



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

Kraków, 12.04.2021 r.

**Recenzja rozprawy doktorskiej mgr inż. Adama Dendka zatytułowanej  
„Machine learning based long-lived particle reconstruction algorithm for  
Run 2 and upgrade LHCb trigger and a flexible software platform for the  
UT detector read-out chip emulation”**

Pan mgr inż. Adam Dendek przedstawił do recenzji rozprawę doktorską zatytułowaną „Machine learning based long-lived particle reconstruction algorithm for Run 2 and upgrade LHCb trigger and a flexible software platform for the UT detector read-out chip emulation”, podsumowującą pracę jaką wykonał w ramach współpracy LHCb, a dotyczącą oprogramowania systemu odczytu oraz szybkiej rekonstrukcji w modernizowanym systemie detekcji śladów cząstek naładowanych eksperymentu. Główne dwa projekty zrealizowane w ramach doktoratu to opracowanie opartego na uczeniu maszynowym szybkiego algorytmu rekonstrukcji rozpadów cząstek długożyciowych, pracującego w czasie rzeczywistym w ramach systemu wyzwalania oraz oprogramowanie do odczytu i monitorowania danych surowych z przyszłego paskowego detektora krzemowego Upstream Tracker (UT).

Eksperyment LHCb to wysoko wyspecjalizowany jednoramienny spektrometr „do przodu” pokrywający obszar pseudopospieszności od około 2 do 5, pozwalający na precyzyjną detekcję i ekskluzywną rekonstrukcję rozpadów ciężkich hadronów, w tym rozpadów kaskadowych z udziałem długożyciowych hadronów takich jak mezon  $K^0$  czy barion  $\Lambda^0$ . Fizyka zapachu jest jednym z kluczowych narzędzi do precyzyjnej weryfikacji i pomiaru parametrów Modelu Standardowego cząstek elementarnych (MS), zarazem umożliwiającą pośrednie poszukiwanie zjawisk wykraczających poza MS poprzez obserwację odstępstw od przewidywań tej ostatniej. LHCb korzystając z wysokiej świetlności zderzacza LHC nie podlega przy tym ograniczeniu do produkcji par podstawowych (naładowanych lub neutralnych) mezonów  $B$ , typowemu dla fabryk  $B$ , czyli zderzaczy  $e^+e^-$  pracujących w rezonansie  $Y(4S)$ .

Poza systemem kalorymetrów, komór mionowych oraz rozbudowanym systemem detektorów typu RICH służących do identyfikacji cząstek, sercem spektrometru LHCb jest krzemowy system detekcji śladów naładowanych. Patrząc od strony punktu oddziaływania, składa się on kolejno z pikselowego detektora wierzchołka (VELO) oraz dwóch krzemowych spektrometrów paskowych (stacje TT oraz T1-3) usytuowanych odpowiednio przed i za magnesem zakrzywiającym.

To właśnie szybkie rozpoznawanie i rekonstrukcja śladów w krzemowych detektorach paskowych jest tematem pierwszego z przedstawionych projektów. Oprogramowanie opracowane przez autora stanowi element systemu wyzwalania wysokiego poziomu (HLT). Stąd, poza oczywistym wymaganiem wysokiej wydajności rekonstrukcji oraz minimalnego udziału błędnie zrekonstruowanych trajektorii (ghost tracks), algorytm taki musi być szybki, aby nie obciążać budżetu czasu obliczeniowego systemu wyzwalania. Stąd, autor przeprowadził wnikliwą analizę możliwości użycia algorytmów uczenia maszynowego (ML). ML staje się coraz powszechniejsze w złożonej analizie danych, gdzie klasyfikacja lub regresja oparta jest o wielowymiarową przestrzeń, często nietrywialnie skorelowanych zmiennych. Metody uczenia maszynowego znane są ze swojej skuteczności oferując bliską optymalną wydajność. Raz nauczonego algorytm, jest przy tym egzekwowany szybko co stanowi wielki atut w zastosowaniu do układów czasu rzeczywistego. Niemniej, skuteczność ML zależy od dobrego dopasowania architektury algorytmu uczącego się (tzw. hyperparameters) i samo w sobie stanowi odrębną dziedzinę wiedzy, żeby nie powiedzieć sztuki.



# INSTYTUT FIZYKI JĄDROWEJ im. Henryka Niewodniczańskiego POLSKIEJ AKADEMII NAUK

Drugim z przedstawionych projektów jest oprogramowanie służące do zbierania i monitorowania danych surowych z nowego detektora paskowego Upstream Tracker (UT) który, w ramach prowadzonej obecnie modernizacji eksperymentu LHCb, zastąpi stację TT. Zostało ono zaimplementowane w ramach oficjalnego pakietu oprogramowania eksperymentu t.j. szkieletu *Gaudi*. Jak dotąd posłużyło do analizy danych zbieranych podczas testów prototypowych detektorów UT na wiązkę, a w przyszłości ma stanowić element monitorujący poprawność działania i jakość danych podczas regularnych sesji naświetlań (physics runs). Testy prototypowych modułów UT na wiązkę wykorzystywały inną niż docelowa elektronikę odczytu (front-end) która, w szczególności, pozbawiona była bloku procesowania sygnału cyfrowego (DSP). Funkcje tego ostatniego pełniła emulacja napisana przez autora. To zadanie doskonale wpisało się w tematykę badań grupy AGH która, między innymi, jest autorem projektu docelowego układu scalonego odczytu UT, SALT.

Przedstawiona do recenzji rozprawa pana mgr inż. Dendka składa się z krótkiego wprowadzenia, sześciu głównych rozdziałów, podsumowania oraz bogatej bibliografii zawierającej 141 pozycje. Praca napisana jest w języku angielskim z dodatkiem zwięzłego streszczenia w języku polskim.

Rozdział 1 zawiera opis układu eksperymentalnego którego dotyczy rozprawa doktorska. Rozpoczyna go krótki wstęp na temat laboratorium CERN i opis zespołu akceleratorów ze szczególnym uwzględnieniem Wielkiego Zderzacza Hadronów (ang.: LHC), wraz z dyskusją podstawowych parametrów zderzeń proton-proton w eksperymencie LHCb. Centralna część rozdziału poświęcona jest opisowi spektrometru LHCb w jego obecnym kształcie, dokładniej do końca tzw. Run 2. Przedstawione są wszystkie budujące go podsystemy, a to krzemowy system detekcji śladów (pikselowy VELO, oraz paskowe TT przed i T1, T2, T3 za magnesem zakrzywiającym) uzupełniony gazowym detektorem dryfowym OT, magnes zakrzywiający, podwójny system detektorów RICH służących do identyfikacji cząstek naładowanych, kalorymetry i wreszcie system wyzwalania (ang.: trigger). Opis tego ostatniego jest szczególnie rozbudowany, jako że główna część projektu doktorskiego dotyczy szybkiej rekonstrukcji śladów stanowiącej integralną składową systemu wyzwalania wysokiego poziomu (ang.: HLT). Rozdział wieńczy dyskusja przedstawiająca motywację oraz zakres modernizacji spektrometru przygotowujących eksperyment do Run 3. Znaleźć tam można między innymi opis przyszłego detektora UT oraz architektury zaprojektowanego przez zespół AGH układu scalonego front-end SALT.

Rozdział 2 poświęcony jest w całości głównym aspektom programu fizycznego LHCb. Począwszy od dyskusji podstawowych symetrii przyrody, poprzez krótki opis Modelu Standardowego, koncentruje się na przedstawieniu sektora oddziaływań słabych odpowiedzialnych za mieszanie kwarków i łamanie symetrii ładunkowo-przestrzennej CP.

Rozdział 3 zawiera rozbudowaną dyskusję algorytmów uczenia maszynowego. To bardzo szybko rozwijająca się dziedzina informatyki znajdująca coraz szersze zastosowania również w eksperymentalnej fizyce cząstek. Najkrócej, zamiast gotowego przepisu na rozwiązanie danego problemu, algorytm uczy się go rozwiązywać na podstawie przykładów. Poza wprowadzeniem pojęć niezbędnych w dalszej części pracy, rozdział ten posiada niekwestionowany walor dydaktyczny. Czytelnik dowiaduje się o podstawowych gałęziach rozwoju uczenia maszynowego, jego historii i zastosowaniach. Sporo miejsca poświęca autor teorii oceny metod uczenia maszynowego, w tym takim pojęciom jak macierz błędu czy krzywa ROC (Receiver Operating Characteristics) oraz znaczeniu ROC AUC. Kolejnymi zagadnieniami są optymalizacja procesu uczenia, w szczególności optymalizacja architektury trenowanego algorytmu, na przykład optymalizacja bayesowska. W rozdziale omówione są cztery rodzaje algorytmów uczących się wykorzystanych następnie do analizy problemu rekonstrukcji śladów naładowanych w detektorze a konkretnie do klasyfikacji poprawnie zidentyfikowanych śladów w odróżnieniu od śladów będących wynikiem błędnie przypisanych punktów pomiarowych. Autor przeanalizował klasyfikatory *k-Nearest Neighbours* (kNN), *Logistic Regression* (LR), *Boosted Decision Trees* (BDT) oraz *Deep Neural Network* (dNN).

Rozdział 4 stanowi rdzeń rozprawy. Rozpoczyna go szczegółowe omówienie procedury rekonstrukcji śladów w eksperymencie, w szczególności tzw. 'downstream tracks', czyli śladów rozpoczynających się za detektorem VELO w rezultacie rozpadów cząstek długożyciowych. Autor przeprowadza czytelnika przez kolejne etapy rekonstrukcji śladu począwszy od tzw. załączka poprzez kojarzenie kolejnych sygnałów w detektorze po pomiar parametrów kinematycznych i szacowanie jego wiarygodności. Ze względu na duże tło kombinatoryczne kluczowym jest poprawna kategoryzacja powstałych kandydatów na ślady cząstek. Co więcej, w przypadku rekonstrukcji w czasie rzeczywistym HLT, trzeba to robić szybko. To zadanie autor powierzył algorytmom uczenia maszynowego



# INSTYTUT FIZYKI JĄDROWEJ im. Henryka Niewodniczańskiego POLSKIEJ AKADEMII NAUK

wykorzystującym podstawowe parametry kinematyczne,  $\chi^2/\text{ndf}$  liczbę skojarzonych sygnałów itp. Optymalizacji pod względem uzyskanej separacji poprawnych od błędnie zrekonstruowanych śladów jak i zbieżności samego procesu uczenia poddane zostały cztery w/w klasyfikatory. Na tej podstawie, algorytm dNN został wybrany jako najefektywniejszy. Własności nowego algorytmu rekonstrukcji (Long-Lived Tracking) takie jak wydajność oraz skuteczność odrzucania błędnych kandydatów została szczegółowo przeanalizowana tak na danych symulowanych jak i danych rzeczywistych ze zderzeń wykorzystując zrekonstruowane rezonanse  $K_s^0$  oraz  $\Lambda^0$ . W obu przypadkach autor zaobserwował znaczącą poprawę w stosunku do dotychczasowego algorytmu (PatDownstream). Co ważne, uzyskano bardzo istotne (45%) przyspieszenie działania algorytmu rekonstrukcji. Autor zaproponował też dalsze kierunki możliwej poprawy poprzez zastosowanie innej architektury sieci (RNN) lub zmodyfikowanych funkcji straty.

Rozdział 5 zawiera zwięzły opis obiektowo zorientowanego oprogramowania przygotowanego przez autora do odczytu prototypowych modułów UT. Detektory te nie posiadały jeszcze projektowanego dla nich układu scalonego SALT, a realizowane przez niego algorytmy na etapie procesowania sygnału cyfrowego (ang.: DSP) zostały przeniesione do warstwy softwarowej zaimplementowanej w szkielecie *Gaudi* stanowiącym środowisko oprogramowania eksperymentu LHCb. Rozdział rozpoczyna omówienie podstaw działania półprzewodnikowych detektorów krzemowych wraz z dyskusją procesów rządzących oddziaływaniem cząstek naładowanych z materią. Emulacja działania bloku DSP realizuje wyznaczanie i odejmowanie piedestałów, wyznaczanie poziomu szumu wspólnego oraz indywidualnych kanałów odczytu, wreszcie formowanie klastrów sygnałowych.

Rozdział 6 opisuje analizę danych prototypowych modułów UT z wykorzystaniem oprogramowania przygotowanego przez doktoranta. Po opisanie układu eksperymentalnego i podstawowych etapów analizy danych przedstawiona są wyniki analizy rozkładu ładunku zebranego na dwóch sąsiadujących paskach odczytu. Wyznaczenie zależności tzw. „charge sharing” od punktu przejścia cząstki i konta w stosunku do płaszczyzny detektora pozwala na optymalizację rozdzielczości przestrzennej detektora.

Rozdział 7 zawiera krótkie podsumowanie rozprawy.

Praca jest poprawnie zredagowana, a organizacja materiału jest w zasadzie przejrzysta, choć n.p. ostatni akapit podrozdziału 4.2 wolałbym zobaczyć w rozdziale 2 dyskutującym program fizyczny LHCb. Czasami, jednak czytelnik odnosi wrażenie jakby zabrakło końcowej korekty edycyjnej i dotyczy to zarówno licznych drobnych literówek jak i pewnych zaniedbań redakcyjnych. Rozprawa zawiera sporo informacji, po części o charakterze ogólnym dotyczącym technik eksperymentalnych, fizyki oddziaływań słabych czy współczesnych metod uczenia maszynowego. Całość jest opatrzona bogatą bibliografią. Szkoda więc że nie jest ona uporządkowana w kolejności odwołań, co jest dosyć standardowym oczekiwaniem w stosunku do tego typu dokumentów. Szkoda też, że zabrakło w niej publikacji Emmy Noether z 1918 roku, której twierdzenie stanowi kanwę rozważań rozdziału 2. Również w wykazie publikacji na wstępie rozprawy pierwsza pozycja nie jest datowana. Zaś z wykazu kontrybucji konferencyjnych trudno się zorientować jaki był ich charakter.

Rozdział 4 rozprawy nie wyjaśnia, dlaczego optymalizacja uczenia maszynowego dotyczy jedynie architektury algorytmu uczącego, a nie rozpoczyna się od szerszej dyskusji zmiennych wejściowych. Zmienne użyte do dyskryminacji wymienione są na początku podrozdziału 4.7, ale zasadność tego wyboru nie jest skomentowana. Dla przykładu, zmienne  $p$ ,  $p_T$ ,  $t_x$  i  $t_y$  wydają się być w 100% zależne. Podobnie, użycie zmiennych takich jak  $p_T$  czy  $\eta$  może skutkować niejednorodnością wydajności rekonstrukcji w funkcji tych parametrów kinematycznych. Efekt taki można pod pewnymi warunkami zaakceptować, niemniej sam fakt zasługuje na wzmiankę. W podrozdziale 4.13 wymienione są zmienne wejściowe do głębokiej sieci neuronowej (dNN) uznanej za najefektywniejszy algorytm. Lista ta różni się nieco od wzmiankowanej powyżej, dodatkowo większość zmiennych jest zlogarytmowana. Ten fakt nie jest w żaden sposób skomentowany. Swoją drogą, bezpośrednie porównanie i sposób wyłonienia najlepszej architektury zostało przemilczane.

Czy znane jest źródło naprzemiennej struktury przesłuchu pomiędzy sąsiednimi kanałami widocznej na rysunku 6.3.1? To nie uwaga do autora, raczej ciekawość recenzenta, która nie została zaspokojona w tekście rozprawy. Formuła funkcji błędu we wzorze 6.4 jest niepoprawna (granice całkowania oraz brak pierwiastka z  $\pi$ ).

Niektóre rysunki są niedostatecznie opisane. Dla przykładu, nie dowiadujemy się co oznaczają czerwone i czarne punkty na rys. 1.4.4, a rysunek 3.3.7 jest nie tylko niezrozumiale opisany, ale wydaje się nie znajdować żadnego odniesienia w tekście. Rysunek 4.7.1 jest zupełnie nieczytelny (rozdzielczość),



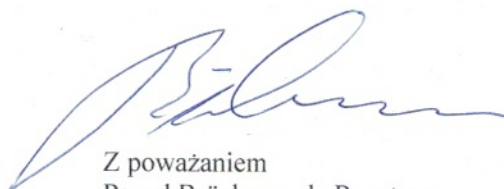
INSTYTUT FIZYKI JĄDROWEJ  
im. Henryka Niewodniczańskiego  
POLSKIEJ AKADEMII NAUK

a rys. 4.7.3 jest trudno czytelny ze względu na rozmiar opisów na osiach, a punkty pomarańczowe najprawdopodobniej przykrywają punkty niebieskie na rozkładach dwuwymiarowych czyniąc je mało użytecznymi. Najmniej fortunnie ma się rzecz z rysunkami w rozdziale 5. Tabelki z informacją statystyczną na rysunku 5.3.5 są nieczytelne, a podpis pod rysunkiem 5.3.6 powinien tłumaczyć różnice pomiędzy zamieszczonymi panelami. Niejasnym jest, dlaczego rys. 5.3.7 pokazuje akurat sześć bardzo podobnie wyglądających rozkładów, bo ich tabelki z informacją statystyczną są nieczytelne. Podobnie, podpis pod rysunkiem 5.3.9 jest szczątkowy i nie wyjaśnia znaczenia prezentowanych wyników. Dodatkowo, rysunki 5.3.8 i 5.3.9 posiadają referencje do publikacji [11] i nie jest jasne czy zostały otrzymane jako produkt omawianego pakietu monitorującego dane surowe UT.

Drobnych literówek czy potknięć językowych nie wymieniam. Napomknę tylko o jednej. Fragment tytułu rozprawy brzmiący „...and upgrade LHCb trigger” jest niegramatyczny. Spodziewałbym się albo „upgraded LHCb trigger” albo „upgrade of the LHCb trigger” w zależności od faktycznej intencji autora.

Powyższe uwagi nie mają, jednakże, wpływu na pozytywną ocenę pracy, która dokumentuje rzetelnie zrealizowane projekty z dziedziny uczenia maszynowego oraz programowania w czasie rzeczywistym mające bezpośrednie i wartościowe zastosowanie w eksperymencie LHCb.

Podsumowując, pan mgr inż. Adam Dendek zrealizował dwa autorskie projekty polegające na napisaniu pakietów oprogramowania z zastosowaniem do akwizycji danych z krzemowych detektorów paskowych oraz szybkiej rekonstrukcji śladów naładowanych w eksperymencie LHCb. Rezultaty obu projektów już znalazły zastosowanie w eksperymencie służąc bądź zbieraniu i monitorowaniu danych z testów na wiązce, bądź stanowiąc integralny element usprawnionego systemu wyzwiania wysokiego poziomu. Dokumentując swoją pracę doktorant wykazał się dogłębnym zrozumieniem złożoności współczesnych metod uczenia maszynowego i znajomością programu fizycznego eksperymentu, wreszcie precyzyjnie wyjaśnił zasadę działania półprzewodnikowych detektorów krzemowych wraz ze szczegółami funkcjonowania elektroniki odczytu front-end oraz wstępnej obróbki danych surowych. Nie mam wątpliwości, że przedstawiona praca spełnia wymagania stawiane pracom doktorskim i wnioskuję do Rady Dyscypliny Nauki Fizyczne AGH o dopuszczenie pana mgr inż. Adama Dendka do dalszych etapów przewodu doktorskiego.



Z poważaniem  
Paweł Brückman de Renstrom