MSSM 4b Search in CDF Run II

Lauren Tompkins,[1] Advisor Jim Siegrist,[1] and Advisor Young-Kee Kim[2]

[1] *UC Berkeley and Lawrence Berkeley National Laboratory*

[2] *University of Chicago*

(Dated: May 14, 2004)

Abstract

This thesis is a status report on a search for a Minimal Supersymmetric Standard Model (MSSM) Higgs Boson, the h or A, in 130 pb^{-1} of CDF Run II data collected at the TeVatron at Fermi National Accelerator Laboratory. We look for new particles produced in 1.8 TeV proton anti-proton collisions through a 4 b quark decay channel, the signature of which are four energetic b jets. This analysis currently searches for the boson at $\tan\beta = 50$, a particular value of a free parameter of the MSSM theory. We give a cursory overview of the analysis and presents a detailed account of the progress of determination of the background sources to the measurement with a particular emphasis on the Quantum Chromodynamical (QCD) multi-jet background.

INTRODUCTION

Field theories in modern physics are built on the concept of gauge invariance. The Standard Model, which governs the world of fundamental particles, is no exception. To describe the interactions of fundamental particles the Standard Model includes the strong nuclear force and the electroweak force, the unification of the weak nuclear and electromagnetic forces. For these interactions to be gauge invariant, they require that all particles be massless, but as experiment and daily experience show us, fundamental particles are certainly not massless! In order to fix this problem, termed "spontaneous symmetry breaking", we introduce a scalar field, the Higgs field, which allows for massive particles. The search for the Higgs Boson, the particle associated with the Higgs field, is currently driving high energy particle experiment. However, in its current formulation, the Standard Model is theoretically unsatisfying and consequently alternatives are under investigation. One such alternative is the MSSM (Minimal Supersymmetric Standard Model), which proposes, among other changes to the Standard Model, 5 Higgs bosons: 3 neutral (h, A, H^0) and 2 charged ($H^\pm$) particles.

Using the CDF detector at the TeVatron, a 1.8 TeV proton anti-proton collider located in Batavia, IL, I am participating in a search a light neutral MSSM Higgs boson, h or A. The h is created at lowest order by the annihilation of at quark and an anti-quark producing a gluon which produces a b and a anti-b quark. One of the b or anti-b radiates an h boson, which in turns decays to a b and anti-b quark. Experimentally we look for the signatures of the 4 b quarks, termed jets, and try to reconstruct the mass of the h based on the energy of the b quarks, or equivalently, the energy of the jets. I present studies of the mass of jet pairs for our current CDF Run II data sample with a specific emphasis on determining the QCD (Quantum Chromodynamical) background to our measurement.

Section gives a theoretical overview to the Standard Model and MSSM Higgs mechanisms. The TeVatron and CDF detector are described in Section . Section presents the data set and event selection and Section **??** explains the techniques used in this analysis. In Section we present preliminary results of the analysis and lastly, in Section **??** we draw conclusions.

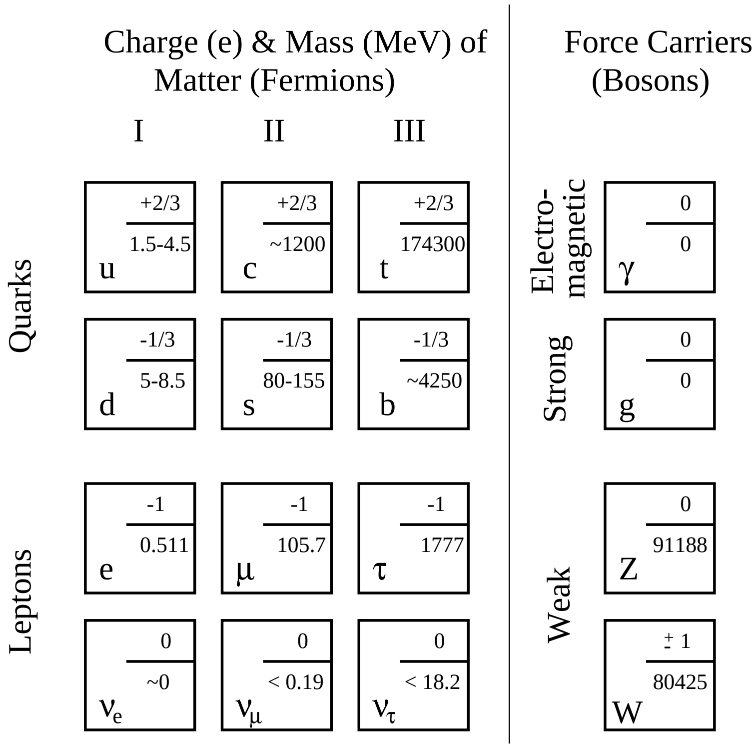

FIG. 1: The Standard Model of Particle Physics. Each particle is accompanied by its $\frac{charge}{mass(inMeV)}$

THEORY

The Standard Model of Particle Physics is a good physical theory which accurately describes the interactions of matter up to the TeV scale, our current energy frontier. It's basic building blocks are the fermions, quarks and leptons, and the bosons, force carriers. Fermions are half integer spin particles which form 'normal' matter. Bosons are of a different nature, acting as exchange particles for three of the four known forces. Figure 1 depicts the Standard Model.

The three forces incorporated in the Standard Model are the electromagnetic force (mediated by photon), weak nuclear ($W^{\pm}, Z$ bosons) and strong nuclear forces (gluons). These interactions are represented by the product of three gauge groups: $\mathrm{SU}(3)_{color} \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$, with quarks making up color triplets of $\mathrm{SU}(3)_{color}$, left handed doublets of $\mathrm{SU}(2)_L$, and right handed singlets, while leptons are only found in the doublets and singlets. Each quark doublet contains an up-type and a down-type element which couple differently to the various fields through different quantum numbers. Figure 2 give examples of the first generation of multiplets.

$$\begin{pmatrix} u \\ d' \end{pmatrix}^{1/6}_{L,b} \quad d^{-1/3}_{R,b}$$

FIG. 2: Examples of a left handed quark doublet and right handed singlet

We can represent the interactions of these fields in a Lagrangian which accounts for each of the forces described above. The simplest Lagrangian you could think of to describe the standard model has terms for the boson interactions, the left handed electroweak and strong interactions and the right handed electroweak and strong interactions. Unfortunately this equation is not adequate because it violates the principle of gauge invariance.

Gauge invariance is the cornerstone of modern field theories. It is the requirement that the laws of physics (i.e. the field theories) respect the symmetries of the theory. For example, the familiar gauge transformations for electromagnetism allow for the time derivative of a scalar field to be added to the scalar potential and the gradient of the same field to be added to the vector potential without changing the physical electromagnetic fields. This section describes how the Higgs mechanism saves gauge invariance, both in Standard Model and non-Standard Model theories, and it describes the theory behind Higgs production and QCD background.

The Higgs Mechanism

If we try to make a Lagrangian which respects the symmetries of the groups outlined above we find that we cannot make one with non-zero particle masses. In order to remedy this problem we can add a scalar field to the Lagrangian, which we will term the Higgs field, which is an SU(2) complex doublet with four components, two imaginary and two real. If we introduce this scalar field to the Lagrangian and give it a non-zero vacuum expectation value then we find that we can have massive bosons and particles. In fact, we naturally get massive $W^{\pm}$ and Z bosons while retaining a massless photon.

To understand better how the Higgs mechanism gives mass we can use an analogy to superconductivity. In a superconductor magnetism effectively becomes short-ranged as it is expelled from the conductor. It manages to penetrate the superconductor over a short distance but dies off exponentially. We can interpret this phenomenon as the photon, the

$$\begin{pmatrix} H^+ \\ H_0 \end{pmatrix}$$

FIG. 3: Standard Model Higgs Doublet

mediator of the interaction, having acquired a mass. To see how this occurs we note that in a superconductor there is a weak attraction mediated by phonons of electron with have energies near the Fermi surface and these electrons form a loosely bound state termed a Cooper pair. At sufficiently low temperatures the Cooper pairs form a Bose-Einstein Condensate which causes the magnetic field to become short ranged. One can solve for the exponential decay of the vector potential in the superconductor to prove the damping of the force, or more intuitively one can think of the photon as bouncing around the cooper pairs, always interacting, becoming short ranged.

How does this relate to the Higgs mechanism? If we imagine that there exists a condensate everywhere in the universe we can construct a mechanism which gives us massive particles. The existence of massive particles hints that the condensate should act somewhat like neutrinos, interacting only with the W and Z bosons and not the photon. Therefore we give the Higgs boson doublet the same quantum numbers as the neutrino doublet, namely an isodoublet of hypercharge -1/2. The coupling of this condensate to the W and Z bosons is fixed by the weak interaction to give us the W and Z masses, which we can calculate. The couplings for the rest of the fermions are free parameters in the theory who's values are determined by the observed mass of the particles. These couplings are termed Yukawa couplings.

MSSM

Although the Standard Model and its associated Higgs Mechanism described above naturally generates massive particles and preserves gauge invariance, it is not a theoretically satisfying description of particle interactions. Its list of problems includes bizarre assignments of quantum numbers, such as hyper charge, the varying strength scale of each of the three interactions, electroweak symmetry breaking, existence of three generations of matter, the mass hierarchy of particles which ranges for 6 orders of magnitude, not including neutrinos! . . . . the list goes on. [**?**].

Supersymmetry is a proposition which attempts to remedy some of the problems of the Standard Model through a means not unlike that which Dirac used to solve the crisis of electron self-energy. Classical electrodynamics dictates that an electron in the vacuum will have a self energy proportional to $(1/r_e)$ where r_e is the 'radius of the electron'. Taking the experimental limit on the radius of the electron to be 10^{-17} cm this implies an energy many orders of magnitude larger than the rest mass energy of the electron. This discrepancy was reconciled by the proposal of positrons and Quantum Electro Dynamics which allowed for the positrons to cancel the self-energy of the electron.

The standard model Higgs boson has a similar energy problem which results from loops of virtual particles in Higgs feynman diagrams. To solve the problem we introduce a superpartner to each particle. This superpartner produces the symmetry from which SUSY gets its name. Supersymmetry contains a symmetry between fermions and bosons by postulating that each particle is found in a supermultiplet which contains both fermions and bosons. Hence for each fermionic Standard Model particle we have an associated bosonic supersymmetric particle. These supermultiplets have the nice properties that they are renormalizable (i.e. don't contain infinities). Unlike the positron, however, the superpartner does not have the same mass as its particle partner.

The multiplet structure requires us to now have two Higgs doublets in the SUSY theory, one of which gives mass to the down-type quarks, the other of which gives mass to the up-type quarks. The two doublets allow for eight degrees of freedom (four phases and four amplitudes). Three of the degrees give mass to the known massive bosons, $W^{\pm}, Z$ and the five remaining degrees go to five Higgs bosons, $h, A, H, H^{\pm}$. The H and h are CP even scalars, while the A is a CP odd pseudoscalar.

The MSSM Higgs masses can be parameterized by two free variables, assuming we know well the mass of the Z boson. The first free parameter is the ratio of the vacuum expectation values of the two doublets:

$$\tan\beta = v_2/v_1 \tag{1}$$

The other is the mass of either the mass of the A or mass of the h. For example, if we choose the m_A as our free parameter we can write

$$m_h^2 = \frac{1}{2}(m_A^2 + m_Z^2 - \sqrt{(m_A^2 + m_Z^2)^2 - 4m_A^2 m_Z^2 \cos^2 2\beta}) \tag{2}$$

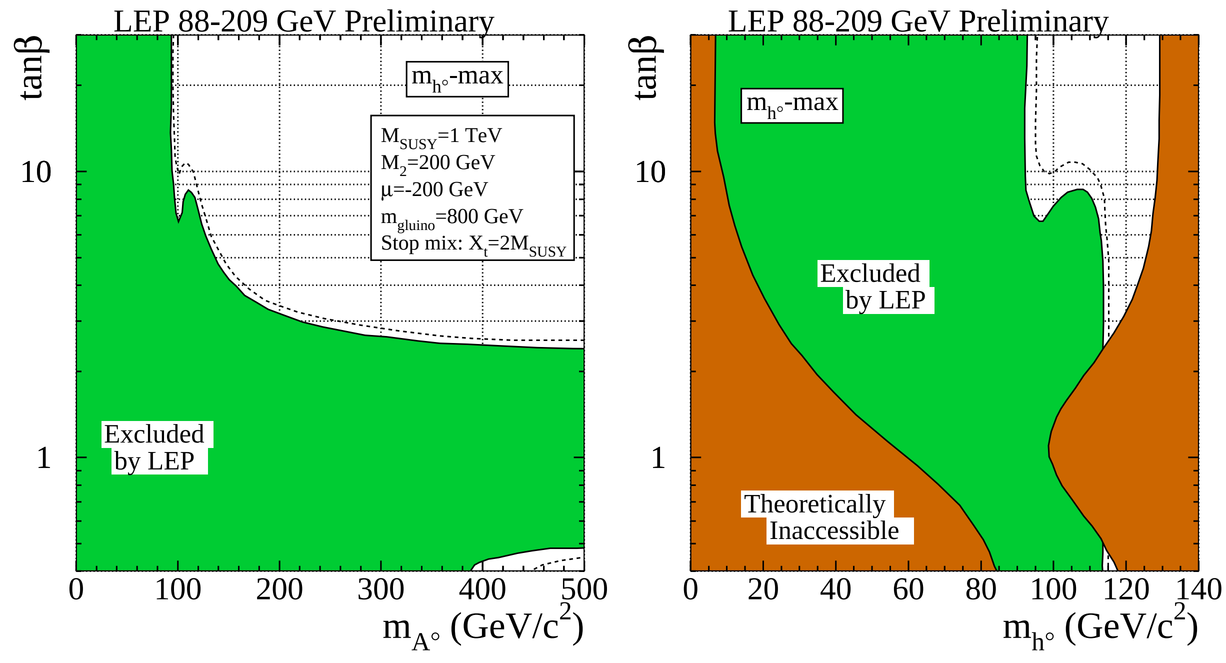

FIG. 4: Current status of MSSM parameter space

The MSSM parameter space has been explored at high energy physics experiments for the past 20 years. Figure 4 shows the current status of MSSM parameter space. Most of the limits derive from direct searches for Higgs bosons produced usually in association with a vector boson at either the TeVatron or at LEP at CERN.

MSSM Higgs Boson Production at the TeVatron

In this analysis we look for h/A. Figure 5 shows branching ratios for different decay modes of the lightest of MSSM HIggs particles. Our analysis searches for Higgs produced in association with two b quarks, decaying to two b quarks. The Higgs and bs are either produced from $q\bar{q}$ annihilation or gg fusion. Why do we choose this channel? Our motivation lies mainly in the fact that the coupling of the Higgs particles to down type quarks is enhanced by a factor of $\tan\beta$ over the standard model. Figure 6 shows the enhancement over the standard model of different interaction vertices for up and down type quarks.

Also, in the region of parameter space still available to the MSSM, a higher $\tan\beta$ is favored, along with a m_h within the energy range of the TeVatron which makes the $bbh \rightarrow bbbb$ a particularly strong signal. Figure 9 shows various production mechanisms for the a light Higgs with a four b quark final state. Additionally, this decay mode allows us to explore

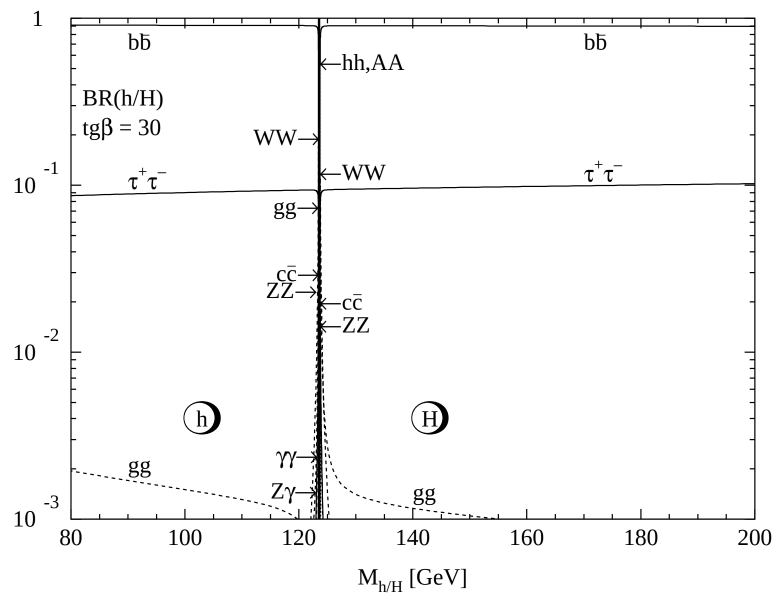

FIG. 5: Various branching fractions for the h and H [3]

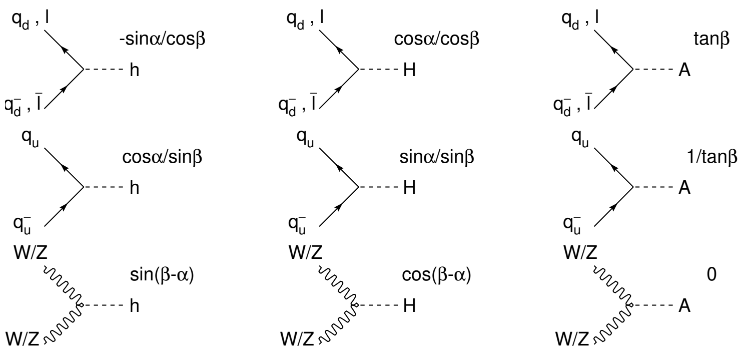

FIG. 6: Enhancements of the MSSM interaction vertex over the Standard Model for various Higgs production processes.

a region of parameter space which is orthogonal to that explored by LEP (see Figure 7).

Figure 8 shows the predicted cross sections for this decay channel at a $\tan\beta$ of 50 and an M_{top} of 175 GeV. (Due to virtual top quark loops in the Higgs diagrams, the Higgs mass is dependent fairly strongly on the mass of the top quark and as CDF and D0 give better values for the top quark mass our predictions of the Higgs mass narrow.)

The $gg(q\overline{q}) \to b\overline{b}A(h) \to bb\overline{b}\overline{b}$ interaction can occur in many different possible interactions. In general we have the two gluons or quarks fusing to form a b and $\overline{b}$ pair. One of the two b quarks radiates a Higgs particle which decays to a $b\overline{b}$ pair. Each of the b quarks produces a

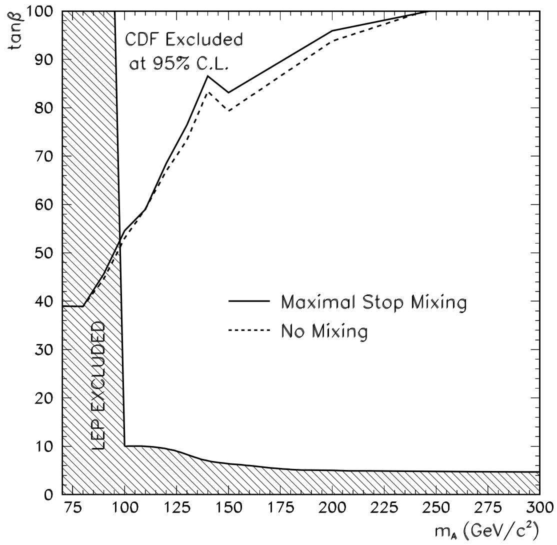

FIG. 7: CDF Run I ($h \to b\bar{b}$) and LEP combined limits

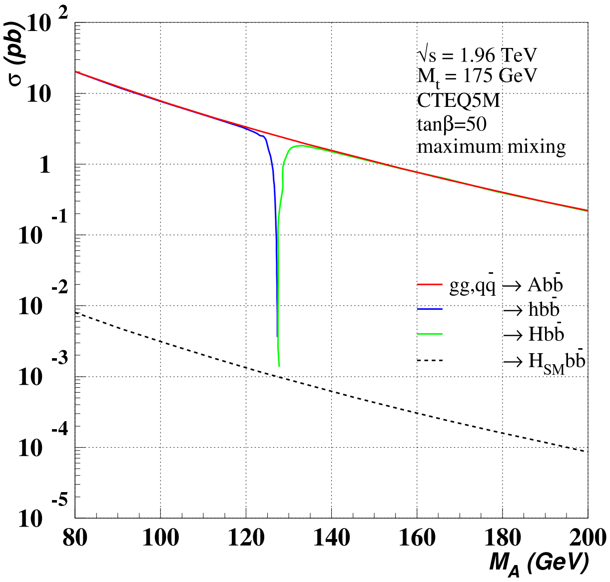

FIG. 8: Projected TeVatron Cross Sections for $bb\overline{bb}$ for Standard Model Higgs, and MSSM A, h, H

jet detectable at CDF. Figure 9 lists a few of the possible Higgs production events.

QCD Background

From a theoretical standpoint the 4b MSSM Higgs signal is a pleasing channel because it has a high probability of being present if the MSSM parameters fall in a reasonable region

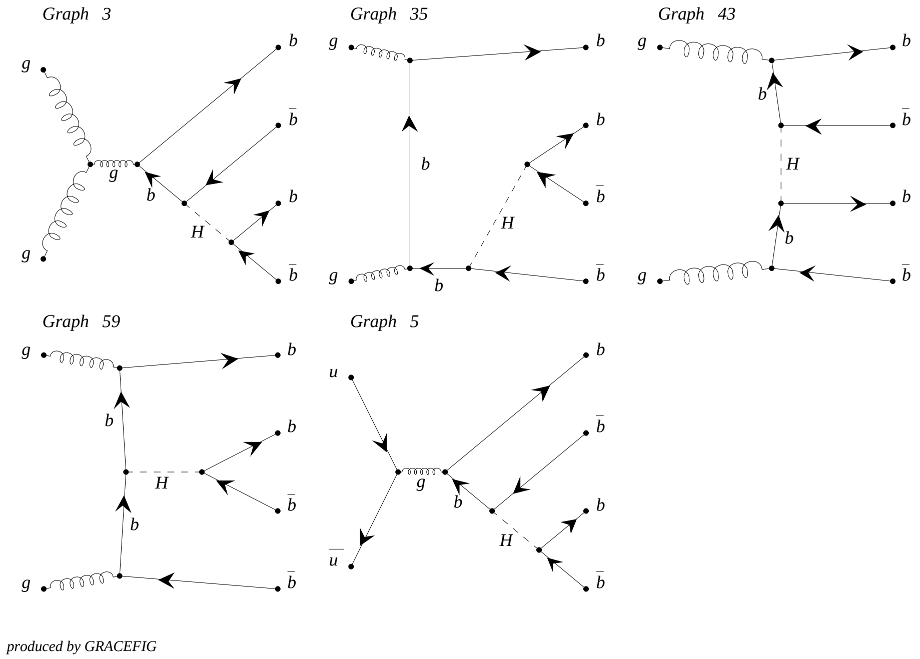

FIG. 9: Various Higgs production mechanisms[3]

of parameter space. Experimentally it is a challenging signal because the background from QCD quark production dominates the energy spectrum. We are particularly concerned with heavy flavor (b, c quarks) which fake our signal events. Since the TeVatron is a hadron collider QCD processes dominate over electroweak ones. Kinematically 4 b quark events are rare because the mass of the b is two orders of magnitude greater than those of the u and d quarks. However, since the QCD cross section dwarfs the electroweak cross section, it is a significant background. Production mechanisms are shown in Figure 10. Monte Carlo simulation models accurately the kinematic distributions for jet events however the overall normalization is unreliable. Therefore we cannot make a theoretical prediction of our background size, rather we estimate it from data as described in Section .

THE TEVATRON AND THE CDF DETECTOR

This analysis uses data collected by the CDF detector at the TeVatron in Batavia, Illinois. The TeVatron is the highest energy particle accelerator in the world today, colliding protons and anti-protons at a center of mass energy of 1.96 TeV at instantaneous luminosities exceeding $5E31\, pb^{-1}s^{-1}$ ($1b = 1barn = 10^{-28}m^2$). These extremely high energies create the necessary conditions for producing the Higgs bosons and other massive particles.

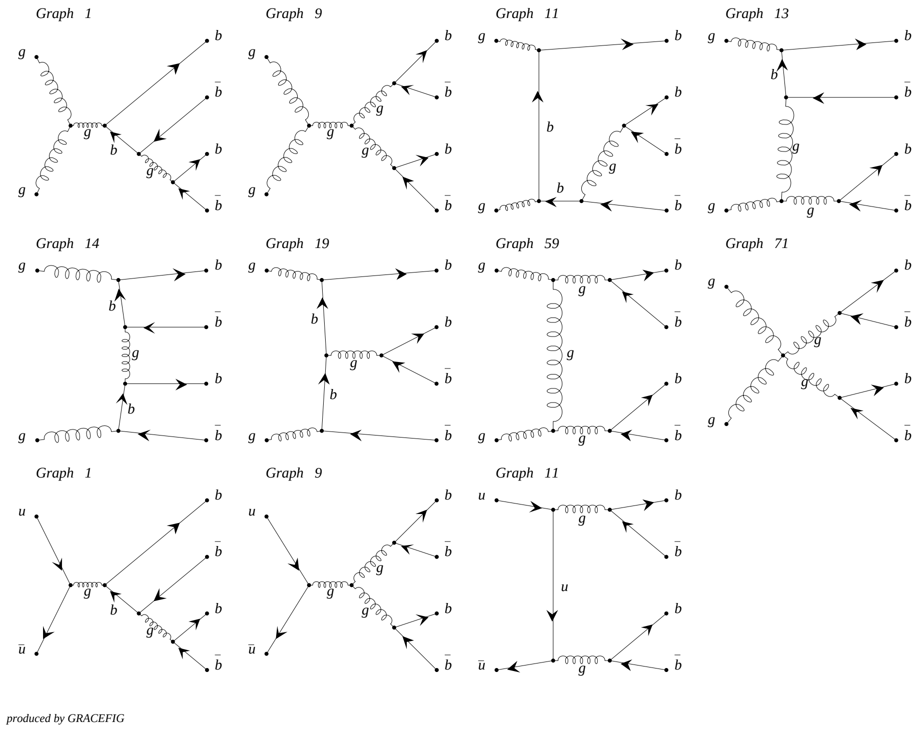

FIG. 10: Various QCD $4b$ production mechanisms

FermiLab and the Tevatron

Fermi National Accelerator Laboratory (FermiLab) is a Department of Energy national laboratory run by the Universities Research Association. The Tevatron uses five stages of acceleration to achieve 980 GeV protons. In the first stage a Cockroft-Walton is used to ionize hydrogen gas, producing H^- ions. These ions then pass through the linear accelerator (LINAC) which uses oscillating electromagnetic fields to accelerate the ions to 400 MeV. At this point the protons are stripped of their electron by passing them through a carbon foil. The beam of protons are then injected into the Booster, a 75 meter synchrotron, accelerating the protons to 8 GeV. Then these protons are collected into bunches of $6 \cdot 10^{10}$ protons before entering the Main Injector. The Main Injector is another stage in the acceleration process where the protons acquire energies of 150 GeV. From this point the protons enter the Tevatron, achieving 980 GeV through the use of superconducting magnets which generate the necessary magnetic fields. The anti-protons are created from 120 GeV protons from the Main-Injector. Bunches of $5 \cdot 10^{12}$ protons are siphoned off to collide with a nickel target, which produces, among other particles, anti-protons. Anti-protons are selected with

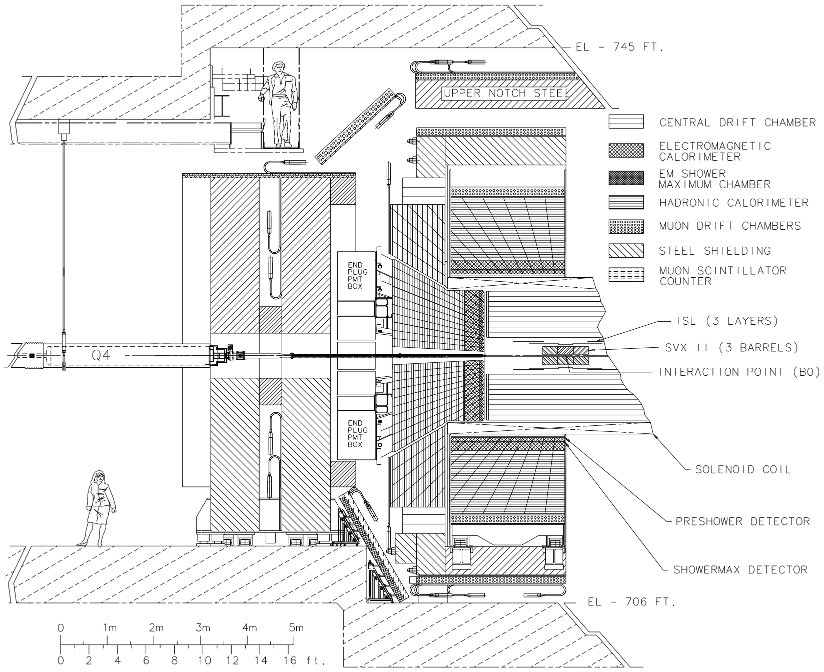

FIG. 11: The CDF detector

an average momentum of 8 GeV and send to the Debuncher and Accumulator which stack bunches of anti-protons which are accelerated to 150 GeV and sent to a counter-rotating beam in the TeVatron for use in collisions. Quadrupole focusing magnets bring the two beams together for collisions at two points along ring, D0, and B0, the center of the CDF detector. The interaction region has a 30 cm spread around the center of the CDF detector and the collisions occur approximately every 396 ns.

The CDF Detector

The CDF detector is an azimuthally symmetric particle detector capable of making tracking, timing and calorimetric measurements. A schematic of the detector is shown in Figure 11. The central region of the detector is approximately 6 m long and covers $|\eta| < 1$ where $\eta = -\ln(\tan(\theta/2))$ and θ is the polar angle measured from the proton direction. r and ϕ define the coordinate system transverse to the beam pipe.

The inner most layers of the CDF detector consist of the tracking systems. Closest to the beam pipe lie the silicon systems, Layer 00, SVX II and the ISL (intermediate silicon layer). The SVX II, designed in large part at LBNL, is comprised of 3 cylindrical barrels

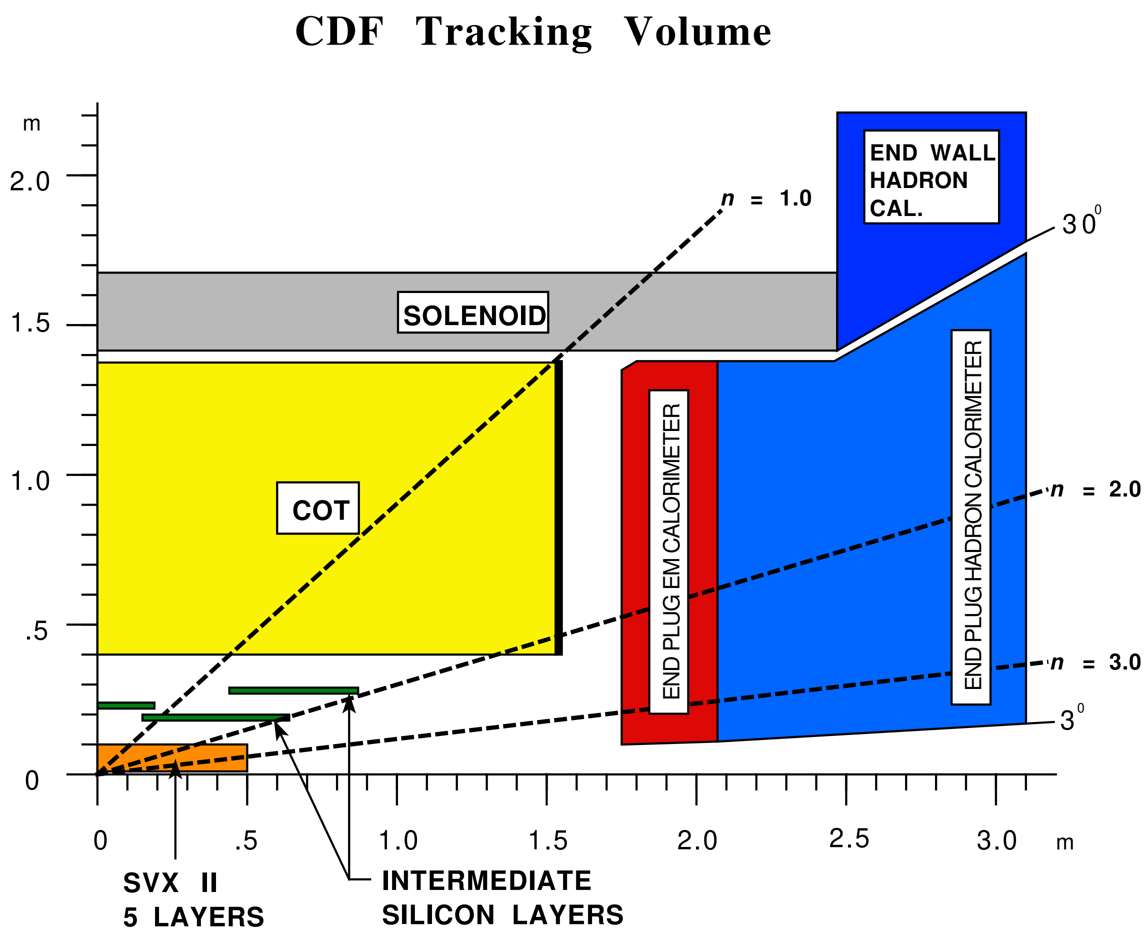

FIG. 12: CDF Tracking Layout

which cover 96 cm of the interaction region and $|\eta| < 2.0$. Each barrel has five layers of double sided silicon microstrip detectors which provide measurements of $r - \phi$ position on one side of the strip and stereo angle on the other side for 3D hit resolution. The silicon system is used for precise tracking and displaced vertex resolution. It provides the ability to b-tag (introduced in techniques section) which is crucial to our analysis.

The intermediate silicon layer (ISL) provides extended track coverage in the forward regions of the detector to compensate for the lack of other tracking in that region. The design of the ISL is fairly similar to the SVX with double sided silicon microstrip detectors providing 3D track resolution. Figure 12 shows a schematic of the tracking volume.

The main tracking engine of the CDF detector is the Central Outer Tracker (COT) a 3 meter long drift chamber which provides tracking for $|\eta| < 1.0$ and $44\ cm < r < 132 cm$. The COT has 4 axial and 4 stereo superlayers filled with gas consisting of 12 sense wires each to provide 3D tracking.

The entire tracking system is contained in a magnetic field of 1.4 Tesla provided by superconducting solenoid at a radius of 132 cm from the beam pipe. The magnetic field allows for precise momentum resolution of charged particles.

Outside of the tracking systems sit the Electromagnetic and Hadronic calorimeters. The electromagnetic and hadronic calorimeters are made of scintillating material and are divided into central ($|\eta| < 1.0$) and plug ($1.0 < |\eta| < 3.0$) regions. The electromagnetic calorimeter uses a lead-scintillator combination whereas the hadronic calorimeter uses iron-scintillator. The calorimeters provide precise information on the energy of particles and allow for easy detection of electrons and hadrons.

Lastly, CDF has muon detector which are situated at the furthest radii from the beam pipe and are comprised of steel absorbers and proportional counters and scintillators to create muon track stubs.

The data acquisition system of CDF is a massive hardware and computing effort to make sense of the data streaming through the CDF detector. The system must reduce an approximately 2.5 MHz collision to a 70 Hz readout rate, which it achieves through a trigger system to select interesting events in 3 levels. Figure 13 shows the flow chart for CDF data.

The data flows from front-end/Level 1 (40 kHz accept rate) hardware to the Level 2 (300 Hz accept rate) trigger electronics and VME Readout Boards then to the Level 3 trigger processing farm, which fully analyze each event which has passed the previous triggers. Those events selected by the Level 3 trigger (approx 50 Hz) are then sent to the data logger which writes some data to tape and uses some for online monitoring of the detector systems. Level 1 uses single trigger calorimetry towers, muon chamber information and COT basic tracking to identify physics objects such as electron, muons, jets and tracks. Level 2 partially reconstructs each event and applies detector specific identification cuts to better define the physics objects passed in Level 1. Additionally, at Level 2 the Secondary Vertex Tracker (SVT) uses the basic tracks found by the Level 1 tracker (XFT) and matches them to hits in the silicon system. It then reconstructs silicon tracks and searches for two track vertices that are displaced by 100 μm to 1 mm from the beam spot. The SVT is designed to tag heavy flavor particles which have a finite lifetime visible in the silicon detector. Level 3 uses the same code as offline analysis to fully reconstruct each event and make final trigger decisions.

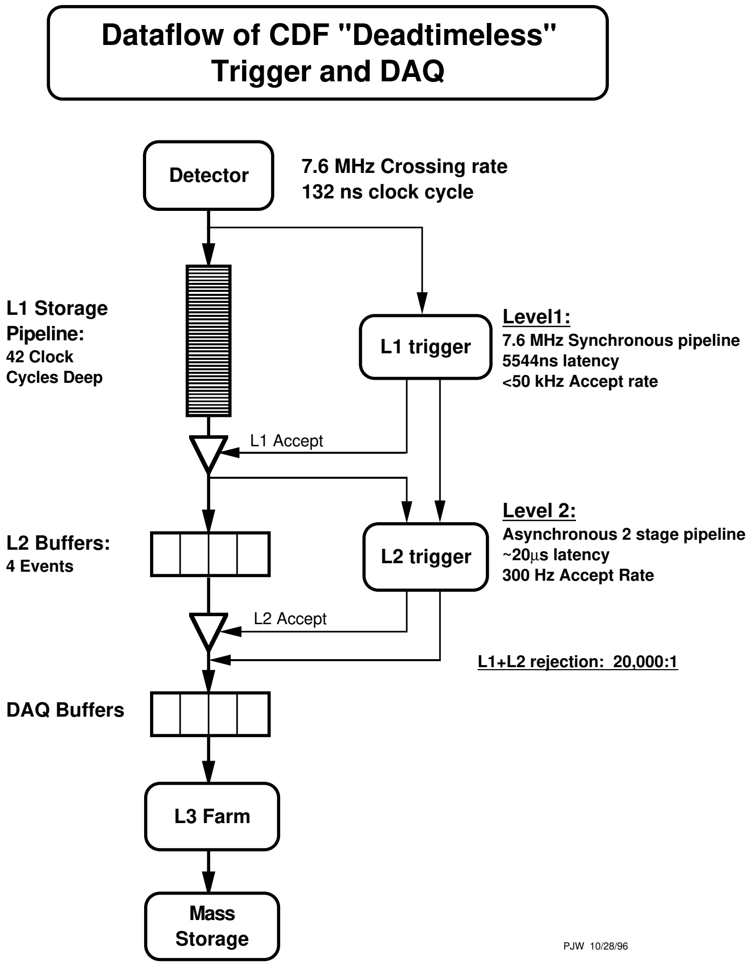

FIG. 13: CDF Data Flow

DATA SET AND EVENT SELECTION

The data used in this analysis is a subset of the 160 pb^{-1} CDF data collected in Run II from proton anti-proton collisions. The events used in the final analysis were selected in two steps: trigger and event selection. The trigger reduces the data to events that are of interest in this analysis in real time. The event selection occurs after the data is written to tape as we try to refine our data set.

Signal Monte Carlo

To determine our trigger and selection criteria we first study the $b\bar{b}h \to b\bar{b}b\bar{b}$ process in Monte Carlo simulation. We use Pythia $gg \to Ab\bar{b}$ and $q\bar{q} \to Ab\bar{b}$ events generated at $\tan\beta = 50$ and $m_A = 90$ GeV. Figures 14 and 15 show various kinematic and geometric

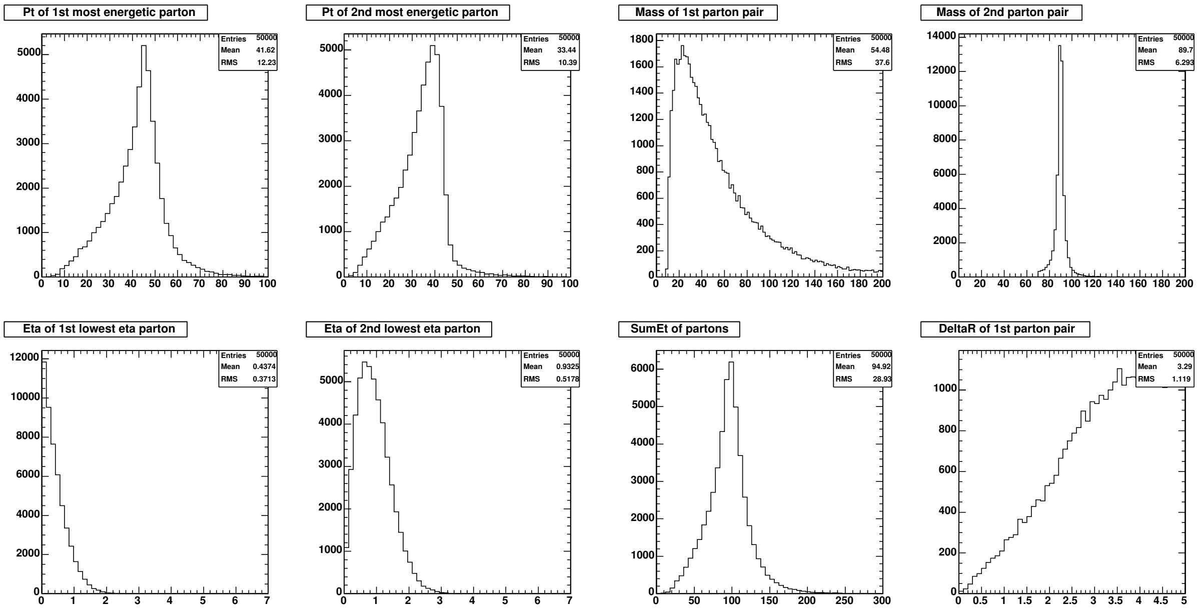

FIG. 14: Plots of kinematic variables for ggA Monte Carlo generated at $\tan\beta = 50$ and $m_A = 90$

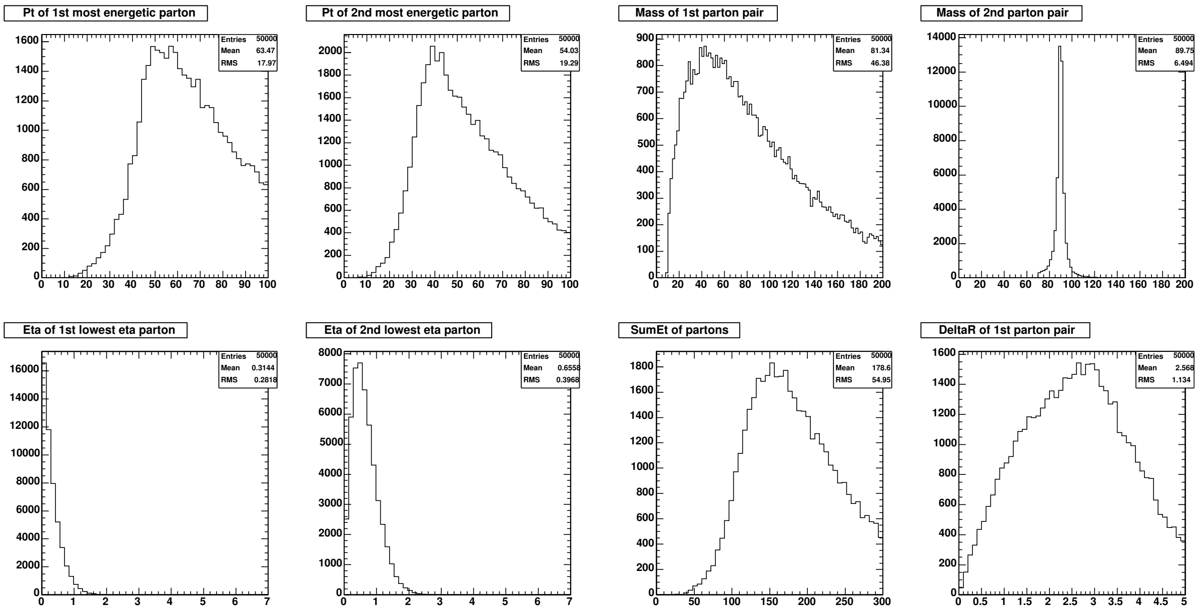

FIG. 15: Plots of kinematic variables for qqA Monte Carlo generated at $\tan\beta = 50$ and $m_A = 90$

distributions of our signal events.

TABLE I: MultiJet Trigger

Level 1	$\geq$ 1 jet $E_T > 10$ $\Sigma E_T > 90$ GeV
Level 2	SVT Tracks ≥ 2 $P_T > 2.0$ GeV $100\mu\mathrm{m} < d_0 < 1$ mm
Level 3	$\geq$ 3 jets $E_T > 10$ GeV $\Sigma E_T > 100$ GeV

Trigger

We have implemented a trigger designed record events that contain four or more high transverse energy jets, at least two of which have displaced vertices. The trigger is summarized in Table I.

At Level 1 we require one jet with a minimum of 10 GeV of transverse energy (E_t) and a total transverse energy in the event of at least 90 GeV. Using 90 GeV signal Monte Carlo we find Level 1 to be 5.7% (68.8%) efficient for $ggA(qqA)$ (see Figure 16). Level 2 exploits our Secondary Vertex Tracker (SVT) requiring 2 SVT tracks of transverse momentum (p_T) greater than 2 GeV and an impact parameter (d_0) greater than 100 microns and less than 1 millimeter. This portion of the trigger is designed to select b jets which have displaced vertices due to the long lifetime of the b. Level 2 is 68.1% (82.2 %) efficient for $ggA(qqA)$ as shown in Figure 17. At Level 3 the requirements are more sophisticated because the L3 computing farm can do event reconstruction. We require three jets with E_T greater than 10 GeV and a ΣE_T greater than 100 GeV. The Level 3 cuts leave us a 1.5%(37.7 5%) efficiency for our trigger on $ggA(qqA)$. This is not an ideal efficiency, so we are investigating changing the cuts to better reflect our signal kinematics (see Figure 18). At the same time, we can see that the trigger does do a decent job of cutting out our background relative to the signal. When looking at the fraction of signal to background that passes each level we see that we gain each time. The results are summarized in Table II.

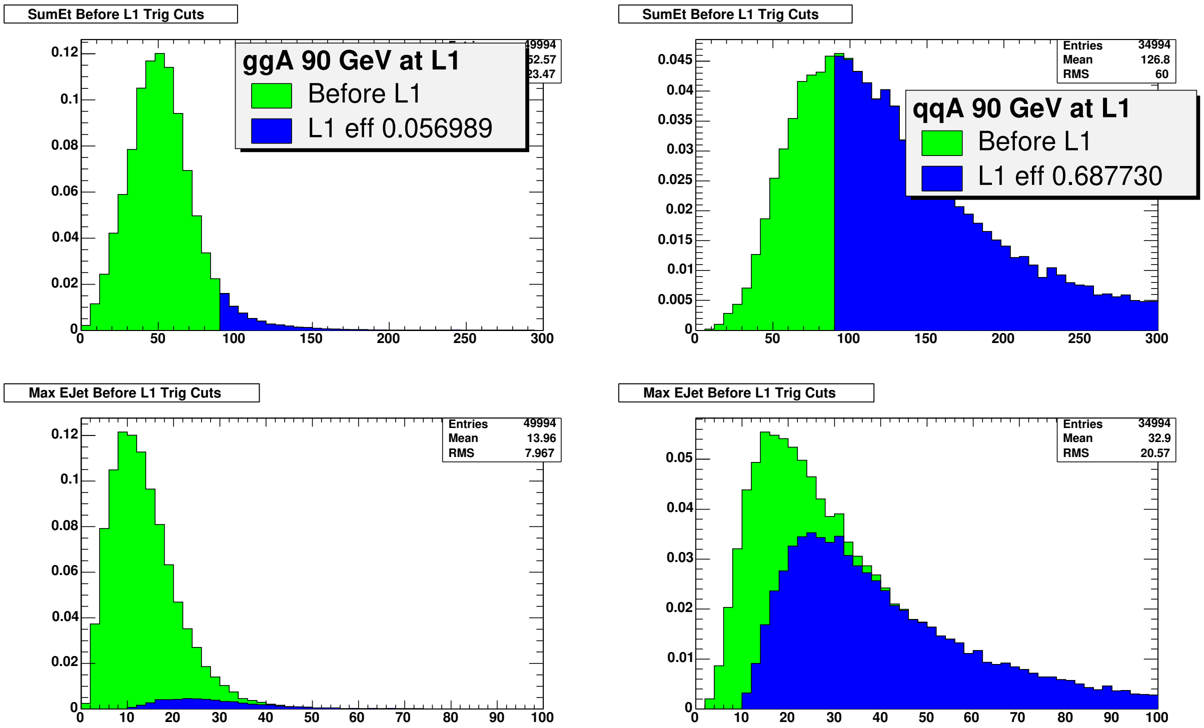

FIG. 16: Level 1 trigger efficiency for ggA and $q\bar{q}A$ 90 GeV signal Monte Carlo ($\tan\beta = 50$)

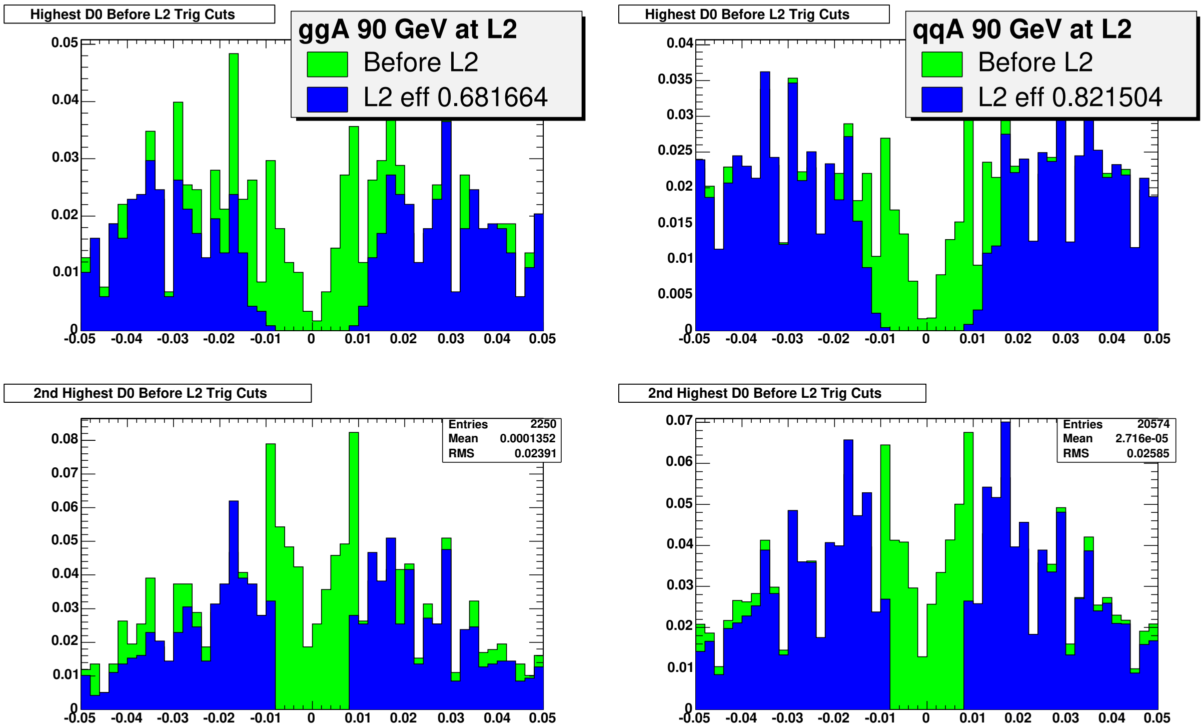

FIG. 17: Level 2 trigger efficiency for ggA and $q\bar{q}A$ 90 GeV signal Monte Carlo ($\tan\beta = 50$)

TABLE II: MultiJet Trigger Efficiency on ggA and qqA

	$m_A = 90\mathrm{GeV}$, $\tan\beta = 50$	
Requirement	$gg \rightarrow bbA$	$qq \rightarrow bbA$
Passes L1&L2&L3	1.5%	37.7%
2+ offline b-tags	32%	47.5%
3+ offline b-tags	9.5%	18.5%
Total eff	0.14%	7.0%
gg+qq eff	0.2%	

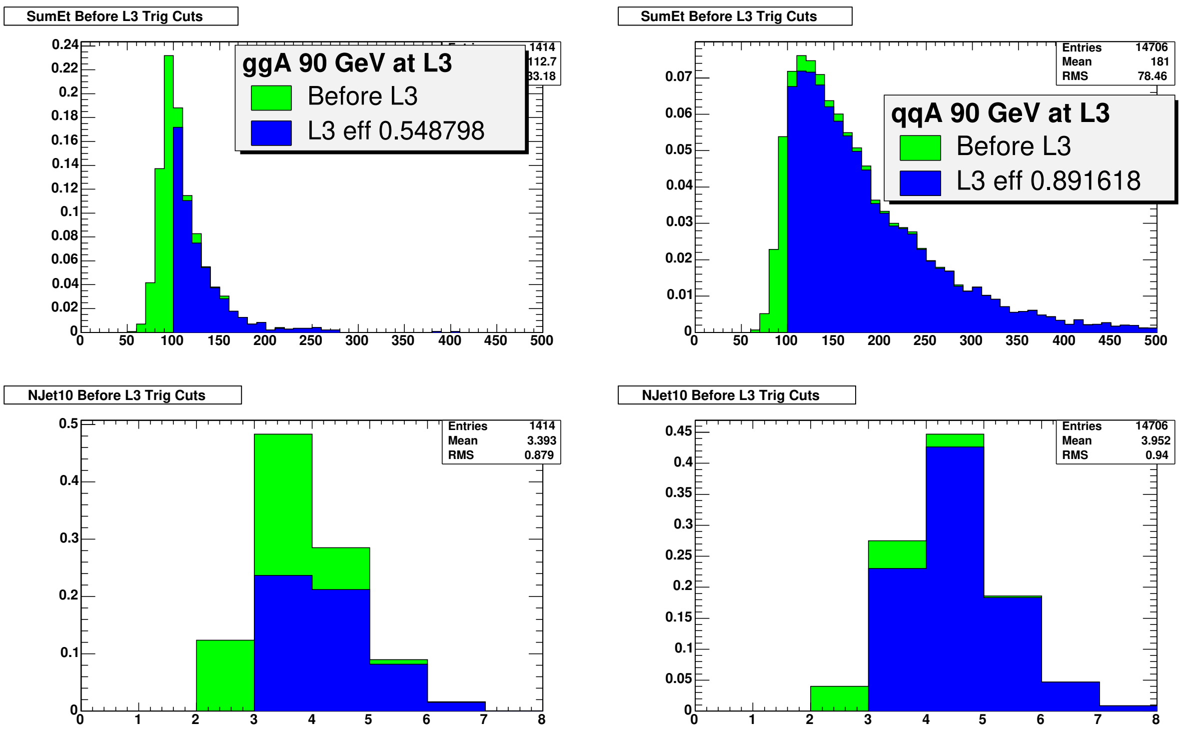

FIG. 18: Level 3 trigger efficiency for ggA and $q\bar{q}A$ 90 GeV signal Monte Carlo ($\tan\beta = 50$)

Event Selection

In addition to the trigger level requirements, we also make more sophisticated analysis cuts to further eliminate background to our signal. First we do a general clean up of the data by rejecting events with cosmic rays or identified electrons or muons since leptons do not appear in the final state of our signal. We then require a $\Sigma E_T > 125$ GeV and at least three b-tagged central jets. In the future we plan to also make an optimized mass dependent jet E_T cut on the first, second and third highest E_T jets which accounts for differing kinematics associated with different possible A masses.

Background

This analysis has four primary backgrounds

$$\begin{array}{ll} QCD & b\bar{b}(c\bar{c})b\bar{b}(c\bar{c}), b\bar{b}(c\bar{c})b(c)j \\ t\bar{t} \rightarrow WbWb & b\bar{b}cjjj, b\bar{b}cjcj \\ W(Z)b\bar{b}/c\bar{c} & \\ fakes & jjjj, b(c)jjj, b\bar{b}(c\bar{c})jj \end{array}$$

We model the QCD and electroweak processes with Monte Carlo and determine the fake tag distributions from data. The QCD background is our primary background to the analysis. We generate $b\bar{b}b\bar{b}$ Monte Carlo events using GR@PPA for a cross section of 503.976 picobarns. Figure 19 show kinematic and geometric variables comparable to those for our signal. To measure the double tag background we also examine Monte Carlo of events containing two b quarks and 2 light flavored jets, with a minium parton energy cut of 15 GeV. We term this Monte Carlo bb2p15.

TECHNIQUES

In order to perform this analysis we employ a wide range of techniques to help us extract a signal from the data. This section describes in detail each of those techniques.

B-Tagging

B-Tagging is our most powerful analysis tool for this search. Since our signal has four b quarks in the final state, if we can distinguish b jets from light flavored jets we can gain a considerable advantage over the massive multi-jet background. B-tagging makes use of our silicon detector's fine track resolution to reconstruct displaced vertices which occur when neutral heavy particles, such as B^0 mesons, decay to charged daughters. B mesons have on average a lifetime of 1.56 picoseconds, which in the reference frame of the detector correspond

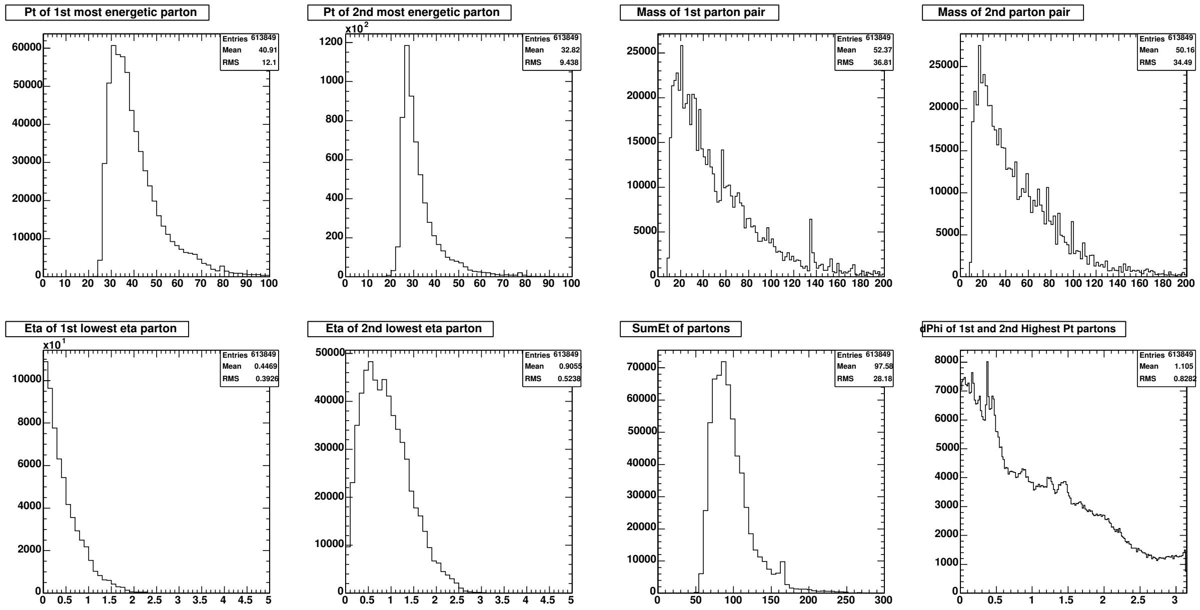

FIG. 19: Plots of kinematic variables for QCD background Monte Carlo

to distances on the order of a milimeter (distance of flight before decay). By contrast light flavored mesons such as the K^0 and Λ^0 particle have no measurable lifetime. Medium sized particles, such as the tau lepton and charm quark have lifetimes half as long as the B meson lifetime and contribute to the number of mistakenly tagged (or mistagged) jets. B-tagging works as follows: the offline reconstruction code searches all combinations of two or more good tracks for a displaced vertex. Once a pair is found, its L_{xy} is calculated as the distance from the interaction point to the vertex. Also, an error, σ_{xy} is calculated with typical values around 150 μm. If $L_{xy}/\sigma_{xy} > 3$ and the tracks are associated with a jet, then the jet is declared as b-tagged. The sign of the tag is determined by the orientation of the vertex to the jet direction. If the vertex and the jet are oriented in the same direction we have a positive tag, if they are reversed it is a negative. Figure 20 illustrates these two cases.

It is very important for us to know our efficiency for finding events if we are trying to extract a signal, correspondingly we must know our b-tagging efficiency for single, double and triple b-tagged events. The single tag efficiency for high energy b jets is roughly 50% as estimated from top tagging efficiencies.

The double tag efficiency is not simply the square of the single tag efficiency for several reasons. First, b quarks are produced in pairs, hence if you tag one jet there is a higher probability that there is a second b jet in the event. Secondly the efficiency per jet is

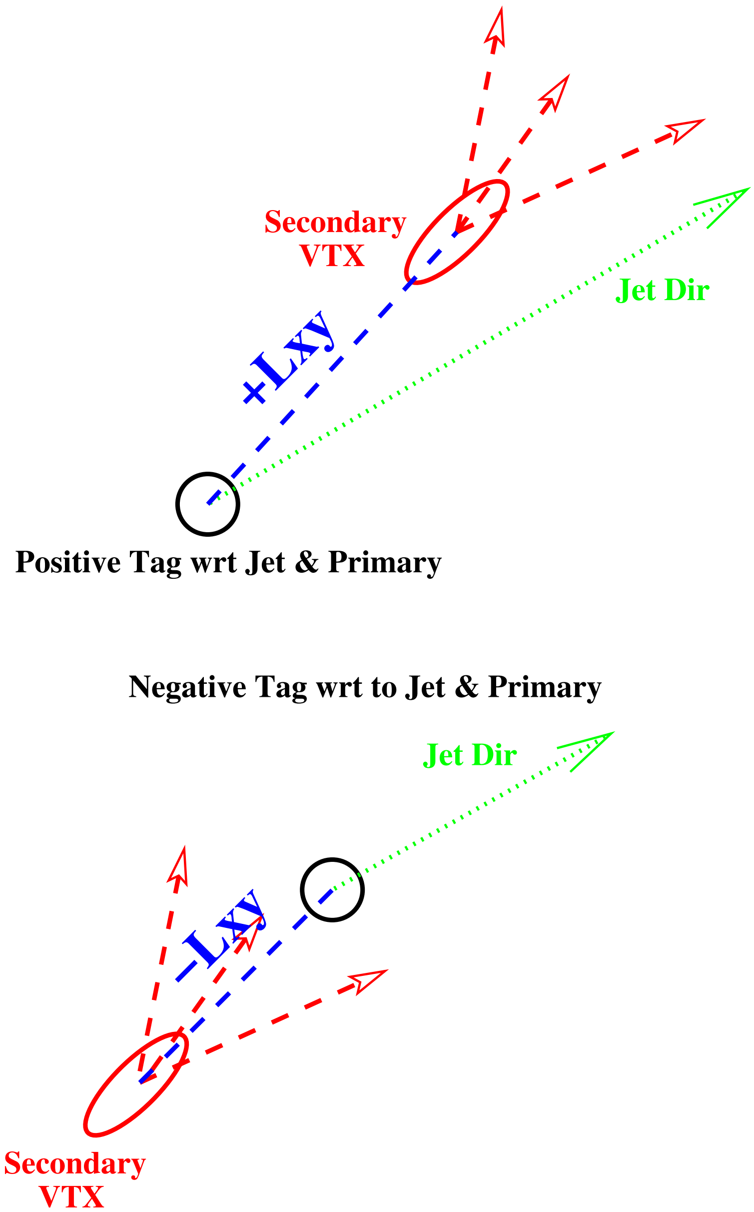

FIG. 20: Positive and Negative b-tags

dependent on various jet characteristics such as energy and position in the detector which can be related for two jets in dijet events. We determine our double tag effiency for our signal from Monte Carlo to be 32% for gg production and 47.5% for $q\overline{q}$ production. Similarly we measure the triple tag efficiency for our signal to be 9.5% for gg production and 7.0% for $q\overline{q}$ production.

FIG. 21: Observed and predicted fake tag distributions

Fake Tags

Unfortunately our Monte Carlo does not measure the fake tag rate very accurately. In order to determine it we look at the negative tag rate in the inclusive jet data. We then parameterize the negative tag rate in terms of E_T of the jet, to account for better vertex method for higher E_T jets, ΣE_T of jets in the event to remove some of the jet characteristic dependence, N_{good}^{trk} (Number of good tracks in event) to account for tracking efficiency, and ϕ, η to factor in geometric effects in the detector. We assume that the number of negative tags is equal to the number of *fake* positive tags and use the combination of the two to calculate our fake tag rate. We then need to figure out what is our rate to mistag a jet in events with more than one tag.

In [2] it is shown that the number of fake tags in a double tag sample is given by

$$N_{++(fakes)} = N_{+-} - N_{--} \tag{3}$$

where $N_{++(fakes)}$ is the number of expected *fake* double positively tagged events and N_{+-} and N_{--} are the number of events with a two negative and a negative and positive tag, respectively. Similarly, for triple tagged events it can be shown that

$$N_{+++(fakes)} = N_{++-} - N_{+--} + N_{---} \tag{4}$$

We apply the above formualas to our dijet and signal data and predict the fake positive tag distributions. Figure 21 shows a comparison between the predicted and observed negative and positive negative jet combinations for the dijet data. Figure **??** show the expected fake tag distributions generated by our mistag matrix for the double tag invariant mass specturm.

Fit Technique

Monte Carlo accurately represents the kinematic distributions of jet events however it does not well model the overall cross section of such events. In order to determine the fraction of QCD triple tagged events in our data sample we must take the normalization from data. This is achieved by first fitting the double tag QCD fraction in a data sample of double-tagged jets and then using the ratio of double to triple tagged events in Monte Carlo to scale our triple tag fraction. The justification behind this method is that in the double tag sample we are completely dominated by QCD events, hence our signal is negligible and the fraction we fit will accurately describe the QCD fraction. Additionally the ratio of double to triple tag QCD events in Monte Carlo agrees well with data.

To fit the double tag fraction we utilize the fact that we have the double tag invariant mass spectrum from data and the double tag mass invariant mass spectrums from each of our monte carlo sources. We employ an extended maximum likelihood technique to determine the fraction each of the Monte Carlo sources (which are termed templates) contribute to the data sample. We choose an extended maximum likelihood over the usual likelihood to account for finite monte carlo statistics on a bin by bin basis which was developed by Roger Barlow and Christine Beeston[**?**]. The method works as follows:

For each data bin we assume that some function f_i which determines the number of predicted events for that bin and is dependent on each of the templates (MC sources). We have that

$$f_i = N_D \sum_j P_j A_{ij}/N_j = \sum_j p_j A_{ij} \tag{5}$$

where N_D is the total number of data events, P_j is the proportion of source j in the data, N_j is the number of events from source j and A_{ji} is the expected number of events from source j in bin i. Using this function (which is only a function of i since j is summed over) we can use Poisson statistics to find the probability of observing d_i data events in bin i

$$P = \exp -f_i \frac{f_i^{d_i}}{d_i}. \tag{6}$$

Then we can determine the p_j by minimizing the logarithm of the total likelihood

$$\ln \mathcal{L} = \sum_i d_i \ln f_i - f_i \tag{7}$$

Frank Filthaut wrote the code to perform this likelihood fit in ROOT[4]. We have modified this code to add a gaussian constraint to certain Monte Carlo templates where the fraction is known to some certain degree of accuracy. Our likelihood function now becomes

$$\ln \mathcal{L} = \sum_i d_i \ln f_i - f_i + \sum_{j \in S} -\frac{1}{2} \frac{(P_j - P_{j0})^2}{\sigma_j^2} \tag{8}$$

where S is the set of templates where we have a constraint on the overall fraction and σ_j is the uncertainty on fraction j and P_{j0}) is the expected fraction of source j.

We have also determined that it is necssary to put a constraint on the relative fraction of two of the background sources of our measurement. We plan to implement this constraint by parameterizing one fraction in terms of the other and requiring that their ratio remains constant. This procedure is explained more concretely in the next section.

RESULTS

Our analysis is currently working on determining the fraction of QCD events in the double tag sample. We are performing a "blinded" analysis, i.e. we are not examining the data until we are ready to make the final measurement. The results presented here concern background measurements. As listed in Section there are four primary background sources to this measurement. We have suceeded in determining the fake tag fraction, as shown in Section , and due to the efforts of LBL group members, we have an accurate measurement of the top quark cross section which gives us a constraint on the possible $t\bar{t}$ contribution to our data.

As mentioned previously, we cannot take the heavy flavor jet cross section from Monte Carlo because it is inaccurate so we plan to measure the cross section in data from the double tag sample and apply it, modified by the ratio of double to triple tag events, to the triple tag sample.

We measure the double tag QCD fraction by using the extended maximum likelihood method described above. We use bb2p15, bbbb, fake tag and inclusive top Monte Carlo as our template sources. For the double tag spectra we ignore the Electroweak (W,Z) background and Higgs signal because their cross sections are so much lower than that of the QCD double tag cross section that they are negligible. Figure 22 shows the unnormalized dijet invariant mass spectra for each of the MC templates used in the fit. We put constraints using the top cross section on the number of inclusive top events, using the uncertainty on the cross section to derive the uncertainty on the fraction. We also constrain the fake tag fraction to be .185 given that we know the fraction of mistagged events. We leave the QCD backgrounds unconstrained. Using these templates and constraints we are able to make a preliminary fit to the dijet invariant mass spectrum. Table **??** summarizes the results and Figure 23 shows a plot of the predicted distributions.

In order to make the fit more accurate we would like to add a constraint that fixes the

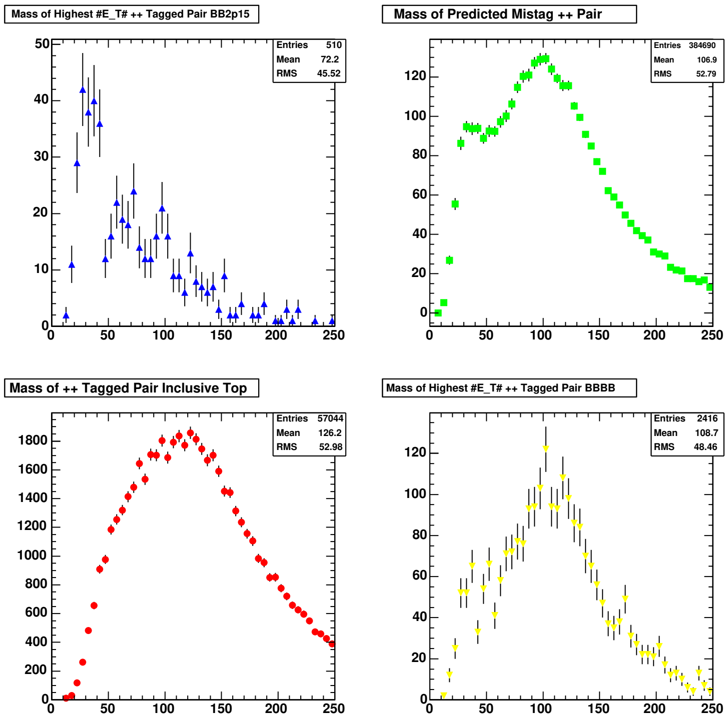

FIG. 22: Double tag dijet invariant mass spectrum for each of the background sources used as templates: bb2p15(top left), fake tags (top right), inclusive top events (bottom left) and QCD 4b events (bottom right).

TABLE III: Preliminary Double Tag Results

Template	Constraint	Fraction	Error
bb2p15	none	.7534	.05446
inclusive top		.004989	.0005750
fake tags	.185	.1828	.01878
bbbb	none	.058897	.04710

ratio of *bb* events to *bbbb* events, which is known. We are in the process of implementing this technique and will measure the triple tag fraction when we have confidence in the double tag results.

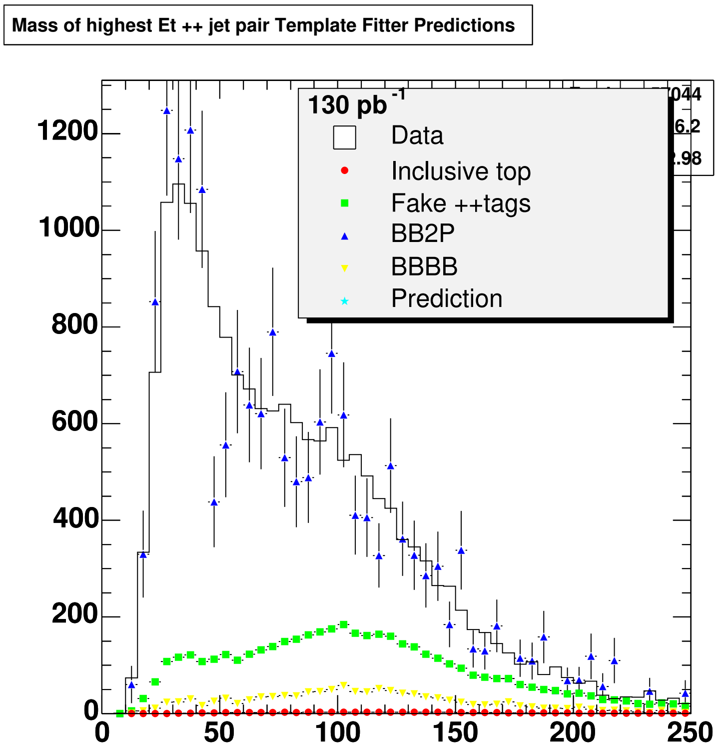

FIG. 23: Preliminary fit to dijet invariant mass spectrum.

CONCLUSIONS

We have almost achieved the understanding necessary to 'open the box' and make a qualitative search for the MSSM Higgs bosons. We understand our background and Monte Carlo and have almost completed work on our fitting tools. Once complete we will be able to make a measurement of the triple tag background and apply it to the data to search for an excess. Hopefully then we will learn more about the MSSM parameter space and the Higgs bosons.

With the rapid approach of the completion of the LHC the search for new physics become more important because if it exists we should see hints of it in current data. Searches such as this one will lay the foundation for work to be done in the next generation of high energy experiments.

I would like to thank the MSSM 4b group: Aaron Dominguez, Pete McNamara, Jared Yakomota, Steve Worm and Kevin Burkett, in particular Aaron Dominguez for his help and guidance. I would also like to thank Young-Kee Kim and Jim Siegrist for their advising on

this thesis.

---

[1] H. Murayama. *Supersymmetry Phenomenology* hep-ph/0002232

[2] J.F. Troconiz, J.A. Valls, and R. Vilar, *Update in the Search for a Neutral Heavy Scalar in the Multijet Data Sample*, CDF Note 4129. J.F. Troconiz, J.A. Valls, and R. Vilar, *Special Issues Relative to the Multijet Higgs Analysis*, CDF Note 4131.

[3] M. Carena, J. S. Conway, H. E. Haber, J. D. Hobbs, *Report of the Higgs Working Group of the Tevatron Run 2 SUSY/Higgs Workshop*, Fermilab-Conf-00/279-T, hep-ph/0010338

[4] F. Filhaut. Root Classes and Members Reference Guide version 3.05. http://root.cern.ch

[5] A. Dominguez. *Search for Neutral Higgs Bosons in e+e- Interactions at center-of-mass energies between 130 GeV and 183 GeV.* Ph.D. Disseration. University of California at San Diego. October, 1998.