The Extraction of the $\bar{d}/\bar{u}$ Ratio in Nuclear Media

by

Marshall Scott

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Applied Physics)
in The University of Michigan
2020

Doctoral Committee:

Professor Wolfgang Lorenzon, Chair
Professor Christine Aidala
Professor Jianming Qian
Professor Gregory Tarle

Marshall Scott

scottmar@umich.edu

ORCID iD:0000-0003-1105-1033

Acknowledgments

First and foremost I would like to thank my parents, Billy Scott and Alyce Coffey, and my brother Langston Scott for always being there for me and supporting me throughout my life. Words cannot describe how fortunate I am to have these three pillars in my life. I would also like to acknowledge my extended family. A heartfelt thank you goes out to Tristan Geis and Thu Huynh for your love, kindness, and for being my home away from home for all of these years. I would also like to thank Rachel Moss, Liz Lindly, Daniel and Laurelyn Leimer, Kathleen Stafford, Alan McCray, and Micheal and Lauren Baird for your deep, loving friendship. I don't know where I would be without Erika Ellis, Ameé "A. J. Maloy" Hennig, Jazmin Berlanga Medina, and Amy Ireton. Your insight, thoughtfulness, and support has often been the light at the end of a very dark tunnel. I would also like to thank Jerry and Leete Kendrick for always making me feel like family and Anastasia Rumsey and Leta Woodruff for the warmth and love that you have shown me over the years. A great thanks goes to Anthony Della Pella and the burger group: Will and Jade Clark, Jonathan Guzman, Jay Barraza, Jenia Rousseva, and Jeremy Waters, for all laughs and good times in graduate school.

A special thanks goes to the Michigan group, especially Wolfgang Lorenzon, whose mentorship, guidance, patience, and understanding has been indispensable throughout my graduate program. I would also like to thank Daniel Morton, who has personally mentored and helped me become the researcher I am today, and Maris Arthurs for his boundless kindness and friendship.

It goes without mentioning that I would not be where I am without the tireless efforts of the Applied Physics department. I am forever indebted to Cagliyan Kurdak, Cynthia McNabb, and

Lauren Segall for always making me feel welcome, valued, and being there to help with navigating grad school.

I am deeply indebted to the SeaQuest collaboration. To Bryan Ramson, Paul E. Reimer, Donald Geesaman, Charles "Chuck" Brown, Yen-Chu "Andrew" Chen, and Kenichi Nakano I offer sincere gratitude for always being open for discussions and assistance, and for helping become a better researcher. Finally, I would like to thank Arun Tadepalli for always being open for anything from an idle chat to an in depth discussion. You bring a depth of understanding, guidance, and humor that has been inestimable.

Last, but certainly not least, I would like to thank Gay and John Stewart, and Daniel Kennefick for helping me start this journey to my doctorate by fostering my undergraduate physics growth. I would not be here today without the physics I studied in Gay and John's classes, the guidance Gay provided throughout my undergraduate career, and the research I did with Daniel. I would like to close these acknowledgements by thinking my thesis committee, Wolfgang Lorenzon, Christine Aidala, Jianming Qian, and Gregory Tarle, for helping me finish this journey.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

This work examines the momentum distribution of the "sea" quarks within the proton and how these distributions are modified in the nuclear medium. The nucleon sea, which is made of gluons and quark-antiquark pairs, possesses an excess of anti-down quarks to anti-up quarks that is not readily explained by perturbative QCD. The distributions of the antiquarks are probed through the Drell-Yan process using the collision of 120 GeV protons on stationary liquid and solid nuclear targets. The cross sections are then derived by extracting the distributions with analysis constraints and machine learning tools to remove charmonium and combinatoric backgrounds. A cross section ratio of the liquid deuterium to liquid hydrogen is used to extract the free proton asymmetry. The modified free proton asymmetry is then used in conjunction with the nuclear to liquid hydrogen cross section ratios to extract the bound proton asymmetry ratio for carbon, iron, and tungsten. In addition to the asymmetry ratio extraction, the nuclear dependence of the Drell-Yan cross section ratio and a comparison between the standard analysis method and machine learning is explored.

Chapter 1:

Introduction

The idea that matter is ultimately composed of atoms, or *atomos*, meaning indivisible in Greek, has been an important notion since the time of the ancient Greeks and Indians. Particle physics is tasked with probing the structure of matter to ever smaller dimensions to arrive at the "true" *atomos* of nature. Until the 1800s Atomism was mostly a philosophical theory, but in 1805, John Dalton proposed that all substances consisted of a combination of different elements, with each element being composed of a single type of atom. This proposition explained the relationships between the weights of substances in chemical reactions [1]. In 1827, Robert Brown, observing the chaotic motion of pollen grains suspended in water, discovered the eponymous Brownian motion. Einstein later described this motion as the interaction of atoms with the grain and Jean Baptiste Perrin used this work to show that Brownian motion is directly related to the kinetic theory of gases. This concordance between the average energy of the grains and the average energy predicted by the kinetic theory furthered the claim that atoms are real [2].

In 1897 the electron was discovered by J. J. Thompson through the study of cathode rays [3]. He realised that their charge to mass ratio was significantly higher than that of any known ions and that they had the same properties regardless of their atomic origin, suggesting that they are their own fundamental particle. Thompson proposed the Plum Pudding model of the atom, where a nucleus is a mass of positive charge with electrons inside. Ernest Rutherford would later discover the nucleus of an atom, as a hard, compact positive center inside each atom, in the early 1900s from his famous Gold Foil Experiment [3]. This disproved the Plum Pudding model and

Rutherford would later show that the hydrogen nucleus can be found inside other nuclei, thereby discovering the proton. The last major constituent of the atom, the neutron, was found by James Chadwick through the study of neutral, massive penetrating radiation in 1932 [4].

That is where the story could have stopped, with the *atoms* of nature being the proton, neutron and the electron, however the existence of radiation emanating from certain atoms provided a hint that there was more to be discovered. With the dawn of particle accelerators and bubble chambers, a whole new world of fundamental particles were discovered and the indivisibility of the proton and neutron was challenged.

1.1 Eightfold Way

One of the remaining problems was to explain how the nucleus remained together under the electromagnetic repulsion of the protons. The force holding the nucleus together, the strong force, is short ranged and was proposed as the exchange of pions by Hideki Yukawa in the 1930s[5]. In the next 20 years with the advent of particle accelerators colliding protons and electrons into targets and cosmic rays interactions inside bubble chambers, a veritable zoo of particles was discovered. From this zoo emerged two main classes of particles, leptons and hadrons. Leptons included the electron, neutrino, and muon, and the hadrons were reserved for the heavier particles. Hadrons, which experience the strong force, included the mesons and the baryons, which received their names from the Greek for "middle" and "heavy" due the relative weight of the particles.

During this period, a number of heavier baryons and mesons were discovered. What was interesting about them is that they were produced in abundance and quickly, in about 10^{-23}s, but decayed slowly, on the order of 10^{-10}s. This suggested that their production and decay modes involved different interactions or forces. A quantum number, strangeness (S), was devised by Murry Gell-Mann and Kazuhiko Nishijima to help with the situation, with S being conserved in their creation and violated in their decay [6]. Gell-Mann and Yuval Ne'eman independently arranged the known particles in geometric structures using their masses, electric charges, and strangenesses

in the early 1960s [7]. This scheme that they devised was two sets of octets, the meson octet which held particles of zero spin, and the baryon octet which contained particles of spin $\frac{1}{2}$, and a baryon decuplet with spin $\frac{3}{2}$; all of which are shown in Figs. 1.1a, 1.1b, and 1.1c.

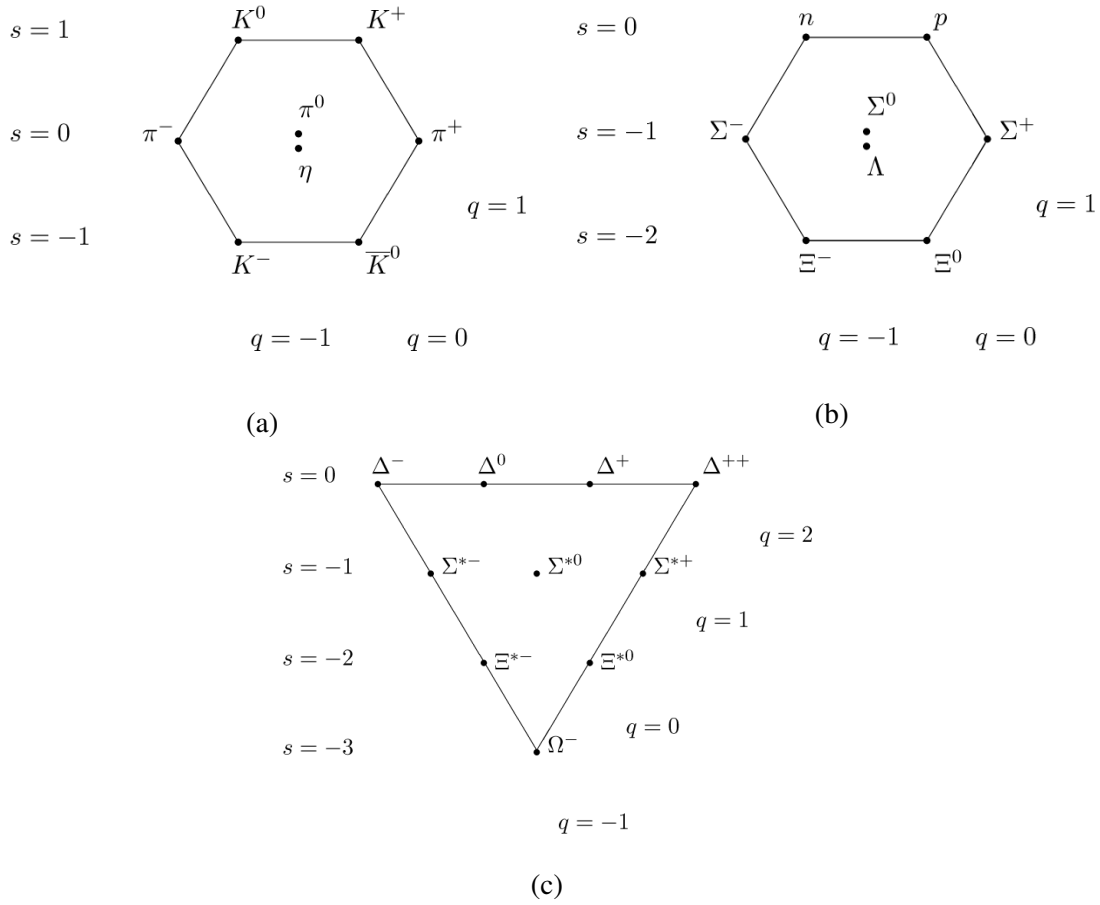


Figure 1.1: The Eightfold Way: a) the meson octet [8], b) the baryon octet [9], and c) the baryon decuplet [10]. Horizontal rows correspond to S values of the particles and the diagonal rows correspond their electric charges.

This structure of octets, dubbed the Eightfold Way, could also predict the mass of yet to be discovered particles through the Gell-Mann-Okubo mass formula [11]. This theory predicted the charge, strangeness, mass, and decay of the Ω^-, which was later discovered in 1964. Cabibbo also used the Eightfold Way to predict certain baryonic decays [12].

1.2 Quark Model

With the success of the Eightfold Way, George Zweig and Gell-Man both independently in 1964 proposed that the baryons and mesons are really combinations of a three more fundamental particles called quarks, with the quarks being the up (u), down (d), and strange (s) [13]. This quark model was thought to be a mathematical abstraction, since no free quarks were observed and their existence necessitated fractional electric charge and baryons number, something that had yet to be seen. In 1968, a series of experiments at the Stanford Linear Accelerator Complex (SLAC) showed that the proton did contain some substructure [14]. This wasn't immediately seen as complete evidence of quarks, but as evidence of the existence of partons, the point like constituents of protons and neutrons. It would take the notion of color charge, in analogy with electric charge, and the discovery of the J/Ψ particle in 1974 to cement the confidence in the model [15, 16]. Color charge, which comes in three colors "red", "blue", and "green", was proposed as a solution of the problem of baryons, such as the Ω^-, having spin $\frac{1}{2}$ and being composed of the same flavor of quark. The Pauli exclusion principle disallows particles to have the same set of quantum numbers, but with the advent of color charge this is no longer violated. The J/Ψ particle has an anomalously large lifetime, on the order of 1,000 times longer than similarly massed particles. The quark model proposed that the J/Ψ is a meson made of a fourth quark, c or charm, and its antiquark partner $\bar{c}$. This suggestion and the prediction that new mesons and baryons containing c and $\bar{c}$ would be found bolstered evidence for the model. These particles were later found in the 1970s. In the late 1970s the b or bottom quark was discovered, and in 1995 the t or top quark was discovered.

1.3 Quantum Chromodynamics

The modern treatment of particle physics relies on the Lagrangian formalism, whereby the Lagrangian is a mathematical structure that details the degrees of freedom and interactions of the particles within a theory. Quantum Chromodynamics (QCD) is the study of the strong interaction

and its Lagrangian,

$$\begin{aligned}\mathcal{L}_{QCD} &= \sum_q \bar{\psi}_{q,a}\left(i\gamma^\mu\partial_\mu\delta_{ab} - g_s\gamma^\mu t^C_{ab}\mathcal{A}^C_\mu - m_q\delta_{ab}\right)\psi_{q,b} - \frac{1}{4}F^A_{\mu\nu}F^{A\mu\nu} \\ F^A_{\mu\nu} &= \partial_\mu\mathcal{A}^A_\nu - \partial_\nu\mathcal{A}^A_\mu - g_s f_{ABC}\mathcal{A}^B_\mu\mathcal{A}^C_\nu,\end{aligned} \tag{1.1}$$

with the sum over quark flavors, q, and implied Einstein summation. The $\psi_{q,a}$ are the quark field Dirac spinors with Latin color indices and the γ^μ are the standard Dirac gamma matrices. There are six known quarks, u, d, s, c, b, and t, with their antiquark pairs and 8 gluons, which mediate the strong interaction. The quarks and antiquarks can possess color or anticolor, while the gluon possesses both color and anticolor, but not of the same kind.

g_s is the QCD coupling constant and it is related to the QCD structure constant by $\alpha_s = \frac{g_s^2}{4\pi}$. $\alpha_s(\mu_R^2)$, like all parameters in quantum field theories, are renormalization scale dependent. Quantum field theories are only effective, in the sense that they describe phenomena, up to an energy scale, μ_R^2. Within that scale the parameters of the theory take their respective renormalized values, but their values outside of that scale can change. The renormalization group equation encodes the scale dependence of α_s,

$$\mu_R^2\frac{d\alpha_s}{d\mu_R^2} = -\left(b_0\alpha_s^2 + b_1\alpha_s^3 + b_2\alpha_s^4 + ...\right), \tag{1.2}$$

where the b coefficients are functions of the number of "active" quarks, quarks whose mass is less than the scale, and the coefficients greater than b_1 are model dependent.

Unlike the other β functions in Quantum Electrodynamics (QED) or Quantum Flavordynamics (QFD), the right hand side of Eq. 1.2, the QCD β function contains a negative sign, signifying that α_s decreases logarithmically with μ_R [18]. This means that the higher the scale the smaller the coupling as seen in Fig. 1.2. This leads to asymptotic freedom or color confinement, the relegation of quarks to color neutral, or singlet, hadrons. In particle physics parlance, distance is inversely proportional to energy, so higher energies corresponds to smaller distances and vice versa. Therefore, at smaller distances the α_s is weak and the quarks and gluons are more or less "free",

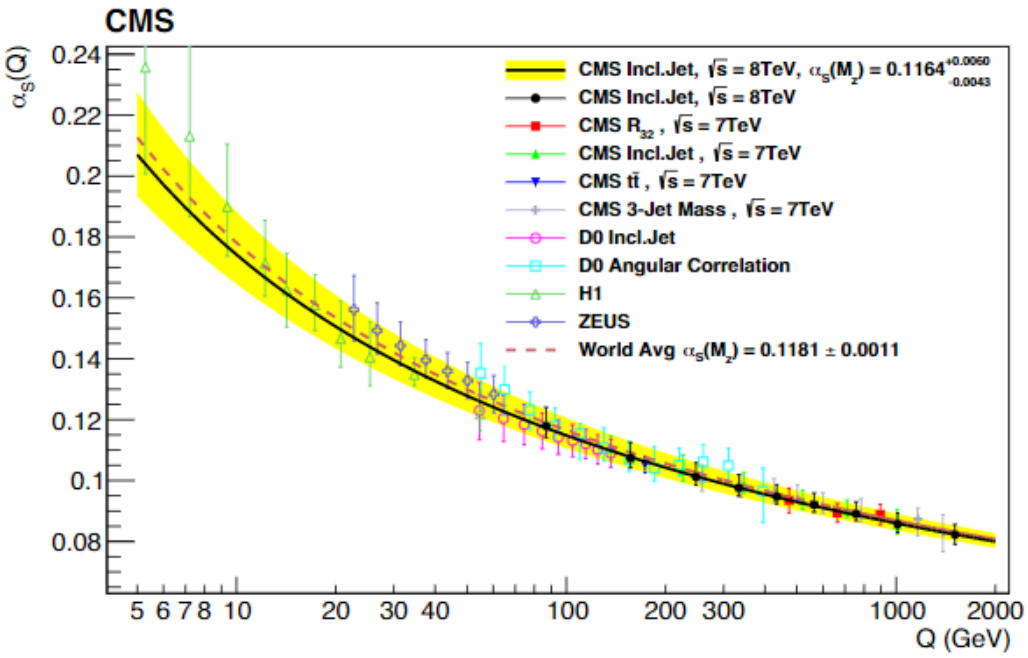


Figure 1.2: Depiction of the slow variation of α_s with respect to Q compiled from CMS, D0, H1 and ZEUS data [17]. This variation in the coupling strength with energy is referred to as the running of the coupling.

but at larger distances the coupling increases so much so that naked quarks and gluons hadronize, i.e. the quarks and gluons generate $q\bar{q}$ pairs from the vacuum that combine into hadrons.

The m_q term in $\mathcal{L}_{QCD}$ is the quark mass, which is not as straightforward to measure as the masses of other particles. All quarks are confined in hadrons save for the top, which decays before it has time to hadronize. Moreover, there can be significant dressing of the quarks inside hadrons, meaning that the masses of hadrons are not the simple summation of the masses of the constituent quarks. For instance, from lattice QCD calculations only 9% of the mass of the proton is due to the masses of the two up quarks, the one down quark, and the strange quarks and antiquarks in the sea. The rest is due to the interaction between the quarks and the gluons [19]. There are two ways to talk about quark masses, pole mass and renormalization mass. The pole mass is the mass that physicists intuitively think of when discussing the mass of particles, while the renormalization mass is model and renormalization scale dependent. QCD commonly uses the $\overline{MS}$ renormalization scheme, consequently the renormalization mass is quoted as $\bar{m}_q\,(\mu_R^2)$, with the scale in parentheses.

A listing of the quark masses and charges is given in Table 1.1.

Quark Flavor	Mass (MeV/c^2)	Charge (e)
u	$2.16^{+0.49}_{-0.26}$	$+\frac{2}{3}$
d	$4.67^{+0.48}_{-0.17}$	$-\frac{1}{3}$
s	93^{+11}_{-5}	$-\frac{1}{3}$
c	1270 ± 20	$+\frac{2}{3}$
b	4180^{+30}_{-20}	$-\frac{1}{3}$
t	173100 ± 900	$+\frac{2}{3}$

Table 1.1: Quark masses and charges. The u, d, and s masses are given at $\mu_R = 2\ GeV$, the c and b masses at $\mu_R = m_c,\ m_b$ respectively, and the mass of the top is the pole mass [20].

The last major aspect of the Lagrangian is the interaction terms, which are illustrated by the Feynman diagrams in Fig. 1.3. QCD has the standard vector boson-particle interaction term of the form $\bar{\psi} g \gamma \mathcal{A} \psi$, but instead of g representing a charge, like in QED, g represents a charge and the color matrix. t^A_{ab} are the color matrices, with the associated commutation relation,

$$\left[t^A, t^B\right] = i f_{ABC} t^C, \tag{1.3}$$

representing the nature of gluon exchange to rotate the color of the quark in the SU(3) space [20]. QCD is unique among the quantum field theories in the Standard Model (SM), in that it possess a SU(3) symmetry with the quarks, antiquarks, and gluons transforming in different representations of the group. Moreover, QCD is non-Abelian; the gluon fields, $\mathcal{A}^A_\mu$, do not commute, leading to self-interaction terms, which can be seen in Figs. 1.3b and 1.3c.

To make QCD predictions one would use the same perturbative machinery of quantum field theories with the summation of Feynman diagrams, calculation of matrix elements, integration over loop momenta, etc. But this fails to make complete predictions in QCD due to color confinement. To ameliorate this situation, a general process, such as a cross section (CS), is factored into a partonic sub-process, calculated using perturbative QCD (pQCD), convoluted with a non-perturbative set of parton distribution functions (PDFs) [21].

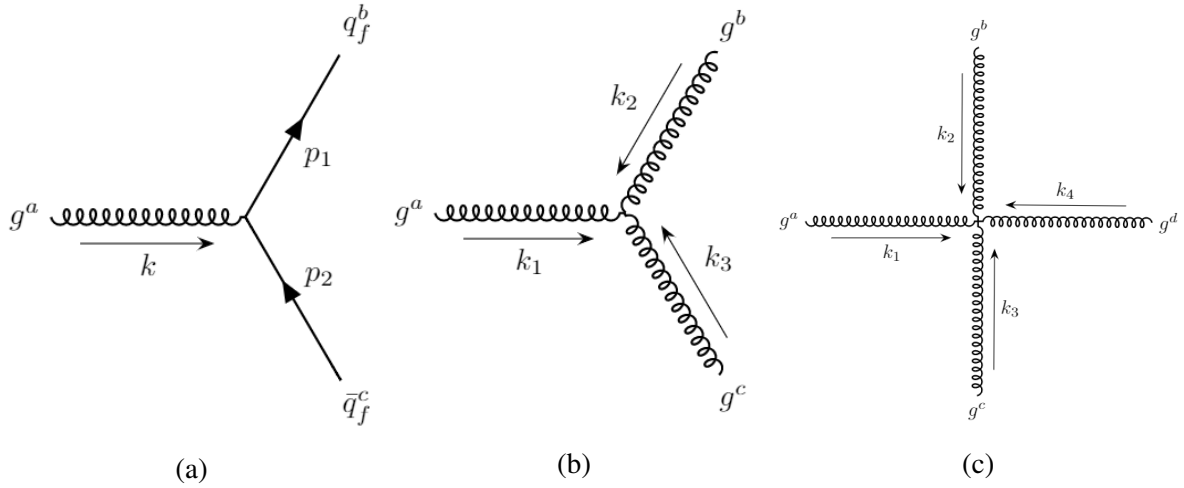


Figure 1.3: QCD Feynman Diagrams: a) the quark-gluon vertex, b) 3-gluon vertex, c) 4-gluon vertex. The Latin indices denote the color of the particles and the p's and k's denote particle momenta.

Factorization is made possible due the existence of two different energy scales, the hard process scale (Q^2) or factorization scale, and the characteristic hadronic scale. The hard process scale is the energy scale of the momentum transfer or the invariant mass of the propagator, which is greater than 1 GeV. The hadronic scale is the typical scale of the divergence of α_s, denoted as Λ_{QCD}. Λ_{QCD}, with a value approximating 200 MeV, arises as the constant of integration in the integral of Eq. 1.2 and is consequently μ_R dependent. Energy and time are inversely proportional from Heisenberg's Uncertainty principle, $\Delta E \Delta t \geq \hbar/2$, so the higher the hard process scale, the shorter the time. If we boost into the center of mass system, the hard process interaction happens at a moment where the partons of the hadron are essentially static and free from their neighbors, therefore the fully hadronic interaction devolves into an interaction between single partons. The factorization scale and the renormalization scale have different origins and subsequently do not need to be the same [22], but in practice are of the same order, and in this thesis will be treated as the same.

It is instructive to note that the PDFs contain the "ignorance" about what is beyond the factorization scale and are more akin to theoretical constructs used to interpret experimental observations than as experimental observables proper [22]. They are explicitly number densities, in that

$q_f\left(x, Q^2\right) dx$ describes the number of quarks of flavor f in the band dx, with x being the Bjorken x scaling variable described in detail in the next section. x can be thought of as the momentum fraction of the struck parton, however in terms of nuclear PDFs (nPDFs), x takes on the description of the fraction of the average momentum of a nucleon within the nucleus. Consequently, x in nPDFs has a range of 0 to A, not 0 to 1. At the same time, nPDFs devised by collaborations frequently restrict x to be within unity [23, 24], therefore care must be given when comparing and discussing different nPDFs and what they measure. The full integral of a PDF within a hadron gives the number of quarks of that flavor, for instance in the proton we know that the number of u quarks and d quarks are 2 and 1 respectively,

$$\begin{aligned} N_u^p &= \int_0^1 \left[u\left(x, Q^2\right) - \bar{u}\left(x, Q^2\right)\right] dx = 2 \\ N_d^p &= \int_0^1 \left[d\left(x, Q^2\right) - \bar{d}\left(x, Q^2\right)\right] dx = 1, \end{aligned} \tag{1.4}$$

with antiquarks counting as "negative" quarks in the sum. The total integrated number of quarks of a flavor within a hadron are referred to as the valence quarks, while the other quarks, antiquarks, and partons are said to be in the "sea" or "ocean".

The total number of PDFs that can be used to accurately describe the structure of a hadron can include all physically possible partons. Their values are model, scheme, and scale dependent. PDFs are found by a global minimization, usually using a χ^2 minimization and Hessian matrix error analysis [26, 27], between the theoretical CS or cross section ratio (CSR) using expressions that explicitly depend on the PDFs and experimental data along with measurement uncertainties. The Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) equations are used to evolve the PDFs and any Q^2 dependent QCD observables from the lower Q^2 values [28], usually around 2 GeV/c^2 to the required Q^2 for the experiment. These equations evolve the quark PDFs by calculating the change in the PDF due to the generation of gluons and the change in the gluon PDF by the generation of quark-antiquark pairs. PDFs are universal, in that the PDF values calculated from one physical process can be used for a different process. This is really a statement to the validity

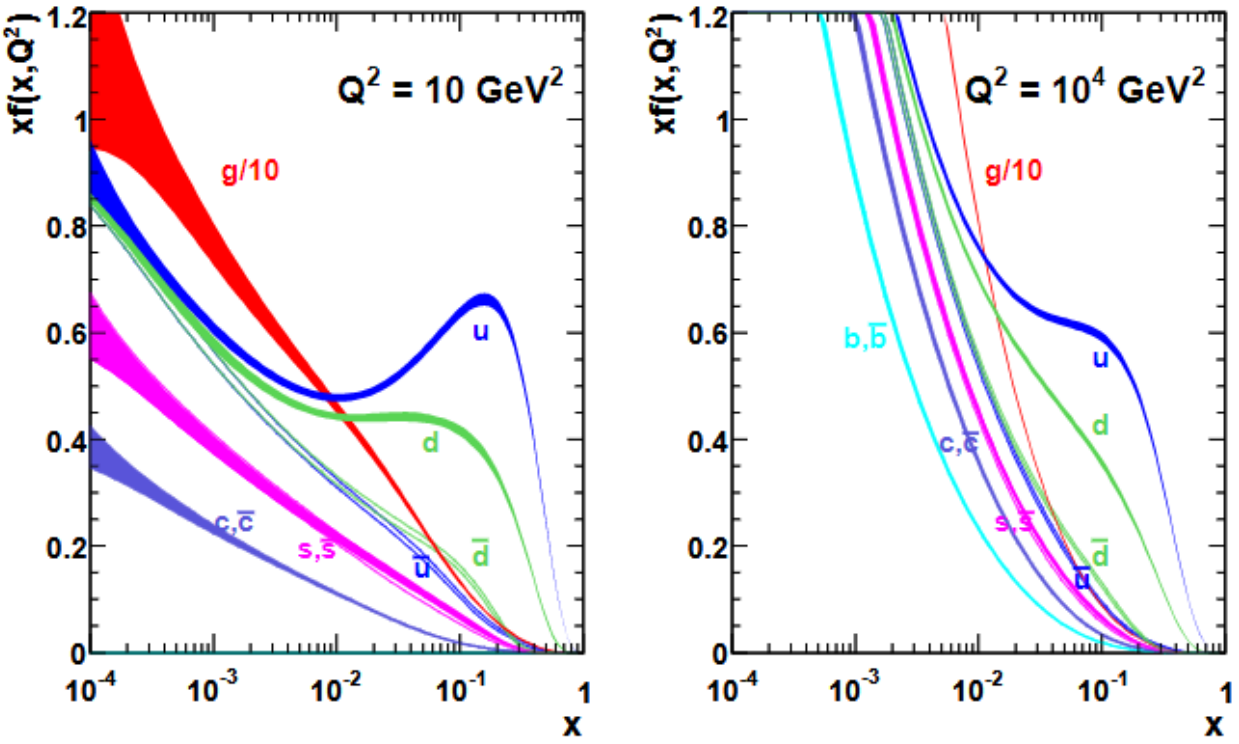


Figure 1.4: MSTW 2008 NLO parton distribution functions at $Q^2 = 10\ GeV^2$ on the left panel and $Q^2 = 10^4\ GeV^2$ on the right panel [25]. The x scaled PDFs are plotted versus x for ease of depiction and the bands correspond to the 68% confidence interval. The increase in band width at low x is due to the lack of experimental constraints on the PDFs in that region. The quark and antiquark PDFs for the heavy quarks, quarks with masses greater than the s, are assumed to be the same.

of factorization, since any process dependence is factored into the hard scattering CS. Different collaborations, such as CTEQ, MSTW, EPPS, DSSZ, etc., publish their values with uncertainties.

1.4 Deep Inelastic Scattering and Drell-Yan

Using QCD to resolve the structure of nucleons principally involves interacting with the partons inside the nuclei with a probe. There are two main ways to investigate this underlying structure: deep inelastic scattering (DIS) and Drell-Yan (DY). DIS nomenclature comes from the fact that the lepton has sufficient energy to peer into the hadron, i.e. energy greater than Λ_{QCD}, inelastic because the momentum exchange is sufficient to break up the hadron, and scattering due to the lepton exchanging a vector boson with parton inside the hadron.

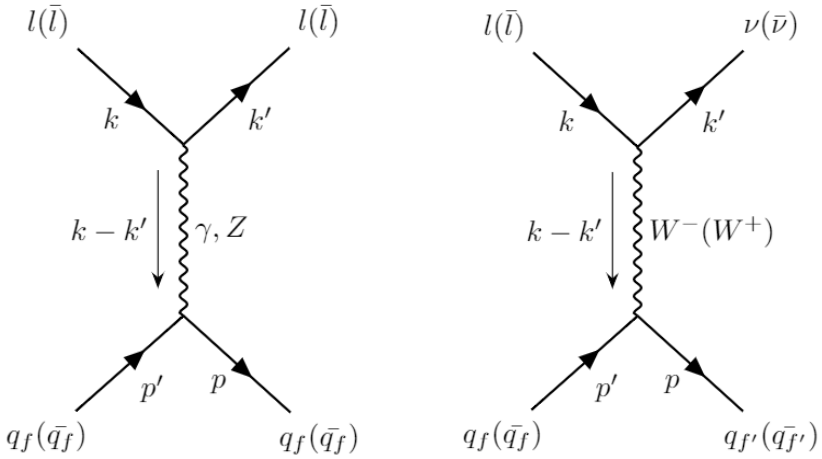
Schematically, DIS is represented as,

$$l + N \xrightarrow{\gamma^*, Z^*} l + X \tag{1.5}$$

for neutral current and

$$l + N \xrightarrow{W^*} \nu + X \tag{1.6}$$

for the charged current. The associated Feynman diagrams are shown in Figs. 1.5a and 1.5b.



(a) (b)

Figure 1.5: DIS Feynman Diagrams: with the left panel depicting the neutral current interaction and the right panel depicting the charged current interaction. In the neutral current diagram, $l(\bar{l})$ can be any charged lepton for γ exchange and any lepton for Z exchange. For the charged current diagram, $l(\bar{l})$ must be charged and $W^{\pm}$ exchange changes the (anti)quark flavor and the (anti)lepton to a (anti)neutrino.

The polarization averaged neutral current differential DIS CS is the sum of two non-perturbative structure functions that characterize the partonic structure of the hadron,

$$\frac{d^2\sigma^{NC}}{dxdQ^2} = \frac{4\pi\alpha_e^2}{xQ^4}\left[\left(1 - y - \frac{(Mxy)^2}{Q^2}\right)F_2\left(x, Q^2\right) + xy^2\left(1 - \frac{2m_l^2}{Q^2}\right)F_1\left(x, Q^2\right)\right] \tag{1.7}$$

[29]. There are several kinematic terms that require unpacking in Eq. 1.7 :

- $Q^2 = -q^2 = (k - k')^2$: the squared momentum transfer between the incoming lepton and the struck quark or antiquark

- α_e : electromagnetic structure constant
- $x = \frac{Q^2}{P \cdot q}$: the Bjorken scaling variable, to leading order (LO) in the infinite momentum frame, it is the momentum fraction carried by the struck parton
- M : mass of the nucleon
- m_l : mass of the lepton
- $y = \frac{P \cdot q}{P \cdot k}$: ratio of transferred energy to the initial lepton energy
- F_2 : the electromagnetic structure function associated with the absorption of a transversely polarized photon
- F_1 : the electromagnetic structure function associated with the absorption of a longitudinally polarized photon.

To leading twist, i.e. with $1/Q^2$ corrections, F_1 and F_2 are related in the Callan-Gross relation,

$$F_2 = 2xF_1, \tag{1.8}$$

with the full expression being,

$$F_2 = 2xF_1 \frac{1+R}{1+2Mx/\nu}. \tag{1.9}$$

R is the ratio of the absorption of longitudinally polarized photons to transversely polarized photons and ν is the energy exchanged between the lepton and the parton. R has been shown experimentally to be small, further confirming that quarks are spin $\frac{1}{2}$ particles [30]. F_2 is proportional to the sum of the quark and antiquark PDFs,

$$F_2 = \sum_i^{nf} e_i^2 x \left[q_i\left(x, Q^2\right) + \bar{q}_i\left(x, Q^2\right) \right], \tag{1.10}$$

with the sum over active quark flavors. It is through this structure function that Bjorken scaling,

portrayed in Fig. 1.6, the dependence of PDFs on x and logarithmically on Q^2, has been demonstrated.

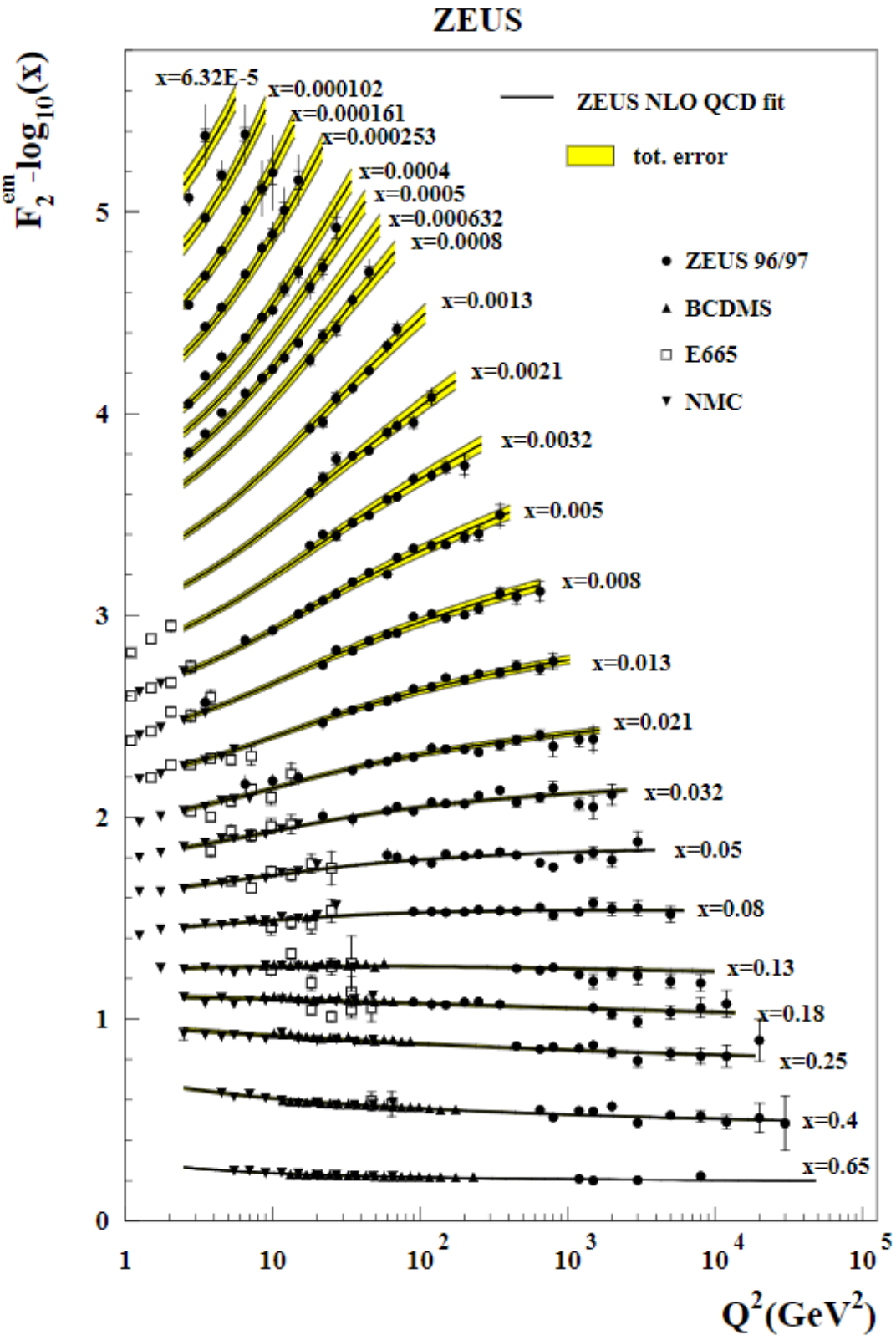


Figure 1.6: Illustration of the relationship between F_2, scaled by $-log_{10}(x)$, and Q^2 derived from ZEUS, BCDMS, E665, and NMC data [31]. For small x, this scaled F_2 shows an almost linear relationship, but for the $x \gtrsim 0.08$ the slope is flat. The dependence of PDFs and structure functions on x and not Q^2 is the signature of Bjorken scaling.

It was thought that the main mechanism generating the hadronic sea was the perturbative splitting of the gluons being exchanged between the valance quarks. This splitting should be symmetric in u and d and their antiparticles, due to isospin symmetry. Isospin or charge symmetry is the statement that since the strong interaction is blind to quark flavor and electric charge, and the u and d masses are similar within the hadronic scale, the u and $\bar{u}$ distributions inside the proton are the same as the d and $\bar{d}$ inside the neutron. However, the New Muon Collaboration (NMC) found evidence suggesting an asymmetry in the sea and consequently a non-perturbative component to the generation of sea quarks. NMC used muon DIS scattering on protons and neutrons and saw a violation of the Gottfried Sum rule [32],

$$S_G = \int_0^1 \frac{1}{x} \left[F_2^p(x) - F_2^n(x)\right] dx. \tag{1.11}$$

Taking the $d(x)$ and $\bar{d}(x)$ in the proton to be the same as the $u(x)$ and $\bar{u}(x)$ in the neutron and vice versa, i.e. isospin symmetry, the sum reduces to

$$\begin{aligned} S_G &= \frac{1}{3}\int_0^1 \left[u(x) - \bar{u}(x)\right] dx - \frac{1}{3}\int_0^1 \left[d(x) - \bar{d}(x)\right] dx + \frac{2}{3}\int_0^1 \left[\bar{u}(x) - \bar{d}(x)\right] dx \\ &= \frac{1}{3} + \frac{2}{3}\int_0^1 \left[\bar{u}(x) - \bar{d}(x)\right] dx. \end{aligned} \tag{1.12}$$

The sum rule was expected to yield $\frac{1}{3}$, with the sea quarks being thought to be flavor symmetric in $\bar{d}$ and $\bar{u}$, but NMC found [33],

$$S_G = 0.240 \pm 0.016, \tag{1.13}$$

which is significantly below the expectation and suggestive a sizeable difference between $\bar{d}$ and $\bar{u}$. At next-to-leading order (NLO), QCD processes that contribute at the α_s^2 level, the DGLAP equations do generate an asymmetry, however it is small compared to the NMC results [34]. This was further explored by the HERMES collaboration who extracted $\bar{d} - \bar{u}$ using semi-inclusive DIS of positrons on liquid hydrogen (LH_2) and liquid deuterium (LD_2). Using the pion yields (N^π) a

ratio,

$$r\left(x,z\right)=\frac{N_{p}^{\pi^{-}}-N_{n}^{\pi^{-}}}{N_{p}^{\pi^{+}}-N_{n}^{\pi^{+}}}$$
$$z=\frac{E^{\pi}}{\nu}, \tag{1.14}$$

was constructed and the asymmetry was extracted in the range of $0.02<x<0.3$ with the integrated result,

$$\int_{0.02}^{0.3}\left[\bar{d}\left(x\right)-\bar{u}\left(x\right)\right]dx=0.107\pm0.021\pm0.017, \tag{1.15}$$

confirming a significant asymmetry.

Neutral current DIS is blind to quark and antiquark flavor, but charged current interactions and DY can elucidate them. DY, first proposed by Sidney Drell and Tun-Mow Yan in 1970, can be thought of as a rotation of the DIS Feynman diagram, seen in Fig. 1.7, with the annihilation of a

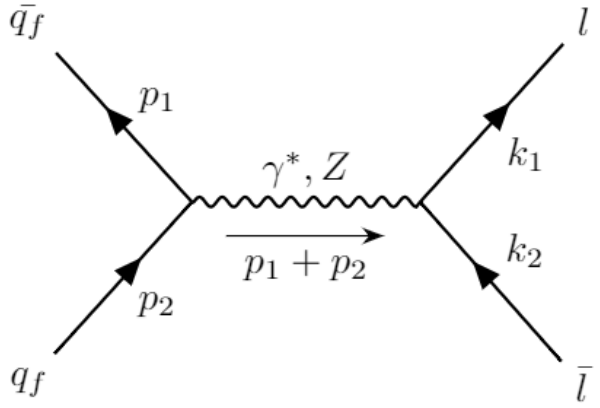


Figure 1.7: The Drell-Yan Feynman diagram showing the annihilation of a $q\bar{q}$ pair into a dilepton pair by way of an intermediate neutral vector boson. The k's and p's refer to particle momenta.

$q\bar{q}$ pair into a virtual γ or Z, which decays into a dilepton pair [35],

$$q_f+\bar{q_f}\xrightarrow{\gamma^*(Z)}l+\bar{l}. \tag{1.16}$$

The LO differential DY CS,

$$\frac{d^2\sigma\left(x_A, x_B, Q^2\right)}{dx_A dx_B} = \frac{4\pi\alpha^2}{9sx_Ax_B}\sum_{i=1}^{n_f} e_i^2\left[q_i^A\left(x_A, Q^2\right)\bar{q}_i^B\left(x_B, Q^2\right) + \bar{q}_i^A\left(x_A, Q^2\right)q_i^B\left(x_B, Q^2\right)\right], \tag{1.17}$$

with s being the center of mass momentum of the system squared,

$$\begin{aligned} s &\equiv \left(p_1 + p_2\right)^2 \\ s &= \frac{Q^2}{x_Ax_B}, \end{aligned} \tag{1.18}$$

Q^2 is the invariant mass of the vector boson and the dilepton pair, and $x_A\,(x_B)$ are the Bjorken x of hadron A(B). Experimentally, the DY CS is about a factor of two larger than the LO CS would suggest, and the transverse momentum distribution of the leptons is broader than initially expected. This k factor, the ratio of the experimental CS to the theoretical LO CS, and transverse momentum broadening suggested significant contributions from NLO diagrams. When these six diagrams are included, the discrepancy between theory and experiment is greatly reduced [36]. The DY x distribution is featureless, with most collaborations seeking measurements outside of the charmonium resonances. Fig. 1.8 shows the dimuon mass spectrum with meson and Z boson resonances.

In 1989, using DY to further investigate the structure of the nucleon, Heinrich *et al.* used pion induced DY on a tungsten target taking the CSR of π^+ to π^-, showing that the ratio of the sea quarks to valence quarks is a decreasing function of x, which can be seen in Fig. 1.9.

Further evidence of the asymmetric sea was found by the NA51 and NuSea collaborations. NA51 measured the DY asymmetry between σ_{pp} and σ_{pn},

$$A_{DY} = \frac{\sigma_{pp} - \sigma_{pn}}{\sigma_{pp} + \sigma_{pn}}, \tag{1.19}$$

yielding a $\bar{u}/\bar{d}$ ratio of 0.51 at $x = 0.18$ [39]. NuSea (FNAL E866), the predecessor to SeaQuest (FNAL E906), used the 800 GeV/c Fermilab proton beam incident on LH_2, LD_2, Fe, Be, and W

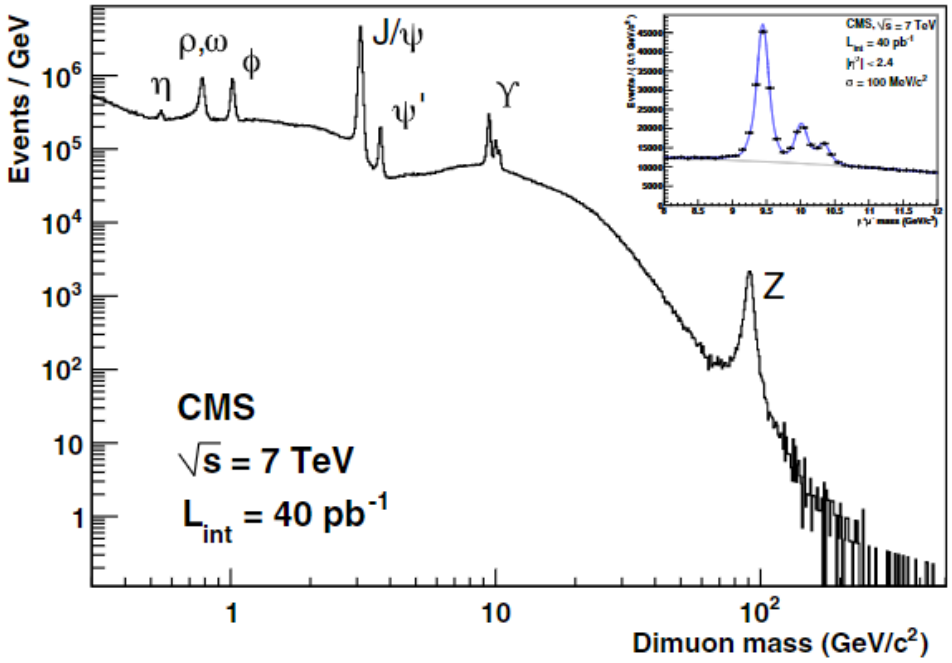


Figure 1.8: Dimuon mass distribution from the CMS [37]. The particle resonances correspond to the light quark mesons in the $m < 1$ GeV/c^2, the two charmonium mesons, J/Ψ and Ψ' near $m \approx 3.1$ GeV/c^2, the Υ, a $b\bar{b}$ meson near 9.4 GeV/c^2, and finally the Z boson at around 90 GeV/c^2. The bulk of the DY research is in the region between the charmonium resonances and the Υ resonance.

[40, 41, 42, 43]. Using Eq.1.17, isospin symmetry, and assuming negligible nuclear dependence in deuterium, i.e. $\sigma_{pD} = \sigma_{pp} + \sigma_{pn}$, the mass number normalized CSR (R_{pA}) is directly related to the sea asymmetry

$$
\begin{aligned}
R_{pD} = \frac{\sigma_{pD}}{2\sigma_{pp}} &\approx \frac{1}{2}\frac{\left[1+\frac{1}{4}\frac{d(x_1)}{u(x_1)}\right]}{\left[1+\frac{1}{4}\frac{d(x_1)}{u(x_1)}\frac{\bar{d}(x_2)}{\bar{u}(x_2)}\right]}\left[1+\frac{\bar{d}\,(x_2)}{\bar{u}\,(x_2)}\right] \\
R_{pD} &\approx \left[1+\frac{\bar{d}\,(x_2)}{\bar{u}\,(x_2)}\right].
\end{aligned}
\tag{1.20}
$$

This holds in the limit that $x_1 \gg x_2$, with x_1 and x_2 being the beam and target x's respectively. The actual asymmetry is extracted using PDFs in an iterative process. The results, shown in Figs. 1.10a and 1.10b, demonstrate a CSR and asymmetry largely inconsistent unity. It is notable that the depression of the CSR and asymmetry below unity in the last x_2 bin is discordant with the PDF

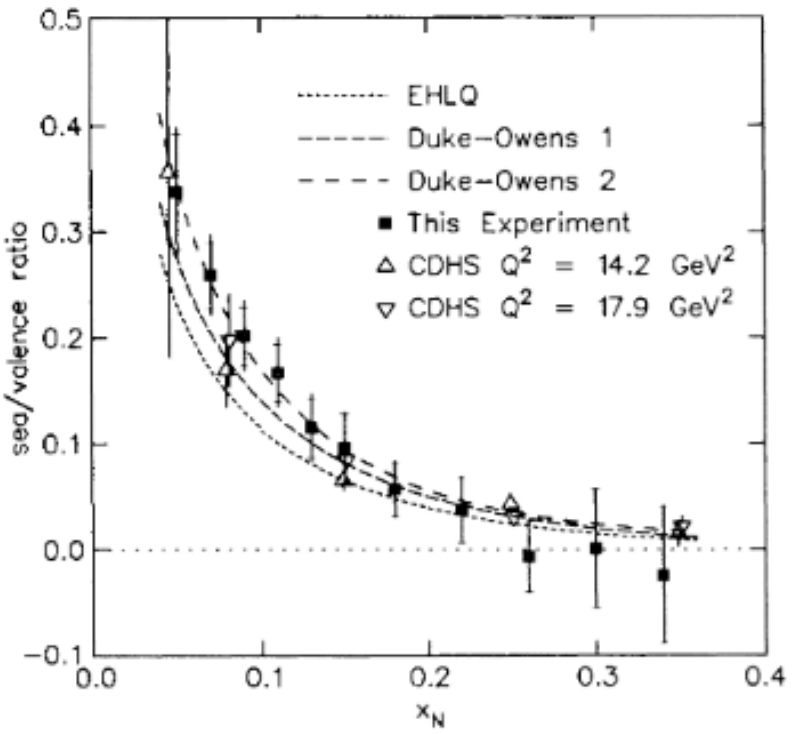


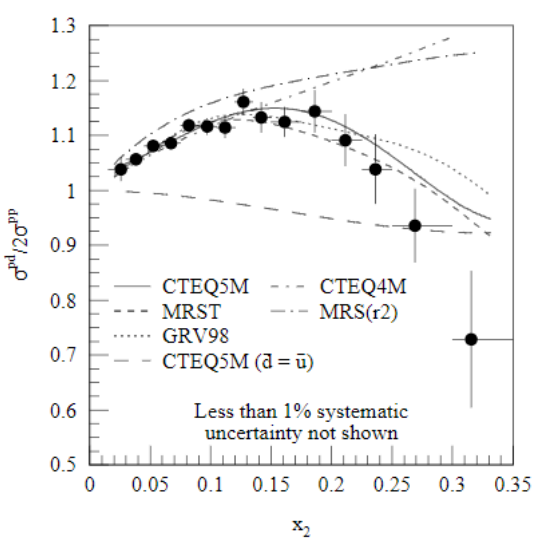
Figure 1.9: Ratio of the sea PDFs to the valence PDFs taken from $\pi^{\pm}$ scattering on a tungsten target with the theory curves from Duke-Owens, PDFs from EHLQ, and iron-neutrino DIS on iron from CDHS [38]. This ratio assumes the equivalence of the quark and antiquark PDFs in the sea.

fits and the results of the preliminary E906 work.

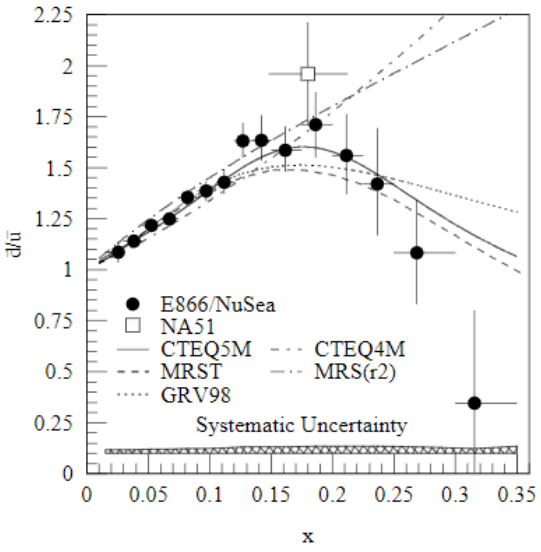
Many different models were proposed to explain the origin of this sea asymmetry, with the two main types of theories being the statistical model and the pion cloud model. The statistical model treats the quark and antiquark PDFs as Fermi-Dirac distributions and these distributions then evolve through the DGLAP equations. In Bourreley and Soffer's model, the PDFs contain a helicity dependent and helicity independent components,

$$\begin{aligned} xq^h\left(x,Q_0^2\right) &= \frac{A_q X_{0q}^h x^{b_q}}{exp\left[\left(x-X_{0q}^h\right)/\bar{x}\right]+1} + \frac{\tilde{A}_q x^{\bar{b}_q}}{exp\left(x/\bar{x}\right)+1} \\ x\bar{q}^h\left(x,Q_0^2\right) &= \frac{\bar{A}_q\left(X_{0q}^{-h}\right)^{-1} x^{\bar{b}_q}}{exp\left[\left(x+X_{0q}^{-h}\right)/\bar{x}\right]+1} + \frac{\tilde{A}_q x^{\bar{b}_q}}{exp\left(x/\bar{x}\right)+1}, \end{aligned} \tag{1.21}$$

with the $X_{0q}^{\pm}$ behaving like thermodynamic potentials of helicity $h = \pm$ and $\bar{x}$ taking the role of a universal temperature [44]. By fitting the polarized and unpolarized data, an asymmetry is found



(a)



(b)

Figure 1.10: Left panel: DY CS from E886 with various PDF predictions as a function of x_2. Right panel: the extracted $\bar{d}/\bar{u}$ ratio from E866 along with NA51's ratio and asymmetry predictions from the same PDF sets [41].

in accord with E886's data for $x_2 < 0.25$.

The pion cloud model posits that the proton and neutron are actually Fock decompositions of pions and more massive baryons,

$$|p\rangle = \sqrt{Z}\,|p\rangle_0 + \sum_{B=N,\Delta} \int d\Omega_{\pi B}\,|\pi B\rangle\,\langle \pi B|_0\,, \tag{1.22}$$

with the sum over baryons and pions so that the sum of the baryon charge and the pion charge is equivalent to the proton charge and the integral over the phase space [45]. The PDFs then become a sum of the bare PDF and convolutions of the pion-baryon momentum distributions and their respective PDFs,

$$q_N^f(x) = Z q_{N0}^f(x) + \sum_{B=N,\Delta} f_{\pi B} \otimes q_\pi^f + \sum_B f_{B\pi} \otimes q_B^f, \tag{1.23}$$

with convolution integral,

$$f_{\pi B} \otimes q_{\pi}^{f} \equiv \int_{x}^{1} \frac{dy}{y} f_{\pi B}\left(y\right) q_{\pi}^{f}\left(x/y\right). \tag{1.24}$$

Alberg and Miller's pion model and Bourreley and Soffer's model both agree with E866's asymmetry ratio data in the $x_2 < 0.2$ region. Furthermore, Alberg and Miller's model is consistent with the $\bar{d}-\bar{u}$ distribution to $x_2 \approx 0.3$ [45]. A more in depth review of the asymmetry models including helicity dependencies and transverse momentum dependence can be found Refs. [29] and [46].

1.5 Nuclear Dependence

As a general notion, nuclear dependence is seen as the modification of hadronic observables due to the influence of other hadrons; namely the difference between free hadrons and bound hadrons. This modification carries over to the PDFs that comprise the observables. As an anstaz, the nucleus was thought to behave as a collection of free protons and neutrons, with the hadronic CS being the sum of the free proton and free neutron CSs weighted by the number of protons and neutrons and a similar expression for the PDFs,

$$\begin{aligned} \sigma_A &\approx Z\sigma_p + (A-Z)\sigma_n \\ q_f^A\left(x,Q^2\right) &\approx Z q_f^p\left(x,Q^2\right) + (A-Z) q_f^n\left(x,Q^2\right). \end{aligned} \tag{1.25}$$

There was also the thought that the mass number normalized CSR,

$$R_{pA} = \frac{\sigma_{pA}/A}{\sigma_{pD}/2}, \tag{1.26}$$

would be expected to be approximately unity over various kinematic variables. When using an isoscalar target in the denominator, the DY R_{pA} is related to the $\bar{d}/\bar{u}$ asymmetry,

$$R_{pA} \approx 1 + \frac{A-2Z}{A}\frac{\bar{d}-\bar{u}}{\bar{d}+\bar{u}}, \tag{1.27}$$

holding NLO corrections cancel in the ratio, which has been shown to be true to the few percent level [47]. Experimentally, this notion has shown to be naive and the extracted nPDFs differ from this simple dependence as seen in Fig. 1.11, which depicts the ratio of the ^{207}Pb nPDF to the anstaz clearly illuminating the nuclear alteration of the PDFs.

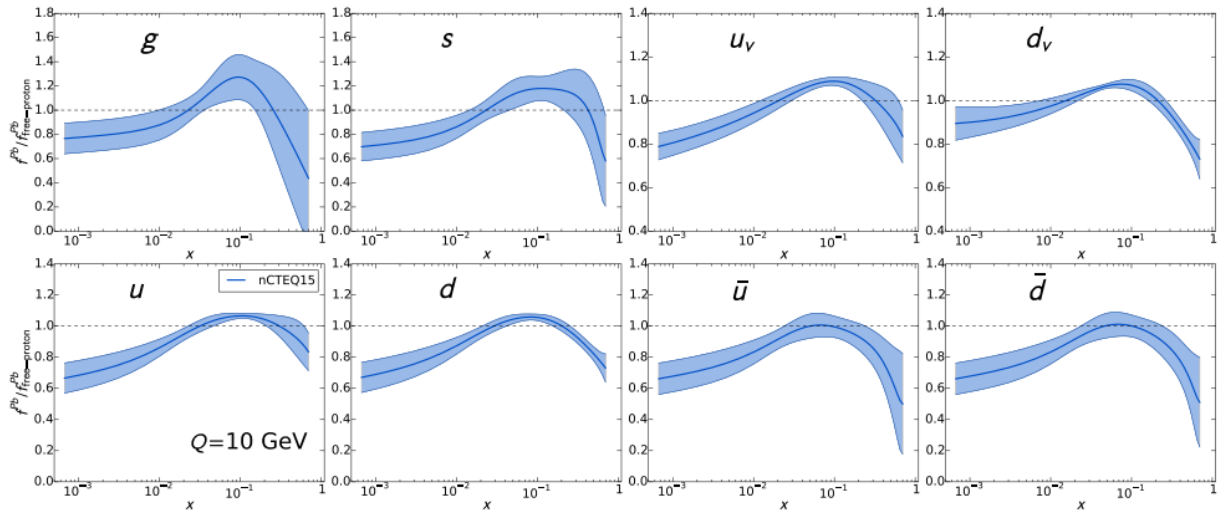


Figure 1.11: The panels describe the ratio of the nCTEQ15 PDF for ^{207}Pb to the naive sum of Z weighted free proton PDF and $(A-Z)$ weighted neutron PDF at $Q = 10\ GeV$ [23]. The subscript "v" denotes the valence PDFs and the u and d nPDFs contain valence and sea contributions. The $\bar{s}$ nPDF is the same as the s nPDF.

The principal way to study nuclear dependence is measuring the R_{pA} distribution, whereby experiments take the ratio with an isoscalar target in the denominator, such as D or C, and perform an isoscalar correction to the numerator to better delineate any nuclear dependence between two collections of deuterium nuclei [48]. A variation between unity and experiment was first shown by the European Muon Collaboration (EMC) when measuring the R_{pA} for Fe to D as a function of x_2. In addition to x_1 and x_2, nuclear dependence is also studied in p_T and x_F, with p_T being the transverse momentum of the products and $x_F = x_1 - x_2$.

There is limited R_{pA} DY data so the conclusions of x_F and x_1 dependence are restricted to E886 and E772. The x_F and x_1 distributions in E772 and E886 show a gradual decrease in R_{pA} as x_F and x_1 increase as can be seen in Fig. 1.12. Part of this increase is due to shadowing at high x_1, which corresponds to low x_2 [43]. There is an A dependence in the ratios as the ratio for W in

each x_1 bin is lower than the corresponding bin in Fe. This is also due in large part to shadowing, which will be explained in more depth in Section 1.5.1.

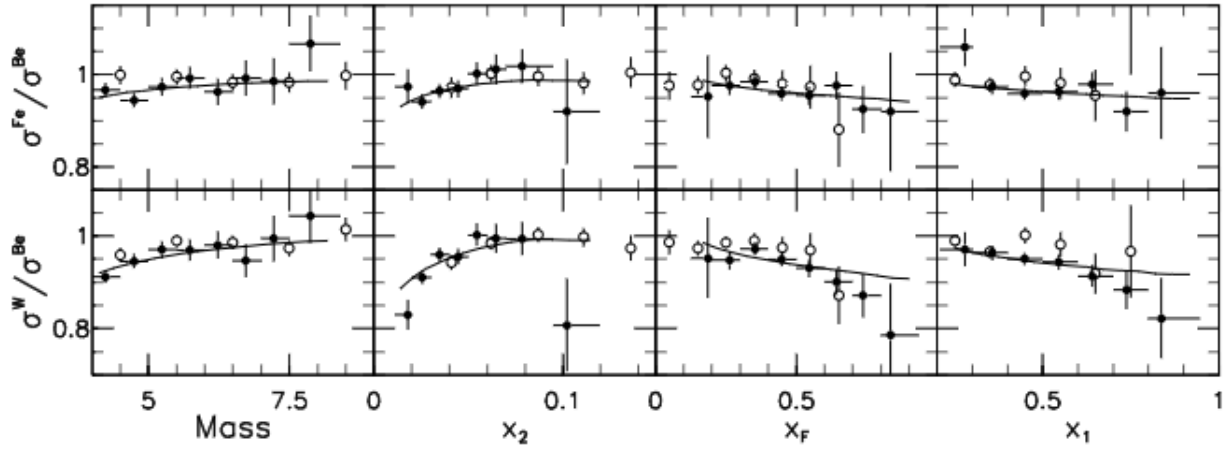


Figure 1.12: R_{pA} for iron and tungsten to beryllium with E866 and E772 being represented as solid and open dots respectively [43]. The ratios cover the mass, x_1, x_2, and x_F kinematic variables and are accompanied by the LO cross section ratio predictions from MRST and EKS98 PDF sets.

p_T dependence is related to the initial state interactions of the incoming parton with the nuclear medium of the target. Subsequently, DY is a clean probe to investigate this interaction since the final state particles, leptons, are color singlets. p_T studies by E866 and E772 demonstrate a rise in R_{pA} at increasing p_T [43]. In addition to a simple p_T distribution, measurements of the difference between the p_T^2 for nuclear targets and deuterium,

$$\Delta\langle p_T^2\rangle \equiv \langle p_T^2\rangle\,(A) - \langle p_T^2\rangle\,(D)\,, \tag{1.28}$$

have also been made by NA10 and E772. Using unpublished E772 data, McGaughley *et al.* showed that Δp_T^2 increases with A [36]. This p_T broadening, with $p_T^2 \propto A^{1/3}$, can further be related to the energy loss of the incident parton.

The x_2 distribution, which has been studied in DIS and DY, can be broken into four main regions: shadowing, antishadowing, EMC effect, and Fermi motion. The regions and the demarcations between them are shown in Fig. 1.13.

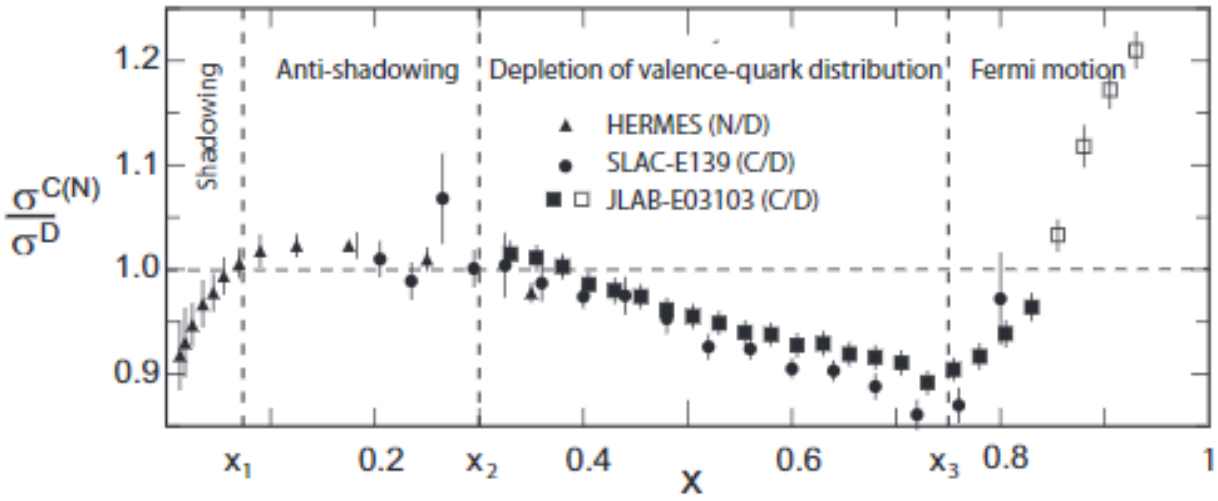


Figure 1.13: R_{pC} x_2 distribution delineating the shadowing ($x \lesssim 0.06$), anti-shadowing ($0.06 \lesssim x \lesssim 0.3$), EMC ($0.2 \lesssim x \lesssim 0.7$), and Fermi motion regions ($x \gtrsim 0.7$). The data is comprised from HERMES, SLAC, and JLAB [49].

1.5.1 Shadowing and Antishadowing

Shadowing and antishadowing are defined as regions in x with ranges of $x \lesssim 0.06$ and $0.06 \lesssim x \lesssim 0.3$ respectively. In the shadowing region R_{pA} decreases below unity, while in the antishadowing region it rises above unity. There are two main explanations for these effects: one is a change in the parton PDFs at low x and the other is an interference between scattering modes. In the first explanation, if we boost to the infinite momentum frame, the longitudinal length of the nucleus is approximately $1/p$, while the lengths of the partons are $1/xp$. If x is small enough, the parton's length extends over the length of the nucleus and the gluons, quarks, and antiquarks that "belonged" to a nucleon now behave as if they came from the entire nucleus. If these partons have the same impact parameter, the spatial overlap can cause a decrease in quark and antiquark PDFs through annihilation [50]. The vector boson experiences a rarefied partonic density and hence the interaction is shadowed [51]. Since this shadowing is inversely related to x, at higher x values the shadowing decreases and gives rise to the antishadowing region.

In the interference picture, shadowing and antishadowing are the consequence of interference between the single nucleon scattering and multiple nucleon scattering. In DIS, the interacting vec-

tor boson fluctuates into a $q\bar{q}$ pair and has a probability of scattering elastically with a surface nucleon and then scattering inelastically a nucleon deeper inside the nucleus. It also has a probability of singly scattering inelastically with an internal nucleon. These two possibilities possess opposite phases in the shadowing region, which leads to a decrease in the CS [52]. This can also be thought of as decrease in the incoming flux that can undergo DIS. In the antishadowing region, the phase difference is complex, which leads to constructive interference and R_{pA} rises above unity [52]. These effects are not limited to higher A nuclei, but have been demonstrated to exist even deuterium as seen in Fig. 1.14.

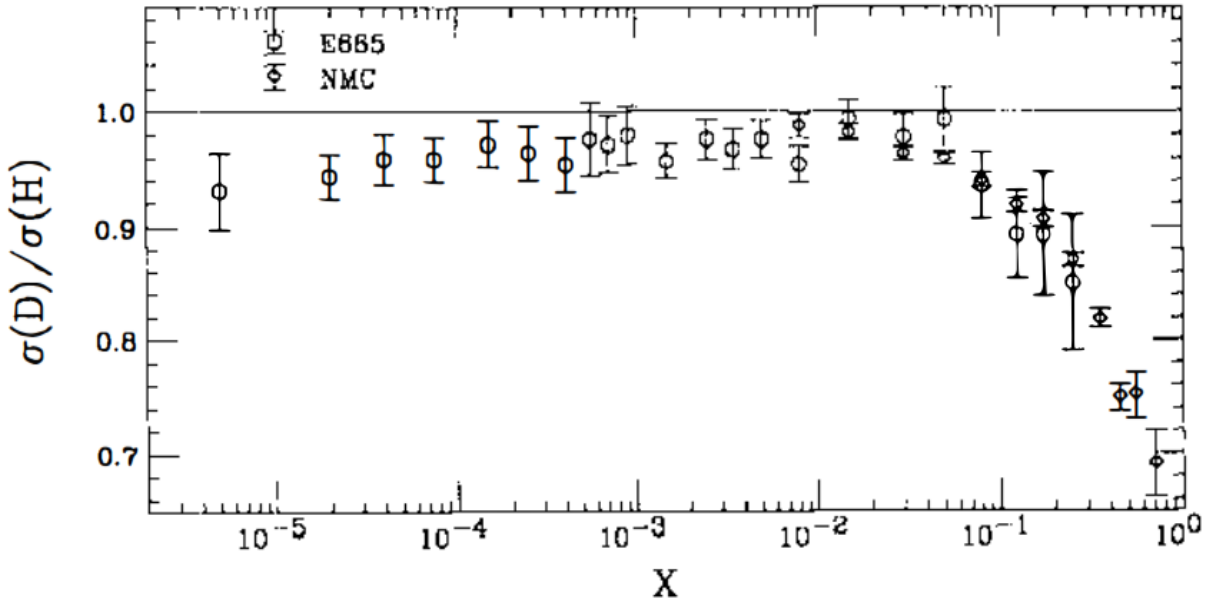


Figure 1.14: The R_{pD} ratio as a function of x_2 using data from E665 and NMC. Deuterium shadowing is seen in the lower x_2 region [30].

1.5.2 EMC Effect

There is some ambiguity in the terminology of using the EMC effect, with some authors using it to refer to the departure of the R_{pA} from unity in x and others, including this thesis, labeling just the $0.2 \lesssim x \lesssim 0.7$ region as the EMC effect. In the region $x > 0.2$, R_{pA} was expected to approximate unity save for Fermi motion due to the dominance of the valence PDFs, but the EMC collaboration

found a marketed departure from this notion [53]. EMC measured the F_2 ratio between Fe and D through muon DIS and found a depression in the ratio below unity for $x > 0.35$. Soon after their data was published a reanalysis of old empty target data at SLAC confirmed the effect [54]. In that analysis the empty target was steel and the ratio was done to deuterium. Later in that year a follow up paper investigated the ratio between aluminium and deuterium, as seen in Fig. 1.15, further confirming the EMC results.

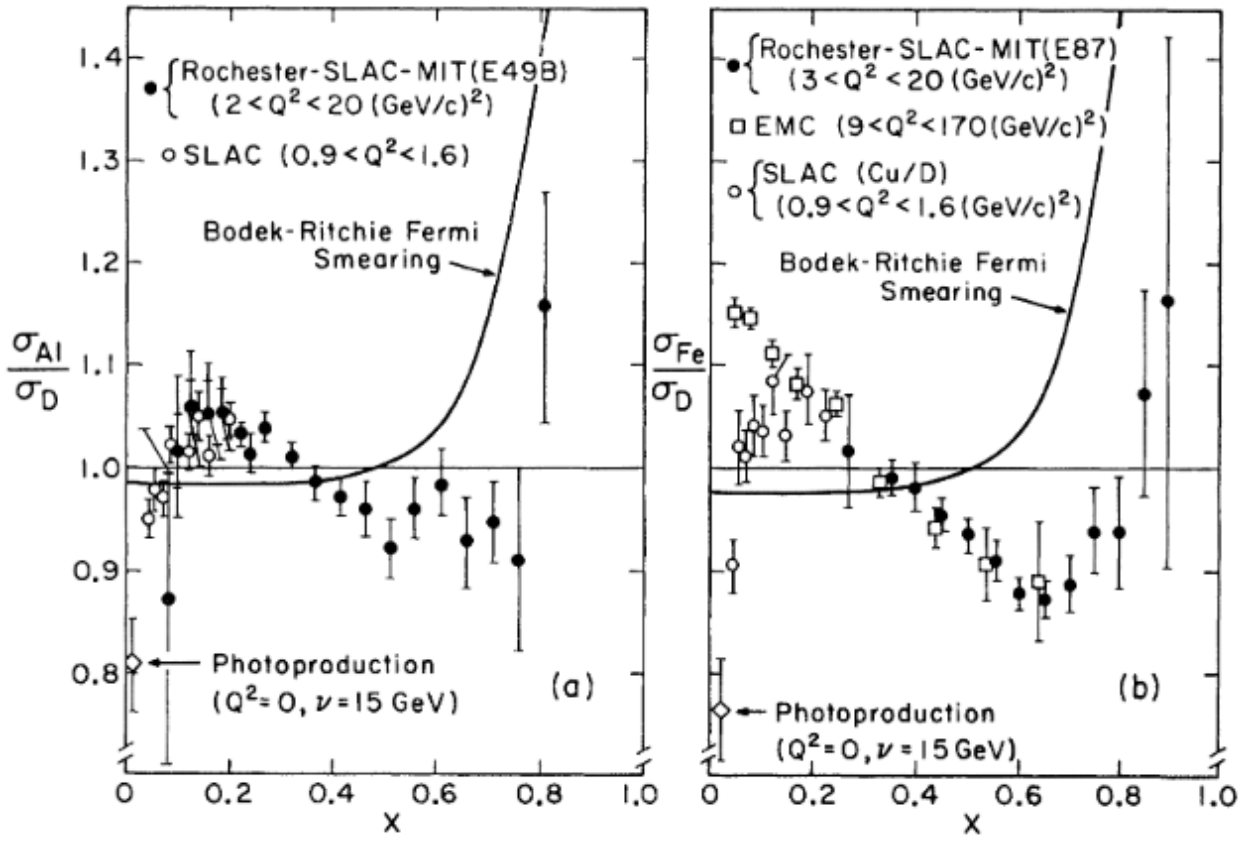


Figure 1.15: The left panel contains x_2 distribution for R_{pAl} from SLAC and the right panel has R_{pFe} ratios for SLAC and EMC [55]. The data is also plotted with a Bodek-Ritchie Fermi Smearing curve depicting the expected ratio if only Fermi motion effects were present and some ancillary R_{pCu} data from SLAC is plotted for comparison.

This decrease in R_{pA}, in the range $0.2 \lesssim x \lesssim 0.7$, would go on to be duplicated in further experiments, such as BCDMS, NMC, and FNAL 665 as outlined in Norton's review [56]. This wane in the ratio is puzzling since in this region the sea quarks and gluon PDFs are small, suggesting that there is a modification of the valence distributions in nuclear media. It does not appear to be

an alteration in the gluon PDFs, since the net charge-weighted momentum fraction carried by the quarks and antiquarks,

$$\Delta \equiv \int \left[F_2^{Fe}\left(x, Q^2\right) - F_2^D\left(x, Q^2\right)\right] dx, \tag{1.29}$$

is zero within errors [57]. There is a small A dependence in this region, but the effect is thought to correlate more with the local density of the nucleus [58].

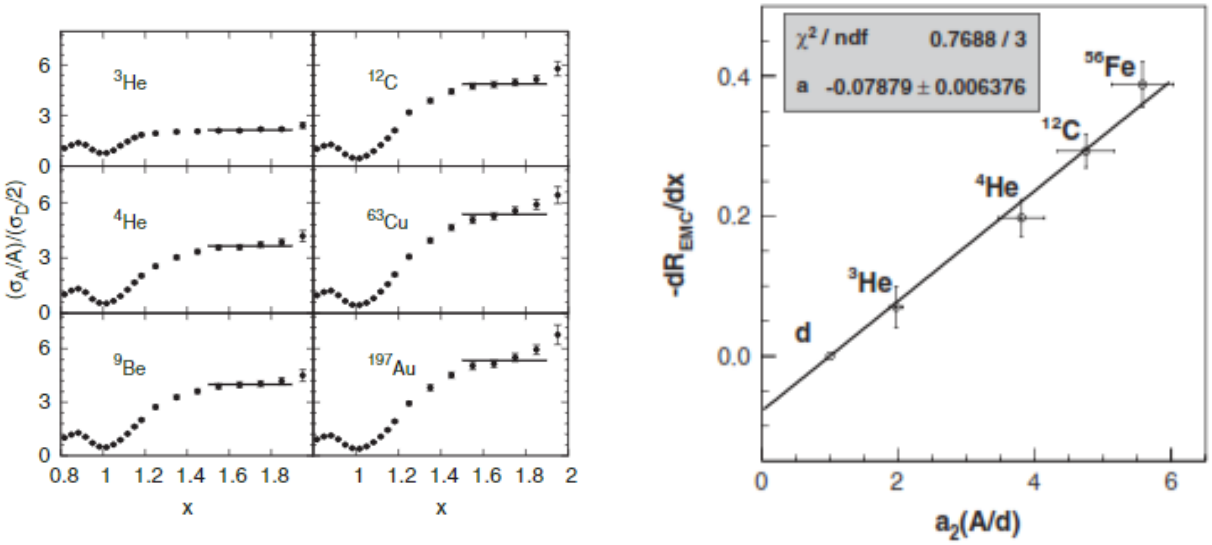
Many different models have been proposed to explain the EMC effect, with the main approaches being binding corrections, pion enhancement, multiquark clusters, and dynamical rescaling. In the binding models the x of the stuck parton is modified due the binding energy of the nucleon-nucleus system. The binding energy causes the mass to become off shell and this reduced mass increases x. This shift in x means that R_{pA} is the ratio of PDFs at two different x's; x calculated by the experiment aligns with the x in the deuterium PDF, but differs from the "true" x in the nuclear PDFs. This difference yields a reduction in R_{pA}. This effect is analogous to the off shell effect in Ref. [59].

Pion enhancement models seek to modify F_2 as in Ref. [60] or the quark and antiquark PDFs as in Ref. [61] to include pions within the nucleus outside of those used to bind the nucleons together. Since pions explicitly contain antiquarks, the antiquark PDFs are increased and with Δ being zero, R_{pA} would decrease in this range. Multiquark cluster models are similar to pion enhancement models, in that there is adjustment in F_2. Here F_2^A becomes the sum of the nucleon F_2 and a combination of quark bag structure functions. These quark bags come in multiples of 3, to keep the nucleus a color singlet. Due to the bags having more quarks, the x distribution of bag can be extended from $0 < x < N/3$, with N being the number of quarks in the bag. The nature of there being some probability of x existing beyond unity, requires softening of the x distribution in the EMC region, which leads to lower R_{pA} [62].

The conception of dynamical rescaling comes from the observation that the nuclear structure function of iron at Q^2 was similar to structure function of deuterium at $\xi^A\left(Q^2\right) Q^2$, with $\xi^A\left(Q^2\right) > 1$. This effect is attributed to the increase in the quark confinement radius in nucleons over free nuclei [63]. The confinement scale is related to the factorization scale, with $\mu_A < \mu_N$, suggesting

that the interaction of multiple nucleons generates more DGLAP evolution at the same Q^2 value.

Lastly, there is a connection between the CSR at $x \approx 0.5$ and short range correlations (SRCs). Short range correlations are correlations between nucleons that lead to high relative momenta, but low total momentum. This pairing pushes the x values of the partons within the nuclei in opposite directions, so that if one member has an x within the EMC region, the other has an $x > 1$. This phenomena is not limited to just pairs of nucleons as there is evidence of triads of nucleons as well [64]. Higinbotham *et al.* were the first to suggest that there was a connection between the EMC region and SRC plateaus, since they both share local density dependencies [65]. Weinstein *et al.* took the negative slope in the EMC region and plotted it against the "SRC scale factor", which is the CSR in the plateau region seen in Fig. 1.16a, revealing a strong linear correlation between them as seen in Fig. 1.16b. This further suggests a relationship between high momentum nucleons in the $x > 1$ region and nucleons in the EMC region, which is explored in depth in Ref. [66].



(a) (b)

Figure 1.16: Left panel: R_{pA} as a function of x_2 for ^{3}He, ^{4}He, C, Cu, Be, and Au with associated plateau regions [64]. Right panel: the plot of the negative slope in the EMC region to the SCR scale factors [67].

1.5.3 Fermi motion

The last region, Fermi motion, is characterized by low PDFs and structure functions, but a rise in R_{pA} in the region $x \gtrsim 0.7$. This is due to the interaction between the neighboring nucleons and the struck nucleon. Due to the Heisenberg uncertainty principle, the confinement of the nucleon by its neighbors into a smaller volume leads to a larger uncertainty in its momentum. This in turn enhances the probability of striking a high x parton. R_{pA} waxes due to deuterium, carbon, or whatever target in the numerator possessing less nucleons, and therefore less possible Fermi motion, than the target in the denominator. A more through discussion of these models can be found in Refs. [30, 56, 63].

1.6 SeaQuest Predictions

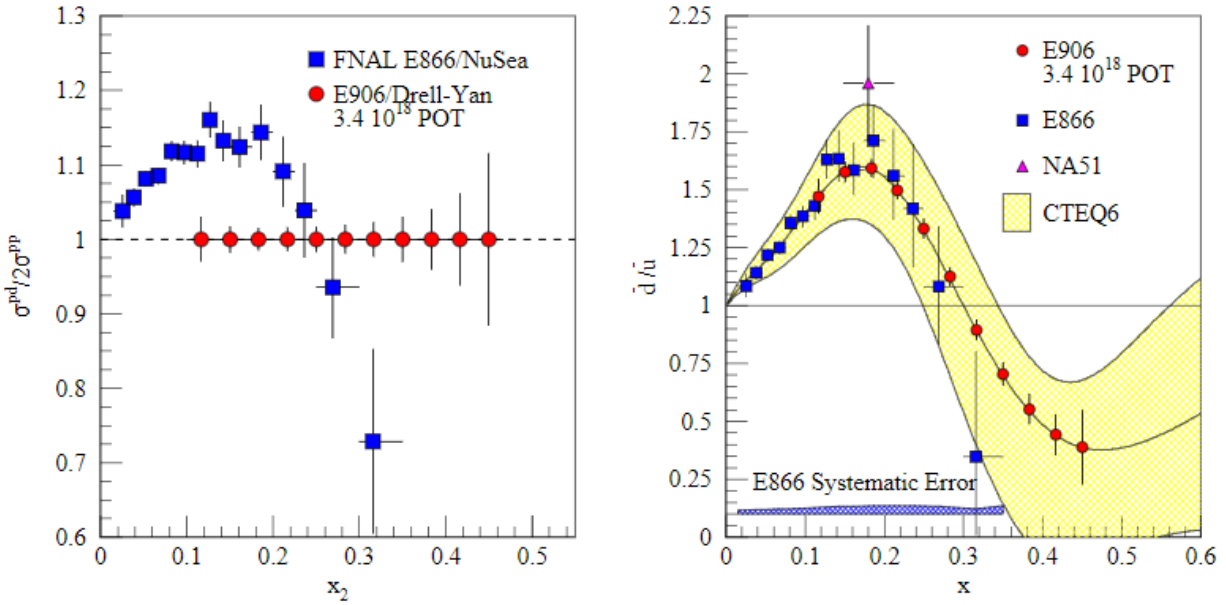


Figure 1.17: Left panel: E866 CSR and the SeaQuest initial CSR proposal. Right panel: extrapolated $\bar{d}/\bar{u}$ ratio from E866 accompanied by the NA51 result and the SeaQuest predicted ratio following the CTEQ6 PDF fit [68].

FNAL E906 or SeaQuest is the successor of E866 with an overlap in E866's data in the $0.1 < x_2 < 0.3$ range and an extension in the DY CSR data to $x_2 = 0.45$ as seen in Fig. 1.17. SeaQuest,

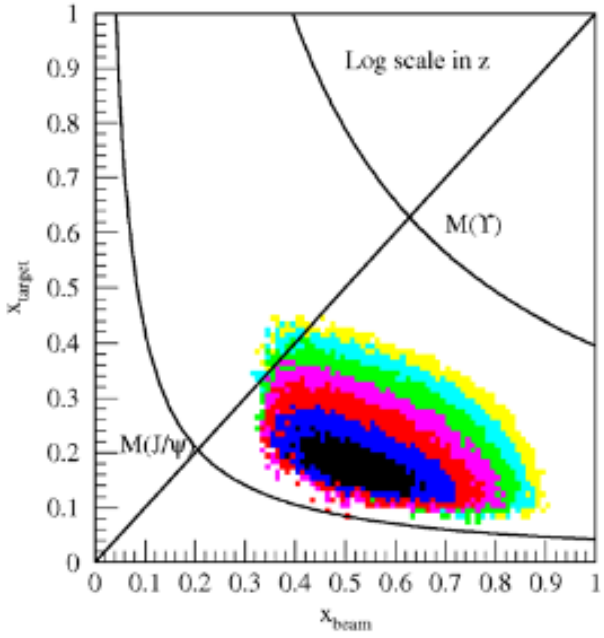


Figure 1.18: SeaQuest acceptance in the $x_1 - x_2$ plane with darker colors representing increased acceptance [68].

like E866, is a fixed target DY experiment seeking to measure the DY CS and extrapolate the sea asymmetry ratio. Due to the lower energy of the Fermilab beam, 120 GeV compared to E866's 800 GeV beam, the measured CS is expected to be ≈ 7 times larger. Moreover, the lower energy allows for 7 times the luminosity and the J/Ψ, one of the dominate backgrounds, is reduced by a factor of 10 due to the production CS being proportional to energy. These conditions conspire to increase the yield of high mass dimuons by a factor of 50 over E866 [68]. Moreover, many of the models seeking to explain the asymmetry predict an asymmetry beyond $x_2 \approx 0.2$ greater than unity in conflict with E886's data as outlined in Section 1.4. SeaQuest is equipped to provide data with lower statistical errors in this region to ameliorate the discrepancy.

In addition to liquid hydrogen and deuterium targets, SeaQuest also contains Fe, C, and W nuclear targets. With the x_2 acceptance range of the experiment, depicted in Fig. 1.18, SeaQuest also overlaps with E772's Fe, C, and W R_{pA} values, shown in Fig. 1.19, and can extend the DY data to $x_2 = 0.45$. This will provide much needed DY data in the EMC ratio region and bolster p_T broadening studies. The SeaQuest predicted R_{pFe} and R_{pW} are shown in Figs. 1.20a and 1.20b.

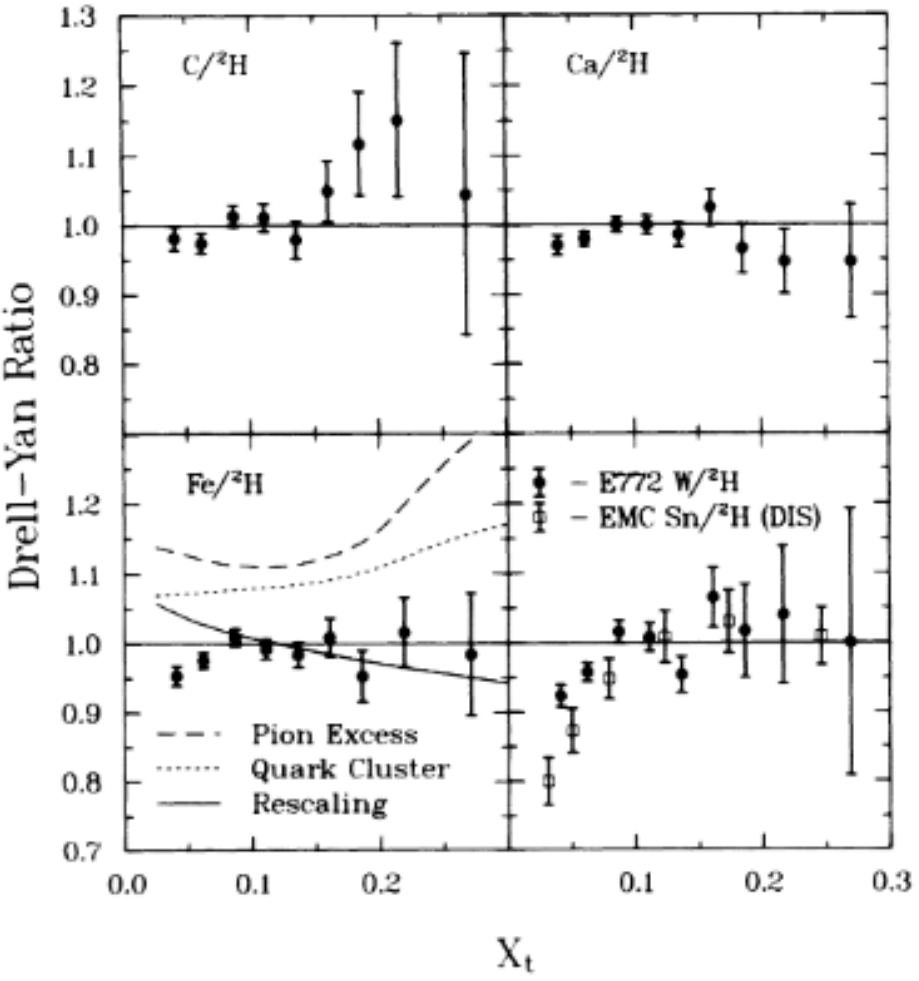
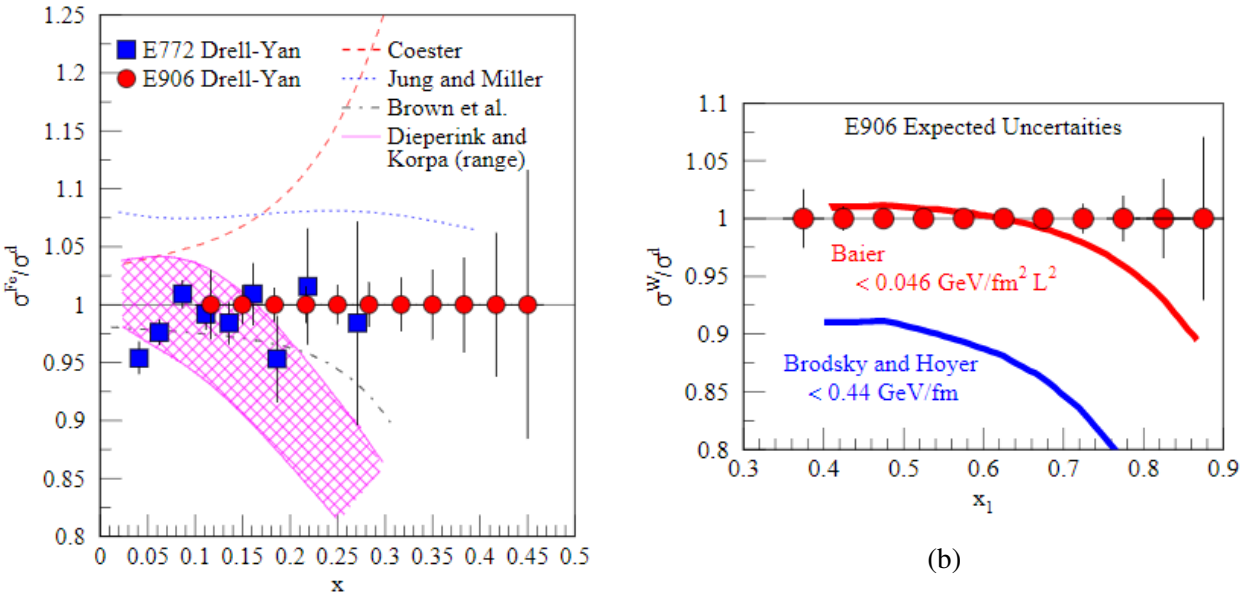


Figure 1.19: E772 R_{pA} ratios for C, Ca, Fe, and W with the EMC collaboration's R_{pSn} ratio [69]. The pion model, quark cluster model, and resealing model predictions are also plotted in the R_{pFe} panel.

This thesis seeks to study the nuclear dependence of the asymmetry of the sea. To that end the CSR of the nuclear targets to LD_2 are also analyzed along with the ratios to LH_2. Furthermore, machine learning and the standard analysis techniques are used as complementary analyses to derive the results.



(a) (b)

Figure 1.20: Left panel: E906 and E866 R_{pFe} as a function of x_2 with the model predictions of Coester, Jung and Miller, Brown *et al.*, and Dieperink and Korpa [68]. Right panel: R_{pW} as a function of x_1 with the energy loss model predictions of Baier and Brodsky and Hoyer.

Chapter 2:

SeaQuest Spectrometer

The SeaQuest experiment is housed in NM4, the former the KTeV building at Fermilab, receiving the 120 GeV Fermilab Main Injector (MI) proton beam. The proton beam begins life as hydrogen gas impinging off a molybdenum cathode. This generates H^- ions that are then accelerated to an energy of 35 KeV. This beam of ions travels into the radio-frequency quadrupole (RFQ), which accelerates the beam to 750 KeV and induces its bunch structure [70, 71]. This beam then passes into the Linear Accelerator (Linac), which uses drift tube cavities to accelerate the beam to 116 MeV and then side-coupled cavities accelerate the beam to 400 MeV. Next the Booster, a synchrotron accelerator, takes the beam and collides it with a carbon foil to strip its electrons off completely. This pure proton beam is then accelerated to 8 GeV through RF cavities and its final bunch structure, 53.1 MHz, is set [72]. Lastly, the beam travels from the Booster to the MI, where the beam is accelerated to 120 GeV and delivered to SeaQuest. A schematic of this process is shown in Fig. 2.1. For spatial reference throughout this discussion, the beam travels in the positive z-direction with upstream denoting being closer to the MI and downstream being the converse, the positive x-direction is to the left when looking down the beam direction, and consequently the positive y-direction is up.

The beam enters SeaQuest and travels through the targets generating dimuon pairs and hadronic detritus. The products travel through an iron core magnet, called FMag, that also serves as a beam dump. After passing through this magnet, the remaining particles traverse a set of hodoscopes and wire chambers and enter the open air analysis magnet, called KMag. This magnet provides a

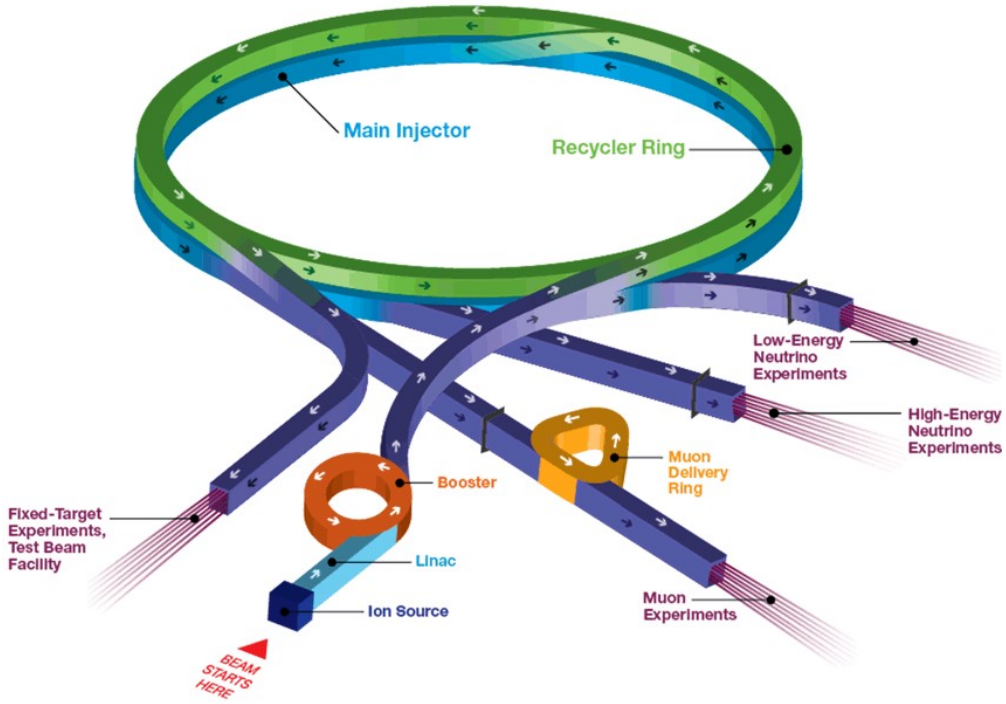


Figure 2.1: Fermilab Accelerator Complex illustrating the Main Injector, Booster, and beam lines [73].

momentum kick and the particles travel through a second set of hodoscopes and wire chambers. In the last step, the particles travel through a third and fourth set of hodoscopes and wire chambers with an iron wall to absorb any lingering hadrons in between. Lastly, the dimuons and pass through the proportional tubes. The data from the detectors is read out, muon tracks are reconstructed, and kinematic information derived. A diagram illustrating the beam passage through SeaQuest a can be seen in Fig. 2.2.

2.1 Beam

In every minute of operation, the beam delivered to SeaQuest is extracted in a 4.1 second spill. This is made possible by the alternating gradient focusing of the quadrupole magnets in the MI. This focusing of the beam engenders betatron motion, a transverse oscillation of the beam particles, and through betatron resonance the beam can be extracted [75]. The extracted beam contains the RF structure of the MI, with each RF ”bucket” being 2 ns long and each bucket being delivered every

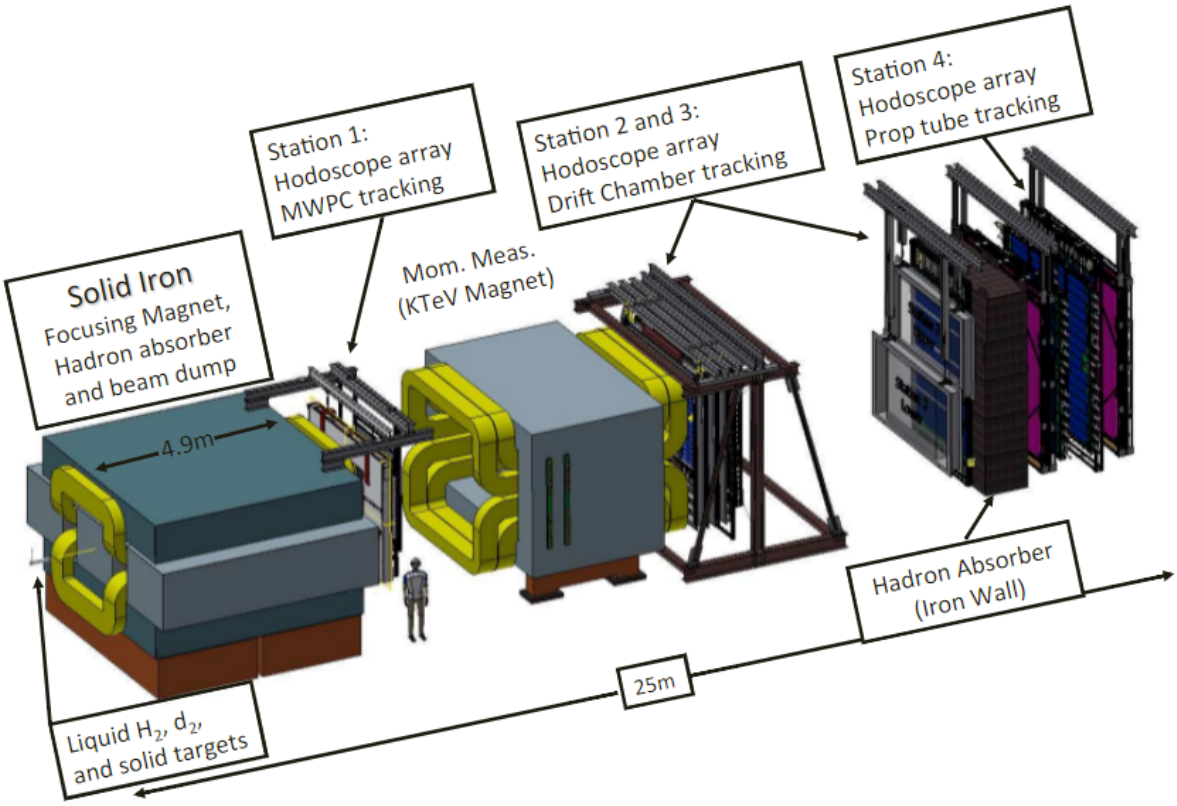


Figure 2.2: A schematic legend of the SeaQuest Spectrometer [74].

18.8 ns. The RF buckets are delivered in batches of 84 buckets and there are 7 of these batches in each MI turn for a total of 588 buckets. Only 492 of those buckets contain protons, with the other buckets being used to inject protons into the MI from the Booster or to allow buckets to be aborted during a run. During the 4.1s spill, the MI goes through 369,000 turns.

Unfortunately, the intensity of the beam can vary significantly within a spill, sometimes an order of magnitude difference between adjacent buckets. These high intensity buckets lead to lower trigger efficiency and it becomes more difficult for the tracking software to reconstruct events. SeaQuest is optimized for high mass, high p_T dimuon pairs, however, the dominant background, a combinatoric background of random single muon tracks that are reconstructed to mimic "true" dimuon pairs, is significantly correlated with intensity. This increased background at higher intensity dominates the possible triggers and gathering useful signal triggers becomes difficult. To combat this, a Beam Intensity Monitor (BIM) was developed to veto high intensity RF buckets.

The BIM generates an inhibit window, ± 9 RF buckets wide, around an RF bucket that violates a programmable threshold; this threshold is usually set between 65,000 and 95,000 protons per bucket [74]. The counting of the protons is done through the use of a Cerenkov detector. This detector uses an 80% Ar and 20% CO_2 gas mixture held at atmospheric pressure as the Cerenkov radiator [76]. The incident beam generates the Cerenkov radiation due to the protons travelling faster than the speed of light in the gas medium and this radiation is reflected off of a mirror into a photomultiplier tube (PMT). The PMT generates electrons based upon the number of photons received and this current is then relayed into a QIE (charge integrated encoder) chip, which transmits a signal every 18.8 ns, coinciding with the MI RF beam structure. The 8 bit output signal is binned and consequently contains a 1% RMS uncertainty [74].

2.2 Targets

SeaQuest uses two cryogenic targets, LH_2 and LD_2, and an empty flask for background subtraction. The flasks are 50.8 cm in length, 7.62 cm in diameter, and contain a volume of 2.2 L. In addition,

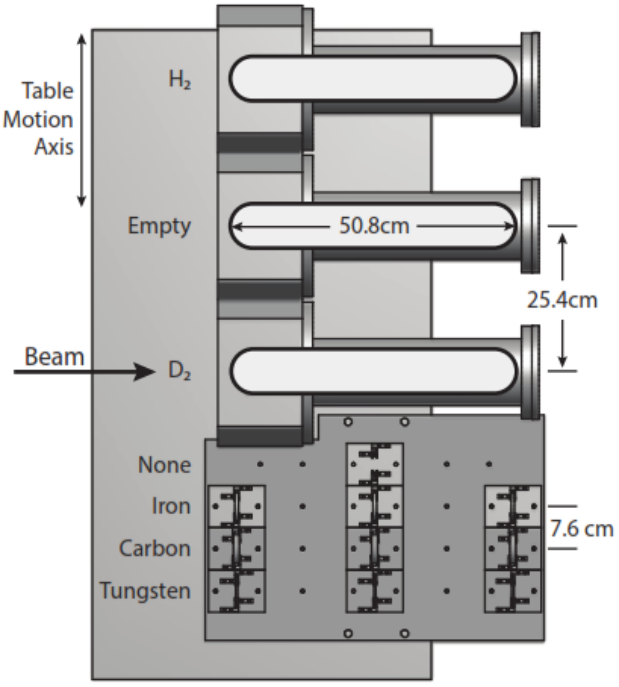


Figure 2.3: Schematic legend of the target table showing the LH_2, LD_2, and EMPTY flasks and the C, Fe, and W nuclear targets along with the NONE target [77].

there are three nuclear targets: carbon, iron, and tungsten, and a "none" target used for background normalization purposes. The solid targets are each divided into three 5.08 cm diameter disks with a distance of 25.4 cm between them, except for a time during Run II where the iron disk spacing was 17.1 cm. The solid target spacing was done to emulate the length of the liquid flasks and thereby minimize any acceptance differences between them.

The targets are housed on a target table, as shown in Fig. 2.3, that is positioned 130 cm upstream of the opening of the iron core analysis magnet. The specifications of the target are detailed in Table 2.3. The purity of hydrogen was reported to be ultra high purity, that is 99.999%, by Matheson, but there was some HD contamination in the deuterium gas [78]. During Runs I, II, and during $\approx$ 70% of Run III deuterium was supplied by FNAL, which contained an estimated 5-9% contamination, with the rest of Run III using ultra purity deuterium. The average composition of deuterium during the contamination period has been proposed to be 95.8 $\pm$ 0.2 % D and 4.2 % H, which corresponds to 91.8% D_2 and 8.2% HD [79].

Target Position	Material	Density (g/cm^3)	Thickness (cm)	Material Int. Length (g/cm^2)	Number of Int. Lengths	Spills/ Cycle
1	LH_2	0.071	50.8	52.0	0.069	10
2	EMPTY	-	-	-	0.0016	2
3	LD_2	0.163	50.8	71.8	0.120	5
4	NONE	-	-	-	0	2
5	Fe	7.87	1.905	132.0	0.114	1
6	C	1.80	3.322	85.8	0.209	2
7	W	19.30	0.953	192.0	0.096	1

Table 2.1: Specifications for the liquid and nuclear targets [80]. Int. is an abbreviation for Interaction. The Spills/Cycle represents an average, for there were changes in the absolute numbers over different data taking runs. There is a slight number of interaction lengths for the EMPTY flask due to the interactions with the iron flask walls.

2.3 Magnets

FMag, shown in Fig. 2.4, is a 43.2 cm $\times$ 160 cm $\times$ 503 cm iron core magnet downstream of the target system. It is run at 2,000 A and generates a magnetic field of 1.8 T, which provides a momentum kick of 3.07 GeV/c [74]. FMag also contains a 5 cm diameter $\times$ 25 cm deep hole to improve target-dump separation. The iron core also serves as a beam dump, which helps to absorb most hadronic products. KMag is an open air magnet downstream of FMag that provides a momentum kick of 0.39 GeV/c from the 0.4 T magnetic field generated by the 1,600 A current. The magnetic fields were checked with Hall probes and calibrated using the reconstruction of the J/Ψ mass [81, 82]. For most of the data taking the magnetic fields of FMag and KMag were parallel, with the fields being aligned in the y-direction. Due to this alignment the $y-z$ plane is referred to the non-bend plane, since the magnetic field would not induce deflection in plane, and the $x-z$ plane is referred to as the bend plane.

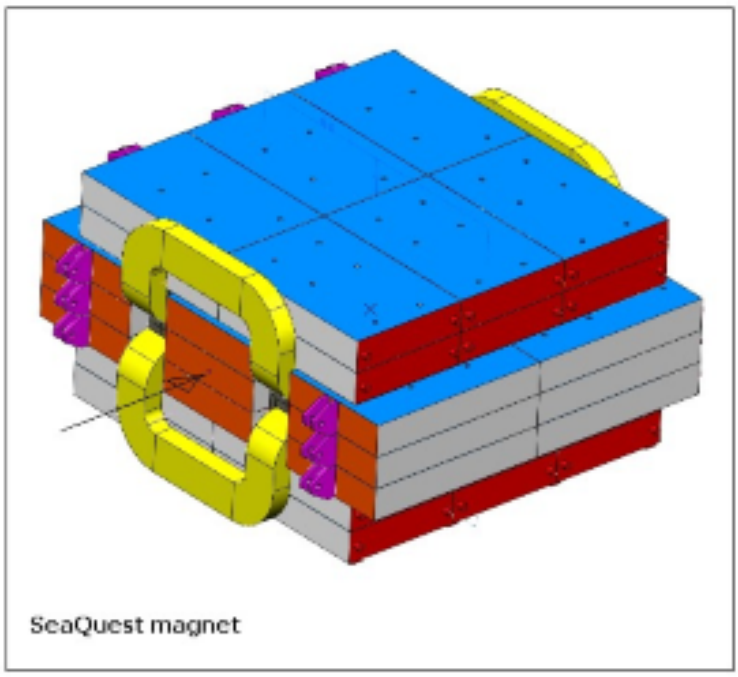


Figure 2.4: Drawing of FMag illustrating the iron slabs and aluminium coils [77].

2.4 Hodoscopes and Wire Chambers

There are 4 hodoscope stations as seen in Fig. 2.2. Stations 1 and 2 used scintillator paddles from HERMES and Stations 3 and 4 used Eljen EJ-200 panels [83, 84]. The vertically aligned hodoscopes provide x-direction resolution and hodoscopes aligned horizontally resolve the vertical direction. The hodoscopes within a plane have a slight overlap of 2-3 mm to improve efficiency at the cost of reduced background/signal resolution [83]. Each hodoscope panel is composed of scintillator material that emits photons when charged particles pass through it. These photons travel throughout the material and into the PMTs via waveguides at the end as seen in Fig. 2.5. Stations 1 and 2 have single $x - y$ planes of hodoscopes with 2.54 cm PMTs and Stations 3 and 4 have 5.1 cm PMTs [85, 74]. Stations 3 and 4 have one x-plane of hodoscopes each and Station 4 has two y-planes with PMTs at both sides. Despite there being y-plane hodoscopes for vertical resolution, they were not used for triggering. The specifications of the hodoscopes are listed in Table 2.2.

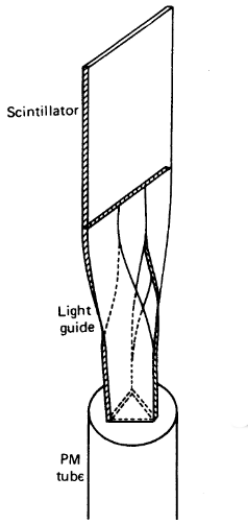


Figure 2.5: Schematic view of the end of a hodoscope paddle showing the wave guide connecting the scintillator material to the PMT [83].

Plane	Number	Length (cm)	Width (cm)	Thickness (cm)	Array Width (cm)	Mean Efficiency
1Y	20 X 2	78.7	7.32	0.64	140	-
1X	23 X 2	69.9	7.32	0.64	161	0.978
2Y	19 X 2	132.0	13.0	0.64	241	-
2X	16 X 2	152.0	13.0	0.64	203	0.989
3X	16 X 2	167.6	14.3	1.3	224	0.959
4Y1	16 X 2	152.4	23.16	1.3	366	-
4Y2	16 X 2	152.4	23.16	1.3	366	-
4X	16 X 2	182.9	19.33	1.3	305	0.979

Table 2.2: Hodoscope specifications [74]. The first number in Plane designates the station number and the letter the plane. The efficiencies of the Y hodoscopes are not listed since they were not used for triggering.

The PMTs generate current based upon the cascade of electrons. An incoming photon impinges on the photocathode, thereby generating an electron via the photoelectric effect. This electron passes through a focusing electrode, which accelerates it toward a primary dynode. Dynodes, electrodes held at a high voltage, release electrons once struck by electrons coming from the previous dynode. The electrons are accelerated by the electric field and in turn generate more electrons

once they strike secondary dynodes. This process generates an avalanche of electrons which finally strike the anode to generate a readout current. The gain of the PMTs, the final number of electrons divided by the number of incident electrons, was adjusted channel by channel to optimize trigger efficiency.

2.5 Drift Chambers

Stations 1, 2, and 3 each have a series of drift chambers. Each chamber holds 6 planes of wires: the X and X' planes are vertical, while the U and U' and V and V' planes are rotated by 14 degrees to the left and right, respectively. The primed planes are offset longitudinally by half a drift cell to remove the left-right ambiguity. Chamber nomenclature is done by having "DC" followed by the station numbers. DC3 is divided into two, DC3p and DC3m, which cover the upper half and lower half respectively. During data set 4, DC1 was replaced with DC1.2 and renamed DC1.1. Later DC1.1 was reinstalled upstream of DC1.2. DC3m was used for commissioning and data set 1, but was later replaced with DC3m.2, which is 25 cm wider on each side, thereby increasing the x_2 acceptance by 20% for $x_2 \approx 0.3$ and 10% for $x_2 \approx 0.4$. All of the chambers used a gas mixture of Ar:CH_4:CF_4 $(88:8:4)$, with the exception of DC1.2, which used Ar:CF_4:C_4H_{10}:$C_3H_8O_2$ $(81:5:12:2)$ [74].

The chambers consist of a series of gold plated tungsten wires held under tension. A voltage applied generates an electric field that envelopes the entire structure. Once a incident particle collides with a gas molecule with sufficient energy to ionize the gas, the ions drift under the electric field toward the wires. The electron drift velocity is significantly faster than that of the positive ions. The current caused by the drifting electrons and any secondary and tertiary electrons colliding with the anode is recorded, signifying a "hit". Using the time between the particle detection from the hodoscopes or proportional tubes and the drift chamber along with the electron mobility within the gas, the distance from the wires can be calculated [87]. The position resolution of each drift plane is at least 400 μm, which corresponds to a momentum resolution of Δp/p of 0.03·p (GeV/c).

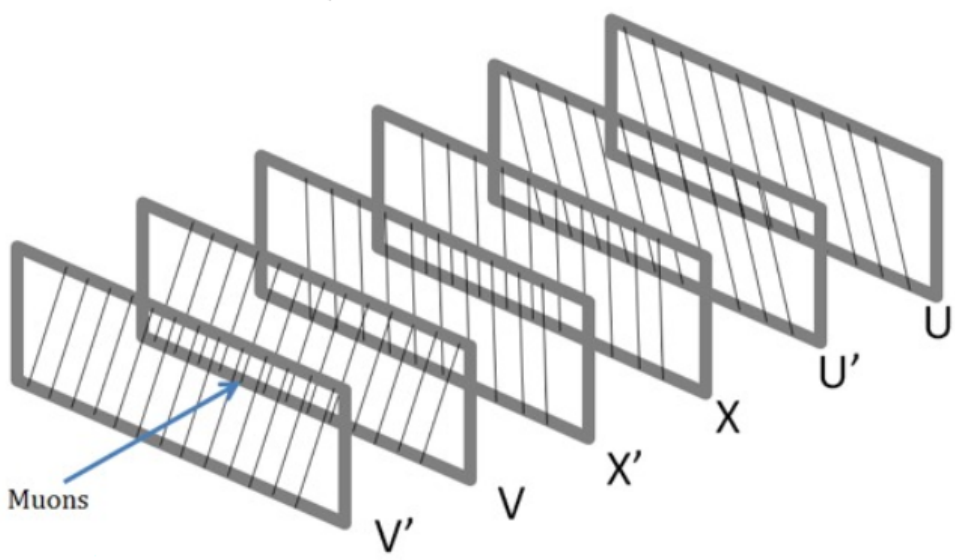


Figure 2.6: Drift plane layout of a chamber showing the X, U, and V planes along with their primes counterparts [86].

In addition, the single plane efficiency is greater than 95%, yielding a tracking efficiency of 90% [86].

The efficiency of the chambers is not constant throughout the chamber, with the efficiency at the center of the planes being higher than the efficiency the edges of the planes. This effect is due to after-pulse removal and the high hit rates of the detectors [88]. Detector efficiency is defined as

$$\epsilon = \frac{N_{eff}}{N_{eff} + N_{ineff}}, \tag{2.1}$$

with N_{eff} is the number of efficient tracks and N_{ineff} is the number of inefficient tracks. An efficient track is a path through the spectrometer that has 18 associated hits, i.e. each wire plane has detected a hit within 3 position resolutions of that plane [89]. An inefficient track for a plane is when it lacks an associated hit, while all other planes have associated hits. Multiple hits on the same wire can also lead to inefficient tracks. When this happens pulse removal is preformed. Pulse removal, as described in Ref. [90], is when a wire experiences one hit from an associated track and an extra uncorrelated hit. If the extra hit is detected on the wire before the hit from the associated track is, the second hit, which corresponds to a muon track, is removed and if the primary hit

is outside the resolution window, the hit is seen as an uncorrelated hit. The likelihood of this happening increases with intensity. If the wire hit rate is too high the muon may come before the positive ions leave the sense wire. This would mean that the electric field is locally weaker than the normal field and therefore the gain is reduced, leading to weaker signals and possibly no detection [88].

Detector	Left Edge	Center	Right Edge
D1	94.2 ± 3.5 %	99.9 ± 0.2 %	97.5 ± 1.9 %
D2	96.9 ± 0.9 %	98.8 ± 0.7 %	98.3 ± 0.5 %
D3p	96.9 ± 0.6 %	98.1 ± 1.1 %	98.1 ± 0.6 %
D3m	98.8 ± 0.8 %	99.4 ± 1.0 %	99.4 ± 0.5 %

Table 2.3: Listing of the drift chamber efficiencies of different chamber parts from Run II [88].

2.6 Muon Identification and Proportional Tubes

Downstream of Station 3 is a muon identification wall, a set of hodoscopes, and a set of proportional tubes. The muon identification wall, a 1-meter thick iron wall, serves to block most remaining hadronic products and to identify muon tracks. Muons experience multiple scattering, a characteristic momentum based deflection when passing through dense materials as shown in Fig. 2.7 [91].

The proportional tubes are arranged in four layers of planes with each plane consisting of nine proportional tube modules. Each module contains sixteen 3.66 m long and 5.08 cm diameter wide tubes, staggered in two sub-layers [93]. The first and fourth planes of tubes are aligned horizontally to provide y-position resolution and the second and third planes are aligned vertically for x-position resolution. A schematic representation of this design is in Figs. 2.8a and 2.8b.

The proportional tubes behave as a single wire chamber, by using the same gas mixture as the chambers and reading out the current from the ionized gas caused by a muon track. However, the maximum drift time is larger than that seen in the chambers, with a maximum drift time of 650 ns.

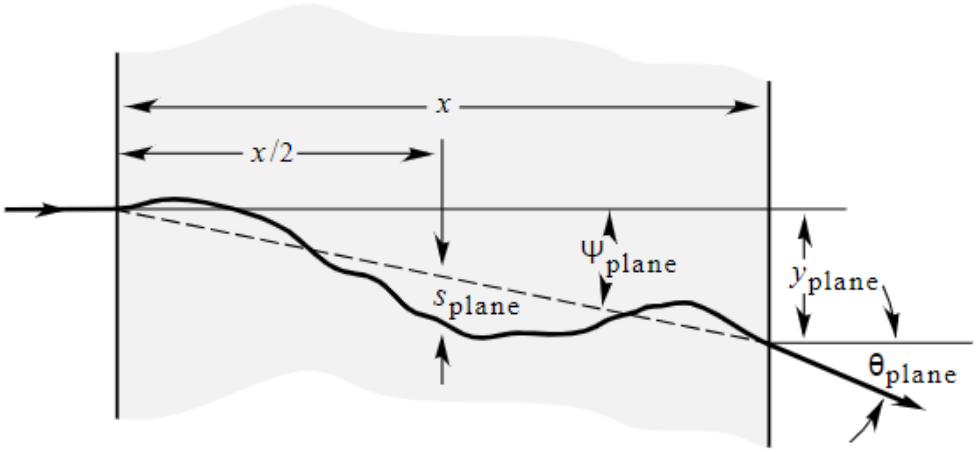
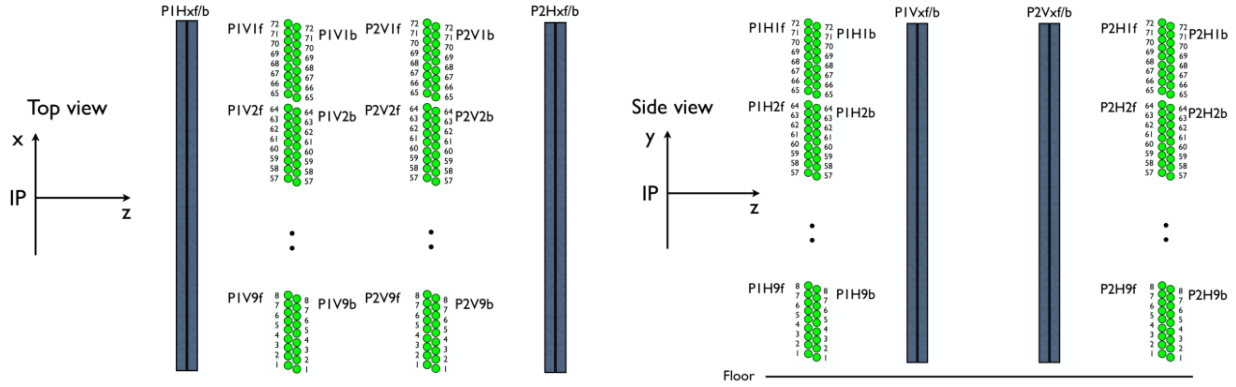


Figure 2.7: Schematic depiction of Coulomb scattering of a particle through matter [92].



(a) (b)

Figure 2.8: Schematic representation of proportional tubes. Left panel: bend plane view of the proportional tubes. Right panel: non-bend plane view of the proportional tubes [74].

A muon track is identified by 8 hits from the 4 proportional tube planes.

2.7 Trigger System

The trigger system consists of 9 CAEN V1495 VME modules with associated Field Gate Programmable Arrays (FPGAs) and a "Trigger Supervisor". This system is optimized for high mass dimuon detection, i.e. mass between 4.0 GeV/c^2 and 9 GeV/c^2. This means that tracks that correlate with events outside this mass range are suppressed. Within this trigger structure there are 3 levels: Level 0, Level 1, and Level 2. Level 0 operates on two possible modes: "Production" mode, which is the standard operating mode, and "Pulser" mode, where Level 0 generates hit patterns for testing. In the Production mode, the four V1495 modules, each connected to a hodoscope quadrant (upper bend plane, lower bend plane, upper non-bend plane, and lower bend plane), sends out the signals from the hodoscopes to Level 1 [74]. Level 1 also contains four V1495 modules each connected to a Level 0 module. Each module compiles four hit track candidates from each quadrant, groups them into bins, and outputs the information to Level 2 in the form of a 32 bit string. At Level 2 all the possible pairs of candidate tracks from Level 1 are generated given predefined selection rules and five output triggers are sent to the Trigger Supervisor.

Levels 0 - 2 all contain the same firmware: Time-to-Digital Converter (TDC) block, delay adjustment pipeline, and trigger matrix. However, Level 0 does not use its trigger matrix and contains some Pulser-mode firmware. The TDC block digitizes up to 96 channel input signals from the hodoscopes and stores the time of each hit with an appropriate offset for alignment with the MI RF clock [94]. The blocks generate a 212.4 MHz clock from the 53.1 MHz MI RF clock using a Phase Locked Loop with a 16/4 ratio to achieve TDC timing resolution of 1.177 ns by sampling all four clocks with each input signal [74]. Each of these TDC bins can only hold one hit. So only the latest hit is stored.

Each TDC sends the timing data, stored in 18.8 ns bins, to a set of RAM blocks, which serve as pipeline and event storage. Each channel can receive a relative delay value of 0-285 ns, i.e. 0

to 255 TDC bins, while the channel-by-channel reading and writing at the output port compensate for any individual offset, insuring simultaneous data output [94]. Once a global trigger is received, the pipeline ceases and the all TDC bins for each of the 96 channels are copied to the buffer. Each buffer contains a maximum of 256 hits and is read in order of the latest time slot; therefore any hits beyond 256 are not recorded [94]. The trigger matrix is comprised of a series of look up tables of logical elements. These logical elements are sets of "AND" and "OR" statements in reference to hodoscope element hits. The sets of hit patterns or roads are found from a FORTRAN based Monte Carlo (MC) simulation where the detector geometry, magnetic fields of FMag and KMag, and particle event generators of DY and charmonium are used to optimize DY acceptance. This Fast MC simulates the momentum and position of the muon tracks in a simplified manner, while the Full MC uses a Geant4 based physics simulation for "true" MC studies. There were 5 trigger matrices used in SeaQuest each with a different set of look up tables as shown in Table 2.4.

Matrix Name	Side	Charge	p_x Requirement	Trigger Specification
MATRIX 1	TB/BT	+-/-+	-	Main physics
MATRIX 2	TT/BB	+-/-+	-	Same-Side
MATRIX 3	TB/BT	++/–	-	Like-Charge
MATRIX 4	T/B	+/-	-	Singles
MATRIX 5	T/B	+/-	$p_x > 3$ GeV/c	High-p_T singles

Table 2.4: Trigger Matrix specifications [74]. "T" and "B" referring to top and bottom track coincidences. This means for instance, that MATRIX 1 finds coincidences of TB/BT and +-/+- signifying that a positive(negative) muon track in the top(bottom) part of the spectrometer is reconstructed as a pair.

2.8 Data Acquisition

Once the muons have traversed through the spectrometer and passed any trigger logic, their information from the hodoscopes, wire chambers, and proportional tubes must be read out and stored for analysis. To this end SeaQuest employs a data acquisition (DAQ) system and a slow control

system. There are three separate DAQs at SeaQuest: the Event DAQ, Scaler DAQ, and Beam DAQ. The Event DAQ is tasked with gathering information associated with event triggers, the Scaler DAQ monitors beam quality, and the Beam DAQ interfaces with the QIE board to read out intensity information.

2.8.1 Event DAQ

The Event or Main DAQ consists of 14 VME crates and one Trigger Supervisor (TS). Inside each crate is a VME CPU, a trigger interface card (TIR), and a set of TDCs. The TS can receive up to 12 different triggers with five of the triggers using the trigger logic from matrices, another five using the NIM modules, and the last two being the beginning of spill (BOS) and end of spill (EOS) triggers from the accelerator division [95].

The VME CPU or Readout Controller (ROC) reads out the TDCs when the TIR receives a trigger. The TS accepts triggers from the NIM module or Level 2 v1495 boards, sends signals to the other 14 crates via the TIRs, and a signal to the QIE system to record the intensity information in the 12-16 RF bucket window around the trigger. The TIRs then induce the ROCs to read out the events from the TDCs. Once the ROCs are finished they notify the TIR. The TIRs then sends information to the TS that they are finished and the TS waits until it receives the last TIR signal [96]. The ROC data is readout to the CODA (Cebaf Online Data Acquisition) program for later analysis and the QIE information read by a scaler-latch located on one of the crates. Once all of the TIRs have responded the TS can accept another trigger and this process can start again. The whole readout took approximately 150 μs during data sets 1-5 and was reduced by a factor of 5 for data sets 6 and 7 [74].

All of the triggers are not treated equally and some are scaled down by a certain factor. In practice this means that only a fraction of the non MATRIX 1 triggered events are accepted in order to keep their trigger rate less than 10 % of MATRIX 1's rate [97].

2.9 Scaler DAQ

The Scaler DAQ was designed to monitor the quality of the beam in real time and contains four different VME scalers. Scaler 1 is triggered in conjunction with a 7.5 kHz gate generator and beam spill signal. This scaler records the response from two hodoscopes and counts the intensity and squared intensity for duty factor calculations. The other scalers are triggered by the BOS or EOS signals, with the BOS initiating the counting of the spill-level rates and the EOS ending the count, reading out the data, and flushing the buffers [98]. The three pieces of information recorded are the number of Event DAQ triggers, beam intensity, and the hodoscope array rates.

2.10 Beam DAQ

The Beam DAQ reads the QIE information from the Cerenkov detector once the EOS signal is received. During each spill five pieces of information are recorded: the RF bucket intensity, the number of protons inhibited due to the BIM veto, the number of protons missed due to the trigger system being busy, the sum of the beam intensity, and sum of the squared intensity. To avoid double counting, the number of protons missed due to the trigger being busy ignores vetoed protons. The beam information is read out to a TDC and can be accessed through an Ethernet connection for real time analysis [95].

2.11 Slow Controls

The last component of the DAQ system is the slow controls, which is a set of coding scripts connecting to the EPICS (Experimental Physics and Industrial Control System) software program. The slow controls retrieve data on time scales longer than a spill and monitor data that describes the accelerator, target, and environmental conditions of the spectrometer apparatus [99]. The accelerator information details the intensity and quality of the beam, configuration of the accelerator, and the conditions of the magnets. Using an EPICS interface, the target information, including

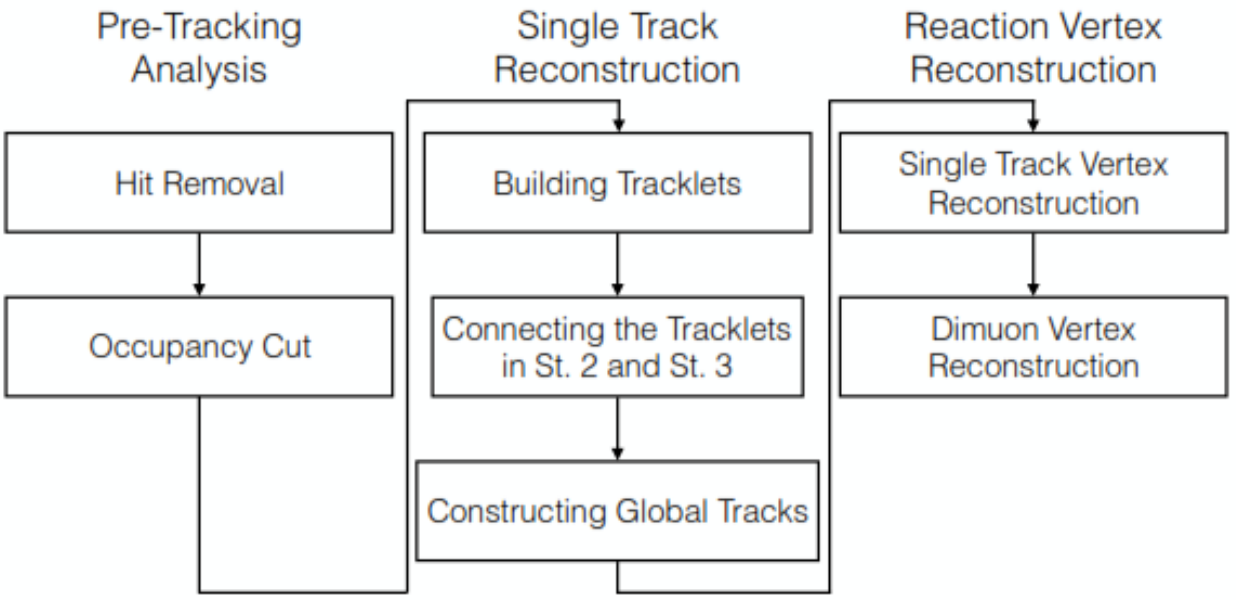


Figure 2.9: Flow diagram of the dimuon reconstruction process [90].

target rotation status and the temperature and pressure of the cryogenic targets, are recorded. The ambient temperature, humidity, and pressure of the experimental hall is monitored by a digital multi-meter and read out over Ethernet.

2.12 Tracking and Reconstruction

To generate the actual muon tracks from the hodoscope hit patterns and chamber hits, the tracking software, kTracker, is implemented. There are three main steps to tracking: pre-tracking analysis, single track reconstruction, and reaction dimuon reconstruction. A flow diagram illustrating this process is seen in Fig. 2.9.

2.12.1 Pre-tracking

Pre-tracking is the process of removing hits and high chamber occupancy events. There are three main types of hit removals: out-of-time hits, after-pulse hits, and cluster hits. Out-of-time hits are when a hit is recorded outside of the TDC window, which are removed since they cannot be properly associated with a track. After-pulse hits, which were described in Section 2.5, are

multiple hits on the same wire within the drift time window. Only the first hit will be used for track construction. Cluster hits are when a portion of a plane of wires are all excited. The extra hits could be a consequence of electronic noise, delta rays, or cell edge effects. All of these types of hits are removed [100].

If the intensity is too high, kTracker becomes less efficient, so a set of chamber occupancy cuts was developed to remove tracks associated with high intensity environments. This process is a local chamber occupancy cut in addition to the BIM veto and is applied after the hit removal. The occupancy cuts for Stations 1, 2, 3p, and 3m are 320, 160, 150, and 150 respectively [90].

2.12.2 Single track reconstruction

To make a full track, partial tracks, also called tracklets, in each station must be combined in a systematic way. Track construction starts with the generation of tracklets, which are the miniature tracks within a drift chamber. The process is initiated by finding a "hit pair" between struck X and X' wires where the difference in the elementID number, the number associated with the wire element within the chamber, is less than two. Physically this corresponds to a path between the adjacent X and X' planes with a small slope. From here a window of possible U hits defined by the geometry and maximum slope is created. After the U window is generated a final V window is found as depicted in Fig. 2.10. Due to the various windows of possible tracklets, multiple triplets may use the same set of hits. However, the total number is reduced by constraints eliminating triplets that: cannot be extrapolated to a neighboring triggered x-hodoscope, cannot be extrapolated back to the target, contain less than four hits, possess a $\chi^2 \geq 15$, or the number of hits within each individual window is one or two [90].

Once the tracklets are generated at Stations 2 and 3, they must be combined. All possible combinations of the two sets of tracks are attempted, with combinations discarded where the slopes and intersections are too large. Combinations were the X hits do not point toward the target or valid triggered proportional tubes are also rejected [100]. After these selection criteria an iterative χ^2

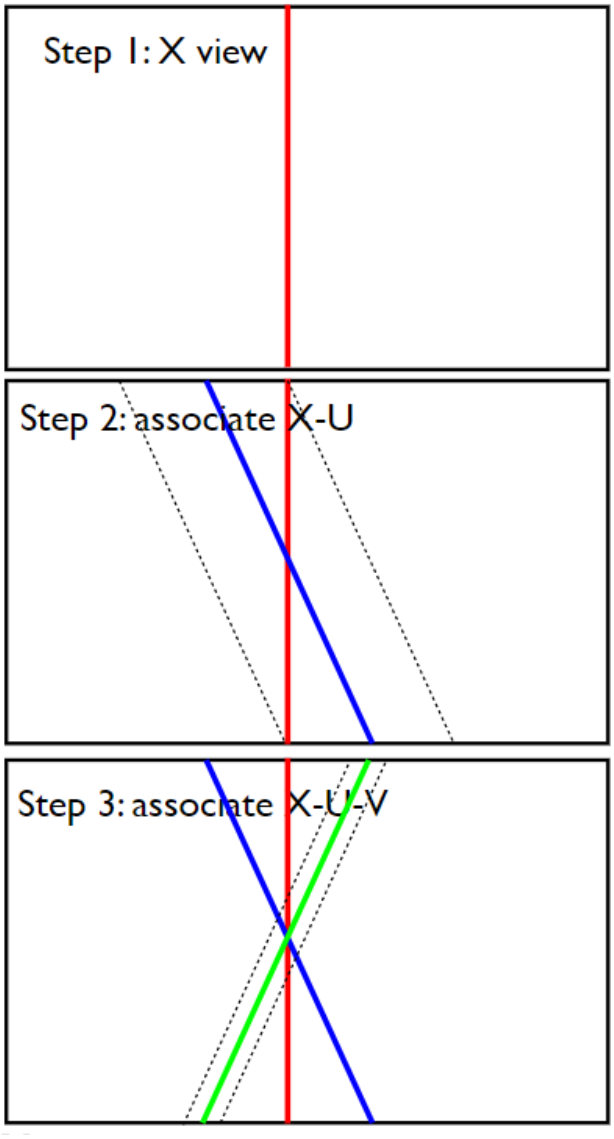


Figure 2.10: Schematic representation of the unprimed wire planes showing the construction of the U and V windows. Top panel: view of the X sense wire. Middle panel: construction of the U window. Bottom panel: construction of the V window [100].

and bad hit removal process is preformed. The χ^2 is defined as

$$\chi^2 = \sum^{N_{hits}} \frac{(w_h - w_t)^2}{r}, \tag{2.2}$$

with w_h and w_t are the position of the hit wire and the position of the estimated track, respectively and r is the local plane resolution [90]. Here an initial χ^2 test is preformed and tracks with $\chi^2 < 3,000$ are retained and then a "bad hit", a hit with more than three chamber resolutions away from the track, is removed and the χ^2 is recalculated. This process continues until all of the bad hits are removed and only tracks with $\chi^2 < 30$ remain. If two tracks contain one third of the same hits, the track with the smallest χ^2 is kept [100]. The tracks are also propagated to Station 4 to combine with proportional tube tracklets. The multiple scattering of muons inside the muon identification wall is corrected and the resultant tracks are constrained to the deflection angle be less than 30 mrad.

To merge the track with the Station 1 tracklet into a global track, a sagitta calculation is done. Sagitta is defined as the perpendicular distance between the center of the arc of a circle to the base of arc. Treating the position of the track at Stations 1 and 2 as the centers of two arcs and a line between the track at Station 3 and target as the base, two separate sagittas can be calculated as seen in Fig. 2.11. The sagitta is based upon the momentum of the track and the bend it feels due to the magnetic field, but the ratio of the sagittas of Station 1 to Station 2 is momentum independent and a sagitta ratio window of 1.85 ± 0.25 is used as the selection window for creation of the global track. The final global track then undergoes the same process of χ^2 and bad track removal as the previous Station 2-4 track, yielding a global track with a $\chi^2 < 20$, a p_T between 5 and 100 GeV/c, and comprising at least 14 hits [90].

The global track is then extrapolated through FMag and into the target. To construct the path through FMag, the energy loss in the iron and the bending due to the magnetic field must be taken into account. For this procedure FMag is divided into 100 slabs and the bending and energy loss is calculated in each slab using the results from the previous slab to reconstruct the momentum

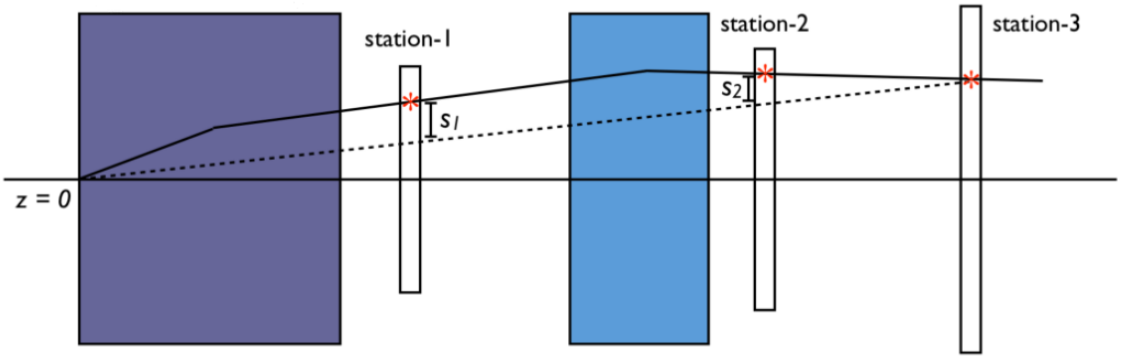


Figure 2.11: Schematic representation of sagitta calculation derived from the track positions at Stations 1, 2, and 3 [101].

and positron of the new slab. This process continues into the target region, wherein the target is divided into 2.5 cm slices. The final z-position of the track is taken to be the z-position of the slab which minimizes the distance between the track and the z-axis. A more detailed description of the energy loss calculations and single vertex construction can be found in Ref. [90].

2.12.3 Dimuon vertex construction

All of the single tracks are sorted into positive and negative tracks representing μ^+ and μ^-, respectively. Within the event window all combinations of the positive and negative tracks are combined and tested to see if they can produce a valid dimuon pair. The dimuon vertex position is found by searching for the target slice that minimizes the difference between the tracks and the distance between the z-axis and the tracks [90]. The preliminary vertex, along with the tracks are fed into a Kalman-Filter. A Kalman-Filter is a linear estimator that predicts where the next portion of the track should be based upon the previous track position, corrects the predicted position by minimizing the χ^2, and then smooths the tracks by back-propagating through the spectrometer. This process continues until the χ^2 converges, thereby finding the dimuon vertex [102].

Chapter 3:

Analysis

3.1 Data Taking Periods

The analysis in this thesis is based upon the data taking in two major periods: Run II and Run III. A Run is a large period of data taking which encompasses many roadsets, which are sets of individual runs of data taking within a determined spill range. A run of data is between 1 and 1.5 hours of beam time, which necessarily contains many spills. The five roadsets accompanied by their spill ranges are in Table 3.1, with roadsets 57 and 59 coming from Run II and roadsets 62 - 70 coming from Run III. The beam offset is the vertical offset between the $yD = 0$ and the average value of yD in each roadset [103], with yD being the y-position at $z = 32$ cm.

Roadset	Run range	Spill range	Beam Offset (cm)	Magnetic Field Orientation
57	8,912 – 10,420	310,955 – 370,099	0.4	B
59	10,421 – 10,912	370,110 – 388,469	0.4	B
62	11,075 – 12,435	409,547 – 482,571	1.6	B
67	12,525 – 15,789	484,746 – 676,223	1.6	B flipped
70	15,793 – 16,076	676,498 – 696,454	1.6	B flipped

Table 3.1: Table of roadsets with their respective run ranges, spill ranges, beam offsets, and magnetic field orientations. B flipped is the orientation opposite to the B field in roadsets 57, 59, and 62.

There are two measurements of the number of "good" protons that are delivered to the targets:

raw and live, with raw used in this analysis. Good here refers to spills that pass a number of quality cuts as outlined in Ref. [104]. Raw refers to the total number of protons delivered to the targets, while the live protons is defined as

$$Live_{G2SEM} \equiv \frac{G2SEM * (QIEsum - QIE_{Inhibit} - QIE_{Busy})}{QIEsum}, \tag{3.1}$$

with $G2SEM$ being the intensity per spill measured in protons from the secondary emission monitor in the G2 enclosure, $QIEsum$ is the total intensity in QIE units, the $QIE_{inhibit}$ is the number of inhibited buckets, and the QIE_{Busy} is the QIE intensity of the beam when the DAQ was busy [105]. The number of "good" protons delivered to the target (POT) is listed in Tables 3.2 - 3.7, taken from the R008 production list, which is an updated version of those seen in Ref. [104]. Using Raw POT or Live POT for the normalizations in the CSR calculations did not generate significant differences.

Target Position	Target	Total Raw POT	Total Live POT
-	All	1.81844×10^{17}	9.26687×10^{16}
1	LH_2	7.82876×10^{17}	4.01315×10^{16}
2	EMPTY	8.19295×10^{17}	4.46133×10^{16}
3	LD_2	4.13282×10^{17}	2.01604×10^{16}
4	NONE	1.24442×10^{17}	6.60844×10^{16}
5	Fe	8.23124×10^{17}	4.14517×10^{16}
6	C	2.49648×10^{17}	1.28741×10^{16}
7	W	8.39496×10^{17}	4.28783×10^{16}

Table 3.2: Table of POT delivered to the individual targets in Roadset 57.

3.2 Data Selection

Once the events have been recorded and dimuons reconstructed, the raw events must undergo constraints or cuts. For this analysis a standard set of analysis cuts (STD) and a set of machine

Target Position	Target	Total Raw POT	Total Live POT
-	All	5.17261×10^{16}	2.14969×10^{16}
1	LH_2	2.24758×10^{16}	9.42881×10^{15}
2	EMPTY	2.26658×10^{15}	1.01688×10^{15}
3	LD_2	1.11329×10^{16}	4.32787×10^{15}
4	NONE	4.46670×10^{15}	2.00673×10^{15}
5	Fe	2.31845×10^{15}	9.41273×10^{14}
6	C	6.78338×10^{15}	2.83152×10^{15}
7	W	2.28232×10^{15}	9.43782×10^{14}

Table 3.3: Table of POT delivered to the individual targets in Roadset 59.

Target Position	Target	Total Raw POT	Total Live POT
-	All	2.94576×10^{17}	1.35076×10^{17}
1	LH_2	1.25817×10^{17}	5.82119×10^{16}
2	EMPTY	2.43685×10^{16}	1.18501×10^{16}
3	LD_2	6.17881×10^{16}	2.70679×10^{16}
4	NONE	2.56435×10^{16}	1.23643×10^{16}
5	Fe	1.50687×10^{16}	6.62631×10^{15}
6	C	2.67396×10^{16}	1.22319×10^{16}
7	W	1.51500×10^{16}	6.72392×10^{15}

Table 3.4: Table of POT delivered to the individual targets in Roadset 62.

learning (ML) cuts are compared. In addition to the constrained experimental data (EXP) and a combinatoric background generated from the experimental data (JMIX), there were several MCs used to simulate the signal and the backgrounds: DY, J/Ψ (JPSI), Ψ' (PSIP), and a MC simulating DY originating from the dump (DUMP).

3.2.1 Standard Analysis Constraints

The STD cuts were developed as a rectangular based set of cuts motivated by physics, MC, and spectrometer knowledge. Rectangular in the sense that variable space that the constrained events

Target Position	Target	Total Raw POT	Total Live POT
-	All	8.12869×10^{17}	3.81306×10^{17}
1	LH_2	3.42718×10^{17}	1.61288×10^{17}
2	EMPTY	7.46929×10^{16}	3.66565×10^{16}
3	LD_2	1.71274×10^{17}	7.70196×10^{16}
4	NONE	7.45034×10^{16}	3.65750×10^{16}
5	Fe	3.73450×10^{16}	1.71730×10^{16}
6	C	7.46692×10^{16}	3.50934×10^{16}
7	W	3.76658×10^{16}	1.75003×10^{16}

Table 3.5: Table of POT delivered to the individual targets in Roadset 67.

Target Position	Target	Total Raw POT	Total Live POT
-	All	8.95383×10^{16}	4.20655×10^{16}
1	LH_2	3.86889×10^{16}	1.82450×10^{16}
2	EMPTY	7.87702×10^{15}	3.84128×10^{15}
3	LD_2	1.95208×10^{16}	8.85504×10^{15}
4	NONE	7.78976×10^{15}	3.79717×10^{15}
5	Fe	3.96850×10^{15}	1.83760×10^{15}
6	C	7.78891×10^{15}	3.66525×10^{15}
7	W	3.90442×10^{15}	1.82418×10^{15}

Table 3.6: Table of POT delivered to the individual targets in Roadset 70.

inhabit is some higher dimensional rectangle. The cuts began as entry 2111 on the document database and with each change the version of the cuts was updated until the final version, the forty-second iteration, was reached. Specifically the cuts, outlined in Ref. [106] and detailed in Appendix A, were constructed by analyzing the experimental data and a set of CTEQ6 MCs [90]. This methodology carries the advantage of using external physics knowledge and spectrometer information to form appropriate data restrictions, such as removing data in low acceptance regions. Though the STD methodology has its merits, there is also the possibility of bias due to preconceptions of how data should appear or how dependencies should form. Moreover, there is

Target Position	Target	Total Raw POT	Total Live POT
-	All	1.43055×10^{18}	6.72613×10^{17}
1	LH_2	6.07987×10^{17}	2.87305×10^{17}
2	EMPTY	1.17398×10^{17}	5.78261×10^{16}
3	LD_2	3.05044×10^{17}	1.37431×10^{17}
4	NONE	1.24848×10^{17}	6.13516×10^{16}
5	Fe	6.69319×10^{16}	3.07234×10^{16}
6	C	1.40946×10^{17}	6.66962×10^{16}
7	W	6.73975×10^{16}	3.12800×10^{16}

Table 3.7: Table of POT delivered to the individual targets for roadsets 57, 59, 62, 67, and 70.

problem of pareidolia, seeing patterns in random data, and the conceptual difficulty of analyzing higher dimensional correlations and spaces. Lastly, there is the issue that rectangular based cuts are necessarily symplectic, meaning that whole sections of events are removed at the edges of the hyper-rectangle instead of acutely excising collections of events within the variable space.

3.2.2 Machine Learning Constraints

Data driven techniques, such as ML, do not contain the same biases since they do not rely on a background of physics or spectrometer information nor do they contain any preconceived notions of how the data should appear. In addition, ML can generate sets of constraints that need not be connected, effectively creating sets of rectangular cuts with Boolean statements in a transformed space. On the other hand, ML is not a perfect analysis technique and may be liable to other issues such as overtraining and poor optimization. On a similar vein, though the ML cuts are entirely reproducible from the same procedure, given a consistent methodology, different algorithms, pre-processing, and different parameterizations can yield different results. Lastly, there is no simple interpretation of the results or the methods, and the learning is only as good as the data used to construct the algorithms.

The goal of using ML was out compete the STD cuts by retaining DY events that the STD

cuts removed and by removing as much or more background events. The preliminary cuts used for the ML were based upon an earlier version of the STD cuts, the thirty-second vs. forty-second iteration. From that start, each individual cut was relaxed until 98% of each target MC passed the cut, with the full set of cuts listed in Appendix B.

Multiple ML techniques were studied with various architectures including linear discriminators, automated rectangular cuts, neural networks, support vector machines, Probability Density Estimators Foams (PDEFs) and Boosted Decision Trees (BDTs). Deep Neural Nets were also experimented with, but performed poorly due to the relative independent input number, training sample quality, and training sample size. The strongest classifiers were the BDTs and the PDEFs. While it is possible to perform a similar ML analysis using a different classifier, BDTs are uniquely suited due to the conceptual simplicity of their classification, their training and testing speed, and robustness to weak variables.

3.2.2.1 Boosted Decision Trees

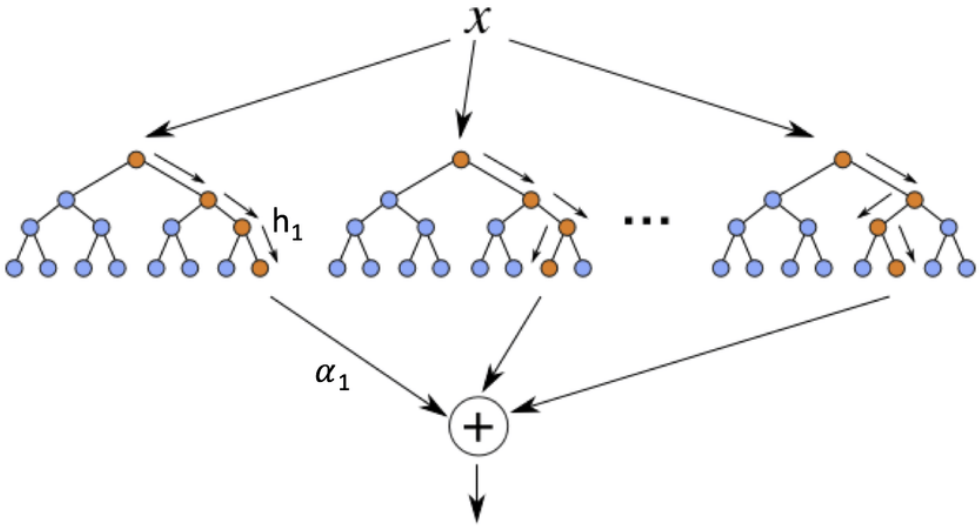


Figure 3.1: Schematic of a BDT regression illustrating the trees and their subsequent nodes from Ref. [107].

A BDT is a collection of binary decision trees that are used together in a classification forest. Data is divided randomly into a training sample and a testing sample. A sample of events in the

training set is selected to become the root node of one of the trees. That sample is then split based upon some splitting criteria into two daughter nodes. Then those daughter nodes are subsequently split; this process continues until the requisite depth is reached. A schematic of this process is illustrated in the Fig. 3.1. The splitting is accomplished with a separation function that is symmetric in signal and background such as

- Gini Index: $p \cdot q$
- Cross entropy: $-p \cdot \ln p - q \cdot \ln q$
- Misclassification of error: $1 - max(p, q)$
- Statistical significance: $\frac{S}{\sqrt{S+B}}$,

with p and q being the fraction of the events that are signal (S) and background (B), respectively [108]. Branches are split based upon which variable optimizes the information gain between the parent node and the sum of the two daughter nodes multiplied by the fraction of the number of events passing that cut,

$$Gain = F(p,q)_{Total} - \sum_{node_i, A} \frac{N_A}{N_{node_i}} \cdot F(p,q)_{node_i}. \tag{3.2}$$

$F(p, q)$ is the splitting algorithm, A being whatever variable is used to split at this node depth, and N_A and N_{node} being the number of events that passed/failed the cut and the number of events at that depth, respectively. The actual cut value is found by searching through the variable's range within that node, i.e. a variable may appear in multiple nodes within a tree, but the ranges may differ due to previous node's cuts.

The response of the forest is a weighted voting, or boosting, of the constituent trees. There are three main boosting algorithms: adaptive boosting, gradient boosting, and bagging [108]. In adaptive boosting, after a tree is constructed, miscalculated events, i.e. signal events calculated to be background or vice versa, are reweighted and used in a new tree. This process continues as trees are generated to work on increasingly difficult events. For gradient boosting a loss function

between the predictions and input data is generated after the tree is created and a gradient of this function is performed. A regression tree is then created whose output nodes are matched to the average value of the gradient in each region with the tree structure. This region is a hyper-cube of cuts that the path down the tree ”creates”. The loss function is then updated with a weight multiplied by the new regression tree and this process iterates. Lastly, bagging is simply a re-sampling of the data, through either explicitly picking a different random sample of the data or through reweighing events randomly. A new decision tree is then generated after each ”bag”.

The number of trees in the forest and the maximum depth are initialized before classification and the ML results are stored in a weight file. This weight file contains the blueprint of the information learned by the ML algorithm and is used to evaluate new events. To assess new data, the weight file takes in data and using the previous learning, rates the event on a scale of -1 to 1. The closer the value is to -1 the more background-like it is and the closer it is to 1 the more signal-like it is. The optimal value is found by maximizing the statistical significance.

For this analysis adaptive boosting was used and the BDT specifications within the TMVA ROOT5 package were optimized. The number of trees and the depth were studied and thirty-five trees with a maximum depth of four worked best in terms of overfitting protection, background rejection, signal retention, and receiver operator characteristic (ROC). Overfitting and ROC are discussed in more depth in Section 3.2.2.3.3.

3.2.2.2 Variable Selection

One of the first steps in the analysis is the creation of a variable list. This analysis used multiple BDTs to first generate a series of target and background specific sets of variables and then sought to find the best combination of these sets to build a better classifier. Beyond the defaults given for running BDTs by TMVA, the forest contained thirty five trees with a maximum a depth of four, and cross entropy was used as the splitting algorithm. These parameters have shown to be a fairly good range from a study of BDT specifications.

Initially the data cuts contained a mass greater than 4.0 GeV/c^2, only used Roadset 67, and the

JPSI and PSIP data sets were combined due to low statistics. However, for the final analysis there was no additional mass cut, the full run range was used, and the JPSI and PSIP data sets were not combined. Data was downloaded as dimuons containing the full kTrack information from both tracks that were used to construct the dimuon pair. Then once downloaded, the dimuon was converted into the two kTracks that shared the same dimuon information. From here a pairwise product and division of all of the kinematic variables was done. This large set of possible variables served as the seed for the analysis. It should be noted that performing BDTs with just the dimuons yielded worse results than using the kTracks, suggesting that the ML is more sensitive to bad tracks than bad dimuons. Different types of variable sets were analyzed based upon how well they discriminated between signal and background.

3.2.2.3 Classification Power

When running any classifier in TMVA a ranking of the variables by their classification power is printed to the screen. There are two lists of rankings: classifier independent and classifier dependent. From these lists three main ranking lists were developed: independent, BDT dependent, and Fisher dependent. Fisher dependent rankings were based upon the Fisher discriminant algorithm. This algorithm seeks to minimize the in-group variance, while maximizing the intergroup variance.

3.2.2.3.1 Independent and Fisher Dependent Rankings In batches of 150 variables a BDT or a Fisher discriminant classification was run, and the rankings were recorded. This eventual set of variables was run through a program that only recorded nonlinear variables that held more discrimination power than their component variables, e.g. if $mass * x_T > mass$ and $mass * x_T > x_T$, $mass * x_T$ was kept. This lead to several hundred variables per target per background, and from here the 150 most powerful variables were fed through a program that calculated the average correlation and the communality of the variables.

Using TMVA to generate a correlation matrix between the variables, the average absolute value of the correlation per variable was calculated. The correlation matrix was generated from the signal

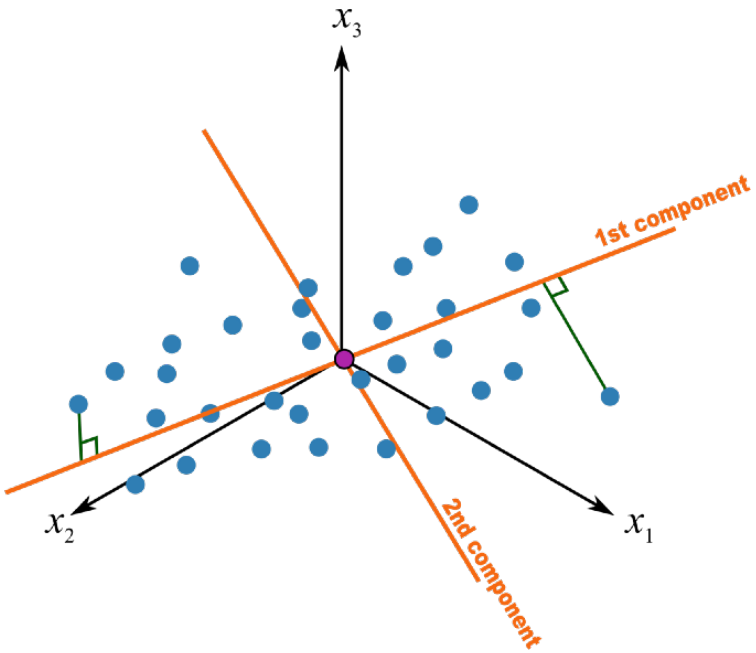


Figure 3.2: Schematic of a 3D principal component analysis illustrating the first two principal components [110].

and background data sets that were normalized so that they had the same average weight. The MC contains weighted events, so constructing the matrices would potentially skew toward the MC and away from the experimental data and JMIX due to their far higher weighted event integrals. Subtracting cross correlations was attempted, but the sum of the cross correlations was greater than the sum of the correlations, so it was not used. Using the correlation matrix and Ref. [109], the communality of the variables based upon the first 10 principal factors was calculated.

Principal Component Analysis (PCA) is the analysis of a data set based upon the variance of the data. The data set is rotated into a component space where the initial direction within the space is along the direction of greatest variance. Each subsequent direction is orthogonal to the previous space of directions and lies along the direction of most variance of the remaining data. The directions within this rotated space are the principal components. An example of a PCA is shown in Fig. 3.2.

The communality of a variable is a measure of how much of the variance of a variable is due to the number of principal components selected, e.g if there were 10 components used and if mass had a communality of 70, then 70% of the variance of the mass is accounted for by those 10

components.

The best variables for classification have large discrimination power, while low correlation/covariance and high communality. To whittle down the variables to a working set, a coefficient of merit, variable effectiveness (V_E), was constructed for correlation and communality,

$$\begin{aligned} V_{E,Corr.} &= \frac{100}{Correlation} \\ V_{E,Comm.} &= \frac{100}{100 - Communality}. \end{aligned} \tag{3.3}$$

The 20 variables per target per background with the greatest V_E were retained. In addition, another correlation method was used to develop an ancillary ranking of variables. In this method the variable with the highest $V_{E,Corr}$ is selected and then the variable who has the highest $V_{E,Corr}$ in relation to that first variable is selected, and then a third who has the highest $V_{E,Corr}$ with respect to the first two is selected. This process iterates until all of the variables are ranked.

3.2.2.3.2 BDT Dependent For BDT dependent rankings, the kinematic variables without the nonlinear combinations were used to develop a tree like structure where the groups of variables are continually split in half in each subsequent level. Based upon the nature of the dependent rankings, the actual ranking can vary depending on the other variables in the set. So splitting the variables into this structure and comparing the rankings along with different groups gives a better notion of which variables are most powerful. This process was done until a depth of four and then within each grouping the ranking was divided by the average ranking of that group. From some previous work [111], it turns out that the ratio of ranking to average ranking is constant after the root node. Only variables above the average ranking were retained and a pairwise product and division of all of the variables was computed.

A secondary set of BDTs were then run. In this run each batch of variables contained all of the linear variables and some number of nonlinear variables until the maximum number of variables in the batch was no larger than 50. 50 variables were chosen to minimize any possible dimensionality issues that may arise. After these were run, the discrimination power of all of the variables was

divided by the average power within the set and the average power of the linear variables was replaced by their respective average from each batch. Finally the top 20 variables were retained.

3.2.2.3.3 Evaluation After the variable selection, a collection of the different variable sets were run and evaluated in their ability to retain signal, remove background, and for how well they classify in general. The variable sets were:

- Independent Sets
  - Correlation (IC) : the individual correlation variables
  - Correlation 2 (IC2) : the individual new correlation variables
  - Communality (IM): the individual communality variables
  - Individual Power (II): individual most powerful variables
- Combined Independent Sets per Target
  - All Correlation (ICA) : all of the correlation variables
  - All Communality (IMA): all of the communality variables
  - All Correlation 2 (IC2A): all of the secondary correlation variables
  - All Power (IIA): all the most powerful variables
- Fisher Dependent Sets
  - Correlation (FC): the individual correlation variables
  - Communality (FM): the individual communality variables
  - Correlation 2 (FC2): the individual new correlation variables
  - Individual Power (FI): individual most powerful variables
- Fisher Combined Independent Sets per Target
  - All Correlation (FCA): all of the correlation variables

- All Communality (FMA): all of the communality variables
- All Correlation 2 (FC2A): all of the secondary correlation variables
- All Power (FA): all the most powerful variables

– BDT Dependent Sets

- Individual Power (BI): individual most powerful variables
- All Power (BA): all the most powerful variables per target

– Reference Sets

- Old All (OA): all of the kDimuon, kTracker, and STD analysis derived variables
- Old Correlation (OC): a reduced set of variables derived by looking at correlation matrix.

The results were sorted into three groups: Best Background (BB), Signal Retention (SR), and ROC (RC), representing the variable sets for each target that removed the most background, retained the most signal, and were the best at classifying, respectively. ROC is a measure of how well a classifier can separate signal from background. It is the integral of the background rejection vs. signal efficiency curve, illustrated in Fig. 3.3.

The greater the absolute value of the integral the more powerful the classifier, with a diagonal line signifying random chance. All of the data files were run through the weight files in each of the sets, i.e. the experimental data for the LH_2 target was run through the BB weight files for DY, DUMP, JMIX, JPSI, and PSIP. The ML values for each weight file along with the maximum efficiency cut (MEC) were stored in a newly created file. The MEC is the cut that maximizes the statistical significance.

3.2.2.4 Minimization Functions

The ML values are then analyzed to find the best set of cuts that maximizes DY and minimizes background. This region in five dimensional ML value space was optimized with a series of fitting

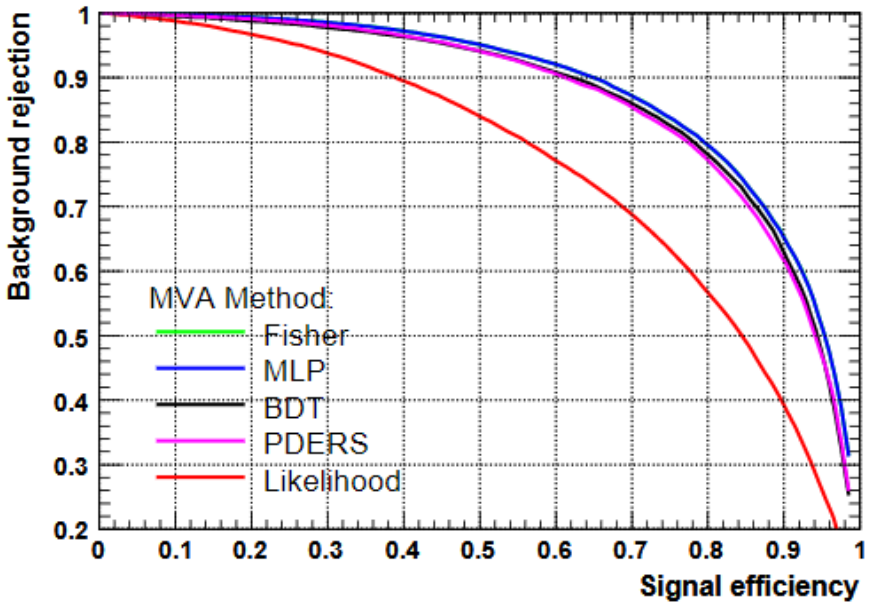


Figure 3.3: An example of the ROC curves for Fisher, MLP, BDT, PDERS, and Likelihood ML algorithms generated from a toy data set [108]. The curves are compiled from the background rejection, how well the classifiers remove background, and signal efficiency, how well the classifiers retain signal.

functions and distribution combinations. The over-line in the $\overline{JMIX}$ and $\overline{DY}$ denote that the distributions were scaled so that their integrals matched that of EXP.

- F_0 : Soft Max function using EXP and JMIX

  $F_0 = \frac{100}{1.0+e^{Data-\overline{JMIX}}}$

- F_1 : Derived Function 1 using EXP and JMIX

  $F_1 = 24 * EXP - 72 * \overline{JMIX} + 3 * EXP * JMIX - EXP^2 - \overline{JMIX}^2$

- F_2 : χ^2 function using EXP and EXP + JMIX

  $F_2 = \sum_i \left(EXP - \overline{JMIX}\right)^2$

- F_3 : Simple Subtraction using EXP and JMIX

  $F_3 = EXP * \left(\overline{JMIX} - EXP\right)$

- F_4 : χ^2 function using EXP and DY MC

  $F_4 = \sum_i \left(EXP - \overline{DY}\right)^2$

- F_5 : Derived Function 2 using EXP, JMIX, and all background MC

  $F_5 = \frac{-EXP*\overline{DY}}{\sum_i Background_i}$

– F_6 : Kolmogorov Smirnov Maximum Distance using EXP and DY MC

The functions were minimized using the CERN MINUIT program in ROOT6 and the best set of ML values characterizing the boundaries of the ML value space were recorded. All of the distributions used the full target range. For the JMIX set, the full occupancy range, i.e. $D1 \in [0-300)$ was used. The JMIX set was developed in three parts: $D1 \in [0-65)$, $D1 \in [65-135)$, and $D1 \in [135-300)$ with the kTracks used to emulate the combinatoric background having occupancies within the specified ranges. The full set was used due to low statistics. The choice to use all of the targets was to not bias the resulting sets in favor of one target or another and due to the low statistics of the JMIX and JPSI sets. The ML values and the data distributions were separated into two sets and combined for each minimization.

In addition to the minimization functions, the data was also constrained with a series of combinations:

– Subtraction : events that failed either ML or dimuon cuts remained

– Union : events that passed either the ML or dimuon cuts remained

– Intersection : the events that passed both ML and dimuon cuts remained.

These combinations served as a further constraint on the ML cuts. The best results are found empirically from the minimization functions and the combinations. The differences in success of the combinations and minimization functions is due to a combination of the ML training and the usage of all of the targets in the minimization functions. By using all of the target data some of the ML variable space will favor one target over another. This is compounded by there being more events in the liquid targets than the solid targets.

3.2.2.5 Final Sets

The final ML sets are the culmination of all of the methods listed in Sections 3.2.2.1 - 3.2.2.4 and found by mass fitting the distributions using ROOFIT and comparing each liquid and nuclear

target to the STD mass fitting results. There was an issue of binning with some of the ML models. The mass fits were initially performed using 50 bins, but when looking at the same fits with 100 bins the quality of the ML results, such as the amount of DY found, changed significantly. This was ameliorated by finding the best set for each target that was comparable to the STD in terms of DY and JMIX, with small to negligible JPSI and PSIP in the DY region, and binning stability. The three bests sets were ML_0, ML_1, and ML_2. The ML_0 set is the compilation of all of the best preforming sets for each individual target and the other two sets were found to have acceptable results for all of the targets, but had more relaxed cuts. In ML_0, the Fe and W targets used ML trained on LD_2 and LH_2, respectively. This suggests that the ML trained on those targets yielded poorer results than expected. For the CSR results and the asymmetry ratio extraction, the ML_2 set is used as the central value and the ML_0 set is used as a ML systematic uncertainty on the choice of ML model. This choice was made due to the ML_2 set faring better in terms of the CSR extrapolation.

Lastly, there is the question of overfitting, which is essentially ML tunnel vision. An algorithm becomes overfit when it as been trained so closely on a collection of data that it cannot adequately apply this learning to new data. This issue was tackled two fold. First the BDT specifications used, the thirty-five trees with a depth of four nodes, was shown to pass the overfitting checks in general in TMVA. This test is a Kolmogorov Smirnov (KS) probability test between two signal distributions and between two background distributions. Secondly, a series of three dimensional KS and χ^2 tests were performed with 25 bins per dimension by taking random samples of data and comparing the ML value distributions. These results were compared to taking a series of dimuon variables such as mass, x_T, x_B, p_T, and dz, and performing the same procedure. The results for the ML values mirrored those of the dimuon variables with χ^2/DOF being around unity and the KS probabilities were well above 0.05. Any lingering ML uncertainties are assumed to be taking care of from the use of an ancillary ML model as a systematic uncertainty on the model choice.

3.2.2.6 ML_0

This set is comprised of the best individual sets for each target as mentioned in Sec. 3.2.2.5. The specifications of the constraints used for each target are are listed in Table 3.8.

Target	Fun.	Comb.	ML Target	Type	DY	JPSI	PSIP	JMIX	DUMP
LH_2	F_2	Uni.	LH_2	BB	OC	OC	OC	OC	OC
LD_2	F_1	Sub.	LD_2	BB	OC	OC	OC	OC	BA
Fe	F_2	Int.	LD_2	BB	OC	OC	OC	OC	BA
C	F_4	Int.	C	SR	OC	OC	OC	OC	OC
W	F_3	Int.	LH_2	BB	OC	OC	OC	OC	OC

Table 3.8: ML_0 specifications detailing the minimization functions (Fun.) used, the set combination (Comb.), the target the ML was derived from (ML Target), the type of variable set (Type), and the individual target variable set for each data set. Uni., Sub., and Int. refer to the union, subtraction, and intersection combination sets, respectively. For the ML type, BB is the best background set and SR is the signal retention set. Lastly, for the individual target sets, OC is the Old Correlation set and BA is the All Power BDT Dependent set.

At first glance it would appear that the specifications of targets are the same, but even those who share ML Target, Type, and distribution sets, the variable spaces are different owning to the different minimization functions and data combinations. Of note is that the liquid targets are favored for training sets based upon this analysis. This is probably a consequence of better statistics for the liquid than the nuclear targets. The resulting ML value cuts are listed in Table 3.9, with the understanding that the ML values must be greater than the cut listed.

3.2.2.7 ML_1 and ML_2

Being that the ML_1 and ML_2 models were developed for all targets, the specifications and cut values are shared. The specifications of the models are listed in Table 3.10 and the cut values are detailed in Table 3.11.

Target	DY_{Val}	$JPSI_{Val}$	$PSIP_{Val}$	$JMIX_{Val}$	$DUMP_{Val}$
LH_2	-0.04820	0.09279	-0.05231	-0.06148	-0.6497
LD_2	-0.04993	-0.05000	-0.05386	-0.05000	0.1768
Fe	-0.05004	-0.05051	-0.05084	-0.05007	-0.05094
C	0.05561	-0.02399	-0.05004	-0.04999	-0.04999
W	-0.9998	-0.05008	-0.05065	-0.05007	-0.05024

Table 3.9: ML_0 cut values for each target.

ML Set	Fun.	Comb.	ML Target	Type	DY	JPSI	PSIP	JMIX	DUMP
ML_1	F_4	Sub.	Fe	RC	OA	BA	OA	BA	OA
ML_2	F_0	Int.	LIQ	BB	OC	OC	OC	OC	OA

Table 3.10: ML_1 and ML_2 specifications detailing the minimization functions (Fun.) used, the set combination (Comb.), the target the ML was derived from (ML Target), the type of variable set (Type), and the individual target variable set for each data set. LIQ is the combination of both liquid target sets. Sub. and Int. are the subtraction and intersection combination sets, respectively. For the ML type, BB is the best background set and RC is the ROC set. Lastly, for the individual target sets, OA is the Old set, OC is the Old Correlation set, and BA is the All Power BDT Dependent set.

ML Set	DY_{Val}	$JPSI_{Val}$	$PSIP_{Val}$	$JMIX_{Val}$	$DUMP_{Val}$
ML_1	-0.1756	-0.1682	-0.9505	-0.7715	-0.2874
ML_2	-0.05142	-0.5698	-0.05001	-0.05000	-0.05018

Table 3.11: ML_1 and ML_2 cut values.

3.3 Dimuon Yields

The cuts listed in the Appendices A and B are used for the initial downloading from the MySQL database servers, but for final CSR analysis, some additional constraints were applied. A listing of these shared cuts are shown in Table 3.12, and a description of all of the variables can be found in Ref. [112].

$costh$ is the cosine of the angle between the line connecting the dimuon pair in the dimuon rest

Variable	Constraint Range
mass	$[4.5 - 8.8)$
x_T	$[0.13 - 0.45)$
x_B	$[0.0 - 1.0)$
x_F	$[-0.1 - 0.95)$
$costh$	$[-0.5 - 0.5]$
$D1$	$[0 - 400)$
$D2$	$[0 - 400)$
$D3$	$[0 - 400)$
$D1 + D2 + D3$	$[0 - 1{,}000)$
I_T	$[0.0 - 64{,}000.0)$

Table 3.12: Table of ancillary constraints for the CSR calculation. See text for details.

frame and the line bisecting that muon line [113]. $D1$, $D2$, and $D3$ are related to the occupancies of chambers DC1, DC2, and the sum of DC3p and DC3m, respectively. Trigger intensity (I_T) is a scaled measure of the intensity in the RF bucket the event originated from and has the units of number of protons per spill ($N_p/spill$). Explicitly,

$$I_T \equiv \frac{(RF_0 - pedestal) * G2SEM}{QIEsum - N_{RF} * N_{turns} * pedestal}, \tag{3.4}$$

with RF_0 being the intensity in QIE units of the bucket that the triggered event came from [114], pedestal is the mean RF_0 of the spill, which was set to be 34 [115], N_{RF} is the total number of RF buckets per MI turn (588), and N_{turns} is the number of MI turns (369,000).

For the raw dimuon yields, all of the relevant cuts for STD and ML are applied. The yields are shown in terms x_T and I_T to reflect the eventual CSR calculations.

3.3.1 Standard Analysis Dimuon Yield

In Tables 3.13 - 3.18 are the dimuon yields for each target, calculated with the STD constraints and the ancillary constraints listed in Table 3.12. The tables are delineated by x_T bin and the eight I_T bins.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	434	9	523	8	205	188	227
[8,000 − 16,000)	536	21	618	8	277	242	285
[16,000 − 24,000)	485	10	563	7	220	209	279
[24,000 − 32,000)	407	14	496	7	152	173	240
[32,000 − 40,000)	273	7	333	9	156	151	168
[40,000 − 48,000)	213	7	247	5	109	98	108
[48,000 − 56,000)	128	6	166	2	69	59	113
[56,000 − 64,000)	122	3	123	1	36	50	50
[0 − 64,000)	2598	77	3069	47	1224	1170	1470

Table 3.13: STD dimuon yields per I_T bin for $x_T \in [0.13 - 0.16)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	874	16	1104	15	430	387	502
[8,000 − 16,000)	1221	34	1354	17	582	503	625
[16,000 − 24,000)	1007	30	1207	13	505	441	605
[24,000 − 32,000)	890	12	951	15	391	357	507
[32,000 − 40,000)	610	22	669	18	324	248	376
[40,000 − 48,000)	458	10	480	9	193	169	255
[48,000 − 56,000)	352	7	332	3	158	142	180
[56,000 − 64,000)	231	9	233	6	95	84	128
[0 − 64,000)	5643	140	6330	96	2678	2331	3178

Table 3.14: STD dimuon yields per I_T bin for $x_T \in [0.16 - 0.195)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	1021	21	1226	19	510	452	600
[8,000 − 16,000)	1301	42	1654	21	684	554	872
[16,000 − 24,000)	1258	39	1508	21	641	561	777
[24,000 − 32,000)	1070	46	1147	26	508	456	614
[32,000 − 40,000)	798	26	807	27	383	356	491
[40,000 − 48,000)	562	21	629	14	290	257	350
[48,000 − 56,000)	406	14	408	8	182	162	221
[56,000 − 64,000)	309	9	272	7	109	130	148
[0 − 64,000)	6725	218	7651	143	3307	2928	4073

Table 3.15: STD dimuon yields per I_T bin for $x_T \in [0.195 - 0.24)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	705	19	879	19	395	295	429
[8,000 − 16,000)	945	29	1099	10	427	388	541
[16,000 − 24,000)	882	20	977	16	421	404	533
[24,000 − 32,000)	751	34	799	20	364	324	438
[32,000 − 40,000)	564	19	574	15	255	232	327
[40,000 − 48,000)	438	19	416	14	168	169	208
[48,000 − 56,000)	300	19	308	5	121	106	168
[56,000 − 64,000)	218	14	226	13	94	90	113
[0 − 64,000)	4803	173	5278	112	2245	2008	2757

Table 3.16: STD dimuon yields per I_T bin for $x_T \in [0.24 - 0.29)$.

3.3.2 Machine Learning Dimuon Yield

For the ML_0 dimuon yield tables, the EMPTY and NONE target yields are reported by using the same ML set used for the LD_2 target. For the CSR per variable bin per intensity bin calculations, the background subtraction and contamination corrections are performed with the same ML set. This means that for the ML_0 R_{pC} ratio, the C and LD_2 distributions were cut with their respective best sets, while the NONE target was cut with the C best set and the LH_2 and EMPTY sets were

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	427	7	499	4	195	178	244
[8,000 − 16,000)	550	10	671	9	288	240	334
[16,000 − 24,000)	494	11	596	13	245	225	305
[24,000 − 32,000)	415	17	508	8	205	168	245
[32,000 − 40,000)	295	12	341	9	151	140	179
[40,000 − 48,000)	236	9	241	5	97	98	148
[48,000 − 56,000)	195	12	165	5	78	74	91
[56,000 − 64,000)	138	10	122	8	46	46	67
[0 − 64,000)	2750	88	3143	61	1305	1305	1169

Table 3.17: STD dimuon yields per I_T bin for $x_T \in [0.29 - 0.35)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	206	8	247	2	98	93	127
[8,000 − 16,000)	298	8	304	3	101	129	167
[16,000 − 24,000)	256	5	276	4	125	99	134
[24,000 − 32,000)	183	6	228	6	93	76	128
[32,000 − 40,000)	148	7	135	2	66	64	76
[40,000 − 48,000)	114	10	111	9	43	33	56
[48,000 − 56,000)	83	5	81	9	29	27	36
[56,000 − 64,000)	61	4	53	5	23	16	36
[0 − 64,000)	1349	53	1435	40	578	537	760

Table 3.18: STD dimuon yields per I_T bin for $x_T \in [0.35 - 0.45)$.

cut with the best LD_2 set. Lastly, the ML was trained with kTracks, so the number of events in the bins can take half-integer values. This did not lead to any significant charge asymmetries in the resulting data. All the data presented for the ML has been scaled by 0.5 to be in line with the dimuon results from the STD cuts. Another way to interpret the results is that each kTrack leading to the dimuon has a weight of 0.5. On the whole, the ML sets have more events than the STD cuts as seen in Section 3.4.

3.3.2.1 Machine Learning Model 0

In Tables 3.19 - 3.24 are the dimuon yields for each target, calculated with the ML_0 constraints and the ancillary constraints listed in Table 3.12. The tables are delineated by x_T bin and the eight I_T bins.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	449	12.5	551	9	195	187.5	239
[8,000 − 16,000)	576.5	23.5	689	16	301.5	255.5	298.5
[16,000 − 24,000)	530.5	11	627	15.5	230	226.5	299
[24,000 − 32,000)	426	17	540	15.5	168.5	182	252.5
[32,000 − 40,000)	314.5	9	356	12.5	155	162.5	169
[40,000 − 48,000)	250	15.5	276.5	10	118.5	106	118
[48,000 − 56,000)	152	7	184.5	5	71	67	119
[56,000 − 64,000)	134	4.5	134.5	6.5	43.5	61	56
[0 − 64,000)	2832.5	100	3358.5	90	1283	1248	1551

Table 3.19: ML_0 dimuon yields per I_T bin for $x_T \in [0.13 - 0.16)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	903	20	1122	25.5	447.5	390	506
[8,000 − 16,000)	1276	39.5	1445	29.5	593	524.5	635.5
[16,000 − 24,000)	1058	36.5	1248	25.5	514.5	463.5	605
[24,000 − 32,000)	965	23.5	1012	27.5	406.5	366.5	513.5
[32,000 − 40,000)	672.5	28	723	24	326	272	372.5
[40,000 − 48,000)	518	16	538.5	16	198.5	198	285.5
[48,000 − 56,000)	380.5	13.5	366.5	17	154.5	154.5	190
[56,000 − 64,000)	264	10.5	244.5	14.5	88	98	136.5
[0 − 64,000)	6037	187.5	6699.5	179.5	2728.5	2467	3244.5

Table 3.20: ML_0 dimuon yields per I_T bin for $x_T \in [0.16 - 0.195)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	1023.5	29.5	1252	24	522.5	460.5	604
[8,000 − 16,000)	1365.5	54	1733.5	27	714	584	861.5
[16,000 − 24,000)	1310.5	58	1586	37	632.5	574.5	763
[24,000 − 32,000)	1152.5	56	1264.5	36	520.5	471.5	621
[32,000 − 40,000)	824.5	39	855	31	391	376.5	490
[40,000 − 48,000)	608	31.5	680	23	301.5	264.5	351
[48,000 − 56,000)	437	21.5	450	21.5	189.5	186	249
[56,000 − 64,000)	352.5	24.5	307	22	125	148.5	161
[0 − 64,000)	7074	314	8128	221.5	3396.5	3066	4100.5

Table 3.21: ML_0 dimuon yields per I_T bin for $x_T \in [0.195 - 0.24)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	713.5	25	912.5	22.5	386.5	321	421.5
[8,000 − 16,000)	998.5	37.5	1168.5	16.5	441	430.5	571
[16,000 − 24,000)	979	42.5	1064.5	30	442	434	566.5
[24,000 − 32,000)	894	57	907.5	34	370.5	338	488
[32,000 − 40,000)	693	34.5	679	32.5	287.5	275	359
[40,000 − 48,000)	529.5	37.5	509.5	34.5	209.5	195.5	248.5
[48,000 − 56,000)	382.5	26.5	378	21	148	141	188.5
[56,000 − 64,000)	299	26	290.5	26	105.5	112	141
[0 − 64,000)	5489	286.5	5910	217	2390.5	2247	2984

Table 3.22: ML_0 dimuon yields per I_T bin for $x_T \in [0.24 - 0.29)$.

3.3.2.2 Machine Learning Model 1

In Tables 3.25 - 3.30 are the dimuon yields for each target, calculated with the ML_1 constraints and the ancillary constraints listed in Table 3.12. The tables are delineated by x_T bin and the eight I_T bins.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	461	7.5	530	8	208	190.5	250.5
[8,000 − 16,000)	619.5	19	739	18	330	264.5	365
[16,000 − 24,000)	620.5	30.5	696.5	24.5	282.5	258	324
[24,000 − 32,000)	550	31	637.5	31	248.5	218.5	268
[32,000 − 40,000)	450	28	477.5	22	189	187.5	220
[40,000 − 48,000)	382	24.5	361	19	134	155.5	187
[48,000 − 56,000)	324	23	272.5	32.5	96.5	120	129.5
[56,000 − 64,000)	267	31.5	211.5	24.5	75	88	94
[0 − 64,000)	5489	286.5	5910	217	2390.5	2247	2984

Table 3.23: ML_0 dimuon yields per I_T bin for $x_T \in [0.29 - 0.35)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	231.5	7.5	282	2.5	103	96	133
[8,000 − 16,000)	357	16	351	12.5	115	141.5	177
[16,000 − 24,000)	331.5	13	365	11.5	155.5	121.5	164.5
[24,000 − 32,000)	319	23.5	333.5	26	125	111.5	145.5
[32,000 − 40,000)	288.5	21.5	223	16	91.5	101.5	94.5
[40,000 − 48,000)	271.5	28.5	215.5	25.5	73.5	75.5	87.5
[48,000 − 56,000)	204.5	23.5	155.5	28.5	47	59	54
[56,000 − 64,000)	174	18	121	34	41.5	48	51.5
[0 − 64,000)	2177.5	151.5	2046.5	156.5	752	754.5	907.5

Table 3.24: ML_0 dimuon yields per I_T bin for $x_T \in [0.35 - 0.45)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	484	14.5	578.5	12.5	211.5	206	244
[8,000 − 16,000)	628.5	25.5	715.5	19	321	275.5	314.5
[16,000 − 24,000)	563	13	652.5	17.5	253.5	243	320
[24,000 − 32,000)	462.5	20	579.5	19	180.5	190.5	268.5
[32,000 − 40,000)	336	10.5	399	16	176.5	170	185.5
[40,000 − 48,000)	267	16	300	11	124	114.5	125.5
[48,000 − 56,000)	171	7.5	201	5.5	76.5	72	125
[56,000 − 64,000)	140.5	4	139	6	47	66.5	58.5
[0 − 64,000)	3052.5	111	3565	106.5	1390.5	1338	1641.5

Table 3.25: ML_1 dimuon yields per I_T bin for $x_T \in [0.13 - 0.16)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	971	23	1201.5	28	469	415.5	538
[8,000 − 16,000)	1368	47	1551	37.5	621	556	690.5
[16,000 − 24,000)	1167	38.5	1374.5	31.5	557	506.5	661.5
[24,000 − 32,000)	1071.5	29	1091	31	432.5	404	568
[32,000 − 40,000)	742.5	32	797	29.5	349	295	410
[40,000 − 48,000)	559.5	22	585	22.5	216	214.5	310.5
[48,000 − 56,000)	427	12	395	19.5	164	168.5	216
[56,000 − 64,000)	289.5	13.5	271	18.5	101	109.5	147.5
[0 − 64,000)	6596	217	7266	218	2909.5	2669.5	3542

Table 3.26: ML_1 dimuon yields per I_T bin for $x_T \in [0.16 - 0.195)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	1110	32.5	1348	27.5	554	495	645.5
[8,000 − 16,000)	1493.5	60	1862.5	38.5	768.5	639.5	926.5
[16,000 − 24,000)	1467	67.5	1735	49	692	628.5	818.5
[24,000 − 32,000)	1270	61.5	1373.5	42.5	550.5	504.5	688
[32,000 − 40,000)	923.5	45.5	964	39	418	407.5	543
[40,000 − 48,000)	683	40.5	753	25	340.5	287.5	387
[48,000 − 56,000)	493	27	495.5	24.5	203.5	205.5	264.5
[56,000 − 64,000)	395	32.5	349.5	22.5	135	155	178.5
[0 − 64,000)	7835	367	8881	268.5	3662	3323	4451.5

Table 3.27: ML_1 dimuon yields per I_T bin for $x_T \in [0.195 - 0.24)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	795	26.5	1001	23	419.5	348	454.5
[8,000 − 16,000)	1100	42	1267.5	19	473.5	452.5	621
[16,000 − 24,000)	1072	46	1172	38	479.5	465	621.5
[24,000 − 32,000)	988.5	68	999	40.5	408	386	523
[32,000 − 40,000)	740.5	42	738.5	38	312.5	288	400.5
[40,000 − 48,000)	593.5	44	567.5	36	236.5	207	264
[48,000 − 56,000)	430.5	29.5	412.5	21	156.5	152	202.5
[56,000 − 64,000)	335	27	323	32.5	117.5	126.5	153
[0 − 64,000)	6055	325	6481	248	2603	2425	3240

Table 3.28: ML_1 dimuon yields per I_T bin for $x_T \in [0.24 - 0.29)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8{,}000)$	515.5	10	587	9	224	206	272
$[8{,}000-16{,}000)$	683	26	818.5	23.5	360	294	403.5
$[16{,}000-24{,}000)$	689.5	37.5	792	32.5	321.5	290	360
$[24{,}000-32{,}000)$	630	37	731	38	284	238	307
$[32{,}000-40{,}000)$	514	37	527.5	31.5	221.5	214.5	244.5
$[40{,}000-48{,}000)$	433	30	407	25	143.5	173	201.5
$[48{,}000-56{,}000)$	361	26	305	31	113	134	144.5
$[56{,}000-64{,}000)$	297.5	32	235.5	21.5	86.5	105	105
$[0-64{,}000)$	4123.5	235.5	4403.5	212	1754	1654.5	2038

Table 3.29: ML_1 dimuon yields per I_T bin for $x_T \in [0.29 - 0.35)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8000)$	264	8.5	313.5	5	116	106.5	150.5
$[8{,}000-16{,}000)$	414.5	19.5	401	15	129	165	196.5
$[16{,}000-24{,}000)$	406.5	18.5	433.5	11	179.5	142.5	182.5
$[24{,}000-32{,}000)$	397	32	405	32.5	151.5	131.5	172
$[32{,}000-40{,}000)$	376	26.5	271	21	115	128	114.5
$[40{,}000-48{,}000)$	352.5	35	261.5	33	91.5	95	106
$[48{,}000-56{,}000)$	264.5	27.5	207.5	40.5	54	75	73.5
$[56{,}000-64{,}000)$	219	29.5	151.5	48.5	54.5	51	65
$[0-64{,}000)$	2694	197	2444.5	206.5	891	894.5	1060.5

Table 3.30: ML_1 dimuon yields per I_T bin for $x_T \in [0.35 - 0.45)$.

3.3.2.3 Machine Learning Model 2

In Tables 3.31 - 3.36 are the dimuon yields for each target, calculated with the ML_2 constraints and the ancillary constraints listed in Table 3.12. The tables are delineated by x_T bin and the eight I_T bins.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8{,}000)$	462	12	561	9.5	202.5	203.5	240.5
$[8{,}000-16{,}000)$	591.5	26	685	14	311	262	303.5
$[16{,}000-24{,}000)$	534	11	634.5	15	242.5	230	309.5
$[24{,}000-32{,}000)$	441.5	16	551.5	14	172.5	189	265.5
$[32{,}000-40{,}000)$	325.5	9	381.5	13.5	170	163	175.5
$[40{,}000-48{,}000)$	257.5	14	286.5	10	116	112.5	121
$[48{,}000-56{,}000)$	155	6.5	191	3.5	72	65.5	119.5
$[56{,}000-64{,}000)$	140	5.5	131.5	5.5	43.5	62.5	57
$[0-64{,}000)$	2907	100	3422.5	85	1330	1288	1592

Table 3.31: ML_2 dimuon yields per I_T bin for $x_T \in [0.13-0.16)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8{,}000)$	918	20	1141.5	23	451.5	404	513
$[8{,}000-16{,}000)$	1289.5	40	1466	26.5	585	536.5	658
$[16{,}000-24{,}000)$	1085	34.5	1282.5	24.5	539.5	480	624
$[24{,}000-32{,}000)$	976.5	22	1027.5	25	396.5	373	535.5
$[32{,}000-40{,}000)$	685	29.5	736	25.5	328.5	281.5	381.5
$[40{,}000-48{,}000)$	513.5	15.5	539.5	15	206.5	200.5	287.5
$[48{,}000-56{,}000)$	387	10	368	14	159.5	159	209
$[56{,}000-64{,}000)$	254	11	251	13.5	92.5	100.5	140
$[0-64{,}000)$	6108.5	182.5	6812	167	2759.5	2535	3348.5

Table 3.32: ML_2 dimuon yields per I_T bin for $x_T \in [0.16-0.195)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8{,}000)$	1001.5	28.5	1234.5	19.5	501	453	589.5
$[8{,}000-16{,}000)$	1329	43	1696.5	28	701	576.5	843.5
$[16{,}000-24{,}000)$	1270	54	1541	31.5	623	570.5	742
$[24{,}000-32{,}000)$	1096.5	44	1217	31	506	454.5	596.5
$[32{,}000-40{,}000)$	790	34	849	24.5	377	355	486
$[40{,}000-48{,}000)$	564.5	28.5	668.5	15.5	301	256.5	349
$[48{,}000-56{,}000)$	426.5	20	442.5	18	174.5	180.5	233
$[56{,}000-64{,}000)$	345	22	285	14	126.5	137.5	158.5
$[0-64{,}000)$	6823	274	7934	182	3310	2984	3998

Table 3.33: ML_2 dimuon yields per I_T bin for $x_T \in [0.195 - 0.24)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8{,}000)$	723.5	22	910	17	394	322	425.5
$[8{,}000-16{,}000)$	987.5	33	1150.5	13	431	416	572.5
$[16{,}000-24{,}000)$	943.5	34.5	1046.5	24.5	447	422	571
$[24{,}000-32{,}000)$	842.5	45.5	882	26	362	338.5	473
$[32{,}000-40{,}000)$	637	26	639	29	281	253	355
$[40{,}000-48{,}000)$	481	29.5	501	23.5	199.5	173	237
$[48{,}000-56{,}000)$	357	25.5	353.5	16	138	127.5	181
$[56{,}000-64{,}000)$	267	20	258	22	95	106.5	131.5
$[0-64{,}000)$	5239	236	5740.5	171	2347.5	2158.5	2946.5

Table 3.34: ML_2 dimuon yields per I_T bin for $x_T \in [0.24 - 0.29)$.

3.4 STD and ML Yield Comparison

To get a better understanding about the differences between the ML and the STD cuts and how those cuts translate to some of the kinematic variables, a direct comparison is warranted. A comparison of the events that pass both sets of constraints along with the tables illustrating the events that only pass one set of the constraints is listed in Appendix D. In Fig. 3.4 is a comparison of

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	471.5	8.5	550.5	8.5	210	191	257
[8,000 − 16,000)	627	20.5	772	16	334	267.5	377
[16,000 − 24,000)	616	29.5	729.5	20.5	298.5	264.5	336
[24,000 − 32,000)	532.5	26	646.5	26	261	209	273.5
[32,000 − 40,000)	423	29.5	476	22	196.5	194	228
[40,000 − 48,000)	375.5	19	361.5	19	128	148.5	190.5
[48,000 − 56,000)	301	20	256.5	26.5	99.5	111	124.5
[56,000 − 64,000)	236.5	23	207	13.5	72	79	94
[0 − 64,000)	3583	176	3999.5	152	1599.5	1464.5	1880.5

Table 3.35: ML_2 dimuon yields per I_T bin for $x_T \in [0.29 - 0.35)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	250.5	7.5	297	3.5	106.5	100.5	143
[8,000 − 16,000)	372.5	12.5	379.5	12	124	151	188
[16,000 − 24,000)	341	15	395.5	9	165	128.5	170.5
[24,000 − 32,000)	334.5	22	355	26.5	139.5	112.5	157
[32,000 − 40,000)	297.5	18	233.5	14	99	101.5	100
[40,000 − 48,000)	286.5	31	225.5	26	81	73.5	88
[48,000 − 56,000)	205	22.5	177.5	32.5	53.5	64	59
[56,000 − 64,000)	173	18.5	120.5	34.5	41	42	54
[0 − 64,000)	2260.5	147	2184	158	809.5	773.5	959.5

Table 3.36: ML_2 dimuon yields per I_T bin for $x_T \in [0.35 - 0.45)$.

the mass yields between the STD cuts and the three ML models for the LH_2, LD_2, C, Fe, and W targets. The figures are plotted on a log scale to better illustrate the differences between the methods. From the figures it is clear that the ML models have more events than the STD cuts, with the largest differences being found in the LH_2 and LD_2 targets. Also of note is that there is no significant difference in the shapes of the mass distributions.

In Fig. 3.5 is a comparison of the x_T yields between the STD cuts and the three ML models

for the LH_2, LD_2, C, Fe, and W targets. Like the mass distributions, the figures are plotted on a log scale and the yields of the ML models surpass those of the STD cuts. However, unlike with the mass distributions, the ML shapes depart from the STD in the LH_2 and LD_2 targets. The largest difference is seen in the LH_2 target, where there are significantly more events in the $x_T > 0.3$ range. This is probably a consequence of there being more training data for the ML, both experimental and MC, for the liquid targets.

Fig. 3.6 is the last set of comparison plots between the STD cuts and the ML cuts. These plots show the I_T distributions and follow the same trend as the mass and x_T distributions; with ML having larger yields and the liquid targets faring relatively better than the nuclear targets. It should be noted that the MC and the JMIX sets do not have any I_T information, so training on this variable was not possible. They do possess some occupancy information, but that does not correlate well with the intensity information.

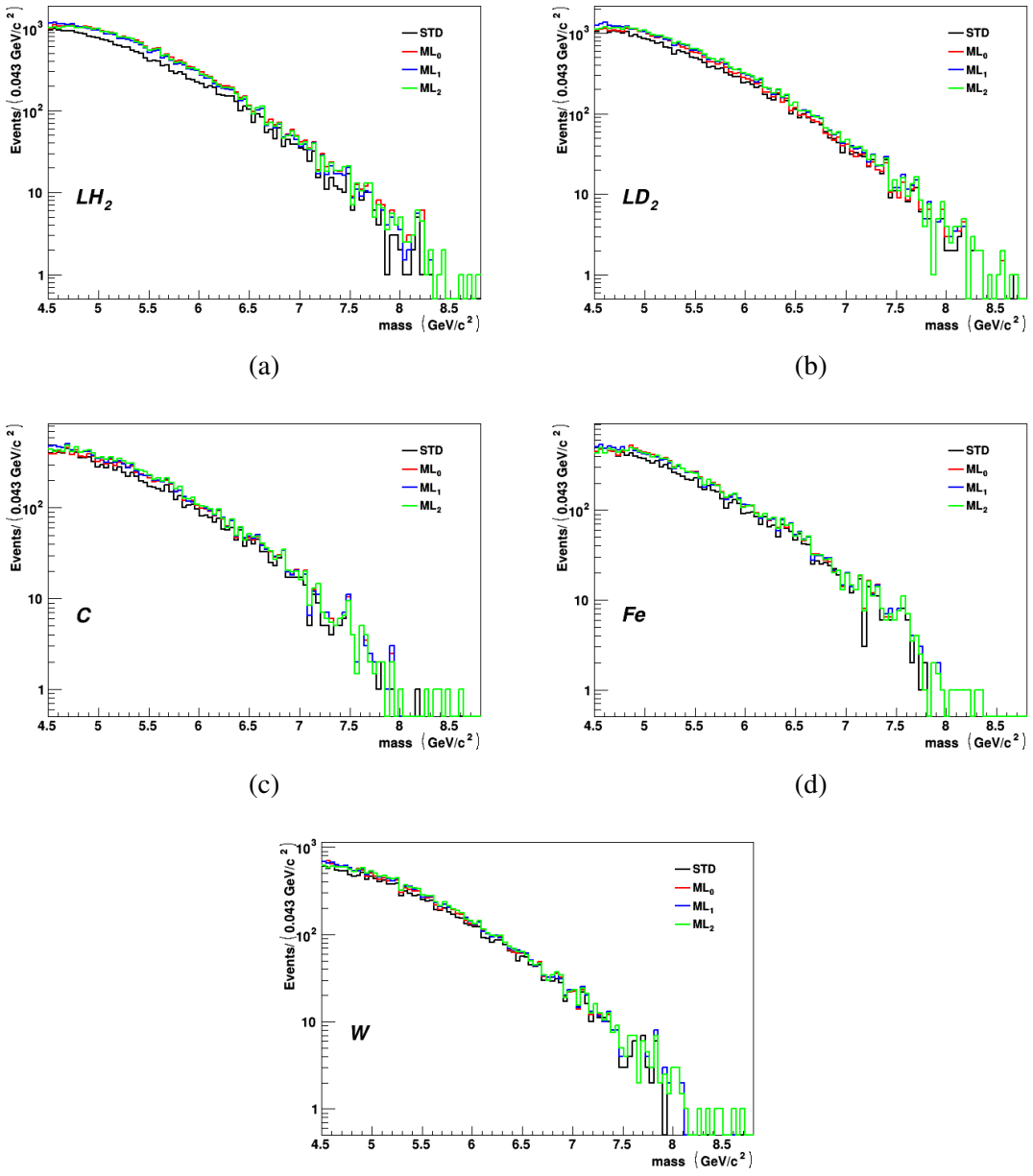


Figure 3.4: Comparison of the mass yields, plotted on a log scale, between the STD, ML_0, ML_1, and ML_2 cuts: a) LH_2 yields, b) LD_2 yields, c) C yields, d) Fe yields, and d) W yields.

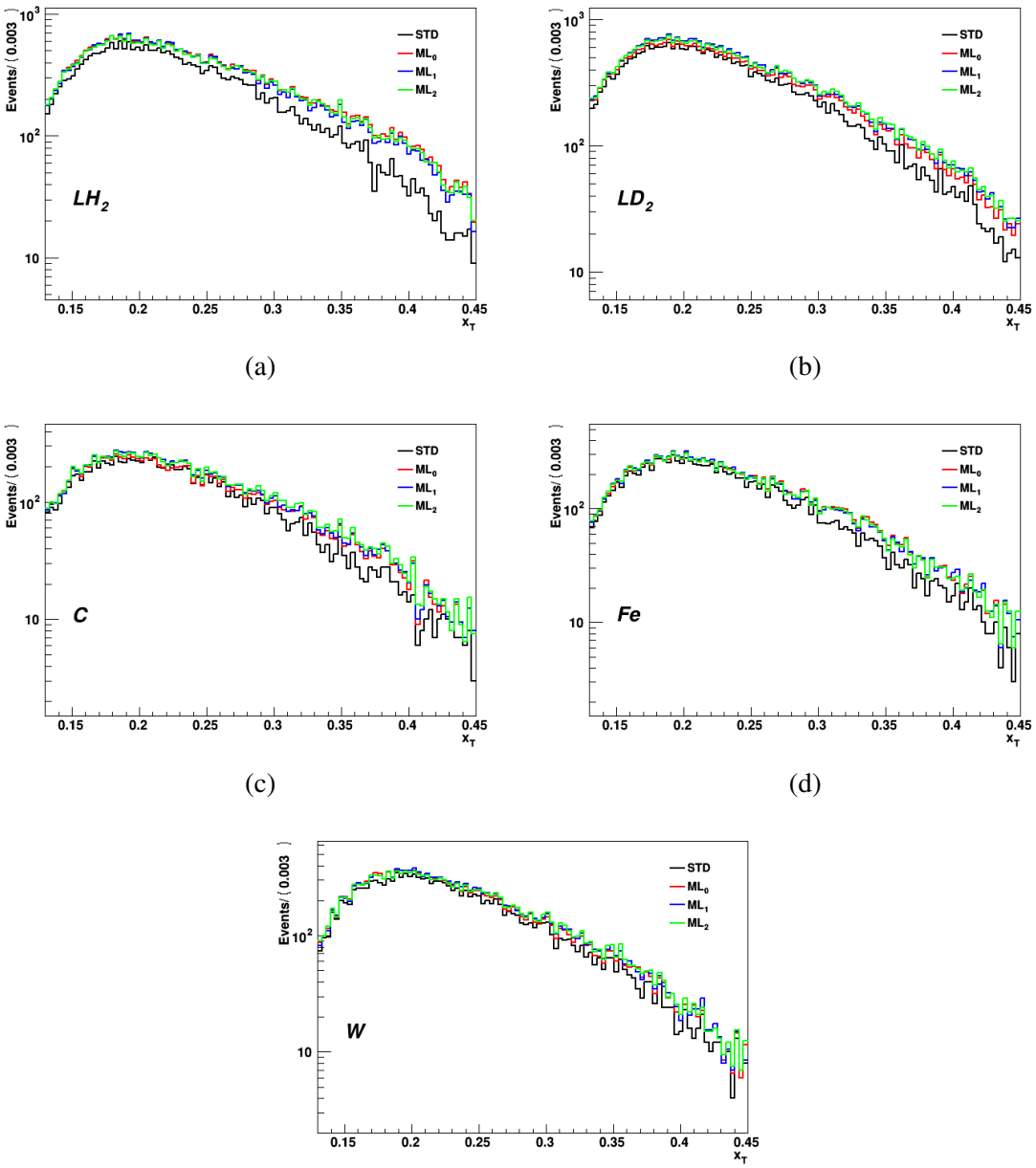


Figure 3.5: Comparison of the x_T yields, plotted on a log scale, between the STD, ML_0, ML_1, and ML_2 cuts: a) LH_2 yields, b) LD_2 yields, c) C yields, d) Fe yields, and d) W yields.

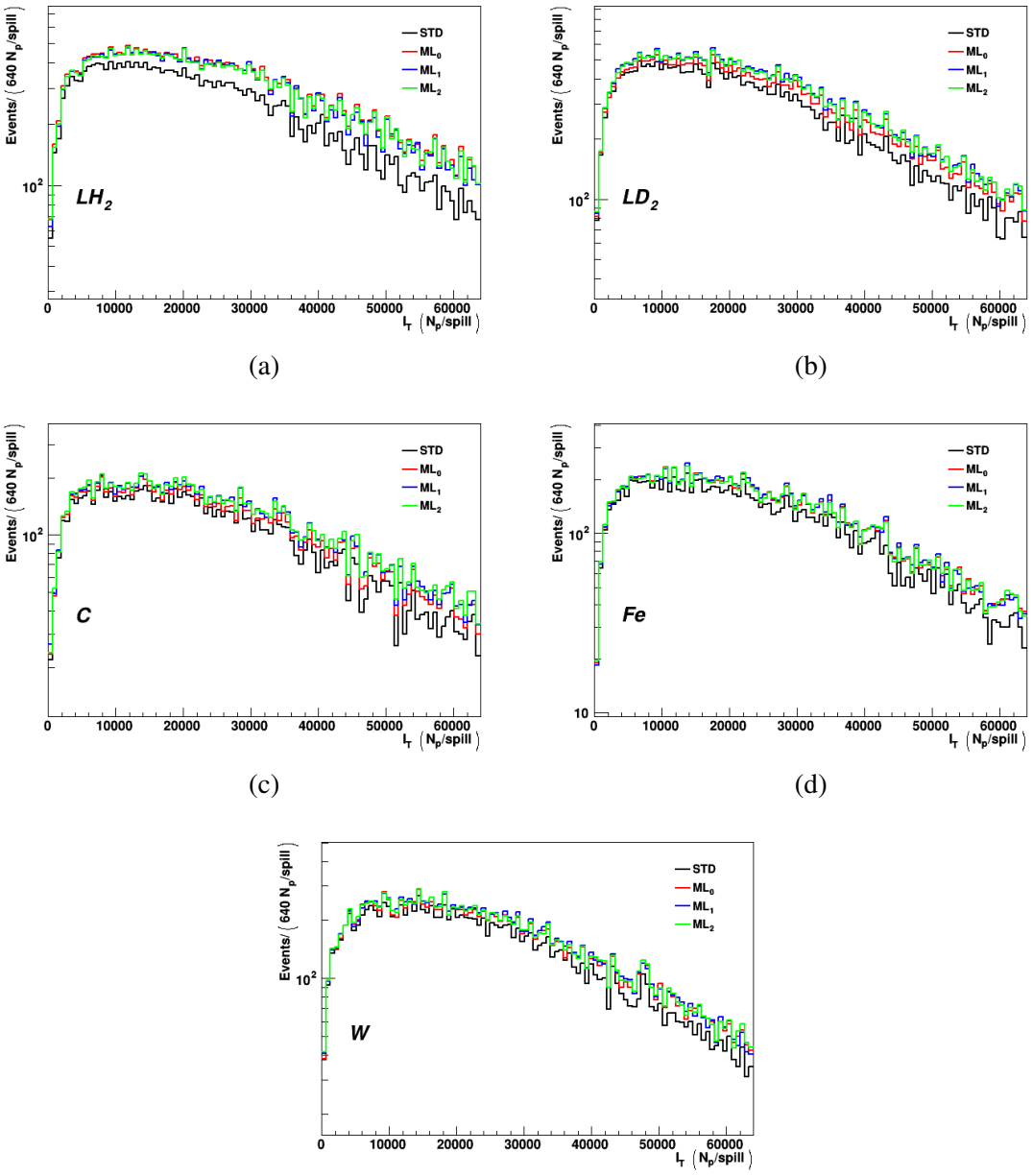


Figure 3.6: Comparison of the I_T yields, plotted on a log scale, between the STD, ML_0, ML_1, and ML_2 cuts: a) LH_2 yields, b) LD_2 yields, c) C yields, d) Fe yields, and d) W yields.

3.5 Cross Section Ratio

To eventually extract the $\bar{d}/\bar{u}$ ratio, the relationship between the measured yields and the DY cross section is needed. The DY yield is a function of the DY cross section and the number of possible interactions the proton has with the target. Mathematically, the yield as defined in Ref. [116] is

$$Y_A = T_A * N_A * POT_A * A_A * \sigma_{pA} * \epsilon_A * \alpha_A / M_A. \tag{3.5}$$

The terms in the expression are defined as

- Y_A: the dimuon yield from target with mass number A
- T_A: the thickness of the target in g/cm^2
- N_A: Avogadro's number, 6.022$\times$10^{23}, in mol^{-1}
- POT_A: the integrated number of protons delivered to the target
- A_A: the beam attenuation through the target
- σ_{pA}: the DY cross section between a proton and the target nucleus
- ϵ_A: the spectrometer efficiency of the target
- α_A: the spectrometer acceptance of the target
- M_A: the mass number of the target nucleus.

T_A is the multiplication of the length of the target by the target's density,

$$T_A = L_A * \rho_A, \tag{3.6}$$

but due to the contamination of the deuterium flask, the thickness of the LD_2 target is a combination

of the POT weighted thicknesses of the LD_2 and LHD,

$$\langle T_D \rangle = L_{flask} * \rho_{LD_2} * \left(\frac{\langle D \rangle + 0.5 * \langle HD \rangle}{\langle V_c \rangle} \right) \tag{3.7}$$

and

$$\langle T_{HD} \rangle = L_{flask} * \rho_{LD_2} \left(\frac{0.5 * \langle HD \rangle}{\langle V_c \rangle} \right), \tag{3.8}$$

respectively, with $\langle V_c \rangle$ being the POT weighted volume fraction for the LD_2. The HD molecule has a larger volume than the D_2 molecule, so a correction for the contamination this has to be taken into account. V_c, the adjusted fraction, is first calculated as

$$\begin{aligned} V_c &= \frac{Frac_{D_2} * V_{D_2} + Frac_{HD} * V_{HD}}{V_{D_2}} \\ &= 0.918 + 0.082 * 1.094 \\ &= 1.008, \end{aligned} \tag{3.9}$$

with the ratio of the HD volume to D_2 volume being 1.094 and the mole fractions during the contamination periods being 0.918 for D_2 and 0.082 for HD, with negligible H_2 contamination. $\langle V_c \rangle$ is found by weighing the V_c by the POT,

$$\begin{aligned} \langle V_c \rangle &= \frac{POT_{cont} * V_c + POT_{pure} * 1.0}{POT_{Total}} \\ &= \frac{(2.21225 \times 10^{17}) * 1.007708 + (8.38729 \times 10^{16})}{3.05044 \times 10^{17}} \\ &= 1.006, \end{aligned} \tag{3.10}$$

with POT_{cont} and POT_{pure} being the POT received during the time when the deuterium was contaminated and when it was pure, respectively.

$\langle D_2 \rangle$ and $\langle HD \rangle$, the POT weighted mole fractions, are calculated to be

$$\begin{aligned} \langle D_2 \rangle &= \frac{(2.21225 \times 10^{17}) * 0.918 + 8.38729 \times 10^{16}}{3.05044 \times 10^{17}} \\ &= 0.9407 \end{aligned} \tag{3.11}$$

and

$$\begin{aligned}\langle HD\rangle &= \frac{(2.21225\times10^{17})*0.082}{3.05044\times10^{17}}\\ &= 0.05947.\end{aligned} \tag{3.12}$$

With the terms in Eqs. 3.7 and 3.8 the thickness of all of the targets can be calculated, which are listed in Table 3.37.

Target	$T_A\ (g/cm^2)$
LH_2	3.597
$\langle\ LD_2\ \rangle$	8.009
$\langle$ LHD $\rangle$	0.2454
C	5.980
Fe	14.99
W	18.39

Table 3.37: Table of the thicknesses of the targets.

As the incident beam traverses the target, its flux is diminished due to scattering with the target material. This diminished flux or attenuation can be characterized as a decrease in the effective length of the target,

$$\begin{aligned}A_A &= (1-exp\left[\lambda_A/L_A\right])*\frac{\lambda_A}{L_A},\\ \lambda_A^{-1} &= \frac{\rho_A*N_A*\sigma_A}{\langle V_c\rangle*M_A},\end{aligned} \tag{3.13}$$

with λ_A being the nuclear interaction length and σ_A being the physical cross section of the atom. For all non-deuterium targets, V_c is unity. For deuterium, the effective attenuation must be constructed as

$$\langle\lambda_D^{-1}\rangle = \frac{\rho_D*\sigma_D*N_A*(1-\langle c\rangle*0.5)}{\langle V_c\rangle*M_D} + \frac{\rho_D*\sigma_H*N_A*\langle c\rangle*0.5}{\langle V_c\rangle*M_D}, \tag{3.14}$$

with the first term being the portion of the λ_D^{-1} coming from interaction of the deuterium nucleus and the second term comes from the hydrogen nucleus interaction [116, 117]. The POT weighted contamination fraction is

$$\begin{aligned} \langle c \rangle &= \frac{0.082 * 2.21225 \times 10^{17}}{3.05044 \times 10^{17}} \\ &= 0.05947. \end{aligned} \tag{3.15}$$

The $\langle c \rangle * 0.5$ terms contain the factor of 0.5 to compensate for the equal contributions of the D and H nuclei in the HD molecule. σ_H and σ_D are 32.2 mb and 46.6 mb, respectively [116]. Table 3.38 details the beam attenuations, nuclear interaction lengths, molar masses, and mass numbers for all of the targets.

Target	A_A	$\lambda_A\ (cm)$	$M_A\ (g/mol)$	A
LH_2	0.9662	734.5	1.008	1
LD_2	0.9451	446.4	2.014	2
C	0.9584	38.83	12.01	12.01
Fe	0.9453	16.77	55.85	55.91
W	0.9536	9.946	183.8	183.9

Table 3.38: Table of beam attenuations (A_A), nuclear interaction lengths (λ_A), molar masses (M_A), and mass numbers (A) for the LH_2, LD_2, C, Fe, and W targets. The nuclear mass numbers are effective mass numbers based upon the densities and compositions of the targets.

3.5.1 Background Subtraction

The DY events generated from the flasks holding the LH_2 and LD_2 targets must be subtracted from their yields; similarly, the DY events generated from the proton beam interacting with the metallic holders must be subtracted from the nuclear target yields. To properly subtract the EMPTY and

NONE events the yields must be scaled by POT$_A$, A_A, ϵ_A, and α_A. They are

$$\begin{aligned} Y'_A &= \frac{Y_A}{POT_A * A_A * \epsilon_A * \alpha_A} \\ &= \frac{T_A * N_A * \sigma_{pA}}{M_A} \end{aligned} \tag{3.16}$$

for all of the targets except for LD_2, which contains the contamination corrections

$$Y'_D = \frac{\langle T_D \rangle * N_A * \sigma_{pD}}{M_D} + \frac{\langle T_{HD} \rangle * N_A * \sigma_{pH}}{M_H}. \tag{3.17}$$

For the subtraction, the efficiencies and acceptances of the liquid targets and the flask are assumed to be the equivalent and the same assumption holds for the nuclear targets and the NONE target. The final normalized DY yields, Y''_A, are

$$Y''_H = \frac{T_H * N_A * \sigma_{pH}}{M_H} - \frac{Y_{EMPTY}}{POT_{EMPTY}}, \tag{3.18}$$

$$Y''_D = \frac{\langle T_D \rangle * N_A * \sigma_{pD}}{M_D} + \frac{\langle T_{HD} \rangle * N_A * \sigma_{pH}}{M_H} - \frac{Y_{EMPTY}}{POT_{EMPTY}}, \tag{3.19}$$

and

$$Y''_N = \frac{T_N * N_A * \sigma_{pN}}{M_N} - \frac{Y_{NONE}}{POT_{NONE}}, \tag{3.20}$$

for the LH_2, LD_2, and nuclear targets, respectively. After rearranging terms and taking the ratio, the CSRs can be calculated as

$$R_{pD} = \frac{\sigma_{pD}/2}{\sigma_{pH}} = \frac{T_H * M_D}{2 * \langle T_D \rangle * M_H} \left(\frac{Y''_D}{Y''_H} - \frac{\langle T_{HD} \rangle}{T_H} \right), \tag{3.21}$$

$$R_{pA} = \frac{\sigma_{pA}/A}{\sigma_{pD}/2} = \frac{2 * \langle T_D \rangle * M_A}{A * T_A * M_D} \left(\frac{Y''_A}{Y''_D - \frac{Y''_H * \langle T_{HD} \rangle}{T_H}} \right), \tag{3.22}$$

and

$$R_{HA} = \frac{\sigma_{pA}/A}{\sigma_{pH}} = \frac{T_H * M_A}{A * T_A * M_H} \left(\frac{Y''_A}{Y''_H} \right), \tag{3.23}$$

with R_{pD} being the liquid CSR used for the extraction of the free proton $\bar{d}/\bar{u}$ ratio, R_{pA} being the nuclear to deuterium CSR used for nuclear dependence, and R_{HA} being the nuclear to hydrogen CSR utilized for the extraction of the bound proton asymmetry ratio.

3.5.2 Intensity Dependence

As mentioned in Section 2.1, as intensity increases the ability of the tracking software to reconstruct proper dimuons is diminished. For dimuon yields this means that the combinatoric background increases with intensity. This background originates when non-associated hits of the hodoscope panels and wire chambers mimic a track that is then reconstructed to form a dimuon pair. There is also the question of the differing efficiencies and acceptances between the targets in the CSR calculations, and that even with the STD or ML cuts, a large portion of the combinatoric background remains.

To derive the CSR an extrapolation in a variable as a function of I_T is used. The understanding is that there is an intensity dependence within the data and this manifests in the CSR. When plotting the CSR of a variable within a bin, this dependence is proportional to a quadratic function. The quadratic nature is due to the probability that a combinatoric background track being reconstructed with a true dimuon track is proportional to the intensity and the probability that two combinatoric tracks are reconstructed to a dimuon pair is proportional to the intensity squared, that is

$$CSR \approx CSR_0 + p_1 * I_T + p_2 * I_T^2. \tag{3.24}$$

An extrapolation of the CSR to $I_T = 0$ would yield CSR_0, the CSR in the absence of any intensity effects.

3.5.2.1 Akaike Information Criterion

Several functions were evaluated as possible functional fits: common polynomial fit 2 (CP2), 2-dimensional polynomial fit 2 (Fit5C), common cosine (CC), 2-dimensional cosine (CC2), common polynomial 1 (CP1), and 2-dimensional fit (Fit2D). These fits are described mathematically as

- CC2: $p_0 * cos\left(\frac{I_T}{p_{10}+p_{11}*V_{ave}}\right)$
- Fit5C: $p_0 + (p_{10} + p_{11} * p_{20}) * I_T + (p_{21} + V_{ave} * p_{20}) * I_T^2$
- Fit2D: $p_0 + (p_{10} + p_{11} * V_{ave}) * I_T$
- CC: $p_0 * cos\left(I_T/p_1\right)$
- CP2: $p_0 + p_1 * I_T + p_2 * I_T^2$
- CP1: $p_0 + p_1 * I_T$,

with the p_0 terms being the CSR_0 and the other polynomial terms being the same in all of the variable bins. V_{ave} is the mean value of the variable within the kinematic bin using both the numerator and denominator distributions to calculate the mean. CP2 and Fit5C were initially proposed as the two functional forms for the I_T dependence and eventually CP2 was chosen by the collaboration with Fit5C bring retained to determine the systematic uncertainties [118, 119]. Part of the reasoning of favoring CP2 was that it possessed more stability between the mass cuts of 4.2 and 4.5 GeV/c^2 and that when performing an Akaike information criterion (AIC), described in detail in Refs.[120, 121], CP2 was significantly favored over Fit5C. The AIC is an objective measure used to discriminate between fit functions that yield similar results. It strongly penalizes fits with larger number of parameters with similar log-likelihoods and is defined as

$$AIC = -2\, ln\left[L\left(y|\tilde{\theta_k}\right)\right] + 2k, \tag{3.25}$$

with $L\left(y|\tilde{\theta_k}\right)$ being the maximum likelihood between the vector of parameter estimates $\tilde{\theta_k}$ and the y vector of measurements, and k is the number of free parameters in the model fit [120]. For

the results presented in Ref. [118], the χ^2 was used instead of the log-likelihood, which should yield similar results in terms of best fit function. The χ^2 and binned log-likelihood minimizations are defined as

$$\chi^2 = \sum_{i}^{n_{bins}} \left(\frac{CSR_{pred,i} - CSR_{mea,i}}{\sigma_{stat_i}} \right)^2 \tag{3.26}$$

and

$$LLV = \sum_{i}^{n_{bins}} CSR_{pred,i} - CSR_{mea,i} + CSR_{mea,i} * ln \left(\frac{CSR_{mea,i}}{CSR_{pred,i}} \right), \tag{3.27}$$

respectively, with $\sigma_{stat,i}$, $CSR_{mea,i}$, and $CSR_{pred,i}$ being the statistical error, measured CSR, and predicted CSR of the ith bin respectively. In Tables 3.39 - 3.43 the AIC for the STD cuts was calculated for the targets listed and for the three pedestals: 30, 34, and 38. The three pedestals were used due to the pedestal systematic error, which is described in more detail in the Results chapter. The mean AICs and their weights along with their standard deviations are also calculated. The higher the weight the more likely the fit function is to the "true" distribution. The ML results mirrored the STD results, so only the STD results are shown. It should be noted that the actual value of the AIC is not important, but the closer it is to negative infinity, the better.

Fit	AIC_{ave}	AIC_{σ}	$Weight_{ave}$	$Weight_{\sigma}$
Fit5C	-40.44	2.182	0.009466	0.000415
CP2	-44.75	2.118	0.08177	0.003040
CC	-47.19	1.966	0.2784	0.03098
CC2	-45.14	1.960	0.09982	0.01097
CP1	-47.95	2.350	0.4042	0.03211
Fit2D	-45.62	2.394	0.1263	0.01291

Table 3.39: Mean AIC, weight, and their standard deviations for the R_{pD} using χ^2 minimization.

Fit	AIC_{ave}	AIC_{σ}	$Weight_{ave}$	$Weight_{\sigma}$
Fit5C	-49.83	21.15	0.002800	0.003918
CP2	-55.88	22.31	0.02663	0.03305
CC	-63.52	24.22	0.2070	0.1004
CC2	-59.40	24.81	0.04081	0.03894
CP1	-65.97	26.02	0.6064	0.1913
Fit2D	-62.41	24.43	0.1163	0.0525

Table 3.40: Mean AIC, weight, and their standard deviations for the R_{HA} using χ^2 minimization.

Fit	AIC_{ave}	AIC_{σ}	$Weight_{ave}$	$Weight_{\sigma}$
Fit5C	-46.04	15.32	0.000008786	0.00001241
CP2	-51.82	14.88	0.0001234	0.00001741
CC	-84.67	4.053	0.7292	0.00003002
CC2	-82.67	4.054	0.2681	0.00009667
CP1	-58.39	12.71	0.002154	0.00003212
Fit2D	-55.52	12.09	0.0003863	0.00005690

Table 3.41: Mean AIC, weight, and their standard deviations for the R_{pD} using χ^2 minimization.

Fit	AIC_{ave}	AIC_{σ}	$Weight_{ave}$	$Weight_{\sigma}$
Fit5C	19.07	0.07517	0.01570	0.00002372
CP2	15.07	0.07605	0.1161	0.0001087
CC	13.06	0.07663	0.3168	0.0001714
CC2	15.06	0.07516	0.1164	0.0001457
CP1	13.06	0.07984	0.3180	0.0003736
Fit2D	15.06	0.07856	0.1169	0.0007714

Table 3.42: Mean AIC, weight, and their standard deviations for the R_{pD} using log-likelihood minimization.

Fit	AIC_{ave}	AIC_{σ}	$Weight_{ave}$	$Weight_{\sigma}$
Fit5C	18.49	0.5263	0.01462	0.0008420
CP2	14.40	0.5028	0.1133	0.002781
CC	12.31	0.5378	0.3216	0.007307
CC2	14.46	0.6377	0.1102	0.01122
CP1	12.29	0.5364	0.3249	0.008339
Fit2D	14.37	0.5691	0.1153	0.004343

Table 3.43: Mean AIC, weight, and their standard deviations for the R_{HA} using log-likelihood minimization.

Fit	AIC_{ave}	AIC_{σ}	$Weight_{ave}$	$Weight_{\sigma}$
Fit5C	18.84	0.2738	0.01392	0.0007550
CP2	14.70	0.3958	0.1103	0.004016
CC	12.48	0.3384	0.3358	0.01485
CC2	14.48	0.3384	0.1235	0.005463
CP1	12.65	0.4038	0.3084	0.001470
Fit2D	14.75	0.2868	0.1079	0.004985

Table 3.44: Mean AIC, weight, and their standard deviations for the R_{pD} using log-likelihood minimization.

From Tables 3.39, 3.40, 3.42, and 3.43, it is clear that CC and CP1 are preferred over the other fits, with the χ^2 showing that the CC is preferred over CP1, while the log-likelihood weighting them equally. For the nuclear dependence tables, Tables 3.41 and 3.44, CC and CC2 dominate for the χ^2 minimization and CC and CP1 dominate for the log-likelihood. This would suggest that a change in functional form is warranted. However, the log-likelihood and the χ^2 do not arrive at the same extracted CSR_0. To characterize the differences between the CSRs, a CSR standard deviation difference was constructed according to

$$D_{CSR} \equiv \frac{|CSR_{ll} - CSR_{\chi^2}|}{\sigma_{com}}, \tag{3.28}$$

with $\sigma_{comp} = \sqrt{\sigma_{ll}^2 + \sigma_{\chi^2}^2}$ and the ll and χ^2 denoting log-likelihood and χ^2, respectively. D_{CSR} gives the number of standard deviations the two CSRs are apart. In the current analysis the log-likelihood had the larger of the two errors across all x_T bins and targets.

It is important to note that the error definition for χ^2 minimization is usually set at unity and for log-likelihood this is usually set at 0.5. When calculating the error on the minimization the parameters are altered one by one until the change in the minimization reaches the error definition. Assuming a χ^2/N_{par} of unity, the χ^2 would be around 50 for the number of x_T and I_T bins, so the error definition of unity is 2% of this χ^2. When running the log-likelihood fits, the results were close to unity, so an error definition of 0.5 yielded significant errors, which is not consistent with the χ^2 results. Therefore the error definition was changed to be 0.02 to emulate the χ^2 errors. The errors do not affect the minimization value, so the differences are independent of definition. The results of the average D_{CSR}, which were only calculated for a pedestal of 34, are listed in Tables 3.45 and 3.46.

From Tables 3.45 and 3.46 it becomes clear that CP2 and Fit5C have the smallest mean D_{CSR} values and the other fit functions suffer from large average deviations in the extracted CSR. It is due to this instability of the fits across χ^2 and log-likelihood minimization's that CP2 was chosen as the functional form of the I_T dependence used for the extrapolation for the x_T CSRs with Fit5C

Target	Fit5C	CP2	CC	CC2	CP1	Fit2D
LD_2	0.2396	0.2514	0.2866	0.3080	0.3378	0.3486
C	0.5267	0.2547	0.7004	0.6821	0.2973	0.5574
Fe	0.4184	0.3190	0.4521	0.4009	0.4930	0.5227
W	0.3552	0.3480	0.3680	0.3556	0.5022	0.5142

Table 3.45: Mean CSR standard deviation difference (D_{CSR}) for the Fit5C, CP2, CC, CC2, CP1, and Fit2D I_T dependence functions for R_{HA} and R_{pD}.

Target	Fit5C	CP2	CC	CC2	CP1	Fit2D
C	0.3365	0.4044	0.7991	0.8160	0.6442	0.6362
Fe	0.1383	0.1368	0.4286	0.4041	0.4282	0.4283
W	0.1302	0.2618	0.1735	0.1843	0.7407	0.7405

Table 3.46: Mean CSR standard deviation difference (D_{CSR}) for the Fit5C, CP2, CC, CC2, CP1, and Fit2D I_T dependence functions for R_{pA}.

being used as a systematic uncertainty on the dependence choice, in accordance with the initial suggestion.

3.6 Error Analysis

For the CSR calculation, an estimate of the statistical error on the ratio is required. The propagation of the statistical error in a CERN ROOT histogram is carried out using implicit Gaussian error propagation between two bins, such that

$$Z = X \pm Y \tag{3.29}$$

and

$$\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2}. \tag{3.30}$$

When scaling a bin by a constant, the bin's height and error are simply scaled by that same value, as in

$$Z = a * X, \tag{3.31}$$

and

$$\sigma_z = a * \sigma_x. \tag{3.32}$$

An issue arises when Z is a nonlinear function of bins, such as $Z = X * Y$ or Z = $\sqrt{X}$. The ROOT analysis package and the prevailing physics wisdom is to treat Z as a function and propagate the error as

$$\sigma_z^2 = \sum_i \left(\frac{\partial Z}{\partial X_i}\right)^2 * (\sigma_{X_i})^2. \tag{3.33}$$

This treats the error in bin Z as a point estimation, i.e. that bin Z is a single point and not a representation of events within the bin. A histogram bin, at least within the default settings of ROOT, represents the number of events that falls within the bounds of the bin with a 1σ confidence interval (CI) about the bin center. That means that in the Gaussian limit the bin would represent 68% of the events sampled from a Gaussian distribution with the bin center as the mean and the standard deviation as σ.

With that in mind, the error on a bin should represent the underlying uncertainty of whatever relationship between Z and its constituents, both of which are Gaussian distributed. For linear expressions, the error is unchanged from Equations 3.30 and 3.32, but for nonlinear expressions, proper nonlinear CIs must be calculated, specifically for CSRs, a CI for a ratio of two distributions is needed. In statistics there are two main methods to calculate the CI of a ratio distribution between two Gaussian distributed random variables: the Delta method and the Fieller method.

3.6.1 Delta Method

This method starts by Taylor expanding the ratio and the variance about the means of the distributions and then takes the expectation of the results [122]. The expectation is achieved by integrating

the expression with the two Gaussians e.g.

$$\begin{aligned} E\left[X\right] &= \int_{-\infty}^{\infty} dX\ X * \frac{1}{\sigma_X\sqrt{2\pi}} * exp\left[-\frac{1}{2}\left(\frac{X-\mu_X}{\sigma_X}\right)^2\right] \\ &= \mu_X, \end{aligned} \tag{3.34}$$

with μ_X being the mean of the X distribution and where the Y Gaussian integral is unity and therefore suppressed due to there not being any explicit X dependence. Defining $Z = X/Y$, with $|$ denoting evaluating what is left of $|$ at μ_X and μ_Y, expanding to second order and taking the expectation yields

$$\begin{aligned} Z &\approx Z| + \left.\frac{\partial Z}{\partial X}\right| * X + \left.\frac{\partial Z}{\partial Y}\right| * Y + \left.\frac{\partial^2 Z}{\partial X^2}\right| * \frac{X^2}{2} + \left.\frac{\partial^2 Z}{\partial Y^2}\right| * \frac{Y^2}{2} + 2 * \left.\frac{\partial^2 Z}{\partial X \partial Y}\right| * \frac{X*Y}{2} \\ &\approx \frac{\mu_X}{\mu_Y} + \frac{X}{\mu_Y} - \frac{\mu_X * Y}{{\mu_Y}^2} + \frac{\mu_X * Y^2}{{\mu_Y}^3} - \frac{X*Y}{{\mu_Y}^2}, \\ E\left[Z\right] &\approx \frac{\mu_X}{\mu_Y} - \frac{COV(X,Y)}{{\mu_Y}^2} + \frac{\mu_X * {\sigma_Y}^2}{{\mu_Y}^3}, \end{aligned} \tag{3.35}$$

and for the expansion of the variance

$$\begin{aligned} {\sigma_Z}^2 &\approx \left(\left.\frac{\partial Z}{\partial X}\right| * (X-\mu_X) + \left.\frac{\partial Z}{\partial Y}\right| * (Y-\mu_Y)\right)^2 \\ &\approx \frac{(X-\mu_X)^2}{{\mu_Y}^2} - \frac{2*(X-\mu_X)(Y-\mu_Y)}{{\mu_Y}^3} + \frac{{\mu_X}^2 * (Y-\mu_Y)^2}{{\mu_Y}^4}, \\ E\left[{\sigma_Z}^2\right] &\approx \frac{{\sigma_X}^2}{{\mu_Y}^2} - \frac{2*\mu_X*COV(X,Y)}{{\mu_Y}^3} + \frac{{\mu_X}^2 * {\sigma_Y}^2}{{\mu_Y}^4}. \end{aligned} \tag{3.36}$$

$COV(X,Y)$ is the covariance between the X and Y distributions, which is expected to be small for the CSRs. The upper limit is then $E[Z] + \sqrt{E[{\sigma_Z}^2]}$ and the lower limit is $E[Z] - \sqrt{E[{\sigma_Z}^2]}$, and the upper(lower) error is taken to be the mean ratio, gathered from the histogram, minus the upper(lower) limit.

3.6.2 Fieller Method

This method takes a somewhat different approach to finding the CI of a ratio distribution. Suppose that there is a linear combination of X and Y called W with $W = X - Z*Y$. From Equations 3.29 - 3.33, W is Gaussian distributed around zero and has a $\sigma_W = \sqrt{{\sigma_X}^2 + Z^2 * {\sigma_Y}^2 + 2 * Z * COV(X,Y)}$. Then the standard normal random variable $\boldsymbol{Z}$, where $\boldsymbol{Z} = \frac{W}{\sigma_W}$, and its square is a chi-squared variable with 1 degree of freedom [123]. This is merely a statement that this function follows a Z or standard score distribution with mean zero. From this the CI is found by finding the Z values that satisfy the inequality

$$\frac{W^2}{\sigma_W^2} < {\boldsymbol{Z}_{\alpha/2}}^2, \tag{3.37}$$

where $1 - \alpha$ is the decimal confidence interval, i.e. $\alpha \approx 0.318$ for the 1σ error and would be 0.05 for a 95% CI. Solving the quadratic equation leads to

$$\begin{aligned} CI = Z + \left(\frac{k}{1-k}\right)\left(Z + \frac{COV(X,Y)}{{\sigma_Y}^2}\right) \\ \pm \left(\frac{\boldsymbol{Z}_{\alpha/2}}{Y\,(1-k)}\right)\left[{\sigma_W}^2 - k\left({\sigma_X}^2 - \frac{COV(X,Y)^2}{{\sigma_Y}^2}\right)\right]^{\frac{1}{2}} \end{aligned} \tag{3.38}$$

with $k = {\boldsymbol{Z}_{\alpha/2}}\frac{{\sigma_Y}^2}{Y^2}$.

Unlike in the Delta method, this result requires samples from the X, Y and Z distributions. Therefore two histograms need to be set side to be filled with the first part of Equation 3.38, and one filled with the second part. These terms are denoted as C and $t * SE$, respectively, in Ref. [124]. The means from these two histograms are then taken to be the results. As in the Delta method, the upper(lower) limit is taken to be the upper limit minus the mean, from the mean Z distribution, and the lower error is the mean minus the lower limit.

3.6.3 Asymmetric Confidence Intervals

Due to the nature of the Fieller or Delta methods, or the CSR calculation in general, an asymmetric CI can arise. Paradoxically, the literature is fairly sparse on how to properly propagate an asym-

metric CI and no nice closed form solutions exist. However, Roger Barlow [125, 126] has done some work that uses the sum of log-likelihoods to estimate the mean and propagated errors. The likelihood is defined as the probability of finding a datum x with parameter θ, $L_i = P(x_i;\theta)$, from some probability distribution. The joint probability, L, is the product of each of these probabilities

$$L(\theta) = \prod_i P(x_i;\theta), \tag{3.39}$$

and

$$lnL(\theta) = \sum_i lnP(x_i;\theta), \tag{3.40}$$

is the log-likelihood being the natural log of the likelihood. In the large N limit, the log-likelihood possesses a parabolic shape [125]. From the curve the maximum can be found by maximizing the function with respect to θ. The θ at this maximum is the estimated parameter. Around this maximum, $ln(\theta)$ is approximated by Taylor expansion as

$$lnL(\theta) \approx lnL(\hat{\theta}) + \frac{1}{2}\frac{\partial^2 lnL}{\partial\theta^2}\bigg|_{\theta=\hat{\theta}} \left(\theta - \hat{\theta}\right)^2 ..., \tag{3.41}$$

with $\hat{\theta}$ being the estimated parameter of interest. Differentiating with respect to θ and setting the left side of the equation equal to zero leaves us with

$$(\theta - \hat{\theta}) = -\left(\frac{\partial lnL(\theta)}{\partial\theta}\right) * \left[\frac{\partial^2 lnL(\theta)}{\partial\theta^2}\right]^{-1}, \tag{3.42}$$

neglecting the $O(3)$ correction. To get the variance the square of each side is taken and the expectation is found

$$E\left[(\theta - \hat{\theta})^2\right] = E\left[\left(-\left(\frac{\partial lnL(\theta)}{\partial\theta}\right) * \left[\frac{\partial^2 lnL(\theta)}{\partial\theta^2}\right]^{-1}\right)^2\right], \tag{3.43}$$

which reduces to

$${\sigma_{\hat{\theta}}}^2 = -\left(E\left[\frac{\partial^2 lnL(\theta)}{\partial\theta^2}\right]\right)^{-1}, \tag{3.44}$$

as seen in Refs. [127] and [128]. This equivalence is the consequence of the nature of logarithms and the expectations of the derivatives of the log-likelihood. For example, take a probability p, its natural logarithm (l), derivative of l with respect to p (S), the derivative of S with respect to p, and the variance of S (σ_S) as

$$l = ln(p), \tag{3.45}$$

$$S = \frac{\partial l}{\partial p}, \tag{3.46}$$

$$S' = (p''/p) - (p'/p)^2, \tag{3.47}$$

and

$$\sigma_S{}^2 = E(S^2) - (E(S))^2, \tag{3.48}$$

respectively. $E(S)$ is equal to zero and so is $E(S')$. Substituting one for the other and recognizing that the expectation of a sum is the same as the sum of the expectation

$$\begin{aligned} \sigma_S{}^2 &= E\left[\left(\frac{p'}{p}\right)^2\right] - E\left[\frac{p''}{p}\right] \\ &= -E\left[\left(\frac{p'}{p}\right)^2 - \frac{p''}{p}\right] \\ &= -E\left[S'\right]. \end{aligned} \tag{3.49}$$

This illustrates the equivalence between the second derivative of the log-likelihood and the variance. This derivation was modeled after Wasserman's [129]. Generally speaking, the second derivative of log-likelihood is the Hessian or information matrix and this derivation can be expanded to multiple dimensions in θ.

To get the 68% CI a line is found that intersects the $lnL(\theta)$ so that the difference between the $lnL(\hat{\theta})$ and the line is $-\frac{1}{2}$; this line, $lnL(\theta) = lnL(\hat{\theta}) - \frac{1}{2}$, intersects at $\theta = \hat{\theta} \pm \sigma_{\hat{\theta}}$. For different CIs one would need $\Delta lnL(\hat{\theta}) = -2$ and $\Delta lnL(\hat{\theta}) = -4.5$ for CIs of 95.5% and 99.7%,

respectively [125].

For measurements where the log-likelihood is not known, one can be constructed from the means, the upper errors, and the lower errors of the data samples. For the symmetric case, the sum of the means and the variances is used to get the final answer, but for the asymmetric case the sum of the log-likelihoods must be used and then the means and errors are extracted. Barlow has devised several functions that have shown to approximate the log-likelihood such as the linear sigma model, the linear variance model, Model 1, and Model 2. These functions are more or less black boxes in the sense that there is no formal procedure or analytical process to derive them, however they are motivated by the underlying means, standard deviations, and variances; and have been shown to approximate the log-likelihood's of systems with known log-likelihoods and converge to a symmetric log-likelihood in the large N limit. As an aside, propagating the upper error and lower error separately in quadrature, which is sometimes done, violates the central limit theorem.

For the linear models, the log-likelihoods of the events are summed. From the composite log-likelihood the maximum becomes the resultant statistic and the upper and lower errors are found from the $\Delta ln(\hat{\theta}) = -\frac{1}{2}$ line [125]. For the other models a series of cumulants, defined in Sections 3.6.3.3 and 3.6.3.4, are found for each event and then summed. The system of equations are then solved numerically to find the mean and errors [126].

3.6.3.1 Linear sigma

In this model the log-likelihood, θ, and weights(w_i) defined as

$$lnL(\hat{\theta}) = -\frac{1}{2}\sum_i \left(\frac{\theta_i}{\sigma_i + \sigma_i'\theta_i}\right)^2, \tag{3.50}$$

$$\theta_i = u\frac{w_i}{\sum_j w_j}, \tag{3.51}$$

and

$$w_i = \frac{(\sigma_i + \sigma_i'\theta)^2}{2\sigma_i}, \tag{3.52}$$

respectively. The u, σ, and σ' terms are described as

$$u = \sum_i \theta_i, \tag{3.53}$$

$$\sigma = \frac{2\sigma^+\sigma^-}{\sigma^+ + \sigma^-}, \tag{3.54}$$

and

$$\sigma' = \frac{\sigma^+ - \sigma^-}{\sigma^+ + \sigma^-}, \tag{3.55}$$

respectively, with $\sigma_+(\sigma_-)$ being the upper(lower) errors.

3.6.3.2 Linear Variance

This model shares the same definitions of u and θ, but has a different log-likelihood and weight

$$\begin{aligned} lnL(\hat{\theta}) &= -\frac{1}{2}\sum_i \left(\frac{{\theta_i}^2}{V_i + V_i'\theta_i}\right)^2 \\ w_i &= \frac{\theta^2}{2V_i + V_i'\theta_i}, \end{aligned} \tag{3.56}$$

with V and V' defined as

$$\begin{aligned} V &= \sigma^+\sigma^- \\ V' &= \sigma^+ - \sigma^-. \end{aligned} \tag{3.57}$$

3.6.3.3 Model 1

For this model the three cumulants of interest are

$$\mu = x_0 + \frac{1}{\sqrt{2\pi}} \left(\sigma^+ - \sigma^- \right), \tag{3.58}$$

$$V = \frac{1}{2} \left(\sigma^{+2} + \sigma^{-2} \right) - \frac{1}{2\pi} \left(\sigma^+ - \sigma^- \right)^2, \tag{3.59}$$

and

$$\gamma = \frac{1}{\sqrt{2\pi}} \left[2 \left(\sigma^{+3} - \sigma^{-3} \right) \pm \frac{3}{2} \left(\sigma^+ - \sigma^- \right) \left(\sigma^{+2} + \sigma^{-2} \right) + \frac{1}{\pi} \left(\sigma^+ - \sigma^- \right)^3 \right]. \tag{3.60}$$

x_0 is the mean of the parent distribution and after solving the system of equations, it represents the estimated mean of the sum of the samples. There is a slight ambiguity in the definition of γ. The "$+$" sign yields the results given in the tables in the paper, however the "$-$" sign is what is written in the expression in the paper.

3.6.3.4 Model 2

The three cumulants for this model are

$$\mu = x_0 + \frac{1}{2} \left(\sigma^+ - \sigma^- \right), \tag{3.61}$$

$$V = \frac{1}{4} \left(\sigma^+ + \sigma^- \right)^2 + \frac{1}{2} \left(\sigma^+ - \sigma^- \right)^2, \tag{3.62}$$

and

$$\gamma = \frac{3}{4} \left(\sigma^+ + \sigma^- \right)^2 \left(\sigma^+ - \sigma^- \right) + \left(\sigma^+ - \sigma^- \right)^3. \tag{3.63}$$

The system of equations in Model 1 or Model 2 are solved for α, σ, and x_0. α and σ are defined as $\frac{1}{2}(\sigma^+ - \sigma^-)$, $\frac{1}{2}(\sigma^+ + \sigma^-)$, respectively. From α and σ, the upper and lower errors can then be derived.

3.6.4 Model Choice

In practice the linear models give similar results, however they can fail if the program does not find a unique solution [130]. This means that one of the linear models should be called first and if that fails Model 2 should be called.

As a final note, due to the nature of the asymmetry calculations, the result, e.g. the value at the maximum likelihood or x_0, can differ from the simple addition of the means of the quantities. This is the case even if the quantities have means of zero. This means that in propagating the asymmetric error the ratio can be shifted.

3.6.5 Error Results

For the purpose of this CSR analysis, the point estimate of the ratio of two bins is carried out with the statistical error of the bins as the standard deviation and the bin content, or bin height, as the "mean" of the bin. It should be noted that using this interpretation means that bin content is what is uncertain, not the mean of the contents within the bin. When the Delta and Fieller methods were compared with the standard method of error propagation, the differences were found to be negligible. Therefore the standard method was employed for this analysis. Moreover, there were no significant asymmetric error bars in any of the CSR calculations so there was no need to use any of the asymmetric error propagation models.

Chapter 4:

Results

4.1 Cross Section Ratio Results

As mentioned in Sections 3.5.2 and 3.5.2.1, the CSR in each x_T bin was extracted from the I_T dependence using CP2 as the functional form of the dependence. The CSR of the nuclear target to deuterium (R_{pA}), was performed to study the nuclear dependence of the DY process. This ratio treats the nuclear CS as a collection of nucleons, with the CS of the nucleon being half the CS of the proton-proton CS and half the proton-neutron CS. Any discrepancy between the ratio and unity signifies a nuclear modification in the underlying PDFs.

The CSR of deuterium to hydrogen (R_{pD}) and the nuclear target to hydrogen CSR (R_{HA}) is defined approximately as,

$$\frac{\sigma_{pA}/A}{\sigma_{pp}} \approx \frac{Z}{A} + \left(\frac{A-Z}{A}\right)\frac{\sigma_{pn}}{\sigma_{pp}}, \tag{4.1}$$

with the assumption of negligible nuclear dependence or a nuclear dependence factored into the PDFs. These CSRs highlight the differences between the free proton and the bound nucleons. From these CSRs, the $\bar{d}/\bar{u}$ ratio can be extracted given PDF, nPDF, and event data. The CSRs were taken as a function of x_T as opposed to x_B, x_F, or another kinematic variable because the PDFs and nPDFs are functions of x_T and Q^2. The extraction explicitly uses the PDFs and nPDFs to construct a CSR.

The extraction was done through an χ^2 minimization, defined in Eq. 3.26, using the CERN ROOT6 MINUIT program. The statistical errors on the value were found by varying the values

until the change in the χ^2 was equal to unity. In addition to the statistical errors on the CSR there were also systematic errors. The systematic errors are related to the uncertainty in the pedestal, normalization, beam, functional form, mass cut, and for the machine learning-the uncertainty in the machine learning cut.

For the pedestal uncertainty, a pedestal value of 34 was used in the extraction. In addition, two separate pedestals of 30 and 38 were used. The uncertainty is taken as the mean squared difference between the CSR at $pedestal = 34$ and the other pedestals, such that

$$\sigma_{ped} = \sqrt{0.5 * (CSR_{34} - CSR_{30})^2 + 0.5 * (CSR_{34} - CSR_{38})^2}. \tag{4.2}$$

For the normalization uncertainty, a simple 2% of the value of the CSR was taken,

$$\sigma_{beam} = 0.02 * CSR. \tag{4.3}$$

This characterizes the uncertainty in Raw POT normalization stemming from the differences in the G2SEM and NIM3 measures [117]. The beam uncertainty arises from the beam missing portions of the nuclear targets due to the beam offsets. It was estimated that when the beam offset is 1.6 cm, as in roadsets 62-70, approximately 1.2% of the beam misses the target [131], so a 0.5% of the CSR was taken as the systematic error for the CSR involving the nuclear targets, such that

$$\sigma_{beam} = 0.005 * CSR. \tag{4.4}$$

The choice of the functional form of the I_T dependence also gives an uncertainty. CP2 was chosen over Fit5C due to the AIC weight of CP2 being significantly larger than Fit5C as seen in Tables 3.39 - 3.44. The difference between the CSRs was taken as the functional form or fit systematic uncertainty, with

$$\sigma_{Fit} = |CSR_{CP2} - CSR_{Fit5C}|. \tag{4.5}$$

The mass uncertainty comes from the changes in the CSR due the change in the mass cut from

4.2 GeV/c^2 to 4.5 GeV/c^2. This difference was studied in Refs. [132, 133]. It was decided by the collaboration to use a mass cut of 4.5 GeV/c^2 and use the CSR difference as the systematic error, such that

$$\sigma_{mass} = |CSR_{4.5} - CSR_{4.2}|. \tag{4.6}$$

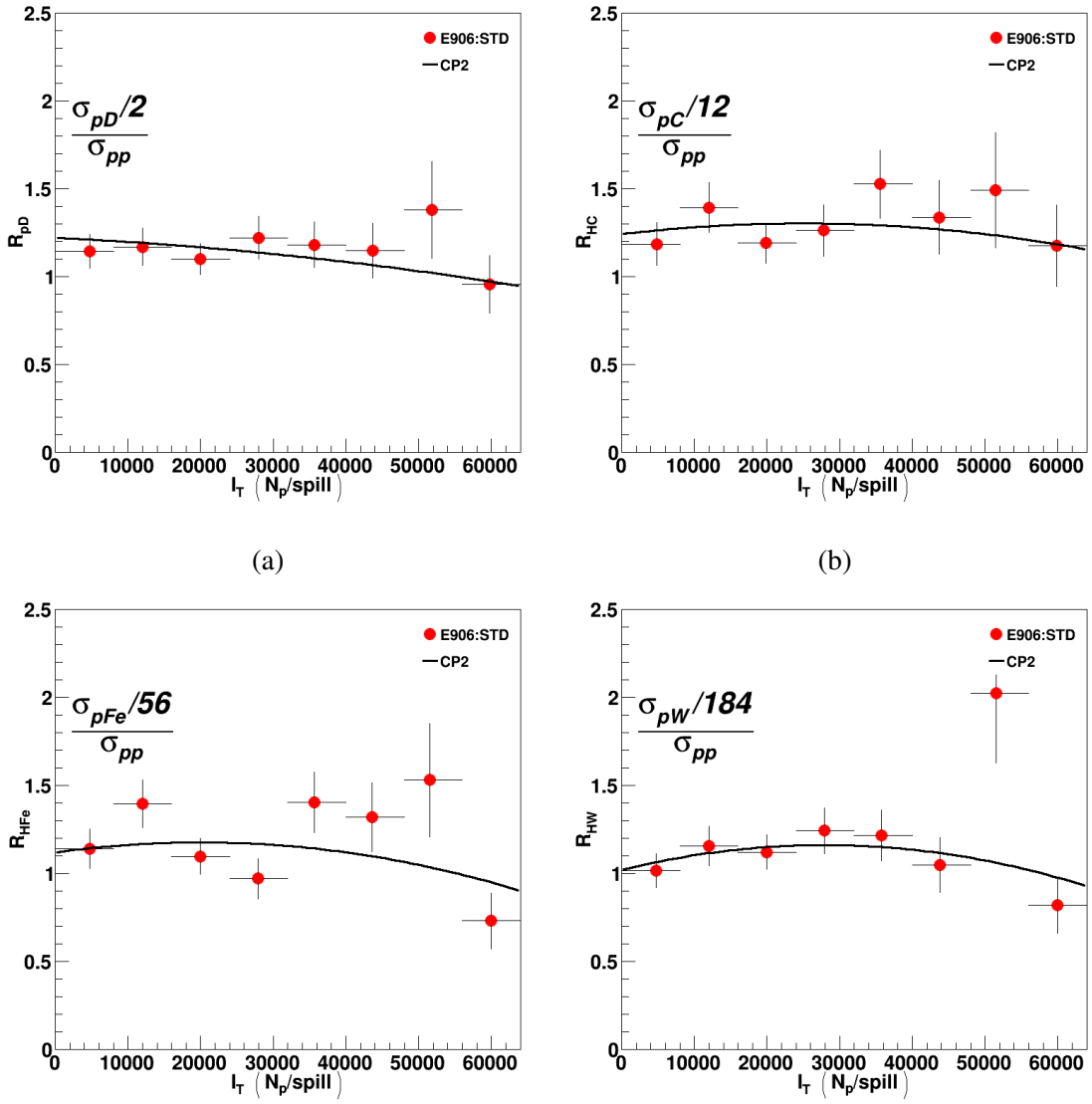
Lastly, to correct for any ML cut uncertainty, ML_0 was used as a secondary set and the difference between ML_2 and ML_0 was taken as the machine learning systematic error, with

$$\sigma_{ML} = |CSR_{ML_2} - CSR_{ML_0}|. \tag{4.7}$$

There are some ancillary systematic errors, such as deuterium contamination, target length difference between the liquid targets, and raw proton accuracy. But they have been shown to be negligible [117].

4.1.1 Standard Analysis Results

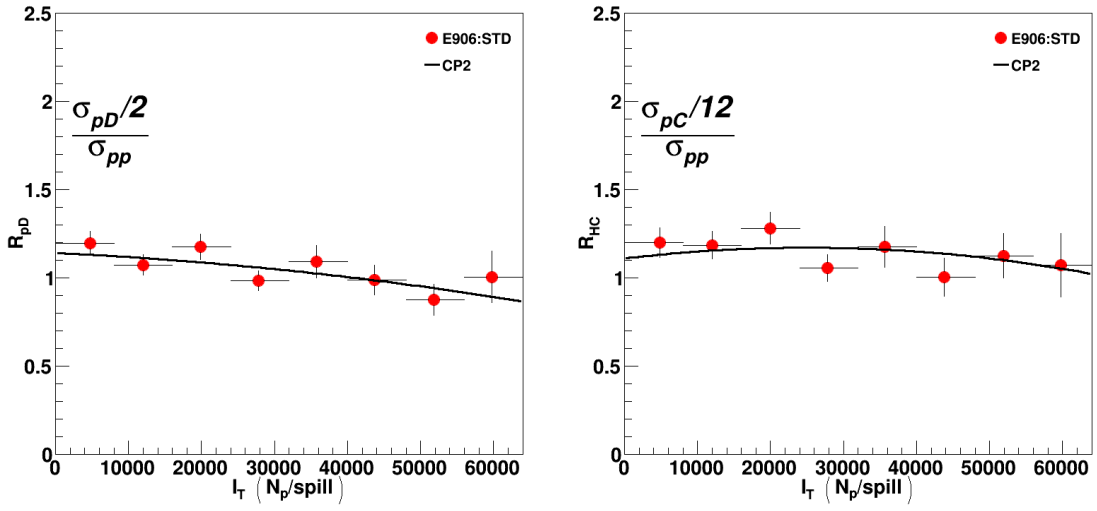
Figs. 4.1 - 4.6 illustrate the I_T dependence within each x_T bin and for the R_{pD} and R_{HA} CSRs using the STD constraints. The nuclear dependence CSR plots are shown in Figs. 4.7 - 4.12. In both sets of figures the CSR is depicted on the vertical axis, the I_T is listed on the horizontal axis, and the statistical errors are shown.



(a) (b)

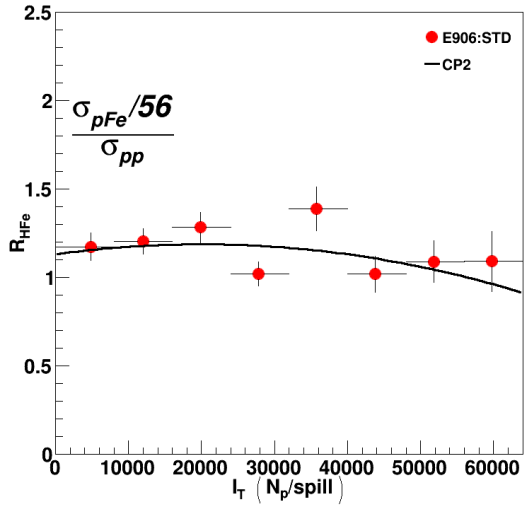
(c) (d)

Figure 4.1: Extracted Target/LH_2 CSR using the STD constraints with parameter values for $x_T \in [0.130-0.160)$: a) the R_{pD} results, b) the R_{HC} results, c) the R_{HFe} results, and d) the R_{HW} results.

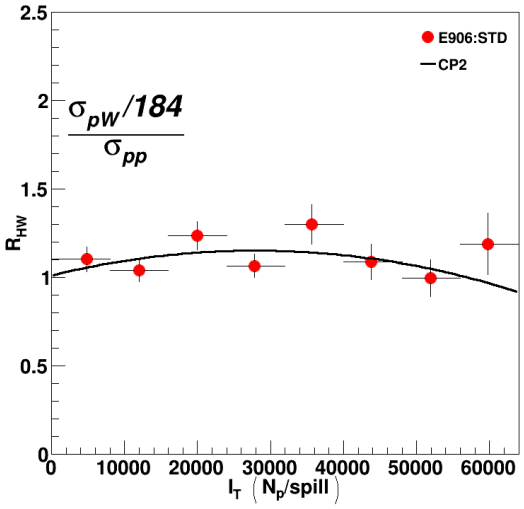


(a)

(b)

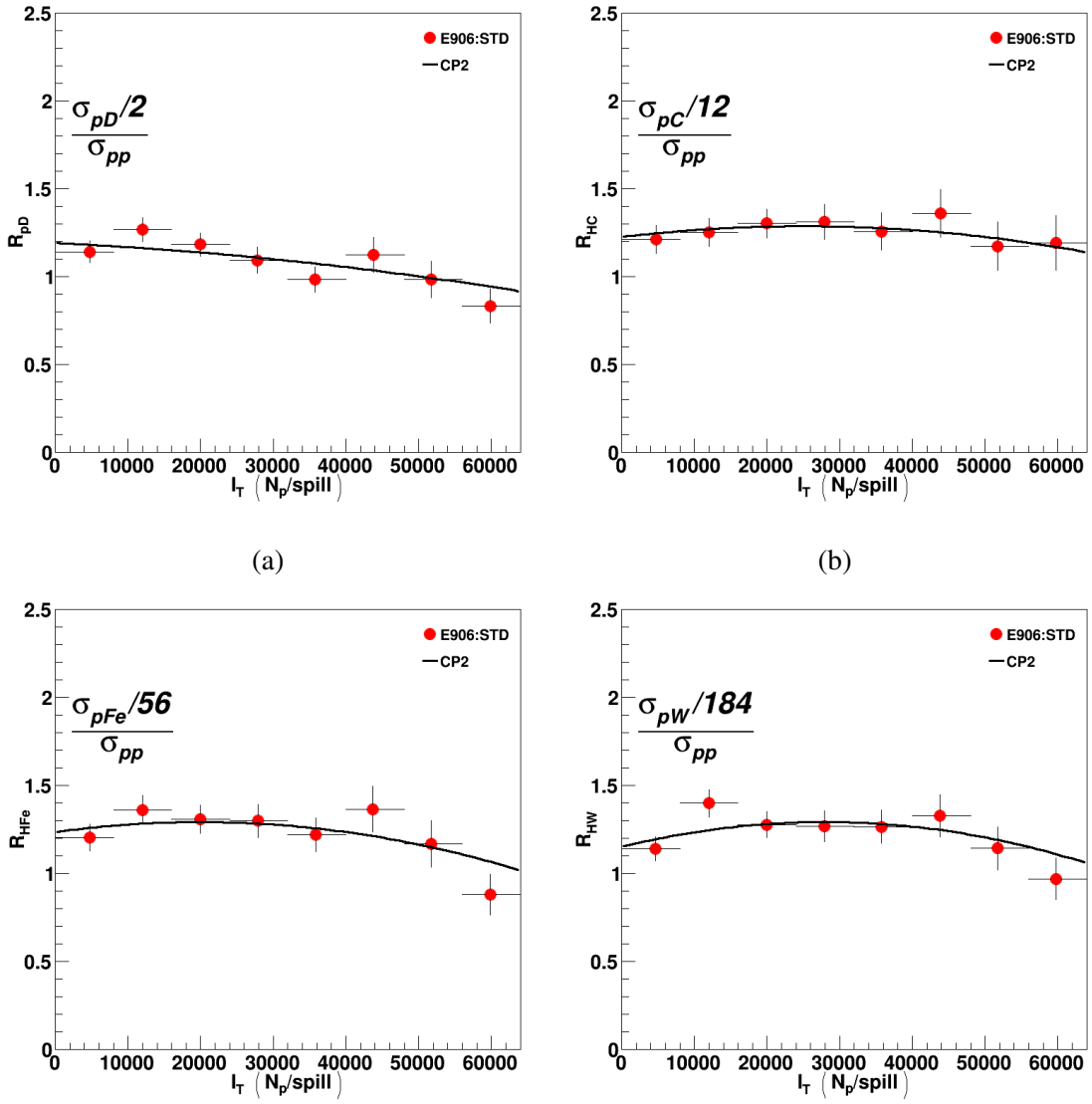


(c)



(d)

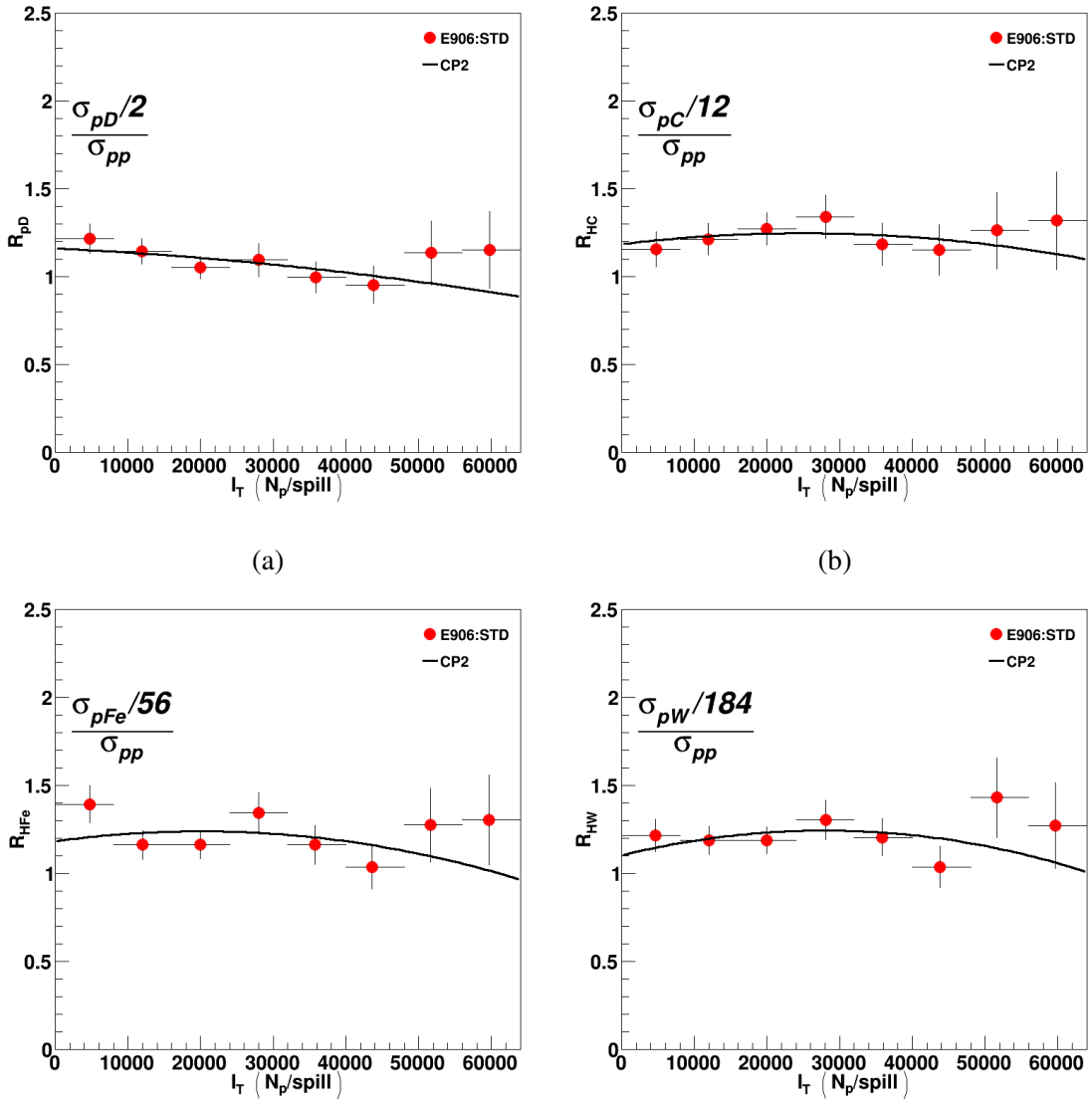
Figure 4.2: Extracted Target/LH_2 CSR using the STD constraints with parameter values for $x_T \in [0.160-0.195)$: a) the R_{pD} results, b) the R_{HC} results, c) the R_{HFe} results, and d) the R_{HW} results.



(a) (b)

(c) (d)

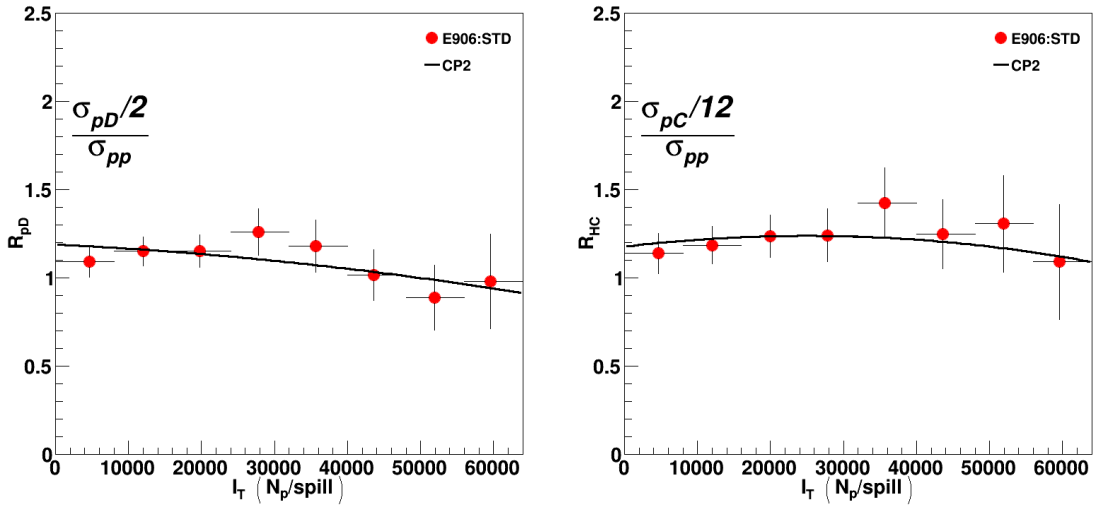
Figure 4.3: Extracted Target/LH_2 CSR using the STD constraints with parameter values for $x_T \in [0.195-0.240)$: a) the R_{pD} results, b) the R_{HC} results, c) the R_{HFe} results, and d) the R_{HW} results.



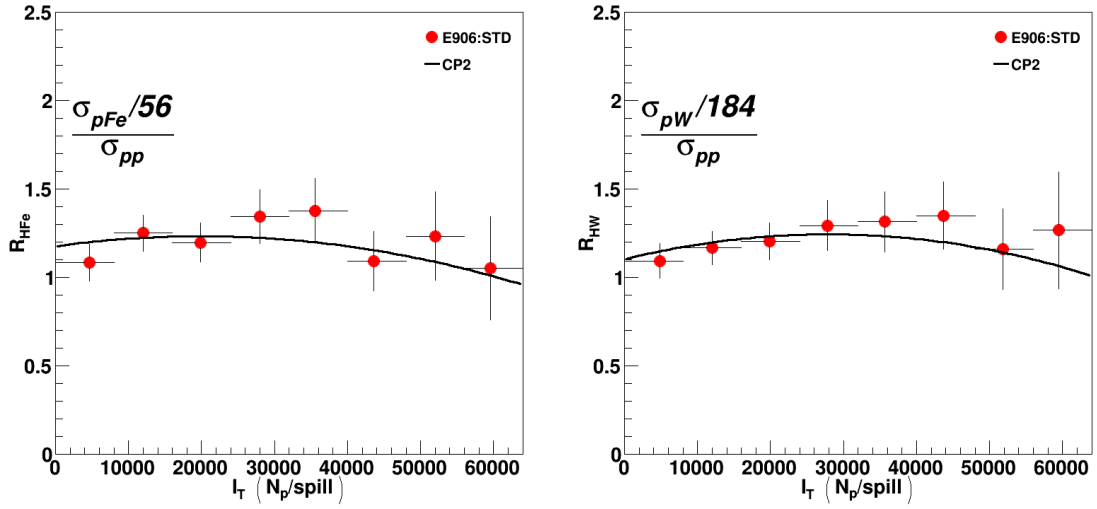
(a) (b)

(c) (d)

Figure 4.4: Extracted Target/LH_2 CSR using the STD constraints with parameter values for $x_T \in [0.240-0.290)$: a) the R_{pD} results, b) the R_{HC} results, c) the R_{HFe} results, and d) the R_{HW} results.

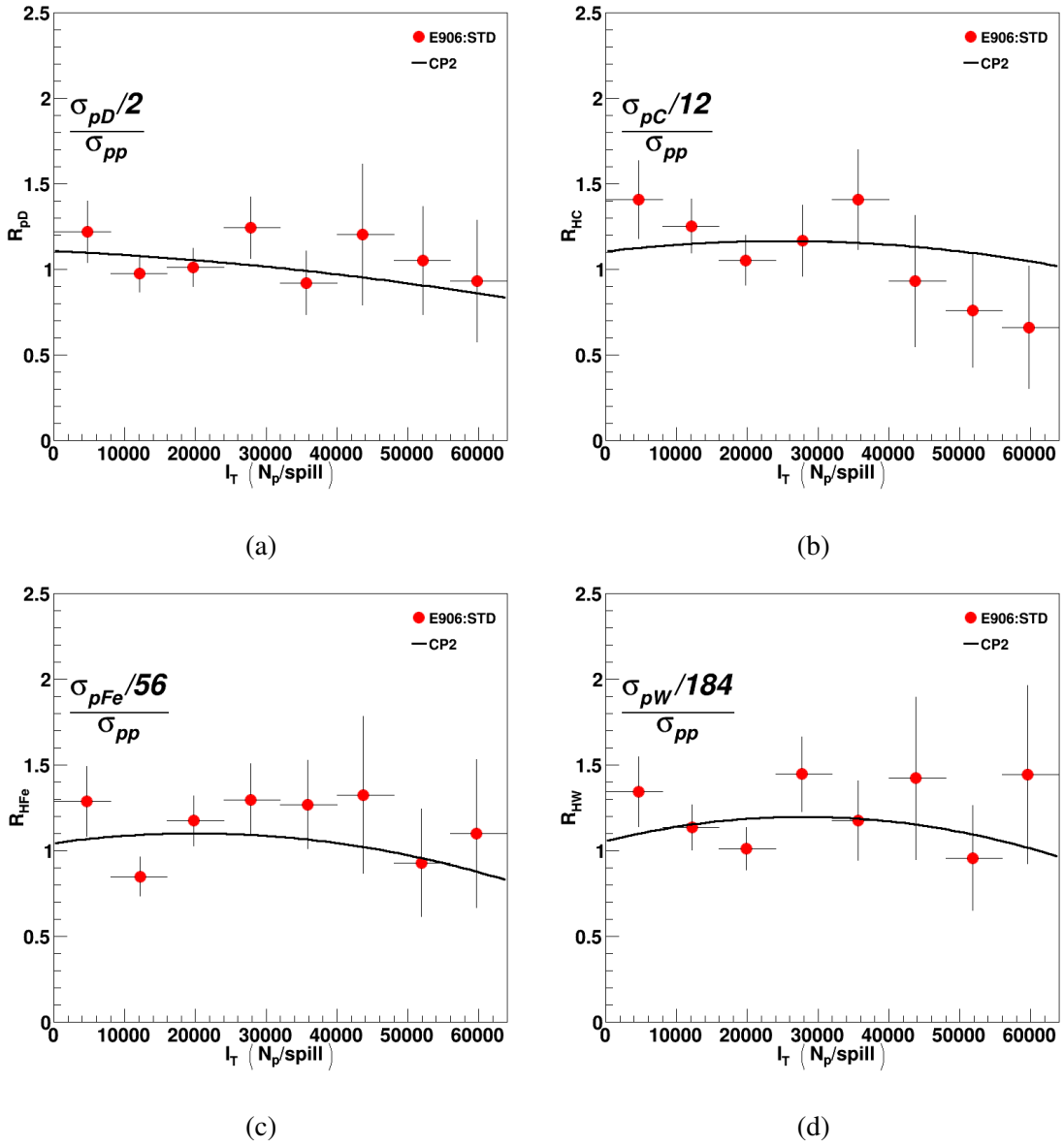


(a) (b)



(c) (d)

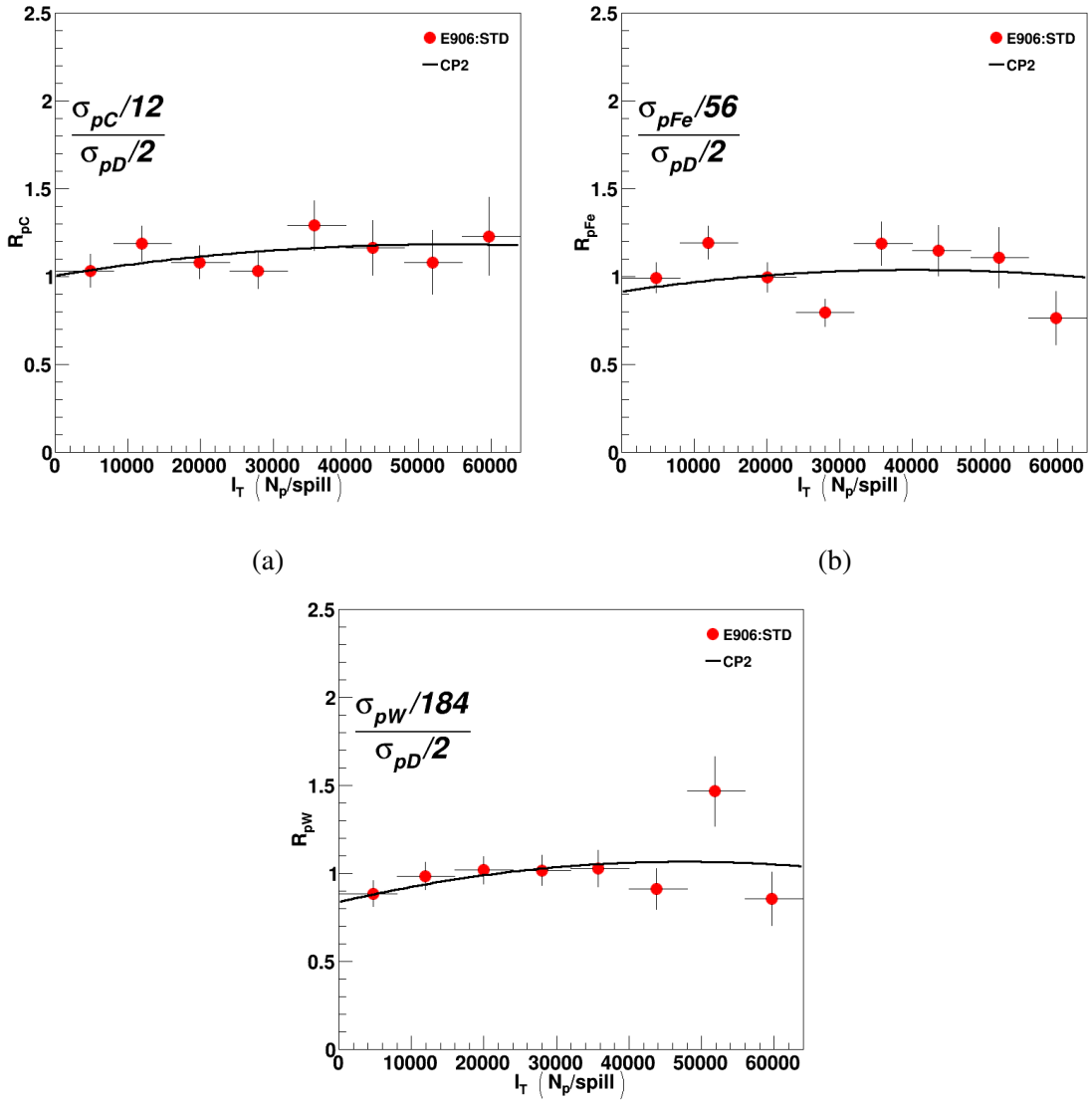
Figure 4.5: Extracted Target/LH_2 CSR using the STD constraints with parameter values for $x_T \in [0.290-0.350)$: a) the R_{pD} results, b) the R_{HC} results, c) the R_{HFe} results, and d) the R_{HW} results.



(a) (b)

(c) (d)

Figure 4.6: Extracted Target/LH_2 CSR using the STD constraints with parameter values for $x_T \in [0.350-0.450)$: a) the R_{pD} results, b) the R_{HC} results, c) the R_{HFe} results, and d) the R_{HW} results.



(a)

(b)

(c)

Figure 4.7: Extracted R_{pA} using the STD constraints with parameter values for $x_T \in [0.130 - 0.160)$: a) the R_{pC} results, b) the R_{pFe} results, and c) the R_{pW} results.

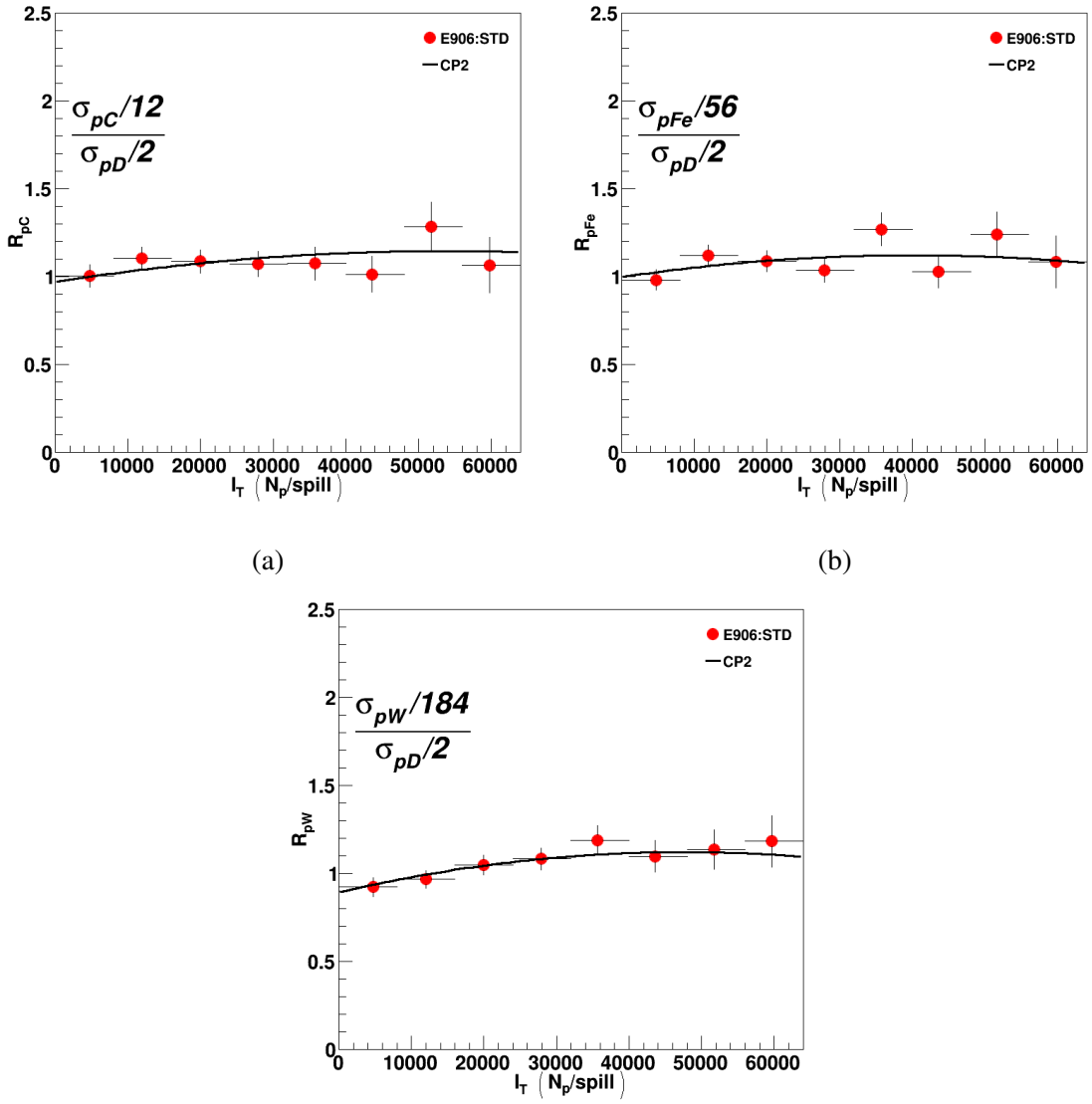


Figure 4.8: Extracted R_{pA} using the STD constraints with parameter values for $x_T \in [0.160 - 0.195)$: a) the R_{pC} results, b) the R_{pFe} results, and c) the R_{pW} results.

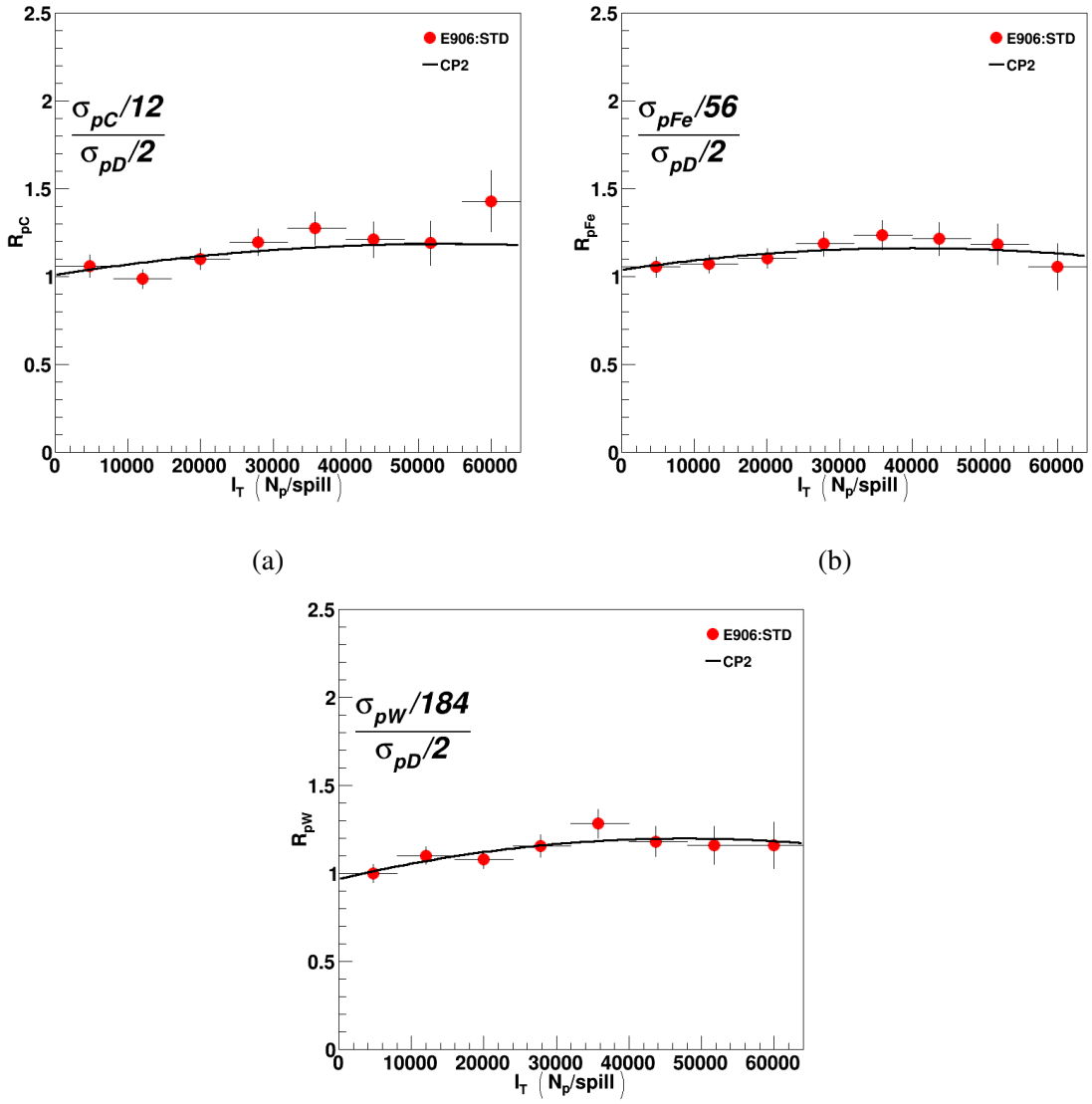
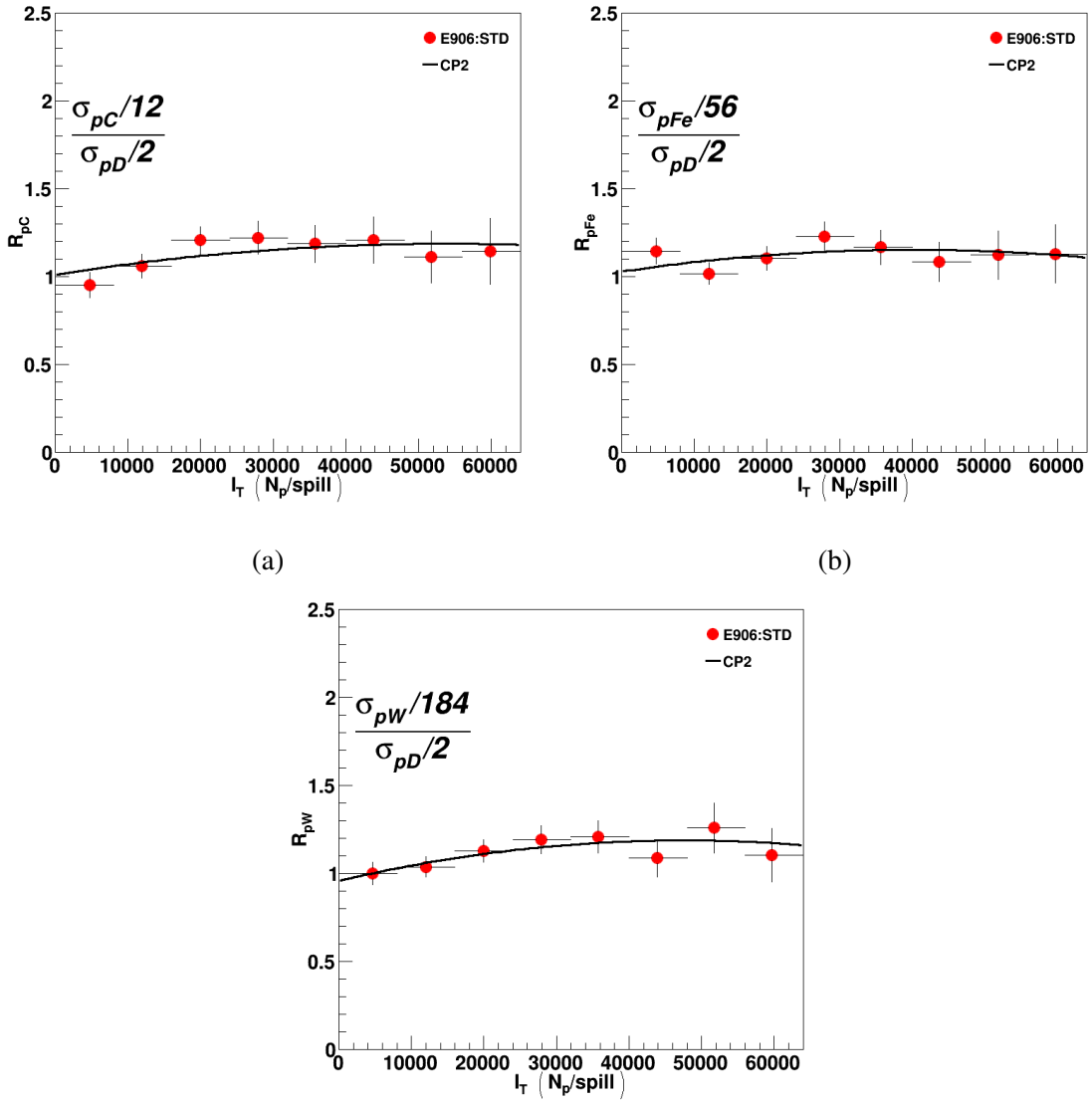


Figure 4.9: Extracted R_{pA} using the STD constraints with parameter values for $x_T \in [0.195 - 0.240)$: a) the R_{pC} results, b) the R_{pFe} results, and c) the R_{pW} results.



(a)

(b)

(c)

Figure 4.10: Extracted R_{pA} using the STD constraints with parameter values for $x_T \in [0.240 - 0.290)$: a) the R_{pC} results, b) the R_{pFe} results, and c) the R_{pW} results.

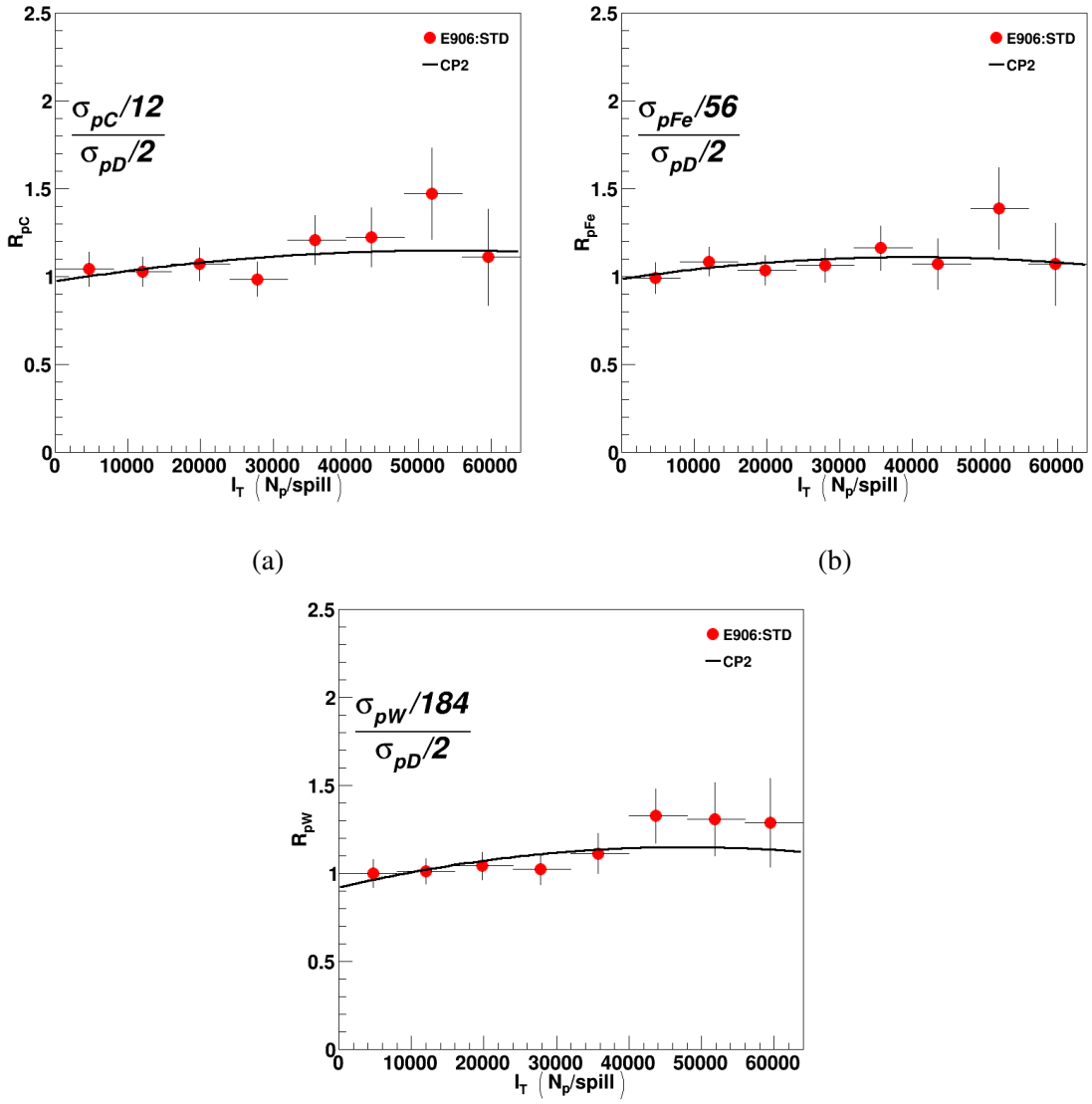
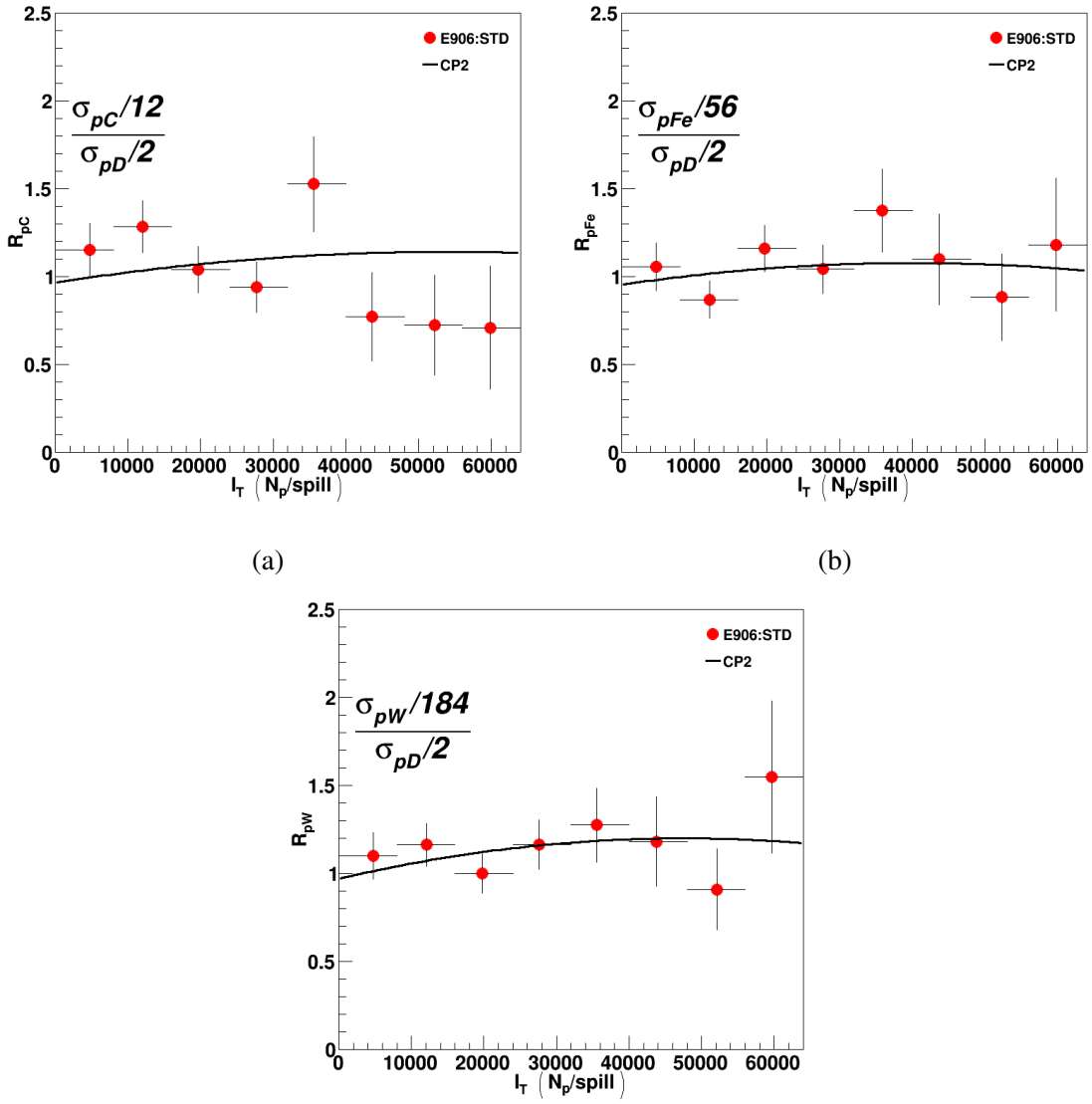


Figure 4.11: Extracted R_{pA} using the STD constraints with parameter values for $x_T \in [0.290 - 0.350)$: a) the R_{pC} results, b) the R_{pFe} results, and c) the R_{pW} results.



(a) (b)

(c)

Figure 4.12: Extracted R_{pA} using the STD constraints with parameter values for $x_T \in [0.350 - 0.450)$: a) the R_{pC} results, b) the R_{pFe} results, and c) the R_{pW} results.

The plots of the extracted STD CSR for the entire x_T range with statistical errors are shown in Fig. 4.13. The extrapolated CSR, the average x_T value in the x_T bin, the statistical errors, and systematic errors for each target for the R_{pD} and R_{HA} CSRs are listed in Appendix C. For the R_{pD} results, the E886 results are also included, showing a significant departure in the $x_T > 0.25$ region. Notably, the R_{pD} results are all above unity.

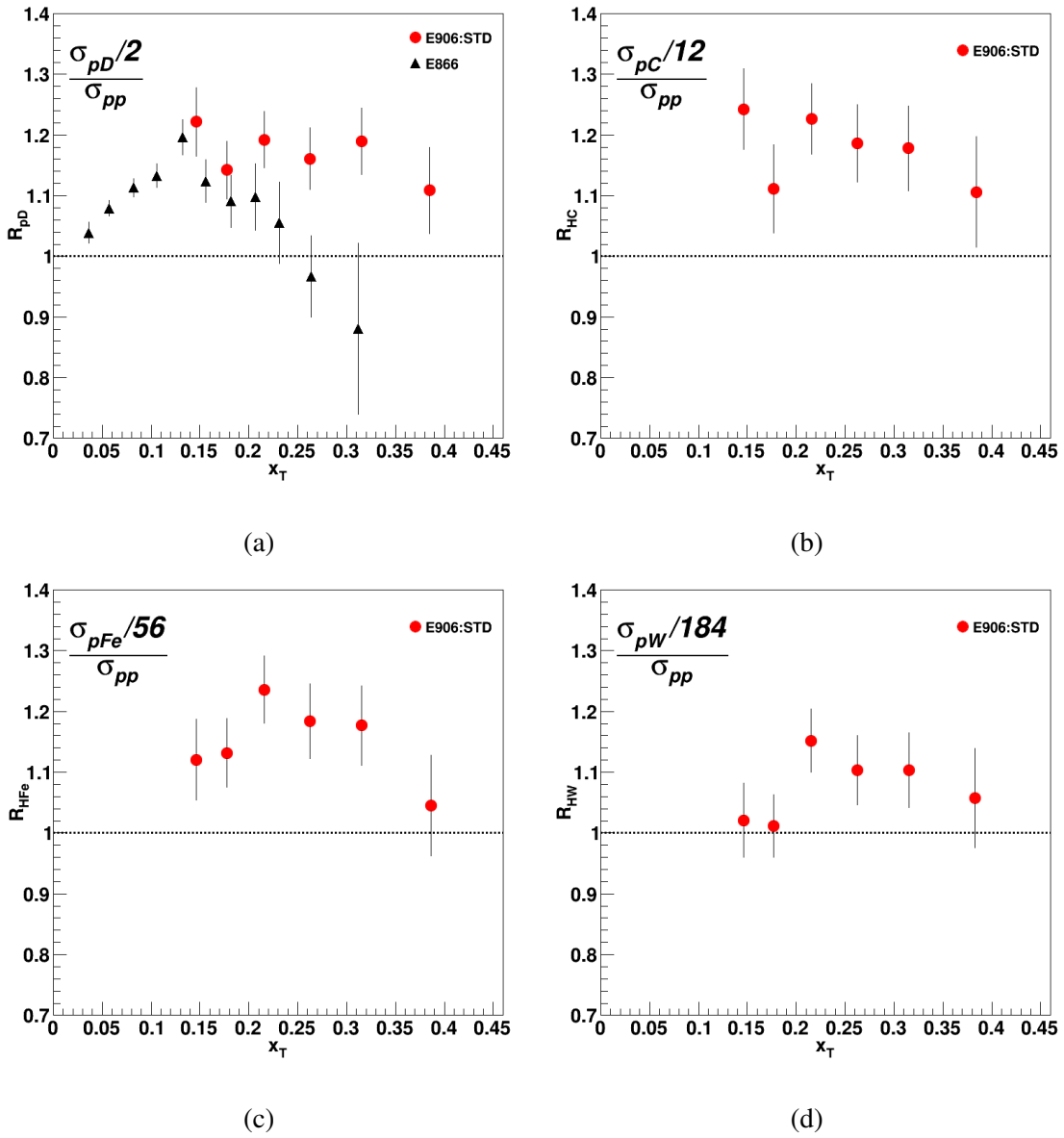


Figure 4.13: Plots of the R_{HA} using the STD constraints: a) the R_{pD} with E886 results from Ref. [40], b) the R_{HC} results, c) the R_{HFe} results, and d) R_{HW} results.

Figure 4.14 shows the plots of the extracted R_{pA} for the STD constraints with associated statistical errors along with the E772 results from Ref.[69] for comparison. The extrapolated CSR, the average x_T value in the x_T bin, the statistical errors, and systematic errors for each target for the R_{pA} CSRs are listed in Appendix C. The E772 results and the STD results are fairly consistent across all targets, as in the results are within the statistical errors of each measurement.

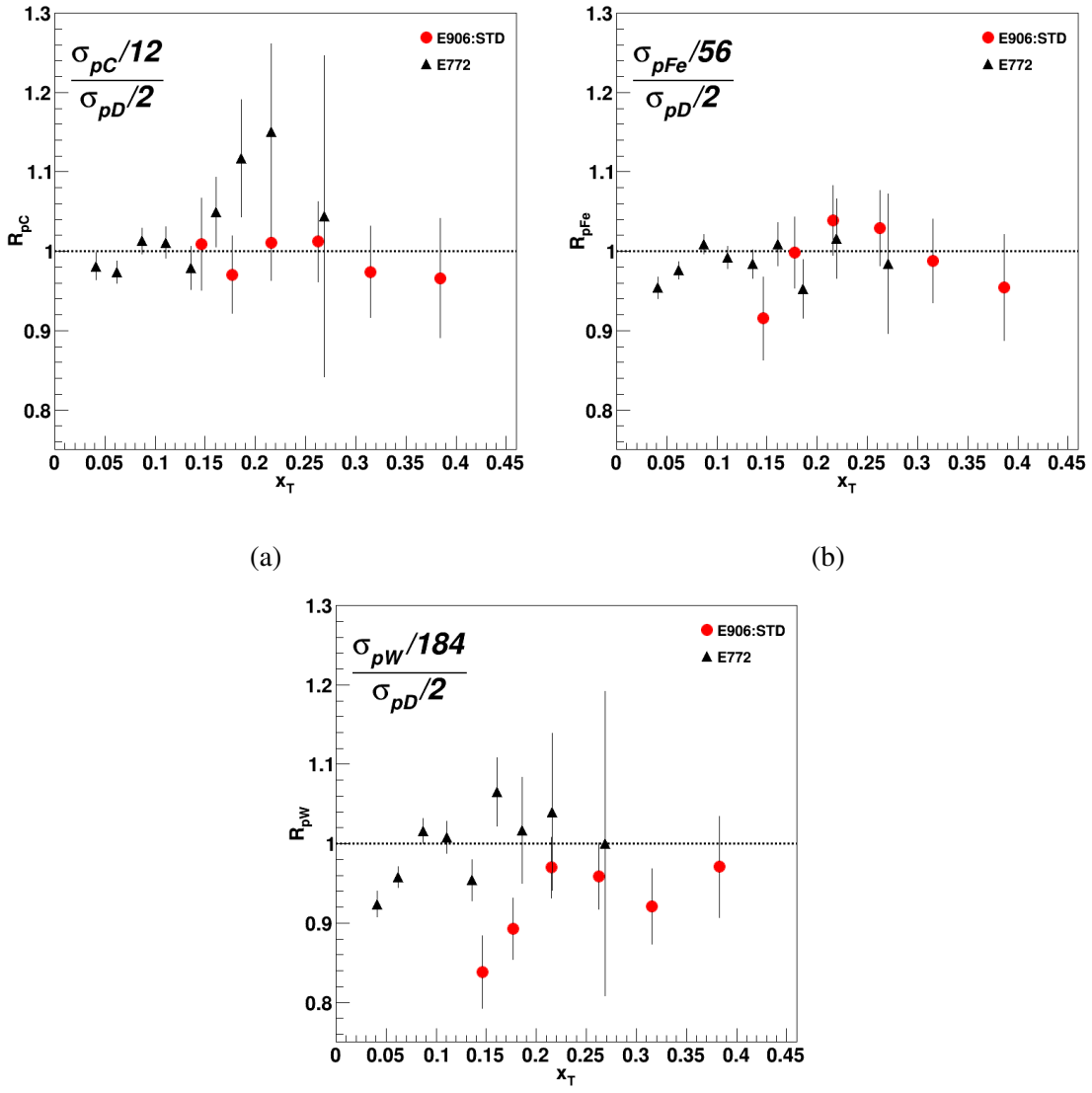


Figure 4.14: Plots of the R_{pA} using the STD constraints with the E772 results from Ref. [69]: a) the R_{pC} results, b) the R_{pFe} results, and c) R_{pW} results.

A comparison of the CSRs with their statistical errors, the quadratic parameter, and the linear parameter for the R_{pD} and R_{HA} are listed in Table 4.1 and the comparison of the R_{pA} values are in Table 4.2. The subscripts in the CSR parameters refer to the extracted CSR in that x_T bin with the bin ranges being $[0.130-0.160)$, $[0.160-0.195)$, $[0.195-0.240)$, $[0.240-0.290)$, $[0.290-0.350)$, and $[0.350-0.450)$ for bins 0 - 5, respectively.

Parameter	R_{pD}	R_{HC}	R_{HFe}	R_{HW}
Quadratic	$-3.7^{+5.5}_{-5.5} \cdot 10^{-11}$	$-9.8^{+7.0}_{-7.0} \cdot 10^{-11}$	$-1.4^{+0.65}_{-0.65} \cdot 10^{-10}$	$-1.8^{+0.62}_{-0.62} \cdot 10^{-10}$
Linear	$-2.0^{+3.3}_{-3.3} \cdot 10^{-6}$	$4.9^{+4.2}_{-4.2} \cdot 10^{-6}$	$5.7^{+4.0}_{-4.0} \cdot 10^{-6}$	$1.0^{+0.37}_{-0.37} \cdot 10^{-5}$
CSR_0	$1.222^{+0.057}_{-0.057}$	$1.243^{+0.073}_{-0.073}$	$1.120^{+0.066}_{-0.066}$	$1.021^{+0.061}_{-0.061}$
CSR_1	$1.142^{+0.047}_{-0.047}$	$1.111^{+0.059}_{-0.059}$	$1.132^{+0.057}_{-0.057}$	$1.012^{+0.052}_{-0.052}$
CSR_2	$1.192^{+0.046}_{-0.046}$	$1.227^{+0.058}_{-0.058}$	$1.236^{+0.056}_{-0.056}$	$1.152^{+0.052}_{-0.052}$
CSR_3	$1.161^{+0.051}_{-0.051}$	$1.186^{+0.064}_{-0.064}$	$1.184^{+0.062}_{-0.062}$	$1.104^{+0.057}_{-0.057}$
CSR_4	$1.190^{+0.055}_{-0.055}$	$1.178^{+0.070}_{-0.070}$	$1.177^{+0.066}_{-0.066}$	$1.104^{+0.062}_{-0.062}$
CSR_5	$1.109^{+0.071}_{-0.071}$	$1.106^{+0.091}_{-0.091}$	$1.045^{+0.083}_{-0.083}$	$1.057^{+0.082}_{-0.082}$
χ^2	29.13	22.04	53.46	37.74
χ^2/DOF	0.7283	0.5511	1.337	0.9435

Table 4.1: Extracted CSR and fit parameters with their associated statistical errors for R_{pD}, R_{HC}, R_{HFe}, and R_{HW} using the STD constraints. The number of degrees of freedom (DOF) was 40.

The R_{pD} CSR using the STD constraints is consistent with the E886 results for $x_T < 0.2$ before they begin to increasingly diverge. Even taking in the statistical and systematic errors, the results are not compatible in the higher x_T range. This divergence is significant due the STD results for all the bins being above unity. Ehlers *et al.*'s model of the nuclear effects in the proton-deuterium DY process are consistent with the R_{pD} results in the $x_T < 0.3$ region [134]. In their model, the neutron is not treated as free, but is subject to off-shell effects and some Fermi smearing. The off-shell effects arise from the struck nucleon being bound to the spectator nucleon. This binding energy requires the nucleon to have a lighter mass than a free nucleon. This in turn modifies the PDFs associated with the free nucleon.

The R_{HA} results are also above unity, but they seem to have a weak A dependence. Across

Parameter	R_{pC}	R_{pFe}	R_{pW}
Quadratic	$6.0\ ^{+6.3}_{-6.3} \cdot 10^{-11}$	$-7.7\ ^{+5.6}_{-5.6} \cdot 10^{-11}$	$-1.0\ ^{+0.49}_{-0.49} \cdot 10^{-10}$
Linear	$6.5\ ^{+3.7}_{-3.7} \cdot 10^{-6}$	$6.2\ ^{+3.3}_{-3.3} \cdot 10^{-6}$	$9.6\ ^{+2.8}_{-2.8} \cdot 10^{-6}$
CSR_0	$1.009\ ^{+0.058}_{-0.058}$	$0.9153\ ^{+0.052}_{-0.052}$	$0.8382\ ^{+0.046}_{-0.046}$
CSR_1	$0.9706\ ^{+0.049}_{-0.049}$	$0.9983\ ^{+0.045}_{-0.045}$	$0.8929\ ^{+0.039}_{-0.039}$
CSR_2	$1.011\ ^{+0.048}_{-0.048}$	$1.039\ ^{+0.044}_{-0.044}$	$0.9699\ ^{+0.039}_{-0.039}$
CSR_3	$1.012\ ^{+0.051}_{-0.051}$	$1.029\ ^{+0.047}_{-0.047}$	$0.9592\ ^{+0.042}_{-0.042}$
CSR_4	$0.9740\ ^{+0.057}_{-0.057}$	$0.9878\ ^{+0.053}_{-0.053}$	$0.9209\ ^{+0.047}_{-0.047}$
CSR_5	$0.9662\ ^{+0.075}_{-0.075}$	$0.9544\ ^{+0.067}_{-0.067}$	$0.9709\ ^{+0.064}_{-0.064}$
χ^2	37.37	41.85	23.83
χ^2/DOF	0.9343	1.046	0.5957

Table 4.2: Extracted CSR and fit parameters with their associated statistical errors for R_{pC}, R_{pFe}, and R_{pW} using the STD constraints. The number of degrees of freedom was 40.

almost each x_T bin, the CSR value decreases as the target's mass number increases, though the differences are largely within the statistical errors. Within the statistical error, the CSR results seem to be stable. The R_{pA} results are consistent with E772's results for carbon and iron. There is tension between the first two x_T bins for tungsten, with the STD results being significantly lower than the E772's. The CP2 fit for all of the CSRs yields a small reduced χ^2 for all of the targets, further suggesting that the intensity dependence for the CSR is a quadratic function in I_T. In addition to the R_{pD} prediction, Ehlers *et al.* also predict the R_{pA} ratio for isoscalar targets, which is only applicable for carbon. Their prediction is at $Q^2 = 54\ GeV^2$, which is not at E906's $Q^2 = 42\ GeV^2$ energy. However, the carbon results are consistent with this prediction. Thus assuming only a slight Q^2 dependence, the STD results may be compatible.

The STD results are consistent with some DY EMC effect. In this context, the EMC effect is a depression in the R_{pA} CSR below unity and the negative slope in the $0.2 \lessapprox x_T \lessapprox 0.7$ region. The actual evidence is not consistent across all of the targets. Tables 4.3 and 4.4 show the results from a linear and a constant fit, respectively. The fit was preformed on the data points in the $x_T > 0.2$ region. From the χ^2/DOF results, the linear fit is favored over the constant fit for carbon and iron,

but not for tungsten. Moreover, the errors on the slope, σ_1, are large so a slope of zero is consistent with the carbon and tungsten results. Incorporating the E772 data, as in Tables 4.5 and 4.6, does not change the trend. Carbon and iron are still more consistent with an EMC effect, with carbon and tungsten's slopes being zero is within the statistical error. More data for $x_T > 0.4$ is needed to make a more substantive declaration of an EMC effect for DY.

Parameter	R_{pC}	R_{pFe}	R_{pW}
p_0	1.083	1.155	0.9926
σ_0	0.1368	0.1235	0.1133
p_1	-0.3119	-0.5169	-0.1356
σ_1	0.4869	0.4367	0.4057
χ^2	0.09551	0.06304	0.6289
χ^2/DOF	0.04776	0.03152	0.3145

Table 4.3: Nuclear CSR fit results for R_{pC}, R_{pFe}, and R_{pW} using the STD constraints for a linear function. The number of degrees of freedom was 2.

Parameter	R_{pC}	R_{pFe}	R_{pW}
p_0	0.9968	1.012	0.9555
σ_0	0.02762	0.02538	0.02270
χ^2	0.5058	1.464	0.7406
χ^2/DOF	0.1686	0.4879	0.2469

Table 4.4: Nuclear CSR fit results for R_{pC}, R_{pFe}, and R_{pW} using the STD constraints for a constant function. The number of degrees of freedom was 3.

4.1.2 Machine Learning Results

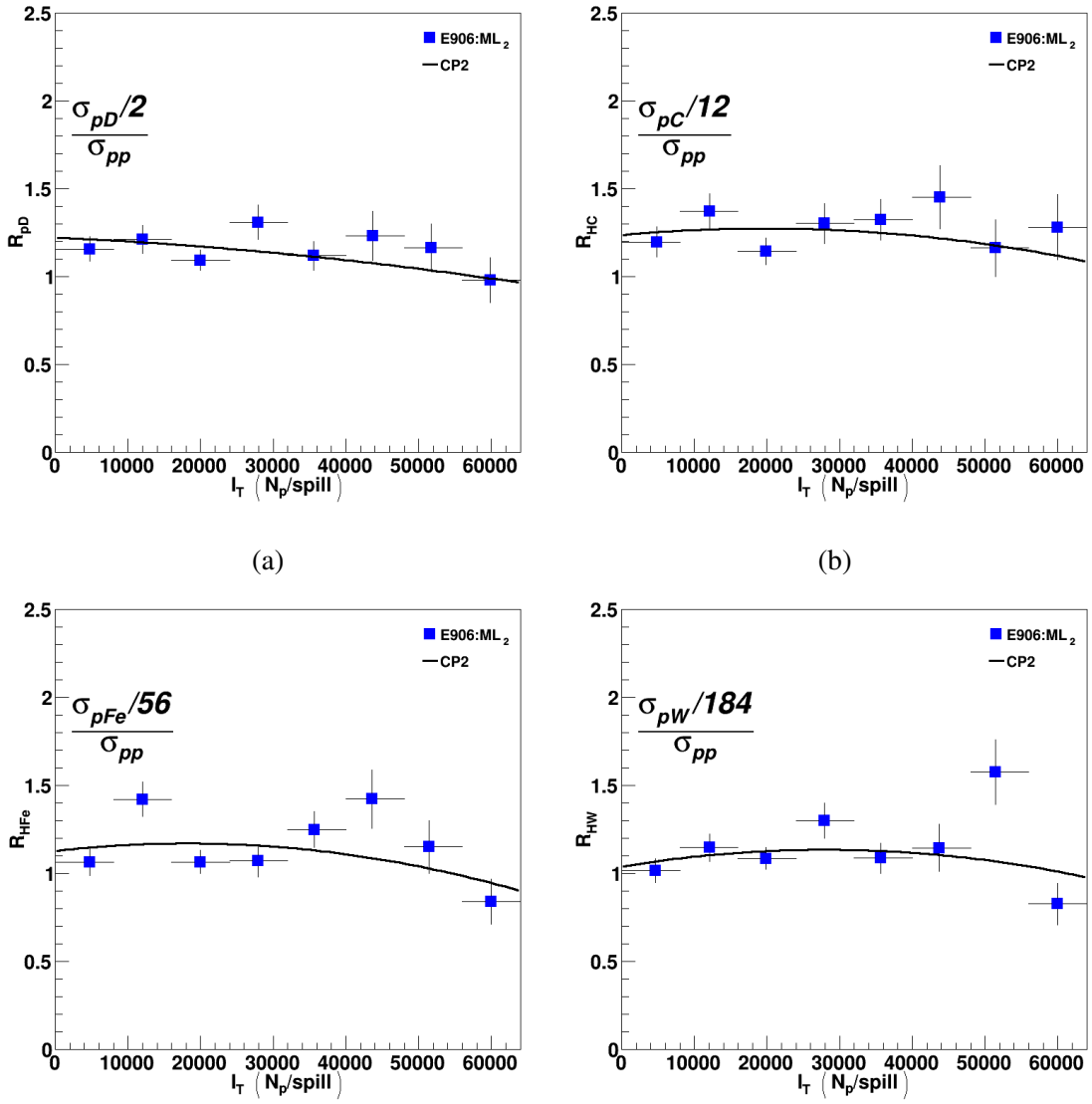
Figures 4.15 - 4.20 illustrate the I_T dependence within each x_T bin for the R_{pD} and R_{HA} CSRs using the ML_2 constraints. The nuclear dependence CSR plots are shown in Figs. 4.21 - 4.26. In both sets of figures the CSR is depicted on the vertical axis, the I_T is listed on the horizontal axis, and the statistical errors are shown.

Parameter	R_{pC}	R_{pFe}	R_{pW}
p_0	1.129	1.129	1.015
σ_0	0.1309	0.1079	0.1092
p_1	-0.4516	-0.4441	-0.2028
σ_1	0.4719	0.3981	0.3954
χ^2	1.472	0.3256	1.213
χ^2/DOF	0.3681	0.08140	0.3032

Table 4.5: Nuclear CSR fit results for R_{pC}, R_{pFe}, and R_{pW} using the STD constraints and the E772 results for a linear function. The number of degrees of freedom was 4.

Parameter	R_{pC}	R_{pFe}	R_{pW}
p_0	1.007	1.011	0.9602
σ_0	0.02656	0.02192	0.02198
χ^2	2.3882	1.570	1.476
χ^2/DOF	0.4776	0.3140	0.2952

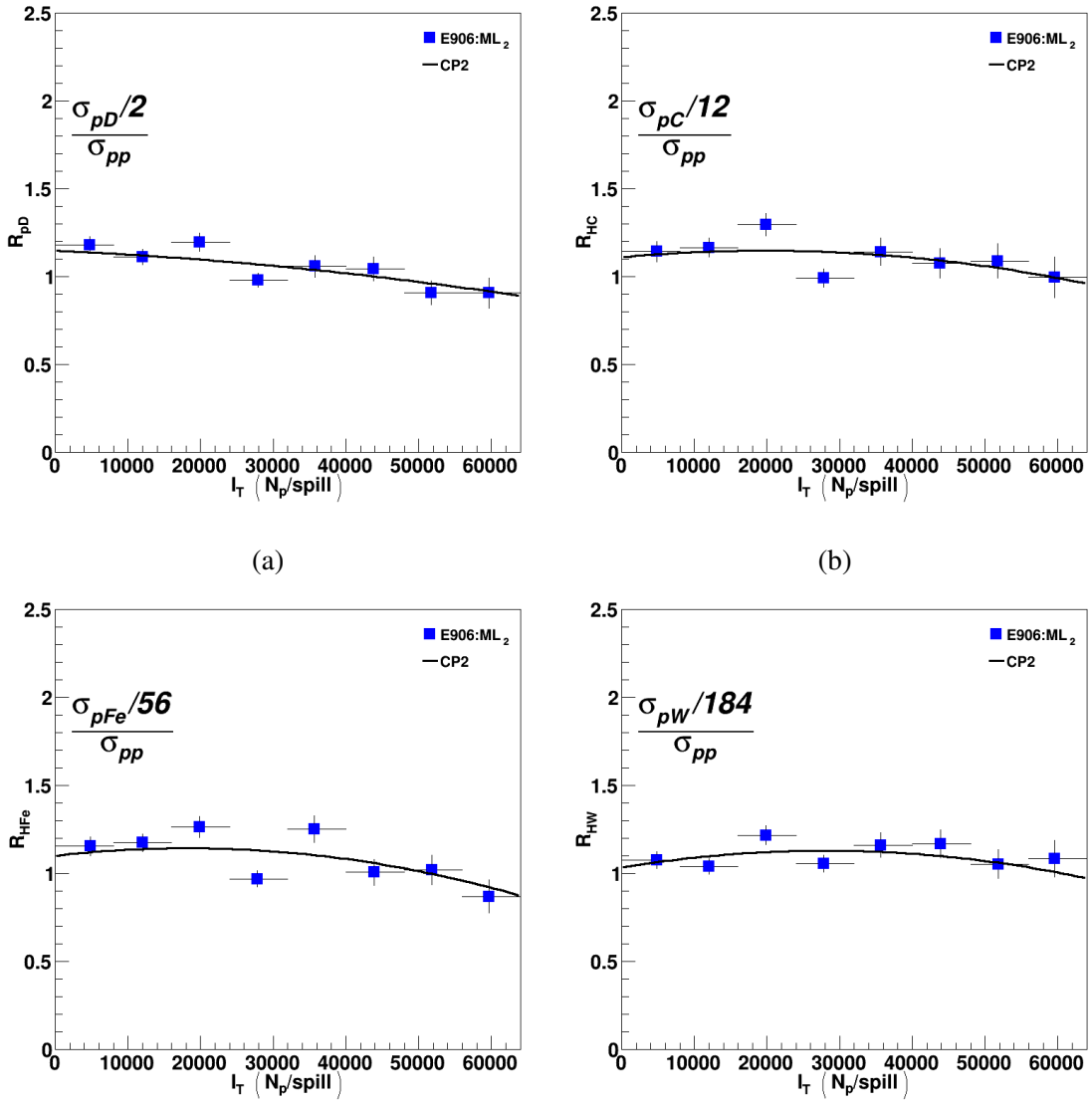
Table 4.6: Nuclear CSR fit results for R_{pC}, R_{pFe}, and R_{pW} using the STD constraints and the E772 results for a constant function. The number of degrees of freedom was 5.



(a) (b)

(c) (d)

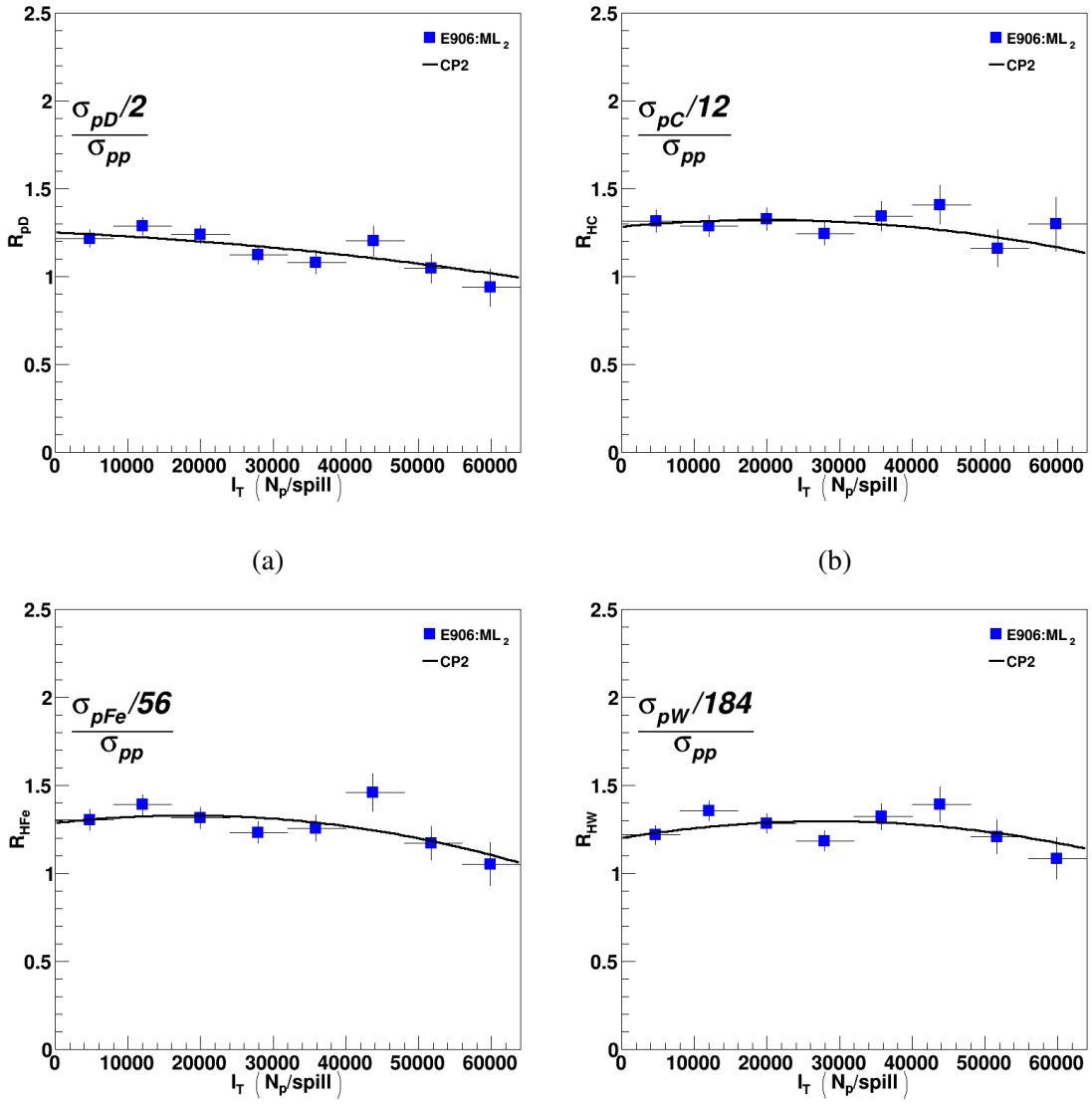
Figure 4.15: Extracted Target/LH_2 CSR using the ML_2 constraints with parameter values for $x_T \in [0.130-0.160)$: a) the R_{pD} results, b) the R_{HC} results, c) the R_{HFe} results, and d) the R_{HW} results.



(a) (b)

(c) (d)

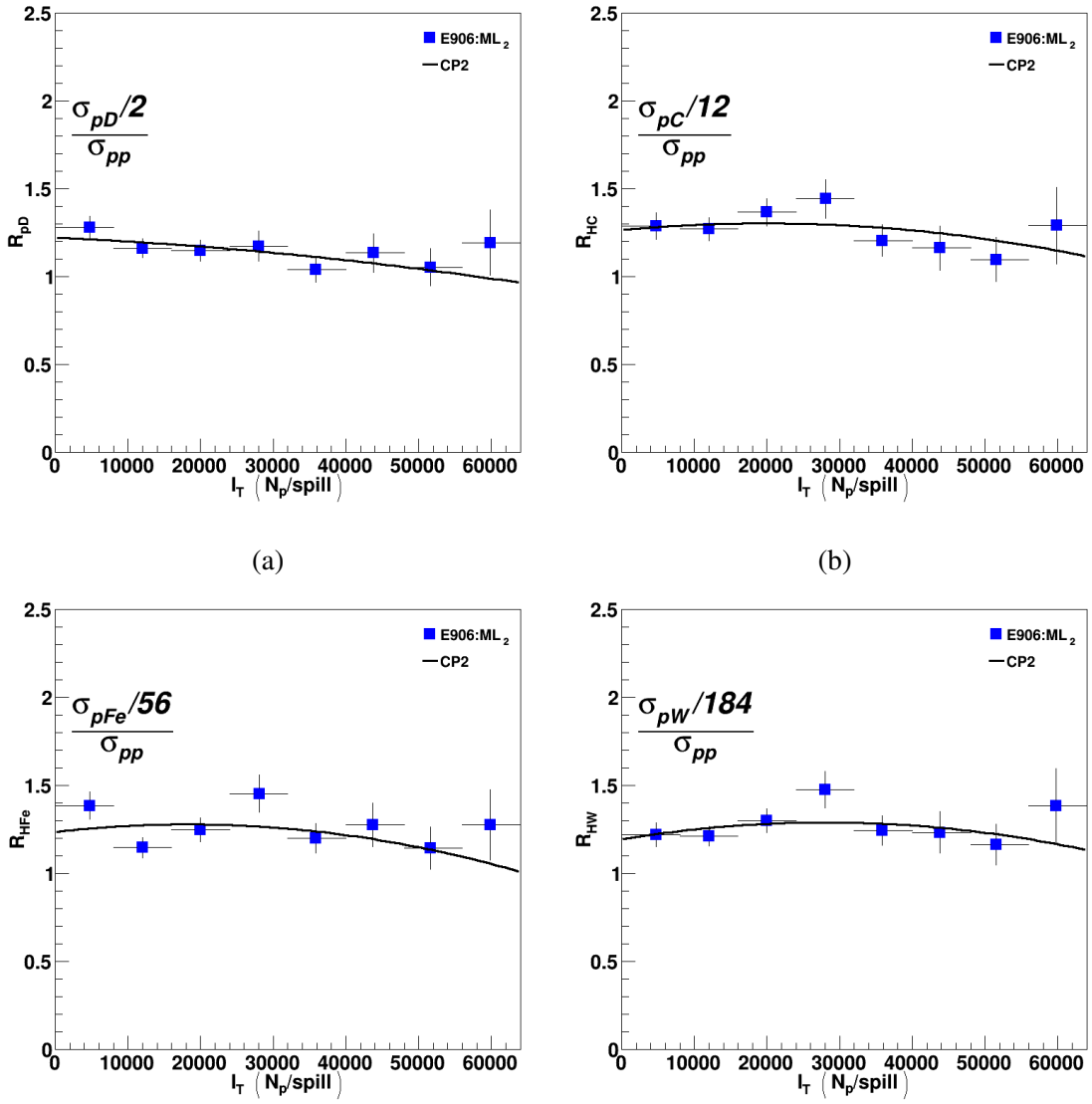
Figure 4.16: Extracted Target/LH_2 CSR using the ML_2 constraints with parameter values for $x_T \in [0.160-0.195)$: a) the R_{pD} results, b) the R_{HC} results, c) the R_{HFe} results, and d) the R_{HW} results.



(a) (b)

(c) (d)

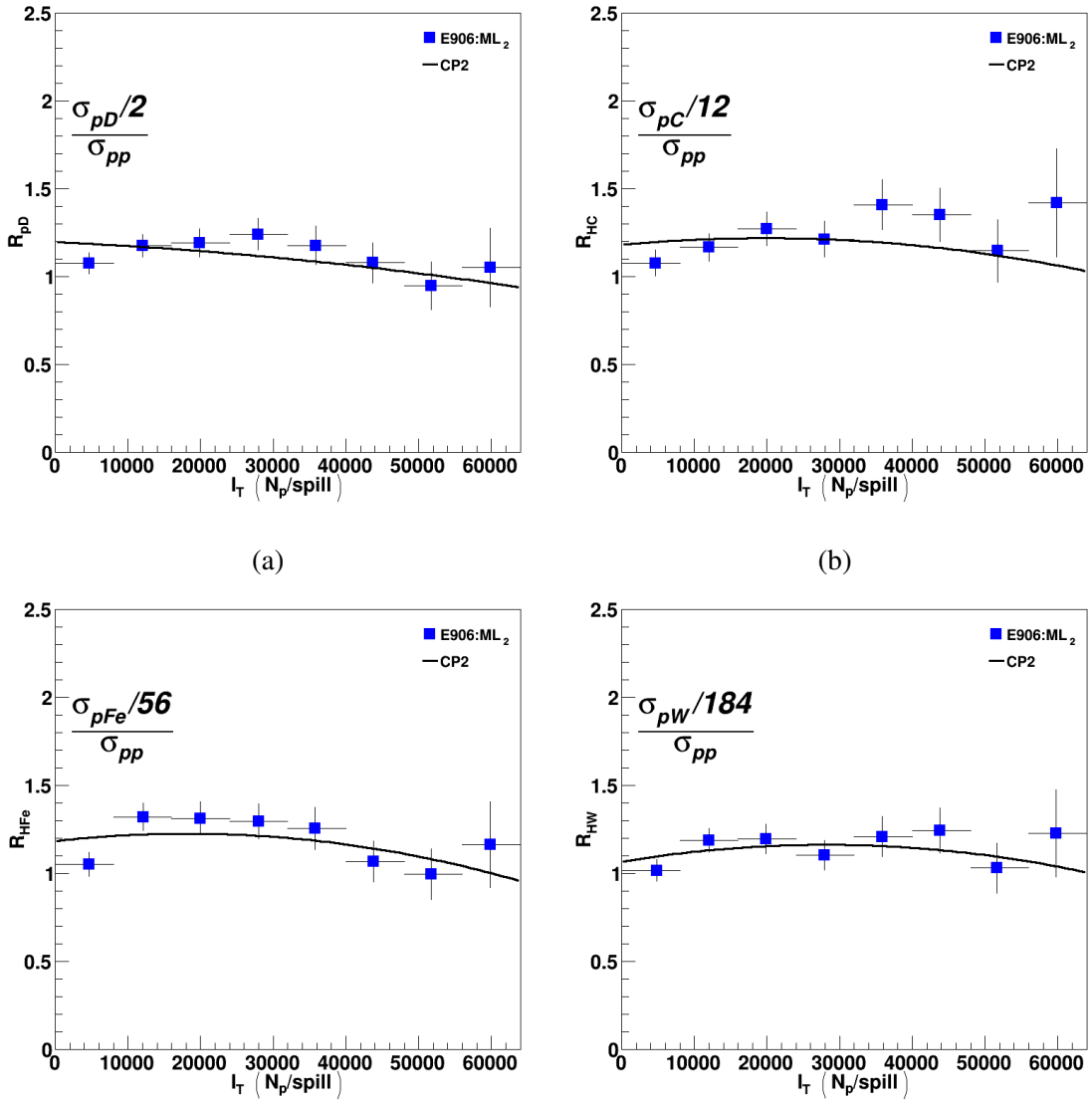
Figure 4.17: Extracted Target/LH_2 CSR using the ML_2 constraints with parameter values for $x_T \in [0.195-0.240)$: a) the R_{pD} results, b) the R_{HC} results, c) the R_{HFe} results, and d) the R_{HW} results.



(a) (b)

(c) (d)

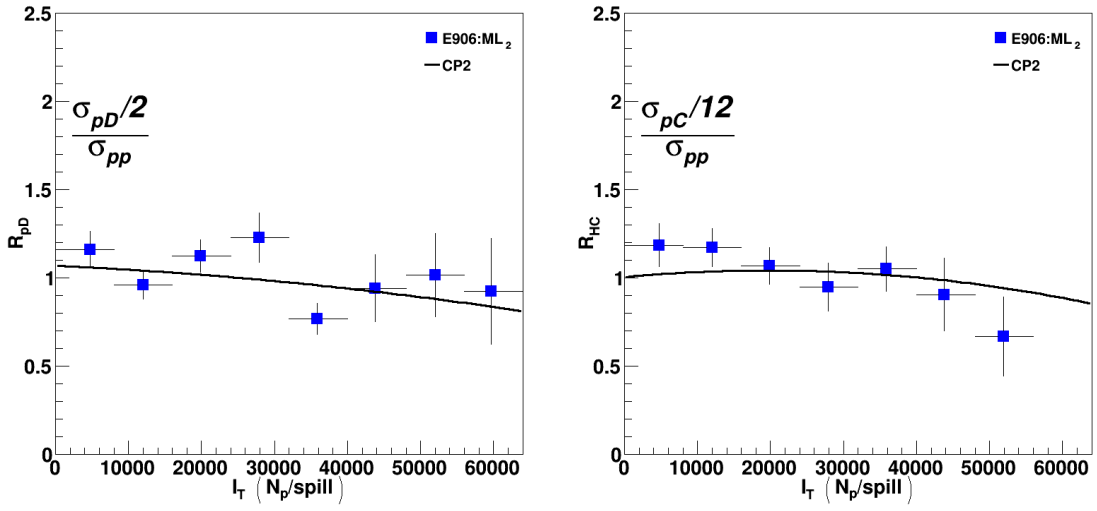
Figure 4.18: Extracted Target/LH_2 CSR using the ML_2 constraints with parameter values for $x_T \in [0.240-0.290)$: a) the R_{pD} results, b) the R_{HC} results, c) the R_{HFe} results, and d) the R_{HW} results.



(a) (b)

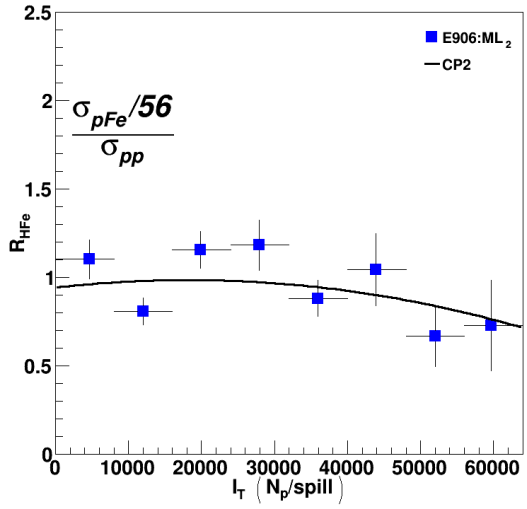
(c) (d)

Figure 4.19: Extracted Target/LH_2 CSR using the ML_2 constraints with parameter values for $x_T \in [0.290-0.350)$: a) the R_{pD} results, b) the R_{HC} results, c) the R_{HFe} results, and d) the R_{HW} results.

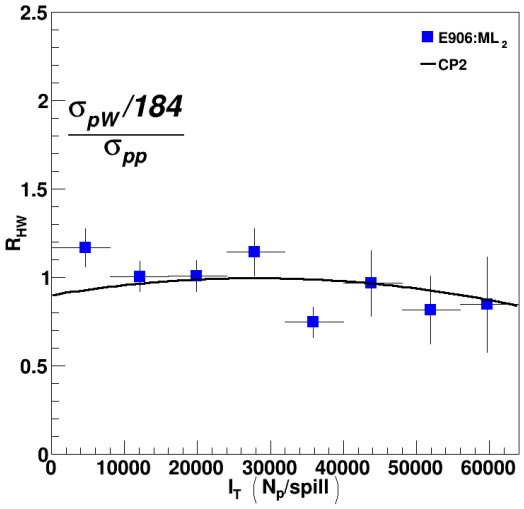


(a)

(b)



(c)



(d)

Figure 4.20: Extracted Target/LH_2 CSR using the ML_2 constraints with parameter values for $x_T \in [0.350-0.450)$: a) the R_{pD} results, b) the R_{HC} results, c) the R_{HFe} results, and d) the R_{HW} results.

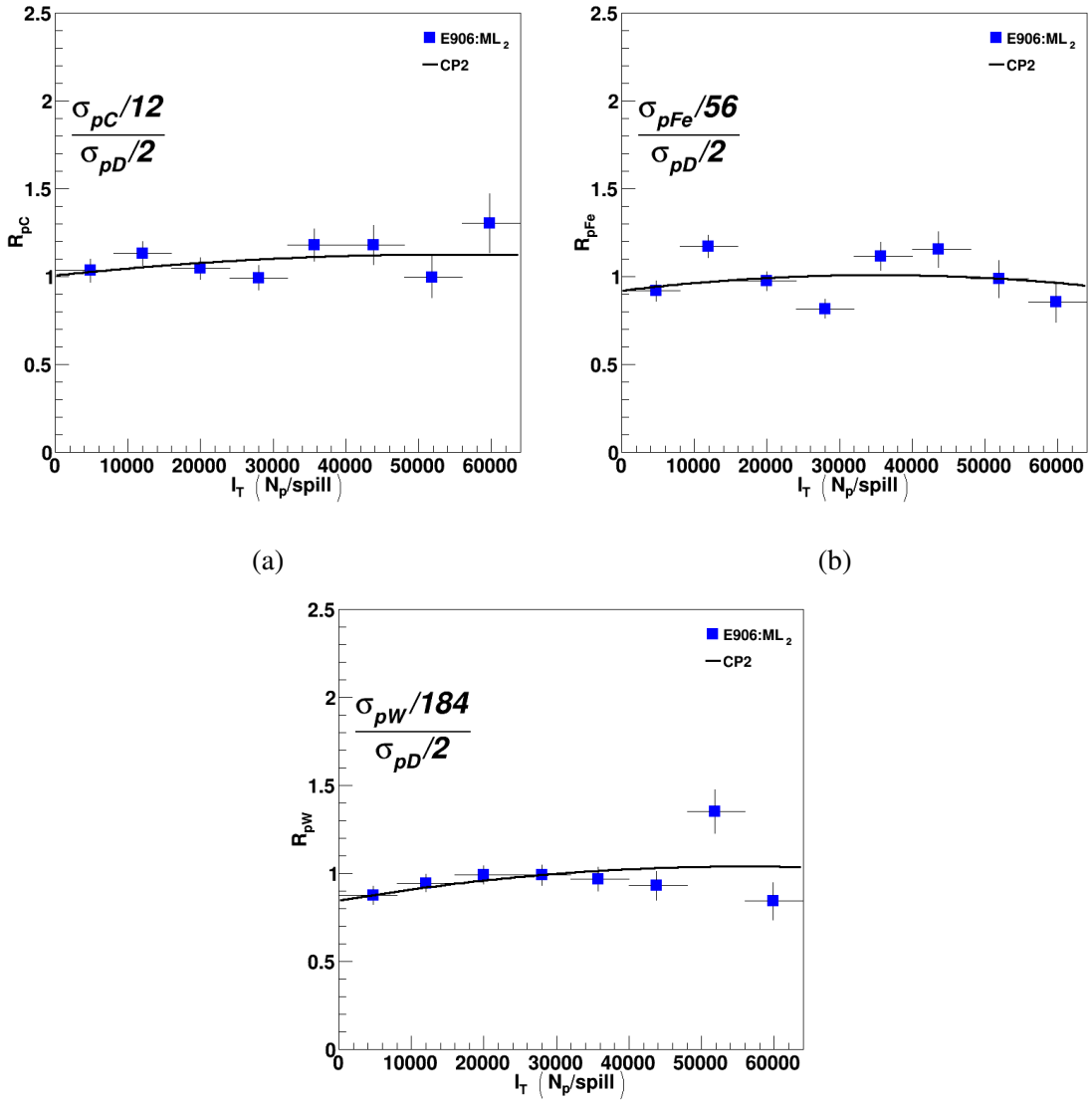


Figure 4.21: Extracted R_{pA} using the ML$_2$ constraints with parameter values for $x_T \in [0.130 - 0.160)$: a) the R_{pC} results, b) the R_{pFe} results, and c) the R_{pW} results.

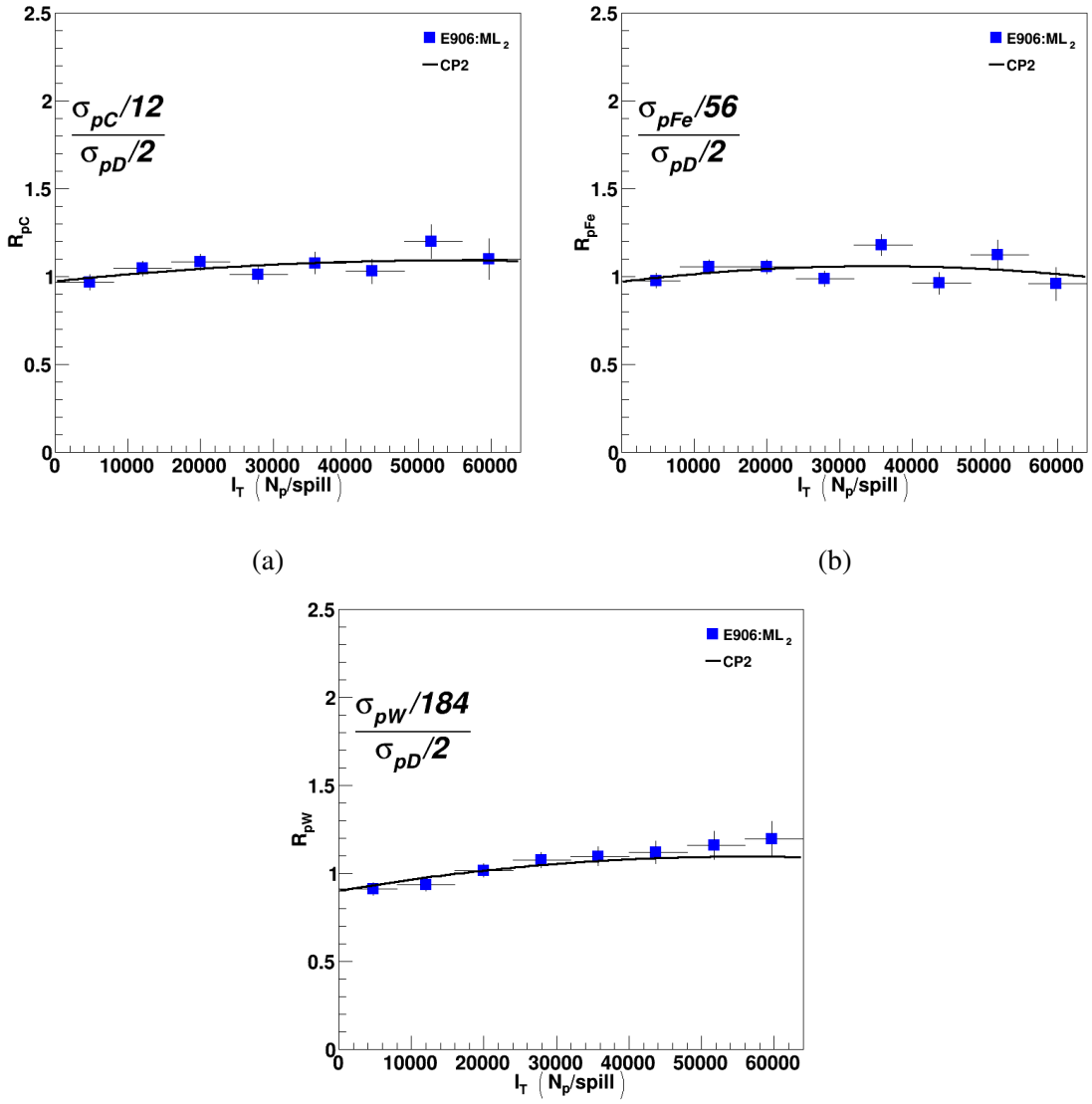


Figure 4.22: Extracted R_{pA} using the ML$_2$ constraints with parameter values for $x_T \in [0.160 - 0.195)$: a) the R_{pC} results, b) the R_{pFe} results, and c) the R_{pW} results.

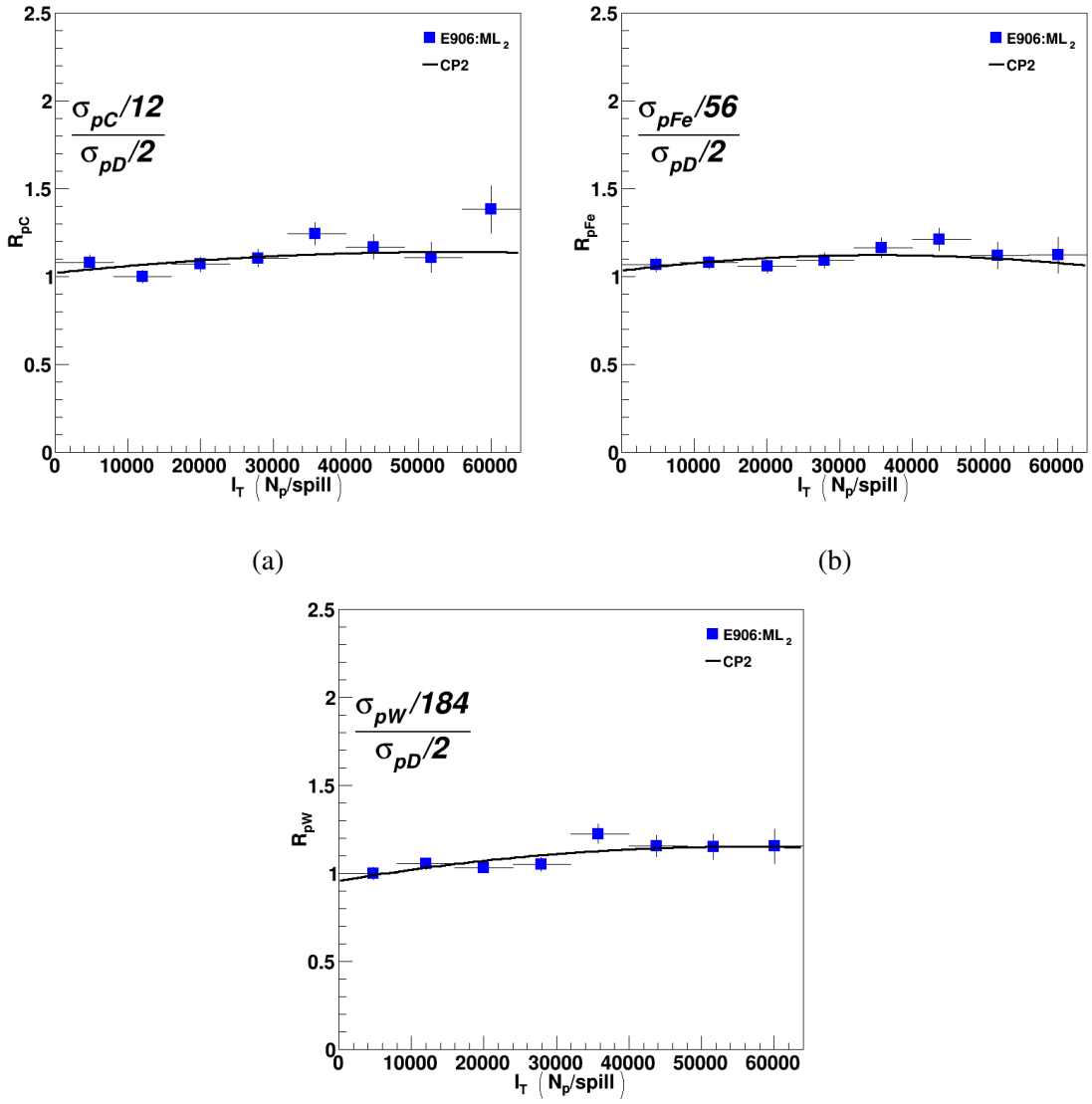


Figure 4.23: Extracted R_{pA} using the ML$_2$ constraints with parameter values for $x_T \in [0.195 - 0.240)$: a) the R_{pC} results, b) the R_{pFe} results, and c) the R_{pW} results.

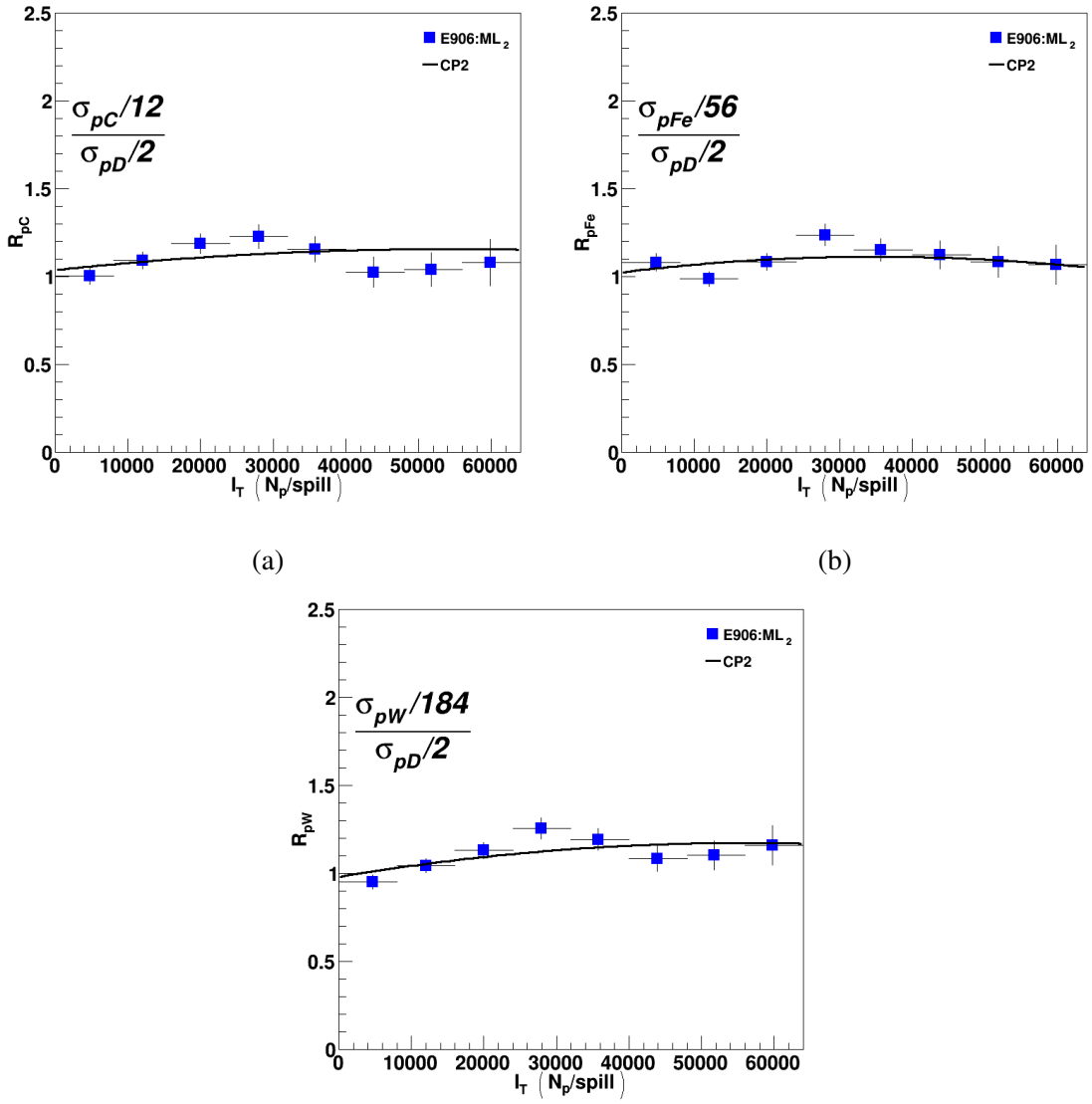


Figure 4.24: Extracted R_{pA} using the ML$_2$ constraints with parameter values for $x_T \in [0.240 - 0.290)$: a) the R_{pC} results, b) the R_{pFe} results, and c) the R_{pW} results.

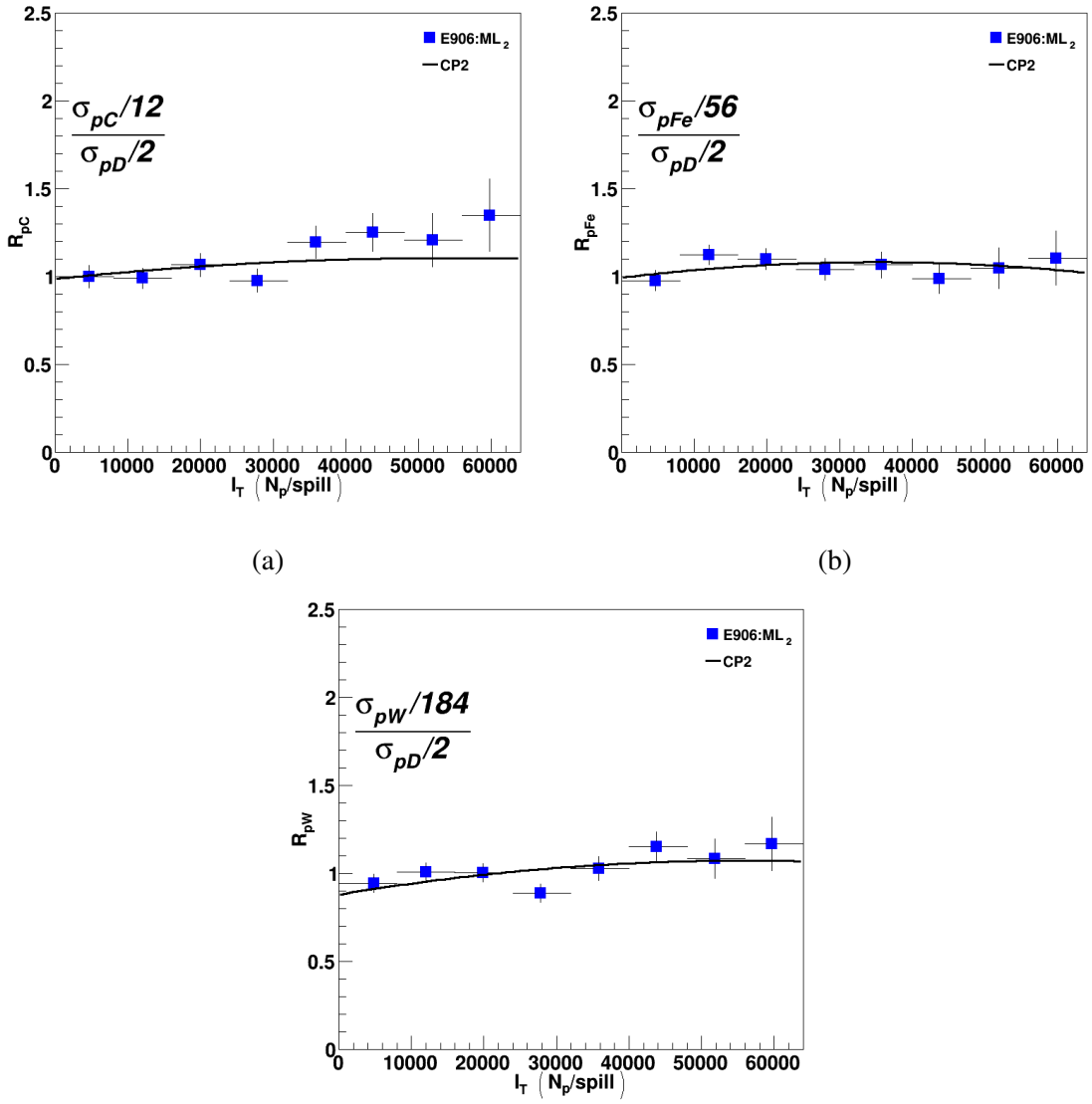


Figure 4.25: Extracted R_{pA} using the ML$_2$ constraints with parameter values for $x_T \in [0.290 - 0.350)$: a) the R_{pC} results, b) the R_{pFe} results, and c) the R_{pW} results.

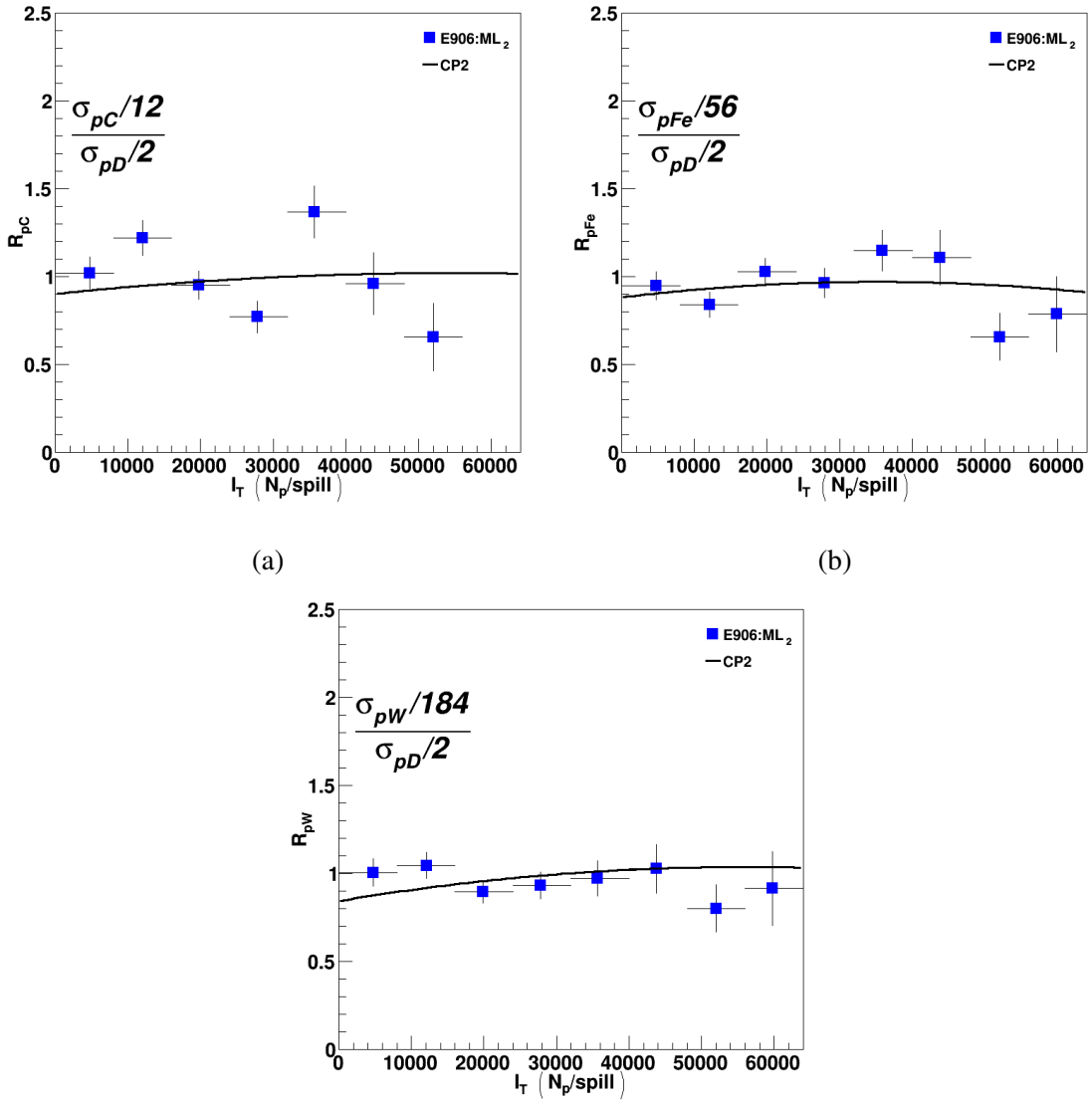


Figure 4.26: Extracted R_{pA} using the ML$_2$ constraints with parameter values for $x_T \in [0.350 - 0.450)$: a) the R_{pC} results, b) the R_{pFe} results, and c) the R_{pW} results.

The plots of the ML_2 extrapolated CSR with the statistical errors are shown in Fig. 4.27. The extrapolated CSR, the average x_T value in the x_T bin, the statistical errors, and systematic errors for each target for the R_{pD} and R_{HA} CSRs are listed in Appendix C. As with the STD R_{pD} results, the E886 results are also included. However, the departure between the E886 and the ML_2 begins slightly earlier at $x_T = 0.2$ instead of $x_T = 0.25$, but the results are still above unity.

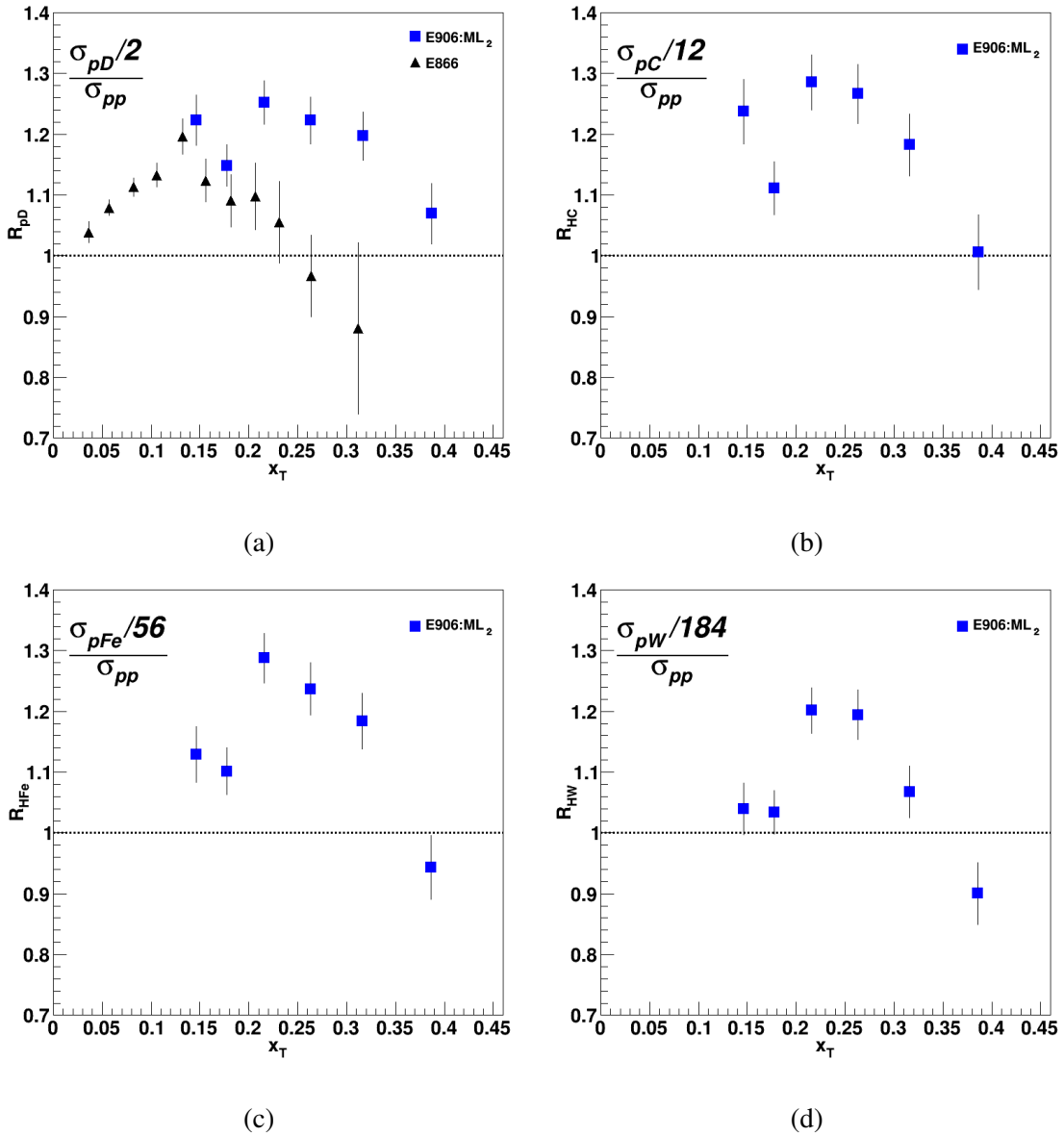


Figure 4.27: Plots of the R_{HA} using the ML_2 constraints: a) the R_{pD} with E886 results from Ref. [40], b) the R_{HC} results, c) the R_{HFe} results, and d) R_{HW} results.

In Fig. 4.28 are the plots of the extracted R_{pA} for the ML_2 constraints with associated statistical errors and the E772 results for comparison. The extrapolated CSR, the average x_T value in the x_T bin, the statistical errors, and systematic errors for each target for the R_{pA} CSRs are listed in Appendix C.

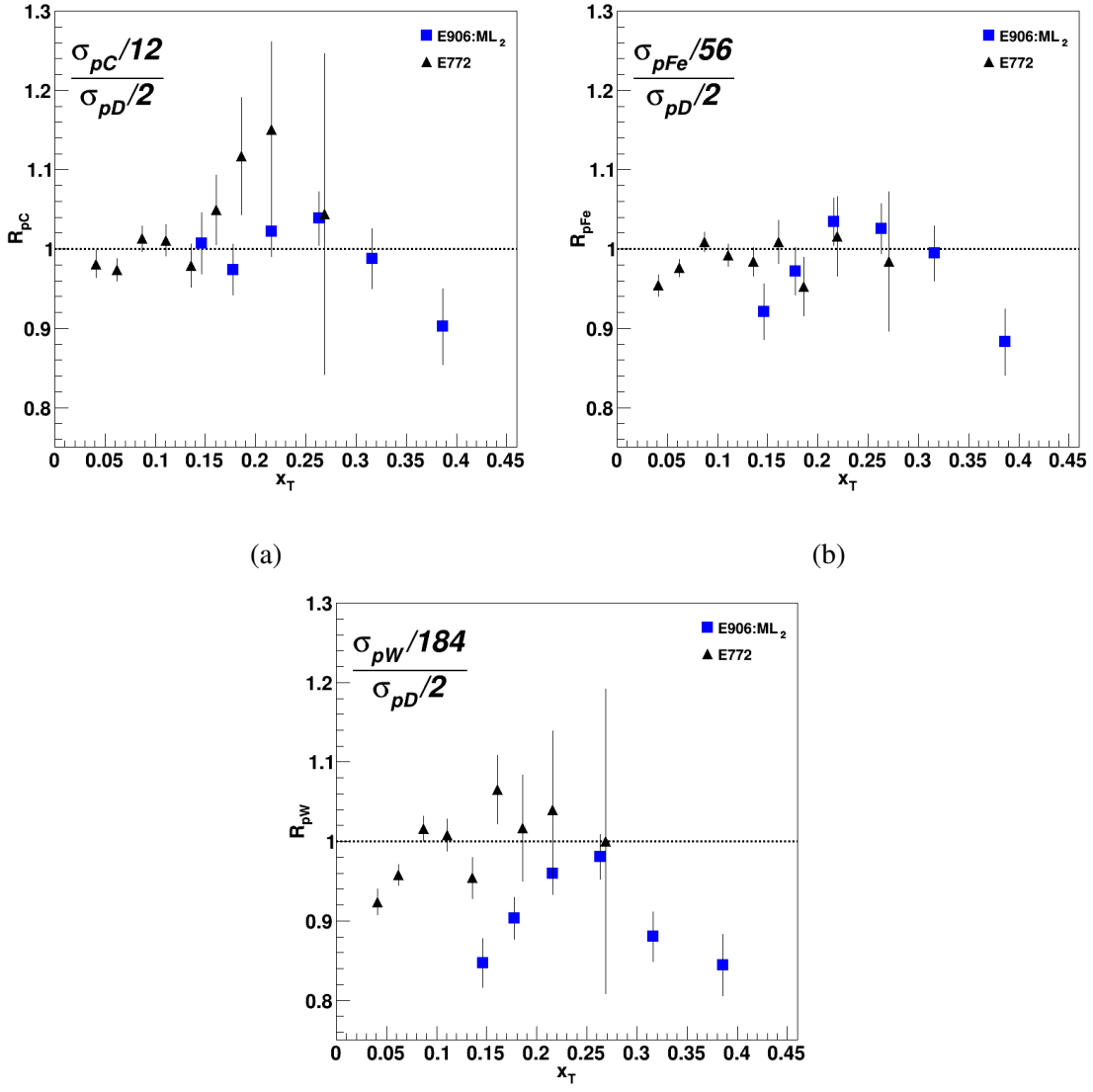


Figure 4.28: Plots of the R_{pA} using the ML_2 constraints with the E772 results from Ref. [69]: a) the R_{pC} results, b) the R_{pFe} results, and c) R_{pW} results.

A comparison of the CSRs with their statistical errors, the quadratic parameter, and the linear parameter for the R_{pD} and R_{HA} are listed in Table 4.7 and the comparison of the R_{pA} values are in Table 4.8. The subscripts in the CSR parameters refer to the extracted CSR in that x_T bin with the bin ranges defined as they were in Section 4.1.1.

Parameter	R_{pD}	R_{HC}	R_{HFe}	R_{HW}
Quadratic	-3.4 $^{+4.2}_{-4.2} \cdot 10^{-11}$	-9.6 $^{+5.6}_{-5.6} \cdot 10^{-11}$	-1.3 $^{+0.45}_{-0.45} \cdot 10^{-10}$	-1.2 $^{+0.44}_{-0.44} \cdot 10^{-10}$
Linear	-1.9 $^{+2.5}_{-2.5} \cdot 10^{-6}$	3.8 $^{+3.3}_{-3.3} \cdot 10^{-6}$	4.7 $^{+2.7}_{-2.7} \cdot 10^{-6}$	6.8 $^{+2.6}_{-2.6} \cdot 10^{-6}$
CSR_0	1.223 $^{+0.041}_{-0.041}$	1.238 $^{+0.053}_{-0.053}$	1.129 $^{+0.046}_{-0.046}$	1.040 $^{+0.042}_{-0.042}$
CSR_1	1.149 $^{+0.034}_{-0.034}$	1.111 $^{+0.044}_{-0.044}$	1.102 $^{+0.039}_{-0.039}$	1.034 $^{+0.036}_{-0.036}$
CSR_2	1.252 $^{+0.036}_{-0.036}$	1.286 $^{+0.045}_{-0.045}$	1.288 $^{+0.041}_{-0.041}$	1.202 $^{+0.038}_{-0.038}$
CSR_3	1.223 $^{+0.039}_{-0.039}$	1.267 $^{+0.049}_{-0.049}$	1.237 $^{+0.043}_{-0.043}$	1.194 $^{+0.041}_{-0.041}$
CSR_4	1.197 $^{+0.040}_{-0.040}$	1.182 $^{+0.051}_{-0.051}$	1.184 $^{+0.046}_{-0.046}$	1.068 $^{+0.042}_{-0.042}$
CSR_5	1.069 $^{+0.049}_{-0.049}$	1.006 $^{+0.061}_{-0.061}$	0.9435 $^{+0.052}_{-0.052}$	0.9006 $^{+0.051}_{-0.051}$
χ^2	47.64	59.73	84.34	58.23
χ^2/DOF	1.191	1.493	2.108	1.456

Table 4.7: Extracted CSR and fit parameters with their associated statistical errors for R_{pD}, R_{HC}, R_{HFe}, and R_{HW} using the ML_2 constraints. The number of degrees of freedom was 40.

As with the R_{pD} STD results, the ML_2 results begin to differ significantly with the E886 results in the higher x_T region. The ML_2 results are also consistent with Ehlers *et al.* in the $x_T < 0.3$ region. The nuclear R_{pA} results also dovetail with the E772 results for carbon and iron, like the STD results. The tungsten results are fairly inconsistent, with the CSRs being consistently lower than the E772's. Like with the STD carbon results, the ML_2 results are consistent with the deuterium to isoscalar CSR prediction of Ehlers *et al.* assuming a small Q^2 dependence.

The ML_2 results are somewhat consistent with an EMC effect, as the slopes from the linear fit are inconsistent with zero. The linear fit has significantly lower χ^2/DOF than the constant fit as can be seen in Tables 4.9 and 4.10. Unlike the STD results, all of the targets exhibit an effect, which is a consequence of the last x_T bin being depressed. Combining the ML_2 results and the E772 results, as in Tables 4.11 and 4.12, does not change this conclusion.

Parameter	R_{pC}	R_{pFe}	R_{pW}
Quadratic	-4.0 $^{+4.0}_{-4.0} \cdot 10^{-11}$	-7.2 $^{+3.8}_{-3.8} \cdot 10^{-11}$	-6.2 $^{+3.4}_{-3.4} \cdot 10^{-11}$
Linear	4.4 $^{+2.3}_{-2.3} \cdot 10^{-6}$	5.1 $^{+2.2}_{-2.2} \cdot 10^{-6}$	6.9 $^{+2.0}_{-2.0} \cdot 10^{-6}$
CSR_0	1.008 $^{+0.039}_{-0.039}$	0.9209 $^{+0.035}_{-0.035}$	0.8472 $^{+0.031}_{-0.031}$
CSR_1	0.9741 $^{+0.032}_{-0.032}$	0.9718 $^{+0.030}_{-0.030}$	0.9034 $^{+0.027}_{-0.027}$
CSR_2	1.022 $^{+0.032}_{-0.032}$	1.035 $^{+0.030}_{-0.030}$	0.9596 $^{+0.027}_{-0.027}$
CSR_3	1.039 $^{+0.034}_{-0.034}$	1.026 $^{+0.032}_{-0.032}$	0.9809 $^{+0.028}_{-0.028}$
CSR_4	0.9876 $^{+0.038}_{-0.038}$	0.9944 $^{+0.035}_{-0.035}$	0.8804 $^{+0.031}_{-0.031}$
CSR_5	0.9021 $^{+0.048}_{-0.048}$	0.8828 $^{+0.042}_{-0.042}$	0.8444 $^{+0.039}_{-0.039}$
χ^2	81.15	41.85	56.07
χ^2/DOF	2.029	1.046	1.402

Table 4.8: Extracted CSR and fit parameters with their associated statistical errors for R_{pC}, R_{pFe}, and R_{pW} using the ML_2 constraints. The number of degrees of freedom was 40.

The χ^2 of the ML_2 is noticeably higher than the STD for all targets. The last x_T bin is responsible for the larger χ^2, since the fit results, once that bin is removed, are more inline with the STD results, as seen in Tables 4.13 - 4.16. The iron and tungsten χ^2/DOFs favor the linear fit, while carbon's results are more consistent with the constant value when using only the ML_2 data. The carbon and iron's respective slopes are consistent with zero, while tungsten shows some EMC effect, but a significantly worse reduced χ^2. When also incorporating E772's data, the results change slightly. All targets favor the linear fit, and thus suggest some EMC effect, but the statistical error on the slopes of iron and carbon still cast some doubt as they are consistent with zero. The conclusion for the ML_2 is the same as with the STD results, more data is needed in the $x_T > 0.4$ region to elucidate if an EMC effect for DY exists.

Parameter	R_{pC}	R_{pFe}	R_{pW}
p_0	1.188	1.229	1.147
σ_0	0.08901	0.07370	0.07370
p_1	-0.6709	-0.8231	-0.7782
σ_1	0.3136	0.2829	0.2586
χ^2	1.505	1.474	2.833
χ^2/DOF	0.7527	0.7370	1.417

Table 4.9: Nuclear CSR fit results for R_{pC}, R_{pFe}, and R_{pW} using the ML_2 constraints for a linear function using the . The number of degrees of freedom was 2.

Parameter	R_{pC}	R_{pFe}	R_{pW}
p_0	1.001	0.9978	0.9295
σ_0	0.01823	0.01686	0.01515
χ^2	6.081	9.940	11.89
χ^2/DOF	2.027	3.313	3.964

Table 4.10: Nuclear CSR fit results for R_{pC}, R_{pFe}, and R_{pW} using the ML_2 constraints for a constant function. The number of degrees of freedom was 3.

Parameter	R_{pC}	R_{pFe}	R_{pW}
p_0	1.205	1.210	1.156
σ_0	0.08717	0.07532	0.07237
p_1	-0.7228	-0.7704	-0.8043
σ_1	0.3090	0.2686	0.2552
χ^2	2.447	1.860	3.302
χ^2/DOF	0.6117	0.4649	0.8254

Table 4.11: Nuclear CSR fit results for R_{pC}, R_{pFe}, and R_{pW} using the ML_2 constraints and the E772 results for a linear function. The number of degrees of freedom was 4.

Parameter	R_{pC}	R_{pFe}	R_{pW}
p_0	1.006	0.9992	0.9324
σ_0	0.01792	0.01572	0.01493
χ^2	7.919	10.09	13.23
χ^2/DOF	1.584	2.018	2.647

Table 4.12: Nuclear CSR fit results for R_{pC}, R_{pFe}, and R_{pW} using the ML_2 constraints and the E772 results for a constant function. The number of degrees of freedom was 5.

Parameter	R_{pC}	R_{pFe}	R_{pW}
p_0	1.100	1.222	1.139
σ_0	0.1289	0.1206	0.1074
p_1	-0.3153	-0.3944	-0.7491
σ_1	0.4912	0.4586	0.4086
χ^2	0.6205	0.6333	2.825
χ^2/DOF	0.6205	0.06333	2.825

Table 4.13: Nuclear CSR fit results for R_{pC}, R_{pFe}, and R_{pW} using the ML_2 constraints for a linear function. The number of degrees of freedom was 1.

Parameter	R_{pC}	R_{pFe}	R_{pW}
p_0	1.018	1.012	0.9449
σ_0	0.01973	0.01844	0.01646
χ^2	1.033	0.8027	6.185
χ^2/DOF	0.5163	0.4013	3.093

Table 4.14: Nuclear CSR fit results for R_{pC}, R_{pFe}, and R_{pW} using the ML_2 constraints for a constant function. The number of degrees of freedom was 3.

Parameter	R_{pC}	R_{pFe}	R_{pW}
p_0	1.128	1.108	1.152
σ_0	0.1260	0.1115	0.1053
p_1	-0.4099	-0.3503	-0.7901
σ_1	0.4824	0.4313	0.4022
χ^2	1.733	3101	3.300
χ^2/DOF	0.3681	0.1034	1.100

Table 4.15: Nuclear CSR fit results for R_{pC}, R_{pFe}, and R_{pW} using the ML_2 constraints and the E772 results for a linear function. The number of degrees of freedom was 3.

Parameter	R_{pC}	R_{pFe}	R_{pW}
p_0	1.023	1.019	0.9478
σ_0	0.01933	0.01698	0.01618
χ^2	2.455	0.9696	7.158
χ^2/DOF	0.6137	0.2424	1.790

Table 4.16: Nuclear CSR fit results for R_{pC}, R_{pFe}, and R_{pW} using the ML_2 constraints and the E772 results for a constant function. The number of degrees of freedom was 4.

Unlike the STD results, the last x_T bin for the R_{HA} CSRs are consistent with being below unity. This behavior is further seen in the nuclear R_{pA} results and suggests strongly that the ML_2 and the ML in general has a bias in the higher x_T region. This is probably a reflection of the yield differences between the STD and the ML_2 as seen in Section 3.4. ML_2 has higher yields and the shapes are similar to the STD, but the LH_2 and LD_2 distributions for x_T and I_T show increasingly higher differences at higher values of x_T and I_T. This probably led to depressions in the extracted CSR for the larger x_T bins. The larger reduced χ^2s also reflect that the ML_2 does not have the same quadratic I_T dependence as the STD. This dependence may be able to be corrected with an acceptance correction, finding a different I_T dependence functional form, or with using more experimental and MC data to better train the ML. An acceptance correction could only potentially correct any x_T correction issues, since I_T shares no kinematic dependence and cannot be readily reproduced by MC.

4.1.3 Comparison

Figures 4.29 and 4.30 illustrate the differences between the CSRs extracted with the STD and the ML_2 constraints. The figures show the extracted CSRs and the E886 results for the R_{pD} CSR and the E772 for the R_{pA} CSRs.

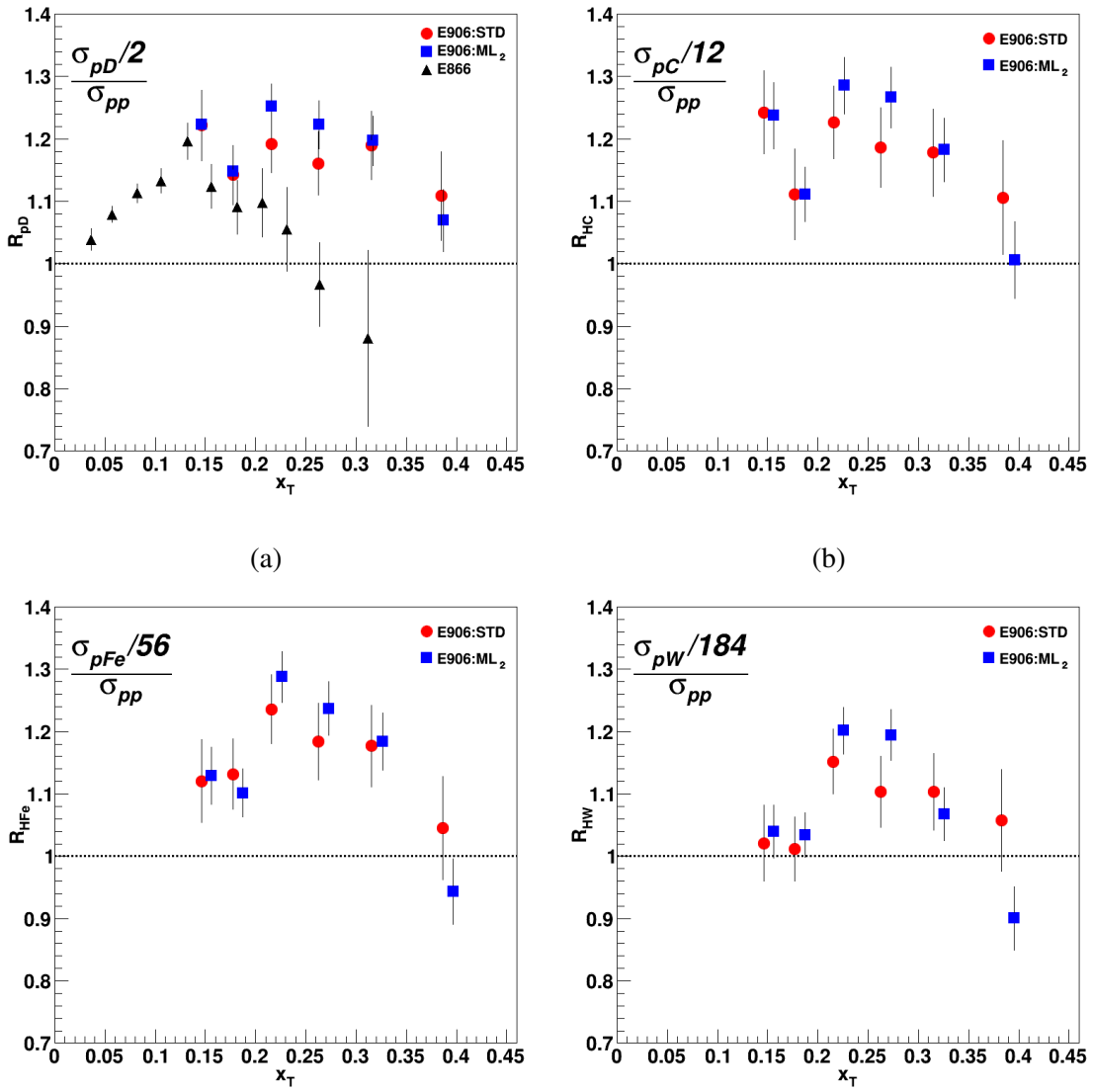
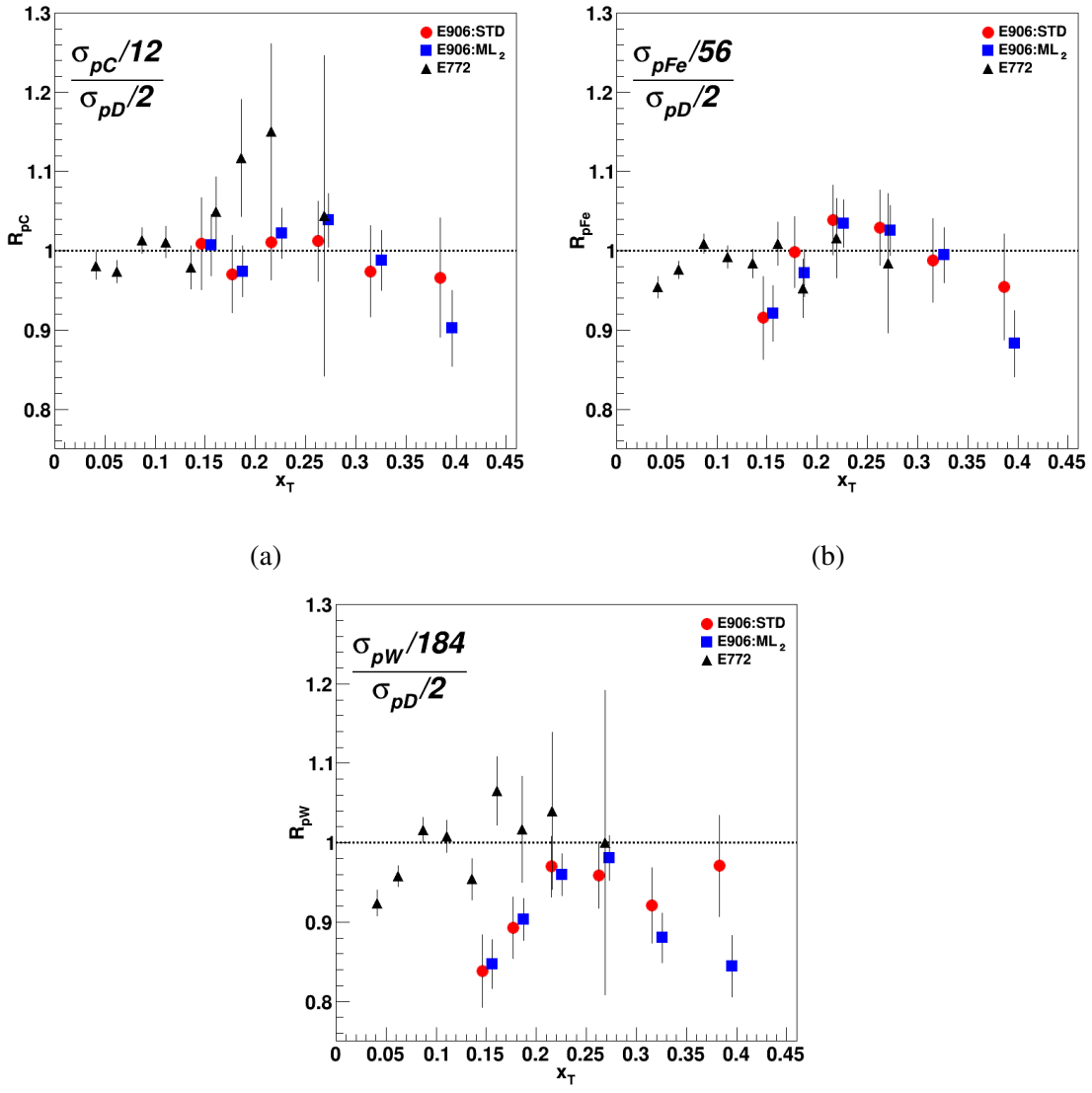


Figure 4.29: Comparison of the extracted R_{pD} CSR with the E886 results and the R_{HA} CSRs between the STD and ML_2 constraints: a) the R_{pD} results, b) the R_{HC} results, c) the R_{HFe} results, and d) the R_{HW} results.



(a)

(b)

(c)

Figure 4.30: Comparison of the extracted R_{pA} CSRs with the E772 results between the STD and ML_2 constraints: a) the R_{pC} results, b) the R_{pFe} results, and c) the R_{pW} results.

For a quantitative measure of the compatibility between the CSRs of the STD and the ML_2 cuts a modification of the CSR standard deviation difference, Eq. 3.28, was constructed

$$D_{CSR} = \frac{|CSR_{STD} - CSR_{ML_2}|}{\sigma_{comp}} \tag{4.8}$$

with $\sigma_{comp} = \sqrt{\sigma^2_{STD} + \sigma^2_{ML_2}}$. This D_{CSR} gives a measure of how compatible the extracted CSRs are, in other words how many standard deviations the STD and the ML_2 are from each other using σ_{comp} as the measure. The results are listed in Tables 4.17 and 4.18. From the tables most of the ML_2 values are within a σ_{comp} of the STD values and those that are a bit beyond are within a σ of the STD when the statistical and systematic errors are propagated in quadrature. The differences that are still significant are R_{HW} for x_T bin 5 and R_{pW} for x_T bins 3 and 5. These bins have significant systematic errors stemming from the σ_{fit} and the σ_{ML}. With that in mind, the STD and ML_2 results can be considered compatible.

Considering the χ^2 for the STD and the ML_2 results in Tables 4.1 - 4.8, it is clear that the STD results contain smaller χ^2 than the ML_2, however ML_2 has smaller statistical errors. The smaller χ^2 suggests that the I_T dependence within the STD is better modeled as the CP2 and a quadratic function in general. For the ML_2, the R_{HFe} and R_{pC} χ^2/DOFs are relatively large, being above two, while the other values are in the 1.0 - 1.5 range. The fact that the STD χ^2 results are more consistent across all of the CSR as opposed to the ML_2, which shows larger values for the R_{pA} and R_{HA} relative to the R_{pD}, further suggests that the ML_2 is biased toward the liquid targets.

In Appendix D, the yield differences between the two sets of constraints are listed. In addition to the total yields for each target, the events that pass only one set of constraints are detailed as are the events that pass both sets of constraints. From these tables it is shown that the ML_2 finds 24% more LH_2 and almost 20% more LD_2 events than the STD constraints. This is followed by the C, Fe, and W results where the ML_2 finds 19%, 14%, and 13% more events than the STD constraints. There is a large difference between the constraint sets in the EMPTY and NONE targets, where the ML_2 finds 90% and 148% more events than the STD constraints. This discrepancy could be

the source for some of the CSR differences due to background subtraction. Since the background is scaled by $POT_{Target}/POT_{Background}$ and the increase in the number of target events is less than the increase in the number of background events, more of the target events would be removed in the ML_2 than in the STD analysis. This would alter the CS before the CSR is calculated and could lead to differences in the extracted CSR.

As an ancillary or main method of analysis, ML has its merits as the R_{pD} results and the liquid target yields illustrate. The cause of the decreases in effectiveness of the ML in the nuclear targets probably has its origins in the nuclear training data and the background target data. Assuming that training and testing of the liquid targets is not fundamentally different than training and testing of the nuclear targets, more beam time dedicated to nuclear targets and a larger number of background distributions could improve the results. Moreover, the MC used to generate the nuclear DY and background MC was the CTEQ6 free proton MC using isospin symmetry for the neutron PDFs. To produce the nuclear CS and event distributions the protons were weighted by Z and the neutrons were weighted by $A - Z$,

$$\sigma_A = a\,(Z)\,\sigma_{pp} + b\,(A - Z)\,\sigma_{pn}, \tag{4.9}$$

with $a\,(Z)$ and $b\,(A - Z)$ being some scaling function directly related to the Z and $A - Z$ of the target, and the CS using the free proton PDFs. If these PDFs were replaced with a nPDF set, such as EPPS16, then the improved MC could lead to better discrimination between the DY and the background.

Another improvement would be to also train on the EMPTY, NONE, and MC based upon these background targets. To improve the combination of the different ML sets, a secondary ML could have been run instead of the minimization functions, which would allow more freedom to remove events. Lastly, an x_T acceptance correction could have been made instead of relying solely on the extrapolation method.

x_T bin	R_{pD}	R_{HC}	R_{HFe}	R_{HW}
$[0.130 - 0.160)$	0.024	0.055	0.102	0.249
$[0.160 - 0.195)$	0.109	0.005	0.437	0.357
$[0.195 - 0.240)$	1.024	0.798	0.754	0.774
$[0.240 - 0.290)$	0.968	1.002	0.709	1.298
$[0.290 - 0.350)$	0.109	0.051	0.091	0.482
$[0.350 - 0.450)$	0.453	0.909	1.041	1.627

Table 4.17: CSR standard deviation difference between the STD and ML_2 constraints for R_{pD}, R_{HC}, R_{HFe}, and R_{HW}.

x_T bin	R_{pC}	R_{pFe}	R_{pW}
$[0.130 - 0.160)$	0.018	0.090	0.164
$[0.160 - 0.195)$	0.060	0.494	0.223
$[0.195 - 0.240)$	0.189	0.079	0.220
$[0.240 - 0.290)$	0.433	0.068	0.431
$[0.290 - 0.350)$	0.198	0.106	0.716
$[0.350 - 0.450)$	0.718	0.907	1.692

Table 4.18: CSR standard deviation difference between the STD and ML_2 constraints for R_{pC}, R_{pFe}, and R_{pW}.

4.2 Asymmetry Ratio Results

For the extraction of the asymmetry ratio, the nCTEQ15 with pion data and the EPPS16 nPDFs are used, and their results are compared. These nPDF sets differ in parameterization of the PDFs, the data selection, kinematic cuts, and initial scale [135]. nCTEQ15 follows a modified version of CTEQ's free proton PDF parameterization

$$xq_i^{p/A}(x, Q_0) = c_0 x^{c1} (1-x)^{c2} e^{c_3 x} \left(1 + e^{c4} x\right)^{c5} \tag{4.10}$$

and

$$\frac{\bar{d}\left(x,Q_0\right)}{\bar{u}\left(x,Q_0\right)} = c_0 x^{c1}\left(1-x\right)^{c2} + \left(1+c_3 x\right)\left(1-x\right)^{c4}, \tag{4.11}$$

with i standing for

$$i = u_v, d_v, g, \bar{u}+\bar{d}, s+\bar{s}, s-\bar{s}. \tag{4.12}$$

$q^{p/A}$ denotes the nPDF of the parton or partonic function found in the bound proton within a nucleus of mass number A. The subscript v denotes the valence nPDFs. The neutron nPDFs are found by isospin symmetry within the valence and sea, meaning that the d, u, $\bar{d}$, and $\bar{u}$ distributions in the neutron are taken to be the u, d, $\bar{u}$, and $\bar{d}$ distributions in the proton, respectively. From there, each c parameter is then expanded to include nuclear dependence,

$$c_j \rightarrow c_j\left(A\right) \equiv c_{j,0} + c_{j,1}\left(1 - A^{-c_{j,2}}\right), \tag{4.13}$$

with j running from 1 to 5. This nuclear dependence is structured to reduce to the free proton PDFs in the $A = 1$ limit [23]. Though the name would suggest that nCTEQ15 would reduce to CTEQ14, it is actually closer to CTEQ6.

EPPS16 takes a different approach and develops nuclear dependence on top of the CTEQ14 free proton PDFs with the modified PDF being

$$q_i^{p/A}\left(x,Q^2\right) \equiv R_i^A\left(x,Q^2\right) q_i^p\left(x,Q^2\right). \tag{4.14}$$

The nuclear dependence is scaled using carbon as a baseline, with

$$\begin{aligned} R_i^{12}\left(x\right) &= \begin{cases} a_0 + a_1\left(x-x_a\right)^2 & x \le x_a \\ b_0 + b_1 x^{\alpha} + b_2 x^{3\alpha} & x_a \le x \le x_e \\ c_0 + \left(c_1 - c_2 x\right)\left(1-x\right)^{-\beta} & x_e \le x \le 1 \end{cases} \\ R_i^A\left(x\right) &= R_i^{12}\left(x\right)\left(\frac{A}{12}\right)^{\gamma_i\left[R_i^{12}-1\right]}, \end{aligned} \tag{4.15}$$

where the i and A dependencies of the parameters are implied; x_a is the antishadowing maximum, x_e is the EMC minimum, and $\alpha = 10x_a$ [24]. As with the nCTEQ15 set, the neutron nPDF is found by isospin symmetry.

For relevance to the $\bar{d}/\bar{u}$ extraction, it should be noted that nCTEQ15 has the ratio as an explicit parameter, while EPPS16 allows $\bar{d}$ and $\bar{u}$ to float independently. Due to this and data differences, there is a difference in the $\bar{d}$ and $\bar{u}$ between them, but it is not significant in the region of SeaQuest's acceptance, i.e. $x_T \in [0.1 - 0.45)$, as can be seen in the Fig. 4.31. It is instructive to note that the uncertainty bands of the nCTEQ15 are significantly smaller than those of the EPPS16. These smaller uncertainty bands are due to the more restrictive constrains on the sea nPDFs in nCTEQ15. EPPS16 has 9 parameters to fit sea nPDFs, while nCTEQ15 has 2. nCTEQ15 constrains only $\bar{d}+\bar{u}$, while EPPS16 allows each each sea quark nPDF to vary [24]. Due to the propagation of the error of ratio distributions, this error band difference carries over to the error bands for the $\bar{d}/\bar{u}$ ratio.

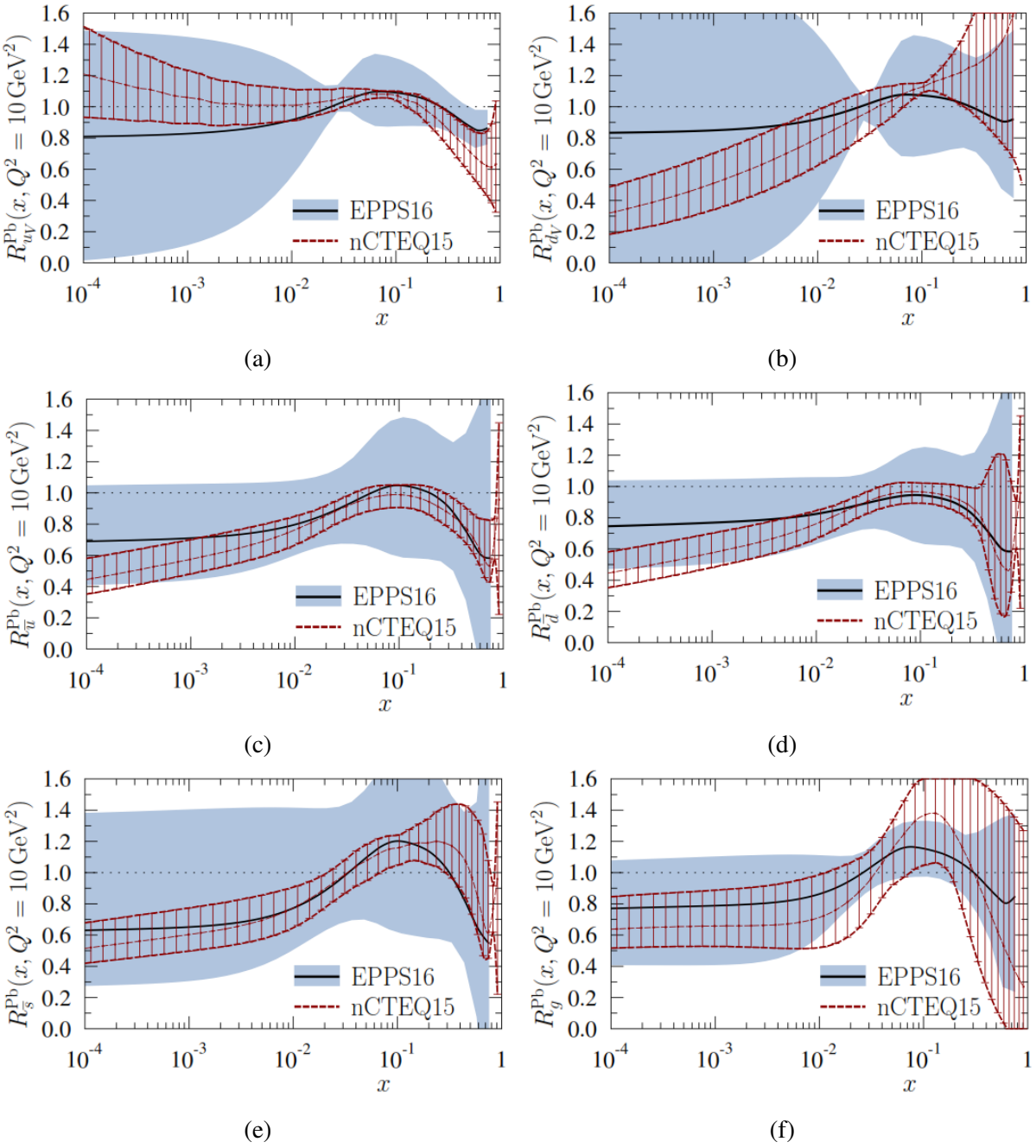


Figure 4.31: Comparison of the ^{207}Pb nPDF modifications, i.e. the ratio of the nPDF to the free proton PDF, at $Q^2 = 10\ GeV^2$ between EPPS16, denoted as black with blue uncertainties, and nCTEQ15, represented as red [24]: a) the u_v results, b) the d_v results, c) the $\bar{u}$ results, d) the $\bar{d}$ results, e) the $\bar{s}$ results, and f) the gluon results. The subscript "v" denotes the valence distributions and the s and $\bar{s}$ distributions are assumed to be the same.

4.2.1 Extraction Methodology

As shown in Eq. 1.17, the full DY cross section is a convolution of the partonic hard scattering cross section and the PDFs. This continues for the nuclear cross section to

$$d\sigma_{AB\to l\bar{l}X} = \sum_{ij} q_i^{A_1 Z_1} \otimes q_j^{A_2 Z_2} \otimes d\hat{\sigma}^{ij\to l\bar{l}X}, \tag{4.16}$$

which for this analysis on the nuclear PDFs becomes

$$d\sigma_{AB\to l\bar{l}X} = \sum_{ij} Z * q_i^p \otimes q_j^{p/A} \otimes d\hat{\sigma}_p^{ij\to l\bar{l}X} + (A-Z) * q_i^p \otimes q_j^{n/A} \otimes d\hat{\sigma}_n^{ij\to l\bar{l}X}. \tag{4.17}$$

This is essentially the sum of the Z weighted proton-proton cross section and the $(A-Z)$ proton-neutron cross section. The nuclear dependence is factored into the nuclear PDFs for both the nCTEQ15 and EPPS16 sets. For the cross section calculations an $A = 1$ nCTEQ15 nPDF was used for the nCTEQ15 calculation and the CTEQ14 PDF was used for the EPPS16 calculation.

The extraction routine is a modification of Paul E. Reimer's with the specifics of the routine outlined in Refs. [136, 137]. The routine takes in the experimental events used to derive the CSRs along with the CSRs, statistical errors, and systematic errors. Using the PDFs from the PDF sets, a modified $\bar{d}$ and $\bar{u}$ are generated as

$$\bar{d} \to \frac{\left(\bar{d}+\bar{u}\right) * \bar{r}}{1.0+\bar{r}} \tag{4.18}$$

and

$$\bar{u} \to \frac{\bar{d}+\bar{u}}{1.0+\bar{r}}, \tag{4.19}$$

with $\bar{r}$ being the $\bar{d}/\bar{u}$ ratio derived by a fitting function using the current event's x_T and the previous values of $\bar{r}$. The PDFs are then fed into a NLO DY cross section program developed by CTEQ. To get the CSR, the proton-neutron CS is calculated using isospin symmetry and the CSR

is constructed as

$$\frac{\sigma_{pA}/A}{\sigma_{pp}} = \frac{Z * \hat{\sigma}_{pp} + (A - Z) * \hat{\sigma}_{pn}}{A * \sigma_{pp}}, \tag{4.20}$$

with the $\hat{\sigma}$ denoting the fact that the proton or neutron is bound and derived using the nPDF. In the calculation of the asymmetry ratio in hydrogen, the proton and the neutron are both treated as free. A χ^2 minimization between the set of CSRs derived by the program and the CSR from the data, given as

$$\chi^2 = \sum_{i}^{n_{bins}} \left(\frac{CSR_{calc,i} - CSR_{mea,i}}{\sigma_{mea,i}} \right)^2, \tag{4.21}$$

with CSR_{calc} being the CSR calculated by the program, CSR_{mea} being the measured CSR from the data, and σ_{mea} being the CSR statistical error, is run by changing the asymmetry ratio until $\chi^2 = 2.5 \times 10^{-7}$ or the number of iterations exceeds 200 [137]. For this analysis, a weighted set of data was used instead of using all of the events. Effectively, a 2D histogram in x_T and x_B with 20 bins a side was generated. The mean x_T and x_B in each bin was found and the number of events within the bin was recorded. Each bin then became a weighted event using the mean values as the x_T and x_B values and the number of events as the weight. When comparing the $\bar{d}/\bar{u}$ ratio for hydrogen using the STD cuts to the results calculated by the collaboration using all of the events, the differences are less than 3% or $0.27 * \sigma$, illustrating that using weighted events does not change the values significantly.

The statistical errors are found by increasing(decreasing) the data CSR by the $\pm$ statistical error and rerunning the minimization. The absolute difference between the $\bar{d}/\bar{u}$ calculated from the CSR central value and the CSR with statistical error is taken to be the error. The same process occurs for the CSR systematic errors.

The free proton asymmetry was first calculated using the CTEQ14 or the nCTEQ15 PDF set. Then the modified $\bar{d}/\bar{u}$ values were used as the input for the free proton PDFs using Eqs. 4.18 and 4.19 for the nuclear asymmetry calculations. The asymmetry derived from the CTEQ14 was used for the EPPS16 set and the values derived from the nCTEQ15 was used for the nCTEQ15 set. All of the PDF and nPDF sets were derived to the NLO level.

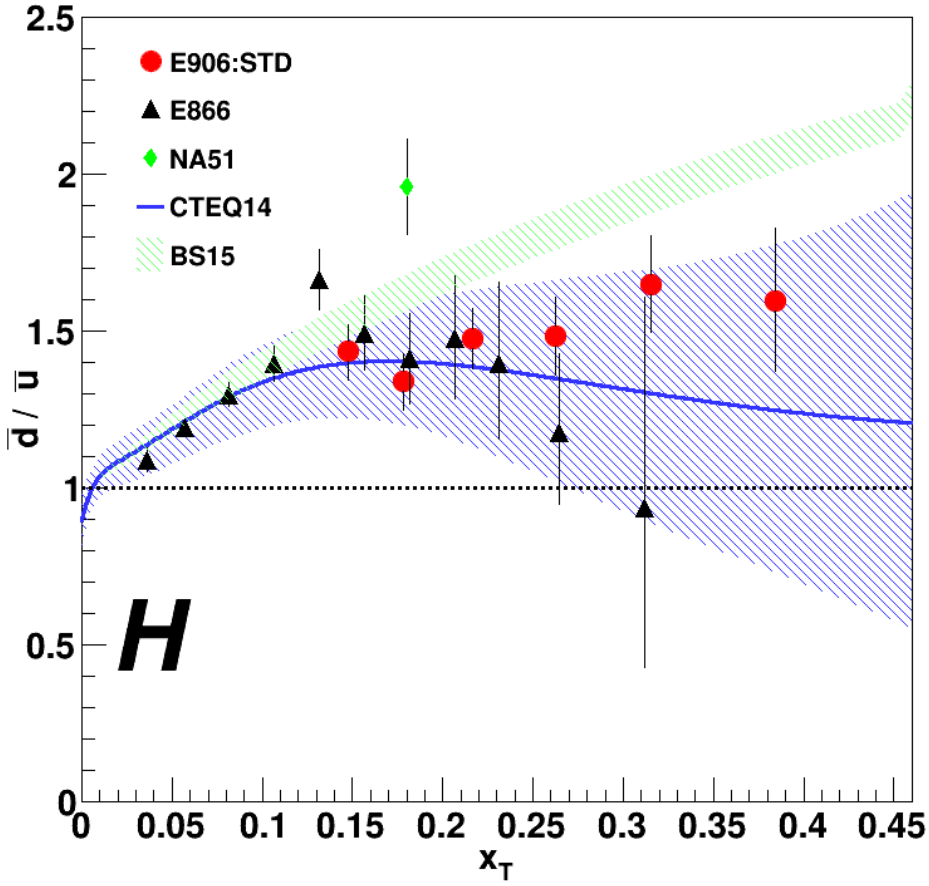


Figure 4.32: STD $\bar{d}/\bar{u}$ ratio for the free proton extracted with the CTEQ14 PDF set accompanied by the CTEQ14 prediction and error band, the theory band from Bourrely and Soffer, and the E886 and NA51 results.

4.2.2 Free Proton Results

The asymmetry results using the CTEQ14 PDF and the nCTEQ15 PDF with the STD constraints are plotted in Figs. 4.32 and 4.33 accompanied by the E866 and NA51 results from Refs. [40] and [39], respectively. The theory prediction of Bourrely and Soffer (BS15), Ref. [138], is also plotted. The tables of the values with errors and the covariance matrix for the extraction from the CTEQ14 and the nCTEQ15 PDF sets are listed in Appendix C.

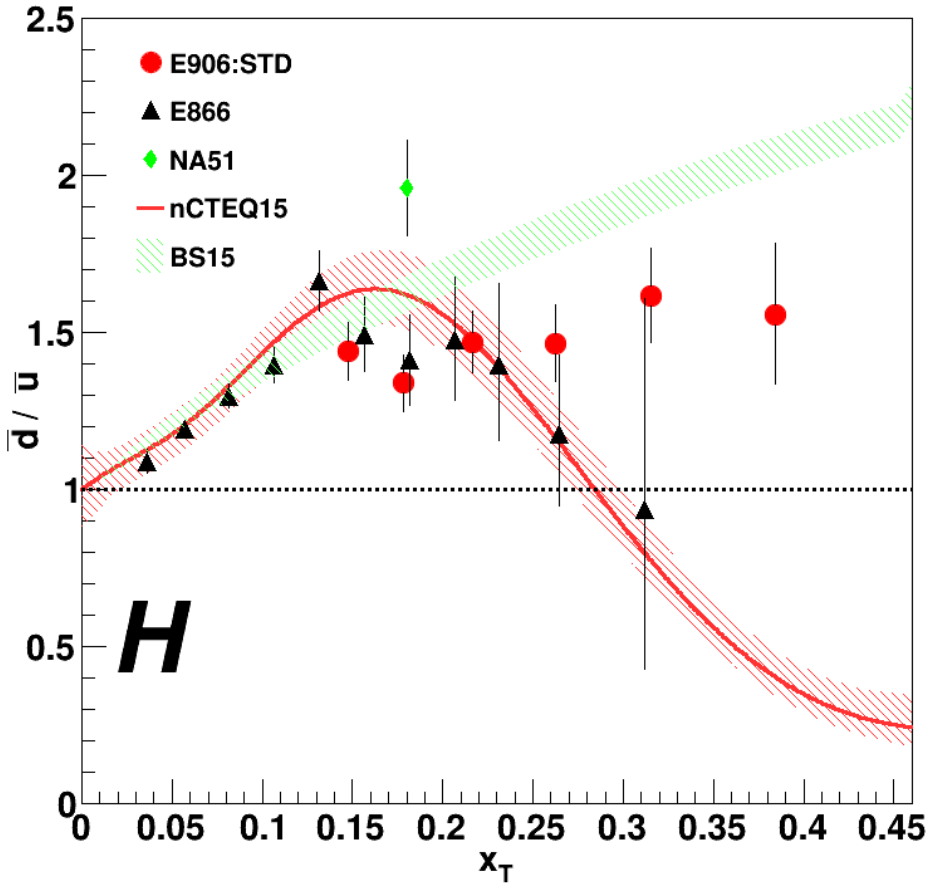


Figure 4.33: STD $\bar{d}/\bar{u}$ ratio for the free proton extracted with the nCTEQ15 PDF set accompanied by the nCTEQ15 prediction and error band, the theory band from Bourrely and Soffer, and the E886 and NA51 results.

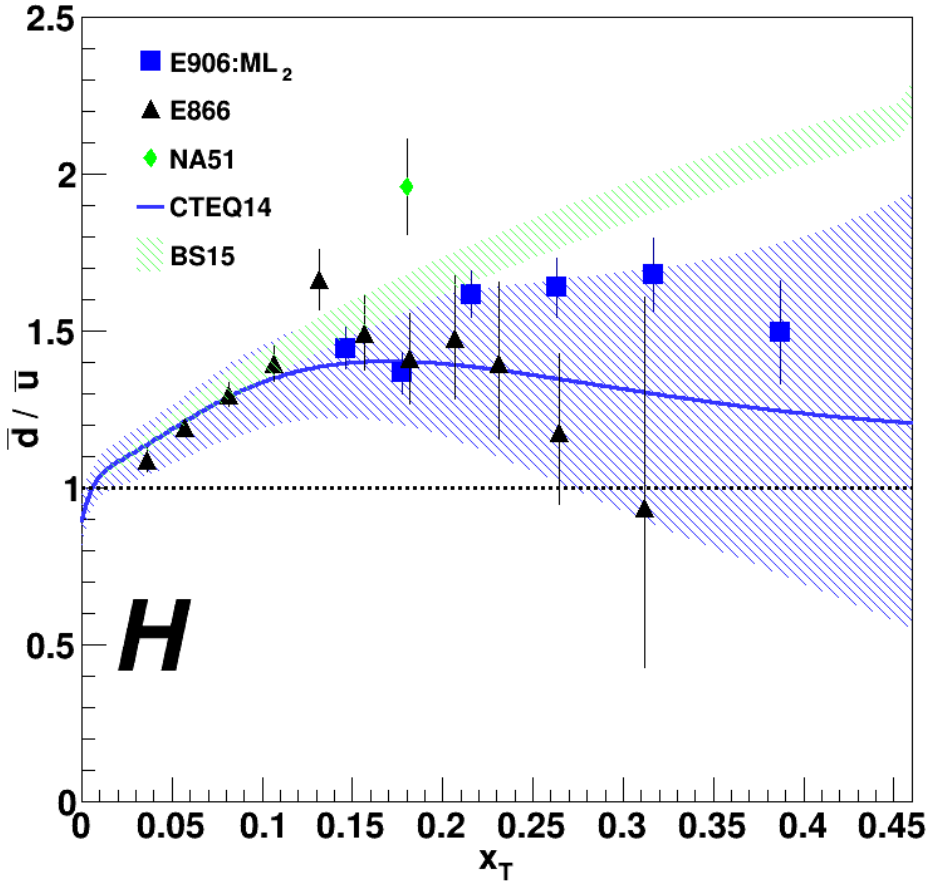


Figure 4.34: ML_2 $\bar{d}/\bar{u}$ ratio for the free proton extracted with the CTEQ14 PDF set accompanied by the CTEQ14 prediction and error band, the theory band from Bourrely and Soffer, and the E886 and NA51 results.

The asymmetry results using the CTEQ14 PDF and the nCTEQ15 PDF with the ML_2 constraints are plotted in Figs. 4.34 and 4.35 accompanied by the E866 and NA51 results with the theory band from Bourrely and Soffer. The tables containing the values, errors, and covariance matrices are listed in Appendix C.

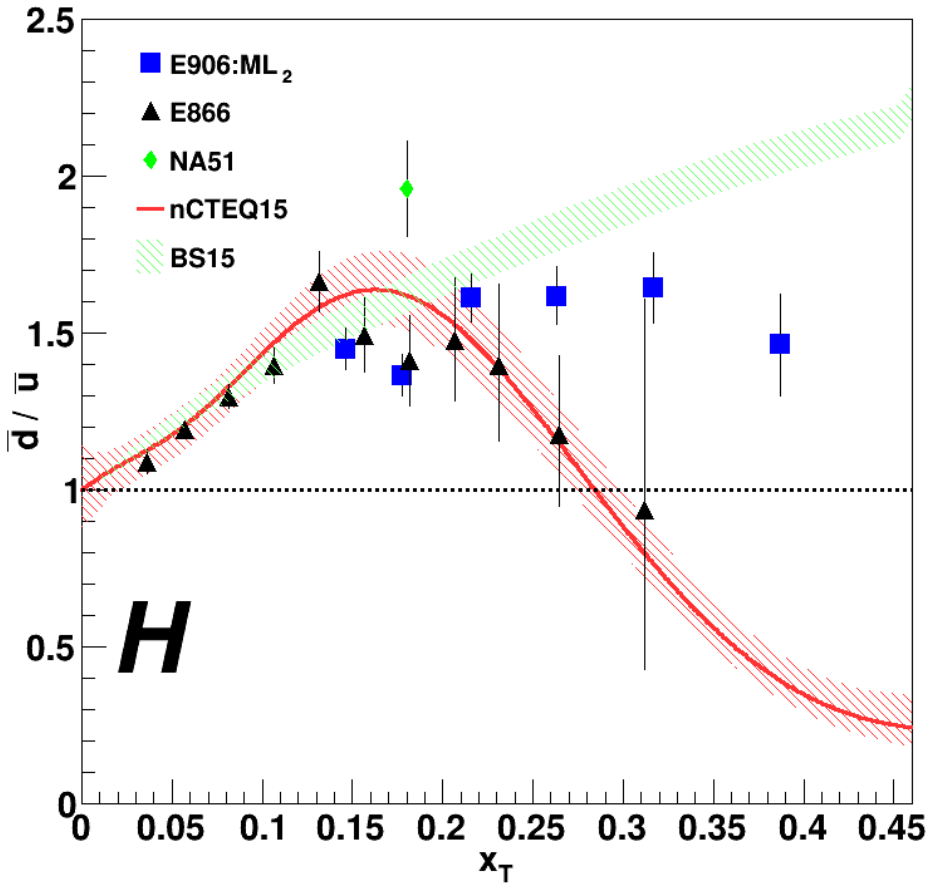


Figure 4.35: ML_2 $\bar{d}/\bar{u}$ ratio for the free proton extracted with the nCTEQ15 PDF set accompanied by the nCTEQ15 prediction and error band, the theory band from Bourrely and Soffer, and the E886 and NA51 results.

From Figs. 4.32 - 4.35, the STD and ML_2 results lie below the BS15 theory band. The ML_2 is more consistent with the prediction since the asymmetry ratio results are higher for the mid x_T range. The EPPS16 prediction and error band is more consistent with the STD and ML_2 results than the nCTEQ15 fit. There seems to be strong bias in the nCTEQ15 prediction due to the E886 results, which causes the prediction to essentially follow those points and predict a ratio less than unity. This bias is seen in all of the nCTEQ15 results.

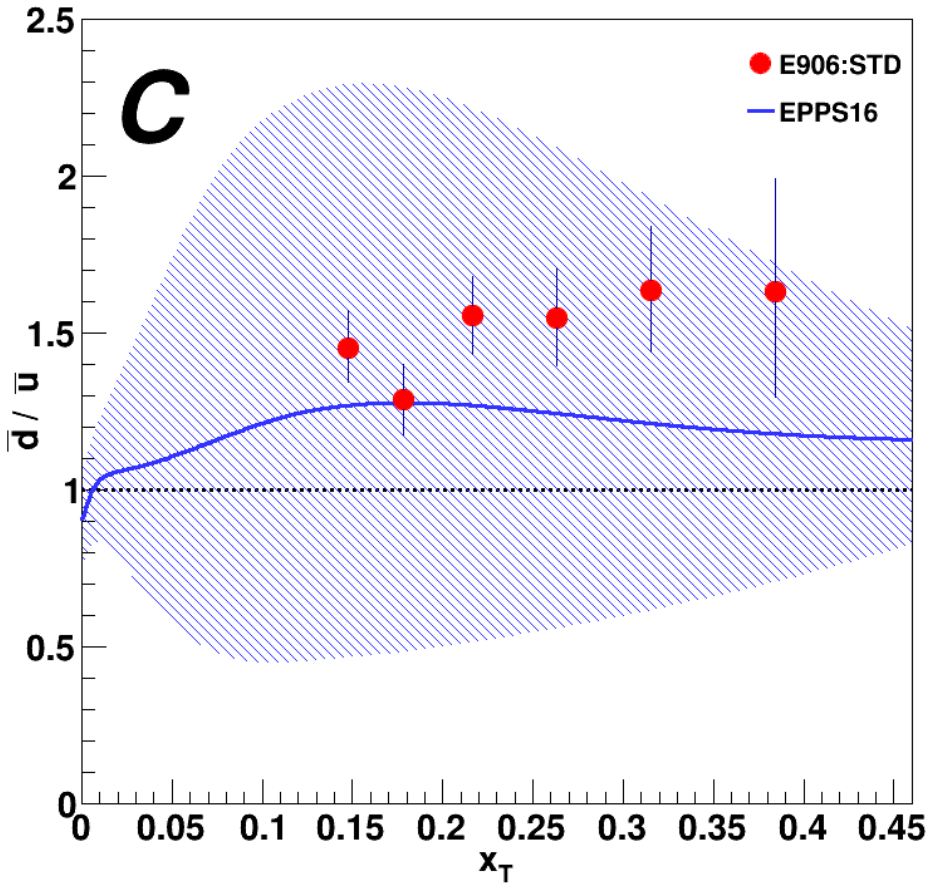


Figure 4.36: STD $\bar{d}/\bar{u}$ ratio for carbon extracted with the EPPS16 nPDF set accompanied by the EPPS16 prediction and error band.

4.2.3 Carbon Results

The asymmetry results using the EPPS16 and nCTEQ15 nPDF sets with the STD constraints are plotted in Figs. 4.36 and 4.37. The tables with the results and errors are detailed in Appendix C.

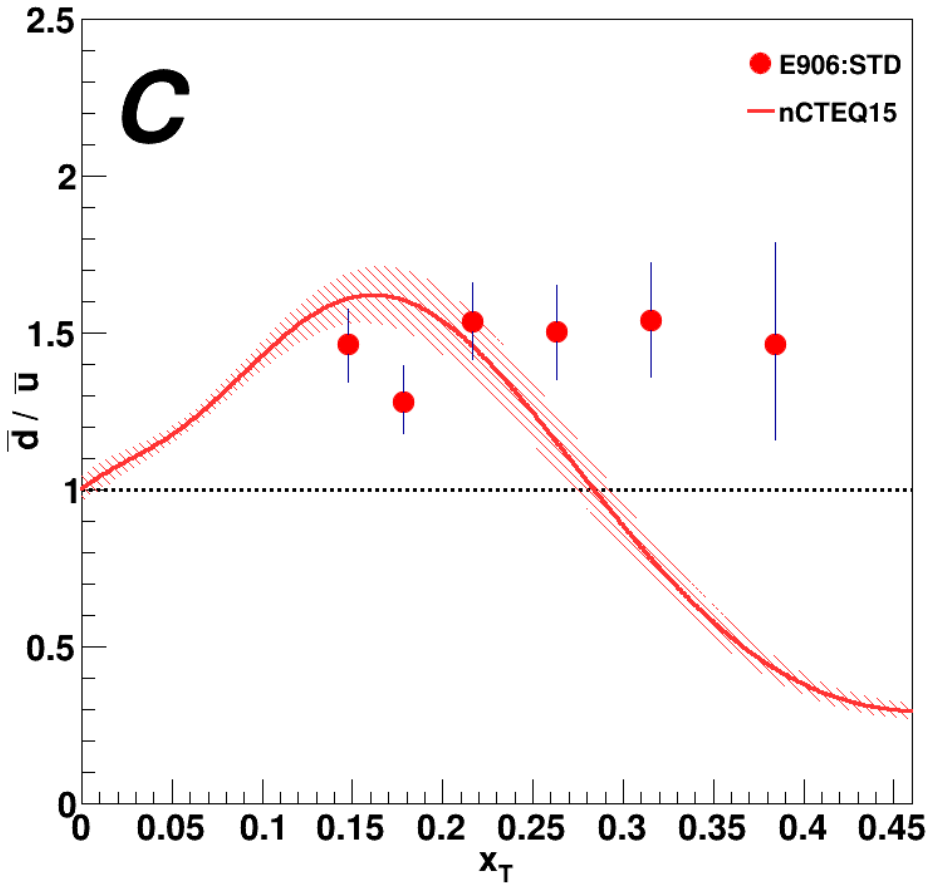


Figure 4.37: STD $\bar{d}/\bar{u}$ ratio for carbon extracted with the nCTEQ15 nPDF set accompanied by the nCTEQ15 prediction and error band.

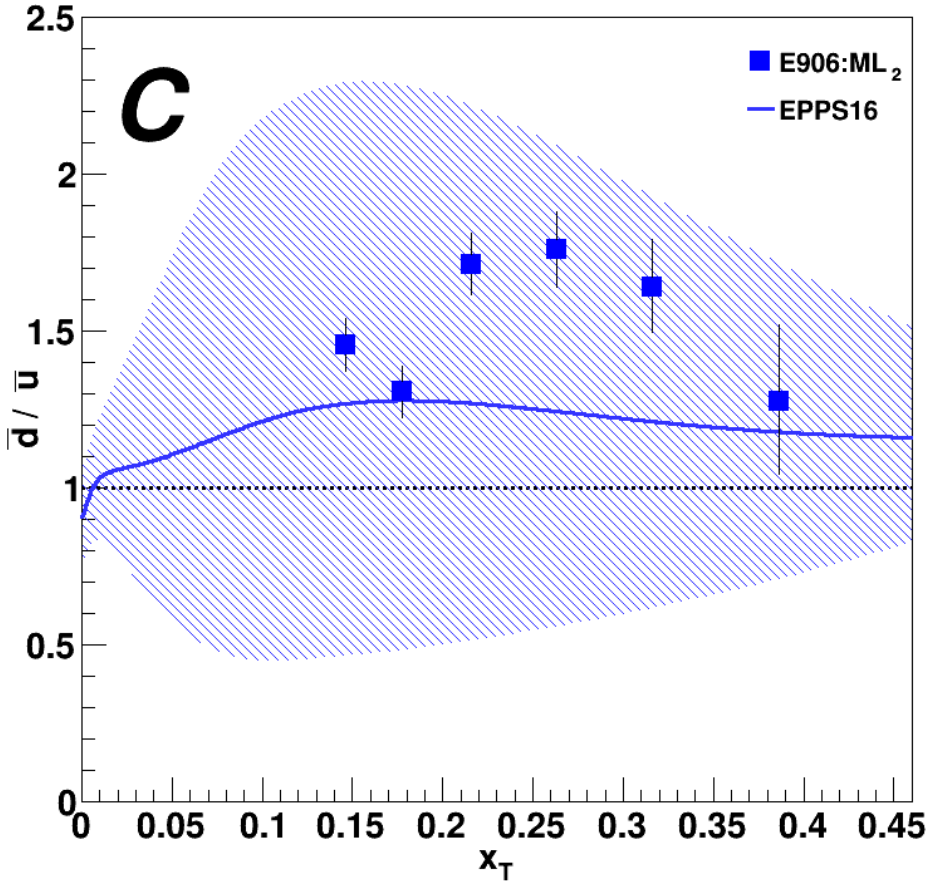


Figure 4.38: ML_2 $\bar{d}/\bar{u}$ ratio for carbon extracted with the EPPS16 nPDF set accompanied by the EPPS16 prediction and error band.

The asymmetry results using the EPPS16 and nCTEQ15 nPDF sets with the STD constraints are plotted in Figs. 4.38 and 4.39. The tables of the extracted values with the statistical and systematic errors, and the covariance matrices are listed in Appendix C.

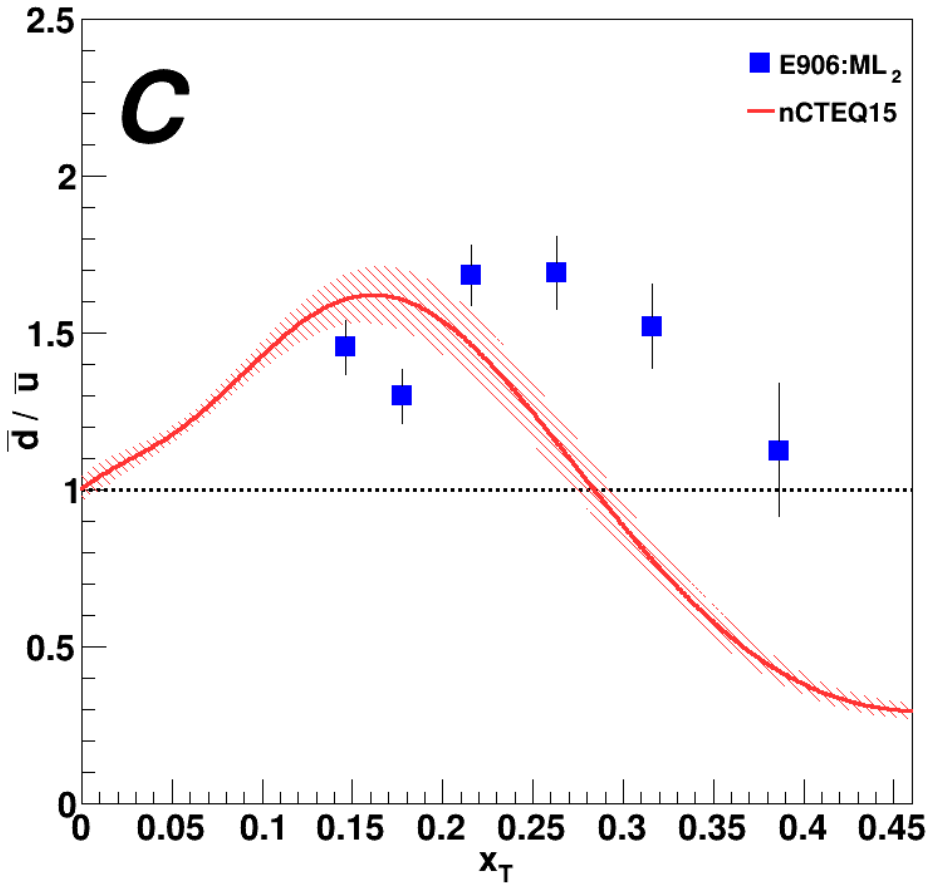


Figure 4.39: ML_2 $\bar{d}/\bar{u}$ ratio for carbon extracted with the nCTEQ15 nPDF set accompanied by the nCTEQ15 prediction and error band.

Like the free proton results, the EPPS16 prediction is more consistent with the STD and the ML_2 results. The asymmetry ratio for carbon is more or less stable for the STD results. The ML_2 shows some stability except for the last x_T bin, which is consistent with the lower R_{HA} CSRs discussed in Section 4.1.3.

4.2.4 Iron Results

The asymmetry results using the EPPS16 and nCTEQ15 nPDF sets with the STD constraints are plotted in Figs. 4.40 and 4.41. The covariance matrices and tables of the extracted values are listed in Appendix C.

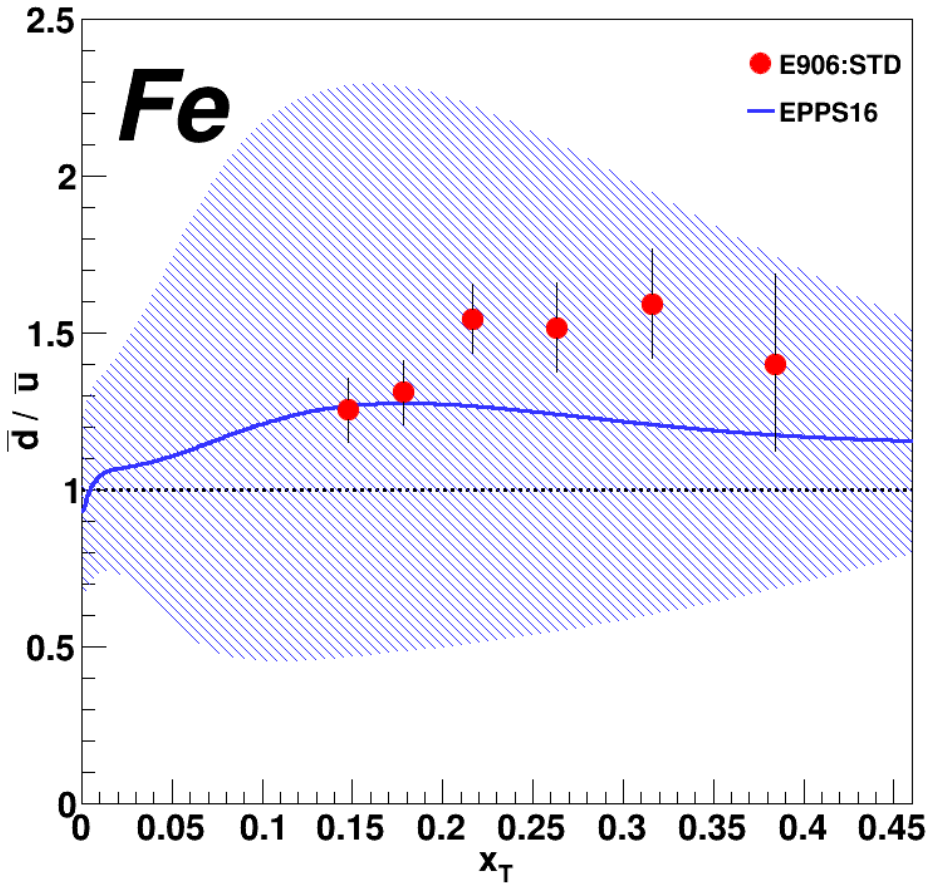


Figure 4.40: STD $\bar{d}/\bar{u}$ ratio for iron extracted with the EPPS16 nPDF set accompanied by the EPPS16 prediction and error band.

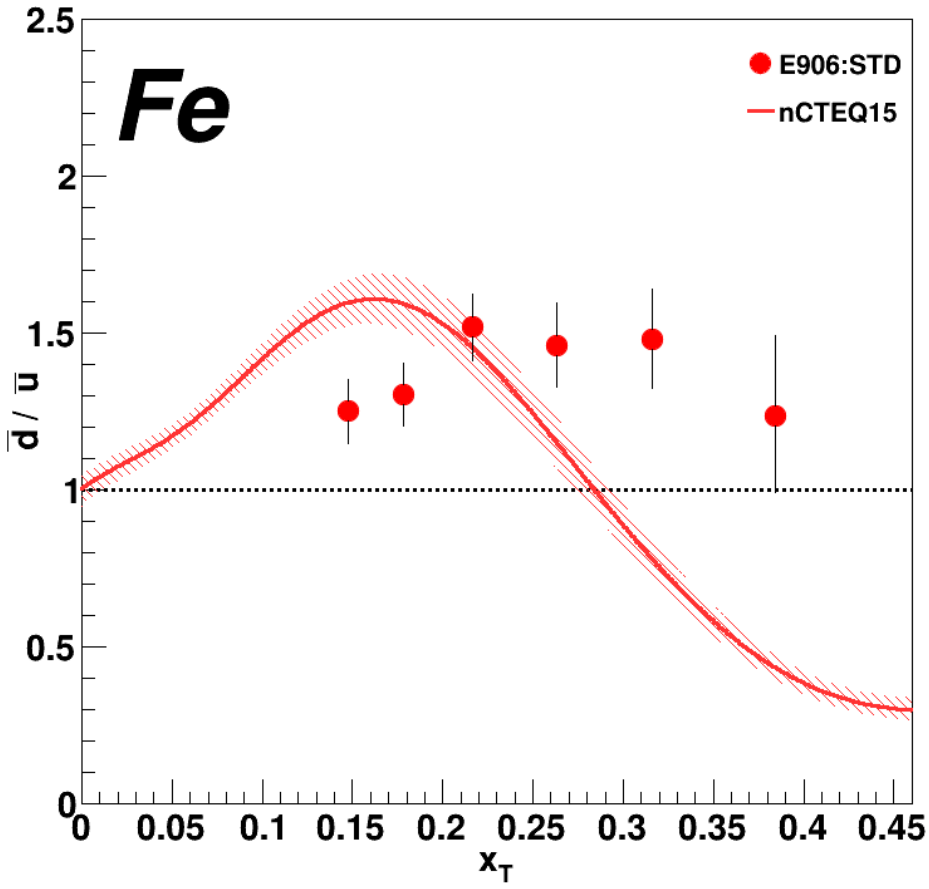


Figure 4.41: STD $\bar{d}/\bar{u}$ ratio for iron extracted with the nCTEQ15 nPDF set accompanied by the nCTEQ15 prediction and error band.

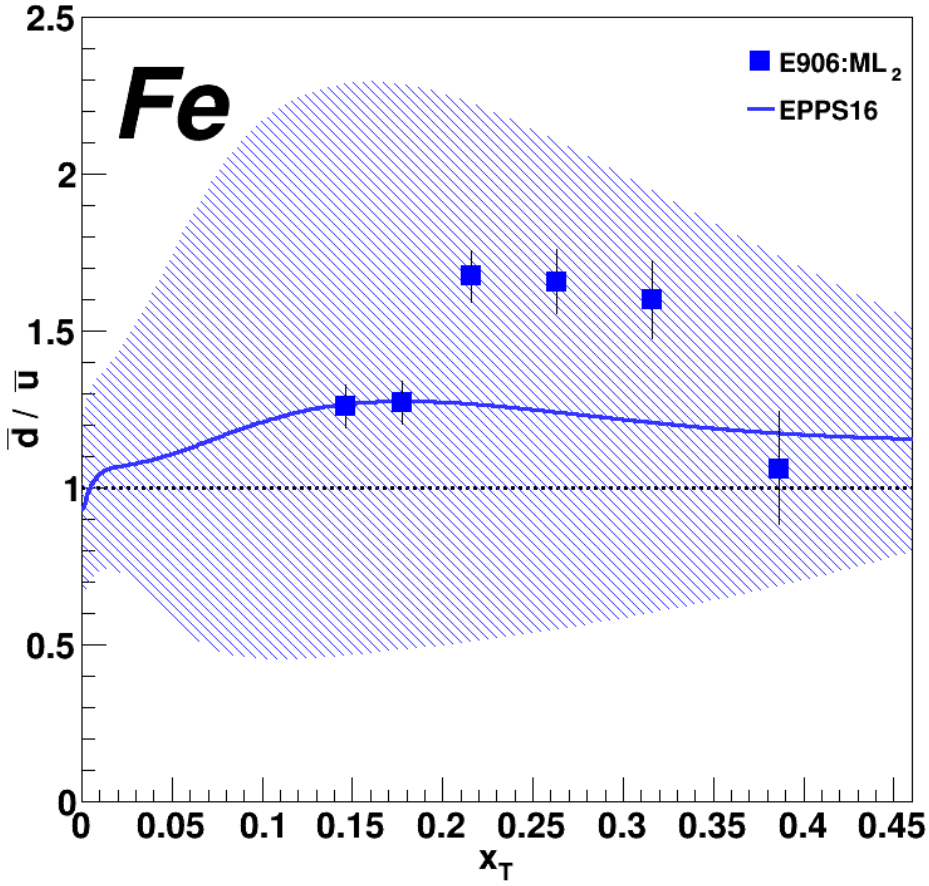


Figure 4.42: ML_2 $\bar{d}/\bar{u}$ ratio for iron extracted with the EPPS16 nPDF set accompanied by the EPPS16 prediction and error band.

The asymmetry results using the EPPS16 and nCTEQ15 nPDF sets with the ML_2 constraints are plotted in Figs. 4.42 and 4.43. Appendix C contains the tables of extracted values, statistical and systematic errors, and the covariance matrices.

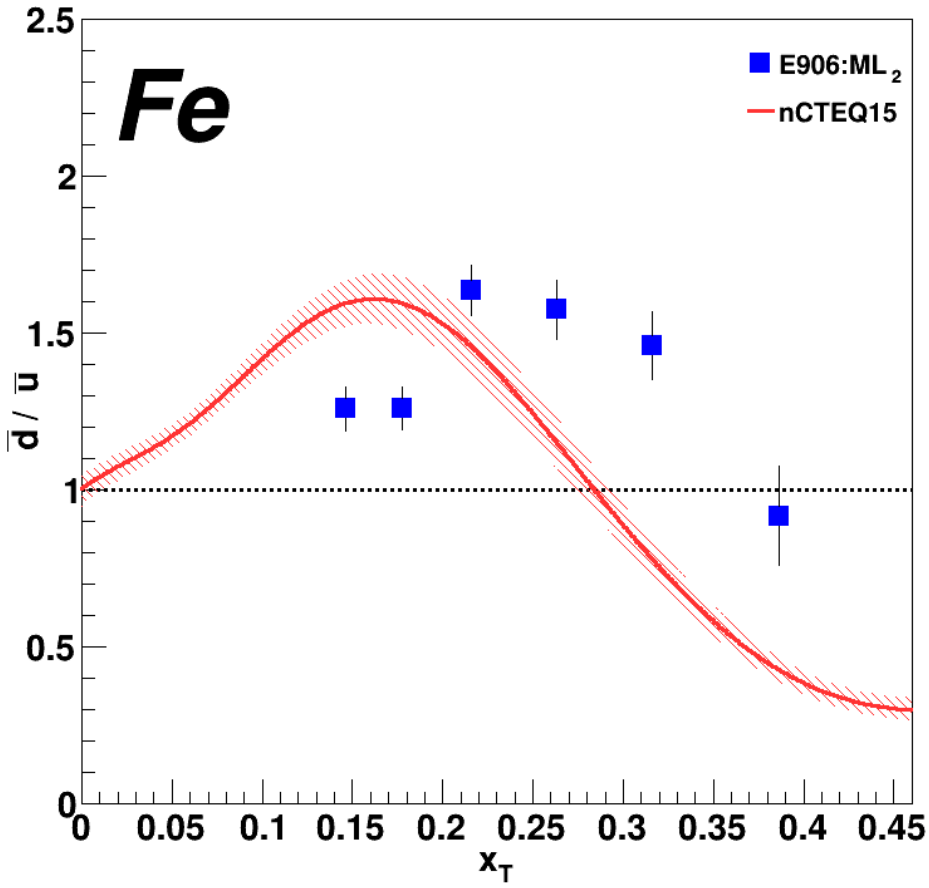


Figure 4.43: ML_2 $\bar{d}/\bar{u}$ ratio for iron extracted with the nCTEQ15 nPDF set accompanied by the nCTEQ15 prediction and error band.

The iron results conform to the EPPS16 predictions fairly well as opposed to the nCTEQ15 prediction. The ratios above $x_T = 0.2$ are stable for the STD and the ML_2 with the exception of the last x_T bin for the ML_2. There is a notable shift in the ratio after $x_T = 0.2$ that is more prominent in the ML_2 than the STD. This also seen in the tungsten results. This is also the transition to the EMC region, suggesting that there may be a connection.

4.2.5 Tungsten Results

The asymmetry results using the EPPS16 and nCTEQ15 nPDF sets with the STD constraints are plotted in Figs. 4.44 and 4.45. The tables of the asymmetry ratio values, systematic and statistical

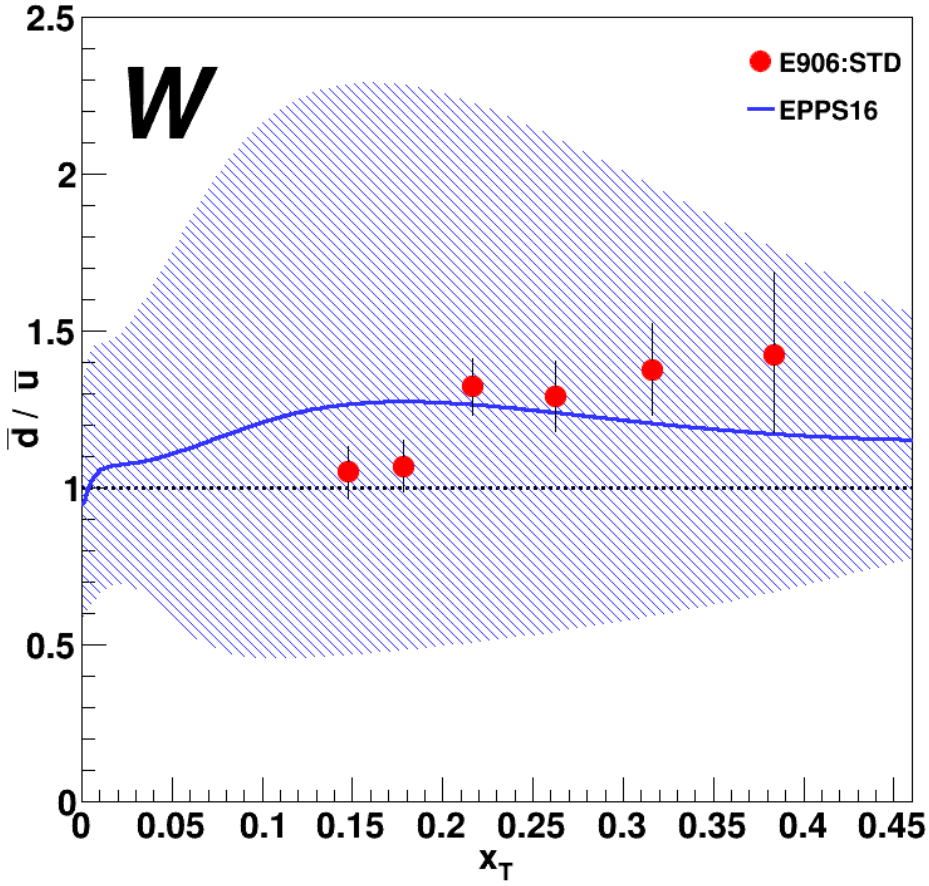


Figure 4.44: STD $\bar{d}/\bar{u}$ ratio for tungsten extracted with the EPPS16 nPDF set accompanied by the EPPS16 prediction and error band.

errors, and the covariance matrices are listed in Appendix C.

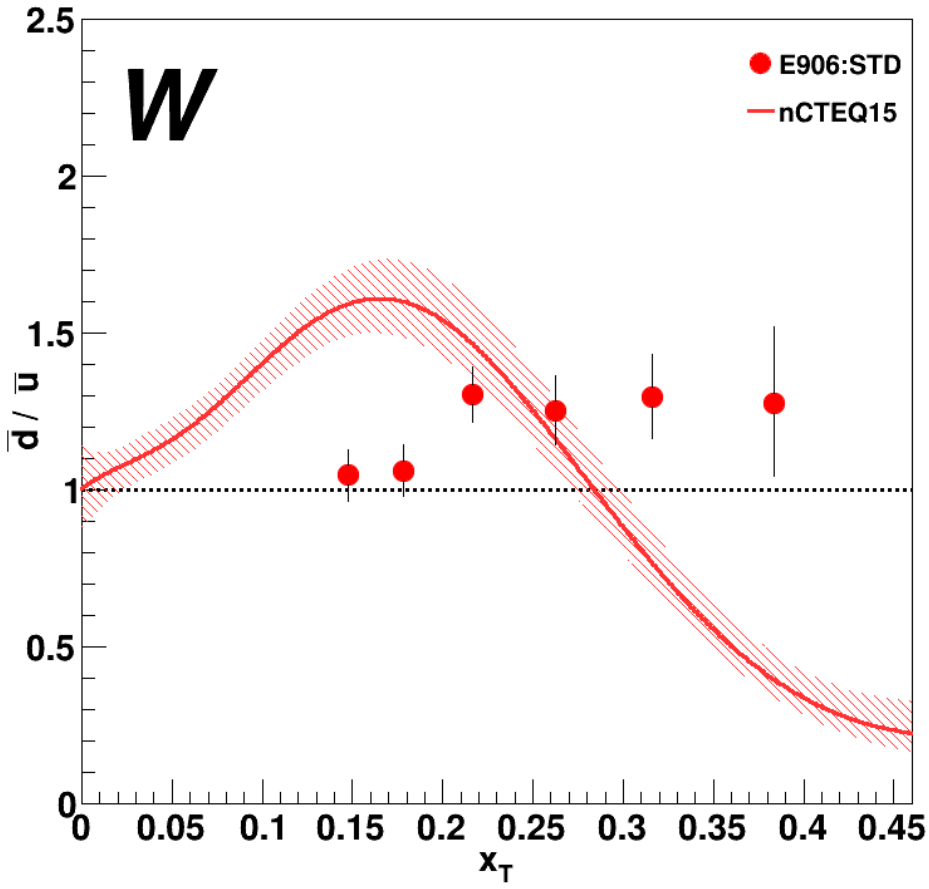


Figure 4.45: STD $\bar{d}/\bar{u}$ ratio for tungsten extracted with the nCTEQ15 nPDF set accompanied by the nCTEQ15 prediction and error band.

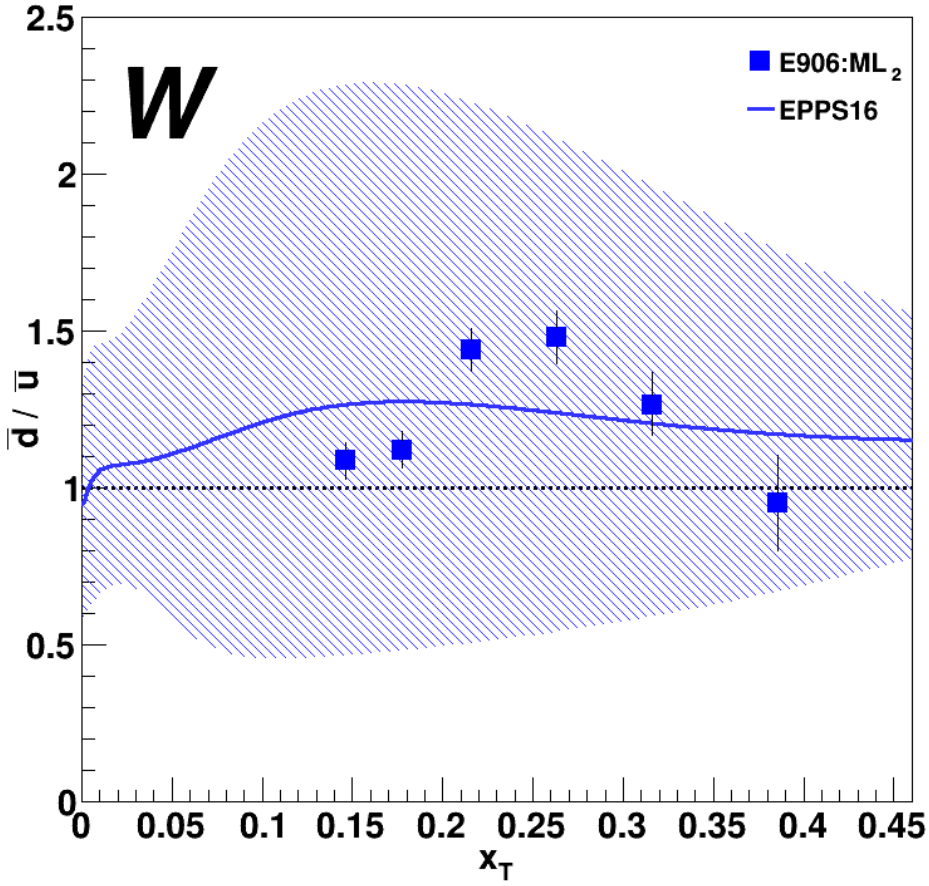


Figure 4.46: ML_2 $\bar{d}/\bar{u}$ ratio for tungsten extracted with the EPPS16 nPDF set accompanied by the EPPS16 prediction and error band.

The asymmetry results using the EPPS16 and nCTEQ15 nPDF sets with the ML_2 constraints are plotted in Figs. 4.46 and 4.47. The extracted values, statistical and systematic errors, and the covariance matrices are listed in Appendix C.

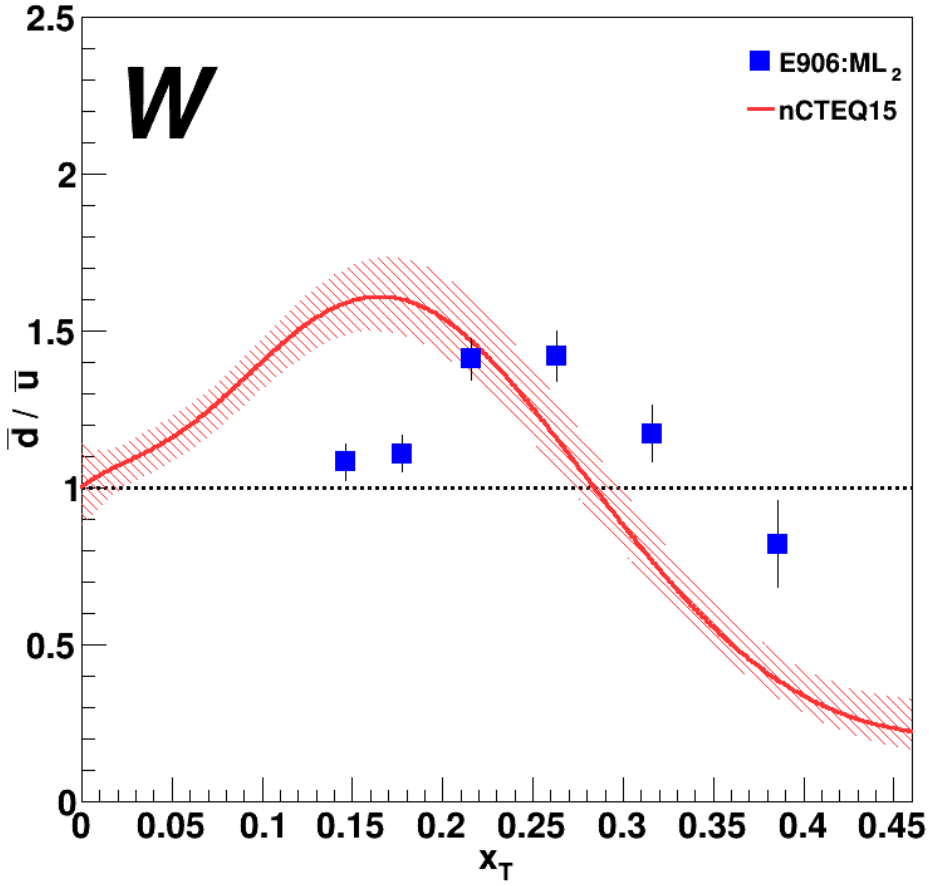


Figure 4.47: ML_2 $\bar{d}/\bar{u}$ ratio for tungsten extracted with the nCTEQ15 nPDF set accompanied by the nCTEQ15 prediction and error band.

The tungsten results follow the same pattern as the iron results, with the STD and ML_2 being consistent with the EPPS16 prediction. There is also the same shift in the asymmetry ratios in the $x_T > 0.2$ region for both the STD and ML_2 results. The ML_2 results show a decrease in the last two x_T bins, which is a consequence of the depressed R_{HA} in those two bins as seen in Table 4.7. Kumano predicted an excess of $\bar{u}$ to $\bar{d}$ for tungsten [139, 34]. In his model, parton recombination, gluons being generated from $q\bar{q}$ pairs at low x_T as mentioned in Section 1.5.1, favors the $d\bar{d}$ pairs over the $u\bar{u}$ pairs due to neutron excess. The neutron has more d quarks and there are more neutrons in heavier A elements, which would lead to an excess of $\bar{u}/\bar{d}$. The resulting asymmetry difference

is defined as

$$\bar{u}^A - \bar{d}^A = -\epsilon w_{\bar{q}}(x, A, Z)\left[\bar{u}(x) - \bar{d}(x)\right], \tag{4.22}$$

with ϵ being the neutron excess defined as

$$\epsilon \equiv \frac{A - 2Z}{A} \tag{4.23}$$

and $w_{\bar{q}}(x, A, Z)$ being the nuclear modification factor [34]. With or without the nuclear modification, the STD and ML_2 results are incompatible with the model for the majority of the iron and tungsten results. The iron and tungsten asymmetry ratios are consistent with $\bar{d}$ superseding $\bar{u}$ and the model predicts the opposite for the entire $x_T > 0.1$ range. Moreover, ϵ is zero for isoscalar targets, which would predict a $\bar{d}/\bar{u}$ of unity, which is inconsistent with the STD and ML_2 carbon results.

Lastly, the tungsten results are consistent with the $\Delta_{UL}(x)$ results from E772 [47]. $\Delta(x)$ is defined as

$$\Delta(x) = \frac{\bar{d} - \bar{u}}{\bar{d} + \bar{u}}, \tag{4.24}$$

with the E772 results extracting $\Delta(x)$ from R_{pW} and tungsten to carbon CSR data. The E772 results, which are shown for the upper 2σ limit (UL) in $\Delta(x)$, show an excess of $\bar{d}$ over $\bar{u}$ for the $x_T \in [0.04 - 0.267)$ range.

4.2.6 Comparison

Figure 4.48 illustrates that the asymmetries extracted by the EPPS16/CTEQ14 and the nCTEQ15 are close and well within the statistical errors for the STD constraints. The nCTEQ15 results are offset by 0.01 for ease of viewing. Though the EPPS16 and nCTEQ15 extractions differ slightly, the EPPS16 results are systematically higher than the nCTEQ15 results. Tables 4.19 and 4.20 detail the $\bar{d}/\bar{u}$ ratio with statistical errors for the free proton, carbon, iron, and tungsten targets using the EPPS16/CTEQ14 and nCTEQ15 PDF sets.

x_T bin	H	C	Fe	W
$[0.130 - 0.160)$	$1.437\ ^{+0.086}_{-0.092}$	$1.454\ ^{+0.117}_{-0.110}$	$1.257\ ^{+0.100}_{-0.105}$	$1.053\ ^{+0.080}_{-0.084}$
$[0.160 - 0.195)$	$1.340\ ^{+0.089}_{-0.090}$	$1.289\ ^{+0.112}_{-0.111}$	$1.315\ ^{+0.099}_{-0.105}$	$1.071\ ^{+0.084}_{-0.080}$
$[0.195 - 0.240)$	$1.477\ ^{+0.095}_{-0.097}$	$1.557\ ^{+0.123}_{-0.122}$	$1.546\ ^{+0.109}_{-0.110}$	$1.323\ ^{+0.089}_{-0.089}$
$[0.240 - 0.290)$	$1.484\ ^{+0.123}_{-0.122}$	$1.550\ ^{+0.156}_{-0.154}$	$1.518\ ^{+0.142}_{-0.140}$	$1.292\ ^{+0.114}_{-0.112}$
$[0.290 - 0.350)$	$1.650\ ^{+0.155}_{-0.151}$	$1.637\ ^{+0.203}_{-0.197}$	$1.593\ ^{+0.177}_{-0.174}$	$1.378\ ^{+0.145}_{-0.143}$
$[0.350 - 0.450)$	$1.599\ ^{+0.229}_{-0.227}$	$1.635\ ^{+0.358}_{-0.339}$	$1.401\ ^{+0.288}_{-0.276}$	$1.427\ ^{+0.260}_{-0.250}$

Table 4.19: Comparison of the $\bar{d}/\bar{u}$ results using the STD constraints and the EPPS16/CTEQ14 PDF set for the free proton, carbon, iron, and tungsten.

x_T bin	H	C	Fe	W
$[0.130 - 0.160)$	$1.442\ ^{+0.092}_{-0.094}$	$1.463\ ^{+0.114}_{-0.118}$	$1.254\ ^{+0.100}_{-0.106}$	$1.049\ ^{+0.080}_{-0.084}$
$[0.160 - 0.195)$	$1.340\ ^{+0.091}_{-0.092}$	$1.283\ ^{+0.116}_{-0.102}$	$1.307\ ^{+0.99}_{-0.103}$	$1.061\ ^{+0.084}_{-0.079}$
$[0.195 - 0.240)$	$1.471\ ^{+0.097}_{-0.097}$	$1.538\ ^{+0.121}_{-0.121}$	$1.521\ ^{+0.106}_{-0.107}$	$1.304\ ^{+0.089}_{-0.088}$
$[0.240 - 0.290)$	$1.466\ ^{+0.123}_{-0.122}$	$1.504\ ^{+0.150}_{-0.149}$	$1.462\ ^{+0.135}_{-0.133}$	$1.253\ ^{+0.110}_{-0.109}$
$[0.290 - 0.350)$	$1.617\ ^{+0.151}_{-0.148}$	$1.542\ ^{+0.184}_{-0.181}$	$1.482\ ^{+0.157}_{-0.157}$	$1.298\ ^{+0.136}_{-0.134}$
$[0.350 - 0.450)$	$1.559\ ^{+0.226}_{-0.223}$	$1.466\ ^{+0.323}_{-0.306}$	$1.236\ ^{+0.256}_{-0.245}$	$1.278\ ^{+0.244}_{-0.233}$

Table 4.20: Comparison of the $\bar{d}/\bar{u}$ results using the STD constraints and the nCTEQ15 nPDF set for the free proton, carbon, iron, and tungsten.

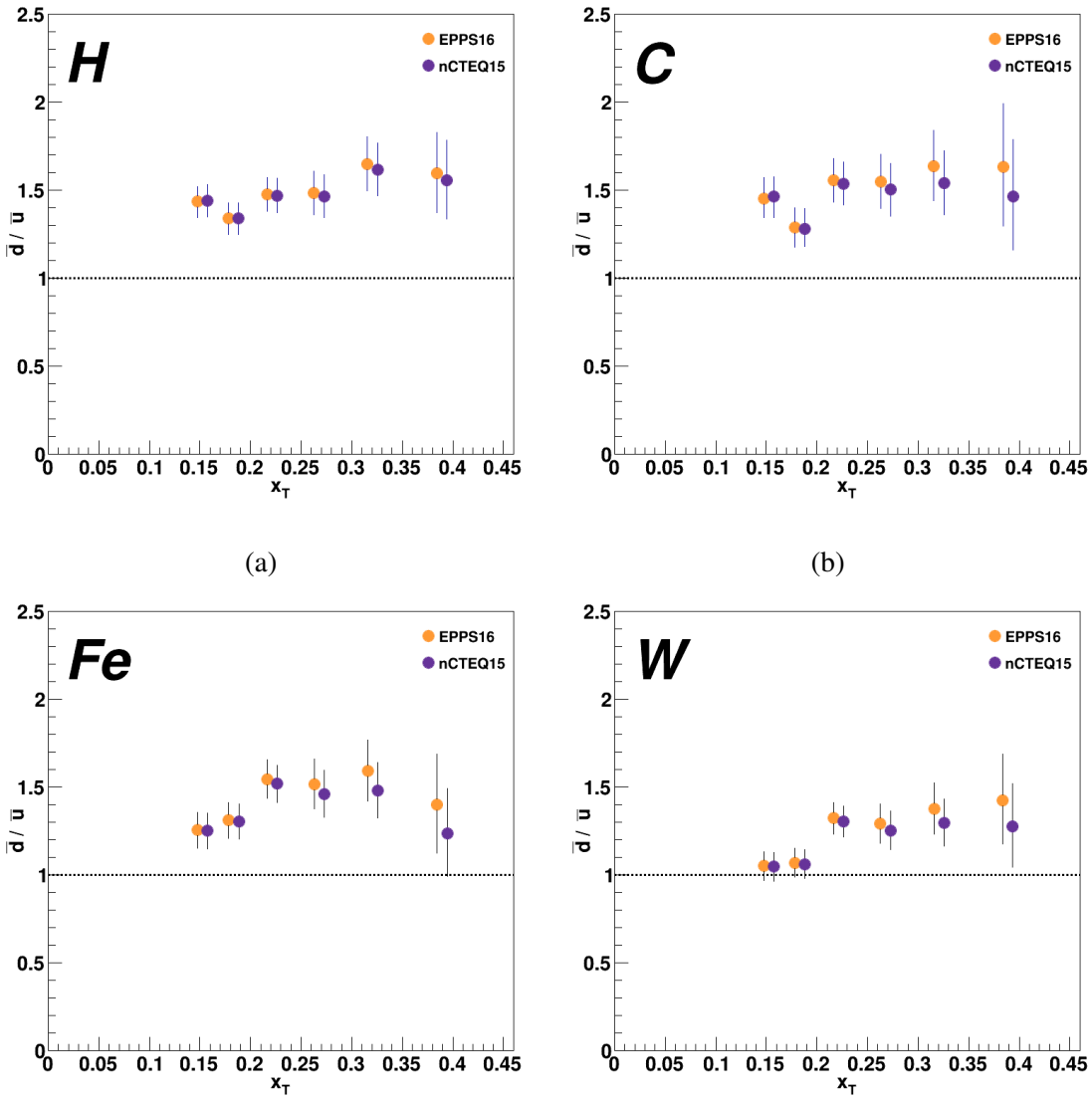


Figure 4.48: Extracted $\bar{d}/\bar{u}$ ratio using the STD constraints comparing the EPPS16/CTEQ14 and nCTEQ15 results: a) the free proton asymmetry ratio, b) the carbon asymmetry ratio, c) the iron asymmetry ratio, and d) the tungsten asymmetry ratio.

For the ML_2 constraints, Fig. 4.49 also showcases the similarity between the asymmetries extracted by the EPPS16/CTEQ14 and the nCTEQ15 PDFs. As with the STD results, the nCTEQ15 results are offset by 0.01. The values for both sets are in Tables 4.21 and 4.22, respectively. These tables detail the $\bar{d}/\bar{u}$ ratio along with their statistical errors for the free proton, carbon, iron, and tungsten targets using the EPPS16/CTEQ14 and nCTEQ15 PDF sets.

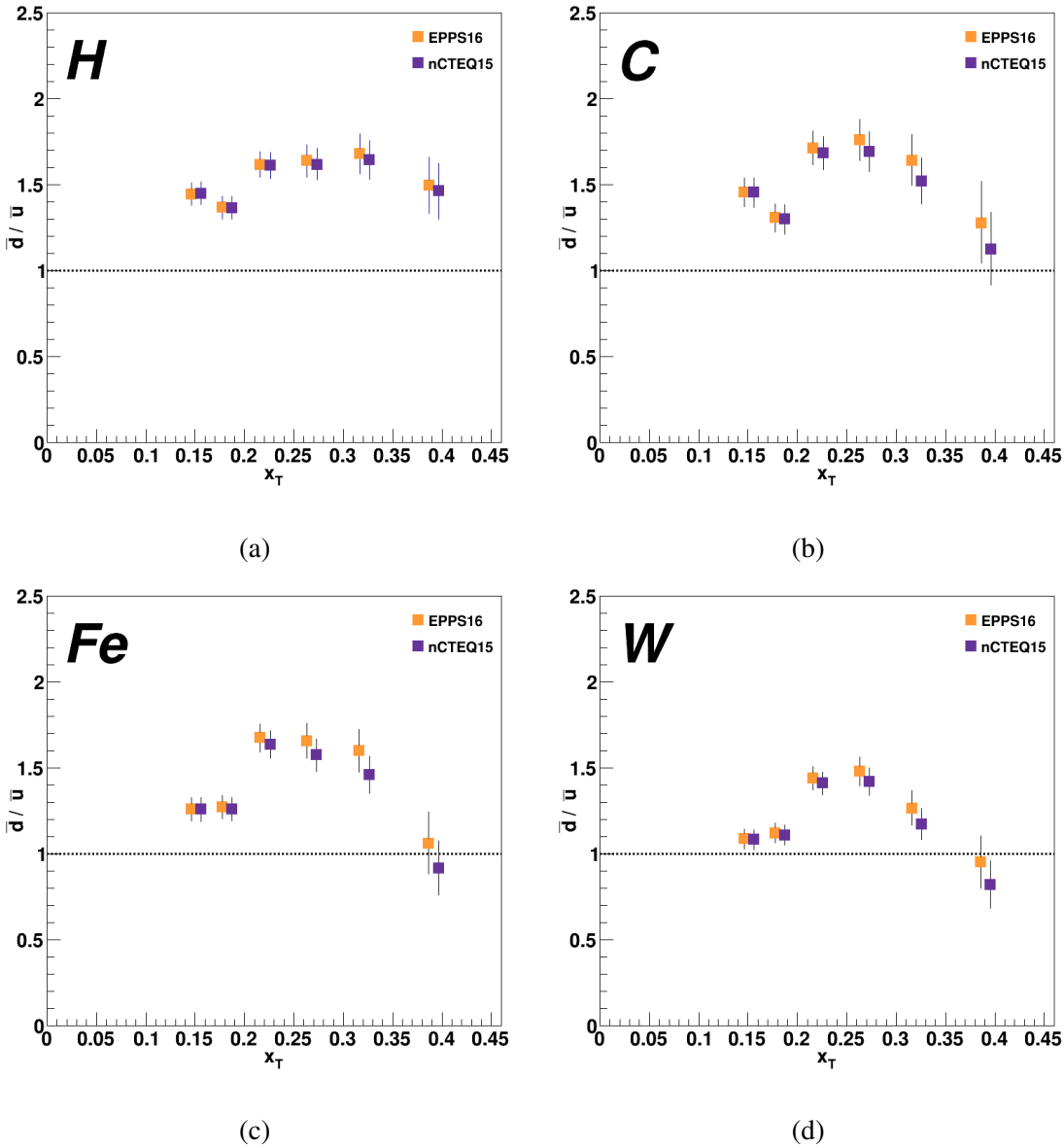


Figure 4.49: Extracted $\bar{d}/\bar{u}$ ratio using the ML_2 constraints comparing the EPPS16/CTEQ14 and nCTEQ15 results: a) the free proton asymmetry ratio, b) the carbon asymmetry ratio, c) the iron asymmetry ratio, and d) the tungsten asymmetry ratio.

x_T bin	H	C	Fe	W
$[0.130 - 0.160)$	$1.446\ ^{+0.067}_{-0.066}$	$1.458\ ^{+0.084}_{-0.086}$	$1.261\ ^{+0.068}_{-0.070}$	$1.090\ ^{+0.056}_{-0.059}$
$[0.160 - 0.195)$	$1.371\ ^{+0.063}_{-0.069}$	$1.310\ ^{+0.080}_{-0.087}$	$1.273\ ^{+0.069}_{-0.069}$	$1.122\ ^{+0.058}_{-0.055}$
$[0.195 - 0.240)$	$1.619\ ^{+0.075}_{-0.074}$	$1.713\ ^{+0.098}_{-0.098}$	$1.676\ ^{+0.082}_{-0.081}$	$1.442\ ^{+0.067}_{-0.068}$
$[0.240 - 0.290)$	$1.639\ ^{+0.094}_{-0.094}$	$1.761\ ^{+0.120}_{-0.122}$	$1.658\ ^{+0.103}_{-0.102}$	$1.482\ ^{+0.083}_{-0.083}$
$[0.290 - 0.350)$	$1.679\ ^{+0.116}_{-0.114}$	$1.642\ ^{+0.151}_{-0.147}$	$1.600\ ^{+0.126}_{-0.123}$	$1.267\ ^{+0.101}_{-0.096}$
$[0.350 - 0.450)$	$1.495\ ^{+0.164}_{-0.164}$	$1.277\ ^{+0.243}_{-0.232}$	$1.061\ ^{+0.183}_{-0.175}$	$0.9515\ ^{+0.154}_{-0.148}$

Table 4.21: Comparison of the $\bar{d}/\bar{u}$ results using the ML_2 constraints and the EPPS16/CTEQ14 PDF set for the free proton, carbon, iron, and tungsten.

x_T bin	H	C	Fe	W
$[0.130 - 0.160)$	$1.447\ ^{+0.068}_{-0.063}$	$1.458\ ^{+0.083}_{-0.087}$	$1.259\ ^{+0.070}_{-0.072}$	$1.085\ ^{+0.056}_{-0.059}$
$[0.160 - 0.195)$	$1.367\ ^{+0.068}_{-0.067}$	$1.301\ ^{+0.083}_{-0.087}$	$1.262\ ^{+0.068}_{-0.069}$	$1.111\ ^{+0.057}_{-0.057}$
$[0.195 - 0.240)$	$1.613\ ^{+0.076}_{-0.076}$	$1.685\ ^{+0.096}_{-0.096}$	$1.637\ ^{+0.080}_{-0.079}$	$1.412\ ^{+0.066}_{-0.066}$
$[0.240 - 0.290)$	$1.618\ ^{+0.094}_{-0.091}$	$1.691\ ^{+0.116}_{-0.115}$	$1.577\ ^{+0.092}_{-0.096}$	$1.420\ ^{+0.081}_{-0.079}$
$[0.290 - 0.350)$	$1.645\ ^{+0.113}_{-0.111}$	$1.521\ ^{+0.137}_{-0.131}$	$1.460\ ^{+0.109}_{-0.108}$	$1.173\ ^{+0.090}_{-0.089}$
$[0.350 - 0.450)$	$1.463\ ^{+0.163}_{-0.161}$	$1.124\ ^{+0.217}_{-0.207}$	$0.9172\ ^{+0.160}_{-0.154}$	$0.8212\ ^{+0.141}_{-0.136}$

Table 4.22: Comparison of the $\bar{d}/\bar{u}$ results using the ML_2 constraints and the nCTEQ15 nPDF set for the free proton, carbon, iron, and tungsten.

Tables 4.19 to 4.22 illustrate that the bound proton asymmetry differs from that of the free proton. To quantify this difference an asymmetry ratio standard deviation difference was constructed,

$$D_{ASY} = \frac{|ASY_{free} - ASY_{bound}|}{\sigma_{comp}}, \tag{4.25}$$

with ASY_{free} being the $\bar{d}/\bar{u}$ ratio from the free proton, ASY_{bound} being the ratio from the bound proton, and σ_{comp} being defined as

$$\sigma_{comp} = \sqrt{\sigma_{free}^2 + \sigma_{bound}^2}, \tag{4.26}$$

with σ_{free} and σ_{bound} being the upper statistical errors from the asymmetry ratio from the free proton and the bound proton, respectively. The results using the EPPS16 nPDFs and the CTEQ14 PDF are listed in Tables 4.23 and 4.24.

x_T bin	C	Fe	W
$[0.130 - 0.160)$	0.204	2.083	4.447
$[0.160 - 0.195)$	0.568	0.277	3.012
$[0.195 - 0.240)$	0.838	0.730	1.610
$[0.240 - 0.290)$	0.533	0.280	1.553
$[0.290 - 0.350)$	0.085	0.368	1.761
$[0.350 - 0.450)$	0.157	0.862	0.749

Table 4.23: Asymmetry ratio standard deviation difference for the STD constraints between the free proton asymmetry ratio derived from the CTEQ14 PDF set and the asymmetry ratio for carbon, iron, and tungsten derived from the EPPS16 nPDF set.

For the STD results, the C and most of the Fe asymmetry ratios are within the statistical error of the free proton asymmetry. This suggests a negligible modification of the ratios. The W results suggest a modification across the board except for the last x_T bin, which is within σ_{comp} due to the large statistical errors in both the free proton and W asymmetry ratios. For the ML_2 cuts, modification of the ratio is seen in all targets, with stronger evidence seen in the tungsten target. There seems to be an A dependence in the deviations, for the deviations increase as A increases for

x_T bin	C	Fe	W
$[0.130 - 0.160)$	0.175	2.757	5.317
$[0.160 - 0.195)$	0.957	1.543	3.951
$[0.195 - 0.240)$	1.251	0.754	2.343
$[0.240 - 0.290)$	1.286	0.194	1.668
$[0.290 - 0.350)$	0.322	0.686	3.547
$[0.350 - 0.450)$	1.335	2.659	3.327

Table 4.24: Asymmetry ratio standard deviation difference for the ML_2 constraints between the free proton asymmetry ratio derived from the CTEQ14 PDF set and the asymmetry ratio for carbon, iron, and tungsten derived from the EPPS16 nPDF set.

most of the x_T bins for the STD and ML_2 results. This pattern could be a consequence of a Pauli blocking like effect within the proton and the neutron.

Pauli blocking is the manifestation of the Pauli exclusion principle from the perturbative generation of $q\bar{q}$ pairs by the gluons. If the gluon is to split into a $q\bar{q}$ pair then the u and d and their antimatter counterparts are more likely to be created than the heavier quarks and antiquarks. In a region of space that already contains u quarks, like the proton, the probability that the gluon splits into a $u\bar{u}$ is diminished since the available set of quantum states for the u quark are already constrained by the presence of extant u quarks. Since the proton contains two valence u quarks, the probability of generating $d\bar{d}$ is higher than $u\bar{u}$ and the converse is true for the neutron. This leads to more $\bar{d}$ relative to $\bar{u}$ for isotopes who have more protons than neutrons and a decreased $\bar{d}/\bar{u}$ ratio for isotopes with more neutrons than protons. Consequently, the carbon and free proton asymmetry ratios are more similar than the free proton and iron or tungsten. Moreover, the modification in tungsten is the largest owning to having significantly more neutrons than protons.

This effect cannot be a simple manifestation of free proton $\bar{d}/\bar{u}$ ratio and isospin symmetry, as in

$$ASY_A = \frac{Z * ASY_{free}}{A} + \frac{A - Z}{A * ASY_{free}}, \tag{4.27}$$

with ASY being the asymmetry ratio and the inverse ratio would be the expected neutron asym-

metry ratio due to isospin symmetry. This measure yields significantly smaller asymmetry ratios that are inconsistent with the STD or ML_2 results with all differences beyond the statistical error. Another explanation could be that the $\bar{d}$ and $\bar{u}$ have different nuclear dependencies, which leads to a difference in the ratio.

Lastly, the STD and ML_2 results need to be compared. The results for the free proton and the carbon, iron, and tungsten bound proton are plotted in Fig. 4.50 and detailed in Table 4.25. The ML_2 results are offset by 0.01 in the figure. Since the differences between the EPPS16 and the nCTEQ15 extracted asymmetries was small, only the EPPS16 set was used for the STD and ML_2 comparison. The differences in the values echo the differences seen in the R_{pD} and R_{HA} ratios in Table 4.17 with the STD and the ML_2 being compatible for most bins. For Table 4.25, D_{ASY} is defined as

$$D_{ASY} = \frac{|ASY_{STD} - ASY_{ML_2}|}{\sigma_{comp}}, \tag{4.28}$$

with ASY_{STD} being the asymmetry ratio calculated from the STD constraints, ASY_{ML_2} being the asymmetry ratio calculated from the ML_2 constraints, and σ_{comp} defined as

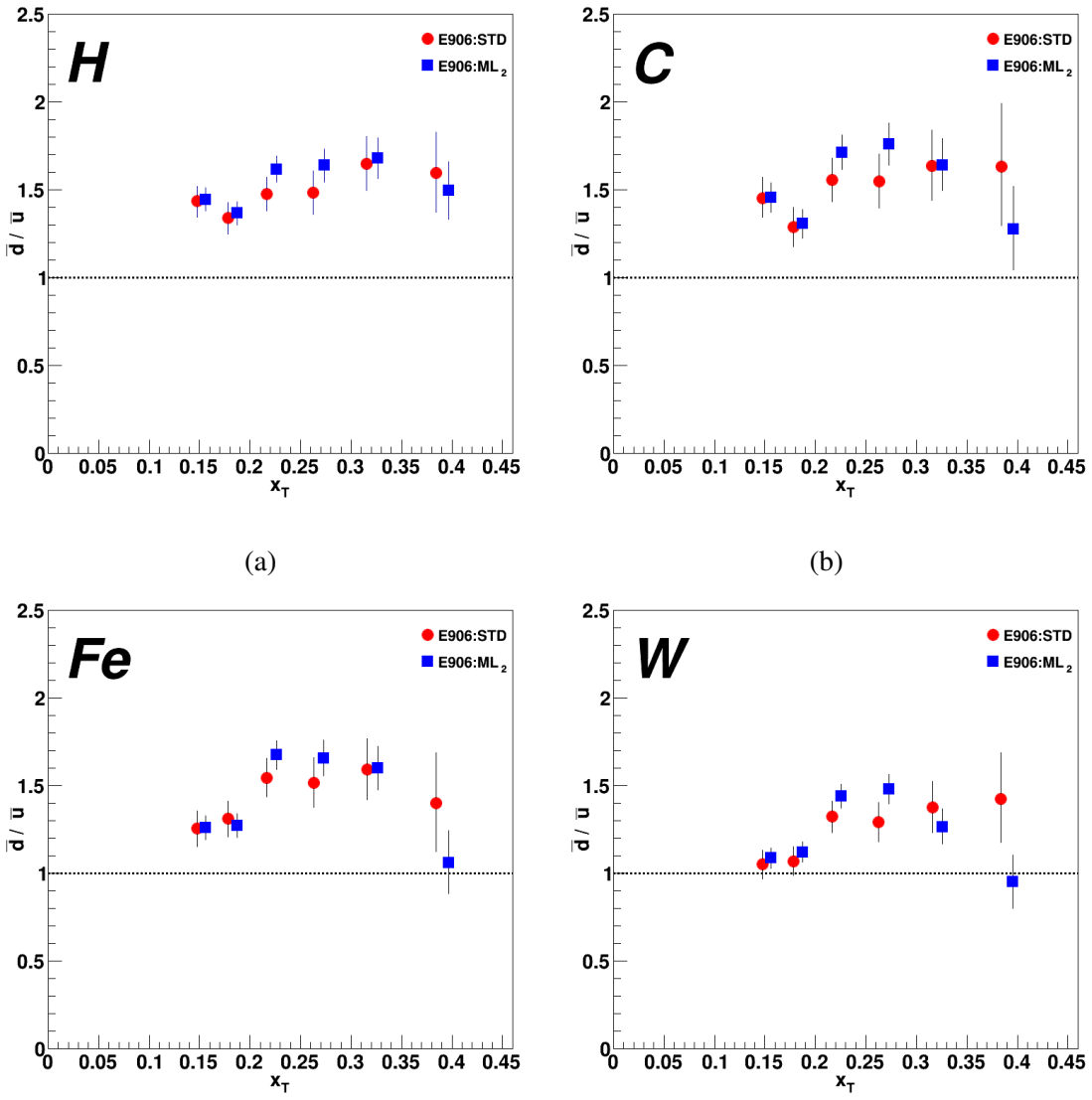
$$\sigma_{comp} = \sqrt{\sigma^2_{STD} + \sigma^2_{ML_2}}, \tag{4.29}$$

with σ_{STD} and σ_{ML_2} being the upper statistical errors on the asymmetry ratio from the STD and ML_2 constraints, respectively.

The D_{ASY} results reverberate with the D_{STD} results in Table 4.17 with most of the asymmetry ratios being within the statistical error of the STD set. The last x_T bin is where they disagree most strongly, which is a remnant with the large CSR disagreements in that bin.

x_T bin	H	C	Fe	W
$[0.130 - 0.160)$	0.086	0.025	0.036	0.375
$[0.160 - 0.195)$	0.283	0.155	0.345	0.501
$[0.195 - 0.240)$	1.168	0.994	0.948	1.070
$[0.240 - 0.290)$	1.002	1.073	0.793	1.345
$[0.290 - 0.350)$	0.151	0.019	0.030	0.628
$[0.350 - 0.450)$	0.366	0.827	0.998	1.573

Table 4.25: Asymmetry ratio standard deviation difference between the STD and the ML_2 constraints using the CTEQ14 PDF set for the free proton and the EPPS16 nPDF set for carbon, iron, and tungsten.



(a) (b)

(c) (d)

Figure 4.50: Comparison of the STD and ML_2 asymmetry ratio results using the CTEQ14 PDF for the free proton and the EPPS16 nPDFs for the nuclear targets: a) the free proton asymmetry results, b) the carbon results, c) the iron results, and d) the tungsten results.

Chapter 5:

Conclusion

In summation, using the Main Injector beam from Fermilab and the SeaQuest spectrometer, the Drell-Yan (DY) cross section ratios of LD_2, C, Fe, and W to LH_2 and the nuclear cross section ratios of C, Fe, and W to LD_2 as functions of x_T have been measured and analyzed. For the analysis, two methods were employed: the standard analysis and machine learning. The standard analysis set of constraints were a set of rectangular-type constraints based upon physics and spectrometer information developed by the SeaQuest collaboration. Multiple machine learning algorithms were studied including Support Vector Machines, Neural Nets, Probability Density Estimator Foams, and Boosted Decisions Trees (BDTs), but BDTs were shown to be optimal for this analysis. A set of BDTs were trained on a relaxed set of constraints and the resulting model, machine learning model 2, increased the DY yield by more than 20%.

The cross section ratios were generated as functions of x_T and I_T with the intensity dependence being described as a quadratic function, common polynomial 2, which was chosen based upon the Akaike Information Criterion results. The parameters of the function were found through a χ^2 minimization and the function at zero intensity was taken to be the cross section ratio. The statistical errors were derived from changing the cross section ratio value until $\Delta\chi^2 = 1$. The systematic errors were derived from the uncertainties in pedestal, beam normalization, intensity dependence, mass constraint, and for the machine learning-machine learning cut.

The DY cross section ratios were measured in the $x_T \in [0.13 - 0.45)$ range. The deuterium to hydrogen results, for both sets of constraints, are consistent with predictions [134] in the $x_T < 0.3$

region and the E886 results in the $x_T < 0.2$ region [40]. The nuclear to LD_2 cross section ratio results are largely consistent with the E772 results for the carbon, iron, and tungsten targets and the carbon results are consistent with the isoscalar cross section ratio prediction [134], assuming a small Q^2 dependence. The standard analysis and machine learning results display a slight EMC effect in the $x_T \in (0.2 - 0.45)$ region, as the cross section ratio distributions are more consistent with a linear functional form than a constant form. More data in the $x_T > 0.4$ region would help to demonstrate whether an EMC effect is seen in DY. Though the standard analysis and the machine learning results are compatible when taking into consideration the statistical and systematic errors, there are some deviations between them especially in regards to some of higher x_T bins. This difference is due to machine learning possessing a bias in the liquid targets that is more apparent in the high x_T and I_T regions. There is also the issue that the machine learning finds more DY coming from the EMPTY and NONE targets, which has led to differences in the cross section and cross section ratio. As an analysis method, machine learning has shown success, but could be improved by training on a more accurate set of monte carlos and on the EMPTY and NONE targets, using an explicit acceptance correction, and by using a secondary machine learning algorithm to construct the models.

From the LD_2 to LH_2 cross section ratio, the $\bar{d}/\bar{u}$ ratio was extracted for the free proton using a set of weighted events. This free proton asymmetry ratio was extracted from the CTEQ14 and nCTEQ15 PDF sets, with both sets yielding comparable values. However, the predicted value of the ratio is more consistent with the CTEQ14 predictions than the nCTEQ15 predictions, which is to be expected since the nCTEQ15 set is closer to CTEQ6, an older CTEQ PDF set. The free proton asymmetry results are all above unity, in direct tension with the E886 results in the $x_T > 0.2$ region. The statistical model prediction of Bourrely and Soffer [44] is closer to the machine learning results than the standard analysis results, but it predicts a larger asymmetry than is observed.

The nuclear asymmetry ratio was extracted using the asymmetry results derived from the free proton and the nuclear to LH_2 cross section ratios. These ratios were largely above unity, which

in the case of the carbon and tungsten results conflict with the prediction of unity for carbon and less than unity for tungsten [34, 139]. However, the tungsten $\bar{d}/\bar{u}$ ratio is consistent with the E772 results [47]. The EPPS16 nPDF, CTEQ14 PDF, and the free proton asymmetry using the CTEQ14 PDF were compared to the nCTEQ15 nPDF and the free proton asymmetry using the nCTEQ15 PDF. As with the free proton asymmetry results, the differences in the nuclear asymmetry ratios between the two sets were negligible, but the EPPS16 nPDF was more consistent with the results. The influence of the E886 results in the nCTEQ15 seems to be a significant cause of the discrepancy. The $\bar{d}/\bar{u}$ ratio for a bound proton shows some modification from that of a free proton with a slight A dependence. The carbon results are closer to the free proton results, then followed by the iron and tungsten. This suggests that the higher the neutron to proton ratio within the target, the lower the asymmetry ratio. This trend could be a consequence of a Pauli-blocking effect or a differing in the nuclear dependencies of the $\bar{d}$ and $\bar{u}$ quarks.

Appendix A:

Standard Analysis Constraints

Listed in the code below is the MYSQL query for all of the analysis constraints gathered from Ref. [106]. A definition of the variables can be found in Ref. [112].

```
schema = R008SCHEMA


FROM
    schema + .kDimuon a
    Join + schema + .kTrack as b using(spillid, eventid, runid)
    Join + schema + .kTrack as c using(spillid, eventid, runid)
    Join + schema + .kEvent as e using(spillid, eventid, runid)


    schema = R007SCHEMA
    Join  + schema + .Occupancy as d using(spillid, eventid, runid)
    Join  + schema + .Event as f using(spillid, eventid, runid)
    Join  + schema + .Spill as g using(spillid, runid, targetPos)
    Join  + schema + .BeamDAQ as h using(spillid)
    Join  + schema + .Beam as j using(spillid, runid)
    Join  + schema + .QIE as i using(spillid, eventid, runid)


WHERE
```

```
a.posTrackID = b.trackID
and a.negTrackID = c.trackID
and g.dataQuality = 0
and f.matrix1 = 1
and j.name = \"S:G2SEM\"
and b.roadID*c.roadID < 0
and a.targetPos>0
and a.dz between -280 and -5
and abs(a.dx)<.25
and abs(a.dy-y_offset)<.22
and abs(a.dpx)<1.8
and abs(a.dpy)<2
and a.dpx*a.dpx+a.dpy*a.dpy<5
and a.dpz between 38 and 116
and a.mass between 4.2 and 8.8
and a.dx*a.dx+(a.dy-y_offset)*(a.dy-y_offset)<.06
and a.xT between 0.05 and 0.58
and a.xF between -.1 and .95
and abs(a.trackSeparation)<270
and a.chisq_dimuon<18
and b.chisq_target<15
and b.pz1 between 9 and 75
and b.numHits>13
and b.$x_T$*b.$x_T$+(b.yT-y_offset)*(b.yT-y_offset)<320
and b.xD*b.xD+(b.yD-y_offset)*(b.yD-y_offset)<1100
and b.xD*b.xD+(b.yD-y_offset)*(b.yD-y_offset)>16
and b.chisq_target<1.5*b.chisq_upstream
```

```
and b.chisq_target<1.5*b.chisq_dump
and b.charge=+1
and b.z0 between -320 and -5
and b.chisq/(b.numHits-5)<12
and b.y1/b.y3<1
and abs(abs(b.px1-b.px3)-.416)<.008
and abs(b.py1-b.py3)<.008
and abs(b.pz1-b.pz3)<.08
and b.y1*b.y3>0
and c.chisq_target<15
and c.pz1 between 9 and 75
and c.numHits>13
and c.xT*c.xT+(c.yT-y_offset)*(c.yT-y_offset)<320
and c.xD*c.xD+(c.yD-y_offset)*(c.yD-y_offset)<1100
and c.xD*c.xD+(c.yD-y_offset)*(c.yD-y_offset)>16
and c.chisq_target<1.5*c.chisq_upstream
and c.chisq_target<1.5*c.chisq_dump
and c.charge=-1
and c.z0 between -320 and -5
and c.chisq/(c.numHits-5)<12
and c.y1/c.y3<1
and abs(abs(c.px1-c.px3)-.416)<.008
and abs(c.py1-c.py3)<.008
and abs(c.pz1-c.pz3)<.08
and c.y1*c.y3>0
and abs(b.chisq_target+c.chisq_target-a.chisq_dimuon)<2
and b.y3*c.y3<0
```

```
and b.numHits+c.numHits>29
and b.numHitsSt1+c.numHitsSt1>8
and abs(b.x1+c.x1)<42
and abs(b.py1)>.02
and abs(c.py1)>.02
and d.D1<400
and d.D2<400
and d.D3<400
and d.D1+d.D2+d.D3<1000
```

Appendix B:

Machine Learning Model 2 Constraints

Listed in the code below is the MYSQL query for all of the analysis constraints for ML_2.

```
schema = R008SCHEMA

FROM
    schema + .kDimuon a
    Join + schema + .kTrack as b using(spillid, eventid, runid)
    Join + schema + .kTrack as c using(spillid, eventid, runid)
    Join + schema + .kEvent as e using(spillid, eventid, runid)

    schema = R007SCHEMA
    Join  + schema + .Occupancy as d using(spillid, eventid, runid)
    Join  + schema + .Event as f using(spillid, eventid, runid)
    Join  + schema + .Spill as g using(spillid, runid, targetPos)
    Join  + schema + .BeamDAQ as h using(spillid)
    Join  + schema + .Beam as j using(spillid, runid)
    Join  + schema + .QIE as i using(spillid, eventid, runid)

WHERE
    a.posTrackID = b.trackID
    and a.negTrackID = c.trackID
    and g.dataQuality = 0
    and f.matrix1 = 1
    and j.name = \"S:G2SEM\"
    and b.roadID*c.roadID < 0
    and a.targetPos>0
    and abs(a.dx)<.36
    and abs(a.dy-y_offset)<.22
    and a.dz between -300 and 2
    and abs(a.dpx)<2 and abs(a.dpy)<2
    and a.dpx*a.dpx+a.dpy*a.dpy<6.
    and a.dpz between 37 and 116
```

```
and a.mass between 4.2 and 8.8
and a.dx*a.dx+(a.dy-y_offset)*(a.dy-y_offset)<.17
and a.xF between -.15 and .95
and a.xT between 0.0 and 0.6
and abs(a.costh)<.58
and abs(a.trackSeparation)<345
and a.chisq\_dimuon<60
and b.chisq\_target<42
and b.pz1 between 9 and 75
and b.numHits>13
and b.xT*b.xT+(b.yT-y_offset)*(b.yT-y_offset)<1350
and b.xD*b.xD+(b.yD-y_offset)*(b.yD-y_offset)<1200
and b.xD*b.xD+(b.yD-y_offset)*(b.yD-y_offset)>10
and b.chisq\_target<2.7*b.chisq\_upstream
and b.chisq\_target<11*b.chisq\_dump
and b.charge=+1
and b.z0 between -350 and 85
and b.chisq/(b.numHits-5)<13
and c.chisq\_target<42
and c.pz1 between 9 and 75
and c.numHits>13
and c.xT*c.xT+(c.yT-y_offset)*(c.yT-y_offset)<1350
and c.xD*c.xD+(c.yD-y_offset)*(c.yD-y_offset)<1200
and c.xD*c.xD+(c.yD-y_offset)*(c.yD-y_offset)>10
and c.chisq\_target<2.7*c.chisq\_upstream
and c.chisq\_target<11*c.chisq\_dump
and c.charge=-1
and c.z0 between -350 and 85
and c.chisq/(c.numHits-5)<13
and abs(b.py1-b.py3)<.008
and abs(c.py1-c.py3)<.008
and abs(b.pz1-b.pz3)<.08
and abs(c.pz1-c.pz3)<.08
and abs(b.chisq_target+c.chisq_target-chisq_dimuon)<17.5
and b.y1*b.y3>0
and c.y1*c.y3>0
and b.y3*c.y3<0
and b.numHits+c.numHits>28
and b.numHitsSt1+c.numHitsSt1>7

if (bFlip)
{
  and abs(b.px1-b.px3+.416)<.008 and abs(c.px1-c.px3-.416)<.008
  and least(b.pz1,c.pz1)-greatest(b.x1,-c.x1)/4.5>11
}
```

```
if (!bFlip)
{
  and abs(b.px1-b.px3-.416)<.008 and abs(c.px1-c.px3+.416)<.008
  and least(b.pz1,c.pz1)-greatest(-b.x1,c.x1)/4.5>11
}.
```

Care was taken for Roadsets 57-62 due to a change in offset and an opposite signed magnetic field in relation to Roadset 67. Roadset 67 was taken as the standard set and the x-position and momentum dependent variables such as: $px0$, $px1$, $px2$, $px3$, pxD, px_T, $x0$, $x1$, $x3$, xD, x_T(kTracker), $thbend$, $roadid$, and dpx; had their signs inverted to mimic Roadset 67. In addition, the y-position variables: dy, $y0$, yD, yT, $y1$, and $y3$ in Roadsets 57 and 62 also had 1.2 cm added to them. This alters the y-offset from 0.4 to 1.6, which is line with Roadset 67. Lastly, after looking at the y-position variables from Roadsets 57-62, there was still an offset. This offset does not appreciably effect any of the liquid ML results, but it was corrected in case the nuclear targets ML weight files depended on the y positions.

A function was fitted using the average values of yT, yD, $y1$, and $y3$ for each kTrack and their respective z-positions for Roadsets 57, 59, 62, and 67. The corrected value of the y-positions was determined as

$$y_{f,RS} = y_{i,RS} + F_{67} - F_{RS}, \tag{B.1}$$

which was implemented as

$$y_f = y_i - y_{offset} + 1.6 + p1 * z_{position} + p0. \tag{B.2}$$

The variables and the parameters for Roadsets 57 and 59 are listed in Table B.1 and those for Roadset 62 are in Table B.2.

Variable	kTrack Charge	Z position	p0	p1
yT	$+$	-1.29	-0.52047	-0.368
yD	$+$	0.42	-0.52047	-0.368
$y1$	$+$	6.00	-0.52047	-0.368
$y3$	$+$	19.10	-0.52047	-0.368
yT	$-$	-1.29	-0.010392	0.24875
yD	$-$	0.42	-0.010392	0.24875
$y1$	$-$	6.00	-0.010392	0.24875
$y3$	$-$	19.10	-0.010392	0.24875

Table B.1: Table of y-position variables and parameters for Roadsets 57 and 59.

Variable	kTrack Charge	Z position	p0	p1
yT	$+$	-1.29	-0.0959305	-0.19249
yD	$+$	0.42	-0.0959305	-0.19249
$y1$	$+$	6.00	-0.0959305	-0.19249
$y3$	$+$	19.10	-0.0959305	-0.19249
yT	$-$	-1.29	0.01240	0.168196
yD	$-$	0.42	0.01240	0.168196
$y1$	$-$	6.00	0.01240	0.168196
$y3$	$-$	19.10	0.01240	0.1681960

Table B.2: Table of y-position variables and parameters for Roadset 62.

Appendix C:

Cross Section and Asymmetry Ratio Data Tables

This appendix is a listing of the CSR and asymmetry ratio tables for the STD and ML_2 constraints. The definitions of the systematic errors are detailed in Section 4.1.

C.1 Standard Analysis Cross Section Results

The extrapolated CSR, the average x_T value in the x_T bin, the statistical errors, and systematic errors for each target for the R_{pD} and R_{HA} CSRs are listed in Tables C.1 - C.4.

Ave. x_T	CSR	σ_{stat}	σ_{ped}	σ_{norm}	σ_{beam}	σ_{fit}	σ_{mass}	σ_{sys}
0.1474	1.222	0.05680	0.007	0.02443	-	0.017	0.044	0.053
0.1782	1.142	0.04718	0.003	0.02285	-	0.011	0.005	0.026
0.2163	1.192	0.04631	0.005	0.02384	-	0.021	0.037	0.045
0.2628	1.161	0.05107	0.004	0.02322	-	0.079	0.023	0.034
0.3156	1.190	0.05461	0.005	0.02379	-	0.017	0.021	0.036
0.3844	1.109	0.07119	0.008	0.02218	-	0.032	0.022	0.045

Table C.1: STD CSR results with statistical and systematic errors for R_{pD}. There is no beam systematic error for the liquid CSR.

Ave. x_T	CSR	σ_{stat}	σ_{ped}	σ_{norm}	σ_{beam}	σ_{fit}	σ_{mass}	σ_{sys}
0.1474	1.243	0.07275	0.008	0.02485	0.006213	0.027	0.046	0.060
0.1782	1.111	0.05946	0.003	0.02222	0.005555	0.017	0.039	0.048
0.2163	1.227	0.05843	0.004	0.02453	0.006133	0.027	0.005	0.026
0.2629	1.186	0.06359	0.004	0.02373	0.005932	0.015	0.002	0.029
0.3157	1.178	0.06954	0.005	0.02356	0.005890	0.033	0.000	0.041
0.3840	1.106	0.09136	0.007	0.02212	0.005529	0.061	0.000	0.065

Table C.2: STD CSR results with statistical and systematic errors for R_{HC}.

Ave. x_T	CSR	σ_{stat}	σ_{ped}	σ_{norm}	σ_{beam}	σ_{fit}	σ_{mass}	σ_{sys}
0.1476	1.121	0.06644	0.009	0.02241	0.005602	0.061	0.049	0.082
0.1784	1.132	0.05660	0.003	0.02263	0.005658	0.038	0.025	0.051
0.2163	1.236	0.05556	0.002	0.02472	0.006179	0.008	0.035	0.044
0.2629	1.184	0.06156	0.004	0.02367	0.005918	0.026	0.045	0.058
0.3159	1.177	0.06563	0.004	0.02354	0.005885	0.059	0.043	0.077
0.3844	1.045	0.08253	0.002	0.02091	0.005226	0.105	0.043	0.115

Table C.3: STD CSR results with statistical and systematic errors for R_{HFe}.

Ave. x_T	CSR	σ_{stat}	σ_{ped}	σ_{norm}	σ_{beam}	σ_{fit}	σ_{mass}	σ_{sys}
0.1476	1.021	0.06070	0.006	0.02042	0.005106	0.026	0.085	0.091
0.1783	1.012	0.05163	0.004	0.02024	0.005060	0.016	0.072	0.077
0.2162	1.152	0.05160	0.004	0.02304	0.005761	0.001	0.007	0.025
0.2628	1.104	0.05690	0.004	0.02207	0.005518	0.020	0.010	0.032
0.3159	1.104	0.06152	0.004	0.02207	0.005518	0.042	0.009	0.049
0.3836	1.057	0.08197	0.009	0.02115	0.005287	0.079	0.010	0.083

Table C.4: STD CSR results with statistical and systematic errors for R_{HW}.

The extrapolated CSR, the average x_T value in the x_T bin, the statistical errors, and systematic errors for each target for the R_{pA} CSRs are listed in Tables C.5 - C.7.

Ave. x_T	CSR	σ_{stat}	σ_{ped}	σ_{norm}	σ_{beam}	σ_{fit}	σ_{mass}	σ_{sys}
0.1473	1.009	0.05809	0.002	0.02018	0.005044	0.004	0.005	0.022
0.1781	0.9706	0.04871	0.001	0.01941	0.004853	0.002	0.028	0.035
0.2163	1.011	0.04760	0.001	0.02022	0.005056	0.000	0.035	0.041
0.2627	1.012	0.05062	0.001	0.02025	0.005061	0.001	0.016	0.026
0.3151	0.9740	0.05741	0.001	0.01948	0.004870	0.001	0.018	0.027
0.3849	0.9662	0.07543	0.003	0.01932	0.004831	0.005	0.019	0.028

Table C.5: STD CSR results with statistical and systematic errors for R_{pC}.

Ave. x_T	CSR	σ_{stat}	σ_{ped}	σ_{norm}	σ_{beam}	σ_{fit}	σ_{mass}	σ_{sys}
0.1475	0.9152	0.05213	0.004	0.01831	0.004576	0.025	0.078	0.084
0.1783	0.9983	0.04461	0.002	0.01997	0.004992	0.014	0.020	0.032
0.2163	1.039	0.04373	0.001	0.02078	0.005194	0.003	0.004	0.022
0.2627	1.029	0.04737	0.002	0.02059	0.005147	0.010	0.018	0.029
0.3153	0.9878	0.05259	0.002	0.01976	0.004939	0.025	0.018	0.037
0.3853	0.9544	0.06706	0.004	0.01909	0.004772	0.043	0.017	0.051

Table C.6: STD CSR results with statistical and systematic errors for R_{pFe}.

Ave. x_T	CSR	σ_{stat}	σ_{ped}	σ_{norm}	σ_{beam}	σ_{fit}	σ_{mass}	σ_{sys}
0.1475	0.8382	0.04574	0.003	0.01676	0.004191	0.035	0.106	0.113
0.1782	0.8929	0.03877	0.001	0.01786	0.004465	0.021	0.059	0.065
0.2162	0.9699	0.03861	0.002	0.01940	0.004849	0.003	0.039	0.044
0.2626	0.9592	0.04164	0.001	0.01918	0.004796	0.018	0.010	0.029
0.3153	0.9209	0.04721	0.001	0.01842	0.004605	0.046	0.009	0.050
0.3844	0.9709	0.06397	0.001	0.01942	0.004855	0.078	0.009	0.081

Table C.7: STD CSR results with statistical and systematic errors for R_{pW}.

C.2 Machine Learning Cross Section Results

The extrapolated CSR, the average x_T value in the x_T bin, the statistical errors, and systematic errors for each target for the R_{pD} and R_{HA} CSRs are listed in Tables C.8 - C.11.

Ave. x_T	CSR	σ_{stat}	σ_{ped}	σ_{norm}	σ_{beam}	σ_{fit}	σ_{mass}	σ_{ML}	σ_{sys}
0.1474	1.223	0.04133	0.003	0.02447	-	0.013	0.043	0.042	0.067
0.1782	1.149	0.03445	0.004	0.02297	-	0.008	0.001	0.066	0.071
0.2165	1.252	0.03572	0.003	0.02504	-	0.002	0.033	0.083	0.092
0.2633	1.223	0.03855	0.002	0.02446	-	0.006	0.025	0.096	0.103
0.3165	1.197	0.04004	0.004	0.02394	-	0.013	0.020	0.092	0.098
0.3869	1.069	0.04944	0.004	0.02139	-	0.026	0.022	0.138	0.144

Table C.8: ML_2 CSR results with statistical and systematic errors for R_{pD}. There is no beam systematic error for the liquid CSR.

Ave. x_T	CSR	σ_{stat}	σ_{ped}	σ_{norm}	σ_{beam}	σ_{fit}	σ_{mass}	σ_{ML}	σ_{sys}
0.1475	1.238	0.05308	0.003	0.02475	0.006188	0.042	0.052	0.000	0.072
0.1782	1.111	0.04395	0.004	0.02223	0.005557	0.024	0.014	0.002	0.036
0.2165	1.286	0.04540	0.003	0.02571	0.006428	0.003	0.014	0.078	0.084
0.2634	1.267	0.04856	0.002	0.02533	0.006333	0.023	0.015	0.099	0.106
0.3167	1.182	0.05080	0.005	0.02365	0.005912	0.054	0.008	0.054	0.081
0.3868	1.006	0.06129	0.005	0.02012	0.005029	0.093	0.010	0.062	0.114

Table C.9: ML_2 CSR results with statistical and systematic errors for R_{HC}.

Ave. x_T	CSR	σ_{stat}	σ_{ped}	σ_{norm}	σ_{beam}	σ_{fit}	σ_{mass}	σ_{ML}	σ_{sys}
0.1476	1.129	0.04589	0.006	0.02258	0.005643	0.048	0.022	0.048	0.075
0.1784	1.102	0.03872	0.004	0.02203	0.005508	0.030	0.015	0.023	0.047
0.2165	1.288	0.04052	0.004	0.02575	0.006438	0.007	0.033	0.023	0.049
0.2634	1.237	0.04338	0.004	0.02474	0.006185	0.022	0.033	0.014	0.049
0.3167	1.184	0.04570	0.006	0.02368	0.005921	0.050	0.031	0.004	0.064
0.3872	0.9435	0.05224	0.009	0.01887	0.004718	0.097	0.033	0.005	0.104

Table C.10: ML_2 CSR results with statistical and systematic errors for R_{HFe}.

Ave. x_T	CSR	σ_{stat}	σ_{ped}	σ_{norm}	σ_{beam}	σ_{fit}	σ_{mass}	σ_{ML}	σ_{sys}
0.1476	1.040	0.04237	0.003	0.02079	0.005198	0.058	0.039	0.030	0.080
0.1782	1.034	0.03574	0.003	0.02068	0.005171	0.037	0.024	0.012	0.050
0.2164	1.202	0.03774	0.001	0.02403	0.006008	0.010	0.020	0.040	0.052
0.2632	1.194	0.04075	0.001	0.02389	0.005972	0.023	0.025	0.091	0.100
0.3166	1.068	0.04222	0.002	0.02135	0.005338	0.056	0.024	0.085	0.107
0.3863	0.9006	0.05071	0.006	0.01801	0.004503	0.116	0.027	0.079	0.144

Table C.11: ML_2 CSR results with statistical and systematic errors for R_{HW}.

The extrapolated CSR, the average x_T value in the x_T bin, the statistical errors, and systematic errors for each target for the R_{pA} CSRs are listed in Tables C.12 - C.14.

Ave. x_T	CSR	σ_{stat}	σ_{ped}	σ_{norm}	σ_{beam}	σ_{fit}	σ_{mass}	σ_{ML}	σ_{sys}
0.1473	1.008	0.03861	0.004	0.02015	0.005038	0.006	0.007	0.035	0.042
0.1782	0.9741	0.03205	0.002	0.01948	0.004871	0.005	0.008	0.060	0.064
0.2165	1.022	0.03169	0.002	0.02044	0.005110	0.001	0.012	0.006	0.025
0.2634	1.037	0.03382	0.002	0.02077	0.005193	0.003	0.007	0.001	0.023
0.3162	0.9876	0.03781	0.003	0.01975	0.004938	0.011	0.008	0.030	0.039
0.3862	0.9021	0.04778	0.000	0.01804	0.004511	0.019	0.008	0.083	0.087

Table C.12: ML_2 CSR results with statistical and systematic errors for R_{pC}.

Ave. x_T	CSR	σ_{stat}	σ_{ped}	σ_{norm}	σ_{beam}	σ_{fit}	σ_{mass}	σ_{ML}	σ_{sys}
0.1475	0.9209	0.03486	0.002	0.01842	0.004605	0.028	0.055	0.016	0.067
0.1783	0.9718	0.02995	0.002	0.01944	0.004859	0.017	0.007	0.032	0.042
0.2165	1.035	0.02988	0.002	0.02069	0.005173	0.004	0.004	0.047	0.053
0.2632	1.026	0.03153	0.002	0.02051	0.005128	0.012	0.011	0.070	0.075
0.3163	0.9944	0.03506	0.002	0.01989	0.004972	0.031	0.014	0.089	0.097
0.3866	0.8828	0.04164	0.001	0.01766	0.004414	0.056	0.013	0.131	0.144

Table C.13: ML_2 CSR results with statistical and systematic errors for R_{pFe}.

Ave. x_T	CSR	σ_{stat}	σ_{ped}	σ_{norm}	σ_{beam}	σ_{fit}	σ_{mass}	σ_{ML}	σ_{sys}
0.1474	0.8472	0.03101	0.004	0.01694	0.004236	0.056	0.074	0.002	0.095
0.1782	0.9034	0.02651	0.001	0.01807	0.004517	0.034	0.023	0.045	0.064
0.2164	0.9596	0.02659	0.001	0.01919	0.004798	0.009	0.009	0.029	0.037
0.2630	0.9809	0.02822	0.002	0.01962	0.004904	0.022	0.002	0.003	0.030
0.3161	0.8804	0.03126	0.001	0.01761	0.004402	0.060	0.006	0.009	0.063
0.3857	0.8444	0.03873	0.001	0.01689	0.004222	0.109	0.006	0.060	0.126

Table C.14: ML_2 CSR results with statistical and systematic errors for R_{pW}.

C.3 Standard Asymmetry Ratio Results

This section details the asymmetry ratio results using the STD constraints for the free proton and the carbon, iron, and tungsten bound proton.

C.3.1 Free Proton Results

Tables C.15 and C.16 contain the asymmetry ratio values with errors and the covariance matrix, respectively, using the CTEQ14 set and Tables C.17 and C.18 have the corresponding values and covariance matrix using the nCTEQ15 set, respectively.

Ave. x_T	$\bar{d}/\bar{u}$	σ^+_{stat}	σ^-_{stat}	σ^+_{sys}	σ^-_{sys}
0.1474	1.437	0.086	0.092	0.101	0.100
0.1782	1.340	0.089	0.090	0.062	0.060
0.2163	1.477	0.095	0.097	0.101	0.101
0.2628	1.484	0.123	0.122	0.084	0.084
0.3156	1.650	0.155	0.151	0.102	0.102
0.3844	1.599	0.229	0.227	0.145	0.145

Table C.15: STD $\bar{d}/\bar{u}$ results for the free proton using the CTEQ14 PDF set along with the statistical and systematic errors.

Bin	0	1	2	3	4	5
Positive Shift						
0	0.086	0.000	0.000	0.000	0.000	0.000
1	0.016	0.089	0.000	0.000	0.000	0.001
2	0.001	0.014	0.095	0.000	0.000	0.001
3	0.002	0.004	0.014	0.123	0.000	0.001
4	0.000	0.001	0.003	0.014	0.155	0.002
5	0.001	0.002	0.007	0.010	0.002	0.229
Negative Shift						
0	0.092	0.000	0.000	0.000	0.000	0.001
1	0.018	0.090	0.000	0.000	0.000	0.001
2	0.004	0.010	0.097	0.000	0.000	0.001
3	0.002	0.004	0.017	0.122	0.000	0.001
4	0.000	0.000	0.004	0.014	0.151	0.001
5	0.000	0.003	0.008	0.013	0.003	0.227

Table C.16: STD $\bar{d}/\bar{u}$ statistical error matrix for the free proton calculated from the CTEQ14 PDF set.

Ave. x_T	$\bar{d}/\bar{u}$	σ^+_{stat}	σ^-_{stat}	σ^+_{sys}	σ^-_{sys}
0.1474	1.442	0.092	0.094	0.102	0.101
0.1782	1.340	0.091	0.092	0.062	0.061
0.2163	1.470	0.097	0.097	0.103	0.102
0.2628	1.466	0.123	0.122	0.085	0.084
0.3156	1.617	0.151	0.148	0.101	0.101
0.3844	1.558	0.226	0.223	0.142	0.142

Table C.17: STD $\bar{d}/\bar{u}$ results for the free proton using the nCTEQ15 nPDF set along with the statistical and systematic errors.

Bin	0	1	2	3	4	5
Positive Shift						
0	0.092	0.000	0.000	0.000	0.000	0.000
1	0.016	0.091	0.000	0.000	0.000	0.000
2	0.003	0.013	0.097	0.000	0.000	0.000
3	0.002	0.003	0.015	0.123	0.000	0.000
4	0.000	0.000	0.003	0.014	0.151	0.002
5	0.001	0.002	0.005	0.007	0.002	0.226
Negative Shift						
0	0.094	0.000	0.000	0.000	0.000	0.000
1	0.018	0.092	0.000	0.000	0.000	0.000
2	0.004	0.010	0.097	0.000	0.000	0.000
3	0.002	0.004	0.017	0.122	0.000	0.000
4	0.000	0.001	0.004	0.014	0.148	0.001
5	0.001	0.001	0.006	0.009	0.003	0.223

Table C.18: STD $\bar{d}/\bar{u}$ statistical error matrix for the free proton calculated from the nCTEQ15 nPDF set.

C.3.2 Carbon Results

Tables C.19 and C.20 contain the asymmetry ratio values with errors and the covariance matrix, respectively using the EPPS16 set and Tables C.21 and C.22 have the corresponding values and covariance matrix for using the nCTEQ15 set, respectively.

Ave. x_T	$\bar{d}/\bar{u}$	σ^+_{stat}	σ^-_{stat}	σ^+_{sys}	σ^-_{sys}
0.1474	1.454	0.117	0.110	0.122	0.109
0.1782	1.288	0.112	0.111	0.101	0.098
0.2163	1.556	0.123	0.122	0.059	0.059
0.2629	1.549	0.156	0.154	0.074	0.075
0.3157	1.637	0.203	0.197	0.130	0.130
0.3840	1.634	0.358	0.339	0.253	0.244

Table C.19: STD $\bar{d}/\bar{u}$ results for carbon using the EPPS16 nPDF set along with the statistical and systematic errors.

Bin	0	1	2	3	4	5
Positive Shift						
0	0.117	0.000	0.000	0.000	0.000	0.000
1	0.027	0.112	0.000	0.000	0.000	0.001
2	0.012	0.013	0.123	0.000	0.000	0.001
3	0.002	0.003	0.017	0.156	0.000	0.001
4	0.000	0.000	0.005	0.018	0.203	0.001
5	0.002	0.004	0.010	0.009	0.013	0.358
Negative Shift						
0	0.110	0.000	0.000	0.000	0.000	0.000
1	0.021	0.111	0.000	0.000	0.000	0.001
2	0.001	0.012	0.122	0.000	0.000	0.001
3	0.004	0.004	0.019	0.154	0.000	0.001
4	0.000	0.002	0.005	0.020	0.197	0.001
5	0.002	0.003	0.012	0.010	0.017	0.339

Table C.20: STD $\bar{d}/\bar{u}$ statistical error matrix for carbon calculated from the EPPS16 nPDF set.

Ave. x_T	$\bar{d}/\bar{u}$	σ^+_{stat}	σ^-_{stat}	σ^+_{sys}	σ^-_{sys}
0.1474	1.463	0.114	0.118	0.118	0.118
0.1782	1.282	0.116	0.102	0.105	0.097
0.2163	1.538	0.121	0.121	0.063	0.061
0.2629	1.503	0.150	0.149	0.077	0.078
0.3157	1.541	0.184	0.181	0.128	0.127
0.3840	1.465	0.323	0.306	0.229	0.221

Table C.21: STD $\bar{d}/\bar{u}$ results for carbon using the nCTEQ15 nPDF set along with the statistical and systematic errors.

Bin	0	1	2	3	4	5
Positive Shift						
0	0.114	0.000	0.000	0.000	0.000	0.000
1	0.019	0.116	0.000	0.000	0.000	0.001
2	0.002	0.016	0.121	0.000	0.000	0.001
3	0.002	0.005	0.018	0.150	0.000	0.001
4	0.001	0.002	0.005	0.017	0.184	0.001
5	0.001	0.001	0.004	0.002	0.023	0.323
Negative Shift						
0	0.118	0.000	0.000	0.000	0.000	0.000
1	0.023	0.102	0.000	0.000	0.000	0.001
2	0.004	0.006	0.121	0.000	0.000	0.001
3	0.002	0.003	0.019	0.149	0.000	0.001
4	0.000	0.001	0.004	0.019	0.181	0.001
5	0.001	0.003	0.005	0.001	0.031	0.306

Table C.22: STD $\bar{d}/\bar{u}$ statistical error matrix for carbon calculated from the nCTEQ15 nPDF set.

C.3.3 Iron Results

Tables C.23 and C.24 contain the asymmetry ratio values with errors and the covariance matrix, respectively using the EPPS16 set and Tables C.25 and C.26 have the corresponding values and covariance matrix for using the nCTEQ15 set, respectively.

Ave. x_T	$\bar{d}/\bar{u}$	σ^+_{stat}	σ^-_{stat}	σ^+_{sys}	σ^-_{sys}
0.1476	1.257	0.100	0.105	0.142	0.146
0.1784	1.314	0.099	0.105	0.098	0.100
0.2163	1.546	0.109	0.110	0.094	0.093
0.2629	1.518	0.142	0.140	0.140	0.140
0.3159	1.593	0.177	0.174	0.228	0.227
0.3844	1.401	0.288	0.276	0.407	0.380

Table C.23: STD $\bar{d}/\bar{u}$ results for iron using the EPPS16 nPDF set along with the statistical and systematic errors.

Bin	0	1	2	3	4	5
Positive Shift						
0	0.100	0.000	0.000	0.000	0.000	0.001
1	0.014	0.099	0.000	0.000	0.000	0.001
2	0.003	0.008	0.109	0.000	0.000	0.001
3	0.002	0.003	0.016	0.142	0.000	0.001
4	0.001	0.001	0.004	0.016	0.177	0.001
5	0.002	0.003	0.010	0.008	0.013	0.288
Negative Shift						
0	0.105	0.000	0.000	0.000	0.000	0.001
1	0.020	0.105	0.000	0.000	0.000	0.001
2	0.005	0.014	0.110	0.000	0.000	0.001
3	0.002	0.004	0.017	0.140	0.000	0.001
4	0.001	0.001	0.004	0.017	0.174	0.001
5	0.008	0.006	0.011	0.010	0.017	0.276

Table C.24: STD $\bar{d}/\bar{u}$ statistical error matrix for iron calculated from the EPPS16 nPDF set.

Ave. x_T	$\bar{d}/\bar{u}$	σ^+_{stat}	σ^-_{stat}	σ^+_{sys}	σ^-_{sys}
0.1476	1.254	0.100	0.106	0.145	0.150
0.1784	1.306	0.099	0.103	0.104	0.104
0.2163	1.520	0.106	0.107	0.100	0.100
0.2629	1.461	0.135	0.133	0.144	0.144
0.3159	1.481	0.157	0.157	0.220	0.219
0.3844	1.236	0.256	0.245	0.361	0.338

Table C.25: STD $\bar{d}/\bar{u}$ results for iron using the nCTEQ15 nPDF set along with the statistical and systematic errors.

Bin	0	1	2	3	4	5
Positive Shift						
0	0.100	0.000	0.000	0.000	0.000	0.000
1	0.013	0.099	0.000	0.000	0.000	0.000
2	0.005	0.009	0.106	0.000	0.000	0.000
3	0.001	0.004	0.016	0.135	0.000	0.000
4	0.000	0.001	0.004	0.015	0.157	0.000
5	0.001	0.002	0.004	0.000	0.022	0.256
Negative Shift						
0	0.106	0.000	0.000	0.000	0.000	0.000
1	0.019	0.103	0.000	0.000	0.000	0.000
2	0.004	0.013	0.107	0.000	0.000	0.000
3	0.002	0.004	0.018	0.133	0.000	0.000
4	0.001	0.001	0.004	0.016	0.157	0.000
5	0.000	0.002	0.004	0.001	0.028	0.245

Table C.26: STD $\bar{d}/\bar{u}$ statistical error matrix for iron calculated from the nCTEQ15 nPDF set.

C.3.4 Tungsten Results

Tables C.27 and C.28 contain the asymmetry ratio values with errors and the covariance matrix, respectively using the EPPS16 set and Tables C.29 and C.30 have the corresponding values and covariance matrix for using the nCTEQ15 set, respectively.

Ave. x_T	$\bar{d}/\bar{u}$	σ^+_{stat}	σ^-_{stat}	σ^+_{sys}	σ^-_{sys}
0.1476	1.053	0.080	0.084	0.147	0.149
0.1783	1.070	0.084	0.080	0.125	0.124
0.2162	1.323	0.089	0.089	0.046	0.047
0.2628	1.292	0.114	0.112	0.068	0.068
0.3159	1.377	0.145	0.143	0.125	0.126
0.3836	1.426	0.260	0.250	0.262	0.252

Table C.27: STD $\bar{d}/\bar{u}$ results for tungsten using the EPPS16 nPDF set along with the statistical and systematic errors.

Bin	0	1	2	3	4	5
Positive Shift						
0	0.080	0.000	0.000	0.000	0.000	0.000
1	0.013	0.084	0.000	0.000	0.000	0.000
2	0.006	0.010	0.089	0.000	0.000	0.001
3	0.001	0.002	0.012	0.114	0.000	0.001
4	0.000	0.000	0.003	0.013	0.145	0.001
5	0.001	0.002	0.007	0.006	0.010	0.260
Negative Shift						
0	0.084	0.000	0.000	0.000	0.000	0.000
1	0.015	0.080	0.000	0.000	0.000	0.000
2	0.003	0.006	0.089	0.000	0.000	0.001
3	0.001	0.003	0.014	0.112	0.000	0.001
4	0.000	0.001	0.004	0.014	0.143	0.000
5	0.001	0.002	0.008	0.006	0.013	0.250

Table C.28: STD $\bar{d}/\bar{u}$ statistical error matrix for tungsten calculated from the EPPS16 nPDF set.

Ave. x_T	$\bar{d}/\bar{u}$	σ^+_{stat}	σ^-_{stat}	σ^+_{sys}	σ^-_{sys}
0.1476	1.048	0.080	0.084	0.148	0.146
0.1783	1.061	0.084	0.079	0.129	0.124
0.2162	1.304	0.089	0.088	0.050	0.050
0.2628	1.253	0.110	0.109	0.071	0.071
0.3159	1.297	0.136	0.134	0.126	0.126
0.3836	1.277	0.244	0.233	0.246	0.235

Table C.29: STD $\bar{d}/\bar{u}$ results for tungsten using the nCTEQ15 nPDF set along with the statistical and systematic errors.

Bin	0	1	2	3	4	5
Positive Shift						
0	0.080	0.000	0.000	0.000	0.000	0.000
1	0.016	0.084	0.000	0.000	0.000	0.000
2	0.005	0.012	0.089	0.000	0.000	0.000
3	0.001	0.004	0.013	0.111	0.000	0.000
4	0.000	0.001	0.003	0.013	0.136	0.001
5	0.000	0.001	0.002	0.001	0.017	0.244
Negative Shift						
0	0.084	0.000	0.000	0.000	0.000	0.000
1	0.013	0.079	0.000	0.000	0.000	0.001
2	0.002	0.006	0.088	0.000	0.000	0.000
3	0.001	0.002	0.014	0.109	0.000	0.000
4	0.000	0.000	0.003	0.013	0.134	0.000
5	0.000	0.001	0.003	0.001	0.020	0.233

Table C.30: STD $\bar{d}/\bar{u}$ statistical error matrix for tungsten calculated from the nCTEQ15 nPDF set.

C.4 Machine Learning Asymmetry Ratio Results

This section details the asymmetry ratio results using the ML_2 constraints for the free proton and the carbon, iron, and tungsten bound proton.

C.4.1 Free Proton Results

Tables C.31 and C.32 contain the asymmetry ratio values with errors and the covariance matrix, respectively for the ML_2 constraints using the CTEQ14 set and Tables C.33 and C.34 have the corresponding values and covariance matrix for using the nCTEQ15 set, respectively.

Ave. x_T	$\bar{d}/\bar{u}$	σ^+_{stat}	σ^-_{stat}	σ^+_{sys}	σ^-_{sys}
0.1474	1.446	0.067	0.066	0.140	0.142
0.1782	1.370	0.063	0.069	0.152	0.158
0.2165	1.618	0.075	0.074	0.213	0.211
0.2633	1.639	0.094	0.094	0.250	0.248
0.3156	1.679	0.116	0.114	0.275	0.274
0.3836	1.495	0.164	0.164	0.480	0.464

Table C.31: ML_2 $\bar{d}/\bar{u}$ results for the free proton using the CTEQ14 PDF set along with the statistical and systematic errors.

Bin	0	1	2	3	4	5
Positive Shift						
0	0.067	0.000	0.000	0.000	0.000	0.001
1	0.012	0.063	0.000	0.000	0.000	0.001
2	0.002	0.002	0.075	0.000	0.001	0.001
3	0.001	0.002	0.011	0.094	0.001	0.001
4	0.000	0.000	0.003	0.012	0.116	0.002
5	0.001	0.003	0.006	0.010	0.002	0.163
Negative Shift						
0	0.066	0.000	0.000	0.000	0.000	0.001
1	0.012	0.069	0.000	0.000	0.000	0.001
2	0.005	0.011	0.074	0.000	0.001	0.001
3	0.002	0.005	0.012	0.094	0.001	0.001
4	0.000	0.002	0.003	0.012	0.114	0.001
5	0.002	0.007	0.008	0.012	0.002	0.164

Table C.32: ML_2 $\bar{d}/\bar{u}$ statistical error matrix for the free proton calculated from the CTEQ14 PDF set.

Ave. x_T	$\bar{d}/\bar{u}$	σ^+_{stat}	σ^-_{stat}	σ^+_{sys}	σ^-_{sys}
0.1474	1.447	0.068	0.063	0.148	0.141
0.1782	1.367	0.068	0.067	0.162	0.159
0.2165	1.613	0.076	0.076	0.223	0.219
0.2633	1.618	0.094	0.091	0.258	0.253
0.3165	1.644	0.113	0.111	0.278	0.278
0.3869	1.463	0.163	0.161	0.478	0.460

Table C.33: ML_2 $\bar{d}/\bar{u}$ results for the free proton using the nCTEQ15 nPDF set along with the statistical and systematic errors.

Bin	0	1	2	3	4	5
Positive Shift						
0	0.068	0.000	0.000	0.000	0.000	0.000
1	0.016	0.068	0.000	0.000	0.000	0.000
2	0.006	0.008	0.076	0.000	0.000	0.000
3	0.002	0.003	0.010	0.094	0.000	0.000
4	0.001	0.001	0.003	0.014	0.113	0.001
5	0.000	0.001	0.004	0.007	0.000	0.163
Negative Shift						
0	0.063	0.000	0.000	0.000	0.000	0.000
1	0.009	0.067	0.000	0.000	0.000	0.000
2	0.001	0.007	0.076	0.000	0.000	0.000
3	0.000	0.000	0.011	0.091	0.000	0.000
4	0.000	0.001	0.003	0.011	0.111	0.001
5	0.001	0.001	0.005	0.008	0.002	0.161

Table C.34: ML_2 $\bar{d}/\bar{u}$ statistical error matrix for the free proton calculated from the nCTEQ15 nPDF set.

C.4.2 Carbon Results

Tables C.35 and C.36 contain the asymmetry ratio values with errors and the covariance matrix, respectively using the EPPS16 set and Tables C.37 and C.38 have the corresponding values and covariance matrix for using the nCTEQ15 set, respectively.

Ave. x_T	$\bar{d}/\bar{u}$	σ^+_{stat}	σ^-_{stat}	σ^+_{sys}	σ^-_{sys}
0.1475	1.457	0.084	0.086	0.134	0.136
0.1782	1.310	0.080	0.087	0.087	0.092
0.2165	1.713	0.098	0.098	0.196	0.197
0.2634	1.760	0.120	0.122	0.275	0.272
0.3167	1.641	0.151	0.147	0.273	0.269
0.3868	1.277	0.243	0.232	0.465	0.420

Table C.35: ML_2 $\bar{d}/\bar{u}$ results for carbon using the EPPS16 nPDF set along with the statistical and systematic errors.

Bin	0	1	2	3	4	5
Positive Shift						
0	0.084	0.000	0.000	0.000	0.000	0.001
1	0.014	0.080	0.000	0.000	0.000	0.001
2	0.001	0.008	0.098	0.000	0.000	0.001
3	0.001	0.002	0.013	0.120	0.000	0.001
4	0.001	0.001	0.004	0.017	0.151	0.001
5	0.002	0.004	0.010	0.010	0.015	0.243
Negative Shift						
0	0.086	0.000	0.000	0.000	0.000	0.001
1	0.020	0.087	0.000	0.000	0.000	0.001
2	0.004	0.012	0.098	0.000	0.000	0.001
3	0.002	0.006	0.014	0.122	0.000	0.001
4	0.001	0.001	0.004	0.018	0.147	0.001
5	0.002	0.002	0.012	0.012	0.019	0.232

Table C.36: ML_2 $\bar{d}/\bar{u}$ statistical error matrix for carbon calculated from the EPPS16 nPDF set.

Ave. x_T	$\bar{d}/\bar{u}$	σ^+_{stat}	σ^-_{stat}	σ^+_{sys}	σ^-_{sys}
0.1475	1.457	0.083	0.087	0.133	0.139
0.1782	1.301	0.083	0.087	0.093	0.095
0.2165	1.685	0.096	0.096	0.208	0.206
0.2634	1.691	0.116	0.115	0.278	0.275
0.3167	1.520	0.137	0.131	0.265	0.257
0.3868	1.124	0.217	0.207	0.414	0.375

Table C.37: ML_2 $\bar{d}/\bar{u}$ results for carbon using the nCTEQ15 nPDF set along with the statistical and systematic errors.

Bin	0	1	2	3	4	5
Positive Shift						
0	0.083	0.000	0.000	0.000	0.001	0.000
1	0.014	0.083	0.000	0.000	0.001	0.000
2	0.004	0.009	0.096	0.000	0.001	0.000
3	0.002	0.004	0.013	0.116	0.001	0.000
4	0.001	0.001	0.004	0.016	0.137	0.000
5	0.002	0.002	0.004	0.002	0.025	0.217
Negative Shift						
0	0.087	0.000	0.000	0.000	0.001	0.000
1	0.018	0.087	0.000	0.000	0.001	0.000
2	0.003	0.009	0.096	0.000	0.001	0.000
3	0.001	0.004	0.014	0.115	0.001	0.000
4	0.002	0.001	0.004	0.016	0.131	0.000
5	0.001	0.001	0.005	0.001	0.027	0.207

Table C.38: ML_2 $\bar{d}/\bar{u}$ statistical error matrix for carbon calculated from the nCTEQ15 nPDF set.

C.4.3 Iron Results

Tables C.39 and C.40 contain the asymmetry ratio values with errors and the covariance matrix, respectively using the EPPS16 set and Tables C.41 and C.42 have the corresponding values and covariance matrix for using the nCTEQ15 set, respectively.

Ave. x_T	$\bar{d}/\bar{u}$	σ^+_{stat}	σ^-_{stat}	σ^+_{sys}	σ^-_{sys}
0.1476	1.261	0.068	0.070	0.131	0.131
0.1784	1.273	0.069	0.069	0.090	0.090
0.2165	1.675	0.082	0.081	0.099	0.097
0.2634	1.657	0.103	0.102	0.118	0.118
0.3167	1.599	0.126	0.123	0.205	0.215
0.3872	1.060	0.183	0.175	0.374	0.341

Table C.39: ML_2 $\bar{d}/\bar{u}$ results for iron using the EPPS16 nPDF set along with the statistical and systematic errors.

Bin	0	1	2	3	4	5
Positive Shift						
0	0.068	0.000	0.000	0.000	0.000	0.000
1	0.013	0.069	0.000	0.000	0.000	0.000
2	0.004	0.010	0.082	0.000	0.000	0.000
3	0.002	0.005	0.012	0.103	0.000	0.001
4	0.000	0.000	0.003	0.013	0.126	0.001
5	0.001	0.003	0.009	0.008	0.014	0.183
Negative Shift						
0	0.070	0.000	0.000	0.000	0.000	0.000
1	0.011	0.069	0.000	0.000	0.000	0.000
2	0.003	0.010	0.081	0.000	0.000	0.000
3	0.001	0.002	0.013	0.102	0.000	0.001
4	0.001	0.000	0.003	0.014	0.123	0.001
5	0.003	0.003	0.011	0.010	0.017	0.175

Table C.40: ML_2 $\bar{d}/\bar{u}$ statistical error matrix for iron calculated from the EPPS16 nPDF set.

Ave. x_T	$\bar{d}/\bar{u}$	σ^+_{stat}	σ^-_{stat}	σ^+_{sys}	σ^-_{sys}
0.1476	1.259	0.070	0.072	0.135	0.135
0.1784	1.262	0.068	0.069	0.106	0.094
0.2165	1.636	0.080	0.079	0.108	0.107
0.2634	1.577	0.092	0.096	0.125	0.127
0.3167	1.460	0.109	0.108	0.198	0.202
0.3872	0.9172	0.160	0.154	0.327	0.300

Table C.41: ML_2 $\bar{d}/\bar{u}$ results for iron using the nCTEQ15 nPDF set along with the statistical and systematic errors.

Bin	0	1	2	3	4	5
Positive Shift						
0	0.070	0.000	0.000	0.000	0.000	0.000
1	0.013	0.068	0.000	0.000	0.000	0.000
2	0.003	0.009	0.080	0.000	0.000	0.000
3	0.001	0.003	0.012	0.092	0.000	0.000
4	0.000	0.001	0.003	0.012	0.109	0.000
5	0.000	0.000	0.004	0.000	0.021	0.160
Negative Shift						
0	0.072	0.000	0.000	0.000	0.000	0.000
1	0.013	0.069	0.000	0.000	0.000	0.000
2	0.003	0.008	0.079	0.000	0.000	0.000
3	0.001	0.002	0.012	0.096	0.000	0.000
4	0.000	0.000	0.003	0.013	0.108	0.000
5	0.001	0.002	0.004	0.001	0.016	0.154

Table C.42: ML_2 $\bar{d}/\bar{u}$ statistical error matrix for iron calculated from the nCTEQ15 nPDF set.

C.4.4 Tungsten Results

Tables C.43 and C.44 contain the asymmetry ratio values with errors and the covariance matrix, respectively using the EPPS16 set and Tables C.45 and C.46 have the corresponding values and covariance matrix for using the nCTEQ15 set, respectively.

Ave. x_T	$\bar{d}/\bar{u}$	σ^+_{stat}	σ^-_{stat}	σ^+_{sys}	σ^-_{sys}
0.1476	1.090	0.056	0.059	0.124	0.126
0.1782	1.121	0.058	0.055	0.090	0.087
0.2164	1.442	0.067	0.068	0.103	0.103
0.2632	1.482	0.083	0.083	0.218	0.215
0.3166	1.266	0.101	0.096	0.289	0.281
0.3863	0.9515	0.154	0.148	0.454	0.403

Table C.43: ML$_2$ $\bar{d}/\bar{u}$ results for tungsten using the EPPS16 nPDF set along with the statistical and systematic errors.

Bin	0	1	2	3	4	5
Positive Shift						
0	0.056	0.000	0.000	0.000	0.000	0.000
1	0.008	0.058	0.000	0.000	0.000	0.000
2	0.002	0.008	0.067	0.000	0.000	0.000
3	0.001	0.004	0.008	0.083	0.000	0.000
4	0.000	0.001	0.002	0.012	0.101	0.001
5	0.001	0.003	0.007	0.006	0.014	0.154
Negative Shift						
0	0.059	0.000	0.000	0.000	0.000	0.000
1	0.009	0.055	0.000	0.000	0.000	0.000
2	0.003	0.007	0.068	0.000	0.000	0.000
3	0.001	0.002	0.010	0.083	0.000	0.000
4	0.000	0.001	0.003	0.013	0.096	0.000
5	0.001	0.003	0.007	0.008	0.011	0.148

Table C.44: ML$_2$ $\bar{d}/\bar{u}$ statistical error matrix for tungsten calculated from the EPPS16 nPDF set.

Ave. x_T	$\bar{d}/\bar{u}$	σ^+_{stat}	σ^-_{stat}	σ^+_{sys}	σ^-_{sys}
0.1476	1.084	0.056	0.059	0.126	0.128
0.1782	1.110	0.057	0.057	0.094	0.092
0.2164	1.411	0.066	0.066	0.112	0.112
0.2632	1.419	0.081	0.079	0.225	0.220
0.3166	1.172	0.090	0.089	0.282	0.278
0.3863	0.8212	0.141	0.136	0.417	0.370

Table C.45: ML_2 $\bar{d}/\bar{u}$ results for tungsten using the nCTEQ15 nPDF set along with the statistical and systematic errors.

Bin	0	1	2	3	4	5
Positive Shift						
0	0.056	0.000	0.000	0.000	0.000	0.000
1	0.008	0.057	0.000	0.000	0.000	0.000
2	0.000	0.011	0.066	0.000	0.000	0.000
3	0.002	0.006	0.010	0.081	0.000	0.000
4	0.000	0.001	0.003	0.009	0.090	0.000
5	0.000	0.000	0.003	0.001	0.016	0.141
Negative Shift						
0	0.059	0.000	0.000	0.000	0.000	0.000
1	0.010	0.057	0.000	0.000	0.000	0.000
2	0.003	0.008	0.066	0.000	0.000	0.000
3	0.001	0.003	0.010	0.079	0.000	0.000
4	0.000	0.000	0.002	0.010	0.089	0.000
5	0.000	0.001	0.003	0.003	0.018	0.136

Table C.46: ML_2 $\bar{d}/\bar{u}$ statistical error matrix for tungsten calculated from the nCTEQ15 nPDF set.

Appendix D:

Event Comparison

This appendix is a listing of the STD and ML_2 yield differences per I_T bin per x_T bin. The yields of the STD and all of the ML models are listed in Section 3.3. The tables were constructed by splitting the STD dimuon events into their separate kTracks, which effectively gives each of the two tracks that were joined to make a dimuon a weight of 0.5. This was already done with the ML models. The set operations between the two constraints were then performed and the resulting events were divided by two to approximate a yield composed of dimuons. Due to the ML models being developed using kTracks, not all of the tracks have their opposite charged pair. This results in some of the bins in the tables having half integer event values, since there can be an odd number of kTracks in a bin before the halving is done.

D.1 Yield Intersection

Tables D.1 - D.6 list the dimuon yields for the events that pass the STD and the ML_2 constraints.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8{,}000)$	401.5	7	496.5	7.5	186.5	175.5	211.5
$[8{,}000-16{,}000)$	490	17.5	590	7.5	256.5	226	265
$[16{,}000-24{,}000)$	444	7.5	533	5.5	206	195.5	264
$[24{,}000-32{,}000)$	361.5	10	469.5	7	135	163.5	224
$[32{,}000-4{,}0000)$	245.5	4.5	306.5	5	145.5	136.5	155.5
$[4{,}0000-48{,}000)$	197.5	7	232.5	5	100	90	100
$[48{,}000-56{,}000)$	116	3	156	2	62	53.5	103.5
$[56{,}000-64{,}000)$	106.5	2	119.5	1	31	46	46.5
$[0-64{,}000)$	2362.5	58.5	2903.5	40.5	1122.5	1086.5	1370

Table D.1: Dimuon yields for events that are in intersection of STD and ML_2 constraints detailed in yield per I_T bin for $x_T \in [0.13 - 0.16)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8{,}000)$	810.5	12	1033.5	13	400	357	462
$[8{,}000-16{,}000)$	1107.5	29	1277	14.5	528.5	466	569.5
$[16{,}000-24{,}000)$	900	23	1125.5	9.5	455	400	552
$[24{,}000-32{,}000)$	788.5	10.5	889.5	13.5	349	315	464.5
$[32{,}000-4{,}0000)$	557.5	21	622	15.5	291	231.5	332.5
$[4{,}0000-48{,}000)$	415	7.5	444.5	6	170.5	149.5	228
$[48{,}000-56{,}000)$	309	6.5	303	3	136	122.5	168
$[56{,}000-64{,}000)$	202	8	208	6	78.5	75	116
$[0-64{,}000)$	5090	117.5	5903	81	2408.5	2116.5	2892.5

Table D.2: Dimuon yields for events that are in intersection of STD and ML_2 constraints detailed in yield per I_T bin for $x_T \in [0.16 - 0.195)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8{,}000)$	900.5	21	1151.5	14	473.5	420.5	549
$[8{,}000-16{,}000)$	1165.5	37.5	1541	17	632	507	779
$[16{,}000-24{,}000)$	1097	30	1388	15.5	553	505	690
$[24{,}000-32{,}000)$	927	35	1049	20	448.5	404	532
$[32{,}000-4{,}0000)$	672.5	21.5	725.5	16.5	329	311.5	434
$[4{,}0000-48{,}000)$	469	14.5	563	9.5	250.5	221	308
$[48{,}000-56{,}000)$	341	12	368	6	160.5	140	195.5
$[56{,}000-64{,}000)$	264	7.5	235.5	6	96	114	128.5
$[0-64{,}000)$	5836.5	179	7021.5	104.5	2943	2623	3616

Table D.3: Dimuon yields for events that are in intersection of STD and ML_2 constraints detailed in yield per I_T bin for $x_T \in [0.195 - 0.24)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8{,}000)$	638	15.5	819.5	15.5	363	278.5	398
$[8{,}000-16{,}000)$	837.5	25	1017	9	386.5	356.5	496
$[16{,}000-24{,}000)$	771	17	894.5	12.5	382	362.5	482
$[24{,}000-32{,}000)$	658.5	26	717.5	15	323	287	399
$[32{,}000-4{,}0000)$	495	14.5	509.5	10	227	200	293.5
$[4{,}0000-48{,}000)$	353.5	15	374.5	11	148	146.5	179.5
$[48{,}000-56{,}000)$	255.5	15	268.5	2.5	106.5	92	150
$[56{,}000-64{,}000)$	179	10.5	192.5	8.5	76	76.5	97.5
$[0-64{,}000)$	4188	138.5	4793.5	84	2012	1799.5	2495.5

Table D.4: Dimuon yields for events that are in intersection of STD and ML_2 constraints detailed in yield per I_T bin for $x_T \in [0.24 - 0.29)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8{,}000)$	396	6	475	4	185	167	227
$[8{,}000-16{,}000)$	501	7.5	641.5	9	273.5	216.5	310
$[16{,}000-24{,}000)$	451	10	558.5	10.5	224	210	274
$[24{,}000-32{,}000)$	370	11.5	479.5	6	185.5	154.5	217
$[32{,}000-4{,}0000)$	258.5	9.5	319.5	8.5	136	127.5	167
$[4{,}0000-48{,}000)$	203	7	225.5	5	88	91	132.5
$[48{,}000-56{,}000)$	161.5	10.5	151.5	3.5	70	65.5	79
$[56{,}000-64{,}000)$	121.5	9	110	6	43	41.5	61.5
$[0-64{,}000)$	2462.5	71	2961	52.5	1205	1073.5	1468

Table D.5: Dimuon yields for events that are in intersection of STD and ML_2 constraints detailed in yield per I_T bin for $x_T \in [0.29 - 0.35)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8{,}000)$	198	7	239.5	2	94.5	87	113.5
$[8{,}000-16{,}000)$	277.5	6.5	295	3	94.5	120.5	151
$[16{,}000-24{,}000)$	236.5	4	269	4	118	92.5	127
$[24{,}000-32{,}000)$	169.5	6	219.5	5.5	86	71.5	117.5
$[32{,}000-4{,}0000)$	139	6.5	127	2	61	61	67.5
$[4{,}0000-48{,}000)$	100	9	105	7.5	41	33	50
$[48{,}000-56{,}000)$	77	3.5	76.5	8	27	27	31
$[56{,}000-64{,}000)$	56	4	50	4	22.5	15	33
$[0-64{,}000)$	1253.5	46.5	1381.5	36	544.5	507.5	690.5

Table D.6: Dimuon yields for events that are in intersection of STD and ML_2 constraints detailed in yield per I_T bin for $x_T \in [0.35 - 0.45)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8{,}000)$	3344.5	68.5	4215.5	56	1702.5	1485.5	1961
$[8{,}000-16{,}000)$	4379	123	5361.5	60	2171.5	1892.5	2570.5
$[16{,}000-24{,}000)$	3899.5	91.5	4768.5	57.5	1938	1765.5	2389
$[24{,}000-32{,}000)$	3275	99	3824.5	67	1527	1395.5	1954
$[32{,}000-4{,}0000)$	2368	77.5	2610	57.5	1189.5	1068	1450
$[4{,}0000-48{,}000)$	1738	60	1945	44	798	731	998
$[48{,}000-56{,}000)$	1260	50.5	1323.5	25	562	500.5	727
$[56{,}000-64{,}000)$	929	41	915.5	31.5	347	368	483
$[0-64{,}000)$	21193	611	24964	398.5	10235.5	9206.5	12532.5

Table D.7: Dimuon yields for events that are in intersection of STD and ML_2 constraints detailed in yield per I_T bin for $x_T \in [0.13 - 0.45)$.

D.2 Standard Analysis

Tables D.8 - D.13 lists the dimuon yields for the events that pass the STD constraints, but failed the ML_2 constraints.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	32.5	2	26.5	0.5	18.5	12.5	15.5
[8,000 − 16,000)	46	3.5	28	0.5	20.5	16	20
[16,000 − 24,000)	41	2.5	30	1.5	14	13.5	15
[24,000 − 32,000)	45.5	4	26.5	0	17	9.5	16
[32,000 − 40,000)	27.5	2.5	26.5	4	10.5	14.5	12.5
[40,000 − 48,000)	15.5	0	14.5	0	9	8	8
[48,000 − 56,000)	12	3	10	0	7	5.5	9.5
[56,000 − 64,000)	15.5	1	3.5	0	5	4	3.5
[0 − 64,000)	235.5	18.5	165.5	6.5	101.5	83.5	

Table D.8: Dimuon yields for events that passed the STD constraints, but failed the ML_2 constraints detailed in yield per I_T bin for $x_T \in [0.13 - 0.16)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	63.5	4	70.5	2	30	30	40
[8,000 − 16,000)	113.5	5	77	2.5	53.5	37	55.5
[16,000 − 24,000)	107	7	81.5	3.5	50	41	53
[24,000 − 32,000)	101.5	1.5	61.5	1.5	42	42	43.5
[32,000 − 40,000)	52.5	1	47	2.5	33	16.5	43.5
[40,000 − 48,000)	43	2.5	35.5	3	22.5	19.5	27
[48,000 − 56,000)	43	0.5	29	0	22	19.5	12
[56,000 − 64,000)	29	1	25	0	16.5	9	12
[0 − 64,000)	553	22.5	427	15	269.5	214.5	286.5

Table D.9: Dimuon yields for events that passed the STD constraints, but failed the ML_2 constraints detailed in yield per I_T bin for $x_T \in [0.16 - 0.195)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8{,}000)$	120.5	0	74.5	5	36.5	31.5	51
$[8{,}000-16{,}000)$	135.5	4.5	113	4	52	47	93
$[16{,}000-24{,}000)$	161	9	120	5.5	87.5	56	87
$[24{,}000-32{,}000)$	143	11	98	6	59.5	52	81
$[32{,}000-40{,}000)$	125.5	4.5	81.5	10.5	54	44.5	57
$[40{,}000-48{,}000)$	93	6.5	66	4.5	39.5	36	42
$[48{,}000-56{,}000)$	65	2	40	2	21	22	25.5
$[56{,}000-64{,}000)$	45	1.5	36.5	1	13	16	19.5
$[0-64{,}000)$	888.5	39	629.5	38.5	363	305	456

Table D.10: Dimuon yields for events that passed the STD constraints, but failed the ML_2 constraints detailed in yield per I_T bin for $x_T \in [0.195 - 0.24)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8{,}000)$	67	3.5	59.5	3.5	32	16.5	31
$[8{,}000-16{,}000)$	107.5	4	82	1	40.5	31.5	44.5
$[16{,}000-24{,}000)$	111	3	82.5	3.5	39	41.5	51
$[24{,}000-32{,}000)$	92.5	8	81.5	5	41	37	40
$[32{,}000-40{,}000)$	69	4.5	64.5	5	28	32	33.5
$[40{,}000-48{,}000)$	84.5	4	41.5	3	20	22.5	28.5
$[48{,}000-56{,}000)$	44.5	4	39.5	2.5	14.5	14	18
$[56{,}000-64{,}000)$	39	3.5	33.5	4.5	18	13.5	15.5
$[0-64{,}000)$	615	34.5	484.5	28	233	208.5	262

Table D.11: Dimuon yields for events that passed the STD constraints, but failed the ML_2 constraints detailed in yield per I_T bin for $x_T \in [0.24 - 0.29)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8{,}000)$	31	1	24	0	10	11	17
$[8{,}000-16{,}000)$	49	2.5	29.5	0	14.5	23.5	24
$[16{,}000-24{,}000)$	43	1	37.5	2.5	21	15	31
$[24{,}000-32{,}000)$	45	5.5	28.5	2	19.5	13.5	27
$[32{,}000-40{,}000)$	36.5	2.5	21.5	0.5	15	12.5	12
$[40{,}000-48{,}000)$	33	2	15.5	0	9	7	15
$[48{,}000-56{,}000)$	33.5	1.5	13.5	1.5	8	8.5	12
$[56{,}000-64{,}000)$	16.5	1	12	2	3	4.5	5.5
$[0-64{,}000)$	287.5	17	182	8.5	100	95.5	143.5

Table D.12: Dimuon yields for events that passed the STD constraints, but failed the ML_2 constraints detailed in yield per I_T bin for $x_T \in [0.29 - 0.35)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8000)$	8	1	7.5	0	3.5	6	13.5
$[8{,}000-16{,}000)$	20.5	1.5	9	0	6.5	8.5	16
$[16{,}000-24{,}000)$	19.5	1	7	0	7	6.5	7
$[24{,}000-32{,}000)$	13.5	0	8.5	0.5	7	4.5	10.5
$[32{,}000-40{,}000)$	9	0.5	8	0	5	3	8.5
$[40{,}000-48{,}000)$	14	1	6	1.5	2	0	6
$[48{,}000-56{,}000)$	6	1.5	4.5	1	2	0	5
$[56{,}000-64{,}000)$	5	0	3	1	0.5	1	3
$[0-64{,}000)$	95.5	6.5	53.5	4	33.5	29.5	69.5

Table D.13: Dimuon yields for events that passed the STD constraints, but failed the ML_2 constraints detailed in yield per I_T bin for $x_T \in [0.35 - 0.45)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8{,}000)$	322.5	11.5	262.5	11	130.5	107.5	168
$[8{,}000-16{,}000)$	472	21	338.5	8	187.5	163.5	253
$[16{,}000-24{,}000)$	482.5	23.5	358.5	16.5	218.5	173.5	244
$[24{,}000-32{,}000)$	441	30	304.5	15	186	158.5	218
$[32{,}000-4{,}0000)$	320	15.5	249	22.5	145.5	123	167
$[4{,}0000-48{,}000)$	283	16	179	12	102	93	126.5
$[48{,}000-56{,}000)$	204	12.5	136.5	7	74.5	69.5	82
$[56{,}000-64{,}000)$	150	8	113.5	8.5	56	48	59
$[0-64{,}000)$	2675	138	1942	100.5	1100.5	936.5	1317.5

Table D.14: Dimuon yields for events that passed the STD constraints, but failed the ML_2 constraints detailed in yield per I_T bin for $x_T \in [0.13 - 0.45)$.

D.3 Machine Learning Model 2

Tables D.15 - D.20 list the dimuon yields for the events that pass the ML_2 constraints, but failed the STD constraints.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	76	6	71.5	3.5	18	28.5	30.5
[8,000 − 16,000)	117	10.5	113.5	10.5	53.5	42.5	46.5
[16,000 − 24,000)	114.5	4	111.5	11	44.5	45.5	51.5
[24,000 − 32,000)	90.5	10.5	93.5	11	44	33	41.5
[32,000 − 40,000)	86.5	4.5	77	10.5	25	28	28.5
[40,000 − 48,000)	62	7.5	59	5	24	25	24.5
[48,000 − 56,000)	55	4	39	4	12	14.5	20.5
[56,000 − 64,000)	37.5	3.5	23	5	16.5	18.5	11
[0 − 64,000)	639	50.5	588	60.5	237.5	235.5	254.5

Table D.15: Dimuon yields for events that passed the ML_2 constraints, but failed the STD constraints detailed in yield per I_T bin for $x_T \in [0.13 - 0.16)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	149	9	150	12	60	50	69.5
[8,000 − 16,000)	248	17	243	20	88.5	81	109.5
[16,000 − 24,000)	225	16	218.5	18	91.5	98.5	101.5
[24,000 − 32,000)	251	15	183.5	19	68	72.5	96.5
[32,000 − 40,000)	173.5	8	148	12	55.5	58.5	66
[40,000 − 48,000)	138.5	14.5	131	16	43	60	76
[48,000 − 56,000)	106.5	8.5	83.5	16.5	31.5	43	44.5
[56,000 − 64,000)	83.5	3	55.5	12	21.5	34	34.5
[0 − 64,000)	1375	91	1213	125.5	459.5	497.5	598

Table D.16: Dimuon yields for events that passed the ML_2 constraints, but failed the STD constraints detailed in yield per I_T bin for $x_T \in [0.16 - 0.195)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	177.5	13	172.5	9.5	74.5	69	87.5
[8,000 − 16,000)	297	23.5	277	21.5	117	114	127.5
[16,000 − 24,000)	312.5	33.5	287.5	27.5	119.5	106	123
[24,000 − 32,000)	316	25	293	21	99.5	97.5	119.5
[32,000 − 40,000)	225	22	207.5	21	77.5	84.5	93
[40,000 − 48,000)	196	24	170	16.5	75.5	66	76.5
[48,000 − 56,000)	142	11.5	121	21	47	54.5	67.5
[56,000 − 64,000)	128.5	20	102.5	19	38.5	42	40.5
[0 − 64,000)	1794.5	172.5	1631	157	649	633.5	735

Table D.17: Dimuon yields for events that passed the ML_2 constraints, but failed the STD constraints detailed in yield per I_T bin for $x_T \in [0.195 - 0.24)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	126.5	10.5	154.5	7	47	63.5	50.5
[8,000 − 16,000)	233.5	15	218	11	76.5	95.5	110
[16,000 − 24,000)	270.5	32	254	22	86	98	120
[24,000 − 32,000)	296.5	46	243.5	25.5	89.5	90.5	117.5
[32,000 − 40,000)	248.5	27.5	219.5	23	80	86	98
[40,000 − 48,000)	217	29	180.5	28	74.5	54.5	84.5
[48,000 − 56,000)	175	14	141.5	22	54	56.5	50
[56,000 − 64,000)	149.5	20.5	116	19.5	38.5	41	52.5
[0 − 64,000)	1717	194.5	1527.5	158	546	585.5	683

Table D.18: Dimuon yields for events that passed the ML_2 constraints, but failed the STD constraints detailed in yield per I_T bin for $x_T \in [0.24 - 0.29)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	103	3	97	6.5	37	34.5	39
[8,000 − 16,000)	179	16	159	18.5	77.5	68.5	81.5
[16,000 − 24,000)	246	28	213	22.5	94.5	76.5	85.5
[24,000 − 32,000)	252.5	24	227.5	28	93	92.5	78
[32,000 − 40,000)	250.5	25.5	207.5	20.5	69	87	76
[40,000 − 48,000)	227	23.5	186.5	19	58.5	81	74
[48,000 − 56,000)	201	21	145.5	32	40.5	64	58.5
[56,000 − 64,000)	169.5	23.5	122	23	39.5	61	42.5
[0 − 64,000)	1628.5	164.5	1358	170	509.5	565	535

Table D.19: Dimuon yields for events that passed the ML_2 constraints, but failed the STD constraints detailed in yield per I_T bin for $x_T \in [0.29 - 0.35)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
[0 − 8,000)	60.5	1.5	63	3	16	18	31.5
[8,000 − 16,000)	127	11	92	12.5	29.5	41	37
[16,000 − 24,000)	148.5	13	151.5	11.5	46.5	49	51.5
[24,000 − 32,000)	206.5	22	174	29	58	53	52
[32,000 − 40,000)	216.5	15.5	142	15	46	58	45
[40,000 − 48,000)	230	27.5	142	25	45.5	54.5	47.5
[48,000 − 56,000)	170	24	116.5	31	26	44	39.5
[56,000 − 64,000)	147	21.5	91.5	41.5	25.5	34	26.5
[0 − 64,000)	1306	136	972.5	168.5	293	351.5	330.5

Table D.20: Dimuon yields for events that passed the ML_2 constraints, but failed the STD constraints detailed in yield per I_T bin for $x_T \in [0.35 - 0.45)$.

Intensity Range	LH_2	EMPTY	LD_2	NONE	Fe	C	W
$[0-8{,}000)$	692.5	43	708.5	41.5	252.5	263.5	308.5
$[8{,}000-16{,}000)$	1201.5	93	1102.5	94	442.5	442.5	512
$[16{,}000-24{,}000)$	1317	126.5	1236	112.5	482.5	473.5	533
$[24{,}000-32{,}000)$	1413	142.5	1215	133.5	452	439	505
$[32{,}000-4{,}0000)$	1200.5	103	1001.5	102	353	402	406.5
$[4{,}0000-48{,}000)$	1070.5	126	869	109.5	321	341	383
$[48{,}000-56{,}000)$	849.5	83	647	126.5	211	276.5	280.5
$[56{,}000-64{,}000)$	715.5	92	510.5	120	180	230.5	207.5
$[0-64{,}000)$	8460	809	7290	839.5	2694.5	2868.5	3136

Table D.21: Dimuon yields for events that passed the ML_2 constraints, but failed the STD constraints detailed in yield per I_T bin for $x_T \in [0.13 - 0.45)$.

D.4 Comparison

Table D.22 summarizes the dimuon yields for the STD and ML_2 analyses. It details the integrated events using the STD and ML_2 constraints, the intersection of the two sets, the set differences, and the ratio of the events passing solely the STD and ML_2 constraints to the union of the sets. This information is listed for all of the targets and illustrates that there is a large overlap of the events that pass both sets of constraints.

Target	STD	ML_2	STD $\cap$ ML_2	STD$\setminus$$ML_2$	$ML_2\setminus$STD	R_{STD}	R_{ML_2}
LH_2	23868	29653	21193	2675	8460	0.74	0.92
EMPTY	749	1420	611	138	809	0.48	0.91
LD_2	26906	32254	24964	1942	7290	0.79	0.94
NONE	499	1238	398.5	100.5	839.5	0.37	0.92
Fe	11336	12930	10235.5	1100.5	2694.5	0.81	0.92
C	10143	12075	9206.5	936.5	2868.5	0.78	0.93
W	13850	15668.5	12532.5	1317.5	3136	0.82	0.92

Table D.22: Comparison of the dimuon yields for the STD and ML_2 constraints for all of the targets. STD and ML_2 are the total event yields using the STD and ML_2 constraints, respectively. STD $\cap$ ML_2 are the events that passed both sets of constraints, STD$\setminus$$ML_2$ denotes the events that passed the STD constraints, but failed the ML_2 constrains, and $ML_2\setminus$STD is the converse. R_{STD} and R_{ML_2} are the ratios of the STD and ML_2 to the union of STD and ML_2(STD $\cup$ ML_2), respectively.

The R_{ML_2} is fairly consistent across all of the targets illustrating that as a share of all of the events passing either of the constraints, the ML_2 finds more than 90% of the events. The STD results are fairly consistent as well with the caveat of the EMPTY and NONE targets. The STD constraints find 73-81% of the events. As mentioned in Section 3.3, the liquid targets see the largest difference between the constraint sets. The ML_2 finds 24% more LH_2 and almost 20% more LD_2 events than the STD constraints. This is followed by the C, Fe, and W results where the ML_2 finds 19%, 14%, and 13% more events than the STD constraints. The large difference in the EMPTY and NONE targets could be the source for some of the CSR differences between

the constraint sets. This is due to background subtraction, since the background is scaled by $POT_{Target}/POT_{Background}$ and the increase in the number of target events is less than the increase in the number of background events, more of the target would be removed in the ML_2 than in the STD. This would alter the CS before the CSR is calculated and could lead to differences in the extracted CSR.

References

[1] Linus Pauling. *General Chemistry*. Dover Publications, 1988. ISBN: 9780486656229.

[2] Charlotte Bigg. "Evident atoms: visuality in Jean Perrin's Brownian motion research". In: *Studies in History and Philosophy of Science Part A* 39.3 (2008). Science and the Changing Senses of Reality CIRCA 1900, pp. 312–322. DOI: https://doi.org/10.1016/j.shpsa.2008.06.003. URL: http://www.sciencedirect.com/science/article/pii/S0039368108000563.

[3] David Griffiths. *Introduction to Elementary Particles*. WILEY-VCH Verlag GmbH Co. KGaA, 2008. ISBN: 9783527406012.

[4] James Chadwick. "Possible Existence of a Neutron". In: *Nature* 129 (1932). DOI: https://doi.org/10.1038/129312a0.

[5] Hideki YUKAWA. "On the Interaction of Elementary Particles. I". In: *Proceedings of the Physico-Mathematical Society of Japan. 3rd Series* 17 (1935), pp. 48–57. DOI: 10.11429/ppmsj1919.17.0_48.

[6] K Nishijima. "From isospin to strangeness". In: *AIP Conference Proceedings*. Vol. 300. 1. American Institute of Physics. 1994, pp. 125–131.

[7] M. Gell-Mann. "A schematic model of baryons and mesons". In: *Physics Letters* 8.3 (1964), pp. 214–215. DOI: https://doi.org/10.1016/S0031-9163(64)92001-3. URL: http://www.sciencedirect.com/science/article/pii/S0031916364920013.

[8] Lauraschudder. *Meson Octet*. 2007. URL: https://commons.wikimedia.org/w/index.php?curid=2379550. License: Creative Commons BY-SA 3.0.

[9] Lauraschudder. *Baryon Octet*. 2007. URL: https://commons.wikimedia.org/w/index.php?curid=2379551. License: Creative Commons BY-SA 3.0.

[10] Lauraschudder. *Baryon Decuplet*. 2007. URL: https://commons.wikimedia.org/w/index.php?curid=2379556. License: Creative Commons BY-SA 3.0.

[11] Susumu Okubo. "Note on unitary symmetry in strong interaction. ii: Excited states of baryons". In: *Progress of Theoretical Physics* 28.1 (1962), pp. 24–32.

[12] Nicola Cabibbo. "Unitary Symmetry and Leptonic Decays". In: *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531. URL: https://link.aps.org/doi/10.1103/PhysRevLett.10.531.

[13] Michael Riordan. "The Discovery of Quarks". In: *Science* 256.5061 (1992), pp. 1287–1293. DOI: 10.1126/science.256.5061.1287. eprint: https://science.sciencemag.org/content/256/5061/1287.full.pdf. URL: https://science.sciencemag.org/content/256/5061/1287.

[14] M. Breidenbach et al. "Observed Behavior of Highly Inelastic Electron-Proton Scattering". In: *Phys. Rev. Lett.* 23 (16 Oct. 1969), pp. 935–939. DOI: 10.1103/PhysRevLett.23.935. URL: https://link.aps.org/doi/10.1103/PhysRevLett.23.935.

[15] J. J. Aubert et al. "Experimental Observation of a Heavy Particle J". In: *Phys. Rev. Lett.* 33 (23 Dec. 1974), pp. 1404–1406. DOI: 10.1103/PhysRevLett.33.1404. URL: https://link.aps.org/doi/10.1103/PhysRevLett.33.1404.

[16] J. Augustin *et al.* "Discovery of a Narrow Resonance in e^+e^- Annihilation". In: *Phys. Rev. Lett.* 33 (23 Dec. 1974), pp. 1406–1408. DOI: 10.1103/PhysRevLett.33.1406. URL: https://link.aps.org/doi/10.1103/PhysRevLett.33.1406.

[17] V. Khachatryan et al. "Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV". In: *Journal of High Energy Physics* 2017.3 (Mar. 2017). DOI: 10.1007/jhep03(2017)156. URL: http://dx.doi.org/10.1007/JHEP03(2017)156.

[18] David J. Gross and Frank Wilczek. "Ultraviolet Behavior of Non-Abelian Gauge Theories". In: *Phys. Rev. Lett.* 30 (26 June 1973), pp. 1343–1346. DOI: 10.1103/PhysRevLett.30.1343. URL: https://link.aps.org/doi/10.1103/PhysRevLett.30.1343.

[19] Yi-Bo Yang et al. "Proton Mass Decomposition from the QCD Energy Momentum Tensor". In: *Physical Review Letters* 121.21 (Nov. 2018). DOI: 10.1103/physrevlett.121.212001. URL: http://dx.doi.org/10.1103/PhysRevLett.121.212001.

[20] M. Tanabashi *et al.* "Review of Particle Physics". In: *Phys. Rev. D* 98 (3 Aug. 2018 and 2019 update), p. 030001. DOI: 10.1103/PhysRevD.98.030001. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[21] John C. Collins, Davison E. Soper, and George Sterman. *Factorization of Hard Processes in QCD*. 2004. arXiv: hep-ph/0409313 [hep-ph].

[22] Joseph F. Owens and Wu-Ki Tung. "Parton distribution functions of hadrons". In: *Ann. Rev. Nucl. Part. Sci.* 42 (1992), pp. 291–332. DOI: 10.1146/annurev.ns.42.120192.001451.

[23] K. Kovařík et al. "nCTEQ15: Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework". In: *Physical Review D* 93.8 (Apr. 2016). DOI: 10.1103/physrevd.93.085037. URL: http://dx.doi.org/10.1103/PhysRevD.93.085037.

[24] Kari J. Eskola et al. "EPPS16: nuclear parton distributions with LHC data". In: *The European Physical Journal C* 77.3 (Mar. 2017). DOI: 10.1140/epjc/s10052-017-4725-9. URL: http://dx.doi.org/10.1140/epjc/s10052-017-4725-9.

[25] A. D. Martin et al. "Parton distributions for the LHC". In: *The European Physical Journal C* 63.2 (July 2009), pp. 189–285. DOI: 10.1140/epjc/s10052-009-1072-5. URL: http://dx.doi.org/10.1140/epjc/s10052-009-1072-5.

[26] J. Pumplin, D. R. Stump, and W. K. Tung. "Multivariate fitting and the error matrix in global analysis of data". In: *Phys. Rev. D* 65 (1 Dec. 2001), p. 014011. DOI: 10.1103/PhysRevD.65.014011. URL: https://link.aps.org/doi/10.1103/PhysRevD.65.014011.

[27] J. Pumplin et al. "Uncertainties of predictions from parton distribution functions. II. The Hessian method". In: *Physical Review D* 65.1 (Dec. 2001). DOI: 10.1103/physrevd.65.014013. URL: http://dx.doi.org/10.1103/PhysRevD.65.014013.

[28] Guido Altarelli and Giorgio Parisi. "Asymptotic freedom in parton language". In: *Nuclear Physics B* 126.2 (1977), pp. 298–318.

[29] Wen-Chen Chang and Jen-Chieh Peng. "Flavor structure of the nucleon sea". In: *Progress in Particle and Nuclear Physics* 79 (Nov. 2014), pp. 95–135. DOI: 10.1016/j.ppnp.2014.08.002. URL: http://dx.doi.org/10.1016/j.ppnp.2014.08.002.

[30] D F Geesaman, K Saito, and A W Thomas. "The Nuclear EMC Effect". In: *Annual Review of Nuclear and Particle Science* 45.1 (1995), pp. 337–390. DOI: 10.1146/annurev.ns.45.120195.002005. eprint: https://doi.org/10.1146/annurev.ns.45.120195.002005. URL: https://doi.org/10.1146/annurev.ns.45.120195.002005.

[31] S. Chekanov et al. "ZEUS next-to-leading-order QCD analysis of data on deep inelastic scattering". In: *Physical Review D* 67.1 (Jan. 2003). DOI: 10.1103/physrevd.67.012007. URL: http://dx.doi.org/10.1103/PhysRevD.67.012007.

[32] Kurt Gottfried. "Sum Rule for High-Energy Electron-Proton Scattering". In: *Phys. Rev. Lett.* 18 (25 June 1967), pp. 1174–1177. DOI: 10.1103/PhysRevLett.18.1174. URL: https://link.aps.org/doi/10.1103/PhysRevLett.18.1174.

[33] P. Amaudruz et al. "Gottfried sum from the ratio F_2^n/F_2^p". In: *Phys. Rev. Lett.* 66 (21 May 1991), pp. 2712–2715. DOI: 10.1103/PhysRevLett.66.2712. URL: https://link.aps.org/doi/10.1103/PhysRevLett.66.2712.

[34] Shunzo Kumano. *Flavor-dependent antiquark distributions in the nucleons and nuclei*. Oct. 2017. URL: http://www.int.washington.edu/talks/WorkShops/int_17_68W/People/Kumano_S/Kumano.pdf.

[35] S. D. Drell and Tung-Mow Yan. "Massive Lepton Pair Production in Hadron-Hadron Collisions at High-Energies". In: *Phys. Rev. Lett.* 25 (1970). [Erratum: Phys. Rev. Lett.25,902(1970)], pp. 316–320. DOI: 10.1103/PhysRevLett.25.316,10.1103/PhysRevLett.25.902.2.

[36] P. L. McGaughey, J. M. Moss, and J. C. Peng. "HIGH-ENERGY HADRON-INDUCED DILEPTON PRODUCTION FROM NUCLEONS AND NUCLEI". In: *Annual Review of Nuclear and Particle Science* 49.1 (Dec. 1999), pp. 217–253. DOI: 10.1146/annurev.nucl.49.1.217. URL: http://dx.doi.org/10.1146/annurev.nucl.49.1.217.

[37] The CMS collaboration. "Performance of CMS muon reconstruction in pp collision events ats = 7TeV". In: *Journal of Instrumentation* 7.10 (Oct. 2012), P10002–P10002. DOI: 10.1088/1748-0221/7/10/p10002. URL: http://dx.doi.org/10.1088/1748-0221/7/10/P10002.

[38] J. G. Heinrich et al. "Measurement of the ratio of sea to valence quarks in the nucleon". In: *Phys. Rev. Lett.* 63 (4 July 1989), pp. 356–359. DOI: 10.1103/PhysRevLett.63.356. URL: https://link.aps.org/doi/10.1103/PhysRevLett.63.356.

[39] A. Baldit et al. "Study of the isospin symmetry breaking in the light quark sea of the nucleon from the Drell-Yan process". In: *Phys. Lett. B* 332 (1994), pp. 244–250. DOI: 10.1016/0370-2693(94)90884-2.

[40] E. A. Hawker et al. "Measurement of the Light Antiquark Flavor Asymmetry in the Nucleon Sea". In: *Physical Review Letters* 80.17 (Apr. 1998), pp. 3715–3718. DOI: 10.1103/physrevlett.80.3715. URL: http://dx.doi.org/10.1103/PhysRevLett.80.3715.

[41] R. S. Towell et al. "Improved measurement of thed/ūasymmetry in the nucleon sea". In: *Physical Review D* 64.5 (Aug. 2001). DOI: 10.1103/physrevd.64.052002. URL: http://dx.doi.org/10.1103/PhysRevD.64.052002.

[42] J. C. Webb et al. *Absolute Drell-Yan Dimuon Cross Sections in 800 GeV/c pp and pd Collisions*. 2003. arXiv: hep-ex/0302019 [hep-ex].

[43] M.A. Vasilev et al. "Parton energy loss limits and shadowing in Drell-Yan dimuon production". In: *Phys. Rev. Lett.* 83 (1999), pp. 2304–2307. DOI: 10.1103/PhysRevLett.83.2304. arXiv: hep-ex/9906010.

[44] Claude Bourrely and Jacques Soffer. *Flavor symmetry breaking of the nucleon sea in the statistical approach*. 2019. arXiv: 1901.03071 [hep-ph].

[45] Mary Alberg and Gerald A. Miller. "Chiral light-front perturbation theory and the flavor dependence of the light-quark nucleon sea". In: *Phys. Rev. C* 100 (3 Sept. 2019), p. 035205. DOI: 10.1103/PhysRevC.100.035205. URL: https://link.aps.org/doi/10.1103/PhysRevC.100.035205.

[46] D F Geesaman and P E Reimer. "The sea of quarks and antiquarks in the nucleon". In: *Reports on Progress in Physics* 82.4 (Mar. 2019), p. 046301. DOI: 10.1088/1361-6633/ab05a7. URL: http://dx.doi.org/10.1088/1361-6633/ab05a7.

[47] P. L. McGaughey *et al.* "Limit on the $\bar{d}/\bar{u}$ asymmetry of the nucleon sea from Drell-Yan production". In: *Phys. Rev. Lett.* 69 (12 Sept. 1992), pp. 1726–1728. DOI: 10.1103/PhysRevLett.69.1726. URL: https://link.aps.org/doi/10.1103/PhysRevLett.69.1726.

[48] Marina Walt, Ilkka Helenius, and Werner Vogelsang. "Open-source QCD analysis of nuclear parton distribution functions at NLO and NNLO". In: *Physical Review D* 100.9 (Nov. 2019). DOI: 10.1103/physrevd.100.096015. URL: http://dx.doi.org/10.1103/PhysRevD.100.096015.

[49] Klaus Rith. *Present Status of the EMC effect*. 2014. arXiv: 1402.5000 [hep-ex].

[50] Jianwei Qiu. "Nuclear shadowing at small values of x". In: *Nuclear Physics B* 291 (1987), pp. 746–764. DOI: https://doi.org/10.1016/0550-3213(87)90494-9. URL: http://www.sciencedirect.com/science/article/pii/0550321387904949.

[51] Nikolai Nikolaev and B. Zakharov. "Colour transparency and scaling properties of nuclear shadowing in deep inelastic scattering". In: *Zeitschrift für Physik C* 49 (Dec. 1991), pp. 607–618. DOI: 10.1007/BF01483577.

[52] Stanley J. Brodsky and Hung Jung Lu. "Shadowing and antishadowing of nuclear structure functions". In: *Phys. Rev. Lett.* 64 (12 Mar. 1990), pp. 1342–1345. DOI: 10.1103/PhysRevLett.64.1342. URL: https://link.aps.org/doi/10.1103/PhysRevLett.64.1342.

[53] J.J. Aubert et al. “The ratio of the nucleon structure functions F2N for iron and deuterium”. In: *Physics Letters B* 123.3 (1983), pp. 275–278. DOI: https://doi.org/10.1016/0370-2693(83)90437-9. URL: http://www.sciencedirect.com/science/article/pii/0370269383904379.

[54] A. Bodek *et al.* “Electron Scattering from Nuclear Targets and Quark Distributions in Nuclei”. In: *Phys. Rev. Lett.* 50 (19 May 1983), pp. 1431–1434. DOI: 10.1103/PhysRevLett.50.1431. URL: https://link.aps.org/doi/10.1103/PhysRevLett.50.1431.

[55] A. Bodek *et al.* “Comparison of the Deep-Inelastic Structure Functions of Deuterium and Aluminum Nuclei”. In: *Phys. Rev. Lett.* 51 (7 Aug. 1983), pp. 534–537. DOI: 10.1103/PhysRevLett.51.534. URL: https://link.aps.org/doi/10.1103/PhysRevLett.51.534.

[56] P R Norton. “The EMC effect”. In: *Reports on Progress in Physics* 66.8 (July 2003), pp. 1253–1297. DOI: 10.1088/0034-4885/66/8/201.

[57] Edmond L. Berger and F. Coester. “Nuclear effects in deep-inelastic lepton scattering”. In: *Phys. Rev. D* 32 (5 Sept. 1985), pp. 1071–1084. DOI: 10.1103/PhysRevD.32.1071. URL: https://link.aps.org/doi/10.1103/PhysRevD.32.1071.

[58] J. Seely *et al.* “New Measurements of the European Muon Collaboration Effect in Very Light Nuclei”. In: *Phys. Rev. Lett.* 103 (20 Nov. 2009), p. 202301. DOI: 10.1103/PhysRevLett.103.202301. URL: https://link.aps.org/doi/10.1103/PhysRevLett.103.202301.

[59] C. A. Garcıa Canal, E. M. Santangelo, and H. Vucetich. “Nucleon Effective Mass and the A Dependence of Structure Functions”. In: *Phys. Rev. Lett.* 53 (15 Oct. 1984), pp. 1430–1432. DOI: 10.1103/PhysRevLett.53.1430. URL: https://link.aps.org/doi/10.1103/PhysRevLett.53.1430.

[60] Hong Jung and Gerald A. Miller. “Pionic contributions to deep inelastic nuclear structure functions”. In: *Phys. Rev. C* 41 (2 Feb. 1990), pp. 659–664. DOI: 10.1103/PhysRevC.41.659. URL: https://link.aps.org/doi/10.1103/PhysRevC.41.659.

[61] A. E. L. Dieperink and C. L. Korpa. “Pions in the nuclear medium and Drell-Yan scattering”. In: *Physical Review C* 55.5 (May 1997), pp. 2665–2674. DOI: 10.1103/physrevc.55.2665. URL: http://dx.doi.org/10.1103/PhysRevC.55.2665.

[62] K.E. Lassila and U.P. Sukhatme. “The EMC effect at all x in the quark cluster model”. In: *Physics Letters B* 209.2 (1988), pp. 343–346. DOI: https://doi.org/10.1016/0370-2693(88)90959-8. URL: http://www.sciencedirect.com/science/article/pii/0370269388909598.

[63] Simona Malace et al. “The challenge of the EMC effect: Existing data and future directions”. In: *International Journal of Modern Physics E* 23.08 (Aug. 2014), p. 1430013. DOI: 10.1142/s0218301314300136. URL: http://dx.doi.org/10.1142/S0218301314300136.

[64] N. Fomin *et, al*. “New Measurements of High-Momentum Nucleons and Short-Range Structures in Nuclei”. In: *Phys. Rev. Lett.* 108 (9 Feb. 2012), p. 092502. DOI: 10.1103/PhysRevLett.108.092502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.108.092502.

[65] D. W. Higinbotham, J. Gomez, and E. Piasetzky. *Nuclear Scaling and the EMC Effect*. 2010. arXiv: 1003.4497 [hep-ph].

[66] Or Hen et al. “Nucleon-nucleon correlations, short-lived excitations, and the quarks within”. In: *Rev. Mod. Phys.* 89 (4 Nov. 2017), p. 045002. DOI: 10.1103/RevModPhys.89.045002. URL: https://link.aps.org/doi/10.1103/RevModPhys.89.045002.

[67] L. B. Weinstein et al. “Short Range Correlations and the EMC Effect”. In: *Phys. Rev. Lett.* 106 (5 Feb. 2011), p. 052301. DOI: 10.1103/PhysRevLett.106.052301. URL: https://link.aps.org/doi/10.1103/PhysRevLett.106.052301.

[68] J. Arrington *et al. Drell-Yan Measurements of Nucleon and Nuclear Structure with the Fermilab Main Injector: E906. Projects Document 395-v2*. Oct. 2006. URL: https://projects-docdb.fnal.gov/cgi-bin/ShowDocument?docid=395.

[69] D.M. Alde *et al.* “Nuclear dependence of dimuon production at 800 GeV”. In: *Phys. Rev. Lett.* 64 (21 May 1990), pp. 2479–2482. DOI: 10.1103/PhysRevLett.64.2479. URL: https://link.aps.org/doi/10.1103/PhysRevLett.64.2479.

[70] Jessica Orwig. *Cockcroft-Walton’s successor: a peep inside the new RFQ and how it works*. Nov. 2012. URL: https://www.fnal.gov/pub/today/archive/archive_2012/today12-11-02_RFQReadmore.html.

[71] Joseph Piergrossi. *RFQ system to take Fermilab into the Intensity Frontier*. Dec. 2012. URL: https://www.fnal.gov/pub/today/archive/archive_2012/today12-08-27.html.

[72] Joey Thompson. “Introduction to colliding beams at Fermilab”. In: (Oct. 1994). DOI: 10.2172/10105360.

[73] *Fermilab’s Accelerator Complex*. Mar. 2019. URL: https://www.fnal.gov/pub/science/particle-accelerators/accelerator-complex.html.

[74] C.A. Aidala *et al.* "The SeaQuest spectrometer at Fermilab". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 930 (2019), pp. 49–63. DOI: https://doi.org/10.1016/j.nima.2019.03.039. URL: http://www.sciencedirect.com/science/article/pii/S016890021930347X.

[75] Adam Watts. *Slow Spill extraction - Beam to SeaQuest*. Aug. 2016. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=1715.

[76] Bryan Ramsom. *E906 Beamline Cherenkov Detector*. June 2013. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=772.

[77] Bryan P. Dannowitz. "Nuclear Dependence of Proton-Induced Drell-Yan Dimuon Production at 120 GeV at Seaquest". PhD thesis. Illinois U., Urbana, 2016. DOI: 10.2172/1294515.

[78] Paul E. Reimer. *Hydrogen and Deuterium Analysis*. Mar. 2017. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=1893.

[79] Don Geesaman. *Handling the liquid target contamination and densities*. Apr. 2019. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=4993.

[80] Paul E. Reimer and Josh Rubin. *Solid targets*. Sept. 2015. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=782.

[81] Don Geesaman. *Provenance of Magnetic Fields used in GMC and Tracker*. Feb. 2018. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=3053.

[82] Kun Liu. *Detector alignment and magnetic field calibration*. Feb. 2015. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=1275.

[83] Naomi Makins. *Station 1 and 2 Hodoscopes*. Apr. 2010. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=233.

[84] Donald Isenhower. *Station 3 and 4 Hodoscope Report*. May 2010. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=229.

[85] Donald Isenhower. *Report on Station 3 and 4 Hodoscopes*. Apr. 2010. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=218.

[86] Kei Nagai. *Performance of Drift Chambers for E906/SeaQuest Drell–Yan Experiment at Fermilab*. Oct. 2014. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=1143.

[87] Catherine Culkin. *Drift Chambers*. June 2013. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=762.

[88] Kei Nagai. *Chamber Efficiency in Run2*. Mar. 2015. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=1321.

[89] Shumpei Nara. *Drift chamber efficiency calculated by Hodoscope tracking*. Aug. 2015. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=1459.

[90] Kei Nagai. "Recent Measurement of Flavor Asymmetry of Antiquarks in the Proton by Drell–Yan Experiment SeaQuest at Fermilab". PhD thesis. Tokyo Inst. Tech, 2017. DOI: 10.2172/1346822.

[91] Zhaohuizi Ji. *Proportional tubes*. July 2019. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=6079.

[92] M. Tanabashi et al. "Review of Particle Physics". In: *Phys. Rev. D* 98.3 (2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001.

[93] Sho Uemura. *Prop tubes and fine-grained hodoscopes*. June 2018. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=3614.

[94] Shiuan-Hal *et al.* "FPGA-based trigger system for the Fermilab SeaQuest experiments". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 802 (Dec. 2015), pp. 82–88. DOI: 10.1016/j.nima.2015.09.001. URL: http://dx.doi.org/10.1016/j.nima.2015.09.001.

[95] Kaz Nakahara and Brian G. Tice. *DAQ documentation*. June 2017. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=1033.

[96] Kun Liu and Su-Yin Wang. *E906 DAQ upgrade project*. Aug. 2016. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=1724.

[97] Shou Miyasaka. "Probing Flavor Asymmetry of Anti-quarks in the Proton by Drell-Yan Experiment SeaQuest". PhD thesis. Tokyo Inst. Tech., 2016. DOI: 10.2172/1325974.

[98] Brian G. Tice. *Scalar DAQ Crash Course*. June 2015. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=1408.

[99] Brian G. Tice. *Beam DAQ Crash Course*. June 2015. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=1410.

[100] Kun Liu. *kTracker step-by-step walkthrough*. Feb. 2015. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=1283.

[101] Kun Liu. *Tracking Overview for DNP 2015*. Nov. 2015. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=1530.

[102] Kun Liu. *Track fitting based on Kalman filter*. Apr. 2013. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=448.

[103] N. Chuck Brown and Arun Tadepalli. *Beam y-offset for different roadsets*. June 2017. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=2198.

[104] Kenichi Nakano. *Spill Selection on Roadsets 57-70 with R006$_V$003*. Mar. 2017. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=1307.

[105] C. David Christian and Brian G. Tice. *Absolute Normalization - Beam Inhibit and DAQ Deadtime*. June 2015. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=1212.

[106] N. Chuck Brown and Arun Tadepalli. *Data selection cuts*. Nov. 2018. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=2111.

[107] Mahsa Shoaran *et al.* "Energy-Efficient Classification for Resource-Constrained Biomedical Applications". In: *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* PP (June 2018), pp. 1–1. DOI: 10.1109/JETCAS.2018.2844733.

[108] A. Hoecker *et,. al*. *TMVA - Toolkit for Multivariate Data Analysis*. 2007. arXiv: physics/0703039 [physics.data-an].

[109] *NIST/SEMATECH e-Handbook of Statistical Methods : Properties of Principal Components*. URL: https://www.itl.nist.gov/div898/handbook/pmc/section5/pmc551.htm.

[110] Nikita Sharma. *Understanding the Mathematics Behind Principal Component Analysis*. Feb. 2020. URL: https://heartbeat.fritz.ai/understanding-the-mathematics-behind-principal-component-analysis-efd7c9ff0bb3.

[111] Marshall Scott. *2019 April Meeting APS Slides: Using Machine Learning Tools to Enhance Background Subtraction at E906/SeaQuest*. Apr. 2019. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=5443.

[112] Kun Liu. *R005 data format, reconstruction status and alignment issue*. Sept. 2015. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=1476.

[113] Evan McClellan. *Angular Distributions of High-Mass Dilepton Production in Hadron Collisions*. June 2016. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=1676.

[114] Kenichi Nakano. *Events with QIE.RF+00 = 0 in Early Part of Run 2*. July 2016. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=1378.

[115] Bryan Kerns. *QIE PEdestal Graphs*. July 2015. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=1424.

[116] Don Geesaman. *Handling the liquid target contamination and densities*. Apr. 2019. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=4993.

[117] Don Geesaman *et al. Nuclear dependence of Anti-Quarks in Nuclei (DY) - analysis note*. Apr. 2020. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=6001.

[118] Daniel H. Morton. *Recent Combined Fit*. Apr. 2020. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=7849.

[119] Paul E. Reimer. *What do we publish?* Apr. 2020. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=7732.

[120] Joseph E. Cavanaugh and Andrew A. Neath. "The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements". In: *WIREs Computational Statistics* 11.3 (2019), e1460. DOI: 10.1002/wics.1460. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wics.1460. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.1460.

[121] H. Akaike. "A new look at the statistical model identification". In: *IEEE Transactions on Automatic Control* 19.6 (1974), pp. 716–723.

[122] A.M.F. Mood, F.A. Graybill, and D.C. Boes. *Introduction to the Theory of Statistics*. McGraw-Hill international editions: Statistics series. McGraw-Hill, 1974. ISBN: 9780070854659. URL: https://books.google.com/books?id=dAjnRwAACAAJ.

[123] Joseph Beyene and Rahim Moineddir. "Methods for confidence interval estimation of a ratio parameter with application to location quotients". In: *BMC Medical Research Methodology* 5:1 (1 2005), pp. 1–7. DOI: 10.1186/1471-2288-5-32.

[124] William P. Dunlap and N. Clayton Silver. "Confidence intervals and standard errors for ratio of normal variables". In: *Behavior Research Methods, Instruments, Computers* 18 (5 1986), pp. 469–471. DOI: 10.3758/BF03201412.

[125] Roger John Barlow. *Practical Statistics for Particle Physics*. 2019. arXiv: 1905.12362 [physics.data-an].

[126] Roger Barlow. *Asymmetric Systematic Errors*. 2003. arXiv: physics/0306138 [physics.data-ar

[127] Sherry Towers. *Maximum Likelihood Estimation (MLE)*. URL: http://www.sherrytowers.com/mle%5C_introduction.pdf.

[128] Mervyn G. Marasinghe. *Maximum Likelihood Estimation*. URL: https://mervyn.public.iastate.edu/stat580/Notes/s09mle.pdf.

[129] Larry Wasserman. *Lecture Notes 9: Asymptotic Theory(Chapter 9)*. URL: http://www.stat.cmu.edu/~larry/=stat705/Lecture9.pdf.

[130] Roger Barlow. *Asymmetry-Master*. 2019. DOI: 10.5281/zenodo.1494909.

[131] Chuck N. Brown *et al. Beam loss from solid targets due to Y-offset*. Oct. 2019. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=6605.

[132] Kenichi Nakano. *Study of Mass-Cut Dependence with Individual Intensity Fit*. Apr. 2020. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=7816.

[133] Jason Dove. *Mass Cut Effects*. Apr. 2020. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=7726.

[134] P. J. Ehlers et al. "Nuclear effects in the proton-deuteron Drell-Yan process". In: *Phys. Rev. D* 90 (1 July 2014), p. 014010. DOI: 10.1103/PhysRevD.90.014010. URL: https://link.aps.org/doi/10.1103/PhysRevD.90.014010.

[135] Hannu Paukkunen. "Nuclear PDFs Today". In: *PoS* HardProbes2018 (2018), p. 014. DOI: 10.22323/1.345.0014. arXiv: 1811.01976 [hep-ph].

[136] Paul E. Reimer. *dbar/ubar cross check*. Mar. 2020. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=7446.

[137] Paul E. Reimer. *Cross Check of dbar/ubar extraction*. Mar. 2020. URL: https://seaquest-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=7692.

[138] Claude Bourrely and Jacques Soffer. "New developments in the statistical approach of parton distributions: Tests and predictions up to LHC energies". In: *Nuclear Physics A* 941 (Sept. 2015), pp. 307–334. DOI: 10.1016/j.nuclphysa.2015.06.018. URL: http://dx.doi.org/10.1016/j.nuclphysa.2015.06.018.

[139] S. Kumano. *Nuclear modification of the flavor asymmetry* $\bar{u} - \bar{d}$. 1996. arXiv: hep-ph/9605450 [hep-ph].