dr hab. Paweł Brückman de Renstrom, prof. IFJ PAN
Instytut Fizyki Jądrowej PAN
ul. Radzikowskiego 152
32-342 Kraków

Krakow, 30.08.2025 r.

Referee report on the doctoral thesis of mgr. inż. Saliha Bashir entitled „Impact of models for multiparton interactions in proton-proton collisions at LHC energies for the prediction of radiation damage of LHCb silicon trackers"

The doctoral thesis submitted by Saliha Bashir is devoted to monitoring and predicting radiation damage induced on the VErtex LOcator (VELO) silicon detector of the LHCb spectrometer by particle fluence from the LHC interactions. The harsh radiation conditions in the immediate vicinity of the LHC interaction point is one of the main challenges for design, construction and operation of silicon precision tracking devices. LHCb is a single-arm forward spectrometer with the pseudorapidity coverage of $2<\eta<4.5$. VELO surrounds the interaction point approaching the beam line at the radius smaller than a centimetre. The forward geometry of the detector makes it subject to particularly high radiation doses caused by large particle fluence at high η.

Significant part of the thesis deals with the pivotal issue of contemporary experimental particle physics, i.e. the modelling of collision physics with Monte Carlo (MC) event generators which generally fail to precisely describe the soft component of the particle flow which critically depends on the details of modelling of multiparton interactions (MPI), parton shower and hadronization. This, in turn is pivotal for predicting radiation damage to silicon using simulation tools such as e.g. FLUKA. In particular, PYTHIA and HERWIG generators considered in the thesis differ sizeably in this respect. A challenging attempt to tune PYTHIA parameters in order to better describe data is undertaken by the author. On the other hand, a novel approach of estimating particle fluence directly from data is studied. Neutrons (especially slow ones) are particularly harmful to semiconductors. Yet, they cannot be directly observed but their flux can be estimated from the observed flux of protons by the virtue of isospin symmetry. Hence it is relevant not only to measure or predict inclusive charged particle fluences but also assess them individually per particle species (pions, kaons and protons). Finally, estimation of particle fluence from silicon hit occupancy was investigated. It has an advantage of being highly inclusive but is inherently agnostic to either particle momentum or its identity.

The thesis edited in English on 175 pages consists of seven chapters and appendices. The clarity of the document would benefit from more careful structuring and wording. The author accounts sometimes in quite a detail how certain bits of analysis were technically executed but fails to state why and put them into a wider context of the project. Similarly, the interconnection between the chapters is not sufficiently explained. I'm also missing a general discussion of what accuracy of the fluence estimation we are after. E.g., is the ~10% inaccuracy relevant for the radiation damage estimation? Which effects are the most important? E.g., is PID mismodelling more of an issue than the modelling of the inclusive fluence? Furthermore, bibliography appears rather clumsy. Some references are missing authors, and experimental collaborations are often wrongly quoted or not quoted at all. Some quotations from the text seem to be misplaced and e.g. refs [1] and [5] are duplicates of each other. I also found figures unreferenced from the text.

The first chapter gives a short introduction to particle physics as described by the Standard Model (SM). Still, I find some statements questionable:
p. 18: "hadrons are significantly larger as compared to quarks" - sounds odd given SM assumes quarks to be point-like,
p. 18: "known as exotic quarks" – from the context I guess it should read "exotic hadrons",
p. 19: "The gauge bosons for the strong force act between the quarks and the eight massless, spin-1 particles called gluons (g).",
p. 21: Gell-Mann-Nishijima formula and the Glashow relation are quoted side-by-side in Eq.1.1 and Eq.1.2 without any distinction. Similarly, "isospin" of strong interactions and the "weak isospin" are not told apart and properly defined.
p. 21: Spin 1/2 particles are not vector but spinor fields, while spin 1 gauge bosons are vector fields.
p. 22: Variations around the minimum of the Higgs field are not responsible for fermion and boson masses. It is rather the coupling to the ground state of the Higgs field (vev).
p. 26: It's unclear why ref. [13] is quoted for experimental confirmations of the SM.

Chapter 2 briefly describes the experimental setup, i.e. the LHC accelerator complex and the principal components of the LHCb spectrometer. This is further complemented by a detailed discussion of categories of tracks reconstructed by LHCb and a discussion of radiation damage in silicon sensors. The section is comprehensive, yet I could not fully agree with the statement from p. 32: that particle identification system "differentiate the primary from secondary vertices". Also, the LHCb magnet has a bending power and not the magnetic field of 4Tm (p. 36). Section 2.4 and Fig. 2.14 therein are very outdated, as if were written in the previous decade. Similarly, in Sec. 2.7.3 one finds the sentence: "During Run 3, it is expected that the fluence annually will reach levels equivalent to the combined fluence from Runs 1 and 2.", which sounds as if it was written before Run 3 startup.

Modelling of *pp* interactions by PYTHIA and HERWIG MC event generators and their direct comparison is given in Chapter 3. Section 3.6 compares directly distributions of key kinematical variables issue of the two generators. However, it remains unsaid to what extent the observed differences are relevant for the problem addressed in the thesis. Consequently, it is harder to appreciate the study of the impact of generator settings on such distributions reported in Section 3.7. Besides, "Lattice QCD" cannot be defined as "perturbative calculation", as stated on p. 52.

Chapter 4 undertakes a difficult attempt to improve the MC modelling of inclusive event distributions by fitting selected PYTHIA parameters to LHCb data using the RIVET and PROFESSOR frameworks. I could not find any account of what selections were applied to data & MC to perform the comparisons, how the MC samples were normalised for cross-section studies nor what systematics were considered. At the top of the page 80 the reader is left confused whether the MPI and CR parameters or the alpha_s and the fragmentation functions have the dominant impact on the relevant distributions.
It remains unexplained if the cross-sections quoted in Table 4.2 are tuned directly or rather via the related generator parameters, (p^T_ref , epsilon and alpha_s) introduced in Section 3.7. In the earlier case, one could expect a perfect match to the yields observed in data, yet this does not seem to be the case. On the other hand, such parameterisation would be specific for the COM considered. The table caption should clearly state what COM energy these cross-sections correspond to. Results of the tuning obtained for 7 and 13 TeV COM are given in Table 4.3. The uncertainties are surprisingly large and do not justify the number of significant digits quoted in the table. What processes (plugins) were used in the tuning?
Fig. 4.4 shows σ_{inel}. It should be explained how it relates to quantities from Tables 4.2.& 4.3. σ_{SD} appears twice in the tables. Is it intentional?

At 7 TeV the fitted tune seems to agree perfectly with data while Sim10 doesn't. Yet, the reported χ^2 values are almost identical. At 13 TeV the situation is opposite and here the χ^2 difference is huge. B.t.w, why did the tune depart from data as the result of the fit?
Table 4.6 lists results of the flavour composition tuning for parameters listed in Table 4.4 and using processes given in Table 4.5. The section lacks a conclusion. In particular, the standard LHCb tuning values are not reported.
From Fig.4.5 it is hard to draw the conclusions given at the end of Sec. 4.5. The difference between the tunes is way smaller than their common discrepancy to data.
Some fit results reported in Table 4.10 hit the boundary of the allowed variation range, which is dubious.
The conclusions of MPI tuning based on Figs. 4.7 through 4.11 show generally better data description with the standard LHCb tune. Is there a reason why the dedicated tuning did not improve the picture?
Which tunes were used for the comparisons reported in Section 4.9?
It is unclear what is meant by the “lifetime of a charged track” on p. 101.
On p. 104 one reads: "...while Pythia is in agreement with the data or is within the uncertainty." Isn't it the same? B.t.w., what uncertainties of the model/simulation have been considered?

A valuable study of the novel method to measure multiplicities of different particle species and particle fluence in LHC Run 2 data is reported in Chapter 5, pivotal for the thesis. Yet, the goals are not sufficiently clearly stated. In consequence, it is not evident whether they have been met. Are the investigated Long Tracks representative for the total flux? Is it enough to measure the long track multiplicities and energy spectra, or is it necessary to involve simulation to assess the total particle fluence?
On p. 126 one finds the sentence: "The values of ϕ_{eq} obtained with the use of reconstructed long tracks are lower than the values obtained in FLUKA simulation." But we don't learn by how much. Why are we concerned with comparing data to MC, then? Are these findings relevant? I was hoping to find a discussion of how MC can be used to compensate for the reconstruction and acceptance inefficiencies.
In Table 5.1 one expects an explanation of why the average number of *pp* collisions per bunch crossing are so different in data and MC. It should also be commented on how relevant this is for the flux comparison.
The discussion of data selections in Section 5.3.1 refers to "background". What does "background" to min-bias events mean? Similarly, in Fig. 5.7, it is not explained what "signal" means.
How is the MC on plots in Figs. 5.4 through 5.6 normalized?
Why is the track multiplicity in Fig. 5.6 so different from the one in Fig. 5.5c?
Why aren’t we seeing full closure in Fig. 5.6b after multiplicity re-weighting?
Section 5.3.4 gives the detailed account of how particles are identified in LHCb, but there is no explanation why this is important for the flux analysis. Such justification would best fit in Section 5.1.
Section 5.4 fails to explain what the MC simulation is used for. For sure, data reflect the actual flux in the detector acceptance. I guess one might need MC to correct for signals which are not reconstructed as legitimate tracks. And perhaps one also wants to establish the actual rates of different particle species, in particular protons (hence deduce also neutrons). But this should be clearly stated.
How is MC normalised in Figs 5.10 - 5.14 and what are the uncertainties considered? Are the plots using the re-weighting introduced on p. 114?
I'd expect Fig 5.10a to be very close to 5.4b. Yet, they look quite different. Why?
On p. 124 one reads: "...the number of pions is more than two orders of magnitude higher than protons and kaons", while from the plots it looks more like one order.

Finally, Chapter 6 describes yet another new way of calculating fluences in the early Run 3 data directly from counting silicon hits. Their potential impact on radiation damage in the VELO detector is discussed as well. The proposed method is an interesting alternative to the one discussed in the previous chapter, but the reconstruction of particle energy and identity are somehow lost in the discussion, while earlier these were claimed to be pivotal for the radiation damage.

Section 6.4 comes only towards the end of the thesis bringing a relevant attempt to set the actual goals of the PhD project. As such it should belong to the introduction and is clearly misplaced here.
Section 6.6 gives an alternative definition of the fluence (p. 140). Is this definition different from the one given in Eq. 2.2? Text says about the total distance traversed by ionising particles, yet the plots present Nhits/cm2 which is not equivalent. One reads "fluence is estimated as track-length density, when a track is counted as a hit in the VELO sensor, multiplied by the sensor width" - isn't the track incidence angle missing in the definition?
Why 6.9a and 6.9b differ by ~one order of magnitude while the data and MC samples are claimed the same (100k events)? How is the plot in Fig. 6.10a normalised?
Caption of Fig. 6.11 states "Fluence as a function of energy". How is the energy dependence represented on the plots? What integrated luminosity does this estimation correspond to?

Chapter 7 briefly concludes on the obtained results.

In summary, the theses reflect substantial amount of work concerning both highly nontrivial MC tuning studies as well as direct data analysis carried out in the demanding environment of a big collaboration. It raises an important question of how the particle fluence can be either predicted from MC or estimated from recoded data in view of assessing radiation damage to silicon sensors and proposes own original solutions. Three methods are examined: MC modelling, counting tracks fully reconstructed in the spectrometer and finally counting hits in the silicon detectors. Both the analysis and the dissertation demonstrate independent work of the doctoral student. The research, while relevant gives the impression of work-in-progress state and the author rightly remarks on multiple occasions that certain aspects deserve further studies. The reader might also expect more conclusive statements on the applicability of the results to the problem of precise determination of radiation damage in the LHCb VELO silicon sensors.
Nonetheless, despite the above imperfections, I confirm that Ms Saliha Bashir satisfies the formal requirements for doctoral candidates according to the appropriate Polish regulations (art. 187 Ustawy - Prawo o szkolnictwie wyższym i nauce (z późn. zm.), z dnia 20 lipca 2018 r).
I recommend admission of Ms Saliha Bashir for the subsequent stages of the procedure, including the public defence.

Kind regards

Signature Not Verified
Digitally signed by Paweł Brückman de Renstrom
Date: 2025.09.01 13:59:44 CEST

Paweł Brückman de Renstrom