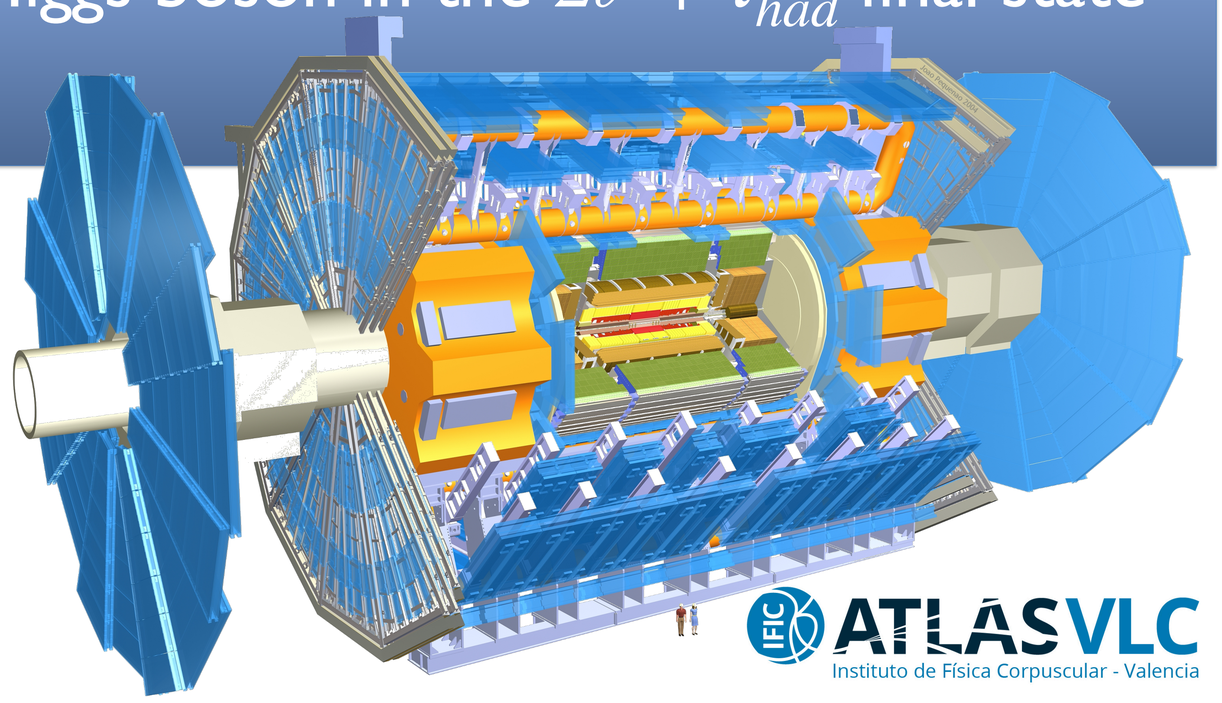
ATLAS Week
CERN (Geneva)
June 20, 2022

Lepton assignment for the associated production of single top quark and Higgs boson in the $2\ell + \tau_{had}$ final state

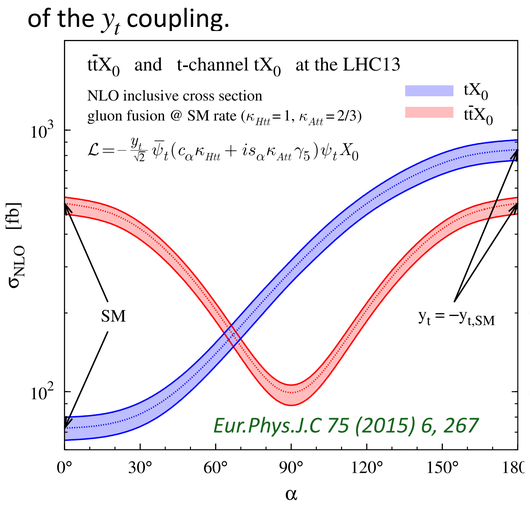
Pablo Martínez-Agulló
Instituto de Física Corpuscular, IFIC (CSIC-UVEG, Valencia, Spain) - Pablo.Martinez.Agullo@ific.uv.es



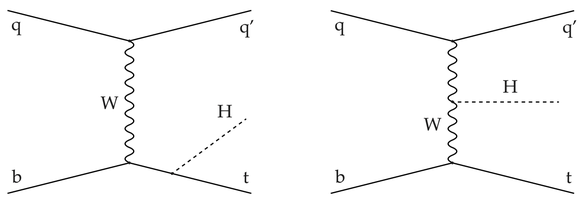
IFIC ATLASVLC
Instituto de Física Corpuscular - Valencia

Theoretical motivation

- In the Standard Model (SM) the Higgs field couples to fermions through a Yukawa interaction with a coupling strength proportional to the mass of the fermion.
- The Yukawa coupling between the **top** quark and the **Higgs** boson (y_t) is the **strongest coupling** of the SM, almost unity.
- The **only process sensitive to** both the **sign** and magnitude of y_t is the associated production of a single top quark and a Higgs boson (tHq). Its observation would allow to probe the CP properties of the y_t coupling.



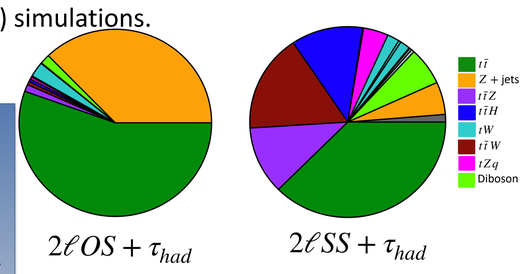
- The two leading order (LO) diagrams of the tHq production (right) **interfere depending on the sign of y_t .**
- The **SM predicts destructive interference** but if the y_t is different from the y_t^{SM}, a larger cross section would be observed (see blue line in the left figure).
- Current observations favour the SM predictions [*Eur.Phys.J.C 81 (2021) 378*].
- The $y_t = -y_t^{SM}$ hypothesis has not been discarded yet.



Dominant LO Feynman diagrams for the production of the tHq events. The Higgs boson can be radiated form either the quark or the *W*.

Background estimation

There are several processes whose final state cannot be distinguished from that of the tHq $2\ell + \tau_{had}$. These are estimated with Monte Carlo (MC) simulations.



$2\ell OS + \tau_{had}$ $2\ell SS + \tau_{had}$

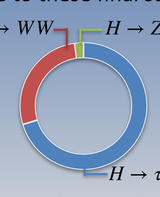
Main backgrounds: top quark-antiquark pair production ($t\bar{t}$) and *Z*+jets. These two processes mimic the signal signature due to the jets faking either the ℓ or the τ_{had} .

Search of the associated *tHq* production

The search of the tHq production, which is exceptionally **challenging** due to the extremely **small inclusive cross-section** (~73 fb) [*arXiv:1610.07922 [hep-ph]*], is performed in several channels which are defined according to the multiplicity of light-flavoured leptons (ℓ) and of hadronically decaying taus (τ_{had}) in the final state.

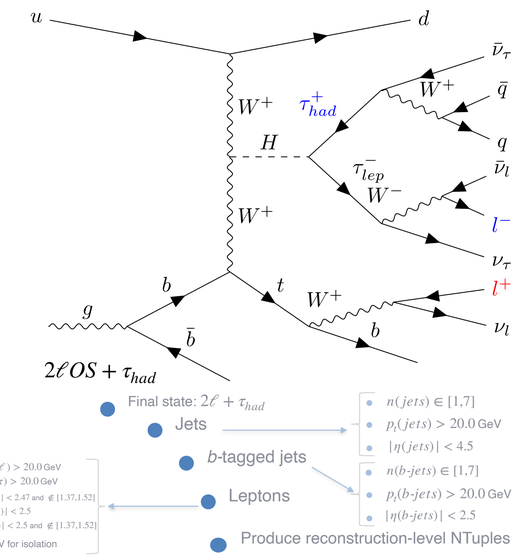
This analysis focuses on the signature of **two light leptons** (electrons or muons) **and one hadronically decaying tau** **($2\ell + \tau_{had}$)** and distinguishes between the channel in which the light leptons have the same ($2\ell SS + \tau_{had}$) or opposed ($2\ell OS + \tau_{had}$) electrical charge.

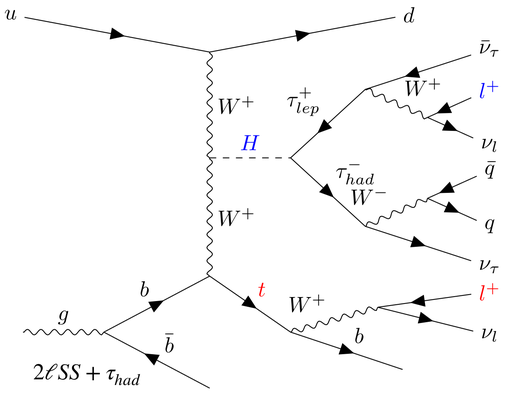
Three Higgs-boson decay channels contribute to these final states:



Relative contribution of each decay mode for the tHq $2\ell + \tau_{had}$

The combined cross-section of the two $2\ell + \tau_{had}$ final state channels is 3.5% of the total tHq production. Therefore, the **separation of signal** events from the main backgrounds is a **crucial** step for this analysis.



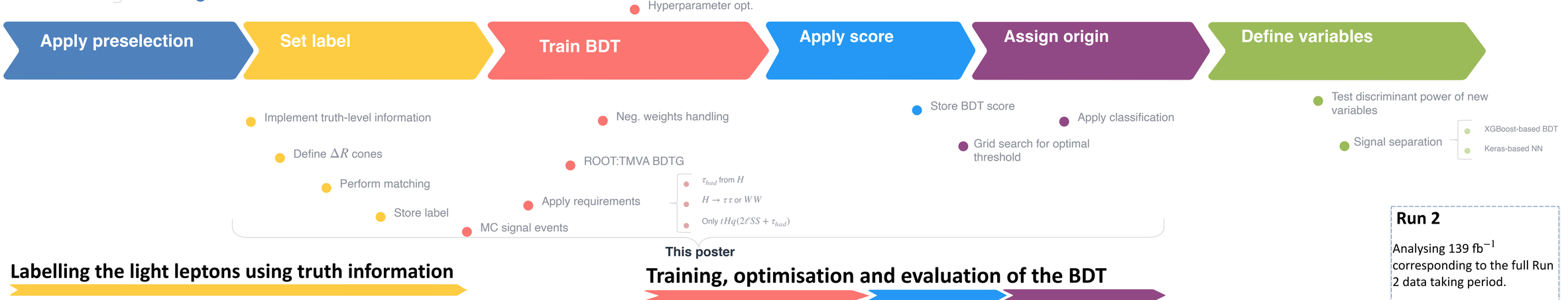


Lepton assignment

The knowledge of which light lepton is originated from the Higgs boson and which from the top quark is **crucial to discriminate** the tHq signal from the background.

- Used to **define reconstruction-level variables**.
- Assuming that the τ_{had} is originated from the Higgs boson decay (true 84% of times).
- $2\ell OS + \tau_{had}$: Unambiguous immediate association. The ℓ with the same electric charge as the τ_{had} is the one originated from the top quark decay.
- $2\ell SS + \tau_{had}$: Impossible to determine a priory the origin of the leptons. Using **machine learning** tools: gradient **Boosted Decision Tree** (BDT).

*Methods based on directly cutting on a set of variables have also been tested. The best of them offers a classification efficiency of 84%.

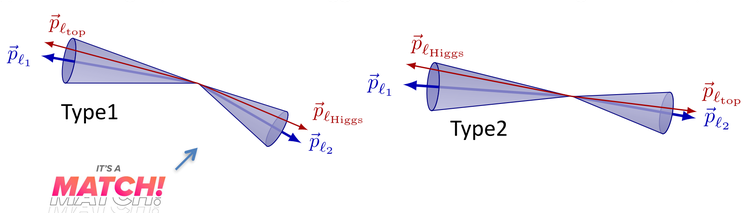


This poster

Labelling the light leptons using truth information

Before performing a supervised training it is necessary to **label the leptons** in the events. To do so, the **truth**-level and **reconstruction**-level information of the events is compared to determine a correspondence between the leading (ℓ_1^{reco}) and subleading (ℓ_2^{reco}) leptons and the lepton from Higgs boson (ℓ_{Higgs}^{truth}), and the lepton from the top quark (ℓ_{Top}^{truth}):

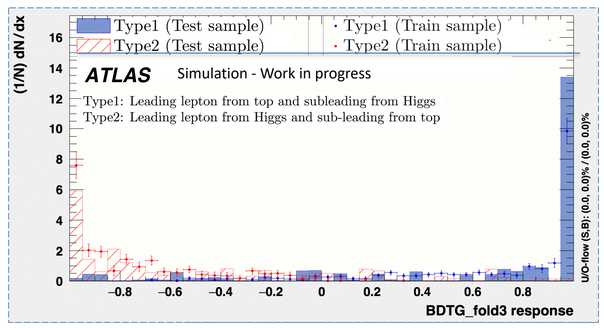
1. Access to parton-level information. At parton level the **origin** of the lepton is **obviously known**.
   - Requirements: τ_{had} from the Higgs decay and $H \to \tau\tau$ or $H \to WW$. The $H \to ZZ$ is neglected.
2. Define $\Delta R < 0.1$ cones around momentum of the the reconstructed lepton ($\vec{p}_{\ell_1}$).
3. Look for a parton level lepton ($\vec{p}_{\ell_{reco}}$) within each cone.
4. Perform matching:
   - Both cones are required to have one and only one lepton of each level.
5. Define type of events:
   - **Type1**: Leading lepton from the top quark and subleading from the Higgs boson (61.1%).
   - **Type2**: Leading lepton from the Higgs boson and subleading from top quark (38.9%).



Training, optimisation and evaluation of the BDT

After labelling the signal samples, a BDT can be trained to identify the origin of the light leptons:

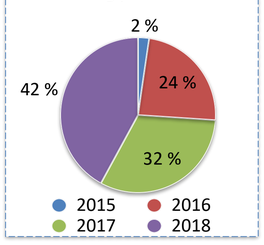
1. Select events for training:
   i. Signal events only: The process to reconstruct is the $tHq(2\ell + \tau_{had})$.
   ii. $2\ell SS + \tau_{had}$
   iii. Demand τ_{had} from Higgs boson.
   iv. Requiere $H \to \tau\tau$ or WW.
2. Ignore negatively weighted events.
3. Rank features: Optimised by ROOT.TMVA.
4. Optimise hyperparameters: Grid serach.
5. Use k-folding for cross-validation (k=5)
   i. Avoids overtraining
   ii. Mitigates the effects of low statistics.
   iii. Results in 5 different BDTs
6. Implement and evaluate the BDT: Implemented in the NTuple software framework where the BDT score is calculated for each event.
7. Once the BDT score is saved, the optimal threshold for classification is found through a scan. Best point around 0.0



$$\text{Efficiency} = \frac{\text{Events properly assigned}}{\text{Total events}} \geq 95\,\%$$

Run 2

Analysing 139 fb^{-1} corresponding to the full Run 2 data taking period.



For more details:

Summary

✦ In order to define variables that allow to create a signal enriched region, it is of primal importance the association of the light leptons to the Higgs-boson and top-quark decay systems.

✦ For the $2\ell SS + \tau_{had}$ a **BDT-based method** has been developed offering a classification **efficiency of more than 95%.**