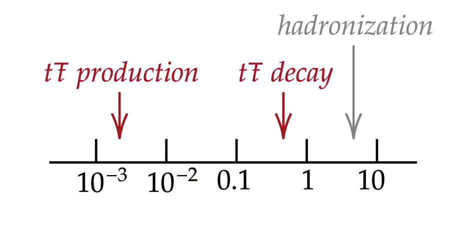
Multivariate analysis to discriminate top quark pair production channels at LHC

Morgan Del Gratta
Done in collaboration with: Prof. Maximiliano Sioli, Prof. Rita Fioresi and Claudio Severi

29 March 2023

The top quark

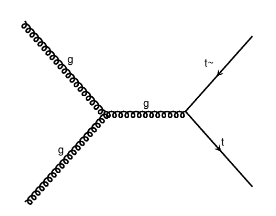
The **quark top** is one of the fundamental fermions of the Standard Model, and is also the most massive one: $m_t = 172.56 \pm 0.4$ GeV. Due to this it has a very short mean life, over ten times smaller than the typical QCD hadronization timescale, and therefore decays before it has time to hadronize.

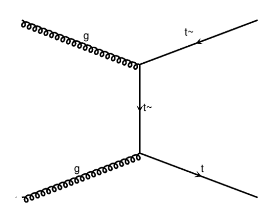


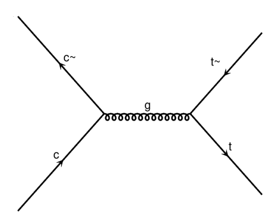
Relevant timescales for the top, in GeV $^{-1}$.

The top quark

At LHC the top t can be produced along its antiparticle, the antitop $\bar{t}$, through strong interaction ($t\bar{t}$ pairs), or by itself from the weak interaction. For $t\bar{t}$ pairs at Leading Order we have two **production channels**: gg and $q\bar{q}$.

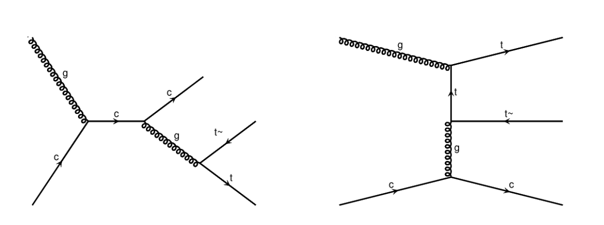






Il quark top

At Next to Leading Order we have an additional production channel: gq.



Different production mechanisms lead to pairs with different characteristics, and it can be useful to have a tool that discriminates pairs on the basis of their production channel.

Multivariate analysis

This type of problem, where one tries to isolate a signal while suppressing a background, is well known in the context of multivariate analysis, and is known as **classification**. In this work: signal $\rightarrow$ events gg, background $\rightarrow$ events $q\bar{q}$ and gq.

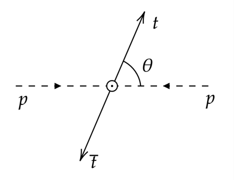
The algorithms have been implemented using the **TMVA** packet, integrated in the ROOT environment. The selected TMVA methods were:

- **Fisher Linear Discriminant**
- **Boosted Decision Trees**
- **Multilayer Perceptron**.

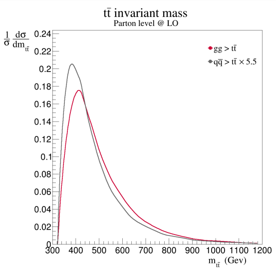
As input for the classifiers we simulated events with production of $t\bar{t}$ pairs. These were generated in two contexts:

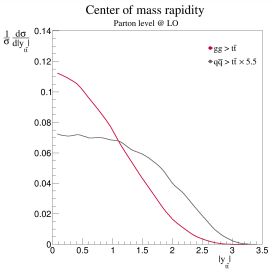
at **Leading Order** at **Parton Level** + at **Next to Leading Order** both at **Parton** and at **Particle Level**.

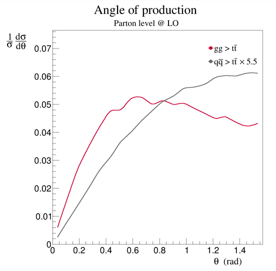
In this context we studied the kinematics of the top-antitop pair. We selected three observables to use in the classification process:



- the invariant mass of the system, $m_{t\bar{t}}$
- the rapidity of the pair $t\bar{t}$, $y_{t\bar{t}}$
- the production angle of the top in the center of mass reference frame, θ.

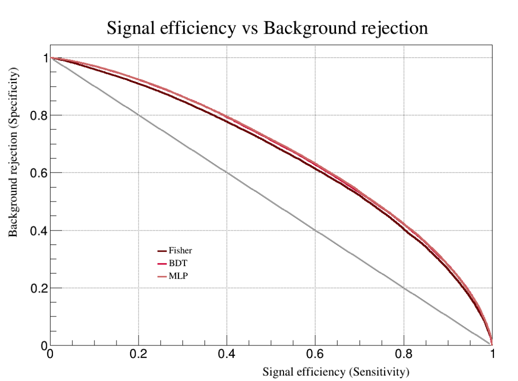






Analysis at Leading Order

At this level the analysis does not give great results, as one can see from the ROCs obtained:



We thus proceed to the analysis at NLO.

Analysis at Next to Leading Order

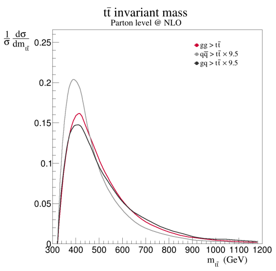
A NLO we can still utilize the previously mentioned variables, as well as two new ones:

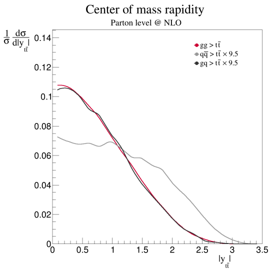
- the transverse momentum of the pair, p_T
- the number of jets in the final state, N_j.

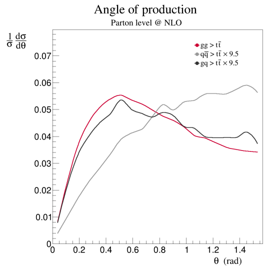
This is due to the fact that one expects a higher gluon radiation with low p_T in gg events than in scattering events between quarks and antiquarks. This provides higher transverse momentum for the pair produced through gluon fusion, as well as a larger number of particles with low transverse momentum in the final state.

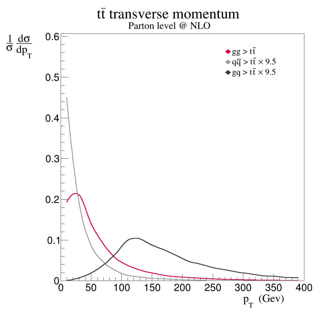
However we need to take into account the third production channel, gq.

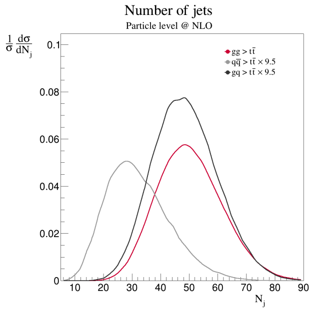
Analysis at Next to Leading Order





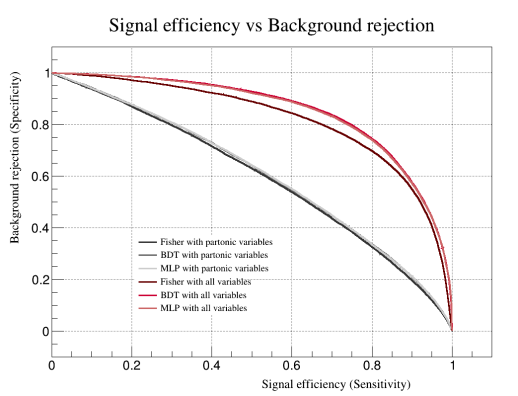






Analysis at Next to Leading Order

With the help of these additional variables the performance increases significantly:



Conclusions

The process of classification leads to good results when analyzing events at **Next to Leading Order**, having:

- kinematic observables of the pair at **Parton Level** + variables linked to their product decays at **Particle Level**.
- algorithm with best performance → **Boosted Decision Trees**
- initial purity sample of **0.81** → final purity sample of **0.92**
- final efficiency of **0.74**.

This analysis is a first step towards the selection of a clean sample of gg events, important for measuring their contribution at LHC and to study in detail spin correlations at TeV energies.