Contents

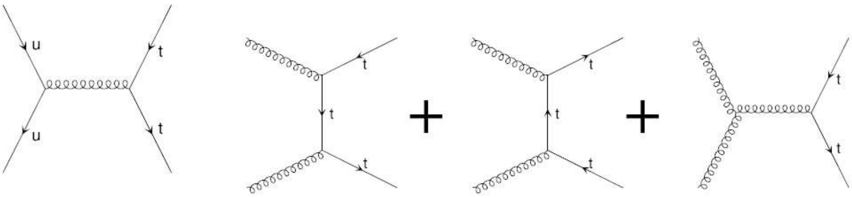
1 Introduction

This thesis deals with the problematic of top quark charge measurement in CDF experiment at Fermilab. The goal is to determine if the top quark observed on Tevatron experiments is the Standard Model particle with the predicted charge 2/3 or it is some exotic 4^{th} generation quark with the charge of -4/3 as suggested by some alternative theories.

The CDF top quark charge measurement uses two decay channels of $t\bar{t}$ pair for the charge estimation. The author's contribution was in so called lepton + jets channel (LJ), but to have the complete picture, also the work done in dilepton channel (DIL) is mentioned.

2 Top quark decay

The top quark was first experimentally observed in the experiments CDF and D0 on Tevatron accelerator at Fermilab in 1995 [1][2]. Top quark is produced mainly by strong interactions in $t\bar{t}$ pairs created by quark-antiquark ($q\bar{q}$) annihilation or gluon fusion (gg) (fig 1). In addition to that, the top quark can be (under the SM) also produced through electroweak interactions.



$q\bar{q} \rightarrow t\bar{t}$ $gg \rightarrow t\bar{t}$

Figure 1: Feynman diagrams for $t\bar{t}$ production via strong interaction

At the time of the discovery, the top mass was determined to be $m_t \approx 174$ GeV/c^2. For this mass the SM predicts the decay width $\Gamma \approx 1.4$ GeV [3] and the half-life $\tau \approx 10^{-24} s$.

2.1 Top quark decay modes

Because of the short life time of the top quark, it decays before it can hadronize. The top events (events with top quark) can be categorized by the final decay product configuration. The dominant is the electroweak decay to b quark and W boson, the electroweak decays to s or d quark are significantly suppressed (the corresponding elements of CMK matrix are small). The W

boson also lives very short time and it is observable only via its decay products. It can decay either into two quarks ($W \rightarrow q\bar{q}$) or leptonicaly ($W \rightarrow l\nu$). From experimental point of view it is convenient to introduce three different categories of the $t\bar{t}$ pair decay (so called decay modes), accordingly to the W bosons decays [5]:

1. Hadronic mode (44% of all $t\bar{t}$ decays) - both of the W bosons from $t\bar{t}$ pair decay hadronicaly
2. Semileptonic mode (30% of all $t\bar{t}$ decays) - one of the W bosons decay hadronicaly, the other one leptonicaly (only decay to e, μ taken into account)
3. Dilepton mode (5% of all $t\bar{t}$ decays) - both of the W bosons decay leptonicaly (only decay to e, μ taken into account)

- The W boson decays into τ ($W \rightarrow \tau\nu$) are not taken into account because the τ can decay both hadronicaly and leptonicaly. The techniques to identify such decays are rather complex, therefore this decay mode is not used for most top quark analyses.

Hadronic mode

Although this decay mode is the most common, its disadvantage is the big QCD background [4]. To select the signal, high transverse momentum p_t of jets is requested. The signal to background ration can be improved with increase of required number of jets and by use of b-tagging (procedure to identify jets coming from b quarks).

Semileptonic mode

This mode is also called the lepton plus jets mode. In this mode, an isolated lepton with high transverse momentum p_t and missing transverse energy (indicating neutrino) is required. The signature of this event is also characteristic by high transverse momentum jets (two coming from W decay and two from b quark decay). The b-tagging requirement is not necessary, on the other hand it helps to significantly reduce the background. The background processes include mainly QCD, W production with many jets and diboson (WW, WZ, ZZ) production.

Dilepton mode

For selection of this mode, two isolated leptons with high transverse momentum and large missing transverse energy (from two neutrinos) are required. Although this decay channel is the rarest, its advantage is a significantly reduced QCD background. The other background processes for this mode include Drell-Yan ($Z^*/\gamma \rightarrow e^+e^-, \mu^+\mu^-$), $Z \rightarrow \tau\tau$ and diboson production.

3 The determination of Top Quark charge

Since the discovery of the top quark, CDF has made several analyses to confirm that it has the properties expected in the SM - decay fraction into different final states, Lorentz structure of the $t \rightarrow bW$ vertex, mass, etc. One of the properties still pending is the charge.

The electromagnetic couplings can be measured using $t\bar{t}\gamma$ events [6], although this measurement needs more data than is available on CDF. Alternatively, one can reconstruct the top charge from the charge of its decay products.

The SM top quark is expected to have the charge $+2/3$. However, there is a theory suggesting another explanation for the CDF Run 1 experimental data [7]. This theory states that the quark with the mass $m \approx 170$ GeV/c^2 discovered at Tevatron is not the expected SM top quark, but an exotic quark of a doublet $(Q_1, Q_4)_R$, where Q_4 has the charge $-4/3$ and Q_1 that mixes with the right component of b-quark, has the charge -1/3. The SM top quark with the charge 2/3 should have the mass $m_t = 274$ GeV/c^2 in this scheme. Using such an exotic quark qives better results in fitting electroweak data from precision measurements at LEP, SLAC and Tevatron colliders.

To reconstruct the top charge from the charges of its decay products in $t\bar{t}$ events, the following ingredients are needed:

- correctly associating the W boson with the right b-jet from the same top quark decay
- determining the charge of the W boson (from leptonic decay $W \rightarrow l\mu$)
- determining the flavor of the b-jet

The necessary steps for obtaining the needed informations are described in this chapter.

3.1 Event selection

For the top charge analysis we have used only the di-lepton and lepton + jets samples. The fully hadronic sample has not been used due to a huge QCD background expected in this case.

3.1.1 Dilepton event selection

The event selection for dilepton channel follows the selection used in the top cross section analysis based on the data coresponding to integrated luminosity $1.2 fb^{-1}$ [8].

- Two isolated tight leptons with $E_T > 20$ GeV with opposite sign.
- At least two jets with high transversal energy, $E_T > 15$ GeV
- high missing transversal energy, $\vec{\not\!\!E}_t > 25$ GeV
- $H_T = p_{Tlep} + E_{Tjet} + \vec{\not\!\!E}_t > 200$ GeV
- Z veto. The event is vetoed if the lepton and another object have an invariant mass with energy between 76 Gev and 106 GeV. If the lepton is an electron, the other object can be an electromagnetic object, a jet or a track of an opposite charge particle. If the lepton is muon, the other object can be a minimum-ionizing track of opposite charge.
- One jet tagged as b-jet using Tight SecVtx tagging algorithm [10].

3.1.2 Lepton + jets event selection

The event selection in the lepton + jets channel follows the selection used in the top cross section analysis based on the data coresponding to integrated luminostity $1.12 fb^{-1}$ [9].

- One tight lepton - electron or muon with $p_T > 20$ GeV.
- $\vec{\not\!\!E}_t > 20$ GeV.
- Dilepton veto. The event is removed if additional loose lepton is found.
- Z veto. The same as in case of dilepton event selection.
- Cosmic veto. The event is vetoed if the lepton is a muon identified as cosmic.

- Conversion veto. The event is vetoed if the lepton is an electron identified as coming from a photon conversion.
- At least 4 tight jets or 3 tight jets and one loose jet .
- Two jets tagged as b-jets using Loose SecVtx tagging algorithm [10].

3.2 Optimization of parameters

3.2.1 Definition of performance factors

The analysis of the top charge consist of several algorithms (lepton and b-jet pairing, b-jet flavor tagging, etc.) each of them containing parameters that need to be optimized. In order to be able to do this, there is a need for quantitative criteria for the best options. The variables that were chosen for this analysis are:

- efficiency - the number of events remaining after certain selection criteria over the number of events available before applying the cut
- purity - the number of events that are correctly identified (based on MC information) over the number of events remaining after the cut

There is a trade-off between efficiency and purity. It is good to have as many events as possible to have a statistically significant measurement, but at the same time, to have a many of wrongly assigned events will dilute the measurement.

3.2.2 Pairing between lepton and b-jet

As was mentioned before, for the reconstruction of top charge, there is a need to correctly associate lepton coming from W boson decay with the b-quark from the same top quark decay. As the event characteristics is not the same for dilepton and lepton + jets channels, different methods for lepton b-jet pairing are used. In dilepton channel, the criteria for correct pair selection is based on invariant mass of the lepton - b-jet system. In lepton + jets channel, the selection is based on complete kinematic fit of the event.

Lepton b-jet pairing in dilepton channel

As was mentioned before, for selection of the right lepton-jet pair in dilepton channel the squared value of invariant mass of the pair M_{lb}^2 was used. Based on the event selection criteria, in the final state of $t\bar{t}$ decay there are two leptons and two or more jets, at least one of them tagged as b-jet. When ordered according to transverse energy, the first two jets with highest E_T are considered as b-jets where one of them is b-tagged. If the b-tagged jet is not among this two jets, the event is not used.

In dilepton events, with 2 b-jets and 2 leptons, there are two possible combinations of lepton b-jet pairing and for each combination two lepton-jet invariant masses can be reconstructed.

To optimize the selection method from the purity point of view, the four values of M_{lb}^2 existing in this case are ordered, and the combination that does not contain the highest of the 4 M_{lb}^2 values is chosen as the right one. To increase the purity, a lower limit for the M_{lb-max}^2 is set. By the M_{lb-max} is denoted the maximal of the 4 lepton b-jet invariant masses in dilepton event. The events where the $M_{lb-max}^2 < 21,000$ GeV$^2/c^4$ are refused, since for these events it is hard to select the right combination.

Lepton b-jet pairing in lepton + jets channel

In the lepton + jets channel a different strategy for the pairing is used. It is based on the event by event kinematic fit that is used in the top quark mass measurements. The essence of this approach is in full reconstruction of event topology.

Generaly in lepton + jets channel there is one lepton and at least four jets, two of them tagged as b-jets in the final state of $t\bar{t}$ decay. In this channel one of the W bosons decays into lepton and neutrino, the other into two light quarks. The top quark, which the leptonicaly decaying W comes from, and its decay products are called leptonic branch of the $t\bar{t}$ decay. The top quark, which the hadronicaly decaying W comes from, and its decay products are called hadronic branch. The b-jets are called leptonic or hadronic according to which branch they belong to.

Ignoring the experimental inormation about b-tagging, there are 12 ways how to choose 2 b-jets from the 4 jets of event. On top of this, the sign of neutrino momentum z-component is not known therefore there are altogether 24 possible topological combinations for arangement of jets (assigning each of the jets the position of either leptonic or hadronic b-jet or one of the light quark jets coming from W decay) and neutrino z-direction. For each of these combinations, the full kinematic fit of the event is done using TopMassFitter (part of the CDF analysis tools package) yielding the value of χ^2 and the kinematic parameters of the event.

As the information about b-tagging is part of the Top Mass Fitter input, in the end each of the 24 combination does have a flag, which tells if the combination satisfies the tagging requirement - meaning that if the jets tagged as b-jets (using SecVtx tagging algorithm) are in position of either leptonic or hadronic b-jet in the final configuration, the combination is labeled as a "tagged" one. In case of this analysis, where two b-tagged jets are required, there are only four tagged combinations. As the right one of them, the combination with lowest value of χ^2 is taken.

3.2.3 Jet flavor tagging algorithm

The next important part of the analysis is the determination of the flavor of b-jet, i.e. to determine whether the jet is from a b-quark or from an $\bar{b}$-quark. There are several methods of obtaining the information on the b-quark flavor.

The approach used in the analysis is to calculate the weighted charge of the tracks coming from b-quark hadronization, as their charges are sensitive to certain degree to the charge of the original b-quark.

Track selection for jet charge calculation

To use the weighting method for the jet charge calculation, the tracks belonging to a given jet should be identified. For the track-to-jet association, the relative angle of track and jet axis was used.

For the jet charge calculation only the tracks from the group of tracks used for the secondary vertex fit (SecVtx tracks) were selected. The cut on $p_T > 1.5$ GeV and the cut on the impact parameter with respect to secondary vertex $|d_0| < 0.15$ cm were set.

Optimization of track weighting

For the tagging of the flavor of the b-jet, the algorithm using weighted charge of the tracks inside b-jet cone was chosen. Using this method, the charge of the jet is calculated as

$$q_{jet} = \frac{\sum_i w_i^\kappa \cdot q_i}{\sum_i w_i^\kappa} \qquad (1)$$

where q_i is the charge of the i-th track, w_i is the appropriate weight assigned to this track and κ is an optimization coefficient. It is also required for each jet to have at least two tracks, satisfying the selection criteria mentioned above, associated with it. For the purpose of the analysis, the weight was chosen in the form of $w_i = \vec{j} \cdot \vec{p}_i$.

3.3 Calibration of jet flavor tagging algorithm

One of the biggest challenges in the top quark charge analysis is to tag correctly the flavor of b-jet. The flavor tagging method was described in the previous section. However, the method is sensitive to details of the fragmentation process, where the MC cannot be fully trusted to reproduce the real situation. Therefore the purity of the algorithm should be obtained from b-jet containing experimental data. It is needed especially because the value of purity and its uncertainty are the parameters for which the top quark charge measurement is the most sensitive. The value of purity serves as one of the input parameters in estimating the statistical significance of observed experimental results.

3.3.1 Jet charge calibration method

To be able to calibrate the jet flavor tagging algorithm, the muon enriched dijet data sample was used. Using an appropriate event selection, a sub-sample with a high population of $b\bar{b}$ pairs can be obtained. In these events, one of the jets is identified by the heavy flavor hadron that decays semileptonicaly, producing the triggered muon. This jet is referred to as *muon* jet. The other jet, opposite to the muon one is referred to as the *away* jet.

If the muon and away jet are indeed from $b\bar{b}$, the charge of the muon should be correlated with the flavor of the away jet (if there is no mixing the μ charge has the opposite sign compared to the b-quark initiating the away jet), which is obtained by applying the flavor tagging algorithm.

A main drawback of this method is in determination of the $b\bar{b}$ dijet fraction ($f_{b\bar{b}}$) from the pairs passing the selection cuts, since muon can be also produced by a charm meson decay or can be a jet, misidentified as muon, paired with a light jet wrongly assigned as an away jet, for example in gluon splitting events.

To take all these effects into account, the observed purity P_{obs} can be expressed as:

$$P_{obs} = f_{b\bar{b}}(1 - f_{secmix})P_{JQ} + f_{b\bar{b}}f_{secmix}(1 - P_{JQ}) + 0.5 \cdot f_{bkgd} \tag{2}$$

where P_{JQ} is the wanted purity of that is the performance of jet flavor tagging algorithm in b-jets, f_{secmix} is the fraction of secondaries or mixing and f_{bckg} is the fraction of background with assigned purity 0.5.

In order to determine the fraction of true $b\bar{b}$ events, the transverse momentum of the muon relative to the muon jet axis ($p_{T,rel}$) and the invariant mass of the secondary vertex on tagged away jet (M_{vtx}) [11] were used.

The $b\bar{b}$ fraction is obtained by fitting the corresponding data spectra on the muon and away jets with templates from MC. The P_{obs} is corrected for non-$b\bar{b}$ contribution. Then the purity of jet flavour tagging algorithm P_{JQ} for signal can be calculated, which is later used in the statistical treatment.

3.3.2 Measuring the $b\bar{b}$ fraction

The problem now is to actually measure the $b\bar{b}$ fraction ($f_{b\bar{b}}$) in the data sample. In order to extract the b content on the muon side, the $p_{T,rel}$ distribution of the muon transverse energy with respect to the jet axis, was used. To obtain the shapes of muon $p_{T,rel}$ distributions used as templates, the heavy flavor enriched MC sample (subsample of di-jet MC, where the presence of muon on generator level was required) was used for the parton matched b and c jets, and the di-jet MC for the light/gluon jets. The $p_{T,rel}$ templates are shown in Figure 2 for three selected intervals of away jet E_T.

For the away jet side, the invariant mass, M_{vtx}, of the secondary vertex tracks belonging to the away jet is reconstructed and used. The M_{vtx} templates were obtained using the di-jet MC sample, where the jets were required to pass the same selection cuts listed for the enriched sub-sample. In the case of light jets only events with no heavy flavor quarks were used (based on MC generator information). Figure 3 shows the M_{vtx} templates for three selected intervals of away jet E_T.

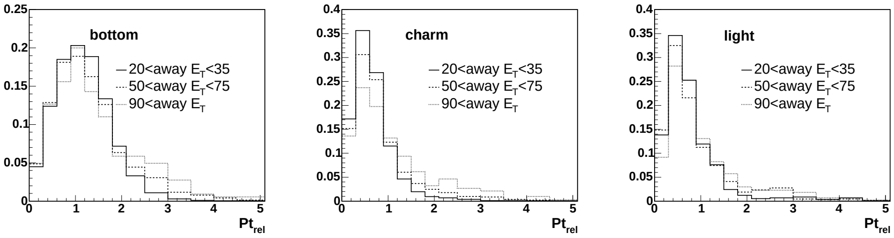


Figure 2: The $p_{T,rel}$ templates for bottom, charm and the light quark/gluon jets obtained from MC for three selected intervals of away jet E_T

3.3.3 Scale factor

Following the above described procedure the muon calibration data of integrated luminosity 1.5 fb^{-1} were fitted and the b fraction of muon jets was found. The $b\bar{b}$ fraction in the experimental sample is obtained for each E_T bin.

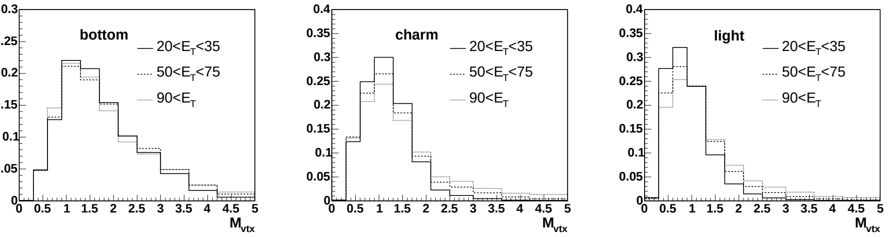


Figure 3: The M_{vtx} templates for bottom, charm and light quark/gluon jets obtained from MC for three selected intervals of away jet E_T

The last ingredient for the calculation of the purity P_{JQ}, is the observed purity P_{obs}. The value of P_{obs} is found as a fraction of the OS events in each bin of the data sample.

To be able to get the jet flavor tagging purity (P_{JQ}) for any sample, in particular for the high E_T b-jets in top events, the scale factor (SF) was calculated as the ratio of the purity P_{obs} in muon calibration data and P_{obs} in a weighted average between a generic and an eriched MC samples (taking into accout the size of the samples).

As the observed purity P_{obs} in MC samples, the weighted average between the jet charge purity observed in the heavy flavor enriched MC samples and in the generic (di-jet) MC one was used.

The purity of the jet flavor tagging algorithm in data is obtained by inserting the P_{obs} values, shown in the previous section, and the $b\bar{b}$ fractions into equation 2.

Figure 4 presents the result for the SF, for the loose tagger, with a fit of the ratio between data and MC with a constant function. A linear fit, used to obtaind an uncertainty due to the E_T dependence is also shown.

3.3.4 Systematic uncertainties

There are several sources of systematic uncertainties related to the procedure used to find the b fraction on the muon jet side and also related to the b fraction on the away jet side. Also an uncertainty due to the E_T dependence is present.

Table 1 shows the systematic uncertainties assigned to SF, from the above sources. The uncertainty due to E_T dependence is also included.

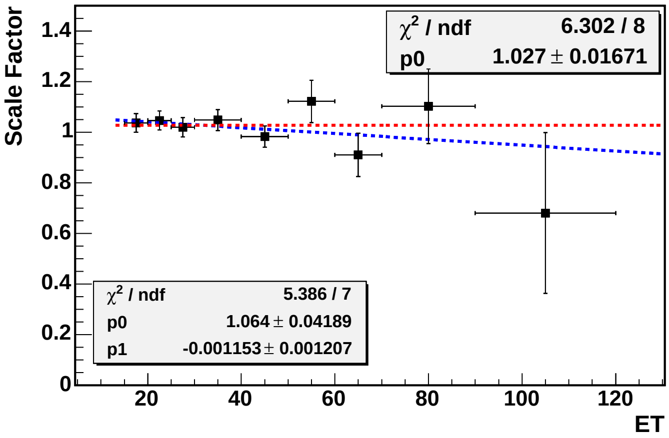


Figure 4: Scale factor as a function of E_T, for the loose tagged jets, calculated from the ratio between the JQ purity in muon calibration data and in a weighted average between a generic and an eriched MC. The red line corresponds to a fit with a constant function, while the blue is the fit with linear function with non-zero slope.

Systematic source	Relative Systematic Uncertainty (%)
tag bias	1.0
non-b	0.1
M_{vtx} fit (bc)	1.
M_{vtx} fit (bl)	0.1
track rec. ineff.	1.5
E_T dependence	1.
total	2.3

Table 1: Relative systematic uncertainties on the scale factor

3.4 Systematics

The uncertainties on number of signal and background events due to systematics are already included on number obtained from cross-section predictions. To get the full systematic uncertainty of the analysis, the systematic uncertainties on the efficiencies and purities of pairing and jet flavor tagging need to be evaluated. Those uncertainties include:

- Jet energy scale (JES)
- Initial/Final state radiation (ISR/FSR)
- Top mass uncertainty
- b-tagging procedure
- parton distribution functions (PDF)
- MC modeling

The combined systematics uncertainty on the efficiencies and purities are calculated by summing each individual contribution in quadrature. Tables 2 summarizes the sources and amount of systematics for lepton + jets and dilepton channels respectively. The systematics uncertainties are shown for the total event selection efficiency, purity of pairing and jet flavor tagging purity. They were determined using standard CDF procedures.

3.5 Expectations

After all the optimizations and background and systematic studies, the final expectation numbers can be obtained. These numbers include the total efficiencies and purities for the signal and background and the final number of the signal and background events.

To get this numbers, the background and signal predictions are taken from the cross-section studies for both dilepton [8] and lepton + jets channel [9]. The number of events expected for the signal and each background with the corresponding efficiencies is shown in Table 3. The uncertainties on N_b and N_s are propagated from the prediction numbers and efficiencies.

After getting the expected number of events for the signal and each backgroud, the number of events with the SM signature and those with the XM signature are estimated using the charge asymmetries obtained from the background studies. For the background where no charge asymmetry is expected, the value 0.5 ± 0.0 is used. For the backgrounds where the charge

all numbers are in %	pairing eff	pairing pur	jetQ eff	jetQ purity
L+J				
ISR/FSR	2.8	0	0	0.4
MC generator	0.6	0	0.1	(1.64)
JES	0.3	0	0	0
PDF	1	0.3	0	0
top mass	1.3	3.3	0.1	0.54
total	3.3	3.4	0.1	0.7
DIL				
ISR/FSR	3.1	0.5	0.3	0.7
MC generator	0	0	1.0	(2.0)
JES	4.4	1.1	0.4	0
PDF	4.0	0.4	0	0
top mass	3	1	0	0
total	7.3	1.6	1.1	1.7

Table 2: Systematics uncertainties for the lepton + jets and dilepton channels.

asymmetry is expected, the actual value from the MC studies is used. The final number of expected SM like and XM like events for each background are shown in Table 4.

When all the purities for each background are collected, they need to be combined into total background purity. The total purity is calculated considering the fraction of asymetric backgrounds in respect to the symetric rest of the background and using their measured purities.

For the signal, the efficiency shown in Table 3 is the combined efficiency of pairing and flavor tagging. The estimation of purity is not so straightforward, since the purity is different for cases when b-tagged jets really correspond to b-quarks, or not. The purities for both of the cases can be found by examining b-jet charges event by event in MC, and after that the difference of flavor tagging algorithm performance between MC and data needs to be taken into account. This study was explained previously, and as a result of that the scale factor $SF_{JQ} = 1.03 \pm 0.01 \pm 0.04$ was obtained. Apart of this, there is another scale factor (SF_{nonb}) to take into consideration, which corresponds to the mistags. To include all the effects, each term that is contributing to N^+ was separated and the expression for the combined purity was constructed:

background	**prediction**	**efficiency**	N_b **or** N_s
L + J			
W+HF	10.23 ± 4.31	0.14 ± 0.01	1.47 ± 0.62
QCD fakes	4.06 ± 4.94	0.15 ± 0.03	0.61 ± 0.75
Diboson	0.95 ± 0.15	0.20 ± 0.03	0.19 ± 0.04
Mistag	2.29 ± 0.68	0.15 ± 0.01	0.33 ± 0.10
Singletop	2.64 ± 0.38	0.21 ± 0.01	0.55 ± 0.08
Total	20.17 ± 6.61	-	3.15 ± 0.99
Signal	138.56 ± 24.02	$0.52^{\pm 0.002(stat)}_{\pm 0.02(sys)}$	72.09 ± 12.73

DIL			
Drell-Yan	$0.51^{+1.02}_{-0.51}$	0.30 ± 0.05	$0.15^{+0.31}_{-0.15}$
Fakes	$2.82^{+5.24}_{-2.82}$	0.25 ± 0.04	$0.71^{+1.39}_{-0.69}$
Diboson	$0.19^{+0.38}_{-0.19}$	0.5 ± 0.08	$0.9^{+0.19}_{-0.09}$
Total	$3.52^{+5.75}_{-2.87}$	-	$0.96^{+1.47}_{-0.73}$
Signal	41.09 ± 3.8	$0.33^{\pm 0.003(stat)}_{\pm 0.02(sys)}$	13.44 ± 1.60

Total Background	$1.404^{+0.961}_{-0.336}$
Total Signal	$45.652^{\pm 0.571(stat)}_{\pm 8.062(sys)}$

Table 3: The background and signal predictions with the top quark charge specific efficiencies for the sample 1.5 fb^{-1}

background	N_b **or** N_s	**purity**	N^+	N^-
L + J (1.5 fb^{-1})				
W+HF	1.47 ± 0.62	0.5 ± 0.0	0.74 ± 0.31	0.74 ± 0.31
QCD fakes	0.61 ± 0.75	$0.504^{+0.001}_{-0.004}$	0.31 ± 0.38	0.30 ± 0.37
Diboson	0.19 ± 0.04	0.5 ± 0.0	0.09 ± 0.02	0.09 ± 0.02
Mistag	0.33 ± 0.10	0.5 ± 0.0	0.17 ± 0.05	0.17 ± 0.05
Singletop	0.55 ± 0.08	0.51 ± 0.01	0.28 ± 0.04	0.27 ± 0.04
Total	3.15 ± 0.99	0.503 ± 0.002	1.59 ± 0.50	1.57 ± 0.49
Signal	72.09 ± 12.73	$0.569^{\pm 0.004(stat)}_{\pm 0.010(sys)}$	41.02 ± 7.28	31.07 ± 5.54

DIL (1.5 fb^{-1})				
Drell-Yan	$0.15^{+0.31}_{-0.15}$	0.5 ± 0.0	$0.08^{+0.15}_{-0.08}$	$0.08^{+0.15}_{-0.08}$
Fakes	$0.71^{+1.43}_{-0.71}$	0.52 ± 0.02	$0.37^{+0.74}_{-0.37}$	$0.34^{+0.69}_{-0.34}$
Diboson	$0.09^{+0.9}_{-0.19}$	0.5 ± 0.0	$0.05^{+0.09}_{-0.05}$	$0.05^{+0.09}_{-0.05}$
Total	$0.96^{+1.47}_{-0.73}$	$0.513^{+0.016}_{-0.014}$	$0.49^{+0.76}_{-0.38}$	$0.47^{+0.71}_{-0.35}$
Signal	13.44 ± 1.60	$0.587^{\pm 0.006(stat)}_{\pm 0.013(sys)}$	7.89 ± 0.96	5.55 ± 0.69

Total Bckg.	$4.11^{+1.77}_{-1.23}$	0.505 ± 0.005	$2.08^{+0.91}_{-0.63}$	$2.04^{+0.86}_{-0.61}$
Total Signal	85.54 ± 12.83	$0.572^{\pm 0.003(stat)}_{\pm 0.008(sys)}$	48.91 ± 7.35	36.62 ± 5.58

Table 4: Purities and number of expected SM like (N^+) and XM like (N^-) events for each background and signal for the sample of 1.5 fb^{-1}

$$\begin{aligned} p_{comb} &= f_{nonb} \cdot SF_{nonb} \cdot p_{nonb} + (1 - f_{nonb} \cdot SF_{nonb}) \cdot \\ &\quad (p_{pair} \cdot p_{JQ} \cdot SF_{JQ} + (1 - p_{pair}) \cdot (1 - p_{JQ} \cdot SF_{JQ})) \end{aligned} \tag{3}$$

where f_{nonb} is the fraction of b-tagged jets which do not really come from b-quark, SF_{nonb} is the scale factor accounting for mistag rate, p_{nonb} is the purity of the non-b events, p_{pair} is the purity of pairing, p_{JQ} is the purity of jet flavor tagging and SF_{JQ} is the scale factor accounting for difference in jet flavor tagging performance between MC and data.

The statistical and systematic uncertainties for the signal purity were calculated also using this formula. Table 5 lists each variable contributing to the P_{obs} calculation with its asigned uncertainties.

	DIL	LJ
f_{nonb}	0.078 ± 0.002	0.077 ± 0.001
SF_{nonb}	1.05 ± 0.05	1.05 ± 0.05
p_{nonb}	0.5 ± 0.01	0.5 ± 0.01
p_{pair}	$0.930 \pm 0.002(stat) \pm 0.015(sys)$	$0.831 \pm 0.001(stat) \pm 0.028(sys)$
p_{JQ}	$0.604 \pm 0.004(stat) \pm 0.010(sys)$	$0.607 \pm 0.002(stat) \pm 0.004(sys)$
SF_{JQ}	$1.01 \pm 0.01(stat) \pm 0.02(sys)$	$1.01 \pm 0.01(stat) \pm 0.02(sys)$

Table 5: Variables used for calculation of signal purity

Finaly, after all the pieces fit togheter, the total amount of background and signal events for both channels can be calculated considering two measurements of b-jet charge per event. The total purities are calculated from the values for each channel, where the amount of each background is taken into account. These values, which are summarized in Table 6 are used as input to the statistical treatment of the analysis.

N_s	171.07 ± 25.66
N_b	8.23 ± 3.55
p_s	$0.572 \pm 0.003(stat) \pm 0.008(sys)$
p_b	0.505 ± 0.005

Table 6: Expected total number of signal and background events and their total purities.

3.6 Statistical treatment

The statistical treatment used for top quark charge measurement follows the method outlined in [13]. The frequentist approach is chosen. The measurement contains several nuisance parameters that parametrize the size of signal and background and their purities, therefor the profile likelihood method was adopted.

3.6.1 Profile likelihood method

The basic assumption is that there is a hypothesis or probability model for the experimental data which depends on a set of parameters of interest $\pi = (\pi_1, \cdots \pi_k)$ and also on a set of additional nuisance parameters $\theta = (\theta_1, \cdots \theta_l)$. If the probability density function for the model is denoted by $f(x|\pi, \theta)$ and there is a set of independent measurements $X = (X_1, \cdots X_n)$ then the likelihood function can be writen as

$$L(\pi, \theta|X) = \prod_i f(X_i|\pi, \theta) \tag{4}$$

A standard technique for constructing confidence intervals is to find a corresponding hypothesis test. The hypothesis test can be expressed as verification of the hypothesis formulated as: $\pi = \pi_0$ versus the hypothesis $\pi \neq \pi_0$ and a test can be based on the profile likelihood ratio:

$$\lambda(\pi_0|X) = \frac{max(L(\pi_0, \theta|X); \theta)}{max(L(\pi, \theta|X); \pi, \theta)} \tag{5}$$

The maximum in the denominator is found over the full space of parameters (over all values of the nusisance parameters and parameters of interest), while the maximum in the numerator is found only over the subspace with $\pi = \pi_0$. Now the ratio λ is a funcion of π_0 and the data X only and not of the nuisance parameters. Acording to the theory $-2\log\lambda$ has an approximate χ^2 distribution with k degrees of freedom.

In the particular case of the top quark charge analysis there is only one parameter of interest $\pi = f_+$ – fracton of events with the signature of the SM hypothesis. The end result of analysis will be a χ^2 distribution as a function of f_+. In the analysis there are four nuisance parameters:

- N_s - expected number of signal events after lepton - jet pairing and flavor tagging
- N_b - expected number of all the background events added together after paring and flavor tagging

- p_s - expected purity of the pairing and flavor tagging methods on the signal
- p_b - expected charge asymmetery for all the backgrounds combined together

These parameters are known with certain uncertainties and this fact should be taken into account at statistical treatment – in particular at finding the profile likelihood 5 the values of nuisance parameters are generated using the corresponding Nuisance parameter distributions.

3.6.2 The analysis deliverables

An importand part of any statistical treatment is a definiton of what the analysis is trying to answer. In case of the top quark charge measurement, the sought answer is if the data is composed of SM top like events or XM like events. There is a need to make sure that the statistical treatment is saying whether the analysis has enough sensitivity to exclude either hypotesis.

The practical goal is to decide which of the two hypothesis (the SM hypothesis versus XM one), treated exclusively, is true. The test static of the analysis, i.e. the random variable used to make this decision, is variable f_+. Taking the SM hypothesis as the null hypothesis of the analysis, its acceptance and rejection regions need to be found, expressed in terms of the test static intervals.

The decission on which of the two hypothesis is true can be expressed in term of significance level (α) and power of the test ($1 - \beta$).

- $\alpha = P(rejectSM|SM)$ – probability of rejecting SM hypothesis, if SM hypothesis is true.
- $\beta = P(acceptSM|XM)$ – probability of accepting SM hypothesis, if XM hypothesis is true.

  The value of $1 - \beta$ is thus called the power of the test to discriminate against the XM hypothesis.

P-value.

One of the basic terms used in statistical testing of a hypothesis is the p-value. It express the probability of obtaining a result at least as extreme as a given data point, provided that the null hypothesis is true

As for any data pair (x^+, x^-) (the number of events with SM and XM signature) the value of the used test static $(f_+)_{rec}$ can be unambiguously reconstructed. Then the p-value can be expressed in terms of reconstructed f_+ instead of (x^+, x^-).

If the $(f_+)_{observ}$ is the value of thest static corresponding to the observed data $(x^+, x^-)_{observ}$ then the p-value can be expressed as:

$$p - value = \int_{-\infty}^{(f_+)_{observ}} g(f_+|SM) \tag{6}$$

As only values of $f_+ < 1$ are taken into account, the so called one-tail p-value is used. The ultimate goal of the analysis is to calculate the p-value for the experimentaly observed pair (x^+, x^-).

With the distributions of f_+ for XM and SM, the pseudo-experiments generated accordingly to the XM can be used, and for each one of them the area under the SM distribution is looked at. The distribution of p-values according to the SM distribution is obtained. After this, a value for α is chosen - the probability of wrongly rejecting SM. The alpha value is wanted to be small, for analysis it was chosen to be 1%. Then, for the chosen value of $\alpha = 0.01$ the value of β is calculated.

Once the value of f_+ for the actual data is obtained, its p-value using the SM distribution is extracted. If the p-value is greater than 0.01, it can be said that the analysis excludes the XM at the confidence level equal to $1 - \beta$ already determined from sensitivity study. If the p-value is less than 0.01, the analysis excludes the SM at the 99% confidence level.

Bayes factor.

Another observable is to look at how likely the SM hypothesis is compared to XM and build the likelihood ratio. In this bayesian approach the systematic uncertainties are integrated over. No minimization is done. The likelihoods at $f_+ = 0$ and $f_+ = 1$ are evaluated and integrated over the nuisance parameters. Then the ratio, called the Bayes Factor (BF), is taken. Looking at the SM and exotic hypotheses as mutually exclusive (either SM or exotics is true) the Bayes factor for these hypotheses can be calculated as follows:

$$F_{Bayes} = \frac{P(x^+, x^-|f_+ = 1)}{P(x^+, x^-|f_+ = 0)} \tag{7}$$

where $P = L_s \cdot G_{N_s} G_{N_b} G_{p_s} G_{p_b}$ (G_X is Gaussian distribution of the nuisance parameter X).

The value obtained can be interpreted as how likely the SM (XM) is, compared to the XM (SM). By taking $2\ln(BF)$ the number similar to a χ^2

is obtained and compared to following scale:

- 0-2: not worth mentioning,
- 2-6: positive evidence,
- 6-10: strong evidence,
- 10-: very strong evidence.

Figure 5 shows dhe distribution of p-values according to SM when XM hypothesis is true. The area under the p-value curve at $\alpha = 0.01$ is $1 - \beta = 0.87$ which is the sensitivity of the analysis.

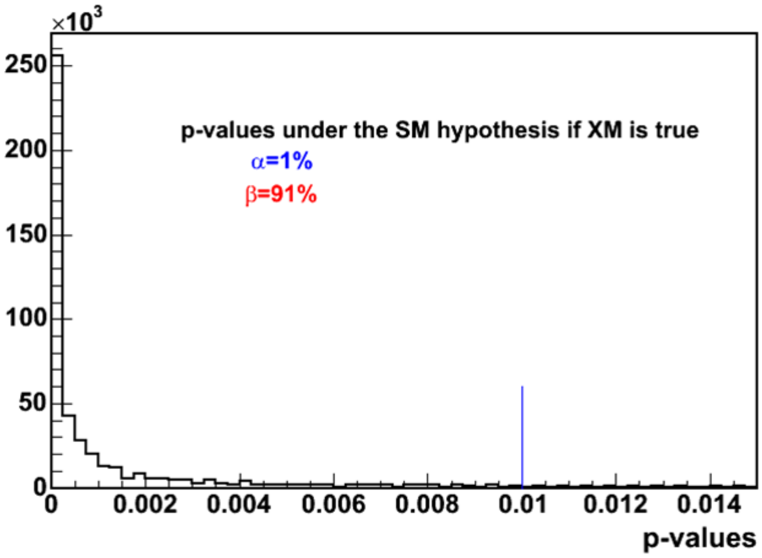


Figure 5: p-value distribution according to the SM if XM is true.

3.7 Results

After finishing all the optimization and calibration studies the experimental data from CDF detector were analyzed and a the statistical significance of observed results was estimated.

In the table 7 the final numbers of observed pairs for dilepton and lepton + jets channels are listed. In dilepton channel, after applying all the selection criteria, there are 26 lepton - b-jet pairs left. 13 of them look like SM and

13 look like XM. In lepton + jets the number of lepton - b-jet pairs after selection criteria have been applied is 199, 111 pairs are SM like and 88 are XM like. Overall there are 225 pairs, from which 124 look like SM and 101 look like XM.

Figure 6 shows the charge distribution of $Q_W * Q_b$ for the MC and experimental data events. In case of SM like pairs, the b-jet charge is anticorrelated with the W charge, in case of XM like pairs, the charges are correlated.

Figure 7 shows the profile likelihood function calculated for actual amount of expected signal and background and final purities. For observed number of SM and XM like pairs in experimental data, the minimum is found at the value of $f_+ = 0.87$. The corresponding p-value under SM hypothesis is $p = 0.31$. As this value is greater than the value of $\alpha = 0.01$ the XM hypothesis is excluded with the confidence of $1 - \beta = 87\%$.

Figure 8 shows the probability distributions for SM and XM hypothesis as the function of f_+. The observed value of $f_+ = 0.87$ is shown.

Using the bayesian approach, the Bayes factor is calculated using the final observed numbers. From the value of $2\log(BF) = 12.01$ using the scale mentioned in previous section, it can be said that the experimental data *very strongly favors* standard model over exotic one.

Yield	**Observed**	**After pairing**	**JQ defined**	**SM**	**XM**
L + J	193	102	199 pairs	111	88
DIL	44	14	26 pairs	13	13
Total	237	116	225 pairs	124	101

Table 7: The final number of events obtained from data

4 Final Summary

This thesis has presented the first CDF top quark charge measurement. The goal was to decide if the top quark discovered on the Tevatron experiments is really the particle predicted by SM with the charge 2/3 or if it is, as suggested by some alternative theories, an exotic quark with the charge -4/3.

The basic idea of the measurement was based on the analysis of the electric charges of the $t\bar{t}$ decay products ($t \rightarrow Wb$). The procedure consists of the following basic steps:

- correct association of W-boson and b-quark from the same top decay (lepton - b-jet invariant mass criteria for the dilepton and kinematic fit for the lepton + jets channel);

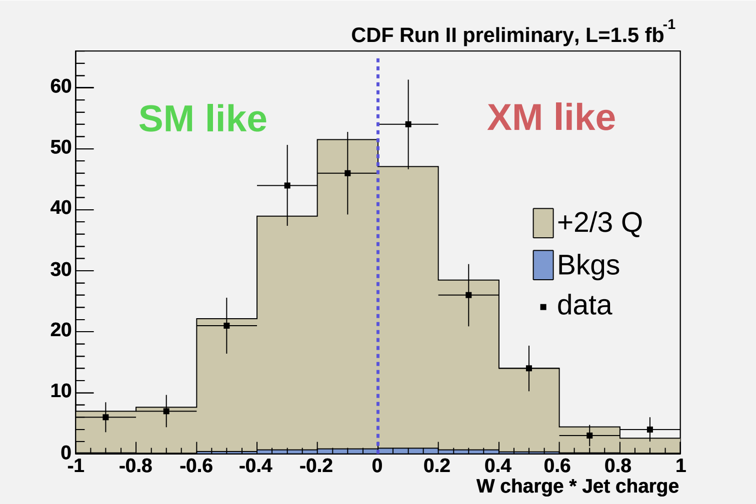


Figure 6: W-charge * Jet charge for $t\bar{t}$ events

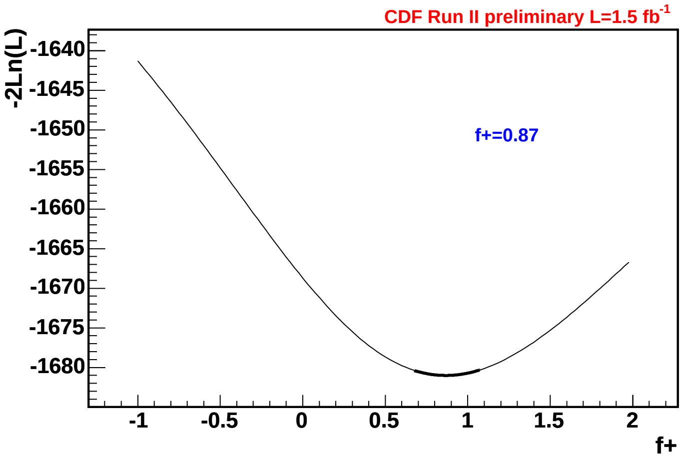


Figure 7: Profile Likelihood function for f_+. For the observed number of SM and XM like pairs $(124, 101)$, the profile likelihood function was calculated for selected values of f_+ covering the whole range of possible f_+ values. The minimum was found for the $f_+ = 0.87$.

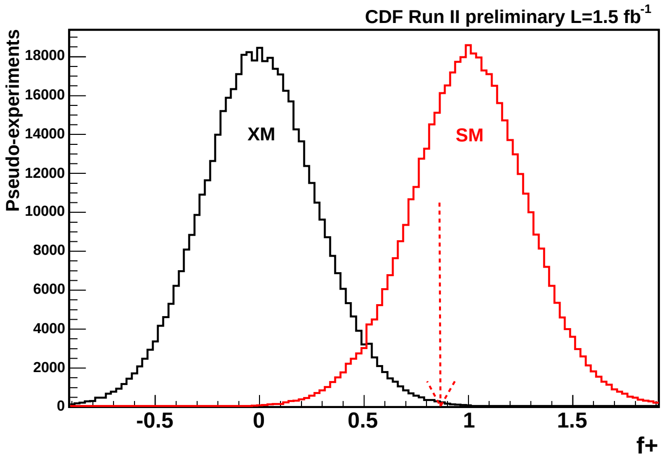


Figure 8: Probability distribution for XM and SM hypothesis. The value of $f_+ = 0.87$ is shown, for which the minimum of the profile likelihood function was found. The area under SM curve up to the this value of f_+ is the $p - value = 0.31$.

- determination of W-boson charge (the sign of the lepton from leptonic decay);
- determination of the charge of b-jet using weighting method.

As the final output of the procedure there are lepton - b-jet pairs with corresponding b-jet charges. Under the SM hypothesis, the charges of associated lepton and b-quark should have the opposite signs while under the XM hypothesis, the signs of the charges should be the same.

In case of the b-jets, the spectrum of charges evaluated by the weighting procedure has a wide spread and the anticorrelation (correlation) of the paired lepton and b-jet's charges in SM (XM) can be violated, but it should be valid for the mean value of the charge spectrum of the paired jet.

After applying the whole procedure, there are 225 pairs, out of which 124 look like SM and 101 look like exotic model.

After the statistical treatment, taking into account all the sensitivity and systematic studies, the result is that the data exclude XM with confidence level of 87%. Using Bayesian approach, it can be said that the experimental data favors strongly SM over XM.

To improve this results, different methods of pairing and b-jet charge determination can be investigated. For pairing one could combine the information from the Top Mass Fitter with the lepton - b-jet invariant mass calculation, to increase the purity of the pairing. For b-jet charge determination, the template method, neural network or semileptonic B-meson decay could be used. However all these methods require higher statistics of experimental data than it is available now. Some of them would become possible when the CDF collects the planed $4-8$ fb$^-1$ amount of data. The others could be applied in the future experiments on LHC, where the amount of $t\bar{t}$ events will be much higher.

References

[1] Abe et al.(CDF collaboration),Phys.Rev.Lett. 74, 2626(1995).

[2] S. Abachi et al.(D0 collaboration),Phys.Rev.Lett. 74, 2632(1995).

[3] A. Denner and T. Sack, Z. Phys. C 46 (1990) 653.

[4] G. Chiarelli, Int. J. Mod. Phys. A 13 (1998) 2883.

[5] T. Affolder et al. (CDF Collaboration), Phys. Rev. D63 (2001) 032003.

[6] M. Ciljak, M. Jurcovicova, S. Tokar, U. Bauer, *Top charge measurement at ATLAS detector*, ATLAS Note, PHYS-2003-35 (2003).

[7] D. Chang, W.F.Chang, E. Ma, Phys.Rev.D59, 091503(1999) and Phys.Rev.D61, 037301(2000).

[8] T. Maki, J. Antos, A. Beretvas, Y.C. Chen, R. Lysak, *Top Mass Measurement id Dileptonic Channel using Template Method in 1.2 fb*$^{-1}$, CDF Note 8727 (2007).

[9] Franklin, Grinstein, Guimaraes da Costa, Sherman, *Measurement of the Top Pair Cross Section in Lepton+Jets in 1.12 fb*$^{-1}$, CDF Note 8767 (2007).

[10] D. Acosta et al. (CDF Collaboration), Pys. Rev D71, 052003 (2005).

[11] T. Wright, *SecVtx B-tag Efficiency using Muon Transverse Momentum*, CDF Note 7448 (2005).

[12] J. Guimaraes da Costa, D. Sherman, *Measurement of the Top Pairl Production Cross Section in 700 pb*$^{-1}$ *with Loose SecVtx*, CDF Note 8120 (2006).

[13] W. A. Rolke, A. M. Lopez, J. Conrad, Nucl. Instrum.Meth. A551 (2005) 493-503.