CDF Note 10979

Measurement of Single Top Production Cross Section in $\not\!\!E_T$ plus Jets Sample with the Full CDF Run II Data Set

The CDF Collaboration
URL http://www-cdf.fnal.gov
(Dated: March 7, 2013)

Top quarks are produced mostly in pairs at the Tevatron through the strong force. However, the electroweak force also allows the production of a single top quark with a cross section that is roughly half the $t\bar{t}$ one. In addition to the lower rate, the less distintinctive signature makes this process much harder to be separated from background. In the past, Tevatron experiments have always been looking for single top production in events where one high energy electron or muon is identified, as expected in the top leptonic decay channel $t \to Wb \to l\nu b$, in order to improve the signal signature over background. We present here a measurement of single top production cross section selecting events consistent with W+jets topology but where no electron or muon has been identifieded, and where the tau lepton in the $t \to Wb \to \tau\nu b$ channel is reconstructed as a jet in the calorimeters. Multivariate analysis techniques are necessary to suppress the large background and to discriminate the single top signal. We use the likelihood profile of this discriminant to estimate an expected production cross section $\sigma^{s+t}_{exp} = 3.2^{+1.4}_{-1.4}$ pb. In the CDF full data set we measure $\sigma^{s+t}_{obs} = 3.0^{+1.5}_{-1.4}$ pb.

Preliminary Results for Winter 2013 Conferences

I. INTRODUCTION

We analyzed electroweak single top production using the full CDF Run II data set collected up to the end of the Tevatron run in September 2011, corresponding to a total integrated luminosity of 9.1 fb^{-1}. The cross section being so small and the processes that mimic the signal so large, CDF and D0 have not been able to measure the cross section of this interesting process with full-prove significance. In order to increase the scanty statistics, we look here at events rejected by previous analyses, i.e. events where there are no identified leptons, or where s are reconstructed as a jet. We thus rely solely on the signature of high P$_T$ jets and large missing transverse energy. Being statistically independent of the lepton+jets sample, this sample provides, albeit with low precision, an independent measurement of the single top production cross section. Eventually, the result of this measurement can be combined with the lepton+jets one and help reaching a solid single top cross section measurement at the Tevatron.

Many Standard Model processes can produce a final state with large missing tranverse energy and jets, such as single top (our signal), top pair production, W/Z + jets, diboson production. In addition, QCD multijet production can mimic this signature due to severely mismeasured jets which appear to have large $\not\!\!E_T$. Since the QCD heavy flavour production cross-section is orders of magnitude higher than that of the signal, it constitutes the biggest background in this search. Additionally light flavour jets can be falsely identified as b-jets (commonly referred to as "mistags").

Having no identified leptons in the final state, backgrounds are many orders of magnitude larger than the signal even after requiring large missing transverse energy and b-tagged jets in the final state. It is thus necessary to develop an event selection which reduces backgrounds to a more manageable size before trying to build a discriminant to measure the single top cross section. The QCD multijet production is the only background where the $\not\!\!E_T$ is mainly instrumental. In a first step we study the kinematics of these events and implement a multivariate technique to cut them out as much as possible.
We develop in a second step a discriminant to act against $t\bar{t}$ background. Top pair production gives a large contribution to total background, especially when both the leading jets in the final states are required to be b-tagged. In addition, the production of real top quarks makes this background more signal like if we take into account variables connected to the top mass, that are instead particularly useful in separating single top from all other non-top backgrounds.
In the final step, we use again a machine learning technique to discriminate the signal from the surviving backgrounds, and scan its output distribution to measure the single top production cross section in the $\not\!\!E_T$+jets final state.

II. EVENT PRE-SELECTION

Since we require large missing transverse energy and no leptons, we use a set of loose lepton identification cuts to make sure that we reject all events with true leptons. To impose track isolation we use a slightly modified version of the commonly used CMIO muon identification cuts [1]. We constructed these cuts to be loose enough to make sure that we exclude the single top signal already considered in the channel with identified leptons.
Jets are reconstructed off line using the JetClu algorithm with a cone in the $\eta-\phi$ space of radius R=0.4. In this analysis we require $\not\!\!E_T > 50$ GeV and $E_T\ (j_1) > 35$ GeV, $E_T\ (j_2) > 25$ GeV. We then apply the trigger parameterization described in [7] to our Monte Carlo simulations. It was observed that the trigger efficiency depends also on the difference in R-space between the two jets. We preserve this choice which maintains the trigger fully efficient: ΔR $>$ 1 (to avoid cluster merging at L2, which would result in a loss of efficiency). The two leading jets are required to be not very forward: $|\eta|(j_i) < 2$, while the leading jet is required to be very central to match the trigger requirements $|\eta|(j_1) < 0.9$. We accept events with 2 or 3 tight jets. In order to improve the signal-to-background ratio further, we need to identify jets originating from a b-quark. We do so employing both the SECVTX and JetProb b-tagging algorithms. We subdivide the sample into three orthogonal tagging categories (S stands for tight SECVTX tag, J stands for JetProb <5%):

- 1S: only one of the 2 leading jets is tagged by SecVTX
- SJ: one of the 2 leading jets is tagged by SecVTX, and the other one is not tagged by SecVTX but by JetProb
- SS: both of the leading jets are tagged by SecVTX.

Most data at this analysis stage are composed of QCD production of two or three jets, where one of the jets is poorly measured, resulting in a large transverse energy imbalance. Due to this mismeasurement, most of the time the poorly measured jet will be the second highest E_T jet; the $\not\!\!E_T$ will as a consequence be aligned to it in the transverse plane. To define the pre-selection 1 region we require $\Delta\phi(\not\!\!E_T, j_2) > 0.4$. We summarize the preselection cuts in the following list;

- lepton veto = use loose identification cuts to reject events with isolated leptons
- $\not\!\!E_T > 50$ GeV
- Number of jets = 2 or 3 and one of the leading jets (j1 or j2) central ($|\eta| < 0.9$). Events with a larger number of jets are rejected
- ΔR$(j_1, j_2) > 1$
- $E_T(j_1) > 35$ GeV, $E_T(j_2) > 25$ GeV
- Events are b-tagged either 1S, SS, or SJ

III. MODELING

To model the single top dynamics and predict both shape and rate, we use a sample generated with Powheg. The official sample is actually composed of three subsamples, one s-channel sample, a LO and a NLO t-channel sample. This process is characterized by an accurately predicted cross section, used to derive its normalization.

We use the LO Pythia cross-sections scaled by a k-factor corresponding to the ratio between the NLO and LO cross-section prediction in MCFM3 to predict both shape and rate of diboson production. The LO MCFM predictions are also consistent with those from Pythia and are reported to be in good agreement with the data. To estimate the contribution of mistagged light flavor diboson events in the tag sample we apply the corresponding mistag matrix to diboson light flavor MC samples, vetoing events with a real b- or c-quark from HEPG. This process is characterized by an accurately predicted cross section, used to derive its normalization.

Top pair-production yields a significant contribution to the background in the pre-selection region, especially in the double-tag sample. Semi-leptonic top decays are energetic, bear large $\not\!\!E_T$ and high jet multiplicity. We use Monte Carlo samples generated for the Top Group to model the shape of this background. The top-antitop events were generated with Pythia with $M_{top} = 172.5$ GeV and normalized with the cross section measured by CDF: 7.71 ± 0.51 pb [3].

V+jets background was generated with Alpgen+Pythia. As for diboson production, the contribution of mistagged light flavor V+jets events was determined by applying the mistag rate matrix to W/Z+lf samples vetoing events with a real b- or c-quark from HEPG. Since the imprecise theoretical prediction for the cross section of this process, we derive its normalization from data.
Due to the large cross-sections, it is practically impossible to generate enough statistics to simulate all QCD processes. To deal with this problem a method for estimating QCD background from data was developed [2]. This technique allows us to estimate not only heavy flavor QCD production, but also processes with a light flavor jet falsely tagged as a b-quark. Additionaly it allows us to model Single SecVTX tagged data sample, which adds additional sensitivity to the analysis. We derive QCD multijet normalization from data.

IV. MULTIJET NEURAL NETWORK, QCDNN

We train an artificial neural network (NN), a multilayer perceptron (MLP) fed with 13 kinematic variables, to separate the signal from the main background: QCD multijet production. We cut on this output at QCDNN > 0.35, to form the pre-selection 2 region. This cut removes almost 65% of the QCD multijet background, while keeping most of the signal (more than 90%).

The 13 input variables are:

- missing transverse momentum, MPT
- missing transverse energy, $\not\!\!E_T$

- difference in ϕ between missing transverse energy $\not\!\!E_T$ and missing transverse momentum MPT, $\Delta\phi(\not\!\!E_T$, MPT)
- maximum difference in R-space between two jets, or all jet pairs
- minimum difference in ϕ between the $\not\!\!E_T$ and each jet
- minimum difference in ϕ between the MPT and the jets, considering all (MPT, j_i) pairings
- maximum difference in ϕ between jet directions, for all jet pairs;
- ratio of MHT (vector sum of tight jet ET) over $\not\!\!E_T$
- $\Delta\phi$ between the direction of the leading jets in jet pair rest frame and the direction of the jet pair boost
- $\not\!\!E_T$ /HT :HT being the scalar sum of the two leading jets E_T;
- $\not\!\!E_T$ significance : $\not\!\!E_T$ over the square root of sum E_T (over all calorimetric activity)
- invariant mass of $\not\!\!E_T$, j_1 and j_2
- Event sphericity : S = 1.5×($\lambda_2 + \lambda_3$), where λ_1 λ_2 and λ_3 are the second and third eigenvalues of the sphericity tensor.

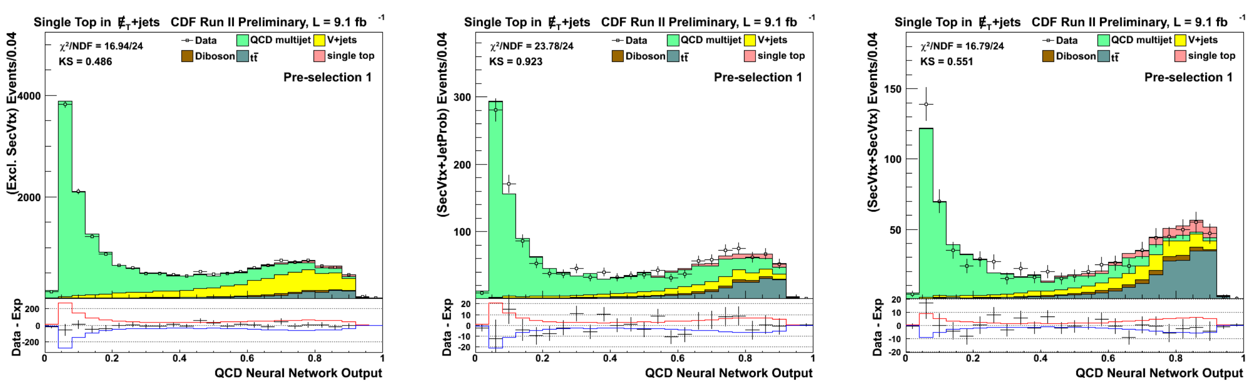

FIG. 1. QCDNN stack plots in 1S, SJ, SS regions with subtraction plots on bottom. KS/χ^2 tests take into account statistical and systematic uncertainties on signal and background processes.

V. $t\bar{t}$ NEURAL NETWORK, TTNN

We train an artificial neural network (NN), a multilayer perceptron (MLP) fed with 12 kinematic variables, to separate the signal from top pair production background. We cut on this output at TTNN > 0.3, to form the signal region. With this cut almost 50% of $t\bar{t}$ background is removed, while keeping most of the signal events that passed the QCDNN cut (more than 90%).

The 12 input variables are:

- HT, the scalar sum of the two leading jets E_T and the $\not\!\!E_T$
- Vectorial sum of the jet pt
- Invariant transverse mass of $\not\!\!E_T$ and the 2 leading jets
- Invariant mass of j_1, j_2 and j_3
- Pt of the first jet
- Aplanarity
- Number of jets in the final state

- Missing Pt
- mht, vector sum of tight jet E_T
- Maximum difference in ϕ between two jets directions, over all jet pairs
- Pt of the second jet
- $\not\!\!E_T$, missing transverse energy

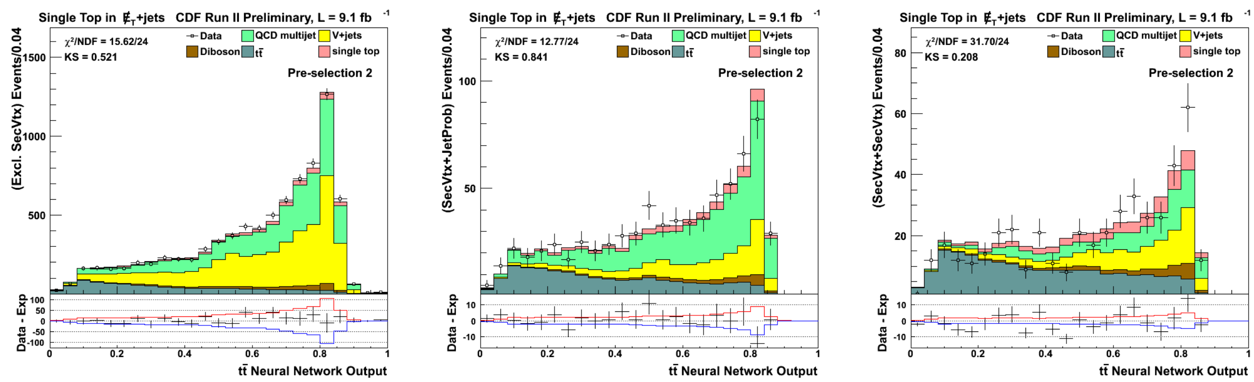

FIG. 2. TTNN stack plots in 1S, SJ, SS regions with subtraction plots on bottom. KS/χ^2 tests take into account statistical and systematic uncertainties on signal and background processes.

VI. FINAL DISCRIMINANT NEURAL NETWORK

7 variables are fed to a MLP trained with events with at least one tight tag (SecVTX) to obtain the final NN discriminant. We then use a likelihood profile of this discriminant to measure the production cross section of single top events. The 7 used variables are:

- QCDNN.singletop=the QCDNN output
- JET.pt2=the transverse momentum of the second jet
- MASS.mvj2=transverse mass of the missing transverse energy and the second jet
- MVJ123=transverse mass of the missing transverse energy and all jets
- MET.ht=sum of jets E_T
- MASS.mvj12=transverse mass of the missing transverse energy and of the two leading jets
- MET.sumjetpt=sum of the jet transverse momenta

VII. SYSTEMATIC UNCERTAINTIES

We consider several systematic uncertainties affecting this analysis. The dominant systematic sources are the uncertainties on multijet normalization (20%) and the background cross sections. We consider V+jets noemalization as unconstrained in the final fit. We also consider uncertainties from the jet energy scale (JES) (1.4 - 12.9%), the luminosity measurement [4] (5%), parton density functions (2%) [5], lepton veto (2%), trigger efficiency (0.4 - 1.5%), b-tag efficiency and ISR/FSR (2% on single top normalization). We also assign systematic uncertainties, based on the variation in the shape of the distribution of kinematic quantities. For Monte Carlo samples, we included $\pm 1\sigma$ variation of the jet energy scale as a shape uncertainty. We also vary the Q^2 scale, a parameter in the perturbative expansion used to calculate the matrix elements in the alpgen generator, to generate the shape uncertainty templates for the W/Z plus jet backgrounds. For QCD multijet events, we vary the tag-rate probability in each bin of the matrix by $\pm 1\sigma$ in predicting the QCD shape as a shape uncertainty.

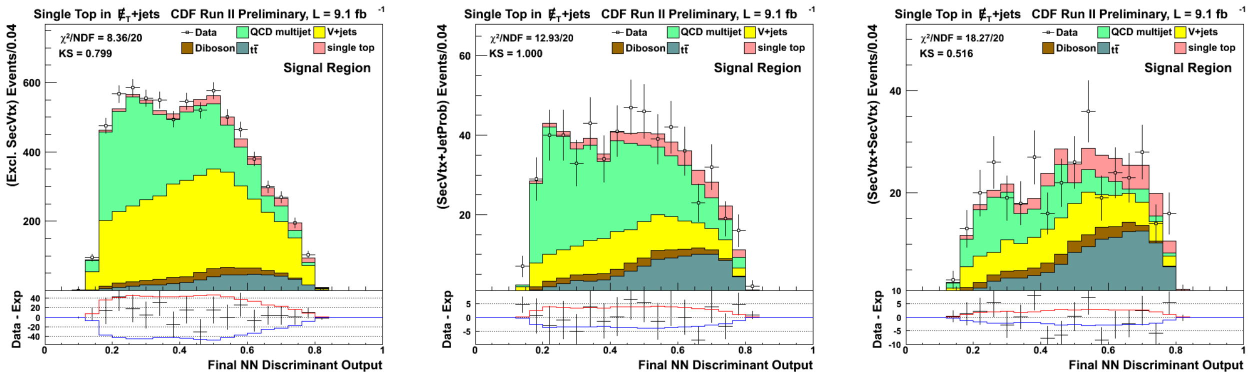

FIG. 3. SIGNN stack plots in 1S, SJ, SS regions with subtraction plots on bottom. KS/χ^2 tests take into account statistical and systematic uncertainties on signal and background processes.

Single top in $\not\!\!E_T$+jets CDF Run II Preliminary, L = 9.1 fb^{-1}			
	1S	SJ	SS
t-channel	170.0 $\pm$ 8.9	8.1 $\pm$ 0.7	8.8 $\pm$ 0.7
s-channel	94.1 $\pm$ 4.9	31.4 $\pm$ 2.6	38.4 $\pm$ 2.8
QCD multijet	2642.1 $\pm$ 191.0	284.0 $\pm$ 21.7	85.9 $\pm$ 8.1
W+jets	2425.1 $\pm$ 212.3	77.0 $\pm$ 8.5	55.7 $\pm$ 5.8
Z+jets	1009.8 $\pm$ 88.1	37.8 $\pm$ 4.1	32.6 $\pm$ 3.4
Diboson	228.5 $\pm$ 27.4	23.6 $\pm$ 3.2	23.4 $\pm$ 3.0
$t\bar{t}$	453.3 $\pm$ 40.3	94.5 $\pm$ 10.3	108.2 $\pm$ 11.1
Expected	7023.0 $\pm$ 360.0	556.4 $\pm$ 27.6	353.1 $\pm$ 17.5
Data	7186	569	351

TABLE I. Event yields in the signal region. Statistical and systematic uncertainties are reported.

VIII. RESULTS

e estimated the expected cross-section performing 20.000 pseudo-experiments with all systematics turned on (we also performed 1.000.000 samplings of the systematics space during the MC integration of the systematics in MCLIMIT). Fig. 4 shows the spread of the outcomes of those pseudo-experiments.

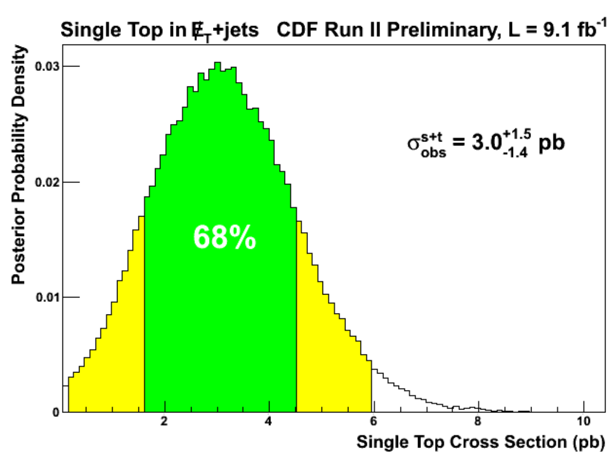

FIG. 4. Posterior probability density distribution

After accounting for statistical and systematic errors, we expect a cross-section of:

$$\sigma^{s+t}_{exp} = 3.2^{+1.4}_{-1.4} \text{ pb}$$

Once this analysis is applied to 9.1 fb^{-1} of CDF Run II data, the observed cross-section is:

$$\sigma^{s+t}_{obs} = 3.0^{+1.5}_{-1.4} \text{ pb}$$

IX. CONCLUSIONS

We have presented a search for combined s- and t-channel electroweak single top production in the $\not{E}_T$+jets channel (events where the lepton from the W decay is either not identified or reconstructed as a jet). We have analyzed 9.1 fb^{-1} of CDF Run II data and measured the single top production cross-section:

$$\sigma^{s+t}_{obs} = 3.0^{+1.5}_{-1.4} \text{ pb} \quad (\sigma^{s+t}_{exp} = 3.2^{+1.4}_{-1.4} \text{ pb})$$

ACKNOWLEDGMENTS

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium für Bildung und Forschung, Germany; the Korean World Class University Program, the National Research Foundation of Korea; the Science and Technology Facilities Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Ministerio de Ciencia e Innovación, and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; the Academy of Finland; the Australian Research Council (ARC); and the EU community Marie Curie Fellowship contract 302103.

[1] A. Apresyan, D. Bortoletto, F. Margaroli, K. Potamianos, Oscar Gonzalez, M.Vidal, Song Ming Wang, Event selection for Higgs search in the $\not{E}_T$ plus jets sample, CDF 9358, 2008

[2] Daniela Bortoletto, Qiuguang Liu, Fabrizio Margaroli, Karolos Potamianos, Background Estimation in $\not{E}_T$ + b-jets Analysis, CDF/ANAL/MISSING ET/CDFR/10452

[3] Evelyn Thomson et. al Combination of CDF and D0 ttbar cross-sections, CDF note 10916

[4] S. Jindariani, S. Klimenko, J. Konigsberg, G. Lungu, V. Necula, L. Pinera, A. Pronko, R. Rossin, A. Sukhanov, D. Tsybychev, S.M. Wang, Luminosity Uncertainty for Run 2 up until August 2004, CDF/ANAL/CDF/7446

[5] Oscar Gonzalez, Carsten Rott, Uncertainties due to the PDFs for the gluino-sbottom search, CDF/PHYS/EXOTIC/CDFR/7051

[6] Sensitivity, Exclusion and Discovery with Small Signals, Large Backgrounds, and Large Systematic Uncertainties, CDF note CDF/DOC/STATISTICS/PUBLIC/8128 (2007)

[7] O. Gonzalez and M. Vidal, Getting ready for the 4 fb-1 dataset: Trigger efficiency of the MET+JETS samples, CDF/PHYS/MISSING_ET/CDFR/9745

Appendix A: Systematic Uncertainties Table

Systematic	Region	Signal	diboson	$t\bar{t}$	V+jets	QCD
Luminosity	1S+SJ+SS	$\pm^{6\%}_{6\%}$	$\pm^{6\%}_{6\%}$	no	no	no
PDF	1S+SJ+SS	$\pm^{2\%}_{2\%}$	$\pm^{2\%}_{2\%}$	no	no	no
Lepton veto	1S+SJ+SS	$\pm^{2\%}_{2\%}$	$\pm^{2\%}_{2\%}$	no	no	no
B-tagging SecVTX	1S	$\pm^{5.2\%}_{5.2\%}$	$\pm^{5.2\%}_{5.2\%}$	$\pm^{5.2\%}_{5.2\%}$	no	no
	SJ	$\mp^{3\%}_{3\%}$	$\mp^{3\%}_{3\%}$	$\mp^{3\%}_{3\%}$	no	no
	SS	$\pm^{10.4\%}_{10.4\%}$	$\pm^{10.4\%}_{10.4\%}$	$\pm^{10.4\%}_{10.4\%}$	no	no
B-tagging JetProb	1S	$\mp^{3\%}_{3\%}$	$\mp^{3\%}_{3\%}$	$\mp^{3\%}_{3\%}$	no	no
	SJ	$\pm^{3.3\%}_{3.3\%}$	$\pm^{3.3\%}_{3.3\%}$	$\pm^{3.3\%}_{3.3\%}$	no	no
	SS	$\mp^{0\%}_{0\%}$	$\mp^{0\%}_{0\%}$	$\mp^{0\%}_{0\%}$	no	no
Cross Section	1S+SJ+SS	no	$\pm^{6\%}_{6\%}$	$\pm^{6.6\%}_{6.6\%}$	no	$\pm^{20\%}_{20\%}$
JES shape/rate	1S	yes/$\pm^{4\%}_{4\%}$	yes/$\pm^{6\%}_{6\%}$	yes/$\mp^{2\%}_{2\%}$	yes/no	no/no
	SJ	yes/$\pm^{4\%}_{4\%}$	yes/$\pm^{5\%}_{5\%}$	yes/$\pm^{1\%}_{1\%}$	yes/no	no/no
	SS	yes/$\pm^{3\%}_{3\%}$	yes/$\pm^{6\%}_{6\%}$	yes/$\pm^{1\%}_{1\%}$	yes/no	no/no
Q^2 scale	1S+SJ+SS	no	no	no	yes (only W+jets)	no
TRF	1S+SJ+SS	no	no	no	no	yes
ISR/FSR	1S+SJ+SS	$\pm^{2\%}_{2\%}$	no	no	no	no
Top mass dependence	1S+SJ+SS	yes	no	yes	no	no
Trigger efficiency	1S+SJ+SS	$\pm^{2\%}_{2\%}$	$\pm^{2\%}_{2\%}$	$\pm^{2\%}_{2\%}$	no	no

TABLE II. Systematic uncertainties for 1S, SJ and SS regions for the templates used in the fits.