Polynomial Discrimination and Weighted Counting in a Standard Model Higgs Boson Search

Sigve Haug
Department of Physics
University of Oslo
Norway

Thesis submitted in partial fulfillment
of the requirements for the degree Doctor Scientarium.

October 2004

Abstract

A particle search in terms of discrimination and hypothesis testing is described. Then polynomial discriminators and weighted counting are used to search for the Standard Model process $e^+e^- \rightarrow H\nu\bar{\nu}$ in the year 2000 data taken by the DELPHI detector at the Large Electron Positron Collider at CERN (LEP). No signal is found. However, using the CL_s convention for hypothesis testing a 95% CL lower limit on the Higgs boson mass was deduced to be 98.4 GeV/c^2.

In opposition to independent discriminations on the same data, minimal and highly hypothesis independent preselections are used. This enables a more standardized statistical treatment. For the chosen observables applied in the discriminations, the sample sizes allow for polynomial discriminators of third order. The discriminating power is comparable or slightly higher than those achieved in the independent searches.

Acknowledgements

Four years of teaching, education and Higgs boson search at the very end of the LEP era were given to me via an assistantship at the Department of Physics at the University of Oslo. The group of experimental particle physics accepted me as a novice on the field who came from the KARMEN experiment and had a unfamiliar background from the German university system. Since then I have crossed many peoples trajectories at the institute, on conferences, at CERN and at a lot of other points. I owe them all something.

In particular I must mention my supervisor Prof. Alex L. Read. Without his position in the field and his expert knowledge this work would evidently not exist. Further Cand. Scient. Børge K. Gjelsten appeared to become a colleague and friend suitable for physics discussions, literate smalltalk and exchange of frustration. Cand. Scient. Yuriy Volodumyrovych Pylypchenko deserves the same characteristic. Mustafa Hussain and all the others in our research group are remembered because of their crazy ideas, generous wallets and outstanding wit. Without my students at the Oslo University College life would have been just that 50%. My friends in Oslo Byforum have certainly contributed to the underestimated non-scientific part of being with their struggle for a better city. Financially my parents have continued where the state gave in. The list continues and the not mentioned are remembered. Of course great parts of this work are based on excellent scientific achievements of many in the DELPHI collaboration.

Finally I express my thanks and love to Magister Christine Brandtner who gave birth to our son Lyder Leander during my work on this thesis. Without her I would still have been a lost ship on a paved ocean.

Contents

List of Figures

List of Tables

Chapter 1

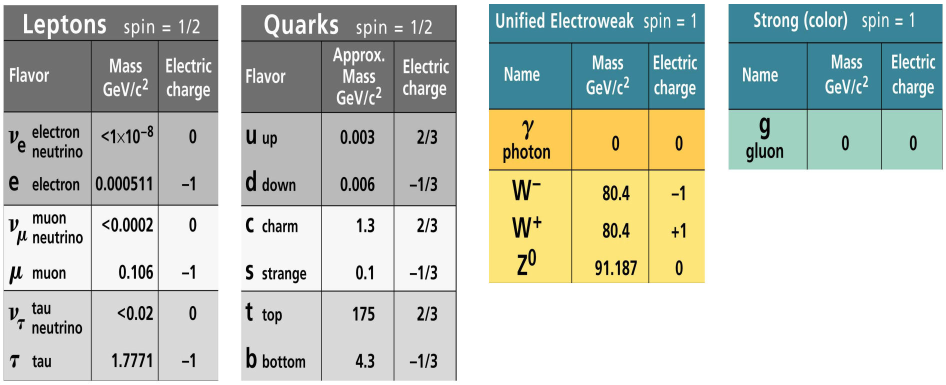
Introduction

The rapid progress and the exploding expansion of information technology have enabled an extensive use of advanced statistical methods in most disciplines. With respect to this trend, the field of elementary particle physics is rather a rule than an exception. In the experimental exploration of the elementary particles and their interactions and in the search for new particles, old established as well as new invented statistical methods have been exploited to the practical limit given by computing power and available data. This is a development whose significance will only increase in the years to come. Hence this work explores the use of advanced statistical methods in the search for the Standard Model Higgs boson.

The human search for the elementary matter that constitutes the universe is a fact as old as the so-called Western civilisation. Since the era of the Greek presocratics in the sixth and the fifth century BC, the representations of the basic pieces of nature and their interactions have evolved, at least in the established history of ideas, from the atomistic concept of Leukippos' and Demokritos' via Aristoteles' fundamental elements, the periodic elements, the discovery of baryons and leptons to the present landscape of elementary particles. Today this evolution of human imagination is crowned by the Standard Model, the relativistic quantum field theory containing all observed elementary matter particles and interactions but gravitation. No observation has yet reached an established acceptance as falsification of this theory. Even the recent discovery of neutrino oscillations whose consequence introduces massive neutrinos, may fit into the scheme of the Standard Model.

But even if the Standard Model stands in almost perfect agreement with present observations, it is not complete by experimental means. The observed particles of the Standard Model are listed in Table 1.1 with some of their properties. As indicated in the table, the Standard Model contains the unified electroweak force which is a symmetry observed to be broken at the GeV scale. The complex Higgs doublet accommodates this symmetry breaking and thus is an essential part of the theory. Its physical particle, the Higgs boson, is in spite of the immense investments still the only unobserved particle in the Standard Model. The reason is likely to be the large mass of this particle, maybe several GeV/c^2 beyond the experimental discovery potential of LEP around 109.5 GeV/c^2 which has been reported by the LEP Higgs Working Group [1].[1] Hence, this working group has combined the search results from the four LEP experiments to deduce a lower exclusion limit for the Standard Model Higgs boson at a mass of 114.4 GeV/c^2 (see Appendix B). This limit from LEP is for the time being the world record. The proton anti-proton collision data to come from the Tevatron during the next years may not enable a discovery beyond the LEP limit, but some improvement of the exclusion limit is expected. [2] In a situation where no evidence of Higgs boson production is found, the data delivered by the Tevatron may be used to exclude the Standard Model Higgs boson up to a mass around 180 GeV/c^2 at a 95% confidence level within 2008 [2].

Table 1.1: Observed fields in the Standard Model and some of their properties. To the left the fermions are grouped into quarks and leptons. To the right the gauge bosons are arranged by the forces they represent. The Higgs boson is still not observed and thus not listed. Table taken from Reference [3].



Leptons spin = 1/2

Flavor	Mass GeV/c^2	Electric charge
ν_e electron neutrino	$<1\times10^{-8}$	0
e electron	0.000511	−1
ν_μ muon neutrino	<0.0002	0
μ muon	0.106	−1
ν_τ tau neutrino	<0.02	0
τ tau	1.7771	−1

Quarks spin = 1/2

Flavor	Approx. Mass GeV/c^2	Electric charge
u up	0.003	2/3
d down	0.006	−1/3
c charm	1.3	2/3
s strange	0.1	−1/3
t top	175	2/3
b bottom	4.3	−1/3

Unified Electroweak spin = 1

Name	Mass GeV/c^2	Electric charge
γ photon	0	0
W^-	80.4	−1
W^+	80.4	+1
Z^0	91.187	0

Strong (color) spin = 1

Name	Mass GeV/c^2	Electric charge
g gluon	0	0

This prospect is fairly interesting because, on the one side, the Tevatron will be able to confirm a nearly two standard deviations effect in the combined LEP results for a Higgs boson with a mass around 116 GeV/c^2, and on the other, because indirect constraints on the Higgs boson mass highly prefer masses below

---

[1] LEP is the abbreviation for the Large Electron Positron Collider which was operated at the European Center for Particle Physics (CERN) from the summer of 1989 till the autumn of 2000. The four experiments ALEPH, DELPHI, L3 and Opal collected e^+e^- collision data.

[2] The Tevatron is the proton anti-proton accelerator at Fermi National Accelerator Laboratory (Fermilab) whose operation started in 1983. It produces collisions for the two experiments CDF and D0. At the Tevatron the bottom quark and the top quark were discovered in 1977 and 1995 respectively. In July 2000 the Fermilab experimenters announced the first direct observation of the ν_τ. The goal is to obtain 15 pb^{-1} by the year 2008 [2].

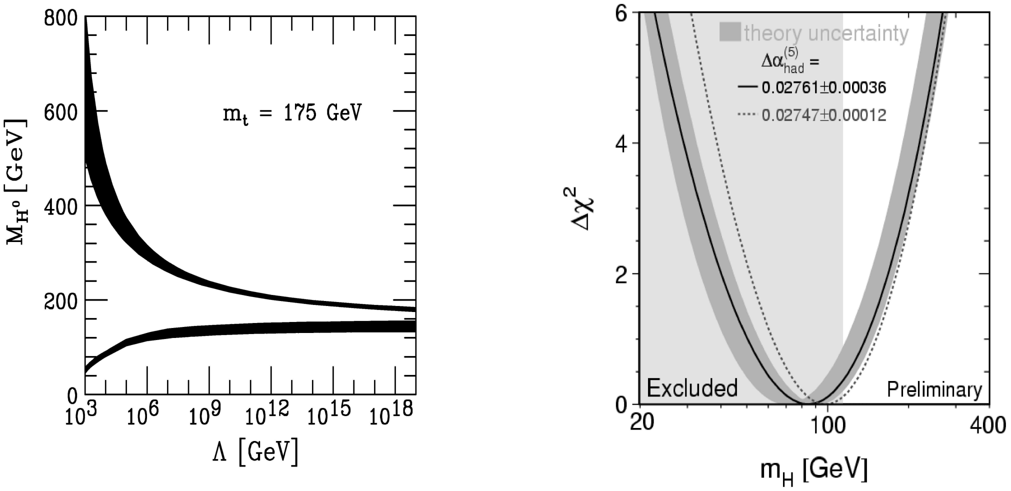


Figure 1.1: To the left bounds on the Higgs boson mass based on self-consistency arguments of the Standard Model are shown [55]. Λ denotes the high energy scale where the model would become non-perturbative or the electroweak potential unstable. To the right the χ^2 of the Higgs boson mass fit to the electroweak precision data is shown [28].

200 GeV/c^2 as shown in Figure 1.1 [6].[3] However, a final answer on the existence of the Higgs boson is first certain with the future data from the LHC at CERN whose collection is expected to start in 2007.[4] A further experimental fact for the incompleteness of the Standard Model is to an increasing degree established by the emerging field of astroparticle physics. According to present models of the universe and recent observations, the matter content of the Standard Model only accounts for roughly 5% of the energy in the universe. Calculations trying to reproduce the spin of observed galaxies, are forced to assume that approximately 25% of the energy is allocated in dark and unknown matter. If the remnant 70% is assigned to the vacuum energy, the matter responsible for the 25% also has to be counted for in a complete model of the fundamental matter constituents of the universe.

The Standard Model is not only incomplete by experimental means, it is also not considered as complete by theoretical means. At the reduced Planck scale $M_P = (8\pi G_{Newton})^{-1/2} = 2.4 \times 10^{18}$ GeV the strength of gravity becomes comparable to the strength of the Standard Model forces and a theory including quantum gravity must be introduced. Furthermore, the huge sensitivity of the Higgs potential and thus the whole mass spectrum of the Standard Model to the ultraviolet momentum cutoff requires a Λ_{UV} at the TeV scale to fit the

---

[3] The Higgs mass is a free parameter in the Standard Model, but affects other measured quantities via loop diagrams.

[4] LHC is the abbreviation for Large Hadron Collider which is replacing LEP at CERN. It will produce proton-proton collisions at a center of mass energy of 14 TeV.

preferred value of the Higgs boson mass around 100 GeV/c^2 (see Figure 1.1). Experimental deviations from the Standard Model predictions and therefore the need for theoretical extensions are expected already at energies accessible by the LHC. In highly speculative models with one or more extra dimensions it has even been shown that the fundamental Planck scale can have its four dimensional equivalent at the LHC scale [4],[7],[8]. The experimental study of supersymmetry, quantum gravity and then even string theory thus seems at least thinkable in the near future and heavily motivates the enormous activity within extensions of the Standard Model.

Both the experimental prospects and the theoretical research, driven by the fundamental questions whose answers the particle physics requires, have and will extend the use of advanced statistical methods within the field. On the one hand the multivariate discrimination methods used to reduce the enormous amount of background at LEP will be developed further. Maybe even the modified frequentist method, which established itself as the appropriate way of setting limits on the Higgs boson masses and testing hypotheses in the Higgs searches at LEP, will be extensively applied in the numerous searches at LHC. The obvious reason for such assumptions is that these methods exploit the data more efficiently. Multivariate discrimination takes into account correlations of the observables, and the modified frequentist method with weighted event counting includes measured information in a more complete manner. Thus the results become more stringent. On the other hand the amount of data to be acquired, being on a petabyte scale, and the enormous parameter spaces of the various hypotheses, i.e. models, to be tested on coming observations will prefer a high level of statistical treatment and automation.

Common to many advanced statistical methods is that they are computing intensive. When correlations are taken into account, larger samples are required. This demands huge simulations and extensive storage space. Then, when the samples are large, the calculation of discriminants becomes time consuming. Hypothesis testing involving more information also tends to increase computing times significantly. And last but not least, the scans of parameter spaces of models like the Minimal Supersymmetric Standard Model enlarge the need for computing capacity. This need is being met by new computer networks on a planetary scale.[5] Combined with higher level languages and intelligent network software, distributed computing will enable the extended use of advanced statistics.[6] It may even extend the way of representing fundamen-

---

[5] The usual supercomputer centers and dedicated UNIX clusters in connection with a high threshold of bureaucracy are being replaced by local LINUX networks networked into international computing networks. At the time of writing some tens of local LINUX workstations were available for this work, more than a hundred in the Condor system of the university and many hundreds via the Nordugrid GRID testbed. In near future the power of thousands and thousands of machines will be under the typing fingers of a scientist.

[6] Increasingly complex implementations demand higher abstraction. The old FORTRAN

tal aspects of nature by allowing us to perform tests of models not previously feasible.

This work explores multivariate discrimination combined with the modified frequentist way of hypothesis testing. Chapter 2 therefore tries to establish an overview of discrimination and hypothesis testing, emphasizing the benefits and the drawbacks of different methods. A special weight is given to the higher order discrimination especially studied in this work. The chapter is not as much original in its content, which relays on well known statistics, as it is in its form. Compressed reviews of the entire statistical procedure of a search are rare, and this second chapter aims to amend this. It serves as a methodological core for the following search. This search is done using the 224 pb^{-1} of data collected by the Detector with Lepton Photon and Hadron Identification (DELPHI) in year 2000 at center of mass energies from 203 GeV to 209 GeV. In the data it was searched for Standard Model Higgs boson events in the $e^+e^- \to H\nu\bar{\nu}$ channel. At high Higgs masses the cross section for this process is challengingly low and thus makes the channel interesting with respect to discrimination as well as weighted event counting. The theoretical and instrumentational aspects of the search are treated in Chapter 3. Chapter 4 then deals with the discrimination of the background and the testing of various hypotheses. Here an attempt was made to go further in the construction of the discriminators than other analyses have done. The existence of independent studies on the same data and the same channel has enabled a performance evaluation, to a certain degree reported in this chapter. The chapter ends with the exclusion limit on the Standard Model Higgs boson mass derived from the hypothesis test in the $e^+e^- \to H\nu\bar{\nu}$ channel. In the concluding Chapter 5 the most important lessons are summarized.

---

libraries in particle physics are now being or have already been replaced by object oriented C++ class libraries. New analysis tools like lhc++, ROOT and JAS written in object oriented languages speed up the feedback by interactivity and transparency. As the samples of data increase in amount and size administrative tasks become more demanding. Scripting languages like Python and Perl thus attain new positions in the field.

Chapter 2

Discrimination and Hypothesis Tests in Searches

"The analysis of search results can be formulated in terms of a hypothesis test. The null hypothesis is that the signal is absent and the alternate hypothesis is that it exists. An analysis of search results is simply a formal definition of the procedure which quantifies the degree to which the hypotheses are favored or excluded by an experimental observation [9]." The interpretation of search results are thus dependent on the formal definition of the search procedure and to a certain extent on the practical implementation of the same. Therefore this chapter attempts to describe the search procedure in general, point out practical challenges, sketch various approaches whose goals are to meet these challenges and to examine the methods used in the Standard Model Higgs boson search presented subsequently.

In this work a search is divided into two steps. The first consists of discriminating the background. This is mostly an inevitable task in High Energy Physics, as well in searches as in measurements. At LEP, being the accelerator whose data will be subjected to this work's particular search, the Higgs boson production cross section is at the order of picobarns, a size up to the third and fourth order of magnitude smaller than other Standard Model processes. Obviously the significance, most often defined as the number of signal events divided by the square root of the number of background events, must be increased significantly before a hypothesis test will make any sense. Hence, discrimination is the essential start whose performance infers the search result. The second step, the hypothesis test, then uses the discriminated samples to state a discovery or an exclusion as stringent as possible. Intuitively, methods that utilize the available information most efficiently are expected to yield the best results in both steps. The following sections are organized according to this division and focus on multivariate techniques.

The first section concerns the discrimination. It starts with a presentation of the mathematical optimal discriminator, a ratio of probability density functions and its practical challenges. These challenges lead to three categories of discrimination methods. Some methods in each category are touched, and special emphasis is given to the discrimination method applied in this work. Then there is given place to a few general remarks on iterative discrimination, selection of observables, over-estimation and cross-validation. The second section presents different approaches to hypothesis testing in order to enable the classification and interpretation of the approach chosen for this work. That approach, the modified frequentist approach, is then explained and commented. In the summary the chapter's most important conclusions are redrawn.

2.1 Discrimination

If two classes 0 and 1 of events are given, the goal of discrimination is to construct a rule that keeps a certain fraction of one class while rejecting as much as possible from the other class. The path to this goal may be deduced from classification theory. In classification theory it is proved that an optimal rule $C(\vec{x})$ for assigning an event $\vec{x}$ to one of two classes, without doubt, is given by

$$C(\vec{x}) = 0 \quad if \quad f(0|\vec{x}) = max(f(0|\vec{x}), f(1|\vec{x})) \tag{2.1}$$

where $\vec{x}$ is the vector of observables spanning the feature space, $f(0|\vec{x})$ and $f(1|\vec{x})$ are the conditional probability density functions, p.d.f.s, for a given $\vec{x}$ to belong to the class 0 or 1.[1] A proof of the optimality of this rule, often called Bayes rule, can be found in e.g. Reference [10]. $C(\vec{x})$ simply tells to assign the event $\vec{x}$ to class 0 if the conditional p.d.f. for this class is greater than the conditional p.d.f. for class 1. It minimizes the error rate, i.e. the frequency of misallocating an event. The rule is in this sense optimal.

In discrimination it is desired to assign events to a class 0 even if Equation 2.1 is not fulfilled, but in a manner which keeps a chosen fraction of signal events in the mentioned optimal sense. This is achieved via a reformulation of Equation 2.1. By introducing the parameter ϵ which represents the desired fraction,

$$D(\vec{x}; \epsilon) = 0 \quad if \quad f(0|\vec{x}) > \epsilon\, f(1|\vec{x}) \tag{2.2}$$

is obtained. This rule allows to keep an arbitrary fraction of class 0 while the rejection of class 1 is optimal. Equation 2.2 is equivalent to $\epsilon f(1|\vec{x}) - f(0|\vec{x}) < 0$ and corresponds to the requirement following from a mandatory assignment, i.e. there is no doubt class, equal errors, i.e. misclassification is equally bad for both classes and the desired goal of constructing a boundary in the feature

---

[1] Definitions of joint, marginal and conditional probability density functions may be found in Reference [4].

space with the biggest possible fraction of class 0 events on one side and the biggest possible fraction of class 1 events on the other side. That this holds is proved e.g. in Reference [13].

The remaining task is then to identify the conditional p.d.f.s. These are known via Bayes' theorem

$$f(k|\vec{x}) = \frac{\sigma_k \; f_k(\vec{x})}{\sum_{l=0}^{1} \sigma_l \; f_l(\vec{x})} \tag{2.3}$$

where k is either 0 or 1 [11].[2] σ_k are the prior probabilities of the classes and the $f_k(\vec{x})$ the joint probability density functions, the likelihoods. Using this, it is clear that Equation 2.2 leads to a ratio of likelihoods.

$$\frac{\sigma_0 \; f_0(\vec{x})}{\sigma_1 \; f_1(\vec{x})} < \beta_\epsilon \tag{2.4}$$

It is then shown that the likelihood ratio is an optimal discriminator if the joint p.d.f.s and the prior probabilities are known. This is not surprising since Equation 2.4 is just the Neyman-Pearson theorem. However, by establishing the task of discrimination in Equation 2.2 the theory of classification may be applied to discrimination.

The general complication in classification, and from now on, in discrimination, is the in principle unknown character of the joint p.d.f. $f_k(\vec{x})$ and the frequency σ_k of class k. However, in High Energy Physics the σ_k is normally known as measured or calculated cross sections for the processes in question. $f_k(\vec{x})$ is in general not known. This appeals to different approaches which may be segmented into three categories. Approaches in the first category tries to estimate $f_k(\vec{x})$ with various techniques. The second category encloses approaches which aim to model $f(k|\vec{x})$ directly. In the third category attempts that takes the consequence of the approximate character of the former two, are grouped. These methods focus straight on the rule itself, ignoring the mathematical foundation of Equation 2.2. Some approaches in each category are reviewed in the following subsections.

2.1.1 Estimation of Joint p.d.f.s

Since the joint p.d.f.s $f_k(\vec{x})$ are in general not known, there exist different ways to estimate them and so obtain a discrimination rule $D(\vec{x};\epsilon)$. It is usual to speak of parametric and non-parametric estimations.

A **parametric estimation** of $f_k(\vec{x})$ is based on distributions described by a set of parameters. For example, an assumption based on normal distributions

---

[2] Bayes' theorem is as usual $P(A|B) = P(A)P(B|A)/P(B)$ where $P(B) = \sum_A P(A)P(B|A)$. Identifying $P(B|A)$ with the joint probability function for B at given A and $P(A)$ with the priori probability for A, i.e. with σ_k, Equation 2.3 is achieved.

with different means $\vec{\mu}_k$, but common covariance matrices, $f_k(\vec{x}) = N(\mu_k, \Sigma)$, leads to a linear rule $f_0(\vec{x})/f_1(\vec{x}) = (\vec{\mu_0}^T - \vec{\mu_1}^T)\Sigma^{-1}(\vec{\mu} - \vec{x})$. $\vec{\mu_0}$, $\vec{\mu_1}$, their mean $\vec{\mu}$ and the covariance matrix Σ are estimated with training samples which may be simulated or measured. For each event $\vec{x}$ a discriminator value can then be calculated after the training. By taking the discriminator values of the signal events, an ϵ or a β_ϵ can be chosen to keep the desired fraction of the signal. All events with discriminator values lower than ϵ or β_ϵ are considered as rejected. An analogous calculation with normal distributions with different means and different covariance matrices $f_k(\vec{x}) = N(\vec{\mu}_k, \Sigma_k)$ leads to a rule quadratic in $\vec{x}$. It is evident that these results are likelihood ratio discriminators which are, as a consequence of Equation 2.2, optimal. Normal distributions are common assumptions in parametric estimations, but other distributions may of course come closer to the reality, depending on the particular situation.

Since parametric estimation does make hard assumptions on the joint p.d.f.s, **non-parametric estimation** may come closer to the real functions. This approach takes direct advantage of the shape of the measured or simulated training samples. However, the construction of p.d.f.s often involves the whole complex of density estimation. The straightforward estimation with histograms suffers from various statistical problems which leads to several improved methods. Kernel estimation and spline interpolation are popular ways of smoothing histograms. A specific case is examined in Appendix A where two-dimensional histograms are smoothed. Compared to the parametric approach, non-parametric estimation requires larger training samples, an unavoidable fact that may be accounted for by reducing the dimension of the feature space.

2.1.2 Estimation of Conditional p.d.f.s

Discrimination aims to reject a biggest possible part of one class while keeping as much as possible of the other class. Thus, one may argue that the joint probability density functions are of subordinate interest. A direct determination of the conditional p.d.f.s does still provide the information whose presence the rule $D(\vec{x};\epsilon)$ is based on. Various approaches to this direct determination exist.

A straightforward and much used discrimination technique consists of sequential cuts. It creates a hypercube in feature space and is normally optimized by intuition and brute force computing. However, this technique may be embedded into the concept of **trees**. Each cut is seen as a node whose conditional p.d.f. becomes

$$f_i(k|\vec{x}) = \frac{Number\ of\ events\ from\ class\ k\ in\ node\ i}{Total\ number\ of\ events\ in\ node\ i} \tag{2.5}$$

A classification tree becomes, in the case of discrimination where there are only two classes, a sequence of nodes whose decisions are to either reject the event or pass it further. The decision $D(\vec{x};\epsilon)$ is based on $f_i(k|\vec{x})$. The obvious question is how to optimize such a tree with respect to the observables used and their cuts. A possible method is to construct a likelihood function for the tree. This function is intuitively binomial

$$L = \prod_{i=1}^{N_{nodes}} f(n_{i_0}|n_i) \propto \prod_{i=1}^{N_{nodes}} (\frac{n_{i_0}}{n_i})^{n_{i_0}} (\frac{n_i - n_{i_0}}{n_i})^{n_i - n_{i_0}} \tag{2.6}$$

where n_i is total number of events in the node, namely the sum of the number of events n_{i_0} and n_{i_1} in each class. If the entropy in a node $S_i = -p_i \log p_i$ with the probabilities p_i is introduced, this relation may be written as

$$\log L = \sum_{i=1}^{N_{nodes}} [n_{i_0} \log(\frac{n_{i_0}}{n_i}) + n_{i_1} \log(\frac{n_{i_1}}{n_i})] = - \sum_{i=1}^{N_{nodes}} n_i [S_{i1} + S_{i2}] \tag{2.7}$$

It is then possible to optimize the tree, or the sequential cuts, by maximizing the log likelihood. It is equivalent to minimizing the impurity, i.e. the entropy in the child nodes. This probabilistic description of sequential cuts provides a way to optimize the cuts, but does not overcome the essential drawback inherent to this technique. Correlations between variables are not taken into account, i.e. the selected region in the feature space is a hypercube.

Another approach to estimation of $f(k|\vec{x})$ directly is based on **nearest neighbour** methods. These methods use local regions in feature space around $\vec{x}$. The conditional p.d.f. may be defined as

$$f_i(k|\vec{x}) = \frac{N^k_{nearest}}{N_{nearest}} \tag{2.8}$$

where $N_{nearest}$ is the number of events nearest to $\vec{x}$ and $N^k_{nearest}$ the number of events belonging to class k nearest to $\vec{x}$. Either the number of nearest events or the local region may be fixed. Further it is necessary to define nearness, i.e. a distance.

A huge set of $f(k|\vec{x})$ estimations is based on **flexible discriminant** analysis. In flexible discriminant analysis a smooth function $g_k(\vec{x})$ may be estimated from the training sample for class k in order to map $\vec{x}$ onto the real axis. In principle there are no restrictions on $g_k(\vec{x})$ and linear rules to arbitrary flexible neural networks can be used.[3] The $f(k|\vec{x})$ may be defined as

---

[3] For a neural network with one hidden layer $g_k(\vec{x}) = g_0(\sum_{j\to k} \bar{\omega}_{jk} g_h(\sum_{i\to j} \omega_{ij} \vec{x}_i))$ is obtained. The functions g_0 and g_h, the number of nodes, the interaction lines and the number of hidden layers must be chosen, the weights $\bar{\omega}$ and ω are trained, i.e. estimated. Because of the enormous freedom neural networks are extremely flexible, but also hard to evaluate.

$$f_i(k|x) = \frac{e^{g_k(\vec{x})}}{\sum_l e^{g_l \vec{x}}} \tag{2.9}$$

where l runs over both classes. If the $g_l(\vec{x})$ are chosen to be linear, the discriminator achieved is also linear and yields the same result as the likelihood ratio of normal distributions with common covariance matrices. However, the approach of flexible discriminant analysis is more general since there may be other joint p.d.f.s that lead to the conditional p.d.f. in Equation 2.9. The parameters of the $g_l(\vec{x})$ can be estimated for example with the Maximum Likelihood method.

2.1.3 Direct Estimation of the Rule $D(\vec{x};\epsilon)$

Equation 2.2 shows how to construct an optimal discriminator. The discriminator is optimal if the conditional probability density functions are exactly known. This is in principle possible via Bayes' theorem and the joint p.d.f.s. Hence, one approach is to estimate these joint p.d.f.s as precisely as possible. A second approach is to estimate the conditional p.d.f.s directly. The third possibility is to estimate a discriminator without the conditional p.d.f.s. This may also lead to an optimal result, or to an estimation as optimal as others. Again there exists a manifold of attempts on this path. Especially flexible discrimant analysis can be used to estimate a function $g_k(\vec{x})$ which maps the event vectors $\vec{x}$ onto the real axis. Then a distance definition may be applied to reject events. In the following an approach used for the Higgs boson search presented in Chapter 4 and Chapter 5 is outlined. This approach might be interpreted as a polynomial generalization of linear and quadratic discriminators.

Now the optimal separation is formulated as a minimal within-variance and simultaneously a maximal-between variance of the class samples. In other words, it is desired to cluster the events from each class and separate the means of the clusters in a most efficient manner in the feature space. These requirements then lead to a surface which separates the events optimally. The surface is approximated arbitrarily closely by polynomials of the event vectors $\vec{x}$.

By following the proof in Reference [13], a linear function can be written as a scalar product of the coefficients $\vec{a}$ and the event vector $\vec{x}$.

$$\epsilon = \vec{a}\vec{x} \tag{2.10}$$

The coefficients may be determined so as to maximize the ratio of between-class variances $\Delta\vec{\mu}$ to within-class variances V , i.e. by differentiation of

$$\frac{(\vec{a}\Delta\vec{\mu})^2}{\vec{a}V\vec{a}} \tag{2.11}$$

with respect to $\vec{a}$. The parameterization of the surface is then obtained and yields

$$\vec{a} \propto V^{-1}\Delta\vec{\mu} \tag{2.12}$$

where V is the covariance matrix and $\Delta\vec{\mu}$ the difference between the sample means of the respective classes. Since only the relative distance is of interest for the discrimination, ϵ may be multiplied by any convenient constant. Without assuming anything about the p.d.f.s, Equation 2.12 is equivalent to a discriminator achieved by the likelihood ratio method for normal joint p.d.f.s with common covariance matrices. It is famously known as the Fischer's discriminant, and analysis based on it, is normally called linear discrimant analysis.

Linear discriminant analysis yields a hyperplane in the feature space and is optimal for normal distributed samples with common covariance matrices. But in general the optimal surface is obviously not a plane. Generalizing the linear expression leads to the following.

$$\epsilon = \vec{x}(\vec{a} + B\vec{x}) \tag{2.13}$$

Expanding $\vec{a}$ to contain the coefficients of the matrix B and $\vec{x}$ the combinations of $x_i x_j$, the equation $\vec{a} = V^{-1}\Delta\vec{\mu}$ can still be used. This generalization gives a quadratic rule which is known to be optimal for normal distributed classes with different covariance matrices (see e.g. Reference [14]).[4] The quadratic rule yields a parabola in the feature space. Its fundamental form is shown in the upper right plot of Figure 2.1, i.e. no interactions, $x_i x_j$ with $i \neq j$, are included and all coefficients are set to one. The linear case which yields a plane is shown in the upper left plot.

Generalizing further, the expansion of $\vec{x}$ now contains the combination with repetitions of the expansion from the second order with the observables x_i. The basic form is shown to the left in the mid row of Figure 2.1. This third order polynomial approximates the surface with a cubic parabola and may be called cubic discriminant analysis. A general discriminator based on Equation 2.12 can be written as

$$\epsilon = \sum_{i=1}^{N_x} x_i(a_{ijk} + \sum_{j=i}^{N_x} x_j(a_{ijk} + \sum_{k=j}^{N_x} x_k(a_{ijk} + \ldots.)))$$

where the number of indexes goes up to the order of the expression and N_x is the number of observables. The coefficients $a_{ijk\ldots}$ are to be estimated. The basic forms of fourth and fifth order polynomial approximations are shown in

---

[4] Due to this fact, a transformation of the samples into normal distributions is often expected to improve the discriminant power at the expense of complication.

the mid right plot and the lower left plot. The sixth plot, to the lower right, is drawn with some interaction between the three observables x_1, x_2 and x_3 to illustrate that more complicated surfaces than the basic forms are very well achievable.

Quadratic discriminant analysis is widely used while higher order polynomials are rare. An obvious reason is the rapidly increasing dimension of the expanded vector $\vec{x}$. Using the intuitive formula for the number of combinations with repetitions, it follows for the dimension of the expanded event vector

$$dim(\vec{x}) = \sum_{k=1}^{n} \binom{N_x+k-1}{k} \tag{2.14}$$

where n is the order of the polynomial and N_x is the number of observables, i.e. the dimension of the feature space. This dependence is illustrated in Figure 2.2. For four different numbers of variables, $N_x = 20, 15, 10$ and 5, the dimension of the expanded $\vec{x}$ is histogrammed versus the order of the polynomial. Noting the logaritmic scale on the vertical axis, it is clear that the dimension explodes with increasing order. A typical Higgs boson search may use twenty event variables which yields $\dim(\vec{x}) = 208$ for a quadratic discriminant. The corresponding covariance becomes a 208×208 matrix. A cubic discriminant includes a 1770×1770 matrix which has to be filled for every event in the analysis. This matrix then has to be inverted. Even with efficient methods like Cholesky factorization this may take significant computing time. Discriminators based on higher order polynomials are thus computational intensive and require large data sets.

As this work was started, a study of existing discrimination methods used in the search for the Standard Model process $e^+e^- \to H\nu\bar{\nu}$ with the DELPHI detector was undertaken. It was found that sequential cuts in combination with a likelihood ratio method and a quadratic discriminant analysis were practiced [15], [16]. Concerning the discrimination, the idea was then to use generalized polynomials, not only such of second order. It seemed natural to approximate the separating surface in the feature space as precisely as the size of the samples would allow. In other words, it seemed natural to increase the order of the polynomial until overestimation occurred. Thus, the evaluation of different discriminators and overtraining had to be given some attention.

2.1.4 Evaluation of Discriminators and Overtraining

The previous subsection intended to establish a scheme which embraces the manifold of approaches to discrimination. The approach which gives the best discrimination may depend on the particular data samples. The choice is thus often a practical choice in which the efficiency of the discriminator seems to be of superior relevance.[5] In the end the discrimination is just a cut on this

---

[5] In general, however, the need for identifying outliers or using unclassified training samples requires knowledge about the joint probability density functions.

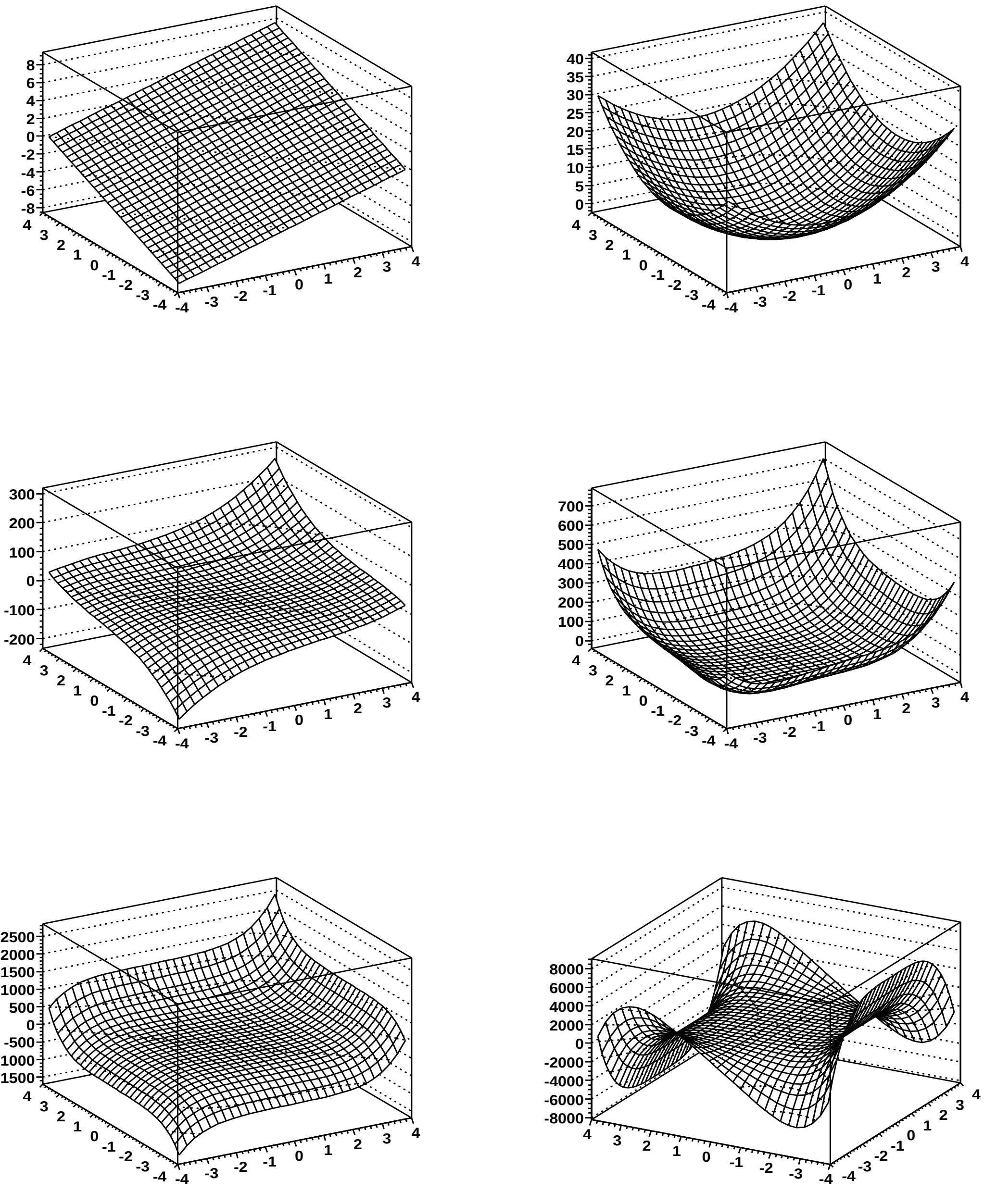


Figure 2.1: Basic surfaces in a 3-dimensional feature space. The surfaces are polynomials of different orders. In the upper left a first, in the upper right a second, in the mid left a third, in the mid right a fourth, in the lower left the fifth order and in the lower right a fifth order with some interaction between the observables are shown.

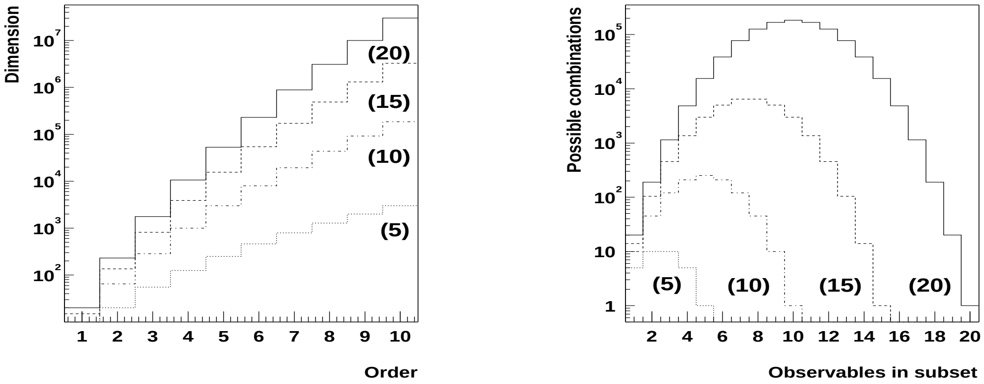


Figure 2.2: Dimension of the expanded event vector $\vec{x}$ versus the order of the polynomial is shown to the left. The solid histogram is for twenty observables, the dashed for fifteen, the dashed dotted for ten and the dotted for five observables. To the right the number of possible combinations of a subset of observables is shown. The solid histogram is for twenty observables in the full set and yields an integral of 1048575 combinations, the dashed for fifteen with an integral of 32753, the dashed dotted for ten with an integral of 1023 and the dotted for five with an integral of 31.

discriminator, resulting in remaining samples which are used to construct p.d.f.s as input for the hypothesis testing. Thus one may judge the discriminator alone on its rejection efficiency, the ratio of the number of background events after the cut to the original number of background events, at a chosen signal efficiency. This is how the **evaluation** of different discriminators were performed in this work (see Chapter 4).

Of course, the rejection efficiency is likely to decrease with increasing order of the polynomial. But at some point **overtraining** enters due to the finite nature of the training samples. This point has to be identified. A much practiced rule of thumb is that the training sample should be five to ten times bigger than the number of parameters to be estimated. For a polynomial discriminator this number is given by

$$N_{par} = k\ dim(\vec{x}) + k\frac{dim(\vec{x})(1 + dim(\vec{x}))}{2} \tag{2.15}$$

where k is the number of classes, here two. The former term represents the number of parameters to be estimated for $\vec{\mu}$ and the latter the number of parameters in the symmetric covariance matrix V. A search with ten variables and two classes, signal and background, then yields 130 parameters for a first order discriminant, 4420 for a quadratic and 82080 for a cubic. However, a discriminator should always be applied to a test sample to ensure there is no overtraining. The signal and rejection efficiencies from the test sample can then be compared to those achieved from the training sample.

The obligatory use of test samples introduces a conflict between the size of the training sample and the size of the test sample. A large test sample reduces the variability of the test but also decreases the accuracy of the estimation. An optimal solution to this conflict is achieved through **cross-validation**. The procedure consists of dividing the complete samples in N subsets. By using the i-th subset as test set for the discriminator estimated via the remaining subsets, a set of efficiencies eff_i is obtained. Repeating this for the remaining N-1 subsets yields then the averaged efficiency

$$eff = \sum_{i=1}^{N} eff_i \tag{2.16}$$

When N approaches the size of the complete sample, the size of the averaged test sample and of the training sample does also. The discriminator then becomes an approximation $\hat{D}(\vec{x}; \epsilon)$ of the discriminator estimated on the complete samples. In this sense cross-validation is an optimal exploitation of the available data. However, when complex discriminators are trained and tested on large samples, the procedure of cross-validation becomes a formidable computing task. The task is however of a parallel nature and therefore the emergence of networked computer networks enables an extended use of this technique.

2.1.5 Iterative Discrimination

When a discriminator $D(\vec{x}, \epsilon)$ is chosen and trained, a background rejection is provided as a function of the parameter ϵ. However, the approximate character of an estimated discriminator does not ensure an optimal separation surface in the feature space. Hence, an iterative discrimination may result in a better performance. For some samples a “cheese plane” technique, where only a tiny fraction of feature space is rejected in each of a large set of iterations, will yield an optimal result. But in practice iteration is again a computing intensive and administrativly complex task. And further the serial nature of iteration reduces the potential gain from the increasing computer parallelization.

In a situation where the background is a composite of several background processes, it may be more efficient to train the discriminator on one process or on one group of processes, discriminate these processes and then repeat the procedure for the remaining processes. For example in a search for $e^+e^- \rightarrow H\nu\bar{\nu}$ it may be natural to divide the background into three groups : Two fermion, four fermion and two photon processes. In Chapter 4 three iterations of discrimination, each trained on one group of background processes, are applied and compared with a discrimination which treats the background processes as a whole. Such "intelligent" iteration, which is a compromise between a sequential cut approach and a pure multivariate discrimination, is expected to gain from both methods' advantages.

2.1.6 Variable Selection

The set of possible discriminators $D(\vec{x}, \epsilon)$ is infinite. And although some discriminators can be proved to be optimal, often they are in practice not. First, the p.d.f.s are not allways known. Second, the observables with the highest discriminating power may be hard to identify. Normally the latter, the selection of observables to be contained in the $\vec{x}$, has the most important influence on the discriminating performance.

In the case where the observables are uncorrelated, redundant information is not expected to decrease the performance of a discriminator. So the use of all possible observables in $\vec{x}$ should be the natural choice as long overtraining is avoided. In general, however, the observables are often correlated and the inclusion of additional ones may very well reduce the performance. Thus all combinations should in principle be tested to find the optimal set of observables. Having a set of typically twenty observables this may be impossible, again with respect to time.[6] As shown in Figure 2.2 all combinations of twenty variables already requires over a million discriminations to be tested. Even the parallel nature and the networked computer networks seem to come short at this point. Some careful selection of variables based on physical "intuition" is still necessary to reduce the set down to a computable size.

In the discrimination applied in this work a combination of "intuition", i.e. careful study of observables, and raw computing is sought (see Chapter 4).

2.2 Hypothesis Testing

The discrimination procedure yields at a chosen signal efficiency three sets of candidate events $\vec{x}_i^{sig}$, $\vec{x}_i^{bgd}$ and $\vec{x}_i^{obs}$ from the simulated and measured samples which are optimized in some way with respect to the signal significance. The probability density functions of $\vec{x}_i^{sig}$ plus $\vec{x}_i^{bgd}$ now represents the signal plus background hypothesis. The p.d.f. of the measured or simulated background samples $\vec{x}_i^{bgd}$ constitute the alternate hypothesis.[7] The search task is to find out to which degree the experimental outcome $\vec{x}_i^{obs}$ disfavours the one or the other hypothesis.

---

[6] The number of possible combinations out of a set of n observables is as always given by $\sum_{k=1}^{n} \binom{n}{k}$.

[7] One differentiates between simple and composite hypotheses. Here the p.d.f.s are determined completely, i.e. the hypotheses are simple.

This is traditionally answered by calculating the probability of the outcome $\vec{x}_i^{obs}$, or more often, of a test statistic with a one to one relation to $\vec{x}_i^{obs}$, in a certain hypothesis. By defining a confidence level, one may reject or accept hypotheses with respect to this level. If marginal p.d.f.s are used to calculate the probabilities, the confidence intervals which respect the confidence level, are called frequentist, or classical confidence intervals. Since the construction of these classical intervals may lead to non-physical results, the Bayesian approach which constructs intervals via conditional probabilities, is sometimes preferred. With this approach non-physical intervals can be avoided by defining the prior probability properly. The frequentist as well as the Bayesian approach are reviewed in References [4] and [17]. However, the arbitrary character of the prior, led to an attempt which tries to eliminate non-physical results within the classical approach. This "unified approach to the classical statistical analysis of small signals" introduces the ratio of the likelihood for the signal plus background hypothesis to the likelihood for the best hypothesis, i.e. the hypothesis which maximizes the likelihood [17]. In addition, the best hypothesis is required to be physical. Experimental outcomes are ordered by decreasing likelihood ratio values and are accepted until the desired confidence level is exceeded.

The Higgs boson searches at LEP have been published using an approach close to this unified method. In fact, the likelihood for the best hypothesis is replaced by the likelihood for the background-only hypothesis. Further the signal is considered as excluded at a confidence level $1 - \alpha$ if a test statistic called CL_s itself is less then α. The search performed in this work follows this LEP tradition by applying the CL_s method, often called the modified frequentist analysis, in combination with the mentioned likelihood ratio. The modified frequentist analysis and its features are thus described in the following subsection. Together with the event discrimination it completes the search.

2.2.1 Modified Frequentist Analysis

The modified frequentist convention introduces the statistic CL_s. A signal plus background hypothesis is regarded as excluded at a confidence level $1 - \alpha$ if

$$CL_s \leq \alpha \tag{2.17}$$

Opposed to the classical procedure it is now not the probability to obtain a lower CL_s, but the CL_s itself which is lower than or equal to α. As will be clear in a sentence, the definition of CL_s ensures conservative limits, i.e. it results in general overcoverage, a coverage probability that is greater than $1 - \alpha$.

The statistic CL_s is defined as the ratio of the p-value for the signal plus background test statistic to the the p-value for the background-only test statistic

$$CL_s \equiv \frac{\mathrm{p-value\ of\ signal\ plus\ background\ hypothesis}}{\mathrm{p-value\ of\ hypothesis\ of\ background\ only}} \tag{2.18}$$

where the words are equivalent to the ratio of CL_{s+b} to CL_b, CL being the probability to obtain a test statistic value X lower than or equal to the observed value X_{obs},

$$CL_{s+b,b} = P_{s+b,b}(X \leq X_{obs}) \tag{2.19}$$

where

$$P_{s+b,b}(X \leq X_{obs}) = \int_0^{X_{obs}} \frac{dP_{s+b,b}}{dX} dX \tag{2.20}$$

and the $dP_{s+b,b}/dX$ is the probability density function.[8] The CL_s is thus a normalization of the p-value for the signal plus background hypothesis to the p-value of the background-only hypothesis. In other words, the CL_s approximates the p-value obtained in a background free experiment. This approximation avoids eventual non-physical rates when background is corrected for in experiments with small or absent signals. This is the modified frequentist procedure which has established some convention in the Higgs boson searches at LEP and which is described in greater detail in Reference [18]. The remaining task is to choose an appropriate test statistic.

2.2.2 The Likelihood Ratio Test Statistic

The likelihood ratio Q is often used as test statistic in hypothesis testing. As a consequence of the Neyman-Pearson theorem, Q maximizes the probability of rejecting a false hypothesis at a specified signal confidence level and the probability to discover a true signal at a specified background confidence level. The proof is reiterated in terms appropriate for new particle searches in Reference [19]. For N_{chan} independent search channels a general Q is given by

$$Q = \frac{\prod_{i=1}^{N_{chan}} \frac{e^{-(s_i+b_i)}(s_i+b_i)^{n_i}}{n_i!}}{\prod_{i=1}^{N_{chan}} \frac{e^{-b_i}b_i^{n_i}}{n_i!}} \frac{\prod_{j=1}^{n_i} \frac{s_i S_i(\vec{x}_{ij})+b_i B_i(\vec{x}_{ij})}{s_i+b_i}}{\prod_{j=1}^{n_i} B_i(\vec{x}_{ij})} \tag{2.21}$$

which is equivalent to

---

[8] Albeit not so often used in particle physics, the Particle Data Group has introduced this notation. p-value can be interpreted as an abbreviation for probability value of a test statistic. For example, an experimental outcome which lies one standard deviation off the expected mean of a normal distribution, has a p-value equal to 31.7%.

$$Q = e^{-s_{tot}} \prod_{i=1}^{N_{chan}} \prod_{j=1}^{n_i} \left(1 + \frac{s_i S_i(\vec{x}_{ij})}{b_i B_i(\vec{x}_{ij})} \right) \tag{2.22}$$

where n_i is the number of observed candidates in the i-th channel, s_i and b_i are the integrated signal and background rates for the i-th channel, $s_{tot} = \sum_{i=1}^{N_{chan}} s_i$ the total signal rate for all channels, and $B_i(\vec{x}_{ij})$ and $S_i(\vec{x}_{ij})$ the probability density functions of the discriminating variables for the background and the signal. These functions are available as the remnants of the simulated signal and background that survived the discrimination. They are usually normalized to one and estimated by kernel techniques, spline interpolation or other methods. Since Q is a product of channels and candidates, a combination of experiments is straight forward.

From Equation 2.21 it is clear that this specific likelihood ratio is not only optimal for searches, but also enables the use of extended event information. The incorporation of the probability density functions has the effect of a **weighted event counting**. The weight is based on the observables contained in the event vectors $\vec{x}_{ij}$. At LEP the Higgs boson searches have used two dimensional event vectors containing the Higgs boson mass estimator and some discriminant variable, often the output of a multivariate discrimination. Some studies on higher dimensional event vectors have been performed on LEP data sets, but “only minor improvements are obtained by extending the likelihood method from two discriminating variables to three and five discriminating variables” [20]. However, in general the use of a larger part of the feature space should have an increased performance potential if the samples are large enough. A further advantage of this particular test statistic is the automatic introduction of systematic uncertainties through the probability density functions.

2.2.3 Limit Computation

At this point the observed test statistic Q_{obs} is calculated by plugging in the observed candidates $\vec{x}_{ij}$ into Equation 2.11 for a chosen signal hypothesis, e.g. for a certain Higgs boson mass. However, the probability density functions of the test statistics $dP_{s+b,b}/dQ$ are needed for the calculation of the confidence limit. Since in general this function is not known, and in particular not in the following search, it is **Monte Carlo** generated. For a large number of simulated experiments, s_i and b_i are generated according to a Poisson distribution and the respective events $\vec{x}_{ij}$ according to the $S_i(\vec{x}_{ij})$ and $B_i(\vec{x}_{ij})$ distributions. Now the CL_s can be computed for all signal hypotheses whose exclusion depends on the fullfillment of Equation 2.17.

2.3 Summary

A search may be seen as a procedure of two steps, discrimination and hypothesis test. Various approaches to both steps have been considered with special emphasis on the techniques used in this work

Before a hypothesis test is performed, the measured data sample is reduced in order to optimize the signal significance. This procedure is called discrimination. A discrimination rule $D(\vec{x};\epsilon)$ based on conditional probability density functions can be proved to be optimal. From this rule the Neyman-Pearson theorem is given and a (likelihood) ratio of joint probability density functions is also an optimal discriminator. Thus either the conditional or the joint p.d.f.s have to be estimated. A third approach is to construct a discriminator rule directly without any probability theory. All discrimination attempts fits into this scheme of three categories. In this work higher order polynomials are used as discriminators. They are derived from a desire of minimizing the within-variance and maximizing the between-variance of the samples. This approach fits into the third category and takes correlations into account, as opposed to sequential cuts and pseudo likelihoods applied by other analyses. The performance is not proved to be optimal and can only be evaluated by comparison.

The hypothesis test may be based on conditional or marginal probability density functions. The classical use of marginal p.d.f.s, the frequentist approach, may result in non-physical intervals. Hence, one may base the test on conditional p.d.f.s, the Bayesian approach. However, this approach includes the arbitrary nature of the prior probability. To overcome the problem of non-physical intervals on the classical path, ordering principles have been proposed. In the modified frequentist analysis these ordering principles are replaced by the CL_s which ensures physical results, to the cost of some overcoverage. This approach has been used in the Higgs boson searches at LEP and is thus adapted to this work. Combined with a likelihood ratio which enables weighted event counting, this is a procedure which provides optimal, but conservative physical results.

Chapter 3

The Standard Model Higgs Boson and the DELPHI Detector

The fascinating path from the electron and positron production through the CERN accelerator complex, the Standard Model theory of electron positron collisions and the DELPHI detection facility have been described hundreds, probably thousands of times in publications and theses. Consequently this chapter is very brief. It only aims to sketch the phenomenological and instrumental framework within which the following Higgs boson search is performed. For detailed information and pedagogical presentations the reader is referred to, for example, [21], [22] and [23].

The CERN accelerator chain as it was in year 2000 is shown in Figure 3.1. Electrons are accelerated to an energy of 200 MeV and steered onto a tungsten target to produce positrons in the LEP Injector Linacs LIL. The particles are fed into the Electron Positron Accumulator EPA and further into the Proton Synchrotron PS where they reach ~3.5 GeV. After a subsequent acceleration in the Super Proton Synchrotron SPS up to around 20 GeV, they are finally led into the Large Electron Positron collider LEP where desired center of mass energies $\sqrt{s}$, two times the beam energy E, are achieved. The radius R of this high technology system is basically constrained by the synchrotron radiation proportional to E^4/R and the aimed energy range. The beam consists of either four or eight bunches of electrons and positrons which are focused to interact in the centers of four detectors whose locations are indicated in Figure 3.1. LEP started to operate in 1989. Until 1995 the center of mass energy was around the Z boson mass, then increased to produce $W^{\pm}$ pairs and search for new particles. In November year 2000, which became the end of the LEP era, a center of mass energy of 209 GeV had been reached.

The event rate for the collisions is as usual given as $\dot{N} = \mathcal{L}\sigma$ where $\mathcal{L}$ is the luminosity and σ is the cross section. If the luminosity is measured, e.g.

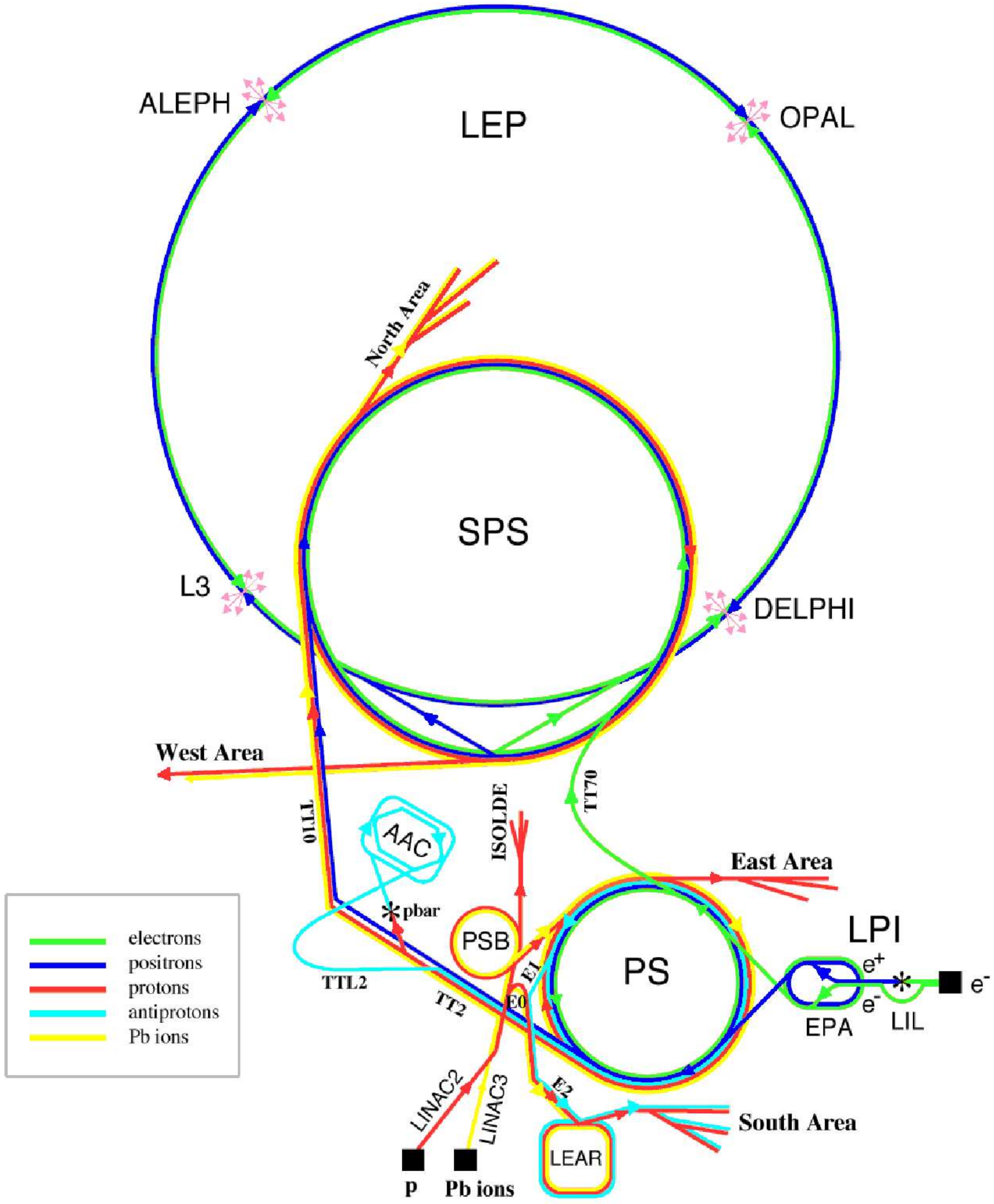


Figure 3.1: Schematic overview of the LEP accelerator complex. Figure taken from Reference [54].

as described in Section 3.2, cross sections can be measured and tested against different hypotheses. The Standard Model Higgs boson properties relevant for the discrimination and hypothesis tests in the following chapters are reviewed in Section 3.1. Then the structure of the DELPHI detector which collected the data subject to the search, is outlined.

3.1 The Standard Model Higgs Boson

In quantum field theory cross sections and decay rates are predicted. These quantities are calculated as functions of the correlation functions which are given by Wick's theorem up to an arbitrary order of precision, provided that the process in question is in the perturbative regime.[1] The action involved, i.e. its Lagrange density, defines the theory.

The Standard Model is a gauge theory with the gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$ which describes the electroweak and strong interactions [24]. The Lagrange density $\mathcal{L}_{SM}$ contains the three generations of fermion fields listed in in Table 1.1 of Chapter 1 and the gauge fields.[2] They have spins, charges and masses. The quarks q and leptons l also have assigned projections of the weak isospin ($I_{e,\mu,\tau,d,s,b} = -1/2$, $I_{\nu,u,c,t} = 1/2$) and are split into right handed singlets and left handed doublets :

$$\begin{pmatrix} \nu_l & q \\ l^- & q\prime \end{pmatrix} = \begin{pmatrix} \nu_l \\ l^- \end{pmatrix}_L, \begin{pmatrix} q \\ q\prime \end{pmatrix}_L, (l^-)_R, (q)_R, (q\prime)_R. \tag{3.1}$$

The right handed fields transform as $SU(2)_L$ singlets and the left handed fields transform as $SU(2)_L$ doublets. For a complete formulation and discussion of $\mathcal{L}_{SM}$ one may for example consult Reference [22]. When $\mathcal{L}_{SM}$ is specified, cross sections, decay rates and equations of motion are in principle computable.

However, $SU(2)_L \times U(1)_Y$ is not a symmetry of the vacuum of the quantum field theory and is spontaneously broken down to the electromagnetic group $U(1)_{EM}$. $SU(3)_C \times SU(2)_L \times U(1)_Y \rightarrow SU(3)_C \times U(1)_{EM}$. This is called the Higgs mechanism [25] and generates masses to the vector bosons $W^\pm$ and Z and the fermions. The symmetry breaking is achieved by introducing a complex scalar field called the Higgs field. Since the masses are experimentally verified, the spontaneous symmetry breaking and the physical massive spin 0 Higgs particle are indispensable constituents of the Standard Model.[3]

---

[1] The standard approach to non-perturbative processes are lattice calculations.

[2] The unphysical scalar gauge fields and the Faddeev-Popov ghosts which have unphysical charges and masses, are omitted.

[3] Alternative resolutions to the introduction of a Higgs field exist. See for example Reference [26] for a review.

Accepting that neutrinos have masses, the Standard Model has 25 input parameters. There are 15 masses, 8 mixing angles and the two couplings for the electroweak and the strong interaction. This number is minimal, albeit not unique. All fields are observed and the massive nature of fifteen of them is verified. There has to be three generations in order to accomplish the observed CP violation, and thus eight mixing angles. Additionally the mass of the Higgs boson increases the set of input parameters till 26. Considering that the Standard Model describes all known matter and that the Higgs boson is the only unobserved constituent of the theory, the enormous interest for Higgs boson searches is evident.

3.1.1 SM Higgs Boson Mass Bounds

Since the input parameters of the Standard Model are not predicted, the Higgs boson mass is unknown, but may be written as a function of the vacuum expectation value ν of the Higgs field

$$m_H = \sqrt{2\lambda}\nu \tag{3.2}$$

which is given by the Fermi coupling and yields $\nu = (\sqrt{2}G_F)^{-1/2} = 246$ GeV. However, self-consistency arguments of the theory may be used to set bounds on the quartic coupling λ and thus m_H.

By considering the one-loop renormalization group equation [21]

$$\frac{d\lambda}{dt} = \beta(t) = \frac{12}{16\pi^2}\lambda^2(t) + \mathcal{O}(\lambda) \tag{3.3}$$

where $t = ln(Q^2/Q_0^2)$, Q_0 is the experimental energy and Q is the cut off energy, an upper limit on the mass is achievable via the so called **triviality** argument [21]. The solution to Equation 3.3 is

$$\frac{1}{\lambda(\nu)} - \frac{1}{\lambda(Q)} = \frac{3}{4\pi^2}ln(\frac{Q^2}{\nu^2}) \Rightarrow \lambda(\nu) = \frac{\lambda(Q)}{1 + \frac{3\lambda(Q)}{2\pi^2}ln(\frac{Q}{\nu})} \tag{3.4}$$

where $Q > Q_0 = \nu$. λ vanishes, becomes "trivial", when the cutoff Q is taken to infinity. If this solution is rewritten as

$$\lambda(Q) = \frac{\lambda(\nu)}{1 + \frac{3\lambda(\nu)}{2\pi^2}ln(\frac{Q}{\nu})} \tag{3.5}$$

it is clear that the self-coupling blows up at a certain Q, the so called Landau pole. In other words, the theory becomes non-perturbative. If $\lambda(Q)$ is taken to infinity, the following relations are obtained

$$\frac{1}{\lambda(\nu)} \geq \frac{3}{2\pi^2}ln(\frac{Q}{\nu}) \Rightarrow \lambda(\nu) \leq \frac{2\pi^2}{3ln(\frac{Q}{\nu})} \Rightarrow m_H^2 \leq \frac{4\nu^2\pi^2}{3ln(\frac{Q}{\nu})} \tag{3.6}$$

A 1 TeV cutoff yields an upper bound about 750 GeV/c^2. Another consistency argument refers to **vacuum stability**. The requirement that the electroweak minimum is absolute, i.e. that the vacuum is stable, limits m_H from below [27]. These theoretical bounds are shown in Figure 1.1 in Chapter 1.

Experimental bounds are obtained indirectly from fits to electroweak observables provided by precision measurements. The Standard Model Higgs boson is sensitive to these values through radiative corrections. The current central value is 91 GeV/c^2 and the 95% limit is around 211 GeV [28]. The χ^2 is shown in Figure 1.1. These indirect experimental bounds favour a light Higgs boson and have motivated the direct Higgs boson searches at LEP.

3.1.2 SM Higgs Boson Production

At LEP electrons were collided with positrons. The Standard Model cross section at tree-level for a Higgs boson production in such collisions has contributions from three diagrams. In Figure 3.2 they are clockwise referred to as the Higgs Strahlung or the Björken process, the WW fusion and the ZZ fusion processes. The last diagram is the associated pair production which is only apparent in the Minimal Supersymmetric Standard Model. Since the search subject to this work is within the Standard Model, and in particular the search for the process $e^+e^- \to H\nu\bar{\nu}$, pair production and ZZ fusion are not treated further.

The cross section for the Higgs Strahlung process at tree-level can be formulated as [29]

$$\sigma(e^+e^- \to HZ) = \frac{G_F^2 m_Z^4}{96\pi s}(\nu_e^2 + a_e^2)\lambda^{1/2}\frac{\lambda + 12m_Z^2/s}{(1 - m_Z^2/s)^2} \tag{3.7}$$

where

$$\lambda = \frac{1}{s^2}[(s - m_H^2 - m_Z^2)^2 - 4m_H^2 m_Z^2],$$

$\sqrt{s}$ is the center of mass energy, $a_e = -1$ and $\nu_e = -1 + \sin^2\theta_W$ where θ_W is the weak mixing angle [30]. The radiative corrections are dominated by photon radiation [31]. However, the contributions to the cross section for the process $e^+e^- \to H\nu\bar{\nu}$, which is the process subject to the search in this work, are received from the Higgs Strahlung with the Z decaying into neutrinos, the WW fusion process and the interference between the two. These contributions make approximately 20% of the Standard Model Higgs boson production. For center of mass energies above the Z resonance the differential cross section for this process takes the form [30]

$$\frac{d\sigma(H\nu\bar{\nu})}{dE_H d\cos\theta} = \frac{G_F^3 m_Z^8 p}{\sqrt{2}\pi^3 s}(3\mathcal{G}_S + \mathcal{G}_I + \mathcal{G}_W) \tag{3.8}$$

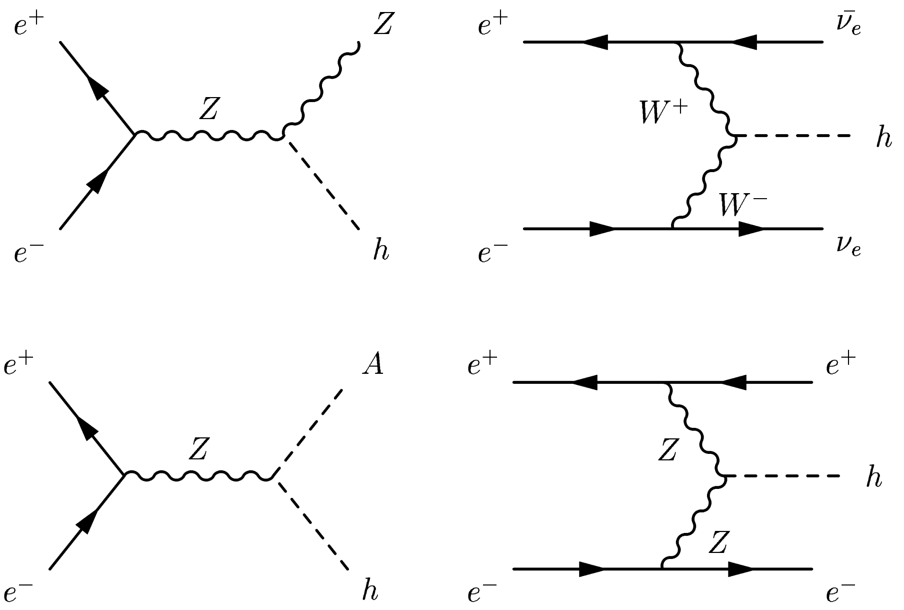


Figure 3.2: The tree-level production diagrams of the Higgs bosons in the Standard Model and the Minimal Supersymmetric Standard Model (MSSM) at LEP. Clockwise from the upper left diagram the Higgs Strahlung, the WW fusion, the ZZ fusion processes and the associated pair production are shown. The pair production is a pure MSSM process.

where the contributions are

$$\mathcal{G}_S = \frac{\nu_e^2 + a_e^2}{96} \frac{s s_\nu + s_1 s_2}{(s - m_Z^2)^2[(s_\nu - m_Z^2)^2 + m_Z^2 \Gamma_Z^2]}$$

$$\begin{aligned}
\mathcal{G}_I = {} & \frac{(\nu_e + a_e)\cos^4\theta_W}{8} \frac{s_\nu - m_Z^2}{(s - m_Z^2)^2[(s_\nu - m_Z^2)^2 + m_Z^2 \Gamma_Z^2]} \\
& \times[2 - (h_1 + 1)\log\frac{h_1 + 1}{h_1 - 1} - (h_2 + 1)\log\frac{h_2 + 1}{h_2 - 1} \\
& +(h_1 + 1)(h_2 + 1)\frac{\mathcal{L}}{\sqrt{r}}]
\end{aligned}$$

$$\begin{aligned}
\mathcal{G}_W = {} & \frac{\cos^8\theta_W}{s_1 s_2 r}\{(h_1 + 1)(h_2 + 1)[\frac{2}{h_1^2 - 1} + \frac{2}{h_2^2 - 1} - \frac{6 s_\chi^2}{r} + \\
& (\frac{3 t_1 t_2}{r} - c_\chi)\frac{\mathcal{L}}{\sqrt{r}}] - [\frac{2 t_1}{h_2 - 1} + \frac{2 t_2}{h_1 - 1} + \\
& (t_1 + t_2 + s_\chi^2)\frac{\mathcal{L}}{\sqrt{r}}]\}
\end{aligned}$$

and several abbreviations have been used. The Higgs boson momentum is $p = \sqrt{E_H^2 - m_H^2}$, the energy of the neutrino pair is $\epsilon_\nu = \sqrt{s} - E_H$ and its invariant mass squared is $s_\nu = \epsilon_\nu^2 - p^2$. The additional terms are :

$$\begin{array}{lclclcl} s_{1,2} & = & \sqrt{s}(\epsilon_\nu \pm p\cos\theta) & , & t_{1,2} & = & h_{1,2} + c_\chi h_{2,1} \\ h_{1,2} & = & 1 + 2m_W^2/s_{1,2} & , & r & = & h_1^2 + h_2^2 + 2c_\chi h_1 h_2 - s_\chi^2 \\ c_\chi & = & 1 - 2ss_\nu/(s_1 s_2) & , & s_\chi^2 & = & 1 - c_\chi^2 \\ \mathcal{L} & = & \log\frac{h_1 h_2 + c_\chi + \sqrt{r}}{h_1 h_2 + c_\chi - \sqrt{r}}. & & & & \end{array}$$

The total cross section for $e^+e^- \to H\nu\bar{\nu}$ is obtained by integration over $-1 < \cos\theta < 1$ and $m_H < E_H < [(\sqrt{s}/2)(1 + (m_H^2/s)]$. It is shown in Figure 3.3 where the effect of interference and photon radiation is illustrated for the $H\nu_e\bar{\nu}_e$ contribution only. For low masses the development is dominated by the Higgs Strahlung process and follows Equation 3.7. At the kinematical limit, $\sqrt{s} - m_Z \sim 115$ GeV, given by the center of mass energy and the Z boson mass, HZ production drops rapidly and the WW fusion becomes dominant. At this point the interference also contributes significantly. Searches around this kinematical limit therefore depend crucially on the inclusion of these processes.

3.1.3 SM Higgs Boson Decay

The width of the Standard Model Higgs boson at masses accessible at LEP is dominated by decays into pairs of fermions and antifermions. The partial width at tree-level for these processes is given as

$$\Gamma(H \to f\bar{f}) = \frac{N_c G_F m_f^2}{4\sqrt{2}\pi} m_H \beta_f^3 \tag{3.9}$$

where N_c is 1 for leptons and 3 for quarks and $\beta_f = \sqrt{1 - 4m_f^2/m_H^2}$ is the velocity of the fermions in the center of mass frame. Higher order corrections to this approximation are described in References [32] and [33]. The branching ratios, $B_i = \Gamma_i/\Gamma_{tot}$, above 0.1% are shown as functions of the Higgs boson mass in Figure 3.4 and are dominated by decays into $b\bar{b}$ pairs at LEP.

The branching ratios for decays into pairs of gauge bosons are also shown in Figure 3.4. Their formulae may be found in the References [21], [34] and [35]. The WW decay becomes dominant around $m_H = 130$ GeV/c^2 as the mass approaches the $2 \times m_W$ threshold. The ZZ decays are less frequent due to the smaller couplings of neutral currents. The very low branching ratio for decays into $\gamma\gamma$ is important for a search for a Higgs boson in the mass range from 110 GeV/c^2 to 150 GeV/c^2 at LHC due to a large background for hadronic decays [36]. Anyway, at LEP these decays are of subdominant importance.

The total width of the Standard Model Higgs boson is shown in the lower graph of Figure 3.4. At $m_H = 40$ GeV/c^2 it is ~1 MeV and then increases,

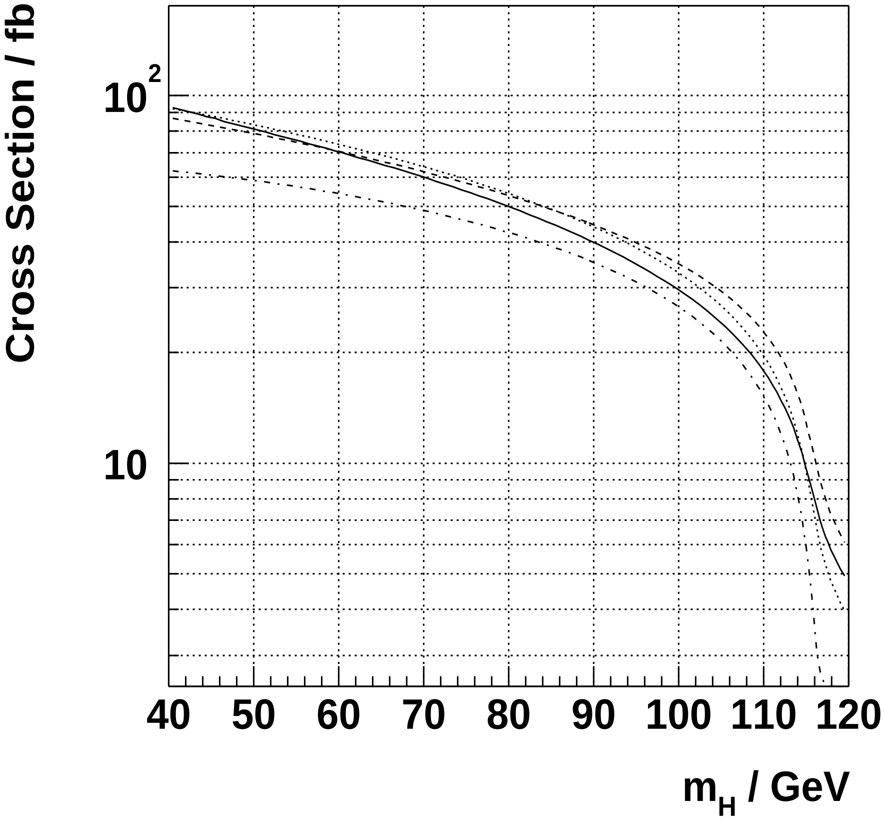
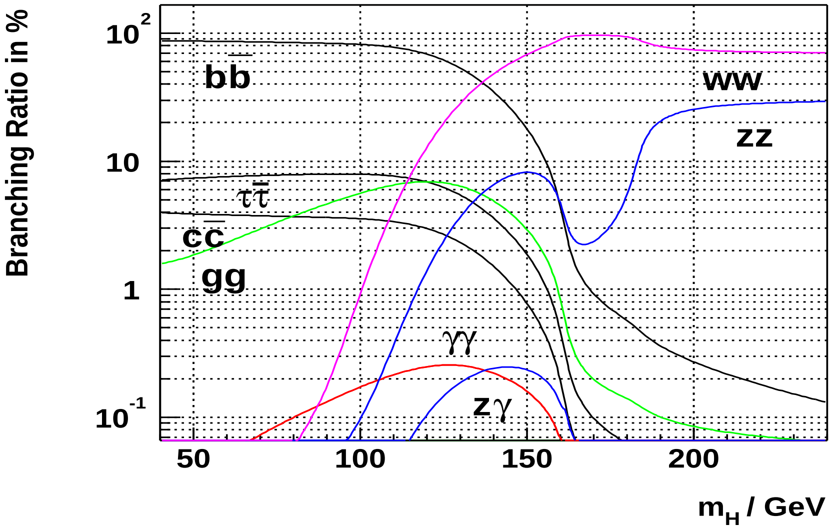


Figure 3.3: The cross section versus the Higgs boson mass for $e^+e^- \to H\nu_e\bar{\nu}_e$ at $\sqrt{s} =$ 206.5 GeV. The Higgs Strahlung cross section is dashed-dotted, the Higgs Strahlung plus boson fusion cross section is dotted, the cross section for Higgs Strahlung, boson fusion and interference is dashed and the latter with radiative corrections is solid. The values are calculated with version 3 of the HZHA generator [39]. Only the $H\nu_e\bar{\nu}_e$ contribution is shown.



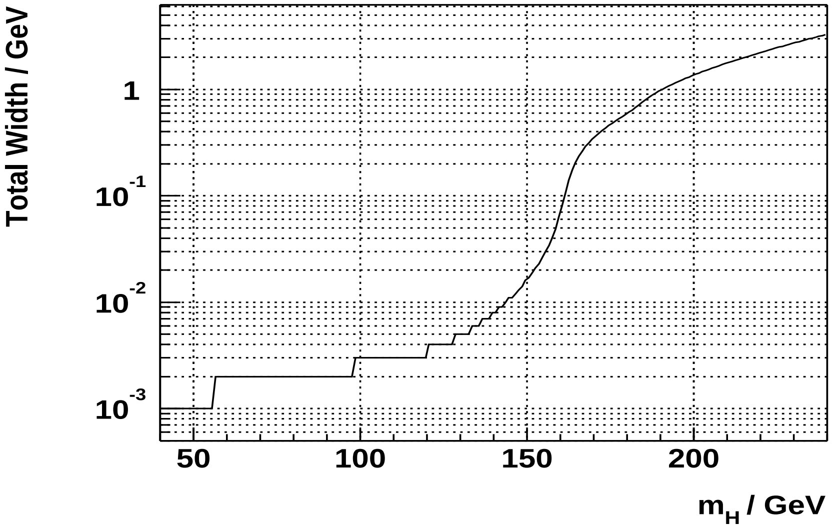


Figure 3.4: In the upper graph the branching ratios for the Standard Model Higgs boson decays is shown in %. In the lower the total width is shown as a function of the mass. The steps at low values only reflect the 10^{-3} precision used. The values are computed with version 3 of the HZHA generator [39].

especially after the contributions from the WW and ZZ decays become dominant in the $m_H = 150$ GeV/c^2 region. At LEP energies the Standard Model Higgs boson is thus relatively narrow with a width around 10 MeV. Further it decays dominantly into $b\bar{b}$ pairs, the next to dominant decay channel being an order of magnitude smaller.

3.2 The DELPHI Detector

LEP produced collisions for four experiments whose detectors were designed to have different advantages. The DELPHI (DEtector with Lepton, Photon and Hadron Identification) detector was a 4π general purpose detector with emphasis on **particle identification** achieved with ring imaging Cherenkov counters, three-dimensional high granularity information in most of its components and high precision vertex determination [37],[23]. The standard cartesian coordinate system was chosen to have its z-axis along the beam, with positive z in the direction of the beam electrons. The x-axis points toward the center of the LEP ring. The y-axis is directed upwards. A polar coordinate system defines the polar angle θ with $\theta = 0$ along the positve z-axis, the azimuth angle ϕ in the plane orthogonal to the beam and the radial coordinate $R = \sqrt{x^2 + y^2}$. The detector is a system of several subdetectors shown schematically in Figure 3.5. The barrel part which covers polar angles from typically 40° to 140° and one of two end caps are visible. A superconducting solenoid provides a highly uniform magnetic field of 1.23 T parallel to the beam. The high backward-forward and azimuthal symmetry reflects the basic distribution of outgoing particles in beam-symmetric collisions.

The **tracking** of charged particles is provided by several subdetectors. The microvertex detector VD is a 3-layer silicon strip detector next to the beam pipe and the collision point. It has a high Rϕ resolution for charged tracks and is essential for the heavy flavour tagging, e.g. b-tagging, which is an important tool in the Higgs searches. Outside the VD the inner detector ID is located. The ID is based on drift chambers and straw tubes. In the barrel outside the ID the time projection chamber TPC, the main tracking device, is situated. Outside the barrel RICH detector, which encloses the TPC, the outer detector OD is installed. Two tracking detectors, the forward chambers FCA and FCB, are placed in the end caps. The muon chambers in the barrel MUB, in the end caps MUF and surrounding the end caps SMC are drift chambers used to detect muons and measure their momenta.

Various **scintillator counters** are used for fast triggering of cosmics and beam events. Additional counters, called hermeticity taggers, were installed to cover gaps in the original set up, especially between the barrel and the end caps. The time-of-flight counters TOF have a detection efficiency for minimum ionizing particles of ~99.9%. The forward hodoscope HOF is located in the

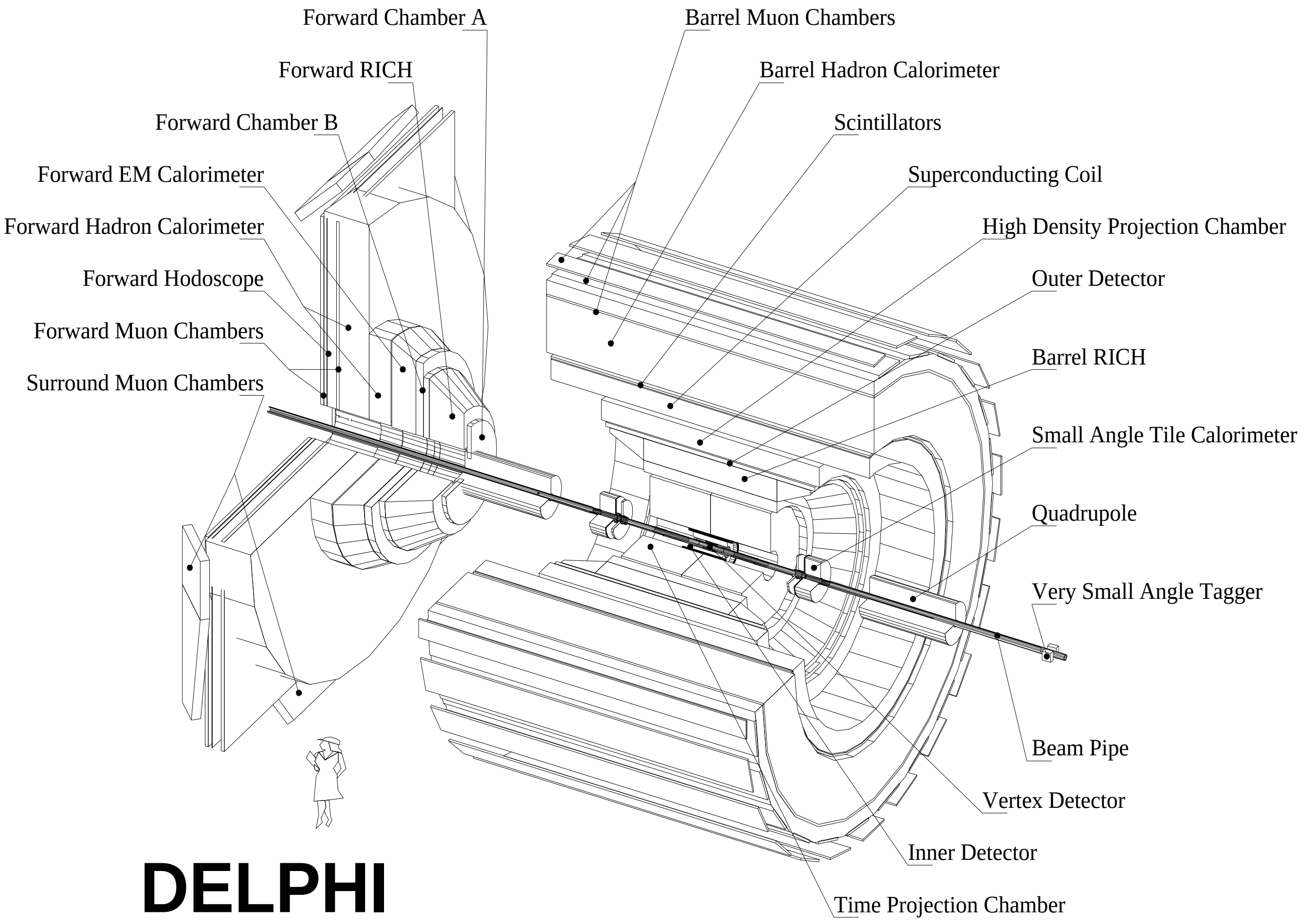


Figure 3.5: Schematic overview of the DELPHI detector system. The barrel and one of two end caps are shown. The subdetectors and the size relative to a scientist are indicated.

gap between each end cap and the second MUF layer. The hermeticity taggers are mainly installed in the z=0 region where many subdetectors in the barrel are divided, and in the $\theta = 40^\circ$ region where the barrel meets the end caps.

The **calorimeters** used to measure electromagnetic and hadronic energy deposits are located both in the barrel and in the end caps. The hadron calorimeter HCAL is a wire chamber detector placed as indicated in Figure 3.5. Its energy resolution is $\sigma_E/E = 0.21/\sqrt{E} \oplus 1.12/\sqrt{E}$. The high density projection chamber HPC is the electromagnetic barrel calorimeter. It is made up of modules of lead and gas to produce electromagnetic showers which are detected via the ionized gas. The energy resolution is $\sigma_E/E = 0.043 \oplus 0.32/\sqrt{E}$. The forward electromagnetic calorimeter FEMC has a Cherenkov lead glass detector in each end cap whose signals are read out by photomultiplyers. The energy resolution is $\sigma_E/E = 0.03 \oplus 0.12/\sqrt{E} \oplus 0.32/E$.

Two other calorimeters are responsible for the **luminosity** measurement whose importance is given by the theoretical formulation of the event rate. Due to its high and theoretically well known cross section at low angles, counting of Bhabha scattering events is used to determine $\mathcal{L}$. The Small Tile angle Calorimeter, the STIC, is the principle luminosity monitor in the DELPHI detector system. It consists basically of two cylinders of lead and scintillator layers located in each end cap. The energy resolution for electrons with 45 GeV is ~3% [23]. The second luminosity monitor is the very small angle tagger, the VSAT. It consists of two modules which is located respectively ouside each end cap close to the beam pipe. Each module is a rectangular W-Si calorimeter stack. The energy resolution at 45 GeV is ~5% and at 95 GeV ~4% ($\sigma_E/E = 0.35/\sqrt{E}$).

The DELPHI **trigger** consists of four levels. The first level trigger T1 accepts events $3.5\mu s$ after the beam cross over signal BCO. It is based on uncorrelated information from the ID, the OD, the FCA, FCB and the scintillator arrays TOF, HOF and those embedded in the HPC, the FEMC and the MUB. The T1 trigger rate is typically ~700 Hz for a normal luminosity of $\sim 1.5 \times 10^{31}$ $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. Since correlations are not taken into account, noise from individual detectors is the dominant background for T1. The second level trigger T2 takes its decision $39\mu s$ after BCO. It includes information from the TPC, the HPC and the MUF whose signals are delayed due to long drift times. At this level correlations between the sub-detectors are introduced and the result is a rate of roughly 4.5 Hz. About 20% of the events are now real physics events. At this level the main background sources are beam gas interactions, synchrotron radiation and cosmic ray events. Level three T3 is a software trigger based on the logic applied at level T2, but using additional information. It reduces the T2 rate by a factor ~2. The last level T4 accepts special topologies of interest. The trigger rate is now roughly at 1 Hz. The combined trigger efficiency for both lepton pairs and hadronic events is very close to 100% [23].

3.3 Summary

The Standard Model may be used to impose bounds on the Higgs boson mass through self-consistency arguments of the theory. Upper bounds achieved in this manner are not within the reach of LEP. However, indirect experimental bounds strongly prefer a Higgs boson mass around 100 GeV/c^2. This has encouraged intensive searches in the data collected by the LEP experiments. In the electron positron collisions produced by LEP, Higgs bosons with masses less than ~120 GeV/c^2 are expected to be created. The $e^+e^- \to H\nu\bar{\nu}$ signal, which accounts for ~20% of the expected Higgs bosons, is dominated by the Higgs Strahlung at low masses. However, at the kinematical limit of the latter fusion and interference contribute at the same order. Considering the dominant decay channel of the Standard Model Higgs boson at LEP energies, the $b\bar{b}$ channel, the 4π DELPHI detector with its b-tagging and particle identification capabilities is suitable for the detection of such events.

Chapter 4

Search for $e^+e^- \to H\nu\bar{\nu}$ with DELPHI

After the classificational review of a search in Chapter 2 and the brief description of the Standard Model and the instrumental aspects of the DELPHI detector in Chapter 3, a search for the Standard Model process $e^+e^- \to H\nu\bar{\nu}$ is now presented.

The chapter is arranged in eight sections. The first completes the description of the experimental signatures which was started in Chapter 3. Both the signal signatures, which define the search channel, and the various background processes are described. Section 4.2 concerns the simulated and the measured data samples. It aims to provide the minimal set of information necessary for a hypothetical reproduction of the following analysis.[1] Section 4.3 reviews the standardized event selection which was used. Section 4.4 presents the preselection used to achieve a certain level of agreement between measured and simulated data before the application of the statistical machinery. Section 4.5 introduces all the variables used in the discrimination and Section 4.6 deals with the discrimination and the hypothesis testing. Higgs bosons are searched for at four different masses, $m_H = 80$ GeV/c^2, $m_H = 90$ GeV/c^2, $m_H = 100$ GeV/c^2, and $m_H = 115$ GeV/c^2. In each search hypothesis tests are performed. Studies of the systematical uncertainties are reviewed in Section 4.7. In Section 4.8 the chapter is summarized and the best limits collected.

---

[1] This is to accommodate the scientific ideal of reproducible results, a normative goal only approximately met in practice.

4.1 The Missing Energy Channel - Signal and Backgrounds

The diagrams for Higgs boson production at tree-level in the Standard Model were shown in Figure 3.2 in Chapter 3. The various decay modes resulting from the different Higgs boson production diagrams define the search channels. As was illustrated in Figure 3.4 about 90% of the Higgs bosons decay to a $b\bar{b}$ quark pair at LEP energies. The Z bosons decay into a $q\bar{q}$ pair in ∼70% of the cases. Together with the Higgs boson decay these decays then ideally, i.e. in cases without so-called jet splitting, cause a four jet signal. This four jet channel has then the highest cross section and was thus the most promising place to look for Higgs bosons at LEP. The channel with the next to the highest cross section is the missing energy channel. About 20% of the Z bosons decay into a $\nu\bar{\nu}$ pair. According to the purely weak cross section for neutrinos, the $\nu\bar{\nu}$ pair escapes the detector and leaves a topology with ideally two acoplanar b-quark jets with the mass of the Higgs boson, accompanied by a large missing energy and a large missing mass close to the Z mass.[2] The following search for $e^+e^- \to H\nu\bar{\nu}$ is done in this channel. For further details on all channels, see Reference [29].

The signal signature in the missing energy channel is then produced by $e^+e^- \to HZ \to H\nu\bar{\nu}$ and includes s-channel Higgs Strahlung and t-channel WW fusion. The Higgs-Strahlung dominates the total cross section far under the kinematic threshold for on-shell ZH production, the center of mass energy minus the Z boson mass. Above the threshold, where the inteference between the two processes WW fusion and the Higgs Strahlung becomes positive, both contributions, inteference included, are of the same order [30]. Thus, at LEP where limits on the Higgs boson mass have been pushed as high as possible, this fact was incorporated in the searches, i.e. in the signal generators [38], [39]. In on-shell Z production the mass of a discovered Higgs boson may be fitted by constraining the missing mass to the Z mass. Although the mass may be measured by the $b\bar{b}$-jets alone, this fit improves the mass resolution.

The possible tree-level background diagrams in the Standard Model are shown in Figure 4.1. Analysing the different possibilities, the background is, with the exception of $ZZ \to b\bar{b}\nu\bar{\nu}$ and the $Z\nu\bar{\nu}$ processes with $Z \to b\bar{b}$, in all other cases (ideally) distinguishable from the signal. As already pointed out in Reference [29], two-fermion processes without initial state radiation and four jet

---

[2] The acoplanarity (angle) of two jets is defined as

$$\pi - \arccos\left(\frac{\vec{p}_1^{\,T} \cdot \vec{p}_2^{\,T}}{|\vec{p}_1^{\,T}||\vec{p}_2^{\,T}|}\right)$$

where $\vec{p}_{1,2}^{\,T}$ are the transverse momenta of the jets, i.e. the momentum projected into the plane ortogonal to the beam axis.

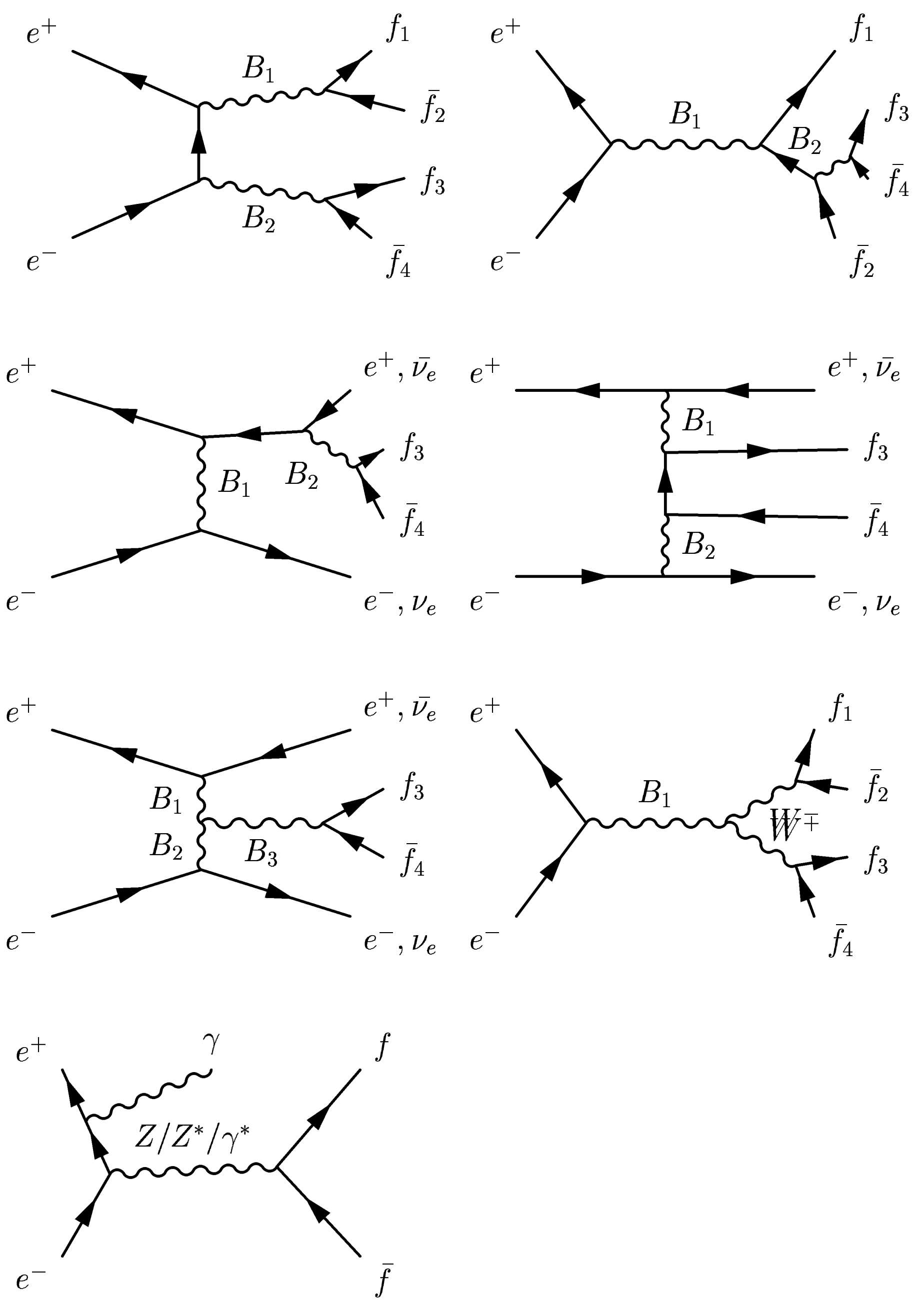


Figure 4.1: Classes of possible tree-level background diagrams at LEP. $B_{1,2,3}$ are Z, γ or $W^\pm$ so as the quantum numbers are conserved. The first row shows the conversion and the annihilation diagrams. The second row contains the Bremsstrahlung and the multiperipheral (two photon) diagrams. The third row shows the non-abelian fusion and annihilation diagrams. The last diagram illustrates the two fermion production processes.

events have no missing energy. Two-fermion processes with initial state radiation have no missing mass. The processes $e^+e^- \rightarrow q\bar{q}(\gamma)$, $WW \rightarrow l\nu + two\ jets$ and those resulting in Ze^+e^- have no missing mass and isolated particles. The processes $e^+e^- \rightarrow q\bar{q}\ (\gamma\gamma)$ and $\gamma\gamma \rightarrow q\bar{q}$ have no missing transverse momentum and a small acoplanarity angle. The processes $e^+e^- \rightarrow (e)\ \nu W$, $WW \rightarrow \tau\nu + two\ jets$ and $ZZ \rightarrow q\bar{q}\nu\bar{\nu}$ have light quark jets. Thus, with a signal with missing mass, missing energy and acoplanarity, there is a potential for successful discrimination of several of the background processes.

The implementations of the theoretical cross sections [29] can be studied in the respective simulation software which was used to generate the samples (see Section 4.2). Extensive studies of the various background processes have been undertaken during the years of LEP operation, and the modeling is accepted to be very accurate. This work relies to a large extent on the fruits of these previous studies.

4.2 Data Samples, Cross Sections and Luminosities

All the high energy data taken with the DELPHI detector in year 2000 have been used. They were recorded at center of mass energies from around 200 GeV up to 209.2 GeV. This data set corresponds to an integrated luminosity of 224.2 pb^{-1} [5]. Since one of the twelve sectors in the Time Projection Chamber, described in Chapter 3, of the DELPHI detector broke down in September 2000, the data is divided into two parts, the amount collected before the breakdown and the amount collected after. The first part corresponds to an integrated luminosity of 164.1 pb^{-1}, the second to 60.1 pb^{-1}. This division of the data led to a separation of the following analysis into two channels with different simulations of signal and background samples. In the following the first is referred to as the E1 channel or E1 sample, the second as the U1 channel or U1 sample.

The Higgs boson production processes and the Standard Model background processes were simulated with the generators shown in Table 4.1. A much used classification into four groups has been adapted. The background is divided into two fermion processes, four fermion processes and two photon processes, respectively denoted with 2f, 4f and 2γ. These classes are divided further according to the simulated samples. The channel names arising will be used as references in the following. The 2FQQ channel contains final states with a quark pair and possible photons. The 2FEE channel contains a e^+e^- pair from Bhabha scattering. The 2FMM and the 2FTT channels denote final states with a $\mu^+\mu^-$ and a $\tau^+\tau^-$ pair respectively. Final states with four fermions originating from charge currents are grouped into 4FCC, those from neutral currents

Table 4.1: Simulated processes at center of mass energy $\sqrt{s} = 206.5$ GeV and the simulation software used. Description of the respective simulation software may be found in the References [39], [42], [43], [44], [45] and [46].

	Channel	Software	Cross section / pb
	signal	HZHA03 with Pythia 6.156	see Table 4.2
2 f	2FQQ	KK2F 4.14, DELPHI int. v. 3	79.237 ± 0.083
	2FEE	BHWIDE 1.01	1008 ± 1
	2FMM	KK2F 4.14, DELPHI int. v. 3	7.572 ± 0.008
	2FTT	KORALZ 4.02	6.703 ± 0.010
4 f	4FCC	WPHACT 21	20.143 ± 0.007
	4FNC	WPHACT 21	7.904 ± 0.023
2 γ	GGQQ	Pythia 6.143	14080 ± 16
	GGLL	WPHACT 21	62.59 ± 0.17

Table 4.2: Signals generated at 206.5 GeV with HZHA version 3.0. The second column shows the cross section at the respective Higgs boson mass, the third the expected number of events at an effective integrated luminosity of 224 pb^{-1}. The fourth and fifth columns contain the number of generated events at the respective masses for the two samples.

m_H / GeVc^{-2}	$\sigma_{e^+e^-\to h\nu\bar{\nu}}$ / pb	N_{exp}	N_{gen}^{U1}	N_{gen}^{E1}
40	0.20480	43.8	4781	5000
50	0.18240	39.0	5000	5000
60	0.16069	34.4	5000	4999
70	0.15201	32.5	5000	4998
80	0.12680	27.1	5000	4998
85	0.11311	24.2	4999	5000
90	0.09942	21.3	4999	5000
95	0.08651	18.5	4999	4997
100	0.07207	15.4	4997	4999
105	0.05618	12.0	4999	4996
110	0.03817	8.2	4996	4498
114	0.02004	4.3	4999	4999
115	0.01489	3.2	4999	4998
116	0.01088	2.3	4997	4999
120	0.00567	1.2	4999	4998

into 4FNC. Two photon processes are called either GGQQ or GGLL, depending on whether they result in hadrons or leptons. The tree-level diagrams for these processes can be identified in Figure 4.1. The signals, $e^+e^- \to H\nu\bar{\nu}$, were generated with Higgs boson masses from 40 GeV/c^2 up to 120 GeV/c^2 in steps shown in Table 4.2. The table contains the generated number of signal events at the different masses and the cross sections. The cross sections for the background processes are shown in Table 4.1. Already the difference between signal and background cross sections, being orders of magnitude, indicates the need for discrimination and a challenging benchmark for testing the multivariate method using a generalized polynomial discriminator. This becomes even more striking when the expected numbers of signal events in Table 4.2 are compared with the expected number of background events in Table 4.3. However, the size of the samples, representing typically 10^2 times the expected number of events, is promising for an effective statistical discrimination procedure.

The fragmentation of the events was done with version 6.156 of Pythia, except for the two photon processes where version 6.143 was used. After the fragmentation the events were fed into the full detector simulation package DELSIM [40]. The configuration of DELSIM took the breakdown of sector six into account. From this point, after the full simulation of the samples, all

Table 4.3: Expected number of events N_{exp}, number of generated events N_{gen} and the ratio of the two N_{gen}/N_{exp} for the E1 and U1 samples.

Channel	N^{E1}_{exp}	N^{E1}_{gen}	$N^{E1}_{gen}/N^{E1}_{exp}$	N^{U1}_{exp}	N^{U1}_{gen}	$N^{U1}_{gen}/N^{U1}_{exp}$
2FQQ	12495.7	732073	58.6	4461.8	596884	133.8
2FEE	158961.6	1986371	12.5	56760.5	993636	17.5
2FMM	1194.1	99998	83.7	426.4	100000	234.5
2FTT	3176.1	73421	69.5	377.5	36220	96.0
4FCC	3176.6	759255	239.0	1134.3	922700	813.5
4FNC	1246.5	423777	339.9	445.1	401997	903.2
GGQQ	2220416	900000	0.4	792844.8	1080000	1.4
GGLL	9870.4	953074	96.6	3524.4	976774	277.1

measured and simulated data were subject to the same treatment respective to the channels.[3]

The uncertainties on the cross sections listed in this section refer to the respective generator software. By assuming an relative error of 1% on all cross sections, a conservative estimate should be ensured.[4]

4.3 Selection and Reconstruction of Hadronic Events

The events which survived the trigger chain described in Section 3.2 were exposed to a loose discrimination suited to select hadronic events. A **standardized** software filter for the year 2000 data was applied. This filter requires the charged particles to have momentum greater than 100 MeV/c and to origin from a defined interaction region. This region was defined to be within $4cm/\sin\Theta$ along the beam axis, Θ being the polar angle of the track, and within 4 cm in the plane orthogonal to the beam axis. Neutral particles are defined either as energy clusters in the calorimeters not associated with charged tracks, or as reconstructed vertices of photon conversions, or as interactions of

---

[3] This work has utilized completely prepared datasets, simulated and recorded by the DELPHI collaboration.

[4] If a conservative uncertainty on the luminosity of 1% is assumed (see next subsection), all event numbers in this section then possess a $\sqrt{0.01^2+0.01^2} = 1.4\%$ uncertainty.

neutral hadrons, or as decays of neutral particles in the tracking volume. In the calorimeters neutral clusters with energy larger than 300 MeV are selected. Except for identified leptons, the $\pi^{\pm}$ mass was used for charged particles. For neutral electromagnetic clusters zero mass was assigned and for neutral hadronic clusters the K^0 mass.

For **jet reconstruction** the DURHAM algorithm was used [41]. Iteratively, the energies $E_{i,j}$ and the momenta of the two tracks in the event with the smallest distance $y_{ij} = 2\min(E_i^2, E_j^2)/E_{vis}^2(1-\cos\alpha_{ij})$, where E_{vis} is the energy of all detected particles' energy and α_{ij} the angle between the tracks, are added up. The algorithm allows for a cut value y_{max}. The procedure is repeated until all particles are associated to one of the required number of jets.

The center of mass energy distributions for both the measured E1 and the measured U1 samples at this level are shown in Figure 4.2. The U1 sample, which was the latest recorded, has a bigger fraction of high energy events than the E1 sample due to the continuous LEP improvements. The plots also show the effect of a **run quality cut**. This cut was designed by the DELPHI collaboration and adapted to this analysis to reject runs with detected problems in the calorimeters and the Time Projection Chamber. Removing 0.39% of the E1 events and 0.63% of the U1 events, this cut leads to effective luminosities of 0.961×164.1 pb^{-1} $= 157.70$ pb^{-1} for the E1 sample and 0.937×60.1 pb^{-1} $= 56.31$ pb^{-1} for the U1 sample. The introduction of effective luminosities is necessary since the run quality cut does not apply to the simulated samples.

The binomial nature of cuts yields binomial uncertainty. The standard deviation is thus given by $\sqrt{\epsilon(1-\epsilon)/N}$, where ϵ denotes the efficiency of the cut and N the number of events before the cut. The run quality cut then yields an uncertainty of 0.09% for the E1 sample and 0.14% for the U1 sample. Considering that the typical uncertainty on the measured luminosities is around 0.5%, a relative uncertainty on the effective luminosities of 1% should ensure a conservative estimation. When referred to luminosities further on, the effective numbers are meant and this **relative uncertainty of 1%** is used.

The sizes of the simulated and measured samples at this level are listed in Table 4.4. The uncertainties include 1% uncertainty for the luminosity and 1% uncertainty for the cross sections.[5] Column three and five show that the signal efficiency is stable at 91% and 92% for all masses. For $e^+e^- \to e^+e^-$,

---

[5] If $N_{Norm} = L\sigma N/N_0$ denotes the normalized number of events after the cut, N is the number of events after the cut, N_0 is the original size of the simulated sample, L the integrated luminosity and σ the cross section, and the well known first order approximation formula for propagation of errors is used, the uncertainty on N_{Norm} becomes

$$\Delta N_{Norm} = \sqrt{(0.01L)^2(\sigma\frac{N}{N_0})^2 + (0.01\sigma)^2(L\frac{N}{N_0})^2 + N(1-\frac{N}{N_0})(\frac{L\sigma}{N_0})^2}$$

where binomial uncertainty $\Delta N_{bin} = \sqrt{N(1-N/N_0)}$ has been used for the cut. The relative uncertainty follows :

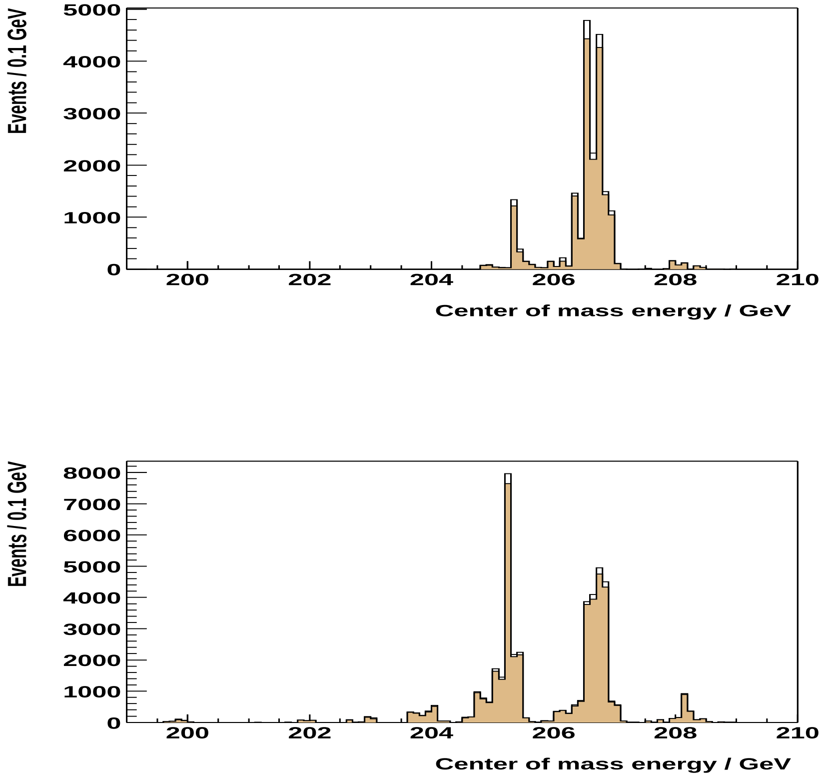


Figure 4.2: Data versus center of mass energy. Top shows the U1 sample, bottom the E1 sample. In both plots the effect of a run quality cut is shown. The shaded plots represent the data after the cut.

Table 4.4: Normalized numbers of events (TE1 and TU1) and effiencies (EFF TE1 and EFF TU1) for the E1 and U1 samples after track selection. 1% uncertainty is assumed for both luminosity and cross sections. One standard deviation binomial uncertainty is used for the cut.

Process	TE1	EFF TE1	TU1	EFF TU1
$m_H = 40$	29.57 ± 0.43	0.92	10.63 ± 0.15	0.92
$m_H = 50$	26.50 ± 0.39	0.92	9.48 ± 0.13	0.92
$m_H = 60$	23.20 ± 0.34	0.92	8.30 ± 0.12	0.92
$m_H = 70$	22.10 ± 0.32	0.92	7.86 ± 0.11	0.92
$m_H = 80$	18.51 ± 0.27	0.92	6.53 ± 0.09	0.92
$m_H = 85$	16.46 ± 0.24	0.92	5.84 ± 0.08	0.92
$m_H = 90$	14.40 ± 0.21	0.92	5.16 ± 0.07	0.92
$m_H = 95$	12.51 ± 0.18	0.92	4.46 ± 0.06	0.92
$m_H = 100$	10.31 ± 0.15	0.91	3.72 ± 0.05	0.92
$m_H = 105$	8.13 ± 0.12	0.92	2.90 ± 0.04	0.92
$m_H = 110$	5.52 ± 0.08	0.92	1.97 ± 0.03	0.92
$m_H = 114$	2.86 ± 0.04	0.91	1.02 ± 0.02	0.91
$m_H = 115$	2.12 ± 0.03	0.91	0.76 ± 0.01	0.91
$m_H = 116$	1.56 ± 0.02	0.91	0.559 ± 0.008	0.91
$m_H = 120$	0.81 ± 0.01	0.91	0.291 ± 0.004	0.91
2FQQ	10859 ± 154	0.86	4197 ± 59	0.93
2FEE	79 ± 3	0.0004	27.3 ± 1.2	0.0005
2FMM	0.18 ± 0.05	0.0001	0.04 ± 0.01	0.0001
2FTT	47 ± 1	0.044	15.4 ± 0.4	0.04
4FCC	2708 ± 38	0.85	676 ± 9.5	0.59
4FNC	414 ± 6	0.33	153 ± 2.2	0.34
GGLL	957 ± 14	0.096	340 ± 4.9	0.096
GGQQ	15592 ± 295	0.007	5580 ± 101	0.007
SUM BKG	30656 ± 335	0.26	10991 ± 118	0.33
DATA	42671	-	18355	-

Table 4.5: The effect of a set of two photon rejection cuts as described in Reference [48]. The first and the third row show the effect on the samples used in this analysis. The second and the fourth row show the effect on the samples in the independent analysis [48].

	Data	Exp. Background	2 Fermion	4 Fermion
E1	13546	13322 ± 21	10000	2946
E1 Ref	13546	13361 ± 11	10023	2964
U1	4569	4269 ± 7	3514	755
U1 Ref	4571	4828 ± 7	3617	1018

$e^+e^- \to \mu^+\mu^-$, $e^+e^- \to \tau^+\tau^-$ and the two photon processes the track selection filter is already very efficient. The two fermion process with quarks in the final state and the four fermion processes are prevailing already at this early level. This corresponds to the expectation which was briefly discussed in Section 4.1. However, the size of the measured samples still exceeds the simulated ones significantly. This indicates that there are measured processes not taken into account by the simulations. To identify and reject the events from these processes one univariate cut is introduced.

4.4 Minimal Univariate Discrimination

Before the univariate discrimination introduced to achieve a reasonable agreement between measured and simulated data is presented, the effect of a set of cuts used in an independent analysis is studied. This set of cuts required for each event at least eight charged tracks, at least one track with a transverse momentum greater than 1.5 GeV/c to originate from the region $z < 1$ mm and $R < 0.2$ mm, the charged energy to be greater than 16% of the center of mass energy, the energy deposit of tracks only detected in the vertex detector and the inner detector to be less than 50% of the charged energy, the transverse energy to be greater than 15% of the center of mass energy, the sum of the absolute values of the momenta along the thrust axis is greater than 25% of the center of mass energy and that the leading electromagnetic shower has more energy than 45% of the center of mass energy. The measured events and the

---

$$\frac{\Delta N_{Norm}}{N_{Norm}} = \sqrt{0.01^2 + 0.01^2 + (1 - \frac{N}{N_0})\frac{1}{N}}$$

normalized expectations are listed in Table 4.5. Implicit it is clear that the two photon background, roughly the difference between the expected background and the sum of the two fermion and the four fermion backgrounds, is reduced significantly from the track selection level shown in Table 4.4. The agreement with the reference numbers is total for the E1 samples. For the U1 samples the reference reports an expected excess at this level, opposed to the present analysis which sees a deficit. However, both deviations are almost equal in size. The analyses are then consistent at this point.

While other searches on the same data samples use an univariate discrimination made up of ∼20 sequential cuts, this search applies only one. The effect of this **one univariate cut** on the E1 samples is shown in Figure 4.3. The upper left graph A shows the visible energy of the events after the event selection treated in the previous section. The measured data is represented by the crosses which also indicate the square root deviation purely due to the bin content. The simulated background is divided into three parts and drawn accumulated. If the upper right graph B is considered, it is clear that a large fraction of the measured events which are not reflected in the simulated samples, are events which have Nip<2. Nip is the number of tracks which have been reconstructed to have a transverse momentum above 1.5 GeV/c and to originate from a region close to the beam crossing point. This region is defined as $z < 1$ mm and $R_\phi < 0.2$ mm. In words this is a 1 mm long cylinder with a 0.2 mm radius centered around the collision point. Requiring at least two tracks of this kind rejects a great part of measured data not contained in the simulations. The events in question are to a large extent assumed to be produced by the so called off-momentum electrons. These are beam particles which have interacted with residual gas particles inside the beam pipe. Due to energy loss via Bremsstrahlung they experience unintended directional changes in the accelerator's magnetic fields. Since these particles do not originate from the beam crossing point, but have left the beam previously, they are heavily rejected by the Nip cut. The visible energy distribution after this cut is shown in the lower left graph C. A large fraction of the low energy excess in the measured sample is now gone. In Figure 4.4 the first three graphs show the same effect for the U1 samples.

The effect of the Nip cut is shown by numbers in Table 4.6. The signals are kept at an efficiency from 81% to 89% in the whole Higgs boson mass range. Apart from the two photon processes which are rejected massively compared to the track selection level shown in Table 4.4, the other background contributions are relatively softly discriminated. **The essential effect is though that large parts of not-simulated background is removed**. This is seen from the difference in the two last rows which now is at ∼ 0.4% and ∼ 1.4% for the E1 and the U1 samples respectively, compared to ∼ 28% and ∼ 40% before the cut.

So far a filter made up of a lax hadronic event selection, a run quality cut

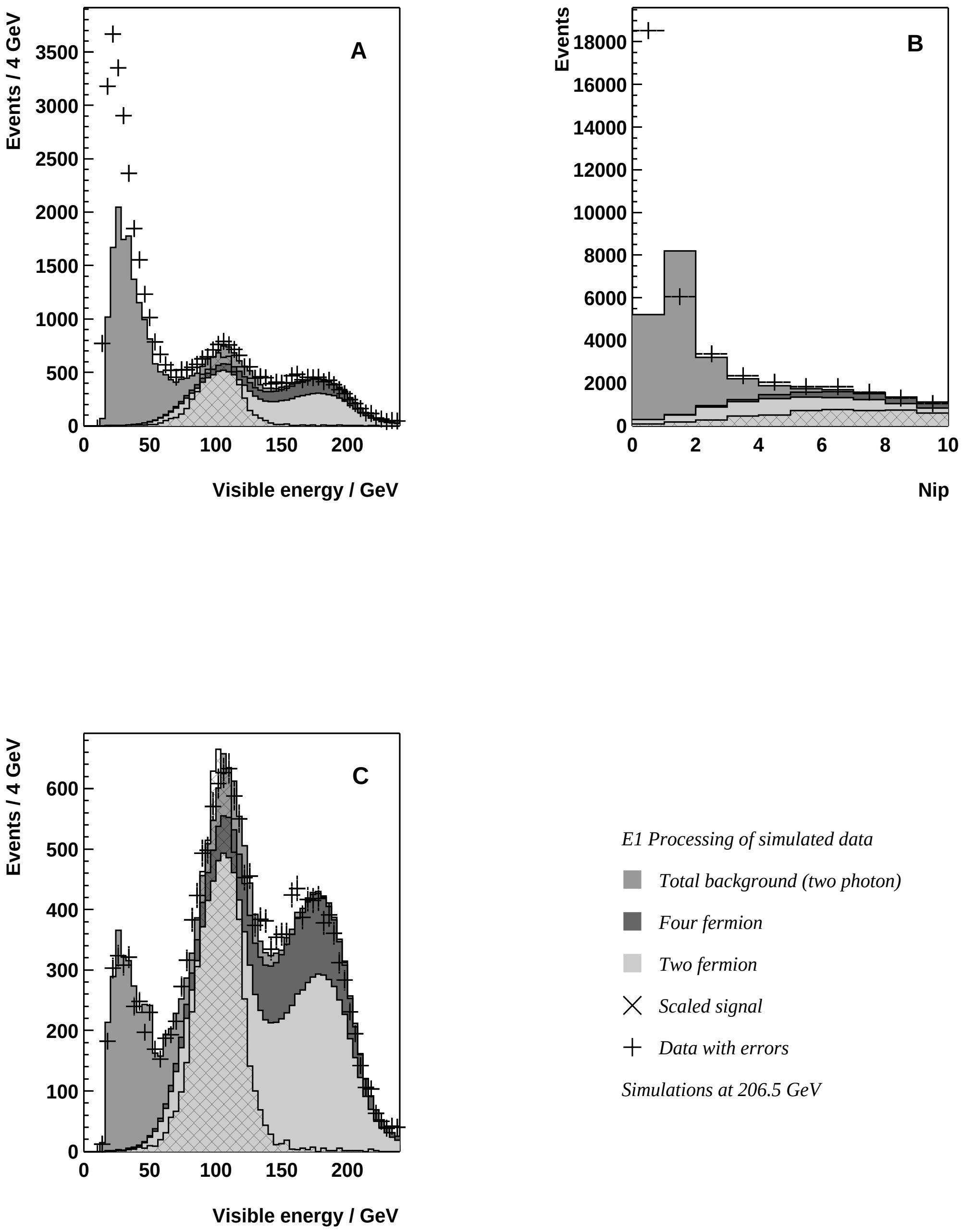


Figure 4.3: Univariate cut on the E1 samples. The visible energy and the the number of tracks from the interaction region, Nip, after the track selection are shown in A and B. The visible energy after the cut $Nip > 1$ is shown in C. The background is divided into three accumulated parts and corresponds to a luminosity of 157.7 pb^{-1}. The signals have a mass of 115 $\mathrm{GeV/c}^2$ and are scaled up to visualize the shapes.

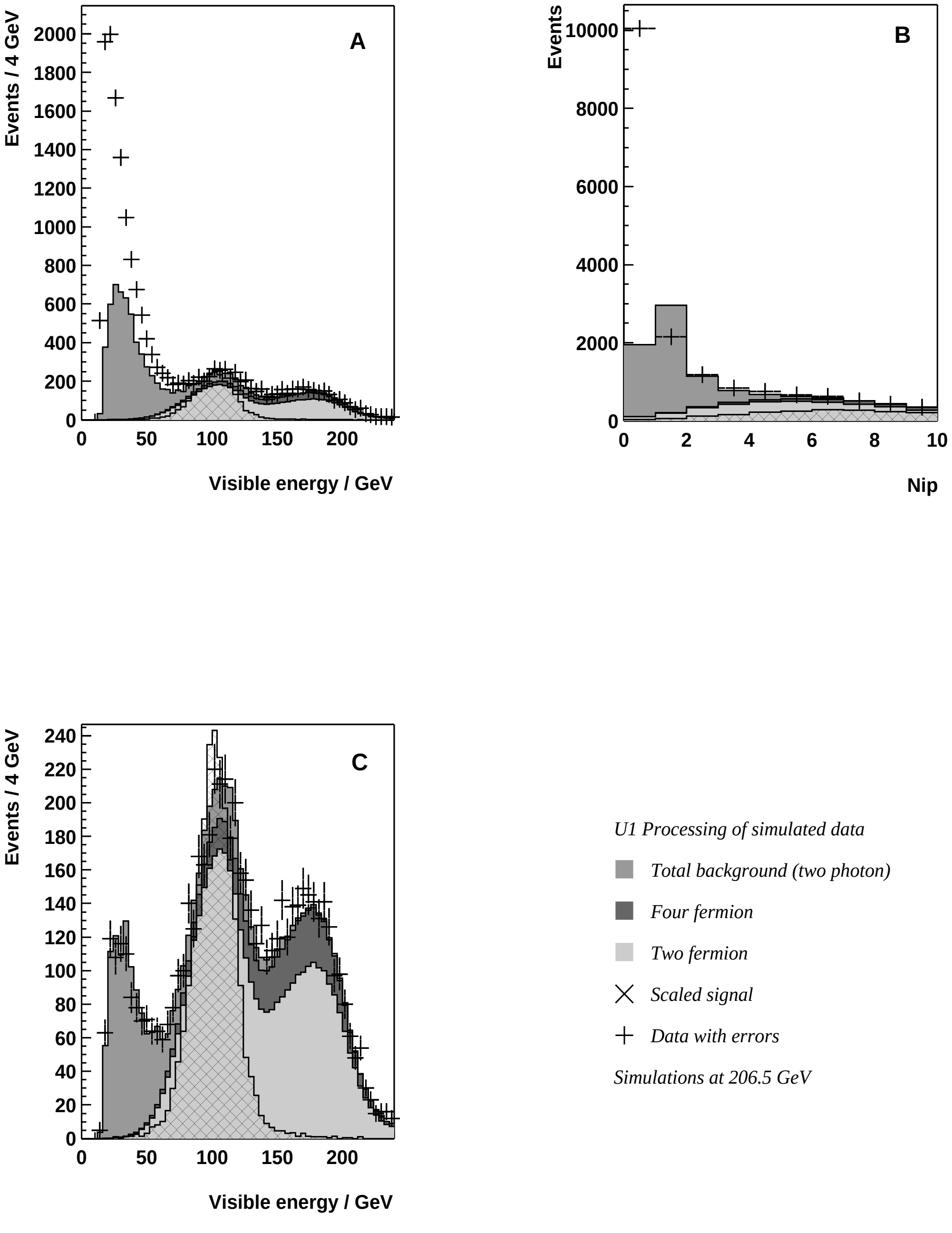


Figure 4.4: Univariate cut on the U1 samples. The histograms correspond to a luminosity of 56.31 pb^{-1}. The signals have a mass of 115 GeV/c^2.

Table 4.6: Normalized numbers of events and effiencies for the E1 and U1 samples after the cut on number of tracks originating from the interaction point $Nip > 1$. For both luminosity and cross sections 1% uncertainty is assumed. One standard deviation binomial uncertainty is used for the cut.

Process	N_{E1}	EFF_{E1}	Process	N_{U1}	EFF_{U1}
$m_H = 40$	26.41 ± 0.41	0.82	$m_H = 40$	9.33 ± 0.14	0.81
$m_H = 50$	24.32 ± 0.37	0.85	$m_H = 50$	8.54 ± 0.13	0.83
$m_H = 60$	21.55 ± 0.33	0.86	$m_H = 60$	7.64 ± 0.11	0.85
$m_H = 70$	20.77 ± 0.31	0.87	$m_H = 70$	7.32 ± 0.11	0.86
$m_H = 80$	17.60 ± 0.26	0.89	$m_H = 80$	6.101 ± 0.093	0.85
$m_H = 85$	15.71 ± 0.23	0.89	$m_H = 85$	5.481 ± 0.083	0.86
$m_H = 90$	13.73 ± 0.20	0.88	$m_H = 90$	4.901 ± 0.074	0.88
$m_H = 95$	11.98 ± 0.18	0.88	$m_H = 95$	4.236 ± 0.064	0.87
$m_H = 100$	9.87 ± 0.14	0.87	$m_H = 100$	3.557 ± 0.053	0.88
$m_H = 105$	7.78 ± 0.11	0.88	$m_H = 105$	2.747 ± 0.041	0.87
$m_H = 110$	5.292 ± 0.080	0.88	$m_H = 110$	1.873 ± 0.028	0.87
$m_H = 114$	2.751 ± 0.041	0.87	$m_H = 114$	0.985 ± 0.014	0.87
$m_H = 115$	2.044 ± 0.030	0.87	$m_H = 115$	0.734 ± 0.011	0.88
$m_H = 116$	1.514 ± 0.022	0.88	$m_H = 116$	0.5359 ± 0.0081	0.87
$m_H = 120$	0.793 ± 0.011	0.89	$m_H = 120$	0.2792 ± 0.0042	0.87
2FQQ	11073 ± 157	0.89	2FQQ	3922 ± 56	0.87
2FEE	3.52 ± 0.53	< 0.01	2FEE	1.31 ± 0.27	< 0.01
2FMM	0.096 ± 0.034	< 0.01	2FMM	0.0127 ± 0.0074	< 0.01
2FTT	23.71 ± 0.67	0.02	2FTT	7.82 ± 0.30	0.02
4FCC	2792 ± 40	0.88	4FCC	670 ± 9	0.59
4FNC	292.2 ± 4.2	0.23	4FNC	136 ± 2	0.30
GGLL	664 ± 10	0.07	GGLL	233 ± 3.4	0.07
GGQQ	3313 ± 102	< 0.01	GGQQ	1105 ± 32	< 0.01
SUM BKG	18161 ± 192	-	SUM BKG	6076 ± 65	-
DATA	18089	-	DATA	6161	-

and one univariate cut has been applied. The run quality cut is reflected in the effective luminosities. The requirement on the event origin was used to achieve roughly an one standard deviation agreement between simulated and measured data. As described in Chapter 2, there should now exist a hope for a multivariate discriminator trained on simulated samples to act as expected on the measured sample. This minimal univariate discrimination constitutes **an approach which differs in at least two ways from other analyses on the same data**. First, this minimal preselection is to a high degree hypotheses independent. The advantage lies in the standardized hypothesis optimization. From now on the statistical procedure takes care of the discrimination and the task of optimization becomes a task of computing. A traditional way of enhancing the signal by many sequential cuts, becomes signal, or hypothesis, specific and reduces the generic potential in a statistical treatment of discrimination. Second, the feature space is larger in order to keep the signal efficiency high for various hypotheses. It follows that the procedure may be more sensitive to discrepancies in the agreement between measured and simulated data. However, this must be checked along the way.

4.5 Variables

After the univariate cut chosen to remove measured events not described by the simulations, the next step is to select the variables which are to be used in the multivariate discrimination. For data coming from complex detectors, in this case from the DELPHI detector [23], hundreds of event variables are constructable. Hence, some physical intuition has to be applied to select a subset of features. In this analysis a subset of twelve variables were chosen based on their apparent discriminating power, where apparent is simply significant visible differences between the simulated signal and background distributions. The distributions of these remaining features are shown in Figures 4.5 and 4.7 for the E1 samples and in Figures 4.6 and 4.8 for the U1 samples. The distributions contain all events which survived the filter examined in the last section. They are normalized to the respective luminosities and the measured distribution is also shown to establish confidence in the simulated samples.

The **missing mass** of an event is defined as $\sqrt{E_{mis}^2 - P_{mis}^2}$ where E_{mis} is the missing energy and P_{mis} is the absolute value of the missing momentum vector. Its distributions for the E1 samples and the U1 samples are plotted in the graphs marked A of Figure 4.5 and 4.6 respectively. Large fractions of all three background classes may be discriminated with this observable, especially the four fermion and the two photon background.

The **acoplanarity** of two jets is defined as $\pi - \arccos(\frac{\vec{p}_1^T \cdot \vec{p}_2^T}{|\vec{p}_1^T||\vec{p}_2^T|})$, where $\vec{p}_{1,2}^T$ are the transverse momenta of the jets, i.e. the momenta projected onto the plane orthogonal to the beam axis. It is shown in the graphs marked B of

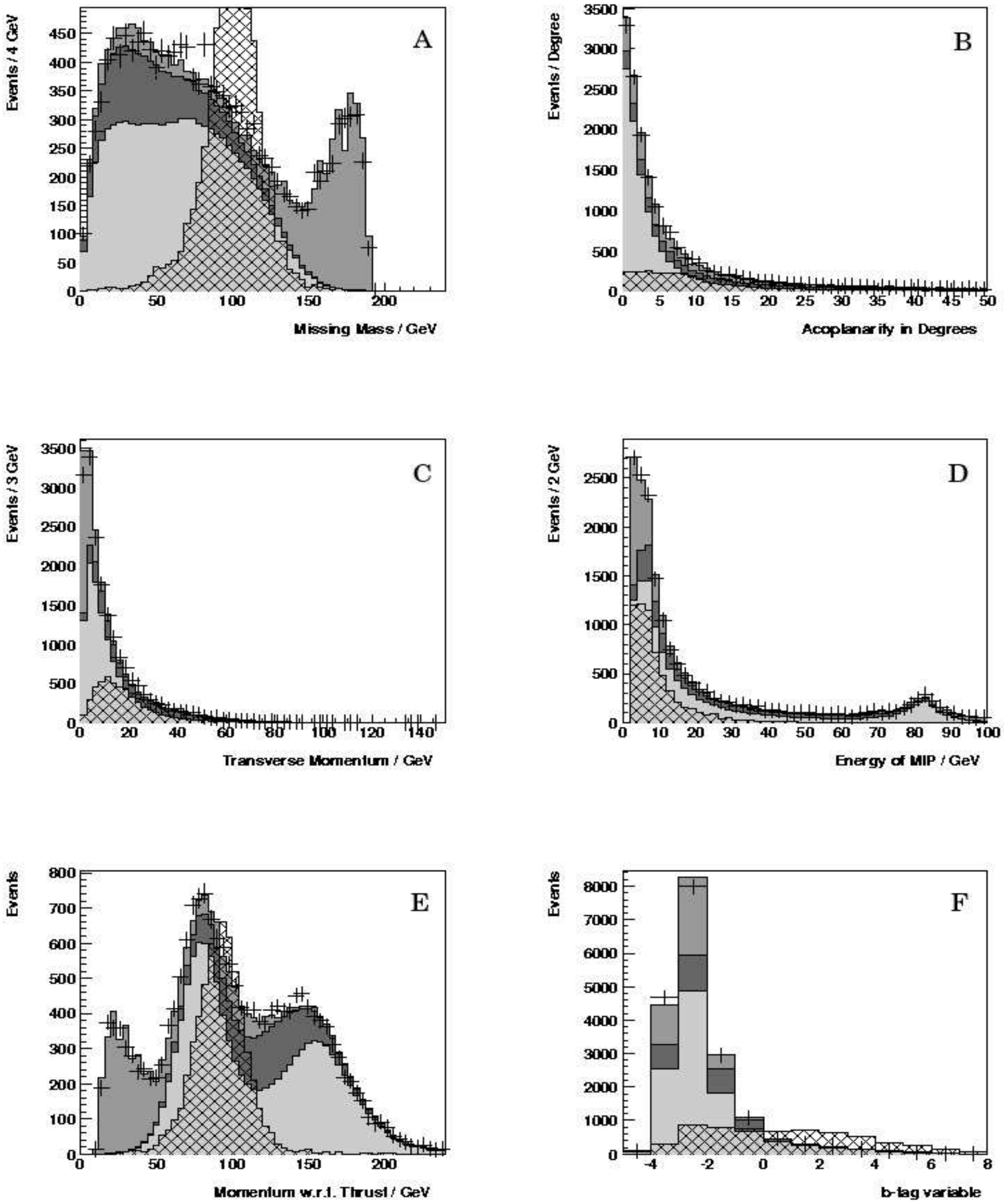


Figure 4.5: Simulated and measured distributions of E1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 157.7 pb^{-1}. The scaled signal has a mass of 115 GeV/c^2. Gray scales as in Figure 4.3.

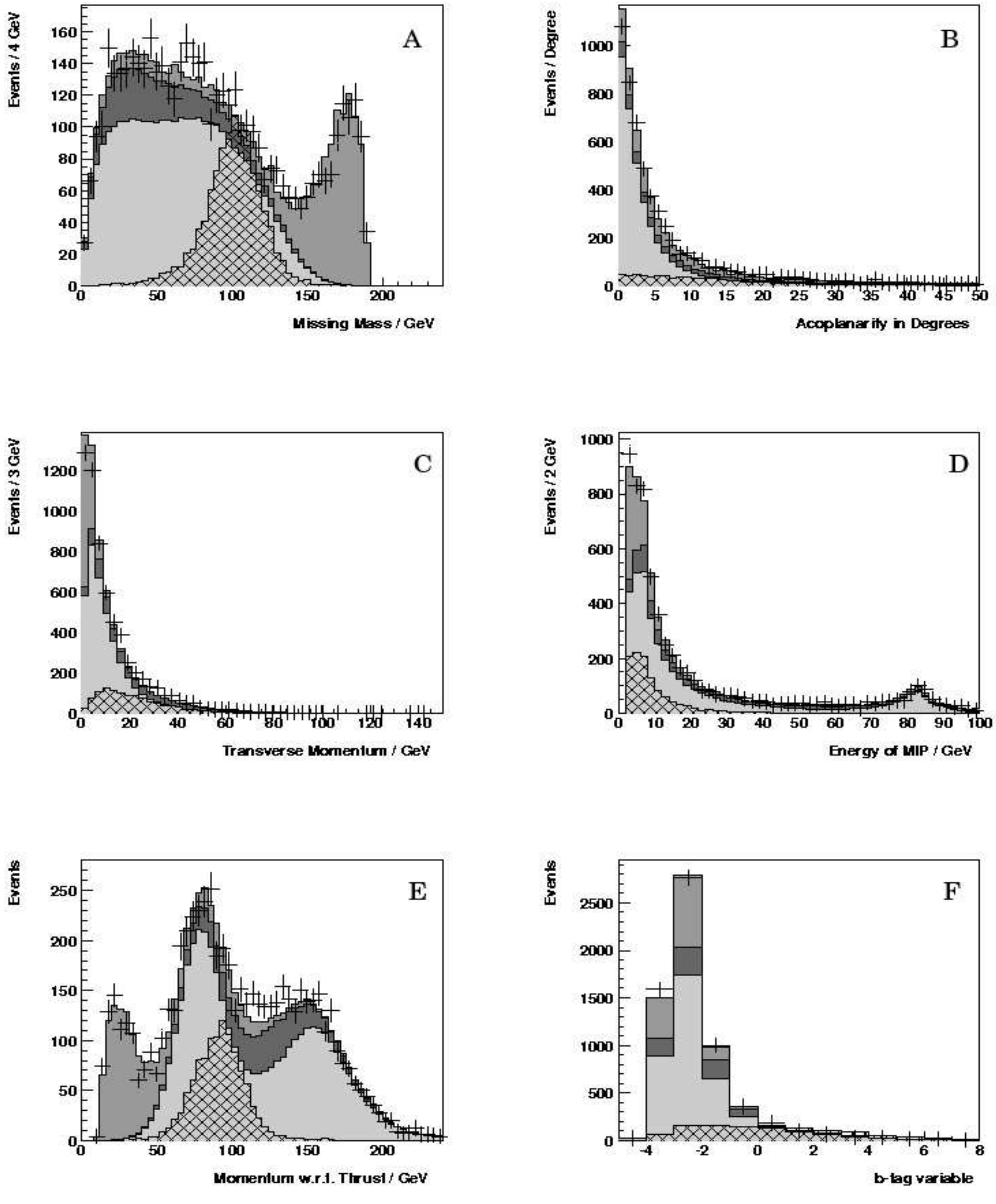


Figure 4.6: Simulated and measured distributions of U1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 56.31 pb^{-1}. The scaled signal has a mass of 115 GeV/c^2. Gray scales as in Figure 4.3.

Figure 4.5 and 4.6 respectively. While the background is peaked at low angles, i.e. the background events are to a high degree events with jets in opposite directions, the signal distribution is relatively flat.

The **transverse momentum** of an event is defined as the visible momentum projected down on the plane orthogonal to the beam axis. Its distributions for the E1 samples and the U1 samples are plotted in he graphs marked C of Figure 4.5 and 4.6 respectively. Large fractions of all three background classes may be discriminated with this observable, especially the two fermion and the two photon background.

The **energy of the most isolated particle**, MIP, in an event is defined via the definition of isolation energy. Isolation energy of a particle i with momentum $\vec{p_i}$ is defined as the energy deposit from other particles j with momentum $\vec{p_j}$ with more than 2 GeV if $\angle(\vec{p_i}, \vec{p_j})$ is between 5 and 60 degrees, or with more than 5 GeV if $\angle(\vec{p_i}, \vec{p_j})$ is between 5 and 25 degrees. The energy of the most isolated particle is then the energy of the particle with the smallest isolation energy. The distributions are shown in the D graphs of Figure 4.5 and 4.6 respectively.

The **momentum w.r.t. thrust** of an event is the sum of all particles' momentum projection onto the thrust axis, using the absolute value. The thrust is defined as in Reference [46]

$$T = max \frac{\sum |\vec{n} \cdot \vec{p_i}|}{\sum |\vec{p_i}|}$$

where $\vec{p_i}$ is the momentum of the particles and the thrust axis $\vec{n}$ is chosen so that T is maximized. The distributions are seen in the E graphs of Figure 4.5 and 4.6.

The **b-tag** of an event is a tag in the two jet configuration, i.e. a variable calculated for each jet when the event is forced into two jets. Specific features of decaying bottom quark hadrons like large mass, long lifetime, i.e. secondary vertex, etc. are used to construct this variable. Detailed descriptions are described in the References [49] and [50]. The distributions are shown in the graphs marked F in Figure 4.7 and Figure 4.8. While the background peaks at values around -2, the signal is to a higher degree flat.

The **minimal** $\angle(jet, \vec{p})$ is the minimal angle between the transverse momentum and the jets. The signal shape of this distribution is convex (from below) while the background shape is concave. The distributions are plotted in the A graphs of Figure 4.7 and 4.8 respectively.

In the B graphs of Figure 4.7 and Figure 4.8 the **transverse energy** of the events are shown. Further the **maximal transverse momentum of tracks**, the **total number of tracks**, the **number of charged tracks** and the **number of leptons** are used to train the discriminator. The distributions are shown in the graphs C, D, E and F of Figure 4.7 and Figure 4.8.

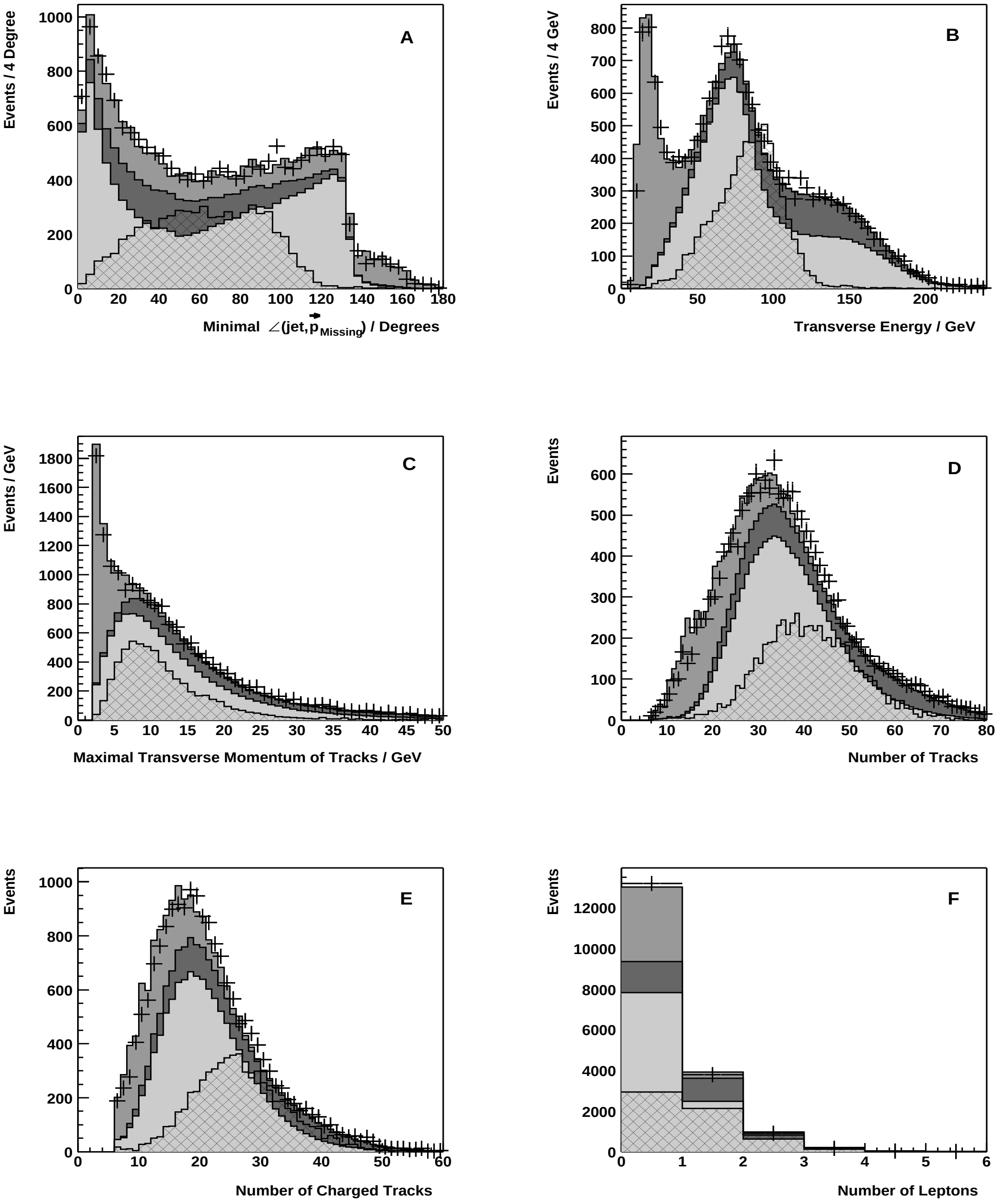


Figure 4.7: Simulated and measured distributions of E1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 157.7 pb^{-1}. The scaled signal has a mass of 115 GeV/c^2. Gray scales as in Figure 4.3.

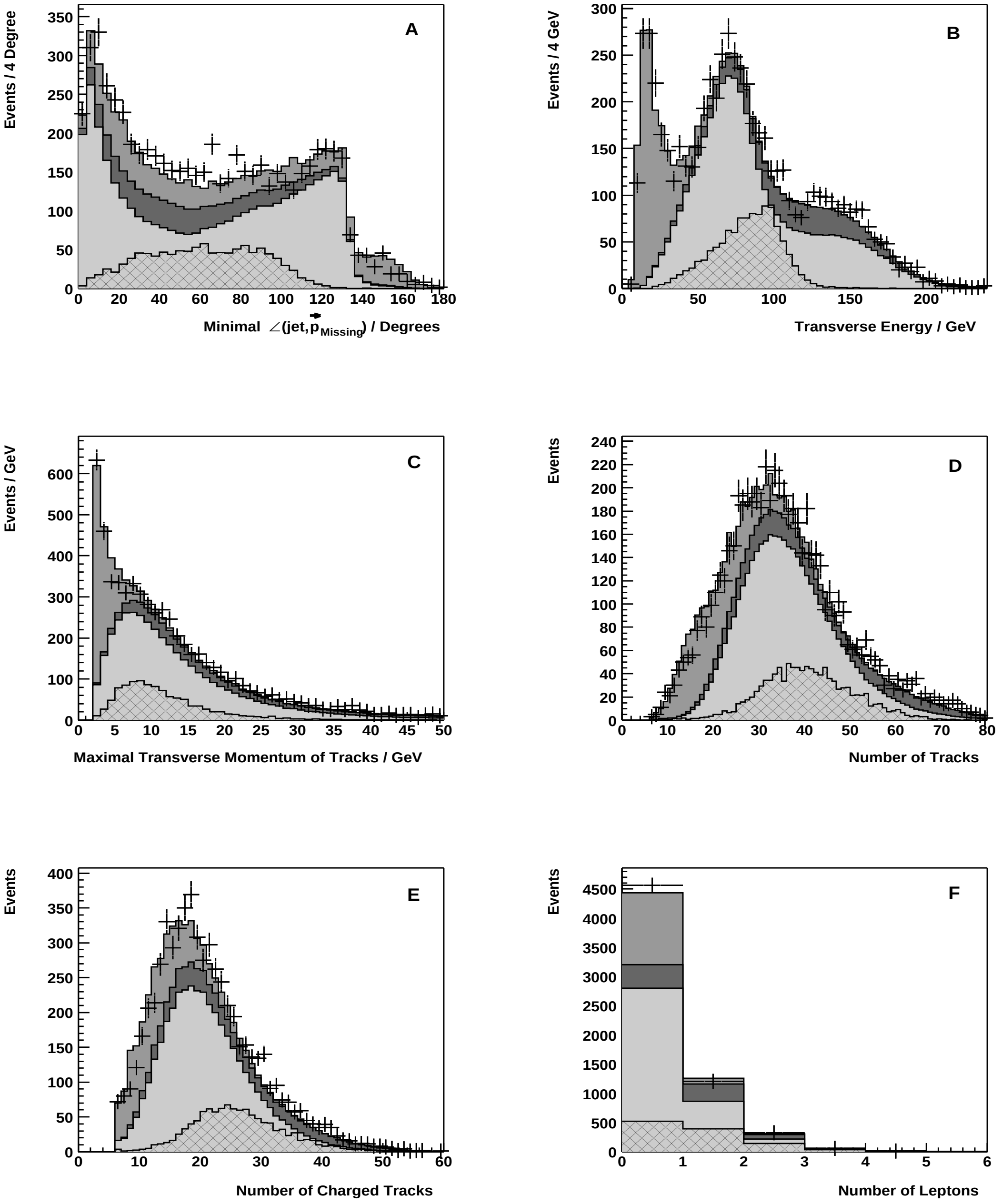


Figure 4.8: Simulated and measured distributions of U1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 56.31 pb^{-1}. The scaled signal has a mass of 115 $\mathrm{GeV/c}^2$. Gray scales as in Figure 4.3.

The selection of observables is not ensured to be optimal. It is a result of a careful study of many different observables and of testing the discriminating power of many of these. An optimal selection can only be ensured by testing all possible combinations. As discussed in 2.1.6 this is a formidable computing task which was not within in the time scale of this work. However, having chosen the observables which span the feature space, the general polynomial discrimination is now straightforward.

4.6 Discrimination and Hypothesis Test

Given the set of event features, i.e. the observables chosen in the previous section, the next step in the analysis is to discriminate the background as efficiently as possible. The general polynomial discriminator as described in Subsection 2.1.3 is applied. Since the most efficient discrimination is expected to be achieved with the highest possible order of the polynomial, the order is increased until overestimation occurs. Due to the complex distributions of the samples an iterative discrimination is applied, again with the highest possible polynomial order. The discrimination in this iterative step may focus on another part in the feature space, rejecting regions not grasped by the first step. The surviving events constitute the input for the hypothesis testing.[6] The search results are reported in this section.

In the following subsection the highest possible order of the polynomials is determined. Then, as pointed out in Section 4.4, the discrimination is pure computing. The subsequent subsections present searches optimized at four different Higgs Boson masses $m_H = 80$ GeV/c^2, $m_H = 90$ GeV/c^2, $m_H = 100$ GeV/c^2, and $m_H = 115$ GeV/c^2. The graphs with background events versus signal efficiencies, distributions of the observables, the candidates with highest weights and the input distributions for the limit calculations are shown. For the optimizations at masses $m_H = 90$ GeV/c^2 and $m_H = 115$ GeV/c^2 independent search results were available [5]. The discrimination power of the method is therefore compared at these optimization points.

At all optimization masses a discrimination with two iterations has been applied. In all cases the same order was used for the polynomials. The expected CL_s were studied at each optimization for the E1 and the U1 samples respectively. The cut on the second iteration polynomial which resulted in the

---

[6] The discrimination procedure was implemented in C/C++. The measured and simulated samples were converted from HBOOK files to ROOT files and read and processed with ROOT class libraries. The polynomial coefficients $\vec{a} = V^{-1}\Delta\vec{\mu}$ were not determined by matrix inversion which is a computing intensive task, but by solving $V\vec{a} = \Delta\vec{\mu}$ via Cholesky factorization. A performance test showed that the use of the object oriented library CLHEP was even faster than the FORTRAN KERNLIB library in this case. This small surprise motivated the choice of language.

Table 4.7: Efficiencies for different Polynomials. The optimization was done at $m_H = 90$ GeV/c^2 and the efficiencies were calculated after the second iteration of discrimination.

P	Order	ϵ_{S1}	ϵ_{S2}	$2\Delta\epsilon_{S1}$	ϵ_B
E1	1.	0.55	0.55	0.02	0.00090
	2.	0.55	0.55	0.02	0.00060
	3.	0.55	0.54	0.02	0.00047
	4.	0.55	0.47	0.02	0.00030
U1	1.	0.55	0.55	0.02	0.00135
	2.	0.55	0.55	0.02	0.00091
	3.	0.55	0.54	0.02	0.00072
	4.	0.55	0.46	0.02	0.00046

lowest expected CL_s was applied. The resulting samples were used as inputs for the hypothesis tests.[7]

According to the Equation 2.21 the quantities occurring in the test statistic Q are inputs for the hypothesis tests.[8] For each channel E1 and U1 the expected numbers of signal and background events are needed. This means as usual the cross sections, the luminosities, the total numbers of generated events and the numbers of events which have survived the discrimination.[9] Thus these numbers are reported for each optimization. The probability density distributions for the background and the signal hypotheses are also required in order to take advantage of the weighted counting. Since the statistics in these distributions were limited, they had to be smooted. How this was done is treated in Appendix A.

4.6.1 Order of the Polynomial Discriminant

In Table 4.7 the signal and background efficiencies, $\epsilon = N/N_{org}$, where N is the number of events at the chosen cut level and N_{org} the number of originally

---

[7] Optimizing each channel separately, as done here, is not ensured to be optimal, but considered as a reasonable approximation.

[8] The calculations in this chapter were all done using a FORTRAN version of the ALRMC software adjusted to the purposes of this analysis [19]. This software has proved itself in the LEP Higgs Working Group calculations of the combined LEP exclusion limits on the Higgs bosons (see e.g. Appendix B).

[9] Expected number of events $N_{exp} = \mathcal{L}\sigma N_{cut}/N_{gen}$ where $\mathcal{L}$ is the integrated luminosity, σ the cross section, N_{cut} the number of events after the cut and N_{gen} the number of generated events.

generated events, for discriminants up to the fourth order after the second iteration are summarized. The cut on the first iteration is done at a 68% signal efficiency. The optimization is done for a Higgs Boson mass $m_H = 90$ GeV/c^2. The first column indicates the processing, the second the iteration and the third the order of the discriminant. The fourth and the fifth columns contain the signal efficiencies for the training ϵ_{S1} and the test ϵ_{S2} sample. ϵ_{S1} is arbitrarily chosen to be ~55%. The sixth column shows the two standard deviations uncertainty $2\Delta\epsilon_{S1}$ on the ϵ_{S1} number. The remaining column holds the efficiency for the total background and describes the discriminating power. As expected it increases with the order. However, for the fourth order discriminants the efficiencies for the test samples are far off the two standard deviations. Due to these very unlikely results it is concluded that the training samples are not big enough to estimate such a large number of parameters and that the signal efficiency is overestimated. Thus the third is taken to be the highest possible order.

4.6.2 Search Optimized at $m_H = 80$ GeV/c^2

Third order polynomials were estimated with a signal mass equal to 80 GeV/c^2. A first iteration cut was performed on the polynomials achieved with the samples resulting from the minimal univariate discrimination presented in Section 4.4. The efficiency for the 80 GeV/c^2 signal was chosen to be 70% for this cut.

The resulting samples after the cut on the first iteration polynomials were used to train second iteration polynomials. The discriminating power of these polynomials is shown in Figure 4.9 and Figure 4.10 for the E1 samples and the U1 samples respectively. Both figures demonstrate that the discriminating power increases with the order of the polynomials. Due to the overtraining studies reported in Subsection 4.6.1, the third order polynomials were used.

The remaining distributions of the observables which were used to estimate the polynomials are shown in Figures 4.11, 4.12, 4.13 and 4.14 for a 70% efficiency for the 80 GeV/c^2 signal. Figure 4.15 shows the distributions of the Higgs boson mass estimator, Constrained Mass, and the Discriminator. The upper two graphs contain the E1 samples, the lower two the U1 samples.

Hypothesis Test and Limits

The study of different cuts on the second iteration polynomials showed that the expected CL_s reached their minima for signal efficiencies ($m_H = 80$ GeV/c^2) of 64% and 46% for the E1 and the U1 samples respectively. With the nomenclature which was established in Equation 2.21, this corresponded to $s_{E1} = 9.21$, $b_{E1} = 88.4$, $s_{U1} = 1.75$ and $b_{U1} = 2.7$. At these efficiencies the number of observed candidates were 87 and 2 respectively.

The candidates with highest $\ln Q$ are listed in Table 4.8. They are ordered with respect to the $\ln Q$ value of the events. While the fifth column contains the cumulative event contributions to the $\ln Q$, the sixth column shows the relative contributions. The event with the highest $\ln Q$ contributes with 12.7%. Among the twenty events listed in Table 4.8, the second and the fifth are events from the U1 sample.

The results for the search optimized at $m_H = 80$ GeV/c^2 are summarized in Figure 4.16. The upper left graph shows the CL_s for for Higgs mass hypotheses from 60 GeV/c^2 to 120 GeV/c^2. The observed results are within one standard deviation of the expected ones. The CL_s approaches one with the decreasing signal cross section.[10] Where the expected median and the observed results intersect the horizontal line at $CL_s = 0.05$, the 95% exclusion limits are found.

$$m_H^{exp} \geq 89.8 \text{ GeV/c}^2$$

$$m_H^{obs} \geq 94.1 \text{ GeV/c}^2$$

The $1 - CL_b$ graph in Figure 4.16 shows that the observed result lies well within one standard deviation from the expected result, the horizontal line at $1 - CL_b = 0.5$. The dotted line indicates the expected result for signal plus background experiment (median). The dotted signal plus background expectation is even at 60 GeV/c^2 far from the discovery convention at 5.7×10^{-7} (five standard deviations). Again the cross section becomes vanishingly small at high masses where there is practically no difference between the two hypotheses.

The $-2 \ln Q_b$ and $-2 \ln Q_{s+b}$ graphs in Figure 4.16 are an alternative way of visualize the confirmation the background only hypothesis. To the left the $-2 \ln Q_b$ shows that the observation is within one standard deviation over the whole mass range. The signal plus background hypothesis is shown in the right graph. The observation is more than two standard deviations off the dashed expectation up to around 95 GeV/c^2. For higher masses the two hypotheses become increasingly similar due to the decreasing signal cross section.

---

[10] Instructive graphs for the understanding of the CL_s and the $1 - CL_b$ are found in Figure 1 of Appendix B.

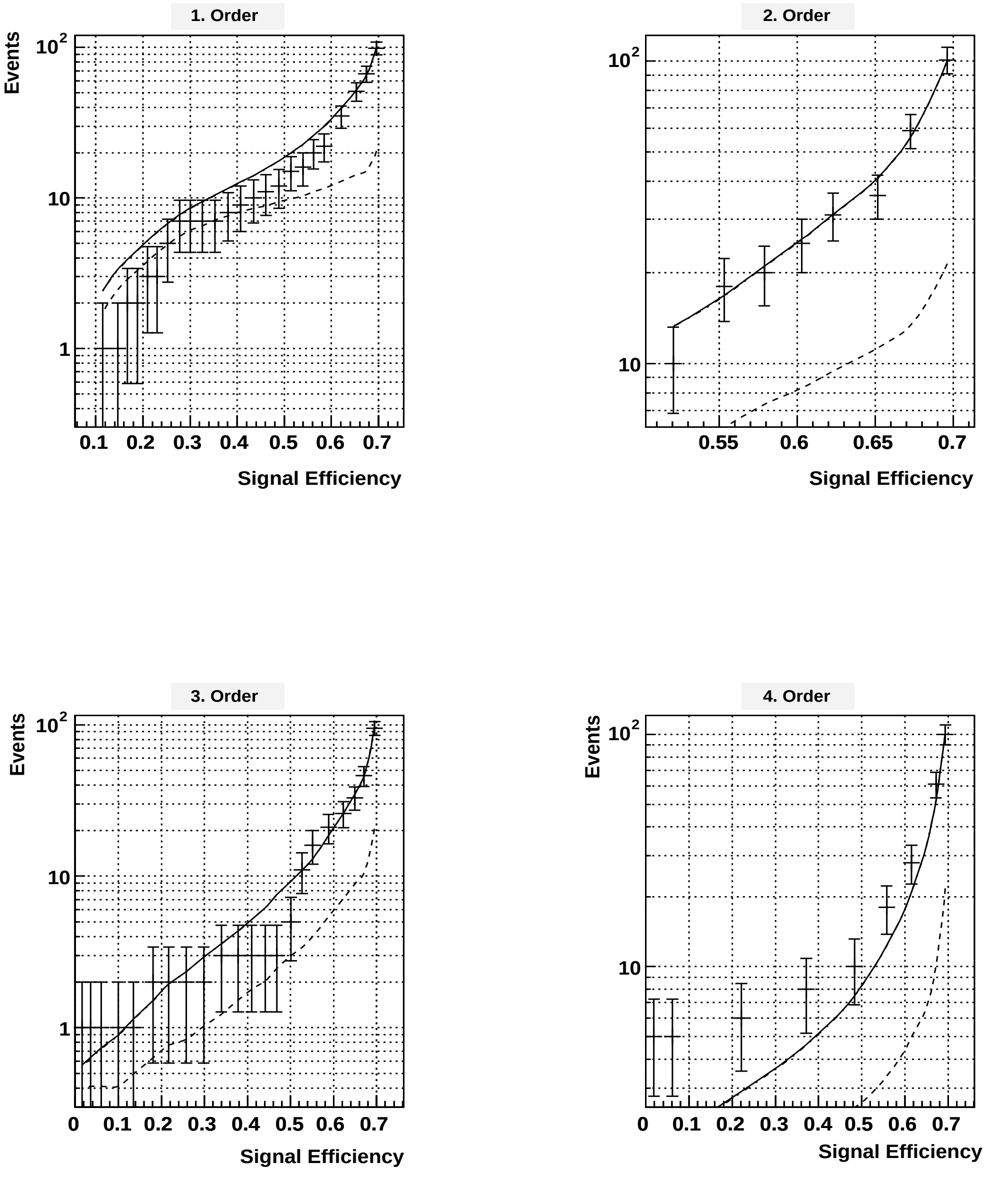


Figure 4.9: E1 Background events versus efficiency for a 80 GeV/c^2 mass signal. The measured data is shown with error bars. The dashed lines are expected background contributions from the two fermion processes. The solid lines represent the expected total background.

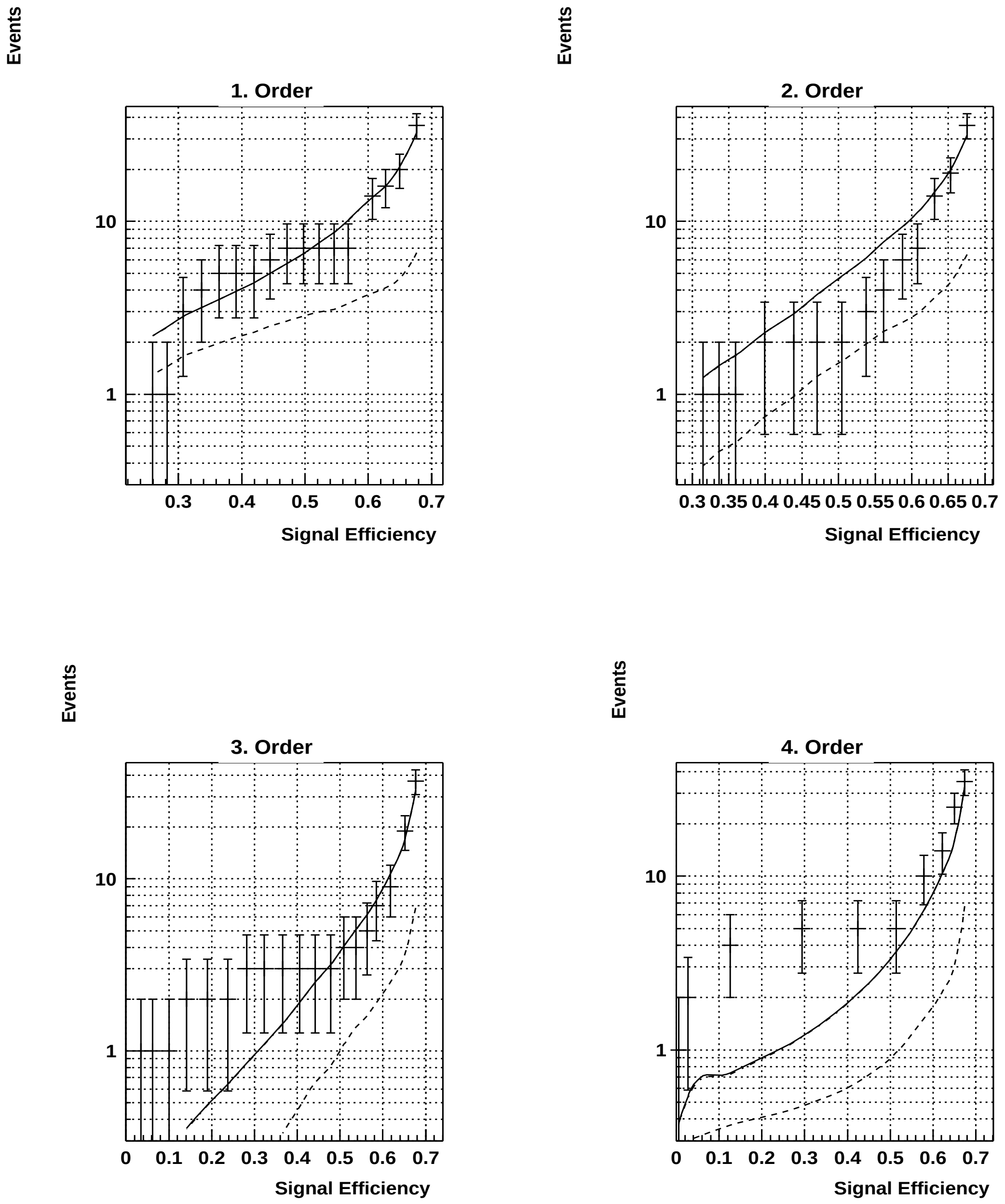


Figure 4.10: U1 Background events versus efficiency for a 80 GeV/c^2 mass signal. The measured data is shown with error bars. The dashed lines are expected background contributions from the two fermion processes. The solid lines represent the expected total background.

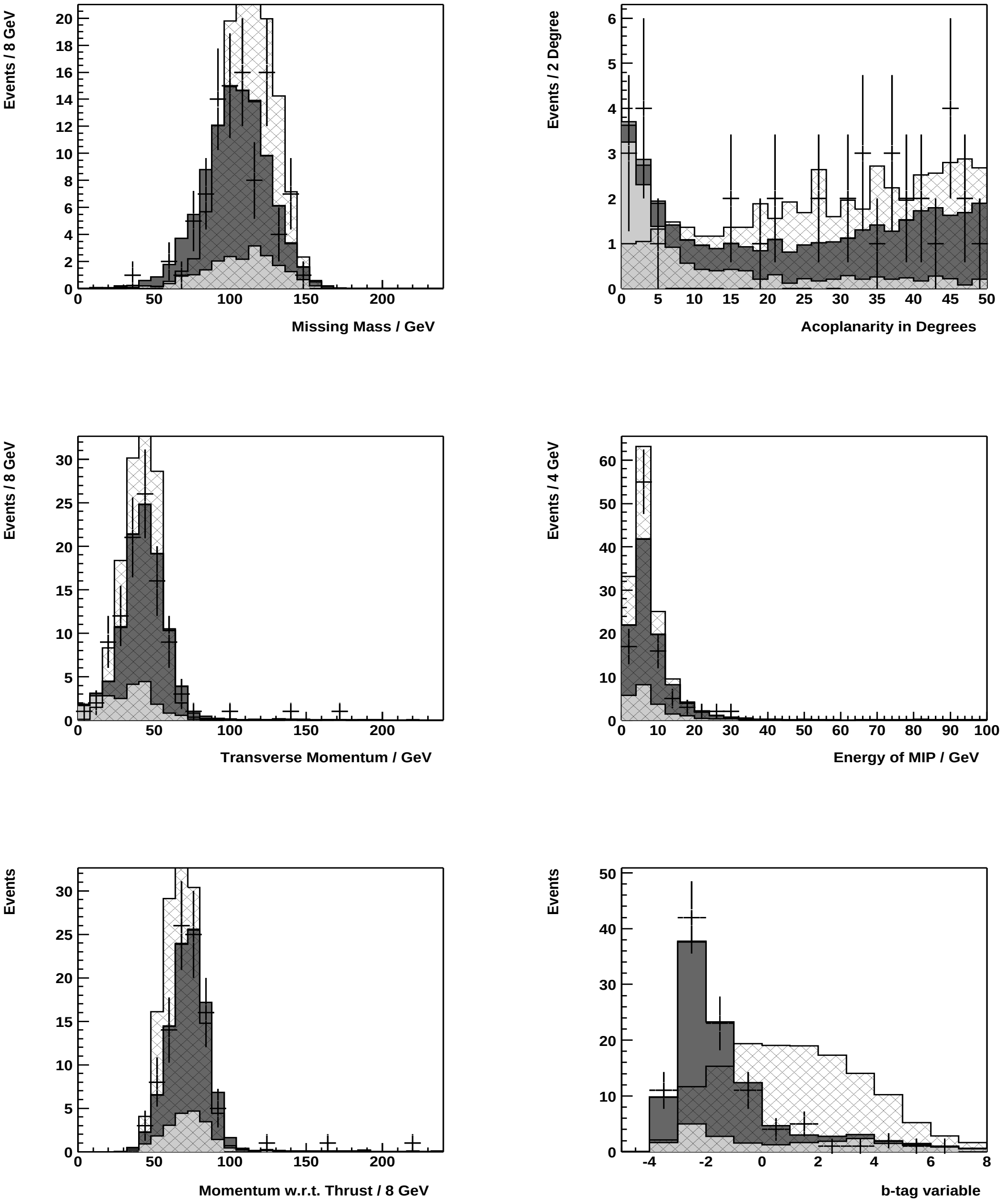


Figure 4.11: Simulated and measured distributions of E1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 157.7 pb^{-1}. The scaled signal has a mass of 80 GeV/c^2. Gray scales as in Figure 4.3.

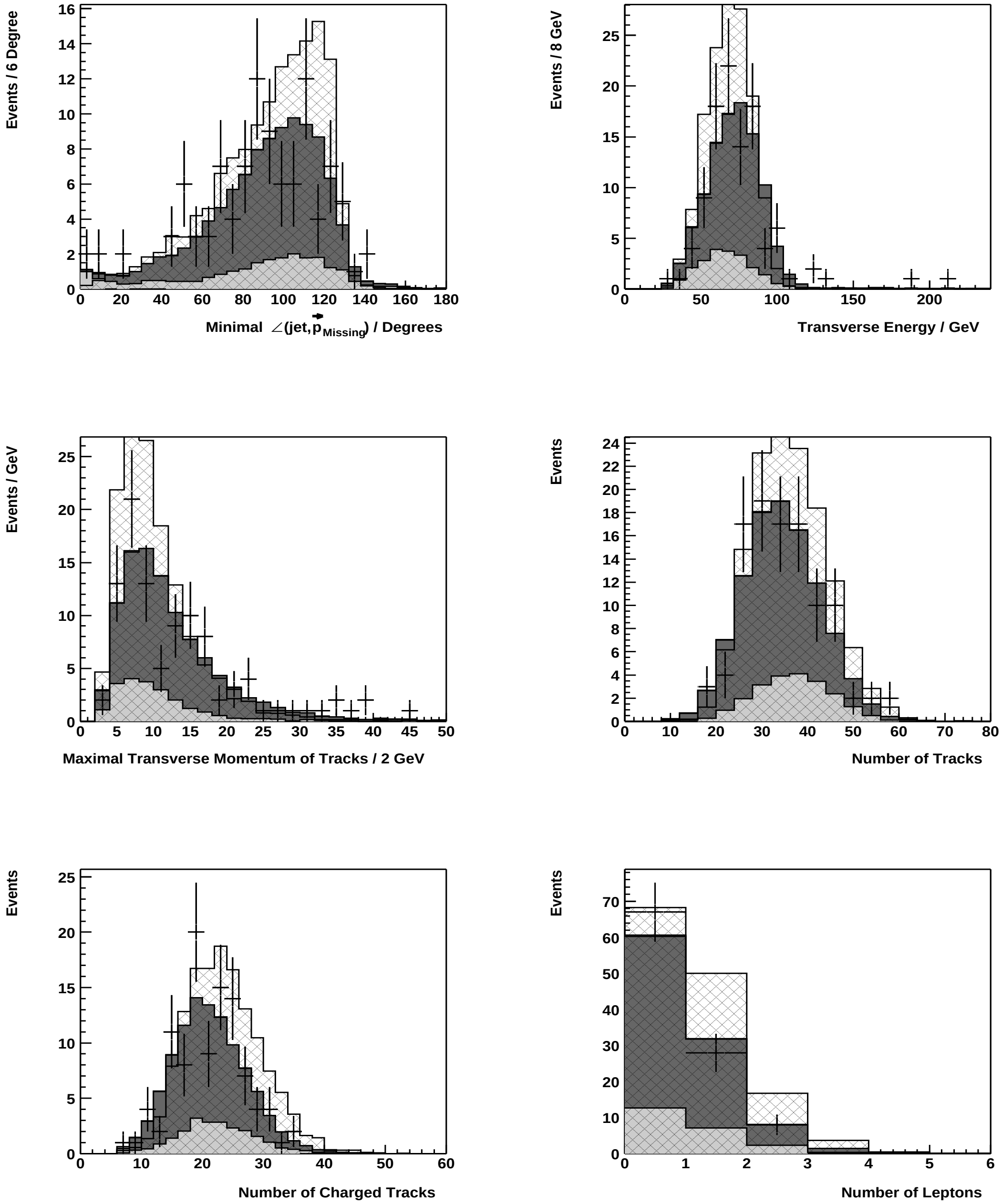


Figure 4.12: Simulated and measured distributions of E1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 157.7 pb^{-1}. The scaled signal has a mass of 80 GeV/c^2. Gray scales as in Figure 4.3.

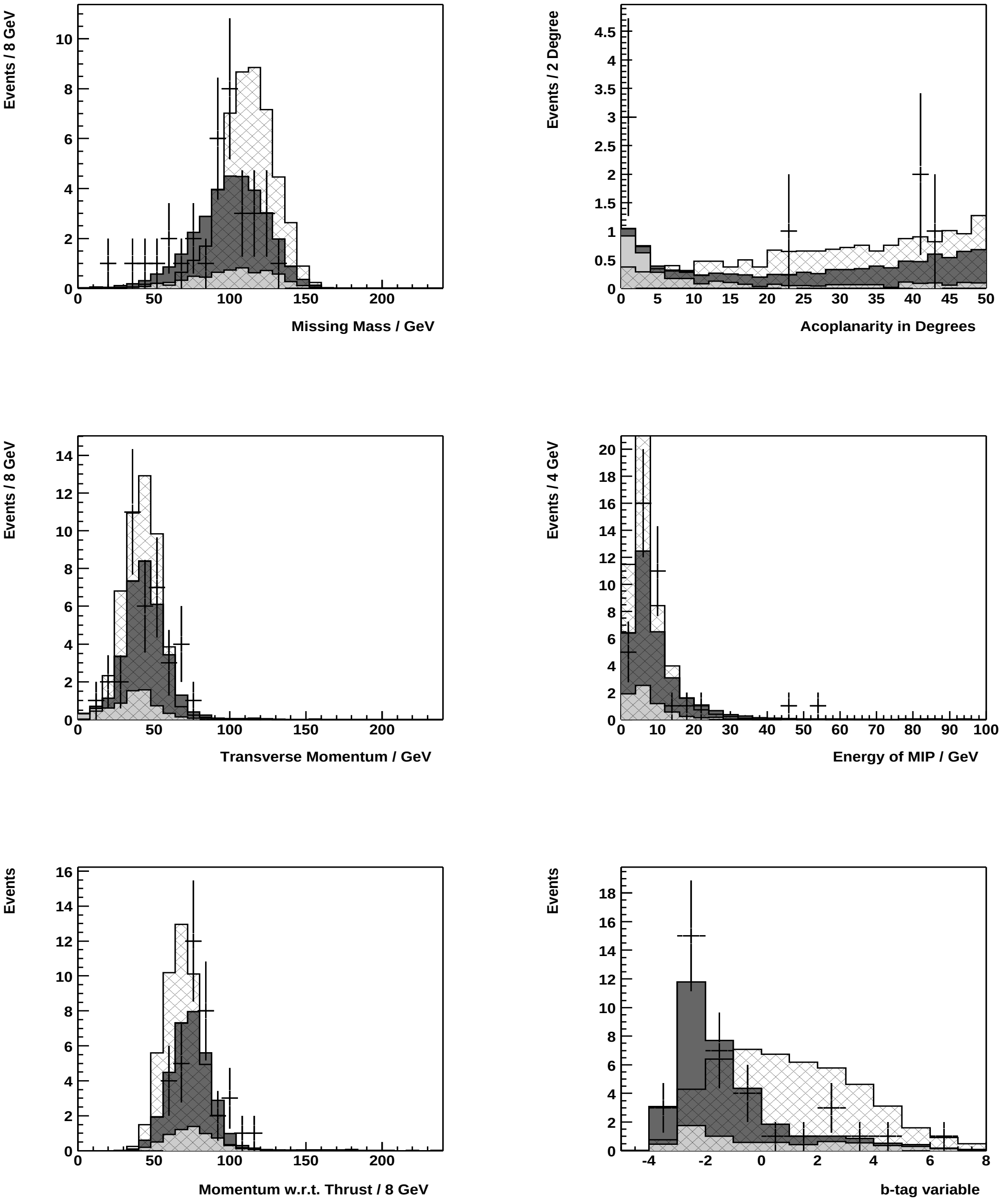


Figure 4.13: Simulated and measured distributions of U1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 56.31 pb^{-1}. The scaled signal has a mass of 80 GeV/c^2. Gray scales as in Figure 4.3.

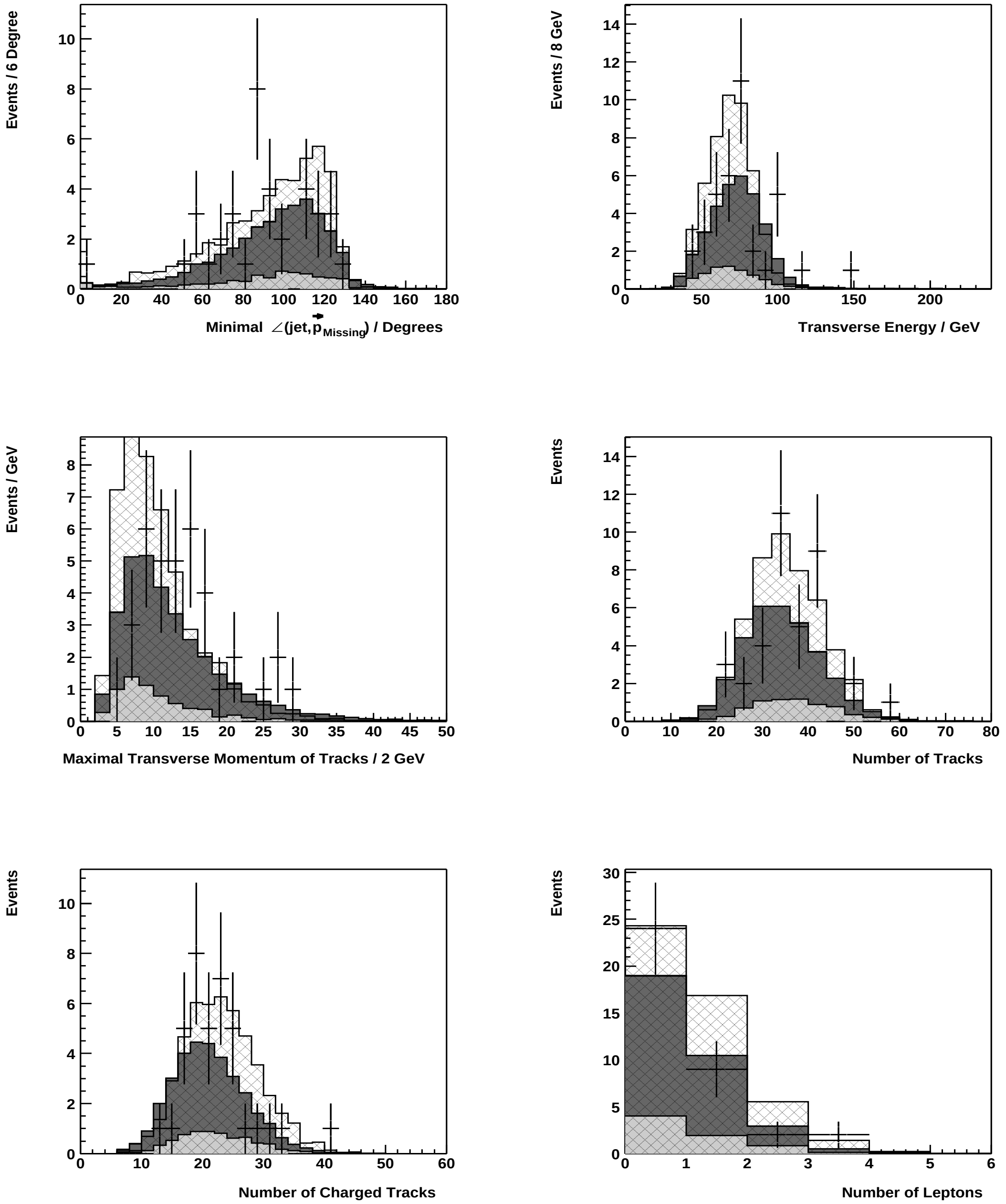


Figure 4.14: Simulated and measured distributions of U1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 56.31 pb^{-1}. The scaled signal has a mass of 80 $\mathrm{GeV/c^2}$. Gray scales as in Figure 4.3.

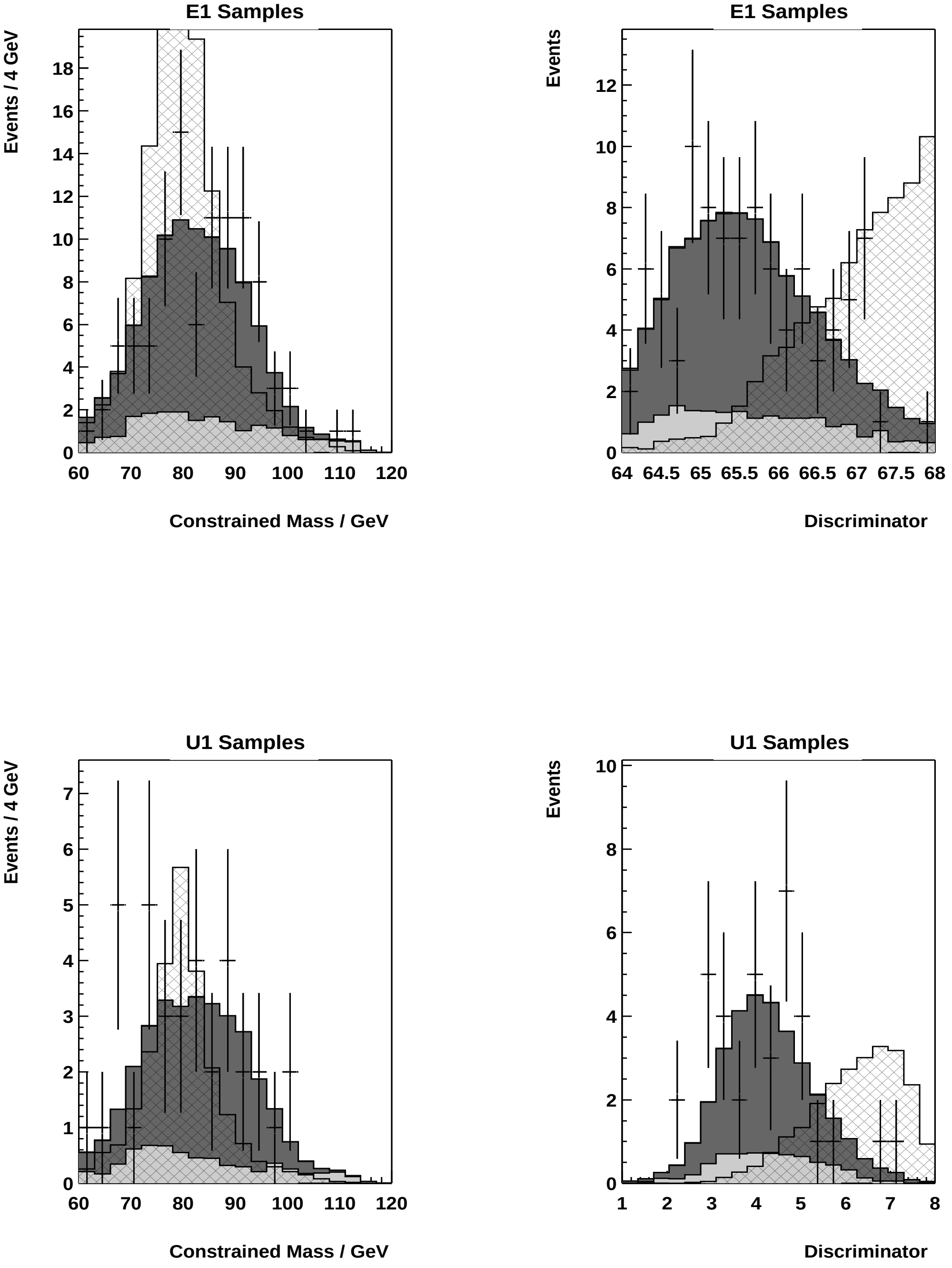


Figure 4.15: Simulated and measured distributions. The background is divided into three accumulated parts. The scaled signals have a mass of 80 GeV/c^2. Gray scales as in Figure 4.3.

Table 4.8: Candidates found in a search optimized at $m_H = 80$ GeV/c^2. $Q = 1 + sS(\vec{x})/bB(\vec{x})$ is the test statistic as defined in Equation 2.21. The candidates are sorted with respect to $\ln Q_n$. The fifth and the sixth columns contain the cumulative ($\sum_{i=1}^{n} \ln Q_i / \sum_{All} \ln Q$) and the relative ($\ln Q_n / \sum_{All} \ln Q$) contributions. The last four columns show the constraint mass, the discriminator value, the acoplanarity and the b-tag value of the event.

n	Run	Event	$\ln Q_n$	Cum	Rel	m_{con}	d_{multi}	$Acop$	b_{tag}
1	110551	16316	1.22	0.127	0.127	59.71	68.51	47.63	0.053
2	117013	2547	0.715	0.201	0.074	66.40	6.735	70.50	2.572
3	114735	2001	0.702	0.274	0.073	95.24	67.95	51.05	-0.350
4	112780	3000	0.678	0.345	0.070	80.61	76.07	0.79	-2.503
5	116369	11333	0.624	0.410	0.065	87.31	23.96	1.29	-3.407
6	111898	2760	0.358	0.447	0.037	75.12	67.29	48.82	1.528
7	110637	20119	0.27	0.476	0.028	77.58	67.11	101.15	0.265
8	111955	5156	0.27	0.504	0.028	79.28	67.12	32.04	4.830
9	114725	15700	0.261	0.531	0.027	84.33	67.08	61.75	4.800
10	111659	20659	0.255	0.557	0.026	87.02	67.03	54.27	-1.263
11	114174	8485	0.249	0.583	0.025	91.69	67.10	57.78	-0.173
12	112048	7786	0.243	0.609	0.025	95.85	67.00	20.30	1.000
13	113235	3486	0.23	0.633	0.023	100.91	67.17	37.76	-0.162
14	111227	4116	0.203	0.654	0.021	66.11	66.86	40.02	0.621
15	113879	4882	0.199	0.675	0.020	73.34	66.88	92.71	1.327
16	109690	1626	0.181	0.693	0.018	90.95	66.86	52.99	0.569
17	111509	14762	0.171	0.711	0.017	97.10	66.87	53.35	8.221
18	113631	11694	0.161	0.728	0.016	104.40	66.91	32.66	-0.556
19	112171	304	0.147	0.743	0.015	68.93	66.76	41.15	1.523
20	113688	5834	0.147	0.759	0.015	61.82	66.69	21.80	-1.610

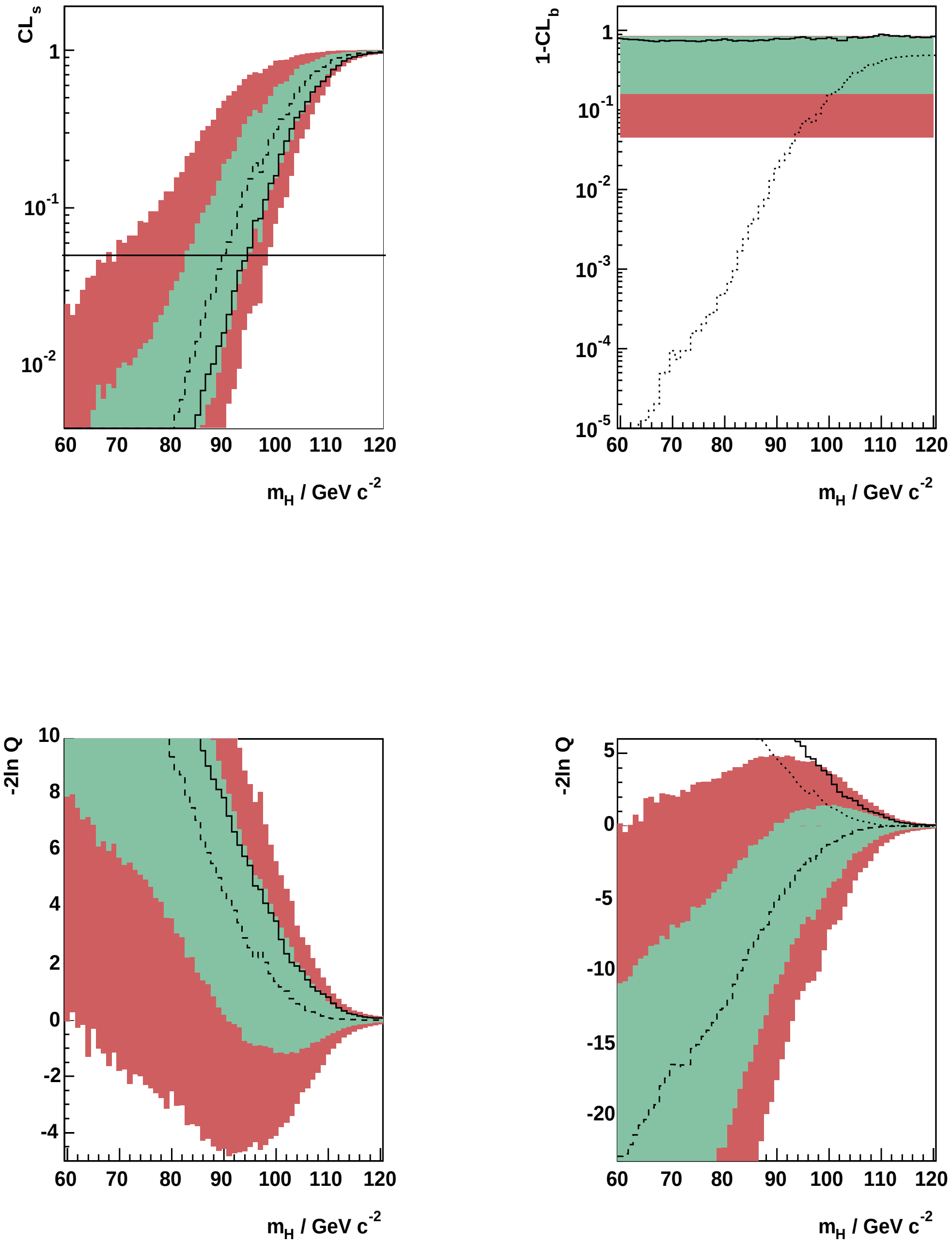


Figure 4.16: Results optimized for m_H = 80 GeV/c^2. The upper graphs show CL_s and $1-CL_b$ for Higgs mass hypotheses from 60 GeV/c^2 to 120 GeV/c^2. The bands correspond to one and two standard deviations around the dashed median expectation for the background. The observed results are solid. The dotted line in the $1-CL_b$ graph represents the median expectation for a signal at the test mass. The lower graphs show $-2\ln Q$ for the background only and the background plus signal hypotheses. The dotted line in the latter graph is the expected median for the background only hypothesis.

4.6.3 Search Optimized at $m_H = 90$ GeV/c^2

Since a discrimination optimized for a Higgs boson mass $m_H = 90$ GeV/c^2 has a decreasing sensitivity for finding Higgs bosons with higher masses, discriminations with heavier optimization masses were computed. Now third order polynomials were estimated with a signal mass equal to 90 GeV/c^2. A first iteration cut was performed on the polynomials achieved with the samples resulting from the minimal univariate discrimination presented in Section 4.4. The efficiency for the 90 GeV/c^2 signal was also chosen to be 70%.

The resulting samples after the cut on the first order polynomials were used to train second iteration polynomials. The discriminating power of these polynomials is shown in Figure 4.17 and Figure 4.18 for the E1 samples and the U1 samples respectively. The distributions of the observables which were used to estimate the polynomials are shown in Figures 4.19, 4.20, 4.21 and 4.22.

There are now about 10 remaining background events in the E1 samples for a 50% efficiency ($m_H = 90$ GeV/c^2). The corresponding numbers for the U1 samples is 5. Compared with the published analysis for low masses, this is a significant improvement. According to Figure 11 in the publication [5] the independent analysis achieves about 25 events at a 50% efficiency for a Higgs boson mass equal to 90 GeV/c^2. Considering that the published analaysis contains an intricate set of sequential cuts and a multivariate likelihood ratio discriminator, the single univariate cut and the polynomial discriminator presented here seem to have a higher discriminating power and a more transparent structure.

Figure 4.23 shows the distributions of the Higgs boson mass estimator, Constrained Mass, and the Discriminator. The upper two graphs contain the E1 samples, the lower two the U1 samples.

Hypothesis Test and Limits

The study of different cuts on the second iteration polynomials showed that the expected CL_s reached their minima for signal efficiencies ($m_H = 90$ GeV/c^2) of 42.4% and 49.1%. With the nomenclature which was established in Equation 2.21, this corresponded to $s_{E1} = 6.64$, $b_{E1} = 4.94$, $s_{U1} = 2.75$ and $b_{U1} = 4.13$. At these efficiencies the number of observed candidates were 6 and 4 respectively.

The candidates with highest $\ln Q$ are listed in Table 4.9. They are ordered with respect to the $\ln Q$. While the fourth column contains the accumulated event contributions to the $\ln Q$, the fifth column shows the individual contributions. The event with the highest $\ln Q$ contributes with 14.2%. Among the ten listed events in Table 4.9, the last four are from the U1 sample. The event number 4 and event number 10 are among the 10 best candidates in the discrimination optimized at $m_H = 80$ GeV/c^2 (see Table 4.8). Further the effect of the new optimization may be seen by calculating the average constrained mass.

For the 10 best candidates obtained by optimizing at $m_H = 80$ GeV/c^2 this is 79.3 GeV/c^2. For the optimization at $m_H = 90$ GeV/c^2 it is 89.6 GeV/c^2.

The results for the search optimized at $m_H = 90$ GeV/c^2 are summarized in Figure 4.24. The upper left graph shows the CL_s for for Higgs mass hypotheses from 70 GeV/c^2 to 120 GeV/c^2. The observed result is solid. The expected median is dashed. It is clear that that the observed result behaves like a perfect background-only experiment. The shape follows from the decreasing signal cross section at higher masses and the decreasing signal efficiency for masses below 85 GeV/c^2. In this case, where the minimal CL_s expected was found for relatively low signal efficiencies, i.e. hard cuts on the discriminators, the efficiency becomes especially sensitive to the signal mass. It is high around the optimized mass, but drops rapidly with increasing distance from this mass.

Again, where the expected median and the observed line intersect the horizontal line at $CL_s = 0.05$, the 95% exclusion limits are found.

$$m_H^{exp} \geq 96.8 \text{ GeV/c}^2$$

$$m_H^{obs} \geq 96.3 \text{ GeV/c}^2$$

The $1 - CL_b$ graph in Figure 4.24 confirms the background-like observation. According to the smaller signal cross section at 90 GeV/c^2, relatively to 80 GeV/c^2, the dotted signal plus background expectation is now more background-like and far from any potential discovery. The $-2\ln Q_b$ and $-2\ln Q_{s+b}$ graphs only emphasize these conclusions.

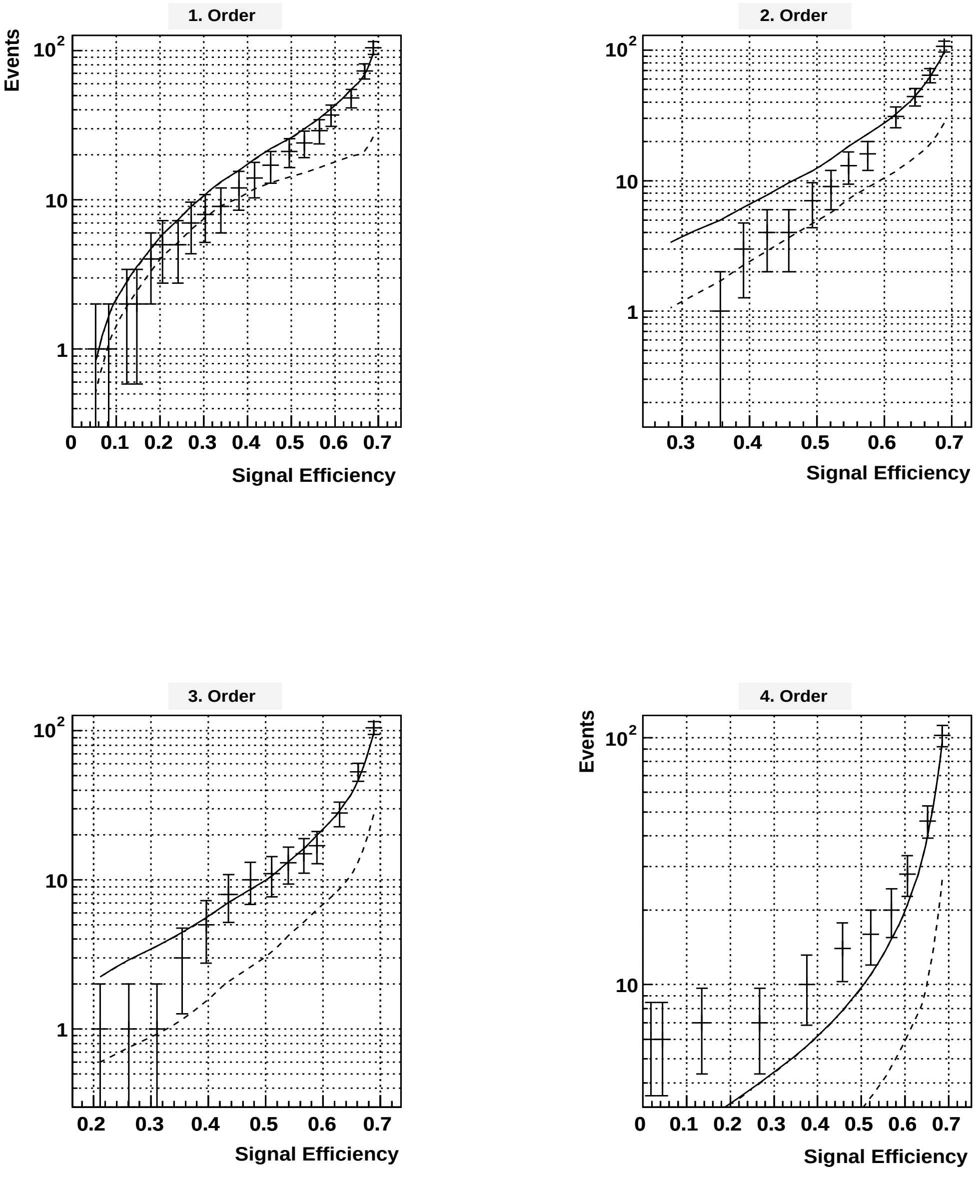


Figure 4.17: E1 Background events versus efficiency for a 90 GeV/c^2 signal mass. The measured data is shown with error bars. The dashed lines are expected background contributions from the two fermion processes. The solid lines represent the total expected background.

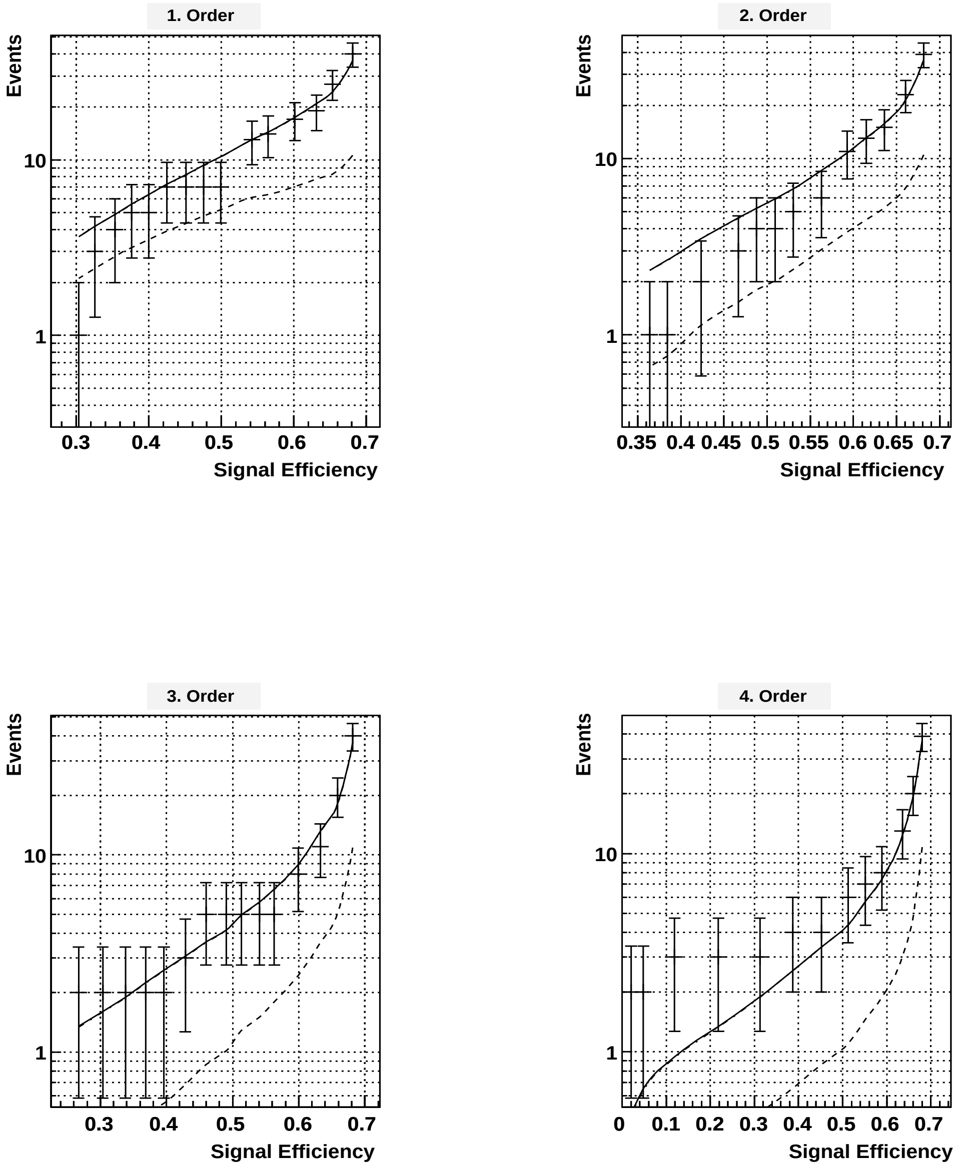


Figure 4.18: U1 Background events versus efficiency for a 90 GeV/c^2 signal mass. The measured data is shown with error bars. The dashed lines are expected background contributions from the two fermion processes. The solid lines represent the total expected background.

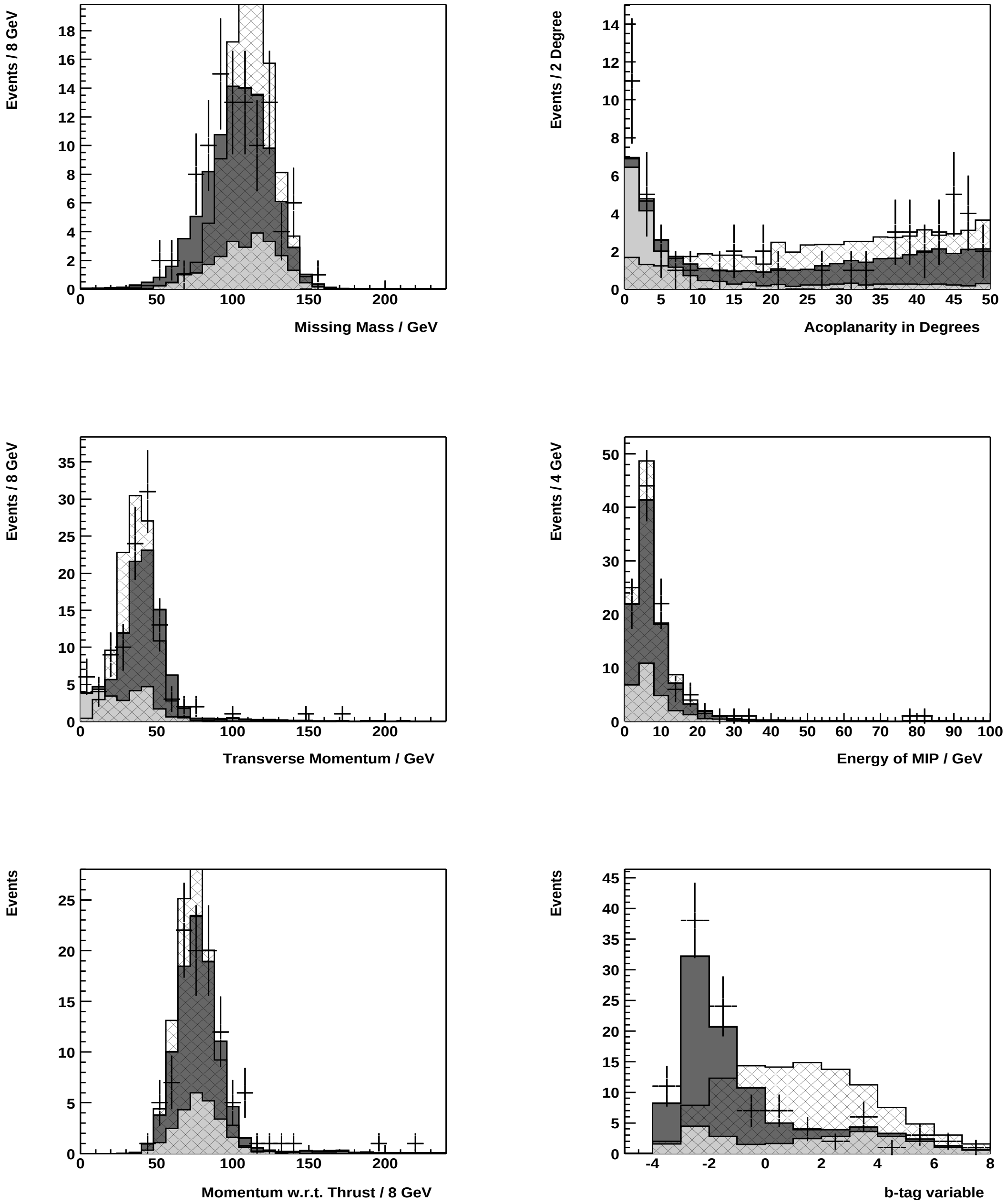


Figure 4.19: Simulated and measured distributions of E1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 157.7 pb^{-1}. The scaled Higgs signal has a mass of 90 $\mathrm{GeV}/\mathrm{c}^2$. Gray scales as in Figure 4.3.

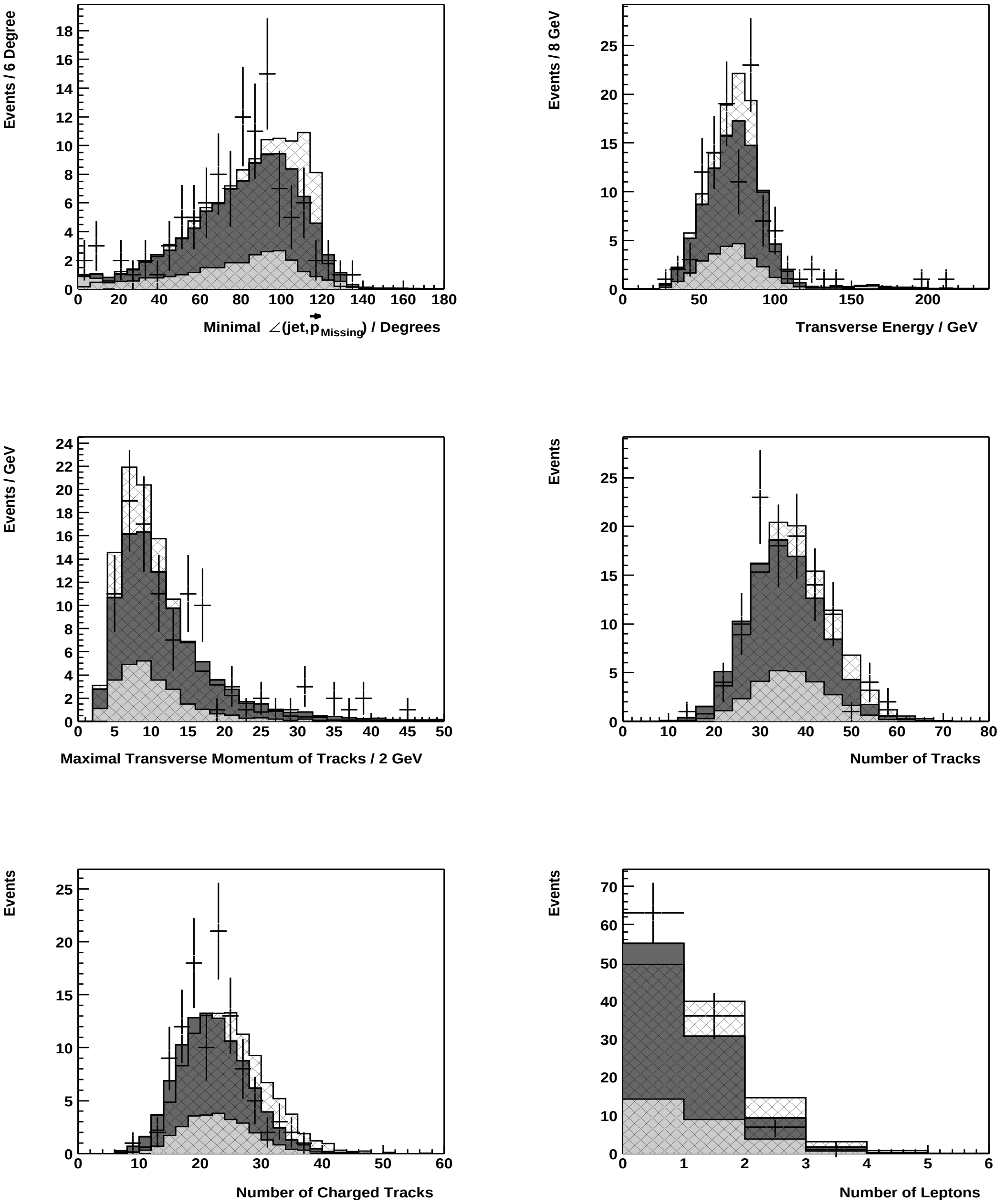


Figure 4.20: Simulated and measured distributions of E1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 157.7 pb^{-1}. The scaled signal has a mass of 90 GeV/c^2. Gray scales as in Figure 4.3.

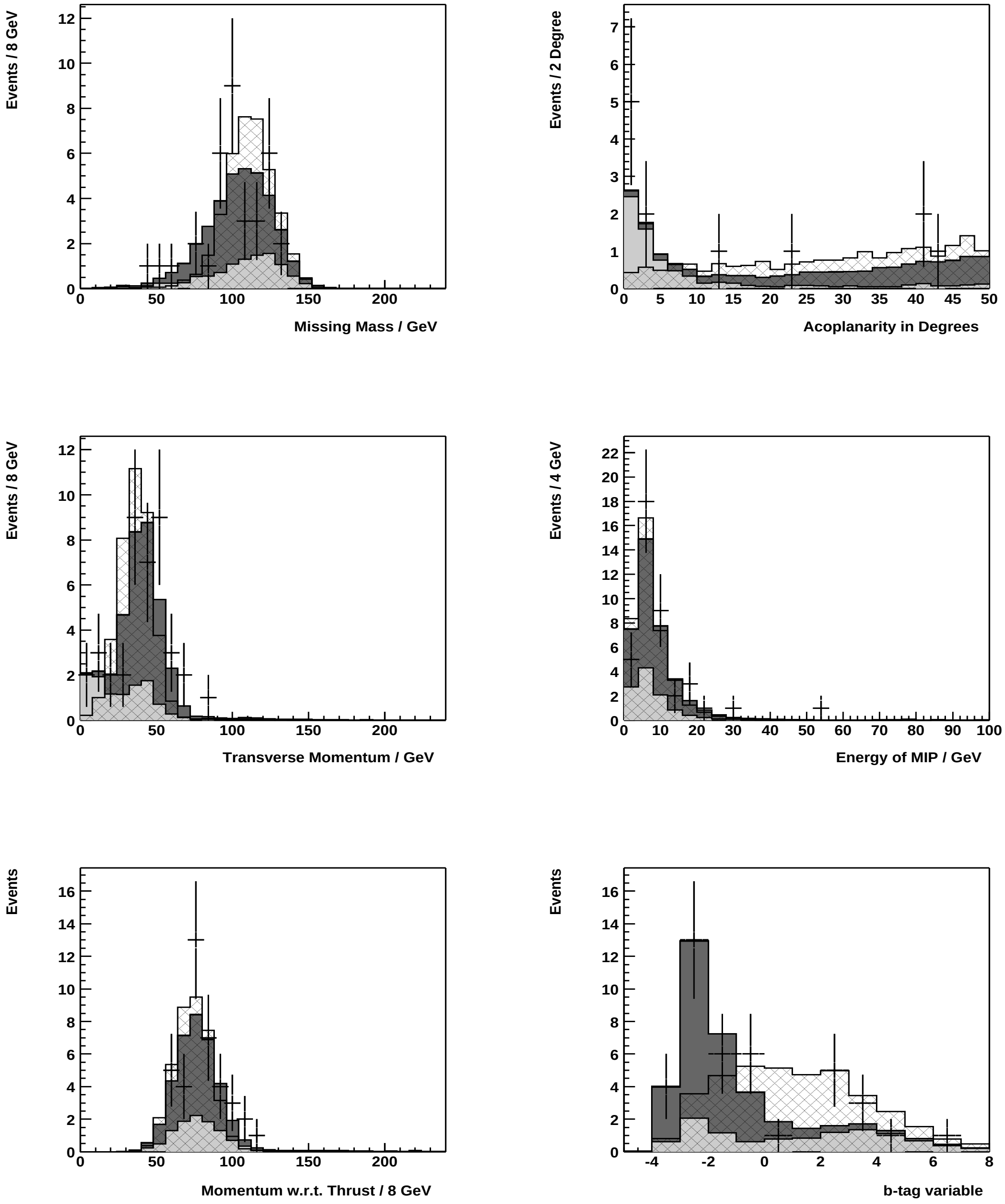


Figure 4.21: Simulated and measured distributions of U1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 56.31 pb^{-1}. The scaled signal has a mass of 90 GeV/c^2. Gray scales as in Figure 4.3.

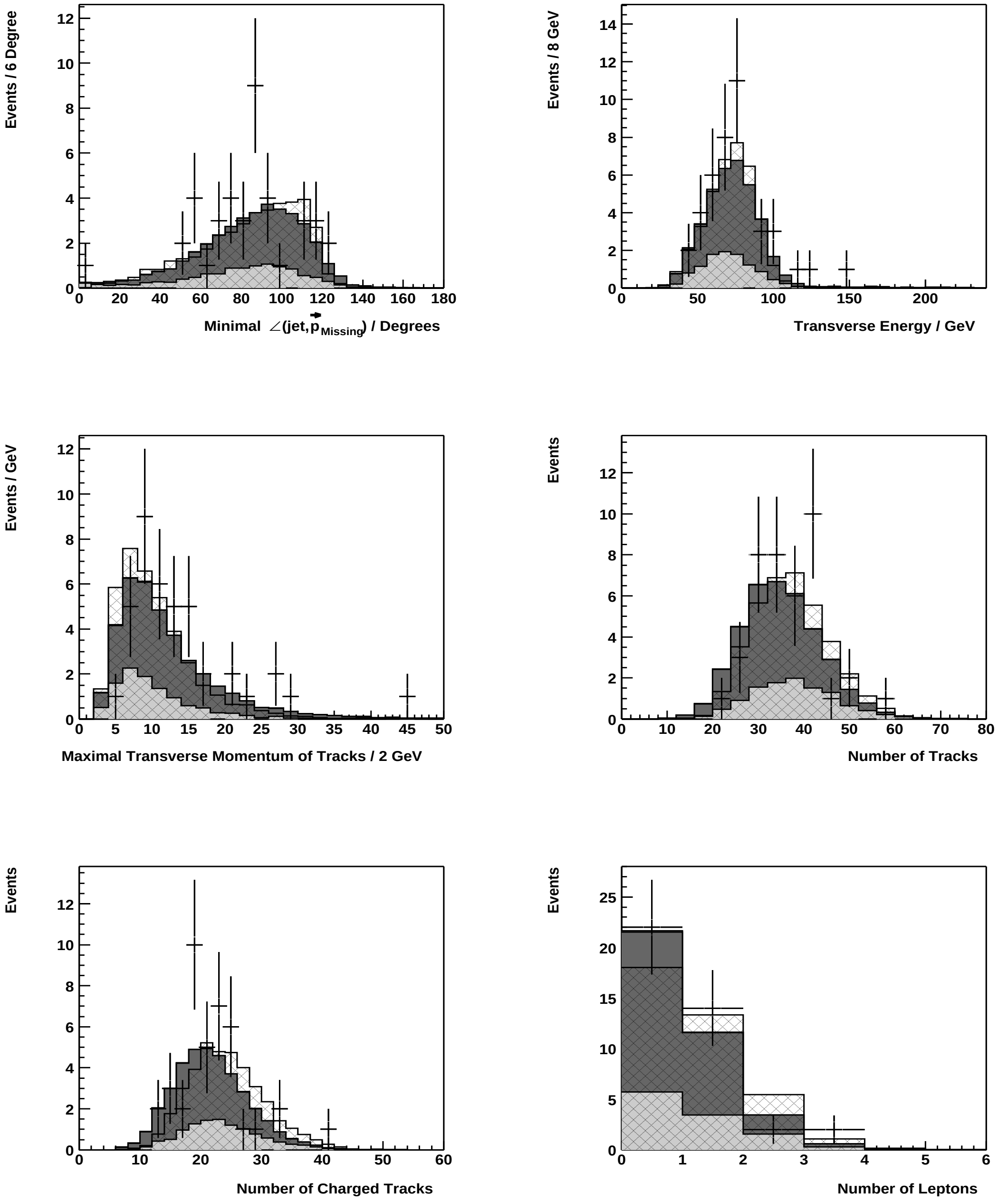


Figure 4.22: Simulated and measured distributions of U1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 56.31 pb^{-1}. The scaled signal has a mass of 90 GeV/c^2. Gray scales as in Figure 4.3.

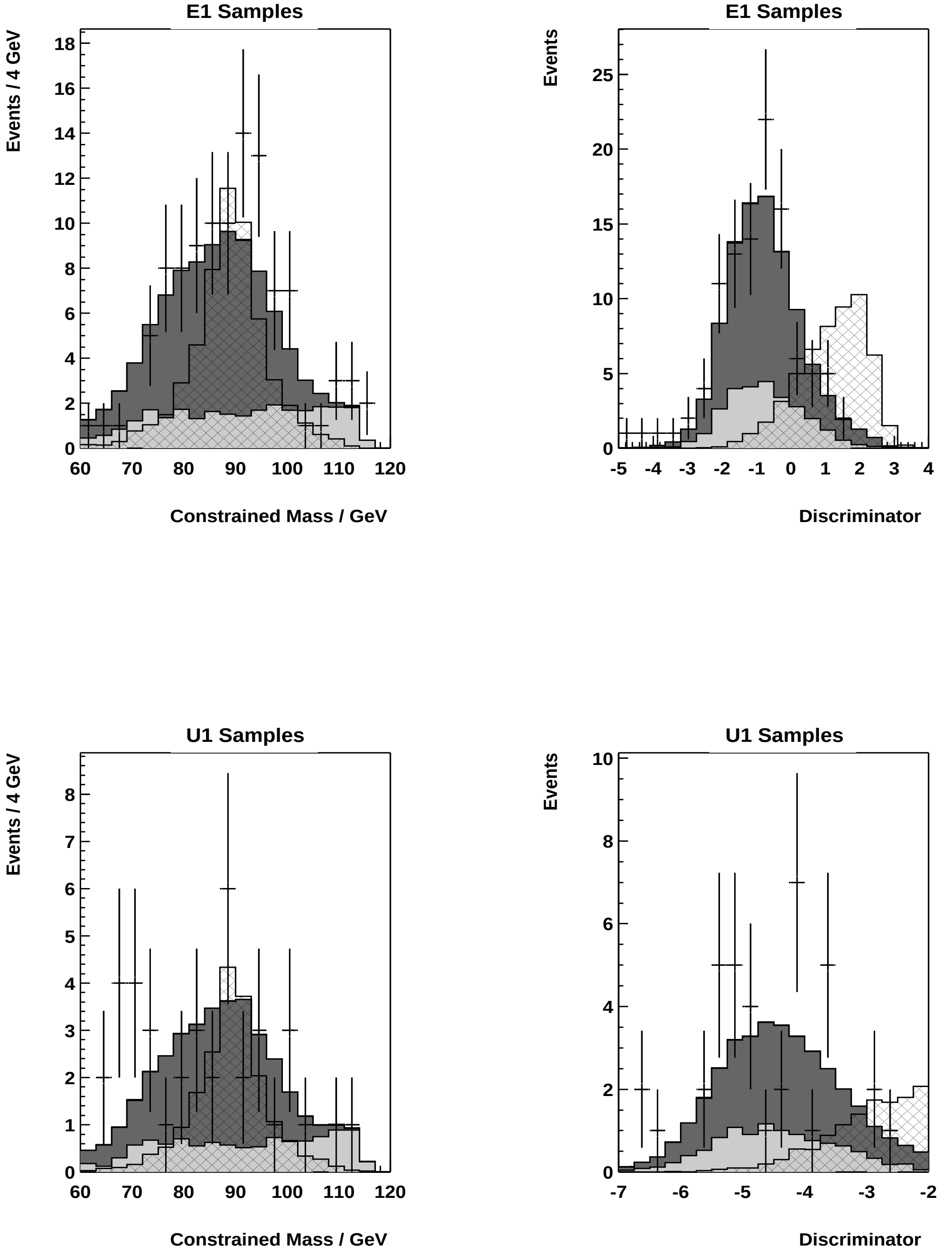


Figure 4.23: Simulated and measured distributions. The background is divided into three accumulated parts. The scaled signals have a mass of 90 GeV/c^2.
Gray scales as in Figure 4.3.

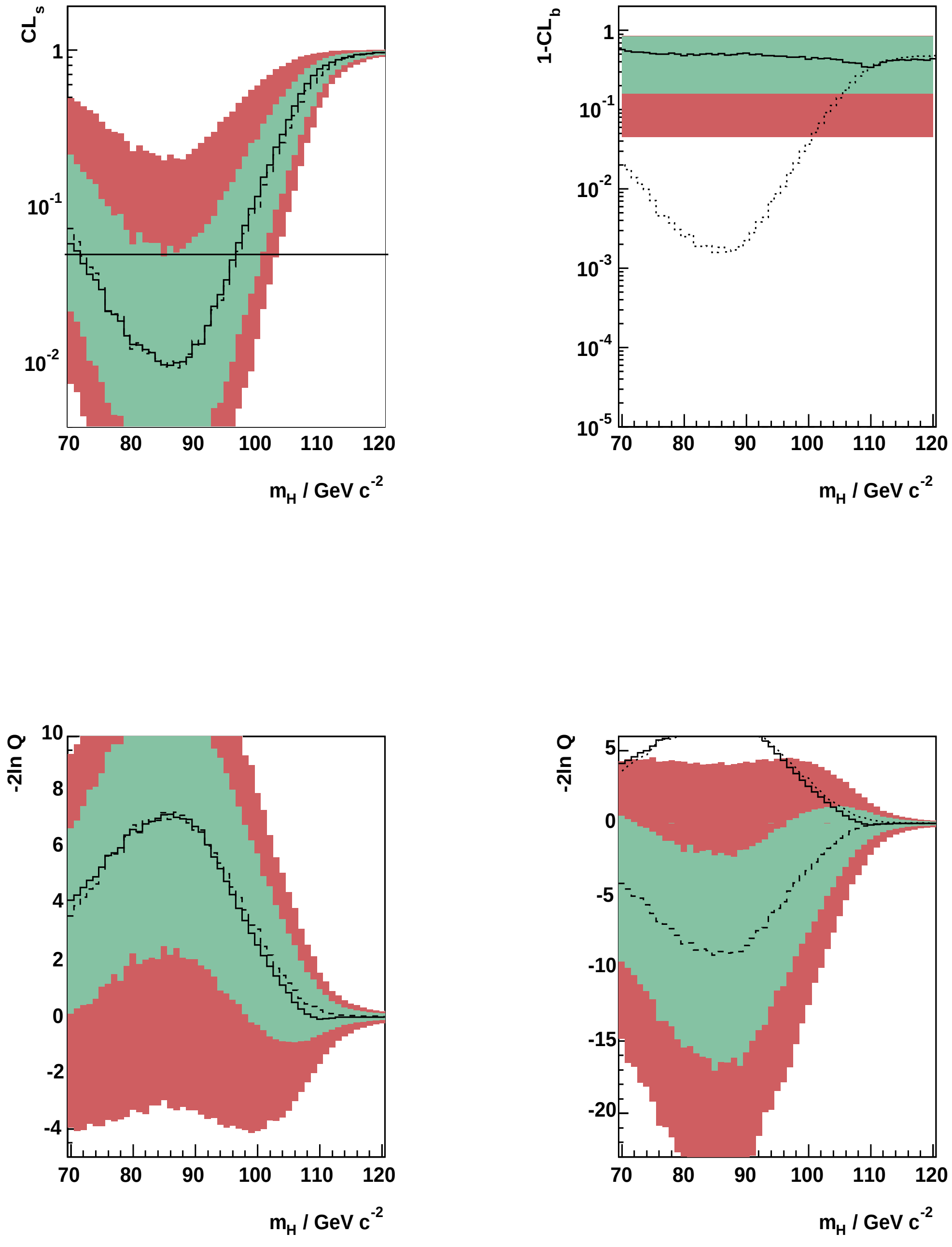


Figure 4.24: Results optimized for $m_H = 90$ GeV/c^2. The upper graphs show CL_s and $1 - CL_b$ for Higgs mass hypotheses from 60 GeV/c^2 to 120 GeV/c^2. The bands correspond to one and two standard deviations around the dashed median expectation for the background. The observed results are solid. The dotted line in the $1 - CL_b$ graph represents the median expectation for a signal at the test mass. The lower graphs show $-2 \ln Q$ for the background only and the background plus signal hypotheses. The dotted line in the latter graph is the expected median for the background only hypothesis.

Table 4.9: Candidates found in a search optimized at $m_H = 90$ GeV/c^2. $Q = 1 + sS(\vec{x})/bB(\vec{x})$ is the test statistic as defined in Equation 2.21. The candidates are sorted with respect to $\ln Q_n$. The fifth and the sixth columns contain the cumulative ($\sum_{i=1}^{n} \ln Q_i / \sum_{All} \ln Q$) and the relative ($\ln Q_n / \sum_{All} \ln Q$) contributions. The last four columns show the constraint mass, the discriminator value, the acoplanarity and the b-tag value of the event.

n	Run	Event	$\ln Q$	Cum	Rel	m_{con}	d_{multi}	$Acop$	b_{tag}
1	111804	2308	0.85	0.141	0.141	77.75	1.41	49.09	3.006
2	111509	14762	0.816	0.277	0.135	97.10	1.41	53.35	8.221
3	114676	486	0.699	0.394	0.116	78.77	1.20	15.96	-2.108
4	114735	2001	0.676	0.506	0.112	95.24	1.14	51.05	-0.350
5	113631	11694	0.65	0.615	0.108	104.4	1.10	32.66	-0.556
6	113235	3486	0.577	0.711	0.096	100.9	1.01	37.76	-0.162
7	117102	2892	0.501	0.794	0.083	93.65	-2.77	22.03	2.602
8	115604	1681	0.471	0.873	0.078	99.03	-2.87	76.99	-2.403
9	117472	2067	0.438	0.946	0.072	82.49	-1.56	41.49	0.014
10	117013	2547	0.323	1.0	0.053	66.40	-2.70	70.50	2.572

4.6.4 Search Optimized at $m_H = 100$ GeV/c^2

The next optimization is done at a Higgs boson mass $m_H = 100$ GeV/c^2. The discriminating power of the polynomials estimated in the second iteration is shown in Figure 4.25 and Figure 4.26. The cut on the first iteration discriminator reduced the signal efficiencies down to 69.1% and 68.7% for the E1 and U1 samples respectively. The third order polynomials were used. The remaining distributions of the observables are shown in Figures 4.27, 4.28, 4.29 and 4.30.

Hypothesis Test and Limits

The study of different cuts on the second iteration polynomials showed that the expected CL_s reached their minima for signal efficiencies ($m_H = 100$ GeV/c^2) of 47.5% and 40.4%. With the nomenclature which was established in Equation 2.21, this corresponded to $s_{E1} = 5.40$, $b_{E1} = 7.99$, $s_{U1} = 1.64$ and $b_{U1} = 1.69$. At these efficiencies the number of observed candidates were 8 and 3 respectively.

The candidates with highest $\ln Q$ are listed in Table 4.10. They are ordered with respect to the $\ln Q$. While the fourth column contains the cumulated event contributions to the $\ln Q$, the fifth column shows the individual contributions. The event with the highest $\ln Q$ contributes with 17.3%. Among the eleven listed events in Table 4.10, the fifth, the sixth and the eight are from the U1 sample. Candidates number 3, 5, 6, 7, 8, and 10 are also among the ten bests in the $m_H = 90$ GeV/c^2 optimization (see Table4.9). For the ten best candidates the average constrained mass is now 94.5 GeV/c^2.

The results for the search optimized at $m_H = 100$ GeV/c^2 are summarized in Figure 4.32. The upper left graph shows the CL_s for for Higgs mass hypotheses from 70 GeV/c^2 to 120 GeV/c^2. The observed result is shown by the solid curve. The expected median is dashed. For this optimization the expected exclusion limit is well above 100 GeV/c^2. But the observed result does not intersect the horizontal line at $CL_s = 0.05$, and a 95% exclusion limit is not obtained. The other graphs show that the background plus signal hypothesis now becomes more and more similar to the background only hypothesis. One expects that searches optimized at higher masses will also not produce any exclusion limits.

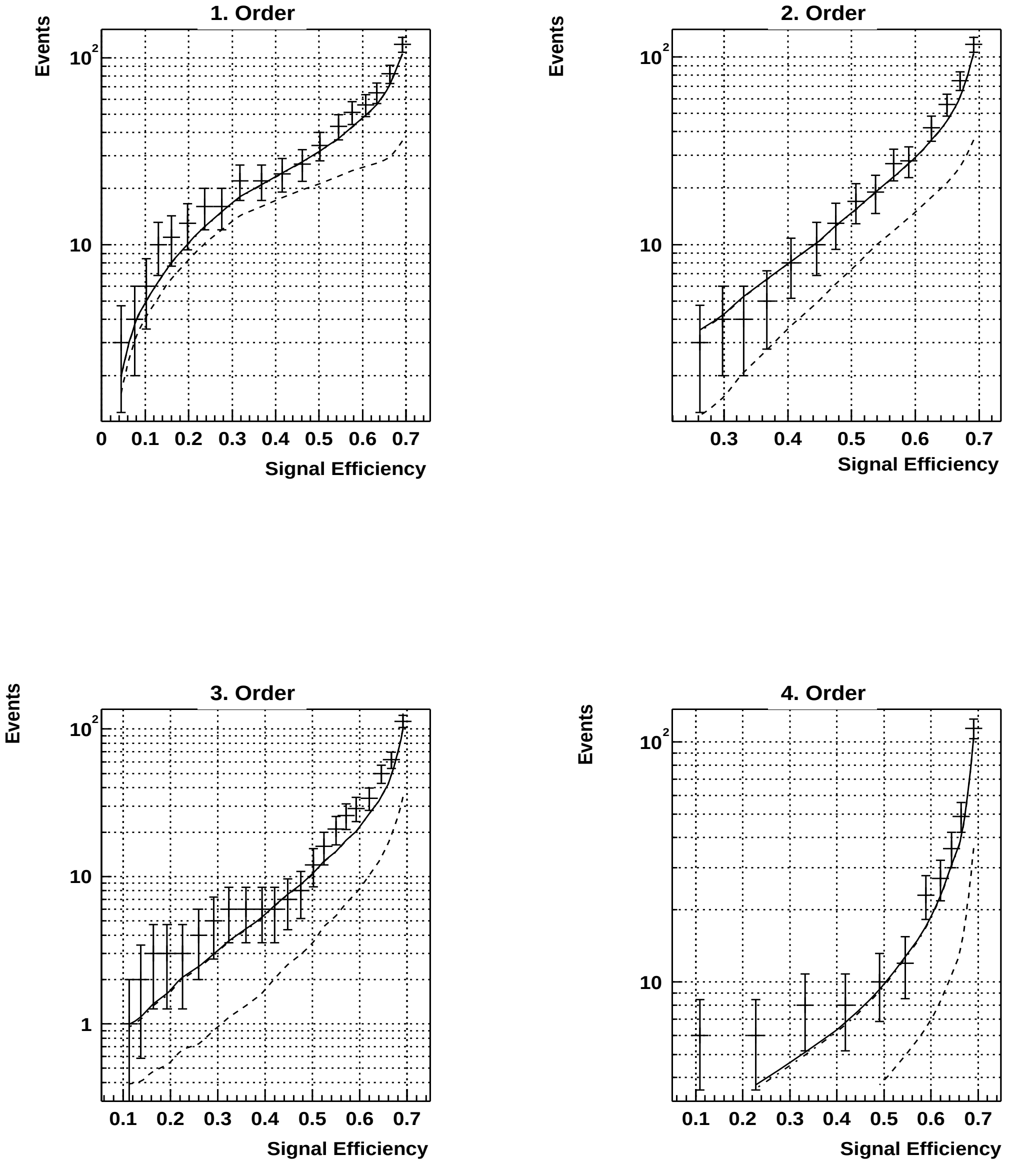


Figure 4.25: E1 Background events versus efficiency for a 100 GeV/c^2 signal mass. The measured data is shown with error bars. The dashed lines are expected background contributions from the two fermion processes. The solid lines represent the total expected background.

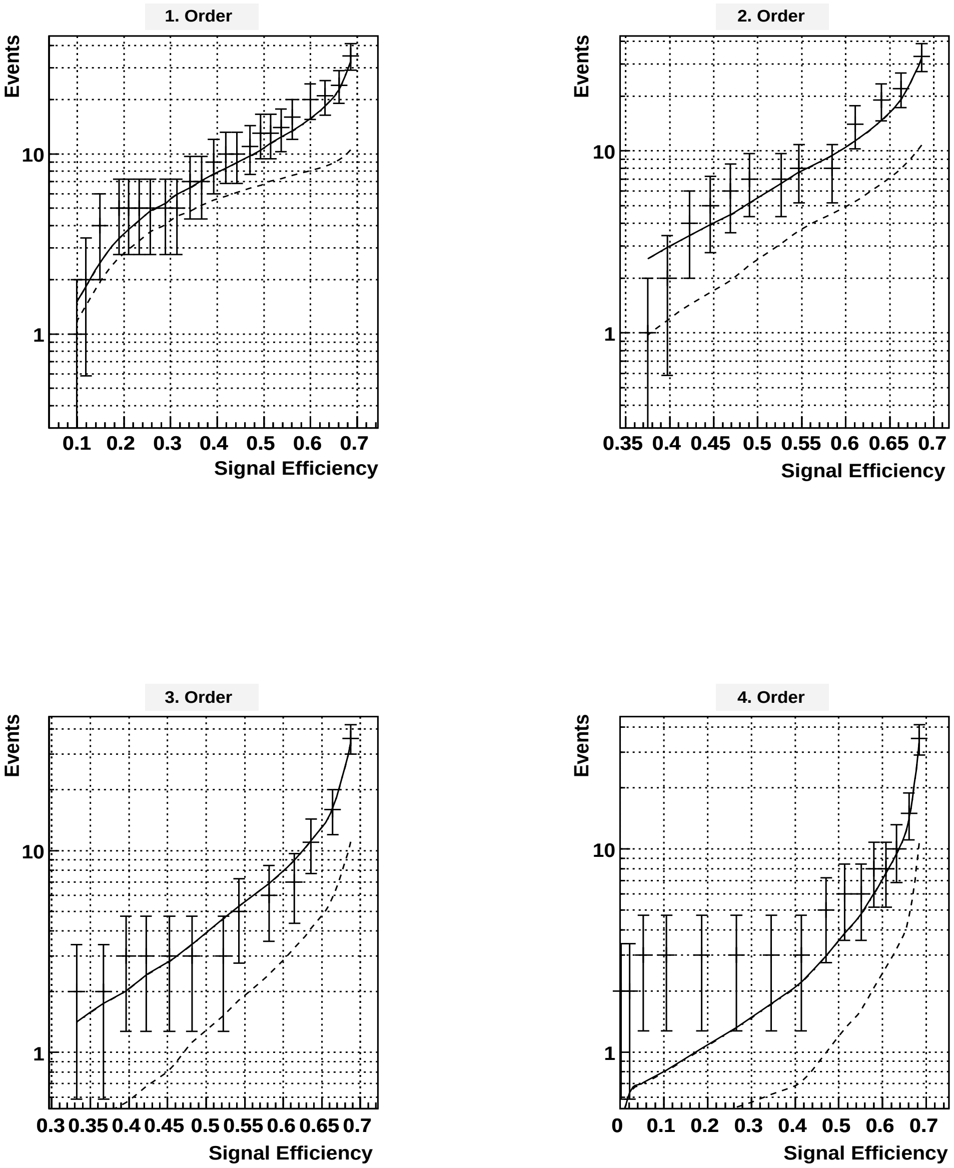


Figure 4.26: U1 Background events versus efficiency for a 100 GeV/c^2 signal mass. The measured data is shown with error bars. The dashed lines are expected background contributions from the two fermion processes. The solid lines represent the total expected background.

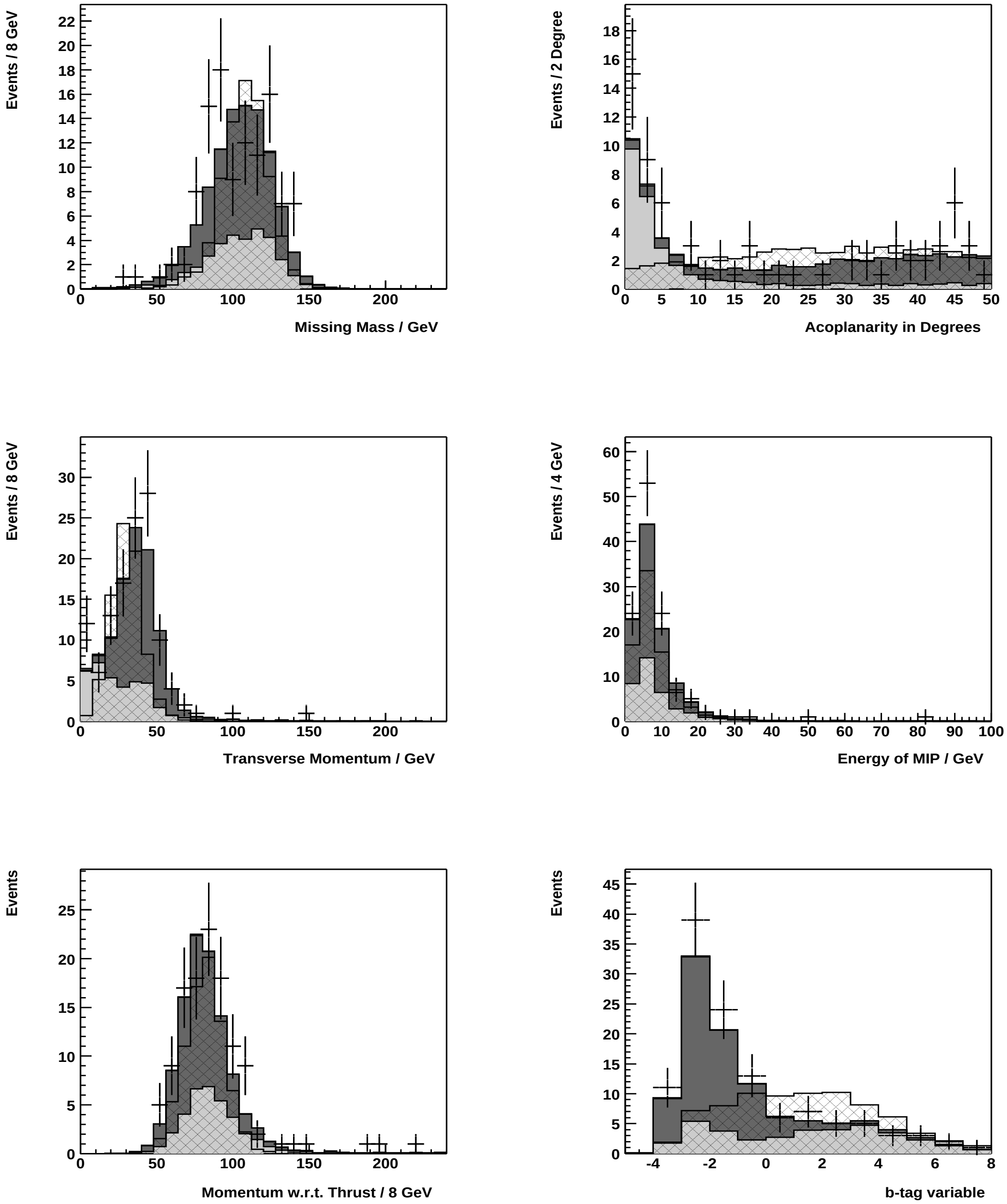


Figure 4.27: Simulated and measured distributions of E1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 157.7 pb^{-1}. The scaled signal has a mass of 100 GeV/c^2. Gray scales as in Figure 4.3.

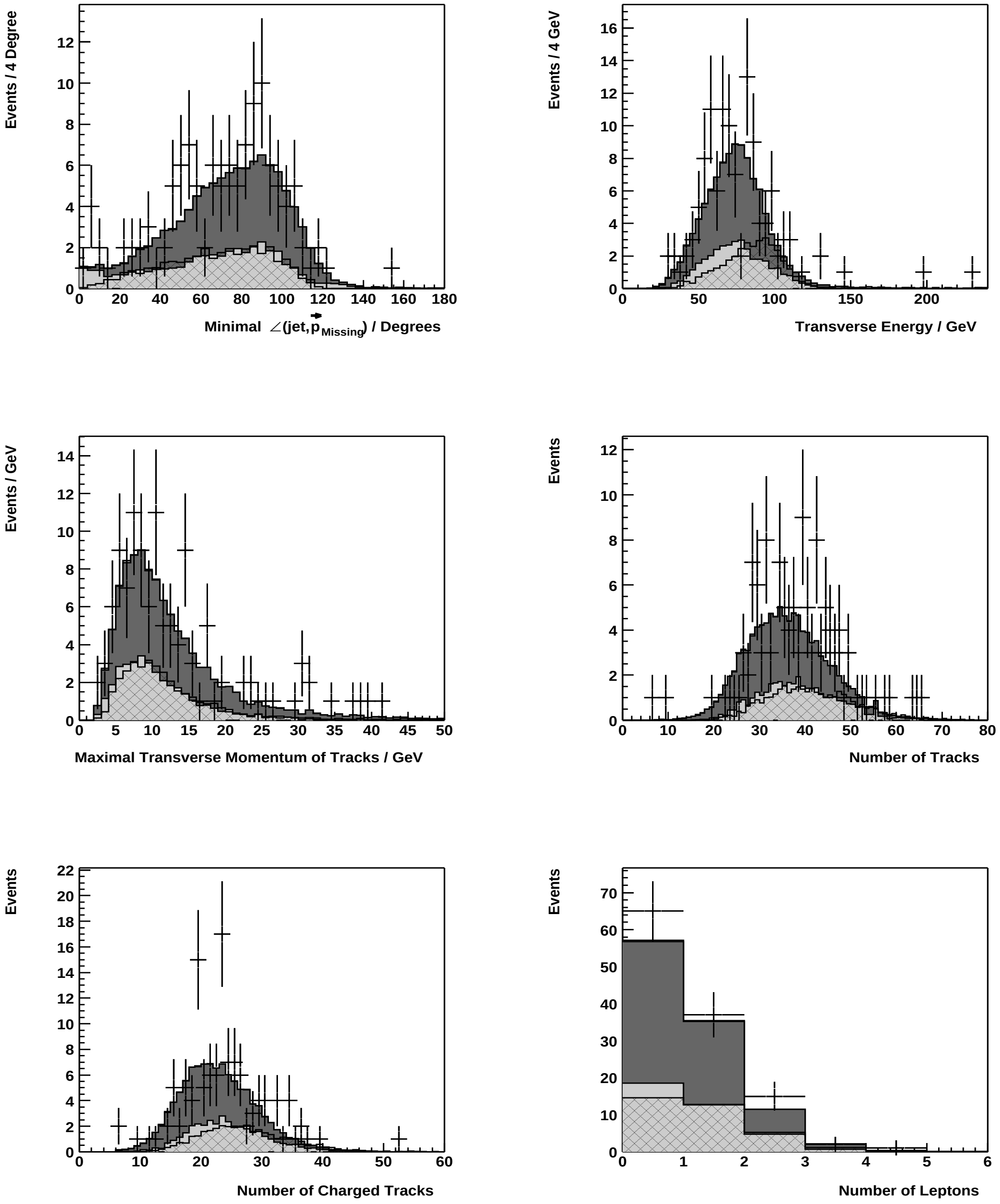


Figure 4.28: Simulated and measured distributions of E1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 157.7 pb^{-1}. The scaled signal has a mass of 100 GeV/c^2. Gray scales as in Figure 4.3.

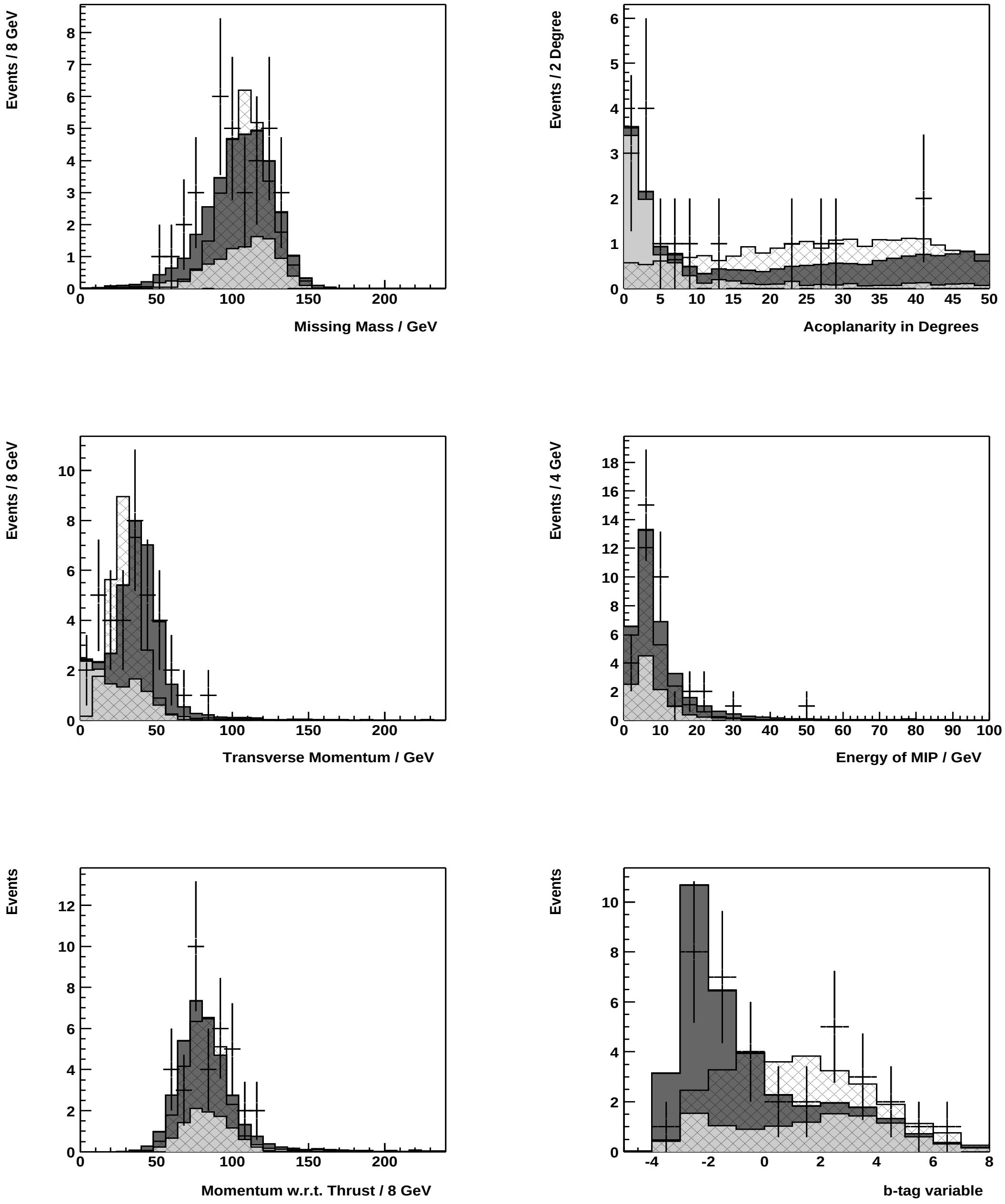


Figure 4.29: Simulated and measured distributions of U1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 56.31 pb^{-1}. The scaled signal has a mass of 100 $\mathrm{GeV/c}^2$. Gray scales as in Figure 4.3.

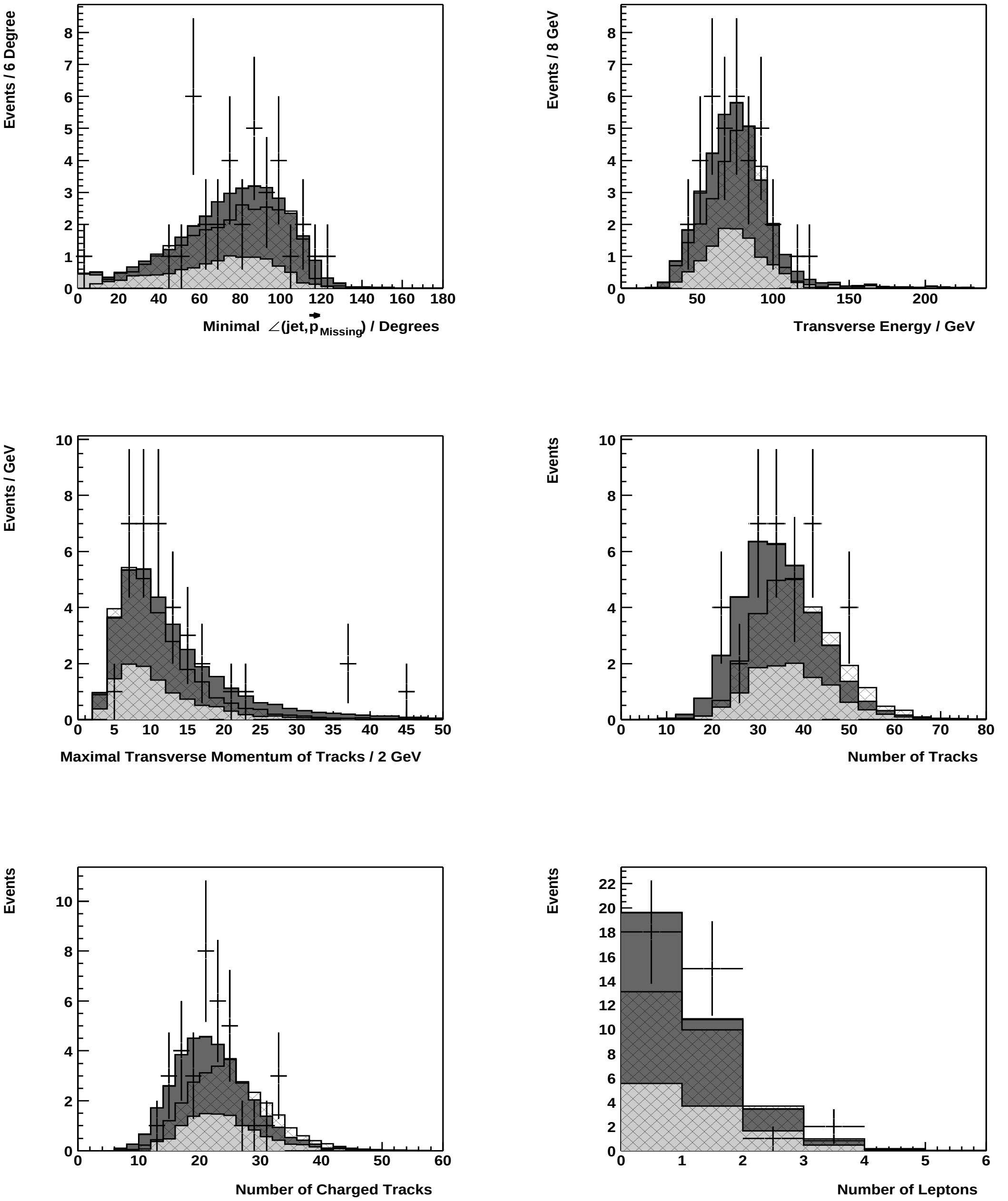


Figure 4.30: Simulated and measured distributions of U1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 56.31 pb^{-1}. The scaled signal has a mass of 100 $\mathrm{GeV/c^2}$. Gray scales as in Figure 4.3.

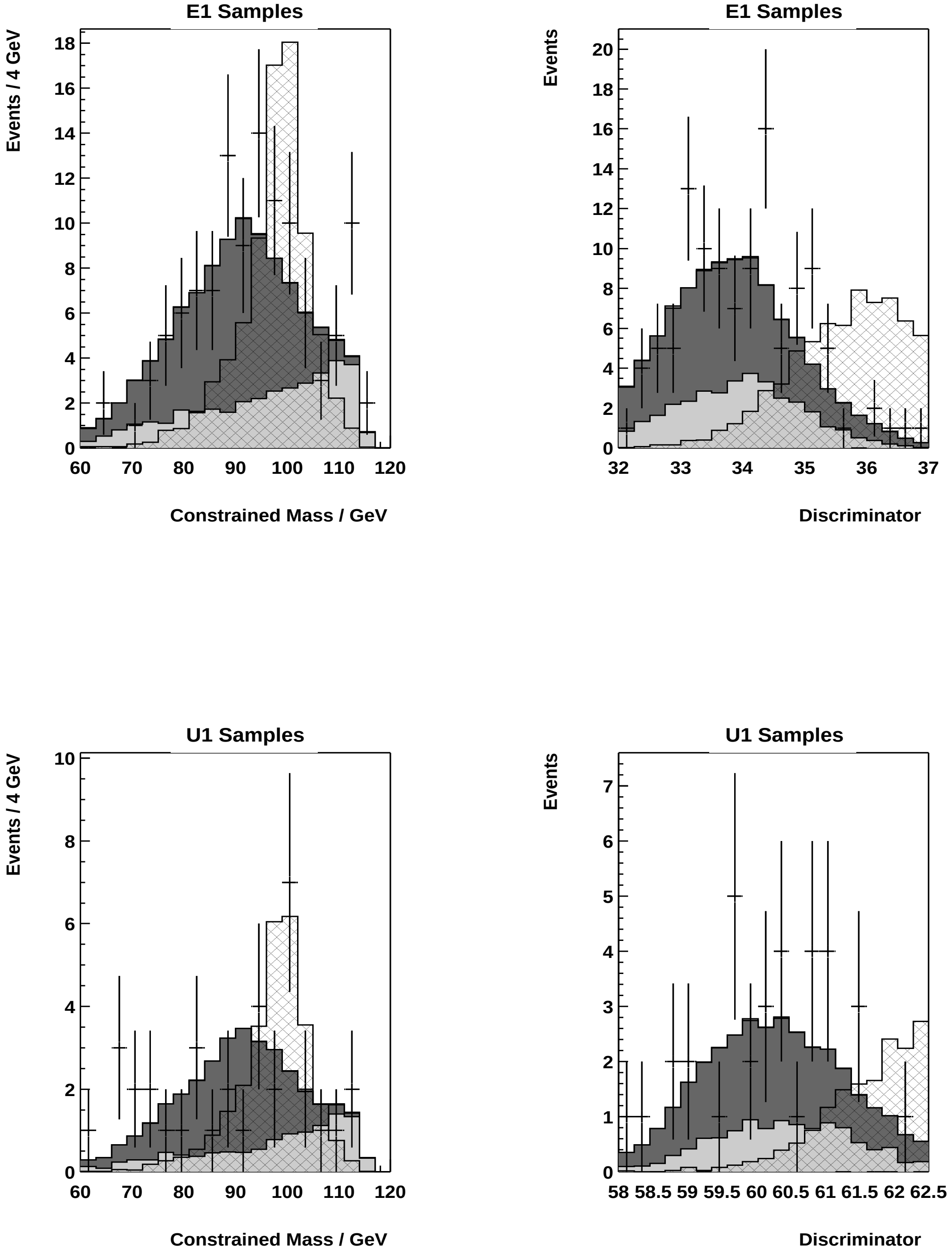


Figure 4.31: The scaled signals have a mass of 100 GeV/c^2.

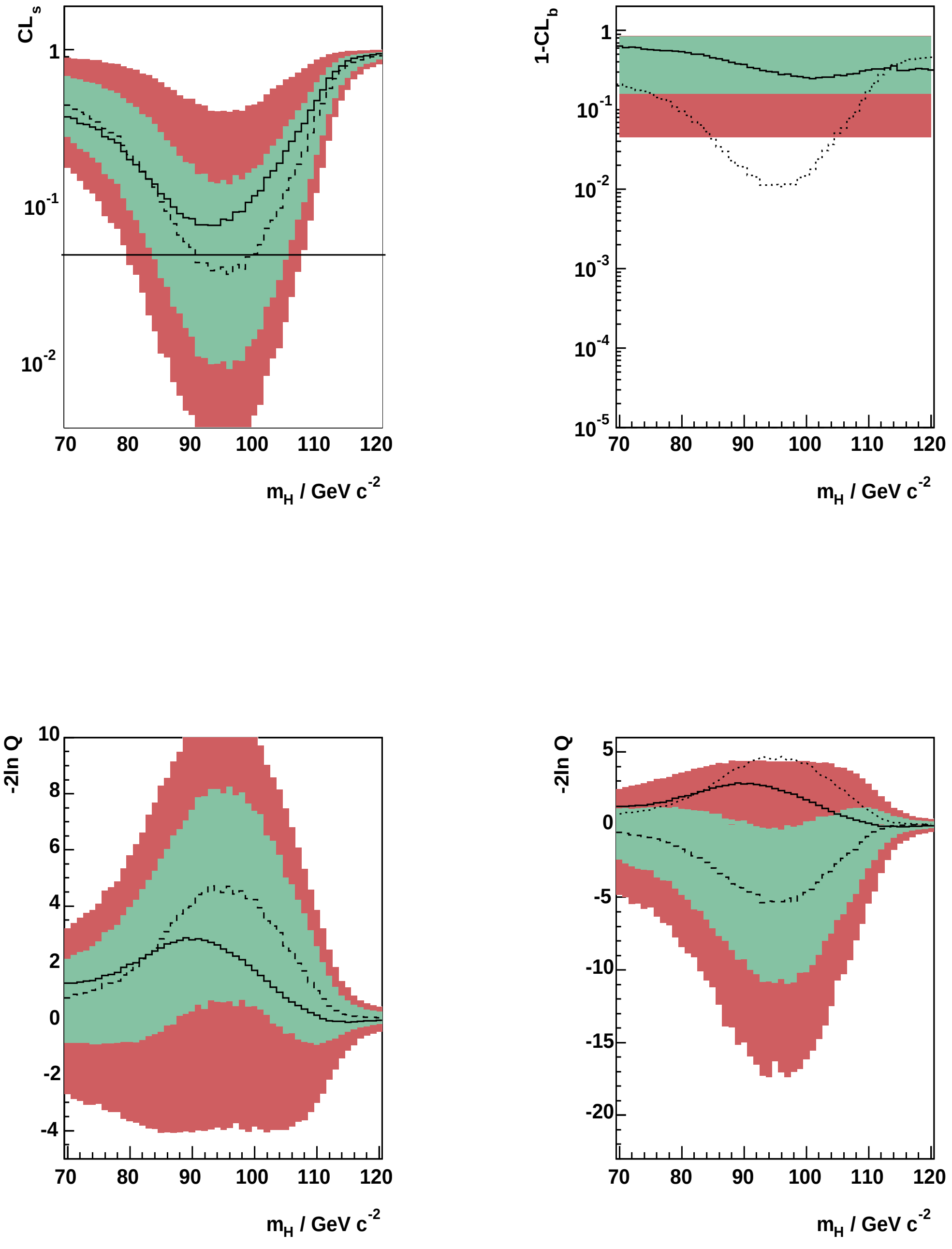


Figure 4.32: Results optimized for $m_H = 100$ GeV/c^2. The upper graphs show CL_s and $1-CL_b$ for Higgs mass hypotheses from 60 GeV/c^2 to 120 GeV/c^2. The bands correspond to one and two standard deviations around the dashed median expectation for the background. The observed results are solid. The dotted line in the $1-CL_b$ graph represents the median expectation for a signal at the test mass. The lower graphs show $-2\ln Q$ for the background only and the background plus signal hypotheses. The dotted line in the latter graph is the expected median for the background only hypothesis.

Table 4.10: Candidates found in a search optimized at $m_H = 100$ GeV/c^2. $Q = 1 + sS(\vec{x})/bB(\vec{x})$ is the test statistic as defined in Equation 2.21. The candidates are sorted with respect to $\ln Q_n$. The fifth and the sixth columns contain the cumulative ($\sum_{i=1}^{n} \ln Q_i / \sum_{All} \ln Q$) and the relative ($\ln Q_n / \sum_{All} \ln Q$) contributions. The last four columns show the constraint mass, the discriminator value, the acoplanarity and the b-tag value of the event.

n	Run	Event	$\ln Q$	Cum	Rel	m_{con}	d_{multi}	$Acop$	b_{tag}
1	111878	1163	1.3	0.173	0.173	113.18	37.30	13.7	-2.206
2	112780	3000	1.04	0.312	0.138	80.61	36.75	0.7	-2.503
3	111509	14762	0.915	0.434	0.122	97.10	36.63	53.3	8.221
4	113302	776	0.667	0.523	0.089	96.09	36.25	23.0	2.958
5	117102	2892	0.658	0.611	0.087	93.65	62.59	22.0	2.602
6	115604	1681	0.643	0.697	0.085	99.03	62.17	76.9	-2.403
7	113631	11694	0.579	0.775	0.077	104.40	36.07	32.6	-0.556
8	117472	2067	0.559	0.849	0.074	82.49	62.51	41.4	0.014
9	111881	10053	0.503	0.917	0.067	77.38	36.12	41.4	-2.584
10	113235	3486	0.331	0.961	0.044	100.91	35.51	37.7	-0.162
11	112048	7786	0.29	1.000	0.038	95.85	35.46	20.3	1.000

4.6.5 Search Optimized at $m_H = 115$ GeV/c^2

Although the search optimized for a mass equal to $m_H = 100$ GeV/c^2 did not result in an exclusion limit, a search optimized at $m_H = 115$ GeV/c^2 was performed. This was mainly done to test the polynomial discriminator. At $m_H = 115$ GeV/c^2, as at $m_H = 100$ GeV/c^2, an independent search is available for comparison.

The discriminating power of the polynomials estimated in the second iteration is shown in Figure 4.33 and Figure 4.34 for the E1 samples and the U1 samples respectively. The cut on the first discriminator resulted in a 69% efficiency for the $m_H = 115$ signals. The distributions of the observables which were used to estimate the second iteration discriminator are shown in Figures 4.35, 4.36, 4.37 and 4.38 for a efficiency of 54 % for a Higgs boson with a 115 GeV/c^2 mass. Figure 4.39 shows the distributions of the Higgs boson mass estimator, Constrained mass, and the Discriminator estimated from the samples which survived the cut on the first iteration discriminator. The upper two graphs contain the E1 samples, the lower two the U1 samples.

As seen in the lower left graphs of Figure 4.33 and Figure 4.34, there are now about $25 + 10$ remaining background events at a 50% efficiency for a Higgs boson mass equal to 115 GeV/c^2. According to Figure 13 in the publication [5] the independent analysis achieves about 35 events at this efficiency. Again, considering that this independent analysis contains an intricate set of sequential cuts and a multivariate likelihood ratio discriminator, the single univariate cut and the polynomial discriminator presented here seem to have a similar discriminating power, but a more transparent structure.

Hypothesis Test and Limits

The study of different cuts on the second iteration polynomials showed that the expected CL_s reached their minima for signal efficiencies ($m_H = 115$ GeV/c^2) of 39.7% and 68.1%. With the nomenclature which was established in Equation 2.21, this corresponded to $s_{E1} = 0.99$, $b_{E1} = 8.79$, $s_{U1} = 0.61$ and $b_{U1} = 52.90$. At these efficiencies the number of observed candidates were 10 and 56 respectively.

The candidates with highest $\ln Q$ are listed in Table 4.11. They are ordered with respect to the $\ln Q$. While the fourth column contains the cumulated event contributions to the $\ln Q$, the fifth column shows the individual contributions. The event with the highest $\ln Q$ contributes with 13.0%. Among the fifteen listed events in Table 4.11, eight are from the U1 sample. This is more than expected as the luminosity of this sample is roughly three times smaller than for the E1 sample.

The results for the search optimized at $m_H = 115$ GeV/c^2 are summarized in Figure 4.40. The upper left graph shows the CL_s for for Higgs mass hypotheses from 70 GeV/c^2 to 120 GeV/c^2. The observed result is shown by the solid curve . The expected median is dashed. In the upper right graph a small excess is seen. The observed curve enters the two sigma band at $m_H \sim 90$ GeV/c^2 and stays there for higher masses. However, this excess is entirely compatible with the background hypothesis and the optimization at this high mass does not allow for any exclusion, neither expected nor observed. The two alternate hypotheses are now nearly indistinguishable.

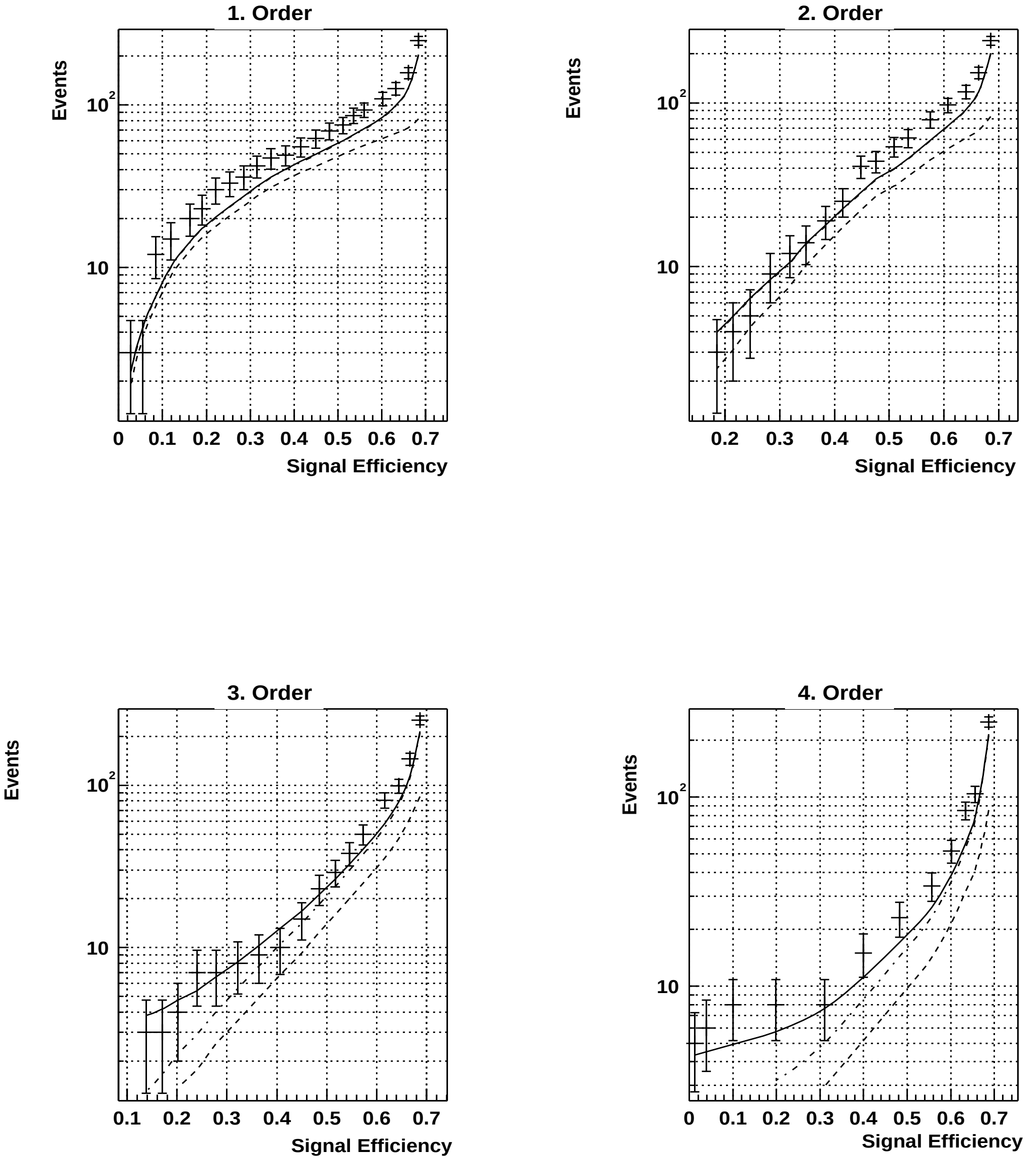


Figure 4.33: E1 Background events versus efficiency for a 115 GeV/c^2 signal mass. The measured data is shown with error bars. The dashed lines are expected background contributions from the two fermion processes. The solid lines represent the total expected background.

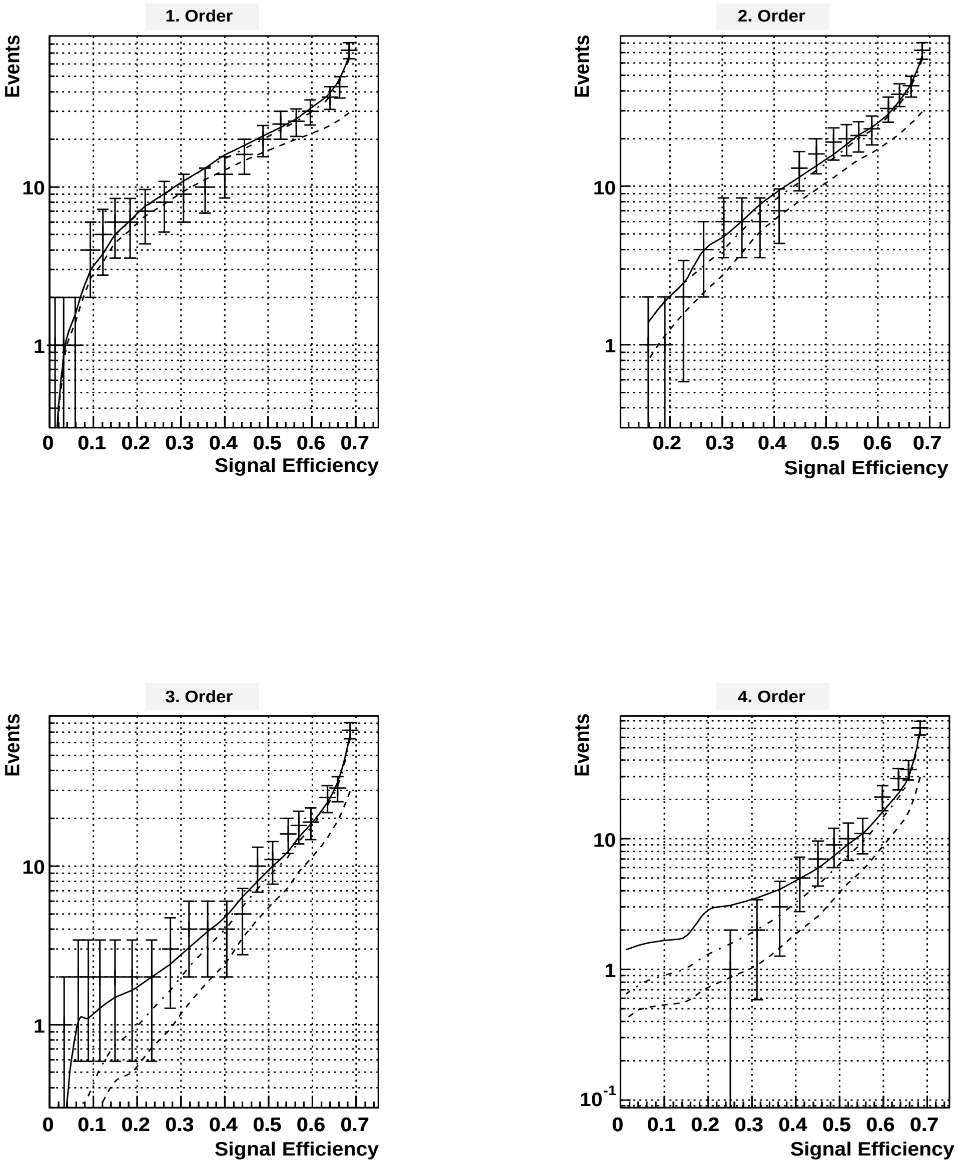


Figure 4.34: U1 Background events versus efficiency for a 115 GeV/c^2 signal mass. The measured data is shown with error bars. The dashed lines are expected background contributions from the two fermion processes. The solid lines represent the total expected background.

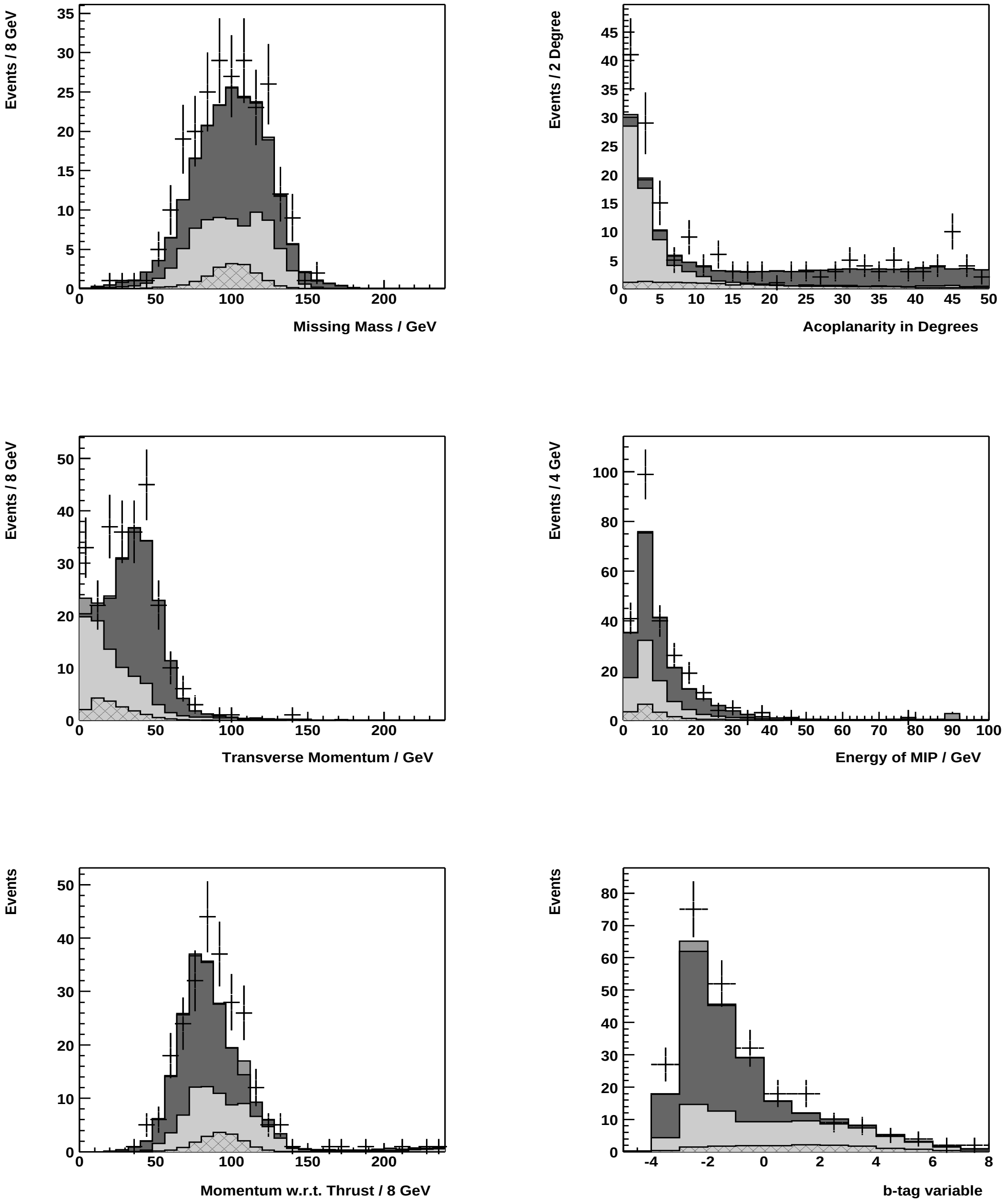


Figure 4.35: Simulated and measured distributions of E1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 157.7 pb^{-1}. The scaled signal has a mass of 115 GeV/c^2. Gray scales as in Figure 4.3.

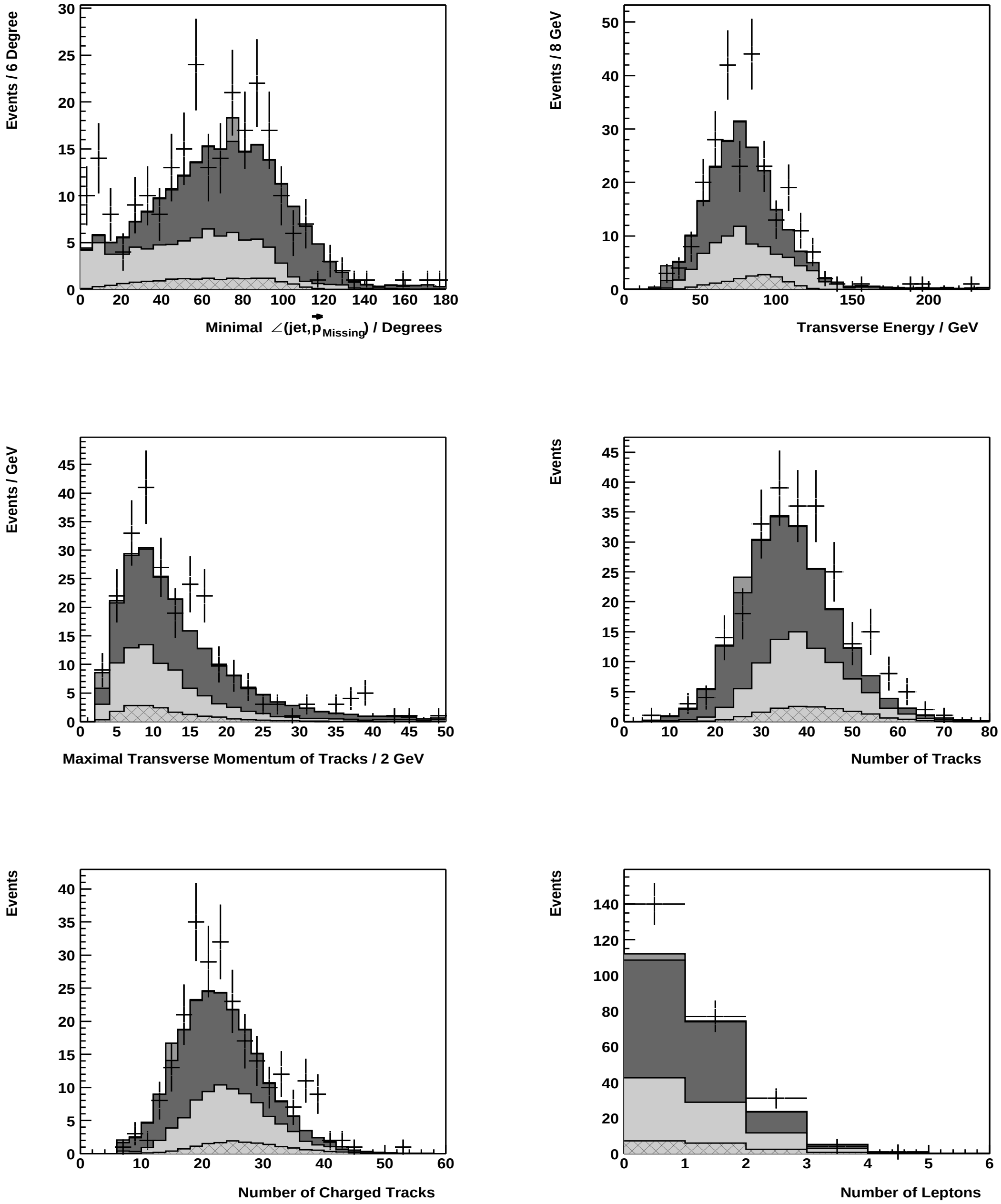


Figure 4.36: Simulated and measured distributions of E1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 157.7 pb^{-1}. The scaled signal has a mass of 115 GeV/c^2. Gray scales as in Figure 4.3.

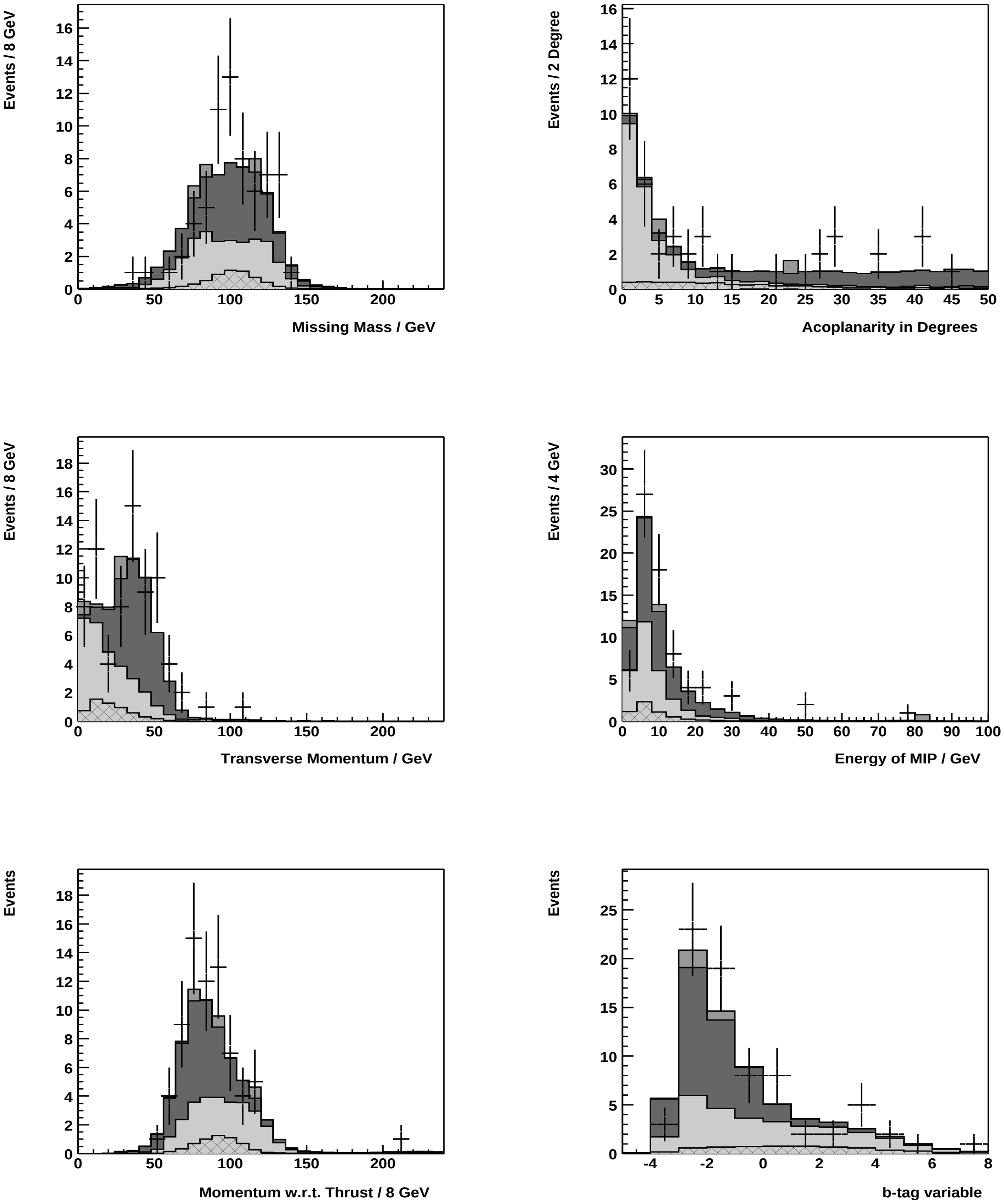


Figure 4.37: Simulated and measured distributions of U1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 56.31 pb^{-1}. The scaled signal has a mass of 115 GeV/c^2. Gray scales as in Figure 4.3.

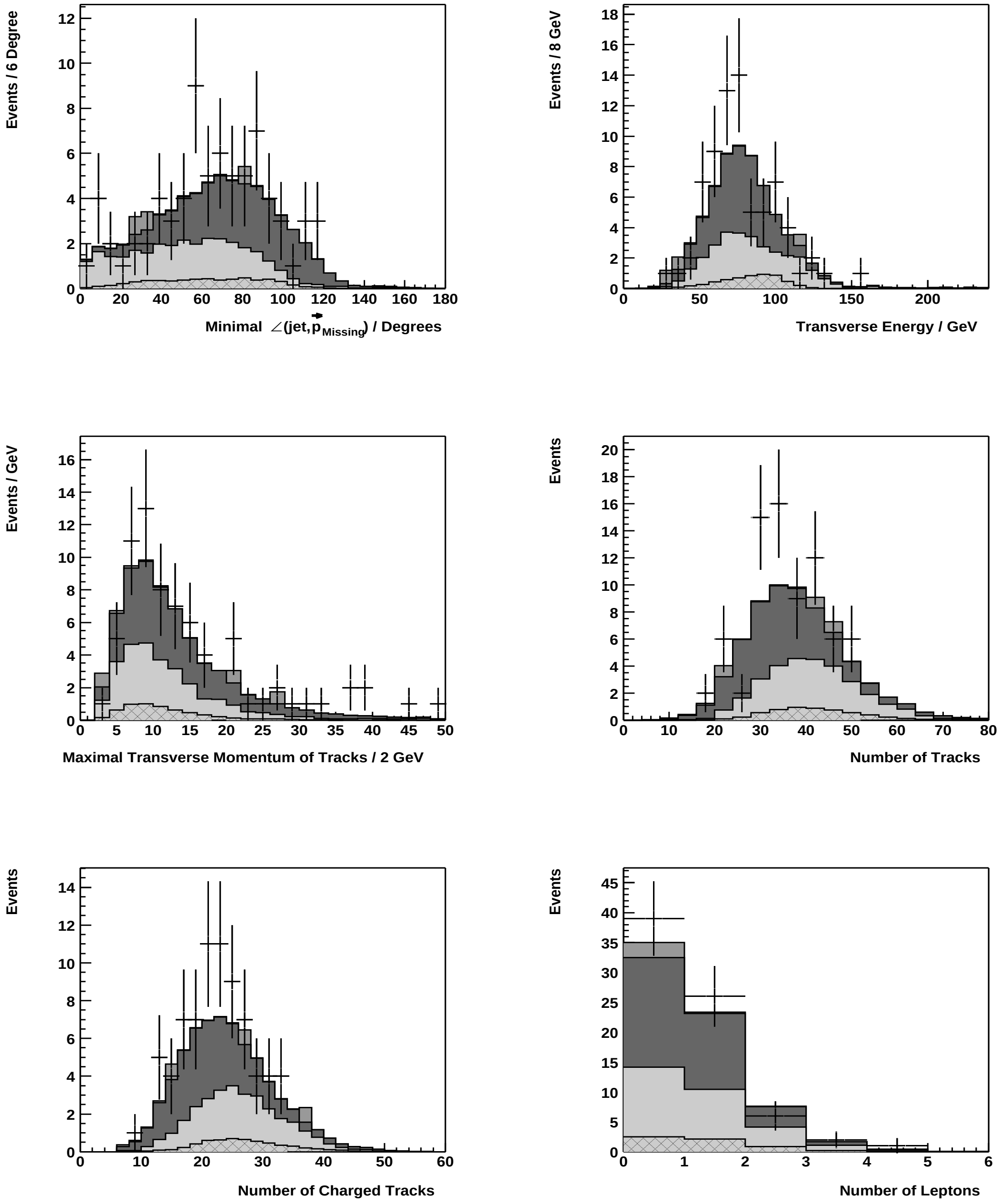


Figure 4.38: Simulated and measured distributions of U1 samples. The background is divided into three accumulated parts and corresponds to a luminosity of 56.31 pb^{-1}. The scaled signal has a mass of 115 GeV/c^2. Gray scales as in Figure 4.3.

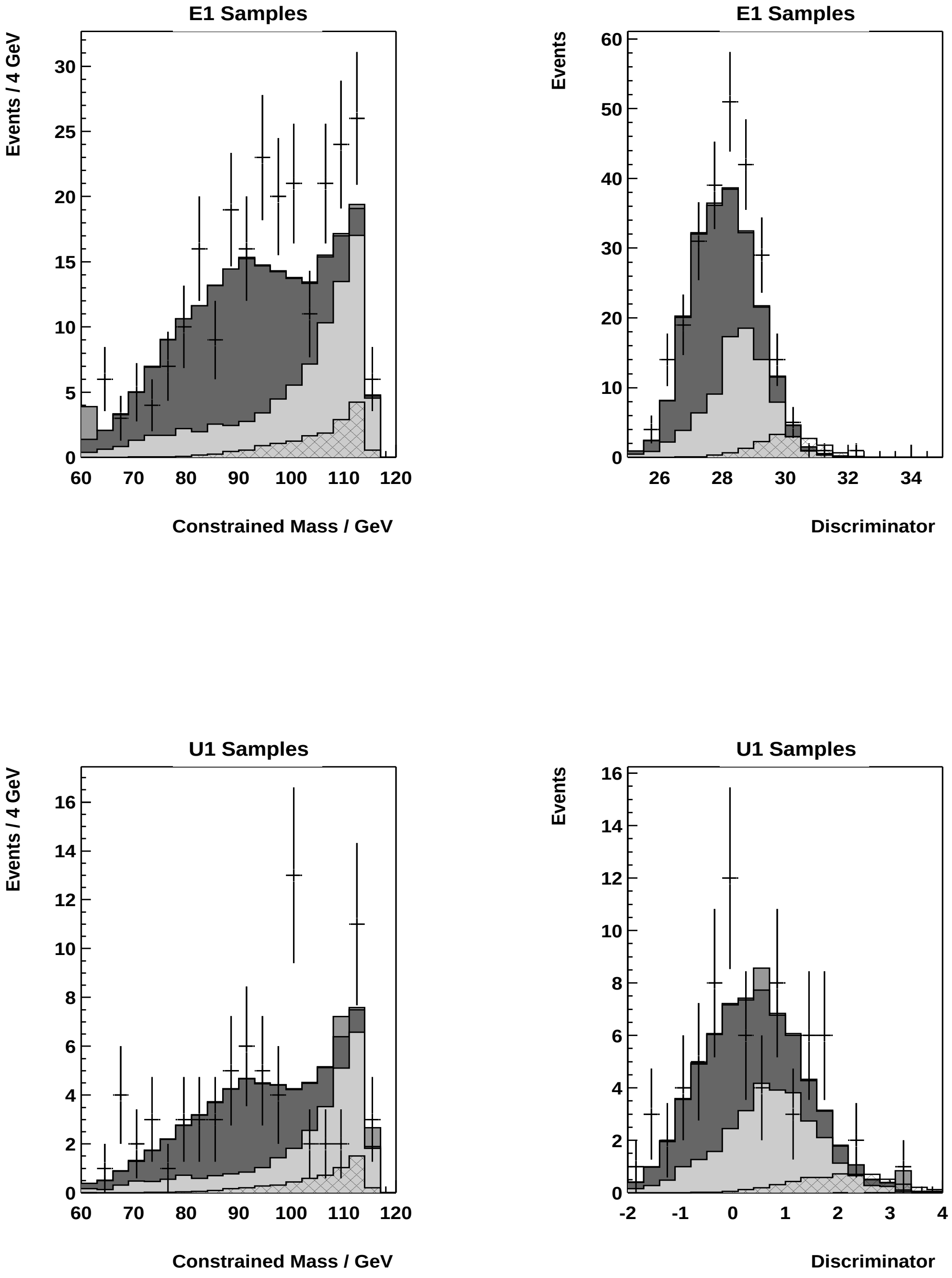


Figure 4.39: Simulated and measured distributions. The background is divided into three accumulated parts. The scaled signals have a mass of 115 GeV/c^2. Gray scales as in Figure 4.3.

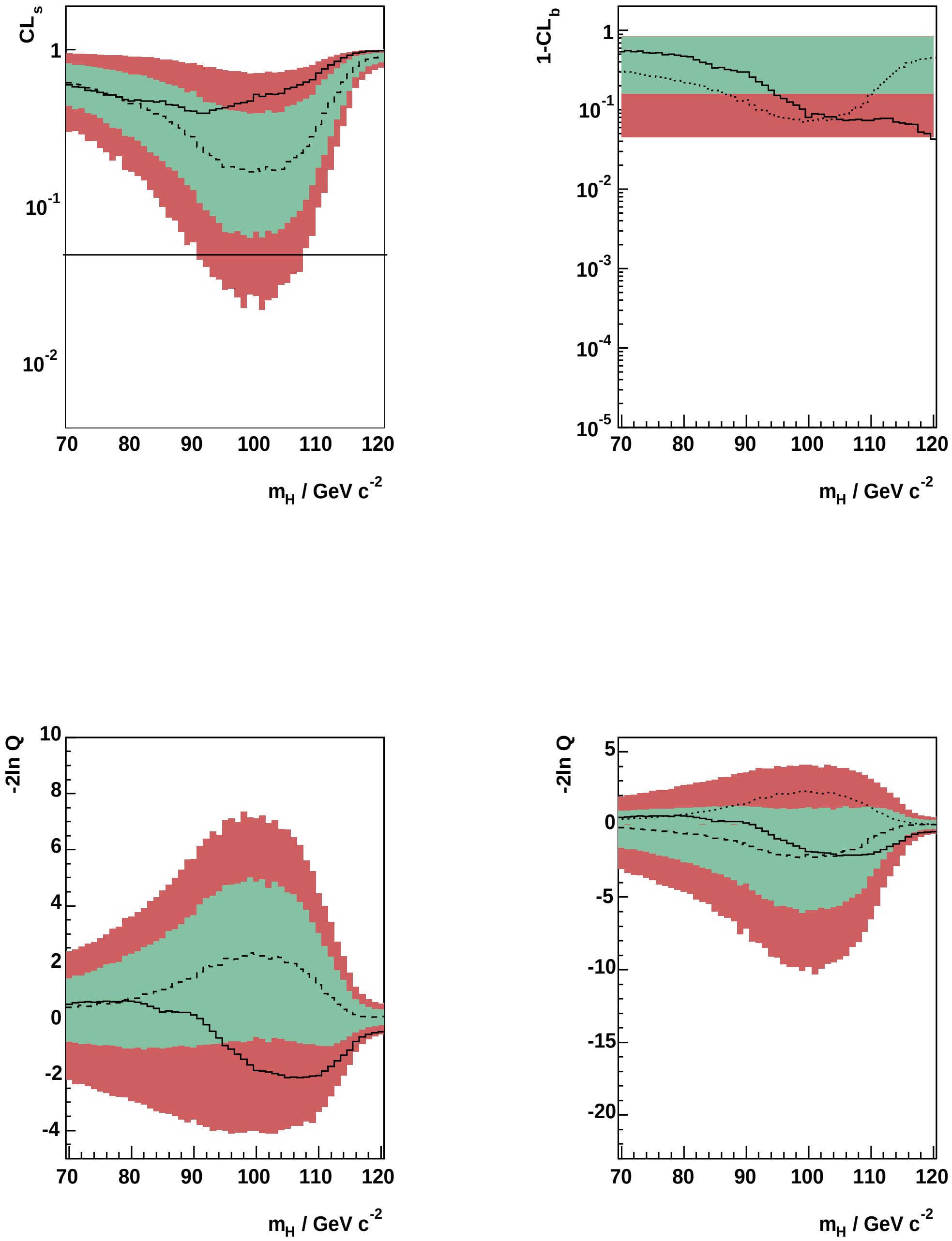


Figure 4.40: Results optimized for $m_H = 115$ GeV/c^2. The upper graphs show CL_s and $1-CL_b$ for Higgs mass hypotheses from 60 GeV/c^2 to 120 GeV/c^2. The bands correspond to one and two standard deviations around the dashed median expectation for the background. The observed results are solid. The dotted line in the $1-CL_b$ graph represents the median expectation for a signal at the test mass. The lower graphs show $-2\ln Q$ for the background only and the background plus signal hypotheses. The dotted line in the latter graph is the expected median for the background only hypothesis.

Table 4.11: Candidates found in a search optimized at $m_H = 115$ GeV/c^2. $Q = 1 + sS(\vec{x})/bB(\vec{x})$ is the test statistic as defined in Equation 2.21. The candidates are sorted with respect to $\ln Q_n$. The fifth and the sixth columns contain the cumulative ($\sum_{i=1}^{n} \ln Q_i / \sum_{All} \ln Q$) and the relative ($\ln Q_n / \sum_{All} \ln Q$) contributions. The last four columns show the constraint mass, the discriminator value, the acoplanarity and the b-tag value of the event.

n	Run	Event	$\ln Q$	Cum	Rel	m_{con}	d_{multi}	*Acop*	b_{tag}
1	113302	776	0.282	0.131	0.130	96.09	30.40	23.04	2.958
2	111298	78	0.266	0.254	0.122	112.45	30.12	1.50	2.398
3	114106	416	0.206	0.350	0.095	114.06	30.04	8.12	2.472
4	114721	2531	0.197	0.441	0.090	107.94	30.48	16.19	-0.749
5	115154	12300	0.151	0.512	0.069	100.40	0.31	29.17	-2.400
6	115035	12411	0.144	0.579	0.066	100.76	0.14	162.45	-1.821
7	109318	6969	0.114	0.632	0.052	68.73	30.48	4.77	7.594
8	117683	119	0.106	0.681	0.048	112.56	1.67	0.79	0.402
9	115053	7234	0.0807	0.718	0.037	111.22	1.73	4.67	0.733
10	111604	21087	0.080	0.756	0.036	111.68	29.87	2.36	1.801
11	117753	954	0.0763	0.791	0.035	114.38	0.97	1.14	-0.165
12	112173	1660	0.0701	0.824	0.032	101.22	30.62	1.85	5.347
13	115306	7000	0.0584	0.851	0.026	71.45	-0.12	118.43	-1.382
14	117317	20381	0.0317	0.865	0.014	99.83	0.78	45.62	0.137
15	115197	2571	0.0235	0.876	0.010	113.40	1.41	3.84	0.150

4.7 Uncertainty Studies

The deduced exclusion limits are based on the inputs which are needed by the test statistic defined in Equation 2.21. Apart from the measured and thus fixed number of events n_i, these are the simulated two dimensional probability density distributions, the efficiencies, i.e. the s_i, and the expected number of background events b_i for each channel i. The remaining task is therefore to identify and estimate the uncertainties in the signal efficiencies and the expected numbers of background events. These are then added in quadrature, and the square rooted sums are then used as uncertainty inputs for the ALRMC package.[11] The uncertainty sources are the following. [12]

- The **statistical uncertainties** are of binomial nature and computed as reported in Section 4.2.
- The systematic uncertainties in the **cross sections** used in the simulations of the various signal and background enter together with the **luminosity** uncertainties when the samples are normalized to obtain the s_i and the b_i. As argued in Section 4.2, a conservative systematic uncertainty of 1% is used for all cross sections and luminosities.
- The systematic **uncertainty in the distributions** used in the three steps of the discrimination are due to the limited modelling of the detector in the simulation set up.

Analyses of all three sources are presented sequentially. Since the best exclusion limit was obtained for an optimization at $m_H = 90$ GeV/c^2, the error analyses were done with this optimization.

In the final limits all uncertainties are included. The following subsections reports how the various sources decrease the limits obtained when the uncertainties are not taken into account.

Limited samples, cross sections and luminosities

The effect of statistical errors and the uncertainties due to the restricted precision in the knowledge of the cross sections and luminosities are summarized in Table 4.12. The chosen discrimination resulted in 6.56 and 4.13 backgrounds

---

[11] As usual the first order approximation yields

$$\sigma_w^2 = (\frac{\partial f}{\partial x})^2\sigma_x^2 + (\frac{\partial f}{\partial y})^2\sigma_y^2 + 2cov(x,y)\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}$$

for a function $w = f(x, y)$ of two variables. The partial derivatives are evaluated at the mean values.

[12] The question of over- and underestimation was treated and studied in Subsection 4.6.1.

Table 4.12: Uncertainties on the expected number of background events N_B and the signal efficiency ϵ_s.

	N_B	ΔN_B	ΔN_B^{stat}	$\Delta N_B^{L,\sigma}$	ΔN_B^{2F}	ΔN_B^{4F}	$\Delta N_B^{2\gamma}$	$\Delta\epsilon_s$
E1	6.557	0.094	0.020	0.092	0.017	0.005	0.010	0.01
U1	4.134	0.059	0.009	0.058	0.008	0.001	0.004	0.01

events for the E1 and U1 samples respectively. ΔN_B is the square rooted sum of ΔN_B^{stat} and $\Delta N^{L,\sigma}$. It is 1.4% of N_B for both the E1 and U1 samples. The signal efficiency uncertainty $\Delta\epsilon_s$ due to these sources is found to be 1%. Without taking these uncertainty sources into account, the limits $m_H^{exp} \geq 96.1$ GeV/c^2 and $m_H^{obs} \geq 98.5$ GeV/c^2 were obtained. Smearing the results corresponding to the uncertainties resulted in

$$m_H^{exp} \geq 95.9 \text{ GeV/c}^2$$

$$m_H^{obs} \geq 98.5 \text{ GeV/c}^2$$

With respect to the unsmeared results, this is a 0.2 GeV/c^2 decrease for the expected limit and no decrease for the observed limit.

Another uncertainty source of statistical character is the bin dependence in the density probability distributions $S_i(\vec{x})$ and $B_i(\vec{x})$. As described in Appendix A, a gaussian kernel smoothing procedure was introduced to control this source. The remaining uncertainty was estimated by computing the CL_s and the exclusion limits for different bin resolutions in the stable region. As seen in Figure A.1 and Figure A.2, the CL_s is stable for smoothed distributions with resolutions of 25×25 bins and upwards. According to the computations, which were performed with distributions with 25×25 bins to 35×35 bins, the remaining fluctuation corresponds to a limit variation of 100 MeV/c^2. To obtain conservative results, the limits are therfore decreased by further 0.1 GeV/c^2.

Systematic Uncertainties due to Finite Detector Modelling

The detector modelling is finite in two senses. There is a finite resolution and there is a finite geometry and computing precision in the simulations. This results in imperfect simulated probability density functions. The correct procedure to estimate the effect of these imperfections would be to sort out all uncertainties on the detector level and calculate how they propagate through the analysis. Since this is impossible in practice, different approximative methods are common.

Table 4.13: Systematic uncertainty estimates on the expected number of background events N_B. Signal efficiency for $m_H = 90$ GeV/c^2 was 0.42 for the E1 Processing (left columns) and 0.61 for the U1 processing (right columns).

E1	N_B	$\Delta N_B/N_B$	U1	N_B	$\Delta N_B/N_B$
	57.23			18.65	
E_{mip}	57.46	0.40		19.85	6.43
M_{vis}	55.59	2.87		20.22	8.42
Acop	55.89	2.34		19.10	2.41
E_{cha}	55.81	2.48		19.34	3.70
P_{thrust}	54.80	4.25		19.47	4.40
b-tag	55.65	2.76		19.19	2.90
Angjm	56.25	1.71		19.48	4.45
E_{tr}	57.29	0.11		19.10	2.41
P_{tr}^{max}	56.29	1.64		19.11	2.47
N_{tracks}	57.23	0.0		18.65	0.0
N_{cha}	57.23	0.0		18.65	0.0
N_{lep}	57.23	0.0		18.65	0.0

Here the bin content of the simulated distributions were rescaled to the bin content of the measured distributions. By repeating the discrimination with one rescaled distribution in turn, twelve deviations in the background expectation were obtained for each processing. In Table 4.13 the relative deviations are listed. The first row contains the results when no rescaling is applied. The other rows show the effect of rescaling for each observable. Especially the columns $\Delta N_B/N_B$ show the percentual deviations from the no-rescaling case.[13] The largest deviation was then taken as the estimation of the systematic uncertainty.

When the search results are smeared according to the systematic uncertainty the limits

$$m_H^{exp} \geq 95.9 \text{ GeV/c}^2$$

$$m_H^{obs} \geq 98.5 \text{ GeV/c}^2$$

are obtained. With respect to the unsmeared results, this is again a 0.2 GeV/c^2 decrease for the expected limit and no decrease for the observed limit.[14] When

---

[13] These computing-intensive studies were performed by using a local Condor computation pool.

[14] A 10 % uncertainty on the expected background for both samples reduces the limits by 0.5 GeV/c^2.

the systematic uncertainties due to finite detector modelling is added in quadrature with the uncertainties treated in the previous subsection and the sum square rooted, 4.5% and 8.5% relative errors are obtained for the number of expected background. The first number refers to the E1 and the second to the U1 samples. On the efficiencies 2% and 4% are obtained. With this total uncertainties final limits are computed to be, when the influence of the density probability distributions are taken into account,

$$m_H^{exp} \geq 95.8 \text{ GeV/c}^2$$
$$m_H^{obs} \geq 98.4 \text{ GeV/c}^2$$

4.8 Summary and Best Limits

This chapter summarized a search for the Standard Model process $e^-e^- \to H\nu\bar{\nu}$. The data measured by the DELPHI detector in year 2000 was analyzed. Signal and background samples, simulated at a center of mass energy equal to 206.5 GeV, served for discriminator and probability density function estimations.

In addition to a standardized track selection commonly used by the DELPHI collaboration, one univariate cut was applied to reduce the background and achieve an acceptable agreement between measured and simulated samples (see Section 4.4). At this level it was aimed at an uniform effect on the signal efficiencies. The result was a highly preselection-independent analysis.

The samples resulting from the univariate discrimination, were then used to estimate third order polynomials. Based on studies of the observable distributions, a subset of twelve were selected. Discriminators optimized for four different signal mass hypotheses, $m_H = 80$ GeV/c^2, $m_H = 90$ GeV/c^2, $m_H = 100$ GeV/c^2 and $m_H = 115$ GeV/c^2, were calculated in two iterations. Cuts on the first iteration discriminators resulted in better discriminating power for the second iteration discriminators. A comparison at $m_H = 90$ GeV/c^2 and $m_H = 115$ GeV/c^2 showed that the discriminating power of this procedure was comparable or even better than the one applied in the published DELPHI search for $e^-e^- \to H\nu\bar{\nu}$.

At all four optimization masses searches based on the CL_s method and weighted counting were performed. Two dimensional Monte Carlo distributions, spanned by the second iteration discriminator and the Higgs boson mass estimator, enabled the weighted counting. The distributions were smoothed with a Gaussian kernel method. The results are summarized in Figure 4.41. The highest exclusion limits were achieved with an optimization at $m_H = 90$ GeV/c^2.

$$m_H^{exp} \geq 95.8 \text{ GeV/c}^2$$

$$m_H^{obs} \geq 98.4\,\mathrm{GeV/c^2}$$

The results in Figure 4.41 show that optimizations at higher masses do enlarge the exclusion potential, i.e. the expected CL_s is lower at high masses. For the $e^+e^- \rightarrow H\nu\bar{\nu}$ channel alone, this does not lead to a 95% exclusion limit. However, in combination with other channels, an optimization at a higher mass should be used.

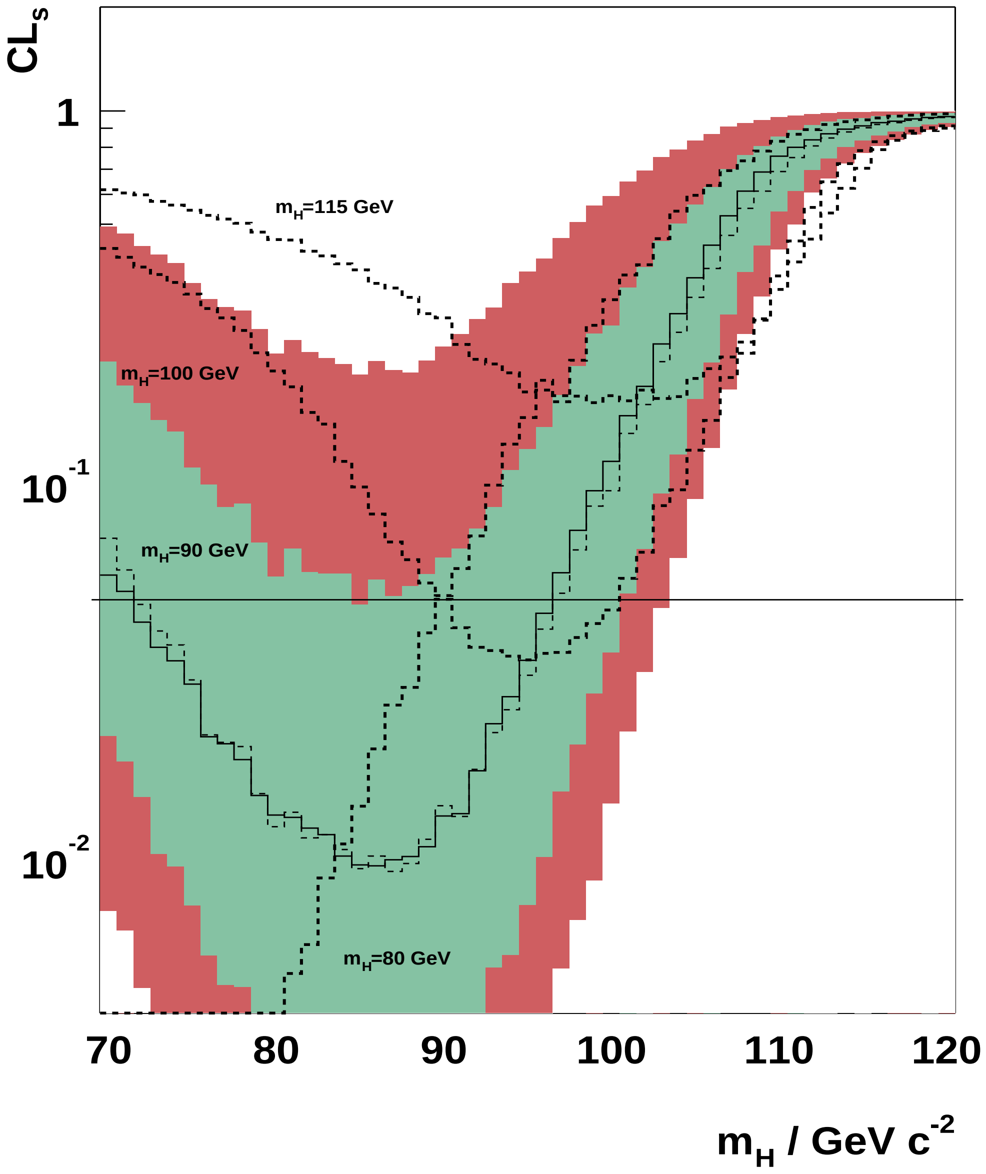


Figure 4.41: CL_s summarized for all four optimization masses. The dashed median expectation for the background for a search optimized at m_H = 90 GeV/c^2 is surrounded by the one and two standard deviation bands. The observed result is shown by the solid curve. Other dashed curves are median expectations for the background marked according to the respective optimization mass.

Chapter 5

Concluding Summary

A search for elementary particles in high energy physics is practically described in terms of discrimination and hypothesis tests. Various approaches within both subfields have been surveyed, and polynomial discriminators and weighted counting in combination with the CL_s method have been used to search for the Standard Model process $e^+e^- \rightarrow H\nu\bar{\nu}$. All high energy data taken by the DELPHI detector in year 2000, the last year of LEP operation, was searched. That is, the highest center of mass energies ever reached in e^+e^- collisions have been analyzed.

There exists an open set of different discriminators. However, all discriminators may be categorized in three classes. Discriminators may be estimated with joint probability functions, with conditional probability functions or by using the data samples directly. Polynomial discriminators belong to the latter class. Polynomial discriminators of first and second order are heavily used and equivalent to gaussian discriminators with equal and unequal covariance matrices. If they are seen in feature space, they represent either a subdividing hyperplane or a subdividing second order hypersurface. Higher orders are rare, but although not proved, in general one may assume that higher order surfaces in feature space do discriminate more efficiently. This has been demonstrated to be true in the discrimination of the background in this search for $e^+e^- \rightarrow H\nu\bar{\nu}$.

The discrimination consisted of one univariate cut and an iterative use of a multivariate method which trained polynomial discriminators. The univariate cut ensured a desired absolute agreement between the simulated and measured samples. The multivariate method estimated the discriminator directly by using a much used criterion : The distance between the means of the hypothesis samples is required to be as large as possible, and the covariance within the samples as small as possible. The available data allowed the estimation of third order polynomials. This was done for four different optimization masses of the Higgs boson, an automatic task of computation since the estimation did not depend on mass-specific univariate cuts. Two iterations were applied in the discrimination. With this procedure a comparable or better discriminating power than those achieved by independent discriminations on the same data, was achieved. This shows that different methods of similar complexity often have comparable capability to exhaust the inherent discrimination potential in the samples.

A hypothesis test may be performed by using either joint or conditional probability density functions. Both approaches need an additional assumption to avoid unphysical results in cases of small signals. The modified frequentist approach is based on joint probability density functions and applies a physical motivated assumption : The p-value of the signal plus background hypothesis is normalized by the p-value of the background only hypothesis. The resulting ratio is used as a test statistic. In the search performed, two dimensional density functions were used in combination with the modified frequentist definition. The best exclusion limits were obtained at an optimization at $m_H = 90$ GeV/c^2.

$$m_H^{exp} \geq 95.8 \text{ GeV/c}^2$$

$$m_H^{obs} \geq 98.4 \text{ GeV/c}^2$$

The uncertainties are included and have reduced the limits by around 0.4 GeV/c^2. The limits are obtained from a search in the $e^+e^- \rightarrow H\nu\bar{\nu}$ alone. A search in the DELPHI data which combines all significant production channels, results in an observed limit for the Standard Model Higgs boson mass equal to 114.3 GeV/c^2 [58]. The limit deduced from the combination of all LEP data is published to be 114.4 GeV/c^2 [58].

As mentioned in the introduction, the experiments at Tevatron may increase the observed exclusion limit. The Tevatron's own 95% exclusion potential is around 180 GeV/c^2. A discovery beyond the LEP exclusion limit is not expected. The answer to the question of the existence of the Higgs boson and the mechanism behind electrowak symmetry breaking is therefore probably left to the LHC experiments to give. The LHC will start its anounced operation in 2007.

Concerning the multivariate methods, the enormous data sets to come and the huge parameter spaces in extended models will certainly motivate further use of automatized multivariate methods on computer grids of planetary scale. The highly speculative character of many of the extended models will also certainly stimulate the use of weighted event counting, the CL_s and similar methods in excluding model parameters.

Appendix A

Density Estimation and Smoothing

According to Equation 2.21, probability density functions $S_{E1}(\vec{x})$, $S_{U1}(\vec{x})$, $B_{E1}(\vec{x})$ and $B_{U1}(\vec{x})$ are used in the computation of the test statistic CL_s. Since these are not analytically known, simulated distributions may be used. In this work simulated two dimensional, i.e. $dim(\vec{x}) = 2$, distributions which contained the Higgs boson mass estimator M_{con} and the multivariate discriminator were employed. However, since these distributions are histograms, the results depend on the chosen bin resolution. To overcome this dependency, a well-tested smoothing procedure was applied.

The effect of the smoothing procedure can be seen in Figure A.1 and Figure A.2. Simulated data sets discriminated to 69% and 68% signal efficiencies were used for the E1 and U1 samples respectively. The optimization of the discriminator was done for a Higgs boson mass hypothesis of 100 GeV/c^2. With the nomenclature which was established in Equation 2.21, this corresponded to $s_{E1} = 7.86$, $b_{E1} = 103.8$, $s_{U1} = 2.79$ and $b_{U1} = 33.7$. For the purpose of testing, a normalized and unbiased subset of the background samples was used to represent a signal-free experiment. By computing the test statistics $1 - CL_b$ and CL_s for probability density distributions, $S_{E1}(\vec{x})$, $S_{U1}(\vec{x})$, $B_{E1}(\vec{x})$ and $B_{U1}(\vec{x})$, with different resolutions, i.e. different number of bins, the drawbacks of histogram estimations became apparent.

In the first row of Figure A.1 and Figure A.2 the bin dependence is illustrated. The shaded bands represent one and two standard deviations around the expected medians and the solid lines the simulated "observations". The $1 - CL_b$ graph shows a background-like "observation" which wiggles around 0.5. The same variation is seen in the CL_s graph. It reflects the fact that the simulated background candidates are fixed while the density resolution varies. The following uncertainty in the resulting CL_s and thus in an exclusion or in a discovery potential is one drawback. The other and more serious is the increasing sensitivity. As seen in the CL_s graph, a higher resolution enables a better discrimination between signal and background and therefore a lower CL_s. However, at some resolution this increased sensitivity is artificial and only reflects fluctuations due to insufficient statistics. These bin dependencies must be avoided.

Since the bin dependencies are consequences of the bin fluctuations in the histograms, improved approaches to density estimations involve some kind of smoothing of the original histograms. Two commonly used methods for such smoothing are kernel estimation and spline interpolation.[1] The former has been used frequently in the Higgs boson searches at LEP. It is based on a gaussian kernel smoothing where each bin in the original histogram is considered as a delta function which then is smeared with a gaussian distribution. Further exponential tails and a reweighting with the smoothed projections of the two dimensional histograms have been used. A study of this approach is described in Reference [52]. A comparison of this techniueqe with a smoothing based on polynomial b-splines [53] resulted in the use of the kernel approach.

The CL_s based on kernel smoothed distributions are shown in the lower rows of Figure A.1 and Figure A.2 for the E1 samples and the U1 samples.[2] The fluctuation is reduced and the decrease in sensitivity is now avoided. All confidence computations in this work were done with kernel smoothed distributions of 30×30 bins.[3] On the mass estimator axis this corresponded to 3.33 GeV/c^2 per bin. On the discriminator axis the resolution varied according to the different optimizations, i.e. different discriminator distributions.

---

[1] There are other techniques. See e.g. Reference [54] for a neural network approximation of histograms.

[2] A smoothing parameter equal to 0.2 was applied.

[3] A FORTRAN code called FUZEXP and written by W.Murray, 18/04/2000, was adapted. The actual smoothing parameter was set to 0.2.

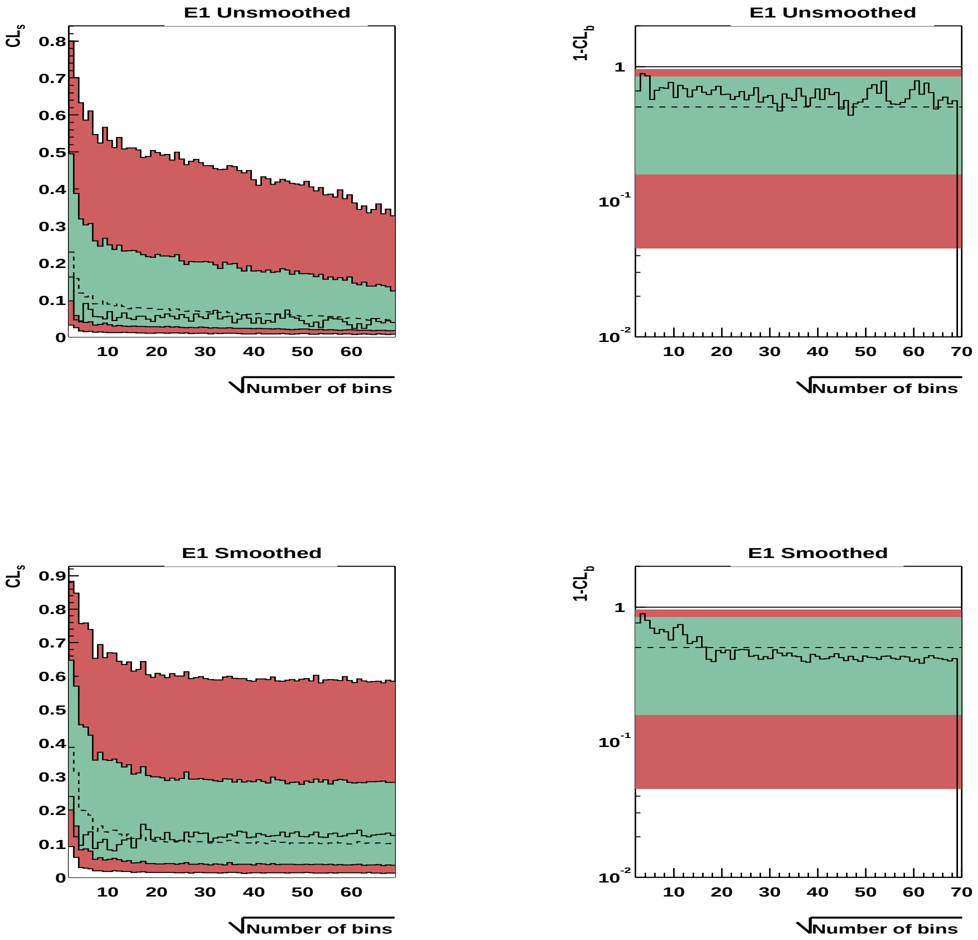


Figure A.1: CL_s and $1 - CL_b$ versus the square root of bins in the two dimensional probability density distributions $S_{E1}(\vec{x})$ and $B_{E1}(\vec{x})$. The upper row shows results calculated with unsmoothed distributions for a simulated background experiment. The mass hypothesis was $m_H = 100$ GeV/c^2. The lower row shows the corresponding results based on smoothed distributions.

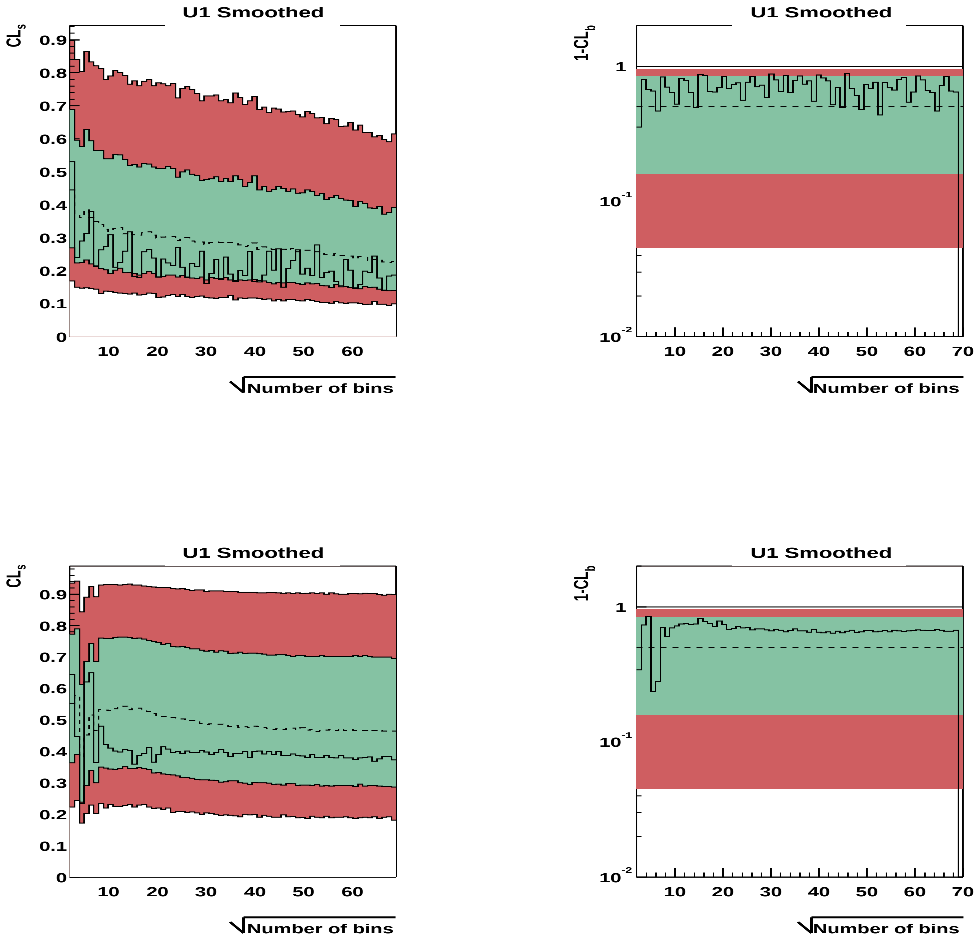


Figure A.2: CL_s and $1 - CL_b$ versus the square root of bins in the two dimensional probability density distributions $S_{U1}(\vec{x})$ and $B_{U1}(\vec{x})$. The upper row shows results calculated with unsmoothed distributions for a simulated background experiment. The mass hypothesis was $m_H = 100$ GeV/c^2. The lower row shows the corresponding results based on smoothed distributions.

Appendix B

Contribution to the Lake Louise Winter Institute 2003 "Particles and Universe"

This appendix includes the proceeding of a contributed talk on the combined results of Higgs searches at LEP. It is published by *World Scientific* in the conference proceedings from 2003.

LEP LIMITS ON HIGGS BOSON MASSES IN THE SM, IN THE MSSM AND IN GENERAL 2HD MODELS

SIGVE HAUG

Department of Physics, University of Oslo
P.O. Box 1048 Blindern, 0316 Oslo, Norway
E-mail: sigve.haug@fys.uio.no

Before shutting down in 2000 the four LEP experiments ALEPH, DELPHI, L3 and OPAL collected a total of 2461 pb^{-1} of data from electron positron collisions at centre of mass energies between 189 GeV and 209 GeV. Combining this data the LEP Higgs Working Group has deduced lower limits for Higgs boson masses at the 95% confidence level. For the Standard Model Higgs boson the limit is at 114.4 $\mathrm{GeV/c^2}$. Limits for the light CP-even and the CP-odd neutral Higgs bosons of the Minimal Supersymmetric Standard Model are also reviewed together with the mass limit on charged Higgs bosons obtained from searches in general two Higgs doublet models at LEP.

1. Introduction

The Standard Model (SM) contains one complex Higgs doublet which accommodates electroweak symmetry breaking and fermion masses. The doublet leads to one physical Higgs boson H^0. Extensions of the SM with two complex Higgs doublets (2HD) have five physical Higgs bosons : a charged pair $H^{\pm}$, one neutral CP-odd scalar A^0 and two neutral CP-even scalars, h^0 and H^0. In the Minimal Supersymmetric Standard Model (MSSM) a 2HD is required.

At the Large Electron Positron Collider (LEP) the four experiments ALEPH, DELPHI, L3 and OPAL collected 2461 pb^{-1} of data from electron positron collisions at centre of mass energies between 189 GeV and 209 GeV. On this basis the LEP Higgs Working Group has deduced lower limits on the Higgs boson masses using the modified frequentist method. In this method a signal plus background hypothesis is considered as excluded at a 95% confidence level when the statistic CL_s, defined as the ratio of the p-value of the signal plus background hypothesis to the p-value of the background only hypothesis, is smaller than or equal to 5%. [1,2]

The search results for neutral Higgs bosons within the SM and the

MSSM are reviewed here. Also the result from the search for charged Higgs bosons within a general 2HD model is considered. The numbers presented are the latest. Final results are expected from the LEP Higgs Working Group by the end of 2003. Further searches for Higgs bosons within other extensions of the SM are reported elsewhere.

2. Limit on the SM Higgs boson mass

The contribution to the SM Higgs production at LEP was dominated by the Higgs Strahlung process $e^+e^- \to H^0Z^0$. In the relevant search channels the WW fusion $e^+e^- \to H\bar{\nu}_e\nu_e$ became significant at the kinematical limit due to the Z^0 mass. The ZZ fusion remained subdominant. The searches were structured in channels based on the relative fractions of the various decay modes of the Z^0 boson. Due to the Higgs boson's prefered decays into b-quark pairs, b-tagging has been essential in the analyses.

The two upper plots in Figure 1 are instructive plots of the probability density functions (pdf) of the constructed observable $-2ln(Q)$ where Q is the likelihood ratio of the two alternate hypotheses. [3] The dashed line is the pdf for the signal plus background hypothesis and the solid for the background only hypothesis. The vertical solid line indicates the observed value. To the left a Higgs mass hypothesis $m_{H^0} = 110$ GeV/c^2 is assumed, to the right $m_{H^0} = 116$ GeV/c^2. The separation between the hypotheses decreases with increasing mass according to the falling cross section. The shaded area on the right side of the observed line is now the p-value for the signal plus background hypothesis. The left lower plot shows the CL_s for Higgs mass hypotheses from 100 GeV/c^2 up to 120 GeV/c^2. In the region where the CL_s is below the horizontal 5% line, the hypothesis is considered excluded. The expected CL_s leads to a limit at 115.3 GeV/c^2. The observed CL_s yields a limit of 114.4 GeV/c^2.

The lower right plot shows the comparability of the data with background only hypotheses, the $1 - CL_b$, in the Higgs mass range from 100 GeV/c^2 to 120 GeV/c^2. The observed solid line lies within the two sigma band in the whole range. The largest deviations appear in the vicinities around 100 GeV/c^2 and 116 GeV/c^2. The dashed-dotted line indicates the $1 - CL_b$ for the signal plus background hypotheses.

3. Limits on the neutral MSSM Higgs boson masses

In the MSSM search a constrained model with a CP conserving Higgs sector and seven parameters, M_{SUSY}, M_2, μ, A, $tan\beta$, m_A and $m_{\tilde{g}}$, is used.

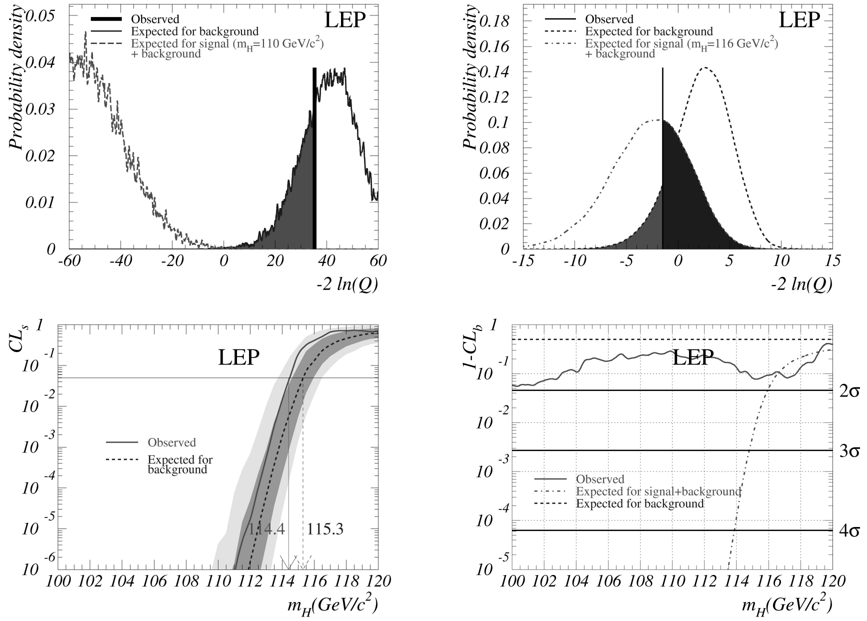


Figure 1. The upper plots show the probability density functions of the observable $-2ln(Q)$ at two different test masses. The lower plots show the CL_s and the $1-CL_b$ for Higgs mass hypotheses from 100 GeV to 120 GeV. The bands in the CL_s plot correspond to one and two standard deviations.

Three benchmark scenarios are examined. The *Max* m_{h^0} scenario yields a maximum mass for the lightest neutral Higgs boson and leads to a conservative mass limit. The *No Mixing* scenario corresponds to the *Max* m_{h^0}, but without mixing in the stop sector, enabling the study of the mixing influence. The *Large* μ scenario ($\mu = 1$ TeV) is designed to illustrate regions in the parameter space where h^0 does not decay into pairs of b-quarks. [4] In addition to the SM processes the MSSM contains the associated pair production $e^+e^- \to A^0h^0$ for the lightest Higgs boson h^0.

In Figure 2 scans in the $tan\beta$ versus m_{h^0/A^0} plane are shown for the *Max* m_{h^0} and the *No Mixing* scenarios. To the left one can see that the lightest CP-even Higgs boson mass is excluded up to 91.0 GeV/c^2 for *Max* m_{h^0} and 91.5 GeV/c^2 for *No Mixing.* The CP-odd Higgs boson is excluded up to 91.9 GeV/c^2 and 92.2 GeV/c^2 respectively. The right plots show that $tan\beta$ is excluded from 0.5 till 2.4 in the *Max* m_{h^0} scenario and from 7.0 till 10.5 in the *No Mixing* scenario.

The *Large* μ scenario is excluded entirely and thus not shown.

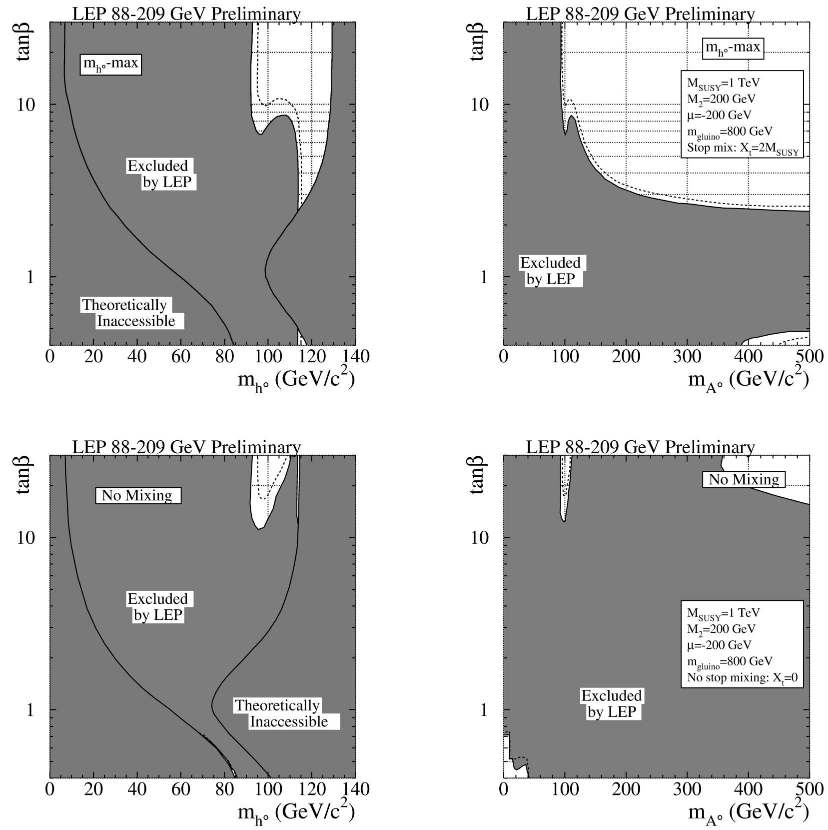


Figure 2. Exclusion scans in the $tan\beta$ versus m_{h^0} plane and in the $tan\beta$ versus m_{A^0} plane for the *Max* m_{h^0} and the *No Mixing* scenarios. The dotted lines indicate the expected exclusion contours. The exclusion level is 95%.

4. Limit on the charged 2HD model Higgs boson mass

In a general two Higgs doublet extension of the SM, the charged Higgs bosons are produced via Z or γ exchange in the process $e^+e^- \to H^+H^-$. The most important decay channels are $(c\bar{s})(\bar{c}s)$, $(\tau^+\nu_\tau)(\tau^-\bar{\nu}_\tau)$ and $(\bar{c}s)(\tau^+\nu_\tau) + (c\bar{s})(\tau^-\bar{\nu}_\tau)$.[5]

In Figure 3 the upper plots show the $1 - CL_b$ and the difference between data and expected background for the L3 experiment only. The case where $Br(H^+ \to \tau^+\nu_\tau) = 0$, is shown. In the vicinity of 68 GeV/c^2 the observed $1 - CL_b$ is more than three standard deviations away from the

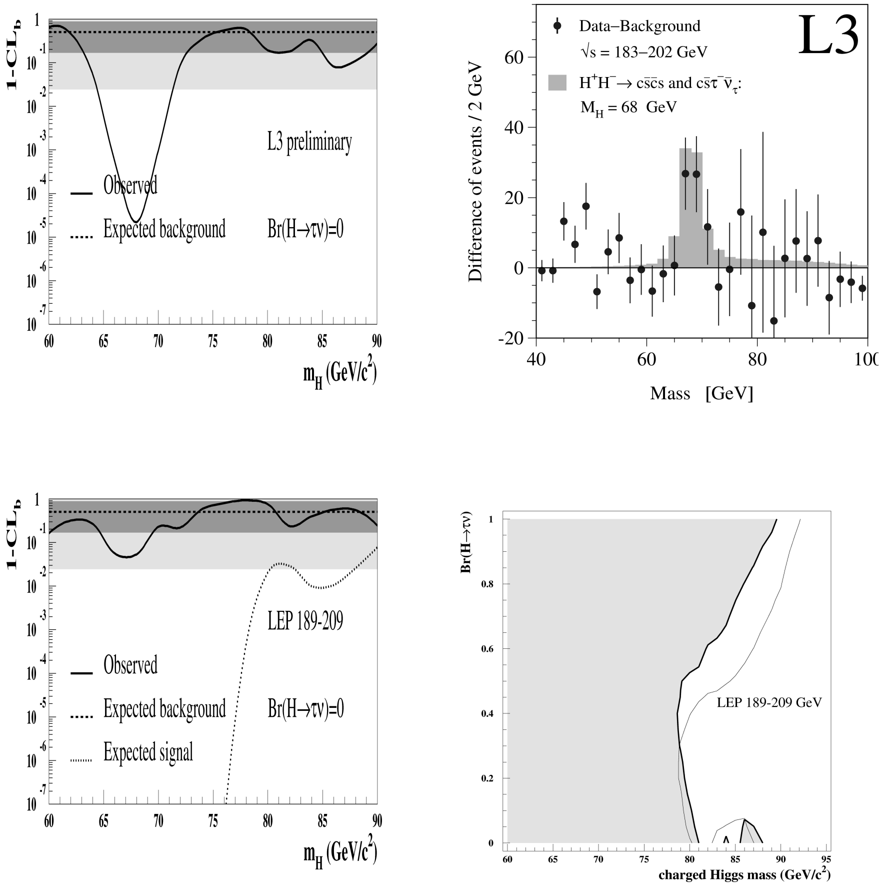


Figure 3. The two upper plots are for L3 alone. The left shows the $1-CL_b$ scan in the non leptonic decay channel. The bands correspond to one and two standard deviations. The mass histogram contains the difference between data and background. The lower plots show combined results from all four experiments. To the left the CL_s scan is plotted, to the right the exclusion scan in the $Br(H^+ \to \tau\nu_\tau)$ versus $m_{H^\pm}$ plane. The exclusion level is 95%.

expected background only hypotheses. The histogram to the right shows the excess responsible for this deviation. However, the other experiments do not confirm this interesting observation. Combined for all four experiments the results are shown in the lower left $1-CL_b$ scan. The lower right exclusion scan shows for which $m_{H^\pm}$ the signal plus background hypotheses are excluded as a function of the $Br(H^+ \to \tau^+\nu_\tau)$. A branching ratio independent lower exclusion limit is obtained at $m_{H^\pm} = 78.6$ GeV/c^2. The exclusion level is 95%. [6]

5. Summary

The LEP Higgs Working Group has calculated lower mass limits on Higgs bosons based on the combined data from the four LEP experiments. The limits for the SM Higgs boson, the neutral MSSM Higgs bosons and the charged Higgs bosons within a general two Higgs doublet extension of the SM, are summarised in Table 1. The results are preliminary and final numbers are expected to appear by the end of 2003. LEP limits on Higgs bosons in other models and scenarios are reported elsewhere.

Table 1. Expected (median) and observed 95% LEP limits on SM, MSSM and 2HD Higgs bosons.

	Expected limit / GeVc^{-2}	Observed limit / GeVc^{-2}	Model
m_{H^0}	115.3	114.4	SM
m_{h^0}	94.6	91.0	MSSM *Max* m_{h^0}
m_{A^0}	95.0	91.9	
tan β	$(0.5, 2.6)$	$(0.5, 2.4)$	
m_{h^0}	95.0	91.5	MSSM *No mixing*
m_{A^0}	95.3	92.2	
tan β	$(0.8, 16.0)$	$(0.7, 10.5)$	
$m_{H^\pm}$	-	78.6	2HD

References

1. A. L. Read, *Modified Frequentist Analysis of Search Results*, in F. James, L. Lyons and Y. Perrin (eds.), *Workshop on Confidence Limits*, CERN Yellow Report 2000-05, available through weblib.cern.ch.
2. K. Hagiwara *et al.*, Phys. Rev. **D66**, 010001 (2002).
3. LEP Higgs Working Group, *Search for the Standard Model Higgs Boson at LEP*, LHWG Note 2002-01.
4. LEP Higgs Working Group, *Searches for the Neutral Higgs Bosons of the MSSM*, LHWG Note 2001-04.
5. A. Djouadi, J. Kalinowsky and P. M. Zerwas, Z. Phys. **C57**, (1993) 569.
6. LEP Higgs Working Group, *Search for Charged Higgs Bosons*, LHWG Note 2001-05.

Bibliography

[1] LEP Higgs Working Group, *Searches for Higgs Bosons*, ALEPH 2000-074 CONF 2000-051, DELPHI 2000-148 CONF 447, L3 Note 2600, OPAL Technical Note TN661.

[2] http://www-bdnew.fnal.gov/pbar/run2b/default.htm.

[3] http://pdg.lbl.gov.

[4] K. Hagiwara *et al.*, Phys. Rev. **D66**, 010001 (2002).

[5] DELPHI Collaboration, *Final results from DELPHI on the searches for SM and MSSM Neutral Higgs Bosons*, submitted to Eur. Phys. J. C., Feb. 2003.

[6] LEP Higgs Working Group, *Search for the Standard Model Higgs Boson at LEP*, LHWG Note 2002-01.

[7] N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, Phys. Lett. **B429**, 263 (1998).

[8] L. Randall and R. Sundrum, Phys. Rev. Lett. **83**, 3370 (1999).

[9] Proceedings of Workshop on Confidence Limits, F. James, L. Lyons and Y. Perrin [Eds.], CERN 2000-005, p. 82.

[10] B. D. Ripley, *Pattern Recognition and Neural Networks*, p. 20, Cambridge University Press, 1996.

[11] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Vol. 1, p. 198, Charles Griffin & Company Ltd, 1968.

[12] — *ibid.*, Vol. 3, p. 314.

[13] — *ibid.*, Vol. 3, p. 315.

[14] — *ibid.*, Vol. 3, p. 322.

[15] D. Jeans, *A Discussion of Fiscriminant Techniques*, DELPHI 2001-135 TRACK 97 (2001).

[16] C. Mariotti and E. Piotto, *Search for neutral Higgs Bosons in* $e^+e^- \rightarrow H\nu\bar{\nu}$ *Channel at* $\sqrt{s} = 189\ GeV$, DELPHI 99-154 PHYS 832 (1999).

[17] G. J. Feldman and R. D. Cousins, Phys. Rev. **D57**, 3873 (1998).

[18] A. L. Read, *Modified frequentist analysis of search results*, in Proceedings of Workshop on Confidence Limits, F. James, L. Lyons and Y. Perrin [Eds.], CERN 2000-005, p. 81.

[19] A. L. Read, *Optimal Statistical Analysis of Search Results based on the Likelihood Ratio and its Application to the Search for the MSM Higgs Boson at* $\sqrt{s}$ *= 161 and 172 GeV*, DELPHI 97-158 PHYS 737 (1997).

[20] J. L. Vinningland, *MSSSM Higgs Mass Limit Studies Using the DELPHI Detector*, Diploma Thesis, Dep. of Physics University of Oslo (2000).

[21] J. Gunion *et. al.*, *The Higgs Hunter's Guide*, Addison-Wesley Publishing Company, 1990.

[22] D. Bardin, *Field Theory and the Standard Model*, in Proceedings of 1999 European School of High-Energy Physics, A. Olchevski [Ed.], CERN 2000-007, p. 5.

[23] DELPHI Collaboration, *Performance of the DELPHI Detector*, Nucl. Instr. Meth. **A378** (1996) 57.

[24] S. Glashow, *Nucl. Phys.* **22** (1961) 579.
S. Weinberg, *Phys. Rev. Lett.* **19** (1967) 1264.
S. Glashow, J. Iliopoulos, L. Maiani, *Phys. Rev* **D2** (1970) 1285.

[25] P. W. Higgs, *Phys. Lett.* **12** (1964) 132. F Englert, R. Brout, *Phys. Rev. Lett.* **13** (1964) 321. G. Guralnik, C. Hagen, T. Kibble *Phys. Rev. Lett.* **13** (1964) 585. P. W. Higgs, *Phys. Rev.* **145** (1966) 1156. T. Kibble, *Phys. Rev.* **155** (1967) 1554.

[26] T. L. Barklow *et al.*, *Strong Coupling and Electroweak Symmetry Breaking*, Working Group Summary Report from the 1996 DPF-DPB Summer Study, Snowmass, Colorado, June 25 - July 12, 1996, SLAC-PUB-7397, CLNS 97/1473 (1997).

[27] J. R. Espinosa and M. Quiros, Phys. Lett. **B353**, (1995) 257-266.

[28] The LEP Electroweak Working Group, http://lepewwg.web.cern.ch/LEPEWWG/.

[29] G. Altarelli, T. Sjostrand and F. Zwirner (ed.), *Cern Yellow Reports : Physics at LEP2*, p. 361, CERN96-01 (1996).

[30] W. Kilian et al., *Higgs strahlung and WW fusion in $e+e-$ collisions* , DESY95-216 (1995).

[31] F. A. Berends, W. L. van Neerven and G.J. Burges, Nucl Phys. **B297** (1998), 429; erratum ibid. **B304** (1988), 921.

[32] E. Gross, B. Kniehl and G. Wolf, Z. Phys. **C63** (1994), 417.

[33] B. Kniehl, Z. Nucl. Phys. Proc. Suppl. **37B** (1994) 268-273.

[34] W. Keung and W. Marciano, Phys. Rev. **D20**, (1984) 248.

[35] A. Vainshtein *et. al.*, Sov. J. Nucl. Phys. **30**, (1979) 711.

[36] U. Fuskeland, *Simulation of a Search for the Standard Model Higgs Boson in the $H \to \gamma\gamma$ channel at LHC/ATLAS*, Diploma Thesis, Univ. of Oslo, 2002.

[37] DELPHI Collaboration, *The DELPHI Detector at LEP*, Nucl. Instr. Meth. **A303** (1991) 233.

[38] T. Fragat, *The inteference term between higgs strahlung and WW fusion in the HZHA generator* , DELPHI99-161 PHYS 836 (1999).

[39] P. Janot, CERN Report 96-01, Vol. 2, p. 309 (1996); Version 3, see http://alephwww.cern.ch/$\tilde{j}$anot/Generators.html.

[40] DELPHI, *DELSIM - DELPHI Event Generation and Detector Simulation User's guide*, 89-67 Prog 142 edition, 1989.

[41] S. Catani *et al.*, Phys. Lett. **B269** (1991) 432.

[42] S. Jadach, B. F. L. Ward and Z. Was, Comp. Phys. Comm. **130** (2000) 260; S. Jadach, B. F. L. Ward and Z. Was, Phys. Rev. **D63** (2001) 113009.

[43] S. Jadach, W. Placzek and B. F. L. Ward, Phys. Lett. **B390** (1997) 298.

[44] S. Jadach, B. F. L. Ward and Z. Was, Comp. Phys. Comm. **124** (2000) 233; S. Jadach, B. F. L. Ward and Z. Was, Comp. Phys. Comm. **79** (1994) 503.

[45] E. Accomando and A. Ballestrero, Comp. Phys. Comm. **99** (1997) 270; E. Accomando, A. Ballestrero and E. Maina, hep-ph/0204052 (2002); A. Ballestrero, R. Chierici, F. Cossutti and E. Migliore, CERN-EP/2002-069.

[46] T. Sjöstrand, Comp. Phys. Comm. **39** (1986) 347; *Pythia 5.7 and Jetset 7.4, Physics and Manual*, CERN-TH 7112/93 (1993).

[47] DELPHI Collaboration, J. Abdallah et al., *Final results on SM and MSSM Neutral Higgs Bosons* , Internal Note, DELPHI 2002-041 CONF 575, 17 July,2002.

[48] DELPHI Collaboration, P. Abreu *et al.*, EP 2003-008 (submitted to Eur. Phys. J. C.).

[49] DELPHI Collaboration, J. Abdallah *et al.*, EP 2002-088 (submitted to Eur. Phys. J. C.).

[50] DELPHI Collaboration, J. Abdallah *et al.*, CERN-EP-98-180.

[51] E. Piotto, *Search for Neutral Higgs Bosons in* $e^+e^- \to H\nu\bar{\nu}$ *Channel with DELPHI Detector at LEP 2*, Tesi di dottorato in Fisica, Università degli studi di Milano, 1999.

[52] W. Murray, V. Ruhlmann-Kleider, *Estimation of Probability Density Functions for the Higgs Search* , DELPHI 2000-067 PROG 240 (2000).

[53] J. Shiers, M. Goosens, *HBOOK Reference Manual, CERN Program Library Short Writeups* , CERN Computing and Networks Divisions (1995).

[54] J. Hansen, *A Search for Neutral Higgs Bosons in the four Jet Channel using the DELPHI Detector at LEP* , unpublished PhD dissertation, University of Oslo 2002, pp. 121 - 122.

[55] T. Hambye and K. Riesselmann, Phys. Rev. **D55**, 7255 (1977).

[56] C. J. Huberty, Applied Discriminant Analysis, John Wiley & Sons, 1994.

[57] T. Y. Young, Classification, estimation, pattern recognition, American Elsevier Publishing Coached., 1974.

[58] LEP Higgs Working Group, *Search for the Standard Model Higgs Boson at LEP*, CERN-EP/2003-011.