The review of PhD thesis

The author of the thesis: Santu Mondal
The title of the thesis: Processes involving top quarks in proton-proton and proton-lead collision

In this thesis, studies about the production of top quarks in proton-proton and proton-lead collisions within the ATLAS experiment at Large Hadron Collider are presented. Two physics analyses are described: the top quark pair with the Higgs boson production and the search for top quark pair production in proton-lead collisions. Both of these topics are very relevant. The measurement of the Higgs boson production is among the key objections of the ATLAS experiment while studies of top quark production within heavy-ion collisions could provide new insights about the quark-gluon plasma.

The main part of the thesis is divided into introduction, five chapters, and conclusion. There is also auxiliary material showing the list of papers the author has contributed, the conference contributions, media appearances, and some additional plots.

The first two chapters provide the introduction to the topic of the thesis. In the initial chapter, the author describes the theoretical basics and the top quark production. Given that the thesis deals with two different processes of top quark production, I'm missing here more detailed comparison of potential top quark production processes, e.g. comparison of the main production processes (single and pair top quark production) with the rare production of top quark (e.g. $t\bar{t}H$). Also, Fig. 1.2 is many years old and much more detailed and recent version exists. In the second chapter, the author describes the LHC accelerator and the ATLAS detector. It is mostly nice description of the detector, especially for the inner detector. On the other hand, there is too brief description of the calorimeter (Sec. 2.3.4) where I'm missing key parameters, e.g. pseudo-rapidity coverage or the energy resolution.

The third chapter describes the author's contribution to b-tagging calibrations within the ATLAS experiment. I think it is a nice attempt to describe the b-tagger algorithms used within the ATLAS. However, I think the calibration procedure is not very well explained for the non-experts and the actual results are not really discussed.

The forth chapter discusses the measurement of $t\bar{t}H$ measurement. This is a very complex analysis. The author focuses on the fakes estimation in two same-sign lepton (e, μ) plus one hadronic tau lepton channel which was author's main contribution to this measurement. The overall description of the measurement is relatively good.

In the fifth chapter is described the measurement of $t\bar{t}$ production in proton-lead collisions. The description here is reasonably detailed. Unfortunately, the figures and tables in Sec. 5.9.1 are not in the order they are mentioned and described in the text, so it is not easy to get oriented in the text.

Overall, I think the description of the results could have been more clear and precise in order to avoid guessing of what the author meant. I have therefore a few main questions and comments to the presented text which I list below:

- It is not clear what was your actual contribution within your ATLAS qualification task. Did you perform the calibrations by yourself by running the jobs and deriving

the scale factors or did you maintain/update the code used for the calibrations or both? Related to that, in Sec. 3.7 and 3.8, the b-tagging scale factors and their uncertainties are plotted but are not discussed at all. It is not clear what is the real purpose of showing the plots. Did you intend to present the actual results of the analysis of measuring the efficiencies or just to demonstrate that with the updated framework you can actually produce these kind of plots? Please, clarify. If it's the first case, then I would be interesting to know e.g. why there are spikes in uncertainties in Fig. 3.9(a) and Fig. 3.10(a) in the 5th and 8th bin while there are no such spikes in Fig. 3.11(a). What has been improved?

- On page 37, you mention 'MV2c10' tagger and latter also present the results for this tagger. However, this tagger was not really defined and explained. I guess it's a special version of 'MV2' tagger. Could you explain what 'c10' means?
- Equations (3.2) and (3.3) both define the weight for the jet. It's not clear why two different formulas are used for the same weight. I can just guess that one is for a b-jet and the other one for a light jet. Could you clarify, please?
- In Sec. 4.7, you mention in the text that in total 877 events are selected in 25 categories, but according to the Fig. 4.4 there are about 1400 events in 2lSS region and about 400 events in 3l region, so these numbers do not match. Can you explain this?
- In Sec. 4.10.2, on page 96, you mention that the template fit method is the nominal method to estimate fakes. However, previously, you described fake factor method for the fakes estimate. It's not clear which method was in the end used in the analysis. Did you do some comparison between these two methods? Or, did you use these two methods for different fakes estimates? Please, clarify.
- Based on Tab. 5.1, $H_T^{l,j}$ has the worst separation power in the dilepton channel and it is not well explained why it is chosen as separating observable (it is just explained why the observable involving MET is not used). Also, $H_T^{l,j}$ is completely missing in the left table for ℓ+jets channel. Could you clarify this?
- Based on description in Sec. 5.7.2 and Tab. 5.4, it's not clear what kind of $t\bar{t}$ modeling uncertainties have been considered. Based on the text, I would expect five different types of uncertainties: parton-shower, matrix-element matching, ISR, FSR, PDF. However, in Tab. 5.4, you mention four different types of uncertainties while using two nuisance parameters for three of them. Could you clarify?
- On page 140, you mention that neither large pulls nor large constraints are observed for the nuisance parameters (NP). I don't think this is a proper statement based on Fig. 5.9 where significant pulls/constraints are shown at least for NPs related to the fakes. You actually mention a pull in the following sentence. Moreover, I think the reason you mention for the pull is actually a reason for the constrain. The reason for pull should be some kind of disagreement between the data and the pre-fit prediction. Therefore, it would be good to see the pre-fit plot and compare to the post-fit plot. Unfortunately, this is not possible because the pre-fit and post-fit plots in Fig. 5.7 and Fig. 5.8 seem to be exactly the same, I guess by mistake.
- What kind of potential improvements would you suggest to both presented measurements ($t\bar{t}$ in p-Pb and $t\bar{t}H$) for LHC Run 3 ($\sim 300/fb$) ?

From a formal point of view, I think there is a relatively large amount of shortcomings and

the thesis would largely benefit from one more careful reading before being submitted. The examples of issues are:

- figures/tables: some are not referenced in the text (e.g. Tab. 4.2, Fig. 5.1); some are swapped (Fig. 4.10 and Fig. 4.11); I think figures in Chap. 2 are taken from some publications or internet sources, so these should be properly cited
- relatively large amount of inaccurate statements, e.g. 'Simulation-to-data scale factors' $\rightarrow$ I believe these are 'data-to-simulation' scale factors; the information in Tab. 2.1 is true only for pp interactions; Fig. 4.1: top,bottom $\rightarrow$ left, right; the last sentence before Sec. 5.5.1 is not precise, since only W+jets and fakes are described
- bibliography: author or journal name is missing (e.g. Ref. [3]-[14],Ref. [160],[162]), titles are not properly listed (e.g. Ref. [2],[69]), collaboration names are not properly spelled out (e.g. Ref. [29]-[36]), Ref. [49] is the same as Ref. [27]
- some shortcuts, labels are not defined (e.g. EV, CDI (page 2), 'EMC' (page 111)) or not properly defined (e.g. in Eq. (1.1), 'B' is used for both baryon number and for 'bottomness')
- repetition of the information, e.g. information in footnote 1 on page 137 was already described in Sec. 5.8, overlap removal information was mentioned in both Sec. 5.4.2 and Sec. 5.4.4
- other small mistakes/typos, e.g. last sentence in Sec. 4.4.2 and next-to-last sentence in Sec. 2.1 are not finished; old information on page 14 ('Currently, there is another planned shutdown from 2019 to 2021.'); typos (page 139: $1.04 \pm 0.9 \rightarrow 1.04 \pm 0.09$, page 17: LHCb $\rightarrow$ LHCf experiment)

The thesis is written in English of a relatively good quality.

In summary, I state that the author described the experimental apparatus, his contribution to b-tagging calibrations within ATLAS experiments, and the procedure and results of two physics analyses. These measurements contribute to knowledge in the particle physics, they are the official results of the ATLAS collaboration, and they were presented as the highlights of the ATLAS collaboration at numerous conferences. The author of the thesis is the primary author of both of these measurements. The thesis met its goals.

Based on the above, I think that even if the formal presentation of the measurements could have been better, the measurements themselves provide a contribution to the given field of science and that the author contributed significantly not only to them but also to the ATLAS experiment. Assuming a successful defense and satisfactory answering to my questions above, I **recommend** that the author is awarded with a PhD degree.

Roman Lysák

August 30, 2024

RNDr. Roman Lysák, Ph.D.
(FZÚ AV ČR)