CERN-THESIS-2005-011
11/01/2005



SEARCHING FOR NEW PHYSICS: CONTRIBUTIONS TO LEP AND THE LHC

by

Kyle S. Cranmer

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Physics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2005

© Copyright by Kyle S. Cranmer 2005

Some Rights Reserved

This work is licensed under the Creative Commons Attribution-ShareAlike License. To view a copy of this license, visit http://creativecommons.org/licenses/by-sa/2.0/ or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

To my parents, Joan and Morris,
and my loving wife, Danielle.

ACKNOWLEDGMENTS

This work could not have come to fruition without the aid and guidance of many people. I am sincerely thankful and deeply indebted to the following people.

I am forever grateful to my wife, Danielle, who has made huge sacrifices in her life to follow me to the other side of the world. She has been loving and patient beyond measure.

I would like to thank my family for their influence was core to my development. My father, Morris, always the scientist, instilled my curiosity for the way things work and amazed me with his plethora of answers. Perhaps it was his inability to answer some questions that really drove me to physics. My mother, Joan, taught me to be intuitive and observant, important qualities for a physicist. Both of my parents have been incredibly supportive of my education, and I am grateful for that. Dylan, my older brother, pursued his dreams with sacrifice and dedication that I have always admired.

A few key teachers and professors have had great influence on my career. Reaching far into the past, Jim Gunnell's reading of *A Brief History of Time* to our class was a watershed event. Irina Lyublinskaya was my first physics teacher, taking me from Newton's laws to Einstein and the Schrödinger equation. At Rice University, Professors Hannu Miettinen and Paul Stevenson introduced me to research in experimental and theoretical particle physics, respectively. Without their key contributions, I would not have made it to where I am today.

I have two good friends that I have known since high school and collaborated with on many projects: R. Sean Bowman and Stephen McCaul. Sean introduced me to Genetic Programming, and he and I wrote PHYSICSGP about a year later. Stephen McCaul essentially taught me how to program. He guided me in the design of my first C++ package (multivariate kernel estimation).

Years later, we collaborated on a library of multivariate visualization and analysis tools. I am grateful to both of them for their friendship and their guidance.

One of my major interests is the statistical interpretation of experimental physics results. My original exposure to this field was with Hannu Miettinen, Bruce Knuteson, and Daniel Whiteson at Rice University. During the LEP era, I learned the LEP statistical procedure from Hongbo Hu, Peter McNamara, and Jason Nielsen. The development of *KEYS* was largely influenced by members of the LEP Higgs Working Group, most notably Chris Tully, Steve Armstrong, Peter McNamara, Jason Nielsen, Arnulf Quadt, Tom Junk, Peter Igo-Kemenes, Tom Greening, and Yuanning Gao. After developing the KEYS package, I met Fred James, who would later become a good friend and mentor. In 2003, I met Louis Lyons who taught me the Neyman construction and encouraged my research in frequentist hypothesis testing with background uncertainty. This was followed by many useful discussions with Bob Cousins and Gary Feldman.

During my original stay at CERN, I enjoyed working with Yuanning Gao, Julian von Wimmersperg, and Yibin Pan on General Search strategies, which was later translated into my interest in QUAERO and VISTA. Steve Armstrong and Jason Nielsen were wonderful officemates that year. Finally, I would be remiss if I did not specifically thank John Walsh for being a wonderful mentor during my sojourn with BaBar.

While in Madison, my fellow graduate student Bill Quayle and I had the pleasure of working Dieter Zeppenfeld on the MadCUP project. This was the beginning of a long and fruitful collaboration, which has extended to Tilman Plehn and David Rainwater.

In the last two years in Genève, I have enjoyed the camaraderie of the Wisconsin group led by Sau Lan Wu: André dos Anjos, Yaquan Fang, Luis Roberto Flores Castillo, Saúl González, Karina Loureiro, Bruce Mellado, Stathes Paganis, Bill Quayle, Alden Stradling, Werner Wiedenmann, and Hiamo Zobernig. Stathes and Werner, in particular, have been valuable resources and pleasures to discuss physics with. I am grateful for the assistance from Annabelle Leung, Alden Stradling, and Neng Xu in the generation of Monte Carlo for this dissertation. I would also like to give a warm thanks to Antonella Lofranco and Catharine Noble who both provided Danielle and I with endless assistance.

I would like to thank many members of the ATLAS collaboration for their assistance in recent years. In particular, I would like to thank Donatella Cavalli for her advice and efforts to address my $\not{p}_T$ problems; Peter Sherwood for introducing me to the ATLAS software; Michael Heldman, Frank Paige, and David Rousseau for their help with ATLAS reconstruction; Elzbieta Richter-Was, Karl Jakobs, and Guillaume Unal for their advice in Higgs searches; and Kétévi Assamagan, Peter Loch, Tadashi Maeno, David Quarrie, and Srini Rajagopalan for their advice in my contributions to ATLAS's analysis model.

During PhyStat2003, I met with my friend Bruce Knuteson who encouraged me to develop the ALEPH interface to QUAERO. This project became a natural extension of my interest in new particle searches and statistics and provided an excellent opportunity to have real data in my thesis. Without Bruce's persistence, this project would not have been realized, and I thank him for that. I would also like to thank Marcello Maggi, Jason Nielsen, Günther Dissertori, Roberto Tenchini, Steve Armstrong, and Patrick Janot for sharing their expertise during the development of the ARCH program.

Of course, none of the research presented in this thesis would have been possible without funding. This work was supported by a graduate research fellowship from the National Science Foundation and US Department of Energy Grant DE-FG0295-ER40896.

Finally, I would like to thank my adviser Sau Lan Wu. Her tireless pursuit for new physics and unwavering support for her group are admirable. It has been a privilege to be in her group and a huge advantage to be based at CERN.

I would like to use one page to point out that if this dissertation was not double-spaced, then it would be fifty pages shorter. This is practice is silly and antiquated in the era of LaTeX typesetting.

DISCARD THIS PAGE

TABLE OF CONTENTS

Page

LIST OF TABLES x

LIST OF FIGURES xi

NOMENCLATURE xviii

ABSTRACT xx

1 Introduction 1

2 Motivation 2

2.1 The Standard Model 3
2.2 Phenomenology 6
2.3 The Phenomenology of the Standard Model Higgs 9
2.4 Results from LEP Higgs Searches 10
2.5 Beyond the Standard Model 14

I Searching For New Physics at LEP **16**

3 The Aleph Detector at LEP 17

3.1 The Large Electron Positron Collider 17
3.2 The Aleph Detector 17

4 Vista@Aleph 20

4.1 Data 21
4.2 Particle Identification 22
4.3 Backgrounds 24
4.4 Comparison of Data and Standad Model Predictions 25
4.5 The $e^{\mp}\mu^{\pm}$ Final State 33

Page

5 Quaero@Aleph . . . 36

5.1 The Quaero Algorithm . . . 36
5.2 TurboSim@Aleph . . . 38
5.3 Systematic Errors . . . 41
5.4 Statistical Interpretation of Quaero Results . . . 41
5.5 Searches Performed with Quaero . . . 43
5.5.1 mSUGRA . . . 44
5.5.2 Excited electrons . . . 47
5.5.3 Doubly charged Higgs . . . 48
5.5.4 Charged Higgs . . . 50
5.5.5 Standard Model Higgs . . . 51
5.6 Summary . . . 52

6 Observations and Conclusions from LEP . . . 53

6.1 Influence of LEP on Preparation for the LHC . . . 53
6.2 Potential for Vista and Quaero at the LHC . . . 54

II Preparing for New Physics at the LHC **56**

7 The ATLAS Detector at the LHC . . . 57

7.1 The Large Hadron Collider at CERN . . . 57
7.1.1 Pile-Up . . . 58
7.1.2 Underlying Event . . . 58
7.2 The ATLAS Detector . . . 59
7.2.1 The Magnet System . . . 59
7.2.2 The Inner Detector . . . 61
7.2.3 Calorimetry . . . 62
7.2.4 The Muon System . . . 64
7.2.5 Trigger and Data Acquisition . . . 66
7.2.6 Fast and Full Simulation of the ATLAS Detector . . . 68

8 Monte Carlo Development for Vector Boson Fusion . . . 71

8.1 The MadCUP Event Generators . . . 72
8.2 Color Coherence . . . 72
8.3 Validation of Color Coherence in External User Processes . . . 74

Appendix

Page

9 Missing Transverse Momentum Reconstruction . . . 77

9.1 Components of $\not{p}_T$ Resolution . . . 77
9.1.1 Calorimeter Response . . . 79
9.1.2 Electronic Noise . . . 79
9.1.3 Geometrical Acceptance . . . 81
9.2 The H1-Style Calibration Strategy . . . 82
9.3 Electronic Noise and Bias . . . 82
9.3.1 Evidence for Bias . . . 83
9.3.2 When Symmetric Cuts Are Asymmetric . . . 84
9.3.3 When Asymmetric Cuts Are Symmetric . . . 85
9.3.4 Local Noise Suppression . . . 85
9.3.5 Estimating the Prior . . . 87
9.3.6 Comparison of Local and Global Noise Suppression . . . 88

10 Vector Boson Fusion $H \to \tau\tau$ . . . 90

10.1 Experimental Signature . . . 90
10.2 Identification of Hadronically Decaying Taus . . . 91
10.3 Electron Identification . . . 92
10.4 Muon Identification . . . 93
10.5 Jet Finding . . . 94
10.6 The Collinear Approximation . . . 96
10.6.1 Jacobian for $M_{\tau\tau}$ . . . 97
10.6.2 A Maximum Likelihood Approach . . . 98
10.7 Central Jet Veto . . . 100
10.8 Background Determination from Data . . . 102
10.9 A Cut-Based Analysis with Fast Simulation . . . 103
10.9.1 Signal and Background Generation . . . 103
10.9.2 List of Cuts . . . 104
10.9.3 Results with Fast Simulation . . . 106
10.10 A Cut-Based Analysis with Full Simulation . . . 108
10.10.1 Signal and Background Generation . . . 108
10.10.2 Results with Full Simulation . . . 109

11 Comparison of Multivariate Techniques for VBF $H \to WW^*$ . . . 112

11.1 Variables . . . 114
11.2 Neural Network Analysis . . . 114
11.2.1 Stability of Results to Different Background Descriptions . . . 115

Appendix

Page

11.3 Support Vector Regression . . . 115
11.4 Genetic Programming . . . 117
11.5 Comparison of Multivariate Methods . . . 117

12 $H \to \gamma\gamma$ Coverage Studies . . . 120

12.1 Systematics for $H \to \gamma\gamma$ . . . 120
12.2 Frequentist Result . . . 122
12.3 Impact of Systematics . . . 124
12.4 Statement on Original Material . . . 124

13 ATLAS Sensitivity to Standard Model Higgs . . . 126

13.1 Channels Considered . . . 126
13.2 Combined Significance . . . 128
13.3 Luminosity Plots . . . 129
13.4 The Power of a 5σ Test . . . 130
13.5 LEP-Style $-2 \ln Q$ vs. m_H Plots . . . 131
13.6 Conclusions . . . 133

LIST OF REFERENCES . . . 135

APPENDICES

Appendix A: Moving LEP-Style Statistics to the LHC . . . 145
Appendix B: Kernel Estimation Techniques . . . 162
Appendix C: Hypothesis Testing with Background Uncertainty . . . 175
Appendix D: Statistical Learning Theory Applied to Searches . . . 188
Appendix E: Genetic Programming for Event Selection . . . 195
Appendix F: The ATLAS Analysis Model . . . 206

DISCARD THIS PAGE

LIST OF TABLES

Table Page

2.1 The fermions of the Standard Model grouped according to family. No right handed neutrinos are included. Braces indicate weak isospin doublets. 5

4.1 Integrated luminosity of the data available in QUAERO@ALEPH for each nominal LEP 2 center of mass energy. 21

4.2 Number of events and observed for $e^{\pm}\mu^{\mp}$ and $e^{\pm}\mu^{\mp}p^{\rm miss}$ final states. 33

9.1 Tabulated values of the ratio $\sigma^{N\Delta}_{noise}/\sigma_{noise}$ in percent, where $\sigma^{N\Delta}_{noise}$ represents the contribution to the $\not{p}_T$ resolution after an $N\Delta$ noise threshold is applied. The quantities f_{sym} and f_{asym} correspond to the symmetric and asymetric cases, respectively. 80

10.1 Cross sections for the signal generated with PYTHIA6.203 104

10.2 Expected number of signal events, background events, and significance with 30 $\rm fb^{-1}$ for various masses. 106

10.3 Signal and Background effective cross-sections after various cuts for $m_H = 130$ GeVwith full simulation. The QCD Zjj background has been scaled by 1.25 to account for the final Electroweak component from fast simulation. 110

11.1 Effective cross-section by channel for each background processes after preselection. . 113

11.2 Expected significance for two cut analyses and three multivariate analyses for different Higgs masses and final state topologies. 118

12.1 Results of the $H \to \gamma\gamma$ coverage study (see text). 125

C.1 The notation used by Kendall for likelihood tests with nuisance parameters 179

DISCARD THIS PAGE

LIST OF FIGURES

Figure Page

2.1 Higgs branching ratios as a function of m_H from M. Spira Fortsch. Phys. 46 (1998) . . 9

2.2 Tree level Feynman diagrams for the Higgsstrahlung (left) and Vector Boson Fusion (right) Higgs production mechanisms from e^+e^- interactions. 10

2.3 Left: the cross section for $e^+e^- \to HZ$ as a function of $\sqrt{s}$ for several Higgs masses as obtained with the HZHA generator. Right: the cross section for $pp \to H + X$ as a function of M_H from M. Spira Fortsch. Phys. 46 (1998). 11

2.4 Feynman diagrams for the Higgs production at the LHC. 11

2.5 The expected and observed evolution of $-2\ln Q$ (left) and CL_s (right) with m_H from all LEP experiments. 12

2.6 The resulting $\Delta\chi^2$ vs m_H from LEP electroweak fits. 13

2.7 Quadratically divergent diagram in Higgs self energy. 13

3.1 An illustration of the LEP tunnel. 18

3.2 An illustration of the ALEPH detector. 18

4.1 A subset of the VISTA@ALEPH comparison of data and Monte Carlo. 27

4.2 A subset of the VISTA@ALEPH comparison of data and Monte Carlo. 28

4.3 A subset of the VISTA@ALEPH comparison of data and Monte Carlo. 29

4.4 A subset of the VISTA@ALEPH comparison of data and Monte Carlo. 30

4.5 A subset of the VISTA@ALEPH comparison of data and Monte Carlo. 31

4.6 Distribution of data-Monte Carlo discrepancy in terms of Gaussian σ (left) and background confidence-level, CL_b (right). The solid curves show the expectated distribution. 32

Figure Page

4.7 Kinematic distributions for the $e^-\mu^+$ final state. 34

4.8 Event displays for two events mis-reconstructed in the $e^\pm\mu^\mp$ final state. 35

5.1 QUAERO's automatic variable selection and choice of binning in the final state $j\not{p}\tau^+$, testing the hypothesis of mSUGRA with $\mu > 0$, $A_0 = 0$, $\tan(\beta) = 10$, $M_0 = 100$, and $M_{1/2} = 120$, using data collected at 205 GeV. 37

5.2 Ten sample lines in the TURBOSIM@ALEPH lookup table, chosen to illustrate TURBOSIM's handling of interesting cases. 40

5.3 Comparison of the output of TURBOSIM@ALEPH (light, green) and ALEPHSIM (dark, red). 42

5.4 Plots of the standard model prediction (dark, red), the querying physicist's hypothesis (light, green), and the ALEPH data (filled circles) for the single most useful variable in all final states contributing more than 0.1 to $\log_{10} Q$. 46

5.5 QUAERO's output ($\log_{10} Q$) as a function of assumed $M_{1/2}$ and M_0, for fixed $\tan\beta = 10$, $A_0 = 0$, and $\mu > 0$. 47

5.6 QUAERO's output ($\log_{10} Q$) as a function of assumed Λ and m_{e^*}, for fixed $f = f' \tan^2\theta_W = 0.28$ and $f_s = 0$ (left). Exclusion contour summarizing a previous OPAL analysis of excited lepton parameter space (right). 49

5.7 QUAERO's output ($\log_{10} Q$) as a function of assumed doubly charged Higgs mass $m_{H^{\pm\pm}}$, in the context of a left-right symmetric model containing a Higgs triplet (left). A previous OPAL analysis is also shown (right). 49

5.8 QUAERO's output ($\log_{10} Q$) as a function of assumed charged Higgs mass $m_{H^\pm}$, in the context of a generic two Higgs doublet model (left). A previous ALEPH result (right). 50

5.9 QUAERO's output ($\log_{10} Q$) as a function of assumed Standard Model Higgs mass m_H (left). Distributions of $-2\ln Q$ from the combined LEP Higgs search. 51

7.1 The LEP tunnel after modifications for the LHC experiments. 57

7.2 An illustration of the ATLAS detector. 60

7.3 An illustration of the ATLAS magnet system. 60

Appendix
Figure Page

7.4 An illustration of the ATLAS inner detector. 62

7.5 An illustration of the ATLAS calorimeter. 63

7.6 An illustration of the ATLAS LAr electromagnetic calorimeter's accordian structure. . 63

7.7 A topological cluster in the barrel (top) and end-cap (bottom). 65

7.8 An illustration of the ATLAS muon spectrometer. 66

7.9 An schematic of the the ATLAS Trigger and data acquisition system. 67

7.10 An ATLANTIS display of a VBF $H \rightarrow \tau\tau$ event simulated with ATLFAST (top) and GEANT3 (bottom). 69

7.11 An ATLANTIS display of a VBF $H \rightarrow \tau\tau$ event simulated with GEANT3 without noise (top) and with noise (bottom). Neither event includes pile-up effects. 70

8.1 Tree-level Feynman diagram for vector boson fusion Higgs production. 71

8.2 A flow diagram for the MadCUP generators. 73

8.3 Illustration of color coherence effects taken from CDF, Phys. Rev. **D**50. 73

8.4 Electroweak and QCD Zjj and $Zjjj$ tree-level Feynman diagrams. 74

8.5 Distribution of η^* taken from Rainwater *et. al.*, Phys. Rev. **D**54. 75

8.6 Distribution of η^* when the third jet is provided from the parton shower of HERWIG (left) and PYTHIA (right). 75

9.1 Parameterization of $\not{p}_x$ in ATLAS detector performance TDR. 78

9.2 These TDR plots show the η-dependence of the sampling and constant terms used to parametrize the hadronic endcap energy resolution to a beam of pions. 78

9.3 Illustration of geometric acceptance corrections to $\not{p}_T$ based on jets. 81

9.4 Distribution of H1-calibrated $\not{p}_T$ minus the Monte Carlo truth $\not{p}_T$ without noise suppression (left) and with a 2Δ asymmetric noise threshold (right) for VBF $H \rightarrow \tau\tau$ events. The 2Δ noise threshold improves the $\not{p}_T$ resolution, but induces a negative bias. 83

Appendix
Figure Page

9.5 The bias on a cell due to an asymmetric (left) or symmetric (right) noise threshold as a function of the true deposited energy. 85

9.6 Estimated true energy as a function of measured energy and $p(E_t = 0)$. 87

9.7 An illustration of cells in the $\eta - \phi$ plane which would be cut by a global 2Δ cut, but would not be cut with the local noise suppression technique. Jet structure can be seen in several areas. 89

9.8 Comparison of $\not{p}_T$ resolution for a global 2Δ noise cut (left) and local noise suppression (right) with GEANT4 and digitized electronic noise. 89

10.1 Schematic representation of a $H \rightarrow \tau\tau \rightarrow lh\not{p}_T$ event. 91

10.2 Left: parton-jet matching efficiencies for fast and full simulation found by Cavasinni, Costanzo, Vivarelli. Right: jet tagging efficiencies based on Monte Carlo truth jets. . 95

10.3 The ratio of reconstructed to truth jet p_T as a function of the true jet's p_T and η 95

10.4 Distribution of signal events in the x_l–x_h plane with no cuts (left) and after the requirements $\not{p}_T > 30$ GeV and $\cos\Delta\phi > -0.9$ (right) with GEANT4 and digitized noise. 97

10.5 Reconstructed Higgs mass for events in the low- and high-purity samples with ATLFAST. 98

10.6 Schematic of the impact of $\not{p}_T$ resolution on the solutions of the x_τ equations. 99

10.7 Distributions of $x_{\tau l}$ and $x_{\tau h}$ for signal events after $\not{p}_T > 30$ GeV and $\Delta\phi_{\tau\tau}$ cuts. Solid filled areas denote unphysical solutions to the x_τ equations. 100

10.8 Distribution of p_T (left) and η^* (right) for the non tagging jets. 101

10.9 Expected Significance for several analysis strategies with 30 $\mathrm{fb}^{-1}$with fast simulation. 107

10.10 $M_{\tau\tau}$ distribution for 30 fb^{-1} obtained with truth $\not{p}_T$. 111

10.11 Expected $M_{\tau\tau}$ distribution for 30 fb^{-1} obtained with fully reconstructed jets, leptons, and a $\not{p}_T$ calculation with local noise suppression. 111

11.1 Tree-level diagram of Vector Boson Fusion Higgs production with $H \rightarrow W^+W^- \rightarrow l^+l^-\nu\overline{\nu}$ 112

11.2 Neural Network output distribution for three different $t\bar{t}$ background samples. 116

Appendix
Figure Page

11.3 The improvement in the combined significance for VBF $H \rightarrow WW$ as a function of the Higgs mass, m_H. 119

11.4 Support Vector Regression and Neural Network output distributions for signal and background for 130 GeV Higgs boson in the $e\mu$ channel. 119

12.1 Left: exponential form used for Toy Monte Carlo. Right: observed number of events in the signal-like region vs. predicted number of events from fit to sideband. The red points represent experiments considered as 3σ discoveries. 121

12.2 Determination of η via a change of variables. 123

13.1 Individual and combined significance versus the Higgs mass hypothesis. 128

13.2 Discovery luminosity versus the Higgs mass hypothesis. 129

13.3 Examples of power for two different signal-plus-background hypotheses with respect to a single background-only hypothesis with 100 expected events (black). 132

13.4 The power (evaluated at 5σ) of ATLAS as a function of the Higgs mass, m_H, for 30 fb^{-1} with and without systematic errors. 132

13.5 A plot of $-2\ln Q$ vs. m_H for 30 fb^{-1} of integrated luminosity. 134

A.1 Left: The pathological behavior of the unmodified Poisson significance calculation (black). It is not only discontinuous, but also increases as the background expectation increases. Continuity is restored with the interpolation (red) provided by the generalized median (right). 152

A.2 Illustration of the numerical "noise" which appears for $\rho(q) \lesssim 10^{-16}$. 153

A.3 Diagram for the Gaussian extrapolation technique. The abscissa corresponds to the histogram bin index of the log-likelihood ratio, in which the 0^{th} bin corresponds to the lower limit $q = -s_{tot}$ (see Equation A.6). 155

A.4 Comparison of the combined significance obtained from various combination procedures. 155

A.5 Utility as a function of the discovery threshold for a channel with an expected 6σ significance when the utility for a Type I error is -17 (top) and -10^5 (bottom). 159

Appendix
Figure Page

A.6 Utility as a function of discovery threshold for a channel with an expected 2σ significance when the utility for a Type I error is -10^5. 161

B.1 The performance of boundary kernels on a Neural Network distribution with a hard boundary 166

B.2 The standard output of the KEYS script. The top left plot shows the cumulative distributions of the KEYS shape and the data. The top right plot shows the difference between the two cumulative distributions, the maximum of which is used in the calculation of the Kolmogorov-Smirnov test. The bottom plot shows the shape produced by KEYS overlayed on a histogram of the original data. 170

C.1 The Neyman construction for a test statistic x, an auxiliary measurement M, and a nuisance parameter b. Vertical planes represent acceptance regions W_b for H_0 given b. The contours of $L(x, M|H_0, b)$ are in shown in color. 180

C.2 Contours of the likelihood ratio (diagonal lines) and contours of $L(x, M|H_0, b)$ (concentric ellipses). 180

C.3 Comparison of the background confidence level, CL_b, as a function of the number of signal events for different experiments and different methods of incorporating systematic error. 183

C.4 Contours of σ_∞^{CH} in the plane of signal-to-background ratio vs. the systematic error α in percent (left) and comparison with the frequentist technique (right). 185

D.1 The VC Confidence as a function of h/l for $l = 10,000$ and $\eta = 0.05$. Note that for $l < 3h$ the bound is non-trivial and for $l < 20h$ is quite tight. 191

D.2 Example of an oriented line shattering 3 points. Solid and empty dots represent the two classes for y and each of the 2^3 permutations are shown. 191

E.1 Signal and Background histograms for an expression. 198

E.2 An example of crossover. At some given generation, two parents (a) and (b) are chosen for a crossover mutation. Two subtrees, shown in bold, are selected at random from the parents and are swapped to produce two children (c) and (d) in the subsequent generation. 199

Appendix
Figure Page

E.3 Monte Carlo sampling of individuals based on their fitness. A uniform variate x is transformed by a simple power to produce selection pressure: a bias toward individuals with higher fitness. 200

E.4 The fitness of the population as a function of time. This plot is analogous to a neural network error vs. epoch plot, with the notable exception that it describes a population and not an individual. In particular, the neural network graph is a 1-dimensional curve, but this is a two dimensional distribution. 202

E.5 An explicit example of the largest polynomial on two variables with degree two. In total, 53 nodes are necessary for this expression which has only 9 independent parameters. 204

F.1 The ATLAS Analysis Event Data Model. 206

DISCARD THIS PAGE

NOMENCLATURE

H_0 The null hypothesis or the "background-only" hypothesis.

H_1 The alternate hypothesis or the "signal-plus-background" hypothesis.

$L(x|H_0)$ The likelihood of observing a generic observable x given the null hypothesis.

$L(x|H_1)$ The likelihood of observing a generic observable x given the alternate hypothesis.

α Probability of Type I error or the *size* of a given hypothesis test. The variable α is also used to refer to background uncertainty (see Appendix C).

β Probability of Type II error. The *power* of a given hypothesis test is defined as $1 - \beta$.

W The acceptance region for the null hypothesis.

Q The likelihood ratio $L(x|H_1)/L(x|H_0)$.

q The log likelihood ratio $q = \ln Q$.

$\rho_{1,H}(q)$ The probability density of q for a given hypothesis H.

m_H The hypothesized mass of the Higgs boson.

M_{xy} The invariant mass of particle x and particle y.

σ This variable usually refers to the standard deviation of some implicit Gaussian distribution, thus it is used in several ways. In the context of the sensitivity of an analysis, the result is usually quoted as $N\sigma$ or $\sigma = N$, where N is given by Equation A.3. In the context of detector performance, σ refers to the resolution of a reconstructed quantity. It is also the symbol that is used for the cross-section of a given particle interaction.

Δ The variable Δ has two meanings in this dissertation. The first is the root mean squared (RMS) electronic noise in a calorimeter cell. The second is the background uncertainty in a frequentist context.

$\not{p}_T$ Missing transverse momentum.

$\mathcal{M}$ The matrix element for a given particle interaction.

η Pseudo-rapidity defined as $\eta = -\ln\tan(\theta/2)$, where θ is the polar angle measured from the beam axis. Also used as a temporary variable in the text.

ϕ The azimuthal angle measured from the x axis or the Higgs scalar field.

$\mathcal{X}$ A test statistic defined for Local Noise Suppression (see Chapter 9).

x_τ The fraction of a tau lepton's momentum carried away by the visible decay product.

$\mathrm{erf}(x)$ The error function defined as $\mathrm{erf}(N) = (2/\sqrt{\pi})\int_N^\infty \exp(-y^2)dy$.

$\Theta(x)$ The Heaviside function, which is zero if x is negative and unity otherwise.

ARCH The general-purpose particle identification program developed for ALEPH data.

SEARCHING FOR NEW PHYSICS:
CONTRIBUTIONS TO LEP AND THE LHC

Kyle S. Cranmer

Under the supervision of Professor Sau Lan Wu
At the University of Wisconsin-Madison

This dissertation is divided into two parts and consists of a series of contributions to searches for new physics with LEP and the LHC. In the first part, an exhaustive comparison of ALEPH's LEP2 data and Standard Model predictions is made for several hundred final states. The observations are in agreement with predictions with the exception of the $e^-\mu^+$ final state. Using the same general purpose particle identification procedure, searches for minimal supergravity signatures, excited electrons, doubly charged Higgs bosons, singly charged Higgs bosons, and the Standard Model Higgs boson were performed. The results of those searches are in agreement with previous ALEPH analyses. The second part focuses on preparation for searches for Higgs bosons with masses between 100 and 200 GeV. Improvements to the relevant Monte Carlo generators and the reconstruction of missing transverse momentum are presented. A detailed full simulation study of Vector Boson Fusion Higgs decaying to tau leptons confirms the qualitative conclusion that the channel is powerful near the LEP limit. Several novel statistical and multivariate analysis algorithms are considered, and their impact on Higgs searches is assessed. Finally, sensitivity estimates are provided for the combination of channels available for low mass Higgs searches. With 30 fb^{-1} the expected ATLAS sensitivity is above 5σ for Higgs masses above 105 GeV.

Sau Lan Wu

ABSTRACT

This dissertation is divided into two parts and consists of a series of contributions to searches for new physics with LEP and the LHC. In the first part, an exhaustive comparison of ALEPH's LEP2 data and Standard Model predictions is made for several hundred final states. The observations are in agreement with predictions with the exception of the $e^-\mu^+$ final state. Using the same general purpose particle identification procedure, searches for minimal supergravity signatures, excited electrons, doubly charged Higgs bosons, singly charged Higgs bosons, and the Standard Model Higgs boson were performed. The results of those searches are in agreement with previous ALEPH analyses. The second part focuses on preparation for searches for Higgs bosons with masses between 100 and 200 GeV. Improvements to the relevant Monte Carlo generators and the reconstruction of missing transverse momentum are presented. A detailed full simulation study of Vector Boson Fusion Higgs decaying to tau leptons confirms the qualitative conclusion that the channel is powerful near the LEP limit. Several novel statistical and multivariate analysis algorithms are considered, and their impact on Higgs searches is assessed. Finally, sensitivity estimates are provided for the combination of channels available for low mass Higgs searches. With 30 fb^{-1} the expected ATLAS sensitivity is above 5σ for Higgs masses above 105 GeV.

Chapter 1

Introduction

This dissertation is somewhat unusual in that it does not focus on one specific measurement or new particle search, as is typical in high energy physics. The bulk of my graduate career belongs to a correspondingly unusual era – the time between data taking at LEP and the LHC – during which the most promising experiments for the direct observation of new physics lie either in the past or the future. Years ago, I decided I had two options: switch topics and produce a thesis that conforms to the expectations of the field; or take advantage of this unique period, address the fundamental issues neglected in the last generation of experiments, and apply what I learn to this new generation of experiments. I chose the latter.

I have tried to take the broadest view possible and reconsider experimentation holistically. The result is a series of contributions to the way we search for new physics. These contributions include the application of an inclusive data analysis strategy to LEP data, improvements to the statistical formalism used by the LEP Higgs working group, theoretical results related to the use of multivariate analysis techniques, and a novel multivariate algorithm.

This dissertation is not merely a collection of potentialities; the results of several new particle searches based on LEP2 data and practical developments in the preparation for data analysis at the LHC are presented.

This dissertation is arranged in two parts with conclusions being drawn at the end of each. The first part focuses on an exhaustive comparison of LEP2 data to Standard Model predictions and the application of the QUAERO analysis procedure to ALEPH's LEP2 data. The second part focuses on the preparation for Higgs searches with ATLAS. In addition there are several appendices that detail advances in statistical and multivariate methods.

Chapter 2

Motivation

Throughout history, fundamental physics has attempted to reduce Nature to her most essential kernels of complexity. Currently, we know of four fundamental forces: *gravity*, the weakest and most familiar force; the *electromagnetic force*, responsible for all of chemistry; the *weak nuclear force*, the short ranged force which powers the sun; and the *strong nuclear force*, which holds the nucleus of an atom together. In addition, we have observed a number of sub-atomic particles; many of which are unstable, but decay with characteristic time scales and kinematic properties [1].

The best known description of gravity is given by Einstein's general theory of relativity, which relates the geometry of space and time to the stress-energy tensor of classical mechanics [2]. The other three forces and all known particles are described in a different formalism, known as Quantum Field Theory (QFT), which is a blend of classical field theory, group theory, and relativistic quantum mechanics. Thus far, all attempts to incorporate General Relativity into the formalism of QFT have failed; sparking interest in new theoretical frameworks such as Superstring theory and M-theory.

Both General Relativity and the Standard Model provide huge predictive power, and both have survived incredibly precise tests of those predictions. Nevertheless, there are a number of reasons to believe that there is a theory more fundamental than the Standard Model, and it is hoped that this theory might unify the description of particles and their interactions in a single "Theory of Everything".

2.1 The Standard Model

> *Symmetry, as wide or as narrow as you may define its meaning is one idea by which man through the ages has tried to comprehend and create order, beauty, and perfection.* - Hermann Weyl

The best known description of the fundamental particles and their interactions, neglecting gravitational effects, is provided by a particular quantum field theory known as the *Standard Model*. It is neither feasible nor appropriate to describe the Standard Model in complete detail in this dissertation; however, it is worth mentioning its most salient features.

Arguably, the most essential component to the Standard Model is the presence the group $U(1)_Y \otimes SU(2)_L \otimes SU(3)_{\text{color}}$ of local gauge symmetries. The group is factorized roughly into the electromagnetic, the weak, and the strong interactions, respectively. Each of these groups is a Lie group, which means that it can either be thought of as a smooth surface obeying group properties or as a particular type of matrix. For instance, the group $U(1)_Y$ can be thought of simply as the rotations of a circle. In this case the symmetry is realized by the the unobservable phase of charged fermions. The unobservable phase is a complex number with modulus of unity, which can be thought of as a 1×1 unitary matrix: hence the name $U(1)$. The circle is manifest as the set of points in the complex plane defined by those local phase transformations $e^{i\alpha(x)}$, where $\alpha(x)$ is an arbitrary function of the space-time coordinate x.

Amazingly, by requiring the electron field to be invariant to local gauge transformations, QFT predicts the existence of an additional field that transforms just as Maxwell's equations of electricity and magnetism. In addition, this field can propagate massless spin-1 particles that interact with charged particles: *i.e.* the photon! In an analogous way, the other symmetries of the Standard Model predict the $W^{\pm}$ and Z bosons of the weak interaction and the gluons of strong interactions.

While there are many ways of formulating Quantum Field Theory, the Lagrangian formalism is the most common and offers several advantages (such as manifest relativistic covariance). In much the same way as classical mechanics, the equations of motion follow from the principle of least action, from which Feynman developed his ubiquitous diagrams [3, 4].

In the 1960's, the theory of Quantum Electrodynamics (QED) was already very successful, and motivated the theoretical community to evolve Fermi's theory of weak interactions into a Yang-Mills theory [5] based on the symmetry group $SU(2)_L$. The immediate problem with this approach was that gauge invariance forbade masses for both the gauge bosons and the leptons. The observation of Peter Higgs was that the gauge invariance could be *spontaneously broken* with the addition of a doublet of complex scalar fields, ϕ, with Lagrangian

$$\mathcal{L}_{\text{Higgs}} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) \; - \; V(\phi) \tag{2.1}$$

where the potential

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \tag{2.2}$$

is the key to spontaneous symmetry breaking [6, 7]. With a plausible mechanism for electroweak symmetry breaking in hand, Glashow [8], Weinberg [9], and Salam [10] proposed a unified electroweak theory of the leptons. This theory retained a massless photon; allowed for massive $W^\pm$ bosons and leptons; predicted the massive, neutral, spin-1 Z boson; and predicted the massive, neutral scalar Higgs boson. The $W^\pm$ and Z bosons were discovered at the CERN SPS by the UA1 and UA2 experiments [11, 12].

At the time of the formulation of the Glashow-Weinberg-Salam theory, two families of leptons were known, and the discovery of the Ω^- [13] had given a great deal of support to the quark model developed by Gell-Mann [14] and, specifically, the notion of color-charge.

In order for the Glashow-Weinberg-Salam theory to be "anomaly free" (*i.e.* conservation laws predicted by Noether's theorem are respected to all orders in perturbation theory) the difference in the sum of charges between the right and left handed doublets must vanish. This property is not satisfied by the leptons alone, but is satisfied if, for each lepton doublet, we include three doublets of quarks – precisely what is provided by three colors. A method of providing electroweak interactions to the quarks while avoiding flavor-changing neutral currents was devised by Glashow, Iliopoulos, and Maiani [15].

The discovery of the tau lepton, by Perl in 1975, provided evidence for a third family of leptons [16]. Moreover, the mass width of the Z boson requires the number of lepton families with

Family	I	II	III	I	II	III
Leptons	$\begin{pmatrix}\nu_e\\ e\end{pmatrix}_L$	$\begin{pmatrix}\nu_\mu\\ \mu\end{pmatrix}_L$	$\begin{pmatrix}\nu_\tau\\ \tau\end{pmatrix}_L$	e_R	μ_R	τ_R
Quarks	$\begin{pmatrix}u\\ d\end{pmatrix}_L$	$\begin{pmatrix}c\\ s\end{pmatrix}_L$	$\begin{pmatrix}t\\ b\end{pmatrix}_L$	u_R d_R	c_R s_R	t_R b_R

Table 2.1 The fermions of the Standard Model grouped according to family. No right handed neutrinos are included. Braces indicate weak isospin doublets.

light neutrinos to be three [17]. After the discovery of the charm [18, 19], bottom [20], and top [21, 22] quarks, we arrive at the what are considered to be the fundamental fermions, seen in Table 2.1. The mass eigenstates of these quarks, however, is not the same as the eigenstates of the weak interaction. The mixing between the two sets of eigenstates is parametrized by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [23, 24]. The CKM matrix plays a fundamental role in the study of CP-violation, in which reactions related to each other by charge conjugation and parity do not proceed at the same rate. Finally, recent experimental evidence [25] shows that neutrinos have mass (and thus right-handed components), and may also experience some mixing. Despite this experimental fact, the right handed neutrinos are not typically included in the Standard Model. Because these right handed neutrinos are color and charge neutral, they only interact gravitationally.

Having seen the success of local gauge theories in the electroweak interaction, the theoretical community constructed a theory of local gauge symmetry based on color charge, known as Quantum Chromodynamics (QCD), to describe the strong interaction. Corresponding to the photon of the electromagnetic interaction, are the eight massless, spin-1 gluons of QCD. Due to the fact that strong interactions are strong and that the symmetry group $SU(3)_{\text{color}}$ is non-Abelian, the dynamics of QCD are incredibly complicated and perturbation theory is not generally applicable. In particular, as colored objects are separated by larger distances, the force between them grows

very rapidly; thus, QCD exhibits a feature known as *confinement*. In contrast, Wilczek, Gross, and Politzer showed that at small distance scales, the theory is asymptotically free [26, 27], which justifies the use of perturbation theory to describe strong interactions with high momentum (or high Q^2) transfer. During such a high Q^2 interaction, colored partons often radiate more colored partons in a *parton shower*. In a process known as *hadronization*, these colored partons group themselves into color neutral objects such as mesons or hadrons. The collection of mesons and hadrons originating from outgoing partons is known as a jet [28]. The three jet events at Tasso were interpreted as $e^+e^- \to q\bar{q}g$: the first direct evidence of the gluon [29, 30].

For completeness, the Standard Model Lagrangian is written explicitly in Equation 2.3.

$$
\begin{aligned}
\mathcal{L}_{SM} = & \underbrace{\frac{1}{4}\mathbf{W}_{\mu\nu}\cdot\mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a}_{\text{kinetic energies and self-interactions of the gauge bosons}} \\
& + \underbrace{\bar{L}\gamma^\mu(i\partial_\mu - \frac{1}{2}g\tau\cdot\mathbf{W}_\mu - \frac{1}{2}g'YB_\mu)L \;+\; \bar{R}\gamma^\mu(i\partial_\mu - \frac{1}{2}g'YB_\mu)R}_{\text{kinetic energies and electroweak interactions of fermions}} \\
& + \underbrace{\frac{1}{2}\left|(i\partial_\mu - \frac{1}{2}g\tau\cdot\mathbf{W}_\mu - \frac{1}{2}g'YB_\mu)\,\phi\right|^2 \;-\; V(\phi)}_{W^\pm, Z, \gamma, \text{and Higgs masses and couplings}} \\
& + \underbrace{g''(\bar{q}\gamma^\mu T_a q)\,G^a_\mu}_{\text{interactions between quarks and gluons}} \quad + \underbrace{(G_1\bar{L}\phi R + G_2\bar{L}\phi_c R + h.c.)}_{\text{fermion masses and couplings to Higgs}}
\end{aligned}
\tag{2.3}
$$

2.2 Phenomenology

While the Lagrangian in Equation 2.3 is the fundamental theoretical object from which the equations of motion are derived, it is not immediately useful for making predictions of observable quantities. From an experimental point of view, one would like a prediction of the rate at which a certain reaction will take place and the shape of various kinematic distributions. The observed rate, R, for a particular interactions is given by

$$R = L\epsilon\sigma, \tag{2.4}$$

where L is the instantaneous luminosity of the colliding beams (with units $\mathrm{cm}^{-2}\mathrm{s}^{-1}$), σ is the cross section for the the interaction (in units of $b = 10^{-24}\mathrm{cm}^2$), and ϵ is the efficiency of observing the given interaction. The quantities L and ϵ are properties of the collider and the detector, respectively; however, the cross-section, σ, can be predicted from theory.

The prediction of the cross-section is obtained from the Feynman diagrammatic approach. Essentially, all Feynman diagrams with the specified initial and final state particles are considered. Each external leg, vertex, and internal propagator corresponds to a matrix of complex numbers, and their product is a single complex number. Of particular importance are the vertex contributions which provide the weak or strong coupling constants, because, in fixed-order perturbation theory, one neglects diagrams above a certain degree in the coupling constants[1]. The sum of this finite subset of Feynman diagrams is called *the matrix element* and is denoted as $-i\mathcal{M}$. The matrix element is incredibly useful due to the equation

$$d\sigma = \frac{|\mathcal{M}|^2}{F}\, dQ, \tag{2.5}$$

which relates the differential cross section to the matrix element, the initial flux of particles, F, and the differential phase space factor, dQ. Both the matrix element and the differential cross section are implicitly functions of the kinematic configuration of the incoming and outgoing legs of the Feynman diagram.

For interactions with leptons or photons in the initial and final state, the procedure described above is quite complete; however, the situation is more complicated for interactions with hadrons in the initial or final state. For instance, we cannot directly use the aforementioned procedure when proton beams collide – as they will at the Large Hadron Collider (LHC). The complications arise due to confinement and the fact that perturbation theory is inapplicable at low-Q^2.

The two major additions to the theoretical framework of the Standard Model are not fundamental additions – though they could be derived from Equation 2.3, in principle – but are phenomenological in nature.

The first addition is the notion of *parton density functions* (PDFs), which quantify what probability to find a particular type of parton inside of the proton as a function of Q^2 and the Bjorken

[1] In this way, one often refers to Leading Order (LO) and Next-to-Leading Order (NLO) calculations

x. The theoretical justification for the PDF approach is due to the *factorization* theorem [31], which roughly states that we can factorize the soft and hard components of QCD at particular factorization scale. Below the factorization scale, the QCD behavior is non-perturbative. However, the measurement of Deep Inelastic Scattering (DIS) in $e^{-}p$ collisions, together with perturbative predictions above the factorization scale, allows us to infer the non-perturbative piece. The *evolution* of the PDFs to different Q^2 and x values is accomplished with the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) [32, 33, 34] and Balitsky-Fadin-Kuraev-Lipatov (BFKL) evolution equations [35, 36, 37]. The measurement of PDFs and the estimation on their uncertainties is an active area of research and is very relevant for searches for new physics at the LHC.

The second phenomenological augmentation to the Standard Model is related to hadronization. As mentioned earlier, colored partons must group themselves together into color neutral objects, such as mesons and hadrons. These colored partons can be produced either from the leading order Feynman diagrams or from initial- and final-state radiation described by the DGLAP equations. The relative rates of the various mesons and hadrons inside of a light-quark, gluon, or heavy-quark initiated jet are described by various phenomenological models, which are tuned to agree with measured jet properties. The implementation of these phenomenological models exists in a number of showering and hadronization generators (SHGs) – most notably PYTHIA and HERWIG [38, 39]. It should be noted that the parton shower does not correspond to a fixed-order in perturbation theory. The essence of DGLAP evolution is to *re-sum* certain leading diagrams to all orders. The resummation provided by PYTHIA is of leading-log (LL) accuracy.

Recently, the process of tree-level matrix element evaluation and high-dimensional phase space integration has been automated [40, 41, 42]. Previously, the same procedure required a custom matrix element and phase-space integration routine to be developed for each reaction under consideration (see Chapter 8). Together with SHGs like PYTHIA and HERWIG, we are able to predict the entire Standard Model at LO and LL accuracy. By the time of the LHC turn-on, it is likely that we will have the same for some extensions to the Standard Model. In addition, there are now major strides in providing a general purpose event generator at NLO [43].

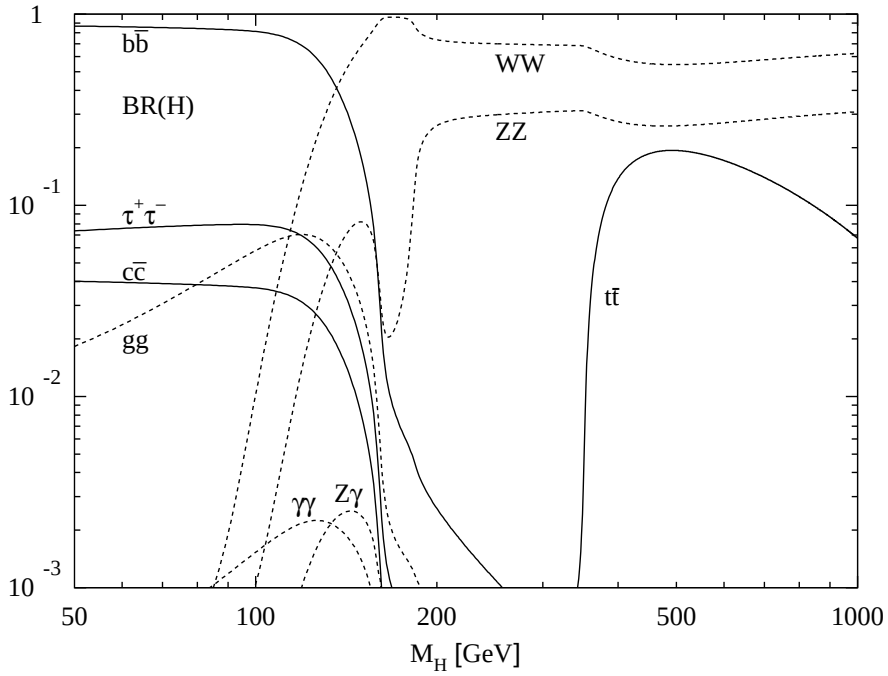


Figure 2.1 Higgs branching ratios as a function of m_H from M. Spira Fortsch. Phys. 46 (1998)

2.3 The Phenomenology of the Standard Model Higgs

The phenomenology of the Standard Model Higgs boson is of particular importance because it is the only particle of the Standard Model that has not been discovered. The Higgs boson mass, m_H, is the only unknown fundamental parameter within the Standard Model.

The decay of the Higgs boson is isotropic in the Higgs rest frame because the Higgs is a scalar particle. The decay of Higgs bosons to the $W^{\pm}$ and Z bosons is related to the couplings implicit in the Higgs mechanism, but the decay to fermions is a typical Yukawa interaction. For Kinematic reasons, the Higgs decay into the $W^{\pm}$ and Z bosons has a rapid increase near $2M_W$, after which it remains fairly constant. Below $2M_W$, we should expect the Higgs to decay primarily to $b\bar{b}$ and $\tau^+\tau^-$ because the Higgs coupling to fermions is proportional to the square of their mass. While the Higgs does not directly couple to either the photon or gluons, top-quark loops provide the decay channels $H \rightarrow \gamma\gamma$ and $H \rightarrow gg$. Finally, we do expect to observe the decay $H \rightarrow t\bar{t}$ if $M_H > 2M_t > 2M_W$. This behavior is summarized in Figure 2.1.

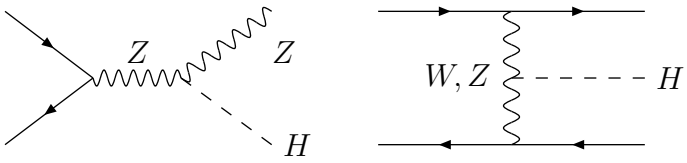


Figure 2.2 Tree level Feynman diagrams for the Higgsstrahlung (left) and Vector Boson Fusion (right) Higgs production mechanisms from e^+e^- interactions.

The s-channel production of Higgs bosons at e^+e^- colliders is very rare due to the low electron mass. However, the so-called Higgsstrahlung (Figure 2.2 left) and vector boson fusion (Figure 2.2 right) provide sufficient rate for a potential discovery. The Higgsstrahlung process dominates the production cross section, which is shown in Figure 2.3 as a function of $\sqrt{s}$ and m_H.

At the LHC, several production modes are available (see Figure 2.4. The most dominant production mode is called gluon fusion (top left), which proceeds through a heavy quark loop. The second dominant process is called vector boson fusion (VBF), in which the Higgs is produced in association with two hard, forward jets (top right)[2]. The search for VBF Higgs is outlined in Chapters 10 and 11. The next most prominent production modes include associated production with a weak boson (bottom left) or two heavy quarks (bottom right). The associated production modes are important because the provide a high-p_T lepton for triggering purposes, thus allowing for $H \rightarrow b\bar{b}$ to be observed at the LHC. The production cross sections as a function of M_H are shown in Figure 2.3.

2.4 Results from LEP Higgs Searches

Searches for the Higgs boson were a major priority for all LEP experiments near the end of LEP2. The LEP Higgs Working Group (LHWG) was formed to combine those results in a consistent statistical framework in order to provide the most powerful indication of discovery or exclusion limits.

[2]VBF Higgs is often denoted as qqH

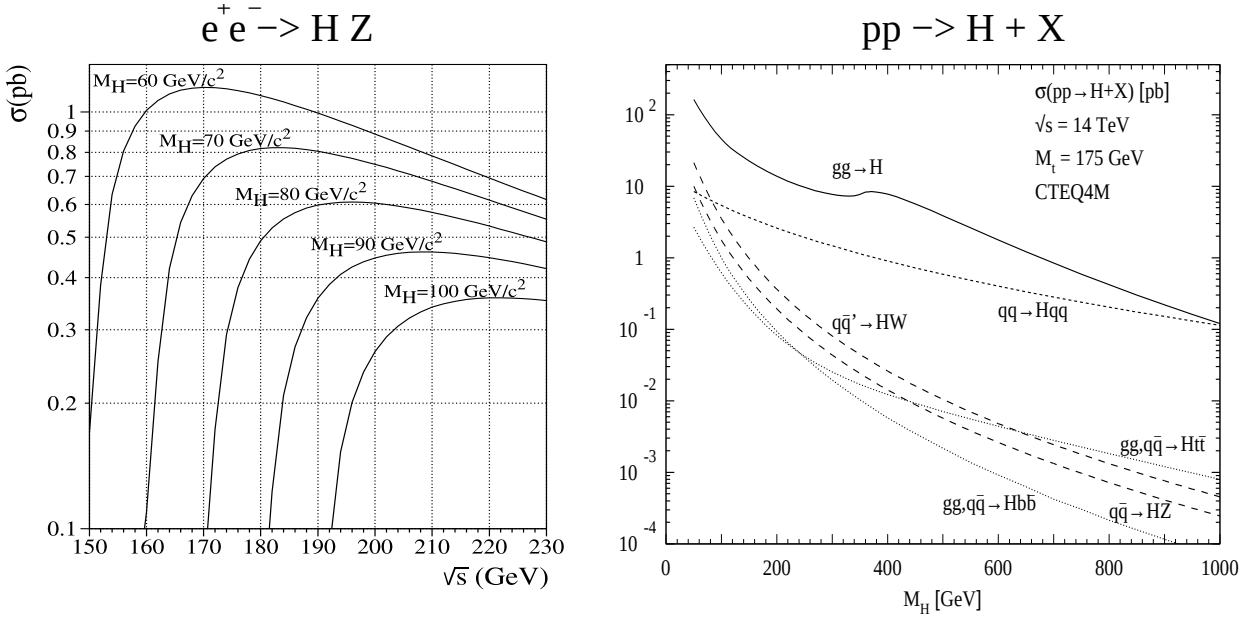


Figure 2.3 Left: the cross section for $e^+e^- \rightarrow HZ$ as a function of $\sqrt{s}$ for several Higgs masses as obtained with the HZHA generator. Right: the cross section for $pp \rightarrow H + X$ as a function of M_H from M. Spira Fortsch. Phys. 46 (1998).

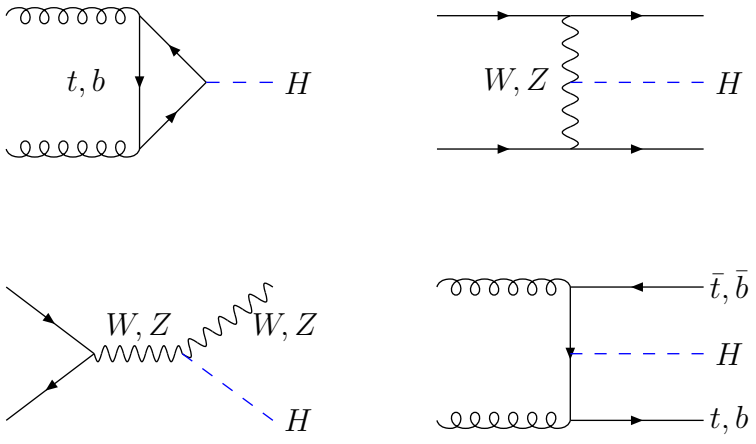


Figure 2.4 Feynman diagrams for the Higgs production at the LHC.

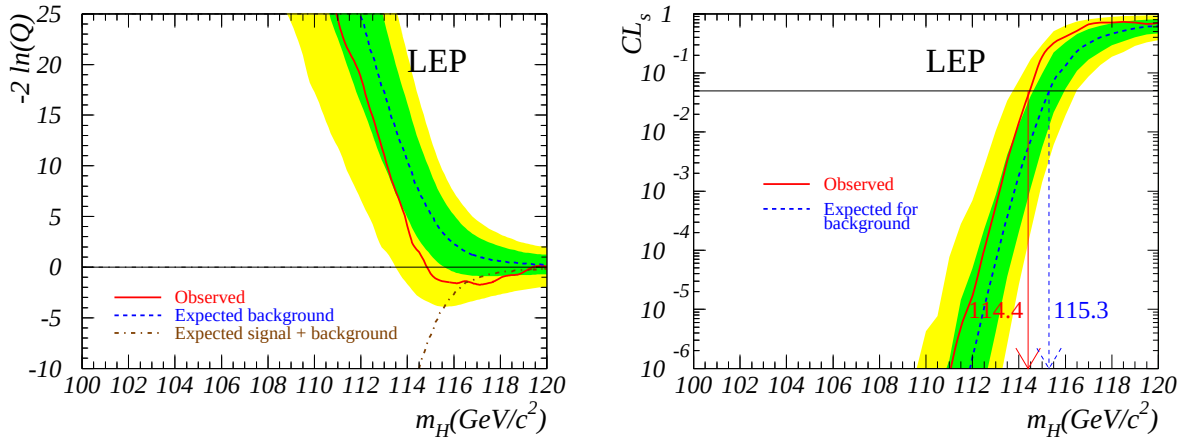


Figure 2.5 The expected and observed evolution of $-2\ln Q$ (left) and CL_s (right) with m_H from all LEP experiments.

The statistical framework of the LHWG is outlined in Appendix A.1. The author contributed the KEYS package to this framework, which is described in Appendix B.

The results of the LHWG are documented in Ref. [44]. The ALEPH collaboration observed an excess [45] in Higgs candidates with a mass near 115 GeV. This excess can be seen in the left of Figure 2.5 where the observed curve drops into the region $-2\ln Q < 0$. The green and yellow bands surrounding the expected background curves correspond to 1σ and 3σ bands. The observed excess is not sufficient to claim a discovery, so an exclusion region was constructed. The right of Figure 2.5 shows that Higgs bosons with mass $m_H < 114.4$ GeV can be excluded at the 95% confidence level.

In addition to direct searches for the Higgs, the LEP electroweak group provided indirect limits on the mass of the Higgs. These indirect searches rely on the fact that the Higgs introduces radiative corrections in the electroweak sector with a leading behavior that is logarithmic in m_H. The corrections are also sensitive to the mass of the top quark, with a leading behavior that is quadratic in m_t. Figure 2.6 shows that a Higgs with mass near 115 GeV is favored and that a Higgs with mass $m_H > 260$ GeV is indirectly excluded at the 95% confidence level.

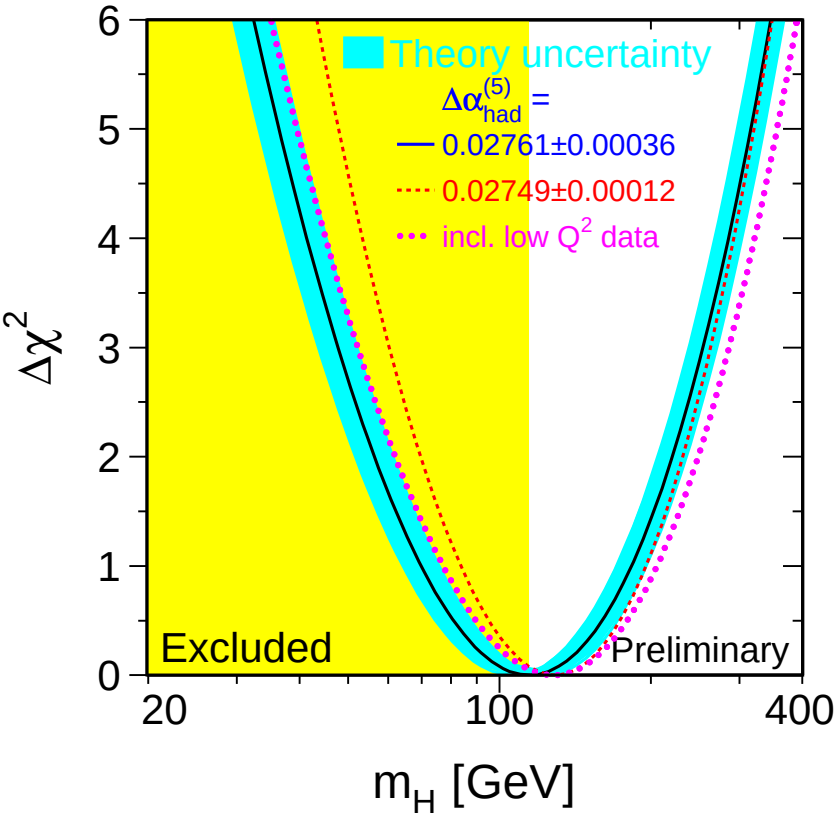


Figure 2.6 The resulting $\Delta\chi^2$ vs m_H from LEP electroweak fits.

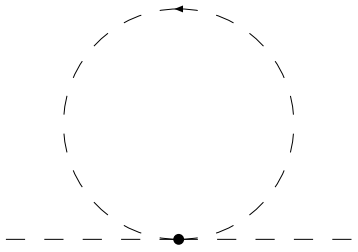


Figure 2.7 Quadratically divergent diagram in Higgs self energy.

2.5 Beyond the Standard Model

The construction of the Standard Model is one of the great achievements of the twentieth century. Unfortunately, the Standard Model itself provides some suggestion that it is not the final theory. In particular, the self-energy contribution of the Higgs boson due to the diagram shown in Figure 2.7 grows quadratically with the cutoff scale Λ. This is a fairly generic behavior for scalar particles; however, Veltman pointed that a particular relationship between the masses of the fermions and the $W, Z,$ and Higgs bosons removes the quadratic divergence at one loop [46]. Unfortunately, Veltman's condition only removes the quadratic divergence if a universal cutoff scale Λ is (arbitrarily) chosen. Andrianov and Rodenberg provided an extended Veltman-like condition by also requiring the vacuum self energies to cancel and for the condition to be only weakly scale dependent [47]. In that case the mass of the top and Higgs are predicted to be $m_t = 177$ GeV and $m_H = 213$ GeV. In both cases, the cancellation of the quadratic divergence requires a *fine tuning* of the fundamental constants. This fine tuning problem is seen as a major flaw of the Standard Model and is oft cited as a motivation for *supersymmetry*.

Supersymmetry is a beautiful theory that postulates a symmetry connecting fermions and bosons. If the symmetry exists at high energies, then additional fermionic loops would cancel the quadratic divergences in the Higgs sector. In addition, supersymmetric theories require two Higgs doublets, which correspond to five physically observable Higgs bosons: the light and heavy CP even, neutral h and H bosons; the pseudoscalar, neutral A boson; and the charged Higgs bosons $H^{\pm}$.

The fact that no supersymmetric particles have yet been observed means that supersymmetry is not an unbroken symmetry in Nature. In the minimal supersymmetric extension to the Standard Model (MSSM) there are no assumptions made to the SUSY-breaking mechanism. Instead, all possible SUSY-breaking terms are considered, which gives rise to more than 100 new, fundamental parameters. Models exist in which the the low-energy parameters are determined from only a

few parameters, which 'live' at a much higher scale, by assuming a specific SUSY-breaking mechanism. These models include minimal-Supergravity (mSUGRA), minimal Gauge Mediated SUSY Breaking (mGMSB) and minimal Anomaly Mediated SUSY Breaking (mAMSB) [48].

A more recent variation on supersymmetry is called Split Supersymmetry [49, 50]. By invoking the recent insight that particular types of string theory have fabulously rich landscapes of long-lived vacuua, they provide an anthropic argument that fine-tuning might not be such an unnatural state of affairs. Amazingly, those theories make predictions that appear to be within the limits of the LHC.

There are innumerable other theories for new physics that have been postulated in recent years. Some propose that the particles we think of as fundamental are actually composite. Others propose large extra dimensions, baby black holes, doubly charged Higgs bosons, etc. One of the goals of this dissertation is to sharpen the tools for searches for new physics in general.

Part I

Searching For New Physics at LEP

Chapter 3

The Aleph Detector at LEP

3.1 The Large Electron Positron Collider

The Large Electron Positron collider (LEP), a 27 km ring with four multipurpose detectors, operated from 1989 to 2000. The LEP accelerator complex is a series of accelerators, shown in Figure 3.1, that brings electrons and positrons from energies of 200 MeV in the Linear Accelerator (LINAC) to 22 GeV in the Super Proton Syncrotron (SPS), which injects them into LEP. During the first phase of LEP operation, which lasted until 1996, the center of mass energy corresponded to the $e^+e^- \rightarrow Z$ resonance, and the physics program was concentrated on precision electroweak physics and B-physics. During the second phase, the center-of-mass energy was increased gradually to 105 GeV per beam and the physics program was more oriented to searches for new physics.

3.2 The Aleph Detector

A detailed description of the ALEPH detector can be found in Ref. [51] and of its performance in Ref. [52]. Charged particles are detected in the central part, which consists of a precision silicon vertex detector (VDET), a cylindrical drift chamber (ITC) and a large time projection chamber (TPC), measuring altogether up to 31 space points along the charged particle trajectories. A 1.5 T axial magnetic field is provided by a superconducting solenoidal coil. Charged particle transverse momenta are reconstructed with a $1/p_T$ resolution of $(6 \cdot 10^{-4} \oplus 5 \cdot 10^{-3}/p_T)\ (\mathrm{GeV}/c)^{-1}$.

In addition to its rôle as a tracking device, the TPC also measures the specific energy loss by ionization $\mathrm{d}E/\mathrm{d}x$. It allows low momentum electrons to be separated from other charged particle species by more than three standard deviations.

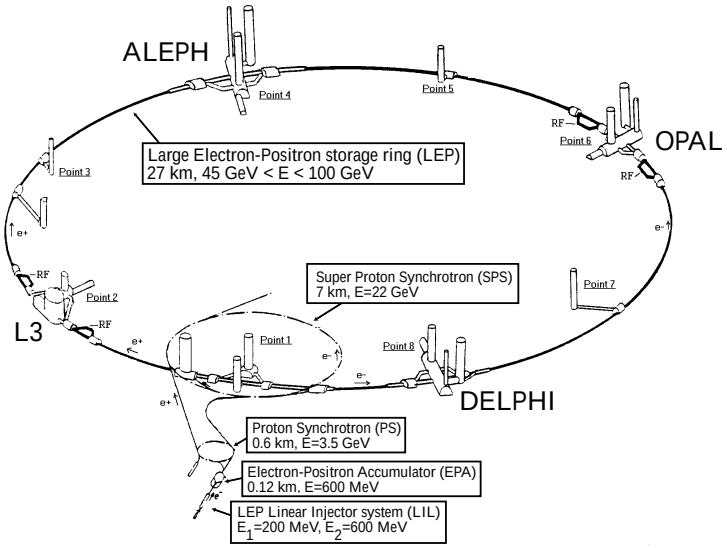


Figure 3.1 An illustration of the LEP tunnel.

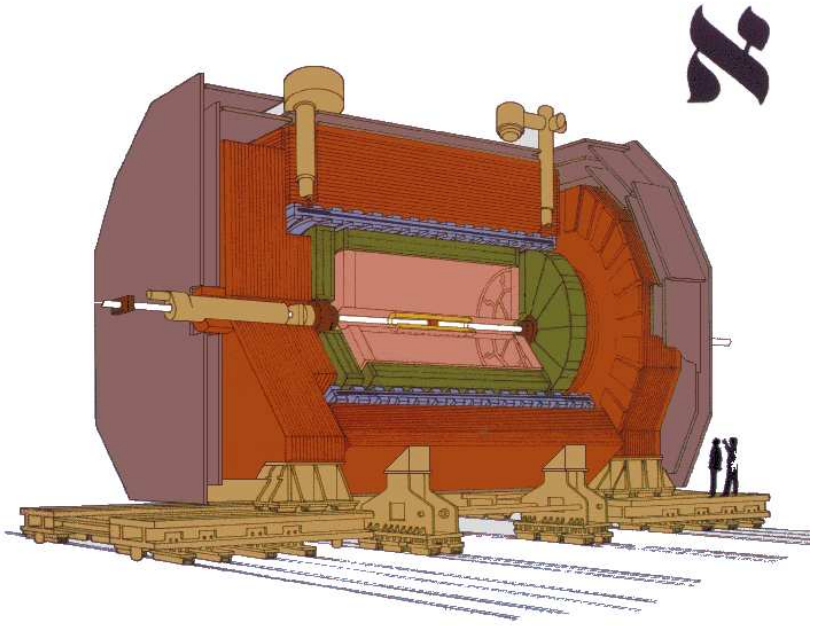


Figure 3.2 An illustration of the ALEPH detector.

Electrons (and photons) are also identified by the characteristic longitudinal and transverse developments of the associated showers in the electromagnetic calorimeter (ECAL), a 22 radiation length thick sandwich of lead planes and proportional wire chambers with fine read-out segmentation. A relative energy resolution of $0.18/\sqrt{E}$ (E in GeV) is achieved for isolated electrons and photons.

Muons are identified by their characteristic penetration pattern in the hadron calorimeter (HCAL), a 1.2 m thick yoke interleaved with 23 layers of streamer tubes, together with two surrounding double-layers of muon chambers. In association with the electromagnetic calorimeter, the hadron calorimeter also provides a measurement of the hadronic energy with a relative resolution of $0.85/\sqrt{E}$ (E in GeV).

Below polar angles of 12° and down to 34 mrad from the beam axis, the acceptance is closed at both ends of the experiment by the luminosity calorimeter (LCAL) [53] and a tungsten-silicon calorimeter (SICAL) [54] originally designed for the LEP 1 luminosity measurement. The dead regions between the two LCAL modules at each end are covered by pairs of scintillators. The luminosity is measured with small-angle Bhabha events with the LCAL with an uncertainty smaller than 0.5%. The Bhabha cross section [55] in the LCAL acceptance varies from 4.6 nb at 183 GeV to 3.6 nb at 207 GeV.

The energy flow reconstruction algorithm, which combines all the above measurements, provides a list of reconstructed objects, classified as charged particles, photons and neutral hadrons, and referred to as energy flow objects in the following [52]. The charged particle tracks used in the present analysis are reconstructed with at least four hits in the TPC, and originate from within a cylinder of length 20 cm and radius 2 cm coaxial with the beam and centered at the nominal collision point.

The ALEPH detector simulation, GALEPH, is performed with Geant3 [56]. The ALEPH reconstruction is known as JULIA [57], and the ALEPH physics analysis package is known as ALPHA [58].

Chapter 4

Vista@Aleph

This chapter describes a particular partitioning of ALEPH's LEP2 data according to identified particles and provides a comparison to Standard Model predictions. The particle identification routine, ARCH, was developed by the author. The comparison with the Standard Model is performed with an algorithm called VISTA developed by Bruce Knuteson. The VISTA algorithm shares a common data format with QUAERO, a framework which was previously used by the DØ collaboration to automate analysis of Tevatron Run I data [59]. QUAERO@ALEPH is described in Chapter 5. The goal of incorporating ALEPH data into the VISTA / QUAERO framework is fourfold:

- to provide a comprehensive comparison between ALEPH's Standard Model Monte Carlo description and the of LEP2 data;
- to use the QUAERO framework to perform several searches for new physics;
- to provide the ALEPH collaboration with a powerful tool in their data archiving effort;
- and to assess the QUAERO framework with searches for new physics at the LHC in mind.

The future use of ALEPH data by former ALEPH members and their collaborators has been anticipated by the ALEPH collaboration and formalized in Ref. [60]. At the time of this writing, both VISTA@ALEPH and QUAERO@ALEPH are password restricted to ALEPH members and documented in Ref. [61].

Sections 4.1-4.3 describe the data collected with the ALEPH detector, the particle identification procedure, and the Standard Model processes that define the reference hypothesis to which

$E_{\rm CM}$ (GeV)	183	189	192	196	200	202	205	207
$\int \mathcal{L}\, dt$ (pb^{-1})	56.82	174.21	28.93	79.83	86.30	41.90	81.41	133.21

Table 4.1 Integrated luminosity of the data available in QUAERO@ALEPH for each nominal LEP 2 center of mass energy.

alternative hypotheses are compared. In Section 4.4, an inclusive comparison of the data and the Standard Model prediction across many final states is presented.

4.1 Data

The approach taken in this chapter and the next is to look at the LEP2 data as inclusively as possible. This approach is complementary to the very exclusive event selection used in most searches for new physics, in which only a small subset of the data is considered.

It is a challenging task to provide a particle identification procedure that works well for *all* events (see Section 4.2). It is even more challenging to provide a Monte Carlo description that describes every triggered event, including events with cosmic origin, beam halo, and beam-gas interactions. Many of these unusual events are removed by requiring either that events are classified as single photon candidates or requiring the event to have one or more tracks with four or more TPC hits, $d_0 < 5$ cm, and $z_0 < 20$ cm.[1] The integrated luminosity corresponding to the ALEPH data satisfying these criteria is listed in Table 4.1. Events are fixed to the nearest of these eight nominal center of mass energies.

In addition, the following criterion exclude events not anticipated in the Standard Model background description. Events containing no object with energy $E > 25$ GeV and $|\cos\theta| < 0.7$ are discarded. Events containing one or more objects with energy $E > 10$ GeV and $|\cos\theta| > 0.9$ are discarded. Events containing one or more photons, missing energy, and no other objects have a large cosmic ray contribution, and are discarded. Events containing leptons separated by greater than 3.13 radians in azimuth are contaminated by cosmic rays and misidentified as e^+e^- events; they are also discarded.

[1] These events are selected with the ALPHA card `CLAS 21,5,6`

With the exception of the above requirements, all events recorded by ALEPH during LEP2 (roughly 6×10^4 events) have been included in VISTA@ALEPH, a comparison between the Standard Model prediction and the data, and QUAERO@ALEPH, an automated search procedure.

4.2 Particle Identification

The data and Monte Carlo events are analyzed with a specific ALPHA analysis algorithm ARCH, developed by the author[2], which identifies electrons ($e^{\pm}$), photons (γ), muons ($\mu^{\pm}$), taus ($\tau^{\pm}$), jets (j), and b-tagged jets (b).

The ARCH algorithm first identifies isolated electrons, photons, and muons from the energy flow objects. Remaining energy flow objects are clustered into mini jets and subjected to isolation and track criteria to identify taus. Energy flow objects not identified as photons, electrons, muons, or taus are clustered into jets, and the heavy flavor content of each jet is tested using QIPBTAG to identify b-jets.

Electrons, photons, muons, and taus were required to have an isolation cone with opening angle greater than 10°. The isolation cone of an object was defined as the cone that includes 5% of the event energy excluding objects within 2° and constituents of the object in question [63].

For the identification of electrons ($e^{\pm}$), complementary measurements of $\mathrm{d}E/\mathrm{d}x$ from the TPC and the longitudinal and transverse shape of the shower of the energy deposition measured in ECAL are used to build the normally distributed estimators R_I, R_L and R_T. These estimators are calibrated as a function of the electron momentum and polar angle for data and simulation using Bhabha events from LEP1 and LEP2, with electron energies from 20 to 100 GeV. To identify a track as an electron, the estimators R_I and R_L are required to be greater than -2.5, while R_T must be greater than -8. In ECAL crack regions, these criteria are supplemented by the requirement that the number of fired HCAL planes does not exceed ten. The measured momentum of the electrons is improved by combining it with the energy deposits in ECAL associated with both the electron and possible bremsstrahlung as it passes through the detector.

[2]The ARCH algorithm grew from an original implementation by Marcello Maggi, which was similar to the algorithm used in Ref. [62]. The ARCH algorithm differs in several respects, with lepton identification completely rewritten.

Photons (γ) are identified via the energy flow object's particle identification information and are required to be isolated.

Muons ($\mu^{\pm}$) are identified using the tracking capability of HCAL and the muon chambers. A road is defined by extrapolating tracks through the calorimeter and muon chambers and counting the number of observed hits on the digital readout strips. To reduce spurious signals from noise, a hit is considered only when fewer than four adjacent strips fire. For a track to be identified as a muon the total number of hits must exceed 40% of the number expected, with hits in at least five of the last ten planes and one of the last three. To eliminate misidentified muons due to hadron showers, cuts are made on the mean cluster multiplicity observed in the last half of the HCAL planes. Within the HCAL and muon chamber crack regions, muons are identified by requiring that the energy deposits in ECAL and HCAL be less than 10% of the track momentum and not greater than 1 and 5 GeV, respectively.

The process of tau ($\tau^{\pm}$) identification begins with the clustering of energy deposits into mini jets using the Jade algorithm [64] with $y_{\rm cut} = 0.001$ relative to the total energy in the event.[3] Isolated mini jets that consist of only one charged track, that consist of two charged tracks with invariant mass less than 2 GeV, or that consist of three charged tracks with invariant mass less than 3 GeV and with total charge ± 1 are identified as taus.

Jets are clustered with the Durham algorithm [66] with $y_{\rm cut} = 0.001$ when constructing the fast simulation described in Section 5.2, and with $y_{\rm cut} = 0.01$ otherwise. Jets containing no charged tracks and not identified as photons are classified as unclustered energy. Isolated jets with exactly two charged tracks consistent with an electron positron pair are identified as photons. Jets with $P_{uds} < 0.01$ from the flavor tagging package QIPBTAG are identified as b jets. Other jets are simply identified as jets.

Missing energy ($\not{p}$) is defined as the negative vector sum of the 4-vectors of the objects identified in the event, neglecting the contribution of energy visible in the detector but not clustered into one of these objects.

[3] Tau identification based on clusters formed with the Jade algorithm [64] with $y_{\rm cut} = (2.7\,{\rm GeV}/E_{\rm vis})^2$, used in the standard Higgs analysis [65], resulted in a larger than desired fraction of jets being misreconstructed as taus.

4.3 Backgrounds

Eight categories of Standard Model processes are generated to serve as the reference model to which hypotheses presented to QUAERO will be compared. Here and below "Standard Model," "background," and "reference model" will be used interchangeably.

$\mathrm{q\bar{q}}$ The process $e^+e^- \to Z/\gamma^* \to q\bar{q}(\gamma)$ is modeled using KK 4.14 [67], with initial state radiation from KK and final state radiation from PYTHIA.

e^+e^- Bhabha scattering and $e^+e^- \to Z/\gamma^* \to e^+e^-(\gamma)$ is modeled using BHWIDE 1.01 [68].

$\mu^+\mu^-$ Pair production of muons, $e^+e^- \to Z/\gamma^* \to \mu^+\mu^-(\gamma)$, is calculated using KK 4.14 [67], including initial and final state radiative corrections and their interference.

$\tau^+\tau^-$ Pair production of taus, $e^+e^- \to Z/\gamma^* \to \tau^+\tau^-(\gamma)$, is calculated using KK 4.14 [67], including initial and final state radiative corrections and their interference.

1ph Single photon production, $e^+e^- \to Z/\gamma^* \to \nu\bar{\nu}(\gamma)$, is included in the background estimate.

Nph Multiphoton production, $e^+e^- \to n\gamma$, with $n \geq 2$, is included in the background estimate.

4f Four fermion events compatible with WW final states are generated using KoralW 1.51 [69], with quarks fragmented into parton showers and hadronized using either PYTHIA 6.1 [38]. Events with final states incompatible with WW production but compatible with ZZ production are generated with PYTHIA 6.1.

2ph Two-photon interaction processes, $\mathrm{e^+e^- \to e^+e^-}X$, are generated with the PHOT02 generator [70]. When X is a pair of leptons, a QED calculation is used with preselection cuts to preferentially generate events that mimic WW production. When X is a multi-hadronic state, a modified version of PYTHIA is used to generate events with the incident beam electron and positron scattered at $\theta < 12^\circ$ and $168^\circ < \theta$, respectively. Events in which the beam electron or positron is scattered through an angle of more than 12° are generated using HERWIG 6.2 [39].

Additional details are available in Refs. [71, 72].

Roughly 47 million Monte Carlo events have been generated and processed through GALEPH (the ALEPH detector simulation), JULIA (the ALEPH reconstruction), and ARCH. The combination of GALEPH, JULIA, and ARCH are denoted for brevity by ALEPHSIM. The ALEPH data events and these Standard Model Monte Carlo events are reduced to 4-vectors of final state objects and stored as text files. At an average of 200 bytes per event, 10 GB is sufficient for their storage. These event sizes make it technically feasible to provide the entire LEP2 and Monte Carlo data sets in a well documented and machine-independent format. These text files are now a part of the ALEPH archival data.

4.4 Comparison of Data and Standad Model Predictions

The first step of both VISTA@ALEPH and QUAERO@ALEPH is to partition all events into exclusive final states based on the particle identification criteria implemented in ARCH. Final states are labeled according to the number and types of objects present, and are ordered according to decreasing discrepancy between the total number of events expected and the total number observed in the data.

Events in the final states e^+j and e^-j come overwhelmingly from e^+e^- with one of the electrons failing electron identification criteria; in this case the jet is promoted to an electron of the opposite sign of the identified electron in the event, and placed in the final state e^+e^-.

Figures 4.1-4.5 summarize the bulk of this comparison.[4] In each figure, the deviation of the data from the Monte Carlo expectation is shown in terms of Gaussian σ. This is achieved in four steps. In the first step, a Poisson distribution is constructed based on the expected number of background events. In the second step, the uncertainty on the number of background events is marginalized according to the Cousins-Highland formalism (see Appendix C). In the third step, the background confidence level, CL_b, is obtained according to Equation A.1. Finally, that CL_b is transformed into σ according to Equation A.3. In this transformation, an excess (deficit) of events becomes a positive (negative) σ value. The error bars on the data correspond to the deviation that

[4] Events with less than five events were excluded from the figures.

would have been observed if the observed number of events, x, fluctuated down to $x - \sqrt{x}$ or up to $x + \sqrt{x}$. The yellow bars, mainly meant to guide the eye, indicate the background uncertainty translated into σ.

The final state $e^-\mu^+$, containing one negatively charged electron and one positively charged muon, is the most discrepant final state observed. In addition, there are four final states with $\sim 3\sigma$ discrepancies. How compelling are these results? Do they indicate a failure of the Standard Model?

One must procede with caution before claiming any failure of the Standard Model. Figures 4.1-4.5 summarize the outcome of 253 nearly independent experiments. One *should expect* to see an experiment with at least a 2.6σ deviation. Figure 4.6 shows the distribution of deviation for 418 final states and the expected normal distribution with a mean of zero and standard deviation of unity. If one does not transform into σ, the distribution of CL_b (also shown in Figure 4.6) should be flat by construction.

The distribution in terms of Gaussian σ is in good agreement with expectations, with the exception of the $e^-\mu^+$ final state. The full comparison, including plots of thousands of kinematic distributions, can be viewed on the web (password restricted to members of the ALEPH collaboration) at Ref. [73]. It is quite impressive that such a wide variety of physics processes can be predicted from a relatively concise Monte Carlo description and analyzed with a single general-purpose algorithm.

Given that the inclusive comparison of data and the Standard Model prediction is quite reasonable, we can confidently proceed to search for specific new physics signatures with the QUAERO algorithm.

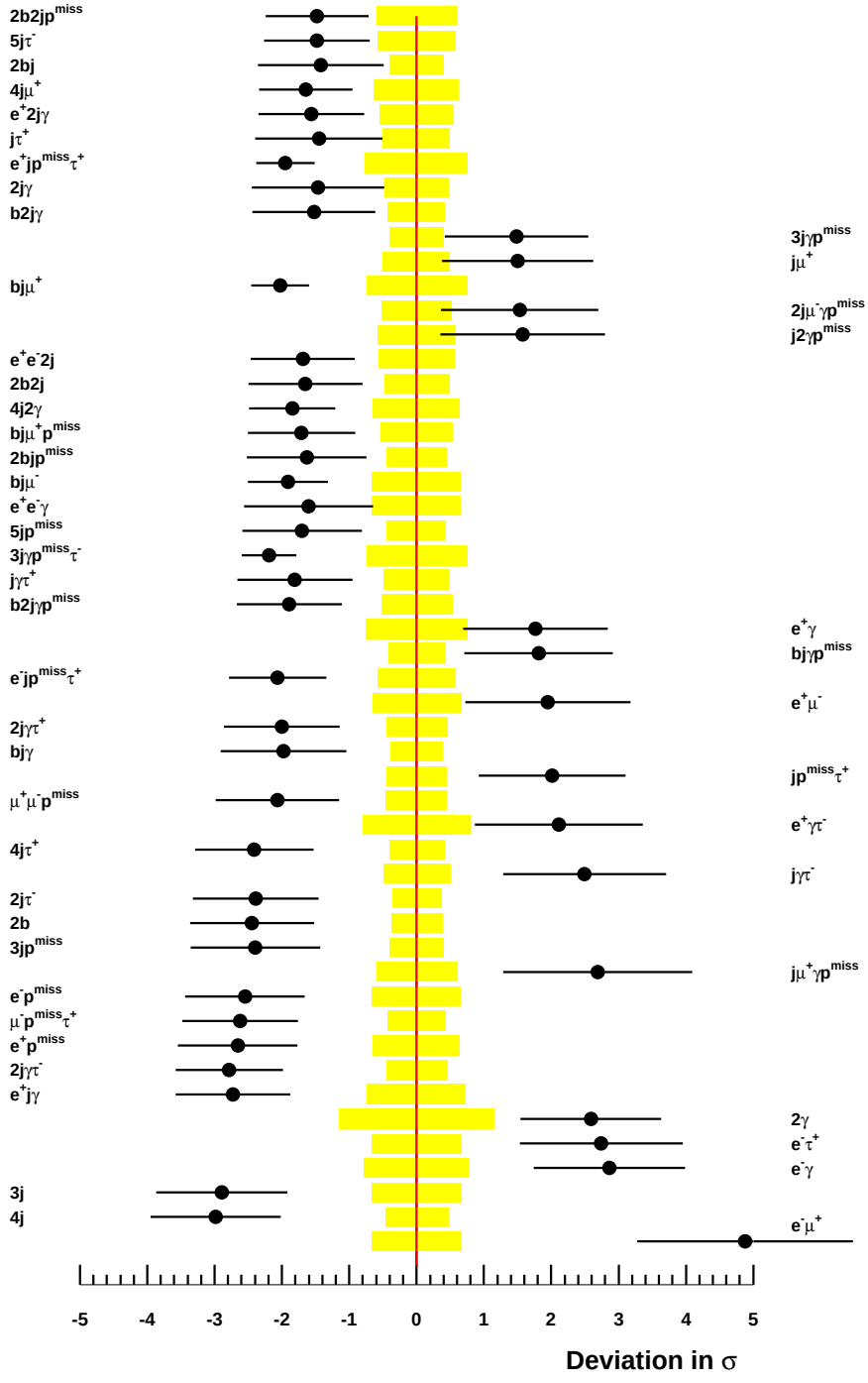


Figure 4.1 A subset of the VISTA@ALEPH comparison of data and Monte Carlo.

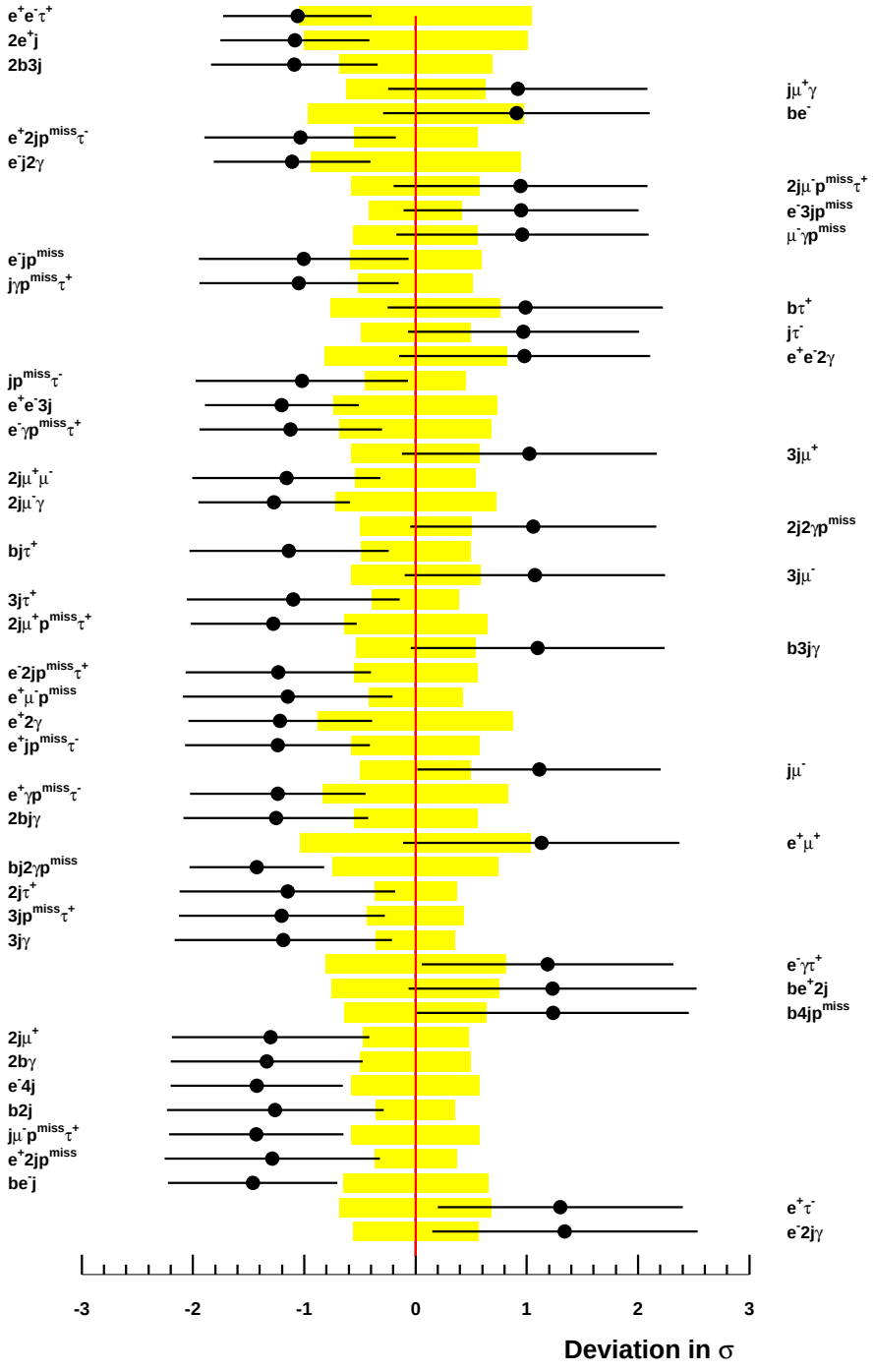


Figure 4.2 A subset of the VISTA@ALEPH comparison of data and Monte Carlo.

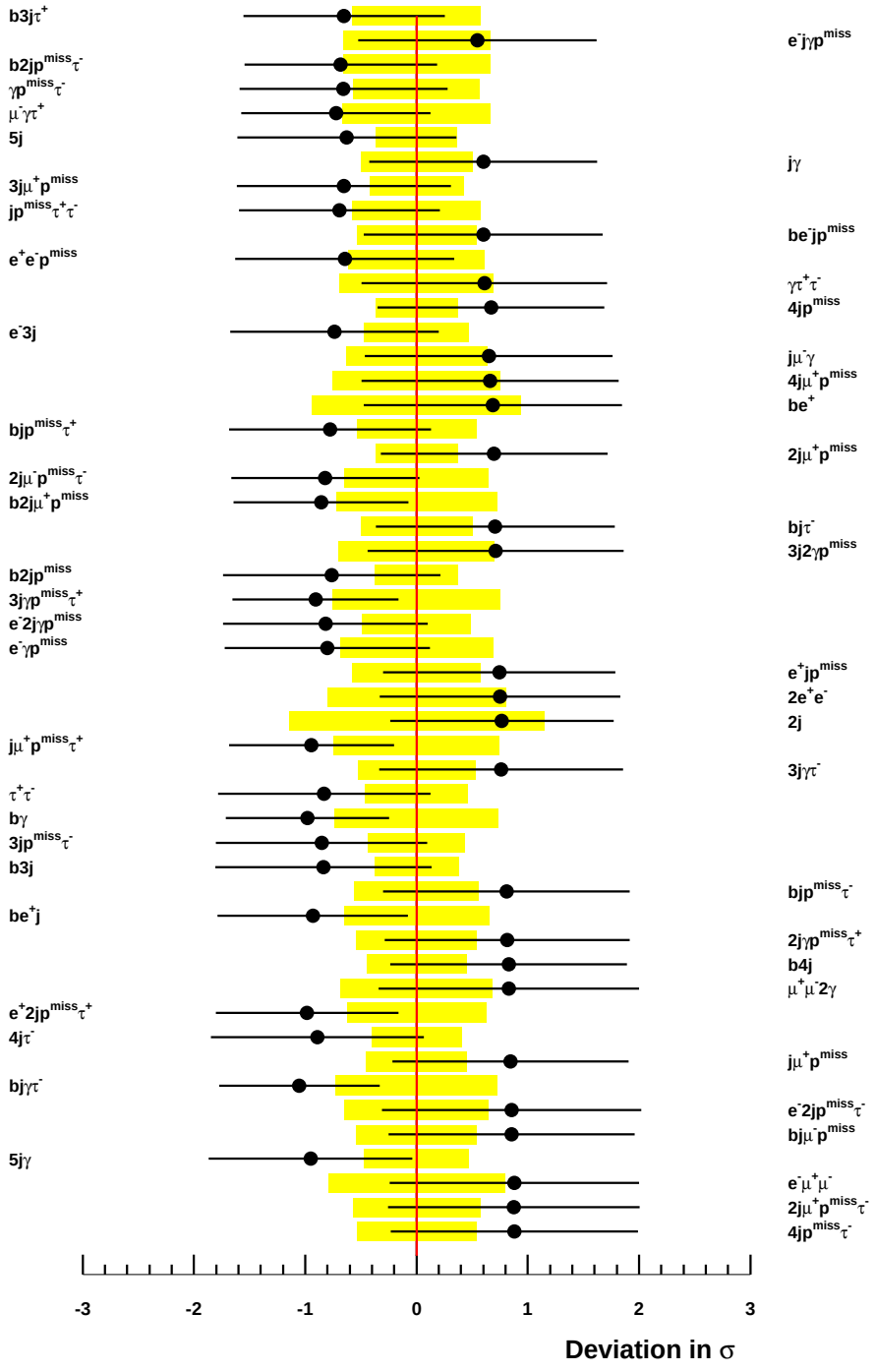


Figure 4.3 A subset of the VISTA@ALEPH comparison of data and Monte Carlo.

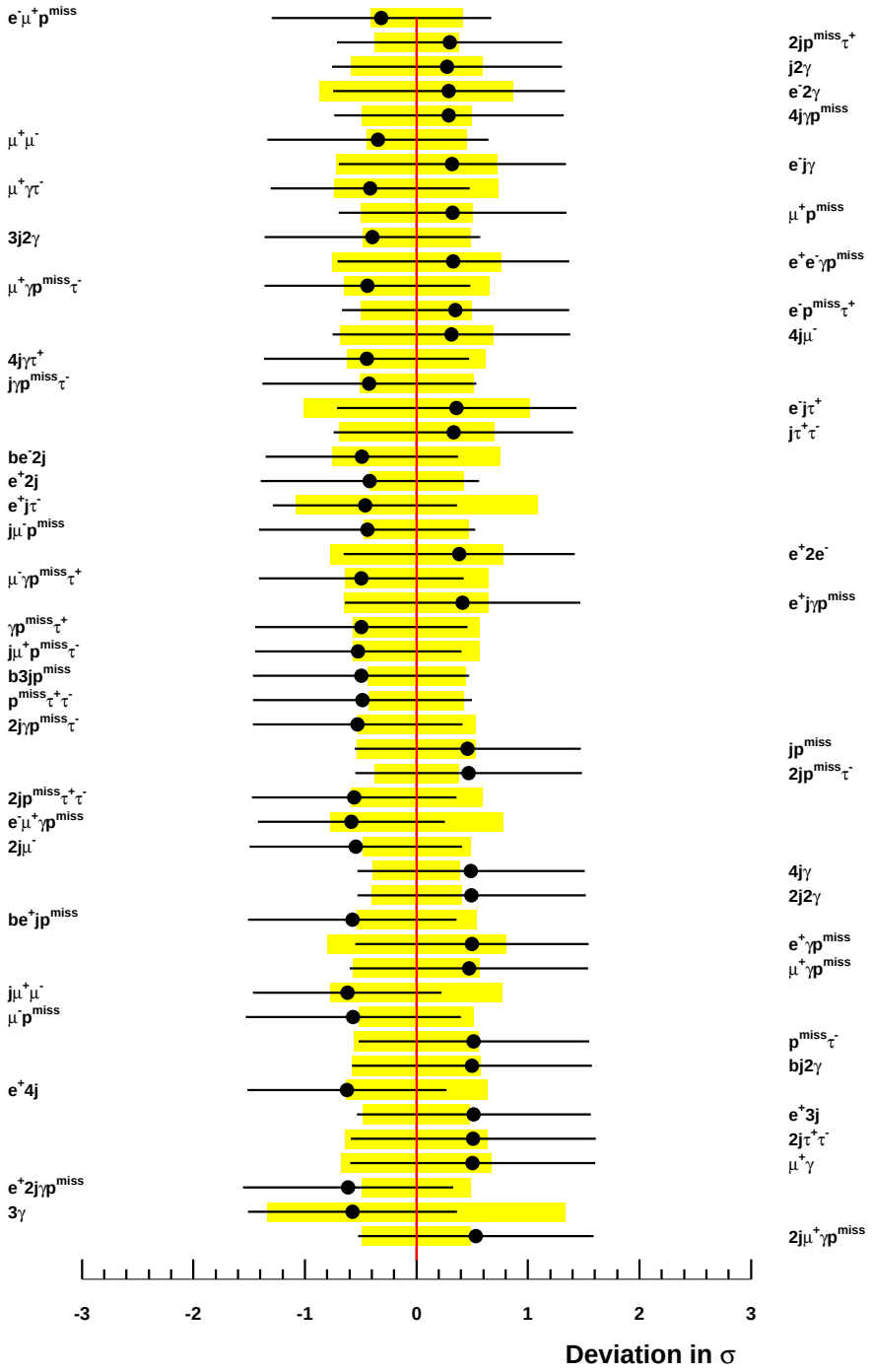


Figure 4.4 A subset of the VISTA@ALEPH comparison of data and Monte Carlo.

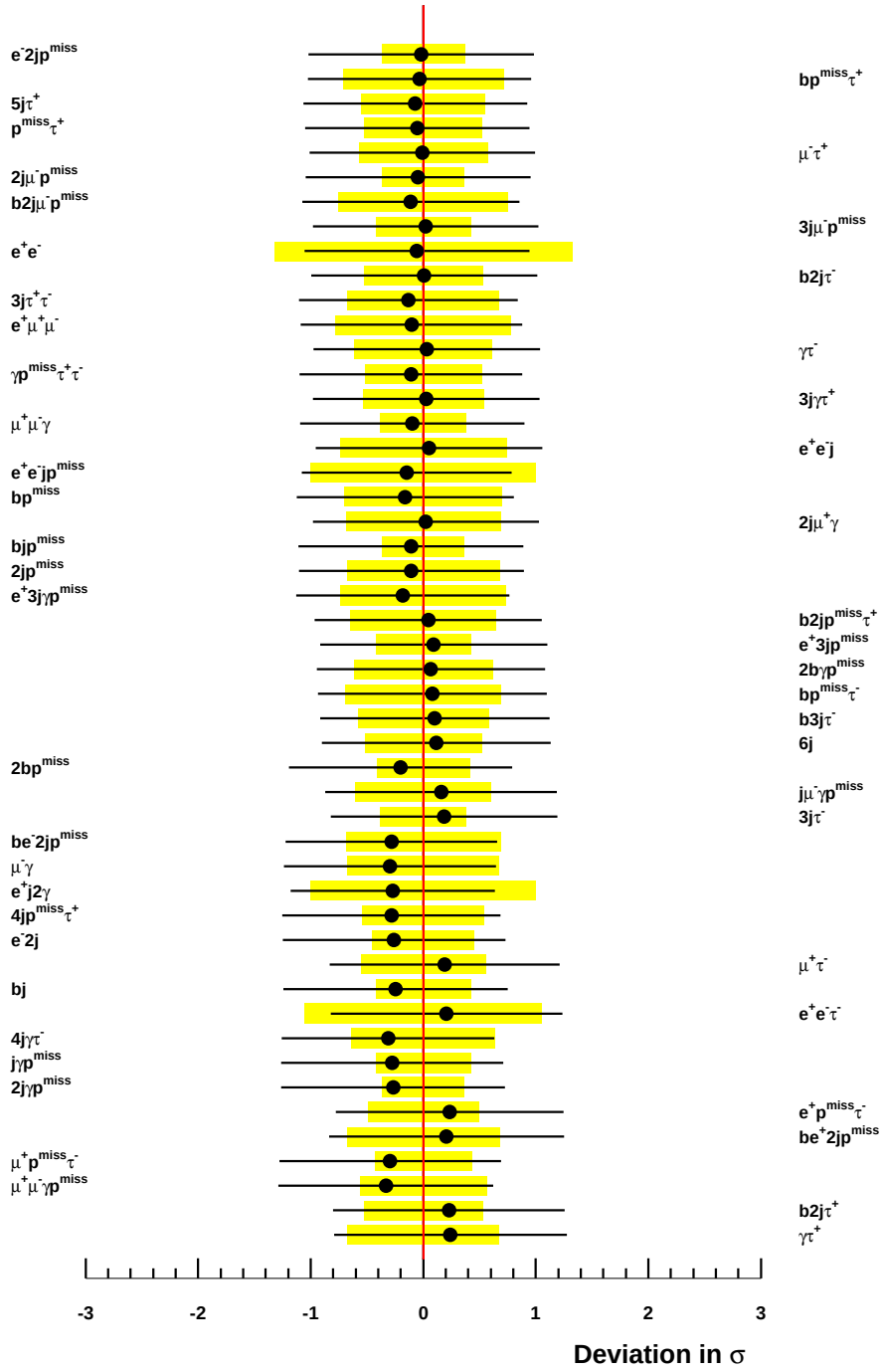


Figure 4.5 A subset of the VISTA@ALEPH comparison of data and Monte Carlo.

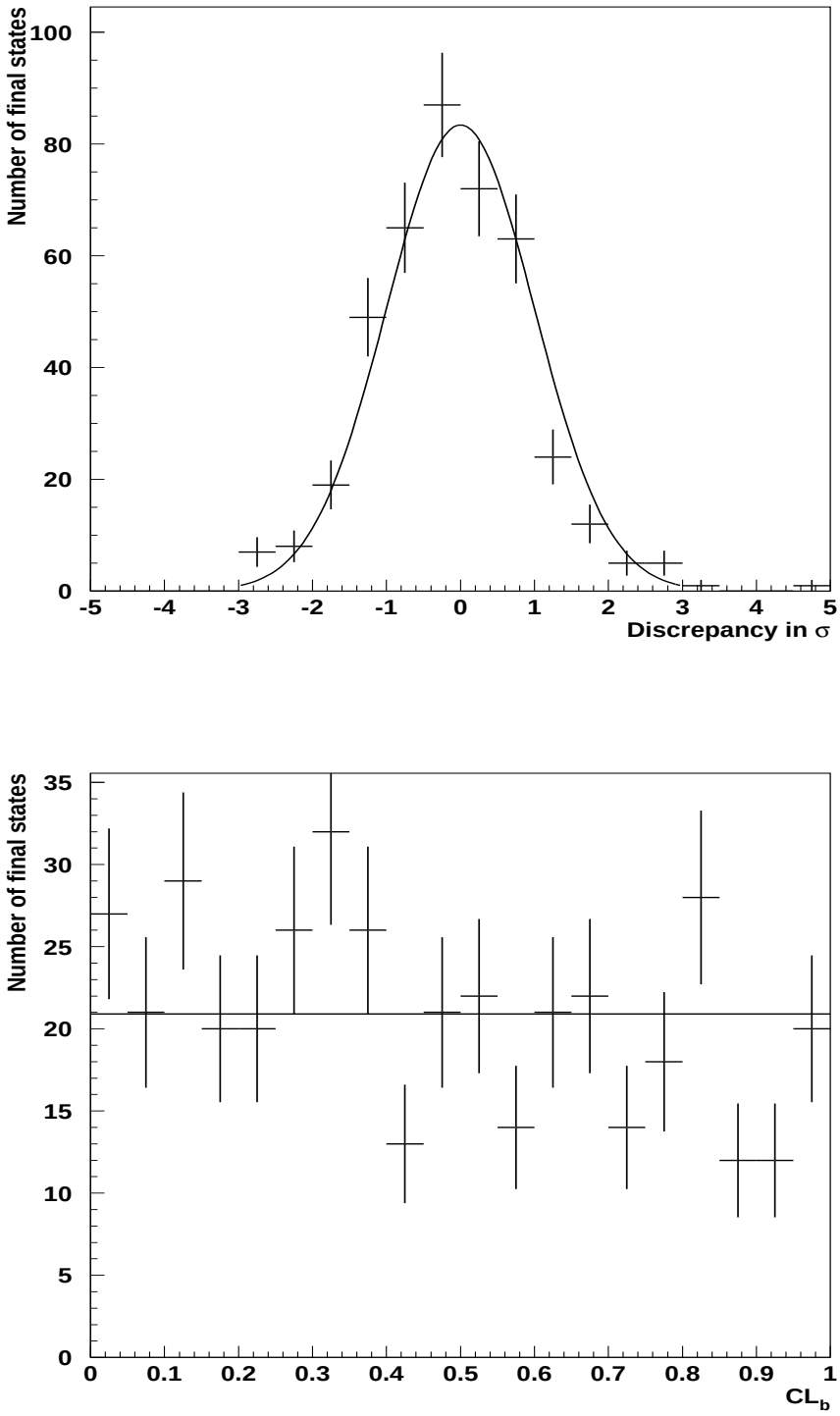


Figure 4.6 Distribution of data-Monte Carlo discrepancy in terms of Gaussian σ (left) and background confidence-level, CL_b (right). The solid curves show the expectated distribution.

Final State	Events Observed	Events Predicted
$e^-\mu^+$	53	20.4 ($\tau^+\tau^-$ = 7.4 , 4f = 7.4 , 2ph = 4.9 , $\mu^+\mu^-$ = 0.6)
$e^+\mu^-$	38	25.1 ($\tau^+\tau^-$ = 10 , 4f = 7.9 , 2ph = 7 , $\mu^+\mu^-$ = 0.2)
$e^-\mu^+p^{\rm miss}$	109	112.3 (4f = 86.9 , $\tau^+\tau^-$ = 24 , 2ph = 0.9 , $\mu^+\mu^-$ = 0.4)
$e^+\mu^-p^{\rm miss}$	99	111.6 (4f = 90.2 , $\tau^+\tau^-$ = 19.7 , 2ph = 1.3 , $\mu^+\mu^-$ = 0.4)

Table 4.2 Number of events and observed for $e^\pm\mu^\mp$ and $e^\pm\mu^\mp p^{\rm miss}$ final states.

4.5 The $e^\mp\mu^\pm$ Final State

The comparison shown in the previous section found that the the $e^-\mu^+$ and $e^+\mu^-$ final states had a larger discrepancy from Standard Model predictions than one would expect. The exact numbers of events observed and predicted (broken down by background contribution) are shown in Table 4.2. The Standard Model prediction does not take into account systematic differences in the particle identification between data and Monte Carlo. The $e^\mp\mu^\pm$ final state was examined in some detail, but an extensive systematic study has not yet been performed . The first potential explanation for the discrepancy from conventional sources comes from systematic differences in the particle identification between data and Monte Carlo. For instance, the sum of events in $e^+\mu^-$ and $e^+\mu^-p^{\rm miss}$ are in agreement between data and Monte Carlo. However, this is not the case for the $e^-\mu^+$ channel, and the distribution of $p^{\rm miss}$ is not steeply falling near the 10 GeV cut on $p^{\rm miss}$.

Figure 4.7 shows the distribution of the electron and muon energies as well as their invariant mass and azimuthal separation. The color codes the various background contributions. The excess appears to have two components: one evenly distributed and one from back-to-back pairs with invariant mass near $\sqrt{s}$. Figure 4.8 shows two events from data classified in one of the $e^\pm\mu^\mp$ final states. The top figure shows two back-to-back pairs of charged particles. The bottom figure shows four charged particles: two of which appear to be muons, and two of which are likely electrons from a photon conversion. The second event is consistant with $e^+e^- \rightarrow Z\gamma \rightarrow \mu^+\mu^-e^+e^-$. Neither of these events is a clean $e^\pm\mu^\mp$ candidate. This systematics of this channel must undergo a detailed study before definitive statments can be made.

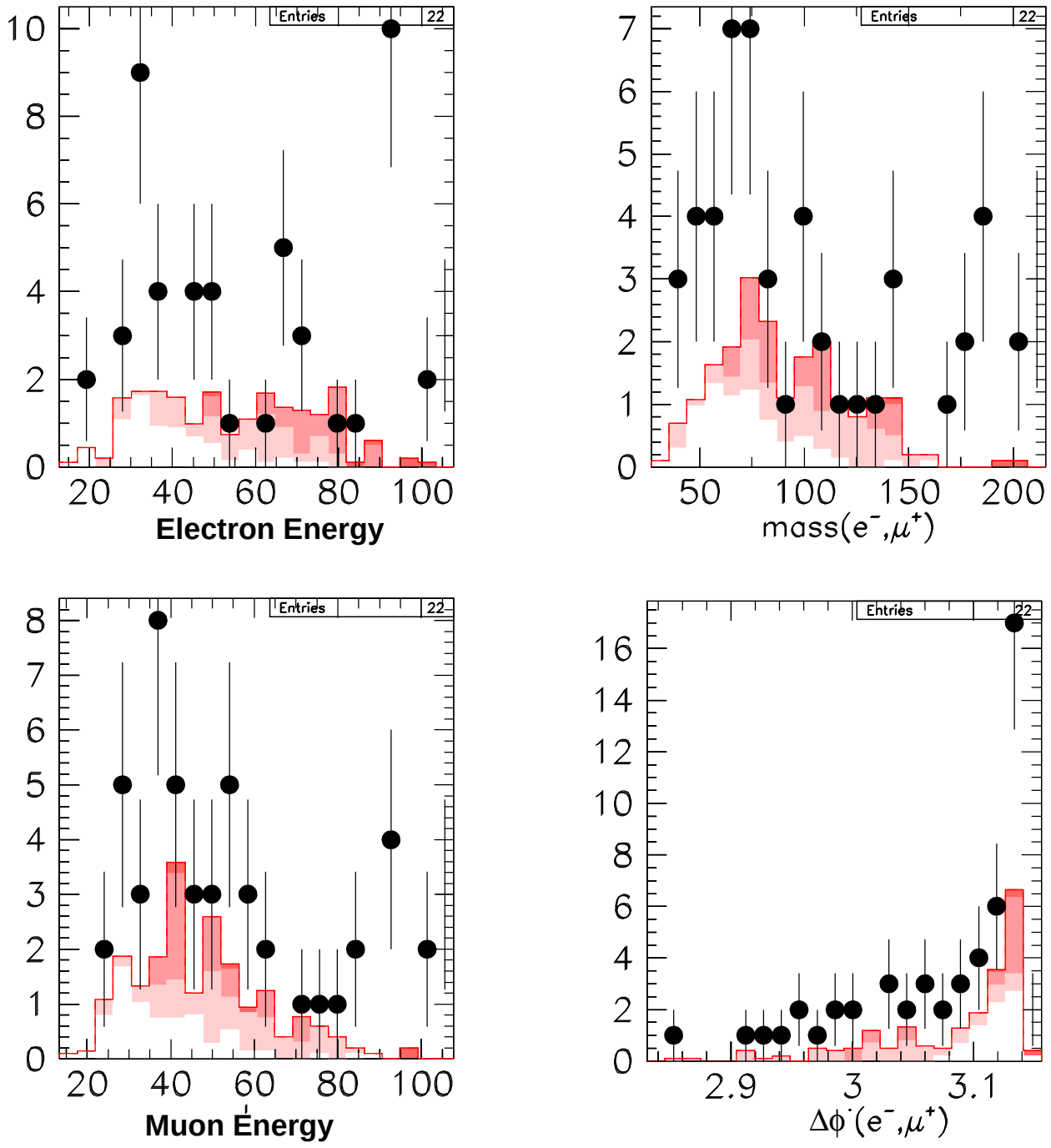


Figure 4.7 Kinematic distributions for the $e^{-}\mu^{+}$ final state.

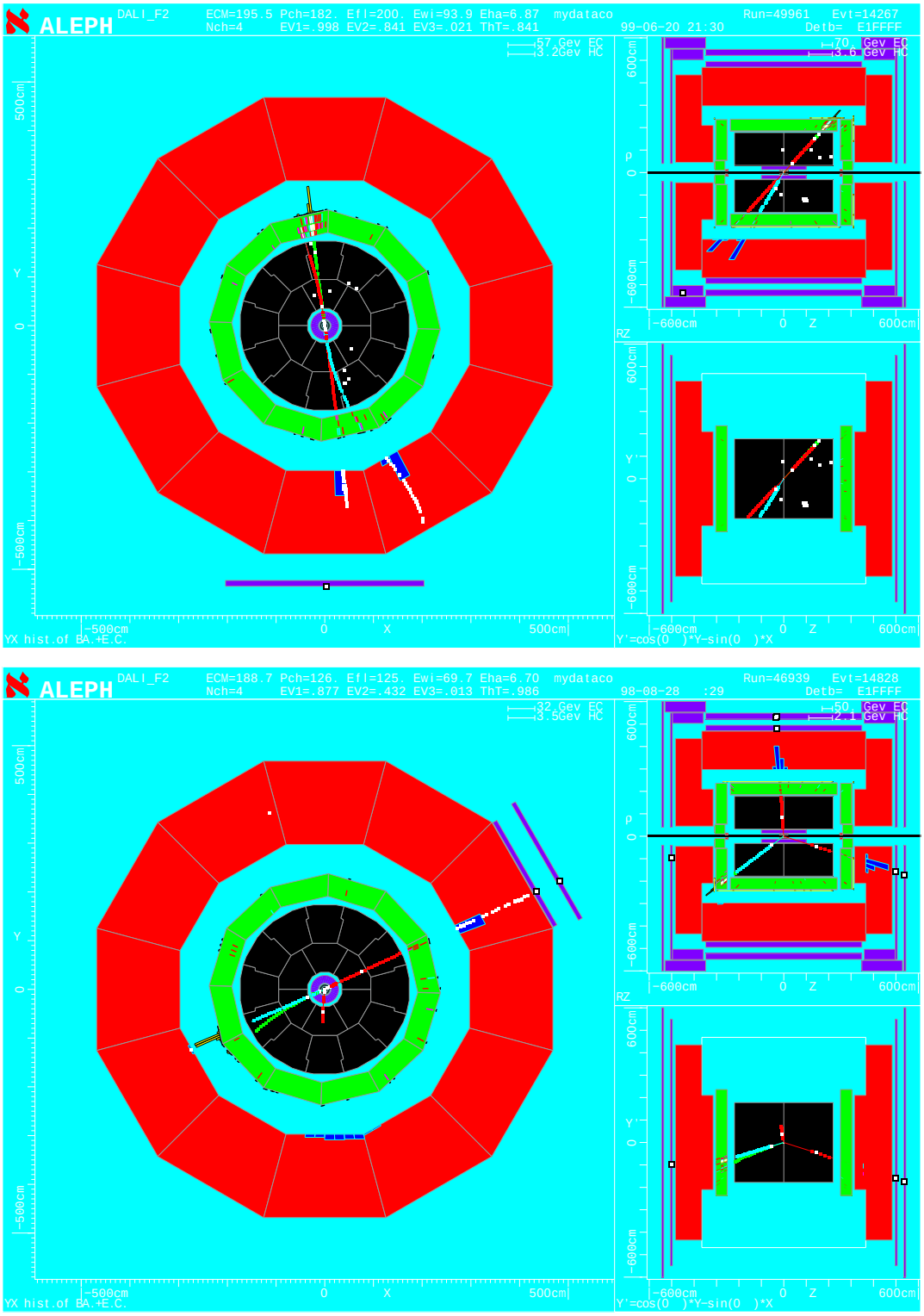


Figure 4.8 Event displays for two events mis-reconstructed in the $e^{\pm}\mu^{\mp}$ final state.

End of tracks or

Chapter 5

Quaero@Aleph

QUAERO is an automated search procedure based on high-level reconstructed objects. QUAERO does not attempt to automate reconstruction or particle identification. Restricting QUAERO's input to only these high-level objects has two consequences. First, it allows for the analysis procedure to be robust and intuitive. If the method attempted to refine particle identification or reconstruction algorithms, then it would be difficult to understand, prone to finding local maxima, and not trustworthy. Secondly, the restriction to high-level objects reduces the power of an analysis performed with QUAERO. Clearly, the typical analysis strategy, which involves a huge amount of time refining particle identification and reconstruction for a particular signature, uses more information and is more powerful. Given these consideration, the relevant question is "Is QUAERO powerful enough?". The answer to this question comes in Section 5.6, after we review the algorithm and its performance in several real-world examples.

Sections 5.1-5.4 describe the QUAERO algorithm, the fast simulation used for signal events, systematics, and statistical interpretation. Section 5.5 contains the results of several analyses that have been performed using QUAERO@ALEPH, allowing comparison to previous ALEPH publications. A summary is given in Section 5.6.

5.1 The Quaero Algorithm

A physicist wishing to test a particular hypothesis against ALEPH data will provide, either in the form of commands to PYTHIA or as a STDHEP file, events predicted by this hypothesis. The response of the ALEPH detector to these signal events is simulated using TURBOSIM@ALEPH.

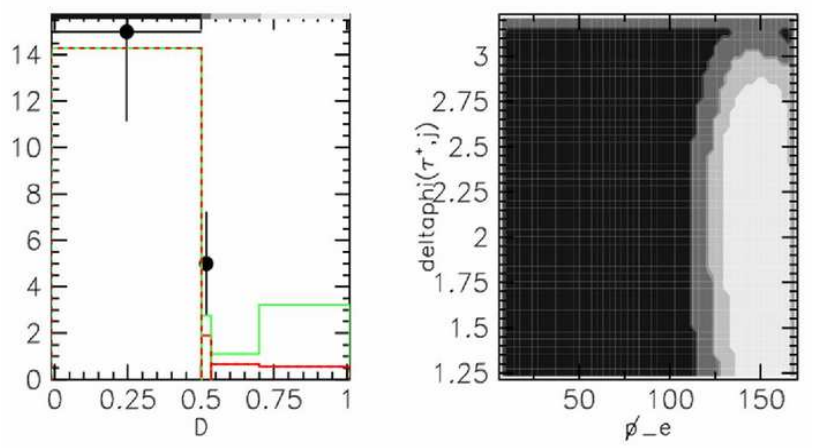


Figure 5.1 QUAERO's automatic variable selection and choice of binning in the final state $j\not{p}\tau^+$, testing the hypothesis of mSUGRA with $\mu > 0$, $A_0 = 0$, $\tan(\beta) = 10$, $M_0 = 100$, and $M_{1/2} = 120$, using data collected at 205 GeV.

Three distinct samples of events exist at this point: the data $\mathcal{D}$; the Standard Model prediction SM; and the hypothesis $\mathcal{H}$, which is the sum of included Standard Model processes and the physicist's signal. In each exclusive final state, a pre-defined list of variables — including object energies, polar angles, and azimuthal angles; angles between object pairs; and invariant mass of object combinations — are ranked according to the difference between the Standard Model prediction and the physicist's hypothesis $\mathcal{H}$. The top d variables in this list are used after removing highly correlated variables, where d is limited to between zero and three by the number of Monte Carlo events available to populate the resulting variable space.

In this variable space, X, multivariate densities are estimated from the Monte Carlo events predicted by SM and $\mathcal{H}$. These densities are used to define a discriminant, D, defined as

$$D(x) = \frac{f_s(x)}{f_s(x) + f_{\mathrm{SM}}(x)}, \tag{5.1}$$

where $x \in X$, $f_s(x)$ and $f_{\mathrm{SM}}(x)$ are probability density functions estimated with a kernel estimation technique similar to those described in Appendix B. This discriminant[1] is one-to-one with the likelihood ratio (see Appendix D). Bins are formed in the variable space with boundaries defined by the contours of $D(x)$. Finally, the likelihood ratio $Q = L(\mathcal{D}|\mathcal{H})/L(\mathcal{D}|\mathrm{SM})$ is determined using

[1] This approach to event selection was the author's first introduction to High Energy Physics, with Hannu E. Miettinen. This approach was used in Refs [74], [75], and [76].

this binning, and systematic errors are integrated numerically. Figure 5.1 shows a two dimensional variable space with contours of $D(x)$ (right) and a histogram of the number of events between these contours (left) for the mSUGRA search described in Section 5.5.

Further details of the QUAERO algorithm are provided in Ref. [77].

5.2 TurboSim@Aleph

To keep QUAERO fast and standalone, ALEPHSIM [2] has been used to construct a fast detector simulation (TURBOSIM@ALEPH). The TURBOSIM algorithm is described in Ref. [78]. This section focuses on the application of this algorithm to the ALEPH detector.

The approach of TURBOSIM is not to model the ALEPH detector independently, but to take full advantage of the effort and expertise that has gone into GALEPH, JULIA, and ARCH. To that end, the events used to define the Standard Model prediction for incorporation into QUAERO have been used to construct a large lookup table of one half million lines mapping particle-level objects to objects reconstructed in the ALEPH detector. Events from all Standard Model background processes at LEP have been incorporated into this table. Sample lines in this table are shown in Figure 5.2. The total table is roughly 100 MB, and as such can be read into memory and searched as a multivariate binary tree. The resulting simulation runs at roughly 10 ms per event.

Particle identification efficiencies are handled through lines in the TURBOSIM@ALEPH table that map a particle level object to no reconstructed level object. Misidentification probabilities are handled through lines that map a particle-level object to a reconstructed-level object of a different type. The merging and overlap of particles is handled by configurations in the table that map two or three particle-level objects to zero or more reconstructed-level objects.

Each line in Figure 5.2 begins with the event's type and run and event number. To the left of the arrow ("->") is a list of nearby particle-level objects; to the right of the arrow is a list of corresponding reconstructed-level objects. The first line shows a b quark incorrectly identified as a jet and a tau, while the second line shows a b quark that has been correctly identified. The third line shows a jet that has been split into two jets; in the fourth line the jet is sufficiently far forward

[2] The combination of GALEPH, JULIA, and ARCH are denoted for brevity by ALEPHSIM.

that it has not been identified. The fifth line shows an electron close to a jet that has been correctly identified as an electron; the sixth line shows an electron close to a jet that has been merged into a single jet. The seventh line shows a correctly reconstructed positron; the eighth line shows a correctly reconstructed muon; the ninth line shows a correctly reconstructed tau. The tenth line shows two nearby jets that are reconstructed as two jets and a low energy photon.

Validation of TURBOSIM@ALEPH has been performed by running a large, independent set of events through both simulations, categorizing the events into exclusive final states, and comparing the distributions of relevant kinematic variables (object momenta, polar angles, and azimuthal angles; angles between object pairs; and invariant masses of all object combinations). The four distributions shown in Figure 5.3 are among the most discrepant of over 3000 distributions and 300 final states.

One must be aware of two important facts when considering Figure 5.3. First, while the events in the comparison are independent from those used to train TURBOSIM@ALEPH, the two distributions are highly correlated.[3] Second, events classified in a particular final state by the full simulation are not necessarily classified in the same final state by TURBOSIM@ALEPH. Somewhat surprisingly, TURBOSIM@ALEPH does quite a good job at reproducing ALEPHSIM.

One should also be aware of the following bias that may be introduced by TURBOSIM's parametric approach that are not introduced by the algorithmic simulation paradigm. When presented with a final state object from a new physics signature, the TURBOSIM lookup table will only have the events used to train it for reference. For instance, a new physics signature at the LHC might involve an electron with an energy of 1 TeV. If the TURBOSIM lookup table did not include a sample of 1 TeV electrons in the Standard Model training set, then it will be biased towards those events used for training. However, at LEP this is not of much concern; largely due to the fact that the entire center of mass energy often is visible in the final state. The sample of four fermion events alone provides a training sample, which nicely covers momentum spectra for each type of final state particle; however, events from each of the eight background processes have been included.

[3] Note that if the same events were processed twice with the full simulation, but with different random number seeds, the resulting comparison would still show some deviations.

```
 1    4f   10.11884     b     73.46    0.61  -1.95   ;   ->   j     45.32    0.54   -1.86
                                                              tau+  25.69    0.66   -2.11    ;
 2    4f   10.20754     b     45.02   -0.29   2.29   ;   ->   b     48.17   -0.30    2.30    ;
 3    4f   10.22333     j     63.92   -0.72  -2.23   ;   ->   j     40.16   -0.75   -2.19
                                                              j     26.78   -0.66   -2.22    ;
 4    4f   10.22324     j     26.23    0.94  -1.82   ;   ->   ;
 5    4f   20.8473      e-    21.12   -0.55   1.39
                        j      5.16   -0.48   1.92   ;   ->   e-    20.69   -0.55    1.4     ;
 6    4f   20.11826     e+    65.36   -0.75  -0.34
                        j     18.05   -0.59  -0.27   ;   ->   j     88.79   -0.75   -0.35    ;
 7    4f   70.17426     e+    68.59    0.23  -0.41   ;   ->   e+    66.62    0.23   -0.41    ;
 8    4f   50.21469     mu-   70.42   -0.51   1.60   ;   ->   mu-   69.05   -0.51    1.60    ;
 9    4f   50.17707     tau+  56.30    0.66   0.80   ;   ->   tau+  57.88    0.66    0.80    ;
10    4f  100.2892      j     46.37    0.00  -2.49
                        j     16.06   -0.17  -2.20   ;   ->   j     23.16    0.02   -2.64
                                                              j     26.96   -0.20   -2.45
                                                              ph     9.68    0.09   -2.34    ;
```

Figure 5.2 Ten sample lines in the TURBOSIM@ALEPH lookup table, chosen to illustrate TURBOSIM's handling of interesting cases.

5.3 Systematic Errors

The experimental sources of systematic error affecting the modeling of these data are detailed below. The evaluation of systematic errors in ALEPHSIM and TURBOSIM@ALEPH is unusual because there is no independent control sample in the data from which to estimate the systematics. The systematics listed below are quite conservative global estimates.

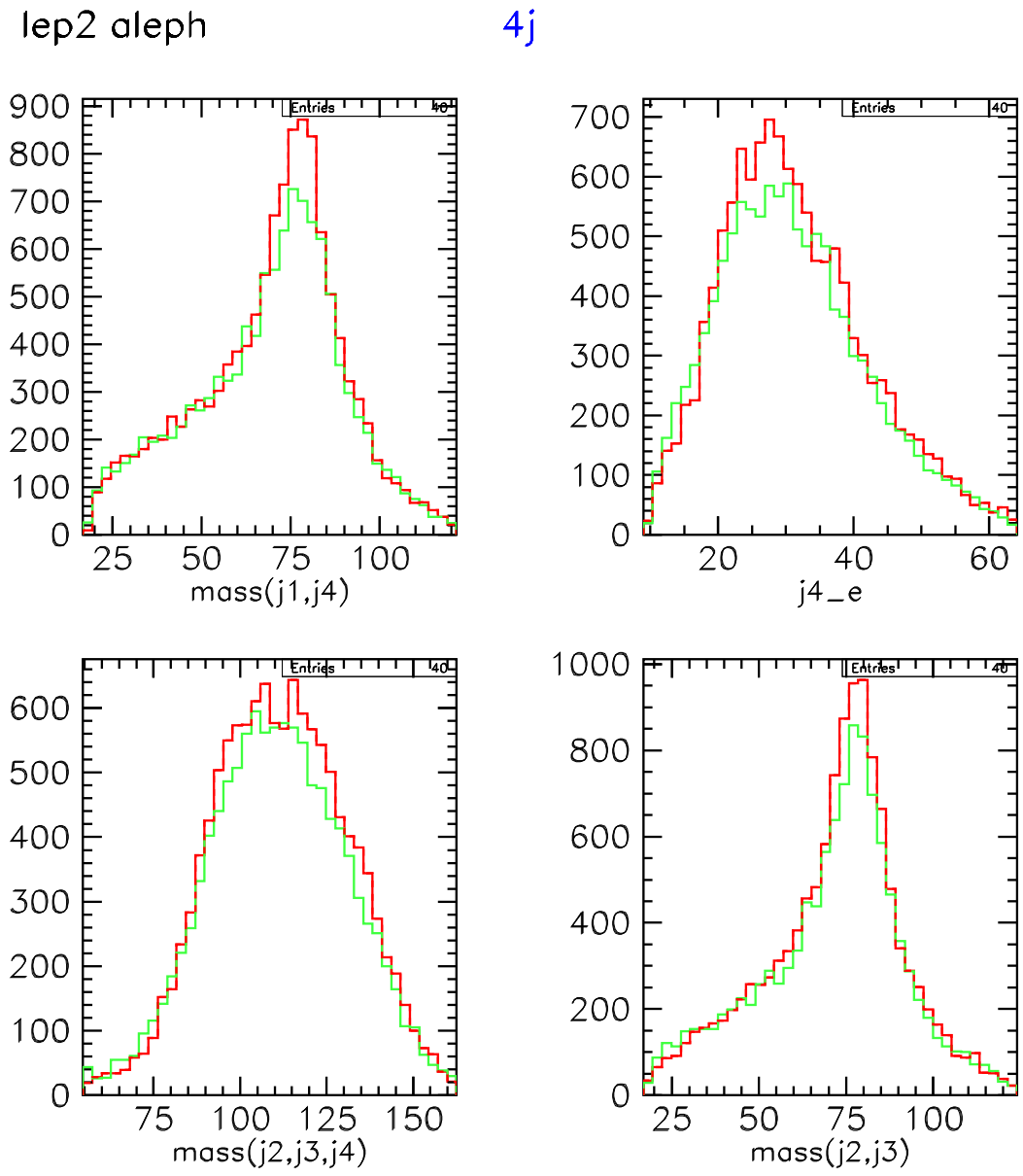
Errors affecting the weight of each Monte Carlo event include a 0.1% uncertainty in the ALEPH LEP2 luminosity. Uncertainties in the energy of each object are specified in addition to uncertainties in the overall event weight. Electrons and photons suffer from an electromagnetic energy scale uncertainty of 1%. Jets suffer from a 2% hadronic energy scale uncertainty, and muons from a momentum uncertainty of 2% . All events processed using TURBOSIM@ALEPH are subjected to an uncertainty on the event weight of 10%.

Although QUAERO allows a full specification of correlated uncertainties, all ALEPH sources of systematic error are treated as uncorrelated. Futhermore, QUAERO@ALEPH does not currently include systematic differences between the data and Monte Carlo's particle identification effeciencies. For instance, events in the final states e^+j and e^-j come overwhelmingly from e^+e^- with one of the electrons failing electron identification criteria; in this case the jet is promoted to an electron of the opposite sign of the identified electron in the event, and placed in the final state e^+e^-.

5.4 Statistical Interpretation of Quaero Results

The result returned by QUAERO is the decimal logarithm of the likelihood ratio Q. The log likelihood ratio was used by the LEP Higgs working group as an intermediate quantity, but it was not the final result *per se*. In a frequentist setting, one would be interested in distributions of Q for the SM and $\mathcal{H}$. Integrals of these distributions define the rates of Type I and Type II errors (see Section A.1).

The likelihood ratio is still useful in a Bayesian context, when considered as an update of betting odds. For instance, the hypothesis of mSUGRA with $\mu > 0$, $A_0 = 0$, $\tan(\beta) = 10$, $M_0 = 100$, and $M_{1/2} = 120$, QUAERO returns $\log_{10} Q = -3.24$ considering only data at 205 GeV.



1

Figure 5.3 Comparison of the output of TURBOSIM@ALEPH (light, green) and ALEPHSIM (dark, red).

If betting odds on this hypothesis were 100:1 against before looking at these data, then these data indicate those odds should be adjusted by an extra factor of $10^{-3.24} \approx 1700$:1 against. Betting odds against this hypothesis after having run this request are now 170000:1. Betting odds against this hypothesis after having run this request using data at all center of mass energies are over one billion to one against.

The estimation of a model parameter is possible with QUAERO. It is accomplished, by maximizing $\log_{10} Q$ (or Q weighted by the prior) with respect to the model parameter, with multiple QUAERO submissions.

Providing an 95% exclusion limit on a model parameter is not possible in a formal frequentist or Bayesian setting with only the result Q. Consider the case of a search with b expected background events and s expected signal events, where $s, b \gg 1$ so that we can make the Gaussian approximation. The the signal hypothesis would be excluded at the 95% level if the number observed events, x, was less than the critical $x^* = s + b - 2\sqrt{s+b}$. The likelihood ratio $Q(x^*)$ depends on the ratio s/b. If $s/b \gg 1$, then $Q(x^*) \to \infty$. If the signal was such that the expected background result would provide a 95% exclusion (*i.e.* $b = x^*$), then $Q(x^*) \to 1/e^2 \approx 0.135$. On the otherhand, if $s/b \to 0$, then $Q(x^*) \to 1$. In the last case, the experiment has no sensitivity, and an exclusion of the signal is equivalent to an exclusion of the background. This unwelcome situation is the motivation for the CL_s method described in Section A.1.6.

Intuitively, an exclusion region corresponds to a region where the signal hypothesis is moderately disfavored by the data. In the remaining sections, $\log_{10} Q = -1$ is used as a rough and convenient choice for the purpose of building intution when comparing with previous results. It should be noted that this notion of exclusion is equivalent to a much higher level of exclusion when $s/b \to \infty$, is more conservative in the case $b = x^*$, and would not exclude a signal when $s/b = 0$.

5.5 Searches Performed with Quaero

QUAERO has been used to test models that have previously been considered at ALEPH, in order to benchmark QUAERO's sensitivity. Additionally, QUAERO has been used to test models

for which ALEPH has no official result. These examples allow us to build some intuition for the QUAERO algorithm: its strengths, and its limitations.

Because the previous results take the form of 95% confidence level exclusions, which cannot be determined from Q, it is difficult to make a direct comparison with previous results. A rough, non-rigorous, but nonetheless useful comparison of the sensitivity of QUAERO's results can be made by comparing $\log_{10} Q = -1$ with the 95% confidence level exclusion limit.

The examples considered in this section include a test of mSUGRA, a search for excited electrons, and searches for doubly charged, singly charged, and neutral Higgs bosons.

5.5.1 mSUGRA

In order to build intuition for the strengths and limitations of QUAERO, QUAERO@ALEPH was first used to test minimal-Supergravity. An interesting feature of the QUAERO analysis strategy, which is true of the other searches as well, is that it spans all final states in one consistant analysis. Typically, the searches for a phenomenologically complicated model examine the different final states independently, and the combination of these individual results is performed as an extra step. It is not always clear that the individual analyses are performed with a consistant set of assumptions; in which case, the combined results is questionable. In contrast, the QUAERO approach is able to span final states, perform a consistant combined analysis, and test the model point in its entirety.

In addition to the familiar parameters of the Standard Model, mSUGRA is defined by four parameters and a sign. Three parameters live at the GUT scale: the scalar mass, M_0 ; the fermionic mass, $M_{1/2}$; and the trilinear couplings, A_0. The remaining parameters are defined at a low-energy scale. These include the ratio of the two vacuum expectation values, $\tan\beta$, and the sign of the supersymmetric Higgs mass parameter μ.

For convenience of comparison to a previous ALEPH result, we take $\tan(\beta) = 10$, $\mu > 0$, and $A_0 = 0$. The mass parameters M_0 and $M_{1/2}$ are allowed to range up to 1 TeV, and between 100 and 200 GeV, respectively. In addition, R-parity conservation is assumed.

QUAERO's automatic variable selection and choice of binning in the final state $j\not{p}\tau^+$, testing the hypothesis of mSUGRA with $\mu > 0$, $A_0 = 0$, $\tan(\beta) = 10$, $M_0 = 100$ GeV, and $M_{1/2} = 120$ GeV, using data collected at 205 GeV is shown in Figure 5.1. Bins of the discriminant (left) correspond to bins in the chosen two-dimensional variable space (right), which is formed by the difference in azimuthal angle between the tau and jet and the missing energy. The vertical axis in the left plot shows the number of events in each bin in the discriminant D. The axes in the right plot have units of radians and GeV. Lighter shades of gray indicate regions preferentially populated by events from the querying physicist's hypothesis, while darker shades of gray indicate regions preferentially populated by events from the Standard Model. The connection between the bins in the two plots is indicated by the shades of gray across the top of the left plot.

QUAERO's selection of bin boundaries near 125 GeV in missing energy is expected, since the stable lightest neutralino carries away much of the energy in the event, as seen in the missing energy distributions for the related final states $e^- j\not{p}$ and $j\mu^+\not{p}$ in the upper right and lower right panes of Figure 5.4. QUAERO also makes use of angular relationships in these events, recognizing that the Standard Model contribution to this final state tends to produce anti-aligned jets and "taus" (usually mistaken jets), while the supersymmetric signal does not.

Figure 5.4 shows the four final states contributing most to the final result are $2j\not{p}$, $e^- j\not{p}$, $j\not{p}\tau^+$, and $j\mu^+\not{p}$. In all four final states QUAERO chooses missing energy ($\not{p}_e$) as a particularly useful variable, since the lightest supersymmetric particle carries away much of the energy in the signal events. The difference in azimuthal angle between the tau and jet in the final state $j\not{p}\tau^+$ edges out missing energy as the most useful variable in this final state because the Standard Model contribution comes from $e^+e^- \to Z/\gamma^* \to q\bar{q}$ when a jet is mistakenly identified as a tau, so that the jet and mistaken tau are back to back in azimuth, while this is not true of the signal processes.

QUAERO's analysis of mSUGRA for fixed $\mu > 0$, $\tan\beta = 10$, and $A_0 = 0$, and in the two-dimensional box defined by $0 < M_0 < 1000$ and $100 < M_{1/2} < 200$, is shown in Figure 5.5. Regions shown in red are disfavored by the data, relative to the Standard Model; deeper shades of red are used for each order of magnitude in the likelihood ratio. Any region favored by the data, relative to the Standard Model, would be shown in green, with deeper shapes of green used for

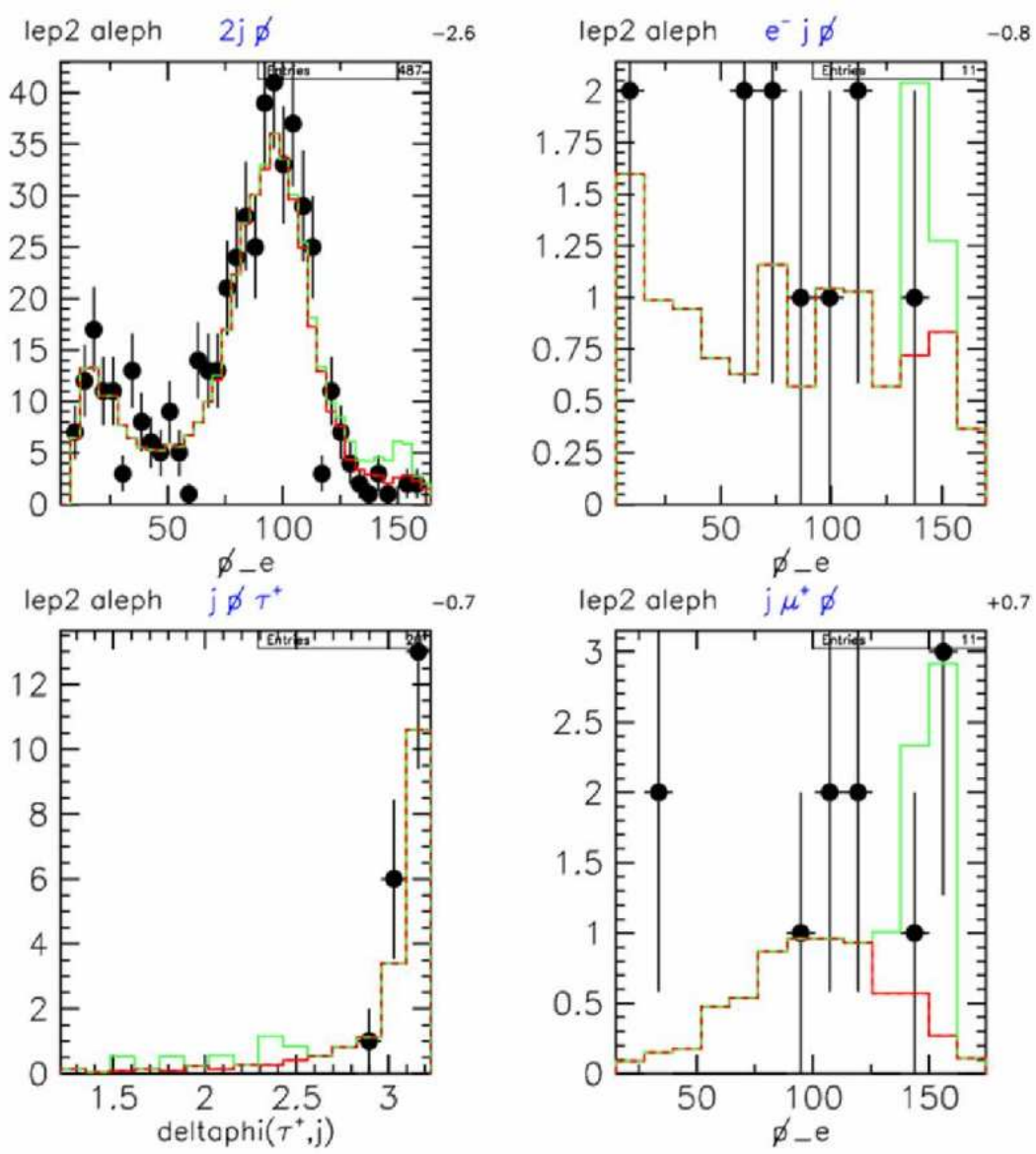


Figure 5.4 Plots of the standard model prediction (dark, red), the querying physicist's hypothesis (light, green), and the ALEPH data (filled circles) for the single most useful variable in all final states contributing more than 0.1 to $\log_{10} Q$.

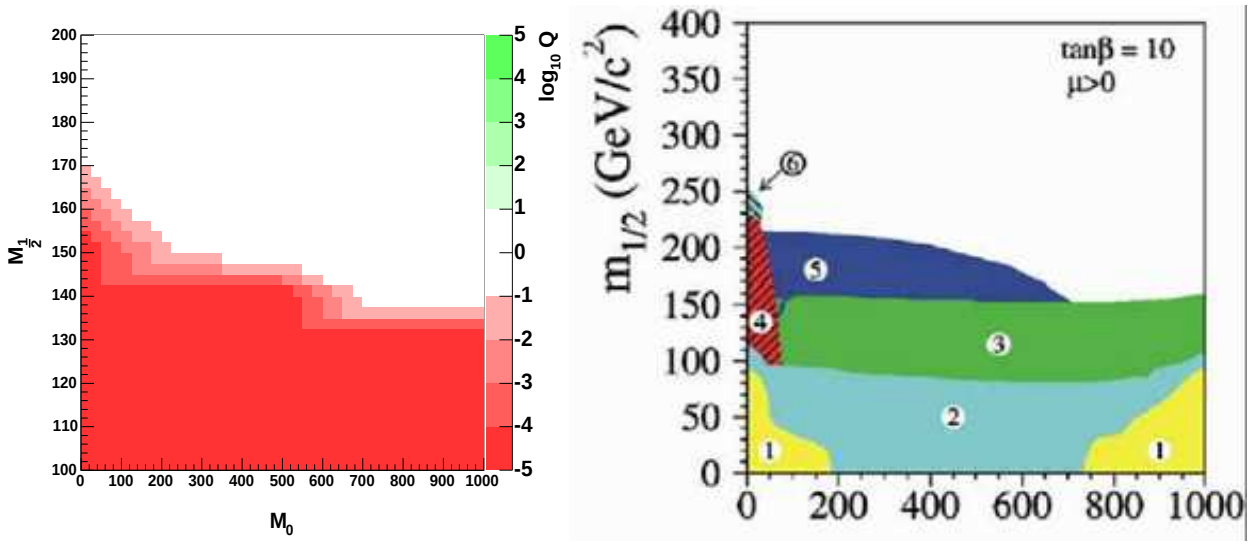


Figure 5.5 QUAERO's output ($\log_{10} Q$) as a function of assumed $M_{1/2}$ and M_0, for fixed $\tan\beta = 10$, $A_0 = 0$, and $\mu > 0$.

each order of magnitude in the likelihood ratio. In all cases QUAERO finds $\log_{10} Q \lesssim 0$, indicating the ALEPH data favor the Standard Model over the provided hypotheses.

As was mentioned in Section 5.4, a 95% confidence level exclusion region does not follow from the contours of the likelihood ratio. However, if we choose the exclusion threshold to be 10:1 against ($\log_{10} Q = -1$), then the portion of the parameter space with $M_{1/2} < 135$ GeVis excluded for values of M_0 up to 1 TeV.

This result is in accord with the result of a previous analysis of this signal, described in Ref. [79], which derived similar limits in this parameter space.

5.5.2 Excited electrons

A search for excited electrons has also been performed using QUAERO. Excited electrons are predicted by a larg class of models in which quarks and leptons are composite objects. These models are attractive because the weak mixing angles and fermion masses become calculable parameters [80]. The relevant parameters of the model are the coupling constants f, f', and f_s –

corresponding to the Standard Model gauge groups $SU(2)$, $U(1)$, and $SU(3)$ – the scale parameter Λ and the excited electron mass m_{e^*}.

Excited electrons have been searched for in a previous analysis from the OPAL collaboration, described in Ref. [81] and shown in the right plot of Figure 5.6. Assuming equality of the $SU(2)$ and $U(1)$ couplings ($f = f'$) and no $SU(3)$ coupling ($f_s = 0$), regions in the parameter space of f/Λ and m_{e^*} above the curves shown in Figure 5.6 are excluded by the OPAL analysis, which also considers the possibility of excited muon and tau leptons. Ref. [81], however, does not provide an easy means for using OPAL data to test the hypothesis of excited electrons under different assumptions on the couplings.

To test a different set of assumptions, the new electroweak couplings are arranged to eliminate the coupling of the excited electron to the Z boson ($f = f' \tan^2 \theta_W$=0.28). The signal was generated with PYTHIA. Details of PYTHIA's excited lepton model can be found in Ref. [38].

A scan was performed on the remaining parameters m_{e^*} and Λ. The result of this scan is shown in Fig. 5.6. Regions shown in red are disfavored by the data, relative to the Standard Model; deeper shades of red are used for each order of magnitude in the likelihood ratio. A few regions of this parameter space are favored by the data relative to the Standard Model; these regions are shown in shades of green.

5.5.3 Doubly charged Higgs

A search for doubly charged Higgs bosons $H^{\pm\pm}$ in a left-right symmetric model with a Higgs triplet has also been performed using QUAERO. The signal was generated with PYTHIA. Taking the masses of the left and right doubly charged Higgs bosons to be equal, the single parameter of this model space is the mass $m_{H^{\pm\pm}}$. Tests of this model space for particular choices of this mass parameter are shown in Figure 5.7.

The result in Figure 5.7 would update betting odds for $m_{H^{\pm\pm}} < 98.5$ GeV to be more than 10:1 against. This can be qualitatively compared with a previous analysis from OPAL, described in Ref. [82] (also shown in Figure 5.7). QUAERO's result is in agreement with the 95% confidence limit of $m_{H^{\pm\pm}} > 98.5$ GeV determined in the previous OPAL analysis.

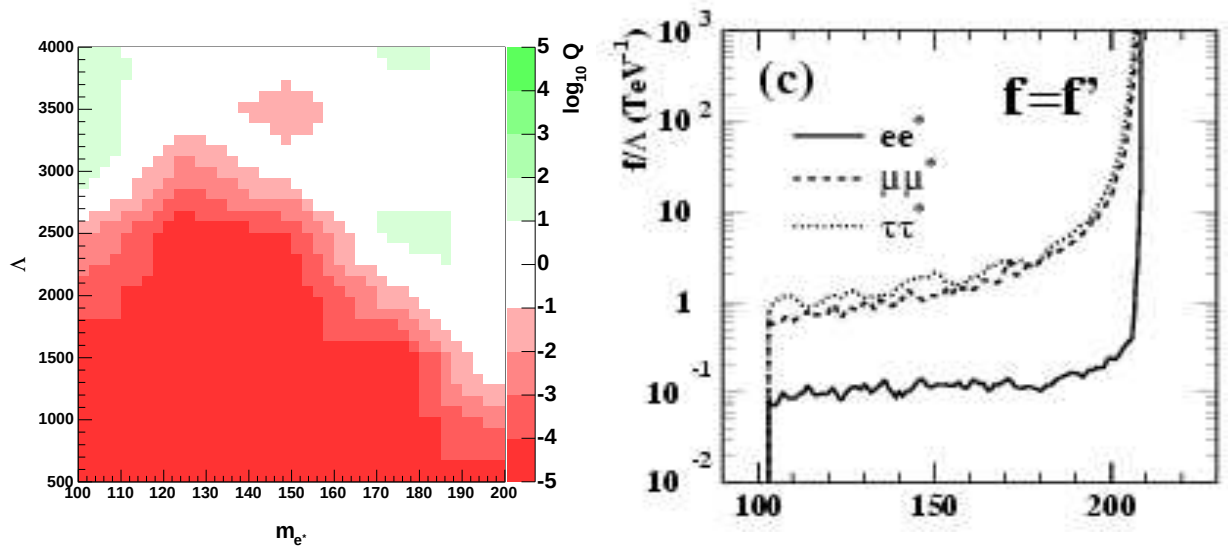


Figure 5.6 QUAERO's output ($\log_{10} Q$) as a function of assumed Λ and m_{e^*}, for fixed $f = f' \tan^2 \theta_W = 0.28$ and $f_s = 0$ (left). Exclusion contour summarizing a previous OPAL analysis of excited lepton parameter space (right).

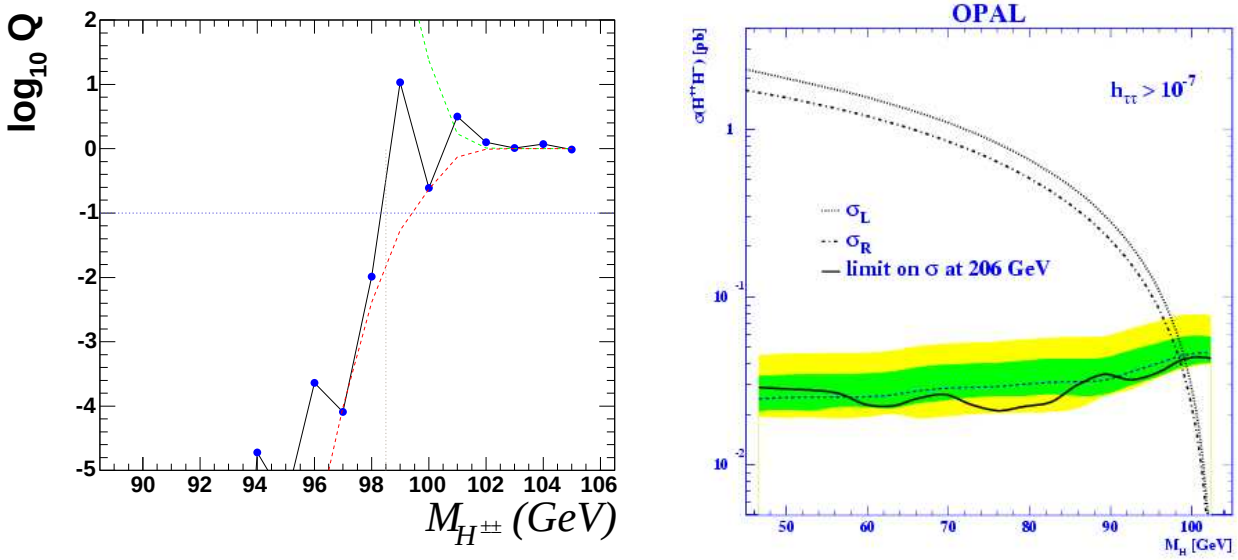


Figure 5.7 QUAERO's output ($\log_{10} Q$) as a function of assumed doubly charged Higgs mass $m_{H^{\pm\pm}}$, in the context of a left-right symmetric model containing a Higgs triplet (left). A previous OPAL analysis is also shown (right).

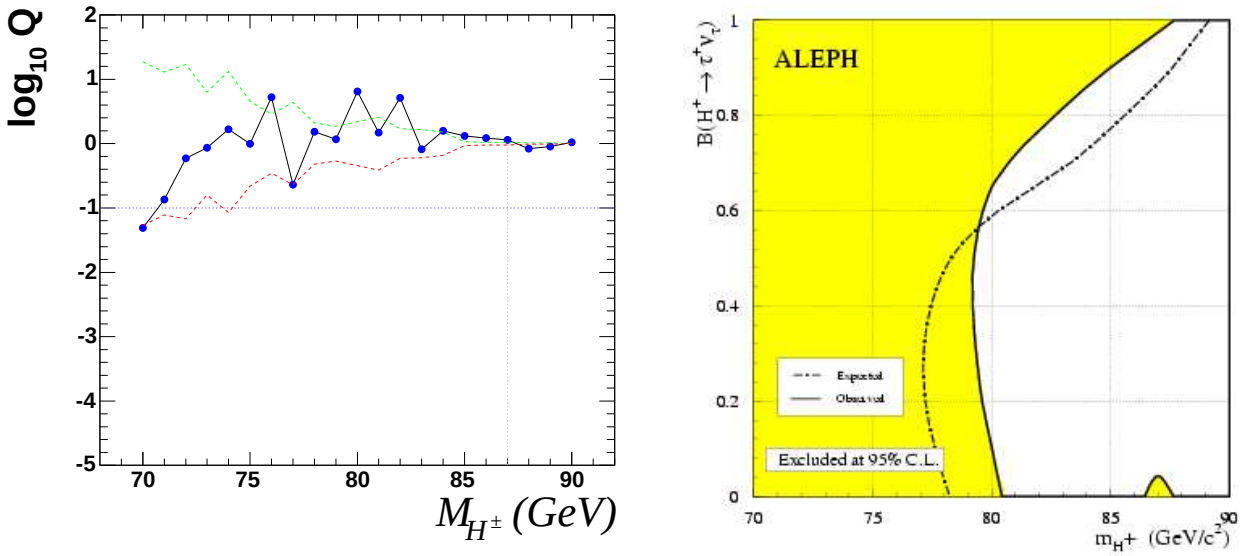


Figure 5.8 QUAERO's output ($\log_{10} Q$) as a function of assumed charged Higgs mass $m_{H^\pm}$, in the context of a generic two Higgs doublet model (left). A previous ALEPH result (right).

5.5.4 Charged Higgs

A search for charged Higgs bosons $H^\pm$ – predicted by generic two Higgs doublet models – has also been performed using QUAERO. Higgs doublet models are found in the MSSM and are strongly motivated. The signal was generated with PYTHIA, and the charged Higgs mass was scanned in the range 70 to 90 GeV. The result of this scan is shown in Figure 5.8 (left).

The vertical, red, dashed line in the left plot marks the exclusion limit from a previous analysis of ALEPH data, described in Ref. [83] (shown in the right plot of Figure 5.8). The horizontal (green) line in the left plot highlights $\log_{10} Q = -1$. The previous analysis allowed the charged Higgs boson branching ratio to tau and tau neutrino to vary; a limit of $m_{H^\pm} > 79.3$ GeV is determined at a confidence level of 95% for any choice of branching ratio of charged Higgs to $\tau\nu_\tau$.

Based on the variations in the observed values of $\log_{10} Q$ and the fact that the expected values $\log_{10} Q$, the QUAERO search for charged Higgs is not very powerful for charged Higgs masses near the mass of the W boson. For lower masses, $M_H < 70$ GeV, the analysis has more sensitivity and disfavors the charged Higgs hypothesis.

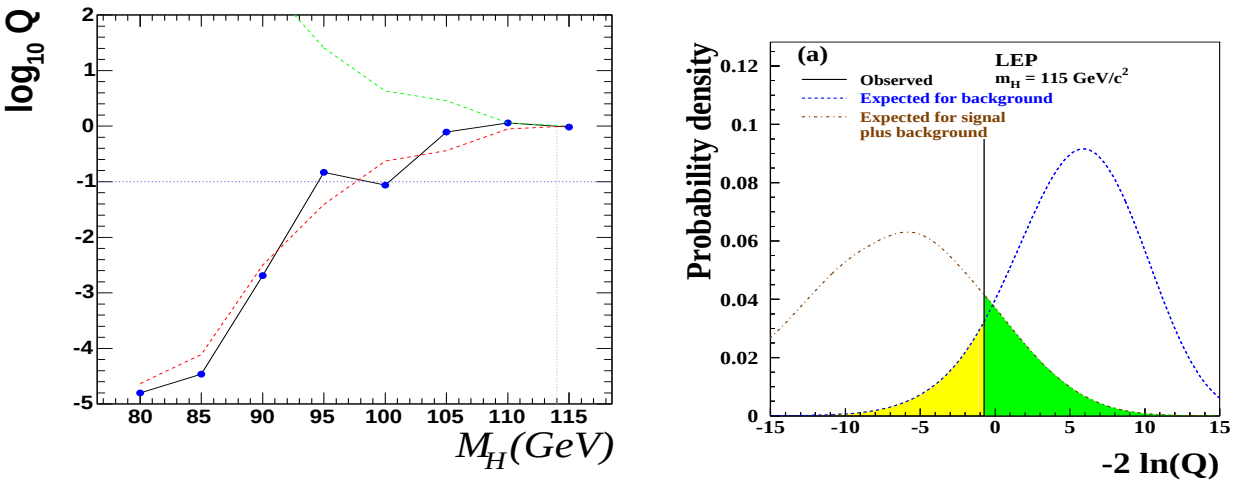


Figure 5.9 QUAERO's output ($\log_{10} Q$) as a function of assumed Standard Model Higgs mass m_H (left). Distributions of $-2 \ln Q$ from the combined LEP Higgs search.

5.5.5 Standard Model Higgs

A search for the Standard Model Higgs boson has also been performed using QUAERO. The signal was generated with PYTHIA including both Higgsstrahlung and weak boson fusion. The interference between these diagrams for the $H\nu\bar{\nu}$ channel was not taken into account.

A scan performed in the mass m_H of the Standard Model Higgs boson results in the output shown in Figure 5.9. QUAERO is able to exclude a Higgs boson with mass $m_H \lesssim 95$ GeV, compared with the previous ALEPH limit of $m_H > 111.5$ GeV. QUAERO's significantly less sensitive result appears to be primarily due to two optimizations employed by the previous ALEPH analysis: (1) a loosening of b-tagging requirements and the inclusion of b-tagging information in the event in a discriminant, and (2) the use of a constrained kinematic fit to the HZ hypothesis, optimized for each m_H. The categorization of events into exclusive final states is sufficiently integral to the existing QUAERO algorithm that allowing the additional flexibility of (1) would require substantial restructuring. The list of variables that QUAERO uses is hardwired and does not adapt to the characteristics of the provided hypothesis, so QUAERO does not have access to the assumed value of m_H that would allow (2). Such deficiencies provide useful direction for future refinements of the QUAERO algorithm.

5.6 Summary

The ALEPH data from LEP2 have been incorporated within QUAERO, an automated analysis algorithm; the resulting prototype is referred to as QUAERO@ALEPH. This short chapter has detailed the data that can be analyzed within QUAERO@ALEPH; the estimation of Standard Model processes used to form the reference model to which hypotheses are to be tested; the systematic uncertainties on the modeling of the ALEPH detector; and the construction of a fast detector simulation, TURBOSIM@ALEPH, which makes use of a large lookup table using events that have been run through ALEPHSIM.

The use of QUAERO@ALEPH has been illustrated with searches for minimal supergravity signatures, excited electrons, doubly charged Higgs bosons, singly charged Higgs bosons, and the Standard Model Higgs boson. QUAERO's results have been found to be in agreement with the previous ALEPH results (when available).

Despite the restriction to high-level objects identified with a general-purpose particle identification procedure, it has been demonstraited that the QUAERO algorithm can be quite sensitive. The favorable performance of QUAERO@ALEPH motivates additional consideration of this and similar algorithms for the LHC.

While no compelling evidence for new physics was found in the searches presented above, QUAERO@ALEPH is now available to ALEPH members and their collaborators as a tool to search archived ALEPH data for other signatures. This tool has three potential uses. First, in the event of an observation of new physics at another detector, QUAERO@ALEPH can be used for a quick confirmation with ALEPH data. Secondly, any authorized user can use QUAERO@ALEPH to perform a quick analysis to assess the sensitivity before beginning a dedicated analysis in the conventional sense. Lastly, in roughly one year, any authorized user will be able to use QUAERO@ALEPH, possibly in conjunction with data from other detectors, to search for new physics and publish their results in accord with the ALEPH statement on archival data.

Chapter 6

Observations and Conclusions from LEP

The LEP2 physics program was a terrific success with an incredible breadth and depth of results. The author made only minimal contributions to the LEP physics program while it was in operation, but was privileged to be a part of the last years of data taking.

6.1 Influence of LEP on Preparation for the LHC

The observed excess of Standard Model Higgs candidates by ALEPH and its consistency with indirect electroweak measurements is one of the most exciting results from LEP. The majority of the work in the next part of this dissertation is devoted to improving ATLAS's sensitivity to a low mass Higgs boson, which are the most challenging to discover at the LHC. If a low mass Higgs boson exists, then it is quite possible that a claim of discovery will require the combination of several channels, the use of multivariate analysis methods, and/or the use of discriminating variables. This is the motivation for the strong emphasis on advanced analysis tools found in this dissertation.

The following practical observations also influenced the work in the following chapter.

- Once data taking begins, it is much more difficult to make fundamental design choices in how one analyzes the data. The limitations in the initial analysis design are often addressed with *ad hoc* solutions, which complicate the final interpretation of results.
- Combining results from different analyses or different experiments is very powerful; however, it is very difficult to do properly without advanced planning.

- Communication with theorists and phenomenologists is key to understand the most effective way to utilize the data. This collaboration must be actively pursued for it to be fruitful.

6.2 Potential for Vista and Quaero at the LHC

The application of VISTA and QUAERO to the ALEPH data was relatively straightforward due to many factors. Being an e^+e^- collider, LEP provided an extremely clean experimental environment, a relatively small set of Standard Model processes to consider, and a relatively short list of theoretical challenges. Also, it should be clear that the quality of the comparison between data and Standard Model was due to the fact that the ALEPH detector was very well understood by the time ARCH was developed. In contrast, the LHC environment is challenging, the number of Standard Model processes is large, and there are large theoretical uncertainties.

The most challenging aspects to the incorporation of ALEPH data into the QUAERO framework was the tuning of ARCH such that it simultaneously provided robust, general purpose particle identification *and* satisfied the demands of TURBOSIM. In particular, TURBOSIM requires a delicate balance between reducing the number of clustered objects in the final state and retaining sufficiently "local" `truth->reconstruction` relationships. Relaxing the tight coupling between the event classification and the fast detector reconstruction will be crucial to the success of these methods at the LHC.

Providing a robust, general purpose particle identification was challenging even without the requirements of TURBOSIM. The strategy of the ALEPH analysis framework was to provide high-level reconstruction objects, such as Energy Flow objects, and to allow the user to make the final particle identification and jet clustering decisions. The rational for this design is that the optimal particle identification procedure is analysis-dependent. The analysis model being developed by ATLAS is still a prototype, but the contents of the Analysis Object Data are based on identified particles (see Appendix F). If this model continues, it will help standardize ATLAS particle identification and facilitate the interface to VISTA and QUAERO. It will also aid in the evaluation of systematic errors in particle identification.

The huge variety of Standard Model processes that will be encountered at the LHC will be a substantial challenge to an inclusive analysis framework like QUAERO. One way to overcome this challenge is to limit the scope of the requests to a subset of final states. Even then, the huge rate of the LHC requires consideration of many backgrounds. Given the amount of work that was necessary to provide the relatively few special purpose Monte Carlo generators used by ALEPH, it is unlikely that a similar approach would work for the LHC. However, the recent development of general purpose Monte Carlo generators may make this feasible. Another difficulty that has plagued VBF analyses is the consistent generation of events without double-counting. This difficulty is related to the matrix element-parton shower matching that has largely been solved by recent programs such as SHERPA. A considerable amount of work will be necessary to understand how to efficiently produce more inclusive Standard Model Monte Carlo data sets.

Lastly, the current implementation of QUAERO, while powerful, is somewhat inflexible. The multivariate analysis procedure QUAERO employs has been tuned to be robust, fast, and powerful, but a user may wish to limit or modify the analysis procedure. Similarly, the result of QUAERO is the decimal log likelihood ratio, which (as was discussed in Section 5.4) does not allow for a frequentist confidence level calculation. Modifications to the QUAERO algorithm that make it more flexible would also increase the chance that it is more widely adopted.

Despite these difficulties, it seems quite possible that ATLAS and CMS could benefit from either the implementation of VISTA and QUAERO or the development of a customized interface that uses some of the ideas of automated analysis. Until the experiments are quite mature, it is difficult to foresee their data being publicly interfaced to QUAERO; though QUAERO does offer a framework with which the experiments might produce combined results. The inclusive view of the data that is available with VISTA is an efficient way to discover deficiencies in Standard Model Monte Carlo description, exceptional cases for general purpose particle identification, and possibly hints of new physics. Furthermore, QUAERO's automated search procedure is very fast and would be useful in navigating complex models if we see hints of new physics.

Part II

Preparing for New Physics at the LHC

Chapter 7

The ATLAS Detector at the LHC

7.1 The Large Hadron Collider at CERN

The Large Hadron Collider (LHC) at CERN is presently under construction in the same tunnel used for LEP, but with an entirely new superconducting magnet system consisting of over 1000 dipoles and 350 quadrapoles. Two additional caverns have been made for the two large, multipurpose detectors ATLAS and CMS. The LHC has been designed to collide 7 TeV proton beams configured in bunches of 10^{11} protons separated by 25 ns with a nominal luminosity of $10^{34}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. In the first few years of operation, the LHC is expected to run in the low-luminosity configuration of $10^{33}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, which will provide approximately $10\ \mathrm{fb}^{-1}$ of data per calender year. The focus of the studies presented in the following chapters is on the first few years of low-luminosity running [84].

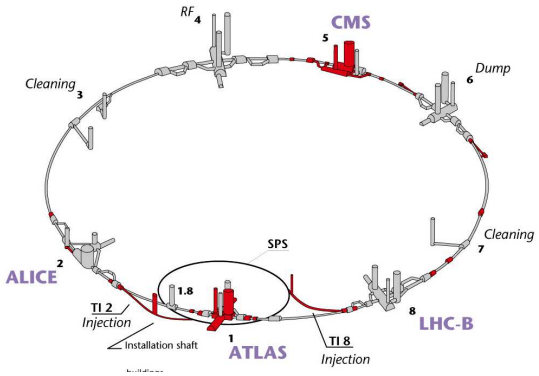


Figure 7.1 The LEP tunnel after modifications for the LHC experiments.

7.1.1 Pile-Up

The total inelastic pp cross-section at $\sqrt{s} = 14\ \mathrm{TeV}$ is about 80 mb, which translates to an average of 2.3 interactions per bunch crossing at low-luminosity [85]. The vast majority of these events are due to "minimum bias interactions" with small transverse momentum that arise from long-range $p - p$ interactions. These minimum bias events can be viewed as a bath of energy superposed on the hard-scattering of physics interest: a phenomenon known as *pile-up*.

The presence of pile-up has had a major impact on the design of the readout electronics for the ATLAS detector. In effect, the pile-up is treated as a type of noise. The minimum bias interactions occasionally do produce events with a more pronounced jet-like structure. These low-p_T jets threaten the efficacy of the central jet veto in vector boson fusion (VBF) Higgs searches. It is expected that the presence of pile-up will not be prohibitive for VBF Higgs searches at low luminosity, but the rate of minimum bias interactions faking central jets does preclude the searches at high-luminosity.

7.1.2 Underlying Event

In addition to minimum bias interactions, there is a different kind of physics background omnipresent at the LHC: the underlying event. Unlike minimum bias interactions, the underlying event arises from the same $p - p$ interaction as the hard scattering of interest.

The underlying event has a hard component that comes from multiple parton interactions. These multiple parton interactions are manifest in the violation of Koba-Nielsen-Olesen scaling [86] that was observed at UA5 and UA1 [87, 88]. These violations grow with increasing $\sqrt{s}$. In addition, multiple parton interactions have been observed at the Tevatron.

A detailed study of the underlying event at the Tevatron can be found in Ref. [89]. Models of the underlying event are available in PYTHIA and in the HERWIG extension JIMMY [90]. The current ATLAS strategy is to tune PYTHIA's phenomenological model to the Tevatron data and extrapolate to the LHC [91]. Because the model in JIMMY is different than PYTHIA's model, JIMMY is being tuned to match PYTHIA's extrapolation.

What can you leave behind
when you're flyin' lightning fast
and all alone?

Only a trace, my friend,
spirit of motion born
and direction grown.

– Townes Van Zandt, *High, Low, and In Between*

7.2 The ATLAS Detector

The ATLAS detector[1] is currently under construction, with parts of the calorimeter and magnet system already in the cavern. The ATLAS detector, shown in Figure 7.2, is incredibly complex and described in exquisite detail in the Techincal Design Reports [92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104]. For completeness, a brief review of the ATLAS detector is described below.

7.2.1 The Magnet System

The ATLAS superconducting magnet system is shown in Figure 7.3. The ATLAS magnet system's unusual configuration and large size make it one of the most challenging engineering feats of the ATLAS detector [97].

The central solenoid provides a 2 T field to the ID in the region ($|\eta| < 1.5$) and is powered by a 8 kA power supply [100]. The central solenoid is housed in the barrel cryostat, between the Inner Detector and the Electromagnetic Calorimeter; thus, the solenoid contributes to up-stream materal and degrades the EMC performance. Special effort has been put into minimizing the up-stream material; in particular, the central solenoid and EMC share one common vacuum vessel.

The barrel torid and two end-cap torids each consist of eight air-core superconducting coils powered by a 21 kA power supply. The large barrel torid is 20 m long and provides bending in the region $|\eta| < 1.3$. The peak magenetic field for the barrel toroid is 3.9 T, and the bending power, which is given by the field integral $\int \mathbf{B}\,\mathrm{d}\mathbf{l}$, ranges from 2 to 6 T-m [98]. The end-cap torid contributes 4 to 8 T-m of bending power in the region $1.6 < |\eta| < 2.7$ [99]. In the overlap region, $1.3 < |\eta| < 1.6$, the bending power, though lower in magnitude, is ensured by an overlap between the barrel and end-cap fields.

[1] ATLAS is A Terrifically Long Acronym Standing for **A** **T**oroidal **L**HC **A**pparatu**S**.

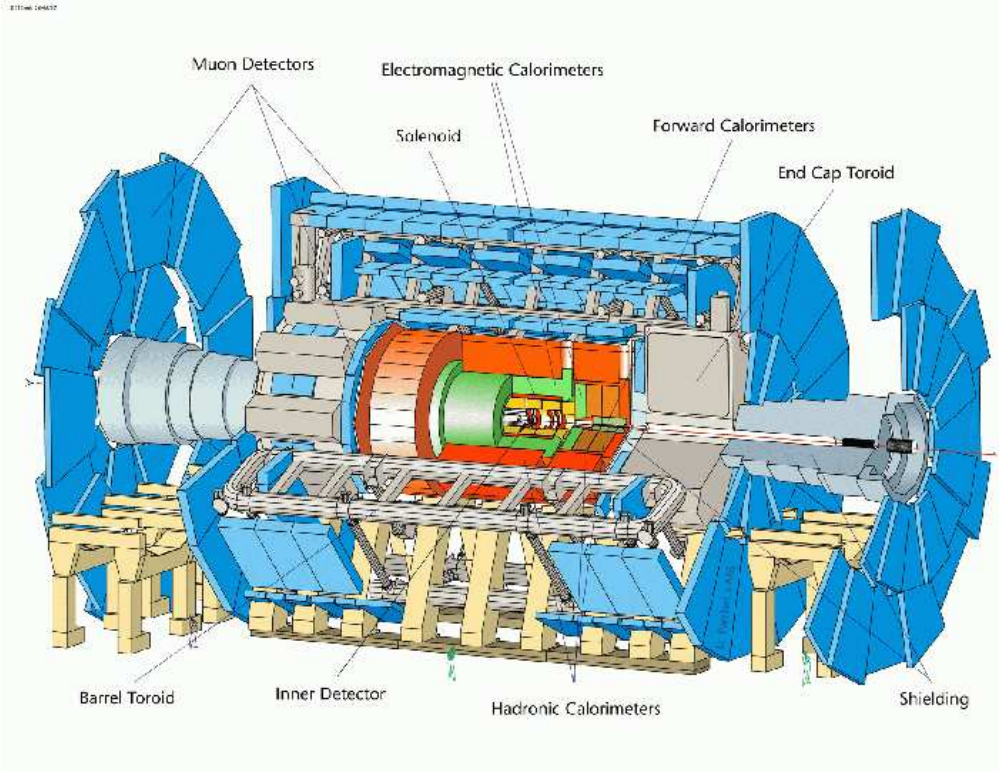


Figure 7.2 An illustration of the ATLAS detector.

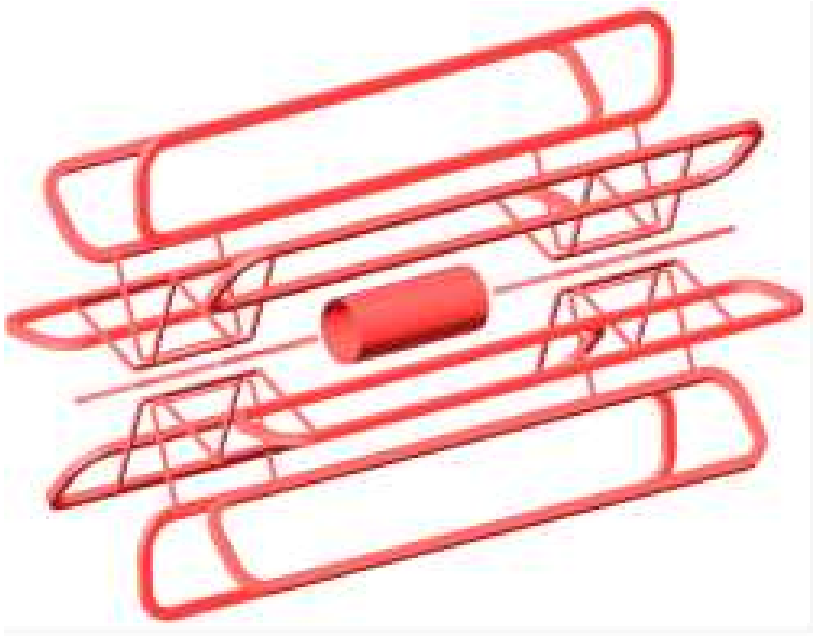


Figure 7.3 An illustration of the ATLAS magnet system.

7.2.2 The Inner Detector

The Inner Detector (ID), shown in Figure 7.4, covers the region $|\eta| < 2.5$ and is immersed in the 2 T magnetic field of the central solenoid [95, 96]. The ATLAS ID is composed of three different sub-detectors:

- **The Pixel Detector** (PD) consists of three barrel layers located at ∼4, 10, and 13 cm from the beam axis and five disks on each side (between radii of 11 and 20 cm). The PD provides a very high granularity set of measurements with about 140 million detector elements, each 50 $\mu\mathrm{m}$ in the $R-\phi$ direction and 300 $\mu\mathrm{m}$ in the z direction. Due to the hostile environment, the chips must be radiation hardened to withstand over 300 kGy of ionizing radiation and over $5 \cdot 10^{14}$ neutrons per cm^2 over ten years of operation. The innermost pixel layer (or B-layer) has been designed to be replaceable in order to maintain the highest possible performance throughout the experiment's lifetime [102].

- **The SemiConductor Tracker** (SCT) is designed to provide eight precision measurements per track. The barrel SCT provide precision points in the $R-\phi$ and z coordinates with eight layers of silicon microstrip detectors (arranged in four pairs, each with small angle stereo to obtain the z measurement). The two end-cap modules are arranged in nine wheels covering up to $|\eta| < 2.5$. In total, it consists of 61 m^2 of silicon detectors with 6.2 million readout channels. The spatial resolution is 16 $\mu\mathrm{m}$ in the $R-\phi$ direction and 580 $\mu\mathrm{m}$ in z.

- **The Transition Radiation Tracker** (TRT) is based on straw detectors, which can operate at very high rates. The TRT provides electron identification capability by using xenon gas to detect transition-radiation photons created in a radiator between the straws. Each channel provides a drift-time measurement with a spatial resolution of 170 $\mu\mathrm{m}$. With a total of 50,000 straws in the barrel region and 320,000 radial straws (arranged in 18 wheels) in the end-caps, the TRT typically provides 36 measurements per track.

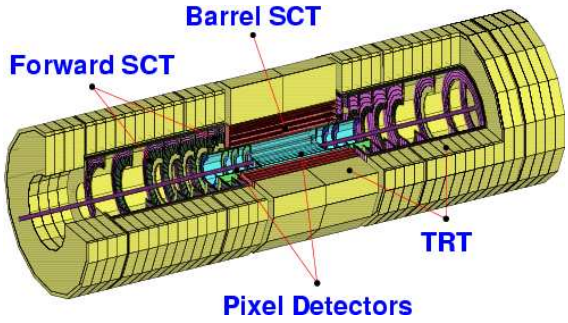


Figure 7.4 An illustration of the ATLAS inner detector.

7.2.3 Calorimetry

The ATLAS calorimeter, shown in Figure 7.5, consists of an electromagnetic calorimeter (EMC) covering the region $|\eta| < 3.2$, a hadronic barrel calorimeter covering the region $|\eta| < 1.7$, hadronic end-cap calorimeters (HEC) covering the region $1.5 < |\eta| < 3.2$, and forward calorimeters covering the region $3.1 < |\eta| < 5$. A concise visual comparison of the different calorimeter sub-systems is given by the two topological clusters shown in Figure 7.7.

The electromagnetic calorimeter (EMC) is a lead-liquid Argon (LAr) sampling calorimeter consisting of a barrel and two end-caps [92, 93]. The barrel consists of two half-barrels, separated by a 6 mm gap. The EMC has has an unusual accordion shape, shown in Figure 7.6, with Kapton electrodes and lead absorber plates. The total thickness of the EMC is ∼25 radiation lengths (X_0). The region $|\eta| < 2.5$ is segmented into three longitudinal segments. The innermost strip section has constant thickness of $\sim 6X_0$ as a function of η and is equipped with narrow strips with a pitch of ~ 4 mm. The middle section is segmented into square towers of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$. The back section has a granularity of 0.05 in η and a thickness varying between 2 and 12 X_0. In total there are nearly 190,000 readout channels.

Because there is about 2.3 X_0 of material before the front face of the calorimeter, a presampler is used to correct for up-stream energy loss. The presampler consists of a 1.1 cm and 0.5 cm active LAr layer in the barrel and end-cap, respectively. In addition to the presampler, a scintillator slab is inserted in the crack region between the barrel and endcap cryostats ($1.0 < |\eta| < 1.6$). In total there are about 10,000 readout channels.

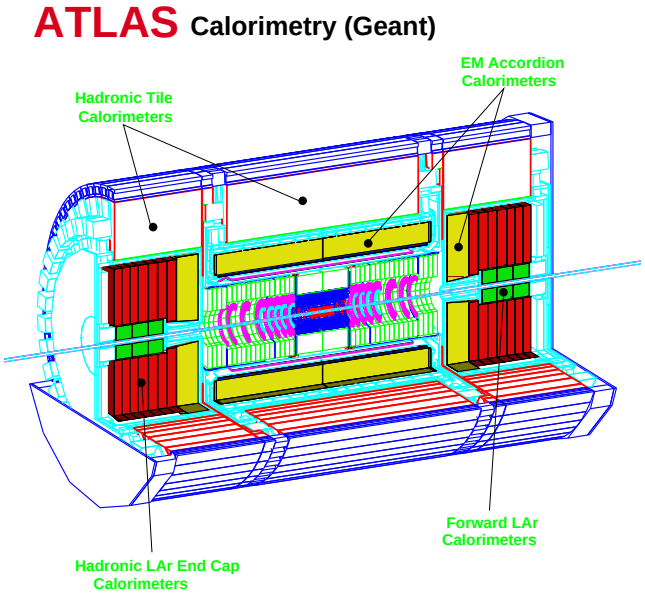


Figure 7.5 An illustration of the ATLAS calorimeter.

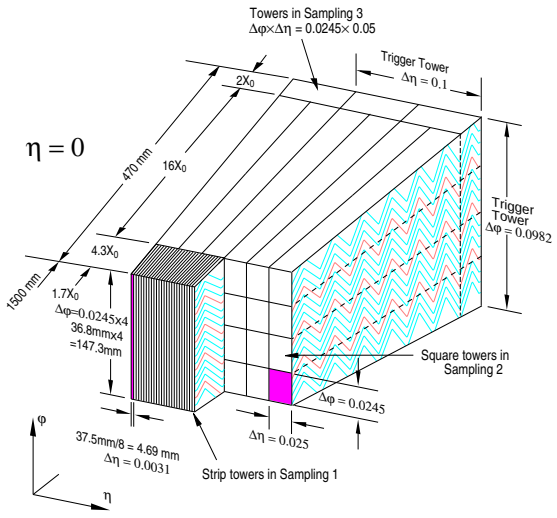


Figure 7.6 An illustration of the ATLAS LAr electromagnetic calorimeter's accordian structure.

The hadronic barrel calorimeter (Tilecal) is composed of a central barrel and two extended barrels. It is based on a novel sampling technique with 3 mm thick plastic scintillator tiles sandwiched between 14 mm thick iron absorption plates. The basic granularity of the Tilecal is $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the first two samplings and 0.2×0.1 in the third. The gap between the barrel and extended barrel is partially instrumented with the Intermediate Tile Calorimeter (ITC). These calorimeters also act as the main flux return for the central solenoid [92, 94].

The hadronic endcap calorimeter (HEC) is a copper-LAr detector with parallel-plate geometry and extends to $\eta = 3.2$ [92, 93]. It is composed of two wheels and has a basic granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ for $1.5 < |\eta| < 2.5$ and 0.2×0.2 for $2.5 < |\eta| < 3.2$.

The forward calorimeter (FCAL) also uses a LAr, but with a high density design due to the high level of radiation it experiences. The FCAL consists of three sections, the first is made of copper and the second two are made of tungsten. Each section consists of concentric rods (cathods) and tubes (annodes) embedded in a matrix. The LAr gap between the rod and tubes form the active medium, which can be as small as 250 μm. The geometry of the FCAL is more natural in an $x - y$ coordinate system; however, the granularity roughly corresponds to $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$. In total there are nearly 4000 readout channels [92, 93].

7.2.4 The Muon System

The ATLAS Muon system, shown in Figure 7.8, provides both a precision muon spectrometer and a stand-alone trigger subsystem [101]. The precision measurements are provided by Monitored Drift Tubes (MDTs) and, in the region $2 < |\eta| < 2.7$, Cathode Strip Chambers (CSCs). The precision measurement is made in a direction parallel to the bending direction: the z coordinate in the barrel and the R coordinate in the end-cap.

The trigger system covers the range $|\eta| < 2.4$ and consist of both Resistive Plate Chambers (RPCs) and Thin Gap Chambers (TGCs). The trigger chambers must have a time resolution better than the LHC bunch spacing of 25 ns, provide triggering with well-defined p_T thresholds, and provide a measurement of the coordinate perpendicular to the precision measurements provided by the MDT or CSC.

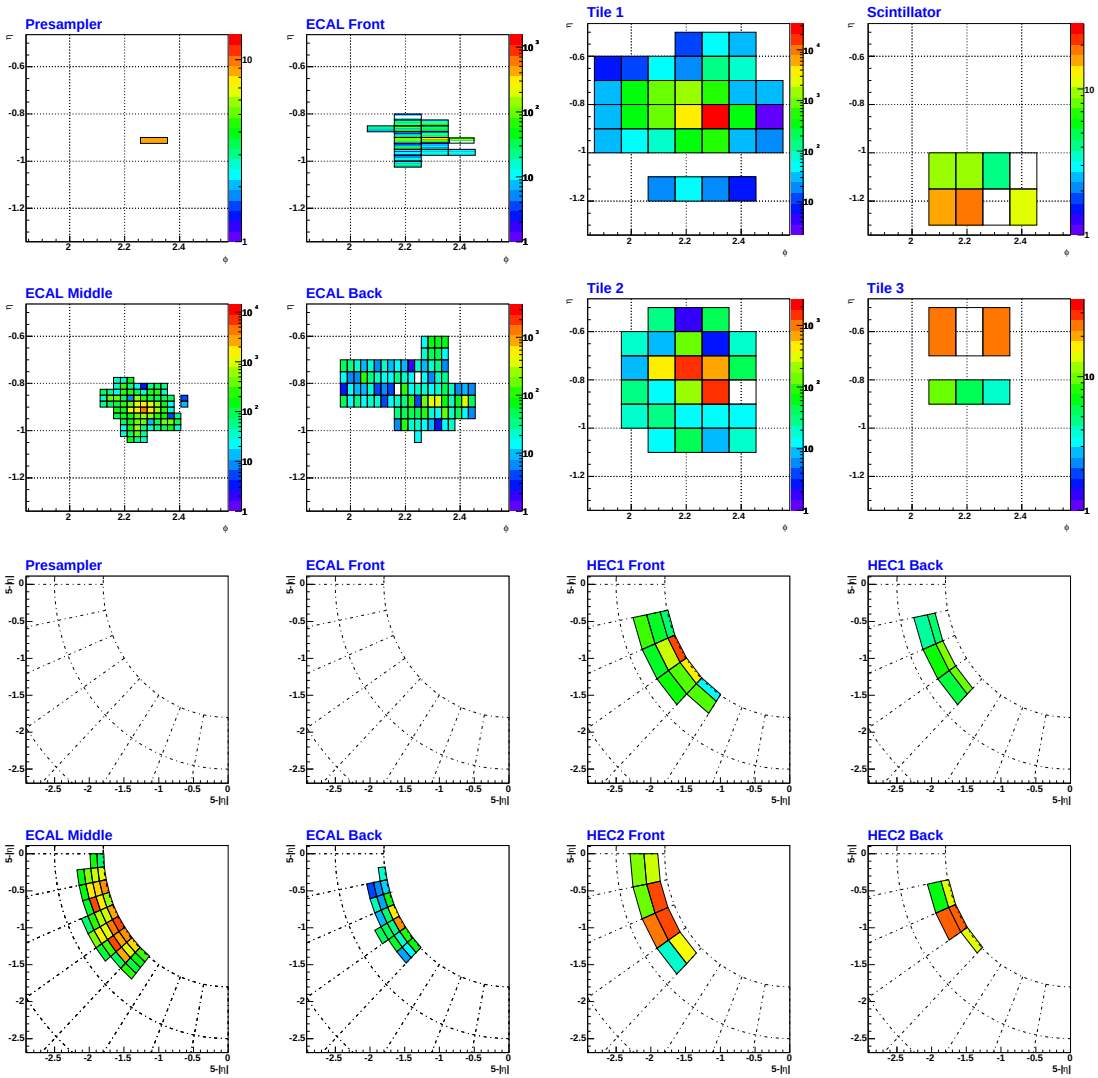


Figure 7.7 A topological cluster in the barrel (top) and end-cap (bottom).

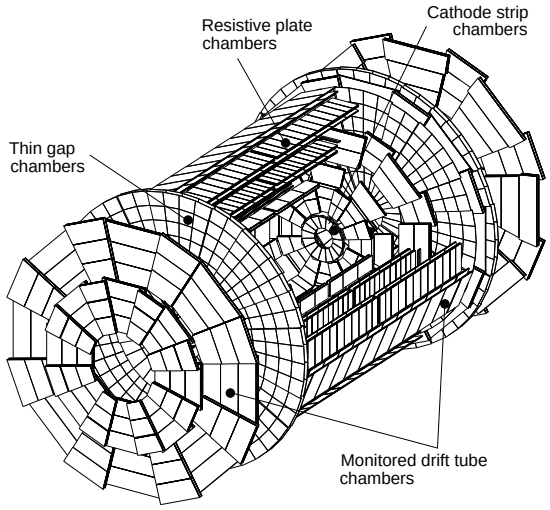


Figure 7.8 An illustration of the ATLAS muon spectrometer.

7.2.5 Trigger and Data Acquisition

The ATLAS trigger and data-acquisition (DAQ) system is based on three levels of online event selection [103, 104, 105]. Starting from an initial bunch-crossing rate of 40 MHz[2], the rate of selected events must be reduced to ~ 100 Hz for permanent storage. In addition to providing a rejection factor of 10^7 against minimum-bias events, interesting hard-scatterings must be retained with a high efficiency.

The level-1 (LVL1) trigger makes an initial selection based on high-p_T muons in the RPCs and TGCs as well as low-granularity calorimeter signatures. These calorimeter signatures include isolated, high-p_T electrons and photons, jets, and τ-jets as well as $\not{p}_T$ and $\sum |E_T|$ (where the sum is over trigger towers). The global LVL1 trigger consists of combinations of these objects in coincidence or veto. Because the pulse shape of the calorimeter signals extends over many bunch crossings, the LVL1 decision is performed with custom integrated circuits, processing events stored in a pipeline with $\sim 2\mu$s latency.

[2] At high-luminosity, the interaction rate is $\sim 10^9$ Hz

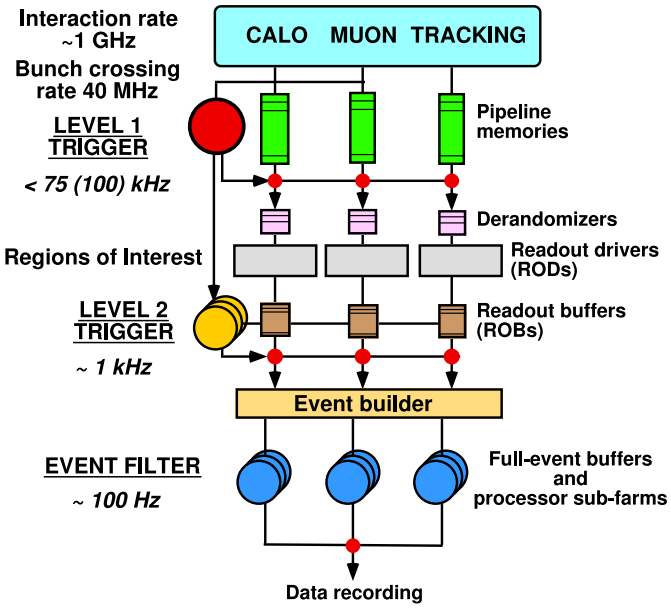


Figure 7.9 An schematic of the the ATLAS Trigger and data acquisition system.

Events selected by LVL1 are read out from the front-end electronics into readout drivers (RODs) and then into readout buffers (ROBs) (See Figure 7.9). If the event is selected by the level-2 (LVL2) trigger (described in the next paragraph), the entire event is transfered by the DAQ to the Event Filter (EF), which makes the third level of event selection.

In principle, the LVL2 trigger has access to all of the event data with full precision and granularity; however, the decision is typically based only on event data in selected *regions of interest* (RoI) provided by LVL1. The LVL2 trigger will reduce LVL1 rate of 75 KHz to ~ 1 kHz with a latency in the range 1-10 ms.

The last stage of online event selection is performed in the Event Filter (EF). The Event Filter utilizes selection algorithms similar to those used in the offline environment. The output rate from LVL2 should be reduced to ~ 100 Hz, depending on the size of the dedicated high-level trigger (HLT) computing cluster available at startup.

7.2.6 Fast and Full Simulation of the ATLAS Detector

Clearly, the success of the ATLAS physics program depends to a large degree on the ability to simulate the ATLAS detector. Moreover, in the context of hypothesis testing, both the null and alternate hypothesis are a convolution of a fundamental theory and a complicated experimental apparatus. The ATLAS collaboration utilizes both a fast and a full simulation of the ATLAS detector for different purposes.

The full simulation of the ATLAS detector is performed with GEANT. The original studies used to produce the numerous technical design reports used GEANT3 almost exclusively. In 2004, ATLAS migrated to GEANT4 after an extensive validation period. The ATLAS detector description is incredibly detailed and undergoes continuous enhancement and bug fixing.

Due to the high energies of incident particles, the presence of pile-up, and the high granularity of the ATLAS detector, full simulation is an incredibly computationally intensive task. The VBF $H \rightarrow \tau\tau$ events studied in Chapter 10 require roughly 15 min per event for simulation on a 2 GHz Pentium II processor. In total, the computing required to simulate the events in this thesis corresponds to more than 100 CPU-years.

While the study of tracking performance, energy resolution, jet clustering, particle identification efficiencies, etc. are solely the domain of the full simulation, a fast simulation is also required. In particular, for searches that have very large reducible backgrounds (the most common being $t\bar{t}$ production) it is not feasible to prototype an analysis with the full simulation. The strategy taken by ATLFAST is to provide a simplified, lower granularity detector description (see Figure 7.10) and parametrize the response of the full simulation as a function of Monte Carlo truth quantities [106]. This approach works well for isolated photons, electrons, and muons. Particular effort has been put into the identification efficiency of forward jets, b-jets, and τ-jets [107]. It has been shown that the efficiencies and rejections obtained from the parametrized particle identification algorithms reproduce full simulation on average, but do not necessarily reproduce the correct correlation to other observables. Lastly, the $\not{p}_T$ calculation in ATLFAST is based on the energy resolution of reconstructed objects, which differs from the cell-based approach used by ATLAS (see Chapter 9).

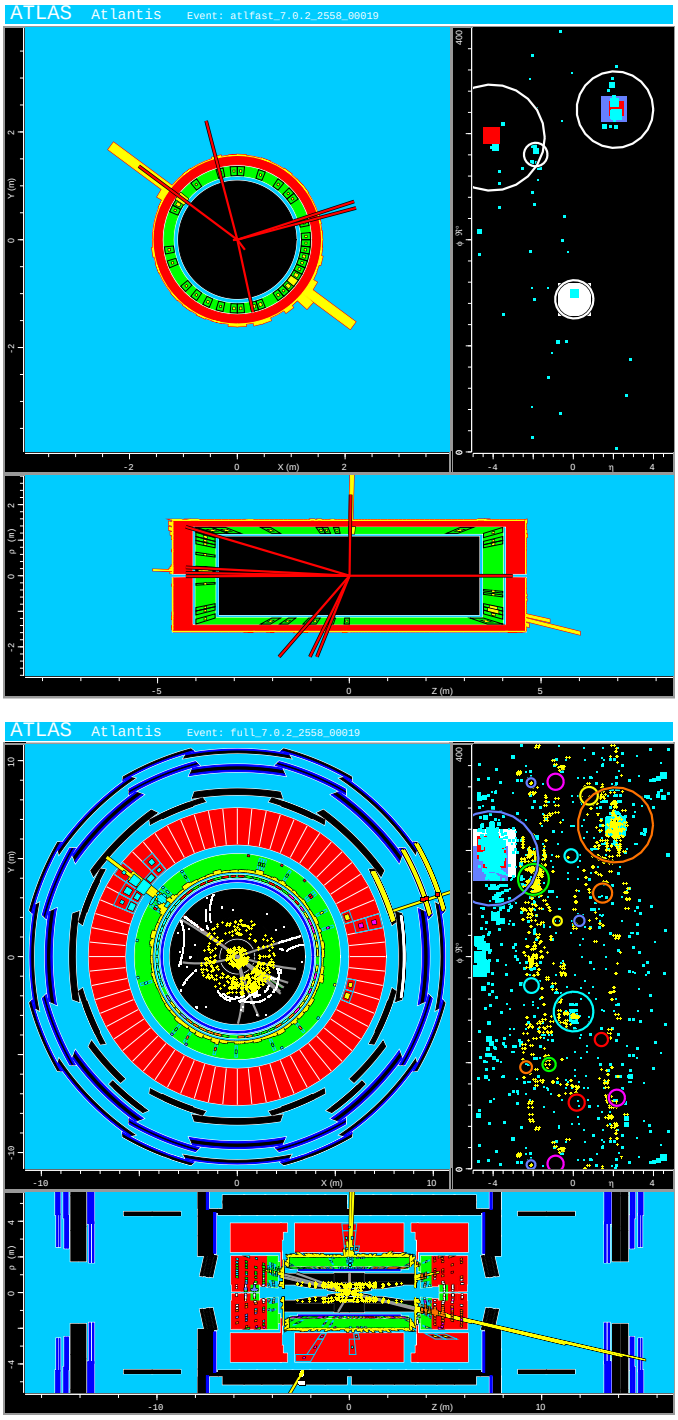


Figure 7.10 An ATLANTIS display of a VBF $H \to \tau\tau$ event simulated with ATLFAST (top) and GEANT3 (bottom).

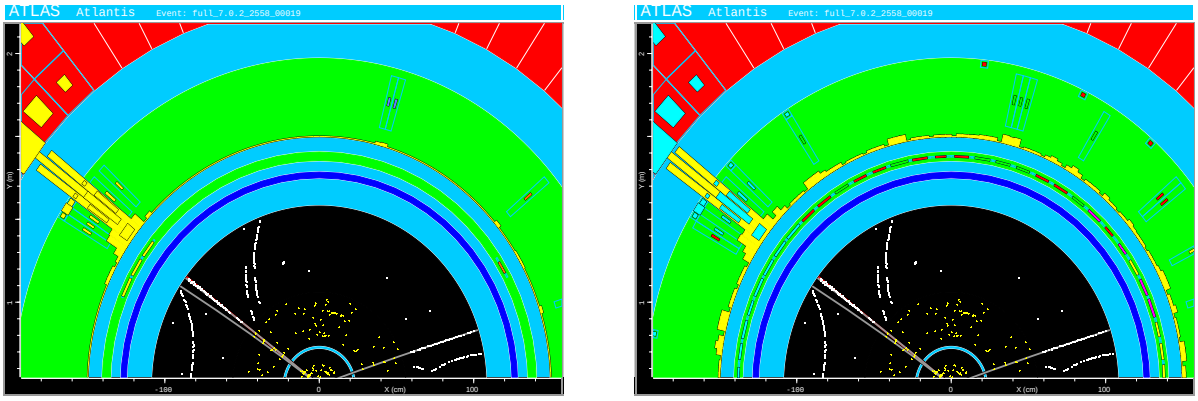


Figure 7.11 An ATLANTIS display of a VBF $H \rightarrow \tau\tau$ event simulated with GEANT3 without noise (top) and with noise (bottom). Neither event includes pile-up effects.

In the chapters that follow, events simulated with GEANT3 include the effect of electronic noise by smearing the energy depositions, or *hits*, according to the RMS electronic noise for that cell. This treatment is a good approximation, but it does not model non-Gaussian tails. Figure 7.11 shows the same VBF $H \rightarrow \tau\tau$ event, simulated in GEANT3, reconstructed with and without electronic noise.

Events simulated with GEANT4 take into account the response of the detector elements, and the ATLAS electronics are simulated to produce a *digitized* output similar to the raw data produced by ATLAS. During the digitization procedure, the electronic noise is applied to each of the calorimeter pulse samples and then processed by the Optimal Filtering Coefficients, thus providing a more realistic description of electronic noise.

Chapter 8

Monte Carlo Development for Vector Boson Fusion

Vector Boson Fusion (VBF) is the second leading production mechanism for the Standard Model Higgs boson at the LHC [108]. The tree-level process is illustrated in Figure 8.1, which shows two final state quark lines in addition to the Higgs boson. These final state quarks give rise to two very energetic and forward jets. It is the presence of these jets and the pattern of additional QCD radiation that provide a powerful handle for background suppression [109, 110].

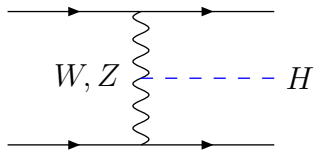


Figure 8.1 Tree-level Feynman diagram for vector boson fusion Higgs production.

An early analysis performed at the parton level with the decays $H \to W^{(*)}W^{(*)}$ and $H \to \tau\tau$ indicated that this process could be a powerful discovery mode in the mass range $115 < M_H < 200\,\mathrm{GeV}$ [111, 112, 113]. Those calculations were performed with a number of special purpose Monte Carlo programs. In order to develop these analyses with a detailed simulation of the ATLAS detector, these programs were interfaced to showering and hadronization generators (such as PYTHIA and HERWIG) as external user processes. The resulting Collection of User Processes is called MadCUP [114].

8.1 The MadCUP Event Generators

The interface to showering and hadronization generators was realized through the HEPEUP and HEPRUP common block structure known as the *les Houches interface* [115]. The major modification necessary to provide this interface was to provide the *color flow* of the partons. The original programs calculated the color-summed amplitude squared, $|\mathcal{M}|^2$, which takes into account interference effects between the (unobservable) color flows under $SU(3)$ with 8 gluons. The MadCUP generators assigned the i^{th} color flow arrangement with probability $|\mathcal{M}_i|^2/|\mathcal{M}|^2$ in the large N limit of $SU(N)$.

A flow chart of the MadCUP event generators is shown in Figure 8.2. The flow chart shows the separation of the color-summed squared matrix element calculation, phase space integration, and color flow selection.

8.2 Color Coherence

QCD predicts that the interference between soft gluons results in a suppression of radiation at large angles. This quantum behavior can be realized in a SHG's final state (time-like) evolution via *angular ordering*. *i.e.* the emission angle in subsequent branchings decreases and the radiation lies within a cone defined by color-flow lines. This has been studied extensively at LEP, and angular ordering works well.

At hadron-hadron colliders there is color in the initial state, thus there is color flow from initial to final state. In fact, the initial and final state radiation cannot be separated in a way that is gauge invariant. QCD gives a prescription for the radiation of the first gluon from a color line. Different color structures add incoherently to $\mathcal{O}(1/N^2)$, thus the intersection of the initial-state cone and the final-state cone is unsuppressed and radiation is contained to their union. Figure 8.3 illustrates the cones defined by angular ordering. Various studies at the Tevatron established that the angular ordering in HERWIG and the time-like evolution, with constraints on the first gluon emission, found in recent versions of PYTHIA both reproduce the most prominent effects of color coherence [116, 117].

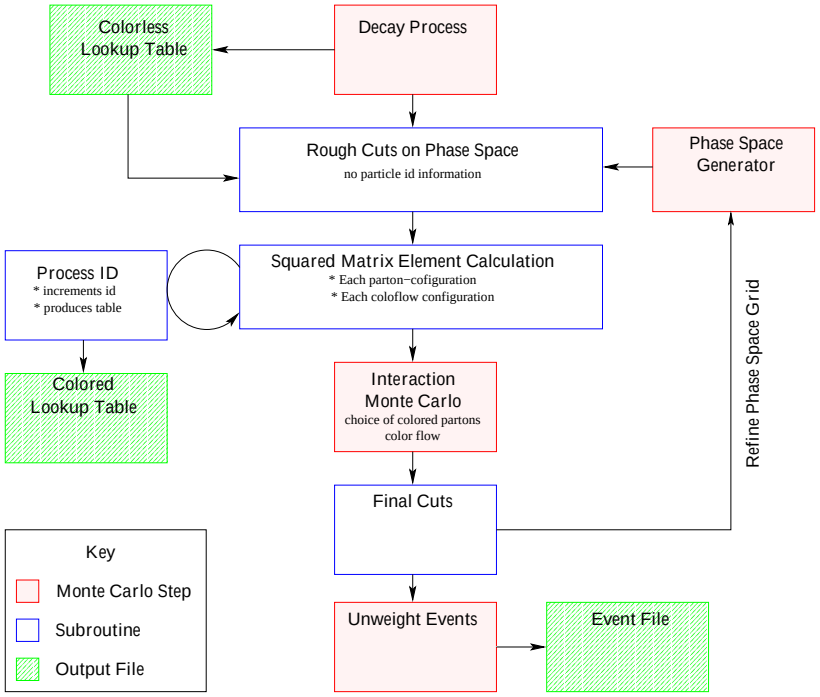


Figure 8.2 A flow diagram for the MadCUP generators.

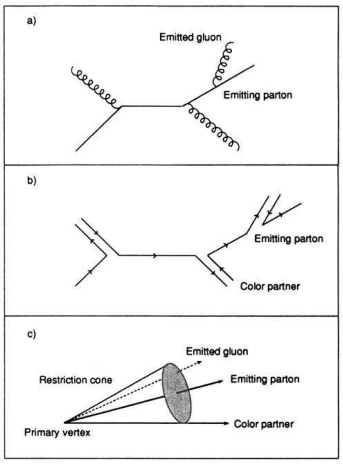


FIG. 2. Angular ordering. (a) Feynman diagram of a final state shower. (b) Color flow in the shower. (c) Partons exiting from the primary vertex. The emitted radiation is constrained around the emitting parton, within a cone defined by the color-connected partner. A similar cone (not shown) confines radiation emitted by the partner.

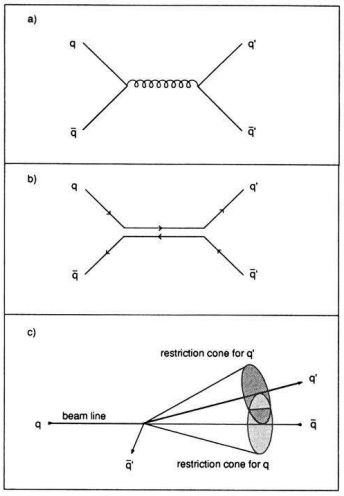


FIG. 3. Phase space constraints for the first gluon emission in a $q\bar{q}$ annihilation. (a) Feynman diagram of the hard scattering. (b) Color lines flow from the initial to the final state, q and q' are color partners. (c) Restriction cones for q and q'. Similar cones can be drawn for $\bar{q}$ and $\bar{q}'$.

Figure 8.3 Illustration of color coherence effects taken from CDF, Phys. Rev. **D**50.

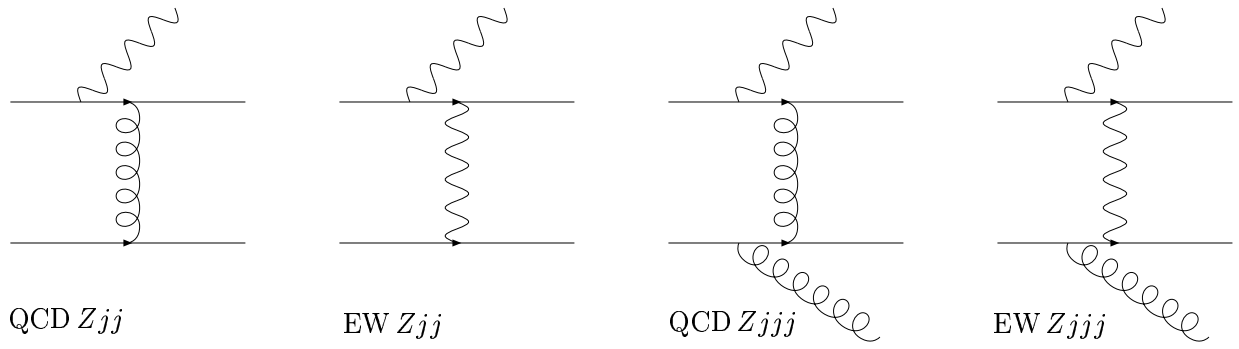


Figure 8.4 Electroweak and QCD Zjj and $Zjjj$ tree-level Feynman diagrams.

8.3 Validation of Color Coherence in External User Processes

A key aspect of the VBF Higgs searches is that the electroweak (EW) nature of the signal (see Figure 8.1) leads to a suppression of QCD radiation between the two tagging jets. In the context of color coherence, there is only one color flow for the signal (at tree level) and the cones of QCD radiation do not include the central region. The irreducible background for the VBF $H \rightarrow \tau\tau$ searches comes from the production of Z in association with two hard, forward jets, Zjj. The production of Zjj can come either from t-channel exchange of either an electroweak boson or a colored parton. These diagrams are shown in Figure 8.4 and are known as EW Zjj and QCD Zjj, respectively. QCD Zjj accounts for most of the cross section due to the fact $\alpha_S >> \alpha$; however, QCD Zjj has color flow configurations in which the cone of radiation does include the central region. The presence of extra QCD radiation between the two tagging jets in the background, but not in the signal, is the motivation for the Central Jet Veto (CJV) found in most VBF analyses.

Studies of the CJV efficiencies for signal and background were carried out in Ref. [109] by constructing a variable sensitive to coherence effects

$$\eta^* = \eta_3 - \frac{\eta_1 + \eta_2}{2}, \tag{8.1}$$

where η_i is the pseudorapidity of the i^{th} jet ordered in decreasing p_T. Using EW and QCD $Zjjj$ Monte Carlo and applying stringent cuts on $\Delta\eta_{12}$ and $M_{j_1 j_2}$, they arrived at the distribution shown in Figure 8.5. Clearly the signal has suppressed jet activity near $\eta^* = 0$, while the QCD background is enhanced near $\eta^* = 0$.

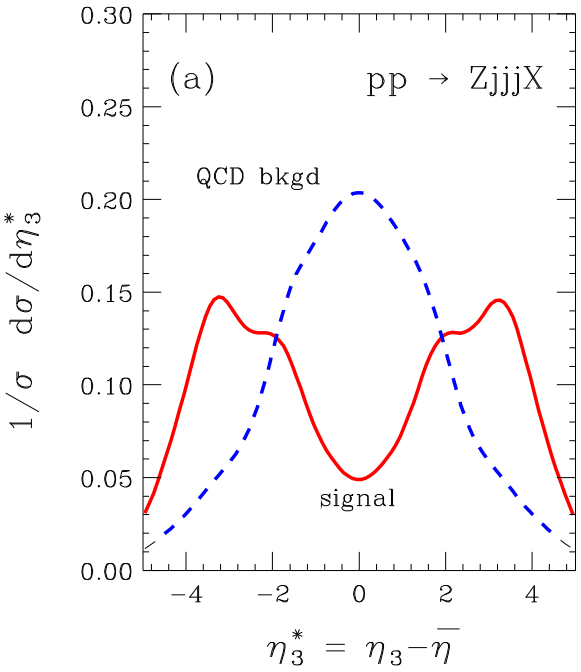


Figure 8.5 Distribution of η^* taken from Rainwater *et. al.*, Phys. Rev. **D**54.

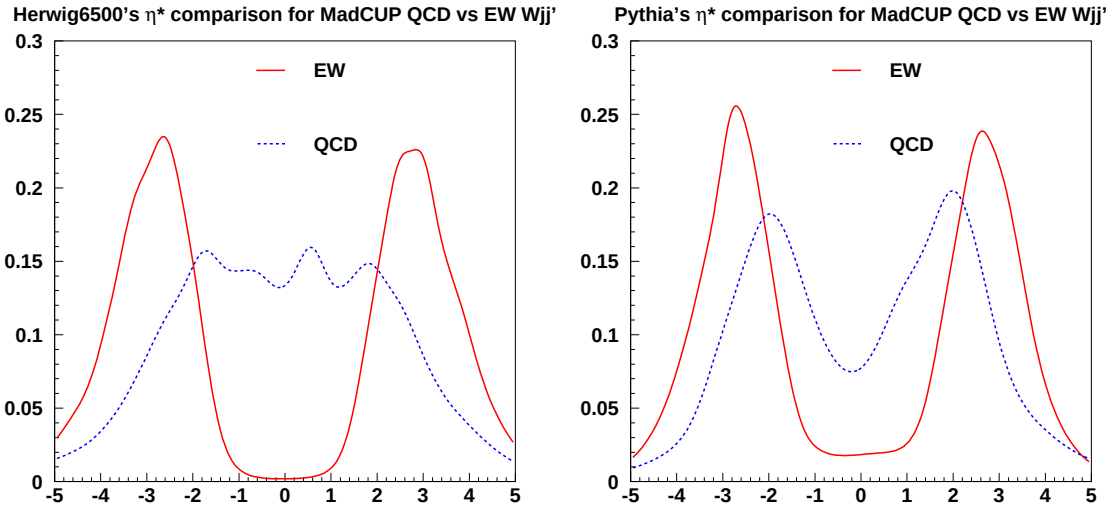


Figure 8.6 Distribution of η^* when the third jet is provided from the parton shower of HERWIG (left) and PYTHIA (right).

In order to test that this behavior survives when the third jet is provided by the parton shower, the EW and QCD Wjj external matrix elements were interfaced to PYTHIA and HERWIG. It was required that the W decay leptonically and that the outgoing partons had $|\eta| < 5.5$, $E_T > 20\ \mathrm{GeV}$, and $\Delta\eta_{12} > 3$ at the generator level. The ATLAS detector was simulated with ATLFAST and it was required that the tagging jets were in opposite hemispheres, had $E_T > 40\ \mathrm{GeV}$, and $\Delta\eta_{12} > 4$. Furthermore, it was required that the third jet had $E_T > 20\ \mathrm{GeV}$. The observed η^* distributions are shown in Figure 8.6. The η^* distributions were estimated using the KEYS package (see Appendix B); the unusual features in HERWIG's QCD η^* distribution are most likely just due to statistical fluctuation.

Clearly, the η^* distribution is much different when the third jet is obtained from the parton shower than when it is obtained with the matrix element. In both PYTHIA and HERWIG, the EW η^* distribution almost vanishes near $\eta^* = 0$. More disturbing is that the QCD distribution is also depleted near $\eta^* = 0$ for PYTHIA. Neither PYTHIA nor HERWIG provide the sharp peak in the QCD η^* that is seen in Figure 8.5.

The conclusion from these studies is that the central jet veto survival efficiency for both EW and QCD backgrounds will be higher when the third jet is obtained with the parton shower than when it is obtained from the matrix element. As a result, the analyses presented in the following chapters will have a higher level of background than expected in Refs. [111, 112, 113].

Currently, the most difficult theoretical challenge for VBF Higgs searches is the consistent description of additional hard jets. New tools like SHERPA [42] will allow for a consistent transition from jets produced from the parton shower to jets produced directly from the matrix element.

Chapter 9

Missing Transverse Momentum Reconstruction

9.1 Components of $\not{p}_T$ Resolution

In order to improve the Higgs mass resolution for VBF $H \rightarrow \tau\tau$ and $H \rightarrow WW$, it is important to understand the major sources of $\not{p}_T$ resolution. If we assume that the detector has a ϕ-symmetry, then we can rewrite the resolution in terms of its Cartesian components: $\sigma(\not{p}_T) = \sqrt{\sigma(\not{p}_x)^2 + \sigma(\not{p}_y)^2}$, with $\sigma(\not{p}_x) = \sigma(\not{p}_y)$. The $\not{p}_T$ resolution is itself due to the convolution of calorimeter energy resolution, electronic noise, and detector coverage effects, which we write suggestively as

$$\sigma(\not{p}_x) = \sigma_{calo} \oplus \sigma_{noise} \oplus \sigma_{geom}. \tag{9.1}$$

The calorimeter component, σ_{calo}, dominates the $\not{p}_T$ resolution. This contribution is usually parametrized as

$$\sigma_{calo} \approx \xi \sqrt{\sum |E_T|}. \tag{9.2}$$

Figure 9.1 shows Monte Carlo simulated $A \rightarrow \tau\tau$ and minimum bias events overlaid on a curve corresponding to $\xi = 0.46$. While this parametrization appears to work well for the high $\sum E_T$ region populated with minimum bias events, it does not fit very well the low $\sum E_T$ region populated with $A \rightarrow \tau\tau$ events. Furthermore, there clearly should be a constant offset for $\sum E_T \rightarrow 0$. These observations motivate a more in depth look at the $\not{p}_T$ resolution based on first principles.

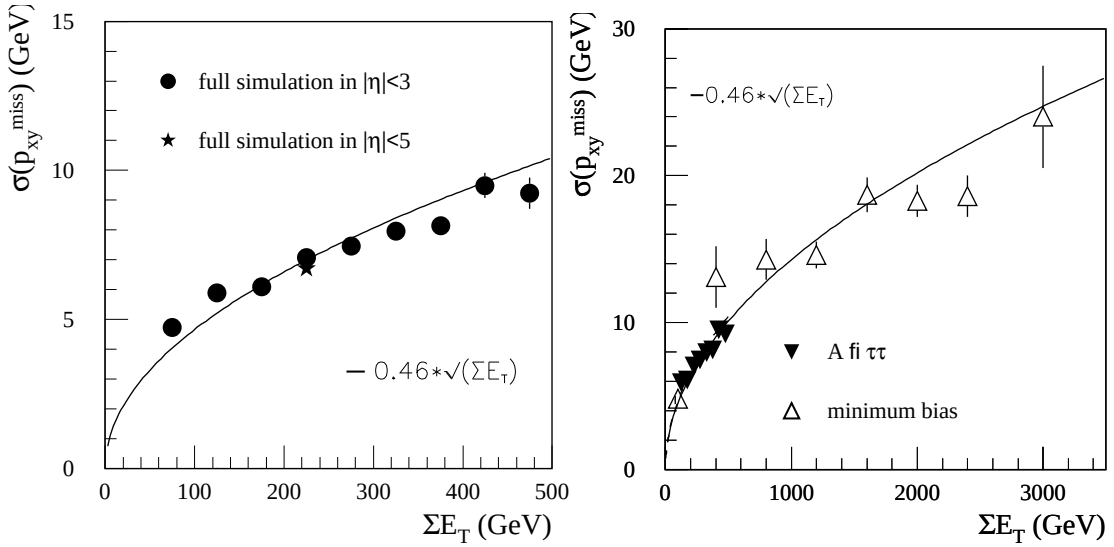


Figure 9.1 Parameterization of $\not{p}_x$ in ATLAS detector performance TDR.

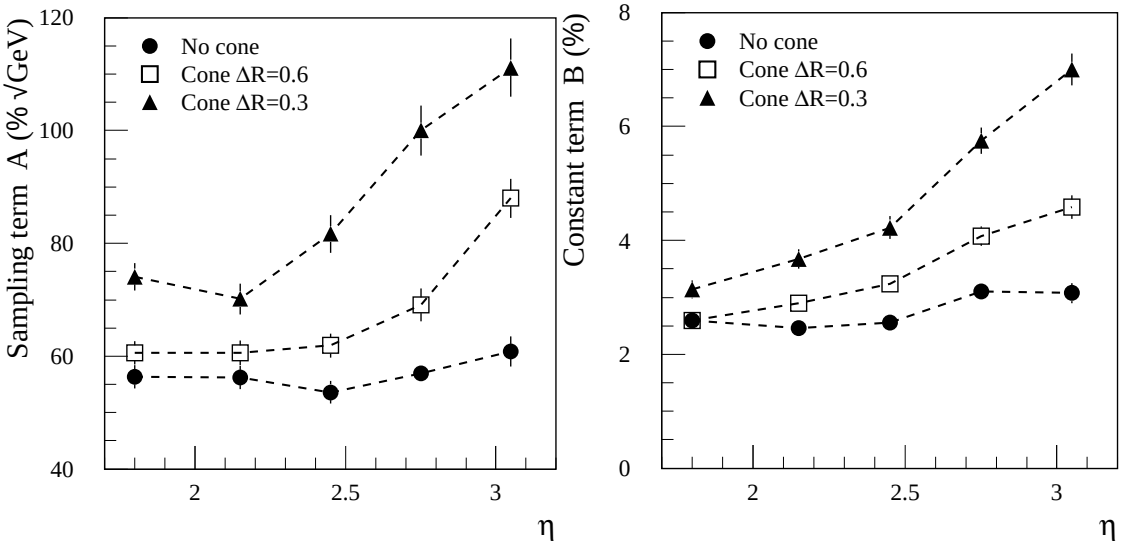


Figure 9.2 These TDR plots show the η-dependence of the sampling and constant terms used to parametrize the hadronic endcap energy resolution to a beam of pions.

9.1.1 Calorimeter Response

As for the calorimeter component, let us consider the particles that are interacting with the detector. It has been observed at the test beam that the calorimeter energy resolution for electrons and pions can be parametrized as

$$\frac{\delta E}{E} = \frac{A}{\sqrt{E}} + B, \tag{9.3}$$

where A and B are referred to as the sampling and constant terms, respectively, and are both η-dependent (see Figure 9.2). Considering these energy measurements as uncorrelated Gaussians with standard deviations given by δE, we can predict σ_{calo} from first principles to be

$$\sigma_{calo} = \sqrt{\sum_{i \in particles} \left(\frac{\delta E_i \cos\phi_i}{\cosh\eta_i} \right)^2}. \tag{9.4}$$

By substituting Eqn. 9.3 into Eqn. 9.4, averaging over ϕ, assuming $A\sqrt{E} \gg BE$, and factoring out $\sum E_T$ one obtains

$$\sigma_{calo} = \sqrt{\sum_{i \in particles} \left(\frac{(A_i\sqrt{E_i} + BE_i)\cos\phi_i}{\cosh\eta_i} \right)^2} \approx \underbrace{\sqrt{\left\langle \frac{A_i^2}{2\cosh\eta_i} \frac{E_{T,i}}{\sum E_T} \right\rangle}}_{\xi} \sqrt{\sum E_T}, \tag{9.5}$$

where ξ^2 is the E_T-weighted average of $A_i^2/2\cosh\eta_i$. Because of the E_T weighting, ξ is not a universal quantity, but depends on the sample of events in which one is interested.

9.1.2 Electronic Noise

Using a similar technique, we can estimate the noise component of the $\not{p}_T$ resolution, σ_{noise}. The noise component is a convolution of noise in each cell and is largely independent of the event properties. From test beam measurements, a database of RMS electronic noise, Δ_i, for each cell has been constructed. Because the noise is modeled as a Gaussian and the convolution of Gaussians is straightforward, it is not difficult to write an analytic expression for σ_{noise}. The difficulty arises from the evaluation of the the precise location of nearly 200,000 calorimeter cells within the detector geometry each with their own Δ_i. By using the aforementioned database, the

GEANT4 detector description, and neglecting the cell-to-cell correlations in the electronic noise, the expression for σ_{calo} and its numerical evaluation are as follows:

$$\sigma_{\rm noise} = \sqrt{\sum_{i \in cells} \left(\frac{\Delta_i \cos\phi_i}{\cosh\eta_i} \right)^2} = 13{\rm GeV}. \tag{9.6}$$

Without intervention σ_{noise} would dominate the $\not{p}_T$ resolution. However, we can reduce σ_{noise} by placing noise threshold on each cell, thus including fewer cells. The noise threshold can either be asymmetric (*e.g.* $E_i > N\Delta_i$) or symmetric (*e.g.* $|E_i| > N\Delta_i$).[1] In both cases, convolution is no longer a straightforward calculation, but can be approximated with toy Monte Carlo or Fourier Transform techniques (see Appendix A). From toy Monte Carlo studies, the reduction in σ_{noise} due to an $N\Delta$ noise threshold is nearly independent of the number of cells included in the convolution (see Table 9.1). In the case of a 1.5Δ asymmetric noise cut, the nominal σ_{calo} is reduced to 4.5 GeV, which is in good agreement with Figure 9.1.

N	Number of Cells	f_{sym} (%)	f_{asym} (%)
1	40	63.3	41.9
1	400	62.8	41.4
1	4000	63.7	41.3
1	40000	65.0	44.0
1.5	4000	51.1	34.7
2	4000	36.6	25.8

Table 9.1 Tabulated values of the ratio $\sigma_{noise}^{N\Delta}/\sigma_{noise}$ in percent, where $\sigma_{noise}^{N\Delta}$ represents the contribution to the $\not{p}_T$ resolution after an $N\Delta$ noise threshold is applied. The quantities f_{sym} and f_{asym} correspond to the symmetric and asymetric cases, respectively.

[1] In the asymmetric case, the energy in a truly empty cell is positively-biased. However, after ϕ-averaging over cells the mean $\not{p}_x$ is not biased.

9.1.3 Geometrical Acceptance

The limited geometric acceptance of the detector contributes the final component of the $\not{p}_T$ resolution, σ_{geom}. Because this component is due to unseen particles, it is difficult to improve. Clearly, σ_{geom} is dependent on the type of events under consideration. From Monte Carlo truth studies on Vector Boson Fusion events, the magnitude of σ_{geom} is roughly 2 GeV, which is mainly due to the forward tagging jets.

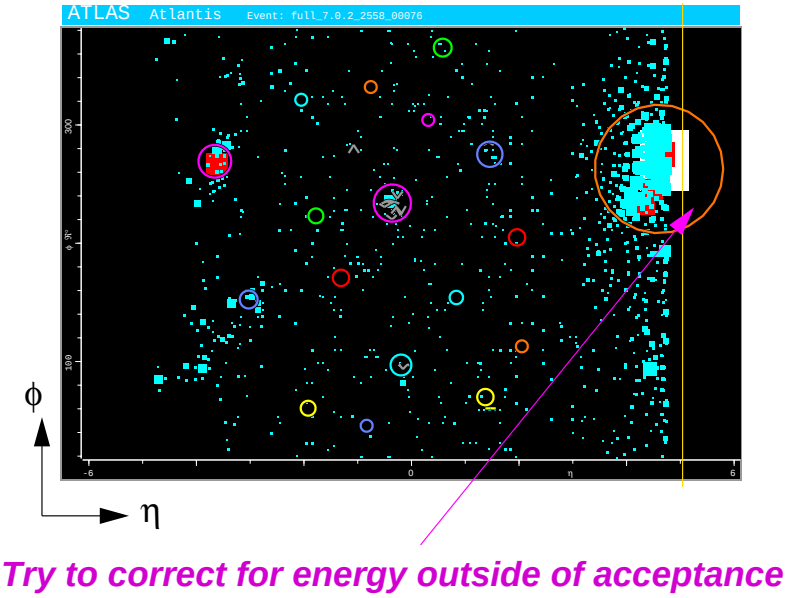


Figure 9.3 Illustration of geometric acceptance corrections to $\not{p}_T$ based on jets.

In an attempt to correct for energy depositions beyond the geometrical acceptance, the author developed a jet-based correction. The jets were considered as homogeneous cones in $\eta - \phi$, and the energy of the jet was corrected for the portion of the cone with $\eta > 5$ (see Figure 9.3). Several refinements to this technique were made, including a correction for the jet barycenter based on the lost energy. None of the refinements showed a significant improvement in the $\not{p}_T$ resolution.

9.2 The H1-Style Calibration Strategy

The ATLAS calorimeter is a non-compensating calorimeter, which means that the response to electromagnetic and purely hadronic components of a shower are not the same. The ratio of the calorimeter response to these components is typically denoted as e/h. By studying the response to pions with Tile calorimeter's barrel module zero in the energy range 10-400 GeV, the fitted value of non-compensation is $e/h = 1.30 \pm 0.01$ [85].

In order to correct for the non-compensation, the energy reconstruction of hadronic interactions must be calibrated from the electromagnetic energy scale. Several methods have been proposed and studied; however, at the time of this writing, the H1-style calibration is the most common strategy. The H1-style strategy takes its name from the H1 experiment at HERA, which corrected the response of individual cells. In this approach, small signals typically have larger corrections.

The extraction of the H1 calibration coefficients is currently performed by minimizing energy resolution of jets. For each jet, the contributing calorimeter cells are partitioned by calorimeter sampling, energy (or energy per unit volume), and pseudorapidity. Because this method is based on jets, the weights are coupled to the jet clustering algorithm. To improve the situation, a similar method has been applied in which the weights are extracted by minimizing the $\not{p}_T$ resolution directly. In that case, the cells are partitioned 18 η- and sampling-dependent sets each with 16 E_T-dependent calibration coefficients [2]. The resulting calibration coefficients range from 0.7 to 6.0.

9.3 Electronic Noise and Bias

As was mentioned in Section 9.1.2, the electronic noise in the calorimeter would dominate the $\not{p}_T$ resolution if noise suppression was not applied. The typical approach for noise suppression is a global noise threshold: *e.g.* a calorimeter cell is removed from the $\not{p}_T$ sum if $E_i < N\Delta_i$. In the remainder of this section it will be demonstrated that global noise suppression induces a bias in $\not{p}_T$ and an alternative local noise suppression strategy will be presented.

[2] The index of the bins is given by $8 + \log_2 |E_T|$, with E_T in units of GeV.

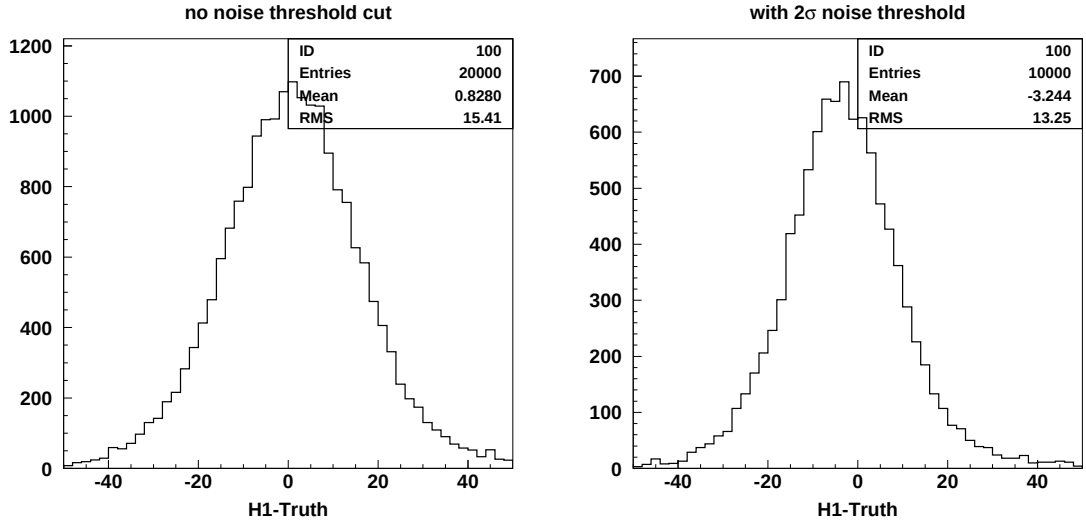


Figure 9.4 Distribution of H1-calibrated $\not{p}_T$ minus the Monte Carlo truth $\not{p}_T$ without noise suppression (left) and with a 2Δ asymmetric noise threshold (right) for VBF $H \rightarrow \tau\tau$ events. The 2Δ noise threshold improves the $\not{p}_T$ resolution, but induces a negative bias.

9.3.1 Evidence for Bias

To demonstrate that global noise suppression induces a bias in $\not{p}_T$, a sample of VBF $H \rightarrow \tau\tau$ events with the GEANT3 simulation of the ATLAS detector (from $-5 < \eta < 5$) including electronic noise have been processed with and without a 2σ asymmetric noise threshold. In Figure 9.4 it can be seen that the $\not{p}_T$ resolution without a noise threshold is about 15 GeV and that the mean of the distribution is slightly biased with respect to the Monte Carlo truth. This bias most likely due to the application of H1 calibration coefficients to calorimeter cells associated to energetic electrons and muons from τ decays. When the noise threshold is applied, the $\not{p}_T$ resolution is improved, but the mean value is shifted by about 3 GeV with respect to the Monte Carlo truth. The additional bias is due entirely to the application of a global noise threshold.

In the $H \rightarrow \tau\tau$ channel, the $\not{p}_T$ comes from the Higgs via τ decays, and the Higgs p_T is balanced against additional jet activity. When the low-energy portion of a jet deposits energy in a calorimeter cell that is comparable to that cell's RMS electronic noise, both symmetric and asymmetric noise thresholds will cause a bias (see Section 9.3.2). The cumulative effect of the

cell-by-cell bias causes a bias in $\not{p}_T$ in the direction of the jets. Because the lepton and τ-jet in $H \rightarrow \tau\tau$ are compact and very energetic, the energy deposition in cells is much larger than the RMS electronic noise; thus, the cells are not biased. This explanation is consistent with observation that the reconstructed Higgs mass in the $H \rightarrow \tau\tau$ channel is always too low.

9.3.2 When Symmetric Cuts Are Asymmetric

A common misconception in the argument presented above is that a symmetric noise threshold will not cause bias. The term "symmetric" is used when the noise threshold is of the form $|E| > N\Delta$; however, this requirement is only symmetric when the true energy is 0 GeV. The presence of real deposited energy in a calorimeter cell breaks this symmetry, and even "symmetric" noise cuts cause bias. To belabor the point just a bit, one can calculate the bias in a cell as a function of the true energy deposited in it, E_t. First, let us model the effect of electronic noise on a true energy deposition with a simple Gaussian form:

$$p(E_{meas}|E_{true}) = \frac{1}{\sqrt{2\pi\Delta^2}} e^{-\frac{(E_{meas}-E_{true})^2}{2\Delta^2}}. \tag{9.7}$$

Next, define the bias to be the average measured energy minus the true deposited energy

$$\mathrm{bias}(E_t) = \int_{-\infty}^{\infty} E \cdot \Theta(E; N\Delta) \cdot p(E|E_t)\, dE \; - E_t, \tag{9.8}$$

where $\Theta(E; N\Delta) = 1$ if the cell survives the noise cut and $\Theta(E; N\Delta) = 0$ if it does not. In the case that the cut is asymmetric, the bias is given by

$$\overline{\mathrm{bias_{asym}(E_t)}} \;=\; \frac{\Delta^2}{\sqrt{2\pi\Delta^2}} e^{-\frac{(N\Delta-E_t)^2}{2\Delta^2}} - \frac{E_t}{2}\left[1 + \mathrm{erf}\left(\frac{N\Delta - E_t}{\sqrt{2\Delta^2}}\right)\right]. \tag{9.9}$$

When the cut is is symmetric the bias is given by

$$\begin{aligned} \overline{\mathrm{bias_{sym}(E_t)}} \;&=\; \frac{\Delta^2}{\sqrt{2\pi\Delta^2}}\left[e^{-\frac{(N\Delta-E_t)^2}{2\Delta^2}} - e^{-\frac{(N\Delta+E_t)^2}{2\Delta^2}}\right] \\ &-\; \frac{E_t}{2}\left[2 + \mathrm{erf}\left(\frac{N\Delta - E_t}{\sqrt{2\Delta^2}}\right) - \mathrm{erf}\left(\frac{N\Delta + E_t}{\sqrt{2\Delta^2}}\right)\right]. \end{aligned} \tag{9.10}$$

The bias as a function of E_t in both cases is shown in Figure 9.5 for several values of N. Note that when $E_t = 0$, the symmetric cut does not cause a bias, but $E_t > 0$ the bias is always negative. For the asymmetric case, the bias is positive for $E_t = 0$ as one would expect. In both cases, when $E \gg N\Delta$ the bias is negligible.

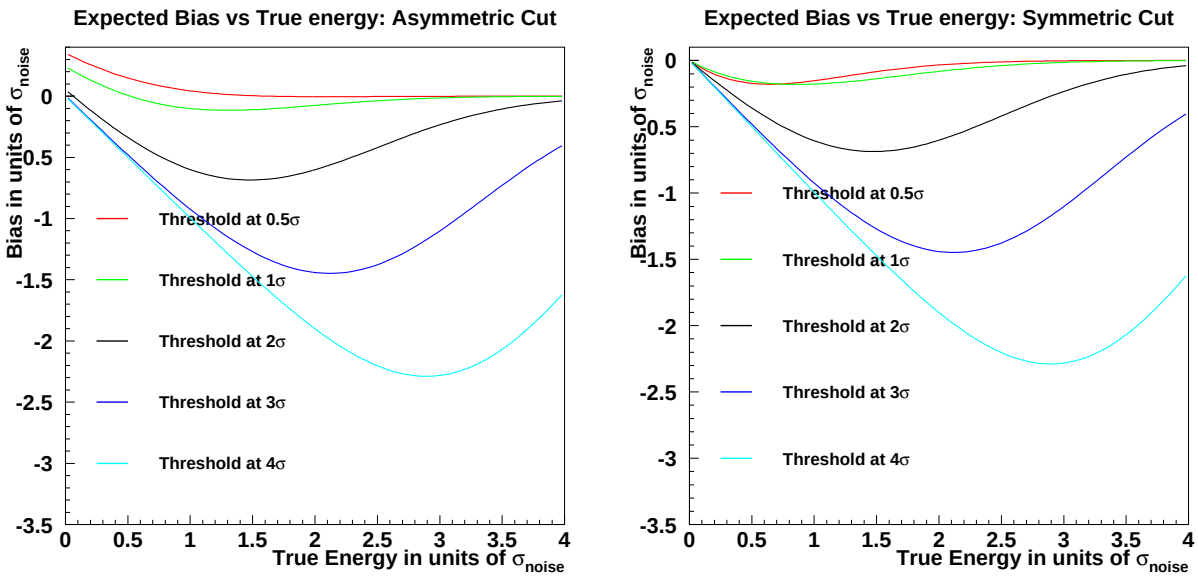


Figure 9.5 The bias on a cell due to an asymmetric (left) or symmetric (right) noise threshold as a function of the true deposited energy.

9.3.3 When Asymmetric Cuts Are Symmetric

In the previous section, we demonstrated that both asymmetric and symmetric noise thresholds can cause bias at the cell level. However, the ϕ-symmetry of a completely empty and noisy detector results in an unbiased estimate of $\not{p}_T$ even for an asymmetric cut. This argument holds true for both symmetric and asymmetric cuts as long as the ϕ-symmetry exists. The presence of a real energy deposition can destroy the ϕ-symmetry, allowing for the bias seen in Section 9.3.1.

9.3.4 Local Noise Suppression

As justified in the previous sections, the removal of calorimeter cells with small, positive deposited energy causes a bias in $\not{p}_T$. On the other hand, without some form of noise suppression, the $\not{p}_T$ resolution is prohibitively poor. Ideally, we would implement a local noise suppression that does not remove cells with true energy depositions, but does remove cells only containing noise. This section and the next describe an attempt at such an algorithm.

The first step of the Local Noise Suppression (LNS) algorithm is to use Bayes' theorem to estimate the true energy in a calorimeter cell from it's measured energy and some *a priori* probability distribution for the cell's true energy. In particular, the *a posteriori* probability on the cell's true energy is given by Equation 9.7 together with

$$p(E_{true}|E_{meas}) = \frac{p(E_{meas}|E_{true})p(E_{true})}{p(E_{meas})}. \tag{9.11}$$

Clearly, the *a priori* distribution, $p(E_t)$, should give special preference to $E_t = 0$ because most calorimeter cells are, in fact, empty. If, in addition, we want the property that we recover the measured energy (in the form of a mean, median, or maximum likelihood estimator) when $E_{meas} >> \Delta$, then we are almost forced into a flat prior on E_t.[3] Thus, for what follows, we will model the *a priori* distribution as:

$$p(E_{true}) = a_0\delta(0) + \text{flat prior elsewhere}, \tag{9.12}$$

where δ is the Dirac δ-function and the coefficient a_0 represents the preferential treatment of $E_t = 0$. The estimation of a_0 is the subject of the next section.

With Equations 9.7 and 9.12 in hand, we can evaluate Equation 9.11. The *a posteriori* probability is not the end goal; instead, we desire an estimate of the true energy given the measured energy and a_0. An approximate solution to this problem is simply the weighted average of the energy at $E_t = 0$ and $E_t = E_m$, *viz.*

$$\hat{E}(E_m; a_0) = \frac{0 \cdot a_0 e^{-E_m^2/2\Delta^2} + (1 - a_0)E_m}{(1 - a_0) + a_0 \cdot e^{-E_m^2/2\Delta^2}}. \tag{9.13}$$

The estimate of the true energy as a function of the measured energy is plotted in Figure 9.6 for several values of $a_0 = p(E_t = 0)$. Notice that Equation 9.13 has the following properties:

- There are no discontinuities in the estimate.
- $\lim_{E_m \to \infty} \hat{E}(E_m; a_0) = E_m$ when $a_0 \neq 1$

[3] It is tempting to estimate $p(E_t)$ from Monte Carlo, but it probably will not satisfy the aforementioned property because the energy distribution has an incredibly complex structure when considered as a joint distribution over all cells.

- $\lim_{a_0 \to 0} \hat{E}(E_m; a_0) = E_m$
- $\lim_{a_0 \to 1} \hat{E}(E_m; a_0) = 0$

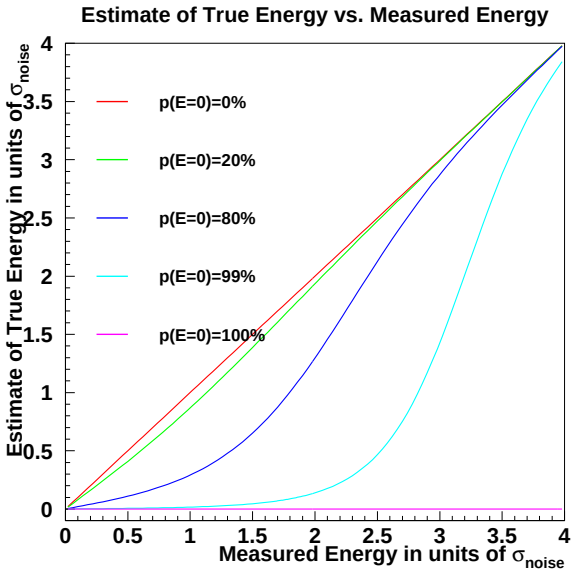


Figure 9.6 Estimated true energy as a function of measured energy and $p(E_t = 0)$.

9.3.5 Estimating the Prior

One can think of Equation 9.13 as a generalization of the asymmetric noise threshold (with no discontinuities and a some statistical justification), but it does not solve the problem of bias in $\not{p}_T$ if a_0 is a global quantity. The essence of the local noise suppression method is that the noise cut is *local* – automatically adjusting to the topology of each event.

A powerful indication that a cell is empty would be if each of the cell's neighbors were empty. Because each of the neighboring cells also have electronic noise, we must introduce some intelligent way to make that inference. Assume that all of the neighbors of a cell are empty, then the quantity

$$\mathcal{X} = \sum_{i \in \text{neighbors}} E_i / \Delta_i \tag{9.14}$$

will be distributed as

$$P(\mathcal{X}) = \frac{1}{\sqrt{2\pi N_n}} e^{-\mathcal{X}^2/2N_n}, \tag{9.15}$$

where N_n is the number of neighboring cells. Within the ATLAS software, it is quite easy to get 2-dimensional and 3-dimensional neighbors even across calorimeter samplings.

We have investigated two methods to estimate a_0 for a given cell:

$$a_0 = a_{max} \cdot \int_{\mathcal{X}}^{\infty} P(\mathcal{X}')d\mathcal{X}' \tag{9.16}$$

and

$$a_0 = a_{max} \cdot P(\mathcal{X})/P(\mathcal{X} = 0), \tag{9.17}$$

where a_{max} is a global parameter used to avoid $a_0 = 1$, in which case $\hat{E}(E_m; a_0)$ always vanishes.

9.3.6 Comparison of Local and Global Noise Suppression

The essence of the local noise suppression strategy is not energy rescaling found in Equation 9.13, but the fact that the noise suppression is based on neighboring cells. Figure 9.7 shows, for an example event, those cells that would be cut by a global 2Δ noise cut, but would not be cut with the local noise suppression strategy. Most of those cells are highly correlated to the physics event and jet structure is visible. In addition their are a few randomly distributed cells for which $\mathcal{X}$ had a large upward fluctuation. The cells correlated to the physics event are precisely the cells, which when cut, produce a bias in $\not{p}_T$.

Figure 9.8 compares the $\not{p}_T$ resolution (defined as the difference in magnitude for reconstructed and truth) for the global 2Δ (left) and local noise suppression (right) strategies. Both plots are made using the VBF $H \rightarrow \tau\tau$ sample described in Section 10.10.1. The local noise suppression improves the resolution from 17 GeV to 14 GeV and reduces the bias by 88%.

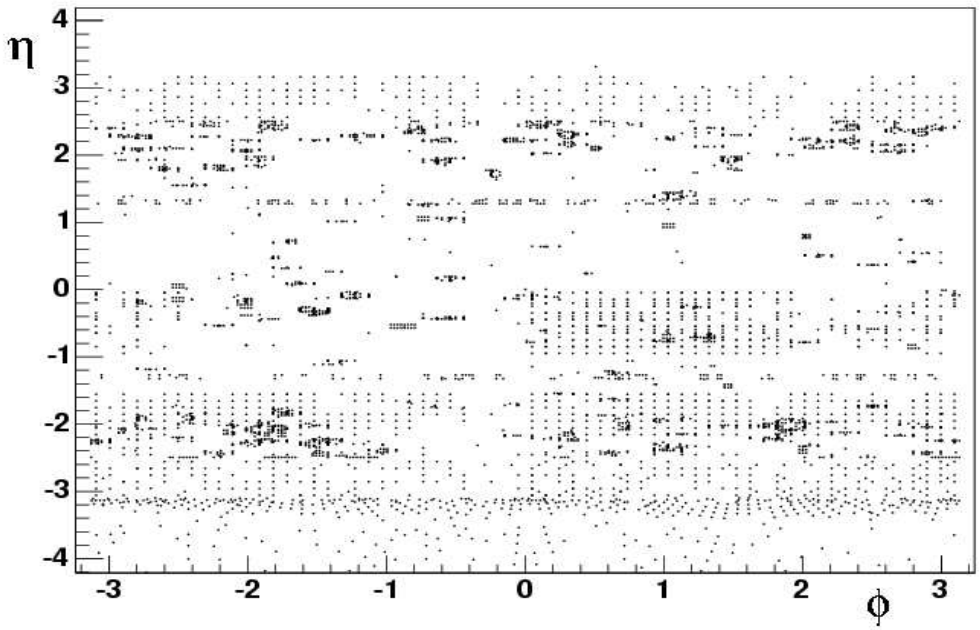


Figure 9.7 An illustration of cells in the $\eta - \phi$ plane which would be cut by a global 2Δ cut, but would not be cut with the local noise suppression technique. Jet structure can be seen in several areas.

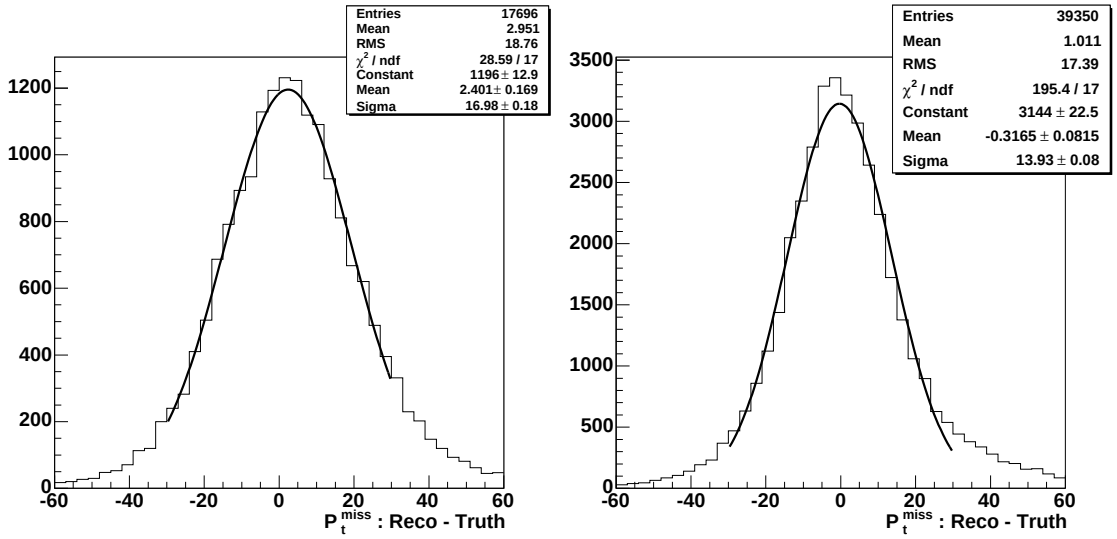


Figure 9.8 Comparison of $\not{p}_T$ resolution for a global 2Δ noise cut (left) and local noise suppression (right) with GEANT4 and digitized electronic noise.

Chapter 10

Vector Boson Fusion $H \rightarrow \tau\tau$

Vector Boson Fusion (VBF) is the second leading production mechanism for Higgs at the LHC (see Section 2.3). Near the LEP Higgs limit of $m_H \approx 114\ \mathrm{GeV}$, the Higgs primarily decays to fermions. In order to trigger on Higgs events, the final state must include a high-p_T lepton or photon. Thus, the three most powerful channels in this low mass range come from $H \rightarrow \gamma\gamma$, $t\bar{t}H$ with $H \rightarrow b\bar{b}$ and at least one top quark decaying leptonically, and VBF $H \rightarrow \tau\tau$ with at least one tau decaying leptonically.

The fully-leptonic and semi-leptonic VBF $H \rightarrow \tau\tau$ analyses are very similar; however, we shall focus on the semi-leptonic (a.k.a lepton-hadron) channel in the remainder of this section. The lepton-hadron channel accounts for 45% of the $H \rightarrow \tau\tau$ signal and offers *bona fide* mass reconstruction. In the MSSM this channel is very important due to the enhanced branching ratio of Higgs to tau leptons and the complementarity of the light and heavy neutral Higgs bosons in the $M_A - \tan\beta$ plane [118].

10.1 Experimental Signature

The experimental signature of all VBF Higgs channels consists of two forward, high-p_T tagging jets with large η-separation and little jet activity between the tagging jets. The Higgs is usually produced in the central rapidity region with significant p_T, thus the decay products tend to lie between the two tagging jets. In the $H \rightarrow \tau\tau$ channels, there is also significant $\not{p}_T$ due to the tau decays. A schematic representation of a VBF $H \rightarrow \tau\tau \rightarrow lh\not{p}_T$ event is shown in Figure 10.1.

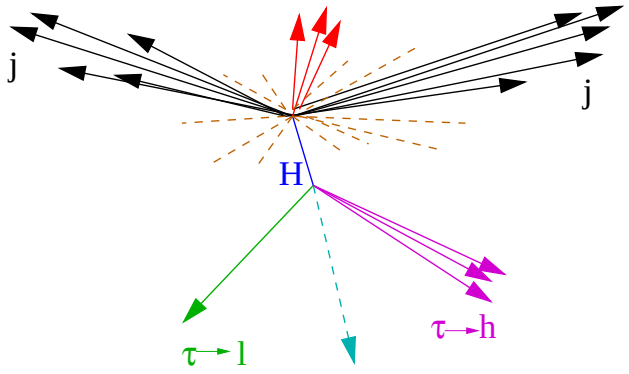


Figure 10.1 Schematic representation of a $H \rightarrow \tau\tau \rightarrow lh\not{p}_T$ event.

10.2 Identification of Hadronically Decaying Taus

The identification of hadronic tau decays and the rate of their mis-identification due to parton-initiated jets has been studied in Refs. [119, 120]. The first tau identification strategies employed by ATLAS were seeded with high-p_T clusters that were then matched to tracks. Currently, there is active involvement in tau identification seeded with tracks [121]. The results shown below are all based on the cluster-seeded approach.

Historically in ATLAS τ-jet separation has been achieved through the electromagnetic radius of the shower , R_{EM}, the isolation fraction ΔE_T^{12} (see below), and the number of tracks matching the calorimeter cluster. Based on these discriminating variables, the tau identification efficiency, ϵ_τ, and jet rejection, R_j, have been parametrized for use in ATLFAST. In the first fast simulation studies [122, 123, 124], the tau efficiency was set to $\epsilon_\tau = 50\%$, which corresponds to a jet rejection of about $R_j \sim 100$.

More recently, the tau-jet separation methods have been extended into an approach which uses five continuous and three discrete discriminating variables to construct a log likelihood-ratio q_τ. The variables used to construct the likelihood-ratio are:

- the electromagnetic radius, R_{EM}, of the cluster,
- the Isolation Fraction, ΔE_T^{12}, defined as the ratio of the difference between energies in cones of size $\Delta R = 0.1$ and 0.2 to the total E_T of the τ-jet,

- the variance in η of the E_T-weighted strips,
- the ratio of the cluster's p_T to p_T of the highest p_T matched track,
- the signed impact parameter of the highest p_T matched track,
- the number of tracks,
- the number of cells in the electromagnetic strip matching the cluster,
- the sum of the charges of the matched tracks.

In the full simulation studies presented below, the tau-jet separation requirement is $q_\tau > 1$, which corresponds to $\sim 70\%$ efficiency and a jet rejection of > 100 for $p_T > 40$ GeV.

The τ-jets in ATLFAST are not treated differently than other hadronic jets; however, the τ-jets are specifically calibrated using an H1-style calibration strategy in the full simulation. The H1-style calibration provides a τ-jet p_T resolution of $\sigma/E = 80\%/\sqrt{E} + 1.4\%$ [125]. Recent results show that a dedicated energy-flow algorithm can significantly improve the resolution in the low p_T region. Lastly, in the full simulation results, the η and ϕ of the τ-jet come from the matched track and the p_T is taken from the H1-calibrated cluster.

10.3 Electron Identification

Electron identification in ATLAS is achieved through both cluster-seeded and track-seeded approaches. The seed clusters are found with a sliding-window in the electromagnetic calorimeter and associated to matched tracks. These clusters have been well studied and have a suite of calibrations applied to them. The electromagnetic clusters found from the track-seeded approach have a different topology and do not have the same suite of calibrations available. Thus, if two electron candidates have the same matched track, the candidate found with the cluster-seeded approach is chosen.

In order to provide good electron-jet separation, a number of cuts on the electromagnetic shower shape and track quality have been applied. The shower shape cuts are η-dependent and include:

- a cut on fraction of energy deposited in first sampling;
- a cut on hadronic energy;
- a cut on the ratio of energy in a 3x7 window to the energy in a 7x7 window in the second sampling;
- a cut on shower width in second sampling;
- a cut on the ratio of the energy of the second highest energy cell in the first sampling to a 3x7 cluster;
- a cut on the difference in energy between the first and second highest energy cells in first sampling;
- a cut on the total width in first sampling;
- a cut on the width in first sampling.

The track quality cuts include at least 1 b-layer hit, 1 pixel-layer hit, 7 precision hits, a transverse impact parameter less than 200 $\mu\mathrm{m}$, and a track match with $\Delta\eta < 0.02$ and $\Delta\phi < 0.05$. In addition, the ratio of the cluster energy to the track momentum, E/p, must be between 70 and 400%. Due to problems with offline release 9.0.0 of the ATLAS software, the requirement that at least 10% of the TRT hits must be high-threshold was not applied.

10.4 Muon Identification

Muons are identified with the a combination of the muon spectrometer, the inner detector, and calorimetry. In the full simulation results, the MOORE package has been used for tracks in the muon spectrometer. These tracks are then extrapolated to the interaction point and a list of combined muon candidates is formed; one for each matched inner detector track. A global χ^2 for the muons is constructed taking into account the complex magnetic field and multiple scattering effects. The MUID package has been used to perform this combined fit. The combined muon candidate with the best χ^2 is chosen for each track in the muon spectrometer.

Because the $\not{p}_T$ is calculated from a loop over calorimeter cells and muons leave little energy in the calorimeter, the $\not{p}_T$ must be corrected for all identified muons. A correction based on the momentum of the combined muon would double-count the energy deposition in the calorimeter, so instead the momentum of the track in the muon spectrometer is used for this correction. This is possible because the muon spectrometer provides an independent momentum measurement after the muon has transversed the hadronic calorimeter. The energy deposited from the muon is included in the $\not{p}_T$ calculation; however, those cells are currently calibrated using H1-style calibration coefficients derived from samples of hadronic jets. The proper treatment of the muons in the context of $\not{p}_T$ is an area of ongoing activity.

10.5 Jet Finding

The ATLAS software currently implements two jet clustering algorithms: a cone algorithm with split & merge and a k_T algorithm. The k_T algorithm has been chosen for this analysis due to its infrared safety and its more robust theoretical interpretation. The input to the k_T algorithm is a collection of $\eta - \phi$ projective *towers*. Due to the subtraction of the electronic noise pedestal, it is possible that these towers have negative energy. Thus, the negative energy towers are merged with neighboring towers until all their energies are positive. In the future, jets based on clusters and more sophisticated noise suppression techniques will be employed. This should help improve both the jet energy resolution and the forward jet tagging efficiency.

In Ref. [107], Cavasinni, Costanzo, and Vivarelli studied the efficiency of the forward jet tagging. Their procedure involved finding the rate that a parton matched a reconstructed jet within a cone of $\Delta R = 0.2$ for fast and full simulation; denoted ϵ_{q-j}^{fast} and ϵ_{q-j}^{full}, respectively (see Figure 10.2 left). Those authors provided a routine called DICECORR.F, which was used to account for the probability that an ATLFAST jet would be reconstructed in the full simulation. The correction used was the ratio of parton-jet matching efficiencies $\epsilon_{q-j}^{full} \,/\, \epsilon_{q-j}^{fast}$. For $P_T > 30$ GeV the probability an ATLFAST jet would be reconstructed in the full simulation is $> 88\%$ for $|\eta| < 4.5$.

Defining the jet tagging efficiency with respect to the parton before parton shower introduces unnecessary complications in the interpretation of forward tagging. The author carried out a similar

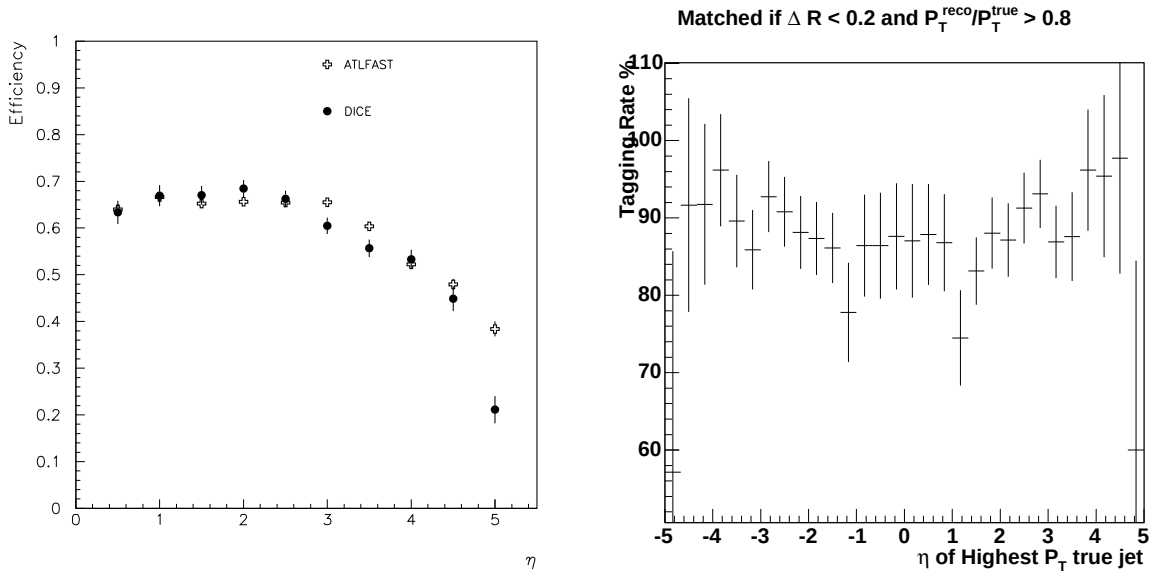


Figure 10.2 Left: parton-jet matching efficiencies for fast and full simulation found by Cavasinni, Costanzo, Vivarelli. Right: jet tagging efficiencies based on Monte Carlo truth jets.

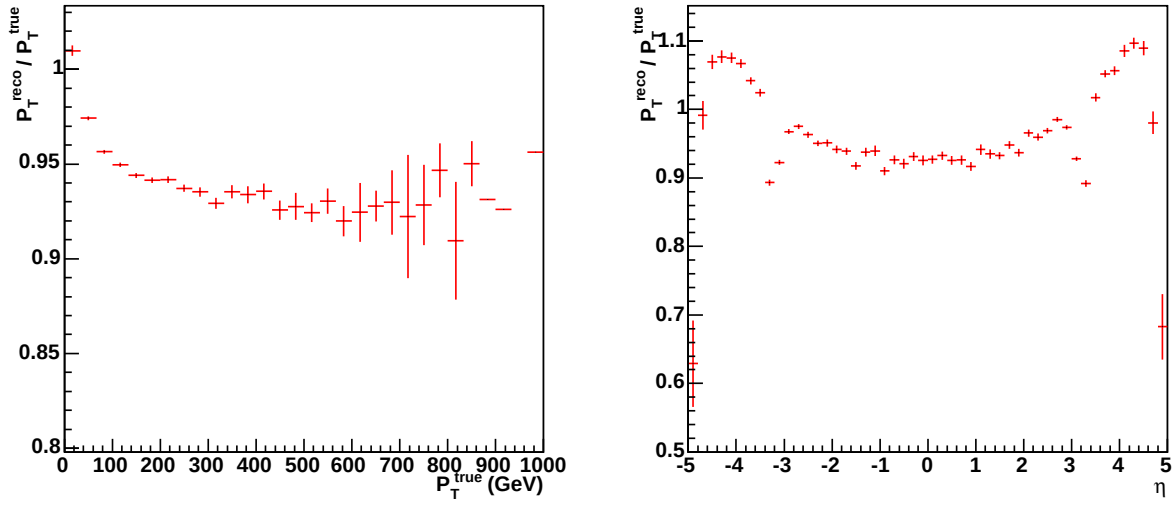


Figure 10.3 The ratio of reconstructed to truth jet p_T as a function of the true jet's p_T and η

study with GEANT4 that defined the jet tagging efficiency with respect to *Monte Carlo truth jets*. Monte Carlo truth jets are sets of final state particles immediately after hadronization clustered by the same cone or k_T algorithm. A Monte Carlo truth jet was considered to be matched if a reconstructed jet fell in a cone with $\Delta R < 0.2$ and the reconstructed jet p_T at least 80% of the truth jet energy p_T.

The simulation of the relative response of the electromagnetic and hadronic calorimetry is significantly different between GEANT3 and GEANT4. The extraction of H1-style calibration weights for the GEANT4 simulation is ongoing and active project within the ATLAS collaboration. At the time of this writing, the best available H1-style calibration coefficients are known as the `G4Beta2` weights. In Figure 10.3 the jet p_T linearity as a function of p_T and η.

10.6 The Collinear Approximation

In order to reconstruct the Higgs boson in this channel, we must account for the momentum of the neutrinos produced from τ decays. Because the neutrinos do not interact significantly with the detector, their momentum can be inferred from the missing transverse momentum, $\not{p}_T$ in the event. The $\not{p}_T$ reconstruction is the most experimentally challenging aspect of the $H \rightarrow \tau\tau$ analysis, and a detailed description of the $\not{p}_T$ reconstruction techniques can be found in Chapter 9.

For very high momentum τ's, one can approximate the direction of the neutrinos to be collinear with the visible τ decay products. This "collinear approximation" fixes the direction of the neutrinos, but not the fraction τ's momentum that they carry away by the neutrino. These two fractions can be determined from the two components of the missing transverse momentum: $\not{p}_x$ and $\not{p}_y$. In particular, let h and l denote the momentum of the hadronic and leptonic τ decay products, respectively. Then the fraction of the corresponding τ's momentum carried away by the lepton ($x_{\tau l}$) or hadron ($x_{\tau h}$) is given by the following relationships.

$$\begin{aligned} x_{\tau h} &= \frac{h_x l_y - h_y l_x}{h_x l_y + \not{p}_x l_y - h_y l_x - \not{p}_y l_x} = \frac{n_\tau}{D_h} \\ x_{\tau l} &= \frac{h_x l_y - h_y l_x}{h_x l_y - \not{p}_x h_y - h_y l_x + \not{p}_y h_x} = \frac{n_\tau}{D_l} \end{aligned} \tag{10.1}$$

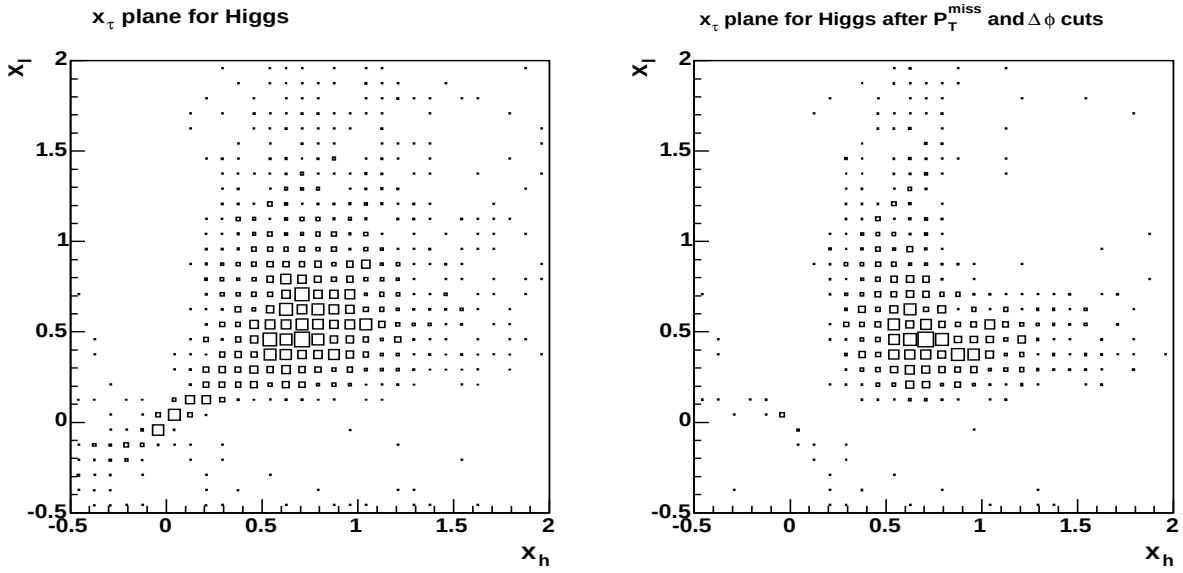


Figure 10.4 Distribution of signal events in the x_l–x_h plane with no cuts (left) and after the requirements $\not{p}_T > 30$ GeV and $\cos\Delta\phi > -0.9$ (right) with GEANT4 and digitized noise.

where n_τ, D_h, and D_l have been introduced for later convenience. The distribution of the x_τ variables is shown in Figure 10.4. Finally, it is possible to reconstruct the Higgs mass in the collinear approximation

$$M_{\tau\tau} = \sqrt{2(E_l + E_{\nu_l})(E_h + E_{\nu_h}) \cdot (1 - \cos\theta)} = \frac{M_{lh}}{\sqrt{x_{\tau h} x_{\tau l}}} = \frac{M_{lh}}{n_\tau} \cdot \sqrt{D_h D_l}, \qquad (10.2)$$

where θ is the opening angle between the two τ's.

10.6.1 Jacobian for $M_{\tau\tau}$

When the azimuthal separation of the two τ's, $\Delta\phi_{\tau\tau}$ approaches π, then the solution becomes unstable; thus it is required that events have $\cos\Delta\phi_{\tau\tau} > -0.9$. Similarly, when the two decay products have large invariant mass, $M_{\tau\tau}$ is sensitive to the $\not{p}_T$ resolution even though the solution of the x_τ equations may be fairly insensitive. To summarize the complicated relationship between $M_{\tau\tau}$ and $\not{p}_T$, it is useful to calculate the Jacobian transformation, J, from $\not{p}_x$ ($\not{p}_y$) to $M_{\tau\tau}$. This Jacobian factor is a function of l, h, and $\not{p}_T$, and transforms the $\not{p}_x$ ($\not{p}_y$) resolution into the resolution

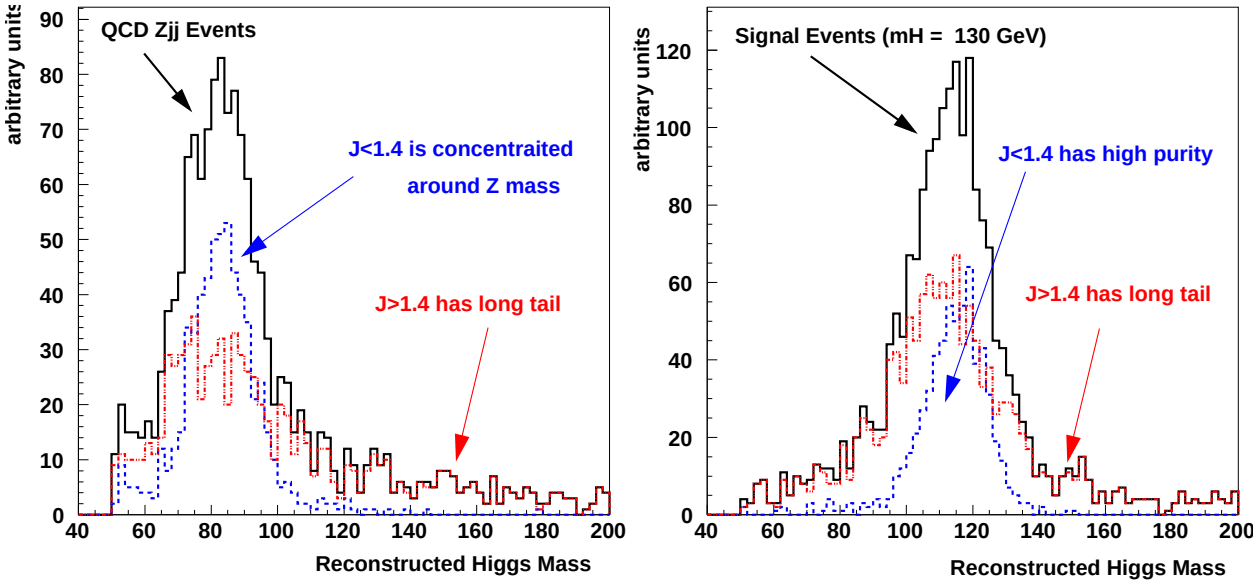


Figure 10.5 Reconstructed Higgs mass for events in the low- and high-purity samples with ATLFAST.

of $M_{\tau\tau}$.

$$\Delta M_{\tau\tau} = \Delta \not{p}_{x/y} \cdot \underbrace{\frac{M_{lh}}{2n_\tau\sqrt{D_h D_l}} \cdot \sqrt{(l_y D_l - h_y D_h)^2 + (h_x D_h - l_x D_l)^2}}_{J} \tag{10.3}$$

When J is small, $M_{\tau\tau}$ is not very sensitive to a mis-measurement of $\not{p}_T$; conversely, when J is large. It is possible that the statistical significance of the channel could be improved by using J to define low- and high-purity samples of events. To illustrate this point, signal and $Z \rightarrow \tau\tau$ simulated with ATLFAST have been partitioned into events with $J < 1.4$ and $J > 1.4$. Figure 10.5 shows a strong suppression in the $Z \rightarrow \tau\tau$ mass tail for the high-purity sample.

10.6.2 A Maximum Likelihood Approach

In $H \rightarrow \tau\tau$ and $Z \rightarrow \tau\tau$ events, the true $\not{p}_T$ should lie in the azimuthal wedge bounded by the visible decay products (see light green region in Figure 10.6). Due to the significant $\not{p}_T$ resolution, it is quite common that the reconstructed $\not{p}_T$ lies outside that wedge (dark red region), which leads to unphysical solutions to the x_τ equations. It is common practice to reject events with unphysical

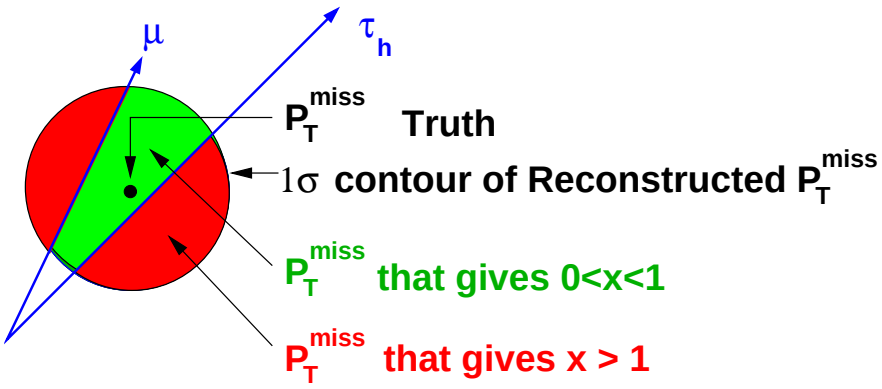


Figure 10.6 Schematic of the impact of $\not{p}_T$ resolution on the solutions of the x_τ equations.

solutions to the x_τ equations. When $\Delta\phi_{\tau\tau}$ is small, the chance that the solution to the x_τ equations is unphysical is quite large, which leads to a large loss in signal efficiency.

The author has developed a maximum likelihood technique in which it is possible to recover the lost signal efficiency while retaining the rejection against backgrounds inconsistent with $X \rightarrow \tau\tau$. The procedure consists of a scan over the $x_{\tau l}$ and $x_{\tau h}$ variables in the physical region. Each hypothesized value of $x_{\tau l}$ and $x_{\tau h}$ corresponds to a hypothesized $\not{p}_T$ given by:

$$\frac{1-x_{\tau l}}{x_{\tau l}} \cdot \vec{l} + \frac{1-x_{\tau h}}{x_{\tau h}} \cdot \vec{h} = \not{p}^{hypo}. \tag{10.4}$$

The consistency of the hypothesized $\not{p}_T$ to the reconstructed $\not{p}_T$ is provided by the χ^2 distribution with two degrees of freedom, where the χ^2 is defined as

$$\chi^2 = \left(\frac{\not{p}_x^{hypo} - \not{p}_x^{reco}}{\sigma(\not{p}_x)}\right)^2 + \left(\frac{\not{p}_y^{hypo} - \not{p}_y^{reco}}{\sigma(\not{p}_y)}\right)^2. \tag{10.5}$$

The point in the x_τ plane that minimizes the χ^2 – or maximizes likelihood – is equivalent to the solution to the x_τ equations when the solution is physical [1]. When the solution of the x_τ equations is unphysical, the maximum likelihood approach maintains a physical value of the x_τ's and the consistency of the event with $X \rightarrow \tau\tau$ is provided by the χ^2. One can cut arbitrarily hard on the χ^2 to remove unwanted backgrounds, or relax the cut to increase signal efficiency.

[1] In that case $\chi^2 = 0$.

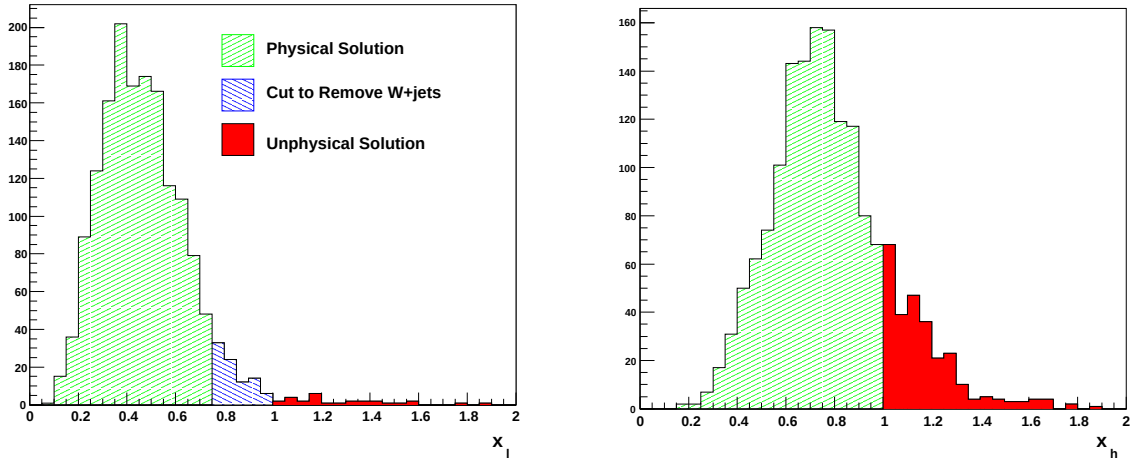


Figure 10.7 Distributions of $x_{\tau l}$ and $x_{\tau h}$ for signal events after $\not{p}_T > 30$ GeV and $\Delta\phi_{\tau\tau}$ cuts. Solid filled areas denote unphysical solutions to the x_τ equations.

The distribution of $x_{\tau l}$ and $x_{\tau h}$ for events with $\chi^2 < 1$ is shown in Figure 10.7. Events with $x_{\tau l} > 0.75$ are normally rejected to reduce W+jets backgrounds, so that cut is not relaxed. However, 27% of signal with $\chi^2 < 1$ can be recovered with the maximum likelihood approach.

10.7 Central Jet Veto

As explained in Chapter 8, one of the key features of Vector Boson Fusion Higgs production is the suppression of QCD radiation between the two forward tagging jets. In contrast, the QCD background Z+jets is expected to have relatively enhanced QCD radiation between the tagging jets (see Figure 8.5). This is the motivation for the central jet veto (CJV).

Because the CJV is sensitive to pile-up and jet clustering effects, it was studied in the full simulation by Cavasinni, Costanzo, and Vivarelli in Ref. [107]. The result of that study was a parametrization of the rate of "fake" central jets from minimum bias interactions as a function of the p_T threshold on central jets. In order to apply that correction to the fast simulation – which does not model pile-up or electronic noise – the definition of central jets was changed from any jet between the tagging jets to any jet with $|\eta| < 3.2$.

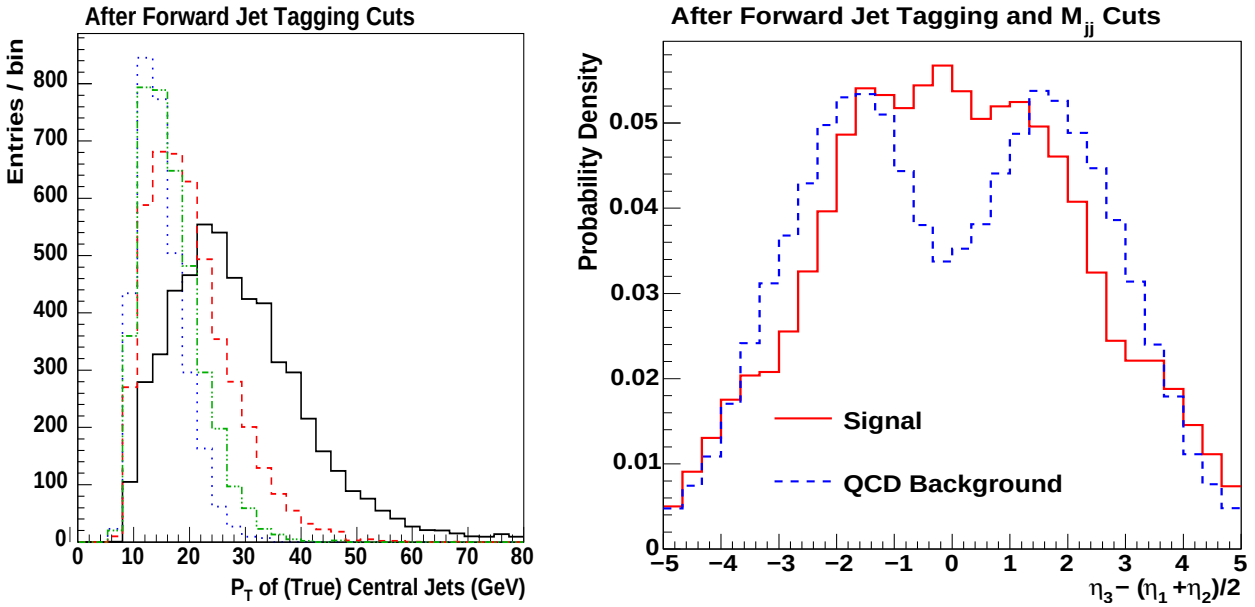


Figure 10.8 Distribution of p_T (left) and η^* (right) for the non tagging jets.

The CJV is also sensitive to the underlying event, which has a large uncertainty in current simulations (see Section 7.1.2). At the time of the fast simulation studies, the underlying event was modeled with PYTHIA's default settings. After tuning PYTHIA's more elaborate multiple-interaction model to the Tevatron data, simulations of the underlying event predicted many more high-p_T particles. The Monte Carlo truth jets (based on final state Monte Carlo truth particles and clustered with the k_T algorithm) show many more high-p_T central jets than in the previous fast simulation studies. The left inset of Figure 10.8 shows the p_T of the third, fourth, fifth, and sixth hardest jets [2] in $|\eta| < 3.2$ for signal events after applying the forward jet tagging requirements.

In Chapter 8 we discovered that the η^* distribution for the third highest p_T jet was depleted near $\eta^* = 0$ for QCD Zjj – in contrast to QCD $Zjjj$ matrix element calculations. This was due the treatment of color coherence in PYTHIA. This behavior is still visible in the full simulation as seen in the Figure 10.8 (right). Also in contrast to expectations, the signal is relatively enhanced near $\eta^* = 0$. The enhancement in the signal includes jets from PYTHIA's parton shower, underlying event, and electronic noise artifacts.

[2] Jets were required to have $p_T > 8$ GeV.

Because the underlying event has such high uncertainty and the η^* distribution is not trustworthy, we have moved the central jet veto to the end of the list of cuts for the full simulation analysis and do not include it in the expected significance calculation. In the future, Monte Carlo generators such as SHERPA may provide a consistent treatment of the QCD Z+jets background. Most likely, the underlying event, minimum bias interactions, etc. will need to be studied with ATLAS data before the central jet veto can be included into the analysis.

10.8 Background Determination from Data

Before one can claim the discovery of Higgs boson in this channel, we must thoroughly demonstrate that we understand the backgrounds and estimate our uncertainty. Monte Carlo simulations provide a direct prediction from quantum field theory to the distributions of measured quantities; however, the Monte Carlo methods have well known limitations. These limitations come from higher-order corrections to matrix element calculations, uncertainty in phenomenological models, and imperfect knowledge of the detector response. As a result, it is desirable to obtain a prediction of the background directly from the data.

Because the background is dominated by $Z \rightarrow \tau\tau$, a great deal of the background properties can be studied with $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$. In particular, the $\not{p}_T$ should vanish in these two channels, which allows one to study the $\not{p}_T$ resolution and any potential biases in the presence of forward tagging jets. Furthermore, the cross section drops rapidly as one increases the cut on $\Delta\eta_{jj}$. Except for non-trivial effects from the trigger, the $\not{p}_T$ performance and $\Delta\eta_{jj}$ shape should carry over to $Z \rightarrow \tau\tau$. It is anticipated that the shape of the irreducible background will be well understood from these studies.

The reducible $t\bar{t}$ and W+jets backgrounds can also be studied with data. One difficulty in estimating the reducible background is the fact that the jet rejection for a fixed tau efficiency depends on the physics process. This process-dependent rejection is due, in part, to the differences in fragmentation of light-quark, b-quark, and gluon initiated jets. Fortunately, these backgrounds are expected to be quite small, so a relatively large uncertainty is tolerable.

Between now and the turn on of the LHC, a large effort is needed to establish the details of how we will determine the background from data. The basic control samples and kinematic extrapolations have been established, but the impact of trigger biases and process-dependent tau-jet separation still need to be understood.

10.9 A Cut-Based Analysis with Fast Simulation

The cut analysis for $H \rightarrow \tau\tau \rightarrow lh\not{p}_T$ was outlined in Ref. [113] with a parton-level analysis including a reasonable estimate of the ATLAS detector performance. This analysis was revisited by the ATLAS collaboration once the necessary Monte Carlo generators were interfaced with showering and hadronization generators and ATLFAST [122, 123, 124]. The cuts used in [122] will be referred to as "Mazini Cuts". The cuts consist of basic visibility requirements, cuts to reduce backgrounds which mimic τ decays, and jet requirements to reduce the $Z \rightarrow \tau\tau$ background. The basic visibility requirements include the presence of high p_T leptons and the identification of a τ hadron. Fake tau backgrounds are suppressed with tau-jet separation cuts and by requiring consistency with the signal in the collinear approximation. Cuts on the transverse mass of the lepton and $\not{p}_T$ reduce $W+$jets backgrounds. The remaining cuts on jet activity take advantage of the two hard forward jets present in the signal and the suppression of hadronic activity between them.

10.9.1 Signal and Background Generation

The basic Matrix Element calculations used in Ref. [113] were interfaced to PYTHIA for the studies Ref. [122, 123, 124]. These interfaces were later improved with the MadCUP project. The difficulties in event unweighting outlined in Ref. [122] were not encountered during the Monte Carlo generation for the studies presented here.

In these studies the QCD and EW Zjj were generated with the MadCUP generator, interfaced to PYTHIA6.203, TAUOLA, and PHOTOS. The CTEQ5L parton distribution functions were used. The cross-section for these channels is sensitive to the minimum allowed $M_{\tau\tau}$ and p_T of the tagging jets. For the generation of these events it was required at the parton level that $|M_Z - M_{\tau\tau}| < 50$ GeV and that outgoing partons had $p_T > 20$ GeV. With these parton-level cuts, the cross-section

m_H (GeV)	110	120	130	140	150
$\sigma\, BR(H \to \tau\tau)$ (fb)	306	259	195	124	62.2

Table 10.1 Cross sections for the signal generated with PYTHIA6.203

times the branching ratio for $Z \to \tau\tau$ was 106.4 pb for QCD Zjj and 632 fb for EW Zjj. In previous studies, the W+jets background was found to be negligible after the cuts listed below. The $t\bar{t}$ background was generated with PYTHIA with one top decaying as $t \to Wb \to l\nu$ (with $l = e, \mu$) and the other without restriction. This decay configuration corresponds to a cross-section of 210 pb.

The $t\bar{t}$ background has been studied extensively within the context of VBF. Because of a b-jet veto, about 80% of the $t\bar{t}$ background has one tagging jet from a non-tagged b-jet and the other from a gluon-initiated jet. PYTHIA is not well suited to generate the hard forward gluon-initiated jet; instead, a matrix element calculation is more appropriate. Unfortunately, the p_T threshold of the tagging jet is low enough that the perturbative calculations are not trustworthy. It was demonstrated, after these studies were performed, that the generator MC@NLO is able to provide a consistent description of the $t\bar{t}$+jets background for VBF channels. Fortunately, the $t\bar{t}$ channel is not the dominant background for this channel, so the conclusions are robust against the expected 20% increase in $t\bar{t}$ background in the signal-like regions of phase-space.

The signal was generated with PYTHIA6.203, which predicts the cross-section in Table 10.1 when gluon-gluon fusion is not included in the generation.

10.9.2 List of Cuts

Some small changes were made to the cuts outlined by Mazini and Azuelos. First the veto on any jet with $p_T > 20\ \mathrm{GeV}$ between the two tagging jets was changed to the fixed range $|\eta| < 3.2$. This change was made so that estimates on the rate at which minimum bias events produce such jets – which were performed in the full simulation – could be included in the fast simulation [107]. Secondly, a b-jet veto was added to suppress the $t\bar{t}$ background further. Lastly, the rapidity gap

criterion was changed from $\Delta\eta_{jj} > 4.4$ to $\Delta\eta_{jj} > 3.8$, which provides a larger signal efficiency and slightly higher expected significance. The final set of cuts is as follows:

- At least one electron with $p_T^e > 25$ GeV or one muon with $p_T^\mu > 20$ GeV is required for the trigger.
- In order to achieve sufficient jet rejection, it is required that the hadron has $p_T^h > 40$ GeV.
- The forward tagging requirement includes the identification of two high p_T jets. The transverse momentum on the highest p_T jet must satisfy $p_{T_{j1}} \geq 40$ GeV and the second highest p_T jet must satisfy $p_{T_{j2}} \geq 20$ GeV. The jets are required to be in opposite hemispheres and separated by $\Delta\eta_{jj} \geq 3.8$.
- In order to avoid a singularity in the collinear approximation, the τ-decay products must be separated in azimuth such that $\cos\phi_{lh} > -0.9$. Consistency with the $H \rightarrow \tau\tau$ decay requires that $0 < x_l < 1.$ and $0 < x_h < 1.$ To improve the rejection against $W+$jets, the lepton requirement is increased to $x_l < 0.75$.
- A cut, $M_T(l,\not{p}_T) < 30$ GeV, on the transverse mass of the lepton and $\not{p}_T$strongly suppresses $W+$jets and $t\bar{t}$, where $M_T(l,\not{p}_T) = \sqrt{(|\not{p}_T| + |p_T^l|)^2 - (\not{p}_T + p_T^l)^2}$
- The neutrinos from the tau decays together with the requirement that the taus are not back-to-back provides significant $\not{p}_T$. Thus it is required that $\not{p}_T > 30$ GeV.
- The Zjj background can be further suppressed with the cut $m_{jj} > 700$ GeV.
- The electroweak nature of the signal suggests little jet activity between the tagging jets. The central jet veto rejects any event with a jet of $p_{Tj}^{veto} > 20$ GeV and $|\eta_j^{veto}| < 3.2$.
- The invariant mass of the two taus provides the final discrimination between Higgs and $Z \rightarrow \tau\tau$ backgrounds. The significance calculated from the Poisson distributions of the number of expected events, denoted σ_P, only includes events in the range $m_H - 10 < M_{\tau\tau} < m_H + 15$. This requirement is removed when calculating the significance with likelihood techniques, which is denoted as σ_L.

m_H GeV	Signal	QCD Zjj ($l = e$)	QCD Zjj ($l = \mu$)	EW Zjj	$t\bar{t}$	σ_P	σ_L
110	20.64	9.096	12.573	3.957	.2349	3.50	4.76
120	18.01	2.263	3.579	1.377	.1788	4.94	5.99
130	13.72	1.138	1.468	.6133	.1484	5.03	6.03
140	8.635	.6917	.6757	.3763	.0915	3.97	4.56
150	4.085	.4698	.3923	.1864	.3453	2.20	3.14

Table 10.2 Expected number of signal events, background events, and significance with 30 fb^{-1} for various masses.

10.9.3 Results with Fast Simulation

The number of signal and background events after all cuts and for various hypothetical Higgs masses is shown in Table 10.2. Using $M_{\tau\tau}$ as a discriminating variable improves the significance of this channel because it takes advantage of the shape of the mass peak. The Poisson and likelihood based significance calculations are also shown in Table 10.2. A comparison of the significance with Mazini's cuts, these new cuts, and the likelihood approach are shown in Figure 10.9.

The determination of the background will be performed with data using independent control samples. As a result, the uncertainty on the background is related to the uncertainty in the extrapolation of those control samples to the signal-like region and the statistical fluctuations in that control sample. If we assume a conservative 10% background uncertainty (as was done in Ref. [124]) and use the Cousins-Highland formalism for incorporating that uncertainty into the significance calculation (see Appendix C), then the significance for $m_H = 130$ GeV drops from 5.03 to 4.89. Clearly, the high s/b affords this channel a great deal of robustness against the uncertainty in the normalization of background. On the other hand, these channels are more sensitive to uncertainty in the shape of the background, especially for lower masses. A more detailed discussion of the background determination can be found in Section 10.8.

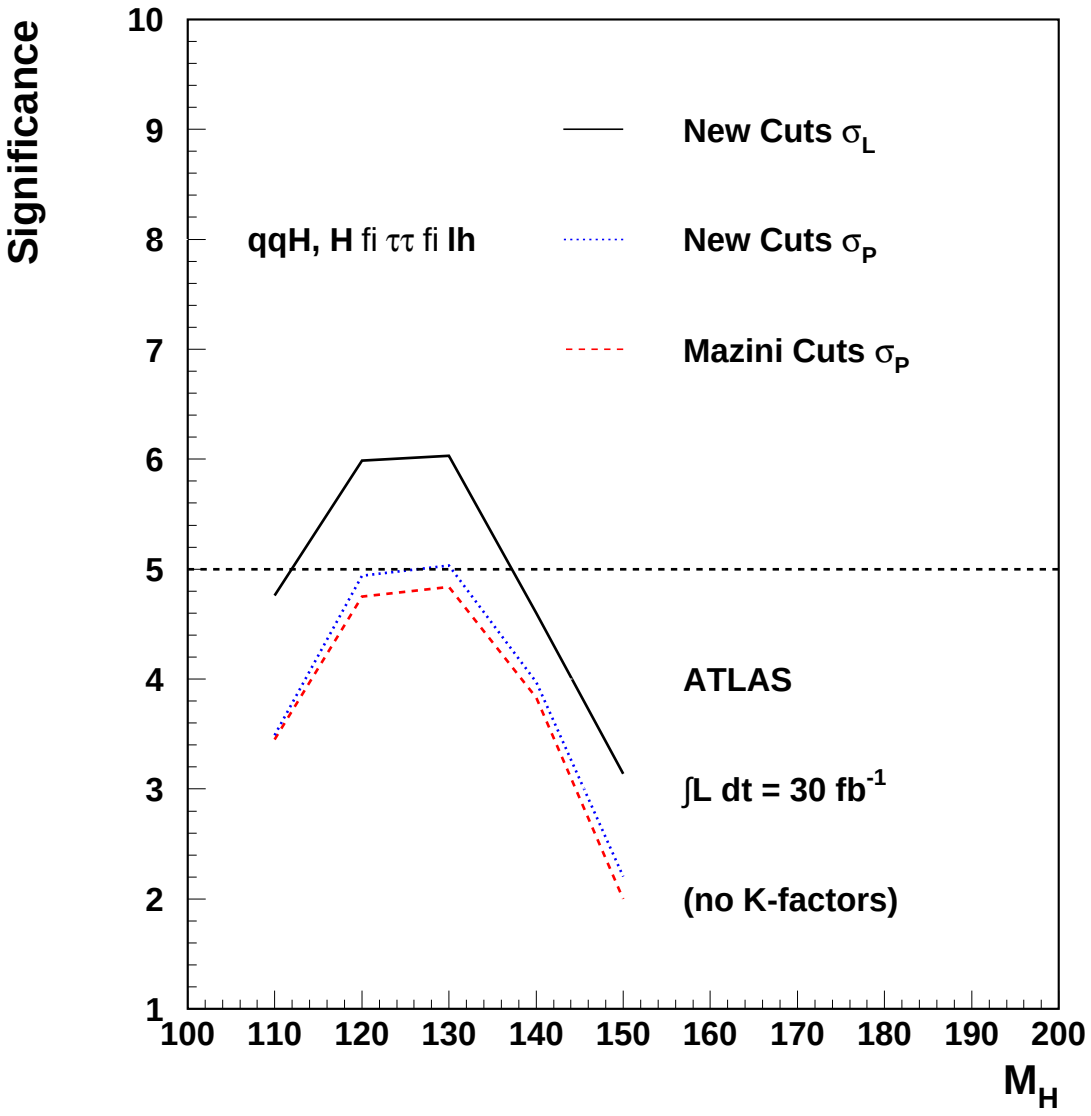


Figure 10.9 Expected Significance for several analysis strategies with 30 $\mathrm{fb}^{-1}$with fast simulation.

With 30 fb^{-1} one can expect a 5σ discovery if the Higgs mass is between 120 and 130 GeV. By using the $M_{\tau\tau}$ as a discriminating variable, the expected significance is enhanced by 20%.

10.10 A Cut-Based Analysis with Full Simulation

In order to verify the results from fast simulation, the previous analysis was repeated with full simulation. Because of the outstanding issues with matrix element-parton shower matching and underlying event discussed in Section 10.7, the central jet veto was removed from the list of cuts. Also, the improved tau-jet separation means that for the same jet rejection, one can achieve a higher tau efficiency. The working point chosen for these studies was a log-likelihood ratio greater than 1, which corresponds to $\sim 70\%$ efficiency for $p_T > 40$ GeV. As will be demonstrated, the most significant differences between the fast and full simulation results are due to a degraded $\not{p}_T$ resolution in the full simulation.

10.10.1 Signal and Background Generation

The signal and background were generated in a way similar to ATLAS's Data Challenge 2. The full event generation includes

- Parton shower and hadronization with the PYTHIAimplemented in offline release 8.0.7, including the Data Challenge 2 tunings of the underlying event, TAUOLA, and PHOTOS.
- A filter that requires at least one electron or muon within $|\eta| < 3.2$ and $p_T > 20$ GeVis applied at the particle-level.
- GEANT4 simulation of the ATLAS detector with offline release 8.0.7.
- Realistic electronic noise included in the *digitization* of the events with offline release 8.0.7.
- Reconstruction with offline relase 9.0.0 and a modified `MissingET` package that provides the `G4Beta2` H1-style hadronic calibration weights and the local noise suppression described in Section 9.3.4.
- Event Summary Data (ESD) and Analysis Object Data (AOD) production with offline release 9.0.0 (see Appendix F).

Because this channel's background contribution is dominated by QCD Zjj and the computational resources for GEANT4 simulation are formidable, only the QCD Zjj and signal were simulated. Furthermore, to expedite the generation of the background, stringent parton-level cuts were placed on the background. It was required that the outgoing quarks have $p_T > 18$ GeV, the minimum separation between them to be $\Delta\eta_{jj} > 3.6$, and their invariant mass be $M_{jj} > 650$ GeV. It was also required that the τ's from the Higgs have $p_T > 18$ GeV. This corresponds to a cross-section of 2928 fb. Just over 340,000 events (117 fb^{-1}) were simulated. In order to account for the contribution of EW Zjj – as predicted from the fast simulation studies – the QCD Zjj cross-section has been increased by 25% in the results reported in the next section.

The signal was only generated for $m_H = 130\ \mathrm{GeV}$. Using HDECAY the cross-section for VBF $H \rightarrow \tau\tau$ at this mass is predicted to be 214 fb. Just over 72,000 events were simulated, corresponding to 340 fb^{-1}.

10.10.2 Results with Full Simulation

The effective cross section for the signal and Zjj background are shown in Table 10.3 along with the efficiency of each cut. Because of the degraded $\not{p}_T$ resolution, the mass window was widened from $120\ \mathrm{GeV} < M_{\tau\tau} < 145\ \mathrm{GeV}$ to $110\ \mathrm{GeV} < M_{\tau\tau} < 150\ \mathrm{GeV}$. Even with this wider mass window, the signal efficiency in this full simulation study is only 75% of what was predicted from fast simulation studies. The requirement that $0 < x_{\tau l} < 0.75$ and $0 < x_{\tau h} < 1$ is sensitive to the degraded $\not{p}_T$ resolution and is the cut with the largest discrepancy between full and fast simulation. In the fast simulation 65% of the signal survived the cuts labeled "collinear approximation", while in the full simulation studies only 38% survives. This low efficiency motivates the maximum likelihood approach outlined in Section 10.6.2.

In addition to a lower signal efficiency, the tails of the $M_{\tau\tau}$ distribution are worse in the full simulation studies. This can be understood as another artifact of degraded $\not{p}_T$ performance.

The distribution of $M_{\tau\tau}$ for signal and Zjj background is shown in Figure 10.10 using the Monte Carlo truth $\not{p}_T$. In that figure, a convincing signal is seen well-separated from the Zjj background. The effect of the 14 GeV $\not{p}_T$ resolution is responsible for the difference between

cut	signal (fb)	ϵ %	Zjj (fb)	ϵ %
after trigger	38.0	17.8	602	16.5
after hadronic τ ID	7.78	20.3	92.5	16.2
after forward Jet Tagging	2.19	28.1	28.6	31.0
after collinear approximation *	0.83	38.2	9.83	34.3
after $M_T(l, P_t^{miss})$ cut	0.59	70.9	8.19	83.3
after P_T^{miss} cut	0.46	77.5	6.44	78.7
after M_{jj} cut	0.38	82.6	5.63	87.5
after $110 < M_{\tau\tau} < 150$	0.34	89.8	0.37	6.5
after Central Jet Veto	0.18	53.0	0.16	42.9

Table 10.3 Signal and Background effective cross-sections after various cuts for $m_H = 130$ GeVwith full simulation. The QCD Zjj background has been scaled by 1.25 to account for the final Electroweak component from fast simulation.

Figure 10.10 and Figure 10.11. If one simply counts the number of events in the mass window the expected significance of this channel is $\sigma_P = 2.4$. If one includes the information in the shape of the $M_{\tau\tau}$ distribution, then the expected significance is $\sigma_L = 3.6$. If one employs the maximum likelihood technique in the collinear approximation, the signal efficiency can be improved by about 25%, and the expected significance is $\sigma_L = 4.2$.

These preliminary results with full simulation do not confirm the results of the fast simulation; however, they do support the conclusion that VBF $H \rightarrow \tau\tau$ as a very powerful channel near the LEP limit. It is clear that the dominant experimental issue is the performance of $\not{p}_T$, which impacts both signal efficiency and invariant mass resolution. The maximum likelihood approach to the collinear approximation does help to recover some lost signal, but it does not improve the mass resolution. The background determination from data and improvements to $\not{p}_T$ are the areas that need the most attention in the coming years.

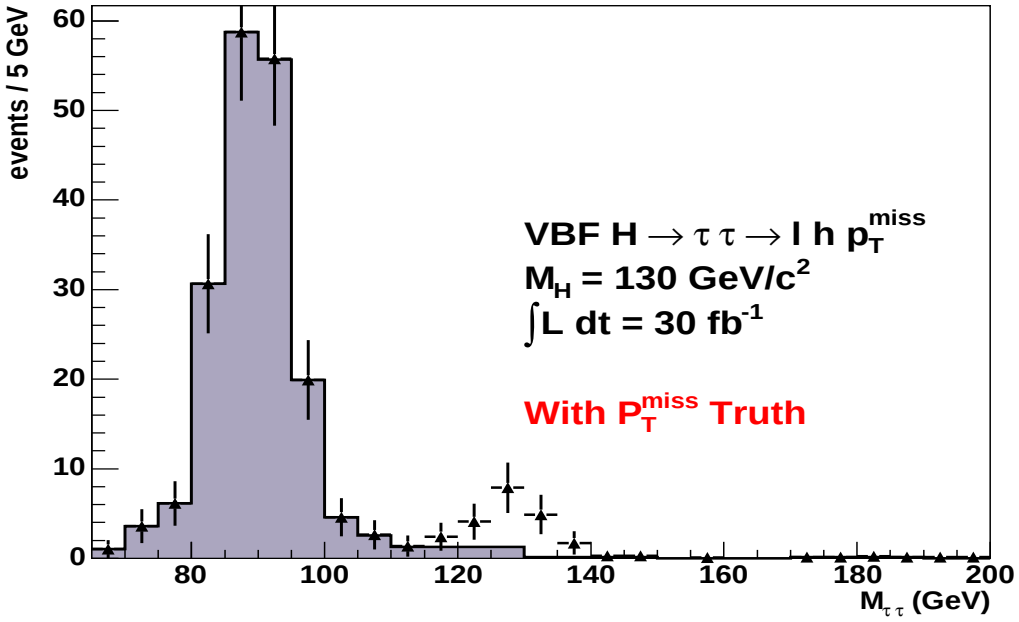


Figure 10.10 $M_{\tau\tau}$ distribution for 30 fb^{-1} obtained with truth $\not{p}_T$.

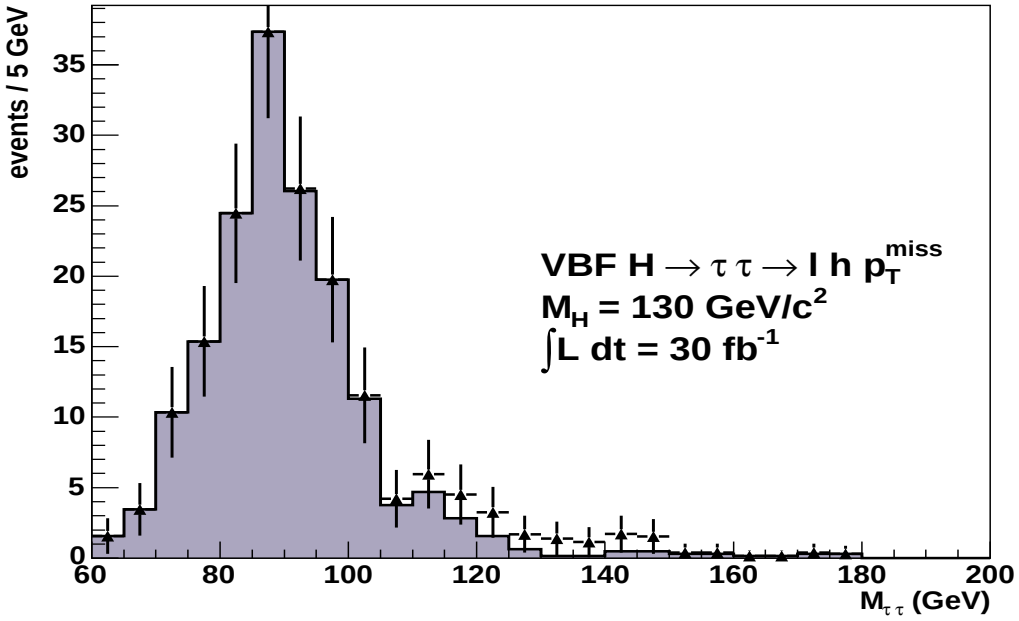


Figure 10.11 Expected $M_{\tau\tau}$ distribution for 30 fb^{-1} obtained with fully reconstructed jets, leptons, and a $\not{p}_T$ calculation with local noise suppression.

Chapter 11

Comparison of Multivariate Techniques for VBF $H \to WW^*$

In this chapter we consider a the potential for multivariate analysis in the search for the Higgs boson at the LHC. While there are many channels available, the recent Vector Boson Fusion analyses offer sufficiently complicated final states to warrant the use of multivariate algorithms. The decay channel chosen is $H \to W^+W^- \to e^{\pm}\mu^{\mp}\nu\overline{\nu},\ e^+e^-\nu\overline{\nu},\ \mu^+\mu^-\nu\overline{\nu}$. These channels will also be referred to as $e\mu$, ee, and $\mu\mu$, respectively.

Originally, these analyses were performed at the parton level and indicated that this process could be the most powerful discovery mode at the LHC in the range of the Higgs mass, m_H, $115 < m_H < 200\,\mathrm{GeV}$ [112]. These analyses were studied specifically in the ATLFAST environment using a fast simulation of the detector [106]. Two traditional cut analyses, one for a broad mass range and one optimized for a low-mass Higgs, were developed and documented in References [126] and [124].

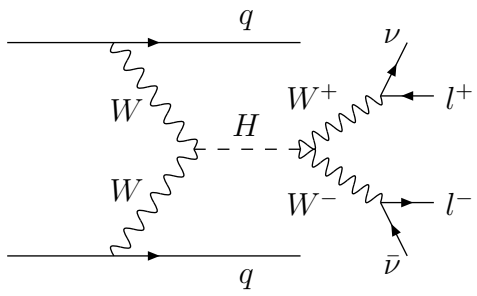


Figure 11.1 Tree-level diagram of Vector Boson Fusion Higgs production with $H \to W^+W^- \to l^+l^-\nu\overline{\nu}$

Process	$e\mu$ $\sigma_{\rm eff}$ (fb)	ee $\sigma_{\rm eff}$ (fb)	$\mu\mu$ $\sigma_{\rm eff}$ (fb)
$t\bar{t}$	12.79	4.75	5.22
EW $WW + \rm jets$	1.05	0.39	0.50
QCD $WW + \rm jets$	1.56	0.52	0.61
EW $Z + \rm jets$	0.12	0.04	0.07
QCD $\gamma^*/Z + \rm jets$	5.40	2.22	2.70

Table 11.1 Effective cross-section by channel for each background processes after preselection.

Figure 11.1 illustrates the complexity of the final state for which we search. Angular correlations between the W decay products impact all variables derived from leptons and the missing transverse momentum. These relationships simultaneously make the analysis challenging and provide the handles with which to reject background. Furthermore, these relationships invite multivariate techniques capable of exploiting correlations among a number of variables.

In total, three multivariate analyses were performed:

- a Neural Network analysis using back-propagation with momentum,
- a Support Vector Regression analysis using Radial Basis Functions, and
- a Genetic Programming analysis using the software described in Appendix E.

Each of the analyses used the same variables, preselection, background generation, and ATLFAST simulation documented in Ref. [127]. In Section 11.2 we describe briefly the neural network analysis presented in Ref. [127]. In Section 11.3 and Section 11.4 we describe the Support Vector Regression and Genetic Programming analyses, respectively. Finally, in Section 11.5 we compare the three methods.

11.1 Variables

Each of the three analyses used the same seven input variables. The approach of the original Neural Network study was to present a multivariate analysis comparable to the cut analysis presented in [126]. Thus, the analysis was restricted to kinematic variables which were used or can be derived from the variables used in the cut analysis.

The variables used were:

- $\Delta\eta_{ll}$ - the pseudorapidity difference between the two leptons,
- $\Delta\phi_{ll}$ - the azimuthal angle difference between the two leptons,
- M_{ll} - the invariant mass of the two leptons,
- $\Delta\eta_{jj}$ - the pseudorapidity difference between the two tagging jets,
- $\Delta\phi_{jj}$ - the azimuthal angle difference between the two tagging jets,
- M_{jj} - the invariant mass of the two tagging jets, and
- M_T - the transverse mass.

The transverse mass is defined as

$$M_T = \sqrt{\left(E_T^{ll} + E_T^{\nu\bar{\nu}}\right)^2 - \left(\overrightarrow{P}_T^{ll} + \overrightarrow{\not{p}_T}\right)^2} \tag{11.1}$$

where

$$E_T^{ll} = \sqrt{\left(\overrightarrow{P}_T^{ll}\right)^2 + M_{ll}^2} \qquad \text{and} \qquad E_T^{\nu\bar{\nu}} = \sqrt{\left(\overrightarrow{\not{p}_T}\right)^2 + M_{ll}^2}. \tag{11.2}$$

11.2 Neural Network Analysis

After thorough testing of the options provided in and SNNS [128] we found back-propagation with momentum[1] to be an efficient algorithm – in agreement with our groups previous experience [129, 130, 131]. In an independent analysis, we used the MLPFIT [132] package with

[1] A learning parameter $\eta = 0.01$ and momentum term $\mu = 0.01$ were used

the Broyden-Fletcher-Goldfarb-Shanno (BFGS) learning method [133]. The two methods agreed within the expected and variability of different training runs [127]. The results from the MLP analysis are reported below.

Figure 11.5 (right) illustrates the discriminating power of a neural network output trained for the $e\mu$ channel with a Higgs mass between $115 - 130\ \mathrm{GeV}$. In each case, the signal is concentrated near 1, while the background peaks near 0.

11.2.1 Stability of Results to Different Background Descriptions

We now turn briefly to the stability of the neural networks with respect to theoretical uncertainties in the Monte Carlo. We have considered two different parton shower models and two different matrix elements for the $t\bar{t}$ background. The identical neural network and cut analysis were used to estimate the effective cross-section for the different $t\bar{t}$ background samples[2]. In contrast to leading order uncertainties in the cross-sections, the parton shower uncertainties do not necessarily apply equally to the cut and neural network analysis.

In order to estimate the sensitivity to the parton shower model, we have used PYTHIA and HERWIG interfaced to an external matrix element calculation provided by the MadCUP project [114]. The use of a common external matrix element sample allows for the isolation of the systematic uncertainty due to the parton shower model.

The use of the neural network output as a discriminating variable is contingent upon the stability in its shape. Figure 11.2 illustrates the stability of the neural network output for the three different $t\bar{t}$ background samples considered. While the effective cross-sections for the three samples differ significantly, the shape of the neural network output appears to be quite stable.

11.3 Support Vector Regression

Support Vector Regression is a relatively new multivariate analysis technique, which has become quite popular among computer scientists due to its nice theoretical properties. The machine

[2]The 20% increase in the $t\bar{t}$ cross-section used in Reference [126] to account for finite width effects has been neglected in this section.

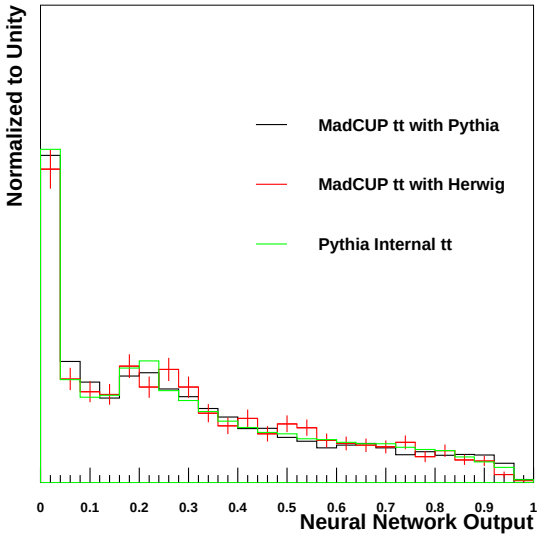


Figure 11.2 Neural Network output distribution for three different $t\bar{t}$ background samples.

learning formalism that it is based on is discussed in Appendix D. By using Support Vector Regression (SVR) instead of Support Vector Classification (SVC), each event is assigned a regression coefficient that is similar to a neural network output. It is often pointed out to physicists that use SVR that SVC offers a superior classification. It should be pointed out that we are not directly interested in a rate of mis-classification, but instead are interested in statistical sensitivity. The statement that SVC is superior to SVR is only true if one does not use the information in the shape of the regression distribution. Furthermore, it is much more practical to optimize the critical boundary between signal and background by cutting on the regression coefficient.

For the Support Vector Regression (SVR) analysis, the BSVM-2.0 [134] library was used. The only parameters are the cost parameter (set to $C = 1000$), the kernel size (set to the default $1/N_{var}$), and the kernel function. BSVM does not support weighted events, so an "unweighted" signal and background sample was used for training. Because the trained machine only depends on a small subset of "Support Vectors", performance is fairly stable after only a thousand or so training samples. In this case, 2000 signal and 2000 background training events were used.

11.4 Genetic Programming

The Genetic Programming approach is a novel multivariate algorithm developed by the author and R. Sean Bowman. It is documented in Appendix E and in Ref. [135].

For this analysis only one island was used with an initial population of 400 individuals. The selection pressure, α, was set to 1.5; the probability that an individual experienced a mutation was 20%; and the probability that an individual performed a cross-over with another individual was 60%.

11.5 Comparison of Multivariate Methods

Ref. [127] summarizes the sensitivity of the neural network based analyses in the Higgs mass range $115 - 130\ \mathrm{GeV}$ for the $e\mu$, ee, and $\mu\mu$ channels with 30 fb^{-1} of data. Combined significance values are obtained from likelihood ratio techniques described in Appendix A and Ref. [136].

Figure 11.3 also shows improvements to the sensitivity of this channel from the use of neural networks, the use of discriminating variables, and the combined improvement. Each of the three improved analyses are compared to the cut-based analysis with number counting (Poisson statistics). The black (solid) line shows the improvement due to the use of the likelihood ratio with the transverse mass as a discriminant variable. The red (dashed) line shows the improvement due to the use of a neural network for event selection. The green (dotted) line shows the improvement due to the use of the likelihood ratio with the neural network output as a discriminant variable. The neural network based analysis without the use of discriminating variables achieves a 20-40% improvement over the cut-based analysis. This is due to the exploitation of correlations between the variables (recall, the same variables are used in both analyses). Furthermore, the use of discriminating variables in the confidence level calculation improves the significance by about an additional 15%.

Because the Genetic Programming analysis was configured to optimize the Poisson significance (which does not use any shape information), it is not possible to compare the significance

	Ref. Cuts	low-m_H Opt. Cuts	NN	GP	SVR
120 ee	0.87	1.25	1.72	1.66	1.44
120 $e\mu$	2.30	2.97	3.92	3.60	3.33
120 $\mu\mu$	1.16	1.71	2.28	2.26	2.08
Combined	2.97	3.91	4.98	4.57	4.26
130 $e\mu$	4.94	6.14	7.55	7.22	6.59

Table 11.2 Expected significance for two cut analyses and three multivariate analyses for different Higgs masses and final state topologies.

among these methods when the significance calculation takes advantage of discriminating variables.

Both NN and SVR methods produce a function which characterizes the signal-likeness of an event. A separate procedure is used to find the optimal cut on this function which optimizes the Poisson significance, σ_P. Fig. 11.5 shows the distribution of the SVR (left) and NN (right) output values. The optimal cut for the SVR technique is shown as a vertical arrow.

Tab. 11.5 compares the Poisson significance, σ_P, for a set of reference cuts, a set of cuts specifically optimized for low-mass Higgs, Neural Networks, Genetic Programming, and Support Vector Regression. It is very pleasing to see that the multivariate techniques achieve similar results. Each of the methods has its own set of advantages and disadvantages, but taken together the methods are quite complementary.

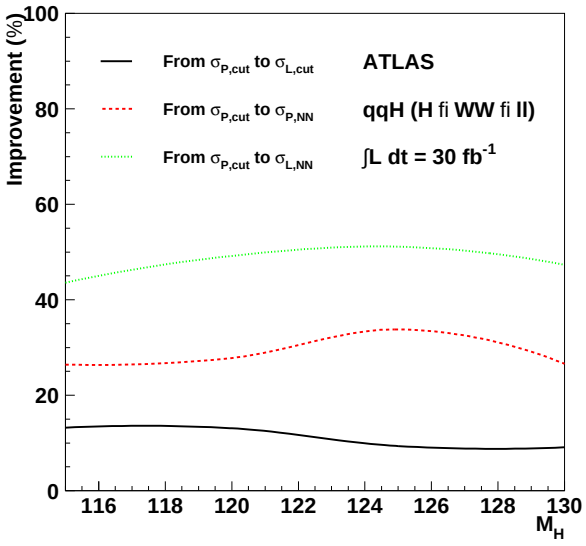


Figure 11.3 The improvement in the combined significance for VBF $H \rightarrow WW$ as a function of the Higgs mass, m_H.

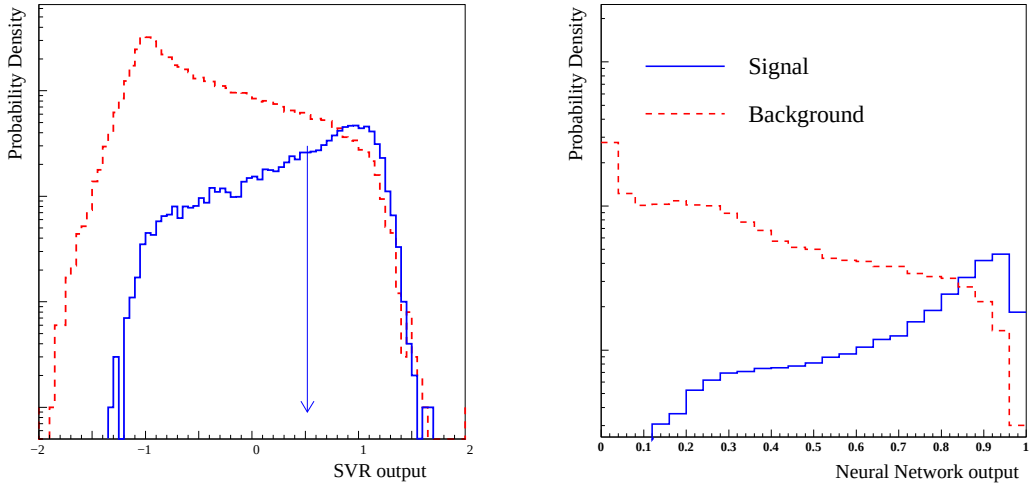


Figure 11.4 Support Vector Regression and Neural Network output distributions for signal and background for 130 GeV Higgs boson in the $e\mu$ channel.

Chapter 12

$H \rightarrow \gamma\gamma$ Coverage Studies

If at first it doesn't fit, fit, fit again. –John McPhee

12.1 Systematics for $H \rightarrow \gamma\gamma$

The inclusive $H \rightarrow \gamma\gamma$ analysis has a huge continuum background with a simple shape. The strategy for this channel has been to use the sidebands to extract the number of expected background events in the signal-like region. The fit takes into account both the electromagnetic energy-scale uncertainty and the cross-section uncertainty. Given this channel's low s/b, the uncertainty on the background must be less than about 0.2% for it to be a discovery channel. Using the sideband technique, the uncertainty on the expected background is expected to be negligible. As we will show below, the uncertainty is not negligible, but the method is robust against uniform energy scale uncertainties.

The use of a fitted background as a substitute for a true prediction could be substantiated if the fits to both Monte Carlo background events and data provide an acceptable χ^2. If the χ^2 is not acceptable, then the parametric form either needs to be rejected or extended. Extending the parametric form of the continuum background will most likely result in a higher uncertainty in the prediction of background events in the signal-like region. Given the low tolerance of the $H \rightarrow \gamma\gamma$ analysis to background uncertainty and the huge expected background, it needs to be confirmed that an acceptable parametric form and small background uncertainty can be achieved simultaneously.

As a first step in this direction, a coverage study for $H \rightarrow \gamma\gamma$ has been performed. A toy Monte Carlo was used to generate a number of experiments with a $M_{\gamma\gamma}$ spectrum given by a

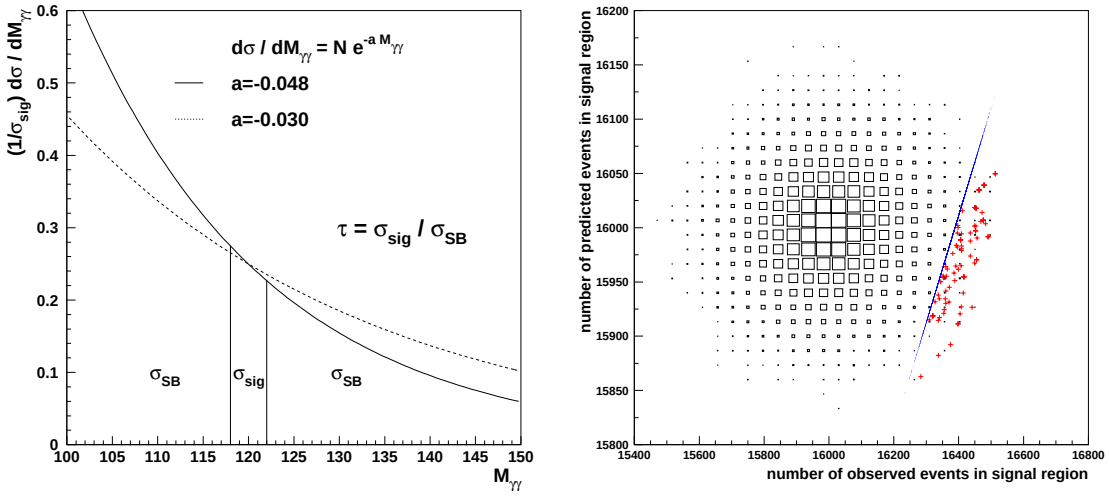


Figure 12.1 Left: exponential form used for Toy Monte Carlo. Right: observed number of events in the signal-like region vs. predicted number of events from fit to sideband. The red points represent experiments considered as 3σ discoveries.

simple exponential form (see Figure 12.1). The exponential was normalized such that an average of 16000 events were generated in the mass window $118 < M_{\gamma\gamma} < 122$ GeV and sampled in the range 100-150 GeV. We then varied the exponent in the range $[-0.048, -0.030]$ GeV^{-1} in steps of 0.002 GeV^{-1}. In total, nearly a million MINUIT fits were performed.

We arrived at several interesting results. First, we found that a *modified* least squares fit to a binned $M_{\gamma\gamma}$ spectrum leads to a predicted background that is biased by about 10 events (which is about a 0.08σ effect). Second, we found that the variation in predicted background events, δb, from the true value was about 38 events across the range of exponents tested. Because the same exponential form was used to generate the toy Monte Carlo, it is not surprising that the background uncertainty is exactly what one would expect from the number of events in the side band region. For convenience let us use τ to denote the ratio of the cross section in the signal like region, σ_{sig}, to the cross section in the sidebands, σ_{SB}. Thus the background uncertainty due to the statistical fluctuations in the sideband region is given by $\delta b \approx \tau\sqrt{N_{SB}}$.

By applying the Cousins-Highland formalism in the case that the background uncertainty given by $\delta b \approx \tau\sqrt{N_{SB}}$ and $b = \tau N_{SB}$, one arrives at the following result:

$$\sigma^{CH} = \frac{s}{\sqrt{b(1+\alpha^2 b)}} = \frac{s}{\sqrt{b(1+\tau)}} \tag{12.1}$$

As a caveat to Section C.3, when the background uncertainty is dominated by statistical error in an axillary experiment, the relative error α is reduced with luminosity; and the saturation of significance does not occur.

12.2 Frequentist Result

Let M be the expected number of background events in the signal like region extrapolated from some sideband measurement. Let x be the number of observed events in the signal like region. Let τ be the ratio of the number of events expected in the side-band, N_{SB} to the number of (background) events expected in the signal like region, M (*viz.* $\tau = M/N_{SB}$).

In the case that the background is a smooth distribution with some assumed parametric form, we can fit the side-bands. Typically the relative error on the fitted parameters will be $1/\sqrt{N_{SB}}$, thus the variation of M will be $\tau\sqrt{N_{SB}}$. Additionally, the Poisson fluctuations of x predict a variation in x to be $\sqrt{M}$, for large M.

For each value of the parameters of H_0 there is a distribution $L(x, M|H_0, b)$. If we can find a region W with the property

$$\int_W L(x, M|H_0, b)dxdM = 1 - \alpha, \tag{12.2}$$

for every value of the nuisance parameter b, then we have a similar test which should provide the correct coverage. For W of the form $W = \{x, M|x < M + \eta\sqrt{M}\}$, the challenge is to find the η which satisfies Equation 12.2. If we write the boundary as a function $x(M) = M + \eta\sqrt{M}$, and expand it about M_0, then the linear form of the boundary is

$$x(M) \approx M + \eta\sqrt{M_0} + (M - M_0)\underbrace{\left(1 + \frac{\eta}{2\sqrt{M_0}}\right)}_{m^{-1}} \tag{12.3}$$

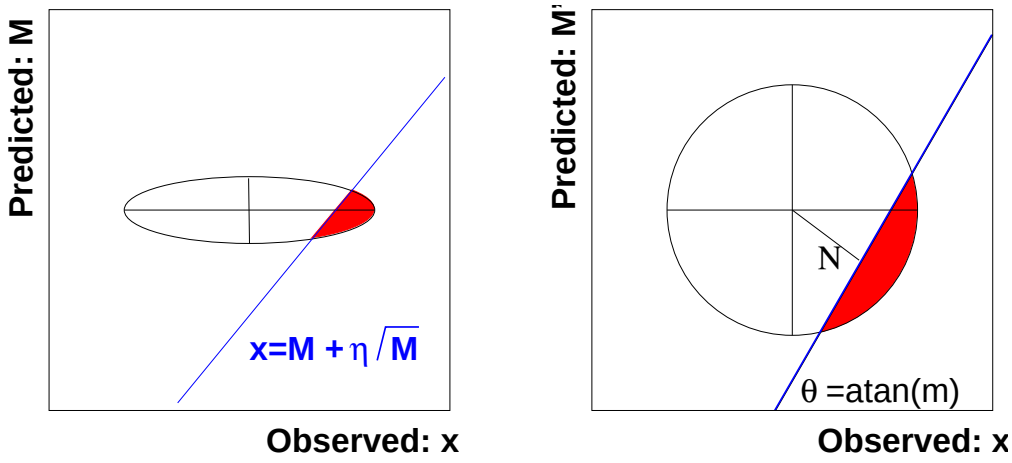


Figure 12.2 Determination of η via a change of variables.

Considering contours of $L(x, M|H_0, b)$ as ellipses with eccentricity

$$\epsilon = \frac{\Delta M}{\Delta x} = \frac{\tau\sqrt{N_{SB}}}{\sqrt{M}} = \sqrt{\tau} \tag{12.4}$$

and the critical boundary as lines with slope $m = ((1+\frac{\eta}{2\sqrt{M_0}})^{-1/2}$, the goal is to find η that satisfies Equation 12.2. By a change of variables $M' = M/\epsilon$, the contours of $L(x, M|H_0, b)$ become circles and the critical boundary has slope $m' = m/\epsilon$. In this new space, the coverage requirement is satisfied if the perpendicular bisector, has a length N (in number of Gaussian σ) that corresponds to α. Here we have $\theta = \tan^{-1} m/\epsilon$ and $\eta = N/\sin\theta$. Note, the x-direction was not modified by the change of coordinates. We can re-write $\eta = N/\sin(\tan^{-1}(m/\epsilon)) = N\sqrt{1+\epsilon^2/m^2)}$.

In the case $m = 1$, we recover the Cousins-Highland result $\eta = N\sqrt{1+\epsilon^2}$.

We have also derived a fully frequentist result applicable when both the number of predicted and observed events are very large

$$\sigma^F = \underbrace{\frac{s}{\sqrt{b(1+\tau/m^2)}}}_{\text{new result}} \quad \text{where} \quad m = \left(1 + \frac{\sigma_0}{2\sqrt{b}}\right)^{-1}, \tag{12.5}$$

where σ_0 is the desired significance of the test. The quantity m, which is less than unity, can be seen as a correction to the Cousins-Highland result. In the case of $H \to \gamma\gamma$, the correction is minuscule, and the Cousins-Highland result is an excellent approximation.

12.3 Impact of Systematics

Let us examine the impact of background uncertainty on the $H \rightarrow \gamma\gamma$ significance. If we assume that the background uncertainty is negligible, then for $M_H = 120$GeV, $\sigma = s/\sqrt{b} = 3.2\sigma$. In our studies, the sideband region ranged from 100-150 GeV (which is quite large), thus $\tau = 8\%$ and $\sigma^{CH} = 3.1\sigma$. However, if we use the sideband region found in the TDR which ranged from 105-135 GeV, $\tau = 25\%$ and $\sigma^{CH} = 2.9\sigma$.

In Table 12.1, we show the probability to claim a 3σ discovery given background-only experiments for several methods. The chance for such a discovery should be 0.135% via Equation A.3. The label BINNED refers to the scenario in which the background is estimated from a χ^2 fit to a binned $M_{\gamma\gamma}$ spectrum (which causes a bias in the expected number of events). The label UNBINNED refers to the unbinned extended likelihood fit procedure, for which no bias was observed. In both the BINNED and UNBINNED cases, an experiment was classified as a discovery if the number of observed events was greater than $b+3\sqrt{b}$. The labels COUSINS-HIGHLAND and FREQUENTIST refer to the unbinned case when discovery was claimed with Equation 12.1 and Equation 12.5, respectively. The statistical error on the entries is approximately 10% for each exponent, and 3% for the sum over all exponents.

12.4 Statement on Original Material

The work in this Chapter was motivated by discussions with Stathes Paganis, and he is responsible for the framework that produced the toy Monte Carlo and performed MINUIT fits. The author is responsible for assessing the methods in terms of coverage, the results shown in figures and tables, and the frequentist approach in Section 12.2.

Exponent	BINNED	UNBINNED	COUSINS-HIGHLAND	FREQUENTIST	N_{exper}
-0.030	0.18%	0.12%	0.08%	0.08%	83726
-0.032	0.19%	0.15%	0.10%	0.10%	46835
-0.034	0.38%	0.33%	0.27%	0.27%	84548
-0.036	0.19%	0.18%	0.16%	0.16%	84628
-0.038	0.21%	0.17%	0.11%	0.11%	84294
-0.040	0.27%	0.23%	0.16%	0.16%	90020
-0.042	0.22%	0.18%	0.10%	0.10%	90020
-0.044	0.21%	0.16%	0.10%	0.10%	90020
-0.046	0.32%	0.27%	0.15%	0.15%	85630
-0.048	0.23%	0.19%	0.12%	0.12%	78514
all	0.25%	0.21%	0.14%	0.14%	818235

Table 12.1 Results of the $H \to \gamma\gamma$ coverage study (see text).

Chapter 13

ATLAS Sensitivity to Standard Model Higgs

In this chapter we present an assessment of the sensitivity of the ATLAS detector to the Standard Model Higgs boson based on the low-mass Higgs studies reported in Ref [124] and the statistical procedures outlined in Appendices A and C. The majority of the analyses considered have been developed with ATLFAST, with the most relevant aspects studied, to varying degrees, in the full simulation. In the coming years, the main focus of the ATLAS collaboration on physics analysis will be on confirming the potential of these analyses in the full simulation, prioritizing the physics program for early discovery, and outlining a detailed physics commissioning schedule.

13.1 Channels Considered

For the combinations presented below, we use the results of a the recent ATLAS scientific note [124]. We have not used more recent results intentionally, in order to focus attention on the combination procedure. For completeness we provide a brief description and relevant references to the channels considered in this combination.

- Vector Boson Fusion $H \rightarrow WW^{(*)}$: This is the dominant discovery channel across most of the mass range considered because of its large cross-section (relative to other VBF processes) and its high signal-to-background ratio. However, because of the presence of two neutrinos in the final state, bona fide mass reconstruction is not possible. We consider a 10% systematic uncertainty on the background normalization [124]. Due to the complementarity of the event selection for this channel and the inclusive $H \rightarrow WW^{(*)} \rightarrow ll\not{p}_T$, the overlap of events selected by the two channels has been found to be negligible.

- Vector Boson Fusion $H \rightarrow \tau^+\tau^-$: This is a very powerful channel for masses around 110-140 GeV. In contrast to the VBF $H \rightarrow WW$ channel, mass reconstruction can be performed here with a mass resolution on the order of 10%. We consider a 10% systematic uncertainty on the background normalization.

- $t\bar{t}H(H \rightarrow b\bar{b})$: This analysis is very important near the LEP exclusion limit where the branching ratio of $H \rightarrow bb$ is very large. Based on reference [124], we apply a uniform 5% uncertainty on this channel when calculating the expected significance, and not the larger systematic found in [137]. If the results in Ref. [137] are used, it is impossible to reach a 5σ significance level with 10% background uncertainty due to the relatively low signal-to-background ratio (see Figure C.4).

- $H \rightarrow \gamma\gamma$: This analysis requires excellent understanding of the electromagnetic energy scale and a low systematic error on the background due to the very low signal-to-background ratio. The systematic error in this channel is considered to be negligible.

- $H \rightarrow WW^* \rightarrow l\nu l\nu$: This analysis is complementary to the $H \rightarrow ZZ^* \rightarrow 4l$ analysis near a Higgs mass of 170 GeV. In contrast to the Vector Boson Fusion analysis, the production mode for the inclusive analysis is dominated by gluon-gluon fusion. The complementary jet requirements are responsible for removing potential overlap with the Vector Boson Fusion analysis. We use the a 5% systematic uncertainty on the background normalization.

- $H \rightarrow ZZ^{(*)} \rightarrow 4l$: Sometimes referred to as the "golden channel", this channel has been the dominant discovery mode for ATLAS across a very large range of masses. Though it no longer has the highest expected significance, this channel offers a stunning mass peak and will be pivotal to the discovery of a Higgs with a mass above 200 GeV. No systematic uncertainty in the background was included.

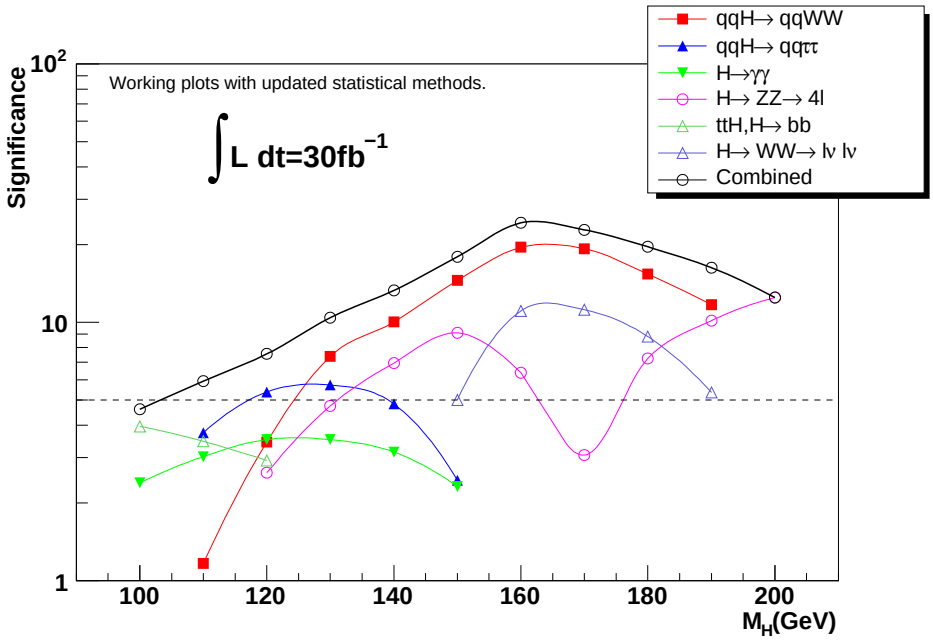


Figure 13.1 Individual and combined significance versus the Higgs mass hypothesis.

13.2 Combined Significance

In this section we present the combined significance of the ATLAS detector (see Figure 13.1). These combinations were made with a consistent treatment which uses the likelihood ratio as a test statistic and the Cousins-Highland formalism to incorporate systematic errors. The combination corresponds to 30 fb^{-1} of integrated luminosity.

The combined significance for 30 fb^{-1} of integrated luminosity is expected to be above 5σ for $m_H \gtrsim 105$ GeV – below the LEP limit [138]. The combined significance is dominated by $ttH(H \to bb)$ and VBF $H \to \tau\tau$ at low masses, VBF $H \to WW$ for intermediate masses, and $H \to ZZ$ for higher masses. Near the LEP limit, several channels are required and available to observe the Higgs.

Recall that an expected significance of 5σ means that there is only a 50% chance to observe an effect in excess of 5σ if the Higgs is indeed there (see Section 13.4). In the 50% of cases in which the effect is less than 5σ, we can be quite confident that the effect will still be in excess of 3 or 4σ.

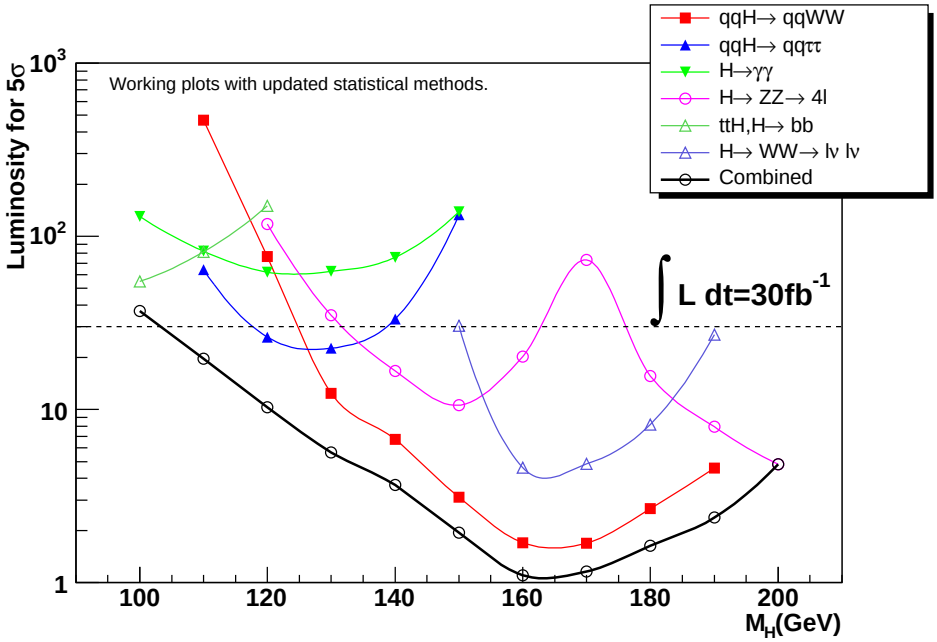


Figure 13.2 Discovery luminosity versus the Higgs mass hypothesis.

13.3 Luminosity Plots

In this section we present the "Luminosity Plot" (see Figure 13.2). We define the discovery luminosity $L^*(m_H)$ to be the integrated luminosity necessary for the expected significance at a Higgs mass m_H to reach 5σ. The discovery luminosity is an informative quantity; however, it must be interpreted with some care:

- Collecting an integrated luminosity equal to the nominal discovery luminosity $L^*(m_H)$ does not guarantee that a discovery will be made if the Higgs is indeed present at the corresponding mass m_H. Instead, with $L^*(m_H)$ of data, the median of the expected signal-plus-background will be at the 5σ level – which corresponds to a 50% chance of discovery. See Section 13.4 for more details.

- In practice an analysis' cuts, systematic error, and signal and background efficiencies are luminosity-dependent quantities. When we make the "luminosity" plot, we treat the analysis as constant. We must interpret the discovery luminosity with some care and realize that

beyond 30 fb^{-1} we move from a low- to a high-luminosity environment. We also must realize, though this is fairly obvious, that a discovery luminosity of 1 fb^{-1} does not mean the first 1 fb^{-1} of data, but at least 1 fb^{-1} of well-understood data that is consistent with the analysis' assumptions.

It is also worth pointing out that there does not always exist a luminosity for which the significance reaches 5σ (see Section C.3). For instance, the $t\bar{t}H(H \to bb)$ analysis is not a discovery channel if we hold fixed the signal-to-background ratio and the systematic error (at 10%) as we increase the luminosity. The background uncertainty in this channel includes statistical fluctuation in the control sample that is used determine the background [137]. As more data are accumulated, the background uncertainty will reduce. Thus for Figure 13.2 we have assumed a 5% systematic error for the $t\bar{t}H(H \to bb)$ analysis.

13.4 The Power of a 5σ Test

The traditional plot that is used to summarize the ATLAS discovery potential is the combined significance shown in Section 13.2; however, as noted in Section A.2.1 and in [136], this plot becomes very difficult to make in a consistent way when the significance goes beyond about 8σ. Furthermore, the plot itself starts to lose relevance when the significance is far above 5σ. The Luminosity Plot shown in Section 13.3 is another possible way of showing the ATLAS discovery potential, but as was discussed it must be interpreted with some care. In this section we introduce a third illustration of the ATLAS discovery potential which is related to the probability of a "false-negative" or Type II error: the *power*.

First, it should be noted that the significance plot measures the separation between the *medians* of the background-only and signal-plus-background hypotheses. Thus, when we see the significance curve cross the 5σ line (at some mass m_H^*) there is only a 50% chance that we would observe a 5σ effect if the Higgs does indeed exist at that mass. In practice, we claim a discovery if the observed data exceeds the 5σ critical region, and do not claim a discovery if it doesn't. The meaning of the 5σ discovery threshold is a convention which sets the probability of a "false-positive" or

Type I error to be $2.85 \cdot 10^{-7}$. With that in mind, the idea that the significance is 12σ at $m_H = 140$ GeV is irrelevant. What is relevant is the probability that we will claim discovery of the Higgs if it is indeed there: that quantity is called the power. The power is defined as $1 - \beta$, where β is the probability of Type II error: the probability that we reject the signal-plus-background hypothesis when it is true [139].

Consider Figure 13.3 with a background expectation of 100 events. The black vertical arrow denotes the 5σ discovery threshold. The red curve shows the distribution of the number of expected events for a signal-plus-background hypothesis with 150 events. Normally we would say the expected significance is 5σ for this hypothesis; however, we can see that only 50% of the time we would actually claim discovery. The blue curve shows the distribution of the number of expected events for a signal-plus-background hypothesis with 180 events. Normally we would say the expected significance is 8σ for this hypothesis; however, a more meaningful quantity - the power - is the probability we would claim discovery which, in this case, is about 98%.

When we use the likelihood ratio as a test statistic, the abscissa in Figure 13.3 is no longer a number of expected events; but instead it is the log-likelihood ratio q. Figure 13.4 shows the power of ATLAS' combined Higgs searches as a function of mass.

13.5 LEP-Style $-2 \ln Q$ vs. m_H Plots

The last plot which we present to summarize the ATLAS discovery potential is the LEP-style $-2 \ln Q$ vs. m_H plot. For each Higgs mass, one can imagine the two distributions of the log-likelihood ratio, $\rho_b(q)$ and $\rho_{s+b}(q)$. Figure 13.5 traces out the median of both of these distributions, and shows the 3σ and 5σ contours around the background only median. From this plot, one obtains the same general information as the significance plot. This plot would also need bands around the signal-plus-background curve for the power to be deduced. This plot is much easier to make than the significance plot, but is considerably harder to interpret for non-experts.

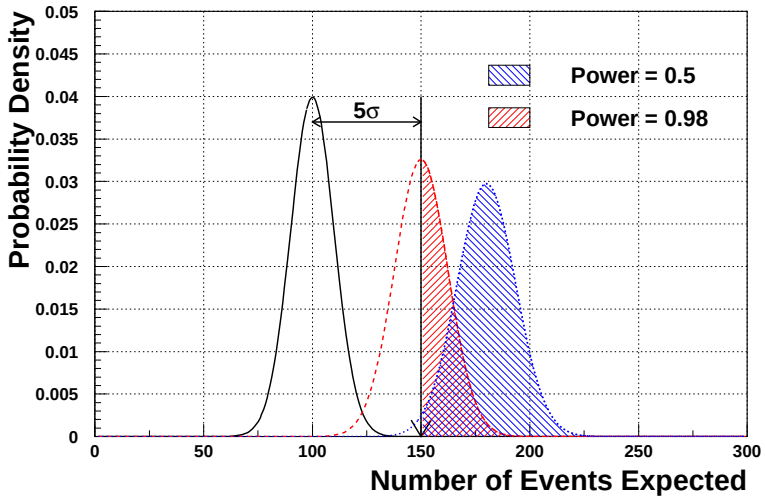


Figure 13.3 Examples of power for two different signal-plus-background hypotheses with respect to a single background-only hypothesis with 100 expected events (black).

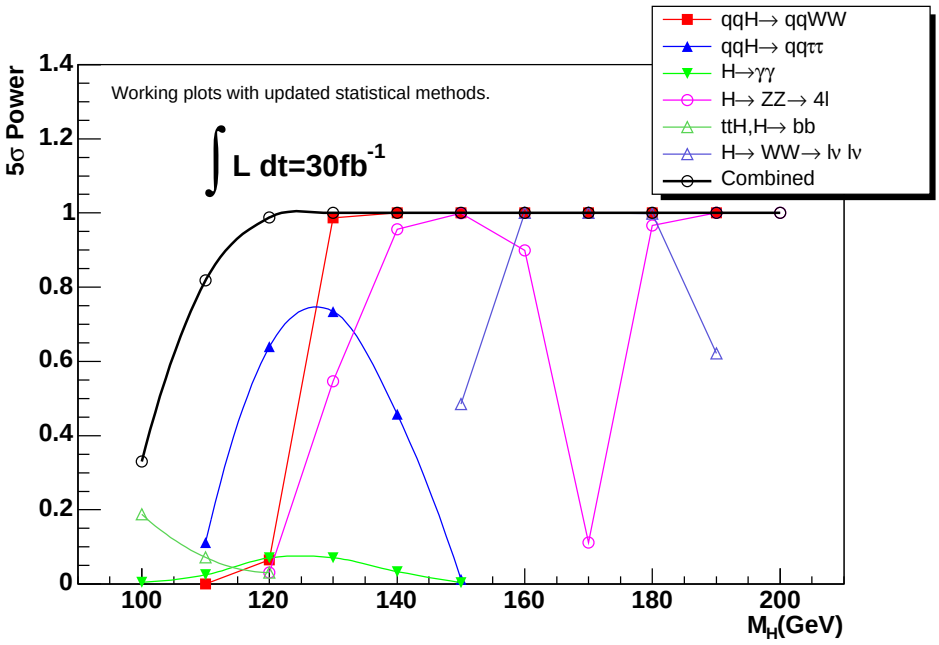


Figure 13.4 The power (evaluated at 5σ) of ATLAS as a function of the Higgs mass, m_H, for 30 fb^{-1} with and without systematic errors.

The reason that the ordinate is $-2\ln Q$, instead of $\ln Q$, is that $-2\ln Q$ is approximately equal to the difference in χ^2 when the data configuration is compared to the background-only and signal-plus-background hypotheses. The line at $-2\ln Q = 0$ corresponds to a data configuration that is ambivalent between the two hypotheses.

13.6 Conclusions

As was mentioned earlier, the majority of the analyses considered have been developed with ATLFAST, with the most relevant aspects studied, to varying degrees, in the full simulation. In the coming years, the main focus of the ATLAS collaboration on physics analysis will be on confirming the potential of these analyses in the full simulation, prioritizing the physics program for early discovery, and outlining a detailed physics commissioning schedule.

Assuming the signal-plus-background hypothesis, ATLAS has at least a 50% chance to claim a 5σ discovery if $m_H \gtrsim 105$ GeV with just 30 fb^{-1} of data (see Figure 13.1). If the Higgs is heavier than 120 GeV, we can be roughly 95% confident that ATLAS will be able to claim a 5σ discovery with 30 fb^{-1} (see Figure 13.5). Multiple channels are available for almost all values of m_H, allowing for a robust discovery and the potential for coupling measurements.

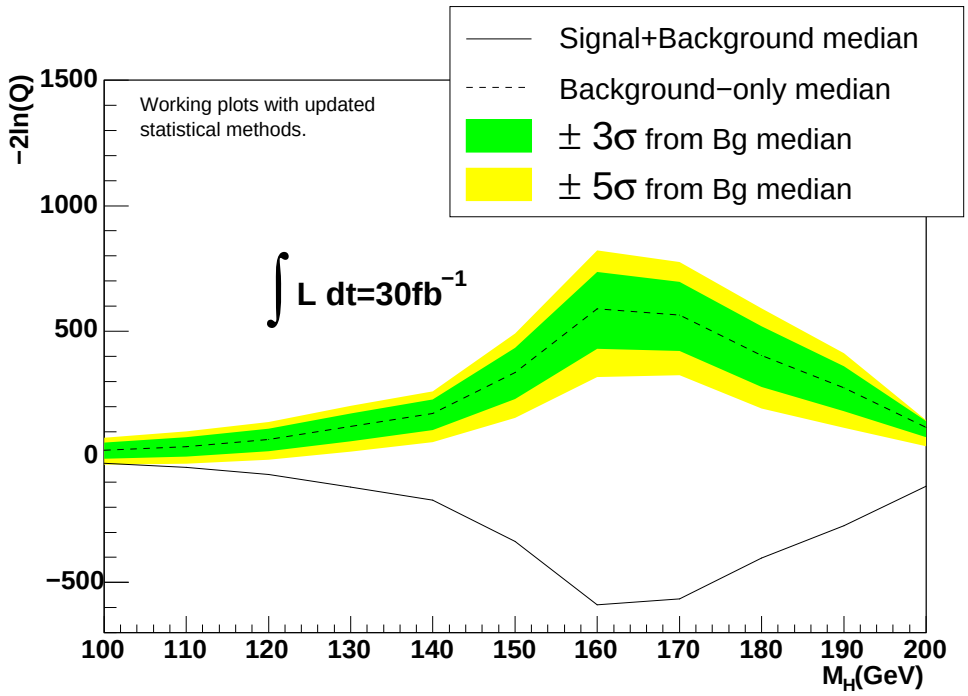


Figure 13.5 A plot of $-2 \ln Q$ vs. m_H for 30 fb^{-1} of integrated luminosity.

LIST OF REFERENCES

[1] S. Eidelman et al. Review of particle physics. *Phys. Lett.*, B592:1, 2004.

[2] Albert Einstein. The foundation of the general theory of relativity. *Annalen Phys.*, 49:769–822, 1916.

[3] R. P. Feynman. Space-time approach to nonrelativistic quantum mechanics. *Rev. Mod. Phys.*, 20:367–387, 1948.

[4] R. P. Feynman. Mathematical formulation of the quantum theory of electromagnetic interaction. *Phys. Rev.*, 80:440–457, 1950.

[5] Chen-Ning Yang and R. L. Mills. Conservation of isotopic spin and isotopic gauge invariance. *Phys. Rev.*, 96:191–195, 1954.

[6] Peter W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys. Lett.*, 12:132–133, 1964.

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

[8] S. L. Glashow. Partial symmetries of weak interactions. *Nucl. Phys.*, 22:579–588, 1961.

[9] S. Weinberg. 19:1264, 1967.

[10] A. Salam. Elementary Particle Theory,. Almqvist and Wiksells, Stockholm, 1968.

[11] G. Arnison et al. Experimental observation of lepton pairs of invariant mass around 95-gev/c^2 at the cern sps collider. *Phys. Lett.*, B126:398–410, 1983.

[12] G. Arnison et al. Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s}$ = 540-gev. *Phys. Lett.*, B122:103–116, 1983.

[13] V. E. Barnes et al. Observation of a hyperon with strangeness -3. *Phys. Rev. Lett.*, 12:204–206, 1964.

[14] Murray Gell-Mann. A schematic model of baryons and mesons. *Phys. Lett.*, 8:214–215, 1964.

[15] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak interactions with lepton - hadron symmetry. *Phys. Rev.*, D2:1285–1292, 1970.

[16] Martin L. Perl et al. Evidence for anomalous lepton production in e+ e- annihilation. *Phys. Rev. Lett.*, 35:1489–1492, 1975.

[17] D. Decamp et al. A precise determination of the number of families with light neutrinos and of the z boson partial widths. *Phys. Lett.*, B235:399, 1990.

[18] J. J. Aubert et al. Experimental observation of a heavy particle j. *Phys. Rev. Lett.*, 33:1404–1406, 1974.

[19] J. E. Augustin et al. Discovery of a narrow resonance in e+ e- annihilation. *Phys. Rev. Lett.*, 33:1406–1408, 1974.

[20] 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.

[21] S. Abachi et al. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, 1995.

[22] F. Abe et al. Observation of top quark production in anti-p p collisions. *Phys. Rev. Lett.*, 74:2626–2631, 1995.

[23] N. Cabibbo. Unitary symmetry and leptonic decays. *Phys. Rev. Lett.*, 10:531–532, 1963.

[24] M. Kobayashi and T. Maskawa. Cp violation in the renormalizable theory of weak interaction. *Prog. Theor. Phys.*, 49:652–657, 1973.

[25] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, 1998.

[26] D. J. Gross and Frank Wilczek. Asymptotically free gauge theories. 2. *Phys. Rev.*, D9:980–993, 1974.

[27] H. David Politzer. Reliable perturbative results for strong interactions? *Phys. Rev. Lett.*, 30:1346–1349, 1973.

[28] George Sterman and Steven Weinberg. Jets from quantum chromodynamics. *Phys. Rev. Lett.*, 39:1436, 1977.

[29] Sau Lan Wu and Georg Zobernig. A method of three jet analysis in e+ e- annihilation. *Zeit. Phys.*, C2:107, 1979.

[30] R. Brandelik et al. Evidence for planar events in e+ e- annihilation at high- energies. *Phys. Lett.*, B86:243, 1979.

[31] John C. Collins, Davison E. Soper, and George Sterman. Factorization for short distance hadron - hadron scattering. *Nucl. Phys.*, B261:104, 1985.

[32] V. N. Gribov and L. N. Lipatov. *Yad. Fiz.*, 15:1218, 1972.

[33] G. Altarelli and G. Parisi. *Nucl. Phys.*, B126:298.

[34] Y. L. Dokshitzer. *Sov. Phys. JETP*, 46:641.

[35] E. A. Kuraev, L. N. Lipatov, and Victor S. Fadin. Multi - reggeon processes in the yang-mills theory. *Sov. Phys. JETP*, 44:443–450, 1976.

[36] E. A. Kuraev, L. N. Lipatov, and Victor S. Fadin. The pomeranchuk singularity in non-abelian gauge theories. *Sov. Phys. JETP*, 45:199–204, 1977.

[37] I. I. Balitsky and L. N. Lipatov. The pomeranchuk singularity in quantum chromodynamics. *Sov. J. Nucl. Phys.*, 28:822–829, 1978.

[38] Torbjorn Sjostrand, Leif Lonnblad, and Stephen Mrenna. PYTHIA 6.2: Physics and Manual; hep-ph/0108264 (2001). 2001.

[39] G. Corcella et al. *JHEP*, 0101:10, 2001.

[40] Fabio Maltoni and Tim Stelzer. Madevent: Automatic event generation with madgraph. *JHEP*, 02:027, 2003.

[41] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, Roberto Pittau, and Antonio D. Polosa. Alpgen, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 07:001, 2003.

[42] Tanju Gleisberg et al. Sherpa 1.alpha, a proof-of-concept version. *JHEP*, 02:056, 2004.

[43] Stefano Frixione and Bryan R. Webber. The mc@nlo event generator. 2002.

[44] Search for the standard model Higgs boson at LEP. *Phys. Lett.*, B565:61–75, 2003.

[45] R. Barate et al. ALEPH Collaboration. Observation of an excess in the search for the standard model higgs. B495:1, 2000.

[46] M. J. G. Veltman. The infrared - ultraviolet connection. *Acta Phys. Polon.*, B12:437, 1981.

[47] Alexander A. Andrianov, R. Rodenberg, and N. V. Romanenko. Fine tuning in one higgs and two higgs standard model. *Nuovo Cim.*, A108:577–588, 1995.

[48] A. Dedes, S. Heinemeyer, S. Su, and G. Weiglein. The lightest higgs boson of msugra, mgmsb and mamsb at present and future colliders: Observability and precision analyses. *Nucl. Phys.*, B674:271–305, 2003.

[49] Nima Arkani-Hamed and Savas Dimopoulos. Supersymmetric unification without low energy supersymmetry and signatures for fine-tuning at the lhc. 2004.

[50] N. Arkani-Hamed, S. Dimopoulos, G. F. Giudice, and A. Romanino. Aspects of split supersymmetry. 2004.

[51] ALEPH Collaboration. ALEPH: A detector for electron-positron annihilations at LEP. Nucl. Instrum. Methods, A294:121, 1990.

[52] ALEPH Collaboration. Performance of the ALEPH detector at LEP. Nucl. Instrum. Methods, A360:481, 1995.

[53] ALEPH Collaboration. Measurement of the absolute luminosity with the ALEPH detector. *Z. Phys.*, C53:375, 1992.

[54] D. Bederede et al. Sical: a high precision silicon-tungsten calorimeter for aleph. Nucl. Instrum. Methods, A365:117, 1995.

[55] S. Jadach et al. Comp. Phys. Comm., 102:229, 1997.

[56] ALEPH Collaboration. http://aleph.web.cern.ch/aleph/Aleph_Light/galeph.html.

[57] ALEPH Collaboration. http://aleph.web.cern.ch/aleph/Aleph_Light/julia.html.

[58] ALEPH Collaboration. http://aleph.web.cern.ch/aleph/Aleph_Light/alpha.html.

[59] DØ Collaboration. Search for new physics using QUAERO: a general interface to DØ event data. *Phys. Rev. Lett.*, 87:231801, 2001.

[60] ALEPH Collaboration. Statement on the use of Aleph data for long-term analyses. 2003.

[61] K. Cranmer and B. Knuteson. QUAERO@ALEPH. Aleph-2004-009, 2004.

[62] A. Heister et al. Measurement of w pair production in e+ e- collisions at centre-of-mass energies from 183-gev to 209-gev. CERN-PH-EP-2004-012.

[63] Thomas Charles Greening. Search for the standard model higgs boson in topologies with a charged lepton pair at a center-of-mass energy of 188.6-gev with the aleph detector. ALEPH *Thesis*, 1999. UMI-99-27293.

[64] W. Bartel et al. *Z. Phys.*, C33:23, 1986.

[65] Jason Nielsen. Observation of an excess in the search for the standard model Higgs boson at ALEPH. ALEPH *Thesis*, 2001. UMI-30-20685.

[66] Y. Dokshitzer. *J. Phys.*, G17:1441, 1991.

[67] Z. Wąs S. Jadach, B. F. L. Ward. Comp. Phys. Comm., 130:260, 2000.

[68] S. Jadach et al. Phys. Lett. , B390:298, 1997.

[69] S. Jadach et al. *Comput. Phys. Commun.*, page 475., 2001.

[70] J. A. M. Vermaseren. 1980.

[71] ALEPH Collaboration. Status of Aleph Monte Carlo Production. 2000.

[72] ALEPH Collaboration. Measurement of W-pair production in e^+e^- at centre-of-mass energies from 183 to 209 GeV. *Eur.Phys.J.*, C, 2004.

[73] K. Cranmer, M. Maggi, B. Knuteson. 2004. http://mit.fnal.gov/ knuteson/Quaero/quaero/doc/devel/aleph/data/.

[74] L. Hölmstrom *et. al.* A new multivariate technique for top quark search. *Comput. Phys. Commun.*, 88:195–210, 1995.

[75] H. Miettinen and G. Epply. Possible hint of top $\rightarrow e^+ \not{E}_t$ + jets. DØ Note 002145 (1994).

[76] H. Miettinen. Top quark results from dØ. DØ Note 002527 (1995).

[77] B. Knuteson. 2003. The QUAERO Algorithm; http://mit.fnal.gov/ knuteson/Quaero/quaero/doc/algorithm/algorithm.ps.

[78] B. Knuteson. 2004. TURBOSIM: A Self-Tuning Fast Detector Simulation; http://mit.fnal.gov/ knuteson/papers/turboSim.ps.

[79] ALEPH Collaboration. *Phys. Lett. B*, 583:247–263, 2004.

[80] K. Hagiwara, D. Zeppenfeld, and S. Komamiya. Excited lepton production at lep and hera. *Z. Phys.*, C29:115, 1985.

[81] OPAL Collaboration. *Phys. Lett. B*, 544:57–72, 2002.

[82] OPAL Collaboration. *Phys. Lett. B*, 526:221–232, 2002.

[83] ALEPH Collaboration. *Phys. Lett. B*, 543:1–13, 2002.

[84] LHCC. LHC Proton Parameters for First Year of Operation: Version 2. http://bruening.home.cern.ch/bruening/lcc/WWW-pages/first_year_parameter.htm.

[85] ATLAS Collaboration. Detector and physics performance technical design report. CERN-LHCC/99-14 (1999).

[86] Z. Koba, Holger Bech Nielsen, and P. Olesen. Scaling of multiplicity distributions in high-energy hadron collisions. *Nucl. Phys.*, B40:317–334, 1972.

[87] G. Arnison et al. Transverse momentum spectra for charged particles at the cern proton anti-proton collider. *Phys. Lett.*, B118:167, 1982.

[88] G. J. Alner et al. Scaling violation favoring high multiplicity events at 540- gev cms energy. *Phys. Lett.*, B138:304, 1984.

[89] D. Acosta et al. The underlying event in hard interactions at the tevatron anti-p p collider. 2004.

[90] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour. Multiparton interactions in photoproduction at hera. *Z. Phys.*, C72:637–646, 1996.

[91] C. M. Buttar, D. Clements, I. Dawson, and A. Moraes. Simulations of minimum bias events and the underlying event, mc tuning and predictions for the lhc. *Acta Phys. Polon.*, B35:433–441, 2004.

[92] ATLAS Collaboration. Calorimeter Performance Technical Design Report. CERN/LHCC/96-40.

[93] ATLAS Collaboration. Liquid Argon Calorimeter Performance Technical Design Report. CERN/LHCC/96-41.

[94] ATLAS Collaboration. Tile Calorimeter Performance Technical Design Report. CERN/LHCC/96-42.

[95] ATLAS Collaboration. Inner Detector Technical Design Report Volume 1. CERN/LHCC/97-16.

[96] ATLAS Collaboration. Inner Detector Technical Design Repor Voume 2t. CERN/LHCC/97-17.

[97] ATLAS Collaboration. Magnet System Technical Design Report. CERN/LHCC/97-18.

[98] ATLAS Collaboration. Barrel Toroid Technical Design Report. CERN/LHCC/97-19.

[99] ATLAS Collaboration. End-Cap Toroids Technical Design Report. CERN/LHCC/97-20.

[100] ATLAS Collaboration. Central Solenoid Technical Design Report. CERN/LHCC/97-21.

[101] ATLAS Collaboration. Muon Spectrometer Technical Design Report. CERN/LHCC/97-22.

[102] ATLAS Collaboration. Pixel Detector Technical Design Report. CERN/LHCC/98-13.

[103] ATLAS Collaboration. First-Level Trigger Technical Design Report. CERN/LHCC/98-14.

[104] ATLAS Collaboration. High-Level Trigger Data Acquisition and Controls Technical Design Report. CERN/LHCC/2003-22.

[105] ATLAS Collaboration. Computing Technical Design Report. CERN/LHCC/96-43.

[106] D. Froidevaux E. Richter-Was and L. Poggioli. Atlfast2.0 a fast simulation package for atlas. ATLAS Internal Note ATL-PHYS-98-131.

[107] I. Vivarelli V. Cavasinni, D. Costanzo. Forward tagging and jet veto studies for Higgs events produced via vector boson fusion. ATLAS communication ATL-COM-CAL-2002-003 (2002).

[108] M. Spira. Fortsch. Phys. **46** (1998).

[109] David L. Rainwater, R. Szalapski, and D. Zeppenfeld. Probing color-singlet exchange in z + 2-jet events at the lhc. *Phys. Rev.*, D54:6680–6689, 1996.

[110] D. L. Rainwater and D. Zeppenfeld. Observing $H \to W^{(\star)}W^{(\star)} \to e^{\pm}\mu^{\pm}\not{p}_T$ in weak boson fusion with dual forward jet tagging at the CERN LHC. hep-ph/9906218 (1999).

[111] N. Kauer et al. $H \to WW$ as the discovery mode for a light Higgs boson. B503:113, 2001.

[112] D. Rainwater and D. Zeppenfeld. Observing $H \to W^{(\star)}W^{(\star)} \to e^{\pm}\mu^{\pm}\not{p}_T$ in weak boson fusion with dual forward jet tagging at the CERN LHC. D60:113004, 1999.

[113] David L. Rainwater, D. Zeppenfeld, and K. Hagiwara. Searching for $h \to \tau\tau$ in weak boson fusion at the lhc. *Phys. Rev.*, D59:014037, 1999.

[114] D. Zeppendeld et al. The home page of the madcup project. http://pheno.physics.wisc.edu/Software/MadCUP/.

[115] E. Boos et al. Generic user process interface for event generators. 2001.

[116] The CDF Collaboration. *Phys.Rev.*, D50:5562–5579, 1994.

[117] The D0 Collaboration. *Phys.Lett.*, B414:419–427, 1997.

[118] T. Plehn, David L. Rainwater, and D. Zeppenfeld. A method for identifying $H \to \tau\tau \to e\mu p_T^{miss}$ at the cern lhc. *Phys. Rev.*, D61:093005, 2000.

[119] D. Cavalli *et. al.* ATL-PHYS-94-051, ATL-PHYS-2003-009.

[120] L. Vacavant I. Hinchliffe, F.E. Paige. ATL-COM-PHYS-2002-037.

[121] F. Tarrade E. Richter-Was, H. Przysiezniak. ATL-PHYS-2004-030.

[122] G. Azuelos and R. Mazini. earching for $H \to \tau^+\tau^- \to l\nu_l\nu_\tau + hx$ by vector boson fusion in ATLAS. ATLAS internal note ATL-PHYS-2003-004.

[123] J. Kanzaki T. Takamoto, S. Asai and R. Tanaka. Study of $H \to \tau\tau$ (lepton and hadron mode) via vector boson fusion in ATLAS. ATLAS internal note ATL-PHYS-2003-007.

[124] S. Asai et al. Prospects for the search of a standard model Higgs boson in ATLAS using vector boson fusion. *Eur. Phys. J.*, C3252:19–54, 2004.

[125] M. Heldman. Private Communication and various ATLAS presentations.

[126] K. Cranmer, B. Mellado, W. Quayle, and Sau Lan Wu. Search for Higgs bosons decay $H \to W^+W^- \to l^+l^-\not{p}_T$ for $115 < M_H < 130\,\mathrm{GeV}$ using vector boson fusion. ATLAS note ATL-PHYS-2003-002 (2002).

[127] W. Quayle, and Sau Lan Wu K. Cranmer, P. McNamara, B. Mellado, Y. Pan. Neural network based search for Higgs bosons decay $H \to W^+W^- \to l^+l^-\not{p}_T$ for $115 < M_H < 130\,\mathrm{GeV}$. ATLAS note ATL-PHYS-2003-007 (2002).

[128] The homepage of the Stuttgart Neural Network Simulator (SNNS). http://www-ra.informatik.uni-tuebingen.de/SNNS.

[129] L. Bellantoni *et. al.* Using neural networks with jet shapes to identify b jets in e+ e- interactions. *Nucl. Instrum. Meth.*, A310:618–622, 1991.

[130] D. Decamp et. al. Search for the neutral Higgs bosons of the MSSM and other two doublet models. *Phys. Lett.*, B265:475–486, 1991.

[131] D. Buskulicet al. Search for the standard model Higgs boson. *Phys. Lett.*, B313:299–311, 1993.

[132] The homepage of mlpfit. http://schwind.home.cern.ch/schwind/MLPfit.html.

[133] R. Fletcher. *Practical Methods of Optimization, second eddition*. Wiley, New York, 1987.

[134] The BSVM library. http://www.csie.ntu.edu/˜cjlin/bsvm.

[135] K. Cranmer and R.S. Bowman. PhysicsGP: A genetic programming approach to event selection. *submitted to Comput. Phys. Commun.*

[136] K. Cranmer, P. McNamara, B. Mellado, W. Quayle, and Sau Lan Wu. Confidence level calculations for $H \to W^+W^- \to l^+l^-\not{p}_T$ for $115 < M_H < 130\,\mathrm{GeV}$ using vector boson fusion. ATLAS communication ATL-COM-PHYS-2002-049 (2002).

[137] J. Cammin and M. Schumacher. The ATLAS discovery potential for the channel $ttH, (H \to bb)$. ATLAS Note ATL-PHYS-2003-024 (2003).

[138] LEP Higgs Working Group. Search for the standard model higgs boson at lep. *Phys. Lett.*, B565:61–75, 2003.

[139] J.K Stuart, A. Ord and S. Arnold. *Kendall's Advanced Theory of Statistics, Vol 2A (6th Ed.)*. Oxford University Press, New York, 1994.

[140] LEP Higgs Working Group. Lower bound for the SM Higgs boson mass: combined result from the four LEP experiments. CERN/LEPC 97-11 LEPC/M 115.

[141] T. Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth.*, A434:435–443, 1999.

[142] R.D. Cousins and V.L. Highland. Incorporating systematic uncertainties into an upper limit. *Nucl. Instrum. Meth.*, A320:331–335, 1992.

[143] J. Nielsen H. Hu. Analytic confidence level calculations using the likelihood ratio and fourier transform. "Workshop on Confidence Limits", Eds. F. James, L. Lyons and Y. Perrin, CERN 2000-005 (2000), p. 109.

[144] A.L. Read. Modified frequentist analysis of search results (the cl(s) method). "Workshop on Confidence Limits", Eds. F. James, L. Lyons and Y. Perrin, CERN 2000-005 (2000), p. 81".

[145] K. Cranmer. Kernel estimation in high-energy physics. *Comput. Phys. Commun.*, 136:198–207, 2001.

[146] ATLAS Collaboration. Detector and physics performance technical design report (volume ii). CERN-LHCC/99-15 (1999).

[147] D. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley and Sons Inc., 1992.

[148] I. Abramson. On bandwidth variation in kernel estimates: A square root law. *Ann. Statist.*, 10:1217–1223, 1982.

[149] F. James and M. Roos. Errors on ratios of small numbers of events. *Nucl. Phys.*, B 172:475–480, 1980.

[150] R.D. Cousins. Improved cenral confidence intervals for the ratio of Poisson means. *Nucl. Instrum. and Meth. in Phys. Res.*, A 417:391–399, 1998.

[151] K. Cranmer. Frequentist hypothesis testing with background uncertainty. PhyStat2003 physics/0310108 (2003).

[152] Gary J. Feldman and Robert D. Cousins. A unified approach to the classical statistical analysis of small signals. *Phys. Rev.*, D57:3873–3889, 1998.

[153] J. Feldman, Gary. Multiple measurements and parameters in the unified approach, 2000. Workshop on Confidence Limits, FermiLab.

[154] V. Vapnik and A.J. Chervonenkis. The uniform convergence of frequencies of the appearance of events to their probabilities. *Dokl. Akad. Nauk SSSR*, 1968. in Russian.

[155] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 2nd edition, 2000.

[156] E. Sontag. VC dimension of neural networks. In C.M. Bishop, editor, *Neural Networks and Machine Learning*, pages 69–95, Berlin, 1998. Springer-Verlag.

[157] J.R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, 1992.

[158] J.K. Kishore et. al. Application of genetic programming for multicategory pattern classification. *IEEE Transactions on Evolutionary Computation*, 4 no.3, 2000.

[159] K. Cranmer. Multivariate analysis and the search for new particles. *Acta Physica Polonica B*, 34:6049–6069, 2003.

[160] K. Cranmer. Multivariate analysis from a statistical point of view. PhyStat2003 physics/0310110 (2003).

[161] R. D. Field and Y. A. Kanev. Using collider event topology in the search for the six-jet decay of top quark antiquark pairs. *hep-ph/9801318*, 1997.

[162] S. Luke. Two fast tree-creation algorithms for genetic programming. *IEEE Transactions on Evolutionary Computation*, 2000.

[163] D. Andre and J.R. Koza. Parallel genetic programming on a network of transputers. In Justinian P. Rosca, editor, *Proceedings of the Workshop on Genetic Programming: From Theory to Real-World Applications*, pages 111–120, Tahoe City, California, USA, 9 1995.

[164] P.J. Werbos. *The Roots of Backpropagation*. John Wiley & Sons., New York, 1974.

[165] D.E. Rumelhart *et. al. Parallel Distributed Processing Explorations in the Microstructure of Cognition*. The MIT Press, Cambridge, 1986.

[166] G. Punzi. Sensitivity of searches for new signals and its optimization. In *PhyStat2003*, 2003.

Appendix A: Moving LEP-Style Statistics to the LHC

A.1 The LEP Statistical Framework

In the final years of data taking at LEP, the LEP Higgs Working Group (LHWG) was formed to combine the results from ALEPH, OPAL, DELPHI, and L3. Key to the success of this combination was a consistent statistical framework between the experiments. The basis of the framework was simple hypothesis testing as viewed in the Neyman-Pearson theory (see Section A.1.1). The framework was extended to include systematic errors with the Cousins-Highland technique (see Appendix C) and modified to protect from undesirable limit setting scenarios with the CL_S method (see Section A.1.6).

A.1.1 The Neyman-Pearson Theory

The Neyman-Pearson theory [139] begins with two Hypotheses: the null hypothesis H_0 and the alternate hypothesis H_1. In the case of a new particle search H_0 is identified with the currently accepted theory (*i.e.* the Standard Model) and is usually referred to as the "background-only" hypothesis. Similarly, H_1 is identified with the theory being tested (*i.e.* Standard Model with Higgs boson at some specified mass m_H) usually referred to as the "signal-plus-background" hypothesis. With these two hypotheses one is able to describe, through theoretical calculations and detector simulation, the probability distribution of physical observables $x \in I$, written as $L(x|H_0)$ and $L(x|H_1)$. Next, one defines a region $W \in I$ such that if the data fall in W we accept the null hypothesis (and reject the alternate hypothesis). Similarly, if the data fall in $I - W$ we reject the null hypothesis and accept the alternate hypothesis. Recognize that if the null hypothesis is true, then there exists a chance that the data could fall in $I - W$ and we reject H_0 even though it is true – we commit a *Type I error*. The probability to commit a Type I error is called the *size* of the test by statisticians, but is commonly referred to as the background confidence level CL_b in particle physics. The size of the test is given by

$$\alpha \equiv CL_b = \int_{I-W} L(x|H_0)\, dx. \tag{A.1}$$

Similarly, if the alternate hypothesis is true, the data could fall in W, in which case we accept H_0 even though it is false – we commit a *Type II error*. The probability to commit a Type II error is given by

$$\beta = \int_W L(x|H_1)\, dx. \tag{A.2}$$

Also of importance is the notion of *power* $= 1 - \beta$, which can be interpreted as the chance that one accepts H_1 when it is true.

In particle physics, the discovery criterion is often referred to as the 5σ requirement (see Section A.3.1). In general cases, the signal and background distributions are not Gaussian, though the expression of the signal significance in terms of Gaussian significance is intuitive. The background confidence level can be converted into an equivalent number of Gaussian standard deviations in by finding the value x which forms a one-sided confidence interval with confidence level of interest. In particular, we want the value of N which satisfies

$$\alpha = \frac{1 - \mathrm{erf}(N/\sqrt{2})}{2}, \tag{A.3}$$

where $\mathrm{erf}(N) = (2/\sqrt{\pi}) \int_N^\infty \exp(-y^2) dy$. Using this convention 5σ corresponds to $\alpha = 2.9 \cdot 10^{-7}$ (and not the more familiar $\alpha = 5.8 \cdot 10^{-7}$).

The central result of the Neyman-Pearson theory is the Neyman-Pearson lemma, which tells us how to chose an acceptance region W. The Neyman-Pearson lemma states that holding α fixed, the region W that maximizes the power is bounded by a contour of the Likelihood ratio

$$W = \left\{ x \;\middle|\; \frac{L(x|H_1)}{L(x|H_0)} < k_\alpha \right\}, \tag{A.4}$$

where k_α is a constant chosen to satisfy Equation A.1.

The formalism here is that which was used by the LEP Higgs working group [140, 141]: it is a classical, or *frequentist*, technique. In order to include systematic errors, the Cousins-Highland approach has been adopted [142]. Furthermore, specific numerical techniques used at ALEPH, which perform convolutions using the Fourier Transform are utilized [143].

A.1.2 The Likelihood Ratio as a Test Statistic

As a consequence of the Neyman-Pearson lemma, the likelihood ratio was used by LHWG to combine channels [144]. In the case of a number counting experiment, x is simply the number of observed events, $L(x|H)$ is a Poisson distribution, and Q can be written as

$$Q(x) = \frac{L(x|H_1)}{L(x|H_0)} = \frac{e^{-(s+b)}(s+b)^x/x!}{e^{-b}b^x/x!} = e^{-s}\left(1+\frac{s}{b}\right)^x, \tag{A.5}$$

where s and b are the expected number of signal and background events respectively.

For convenience, the natural logarithm of this expression,

$$q(x) = \ln Q(x) = -s + x\ln\left(1+\frac{s}{b}\right) \tag{A.6}$$

is often used instead. It can immediately be seen that this expression consists of an offset ($-s$) and a term proportional to the number of events observed. This proportionality factor can be considered to be an event weight, though in this simple example, all events are given the same weight.

A.1.3 Combining Channels and the Likelihood Ratio

To combine two channels, one simply multiplies the likelihood ratios together (or adds the log-likelihood ratios). For N_{ch} channels, this becomes

$$q(x) = \ln Q(x) = -\sum_{i=1}^{N_{ch}} s_i + \sum_{i=1}^{N_{ch}} x_i \ln\left(1+\frac{s_i}{b_i}\right) \tag{A.7}$$

where s_i, b_i and x_i are the signal expectation, background expectation and number of events observed for the i^{th} channel and the generic observable x is now a point in $\mathbb{R}^{N_{CH}}$. Equation A.7 consists of an offset, which is the total signal expectation for all channels, and a sum over candidates, where each candidate is given a weight dependent on its channel's purity.

As in the single channel case, the confidence level can be computed using the Poisson probabilities for observing various numbers of events. With multiple channels, however, this is more complicated, as it requires multiple convolutions. For instance, the probability density function (pdf) for q coming from the combination of two channels A and B is given by

$$\rho_{AB}(q) = \int_{-\infty}^{\infty} \rho_A(q')\rho_B(q-q')dq'. \tag{A.8}$$

As a result, the multi-channel probability distribution is usually computed with Monte Carlo techniques. Monte Carlo techniques, however, have the drawback that it is quite time consuming to generate a sufficiently large sample when computing significances larger than a few standard deviations and the number of expected events is quite large. Fortunately, one can make use of analytic methods, which perform the convolution via fast Fourier Transform (FFT), to compute the multi-channel probability distribution quickly and accurately [143]. More details are given in Section A.2.1.

A.1.4 Discriminating Variables

From a statistical point of view, calculating the likelihood with a discriminating variable is the continuous limit of combining multiple channels (see Equation A.7). Just as there were channels with low and high purity, there are regions in the discriminating variable with low and high purity. In LEP Higgs searches, the discriminant variable was typically the reconstructed Higgs mass, a neural network output, or a b-tagging variable (see Figure B.2). From Monte Carlo, it is possible to construct estimates of the signal and background pdf's $f_s(x)$ and $f_b(x)$, respectively.

For a single event with $x = x_i$, the log-likelihood ratio generalizes in a straight forward manner,

$$q(x_i) = \ln Q(x_i) = -s + \ln\left(1 + \frac{s f_s(x_i)}{b f_b(x_i)}\right). \tag{A.9}$$

In this way, $f_s(x)$ and $f_b(x)$ are mapped into an expected distribution of $q(x)$. For the background-only hypothesis, $f_b(x)$ provides the probability of corresponding values of q needed to define the single event pdf ρ_1[1].

$$\rho_{1,b}(q_0) = \int f_b(x)\,\delta(q(x) - q_0)dx \tag{A.10}$$

A.1.5 KEYS

The description of $f_s(x)$ and $f_b(x)$ is another area of concern. While a histogram will suffice, the discontinuities in the pdf are not desirable. Furthermore, the binning of the histogram can

[1] The integral is necessary because the map $q(x) : x \to q$ may be many-to-one.

produce quite different descriptions of the underlying pdf. These effects lead to a systematic associated with the binning. Some experiments employed the PAW utility SMOOTH to remove the discontinuities; however, this method was plagued with other undesirable effects (see Section B.5).

To alleviate these problems, the author developed KEYS, a package which constructs probability density estimates with kernel estimation techniques. KEYS is described in detail in Appendix B. The LEP Higgs working group adopted the use of KEYS and cited Reference [145] in their final results.

A.1.6 The CL_s Method

The CL_S method was developed by Alex Read in order to avoid excluding the signal hypothesis when the signal and background would both be excluded [144]. The quantity CL_S is defined as

$$CL_s = \frac{CL_{s+b}}{CL_b} \tag{A.11}$$

and does not correspond to a probability. Instead, CL_S is a ratio of frequentist probabilities. The LEP exclusion was based on the requirement $CL_S < 5\%$ (see Figure 2.5).

A.2 An Implementation for the LHC

The author developed a C++ package primarily designed to assess the discovery potential of ATLAS. The focus of this package is on hypothesis testing and not on limit setting. The package was developed for studies of the ATLAS detector's potential to discover the Standard Model Higgs Boson. During the development of this package, several technical challenges were encountered which were not relevant at the LEP experiments.

The package includes a number of useful functions as well as a number command line interfaces which calculates the significance in terms of Gaussian "sigma". There are four main components to the package:

- **PoissonSig** Used to calculate the significance of a number counting analysis.
- **PoissonSig_syst** Used to calculate the significance of a number counting analysis including systematic error on the background expectation.

- **Likelihood** Used to calculate the combined significance of several search channels or to calculate the significance of a search channel with a discriminating variable.
- **Likelihood_syst** Used to calculate the combined significance of several search channels including systematic errors associated with each channel.

The package also includes tools to aid in calculating the luminosity necessary to achieve the 5σ discovery threshold, the power of a test, and contours of $-2\ln Q$ like those found in Chapter 13.

A.2.1 The Fourier Transform Technique

For multiple events, the distribution of the log-likelihood ratio must be obtained from repeated convolutions of the single event distribution [143]. In the Fourier domain, denoted with a bar, the distribution of the log-likelihood for n particles is

$$\overline{\rho_n} = \overline{\rho_1}^n \tag{A.12}$$

Thus the expected log-likelihood distribution for background takes the form

$$\rho_b(q) = \sum_{n=0}^{\infty} \frac{e^{-b} b^n}{n!} \rho_{n,b}(q) \tag{A.13}$$

which in the Fourier domain is simply

$$\overline{\rho_b(q)} = e^{b[\overline{\rho_{1,b}(q)}-1]}. \tag{A.14}$$

For the signal-plus-background hypothesis we expect s events from the $\rho_{1,s}$ distribution and b events from the $\rho_{1,b}$ distribution which leads to

$$\rho_{s+b}(q) = \sum_{n=0}^{\infty} \frac{e^{-b} b^n}{n!} \rho_{n,b}(q) + \sum_{n=0}^{\infty} \frac{e^{-s} s^n}{n!} \rho_{n,s}(q). \tag{A.15}$$

In the Fourier domain ρ_{s+b} is simply

$$\overline{\rho_{s+b}(q)} = e^{b[\overline{\rho_{1,b}(q)}-1]+s[\overline{\rho_{1,s}(q)}-1]}. \tag{A.16}$$

Perhaps it is worth noting that $\overline{\rho(q)}$ is actually a complex valued function of the Fourier conjugate variable of q. Thus numerically the exponentiation in Equation A.14 requires Euler's formula $e^{i\theta} = \cos\theta + i\sin\theta$[2].

Numerically these computations are carried out with the Fast Fourier Transform (FFT). The FFT is performed on a finite and discrete array, beyond which the function is considered to be periodic. Thus the range of the ρ_1 distributions must be sufficiently large to hold the resulting ρ_b and ρ_{s+b} distributions. If they are not, the "spill over" beyond the maximum log-likelihood ratio q_{max} will "wrap around" leading to unphysical ρ distributions. Because the range of ρ_b is much larger than $\rho_{1,b}$ it requires a very large number of samples to describe both distributions simultaneously. The nature of the FFT results in a number of round-off errors and limit the numerical precision to about 10^{-16} – which are significant for consistently describing the significance beyond about 8σ. Extrapolation techniques and Arbitrary Precision calculations can overcome these difficulties and are the subject of Sections A.2.3 and A.2.4, respectively.

A.2.2 Interpolation

In a number counting experiment the background confidence level calculation for an observation will be based on an integer-valued observed number of events N and a real-valued expected number of events b. In this case the CL_b will be given by

$$CL_b = \sum_{i=N}^{\infty} P(i;b) = \sum_{i=N}^{\infty} e^{-b} b^i / i!. \tag{A.17}$$

However, when assessing the discovery potential for a future experiment, we may expect a real-valued number of observed events. Initially, the `PoissonSig` program was written such that it would find the median of the Poisson distribution associated with the signal-plus-background distribution (an integer) and then use that as N in the equation above. This leads to the pathological behavior seen in Figure A.1: the significance is not only discontinuous, but also increases as the background expectation increases. Let us consider the behavior for 3 signal events in the case of

[2] Can't resist pointing out $e^{i\pi} + 1 = 0$

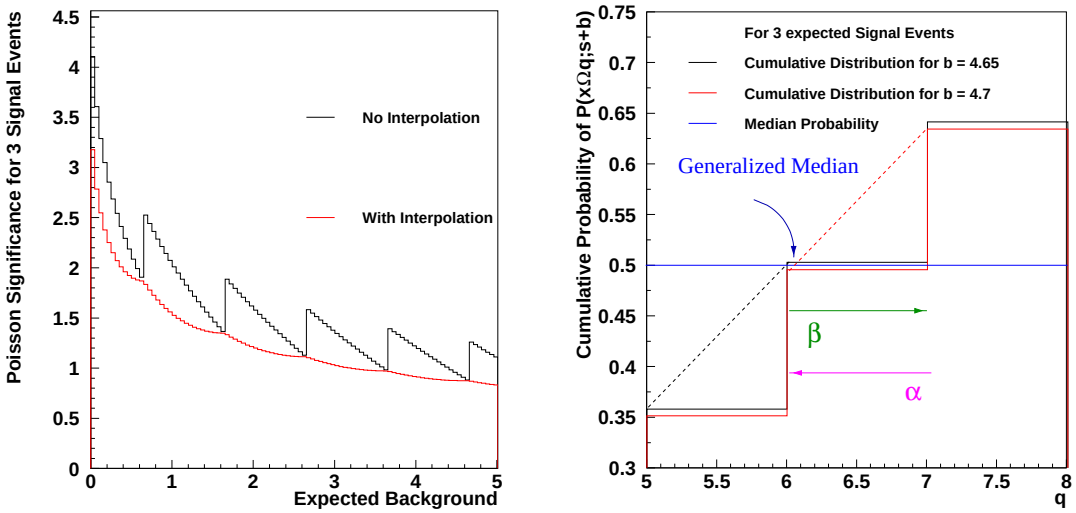


Figure A.1 Left: The pathological behavior of the unmodified Poisson significance calculation (black). It is not only discontinuous, but also increases as the background expectation increases. Continuity is restored with the interpolation (red) provided by the generalized median (right).

4.65 and 4.7 background events. Figure A.1 shows the cumulative distribution of the signal-plus-background distribution is hardly changed between these two points; however, the median changes discontinuously due to the discreteness of the Poisson distribution. Thus for 4.65 background events $N = 6$ and for 4.7 background events $N = 7$. Thus for 4.7 background events the CL_b is less (the significance is higher).

By simply interpolating the cumulative probability and finding its intersection with $1/2$, we can produce a generalized median that changes continuously. With the generalized median of the signal-plus-background distribution we wish to evaluate CL_b. Because the Poisson distribution is discrete, we must also generalize the CL_b calculation. This is done as follows:

- Let x_0 be the last integer with $P(x \leq x_0; s + b) < 1/2$.
- Linearly Between Interpolate x_0 & $x_0 + 1$ to find β & α.
- Generalize the median as $\tilde{\mu} = x_0 + \beta$.
- Generalize CL_b as $P(x \geq \tilde{x}; b) := \alpha P(x \geq x_0; b) + \beta P(x \geq x_0 + 1; b)$

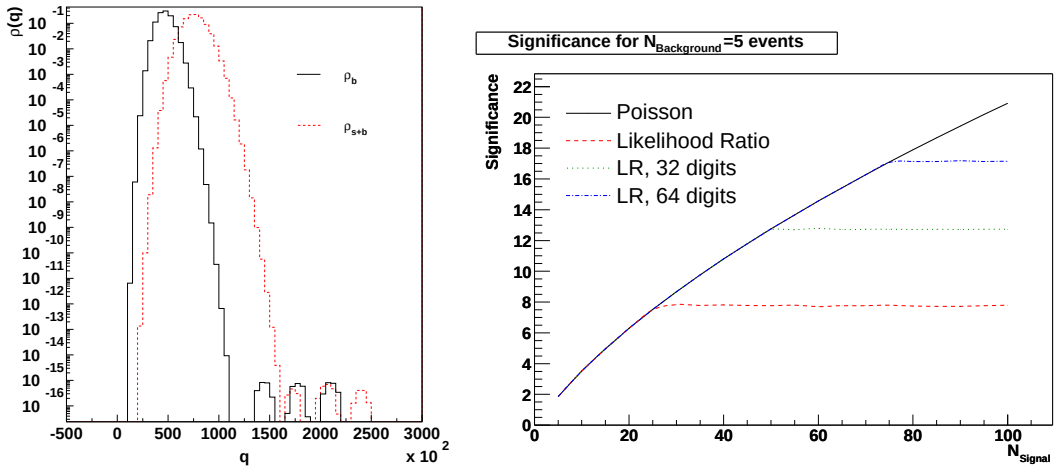


Figure A.2 Illustration of the numerical "noise" which appears for $\rho(q) \lesssim 10^{-16}$.

The same situation occurs in the case of the a likelihood ratio calculation, however; the values of the likelihood ratio need not be integer-valued. Computationally, the ρ_{s+b} distribution is a histogram possibly with many empty bins between the adjacent non-empty bins q_0 and q_1. Thus one must slightly modify the interpolation algorithm above such that $\alpha, \beta \in [0, 1]$, $x_0 \rightarrow q_0$ and $x_0 + 1 \rightarrow q_1$.

A.2.3 Extrapolation

The numerical limitations in the Fourier Transform Technique (introduced in Section A.2.1) are the result of many round-off errors in the FFT. Figure A.2 illustrates a representative ρ_b and ρ_{s+b} distributions spanning over 16 orders of magnitude. It is apparent that the the numerical precision is a limitation when the median of the signal-plus-background distribution is located in these unreliable regions. For double precision floating point numbers, these effects limit the ability to calculate significances above about 8σ. In Section A.2.4 we discuss a solution to this problem in which the FFT is implemented with an arbitrary precision library; however, this method is excruciatingly slow and memory intensive. Thus, in this section various extrapolation techniques are described.

The first extrapolation technique to be applied was a simple "Gaussian extrapolation" in which the ρ_b distribution was described by a Gaussian with the same mean μ_b and standard deviation σ_b

(not really a fit in the common sense of the word). In this case the significance was simply quoted as $\sigma = (\tilde{\mu} - \mu_b)/\sigma_b$ (see Figure A.3). For calculations with many events, the Gaussian approximation is expected to be valid. Because the Gaussian distribution allows for $\rho_b(q < -s_{tot}) > 0$ we expect the Gaussian extrapolation technique to overestimate the significance in general. This behavior can be seen in Figure A.4.

The second method we studied was based on a Poisson fit to the ρ_b distribution. The Poisson distribution has the desirable properties that it will have no probability below the hard limit and that its shape is more appropriate. However, the Poisson distribution is a discrete distribution thus we must find some affine transformation between the space of the log-likelihood ratio and the space of the Poisson distribution. This is accomplished as follows: First we use the identity that for a Poisson distribution the $P(x;\mu)$ the mean is given by μ and the variance is given by μ. Next we assume that our distribution $\rho_b(q)$ takes the form of a Poisson with $q = \alpha x$, which forces $\mathrm{mean}(\rho_b) = \alpha\mu$ and $\mathrm{var}(\rho_b) = \alpha^2\mu$. This gives us two equations which we can use to solve for μ and α. With those parameters, the median of the signal-plus-background distribution and the mean of the background-only distribution can be transformed via α to produce the corresponding Poisson significance.

Figure A.4 offers a comparison of these methods for an example ATLAS Higgs combined significance calculation. For reference a (green-dotted) curve obtained from adding in quadrature (green dotted line) is included. The red dashed line corresponds to the unmodified likelihood ratio which can not produce significance values above about 8σ. The Gaussian extrapolation technique tends to overestimate the significance, while the Poisson extrapolation is well behaved across the entire mass range. The VBF channels and the channels discussed in [146] are used for this combination. This figure is meant to demonstrate the different methods of combination and does not include updated numbers for non-VBF analyses. No systematic errors on background normalization have been included.

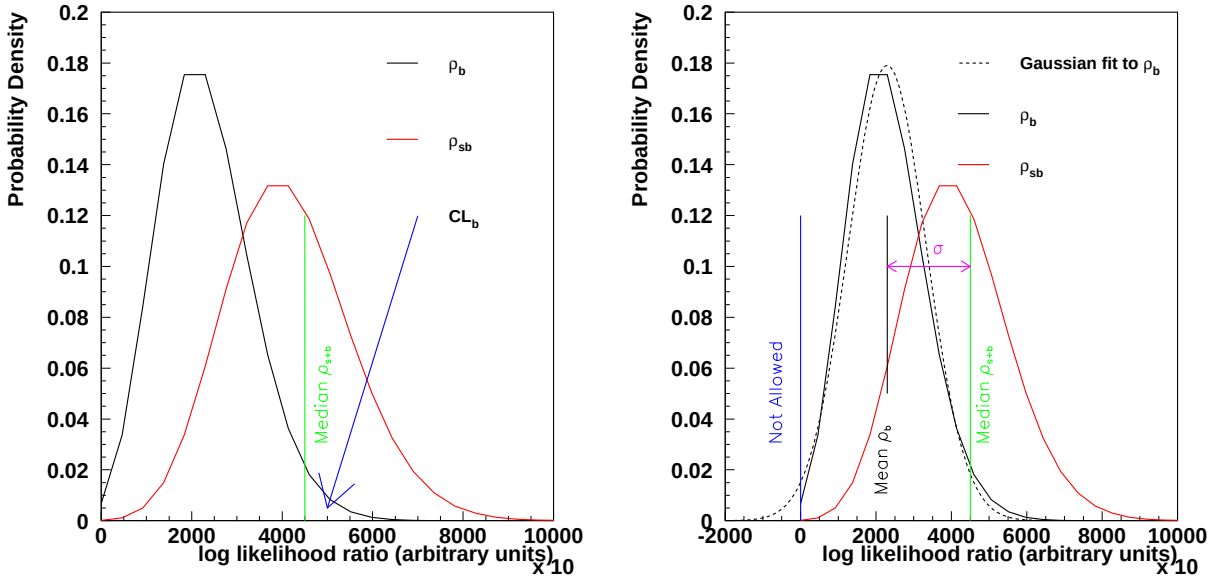


Figure A.3 Diagram for the Gaussian extrapolation technique. The abscissa corresponds to the histogram bin index of the log-likelihood ratio, in which the 0^{th} bin corresponds to the lower limit $q = -s_{tot}$ (see Equation A.6).

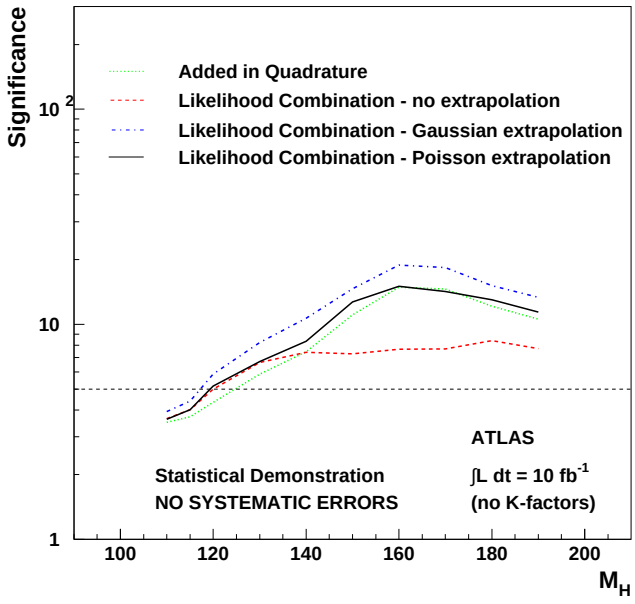


Figure A.4 Comparison of the combined significance obtained from various combination procedures.

A.2.4 Accessing Low CL_b with Arbitrary Precision Libraries

Figure A.2 demonstrates the problem and its solution; it shows the expected significance versus the number of expected signal events for a number-counting experiment with 5 expected background events. For a single-channel number-counting analysis, the CL_b can be calculated from the Poisson distribution $P(n;b)$ directly, and no FFT need be performed (the black curve). However, the calculation can also be done with likelihood ratio techniques, and the results should agree exactly. The red curve was obtained from a likelihood ratio calculation performed with double-precision numbers using the FFTW library for the FFT. From the figure, it is clear that it agrees very well with the exact calculation until the significance approaches about 8σ, where the numerical noise starts to dominate. The green and blue curves show the results of the same calculation performed with 32 digit and 64 digit CLN numbers, respectively. The result is clear and unsurprising: using higher precision numbers to calculate the likelihood ratio probability distribution reduces the numerical noise and makes the calculation of the confidence level (and significance) reliable to much more extreme values.

One might protest that above 5σ we are not interested in the precise value of the significance and that this exercise is purely academic. We refer the interested reader to Section 13.4 for a different summary of the ATLAS discovery potential based on the notion of power.

A.3 Why 5σ?

A.3.1 Decision Making and Utility

Once one specifies the size, α, of the test the power of the test is determined from $L(x|H_0)$ and $L(x|H_1)$. How one chooses the size of the test, however, transcends the Neyman-Pearson theory. Typically, scientists retreat to conventional values such as $\alpha = 0.05$ (which corresponds to a 95% confidence) or 5σ in the case of particle physics. These choices are essentially arbitrary, but that need not be the case.

For example, if the discovery threshold were 100σ, then we would never be able to claim a discovery – which would clearly be of little utility. Similarly, if the threshold were 1σ, then we

would often commit a Type I error – and no one would trust our results. So let us consider arbitrary (positive) utility for discovery or limit setting and (negative) utility for committing a Type I or Type II error. Additionally, we could generalize the accept/exclude logic so that the a the size of the test for discovery is α and for limit setting is α'. In that case there is a possibility that we neither claim discovery nor do we exclude the alternate hypothesis. The lack of a result also has some (negative?) utility. Given that notion of utility we can write:

$$U(H_0) = (1-\alpha') \cdot U(\text{Type I}) + \beta' \cdot U(\text{Limit}) + (\beta - \beta') \cdot U(\text{No Result}) \tag{A.18}$$

and

$$U(H_1) = (1-\beta) \cdot U(\text{Discovery}) + \beta' \cdot U(\text{Type II}) + (\beta - \beta') \cdot U(\text{No Result}). \tag{A.19}$$

One must be careful at this stage. It is quite tempting to add these two utility functions since only one hypothesis can be true. In a Bayesian setting one could introduce $p(H_1)$ and $p(H_0)$ and construct an ultimate $U = p(H_0)U(H_0) + p(H_1)U(H_1)$, but that is not allowed in a frequentist formalism.

Instead we have something more akin to game theory. We must choose a strategy (*i.e.* a discovery threshold in σ) for which we know the payoff under two of our opponents plays. What is unusual, is that our opponent is Nature, and we do not consider her to be diabolical. The *minimax* theorems of game theory only enter if the opponent is also aware of the payoff table and attempts to maximize his/her payoff. Games of this type are called, appropriately, *games against Nature*. There is no equivalent to the *minimax* condition that is not in some way *ad hoc* or Bayesian. Nonetheless, we can say something about the particular case of particle physics.

Let us consider an example of a number counting experiment with 100 expected background events and 60 expected signal events. Traditionally, one would say that this experiment has an expected significance of $s/\sqrt{b} = 6\sigma$. For clarity we consider $\alpha = \alpha'$ (which implies $\beta = \beta'$). For the purpose of making figures, we arbitrarily choose $U(\text{Discovery}) = 12$, $U(\text{Limit}) = 2$, $U(\text{Type I}) = -8$, and $U(\text{Type II}) = -17$. The units of the utility are arbitrary, but they could be, for example, next year's funding for particle physics, the number of faculty appointments, or the

contribution to the gross national product from technology transfer. Instead, it is the ratios of these numbers that drives the strategy.

In Figure A.5 shows the Utility as a function of the discovery threshold in σ. In the top plot, one can see that beyond about 2σ that $U(H_0)$ approaches $U(\text{Limit})$ because the chance of committing a Type I error is quite small and the chance of setting a limit is quite large. Similarly, discovery is nearly assured (the power is high) until about 4σ. Both curves have a sigmoidal shape, which can be characterized by their lower and upper plateaus.

Let us define the plateau points to be the discovery threshold that gives a utility $1-\epsilon$ of its asymptotic value. Via Equation A.18 we arrive at the condition

$$\alpha^+ = \epsilon \left[1 - \frac{U(\text{Type I})}{U(\text{Limit})}\right]^{-1}. \tag{A.20}$$

If the penalty for claiming false discovery was much larger, say $U(\text{Type I}) = -10^5$, then the curve would look like the bottom of Figure A.5. Rewriting Equation A.20 we arrive at

$$1 - \frac{\epsilon}{\alpha^+} = \frac{U(\text{Type I})}{U(\text{Limit})}. \tag{A.21}$$

Ideally, the field would establish these utilities instead of working with the purely conventional 5σ requirement. Since that is not the case, it is reasonable to ask "what is this ratio of utilities which justifies a 5σ discovery threshold?" If we take $\epsilon = 1\%$ and $\alpha = 10^{-7}$, then $|U(\text{Type I})/U(\text{Limit})| > 10^5$. Perhaps this ratio is reasonable, perhaps not, but it is the ratio under which we operate today.

From Equation A.19 we can derive the plateau points for the alternate hypothesis.

$$(1-\beta^+) = \epsilon \left[1 - \frac{U(\text{Discovery})}{U(\text{Type II})}\right]^{-1} \tag{A.22}$$

and

$$\beta^- = \epsilon \left[1 - \frac{U(\text{Type II})}{U(\text{Discovery})}\right]^{-1} \tag{A.23}$$

As mentioned above, there is no equivalent to the *minimax* theorem for games against nature; however, there are some special cases in this context.

In the first case, corresponding to the top of Figure A.5, is when $\alpha^+ < \beta^-$. In that case and only that case do the utility for both hypotheses reach their positive plateaus simultaneously. Essentially any discovery threshold in the range between α^+ and β^- is equivalent in terms of utility.

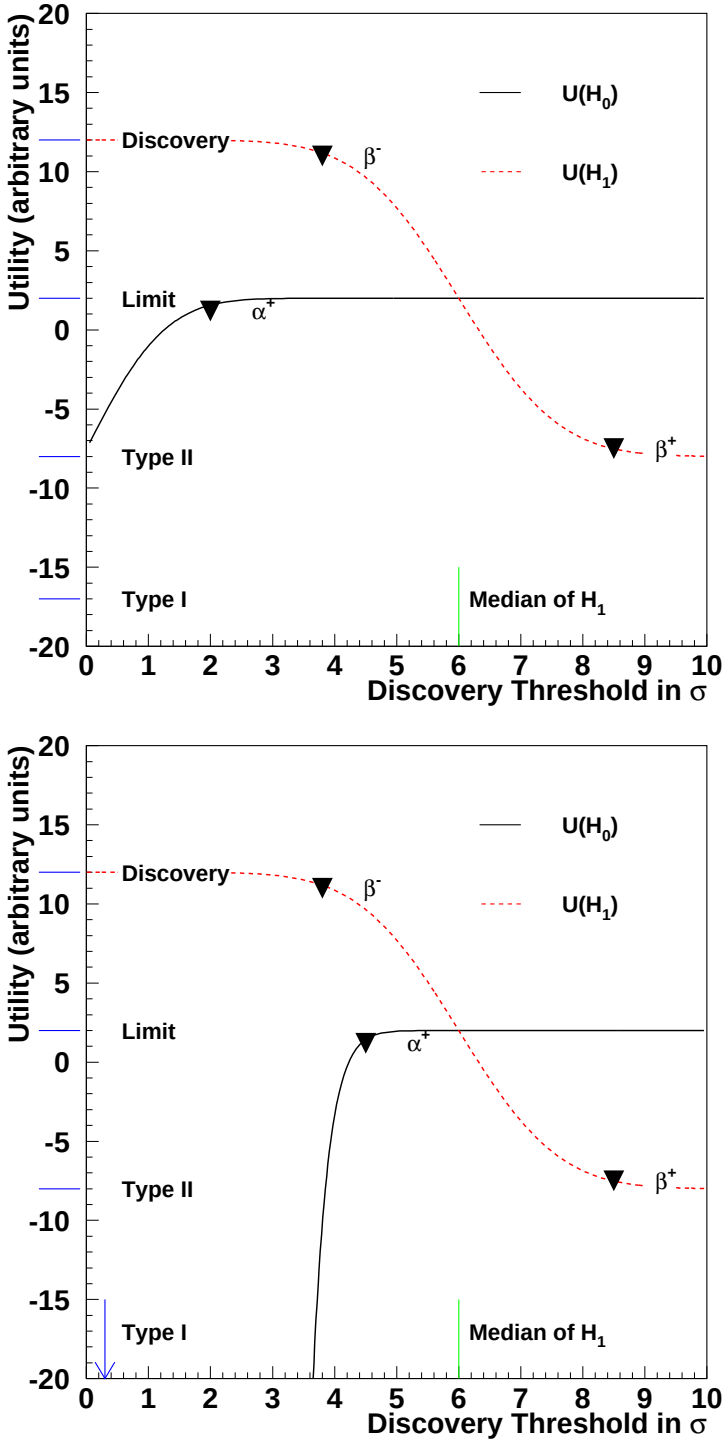


Figure A.5 Utility as a function of the discovery threshold for a channel with an expected 6σ significance when the utility for a Type I error is -17 (top) and -10^5 (bottom).

The second case, corresponding to the bottom of Figure A.5, is more typical for the LHC. The penalty of Type I error is quite large, so that α^+ corresponds to roughly 5σ. Even though the expected significance is 6σ, the probability of discovery starts to drop off around 3.5σ. A reasonable choice for a discovery threshold would be α^+, because $U(H_0)$ has plateaued and beyond that $U(H_1)$ only decreases. While it seems unlikely that one would prefer the slightly larger potential payoff and much larger penalty of β^- to α^+, that argument implicitly relies on one's prior belief of the two hypotheses. If one was very sure of H_1, then he/she might reasonably choose β^-.

A reasonable condition for choosing a discovery threshold would be to maximize the minimum potential payoff. While this sounds like the *minimax* theorem, it does not stem from the same logic. In that case (keeping $\alpha = \alpha'$) we arrive at:

$$(1-\beta)\cdot U(\text{Discovery}) + \beta \cdot U(\text{Type II}) = (1-\alpha)\cdot U(\text{Limit}) + \alpha \cdot U(\text{Type I}). \qquad \text{(A.24)}$$

Recall that β is a function of α once one has specified $L(x|H_0)$ and $L(x|H_1)$.

Finally, let us consider the same utility function as used in the bottom of Figure A.5 for a channel with an expected significance of 2σ. In that case, $\alpha^+ > \beta^+$ and there is no region in which both utilities are positive (see Figure A.6). If one chose to play this game, he/she would choose between two rather grim situations. Physically, a limit would not be very satisfying, because if the signal were there one would likely commit a Type II error. The situation is similar for discovery. Instead of justifying an optimal discovery threshold, the author suggests to not play this game.

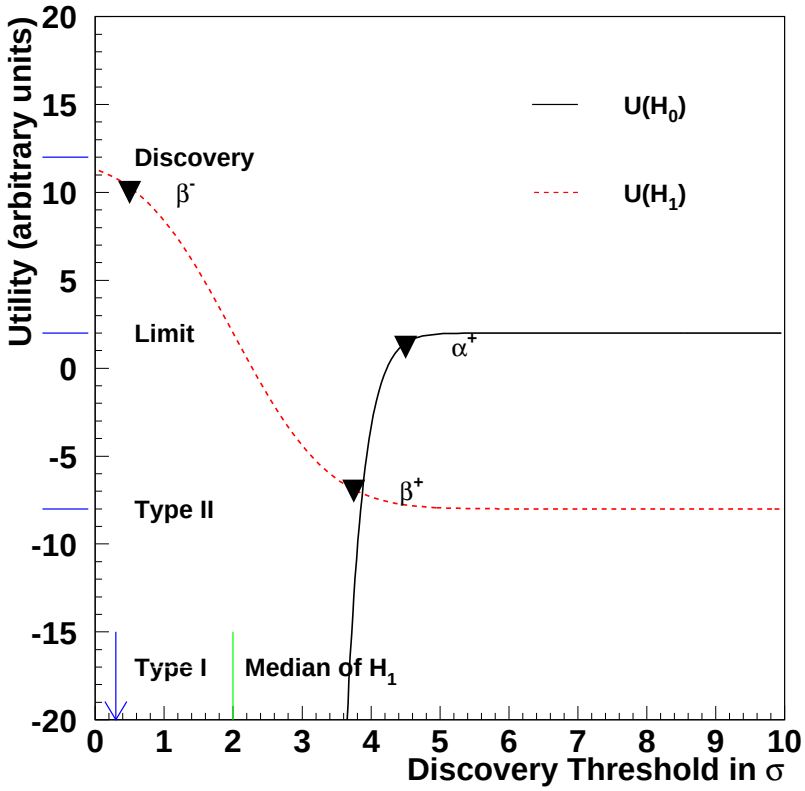


Figure A.6 Utility as a function of discovery threshold for a channel with an expected 2σ significance when the utility for a Type I error is -10^5.

Appendix B: Kernel Estimation Techniques

Perhaps the most common practical duty of a particle physicist is to analyze various distributions from a set of data $\{t_i\}$. The typical tool used in this analysis is the histogram. The role of the histogram is to serve as an approximation of the parent distribution, or probability density function (pdf) from which the data were drawn. While histograms are straightforward and computationally efficient, there are many more sophisticated techniques which have been developed in the last century. One such method, kernel estimation, grew out of a simple generalization of the histogram and has proved to be particularly well-suited for particle physics.

In order to produce continuous estimates $\hat{f}(x)$ of the parent distribution from the *empirical probability density function* $\mathrm{epdf}(x) = \sum_i \delta(x - t_i)$, several techniques have been developed. These techniques can be roughly classified as either parametric or non-parametric. Essentially, a parametric method assumes a model $\hat{f}(x; \vec{\alpha})$ dependent on the parameters $\vec{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \dots)$. The specification of this model is "entirely a matter for the practical [physicist][1]". The goal of a parametric estimate is to optimize the parameters α_i with respect to some goodness-of-fit criterion (*e.g.* χ^2, log-Likelihood, *etc*...). Parametric models are powerful because they allow us to infuse our model with our knowledge of physics. While parametric methods are very powerful, they are highly dependent on the specification of the model. Parametric methods are clearly not practical for estimating the distributions from a wide variety of physical phenomena.

The goal of non-parametric methods is to remove the model-dependence of the estimator. Non-parametric estimates are concerned directly with optimizing the estimate $\hat{f}(x)$. The prototypical non-parametric density estimate is the histogram[2]. Somewhat counterintuitively, non-parametric methods typically involve a large - possibly infinite - number of "parameters" (better thought of as degrees of freedom). Scott and Terrell supplied a more concrete definition of a non-parametric estimator, "Roughly speaking, non-parametric estimators are asymptotically local, while parametric estimators are not."[147] That is to say, the influence of a data point t_i on the density at x should

[1] From a debate between R.A. Fisher and Karl Pearson

[2] The name 'histogram' was coined by Karl Pearson

vanish asymptotically (in the limit of an infinite amount of data) for any $|x - t_i| > 0$ in a non-parametric estimate. The purpose of this paper is to introduce the notion of a kernel estimator and the inherent advantages it offers over other parametric and non-parametric estimators.

B.1 Kernel Estimation

The notion of a kernel estimator grew out of the asymptotic limit of Averaged Shifted Histograms (ASH). The ASH is a simple device that reduces the binning effects of traditional histograms. The ASH algorithm is as follows: First, create a family of N histograms, $\{H_i\}$, with bin-width h, such that the first bin of the i^{th} histogram is placed at $x_0 + ih/N$. Because x_0 is an artificial parameter, each of the H_i is an equally good approximation of the parent distribution. Thus, an obvious estimate of the parent distribution is simply the average of the H_i, hence the name 'Average Shifted Histogram'. Note that resulting estimate (with N times more bins than the original) is not a true histogram, because the height of a 'bin' is not necessarily equal to the number of events falling in that bin. However, it is a superior estimate of the parent distribution, because the dependence of initial bin position is essentially removed. In the limit $N \to \infty$ the ASH is equivalent to placing a triangular shaped *kernel* of probability about each data point t_i [147].

B.1.1 Fixed Kernel Estimation

In the univariate case, the general kernel estimate of the parent distribution is given by

$$\hat{f}_0(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - t_i}{h}\right), \tag{B.1}$$

where $\{t_i\}$ represents the data and h is the smoothing parameter (also called the *bandwidth*). Immediately we can see that our estimate $\hat{f}_0$ is bin-independent regardless of our choice of K. The role of K is to spread out the contribution of each data point in our estimate of the parent distribution. An obvious and natural choice of K is a Gaussian with $\mu = 0$ and $\sigma = 1$:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{B.2}$$

Though there are many choices of K, Gaussian kernels enjoy the attributes of being positive definite, infinitely differentiable, and defined on an infinite support. For physicists this means that our estimate $\hat{f}_0$ is smooth and well-behaved in the tails.

Now we concern ourselves with the choice of the bandwidth h. In Equation B.1, the bandwidth is constant for all i. Thus, $\hat{f}_0$ is referred to as the *fixed kernel estimate*. The role of h is to set the scale for our kernels. Because the kernel method is a non-parametric method, h is completely specified by our data set $\{t_i\}$. In the limit of a large amount ($n \to \infty$) of normally distributed data [147], the *mean integrated squared error* of $\hat{f}_0$ is minimized when

$$h^* = \left(\frac{4}{3}\right)^{1/5} \sigma n^{-1/5}. \tag{B.3}$$

Of course, we rarely deal with normally distributed data, and, unfortunately, the optimal bandwidth h^* is not known in general. In the case of highly bimodal data (*e.g.* the output of a neural network discriminate), the standard deviation of the data is not a good measure for the scale of the true structure of the distribution.

B.1.2 Adaptive Kernel Estimation

An astute reader may object to the choice of h^* given in Equation B.3 on the grounds of self-consistency - non-parametric estimates should only depend on the data locally, and σ is a global quantity. In order for the estimate to handle a wide variety of distributions as well as depend on the data only locally, we must introduce *adaptive kernel estimation*. The only difference in the adaptive kernel technique is that our bandwidth parameter is no longer a global quantity. We require a term that acts as σ_{local} in Equation B.3. Abramson [148] proposed an adaptive bandwidth parameter given by the expression

$$h_i = h/\sqrt{f(t_i)}. \tag{B.4}$$

Equation B.4 reflects the fact that in regions of high density we can accurately estimate the parent distribution with narrow kernels, while in regions of low density we require wide kernels to smooth out statistical fluctuations in our empirical probability density function. Technically we are left with two outstanding issues: i) the expression for h_i given in Equation B.4 references the *a priori*

density, which we do not know, and ii) the optimal choice of h has still not been specified. Clearly, $h^* \propto \sqrt{\sigma}$, because of dimensional analysis. Additionally, $\hat{f}_0$ is our best estimate of the true parent distribution. Thus we obtain

$$\hat{f}_1(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i}K\left(\frac{x-t_i}{h_i}\right), \tag{B.5}$$

with

$$h_i^* = \rho\left(\frac{4}{3}\right)^{1/5}\sqrt{\frac{\sigma}{\hat{f}_0(t_i)}}n^{-1/5}. \tag{B.6}$$

The adaptive kernel estimate can be thought of as a "second iteration" of the general kernel estimation technique. In practice, the adaptive kernel technique almost completely removes any dependence on the original choice of the bandwidth in the fixed kernel estimate $\hat{f}_0$. Furthermore, the adaptive kernel deals very well with multi-modal distributions. In extreme situations (*i.e.* when the scale of the local structure of the data $\sigma_{\rm local}$ is more than about two orders of magnitude smaller than the standard deviation σ of the data) the factor ρ in Equation B.6 should be adjusted from its typical value of unity. In that case

$$\rho = \sqrt{\frac{\sigma_{\rm local}}{\sigma}}. \tag{B.7}$$

We have concluded the construction of a non-parametric estimate $\hat{f}_1$ of an univariate parent distribution based on the empirical probability density function. Our estimate is bin-independent, scale invariant, continuously differentiable, positive definite, and everywhere defined.

B.1.3 Boundary Kernels

Both the fixed and adaptive kernel estimates assume that the domain of the parent distribution is all of $\mathbb{R}$. However, the output of a neural network discriminant, for example, is usually bounded by $0 < x < 1$, where $f(x \leq 0) = f(x \geq 1) \equiv 0$. In order to avoid probability from "spilling out" of the boundaries we must introduce the notion of a *boundary kernel*. Without boundary kernels, our estimate will not be properly normalized and underestimate the true parent distribution close to the boundaries.

Boundary kernels modify our traditional Gaussian kernels so that the total probability in the allowed regions is unity. Clearly, our kernel should smoothly vary back to our original Gaussian

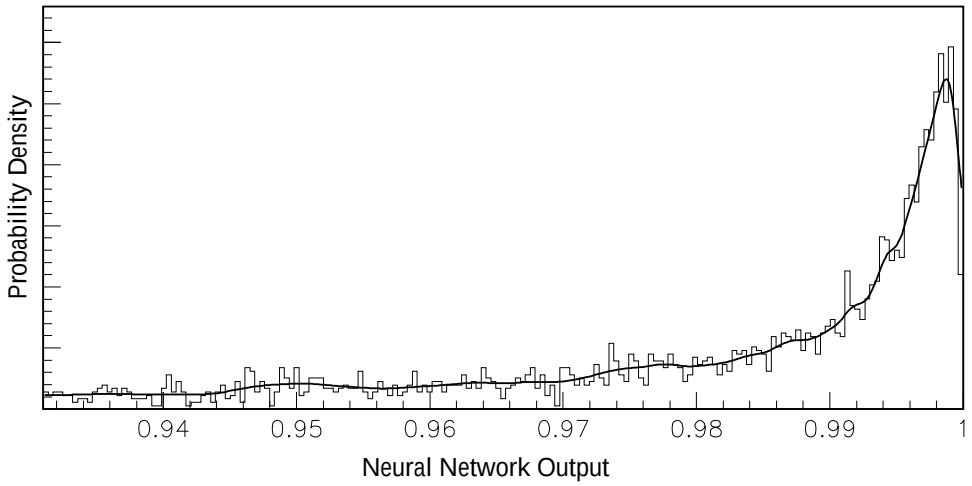


Figure B.1 The performance of boundary kernels on a Neural Network distribution with a hard boundary

kernels as we move far from the boundaries. This constraint quickly reduces the kinds of boundary kernels we need consider. Though a large amount of work has been put forward to introduce kernels which preserve the criteria

$$\int_{-\infty}^{\infty} tK(t)dt = 0, \tag{B.8}$$

these methods are not well suited for physics applications. The primary problem is that the parametrized family of boundary kernels may contain kernels that are not positive definite - which negates their applicability to physics. Also, boundary kernels satisfying Equation B.8 systematically underestimate the parent distribution at a moderate distance from the boundary and overestimate very near the boundary.

An alternate solution to the boundary problem is to simply reflect the data set about the boundary [147]. In that case, the probability that spills out of the boundary is exactly compensated by its mirror.

B.2 Multivariate Kernel Estimation

The general kernel estimation technique generalizes to d-dimensions [147]. One choice for the d-dimensional kernel is simply a product of univariate kernels with independent smoothing parameters. The following discussion will be restricted to the context of such product kernels.

B.2.1 Covariance Issues

When dealing with multivariate density estimation, the covariance structure of the data becomes an issue. Because the covariance structure of the data may not match the diagonal covariance structure of our kernels, we must apply a linear transformation which will diagonalize the covariance matrix Σ_{jk} of the data. Ideally, the transformation would remain a local object; however, in practice such non-linear transformations may be very difficult to obtain. In the remainder of this paper, the transformation matrix will be referred to as A_{jk}, and the $\{\vec{t}_i\}$ will be assumed to be transformed.

B.2.2 Fixed Kernel Estimation

For product kernels, the fixed kernel estimate is given by

$$\hat{f}_0(\vec{x}) = \frac{1}{nh_1 \dots h_d} \sum_{i=1}^{n} \left[\prod_{j=1}^{d} K\left(\frac{x_j - t_{ij}}{h_j} \right) \right]. \tag{B.9}$$

In the asymptotic limit of normally distributed data, the *mean integrated squared error* of $\hat{f}_0$ is minimized when

$$h_j^* = \left(\frac{4}{d+2} \right)^{1/(d+4)} \sigma_j n^{-1/(d+4)}. \tag{B.10}$$

B.2.3 Adaptive Kernel Estimation

The adaptive kernel estimate $\hat{f}_1(\vec{x})$ is constructed in a similar manner as the univariate case; however, the scaling law is usually left in a general form. Because most multivariate data actually lies on a lower dimensional manifold embedded in the input space, the effective dimensionality d' must be found by maximizing some measure of performance or making some assumption. Thus

the multivariate adaptive bandwidth is usually written

$$h_i = h f^{-1/d'}(\vec{t}_i). \tag{B.11}$$

Though $d' \approx d$, the precise value depends on the problem. Note that the form of h_i given in Equation B.11 is independent of j, thus it produces spherically symmetric kernels. This is clearly not optimal. Furthermore, when $d' \neq d$ the optimal value of h may vary wildly. This is because the units are no longer correct and (d/d') powers of scale factors are introduced by $f^{-1/d'}$. Both of these problems may be remedied with the introduction of a natural length scale associated with the data: the geometric mean of the standard deviations of the transformed $\{t_i\}$, $\sigma = \det(A\Sigma A^T)^{1/2d}$. In the absence of local covariance information, the best we can do is assume that the h_j are proportional to σ_j and inversely proportional to $f^{1/d'}$. Thus we arrive at

$$h^*_{ij} = \left(\frac{4}{d+2}\right)^{1/(d+4)} n^{-1/(d+4)} \left(\frac{\sigma_j}{\sigma}\right) \sigma^{(1-d/d')} f^{-1/d'}(\vec{t}_i), \tag{B.12}$$

which is produces estimates that are invariant under linear-transformation of the input space when the covariance matrix is diagonalized.

B.2.4 Multivariate Boundary Kernels

Just as in the univariate case, it is possible that the physically realizable domain of our parent distribution is not all of $\mathbb{R}^d$, but instead a bounded subspace of $\mathbb{R}^d$. Typically, this situation arises when one of the components of the sample vector is bounded in the univariate sense (*i.e.* $t_j < x_j^{\max}$). However, once we diagonalize the covariance matrix of our data the boundary condition will take on a new form in the transformed coordinates. In general, any linear boundary in our original coordinates x_j can be expressed as $c_j x_j = C$, where c_j is the unit-normal to the $(d-1)$-dimensional hyperplane in our d-dimensional domain and C is the distance between the origin and the point-of-closest approach. After transforming to a set of coordinates $x'_j = A_{jk}x_k$, in which the $\{\vec{t}_i\}$ have diagonal covariance, our boundary condition is given by
$d_j x'_j = c_k A^{-1}_{kj} A_{jk} x_k = C$. Thus, for each boundary one must introduce a reflected sample $\{t^{\rm refl}_i\}$ with

$$t^{\rm refl}_{ij} = t_{ij} + 2(C - d_k t_{ik}) d_j, \tag{B.13}$$

in order to rectify the probability that spilled into unphysical regions.

B.2.5 Event-by-Event Weighting

In high-energy physics it is often necessary to combine data from heterogeneous sources (*e.g.* independently produced Monte Carlo data sets which together comprise the Standard Model expectation). In general one would like to estimate the parent distribution from a more general *empirical probability density function* $\mathrm{epdf}(x) = \sum_i w_i \delta(x - t_i)$, where w_i represents the weight or *a posteriori* probability of the i^{th} event. In the case of combining various Monte Carlo samples, one must reweight all events of a sample to some common luminosity (say, 1 pb^{-1}) before combining them. Thus for a Monte Carlo sample with n_{MC} events and cross-section σ_{MC} each event must be weighted with $w_i = 1\ (\mathrm{pb}^{-1})/\mathcal{L}_{\mathrm{eff}} = \sigma_{MC}/n_{MC}$

The covariance matrix of the weighted sample must be generalized as follows:

$$\Sigma_{jk} = \sum_{i=1}^{n} \frac{(t_{ij} - \mu_j)(t_{ik} - \mu_k)}{n} \longrightarrow \tilde{\Sigma}_{jk} = \sum_{i=1}^{n} w_i \frac{(t_{ij} - \tilde{\mu}_j)(t_{ik} - \tilde{\mu}_k)}{\tilde{n}}, \tag{B.14}$$

where $\tilde{n} = \sum_{i=1}^{n} w_i$ and $\tilde{\mu} = \sum_{i=1}^{n} w_i t_i / \tilde{n}$. Then our estimate is simply given by

$$\hat{f}_1(\vec{x}) = \frac{1}{\tilde{n}} \sum_{i=1}^{n} w_i \left[\prod_{j=1}^{d} \frac{1}{\tilde{h}_{ij}} K\left(\frac{x_j - t_{ij}}{\tilde{h}_{ij}} \right) \right]. \tag{B.15}$$

B.3 Use of Kernel Estimation at LEP

As was discussed in the previous Section A.1.5, the LEP Higgs statistical technique took advantage of the shape of signal and background to improve the sensitivity of the searches. The author developed the KEYS package to implement adaptive, univariate kernel estimation for use by the LEP Higgs Working Group.

Figure B.2 shows the standard output of the KEYS for the four jet Higgs channel with a mass of 85 GeV when the reconstructed Higgs mass was used as a discriminating variable.

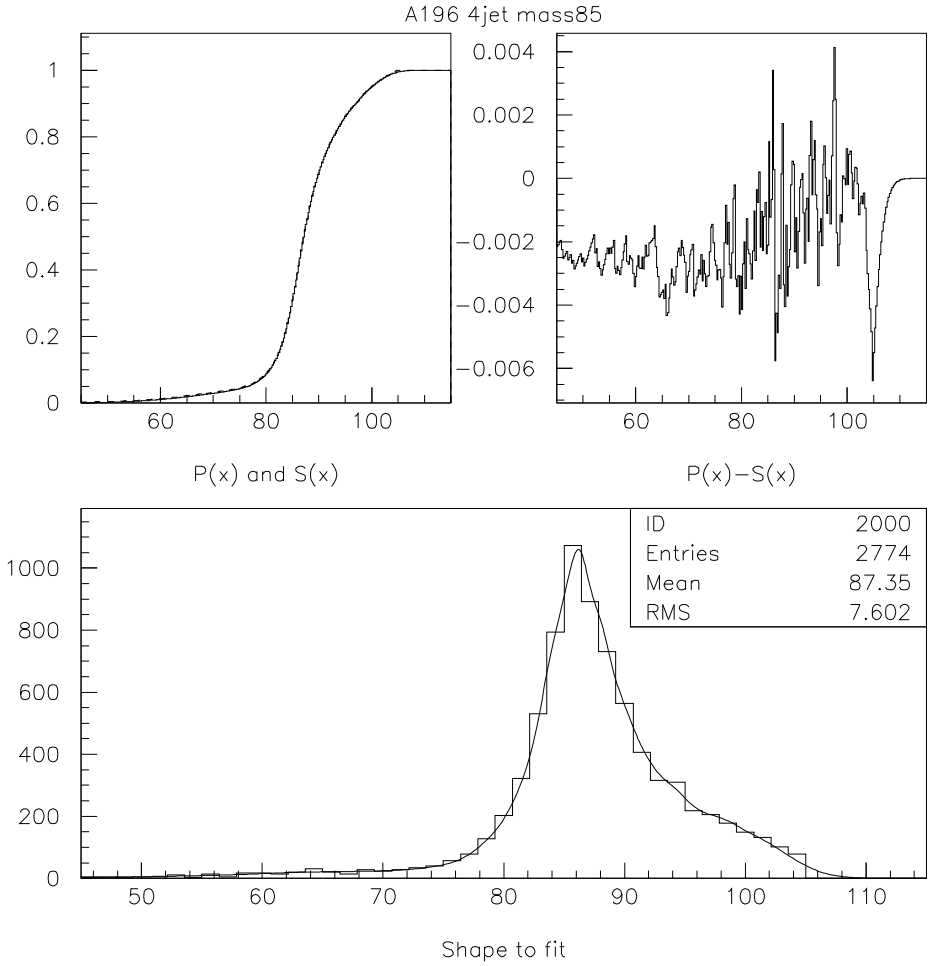


Figure B.2 The standard output of the KEYS script. The top left plot shows the cumulative distributions of the KEYS shape and the data. The top right plot shows the difference between the two cumulative distributions, the maximum of which is used in the calculation of the Kolmogorov-Smirnov test. The bottom plot shows the shape produced by KEYS overlayed on a histogram of the original data.

B.4 Use of Kernel Estimation at BaBar

Another context in which kernel estimation has been applied is the measurement of physical constants via maximum likelihood fitting. Traditionally, the log-Likelihood $\log \mathcal{L} = \sum_i \log f(t_i; \vec{\alpha})$ is maximized with respect to the parameters $\vec{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \dots)$. In this context, $f(t_i; \vec{\alpha})$ is a parametrized model of the physical situation. In practice not all of the α_j are 'floated' or varied in the maximization routine, but instead many parameters are 'fixed' from some independent measurement. While this model incorporates empirical or theoretical information, it may make unwanted assumptions about our data.

For an example, let us consider the measurement of $\sin 2\beta$ at a B factory. The probability density of a CP decay recoiling from a tagged B ($\bar{B}$) meson is given by

$$f(t; \beta) = e^{-\Gamma|t|}(1 \pm \sin 2\beta \sin \Delta m t), \tag{B.16}$$

where t is the time difference between the decay of the CP state and the recoiling B ($\bar{B}$) tagged meson with $\Delta z = \gamma\beta c t$. However, in an experiment we must take into account the mistag rate w and the resolution of Δz. The standard prescription is to measure w and parametrize the resolution distribution $R(\Delta z_{\rm true} - \Delta z_{\rm reco})$ with a single (or double) Gaussian with bias δ and variance σ. The final probability distribution is obtained via a convolution with the resolution function and is of the form $f(t; w, \delta, \sigma, \beta) = R(\delta, \sigma) \otimes f(t; w, \beta)$. Now with w, δ, and σ 'fixed' we must 'float' β to make our measurement [**?**]. Here the form of R, while justified, will have a systematic influence on the measured value of $\sin 2\beta$. If, on the other hand, the resolution function R was estimated via a non-parametric means (*i.e.* kernel estimation techniques), then there would be no artificial influence on the measurement and non-trivial resolution effects would be taken into account automatically.

B.5 Comparison with SMOOTH

It seems appropriate to put kernel estimation techniques in a proper setting before concluding with a discussion of their inherent benefits. Kernel estimation techniques may be applied to situations in which a parametric estimates are popular. Instead, let us consider perhaps the most widely

used non-parametric density estimation technique in high-energy physics: PAW's SMOOTH utility.

B.6 SMOOTH

A full development of the HQUADF function that is used by PAW's SMOOTH utility is beyond the scope of this paper. However, a brief outline of the algorithm is presented. First and foremost, it is important to realize that SMOOTH operates on histograms and not on the original data set $\{t_i\}$. Thus, SMOOTH is dependent on the original binning of the data. SMOOTH was introduced by this journal in John Allison's 1993 paper [**?**]. We will restrict ourselves to the univariate case. Essentially SMOOTH works by finding the bins l of *significant variation* in the histogram $\{h_l\}$ and then using those points to construct a smoothed linear interpolation. Bins of significant variation are those which satisfy $S_l > S^*$, where S^* is a user-defined significance threshold and

$$S_l = \left| \frac{h_{l+1} - 2h_l + h_{l-1}}{\sqrt{Var(h_{l+1}) + 4Var(h_l) + Var(h_{l-1})}} \right|. \tag{B.17}$$

With the points of significant variation $\{x_l\}$ in hand, the smoothed shape is given by

$$s(x) = \sum_l a_l \phi_l(|x - x_l|), \tag{B.18}$$

where $\phi_l(r) = \sqrt{r^2 + \Delta_l^2}$ are the radial basis functions. The Δ_l are user-defined smoothness parameters (radii of curvature). The a_l are found by minimizing the χ^2 between $s(x)$ and the original histogram. As Allison pointed out "lower χ^2 can be obtained by reducing the cut on S_l at the expense of following more of what *might* only be statistical fluctuations." By a different choice of S^* and Δ, the user has the power to magnify or remove statistical fluctuation in the data.

B.7 Comparison

Despite the user-specified parameters S^* and Δ, SMOOTH is a non-parametric estimate of a probability density function based on a set of data. The primary differences between SMOOTH and kernel estimates are their approach and their rigor. While kernel estimates are bin-independent

constructions of the estimate, SMOOTH is a parameter-dependent *fit* of the estimate to a user-provided histogram. Practically speaking, kernel estimates are based on well defined statistical techniques and SMOOTH's estimates are adjusted by eye allowing for user bias and large systematic uncertainty.

B.8 Systematic Errors

When kernel estimation techniques are applied to confidence level calculations or parameter estimation, systematic effects become of particular importance. One may loosely classify the systematic errors associated with probability density estimation as either inherent or user-related errors. In its pure form kernel estimation techniques are entirely deterministic and have no user-specified parameters. If one decides to free the value of ρ from its nominal value of unity (see Equation B.7) or allow $d' \neq d$, then user-related systematic error are introduced. For SMOOTH, the user-related parameters S^* and Δ can not be avoided. In addition to the possible user-related systematic errors, there are inherent systematic errors introduced by any probability density estimation technique. For parametric estimates, this inherent systematic is related to the quality of the model; while for non-parametric estimates, this inherent systematic is related to the flexibility of the technique. The development of kernel estimation techniques has been directly focused on flexibility and the minimization of a particular choice of inherent systematic error: the asymptotic mean integrated squared error [147].

In practice, an experimentalist will want to choose their own estimate of the inherent systematic error (*e.g.* the effect on the measured value of a parameter or 95% confidence level limit). This can be done in a variety of ways that effectively reduce to producing a family of estimates from independent samples of the same parent distribution. This family may be obtained by simply splitting up the data or via toy Monte Carlo simulation. Because the systematic error introduced by the estimation technique is a function(al) of the sampled parent distribution (which is unknown), the estimate itself is the best available choice of the parent distribution to be sampled in a Monte Carlo study.

B.9 Remarks

Obviously, kernel estimation techniques are very powerful and very relevant to high-energy physics. While these techniques have been applied to a wide range of analyses, they seem to be largely unknown by the community.

Appendix C: Hypothesis Testing with Background Uncertainty

In Appendix A we outlined the LEP statistical formalism in the absence of uncertainty on signal and background. In this Appendix, we shall compare several ways of incorporating background uncertainty into the significance calculation.

One encounters both philosophical and technical difficulties when one tries to incorporate uncertainty on the predicted values s and b found in Equation A.16. In a frequentist formalism the unknown s and b become *nuisance parameters*. In a Bayesian formalism, s and b can be marginalized by integration over their respective priors. At LEP the practice was to smear ρ_b and ρ_{s+b} by integrating s and b with a multivariate normal distribution for a posterior. This smearing technique is commonly referred to as the Cousins-Highland technique, and it is has some Bayesian aspects.

In the Section C.1, the Cousins-Highland technique that was implemented by the author into the programs `PoissonSig_syst` and `Likelihood_syst` is presented and critiqued in the context of the LHC. After a brief discussion of nuisance parameters and the Neyman construction, a fully frequentist technique, described in Ref. [139] and implemented by the author, is detailed in Section C.2. In Section C.2.4 other methods for incorporating background uncertainty are outlined. In the remainder of this appendix we compare the various methods in terms of their limiting behavior and a specific example.

C.1 The Cousins-Highland Technique

The Cousins-Highland approach to hypothesis testing is quite popular [44] because it is a simple smearing on the nuisance parameter [142]. In particular, the background-only hypothesis $L(x|H_0, b)$ is transformed from a compound hypothesis with nuisance parameter b to a simple hypothesis $L'(x|H_0)$ by

$$L'(x|H_0) = \int_b L(x|H_0, b)L(b)db, \tag{C.1}$$

where $L(b)$ is typically a normal distribution.

The problem with this method is largely philosophical: $L(b)$ is meaningless in a frequentist formalism. In a Bayesian formalism one can obtain $L(b)$ by considering $L(M|b)$ and inverting it

with the use of Bayes's theorem and the *a priori* likelihood for b. Typically, $L(M|b)$ is normal and one assumes a flat prior on b.

In order to extend the formalism to multiple channels, we introduce the vector quantity $\mathbf{u}$, where u_i is the number of expected events in the i^{th} channel.[1] In general we need a multivariate probability density function to $L(\mathbf{u})$ to accommodate correlated systematic uncertainty between the channels. For instance, if our b-tagging has some uncertainty, then that effect will propagate to the various channels which use b-tagging in a correlated fashion.

To take advantage of the results of Appendix A, we let $\rho_{\mathbf{u}}(q)$ be a generic distribution of the log-likelihood (or any other test-statistic) when we expect u_i events in the i^{th} channel, then the general form of the Cousins-Highland approach to incorporating systematic error is given by[2]:

$$\rho(q) = \int_{u_i \geq 0} \rho_{\mathbf{u}}(q) L(\mathbf{u}) d\mathbf{u} \tag{C.2}$$

The most common form of $L(\mathbf{u})$ is a Gaussian distribution. If we include a correlated error matrix $S_{ij} = \langle u_i - \langle u_i \rangle \rangle \langle u_j - \langle u_j \rangle \rangle$ then Equation C.2 takes the form:

$$\rho(q) = \int_{u_1 \geq 0} \cdots \int_{u_N \geq 0} \rho_{\mathbf{u}}(q) \left(\frac{1}{\sqrt{2\pi}} \right)^N \left(\frac{1}{\sqrt{|S|}} \right) e^{\sum_{i,j=1}^{N} -\frac{1}{2}(u_i - \langle u_i \rangle) S_{ij}^{-1} (u_j - \langle u_j \rangle)} du_i \tag{C.3}$$

Reference [143] provides an analytic expression for the resulting log-likelihood ratio distribution including a correlated error matrix; however, this equation was obtained with an integration over negative numbers of expected events and does not hold.

While the Gaussian form is quite popular, it is not necessarily the most justified. In particular, we impose that the expected number of events satisfies $u_i \geq 0$ while for a Gaussian distribution there is a finite probability of $u_i < 0$. Furthermore, these errors/uncertainties may not be normally distributed even near their mean.

[1] The variable b in the previous section, corresponds to some component of $\mathbf{u}$.

[2] This integral is often referred to as a convolution, however $\rho_{\mathbf{u}}(q)$ is also a function of $\mathbf{u}$ so formally it is not. This integral can be performed with Monte Carlo techniques for an arbitrary $L(\mathbf{u})$.

C.2 Frequentist Methods

Once an analysis has been frozen, the effective cross-section for the background (equivalently, the expected background, b) in that phase-space region is fixed. While the true background might be unknown, in Nature it assumes a unique true value, b_t. To incorporate background uncertainty into a frequentist calculation requires the addition of a *nuisance parameter*. One can not refer to a probability measure on the nuisance parameter; one can only refer to the likelihood of an auxiliary measurement, M, given some value of the nuisance parameter (denoted $L(M|b)$). In practice, the auxiliary measurement is a side-band measurement or a control-sample used to normalize a related Monte Carlo prediction.

The logic for the frequentist method is to simultaneously consider the auxiliary measurement, M, and the test statistic, x, for each value of the nuisance parameter. One performs the Neyman construction: *i.e.* builds $N\sigma$ acceptance regions in the $M - x$ space for the background-only hypothesis. If the measurements M^*, x^* do not fall in the acceptance region, the background-only hypothesis is not consistent with the data at the $N\sigma$ level – which is equivalent to the condition for discovery when $N = 5$.

This technique has been applied to High-Energy Physics in Refs. [149, 150] when the distribution of M and x are both Poissonian. This technique was extended to arbitrary distributions $L(x, M|b)$ by the author and presented at the PhysStat2003 conference [151]. This method relies on the full Neyman construction and uses a likelihood ratio similar to the profile method as an ordering rule. In this formalism, channels with few events are more severely impacted by a systematic uncertainty at the level of 10% than when they are treated with the Cousins-Highland technique. This method is considerably more difficult to implement, and no general-purpose software has been developed.

C.2.1 Nuisance Parameters

Within physics, the majority of the emphasis in statistics has been on limit setting – which can be translated to hypothesis testing through a well known dictionary [139]. When one includes nuisance parameters θ_s (parameters that are not of interest or not observable to the experimenter) into the calculation of a confidence interval, one must insure coverage for every value of the nuisance parameter. When one is interested in hypothesis testing, there is no longer an explicit physics parameter θ_r to cover. Instead one must insure the rate of Type I error is bounded by some predefined value. Analogously, when one includes a nuisance parameters in the null hypothesis, one must insure that the rate of Type I error is bounded for every value of the nuisance parameter. Ideally one can find an acceptance region W which has the same size for all values of the nuisance parameter (*i.e.* a *similar test*). Furthermore, the power of a region W also depends on the nuisance parameter; ideally, we should like to maximize the power for all values of the nuisance parameter (*i.e. Uniformly Most Powerful*). Such tests do not exist in general.

C.2.2 The Neyman-Construction

Usually one does not consider an explicit Neyman construction when performing hypothesis testing between two simple hypotheses; though one exists implicitly. Because of the presence of the nuisance parameter, the implicit Neyman construction must be made explicit and the dimensionality increased. The basic idea is that for each value of the nuisance parameters θ_s, one must construct an acceptance interval (for H_0) in a space which includes their corresponding auxiliary measurements M, and the original test statistic which was being used to test H_0 against H_1. In Appendix **??**, the test statistic was the log-likelihood ratio q. In the following, we will consider an abstract test statistic denoted as x and the expected background rate b as the nuisance parameter.

Let us consider a three-dimensional construction with b, M, and x. For each value of b, one must construct a two-dimensional acceptance region W_b of size α (under H_0). An example construction can be seen in Figure C.1. If an experiment's data (x_0, M_0) fall into an acceptance region W_b, then one cannot exclude the null hypothesis with $100(1-\alpha)\%$ confidence. Conversely, to reject the null hypothesis (*i.e.* claim a discovery) the data must not lie in any acceptance region

Variable	Meaning
θ_r	physics parameters
θ_s	nuisance parameters
$\hat{\theta}_r, \hat{\theta}_s$	unconditionally maximize $L(x\|\hat{\theta}_r, \hat{\theta}_s)$
$\hat{\hat{\theta}}_s$	conditionally maximize $L(x\|\theta_{r0}, \hat{\hat{\theta}}_s)$

Table C.1 The notation used by Kendall for likelihood tests with nuisance parameters

W_b. In other words, to claim a discovery, the confidence interval for the nuisance parameter(s) must be empty (when the construction is made assuming the null hypothesis).

C.2.3 Kendall's Ordering Rule

The basic criterion for discovery was discussed abstractly in the previous section. In order to provide an actual calculation, one must provide an ordering rule: an algorithm which decides how to chose the region W_b. Recall, that there the constraint on Type I error does not uniquely specify an acceptance region for H_0. In the Neyman-Pearson lemma, it is the alternate hypothesis H_1 that breaks the symmetry between possible acceptance regions. The likelihood ratio is used as an ordering rule in the unified approach [152].

At the Workshop on conference limits at FermiLab, Gary Feldman showed that Unified Method with Nuisance Parameters is in Kendall's Theory (the chapter on Likelihood ratio tests & test efficiency) [153]. The notation used by Kendall is given in Table C.1. Also, Kendall identifies H_0 with $\theta_r = \theta_{r0}$ and H_1 with $\theta_r \neq \theta_{r0}$.

Let us briefly quote from Kendall:

> "Now consider the Likelihood Ratio
>
> $$l = \frac{L(x|\theta_{r0}, \hat{\hat{\theta}}_s)}{L(x|\hat{\theta}_r, \hat{\theta}_s)} \tag{C.4}$$
>
> Intuitively l is a reasonable test statistic for H_0: it is the maximum likelihood under H_0 as a fraction of its largest possible value, and large values of l signify that H_0 is reasonably acceptable."

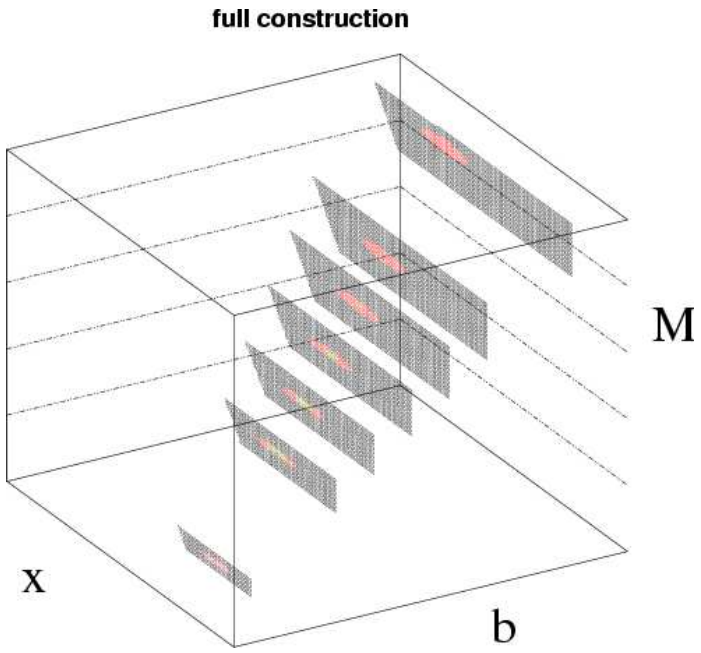


Figure C.1 The Neyman construction for a test statistic x, an auxiliary measurement M, and a nuisance parameter b. Vertical planes represent acceptance regions W_b for H_0 given b. The contours of $L(x, M|H_0, b)$ are in shown in color.

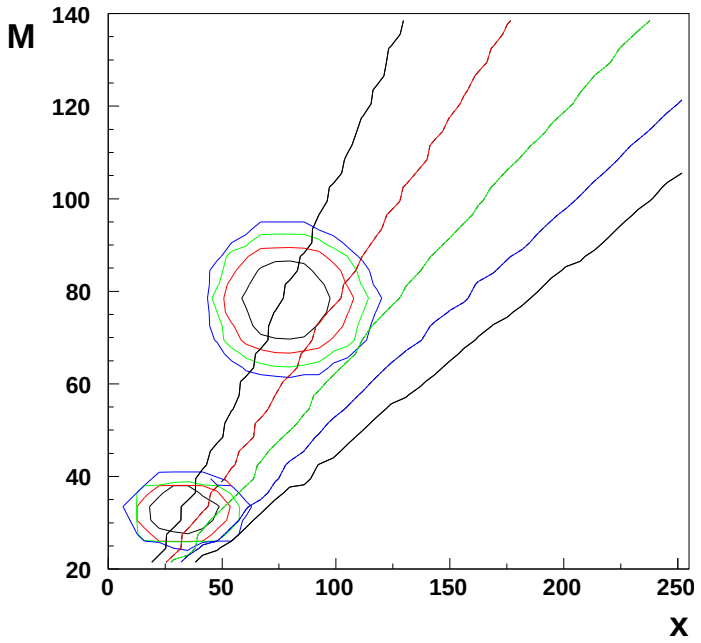


Figure C.2 Contours of the likelihood ratio (diagonal lines) and contours of $L(x, M|H_0, b)$ (concentric ellipses).

Figure C.2 shows contours of the likelihood ratio defined in Equation C.4 as diagonal lines in the $M - x$ plane for the example considered in Section C.4. Contours of the likelihood $L(x, M|H_0, b)$ are shown as concentric ellipses. By specifying the size of the test, one implicitly specifies the likelihood ratio contour which bounds the acceptance region W_b.

Feldman uses this chapter as motivation for the profile method (see Section C.2.4.2), though in Kendall's book the same likelihood ratio is used as an ordering rule *for each value of the nuisance parameter*.

The author tried simple variations on this ordering rule before rediscovering it as written. It is worth pointing out that Equation C.4 is independent of the nuisance parameter b; however, the contour of l_α which provides an acceptance region of size α is not necessarily independent of b. It is also worth pointing out that $\hat{\theta}_r$ and $\hat{\theta}_s$ does not consider the null hypothesis – if it did, the region in which $l = 1$ may be larger than $(1 - \alpha)$. Finally, if one uses θ_s instead of $\hat{\theta}_s$ or $\hat{\hat{\theta}}_s$, one will not obtain tests which are (even approximately) similar.

C.2.4 Other Frequentist Methods

C.2.4.1 The Ratio of Poisson Means

A fully frequentist method for the specific case in which M and x are both Poisson distributed is based on the ratio of their means. In that case, one considers a background and a signal process, both with unknown means. By making "on-source" (*i.e.* x) and "off-source" (*i.e.* M) measurements one can form a confidence interval on the ratio $\lambda = s/b$. If the $100(1 - \alpha)\%$ confidence interval for λ does not include 0, then one could claim discovery. This approach does take into account uncertainty on the background; however, it is restricted to the case in which $L(M|b)$ is a Poisson distribution.

There are two variations on this technique. The first technique has been known for quite some time and was first brought to physics in Ref. [149]. This approach conditions on $x + M$, which allows one to tackle the problem with the use of a binomial distribution. Later, Cousins improved on these limits by removing the conditioning and considering the full Neyman construction [150]. Cousins paper has an excellent review of the literature for those interested in this technique.

C.2.4.2 The Profile Method

As was mentioned in Section C.2.1, the likelihood ratio in Equation C.4 is independent of the nuisance parameters. If it were not for the violations in similarity between tests, one would only need to perform the construction for one value of the nuisance parameters. Clearly, $\hat{\hat{\theta}}_s$ is an appropriate choice to perform the construction. This is the logic behind the profile method. It should be pointed out that the profile method is an approximation to the full Neyman construction; though a particularly good one.

The main advantage to the profile method is that of speed and scalability. Instead of performing the construction for every value of the nuisance parameters, one must only perform the construction once. For many variables, the fully frequentist method is not scalable if one naïvely loops over a fixed grid. However, Monte Carlo sampling the nuisance parameters does not suffer from the curse of dimensionality and serves as a more robust approximation of the full construction than the profile method.

C.2.4.3 "Plus-Minus" Method

Another method that has been used to incorporate systematic error in a frequentist setting is to simply increase the background expectation by some amount and decrease the signal expectation by some amount. For lack of a better name, this method will be referred to as the "plus-minus" method.

There are a few variants on this procedure: *e.g.* not varying the signal expectation. The "plus-minus" and Cousins-Highland methods are truly distinct. Neither method is "more conservative" in general – depending on the number of events and the systematic error either method may produce a lower significance. A comparison of the background confidence level, CL_b, for the two methods in two different scenarios is presented in Figure C.3. The left plot corresponds to an experiment with 100 background events and 10% background uncertainty. The right plot corresponds to an experiment with 35 background events with a 5% background uncertainty. The curves show that without systematic error (green solid line) the CL_b is the lowest (highest significance). Depending

on the experiment, the use of the Cousins-Highland technique (black dashed line) can produce a lower or higher CL_b than the "plus-minus" method (red dotted line).

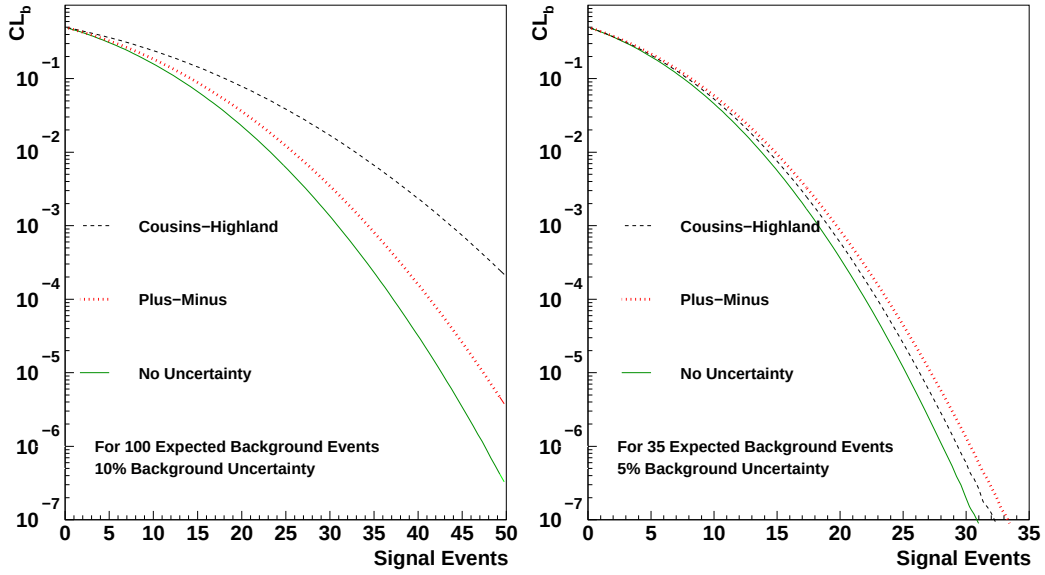


Figure C.3 Comparison of the background confidence level, CL_b, as a function of the number of signal events for different experiments and different methods of incorporating systematic error.

C.2.4.4 Ad Hoc Acceptance Regions

In Section 12.2 a fully frequentist method was presented with an *ad hoc* form for the acceptance region in the $M - x$ plane that was well-suited for that specific problem. Instead of defining the acceptance region with respect to the likelihood ratio, it was simply observed that the contours in Figure C.2 were nearly linear. This motivated the choice of acceptance regions of the form $W = \{x, M | x < M + \eta\sqrt{M}\}$, which are nearly similar (with corrections of order ϵ). Furthermore, the value of η that provides the correct size can be found analytically. From a formal perspective, there is no problem using an *ad hoc* acceptance region – it just might not be the most powerful.

C.3 Saturation of Significance

An important feature of the incorporation of systematic error with the Cousins-Highland technique is the saturation of signal significance. What is meant by "saturation" is that the significance reaches an asymptotic value as more data are collected (holding the systematic error fixed). Consider a channel in which the background normalization has a relative uncertainty α. In the limit of large integrated luminosity, the natural statistical variation becomes negligible compared to the systematic error, and the signal significance approaches a constant σ_∞.

The Cousins-Highland Technique

In the Cousins-Highland technique, one can calculate the saturation exactly. As more events are collected, the Gaussian approximation of a Poisson distribution is valid, thus

$$\sigma_\infty^{CH} = \lim_{L\to\infty} \frac{s}{\sqrt{b(1+\alpha^2 b)}} = \frac{s/b}{\alpha}. \tag{C.5}$$

The above equation is somewhat misleading because there are implicit restrictions on the Cousins-Highland approach with a Gaussian $L(\mathbf{u})$. To claim an observation is inconsistent with the background at the $N\sigma$ level, we must be able to describe the background at the $N\sigma$ level. If the $N\sigma$ contours of our background description include the unphysical prediction $b < 0$, we know our background description is failing. Thus the Cousins-Highland approach is internally inconsistent for $\alpha \gtrsim 1/N\sigma > 1/\sigma_\infty^{CH}$. In particular, the background must be known within 20% to achieve a 5σ effect.

The Frequentist Technique

The frequentist significance calculation is very difficult; however, the limiting behavior of σ_∞ can be derived geometrically. Using the likelihood-ratio as an ordering rule, observing that this produces approximately similar tests, and observing that the contours of this likelihood have a simple form one arrives at:

$$\frac{s}{M} = \frac{\sigma_\infty^F \Delta}{1-\sigma_\infty^F \Delta}. \tag{C.6}$$

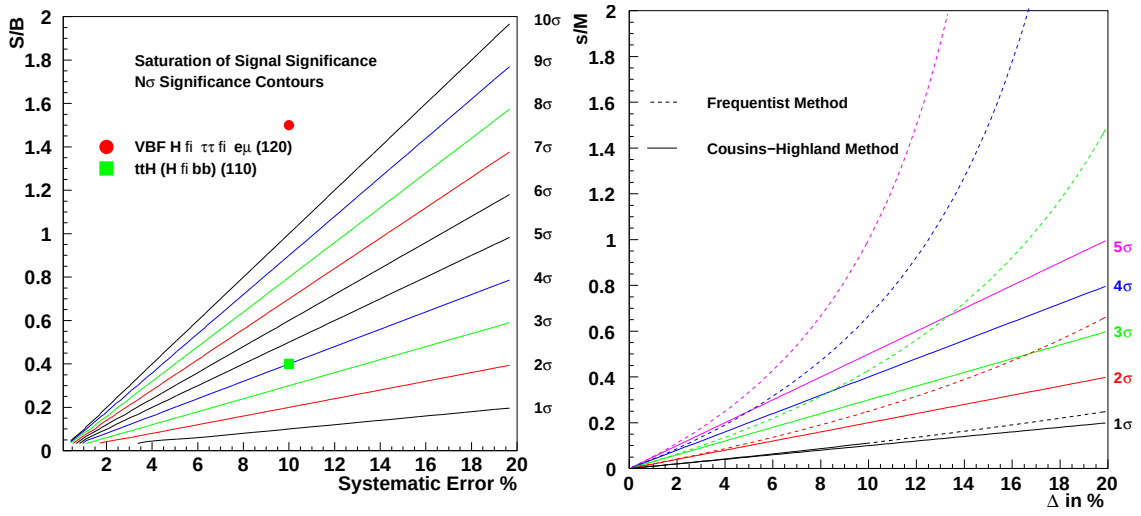


Figure C.4 Contours of σ_{∞}^{CH} in the plane of signal-to-background ratio vs. the systematic error α in percent (left) and comparison with the frequentist technique (right).

The frequentist method shows explicitly that as $\Delta \rightarrow 1/\sigma_{\infty}$, the required $s/M \rightarrow \infty$. This equation can be rewritten as

$$\sigma_{\infty}^{F} = \frac{s}{\Delta(s+M)} = \frac{s/M}{\Delta(1+s/M)}. \tag{C.7}$$

Figure C.4 compares the contours of σ_{∞} for the Cousins-Highland and fully frequentist methods.

The Plus-Minus Technique

It is worth noting that this saturation feature is not present in the "plus-minus" method. In that case,

$$\sigma_{\infty}^{\pm} = \lim_{L\rightarrow\infty} \frac{s(1-\alpha)}{\sqrt{b(1+\alpha)}} = \infty. \tag{C.8}$$

This behavior is not surprising, the "plus-minus" method is equivalent to assuming one knows the background exactly (it just happens to be more than one originally expected).

C.4 An Example

Let us consider the case when the nuisance parameter is the expected number of background events b and M is an auxiliary measurement of b. Furthermore, let us assume that we have a absolute prediction of the number of signal events s. For our test statistic we choose the number of events observed x which is Poisson distributed with mean $\mu = b$ for H_0 and $\mu = s + b$ for H_1. In the construction there are no assumptions about $L(M|H_0, b)$ – it could be some very complicated shape relating particle identification efficiencies, Monte Carlo extrapolation, etc. In the case where $L(M|H_0, b)$ is a Poisson distribution, other solutions exist (see Section **??**). For our example, let us take $L(M|H_0, b)$ to be a Normal distribution centered on b with standard deviation Δb, where Δ is some relative systematic error. Additionally, let us assume that we can factorize $L(x, M|H, b) = L(x|H, b)L(M|b)$ (where H is either H_0 or H_1).

The Frequentist Approach with Kendall's Ordering Rule

For our example problem, we can re-write the ordering rule in Equation C.4 as

$$l = \frac{L(x, M|H_0, \hat{\hat{b}})}{L(x, M|H_1, \hat{b})}, \tag{C.9}$$

where $\hat{b}$ conditionally maximizes $L(x, M|H_1, b)$ and $\hat{\hat{b}}$ conditionally maximizes $L(x, M|H_0, b)$.

Now let us take $s = 50$ and $\Delta = 5\%$, both of which were determined from Monte Carlo. In our toy example, we collect data $M_0 = 100$. Let $\alpha = 2.85 \cdot 10^{-7}$, which corresponds to 5σ. The question now is how many events x must we observe to claim a discovery?[3] The condition for discovery is that (x_0, M_0) do not lie in any acceptance region W_b. In Fig. C.1 a sample of acceptance regions are displayed. One can imagine a horizontal plane at $M_0 = 100$ slicing through the various acceptance regions. The condition for discovery is that $x_0 > x_{\max}$ where $x_{\max}$ is the maximal x in the intersection.

There is one subtlety which arises from the ordering rule in Equation C.9. The acceptance region $W_b = \{(x, M) \mid l > l_\alpha\}$ is bounded by a contour of the likelihood ratio and must satisfy the constraint of size: $\int_{W_b} L(x, M|H_0, b) = (1 - \alpha)$. While it is true that the likelihood is

[3] In practice, one would measure x_0 and M_0 and then ask, "have we made a discovery?". For the sake of explanation, we have broken this process into two pieces.

independent of b, the constraint on size *is* dependent upon b. Similar tests are achieved when l_α is independent of b. The contours of the likelihood ratio are shown in Fig. C.2 together with contours of $L(x, M|H_0, b)$. Contours of the likelihood $L(x, M|H_0, b)$ are shown as concentric ellipses for $b = 32$ and $b = 80$. While tests are roughly similar for $b \approx M$, similarity is violated for $M \ll b$. This violation should be irrelevant because clearly $b \ll M$ should not be accepted. This problem can be avoided by clipping the acceptance region around $M = b \pm N\Delta b$, where N is sufficiently large (≈ 10) to have negligible affect on the size of the acceptance region. Fig. C.1 shows the acceptance region with this slight modification.

In the case where $s = 50$, $\Delta = 5\%$, and $M_0 = 100$, one must observe 167 events to claim a discovery. While no figure is provided, the range of b consistent with $M_0 = 100$ (and no constraint on x) is $b \in [68, 200]$. In this range, the tests are similar to a very high degree.

The Profile Method

In the example above with $x_0 = 167$, $M_0 = 100$, the construction would be made at $b = \hat{\hat{b}} = 117$ which gives the identical results of the fully frequentist method with the likelihood ratio as an ordering rule.

The Cousins-Highland Technique

In the case where $s = 50$, $L(b)$ is a normal distribution with mean $\mu = M_0 = 100$ and standard deviation $\sigma = \Delta M_0 = 5$, one must observe 161 events to claim a discovery. Initially, one might think that 161 is quite close to 167; however, they differ at the 4% level and the methods are only considering a $\Delta = 5\%$ effect. Still worse, if H_0 is true (say $b_t = 100$) and one can claim a discovery with the Cousins-Highland method ($x_0 > 161$), the chance that one could not claim a discovery with the fully frequentist method ($x_0 < 167$) is $\approx 95\%$. Similarly, if H_1 is true and one can claim a discovery with the Cousins-Highland method, the chance that one could not claim a discovery with the fully frequentist method is $\approx 50\%$. Even practically, there is quite a difference between these two methods.

Appendix D: Statistical Learning Theory Applied to Searches

Multivariate Analysis is an increasingly common tool in experimental high energy physics; however, most of the common approaches were borrowed from other fields. Each of these algorithms were developed for their own particular task, thus they look quite different at their core. It is not obvious that what these different algorithms do internally is optimal for the the tasks which they perform within high energy physics. It is also quite difficult to compare these different algorithms due to the differences in the formalisms that were used to derive and/or document them. In Section D.1 we introduce a formalism for a *Learning Machine*, which is general enough to encompass all of the techniques used within high energy physics. We review the statistical statements relevant to new particle searches and translate them into the formalism of statistical learning theory.

D.1 Formalism

Formally a Learning Machine is a family of functions $\mathcal{F}$ with domain I and range O parametrized by $\alpha \in \Lambda$. The domain can usually be thought of as, or at least embedded in, $\mathbb{R}^d$ and we generically denote points in the domain as x. The points x can be referred to in many ways (*e.g.* patterns, events, inputs, examples, ...). The range is most commonly $\mathbb{R}$, $[0,1]$, or just $\{0,1\}$. Elements of the range are denoted by y and can be referred to in many ways (*e.g.* classes, target values, outputs, ...). The parameters α specify a particular function $f_\alpha \in \mathcal{F}$ and the structure of $\alpha \in \Lambda$ depends upon the learning machine [154, 155].

In the modern theory of machine learning, the performance of a learning machine is usually cast in the more pessimistic setting of *risk*. In general, the risk, R, of a learning machine is written as

$$R(\alpha) = \int Q(x, y; \alpha)\, p(x, y) dx dy \tag{D.1}$$

where Q measures some notion of *loss* between $f_\alpha(x)$ and the target value y. For example, when classifying events, the risk of mis-classification is given by Eq. D.1 with $Q(x, y; \alpha) = |y - f_\alpha(x)|$.

Similarly, for regression[1] tasks one takes $Q(x, y; \alpha) = (y - f_\alpha(x))^2$. Most of the classic applications of learning machines can be cast into this formalism; however, searches for new particles place some strain on the notion of risk.

D.1.1 Machine Learning

The starting point for machine learning is to accept that we might not know $p(x, y)$ in any analytic or numerical form. This is, indeed, the case for particle physics, because only $\{(x, y)_i\}$ can be obtained from the Monte Carlo convolution of a well-known theoretical prediction and complex numerical description of the detector. In this case, the learning problem is based entirely on the training samples $\{(x, y)_i\}$ with l elements. The risk functional is thus replaced by the *empirical risk functional*

$$R_{\text{emp}}(\alpha) = \frac{1}{l} \sum_{i=1}^{l} Q(x_i, y_i; \alpha). \tag{D.2}$$

One then must try to approximate $f_{\alpha_0} \in \mathcal{F}$, that minimizes the true risk, by the function f_{α_l}, that minimizes the empirical risk. This is approach is called the *empirical risk minimization* (ERM) inductive principle.

Vapnik outlines the four parts of learning theory in [155]:

1. What are the (necessary and sufficient) conditions for consistency of a learning process based on the ERM principle?

2. How fast is the rate of convergence of the learning process?

3. How can one control the rate of convergence (the generalization ability) of the learning process?

4. How can one construct algorithms that can control the generalization ability?

[1] During the presentation, J. Friedman did not distinguish between these two tasks; however, in a region with $p(x, 1) = b$ and $p(x, 0) = 1 - b$, the optimal $f(x)$ for classification and regression differ. For classification, $f(x) = \{1 \text{ if } b > 1/2, \text{else } 0\}$, and for regression the optimal $f(x) = b$.

Answering question (1) is achieved by considering the notion of non-trivial consistency. The details of the discussion are beyond the scope of this note, but consistency is essentially a guarantee that with an infinite amount of training data ($l \rightarrow \infty$) the ERM principle will produce a function with equal risk to f_{α_0}. Interestingly, the necessary and sufficient conditions for non-trivial consistency are analogous to Popper's theory of non-falsifiability in the philosophy of science. In particular, Vapnik introduces a quantity h that is a property of a learning machine $\mathcal{F}$ and called the Vapnik-Chervonenkis (VC) dimension. Simply put, the conditions for (1) are that h is finite.

The VC dimension of $\mathcal{F}$ is defined as the maximal cardinality of a set which can be shattered by $\mathcal{F}$. "A set $\{x_i\}$ can be shattered by $\mathcal{F}$" means that for each of the 2^h binary classifications of the points $\{x_i\}$, there exists a $f_\alpha \in \mathcal{F}$ which satisfies $y_i = f_\alpha(x_i)$. A set of three points can be shattered by an oriented line as illustrated in Figure D.2. Note that for a learning machine with VC dimension h, not every set of h elements must be shattered by $\mathcal{F}$, but at least one.

The answer to question (2) is the surprising result that there are bounds on the true risk $R(\alpha)$, which are independent of the distribution $p(x, y)$. In particular, for $0 \leq Q(x, y; \alpha) \leq 1$

$$R(\alpha) \leq R_{\text{emp}}(\alpha) + \sqrt{\left(\frac{h(\log(2l/h) + 1) - \log(\eta/4)}{l}\right)}, \tag{D.3}$$

where h is the Vapnik-Chervonenkis (VC) dimension and η is the probability that the bound is violated. As $\eta \rightarrow 0$, $h \rightarrow \infty$, or $l \rightarrow 0$ the bound becomes trivial.

Equation D.3 is a remarkable result which relates the number of training examples l, the fundamental property of the learning machine h, and the risk R independent of the unknown distribution $p(x, y)$. The bounds provided by Equation D.3 are relatively week due to their stunning generality. More important than their weakness, is the realization that with an independent testing sample one can evaluate the true risk arbitrarily well. This testing sample, by definition, is not known to the algorithm, so the bound is useful for the design of algorithms encountered in the 4^{th} part of Vapnik's theory. Neural Network and most other methods, however, rely on an independent testing sample to aid in their design.

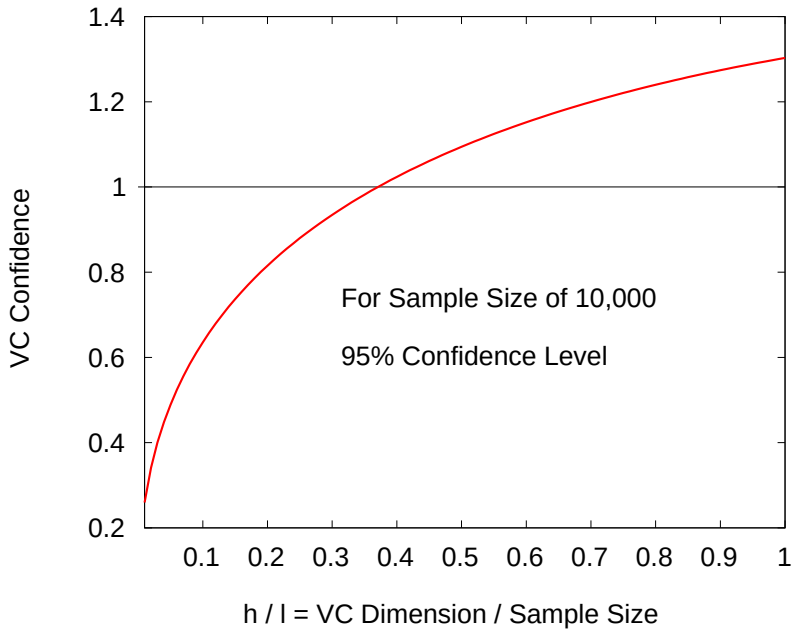


Figure D.1 The VC Confidence as a function of h/l for $l = 10,000$ and $\eta = 0.05$. Note that for $l < 3h$ the bound is non-trivial and for $l < 20h$ is quite tight.

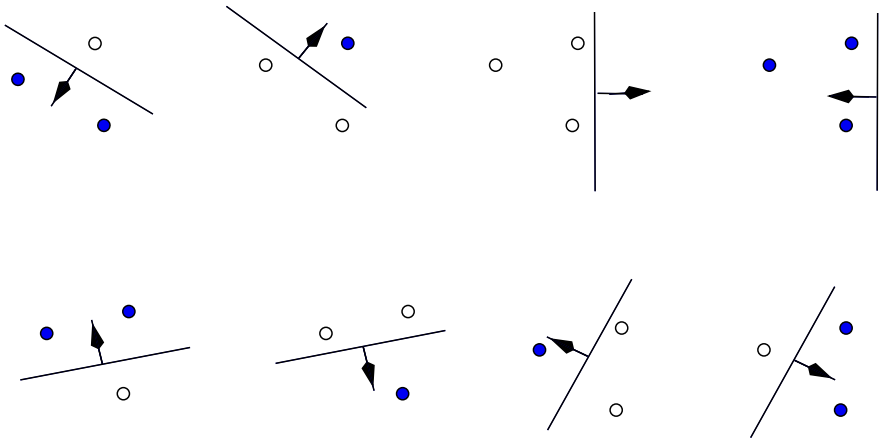


Figure D.2 Example of an oriented line shattering 3 points. Solid and empty dots represent the two classes for y and each of the 2^3 permutations are shown.

D.2 The Neyman-Pearson Theory in the Context of Risk

In Section D.1 we provided the loss functional Q appropriate for the classification and regression tasks; however, we did not provide a loss functional for searches for new particles.

Once the size of the test, α, has been agreed upon, the notion of risk is the probability of Type II error β. In order to return to the formalism outlined in Section D.1, identify H_1 with $y = 1$ and H_0 with $y = 0$. Let us consider learning machines that have a range $\mathbb{R}$ which we will compose with a step function $\tilde{f}(x) = \Theta(f_\alpha(x) - k_\alpha)$ so that by adjusting k_α we insure that the acceptance region W has the appropriate size. The region W is the acceptance region for H_0, thus it corresponds to $W = \{x|\tilde{f}(x) = 0\}$ and $I - W = \{x|\tilde{f}(x) = 1\}$. We can also translate the quantities $p(x|H_0)$ and $p(x|H_1)$ into their learning-theory equivalents $p(x|0) = p(x,0)/p(0) = \delta(y)p(x,y)/\int p(x,0)dx$ and $\delta(1-y)p(x,y)/\int p(x,1)dx$, respectively. With these substitutions we can rewrite the Neyman-Pearson theory as follows. A fixed size gives us the global constraint

$$\alpha = \frac{\int \Theta(f_\alpha(x) - k_\alpha)\ \delta(y)\ p(x,y))dxdy}{\int p(x,0)dx} \tag{D.4}$$

and the risk is given by

$$\begin{aligned} \beta &= \frac{\int [1 - \Theta(f_\alpha(x) - k_\alpha)]\ p(x,1)dx}{\int p(x,1)dx} \\ &\propto \int \Theta(f_\alpha(x) + k_\alpha)\ \delta(1-y)\ p(x,y)dxdy. \end{aligned} \tag{D.5}$$

Extracting the integrand we can write the loss functional as

$$Q(x,y;\alpha) = \Theta(f_\alpha(x) + k_\alpha)\ \delta(1-y). \tag{D.6}$$

Unfortunately, Eq. D.1 does not allow for the global constraint imposed by k_α (which is implicitly a functional of f_α), but this could be accommodated by the methods of Euler and Lagrange. Furthermore, the constraint cannot be evaluated without explicit knowledge of $p(x,y)$.

D.3 Asymptotic Equivalence

Certain approaches to multivariate analysis leverage the many powerful theorems of statistics assuming one can explicitly refer to $p(x,y)$. This dependence places a great deal of stress on

the asymptotic ability to estimate $p(x, y)$ from a finite set of samples $\{(x, y)_i\}$. There are many such techniques for estimating a multivariate density function $p(x, y)$ given the samples [147, 145]. Unfortunately, for high dimensional domains, the number of samples needed to enjoy the asymptotic properties grows very rapidly; this is known as the *curse of dimensionality*.

In the case that there is no (or negligible) interference between the signal process and the background processes one can avoid the complications imposed by quantum mechanics and simply add probabilities. This is often the case with searches for new particles, thus the signal-plus-background hypothesis can be rewritten $p(x, |H_1) = n_s p_s(x) + n_b p_b(x)$, where n_s and n_b are normalization constants that sum to unity. This allows us to rewrite the contours of the likelihood ratio as contours of the signal-to-background ratio. In particular the contours of the likelihood ratio $p(x|H_1)/p(x|H_0) = k_\alpha$ can be rewritten as $p_s(x)/p_b(x) = (k_\alpha - n_b)/n_s = k'_\alpha$.

D.4 Direct vs. Indirect Methods

The loss functional defined in Eq. D.6 is derived from a minimization on the rate of Type II error. This is logically distinct from, but asymptotically equivalent to, approximating the likelihood ratio. In the case of no interference, this is logically distinct from, but asymptotically equivalent to, approximating the signal-to-background ratio. In fact, most multivariate algorithms are concerned with approximating an auxiliary function that is one-to-one with the likelihood ratio. Because the methods are not directly concerned with minimizing the rate of Type II error, they should be considered *indirect methods*. Furthermore, the asymptotic equivalence breaks down in most applications, and the indirect methods are no longer optimal. Neural networks, kernel estimation techniques, and support vector machines all represent indirect solutions to the search for new particles. The Genetic Programming (GP) approach presented in Appendix E is a *direct method* concerned with optimizing a user-defined performance measure.

D.5 VC Dimension of Neural Networks

In order to apply Eq. D.3, one must determine the VC dimension of neural networks. This is a difficult problem in combinatorics and geometry aided by algebraic techniques. Eduardo Sontag has an excellent review of these techniques and shows that the VC dimension of neural networks can, thus far, only be bounded fairly weakly [156]. In particular, if we define ρ as the number of weights and biases in the network, then the best bounds are $\rho^2 < h < \rho^4$. In a typical particle physics neural network one can expect $100 < \rho < 1000$, which translates into a VC dimension as high as 10^{12}, which implies $l > 10^{13}$ for reasonable bounds on the risk. These bounds imply enormous numbers of training samples when compared to a typical training sample of 10^5. Sontag goes on to show that these shattered sets are incredibly special and that the set of all shattered sets of cardinality $\mu > 2\rho + 1$ is measure zero in general. Thus, perhaps a more relevant notion of the VC dimension of a neural network is given by μ.

D.6 Conclusions

Multivariate algorithms are obviously an extremely useful tool in data analysis. The more germane concern for physicists is what are the relevant properties of a multivariate algorithm for their particular application. In this note we have considered three common applications: classification, regression, and the search for new particles. For the three main approaches to multivariate analysis, we have distinguished between their asymptotic and non-asymptotic properties, established relationships among the approaches, and presented the key theorems in their fundamental theories. Particular emphasis has been placed on the Neyman-Pearson setting for the interpretation of searches for new particles and the development of an appropriate notion of risk. We have considered several common multivariate algorithms and indicated their strengths and weaknesses. The final conclusions as to which multivariate algorithms are most appropriate for a given task will remain as much an experiment in human psychology as mathematical rigor.

Appendix E: Genetic Programming for Event Selection

The use of Genetic Programming for classification is fairly limited; however, it can be traced to the early works on the subject by Koza [157]. More recently, Kishore *et al*. extended Koza's work to the multicategory problem [158]. To the best of the author's knowledge PHYSICSGP, the implementation documented here and in Ref [135], is the first use of Genetic Programming within High Energy Physics[1]. PHYSICSGP was developed in collaboration with R. Sean Bowman.

In Section E.1 we provide a brief history of evolutionary computation and distinguish between Genetic Algorithms (GAs) and Genetic Programming (GP). We describe our algorithm in detail for an abstract performance measure in Section E.2 and discuss several specific performance measures in Section E.3.

Close attention is paid to the performance measure in order to leverage recent work applying the various results of statistical learning theory in the context of new particle searches. This recent work consists of two components. In the first, the Neyman-Pearson theory is translated into the *Risk* formalism [159, 160]. The second component requires calculating the Vapnik-Chervonenkis dimension for the learning machine of interest. In Section E.3.1, we calculate the Vapnik-Chervonenkis dimension for our Genetic Programming approach.

E.1 Evolutionary Computation

In Genetic Programming, a group of "individuals" evolve and compete against each other with respect to some performance measure. The individuals represent potential solutions to the problem at hand, and evolution is the mechanism by which the algorithm optimizes the population. The performance measure is a mapping that assigns a fitness value to each individual. GP can be thought of as a Monte Carlo sampling of a very high dimensional search space, where the sampling is related to the fitness evaluated in the previous generation. The sampling is not ergodic – each

[1] About two years after the initial development of PHYSICSGP, Eric Vaandering presented a very similar implementation of genetic programming at CHEP2004. Dr. Vaandering's work appears to be an independent development reaching similar conclusions.

generation is related to the previous generations – and intrinsically takes advantage of stochastic perturbations to avoid local extrema[2].

Genetic Programming is similar to, but distinct from Genetic Algorithms (GAs), though both methods are based on a similar evolutionary metaphor. GAs evolve a bit string which typically encodes parameters to a pre-existing program, function, or class of cuts, while GP directly evolves the programs or functions. For example, Field and Kanev [161] used Genetic Algorithms to optimize the lower- and upper-bounds for six 1-dimensional cuts on Modified Fox-Wolfram "shape" variables. In that case, the phase-space region was a pre-defined 6-cube and the GA was simply evolving the parameters for the upper and lower bounds. On the other hand, our algorithm is not constrained to a pre-defined shape or parametric form. Instead, our GP approach is concerned directly with the construction and optimization of a nontrivial phase space region with respect to some user-defined performance measure.

In this framework, particular attention is given to the performance measure. The primary interest in the search for a new particle is hypothesis testing, and the most relevant measures of performance are the expected statistical significance (usually reported in Gaussian sigmas) or limit setting potential. The different performance measures will be discussed in Section E.3, but consider a concrete example: $s/\sqrt{b}$, where s and b are the number of signal and background events satisfying the event selection, respectively.

E.2 The Genetic Programming Approach

While the literature is replete with uses of Genetic Programming and Genetic Algorithms, direct evolution of cuts appears to be novel. In the case at hand, the individuals are composed of simple arithmetic expressions, f, on the input variables $\vec{v}$. Without loss of generality, the cuts are always of the form $-1 < f(\vec{v}) < 1$. By scaling, $f(\vec{v}) \to af(\vec{v})$, and translation, $f(\vec{v}) \to f(\vec{v}) + b$, of these expressions, single- and double-sided cuts can be produced. An individual may consist of one or more such cuts combined by the Boolean conjunction `AND`. Fig. E.1 shows the signal and

[2] These are the properties that give power to Markov Chain Monte Carlo techniques.

background distributions of four expressions that make up the most fit individual in a development trial.

Due to computational considerations, several structural changes have been made to the naïve implementation. First, an Island Model of parallelization has been implemented (see Section E.2.5). Secondly, individuals' fitness can be evaluated on a randomly chosen sub-sample of the training data, thus reducing the computational requirements at the cost of statistical variability. There are several statistical considerations which are discussed in Reference [135].

E.2.1 Individual Structure, Mutation, and Crossover

The genotype of an individual is a collection of expression trees similar to abstract syntax trees that might be generated by a compiler as an intermediate representation of a computer program. An example of such a tree is shown in Fig. E.2a which corresponds to a cut $|4.2v_1 + v_2/1.5| < 1$. Leafs are either constants or one of the input variables. Nodes are simple arithmetic operators: addition, subtraction, multiplication, and safe division[3].When an individual is presented with an event, each expression tree is evaluated to produce a number. If all these numbers lie within the range $(-1, 1)$, the event is considered signal. Otherwise the event is classified as background.

Initial trees are built using the PTC1 algorithm described in [162]. After each generation, the trees are modified by *mutation* and *crossover*. Mutation comes in two flavors. In the first, a randomly chosen expression in an individual is scaled or translated by a random amount. In the second kind of mutation, a randomly chosen subtree of a randomly chosen expression is replaced with a randomly generated expression tree using the same algorithm that is used to build the initial trees.

While mutation plays an important rôle in maintaining genetic diversity in the population, most new individuals in a particular generation result from crossover. The crossover operation takes two individuals, selects a random subtree from a random expression from each, and exchanges the two. This process is illustrated in Fig. E.2.

[3]Safe division is used to avoid division by zero.

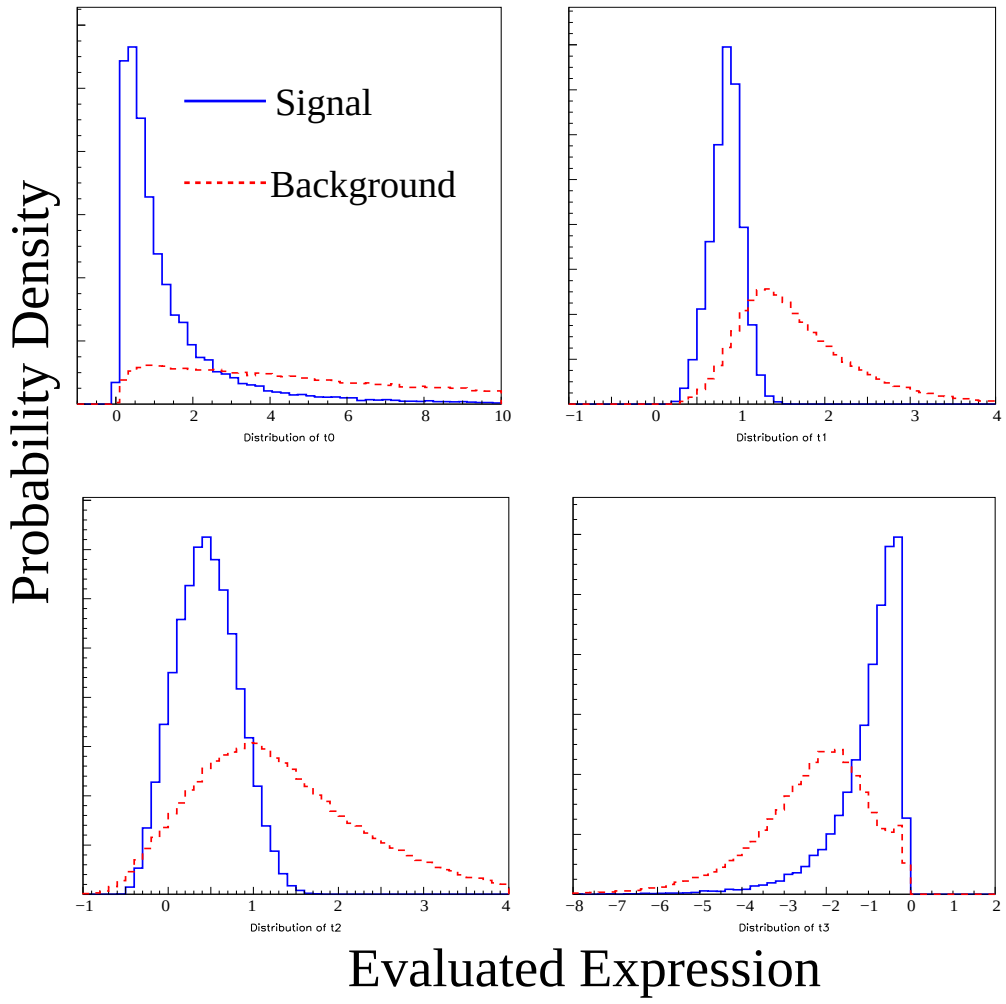


Figure E.1 Signal and Background histograms for an expression.

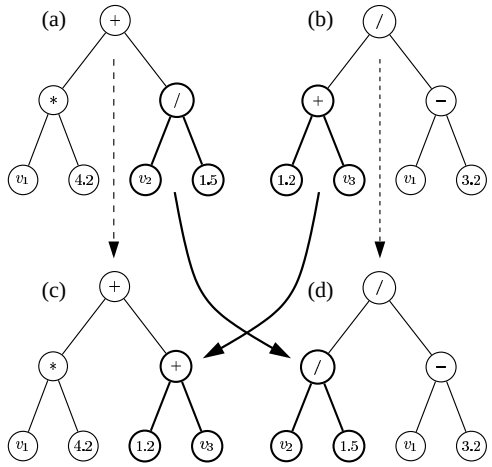


Figure E.2 An example of crossover. At some given generation, two parents (a) and (b) are chosen for a crossover mutation. Two subtrees, shown in bold, are selected at random from the parents and are swapped to produce two children (c) and (d) in the subsequent generation.

E.2.2 Recentering

Some expression trees, having been generated randomly, may prove to be useless since the range of their expressions over the domain of their inputs lies well outside the interval $(-1, 1)$ for every input event. When an individual classifies all events in the same way (signal or background), each of its expressions is translated to the origin for some randomly chosen event exemplar $\vec{v_0}$, *viz.* $f(\vec{v}) \rightarrow f(\vec{v}) - f(\vec{v_0})$. This modification is similar to, and thus reduces the need for, normalizing input variables.

E.2.3 Fitness Evaluation

Fitness evaluation consumes the majority of time in the execution of the algorithm. So, for speed, the fitness evaluation is done in C. Each individual is capable of expressing itself as a fragment of C code. These fragments are pieced together by the Python program, written to a file, and compiled. After linking with the training vectors, the program is run and the results communicated back to the Python program using standard output.

The component that serializes the population to C and reads the results back from the generated C program is configurable, so that a user-defined performance measure may be implemented.

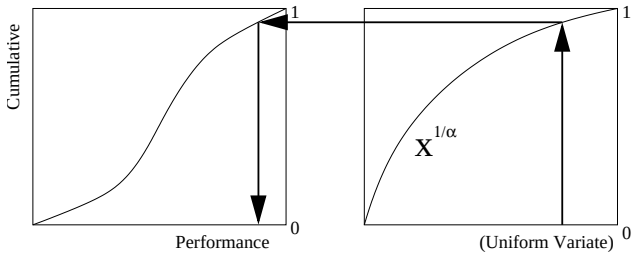


Figure E.3 Monte Carlo sampling of individuals based on their fitness. A uniform variate x is transformed by a simple power to produce selection pressure: a bias toward individuals with higher fitness.

E.2.4 Evolution & Selection Pressure

After a given generation of individuals has been constructed and the individuals' fitnesses evaluated, a new generation must be constructed. Some individuals survive into the new generation, and some new individuals are created by mutation or crossover. In both cases, the population must be sampled randomly. To mimic evolution, some *selection pressure* must be placed on the individuals for them to improve. This selection pressure is implemented with a simple Monte Carlo algorithm and controlled by a parameter $\alpha > 1$. The procedure is illustrated in Fig. E.3. In a standard Monte Carlo algorithm, a uniform variate $x \in [0, 1]$ is generated and transformed into the variable of interest by the inverse of its cumulative distribution. Using the cumulative distribution of the fitness will exactly reproduce the population without selection pressure; however, this sampling can be biased with a simple transformation. The right plot of Fig. E.3 shows a uniform variate x being transformed into $x^{1/\alpha}$, which is then inverted (left plot) to select an individual with a given fitness. As the parameter α grows, the individuals with high fitness are selected increasingly often.

While the selection pressure mechanism helps the system evolve, it comes at the expense of genetic diversity. If the selection pressure is too high, the population will quickly converge on the most fit individual. The lack of genetic diversity slows evolutionary progress. This behavior can be identified easily by looking at plots such as Fig. E.4. We have found that a moderate selection pressure $\alpha \in [1, 3]$ has the best results.

E.2.5 Parallelization and the Island Model

GP is highly concurrent, since different individuals' fitness evaluations are unrelated to each other, and dividing the total population into a number of sub-populations is a simple way to parallelize a GP problem. Even though this is a trivial modification to the program, it has been shown that such coarse grained parallelization can yield greater-than-linear speedup [163]. Our system uses a number of Islands connected to a Monitor in a star topology. CORBA is used to allow the Islands, which are distributed over multiple processors, to communicate with the Monitor.

Islands use the Monitor to exchange particularly fit individuals each generation. Since a separate monitor process exists, a synchronous exchange of individuals is not necessary. The islands are virtually connected to each other (via the Monitor) in a ring topology.

E.3 Performance Measures

The Genetic Programming approach outlined in the previous section is a very general algorithm for producing individuals with high fitness, and it allows one to factorize the definition of fitness from the algorithm. In this section we examine the function(al) which assigns each individual its fitness: the performance measure.

Before proceeding, it is worthwhile to compare GP to popular multivariate algorithms such as Support Vector Machines and Neural Networks. Support Vector Machines typically try to minimize the risk of misclassification $\sum_i |y_i - f(\vec{v}_i)|$, where y_i is the target output (usually 0 or -1 for background and 1 for signal) and $f(\vec{v}_i)$ is the classification of the i^{th} input. This is slightly different than the error function that most Neural Networks with backpropagation attempt to minimize: $\sum_i |y_i - f(\vec{v}_i)|^2$ [164, 165]. In both cases, this performance measure is usually hard-coded into a highly optimized algorithm and cannot be easily replaced. Furthermore, these two choices are not always the most appropriate for High Energy Physics, as discussed in Section D.4.

The most common performance measure for a particle search is the Gaussian significance, $s/\sqrt{b}$, which measures the statistical significance (in "sigmas") of the presence of a new signal. The performance measure $s/\sqrt{b}$ is calculated by determining how many signal events, s, and

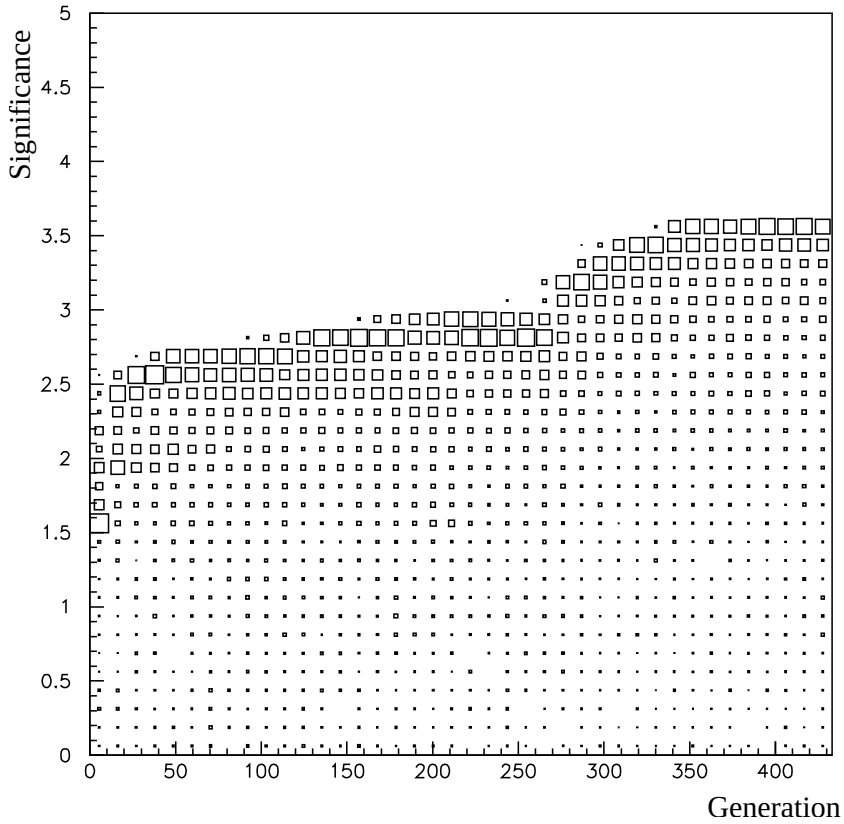


Figure E.4 The fitness of the population as a function of time. This plot is analogous to a neural network error vs. epoch plot, with the notable exception that it describes a population and not an individual. In particular, the neural network graph is a 1-dimensional curve, but this is a two dimensional distribution.

background events, b, a given individual will select in a given amount of data (usually measured in fb^{-1}).

The $s/\sqrt{b}$ is actually an approximation of the Poisson significance, σ_P, the probability that an expected background rate b will fluctuate to $s+b$. The key difference between the two is that as $s, b \to 0$, the Poisson significance will always approach 0, but the Gaussian significance may diverge. Hence, the Gaussian significance may lead to highly fit individuals that accept almost no signal or background events.

The next level of sophistication in significance calculation is to include systematic error in the background only prediction b. These calculations tend to be more difficult and the field has not adopted a standard (see Section C). It is also quite common to improve the statistical significance of an analysis by including a discriminating variable (see Section A.1.4).

In contrast, one may be more interested in excluding some proposed particle. In that case, one may wish to optimize the exclusion potential. The exclusion potential and discovery potential of a search are related, and G. Punzi has suggested a performance measure which takes this into account quite naturally [166].

Ideally, one would use as a performance measure the same procedure that will be used to quote the results of the experiment. For instance, there is no reason (other than speed) that one could not include discriminating variables and systematic error in the optimization procedure (in fact, the author has done both).

E.3.1 VCD for Genetic Programming

The VC dimension, h, is a property of a fully specified learning machine. It is meaningless to calculate the VCD for GP in general; however, it is sensible if we pick a particular genotype. For the slightly simplified genotype which only uses the binary operations of addition, subtraction, and multiplication, all expressions are polynomials on the input variables. It has been shown that for learning machines which form a vector space over their parameters,[4] the VCD is given by the

[4] A learning machine, $\mathcal{F}$, is a vector space if for any two functions $f, g \in \mathcal{F}$ and real numbers a, b the function $af + bg \in \mathcal{F}$. Polynomials satisfy these conditions.

$$
\begin{aligned}
f(x,y;\alpha) = a_1 \quad & +a_2 \cdot x \quad & + a_3 \cdot y \\
+ a_4 \cdot x \cdot x \quad & +a_5 \cdot x \cdot y \quad & + a_6 \cdot y \cdot y \\
+ a_7 \cdot x \cdot x \cdot y \quad & +a_8 \cdot x \cdot y \cdot y \quad & + a_9 \cdot x \cdot x \cdot y \cdot y
\end{aligned}
$$

Figure E.5 An explicit example of the largest polynomial on two variables with degree two. In total, 53 nodes are necessary for this expression which has only 9 independent parameters.

dimensionality of the span of their parameters [156]. Because the Genetic Programming approach mentioned is actually a conjunction of many such cuts, one also must use the theorem that the VCD for Boolean conjunctions, b, among learning machines is given by $\mathrm{VCD}(b(f_1,\ldots,f_k)) \leq c_k \max_i \mathrm{VCD}(f_i)$, where c_k is a constant [156].

If we placed no bound on the size of the program, arbitrarily large polynomials could be formed and the VCD would be infinite. However, by placing a bound on either the size of the program or the degree of the polynomial, we can calculate a sensible VCD. The remaining step necessary to calculate the VCD of the polynomial Genetic Programming approach is a combinatorial problem: for programs of length L, what is the maximum number of linearly independent polynomial coefficients? Fig. E.3.1 illustrates that the smallest program with nine linearly independent coefficients requires eight additions, eighteen multiplications, eighteen variable leafs, and nine constant leafs for a total of 53 nodes. A small Python script was written to generalize this calculation.

The Genetic Programming approach with polynomial expressions has a relatively small VCDs (in our tests with seven variables nothing more than $h = 100$ was found) which affords the relevance of the upper-bound proposed by Vapnik.

E.4 Summary

We have presented an implementation of a Genetic Programming system specifically applied to the search for new particles. In our approach a group of individuals competes with respect to

a user-defined performance measure. The genotype we have chosen consists of Boolean conjunctions of simple arithmetic expressions of the input variables required to lie in the interval $(-1, 1)$. Our implementation includes an island model of parallelization and a recentering algorithm to dramatically improve performance. We have emphasized the importance of the performance measure and decoupled fitness evaluation from the optimization component of the algorithm. In Chapter 11 we demonstrated that this method has similar performance to Neural Networks (the de facto multivariate analysis of High Energy Physics) and Support Vector Regression. We believe that this technique's most relevant advantages are

- the ability to provide a user-defined performance measure specifically suited to the problem at hand,
- the speed with which the resulting individual / cut can be evaluated,
- the fundamentally important ability to inspect the resulting cut, and
- the relatively low VC dimension which implies the method needs only a relatively small training sample.

Appendix F: The ATLAS Analysis Model

During the spring and summer of 2004, the author was actively involved in the development of the ATLAS analysis model. The analysis model is still under development, but we shall briefly describe the initial implementation provided for offline release 9.0.0 of the ATLAS software. The data model involves a hierarchy of detail starting with the Raw data, moving to the Event Summary Data (ESD), the Analysis Object Data (AOD), and finally Tags. The ESD is essentially the output of reconstruction and is expected to be available at the Tier1 computing centers. The AOD is expected to be the primary data format for analysis. The Tags will provide only minimal data from which one can find interesting events in the AOD.

The author was primarily involved in the development of the AOD particle classes. Those classes are shown in Figure F.1. One feature of the design is that the AOD is able to navigate back to the original ESD file from which it was created. For instance, this would allow someone to query an Electron object for the calorimeter cluster from which it was created.

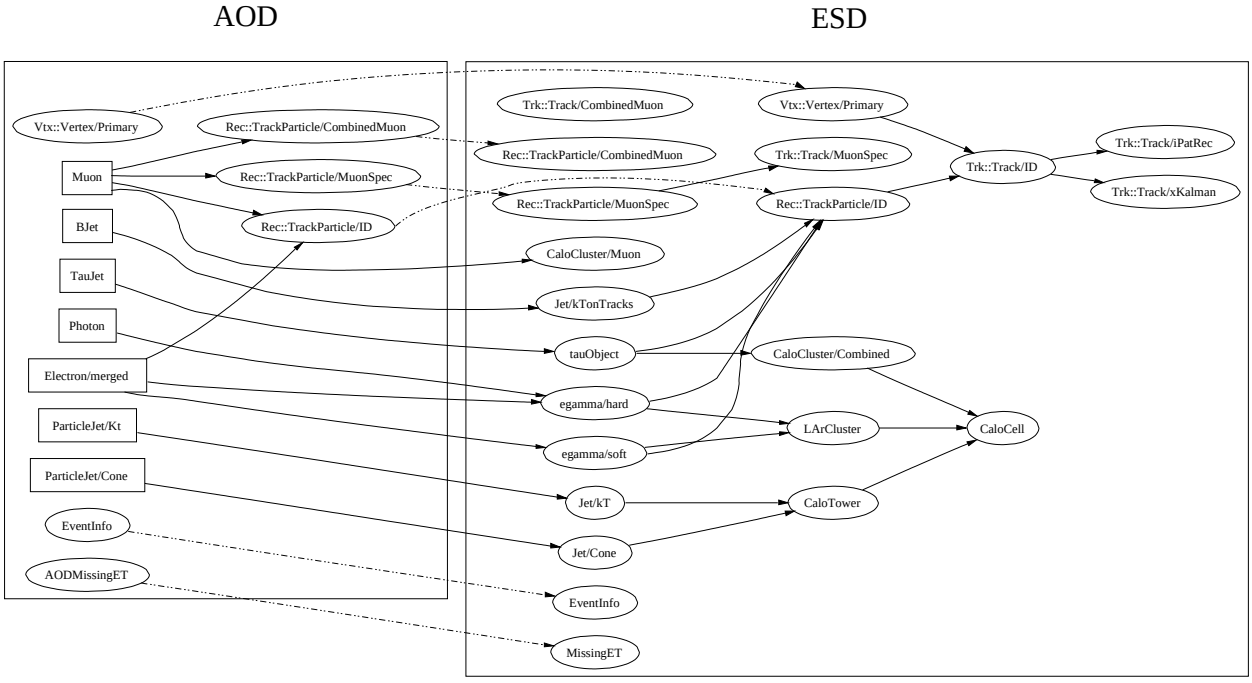


Figure F.1 The ATLAS Analysis Event Data Model.