Signal modeling systematic uncertainties at Tevatron

Viatcheslav Shary
for the CDF and D0 collaborations
CEA, IRFU, SPP
Centre de Saclay,
91191 Gif-sur-Yvette, France

Abstract. An overview of the signal modeling systematic uncertainties in the top quark measurements at the Tevatron is given. We discuss approaches developed by the D0 and CDF collaborations for the estimation of these systematic uncertainties. Uncertainties related to the initial and final state radiation, higher order corrections, choice of the hadronization model, color reconnection and choice of the parton distribution functions set are considered in this contribution.

1. Introduction
Since the top quark discovery at the Tevatron in 1995, the top quark measurements reach an extremely high precision. The most important example is a top quark mass measurement with a precision of 0.56 % [1]. The improvement in the uncertainty is explained by the large statistics accumulated both by D0 [2] and CDF [3] experiments and by the advance in the analysis techniques. Limitation in the precision comes mainly from the systematic uncertainties and the largest component is often the uncertainty of the top quark production simulation or simply signal modeling uncertainty. The following effects are considered when this uncertainty is estimated at the Tevatron.

- Uncertainty due to the initial and final-state radiation simulation.
- Uncertainty related to the higher-order QCD corrections.
- Uncertainty due to the color reconnection effects.
- Uncertainty due to the simulation of jet hadronization and underlying events.
- Uncertainty due to the choice of the parton distribution functions.
- Uncertainty due to the multiple interactions model.
- Uncertainty due to the b-quark jet hadronization simulation.

In the following sections we discuss in details these sources of uncertainties as well as approaches developed by D0 and CDF experiments to evaluate the size of uncertainties.

2. Simulation of the $\bar{t}t$ Production and Uncertainties Estimation
The $\bar{t}t$ production is simulated differently at D0 and CDF experiments. For many analyses CDF use PYTHIA [4] generator both for the matrix element simulation and simulation of the showering and hadronization effects together with the CTEQ5L parton density function (PDF) set [5]. D0 in most of the cases use ALPGEN [6] for the ME simulation and PYTHIA for the
showering and hadronization. CTEQ6L1 PDF set [7] is used by D0. CDF set the factorization and renormalization scale for events generator to $Q^2 = m_t^2 + p_1^2 + (p^2 + p_0^2)$, where $p_1$ is the transverse momentum characterizing the scattering process, $p_1^2, p_2^2$ are the virtualities of the incoming partons. D0 set both scales to $Q^2 = m_t^2 + \langle p_1^2 \rangle$, where $\langle p_1^2 \rangle$ is the average of the square of transverse momentum of all light partons produced in association with the $t\bar{t}$ pair. Some analysis use different generators. For example, for the recent single top analysis CDF uses the next-to-leading order (NLO) POWHEG [8] ME generator and PYTHIA for showering. The single top analysis at D0 uses the COMPHEP-based effective NLO generator SINGLETOP [9]. Simulated “underlying events” are added to each hard collision. For underlying event generation CDF uses the PYTHIA tune A [10], while D0 uses a modified version of the same tune.

### 2.1. Uncertainties due to Initial and Final State Radiation

Additional radiation in the initial state (ISR) or final state (FSR) affects the simulation of the number of jets as well as a the $t\bar{t}$ pair transverse momentum ($p_T$) simulation. Estimation of this uncertainty is based on the study made by CDF using Drell-Yan dilepton events [11]. These events produced by the quark-antiquark annihilation, as the most of $t\bar{t}$ events, but have no FSR radiation. The distribution of the mean $p_T$ of the dilepton pairs in data shows the logarithmic rise with the squared invariant mass and compatible with the variation of $\Lambda_{QCD}$ between 73 and 292 MeV and variation of the $Q^2$ scale between 0.5 and 2, after extrapolation to the $t\bar{t}$ scale. In PYTHIA the ISR and FSR radiation is controlled by the same equation, so variation of $\Lambda_{QCD}$ and $Q^2$ parameters modifies both ISR and FSR radiation in $t\bar{t}$ events simulation. In top quark mass measurement this uncertainty is found to be small 0.09% for D0 and 0.15% for CDF and overlap with other uncertainties (out-of-cone jet energy correction or higher order QCD correction) is not excluded.

Both experiments compare the reconstructed $p_T$ distribution of the $t\bar{t}$ pairs (extremely sensitive to ISR radiation) between data and MC in the context of the $t\bar{t}$ forward-backward asymmetry measurement, where it has a large effect on the measured asymmetry values. For example, D0 founds that this distribution is different between default MC@NLO simulation and data [12], while CDF comparison between POWHEG simulation and data shows that they are in a good agreement [13]. Unfortunately in both cases the reconstructed $t\bar{t}$ $p_T$ distributions are not sensitive to the parton level $p_T$, because reconstructed effect are too large and distribution is “smeared” and couldn’t be use to constrain the ISR uncertainty.

### 2.2. Higher Order Uncertainties

For most of the CDF and D0 analyses use tree level leading order MC generators, so it is necessary to estimate uncertainties related to the higher order QCD corrections. CDF estimates this uncertainties by reweighting the gluon fusion fraction in PYTHIA from the value 5% default for the $t\bar{t}$ production to the NLO estimation of the gluon fraction 20% [14]. For the top quark mass measurement this uncertainties found to be 0.02%. D0 uses different approach. It compares the ALPGEN+HERWIG [15] simulation of $t\bar{t}$ pairs with the simulation using MC@NLO+HERWIG. In top quark mass measurement this estimation leads to the uncertainty of 0.14% for D0. CDF also measure the difference in the top quark mass using the MC@NLO+HERWIG simulation, but concludes that this uncertainties overlap with the uncertainties in ISR and FSR.

### 2.3. Color Reconnection Uncertainties

“Color reconnection” refers to the process of interaction between final state partons of $t\bar{t}$ hard scattering event and color remnants of proton and antiproton. This effect has poor constrains from the experimental data, that is why both collaborations use two different versions of PYTHIA generator to estimate this effect. First is the version with angular ($Q^2$) ordering for jet showers with the A-pro underlying-event model (modified version of the tune A). Second is the ACR-pro


version. This version is identical to A-pro, but it includes the color reconnection effects in the model [16]. For the top quark mass measurement the estimated uncertainty is 0.32% for CDF and 0.16% for D0.

2.4. Hadronization and Underlying Events Uncertainties

In order to estimate hadronization and underlying events effect D0 and CDF compare two generators with different hadronization models, namely PYTHIA which uses the string “Lund model” [4] and HERWIG with hadronization based on the “cluster model” [15]. CDF compares PYTHIA with tune A for the underlying event model with HERWIG with tuned implementation of the underlying event generator JIMMY. D0 compare the set of ALPGEN events hadronized with PYTHIA to the same set of ALPGEN events, but hadronized with HERWIG. Hadronization and underlying event model uncertainty is one of the main source of the systematic uncertainty for the top quark mass measurement. CDF estimation gives 0.40% while D0 measure 0.33%. An attempt has been made to choose “the best” generator or tune using analysis of the jet shapes in QCD events [17], but it was not possible to have and additional constrains from this study to the hadronization uncertainty. In the same time this approach is revealed to be extremely useful to detect the “pathological” behaviour of the new tunes as it happened for the difference in the reconstructed top quark mass observed between \(Q^2\)-ordered and \(p_T\)-ordered models of parton showers in [16]. It has been shown in [18] that these new \(p_T\)-ordered models don’t reproduce shape of the light-quarks and gluon jets. So additional efforts were made in tuning \(p_T\)-ordered models. Further studies of jet shape using the full Tevatron statistics may be useful to constrain better hadronization systematic uncertainty directly from data. The new \(p_T\)-ordered tunes (labeled Perugia-2011) [19] need to be compare with “default” tunes used by D0 and CDF.

2.5. Uncertainties Related to Parton Distribution Functions

CDF estimates an uncertainty due to the choice of the parton distribution functions by variation of 20 eigenvectors in CTEQ6M set. Differences due to these variations are summed up in quadrature. Additionally a difference between two different set of PDFs, MRST98L [20] and the default one CTEQ5L is estimated. The largest uncertainty between these two estimations is added in quadrature to the variation due to the different \(\Lambda_{QCD}\) value in MRST PDF set. D0 estimates this uncertainty by reweighting PYTHIA simulation using the quadratic sum of 20 eigenvectors variations in CTEQ6M. In the top quark mass measurement this uncertainty is estimated to be 0.08% for CDF and 0.14% for D0.

2.6. Multiple Interaction Model Uncertainties

At the Tevatron D0 and CDF experiments have a significant probability that additional \(p\bar{p}\) interactions happen simultaneously with the hard scattered one and a typical average value of number of primary vertices is around 2. To simulate this effect CDF overlays the Poisson-distributed events simulated with PYTHIA. D0 overlays the real data events registered with “zero-bias” trigger. The mean value of number of multiple interactions increases with the instantaneous luminosity increase. This mean value chosen at the moment of MC events generation may be different (usually smaller) than the mean value of the analysis data set. Reweighting of the simulation distribution of number of primary vertices to the distribution observed in data, allows to estimate the uncertainty due to the imperfect simulation of this effect. Such approach is adopted by the CDF. D0 uses the similar approach, but instead of number of primary vertices distribution, D0 reweights the the instantaneous luminosity spectrum. Usually, uncertainty due to the multiple interaction model is small. For example, in the top quark mass measurement this uncertainty estimated to be 0.06% for CDF and 0.03% for D0.
2.7. b-quark Jet Hadronization Uncertainty

CDF estimates the uncertainty due to the b-quark jets simulation by varying the parameters of the default PYTHIA b-quark jets fragmentation based on the Bowler model \cite{21}, where the Bowler fragmentation-function parameter \( r_q = 1.0 \) and Lund fragmentation function parameters \( a = 0.3, b = 0.58 \). The SLD parametrization \( (r_q = 0.980 \pm 0.010, a = 1.30 \pm 0.09, b = 1.58 \pm 0.09) \) and LEP parametrization \( (r_q = 0.897 \pm 0.013, a = 1.03 \pm 0.08, b = 1.31 \pm 0.08) \) \cite{22} is used to estimate the uncertainty. D0 uses the default parametrization from LEP and compares to the simulation with the SLD parametrization. The resulting uncertainties are usually small. For example, for the top quark mass measurements they found to be 0.09% for CDF and 0.03% for D0.

3. Conclusion

In conclusion, we want to remark that signal modeling uncertainties play an important role in the precise top quark measurement. The estimation of these uncertainties often based on the comparison between different generators or generator tunes. The validity of such comparison need to be verified with data when possible. This is especially valid for the case of the new Perugia 2011 \( p_T \)-ordered PYTHIA tunes, where the bias in the reconstructed top quark mass need to be compared with the bias observed in tunes traditionally used by the D0 and CDF. Further improvement of the signal modeling uncertainties estimation could be achieved by using the new tunes for the better estimation of the modeling effects (for example, new PYTHIA tunes for the color reconnection effects). Systematic uncertainties should be constrained with Tevatron data, for example, ISR/FSR estimation could be improved by using the full statistics of Drel-Yan events, the jet hadronization uncertainties could be constrained by the jet shape studies. Additionally, the overlap between different systematic uncertainties need to be studied and reduced when possible. As an example, the higher order QCD uncertainties estimation may overlap with ISR/FSR, ISR/FSR uncertainty estimation may overlap with out-of-cone jet energy scale systematic uncertainty.

Those improvements are still possible and need to be implemented in the final set of the Tevatron measurements, in particular in the top quark mass and cross section measurements, in studies of the top quark properties: forward-backward charge asymmetry, spin correlation, etc.

References

\[1\] Aaltonen T et al. (CDF and D0 Collaborations) 2012 Phys. Rev. D 86 092003
\[2\] Abazov V et al. (D0 Collaboration) 2006 Nucl.Instrum.Meth. A565 463–537 (Preprint physics/0507191)
\[3\] Acosta D et al. (CDF Collaboration) 2005 Phys. Rev. D 71 052003 (Preprint hep-ex/0410041)
\[4\] Sjostrand T, Mrenna S and Skands P Z 2006 JHEP 0605 026 (Preprint hep-ph/0603175)
\[5\] Lai H et al. (CTEQ Collaboration) 2000 Eur.Phys.J. C12 375–392 (Preprint hep-ph/9903282)
\[6\] Mangano M L et al. 2003 JHEP 0307 001 (Preprint hep-ph/0206293)
\[7\] Pumplin J, Stump D, Huston J, Lai H, Nadolsky P M et al. 2002 JHEP 0207 012 (Preprint hep-ph/0201195)
\[8\] Alioli S, Nason P, Oleari C and Re E 2009 JHEP 0909 111 (Preprint 0907.4076)
\[9\] Boos E, Bunichev V, Dudko L, Savrin V and Sherstnev A 2006 Phys.Atom.Nucl. 69 1317–1329
\[10\] Affolder T et al. (CDF Collaboration) 2002 Phys.Rev. D65 092002
\[11\] Abulencia A et al. (CDF Collaboration) 2006 Phys.Rev. D73 032003 (Preprint hep-ex/0510048)
\[12\] Abazov V M et al. (D0 Collaboration) 2011 Phys.Rev. D84 112005 (Preprint 1107.4995)
\[13\] The CDF Collaboration 2005 Preprint CDF note 10807
\[14\] Aaltonen T et al. (CDF Collaboration) 2009 Phys.Rev. D79 031101 (Preprint 0807.4262)
\[15\] Corcella G et al. 2001 JHEP 01 010 (Preprint hep-ph/0011363)
\[16\] Skands P Z and Wicke D 2007 Eur.Phys.J. C52 133–140 (Preprint hep-ph/0703081)
\[17\] Acosta D et al. (CDF Collaboration) 2005 Phys.Rev. D71 112002 (Preprint hep-ex/0505013)
\[18\] Galtier A B March 2010 MC@LHC workshop, URL http://indico.cern.ch/conferenceDisplay.py?confId=74601
\[19\] Skands P Z 2010 Phys.Rev. D82 074018 for the description of Perugia 2011 tunes, see archive version. (Preprint 1005.3457)
[20] Martin A D, Roberts R, Stirling W J and Thorne R 1998 *Eur.Phys.J.* C4 463–496 mRST98 (*Preprint* hep-ph/9803445)

[21] Bowler M 1981 *Z.Phys.* C11 169

[22] Peters Y, Hamacher K and Wicke D (D0 Collaboration) 2006