Romanian Reports in Physics, Vol. 64, No. 4, P. 945–956, 2012

INITIAL AND FINAL STATE RADIATION STUDIES FOR TOP QUARK MASS RECONSTRUCTION IN DILEPTONIC CHANNEL FOR $\sqrt{s} = 7$ TeV P-P COLLISIONS

VALENTINA TUDORACHE, MIHAI CUCIUC

"Horia Hulubei" National Institute for Nuclear Physics and Engineering, Particle Physics Department
P.O.Box MG-6, RO-077125, Bucharest-Magurele, Romania
E-mail: vtudorache@ifin.nipne.ro, E-mail: constantin.mihai.cuciuc@cern.ch

Received March 19, 2012

Abstract. The influence of initial state radiation and final state radiation on the top quark mass reconstruction in the dileptonic channel is investigated. To reconstruct the top mass the m_{T2} variable is used. Studies are performed on simulated samples produced with two Monte Carlo generators, PYTHIA 8 and POWHEG, and the detector response is modelled using the DELPHES framework. Quantitative results on Initial State Radiation, Final State Radiation and Leading Order/Next to Leading Order comparative analysis are shown.

Key words: top quark, mass reconstruction, Monte Carlo generators, ISR, FSR, likelihood method.

1. INTRODUCTION

The top quark, the most massive elementary particle, has been observed at Fermilab in 1995 with a mass approximately 35 times larger than the mass of its isospin partner, the bottom quark. The discovery of this particle completed the third family of elementary particles predicted by the Standard Model (SM). More knowledge about the top mass will provide constraints on the Higgs boson mass, the last missing piece of the SM. Due to its large mass the top quark decays in a short time without forming hadrons. In the SM top quark decays exclusively into a bottom quark and a W boson.

The topology of top quark pair events produced in proton-proton collisions $pp \to t\bar{t} \to bW^+\bar{b}W^+$ depends on the W decay which can be leptonic $W \to l\bar{\nu}$ or hadronic $W \to q\bar{q}'$. The top-antitop events are divided in three channels [3]:

- the fully leptonic (dileptonic) channel, representing 1/9 of the $t\bar{t}$ events; both W bosons are decaying into a lepton-neutrino pair, resulting in an event with two charged leptons, two neutrinos and two b quarks;

- the fully hadronic channel, representing 4/9 of the $t\bar{t}$ events; both W are decaying hadronically which gives two b quarks and four light quarks from W decay;
- the semi-leptonic channel, representing 4/9 of the $t\bar{t}$ events; one W is decaying leptonically and the other W is decaying hadronically.

In this study we will investigate only the dileptonic decay channel. This channel has two undetectable (missing) neutrinos making the reconstruction of the top mass very difficult. The m_{T2} method designed for channels with two missing particles is used to reconstruct the top quark mass [1, 2].

Top quark pairs are produced by proton-proton collisions through gluon-gluon and quark-quark scattering (Fig. 1). The Large Hadron Collider (LHC) will produce millions of $t\bar{t}$ pairs in a sample of 10 fb^{-1} for a cross section of 833 pb at Next to Leading Order (NLO) calculations [3].

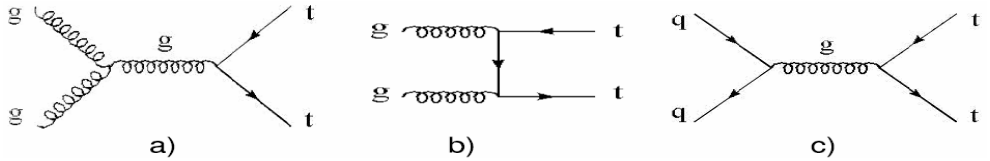

Fig. 1 – Top pair production mechanisms at lowest order: gluon-gluon scattering diagrams (a), (b) and quark-quark scattering diagram (c).

Heavy particles production, as is the case of the top quark, is associated with multiple quark and gluon emissions from incoming partons, called initial state radiation (ISR), and from outgoing partons, called final state radiation (FSR). Because their production and decay processes involve many coloured particles, the top quark mass reconstruction becomes rather difficult.

The $t\bar{t}$ samples were generated with PYTHIA 8 [4] and POWHEG [5]. Switching on and off both ISR and FSR, a comparative analysis between these samples was performed. In order to reconstruct the top mass we will use the end point fit and the likelihood binned fit.

2. THE MT2 VARIABLE FOR TOP MASS

To determine the top quark mass we are using the Cambridge m_{T2} variable, which is not quite a variable due the missing particle dependence mass. The definition for m_{T2} corresponds much more to a function [1, 2].

The m_{T2} variable for each event is given by the formula:

$$m_{T2}\left(m_{inv}\right)=\min_{p_T^{inv(1)}+p_T^{inv(2)}=p_T^{miss}}\left\{\max\left[m_T\left(m_{inv};\ p_T^{inv(1)}\right),\ m_T\left(m_{inv};\ p_T^{inv(2)}\right)\right]\right\}, \quad (1)$$

where the m_T represents the transverse mass of the system, defined:

$$m_T^2\left(m_{inv};p_T^{inv}\right)=m_{vis}^2+m_{inv}^2+2\left(E_T^{vis}E_T^{inv}-P_T^{vis}P_T^{inv}\right), \quad (2)$$

where m_{inv} and p_T^{inv} represent the undetectable mass and transverse momentum respectively, while m_{vis} is the invariant mass of the visible system which in our case is composed of a bottom quark and a lepton. The transverse energy of the visible system and the neutrino are given by the formulae:

$$E_T^{vis}=\sqrt{(|p_T^{vis}|^2+m_{vis}^2)}\ \ E_T^{inv}=\sqrt{(|p_T^{vis}|^2+m_{vis}^2)}. \quad (3)$$

The m_{T2} variable is defined event by event and depends not only on the kinematic variables m_{vis} and p_T, but also on the undetectable mass. To solve equation (1) we have to assign a trial mass to the undetected particle. While the total missing momentum can be determined by experimental measurements, it is also necessary to impose some restrictions on the undetected particle.

The kinematic constraints are given by the formula $p_T^{inv(1)}+p_T^{inv(2)}=p_T^{missing}$.

Furthermore, we introduce the m_{T2} variable for our system $t\to bl\nu$ replacing the $m_{inv,}\ p_T^{inv}, E_T^{inv}$ with $m_{\nu,}\ p_T^{\nu}, E_T^{\nu}$ and $m_{vis,}\ p_T^{vis}, E_T^{vis}$ with $m_{bl,}\ p_T^{bl}, E_T^{bl}$ obtaining the transverse mass:

$$m_T^2=m_{bl}^2+m_\nu^2+2(E_T^{bl}E_T^{\nu}-p_T^{bl}p_T^{\nu}). \quad (4)$$

The direction of the beam is chosen along z axis, so $P_T=(P_x,P_y)$, the transverse energy is:

$$E_T^{bl}=\sqrt{(|p_T^{bl}|^2+m_{bl}^2)}\ \ E_T^{\nu}=\sqrt{(|p_T^{bl}|^2+m_\nu^2)}. \quad (5)$$

In this case the system is composed of a b-l pair and a neutrino particle. Due to the lack of total momentum measurement in the transversal plane, the longitudinal momentum of the two neutrinos can be chosen arbitrarily, as we point out below.

$$p_T^{\nu(1)}+p_T^{\nu(2)}=p_T^{missing}. \quad (6)$$

The m_{T2} variable depends on the trial neutrino mass, therefore fixing the value of m_ν we can obtain a distribution of the m_{T2} variable whose end point represents the mass of top quark computed as [1, 2]:

$$m_t=m_{T2}^{max}(m_\nu)=\max(m_{T2}) \quad (7)$$

In the assumption that the neutrino mass is zero, we could compute m_{T2} using eq. (1) and choose the smallest value of the m_{T2} variable for each event. The results of this method are shown and discussed in the next sections.

3. SIMULATION FRAMEWORK

In this section we are briefly presenting the strategy of PYTHIA 8 and POWHEG Monte Carlo generators used to generate top-antitop events which are processed with a fast detector simulation program named DELPHES [6]. The analysis is performed using the ROOT [7] framework.

3.1. PYTHIA 8

In all the studies discussed in this paper, the generic conditions were chosen to match those at the LHC. Dileptonic samples were generated using the PYTHIA 8 event generator for colliding protons at a center of mass energy of 7 TeV using the CT10 [8] parton distribution functions (PDF) and the renormalisation scale is set to be the arithmetic mean of the squared transverse masses of the two outgoing particles [4]. High-energy hadronic collisions are much more complex than electron-positron collisions due to beam-beam remnants and ISR from incoming partons and FSR from the top decay products. ISR, FSR and beam remnants are components of the underlying events. Also, multiple interactions can occur between pairs of incoming partons which contribute to the underlying events.

PYTHIA 8 provides a more accurate description of the parton shower and the multiple hard interactions than PYTHIA 6 [9] having the added advantage of it being written in C++ [10]. In PYTHIA 8, parton showers are ordered in transverse momentum squared [11] while in PYTHIA 6, they were ordered in squared mass [9]. The evolution of ISR, FSR showers and multiple parton interaction (MPI) depend on the transverse momentum [12, 13]. The evolution equation for all the processes interleaved is given by the relation (8):

$$\frac{dP}{dp_\perp}=\left(\frac{dP_{MPI}}{dp_\perp}+\sum\frac{dP_{ISR}}{dp_\perp}+\sum\frac{dP_{FSR}}{dp_\perp}\right)\exp\left\{-\int_{p_\perp}^{p_\perp^{max}}\left(\frac{dP_{MPI}}{dp_\perp'}+\sum\frac{dP_{ISR}}{dp_\perp'}+\sum\frac{dP_{FSR}}{dp_\perp'}\right)dp_\perp'\right\}. \quad (8)$$

Multiple interactions and ISR are competing processes and can modify the total energy by further MPI and ISR branching, while FSR redistributes the energy among the outgoing partons. In eq.8 $\frac{dP_{MPI}}{dp_\perp}$ and $\frac{dP_{FSR}}{dp_\perp}$ represents the ISR sum over all the incoming partons and the FSR sum over all kicked-out partons respectively; $\frac{dP_{MPI}}{dp'_\perp}$ denotes the multiple parton interactions occurred in process;

$p_{\perp}^{max}$ is the transverse momentum of the previous step of the process. This sequence of hard interactions can be repeated until a lower cut-off is reached, below which any interaction or further radiative branching is forbidden.

3.2. POWHEG

The POWHEG (Positive Weight Hardest Emission Generator) is a dedicated Monte Carlo generator program to produce heavy quarks in hadronic collisions. It describes with NLO accuracy the process in QCD and can be interfaced to Shower Monte Carlo programs such as HERWIG [14], PYTHIA and SHERPA [15]. The interface between Monte Carlo programs and POWHEG-hvq is done using the Les Houches Event File (LHEF) format [16].

In our study, $t\bar{t}$ samples are generated using POWHEG-hvq interfaced to PYTHIA 8, both programs using the same parton distribution functions, through the LHAPDF library [17]. The dileptonic samples are generated at LHC energies, $\sqrt{s} = 7\text{TeV}$ using the CT10 parton distribution function.

The POWHEG package is used to generate the first hardest emission according to the NLO matrix elements, being independent from the subsequent shower. The highest p_T is considered to be that one which triggers the parton branching. Thus, the highest transverse momentum ISR is computed with respect to the beam axis. The POWHEG output is interfaced to PYTHIA which is ordered in transverse momentum, offering the possibility to require some condition to the p_T scale of shower evolution. The subsequent shower should start if $p_T^{evol} > p_T^{POWHEG}$. The initial and final state radiation can occur also when the shower starts, then continue to develop downward [13]. All the information about the shower evolution is set in PYTHIA 8.

3.3. DETECTOR SIMULATION FRAMEWORK

The output information from PYTHIA is stored using the HepMC [18] file format and further processed with the fast detector simulation program DELPHES which is a good approximation of the ATLAS detector. The processed data are saved in a ROOT ntuple which contains a generator-level tree and an analysis-object tree. The output observables such as missing transverse energy (MET), collections of jets and leptons, are accessible in the *Analysis* tree.

In DELPHES, the isolation of charged leptons demands that the particle has $p_T > 2$ GeV/c within a cone of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\varphi^2}$ centred on the cell associated to the charged lepton. The information about MET is provided by the calorimeter cells only, while the jet reconstruction uses the Anti-kt algorithm with a default value of cone size $\Delta R = 0.4$, and the *b*-tagging efficiency is below the expected 40%.

The reconstructed objects are selected if they pass the following criteria:

- p_T>10GeV for electrons, muons;
- p_T>20GeV for jets.

For the analysis we were using a sample of 1M dileptonic $t\bar{t}$ events simulated at a center of mass energy of 7 TeV. Since the goal of the analysis is limited to the study of the ISR and FSR effects, no background was generated.

4. METHOD TO DETERMINE THE TOP MASS

4.1. PARTON LEVEL

Before starting the analysis we have to test the m_{T2} method at the parton level with ISR/FSR and without ISR/FSR. The m_{T2} method has the property that its distribution's end point is delimited by $m_t = m_{T2}^{\max}(m_\nu)$. The same features can be seen in Fig. 2 where a significant amount of events reach the end point of the m_{T2} distribution.

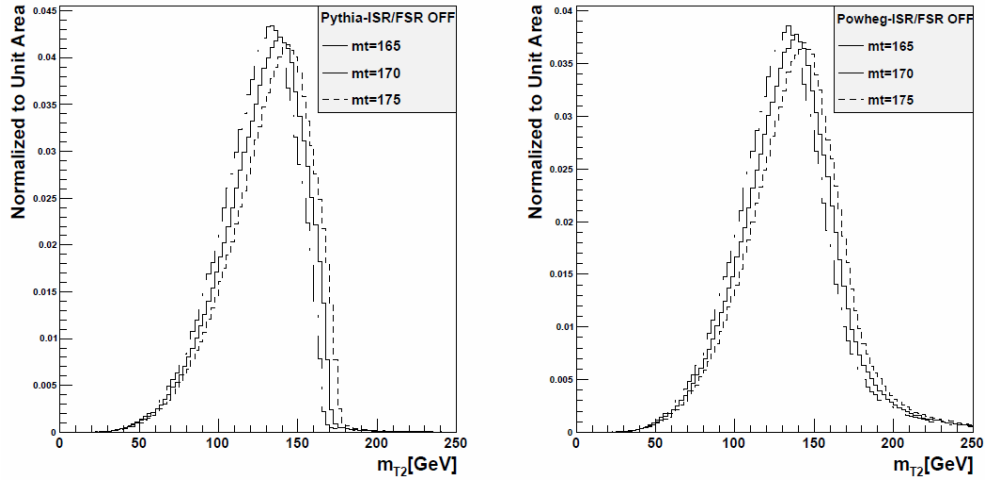

Fig. 2 – The m_{T2} distribution at the parton level for m_t =165 GeV, m_t =170 GeV and m_t =175 GeV when ISR/FSR are off for Pythia (left) and Powheg (right).

In Fig. 2 (left) there are the simulated distributions for m_{T2} for different input top masses, which have an edge point at the top mass for PYTHIA, while the POWHEG distributions (right) not exhibit this behaviour near the generated mass of the top quark. In this case, the lack of the end point is due to POWHEG which describes with next to leading order accuracy the QCD processes. The m_{T2} variable is sensitive to the top mass for both MC generators.

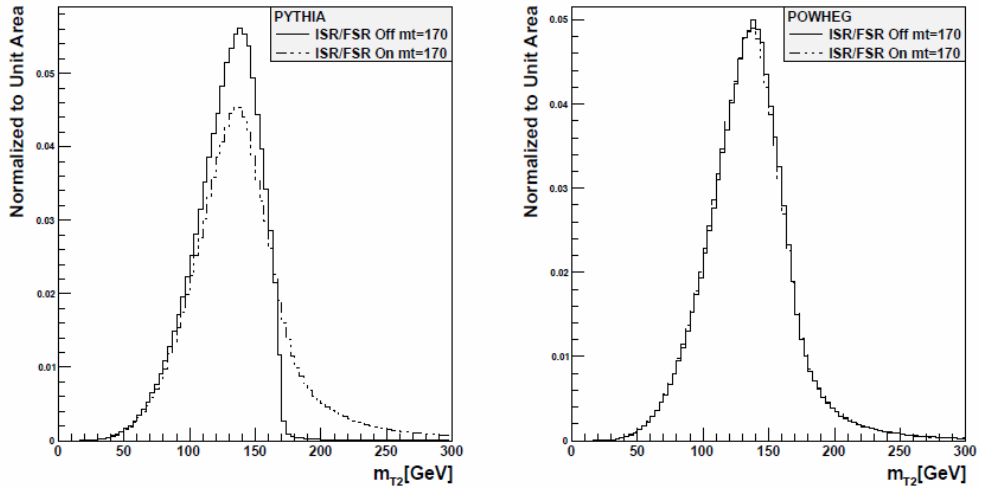

Fig. 3 – The m_{T2} distribution for ISR/FSR turned on and off for m_t = 170 GeV at parton level for PYTHIA generator (left) and POWHEG (right).

In Fig. 3 (left) the dashed and continuous lines show the m_{T2} distributions of data generated in PYTHIA with ISR/FSR on and off, respectively. It is worth mentioning that the small tail which appears in m_{T2} distribution for ISR/FSR off is due to the top quark decay width. ISR and FSR affect the top quark mass reconstruction differently, but their combined effect (when both are switched on) is that the m_{T2} distribution no longer has a clear end point at the generated top quark mass. The production of top quarks is associated with ISR from incoming partons. The ISR jets have similar p_T range as b jets coming from top decay, therefore the ISR jets contribute to the m_{T2} calculation smearing the distribution. In addition, the FSR jets effectively decreases the energy contents of the b jets and also changes the relative angle between the final state jets. The uncertainty due to ISR and FSR affects the m_{T2} distribution and consequently the top mass reconstruction.

Regarding the second generator, POWHEG, one can see in Fig. 3 (right) that both m_{T2} distributions for ISR/FSR switched on and off are similar, therefore ISR/FSR activity has no influence on the m_{T2} distribution.

Comparing PYTHIA 8 with POWHEG, we do notice an important difference between the two of them and this happens due to the fact that PYTHIA is a Leading Order generator while POWHEG is a Next to Leading Order generator. This is explains why the m_{T2} distribution computed for samples generated with PYTHIA has an end point at the input top mass for ISR/FSR off, while this feature is not visible in the POWHEG samples with ISR/FSR off.

4.2. RECONSTRUCTION LEVEL

Using the events which are selected after the specific cuts for a dileptonic topology we obtain the m_{T2} distribution. This will help determine the mass of the top quark from the reconstructed objects. The suitable conditions are found to be:

- at least two b-tagged jets with $p_T > 25$ GeV and $|\eta|<2.5$;
- only two isolated leptons of opposite charge with $p_T > 25$ GeV and $|\eta|<2.5$;
- $|m_{ll}- m_Z|>5$ GeV and $E_T^{miss}>40$ GeV.

The method to compute m_{T2} variable is the same as we described in the section 2; the first step is to identify all possible configurations for b jets and leptons. There are two combinations of b-l paring: (b_1l_1,b_2l_2) and (b_1l_2,b_2l_1), where b_1, b_2 are two b-tagged jets and l_1,l_2 are two isolated leptons. We assume that all events used to compute m_{T2} are correctly selected. The smallest value of m_{T2} is plotted in Fig. 4 for ISR/FSR turned off and on and $m_t = 170$ GeV and $m_v = 0$.

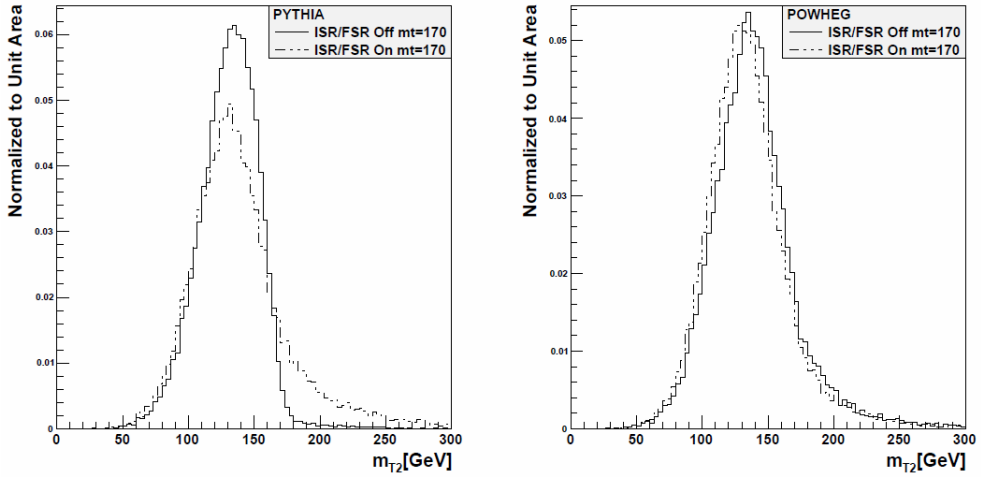

Fig. 4 – The m_{T2} distribution for ISR/FSR turned off and on; m_{T2} is computed for $m_t = 170$GeV.

The influence of ISR and FSR could be observed at the reconstruction level by smearing of the end point in PYTHIA, while in POWHEG both distributions look alike. As it was mentioned at the parton level, ISR jets can be misidentified with the jets coming from top decay, while FSR jets influence the transverse momentum of the b jets and also the angle between them. The kinematics of the final state are modified by ISR/FSR leading to a distorted m_{T2} distribution. The effects due to ISR and FSR are very hard to manage.

Next sections present the end point method and the likelihood method used for the estimation of the top quark mass.

4.2.1. End point method

This method consists in fitting the end point of the m_{T2} distribution obtained from events generated in PYTHIA with a top input mass equal to 170 GeV while the ISR/FSR are turned off. The fit is performed using two linear functions, chosen in a suitable range. Figure 5 shows the fit with the two linear functions having the junction point as a free parameter, obtaining a top mass m_t = 172±0.4 GeV. The systematic uncertainty due to ISR/FSR is estimated by measuring the top mass shift between a sample without ISR/FSR and with ISR/FSR (real data). Both ISR/FSR are known to 20% level [19], therefore the systematic errors are considered to be 20% of shift in mass. The obtained value is 1 GeV. To improve this result a likelihood method will be used in the next section.

This method does not work for a m_{T2} distribution obtained from events generated with POWHEG because it does not show a clear edge at the top mass. In the POWHEG framework, the hardest emission is performed first, using full NLO accuracy.

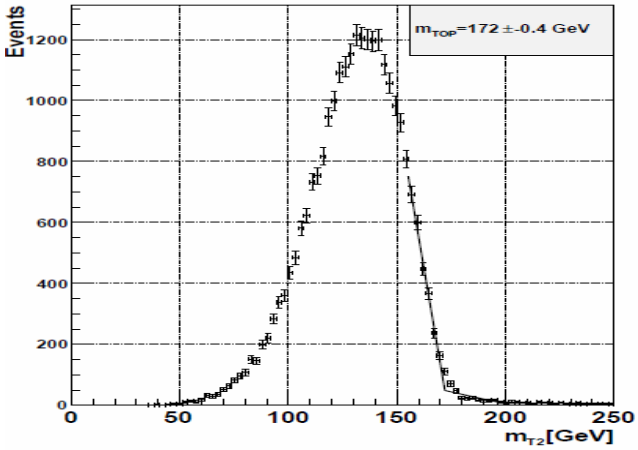

Fig. 5 – The m_{T2} distribution fitted by end point method; the events are generated in PYTHIA with a top input mass of 170 GeV and the conditions of ISR/FSR off.

4.2.2.Likelihood Fit

The m_{T2} sensitivity to top quark mass is used to apply the likelihood method. Ten samples of dileptonic data for different masses in the range of 165-175 GeV with steps of 1 GeV were generated with PYTHIA 8 and POWHEG. Every template is compared to a reference which in our case was chosen to be the one with m_t = 170 GeV. The top mass is extracted from a likelihood fit.

The binned likelihood formula is given by the equation 9 [20] and represents a product of Poisson functions for each bin over N bins:

$$L = \prod_{i=1}^{N} \frac{e^{-p_i} \quad p_i^{k_i}}{k_i\,!}, \tag{9}$$

where k_i and p_i represent the number of data events observed in bin I for nominal data distribution and normalized data, respectively. For each of the ten templates the m_{T2} distribution is matched to the nominal data in the range of $100 < m_{T2} < 200$. The fit is performed with a quadratic function and the minimum of $-\ln L$ represents the mass of top quark

In Figs. 6 and 7 we present the results of the likelihood fit for the m_{T2} distributions for ISR/FSR off and ISR/FSR on for both generators, respectively. L_{max} from the negative logarithm of the likelihood ratio $L / L_{\max}$ is computed as a minimum of the quadratic fit of $-\ln L$ distribution.

The likelihood method used to reconstruct the top mass in the dileptonic channel proves to be a good way to find the top quark mass for both cases. In Fig. 6 the likelihood fits for both generators with and without ISR/FSR are presented.

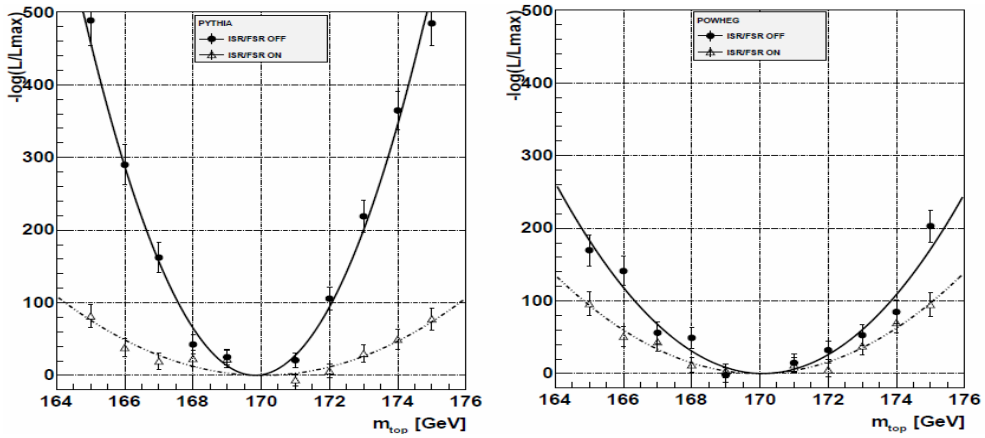

Fig. 6 – Two superimposed negative logarithmic likelihood distributions for ISR/FSR turned on and turned off for PYTHIA and POWHEG.

The top quark mass measurement accuracy is influenced by the existence of the ISR/FSR. Figure 6 shows a greater parable width when ISR/FSR is on, thus providing greater uncertainty for both Monte Carlo generators. The statistical error, less than 0.5 GeV, and the systematic error due to ISR/FSR, computed as we mentioned in the previous section, is 0.08 GeV for PYTHIA 8 and 0.02 GeV for POWHEG.

5. SUMMARY AND CONCLUSIONS

We have investigated the influence of the initial and final state radiation over the top mass reconstruction using the m_{T2} variable in the dileptonic channel. The study was performed using two different Monte Carlo generators, PYTHIA 8 which is a LO MC generator and a NLO MC generator, POWHEG.

The variable m_{T2}, for both MC generators, was depicted for both partonic and reconstruction level, with ISR and FSR switched off and on (Figure 2 and 4). The m_{T2} distribution was distorted by ISR/FSR at both parton and reconstruction levels for PYTHIA 8 and insignificantly for POWHEG.

Two methods were used to determine the ISR/FSR influence on the top mass estimation: the end point fit method and binned likelihood fit. We have shown that the end point fit method can be applied only for PYTHIA 8 data samples, while the binned likelihood can be successfully used for both Monte Carlo generators samples.

We summarize the results on the uncertainty induced by ISR/FSR in Table 1.

Tabel 1

The top mass shift Δm_t and the systematic errors δm obtained for both Monte Carlo generators in the dileptonic channel.

Generators	**Source of uncertainty**	**$\|\Delta m_t\|$(GeV)**		**δm_t(GeV)**	
		End point fit	*Binned Likelihood*	*End point fit*	*Binned Likelihood*
PYTHIA 8	ISR/FSR	5,7	0.4	1	0.08
POWHEG	ISR/FSR	-	0.12	-	0.02

The systematic error induced by ISR/FSR on the reconstruction of the top quark mass, estimated with the likelihood method, is 0.02 GeV for PYTHIA 8 and 0.08 GeV for POWHEG, while in the case of the end point fit method, for PYTHIA 8 only, the mass value is modified by 1 GeV.

Summarizing, the top mass reconstruction is less affected by the activity of ISR/FSR for POWHEG generator, due to the fact that it describes with NLO accuracy the process in QCD. Therefore the POWHEG is recommended for further Monte Carlo data based analysis of the top mass reconstruction algorithms studies.

Acknowledgements. This work was supported by grant POSDRU/88/1.5/S/56668. We acknowledge the support of the Romanian National Authority for Scientific Research (ANCS) through the ATLAS Capacities/Module III CERN project.

REFERENCES

1. C.G. Lester and D.J. Summers, Phys. Lett., **B 463**, 99 (1999).
2. A. Barr, C. Lester, and P. Stephens, J. Phys., **G 29**, 2343 (2003).
3. The Review of Particle Physics, K. Nakamura *et al*., Particle Data Group, J. Phys., **G 37**, 075021 (2010).

4. T. Sjöstrand, S. Mrenna, P.Z. Skands, Comput. Phys. Commun., **178**, 852–867 (2008).
5. P. Nason, J. High Energy Phys. **11**, 040 (2004);
S. Frixione, P. Nason, C. Oleari, J. High Energy Phys., **11**, 070 (2007).
6. S. Ovyn, X. Rouby, V. Lemaıtre, arXiv:0903.2225v3.
7. ROOT homepage, http://root.cern.ch/drupal/
8. *** CT10 Parton Distribution Home Page, http://hep.pa.msu.edu/cteq/public/
9. T. Sjöstrand, S. Mrenna, P.Z. Skands, J. High Energy Phys., **05**, 026 (2006).
10. D. Majumder 1, K. Mazumdar 1, T. Sjöstrand,arXiv:1002.4296v1 [hep-ph].
11. T. Sjöstrand, P.Z. Skands, Eur. Phys. J., **C 39**, 129–154 (2005).
12. T. Sjöstrand, M. van Zijl, Phys. Rev., **D 36**, 2019 (1987).
13. R. Corke, Eur. Phys. J., **C 69** (2010).
14. Manuel Bähr *et al*., Eur. Phys. J., **C 58**, 639–707 (2008).
15. Gleisberg, T. and others, JHEP, **02**, 056 (2004).
16. J. Alwall *et al*., Comput. Phys. Commun., **176**, 300–304 (2007).
17. M.R. Whalley, D. Bourilkov, R.C. Group.
18. *** *HepMC* Event Record, http://lcgapp.cern.ch/project/simu/HepMC/
19. I. Borjanovi´c *et al., Investigation of top mass measurements with the ATLAS detector at LHC*, Eur. Phys. J., **C 39**, S2 63 (2005).
20. W.S. Cho, K. Choi, Y. G. Kim and C. B. Park, arXiv:0804.2185v1.