UNIVERSITY OF CALIFORNIA,
IRVINE

Machine learning applied to parameter spaces of theories beyond the Standard Model

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Physics and Astronomy

by

Jacob Hollingsworth

DISSERTATION COMMITTEE:
Professor Daniel Whiteson, Chair
Professor David Kirkby
Professor Michael Ratz

2021

DEDICATION

To my parents

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Algorithms

Page

ACKNOWLEDGMENTS

Above all, I am thankful for my advisor, Daniel Whiteson, who gave me the freedom to explore and provided much needed course correction when I wandered too far off track. I am also grateful for my collaborators, Philip Tanedo and Michael Ratz, who have made significant contributions to Chapter 4 of this thesis.

I would like to thank the many people that have contributed valuable discussion on the projects presented in this thesis. These include Felix Kling, Johann Brehmer, Kyle Cranmer, Syris Norelli, Stephan Mandt, Babak Shababa, Shimone Whiteson, Benjamin Nachman, Tim Cohen, Jesse Thaler, Tim Tait, Michael Ratz, Philip Tanedo, Jonathan Feng, Anyes Taffard, and the ATLAS SUSY working group. I would also like to thank Michael Ratz and David Kirkby for agreeing to serve on my defense committee.

I appreciate my family for their support and patience during the completion of this thesis. I also would like to thank my friends at UCI and particularly those in my cohort, who were always open to discussing a problem with me. Finally, I would like to thank Chell for their motivation during the writing of this thesis.

Parts of this work received funding from National Science Foundation Graduate Research Fellowship under Grant No. DGE-1321846 and the Physical Sciences Machine Learning Nexus Fellows program at UCI.

This work has benefited greatly from numerous open source software projects without which this thesis would not have been possible. I would particularly like to acknowledge the authors and communities that have contributed to the development of Pytorch [3], Tensorflow [4], SKLearn [5], Pandas [6], Numpy [7], NFlows [8], Madminer [9], Madgraph [10], Pythia [11], Delphes [12], SoftSUSY [13], MicrOmegas [14], SUSY-AI [15], SModelS [16], and CheckMATE [17].

VITA

Jacob Hollingsworth

EDUCATION

Doctor of Philosophy in Physics and Astronomy **2021**
University of California, Irvine *Irvine, CA*

Bachelor of Science in Physics **2016**
Ursinus College *Collegeville, PA*

RESEARCH EXPERIENCE

Graduate Research Assistant **2016–2021**
University of California, Irvine *Irvine, California*

Undergraduate Research Assistant **2013–2016**
Ursinus College *Collegeville, Pennsylvania*

Undergraduate Research Assistant **2015**
Rice University *Houston, Texas*

TEACHING EXPERIENCE

Teaching Assistant **2020–2021**
University of California, Irvine *Irvine, California*

REFEREED JOURNAL PUBLICATIONS

Efficient sampling of constrained high-dimensional theoretical spaces with machine learning **2021**
Jacob Hollingsworth, Philip Tanedo, Michael Ratz, Daniel Whiteson
Accepted with minor revision to The European Physical Journal C

Resonance Searches with Machine Learned Likelihood Ratios **2021**
Jacob Hollingsworth, Daniel Whiteson
Accepted with minor revisions to SciPost Physics

Can exact conditions improve machine-learned density functionals? **2018**
Jacob Hollingsworth, Li Li, Thomas Baker, Kieron Burke
The Journal of Chemical Physics

Geometric representation of spin correlations and applications to ultracold systems **2018**
Rick Mukherjee, Anthony E. Mirasola, *Jacob Hollingsworth*, Ian G. White, Kaden R. A. Hazzard
Physical Review A

Quantum interference in the field ionization of Rydberg atoms **2015**
Rachel Feynman, *Jacob Hollingsworth*, Michael Vennettilli, Tamas Budner, Ryan Zmiewski, Donald P. Fahey, Thomas J. Carroll, Michael W. Noel
Physical Review A

ABSTRACT OF THE DISSERTATION

Machine learning applied to parameter spaces of theories beyond the Standard Model

By

Jacob Hollingsworth

Doctor of Philosophy in Physics and Astronomy

University of California, Irvine, 2021

Professor Daniel Whiteson, Chair

Several theoretical parameter spaces are analysed using techniques from machine learning. First, machine learning is used to leverage high dimensional detector information to constrain the mass and coupling of a simplified model which extends the standard model with a heavy Z' boson. The high dimensional information is seen to improve the exclusion contours of the analysis relative to a low dimensional analysis that instead uses summary statistics. Next, generative models are used to sample a high dimensional parameter space of a supersymmetric model subject to a constraint. Generative models are seen to provide an increase in efficiency of over an order of magnitude in a search for models that satisfy the constraint when compared to random sampling. Finally, we analyse a dataset of supersymmetric theories to understand how they may best discovered in experiment. We propose a set of experiments to be performed at the Large Hadron Collider that are most sensitive to these models.

Chapter 1

Introduction

At the time of writing, the Standard Model of particle physics (SM) is the most successful description of physics at the subatomic level. The SM has endured decades of diverse and sensitive experimental challenges to its legitimacy, with no experiment yet reaching the conventional 5σ significance required to claim a discovery of physics incompatible with the SM. This success comes in spite of numerous very strong challenges to the Standard Model. Issues with the SM include the lack of a consistent description of gravity, the absence of a cold dark matter candidate, and an insufficient explanation for the observed matter antimatter asymmetry.

The SM also contains many inelegant aspects that, while not disqualifying, may hint towards a more fundamental theory. The first of these is the hierarchy problem, where large cancellations are required to achieve the observed Higgs boson mass in the presence of novel physics at high energies. Also included is the strong CP problem, where a parameter in the strong sector of the SM (θ_{QCD}) is found experimentally to be very nearly zero, though no theoretical explanation for this is available.

A last element of the SM that hints towards a more fundamental description of physics is the large number of free parameters, primarily masses and couplings, that are required in order to make quantitative predictions. The SM can be abstractly thought of as a black-box function; mapping input variables to experimental observables. However, this function also requires additional parameters that are unknown and may only be determined in experiment. In total, the SM requires 19 additional parameters, which are given in Table I. These parameters form a 19 dimensional "theory space" of possible universes *a priori*. As experiments are performed to determine the value of a parameter, the dimension of the surviving space consistent with experiment reduces by one until a single point is reached. For the SM, all parameters have been determined in experiment. However, the requirement for such a large number of input parameters may lack the aesthetic appeal that one may expect of a truly fundamental theory.[1]

Theories of physics beyond the Standard Model (BSM) attempt to overcome one or more of these challenges faced by the SM. This is often accomplished by proposing new physics at energy scales that have thus far been unobtainable in experiment. BSM theories typically inherit most or all of the free parameters of the SM and include additional free parameters related to the novel physics introduced by the theory.

There are practical challenges in reducing the theory space of BSM models. Because the added parameters describe physics at scales out of reach of current experiments, it is incredibly difficult to determine their value. One must either carefully analyze all available data for limited direct evidence of BSM physics or very precisely measure low energy observables for possible influence from BSM physics. This is made more difficult in many BSM models where experimental observables are complicated functions of the fundamental theory parameters, so that experimental constraints are nontrivial to apply to the theory space.

[1] Many oppose the legitimacy arguments derived from aesthetics, as there is no reason that a fundamental theory must appeal to a theorist's notion of beauty. Here we do not cast judgement on the validity or invalidity of such complaints, and merely introduce the discussion to later expand on the practical difficulties that accompany the necessity of a large number of free parameters.

Parameter	Description
m_e	Electron mass
m_μ	Muon mass
m_τ	τ mass
m_u	Up quark mass
m_c	Charm quark mass
m_t	Top quark mass
m_d	Down quark mass
m_s	Strange quark mass
m_b	Bottom quark mass
θ_{12}	CKM matrix 12 element
θ_{13}	CKM matrix 13 element
θ_{23}	CKM matrix 23 element
δ^{CP}	CKM matrix phase
g_1	$U(1)$ gauge coupling
g_2	$SU(2)$ gauge coupling
g_3	$SU(3)$ gauge coupling
θ_{QCD}	QCD vacuum angle
v	Higgs vacuum expectation value
m_h	Higgs boson mass

Table 1.1: Free parameters of the Standard Model with a brief description. Most parameters are particle masses and couplings between different particles. Does not include free parameters needed to account for neutrino masses in the SM.

Additionally, BSM parameter spaces are often very high dimensional. This makes standard parameter scan based approaches of finding experimentally consistent regions computationally expensive. As a concrete example, much of this thesis will discuss a BSM model with 19 unknown free parameters. Simply sampling a small, medium, and large value for each parameter would require 3^{19}, or over 1 billion, model evaluations to completely cover the space. This is a manifestation of the curse of dimensionality, where the number of configurations within a space increases exponentially with its dimension.

Recently, machine learning has been incredibly successful at performing diverse tasks in extremely high dimensional spaces. Benefiting from the relative availability of large datasets as well as increasing GPU capabilities, deep learning in particular has been widely adopted for learning from large, high dimensional datasets. As such datasets are often also found in high energy physics, it is natural to turn to deep learning as a potential solution to the model selection problem in high energy physics previously described. The focus of this thesis is to study the use of deep learning as an improvement upon current methods of determining parameters for BSM models.

In Chapter 2, we briefly review the SM and give a full review of the Minimal Supersymmetric Standard Model (MSSM), a theoretically favored high dimensional BSM model. We also introduce the constrained MSSM (cMSSM) and phenomenological MSSM (pMSSM), two lower dimensional MSSM subspaces that will be discussed throughout this work. Towards the end of this discussion, we introduce the concept of simplified models. In the second half of this chapter, we review fundamental aspects of machine learning, with particular emphasis on supervised learning.

In Chapter 3, we use novel machine learning algorithms to determine the mass and coupling of a new particle in a simplified model for a simulated experiment. While the parameter space is only 2 dimensional, machine learning is used to leverage the full dimensionality of the space of observables in the statistical analysis of the data. This allows us to determine the

values for the mass and coupling with greater confidence than standard approaches, which often utilize low dimensional summary statistics in their statistical analyses.

In Chapter 4, we utilize two generative frameworks to directly produce pMSSM models that are consistent with an experimental constraint. This bypasses the need to perform extensive parameter scans in order to construct the region of the parameter space that survives the constraint. Additionally, the generative frameworks are carefully chosen so that the search results are not biased by the use of a generative model, enabling one to confidently perform further searches on the generated data points.

In Chapter 5, we perform a systematic study of the pMSSM parameter space with the hope of guiding experimentalists towards analyses that could best exclude the remaining unexcluded regions of the pMSSM. We first develop a classification scheme for pMSSM model points based upon the mechanism used to achieve the experimentally observed amount of dark matter. We then study the models in each class to determine which analyses could be performed at collider experiments to best exclude that class of models.

Chapter 2

Background

2.1 Particle Physics

2.1.1 The Standard Model

The Standard Model of particle physics posits the matter content of the universe and describes the behavior of this matter with a Lagrangian density constructed from all renormalizable terms that are invariant with respect to local transformations of the gauge group $SU(3)_c \times SU(2)_L \times U(1)_Y$ and global transformations of the Poincaré group. The matter content of the SM and their representation in the gauge group is summarized in Table 2.1. We break the Lagrangian density into 3 distinct parts: the electroweak sector, the quantum chromodynamics sector and the Higgs sector.

$$\mathcal{L}_{SM} = \mathcal{L}_{EW} + \mathcal{L}_{QCD} + \mathcal{L}_{Higgs}. \tag{2.1}$$

Fermions (Spin $\frac{1}{2}$)		
Field	Description	Representation $(SU(3)_c,\, SU(2)_L,\, U(1)_Y)$
Q_i	Left-handed quark doublet	$(\mathbf{3},\, \mathbf{2},\, \frac{1}{3})$
$u_{R,i}$	Right-handed up quark	$(\mathbf{3},\, \mathbf{1},\, \frac{4}{3})$
$d_{R,i}$	Right-handed down quark	$(\mathbf{3},\, \mathbf{1},\, -\frac{2}{3})$
L_i	Left-handed lepton doublet	$(\mathbf{1},\, \mathbf{2},\, -1)$
$\ell_{R,i}$	Right-handed lepton	$(\mathbf{1},\, \mathbf{1},\, -2)$
Bosons (Spin 1)		
Field	Description	Representation $(SU(3)_c,\, SU(2)_L,\, U(1)_Y)$
B	$U(1)_Y$ gauge boson	$(\mathbf{1}, \mathbf{1}, 0)$
W^a	$SU(2)_L$ gauge bosons	$(\mathbf{1}, \bar{\mathbf{3}}, 0)$
G^a	$SU(3)_c$ gauge bosons	$(\bar{\mathbf{8}}, \mathbf{1}, 0)$
Scalars (Spin 0)		
Field	Description	Representation $(SU(3)_c,\, SU(2)_L,\, U(1)_Y)$
H	Higgs field	$(\mathbf{1}, \mathbf{2}, 1)$

Table 2.1: Particle content of the Standard Model. Representations under which the fields transform with respect to gauge transformations are given in the third column. For $U(1)_Y$ transformations, the hypercharge is given instead. The index $i = 1, 2, 3$ and runs over the 3 generations of fermions, which differ only in mass.

This section will review each sector of the SM individually. The focus is primarily on uncovering the appearance and meaning of free parameters in each sector of the theory. This discussion is heavily inspired by the references [18, 19, 20].

The Electroweak Sector

The electroweak weak sector of the Standard Model contains the kinetic terms for fermion fields and gauge boson fields associated with the $SU(2)_L \times U(1)_Y$ symmetry. Before directly discussing the electroweak sector, we first discuss the procedure for constructing gauge invariant Lagrangian density for an arbitrary local, continuous gauge symmetry. This will greatly simplify future discussions of the electroweak and quantum chromodynamics sectors of the Standard Model.

A local, continuous gauge transformation V acts on fermions in the fundamental representation of the gauge group as

$$\psi' = V\psi. \tag{2.2}$$

Kinetic terms for fermions require derivatives of ψ, however, naïve derivatives of ψ are difficult to make gauge invariant due to the appearance of an additional term

$$(\partial_\mu \psi)' = V\partial_\mu \psi + (\partial_\mu V)\,\psi. \tag{2.3}$$

This is easily understood as the derivative of ψ depends on its value in a neighborhood of points around x. So, the transformed derivative must depend on how this neighborhood of points transforms as well and thus cannot be locally gauge invariant. Because of this, the comparator is introduced, which allows distant points to transform in the same way:

$$U(y,x)' = V(y)U(y,x)V(x)^\dagger \tag{2.4}$$

It is easily verified that $\psi(y)$ and $U(y,x)\psi(x)$ transform in the same way and that $U(y,y) = 1$ is required for consistency. The gauge covariant derivative is defined

$$n^\mu D_\mu \psi = \lim_{\epsilon \to 0} \frac{\psi(x+\epsilon n) - U(x+\epsilon n, x)\psi(x)}{\epsilon}. \tag{2.5}$$

This modifies the usual derivative to allow for local gauge invariance. Since the derivative only depends on points in the neighborhood of $\psi(x)$, we expand around $U(x,x) = 1$.

$$U(x+\epsilon n, x) = 1 + ig\epsilon n^\mu A^a_\mu(x) T_a + \mathcal{O}(\epsilon^2) \tag{2.6}$$

where n is a four-vector, g is a scaling constant known as the coupling, T_a are the generators of the Lie group, and A^a_μ is a field that quantifies the transformation associated with rotation

in the a direction of the group in the n^μ direction of spacetime to first order in ϵ. Plugging in, we see that

$$D_\mu \psi = (\partial_\mu - igA^a_\mu T_a)\psi \tag{2.7}$$

A^a_μ can be shown to transform according to

$$\left(A^a_\mu T_a\right)' = V\left(A^a_\mu T_a + \frac{i}{g}\partial_\mu\right)V^\dagger \tag{2.8}$$

by substituting Equation 2.6 into Equation 2.4. Gauge covariant derivatives thus transform in the same manner as the fields on which they act. So, they may be used to easily construct kinetic terms for fermions such as $i\bar{\psi}\gamma^\mu D_\mu \psi$. This also implies that higher order gauge covariant derivatives transform in the same way. This may be used to construct kinetic terms for the newly introduced fields A^a_μ.

One can see that the commutator of gauge covariant derivatives depends only on derivatives of the gauge fields A^a_μ.

$$[D_\mu, D_\nu]\psi = -ig\left(\partial_\nu A_\mu - \partial_\mu A_\nu + gA^a_\mu A^b_\nu f^c_{ab} T_c\right)\psi \tag{2.9}$$

where we have defined $A^a_\mu T_a = A_\mu$ and f^c_{ab} are the structure constants of the Lie group. Evaluating the commutation reveals a tensor $F^i_{\mu\nu} = (\partial_\mu A^i_\nu - \partial_\nu A^i_\mu + gA^a_\mu A^b_\nu f^i_{ab})$ that depends only on derivatives of the gauge field. Substituting this definition into the previous expression shows that

$$[D_\mu, D_\nu] = -igF^i_{\mu\nu}T_i \tag{2.10}$$

One can use this commutation relation to show that terms such as $F^{\alpha\beta}_i F^{\mu\nu,i}$ are gauge invariant. These are gauge field kinetic terms as they depend on the square of the derivative

of the gauge fields. There are two possible Poincaré invariant combinations, $\mathrm{Tr}\left[F^{\mu\nu}F_{\mu\nu}\right]$ and $\epsilon_{\alpha\beta\mu\nu}\mathrm{Tr}\left[F^{\alpha\beta}F^{\mu\nu}\right]$.

These concepts allow us to write down the most general Lagrangian density for this theory that is compatible with gauge and Poincaré symmetries, known as the Yang-Mills Lagrangian [21]:

$$\mathcal{L}_{Yang-Mills} = -\frac{1}{4}F_i^{\mu\nu}F^i_{\mu\nu} - \frac{g^2\theta_{YM}}{32\pi^2}\epsilon^{\alpha\beta\mu\nu}F^i_{\alpha\beta}F_{\mu\nu,i} + \bar{\psi}\left(i\gamma^\mu D_\mu - m\right)\psi. \tag{2.11}$$

We can now turn to the Standard Model gauge groups, beginning with the electroweak $(SU(2)_L \times U(1)_Y)$ sector [22, 23] Left handed fermions are organized into $SU(2)$ doublets

$$Q_i = \begin{bmatrix} u_{L,i} \\ d_{L,i} \end{bmatrix} \qquad\qquad L_i = \begin{bmatrix} \nu_{L,i} \\ \ell_{L,i} \end{bmatrix} \tag{2.12}$$

and right handed fermions are $SU(2)$ singlets. The gauge covariant derivatives for these fermions are given by

$$D_\mu = \begin{cases} \partial_\mu - ig_1 Y B_\mu & SU(2) \text{ singlets} \\ \partial_\mu - ig_1 Y B_\mu - \frac{i}{2}g_2 W^a_\mu \sigma_a & SU(2) \text{ doublets,} \end{cases} \tag{2.13}$$

where Y is the hypercharge of the fermion and the $\frac{\sigma_i}{2}$ appear as these generate transformations in $SU(2)$. The fields W^a_μ are associated with generators of $SU(2)_L$ and B_μ is associated with $U(1)_Y$. The covariant derivatives here must be augmented for $SU(3)$ triplets as well; see the next subsection.

We will adopt the standard notation for the fermion fields, which writes the Weyl spinors of the Standard Model in Dirac spinor notation, where projection to the chirality of the field

is implied. With this, we can rewrite Equation 2.11 as:

$$\mathcal{L}_{EW} = -\frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{4}W^{a}_{\mu\nu}W^{\mu\nu}_{a} + \sum_{f} i\bar{f}\gamma^{\mu}D_{\mu}f, \tag{2.14}$$

where the sum over f runs over all fields $\{L_i, Q_i, u_{R,i}, d_{R,i}, \ell_{R,i}\}$. We remark that terms proportional to θ_{YM} in Equation 2.11 are not present in the electroweak sector as they may be absorbed into fermion field definitions via chiral rotations: $\psi \to \exp(i\gamma^5\theta_{YM})\psi$. Additionally, the mass terms of Equation 2.11 vanish due to the chirality of the fermions. We also remark that one cannot construct Dirac mass terms such as, for example, $m\left(\bar{\ell}_{L,i}\ell_{R,i} + \bar{\ell}_{R,i}\ell_{L,i}\right)$, as these are not $SU(2)_L$ invariant.

Quantum Chromodynamics Sector

The quantum chromodynamics (QCD) sector of the Standard Model pertains to the particle fields that transform non-trivially under the $SU(3)_c$ group, namely the quarks and gluons [24]. Quarks are organized in $SU(3)$ triplets:

$$u_{L/R,i} = \begin{bmatrix} u_{L/R,i,r} \\ u_{L/R,i,g} \\ u_{L/R,i,b} \end{bmatrix} \qquad\qquad d_{L/R,i} = \begin{bmatrix} d_{L/R,i,r} \\ d_{L/R,i,g} \\ d_{L/R,i,b} \end{bmatrix}. \tag{2.15}$$

The gauge covariant derivatives from the previous section are augmented with a new gauge field:

$$D_{\mu} = D_{\mu,EW} - \frac{i}{2}g_3 G^{a}_{\mu}\lambda_a \qquad\qquad SU(3) \text{ triplets}, \tag{2.16}$$

where $D_{\mu,EW}$ refers to D_μ as defined in Equation 2.13 and the λ_a refer to the Gell-Mann matrices, which generate $\frac{\lambda_a}{2}$ rotations $SU(3)$ up to a factor of 2.

We can now write down the QCD Lagrangian by once again referring to Equation 2.11

$$\mathcal{L}_{QCD} = -\frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a - \frac{g^2 \theta_{QCD}}{32\pi^2} \epsilon^{\alpha\beta\mu\nu} G^i_{\alpha\beta} G_{\mu\nu,i}. \tag{2.17}$$

Quark kinetic terms are not included here as they have been included in the electroweak sector of the Lagrangian. For $SU(3)$ gauge symmetries, the θ term no longer vanishes, though it is constrained to be extremely close to 0 by experiment.[1]

The Higgs Sector

The Higgs sector describes the behavior and interactions of the lone scalar field in the Standard Model, the Higgs field [26, 27, 28]. This Higgs field is responsible for breaking the $SU(2)_L \times U(1)_Y$ gauge group to the familiar $U(1)_{EM}$ gauge group. In breaking this symmetry, it also provides mass terms for the fermions of the Standard Model as well as the weak gauge bosons. While the phenomenology of this sector is very rich, for this thesis we do not recount the full consequences of electroweak symmetry breaking in detail. Instead, we focus on constructing the Lagrangian before electroweak symmetry breaking so that the full SM parameter space is introduced.

The Higgs sector introduces a new scalar field Φ, known as the Higgs field. The Higgs field is an $SU(2)$ doublet and can be written in terms of 2 complex scalar fields ϕ^+ and ϕ^0:

$$\Phi = \frac{1}{\sqrt{2}} \begin{bmatrix} \phi^+ \\ \phi^0 \end{bmatrix}. \tag{2.18}$$

[1] There is a long theoretical discussion surrounding the θ_{QCD} term. Here, this term no longer vanishes in the sense that it cannot be absorbed into field redefinitions by performing chiral rotations, as could be done for the analogous $U(1)$ and $SU(2)$ terms. However, it may be written as a total derivative of the Chern-Simons current, and so its effects are invisible in perturbation theory. This term does have non-perturbative effects due to gauge field configurations that cannot be built perturbatively from the vacuum. The strong experimental constraints are derived from one such non-perturbative effect on the neutron electric dipole moment [25].

The most general Lagrangian that is invariant with respect to Lorentz and gauge transformations for Φ (excluding interactions with other fields) is given by

$$\mathcal{L}_{Higgs} \supset |D_\mu \Phi|^2 + \mu^2 \Phi^\dagger \Phi - \lambda \left(\Phi^\dagger \Phi\right)^2 . \tag{2.19}$$

Introducing a scalar field to the theory also allows for many gauge invariant interaction terms. Each of these comes with an associated coupling known as a Yukawa coupling. The full Lagrangian including these interaction terms is:

$$\mathcal{L}_{Higgs} = |D_\mu \Phi|^2 + \mu^2 \Phi^\dagger \Phi - \lambda \left(\Phi^\dagger \Phi\right)^2 + \left(\bar{Q}_i \tilde{\Phi} y_{u,ij} u_{R,j} + \bar{Q}_i \Phi y_{d,ij} d_{R,j} + \bar{L}_i \Phi y_{\ell,ij} \ell_{R,j} + \text{h.c.}\right) \tag{2.20}$$

where $\tilde{\Phi} = i\sigma_2 \Phi^*$. One can check manually that these additional interaction terms are gauge invariant.

Symmetry breaking occurs because the potential for Φ is minimized at some nonzero value $\langle \Phi \rangle$. We may choose a gauge where

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ v \end{bmatrix} . \tag{2.21}$$

Substituting this into the previous equation yields terms:

$$\mathcal{L}_{Higgs} \supset \bar{u}_{L,i} m_{u,ij} u_{R,j} + \bar{d}_{L,i} m_{d,ij} d_{R,j} + \bar{\ell}_{L,i} m_{\ell,ij} \ell_{R,j} + \text{h.c.} \tag{2.22}$$

We observe the appearance of Dirac mass terms for fermions, where we have identified $m_{a,ij} = \frac{y_{a,ij} v}{\sqrt{2}}$ for $a = \{u, d, \ell\}$.

Naïvely, we have introduced 54 new parameters, as each new interaction term is accompanied by a (potentially complex) mass. However, redundancies in field definition and basis rotations cause many of these free parameters to be unphysical. The matrix m_ℓ may be diagonalized by a bi-unitary transformation $m_{\ell,diag} = U_\ell m_\ell V_\ell^\dagger$.[2] Mapping $\ell_{L,i} \to U_{\ell,ij}\ell_{L,j}$ and $\ell_{R,i} \to V_{\ell,ij}\ell_{R,j}$ diagonalizes the lepton mass matrix m_ℓ, and the only physical parameters remaining are the diagonal masses.

For the quark terms, a similar decomposition may not be performed. As for the lepton masses, y_u and y_d may be diagonalized analagously

$$m_{u,diag} = U_u m_u V_u^\dagger \tag{2.23}$$

$$m_{d,diag} = U_d m_d V_d^\dagger. \tag{2.24}$$

However, the matrices for each basis transformation are in general not equal, which has phenomenological consequences. The fields $d_{L,i}$ and $u_{L,i}$ couple to $W^\pm$ bosons in the gauge covariant derivatives as they appear in an SU(2) doublet. Rotating $d_L \to U_d d_L$ and $u_L \to U_u u_L$ causes formerly diagonal interactions in these kinetic terms to gain off-diagonal interactions according to the matrix $U_u^\dagger U_d$. This matrix, known as the Cabbibo-Kobayashi-Maskawa (CKM) matrix, parameterizes the newly introduced off-diagonal couplings that result from the transformation to mass eigenstates. The CKM matrix contains 4 real parameters, $\theta_{12}, \theta_{13}, \theta_{23}, \delta^{CP}$, after the absorption of unphysical relative phases into the definition of quark fields.

Summary and Clarifying Remarks

We have constructed the Standard Model Lagrangian. We have seen that there are 19 free parameters which must be input to the theory. The electroweak sector of the SM

[2] In machine learning this is instead called a singular value decomposition.

Lagrangian contains 2 free parameters: the couplings associated with the $U(1)_Y$ and $SU(2)_L$ gauge groups, g_1 and g_2, respectively. The QCD sector also contains two free parameters: the coupling associated with the $SU(3)$ gauge group g_3 as well as θ_{QCD}. The Higgs sector accounts for the remaining 15 parameters: 4 CKM parameters, 6 quark masses, 3 lepton masses, the higgs mass, and the higgs quartic coupling.[3]

One may solve the Euler-Lagrange equations for the Lagrangian of Equation 2.1 in the usual way to find the equations of motion for the fields of the Standard Model. This, however, finds little practical use when computing quantities of interest. Typically, one is interested in computing amplitudes of the kind:

$$\mathcal{M} = \langle f | \, S \, | i \rangle \tag{2.25}$$

where $|f\rangle$ is a final state defined at $t = \infty$, $|i\rangle$ is an initial state defined at $t = -\infty$, and S is an operator which evolves the initial state forward in time from $t = -\infty$ to $t = \infty$. For instance, this is the case at high energy particle colliders such as the Large Hadron Collider, where the initial state may be two high energy protons and the final state may be some channel of interest to an experimentalist. These computations are efficiently performed using the Feynman rules of a theory, which graphically represent terms in a perturbative expansion for such amplitudes. In this thesis, it is not crucial to detail the Feynman rules of the Standard Model (or any other model). These computations are instead abstracted away as a black box by numerous computer programs.

However, it is important to note that problems arise when computing Feynman diagrams containing loops. Such diagrams often lead to divergent integrals, despite appearing initially as higher order terms in a perturbative expansion. These divergences are remedied by recognizing that the Lagrangian has thus far been written in terms of bare parameters and fields.

[3]The parameterization of the Higgs sector is not unique. A full parameterization requires specifying 2 quantities from $\{m_h, \lambda, \mu^2, v\}$.

A bare quantity defines its tree-level value, or its value when loop corrections are not taken into account. Such a quantity cannot be measured as loop diagrams may not be "turned off" in experiment.

Instead, we understand that infinities in the perturbative expansion must be unphysical and write bare quantities as the sum of a physical (renormalized) part and counterterms which cancel these divergent terms. This procedure causes renormalized quantities to attain a dependence on an unphysical mass scale, known as the renormalization scale. Requiring observables to be independent of renormalization scale yields β-functions which describe the dependence of parameters on the scale of the theory.[4] The parameters of the Lagrangians introduced in this thesis should be interpreted as renormalized parameters defined at a fixed mass scale.

2.1.2 Supersymmetry

Supersymmetry (SUSY) is a spacetime symmetry that allows for rotations between fermions and bosons. Such a symmetry is motivated by the hierarchy problem. Loop corrections to the Higgs boson mass cause the appearance of terms that diverge quadratically with the scale of any new physics, where fermions and bosons give opposite sign contributions [29]. Equal quadratically divergent contributions from a fermion and boson are thus a natural solution to the hierarchy problem. We review supersymmetry in order to later introduce the Minimal Supersymmetric Standard Model (MSSM). This discussion draws heavily from References [19, 29, 30].

The simplest supersymmetric Lagrangian contains a Weyl fermion ψ, a complex scalar field ϕ, and a non-propagating auxillary field F, which must be introduced to allow for rotations between ψ and ϕ. This is because ψ contains 4 degrees of freedom off-shell whereas ϕ only

[4] The evolution of parameters with renormalization scale is described as parameter "running." β-functions are also referred to as renormalization group equations.

contains 2, and so F must possess 2 degrees of freedom to account for the difference. The Lagrangian is known as the Wess-Zumino model:

$$\mathcal{L}_{WZ} = F^* F + |\partial_\mu \phi|^2 + i\bar{\psi}\sigma^\mu \partial_\mu \psi. \tag{2.26}$$

Under a supersymmetry transformation parameterized by an infinitesimal 2 Grassmann number valued component object ϵ which transforms as a Weyl spinor under Lorentz transformation, the fields transform as:

$$\delta_\epsilon \phi = \epsilon\psi \tag{2.27}$$

$$\delta_\epsilon \psi_\alpha = -i\left(\epsilon\sigma^\mu\right)_\alpha \partial_\mu \phi \tag{2.28}$$

$$\delta_\epsilon F = -i\bar{\epsilon}\bar{\sigma}^\mu \partial_\mu \psi. \tag{2.29}$$

One can show that $\delta\mathcal{L}_{WZ}$ is given by a total derivative, and so the action is unchanged upon integration over spacetime.

One may also extend this symmetry to vector bosons. To each vector boson A^a_μ, we associate a spin-$\frac{1}{2}$ fermion λ^a and an auxillary field D^a. Note that, because A^a_μ only has 3 off-shell degrees of freedom, D^a must be a real scalar field. A supersymmetric Lagrangian for these fields is given by

$$\mathcal{L} = -\frac{1}{4} F^a_{\mu\nu} F^{\mu\nu}_a + D^2 + i\bar{\psi}\sigma^\mu D_\mu \psi. \tag{2.30}$$

This Lagrangian is invariant up to a total derivative under supersymmetric transformations:

$$\delta_\epsilon A^a_\mu = -\frac{1}{\sqrt{2}}\left(\bar{\epsilon}\bar{\sigma}_\mu\lambda^a + \bar{\lambda}^a\bar{\sigma}^a_\mu\epsilon\right) \tag{2.31}$$

$$\delta_\epsilon \lambda^a_\alpha = -\frac{1}{2\sqrt{2}}\left(\sigma^\mu\bar{\sigma}^\nu\right)_\alpha F^a_{\mu\nu} + \frac{1}{\sqrt{2}}\epsilon_\alpha D^a \tag{2.32}$$

$$\delta_\epsilon D^a = -\frac{i}{\sqrt{2}}\left(\bar{\epsilon}\bar{\sigma}^\mu D_\mu\lambda^a - D_\mu\bar{\lambda}^a\bar{\sigma}_\mu\epsilon\right). \tag{2.33}$$

It is typically more convenient to label related fields as components of a single multiplet (ϕ, ψ, F) or (A_μ, λ, D). This is done more naturally after introducing superspace coordinates. Superspace coordinates augment the 4 typical (commuting) spacetime coordinates t, x, y, z with 4 additional (anti-commuting) coordinates θ_α, $\bar{\theta}_{\dot{\alpha}}$, where $\alpha, \dot{\alpha} = 1, 2$. Fields over this space are known as superfields.

Because θ and $\bar{\theta}$ are anti-commuting, the Taylor expansion of a superfield S over these coordinates terminates after only a few terms:[5]

$$S = \varphi(x_\mu) + \theta\eta(x_\mu) + \bar{\theta}\bar{\chi}(x_\mu) + \bar{\theta}\sigma^\nu\theta V_\nu(x_\mu) + \\ \theta^2 N(x_\mu) + \bar{\theta}^2 M(x_\mu) + \theta^2\bar{\theta}\bar{\alpha}(x_\mu) + \bar{\theta}^2\theta\beta(x_\mu) + \bar{\theta}^2\theta^2\zeta(x_\mu) \tag{2.34}$$

One can study the transformation of the component fields (over spacetime) with respect to SUSY transformations of the coordinates:

$$\theta^\alpha \to \theta^\alpha + \epsilon^\alpha \tag{2.35}$$

$$\bar{\theta}_{\dot{\alpha}} \to \bar{\theta}_{\dot{\alpha}} + \bar{\epsilon}_{\dot{\alpha}} \tag{2.36}$$

$$x^\mu \to x^\mu + i\epsilon\sigma^\mu\bar{\theta} + i\bar{\epsilon}\bar{\sigma}^\mu\theta \tag{2.37}$$

[5]We choose a brave notation that does not differentiate between superfields and fields over spacetime. The differentiation between fields and superfields should be clear from context.

to see that, in particular:

$$\delta_\epsilon \varphi(x_\mu) = \frac{1}{\sqrt{2}}(\epsilon\eta + \bar{\epsilon}\bar{\chi}) \tag{2.38}$$

$$\delta_\epsilon \eta(x_\mu) = \sqrt{2}\epsilon_\alpha N + \frac{1}{\sqrt{2}}(\sigma^\mu \bar{\epsilon})_\alpha (V_\mu - i\partial_\mu \varphi) \tag{2.39}$$

$$\delta_\epsilon N(x_\mu) = \frac{1}{\sqrt{2}}\left(\bar{\epsilon}\bar{\alpha} - \frac{i}{2}\bar{\epsilon}\bar{\sigma}^\mu \partial_\mu \eta\right) \tag{2.40}$$

These are identical to the transformations $\delta_\epsilon\phi, \delta_\epsilon\psi_\alpha, \delta_\epsilon F$ for the multiplet of fields in the Wess-Zumino model under the identifications $(\varphi(x_\mu), \eta(x_\mu), N(x_\mu)) = (\phi(x_\mu), \sqrt{2}\psi(x_\mu), F(x_\mu))$, provided that all fields outside of this multiplet are set to 0. This can be enforced by requiring a superfield Φ to satisfy:

$$\bar{D}_\alpha \Phi = 0 \tag{2.41}$$

where $\bar{D}_\alpha = -\frac{\partial}{\partial\bar{\theta}^{\dot{\alpha}}} + i\sigma^\mu_{\alpha\dot{\alpha}}\bar{\theta}^{\dot{\alpha}}\partial_\mu$. Such Φ are known as chiral superfields, and are the most convenient representation of these multiplets when constructing supersymmetric Lagrangians. Φ takes a particularly nice form under a coordinate transformation $y_\mu = x_\mu - i\bar{\theta}\sigma_\mu\theta$, where now $\Phi(y_\mu) = \phi(y_\mu) + \sqrt{2}\theta\psi(y_\mu) + \theta^2 F(y_\mu)$.

We may also represent the multiplet (A_μ, λ, D) as a superfield. Once again, a general superfield has redundant components. To represent a multiplet containing a vector boson, we enforce the superfield V to be real:

$$V = V^* \tag{2.42}$$

Such superfields are known as vector superfields. From Equation 2.34 we see that, for a vector superfield,

$$\phi(x_\mu) = \phi(x_\mu)^* \tag{2.43}$$

$$\chi(x_\mu) = \eta(x_\mu) \tag{2.44}$$

$$v_\nu(x_\mu) = v_\nu(X_\mu)^* \tag{2.45}$$

$$N(x_\mu) = M(x_\mu)^* \tag{2.46}$$

$$\alpha(x_\mu) = \beta(x_\mu) \tag{2.47}$$

$$\zeta(x_\mu) = \zeta(x_\mu)^* \tag{2.48}$$

With an eye towards the set of fields (A_μ, λ, D) that this vector superfield to represent, we make the definitions:

$$\chi_\alpha = \lambda_\alpha - \frac{i}{2}\left(\sigma^\mu \partial_\mu \bar{\eta}\right)_\alpha \tag{2.49}$$

$$v_\mu = A_\mu \tag{2.50}$$

$$\zeta = \frac{1}{2}D - \frac{1}{4}\partial_\mu \partial^\mu \varphi \tag{2.51}$$

Thus, the vector superfield can be expressed in terms of its component fields:

$$V = \varphi + \theta\chi + \bar{\theta}\bar{\chi} + \bar{\theta}\sigma^\mu\theta A_\mu + \theta^2 N + \bar{\theta}^2 N^* + \bar{\theta}^2\theta\left(\lambda - \frac{i}{2}\sigma^\mu\partial_\mu\bar{\chi}\right) + \theta^2\bar{\theta}\left(\bar{\lambda} - \frac{i}{2}\bar{\sigma}^\mu\partial_\mu\chi\right) + \theta^2\bar{\theta}^2\left(\frac{1}{2}D - \frac{1}{4}\partial^\mu\partial_\mu\varphi\right). \tag{2.52}$$

Similar to the chiral superfield case, one can confirm that the fields (A_μ, λ, D) transform as expected under a SUSY transformation, provided that terms containing φ, χ and N are ignored. The presence of these extra component fields is due to gauge freedom. One may choose a gauge (known as Wess-Zumino gauge) where the fields φ, χ, N disappear. A more complete discussion of gauge transformations for superfields is given at the end of this section.

Superfields provide a compact representation of the matter content in a SUSY theory. We now turn our attention to constructing SUSY Lagrangians from these superfields. We saw in Equation 2.29 that the F term of a chiral superfield transforms as a total derivative under SUSY transformations. It can be shown that any analytic function W of a chiral superfield is again a chiral superfield. So, the F component of an analytic function $W(\Phi)$ is a candidate term for a SUSY Lagrangian, as the action of such a term remains invariant under SUSY transformations. The F component may be extracted by either differentiation or integration twice with respect to θ, as these operations are equivalent for Grassman variables:

$$\mathcal{L} = \int d^2\theta \, W(\Phi) + h.c. \tag{2.53}$$

The function W is known as the superpotential.

Kinetic terms can be attained by extracting the $\theta^2\bar{\theta}^2$ term of a vector superfield given by a real function of chiral superfields, which similarly leaves the action invariant under SUSY transformations. This term is known as the Kahler potential or as a D term, since the $\theta^2\bar{\theta}^2$ of a vector superfield contains the D field. The canonical Kahler potential is given by:

$$K = \int d\theta^2 d\bar{\theta}^2 \Phi^* \Phi \tag{2.54}$$

In order to develop a SUSY theory with interactions mediated by gauge bosons, we must specify the the behavior of superfields with respect to gauge transformations. Chiral and vector superfields transform under gauge transformation according to:

$$\Phi \rightarrow \exp(ig\Lambda)\Phi \tag{2.55}$$

$$e^{gV} \rightarrow e^{ig\Lambda} e^{gV} e^{-ig\Lambda^\dagger}. \tag{2.56}$$

where $V = V_a T^a$, $\Lambda = \Lambda_a T^a$ and Λ_a are chiral superfields. Alternative choices, such as a real-valued function or general superfield, would cause Φ to no longer be a chiral superfield after gauge transformation. One can confirm that the familiar component fields of Φ and V transform in the same way as in non-SUSY gauge theories.

The canonical Kahler potential must be modified in order to maintain gauge invariance. As in non-SUSY theories, the mediating vector bosons are utilized to cancel novel terms that appear due to gauge transformation. The gauge invariant Kahler potential is given by:

$$K = \Phi^* e^{gV} \Phi \tag{2.57}$$

Lastly, we must include kinetic terms for vector superfields. The chiral superfield

$$\mathcal{W}_\alpha = -\frac{1}{4} \bar{D}^2 e^{-V} D_\alpha e^V \tag{2.58}$$

is analogous to the field strength tensor of non-SUSY theories and is referred to as the supersymmetric field strength. The D term of $\mathrm{Tr}[\mathcal{W}^\alpha \mathcal{W}_\alpha]$ leaves the action invariant under SUSY transformations, and serves as a natural candidate to form gauge boson kinetic terms.

2.1.3 The Minimal Supersymmetric Standard Model

The Minimal Supersymmetric Standard Model (MSSM) is the minimal (in terms of matter content) extension of the Standard Model that is supersymmetric. The MSSM proposes new supersymmetric partners for each Standard Model particle seen in the previous section. Additionally, two Higgs doublets are required in order to have consistent electroweak symmetry breaking. The MSSM retains the $SU(3)_c \times SU(2)_L \times U(1)_Y$ gauge symmetry of the Standard Model. The new fields as well as their charges with respect to the $U(1)_Y$ symmetry are given

Chiral Superfields		
Field	Description	Representation $(SU(3)_c,\, SU(2)_L,\, U(1)_Y)$
Q_i	Left-handed quark doublet	$(\mathbf{3},\, \mathbf{2},\, \frac{1}{3})$
$\bar{u}_{R,i}$	Left-handed anti-up quark	$(\bar{\mathbf{3}},\, \mathbf{1},\, -\frac{4}{3})$
$\bar{d}_{R,i}$	Left-handed anti-down quark	$(\bar{\mathbf{3}},\, \mathbf{1},\, \frac{2}{3})$
L_i	Left-handed lepton doublet	$(\mathbf{1},\, \mathbf{2},\, -1)$
$\bar{\ell}_{R,i}$	Left-handed anti-lepton	$(\mathbf{1},\, \mathbf{1},\, -2)$
H_u	Left-handed up Higgs doublet	$(\mathbf{1},\, \mathbf{2},\, 1)$
H_d	Left-handed down Higgs doublet	$(\mathbf{1},\, \mathbf{2},\, -1)$
Vector Superfields		
Field	Description	Representation $(SU(3)_c,\, SU(2)_L,\, U(1)_Y)$
B	$U(1)_Y$ vector superfield	$(\mathbf{1}, \mathbf{1}, 0)$
W^a	$SU(2)_L$ vector superfields	$(\mathbf{1}, \bar{\mathbf{3}}, 0)$
G^a	$SU(3)_c$ vector superfields	$(\bar{\mathbf{8}}, \mathbf{1}, 0)$

Table 2.2: Particle content of the MSSM. Representations under which the fields transform with respect to gauge transformations are given in the third column. For $U(1)_Y$ transformations, the hypercharge is given instead. The index $i = 1, 2, 3$ and runs over the 3 generations of fermions, which differ only in mass.

in Table 2.2. The Lagrangian is composed of the superfield kinetic terms introduced in the previous section, the superpotential, and a SUSY breaking sector.

For the fields of the MSSM, the most general renormalizable gauge invariant superpotential is given by:

$$W_{RPV} = \mu H_u H_d + Y_u^{ff'} H_u Q_f \bar{u}_{R,f'} + Y_d^{ff'} H_d Q_f \bar{d}_{R,f'} + Y_L^{ff'} H_u L_f \bar{\ell}_{R,f'} + \lambda_{ijk} L_i L_j \ell_{R,k} + \lambda'_{ijk} Q_i L_j \bar{d}_{R,k} + \lambda''_{ijk} \bar{u}_{R,i} \bar{d}_{R,j} \bar{d}_{R,k} + m_i L_i H_u. \quad (2.59)$$

The last 4 terms in this expression are problematic as they lead to tree-level proton decay. Proton lifetimes have been determined experimentally to be very large, and so these couplings must either be 0 or extremely suppressed. In most discussions of the MSSM, as well as those undertaken in this thesis, we impose a Z_2 symmetry known as R-parity. R-parity can be understood as mapping particle fields $\psi \to +\psi$ and sparticle fields $\tilde{\psi} \to -\tilde{\psi}$. This symmetry

causes the problematic terms of Equation 2.59 to dissapear, leaving:

$$W_{MSSM} = \mu H_u H_d + Y_u^{ff'} H_u Q_f \bar{u}_{f'} + Y_d^{ff'} H_d Q_f \bar{d}_{f'} + Y_L^{ff'} H_u L_f \bar{e}_{f'}. \tag{2.60}$$

Imposing R-parity has significant phenomenological implications. First, we remark that each term in the Lagrangian must have an even number of superpartners. Thus, production of SUSY particles at experimental colliders must occur in pairs. Additionally, a SUSY particle undergoing decay must produce at least one SUSY particle in the final state. Consequently, the lightest supersymmetric particle (LSP) must be stable, as there is no lighter supersymmetric particle that may appear in the final state. Due to its longterm stability, the LSP is a natural dark matter candidate.

It is clear that supersymmetry must be broken at low energies, as sparticles have yet to be detected in experiment, though unbroken SUSY would yield sparticles at the same mass as their easily detected SM counterpart. Supersymmetry may be broken either spontaneously or explicitly. The MSSM explicitly breaks supersymmetry by including all non-SUSY invariant terms that are invariant under transformations of the MSSM gauge group and that do not reintroduce quadratic divergences to the Higgs mass.[6] Terms that do not introduce quadratic divergences to the masses of scalar particles are termed "soft", and hence this sector of the Lagrangian is known as the soft SUSY breaking Lagrangian.

$$\begin{aligned}
\mathcal{L}_{soft} = &-\frac{1}{2}\left(M_1 \tilde{B}\tilde{B} + M_2 \tilde{W}^a \tilde{W}_a + M_3 \tilde{g}^a \tilde{g}_a\right) \\
&- \left(A_u^{ff'} \tilde{Q}_f h_u \tilde{\bar{u}}_{f'} + A_d^{ff'} \tilde{Q}_f h_d \tilde{\bar{d}}_{f'} + A_e^{ff'} \tilde{\ell}_f h_d \tilde{\bar{e}}_{f'}\right) \\
&- \left(\tilde{Q}_f^\dagger m_{q,ff'} \tilde{Q}_{f'} + \tilde{\ell}_f^\dagger m_{\ell,ff'} \tilde{\ell}_{f'} + \tilde{\bar{u}}_f^\dagger m_{u,ff'} \tilde{\bar{u}}_{f'} + \tilde{\bar{d}}_f^\dagger m_{d,ff'} \tilde{\bar{d}}_{f'} + \tilde{\bar{e}}_f^\dagger m_{e,ff'} \tilde{\bar{e}}_{f'}\right) \\
&\qquad - \left(m_{h_u}^2 |h_u|^2 - m_{h_d}^2 |h_d|^2 - B_\mu h_u h_d\right) + h.c.
\end{aligned} \tag{2.61}$$

[6] The MSSM in this form is also sometimes referred to as the unconstrained MSSM, or uMSSM.

We can clearly see that these terms are not invariant under SUSY transformation as they are written in terms of sparticle fields rather than superfields. The first line provides mass terms to the gauginos. The second line describes trilinear interactions between sparticle fields. The third line provides masses and mixings to sfermions fields. The last line describes masses and mixings between Higgs bosons.

The bino, wino, and higgsino fields are not mass eigenstates of the theory. Instead, they form a multiplet $\psi^0 = (\tilde{B}, \tilde{W}, \tilde{h}_d^0, \tilde{h}_u^0)$ with mass terms given by -$\frac{1}{2}$ $(\psi^0)^\dagger$ $M_{\chi^0}\psi^0 + h.c.$ where

$$M_{\chi^0} = \begin{pmatrix} M_1 & 0 & -g_1 v_d/\sqrt{2} & g_1 v_u/\sqrt{2} \\ 0 & M_2 & g_2 v_d/\sqrt{2} & -g_2 v_u/\sqrt{2} \\ -g_1 v_d/\sqrt{2} & g_2 v_d/\sqrt{2} & 0 & -\mu \\ g_1 v_u/\sqrt{2} & -g_2 v_u/\sqrt{2} & -\mu & 0. \end{pmatrix} \tag{2.62}$$

The off-diagonal higgsino-bino and higgsino-wino couplings result from electroweak symmetry breaking and derive from Higgs-higgsino-gaugino interactions. The mass matrix is Hermitian and may thus be diagonalized by a single matrix, which can then be absorbed into the definition of ψ^0 to yield mass eigenstates.

The mass eigenstates of this matrix are the neutralinos, which form from a superposition of bino, wino, and higgsino states. They are referred to by their mass ordering χ_i^0 where $i = \{1, 2, 3, 4\}$. The neutralinos are electrically neutral and χ_1^0 is often the lightest sparticle in the spectrum and thus serves as a promising cold dark matter candidate. In large regions of the parameter space, the neutralinos take on nearly pure states. In such cases, the neutralino is often referred to as its primary component. The charged wino and charged higgsino fields are similarly not mass eigenstates of the theory. Instead, these states combine to form a multiplet $\psi^\pm = (\tilde{W}^+, \tilde{h}_u^+, \tilde{W}^-, \tilde{h}_d^-)$ which is then diagonizalized to form two mass eigenstates, referred to as charginos.

A careful accounting of all new parameters introduced by the MSSM shows that 124 parameters are required in order to make quantitative predictions [31]. The MSSM inherits 18 parameters from the Standard Model and contains 1 parameter analogous to the SM Higgs mass. This leaves 105 genuinely novel parameters. This high dimensionality prohibits extensive study of the parameter space.

Some of the terms in Equation 2.61 cause phenomenologically unviable processes at tree level. The new phases of the theory cause novel CP violating processes. Off diagonal elements in mass matrices and trilinear couplings introduce flavor changing neutral currents. Additionally, mass differences between first and second generation sparticles cause large kaon mixing. All of these processes are heavily constrained by experiment. The phenomenological MSSM (pMSSM) enforces these experimental constraints by restricting to the subspace of the MSSM where the coefficients of these terms are set to 0 [32]. This reduces the space to 19 novel parameters given in Table 2.3.

One can still account for theories of spontaneously broken supersymmetry since Equation 2.61 contains all possible terms that may break SUSY. A theory of spontaneously broken supersymmetry will simply constrain the parameters appearing in these terms. The constrained MSSM (cMSSM) is a theory where SUSY breaking is mediated via gravitational interactions between MSSM fields and a hidden sector of fields which break SUSY and are singlets with respect to the MSSM gauge group.[7] The hidden sector is posited to be insensitive to MSSM gauge quantum numbers so that the soft Lagrangian only depends on a common gaugino mass, a common sfermion mass, and a common trilinear coupling. These universal parameters, along with the higgsino mixing parameter, form the cMSSM parameter space. The theory may then be reparameterized in terms of the ratio of Higgs vacuum ex-

[7]The cMSSM may refer to multiple theories of supersymmetry breaking which yield identical parameter spaces. Most commonly discussed is minimal supergravity (mSUGRA)[33, 34, 35, 36], but also included is gauge mediated supersymmetry breaking [37, 38, 39] and anomaly mediated supersymmetry breaking.[40, 41] In this thesis, we use cMSSM to refer to the mSUGRA scenario.

Parameter	Description
$\lvert M_1 \rvert$	Bino mass
$\lvert M_2 \rvert$	Wino mass
M_3	Gluino mass
$\lvert \mu \rvert$	Bilinear Higgs mass
$\lvert A_t \rvert$	Trilinear top coupling
$\lvert A_b \rvert$	Trilinear bottom coupling
$\lvert A_\tau \rvert$	Trilinear τ coupling
M_A	Pseudo-scalar Higgs mass
$m_{\tilde{L}_1}$	1st gen. l.h. slepton mass
$m_{\tilde{e}_1}$	1st gen. r.h. slepton mass
$m_{\tilde{L}_3}$	3rd gen. l.h. slepton mass
$m_{\tilde{e}_3}$	3rd gen. r.h. slepton mass
$m_{\tilde{Q}_1}$	1st gen. l.h. squark mass
$m_{\tilde{u}_1}$	1st gen. r.h. u-type squark mass
$m_{\tilde{d}_1}$	1st gen. r.h. d-type squark mass
$m_{\tilde{Q}_3}$	3rd gen. l.h. squark mass
$m_{\tilde{u}_3}$	3rd gen. r.h. u-type squark mass
$m_{\tilde{d}_3}$	3rd gen. r.h. d-type squark mass
$\tan\beta$	Ratio of Higgs VEVs

Table 2.3: Parameters of the pMSSM. "Left-handed" and "right-handed" are abbreviated by l.h. and r.h., respectively.

Parameter	Description
m_0	Universal sfermion mass
$m_{1/2}$	Universal gaugino mass
A_0	Universal trilinear coupling
$\tan\beta$	Ratio of Higgs VEVs
$sign(\mu)$	Sign of higgsino mixing parameter

Table 2.4: The parameters of the cMSSM. The parameter $sign(\mu) = \{-1, 1\}$ is discrete. The remaining parameters are continuous.

pectation values (VEVs) and the sign of the higgsino mixing parameter. This is summarized in Table 2.4.

2.1.4 Simplified Models

Simplified models attempt to present an effective description of physics at a relatively low energy scale in terms of relatively few parameters [42]. There is no illusion that such a model is a fundamental description of physics. In fact, simplified models are often the limit of a fully developed BSM theory, such as the MSSM, where many particles and interactions have been integrated out.

There are multiple benefits to using simplified models as opposed to fully developed BSM theories. Multiple BSM theories may have the same simplified model limit, allowing simplified model studies to cover a multitude of BSM scenarios. Simplified model results are thus said to be “model independent", or independent of the details of any fundamental, higher energy theoretical description.

An additional advantage of simplified models is that, by design, the parameter spaces are low dimensional, containing only the operators and fields that are immediately relevant to the process described by the model. As such, explorations of simplified model phenomenology are computationally cheap and limits are relatively easy to set and communicate when compared to high dimensional parameter spaces.

An example of a simplified model is presented in the next chapter, where we consider a simplified model that describes the addition of a heavy Z' boson to the Standard Model. The Z' boson is posited to have identical quantum numbers to the SM Z boson with larger mass. Such a boson often arises in BSM models [43]. All other particles and interactions of the fundamental theory are assumed to be negligible, so that the only free parameters of the theory are the mass of the new boson and its coupling to SM fermions.

2.2 Machine Learning

2.2.1 Supervised Learning

Assume a set of m pairs $\{(x_i, f(x_i))\}$ is given, where $x_i \sim p(x)$. In general, $x_i \in \mathbb{R}^n$ is referred to as a feature vector and $f(x_i) \in \mathbb{R}^d$ is referred to as the target. The target vector may be continuous or discrete, and the mapping f may be extremely complicated or have no closed form available. The set of $\{(x_i, f(x_i))\}$ pairs is known as the training set. The goal of supervised learning is to construct a map $\hat{f} : \mathbb{R}^n \to \mathbb{R}^d$ such that $\hat{f}(x_i) = f(x_i)$. Moreover, $\hat{f}$ is constructed to generalize well to samples outside of the initial set, so that $\hat{f}(x)$ reasonably approximates $f(x)$ for arbitrary $x \sim p(x)$.

Supervised learning can generally be split into two types of problems: regression and classification. Regression problems attempt to model continuous y whereas classification problems attempt to predict binary (0 or 1) or categorical y. We will review supervised learning in four parts: model definition, optimization, evaluation, and regularization. Importantly, though, these are often logically coupled, and so a linear reading may require revisiting and refining previously introduced concepts. Much of the discussion follows the references [44, 45].

Model Definition

Model definition refers to specifying the form of $\hat{f}$. The simplest example of a machine learning model is nearest neighbor regression, which evaluates $\hat{f}(x) = f(x_j)$, where x_j is the closest point in the training set to x. This is easily generalized to k nearest neighbors regression, where $\hat{f}(x) = \langle f(x_j) \rangle$ and the average is taken over x_j in the set of k points closest to x in the training set.

Machine learning models can be made more expressive by introducing parametric dependencies to $\hat{f}$. The simplest example of such a model is linear regression, where the output is a linear function of x:

$$\hat{f}(x) = w^T x + b. \tag{2.63}$$

The vector w and b are parameters of the model, which may also be referred to as the weights and bias, respectively.[8] For convenient notation, we will assume the feature vector x is augmented with a 1, so that the bias may be included in w. It is clear that the quality of the machine learning model depends crucially on the choice of w. In the next part of this section, we review a method for choosing the optimal weights. The analogous linear model for binary classification problems is logistic regression, which includes a sigmoid function to achieve outputs in the range $(0, 1)$ i.e. $\hat{f}(x) = \sigma(w^T x)$, where:

$$\sigma(z) = \frac{1}{1 + e^{-z}}. \tag{2.64}$$

Linear models are limited by their inability to capture nonlinear dependencies on the features. While the feature vector may be augmented to include nonlinear transformations of features,

[8] A more explicit notation would include the parametric dependence of $\hat{f}$ on the weights and bias i.e. $\hat{f}(x; w, b)$, though this dependence will be left as understood.

this introduces the need for engineering relevant features and increases model complexity which may cause issues with generalization.

Instead, one can choose to employ a nonlinear model. Nonlinear regression models include support vector machines, kernel ridge regression, and gaussian process regression among others. These make use of the kernel trick, where x is implicitly projected into a higher (potentially infinite) dimensional space and inner products between x and x_i in the training set are used in the prediction $\hat{f}$.

Neural networks are an alternative to these methods which scale well with the number of input data. Neural networks transform the feature vector by iteratively performing matrix multiplication and function evaluation. That is:

$$\hat{f}(x) = a(W_D a(W_{D-1} a(...a(W_1 x)))) \tag{2.65}$$

where W_i are real matrices of weights and a is a function, known as the activation function. The feature vector x is also referred to as the input layer and the full evaluation is referred to as the output layer. Applications of a in between these two layers are known as hidden layers. D is referred to as the depth of the neural network. Neural networks with more than 1 hidden layer are typically called deep networks.

Neural networks are often introduced diagrammatically as in Figure 2.1, where the relevant parts have been labelled. Common choices for the activation function are sigmoid, softmax, rectified linear unit (ReLU). It is important to note that a must be nonlinear in order to be able to model nonlinear dependencies on the features.

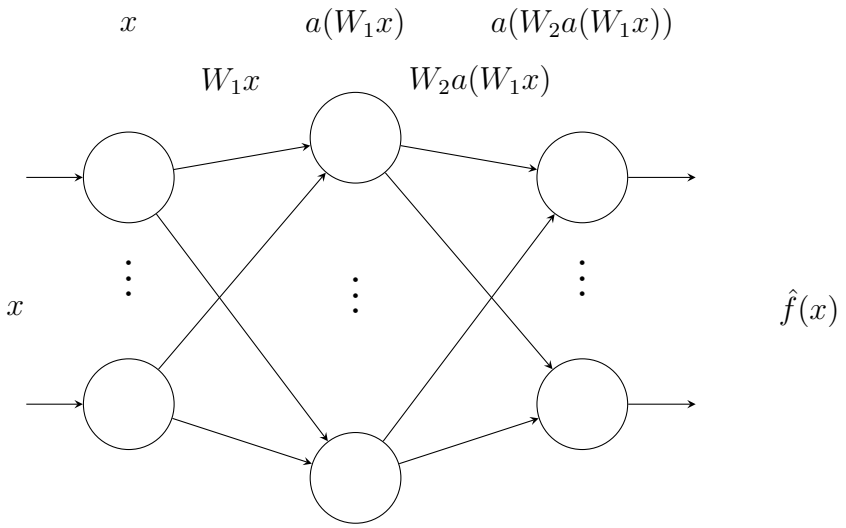

Figure 2.1: A diagram of a neural network.

Name	Problem Type	Function
L2	Regression	$\mathcal{L} = \frac{1}{2m}\sum_{i=1}^{m}\left(\hat{f}(x_i) - f(x_i)\right)^2$
L1	Regression	$\mathcal{L} = \frac{1}{m}\sum_{i=1}^{m}\left\lvert\hat{f}(x_i) - f(x_i)\right\rvert$
Cross Entropy	Classification	$\mathcal{L} = -\frac{1}{m}\sum_{i=1}^{m} y_i^T \log\left(\frac{e^{-f(x_i)}}{\sum_{j=1}^{C}\left(e^{-f(x_i)}\right)_j}\right)$
Cross Entropy	(Binary) Classification	$\mathcal{L} = -\frac{1}{m}\sum_{i=1}^{m}\Big[f(x_i)\log(\hat{f}(x_i)) + (1 - f(x_i))\log(1 - \hat{f}(x_i))\Big]$

Table 2.5: Common loss functions for supervised learning problems and their different use cases. For multi-class cross entropy loss with C classes, y_i is assumed to be one-hot encoded; it is a C dimensional vector that is 1 for the correct class and 0 elsewhere.

Model Optimization

As seen in linear regression and neural networks, the performance of many supervised learning models is critically dependent upon the weights. The process of determining the optimal model weights is known as training. To determine the weights of a model, we first define a loss function $\mathcal{L}$ to quantify how poorly a model performs on the given training set. Common choices for loss functions are summarized in Table 2.5. Once the performance of the model has

been meaningfully quantified, the learning problem is then transformed into an optimization problem. That is, the weights of the model are chosen so that the loss function is minimized.

Loss functions may be minimized either analytically (where available) or numerically. Analytical optimizations may be performed using calculus of variations, which provide closed form solutions to the equations $\frac{\partial \mathcal{L}}{\partial w} = 0$. Solutions exist for linear regression and kernel ridge regression, however, the implementation of analytical solutions typically scale poorly with the amount of training data. Additionally, many models do not have analytical solutions available, such as SVMs and neural networks.

Numerical algorithms are more commonly used to optimize machine learning models. These compute derivatives of the loss function to inform updates to the weights. This is iterated until the model reaches a global or local minimum, or a certain number of iterations is reached. The simplest numerical optimization algorithm is gradient descent, which updates the weights according to:

$$w \leftarrow w - \alpha \frac{\partial \mathcal{L}}{\partial w} \tag{2.66}$$

Gradient descent steps in the direction of the negative gradient until a minimum in the loss function is reached. In this equation, α is a hyperparameter, meaning that it is an input to the model which must be provided beforehand. If α is too small, convergence to the minimum may be very slow, and the model is more likely to become stuck in a local minimum. If α is too large, the model will continuously skip beyond the minimum and the loss will diverge. A detailed method of selecting hyperparameters is given in the portion on regularization.

Many modifications can be made to gradient descent. A common variant of gradient descent is stochastic gradient descent [46], where instead of calculating the derivative of the loss on

the entire training set, only a batch of training set samples is used at a time.[9] This tends to increase the convergence rate of the model, but the final trained model will tend to be slightly displaced from the minimum evaluated on the full training set. Additionally, one can include higher order derivative information, such as the Hessian, in order to incorporate curvature information into the optimization steps [47].

One can also include a momentum term in the weight updates. This results in faster progress towards the minimum as momentum builds up through consecutive steps in the same direction. It also allows the model to coast through flat regions of the loss function and escape from shallow, non-optimal local minima. Finally, one can perform adaptive adjustments to the learning rate over the course of the optimization. Common numerical optimization algorithms include gradient descent, stochastic gradient descent, RMSProp [48], Adadelta [49], Adamax [50], Adam [50], and AMSGrad [51], with Adam being an extremely popular choice for training neural networks.

When optimizing neural networks, all weights cannot be updated simultaneously with a single derivative evaluation. Instead, weights are updated beginning with the output layer and updates are "propagated" back until the input layer is reached. Hence, the algorithm for numerical optimization of neural networks is known as backpropagation [52].

Model Evaluation

Once a model has been chosen and trained, its performance must be evaluated to determine if the training process was successful. A first method of measuring the performance of a model is by plotting the average loss computed on the training set as a function of complete

[9] Some reserve "stochastic gradient descent" for the case where the batch size is 1 and instead use the name "minibatch gradient descent" for batch sizes larger than 1. For this thesis we will refer to both of these cases as stochastic gradient descent.

iterations through the training data, which is known as the learning curve. The average loss is expected to decrease as training continues.

Low training loss alone is not a good performance metric. This is because good performance on the training data may not always imply good performance on data outside of the training set. In fact, since the weights of a model are determined by measuring performance on the training set, the model will tend to perform better on the training set than on out of sample data. To measure the performance of the model on data outside of the training sample, a fraction of the initial set of data points, known as the test set, is set aside and kept hidden during training. Following training, the average loss of the model evaluated on the test set may be evaluated to infer performance on unseen data. One may also evaluate the testing loss at each epoch to plot a learning curve for the test data.

We have already seen that many models require hyperparameters to be input before predictions may be made. Examples include the learning rate α and the size and width of layers in a neural network. Additional sources of hyperparameters will be introduced in the next section on regularization. The set of hyperparameters form a low dimensional space which typically does not have readily available gradients. A tempting way to determine the best choice of hyperparameters is to randomly sample this space to find the configuration with the lowest testing error. However, this will cause the model to overestimate its ability to generalize. Just as the parameters of the model are fit to the training set and so the model typically outperforms on the training set, the hyperparameters in such a case would be fit to the testing set, and so the model will outperform on the testing set.

A solution is to instead introduce a cross validation set. This is a subset of the training set that is withheld during the optimization of weights and is only used to determine the best choice of hyperparameters. The optimization of the cross validation error within the space of hyperparameters may be performed either by grid search, random scan, or other derivative-free optimization algorithms, such as genetic algorithms [53] or bayesian optimization [54].

This is known as hold out cross validation. More involved cross validation procedures, such as k folds cross validation, repeated k folds cross validation, and nested k folds cross validation aim to increase the data efficiency of this process. In typical deep learning applications, hold out cross validation is often preferred as the computational cost of alternative cross validation methods far outweighs the benefits of increased data efficiency.

For regression problems, low mean squared error or mean absolute error on the test set is typically sufficient to determine whether a supervised model has been successfully trained. For classification problems, the cross entropy loss is typically used as a performance metric because it is sensitive not only to the accuracy of a prediction, but also the confidence in that prediction. It is often also advantageous to analyze elements of the confusion matrix and functions of these elements, depending on the intended use of the classifier. These include ROC curves, precision recall curves, AUC, and F score among others [55].

Regularization

Thus far, the training procedure defines the best model as that which optimizes a performance measure on the training set. However, the goal of supervised learning includes the ability for models to generalize well to unseen data. The optimization of weights with no regard towards ability to generalize often causes the trained model to perform very well on the training set but generalize poorly to data in the testing set. This is known as overfitting, or the model is said to have high variance. Regularization of a machine learning model is any attempt to increase the ability of the model to generalize to data beyond the training set, potentially at the expense of performance on the training set.

One method of regularization is to include an additional term in the loss function to incentivize the model to choose smaller weights, as smaller weights will tend to produce smoother functions. For instance, instead of optimizing a loss function described in Table 2.5, one may

optimize $\mathcal{L}_{regularized} = \mathcal{L} + \lambda w^T w$. This type of regularization is known as weight decay. One may also choose to use an L1 norm instead: $\mathcal{L}_{regularized} = \mathcal{L} + \lambda \sum_i |w_i|$.[10] The regularization parameter λ is a hyperparameter that must be input to the model. Very large λ will cause all weights to tend toward 0. Such models will underfit the training data, or will fail to capture variance in the training set and would be said to have high bias. Very small λ will only minimally regularize the model and would tend to overfit the training data.

Alternative methods of regularization are model specific. For instance, one may randomly exclude neurons in a neural network while proportionally scaling the output of surviving neurons, as in dropout [56]. Batch normalization is also believed to have a regularizing effect on neural networks [57].

Finally, one may regularize a model by altering the training procedure. One example of this is data augmentation, where perturbations or symmetric transformations of training data are performed in order to augment the training set. Another is early stopping, where cross validation error is computed at each epoch of training, and the model from the epoch with lowest cross validation error is selected as the final model.

[10] For both regularized loss functions, the bias term is not included in the sum over weights.

Chapter 3

Resonance searches with Machine Learned Likelihood Ratios

This chapter is heavily based on work previously published in collaboration with Daniel Whiteson [58].

3.1 Introduction

A primary focus of the physics program at the Large Hadron Collider is the search for beyond the standard model (BSM) physics. Many BSM models predict the existence of heavy, short lived particles that rapidly decay to familiar standard model particles, leaving a telling experimental signature known as a resonance. Historically, the discovery of new resonances has revolutionized our understanding and confidence in models of particle physics [59, 60, 61, 62, 63, 64, 65]. A similar discovery made in the current age would likely have a similar impact.

Many searches for BSM resonances involve fitting signal and background spectra in the invariant mass of the final state particles, allowing one to construct a likelihood ratio [66, 67, 68, 69]. Recently, this has been greatly improved by utilizing machine learning at various steps in the process [70, 71]. However, relying solely on the invariant mass neglects a great deal of the available information when determining the likelihood ratio of an event. The full event information is given by the set of four-momenta of every particle in the final state. Summarizing this relatively high dimensional information with invariant mass plausibly results in a significant amount of information loss.

Certain methods have been developed that aim to overcome this information loss [72, 73, 74, 75, 76, 77]. The matrix-element method searches for new particles by approximating the detector response with a simple transfer function and marginalizing matrix elements over the unseen detector degrees of freedom [78, 79, 80]. This approach uses multivariate detector level output in determing the likelihood ratio, but critically relies on the choice of transfer functions and can often be very computationally expensive to perform.

In this chapter, we implement a new method of analysis, originally applied to effective field theories (EFTs), which machine learns a likelihood ratio from latent information that is extracted from simulations [81, 82, 83, 9]. Similar to the matrix element method, this new method utilizes multivariate output of a detected event and so it is expected to provide similar improvements in performance. However, it removes the need to approximate the detector using transfer functions, and thus may be viewed as a direct improvement over the matrix element method.

Transferring the methods of Refs. [81, 82, 83, 9] from EFTs to resonance searches is non-trivial for several reasons. First, EFTs have a morphing structure, allowing squared matrix elements to be written as simple polynomials of the parameters of the theory [84, 81]. Consequently, by evaluating the squared matrix element of an event at a handful of "benchmark" points in the parameter space, one is able to infer the squared matrix element for arbitrary

theory parameters, affording significant improvements in computational efficiency. However, resonanace searches do not have a complete morphing structure, as the squared matrix element typically has a non-polynomial dependence on mass. As such, the squared matrix element of an event must be evaluated at every point in the parameter space.

Additionally, the most successful methods for EFTs rely on using the derivative of the log likelihood ratio with respect to the theory parameters, known as the score, to successfully train the machine learning models. Scores are readily available in EFT calculations, but must be numerically calculated for resonance searches, which greatly increases the computational cost of the analysis. As a result, we are constrained to use the less sample efficient methods, which do not rely on this information during training. This work is the first to show that these difficulties can be overcome, and that these methods may still provide substantial improvements over traditional methods beyond the context of EFTs.

In Section II, we review the theory behind performing a resonance search and review the new method that we use to calculate the likelihood ratio. In Section III, we will introduce a simple BSM model and detail how the new framework will be used to search for a resonance in the model. In Section IV, we show the results of our resonance search. In Section V, we discuss the implications of this work and possible future directions.

3.2 Method

Let θ denote a set of theory parameters. Using a standard suite of programs, we can produce a set of simulated events $\mathcal{X} = \{x \sim p(x|\theta)\}$, where x is a random variable of detector level observables and is used interchangably as a realization of this random variable. However, the inverse problem of determining which θ produced a set of events $\mathcal{X}$ is extremely difficult. The Neyman-Pearson lemma shows that the optimal discriminator between two competing

theories, parameterized by θ and θ_0 respectively, is the likelihood ratio:

$$r(x|\theta,\theta_0) = \frac{p(x|\theta)}{p(x|\theta_0)} = \frac{\int dz\, p(x|z)p(z|\theta)}{\int dz\, p(x|z)p(z|\theta_0)}, \tag{3.1}$$

where z is the latent parton level four momenta of all particles in the final state, which is unobservable in experiment.

For typical showering and detector simulations, $p(x|z)$ is intractible due to the extremely large number of ways an event may shower and be detected. For this reason, $r(x|\theta,\theta_0)$ is typically calculated by attempting to approximate $p(x|\theta)$ and $p(x|\theta_0)$ directly, using histograms of x. The number of data points required to adequately populate the bins of the histogram scales exponentially with the dimension of x, which is known as the curse of dimensionality. As a result, this approach becomes impractical as the dimension of x becomes moderately large. As such, histograms usually are constructed in only one or two summary statistics of x, which may result in information loss relative to higher dimensional approaches.

A recent study with effective field theories has offered an alternative method of calculating $r(x|\theta,\theta_0)$ as a function of the detector level output [81]. The remainder of this section follows the discussions in Refs. [81, 82, 83, 9] very closely. We first consider the joint likelihood ratio:

$$r(x,z|\theta,\theta_0) = \frac{p(x|z)p(z|\theta)}{p(x|z)p(z|\theta_0)} = \frac{p(z|\theta)}{p(z|\theta_0)}. \tag{3.2}$$

All intractable parts of the joint likelihood ratio cancel, and for simulated events, we can calculate $r(x,z|\theta,\theta_0)$ in terms of tractible matrix elements as

$$r(x,z|\theta,\theta_0) = \frac{\sigma(\theta)^{-1}|\mathcal{M}(z|\theta)|^2}{\sigma(\theta_0)^{-1}|\mathcal{M}(z|\theta_0)|^2}, \tag{3.3}$$

where σ gives the cross section as a function of theory parameters and $\mathcal{M}$ gives the matrix element as a function of parton level momenta and theory parameters. The joint likelihood ratio cannot be calculated for observed data, as it requires knowledge of z which is not well defined in experiment. However, in the following paragraphs, we demonstrate that machine learning the joint likelihood ratio of simulated events will result in the true likelihood ratio in Eq. (3.1) under a specific set of conditions.

Consider a function $\hat{r}(x|\theta,\theta_0)$ that attempts to predict $r(x,z|\theta,\theta_0)$ given only information of x. We can quantify the error of the function when evaluated on a set of data $(x,z) \sim p_{train}(x,z)$ with the functional:

$$\mathcal{L}[\hat{r}] = \int\int dx\, dz\, p_{train}(x,z)\, |\hat{r}(x|\theta,\theta_0) - r(x,z|\theta,\theta_0)|^2 \,. \tag{3.4}$$

Optimizing this loss requires varying the functional with respect to $\hat{r}(x|\theta,\theta_0)$:

$$\frac{\delta\mathcal{L}}{\delta\hat{r}} = 2\int dz\, p_{train}(x,z)\, (\hat{r}(x|\theta,\theta_0) - r(x,z|\theta,\theta_0)) \tag{3.5}$$

$$= 2p_{train}(x)\hat{r}(x|\theta,\theta_0) - 2\int dz\, p_{train}(x,z) r(x,z|\theta,\theta_0). \tag{3.6}$$

The loss is extremized by setting this variation to 0, yielding:

$$\hat{r}(x|\theta,\theta_0) = \frac{1}{p_{train}(x,z)} \int dz\, p_{train}(x,z) r(x,z|\theta,\theta_0). \tag{3.7}$$

This result holds for an arbitrarily sampled training set. The utility of this solution is clear when we choose the training distribution so that $p_{train}(x,z) = p(x,z|\theta_0)$. When the data is

sampled with this distribution, Equation 3.7 greatly simplifies:

$$\hat{r}(x|\theta,\theta_0) = \frac{1}{p(x|\theta_0)} \int dz\, p(x,z|\theta_0) \frac{p(x,z|\theta)}{p(x,z|\theta_0)} \tag{3.8}$$

$$= \frac{p(x|\theta)}{p(x|\theta_0)} \tag{3.9}$$

Thus, using a deep neural network to represent $\hat{r}(x|\theta,\theta_0)$, we can use standard optimization techniques employed in machine learning to train an estimator that will converge to the familiar likelihood ratio of Equation 3.1 in the limit of infinite data, an infinitely large neural network, and perfect loss optimization. In realistic implementations, deviations from the true likelihood ratio may occur due to the effect of finite datasets, finite neural networks, and inefficient optimization. This likelihood depends only on x, and so it can be evaluated for simulated and observed events alike.

We can use the joint likelihood ratio to construct alternative, more complicated loss functionals that similarly converge to the true likelihood ratio. A summary of these methods as well as their different properties is found in the references [81, 83]. In this work, we find the $\hat{s}(x|\theta,\theta_0)$ that minimizes the ALICE (approximate likelihood with improved cross-entropy estimator) loss functional, which is given by:

$$\mathcal{L}[\hat{s}] = -\int\int dx\, dz\, p(x,z|\theta_0) \big[s(x,z|\theta,\theta_0) \log(\hat{s}(x|\theta,\theta_0)) + (1 - s(x,z|\theta,\theta_0)) \log(1 - \hat{s}(x|\theta,\theta_0)) \big], \tag{3.10}$$

where we have defined

$$s(x,z|\theta,\theta_0) = \frac{1}{1 + r(x,z|\theta,\theta_0)}. \tag{3.11}$$

With $\hat{s}(x|\theta,\theta_0)$, we can form an estimator for the likelihood ratio

$$\hat{r}(x|\theta,\theta_0) = \frac{1-\hat{s}(x|\theta,\theta_0)}{\hat{s}(x|\theta,\theta_0)}, \tag{3.12}$$

which similarly converges to the true likelihood ratio. The minimization is performed over a balanced training set, with an equal number of events drawn from $p(x|\theta)$ and $p(x|\theta_0)$.

Given the likelihood ratio of each event, the likelihood ratio over $\mathcal{X}$ is trivially found:

$$r(\mathcal{X},\theta,\theta_0) = \prod_{x\in\mathcal{X}} r(x|\theta,\theta_0) \tag{3.13}$$

There are a handful of ways to diagnose mismodelling and poor convergence. Errors that result from simulations mismodelling the physical process can be addressed by introducing nuisance parameters and profiling over them [9, 85]. Errors due to poor convergence of the neural network can be addressed by increasing network complexity, increasing the training set size, or tuning learning parameters. Among other methods, poor convergence can be diagnosed by extensively sampling many $p(x|\theta)$ and $p(x|\theta_0)$ and visually comparing the distributions $p(x|\theta)$ and $\hat{r}(x|\theta,\theta_0)p(x|\theta_0)$. However, in higher dimensions, it may be necessary to compare distributions of summary statistics instead [81, 85].

Built into this approach, as well as many standard approaches, is the assumption that the dataset $\mathcal{X}$ is accurately modeled by a simulated distribution $p(x|\theta)$, where θ is possibly augmented by nusiance parameters. As with all simulation-based methods, inaccuracies in the model will lead to systematic uncertainties in the results. We also highlight the assumption that parton level four momenta z can be associated to each event x in the dataset. This assumption is satisfied for Monte Carlo derived background estimates, but would be difficult to satisfy with fully data driven background estimates.

3.3 A Toy Problem

We study a simple model that allows for analytical solutions in order to better present the concepts of the previous section. Consider constructing $p(x|\mu)$ from the distributions:

$$p(x|z) = \frac{1}{\sqrt{2\pi}\sigma_s} e^{-\frac{(x-z)^2}{2\sigma_s^2}} \tag{3.14}$$

$$p(z|\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2}}. \tag{3.15}$$

To complete the analogy to experiments at high energy colliders, $p(z|\mu)$ may model a hard scattering process and the distribution $p(x|z)$ models smearing due to detector effects. For this example, the integral for $p(x|\mu)$ is tractable

$$p(x|\mu) = \int dz\, p(x|z)p(z|\mu) \tag{3.16}$$

$$= \sqrt{\frac{1}{2\pi(\sigma_s^2+1)}} e^{-\frac{(x-\mu)^2}{2(\sigma_s^2+1)}}, \tag{3.17}$$

and so the likelihood ratio may be written

$$r(x|\mu,\mu_0) = e^{\frac{2(\mu-\mu_0)x-(\mu^2-\mu_0^2)}{2(\sigma_s^2+1)}}. \tag{3.18}$$

We now introduce a parameterized model $\hat{r}(x|\theta,\theta_0)$ which attempts to predict the joint likelihood $\frac{p(x|z)p(z|\mu)}{p(x|z)p(z|\mu_0)}$ with only knowledge of x. We choose the model $\hat{r}(x|\theta,\theta_0) = e^{ax+b}$ and optimize the loss

$$\mathcal{L}[\hat{r}] = \int\int dx\, dz\, p_{train}(x,z)\, |\hat{r}(x|\theta,\theta_0) - r(x,z|\theta,\theta_0)|^2 \tag{3.19}$$

with $p_{train}(x,z) = p(x,z|\mu_0)$ to determine the best value of parameters a and b.

The model weights a and b may be learned analytically:

$$\frac{\partial \mathcal{L}}{\partial a} = 0 \implies \mu = \frac{1}{\sqrt{2\pi(\sigma_s^2+1)}} \int dx\, x e^{-\frac{(x-\mu_0)^2}{2(\sigma_s^2+1)}} e^{ax+b} \tag{3.20}$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \implies 1 = \frac{1}{\sqrt{2\pi(\sigma_s^2+1)}} \int dx\, e^{-\frac{(x-\mu_0)^2}{2(\sigma_s^2+1)}} e^{ax+b} \tag{3.21}$$

This system of equations is solved by

$$a = \frac{2(\mu-\mu_0)}{2(\sigma_s^2+1)} \tag{3.22}$$

$$b = \frac{\mu^2-\mu_0^2}{2(\sigma_s^2+1)}. \tag{3.23}$$

Comparison with the analytic solution given in Equation 3.18 shows agreement between the solution derived from direct integration and the optimized machine learning model.

It is worth highlighting the differences in the computations required for each approach. The analytical solution of Equation 3.18 required evaluating the integral appearing in Equation 3.16. The distributions $p(x|z)$ and $p(z|\mu)$ were intentionally defined so that this integral was tractable, but this is not generally the case. A numerical approximation of Equation 3.16 would not be a viable alternative as each x would require its own numerical integral over z.[1]

On the other hand, the machine learning model required only the optimization of Equation 3.19 with respect to the model parameters. An analytic solution to the optimization problem also required evaluating integrals of the sort found in Equation 3.16. However, numerical solutions do not require such integrals. Additionally, since x is integrated instead of being held fixed as before, any integrals appearing in the optimization may be approximated using a finite set of samples as is typical in machine learning. We thus see that the difficult integration problem of the former approach is cast into a familiar optimization problem in the latter approach.

[1]There would also be practical difficulties with such an integral. Typically detector simulations stochastically map z to x. One would need to invert this mapping in order to hold x fixed.

3.4 Experiment

We consider collisions $q\bar{q} \to q\bar{q}$ in the standard model with a massive Z' boson included [86, 87]. The θ that parameterizes this theory is $\theta = (M_{Z'}, g_{Z'})$, where $M_{Z'}$ is the mass of the Z' boson and $g_{Z'}$ is its coupling to standard model quarks. We attempt to find the Z' resonance using two different methods of calculating the likelihood ratio: the ALICE approach described above and the benchmark histogram-based approach, utilized in many current LHC searches [88, 89, 90, 91].

We sample 10^4 events at every point on a grid in θ-space spanning $M_{Z'} \in [275, 325]$ GeV and $g_{Z'} \in [0, 2]$, with grid spacing $\Delta M_{Z'} = 5$ GeV, $\Delta g_{Z'} = .2$, yielding a total of 1.21×10^6 weighted events. We use a constant width $\Gamma_{Z'} = 2.4$ GeV for every theory in the grid. Events are sampled using MadMiner [9], which generates, showers, and detects events using MadGraph v2.6.5, Pythia8 and Delphes, respectively [10, 11, 12]. Jets are reconstructed using the anti-k_{T} algorithm with distance parameter $R = .5$ [92]. Events where more or fewer than two jets are detected are discarded.

For each event, our observables x consist of the four-vector of each reconstructed jet. A cut is placed on the invariant mass so that $m_{jj} \in [150, 450]$ GeV and on the jet transverse momenta such that $p_T > 20$ GeV. For each sample, we calculate the tree-level joint likelihood ratio at every other grid point, which MadMiner performs by using Madgraph's reweight feature [93]. The joint likelihood ratios are then used to unweight the samples in preparation for machine learning.

We focus on a qualitative assessment of the application of machine learned likelihood ratios to resonance searches. As such, we neglect additional complications that may arise in an experimental setting that are not expected to affect qualitative results, such as pile up interactions, trigger strategies, the dependence of $\Gamma_{Z'}$ on $M_{Z'}$, or additional diagrams that

Table 3.1: Table of hyperparamaters used to the train neural network

Hyperparameter	Value
Activation function	tanh
Number of hidden layers	3
Neurons per hidden layer	12
Initial learning rate	2.2×10^{-3}
Final learning rate	10^{-4}
Learning rate decay schedule	Linear
Optimizer	AMSGrad
Batch Size	128
Validation Split	.25
Number of Epochs	100
Training Samples (unweighted)	10^6
θ_0	$(300\text{ GeV}, 2.0)$

may contribute to the detected final state. An interesting direction for future study would be to incorporate these effects into the analysis.

We use MadMiner to train a neural network capable of calculating $r(x|\theta, \theta_0)$ using the ALICE loss functional. We parameterize the dependence on θ, as described in Refs. [81, 94], which allows us to evaluate $r(x|\theta, \theta_0)$ at any θ in the parameter space using a single neural network. The neural network is trained with hyperparameters given in Table 3.1 to minimize the loss in Equation 3.10. We fix $\theta_0 = (300\text{ GeV}, 2.0)$, though results should be independent of this choice.

We also calculate the likelihood ratio from histograms. For this method, we bin events sampled from $p(x|\theta)$ and $p(x|\theta_0)$ in invariant mass, and the likelihood ratio is again given by the ratio of normalized counts in each bin. We use a fixed bin size of 20 GeV at all points in the parameter space.

We separately generate a test dataset $\mathcal{X}_{\text{test}}$ of 10000 events with $\theta_{\text{test}} = (285\text{ GeV}, 1.2)$ and compare the results of a resonance search using our machine learned $\hat{r}(x|\theta, \theta_0)$ to one using $r(x|\theta, \theta_0)$ calculated from histograms. The test set is used to calculate expected p-values for N test events in the asymptotic limit. In this work, we take $N = 50$. To demonstrate the

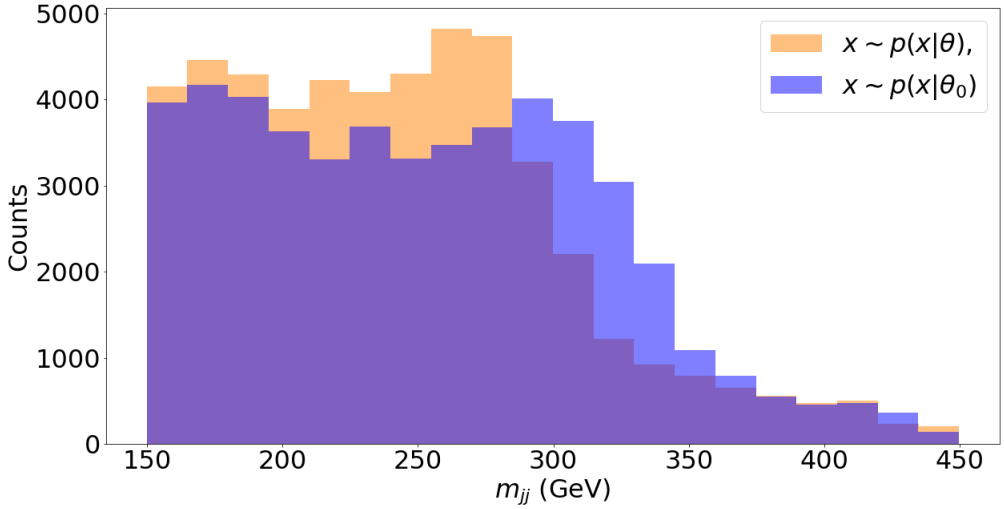

Figure 3.1: Distributions of invariant mass plotted for events drawn from $\theta = (285 \text{ GeV}, 1.0)$ (orange) and $\theta_0 = (315 \text{ GeV}, 1.0)$ (blue). We use these histograms to calculate the likelihood ratio, given by the ratio of counts in each bin.

limit setting abilities of these methods, we follow the example in Ref. [81]. We assume an asymptotically large test set and calculate

$$p_\theta = \exp\left(N \langle \log r(x|\theta, \theta_{\text{MLE}}) \rangle_{x \in \mathcal{X}_{test}}\right), \tag{3.24}$$

where θ_{MLE} is the parameters of the maximum likelihood estimate, which is the θ that maximizes $r(\mathcal{X}_{\text{test}}, \theta, \theta_0)$. This expression takes a simple form because the dimension of our parameter space is two.

Before presenting the results to the full problem described above, we attempt to develop an intuition regarding all of the likelihood ratios discussed thus far. For this, we limit our analysis to the set of events drawn from $\theta = (285 \text{ GeV}, 1.0)$ and $\theta_0 = (315 \text{ GeV}, 1.0)$. In Figure 3.1, we show histograms in invariant mass for events drawn from θ and events drawn from θ_0. The ratio of counts in each bin gives the likelihood ratio if the invariant mass is the only information available. The logarithm of this likelihood ratio is plotted as the grey line in Figure 3.2.

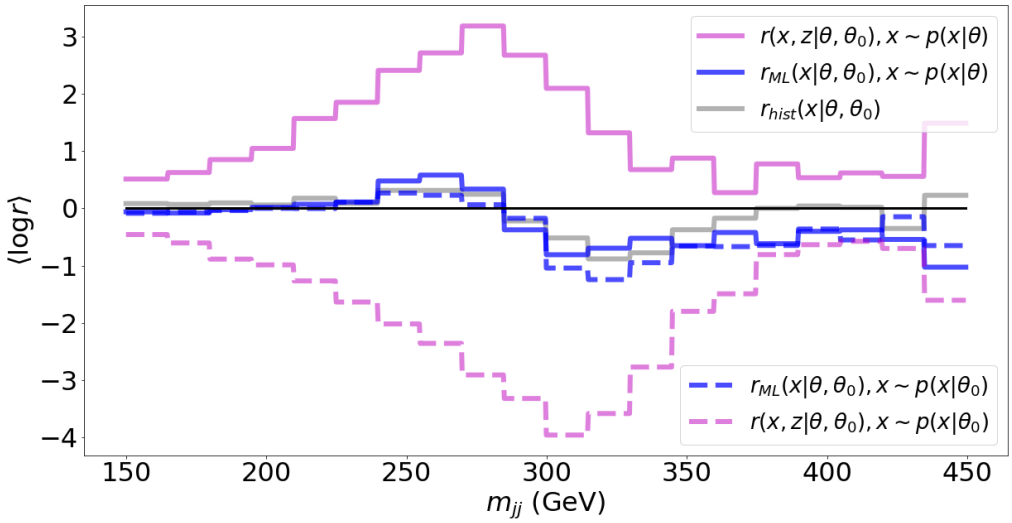

Figure 3.2: A comparison between the log likelihood ratio (grey, calculated using histograms in Figure 3.1), the expected machine learned log likelihood ratio (blue), and the expected log joint likelihood ratio (magenta) for events sampled from $\theta = (285 \text{ GeV}, 1.0)$ (solid) and $\theta_0 = (315 \text{ GeV}, 1.0)$ (dashed). Expectations are calculated with respect to all events that lie within the given invariant mass bin. We remark that the expected value of $r_{\text{hist}}(x|\theta, \theta_0)$ in a bin does not depend on the distribution from which an event is sampled, as only the number of events within each bin will change. The expected machine learned likelihood ratio is closer than the histogram approach to the expected joint likelihood ratio, which represents the optimal expected log likelihood ratio if given complete parton information of every event. The neural network approaches are not well converged at large invariant masses due to the lack of events in this region of the feature space.

The next step is to compare the likelihood ratio calculated from histograms to the machine-learned likelihood ratio as well as the joint likelihood ratio, which benefits from parton-level information. A neural network is used to discriminate between events sampled from θ and θ_0 by minimizing the ALICE loss functional. To compare this to the benchmark result, we show the expected value of the machine learned log likelihood ratio within each invariant mass bin. A full justification of this comparison is provided in section A of the appendix. This expectation is plotted for events sampled from $p(x|\theta)$ (solid) and $p(x|\theta_0)$ (dashed) in blue. The neural network uses multivariate detector level information, which allows it to make more powerful predictions than the histogram based approach, which only uses invariant mass.

We also plot the expected log joint likelihood ratio within each invariant mass bin in magenta for events drawn from $p(x|\theta)$ (solid) and $p(x|\theta_0)$ (dashed). The expected joint likelihood ratio represents the most powerful expected likelihood ratios possible, as it requires knowledge of all considered parton level event information. We see that the machine learning approach is able to use the extra available information to modestly outperform the histogram approach.

3.5 Results

In Figure 3.3, we plot p values as a function of mass and coupling, calculated using the ALICE likelihood ratio. In Figure 3.4, we show the same plot, calculated using histogram based likelihood ratios. While both approaches are capable of selecting the correct region of θ-space, the results are significantly less constrained for the histogram approach than when using the ALICE based likelihood ratio.

We remark that only kinematic information is used to form the likelihood ratios used in Figures 3.3 and 3.4. A full analysis would include total rate information. However, the

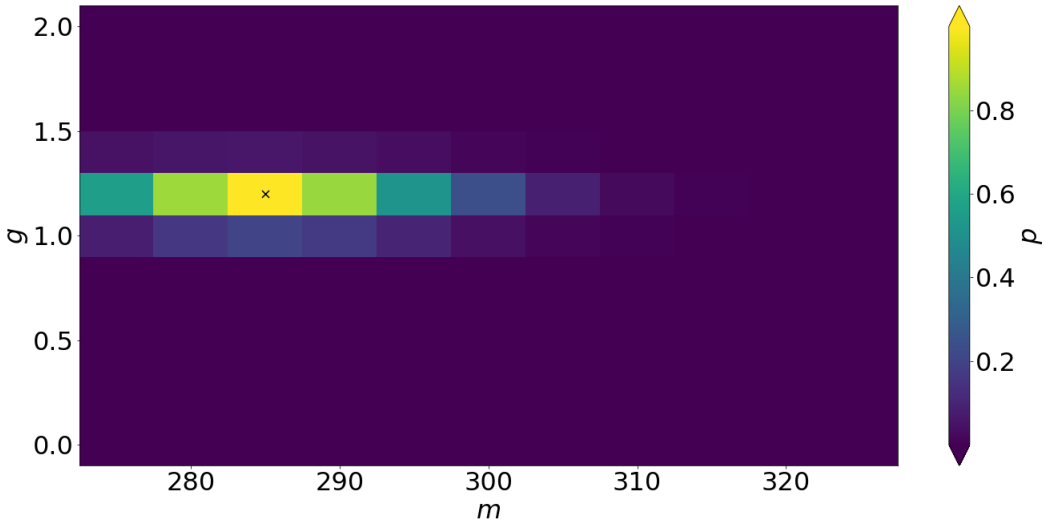

Figure 3.3: We show the expected p values plotted against mass and coupling for an Asimov test set drawn from the theory $(285\ \text{GeV}, 1.2)$. The p values are calculated using our machine learned likelihood ratio. The true value of θ is marked with a black x. In comparison to Figure 3.4, p values are more peaked around the true value of θ.

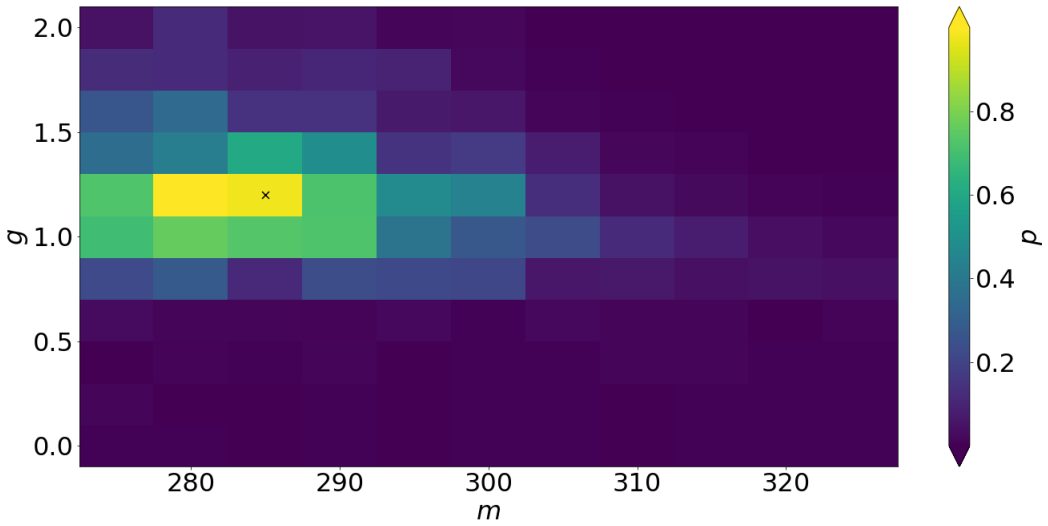

Figure 3.4: We show the expected p values plotted against mass and coupling for an Asimov test set drawn from the theory $(285\ \text{GeV}, 1.2)$. The p values are calculated using our likelihood ratio derived from histograms. The true value of θ is marked with a black x. In contrast to Figure 3.3, p values are more dispersed around the true value of θ.

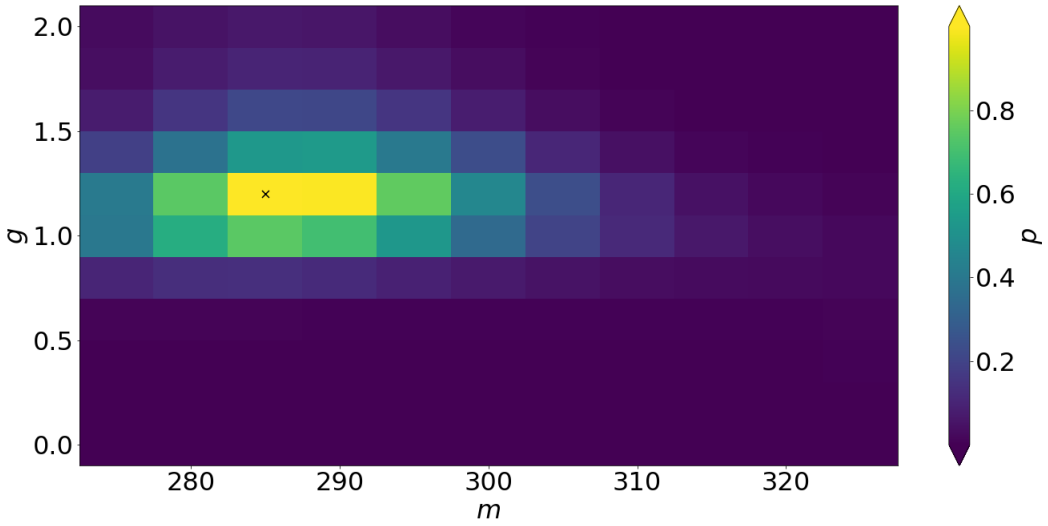

Figure 3.5: We show the expected p values plotted against mass and coupling for an Asimov test set drawn from the theory (285 GeV, 1.2). The p values are calculated using our machine learned likelihood ratio trained only on invariant mass. The true value of θ is marked with a black x.

contribution of rate information to the log likelihood ratio is independent of the method used to calculate the kinematic portion of the log likelihood ratio, and thus is unable to change the relative ordering of the methods.

We believe that the ALICE likelihood ratio is able to outperform histogram based approaches due to the increased information utilized by ALICE. That is, the ALICE likelihood ratio uses information of the full four-momenta of both final state jets. On the other hand, the histogram based approach only has access to the invariant mass of the jet pair and cannot be extended to use additional observables due to the curse of dimensionality.

A possible critique of this interpretation is that the histogram based approach does not perform as well as ALICE because it requires use of a binned PDF, which may result in information loss. To address this concern, we also train a neural network using the ALICE loss functional to predict the likelihood ratio as a function of only invariant mass, instead of the full eight-dimensional dimensional jet four-momenta. This can be considered the

continuous limit of the histogram based likelihood ratio, as it avoids the need for binning while only using the information in invariant mass. The expected p values plotted over θ are shown in Figure 3.5 and the hyperparameters used to train our neural network are shown in Table 3.2.

We see that the machine learned likelihood ratio trained only on invariant mass performs very similarly to the histogram result. This indicates there is not a substantial difference in available information due to binning effects.

In a final attempt to uncover the extra information used by the fully multivariate ALICE likelihood ratio, we train an additional neural network with the ALICE loss functional, but provide the detected invariant mass as well as the difference in pseudorapidity of the two final state jets, denoted Δy_{jj}. The input is thus two dimensional. The hyperparameters used to train the neural network are the same as those in Table 3.2, with the initial and final learning rates adjusted to 2.3×10^{-3} and 4×10^{-5}, respectively.

Because the neural networks for the machine learned likelihoods include θ as an input, they offer a very natural interpolation in θ-space relative to the histogram approach. In Figure 3.6, we use this property to compare the exclusion contours of the fully multivariate, eight dimensional machine learned likelihood ratio (blue), the two dimensional machine learned likelihood ratio (green), and the machine learned likelihood ratio that is only dependent on invariant mass (purple). We see that the neural network trained on two dimensional input performs very similarly to the neural network trained on eight dimensional input, and that both of these significantly outperform the one dimensional approach.

This result demonstrates that the extra information used by the fully multivariate approach is largely or nearly entirely captured in Δy_{jj}. While these exclusion contours depend on the hyperparameter N and will also change if we included rate information into the analysis, the relative ordering will not depend on these factors.

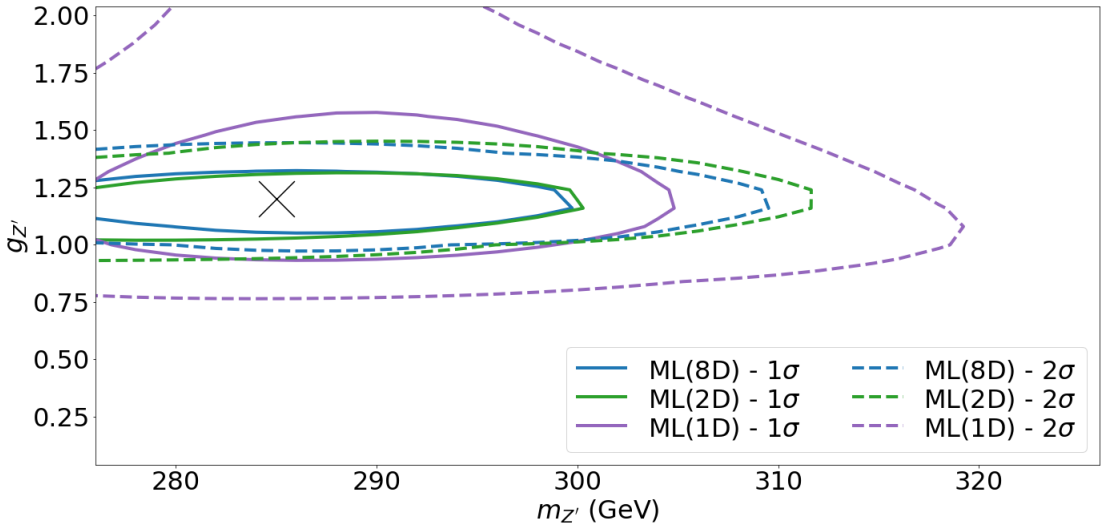

Figure 3.6: The 1σ (solid) and 2σ (dashed) exclusion contours for the machine learned likelihood ratio calculated using the full input (blue), invariant mass and Δy_{jj} (green), and only invariant mass (purple). The black x denotes the true value of θ. The fully multivariate and two dimensional approaches perform very similarly, and both provide more powerful exclusion contours than the approach that uses only invariant mass.

Table 3.2: Table of hyperparamaters used to the train neural network with invariant mass as input.

Hyperparameter	Value
Activation function	tanh
Number of hidden layers	3
Neurons per hidden layer	8
Initial learning rate	10^{-3}
Final learning rate	10^{-5}
Learning rate decay schedule	Linear
Optimizer	AMSGrad
Batch Size	128
Validation Split	.25
Number of Epochs	100
Training Samples (unweighted)	10^6
θ_0	(300 GeV, 2.0)

3.6 Conclusion

In this work, we have performed the first application of a novel likelihood-free inference method, ALICE, beyond the scope of effective field theories. ALICE, along with most methods from this new class of analysis techniques, relies on machine learning latent information extracted from simulations in order to produce a useful likelihood ratio. We have compared the new method to the traditional histogram based approach of performing a resonance search, and have seen dramatic improvement when multivariate detector level event information is included.

ALICE outperforms the histogram approach by providing significantly tighter exclusion contours. We believe that this improvement is similar to the improvement seen when using the matrix-element method, and originates from the greater amount of information that can be meaningfully utilized in multivariate analyses. Since we are now able to compare lower dimensional analyses to a fully dimensional analysis, we were able to conclude that nearly all information is captured in the observables m_{jj} and Δy_{jj} for our simple process. ALICE can be seen as an improvement over the matrix-element method, as it does not require one to approximate the detector using transfer functions.

Possibilities for future work include utilizing the partial morphing structure with the coupling to improve the computational efficiency of this work. One could also numerically evaluate derivatives of the joint likelihood ratio with respect to the mass. In combination with the partial morphing structure in the coupling, this would grant full access to the derivative of the joint likelihood ratio with respect to the theory parameters, which is known as the joint score. For EFTs, methods that include the joint score in the loss function have been more successful than those that neglect this information. It is possible that these results also generalize to resonance searches, and that even more information can be extracted from the detector level four-momenta.

Additionally, one could study if these results generalize to the case where a full treatment of systematic uncertainties in the input is performed. Since these methods scale well to high dimensional problems [95], one could also expand the space of observables to include the four-momenta of all jet constituents, rather than only using reconstructed jet four-momenta. This would potentially increase the information available to the neural network, at the cost of increasing the dimensionality of the problem. Finally, this work may be applied to more complicated resonant processes. We expect that, since we already see improvement in discovery potential in the relatively simple process considered here, more complicated processes may see even greater benefit from the extra information present in the multivariate approach.

Acknowledgements

The authors would like to thank Felix Kling, Johann Brehmer, and Kyle Cranmer for many valuable discussions throughout all stages of this project. The authors would also like to thank Benjamin Nachman and Jesse Thaler for feedback on a preliminary draft of this manuscript. This material is based upon the work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1321846. DW is supported by the Department of Energy Office of Science.

Chapter 4

Efficient Sampling of constrained high dimensional theoretical spaces with machine learning

This chapter is heavily based on work previously published in collaboration with Michael Ratz, Philip Tanedo, and Daniel Whiteson [96].

4.1 Introduction

Models of physics beyond the Standard Model often feature many new parameters that are unknown *a priori* and may only be determined by experiment. However, experimental constraints are not trivial to apply, as they often are expressed in terms of weak scale observables rather than the theory's fundamental parameters. While it is often straightforward (if computationally expensive) to calculate the weak scale observables from the parameters,

the inverse problem is typically intractable. That is, weak scale constraints do not allow for a trivial reduction of the dimensionality of the theory space.

The standard approach is to numerically scan over the theoretical parameters and reject those that are not consistent with experimental data. However, the number of samples required for a brute-force search of the parameter space increases exponentially with its dimension. Thus, particle physicists studying models of new physics are often faced with a computationally intractable task. One may pragmatically restrict to a more tractable subset of parameters based on theoretical prejudice. The danger of this approach is that one may miss viable parameters that are both consistent with experimental observations and generate novel phenomenology.

The MSSM is a well-known example of a new physics model with a large number of free parameters ($\sim$ 100). Most of these parameters are the masses and couplings of the supersymmetric partners of Standard Model particles [29]. This overwhelming dimensionality prohibits a fully general survey of the parameter space. Studies of the MSSM typically restrict to theoretically motivated subspaces [1, 97, 98, 99, 100, 101, 102, 2, 103, 104, 15, 105]. These include the 4+1 dimensional cMSSM as well as the 19 dimensional pMSSM [33, 32]. However, even these reduced spaces are difficult to scan using a brute-force search.

High dimensionality is not the only challenge when scanning the parameters of the MSSM. The fundamental parameters of the theory are defined at some high energy scale and must be evolved to the energy scale of the experiment. This evolution requires one to solve the coupled RGEs for the high-scale parameters over many orders of magnitude to the weak scale. The computational cost of RGE running and calculating experimental observables for a single set of parameters is expensive, $\mathcal{O}$(second) for a modern CPU.

Many recent scans have incorporated machine learning in some capacity to decrease the computational burden of brute-force searching these spaces [15, 105, 101]. These use various

machine learning models to learn the *forward* problem of determining weak scale properties given high-scale parameters. This bypasses the need to perform RGE running and weak scale computations, however one is still faced with the challenge of doing a brute-force search over a high-dimensional parameter space. Machine learning models for the forward problem are thus only a constant improvement in computational time compared to the exponential dependence on the dimension of the space.

In this work, we introduce two methods to efficiently sample high-dimensional parameter spaces subject to constraints at the weak scale. We test these frameworks by sampling regions of the cMSSM and pMSSM parameter spaces that admit a Higgs mass consistent with its experimental value [106, 107]. The first uses a deep neural network to machine-learn the likelihood of an event satisfying this constraint and then samples this likelihood using HMC. The second trains a generative model known as a normalizing flow. We then compare the performance of these frameworks to random sampling.

These methods allows us to directly and quickly generate points in the parameter space that admit a consistent Higgs mass. By solving the inverse problem of sampling high-scale parameters given weak scale properties, we aim to minimize inefficiencies that arise in a brute-force search.

Our presentation is a proof of concept for these generative models and is encouraging for practical applications. For example, the ability to efficiently scan the MSSM parameter space makes it much easier to determine the high-scale parameters that are consistent with a new particle's mass and width if a sparticle is discovered. Alternatively, a trained generative model may permit scans over parameters that are consistent with experimental observations to search for specific theoretical features that one may wish to study, for example: gauge coupling unification, a particular type of dark matter particle, or low fine-tuning measures.

As a demonstration of the efficiency of the generative models, we scan the cMSSM and pMSSM parameter spaces for points that produce the Higgs mass and that saturate the observed dark matter relic density, requiring [108, 109]

$$\begin{aligned} 122 \text{ GeV} < \quad m_h \quad &< 128 \text{ GeV} , \\ 0.08 < \Omega_{\text{DM}} h^2 &< 0.14 . \end{aligned}$$

In this study, the generative models have been trained for consistency with the Higgs mass, not the relic density. We compare a brute-force scan using random sampling to a generative model that has been trained to sample points that admit a consistent Higgs mass. We show that the generative models dramatically increase the sampling efficiency of this scan.

4.2 Methods

4.2.1 Data Generation

The cMSSM contains 4 continuous parameters defined at the GUT scale and 1 discrete sign parameter. These are the universal scalar mass m_0, the universal gaugino mass $M_{1/2}$, universal trilinear coupling A_0, the ratio of Higgs vacuum expectation values $\tan\beta$, and the sign of μ. The pMSSM is the most general subspace of the MSSM that admits first and second generation universality, no new sources of CP violation, and no flavor changing neutral currents [32]. Parameters of the pMSSM are defined at the EW scale. The full list parameters of the pMSSM are listed as part of Table 4.2.

Our datasets are formed by uniform random sampling within bounded regions of the parameter space: cMSSM parameters are sampled at the GUT scale and pMSSM parameters are sampled at the EW scale. Bounds are listed for the cMSSM and the pMSSM in Table 4.1 and Table 4.2, respectively [1, 2], and are chosen to cover large volumes of the parameter

Parameter	Domain	Description
m_0	$[0, 10]$ TeV	Universal scalar mass
$m_{1/2}$	$[0, 10]$ TeV	Universal gaugino mass
A_0	$[-6m_0, 6m_0]$ TeV	Universal trilinear coupling
$\tan\beta$	$[1.5, 50]$	Ratio of Higgs VEVs

Table 4.1: Parameter bounds in the cMSSM scan, following Ref. [1]. A uniform prior is used for all parameters except A_0, where we uniformly sample A_0/m_0.

space that are sensitive to modern collider experiments. For the cMSSM, we fix $\text{sign}(\mu) = 1$. We sample approximately 1.5×10^6 datapoints in the cMSSM and approximately 1.95×10^7 datapoints in the pMSSM. Once sampled, we calculate Higgs masses and relic densities with micrOMEGAs, which internally uses the spectrum generator SoftSUSYv4.1.0 [14, 13].

We apply two theoretical constraints: (*i*) consistent electroweak symmetry breaking and (*ii*) the positivity of all squared masses. In addition to these, we also require that SoftSUSY converges. We do not require that the lightest supersymmetric particle is neutral, though this is the case for 90% of the cMSSM and 99% of the pMSSM parameter points with a consistent Higgs mass.

The theoretical uncertainty in the Higgs mass is significantly larger than its experimental uncertainty [110]. We take the uncertainty in the Higgs mass calculations to be $\sigma_{m_h} = 3$ GeV for all points in the data set [1, 2].

4.2.2 Neural Network

We train the neural network by assigning all points in the dataset a likelihood

$$L(\theta) = \begin{cases} 1 & |m_h(\theta) - m_{h,\text{exp}}| < \sigma_{m_h} \ , \\ 0 & \text{otherwise} \ , \end{cases} \tag{4.1}$$

Parameter	Domain	Description
$\lvert M_1\rvert$	$[.05, 4]$ TeV	Bino mass
$\lvert M_2\rvert$	$[.1, 4]$ TeV	Wino mass
M_3	$[.4, 4]$ TeV	Gluino mass
$\lvert\mu\rvert$	$[.1, 4]$ TeV	Bilinear Higgs mass
$\lvert A_t\rvert$	$[0, 4]$ TeV	Trilinear top coupling
$\lvert A_b\rvert$	$[0, 4]$ TeV	Trilinear bottom coupling
$\lvert A_\tau\rvert$	$[0, 4]$ TeV	Trilinear τ coupling
M_A	$[.1, 4]$ TeV	Pseudo-scalar Higgs mass
$m_{\tilde{L}_1}$	$[.1, 4]$ TeV	1st gen. l.h. slepton mass
$m_{\tilde{e}_1}$	$[.1, 4]$ TeV	1st gen. r.h. slepton mass
$m_{\tilde{L}_3}$	$[.1, 4]$ TeV	3rd gen. l.h. slepton mass
$m_{\tilde{e}_3}$	$[.1, 4]$ TeV	3rd gen. r.h. slepton mass
$m_{\tilde{Q}_1}$	$[.4, 4]$ TeV	1st gen. l.h. squark mass
$m_{\tilde{u}_1}$	$[.4, 4]$ TeV	1st gen. r.h. u-type squark mass
$m_{\tilde{d}_1}$	$[.4, 4]$ TeV	1st gen. r.h. d-type squark mass
$m_{\tilde{Q}_3}$	$[.2, 4]$ TeV	3rd gen. l.h. squark mass
$m_{\tilde{u}_3}$	$[.2, 4]$ TeV	3rd gen. r.h. u-type squark mass
$m_{\tilde{d}_3}$	$[.2, 4]$ TeV	3rd gen. r.h. d-type squark mass
$\tan\beta$	$[1, 60]$	Ratio of Higgs VEVs

Table 4.2: Parameter bounds in the pMSSM scan, following Ref. [2]. A uniform prior is used for all parameters. "Left-handed" and "right-handed" are abbreviated by l.h. and r.h., respectively.

where we ignore a normalization constant. All data points that fail the theoretical constraints are assigned a likelihood of zero.

We use a deep neural network to learn the function $L(\theta)$ [111]. This has two benefits. First, it greatly reduces the time required to evaluate the likelihood of a point. Second, it provides a differentiable interpolation of $L(\theta)$. In the next section we show that HMC requires many evaluations of the likelihood and its gradients. It thus utilizes the full potential of these benefits.

We train a deep neural network $\hat{L}(\theta)$ to minimize the usual L2 loss function

$$\mathcal{L} = |\hat{L}(\theta) - L(\theta)|^2 \, . \tag{4.2}$$

We use a training, validation, and testing split of $0.7, 0.15, 0.15$ respectively for both datasets. Batch norm and dropout layers are used in between each hidden layer of the neural network. Backpropogation is performed using the ADAM optimizer [112].

Some of the pMSSM parameters in Table 4.2 span a disconnected range of positive and negative values, for example M_1, M_2 and μ. We preprocess these parameters by shifting negative values to create a single continuous domain; for example, for μ we shift the negative values by 200 GeV. This has no physical significance and simply prepares the data for input into the neural network. We then standardize each feature. For the cMSSM dataset, we use the feature A_0/m_0 in place of A_0, as this feature is uniformly distributed.

4.2.3 Hamiltonian Monte Carlo

The Hamiltonian Monte Carlo method is a Markov chain Monte Carlo technique that uses an analog of energy conservation to effectively sample the target distribution [113, 114]. To use the method, we first define an auxiliary momentum variable p, where each component

is initially drawn from a normal distribution. Next, we define a potential energy function given by

$$V(\theta) = -\log(\hat{L}(\theta)) \ . \tag{4.3}$$

The kinetic energy function takes the familiar form $T = p^2/2$ where we set the mass to unity, $m = 1$. We then evolve the system from time $t = 0$ to $t = \tau$ according to the Hamiltonian equations of motion

$$\frac{\partial \theta_i}{\partial t} = p_i \ , \qquad\qquad \frac{\partial p_i}{\partial t} = \frac{\nabla \hat{L}(\theta)}{\hat{L}(\theta)} \ . \tag{4.4}$$

We solve these equations using the leap-frog algorithm so that energy is approximately conserved. We take $\theta(\tau)$ as a proposal to add to the Markov chain. The proposal is accepted with probability

$$P = \min\left(1, \frac{e^{-H(\theta(\tau),p(\tau))}}{e^{-H(\theta(0),p(0))}}\right) \ . \tag{4.5}$$

Energy conservation implies that a solution to the the equations of motion should always yield probability 1. However, a rejection step is necessary because we solve these equations numerically. If $\theta(\tau)$ is rejected, then $\theta(0)$ is added to the Markov chain instead. In the limit of an infinite number of samples, the Markov chain converges to a sample of the distribution $\hat{L}(\theta)$. We seed the Markov chain with a random positive sample from the dataset used to train the neural network. We bound the parameter space with hard walls of infinite potential energy.

4.2.4 Normalizing Flows

It is difficult to draw samples from a complicated distribution in a high-dimensional parameter space. On the other hand, it is easy to draw samples from an equally high-dimensional Gaussian distribution. Normalizing flows is a technique that learns an invertible map f from the simple distribution p_Z to the challenging distribution p_Y. One then creates a set of samples from the challenging distribution by mapping easy-to-generate samples:

$$p_Y(y) = p_Z(f^{-1}(y)) \left| \det \left(\frac{\partial f}{\partial y} \right) \right|^{-1}. \tag{4.6}$$

The function f depends on a set of parameters Θ which are learned by maximizing the log likelihood of a training set, $\mathcal{X}$. The loss function for this training is thus

$$\mathcal{L}(\mathcal{X}) = -\sum_{y \in \mathcal{X}} \left(\log \left(p_Z(f^{-1}(y)) \right) - \log \left| \det \left(\frac{\partial f}{\partial y} \right) \right| \right).$$

It is helpful to construct f to be the composition of n successive maps, $f = f_n \circ \cdots \circ f_1$ [111]. Defining $z_{i+1} = f_i(z_i)$ and identifying $y = z_{n+1}$ yields the loss function

$$\mathcal{L}(\mathcal{X}) = -\sum_{y \in \mathcal{X}} \left(\log \left(p_Z(z_1) \right) - \sum_{i=1}^{n} \log \left| \det \left(\frac{\partial z_{i+1}}{\partial z_i} \right) \right| \right).$$

We choose the f_i to be autoregressive transformations. This means that the parameters Θ_i^k that define the function f_i acting on the k^{th} feature z_i^k depends only on the first $(k-1)$ features $z_i^1, \cdots, z_i^{k-1}$:

$$z_{i+1}^k = f_i\big(z_i^k\,;\ \Theta_i^k(z_i^{1:k-1})\big).$$

This structure ensures that the Jacobian matrix $\partial z_{i+1}/\partial z_i$ is lower triangular so that the determinant is simply the product of diagonal elements and may be computed in linear time.

The function $\Theta_i^k\left(z_i^{1:k-1}\right)$ can be represented efficiently with a Masked Autoencoder for Distribution Estimation (MADE) [115]. MADE networks turn off specific internal weights of the neural network so that the autoregressive property is enforced, allowing one neural network to output all model parameters rather than performing a sequential loop over features.

For our application, we choose f_i to be rational-quadratic neural spline flows with autoregressive layers [8]. These are piece-wise monotonic functions defined as the ratio of two quadratic functions on the interval $[-B, B]$, with $K+1$ knots determining the boundaries between bins. Outside of this interval, the transformation is defined to be the identity. These transformations are parameterized by $3K-1$ parameters for each feature, which are K bin heights, K bin widths, and $K-1$ positive derivative values at the knots, as the derivatives are set to 1 at $-B$ and B to ensure a continuous derivative over the domain. Permutation layers are included between rational-quadratic transformation layers. We implement the normalizing flow using the Python package *nflows* [8].

4.3 Results

We analyze the performance of these generative frameworks on the cMSSM and pMSSM datasets described above. The cMSSM is low dimensional and can be scanned relatively well with brute-force search. Thus, we view the cMSSM as a test for the generation methods and the pMSSM as a more practical application. We present the results for the neural network with HMC as well as the normalizing flow side by side. For each method, we generate a dataset of 4×10^5 datapoints.

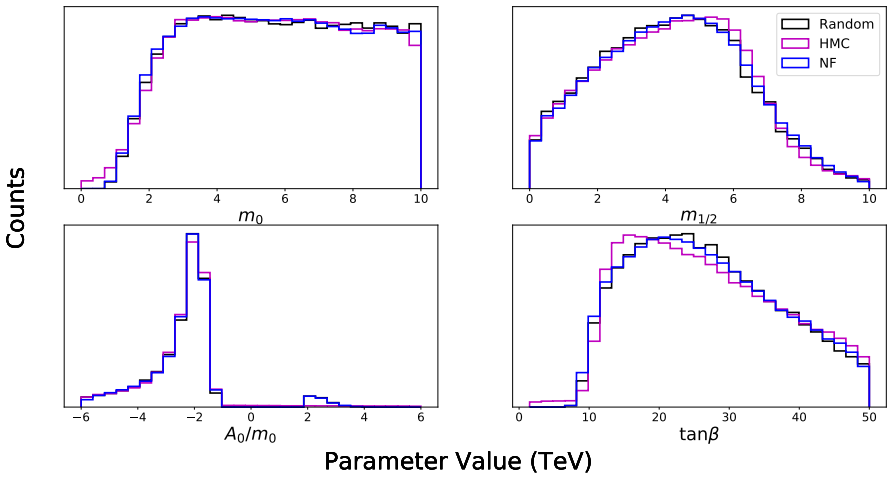

Figure 4.1: Histograms of cMSSM parameters that yield the experimental Higgs mass. We observe good agreement between the random sampling, HMC, and the flow model. BLACK: Data obtained through random sampling with a uniform prior and rejecting points that do not have a consistent Higgs mass. MAGENTA: data sampled with HMC. BLUE: data sampled from the flow model. No rejection step is applied to generated samples.

We present histograms of generated variables to confirm that the distribution of theory parameters is not biased by our generative framework. We also present histograms of m_h to ensure that our generative models sample within the band of permitted Higgs masses and $\Omega_{\text{DM}}h^2$ to provide evidence that the distribution of weak scale quantities match, as these are sensitive to higher order correlations in high energy scale parameters. Finally, we report sampling efficiencies, which are defined as the fraction of the dataset that satisfy a constraint. The hyperparameters used for the supervised neural network, Hamiltonian Monte Carlo, and normalizing flow are given in the Appendix for both datasets.

4.3.1 cMSSM

In Figure 4.1, we compare histograms of the cMSSM parameters at the GUT scale. For both generative models, we see very good agreement between the distribution of generated samples and the distribution of randomly sampled points after the Higgs mass constraint is

applied. Next, we run the parameters to the weak scale in order to perform the combined search for $\Omega_{\rm DM}h^2$ and m_h. In Figure 4.2, we show the distribution of Higgs masses for generated points and randomly sampled points with a rejection step applied. We see that the generative models typically sample within the band of permitted Higgs masses.

As an example application, we show histograms of the dark matter relic density for these datasets in Figure 4.3. We see that the distribution over dark matter relic densities from the generative models appear to accurately reflect the same distribution in the dataset after the Higgs mass constraint is applied. We emphasize that because the RGEs are coupled, weak-scale quantities are generally sensitive to higher-order correlations of the GUT scale parameters, and so matching weak-scale distributions is evidence of matching higher order correlations in the GUT scale parameters. This indicates that the m_h-constrained subspace has been accurately sampled, allowing for an exploration of additional constraints, such as relic density.

In Table 4.3, we compare various statistical properties of random sampling to those of our generative frameworks trained to satisfy the Higgs mass constraint. The first row shows the sampling efficiency with respect to the theoretical constraints mentioned in Section II.A. We see that samples from the generative models are more likely to pass these constraints, as points with a consistent Higgs mass necessarily satisfy the theoretical constraints. The second row shows the sampling efficiency with respect to the Higgs mass constraint. Predictably, the generative models have significantly higher sampling efficiencies than random sampling. We also see that the flow model slightly outperforms the HMC sampling method.

The third row shows the sampling efficiencies with respect to the combined Higgs mass and relic density constraint, where the generative models are still trained to only satisfy the Higgs mass constraint. This simulates a scenario where one would like to study the effect of imposing a new constraint in addition to the constraints that are explicitly trained on. Once again, we see that the generative models have much higher sampling efficiencies, resulting

	Sampling Method		
Constraint	Random	HMC_{m_h}	NF_{m_h}
Theory	0.595	0.859	0.879
Theory $\cap\, m_h$	0.0389	0.723	0.796
Theory $\cap\, m_h \cap\, \Omega_{\text{DM}}h^2$	0.000222	0.00271	0.00456

Table 4.3: Comparison of sampling efficiency in the cMSSM for several methods and several levels of constraints. We compare a brute force random scan (random), Hamiltonian MC of a neural network trained to learn the m_h constraint (HMC_{m_h}), and normalizing flows that incorporate the m_h constraint (NF_{m_h}). The constraints applied are theoretical consistency checks (see text), consistency with the experimental Higgs mass and consistency with the Higgs mass and the dark matter relic density ($\Omega_{\text{DM}}h^2$).

from the high probability that the samples pass the Higgs mass constraint. We see an increase in sampling efficiency of approximately an order of magnitude for both generative frameworks.

4.3.2 pMSSM

Differences between the generative models appear in the higher-dimensional pMSSM. In Figure 4.4, we compare histograms of parameters sampled using brute-force search, HMC and the normalizing flow model. Despite the increased dimensionality, we find very good agreement in the distributions of all parameters.

Figures 4.5 and 4.6 present histograms of m_h and $\Omega_{\text{DM}}h^2$ for the pMSSM. The generative models tend to sample in the band of allowed Higgs masses, with the normalizing flow model matching the brute-force scan well. We see general agreement with the true distribution of dark matter abundances for both generative frameworks, though the HMC samples do not match the brute-force distributions as well as those from the flow model.

Table 4.4 summarizes the performance of our sampling methods in the pMSSM. See Section 4.3.1 for a detailed description of the quantities presented in the table. We find that generative models greatly increase the sampling efficiency relative to a brute-force search. In

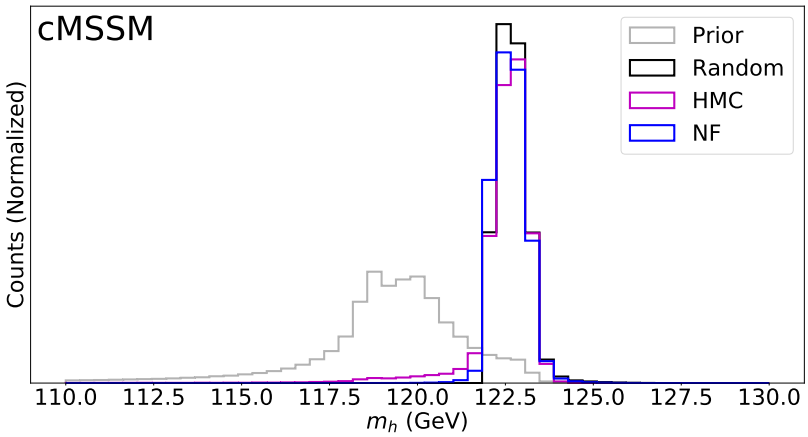

Figure 4.2: Histogram of Higgs masses in the cMSSM for different sampling methods. The generative models are seen to mostly sample points consistent with the Higgs mass constraint. GRAY: data obtained through random sampling with a uniform prior. BLACK: the same randomly sampled data, but points that do not have a consistent Higgs mass are rejected. MAGENTA: data sampled with HMC. BLUE: data sampled with the normalizing flow.

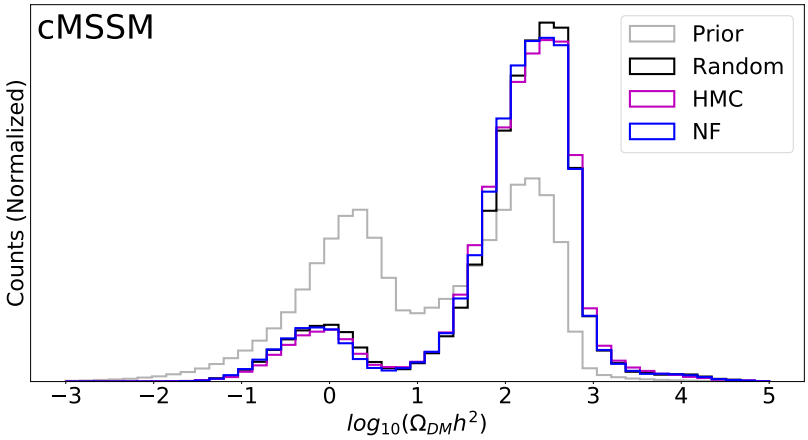

Figure 4.3: Histogram of dark matter thermal relic densities in the cMSSM for different sampling methods. We observe that the distributions of the generative models match the distribution of random sampling, providing evidence that the generative models are able to match higher order correlations in GUT scale parameters. GRAY: data obtained through random sampling with a uniform prior. BLACK: the same randomly sampled data, but points that do not have a consistent Higgs mass are rejected. MAGENTA: data sampled with HMC. BLUE: data sampled with the normalizing flow. Generative models have been trained to satisfy the Higgs mass constraint.

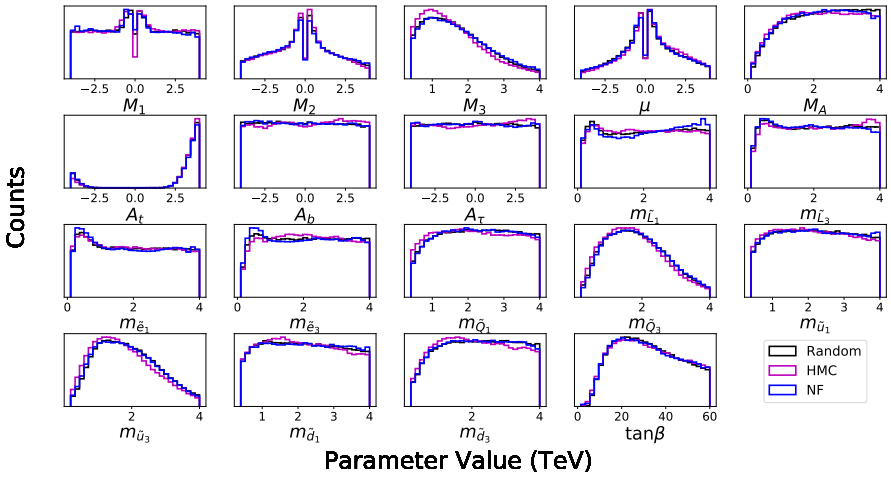

Figure 4.4: Histograms of pMSSM parameters that yield the experimental Higgs mass. We observe good agreement between random sampling, HMC, and the flow model. BLACK: Data obtained through random sampling with a uniform prior and rejecting points that do not have a consistent Higgs mass. MAGENTA: data sampled with HMC. BLUE: data sampled from the flow model. No rejection step is applied to generated samples.

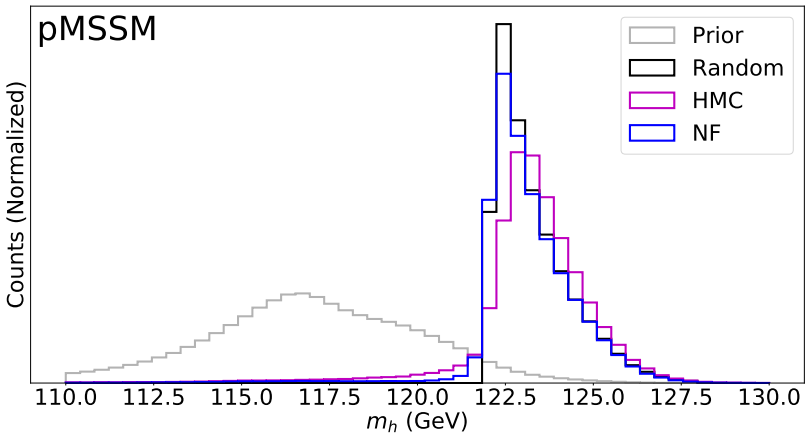

Figure 4.5: Histogram of Higgs masses in the pMSSM. The generative models are seen to mostly sample points consistent with the Higgs mass constraint. GRAY: data obtained through random sampling with a uniform prior. BLACK: the same randomly sampled data, but points that do not have a consistent Higgs mass are rejected. MAGENTA: data sampled with HMC. BLUE: data sampled with the normalizing flow.

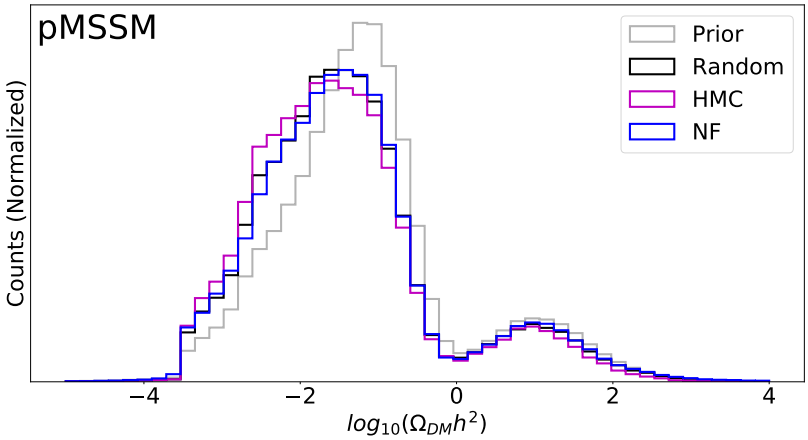

Figure 4.6: Histogram of dark matter thermal relic densities in the pMSSM. We observe that the distributions of the generative models match the distribution of random sampling, providing evidence that the generative models are able to match higher order correlations in EW scale parameters. GRAY: data obtained through random sampling with a uniform prior. BLACK: the same randomly sampled data, but points that do not have a consistent Higgs mass are rejected. MAGENTA: data sampled with HMC. BLUE: data sampled with the normalizing flow. Generative models have been trained to satisfy the Higgs mass constraint.

fact, the improvement in sampling efficiency is much greater than that seen in the cMSSM. This is largely due to the poorer performance of a brute-force search in the higher-dimensional pMSSM.

4.4 Conclusion

We implement two generative frameworks that utilize machine learning in order to increase the sampling efficiency of searches in supersymmetric parameter spaces. These sampling methods offer a more efficient way to search the high-dimensional parameter spaces in models of new particle physics. We compare these generative frameworks to the currently used method of a brute-force search, and have seen orders of magnitude of improvement in the sampling efficiency for both parameter spaces considered here. We show that our generative

	Sampling Method		
Constraint	Random	HMC_{m_h}	NF_{m_h}
Theory	0.479	0.810	0.866
Theory $\cap\, m_h$	0.0189	0.709	0.776
Theory $\cap\, m_h \cap\, \Omega_{\text{DM}} h^2$	0.00165	0.0591	0.0685

Table 4.4: Comparison of sampling efficiency in the pMSSM for several methods and several levels of constraints. Methods compared are brute force random scan, Hamiltonian MC of a neural network trained to learn the m_h constraint (HMC_{m_h}), and normalizing flows that incorporate the m_h constraint (NF_{m_h}). Constraints applied are theoretical consistency checks (see text), consistency with the experimental Higgs mass and consistency with the Higgs mass and the dark matter relic density ($\Omega_{\text{DM}} h^2$).

frameworks are able to sample the underlying data distribution without any evidence of bias or mode collapse.

In the cMSSM, both methods significantly outperformed random sampling, with the flow model slightly outperforming HMC. In the pMSSM the flow model significantly outperforms HMC. This is likely due to the larger dimensionality of the pMSSM. In addition to performance benefits, the flow model is also quicker to train and sample, making it clearly favorable to HMC. However, the HMC framework is more complementary to previous works, as it learns the forward problem of determining likelihoods and uses tested Monte Carlo algorithms to sample this likelihood.

Possibilities for future work include incorporating additional constraints into the generative model. In theory, there is no limit to the number of constraints that can be included into either generative model. However, forming an initial dataset for learning may be difficult when the constraints are very strict. A possible remedy is to train generative models with less restrictive constraints which are then used to produce sizable datasets of points that already satisfy many constraints. This new dataset could then be searched to form a training set for a generative model with increasingly restrictive constraints.

Given the ability of the generative machine learning models to efficiently explore high-dimensional parameter spaces, it will be interesting to apply the techniques described here

to other problems. For instance, one may identify relations that explain why there is a 'little hierarchy' between the electroweak scale and the scale of soft parameters, which go beyond the focus point scenario [116]. In general, one may be able to identify manifolds of viable points in high-dimensional parameter sets, and explore their geometry.

We have shown promising results in subspaces of the MSSM parameter space. These results apply generally to any high-dimensional parameter space with constraints that are computationally expensive to verify. Another direction for future study may be applications to the parameter spaces of even higher-dimensional models of new physics. This includes potentially relaxing constraints built into the pMSSM parameter space, but could also include applications to non-SUSY theories. Finally, one could attempt to further tune the neural network structure and hyperparameters in order to achieve higher sample efficiency than was achieved in this work.

4.5 Acknowledgements

The authors would like to thank Tim Cohen, Syris Norelli, Stephan Mandt and Babak Shahbaba. This material is based upon the work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1321846. DW is supported by the Department of Energy Office of Science. PT is supported by DOE grant DE-SC/0008541. The work of MR is supported by the National Science Foundation under Grant No. PHY-1915005.

Chapter 5

Proposed experimental tests for the phenomenological MSSM

In the previous chapter, large steps were taken towards reducing the sampling inefficiencies inherit to exploring constrained high dimensional parameter spaces. However, generated samples still lie within a high dimensional space and the phenomenology of these samples must be analyzed to determine potential signals that may be observed in experiment. In this chapter, we analyse models satisfying many experimental constraints in order to determine how they may best be observed in experiment. In the future, the sampling methods of the previous chapter may be used to greatly increase the breadth of this search.

This work is still in progress and no methods or results presented here should be considered as final. The results reported were developed in collaboration with Daniel Whiteson and benefited from conversations with Jonathan Feng and Anyes Tafford.

5.1 Introduction

Supersymmetric extensions of the Standard Model serve as a primary candidate for physics beyond the Standard Model (BSM). The Minimal Supersymmetric Standard Model (MSSM) is a theoretically favored supersymmetric model of BSM physics that serves as motivation for many experimental searches. The MSSM presents a diverse range of possible phenomenologies and experimental signatures due to the large number of parameters that it introduces. Searching the MSSM parameter space is a computationally expensive task due to the overwhelming dimensionality of the space.

The phenomenological MSSM (pMSSM) is a lower dimensional subspace of the MSSM that neglects terms that produce tree level flavor changing neutral currents or tree level CP violation, and requires mass universality between first and second generation sparticle masses [32]. This dramatically reduces the dimensionality of the parameter space, requiring only 19 parameters to fully describe the theory. Though still computationally expensive to explore, the pMSSM has been scanned many times, most notably by the ATLAS collaboration in light of their 8 TeV data [103]. These searches have revealed large regions of the parameter space that remain unexcluded by current experimental constraints.

In this work, we determine which experiments have the greatest sensitivity to the remaining unexcluded regions of the pMSSM. Our focus is on collider experiments performed at the Large Hadron Collider (LHC). We first generate a set of pMSSM models that are consistent with the following constraints:

1. Higgs mass consistent with experiment to within 3 GeV [106, 107]
2. Dark matter relic density consistent with experiment to within .03 [108, 109]
3. The lightest supersymmetric particle (LSP) is a neutralino.

4. Consistent with all current collider constraints.

Inspired by parameter scans in the constrained MSSM [1], we develop a classification scheme which partitions the pMSSM parameter space according to the mechanism by which the experimentally observed dark matter relic density is attained. We then decompose models into their simplified model spectrum to determine which final state analyses are most relevant to a given model point. We report the final states of highest cross section in the decomposition that do not appear in the list of experimental results. We report these results for each class of models individually, so that we may study the dependence of the experiment's sensitivity on the type of pMSSM model.

5.2 Methods

5.2.1 Parameter Space Scan

We first sample 2×10^7 pMSSM models randomly from a uniform distribution with bounds given in Table 4.2. Each model is then analyzed using SoftSUSYv.4.1.10 and micrOmegasv5.2.6 [13, 99, 14]. From this scan, only models that possess a neutralino LSP, admit a consistent Higgs mass, and saturate the dark matter relic density are selected. Additional theoretical constraints are also implicitly applied at this step. Namely, the model is required to allow for consistent electroweak symmetry breaking and all squared masses must be positive. Models for which SoftSUSY is unable to converge are discarded.

Models that pass these constraints are then evaluated using SUSY-AI [15]. SUSY-AI is a machine learning model which has been trained to classify pMSSM models as consistent or inconsistent with the referenced ATLAS collider studies with 93% accuracy [117, 118, 119,

120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139]. In total, 9228 pMSSM models survive all constraints applied to the space.

For the 9228 surviving pMSSM models, we use SModelS to decompose the model into its simplified model spectrum [16, 140]. The cross section of each simplified model is computed by multiplying the production cross section by all branching ratios appearing in the simplified model. The production cross sections required for this decomposition were computed using Pythia 8 [11] and branching ratios are computed using SoftSUSY. Only strong production cross sections are considered. Simplified models are clustered according their final state and each final state is assigned a cross section by summing the cross sections of simplified models containing that final state. Final states that are invisible to experiment or that contain long lived sparticles are removed.

The final states presented here are the R parity even products of the decay of two sparticles. Decays of unstable SM particles are not considered when evaluating final states, and so the final states computed here may not correspond to the particles that would be directly measured in experiment. Currently, no kinematic information is considered in determining the cross section of the final state. For each model, we log the 20 final states with the largest cross section that are not covered in the ATLAS pMSSM study. We then rank the final states according to the average cross section over all models for which the final state appears.

5.2.2 Classification of pMSSM Models

We classify pMSSM models in terms of the mechanism by which the experimental dark matter relic density is attained. In large regions of the parameter space, the lightest neutralino takes on a nearly pure state. A nearly pure higgsino or wino LSP will strongly interact with other matter, and so will tend to undersaturate the relic density. On the other hand, a nearly pure bino LSP will generally only interact weakly with other matter, and so will tend

to oversaturate the relic density. An exception to this is when a mechanism is present which enhances interaction cross sections. Such mechanisms include a resonant s-channel annihilation cross-section (pole annihilation), a nearly degenerate next-to-lightest supersymmetric particle (NLSP) which allows for large LSP-NLSP interaction cross-sections (coannihilation), and a mixed LSP which contains non-negligible wino or higgsino components.

The classification scheme presented here is inspired by the classification of models in the cMSSM [1]. The cMSSM is typically partitioned into 5 classes:

- Z/h pole annihilation
- A_0 pole annihilation
- $\tilde{t}$ coannihilation
- $\tilde{\tau}$ coannihilation
- Well-tempered / Focus point

Pole annihilation classes are labelled by the mediator of the resonant s-channel annihilation process. Coannihilation classes are labelled by the nearly degenerate NLSP that coannihilates with the LSP [141, 142]. Well-tempered and focus point refers to models in which the LSP acquires a significant higgsino component [143, 116]. Since phenomenology varies greatly across each class and relatively little within a class, it is common to consider the phenomenology of each class in isolation.

We extend this concept to the pMSSM with the following classification:

- Z/h pole annihilation
- A_0 pole annihilation
- $\tilde{\nu}_1$ coannihilation
- $\tilde{\nu}_3$ coannihilation
- $\tilde{Q}_1$ coannihilation
- $\tilde{L}_1$ coannihilation
- $\tilde{u}_R$ coannihilation
- $\tilde{d}_R$ coannihilation
- $\tilde{e}_R$ coannihilation
- $\tilde{t}$ coannihilation
- $\tilde{b}$ coannihilation
- $\tilde{\tau}$ coannihilation
- $\tilde{W}$ coannihilation
- $\tilde{g}$ coannihilation
- $\tilde{\nu}_\tau\tilde{\tau}$ pair annihilation
- $\tilde{g}\tilde{u}_R$ pair annihilation
- $\tilde{g}\tilde{d}_R$ pair annihilation
- $\tilde{g}\tilde{Q}_1$ pair annihilation
- $\tilde{g}\tilde{t}$ pair annihilation
- $\tilde{g}\tilde{b}$ pair annihilation
- $\tilde{t}\tilde{b}$ pair annihilation
- $\tilde{B}/\tilde{H}$
- $\tilde{B}/\tilde{W}$

Pole annihilation and coannihilation classes are labelled as in the cMSSM. Because sparticle masses are less constrained in the cMSSM than in the pMSSM, there are more potential coannihilation partners present in the pMSSM. $\tilde{B}/\tilde{H}$ and $\tilde{B}/\tilde{W}$ refer to models where a bino LSP contains a higgsino or wino (respectively) admixture. Pair annihilation labels models where two sparticles are nearly degenerate with the each other so that interactions between them dominate contributions to the effective LSP annihilation cross section. A schematic of these mechanisms is shown in Figure 5.1.

Models are first classified according to the components of the LSP. Models containing LSPs that are less than 96% bino are classified as $\tilde{B}/\tilde{H}$ if the Higgsino component is larger than the wino component, and $\tilde{B}/\tilde{W}$ otherwise. The remaining models are classified by extracting the

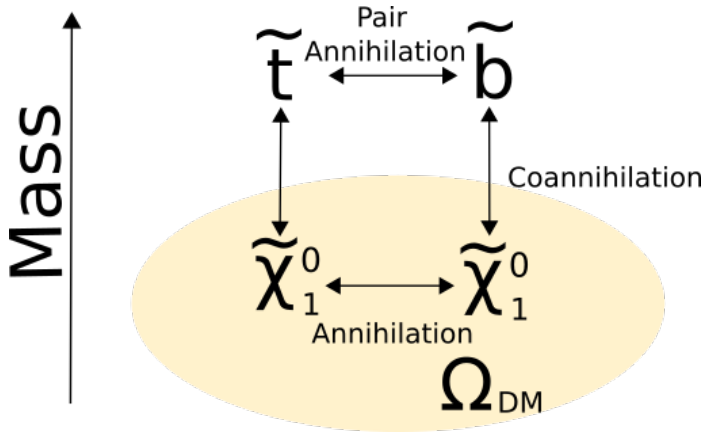

Figure 5.1: A schematic of dark matter mechanisms for bino LSPs presented here. Annihilation occurs between two neutralino LSPs. Coannihilation occurs between an LSP and a slightly more massive sparticle, in this case the $\tilde{t}$ or $\tilde{b}$ squarks. Pair annihilation occurs between two slightly more massive sparticles. This contributes to the effective LSP annihilation cross section as the more massive sparticles are unable to decay to neutralinos after pair annihilating to an SM final state.

normalized contribution of each channel to the effective LSP annihilation cross section from micrOmegas. The initial state of each channel is associated with a class from the above list. As an example, the initial state $\tilde{\chi}_1^0\tilde{t}$ is associated with stop coannihilation. The full list of initial states that define each class is available in Appendix C. At this step, pole annihilation is treated as a single class. For each model, the normalized contributions to the relic density of channels with initial states from the same class are summed and the model is assigned to the class for which this sum is the largest.

Pole annihilation models with LSP mass less than 70 GeV are labelled Z/h pole annihilation. Remaining pole annihilation models are labelled A_0 pole annihilation after verifying that the A_0 mass is within 200 GeV of twice the LSP mass. Large annihilation cross sections that do not satisfy either of these condition are due to t channel sfermion exchange.[1] We choose to label these diagrams as coannihilation, since t channel sfermion exchange only becomes

[1]A last alternative would be s channel exchange of a heavy Higgs H, however, this does not occur for any points in our data set. More precisely, all points for which $|2m_{\tilde{\chi}_0^1} - m_{A_0}| > 200$ GeV also satisfy $|2m_{\tilde{\chi}_0^1} - m_H| > 200$ GeV.

large when the mass of the sfermion is near that of the LSP, i.e. coannihilation processes are large.

This algorithm is summarized below. Performing classification in this way is advantageous as it avoids the need for arbitrary mass cutoffs and naturally handles regions of the parameter space where multiple mechanisms may overlap, which is more abundant in the pMSSM than the cMSSM due to the larger dimensionality. We have checked the consistency of this classification scheme with previous classification algorithms in the cMSSM and observe general agreement in Figure 5.2.

for θ *in dataset* **do**

  if $|Z_B(\theta)|^2 < .96$ **then**

    if $|Z_W(\theta)|^2 > |Z_{H_u}(\theta)|^2 + |Z_{H_d}(\theta)|^2$ **then**
      $c(\theta) = \mathrm{B/W}$

    else
      $c(\theta) = \mathrm{B/H}$

    end

  else

    for *each class* c **do**
      $\Sigma_c = Sum(\text{contributions from class } c \text{ to } \Omega_{DM}^{-1})$

    end

    $c(\theta) = argmax(\Sigma_c)$ **if** $c(\theta) = pole\_annihilation$ **then**

      if $m_{\chi_1^0} < 70\ GeV$ **then**
        $c(\theta) = \mathrm{Z/h}$ pole annihilation

      else if $|2m_{\chi_1^0} - m_{A^0}| < 200\ GeV$ **then**
        $c(\theta) = A^0$ pole annihilation

      else

        `/* t-channel sfermion exchange */`

        $c(\theta) =$ class with next largest Σ_c

      end

  end

end

5.3 Results

All results are presented for the set of models which pass the set of constraints previously listed. It is important to reiterate that the results here are a work in progress; they are

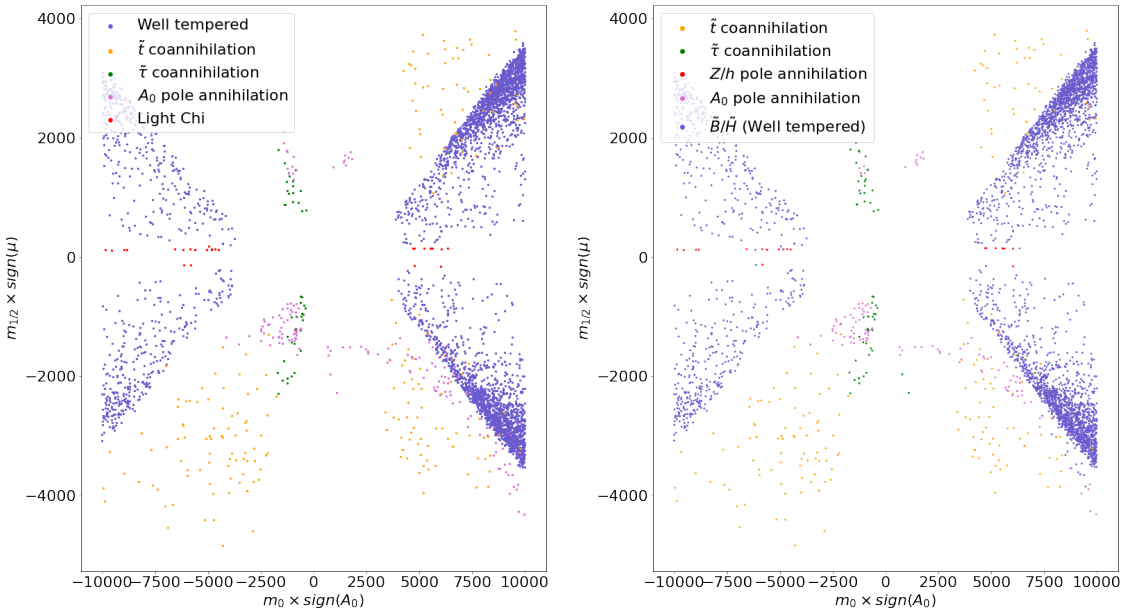

Figure 5.2: For a data set of constrained MSSM models, we plot a 2 dimensional slice in the parameter space as in Figure 2.1 of Reference [1]. As in the reference, models are required to possess the correct Higgs mass, relic density, a neutralino LSP, and charginos heavier than 100 GeV. On the left, we use the classification algorithm presented in the reference. On the right, we use the newly proposed classification algorithm. We see general agreement in the cMSSM. The new classification scheme easily generalizes to the pMSSM, whereas the classification scheme in the reference does not.

not peer reviewed and are likely to change in the future. It may be better to interpret this section as the form of results that we seek, rather than emphasizing a specific result that is likely to change as our methods improve.

5.3.1 Collider Studies

We present the final states of simplified models with the largest cross section that were not included in current experimental databases in Table 5.1. Results are presented for each class of models individually.

	Final State	σ_{BSM}	$N_{final\_state}/N_{class}$
$\tilde{W}$ coannihilation	MET b b jet jet	83.13 fb	17 / 24
	MET b b l	30.03 fb	16 / 24
	MET ta ta ta	14.65 fb	11 / 24
	MET jet jet ta ta	10.98 fb	16 / 24
	MET b b ta	10.17 fb	16 / 24
$\tilde{L}_1$ coannihilation	MET l	3.92 fb	3 / 18
	MET b b jet	2.44 fb	9 / 18
	MET b b jet jet	0.55 fb	4 / 18
	MET W b jet t	0.26 fb	3 / 18
	MET jet t t	0.21 fb	3 / 18
$\tilde{\nu}_\tau$ coannihilation	MET jet t t	1.24 fb	5 / 16
	MET b b jet	0.89 fb	7 / 16
	MET W b jet t	0.35 fb	3 / 16
	MET jet jet t t	0.16 fb	3 / 16
	MET b b jet jet	0.15 fb	3 / 16

$\tilde{g}$ coannihilation	MET b b jet jet	37.69 fb	32 / 64
	MET b b jet jet jet	8.51 fb	61 / 64
	MET b b jet jet jet jet	1.55 fb	38 / 64
	MET W jet jet jet	0.68 fb	4 / 64
	MET b b b b b b	0.50 fb	17 / 64
$\tilde{g}\tilde{u}_R$ pair annihilation	MET b jet jet jet	0.57 fb	1 / 2
	MET b b jet jet jet jet	0.30 fb	1 / 2
	MET W jet jet jet	0.12 fb	1 / 2
	MET W jet jet	0.07 fb	1 / 2
	MET higgs jet jet	0.01 fb	1 / 2
$\tilde{g}\tilde{t}$ pair annihilation	MET b b jet jet	11.37 fb	1 / 3
	MET b b jet jet jet	2.03 fb	2 / 3
	None		
	None		
	None		
$\tilde{g}\tilde{b}$ pair annihilation	MET b b jet	0.54 fb	2 / 2
	MET b b jet jet	0.11 fb	1 / 2
	MET W b b	0.05 fb	1 / 2
	None		
	None		

$\tilde{Q}_1$ coannihilation	MET b b jet	1.07 fb	75 / 179
	MET W W b b	0.97 fb	4 / 179
	MET b jet jet jet t	0.78 fb	4 / 179
	MET b b jet jet	0.76 fb	18 / 179
	MET Z Z b b	0.69 fb	5 / 179
$\tilde{u}_R$ coannihilation	MET jet jet ta ta	4.45 fb	7 / 137
	MET W jet jet	1.94 fb	29 / 137
	MET b b jet	1.23 fb	57 / 137
	MET jet t t	1.23 fb	48 / 137
	MET W W t t	1.08 fb	5 / 137
$\tilde{d}_R$ coannihilation	MET W jet jet	3.63 fb	37 / 155
	MET higgs jet jet	2.37 fb	20 / 155
	MET b jet jet t	1.95 fb	16 / 155
	MET W higgs jet	1.39 fb	6 / 155
	MET b b jet jet jet	1.28 fb	6 / 155
$\tilde{\ell}_R$ coannihilation	MET W jet jet	3.72 fb	4 / 17
	MET b jet jet t	1.62 fb	3 / 17
	MET Z jet jet	1.49 fb	4 / 17
	MET W b t	0.82 fb	4 / 17
	MET b b jet	0.59 fb	7 / 17
$\tilde{b}_1$ coannihilation	MET b b b jet jet t	21.77 fb	3 / 101
	MET W W b b	13.46 fb	14 / 101
	MET b b jet jet t t	12.79 fb	5 / 101
	MET W b b jet jet	8.29 fb	4 / 101
	MET b b jet t t	3.42 fb	3 / 101

$\tilde{t}_1$ coannihilation	MET b b jet jet jet	14.54 fb	8 / 105
	MET b b jet jet jet jet	9.65 fb	3 / 105
	MET b b jet	3.67 fb	27 / 105
	MET jet jet ta ta	2.67 fb	3 / 105
	MET b b jet jet	2.43 fb	19 / 105
$\tilde{\tau}_1$ coannihilation	MET jet jet ta ta	1.15 fb	7 / 27
	MET jet jet ta ta ta	0.81 fb	3 / 27
	MET b t ta	0.34 fb	4 / 27
	MET W ta ta	0.29 fb	3 / 27
	MET b jet t ta	0.21 fb	3 / 27
$\tilde{t}\tilde{b}$ pair annihilation	MET b b jet	2.04 fb	10 / 20
	MET b b jet jet	0.99 fb	7 / 20
	MET b b jet jet jet	0.87 fb	2 / 20
	MET b jet jet t	0.51 fb	4 / 20
	MET b b t t	0.03 fb	2 / 20
Z/h pole annihilation	MET W b t	820.82 fb	6 / 12
	MET W W b	79.63 fb	3 / 12
	MET W W W Z	12.45 fb	3 / 12
	MET W W W higgs	9.80 fb	3 / 12
	MET W W Z higgs	4.86 fb	3 / 12
A_0 pole annihilation	MET b b jet	3.83 fb	35 / 139
	MET jet t t	2.21 fb	20 / 139
	MET W jet jet	2.05 fb	38 / 139
	MET b jet jet t	2.05 fb	7 / 139
	MET b b jet jet	1.59 fb	14 / 139

$\tilde{B}/\tilde{H}$	MET W W W Z	30.97 fb	3 / 6369
	MET jet jet ta ta ta	26.34 fb	4 / 6369
	MET W W Z Z	19.29 fb	3 / 6369
	MET W W Z	17.63 fb	3 / 6369
	MET W jet jet ta ta	16.15 fb	3 / 6369
$\tilde{B}/\tilde{W}$	MET ta ta ta	2.67 fb	4 / 1120
	MET jet jet l ta	2.08 fb	11 / 1120
	MET b jet jet t ta ta	1.76 fb	3 / 1120
	MET b b l	1.50 fb	5 / 1120
	MET W Z jet jet jet jet	1.46 fb	4 / 1120

Table 5.1: Table of most relevant R parity even final states for each class of dark matter models ordered according to the BSM contributions to the final state. τ leptons in the final state are labelled ta. The last column of the table gives the number of models within the class for which the final state appears over the total number of models in the class.

As the methods used to attain these results are still evolving, we will simply provide general commentary rather than an in depth analysis of the specific results appearing here. A weakness of the results presented here is that the R parity even final state of the decay of two sparticles is not always equivalent to the final state seen in experiment, due to the instability of heavy SM particles.

For example, many classes are seen to be sensitive to final states containing W bosons. In experiment, W bosons decay in the beam pipe and may not be directly detected. Instead, they may detected through their leptonic decay channel, and so the final state may be covered by one of the ATLAS analyses. A specific instance of this is the MET W jet jet final state appearing as the third most relevant final state for A_0 pole annihilation models. In experiment, this may be detected as lepton + jets + MET, which is covered by ATLAS current ATLAS analyses [119].

Another weakness of these results is the lack of kinematic information. Large cross sections are seen for some processes, however, it is not known whether these events would be accessible at the LHC. For instance, final state particles may be too soft to be detected in experiment. Incorporating kinematic information into the results would greatly increase the utility of Table 5.1.

Many classes are sensitive to final states of 1-2 b quarks + jets + MET. As b quarks quickly hadronize, they are detected as jets and so these final states contribute to the jets + MET analyses [117, 118]. As only unexcluded models are shown, BSM contributions to the jets + MET final state could not have been significant enough to exclude the model. Thus, we include specifically 1-2 b + jets + MET final states in our results as the sensitivity to these channels may be increased by an analysis that specifically selects for the relevant number of b-tagged jets.

5.4 Conclusion

We have presented an analysis of pMSSM models containing a neutralino LSP, a Higgs boson mass consistent with experiment, a dark matter relic density that saturates the experimental value, and that are not excluded by analyses performed on 8 TeV ATLAS data. We have developed a classification scheme for these models based on the mechanism by which the experimental dark matter relic density is achieved. We have analyzed the sensitivity of these models to potential collider analyses and have presented the most abundant final states for each class of models that have not been already searched for in experiment.

There are numerous ways in which these results may be improved. First, the number of models available for some classes of dark matter is quite low. This can be improved by using a generative model to sample models satisfying the set of constraints considered here

[96]. Additionally, some of the models that are labelled as unexcluded by SUSY-AI may be deemed excluded by a full analysis, as the classification accuracy of SUSY-AI is only 93%. Using class labels derived from a Monte Carlo simulation, as is done by CheckMATE, would avoid this potential issue [17]. Another way in which these results could be improved is the inclusion of t channel interactions between decaying sparticles in simplified models, which would result in non-MET final states. Currently, the simplified model spectrum is computed by SModelS, which neglects these interactions.

Removing final states which decay to states already covered by ATLAS analyses would increase the utility of results. Currently, these final states may simply be skipped when reading Table 5.1. Another possible direction for future study would be to analyse the direct and indirect dark matter detection rates of all pMSSM models in the dataset, comparing to current and planned exclusion limits [144].

Bibliography

[1] Timothy Cohen and Jay G. Wacker. Here be dragons: the unexplored continents of the cmssm. *Journal of High Energy Physics*, 2013(9), Sep 2013.

[2] M. Cahill-Rowley, J. L. Hewett, A. Ismail, and T. G. Rizzo. pmssm studies at the 7, 8 and 14 tev lhc, 2013.

[3] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[4] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[5] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[6] The pandas development team. pandas-dev/pandas: Pandas, February 2020.

[7] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi,

Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[8] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. 2019.

[9] Johann Brehmer, Felix Kling, Irina Espejo, and Kyle Cranmer. MadMiner: Machine learning-based inference for particle physics. 2019.

[10] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[11] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. A brief introduction to pythia 8.1. *Computer Physics Communications*, 178(11):852 – 867, 2008.

[12] The DELPHES 3 collaboration, J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. Delphes 3: a modular framework for fast simulation of a generic collider experiment. *Journal of High Energy Physics*, 2014(2):57, Feb 2014.

[13] B.C. Allanach. Softsusy: A program for calculating supersymmetric spectra. *Computer Physics Communications*, 143(3):305–331, Mar 2002.

[14] G. Belanger, F. Boudjema, A. Pukhov, and A. Semenov. micromegas : a tool for dark matter studies, 2010.

[15] Sascha Caron, Jong Soo Kim, Krzysztof Rolbiecki, Roberto Ruiz de Austri, and Bob Stienen. The bsm-ai project: Susy-ai–generalizing lhc limits on supersymmetry with machine learning. *The European Physical Journal C*, 77(4):257, Apr 2017.

[16] Federico Ambrogi, Sabine Kraml, Suchita Kulkarni, Ursula Laa, Andre Lessa, Veronika Magerl, Jory Sonneveld, Michael Traub, and Wolfgang Waltenberger. Smodels v1.1 user manual: Improving simplified model constraints with efficiency maps. *Computer Physics Communications*, 227:72–98, Jun 2018.

[17] Manuel Drees, Herbi Dreiner, Jong Soo Kim, Daniel Schmeier, and Jamie Tattersall. Checkmate: Confronting your favourite new physics model with lhc data, 2014.

[18] Michael E Peskin and Daniel V Schroeder. *An introduction to quantum field theory*. Westview, Boulder, CO, 1995.

[19] Michael Ratz. Elementary particle physics for anteaters. 2018.

[20] Matthew D. Schwartz. Quantum Field Theory and the Standard Model. 2013.

[21] C. N. Yang and R. L. Mills. Conservation of isotopic spin and isotopic gauge invariance. *Phys. Rev.*, 96:191–195, Oct 1954.

[22] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579–588, 1961.

[23] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.

[24] Harald Fritzsch and Murray Gell-Mann. Current algebra: Quarks and what else? *eConf*, C720906V2:135–165, 1972.

[25] John Terning. TASI 2002 lectures: Nonperturbative supersymmetry. In *Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 2002): Particle Physics and Cosmology: The Quest for Physics Beyond the Standard Model(s)*, pages 343–443, 6 2003.

[26] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[27] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[28] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964.

[29] Stephen P. Martin. A Supersymmetry primer. *Adv. Ser. Direct. High Energy Phys.*, 21:1–153, 2010.

[30] Yuri Shirman. Introduction to supersymmetry and supersymmetry breaking. *The Dawn of the LHC Era*, Aug 2010.

[31] Howard E. Haber. The Status of the minimal supersymmetric standard model and beyond. *Nucl. Phys. B Proc. Suppl.*, 62:469–484, 1998.

[32] A. Djouadi et al. The Minimal supersymmetric standard model: Group summary report. In *GDR (Groupement De Recherche) - Supersymetrie*, 12 1998.

[33] A. H. Chamseddine, R. Arnowitt, and Pran Nath. Locally supersymmetric grand unification. *Phys. Rev. Lett.*, 49:970–974, Oct 1982.

[34] R. Barbieri, S. Ferrara, and C.A. Savoy. Gauge models with spontaneously broken local supersymmetry. *Physics Letters B*, 119(4):343–347, 1982.

[35] Lawrence Hall, Joe Lykken, and Steven Weinberg. Supergravity as the messenger of supersymmetry breaking. *Phys. Rev. D*, 27:2359–2378, May 1983.

[36] Nobuyoshi Ohta. GRAND UNIFIED THEORIES BASED ON LOCAL SUPERSYMMETRY. *Prog. Theor. Phys.*, 70:542, 1983.

[37] Michael Dine and Willy Fischler. A phenomenological model of particle physics based on supersymmetry. *Physics Letters B*, 110(3):227–231, 1982.

[38] Chiara R. Nappi and Burt A. Ovrut. Supersymmetric extension of the su(3)×su(2)×u(1) model. *Physics Letters B*, 113(2):175–179, 1982.

[39] M Claudson L Alvarez-Gaume and M B Wise. Low-energy supersymmetry, 1982.

[40] Lisa Randall and Raman Sundrum. Out of this world supersymmetry breaking. *Nuclear Physics B*, 557(1-2):79–118, Sep 1999.

[41] Gian F Giudice, Markus A Luty, Hitoshi Murayama, and Riccardo Rattazzi. Gaugino mass without singlets. *Journal of High Energy Physics*, (12), 12 1998.

[42] Daniele Alves, Nima Arkani-Hamed, Sanjay Arora, Yang Bai, Matthew Baumgart, Joshua Berger, Matthew Buckley, Bart Butler, Spencer Chang, Hsin-Chia Cheng, and et al. Simplified models for lhc new physics searches. *Journal of Physics G: Nuclear and Particle Physics*, 39(10):105005, Sep 2012.

[43] Thomas Gerard Rizzo. Z' phenomenology and the lhc. 2006.

[44] Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, Cambridge, MA, USA, 2016. `http://www.deeplearningbook.org`.

[45] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[46] J. Kiefer and J. Wolfowitz. Stochastic Estimation of the Maximum of a Regression Function. *The Annals of Mathematical Statistics*, 23(3):462 – 466, 1952.

[47] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, New York, NY, USA, second edition, 2006.

[48] T. Tieleman and G. Hinton. Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning, 2012.

[49] Matthew D. Zeiler. Adadelta: An adaptive learning rate method, 2012.

[50] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[51] Sashank J. Reddi, Satyen Kale, and Sanjiv Kumar. On the convergence of adam and beyond. In *International Conference on Learning Representations*, 2018.

[52] Henry J Kelley. Gradient theory of optimal flight paths. *Ars Journal*, 30(10):947–954, 1960.

[53] A. S. Fraser. Simulation of genetic systems by automatic digital computers i. introduction. *Australian Journal of Biological Sciences*, 10(4):484–491, 1957.

[54] J. Močkus. On bayesian methods for seeking the extremum. In G. I. Marchuk, editor, *Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7, 1974*, pages 400–404, Berlin, Heidelberg, 1975. Springer Berlin Heidelberg.

[55] Tom Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006. ROC Analysis in Pattern Recognition.

[56] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.

[57] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, page 448–456, 2015.

[58] Jacob Hollingsworth and Daniel Whiteson. Resonance searches with machine learned likelihood ratios, 2020.

[59] J. J. Aubert, U. Becker, P. J. Biggs, J. Burger, M. Chen, G. Everhart, P. Goldhagen, J. Leong, T. McCorriston, T. G. Rhoades, M. Rohde, Samuel C. C. Ting, Sau Lan Wu, and Y. Y. Lee. Experimental observation of a heavy particle j. *Phys. Rev. Lett.*, 33:1404–1406, Dec 1974.

[60] J. E. Augustin, A. M. Boyarski, M. Breidenbach, F. Bulos, J. T. Dakin, G. J. Feldman, G. E. Fischer, D. Fryberger, G. Hanson, B. Jean-Marie, R. R. Larsen, V. Lüth, H. L. Lynch, D. Lyon, C. C. Morehouse, J. M. Paterson, M. L. Perl, B. Richter, P. Rapidis, R. F. Schwitters, W. M. Tanenbaum, F. Vannucci, G. S. Abrams, D. Briggs, W. Chinowsky, C. E. Friedberg, G. Goldhaber, R. J. Hollebeek, J. A. Kadyk, B. Lulu, F. Pierre, G. H. Trilling, J. S. Whitaker, J. Wiss, and J. E. Zipse. Discovery of a narrow resonance in e^+e^- annihilation. *Phys. Rev. Lett.*, 33:1406–1408, Dec 1974.

[61] S. W. Herb et al. Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions. *Phys. Rev. Lett.*, 39:252–255, 1977.

[62] G. Arnison et al. Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c^2 at the CERN SPS Collider. *Phys. Lett.*, 126B:398–410, 1983.

[63] P. Bagnaia et al. Evidence for Z0 $\rightarrow$ e^+ e^- at the CERN anti-p p Collider. *Phys. Lett.*, 129B:130–140, 1983.

[64] G. Aad *et al.* (ATLAS Collaboration). Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1 – 29, 2012.

[65] S. Chatrchyan *et al.* CMS Collaboration. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30 – 61, 2012.

[66] C. P. Shen *et al.* (The Belle Collaboration). Evidence for a new resonance and search for the $y(4140)$ in the $\gamma\gamma \rightarrow \phi j/\psi$ process. *Phys. Rev. Lett.*, 104:112004, Mar 2010.

[67] T. Aaltonen *et al.* (CDF Collaboration). Search for new particles decaying into dijets in proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Rev. D*, 79:112002, Jun 2009.

[68] V. M. Abazov *et al.* (The D0 Collaboration). Measurement of dijet angular distributions at $\sqrt{s} = 1.96$ TeV and searches for quark compositeness and extra spatial dimensions. *Phys. Rev. Lett.*, 103:191803, Nov 2009.

[69] V. Khachatryan *et al.* (CMS Collaboration). Transverse-momentum and pseudorapidity distributions of charged hadrons in *pp* collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev. Lett.*, 105:022002, Jul 2010.

[70] Meghan Frate, Kyle Cranmer, Saarik Kalia, Alexander Vandenberg-Rodes, and Daniel Whiteson. Modeling Smooth Backgrounds and Generic Localized Signals with Gaussian Processes. 2017.

[71] P. Baldi, P. Sadowski, and D. Whiteson. Searching for exotic particles in high-energy physics with deep learning. *Nature Communications*, 5:4308 EP –, Jul 2014. Article.

[72] Anders Andreassen and Benjamin Nachman. Neural networks for full phase-space reweighting and parameter tuning, 2019.

[73] Anders Andreassen, Patrick T. Komiske, Eric M. Metodiev, Benjamin Nachman, and Jesse Thaler. Omnifold: A method to simultaneously unfold all observables, 2019.

[74] Jack Collins, Kiel Howe, and Benjamin Nachman. Anomaly detection for resonant new physics with machine learning. *Phys. Rev. Lett.*, 121:241803, Dec 2018.

[75] Jack H. Collins, Kiel Howe, and Benjamin Nachman. Extending the search for new resonances with machine learning. *Phys. Rev. D*, 99:014038, Jan 2019.

[76] Anders Andreassen, Benjamin Nachman, and David Shih. Simulation assisted likelihood-free anomaly detection, 2020.

[77] Benjamin Nachman and David Shih. Anomaly detection with density estimation, 2020.

[78] K. Kondo. Dynamical Likelihood Method for Reconstruction of Events With Missing Momentum. 1: Method and Toy Models. *J. Phys. Soc. Jap.*, 57:4126–4140, 1988.

[79] K. Kondo. Dynamical likelihood method for reconstruction of events with missing momentum. 2: Mass spectra for $2 \to 2$ processes. *J. Phys. Soc. Jap.*, 60:836–844, 1991.

[80] V. M. Abazov *et al.* (D0 Collaboration). A precision measurement of the mass of the top quark. *Nature*, 429(6992):638–642, 2004.

[81] Johann Brehmer, Kyle Cranmer, Gilles Louppe, and Juan Pavez. A guide to constraining effective field theories with machine learning. *Phys. Rev. D*, 98:052004, Sep 2018.

[82] Johann Brehmer, Kyle Cranmer, Gilles Louppe, and Juan Pavez. Constraining effective field theories with machine learning. *Phys. Rev. Lett.*, 121:111801, Sep 2018.

[83] Markus Stoye, Johann Brehmer, Gilles Louppe, Juan Pavez, and Kyle Cranmer. Likelihood-free inference with an improved cross-entropy estimator. 2018.

[84] A morphing technique for signal modelling in a multidimensional space of coupling parameters. Technical Report ATL-PHYS-PUB-2015-047, CERN, Geneva, Nov 2015.

[85] Johann Brehmer and Kyle Cranmer. Simulation-based inference methods for particle physics, 2020.

[86] David Curtin, Rouven Essig, Stefania Gori, and Jessie Shelton. Illuminating dark photons with high-energy colliders. *Journal of High Energy Physics*, 2015(2), Feb 2015.

[87] David Curtin, Rouven Essig, Stefania Gori, Prerit Jaiswal, Andrey Katz, Tao Liu, Zhen Liu, David McKeen, Jessie Shelton, Matthew Strassler, and et al. Exotic decays of the 125 gev higgs boson. *Physical Review D*, 90(7), Oct 2014.

[88] The ATLAS collaboration. Search for dark matter in events with a hadronically decaying vector boson and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. 2018.

[89] Morad Aaboud et al. Search for dark matter and other new phenomena in events with an energetic jet and large missing transverse momentum using the ATLAS detector. *JHEP*, 01:126, 2018.

[90] Albert M Sirunyan et al. Search for low-mass resonances decaying into bottom quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev.*, D99(1):012005, 2019.

[91] Albert M. Sirunyan et al. Searches for pair production of charginos and top squarks in final states with two oppositely charged leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 11:079, 2018.

[92] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-ktjet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, Apr 2008.

[93] Olivier Mattelaer. On the maximal use of Monte Carlo samples: re-weighting events at NLO accuracy. *Eur. Phys. J.*, C76(12):674, 2016.

[94] Pierre Baldi, Kyle Cranmer, Taylor Faucett, Peter Sadowski, and Daniel Whiteson. Parameterized neural networks for high-energy physics. *The European Physical Journal C*, 76(5):235, Apr 2016.

[95] Johann Brehmer, Siddharth Mishra-Sharma, Joeri Hermans, Gilles Louppe, and Kyle Cranmer. Mining for dark matter substructure: Inferring subhalo population properties from strong lenses with machine learning. *The Astrophysical Journal*, 886(1):49, nov 2019.

[96] Jacob Hollingsworth, Michael Ratz, Philip Tanedo, and Daniel Whiteson. Efficient sampling of constrained high-dimensional theoretical spaces with machine learning, 2021.

[97] Diptimoy Ghosh, Monoranjan Guchait, Sreerup Raychaudhuri, and Dipan Sengupta. How constrained is the constrained mssm? *Physical Review D*, 86(5), Sep 2012.

[98] Chengcheng Han, Ken-ichi Hikasa, Lei Wu, Jin Min Yang, and Yang Zhang. Status of cmssm in light of current lhc run-2 and lux data. *Physics Letters B*, 769:470–476, Jun 2017.

[99] B. C. Allanach. Impact of cms multi-jets and missing energy search on cmssm fits. *Physical Review D*, 83(9), May 2011.

[100] O. Buchmueller, R. Cavanaugh, A. De Roeck, M. J. Dolan, J. R. Ellis, H. Flächer, S. Heinemeyer, G. Isidori, J. Marrouche, D. Martínez Santos, K. A. Olive, S. Rogerson, F. J. Ronga, K. J. de Vries, and G. Weiglein. The cmssm and nuhm1 after lhc run 1. *The European Physical Journal C*, 74(6):2922, Jun 2014.

[101] Michael Bridges, Kyle Cranmer, Farhan Feroz, Mike Hobson, Roberto Ruiz de Austri, and Roberto Trotta. A coverage study of the cmssm based on atlas sensitivity using fast neural networks techniques. *Journal of High Energy Physics*, 2011(3), Mar 2011.

[102] M. Cahill-Rowley, J. L. Hewett, A. Ismail, and T. G. Rizzo. Lessons and prospects from the pmssm after lhc run i. *Phys. Rev. D*, 91:055002, Mar 2015.

[103] G. Aad, B. Abbott, J. Abdallah, O. Abdinov, R. Aben, M. Abolins, O. S. AbouZeid, H. Abramowicz, H. Abreu, and et al. Summary of the atlas experiment's sensitivity to supersymmetry after lhc run 1 — interpreted in the phenomenological mssm. *Journal of High Energy Physics*, 2015(10), Oct 2015.

[104] V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, and et al. Phenomenological mssm interpretation of cms searches in pp collisions at $\sqrt{s} = 7$ and 8 tev. *Journal of High Energy Physics*, 2016(10), Oct 2016.

[105] B.S. Kronheim, M.P. Kuchera, H.B. Prosper, and A. Karbo. Bayesian neural networks for fast susy predictions. *Physics Letters B*, 813:136041, Feb 2021.

[106] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A.A. Abdelalim, O. Abdinov, R. Aben, B. Abi, M. Abolins, and et al. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1–29, Sep 2012.

[107] S. Chatrchyan, V. Khachatryan, A.M. Sirunyan, A. Tumasyan, W. Adam, E. Aguilo, T. Bergauer, M. Dragicevic, J. Erö, C. Fabjan, and et al. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30–61, Sep 2012.

[108] D. N. Spergel, L. Verde, H. V. Peiris, E. Komatsu, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, and et al. First-year wilkinson microwave anisotropy probe (wmap) observations: Determination of cosmological parameters. *The Astrophysical Journal Supplement Series*, 148(1):175–194, Sep 2003.

[109] C. L. Bennett, D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, N. Odegard, K. M. Smith, R. S. Hill, B. Gold, M. Halpern, E. Komatsu, M. R. Nolta, L. Page, D. N. Spergel, E. Wollack, J. Dunkley, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright. NINE-YEAR WILKINSON MICROWAVE ANISOTROPY PROBE (WMAP) OBSERVATIONS: FINAL MAPS AND RESULTS. *The Astrophysical Journal Supplement Series*, 208(2):20, sep 2013.

[110] Peter Athron, Jae-hyeon Park, Tom Steudtner, Dominik Stöckinger, and Alexander Voigt. Precise higgs mass calculations in (non-)minimal supersymmetry at both high and low scales. *Journal of High Energy Physics*, 2017(1), Jan 2017.

[111] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.

[112] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. 2017.

[113] Radford M. Neal. Mcmc using hamiltonian dynamics, 2012.

[114] Michael Betancourt. A conceptual introduction to hamiltonian monte carlo, 2018.

[115] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. 2018.

[116] Jonathan L. Feng, Konstantin T. Matchev, and Takeo Moroi. Focus points and naturalness in supersymmetry. *Phys. Rev. D*, 61:075005, 2000.

[117] The ATLAS collaboration. Search for squarks and gluinos with the atlas detector in final states with jets and missing transverse momentum using $\sqrt{s} = 8$ tev proton-proton collision data. *Journal of High Energy Physics*, 2014(9):176, Sep 2014.

[118] The ATLAS collaboration. Search for new phenomena in final states with large jet multiplicities and missing transverse momentum at $\sqrt{s} = 8$ tev proton-proton collisions using the atlas experiment. 2013, October 2013.

[119] The ATLAS collaboration. Search for squarks and gluinos in events with isolated leptons, jets and missing transverse momentum at $\sqrt{s} = 8$ tev with the atlas detector. *Journal of High Energy Physics*, 2015(4):116, Apr 2015.

[120] The ATLAS collaboration. Search for supersymmetry in events with large missing transverse momentum, jets, and at least one tau lepton in 20 fb^{-1} of $\sqrt{s}$= 8 tev proton-proton collision data with the atlas detector. *Journal of High Energy Physics*, 2014(9):103, Sep 2014.

[121] The ATLAS collaboration. Search for supersymmetry at $\sqrt{s}$= 8 tev in final states with jets and two same-sign leptons or three leptons with the atlas detector. *Journal of High Energy Physics*, 2014(6):35, Jun 2014.

[122] The ATLAS collaboration. Search for strong production of supersymmetric particles in final states with missing transverse momentum and at least three b-jets at $\sqrt{s} = 8$ tev proton-proton collisions with the atlas detector. *Journal of High Energy Physics*, 2014(10):24, Oct 2014.

[123] The ATLAS collaboration. Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at $\sqrt{s} = 8$ tev with the atlas detector. *The European Physical Journal C*, 75(7):299, Jul 2015.

[124] The ATLAS collaboration. Search for direct top-squark pair production in final states with two leptons in pp collisions at $\sqrt{s}$= 8 tev with the atlas detector. *Journal of High Energy Physics*, 2014(6):124, Jun 2014.

[125] The ATLAS collaboration. Search for direct pair production of the top squark in all-hadronic final states in proton-proton collisions at $\sqrt{s}$= 8 tev with the atlas detector. *Journal of High Energy Physics*, 2014(9):15, Sep 2014.

[126] ATLAS Collaboration. Search for pair-produced third-generation squarks decaying via charm quarks or in compressed supersymmetric scenarios in *pp* collisions at $\sqrt{s} = 8$ tev with the atlas detector. *Phys. Rev. D*, 90:052008, Sep 2014.

[127] The ATLAS collaboration. Search for direct top squark pair production in events with a z boson, b-jets and missing transverse momentum in $\sqrt{s} = 8$ tev pp collisions with the atlas detector. *The European Physical Journal C*, 74(6):2883, Jun 2014.

[128] The ATLAS collaboration. Search for direct third-generation squark pair production in final states with missing transverse momentum and two b-jets in $\sqrt{s} = 8$ tev pp collisions with the atlas detector. *Journal of High Energy Physics*, 2013(10), Oct 2013.

[129] The ATLAS collaboration. Search for direct pair production of a chargino and a neutralino decaying to the 125 gev higgs boson in $\sqrt{s} = 8$ tev pp collisions with the atlas detector. *The European Physical Journal C*, 75(5):208, May 2015.

[130] ATLAS Collaboration. Search for supersymmetry in events with four or more leptons in $\sqrt{s} = 8$ tev *pp* collisions with the atlas detector. *Phys. Rev. D*, 90:052001, Sep 2014.

[131] The ATLAS collaboration. Search for the direct production of charginos, neutralinos and staus in final states with at least two hadronically decaying taus and missing transverse momentum in pp collisions at $$ \sqrt{s} = 8$$ $tevwiththeatlasdetector. Journal of High Energy Physics, 2014(10) : 96, Oct 2014$.

[132] The ATLAS Collaboration. Search for charginos nearly mass degenerate with the lightest neutralino based on a disappearing-track signature in *pp* collisions at $\sqrt{s}$=8 tev with the atlas detector. *Phys. Rev. D*, 88:112006, Dec 2013.

[133] The ATLAS Collaboration. Searches for heavy long-lived sleptons and r-hadrons with the atlas detector in pp collisions at s=7 tev. *Physics Letters B*, 720(4):277–308, 2013.

[134] The ATLAS Collaboration. Searches for heavy long-lived charged particles with the atlas detector in proton-proton collisions at $\sqrt{s} = 8$ tev. *Journal of High Energy Physics*, 2015(1):68, Jan 2015.

[135] The ATLAS Collaboration. Search for neutral higgs bosons of the minimal supersymmetric standard model in pp collisions at $\sqrt{s} = 8$ tev with the atlas detector. *Journal of High Energy Physics*, 2014(11):56, Nov 2014.

[136] The ATLAS Collaboration. Search for direct production of charginos and neutralinos in events with three leptons and missing transverse momentum in $\sqrt{s}$= 8 tev pp collisions with the atlas detector. *Journal of High Energy Physics*, 2014(4):169, Apr 2014.

[137] The ATLAS Collaboration. Search for direct production of charginos, neutralinos and sleptons in final states with two leptons and missing transverse momentum in pp collisions at $\sqrt{s}$= 8tev with the atlas detector. *Journal of High Energy Physics*, 2014(5):71, May 2014.

[138] The ATLAS Collaboration. Search for top squark pair production in final states with one isolated lepton, jets, and missing transverse momentum in $\sqrt{s}$= 8 tev pp collisions with the atlas detector. *Journal of High Energy Physics*, 2014(11):118, Nov 2014.

[139] The ATLAS Collaboration. Atlas run 1 searches for direct pair production of third-generation squarks at the large hadron collider. *The European Physical Journal C*, 75(10), Oct 2015.

[140] Federico Ambrogi, Sabine Kraml, Suchita Kulkarni, Ursula Laa, Andre Lessa, and Wolfgang Waltenberger. On the coverage of the pmssm by simplified model results. *The European Physical Journal C*, 78(3), Mar 2018.

[141] P. Binetruy, G. Girardi, and P. Salati. Constraints on a System of Two Neutral Fermions From Cosmology. *Nucl. Phys. B*, 237:285–306, 1984.

[142] Kim Griest and David Seckel. Three exceptions in the calculation of relic abundances. *Phys. Rev. D*, 43:3191–3203, May 1991.

[143] Jonathan L. Feng, Konstantin T. Matchev, and Takeo Moroi. Multi - TeV scalars are natural in minimal supergravity. *Phys. Rev. Lett.*, 84:2322–2325, 2000.

[144] E. Aprile, J. Aalbers, F. Agostini, M. Alfonsi, L. Althueser, F.D. Amaro, M. Anthony, V.C. Antochi, F. Arneodo, L. Baudis, and et al. Constraining the spin-dependent wimp-nucleon cross sections with xenon1t. *Physical Review Letters*, 122(14), Apr 2019.

Appendix A

Appendix

A.1 Appendix A

We justify the comparison of the expected value of the machine learned log likelihood ratio within each invariant mass bin used in Chapter 3. Consider a test set $\mathcal{X}_{\text{test}}$ of N events drawn from the distribution $p_{test}(x)$ where N is large. Using only invariant mass (denoted m_{jj}) to find the log likelihood ratio, we have

$$\log r(\mathcal{X}_{\text{test}}, \theta, \theta_0) = N \int dm_{jj}\, p_{test}(m_{jj}) \log r(m_{jj}|\theta, \theta_0). \tag{A.1}$$

A binned form of the expression $\log r(m_{jj}|\theta, \theta_0)$ is plotted as the grey line in Figure 3.2.

Using higher dimensional observations $x = (m_{jj}, x')$ to find the log likelihood ratio, we instead have:

$$\log r(\mathcal{X}_{\text{test}}, \theta, \theta_0) = N \int dx\, p_{test}(x) \log r(x|\theta, \theta_0),$$

which can be manipulated to take the same form as Equation A.1:

$$\log r(\mathcal{X}_{\text{test}}, \theta, \theta_0) = N \int dm_{jj}\, p_{test}(m_{jj}) \Big[\int dx'\, p(x'|m_{jj}) \log r(x', m_{jj}|\theta, \theta_0) \Big]. \tag{A.2}$$

The bracketed expression is approximated by taking the expectation within bins of invariant mass using the same binning as the previous case. This is plotted as the blue and magenta lines in Figure 3.2. We see that the bracketed expression is analogous to $\log r(m_{jj}|\theta, \theta_0)$ when using higher dimensional data to calculate the log likelihood ratio.

A.2 Appendix B

We present the hyperparameters for our machine learning models found in Chapter 4 in Table A.1.

A.3 Appendix C

For each class of pMSSM models introduced in Chapter 5, we list the initial states of contributing channels to the effective LSP annihilation cross section that are included in this class.

	Parameter	cMSSM	pMSSM
Supervised NN	Learning rate	0.001	0.0001245
	Hidden layers	5	10
	Nodes per layer	49	154
	Dropout	0.5	0.0
	Activation function	Sigmoid	Sigmoid
	Optimizer	ADAM	ADAM
	Batch size	128	128
	Epochs	50	50
HMC	Step size	0.025	0.008
	Number of steps	12	12
	Mass	1.0	1.0
	Chain length	5000	5000
	Burn-in steps	1000	1000
	Number of chains	100	100
NF	Num transforms	3	3
	Batch size	1024	1024
	Epochs	300	300
	B	2.0	2.0
	NN hidden features	64	64

Table A.1: Hyperparameters used for the machine learning models for to the cMSSM and pMSSM datasets.

Class	Inital States
Pole Annihilation	$\tilde{\chi}_1^0\tilde{\chi}_1^0$
$\tilde{W}$ Coannihilation	$\tilde{\chi}_1^0\tilde{\chi}_2^0$, $\tilde{\chi}_1^0\tilde{\chi}_1^\pm$, $\tilde{\chi}_1^\pm\tilde{\chi}_1^\pm$, $\tilde{\chi}_1^\pm\tilde{\chi}_2^0$
$\tilde{L}_1$ Coannihilation	$\tilde{\nu}_e\tilde{\nu}_e$, $\tilde{\nu}_\mu\tilde{\nu}_\mu$, $\tilde{e}_L\tilde{e}_L$, $\tilde{\mu}_L\tilde{\mu}_L$, $\tilde{\nu}_e\tilde{\nu}_m u$, $\tilde{\nu}_e\tilde{e}_L$, $\tilde{\nu}_e\tilde{\mu}_L$, $\tilde{\nu}_\mu\tilde{e}_L$, $\tilde{\nu}_\mu\tilde{\mu}_L$, $\tilde{e}_L\tilde{\mu}_L$, $\tilde{\chi}_1^0\tilde{\nu}_e$, $\tilde{\chi}_1^0\tilde{\nu}_\mu$, $\tilde{\chi}_1^0\tilde{e}_L$, $\tilde{\chi}_1^0\tilde{\mu}_L$
$\tilde{\nu}_\tau$ Coannihilation	$\tilde{\nu}_\tau\tilde{\nu}_\tau$, $\tilde{\chi}_1^0\tilde{\nu}_\tau$
$\tilde{g}$ Coannihilation	$\tilde{g}\tilde{g}$, $\tilde{\chi}_1^0\tilde{g}$
$\tilde{Q}_1$ Coannihilation	$\tilde{u}_L\tilde{u}_L$, $\tilde{c}_L\tilde{c}_L$, $\tilde{d}_L\tilde{d}_L$, $\tilde{s}_L\tilde{s}_L$, $\tilde{u}_L\tilde{d}_L$, $\tilde{u}_L\tilde{c}_L$, $\tilde{u}_L\tilde{s}_L$, $\tilde{d}_L\tilde{c}_L$, $\tilde{d}_L\tilde{s}_L$, $\tilde{c}_L\tilde{s}_L$, $\tilde{\chi}_1^0\tilde{u}_L$, $\tilde{\chi}_1^0\tilde{c}_L$, $\tilde{\chi}_1^0\tilde{d}_L$, $\tilde{\chi}_1^0\tilde{s}_L$
$\tilde{u}_R$ Coannihilation	$\tilde{u}_R\tilde{u}_R$, $\tilde{u}_R\tilde{c}_R$, $\tilde{c}_R\tilde{c}_R$, $\tilde{\chi}_1^0\tilde{u}_R$, $\tilde{\chi}_1^0\tilde{c}_R$
$\tilde{d}_R$ Coannihilation	$\tilde{d}_R\tilde{d}_R$, $\tilde{d}_R\tilde{s}_R$, $\tilde{s}_R\tilde{s}_R$, $\tilde{\chi}_1^0\tilde{d}_R$, $\tilde{\chi}_1^0\tilde{s}_R$
$\tilde{\ell}_R$ Coannihilation	$\tilde{e}_R\tilde{e}_R$, $\tilde{e}_R\tilde{\mu}_R$, $\tilde{\mu}_R\tilde{\mu}_R$, $\tilde{\chi}_1^0\tilde{e}_R$, $\tilde{\chi}_1^0\tilde{\mu}_R$
$\tilde{b}$ Coannihilation	$\tilde{b}\tilde{b}$, $\tilde{\chi}_1^0\tilde{b}$
$\tilde{t}$ Coannihilation	$\tilde{t}\tilde{t}$, $\tilde{\chi}_1^0\tilde{t}$
$\tilde{\tau}$ Coannihilation	$\tilde{\tau}\tilde{\tau}$, $\tilde{\chi}_1^0\tilde{\tau}$
$\tilde{\nu}_\tau\tilde{\tau}$ Pair Annihilation	$\tilde{\tau}\tilde{\nu}_\tau$
$\tilde{g}\tilde{u}_R$ Pair Annihilation	$\tilde{g}\tilde{u}_R$, $\tilde{g}\tilde{c}_R$
$\tilde{g}\tilde{d}_R$ Pair Annihilation	$\tilde{g}\tilde{d}_R$, $\tilde{g}\tilde{s}_R$
$\tilde{g}\tilde{Q}_1$ Pair Annihilation	$\tilde{g}\tilde{u}_L$, $\tilde{g}\tilde{c}_L$, $\tilde{g}\tilde{d}_L$, $\tilde{g}\tilde{s}_L$
$\tilde{g}\tilde{t}$ Pair Annihilation	$\tilde{g}\tilde{t}_1$
$\tilde{g}\tilde{b}$ Pair Annihilation	$\tilde{g}\tilde{b}_1$
$\tilde{t}\tilde{b}$ Pair Annihilation	$\tilde{t}_1\tilde{b}_1$

Table A.2: Initial states of contributing channels to DM relic density associated with each class of pMSSM models. No differentiation is made between particle and antiparticle states.