

# Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

**Autor práce** Bc. Tomáš Čelko

**Název práce** Clustering hits and predictions in data from TimePix3 detectors

**Rok odevzdání** 2023

**Studijní program** Informatika **Studijní obor** Umělá inteligence

**Autor posudku** RNDr. František Mráz, CSc. **Role** vedoucí

**Pracoviště** KSVI MFF UK

## Text posudku:

Diplomová práce Bc. Tomáše Čelky sa zaoberá spracovaním dát z detektoru častíc TimePix3. Takýto detektor detekuje takzvané hity, čo sú záznamy o pozícii pixelu, ktorý zasiahla častica, kedy sa to stalo a koľko energie sa zachytilo v pixele. Takýto detektor dokáže produkovať veľké množstvo záznamov o hitoch pričom záznamy nie sú časovo utriedené. Na ďalšie spracovanie, je nutné hity pospojovať do takzvaných klasterov. Jeden klaster potom popisuje dráhu jednej častice, alebo viacerých častíc, ak sa častica rozpadla. Parametre dráh častíc potom môžu slúžiť na zistenie smeru letu, druhu častice a pod.

Detektor TimePix3 dokáže produkovať desiatky miliónov záznamov o hitoch za sekundu. existujúce klastrovacie programy dokážu spracovať asi 3 milióny hitov za sekundu. Cieľom autora bolo zvýšiť rýchlosť klastrovania využitím paralelizácie na viacjadrových procesoroch a pomocou filtrovania založeného na strojovom učení.

Autor práce urobil veľmi podrobnú analýzu toho, akým spôsobom je možné využiť viacjadrové procesory na spracovanie takýchto dát s cieľom je čo najrýchlejšie zhlukovať jednotlivé hity do klastrov. Bc. Tomáš Čelko veľmi podrobne matematicky analyzoval možné cesty paralelizácie klastrovania dát z detektoru TimePix3. Navrhol niekoľko spôsobov ako rozdeliť toto spracovanie či už formou pipeline alebo rozdelením dát na niekoľko prúdov podľa pozície hitu na detektore alebo podľa času zachytenia častice. Paralelné klastrovanie prináša problém, že pri rozdelení dát do samostatne spracovávaných prúdov a blokov sa stáva, že niektoré klastre budú rozdelené do rôznych blokov. Takéto klastre je treba nájsť a spojiť dohromady do výsledných klastrov.

Autor navrhol niekoľko metód ako pri spracovaní rozdeľovať hity podľa času alebo polohy na detektore. Z výsledkov vyplynulo, že je lepšie deliť dáta na bloky podľa času. Na základe teoretickej analýzy, diplomant implementoval klastrovací algoritmus a aplikoval ho na dáta zo skutočných detektorov. Pre kontrolu správnosti, porovnal výstupy z jeho nového klastrovacieho algoritmu s výstupmi starších klastrovacích algoritmov. Ukázal, že ním navrhnuté metódy produkujú správne

klastre a pritom dokážu zvýšiť rýchlosť klastrovania 4 až 8 násobne.

Aby autor ešte viacej zrýchlil proces klastrovania, využil metódy strojového učenia na analýzu toku dát z detektoru tak, že keď detekuje nejaké zaujímavé vlastnosti, ako napríklad významné zmeny v toku dát, tak zapne klastrovanie. Naopak, keď tok dát je “nezaujímavý”, tak sa klastrovanie zastaví. Týmto spôsobom je možné nielen spracovať väčší počet hitov v kratšom čase ale šetriť aj miesto na ukladanie dát pretože tie “nezaujímavé” klastre sa vôbec neukladajú.

Práca je pomerne rozsiahla, ale aj tak sú tam časti, ktoré by mohli byť urobené podrobnejšie. V experimentálnej časti autor porovnáva výkon svojich algoritmov klastrovania na osobnom počítači so 4-jadrovým procesorom a na servere s 16-jadrovým procesorom. To sú správne zvolené cieľové platformy, no vzhľadom k vývoji hardware by bolo užitočné vedieť ako počet jadier procesora ovplyvňuje rýchlosť klastrovania pre rôzne konfigurácie výpočtových procesov.

Práca je napísaná výbornou angličtinou. Text je veľmi dôkladný, dobre štruktúrovaný a zrozumiteľný. Časti práce už sú súčasťou 2 publikácii a jednej publikácie, ktorá je v príprave. Podľa môjho názoru je práca na veľmi vysokej úrovni. Jej výsledky budú používané v experimentálnej fyzike vysokých energií a prispievajú k zlepšeniu všetkých algoritmov ktoré nadväzujú na takto konštruované dáta, ako je napríklad klasifikácia detekovaných častíc. Doporučujem preto, aby bola uznaná ako diplomová práca.

**Práci doporučuji k obhajobě.**

**Práci navrhuji na zvláštní ocenění.**

Prácu doporučujem do súťaže Cena Wernera von Siemensu pre jej kvalitu (3 publikácie) a pre jej význam pri spracovaní dát z časticovej fyziky, napr. z urýchľovača v CERN.

V Praze dne 26. 8. 2023

Podpis: