

prof. dr hab. inż. Krzysztof Korcyl

Kraków, 15.11.2024

Oddział Fizyki i Astrofizyki Cząstek

Zakład Fizyki Procesów Dyfrakcyjnych

Instytut Fizyki Jądrowej PAN

## RECENZJA

Przedmiotem recenzji jest praca doktorska mgr Michała Mazurka pt. „New simulation software and machine learning technologies in the LHCb experiment to evaluate physics performance of Run 3” przygotowana pod kierownictwem promotora prof. Wojciecha Wiślickiego oraz dr Wojciecha Krzemienia jako promotora pomocniczego w Zakładzie Fizyki Wielkich Energii w Departamencie Badan Podstawowych Narodowego Centrum Badań Jądrowych w Świerku.

Praca liczy 85 stron, 195 pozycji literatury (większość odnosząca się do publikacji wyników analiz fizycznych) oraz 4 załączniki zawierające głównie histogramy z prac własnych autora dodatkowo ilustrujące wyniki prowadzonych badań i pomiarów. Praca jest napisana w języku angielskim.

W rozdziale 1 doktoratu przedstawione zostały pokrótce ogólne założenia eksperymentów fizyki wysokich energii. Autor omawia rolę modeli teoretycznych oddziaływań cząstek elementarnych, znaczenie pomiarów eksperymentalnych, lecz szczególną uwagę poświęca symulacjom numerycznym poszczególnych składników doświadczenia fizycznego w fizyce cząstek. W części dotyczącej Modelu Standardowego, czyli matematycznego sformułowania oddziaływań cząstek elementarnych bardzo skrótowo, lecz poprawnie wprowadzone zostają pojęcia cząstek elementarnych i omówione ich typy. Następnie opisana została rola różnego rodzaju detektorów oraz sposobu zbierania i gromadzenia informacji o rejestrowanych sygnałach oddziaływań cząstek. Wspomniane zostały instalacje eksperymentalne z różnych ośrodków naukowych, nie tylko te z CERNu. Rola symulacji numerycznych opisana została w rozdziale 1.3. Algorytmy takie, potocznie nazywane Monte Carlo z powodu korzystania w swoich założeniach z liczb pseudolosowych, pozwalają na zrekonstruowanie oczekiwanych końcowych rozkładów statycznych zarówno modeli teoretycznych (generatory zdarzeń) jak i odpowiedzi różnych elementów podzespołów detektorowych (symulatory detektorów). Przedstawione ogólne wprowadzenie teoretyczne stanowi punkt wyjścia do dyskusji osiągnięć Autora omówionych w kolejnych rozdziałach. Bardziej pogłębiona dyskusja procesów fizycznych o szczególnym znaczeniu dla detektora LHCb znajduje się w rozdziale 2.3. Autor koncentruje się na rozpadach mezonów B i D poprzez które możliwe jest zmierzenie wartości podstawowych parametrów Modelu Standardowego takich jak np. elementy macierzy CKM. Opisuje również rozpady rzadkie oraz wspomina o spektroskopii hadronów z uwzględnieniem nowych stanów egzotycznych zarejestrowanych przez LHCb. Całość pokazuje podstawowe rozeznanie Autora w zagadnieniach teoretycznych fizyki cząstek elementarnych oraz konstrukcji detektorów.

Rozdział 3 poświęcony jest oprogramowaniu wykorzystywanemu do symulacji procesów fizycznych i symulacji interakcji wygenerowanych cząstek z materiałem detektorów. Platforma programistyczna GAUSS (framework) używana w eksperymencie LHCb w naświetlaniach Run 1 oraz 2 okazała się być wąskim gardłem w szybkości generowanych przypadków przy planowanym zwiększeniu świetlności LHC w Run3 i zwiększeniu liczby równoczesnych oddziaływań w czasie przecinania się wiązek akceleratora. Prace w których Autor uczestniczył polegały na wprowadzeniu pośredniej warstwy oprogramowania GAUSSINO, która umożliwiła odseparowanie zależności związanych z detektorem LHCb od czystych symulacji fizycznych. Wprowadzenie GAUSSINO wymagało rozwiązania szeregu problemów wynikających z uzgodnienia interfejsów i synchronizacji przypadków z programami generującymi przypadki, zajmującymi się rozpadem cząstek, wprowadzaniem szumu, nakładaniem się przypadków itp.. Dodatkowa warstwa umożliwiła wprowadzenie i wykorzystanie nowoczesnych technologii oprogramowania jak wielowątkowości oraz uczenia maszynowego do prowadzenia tzw. szybkich symulacji wykorzystujących sparametryzowane odpowiedzi detektora. Imponująco wygląda zwiększenie przepustowości przetwarzania przypadków w kalorymetrze dla szybkich symulacji wykorzystujących techniki uczenia maszynowego w porównaniu do detalicznych symulacji opartych o pakiet GEANT4. Osobnym zagadnieniem jest przygotowanie danych do etapu uczenia modelu sieci neuronowej, w którym bardzo pomocna jest wizualizacja generowanych danych. W zakończeniu rozdziału 3 Autor omawia prace związane z interfejsowaniem platformy GAUSSINO do różnych środowisk graficznych umożliwiających śledzenie jakości generowanych danych.

Podsumowując rozdział trzeci można stwierdzić, że zakres tematyczny i opisanie szczegółowych aspektów omawianego przebudowania oprogramowania symulacyjnego dla wprowadzenia dodatkowej warstwy umożliwiającej wykorzystanie nowoczesnych technik programowych potwierdza umiejętności Autora do samodzielnego prowadzenia pracy naukowej. Samodzielność ta widoczna jest także w formie treści opisującej badane zagadnienie. Przypomina ona techniczne spotkania grupy współpracujących osób, które używają sobie znanych skrótów i założeń świadczących o głębokiej znajomości omawianego zagadnienia. Z drugiej strony forma ta jest pewnym utrudnieniem dla osoby spoza zainteresowanego kręgu – np. recenzenta.

Najciekawszym rozdziałem jest rozdział 4, w którym Autor prezentuje oryginalne rozwiązania unowocześnienia oprogramowania symulacyjnego wykorzystujące nowoczesne techniki uczenia maszynowego. Planowane unowocześnienie LHC wiąże się z niespotykanym dotychczas wzrostem ilości danych i potrzebnej mocy obliczeniowej. Nie opracowano dotychczas jednolitego sposobu obsługi tych wyzwań. Wszelkie próby wykorzystania nowoczesnych technik obliczeniowych z wykorzystaniem uczenia maszynowego przede wszystkim noszą znamiona oryginalności i otwierają pole do interesujących badań i eksperymentów.

Autor omawia w pracy trzy miejsca w których uczenie maszynowe zostało użyte i prezentuje uzyskane wyniki: zintegrowanie platformy GAUSSINO ze środowiskami uczenia maszynowego dla szybkich symulacji, wykorzystanie generatywnej sztucznej inteligencji do produkcji przypadków służących do symulacji odpowiedzi detektora oraz wykorzystanie algorytmów rozpoznawania obiektów z uczenia maszynowego do rekonstrukcji klastrów w kalorymetrze. Dwie platformy uczenia maszynowego PyTorch oraz ONNXRuntime zostały wykorzystane do zademonstrowania działającego interfejsu do GAUSSINO wspierającego przetwarzanie wielowątkowe. Interfejs ten został następnie wykorzystany w uruchomieniu modelu wariacyjnego autoenkodera (VAE) do generowania depozytów energii w kalorymetrze. Autor przedstawia dwa parametry, całkowitą energię oraz liczbę hitów, które zostały dodane do VAE aby ustabilizować przetwarzanie po dekodrze. Pozwoliło to uzyskać lepszą

rekonstrukcję depozytów energii w porównaniu do klasycznego VAE. Po wytrenowaniu modelu dokonano jego weryfikacji, która polegała na porównaniu wyników symulacji z pełną symulacją GEANT4, gdzie uzyskano zgodność zrekonstruowanej energii przy około 8-mio krotnie zwiększonej częstotliwości generowanych przypadków.

Ponieważ rekonstrukcja klastrów w kalorymetrze pochłania około 25% czasu symulacji na CPU, Autor przedstawia studia nad wykorzystaniem konwolucyjnych sieci neuronowych do rozpoznawania obiektów w mapach hitów rejestrowanych w detektorach. W badaniach został wykorzystany model sieci neuronowej realizujący algorytm YOLOv3 z wykorzystaniem zdeponowanej energii jako dodatkowym parametrem opisującym prostokąty ograniczające, który był trenowany na  $10^5$  przypadkach minimum-bias. Testy zostały przeprowadzone na próbce 2000 przypadków, a wyniki porównane z klasycznym algorytmem rozpoznawania klastrów CELLULAR AUTOMATION. Wyniki pokazują lepszą wydajność YOLOv3 w rozpoznawaniu klastrów, która słabiej radzi sobie z rozpoznawaniem energii generując dodatkowo nieistniejące klastry (ghost) lub nie rekonstruując wszystkich (missed). W części końcowej rozdziału Autor wskazuje na znaczenie odpowiedniego przygotowania próbek wykorzystywanych w trenowaniu sieci i możliwym wykorzystaniu uniwersalności GAUSSINO, które może być użyte w produkcji próbek trenujących dla modeli pracujących dla bardzo prostych zestawów detektorów ale równocześnie bardziej złożonych konstrukcji. W pierwszym przypadku można skoncentrować się na badaniu wpływu zewnętrznych ograniczeń (geometria detektorów, środowisko).

Podsumowując zawartość rozdziału czwartego można stwierdzić że Doktorant zaproponował i przeprowadził badania oryginalnego rozwiązania problemu zwiększenia wydajności oprogramowania symulacyjnego LHCb poprzez wykorzystanie nowoczesnych technik obliczeniowych opartych o elementy uczenia maszynowego.

Ostatni rozdział piąty poświęcony jest analizie fizycznej wyników symulacji uzyskanych z wykorzystaniem szybkich symulacji opartych o elementy uczenia maszynowego z wynikami z dokładnych symulacji opartych o GEANT4. Do porównania zostały wybrane 4 kanały rozpadu mezonu B w dwóch scenariuszach symulacyjnych. Pierwszy scenariusz pokazuje wyniki porównania rozkładów masy niezmienniczej mezonu B oraz rozkładów momentu produkowanych cząstek z symulacji wykorzystujących nieskalibrowany model oparty o wariacyjny autoenkoder. Wykorzystując n-tuple budowane z rekonstruowanych wielkości Autor dochodzi do wniosku, że widoczne różnice są efektem słabego wytrenowania modelu autoenkodera i w kolejnym scenariuszu symulacyjnym przedstawia wyniki uzyskane dla skalibrowanego modelu VAE, w którym zostały użyte dwa dodatkowe parametry modyfikujące wcześniej zaproponowane rozszerzenie VAE. Wyniki drugiego scenariusza poprawiają wyniki z szybkich symulacji i praktycznie zgadzają się z wynikami z symulacji detalicznych. W konkluzji tego rozdziału i całej pracy Autor zwraca uwagę na wpływ jakości modelu, odpowiedniego przygotowania próbek treningowych i precyzyjnego dostrojenia modelu na uzyskanie wysokiej jakości i wydajności procesu symulacji wykorzystującego elementy uczenia maszynowego.

Przedstawiony materiał świadczy o dobrej znajomości zagadnień teoretycznych fizyki cząstek elementarnych, potwierdza zdolności do samodzielnego prowadzenia badań w budowie systemów symulacji eksperymentów fizycznych, a także demonstrowa zdolności i umiejętności do wprowadzania nowatorskich rozwiązań. Podsumowując, recenzowana praca spełnia kryteria właściwe dla rozpraw doktorskich i wnosi o jej przyjęcie oraz dopuszczenie do publicznej obrony.

## Uwagi do manuskryptu.

- Brak jasno postawionej tezy, którą praca przedstawiona w manuskrypcie powinna uzasadnić. Pośrednio można się jej doszukać w tytule pracy jednak nie ma w nim celu wykorzystania opisywanych technologii.
- Brak precyzyjnego określenia wkładu własnego Autora w omawianym projekcie. Podobnie jak w poprzednim punkcie można się tego domyślić ze szczegółowych opisów wykonywanych badań oraz przeglądu zamieszczonej literatury. Wiadomym jest również, że przygotowanie oprogramowania dla tak dużych i skomplikowanych systemów jak symulacja eksperymentu na akceleratorze LHC wymaga współpracy dziesiątek programistów i fizyków, więc w pracy doktorskiej podsumowującej pracę doktoranta w takim systemie koniecznie powinien się znaleźć opis zakresu prac własnych.
- strona 48 – brakuje tekstowego opisu wyników porównania pyTorch/ONNXR – są dwa obrazki ze stosunkiem czasów wykonania oraz zaalokowanej pamięci ale nie ma słowa podsumowania (np. PyTorch lepszy przy większej liczbie wątków (scheduler lepiej pracuje?)). W tekście jest informacja, że pomiary dokonano na architekturze oferującej 40 wątków – jak zatem uzyskano dane do wykresu dla pozycji 40-40 (inter/intra threads).
- dopiero na stronie 31 jest definicja fast simulations – a termin ten jest używany wielokrotnie wcześniej przed tym objaśnieniem – nie ma definicji ultra-fast simulations.
- strona 49 – rysunek nieczytelny – co jest A a co B – z podpisu wynika że oba są fast simulations, a w tekście jest napisane że jeden z nich ma być GEANT4 detailed
- strona 61 – na rysunkach 4.16 jest old reco clusters/energy – co to jest old reco ? rozumiem że predicted jest YOLO
- niewygodna synchronizacja tekstu i odpowiadających mu obrazków prowadząca do konieczności wielokrotnego przewracania stron w poszukiwaniu obrazka opisywanego na danej stronie (bardzo często znajduje się on kilka stron dalej). Podpisy są lakoniczne i oczywiste – typu: zależności a od b gdzie a i b to opisy osi wykresu. Brakuje zdań naprowadzających na cel zamieszczenia danego rysunku, a nawet jeżeli on jest w tekście to parę stron wcześniej. Brak podsumowującego komentarza jest szczególnie widoczny w licznych rysunkach zamieszczonych w kilku dodatkach w zakończeniu manuskryptu.