UNIVERSITÀ DI PISA
FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
DIPARTIMENTO DI FISICA “E. FERMI”

CORSO DI LAUREA IN FISICA
Anno Accademico 2004 – 2005

Elaborato Finale:

Methods for the Measurement of the Top Quark Mass

Candidato

Nicola Pozzobon

Relatore

prof. Giorgio Bellettini

Contents

Introduction

In 1995 the top quark was discovered at Fermi National Accelerator Laboratory. Ten years later physicists still produce a remarkable effort to measure its mass. Why is this measure so important? And which techniques are employed?

To understand these topics, the candidate attended a three-weeks training in Fermilab, supervised by physicists of INFN (section of Pisa) and JINR (Dubna), where he got in contact with the CDF detector, the physics program and the offline analysis resources at CDF. Then he carried on different tests on a particular selection, called “cleaning”, that could be used in top mass analysis involving the so-called template method, using a set of Montecarlo simulated events. The aim of these tests was to check the effects of cleaning, particularly to know how much the top mass resolution is improved, and how often the event is interpreted in the proper way. If cleaning events improved the fraction of correctly reconstructed masses or the resolution on the top mass, it could be useful in the study of top candidate events which do not show some crucial features, such as the identification of jets as b-jets.

This report presents the preliminary results of such test, showing how they are unsatisfactory about reduction of the spread of the mass distribution but encouraging in terms of “topological matching” and parton-to-jet assignment.

1. The Top Quark

1.1 The Top Quark in Standard Model

The top quark, first seen as an evidence in 1994 and later confirmed as a discovery at the Fermi National Accelerator Laboratory in 1995, is one of fundamental constituents of matter and is believed to be the last of the 6 existing quarks. According to the Standard Model, all matter is composed of 12 spin $\frac{1}{2}$ particles (fermions): 6 quarks (carrying fractional charge) and 6 leptons (carrying integer or zero charge). Each fermion is associated to its antiparticle, carrying opposite charge and magnetic moment, but with the same quantum numbers and mass.

As soon as it was understood that the narrow resonances of mass $\simeq 9.5$ GeV discovered at Fermilab in 1977 in a μ pair production experiment were bound states of a fifth quark, the b quark, it appeared clear the its partner in a weak isospin doublet ought to exist. Experience shows that weak decays of the second doublet quarks (strange and charm quarks) are dominated by charged current (charge-exchange) decays. Flavour Changing Neutral Current (FCNC) decays like $s \to d$ or $c \to u$ are strongly suppressed. This is explained in the Standard Model by a compensation of decay contributions to the weak doublet eigenstates which mix into physical mass eigenstates. The occurrence of weak eigenstates as doublets is essential for this mechanism to work(1). It was soon observed that even in the b quark decay FCNC were suppressed: only $b \to c$ decays were observed. As a consequence, the b quark ought to be the lower member of a third doublet.

In the 80's electron-positron annihilation experiments at the PETRA storage ring (DESY, Hamburg) showed a clear forward-backward asymmetry in the bb final state jets polar distribution. A $\cos\theta$ term is indeed expected in the center of momentum system angular distribution of this process with coefficient proportional to the third component of the weak isospin of the b quark(2). A fit to the data indicated $I_3^b = 0.5 \pm 0.05$(3). The b quark was thus proved to be a member of a weak doublet. However, there was no value imposed by the SM to the top quark mass. Indirect information was obtained by fits to a number of observables whose value is sensitive to virtual top quark exchange loops. In 1995 a value around 175 GeV was indicated.

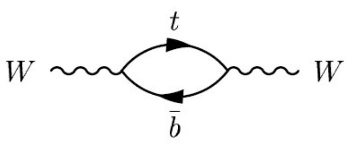
---

(1) A clear explanation of FCNC suppression and GIM mechanism can be found in [PER].

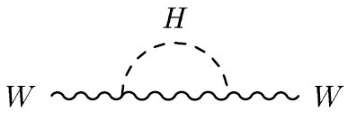
(2) More details on this dependence can be found in [PDG].

(3) I_3^b is the z-axis component of b-quark weak isospin.

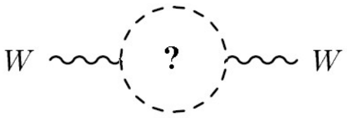
1.2 Why an Accurate Measurement of the Top Quark Mass?



$\Delta M_W \propto M_t^2$



$\Delta M_W \propto \ln M_H$



New Physics

Top exchange loops affect the physical W mass to an extent which depends on the square of the top mass. Through the same mechanism (although with logarithmic terms) M_W depends also on the Higgs boson mass. Within the SM on comparing M_W versus M_t one can thus obtain an indirect information on the Higgs boson mass. Direct searches for Higgs production at LEP2 have excluded $M_H < 114.4$ GeV at 95% confidence level. While waiting for the discovery of the Higgs, information on the upper limit of M_H is the best possible progress in determining this basic parameter of the SM.

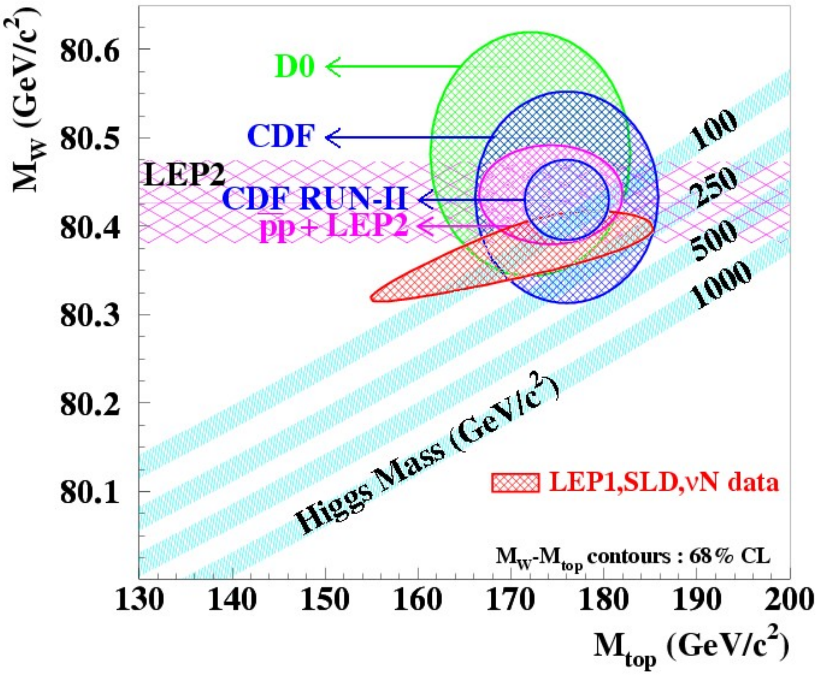


Figure 1.1: Relations linking M_t, M_H and M_W, comparing LEP2 results, CDF and DØ results, and Tevatron RUN II expectations; LEP 1 spot is not symmetric because M_t was only indirectly determined (See Figure 1.2).

At the end of RUN I in 1997 the top mass was measured, by adding CDF and DØ information, as $M_t = 178 \pm 4.3$ GeV. The derived information on M_H is interesting but still numerically rather poor:

$$M_H = 117^{+67}_{-45} \text{ GeV} \Longrightarrow M_H < 251 \text{ GeV (95\% CL)} \tag{1.1}$$

In RUN II both CDF and DØ are expecting to integrate a luminosity – see (2.1) – from 4 to 6 fb^{-1}. Already with $\simeq 2$ fb^{-1} the uncertainty on the top mass is expected to be reduced (as allowed by the systematic errors) to about 2 GeV, providing a much more significant information on M_H. As the top is the only fermion with mass of the same order of magnitude as the masses of W and Z, which are the mediators of EW interactions, there is the hope that some top parameters, or some of its production or decay properties, will deviate from SM predictions. As it will be shown in Section 4.4, in order to derive the top mass one must reconstruct the complete kinematics of the $t\bar{t}$ final state. This is done under several SM dependent assumptions, in particular assuming dominant $t \rightarrow Wb$ decay with width:

$$\Gamma(t \rightarrow Wb) \simeq 170 \text{ MeV} \cdot \left(\frac{M_t}{M_W}\right)^3 . \tag{1.2}$$

The predicted top width depends strongly on the top mass and implies a top lifetime $\tau_t \sim 10^{-24}$ s, an order of magnitude lower than the hadronization time $\tau_{QCD} \sim 10^{-23}$ s. One concludes that the top quark will decay as a free particle, before being bound in a physical hadron with other lighter quarks. This would allow to measure its mass directly, while this is impossible for all other quarks. Anomalies in the reconstructed $t\bar{t}$ states can shed light on the extent to which this unique prediction is valid and possibly give indications on whether the top width is related to its mass precisely as predicted by the SM. Finally, by fully reconstructing the $t\bar{t}$ events one will get detailed information on its production and decay kinematics, such as its polarization, to be compared with SM expectations.

1.3 Status of the Measure of the Top Quark Mass

At the end of RUN I, the top mass was determined using a maximum likelihood procedure(4). There were 91 candidate $t\bar{t}$ events but only 77 with exactly 4 jets selected. Only 22 out of those passed a cut on background probability, with a measured $M_t = 178.0 \pm 4.3$ GeV.

CDF people carried on three independent analyses on the dilepton channel, obtaining consistent results. Unfortunately this channel has few statistics and is underconstrained, which leads to request in a dilepton $t\bar{t}$ candidate event a great amount of missing energy. Imposing some reasonable constraints (such as assuming different values of M_t and computing the probability of the event, scanning the azimuthal distribution of fitted neutrinos or imposing that t and $\bar{t}$ have null total momentum) the result was: $M_t = 168.1^{+11.0}_{-9.8}(\text{stat}) \pm 8.6(\text{sys})$ GeV [VEL].

The main way of measuring the t mass is the template method, where events with more than 4 jets are allowed. At least 4 jets must have $E_T > 8$ GeV and b-tagging is required. Then kinematical constraints are applied, and a kinematic χ^2 is minimized taking M_t as a free parameter. Then templates are built from Montecarlo samples generated with different top mass as input and considering background too: the reconstructed from data mass is compared to the templates

---

(4) Many technical aspects of this Section are explained in Sections 3.1 and 4.4.

results and a likelihood fit is then applied. The top mass measured with 28 candidate b-tagged $t\bar{t}$ events was $M_t = 174.9^{+7.1}_{-7.7}(\mathrm{stat}) \pm 6.5(\mathrm{sys})$ GeV. If two tags are requested, the result is (11 events) $M_t = 180.9^{+6.4}_{-6.0}(\mathrm{stat}) \pm 5.8(\mathrm{sys})$ GeV.

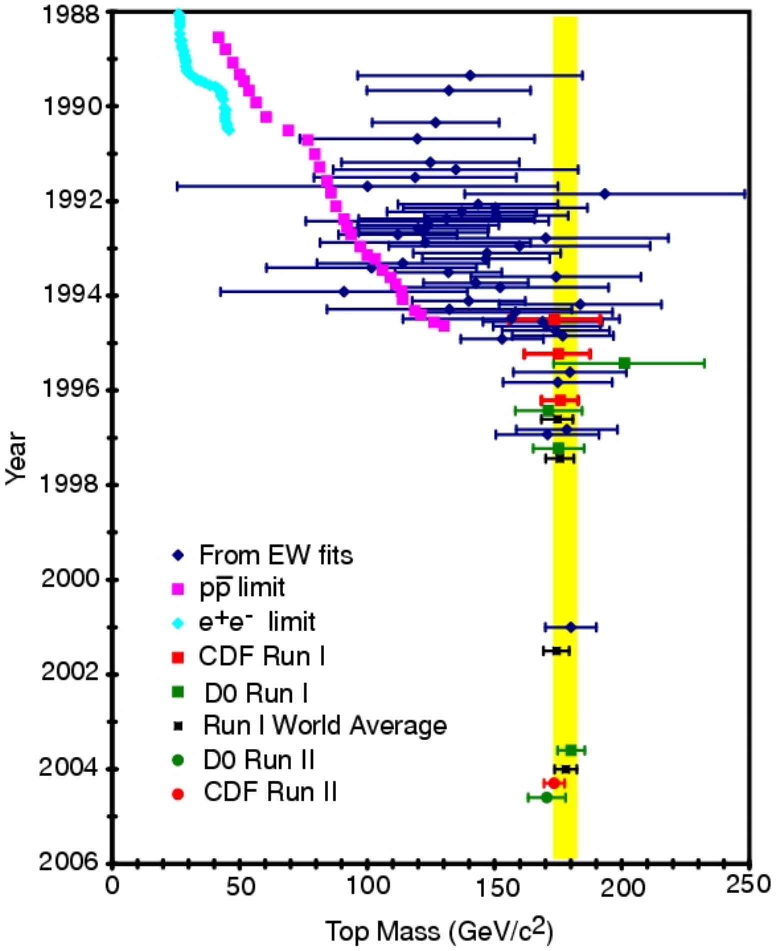


Figure 1.2: Evolution in time of the known value of M_t, from indirect evaluations – before 1994 – to direct measurements [VEL].

Combining all these results of the "standard" template method analysis, one obtains $M_t = 176.7^{+4.9}_{-4.1}(\mathrm{stat}) \pm 6.6(\mathrm{sys})$ GeV. The multivariable method is another template method used at CDF. The candidate $t\bar{t}$ event must be b-tagged, but energy scale is not fixed (in fact it changes according to $W \to q\bar{q}$ reconstruction) and the most probably the χ^2 is associated to a correct parton-to-jet assignment, the heavier the weight of the template is. This way a $M_t = 179.6^{+6.4}_{-6.3}(\mathrm{stat}) \pm 6.6(\mathrm{sys})$ GeV was obtained [VEL].

DØ collaboration studied top mass in the lepton plus jets channel with large p_T and missing E_T, always without any b-tagging operation but with a geometrical discriminant. The result of all DØ studies (191 candidate) gave $M_t = 177.5 \pm 5.8(\mathrm{stat}) \pm 7.1(\mathrm{sys})$ GeV [VEL].

2. The Tevatron Collider and CDF

2.1 The Tevatron

The Tevatron is at present day (year 2005) the highest energy particle accelerator. It is housed in the Fermi National Accelerator Laboratory in Batavia, IL, about 50 km west of Chicago, USA. In this collider, protons and antiprotons are accelerated in opposite directions in a 1 km radius ring and are brought to head-on collision in two points (called B0 and D0) where the CDF and DØ detectors are located, at an energy in the center of momentum system of 1.96 TeV. In a previous run ended in 1996 (RUN I) the collision energy was slightly less, 1.8 TeV.

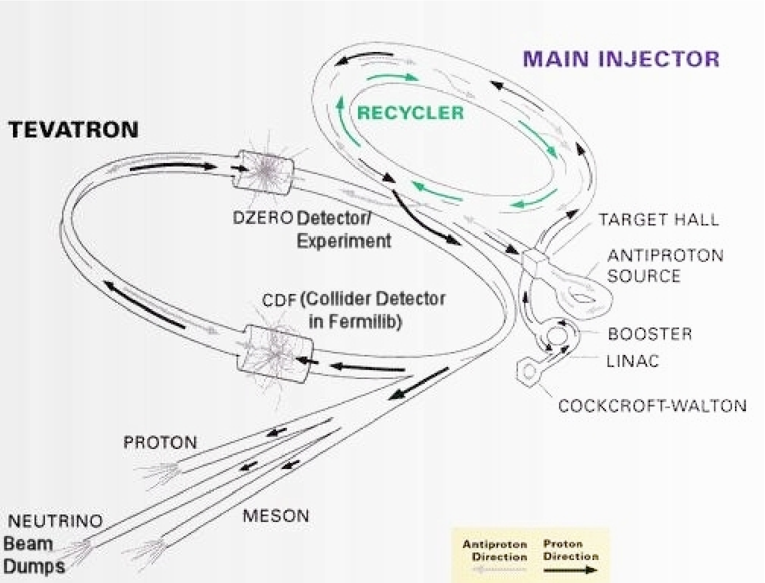


Figure 2.1: Accelerator Chain at Fermilab during Tevatron RUN II.

The current operative phase of the Tevatron is the so-called RUN II, which started in 2001 after a period of upgrades of the accelerator, CDF and DØ detectors. The beams of protons and antiprotons are structured in “trains” of 36 bunches rotating at a frequency of 47.7 kHz and producing at the collision points $5 \cdot 10^6$ bunch-bunch crossings per second with 392 ns separation between crossings. Each bunch contains typically $3 \cdot 10^{11}$ protons or $3 \cdot 10^{10}$ antiprotons. A continuous effort is being made to increase these numbers, in particular the current in the antiproton beam, since they determine the collider luminosity $\mathcal{L}^{(1)}$ and the interaction rate $\mathrm{d}N/\mathrm{d}t$ for any

---

(1) $[\mathcal{L}] = L^{-2} \cdot T^{-1}$.

process of interest. If a process has cross section σ the event rate $\mathrm{d}N/\mathrm{d}t$ is then:

$$\frac{\mathrm{d}N}{\mathrm{d}t}(t) = \sigma \cdot \mathcal{L}(t) \tag{2.1}$$

The collected data samples are proportional to the time-integrated luminosity ($N \propto \int \mathcal{L}(t)\mathrm{d}t$). Data are collected during runs of typical duration of 10 to 20 hours, with maximum instantaneous luminosity at injection and decreasing luminosity with time. At the end of a run the luminosity is reduced typically to $\sim 10\%$ of the initial value. At present (29th April 2005) the record Tevatron instantaneous luminosity has been $\mathcal{L}_{record} \simeq 1.26 \cdot 10^{32}$ cm^{-2} s^{-1} [TEV].

While large currents of protons are easily produced, production of antiproton beams is a very difficult task. The 120 GeV proton beam of the Fermilab Main Injector is extracted and directed to a “source target” where antiprotons are produced, collected at 8 GeV and “cooled” (compacted in phase space) in a system of three rings (debuncher-accumulator-recycler). When enough current is stored and cooled, the antiproton beam is extracted, accelerated to 120 GeV and injected into the Tevatron where it is accelerated together with the opposite proton beam to the collision energy of $\simeq 1$ TeV. A detailed description of this system will not be given in this report. A schematic layout of the proton and antiproton sources and of the beam transport system to the Tevatron is shown in Figure 2.1.

2.2 The CDF RUN II Detector

CDF is a general purpose full-coverage particle detector surrounding the Tevatron vacuum pipe at the B0 collision point. It is composed by an approximately cylindrical multi-layered silicon tracker just outside the pipe, a cylindrical drift chamber (Central Outer Tracker) at a radial distance from 30 to 150 cm inside a solenoid magnet providing a longitudinal field $B \simeq 1.4$ T. Around the COT and just inside the solenoid coil is located a layer of time-of-flight plastic scintillation counters. Electromagnetic and hadronic calorimeter structured in projective towers are arranged outside the magnet coil and backed by muon detectors. A 3D section of CDF is drawn in Figure 2.6.

Silicon Tracking System

The CDFII silicon tracking system consists of three sub-units: an innermost silicon detector single layer glancing the Tevatron pipe, called Layer 00, a five-layered approximately cylindrical Silicon Vertex Detector (SVX II), and a double-layered Intermediate Silicon Layers (ISL). Figure 2.2 shows the cross section of the full silicon tracker. SVX II is structured in three longitudinal barrels of total length 96 cm, covering tracking range of $|\eta| < 2$(2). As only the $|\eta| < 1$ region is fully covered by the COT, tracking at $1 < |\eta| < 2$ is obtained by the double-layered ISL. The innermost layer L00 has radius ~ 1.5 cm and the outermost ISL layer is 28 cm far from the beampipe. The precision of position measurements of this apparatus is excellent: 12 μm for SVX II and 16 μm for ISL (both values are given for measurements on the axial direction), allowing all together an

---

(2) η is the so-called pseudorapidity, defined as $-\ln(\tan(\theta/2))$, where θ is the polar angle with the z-axis collinear with the beampipe.

impact parameter resolution for energetic tracks of $\simeq 40\ \mu$m [TDR].

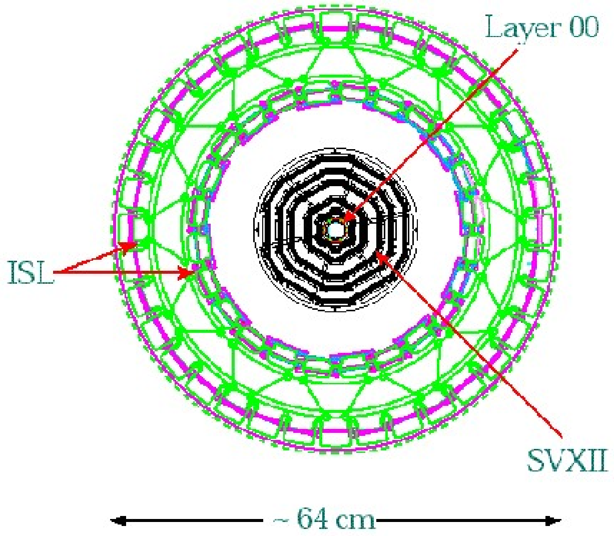


Figure 2.2: Silicon Tracker cross section.

Central Outer Tracker

The COT is a cylindrical open cell drift chamber covering the pseudorapidity range $|\eta| < 1$. It provides 96 points between 44 and 132 cm with a resolution of $\sim 140\ \mu$m in the $\hat{\phi}$ direction (normal to the radial direction). The total number of wire layers is 96, grouped 12 by 12 in 8 superlayers. The electrical field in the drift cells is inclined with respect to the $\hat{\phi}$ direction in order to allow the ionization electrons, produced in the gas mixture (Ar-Ethane-CF_4), to drift in this direction under the combined effect of the electrical field and of the solenoid magnetic field.

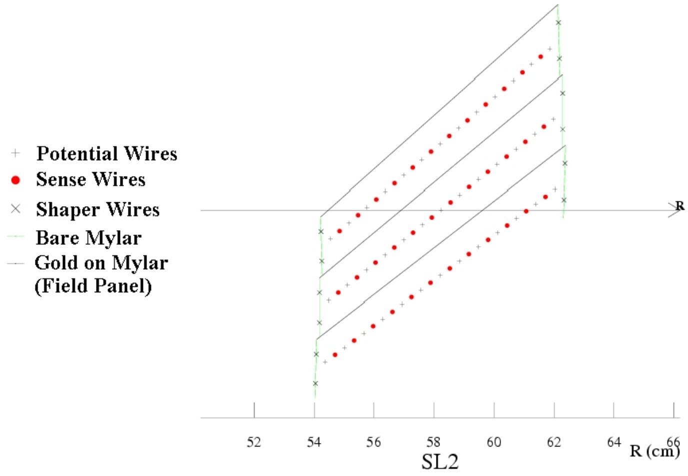


Figure 2.3: Wires layout in a COT Superlayer 2 cell.

A Time of Flight Detector (TOF) is placed in the few centimeter clearance just outside of the COT and inside the magnet coil. It is a single layer of plastic scintillator bars viewed by photomultipliers at both ends. The time resolution is $\sim$ 120 ps [TDR].

Calorimetry

The tracking system (Silicon plus COT) and the solenoid providing the magnetic field are surrounded by the central calorimeter, and closed at the ends by two end-plug calorimeters. The magnetic flux is returned through the end-plugs and through two poles behind the central calorimeters, in such a way that this calorimeter is outside the magnetic circuit. At all angles the calorimetry is split into two longitudinal compartments, a front Pb-scintillator sandwich electromagnetic compartment and a rear Fe-scintillator hadronic compartment. The central calorimeter is divided into 24 azimuthal wedges containing 9 projective towers each, which cover 0.11 in $|\eta|$. The end plug calorimeters are also structured in projective towers in a similar way [TDR].

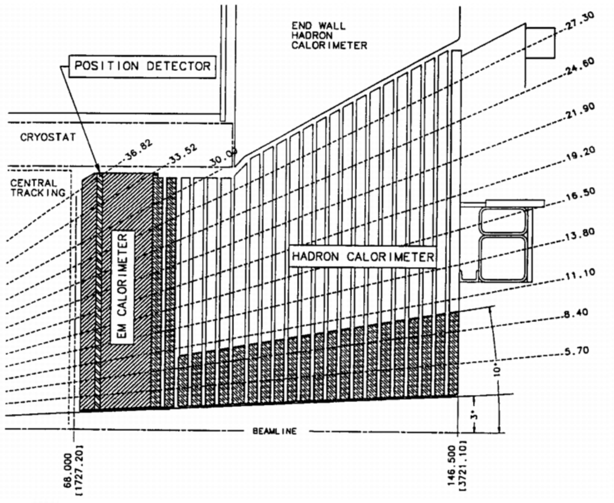


Figure 2.4: End Plug Calorimeter.

Muon Detectors

Large transverse momentum muons are as important as electrons in top studies. The CDF calorimeters are surrounded by muon tracking chambers and trigger scintillation counters up to $|\eta| < 2.4$. These are large systems composed of several structures covering the outer detector as well as possible [TDR].

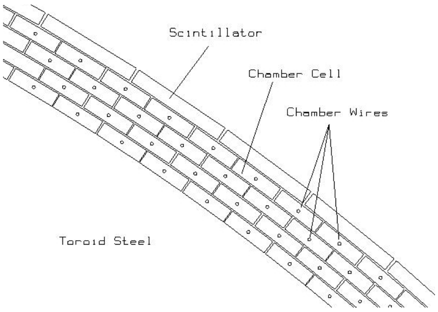


Figure 2.5: Structure of a Muon Detector.

CDF Trigger

At the Tevatron collider the top pair production cross section is $\sim$ 6 pb [TXS], to be compared to the total inelastic cross section of $\sim$ 60 mb [PDG]. An efficient and selective trigger is therefore vital in searching for top events among a 10^{10} larger rate of mostly uninteresting events. CFD performs studies of a large variety of physics besides top, by employing over 50 different trigger channels tuned to select processes of specific interest (multi-jet events, beauty hadron production . . .). Each trigger chain is allowed to run at a maximum rate (its allocated “budget”), in such a way that the overall dead time of the data acquisition system is small and the most interesting channels are collected with nearly full efficiency. The RUN II DAQ electronics can collect events at $\simeq$ 75 Hz, with a dead time limited to $\sim$ 5% [TRI], [TDR].

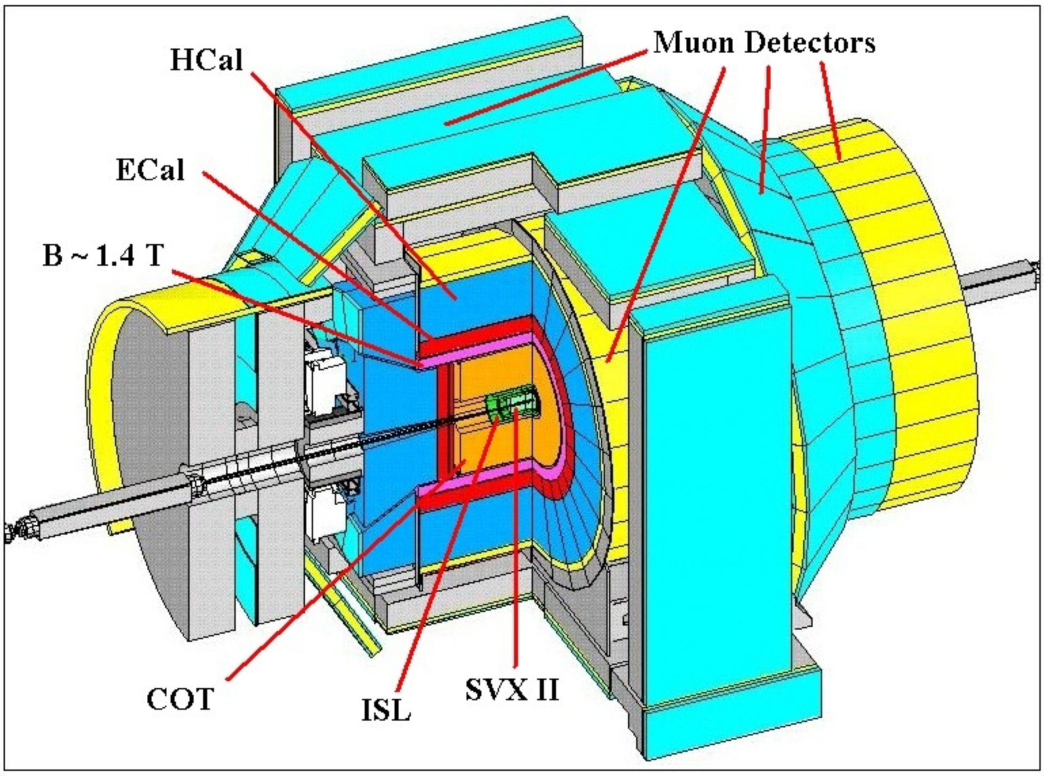


Figure 2.6: Scheme of CDF RUN II Detector.

3. Features of $t\bar{t}$ Events

3.1 The Lepton plus Jets Channel

In a $p\bar{p}$ collision, a primary $q\bar{q}$ pair can annihilate ($\sim$ 85% of cases) or a gg pair can fuse ($\sim$ 15% of cases) producing a $t\bar{t}$ pair. As each top quark decays as $t \to Wb$, the $t\bar{t}$ pair will end into three different final states depending on the decay modes of the two W:

- the dilepton channel, when both W decay leptonically ($W \to e\nu$ or $W \to \mu\nu$)$^{(1)}$;
- the lepton plus jets channel, when one W decays leptonically;
- the full-hadronic channel, when both W decay hadronically into two quarks.

The following study addressed the lepton plus jets channel as a compromise between a fair branching ratio ($\sim$ 30%) and limited background. The decay chain leading to a final state of this kind is:

$$\begin{aligned}
&p\bar{p} \to t\bar{t} + \text{“underlying event”} \\
&t \to bW^{+} \to b + q\bar{q} \to \text{jet}_b + \text{jet}_q + \text{jet}_{\bar{q}} \\
&\bar{t} \to \bar{b}W^{-} \to \text{jet}_{\bar{b}} + \ell\bar{\nu}_\ell
\end{aligned} \tag{3.1}$$

where q and $\bar{q}$ are $u\bar{d}$ or $c\bar{s}$ pairs and ℓ is an electron or a muon (the roles of t and $\bar{t}$ can be exchanged). In the dilepton channel both W decay in $\ell+\nu$ so that the system of equations to solve to reconstruct the event is underconstrained, and needs fitting the directions of both neutrinos that escape from the detector. In the lepton plus jets channel there is only one not measured final state vector and the event kinematics can be fully reconstructed – see Section 4.4 [VEL].

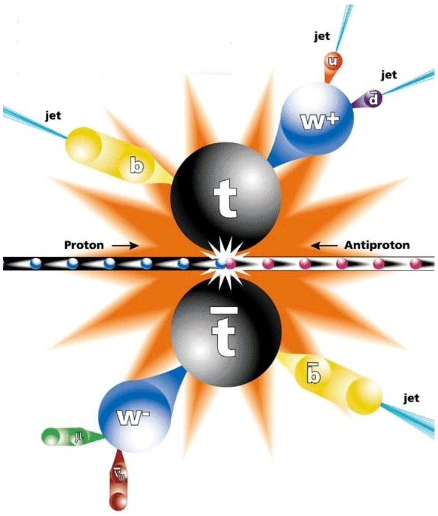


Figure 3.1: The lepton plus jets channel.

3.2 Jets in $t\bar{t}$ Events

The final state quarks in top decays manifest themselves as “jets” of hadrons, which are narrow bunches of energetic particles produced by the “fragmentation” of a quark around its direction of flight. They are identified and reconstructed from information provided by calorimeters, so that another way to think about jets is, in fact, to refer to them as fluxes of energy in a precise region of space.

---

(1) τ leptons are difficult to detect and are not considered in this work.

Reconstruction of Jets

Several algorithms have been developed at CDF to reconstruct jets: the main one, which gives better resolution in clustering, and which has been adopted for this study, is named `JetClu`. In `JetClu` the jet is defined as a flare of energy within a cone in the $\eta-\phi$ space whose size is defined by its radius as

$$R=\sqrt{\Delta\eta^2+\Delta\phi^2}. \tag{3.2}$$

Cone sizes currently used at CDF are $R \in \{0.4,\ 0.7,\ 1.0\}$, depending on the type of analysis. In multijet events like top candidates the default radius is $R=0.4$.

In order to find jets in an event, calorimeter towers which returned a transverse energy $E_T > 1$ GeV(2) are selected and classified as possible seeds of a jet. Towers above some E_T threshold (0.1 GeV by default) around a seed are grouped in preclusters, except for towers outside a 7×7 rectangle at whose center the seed is. Towers outside a precluster are used to build a new one. The centroid of a precluster is found via a weighted mean over E_T, and the cone is then defined around that centroid. Then a new centroid of the cluster is computed taking into account only towers included in the cluster, and the algorithm iterates until the list of tower assignments remains unchanged. Jet parameters are recomputed after all towers are uniquely assigned(3) [JET].

Tagging of *b* Jets

In the lepton plus jets channel, as in (3.1), two out of four jets are the result of the fragmentation of a b and a $\bar{b}$. It is very important to identify these jets and distinguish them from light quark and gluon jets since b jets are relatively less likely in background events. Moreover, since identified b jets are not considered as possible prongs from W decay the number of jet-to-parton combinations(4) to be analyzed in the event reconstruction is reduced (in fact, the misidentification probability is in average $\simeq 0.7\%$ [TXS], [TAG]). The identification of b jets is called b-tagging. The two b-tagging algorithms used at CDF in the top analysis are the Soft Lepton Tagger (SLT) and the Secondary Vertex Tagger (SECVTX). SLT looks for e or μ leptons of relatively large p_T ($p_T > 2$ GeV) within the jet cone as they are likely to be produced in the decay of B-hadrons in the jet. SECVTX looks for vertices with at least 2 tracks originating in a point, called “secondary vertex”, whose distance from the primary one is much larger than the uncertainty in the position reconstruction, as in Figure 3.2.

---

(2) $E_T = \sum E_i \sin\theta_i$, where θ_i is the angle between the beampipe axis and the line linking the nominal interaction point to the centroid of a calorimeter tower.

(3) One tower should not be assigned to two or more jets.

(4) See Section 4.4.

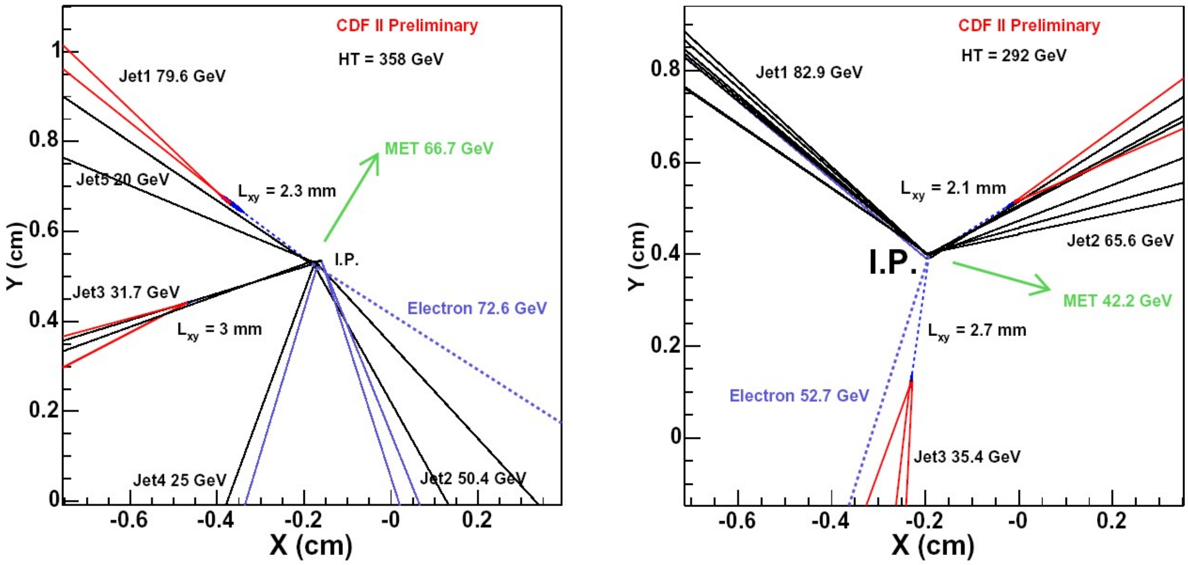


Figure 3.2: Example of secondary vertices found by SECVTX algorithm: [L] run 167551, event 7969376; [R] run155145, event 132579. IP is the interaction point (the primary vertex), MET is the missing E_T (see Section 4.4) and b-tagged jets are drawn in red [YOR].

4. Study of the Performances of Cleaning

4.1 Goal of this Work

The uncertainty in the top mass can be reduced if a sample with a better signal to background ratio is selected with a limited loss in statistics. An example of such a method is to require that one jet in the event should be *b*-tagged. For this report, the candidate was given the task of looking for an alternative method of improving the top fraction by using a particular kinematical selection based on jet shapes⁽¹⁾. This method, named “event cleaning”, will be presented and its preliminary performance will be reported below.

4.2 Description of the Analysis Strategy

The study was limited to the lepton plus jets channel, testing on simulated events if a “cleaning process” of events, if added to selections already used at CDF, can improve analysis performances. These improvements can be seen in three ways: a decreasing of systematic errors, a better consistence with the “real mass” of the reconstructed one and a better interpretation of jets in selected events.

As a start, from Montecarlo simulations one determines the cuts to select events which have the requested $t\bar{t}$ topology. By making use of some additional information when available, such as if some jet is *b*-tagged, the events are reconstructed in the $t\bar{t}$ hypothesis. Events in the lepton plus jets sample are selected from the large p_T electron and muon trigger streams. The most important analysis cuts for top mass studies are:

1. Lepton isolation, which means that, considering a cone with $R = 0.4$ in the $\eta - \phi$ space centered on the lepton (η, ϕ), the total momentum of other charged tracks inside the cone should be less than 5% of the lepton momentum;

2. Lepton $p_T > 20$ GeV;

3. $\not{E}_T > 20$ GeV⁽²⁾;

4. Number of jets ≥ 4, as the study is addressed to the lepton plus jets channel;

5. Jet $E_T \geq 15$ GeV;

---

(1) Similar tests were carried on for RUN I by Stefano Bettelli in 1996 [BET].

(2) The quantity $\not{E}_T$ is the so-called “missing E_T”, defined as $\not{E}_T = -\sum_i p_{Ti}$ (from the conservation of total momentum, which is null).

6. Jet centroid should have $|\eta| < 2$, which is the coverage of ISL.

As this work is not a measurement of M_t but only a preliminary study of cleaning performances, only cuts number 4 and 5 were applied according as described in Section 4.5.

The events are reconstructed in the $t\bar{t}$ hypothesis by assuming that the four more energetic (in E_T) jets correspond to the four top decay partons⁽³⁾. From energy-momentum conservation, one can write down 20 scalar equations with 18 unknowns once the W mass is assumed to be known and the mass of the two final state t and $\bar{t}$ are assumed to be equal:

$$\begin{array}{rcl} p_t^\mu & = & p_{W^+}^\mu + p_b^\mu \\ p_{\bar{t}}^\mu & = & p_{W^-}^\mu + p_{\bar{b}}^\mu \\ p_{W^\pm}^\mu & = & p_{\ell^\pm}^\mu + p_\nu^\mu \\ p_{W^\mp}^\mu & = & p_q^\mu + p_{\bar{q}}^\mu \\ (\sqrt{s},0,0,0) & = & p_t^\mu + p_{\bar{t}}^\mu + p_X^\mu \\ M_t & = & M_{\bar{t}} \\ M_{W^\pm} & = & 80.4\ \mathrm{GeV} \end{array} \tag{4.1}$$

After defining the event as $p\bar{p} \to t\bar{t} + X$, where X is the additional system ("underlying event") to the six $t\bar{t}$ decay vectors, and making use of the constraints on masses, these equations reduce to six effective scalar equations:

$$\begin{array}{rcl} p_{Tt} + p_{T\bar{t}} + p_{TX} & = & 0 \\ M_{W^\pm}^2 & = & p^\mu p_\mu(\ell + \nu) \\ M_{W^\mp}^2 & = & p^\mu p_\mu(\mathrm{jet}_q + \mathrm{jet}_{\bar{q}}) \\ M_t^2 & = & p^\mu p_\mu(W^+ + \mathrm{jet}_b) \\ M_{\bar{t}}^2 & = & p^\mu p_\mu(W^- + \mathrm{jet}_{\bar{b}}) \\ p_{T\nu} & = & \not{E}_T \end{array} \tag{4.2}$$

Since the not measured quantities are four (the top mass and the three components of the neutrino momentum), the system is over-constrained and can be solved by a two-dimensional χ^2 minimization as a function of the top mass and of the longitudinal momentum of the neutrino (with its sign).

In a lepton plus jets $t\bar{t}$ event, there are at least 4 jets, which should be assigned, at a calorimeter level, to their "parent" partons (called primary partons) according to 12 different combinations⁽⁴⁾. All possible assignments of the 4 jets to the 4 partons are tried. If one jet is b-tagged, it is assumed to be a b-jet and the combinations are reduced to 6. If two jets are b-tagged, the combinations are 2. The fitted neutrino p_z can assume two values, according to the previous paragraph, so that the total number of combinations is 24 (12 or 4 if the event is b-tagged). The reconstruction of M_t is done starting from the conservation of energy, according to whom the following χ^2 (4.3)⁽⁵⁾ can

---

(3) Montecarlo studies indicate that, while being the best assumption, it is correct only in $\sim 50\%$ of cases [TCS].

(4) The total number of combinations is $4! = 24$, but to reconstruct M_t the jets deriving from light quarks can be exchanged without differences in the result $\Longrightarrow 4!/2 = 12$.

(5) Capital letters refer to fitted or already known values, lower-case to reconstructed ones; X is the non clustered energy ("underlying event"); σ are uncertainties on measured energies, while Γ are Breit-Wigner widths of particles ($\Gamma_W \approx 2$ GeV is measured [PDG], but $\Gamma_t \approx 1.5$ GeV is assumed by theory [WID] – See Section 1.2).

be assigned to each t mass reconstructed in a particular combination:

$$\chi^2 = \left(\frac{X-x}{\sigma_x}\right)^2 + \Sigma_{\ell,\mathrm{jets}}\left(\frac{E_T-e_T}{\sigma_{e_T}}\right)^2 + \left(\frac{M_W-m_{jj}}{\Gamma_W}\right)^2 + \left(\frac{M_W-m_{\ell\nu}}{\Gamma_W}\right)^2 + \left(\frac{M_t-m_{jjj}}{\Gamma_t}\right)^2 + \left(\frac{M_t-m_{\ell\nu j}}{\Gamma_t}\right)^2 \tag{4.3}$$

Each fit returns one value for the top mass and for the square of the longitudinal neutrino momentum. Only solutions with an acceptable χ^2 are retained(6), and the mass returned by the solution with the lowest χ^2 is taken as the event mass and entered into a mass distribution. The experimental top mass is obtained by a two component maximum likelihood fit to this distribution in terms of simulated signal and background "templates". The signal template depends on the assumed top mass and the one which fits the distribution best gives the measured top mass, while the associated width of the χ^2 distribution gives the measurement error.

As stated before, the aim of this work is to see which improvements a "cleaning process" can give. This kind of selection starts with ordering all jets in decreasing E_T: the first 4 are called "leading jets". This is done for jets with $R = 0.4$ and $R = 0.7$. Then each 0.4-jet is matched to the closest (in the $\eta - \phi$ space with a distance defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$) 0.7-jet; if the leading jets are matched one to each other, respecting the order (the first in the 0.4 list is matched to the first in the 0.7 list, and so on ...) the event is called clean. The aim of cleaning is to select events with well-separate jets and with the lowest number of interfering particles. It can be seen as a test of the "goodness" of the reconstruction of jets too. As leading jets are supposed to be those in which primary partons fragment, the b-tagging takes the form of a request of at least one tagged jet, that must be one of leading ones (b-tagging is referred to 0.4 jets).

4.3 Why to Clean Events

A major source of errors in the top mass obtained by this process is due to the wrong jet-to-parton assignments that are chosen as best solutions. These errors are born both because a gluon radiation jet may enter among the 4 leading jets, and because a wrong combination of the right jets may be chosen by the best χ^2 criterion. By b-tagging the jets and forcing the assignment of a b-tagged jet to a b-parton in the fit one reduces the second source of errors, however, it would be important to reduce also the first one. This is the attempt carried out with this work. We have studied on simulations whether by requesting the jets in the event to be well separated and uniquely defined (independent of the adopted cone size) would make it more likely that the four leading jets are those generated by the four final state partons of the $t\bar{t}$ system. This "cleaning" is made before assigning an order of jet-to-parton assignment and therefore can be applied to the b-tagged events as well. However, it would be particularly important for 0-tags events where the combination errors are maximal, and the study was tuned to this sample. Improvements obtained by cleaning would be gauged eventually by reduced systematic errors and by smaller spread of the values of the reconstructed mass. However, a better selection of the 4 jets attributed to top decays would be a first step towards that goal.

---

(6) The acceptable χ^2 range is loosely defined. Simulations indicate that a cut at $\chi^2 < 10$, which was used for this work, accepts nearly 100% of the correct combinations [BRU], [ZT2].

4.4 Study of HERWIG $t\bar{t}$ Events ($M_t = 175$ GeV)

To simulate $t\bar{t}$ events the HERWIG generator is commonly used at CDF (Hadron Emission Reactions With Interfering Gluons [HER]). A sample of 10955 events with input $M_t = 175$ GeV was generated, on which the effects of cleaning were tested. All tests were performed on three different samples: the pretagged sample, the tagged sample and the 0-tags sample.

- The pretagged sample contains all events with at least 4 jets per description(7) (because we are dealing with HERWIG lepton plus jets events) reconstructed by `JetClu` algorithm: these jets must have $E_T \geq 15$ GeV, which is a "standard" selection used at CDF, from MC studies.
- The tagged sample contains those pretagged candidates where at least one leading jet is b-tagged by the SECVTX algorithm in a $R = 0.4$ cone. As the tag probability in gluon or light quark jets is very small ($< 1\%$ per jet [TAG], [TXS]), in the event reconstruction the b-tagged jet is assigned to a b quark by default. In this case, the mass of the event can be not the lowest-χ^2 one, or can have a $\chi^2 \geq 10$ (in this case the event is rejected). The test on the tagged sample will be treated in Appendix A.
- The 0-tags sample is the complement of the tagged sample: it is composed by pretagged events which show no b-tags at all, and masses are selected as in the pretagged case(8).

4.5 Description of Tests

The tests were carried on the main sample and on sub-samples obtained after applying a number of additional selection cuts, each one stricter than the previous one. The basic selection (labeled BC, and understood if label is omitted) keeps all HERWIG events which allow at least one reconstruction with $\chi^2 < 10$ within those featuring at least 4 jets (with $E_T \geq 15$ GeV) per description. The BC selection is found to be 84.4% efficient and is common to all sub-samples. All efficiencies will be quoted relative to the BC sample in the following.

The 4 "semi-exclusive" jets selection (4sXJ) adds to BC the request of having 4 and only 4 jets in the 0.4 description with $E_T \geq 15$ GeV. The 4 exclusive jets selection (4XJ) adds to 4sXJ the request of having no jets in the $(8, 15)$ GeV E_T region. These additional selections are chosen such as to help cleaning which would follow. However, such cuts will cause a loss of statistics: one must account for this loss for a fair evaluation of the effects of cleaning. A parallel selection can be applied while studying MC events: the "4 partons matched" selection (M4q). This is a topological request on the decay tree of the event: all the primary partons must have a direction in the $\eta - \phi$ space compatible with the direction of a leading jet; in a figure-like explanation, the parton must be "inside" the jet, so that the request to have a match is then $\Delta R < 0.4$, and each parton must be matched to a different leading jet (0.4 jets were used). Both 4sXJ and 4XJ selections hope to reduce the number of events with radiating gluons. The M4q selection is a way to

---

(7) In this work it can be either `JetClu` $R = 0.4$ or `JetClu` $R = 0.7$.

(8) There are events with more than 4 large E_T jets with a b-tag in a non-leading jet. This sample deserves attention and could be considered in future studies.

test if 4sXJ and 4XJ achieve their aim.

The last kind of selection is the “fitter selection” (F), which keeps only events where the jet configuration (which is the result of parton-to-jet assignment) that gives the selected mass is the same reconstructed “by hand” looking at the decay tree and assigning the jet to the closest parton: this procedure of matching is the same of the M4q. The difference between the two selections is that F requests the partons to be matched in the correct order, while M4q does not care about the parton-to-jet assignment configuration. All the templates of these selections were compared to the one restricted by cleaning (C) to see the effects of cleaning itself.

Three kinds of parameters have been used to evaluate the effects of cleaning on simulated top events:

1. efficiency of the cuts defining the sample to be cleaned;
2. frequency of jets matching to the top primary partons, before and after cleaning;
3. width of the top mass template before and after cleaning.

The **absolute efficiency** (Ef) is the ratio between the number of events that pass the particular selection and the number of BC pretagged events; the **relative efficiency** (rEf) is the ratio between the number of events that pass the particular selection and the number of pretagged events in the specific selection (BC, 4sXJ, 4XJ without applying C): this parameter is basic to evaluate effects of cleaning, particularly if compared with *b*-tagging, which is the main tool in top mass studies. The **good match fraction** (*GM*) is the ratio between the number of events in the particular selection while applying M4q, in which the four leading jets match the primary partons, and the number of pretagged BC events; the **relative good match fraction** (*rGM*) is the fraction of events passing the M4q cut after an additional selection is made. The **fitter efficiency** (fEf) is the rate of events where the parton-to-jet assignment is correct (or better, the one corresponding to the selected mass is the same reconstructed “by hand”). The most significant information in the reconstructed top mass distributions is its RMS: the narrower the distribution, the better the measurement is. It is also desirable that the mean of the distribution be close to the input mass, although this is not essential as long as the correlation of the output to the input mass is linear and known [ZT1], [BRU].

4.6 Pretagged Sample

In the pretagged sample the good match fraction (*GM*) is 58.0%. The number of F events corresponds to a fitter efficiency (fEf) of 31.8%. The mass distributions in the full sample and in its correctly fitted events are shown in Figure 4.1. The width of the distribution is reduced appreciably by F, from 34.8 to 28.9 GeV. The average mass is shifted by less than 1 GeV.

If cleaning C is applied, the good match rate *GM* is increased from 58.0% up to 63.6%. However, the C efficiency is 42.6%. The correct ordering rate fEf found by the fit increases from 31.8% to 35.9%, with a C efficiency of 48.0%. The mass distributions for the full sample and for F events after cleaning are shown in Figure 4.2. One observes an improvement in the width from

28.9 GeV to 26.5 GeV in the F sample, but no improvement in the full sample.

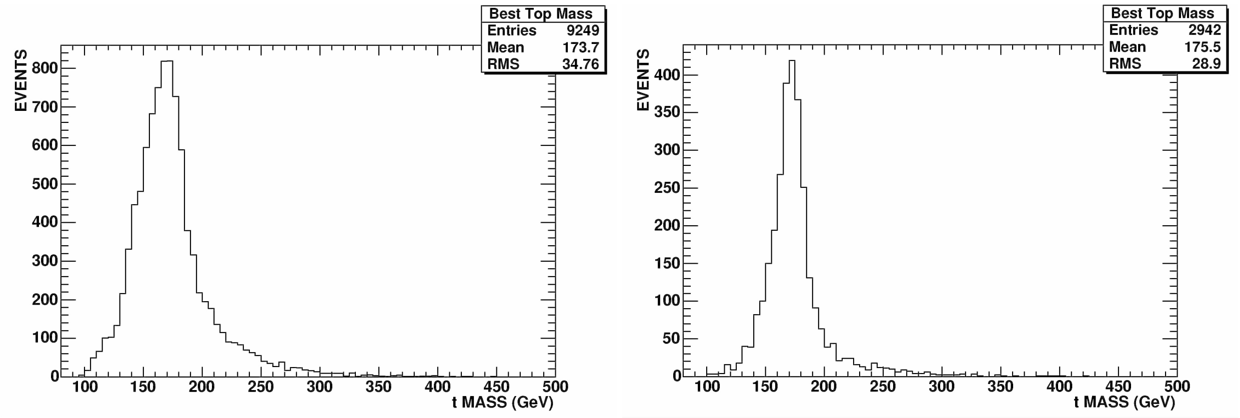


Figure 4.1: [L] t mass distribution relative to pretagged BC events; [R] t mass distribution relative to pretagged BC+F events.

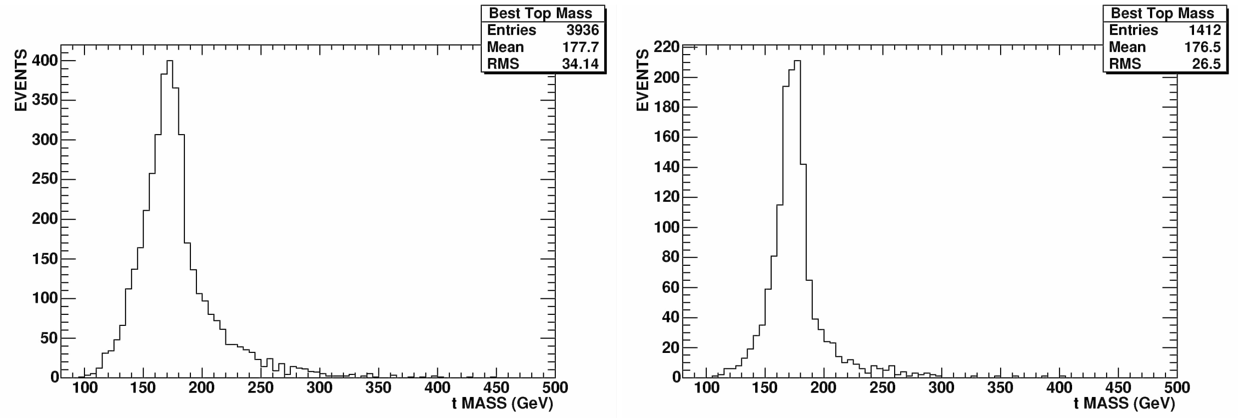


Figure 4.2: [L] t mass distribution relative to pretagged BC events; [R] t mass distribution relative to pretagged BC+F events. In both cases, cleaning is applied.

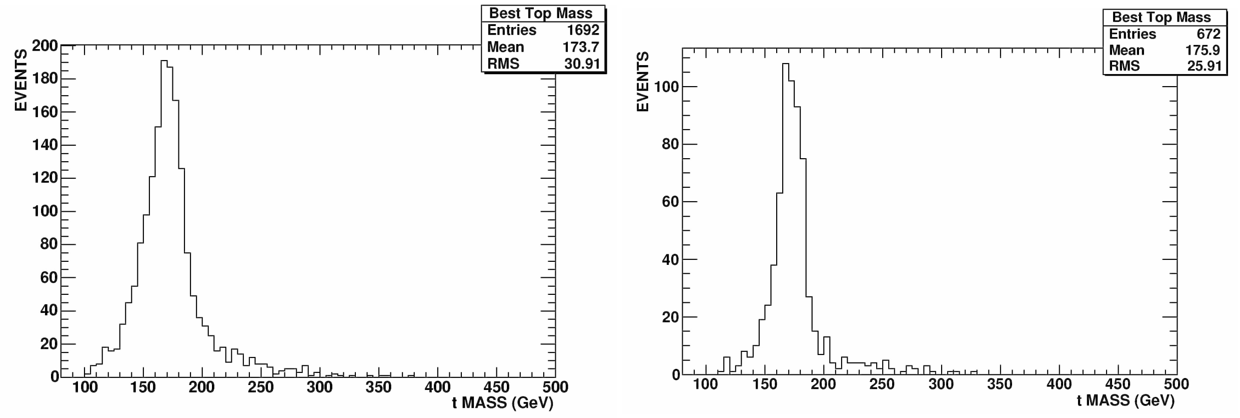


Figure 4.3: [L] t mass distribution relative to pretagged 4XJ events; [R] t mass distribution relative to pretagged 4XJ+F events.

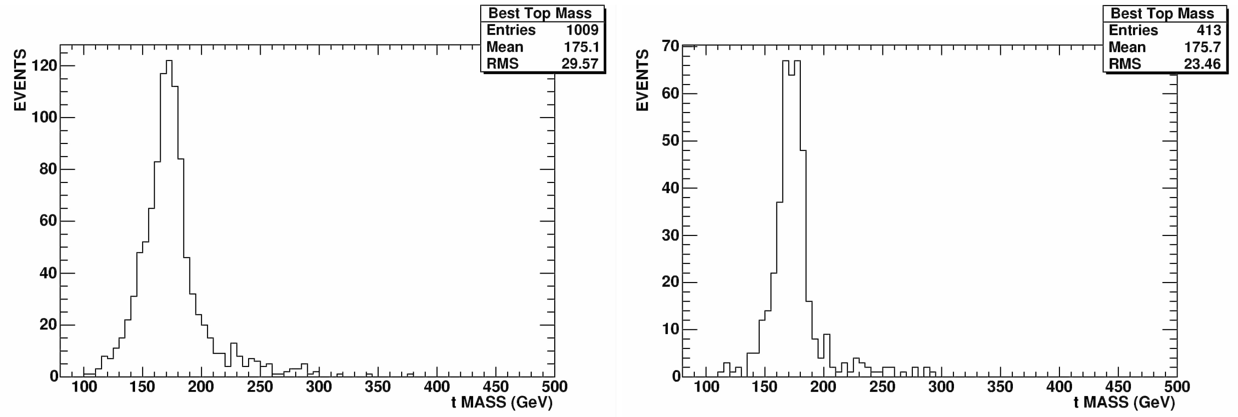


Figure 4.4: [L] t mass distribution relative to pretagged 4XJ events; [R] t mass distribution relative to pretagged 4XJ+F events. In both cases cleaning is applied.

pretagged								
set	# events	Ef	rEf	GM	rGM	fEf	M_t (GeV)	RMS (GeV)
BC	9249	1.000	1.000				173.7	34.8
BC+C	3936	0.426	0.426				177.7	34.1
BC+F	2942	0.318	1.000			0.318	175.5	28.9
BC+C+F	1412	0.153	0.480			0.359	176.5	26.5
M4q	5364	0.580	1.000	0.580			171.8	29.7
M4q+C	2503	0.271	0.467	0.636			173.8	28.1
M4q+F	2480	0.268	1.000	0.843		0.462	175.4	27.6
M4q+C+F	1232	0.133	0.497	0.873		0.492	175.9	25.5
4sXJ	3077	0.333	1.000				172.8	33.2
4sXJ+C	1597	0.173	0.519				176.3	33.4
4sXJ+M4q	2035	0.220	1.000	0.220	0.661		172.1	29.3
4sXJ+M4q+C	1134	0.123	0.557	0.288	0.710		173.5	28.2
4sXJ+F	1053	0.114	1.000			0.342	174.9	28.0
4sXJ+C+F	599	0.065	0.569			0.375	176.1	25.9
4XJ	1692	0.183	1.000				173.7	30.9
4XJ+C	1009	0.109	0.596				175.1	29.6
4XJ+M4q	1300	0.141	1.000	0.141	0.768		173.1	27.4
4XJ+M4q+C	802	0.087	0.617	0.204	0.795		173.1	25.5
4XJ+F	672	0.073	1.000			0.397	175.9	25.9
4XJ+C+F	413	0.045	0.615			0.409	175.7	23.4

Table 4.1: Effects of cleaning in the pretagged sample of a 10955 HERWIG $t\bar{t}$ (175 GeV) lepton plus jets events.

Since the F sample cannot be obtained by a cut on the data, the hope for progress must be moved to the 4sXJ and to the 4XJ samples. For sake of brevity, only 4XJ results will be explained. The mass distributions for 4XJ events and for 4XJ+F events are shown in Figure 4.3 The same samples after cleaning are shown in Figure 4.4. The 4XJ selection applied to the pretagged sample has an efficiency (Ef) of 18.3% (1692 events out of 9249). The C selection on the 4XJ sample has an efficiency (*rEf*) of 59.6% (1009 out of 1692). In the 4XJ sample cleaning gives a just appreciable improvement in resolution, since the distribution width changes from 30.9 GeV to 29.6 GeV – see Section 4.7. The effect is more significant in the matched F sample, since the width decreases from 25.9 GeV to 23.4 GeV after cleaning. The improvement in matching is now given by *rGM*, whose change is from 76.8% to 79.5%, while the efficiency of the fitter *fEf* changes from 39.7% to 40.9%. All results (including those of 4sXJ events) are summed up in Table 4.1.

4.7 0-tags Sample

The 0-tags sample contains 3494 events, the 38% of the total. This is an important fraction which justifies the effort to make good use of it. The mass distributions for this sample and for its matched F events are shown in Figure 4.5. Figure 4.6 shows the same distributions after cleaning. The efficiency of cleaning is 42.7% (1492 events out of 3494), and the resolution improves minimally from 35.8 to 34.8 GeV. Simulation shows that cleaning performs better in the matched F event sample, where the width improves from 28.4 GeV to 24.9. All together no significant improvement is predicted in this sample. There are improvements in *fEf* (which increases from 31.5% to 36.7%) and *GM* (from 54.1% 61.5%). The next step should be to study the exclusive 4XJ events in the 0-tags sample.

Within the 3494 events of the 0-tags sample there are 618 4XJevents (17.6%). The mass distributions for 4XJ and for 4XJ+F events are shown in Figure 4.7. The same distributions after cleaning are shown in Figure 4.8. Cleaning improves the mass resolution of the 4XJ sample from 32.6 GeV to 31.1 GeV. We have not studied the uncertainties on the mean and the RMS of the distribution in detail, but they are of the order of 0.5 GeV and this small improvement is significant. The predicted improvement in the F sample is very similar, from 26.9 GeV to 25.3 GeV. *fEf* increases from 43.9% to 46.3% and *rGM* from 77.3% 81.2%.

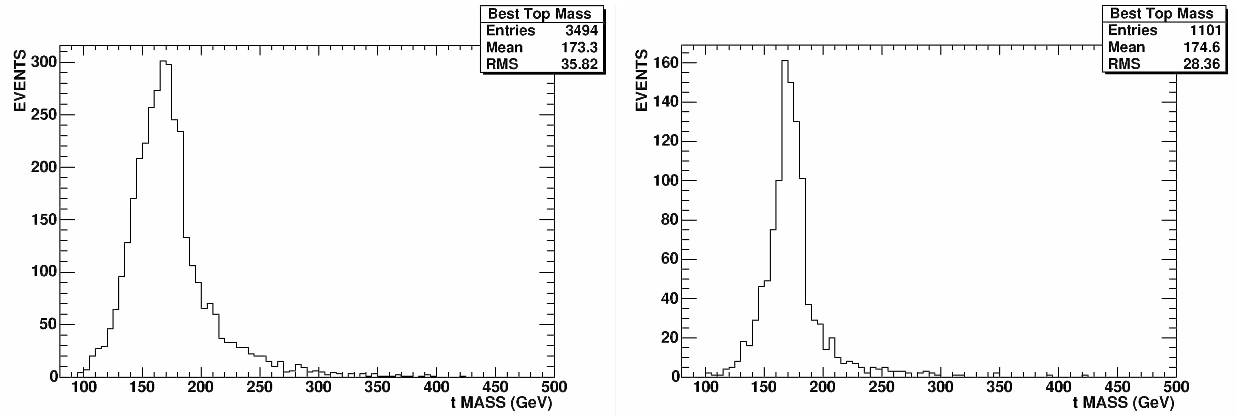


Figure 4.5: [L] t mass distribution relative to 0-tags BC events; [R] t mass distribution relative to 0-tags BC+F events.

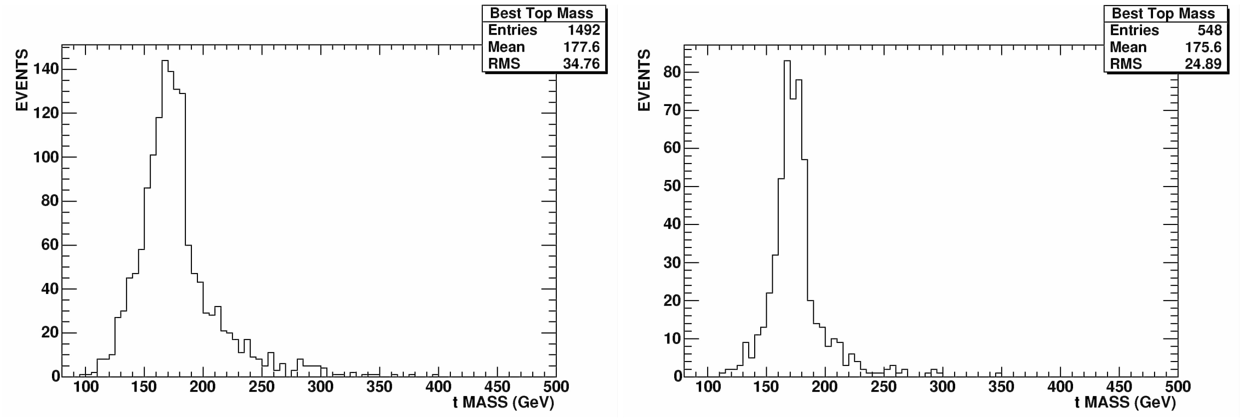


Figure 4.6: [L] t mass distribution relative to 0-tags BC events; [R] t mass distribution relative to 0-tags BC+F events. In both cases cleaning is applied.

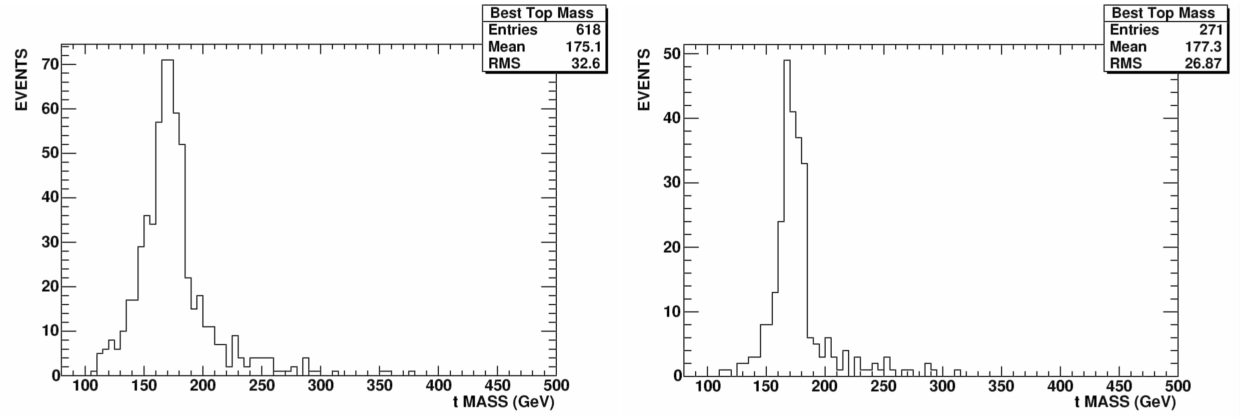


Figure 4.7: [L] t mass distribution relative to 0-tags 4XJ events; [R] t mass distribution relative to 0-tags 4XJ+F events.

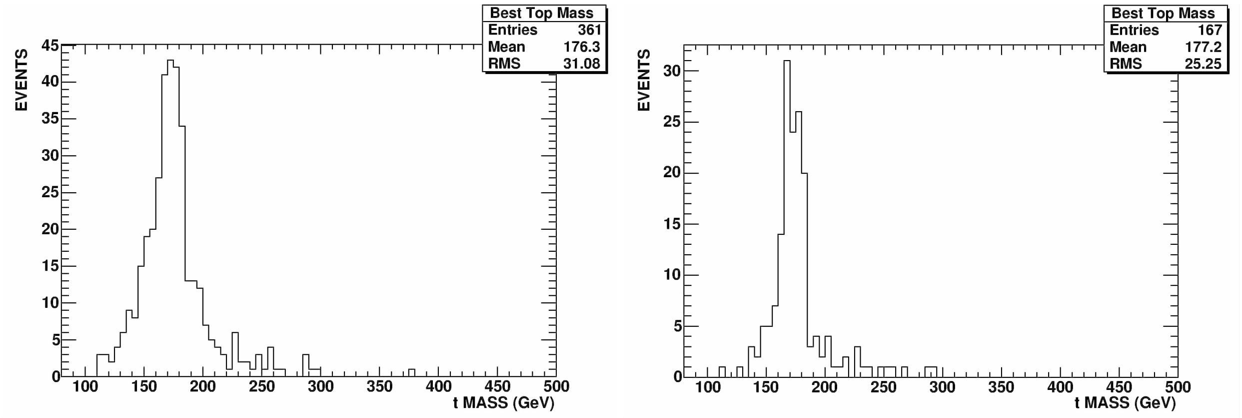


Figure 4.8: [L] t mass distribution relative to 0-tags 4XJ events; [R] t mass distribution relative to 0-tags 4XJ+F events. In both cases cleaning is applied.

0-tags								
set	# events	Ef	rEf	GM	rGM	fEf	M_t (GeV)	RMS (GeV)
BC	3494	0.378	0.378				173.3	35.8
BC+C	1492	0.161	0.161				177.6	34.8
BC+F	1101	0.119	0.374			0.315	174.6	28.4
BC+C+F	548	0.059	0.186			0.367	175.6	24.9
M4q	1891	0.204	0.353	0.541			171.6	29.6
M4q+C	917	0.099	0.171	0.615			173.9	28.5
M4q+F	888	0.096	0.358	0.807		0.470	174.5	25.9
M4q+C+F	462	0.050	0.186	0.843		0.504	175.0	23.8
4sXJ	1160	0.125	0.377				172.9	34.4
4sXJ+C	607	0.066	0.197				176.3	32.8
4sXJ+M4q	743	0.080	0.365	0.213	0.641		173.1	30.5
4sXJ+M4q+C	424	0.046	0.208	0.284	0.699		174.6	29.3
4sXJ+F	422	0.046	0.401			0.364	176.0	29.2
4sXJ+C+F	248	0.027	0.236			0.409	176.0	24.9
4XJ	618	0.067	0.365				175.1	32.6
4XJ+C	361	0.039	0.213				176.3	31.1
4XJ+M4q	478	0.052	0.368	0.137	0.773		175.1	29.4
4XJ+M4q+C	293	0.032	0.225	0.196	0.812		175.6	28.3
4XJ+F	271	0.029	0.403			0.439	177.3	26.9
4XJ+C+F	167	0.018	0.249			0.463	177.2	25.3
AT+4XJ+C	361	0.039	0.213				176.3	31.1
AT+4XJ+C+F	167	0.018	0.249			0.463	177.2	25.3

Table 4.2: Effects of cleaning in the 0-tags sample of a 10955 HERWIG $t\bar{t}$ (175 GeV) lepton plus jets events set.

A sort of “artificial tagging” (AT) was attempted: in clean and 4XJ events the leading jet was forced to be tagged, with the additional request that the ratio between E_T of the first leading jet and the fourth one ought to be greater than 2 (from MC studies), and the mass selection was done as for the tagged sample. Nothing changes from 4XJ+C events: the number of selected events is the same, and the mass distribution too. This means that the combined selection (4XJ+C) selects events where the lowest-χ^2 mass corresponds to the “first leading-to-primary *b*” assignment. All results (including those of 4sXJ events) are summed up in Table 4.2.

5. Concluding Remarks

5.1 The Problem of Statistics

We have addressed the possible improvements obtained with cleaning and found a small progress in the 4XJ sample. However, the 4XJ selection shows a great loss in statistics, reducing the sample of a factor of ~ 5. If cleaning is then applied, the reduction of the sample is very different in the full set of events and in 4XJ events: in the first case cleaning rejects 56-57% of events, but in the second case it rejects only 39-42% of events.

The 4XJ sample is the one where cleaning has better performed, and it has been used as a "starting sample" for top mass studies. As the 4XJ selection rejects a very significant fraction of events ($\sim$ 80%), a "clean sample" could be an alternative starting point if other selections less expensive than the 4XJ are found. For any selection rejecting a fraction of events to be worth, the loss in statistics should be more than compensated by an improved sample quality, so that the measurement can can lead to a better result.

The method in its present version looks unsatisfactory in terms of top mass resolution. We can hope to be able to find a positive conclusion because cleaning causes a progress in a number of significant parameters:

- the topological matching is increased in all the samples, particularly in the 0-tags sample, where $\frac{\Delta rGM}{rGM} = +13.7\%$ (in pretagged and tagged events it is, respectively, $+9.7\%$ and $+6.0\%$), which means that radiating gluons phenomena in the fragmentation of primary partons is minimized;

- the rate of correct parton-to-jet assignments is increased in all the samples, with an again excellent performance on 0-tags events, where $\frac{\Delta fEf}{fEf} = +16.5\%$ (in pretagged and tagged events it is, respectively, $+12.9\%$ and $+8.9\%$);

Table 5.1 sums up the relevant changes in fundamental parameters of goodness, including those of tagged events.

changes in relevant parameters			
sample	pretagged	tagged	0-tags
BC			
rGM	0.580	0.648	0.541
rGM with C	0.636	0.687	0.615
fEf	0.318	0.506	0.315
fEf with C	0.359	0.551	0.367
4XJ			
rGM	0.768	0.794	0.773
rGM with C	0.795	0.811	0.812
fEf	0.397	0.587	0.439
fEf with C	0.409	0.608	0.463

Table 5.1: Changes of relevant parameters to evaluate the effects of cleaning on a HERWIG $t\bar{t}$ single lepton events sample.

5.2 Possible Refinements

The next step of this work will be a revision of the software used for the analysis. This step will be made in order to understand if some problems encountered while doing tests were due to computing reasons or to physical ones.

A problem that was not addressed in this work was how kinematical selections like cleaning affect background events. Particularly in the 0-tags sample, where background is expected to be 50% [ZT2], kinematical selections will be aimed primarily to improve the signal to background ratio. Besides 4XJ, a number of kinematical selections can be tried addressing the energy distribution of jets and their angular correlations, which might be not as expensive as 4XJ in terms of statistics. By applying cleaning one might find a much more significant progress in a realistic signal plus background sample than found in our study of a pure $t\bar{t}$ sample.

A separate line of research will address b-tagged events. The impact of cleaning on tagged top events has already been quickly looked at and is reported in Appendix A. While cleaning was originally conceived to extend the event sample for the top mass measurement, it might turn out to be more important as a means to improve the sample purity and event quality in the b-tagged sample.

A. Study of Tagged Events

A.1 Tagged Sample

As the *b*-tagging is one of main tools in top events identification and in top mass studies, here the test are done over the tagged subset of studied events. The considered *b*-tagging is the SECVTX one, and the masses are selected out of those where the *b*-to-jet assignment is correct (the *b* is assigned to the *b*-tagged jet). In this case the selected mass may be not the lowest-χ^2 one and this χ^2 may be greater than 10. If one thinks in terms of configurations instead of events, the tagged sample is not a strictly contained subset of the pretagged sample, but they overlap only when the selected mass is the same. This behavior can explain some inconsistencies, such as the followings:

- if the number of 0-tags candidates is summed with the number of tagged candidates the total is very lower than the number of pretagged candidates;

- if the number of events with a correct parton-to-jet assignment (F) in the tagged sample is summed to the one of the 0-tags sample one gets a number greater than the one of F pretagged events.

The results of these studies have already been anticipated in Table 5.1.

In BC tagged events the *GM* fraction is 64.8% and the computed fEf is 50.6%. As the *b*-to-*b*-tagged jet assignment is forced, the fEf is sensibly greater than in the previous cases. The *GM* is greater too, but the difference is smaller. After cleaning, the fEf is increased from 50.6% to 55.1%, and the *GM* is increased from 64.8% to 68.7%. Also in this case the fractions are greater than those of the pretagged and the 0-tags samples. The mass distributions for BC events and BC+F events are reproduced in Figure A.1. The same distributions after cleaning are reproduced in Figure A.2.

As for pretagged and 0-tags events, the test was performed on 4XJ events too. In this case *rGM*, when cleaning is not requested, is 79.4% and fEf is 58.7% (both values are once more greater than in pretagged and 0-tags events). If the request of having clean 4XJ events is made, *rGM* increases from 79.4% up to 81.1%, and fEf is still greater than in merely 4XJ events: 60.8% instead of 58.7%. The mass distributions for 4XJ events and for 4XJ+F events are reproduced in Figure A.3 (before cleaning) and in Figure A.4 (after cleaning). All results (including those of 4sXJ events) are summed up in Table A.1.

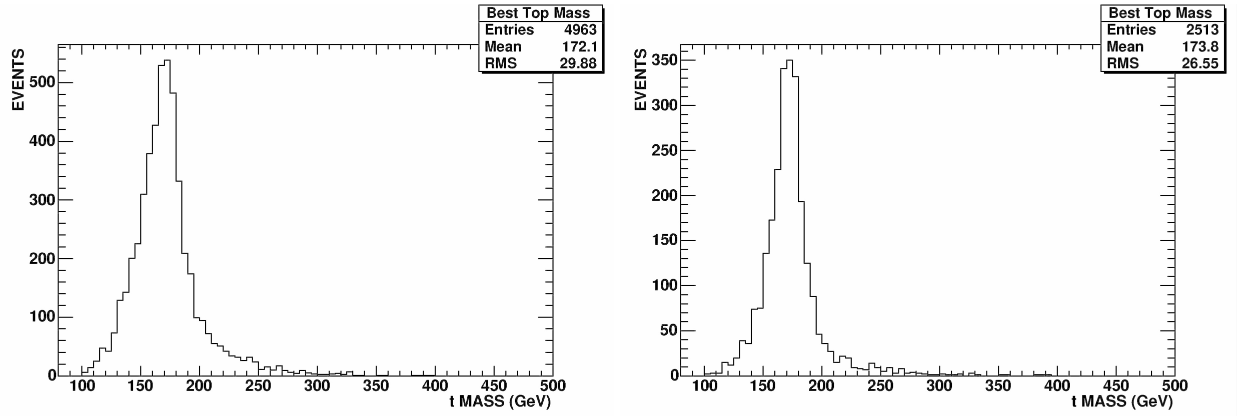


Figure A.1: [L] t mass distribution relative to tagged BC events; [R] t mass distribution relative to tagged BC+F events.

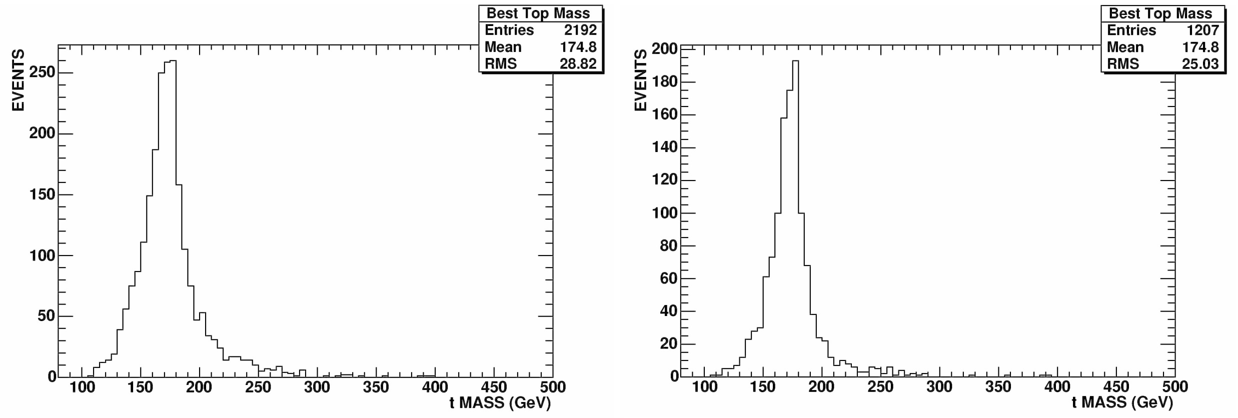


Figure A.2: [L] t mass distribution relative to tagged BC events; [R] t mass distribution relative to tagged BC+F events. In both cases cleaning is applied.

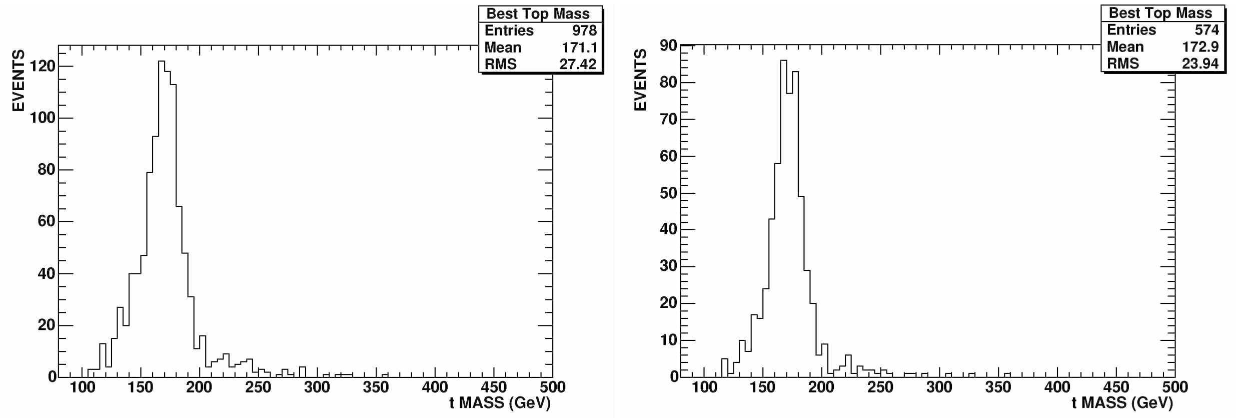


Figure A.3: [L] t mass distribution relative to tagged 4XJ events; [R] t mass distribution relative to tagged 4XJ+F events.

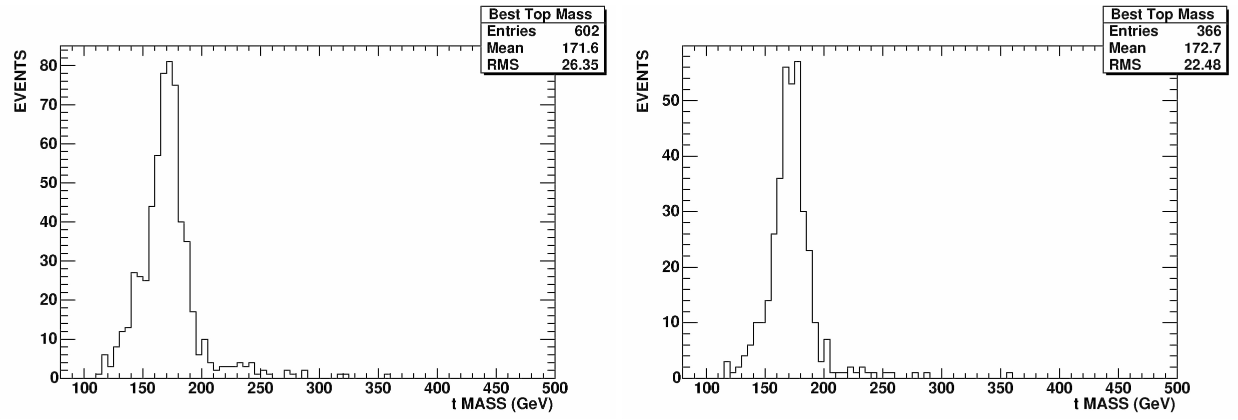


Figure A.4: [L] t mass distribution relative to tagged 4XJ events; [R] t mass distribution relative to tagged 4XJ+F events. In both cases cleaning is applied.

tagged								
set	# events	Ef	rEf	GM	rGM	fEf	M_t (GeV)	RMS (GeV)
BC	4963	0.537	0.537				172.1	29.9
BC+C	2192	0.237	0.237				174.8	28.8
BC+F	2513	0.272	0.854			0.506	173.8	26.6
BC+C+F	1207	0.131	0.410			0.551	174.8	25.0
M4q	3215	0.348	0.599	0.648			171.5	27.0
M4q+C	1505	0.163	0.281	0.687			173.1	25.3
M4q+F	2130	0.230	0.859	0.848		0.663	173.8	25.3
M4q+C+F	1048	0.113	0.423	0.868		0.696	174.5	23.4
4sXJ	1681	0.182	0.546				170.9	28.7
4sXJ+C	896	0.097	0.291				173.0	29.3
4sXJ+M4q	1201	0.130	0.590	0.242	0.714		171.1	26.2
4sXJ+M4q+C	676	0.073	0.332	0.308	0.754		172.0	26.0
4sXJ+F	909	0.098	0.863			0.541	172.0	25.3
4sXJ+C+F	519	0.056	0.493			0.579	173.5	24.7
4XJ	978	0.106	0.578				171.1	27.4
4XJ+C	602	0.065	0.356				171.6	26.4
4XJ+M4q	777	0.084	0.598	0.157	0.794		171.1	24.4
4XJ+M4q+C	488	0.053	0.375	0.223	0.811		170.8	22.5
4XJ+F	574	0.062	0.854			0.587	172.9	23.9
4XJ+C+F	366	0.040	0.545			0.608	172.7	22.5

Table A.1: Effects of cleaning in the tagged sample of a 10955 HERWIG $t\bar{t}$ (175 GeV) lepton plus jets events set.

A.2 Comments

The tagged sample is not saved from the loss in statistics when the 4XJ selection is done. In this case the loss in statistics is slightly less ($\sim 1-2\%$) than in the pretagged or 0-tags sample. In the full tagged sample, the loss in statistics due to 4XJ is 80.3%, and it is only 72.5% in clean and tagged events. If one looks at F events, the loss in statistics is 77.2% (69.7% if the C request is added).

Also in this case the improvement in resolution is not great, both for BC and 4XJ events, and cannot be considered satisfactory.

B. Is Cleaning Independent from Other Selections?

Another interesting aspect of these tests is the independence of cleaning from other selections (if we consider separately cleaning and one or more other selections, do we get the expected number of events if they are combined?). In the following Sections, comparisons with *b*-tagging and 4XJ are shown and explained.

B.1 Cleaning and *b*-tagging are Independent on Each Other

Looking at Tables 4.1 and A.1, and interpreting rEf efficiencies as probabilities, the probability of having a clean events among pretagged is 42.6%. The probability of having a *b*-tagged event (the efficiency of *b*-tagging) is 53.7%. If they were independent, their product should give the probability of having a clean and *b*-tagged event, which is 22.8%, corresponding to $\sim$ 2110 events. The found number of tagged C events is 2182, corresponding to 23.7% rEf.

If 4XJ events are considered, the tagging rEf is 57.8%, and the cleaning rEf is 59.6%, their product is then 34.4%, corresponding to $\sim$ 580 expected events. The computed number of events is 602, corresponding to 35.2% of rEf.

Then, if F events are took under exam, the tagging rEf is 85.4%, the cleaning rEf is 48.0%, corresponding to a product rEf of 41.0%. The expected number of events is then $\sim$ 1210, while the computed number is 1207 ($rEf = 41.0\%$).

If one looks only at those selections that are possible while studying real data, cleaning and tagging are independent enough: in fact the difference $\sim$ 1% in expected and "measured" rEf is not so great to state they are dependent. This very small difference vanishes in events with a correct parton-to-jet assignment (F), and the natural conclusion is the independence of cleaning and *b*-tagging.

B.2 Cleaning and 4XJ are Closely Correlated

The logic behind the following considerations is the same as in previous Section. In this case, rEf of the compound selection can be different from the one listed in summary Tables (in fact, they are differently defined): in this Section rEf of the compound selection should be then understood as the "probability" of finding a 4XJ+C event out of those under exam. In the pretagged sample, as already explained, the cleaning rEf is 42.6%; the $\mathsf{4XJ}$ relative efficiency is 18.3%, leading to an expected efficiency (in case of independence) of 7.8% ($\sim$ 720 expected events). The computed number of events is 1009 ($rEf = 10.9\%$). In the F events subset, the expected number of events is ~ 320 ($rEf = 11.0\%$), while the effective number of candidate events is 413 ($rEf = 14.0\%$).

Proceeding in the same way, in the tagged sample the expected number of clean and $\mathsf{4XJ}$ events is ~ 430 ($rEf = 8.7\%$), against a “measured” number of candidates equal to 602 ($rEf = 12.1\%$). In the F subsample, the expected number of events is ~ 280 ($rEf = 11.0\%$), while the effective one is 366 ($rEf = 14.6\%$). In the 0-tags sample the expected number of events are, respectively for the full sample and the F one, ~ 260 and ~ 135 ($rEf = 7.5\%$ and $rEf = 12.2\%$), while the effective ones are 361 and 167 ($rEf = 10.3\%$ and $rEf = 15.2\%$).

In any case, cleaning and $\mathsf{4XJ}$ are not independent selections, but they can be still complementary; moreover, relative efficiencies follow the laws of conditioned probability:

$$\mathbb{P}(\mathsf{C}|\mathsf{4XJ}) = \frac{\mathbb{P}(\mathsf{4XJ}+\mathsf{C})}{\mathbb{P}(\mathsf{C})} \simeq \frac{rEf(\mathsf{4XJ}+\mathsf{C})}{rEf(\mathsf{C})} \tag{B.1}$$

The computed fraction for the pretagged sample is 59.6%, it is 61.5% for the tagged one and 58.4% for the 0-tags one. The mutual dependence of $\mathsf{4XJ}$ and C could have been already expected after the study of the loss in statistics due to $\mathsf{4XJ}$, which was strongly lower in clean events.

Bibliography

[BET] Stefano Bettelli.
Metodi ottimali per la misura della massa del top quark.
Tesi di Laurea, Università di Pisa, A.A. 1995-1996.

[BRU] Erik M. Brubaker.
A Measurement of the Mass of the Top Quark in Lepton + Jets Events at CDF.
Ph.D. Thesis, University of California, Berkeley, 2005.
CDF internal note CDF/THESIS/TOP/PUBLIC/7624, 3rd May 2005.

[HER] G.Corcella *et al.*
HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons.
JHEP **101**, 10 (2001).
url: http://www.arxiv.org/abs/hep-ph/0011363

[JET] The CDF Collaboration – F. Abe *et al.*
Topology of three-jet events in $p\bar{p}$ *collisions at* $\sqrt{s} = 1.8$ *TeV.*
Phys. Rev. D **45**, 1448 (1992).
url: http://prola.aps.org/abstract/PRD/v45/i5/p1448_1

[KIN] J. Antoš, A. Beretvas, Y.C. Chen, R. Lysák, M. Šiket, G.P. Yeh.
Top Quark Mass from a Full Kinematic Analysis of Dilepton Events in RUN II.
CDF internal note CDF/PHYS/TOP/CDFR/6465, 23rd May 2003.

[MEH] Andrew Mehta.
Jet Resolution Studies of the H1 Algorithm to Combine Tracks and Calorimeter Towers.
CDF internal note CDF/PHYS/JET/GROUP/7468, 4th February 2005.

[OFF] The CDF Collaboration.
CDF Offline Analysis Resources.
url: http://cdfcodebrowser.fnal.gov/CdfCode/

[PDG] Particle Data Group – S. Eidelman *et al.*
Phys. Lett. B **592**, 1 (2004).
url: http://pdg.lbl.gov/2004/tables/contents_tables.html

[PER] Donald H. Perkins.
Introduction to High Energy Physics.
Cambridge UP, 2000.

[PRD] The CDF Collaboration – D. Acosta *et al.*
Top Quark Mass Measurement Using the Template Method in the Lepton + Jets Channel at CDF II.
To be published (draft 5th July 2005).
url: http://www-cdf.fnal.gov/~brubakee/internal/topMassPRD/topMassPRD.v2.0.ps

[TAG] Anant Gajjar, Ronan McNulty, Andrew Mehta, Tara Shears.
Measurement of the Inclusive $b\bar{b}$ *Jet Production Cross Section.*
CDF internal note CDF/PHYS/JET/PUBLIC/7601, 23rd April 2005.

João Guimarães da Costa, Sal Rappoccio, Daniel Sherman.
Measurement of the SecVtx Mistag Asymmetry in 5.3.3
CDF internal note CDF/ANAL/SEC_VTX/CDFR/7585, 12th April 2005.

Monica D'Onofrio.
Preliminary results on inclusive b-jet cross section.
CDF internal note CDF/PHYS/JET/PUBLIC/6984, 2nd May 2005.

http://www-cdf.fnal.gov/physics/new/top/public/btag/

[TCS] Giorgio Bellettini, Julian Budagov, Guram Chlachidze, George Velev.
Top Mass Analysis using 3 best χ^2 *Permutations.*
CDF internal note CDF/PHYS/TOP/GROUP/6770, 19th November 2003.

[TDR] The CDF Collaboration – F. Abe *et al.*
The CDF II Detector Technical Design Report.
FERMILAB-Pub-96/390-E, 1996.
url: http://www-cdf.fnal.gov/upgrades/tdr/tdr.html
The CDF IIb Detector Technical Design Report.
September 2002.
url: http://www-cdf.fnal.gov/upgrades/run2b/Documents/Documents.html

[TEV] Fermilab Accelerator Division.
Accelerator Division web page.
url: http://www-bd.fnal.gov/

[TRI] Cheng-Ju Lin.
Introduction to CDF L1/L2 Triggers.
Roberto Carosi.
CDF online SVT Trigger.
Lecture Series on CDF Detector, 18th November 2004.
url: http://www-cdf.fnal.gov/internal/WebTalks/Archive/0411/

[TXS] The CDF Collaboration – D. Acosta *et al.*
Measurement of the $t\bar{t}$ *production cross section in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.96$ *TeV using lepton + jets events with secondary vertex b-tagging.*
Phys. Rev. D **71**, 52003 (2005).
url: http://arxiv.org/abs/hep-ex/0410041

[VEL] Gueorgui Velev.
Top Mass Measurement at Tevatron Run II.
CDF internal note CDF/PUB/TOP/PUBLIC/7674, 3rd June 2005.
Les Réncontres de Physique de la Vallée d'Aoste, La Thuile, 4th March 2005.
url: http://www.pi.infn.it/lathuile/2005/programme.htm

[WID] A.H. Hoang, T. Teubnerb.
Top Quark Pair Production Close to Threshold: Top Mass, Width and Momentum Distribution.
Phys. Rev. D **60**, 114027 (1999).
url: http://www.arxiv.org/abs/hep-ph/9904468

[YOR] Kohei Yorita.
Measurement of the Top Quark Mass by Dynamical Likelihood Method using the Lepton plus Jets Events.
Ph.D. Thesis, Waseda University, Tokyo, 2005.
CDF internal note CDF/THESIS/TOP/PUBLIC/7631, 5th May 2005.

[ZT1] Giorgio Bellettini, Julian Budagov, Guram Chlachidze, George Velev.
Top Quark Mass Measurement in Non-tagged Lepton+Jets events at CDF.
CDF internal note CDF/PHYS/TOP/PUBLIC/7245, 17th September 2004.

[ZT2] Giorgio Bellettini, Julian Budagov, Guram Chlachidze, George Velev.
Top Mass Reconstruction in Non-tagged Lepton+Jets events.
CDF internal note CDF/PHYS/TOP/GROUP/7063, 20th July 2004.

Acknowledgments

This job was possible thanks to many people who helped me: first of all my advisor Giorgio Bellettini, who trusted me allowing this amazing experience happen; Guram Chlachidze and the Dubna CDF Group, who supervised my work in Fermilab; Giorgio Chiarelli and Antonio Sidoti for the help they gave me during my first struggles with CDF offline; Gueorgui Velev for his talk in La Thuile; Fiodar Prokoshin and Giuseppe Latino for many clearing discussions on CDF software and Jet Reconstruction; Monica D'Onofrio for information about fake rates in *b*-tagging; Roberto Rossin, Alberto Belloni, Michele Giunta, Diego Tonelli, Stefano Torre and Pierluigi Catastini for many small hints, and not only . . .

Pisa, 13th July 2005