# The Radiance of the Gluon Spin: Constraining the Proton Spin Structure with the Direct Photon Double Helicity Asymmetry

by

Michael Betancourt

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2011

© Massachusetts Institute of Technology 2011. All rights reserved.

Author . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Department of Physics
November 15, 2011

Certified by . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Robert Redwine
Professor of Physics
Thesis Supervisor

Accepted by . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Krishna Rajagopal
Associate Department Head for Education

# The Radiance of the Gluon Spin: Constraining the Proton Spin Structure with the Direct Photon Double Helicity Asymmetry

by

Michael Betancourt

Submitted to the Department of Physics
on November 15, 2011, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Physics

## Abstract

Although fundamental to the observable universe, the proton is not elementary. Rather the particle is a bound state of three valence quarks and the QCD vacuum that condenses around them, its properties an amalgamation of those underlying degrees of freedom. Naive expectations presume that contributions from the valence quarks dominate these properties, but the deep inelastic scattering (DIS) experiments which first investigated the proton structure in detail revealed the importance of the vacuum. In particular, polarized DIS measurements uncovered a surprisingly inadequate quark polarization, necessitating significant contributions to the proton spin from elsewhere. The total spin of the gluon field confining the quarks is one possibility, but a contribution only weakly constrained by the electromagnetic probes of DIS.

An observable far more sensitive to contributions from the gluon field can be found in the collision of two polarized protons. By correlating the incident proton helicities with final-states originating from an initial-state gluon, the double-helicity asymmetry directly probes the underlying gluon polarization and provides much stronger experimental constraints. Asymmetries measured with hadronic final-states have already improved the understanding of the proton spin structure significantly, but with accumulating statistics these measurements will eventually be limited by systematic uncertainties. Although direct photons are rare in the hadronic environment, the simplicity of the resulting asymmetry ultimately promises a more precise probe of the gluon polarization.

Located at the Relativistic Heavy Ion Collider (RHIC), the only facility in the world capable of accelerating and colliding polarized proton beams, the Solenoidal Tracker at RHIC (STAR) provides the large acceptance electromagnetic calorimetry and charged particle tracking critical for measuring direct photons and, subsequently, their asymmetry. Utilizing data from the 2009 running period with intricate simulation, state-of-the-art statistical methods have been developed to tease out the rare photon signal from an overwhelming hadronic background to enable the first direct photon measurements at STAR.

This thesis details the construction of the unpolarized cross section and an initial double-helicity asymmetry, proving the feasibility of the direct photon program at the experiment.

Thesis Supervisor: Robert Redwine
Title: Professor of Physics

*For my father, who never had the chance to be curious.*

# Preface

Because of the difficulty of the analysis, this thesis utilizes a significant amount of statistical methodology with which the reader may be unfamiliar. Appendix C briefly introduces the theory behind these techniques, but for a more thorough understanding the reader is encouraged to consult the references included therein. In order to maintain a manageable scope throughout the work, specific implementation details are not discussed.

This thesis references many documents accessible only to members of the STAR collaboration. Because they contain critical information not available in the public domain, they are included for completeness. To avoid any confusion, these references are denoted with a † in the bibliography.

# Acknowledgments

Despite the frustration, anxiety, and distress typical of the graduate student experience, my years at MIT have been great because of the people with whom I've had the great fortune to share my time.

First and foremost I am indebted to my advisor, Bob Redwine. Since I first began my graduate work, Bob has been a dependable resource for encouragement, guidance, and amusing anecdotes: I could not be happier that I'll never be able to commiserate over terrible advisors.

My thanks to everyone who has been a part of the MIT RHIC Spin group through the years, particularly Bernd Surrow for giving me the freedom to pursue an unconventional analysis, Alan Hoffman and his frustratingly insightful perspectives on the graduate experience, Matt Walker and his refusal to accept inferior science, and Joe Seele who continuously challenged me if only because I needed it. I would be remiss if I failed to acknowledge the support of Hal Spinka, Jason Webb, and Jérôme Lauret in the STAR Collaboration, or the continued help from the dedicated staff of the Laboratory for Nuclear Science and the MIT Department of Physics.

I can only hope that my future endeavors will afford me the opportunity to meet people as engaging, passionate, and thoughtful as the graduate students at MIT, Boston, and beyond. Tim Barnes, Sean Clarke and Val Hellstein, Kat Deck, Lu Feng, Eric and Stephanie Fitzgerald, Kasia Gora, Dave Guarrera, Bernadette Heyburn, Steve Jaditz, Chris Jones, Aviv Keshet, Adrian Liu, Andy Lutomirski, John McGee, Jenny Meyer, Elan Pavlov, Leslie Rogers, John Rutherford, Axel Schmidt, Naveen Sinha, Nic Smith-Lefebvre, Leo Stein, Chun-Yi Sum, Kazu Terao, Victor Tsai, Sarah Vigeland, Chris Williams, Phil Zukin: I'll miss you guys.

Lastly, I thank my parents for provided me the determination and the opportunity to pursue my passions.

# Contents

Preface . . . . . . . . . vi
Acknowledgements . . . . . . . . . vii

**1 Why Does the Proton Spin? 1**
1.1 Modeling the Proton . . . . . . . . . 2
1.1.1 Static Quarks . . . . . . . . . 2
1.1.2 To Infinity and Bjorken:
Deep Inelastic Scattering and the Parton Model . . . . . . . . . 3
1.1.3 Quantum Chromodynamics . . . . . . . . . 14
1.2 Probing the Gluon Spin . . . . . . . . . 20
1.2.1 Gluon Gazing . . . . . . . . . 24

**2 Colliding Polarized Protons**
**and Observing the Aftermath 29**
2.1 The Relativistic Heavy Ion Collider . . . . . . . . . 31
2.1.1 Beam Acceleration . . . . . . . . . 31
2.1.2 Beam Steering . . . . . . . . . 32
2.1.3 Beam Polarization . . . . . . . . . 32
2.1.4 Beam Structure . . . . . . . . . 34
2.2 STAR . . . . . . . . . 35
2.2.1 The “Solenoid” . . . . . . . . . 38
2.2.2 The “Tracker” and Other Detector Subsystems . . . . . . . . . 38
2.2.3 Trigger System and Data Acquisition . . . . . . . . . 43

**3 Data Analysis 47**

3.1 Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 47

3.2 Simulation . . . . . . . . . . . . . . . . . . . . . . . . . . . 49

3.2.1 PYTHIA . . . . . . . . . . . . . . . . . . . . . . . . . 49

3.2.2 GEANT . . . . . . . . . . . . . . . . . . . . . . . . . 52

3.2.3 Reconstruction . . . . . . . . . . . . . . . . . . . . . 53

3.2.4 Filtering . . . . . . . . . . . . . . . . . . . . . . . . . 53

3.3 Preprocessing . . . . . . . . . . . . . . . . . . . . . . . . . 54

3.3.1 Event Selection . . . . . . . . . . . . . . . . . . . . . 56

3.3.2 Photon Candidates . . . . . . . . . . . . . . . . . . . 56

3.3.3 Discriminants . . . . . . . . . . . . . . . . . . . . . . 57

3.4 Analysis Techniques . . . . . . . . . . . . . . . . . . . . . 62

3.4.1 Direct Photon Extraction . . . . . . . . . . . . . . . 63

3.4.2 Deconvolution . . . . . . . . . . . . . . . . . . . . . . 74

3.4.3 Integrated Analysis . . . . . . . . . . . . . . . . . . . 78

**4 Results 79**

4.1 Cross Section . . . . . . . . . . . . . . . . . . . . . . . . . 79

4.1.1 Efficiencies . . . . . . . . . . . . . . . . . . . . . . . . 80

4.1.2 Luminosity . . . . . . . . . . . . . . . . . . . . . . . . 81

4.1.3 Systematic Errors . . . . . . . . . . . . . . . . . . . . 83

4.1.4 Measurement . . . . . . . . . . . . . . . . . . . . . . 84

4.2 Asymmetry . . . . . . . . . . . . . . . . . . . . . . . . . . 84

4.2.1 Polarization . . . . . . . . . . . . . . . . . . . . . . . 87

4.2.2 Relative Luminosity . . . . . . . . . . . . . . . . . . 90

4.2.3 False Asymmetries . . . . . . . . . . . . . . . . . . . 91

4.2.4 Systematic Errors . . . . . . . . . . . . . . . . . . . . 92

4.2.5 Measurement . . . . . . . . . . . . . . . . . . . . . . 94

**5 Conclusion 97**

**A Twist in QCD 99**

**B Discriminant Distributions 101**

**C Inference 113**

C.1 Classical Statistics . . . . . . . . . . . . . . . . . . . . . . . . . . . 114

C.2 Bayesian Inference . . . . . . . . . . . . . . . . . . . . . . . . . . . 115

C.2.1 Marginalization . . . . . . . . . . . . . . . . . . . . . . . . . 116

C.2.2 Model Comparison . . . . . . . . . . . . . . . . . . . . . . . 117

C.2.3 Markov Chain Monte Carlo . . . . . . . . . . . . . . . . . . . 118

**D Gaussian Processes 121**

D.1 Linear Bayesian Regression . . . . . . . . . . . . . . . . . . . . . . . 122

D.2 Nonlinear Bayesian Regression . . . . . . . . . . . . . . . . . . . . . 124

D.3 Gaussian Processes . . . . . . . . . . . . . . . . . . . . . . . . . . . 126

D.3.1 Model Selection . . . . . . . . . . . . . . . . . . . . . . . . . 128

# List of Figures

1-1 Leading-order description of deep inelastic scattering: an incident lepton emits a virtual photon that is absorbed by a nucleon target. . . 4

1-2 The hadronic tensor simplifies dramatically when an application of the optical theorem replaces the squared photon-proton vertex (a) with the imaginary part of the forward Compton amplitude (b), where the proton absorbs a photon at $\vec{x} = \vec{0}$ before emitting a photon at $\vec{x} = \vec{\xi}$. 7

1-3 In the parton model the incident photon scatters not from the proton as a whole but from a point particle and the leading twist behavior of the general forward Compton amplitude (a) reduces to the "handbag" diagram (b). . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

1-4 Results from the E80 and E130 asymmetry measurements. The curve extrapolates to low $x$ with motivation from Regge theory and integrates to give $\int \mathrm{d}x \, g_1^p = 0.17 \pm 0.05$, in agreement with the Ellis-Jaffe sum rule [1]. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

1-5 Results from the EMC asymmetry measurements. The measured $g_1^{ep}$ falls below the E80/E130 extrapolations at small $x$, with $\int dx \, g_1^{ep}$ then deviating from the Ellis-Jaffe result [2]. . . . . . . . . . . . . . . . . 13

1-6 Quark (separated into valence and sea contributions) and gluon polarized distribution functions are inferred from a NLO global analysis of deep inelastic scattering data [3]. . . . . . . . . . . . . . . . . . . . 21

1-7 Cartoon of factorization in proton-proton collisions. Soft, non-perturbative physics is absorbed into the parton distribution functions $f_i$ and $f_j$ while harder perturbative physics gives the partonic cross section $\sigma_{ij}$. $D$ encodes the non-perturbative physics inherent in any fragmentation of final state partons. . . . . . . . . . . . . . . . . . . . . . . . . . 22

1-8 Sea and gluon polarized distribution functions inferred from a global NLO analysis of deep inelastic scattering with hadronic double-helicity asymmetry measurements [4]. Notice the significant improvement in the gluon polarization compared to Figure 1-6. . . . . . . . . . . . . 23

1-9 Leading-order predictions of the double-helicity asymmetry $A_{\mathrm{LL}}$ at $\sqrt{s} = 200$ GeV and central rapidity, given various models of the gluon polarized distribution function [4, 5] for (a) direct photons and (b) jets. The direct photon asymmetry is not only larger than the jet asymmetry, but also provides better model discrimination as the different predictions do not overlap. . . . . . . . . . . . . . . . . . . . . . . 24

1-10 Leading-order direct photon diagrams; time runs from left to right. At moderate $x$ prompt photon production is dominated by the third diagram which is directly sensitive to the polarized gluon distribution function. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25

1-11 Full resummed, next-to-leading-order calculation of the direct photon cross section with a Frixione cone radius of 0.3 provided by Werner Vogelsang based on code originally developed for [6]. . . . . . . . . . 28

2-1 A schematic of the Relativistic Heavy Ion Collider and associated facilities. The ring itself consists of six sextants, each containing an interaction point here labeled with an X (note that, as they pass through each interaction point, the blue and yellow beams weave through each other). In addition to the Siberian Snakes that maintain the transverse polarization of the beam, the interaction points servicing the STAR and PHENIX experiments are equipped with spin rotators to enable the collision of longitudinal beams and hence the study of double-helicity asymmetries. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30

2-2 The four superconducting dipoles in the full Siberian snakes run the beam through a trajectory that rotates the polarization (black) before return to the original beam axis (blue). Adapted from [7]. . . . . . . 33

2-3 A schematic of the STAR detector with the Time Projection Chamber (blue) and Barrel Electromagnetic Calorimeter (red) highlighted. . . 36

2-4 A cartoon of the STAR detector and the STAR coordinate system as viewed from center of RHIC. Adapted from [8]. . . . . . . . . . . . . 37

2-5 A view of a Barrel Electromagnetic Calorimeter (BEMC) module along the $z$ axis. Layers of lead and plastic scintillator held together with steel compression straps form the calorimeter, with epoxy filled channels separating the individual towers. The module itself mounts into the experiment with a rail system. Adapted from [9]. . . . . . . . . . 40

2-6 Expanded view of the Barrel Shower Maximum Detector. The sense gas and anode wires are sheathed in long cells formed in the extruded aluminum separating the two planes of cathode strips, the $\eta$ strips perpendicular cells and the $\phi$ strips parallel. Adapted from [9]. . . . 43

3-1 The modes of the normalized signal and background shapes from the full joint posterior produced distributions of the master discriminant that were qualitatively consistent with the data. A more careful analysis of the residuals, however, revealed non-trivial systematic discrepancies for some of the phase space bins. . . . . . . . . . . . . . . . . 70

3-2 Marginalizing the joint posterior gave (a) the posterior for the signal and background yields and then (b) the posterior for the signal yield alone. The marginalized signal posteriors in each phase space bin were consistent with gaussian distributions, justifying their characterization with only a mean and variance. . . . . . . . . . . . . . . . . . . . . . . 70

3-3 The posterior extraction residuals, defined as the difference between the inferred signal yield and the true yield and represented here by the mean and 68% credibility interval, were consistent with zero even when the signal in the pseudo-data was reweighted by (a) zero, (b) one, (c) two, and (d) a $E_T$-dependent weight that corrected the direct photon cross section assumed in PYTHIA to that seen in the Run 9 data. Because the same partition of the simulation was used and the reweighting does not change the effective sample size of the pseudo-data, the statistical fluctuations are indeed expected to be identical in each test, except for the case of no signal where the positivity constraint biases the fluctuations. . . . . . . . . . . . . . . . . . . . . . . . . 71

3-4 Discrepancies between the data and simulation were significant at (a) lower energies where they manifested as systematic behavior in the signal extraction residuals, but not at (b) higher energies. . . . . . . 72

3-5 The systematic behavior of the residuals was consistent across the phase space bins, partitioning into three regions across the support of the discriminant. . . . . . . . . . . . . . . . . . . . . . . . . . 73

4-1 The inclusive direct photon cross section is in excellent agreement with the resummed NLO calculation, justifying the use of the perturbative theory in constraining the gluon polarization with the direct photon asymmetry. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 86

4-2 The blue and yellow beam polarizations are measured for each fill independently, with corresponding statistical and systematic uncertainties. Only the former are shown here, as the latter are superseded by the constant correlated systematic uncertainty. . . . . . . . . . . . . . . 88

4-3 The relative luminosity $R_3$ was computed for each run within every fill. Note that the statistical uncertainties were of the order $10^{-3}$ and cannot be seen on the scale of the figure. . . . . . . . . . . . . . . . 91

4-4 Constant fits to the false asymmetries were all consistent with zero, showing no evidence of any systematic bias in the experimental configuration: clockwise from the top-left, (a) $A_Y$, (b) $A_B$, (c) $A_{LS}$, and (d) $A_{UL}$. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 93

4-5 The direct photon double-helicity asymmetry only mildly discriminates between different models of the gluon polarization due to the limited statistics of the measurement. . . . . . . . . . . . . . . . . . . . . . . 96

A-1 Factorization theorems typically hold at all orders in the perturbative expansion of the hard interaction, but only at leading twist where the proton is probed locally. This includes higher-order, leading-twist contributions such as (a) but not higher twist contributions such as (b). 100

A-2 The twist of any contribution to the hadronic tensor is given by the number of the fields interacting with the soft structure of the proton. One leading twist 2 contribution is given by two quarks fields (a) while a higher twist 4 contribution features four quark fields. . . . . . . . . 100

# List of Tables

3.1 Details of the L2BGamma trigger and its dependencies at L0. Note that, because the online triggers did not have the final BEMC calibration, all thresholds are approximate. . . . . . . . . . . . . . . . . . . . . . 48

3.2 A photon-specific quality assessment began with the removal of all data failing these criteria, as defined in [10]. . . . . . . . . . . . . . . . . . . 49

3.3 The photon simulation included both direct photon subprocesses and QCD subprocesses. Note that quark-gluon Compton scattering (29) is by far the most dominant of the signal processes. . . . . . . . . . . . 50

3.4 Each component of the Pythia vertex within the STAR detector geometry is sampled independently from a Gaussian distribution. All values are in centimeters. . . . . . . . . . . . . . . . . . . . . . . . . . . 51

3.5 The decays of various hadrons are suppressed within Pythia so that any decay vertex displaced from the hard interaction can be correctly reconstructed in the second phase of the simulation. . . . . . . . . . . 52

3.6 Simulation timestamps were randomly assigned to events in order to mimic the time variations seen in the data, in particular the status tables of the BEMC. . . . . . . . . . . . . . . . . . . . . . . . . . . . 54

3.7 With both filters, the full simulation reached at least 1.87 $\text{pb}^{-1}$ of integrated luminosity across all the relevant phase space. Events in partitions with higher integrated luminosity are weighted to ensure that the effective luminosity is uniform. Note that `CKIN(4)`$= -1$ removes any upper bound on the partonic transverse momentum. . . . . . . . 55

3.8 Hyperparameter optimization within the evidence approximation also implements *automatic relevance detection* which ranks each input variable in order of importance to the output discriminant. Due to redundancies amongst the variables there are often multiple optima and selecting only one can introduce superficial differences between variables. In the particular optimum used for this analysis, this is evident in the higher relevance of the BSMD-$\eta$ variables compared to their BSMD-$\phi$ equivalents. As expected, the shower shape discrimination and cone variables most sensitive to the dominant neutral pion background proved to be most important to the construction of the gaussian process. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66

3.9 Each discriminant was first transformed so that its distribution in data was approximately gaussian. The small number, $\epsilon = 10^{-5}$, avoided numerical singularities. . . . . . . . . . . . . . . . . . . . . . . . 67

3.10 The systematic deviation of the residuals from pure sampling uncertainties separated into three dominant behaviors, a constant deficit as negative $y$, a linear excess at central $y$, and a small deficit at positive $y$. A separate functional form was fit in each of these regions to isolate the systematic behavior from the noise, except in the last region where there are not enough bins to constrain any fit. . . . . . . . . . . . . 74

4.1 The posteriors for the total efficiency in each bin was modeled with independent beta distributions. . . . . . . . . . . . . . . . . . . . . 82

4.2 The cross section posterior was approximated as a multivariate gaussian distribution, with systematic model errors considered only for the uncorrelated variances. . . . . . . . . . . . . . . . . . . . . . . . . 85

4.3 The relative luminosity could not be satisfactorily calculated for some bunch crossings, which must then be removed from any analysis using polarization information. . . . . . . . . . . . . . . . . . . . . . . . 90

4.4 Four false asymmetries can arise from asymmetries in the experimental setup, but not in QCD. . . . . . . . . . . . . . . . . . . . . . . . . . . . 93

4.5 As with the unpolarized cross section, the asymmetry was approximated as a multivariate gaussian while systematic errors were either perfectly uncorrelated or correlated. Note that the elements of the covariance matrix are scaled by $10^3$ for display. . . . . . . . . . . . . . . 95

A.1 Each fundamental operator in QCD contributes to the twist of the forward Compton amplitude based on its operator structure. . . . . . 99

C.1 Transformations and marginalization are trivial to implement once a distribution has been represented by samples. . . . . . . . . . . . . . . 119

# Chapter 1

# Why Does the Proton Spin?

Like the electron, the proton is ubiquitous in the observable universe, essential to structure on nuclear, atomic, and even cosmological scales. Both particles have no obvious structure of their own and can be entirely characterized by a simple set of quantum numbers such as mass, spin, and coupling to external fields. Unlike the electron, however, the proton is not an elementary particle – the simple global description emerges from a rich internal structure.

The model of the proton has evolved with the increasing precision of experimental probes, early theories based on static quark configurations superseded by the ascent of a full quantum field theory as demanded by the measurements. Investigating the origin of the proton spin has been particularly critical to revealing the depth and complexity of its internal structure.

Today the development of collider technology allows for new means of studying the proton and its constituents, namely the gluons that bind the particle together. Measuring the aftermath of high-energy collisions between polarized protons directly probes the gluon polarization and pushes the understanding of the proton spin further. In particular, the rare high-energy photons that emanate from these collisions are uniquely sensitive to how the gluons spin.

# 1.1 Modeling the Proton

## 1.1.1 Static Quarks

By the 1950s particle physics faced the daunting task of classifying the myriad of particles being discovered in cosmic ray and accelerator experiments. Each of the baryons could be described by two quantum numbers, charge and strangeness, with systematic patterns that intimated an underlying structure [11]. Gell-Mann offered the first successful classification scheme with his Eightfold Way.

Each baryon, he posited, fell into an irreducible representation of $SU(3)$ with charge and strangeness given by the two Casimir invariants of the group. This approach offered eight and ten dimensional mulitplets with charge and strangeness patterns that naturally described the mess of hadrons and mesons in the literature. Holes in the patterns were quickly filled as experimentalists continued to find more and more particles.

A seductive feature of Gell-Mann's use of group theory was that the irreducible representations could be built out of repeated application of the three dimensional, fundamental representations of $SU(3)$ [12]. The one, eight, and ten dimensional hadron multiplets, for example, are the product of three fundamental representations,

$$3 \otimes 3 \otimes 3 = 10 \oplus 8 \oplus 8 \oplus 1.$$

Likewise, the one and eight dimensional meson multiplets could be produced from a fundamental representation and its conjugate,

$$3 \otimes \bar{3} = 8 \oplus 1.$$

While first considered a mathematical curiosity, the fundamental representations quickly became a speculative playground. Were they representative of physical states, the entire spectrum of hadrons and mesons would descend from only three elementary particles frivolously dubbed quarks. The physical interpretation became even more alluring when spin was added to the picture.

Endowing the fundamental representations with spin, the total spin of each composite hadron falls into tensor products of the $SU(2)$ representations for each quark:

$$\frac{1}{2} \otimes \frac{1}{2} \otimes \frac{1}{2} = \frac{3}{2} \oplus \frac{1}{2}$$

for baryons and

$$\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0$$

for mesons. Not only do the irreducible representations of these tensor products parallel the spin spectrum of observed hadrons, their structure also introduces relationships between the baryon magnetic moments that closely matches that seen in the laboratory.

The so-called static quark model proved to be dramatically successful in the decades that followed.

### 1.1.2 To Infinity and Bjorken: Deep Inelastic Scattering and the Parton Model

As collider technology became more sophisticated, scattering experiments advanced beyond the global structure of the hadron spectrum and began to probe the internal structure of the particles. These probes were highly relativistic, however, in contrast to the inherently non-relativistic approximations of the static quark model . Would the success of the quark model carry over to this new deeply inelastic scattering regime?

In deep inelastic scattering (DIS) a beam of high energy leptons probes a nucleon, here a proton, target [13, 14]. The kinematics of inclusive DIS (Figure 1-1), where

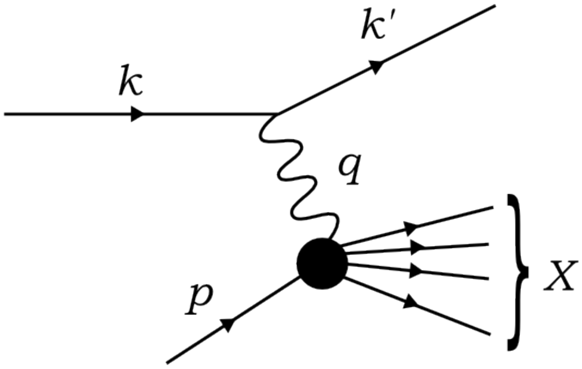


Figure 1-1: Leading-order description of deep inelastic scattering: an incident lepton emits a virtual photon that is absorbed by a nucleon target.

none of the proton remnants are measured, is fully described by two Lorentz invariant quantities: the momentum transfer

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

and the Bjorken scaling variable

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

Summing over all possible hadronic final states $X$, the unpolarized cross section for the interaction is given by

$$\begin{aligned} \mathrm{d}\sigma &\propto \sum_{\lambda\lambda'\sigma} \sum_{X} (2\pi)^4 \, \delta^4(k + p - k' - p_X) \, |\mathcal{M}_{\lambda\lambda'\sigma}|^2 \\ &\propto \sum_{\lambda\lambda'\sigma} \sum_{X} (2\pi)^4 \, \delta^4(q + p - p_X) \, |\mathcal{M}_{\lambda\lambda'\sigma}|^2 \\ &\propto \sum_{\lambda\lambda'\sigma} \sum_{X} \int \mathrm{d}^4\xi \; e^{i(q+p-p_X)\cdot\xi} \, |\mathcal{M}_{\lambda\lambda'\sigma}|^2 \,. \end{aligned}$$

where $\lambda$, $\lambda'$, and $\sigma$ are the polarizations of the initial state lepton, final state lepton, and nucleon target, respectively.

At the lowest order in perturbation theory, the amplitude $\mathcal{M}_{\lambda\lambda'\sigma}$ is of the form

$$\mathcal{M}_{\lambda\lambda'\sigma} \propto \langle k, \lambda | e\gamma^{\mu} | k', \lambda' \rangle \frac{g_{\mu\nu}}{q^2} \langle p, \sigma | e J^{\nu}_{\mathrm{EM}}(0) \, | X \rangle$$

where $J^{\nu}_{\mathrm{EM}}(0)$ is the electromagnetic current of the proton evaluated at the point of contact, taken here as $\vec{x} = \vec{0}$ without loss of generality.

Substituting this amplitude, the cross section becomes

$$\begin{aligned}
\mathrm{d}\sigma &\propto \sum_{\lambda\lambda'\sigma}\sum_{X}\int \mathrm{d}^4\xi\, e^{i(q+p-p_X)\cdot\xi}\left|\langle k,\lambda|e\gamma^{\mu}|k',\lambda'\rangle\frac{g_{\mu\nu}}{q^2}\langle p,\sigma|eJ^{\nu}_{\mathrm{EM}}(0)\,|X\rangle\right|^2 \\
&\propto \sum_{\lambda\lambda'\sigma}\sum_{X}\int \mathrm{d}^4\xi\, e^{i(q+p-p_X)\cdot\xi}\times \\
&\qquad [\langle k,\lambda|\gamma^{\mu}|k',\lambda'\rangle\langle k',\lambda'|\gamma^{\nu}|k,\lambda\rangle]\,\frac{e^2}{Q^4}\,\left[\langle p,\sigma|J^{\mathrm{EM}}_{\mu}(0)\,|X\rangle\langle X|J^{\mathrm{EM}}_{\nu}(0)\,|p,\sigma\rangle\right] \\
&\propto \left[\sum_{\lambda\lambda'}\langle k,\lambda|\gamma^{\mu}|k',\lambda'\rangle\langle k',\lambda'|\gamma^{\nu}|k,\lambda\rangle\right]\times\frac{e^2}{Q^4}\times \\
&\qquad \left[\sum_{\sigma}\sum_{X}\int \mathrm{d}^4\xi\, e^{i(q+p-p_X)\cdot\xi}\langle p,\sigma|J^{\mathrm{EM}}_{\mu}(0)\,|X\rangle\langle X|J^{\mathrm{EM}}_{\nu}(0)\,|p,\sigma\rangle\right]
\end{aligned}$$

or

$$d\sigma \propto l^{\mu\nu}\times\frac{e^2}{Q^4}\times W_{\mu\nu}.$$

The leptonic tensor,

$$l_{\mu\nu} = \sum_{\lambda\lambda'}\langle e|\gamma_{\mu}|e'\rangle\langle e'|\gamma_{\nu}|e\rangle,$$

is readily computed with Dirac spinor algebra, but the hadronic tensor encoding the non-perturbative structure of the proton,

$$W_{\mu\nu} = \frac{1}{4\pi}\sum_{\sigma}\sum_{X}\int \mathrm{d}^4\xi\, e^{i(q+p-p_X)\cdot\xi}\langle p,\sigma|J^{\mathrm{EM}}_{\mu}(0)\,|X\rangle\langle X|J^{\mathrm{EM}}_{\nu}(0)\,|p,\sigma\rangle,$$

requires a bit more care.

Note that, although the electromagnetic interaction is local in position space, the proton and probing photon are in momentum eigenstates and the exact position of the interaction is smeared out by the uncertainty principle: the photon probes the entire neighborhood of the proton instantaneously. This non-locality is implied by the delta function enforcing energy and momentum and can be made more explicit by factoring the exponential,

$$\begin{aligned} W_{\mu\nu} &= \frac{1}{4\pi} \sum_{\sigma} \sum_{X} \int \mathrm{d}^4\xi \, e^{i(q+p-p_X)\cdot\xi} \langle p, \sigma | J_{\mu}^{\mathrm{EM}}(0) \, | X \rangle \langle X | J_{\nu}^{\mathrm{EM}}(0) \, | p, \sigma \rangle \\ &= \frac{1}{4\pi} \sum_{\sigma} \sum_{X} \int \mathrm{d}^4\xi \, e^{iq\cdot\xi} \, e^{(p-p_X)\cdot\xi} \langle p, \sigma | J_{\mu}^{\mathrm{EM}}(0) \, | X \rangle \langle X | J_{\nu}^{\mathrm{EM}}(0) \, | p, \sigma \rangle , \end{aligned}$$

to give a translation operator that moves the position of one vertex relative to the other,

$$\begin{aligned} W_{\mu\nu} &= \frac{1}{4\pi} \sum_{\sigma} \sum_{X} \int \mathrm{d}^4\xi \, e^{iq\cdot\xi} \langle p, \sigma | e^{(p-p_X)\cdot\xi} J_{\mu}^{\mathrm{EM}}(0) \, | X \rangle \langle X | J_{\nu}^{\mathrm{EM}}(0) \, | p, \sigma \rangle \\ &= \frac{1}{4\pi} \sum_{\sigma} \sum_{X} \int \mathrm{d}^4\xi \, e^{iq\cdot\xi} \langle p, \sigma | J_{\mu}^{\mathrm{EM}}(\xi) \, | X \rangle \langle X | J_{\nu}^{\mathrm{EM}}(0) \, | p, \sigma \rangle . \end{aligned}$$

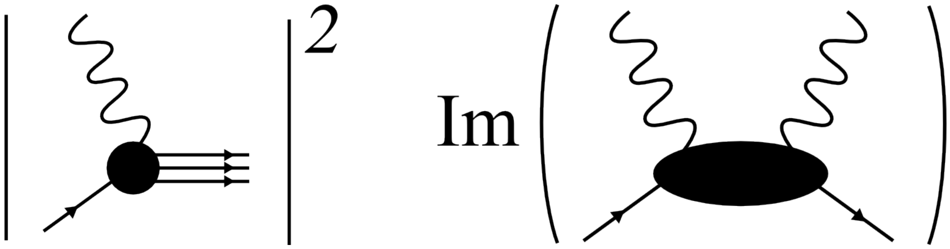
Summing over the proton spin states then yields

$$W_{\mu\nu} = \frac{1}{4\pi} \sum_{X} \int \mathrm{d}^4\xi \, e^{iq\cdot\xi} \langle p | J_{\mu}^{\mathrm{EM}}(\xi) \, | X \rangle \langle X | J_{\nu}^{\mathrm{EM}}(0) \, | p \rangle .$$

Finally, by appealing to the optical theorem, the summation over all possible final states can be replaced with the imaginary part of the forward Compton amplitude,

$$W_{\mu\nu} = \frac{1}{4\pi} \int d^4\xi \, e^{iq\cdot\xi} \langle p | [ J_{\mu}^{\mathrm{EM}}(\xi) \, , J_{\nu}^{\mathrm{EM}}(0) ] | p \rangle .$$

This forward structure of the hadronic tensor is both mathematically and conceptually simpler: the proton inelastically scattering off of the incident photon is replaced by the elastic absorption and emission of the photon (Figure 1-2).



(a) (b)

Figure 1-2: The hadronic tensor simplifies dramatically when an application of the optical theorem replaces the squared photon-proton vertex (a) with the imaginary part of the forward Compton amplitude (b), where the proton absorbs a photon at $\vec{x} = \vec{0}$ before emitting a photon at $\vec{x} = \vec{\xi}$.

**The Infinite Momentum Frame**

The structure of the hadronic tensor simplifies dramatically for an observer traveling collinear to the photon probe at the speed of light. Instead of taking the limit of an infinite boost along $x^3$, however, consider the change of variables [15]

$$\begin{aligned} x^+ &= \tfrac{1}{\sqrt{2}}\left(x_0 + x_3\right) \\ x^- &= \tfrac{1}{\sqrt{2}}\left(x_0 - x_3\right) \ . \\ x^\perp &= \left(x_1, x_2\right) \end{aligned}$$

Taking $x^+$ as a time coordinate with $x^-$ as the position orthogonal to $x^\perp$ produces dynamics equivalent to those in an infinite momentum frame without having to explicitly consider the limiting boost. These light-cone coordinates directly parameterize the null geodesics, or the light cone, from the initial photon-proton interaction at $\vec{x} = 0$.

In the infinite momentum frame the photon four-momentum becomes

$$q^+ = -\frac{Mx}{\sqrt{2}},\ q^\perp = 0,\ q^- = \frac{Q^2}{\sqrt{2}Mx}.$$

Deep inelastic scattering is now formally defined as the Bjorken limit of the lepton-nucleon interaction, where $x$ remains constant as $Q^2 \to \infty$. On the light cone the energy $q^+$ remains constant while the momentum $q^-$ diverges.

The consequence for the hadronic tensor is immediate when the integral is taken in light cone coordinates,

$$W_{\mu\nu} = \frac{1}{4\pi} \int \mathrm{d}\xi^{-} \, e^{iq^{+}\xi^{-}} \int \mathrm{d}\xi^{+} \mathrm{d}^{2}\xi_{\perp} \, e^{iq^{-}\xi^{+}} \langle p | \left[ J_{\mu}^{\mathrm{EM}}(\xi) \, , J_{\nu}^{\mathrm{EM}}(0) \right] | p \rangle .$$

Because $q^{-}$ diverges, the exponential weight $\exp\left(iq^{-}\xi^{+}\right)$ rapidly oscillates and the integral vanishes unless $\xi^{+} \to 0$; consequently the hadronic tensor is dominated by the singular behavior of the correlation function at $\xi^{+} = 0$. Although the hadronic tensor is inherently non-local, at high energies its structure is determined largely by the physics of vanishing displacements on the light cone.

Given the particular behavior of the singularity, the operator product expansion [16, 17] can now be applied to replace the correlation function with an infinite sum of non-singular, local operators,

$$\mathcal{O}_{1}\left(\xi^{+}\right) \mathcal{O}_{2}\left(0\right) \to \sum_{n} C_{12}^{n}\left(\xi^{+}\right) \mathcal{O}_{n}\left(0\right) ,$$

where all of the singular behavior of the correlation function as $\xi^{+} \to 0$ is encoded in the coefficients of the expansion.

Each term in the operator product expansion is ordered by twist,

$$t = d - s$$

where $d$ is the mass dimension of the operator $\mathcal{O}_{n}$ and $s$ is its spin. As evident from dimensional analysis, the higher twist operators must carry additional factors of $Q^{2}$ in their coefficients and are consequently dramatically suppressed when the momentum transfer is large. The operator product expansion formalizes the lightcone dominance of the hadronic tensor in the Bjorken limit.

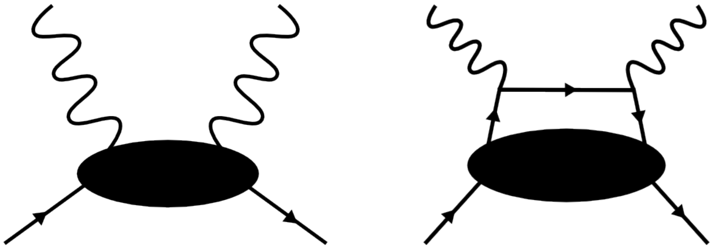

(a) (b)

Figure 1-3: In the parton model the incident photon scatters not from the proton as a whole but from a point particle and the leading twist behavior of the general forward Compton amplitude (a) reduces to the "handbag" diagram (b).

### The Parton Model

Motivated by the static quark model, the parton model assumes that the proton is comprised of a sea of non-interacting point particles known as partons. In this model the incident photon scatters from a single parton (Figure 1-3) and the leading twist contribution to the hadronic tensor becomes [18]

$$W_{\mu\nu} \propto \sum_i \int \mathrm{d}^4 k \, e_i^2 f_i(p,k) \, w_{\mu\nu}(q,k) \, \delta\big[(k+q)^2\big]$$

where the $i$ runs over quark flavors and $w_{\mu\nu}$ describes the vertex of the virtual photon and a massless quark with momentum $k$.

Now consider the fraction of light cone momentum carried by each quark, $x_i = k^+/p^+$. Expanding the argument of the delta function,

$$\begin{aligned} (k+q)^2 &= 2k \cdot q + 2q^+ q^- \\ &= 2x_i p \cdot q + 2q^+ q^- \\ &= x_i \frac{Q^2}{x} - Q^2 \\ &= \frac{Q^2}{x}\left(x_i - x\right), \end{aligned}$$

gives

$$\delta\big[(k+q)^2\big] = \frac{x}{Q^2}\delta(x_i - x) \, .$$

On the light cone the Bjorken scaling variable $x$ is not only a convenient relativistic invariant but also assumes a neat physical interpretation as the fraction of light cone momentum, $x = k^+/p^+$, of the struck parton.

With this new interpretation of $x$, the hadronic tensor becomes

$$W_{\mu\nu} \propto \sum_i \int \mathrm{d}^4k \, f_i(p,k) \, w_{\mu\nu}(q,k) \, \frac{x}{Q^2} \delta(x_i - x)$$

or, after some algebra,

$$W_{\mu\nu} \propto \sum_i \frac{2e_i^2}{Q^2} \int \mathrm{d}x \, w_{\mu\nu}(q,k) \, q_i(x) \, \delta(x_i - x) \, ,$$

where the quark distribution function $q_i(x)$ is the probability of finding a parton of type $i$ with momentum fraction $x = k^+/p^+$ in the proton and $w_{\mu\nu}$ is the squared vertex of a photon-parton interaction.

Taking the partons as spin 1/2 quarks, $w_{\mu\nu}$ is readily calculated and the resulting hadronic tensor becomes independent of the momentum transfer $Q^2$, scaling with only $x$. Observation of this "Bjorken scaling" in deep inelastic scattering experiments provided the first strong validation of a physical interpretation to the quark model; quarks were not simply a mathematical convenience but true constituents of the proton.

### The Ellis-Jaffe Sum Rule

Spin enters into the deep inelastic scattering framework when polarized processes are considered. Following the unpolarized reasoning but without the sum over the hadronic spin states yields a polarized hadronic tensor $W^{\sigma}_{\mu\nu}$ and the polarized quark distribution functions,

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

where $\uparrow$ and $\downarrow$ denote the spin projection along and opposite $x^{-}$, respectively. The total spin[1] of the quarks within the proton is then given by

$$\Sigma = \sum_i \int dx \, \Delta q_i(x) \, .$$

Enforcing $SU(3)$ flavor symmetry and assuming that the polarization of strange quarks is negligible, Ellis and Jaffe derived a sum rule for the polarized quark distributions in terms of the proton axial charge [19]

$$\begin{aligned} \int \mathrm{d}x \, g_1^p(x) &\equiv \frac{1}{2} \left( \sum_i \int \mathrm{d}x \, e_i^2 \, \Delta q_i(x) \right) \\ &\approx 0.15 \, g_A . \end{aligned}$$

Taking $g_A$ extracted from beta decay measurements, the Ellis-Jaffe sum rule gives [2]

$$\int \mathrm{d}x \, g_1^p \approx 0.17$$

with

$$\Sigma \approx 0.60,$$

significantly less than the non-relativistic result of unity from the static quark model.

While Ellis and Jaffe immediately noted the possible weakness in their assumptions, the deviation of their sum rule from the static quark model verified the importance of relativistic effects and substantiated more complex models of the proton [21].

### The Spin Crisis

Not long after Ellis and Jaffe offered their sum rule, experiments across the world began to measure asymmetries sensitive to the polarized distribution functions [22].

---

[1] Technically the polarized distribution functions define helicity with respect to the direction of the incident photon and not the spin along an arbitrary axis.

[2] For completeness the results here include small corrections from a complete QCD treatment. See, for example, [20].

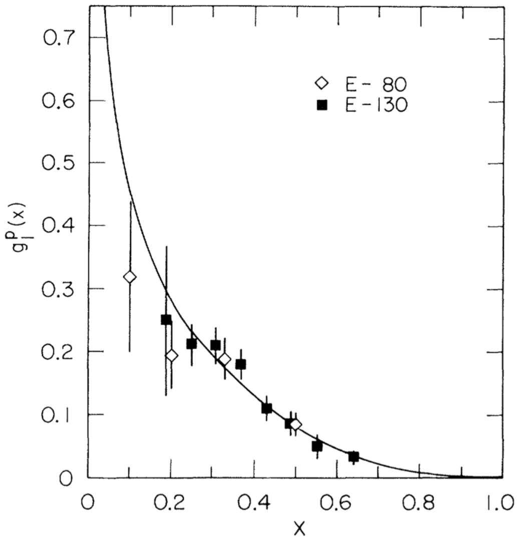


Figure 1-4: Results from the E80 and E130 asymmetry measurements. The curve extrapolates to low $x$ with motivation from Regge theory and integrates to give $\int \mathrm{d}x\, g_1^p = 0.17 \pm 0.05$, in agreement with the Ellis-Jaffe sum rule [1].

The E80 and E130 experiments at SLAC produced the first results [1], measuring asymmetries from the collision of polarized electrons and polarized butanol targets to constrain $g_1^p(x)$ at high $x$ (Figure 1-4). These early experimental results were consistent with the Ellis-Jaffe sum rule but the precision was not sufficient to provide a strong test.

The EMC collaboration pushed low into $x$, utilizing high energy muons on various proton targets [2]. Contrary to the earlier extrapolations, the measured contributions from lower $x$ proved to be insufficient and measurements began to deviate from the Ellis-Jaffe prediction (Figure 1-5).

Adding the total statistical and systematic uncertainties in quadrature, the first measurement published by EMC gave

$$\int \mathrm{d}x\, g_1^p = 0.126 \pm 0.018,$$

which implies a quark polarization of

$$\Sigma = 0.13 \pm 0.19.$$

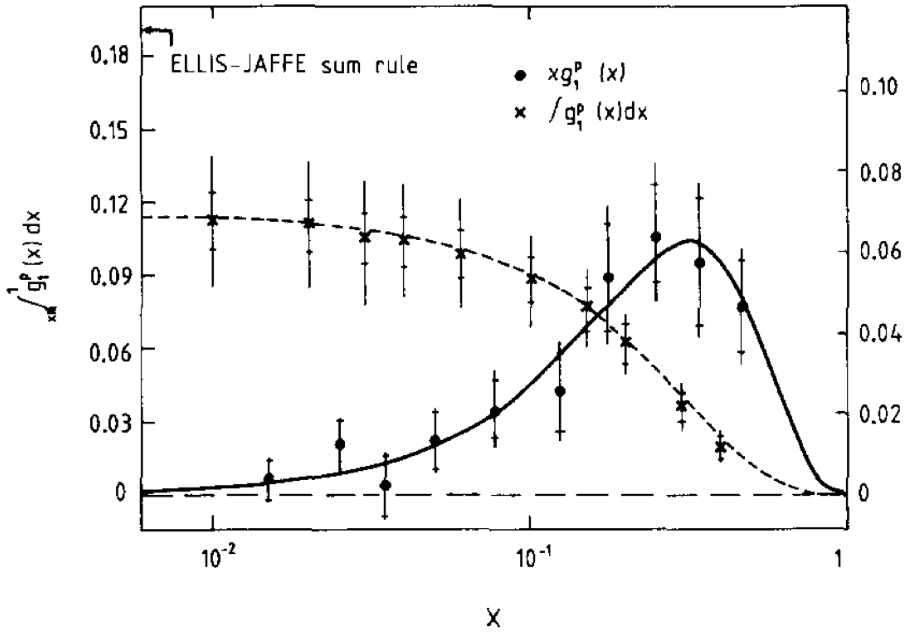


Figure 1-5: Results from the EMC asymmetry measurements. The measured $g_1^{ep}$ falls below the E80/E130 extrapolations at small $x$, with $\int dx\, g_1^{ep}$ then deviating from the Ellis-Jaffe result [2].

Not only did the measurement conflict with the Ellis-Jaffe sum rule, it was consistent with a proton where the quarks were not polarized at all!

Soon afterwards the SMC collaboration provided a second measurement [23] that pushed the quark polarization higher,

$$\int \mathrm{d}x\, g_1^p = 0.136 \pm 0.015$$

with

$$\Sigma = 0.22 \pm 0.10,$$

and helped to alleviate the concern of fully unpolarized quarks. Although the measurements were no longer sufficiently extreme to be worthy of a crisis, the surprisingly small quark polarization they implied left many wondering about the residual spin puzzle: from where does the remaining spin of the proton originate?

### 1.1.3 Quantum Chromodynamics

While the asymmetries were being measured, unpolarized physics was confirming the success of quantum chromodynamics (QCD) as a model of the strong interaction binding together hadrons. An $SU(3)$ gauge theory with quark and gluon degrees of freedom, QCD [17] is characterized by the Lagrangian density

$$\mathcal{L} = \bar{\psi}_k \left( i\gamma^\mu D_\mu - m_k \right) \psi_k - \frac{1}{2} F_a^{\mu\nu} F_{\mu\nu}^a$$

where $D_\mu$ is the gauge covariant derivative

$$D_\mu = \partial_\mu - igT^a A_\mu^a$$

and $F_a^{\mu\nu}$ is the set of color gauge connections

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f_{abc} A_\mu^b A_v^c.$$

As with quantum electrodynamics, we can define gauge electric and magnetic fields. Note that there will be eight pairs of fields, one for each color index $a$,

$$E_a^i = F_a^{0i}$$

and

$$B_a^i = \epsilon^{ijk} F_a^{jk}$$

To unburden the notation, the color indices $a$ will be dropped from here on in, with a sum over the color indices implied by the trace operator, e.g.

$$\mathrm{Tr} F^{\mu\nu} F_{\mu\nu} \equiv F_a^{\mu\nu} F_{\mu\nu}^a.$$

Now the classical QCD Lagrangian admits a gauge-invariant energy-momentum tensor

$$T^{\mu\nu} = \frac{i}{4}\bar{\psi}_k \left(\gamma^\mu D^\nu + \gamma^\nu D^\mu\right) \psi_k + \text{h.c.} + \text{Tr}\left(F^{\mu\alpha}F^{\nu}_{\alpha} - \frac{1}{4}g^{\mu\nu}F^{\alpha\beta}F_{\alpha\beta}\right)$$

that is both conserved, $\partial_\mu T^{\mu\nu} = 0$ and symmetric $T^{\mu\nu} = T^{\nu\mu}$.

Likewise, a conserved angular momentum tensor can be defined

$$M^{\alpha\mu\nu} = T^{\alpha\nu}x^\mu - T^{\alpha\mu}x^\nu$$

with the conserved angular momentum current

$$J^i = \frac{1}{2}\epsilon^{ijk}\int d^3x\, M^{0jk}.$$

After appropriately quantizing QCD, the angular momentum current $\mathbf{J}$ becomes a local operator whose expectation value in a polarized proton state must equal the total spin of the proton. Examining the structure of the current, in particular any factorization between the quark and gluon operators, offers a new approach to understanding the spin puzzle. Relating this new perspective to the parton model, however, requires a careful understanding of the light-cone behavior of both.

### Resolving the Spin Puzzle With QCD

In order to avoid complications arising from unphysical degrees of freedom, the canonical quantization of QCD is typically performed in a physical gauge. This quantization defines the quark and gluon degrees of freedom of theory and, consequently, the fundamental constituents of the proton.

Quantizing in $A^0 = 0$ gauge, the angular momentum operator factors into three gauge-invariant, local operators [24]:

$$\hat{J}^3 = \frac{1}{2}\hat{\Sigma} + \hat{L}_q + \hat{J}_g,$$

where

$$\begin{aligned}\hat{\Sigma} &= \int d^3x\, \bar{\psi}_k \gamma^3 \gamma_5 \psi_k \\ \hat{L}_q &= \int d^3x\, \bar{\psi}_k \left(\mathbf{x} \times i\mathbf{D}\right)^3 \psi_k \\ \hat{J}_g &= \int d^3x\, \bar{\psi}_k \left(\mathbf{x} \times (\mathrm{Tr}(\mathbf{E} \times \mathbf{B}))\right)^3 \psi_k\end{aligned}$$

The first term, $\hat{\Sigma}$, generates internal rotations of the quark fields and is therefore identified with the contribution of quark spin to the total angular momentum operator. Likewise, $\hat{L}_q$ generates spatial rotations of the quark fields while $\hat{J}_g$ generates both internal and spatial rotations of the gluon fields[3]. Care must be taken with the interpretation of these operators, however, as the interactions concealed in the covariant derivatives mix the quark and gluon fields under each of the rotations.

Applying this factorization to a proton polarized along the arbitrary aixs $x^3$ gives

$$\langle p, \sigma | \hat{J}^3 | p, S \rangle = \langle p, S | \left( \frac{1}{2}\hat{\Sigma} + \hat{L}_q + \hat{J}_g \right) | p, \sigma \rangle$$

or

$$\frac{1}{2} = \frac{1}{2}\langle \hat{\Sigma} \rangle + \langle \hat{L}_q \rangle + \langle \hat{J}_g \rangle.$$

In addition to the quarks, QCD introduces contributions from both the gluons and the dynamic angular motion of the two fields.

Likewise, QCD can be quantized on the light cone with $A^+ = 0$. In the infinite momentum frame,

$$J^- = \int d^2x_\perp \, dx_- \, M^{+12},$$

which factors [20, 28] into

$$J^- = \frac{1}{2}\hat{\Sigma} + \hat{\mathcal{L}}_q + \Delta\hat{G} + \hat{\mathcal{L}}_g$$

---

[3] Recent research has suggested that this decomposition, in particular $\hat{J}_g$, can be further separated into frame and gauge invariant terms [25, 26, 27], but the work is still ongoing with many details still to be worked out

where

$$\begin{aligned}
\hat{\Sigma} &= \int d^2x_\perp \, dx_- \, \bar{\psi}_k^+ \gamma_5 \psi_k^+ \\
\hat{\mathcal{L}}_q &= \int d^2x_\perp \, dx_- \, \bar{\psi}_k^+ \left( \vec{x} \times i\vec{\mathcal{D}} \right)^- \psi_k^+ \\
\Delta\hat{G} &= \int d^2x_\perp \, dx_- \, \mathrm{Tr}(\epsilon^{+-ij} F^{+i} A^j) \\
\hat{\mathcal{L}}_g &= \int d^2x_\perp \, dx_- \, \mathrm{Tr}\left( 2F^{+j} \left( \vec{x} \times i\vec{\mathcal{D}} \right)^- A^j \right).
\end{aligned}$$

Note that the covariant derivatives $\vec{\mathcal{D}}$ are in terms of a residual light cone gauge and are not the full covariant derivatives of QCD [28].

As in the previous factorization, each of these terms is a gauge-invariant, local operator that generates internal and spatial rotations of the quark and gluon fields. The quark spin operator $\hat{\Sigma}$ is identical, modulo the coordinate transformation, but on the light cone the quark angular momentum slightly simplifies and the internal and external rotations of the gluons separate.

While these factorizations elucidate the operator structure of each quantization, it still remains to be seen how these operators, and their underlying quantizations, relate to the parton distribution functions in DIS.

### QCD and the Parton Model

If QCD does indeed define the physics within the proton then the light cone behavior of the polarized hadronic tensor, $W^{\sigma}_{\mu\nu}$, should be described entirely by the theory.

Because of asymptotic freedom, the incident photon with large momentum transfer interacts with only weakly coupled degrees of freedom, and a parton-model-like factorization between the soft physics of the proton structure and the perturbative physics of the hard photon-quark interaction should be possible. With a theory as sophisticated as QCD, however, a formal separation of the two regimes is incredibly complex, especially considering the canceling divergences arising at higher orders in the perturbative expansion of the photon-proton interaction.

Formal factorization procedures [29] introduce an energy scale, $\mu_F$, absorbing divergences below this scale into non-perturbative distribution functions for the quark and gluon degrees of freedom while leaving the rest to the perturbative calculation. Provided that the details have been handled correctly, these factorized divergences cancel exactly and admit well-behaved calculations. Remarkably, the distribution functions defined this way also manifest universality: factorizing different interactions yields the same distribution functions for the same target nucleon.

Requiring that any physical observable is independent of this scale, however, induces a relationship between the distribution functions $q_i(x, \mu)$ in the model. The resulting DGLAP evolution equations [14] convolve the distributions when the scale is varied and introduce small corrections dependent on the momentum transfer of the interaction. Note that this convolution includes all distribution functions allowing, for example, the gluon distribution function to be constrained with the $Q^2$ dependence of DIS observables even though the leptonic probe does not couple to the gluons directly.

With the introduction of a factorization scale $\mu_F$ the form of the hadronic tensor in QCD mirrors that of the parton model,

$$W^{\sigma}_{\mu\nu} \propto \sum_i e_i^2 \int \mathrm{d}x \, w^{\sigma}_{\mu\nu}(q, k, \mu_F) \, \Delta q_i(x, \mu_F) \, \delta(x_i - x) \, .$$

From the perspective of a quantum field theory, the distribution function is the amplitude for creating a definite helicity quark and propagating it along the light cone before annihilation. In QCD this amplitude is given by the quark bilinear [30]

$$\Delta q_i(x, \mu) = \frac{1}{2} \int \frac{\mathrm{d}\lambda}{2\pi} e^{i\lambda x} \, \left. \langle p, \sigma | \bar{\psi}_i(0) \, U(0, \lambda n) \, \gamma^{\mu} n_{\mu} \gamma_5 \psi_i(\lambda \vec{n}) \, | p, \sigma \rangle \right|_{\mu}$$

where $\vec{n}$ is a null vector and $U(0, \lambda n)$ is a Wilson link ensuring gauge invariance through the propagation. The renormalization scale, $\mu$, of the bilinear is equivalent to the factorization scale $\mu_F$ defining the soft physics encoded in the distribution.

Although the bilinear is inherently a non-linear correlation, its moments over $x$ are local [31]. In particular, the first moment of the quark distribution function is equal to the expectation of the quark spin operator from the QCD angular momentum current,

$$\Sigma \equiv \langle \hat{\Sigma} \rangle = \int \mathrm{d}x \, \Delta q(x, \mu) \, .$$

The moments of the gluon distribution are more subtle. As with the quarks, the gluon distribution is given by the amplitude for the creation, propagation, and annihilation of a definite helicity gluon:

$$\Delta g(x, \mu) = \frac{i}{2x} \int \frac{\mathrm{d}\lambda}{2\pi} e^{i\lambda x} \; \langle p, \sigma | F^{+\alpha}(0) \, U(0, \lambda n) \, \tilde{F}^{+}_{\alpha}(\lambda n) \, | p, \sigma \rangle \Big|_{\mu} \, .$$

where $\tilde{F}_{\alpha\beta} = \frac{1}{2} \epsilon_{\alpha\beta\mu\nu} F^{\mu\nu}$. Because of the additional factor of $x$ in the denominator only the higher moments

$$\int \mathrm{d}x \, x^{n-1} \Delta g(x, \mu) \, , \; n > 1$$

are local. The first moment is non-local and cannot be identified with the expectation of any of the local QCD operators derived above.

There is, however, one exception. In light-cone gauge the first moment collapses to a local amplitude and can be associated with the gluon spin operator from the light-cone factorization,

$$\Delta G \equiv \langle \Delta \hat{G} \rangle = \int \mathrm{d}x \, \Delta g(x, \mu) \, .$$

Indeed, there is a particular correspondence between the light-cone gauge and the parton model. Defined at leading twist, partons are the degrees of freedom along the light cone. Just as light-cone coordinates are natural parameterizations of the light cone, light-cone quantization naturally defines the analogous degrees of freedom in

QCD and, as noted previously, light-cone quantization requires the light-cone gauge to project out any unphysical degrees of freedom.

For high energy probes the light-cone decomposition of the proton spin is natural, and contributions from the quark and gluon spin correspond to well behaved moments of the respective distribution functions. With the quark distribution already well constrained from DIS measurements, the next step in deciphering the spin puzzle is constraining the gluon distribution function.

## 1.2 Probing the Gluon Spin

Previous constraints of $\Delta g(x)$ from DIS experiments took advantage of DGLAP evolution distribution functions to infer $\Delta g(x)$ from the small violation of Bjorken scaling in the polarized quark distribution functions. Because of the small $Q^2$ coverage of these experiments, however, the resulting constraints on the gluon polarization are limited (Figure 1-6).

In order to constrain $\Delta g(x)$ directly, one has to replace the polarized electron beam in DIS, and its electromagnetic interaction, with a strongly interacting probe. Without free quarks, such measurements have became feasible only with the development of a polarized proton-proton collider; the correlations between hard interactions induced by a gluon in one of those protons with the initial proton helicities is immediately sensitive to the gluon polarization.

In particular the asymmetry between interactions with the two proton helicities aligned verses antialigned, or the double-helicity asymmetry[4],

$$A_{\mathrm{LL}} = \frac{\sigma^{\upharpoonleft\upharpoonright} - \sigma^{\upharpoonleft\downharpoonright}}{\sigma^{\upharpoonleft\upharpoonright} - \sigma^{\upharpoonleft\downharpoonright}}$$

---

[4] Double "spin" asymmetry is often used interchangeably, but as the sensitivity to $\Delta g(x)$ requires the protons to be in helicity eigenstates, helicity is the more appropriate term.

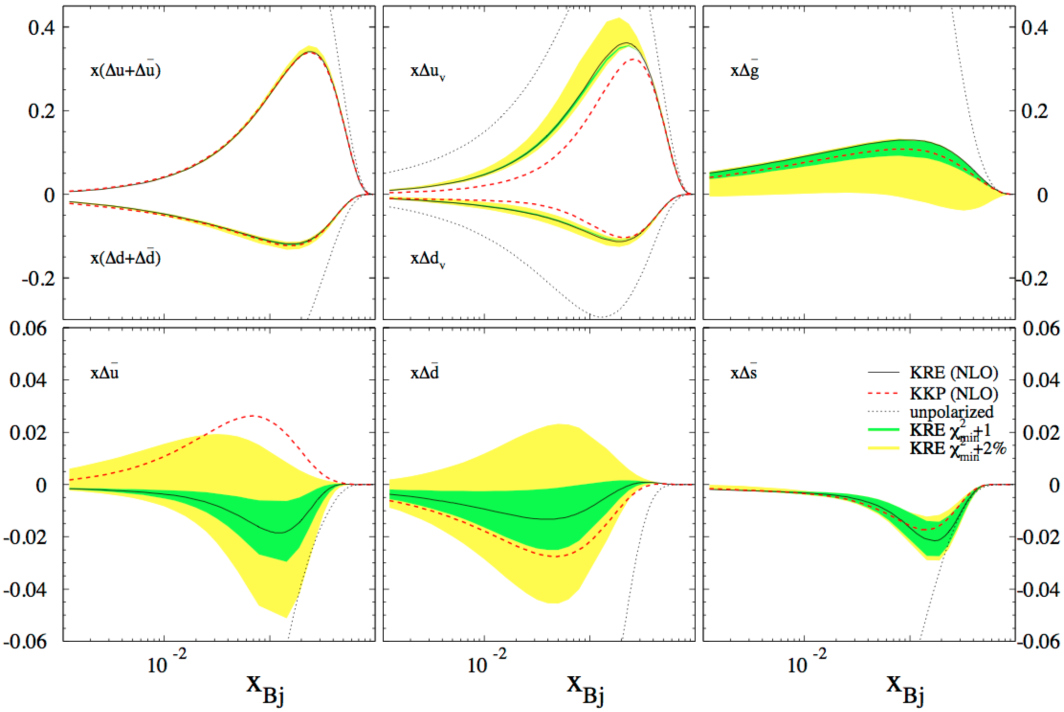


Figure 1-6: Quark (separated into valence and sea contributions) and gluon polarized distribution functions are inferred from a NLO global analysis of deep inelastic scattering data [3].

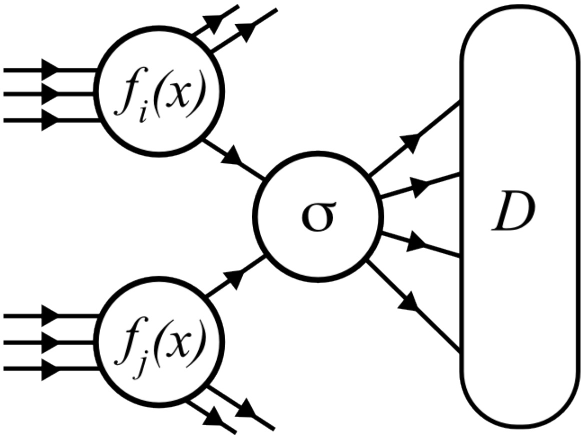


Figure 1-7: Cartoon of factorization in proton-proton collisions. Soft, non-perturbative physics is absorbed into the parton distribution functions $f_i$ and $f_j$ while harder perturbative physics gives the partonic cross section $\sigma_{ij}$. $D$ encodes the non-perturbative physics inherent in any fragmentation of final state partons.

provides a powerful probe of the gluon polarization [32]. Assuming a leading twist factorization, the leading order contribution to the double-helicity asymmetry becomes

$$A_{\mathrm{LL}} = \frac{\sum_{i,j} \int \mathrm{d}y \, \Delta f_i(y) \, \Delta f_j(y) \, \Delta\sigma_{ij}(y) \, D_f(y)}{\sum_{i,j} \int \mathrm{d}y \, f_i(y) \, f_j(y) \, \sigma_{ij}(y) \, D(y)}$$

where $D_f$ refers to any jet or hadron fragmentation in the final state and $y$ denotes any kinematic variables describing the hard interaction (Figure 1-7). At higher orders the dependence of the asymmetry on the polarized distribution functions becomes more complicated, but a global NLO analysis provides the necessary framework for extracting constraints from the measured asymmetry.

Note that, while the assumption is often taken for granted, no theoretical proof of factorization exists for arbitrary final states in proton-proton collisions. Recent applications of soft collinear effective theories [33] have shown that factorization in hadronic collisions is far more subtle than in DIS, requiring an additional scale between the traditional soft distribution functions and hard parton scattering. For what kinematics these corrections are relevant remain to be seen.

Of the many possible hard interactions to study, hadronic channels are the obvious first choice. Because of the large hadronic cross sections these channels are copiously

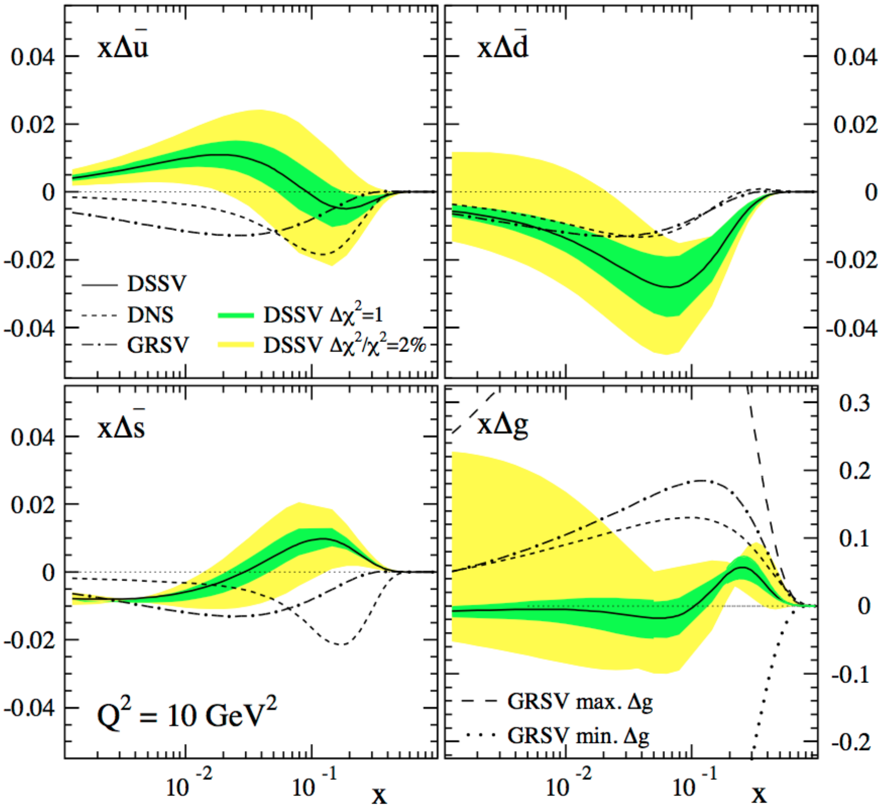
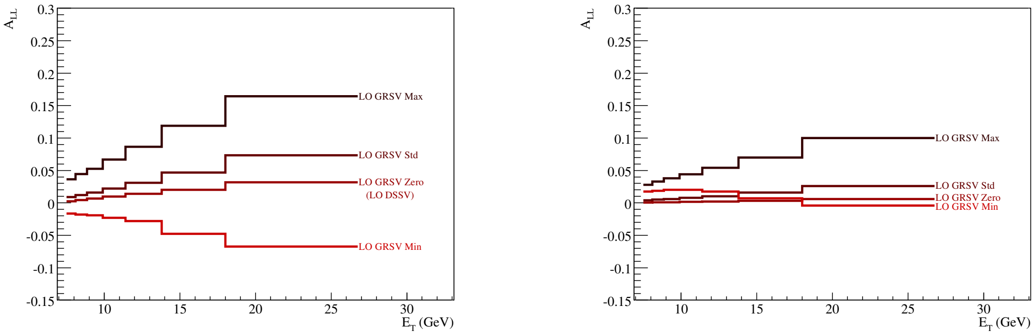


Figure 1-8: Sea and gluon polarized distribution functions inferred from a global NLO analysis of deep inelastic scattering with hadronic double-helicity asymmetry measurements [4]. Notice the significant improvement in the gluon polarization compared to Figure 1-6.

produced in the proton collisions and provide abundant statistics for the resulting asymmetry measurement. Indeed, hadronic measurements have dominated the early asymmetry results and provided the strongest constraints on $\Delta g(x)$ (Figure 1-8).

Eventually, however, the information given by additional data is lost in the systematic uncertainties inherent to hadronic measurements. Inclusive hadron measurements, for example, are dependent on uncertain fragmentation functions to connect back to the hard interaction and full jet reconstruction is limited by hadronization uncertainties.

Moreover, the asymmetry in these channels is diluted by contributions from interactions with no gluon at all. Because any hadronic final state could have come from a



(a) (b)

Figure 1-9: Leading-order predictions of the double-helicity asymmetry $A_{\mathrm{LL}}$ at $\sqrt{s} =$ 200 GeV and central rapidity, given various models of the gluon polarized distribution function [4, 5] for (a) direct photons and (b) jets. The direct photon asymmetry is not only larger than the jet asymmetry, but also provides better model discrimination as the different predictions do not overlap.

quark-quark, quark-gluon, or gluon-gluon interaction the double-helicity asymmetry becomes (schematically)

$$A_{\mathrm{LL}} \sim \left(\Delta q\right)^2 + \Delta q \Delta g + \left(\Delta g\right)^2 .$$

If $\Delta q(x)$ is large within the kinematics of the measurement then the sensitivity of the asymmetry to the gluon polarization weakens.

### 1.2.1 Gluon Gazing

There is one hard interaction without such dilution. At leading order, final state photons are produced through three possible interactions (Figure 1-10). Without an antiproton beam, the overall signal is dominated by quark-gluon Compton scattering and the photon asymmetry is dominated by the linear quark-gluon term,

$$A_{\mathrm{LL}} \sim \Delta q \Delta g,$$

significantly improving the sensitivity to $\Delta g$ (Figure 1-9).

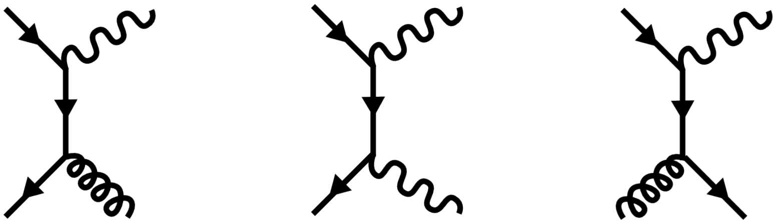


Figure 1-10: Leading-order direct photon diagrams; time runs from left to right. At moderate $x$ prompt photon production is dominated by the third diagram which is directly sensitive to the polarized gluon distribution function.

In addition to the unique sensitivity to $\Delta g(x)$, photons boast an extremely clean experimental signature. With no possible decays, photons propagate from the collision largely unimpeded until they interact electromagnetically with the detector, providing a direct probe of the hard interaction and, consequently, the initial state gluon. The final asymmetry is free of the systematics induced by the intricacies of the finals states in the hadronic channels.

The power of the photon as a probe of the gluon polarization comes at the price of an electromagnetic vertex in place of a strong vertex in the leading order diagram. Consequently, high energy photon production is dramatically suppressed relative to the hadronic channels,

$$\frac{\Gamma_{\text{photon}}}{\Gamma_{\text{hadron}}} \propto \left( \frac{\alpha_{\text{EM}}(Q^2)}{\alpha_{\text{s}}(Q^2)} \right)^2 \ll 1.$$

Not only are the photons rare, and hence statistically limited, but the suppression relative to the hadronic interactions drastically reduces the ability to isolate them from the overwhelming hadronic activity. The total photon yield is given by contributions from the final state, or prompt, photons and the photons produced by the decay of hadrons,

$$\sigma_\gamma = \sigma_{\text{prompt}} + \sigma_{\text{decay}}.$$

The final state photons are themselves comprised as photons from the hard interaction and those from the collinear radiation, or fragmentation, of photons from final state quarks [5] ,

$$\sigma_{\text{prompt}} = \sigma_{\text{direct}} + \sigma_{\text{frag}}.$$

Only the contribution from direct photons is sensitive to the gluon polarization and the remaining terms must be carefully excluded from the measurement lest the sensitivity of the asymmetry be diluted.

The decomposition of the "prompt" photons becomes theoretically unsound, however, when moving to higher-order calculations. Fragmentation photons are typically removed from calculations with the use of hadron isolation criteria, but naive isolation cuts spoil the stability of beyond-leading-order calculations by removing the soft gluon radiation necessary to cancel infrared divergences. Weaker isolation criteria admit stable calculations but cannot fully separate the fragmentation photons, introducing a dependence photon fragmentation functions and their accompanying uncertainties.

### The Frixione Isolation Criteria

Alternatively, Frixione proposed a more sophisticated definition of a direct photon at all orders in perturbation theory utilizing a soft isolation criteria [34],

$$\sum_i E_i \theta(\delta - \Delta R_i) \leq \mathcal{H}(\delta)$$

where $i$ runs over hadrons and $\Delta R_i$ is the angular distance between the $i$th hadron and the photon. The isolation function $\mathcal{H}(\delta)$ has the defining property

$$\lim_{\delta \to 0} \mathcal{H}(\delta) = 0$$

so that a hard cut is placed on only hadrons exactly collinear with the photon.

---

[5] Unfortunately, the definitions of "prompt" and "direct" are not consistent within the literature and the two are often exchanged, even amongst articles from the same author. The notation taken here will be followed consistently throughout this work.

With Frixione's isolation criteria, higher-order calculations of direct photon interactions, and consequently better $\Delta g(x)$ constraints from photon asymmetries, are feasible without the need for photon fragmentation functions. Progress has even been made towards proving factorization with these final-state photons [35].

Provided that they can be isolated from fragmentation and decay contributions experimentally, the direct photon channel as defined by Frixione offers a uniquely sensitive look into the gluon polarization.

### Theoretical Calculations

Resummed, next-to-leading order calculations of both the direct photon cross section and double-helicity asymmetry have been performed by Werner Vogelsang [36]. The unpolarized cross section calculation (Figure 1-11) admits a comparison to data and a crucial verification that the perturbative physics separating the final state photon from the initial state gluon are understood and do not confound the sensitivity of the asymmetry measurement. Once the unpolarized calculation has been validated, the asymmetry calculations provide the means for extracting the gluon polarization from the data.

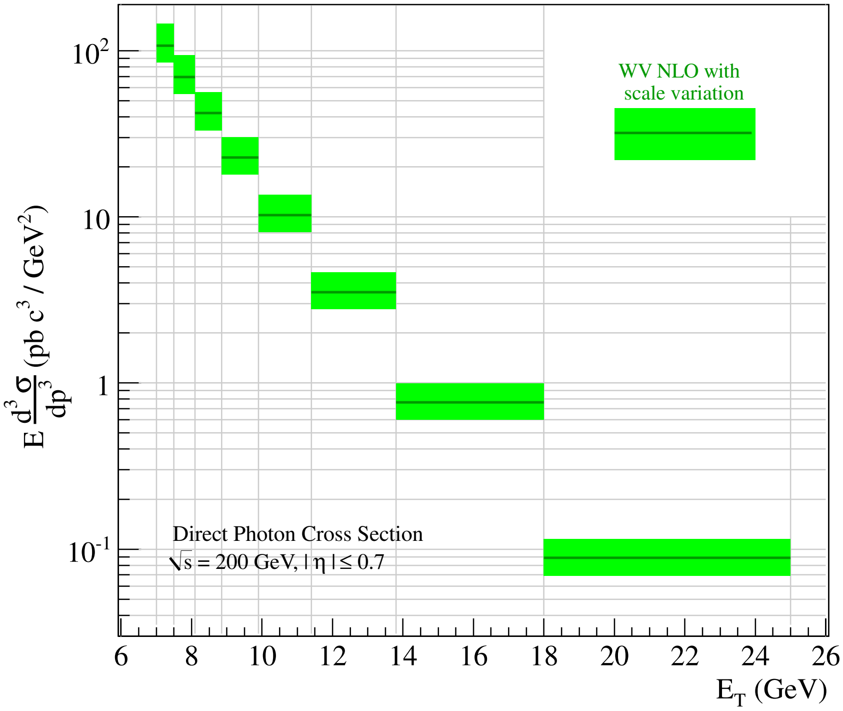


Figure 1-11: Full resummed, next-to-leading-order calculation of the direct photon cross section with a Frixione cone radius of 0.3 provided by Werner Vogelsang based on code originally developed for [6].

# Chapter 2

# Colliding Polarized Protons and Observing the Aftermath

The critical requirement in the measurement of a double-helicity asymmetry is the ability to collide polarized proton beams. Unlike polarized DIS experiments where the hadron target is stationary and relatively straightforward to polarize, these beams must remain polarized through the acceleration to high energies. Such a facility did not exist until the inception of the Relativistic Heavy Ion Collider (RHIC) in 2000 [37] and, today, RHIC has enabled many asymmetry measurements through a diversity of channels, placing the strongest constraints on the gluon polarization of the proton.

Various experiments have utilized RHIC since its inception. Most focused on the collision of heavy ions in the study of QCD at high densities and temperatures, but only two have taken advantage of the polarized proton beams unique to RHIC: STAR and PHENIX [38]. This analysis provides the first measurement of a direct photon asymmetry at STAR.

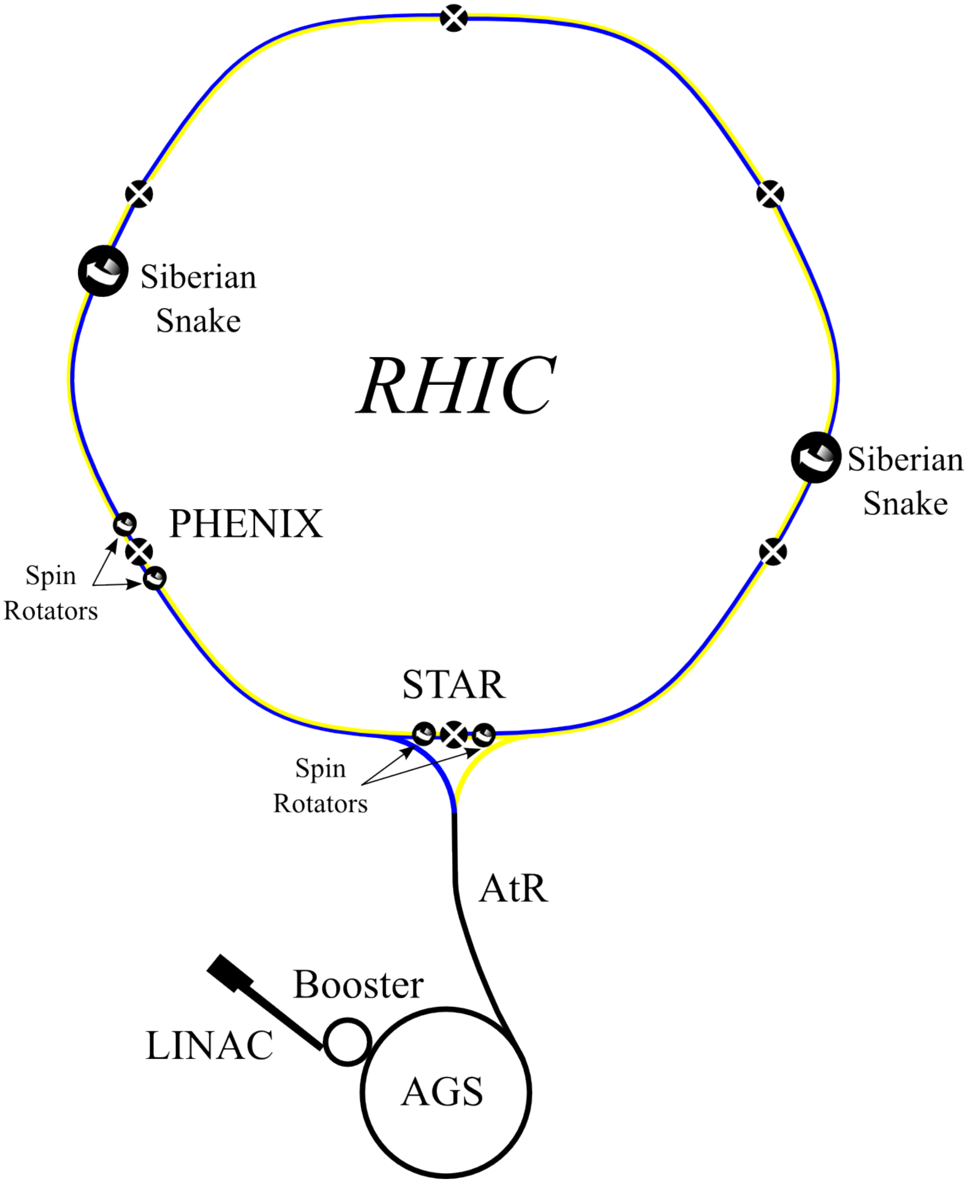

Siberian Snake
RHIC
Siberian Snake
PHENIX
Spin Rotators
STAR
Spin Rotators
AtR
Booster
LINAC
AGS

Figure 2-1: A schematic of the Relativistic Heavy Ion Collider and associated facilities. The ring itself consists of six sextants, each containing an interaction point here labeled with an X (note that, as they pass through each interaction point, the blue and yellow beams weave through each other). In addition to the Siberian Snakes that maintain the transverse polarization of the beam, the interaction points servicing the STAR and PHENIX experiments are equipped with spin rotators to enable the collision of longitudinal beams and hence the study of double-helicity asymmetries.

## 2.1 The Relativistic Heavy Ion Collider

Over 3.8 km in circumference, RHIC (Figure 2-1) is at once a high-energy synchrotron, storage ring, and collider facility [39]. The ring consists of two counterrotating beams, Blue traveling clockwise and Yellow counter-clockwise, each capable of being filled with ion species ranging from protons to gold nuclei. With the addition of spin manipulation technology, RHIC is also able to accelerate and store beams of polarized protons, admitting the polarized proton collisions necessary for the asymmetry measurement [7].

### 2.1.1 Beam Acceleration

Both proton beams begin with a 200 MeV linear accelerator, or LINAC, supplied by a polarized $H^-$ source. The LINAC beam feeds the Booster synchrotron where the electrons are stripped upon injection and the remaining protons accelerated to 1.5 GeV before entering the Alternating Gradient Synchrotron (AGS). Here the beam, consisting of approximately $10^{11}$ protons, is bunched and then accelerated to 24.3 GeV. The bunched protons are injected into RHIC through the AGS-to-RHIC transfer line, filling the ring one bunch at a time.

Once both beams have been filled they are accelerated to their final energies, typically 100 or 250 GeV in each beam to give 200 and 500 GeV collision energies, respectively. As the beams collide, the protons are depleted and the luminosity falls; once the luminosity drops to an undesired level the beams are dumped and the ring refilled.

### 2.1.2 Beam Steering

Through acceleration and storage in the main RHIC ring, the bunched beams are steered by 66 FODO[1] cells, 11 in each of the RHIC sextant arcs. Each FODO cell contains a focusing and defocusing quadrapole magnet, each surrounding by a fine tuning sextupole and corrector magnets, and two superconducting dipole magnets that maintain the arc of the beam.

Until collisions are desired, the two beams are kept isolated in separate beam pipes. Interaction points are created with a series of dipole magnets that steer the beams into each other: additional quadrapole magnets first focus each beam to increase the collision luminosity and then a series of dipole magnets chicane the beams into a direct collision path. After the collision the beams experience the reverse steering, first returned to their original orbits by the dipoles and then defocused by the quadrapoles.

### 2.1.3 Beam Polarization

As noted above, the proton acceleration sequence begins with a polarized source. The Optically Pumped Polarized Ion Source (OPPIS) transfers polarization from an incident laser onto the nuclei of a neutral hydrogen beam, producing a jet of hydrogen with over 80% polarization [40]. Polarization from electrons in an optically pumped rubidium vapor is first transferred to the hydrogen via collisional processes and then to the nuclear protons via a Sona transition [41] before the incorporation of additional electrons in preparation for the LINAC. The difficulty for RHIC is maintaining this polarization through each stage of the acceleration, especially through spin resonances that can rapidly depolarize the beam.

---

[1] FODO is accelerator physics short hand for "focusing, zero, defocusing, zero", referring to the quadrapole magnets that focus and defocus the beam and the two dipoles that turn the beam but otherwise provide no focusing.

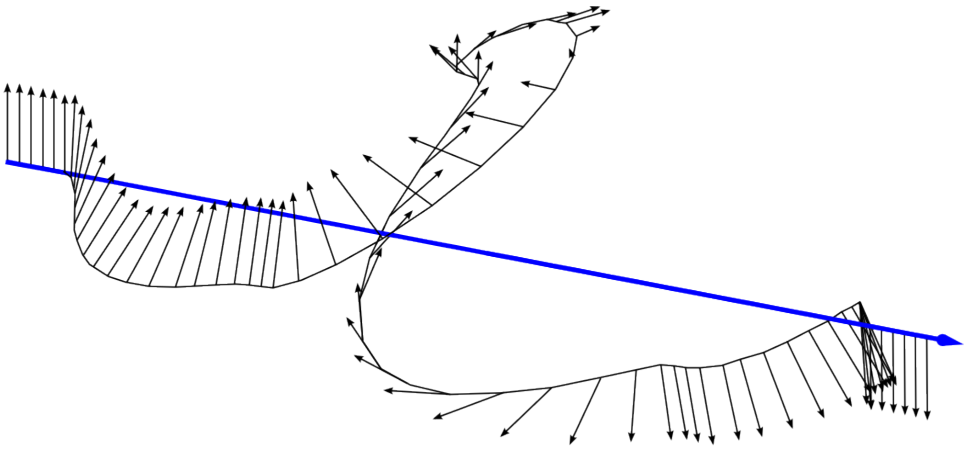


Figure 2-2: The four superconducting dipoles in the full Siberian snakes run the beam through a trajectory that rotates the polarization (black) before return to the original beam axis (blue). Adapted from [7].

The preservation of the beam polarization is feasible only with the introduction of the Siberian snake [42], an assembly of four dipole magnets that rotates the transverse component of the polarization around the beam axis by 180 degrees (Figure 2-2). By placing snakes on opposite sides of the RHIC ring, the beams spend half of each rotation in one polarization and the half in the other; small perturbations to the spin cancel through a full rotation and, provided stronger depolarizing resonances are avoided, the transverse polarization is maintained through acceleration and then storage. In addition to the two full Siberian snakes in RHIC, the AGS also contains two partial snakes to help conserve the polarization through the preliminary acceleration.

Once in storage mode, the beams are fully energized and held with stable transverse polarization. The double-helicity asymmetry, however, requires that the beams be longitudinally polarized during each collision. This is accomplished with the use of spin rotators, a configuration of two dipole magnets that rotate the beam polarizations from transverse to longitudinal (and then back) on each side of the interaction point, in a manner similar to the Siberian snakes.

Because of inevitable polarization losses, continuous monitoring of the beam polarization is critical to accurate asymmetry measurements. Inserting a hydrogen target into the beam and measuring elastic scattering with a Breit-Rabi polarimeter provides a precise determination of polarization, but the measurement is time consuming and starves the experiments of valuable luminosity. Elastic scattering from a carbon target provides a faster measurement, but ultimately only a relative one due to unknown nuclear physics. RHIC takes advantage of both strategies, frequently inserting a thin carbon wire into the beam to track the relative polarization with a hydrogen-jet target occasionally providing absolute measurements [43].

### 2.1.4 Beam Structure

The beams in RHIC maintain the bunched structure in which they are filled. Each consists of 120 bunches, all but 8 of which are filled from the AGS: those 8 empty bunches form an abort gap in each beam, providing adequate time to steer the protons into a beam dump in the case of an emergency abort. Once the beams are full and accelerated to the desired energy, they are "cogged" by adjusting their relative phase to ensure that the bunches overlap at six desired interaction points. Once cogged there are 120 bunch crossings between the beams at each interaction point, 16 of which contain at least one empty bunch from the abort gaps[2], each crossing separated by 109 ns.

In order to avoid any time-varying systematic in the beam polarizations the bunch crossings must rapidly cycle through the possible initial state helicity configurations, $h_B h_Y = \{++, +-, -+, --\}$. This is done by rapidly reversing the polarization of each bunch; typically one beam is filled with the polarization switched after every bunch crossing and the other after every pair of bunches, with a possible phase between the patterns. Although this produces all four possible initial state configurations uniformly, there is no guarantee that the collisions will be uniformly distributed amongst the configurations: the abort gaps and other bunch crossing inefficiencies bias the distribution of initial states[3], and only by recording the relative frequency (or luminosity) of each initial state will a proper asymmetry be viable.

---

[2] At one interaction point the two abort gaps overlap and only 8 bunch crossings are empty. This interaction point is reserved for the PHENIX experiment which can handle the additional luminosity.

## 2.2 STAR

Located at the 6 o'clock interaction point, The Solenoid Tracker at RHIC (STAR) is a suite of large acceptance detectors suited for the high multiplicity environment of heavy ion collisions [8]. With precision tracking and expansive electromagnetic calorimetry, however, the experiment is also amenable to recording the aftermath of high energy proton collisions (Figure 2-3).

STAR defines a local coordinate system from the perspective of an observer at the center of RHIC facing South (Figure 2-4). The beam axis defines $z$, the Blue beam entering from negative $z$ and Yellow from positive, and $y$ is given by the vertical; $x$ lies along the line of sight, constrained by the right-handed coordinate system. More appropriate to high energy collisions is the $(r, \phi, \eta)$ coordinate system[4],

$$\begin{aligned} r &= \sqrt{x^2 + y^2} \\ \phi &= \arctan\left(y/x\right), \\ \eta &= -\log\left[\tan\left(\theta/2\right)\right] \end{aligned}$$

where $\theta$ is the usual polar angle,

$$\theta = \arctan\left(r/z\right).$$

---

[3] Incidentally, this means that summing over the polarization does not produce a perfectly unpolarized beam. Unpolarized measurements technically require the same relative corrections as any asymmetry, although any such correction is negligible given the small asymmetries measured so far at RHIC.

[4] Because the position of a collision is distribution around the nominal interaction point, there are technically two coordinates systems: one defined with respect to the detector and one defined event by event with respect to the interaction vertex. Measurements are defined in terms of the latter.

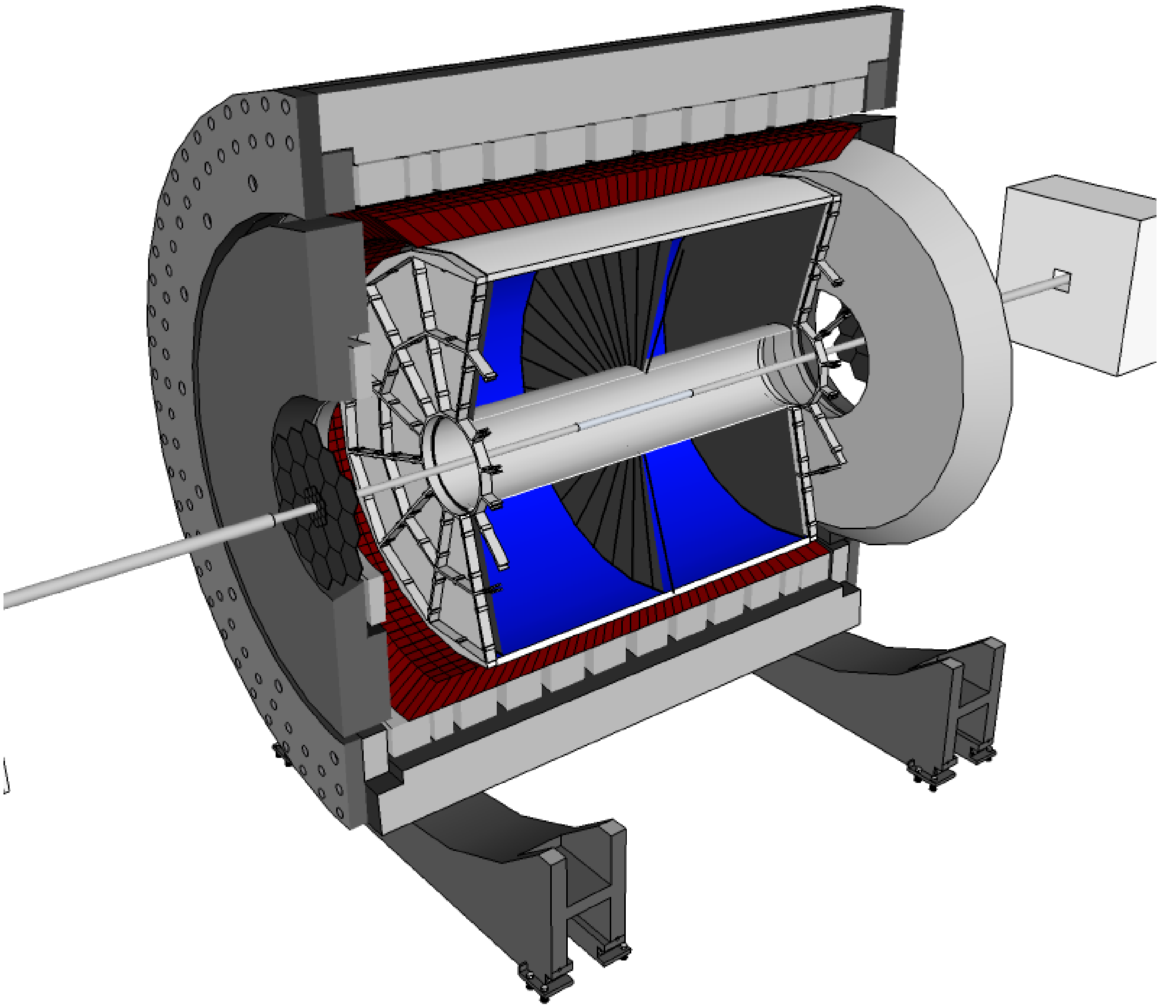


Figure 2-3: A schematic of the STAR detector with the Time Projection Chamber (blue) and Barrel Electromagnetic Calorimeter (red) highlighted.

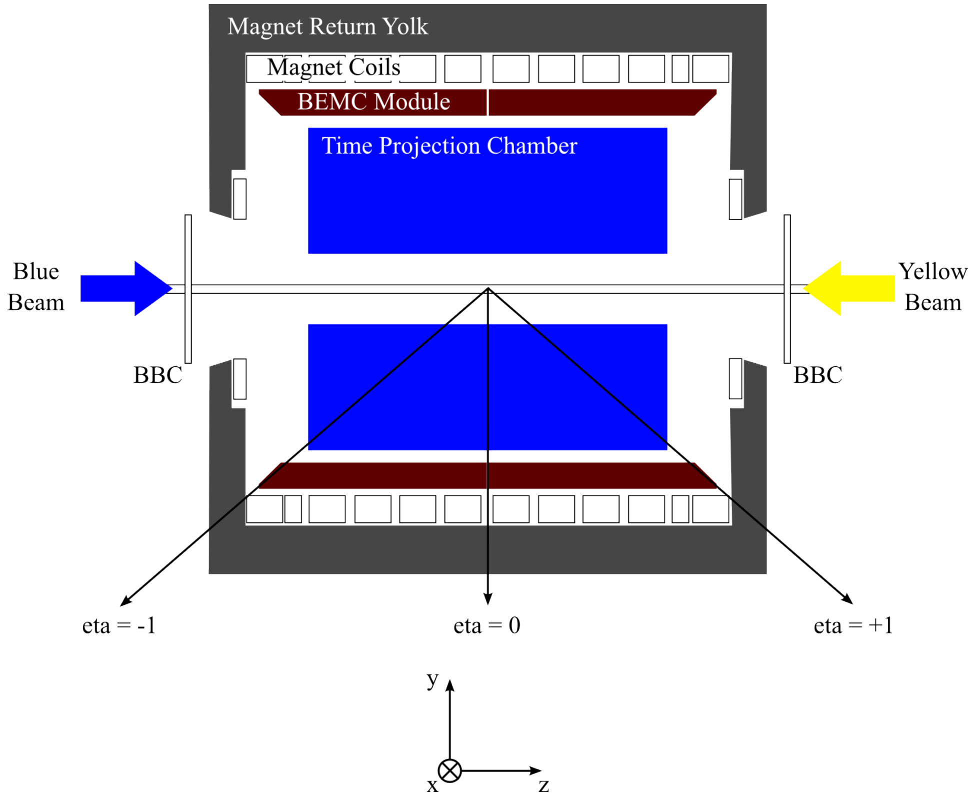

Magnet Return Yolk
Magnet Coils
BEMC Module
Time Projection Chamber
Blue Beam
Yellow Beam
BBC
BBC
eta = -1
eta = 0
eta = +1
y
x
z

Figure 2-4: A cartoon of the STAR detector and the STAR coordinate system as viewed from center of RHIC. Adapted from [8].

### 2.2.1 The “Solenoid”

Critical to many of the detectors at STAR, the magnet system provides a maximum 0.5 T solenoidal field (with reversible orientation) uniform over much of the experiment [44]. The cylindrical, water-cooled coils and steel return yolk surround the experiment, leaving 5 m within the inner diameter for the detector subsystems.

### 2.2.2 The “Tracker” and Other Detector Subsystems

The detector subsystems are dominated by STAR’s namesake, the Time Projection Chamber, which provides tracking and particle identification at central rapidity. The massive tracker volume is surrounded by calorimetry, forming a barrel. Placed at forward and backward rapidity, the Beam-Beam Counters monitor the collisions at the interaction point, providing corrections and systematic checks to the measurements.

Additional detectors provide calorimetric coverage at forward rapidity and enhance tracking and particle identification at central rapidity.

#### The Time Projection Chamber

Over 4 m in length and 4 m in diameter, the Time Projection Chamber (TPC) is a cylindrical gas volume surrounding the interaction point at STAR [45]. Charged particles emanating from the beam collisions pass through the volume and ionize the gas, a mixture of 90% argon and 10% methane (for quenching fluorescence) known as P10. A highly uniform electric field drifts the ionization products to anode wires along the volume’s endcaps that form a multiwire proportion chamber: the avalanche formed as the electrons are drawn onto the anode wires induces an image current on nearby cathode pads which is integrated and digitized to provide a two dimensional projection of the ionization. The uniformity of the electric field, and hence the drift velocity, admits a reconstruction of $z$ from the total drift time, providing full three dimensional reconstruction of the initial ionization.

Interpolating between the measured ionization reproduces the paths of the original charged particles and, due to the bending in the magnetic field, their momentum and charge. Moreover, the energy loss of the particle, and consequently the magnitude of the ionization along the reconstructed track, depends on the particle velocity and provides discrimination between different species of charged particles. The TPC admits precision reconstruction for incident particles spanning $|\eta| \lesssim 1.2$ and full azimuth, with momenta between 100 MeV/c and 30 GeV/c. Considering all such tracks in a single event also constrains the position of the initial collision, otherwise known as the vertex.

With a 135 V/m electric field, the ionized electrons travel at $\sim 4.5\,\mathrm{cm}/\mu\mathrm{s}$, requiring over $400\,\mu\mathrm{s}$ for a single event to clear the gas volume. Over 400 bunch crossings pass in this time and the readout of a single event is pervaded by tracks belonging to collisions both before the event and after. These pileup tracks are a relative nuisance at small luminosities, but a serious issue as the collision rate increases. The ionized nuclei are even more problematic - because of their significantly larger mass, the nuclei take longer to drift and form a cloud of positive charge near the central cathode. This space charge perturbs the drift of the ionized electrons, biasing the reconstruction of the underlying track unless carefully accounted for in the TPC calibration.

### Barrel Electromagnetic Calorimeter

Immediately surrounding the TPC, the 120 modules of the Barrel Electromagnetic Calorimeter (BEMC) provide calorimetry between $-1 < \eta < 1$ with full azimuthal coverage [9].

Spanning $\Delta\phi = 1$ and $0 < |\eta| < 1$, each module consists of alternating layers of lead and plastic scintillator megatiles held together with steel compression straps.

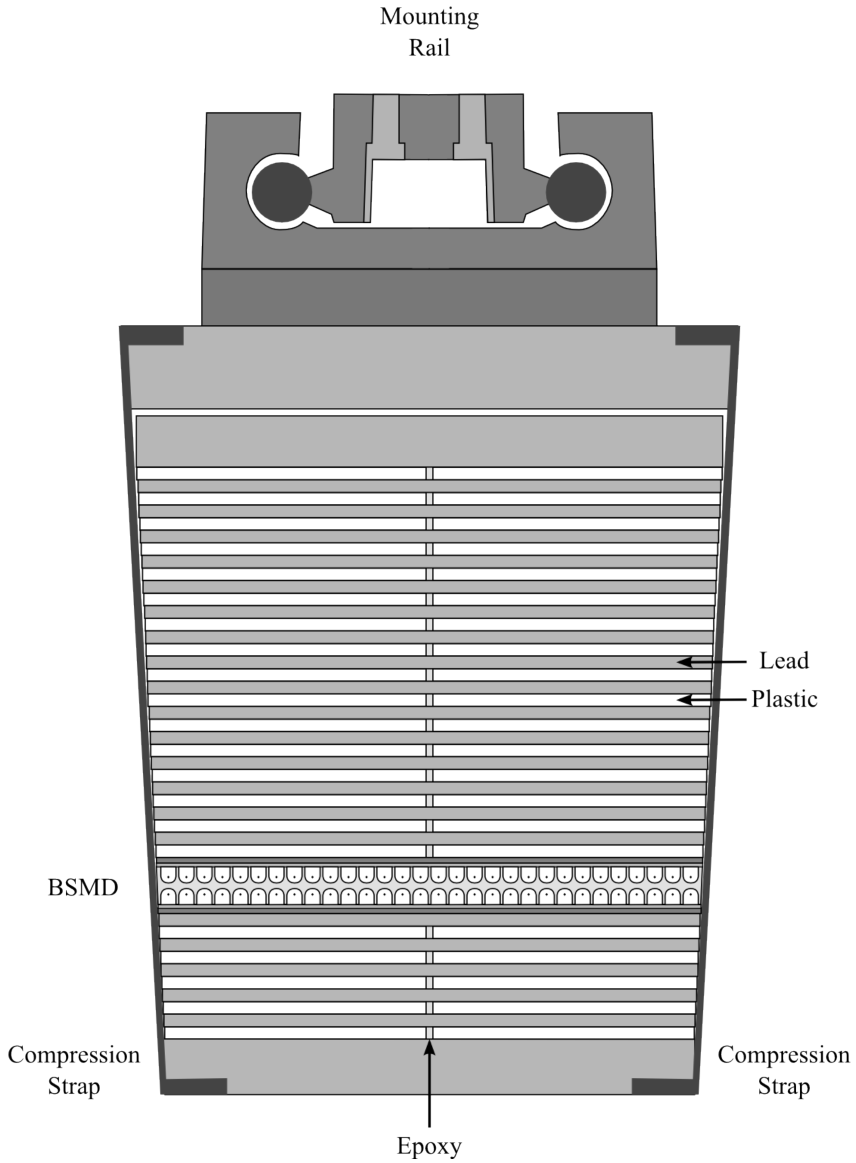

Mounting Rail
Lead
Plastic
BSMD
Compression Strap
Compression Strap
Epoxy

Figure 2-5: A view of a Barrel Electromagnetic Calorimeter (BEMC) module along the $z$ axis. Layers of lead and plastic scintillator held together with steel compression straps form the calorimeter, with epoxy filled channels separating the individual towers. The module itself mounts into the experiment with a rail system. Adapted from [9].

The 20 lead and 21 plastic layers are all uniformly 5 mm thick, except for the first two layers of plastic which are slightly thicker at 6 mm (Figure 2-5).

Thin channels in the plastic form individual towers, with optically opaque epoxy isolating the scintillation light. Wavelength shifting fibers running just inside of these channels gather the light in each division, routing fibers from the same tower to a single phototube outside of the magnet before being integrated and digitized[5]. Each module contains 40 such towers, each spanning $\Delta\phi = \Delta\eta = 0.05$ with an approximate depth of 20 radiation lengths; the summed modules combining to give 4800 towers in a projective geometry.

Set almost entirely by the structure of the sampling calorimeter, the energy resolution of the BEMC is given by

$$\sigma_E = 14\%/\sqrt{E} \oplus 1.5\%.$$

The absolute scale of the calorimeter gains is determined with a rigorous calibration procedure that equalizes the response to incident electrons and the momentum as measured by the TPC. With sufficient statistics this allows for each tower to be calibrated independently, although in practice the calibration has been limited to small patches of towers.

### Barrel Shower Maximum Detector

Because the Molière radius of Pb is only 1.6 cm, the 10 cm × 10 cm towers can fully contain the showers of multiple incident particles. Without finer spatial resolution the BEMC is unable to discriminate between the single photon and the two neighboring photons from a neutral pion decay, let alone determine any details of the single shower.

---

[5] The light from the first two plastic layers is partially diverted to a separate readout, forming a preshower detector not used in this analysis. The addition thickness of these layers ensures that the partition does not spoil the uniformity of the integrated layers.

In order to provide the necessary segmentation, the Barrel Shower Maximum Detector (BSMD) was introduced into the BEMC. The BSMD consists of 20 mm multi-wire proportion counters placed into each module after the first 10 megatiles, about five radiation lengths into the calorimeter where electromagnetic showers are at their widest. Incident electromagnetic showers ionize a mix of argon and CO2 gas, inducing a signal on orthogonal planes of cathode strip readouts as the ionization avalanches towards anode wires (Figure 2-6).

The thin readout strips provide spatial resolution on the order of the Molière radius and dramatically improved shower reconstruction. Segmentation in $\eta$ is provided with 300 strips running along the length of every module. Each strip is 1.46 cm wide (increasing to 1.88 cm for $|\eta| > 0.5$) and spans the full width of the module. The 300 strips in the orthogonal $\phi$ plane are each 1.33 cm in width, arranged into 20 rows of 15 strips, each $\eta = 0.1$ long. Note that in this configuration each row of $\phi$ strips is covered with 15 $\eta$ strips, the overlap aligning with four towers in the module.

With 64000 total cathode strips, the readout of the BSMD is limited by practical constraints. In particular, the digitization of each signal is limited to a 10-bit ADC and the signal amplification has to be tuned between higher resolution (larger gains amplifying smaller energy depositions and providing greater detail into the shape of the incident shower) or larger dynamic range (higher gains also push the peak of the shower past the range of the ADC, saturating the response and truncating shower). Note that the behavior of the strips at saturation provides a convenient, though crude, means for calibrating the absolute scale of the BSMD.

### Beam-Beam Counters

Flanking the barrel at $3.4 < |\eta| < 5.0$, the two Beam-Beam Counters (BBCs) are each comprised of 18 hexagonal tiles of plastic scintillator, 6 smaller tiles nested within 12 larger tiles [46]. The 6 smaller tiles provide a convenient minimum bias condition for

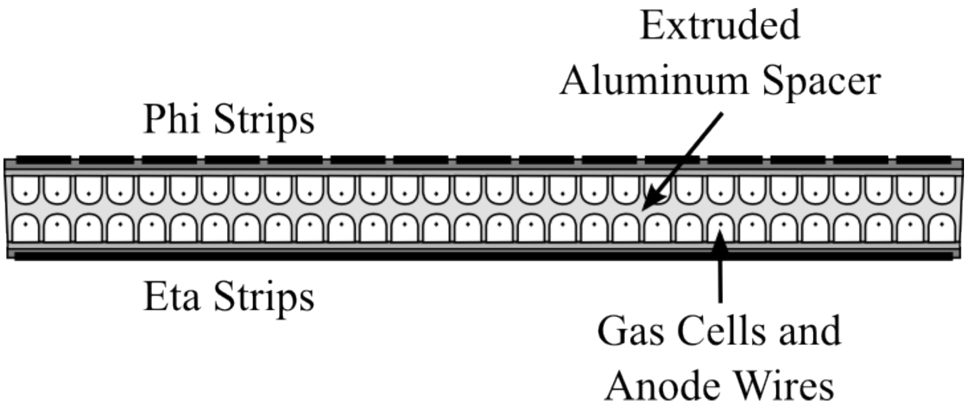


Figure 2-6: Expanded view of the Barrel Shower Maximum Detector. The sense gas and anode wires are sheathed in long cells formed in the extruded aluminum separating the two planes of cathode strips, the $\eta$ strips perpendicular cells and the $\phi$ strips parallel. Adapted from [9].

collisions: the coincident signal in the BBCs signifying a spray of charged particles from the remnants of a proton-proton collision, with no other restriction placed on the details at mid-rapidity. The BBCs also provide a convenient means of measuring the relative luminosity of the beam spin states recorded by the detector, allowing any measured asymmetry to be corrected for biases in the spin states over the course of a fill.

### 2.2.3 Trigger System and Data Acquisition

At every bunch crossing a global timing signal prompts detectors with fast readouts (such as the BEMC) to begin recording data. The response of these fast detectors provides crucial information for the rest of the experiment, particularly whether or not to trigger the readout of slow detectors (namely the TPC) and save the event information to disk.

#### The Trigger Hierarchy

The STAR trigger system [47] consists of four hierarchical layers, L0, L1, L2, and L3, with events written to disk only after satisfying all four layers.

L0 runs off of local data storage and manipulation boards (DSMs) which accumulate the raw calorimeter readout and forward summary statistics (in particular maxima and summations of tower ADCs in given regions) to the trigger control unit (TCU). This enables the TCU to make a trigger decision based on fast decision tree algorithms, largely limited to single towers, small groups, or large patches of towers passing set ADC thresholds. Any of the multiple trigger criteria will satisfy the L0 requirement, signaling the readout of the slow detectors and advance the system to the next layer.

While the slow detectors are read out, more time-consuming algorithms can run in the L1 layer. Rejection here aborts the readout and resets the system, reducing the deadtime of the experiment. L1 triggers have not been utilized for many years, and the L1 layer passed all events to L2 in Run 9.

The L2 layer runs off a suite of commercial PCs with fast access to the raw calorimeter data. With more processing power and global calorimeter information, algorithms in the L2 layer are significantly more sophisticated than those in the previous layers and enable a suite of efficient triggers targeted at specific analyses.

Only at the final layer is information from the slow detectors made available. The L3 layer incorporates approximate tracking information into the trigger decisions, admitting algorithms dedicated to unique charged particle signatures such as those characteristic of heavy flavor decays. Although not used for proton collisions, these triggers have become powerful in heavy ion analyses with recent upgrades to the fast tracking software.

### Data Acquisition

Once an event has been accepted by the trigger system, the STAR Data Acquisition System [48] formats the information from each detector subsystem before writing the data to tape at the RHIC Computing Facility [49], where it is later made available for analysis by individual collaborators. Data from a given fill is separated into individual *runs* of a few hundred thousand events, not to be confused with the labeling of the yearly running periods.

# Chapter 3

# Data Analysis

For all of the theoretical advantages, direct photons offer little until they can be isolated from the overwhelming QCD background. In addition to the data itself, a successful measurement requires an accurate simulation to elucidate the detector signatures of signal and background and what information may be able to discriminate between them. Reducing the event-wide detector responses into smaller clusters, being careful to not discard any of the potential discrimination, yields candidates photons from which the true photon spectrum can be inferred with a suite of powerful analysis techniques.

## 3.1 Data

The data for this measurement were collected in the ninth running of the RHIC collider, beginning in early 2009 and continuing through the summer until early July. For almost the entire running period the facility collided polarized protons at a center of mass energy $\sqrt{s} = 200\,\mathrm{GeV}$, $100\,\mathrm{GeV}$ in each beam.

| Trigger | Level | Offline ID | Thresholds |
|---|---|---|---|
| BHT2 | L0 | 240540 | 4.2 GeV in at least one BEMC tower |
| JP1 | L0 | 240411 | 5.5 GeV in at least one of 12<br>predetermined partitions of 400 BEMC towers |
| L2BGamma | L2 | 240620 | 4.5 GeV BEMC tower seed,<br>6 GeV in four tower cluster<br>5% Random accept |

Table 3.1: Details of the L2BGamma trigger and its dependencies at L0. Note that, because the online triggers did not have the final BEMC calibration, all thresholds are approximate.

While a variety of L0 and L2 triggers were operation in Run 9, only the L2BGamma trigger (Table 3.1) will be considered here. All events passing a simple high-tower trigger or jet-patch trigger at L0 were passed to the L2BGamma trigger, which executed a crude seed-based cluster finder in the BEMC.

All events passing the L2BGamma trigger were collected into an exclusive set of DAQ files known as the *st_gamma* stream[1]. The raw data from all files in the *st_gamma* stream were reconstructed into the detector responses pertinent to user analysis within the P10ic production series [50]. Of all the events in the *st_gamma* stream, only those events from runs configured for physics, in particular *production2009_200Gev_Single*, were included.

As with previous running periods, the gross quality of the polarized data in Run 9 was assessed through a brute force examination procedure [10]. This quality control determined master priority lists used by the majority of analyses [51], but the particular criteria used for these lists were more strict than necessary for a photon measurement at central rapidity. Consequently, custom criteria were developed for the photon measurement (Table 3.2), supplemented with specifications from a more detailed assessment of detector subsystem performance [52] and beam tagging [53].

---

[1] Technically there were two streams, *st_gamma* and *st_gamma_adc*. The latter contained events where the BSMD response was not zero-suppressed in hardware, while the former contained only zero-suppressed events.

| Code | Description |
|---|---|
| Cx | TPC Failure |
| Bx | BEMC Failure |
| Sx | BSMD Failure |
| Dt | Run stopped prematurely due to TPC problems |
| Db | Run stopped prematurely due to BEMC problems |
| Dm | Run stopped prematurely due to magnet problems |
| Dq | Run stopped prematurely due to DAQ problems |
| XX | Run failed |
| ?? | Unknown status |

Table 3.2: A photon-specific quality assessment began with the removal of all data failing these criteria, as defined in [10].

## 3.2 Simulation

In order to fully understand the measurement process – in particular the response of the STAR detector to incident collisions – the data must be accompanied by a full simulation. Simulation of proton-proton collisions at STAR generally proceeds in three stages: the collision itself is modeled with the Pythia Monte Carlo generator, the interaction of the collision with the STAR detector is modeled with Geant, and finally the output of the simulated detector responses is reconstructed into the same format as the data. In this latter stage detector imperfections, readouts, and calibrations can be modeled.

### 3.2.1 PYTHIA

Pythia [54] is a Monte Carlo generator capable of simulation a large class of particle interactions. Formally, the algorithm attempts to draw samples from the space of possible final states using leading-order matrix elements of the hard interaction with a phenomenological procedure for incorporating higher order effects of initial-state

| **Subprocess ID** | **Interaction** | **Description** |
|---|---|---|
| 14 | $f_i\bar{f}_i \to g\gamma$ | Photon |
| 18 | $f_i\bar{f}_i \to \gamma\gamma$ | Photon |
| 29 | $f_i g \to f_i\gamma$ | Photon |
| 114 | $gg \to \gamma\gamma$ | Photon |
| 115 | $gg \to g\gamma$ | Photon |
| 11 | $f_i f_j \to f_i f_j$ | QCD |
| 12 | $f_i\bar{f}_i \to f_k\bar{f}_k$ | QCD |
| 13 | $f_i\bar{f}_i \to gg$ | QCD |
| 28 | $f_i g \to f_i g$ | QCD |
| 53 | $gg \to f_k\bar{f}_k$ | QCD |
| 68 | $gg \to gg$ | QCD |

Table 3.3: The photon simulation included both direct photon subprocesses and QCD subprocesses. Note that quark-gluon Compton scattering (29) is by far the most dominant of the signal processes.

radiation, final-state radiation, and hadronization. The simulation for this measurement utilized version `6.4.23` of the code, in particular the `Pro-pT0` tune of the phenomenological parameters from the Professor collaboration [55].

In order to model the data, the simulation included both direct photon subprocesses and QCD $2 \to 2$ subprocesses (Table 3.3). Note that these direct photon processes are essentially leading-order, but the higher-order effects added through the phenomenological modeling may contradict the formal definition of a direct photon from Frixione. Consequently, an additional Frixione condition with $R = 0.3$ is required before a direct photon event is classified as signal to ensure consistency with explicit theoretical definition.

The total simulation was partitioned into subsets by the PYTHIA variables `CKIN(3)` and `CKIN(4)`, which place lower and upper thresholds, respectively, on the final-state transverse momentum. This subdivision allowed a more even sampling of the final-state phase space, but not without consequence. The `CKIN` variables are defined in the partonic center of mass frame and not in the detector frame: the hard constraints

| **Component** | **Mean** | **Standard Deviation** |
|---|---|---|
| $v_x$ | 0.3 | 0.01 |
| $v_y$ | 0.0 | 0.01 |
| $v_z$ | -5.96 | 60 |

Table 3.4: Each component of the Pythia vertex within the STAR detector geometry is sampled independently from a Gaussian distribution. All values are in centimeters.

in the partonic frame become soft constraints in the detector frame. In order to avoid any bias by excluding events leaking across the constraint, the final simulation had to be sampled from a partonic phase space much larger than that defined in the detector frame.

Once the evolution of the final state has concluded, a list of the final-state particles and their kinematics is introduced into the next phase of the simulation where the response of the STAR detector is modeled. These particles are all assumed to originate from a single vertex, but the location of that vertex relative to the detector is chosen to conform with the spatial interaction distribution seen in data (Table 3.4). Note that various hadronic decays were suppressed within Pythia so that the hadrons would decay only in the detector reconstruction where their decay vertices can be displaced from the Pythia vertex (Table 3.5).

Because the simulated vertex distribution was chosen when only preliminary vertex data were available, the disagreement between the reconstructed vertex distributions in the data and simulation was appreciable. The discrepancy of the $z$ component of the vertex is particularly important, as the location along the beam axis defines the incidence of the final state particles and hence the detector response. A simple reweighting of the simulated events as a function of the reconstructed $v_z$, however, successfully brought the two distributions into agreement and removed the possible bias in the simulation.

| **Hadron** | **GEANT ID** | **PYTHIA Switch** |
|---|---|---|
| $\pi^0$ | 111 | `MDCY(102, 1) = 0` |
| $\pi^+$ | 211 | `MDCY(106, 1) = 0` |
| $\eta$ | 221 | `MDCY(109, 1) = 0` |
| $K^+$ | 321 | `MDCY(116, 1) = 0` |
| $K_\mathrm{S}$ | 310 | `MDCY(112, 1) = 0` |
| $K_\mathrm{L}$ | 130 | `MDCY(105, 1) = 0` |
| $\Lambda^0$ | 3122 | `MDCY(164, 1) = 0` |
| $\Sigma^0$ | 3212 | `MDCY(167, 1) = 0` |
| $\Sigma^+$ | 3112 | `MDCY(162, 1) = 0` |
| $\Sigma^-$ | 3222 | `MDCY(169, 1) = 0` |
| $\Xi^-$ | 3312 | `MDCY(172, 1) = 0` |
| $\Xi^0$ | 3322 | `MDCY(174, 1) = 0` |
| $\Omega^-$ | 3334 | `MDCY(176, 1) = 0` |

Table 3.5: The decays of various hadrons are suppressed within Pythia so that any decay vertex displaced from the hard interaction can be correctly reconstructed in the second phase of the simulation.

### 3.2.2 GEANT

Once the Pythia final state has been placed into the STAR geometry, the particles are allowed to propagate and their interaction with the detector simulated with Geant3 [56]. The STAR geometry, comprised of each detector subsystem and much of the support structure, is modeled with the GSTAR framework [57].

Significant improvements were made in the development of the Run 9 STAR geometry. In addition to the the introduction of various missing details, notable errors in the placement of the BEMC were removed [58]. The importance of the BEMC to the photon measurement warranted an elaborate study to validate the corrections and ensure that no large problems remained [59]. Those same studies also warranted the reduction of thresholds for the propagation of electromagnetic particles within Geant, which especially improved the simulation of the BSMD.

### 3.2.3 Reconstruction

Raw Geant output is processed much the same way as data: any higher-level reconstruction such as TPC tracking is performed and all of the remaining detector information reduced into a user-amenable format. Reconstructing simulation, however, also requires the use of slow simulations. These afterburners model the detector readouts, simulating the digitization of the readout with the introduction of electronic noise and other effects. It is in this stage where final detector calibrations and status tables are set.

Reconstruction in both data and simulation is performed with the Big Full Chain (BFC), a massive ROOT macro that admits a wide range of arguments for customizing the process [60]. The photon simulation was reconstructed with the following arguments:

```
DBV20100601,tpcrs,y2009a,MakeEvent,ITTF,FiltGamma,NoSsdIt,
NoSvtIt,Idst,BAna,l0,Tree,logger,Sti,VFPPVnoCTB,eemcDb,beamLine,
tpcDB,TpcHitMover,TpxClu,bbcSim,tofsim,tags, emcY2,EEfs,evout,
-dstout,IdTruth,geantout,big,fzin,MiniMcMk,clearmem,Corr4,
OSpaceZ2,OGridLeak3D.
```

One final argument varied over the course of simulation; `stdYYYYMMDD.HHMMSS` determines the timestamp for the reconstructed simulation, and consequently which calibrations and status tables are applied. The argument was randomly sampled for each simulated event in order to match the gross time variations seen in the data (Table 3.6).

### 3.2.4 Filtering

Unfortunately, taking a sampled event through the full simulation chain can be quite costly and the statistical power of any naive simulation sampling from the whole of

| **BFC Option** | **Probability** |
|---|---|
| `std20090518.114745` | 0.415 |
| `std20090607.214745` | 0.120 |
| `std20090621.124745` | 0.465 |

Table 3.6: Simulation timestamps were randomly assigned to events in order to mimic the time variations seen in the data, in particular the status tables of the BEMC.

phase space will be negligible compared to the collected data. Not all of the direct photon and QCD phase space is relevant to the photon measurement, however. The detector response to most QCD events looks nothing like a direct photon, and much of the expanded partonic phase space still lies outside of the analysis thresholds. Filtering events irrelevant to the photon measurement before they consume computational resources affords the generation of many more simulated events without the introduction of biases or other artifacts [61].

The photon simulation used two levels of filtering. Immediately after the PYTHIA event had been placed into the STAR geometry, the `StMCFilter::StBemcGammaFilter` rejected the event if none of the potential detector responses resembled that of a photon. Once GEANT had processed an event but before the computationally demanding elements of the reconstruction, the `StGammaFilterMaker` made a final acceptance decision. Rigorous testing ensured that the power of the filters did not come at the expense of possible biases [62].

Together, the two filters allowed for a simulation with almost $2\,\mathrm{pb}^{-1}$ of integrated luminosity (Table 3.7), the largest simulation ever produced for the STAR experiment. The actual computation was carried out at the MIT Tier2 cluster, part of the Bates High Performance Computing Center, via the Open Science Grid [63].

## 3.3 Preprocessing

Attempting to analyze each event in its entirety is impractical; each event must instead be reduced to only the most relevant information. This process began by

| **Partition Name** | **Subprocesses** | **CKIN(3) (GeV)** | **CKIN(4) (GeV)** | **Total Filter Acceptance** | **Integrated Luminosity ($\mathrm{pb}^{-1}$)** |
|---|---|---|---|---|---|
| rcf10270 | Photon | 2 | 3.5 | 0.033% | 25.0 |
| rcf10271 | Photon | 3.5 | 4.5 | 0.531% | 25.4 |
| rcf10272 | Photon | 4.5 | 6.5 | 4.61% | 25.0 |
| rcf10273 | Photon | 6.5 | 8.5 | 23.2% | 25.1 |
| rcf10274 | Photon | 8.5 | 10 | 39.1% | 25.2 |
| rcf10275 | Photon | 10 | 15 | 47.0% | 24.9 |
| rcf10276 | Photon | 15 | 25 | 60.1% | 24.4 |
| rcf10277 | Photon | 25 | 40 | 74.4% | 30.6 |
| rcf10278 | Photon | 40 | -1 | 79.1% | 2700 |
| rcf10279 | QCD | 2 | 3.5 | 0.00012% | 2.13 |
| rcf10280 | QCD | 3.5 | 4.5 | 0.0018% | 1.87 |
| rcf10281 | QCD | 4.5 | 6.5 | 0.020 % | 1.95 |
| rcf10282 | QCD | 6.5 | 8.5 | 0.200 % | 1.98 |
| rcf10283 | QCD | 8.5 | 10 | 0.763% | 2.00 |
| rcf10284 | QCD | 10 | 15 | 2.54% | 2.03 |
| rcf10285 | QCD | 15 | 25 | 12.6% | 2.03 |
| rcf10286 | QCD | 25 | 40 | 41.2% | 2.01 |
| rcf10287 | QCD | 40 | -1 | 67.1% | 2.76 |

Table 3.7: With both filters, the full simulation reached at least 1.87 $\mathrm{pb}^{-1}$ of integrated luminosity across all the relevant phase space. Events in partitions with higher integrated luminosity are weighted to ensure that the effective luminosity is uniform. Note that `CKIN(4)`$= -1$ removes any upper bound on the partonic transverse momentum.

applying final event-level constraints before constructing photon candidates with each event. These candidates were further reduced to a list of discriminants characterizing each potential photon.

### 3.3.1 Event Selection

In order to remove any stray events present in the *st_gamma* stream, the analysis began with an absolute requirement of the L2BGamma trigger. The trigger logic was also emulated in software, allowing thresholds computed from final calibrations as well as carefully constructed status tables; any event not satisfying the software trigger was rejected.

All reconstructed vertices in each event are ranked via heuristics designed to isolate the primary vertex associated with the hard interaction. Any events without a primary vertex of positive rank are indicative of possible biases in the reconstruction, and were immediately removed. The primary vertex was also required to be central, with $|v_z| \leq 60\,\mathrm{cm}$.

A final restriction was placed on events with poor collider information, such as the exact bunch crossing of the incident collision and the local beam polarization.

Because simulated events lack hardware trigger and beam information, only the simulated trigger and vertex conditions are required for the simulation.

### 3.3.2 Photon Candidates

High-energy photons emanating from the hard interaction interact electromagnetically and, consequently, their detector signature is an isolated deposition of energy in the BEMC.

Construction of photon candidates began with a seed-based clustering algorithm in the calorimeter, requiring first a seed tower above a given threshold and then the cluster formed by the seed and its neighboring eight towers above another. Any tower with bad status amongst these neighboring towers immediately invalidated the cluster as a potential candidate[2]. For this analysis the seed energy threshold was set to 5.25 GeV, with the total cluster requiring at least 7 GeV in transverse energy.

Once a cluster was found, the candidate is formed by collecting the information from the nine BEMC towers along with the corresponding BSMD strips as well as any TPC tracks and BEMC towers within a given neighborhood. Neighborhoods were defined as cones in $\eta$-$\phi$ space,

$$\Delta R \equiv \sqrt{\left(\eta - \eta_{\text{candidate}}\right)^2 - \left(\phi - \phi_{\text{candidate}}\right)^2},$$

where the candidate coordinates are taken as the energy-weighted mean of the individual tower coordinates relative to the center of the detector. Specifically, all strips within $\Delta R = 0.06$ and all tracks and towers within $\Delta R = \pi/2$ were associated with the candidate.

Finally, a few additional requirements were imposed on the candidates. Any candidate with a failure of either BSMD plane (defined as bad status in more than half of the associated strips) or a single TPC track projecting to the the BEMC cluster was removed. Additionally, a small region of the BSMD exhibited odd behavior throughout the run, invalidating candidates within the range ($0.5 \leq \eta \leq 0.57$, $2.4 \leq \phi \leq 2.8$).

### 3.3.3 Discriminants

Although smaller than the event from which they're constructed, the photon candidates were still ungainly and the information they contained largely incoherent and insensitive to the difference between direct photons and the massive backgrounds. Carefully designed discriminants, however, isolated the potentially sensitive properties of each cluster into a much smaller set of variables. The success of such discriminants relied on minimizing the loss of information while avoiding regions of phase space where simulation proved to be inaccurate.

---

[2] A collection of mildly misbehaving towers (Software IDs 30, 95, 796, 1130, 1294, 1348, 1709, 1823, 2043, 2445, 3028, 3692, 3702, 4013, 4299) not included in the status tables was explicitly treated as bad for clustering and the subsequent candidate construction.

Distributions of each discriminant in data and simulation, and the decomposition into signal and background within the latter, are included in Appendix B.

### Kinematic Variables

The gross kinematics of the cluster are not particularly helpful in isolating the direct photon signal, but their correlations with other discriminants proved useful. The position of the cluster relative to the detector, for example, allowed other discriminants to be sensitive to the boundaries of the BEMC. To this end three kinematic variables were included amongst the discriminants.

**Et:** Transverse energy (GeV) of the candidate relative to the primary vertex
**eSeed:** Energy (GeV) of the cluster seed tower
**dEta:** Pseudorapidity of the cluster relative to the detector

### Isolation Variables

One of the benefits of a direct photon signal is that the resulting detector signature is clean; not only does this improve reconstruction precision but also helps to separate signal and background. Embedded in jets, hadrons are usually accompanied by neighboring particles leaving tracks in the TPC and depositing energy in surrounding BEMC towers. Photons, however, are accompanied only by occasional remnants from the collision known as the underlying event. Quantifying the isolation of a candidate from both neutral and charged energy provided powerful discrimination.

| | |
|---|---|
| **sumConeTowerEtdEt:** | Summed transverse energy of all towers within a cone of $\Delta R = 1.0$ around the candidate, normalized to the summed transverse energy of all towers in the candidate itself. |
| **sumConeTrackPtdEt:** | Summed transverse momentum of all tracks within a cone of $\Delta R = 1.0$ around the candidate, normalized by the summed transverse energy of all towers in the candidate. |
| **sumCandidateTrackPtdEt:** | Summed transverse momentum of all tracks projecting to the candidate, normalized by the summed transverse energy of all towers in the candidate. |

In order to avoid discrepancies between the data and simulation, a 1 GeV threshold was imposed on all tracks and towers entering into the cone sums of the first two discriminants..

### Shower Shape Variables

Direct photons are so rare compared to the background that even unusual QCD processes compete. Consider, for example, a final-state parton that hadronizes with almost all of its energy fragmenting into a single neutral pion. The pion subsequently decays into two photons with little separation, leaving a lone energy deposition in the BEMC. Without an accompanying shower of hadrons, however, the neutral pion is impossible to distinguish from a true direct photon unless the two photons can be resolved. Using both BEMC and BSMD information, shower shape variables discriminated between electromagnetic showers with one or two components.

| | |
|---|---|
| **f:** | Seed energy normalized to the summed transverse energy of all towers above 250 MeV in the candidate. |
| **gEta:** | Maximum strip energy normalized to the summed energy of all candidate strips in the $\eta$ plane of the BSMD. |
| **gPhi:** | Maximum strip energy normalized to the summed energy of all candidate strips in the $\phi$ plane of the BSMD. |

Although these BSMD ratios provided some discrimination, they considered only the aggregate response of the BSMD. Taking full advantage of the BSMD required modeling the shower shapes in order to admit Bayesian model selection between the single-shower and double-shower hypotheses[3].

The BSMD measures the transverse profile of an electromagnetic shower, which is qualitatively described by a central core with diffuse tails. The quantitative properties of this shape, however, are more subtle; Gaussian profiles have proven to be too narrow in practice, motivating the use of mixture models and other complicated schemes. A robust alternative is the scaled Cauchy distribution,

$$E(x) = \frac{\epsilon}{\pi\gamma} \frac{1}{1 + \left( \frac{x-\chi}{\gamma} \right)^2},$$

where $\epsilon$ is the total energy, $\gamma$ the half-width at half-maximum, and $\chi$ the center. This Cauchy forms offers the heavier tails required by data without the need for superfluous degrees of freedom.

In order to avoid showers fitting to noise or other detector artifacts, the priors on the Cauchy parameters were chosen to penalize showers with nonphysical energies and widths. The total energy and width followed relatively diffuse Gamma priors,

$$p(\epsilon) = \frac{b^a}{\Gamma(a)} \epsilon^{a-1} e^{-b\epsilon};\ a = 1,\ b = \frac{1}{15} \mathrm{keV}^{-1}$$

$$p(\gamma) = \frac{b^a}{\Gamma(a)} \gamma^{a-1} e^{-b\gamma};\ a = \frac{81}{49},\ b = \frac{36}{49} \mathrm{cm}^{-1},$$

while the position was given a broad Gaussian prior,

$$p(\chi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[ -\frac{1}{2} \left( \frac{\chi}{\sigma} \right)^2 \right];\ \sigma = 7\,\mathrm{cm}.$$

---

[3] For more on Bayesian model selection, see Appendix C.

Ignoring the skewness of the Gamma distributions, these priors were schematically

$$\begin{aligned}
\epsilon &\sim 15 \pm 15\,\mathrm{keV} \\
\gamma &\sim 2.25 \pm 1.75\,\mathrm{cm} \\
\chi &\sim 0 \pm 7\,\mathrm{cm}.
\end{aligned}$$

Lastly note that the width and position priors were specified in terms of transverse location along the respective BSMD plane, centered at the energy-weighted mean of the shower.

Because the BSMD cathodes are sensitive not to point samples of the incident showers but to integral energies, the likelihood was built up from integrals of the shower over the width of each strip,

$$\begin{aligned}
\tilde{E}_i &= \int_{x_i}^{x_{i+1}} \mathrm{d}x\, E(x) \\
&= \frac{\epsilon}{\pi} \left[ \arctan\left( \frac{x_{i+1} - \chi}{\gamma} \right) - \arctan\left( \frac{x_i - \chi}{\gamma} \right) \right];
\end{aligned}$$

the extension to two showers is immediate. The measurement model itself assumed the gaussian limit of Poisson noise similar to that seen in a calorimeter,

$$\begin{aligned}
p(\mathbf{E}|\epsilon,\gamma,\chi) &= \prod_i p(E_i|\epsilon,\gamma,\chi) \\
&= \left( 2\pi\delta\tilde{E}_i^2 \right)^{-\frac{1}{2}} \exp\left[ -\frac{1}{2} \frac{\left( E_i - \tilde{E}_i \right)^2}{\delta \cdot \tilde{E}_i^2} \right],
\end{aligned}$$

where $\delta = 0.125$ roughly matched the fluctuations seen in data and simulation. Noise models more appropriate for energy deposition in gas are awkward to implement computationally and don't provide qualitatively different results than the calorimeter model.

With the model completely specified by the priors and likelihood, the Bayesian evidence for the one and two shower models

$$Z_1 = \int \mathrm{d}\phi \, p(\mathbf{E}|\phi) \, p(\phi)$$

$$Z_2 = \int \mathrm{d}\phi_1 \mathrm{d}\phi_2 \, p(\mathbf{E}|\phi_1, \phi_2) \, p(\phi_1) \, p(\phi_2)$$

$$\phi = \{\epsilon, \gamma, \chi\} \, ,$$

was computed with nested sampling [64]. An efficient implementation of nested sampling using constrained Hamiltonian Monte Carlo was developed especially for this purpose [65].

The log of ratio of the resulting evidences, known as the log-odds, supplies the actual model selection: when the ratio is positive the data prefers the single shower model, while negative values support the double shower model.

| | |
|---|---|
| **dLogZeta** | The Bayesian log-odds between a single shower hypothesis and a double shower hypothesis in the $\eta$ plane of the BSMD. |
| **dLogZphi** | The Bayesian log-odds between a single shower hypothesis and a double shower hypothesis in the $\phi$ plane of the BSMD. |

Note that the log-odds were not used to provide a hard constraint on the data, but rather entered into the analysis the same as all of the other discriminants. Consequently, the precise validity of the assumptions in this shower model was not particularly relevant; provided that the model captured the qualitative features of the shower the resulting discriminant would provide the desired sensitivity.

## 3.4 Analysis Techniques

With each real and simulated event reduced to a manageable number of discriminants, the analysis continued to the extraction of the direct photon signal from the overwhelming QCD background. The limited sensitivity of the variables, however, required a careful use of the data in order to take advantage of all available information.

Even a successfully extracted spectrum, however, would be dependent on the reconstructed energy of the photon candidate and not on the true energy. A faithful comparison to theoretical calculations required a proper deconvolution.

### 3.4.1 Direct Photon Extraction

Ideally, the discriminants are sufficiently informative that the signal and background can be partitioned with cuts on the individual variables. In more difficult cases the correlations can be taken into account in order to isolate the two distributions, and in the most difficult cases the two overlap: they are fundamentally irreducible. Unfortunately, the photon discriminants were of the latter case: even taking into account the correlations between the variables, the signal and background could not be separated.

Cuts could still be made, but only at the expense of introducing systematics from signal efficiency and background contamination. The overlap of the direct photon and QCD distributions was so severe, however, that cuts could not separate a region where the signal was even comparable to background and the corrections would have become highly dependent on uncertain properties of the simulation.

Inferring the shape of the signal and background distributions from simulation offered an alternative strategy. Fitting the two shapes to data gave the photon signal even in the presence of significant overlap, and the residuals provided bounds on any bias present in the simulation.

### Creating a Master Discriminant

Incorporating correlations into the classification of events is a fundamental element of machine learning [66]. Here events with known classifications are used to determine a mapping from the original discriminant space to a one-dimensional space where the signal and background distributions are optimally separated, at least within the limitations of the particular algorithm.

The theory of gaussian processes admits a particularly powerful algorithm that, unlike most machine learning algorithms popular in physics, is both fully Bayesian and highly analytic [67]; for details, see Appendix D. A gaussian process induces a mapping from the input discriminants $\tilde{\mathbf{x}}$ to a master discriminant $\tilde{y}$ as a gaussian distribution[4] conditioned on the training data $\{y_i, \mathbf{x}_i\}$,

$$p(\tilde{y}|\mathbf{y}) = \mathcal{N}\big(y|\mathbf{k}^T \mathbf{C}_N^{-1} \mathbf{y}, c - \mathbf{k}^T \mathbf{C}_N^{-1} \mathbf{k}\big) \,,$$

where

$$y_i = \begin{cases} +1, & i \in \text{Signal} \\ -1, & i \in \text{Background} \end{cases} \,,$$

$$k_i = k(\mathbf{x}_i, \tilde{\mathbf{x}}) \,,$$

$$c = k(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}) + \beta^{-1},$$

and

$$(\mathbf{C}_N)_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) + \beta^{-1} \delta_{ij}.$$

The kernel, or covariance function, $k(\mathbf{u}, \mathbf{v})$ determines the variability of the mapping while $\beta$ admits a certain amount of uncertainty in the training data. Of the many possible choices of kernel, this analysis used the robust squared exponential function,

---

[4] To be precise the mean serves as the desired many-to-one mapping, but the variance can be extremely useful in, for example, quantifying the limitations of the training data. It was not considered for this analysis.

$$k(\mathbf{u}, \mathbf{v}) = \theta_1 \exp \left( \sum_{k=1}^{d} \rho_k \left| \mathbf{u} - \mathbf{v} \right|_k^2 \right) + \theta_2.$$

Because of the consistent probabilistic construction of the gaussian process, the hyperparameters, $\{\theta_i, \rho_i\}$, and the training noise, $\beta$, can themselves be inferred from the data through marginalization. Given computational limitations this inference is often limited to the evidence approximation [67], where an optimization algorithm finds the best values for the hyperparameters given the training data. Note that the carefully placed $\{\rho_i\}$ in the squared exponential kernel implement automatic relevance detection [68], where the optimization also determines the importance of each input discriminant to the master discriminant $y$.

The training data consisted of roughly 10,000 signal events and 10,000 background events randomly sampled from the full simulation[5]. Optimal hyperparameters were chosen from an ensemble of non-linear conjugate gradient optimizations, each initiated with a random seed. The resulting relevances conformed to expectation (Table 3.8).

Note that the gaussian process algorithm assumes that all conditioning variables themselves follow Gaussian distributions, which was clearly not true of the discriminants. In order to improve the performance of the gaussian process, each input discriminant was transformed by hand so that the resulting distribution over the data was approximately gaussian (Table 3.9). This mimics a more sophisticated procedure known as warped gaussian processes [69].

### Fitting the Master Discriminant

Mapping the whole of simulation onto the master discriminant, $y$, demonstrated the substantial overlap between the signal and background distributions; the input discriminants simply didn't have the sensitivity to isolate the two distributions. The

---

[5] The number of events was limited by the computational burden of the matrix calculations implicit in the gaussian process mapping: $20,000^2$ doubles just fits into the maximum available RAM on a 32-bit operating system

| **Discriminant** | **Relevance** ($\rho_i$) | **Rank** |
|---|---|---|
| Et | 0.014 | 10 |
| eSeed | 0.047 | 8 |
| dEta | 0.015 | 9 |
| sumConeTowerEtdEt | 0.400 | 2 |
| sumConeTrackPtdEt | 0.166 | 5 |
| sumCandidateTrackPtdEt | 0.055 | 7 |
| f | 0.216 | 4 |
| gEta | 0.112 | 6 |
| gPhi | 0.012 | 11 |
| dLogZeta | 0.622 | 1 |
| dLogZphi | 0.225 | 3 |

Table 3.8: Hyperparameter optimization within the evidence approximation also implements *automatic relevance detection* which ranks each input variable in order of importance to the output discriminant. Due to redundancies amongst the variables there are often multiple optima and selecting only one can introduce superficial differences between variables. In the particular optimum used for this analysis, this is evident in the higher relevance of the BSMD-$\eta$ variables compared to their BSMD-$\phi$ equivalents. As expected, the shower shape discrimination and cone variables most sensitive to the dominant neutral pion background proved to be most important to the construction of the gaussian process.

| **Discriminant** | **Nominal Support** | **Transformation** |
|---|---|---|
| Et | [7, $\infty$) | $(x-7)^{\frac{1}{3}}$ |
| eSeed | [5.25, $\infty$) | $(x-5.25)^{\frac{1}{3}}$ |
| dEta | (-1, 1) | $\log\left(\frac{1-x}{1+x}\right)$ |
| sumConeTowerEtdEt | [1, $\infty$) | $(x-1+\epsilon)^{\frac{1}{3}}$ |
| sumConeTrackPtdEt | [0, $\infty$) | $(x+\epsilon)^{\frac{1}{3}}$ |
| sumCandidateTrackPtdEt | [0, $\infty$) | $(x+\epsilon)^{\frac{1}{3}}$ |
| f | [0, 1] | $\left(\frac{1-x}{x}+\epsilon\right)^{\frac{1}{3}}$ |
| gEta | [0, 1] | $\left(\frac{1-x}{x}+\epsilon\right)^{\frac{1}{3}}$ |
| gPhi | [0, 1] | $\left(\frac{1-x}{x}+\epsilon\right)^{\frac{1}{3}}$ |
| dLogZeta | $(-\infty, \infty)$ | None |
| dLogZphi | $(-\infty, \infty)$ | None |

Table 3.9: Each discriminant was first transformed so that its distribution in data was approximately gaussian. The small number, $\epsilon = 10^{-5}$, avoided numerical singularities.

master discriminant also serves as a map to a one-dimensional space, however, where inferring the signal from a fit of signal and background shapes becomes a practical alternative.

First, events from both data and simulation were partitioned into the desired phase space bins. Within each bin the data, simulation tagged as signal, and simulation tagged as background were then all binned in the master discriminant.

The signal and background shapes were inferred from the resulting simulation histograms. An application of Bayes' Theorem gave the posterior of the signal shape $\mathbf{s}$ conditioned on the weighted and unweighted signal counts in each master discriminant bin, $\tilde{\mathbf{N}}^s$ and $\mathbf{N}^s$ respectively,

$$p\Big(\mathbf{s}|\tilde{\mathbf{N}}^s, \mathbf{N}^s\Big) = p\Big(\tilde{\mathbf{N}}^s, \mathbf{N}^s|\mathbf{s}\Big)\, p(\mathbf{s})\,,$$

where the likelihood was approximated as (see, for example, [70])

$$p\left(\tilde{\mathbf{N}}^s, \mathbf{N}^s | \mathbf{s}\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi s_i \tilde{N}_i^s / N_i^s}} \exp\left[ -\frac{1}{2} \frac{\left(\tilde{N}_i^s - s_i\right)^2}{s_i \tilde{N}_i^s / N_i^s} \right]$$

and a Dirichlet prior ensured that the shape normalized to unity,

$$p(\mathbf{s}) = \mathrm{Dir}(\mathbf{s}|\mathbf{1}) \,.$$

The posterior for the background shape was analogously constructed.

Taking a point estimate of the shapes, the normalization of both signal and background within a given phase space bin could then be inferred from the data histogram $\mathbf{x}$,

$$\begin{aligned} p(\lambda_s, \lambda_b | \mathbf{x}, \mathbf{s}, \mathbf{b}) &= p(\mathbf{x} | \lambda_s, \lambda_b, \mathbf{s}, \mathbf{b}) \, p(\lambda_s, \lambda_b) \\ &= \left[ \prod_{i=1}^{n} \frac{\lambda_i^{x_i} e^{-\lambda_i}}{x_i!} \right] \cdot p(\lambda_s, \lambda_b) \end{aligned}$$

where

$$\lambda_i = \lambda_s s_i + \lambda_b b_i$$

and

$$p(\lambda_s, \lambda_b) \propto 1.$$

Incorporating the uncertainty of the shapes, however, required building a joint posterior (Figure 3-1)

$$\begin{aligned} &p\left(\lambda_s, \lambda_b, \mathbf{s}, \mathbf{b} | \mathbf{x}, \tilde{\mathbf{N}}^s, \mathbf{N}^s, \tilde{\mathbf{N}}^b, \mathbf{N}^b\right) = \\ &\qquad\qquad p(\mathbf{x} | \lambda_s, \lambda_b, \mathbf{s}, \mathbf{b}) \, p\left(\mathbf{s} | \tilde{\mathbf{N}}^s, \mathbf{N}^s\right) p\left(\mathbf{b} | \tilde{\mathbf{N}}^b, \mathbf{N}^b\right) p(\lambda_s, \lambda_b) \end{aligned}$$

and marginalizing over the shapes explicitly (Figure 3-2)

$$p\Big(\lambda_s, \lambda_b | \mathbf{x}, \tilde{\mathbf{N}}^s, \mathbf{N}^s, \tilde{\mathbf{N}}^b, \mathbf{N}^b\Big) = \int \mathrm{d}\mathbf{s} \, \mathrm{d}\mathbf{b} \, p\Big(\lambda_s, \lambda_b, \mathbf{s}, \mathbf{b} | \mathbf{x}, \tilde{\mathbf{N}}^s, \mathbf{N}^s, \tilde{\mathbf{N}}^b, \mathbf{N}^b\Big) \, .$$

Similarly, the posterior for the total signal yield alone was given by further marginalizing over the total background yield,

$$p\Big(\lambda_s | \mathbf{x}, \tilde{\mathbf{N}}^s, \mathbf{N}^s, \tilde{\mathbf{N}}^b, \mathbf{N}^b\Big) = \int \mathrm{d}\lambda_b \, p\Big(\lambda_s, \lambda_b | \mathbf{x}, \tilde{\mathbf{N}}^s, \mathbf{N}^s, \tilde{\mathbf{N}}^b, \mathbf{N}^b\Big) \, .$$

The construction of the joint normalization/shape posteriors and subsequent marginalizations was performed with Markov Chain Monte Carlo [66, 68], in particular the powerful Hamiltonian Monte Carlo framework [71]. Special methods were developed to ensure that the unique support of the Dirichlet prior on both shapes was preserved in the sampling [72].

Note that, because the shapes and normalizations are factored in the posterior, the extraction is manifestly insensitive to the relative yields of signal and background assumed in the simulation. This indifference is immediately evident within simulation when the ratio of signal to background is different for the sample used to infer the shapes and the sample treated as data (Figure 3-3).

### Systematics

This signal extraction approach is fundamentally dependent upon the assumed validity of simulation, not unlike any cut-based scheme requiring signal efficiency and background contamination corrections. Unlike the cut-based scheme, however, this approach offers a natural means of validating that assumption.

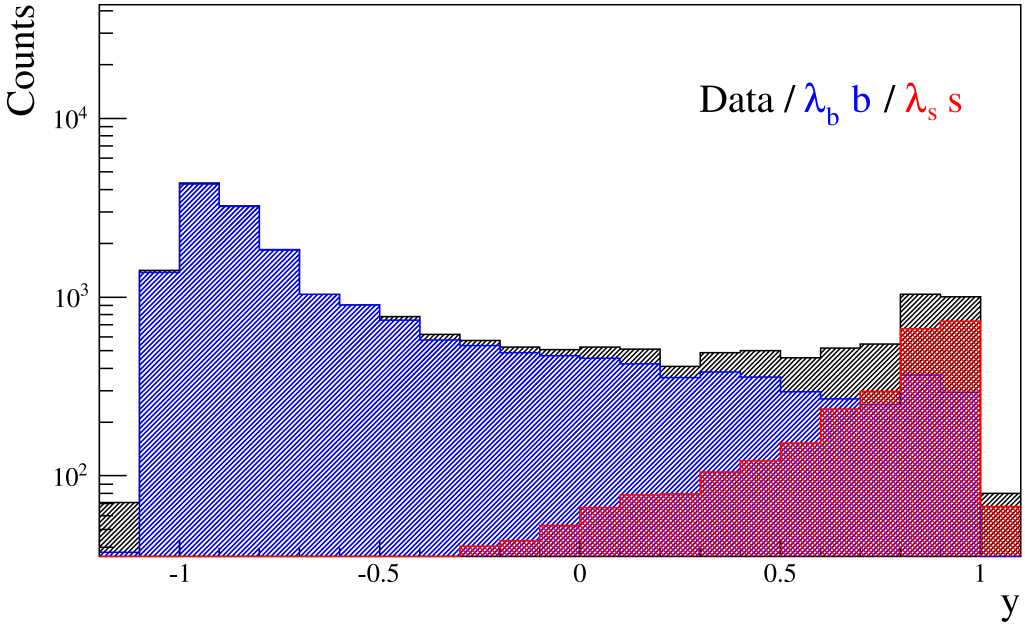
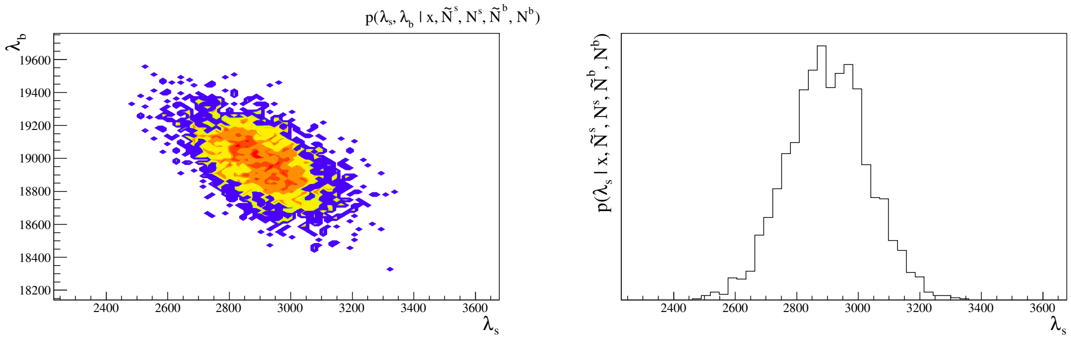
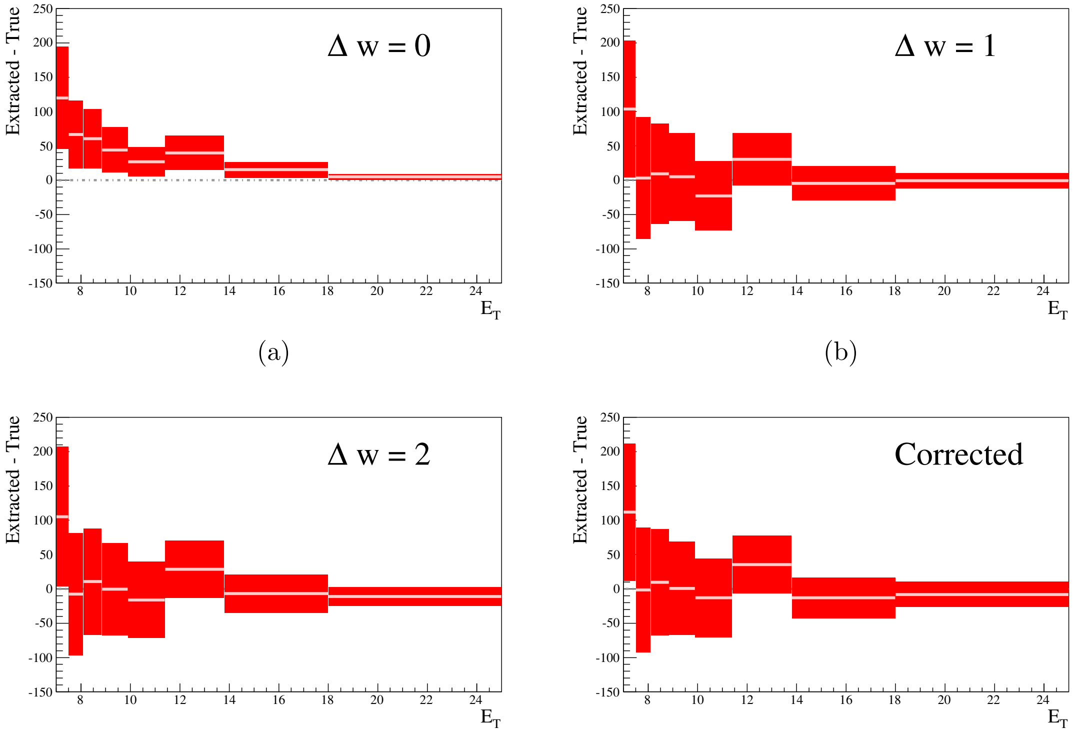


Figure 3-1: The modes of the normalized signal and background shapes from the full joint posterior produced distributions of the master discriminant that were qualitatively consistent with the data. A more careful analysis of the residuals, however, revealed non-trivial systematic discrepancies for some of the phase space bins.



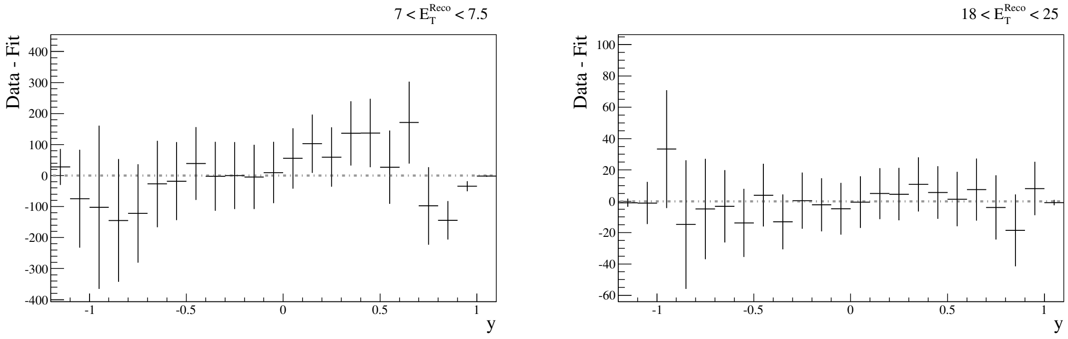
(a) (b)

Figure 3-2: Marginalizing the joint posterior gave (a) the posterior for the signal and background yields and then (b) the posterior for the signal yield alone. The marginalized signal posteriors in each phase space bin were consistent with gaussian distributions, justifying their characterization with only a mean and variance.



(a)
(b)
(c)
(d)

Figure 3-3: The posterior extraction residuals, defined as the difference between the inferred signal yield and the true yield and represented here by the mean and 68% credibility interval, were consistent with zero even when the signal in the pseudo-data was reweighted by (a) zero, (b) one, (c) two, and (d) a $E_T$-dependent weight that corrected the direct photon cross section assumed in PYTHIA to that seen in the Run 9 data. Because the same partition of the simulation was used and the reweighting does not change the effective sample size of the pseudo-data, the statistical fluctuations are indeed expected to be identical in each test, except for the case of no signal where the positivity constraint biases the fluctuations.



(a) (b)

Figure 3-4: Discrepancies between the data and simulation were significant at (a) lower energies where they manifested as systematic behavior in the signal extraction residuals, but not at (b) higher energies.

Provided the simulated signal and background shapes are consistent with reality, the fit spectra should agree with the data spectra up to the expected sampling uncertainties; in particular, the fit residuals[6] should be independently distributed around zero. Any systematic deviation from this behavior indicates deficiencies in the simulation, and a careful quantitative study can be used to bound the possible bias of the extracted signal. These bounds encompass all possible sources of data/simulation discrepancy from missing higher-order effects to incomplete detector reconstruction to calibration biases.

Consider the fit residuals from a low $E_T$ phase space bin and a high $E_T$ bin (Figure 3-4): the latter was consistent with statistical fluctuations and validated the simulation at high energies, but the former showed significant systematic deviations. The residuals were partitioned into three regions of particular behavior and, in order to separate the systematic effects from the expected noise, each region was fit to a functional form (Figure 3-5).

At positive values of the output discriminant the residuals showed a small deficit, which would be, in the worst case, due entirely to a deficit in the signal. At more neutral values there was a clear linear trend resulting in an excess towards the signal;

---

[6] In particular, the residuals are defined as the difference between the data and the posterior mode of the summed signal and background spectra and the associated uncertainties given by the likelihood alone. This approximates a posterior predictive check [73].

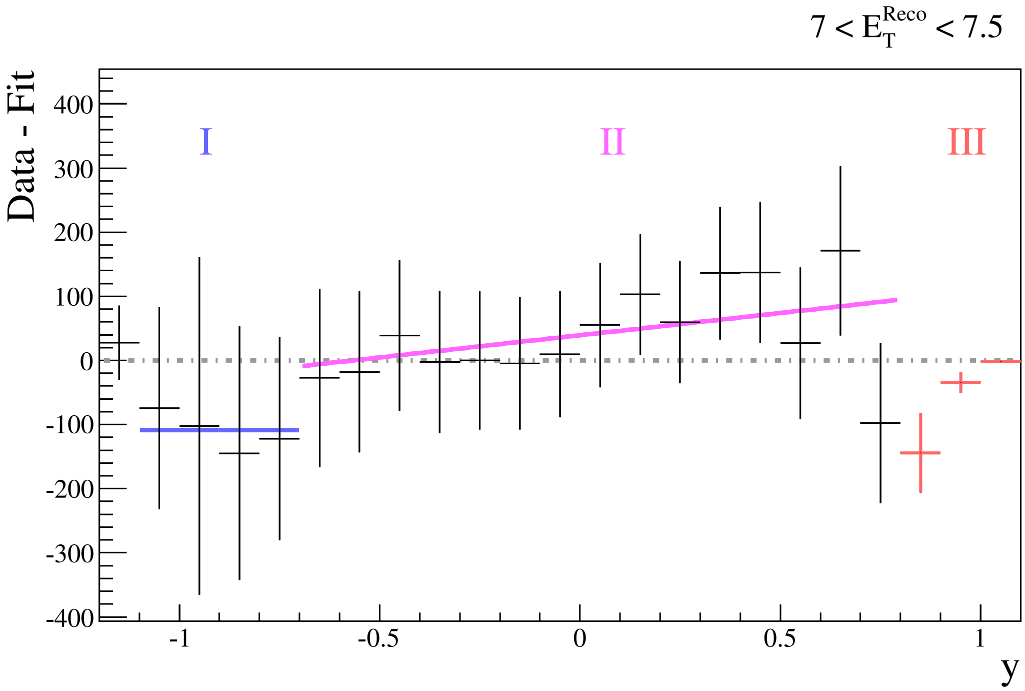


Figure 3-5: The systematic behavior of the residuals was consistent across the phase space bins, partitioning into three regions across the support of the discriminant.

again the conservative approach placed the responsibility wholly on the signal, in this case a surplus. Towards the background dominated region of negative $y$ the residuals again showed a deficit but the effect on the signal extraction was more subtle as the deficit was far from the signal distribution. Here the cautious approach considered the fault to lie in the background, with the signal absorbing any changes from scaling the background to correct the deviation (Table 3.10). In the end, all excesses and deficits were added in quadrature to produce low and high systematics, respectively.

This general behavior was consistent across all of the lower energy bins, gradually falling in magnitude until indistinguishable from noise at higher energies. Consequently the same regions and functional forms were used for computing the systematic bounds across all of phase space.

| **Region** | **Interval** | **Functional Form** | **Systematic Bound** | **Typical Behavior** |
|---|---|---|---|---|
| I | $(-1.2, -0.7)$ | $c$ | $\frac{c}{\lambda_b \max b_i} \cdot \sum_{i \vert y_i > 0} \lambda_b b_i$ | Deficit |
| II | $(-0.7, 0.8)$ | $mx + b$ | $\int_0^{0.8} \mathrm{d}y \ (mx + b)$ | Excess |
| III | $(0.8, 1.1)$ | None | $\sum_{i \vert y_i > 0.8} r_i$ | Deficit |

Table 3.10: The systematic deviation of the residuals from pure sampling uncertainties separated into three dominant behaviors, a constant deficit as negative $y$, a linear excess at central $y$, and a small deficit at positive $y$. A separate functional form was fit in each of these regions to isolate the systematic behavior from the noise, except in the last region where there are not enough bins to constrain any fit.

### 3.4.2 Deconvolution

Because the phase space bins are defined by the reconstructed energy of the photon candidates, the signal extraction procedure produced a reconstructed spectrum, $\mathbf{d}$, and not the true spectrum, $\mathbf{t}$, necessary for comparisons with theory. The detector resolution separating true energy from reconstructed energy induced a convolution between the two spectra; ignoring uncertainties such a convolution is typically written as,

$$\mathbf{d} = \mathbf{H}\mathbf{t},$$

where $\mathbf{H}$ is the *transfer matrix* encoding the specifics of the convolution. In particular, the element $H_{ij}$ gives the probability of an even in the $j$th true bin falling into the $i$th reconstructed bin.

Any attempt to invert the system directly fails spectacularly. Because the transfer matrix is, in general, not square the usual matrix inverse does not exist and, even if the matrix were square, any numerical solution is ill-posed: there is an infinite number of solutions satisfying the system of equations and inversion is unstable.

In order to pose the problem well, additional information must be introduced to reduce the space of solutions. Typically this information comes in terms of an assumption about the properties of the true spectrum, such as smoothness, implicit in the choice of a particular functional form. The choice of a functional form regularizes the problem and admits a clean solution, but care must be taken to ensure that the choice does not bias the deconvolution.

Given the construction of a transfer matrix and the choice of an appropriate functional form, a joint posterior was built over all free parameters in the system and marginalized to give a posterior for the true spectrum alone.

### The Transfer Matrix

The construction of the transfer matrix relies, in general on the simulated detector response. Comparing the true energy of direct photons incident on the BEMC with their known energy, for example, furnishes the value

$$H_{ij} = \lim_{n_j \to \infty} \frac{n_{i|j}}{n_j}.$$

Without infinite simulation statistics, however, the estimates of the transfer matrix are imprecise, with the uncertainty correlated along each column. These finite counts were incorporated into the deconvolution with a full probabilistic model for each column $\mathbf{h}_j \equiv H_{\cdot j}$,

$$\begin{aligned} p(\mathbf{h}_j | \mathbf{s}_j, \beta) &\propto p(\mathbf{s}_j | \mathbf{h}_j)\, p\left(\mathbf{h}_j | \boldsymbol{\beta}\right) \\ &\propto \mathrm{Mult}(\mathbf{s}_j | \mathbf{h}_j)\, \mathrm{Dir}(\mathbf{h}_j | \boldsymbol{\beta}) \\ &= \mathrm{Dir}(\mathbf{h}_j | \mathbf{s}_j + \boldsymbol{\beta})\,, \end{aligned}$$

where $\mathbf{s}_j$ is the spectrum of simulation events falling into the $j$th true bin[7]. The Dirichlet hyperparameters were taken to be $\beta_i = 10^{-3}$.

---

[7] Because of the binning by true energy, the simulation weights can be ignored to first order. This admits the Multinomial likelihood instead of the more complicated importance sampling approximation used in the signal extraction.

Defining $\mathbf{S} \equiv \{\mathbf{s}_j\}$, the distribution for the entire transfer matrix became

$$p(\mathbf{H}|\mathbf{S}, \boldsymbol{\beta}) = \prod_{j=1}^{N+2} \mathrm{Dir}(\mathbf{h}_j | \mathbf{s}_j + \boldsymbol{\beta}) \, .$$

Note that the transfer matrix is not square, as events falling below and above the nominal energy thresholds can reconstruct in the measured spectrum. Because the spectra of interest are steeply falling, this leakage is significant at the low energy threshold and cannot be ignored lest the events be forced into the analysis window and bias the answer.

### Modeling the True Spectrum

Ultimately the deconvolution is limited by the choice of a functional form for the true spectrum. Parametric functions are well understood and easy to manipulate, but they can be overly restrictive when applied to general deconvolution problems. While low-order polynomials adequately span the solution space for soft spectra, for example, they cannot fully cover the space of solutions for harder spectra with exponential or power-law behavior and would lead to a biased deconvolution. Non-parametric functions[8], however, have the flexibility to span both solution spaces at once.

Ubiquitous in non-parametric methods, kernel density estimators construct a function as a sum of independent kernels,

$$f(x) = \sum_{i=1}^{M} a_i k(x, \theta_i) \, ,$$

Gaussian kernels in particular provide a robust form for smooth functions,

$$f(x) = \sum_{i=1}^{M} a_i \left( 2\pi w^2 \right)^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( \frac{x - r_i}{w} \right)^2 \right] .$$

---

[8] Non-parametric does not refer to the absence of parameters but rather a sufficiently large number of parameters such that the data can determine the necessary structure of the function itself. Examples include splines and the gaussian processes considered above.

The kernels are distributed across phase space as necessary, often on a uniform grid, with the width $w$ determining the overall smoothness of the function. Here the width was included in the model posterior and then marginalized, allowing the data to determine the necessary smoothness.

Besides a preference for smooth solutions, there is an additional piece of information that can help to regularize the problem at hand. Convolution redistributes the events but it cannot create or destroy them: the total integral of the true spectrum must be conserved. This was incorporated into the kernel density estimator by rewriting the amplitudes as

$$a_i = n\hat{a}_i,$$

where the total integral $n$ was estimated from the measured spectrum,

$$p(n|\mathbf{d}, \boldsymbol{\sigma}) = \mathcal{N}\left( n \middle| \sum_i d_i, \sum_i \sigma_i^2 \right),$$

and the relative amplitudes were modeled with a Dirichlet prior to ensure the integral conservation,

$$p(\hat{\mathbf{a}}) = \mathrm{Dir}(\hat{\mathbf{a}}|\boldsymbol{\alpha})\,.$$

The Dirichlet hyperparameters were taken to be $\alpha_i = 1$.

### Measurement Model

Given the non-parametric model of the true spectrum, the binned spectrum was reconstructed by integrating over each bin,

$$t_i = \int_{E_T^i}^{E_T^{i+1}} \mathrm{d}E_T\, f\left(E_T\right).$$

An application of the transfer matrix gave the reconstructed spectrum,

$$m_i = \sum_{j=1}^{N+2} H_{ij} t_j,$$

from which the data were assumed to be distributed,

$$p(\mathbf{d}|\boldsymbol{\sigma}, \mathbf{a}, \mathbf{H}, w) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[ -\frac{1}{2} \left( \frac{m_i - d_i}{\sigma_i} \right)^2 \right].$$

With the above considerations, the posterior over all uncertain parameters became

$$p(\hat{\mathbf{a}}, n, \mathbf{H}, w|\mathbf{d}, \boldsymbol{\sigma}, \mathbf{S}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \propto p(\mathbf{d}|\boldsymbol{\sigma}, \mathbf{a}, \mathbf{H}, w)\, p(n|\mathbf{d}, \boldsymbol{\sigma})\, p(\mathbf{H}|\mathbf{S}, \boldsymbol{\beta})\, p(w)\,,$$

where all but $p(w)$ have been defined above. The prior width was taken to be uniform.

While the joint posterior is built from the parameters of the kernel density estimator, the desired posterior was really of the binned true spectrum $\mathbf{t}$. Because this spectrum was deterministically related to the kernel density estimator, however, the posterior for the true spectrum was immediately given by a change of variables and a subsequent marginalization,

$$p(\mathbf{t}|\mathbf{d}, \boldsymbol{\sigma}, \mathbf{S}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \int \mathrm{d}\mathbf{q}(\hat{\mathbf{a}}, n, \mathbf{H}, w)\ p(\mathbf{t}(\hat{\mathbf{a}}, n, \mathbf{H}, w)\,, \mathbf{q}(\hat{\mathbf{a}}, n, \mathbf{H}, w)\,|\mathbf{d}, \boldsymbol{\sigma}, \mathbf{S}, \boldsymbol{\alpha}, \boldsymbol{\beta})\,.$$

Once again the computationally challenging task of constructing the joint posterior and then marginalizing the unnecessary parameters was handled with Hamiltonian Monte Carlo methods.

### 3.4.3 Integrated Analysis

Given the above methods to extract and then deconvolve the direct photon spectrum, the final analysis was straightforward. The kinematic phase spaces of both the data and simulation were first partitioned and the direct photon yield inferred within each bin. This extraction yielded a raw photon spectrum in terms of the reconstructed transverse energy, providing the necessary information for the inference of the direct photon spectrum in terms of the true incident energy.

# Chapter 4

# Results

With procedure to extract the direct photons from the data and properly deconvolve the resulting spectra, the direct photon measurements could finally be constructed. In order to ensure an understanding of the hard interaction and the ability to infer the initial-state gluon from the final state photon, an unpolarized cross section was first composed and compared to the perturbative QCD calculation. Confident of the accuracy of the calculations, the extracted signal was then manipulated into the double-helicity asymmetry sensitive to the underlying gluon polarization of the proton.

## 4.1 Cross Section

The relativistically-invariant cross section, differential over the kinematic phase space, is defined as

$$E\frac{\mathrm{d}^3\sigma}{\mathrm{d}p^3} = \frac{\mathrm{d}^3\sigma}{E_T\,\mathrm{d}E_T\,\mathrm{d}\eta\,\mathrm{d}\phi}.$$

Measurements averaged the fully differential cross section over given partitions of phase space,

$$\begin{aligned}\frac{1}{\left(\delta E_T\right)_i}\frac{1}{\left(\delta\eta\right)_i}\frac{1}{\left(\delta\phi\right)_i}\int_{\left(\delta E_T\right)_i}\int_{\left(\delta\eta\right)_i}\int_{\left(\delta\phi\right)_i} \mathrm{d}E_T\,\mathrm{d}\eta\,\mathrm{d}\phi\,\frac{\mathrm{d}^3\sigma}{E_T\,\mathrm{d}E_T\,\mathrm{d}\eta\,\mathrm{d}\phi}&\\ =&\frac{\left\langle\sigma\right\rangle_i}{\left\langle E_T\right\rangle_i\left(\delta E_T\right)_i\left(\delta\eta\right)_i\left(\delta\phi\right)_i},\end{aligned}$$

where the cross section averaged over the $i$th partition, $\left\langle\sigma\right\rangle_i$, is defined in terms of the experimental values

$$\left\langle\sigma\right\rangle_i = \frac{1}{\epsilon_i}\frac{\lambda_i}{\tilde{\mathcal{L}}}.$$

The efficiency $\epsilon_i$ accounts for any data lost in the measurement of the deconvolved direct photon yield $\lambda_i$, while the integrated luminosity $\tilde{\mathcal{L}}$ normalizes the counts to the delivered beam intensity.

Here the phase space was divided into eight partitions, specified by eight bins of transverse energy and a single bin over $\eta$ and $\phi$. The boundaries of the transverse energy bins, in GeV, were $\{7,\ 7.5,\ 8.1,\ 8.85,\ 9.9,\ 11.4,\ 13.8,\ 18,\ 25\}$ with $-0.7 \leq \eta \leq 0.7$ and full azimuthal coverage, $-\pi \leq \phi \leq \pi$. Note that $\eta$ here refers to the angle relative to the vertex and not the detector geometry: given the restriction to vertices within 60 cm of the nominal interaction point, the photons were confined to $-0.9 \leq \eta_{\mathrm{D}} \leq 0.9$ so that the boundary of the BEMC was not an issue.

Once the experimental parameters entering into the measurement were calculated for each partition of phase space and their uncertainties understood, the full cross section was constructed and finally compared to the theoretical calculation.

### 4.1.1 Efficiencies

No matter the sophistication of the signal extraction, the measurement of direct photons will always suffer inefficiencies in the collection and reduction of the data itself. Events without a central reconstructed vertex, events not firing the online trigger, or even events failing to reconstruct as a photon candidate are all lost in the process but must be considered in the final measurement.

The efficiencies were calculated together by comparing the propagation of simulated direct photons through the detector model and subsequent analysis, save for the inefficiency of the requirement of a central vertex along the beamline. This vertex efficiency was computed analytically, assuming a gaussian distribution of the reconstructed vertex position along the beamline fit to the central data.

Note that efficiencies are fundamentally an inverse binomial process, with the raw data $N_i$ representing only the successful trials and the total number of events $N'_i$, let alone the true yields $\lambda'_i$, unknown. Given reasonable statistics, however, the proper treatment of this process converges to the gaussian typically assumed in physics analyses,

$$p(\lambda'_i|\lambda_i, \epsilon_i) \to \mathcal{N}\left( \lambda'_i \,\middle|\, \frac{\lambda_i}{\epsilon_i}, \frac{\sigma^2_{\lambda_i}}{\epsilon_i^2} \right),$$

where $\lambda_i$ is the yield estimated from the raw counts $N_i$ alone.

Assuming a beta prior conjugate to the inverse binomial likelihood, the posterior for each efficiency was given by independent beta distributions (Table 4.1)

$$p(\epsilon_i|\alpha_i, \beta_i) = \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)\,\Gamma(\beta_i)} \epsilon_i^{\alpha_i - 1} \left(1 - \epsilon_i\right)^{\beta_i - 1}.$$

Convolving the efficiency distributions with the corrected yields was a straightforward application of Monte Carlo integration.

### 4.1.2 Luminosity

Once corrected, the binned counts become average cross sections by normalizing to the component of the beam intensity responsible for hard interactions: the intensity of the beam overlap, or luminosity.

| $\mathbf{E_T}$ (GeV) | **Efficiency** | | | |
|---|---|---|---|---|
| | $\boldsymbol{\alpha_i}$ | $\boldsymbol{\beta_i}$ | **Mean** | **Variance** |
| 7 - 7.5 | 865.2 | 3540.9 | 0.1964 | $3.581 \times 10^{-5}$ |
| 7.5 - 8.1 | 1216.3 | 2873.7 | 0.2974 | $5.108 \times 10^{-5}$ |
| 8.1 - 8.85 | 1167.0 | 2286.4 | 0.3379 | $6.477 \times 10^{-5}$ |
| 8.85 - 9.9 | 1033.1 | 1904.0 | 0.3518 | $7.761 \times 10^{-5}$ |
| 9.9 - 11.4 | 793.4 | 1391.7 | 0.3631 | $1.058 \times 10^{-4}$ |
| 11.4 - 13.8 | 551.6 | 919.7 | 0.3749 | $1.592 \times 10^{-4}$ |
| 13.8 - 18 | 280.0 | 444.5 | 0.3865 | $3.268 \times 10^{-4}$ |
| 18 - 25 | 72.3 | 115.7 | 0.3847 | $1.253 \times 10^{-3}$ |

Table 4.1: The posteriors for the total efficiency in each bin was modeled with independent beta distributions.

From the collider perspective the luminosity accumulated over the entire run is given by

$$\tilde{\mathcal{L}} = \int \mathrm{d}t \, \mathcal{L} = \int \mathrm{d}t \, f \frac{N_Y N_B}{A},$$

where $N_Y$ and $N_B$ are the beam multiplicities within the overlap region $A$, and $f$ is the bunch crossing frequency. When constructing a measurement, however, it is easier to estimate the integrated luminosity using a process with a known cross section,

$$\tilde{\mathcal{L}} \approx \frac{N}{\sigma},$$

where the counts $N$ span all runs used in the analysis.

The cross section for events triggered by a minimum bias condition, defined as a coincidence between hits in the BBCs, was measured in Run 2 with a 7% relative uncertainty [74],

$$\sigma_{\mathrm{MB}} \sim 26.1 \pm 1.8 \, \mathrm{mb}.$$

In Run 9, the BBCMB-Cat3 [75] trigger utilized the minimum bias condition while reading out the same suite of detectors as L2BGamma, ensuring equal dead time and minimizing any bias in the calculation.

Integrating the minbias triggers over all runs in the *st_gamma* streams gave[1]

$$\tilde{\mathcal{L}} \approx 13.5\,\mathrm{pb}^{-1},$$

with the relative error dominated by the cross section uncertainty.

Any bias from residual differences in the L2BGamma and BBCMB-Cat3 definitions, or changes in the BBC response since the cross section was measured, was assumed to be negligible compared to the original cross section uncertainty, but the total luminosity uncertainty was increased to 10% as a conservative measure.

### 4.1.3 Systematic Errors

In addition to the luminosity uncertainty, two additional sources of systematic uncertainty contributed to the final measurement: bias of the signal extraction and dependence of the deconvolution and efficiency corrections on the BEMC calibration. Both were incorporated by repeating the analysis multiple times, in the former case only the steps after the signal extraction but in the latter case everything from candidate reconstruction, with the relevant parameters shifted by the estimated systematics[2] and the resulting cross section compared to the nominal value.

The shifts were treated independently and, without any model of the distribution of these effects, the contributions combined in quadrature. Note that, per STAR convention, the luminosity was treated separately and not explicitly added to the total systematic error.

---

[1] Because of resource limitations, the minimum bias triggers were highly prescaled: only a small fraction of events satisfying the minimum bias condition are actually accepted. The raw trigger counts were corrected for this prescale factor in order ensure an accurate calculation.

[2] The quoted BEMC calibration systematic was known to be conservative and, indeed, the data/simulation agreement significantly deteriorated when the perturbed calibrations were used. Because the overall systematic error of the measurements is dominated by this calibration, the final values should be taken with some reservation.

### 4.1.4 Measurement

Following through the analysis yielded the final values for the cross section measurement (Table 4.2).

The results compare well to the NLO calculation (Figure 4-1), especially considering the experimental and theoretical uncertainties, giving confidence in the ability of the same theoretical tools to constrain the gluon polarization from the measured asymmetry.

## 4.2 Asymmetry

The double-helicity asymmetry is formally defined as the ratio of the polarized and unpolarized invariant cross sections,

$$A_{\mathrm{LL}} = \frac{E\left(\mathrm{d}^3 \Delta\sigma / \mathrm{d}^3 p\right)}{E\left(\mathrm{d}^3 \sigma / \mathrm{d}^3 p\right)} = \frac{\left(\mathrm{d}^3 \sigma^{\uparrow\uparrow} / \mathrm{d}^3 p\right) - \left(\mathrm{d}^3 \sigma^{\uparrow\downarrow} / \mathrm{d}^3 p\right)}{\left(\mathrm{d}^3 \sigma^{\uparrow\uparrow} / \mathrm{d}^3 p\right) + \left(\mathrm{d}^3 \sigma^{\uparrow\downarrow} / \mathrm{d}^3 p\right)}.$$

Averaging over a phase space partition, the asymmetry becomes

$$\left(A_{\mathrm{LL}}\right)_i = \frac{\left\langle \sigma^{\uparrow\uparrow} \right\rangle_i - \left\langle \sigma^{\uparrow\downarrow} \right\rangle_i}{\left\langle \sigma^{\uparrow\uparrow} \right\rangle_i - \left\langle \sigma^{\uparrow\downarrow} \right\rangle_i},$$

or, for an experiment with perfect beams,

$$\left(A_{\mathrm{LL}}\right)_i = \frac{\lambda_i^{\uparrow\uparrow} - \lambda_i^{\uparrow\downarrow}}{\lambda_i^{\uparrow\uparrow} + \lambda_i^{\uparrow\downarrow}}.$$

Note that only the deconvolved direct photo yields $\lambda_i$ remain; the efficiencies and luminosity cancel in the ratio, dramatically reducing sources of systematic bias.

| $E_T$ (GeV) | Cross Section (pb $c^3$ / GeV$^2$) | | | | | | | | | Systematic Errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Covariance | | | | | | | | Signal Extraction | BEMC Calibration | Luminosity |
| | | 7 - 7.5 | 7.5 - 8.1 | 8.1 - 8.85 | 8.85 - 9.9 | 9.9 - 11.4 | 11.4 - 13.8 | 13.8 - 18 | 18 - 25 | | | |
| 7 - 7.5 | 115.076 | 586.198 | -106.973 | -69.304 | 52.563 | -18.26 | 3.604 | -0.963 | 0.209 | +0.348<br>-2.053 | +18.914<br>-16.545 | ± 10% |
| 7.5 - 8.1 | 58.68 | -64.181 | 48.142 | -5.149 | -7.659 | 3.235 | -1.039 | 0.13 | 0.043 | +0.385<br>-0.417 | +6.936<br>-5.578 | |
| 8.1 - 8.85 | 46.88 | -22.541 | -2.792 | 14.082 | -6.445 | 1.006 | -0.056 | -0.139 | 0.017 | +1.29<br>-0.301 | +3.919<br>-2.516 | |
| 8.85 - 9.9 | 26.505 | 7.128 | -1.732 | -2.688 | 3.688 | -1.155 | 0.09 | 0.024 | -0.018 | +0.46<br>-0.24 | +2.428<br>-3.037 | |
| 9.9 - 11.4 | 12.555 | -0.941 | 0.277 | 0.159 | -0.439 | 0.537 | -0.109 | -0.002 | 0.004 | +0.091<br>-0.069 | +1.27<br>-1.158 | |
| 11.4 - 13.8 | 4.159 | 0.051 | -0.025 | -0.003 | 0.009 | -0.031 | 0.052 | -0.007 | 0 | +0.032<br>-0.011 | +0.235<br>-0.254 | |
| 13.8 - 18 | 1.064 | -0.003 | 0 | -0.002 | 0 | -0.001 | -0.002 | 0.003 | -0.001 | +0.003<br>-0 | +0.095<br>-0.106 | |
| 18 - 25 | 0.124 | 0 | 0 | 0 | -0.001 | 0 | 0 | -0.001 | 0.001 | +0<br>-0 | +0.009<br>-0.015 | |

Table 4.2: The cross section posterior was approximated as a multivariate gaussian distribution, with systematic model errors considered only for the uncorrelated variances.

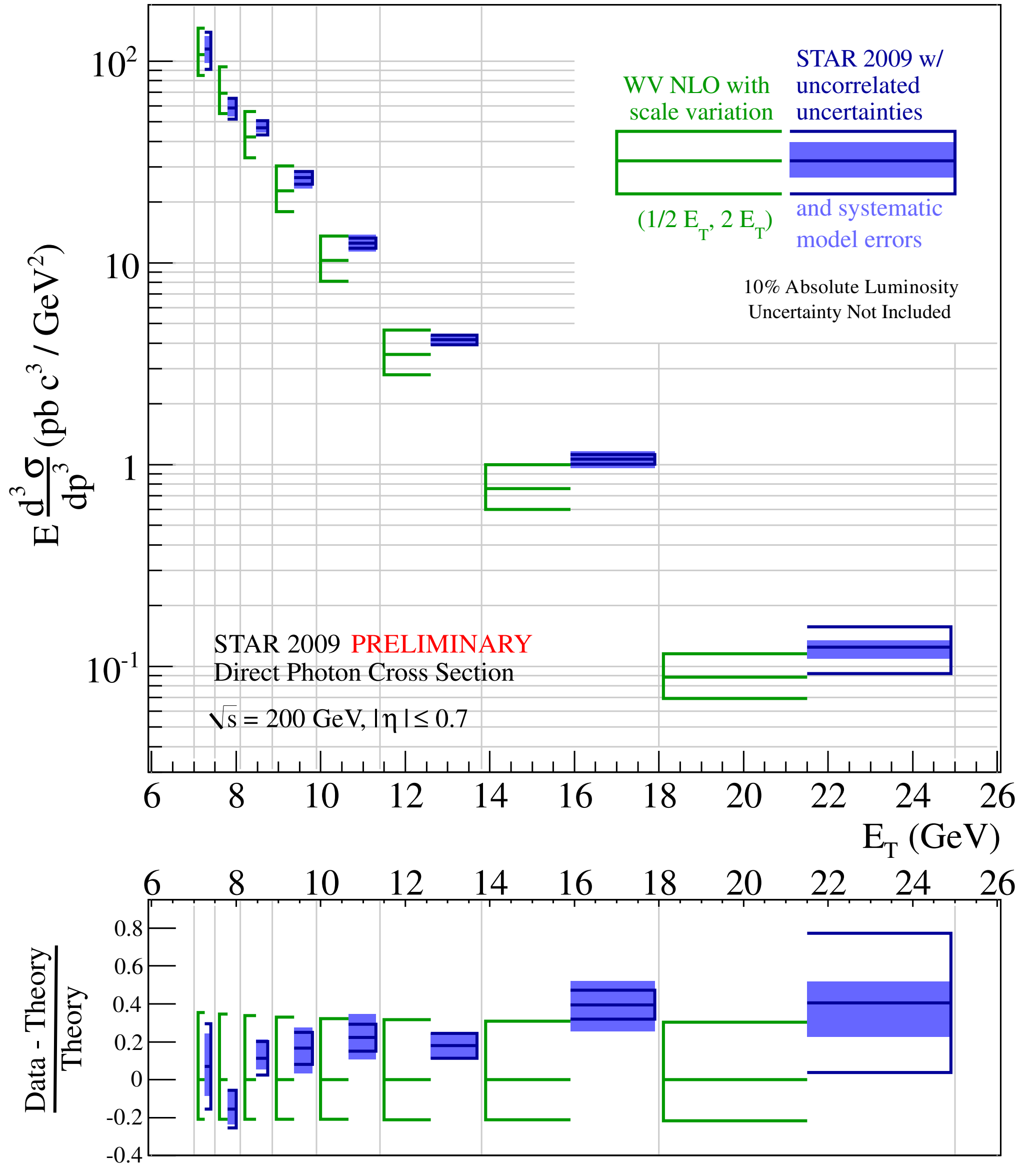


Figure 4-1: The inclusive direct photon cross section is in excellent agreement with the resummed NLO calculation, justifying the use of the perturbative theory in constraining the gluon polarization with the direct photon asymmetry.

In practice the asymmetry must be corrected for imperfect polarization and unbalanced spin states,

$$\left(A_{\mathrm{LL}}\right)_i = \frac{1}{P_Y(t)\,P_B(t)}\,\frac{\lambda i^{\uparrow\uparrow} - R_3(t)\,\lambda_i^{\uparrow\downarrow}}{\lambda_i^{\uparrow\uparrow} + R_3(t)\,\lambda_i^{\uparrow\downarrow}},$$

where $R_3$ is the relative luminosity of the aligned and anti-aligned helicity configurations,

$$R_3(t) \equiv \frac{\mathcal{L}^{\uparrow\uparrow}(t)}{\mathcal{L}^{\uparrow\downarrow}(t)},$$

and $P_Y, P_B$ are the polarizations of the yellow and blue beams, respectively. These time-dependent corrections must be treated particularly carefully as they do not in general commute with the signal extraction; both the polarization and the relatively luminosity are measured throughout the data collection, the former by fill and the latter by run, but the signal extraction is performed only on the whole of the data at once.

Given a means to incorporate these corrections into the analysis and rigorous constraints of any possible systematic errors, the direct photon double-helicity asymmetry was composed for each phase space partition. This measurement used the same partitions defined for the cross section, save for the removal of the lowest $E_T$ bin due to excessive uncertainties.

### 4.2.1 Polarization

Both beam polarizations are measured by fill [76] (Figure 4-2) and the asymmetry is more precisely written as a sum over each fill $j$,

$$\left(A_{\mathrm{LL}}\right)_i = \frac{\sum_j \left(\lambda_{ij}^{\uparrow\uparrow} - R_3(t)\,\lambda_{ij}^{\uparrow\downarrow}\right)}{\sum_j \left(P_Y P_B\right)_j \left(\lambda_{ij}^{\uparrow\uparrow} + R_3(t)\,\lambda_{ij}^{\uparrow\downarrow}\right)},$$

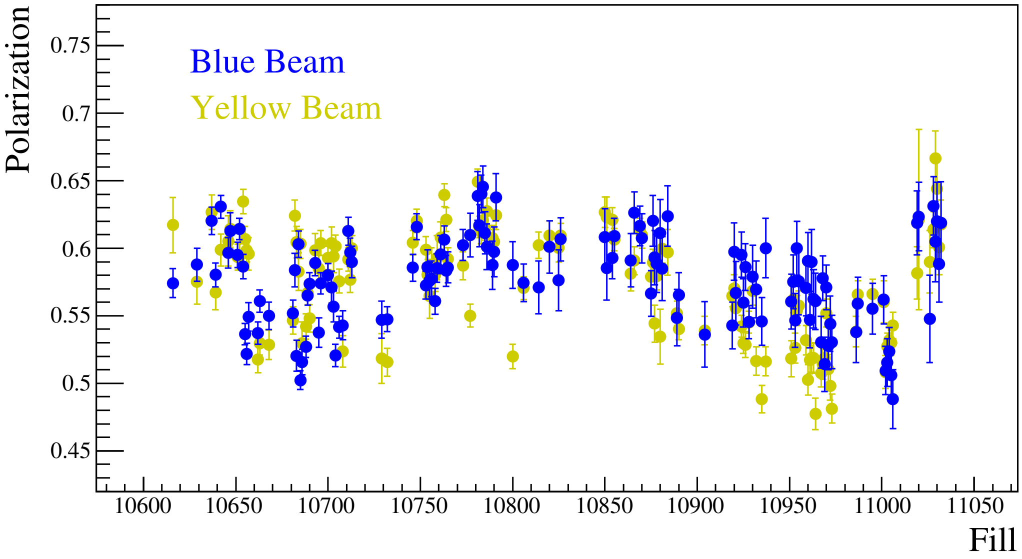


Figure 4-2: The blue and yellow beam polarizations are measured for each fill independently, with corresponding statistical and systematic uncertainties. Only the former are shown here, as the latter are superseded by the constant correlated systematic uncertainty.

or equivalently

$$\begin{aligned}(A_{\mathrm{LL}})_i &= \frac{\sum_j \left(\lambda_{ij}^{\uparrow\uparrow} + R_3(t)\,\lambda_{ij}^{\uparrow\downarrow}\right)}{\sum_j \left(P_Y P_B\right)_j \left(\lambda_{ij}^{\uparrow\uparrow} + R_3(t)\,\lambda_{ij}^{\uparrow\downarrow}\right)} \frac{\sum_j \left(\lambda_{ij}^{\uparrow\uparrow} - R_3(t)\,\lambda_{ij}^{\uparrow\downarrow}\right)}{\sum_j \left(\lambda_{ij}^{\uparrow\uparrow} + R_3(t)\,\lambda_{ij}^{\uparrow\downarrow}\right)} \\ &= \frac{\sum_j \left(\lambda_{ij}^{\uparrow\uparrow} + R_3(t)\,\lambda_{ij}^{\uparrow\downarrow}\right)}{\sum_j \left(P_Y P_B\right)_j \left(\lambda_{ij}^{\uparrow\uparrow} + R_3(t)\,\lambda_{ij}^{\uparrow\downarrow}\right)} \frac{\left(\lambda_{i}^{\uparrow\uparrow} - R_3(t)\,\lambda_{i}^{\uparrow\downarrow}\right)}{\left(\lambda_{i}^{\uparrow\uparrow} + R_3(t)\,\lambda_{i}^{\uparrow\downarrow}\right)} \\ &= \frac{1}{\langle P_Y P_B \rangle} \frac{\left(\lambda_{i}^{\uparrow\uparrow} - R_3(t)\,\lambda_{i}^{\uparrow\downarrow}\right)}{\left(\lambda_{i}^{\uparrow\uparrow} + R_3(t)\,\lambda_{i}^{\uparrow\downarrow}\right)},\end{aligned}$$

where $\langle P_Y P_B \rangle$ is the effective polarization over the entire run,

$$\langle P_Y P_B \rangle \equiv \frac{\sum_j \left(P_Y P_B\right)_j \left(\lambda_{ij}^{\uparrow\uparrow} + R_3(t)\,\lambda_{ij}^{\uparrow\downarrow}\right)}{\sum_j \left(\lambda_{ij}^{\uparrow\uparrow} + R_3(t)\,\lambda_{ij}^{\uparrow\downarrow}\right)}.$$

Assuming uniform helicity configurations, $R_3(t) = 1$ and the yields become

$$\begin{aligned} \lambda_{ij}^{\upharpoonleft\upharpoonright} + R_3(t)\, \lambda_{ij}^{\upharpoonleft\downharpoonright} &= \lambda_{ij}^{\upharpoonleft\upharpoonright} + \lambda_{ij}^{\upharpoonleft\downharpoonright} \\ &= \lambda_{ij}, \end{aligned}$$

with the resulting effective polarization

$$\langle P_Y P_B \rangle \approx \frac{\sum_j \left( P_Y P_B \right)_j \lambda_{ij}}{\sum_j \lambda_{ij}}.$$

Corrections to this approximation take the form

$$\left( R_3(t) \, \frac{R_3\left(r\right) - 1}{R_3\left(t\right) + 1} \right)^2,$$

which were only of the order 1% for the typical relative luminosities encountered through the run and consequently negligible compared to the systematic uncertainties in the polarization measurements themselves. Note that, given this assumption, the polarization is uniform across phase space and the sample average can be computed without regarding any particular partition,

$$\langle P_Y P_B \rangle \equiv \frac{\sum_j \left( P_Y P_B \right)_j \lambda_j}{\sum_j \lambda_j}.$$

Taking the multiplicity of all L2BGamma triggered events in each fill as a proxy for the yields $\lambda_j$ and marginalizing over the statistical uncertainties of the polarizations gave the effective polarization

$$\langle P_Y P_B \rangle = 0.32,$$

where the residual statistical uncertainty was vanishing.

The fill-dependent systematic uncertainties for each beam polarization were dominated by the correlated systematic uncertainty of the product $\left( P_Y P_B \right)_j$, quoted at 8.8% independent of fill which immediately implies the same 8.8% systematic on the

| **Excluded Bunch Crossing** | **Complication** |
|---|---|
| 20, 60 | Kicker Bunches |
| 78, 79, 80 | Peculiar Rate Structure |
| 31-39, 111-119 | Abort Gaps |

Table 4.3: The relative luminosity could not be satisfactorily calculated for some bunch crossings, which must then be removed from any analysis using polarization information.

effective polarization. Because this systematic dominated both the statistical uncertainty and approximation errors, the latter were ignored.

### 4.2.2 Relative Luminosity

Relative luminosities of the various beam helicity configurations were calculated for each run using the BBC coincidences corrected for various detector effects [77, 78, 79] (Figure 4-3). Various bunch crossings, such as in and around the abort gaps, were removed from these calculations and consequently avoided in the polarized measurement as well (Table 4.3).

The run-by-run variation of these measurements were incorporated into the analysis by weighting all events from anti-aligned collisions with the corresponding relative luminosity before signal extraction and deconvolution, in which case the asymmetry becomes

$$(A_{\mathrm{LL}})_i = \frac{1}{\langle P_Y P_B \rangle} \frac{\lambda_i^{\uparrow\uparrow} - \left( R_3 N^{\uparrow\downarrow} \right)_i}{\lambda_i^{\uparrow\uparrow} + \left( R_3 N^{\uparrow\downarrow} \right)_i}.$$

Statistical uncertainties in the relative luminosity were marginalized from the analysis by repeating the signal extraction multiple times, sampling $R_3$ from its measured distribution independently for each event, and then averaging the resulting posteriors. The systematic error of the relative luminosity induced a correction $\Delta R / R \sim 10^{-3}$ to the final asymmetry, a negligible contribution that will not be considered further.

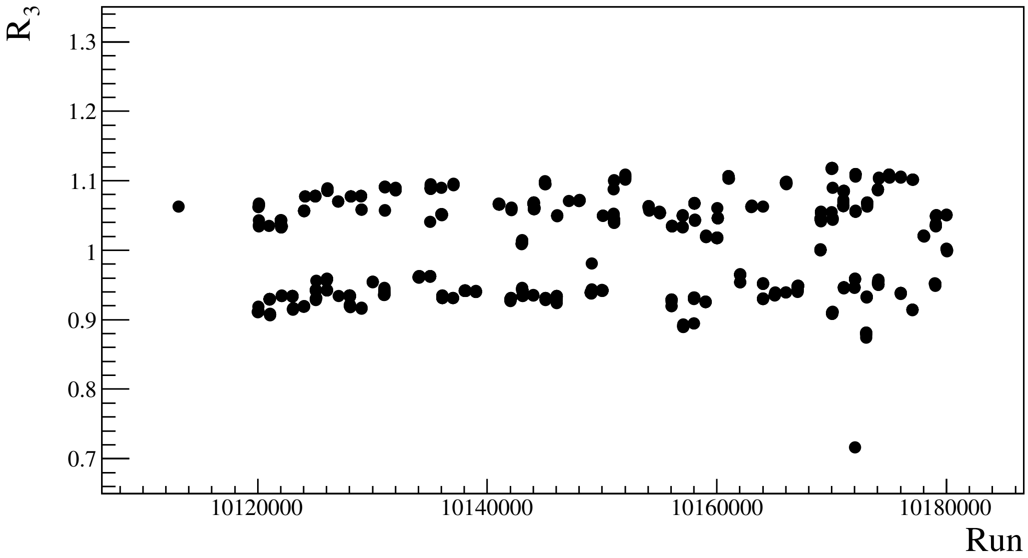


Figure 4-3: The relative luminosity $R_3$ was computed for each run within every fill. Note that the statistical uncertainties were of the order $10^{-3}$ and cannot be seen on the scale of the figure.

### 4.2.3 False Asymmetries

While not the only asymmetry one can measure with longitudinally-polarized beams at RHIC, the double-helicity asymmetry is the only one with a physical origin at $\sqrt{s} = 200\,\mathrm{GeV}$. The remaining false asymmetries arise only from systematic effects in the beam orientation or detector response and serve as critical checks on possible biases in the experiment, especially as the precision in the asymmetry improves.

These false asymmetries consider more sophisticated polarization configurations and, consequently, relative luminosities. The four initial state configurations,

$$h_B h_Y = \{++, +-, -+, --\}$$

admit six relative luminosities,

$$\begin{aligned}
R_1 &= \frac{\mathcal{L}^{++} + \mathcal{L}^{+-}}{\mathcal{L}^{-+} + \mathcal{L}^{--}} \\
R_2 &= \frac{\mathcal{L}^{++} + \mathcal{L}^{-+}}{\mathcal{L}^{+-} + \mathcal{L}^{--}} \\
R_3 &= \frac{\mathcal{L}^{-+} + \mathcal{L}^{+-}}{\mathcal{L}^{++} + \mathcal{L}^{--}} \\
R_4 &= \frac{\mathcal{L}^{++}}{\mathcal{L}^{--}} \\
R_5 &= \frac{\mathcal{L}^{+-}}{\mathcal{L}^{--}} \\
R_6 &= \frac{\mathcal{L}^{-+}}{\mathcal{L}^{--}}.
\end{aligned}$$

Note that, in this notation,

$$\begin{aligned}
\mathcal{L}^{\uparrow\uparrow} &= \mathcal{L}^{++} + \mathcal{L}^{--} \\
\mathcal{L}^{\uparrow\downarrow} &= \mathcal{L}^{+-} + \mathcal{L}^{-+}.
\end{aligned}$$

Aside from the double-helicity asymmetry there are four relevant false asymmetries, two single helicity and two double-helicity (Table 4-4). The construction of each false asymmetry proceeded as per the double-helicity asymmetry, taking effective polarizations and incorporating the relative luminosities as weights of the raw events (systematics were not considered). Once deconvolved, the means and uncorrelated variances of each false asymmetry posterior were integrated into a simple least squares fit to a constant (Figure 4-4). Each fit was consistent with zero, confirming the absence of any systematic bias contributing to the double-helicity asymmetry.

### 4.2.4 Systematic Errors

Aside from the polarization and disregarding relatively luminosity systematics, there is one additional source of systematic error unique to the asymmetry. To this point the

| **Asymmetry** | **Description** | **Definition** |
|---|---|---|
| $A_{\mathrm{Y}}$ | Yellow Beam Single Spin | $\frac{1}{P_Y}\frac{(\lambda^{++}+\lambda^{+-})-R_1(\lambda^{-+}+\lambda^{--})}{(\lambda^{++}+\lambda^{+-})+R_1(\lambda^{-+}+\lambda^{--})}$ |
| $A_{\mathrm{B}}$ | Blue Beam Single Spin | $\frac{1}{P_B}\frac{(\lambda^{++}+\lambda^{-+})-R_2(\lambda^{+-}+\lambda^{--})}{(\lambda^{++}+\lambda^{-+})+R_2(\lambda^{+-}+\lambda^{--})}$ |
| $A_{\mathrm{LS}}$ | Like-Sign | $\frac{1}{P_Y P_B}\frac{\lambda^{++}-R_4\lambda^{--}}{\lambda^{++}+R_4\lambda^{--}}$ |
| $A_{\mathrm{US}}$ | Unlike-Sign | $\frac{1}{P_Y P_B}\frac{R_6\lambda^{+-}-R_5\lambda^{-+}}{R_6\lambda^{+-}+R_5\lambda^{-+}}$ |

Table 4.4: Four false asymmetries can arise from asymmetries in the experimental setup, but not in QCD.

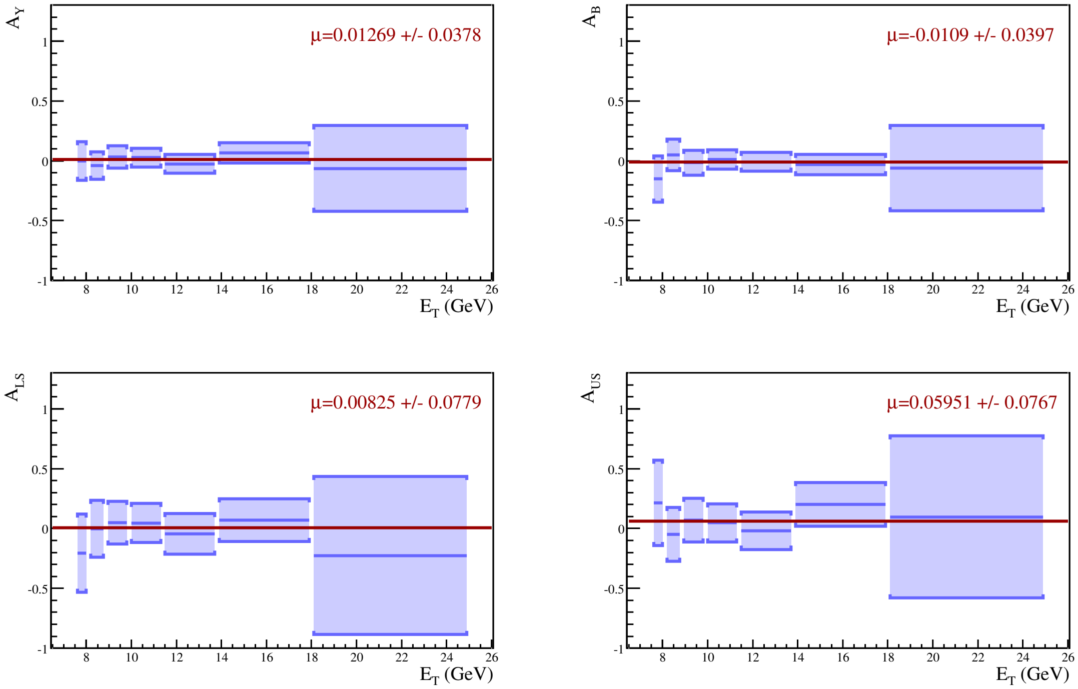


Figure 4-4: Constant fits to the false asymmetries were all consistent with zero, showing no evidence of any systematic bias in the experimental configuration: clockwise from the top-left, (a) $A_Y$, (b) $A_B$, (c) $A_{LS}$, and (d) $A_{UL}$.

beams have been assumed to be polarized exactly longitudinally, but imperfections in the spin flippers may admit a small transverse component to the beams which itself induces false asymmetries. With no careful study of this effect in 2009, the systematic error from non-longitudinal beam components was conservatively approximated as $\delta_{\mathrm{NL}} = 0.02 \cdot A_{\mathrm{LL}}$.

Once again, most of the systematic errors were combined in quadrature. The correlated polarization systematic was excluded from the total and quoted separately.

### 4.2.5 Measurement

A posterior for the double-helicity asymmetry was constructed from the extracted, deconvolved spectra $\lambda_i^{\upharpoonleft\upharpoonright}$ and $\left(R_3 \lambda^{\upharpoonleft\downharpoonright}\right)_i$ and the effective polarization $\langle P_Y P_B \rangle$. The transformation of the posterior distribution itself was computed with Monte Carlo methods, assuming that the posterior was adequately described by a multivariate gaussian (Table 4.5).

After the considerable effort to isolate the direct photons, they finally offered their first glimpse of the gluon polarization (Figure 4-5). The measured asymmetry is plotted along with four leading-order models of the gluon spin [4, 5], a comparison which unfortunately highlights the fundamental limitation of direct photons as a tool for studying the proton: statistics. Indeed, with only 13.5 $\mathrm{pb}^{-1}$, direct photon measurements are not expected to be competitive to the ample hadronic asymmetries.

The comparison, however, also shows the promise of the measurement. Already the direct photon asymmetry is strongly inconsistent with large gluon polarization, and the incorporation of more data along with improvements in beam polarization promises significantly more discrimination amongst the other models before being limited by the systematic errors.

| $E_T$ (GeV) | Mean | Asymmetry Covariance $\times 10^3$ | | | | | | | Systematic Errors | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 7.5 - 8.1 | 8.1 - 8.85 | 8.85 - 9.9 | 9.9 - 11.4 | 11.4 - 13.8 | 13.8 - 18 | 18 - 25 | Signal Extraction | BEMC Calibration | Non-Longitudinal Beam Component | Relative Luminosity | Polarization |
| 7.5 - 8.1 | -0.048 | 87.098 | -6.987 | -11.284 | 3.087 | -1.712 | -3.266 | 4.033 | +0<br>-0.02 | +0.015<br>-0.016 | $\pm 11.75 \times 10^{-4}$ | $\pm 15.41 \times 10^{-4}$ | ± 8.8% |
| 8.1 - 8.85 | 0.021 | -6.987 | 39.34 | -18.113 | 2.201 | -0.458 | -0.407 | 1.524 | +0.018<br>-0 | +0<br>-0.028 | $\pm 5.48 \times 10^{-4}$ | | |
| 8.85 - 9.9 | 0.032 | -11.284 | -18.113 | 26.81 | -8.968 | 0.525 | 0.413 | -3.773 | +0<br>-0.009 | +0.02<br>-0.002 | $\pm 8.16 \times 10^{-4}$ | | |
| 9.9 - 11.4 | -0.004 | 3.087 | 2.201 | -8.968 | 19.764 | -5.801 | -0.066 | 0.751 | +0.011<br>-0.007 | +0.026<br>-0.024 | $\pm 0.89 \times 10^{-4}$ | | |
| 11.4 - 13.8 | -0.117 | -1.712 | -0.458 | 0.525 | -5.801 | 19.138 | -4.282 | -0.308 | +0.01<br>-0.017 | +0.058<br>-0.096 | $\pm 29.05 \times 10^{-4}$ | | |
| 13.8 - 18 | -0.055 | -3.266 | -0.407 | 0.413 | -0.066 | -4.282 | 21.817 | -14.961 | +0.007<br>-0 | +0.018<br>-0.039 | $\pm 13.54 \times 10^{-4}$ | | |
| 18 - 25 | -0.123 | 4.033 | 1.524 | -3.773 | 0.751 | -0.308 | -14.961 | 384.894 | +0.034<br>-0 | +0<br>-0.042 | $\pm 30.51 \times 10^{-4}$ | | |

Table 4.5: As with the unpolarized cross section, the asymmetry was approximated as a multivariate gaussian while systematic errors were either perfectly uncorrelated or correlated. Note that the elements of the covariance matrix are scaled by $10^3$ for display.

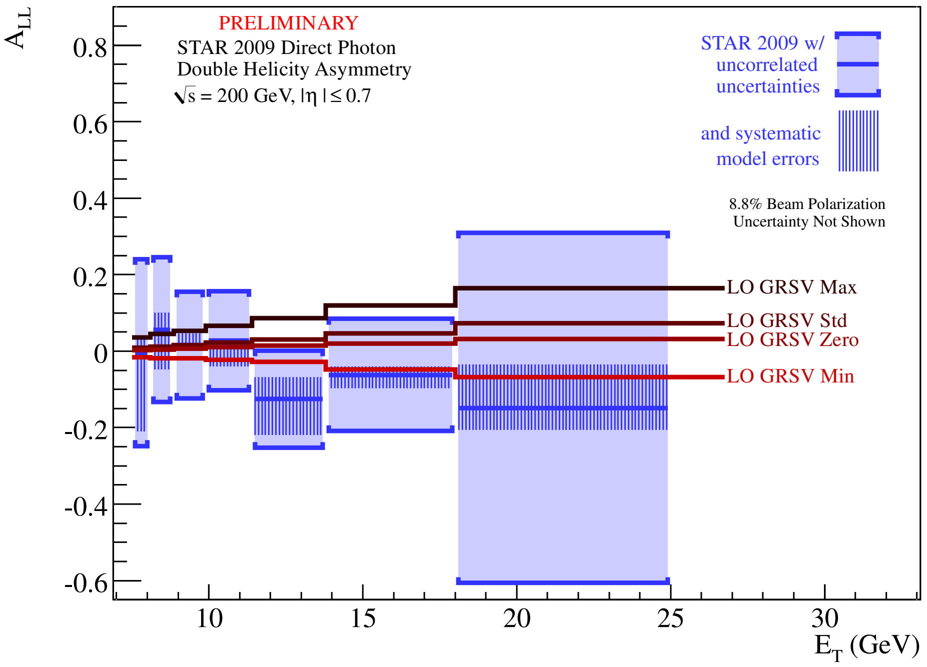


Figure 4-5: The direct photon double-helicity asymmetry only mildly discriminates between different models of the gluon polarization due to the limited statistics of the measurement.

# Chapter 5

# Conclusion

Ultimately limited by the accumulated luminosity at STAR, the direct photon asymmetry cannot offer any novel constraints of the gluon polarization within the proton. The weak constraints it does provide, however, are consistent with hadronic measurements of the gluon polarization, at least across the partonic phase space to which the experiment is sensitive. The Run 9 hadronic asymmetries [80, 81] provide the highest precision constraints to date and suggest that the gluon polarization may be as large as that of the quarks, although still not large enough to account for all of the proton spin [82].

Utilizing higher-energy beams and detectors at larger rapidity, future measurements at RHIC will expand the constraints to smaller $x$, with a future electron-ion collider offering the hope of studying the asymptotic behavior of the gluon polarization as $x$ approaches zero [83]. Given conservative assumptions on the asymptotic behavior of the polarization, a large gluon polarization secluded in these unexplored regions would be unexpected, and the spin structure of the proton remains as puzzling as ever.

Angular momenta of the QCD fields offer the obvious resolution, but the viability of constraining these contributions in deep inelastic scattering is questionable. The problem is that the light-cone dominance that simplifies deep inelastic scattering so greatly also dramatically suppresses the angular momenta operators and hence their influence on any observables,

$$\lim_{x^+ \to 0} \left( \vec{x} \times i\mathcal{D} \right)^- = 0.$$

By construction, deeply virtual compton scattering is sensitive to physics away from the light-cone and, consequently, the angular momenta [84]. Away from the light-cone, however, the *generalized parton distribution functions* constrained by these measurements must be defined in terms of fields quantized in $A_0 = 0$ gauge and not the light-cone quanta specifying the canonical understanding of partons. Whether these deeply virtual measurements can be reconciled with those from the deep inelastic scattering regime remains a mystery.

Thus the spin puzzle endures, furnishing more questions than answers and maturing our understanding of QCD and the observable universe.

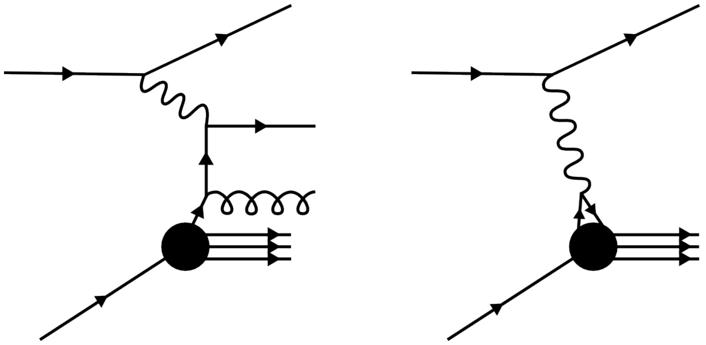
# Appendix A

# Twist in QCD

Due to the tremendous complexity of QCD, factorization has been explicitly proved for only a few interactions, and even then the proofs apply only to the leading twist contributions in the operator product expansion (Fig A-1). Distinguishing between contributions from beyond-leading-order in the perturbative expansion and higher twist is elucidated by considering the operator structure of the theory [13]. Quark fields $\psi$, gluon fields $F^{\mu\nu}$, and covariant derivatives $D^{\mu}$ all contribute to the twist of the final observable (Table A.1) and the twist of a given diagram is readily calculated by simply counting these fields (Figure A-2).

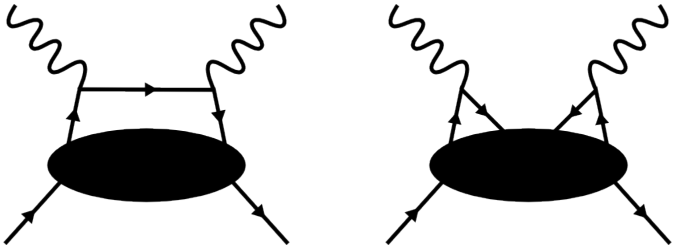
| | $\boldsymbol{\psi}$ | $\mathbf{F}^{\boldsymbol{\mu\nu}}$ | $\mathbf{D}^{\boldsymbol{\mu}}$ |
|---|---|---|---|
| **d** | 3/2 | 2 | 1 |
| **s** | 1/2 | 1 | 1 |
| **t** | 1 | 1 | 0 |

Table A.1: Each fundamental operator in QCD contributes to the twist of the forward Compton amplitude based on its operator structure.



(a) (b)

Figure A-1: Factorization theorems typically hold at all orders in the perturbative expansion of the hard interaction, but only at leading twist where the proton is probed locally. This includes higher-order, leading-twist contributions such as (a) but not higher twist contributions such as (b).



(a) (b)

Figure A-2: The twist of any contribution to the hadronic tensor is given by the number of the fields interacting with the soft structure of the proton. One leading twist 2 contribution is given by two quarks fields (a) while a higher twist 4 contribution features four quark fields.

# Appendix B

# Discriminant Distributions

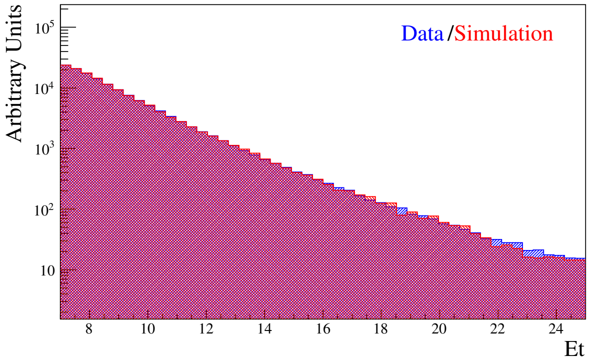
Because the signal extraction and efficiency corrections were dependent on simulation, agreement between the data and the simulation is critical to the success of the analysis. This appendix considers the comparison of the measured and simulated distributions for each discriminant included in the analysis, as well as the signal and background distributions within simulation demonstrating each variable's sensitivity to direct photons.

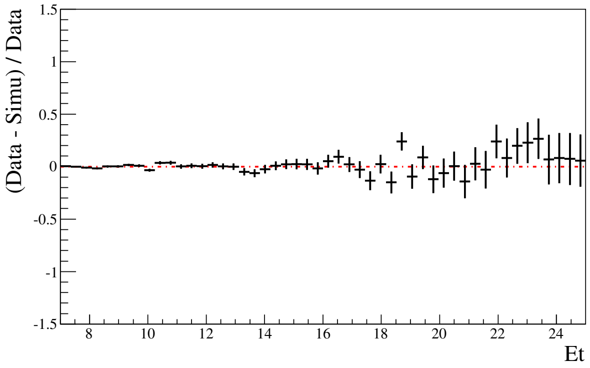
The simulation was first weighted to ensure constant integrated luminosity across the entirely of phase space before a vertex dependent correction was applied as detailed in Chapter 3. For the comparison an additional global weight was introduced to ensure the same number of effective events in both the data and simulation samples.

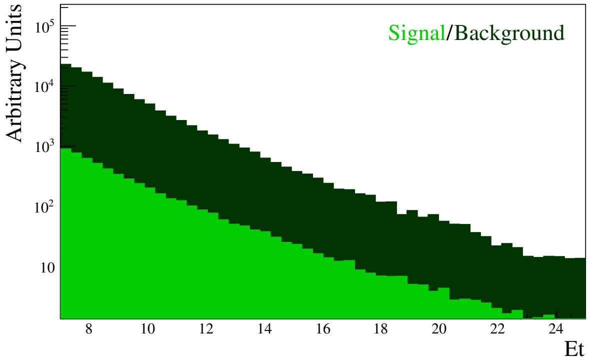
Note that the transverse energy dependence of the direct photon cross section within Pythia was known to be incorrect, in particular the cross section was significantly underestimated at higher energies, and some discrepancy is expected in regions of the distributions where the signal contribution is non-trivial. This did not affect the analysis, however, which was carefully designed to be independent of the signal to background ratios defined in the simulation.

# Et

Transverse energy (GeV) of the candidate relative to the primary vertex.

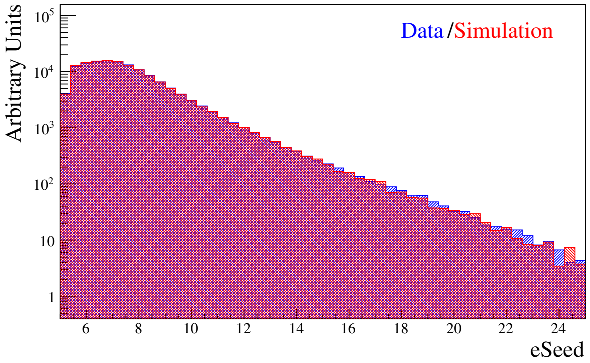


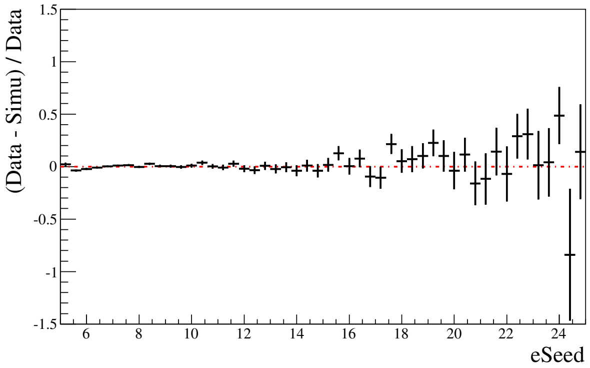


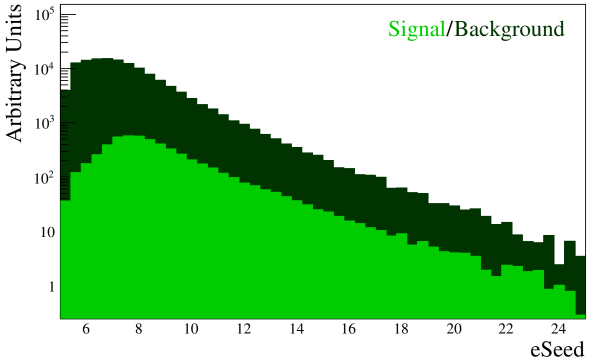


# Eseed

Energy (GeV) of the cluster seed tower.

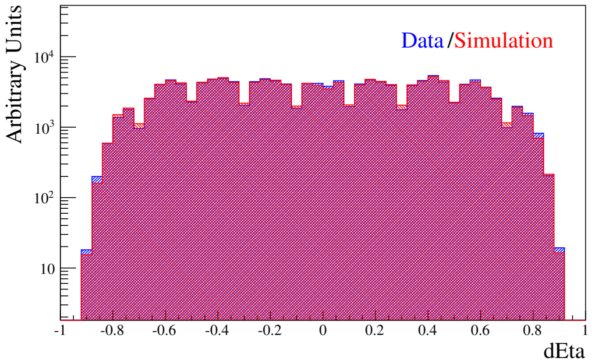


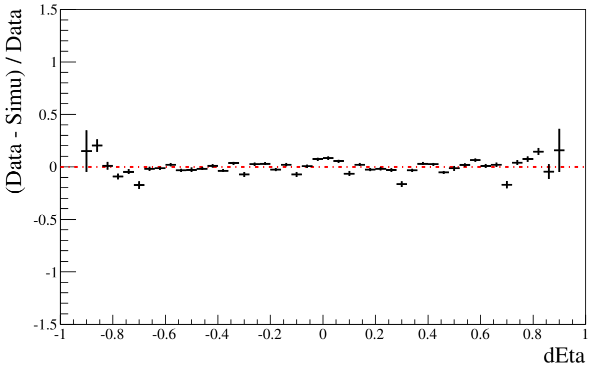


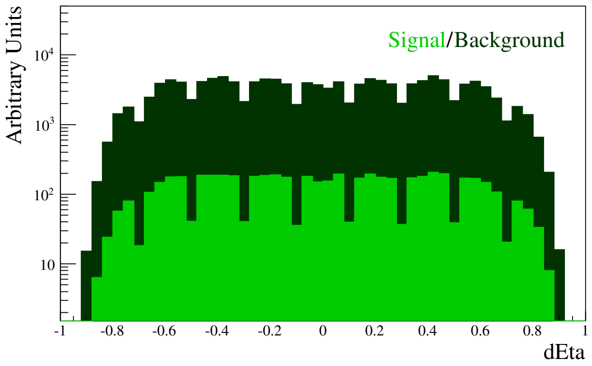


# dEta

Pseudorapidity of the cluster relative to the detector.

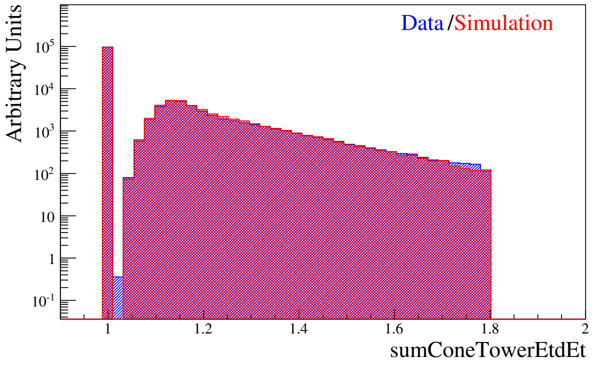


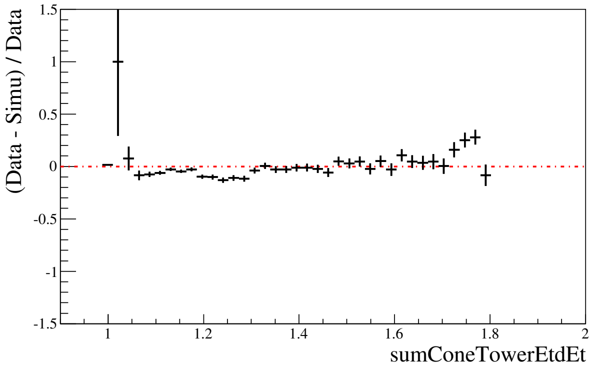


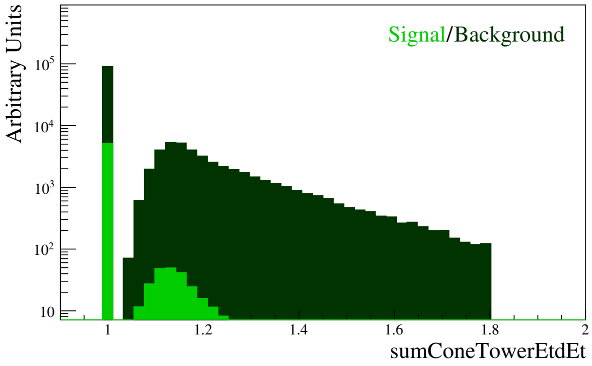


# sumConeTowerEtdEt

Summed transverse energy of all towers within a cone of $\Delta R = 1.0$ around the candidate, normalized to the summed transverse energy of all towers in the candidate itself.

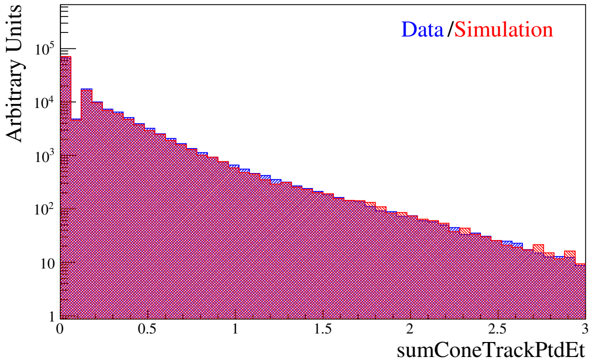


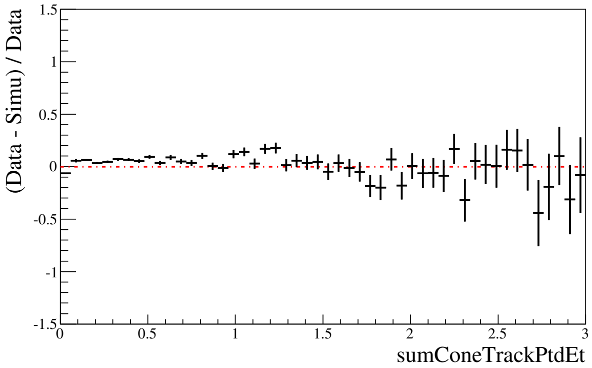


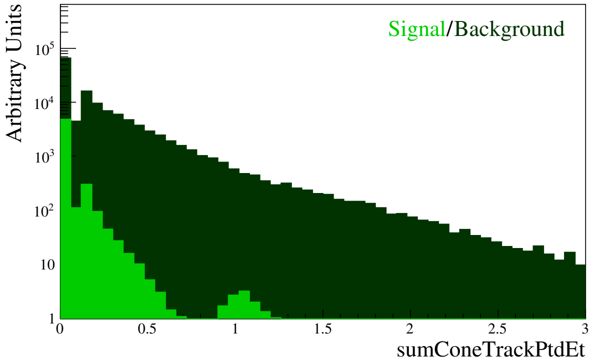


# sumConeTrackPtdEt

Summed transverse momentum of all tracks within a cone of $\Delta R = 1.0$ around the candidate, normalized by the summed transverse energy of all towers in the candidate.

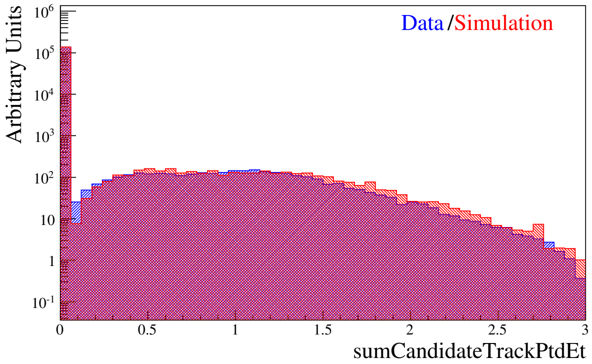


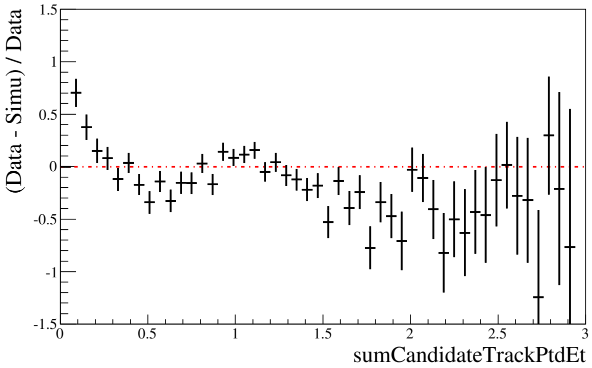


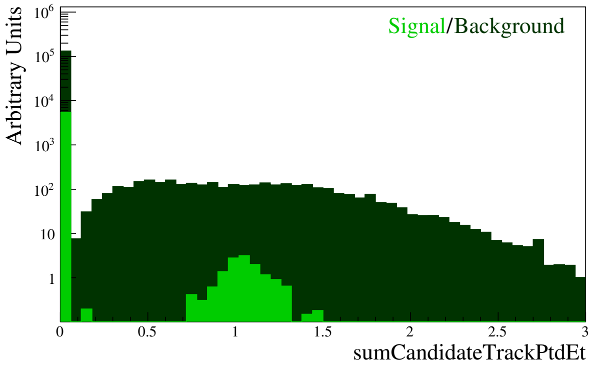


# sumCandidateTrackPtdEt

Summed transverse momentum of all tracks projecting to the candidate, normalized by the summed transverse energy of all towers in the candidate.

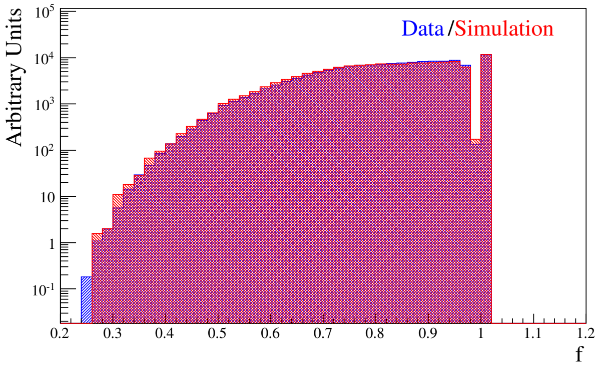


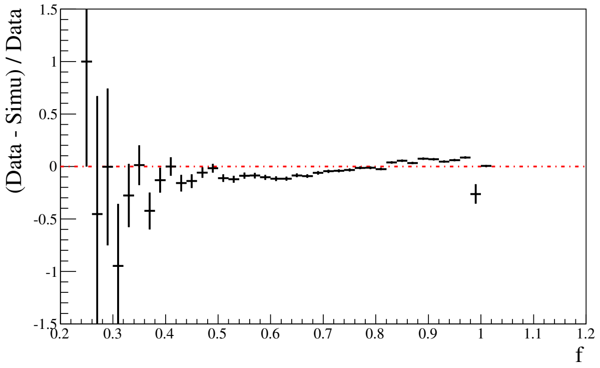


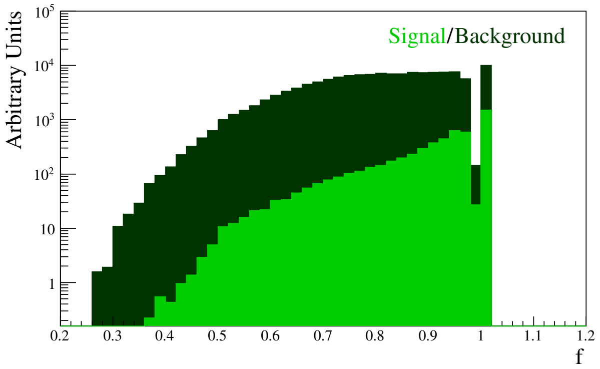


# f

Seed energy normalized to the summed transverse energy of all towers above 250 MeV in the candidate.

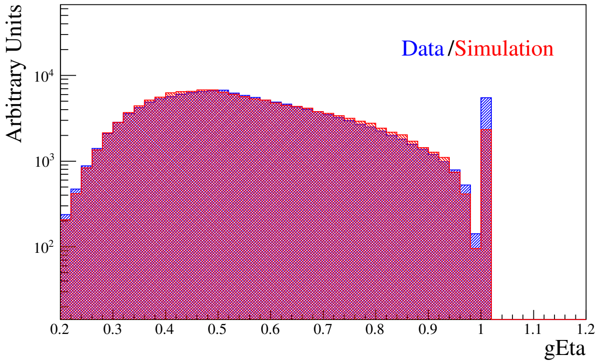


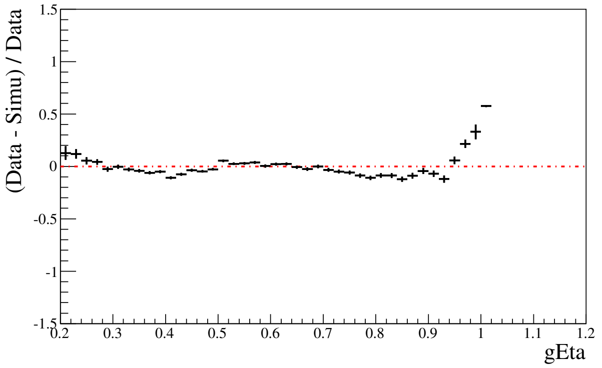


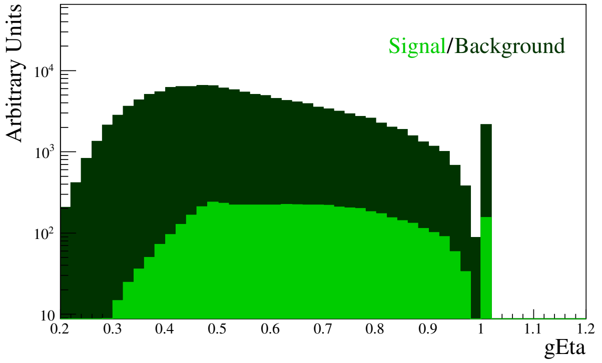


# gEta

Maximum strip energy normalized to the summed energy of all candidate strips in the $\eta$ plane of the BSMD.

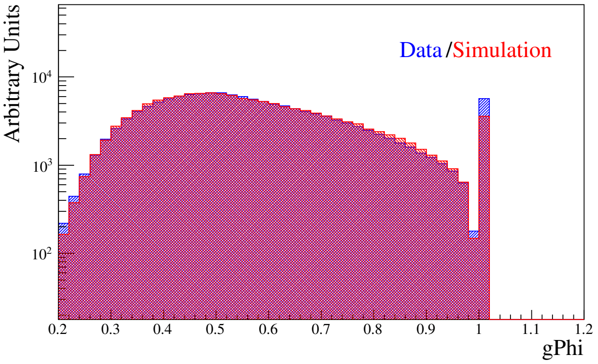


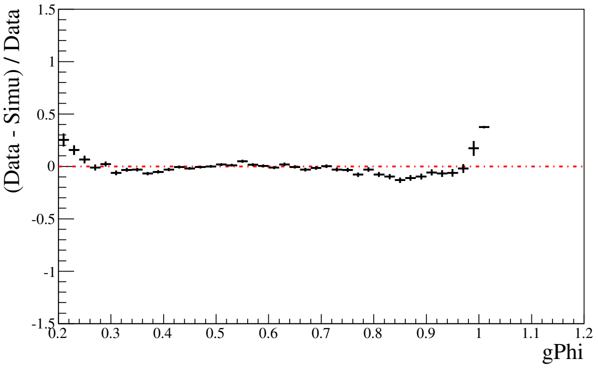


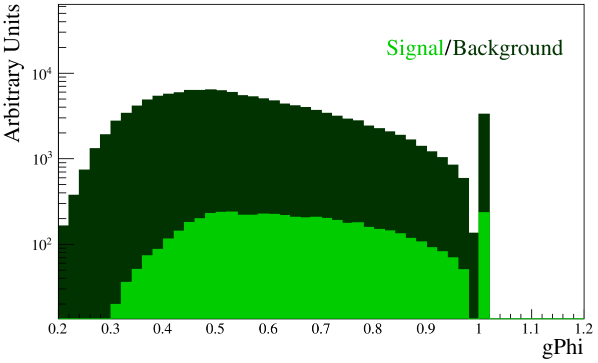


# gPhi

Maximum strip energy normalized to the summed energy of all candidate strips in the $\phi$ plane of the BSMD.

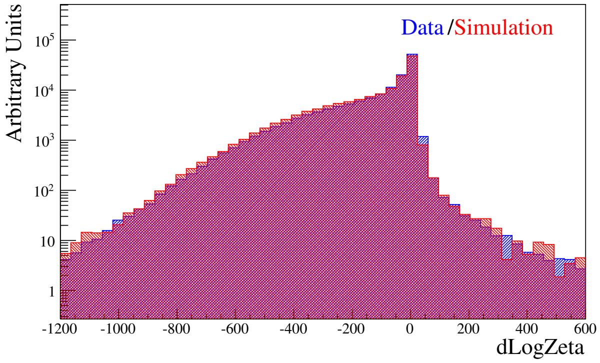


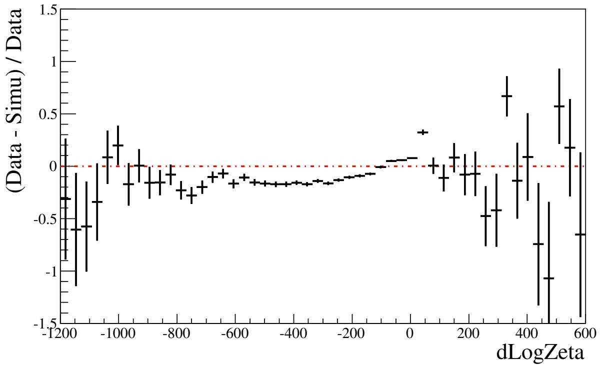


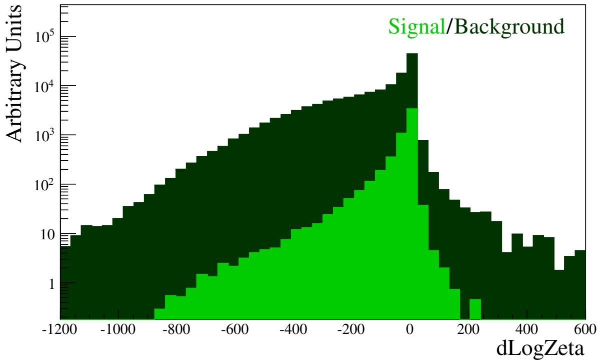


# dLogZeta

The Bayesian log-odds between a single shower hypothesis and a double shower hypothesis in the $\eta$ plane of the BSMD. The single shower model is preferred for dLogZeta $> 0$.

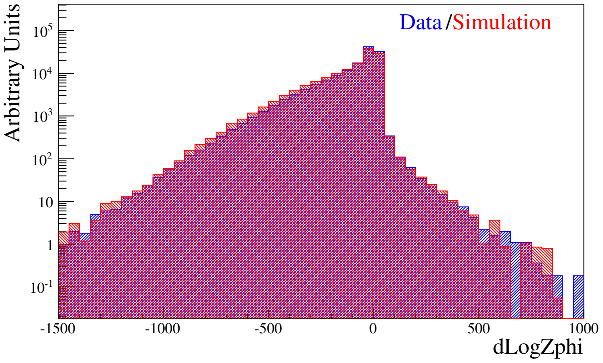


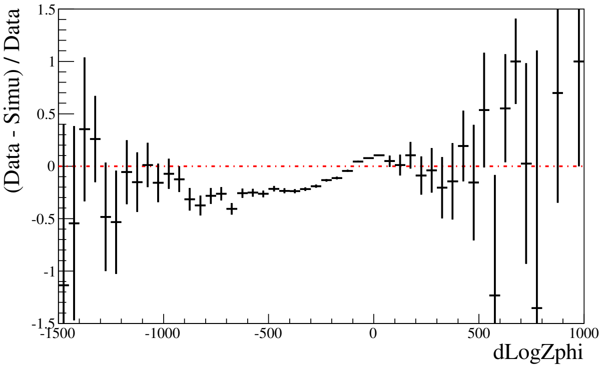


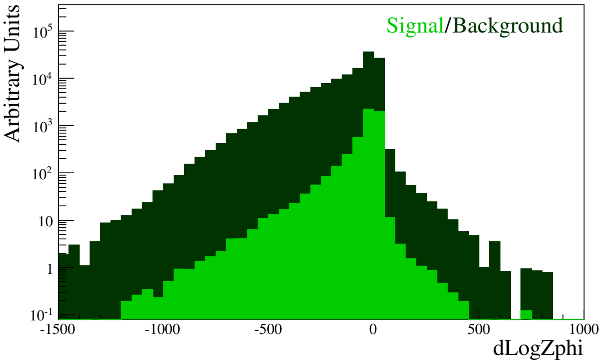


# dLogZphi

The Bayesian log-odds between a single shower hypothesis and a double shower hypothesis in the $\phi$ plane of the BSMD. The single shower model is preferred for dLogZphi $> 0$.







# Appendix C

# Inference

Experimental physics requires not only an apparatus to collect data, but also a means to analyze that data. Such inference involves the construction of models to describe how the data are generated from the underlying physics, and the the constraint of those models given the data at hand.

Classical methods of inference developed at the beginning of the twentieth century have achieved orthodoxy in many fields, including much of physics, but by construction they cannot answer many of the questions arising in the typical analysis. Developed long before the advent of "classical statistics" but becoming popular only in recent decades, Bayesian methods offer a consistent framework for inference where these queries have natural solutions.

The statistics and inference literature is vast and often intimidating to the physicist, but a few references are particularly welcoming. Bishop [66], MacKay [68], and Sivia [85] provide clear, pedagogical introductions to Bayesian inference with particular focus on implementations and other practical considerations. More intense works, Bernardo [86] and Jaynes [87] consider the matters of mathematical and philosophical rigor relevant not so much to immediate application but fundamental to understanding of the ultimate power and limitations of inference.

## C.1 Classical Statistics

Classically, probabilities are defined as long-term frequencies: the probability of an event $i$ is the total number of trials where the event is realized normalized to the total number of trials as that total approaches infinity.

$$p_i = \lim_{n \to \infty} \frac{n_i}{n}.$$

Canonical examples include the probabilities of rolling each face on a die or drawing a particular card from a deck.

From this perspective, probabilities are defined over ensembles of data sets, where any ensemble is specified by defining the probabilities of every possible realization, $\mathcal{D}$, of the data dependent on some parameters, $\alpha$, with a sampling distribution, $p(\mathcal{D}|\alpha)$. Given a sampling distribution, the parameters $\alpha$ can be inferred from any measurement by constructed functions of the data that, on average, estimate the true value.

The true value is estimated directly, for example, with an *estimator*: any function of the data $\hat{\alpha}(\mathcal{D})$ such that

$$\int \mathrm{d}\mathcal{D} \, \hat{\alpha}(\mathcal{D}) \, p(\mathcal{D}|\alpha) \approx \alpha.$$

Alternatively, the true value can be bracketed with two functions $\hat{\alpha}_1(\mathcal{D})$ and $\hat{\alpha}_2(\mathcal{D})$, known as a *confidence interval*, such that

$$p_{\mathcal{D}}(\hat{\alpha}_1(\mathcal{D}) \leq \alpha \leq \hat{\alpha}_2(\mathcal{D})) = \beta,$$

for some predefined probability $\beta$.

Ultimately these methods are defeated by their own construction. In practice only a single realization of the data is available and, unless the ensemble is nearly degenerate and all realizations are almost indistinguishable, inferences relying on the entire ensemble can be misleading. Although the realizations typically become degenerate as the size of the data set increases, this asymptotic behavior depends on the complexity of the sampling distribution. Determining when a data set is large enough for these issues to be inconsequential is a notoriously difficult problem.

Even if the ensemble is known to be degenerate, the ensemble assumption itself restricts the scope of possible inferences to within the a single sampling distribution. Two different models, for example, cannot be directly compared to determine which best fits the measured data. Classical statistics can only judge the consistency of the data within each model individually. In addition, a sampling distribution with parameters irrelevant to the physics, also known as nuisance parameters, admits inference only on the whole of the parameters at once. No consistent means of reducing the inference to only the relevant parameters exists in the classical framework.

## C.2 Bayesian Inference

While the classical approach is ultimately limited by its dependence on the ensemble assumption and hence the interpretation of probabilities as frequencies, the scope of probability is far more general than mere frequencies. As first shown by Cox [88], and later refined by Knuth and Skilling [89], any strictly ordered algebra respecting associativity and independence induces a probability calculus with measures. In other words, any consistent degree of belief must be described by the mathematics of probability theory.

With this more general interpretation the parameters themselves can be treated probabilistically, and the sampling distribution of classical statistics induces a distribution over the parameters through Bayes' Theorem,

$$p(\alpha|\mathcal{D},H)=\frac{p(\mathcal{D}|\alpha,H)\,p(\alpha|H)}{p(\mathcal{D}|H)}$$

where $H$ denotes any implicit assumptions, such as the explicit form of each term in the construction.

Encoding all available information about the parameters, the posterior $p(\alpha|\mathcal{D}, H)$ is comprised of three terms: the sampling distribution, or likelihood $p(\mathcal{D}|\alpha, H)$, defines how the data is generated from the underlying parameters; the prior $p(\alpha|H)$ encapsulates all information about the parameters available before the data is realized[1]; and finally the evidence

$$p(\mathcal{D}|H) = \int \mathrm{d}\alpha \, p(\mathcal{D}|\alpha, H) \, p(\alpha|H)$$

ensures that the posterior is a properly normalized probability distribution.

In this more general framework, inference is conditional only on the data and the assumptions $H$ implicit in the assembly of the posterior[2], not on any awkward construction such as the existence of ensembles. This comprehensive approach, however, admits more than just inference of the parameters in a static model. Not only do the rules of probability allow the consistent mapping of posteriors as the underlying model is transformed, they can also be applied to the models themselves to define model comparison.

Although elegant and powerful, Bayesian inference is hampered by computational limitations. The advent of techniques such as Markov chain Monte Carlo, however, has triggered dramatic growth in the practical potential of these methods.

### C.2.1 Marginalization

Unlike the estimators of classical statistics, posteriors can be transformed without the loss of information. Given a posterior for the parameters $\alpha$, for example, the posterior for the transformed space $\gamma(\alpha)$ is immediate,

---

[1] Note that if the prior is uniform then the posterior becomes proportional to the sampling distribution and many of the classical estimators can be interpreted as point estimates of the Bayesian posterior. This implicit uniform prior can be inappropriate in many cases, but not as limiting as the loss of information from replacing a full distribution with a point estimate.

[2] Similar assumptions are implicit in the choice of sampling distribution in the classical analysis as well, but rarely considered.

$$p(\gamma|\mathcal{D}, H) = p(\alpha(\gamma)\,|\mathcal{D}, H) \left| \frac{\partial \alpha}{\partial \gamma} \right| .$$

Undesired parameters, such as the nuisance parameters introduced above, can simply be integrated out, or *marginalized*,

$$\begin{aligned} p(\alpha_1|\mathcal{D}, H) &= \int \mathrm{d}\alpha_2 \, p(\alpha_1, \alpha_2|\mathcal{D}, H) \\ &\propto \int \mathrm{d}\alpha_2 \, p(\alpha_1|\alpha_2, \mathcal{D}, H) \, p(\mathcal{D}|\alpha_2, H) \, p(\alpha_2, H) \, . \end{aligned}$$

Note the implicit presence of the evidence in the integral; marginalization incorporates not just the prior uncertainty of the nuisance parameters, $p(\alpha_2, H)$, but also the dependence of the nuisance parameters on the data, $p(\mathcal{D}|\alpha_2, H)$. When this dependence is non-trivial, the data inform the nuisance parameters, reducing their uncertainty and, consequently, improving the marginalized posterior.

### C.2.2 Model Comparison

Consider a space consisting of models defined by their assumptions, $H_i$. No different from the inference of the parameters within a given model, inference on the model space is given by Bayes' Theorem:

$$p(H_i|\mathcal{D}) = \frac{p(\mathcal{D}|H_i) \, p(H_i)}{\sum_j p(\mathcal{D}|H_j) \, p(H_j)} .$$

Note that the likelihood of a model is simply its evidence,

$$p(\mathcal{D}|H_i) = \int \mathrm{d}\alpha \, p(\mathcal{D}|\alpha, H) \, p(\alpha, H) \, ,$$

so that model inference immediately follows once a model prior has been assigned and the individual models have been fully analyzed.

The comparison of two models to data, for example, condenses into the ratio of model posteriors, or odds,

$$\begin{aligned} O_{12}(\mathcal{D}) &= \frac{p(\mathcal{D}|H_1)}{p(\mathcal{D}|H_2)} \\ &= \frac{p(\mathcal{D}|H_1)\,p(H_1)}{\sum_j p(\mathcal{D}|H_j)\,p(H_j)} \frac{\sum_j p(\mathcal{D}|H_j)\,p(H_j)}{p(\mathcal{D}|H_2)\,p(H_2)} \\ &= \frac{p(\mathcal{D}|H_1)\,p(H_1)}{p(\mathcal{D}|H_2)\,p(H_2)} \end{aligned}$$

where the first model provides a better fit to the data if $O_{12}(\mathcal{D}) > 1$ and the second model proves superior if $O_{12}(\mathcal{D}) < 1$.

### C.2.3 Markov Chain Monte Carlo

Although Bayes' Theorem gives an explicit form for the posterior, Bayesian inference is often difficult to implement in practice because of the need to integrate over the parameter space. For example, even if the normalization isn't relevant and the evidence can be ignored, characterizing the posterior with moments or quartiles still requires integrals, as does marginalization and model comparison. Analytic integrals are rare and largely confined to simple problems, and standard numerical integration methods are infeasible when considering more than a few parameters.

Monte Carlo integration, however, is a general method that does not scale poorly with the dimensionality of the parameter space. Here expectations are approximated by summing the desired function over a finite set of samples drawn from the posterior,

$$\int \mathrm{d}\alpha f(\alpha)\,p(\alpha|\mathcal{D}, H) \approx \frac{1}{n} \sum_i f(\alpha_i)\,,\ \alpha_i \sim p(\alpha|\mathcal{D}, H)\,.$$

As the number of independent samples $n$ increases, the error of the approximation converges to zero independently of the dimension of the parameter space, ensuring the practicality of the method even on large spaces.

| | **Nominal** | **Transformed** | **Marginalized** |
|---|---|---|---|
| **Distribution** | $p(\alpha, \beta)$ | $p(\gamma(\alpha, \beta), \delta(\alpha, \beta)) \left\| \frac{\partial(\gamma, \delta)}{\partial(\alpha, \beta)} \right\|$ | $\int \mathrm{d}\beta \, p(\alpha, \beta)$ |
| **Samples** | $\{\alpha_i, \beta_i\}$ | $\{\gamma(\alpha_i, \beta_i), \delta(\alpha_i, \beta_i)\}$ | $\{\alpha_i\}$ |

Table C.1: Transformations and marginalization are trivial to implement once a distribution has been represented by samples.

A general approach to Bayesian inference is then to consider not the posterior explicitly, but rather a set of samples, $\{\alpha_i\}$ drawn from it. Not only can any expectation be immediately evaluated[3], transformations and marginalization become trivial (Table C.1).

Because they require global knowledge of the distribution, independent samples cannot be drawn from complicated posteriors without unrealistic resources. Markov chain Monte Carlo offers a compromise, generating dependent samples from the posterior using only local properties of the distribution and hence a far smaller computational burden. Once the dependence of the samples is understood, Markov chain Monte Carlo becomes a powerful tool for building and learning the complex models of pioneering scientific analyses [90].

---

[3] The one exception is the evidence integral, whose Monte Carlo variance diverges. These integrals can be calculated, however, with a modified Monte Carlo technique known as nested sampling [64].

# Appendix D

# Gaussian Processes

Many common problems in data analysis, from multivariate regression to machine learning classification, can be reduced to the estimation of a one-dimensional function $y(\mathbf{x})$ subject to gaussian noise,

$$t = y(\mathbf{x}) + \epsilon$$

with

$$\epsilon \sim \mathcal{N}\big(0, \beta^{-1}\big) \, ,$$

given only a set of measurements $\{t_i, \mathbf{x}_i\}$. A Bayesian treatment assuming a linear function $y(\mathbf{x})$ is straightforward and immediately admits a non-linear, non-parameteric extension. Based on kernel methods, this extension offers a powerful, comprehensive approach to the estimation problem with all of the benefits of consistent probabilistic interpretation.

The discussion here follows from the excellent treatments in Bishop [66] and Rasmussen [67]. A more cursory but less imposing discussion can be found in MacKay [68].

## D.1 Linear Bayesian Regression

The simplest approach to the estimation problem is to model the function as a linear combination of the input variables,

$$y = \mathbf{w}^T \mathbf{x}.$$

Assuming a gaussian prior,

$$p(\mathbf{w}) = \mathcal{N}\left(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbb{I}\right),$$

the posterior distribution for the weights given a measurement becomes

$$\begin{aligned} p(\mathbf{w}|\mathbf{t}) &= \frac{p(\mathbf{t}|\mathbf{w})\, p(\mathbf{w})}{p(\mathbf{t})} \\ &= \left[\prod_{i=1}^{N} \mathcal{N}\left(t_n|\mathbf{w}^T\mathbf{x}_n, \beta^{-1}\right)\right] \mathcal{N}\left(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbb{I}\right) \\ &= \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N), \end{aligned}$$

where

$$\mathbf{m}_N = \beta \mathbf{S}_N \mathbf{X}^T \mathbf{t},$$

$$\mathbf{S}_N^{-1} = \alpha\mathbb{I} + \beta \mathbf{X}^T \mathbf{X},$$

and

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_N^T \end{pmatrix}.$$

The distribution for a new point $t$ marginalized over the weights is then

$$\begin{aligned} p(t|\mathbf{t}) &= \int d\mathbf{w}\, p(t|\mathbf{w})\, p(\mathbf{w}|\mathbf{t}) \\ &= \int d\mathbf{w}\, \mathcal{N}\big(t|\mathbf{w}^T\mathbf{x}, \beta^{-1}\big)\, \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N) \\ &= \mathcal{N}\big(t|\mathbf{m}_N^T\mathbf{x}, \sigma_N^2(\mathbf{x})\big)\,, \end{aligned}$$

where

$$\sigma_N^2(\mathbf{x}) = \beta^{-1} + \mathbf{x}^T\mathbf{S}\mathbf{x}.$$

Note that the mean of the predictive distribution is a linear combination of the measured function values,

$$\begin{aligned} \langle t \rangle &= \mathbf{m}_N^T\mathbf{x} \\ &= \mathbf{x}^T\mathbf{m}_N \\ &= \mathbf{x}^T\beta\mathbf{S}_N\mathbf{X}\mathbf{t} \\ &= \sum_{i=1}^{N} \big(\beta\mathbf{x}^T\mathbf{S}_N\mathbf{x}_n\big)\, t_n. \end{aligned}$$

Defining the equivalent kernel

$$k(\mathbf{x}, \mathbf{y}) = \beta\mathbf{x}^T\mathbf{S}_N\mathbf{y},$$

the predictive mean becomes

$$\langle t \rangle \equiv \sum_{i=1}^{N} k(\mathbf{x}, \mathbf{x}_n)\, t_n.$$

Moreover, the variance is also specified in terms of $k(\mathbf{x}, \mathbf{y})$,

$$\sigma_N^2(\mathbf{x}) = \beta^{-1}\left[1 + k(\mathbf{x}, \mathbf{x})\right],$$

as can the covariance between any two of the measured points

$$\mathrm{cov}[t(\mathbf{x})\,,t(\mathbf{y})] = \beta^{-1}k(\mathbf{x},\mathbf{y})\,.$$

The equivalent kernel $k(\mathbf{x},\mathbf{y})$ completely characterizes the predictive distribution. If the kernel of an unknown point $\mathbf{x}$ and a known point $\mathbf{x}_n$ is large, then the two points are highly correlated and the latter will contribute strongly to the predicted mean, hence the equivalent kernel defines an effective geometry in the target space.

## D.2 Nonlinear Bayesian Regression

The linear model is immediately extended by introducing basis functions that map the linear feature space into a nonlinear one,

$$y = \mathbf{w}^T\boldsymbol{\phi}(\mathbf{x})\,.$$

Note that no orthogonality or completeness requirement is made of the basis functions, and the dimension of basis functions does not have to match that of the input vector $\mathbf{x}$.

Evaluation of the nonlinear model parallels the linear model. The posterior distribution for the weights becomes

$$p(\mathbf{w}|\mathbf{t}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_N,\mathbf{S}_N)\,,$$

where

$$\mathbf{m}_N = \beta\mathbf{S}_N\boldsymbol{\Phi}^T\mathbf{t}$$

and

$$\mathbf{S}_N^{-1} = \alpha\mathbb{I} + \beta\boldsymbol{\Phi}^T\boldsymbol{\Phi}.$$

The design matrix $\mathbf{\Phi}$ is defined as

$$\mathbf{\Phi} = \begin{pmatrix} \boldsymbol{\phi}^T(\mathbf{x}_1) \\ \vdots \\ \boldsymbol{\phi}^T(\mathbf{x}_n) \end{pmatrix}.$$

Once again marginalizing over the weights, the predictive distribution is given by a gaussian

$$p(t|\mathbf{t}) = \mathcal{N}\big(t|\mu_N(\mathbf{x})\,, \sigma_N^2(\mathbf{x})\big)$$

with mean

$$\begin{aligned} \mu_N(\mathbf{x}) &= \sum_{i=1}^{N} \beta \boldsymbol{\phi}^T(\mathbf{x})\, \mathbf{S}_N \boldsymbol{\phi}(\mathbf{x}_n)\, t_n \\ &= \sum_{i=1}^{N} k(\mathbf{x}, \mathbf{x}_n)\, t_n \end{aligned}$$

and variance

$$\begin{aligned} \sigma_N^2(\mathbf{x}) &= \beta^{-1} + \boldsymbol{\phi}^T(\mathbf{x})\, \mathbf{S}_N \boldsymbol{\phi}(\mathbf{x}) \\ &= \beta^{-1}\left[1 + k(\mathbf{x}, \mathbf{x})\right]. \end{aligned}$$

As in the linear case, the equivalent kernel,

$$k(\mathbf{x}, \mathbf{y}) = \beta \boldsymbol{\phi}^T(\mathbf{x})\, \mathbf{S}_N \boldsymbol{\phi}(\mathbf{y})\,,$$

fully determines the predictive distribution. The ubiquity of the equivalent kernel proves extremely useful upon the realization that a choice of kernel supersedes the choice of basis functions $\boldsymbol{\phi}(\mathbf{x})$: instead of selecting a specific form for the basis functions a priori and deriving the equivalent kernel, one can select the kernel directly and avoid the basis functions entirely. This "kernel trick", however, is more than just a mathematical convenience. Kernels nonlinear in the feature space points $\mathbf{x}$ are equivalent to infinite dimensional basis functions that would be otherwise computationally infeasible.

## D.3 Gaussian Processes

Stochastic processes define probability distributions over infinite sets of random variables [91]. Gaussian processes are stochastic processes fully described by a mean function

$$\mu(\mathbf{x})$$

and covariance function

$$k(\mathbf{x}, \mathbf{y})\,,$$

which can be thought of as the infinite extension of a mean vector and covariance matrix characterizing a finite multivariate gaussian distribution.

An important property of stochastic processes is consistency: any finite subset of random variables must be distributed by the same distribution. For gaussian processes, this implies that any finite sample of points $\{\mathbf{x}_n\}$ is distributed as a multivariate gaussian

$$\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}, \mathbf{K})$$

where the mean and covariance are given by the defining functions

$$\mu_n = \mu(\mathbf{x}_n)\,,$$

and

$$K_{mn} = k(\mathbf{x}_m, \mathbf{x}_n)\,.$$

The consistency requirement ensures that sampling a finite subset of random variables from a gaussian process is equivalent to marginalizing over the infinite set of random variables not sampled.

An alternative approach to the estimation problem is then to model the function $y(\mathbf{x})$ as a gaussian process. For a finite sample of points, the gaussian process implements a prior over functions

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y}|\boldsymbol{\mu}, \mathbf{K})\,.$$

Taking the $\boldsymbol{\mu}$ to be $\mathbf{0}$ by symmetry and assuming a gaussian likelihood, the posterior becomes

$$\begin{aligned} p(\mathbf{t}|\mathbf{y}) &= \frac{p(\mathbf{y}|\mathbf{t})\, p(\mathbf{y})}{p(\mathbf{t})} \\ &= \frac{\left[\prod_{i=1}^{N} \mathcal{N}(\mathbf{y}|\mathbf{t}, \beta^{-1})\right] \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K})}{p(\mathbf{t})} \end{aligned}$$

Note that the evidence can be calculated analytically,

$$\begin{aligned} p(\mathbf{t}) &= \int d\mathbf{y}\, p(\mathbf{t}|\mathbf{y})\, p(\mathbf{y}) \\ &= \int d\mathbf{y} \left[\prod_{i=1}^{N} \mathcal{N}\big(\mathbf{y}|\mathbf{t}, \beta^{-1}\big)\right] \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K}) \\ &= \mathcal{N}(\mathbf{t}|\mathbf{0}, \mathbf{C})\,, \end{aligned}$$

where

$$\mathbf{C} = \mathbf{K} + \beta^{-1}\mathbb{I}.$$

The predictive distribution for a new point $\mathbf{x}_{N+1}$ is found by first computing the evidence

$$p(\mathbf{t}_{N+1}) = \mathcal{N}(\mathbf{t}|\mathbf{0}, \mathbf{C}_{N+1})\,.$$

and then conditioning on the existing measurements,

$$p(t_{N+1}|\mathbf{t}) = \int d\mathbf{t}\, \mathcal{N}(\mathbf{t}|\mathbf{0}, \mathbf{C}_{N+1})\,.$$

After first partitioning the covariance matrix

$$\mathbf{C}_{N+1} = \begin{pmatrix} \mathbf{C}_N & \mathbf{k} \\ \mathbf{k}^T & c \end{pmatrix},$$

with

$$k_n = k(\mathbf{x}_n, \mathbf{x}_{N+1})$$

and

$$c = k(\mathbf{x}_{N+1}, \mathbf{x}_{N+1}) + \beta^{-1},$$

the predictive distribution becomes

$$p(t_{N+1}|\mathbf{t}) = \mathcal{N}\left(\mathbf{k}^T\mathbf{C}_N^{-1}\mathbf{t}, c - \mathbf{k}^T\mathbf{C}_N^{-1}\mathbf{k}\right).$$

The gaussian process result is equivalent to the nonlinear Bayesian regression result for a slightly modified definition of the kernel and the choice of a kernel is then equivalent to the choice of a covariance function, only a wealth of understanding exists for the latter [92]. Utilizing this knowledge, covariance functions can be chosen implementing systems ranging from chaotic to smooth to periodic. There even exist covariance functions that implement infinitely large neural networks [93].

### D.3.1 Model Selection

In practice, any choice of covariance function will be parametric. A popular choice, for example, is the squared exponential,

$$\begin{aligned} C(\mathbf{x}, \mathbf{y}) &= k(\mathbf{x}, \mathbf{y}) + \beta^{-1}\delta(\mathbf{x} - \mathbf{y}) \\ &= \theta_1 \exp\left(\sum_{k=1}^{n} \rho_k |\mathbf{x} - \mathbf{y}|_k^2\right) + \theta_2 + \beta^{-1}\delta(\mathbf{x} - \mathbf{y}). \end{aligned}$$

Optimal choices for the hyperparameters $\boldsymbol{\alpha} = \{\theta_i, \rho_i, \beta\}$ are not immediately obvious, but the Bayesian formalism provides a powerful method for learning the hyperparameters from the data itself.

In the philosophy of Bayesian model selection, the optimal hyperparameters will maximize the evidence, or equivalently the log evidence

$$\log p(\mathbf{t}|\boldsymbol{\alpha}) = -\frac{1}{2} \log |\mathbf{C}_N| - \frac{1}{2} \mathbf{t}^T \mathbf{C}_N^{-1} \mathbf{t} - \frac{N}{2} \log (2\pi) \, .$$

Conveniently, the gradient of the log evidence with respect to the hyperparameters itself can be analytically evaluated

$$\frac{\partial}{\partial \alpha_i} \log p(\mathbf{t}|\boldsymbol{\alpha}) = -\frac{1}{2} \mathrm{Tr} \left( \mathbf{C}_N^{-1} \frac{\partial \mathbf{C}_N}{\partial \alpha_i} \right) + \frac{1}{2} \mathbf{t}^T \mathbf{C}_N^{-1} \frac{\partial \mathbf{C}_N}{\partial \alpha_i} \mathbf{C}_N^{-1} \mathbf{t}.$$

With these results efficient search algorithms, such as nonlinear conjugate gradients, can be used to explore the hyperparameter phase space and find a maximum, or at least a good approximation to one.

Optimizing the hyperparameters provides more than just improved predictions. Given a particular covariance function, the optimal hyperparameters provide valuable insight into the measured data. In the squared exponential, for example, the $\rho_i$ directly control the contribution of the $i$th variable to the predictive distribution and, as the evidence is optimized, variables irrelevant to the final result with have their respective $\rho_i$ driven to zero. Comparing the $\rho_i$ implements *automatic relevance detection* where the algorithm itself determines the relevance of each input variable from the data alone.

# Bibliography

[1] SLAC E130, Phys.Rev.Lett. **51**, 1135 (1983).

[2] EMC, Physics Letters B **206**, 364 (1988).

[3] D. de Florian, G. A. Navarro, and R. Sassot, Phys. Rev. **D71**, 094018 (2005), hep-ph/0504155.

[4] D. de Florian, R. Sassot, M. Stratmann, and W. Vogelsang, Phys. Rev. Lett. **101**, 072001 (2008), 0804.0422.

[5] M. Glück, E. Reya, M. Stratmann, and W. Vogelsang, Phys. Rev. D **63**, 094005 (2001).

[6] L. E. Gordon and W. Vogelsang, Phys. Rev. D **50**, 1901 (1994).

[7] I. Alekseev *et al.*, Nucl. Instrum. Methods A **499**, 392 (2003).

[8] K. H. Ackermann *et al.*, Nucl. Instrum. Methods A **499**, 624 (2003).

[9] M. Beddo *et al.*, Nucl. Instrum. Methods A **499**, 725 (2003).

[10] J. Sowinski, 2009 pp Run QC, `http://www.star.bnl.gov/protected/spin/sowinski/runQC_2009/`, 2009, (Accessed August 8, 2011) †.

[11] D. Griffiths, *Introduction to Elementary Particle Physics* (WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2004).

[12] H. Georgi, *Lie Algebras in Particle Physics* (Westview Press, Colorado, 1999).

[13] A. V. Manohar, (1992), hep-ph/9204208.

[14] R. Devenish and A. Cooper-Sarkar, *Deep Inelastic Scattering* (Oxford University Press, New York, 2004).

[15] J. B. Kogut and D. E. Soper, Phys.Rev. **D1**, 2901 (1970).

[16] R. L. Jaffe, (1996), hep-ph/9602236.

[17] M. E. Peskin and D. V. Schroeder, *And Introduction to Quantum Field Theory* (Westview Press, 1995).

[18] R. G. Roberts, *The structure of the proton* (Cambridge University Press, New York, 1993).

[19] J. R. Ellis and R. L. Jaffe, Phys.Rev. **D9**, 1444 (1974).

[20] R. Jaffe and A. Manohar, Nucl.Phys. **B337**, 509 (1990), Revised version.

[21] A. Chodos, R. L. Jaffe, K. Johnson, C. B. Thorn, and V. F. Weisskopf, Phys. Rev. D **9**, 3471 (1974).

[22] E. Hughes and R. Voss, Ann.Rev.Nucl.Part.Sci. **49**, 303 (1999).

[23] SMC, D. Adams *et al.*, Physics Letters B **329**, 399 (1994).

[24] X.-D. Ji, Phys. Rev. Lett. **78**, 610 (1997), hep-ph/9603249.

[25] X.-S. Chen, X.-F. Lü, W.-M. Sun, F. Wang, and T. Goldman, Phys. Rev. Lett. **100**, 232002 (2008).

[26] M. Wakamatsu, Phys. Rev. **D83**, 014012 (2011), 1007.5355.

[27] Y. Hatta, (2011), 1101.5989.

[28] S. Bashinsky and R. L. Jaffe, Nucl. Phys. **B536**, 303 (1998), hep-ph/9804397.

[29] J. C. Collins, D. E. Soper, and G. F. Sterman, Adv.Ser.Direct.High Energy Phys. **5**, 1 (1988), hep-ph/0409313.

[30] B. W. Filippone and X.-D. Ji, Adv. Nucl. Phys. **26**, 1 (2001), hep-ph/0101224.

[31] D. E. Soper, Nuclear Physics B - Proceedings Supplements **53**, 69 (1997), Lattice 96.

[32] G. Bunce, N. Saito, J. Soffer, and W. Vogelsang, Ann.Rev.Nucl.Part.Sci. **50**, 525 (2000), hep-ph/0007218.

[33] I. W. Stewart, F. J. Tackmann, and W. J. Waalewijn, Phys. Rev. **D81**, 094035 (2010), 0910.0467.

[34] S. Frixione, Phys.Lett. **B429**, 369 (1998), hep-ph/9801442.

[35] T. Becher and M. D. Schwartz, JHEP **1002**, 040 (2010), 0911.0681.

[36] W. Vogelsang, Personal Communication, 2011.

[37] M. Harrison, T. Ludlam, and S. Ozaki, Nucl. Instrum. Methods A **499**, 235 (2003).

[38] T. Ludlam, Nucl. Instrum. Methods A **499**, 428 (2003).

[39] H. Hahn *et al.*, Nucl. Instrum. Methods A **499**, 245 (2003).

[40] A. Zelenski *et al.*, Optically-pumped polarized h- ion sources for rhic and hera colliders, in *Particle Accelerator Conference, 1999. Proceedings of the 1999* Vol. 1, pp. 106 –108 vol.1, 1999.

[41] L. W. Anderson *et al.*, Phys. Rev. Lett. **52**, 609 (1984).

[42] W. Mackay *et al.*, Commissioning and future plans for polarized protons in rhic, in *Particle Accelerator Conference, 2001. PAC 2001. Proceedings of the 2001* Vol. 1, pp. 24 –26 vol.1, 2001.

[43] K. O. Eyser *et al.*, AIP Conf. Proc. **915**, 916 (2007), nucl-ex/0612020.

[44] F. Bergsma *et al.*, Nucl. Instrum. Methods A **499**, 633 (2003).

[45] M. Anderson *et al.*, Nucl. Instrum. Methods A **499**, 659 (2003).

[46] The STAR Collaboration, J. Kiryluk, (2005), hep-ex/0501072.

[47] F. S. Bieser *et al.*, Nucl. Instrum. Methods A **499**, 766 (2003).

[48] J. M. Landgraf *et al.*, Nucl. Instrum. Methods A **499**, 762 (2003).

[49] B. G. Gibbard and T. G. Throwe, Nucl. Instrum. Methods A **499**, 814 (2003).

[50] L. Didenko, Production Options, `http://www.star.bnl.gov/devcgi/dbProdOptionRetrv.pl`, 2011, (Accessed August 8, 2011) †.

[51] B. Page, Prioritized Run List Production Primer, `http://drupal.star.bnl.gov/STAR/blog/pagebs/2011/jan/13/prioritized-run-list-production-primer`, 2011, (Accessed August 8, 2011) †.

[52] G. Webb, Run QA of pp 2009 200GeV, `http://drupal.star.bnl.gov/STAR/blog/gdwebb/2010/mar/19/run-qa-pp-2009-rff-200gev`, 2010, (Accessed August 8, 2011) †.

[53] R. Fatemi, 2009 200 GeV SpinDB (full MuDst), `http://drupal.star.bnl.gov/STAR/blog/rfatemi/2010/mar/23/2009-200-gev-spindb-full-mudst`, 2010, (Accessed August 8, 2011) †.

[54] T. Sjostrand, S. Mrenna, and P. Z. Skands, JHEP **05**, 026 (2006), hep-ph/0603175.

[55] A. Buckley, H. Hoeth, H. Lacker, H. Schulz, and J. E. von Seggern, Eur. Phys. J. **C65**, 331 (2010), 0907.2973.

[56] M. Goossens, *GEANT: Detector Description and Simulation Tool, long writeup W5013; March 1994* CERN Program Library (CERN, Geneva, 1993).

[57] M. Potehkin, GSTAR, `http://www.star.bnl.gov/public/comp/simu/gstar/gstar.html`, 2003, (Accessed August 8, 2011).

[58] S. Trentalange, BTOW/BSMD Geometry Changes, `http://drupal.star.bnl.gov/STAR/blog-entry/trent/2008/sep/15/btwo-geometry-changes`, 2008, (Accessed August 8, 2011) †.

[59] M. Betancourt, BEMC Geometry Study, `http://drupal.star.bnl.gov/STAR/blog/betan/2009/sep/14/bemc-geometry-study`, 2009, (Accessed August 8, 2011) †.

[60] BFChain Options, `http://www.star.bnl.gov/cgi-bin/protected/cvsweb.cgi/~checkout~/StRoot/StBFChain/doc/index.html`, 2011, (Accessed August 8, 2011) †.

[61] M. Betancourt *et al.*, STAR Note **SN0523** (2010).

[62] M. Betancourt, 2009 BEMC Photon Filter Validation and Simulation Production Request, `http://drupal.star.bnl.gov/STAR/blog/betan/2010/aug/04/2009-bemc-photon-filter-validation-and-simulation-production-request`, 2009, (Accessed August 8, 2011) †.

[63] R. Pordes *et al.*, Journal of Physics: Conference Series **125**, 012070 (2008).

[64] J. Skilling, Nested sampling, in *Maximum Entropy and Bayesian methods in science and engineering*, edited by G. Erikson, J. T. Bercher, & C. R Smith, , American Institute of Physics Conference Series Vol. 735, pp. 395–405, AIP Press, 2004.

[65] M. Betancourt, Nested sampling with constrained hamiltonian monte carlo, in *Maximum Entropy and Bayesian methods in science and engineering*, edited by A. Mohammad-Djafari, J.-F. Bercher, & P. Bessiére, , American Institute of Physics Conference Series Vol. 1305, pp. 165–172, AIP Press, 2011, 1005.0157.

[66] C. Bishop, *Pattern Classification and Machine Learning* (Springer, New York, 2007).

[67] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning* (The MIT Press, Cambridge, Massachusetts, 2006).

[68] D. MacKay, *Information Theory, Inference, and Machine Learning* (Cambridge University Press, New York, 2003).

[69] E. Snelson, C. E. Rasmussen, and Z. Ghahramani, Warped gaussian processes, in *Advances in Neural Information Processing Systems 16*, edited by S. Thrun, L. Saul, and B. Schölkopf, MIT Press, Cambridge, MA, 2004.

[70] M. J. Betancourt, ArXiv e-prints (2010), 1009.5604.

[71] R. Neal, Mcmc using hamiltonian dynamics, in *Handbook of Markov Chain Monte Carlo*, edited by S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, CRC Press, New York, 2011.

[72] M. J. Betancourt, ArXiv e-prints (2010), 1010.3436.

[73] A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian Data Analysis* (Chapman & Hall/CRC Press, Boca Raton, Florida, 2004).

[74] STAR Collaboration, J. Adams *et al.*, Phys. Rev. Lett. **91**, 172302 (2003).

[75] J. Dunlop, p+p Run 9 Trigger FAQ, `http://orion.star.bnl.gov/protected/common/common2009/trigger2009/triggers2009.html`, 2011, (Accessed August 24, 2011) †.

[76] I. Alekseev *et al.*, Run-09 pC Polarimeter Analysis, `http://www4.rcf.bnl.gov/~cnipol/pubdocs/Run09Offline/pC_2009.pdf`, 2010, (Accessed August 24, 2011).

[77] J. Seele, Relative Luminosity Studies - IV, `http://drupal.star.bnl.gov/STAR/blog/seelej/2011/jan/11/relative-luminosity-studies-iv`, 2011, (Accessed August 24, 2011) †.

[78] J. Seele, Relative Luminosity Studies - VII, `http://drupal.star.bnl.gov/STAR/blog/seelej/2011/jan/27/relative-luminosity-studies-vii`, 2011, (Accessed August 24, 2011) †.

[79] J. Seele, Relative Luminosity Studies - VIII, `http://drupal.star.bnl.gov/STAR/blog/seelej/2011/feb/01/relative-luminosity-studies-viii`, 2011, (Accessed August 24, 2011) †.

[80] STAR, P. Djawotho, (2011), 1106.5769.

[81] STAR, M. Walker, (2011), 1107.0917.

[82] B. Surrow, Personal Communication, 2011.

[83] D. Boer *et al.*, (2011), 1108.1713.

[84] X.-d. Ji, Nucl. Phys. Proc. Suppl. **119**, 41 (2003), hep-lat/0211016.

[85] D. Sivia and J. Skilling, *Data Analysis* (Oxford University Press, New York, 2008).

[86] J. M. Bernardo and A. Smith, *Bayesian Theory* (John Wiley and Sons, New York, 2004).

[87] E. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, New York, 2003).

[88] R. T. Cox, American Journal of Physics **14**, 1 (1946).

[89] K. H. Knuth and J. Skilling, ArXiv e-prints (2010), 1008.4831.

[90] S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, editors, *Handbook of Markov Chain Monte Carlo* (CRC Press, New York, 2011).

[91] A. Papoulis and S. Pillai, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York, 2002).

[92] P. Abrahamsen, Norwegian Computing Center Technical Report **917** (1997).

[93] R. Neal, *Bayesian Learning For Neural Networks* (Springer, New York, 1996).