Dr hab. Anna Kaczmarska
Institute of Nuclear Physics PAN
Radzikowskiego 152
31-342 Krakow
Anna.Kaczmarska@ifj.edu.pl

Kraków, 04.09.2025

**Review of the doctoral dissertation by Saliha Bashir, M. Sc.
entitled "Impact of models for multiparton interactions in proton-proton collisions at LHC energies for the prediction of radiation damage of LHCb silicon trackers" prepared under the supervision of dr hab. inż. Agnieszka Obłąkowska-Mucha**

The dissertation investigates several aspects of radiation damage prediction in one of the LHCb experiment's subdetectors, the Vertex Locator (VELO). Owing to its position closest to the interaction point, the VELO is subject to an exceptionally high flux of traversing particles and is therefore particularly prone to radiation-induced degradation. Predictions and evaluations of radiation damage are typically based on simulations of particle fluence. For such simulations to be reliable, the generated particle multiplicities in proton–proton (pp) collisions, together with their composition, energy spectra, and angular distributions, must be consistent with experimental data. Different physics models implemented in event generators provide different predictions for these quantities, with particularly pronounced discrepancies for low-momentum particles produced at high rapidity. This dissertation presents a comparative study of two general-purpose generators, *Pythia* and *Herwig*, in their estimation of particle multiplicities in minimum-bias events. Furthermore, it examines the feasibility of tuning *Pythia* parameters to achieve improved agreement with data collected by the LHCb experiment during Run 1 and Run 2. This thesis also presents an alternative, data-driven approach to fluence estimation based on minimum-bias events from LHCb Run 2 and Run 3.

The dissertation, written in English, contains seven chapters, one appendix, and a bibliography of 129 items.
The First Chapter provides an introductory discussion of the Standard Model.
Comments to the chapter:
- I was a bit confused to read that "baryons comprising of three colorless quarks".
- In the section describing electroweak interactions, it would be appropriate to include a description of the Brout–Englert–Higgs mechanism, accompanied by proper citations to the original works. Furthermore, when addressing measurements of the CKM matrix elements, results from the Belle experiment should also be mentioned to provide a more comprehensive overview of the current experimental landscape.

Chapter 2 describes the Large Hadron Collider (LHC) and the LHCb experiment. It provides an overview of the LHCb subdetectors, the LHCb upgrade program, and the trigger system at a rather general level. The chapter also outlines the reconstruction procedures for different track types. Finally, it introduces the operating principles of silicon detectors and discusses the mechanisms of radiation damage.
Comments to the chapter:
- As the thesis makes use of data from the LHC Runs 2 and 3, the description of the detector should cover both its pre- and post–Upgrade I configurations.
- The description of track reconstruction lacks a clear indication of the track reconstruction sequence to avoid reconstructing the same track twice.

- Figure 2.14 contains an outdated schedule.

Chapter 3 presents a comparison performed by the Author between two general-purpose event generators, *Pythia* and *Herwig,* using prepared samples of pp collisions at $\sqrt{s} = 7$ TeV and 13 TeV with default configurations. The comparison focuses on key observables for the fluence: energy, momentum, pseudorapidity (η), charged particle multiplicity, and particle composition. As expected, due to the different models used for hadronization and parton showers, significant differences can be observed. The Chapter also contains studies on the impact of modifying generator parameters relevant to multiparton interactions on particle production, as well as a comparison of proton and neutron production.

Comments to the chapter:

- Is there an explanation for why *Pythia* produces more pions than *Herwig* but fewer protons?
- Fig. 3.9 (d): Why does the *Pythia* multiplicity distribution exhibit a two-peak structure, one broad and another peaked at 5?
- *In all comparisons, uncertainties are missing.* Were they estimated?
- *What sample was used for the studies presented in Sec. 3.8?*
- What does the $l \rightarrow \gamma$ process refer to?

In Chapter 4, the tuning of various parameters in *Pythia*, including cross-sections, colour-reconnection schemes, flavour composition, and multiparton interaction parameters, is presented. For this purpose, the *Rivet* and *Professor* software tools are employed to improve the description of Run 1 and Run 2 data relative to the official LHCb tunes. Despite the author's efforts, these studies are not conclusive. The obtained results do not show better agreement with data, and further work is therefore required, also including Run 2 and Run 3 data. Finally, the *Pythia* and *Herwig* samples, generated with the LHCb-specific settings, are compared to the LHCb data at $\sqrt{s} = 7$ TeV. As anticipated, *Pythia* yields the most accurate description of the recorded events.

Comments to the chapter:

- Tab. 4.2: For which $\sqrt{s}$ values are the LHCb tunes presented? What are the motivations for the minimum and maximum values used in the tuning?
- Tab. 4.3: It is not clear why the tuning presented here, including the adjustment of additional parameters, does not achieve at least the level of agreement obtained with the official LHCb tuning. The results almost consistently exhibit poorer agreement with the data. Furthermore, many of the fitted parameters are found to lie close to the boundaries imposed in the fit, which raises concerns regarding the validity and robustness of the fitting procedure.
- Fig. 4.4: It is unclear why the presented values cannot be reproduced using the cross-sections from Tables 4.2 and 4.4 ($\sigma_{tot} - \sigma_{el}$).
- Tab. 4.6 and 4.8: What were the default parameter values and ranges of parameters used in the fit?
- Tab. 4.10: No χ^2/ndof values are provided for those fits - the same for Fig. 4.8.
- Fig. 4.8: It is not clear why the Test_tune, which is based solely on the LHCb plugins, shows better agreement with ATLAS and CMS data than with the LHCb data.
- Fig. 4.12 Are the colour-reconnection (CR) models used with the default parameters, or are the parameters previously optimized for CR1 applied?

Chapter 5 presents charged hadron multiplicity measurement, based on minimum-bias LHCb data collected during dedicated Run 2 periods in 2018. The principal aim of this analysis is to investigate whether reconstructed and identified charged hadrons can be employed as a reliable proxy for estimating the present level of radiation damage in the VELO sensors. The results reveal notable discrepancies between experimental data and MC predictions, particularly in the multiplicity distributions of heavier hadrons. These findings underscore the necessity for further modelling improvements. In the concluding section, the fluence is determined in bins of η and longitudinal position (z) for a subset of VELO sensors. Although this estimation remains preliminary and approximate, it might be valuable for applications such as the online monitoring of the VELO detector, as it can show relative differences.

Comments to the chapter:

- It is not clear what the final selection is. For example, in Table 5.2, it is stated that track χ^2/ndof < 4, while in the text it is < 3. In this table, p < 100 GeV/c is given, whereas in the text, an additional cut on track momentum p > 4 GeV/c is mentioned.
- It is mentioned that to give the best description of data reweighting of both the number of tracks and reconstructed PVs should be used, but it seems that only reweighting on the number of VELO tracks is applied. The closure of Fig. 5.6 (b) is rather poor.
- Why do the produced MC samples exhibit such different μ values compared to the data?
- How good is the PID reproduced by MC? Are there any corrections/scale factors used?
- Fig. 5.9 (a): Please, explain the behaviour of purity for kaons and protons more clearly.
- Are the results in Section 5.4 obtained after reweighting from Eq. 5.1? Why does Fig. 15.3 show much better agreement with data than Fig. 5.6 (b)?. Also, I disagree with the statement about Fig. 5.13: "*There is a difference between the multiplicity of pions, which is underestimated in the MC sample at lower multiplicity, but overestimated at higher multiplicity values. However, for kaons, and protons, there is not a drastic*". I would say it is quite good agreement for pions, but quite poor agreement for kaons and protons.
- It is not stated if the fluence shown in Figs 5.15 and 5.17 is calculated based on MC or data. Also, Fig. 5.15 and 5.17 show the opposite effect of the fluence changes with η.
- "*The values of ϕ_{eq} obtained with the use of reconstructed long tracks are lower than the values obtained in FLUKA simulation*". Those values should be mentioned in the text or a reference given.

Chapter 6 introduces an alternative data-driven approach to estimate the particle fluence based on hit counts within the VELO detector. A detailed description of the VELO detector is included; however, its placement would be more appropriate in Section 2.3. The effects of radiation damage are also discussed. The analysis is based on Run 3 data collected by the LHCb experiment between July and September 2024. For comparison, an MC sample generated with *Pythia* is employed. The number of hits in various detector modules is compared between data and simulation. Additionally, for one module, the temporal decrease in the number of recorded hits observed in the data is shown. Its change over time can be attributed to radiation damage, which progressively reduces the efficiency and sensitivity of the silicon sensors. Finally, the fluence is estimated through the track-length density method and presented as a function of the radial coordinate, in different η intervals and z position of modules.

Comments to the chapter:

- Fig. 6.3 (b): red and black lines are not described.
- Was the pileup reweighting done? MC and data have different μ values.
- The difference between the old and the new definition of fluence needs more explanation.

- Fig. 6.11: What causes these small bumps/kicks in the hit distribution for r close to 2 cm and ~3.5?

Chapter Seven summarises the obtained results.

The thesis editing is poor and requires significant improvements. There are too many mistakes to list in full, but the main issues include the following:

- Phrases and sentences are often repeated, sometimes even within the same paragraph.
- A lot of figures are carelessly prepared, with distributions cut off by too short axes ranges and legends that are unreadable.
- Figure captions are frequently cryptic and inconsistent. Sometimes incorrect (Fig. 6.11).
- Acronyms are repeatedly reintroduced (for instance, MC is defined 11 times).
- Numbers in tables are often left unrounded.
- Many important terms are neither explained nor properly cited, while others are cited excessively—sometimes with different references each time.
- The same figure (the VELO detector schema) appears three times throughout the thesis.
- Some figures (Fig. 1.2, 2.8 (a), 5.17) are never referenced in the text.
- The bibliography is disorganized. Probably, incorrect source file types were used, resulting in corrupted author entries (e.g., *A. Collaboration* instead of *ATLAS Collaboration*).
- The Author often gives references to Ph.D. theses rather than original works.
- Several references are inappropriate (such as the one for the ATLAS experiment), and some are duplicated ([5] and [1], [53] and [52]).

In summary, the dissertation of Saliha Bashir addresses the challenge of tuning the *Pythia* event generator to better describe LHCb data, particularly for variables sensitive to multiparton interactions. The author also compares *Pythia*'s predictions with those of *Herwig*. Although the results did not lead to improvements in the current tuning, the work established a methodology based on advanced techniques that can be applied to further optimisation. The second part of the dissertation introduces an innovative data-driven approach to estimate particle fluence within VELO, enabling fast tracking of radiation damage over time. Although methodologically challenging, this novel strategy demonstrates the possibility of real-time monitoring techniques that can complement and enhance traditional simulation-based methods.

The author's independent contribution is clearly defined. It is also evident that a considerable amount of technical and useful work has been carried out for the experiment, even if the results are not yet conclusive. Such work in large-scale experiments is often unspectacular and challenging, yet it remains essential for collaborative analyses.

Despite the above comments, I conclude that the presented work meets the requirements of a doctoral thesis, and I recommend that Saliha Bashir be admitted to the subsequent stages of the procedure for awarding the doctoral degree.

Anna Kaczmarska
Institute of Nuclear Physics PAN

A—h