GOETHE
UNIVERSITÄT
FRANKFURT AM MAIN

Johann Wolfgang Goethe-Universität Frankfurt am Main
Fachbereich 13 Physik • Institut für Kernphysik

Prof. Dr. Alberica Toia
Institut für Kernphysik
Goethe-Universität
Max-von-Laue-Str.1
60438 Frankfurt am Main

Tel. (IKF): +49 (0) 69 798 47078
Tel. (GSI): +49 (0) 6159 712 660
Telefax: +49 (0) 69 798 47100
E-Mail: a.toia@gsi.de

Frankfurt, 15 January 2024

Report on the Ph.D. Thesis „Open charm production at STAR"
by Lukas Kramarik

To whom it may concern

I read with large interest the Ph.D. thesis of Lukas Kramarik on open charm production in the STAR experiment. Heavy-flavor measurements are of great importance in heavy-ion collisions because they are produced in the initial stage of the collision and therefore their proagation in the Quark-Gluon Plasma provides important insights in the properties of QCD at high-energy density. Measurements in p+p collisions are essential as elementary reaction baseline, and provide a powerful test of perturbative QCD, while d+Au collisions, where the presence of hot and dense medium is not expected, provide a reference for the cold nuclear matter effects, due to the presence of the nuclei themselves in the initial stages of the collision and the dense environment after hadronicazion. They are interesting in their own right and important for the understanding and the interpretation of the results in heavy ion collisions.

The study of open charm was carried on by the candidate using the measurement of D^0s mesons. In his thesis, Lukas Kramarik presents the analysis of D^0 via their hadronic decay into an unlike sign pair of charged kaon and pion, performed in the STAR experiment in d+Au collisions at $\sqrt{s}$=200 GeV with the data set collected in 2016. Thanks to the Heavy Flavor Tracker, the excellent impact parameter resolution allowed a good measurement of the topological properties of D mesons decays, used for the signal extraction. The aim of the thesis is to report the pT differential invariant cross section of D^0 in three p_T ranges (1-2, 2-3, 3-5 GeV/c), which overlaps and partially extend the previous STAR measurements which reached only 3 GeV/c. The analysis used machine learning algorithms for the signal extraction, specifically the boosted decision tree, however also the random forest and deep neural network were studied in simulations. These results are compared to the previous d+Au measurements from 2003, and to the results in p+p collisions, resulting in a nuclear modification factor consistent within uncertainties to the one of light hadrons, ie at unity or slightly above unity for $p_T > 2$ GeV/c, or below unity for $p_T < 2$ GeV/c, consistent to the suppression observed in central and peripheral Au+Au collisions.

Campus Riedberg - Max-von-Laue-Straße 1 - 60438 Frankfurt am Main

The candidate contributed to other analyses in STAR (among the others, D^0 and D^* measurements in p+p collisions at $\sqrt{s}$=500 GeV) which are not represented in this document.

The thesis consists of five chapters and it is followed by the candidate's presentations and publications with author's primary contribution. The introduction provides the general context to the activity of Lukas Kramarik during his Ph.D. The first chapter gives an introduction to the relevant field of studies which includes a short qualitative summary of the Standard Model, the elementary particles and their fundamental interactions, QCD physics, the Phase Diagram and the formation of a hot and dense Quark Gluon Plasma in heavy ion collisions, its properties and potential signatures, Cold Nuclear Matter effects in p/d +A collisions, including modification of the Parton Distribution Functions in nuclear matter (nPDF), saturation, Cronin effect, and potential differences in the hadronization mechanisms; more emphasis is given to summarize the existing measurements of charmed hadrons, at RHIC and LHC. Chapter two presents the machine-learning methods used for signal extraction, and shows their performance on the extraction of the D^0 meson: specifically simulations are done with HIJING event generator and transported with a GEANT model of the STAR apparatus, data sets for train, test and validation are prepared and topological variables are defined. The performance of the different algorithms, namely random forests, deep neural networks and boosted decision trees are evaluated with the receiver operating characteristic curves. The candidate, who is also author of a separate publication which details the methods, demonstrates a very good understanding of the field and their application in the reconstruction of a complex topological based reconstruction.
Chapter three is dedicated to the description of the RHIC accelerator and the STAR experiment, in particular of the detectors, with a special emphasis on the Heavy Flavor Tracker, the tracking, the particle identification and the trigger system utilized in the analysis reported in this thesis.
Chapter four provides a detailed description of the analysis, the D^0 reconstruction in d+Au collisions. The main steps, starting from the event and track selection to the raw yield extraction and calculation of the topological variables on which the Boosted Decision Tree algorithm is applied for the reconstruction of the D^0 signal; finally the methods for the calculation of the reconstruction efficiency, using fast simulations to convolute single particle efficiency, as well as for training of the BDT, and full-event simulations and embedding, used to estimate the reconstruction performance in the specific conditions of the data, are well described. The challenge of this analysis is to extract a signal over a large background, which is done using machine-learning algorithms. For this is crucial to verify that test and training distributions do not have significantly different shapes, that allows to select a reliably significant cut for signal and background. Finally the peak is fit by the sum of Gaussian and linear functions and the yield is extracted by integration of the fit function or by bin counting. Efficiency corrections are described in great detail, and all contributions are presented, together with the systematic uncertainties, which are dominated by the response of the Boosted Decision Tree and the raw yield extraction. Overall the analysis is robust and sound and the candidate demostrates to master it well. The last sub-section finally presents the invariant yield and the nuclear modification factor of D^0 meson and compares to the previous publiaction of the d+Au data (2003), to central ad peripheral Au+Au data, to ALICE measurements of p+Pb. The p_T-differential cross sections in d+Au collisions are compatible with the fit to the p+p distribution and with the previous publication of 2003 within uncertainties. The results presented in this thesis extend the previous results to the high pT region and have significantly smaller uncertainties than those measured in 2003. These data also serve as reference for estimating the Cold Nuclear Matter effects with the nuclear modification factor R_{AA}, which is around unity or sligthly above for $p_T > 2$ GeV/c showing a consistent trend with light hadrons, or below unity for $p_T < 2$ GeV/c, as in central and peripheral Au+Au collisions, suggesting that CNM effects could explain the R_{AA} at low p_T.

This thesis is a concise document, overall well written. The candidate demonstrates a solid expertise in the different parts of the analysis, and a good understanding of heavy flavor physics. This is a challenging analysis that requires a very good understanding of the data sets, a good knowledge of the various sub-detectors involved in the analysis, and their perfomance during data taking, and the machine-learning techniques employed for the signal extraction. Given the large amount of background, and the employ of ML methods, the assessment of the systematic uncertainties is challenging. The achieved extension of the measurements to the higher momentum range is a valuable result.

I suggest in the following few points that would benefit from a further discussion:
- in general a bit more discussion on the interpretation of the results would be appreciated: what do we learn from the agreement with light hadrons at $p_T > 2$ GeV/c? What do we learn from the ageement with Au+Au data for $p_T < 2$ GeV/c? And besides the comparison with other data sets, can the data be compared to theoretical calculations? Can we learn something more quantitative about CNM?
- given the existence of another data sets of d+Au collisions at $\sqrt{s}$=200 GeV, the direct comparison of the two data sets would benefit from more details, eg: what is the event statistics of the data sets? Why do the results of this thesis not extend to the low p_T region? What is the agreement of the results in the p_T region common to the two data sets? I think especially this point should be addressed critically. The author comments that the systematic uncertainty is much reduced compared to the old data set: what is the contribution with the main impact?
- the usage of machine learning is cetainly interesting and innovative. It would be nice though to see a comparison with a traditional analysis with individual cuts, what it improves, what is more challenging, ...
- in the assessment of the systematic uncertainties, which fraction of the uncertainty is common to the analysis in the different systems p+p, d+Au, Au+Au? How much does it cancel in the R_{AA} calculation;
- some quantitative comparison of the D^0 mesons with the non-photonic electrons would be desirable.

To conclude, I recommend this thesis to be defended. The work of Lukas Kramarik provides new measurements of D^0 meson reconstruction in the RHIC energy regime, allowing a quantitative test of pQCD calculations and of energy loss models in strongly interacting nuclear matter.

Your sincerely,
Prof. Dr. Alberica Toia

Campus Riedberg - Max-von-Laue-Straße 1 - 60438 Frankfurt am Main