JOINT INSTITUTE FOR NUCLEAR RESEARCH
Dzhelepov Laboratory of Nuclear Problems

FINAL REPORT ON THE START PROGRAMME

Estimation of misidentified of hadronic tau lepton using FakeFactor method in proton–proton collisions at $\sqrt{s} = 13$ TeV at the LHC

ATLAS experiment

Supervisor: PhD Huseynov Nazim
Student: Mammadova Aliya, Azerbaijan
Lomonosov Moscow State University BB
Participation period: Aug.14 – Sep.24, 2022

Dubna, 2022

Contents

1 Introduction

The main mechanism for the production of the Higgs boson at the Tevatron and LHC hadron colliders is the fusion of colliding gluons into a Higgs boson through a triangular loop diagram of quarks. The main processes of the production of the Higgs boson, studied at the Large Hadron Collider[4]:

- Gluon Fusion: $gg \rightarrow H$
- WW-, ZZ-Fusion: $W^+W^- \rightarrow H, \;\; ZZ \rightarrow H$
- Higgs boson emission from W- and Z- bosons: $q\overline{q} \rightarrow WH, q\overline{q} \rightarrow ZH$
- Emission of the Higgs boson from the top-quark: $q\overline{q} \rightarrow t\bar{t}H, \;\; gg \rightarrow t\bar{t}H,$

We investigate the associative production of the Higgs boson with the two top quarks in the 2lSS1tau channel, it is produced together and decays into two tau leptons. The Feyman diagram of this process is shown in Figure 1. Tau leptons can decay leptonically to an electron or muon (and associated neutrinos) but such decays are very difficult to distinguish from prompt leptons. The study of tau leptons occupies an important place in the ATLAS physics program. The heaviest lepton (with a mass of 1.78 GeV) appears in many final states of physics beyond the Standard Model (SM). In the Standard Model, the Higgs boson is predicted

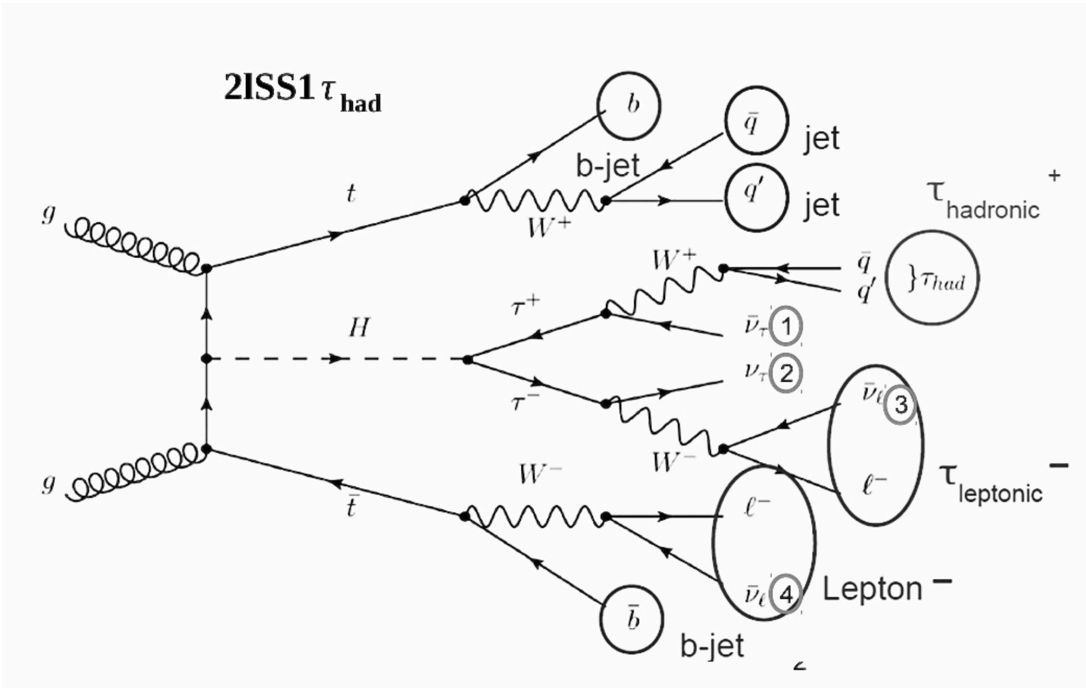

Figure 1: Feyman process diagram ttH

to couple more strongly to the heviest fermions, especially to the top quark. The top-quark Yukawa coupling, expected to be of order unity, can be probed directly by measuring the cross section for associated production of a Higgs boson with a top-quark pair ($t\bar{t}H$).

2 The Standard Model

The Standard Model (SM) is a quantum field theory that describes not only the dynamics of elementary particles but also particle decays and the production of new particles out of a vacuum,which corresponds to the energy ground state of the theory. The elementary particles are characterized by their mass, electric charge, and spin angular momentum $s\bar{h}(\bar{h} = h/2\pi)$, where s is the spin quantum number or spin for short and h is Planck's constant. In the SM, particles are divided into two categories: particles with half-integer spin are called fermions, and those with integer spin are called bosons. There are three types of families of particles: two types of fermions called leptons and quarks with spin–$\frac{1}{2}$, and one family of bosons called gauge bosons with spin–1. These particles and the corresponding anti-particles constitute all matter in the Standard Model.[2]

The theory of the Standard Model describes a collection of fundamental point-like particles and mechanisms which are the foundation of all visible matter and three of the four fundamental forces, namely the electromagnetic force, the weak force and the strong force. Gauge bosons mediate the three fundamental forces named before and the Higgs boson is a consequence of the Higgs mechanism, which is the part of the Standard Model that explains the masses of fundamental particles.

3 The LHC and the ATLAS

The Large Hadron Collider (LHC) is the most powerful particle collider in the world, capable of colliding proton-beams at center-of-mass energies of up to 14 TeV. The acceleration of the particles takes place in the 27 km diameter ring of the LHC and its three primary preaccelerators. The protons beams consist of bunches, 2808 per beam, with a bunch-spacing of 25 ns and each bunch containing 1.1×10^{11} protons. In order to idedentify the approximately 1000 particles that emerge per bunch from the collision-point, large detectors are needed. ATLAS (A Toroidal LHC ApparatuS) is, together with CMS (Compact Muon Solenoid), one of the general-purpose detectors of the LHC, on whose interaction-point it is centered. It has been built to analyze and identify the particles that are emerging in the pp-collisions.[3]

Beams are focussed and brought to collision at four different interaction points along the accelerator ring and at each interaction point one experiment is located. Beams consist of protons which are grouped into bunches composed of approximately 1.3×10^{11}protons per bunch.

4 The top quark

We are focused on investigation of three particles, this is the Higgs boson, which we talked about earlier, this is the tau lepton and the top quark. In the Standard Model, the t quark has the same quantum numbers and interacts in the same way as the $u-$ and $c-$ quarks. It is a fermion with spin 1/2 and a $b-$quark partner in a weak isotopic spin, has an electric charge $Q_t = +2/3$ and is described by the Dirac field. The lifetime of the t-quark is about 5×10^{-25} seconds.The t quark is the most massive of all particles in the Standard Model:

$$m_t = (173,1 \pm 0,6)\ \ GeV$$

In the SM, the masses of all particles are formed due to the interaction with the Higgs scalar field condensate. The interaction of the t-quark with the Higgs field is a Yukawa-type interaction of the form $\Lambda_f = y_t \bar{f} f H$ with constant $y_t = \sqrt{2} m_f / v_{ew}$, where $v_{ew} \approx 246 GeV$ the value of the vacuum average given by the Fermi constant G_F

Main mechanisms of formation of top quarks.

Within the framework of the SM, the main (in terms of the magnitude of the corresponding cross sections) mechanisms for the formation of t-quarks in hadronic interactions are gluon-gluon and quark-antiquark annihilations, leading to the formation of a $t\bar{t}$ pair[1]:

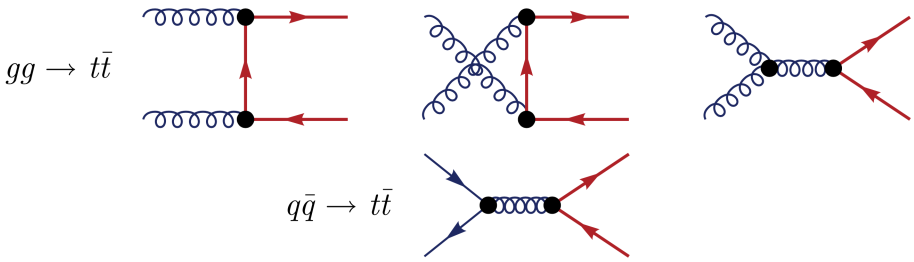

Figure 2: Diagrams describing the production of a pair of $t\bar{t}$ quarks

$$gg \to t\bar{t}$$
$$q\bar{q} \to t\bar{t}$$

The next largest cross section is the processes of electroweak production of the t-quark. These processes are usually classified according to the square of the 4-momentum of the virtual W-boson involved in the process:

$$qb \to tq\prime,\ p_W^2 < 0 : \text{t-channel},$$

$$q\bar{q}\prime \to t\bar{b},\ p_W^2 > 0 : \text{s-channel},$$
$$gb \to tW,\ p_W^2 = M_W^2 : \text{tW-channel}.$$

5 τ leptons

Six particles that do not participate in strong interactions form a class of leptons. These are electron e^-, negatively charged muon μ^-, negatively charged τ^-lepton and three neutral particles - electron neutrino ν_e, muonic neutrino ν_μ and tau-neutrino ν_τ.

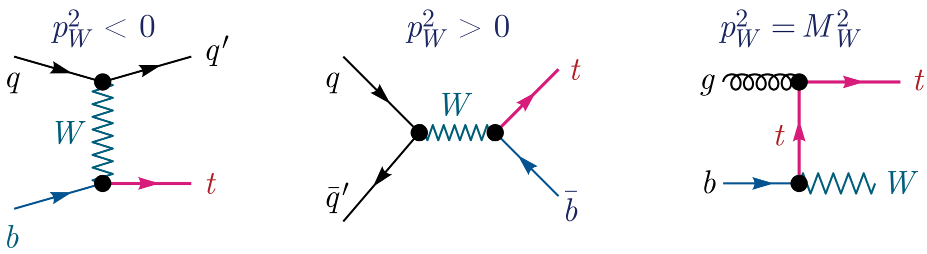

Figure 3: Diagrams describing the processes of electroweak (single) production of t-quarks

Leptons are considered structureless particles. Their size is $< 10 - 17cm$. The study of the properties of leptons shows that they are grouped in pairs. Each pair consists of a negatively charged lepton and a neutrino. Thus, 6 leptons form 3 generations. All leptons have spin J = 1/2, they are fermions. Charged leptons (e^-, μ^-, τ^-) participate in electromagnetic and weak interactions. Neutral leptons $(\nu_e, \nu_\mu, \nu_\tau)$ participate only in weak interactions. Every lepton has an antiparticle. These are the positron e^+, the positively charged muon μ^+, the positively charged taon τ^+ and the three types of antineutrinos antineutrino e, antineutrino μ, antineutrino τ. Each generation of leptons has its own lepton charge, respectively L_e, L_μ, L_τ. The lepton charge, like the usual electric charge, is conserved and additive, i.e. the charge of a system of leptons is equal to the sum of the lepton charges of individual leptons and must be the same before and after the completion of any process.

τ^-lepton is negatively charged, $\tau^+$is positively charged. They are respectively a particle and an antiparticle. τ-lepton as a result of weak interaction, which occurs under the action of the W–boson, turns into τ-neutrino ν_τ. The W-boson decays into one of the following pairs of particles:

- electron e^-, electron antineutrino antineutrino e,
- negatively charged muon μ^-, muon antineutrino antineutrino μ,
- d quark, anti u antiquark.

Because quarks and antiquarks in the free state are not observed, but are part of hadrons, hadrons are formed at the decay point of the W-boson, which are observed as a result of the decay of the $\tau-$ lepton. The $\tau-$lepton has a lifetime $\tau \approx 2.9 \times 10^{-13} s$ and therefore, as a rule, is registered through its decay channels.

6 Fake tau estimation

'Fake' leptons are originating in Hadron decays, Photon Conversions, or misidentified objects. Leptons with other origins (not Fake) are classified as 'Prompt'. This classification is done on truth-level and is only possible in MC samples.We will use the TRexFitter package to display the signal, background, and data. Let's say we have two regions: Signal region and Control region. We will evaluate fake tau in two ways, one of them is the evaluation using logical triggers we require τ candidates to have $p_T > 25GeV$. In the signal region, we require the presence of two leptons of the same sign and we set the cut to quality of leptons reconstruction. First, we will perform an analysis using a trigger system on Monte Carlo data in the 2lSS1tau channel, then we will consider the Fake Factors method, where we will use both experimental data and Monte Carlo data. The fundamental idea of the fake factor method is simple: select a control sample of events enriched in the background being estimated, and then use an extrapolation factor to relate these events to the background in the signal region. In the fake factor (FF) method, background processes where a quark- or gluon-initiated jet is reconstructed and identified as a τ candidate are estimated from data. The fake factor method is universal and precise.[6]

The method is data-driven provided the control sample is selected in data, and the extrapolation factor is measured with data. The control region is defined in order to select the background being estimated. The type of background considered with the fake factor method arises from particle misidentification.

The coefficients are calculated according to the following formulas:

$$FF = \frac{N_{pass}ID - N_{pass}ID, true\ \tau}{N_{failed}ID - N_{failed}ID, true\ \tau}$$

$$N^{SR}_{pass\ ID} = (N^{SR}_{failed\ ID} - N^{SR}_{failed\ ID, true\ \tau}) \cdot FF$$

In the numerator, the first term is the data that went through the Recurrent Neural Network(RNN), in our case tausJetRNNigMedium0==1 and went through the trigger, the second term is the Monte Carlo data that went through the trigger, RNN and truth information. In the denominator, the first term did not pass the RNN, and the second is the Monte Carlo data, which also did not pass the RNN, but passed the truth information and the trigger. Now, having defined all the variables, we can proceed to the calculations.

7 Results

In this work, the distributions of all the variables used were obtained, using the tauisHadronic variable, the number of fake and real tau was determined. Tauishadronic is a boolean value, it takes the value of either 0 or 1, and if it takes the value 0, it will return false, otherwise true. With the help of this variable and the rest of the cuts, the values presented in Table 5 were obtained, as can be seen from the table, the value of the ttH signal in the signal region is 14.3. The remaining Samples are the main backgrounds, which were calculated separately to reduce the computation time.

The next step, after creating new regions in the config file, we calculated the variables we need to calculate the coefficients F for Fake Factor method. Variable values already take into account non-main and other backgrounds such as VV, tZ, WtZ,

Sample	Fake tau yields	Real tau yields	Hadronic tau yields	Fake tau Fraction(%)
ttH	3.1	11.3	14.3	21,6
ttW	6.4	4	10.4	61,5
ttZ	4.3	10.1	14.4	29,8
ttbar	3.7	3.4	7.1	52,1

Figure 4: Signal Region (SR)

1 prong

Control Region (CR)		Signal Region (SR)		DATA	
N_passRNN_Matching_1prong	36,545	N_failedRNN_Matching_1prong	3,382	N_DATA_failedRNN_1prong **(SR)**	15
N_failedRNN_Matching_1prong	4,488	N_passRNN_Matching_1prong	22,057	N_DATA_passRNN_1prong **(SR)**	29
N_failedRNN_noMatching_1prong	8,068	N_failedRNN_noMatching_1prong	5,599	N_DATA_passRNN_1prong **(CR)**	56
N_passRNN_noMatching_1prong	13,103	N_passRNN_noMatching_1prong	8,912	N_DATA_failedRNN_1prong **(CR)**	8
Total			39,950		
Fakes	64,352				

3 prong

Control Region (CR)		Signal Region (SR)		DATA	
N_passRNN_Matching_3prong	14,38	N_failedRNN_Matching_3prong	1,926	N_DATA_failedRNN_3prong **(SR)**	3
N_failedRNN_Matching_3prong	2,611	N_passRNN_Matching_3prong	6,213	N_DATA_passRNN_3prong **(SR)**	8
N_failedRNN_noMatching_3prong	4,807	N_failedRNN_noMatching_3prong	3,440	N_DATA_passRNN_3prong **(CR)**	14
N_passRNN_noMatching_3prong	2,878	N_passRNN_noMatching_3prong	2,087	N_DATA_failedRNN_3prong **(CR)**	3
Total			13,666		
Fakes	1,048				

Figure 5: Values for calculating the coefficients FF and N^{SR}_{passID}

tW and others After obtaining the desired values, we can get the FF coefficient for one prong, which will be

$$FF_{1prong} = 5,539$$

and the FF coefficient for 3 prongs, which will be

$$FF_{3prongs} = 0,976$$

based on the found values, we we can compute

$$N^{SR}_{passID} = 65,4.$$

Figures 7, 8 and 9 show the results of the fake evaluation without RNN and Fake Factor. As can be seen from the graphs, the signal value in the Signal region is 8.9, and in the Contol region 14.3 RNN(Recurrent neural networks) has 4 work points, in all calculations we used Medium work point.

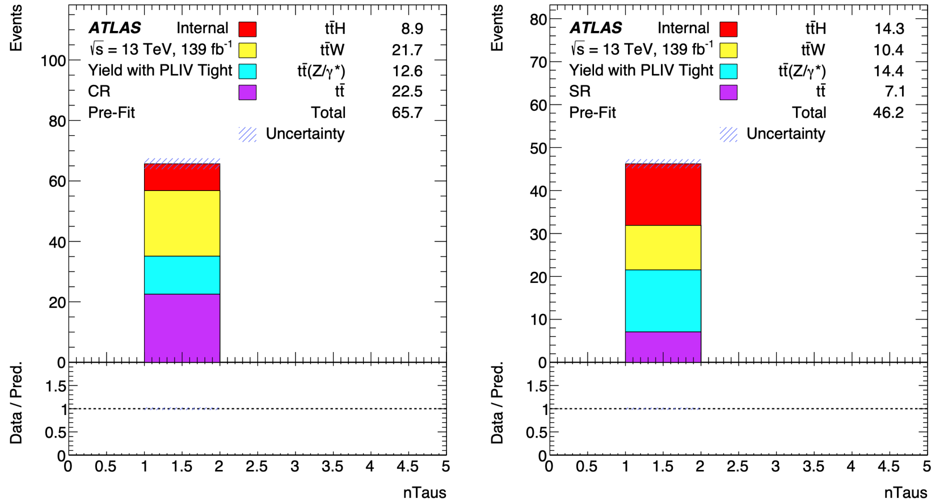

Table 1: Hadronic tau yields in Signal and in Control Regions

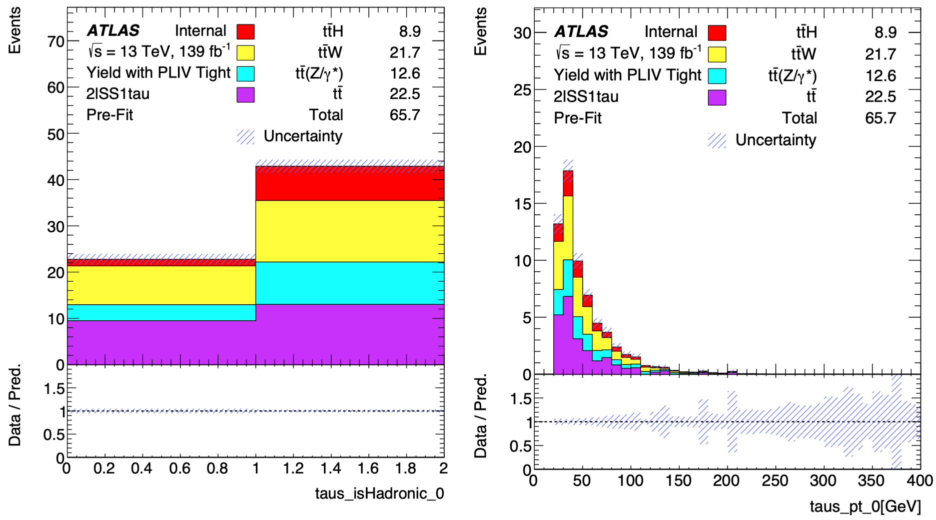

Table 2: Variable tau is Hadronic and Momentum distribution

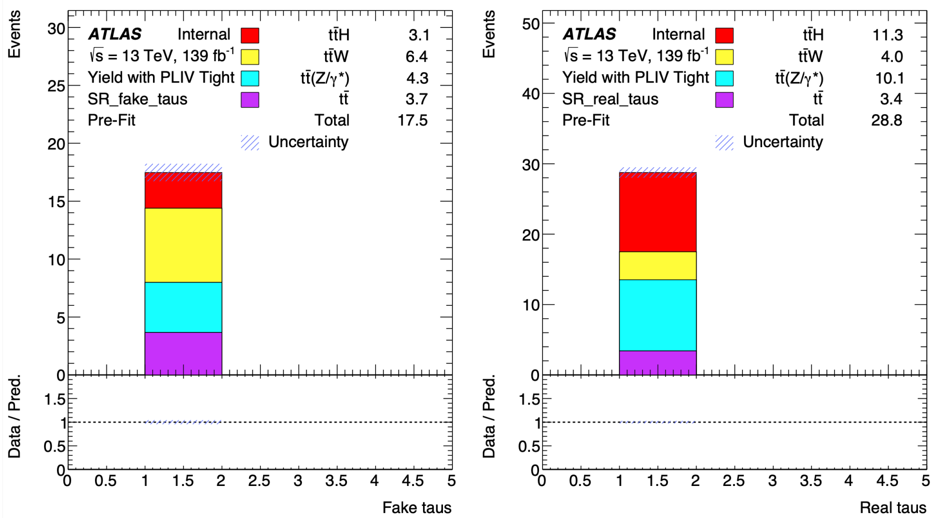

Table 3: Estimation of real and fake tau in the Signal Region(SR)

8 Conclusion

In this work, the fake tau leptons in the 2lSS1tau channel were estimated using the Fake factor data-driven method. Which gave the value $N^{SR}_{passID} \approx 65,4$. The work

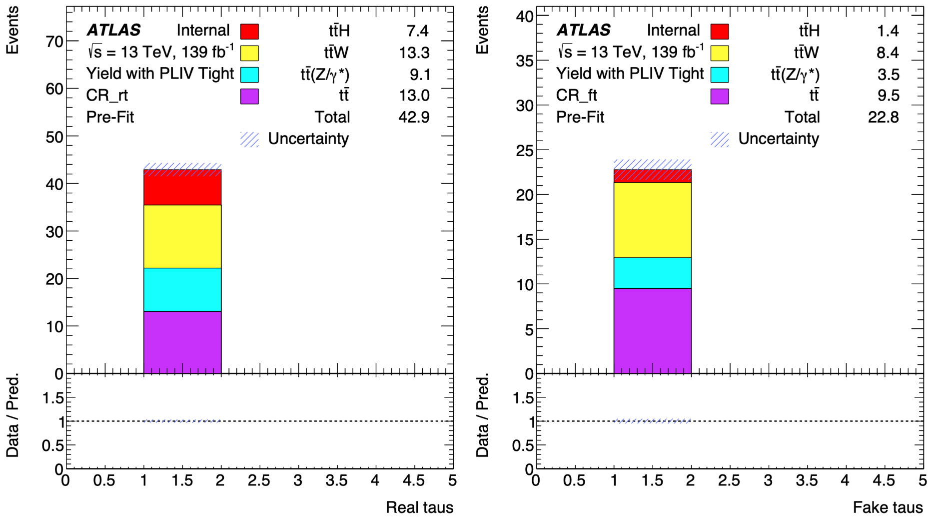

Table 4: Estimation of real and fake tau in the Control Region(CR)

was done using the TRExFitter package. This analysis needs to be extended, this will be a task for further research.

References

[1] Top Quark. Results and prospects *D.V.Skobeltsyn Research Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, Institute of High Energy Physics of the National Research Center "Kurchatov Institute", Protvino, Russia*, *E.Boos, L.Dudko, P.Mandrik, S.Slabospitsky*, (2019)

[2] Search for the Standard Model Higgs boson produced in association with a top quark pairin multilepton final states at $\sqrt{s} = 13TeV$ with the ATLAS detector *Czech Technical Universityin Prague*, *Babar Ali*, (2019)

[3] Study of Higgs boson production inassociation with $t\bar{t}$ quark pairs at $\sqrt{s} = 13TeV$ with the ATLAS detector *Masterarbeit in Physik vorgelegt dem Fachbereich Physik, Mathematik und Informatik (FB 08) der Johannes Gutenberg-Universität Mainz*, *Julian Fischer, Prof. Dr. Stefan Tapprogge, Prof. Dr. Anthony Doyle* , (2018)

[4] Study of the associative production of the Higgs boson together with the top quark in the ATLAS experiment *Moscow Institute of Physics and Technology*, *Tropina A.D.*, (2022)

[5] Chapter 8 The Fake Factor Method, 64 pages

[6] Data-driven estimation of fake τ background in Higgs searches in ATLAS *Institute of Nuclear Physics, Polish Academy of Sciences Marzieh Bahmani*