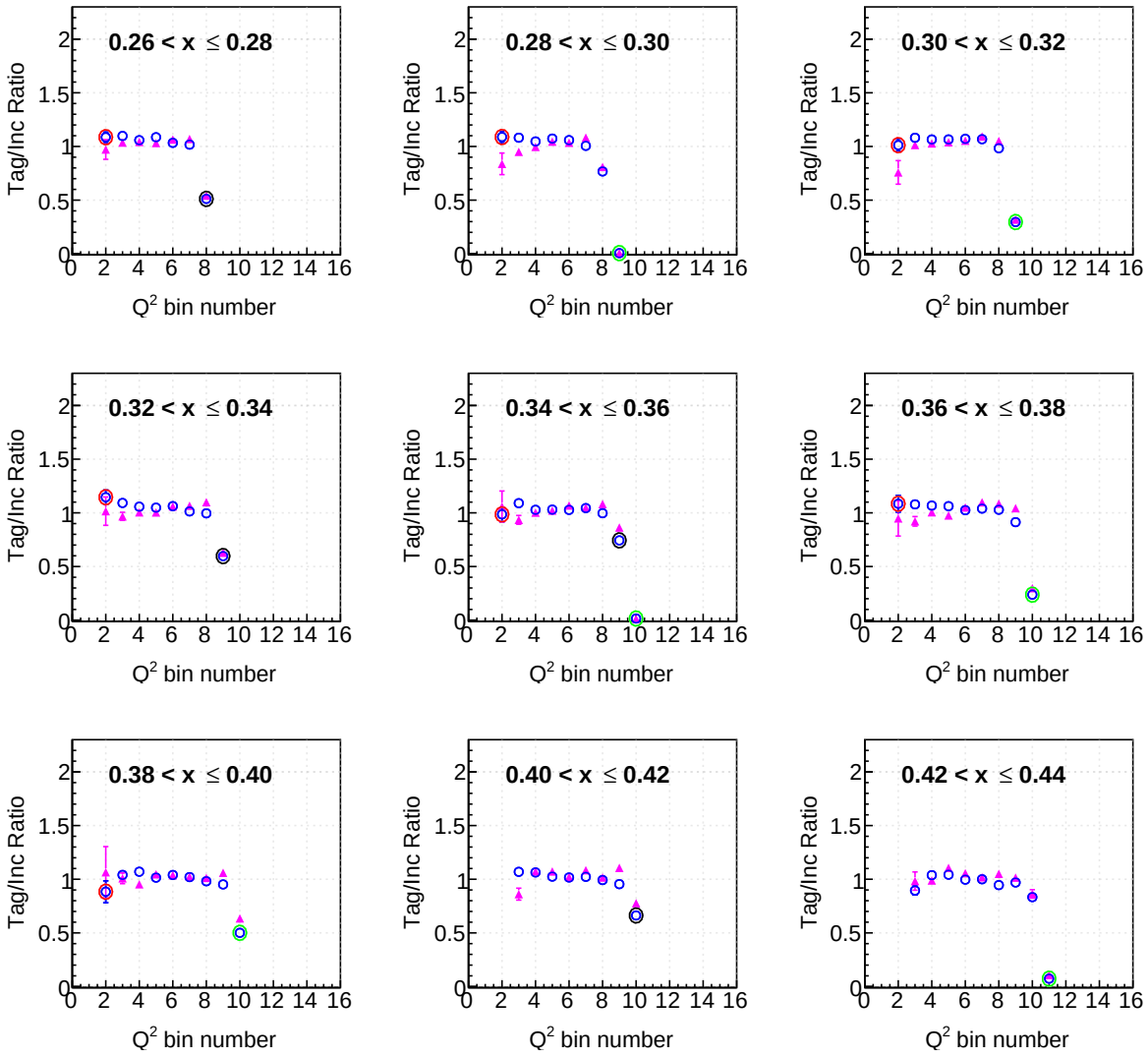



FIG. 120. Tag to inclusive ratio for both data and simulation showing filtered bins. Continuation of Fig. 119

.

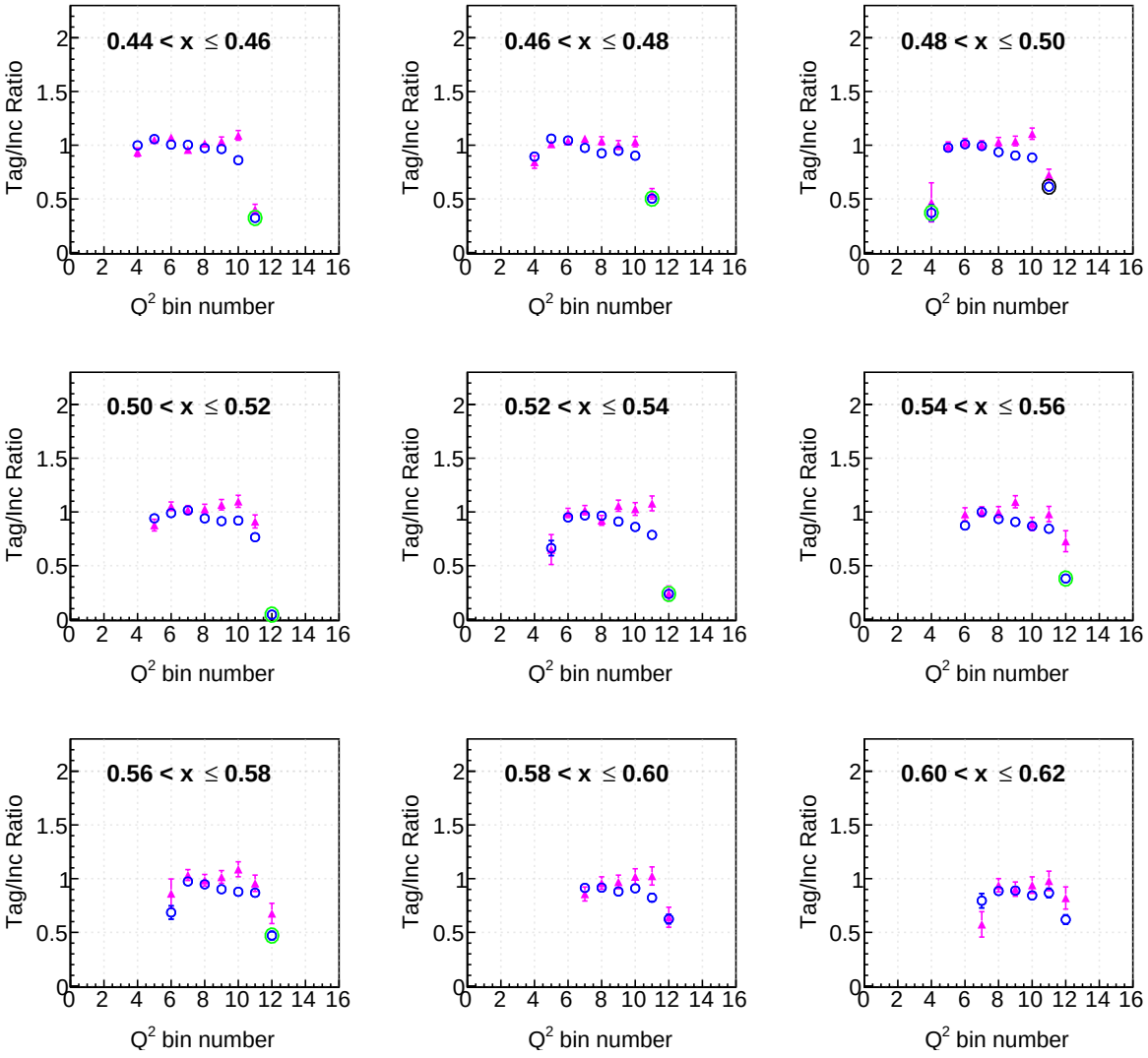



FIG. 121. Tag to inclusive ratio for both data and simulation showing filtered bins. Continuation of Fig. 120

.

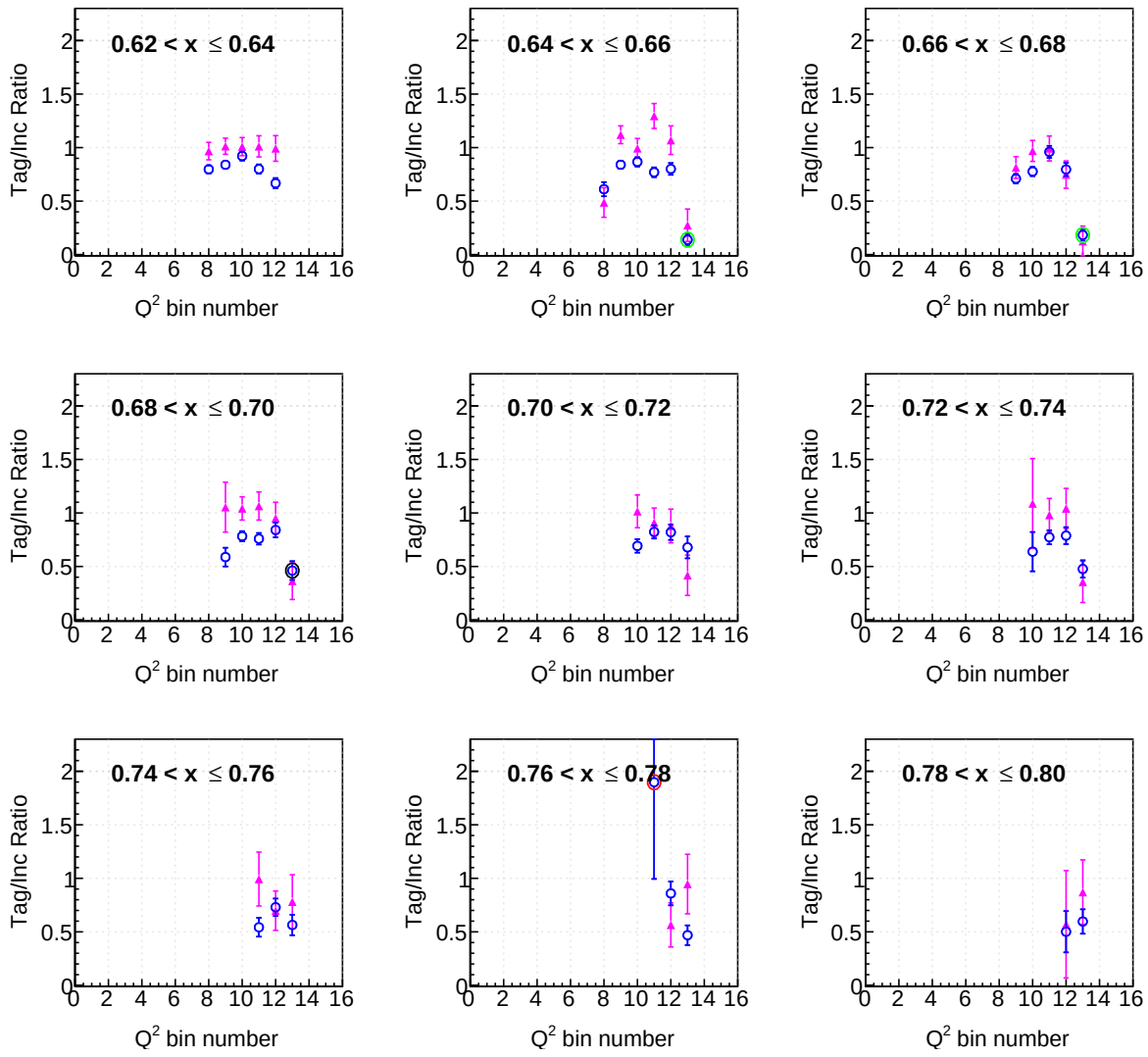



FIG. 122. Tag to inclusive ratio for both data and simulation showing filtered bins. Continuation of Fig. 121

.

## 9.4 MERGING EXPERIMENTAL DATASETS

The BONuS12 RTPC had an HV change during the Summer 2020 period, which affected the measured proton-tagged nDIS yield. This HV change was carried out around the middle of the Summer 2020 period, at run number 12600, for the purpose of reducing HV trips and

making the RTPC blind to the higher energy recoils and the minimum ionizing particles. Before we can merge those two halves of the experiment in a single dataset, we need to show that these two datasets are statistically compatible. The same selection criteria for the inclusive $(D(e,e')X)$ and proton-tagged $(D(e,e'p_s)X)$ DIS events have been carried out over the two datasets before and after the HV change. Then, after binning the data in the 2D bins of $x$ and $Q^2$ for the inclusive channel and $x^*$ and $Q^2$ for the tagged channel, the ratio of tagged to inclusive events in each $x/x^*$ bin, after integrating over $Q^2$, is calculated for data before and after the HV change. The ratios were cross-normalized to their values for $x/x^*$ in the range from 0.26 to 0.34. The two results are shown in Fig. 123. A so-called "t-test" has been applied to the two measured ratios to test the data compatibility before merging the two datasets. In this test, the $t$ variable is defined as:

$$t = \frac{r_1 - r_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}, \tag{137}$$

where $r_1$ and $r_2$ are the tagged to the inclusive yield ratios for the datasets before and after the HV change, respectively, with $\sigma_1^2$ and $\sigma_2^2$ are the variances in the ratios.

The "t-test" results between the points of the two measured ratios are presented in Fig. 124. Ideally, we would expect a normal distribution with a mean of 0 and a standard deviation of 1. From the figure, one sees a normal distribution with mean$\sim$ 0.003468 and $\sigma$ $\sim$ 1.059. Given that the standard error in the mean is defined as

$$\text{Standard Error} = \frac{\sigma}{\sqrt{n}} = \frac{1.059}{\sqrt{156}} = 0.084.$$

Hence, the mean is within 1 Standard Error of Mean (SEM) from the expected mean of 0. This shows there is not enough evidence to suggest a statistically significant difference between the two datasets. Hence, the results from the two sets of the experiment are compatible and can be safely merged together into one set. So, we combine the tagged to the inclusive ratio for two different halves of the experiment using the statistical weighting of the ratios. The combined tagged-to-inclusive ratio for the entire data set can be defined as,

$$r_{weighted} = \frac{r_1 \cdot \frac{1}{\sigma_1^2} + r_2 \cdot \frac{1}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}, \tag{138}$$

$$\sigma_{weighted} = \sqrt{\frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}}. \tag{139}$$

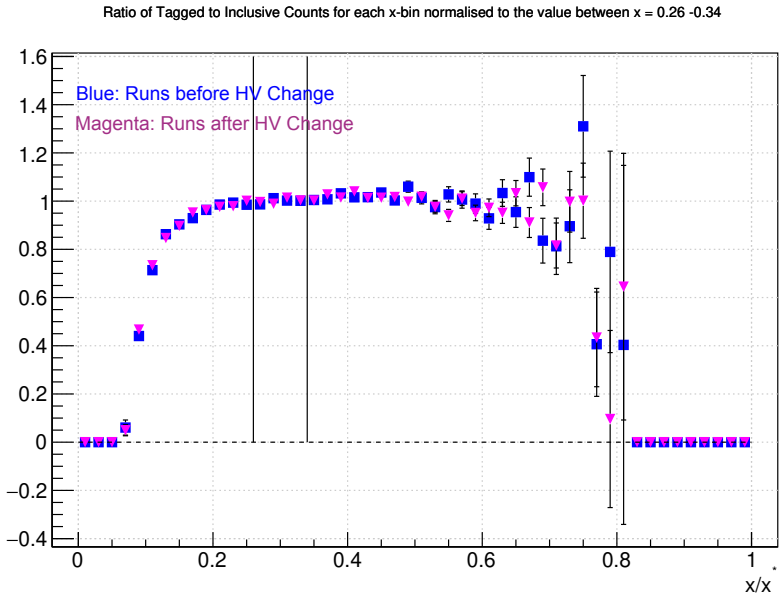



FIG. 123. The yield ratio of tagged to inclusive data as a function of $x/x^*$ for data before (in blue) and after (in magenta) the BONuS12 HV change. The black vertical lines indicate the range in $x$, 0.26 - 0.34, which is used to normalize the two ratios.

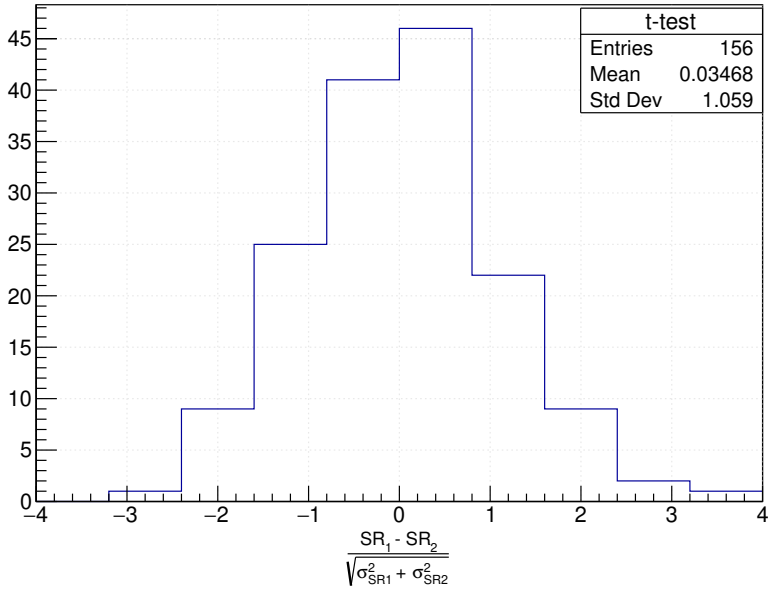



FIG. 124. The "t-test" of super ratio (tagged to inclusive yield ratios of data to MC) before and after the BONuS12 HV change.

## 9.5 FINAL RESULT

Finally, after applying the necessary electron selection cuts, fiducial cuts, and file selections, we made sure the data was of high quality. We then carefully selected RTPC tracks, and applied some corrections (e.g. momentum corrections), and file selections. Then, we also sampled accidental backgrounds and applied corrections for contamination to subtract the backgrounds.

With all these steps completed, we present the final extracted ratios $F_2^n/F_2^d$ in 4 different $Q^2$ bins [see Fig. 125]. Fig. 126 shows the same ratio integrated overall $Q^2$. Then we finally obtain $F_2^n/F_2^p$ using,

$$\left(\frac{F_2^n}{F_2^p}\right)^{\text{true}} = \left(\frac{F_{2n}}{F_{2d}}\right)^{\text{true}} * \left(\frac{F_{2d}}{F_{2p}}\right)^{\text{fit}}. \tag{140}$$

The final extracted $F_{2n}/F_{2p}$ ratios are the main result of this analysis. At low $Q^2$, we observe a slight deviation from the model. We have identified some issues with the simulation that might have caused this, and there is room for improvement, which is currently in progress. At very high $Q^2$, there is no accurate model available, so with this experimental data, we hope to provide important data points to theorists for future model improvements. The comparisons from different theoretical model extractions and experiments are laid out along with our extracted results in Fig. 127. We see that our result is in close agreement with the BONuS6 result, which was the previous predecessor of the BONuS12 experiment. It was also done at Hall B of Jefferson Lab around 2006. The beam energy available was much lower than it is with the BONuS12 experiment, so we see no data points above 0.6. The result from the MARATHON Experiment conducted at Hall C of Jefferson Lab is also presented for comparison. The normalization point for MARATHON and our BONuS12 experiment is different. Hence, we see a shift between the two datasets, but even if we scale our datasets for the same normalization, we see the tension between the two datasets. We are still investigating our final extracted results. So, we cannot comment on this yet. Similarly, the Jefferson Lab Angular Momentum (JAM) Collaboration result is also compared. The collaboration uses the world data from different experimental facilities on different kinematical regions and performs analysis to extract the result. Finally, When we extrapolated the $F_2^n/F_2^p$ ratio obtained in our experiment, it approached agreement with the DSE realistic model.

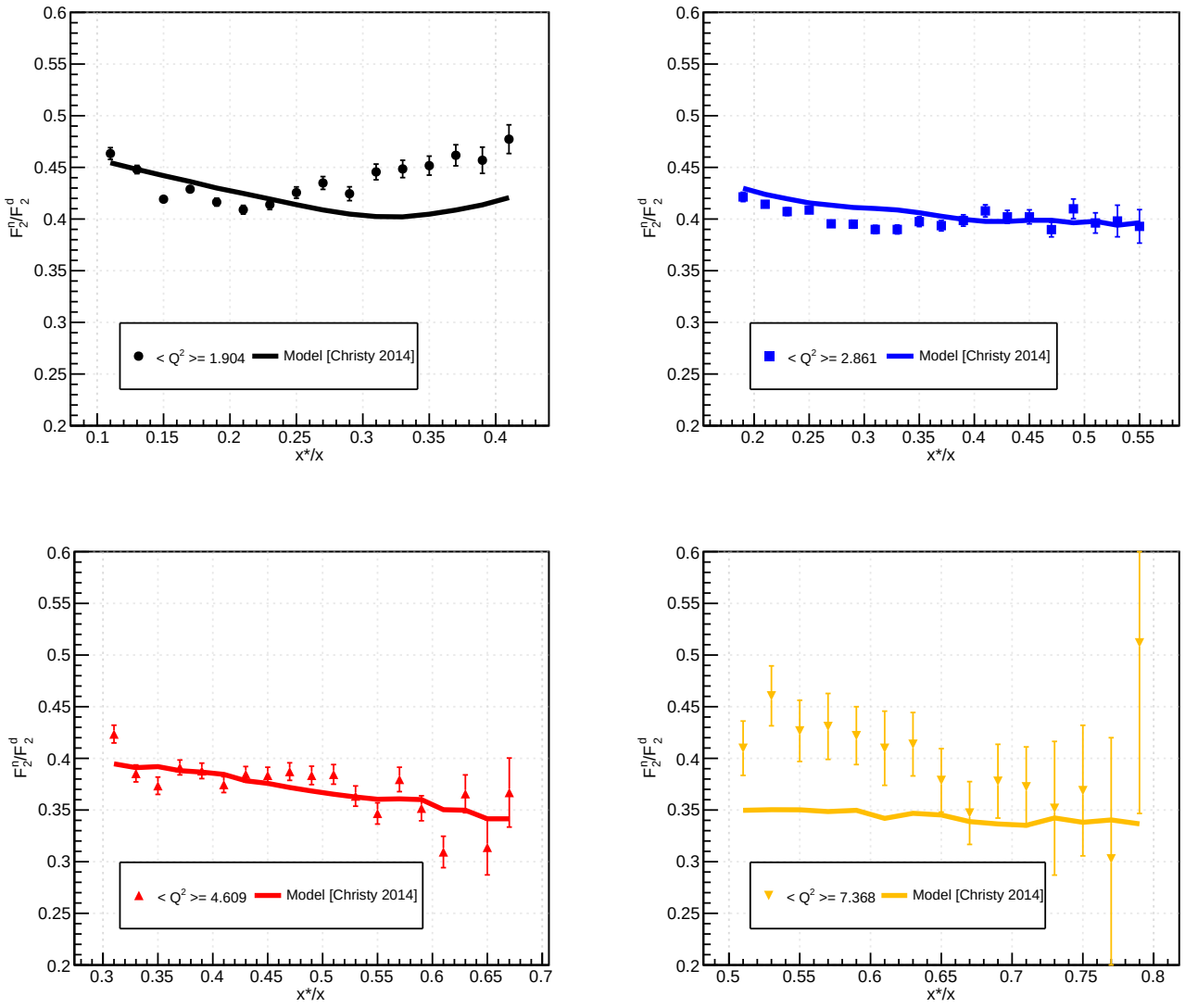



FIG. 125. The Structure Function ratios $F_2^n/F_2^d$ for four bins in $Q^2$.

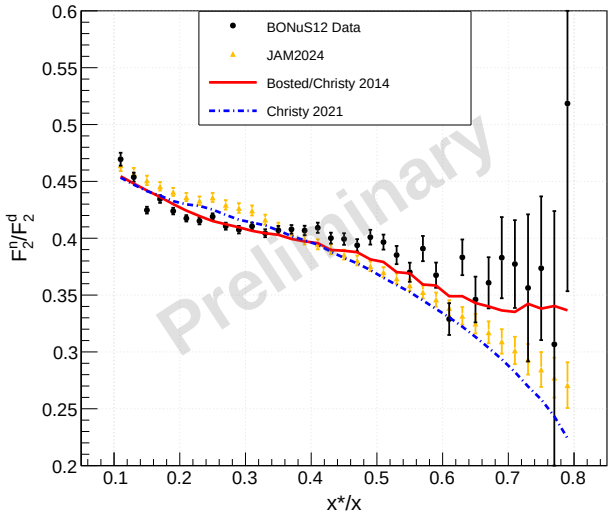



FIG. 126. The Structure Function ratios $F_2^n/F_2^d$ from BONuS12 experiment.

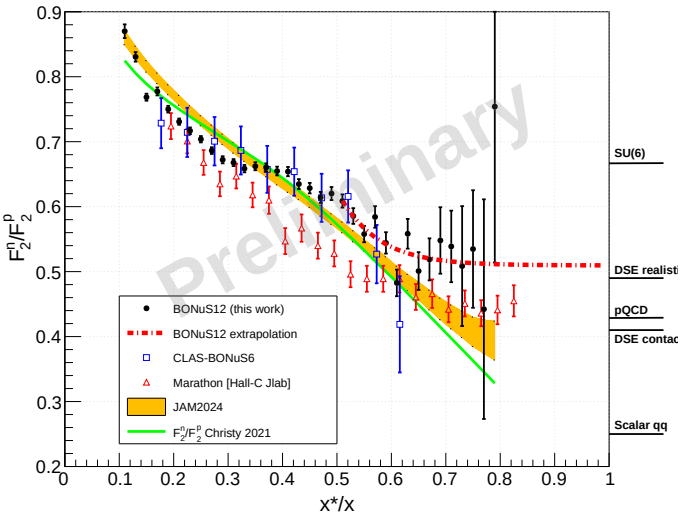



FIG. 127. The Structure Function ratios $F_2^n/F_2^p$ from BONuS12 experiment. Results from different analyses are presented in the legend. BONuS6 [20] and [21], MARATHON [40], and JAM [41].

## 9.6 CONCLUSION AND FUTURE PERSPECTIVE

The primary goal of the BONuS12 experiment was to extract the model-independent structure function ratio $F_2^n/F_2^p$. Since a free neutron target is not available, we took a deuterium target and used spectator tagging to effectively create a quasi-free neutron target. To minimize the effects of target fragmentation and final-state interactions (FSIs), we chose our kinematics region to low momentum spectators and relatively backward-moving spectators.

In addition to achieving our primary physics objectives, this experiment successfully designed, constructed, and demonstrated the use of a low-momentum GEM-based RTPC. This accomplishment may open new avenues in the development of low-momentum detectors for physics experiments and expand the applications of time projection chambers (TPCs).

Despite these achievements, there remain several areas for further study. One important direction is to explore different models of final-state interactions and target fragmentation to better understand their effects on the extracted structure function. Additionally, comparing the results using different event generator models as input will provide insights into the model dependencies and potential systematic uncertainties in the extraction of $F_2^n$. Furthermore, studying the dependence of $F_2^n$ on different spectator momenta $p_s$ and spectator angles $\theta_{pq}$ will help study the validity of the spectator model and the impacts of this on our measurements.

Another important check is to evaluate the effect of the deep inelastic scattering (DIS) cut on the extracted results and ensure that it does not introduce biases. Additionally, investigating whether there is any remaining resonance contribution for $W^* > 1.8$ will be crucial in confirming the validity of the DIS region selection.

As mentioned at the beginning of this thesis, the analysis presented here is based on data from the Summer 2020 run. However, we still have results to analyze from the Spring 2020 run, which will provide further insights and cross-checks. Additionally, ongoing improvements in the simulated data analysis will further refine our understanding and improve our extracted results. With these developments, we expect to achieve even more precise results for publication.

Along with the upcoming Electron-Ion Collider (EIC) and planned Jefferson Lab upgrades to 22 GeV beam energy in the near future, there will be new opportunities to extend our understanding of the neutron structure function at even higher $x \sim 0.9$. These future experiments will provide valuable insights into nucleon structure in the high-$x$ region, further testing our current models and improving our understanding of non-perturbative QCD effects.

Finally, the RTPC technology demonstrated in this experiment could be further optimized and explored for broader applications in nuclear and particle physics.

# CHAPTER 10

# SYSTEMATIC UNCERTAINTIES

Systematic uncertainties are the possible bias on the extracted results that might be added due to some sources during the data taking, data reconstruction, or during the physics analysis phase. As one cannot be sure if the extracted results have biases, it is still possible to associate additional uncertainties with the results that might emerge from systematic sources. In the analysis presented in this dissertation, the usage of the yield ratios as the observable minimizes the systematic uncertainties that are associated with the normalization and the inefficiencies that appear in the cross-sections but cancel out in the ratios. However, some sources may still affect the results systematically. In this section, the remaining possible systematic resources have been studied and their associated uncertainties on the extracted super-ratios have been evaluated.

## 10.1 SYSTEMATIC UNCERTAINTY QUANTIFICATION

For RGF, we measure the ratio of the tag to inclusive events in a 2D bin of $Q^2$ and $x/x^*$. To calculate the ratio, we apply a standard set of cuts on certain track quality and kinematic phase space to remove contamination and backgrounds. To quantify systematic uncertainty in the ratios, we vary specific cuts and observe how the ratio changes relative to standard cuts. For example, to study the systematic uncertainty introduced by $\Delta vz$ cut, we proceed as follows:

The standard nominal $\Delta v_z$ cut for RGF is:

$$\Delta v_z = \langle v_z \rangle \pm 2\sigma.$$

We call this ratio $R_2$. We then change this cut to $2.5\sigma$ to study the systematics induced by the $\Delta vz$ cut, and the new ratio is $R_{2.5}$. A wider cut of $2.5\sigma$ increases the number of events passing this cut. This cut affects only tagged event selection. Although the number of accidental backgrounds increases, we assume the fraction of background under the signal peak remains unchanged. The background subtraction procedure should adequately account

for this. Finally, we normalize the ratio to 1, ensuring the change in event counts is managed. After applying the new cut, we define:

$$\Delta R(x_i, Q_i^2) = R_{2.5} - R_2, \tag{141}$$

$$\sigma_{\Delta R}(x_i, Q_i^2) = \sqrt{\sigma_{R_{2.5}}^2 + \sigma_{R_2}^2}. \tag{142}$$

Here, $\Delta R(x_i, Q_i^2)$ represents the change in ratios, while $\sigma_{\Delta R}(x_i, Q_i^2)$ gives the statistical uncertainty in $\Delta R$. We aim to estimate the percentage of change attributed to statistical fluctuation. Systematic uncertainty for each $(x_i, Q_i^2)$ bin is defined as:

$$\text{Systematic Uncertainty}(x_i, Q_i^2)[\%] = \frac{\Delta R(x_i, Q_i^2)}{R_2(x_i, Q_i^2)} \times 100. \tag{143}$$

Now to calculate the total systematic uncertainity we weighted each bin by the square of inverse of their statistical uncertainity (inverse variance weighting).

$$\text{Total Systematic Uncertainty}[\%] = \frac{\sum_i \Delta r_i(x_i, Q_i^2) \cdot \frac{1}{\sigma_i^2(x_i, Q_i^2)}}{\sum_i \frac{1}{\sigma_i^2(x_i, Q_i^2)}} \times 100 \tag{144}$$

For combining all $x$ and $Q^2$ bins, we average out this percentage difference over all bins and present the result,

**Studies done using Experimental Data**

- Particle misidentification [ $\pi^-$ misidentified as electron].
- Pair-Symmetric Background Correction.
- Goodness of track reconstruction in RTPC.
- Accidental Backgrounds in RTPC.
- Acceptance of RTPC along the length of the detector.
- He-contamination correction.

### 10.1.1 MISIDENTIFIED PIONS

This source is associated with negative pions that have been misidentified as the trigger electron in an event and passed all the fiducial, energy, and the DIS selection cuts. In order

to study this source, one would check the PID distribution for the negative first particles in the reconstructed events, which is shown in Fig. 128. One approach to find how many pions have been misidentified as an electron is by checking how many pions have been identified as pions (PID =-211) and passed all the selection cuts of an electron. Fig. 129 presents the number pf photo-electron (*nphe*) distribution for the the $\pi^-$ reconstructed particles at different stages of the selection cuts (see the captions for details on the cuts). From this latter figure, one can see that less than 1% of the negative reconstructed pions have passed the selection cuts for the DIS electrons. Considering the pion to the electron multiplicity from Fig. 128, which is less than 0.1%, one can give a conservative 0.1% as a upper estimate to the misidentified pions as electrons in our sample.

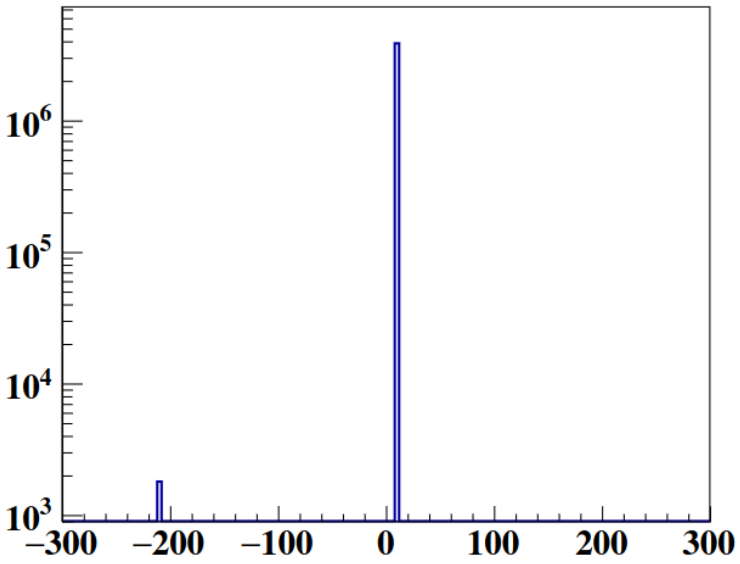



FIG. 128. PID for negatively charged and first particle in an event

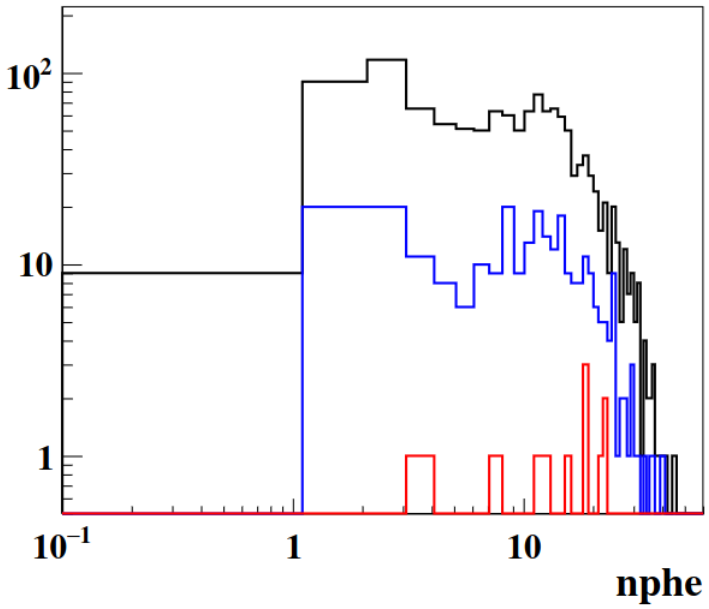



FIG. 129. Distributions of the number of photo-electrons (*nphe*) produced by the $\pi^-$ particles at different stages of selection. In black: after selecting PID =-211 and being the first particle in an event. In blue: after adding $v_z$ and fiducial cuts. In red: after applying the sampling fraction and energy cuts.

### 10.1.2 PAIR-SYMMETRIC BACKGROUND CORRECTION

This source is associated with the bin-by-bin correction factor for the electrons that originated from an electron-positron annihilation pair rather than an incident electron that scattered from the bound neutron. To evaluate the systematic error that is introduced by this correction, the bin-by-bin correction factor was doubled, and the change in the tagged to the inclusive ratio induced by this change was adopted as the systematic uncertainty from this source.

### 10.1.3 DELTA-VZ SYSTEMATICS

For the BONuS12 experiment, we have applied a sector-dependent delta-vz cut of 2 $\sigma$ around the mean value (see chapter 5). To study the systematic uncertainty introduced by this cut, the cut was changed to 2.5 $\sigma$, and yield ratios were compared. The percentage

difference between the two calculated ratios is considered as the associated systematic error that emerged from this cut.

### 10.1.4 TDIFF SYSTEMATICS

Similarly, $t_{diff}$ file-dependent cut was 1.9 $\sigma$ around the mean. This cut was changed to 2.5 $\sigma$ to evaluate the systematic uncertainty that is associated with this cut.

### 10.1.5 RTPC ACCEPTANCE

To study the systematic uncertainty that emerged from the fiducial cut along the RTPC, the $z$-vertex cut was changed from (-210, 180) mm to (-180, 150) mm. The percentage difference between the two ratios is considered as the associated systematic uncertainty.

### 10.1.6 HELIUM CONTAMINATION CORRECTION

A correction factor was introduced to correct for the $^4$He contamination in the $D_2$ target based on the total number of tritons/Helium-3 in $D_2$ runs compared to $^4$He runs. To study the systematic effect introduced by this factor, the analysis has been repeated using a 50% increase in this contamination factor. We implemented this He contamination factor for both inclusive and tagged data. The percentage difference between the two ratios is considered as the emerging systematic uncertainty from this correction.

### 10.1.7 NORMALIZATION

The most direct result of our analysis is the ratio of neutron to deuteron structure functions, $F_2^n/F_2^d$. As explained before, we do not have a sufficiently precise determination of the reconstruction efficiency of the RTPC for spectator protons in the VIP kinematic range to determine this ratio with absolute normalization. (We also do not know the efficiency- and acceptance-weighted integral over the deuteron wave function with sufficient precision). Therefore, we arbitrarily normalize our results to the ratio $F_2^n/F_2^d = 0.419$ determined from world data and averaged over the interval $0.28 < x < 0.32$. The choice of this normalization region is driven by the expectation (based on existing nuclear data) that the EMC ratio in this region is very close to 1, i.e., nuclear corrections are minimal. Furthermore, we compared three different parametrizations of the world data (Bosted/Christy 2014; Christy 2023, and JAM) and found they agree to within 1.3% with each other. For this reason, we assume an overall normalization uncertainty of 1.3%. Obviously, our results can be compared also

based on different parametrizations by scaling them with $F_2^n/F_2^d(x = 0.3)/0.419$.

As final result of our analysis we also present the ratio $F_2^n/F_2^p$ which is more directly related to ratio of $d$ over $u$ quark PDFs. To convert to this quantity, we multiply our data with the ratio $F_2^d/F_2^p$, again taken from parametrizations of world data. We note that this ratio is not dependent on any treatment of nuclear effects (at least in principle), since both quantities are directly measured. While the three parametrizations agree within 1.7% with each other at $x = 0.3$, they deviate more at higher $x$, up to 4.4% at $x = 0.75$. We, therefore, assume an additional normalization uncertainty of the form $\sigma = 0.5\% + x * 5\%$.

### 10.1.8 THEORETICAL UNCERTAINTY

Our entire method of extracting ratios of $F_2^n$ to other structure functions rests on the assumption that the tagged events we observe contain pure spectator protons only. We explain in the introduction why this is a reasonably good approximation within the "VIP" kinematic range we select for protons in the RTPC, but there are obviously potential deviations from this picture, in particular, due to in-medium modifications of the struck neutron and final state interaction between it and the "spectator" proton. Instead of trying to quantify these uncertainties here, we prefer to quote our results as they are and leave it to future theoretical analyses to quantify (and potentially correct for) these effects.

| **Source of Systematics** | **Systematic Uncertainty** |
|---|---|
| Pion Contamination | 0.1 % |
| Pair Symmetric background | 0.8 % |
| Deltavz cut | 1.9 % |
| $T_{diff}$ cut | 1.6 % |
| RTPC acceptance | 1.6 % |
| Target Contamination | 0.1 % |
| Overall Normalization | 0.5 % + x*5 % |

TABLE 5. Summary of Systematic Uncertainties

# REFERENCES


[1] H. F. Coward, Journal of Chemical Education **4**, 23 (1927).

[2] E. Rutherford, Philosophical Magazine **21**, 669 (1911).

[3] J. Chadwick, Nature **129**, 312 (1932).

[4] J. D. Cockcroft and E. T. S. Walton, Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences **136**, 619 (1932).

[5] R. Hofstadter, Rev. Mod. Phys. **28**, 214 (1956).

[6] M. Gell-Mann, Physics Letters **8**, 214 (1964).

[7] G. Zweig, *An SU(3) model for strong interaction symmetry and its breaking. version 2*, tech. rep. 8419/TH.412 (CERN, 1964).

[8] Wikipedia contributors, *Standard model*, Accessed: 2024-07-11, 2024.

[9] G. Aad et al., Phys. Rev. Lett. **114** (2015).

[10] M. N. Rosenbluth, Phys. Rev. **79**, 615 (1950).

[11] C. Perdrisat, V. Punjabi, and M. Vanderhaeghen, Progress in Particle and Nuclear Physics **59**, 694 (2007).

[12] J. D. Bjorken, Phys. Rev. **179**, 1547 (1969).

[13] G. Altarelli and G. Parisi, Nuclear Physics B **126**, 298 (1977).

[14] V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15**, 438 (1972).

[15] Y. L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977).

[16] J. GAYLER, "Proton structure function measurements from hera", in *Multiparticle dynamics* (), pp. 274–279.

[17] F. M. Gonzalez et al., Phys. Rev. Lett. **127**, 162501 (2021).

[18] A. Accardi et al., Phys. Rev. D **93**, 114017 (2016).

[19] V. Palli et al., Phys. Rev. C **80**, 054610 (2009).

[20] S. Tkachenko et al., Phys. Rev. C **89**, 045206 (2014).

[21] I. Niculescu et al., Phys. Rev. C **91**, 055206 (2015).

[22] C. W. Leemann, D. R. Douglas, and G. A. Krafft, Annual Review of Nuclear and Particle Science **51**, 413 (2001).

[23] B. Mecking and et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **503**, 513 (2003).

[24] V. Burkert and et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **959**, 163419 (2020).

[25] R. Fair et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **962**, 163578 (2020).

[26] M. Mestayer et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **959**, 163518 (2020).

[27] M. Ungaro et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **957**, 163420 (2020).

[28] D. Carman et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **960**, 163629 (2020).

[29] D. Carman et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **960**, 163626 (2020).

[30] G. Asryan et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **959**, 163425 (2020).

[31] P. Chatagnon et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **959**, 163441 (2020).

[32] D. R. Nygren, *The time-projection chamber: a new 4. pi. detector for charged particles*, tech. rep. SLAC-PEP-NOTE-144; ON: DE88014821; TRN: 88-030449 (Stanford Linear Accelerator Center, Menlo Park, CA (USA); Lawrence Berkeley Lab., CA (USA), Jan. 1974).

[33] F. Sauli, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **805**, Special Issue in memory of Glenn F. Knoll, 2 (2016).

[34] H. Fenker et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **592**, 273 (2008).

[35] V. Ziegler et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **959**, 163472 (2020).

[36] Jefferson Lab CLAS12 Collaboration, *CLAS12 EventBuilder*, `https://clasweb.jlab.org/wiki/index.php/CLAS12_EventBuilder`, Accessed: 2024-10-27.

[37] I. Albayrak et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **1062**, 169190 (2024).

[38] R. B. Wiringa, V. G. J. Stoks, and R. Schiavilla, Phys. Rev. C **51**, 38 (1995).

[39] P. E. Bosted and M. E. Christy, Phys. Rev. C **77**, 065206 (2008).

[40] D. Abrams et al., Phys. Rev. Lett. **128**, 132003 (2022).

[41] C. Cocuzza et al., Phys. Rev. Lett. **127**, 242001 (2021).

# APPENDIX

## FIDUCIAL CUTS

### DC FIDUCIAL CUT

The fiducial cut for the Drift Chamber (DC) is defined by a sector-dependent set of constraints on the azimuthal angle $\phi$ as a function of the polar angle $\theta$. These constraints ensure that the selected tracks fall within the detector's acceptance regions. For a given sector, the fiducial cut limits are determined using the following parametric equations,

$$\phi_{\min} = p_0 + p_1 \ln(\theta) + p_2\theta + p_3\theta^2, \tag{145}$$

$$\phi_{\max} = q_0 + q_1 \ln(\theta) + q_2\theta + q_3\theta^2, \tag{146}$$

where:

- $\phi_{\min}$ and $\phi_{\max}$ are the minimum and maximum limits for the azimuthal angle $\phi$, respectively in the DC local coordinate system.
- $\theta$ is the polar angle of the track in the DC local coordinate system.
- $p_0, p_1, p_2, p_3$ and $q_0, q_1, q_2, q_3$ are sector-dependent parameters that define the shape of the fiducial region.

#### Sector-Dependent Parameters

The values of $p_0, p_1, p_2, p_3$ and $q_0, q_1, q_2, q_3$ for each sector are given in Table 6.

| Sector | $p_0$ | $p_1$ | $p_2$ | $p_3$ | $q_0$ | $q_1$ | $q_2$ | $q_3$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 53.82 | -41.17 | 2.786 | -0.02561 | -56.85 | 42.17 | -2.82 | 0.02603 |
| 2 | 100.02 | -33.13 | 2.135 | -0.01938 | 3.773 | 41.85 | -2.802 | 0.0259 |
| 3 | 175.1 | -41.59 | 2.8 | -0.02575 | 74.24 | 34.97 | -2.172 | 0.01951 |
| 4 | 232.6 | -36.89 | 2.089 | -0.01731 | -230.9 | 38.05 | -2.41 | 0.02157 |
| 5 | -71.26 | -37.46 | 2.413 | -0.02162 | -176.7 | 41.76 | -2.756 | 0.02531 |
| 6 | -16.49 | -34.21 | 2.079 | -0.01754 | -113.4 | 39.56 | -2.535 | 0.02277 |

TABLE 6. DC Fiducial cut parameters for each sector.

## PCAL FIDUCIAL CUT

The following table presents the fiducial cut thresholds for the six sectors of the detector. Each sector has corresponding minimum values for $u$, $v$, and $w$.These values are used in the fiducial cut condition to ensure valid hits in the PCAL detector.

| Sector | $u_{\min}$ | $v_{\min}$ | $w_{\min}$ |
|---|---|---|---|
| 1 | 33 | 13 | 7 |
| 2 | 27 | 8 | 14 |
| 3 | 37 | 14 | 14 |
| 4 | 17 | 14 | 6 |
| 5 | 32 | 14 | 13 |
| 6 | 22 | 8 | 6 |

## MOMENTUM-DEPENDENT SAMPLING FRACTION CUT

The momentum-dependent sampling fraction (SF) cut is applied using the following parameters for each sector. The parameters for the mean (mean($p$)) and standard deviation ($\sigma(p)$) are given as polynomial coefficients, which are used to evaluate the mean and sigma as functions of the momentum $p$.

**Mean and Standard Deviation Parameters**

The following tables provide the polynomial coefficients for the mean and standard deviation of the sampling fraction for each sector.

**Mean Parameters (mean$(p)$)**

$$\text{mean}(p) = a_0 + a_1 p + a_2 p^2, \tag{147}$$

| Sector | $a_0$ | $a_1$ | $a_2$ |
|---|---|---|---|
| 1 | 0.213784 | 0.00722088 | -0.000725496 |
| 2 | 0.216598 | 0.00732737 | -0.000684923 |
| 3 | 0.213267 | 0.0092289 | -0.000951656 |
| 4 | 0.230588 | 0.0054414 | -0.00052015 |
| 5 | 0.215682 | 0.00698124 | -0.000641261 |
| 6 | 0.220351 | 0.00653102 | -0.000698198 |

TABLE 7. Polynomial coefficients for the mean parameters by sector.

**Sigma Parameters ($\sigma(p)$)**

$$\sigma(p) = b_0 + b_1 p + b_2 p^2. \tag{148}$$

| Sector | $b_0$ | $b_1$ | $b_2$ |
|---|---|---|---|
| 1 | 0.0199171 | -0.000712994 | 1.57644e-05 |
| 2 | 0.0193179 | -0.00014262 | -5.88923e-05 |
| 3 | 0.0199035 | -0.00125284 | 5.1824e-05 |
| 4 | 0.0253942 | -0.00291724 | 0.000173601 |
| 5 | 0.0217675 | -0.00220587 | 0.000176306 |
| 6 | 0.0208277 | -0.00148967 | 5.84997e-05 |

TABLE 8. Polynomial coefficients for the sigma parameters by sector.

**SF Cut Definition** The momentum-dependent sampling fraction cut is applied using the mean and sigma parameters for each sector, as described in the previous section. The cut is defined as follows:

$$\text{Lowerbound} = \text{mean}(p) - 3 \cdot \sigma(p), \tag{149}$$

$$\text{Upperbound} = \text{mean}(p) + 3 \cdot \sigma(p). \tag{150}$$

An event passes the cut if the sampling fraction $SF$ lies between these bounds:

$$\text{Sampling Fraction cut} = (\text{SF} \geq \text{Lowerbound}) \quad \text{and} \quad (\text{SF} \leq \text{Upperbound}). \tag{151}$$

# VITA

Madhusudhan Pokhrel
Department of Physics
Old Dominion University
Norfolk, VA 23529

## EDUCATION

2017-2025 PhD. Candidate (Physics), Old Dominion University, Norfolk, Virginia, USA.
2019 Master of Science (Physics), Old Dominion University, Norfolk, Virginia, USA.
2016 Master of Science (Physics), Tribhuvan University, Kathmandu, Nepal.
2013 Bachelor of Science (Physics), Tribhuvan University, Kathmandu, Nepal.

## AWARDS AND SCHOLARSHIPS

1. Brian Donald Diederich and Flavia Alexandra Osorhean Scholarship for Doctoral Research in Nuclear Physics, Old Dominion University, March 2023.
2. Dominion Scholarships, Old Dominion University, Jan 2025.

Typeset using LaTeX.