# A Quasi-Model-Independent Search Strategy for New High-$p_T$ Physics at Tevatron Run II

by

Jang Woo Lee

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

Master of Science in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

[September 2004]
August 2004



Author ..............................................................
Department of Physics
August 5, 2004

Certified by ..............................................................
Bruce O. Knuteson
Assistant Professor
Thesis Supervisor

Accepted by ..............................................................
Thomas J. Greytak
Associate Department Head for Education

# A Quasi-Model-Independent Search Strategy for New High-$p_T$ Physics at Tevatron Run II

by

Jang Woo Lee

Submitted to the Department of Physics
on August 5, 2004, in partial fulfillment of the
requirements for the degree of
Master of Science in Physics


## Abstract

Compelling arguments suggest the presence of new physics at the electroweak scale, an energy regime accessible to the Fermilab Tevatron. Unfortunately, predictions for the form that new physics will take are all over the map. A quasi-model-independent search strategy for new physics (SLEUTH) was introduced in Tevatron Run I to allow a probe of high-$p_T$ hadron collider data. This dissertation prescribes the SLEUTH algorithm that will be used in Tevatron Run II. Improvements over the Run I algorithm are discussed, and an intuition is developed for SLEUTH's performance by considering a number of beyond-the-standard-model examples. This dissertation defines *a priori* the SLEUTH algorithm, before its application to Tevatron Run II data, allowing an unbiased and rigorous *a posteriori* measure of the "interestingness" of any observed signal. This algorithm will also be used with small modifications at the Large Hadron Collider at CERN. In addition, a new detector simulation, TurboSim, was tested for physics analysis use for the first time.


Thesis Supervisor: Bruce O. Knuteson
Title: Assistant Professor

# Acknowledgments


To

Ho - Yeon, my wife, who endured;

Sangmin, my son, who gave me a reason;

and

my parents, who made it possible.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Context

The Standard Model is an impressive theory, accurately predicting, or at least accommodating, the results of nearly all particle physics experiments to date. This understanding points to an inconsistency in the theory that suggests the presence of new phenomena slightly above the electroweak scale.

Electroweak symmetry is broken in the standard model when a scalar field (the Higgs field) acquires a vacuum expectation value. Since the quantum corrections to the renormalized mass squared of a scalar field grow as the square of the heaviest energy scale in the theory (naively the Planck scale, of order $10^{19}$ GeV), and since the mass of the standard model Higgs boson is of the order of a few hundred GeV, a fine-tuning at the level of one part in $10^{16}$ is required to keep the Higgs mass at the electroweak scale.

Popular solutions to this hierarchy problem include the introduction of a new symmetry connecting fermionic and bosonic degrees of freedom [1], the introduction of a new strong dynamics [2], and the introduction of additional spatial dimensions. In their most general forms, these classes of models are capable of "predicting" any of many different signatures. Previous searches for these signals have fought to strike a balance between the simultaneous desires to assume as little as possible about the signal and yet achieve "optimal sensitivity" to more specific signals. These are necessarily contradictory objectives.

Many new phenomena have been predicted in addition to those resulting from these proposed solutions to the hierarchy problem. Among them are leptoquarks, proposed in an attempt to explain the relationship between quarks and leptons in the standard model and appearing in many grand unified theories; composite quarks and leptons, in case the "fundamental" particles of the standard model turn out not to be fundamental at scales $\lesssim 10^{-18}$ meters; a fourth generation of quarks or leptons; excited quarks and leptons, in analogy to the excited states of hadrons observed at much lower energies; new heavy gauge bosons, arising from additional gauge symmetries in models extending the $SU(3)_C \times SU(2)_L \times U(1)_Y$ of the standard model; and many others. Of course, Nature may have other ideas. The CDF and DØ collaborations have performed many searches on the data collected during Run I of the Fermilab Tevatron, but have we looked in all the right places?

The Standard Model has so far passed the experimental tests to which it has been subjected. Nonetheless, there is a hole in the Standard Model that indicates there are likely to be new fundamental discoveries at energy scales that our accelerators are just beginning to probe.

Possible findings include

- particles with bare magnetic charge ("magnetic monopoles");
- a new symmetry of Nature ("supersymmetry");
- a new strong force ("technicolor");
- the presence of a new weak force ("heavy gauge bosons");
- the existence of large extra spatial dimensions, curled up on scales smaller than 1 mm ("extra dimensions");
- electrons in excited states ("excited fermions");
- laws of nature that are not isotropic ("non-commutative theories");
- a heavier analog of the electron ("fourth generation of fermions"); and
- evidence pointing to a unification of the strong, weak, and electromagnetic forces ("grand unified theories").

Each of these possibilities represents a class of theories with a number of adjustable parameters. When taken together, the model space is sufficiently large that systematically checking all possibilities is not a viable option. We know only vaguely what it is we should be searching for; equivalently, we are searching for more things than can possibly be tested at one time.

Humans are notoriously good at finding patterns, particularly when dealing with small numbers of events — the history of particle physics is strewn with cases of patterns being mistakenly discerned. An algorithm is required that will enable a general search for all possibilities simultaneously, rigorously taking into account the "trials factor" (a measure of the number of different possibilities considered) when reporting a final number.

# Chapter 2

# SLEUTH

The solution of this problem in the context of particle physics at accelerators that collide protons or their antiparticles is an algorithm called SLEUTH. Consideration of the many possibilities just mentioned naturally leads one to ask whether there is any common feature among them. If such a common feature exists, perhaps it can be searched for in a general way.

It turns out that such a commonality does in fact exist, justifying the following three assumptions:

- the data can be partitioned in such a way that a signal will predominantly appear within a small number of these partitions, or Sleuth Final States;
- interactions signaling the presence of new physics will generally produce objects with larger energy than expected from background processes; and
- a signal is apt to appear as an excess of events — i.e., more events observed in the data than expected from background.

SLEUTH begins by taking all of the data collected in the experiment and partitioning it into categories. This partitioning is orthogonal; each event ends up in one and only one of these categories. The partitioning is chosen such that if new physics appears in the data, it is likely to end up predominantly in a single category. Each category contains a set of "similar" events, in the sense that the events in each category contain the same types of debris from the collision.

Standard object identification criteria are used to identify isolated and energetic electrons ($e^{\pm}$), muons ($\mu^{\pm}$), taus ($\tau^{\pm}$), photons ($\gamma$), jets ($j$), jets from a parent bottom quark ($b$), and missing energy transverse to the beam axis ($\not\!\!E_T$).

The events with the same number of final state objects are grouped together. The only exception are the jets, which are identified as pairs, $jj$. The events with odd-number of jets (2n+1) are grouped together with events with even-number of jets (2n). The reason for this grouping is because the (2n+1)-th jet is assumed to be produced by a gluon radiated by one of the incoming or outgoing quarks (initial state radiation, ISR; or final state radiation, FSR). The jets coming from parent bottom quarks ($b$'s) are also identified as pairs, $bb$. The events with one or more odd-number of well-tagged $b$'s (2n-1) are are grouped together with even-number of $b$'s (2n). The different treatment between jets ($j$) and jets from b quarks ($b$) is due to the lower efficiency of the b-tagging mechanism, which is typically around 30%. In SLEUTH$_{\text{I}}$, $b$'s were not identified as separate objects, but as jets. In SLEUTH$_{\text{II}}$, $b$'s are considered as separate objects when identified.

The variable(s) considered when looking at the event is one of the main difference between SLEUTH$_{\text{I}}$ and SLEUTH$_{\text{II}}$. In SLEUTH$_{\text{I}}$, the variables were: missing $E_T$ of the event ($\not\!\!E_T$), $\sum p_T$ of the $e/\mu$, $\sum p_T$ of the $\gamma/W/Z$, and the $\sum p_T$ of the jets, each considered a separate variable. In SLEUTH$_{\text{II}}$, the single variable considered is the $\sum p_T$ of all the objects in the events. Comparison between the two versions of SLEUTH for the top example indicates that SLEUTH$_{\text{II}}$ is more sensitive to new physics.

## 2.1 $\text{SLEUTH}_\text{I}$

In Refs. [3, 6] a quasi-model-independent search strategy ("$\text{SLEUTH}_\text{I}$"), designed to systematically search for new high $p_T$ physics at a collider experiment sensitive to physics at the electroweak scale was introduced, and it was applied to 32 populated final states in the DØ Run I data. The most interesting final state was found to be $ee4j$, with $\mathcal{P} = 0.04$. Taking into account all final states considered, the final result was $\tilde{\mathcal{P}} = 0.89$. $\mathcal{P}$ and $\tilde{\mathcal{P}}$ are measures of how interesting an excess of events are relative to the Standard Model backgrounds, and are, loosely speaking, probabilities of finding a more interesting result in that final state ($\mathcal{P}$) and in any final state ($\tilde{\mathcal{P}}$), respectively. Detailed descriptions of each value are given in the following pages.

The final outcome was a null result. $\text{SLEUTH}_\text{I}$ did not find any evidence of new physics beyond the Standard Model. $\text{SLEUTH}_\text{I}$ algorithm and results are summarized here only for comparison with $\text{SLEUTH}_\text{II}$, on which this thesis is based. The original publications on $\text{SLEUTH}_\text{I}$ contain more details. [1]

$\text{SLEUTH}_\text{I}$ has three components: the definitions of physical objects and exclusive final states; the choice of variables relevant for each final state; and an algorithm that systematically hunts for an excess in the space of those variables, and quantifies the likelihood of any excess found. We consider each in turn.

### 2.1.1 Final states

The data are partitioned into exclusive final states using standard criteria that identify isolated and energetic electrons ($e$), muons ($\mu$), and photons ($\gamma$), as well as jets ($j$), missing transverse energy ($\not\!\!E_T$), and the presence of $W$ and $Z$ bosons. We expect the first sign of new physics to appear in one of these final states. We analyze each of these final states independently.

[1] Material in this paper regarding $\text{SLEUTH}_\text{I}$ has been taken from various sources including published papers [3, 4, 5, 6] and e-Prints [7]. This summary section on $\text{SLEUTH}_\text{I}$ is largely based on [8].

### 2.1.2 Variables

For each exclusive final state, we consider a small set of variables summarized in Table 2.1.

| If the final state includes | then consider the variable |
|---|---|
| $\not\!\!E_T$ | $\not\!\!E_T$ |
| one or more charged leptons | $\sum p_T^{\ell}$ |
| one or more electroweak bosons | $\sum p_T^{\gamma/W/Z}$ |
| one or more jets | $\sum p_T^{j}$ |

Table 2.1: A quasi-model-independently motivated list of interesting variables for any final state. The set of variables to consider for any exclusive channel is the union of the variables in the second column for each row that pertains to that final state.

### 2.1.3 Algorithm

The SLEUTH$_{\text{I}}$ algorithm requires as input a data sample, a set of events modeling each background process $i$, and the number of background events $\hat{b}_i \pm \delta\hat{b}_i$ from each background process expected in the data sample. From these we determine the region $\mathcal{R}$ of greatest excess and quantify the degree $\mathcal{P}$ to which that excess is interesting. The algorithm itself, applied to each individual final state, consists of seven steps:

1. We construct a mapping from the $d$-dimensional variable space defined by Table 2.1 into the $d$-dimensional unit box (i.e., $[0,1]^d$) that flattens the total background distribution. We refer to $[0,1]^d$ for arbitrary $d$ as the "unit box," meaning unit interval when $d = 1$, unit square when $d = 2$, unit cube when $d = 3$, and so forth. We use this to map the data into the unit box.

2. Regions are then defined about sets of data points using the concept of Voronoi diagrams, an example of which is shown in Fig. 2-1. We define a "region" $R$ about a set of $N$ data points to be the volume within the unit box closer to one of the data points in the set than to any of the other data points in the sample. The arrangement of data points themselves thus determines the regions. A region containing $N$ data points is called an $N$-region.

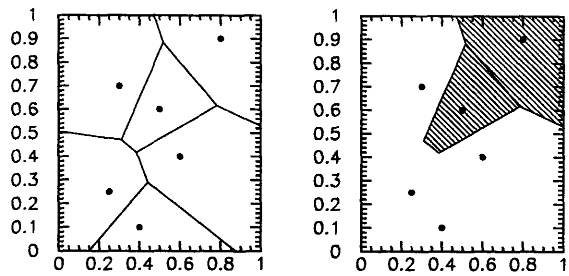


Figure 2-1: The seven black circles in each panel are data points within the unit square. The Voronoi diagram (a) is formed by drawing perpendicular bisectors of imaginary line segments connecting neighboring pairs of points; the resulting lines partition the unit square into regions with the property that each space point inside the region is closer to the single data point inside that region than to any other data point in the square. Regions considered by SLEUTH$_{\rm I}$ are unions of these individual regions, such as the shaded region in (b). Criteria are imposed upon the regions that SLEUTH$_{\rm I}$ is allowed to consider, including the criterion that the region include the "upper right-hand corner" (1,1) of the unit square, as shown in (b).

3. Each region contains an expected number of background events $\hat{b}_R$, numerically equal to the volume of the region $\times$ the total number of background events expected, and an associated systematic error $\delta\hat{b}_R$, which varies within the unit box according to the systematic errors assigned to each contribution to the background estimate. We can therefore compute the probability $p_N^R$ that the background in the region fluctuates up to or beyond the observed number of events. This probability is the first measure of the degree of interest of a particular region.

4. The rigorous definition of regions reduces the number of candidate regions from infinity to $\approx 2^{N_{\rm data}}$ , where $N_{\rm data}$ is the number of events observed in the data in this category. Imposing explicit criteria on the regions that the algorithm is allowed to consider further reduces the number of candidate regions. We apply geometric criteria that favor high values in at least one dimension of the unit box, and we limit the number of events in a region to fifty. The number of remaining candidate regions is still sufficiently large that an exhaustive search

is impractical, and a heuristic is employed to search for regions of excess. In the course of this search, the $N$-region $\mathcal{R}_N$ for which $p_N^R$ is minimum is determined for each $N$, and $p_N = \min_R (p_N^R)$ is noted.

5. In any reasonably-sized data set, there will always be regions in which the probability for $b_R$ to fluctuate up to or above the observed number of events is small. We determine the fraction $P_N$ of *hypothetical similar experiments* (hse's) in which $p_N$ found for the hse is smaller than $p_N$ observed in the data by generating random events drawn from the background distribution and computing $p_N$ by following steps 1–4.

6. We define $P$ and $N_{\min}$ by $P = P_{N_{\min}} = \min_N (P_N)$, and identify $\mathcal{R} = \mathcal{R}_{N_{\min}}$ as the most interesting region in this final state.

7. We use a second ensemble of hse's to determine the fraction $\mathcal{P}$ of hse's in which $P$ found in the hse is smaller than $P$ observed in the data. The most important output of the algorithm is this single number $\mathcal{P}$, which may loosely be said to be the "fraction of hypothetical similar experiments in which you would see an excess as interesting as what you actually saw in the data." $\mathcal{P}$ takes on values between zero and unity, with values close to zero indicating a possible hint of new physics. The computation of $\mathcal{P}$ rigorously takes into account the many regions that have been considered within this final state.

The smallest $\mathcal{P}$ found in the many different final states considered ($\mathcal{P}_{\min}$) determines $\tilde{\mathcal{P}}$, the "fraction of *hypothetical similar experimental runs* (hser's) that would have produced an excess as interesting as actually observed in the data," where an hser consists of one hse for each final state considered. $\tilde{\mathcal{P}}$ is calculated by simulating an ensemble of hypothetical similar experimental runs, and noting the fraction of these hser's in which the smallest $\mathcal{P}$ found is smaller than the smallest $\mathcal{P}$ observed in the data. Because $\tilde{\mathcal{P}}$ depends only on the single final state that defines $\mathcal{P}_{\min}$, correlations among final states may be neglected in this calculation. Like $\mathcal{P}$, $\tilde{\mathcal{P}}$ takes on values between zero and unity, and the potential presence of new high $p_T$ physics would be indicated by finding $\tilde{\mathcal{P}}$ to be small. The difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ is that in computing $\tilde{\mathcal{P}}$ we account for the many final states that have been considered.

### 2.1.4 Results

SLEUTH$_{\mathrm{I}}$'s performance on representative signatures has been studied.[3, 4, 5, 6] When ignorance of $t\bar{t}$ is feigned in the $e\mu X$ final states, we find $\mathcal{P}_{e\mu \not{E}_T 2j} = 1.9\sigma$ in DØ data, correctly suggesting the presence of $t\bar{t}$. Feigning ignorance of $t\bar{t}$ in the $W$+jets-like final states, we find $\mathcal{P}_{\mathrm{min}} > 3\sigma$ in 30% of an ensemble of mock experimental runs on the final states $W\,3j$, $W\,4j$, $W\,5j$, and $W\,6j$. Dedicated searches for the top quark in these channels[12] yield an excess of $2.75\sigma$ in $e\mu\not{E}_T\,2j$, $2.6\sigma$ in $W\,4j(nj)$ with no $b$-tag, and $3.6\sigma$ in $W\,3j(nj)$ with a $b$-tag.

SLEUTH$_{\mathrm{I}}$ was applied to over thirty exclusive final states at DØ, determining $\mathcal{P}$ for each. Upon taking into account the many final states (both populated and unpopulated) that are considered, it was found $\tilde{\mathcal{P}}$=0.89, implying that 89% of an ensemble of hypothetical similar experimental runs would have produced a final state with a candidate signal more interesting than the most interesting observed in these data. This can be interpreted as a null result, i.e. no signature beyond the standard model was found by applying SLEUTH$_{\mathrm{I}}$ to the DØ data from Run I of the Fermilab Tevatron.

## 2.2 Problems of Sleuth$_{\text{I}}$

Sleuth$_{\text{I}}$ had room for improvement, and some of the issues that provided motivation for development of an improved version of Sleuth, Sleuth$_{\text{II}}$, are as follows.

- In the 2-step transformation technique, the "interestingness" of an event is dependent on the relative distribution of the data points.

- The "interestingness" of an event does not depend on the $p_T$, in the case of a final state with a single data point, when it should.

- The integrations for background fluctuation calculation starts half-way between the data points. But the half-way point changes before and after the transformation of the data points. There is an ambiguity in which half-way point should be taken. The Sleuth$_{\text{I}}$ algorithm takes the half-way point after the transformation. In Sleuth$_{\text{II}}$, because it is 1-dimensional, we can start the integration from the data point of interest. Thus, there is no ambiguity in the range.

- There is an ambiguity in the heuristics of selecting the region of interest, which could be questioned if we found something new.

- CPU issue 1: The 2nd step of the transformation involves all the combinations of the data points. The time cost scales worse than $O(N^2)$, where $N$ is the number of Monte Carlo points used to estimate the background. When $N \lesssim O(10^4)$, the time cost becomes prohibitive. We want to reduce the computation time.

- CPU issue 2: When calculating the "interestingness" measure ($\tilde{\mathcal{P}}$) involving the hypothetical similar experimental runs (hser's), if we want a definite number instead of an upper limit, the number of calculations increases with $O(N)$. If an interesting region is found, we want the algorithm to be fast enough to get a definite number instead of an upper limit. And we want the ability to fill in the high-$p_T$ tail of the background, should it be necessary to calculate the definite number.

- When integrating to compute the background fluctuations (the cummulative background distribution function), the integral in two or more dimensions is generally not single-valued and depends on the path of the integration. 1-dimension does not have this problem.

As a solution to these problems, we chose to use in SLEUTH$_{\rm II}$ the sum of the transverse momentum, $\sum p_T$, of all the objects in an event as the single variable characterizing an event, and not other variables, for the following reasons.

- SLEUTH$_{\rm I}$ used the $p_T$ of the objects individually (because the expected values and resolutions are different for leptons, photons, W, Z, jets and $\not\!\!E_T$), but the "interestingness" of an event is represented equally well by the scalar sum of the transverse momentum, $\sum p_T$.

- Invariant mass, for example, is good for finding a single particle or testing a single theory, but it is not so clear when doing a model-independent search.

  ex1) $LQ\ \overline{LQ} \to e^+e^-jj$. It is not clear which objects to take for the invariant mass, just from looking at the final state objects.

  ex2) $t\ \bar{t} \to W\ W\ b\ \bar{b} \to e\ \mu\ \nu_e\nu_\mu\ b\ \bar{b}$. Again the combinatorics is the problem. Also, there is a "trials factor" involved with each "edge" of the region examined. SLEUTH$_{\rm II}$ only has one "edge", the lower $p_T$ bound; a mass window has two.

Whether using a single observable, $\sum p_T$, was a sound choice can only be answered by looking at examples, and comparing the results from the two versions of SLEUTH.

## 2.3 SLEUTH$_{\text{II}}$, in general

### 2.3.1 Definitions of final states

The specification of final states is based on the notions of exclusive channels and standard particle identification. We attempt to label these exclusive final states as completely as possible while maintaining a high degree of confidence in the label.

An event is placed into an exclusive final state according to the following rules:

1. Event requirements.

   (a) Cleanup. Standard cleanup criteria are imposed to ensure that all subsystems were operating properly when the event was recorded. Events failing these criteria are dropped.

   (b) Kinematic criteria. Each event must pass at least one trigger with sufficient room to spare that the trigger may be considered fully efficient for that type of event. Events not satisfying this requirement are dropped.

2. Object identification.

   (a) Final state objects. We identify the number of isolated electrons, muons, photons, and jets observed in the event, and whether there is a significant imbalance in transverse momentum. We treat $\not{E}_T$ as an object in its own right, which must pass certain quality criteria.

   (b) Identification criteria. We define the *identification efficacy* $I$ for particular criteria used to define an object $o$ by considering the inclusive sample $oX$, scaled to 1 fb$^{-1}$ of data. Denote by $f$ ("fakes") the number of events expected in this data sample in which the $o$ is falsely identified, and denote by $\epsilon$ ("efficiency") the probability that an event in which a true $o$ is produced appears in this sample. Define

$$I = \frac{(1\text{pb})(1\text{fb}^{-1})\epsilon}{\sqrt{f}}. \tag{2.1}$$

If standard identification criteria for $o$ exist, we use the standard. If no standard exists, we use the criteria that maximize $I$.

(c) Kinematic criteria. We identify objects with $p_T > 20$ GeV in the fiducial volume of the detector.

(d) Notable omissions. We elect not to identify top quarks, or jets with charm quark parents.

3. Differentiation of jets.

(a) $b\bar{b}$. Our understanding of quark flavor suggests that $b$ quarks should be produced in pairs. We identify $b$ quarks as pairs rather than individually in order to increase the robustness of identification, and to reduce the total number of final states. One extremely well-tagged $b$ jet may be sufficient to consider an event to contain a $b\bar{b}$ pair.

(b) $jj$. Our understanding of color flow suggests that quark jets should appear in pairs. We determine the integer $n$ for which the number of jets in the event is at least $2n$ but not greater than $2n+1$. The event is then said to contain $n$ $jj$ pairs. If there are one or more jets in the event with $20 > p_T^j > 10$ GeV, we include the most energetic of these in our counting of jets. We identify jets as pairs rather than individually in order to reduce the total number of final states, and to keep signal events with one additional radiated gluon within the same final state.

(c) $\tau$. A jet identified as a $\tau$ with high probability will be considered to be a $\tau$.

4. Combinations.

(a) $Z$ boson. We combine an $\ell^+\ell^-$ pair into a $Z$ boson if the invariant mass $m_{\ell^+\ell^-}$ falls within a $Z$ boson mass window ($75 \leq m_{\ell^+\ell^-} \leq 105$ GeV) and the event contains neither significant $\not\!\!E_T$ nor a third charged lepton. (Here and elsewhere "$\ell$" and "lepton" denote an electron, muon, or tau.) If the event contains exactly one photon in addition to a $\ell^+\ell^-$ pair, and contains neither significant $\not\!\!E_T$ nor a third charged lepton, and if $m_{\ell^+\ell^-}$ does not fall within the $Z$ boson mass window, but $m_{\ell^+\ell^-\gamma}$ does, then the $\ell^+\ell^-\gamma$ triplet becomes a $Z$ boson.

(b) $W$ boson. A lepton and $\not\!\!E_T$ become a $W$ boson if the transverse mass $m^T_{\ell\not E_T}$ is within a $W$ boson mass window ($m^T_{\ell\not E_T} \leq 110$ GeV) and the event contains no second charged lepton. Because the $W$ boson mass window is so much wider than the $Z$ boson mass window, no attempt is made to identify radiative $W$ boson decays.

5. Equivalence relations.

(a) Global charge conjugation. Final states that are related through global charge conjugation are considered to be equivalent. Thus $e^+e^-\gamma$ is a different final state than $e^+e^+\gamma$, but $e^+e^+\gamma$ and $e^-e^-\gamma$ together make up a single final state.

(b) Global lesser generation switch. Final states that are related through global switching of the first and second generation are considered to be equivalent. Thus $e\not\!\!E_T\gamma$ and $\mu\not\!\!E_T\gamma$ together make up a single final state. The decision to consider third generation objects ($b$ quarks and $\tau$ leptons) differently from first and second generation objects reflects our prejudice that the third generation may be "special."

6. Evolution with time.

(a) In assessing the most recent, up-to-date SLEUTH result on any particular day, no account should be taken of previous results. In particular, there is no additional trials factor due to the passage of time.

### 2.3.2 Variables

There is strong reason to believe that the physics responsible for electroweak symmetry breaking occurs at the scale of the mass of the Higgs boson, or on the order of a few hundred GeV. Any new massive particles associated with this physics can therefore be expected to decay into objects with large transverse momenta in the final state. We use a single variable: $\sum p_T$, the summed transverse momenta of all objects in the event.

### 2.3.3 Algorithm

The algorithm used to search for regions of excess has been greatly simplified from the algorithm used in $\textsc{Sleuth}_\mathrm{I}$. $\textsc{Sleuth}_\mathrm{II}$ uses only one variable, $\sum p_T$. For each $N$, the single region $\mathcal{R}$ considered consists of those $N$ events in that final state with the largest $\sum p_T$. We set $p_N^R$ equal to the integral of the background from the position of the data point with the $N^{\text{th}}$ largest value of $\sum p_T$ up to infinity. Use of hypothetical similar experiments and the combining of results from different final states into a single $\tilde{\mathcal{P}}$ follows as in $\textsc{Sleuth}_\mathrm{I}$.

$\textsc{Sleuth}_\mathrm{II}$ algorithm involves four major steps:

1. The data are partitioned into exclusive final states according to the objects in the detector, with some modifications as described the previous section on definitions of final states. The purpose of the partition and modifications is to group the events so that a signal beyond the Standard Model will likely appear in a single or a small number of Sleuth Final State.

2. For each event in an exclusive final state (Sleuth Final State), a single variable is calculated: the summed transverse momentum ($\sum p_T$) of all objects in the event. The missing energy ($\not{E}_T$) in an event is considered as an object, and included in the sum, if the missing energy is a significant part of the final state.

3. For each of the final states, the events in the final state are ordered in $\sum p_T$. Regions $R$ are defined in each final state by the semi-infinite intervals ($\sum p_T$, $\infty$) with the lower bound at each data point in the $\sum p_T$ distribution. The interestingness $p_N$ of an arbitrary region containing $N$ data points is defined as the Poisson probability that the integrated background with $\sum p_T$ above the

lowest $\sum p_T$ of the $N$ data points would fluctuate up to or beyond $N$.

$$p_N = \sum_{i=N}^{\infty} \frac{e^{-b_R}(b_R)^i}{i!}, \tag{2.2}$$

where, $b_R$ is the number of background events in the region $R$ with $N$ events.

The most interesting region $\mathcal{R}$ of each final state is determined by the $N$ data points for which $p_N$ is minimal. The fraction $\mathcal{P}$ of hypothetical similar experiments in which a region more interesting than $\mathcal{R}$ would be seen in this final state is determined by performing pseudo experiments, assuming a Poisson distribution of the number of background events. The hypothetical similar experiments are realized by generating a random number of events from a Poisson distribution with the mean equal to the number of background events in that region $\mathcal{R}$. Thus, the values calculated for each final state is the $\sum p_T$ which defines the region $\mathcal{R}$, and the $\mathcal{P}$ which shows how interesting that region $\mathcal{R}$ of that final state is.

$$\mathcal{P} = \frac{1}{N_{hse}} \sum_{i=1}^{N_{hse}} \Theta(p_N^{data} - p_N^{hse}), \tag{2.3}$$

$$\Theta(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0, \end{cases}$$

where, $N_{hse}$ is the number of hypothetical similar experiments, $p_N^{data}$ and $p_N^{hse}$ are the probabilities $p_N$ calculated for data and hypothetical similar experiments.

4. The most interesting region $\tilde{\mathcal{R}}$ from all final states considered is the region $\mathcal{R}$, i.e. final state and $\sum p_T$, with the minimum $\mathcal{P}$. The fraction $\tilde{\mathcal{P}}$ of hypothetical similar experiments in which a region more interesting than $\tilde{\mathcal{R}}$ would be seen in any final state is determined by performing additional pseudo experiments, again assuming a Poisson distribution of the number of background events, taking into account the number of final states considered. For each final state considered, we approximate the probability of seeing a more interesting region by taking the lesser value between the smallest $\mathcal{P}$ from the previous step or

the probability that the number of background events for that region fluctuates above a minimum number of events $n_{min}$. This minimum number of events was changed from one (i.e. an event is seen in that region $\mathcal{R}$) to four as discussed in the following section during this study.

$$\tilde{\mathcal{P}} = 1 - \prod_{j=1}^{N_{fs}} (1 - min( \sum_{i=n_{min}}^{\infty} \frac{e^{-b_R}(b_R)^i}{i!}, \mathcal{P}_{min})), \tag{2.4}$$

where $\mathcal{P}_{min}$ is the minimum $\mathcal{P}$ seen in the data.

SLEUTH takes as inputs the observed data and estimated background, and the outputs are the most interesting regions $\mathcal{R}$ and $\mathcal{P}$ for each final state. The final SLEUTH results are the most interesting region $\tilde{\mathcal{R}}$ seen in the data, given as a final state and a threshold $\sum p_T$, and the fraction $\tilde{\mathcal{P}}$ which is a measure of the interestingness of this region. If the data does not differ much from the expected background, $\tilde{\mathcal{P}}$ will be some relatively large number between zero and one. If data differs significantly from the expected background, $\tilde{\mathcal{P}}$ will be a small number close to zero.

In High Energy Physics, the default threshold for discovery is five standard deviations, corresponding to a probability of approximately $10^{-7}$, in the deviation of observed data from the expected value. The "trials factor", i.e. a factor which reflects how many different places a deviation from expected value could have been searched for, has been argued to be approximately $10^4$ over the course of a given experiment. [9] Because this "trials factor" has already been accounted for in calculating $\tilde{\mathcal{P}}$, we claim that a signal of 5 standard deviations, or a probability of approximately $10^{-7}$, can be seen as equivalent to a $3\sigma$ effect, or a probability of approximately $10^{-3}$, in terms of $\tilde{\mathcal{P}}$. Thus, the threshold for discovery in this paper is taken to be $\tilde{\mathcal{P}} < 10^{-3}$. The exact conversion from units of standard deviations ($\tilde{\mathcal{P}}_{[\sigma]}$) to $\tilde{\mathcal{P}}$ is given by

$$\tilde{\mathcal{P}} = \frac{1}{\sqrt{2\pi}} \int_{\tilde{\mathcal{P}}_{[\sigma]}}^{\infty} e^{-t^2/2} dt \tag{2.5}$$

### 2.3.4 Minimum number of events in each final state

Only final states in which four or more events are observed in the data are considered by SLEUTH. This anticipates the next phase of the analysis, in which an interpretation of the underlying physics is sought. This interpretation requires seeing a pattern or theme among the observed events, and for this several events are needed.

Purely practial considerations also demand that only final states with four or more events be considered. Suppose we find $\mathcal{P} = 10^{-6}$ in the final state $1e^{+}1e^{-}1bb$ ; then in computing $\tilde{\mathcal{P}}$ all final states with $b > 10^{-6}$ must be considered and accounted for. (A final state with $b = 10^{-7}$, on the other hand, counts as only $\approx 0.1$ final states, since the fraction of hypothetical similar experiments in which $\mathcal{P} < 10^{-6}$ in this final state is equal to the fraction of hypothetical similar experiments in which one or more events is seen in this final state, which is $10^{-7}$.) This is a large practical problem, since it requires that all final states with $b > 10^{-6}$ be enumerated and estimated, and it is very difficult to do this believably.

To solve this problem, suppose SLEUTH considers only final states with at least $d_{\mathrm{min}}$ events observed in the data. Suppose further that our goal is to be able to find $\tilde{\mathcal{P}} < \tilde{\mathcal{P}}_{\mathrm{discovery}}$, where we have argued that the appropriate threshold for discovery is $\tilde{\mathcal{P}}_{\mathrm{discovery}} = 10^{-3}$. There will be some number $N_{\mathrm{fs}}(b_{\mathrm{min}})$ of final states with expected number of events $b > b_{\mathrm{min}}$; we must choose $b_{\mathrm{min}}$ to be sufficiently large that all of these $N_{\mathrm{fs}}(b_{\mathrm{min}})$ final states can be enumerated and estimated. The time cost of simulating events is such that the integrated luminosity of our Monte Carlo events is at most 100 times the integrated luminosity of the data; this practical constraint restricts $b_{\mathrm{min}} > 0.01$. The number of SLEUTH Tevatron Run II final states with $b_{\mathrm{min}} > 0.01$ is roughly $N_{\mathrm{fs}}(b_{\mathrm{min}} = 0.01) = 10^{3}$.

We have roughly $\tilde{\mathcal{P}} = \mathcal{P}_{\mathrm{min}} N_{\mathrm{fs}}(b_{\mathrm{min}})$, where $\mathcal{P}_{\mathrm{min}}$ is the smallest $\mathcal{P}$ found in any final state. From the discussion above, we can only be confident that we have correctly computed $\tilde{\mathcal{P}}$ from $\mathcal{P}_{\mathrm{min}}$ if $\mathcal{P}_{\mathrm{min}} > ({b_{\mathrm{min}}}^{d_{\mathrm{min}}})$; if otherwise, final states with $b < b_{\mathrm{min}}$ will need to be accounted for. Thus we can only confidently compute $\tilde{\mathcal{P}}$ if $\tilde{\mathcal{P}} > ({b_{\mathrm{min}}}^{d_{\mathrm{min}}}) N_{\mathrm{fs}}(b_{\mathrm{min}})$.

Solving this inequality for $d_{\min}$, we find

$$d_{\min} > \log_{10}\left(\frac{\tilde{\mathcal{P}}}{N_{\rm fs}(b_{\min})}\right)(\log_{10} b_{\min})\,. \qquad (2.6)$$

Inserting the values above in the right hand side of this expression, we find the requirement $d_{\min} \geq 4$ has been imposed upon us.

### 2.3.5 Comparison between Sleuth$_{\rm I}$ and Sleuth$_{\rm II}$

As a direct comparison between Sleuth$_{\rm I}$ and Sleuth$_{\rm II}$ we compare the results of the top quark example for both versions of Sleuth. From previous Sleuth$_{\rm I}$ result [6], we know that Sleuth$_{\rm I}$ would not have been able to "discover" a top signal when compared to a Standard Model background without the $t\bar{t}$ process. In this study we repeat the same analysis, and show that Sleuth$_{\rm II}$ is able to find the same top signal, demonstrating the improved performance of Sleuth$_{\rm II}$ over Sleuth$_{\rm I}$. Section 5.4 of this thesis looks in detail at the top example.

## 2.4 SLEUTH$_{II}$: Tevatron Run II

For this study, the event requirements, or "triggers" as they are often called, are an isolated and energetic electron, muon, or tau with $p_T > 25$ GeV/c; a photon with $p_T > 200$ GeV/c; a jet or a b-jet with $p_T > 400$ GeV/c. An event is required to have at least one of these objects to be considered as a final state. As the kinematic requirement of an object, we applied a uniform $p_T$ cut of $p_T > 20$ GeV/c for all the objects: electrons, muons, taus, photons, jets and b-tagged jets. This cut is similar to common "standard" cuts applied to these objects in other analysis at the CDF experiment. These two types of cuts meet the kinematic criteria of the event requirements and object identification of Sleuth final states as described previously. An additional overall $\sum p_T > 200$GeV/c cut was applied to the events, limiting our attention to only high-$p_T$ events.

# Chapter 3

# Objectives

The objectives of the study, which this thesis is based on, is two folds. The first and primary objective is to fully develop and test the implementation of $\text{SLEUTH}_{\text{II}}$ [1] , the quasi-model-independent algorithm to find and quantify the excess of data compared to Standard Model background. The secondary objective is to test the usability of a new detector simulation TurboSim [10], which takes parton level events and simulates the detector responses via a look-up table, by-passing the hadron level objects.

The development and testing of SLEUTH is the main focus of this study. At the start of this study, the SLEUTH algorithm was still not fully developed, though the basic algorithm was already implemented. The first part of the development process was to test the algorithm using a detector simulation PGS (Pretty Good Simulation) [11], which provides a rough parametrized detector simulation appropriate for either of the two Tevatron experiments. PGS allowed us to modify, test and debug the algorithm by quickly generating Monte Carlo events of both the Standard Model processes – from which the pseudo data was pulled to simulate Standard Model processes of data, and with which the background was generated – and the non-Standard Model processes which were used as "signals" to test the response of SLEUTH.

One of the early modifications was to add the $p_T$ threshold as an additional cross-check of the SLEUTH result $\mathcal{P}$, so that it could be compared against "signals" for

[1] SLEUTH and $\text{SLEUTH}_{\text{II}}$ are used interchangeably in this paper. Unless refering to both $\text{SLEUTH}_{\text{I}}$ and $\text{SLEUTH}_{\text{II}}$ from the context, SLEUTH will henceforth mean $\text{SLEUTH}_{\text{II}}$.

sensibility of the regions selected by the $p_T$ thresholds. Also added was the output to directly compare the number of events in the data and background above a $p_T$ threshold, again as a sensibility check.

Another early modification made to the algorithm was to add the ability to generate more events in the high-$p_T$ background. The purpose of this was to 1) fill in the high-$p_T$ background should it be necessary to allow SLEUTH to calculate the $\mathcal{P}$ for Final States [2] which did not have enough background above the $p_T$ threshold for comparison with data; and 2) to allow the option of generating minimum number of background events in the first step so that CPU time may be saved, only to fill in the high-$p_T$ tail if it becomes necessary. The two-step approach in the generation of background allows us the option of reducing the number of events needed for the background by roughly a factor of 10, which reduces the CPU time by an order of magnitude. This option will be important especially when a full detector simulation is used, when CPU processing time becomes a big issue.

For the current study, the detector simulation was done mostly by two faster and simplified simulations: PGS (detector response is parameterized) and TurboSim (detector response is tabulated in a look-up table).

The SLEUTH algorithm has been extensively tested with both the PGS and TurboSim detector simulations for bugs. The current implementation has been shown to be stable for both cases.

One of the objectives of this study was to answer the "6 Open Questions" regarding Sleuth as described in the Section 2.3.1 Definitions of final states. The questions are summarized here for clarity.

1) Should we treat jets as pairs: e.g. 3 jets (3j) $\rightarrow$ a jet pair (1jj) ?

2) Should we treat b's as pairs, separate from jets : e.g. 1b $\rightarrow$ 1bb, if well-tagged?

3) Should we combine $l^+l^-$ or $l^+l^-\gamma$ into a Z, if $m_{l^+l^-}$ or $m_{l^+l^-\gamma}$ is between 76 and 106 GeV/c$^2$ ?

[2] "Final States" refers to Sleuth Final States, which are defined by Partition Rule and partitioned accordingly. Each Final State may include two or more actual final states, which are composed of different detector objects.

4) Should we combine $l^+$ missing-$E_T$ ($\not\!\!E_T$) or $l^-$ missing-$E_T$ ($\not\!\!E_T$) into a $W^\pm$, if the invariant mass is less than 110 GeV/c$^2$ ?

5) Should we apply global charge conjugation: e.g. $e^+e^+\gamma = e^-e^-\gamma \neq e^+e^-\gamma$ ?

6) Should we group 1st and 2nd generation together, while treating 3rd generation separately from them: e.g. treat 1$e^+$ ( includes $e^+$, $\mu^+$) and 1$\tau^+$; 1jj and 1bb as Sleuth Final State objects?

The answers to the questions above are obtained by looking at how SLEUTH responds to various examples in the following pages. The basic criteria for determining the answers will be whether the choices increase the likelihood that Sleuth will find a deviation in data from the background, by maximizing the number of data events in a given Final State while keeping the Final States with large background events separate, in addition to the physics motivation of partitioning into the exclusive Final States.

Another part of the study was to finalize the specific partitioning rule for each object into Final States. This included how to treat jets (j's) and jets tagged as coming from b quarks (b's). This is not as simple as it seems at first glance, as we shall see later.

A uniform $p_T > 20$GeV/c cut was applied to all observed objects in the detector. Of particular interest was the $p_T$ cut on missing-$E_T$, or $\not\!\!E_T$ (pmiss), due to limitations in TurboSim to handle $\not\!\!E_T$ (pmiss) correctly. A separate study was done on the $\not\!\!E_T$ (pmiss) to determine the $p_T$ cut. The final version of the partition rule, Partition Rule ver. 2.0, reflects the results of this.

Most of the objects had a uniformly low cut of $p_T > 20GeV/c$, but well above the minimum energy required for the identification of the objects, to include as many objects as possible for a full description of the events. The cut at $p_T > 20GeV/c$ corresponds to the scale at which pQCD (perturbative Quantum Chromodynamics) calculations and experimental jet clustering become dependable. The algorithm is by design (in principle) not sensitive to the exact cuts of the objects, since the deviation of data from background is calculated above a $p_T$ threshold of the $\sum p_T$ typically much larger – the lowest being 200 - 300 GeV/c – than the sum of the cuts of the individual objects. But it was important for the object $\not{E}_T$ (pmiss) in particular. This is related to the detector simulation TurboSim used in this study, and the fact that pmiss is not an observed object in the detector, but one calulated for each event from the $p_T$ of the other objects in the event. The specific cuts for the "composite" objects W and Z also require a more detailed treatment, since the cuts are close to the $p_T$ thresholds, and influences the total number of Final States by partitioning into separate Final States the events which contain them.

The additional trigger cut applied at $\sum p_T = 200\text{GeV/c}$ for this study, reduces the number of events that we need to consider. This trigger cut has marginal effect on the final results of the examples considered, except possibly the vector boson (di-boson production) example.

A secondary objective, to test the usability of a new detector simulation TurboSim, was based on the need to use a fast detector simulation which provided reasonable description of the detector response while allowing quick turn around of the event generation for repeated testing and debugging. Also, the newly developed TurboSim had been shown to agree fairly well with the full detector simulation for various observables in terms of overall distributions as a proof of concept. [10] But it still had not been used in any analysis to demonstrate the usability of the software package. Thus, this study is also a first demonstration of the usability of TurboSim as a detector simulation for an analysis. TurboSim itself has undergone some changes while this study was in progress, and the two versions of it will be refered to in this thesis as TurboSim ver.1 (dated Feb 16) and TurboSim ver.2 (dated Apr 25). The dates refer to the dates of the look-up tables which were used for the study, produced by TurboSim as a result of "training" on existing Monte Carlo samples from CDFSim, the standard full detector simulation at CDF.

# Chapter 4

# Methodology

This thesis is based on the Monte Carlo events generated with the Pythia [14] event generator, which are then passed through the detector simulations, and then processed by different components of an analysis package which is under development. Finally the Sleuth algorithm takes the inputs, partitions them according to the partition rule specific to the type of experiment (i.e. collision) and the Sleuth algorithm, determines which Sleuth Final States are most interesting, and calculates the probability that a similar experiment would yield a more interesting result. Table 4.1 shows a schematic of the flow of information from the event generation stage to the final calculations of $\mathcal{P}$ and $\tilde{\mathcal{P}}$ by Sleuth.

As shown in Table 4.1, the information generated by Pythia passes through many stages and components of the analysis package, and this study includes a Monte Carlo production testing of the whole package at the level and scale expected for the final analysis of data, as well as the testing and debugging of the components and scripts in the various stages.

Especially, the Detector Simulation stage can change by replacing the component with different types of detector simulation program: the first one tested was PGS (used to produce the first $Z'$ example with mass $m_{Z'} = 200$), and TurboSim was used to generate most of the examples.

In effect, this study represents the first test of the actual usability of TurboSim for an analysis, and as such, shares many limitations with the TurboSim component of the package, in addition to whatever limitations the main Sleuth part of the package may have. Specifically, the missing energy of an event is not handled with the resolution comparable to the other objects in an event, as can be seen in the $Z'$ example, and especially the leptoquark example, where we do not expect such widespread existence of $\not{E}_T$ (pmiss). The number of jets and the value of jet $p_T$ becomes less accurate with the increase in the number of near-by jets. Since, with the increasing number of jets in an event, the possibility of jets being close to each other increases, the accuracy of the number of jets and jet $p_T$ at the detector level degrades as the number of jets increases. Another problem in the TurboSim is the handling of b's, jets which are identified as coming from a b quark by the displacement of its vertex from the collision vertex. The details will be discussed in the following sections with examples.

However, the limitations due to TurboSim compared to the full detector simulation, can be improved or eliminated by replacing the Detector Simulation component with a full-detector simulation within CDFSim. The current study is based mostly on TurboSim due to its availability, processing speed, and a practical need to test it in parallel with the testing and development of Sleuth.

Flow of Information

| | |
|---|---|
| 1. [ Event Generator ]<br>( Vista / Pythia )<br><br>- Vista: generateAllSMBkg<br>generateOneSMBkg<br>generateMoreSMBkg<br>- Quaero: generate.f<br>- Pythia | → 2. [ Detector Simulation ]<br>( Vista / TurboSim \| PGS \| CDFSim )<br><br>- PGS (Pretty Good Simulation)<br>or TurboSim : called from simulate<br>or CDFSim<br><br>- Vista: simulate<br>misReconstruct |
| → 3. [ Partitioning I ]<br>( Quaero )<br><br>- Quaero: partition<br>PartitionRule | → 4. [ Partitioning II ]<br>( Quaero / Sleuth )<br><br>- Sleuth: objects2SleuthVariables |
| → 5. [ Calculate Sleuth Variables ]<br>( Sleuth )<br><br>- Sleuth: objects2SleuthVariables | → 6. [ Sleuth Results Calculation ]<br>( Sleuth )<br><br>- Sleuth: sleuth<br>MultiChannelResult |

Table 4.1: The process used in this study to generate MC events, simulate the detector response, and partition them for feeding into SLEUTH. Each step in the process, in brackets, [...], is followed by the program or part of the package under development, in parenthesis (...), used to carry out the step. The name of the functions used in the package follows.

# Chapter 5

# Examples

## 5.1 $Z'$

The $Z'$ is a heavier analogue of the Z boson, and decays into a lepton pair: $e^+e^-$, $\mu^+\mu^-$, $\tau^+\tau^-$, or $\nu\bar{\nu}$ at the generator level. It is the cleanest signal we could hope to generate at a hadron collider such as the Tevatron, and was used in this study to initially debug each component of the package, operation of the components as a whole, and develop tools used in this study. The $Z'$ was also used to get an estimate of the amount of signal events needed to do a "plausible" study of the SLEUTH algorithm at a given integrated luminosity. It was then used to answer some of the open questions, posed at the beginning of the study, regarding the leptons.

### $Z'$ ($\mathrm{m_{Z'}} = 200\mathrm{GeV/c^2}$) in $1\mathrm{pb}^{-1}$ of data

Suppose a $Z'$ with mass $m_{Z'} = 200GeV/c^2$ were produced in $p\bar{p}$ collisions with a production cross section of 300 pb. In 1 $\mathrm{pb}^{-1}$ of data in a simulated experiment, 9 events land in final states considered by SLEUTH. These events are added to an event sample drawn from known Standard Model processes to create a sample of events in which the manifestation of the $Z'$ is hidden. SLEUTH is provided Monte Carlo background events corresponding to 10 $\mathrm{pb}^{-1}$ of known Standard Model processes, reweighted to 1 $\mathrm{pb}^{-1}$ for normalization. SLEUTH is unable to reliably calculate $\mathcal{P}$

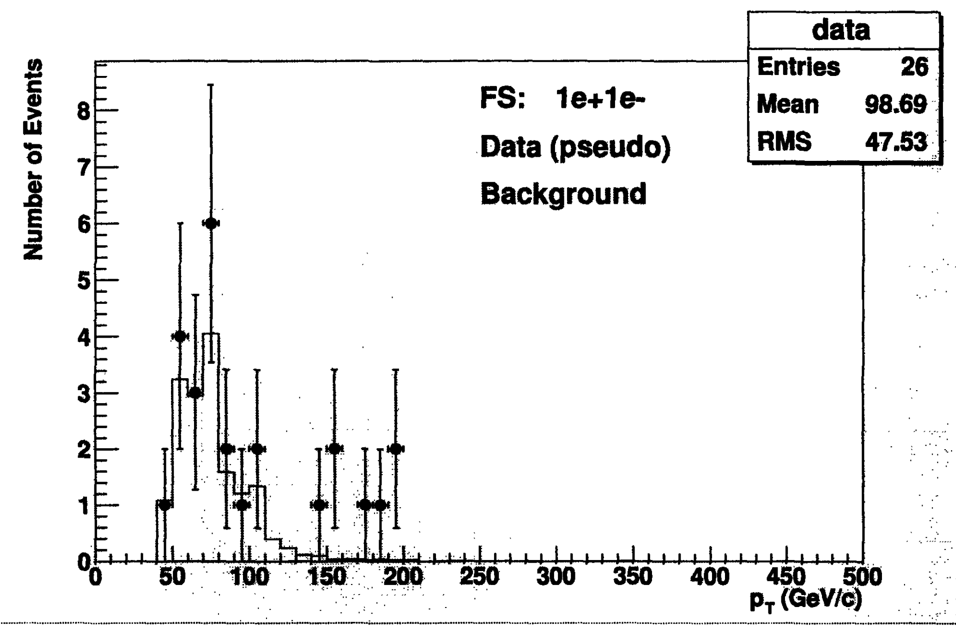


Figure 5-1: Number of events versus $p_T$ for data (signal and peudo data) and SM background for the Sleuth Final State 1e+1e-. The filled blue circles are data points, and the red lines are the expected Standard Model backgrounds as calculated by Pythia. The 7 data events with the highest $p_T$ are the $Z'$ events. $m_{Z'} = 200\text{GeV/c}^2$. Detector simulation by PGS. Using two-step high-$p_T$ background fill-in.

in final states containing a data event lying at larger $\sum p_T$ than the most energetic background event in that final state; in this case additional Monte Carlo is generated to fill in the tail of the background $\sum p_T$ distribution. After this 2nd step, we find that SLEUTH$_{\text{II}}$ can in fact return a meaningful $\mathcal{P}$ value. The capability to generate more MC events in this way is one of the differences between SLEUTH$_{\text{I}}$ and SLEUTH$_{\text{II}}$. We believe this is one of the improvements of SLEUTH$_{\text{II}}$ over SLEUTH$_{\text{I}}$, because we can save CPU time by not blindly generating much more MC events than necessary for the few final states which lack the statistics above a certain $p_T$.

Note in Table 5.1 that SLEUTH combines events with an electron–positron pair with events containing oppositely charged muons into the SLEUTH final state $\ell^+\ell^-$. Beyond-the-standard-model physics that is blind to the difference between the first and second generations are thus put into the same box.

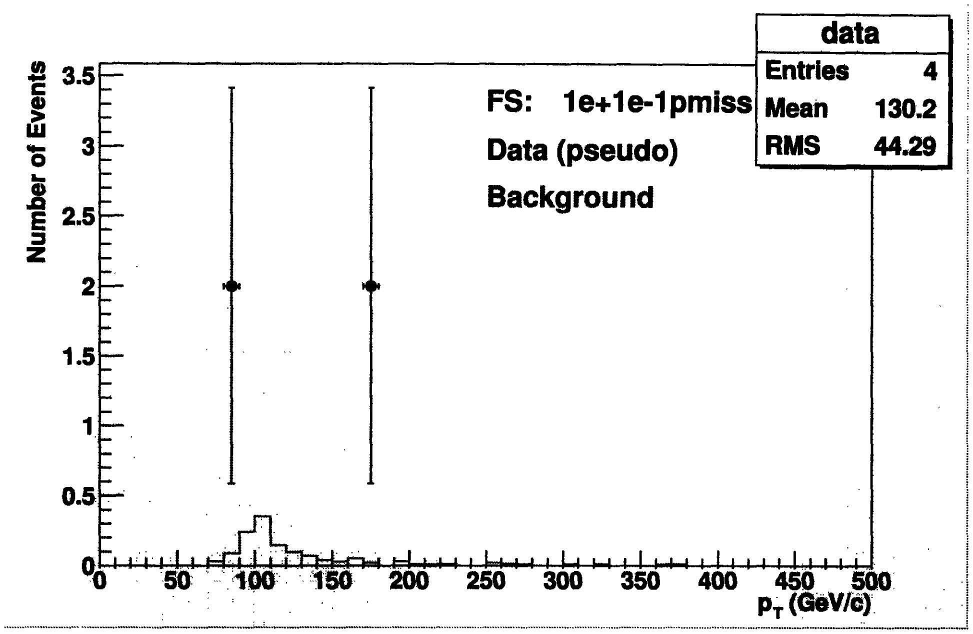


Figure 5-2: Number of events versus $p_T$ for data (signal and peudo data) and SM background. The one $Z'$ signal event had a $p_T = 172$GeV. For this Sleuth FS, Sleuth finds a $\mathcal{P} = 0.03987$, and a $p_T$ threshold of 172 GeV. $m_{Z'} = 200$GeV. Detector simulation by PGS. Using two-step high-$p_T$ background fill-in.

The $\sum p_T$ distribution in the final state $1\ell^+1\ell^-$ (1e+1e-) is plotted in Fig. 5-1. SLEUTH finds $\mathcal{P} < 1.97$e-06, selecting a region $\mathcal{R} = (\sum p_T > 147)$. The number of events expected from Standard Model processes in this region $\mathcal{R}$ is 0.25, with 7 events observed in this pseudo experiment. As expected, SLEUTH finds this final state to have smallest $\mathcal{P}$, the decay $Z' \to e^+e^-$ providing the most obvious signal in the data.

For this example, there were 104 Sleuth Final States with some background events in them, and 23 Sleuth FS with some data events in them. The difference in the number of FS for the two are due to the difference in statistics between the data

| Final state | | Number of Events |
|---|---|---|
| 1e+1e- | | 7 |
| | ( $1e^+1e^-$ | 6 ) |
| | ( $1\mu^+1\mu^-$ | 1 ) |
| 1e+1e-1pmiss | | 1 |
| 1e+1pmiss | | 1 |

Table 5.1: $Z'$ events landing in SLEUTH-considered final states. $m_{Z'} = 200\text{GeV}/c^2$. Detector simulation by PGS.

("pseudo data" and signal), and the MC background. Among the 23 Sleuth FS, $Z'$ events ended up in 3 of them. Sleuth found the 1e+1e- FS as the most interesting, with $\mathcal{P} < 1.97\text{e-}06$. This corresponds to $\tilde{\mathcal{P}} < 0.0002$, exceeding the threshold for discovery.

### $Z'$ ($\mathrm{m}_{Z'} = 400\mathrm{GeV/c^2}$) in $100\mathrm{pb}^{-1}$ of data

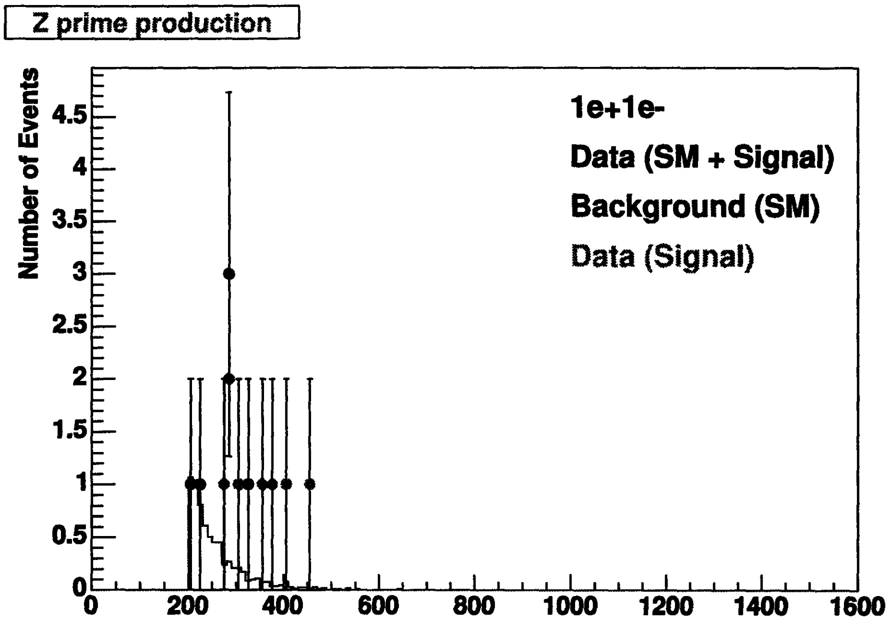


Figure 5-3: Histogram in $\sum p_T$ for data ( $Z'$ signal and peudo data pulled from Standard Model Monte Carlo simulation) and Standard Model background for the Sleuth Final State 1e+1e-. The filled circles are data points, and the solid lines are the expected SM backgrounds as calculated by Pythia. The darker (blue) points are the events from Standard Model Monte Carlo, and the 8 data events are shown in lighter shade (green), with the four highest $\mathrm{p_T}$ events coming from the $Z'$ signal. $\mathrm{m}_{Z'} = 400\mathrm{GeV/c^2}$. Detector simulation by TurboSim ver.1.

Here we generated more plausible signal of $Z'$ with a mass $m_{Z'} = 400\mathrm{GeV}$, a total cross-section of 3pb, corresponding to a $100\mathrm{pb}^{-1}$ of data, which is on the same order of magnitude of the data collected so far by the CDF Collaboration in the Tevatron Run II. A total of 16 events out of 300 generated signal events made it into the Final

| Sleuth Final State | | Number of Events |
|---|---|---|
| 1e+1e- | | 8 events |
| | ( $1e^+1e^-$ | 7 events ) |
| | ( $1\mu^+1\mu^-$ | 1 events ) |
| 1e+1mu-1pmiss | | 1 event |
| 1e+1ph | | 1 event |
| 1e+1ph1jj | | 1 event |
| 1e+1pmiss | | 4 events |
| | ( $1e^+1pmiss$ | 3 events ) |
| | ( $1\mu^+1pmiss$ | 1 events ) |
| 1ph1pmiss1jj | | 1 event |
| Total | | 16 events |

Table 5.2: $Z'$ events which make it into Sleuth Final States. $m_{Z'} = 400\text{GeV}/c^2$. Detector simulation by TurboSim ver.1.

States (FS) as defined for Sleuth, after applying a $\sum p_T > 200 GeV/c$ cut in addition to the cuts from the previous $m_{Z'} = 200$GeV $Z'$ example. We mixed this with "pseudo data" events (events mimicking real data, which are generated with Standard Model Monte Carlo) corresponding to the same 100pb$^{-1}$ of data. We then give SLEUTH$_{\text{II}}$ 10000pb$^{-1}$ of Standard Model background Monte Carlo events, and see if there are any Final States which lack the MC statistics for SLEUTH$_{\text{II}}$ to return a meaningful $\mathcal{P}$ value. For the example considered in this section, we did not have to apply the filling-in of the high-$p_T$ tail. In principle, if there are some Final States for which a meaningful $\mathcal{P}$ value cannot be calculated by SLEUTH$_{\text{II}}$ because of lack of statistics in the Monte Carlo background, we can apply the same method from the previous example to fill in the high-$p_T$ tails of the backgrounds with more events.

The number of events versus $\sum p_T$ for the Sleuth Final State $1\ell^+1\ell^-$ (1e+1e-) is plotted in Fig. 5-3. For this Final State, Sleuth finds $\mathcal{P}$ = 2.2e-4 , and threshold in $\sum p_T$ of 279 GeV. The number of expected background above the threshold $\sum p_T >$ 279 GeV is 1.8 events, with 10 data events observed. Sleuth found this Final State as the most interesting. This was the expected result, since the decay $Z' \to e^+e^-$ is one of the signals of $Z'$ that we look for. It is interesting to note that this $\mathrm{p_T}$ threshold exactly corresponds to the lowest $\sum p_T$ of the $Z'$ signal event we mixed in with the Standard Model pseudo data.

As can be seen in Table 5.2, Sleuth puts the 1e+1e- and $1\mu^+\mu^-$ into the same Final State again. This increases our sensitivity to the signal, when the signal particle can decay into both electron and muon pairs with similar probability. This can be easily understood by considering the fact that the total number of background expected after all the cuts are applied are about 7 events, and only about 2 events above the $\sum p_T$ threshold of 278 GeV, which is the lowest $\sum p_T$ of the $Z'$ signal. Such low background ensures that any additional signal we add by combining the two final states into one Sleuth Final State greatly increases the chances of Sleuth finding the excess of data interesting.

In this example, Sleuth also puts $1\mu^+1pmiss$ into the same Final State as the $1e^+1pmiss$. The merit of combining electrons and muons in the final states into one Sleuth Final State is not so obvious in this case, since SLEUTH$_{\rm II}$ did not find this Final State particularly interesting ($\mathcal{P} = 0.07642$, i.e. probability that a similar experiment would give a more interesting result, meaning at least as many events in this region, is 0.07642). This is not enough to make a statement about the excess indicating a process beyond the Standard Model, since statistical fluctuations in the Standard Model background could cause such a result. Also, the threshold $\sum p_T > 333.6$ found by SLEUTH$_{\rm II}$ does not correspond to the actual lowest $\sum p_T$ of the $Z'$ signal that was introduced, which had $\sum p_T = 283.8 GeV/c$. The 4 $Z'$ signals were lost in the large Standard Model background, which were 114.5 events after all the cuts, and 11.6 events even above the threshold $\sum p_T > 333.6$.

| Sleuth Final State | $\mathcal{P}$ | $p_T$ threshold | $\tilde{\mathcal{P}}$ |
|---|---|---|---|
| 1e+1e- | 0.000224545 | 278.668 | |
| 1e+1mu-1pmiss | 0.171013 | 372.758 | |
| 1e+1ph | 0.165631 | 572.914 | |
| 1e+1ph1jj | 0.0100738 | 339.325 | |
| 1e+1pmiss | 0.0764234 | 333.621 | |
| 1ph1pmiss1jj | 0.397614 | 1424.5 | |
| 1bb1w | 0.298063 | 215.353 | |
| 1e+ | 0.440529 | 301.725 | |
| 1e+1bb | 0.139276 | 261.313 | |
| 1e+1e-1jj | 0.374883 | 399.827 | |
| 1e+1e-1pmiss | 0.180099 | 247.246 | |
| 1e+1e-1pmiss1jj | 0.573888 | 313.025 | |
| 1e+1jj | 0.180018 | 217.878 | |
| 1e+1jj1bb | 0.109409 | 317.048 | |
| 1e+1mu-1pmiss1jj | 0.680272 | 225.201 | |
| 1e+1ph1pmiss | 0.138217 | 770.757 | |
| 1e+1ph1pmiss1jj | 0.928074 | 463.877 | |
| 1e+1ph1pmiss2jj | 0.0449539 | 403.346 | |
| 1e+1ph1pmiss3jj | 0.00297934 | 618.5 | |
| 1e+1ph2jj | 0.262985 | 288.065 | |
| 1e+1pmiss1bb | 0.082713 | 332.623 | |
| 1e+1pmiss1jj | 0.0778513 | 200.878 | |
| 1e+1pmiss1jj1bb | 0.247985 | 294.292 | |
| 1e+1pmiss2jj | 0.418848 | 268.062 | |
| 1e+1tau-1ph1pmiss1jj | 0.0349987 | 316.749 | |
| 1e+1tau-1pmiss | 0.187705 | 301.837 | |
| 1e+2jj | 0.655738 | 282.597 | |
| 1e+2ph | 0.190658 | 212.111 | |
| 1e+2ph1pmiss2jj | 0.041468 | 397.778 | |
| 1jj | 0.382409 | 913.16 | |
| 1jj1w | 0.263158 | 406.052 | |
| 1jj1z | 0.130421 | 423.927 | |
| 1ph | 0.515464 | 568.144 | |
| 1ph1jj | 0.658979 | 493.135 | |
| 1ph1jj1w | 0.2414 | 230.934 | |
| 1ph1pmiss | 0.421053 | 406.288 | |
| 1ph1pmiss1bb | 0.18797 | 1075.48 | |
| 1ph1pmiss1jj | 0.397614 | 1424.5 | |
| 1ph1pmiss2jj | 0.917431 | 1197.39 | |
| 1ph1w | 0.555556 | 203.536 | |
| 1ph2jj | 0.2079 | 455.962 | |
| 1pmiss1jj | 0.0768197 | 947.258 | |
| 1pmiss2jj | 0.600601 | 2322.03 | |
| 1pmiss3jj | 0.387973 | 2195.1 | |
| 1tau-1jj1w | 0.345125 | 239.566 | |
| 1w | 0.211752 | 223.796 | |
| 1z | 0.619195 | 275.89 | |
| 2jj1w | 0.356189 | 298.489 | |
| 2ph1pmiss1jj | 0.597015 | 874.971 | |
| 3ph1pmiss1jj | 0.0434075 | 814.058 | |
| 1e+1e- | 0.000224545 | 278.668 | 0.043 |

Table 5.3: Sleuth results for Final States with the $Z'$ signals (top 6 rows) and other Final States with pseudo data pulled from Standard Model backgrounds. The most interesting Final State was found to be 1e+1e-, as expected, with a $\tilde{\mathcal{P}} = 0.043$. Detector simulation by TurboSim ver.1.

$Z'$ ($\mathrm{m_{Z'}} = 400\mathrm{GeV/c^2}$), **TurboSim ver. 2, Partition Rule ver. 2.0**

| Sleuth Final State | | Number of Events |
|---|---|---|
| 1e+1e- | | 4 events |
| | ( $1e^+1e^-$ | 2 events ) |
| | ( $1\mu^+1\mu^-$ | 2 events ) |
| 1e+1e-1pmiss1tau+ | | 1 event |
| 1e+1ph | | 1 event |
| 1e+1pmiss | | 4 events |
| | ( $1e^+1pmiss$ | 1 events ) |
| | ( $1e^-1pmiss$ | 1 events ) |
| | ( $1j1\mu^-1pmiss$ | 1 events ) |
| | ( $1\mu^-1pmiss$ | 1 events ) |
| 1e+1tau - | | 1 event |
| | ( $1e^+1j1\tau-$ | 1 events ) |
| 1tau-1w | | 1 event |
| | ( $1e^+1pmiss1\tau+$ | 1 events ) |
| Total | | 12 events |

Table 5.4: $Z'$ events which make it into Sleuth Final States. Detector simulation done with TurboSim ver.2 (dated Apr 25). Partition Rule ver. 2.0 was applied for Sleuth.

Tests similar to previous ones were performed using TurboSim ver. 2.0 and Partition Rule ver. 2.0, the final versions used in this study, with $m_{Z'} = 400\mathrm{GeV}$, and total cross section $\sigma_{Z'} = 3\mathrm{pb}$, corresponding to $100\mathrm{pb}^{-1}$ of data. Table 5.4 shows how the $Z'$ signals were partitioned into Sleuth Final States. Table 5.5 is a summary of the SLEUTH results. SLEUTH found the 1e+1e- Final State most interesting, with $\mathcal{P} = 5.7e{-}05$, in the region $\sum p_T > 351GeV$. This indicates that the $Z'$ signal seen in this Final State is very unlikely to be a statistical fluctuation of the Standard Model background. But the final value of $\tilde{\mathcal{P}} = 0.0025$, taking into account all the other Final States searched by SLEUTH indicates that it is at the border of "discovery", and no definite claim can be made in this case.

As can be seen in Figure 5-4, the threshold $\sum p_T > 351GeV$ found by SLEUTH for the 1e+1e- Final State does not correspond to a signal data point, but a pseudo data point slightly below the $Z'$ signal with the lowest $\sum p_T$. It did include all the signal data points in the region of interest. Figure 5-5 shows the $\sum p_T$ distribution of data and background for the 1 $e/\mu$ and $\not\!\!E_T$ (1e+1pmiss) Final State. Notice that most of the $Z'$ signal events were not included in the region of interest ($\sum p_T > 402GeV$), but lost to the large background events around them. The rather large value of $\mathcal{P} = 0.3295$ reflects this. Both Final States had 4 $Z'$ signals mixed with the pseudo data generated from Standard Model processes.

Though we cannot claim a discovery at this point, it is interesting to note that the SLEUTH algorithm can find an excess in data relative to the Standard Model background with only 4 signal events and 6 data points in a region in the case of the 1e+1e- Final State. This is partly due to the fact that the 1e+1e- is such a "clean" Final State with not much Standard Model background around the signal points. It shows where SLEUTH is most effective in finding a deviation from the Standard Model, which is part of the purpose of this study. SLEUTH does a very good job of finding an excess in data when the $\sum p_T$ of the signal is large compared to the background events as in the 1e+1e- Final State. But it does not do such a good job when the signal events are mixed with background events over a wide range of $\sum p_T$ as in the 1e+1pmiss Final State. In other words, the algorithm is less sensitive to an overall normalization difference in the number of events.

This reflects one of the assumptions behind the algorithm, that new physics processes will generally produce objects with larger energy than expected from background processes, which will appear as events with large $\sum p_T$ in the high-$p_T$ end of the distribution. SLEUTH is good at finding those signal events. If a new physics process produces objects which would naturally be in different Final States, i.e. it has many decay chains, SLEUTH will be most likely to find an excess in Final States where the $\sum p_T$ of the signal events is large compared to that of the background events.

Another point of notice with the $Z'$ example is that this example shows whether our choice of partition rule regarding the lepton generation and charge conjugation was a sound one. The partition rule of SLEUTH groups into the same Final State, events which differ only in the 1st and 2nd generation of the leptons, or the charge sign of the leptons. If we had not done this, the events in the 1e+1e- and 1e+1pmiss

Final States could have been thrown away depending on the background data points, because they may not have been able to pass the minimum number of events requirement of the SLEUTH algorithm. But because of the two rules, the signal events will pass the minimum number of events requirement, regardless of the background events in the two Final States.

Most of the initial debugging and testing of various changes made to the analysis package including SLEUTH was performed by looking at the $Z'$ example, due to its clean signature. Many versions of the Partition Rule exists for the $Z'$ example, but obmitted here since they differ only slightly among them in the final SLEUTH results, and do not add to our understanding of the $Z'$ example.

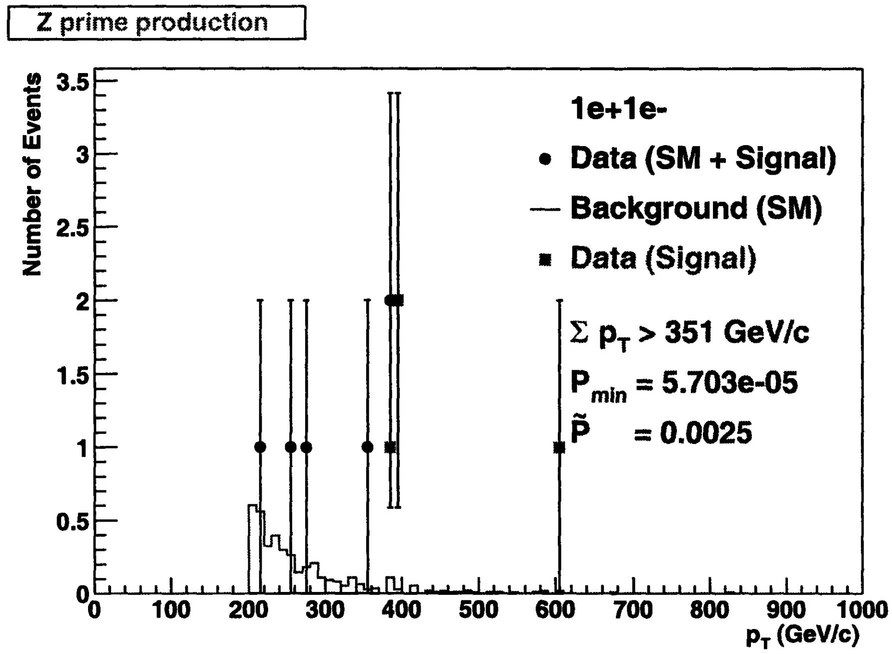


Figure 5-4: Histogram in $\sum p_T$ for data ( $Z'$ "signal" and peudo data pulled from Standard Model Monte Carlo simulation) and Standard Model background for the Sleuth Final State 1e+1e-. The dark (blue) filled circles are data points including both the Standard Model and Signal events, the light (green) squares are "signal" events only, and the solid (red) lines are the expected SM backgrounds as calculated by Pythia. CDF detector simulation done with TurboSim ver.2 (dated Apr 25). Partition Rule ver. 2.0 was applied for Sleuth.

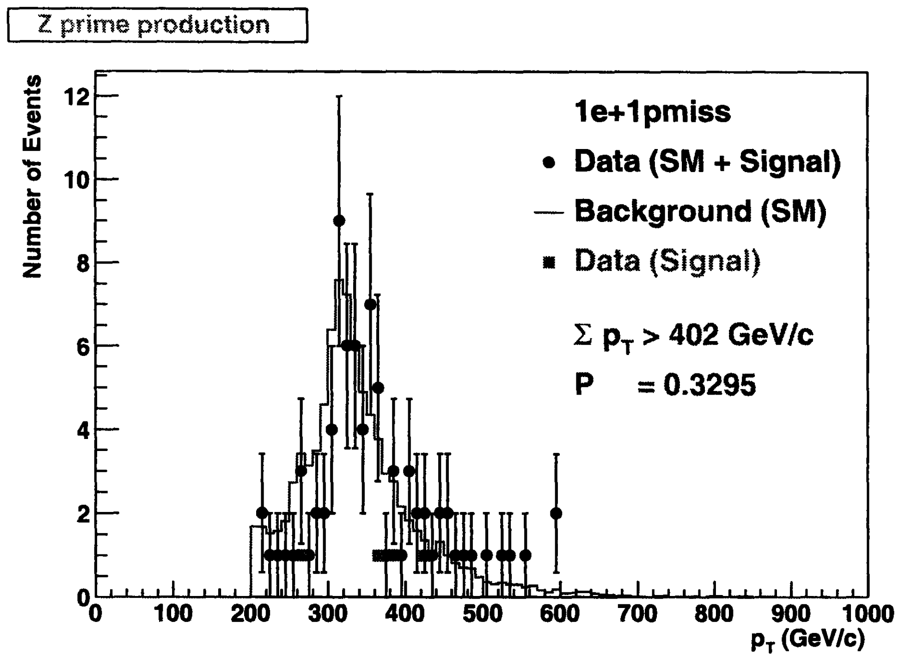


Figure 5-5: Histogram in $\sum p_T$ for data ( $Z'$ "signal" and peudo data pulled from Standard Model Monte Carlo simulation) and Standard Model background for the Sleuth Final State 1e+1pmiss. The dark (blue) filled circles are data points including both the Standard Model and Signal events, the light (green) squares are "signal" events only, and the solid (red) lines are the expected SM backgrounds as calculated by Pythia. CDF detector simulation done with TurboSim ver.2 (dated Apr 25). Partition Rule ver. 2.0 was applied for Sleuth.

| Sleuth Final State | $\mathcal{P}$ | $p_T$ threshold | $\tilde{\mathcal{P}}$ |
|---|---|---|---|
| 1e+1e- | 5.70288e-05 | 350.560 | |
| 1e+1pmiss | 0.329489 | 401.868 | |
| 1e+1ph | 0.0991818 | 200.636 | |
| 1tau+1w | 0.034229 | 266.435 | |
| 1e+ | 0.20202 | 223.318 | |
| 1e+1e-1jj1pmiss | 0.420168 | 363.526 | |
| 1e+1e-1pmiss | 0.505689 | 228.851 | |
| 1e+1jj | 0.606061 | 356.223 | |
| 1e+1jj1ph | 0.239808 | 213.675 | |
| 1e+1jj1pmiss | 0.506971 | 220.704 | |
| 1e+1pmiss1tau+ | 0.405268 | 244.539 | |
| 1jj | 0.909091 | 939.564 | |
| 1jj1ph1w | 0.077821 | 430.643 | |
| 1jj1pmiss | 0.0670466 | 1048.76 | |
| 1jj1tau-1w | 0.0233618 | 273.226 | |
| 1jj1w | 0.165153 | 303.887 | |
| 1jj1z | 0.585652 | 293.418 | |
| 1ph | 0.637959 | 460.802 | |
| 1ph1w | 0.192771 | 203.164 | |
| 1w | 0.959233 | 249.434 | |
| 1z | 0.621118 | 244.288 | |
| 2jj1w | 0.168776 | 334.448 | |
| 1e+1e- | 5.70288e-05 | 350.560 | 0.0025 |

Table 5.5: Sleuth results for Final States with the $Z'$ signals (top 4 rows) and other Final States with pseudo data pulled from Standard Model backgrounds. The most interesting Final State was found to be 1e+1e-, as expected, with a $\tilde{\mathcal{P}} = 0.0025$. The $\sum p_T$ of the $Z'$ signal was found by the Sleuth algorithm to be approximately 350 GeV/c, a value which is close to the invariant mass of the $Z'$, $m_{Z'} = 400$GeV. Detector simulation done with TurboSim ver.2 (dated Apr 25). Partition Rule ver. 2.0 was applied for Sleuth.

## 5.2 $W'$

The $W'$ is a charged, heavy ($M_{W'} > M_W$) vector boson that appears in certain attempts to expand the $SU(2)_L \times U(1)_Y$ electroweak group in the Standard Model, such as the left-right symmetric model [15].

The $W'$ signals for this example was generated with an invariant mass of $m_{W'} = 500GeV/c^2$ and a cross-section of 3 pb. The choice of these numbers were based on the previous searches [16, 17, 18, 19] for the $W'$ boson by the CDF Collaboration. Previous searches returned null results, excluding this mass. But this choice allows comparison with those previous searches. The total number of events used in this example are comparable to [16], within about a factor of 2, though the decay channels considered are quite different. Thus, the hypothetical $W'$ "signal" we are considering is similar in its magnitude to those considered elsewhere.

The generated $W'$ signals are given in Table 5.6 (for Partition Rule ver. 1.0) and in Table 5.8 (for Partition Rule ver. 2.0). Both examples were passed through TurboSim ver. 2.0 for detector simulation, and only differ in the rules applied for the partitioning. Partition Rule ver. 2.0 refers to the final partitioning rule applied to all the final examples considered in this thesis.

As can be seen in the two tables, the number of events which end up in the two final states do not differ in this case between the tables, since they were passed through identical detector simulation and the changes in Partition Rule was in the treatment of jets (j's) and b-tagged jets (b's), in W and Z mass ranges, and $\not{E}_T$ lower bound cut (which had no effect here). Some more detail is given in Table 5.8, to illustrate which final states were partitioned together into a single Sleuth Final State.

| Sleuth Final State | Number of Events |
|---|---|
| 1e+1jj1pmiss | 1 event |
| 1e+1pmiss | 31 events |
| Total | 32 events |

Table 5.6: $W'$ events which make it into Sleuth Final States. CDF detector simulation done with TurboSim ver.2 (dated Apr 25). Partition Rule ver. 1.0 was applied for Sleuth.

| Sleuth Final State | $\mathcal{P}$ | $p_T$ threshold | $\tilde{\mathcal{P}}$ |
|---|---|---|---|
| 1e+1pmiss | $< 5.68721e-06$ | 417.827 | |
| 1e+1jj1pmiss | 0.178333 | 211.495 | |
| 1e+ | 0.623053 | 220.358 | |
| 1e+1e- | 0.405268 | 325.532 | |
| 1e+1e-1jj1pmiss | 0.338983 | 363.526 | |
| 1e+1e-1pmiss | 0.835073 | 228.851 | |
| 1e+1jj | 0.650407 | 307.838 | |
| 1e+1jj1ph | 0.323887 | 206.608 | |
| 1e+1ph | 0.091638 | 200.636 | |
| 1e+1pmiss1tau+ | 0.151688 | 257.818 | |
| 1e+2jj1pmiss | 0.371058 | 334.448 | |
| 1jj | 0.911162 | 939.564 | |
| 1jj1pmiss | 0.0697837 | 1048.76 | |
| 1jj1w | 0.852878 | 225.814 | |
| 1jj1z | 0.959233 | 293.418 | |
| 1ph | 0.671141 | 460.802 | |
| 1w | 0.567376 | 215.748 | |
| 1z | 0.154679 | 248.244 | |
| 1e+1pmiss | $< 5.68721e-06$ | 417.827 | $< 0.0003$ |

Table 5.7: Sleuth results for Final States with the $W'$ signals (top 2 rows) and other Final States with pseudo data pulled from Standard Model backgrounds. The most interesting Final State was found to be 1e+1pmiss, as expected, with a $\tilde{\mathcal{P}} < 0.0003$, exceeding the threshold for discovery. CDF detector simulation done with TurboSim ver.2 (dated Apr 25). Partition Rule ver. 1.0 was applied for Sleuth.

| Sleuth Final State | | Number of Events |
|---|---|---|
| 1e+1jj1pmiss | | 1 event |
| 1e+1pmiss | | 31 events |
| | ( $1e^{+}1pmiss$ | 6 events ) |
| | ( $1e^{-}1pmiss$ | 9 events ) |
| | ( $1\mu^{+}1pmiss$ | 4 events ) |
| | ( $1\mu^{-}1pmiss$ | 7 events ) |
| | ( $1j1\mu^{+}1pmiss$ | 3 events ) |
| | ( $1j1\mu^{-}1pmiss$ | 2 events ) |
| Total | | 32 events |

Table 5.8: $W'$ events which make it into Sleuth Final States. CDF detector simulation done with TurboSim ver.2 (dated Apr 25). Partition Rule ver. 2.0 was applied for Sleuth.

In Table 5.8, we see that the Sleuth Final State 1e+1pmiss includes both the charge conjugated leptons – $e^+$ and $e^-$, and $\mu^+$ and $\mu^-$ – in the same Final State. Because of this choice, we have increased the number of events in the 1e+1pmiss Final State by about a factor of 2, compared to what it would have been had we not combined the two charge signs.. We also see that the Sleuth Final State 1e+1pmiss includes both the 1st and 2nd generation leptons – $e$'s and $\mu$'s – in the same Final State, which also increases the number of events in the 1e+1pmiss Final State by about a factor of 2, compared to what it would have been otherwise. This effectively placed the signal events into one single Final State, thus increasing the likelihood that SLEUTH would find that Final State interesting, compared to the other Final States which does not have signal events in the data. Thus, this example supports the answer to two of the open questions posed at the beginning of this study.

We see that the two partitioning rules combined – placing charge conjugated leptons into the same Final State, and 1st and 2nd generation leptons into the same Final State – have increased the number of events in one "signal" Final State by about a factor of 4. They have also saved the events which would have been discarded by the minimum number of events requirement in each final state, which is 4. The five events with one jet each ($1j1\mu^{+}1pmiss, 1j1\mu^{-}1pmiss$) would have been discarded if they had been partitioned into a Final State by themselves without the charge conjugation rule. But here they contribute to the single 1e+1pmiss SLEUTH Final

| Sleuth Final State | $\mathcal{P}$ | $p_T$ threshold | $\tilde{\mathcal{P}}$ |
|---|---|---|---|
| 1e+1pmiss | $< 4.68274e-06$ | 417.827 | |
| 1e+1jj1pmiss | 0.468933 | 220.704 | |
| 1e+ | 0.20202 | 223.318 | |
| 1e+1e- | 0.193237 | 350.56 | |
| 1e+1e-1jj1pmiss | 0.439078 | 363.526 | |
| 1e+1e-1pmiss | 0.503145 | 228.851 | |
| 1e+1jj | 0.585652 | 356.223 | |
| 1e+1jj1ph | 0.243605 | 213.675 | |
| 1e+1ph | 0.151172 | 200.636 | |
| 1e+1pmiss1tau+ | 0.392927 | 244.539 | |
| 1jj | 0.938967 | 939.564 | |
| 1jj1ph1w | 0.086003 | 430.643 | |
| 1jj1pmiss | 0.0591628 | 1048.76 | |
| 1jj1tau-1w | 0.0226232 | 273.226 | |
| 1jj1w | 0.182815 | 303.887 | |
| 1jj1z | 0.594354 | 293.418 | |
| 1ph | 0.647249 | 460.802 | |
| 1ph1w | 0.191755 | 203.164 | |
| 1tau+1w | 0.032486 | 266.435 | |
| 1w | 0.968523 | 249.434 | |
| 1z | 0.614439 | 244.288 | |
| 2jj1w | 0.17567 | 334.448 | |
| 1e+1pmiss | $< 4.68274e-06$ | 417.827 | $< 0.00023$ |

Table 5.9: Sleuth results for Final States with the $W'$ signals (top 2 rows) and other Final States with pseudo data pulled from Standard Model backgrounds. The most interesting Final State was found to be 1e+1pmiss, as expected, with a $\tilde{\mathcal{P}} < 0.00023$, exceeding the threshold for discovery. CDF detector simulation done with TurboSim ver.2 (dated Apr 25). Partition Rule ver. 2.0 was applied for Sleuth.

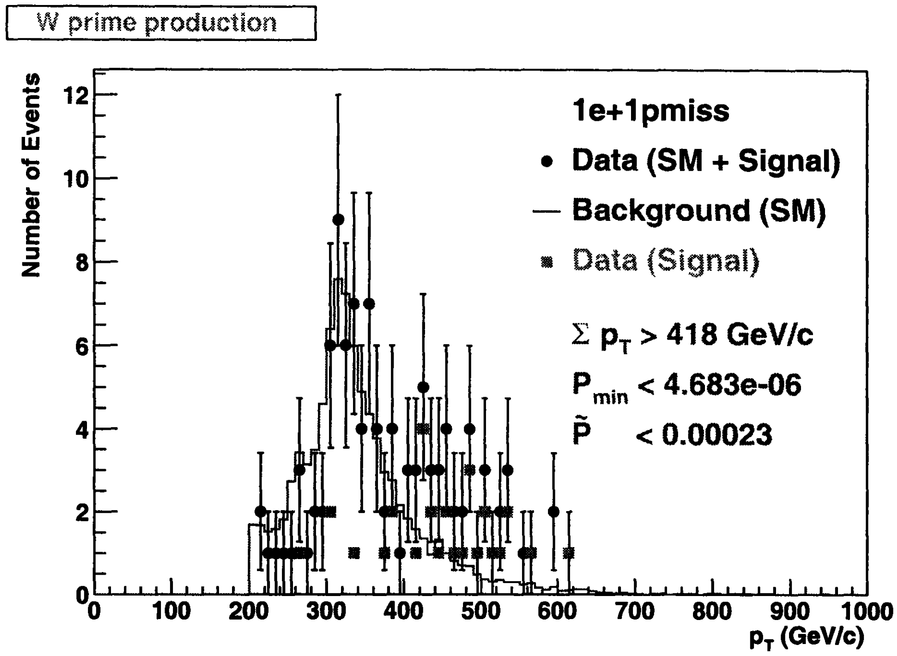


Figure 5-6: Histogram in $\sum p_T$ for data ( $W'$ "signal" and peudo data pulled from Standard Model Monte Carlo simulation) and Standard Model background for the Sleuth Final State 1e+1pmiss. The dark (blue) filled circles are data points including both the Standard Model and "signal" events, the light (green) squares are "signal" events only, and the solid (red) lines are the expected SM backgrounds as calculated by Pythia. CDF detector simulation done with TurboSim ver.2 (dated Apr 25). Partition Rule ver. 2.0 was applied for Sleuth.

State. The $W'$ is a very good example of how the Partition Rule defined for Sleuth does a very good job of putting all the signal events into one Final State.

The two Final States 1e+1pmiss and 1e+1jj1pmiss, which have $W'$ signal in them, are shown in Figure 5-6 and Figure 5-7. The Final State 1e+1jj1pmiss has only a single $W'$ signal event, and the $\sum p_T > 221$ GeV/c threshold found by Sleuth does not correspond to the $W'$ signal, but a fluctuation in data drawn from the Standard Model background. The $\mathcal{P} = 0.4689$ reflects this, indicating that the probability that the background fluctuation could give such a result is 0.4689, or about half the time.

From the Sleuth results in Table 5.9, the Final State 1e+1pmiss shown in Figure 5-6, which has 31 $W'$ signal events in it, has a $\sum p_T > 418$ GeV/c threshold found by Sleuth. This corresponds correctly to the $W'$ signal in the Final State, as shown in the Figure 5-6 by the lighter shaded (green) "Data(Signal)" points, and the $\mathcal{P}$ <

$4.683e-06$ means that the probability that a background could fluctuate to give such a signal is less than $4.683e-06$, which is not very likely. The probability that a hypothetical similar experiment would give a result which is more interesting than what we have is $\tilde{\mathcal{P}} < 0.00023$, which is less than $1/1000$, and corresponds to a discovery.

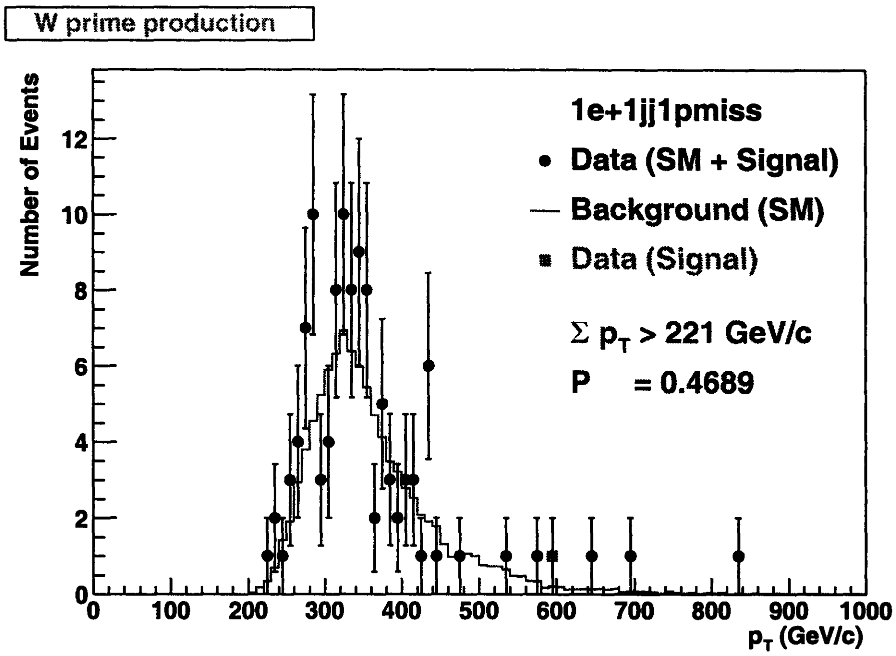


Figure 5-7: Histogram in $\sum p_T$ for data ( $W'$ "signal" and peudo data pulled from Standard Model Monte Carlo simulation) and Standard Model background for the Sleuth Final State 1e+1jj1pmiss. The dark (blue) filled circles are data points including both the Standard Model and "signal" events, the light (green) squares are "signal" events only, and the solid (red) lines are the expected SM backgrounds as calculated by Pythia. CDF detector simulation done with TurboSim ver.2 (dated Apr 25). Partition Rule ver. 2.0 was applied for Sleuth.

## 5.3 Leptoquark

The Leptoquark signals for this example were generated with an invariant mass of $m_{LQ} = 200GeV/c^2$ and a cross-section of 1 pb. The choice of these numbers were based on the previous results from H1 and ZEUS experiments at HERA, and CDF and DØ at Tevatron, where the observation of excess signals or limits put on the mass of the leptoquarks was around $200GeV/c^2$, and the cross-section was less than 5 pb. [20, 21, 22, 23, 24, 25, 26]

The Leptoquark signal generated for this example were pair-produced as LQ and $\overline{\text{LQ}}$, and decays into $e^-qe^+\bar{q}$ at the generator level, where $q$ and $\bar{q}$ are quark and anti-quark. This would ideally show up in the detector as a 1e+1e-1jj Final State, with the pair of jets coming from the quarks. In reality, one of the electrons may not be found by the detector due to efficiency being less than unity, and extra jets coming from the proton remnants or radiated gluons may appear. Any object which is present at the parton level which are not found at the detector level would show up as pmiss, since any missing object would mean the vector sum of the $p_T$ would not be zero.

Table 5.10 shows the number of leptoquark signal events for each of the Final States which have at least one signal events when the detector simulation was done with TurboSim ver.1. Table 5.11 and Table 5.13 shows the same when the detector simulation was done with TurboSim ver.2, the difference being that Table 5.11 is when Partition Rule ver.1 was applied, and Table 5.13 is when Partition Rule ver.2 was applied.

Comparison between the first two tables shows that the number of Final States with 3 or 4 jet pairs in them are much more frequent in the set generated with TurboSim ver. 1 than the set generated with TurboSim ver. 2. The two sets had the same Partition Rule ver. 1 applied to them. This indicates that the jets were simulated with too high $p_T$. From Table 5.10, we see that there are many cases where the objects in a given Final State differ only by the number of jet pairs in the Final

| Sleuth Final State | Number of Events |
|---|---|
| 1e+1e-1jj | 2 |
| 1e+1e-1ph1jj | 1 |
| 1e+1e-1ph1pmiss2jj | 1 |
| 1e+1e-1ph1pmiss2jj1bb | 1 |
| 1e+1e-1ph1pmiss3jj | 3 |
| 1e+1e-1ph1pmiss4jj | 1 |
| 1e+1e-1ph2jj | 2 |
| 1e+1e-1pmiss1jj | 2 |
| 1e+1e-1pmiss1jj1bb | 1 |
| 1e+1e-1pmiss2jj | 4 |
| 1e+1e-1pmiss3jj | 6 |
| 1e+1e-2jj | 4 |
| 1e+1e-3jj | 1 |
| 1e+1e-4jj | 1 |
| 1e+1ph1jj | 4 |
| 1e+1ph1pmiss1jj | 2 |
| 1e+1ph1pmiss1jj1bb | 1 |
| 1e+1ph1pmiss2jj | 7 |
| 1e+1ph1pmiss3jj | 3 |
| 1e+1ph1pmiss4jj | 4 |
| 1e+1ph2jj | 3 |
| 1e+1ph3jj | 1 |
| 1e+1pmiss2jj | 8 |
| 1e+1pmiss3jj | 1 |
| 1e+1pmiss4jj | 1 |
| 1e+1pmiss5jj | 3 |
| 1e+2jj | 2 |
| 1e+2ph1pmiss2jj | 2 |
| 1e+2ph1pmiss3jj | 1 |
| 1e+2ph1pmiss3jj1bb | 1 |
| 1e+2ph2jj | 1 |
| 1e+3jj | 3 |
| 1e+5jj | 1 |
| 1pmiss1jj | 1 |
| 2jj1z | 1 |
| 3ph1pmiss2jj | 1 |
| Total | 82 events |

Table 5.10: Leptoquark events which make it into Sleuth Final States. Detector simulation done with TurboSim v.1 (dated Feb 16)

| Sleuth Final State | Number of Events |
|---|---|
| 1e+1e-1jj1pmiss | 5 |
| 1e+1e-2jj | 7 |
| 1e+1e-2jj1pmiss | 8 |
| 1e+1jj | 1 |
| 1e+1jj1ph | 1 |
| 1e+1jj1ph1pmiss | 1 |
| 1e+1jj1pmiss | 12 |
| 1e+2jj | 4 |
| 1e+2jj1ph | 5 |
| 1e+2jj1ph1pmiss | 3 |
| 1e+2jj1pmiss | 18 |
| 1e+2jj1pmiss1tau+ | 1 |
| 1e+3jj1ph1pmiss | 1 |
| 1jj1pmiss | 1 |
| 2jj1pmiss | 2 |
| 2jj1w | 1 |
| Total | 71 |

Table 5.11: Leptoquark events which make it into Sleuth Final States. Detector simulation done with TurboSim v.2 (dated Apr 25). Partition Rule ver. 1.0 was applied for Sleuth.

| Sleuth Final State | $\mathcal{P}$ | $p_T$ threshold | $\tilde{\mathcal{P}}$ |
|---|---|---|---|
| 1e+1e-1jj1pmiss | $< 5.98192e-06$ | 1486.43 | |
| 1e+1e-2jj | $< 8.12248e-08$ | 979.854 | |
| 1e+1e-2jj1pmiss | $< 8.62824e-08$ | 1456.02 | |
| 1e+1jj | $< 4.49897e-05$ | 1162.97 | |
| 1e+1jj1ph | $< 1.04651e-06$ | 1009.12 | |
| 1e+1jj1pmiss | $< 5.30438e-05$ | 685.824 | |
| 1e+2jj | $< 1.7984e-07$ | 917.298 | |
| 1e+2jj1ph | $< 8.0825e-08$ | 733.62 | |
| 1e+2jj1pmiss | $< 5.27534e-07$ | 717.134 | |
| 1e+2jj1ph | $< 8.0825e-08$ | 733.62 | $< 0.0000055$ |

Table 5.12: Sleuth results for some of the Final States found to be interesting. The most interesting Final State was found to be 1e+2jj1ph, with a $\tilde{\mathcal{P}} < 0.0000055$, exceeding the threshold for discovery.

| Sleuth Final State | Number of Events |
|---|---|
| 1e+1e-1jj1pmiss | 5 |
| 1e+1e-2jj | 6 |
| 1e+1e-2jj1pmiss | 8 |
| 1e+1jj | 1 |
| 1e+1jj1ph | 1 |
| 1e+1jj1ph1pmiss | 1 |
| 1e+1jj1pmiss | 9 |
| 1e+2jj | 5 |
| 1e+2jj1ph | 5 |
| 1e+2jj1ph1pmiss | 3 |
| 1e+2jj1pmiss | 14 |
| 1e+2jj1pmiss1tau+ | 1 |
| 1jj1pmiss | 1 |
| 1jj1w | 3 |
| 2jj1pmiss | 2 |
| 2jj1w | 4 |
| 2jj1z | 1 |
| 3jj1ph1w | 1 |
| Total | 71 |

Table 5.13: Leptoquark events which make it into Sleuth Final States. Detector simulation done with TurboSim ver.2 (dated Apr 25). Partition Rule ver. 2.0 was applied for Sleuth. The effect of this was to change 1 event from 1e+1e-2jj to 2jj1z, 3 events from 1e+1jj1pmiss to 1jj1w, 1 event from 1e+2jj1pmiss to 1e+2jj, 3 events from 1e+2jj1pmiss to 2jj1w, 1 event from 1e+3jj1ph1w to 3jj1ph1w.

| Sleuth Final State | $\mathcal{P}$ | $p_T$ threshold | $\tilde{\mathcal{P}}$ |
|---|---|---|---|
| 1e+1e-1jj1pmiss | $< 1.58327e-05$ | 1486.43 | |
| 1e+1e-2jj | $< 8.01948e-08$ | 979.854 | |
| 1e+1e-2jj1pmiss | $< 8.83873e-08$ | 1456.02 | |
| 1e+1jj | $< 4.26536e-05$ | 1162.97 | |
| 1e+1jj1ph | $< 4.18679e-07$ | 1009.12 | |
| 1e+1jj1pmiss | $< 4.52448e-05$ | 830.237 | |
| 1e+2jj | $< 1.16409e-07$ | 1141.83 | |
| 1e+2jj1ph | $< 7.95616e-08$ | 733.62 | |
| 1e+2jj1pmiss | $< 2.10208e-07$ | 727.794 | |
| 1jj1w | $< 5.22739e-05$ | 742.552 | |
| 2jj1w | $< 5.39492e-07$ | 1027.31 | |
| 1e+2jj1ph | $< 7.95616e-08$ | 733.62 | $< 0.0000052$ |

Table 5.14: Sleuth results for some of the Final States found to be interesting. The most interesting Final State was found to be 1e+2jj1ph, with a $\tilde{\mathcal{P}} < 0.0000052$, exceeding the threshold for discovery.

State, such as the 1e+1e-1ph1pmissNjj, where N is the number of jet pair. The fact that there are cases where the number of events in a Final State with a larger number of jet pairs is greater than the number of events in a similar Final State with smaller number of jet pairs means that the number of jets simulated by TurboSim for this example has some problems, since we expect it to generally drop without a clear physics process to support the larger jet pairs. Also, Final States with 1bb means that they had a jet pair which was identified to be from b quarks. Since we know the Leptoquark signal should not have b's in them, we see that the simulation of b's in TurboSim for the Leptoquark example also has a problem. Note also that many Final States have 1pmiss in them, indicating that the $\not{E}_T$ isn't handled very well by TurboSim. The problems mentioned here were attributed to the lack of statistics of high $p_T$ jet pairs generated with CDFSim to train TurboSim; the needs to refine the handling of $\not{E}_T$ and unclustered energies which make up the $\not{E}_T$, and of jets coming from b quarks. Note that all these problems are closely related to the handling of jets in TurboSim.

The two later tables, Table 5.11 and Table 5.13 with just different Partition Rule applied, have the same number of total events, as they should, with the events redistributed among the Final States. We see that the main effect of the changes made in TurboSim ver. 2 is to reduce the number of events in Final States with the object pmiss, and moved them into Final States with a W or a Z boson. These were two of the goals of the changes made, so we expected this to happen. Whether this is a positive effect needs to be discussed. Leptoquarks do not decay into W or Z bosons at the generator level. But the purpose of the changes were to have as many events which could have come from the W or Z boson in a Final State with a W or Z. Since that means increasing the $p_T$ range of the objects which make up the W or Z, it will have side effects as shown here. But as long as these effects do not influence the final Sleuth outcome in a significant way, i.e. change the most interesting Final State found by Sleuth or the $\tilde{\mathcal{P}}$ value significantly, a minor redistribution of events should be negligible. We see that is the case here by comparing the final Sleuth results from Table 5.12 and Table 5.14. The most interesting Final State is found to be 1e+2jj1ph with $\tilde{\mathcal{P}} < 0.0000055$ for Partition Rule ver. 1 and $\tilde{\mathcal{P}} < 0.0000052$ for Partition Rule ver. 2.

Figure 5-8, Figure 5-9, Figure 5-10 and Figure 5-11 show $\sum p_T$ distribution of four Final States from Table 5.14. Figure 5-8 is for the 1e+2jj1ph Final State. This Final State was found to be the most interesting Final State by Sleuth with $\tilde{\mathcal{P}} < 5.2e-06$, which would be identified as a new discovery. We see that there are extra jets in the events, which were not expected, as well as photons which are mostly electrons mis-identified as photons.

Figure 5-9 is for the 1e+1e-1jj1pmiss Final State, one of the Final States in which we would expect a large leptoquark signal. This is because, except for the $\not{E}_T$ (1pmiss) it is very similar to the 1e+1e-1jj Final State, which is ideally the Final State we would expect most of the signal events to be in. But from Table 5.14 we see that is not the case. The two most populated Final States were 1e+1jj1pmiss shown in Figure 5-10, and 1e+2jj1pmiss shown in Figure 5-11. The events in 1e+1jj1pmiss Final State has one electron which were lost in the detector, and a resulting $\not{E}_T$ (1pmiss). The events in 1e+2jj1pmiss have an extra jet pair in addition to the the missing electron and an extra $\not{E}_T$.

We see that the interpretation of the "signal" in the real data, as to which physics process is responsible for the excess in data, will not be simple in the presence of $\not{E}_T$ due to detector effects or extra jets radiated from the primary partons coming from the interaction. The leptoquark example was the most problematic and difficult to interpret among the various examples included in this study. It was also the example which TurboSim seemed to have the most difficulty in simulating the detector effects, since the $\sum p_T$ of TurboSim ver. 1 had not been convincing, and much more events had $\not{E}_T$ (pmiss) in them for both versions of TurboSim than could be accounted for by simple detector inefficiencies. Still, the outcome of TurboSim can be understood as a need to adjust the components responsible for the jet energies and the handling of $\not{E}_T$ (pmiss), and a need to adjust the efficiency for electrons with large rapidity ($|\eta| \geq 1$).

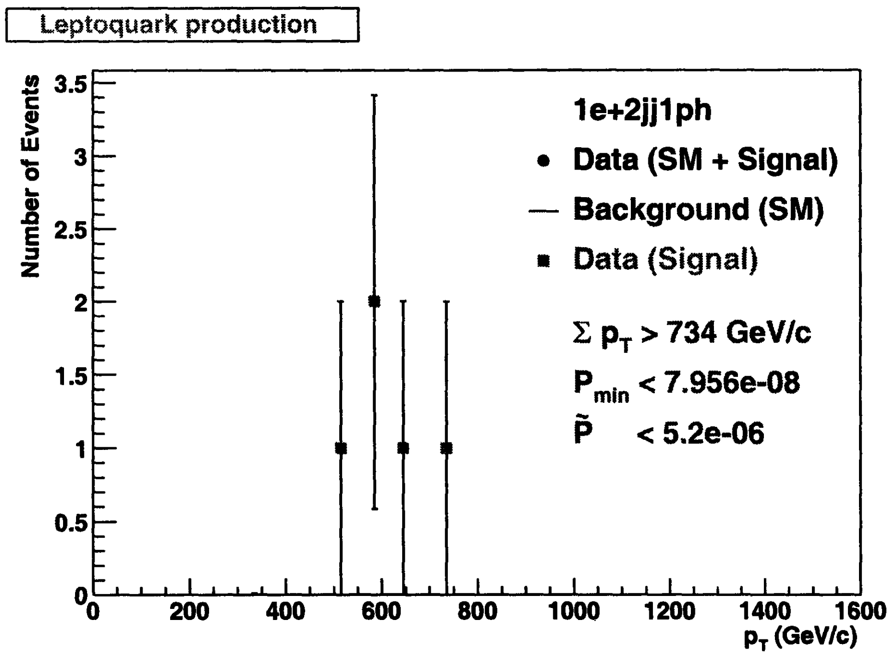


Figure 5-8: Histogram in $\sum p_T$ for data ( Leptoquark "signal" and peudo data pulled from Standard Model Monte Carlo simulation) and Standard Model background for the Sleuth Final State 1e+2jj1ph. The dark (blue) filled circles are data points including both the Standard Model and "signal" events, the light (green) squares are "signal" events only, and the solid (red) lines are the expected SM backgrounds as calculated by Pythia. (CDF detector simulated with TurboSim ver.2. Partition Rule ver. 2.0)

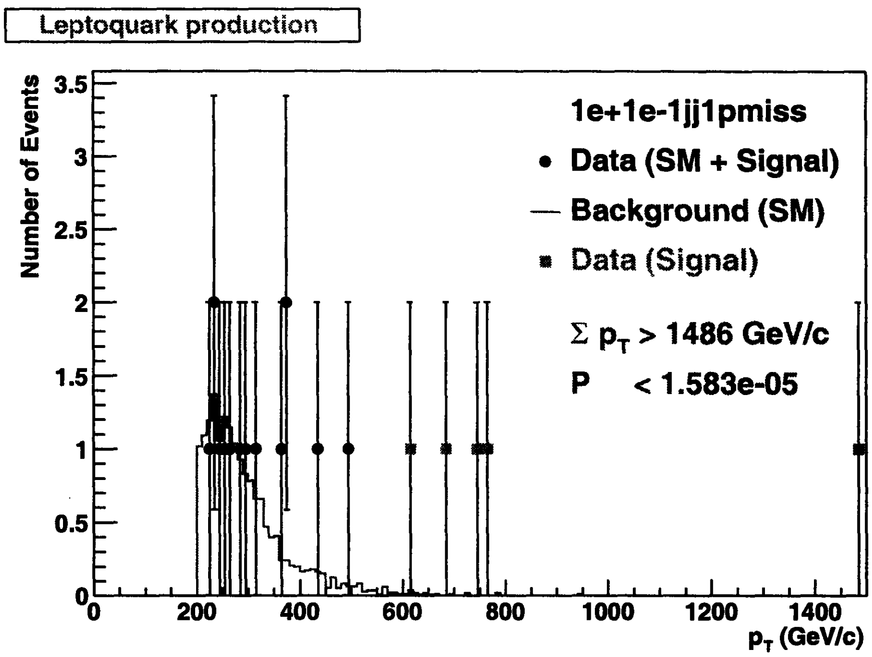


Figure 5-9: Histogram in $\sum p_T$ for data ( Leptoquark "signal" and peudo data pulled from Standard Model Monte Carlo simulation) and Standard Model background for the Sleuth Final State 1e+1e-1jj1pmiss. The dark (blue) filled circles are data points including both the Standard Model and "signal" events, the light (green) squares are "signal" events only, and the solid (red) lines are the expected SM backgrounds as calculated by Pythia. (CDF detector simulated with TurboSim ver.2. Partition Rule ver. 2.0)

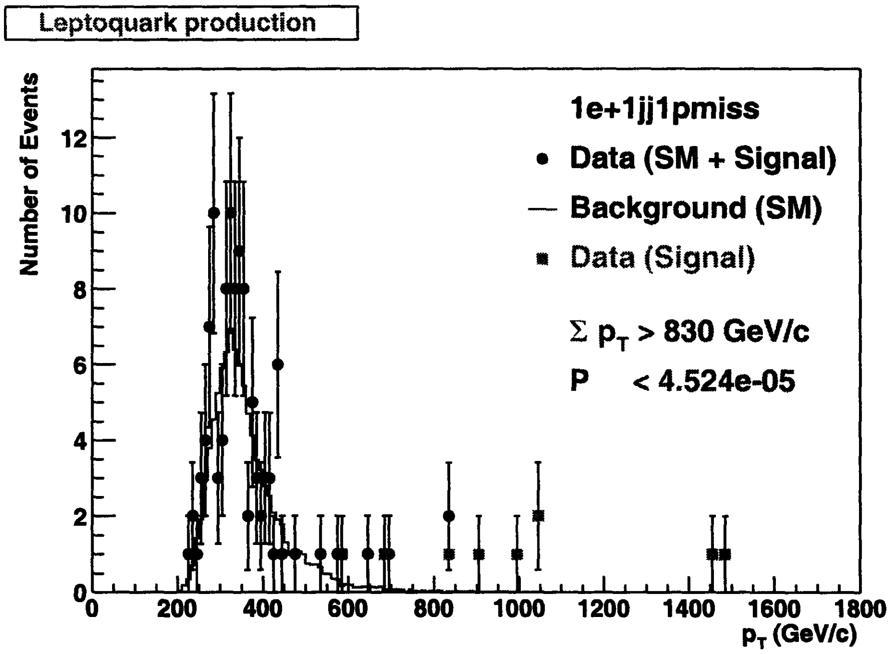


Figure 5-10: Histogram in $\sum p_T$ for data ( Leptoquark "signal" and peudo data pulled from Standard Model Monte Carlo simulation) and Standard Model background for the Sleuth Final State 1e+2jj1ph. The dark (blue) filled circles are data points including both the Standard Model and "signal" events, the light (green) squares are "signal" events only, and the solid (red) lines are the expected SM backgrounds as calculated by Pythia. (CDF detector simulated with TurboSim ver.2. Partition Rule ver. 2.0)

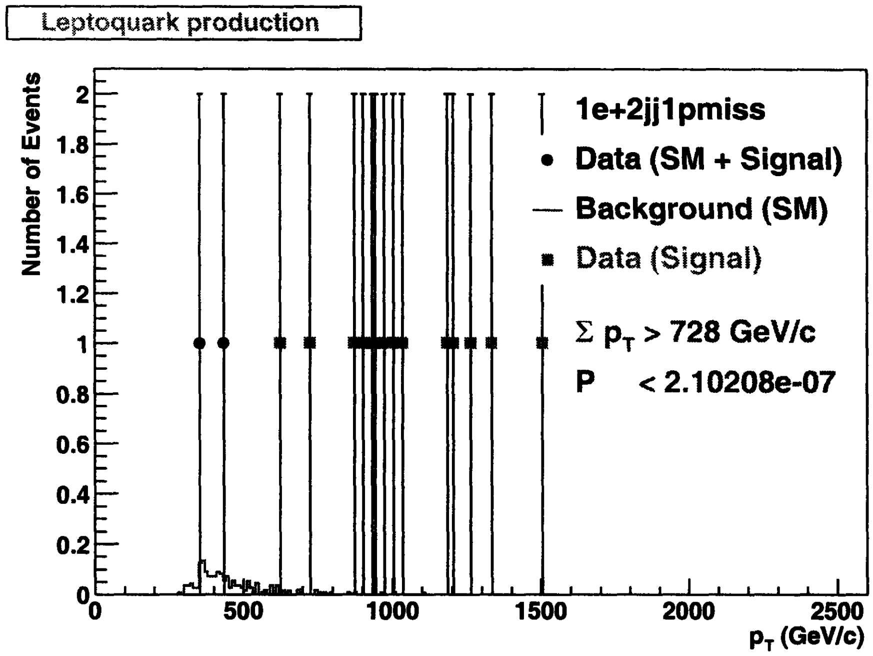


Figure 5-11: Histogram in $\sum p_T$ for data ( Leptoquark "signal" and peudo data pulled from Standard Model Monte Carlo simulation) and Standard Model background for the Sleuth Final State 1e+2jj1ph. The dark (blue) filled circles are data points including both the Standard Model and "signal" events, the light (green) squares are "signal" events only, and the solid (red) lines are the expected SM backgrounds as calculated by Pythia. (CDF detector simulated with TurboSim ver.2. Partition Rule ver. 2.0)

## 5.4 Top quark

We assumed the top quark [12, 13] had not been found yet, and generated 100pb$^{-1}$ of signal events of the top quark with a mass of $m_t = 175\text{GeV}/c^2$, and a total cross-section of 5.53pb. We also generated 100pb$^{-1}$ "pseudo data" events pulled from SM background calculated without the top events, and mixed them together. We then generated an independent set of SM background events without the top. We then gave these to Sleuth as inputs.

The top quark example played the role of a final test of whether Sleuth performs as expected, and finds interesting the Final States which are the signals from the top quark. The expected signal Final States were 1bb1e+1e-1pmiss or 1bb1jj1w. Since the top decays into a b quark and a W boson, and the W decays into lepton and neutrino, or quark pair, the top pair produced would decay into two b's (1bb) and two W's (2w), and the two W's would either show up in the Sleuth Final States as a W and a jet pair (1jj1w); two leptons and a $\not{E}_T$ (1e+1e-1pmiss); or a lepton, $\not{E}_T$ and a jet pair (1e+1jj1pmiss).

Table 5.15 (Partition Rule ver. 1.0) and Table 5.17 (Partition Rule ver. 2.0) show how the top quark signal events were partitioned into Sleuth Final States using the two versions of Partition Rule and TurboSim ver.1. (Turbosim ver. 2 had a problem with handling of b's and could not be used for this example.) Comparison between the two tables shows that the number of Final States with a b pair (1bb) has increased, as intended for Partition Rule ver. 2.0. This is a big improvement over the previous version of Partition Rule. The improvement is more pronounced if the Sleuth results are compared between Table 5.16 and Table 5.18. While none of the Final States in Table 5.16 has a b pair (1bb) in them, three of the Final States in Table 5.18 has b pairs (1bb) in them, and one of those three Final States was found correctly by Sleuth to be the most interesting Final State. The number of Final States with W's has increased also. In Table 5.17, four Final States with the

| Sleuth Final State | Number of Events |
|---|---|
| 1bb1e+1jj | 1 |
| 1bb1e+1jj1pmiss | 1 |
| 1bb1e+1pmiss | 2 |
| 1bb1w | 1 |
| 1e+1e-1jj | 1 |
| 1e+1e-1jj1pmiss | 1 |
| 1e+1jj | 6 |
| 1e+1jj1mu+1pmiss | 1 |
| 1e+1jj1mu- | 1 |
| 1e+1jj1mu-1pmiss | 1 |
| 1e+1jj1ph | 1 |
| 1e+1jj1pmiss | 19 |
| 1e+1ph1pmiss | 1 |
| 1e+2jj | 8 |
| 1e+2jj1mu-1pmiss | 1 |
| 1e+2jj1pmiss | 10 |
| 1e+3jj1pmiss | 1 |
| 1jj1ph1w | 1 |
| 1jj1ph1w | 2 |
| 1jj1w | 4 |
| 2jj1w | 9 |
| 3jj1w | 1 |
| Total | 73 |

Table 5.15: Top quark events which make it into Sleuth Final States. Detector simulation done with TurboSim v.1 (dated Feb 16). Partition Rule ver. 1.0 was applied for Sleuth.

most number of top signal events have W's in them. Also, the most interesting Final State includes a W (1w), correctly indicating that the top signal includes a W in the decay chain. This shows that the final version of Partition Rule (ver. 2.0) is correctly partitioning events, treats jet pairs (1jj) and b pairs (1bb) correctly, and does a good job of identifying a lepton and a neutrino (pmiss) into a W in the decay chain.

The Final States in Table 5.17 (Partition Rule ver. 2.0) with large number of events were, in decreasing order: 1bb1w, 1bb1jj1w, 2jj1w, 1jj1w, 1e+2jj1pmiss, 1bb1e+1pmiss, 1e+1jj1pmiss. The remaining Final States had less than three top signal events in them. Let's look at each of the Final States. 1bb1w is mostly due to loss of one of the jets in the signal. Sleuth Final States partitions an event with a

| Sleuth Final State | $\mathcal{P}$ | $p_T$ threshold | $\tilde{\mathcal{P}}$ |
|---|---|---|---|
| 1e+2jj | $< 4.01496e-05$ | 203.276 | |
| 2jj1w | $< 4.41959e-06$ | 211.162 | |
| 1e+ | 0.83682 | 296.852 | |
| 1e+1e- | 0.688468 | 326.494 | |
| 1e+1e-1jj1pmiss | 0.0847099 | 376.916 | |
| 1e+1e-1pmiss | 0.618238 | 258.875 | |
| 1e+1jj | 0.361337 | 271.245 | |
| 1e+1jj1ph | 0.0533476 | 201.91 | |
| 1e+1jj1ph1pmiss | 0.297177 | 407.251 | |
| 1e+1jj1pmiss | 0.209314 | 248.132 | |
| 1e+1jj1pmiss1tau- | 0.110957 | 266.257 | |
| 1e+1ph | 0.550964 | 210.781 | |
| 1e+1ph1pmiss | 0.157356 | 222.852 | |
| 1e+1pmiss | 0.446429 | 485.882 | |
| 1e+2jj1pmiss | 0.0100397 | 208.217 | |
| 1jj | 0.234467 | 830.275 | |
| 1jj1ph | 0.0942285 | 535.227 | |
| 1jj1ph1pmiss | 0.883002 | 1409.58 | |
| 1jj1pmiss | 0.0653915 | 991.658 | |
| 1jj1w | 0.616333 | 205.239 | |
| 1jj1z | 0.767754 | 204.518 | |
| 1ph | 0.824742 | 676.84 | |
| 1ph1pmiss | 0.439078 | 574.443 | |
| 1w | 0.185357 | 221.281 | |
| 1z | 0.204813 | 244.134 | |
| 2jj1ph1pmiss | 0.544218 | 884.018 | |
| 2jj1w | $< 4.41959e-06$ | 211.162 | $< 0.00031$ |

Table 5.16: Sleuth results for Final States with the top signals and with pseudo data pulled from Standard Model backgrounds. The most interesting Final State was found to be 2jj1w, one of the final states top is expected to decay into, with a $\tilde{\mathcal{P}} < 0.00031$, exceeding the threshold for discovery. Detector simulation done with TurboSim v.1 (dated Feb 16). Partition Rule ver. 1.0 was applied for Sleuth.

| Sleuth Final State | Number of Events |
|---|---|
| 1bb1e+1jj | 1 |
| 1bb1e+1mu+1pmiss | 1 |
| 1bb1e+1mu- | 1 |
| 1bb1e+1mu-1pmiss | 1 |
| 1bb1e+1ph | 1 |
| 1bb1e+1pmiss | 5 |
| 1bb1jj1w | 12 |
| 1bb1ph1w | 1 |
| 1bb1w | 16 |
| 1e+1e-1jj1pmiss | 2 |
| 1e+1jj1pmiss | 4 |
| 1e+2jj | 1 |
| 1e+2jj1mu-1pmiss | 1 |
| 1e+2jj1pmiss | 6 |
| 1e+3jj1pmiss | 1 |
| 1jj1ph1w | 1 |
| 1jj1w | 7 |
| 1ph1w | 1 |
| 2jj1w | 9 |
| 3jj1w | 1 |
| Total | 73 |

Table 5.17: Top quark events which make it into Sleuth Final States. Detector simulation done with TurboSim v.1 (dated Feb 16). Partition Rule ver. 2.0 was applied for Sleuth.

| Sleuth Final State | $\mathcal{P}$ | $p_T$ threshold | $\tilde{\mathcal{P}}$ |
|---|---|---|---|
| 1bb1e+1pmiss | 0.0815661 | 469.661 | |
| 1bb1jj1w | $< 1.90892e-07$ | 256.598 | |
| 1bb1ph1pmiss | 0.266134 | 613.047 | |
| 1bb1w | 0.00313652 | 234.037 | |
| 1e+ | 0.879121 | 228.213 | |
| 1e+1e-1jj1pmiss | 0.0698934 | 376.916 | |
| 1e+1e-1pmiss | 0.854701 | 258.875 | |
| 1e+1jj | 0.472813 | 274.585 | |
| 1e+1jj1ph1pmiss | 0.135823 | 407.251 | |
| 1e+1jj1pmiss | 0.265957 | 263.436 | |
| 1e+1ph1pmiss | 0.164677 | 238.447 | |
| 1e+1pmiss | 0.238095 | 484.609 | |
| 1e+2jj1pmiss | 0.0349498 | 329.029 | |
| 1jj | 0.310078 | 830.275 | |
| 1jj1ph1pmiss | 0.821355 | 1409.58 | |
| 1jj1ph1w | 0.406504 | 201.91 | |
| 1jj1pmiss | 0.0512229 | 991.658 | |
| 1jj1tau-1w | 0.00636851 | 265.076 | |
| 1jj1w | 0.943396 | 201.109 | |
| 1ph | 0.787402 | 676.84 | |
| 1ph1pmiss | 0.821355 | 509.106 | |
| 1ph1w | 0.358423 | 220.734 | |
| 1w | 0.514139 | 295.825 | |
| 1z | 0.0831774 | 225.536 | |
| 2jj1ph1pmiss | 0.643087 | 884.018 | |
| 2jj1w | 0.00112909 | 203.276 | |
| 1bb1jj1w | $< 1.90892e-07$ | 256.598 | $< 2.1e-05$ |

Table 5.18: Sleuth results for Final States with the top signals and with pseudo data pulled from Standard Model backgrounds. The most interesting Final State was found to be 1bbjj1w, one of the final states top is expected to decay into, with a $\tilde{\mathcal{P}} < 0.000021$, exceeding the threshold for discovery. Detector simulation done with TurboSim v.1 (dated Feb 16). Partition Rule ver. 2.0 was applied for Sleuth.

single jet into a Final State with no jet pair. This is a drawback of grouping jets into pairs. But since this acts as a "cleaning up" of the softly radiated extra jets for other signals, it is a trade-off, and cannot be totally eliminated. The Final State 1bb1jj1w is one of the "correct" signals of the top quark signal, and 12 events out of 73 events that pass the basic cuts are partitioned into this Final State. The events in the 2jj1w Final State did not have any of the jets identified as coming from a b quark. Since the detector only has about a 30% efficiency in the identification of b's, and Partition Rule ver. 2.0, requires only one of the b's to make it into the 1bb1jj1w Final State, the ratio of events seems right. In the absense of dependable b identification, this Final State would be one of the top signal Final State. The events in the 1jj1w Final State is similar to 2jj1w, but mostly missing one jet. The events in the 1e+2jj1pmiss, 1bb1e+1pmiss 1e+1jj1pmiss Final States had an electron and a $\not\!\!E_T$ (pmiss) which did not have a transverse mass $m_T$ in the W range, and did not make it into either the 2jj1w, the 1bb1w, 1jj1w Final States, respectively.

Fine tuning the jet cuts to rearrange the number of jets is an attractive option, but this will likely result in Sleuth being more sensitive to one kind of signal (multiple jet pairs vs. single jet pairs), and should not be tuned to a particular signal. Rather, the general goal should be that Sleuth Final States includes as many objects as is reasonably identifiable, i.e. cuts should be as low as possible, and the overall distribution of events in the Final States should indicate where the signal is coming from. Note that the lowering of jet cuts to move events from 1bb1w to 1bb1jj1w for the top quark example would also move events in a different example, such as the Leptoquark, with multiple jet pairs in the same direction, even if the lesser jet pair is the correct signal for that example. The "clean up" effect of jet pairing for the later example would be reduced, and worsen the Sleuth result for that example.

Figure 5-12 and Figure 5-13 shows $\sum p_T$ distribution of two Final States from Table 5.18. These two Final States had the most number of top signal events in Table 5.17. Sleuth found the 1bb1jj1w Final State the most interesting with a $\mathcal{P} < 1.909e-07$ with a $p_T$ threshold of $\sum p_T > 257$. The Final State 1bb1w had the largest number of top signal events among the 20 Final States which had at least

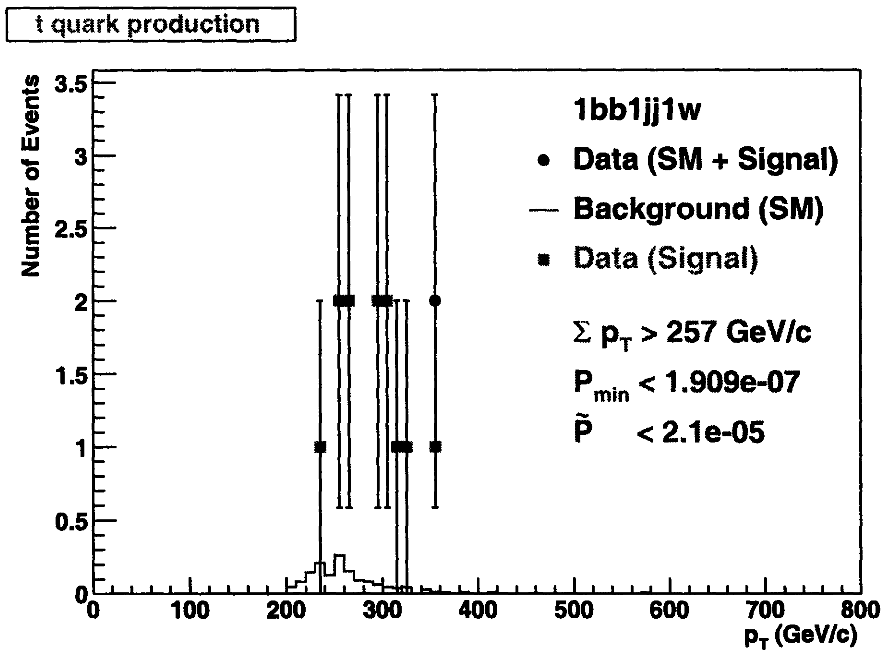


Figure 5-12: Histogram in $\sum p_T$ for data ( top quark "signal" and peudo data pulled from Standard Model Monte Carlo simulation) and Standard Model background for the Sleuth Final State 1bb1jj1w. The dark (blue) filled circles are data points including both the Standard Model and "signal" events, the light (green) squares are "signal" events only, and the solid (red) lines are the expected SM backgrounds as calculated by Pythia. (CDF detector simulated with TurboSim ver. 1. Partition Rule ver. 2.0)

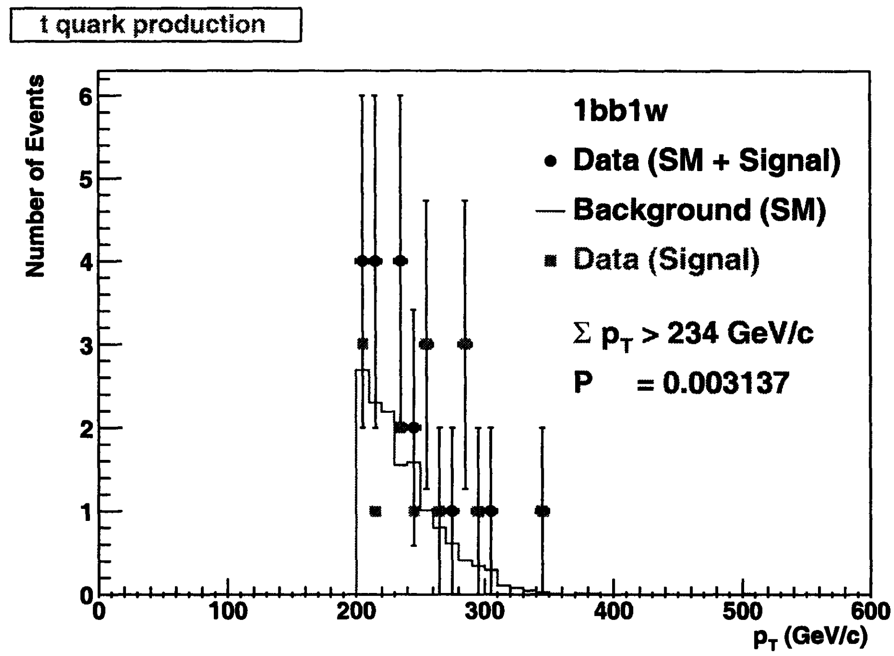


Figure 5-13: Histogram in $\sum p_T$ for data ( top quark "signal" and peudo data pulled from Standard Model Monte Carlo simulation) and Standard Model background for the Sleuth Final State 1bb1w. The dark (blue) filled circles are data points including both the Standard Model and "signal" events, the light (green) squares are "signal" events only, and the solid (red) lines are the expected SM backgrounds as calculated by Pythia. (CDF detector simulated with TurboSim ver. 1. Partition Rule ver. 2.0)

one top signal event. The $p_T$ threshold for the two Final State generally agrees, and gives us an idea of what the mass of the signal would be. From the $p_T$ threshold, we get the approximate lower bound on the transverse mass, i.e. $m_T > 250GeV/c^2$. By looking at the data points ( labeled as Data (SM+Signal) in the two figures), we get an approximate upper bound on the transverse mass, i.e. $m_T < 350GeV/c^2$. We cannot get an estimate on the invariant mass because a $\not{E}_T$ (the W is identified from a lepton and a neutrino, or a $\not{E}_T$ in the detector) is included in the objects which make up the Final State. This is in agreement with the invariant mass of the top quark pair produced in this example, $m_{tt} \approx 350GeV/c^2$. Sleuth correctly finds the 1bb1jj1w Final State most interesting, with $\tilde{\mathcal{P}} < 2.1e - 05$ which exceeds the threshold of discovery. Thus, Sleuth has "found" the top quark, which decays into 1bb1jj1w.

## 5.5 Vector Boson

We generated 100pb$^{-1}$ of signal events of the W and Z bosons pair production, the cross-sections were the default values in Pythia: $f + \bar{f} \rightarrow Z^0 + Z^0$ : 1.1 pb, $f + \bar{f}' \rightarrow Z^0 + W^{\pm}$ : 2.4 pb, $f + \bar{f} \rightarrow W^+ + W^-$ : 8.2 pb, and a total cross-section of 11.7 pb. We also generated 100 pb$^{-1}$ "pseudo data" events pulled from SM background calculated without the top events, and mixed them together. We then generated an independent set of SM background events without those processes, and gave it to Sleuth as inputs.

This example is effectively testing if Sleuth will find the excess in data due to these pair production of W and Z bosons, treating them as if these processes were new physics processes. The purpose of this example is mainly to test how the TurboSim and Sleuth treats processes which include weak bosons in the decay chain, and whether these bosons are correctly identified as such when they are produced in pairs, which will make them more difficult to identify. Since the W and Z can decay into both lepton pairs or quark pairs, the most likely signal Final States should be 1e+1jj1pmiss or 1jj1w for the W, and 1e+1e-1jj or 1jj1z for the Z. And the events should be dominated by the W boson pair production due to the difference in the cross-sections for the processes.

The signal we are considering is the di-boson production of the W and Z, and not the presence of the W or Z bosons. Because we have not eliminated the production of W and Z bosons entirely from the Standard Model background — single W and Z boson production coming from processes such as $f + \bar{f} \rightarrow g + Z^0, f + \bar{f}' \rightarrow g + W^{\pm}$, and decays such as $t \rightarrow b + W^+$ — we do expect the signal to be more difficult to find than the top example.

From Table 5.19, we see that the majority of the signal events were partitioned into 1jj1w and 1e+1jj1pmiss as expected. These two Final States could have been put into one Final State 1jj1w if we had taken the upper bound of the W transverse mass $m_T$ higher than that defined in the Partition Rule ver. 2.0. But that would have

| Sleuth Final State | Number of Events |
|---|---|
| 1e+1e-1pmiss | 1 |
| 1e+1jj | 1 |
| 1e+1jj1mu-1pmiss | 1 |
| 1e+1jj1pmiss | 4 |
| 1e+1mu-1pmiss | 1 |
| 1e+2jj | 1 |
| 1jj1ph1w | 1 |
| 1jj1w | 16 |
| 1jj1z | 1 |
| 1w | 1 |
| 2jj1w | 2 |
| Total | 30 |

Table 5.19: W and Z pair production events which make it into Sleuth Final States. Detector simulation done with TurboSim v.2 (dated Apr 25). Partition Rule ver. 2.0 was applied for Sleuth.

risked contaminating the w's in a Sleuth Final State with 1e+1pmiss which do not come from the decay of a W boson. We have lost 20% of the signal in 1jj1w due to the partitioning of 1e+1pmiss into a 1w using the $m_T$, as defined in Partition Rule ver. 2.0 – if we had not included the "W criteria", the two Final States would have been partitioned together, and we would have gained about 25% increase in the number of events relative to the 1jj1w Final State. That is a trade off for wanting to distinguish between the Final State objects 1e+1pmiss which do come from the decay of a W boson from those that do not come from it. This is a matter of judgement on whether we want the ability to identify 1e+1pmiss from a W or not. If the interpretation of the underlying physics process is improved by this ability, then the "W criteria" should be kept at the cost of about $20-25\%$ loss in the number of events that goes into a Final State which includes 1e+1pmiss. If the improvement in interpretation is not such an important issue, or we want to maximize the number of events in a Final State with 1e+1pmiss to allow more concentration of events into exclusive Final States, then the "W criteria" should be removed from later versions of Partition Rule. Another side to this problem is, whether we expect any new particle with a mass around or below the W mass to decay into a lepton and a neutrino (1e+1pmiss). If the answer is yes, we should discard the "W criteria"; if the answer is no, we should keep it. The likely

| Sleuth Final State | $\mathcal{P}$ | $p_T$ threshold | $\tilde{\mathcal{P}}$ |
|---|---|---|---|
| 1e+ | 0.662252 | 230.602 | |
| 1e+1e-1jj1pmiss | 0.294768 | 513.325 | |
| 1e+1e-1pmiss | 0.34662 | 348.096 | |
| 1e+1jj | 0.0786318 | 361.402 | |
| 1e+1jj1ph | 0.118064 | 200.346 | |
| 1e+1jj1pmiss | 0.285307 | 507.318 | |
| 1e+1ph | 0.883002 | 279.769 | |
| 1e+1pmiss | 0.310078 | 443.196 | |
| 1jj | 0.551724 | 920.362 | |
| 1jj1ph1w | 0.4561 | 234.377 | |
| 1jj1pmiss | 0.686106 | 1091.6 | |
| 1jj1w | 0.121951 | 204.923 | |
| 1jj1z | 0.0237939 | 251.294 | |
| 1ph | 0.397614 | 647.651 | |
| 1ph1w | 0.3663 | 203.175 | |
| 1tau+1w | 0.441989 | 324.022 | |
| 1w | 0.214018 | 299.446 | |
| 1z | 0.641026 | 391.809 | |
| 2jj1w | 0.0604412 | 257.67 | |
| 1jj1z | 0.0237939 | 251.294 | 0.57 |

Table 5.20: Sleuth results for Final States with the W or Z pair production signals and with pseudo data pulled from Standard Model backgrounds. The most interesting Final State was found to be 1jj1z, one of the final states a W and Z couple is expected to decay into, with a $\tilde{\mathcal{P}} < 0.57$. Detector simulation done with TurboSim v.2 (dated Apr 25). Partition Rule ver. 2.0 was applied for Sleuth.

answer to this is no, since we haven't found any new particle yet with a mass around the W mass, which decays into a lepton and a neutrino.

From Figure 5-14 and Figure 5-15, we see that by using the "W criteria", we have effectively divided the W pair production signal at about $\sum p_T = 300 GeV/c$. This has improved the ability of Sleuth to find excess in the data compared to the background for this example, as can be seen by comparing the $\mathcal{P}$ value of the two Final States. We see that the probability of the background to fluctuate up to the number of events in the data is lower in the 1jj1w Final State by about a factor of 2 compared to the 1e+1jj1pmiss Final State. But this does not necessarily guarantee that such an effect will hold for all signals. The opposite would generally be true for signal which is broadly distributed over a wide range of $\sum p_T$ across the W transverse mass cut $m_T < 160GeV/c^2$. This rather wide range, reaching well above the W mass was determined by looking at the resolution of the pmiss coming from a definite signal such as the $W'$ example, which gave a resolution of about $p_T \sim 40GeV/c$ for the detector simulation TurboSim. The Partition Rule ver. 2.0 allowed a lepton and a $\not{E}_T$ (1e+1pmiss) with transverse mass $m_T$ approximately $2\sigma$ above the W mass to be considered a W (1w).

As shown in Table 5.20 and Figure 5-16, Sleuth finds the Final State 1jj1z most interesting with a $p_T$ threshold of $\sum p_T > 251GeV/c$ and $\mathcal{P} = 0.02379$, or probability of the Standard Model background fluctuating up to the events the data are about two out of a hundred cases. The "Z criteria" is not as problematic as the "W criteria" because of two reasons. The Z's in the Sleuth Final States (z) are considered only for lepton pairs (1e+1e-) or lepton pairs and a radiated photon (1e+1e-1ph) which are much more clear objects in the detector than $\not{E}_T$ (pmiss), which could be more than one neutrino. Also the Z's are determined by an invariant mass with a relatively narrow range $76 < m_Z < 106$, with both an upper and lower bound, where as the W's had only an upper bound. Due to the lack of statistics in the Z signals, additional check was done with the Standard Model background events to check that the Z's in the Final States are coming mostly from processes which have Z's at the parton level.

The $\tilde{\mathcal{P}} = 0.57$ given in Table 5.20 for the 1jj1z means over half of hypothetical similar experiments would return results as interesting as this example or more interesting. This definitely does not qualify as a discovery, and would be considered to mean the data and Standard Model background is in agreement. Thus a null result for a search for excess in data compared to the background was returned for this example.

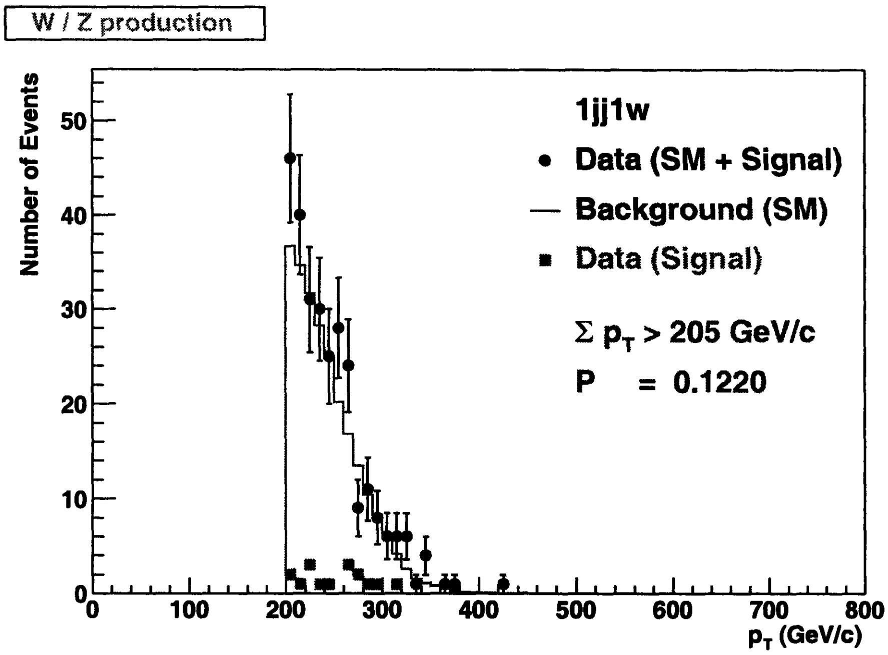


Figure 5-14: Histogram in $\sum p_T$ for data ( W or Z pair production "signal" and peudo data pulled from Standard Model Monte Carlo simulation) and Standard Model background for the Sleuth Final State 1jj1w. The dark (blue) filled circles are data points including both the Standard Model and "signal" events, the light (green) squares are "signal" events only, and the solid (red) lines are the expected SM backgrounds as calculated by Pythia. (CDF detector simulated with TurboSim ver. 2. Partition Rule ver. 2.0)

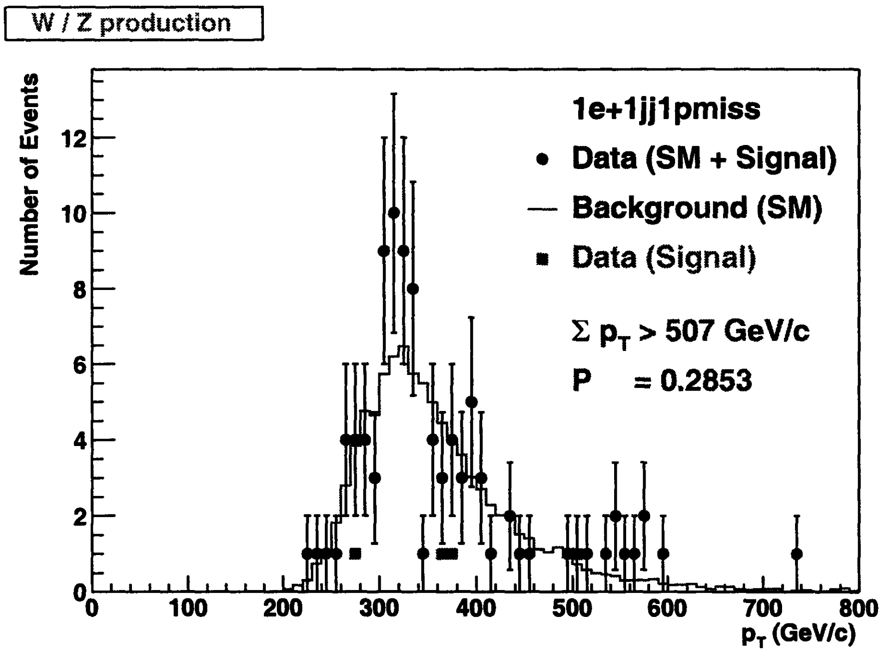


Figure 5-15: Histogram in $\sum p_T$ for data ( W or Z pair production "signal" and peudo data pulled from Standard Model Monte Carlo simulation) and Standard Model background for the Sleuth Final State 1e+1jj1pmiss. The dark (blue) filled circles are data points including both the Standard Model and "signal" events, the light (green) squares are "signal" events only, and the solid (red) lines are the expected SM backgrounds as calculated by Pythia. (CDF detector simulated with TurboSim ver. 2. Partition Rule ver. 2.0)

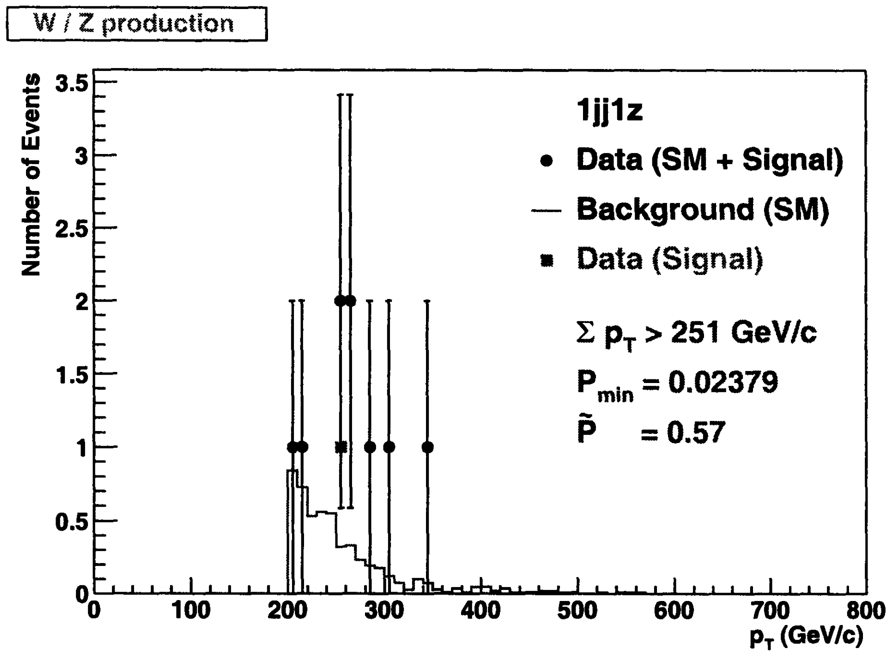


Figure 5-16: Histogram in $\sum p_T$ for data ( W or Z pair production "signal" and peudo data pulled from Standard Model Monte Carlo simulation) and Standard Model background for the Sleuth Final State 1jj1z. The dark (blue) filled circles are data points including both the Standard Model and "signal" events, the light (green) squares are "signal" events only, and the solid (red) lines are the expected SM backgrounds as calculated by Pythia. (CDF detector simulated with TurboSim ver. 2. Partition Rule ver. 2.0)

# Chapter 6

# Answers to Open Questions

The questions posed at the start of this study are whether our treatment of the following items are a good choice. Does each treatment increase the chances of Sleuth finding a beyond-Standard-Model signal? Are the following choices in the Partition Rule good choices, and have they been implemented in Sleuth correctly?

1) Treat jets as pairs: e.g. 3 jets $\rightarrow$ 1jj ?

Answer is yes.

From the examples with multiple jets, such as the leptoquark and top examples, there are $10 - 30\%$ increase in number of events in the Final State with the smaller number of jets, if we treat jets as pairs. In other words, 2jj made up of 4j and 5j sees increase in the number of events, whereas if they were partitioned separately, many of the 5j Final State would have had too few events to be considered separately as a Sleuth Final State (remember Sleuth requires a minimum number of 4 events), and would have been just thrown away. A good example of this is the Leptoquark example. The events in the 1e+2jj1pmiss Final State (TurboSim ver. 2, Partition Rule ver. 2.0), are made up of 14 events of 1e+4j1pmiss (4 jets) and 4 events of 1e+5j1pmiss (5 jets) combined together into 1e+2jj1pmiss (4 jets).

An additional gain by the treatment of jets as pairs is the shorter processing time. It reduces the number of Final States with jets by half – and many Final States have jets in them – which decreases the Sleuth calculation time significantly.

Current implementation in Partition Rule ver. 2.0 correctly finds the Final State 1bb1jj1w as the most interesting Final State for the top quark, with the number of jet pairs consistent with the known decay products.

Thus, we can conclude the choice of jet pairing is a good idea, and the current implementation works correctly.

2) Treat b's as pairs, separate from jets : e.g. 1b $\rightarrow$ 1bb, if well-tagged ?

Answer is yes.

From the top example, we see that the treatment of jets identified as coming from b quarks (b's), as implemented in Partition Rule ver. 2.0, correctly gives the number of b's in the Final States which have b's in them. It also correctly finds the Final State 1bb1jj1w as the most interesting Final State with the number of b pairs consistent with the known decay products.

Thus, we can conclude the choice of b jet pairing is a good idea, and the current implementation works correctly.

3) Combine $l^+l^-$ or $l^+l^-\gamma$ into a Z, if $m_{l^+l^-}$ or $m_{l^+l^-\gamma}$ is between 76 and 106 GeV/c$^2$ ?

Answer is yes.

From the Vector Boson example and looking at the Standard Model background events, we see that combining $l^+l^-$ or $l^+l^-\gamma$ into a Z works correctly. Also, Sleuth returned 1jj1z as the most interesting Final State for the Vector Boson example, which is consistent with the known decay products of a pair of W and Z bosons. The primary gain with this choice is the additional information on whether the lepton pairs, or lepton pairs and a photon, are coming from the decay of a Z boson, which would be valuable in the physics interpretation of the possible processes responsible for the excess found in the data.

Thus, we can conclude that the combining of lepton pairs, or lepton pairs and a photon, into a Z is a good idea, and the current implementation works correctly.

4) Combine $l^+$ missing-$E_T$ or $l^-$ missing-$E_T$ into a $W^{\pm}$, if the transverse mass less than 160 GeV/c$^2$ ?

Answer is yes.

From the Vector Boson example, current implementation of this choice in Partition Rule ver. 2.0 correctly partitions a majority of the signal events into the 1jj1w Final State, which is the correct dominant Final State, consistent with the decay products. A more detailed discussion of the issue was given in a previous section when dealing with the Vector Boson example.

From the top example, current implementation in Partition Rule ver. 2.0 correctly finds the Final State 1bb1jj1w as the most interesting Final State with the treatment of W (1w) as a Sleuth object consistent with the known decay products. The 1bb1jj1w is the expected top signal Final State, and SLEUTH$_{\mathrm{II}}$ found it to be interesting at the level of "a discovery". This is a marked improvement over SLEUTH$_{\mathrm{I}}$, which could not find the top quark signature in a comparable amount of data.

Thus, we can conclude the choice of combining a lepton (1e+) and a $\not\!\!E_T$ (pmiss) into a W (1w) is a good idea, and the current implementation works correctly.

The change in the range of the transverse mass, $m_T < 160\text{GeV/c}^2$, is due to the poor resolution of the $\not\!\!E_T$ of $p_T \sim 40\text{GeV/c}$ when the detector is simulated by TurboSim.

5) Global charge conjugation: e.g. $e+e+\gamma = e-e-\gamma \neq e+e-\gamma$ ?

Answer is yes.

Many examples, such as the top, Vector Boson, W' support this. W' is a good example. 31 events were partitioned into 1e+1pmiss Final State out of 32 events generated. Among the events in the 1e+1pmiss Final State, 13 came from events with e+(mu+), 18 came from events with e-(mu-). The choice of global charge conjugation resulted in concentrating the number of signal events into one Final State, and increasing the number of events in such a Final State by a factor of about 2 in most cases.

We can conclude that the global charge conjugation is a good idea, and the current implementation works correctly.

6) Group 1st and 2nd generation together, while treating 3rd generation separately from them: e.g. treat 1e+ ( includes e+,$\mu$+) and 1$\tau$+; 1jj and 1bb as Sleuth final state objects?

Answer is yes.

Since the jets and b's are discussed separately, let us focus on the leptons. Many examples, such as the Z', W' and top quark, support this choice. The W' example is again a good example. 31 events were partitioned into 1e+1pmiss Final State out of 32 events generated. Among the events in the 1e+1pmiss Final State, 15 are events with a e, 16 are events with $\mu$. The choice of grouping 1st and 2nd generation together resulted in concentrating the number of signal events into one Final State, and increasing the number of events in such a Final State by a factor of about 2 in most cases.

We can conclude that the grouping of 1st and 2nd generation together is a good idea, and the current implementation works correctly.

# Chapter 7

# Results and Conclusion

The results of this study can be summarized as follows:

- The SLEUTH$_{II}$ analysis package has been developed and debugged to work with detector simulations PGS (Pretty Good Simulation) and TurboSim. No known bugs are remaining.

- SLEUTH$_{II}$ has been tested with Monte Carlo examples generated by Pythia and processed with two types of detector simulations: PGS and TurboSim. The algorithm as currently implemented performs as intended.

- TurboSim has been tested for use with a physics analysis for the first time, following the demonstration of its proof of concept. Some problems have been found with the two versions of TurboSim available at the time of the study, especially with the fine tuning of the $p_T$ of the jets in multiple jets (j's); handling of jets identified as coming from b's (b's); compensating for unclustered energies at the detector level (currently TurboSim does not include them) and handling of $\not{E}_T$ (pmiss); and extending the $\eta$ range of the electron identification to reflect the latest developments of the CDF detector. Comparison of the two versions show that the TurboSim ver. 2 performs better than previous version in reproducing the jet $p_T$, but still seems to have problems reproducing the number and $p_T$ of the jets when there are large number of jets in an event. Handling of pmiss is a common problem with both versions.

- Partition Rule (ver. 2.0) has been finalized for $\text{SLEUTH}_{\text{II}}$ at CDF Run II, and confirmed with examples using TurboSim. Possible improvements are a more detailed study of the cuts applied to the objects, especially of the jets using CDFSim, which may give more reliable description of the actual CDF detector. But since the odd number jets, especially the softly radiated 3rd and 5th jets, do not contribute to the number of jets counted in the Sleuth Final States, the details of their cuts should not have significant impact on the number of Final States defined by Partition Rule and considered by Sleuth in its calculations. And in principle, the odd number jets should be included in the calculation of the $\sum p_T$ of an event as much as possible, since higher cut on the "presumed" softer jets are somewhat arbitrary and goes against the basic idea of SLEUTH of including all identifiable objects in an event, and relying on the correct Partition Rule and the Sleuth algorithm to minimize the effects of small variations in the actual cuts.

- The "6 Open Questions" posed at the beginning of the study have been answered. Examples supporting the assumptions of the Partition Rule, as outlined in Section 2.3.1 Definitions of final states, have been provided for the first time.

- Final test of $\text{SLEUTH}_{\text{II}}$ using the top example returned a positive result: $\text{SLEUTH}_{\text{II}}$ "found" the top quark in the 1bb1jj1w Final State with a $p_T$ threshold of $\sum p_T = 256.6$ GeV/c, $\mathcal{P} < 1.91 \times 10^{-7}$ and the overall $\tilde{\mathcal{P}} < 2.1 \times 10^{-5}$, which exceeds the threshold of "discovery". Thus, $\text{SLEUTH}_{\text{II}}$ would have "found" the top, had it not been found already.

- $\text{SLEUTH}_{\text{II}}$ demonstrated improvement over $\text{SLEUTH}_{\text{I}}$ by correctly finding the 1bb1jj1w Final State as the most interesting (i.e. having the minimum $\mathcal{P}$) Final State with the final Sleuth result $\tilde{\mathcal{P}} < 2.1 \times 10^{-5}$, indicating a "discovery" of the top. $\text{SLEUTH}_{\text{I}}$ could not claim a "discovery" of the top quark when run on DØ data with a comparable integrated luminosity.

- It has been shown, for the $Z'$, $W'$, leptoquark and top examples considered in this thesis, that the choice of $\sum p_T$ as the single variable used to characterize an event was a valid choice. We found it to be a good variable which would indicate existence of new physics not included in the background processes.

The details of the $\text{S}\textsc{leuth}_{\text{II}}$ algorithm have been tested, and it has been shown to perform well. $\text{S}\textsc{leuth}_{\text{II}}$ performs better than $\text{S}\textsc{leuth}_{\text{I}}$ in the case of top example, when ignorance of its existence is feigned, by finding the top quark with $\tilde{\mathcal{P}} < 2.1 \times 10^{-5}$. The open questions posed at the beginning of the study regarding the assumptions made in the development of $\text{S}\textsc{leuth}_{\text{II}}$ have been answered, with supporting examples. Some intuition on SLEUTH performance has been gained by looking at 3 examples of beyond-the-standard-model processes, and 2 examples of standard model processes. $\text{S}\textsc{leuth}_{\text{II}}$ is ready for additional testing with the standard full detector simulation CDFSim, and applying to real CDF Run II data.

TurboSim has been tested for its usability for a physics analysis at a hadron collider such as the Tevatron. Initial results show promise of its use, but also revealed some problems as it is currently implemented: the faithful reproduction of the $p_T$ of multiple jets, handling of b-tagged jets, and poor resolution of the missing-$E_T$. The first two problems may be solved by training TurboSim on a better Monte Carlo sample of the full detector simulation; the last problem will require further development.

# Bibliography


[1] See for example H. Haber and G. Kane, Phys. Rept. **117**, 75 (1985); H. Nilles, Phys. Rept. **110**, 1 (1984); P. Nath *et al.*, *Applied $N = 1$ Supergravity*, (World Scientific 1984); X. Tata, in *The Standard Model and Beyond*, edited by J. Kim (World Scientific, 1991), p. 304.

[2] S. Weinberg, Phys. Rev. D **13**, 974 (1976); *ibid.* **19**, 1277 (1979); L. Susskind, *ibid.* **20**, 2619 (1979); S. Dimopoulos and L. Susskind, Nuc. Phys. **B155**, 237 (1979); E. Eichten and K. Lane, Phys. Lett. **90B**, 125 (1980).

[3] DØ Collaboration, B. Abbott *et al.*, Phys. Rev. D **62**, 92004 (2000).

[4] DØ Collaboration, B. Abbott *et al.*, Phys. Rev. D **64**, 12004 (2001).

[5] DØ Collaboration, B. Abbott *et al.*, Phys. Rev. Lett., **86**, 3712 (2001).

[6] B. Knuteson, Ph.D. thesis, University of California at Berkeley, 2000.

[7] B. Knuteson, hep-ex/0105027, (2001); hep-ex/0311059, (2003); hep-ex/0402029, (2004).

[8] B. Knuteson, hep-ex/0105027, (2001).

[9] B. Knuteson, Conference proceeding that includes a discussion of Sleuth: http://arxiv.org/abs/hep-ex/0402029

[10] B. Knuteson, (2004)
http://mit.fnal.gov/~knuteson/HighPt/web/TurboSim.htm
Conference proceeding that includes a discussion of TurboSim:

http://arxiv.org/abs/hep-ex/0402029;

Additional sample TurboSim plots:

http://mit.fnal.gov/~knuteson/Quaero/quaero/doc/devel/cdf/TurboSim/

[11] J. Conway, (1998)
http://www.physics.rutgers.edu/~jconway/research/software/software.html

[12] DØ Collaboration, S. Abachi *et al.*, Phys. Rev. Lett. **79**, 1203 (1997).

[13] CDF Collaboration, T. Affolder *et al.*, Phys. Rev. D **63**, 32003 (2001).
CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett **82**, 271 (1999).
CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett **80**, 2767 (1998).
CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett **80**, 2779 (1998).
CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett **79**, 3585 (1997).
CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett **74**, 2626 (1995).
CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett **73**, 225 (1994).
CDF Collaboration, F. Abe *et al.*, Phys. Rev. D **50**, 2966 (1994).

[14] T. Sjöstrand, Comput. Phys. Commun. **82**, 74 (1994); we used v6.218, (2003).

[15] J. C. Pati and A. Salam, Phys. Rev. D **10**, 275 (1974); R. N. Mohapatra and J. C. Pati, Phys. Rev. D **11**, 2558 (1975); G. Senjaovic and R. N. Mohapatra Phys. Rev. D **12**, 1502 (1975).

[16] CDF Collaboration, D. Acosta *et al.*, Phys. Rev. Lett **90**, 81802 (2003).

[17] CDF Collaboration, T. Affolder *et al.*, Phys. Rev. Lett **87**, 231803 (2001).

[18] CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett **84**, 5716 (2000).

[19] CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett **67**, 2609 (1991).

[20] J. C. Pati and A. Salam, Phys. Rev. D **19**, 275 (1974); H. Georgi and S. Glashow, Phys. Rev. Lett. **32**, 438 (1974); also, see J. L. Hewett and T. G. Rizzo, Phys. Rep. **183**, 193 (1989) and references therein.

[21] H1 Collaboration, S. Aid *et al.*, Phys. Lett. B **369**, 173 (1996); ZEUS Collaboration, M. Derrick *et al.*, Z. Phys. C **73**, 613 (1997).

[22] H1 Collaboration, C. Adloff *et al.*,Z. Phys. C **74**, 191 (1997); ZEUS Collaboration, J. Breitweg *et al.*, Z. Phys. C **74**, 207 (1997).

[23] DØ Collaboration, B. Abbott *et al.*, Phys. Rev. Lett. **79**, 4321 (1997).

[24] CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett. **79**, 4327 (1997).

[25] CDF Collaboration, F. Abe *et al.*, Phys. Rev. D **48**, 3939 (1993).

[26] DØ Collaboration, S. Abachi *et al.*, Phys. Rev. Lett. **72**, 965 (1994).