AUGSTAS VEIKTSPĒJAS SKAITĻOŠANAS RISINĀJUMI AR CUDA TEHNOLOĢIJU

HIGH PERFORMANCE COMPUTING SOLUTIONS WITH CUDA TECHNOLOGY

Autors: **Kaspars Vogulis,** e-pasts: kasixlv@gmail.com,+371 22437106
Zinātniskā darba vadītājs: **Sergejs Kodors, Dr.sc.comp.,**
Rēzeknes Tehnoloģiju akadēmija, Atbrīvošanas aleja 115, Rēzekne

Abstract*: Proposed research is completed to view the data computing capabilities with CUDA technology and video card options. The goal of work is to compare task execution speed using the CPU and CUDA technology.*

Research study concluded that the use of CUDA technology in certain tasks can improve execution time and save system resources. CUDA technology uses the graphic processor (GPU) parallel architecture that allows a large number of tasks to be solved simultaneously and independently of each other.

Keywords: *CUDA technology, computing operations, data operations, graphics processor*

Ievads

Katrs jaunais atklājums informācijas tehnoloģiju sfērā paver iespējas uz jauniem un iepriekš neredzētiem problēmu risinājumiem un gatavo risinājumu optimizāciju. Viens no problēmu risinājuma un optimizācijas veidiem ir paralēlā programmēšana.

Paralēlā programmēšana ļauj izmantot datora arhitektūras sadalījumu un iespējas, piemēram, ar vairāku kodolu procesoriem. Tās principa pamatā ir pieņēmums, ka katru problēmu un uzdevumu var sadalīt mazās, neatkarīgās daļiņās. Šis programmēšanas veids tiek izmantots, lai apstrādātu datus ar video kartes grafisko procesoru ar dažādu tehnoloģiju, piemēram, ar *CUDA* tehnoloģijas palīdzību. Datu apstrāde ar videokarti ļauj ieekonomēt centrālā procesora resursus un paātrināt darba izpildi.

Darba mērķis ir salīdzināt uzdevuma veikšanas ātrumu, izmantojot centrālo procesoru un *CUDA* tehnoloģiju.

Lai sasniegtu mērķi tika izvirzīti šādi uzdevumi:

1. Izpētīt grafiskā procesora būtību un tā pielietošanas iespējas.
2. Izpētīt *CUDA* tehnoloģiju un salīdzināt to ar citām līdzīgām tehnoloģijām.
3. Realizēt algoritmu, kas apstrādā masīvu, izmantojot centrālo procesoru un grafisko procesoru ar *CUDA* tehnoloģijām.
4. Salīdzināt abus algoritmus pēc izpildes ātruma (izpildes laika).

Pētījuma metodes:

1. Aprakstošā jeb monogrāfiskā: literatūras analīze, lai izpētītu paralēlo programmēšanu, grafisko procesoru un *CUDA* tehnoloģijas iespējas.
2. Kvantitatīvā: tiek eksperimentāli salīdzināts uzdevuma veikšanas ātrumu pēc izpildes laika, izmantojot centrālo procesoru un *CUDA* tehnoloģiju.

Hipotēze : skaitļošanas operācijas tiks paveiktas īsākā laika periodā, izmantojot CUDA tehnoloģiju, nekā izmantojot centrālo procesoru.

2.Grafiskais procesors

2.1.Grafiskā procesora priekšrocības un trūkumi

Grafiskie procesori jeb *GPU* (*Graphical Processing Unit*) ir speciālā mikroshēma, kas ir paredzēta taisīt sarežģītus matemātiskus aprēķinus, paredzētus pirmkārt grafikas renderēšanai [1]. Grafiskie procesori sastāv no vairākiem simtiem kodolu, kas var vienlaicīgi izpildīt simtiem datu apstrādes norišu un tie ir optimizēti darbam ar attēliem un grafiku [2].

Grafiskais procesors ir vairāk orientēts uz daudzu darbību veikšanu sekundē, bet

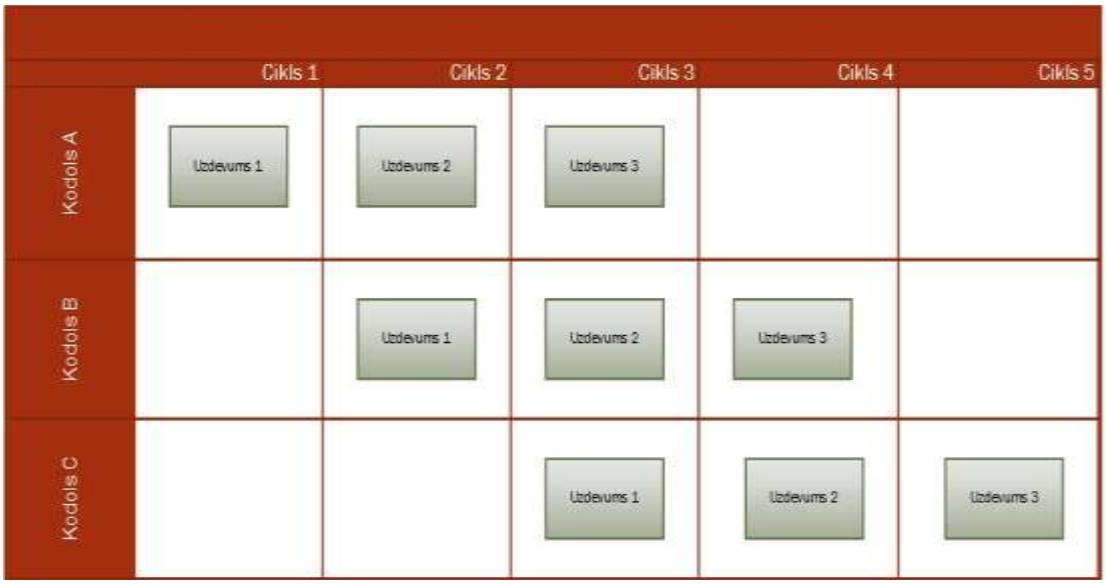
http://dx.doi.org/10.17770/het2017.21.3566

centrālais procesors mēģina samazināt laiku viena uzdevuma veikšanai jeb cik ilgi uzdevums tiek veikts sekundēs [3]. Grafisko procsoru var izmantot vispārējiem uzdevumiem jeb *GPGPU* (*General-purpose computing on graphics processing units*). Tomēr, lai kontrolētu *GPU* darbību un veiktu unikālas instrukcijas, ir jāizmanto papildus tehnoloģijas kas ļauj instruktēt grafisko procesoru par veicamajām darbībām [4].

3. Paralēlā programmēšana

3.1. Grafiskā procesora paralēlās programmēšanas principi

Paralēla programmēšana ietver sevī paralēlās skaitļošanas izmantošanu problēmu risināšanā, kurā uzdevumi tiek sadalīti vairākās smalkās daļās, kuras tiek atrisinātas atsevišķi un vienlaicīgi, no kurām tiek iegūts rezultāts [5]. Paralēlās programmēšanas piemērs ir redzams 1.attēlā:



1.att. **Konveijera sistēmas darbība**

4. *CUDA* tehnoloģija

4.1 *CUDA* tehnoloģiju attīstība un pielietojums

CUDA (*Compute Unified Device Architecture*) ir aplikāciju programmēšanas vide (*API*), kura tika izstrādāta kompānijā *NVIDIA*, kura arī ražo grafiskos procesorus un videokartes. *Nvidia* par savu *CUDA* tehnoloģiju paziņoja 2006. gada novembrī. Citas konkurējošās tehnoloģijas ir *OpenGL, OpenCL, OpenMP, Direct3D, DirectCompute* [6].

Meteoroloģijā ar grafiskā procesora palīdzību tika samazināts laiks līdz pat 20%, lai pareģotu laikapstākļus dažādiem apvidiem ar simulāciju un datu apstrādes palīdzību [7]. *NATO* izmanto zemūdens izpētes tālvadības ierīces, kurās izmanto grafiskos procesorus, lai tas palīdzētu analizēt apkārti un novērot bīstamus objektus un mīnas [8]. Videonovērošanā tiek izmantoti grafiskie procesori ar *CUDA* tehnoloģiju [9].

5. *CUDA* tehnoloģiju salīdzinājums ar procesora iespējā

5.1. Materiāli un metodes

Pētnieciskajā daļā tiks salīdzināts noteikta uzdevuma izpildes ātrums, izmantojot centrālo procesoru (*CPU*) vai grafisko procesoru ar *CUDA* tehnoloģiju. Uzdevums ir dotajam skaitļu masīvam ar elementiem no 0 līdz noteiktam skaitlim *N* aprēķināt kvadrāta pakāpes kāpinājumu, dalījumu pašam ar sevi un izvadīt to. Precizitātes labad, abi kodi tiks veikti dažādos projektos. Netiks mērīts visa algoritma izpildes laiks, bet tikai kāpināšanas laiks, jo masīvu saglabāšanas un izvades metodes krasi atšķiras parastajā koda un *CUDA* kodā.

Testējot *CUDA* tehnoloģiju tiks izmantots gan centrālais procesors, gan videokarte.

Centrālais procesors ļaus izveidot mainīgos, saglabāt un atvēlēt atmiņu, un beigās izvadīt rezultātus uz ekrāna. Aritmētisko daļu paveiks grafiskais procesors ar vairākiem simtiem kodolu. Programmas darbībā, grafiskajā procesorā vienlaikus darbosies *N* skaits kodolu, kuri vienlaicīgi aprēķinās rezultātu. Tad rezultāti tiks apkopoti un pārkopēti uz *CPU* atmiņu.

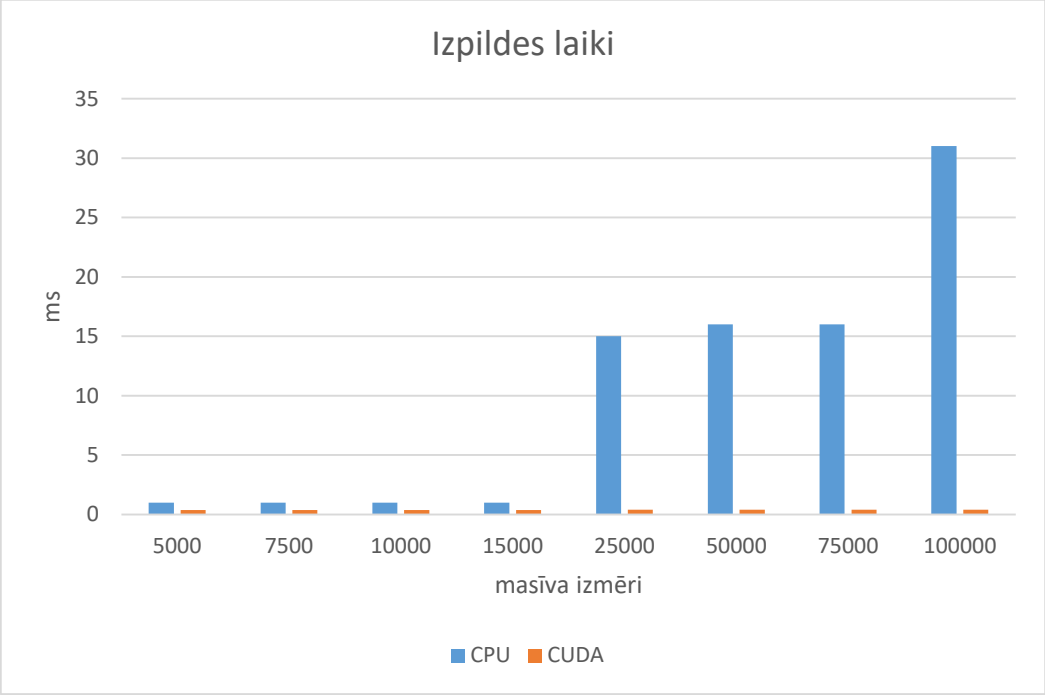
5.2.Rezultāti un rezultātu analīze

CPU un CUDA katram 8 testi ar masīviem ar izmēriem 5000, 7500, 10000, 15000, 25000, 50000, 75000,100000. Katras programmas rezultātus jeb noteikta segmenta izpildes laiks tika saglabāts un apkopots tabulā.

Pirmais tests notika, izmantojot tikai centrālo procesoru. Otrajā testā tikai izmantota *CUDA* tehnoloģija. Vidējie rezultāti ir redzami 2.attēlā un tabulā 5.1.

5.1.tabula

	Laiks (ms)			**Laiks (ms)**	
Izmērs	***CPU***	***CUDA***	**Izmērs**	***CPU***	***CUDA***
5000	1.00	0.37	25000	15.00	0.39
7500	1.00	0.37	50000	16.00	0.40
10000	1.00	0.38	75000	16.00	0.40
15000	1.00	0.38	100000	31.00	0.40



2.att. **Testu rezultāti**

Pēc datiem var spriest, ka grafiskais procesors izpilda savu uzdevumu ātrāk, neatkarīgi no masīva izmēra. *GPU* izpildes ātrumu pārāk stipri neietekmē masīva izmēri, jo kerneļi uzdevumus veic vienlaicīgi un izmantoto kerneļu skaits neietekmē kopējā darba ilgumu. Centrālais procesors ar darbu tika galā daudz lēnāk un tā ātrumu stipri ietekmēja masīva izmēri.

Secinājumi

Grafiskā procesora būtības un pielietošanas izpēte ļāva secināt, ka mūsdienās grafisko procesoru izmanto ne tikai, lai izvadītu un apstrādātu attēlus, grafikas vai animācijas, bet arī datu analīzē, apstrādē un citās skaitļošanas operācijās. Šādas iespējas sniedz dažādas tehnoloģijas, kuras atļauj izmantot grafisko procesoru paralēlo arhitektūru.

CUDA tehnoloģija ir izstrādāta kompānijā *Nvidia* , un paredzēta, lai sniegtu iespēju apstrādāt skaitļošanas operācijas ar *Nvidia* ražotām videokartēm. *CUDA* tehnoloģijai ir daudzas priekšrocības skaitļošanā un izmantošanā, nekā citām tehnoloģijām, tomēr tās nav tik plaši izmantojamas, jo der tikai noteiktām videokartēm.

Veicot praktisku eksperimentu, kurā salīdzināts programmas skaitļošanas daļas izpildes laiks starp centrālo procesoru un grafisko karti, izmantojot *CUDA* tehnoloģiju. Katrā no testiem ar attiecīgo skaitļu skaitu, kāpināšanas un dalīšanas operācijas ātrāk tika veiktas ar *CUDA* tehnoloģiju. Palielinot skaitļošanas operāciju skaitu un masīva elementu daudzumu, centrālā procesora darbību izpildes laiks paildzinājās. Savukārt, *CUDA* tehnoloģiju elementu izmaiņas krasi neietekmēja.

Visi darba uzdevumi tika paveikti un darba mērķis ir sasniegts. Tika pierādīts, ka izmantojot *CUDA* tehnoloģiju, skaitļošanas operācijas tika paveiktas ātrāk, nekā ar centrālo procesoru.

Literatūra

1. Nickolls J., Kirk D. Graphics and Computing GPUs. *Computer organization and design* Amerikas Savienotās Valstis: Elsevier INC.
2. Patterson D.A.,. Hennesy J.L,*Computer organization and design* Amerikas Savienotās Valstis: Elsevier INC. 2014.gads
3. Sanders J., Kandrot E. *CUDA by Example.* Amerikas Savienotās valstis. Nvidia . 2011. gads
4. Nickolls J., Dally W.J. *THE GPU COMPUTING ERA*. 2010.gads. Amerikas Savienotās valstis. IEEE Computer Society. Nvidia. 2010.gads
5. Garland M, Kudlur M., Zheng Y. *Designing a Unified Programming Model for Heterogeneous Machines.* Amerikas Savienotās valstis. Nvidia. 2010.gads
6. Cook S. CUDA Programming. A Developer's Guide to Parallel computing with GPUs. Amerikas Savienotās valstis. Elsevier INC. 2013.gads
7. Computational Fluid Dynamics. Sk. internetā (21.05.2016) http://www.nvidia.com/object/national_center_for_atmospheric_research.html
8. NATO CMRE REVOLUTIONIZES REAL-TIME UNDERSEA MINE DETECTION. Sk internetā (19.05.2016) http://www.nvidia.com/content/tesla/pdf/nato-case-study.pdf
9. Farber R. *CUDA Application Design and Development.*Amerikas Savienotās valstis. Elsevier INC. 2011.gads