

ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ ПАРАЛЕЛЬНИХ ОБЧИСЛЕНЬ З ВИКОРИСТАННЯМ ГРАФІЧНОГО ПРОЦЕСОРА ТА ПЛАТФОРМИ CUDA

З метою покращення ефективності паралельних обчислень розглянуто три методи реалізації — на центральних процесорах, цифрових сигнальних процесорах та графічних процесорах. Розроблено алгоритми та відповідні додатки для кожного пристрою, представлено результати тестування. З метою покращення ефективності паралельної цифрової обробки радіолокаційного сигналу в радіолокаторі для знаходження цілей та їх параметрів запропоновано використання графічних процесорів та програмно-апаратної платформи CUDA. Визначені основні алгоритми для реалізації даної радіотехнічної задачі.

Ключові слова: графічний процесор, ефективність, паралельні обчислення, паралельна цифрова обробка сигналів, радіолокатор.

YU.YU. REUTSKA

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

INCREASING THE EFFICIENCY OF PARALLEL COMPUTING USING GPU AND CUDA PLATFORM

Abstract — The aim is to improve the efficiency of parallel computing for the digital signal processing.

Three methods of the implementing with the central processors, the digital signal processors, and the graphics processors are considered for increasing the efficiency of parallel computing. Algorithms and related with them applications are developed for each device. The test results are presented. In order to increase the efficiency of the parallel radar digital signal processing to find the targets and their parameters the use of GPU and CUDA software and hardware platform it was suggested. Basic algorithms for the realization of the radar problems have been identified.

This method of increasing the efficiency of computing can be implemented both in parallel digital processing of the radar signal and in any devices where there is a capacious computation.

Keywords: graphics processor, GPU, efficiency, parallel computing, parallel digital signal processing, radar.

Вступ

Новостворювані радіотехнічні системи (РТС) повинні мати кращі показники ефективності та якості, ширші функціональні можливості і більшою мірою задовольняти вимогам одержувачів інформації. При створенні РТС намагаються отримати кращі характеристики при певних умовах її використання. Так як вимоги до систем багатоваріантні, і часто в суперечці одна з одною, то бажано оптимізувати систему по узагальненому критерію. Як відмічено в [1], основним критерієм виступає ефективність системи, тобто кількісна характеристика якості виконання системою заданих функцій, що відносяться до енергетичних, інформаційних та економічних затрат. Потужним способом підвищення ефективності функціонування РТС є паралельні обчислення у разі цифрової обробки сигналів (ЦОС). На практиці оцінка якості роботи системи робиться за її основними тактико – технічними характеристиками, а саме за зоною дії, точністю, достовірністю, роздільною здатністю, пропускнуною спроможністю, швидкодією, завадостійкістю, надійністю, доступністю, універсальністю. Ускладнення функцій ЦОС вирішується, головним чином, за рахунок пристроїв цифрової техніки. Кожен клас пристроїв ЦОС має свої переваги і недоліки, які можуть сильно впливати на роботу системи, що побудована на їх основі. Для вибору пристроїв ЦОС повинна враховуватись не тільки продуктивність (швидкодія), але і багато інших параметрів, таких як необхідна кількість для багатоканальної обробки (ціна та енергетичні показники), програмна та апаратна версійність. Вибір оптимального пристрою реалізації паралельних обчислень та створення відповідних оптимальних алгоритмів для кожної конкретної задачі в РТС дозволяє підвищити ефективність системи, особливо швидкодію, понизити вартість, перейти на високошвидкісні методи передачі і обробки інформації, створити інтегровані багатофункціональні комплекси з високим рівнем штучного інтелекту, що адаптивні до заводової обстановки.

Особлива потреба в паралельних обчисленнях у разі цифрової обробки радіолокаційного сигналу для вирішення радіолокаційних задач виникає у разі оцінки роботи імпульсно-доплерівського радару в заводовій обстановці. Наприклад, доцільним є моделювання цифрової обробки радіолокаційного сигналу, який відбитий від цілей на фоні адитивних флукуаційних та пасивних завод для більш точного і високошвидкісного визначення дальності до об'єктів та швидкостей об'єктів.

Основна частина

Аналіз реалізації багатоканальної обробки в сучасних РТС [2] дозволяє зробити висновок, що найбільш розповсюджені три основні методи реалізації ЦОС — на універсальних процесорах, або центральних процесорах (ЦП, англ. Central Processing Unit, CPU), цифрових сигнальних процесорах (ЦСП, англ. Digital Signal Processor, DSP) та програмованих користувачем вентильних матрицях (ПКВМ, англ. Field-Programmable Gate Array, FPGA), що є одним з різновидів програмованої логічної інтегральної схеми (ПЛІС, англ. Programmable Logic Device, PLD).

В [3] розглянуто фактори, що визначають архітектурні рішення ЦОС — пристроїв, їх основні

властивості і області застосування. Сучасна тенденція розвитку ЦСП і ЦП показує, що відмінності між цими двома пристроями зникає. Багатоядерні ЦП мають кращу швидкодію та мають найвищі обсяги продажів процесорів на ринку, але перевагою ЦСП залишаються простота схемотехніки, наявність великої кількості інтерфейсів взаємодії, ціна та енергоспоживання. В [4] проаналізовано ефективність програмованої логічної інтегральної схеми. Програмування ПЛІС є складним завданням, що вимагає наявності спеціальних знань і великих трудовитрат на реалізацію та налагодження, хоча програмування і значно спростилося після появи інструментів алгоритмічного синтезу. Використання алгоритмічного синтезу дозволяє задавати логіку алгоритму в зручній формі, наприклад, на мові С, і дозволяє істотно підвищити гнучкість реалізації конкретного алгоритму. Зараз перевага перед звичайними ПЛІС для вирішення окремих задач ЦОС надається перспективним пристроям – різновидам ПЛІС на основі ПКВМ (англ. FPGA) з гнучкою архітектурою, високим рівнем паралелізму роботи і досить високою ефективністю, особливо при розробці систем, що випускаються малими або середніми серіями. Хоча швидкодія ЦСП велика, але такий процесор може одночасно виконувати лише кілька операцій, тоді як ПКВМ здатні виконувати одночасно практично необмежене число операцій, забезпечуючи високий паралелізм роботи. Але в той же час ПКВМ не можуть виконувати операції з плаваючою точкою, якщо ключова вимога задачі — точність. Також витрачається дуже багато ресурсу ПЛІС для виконання таких операцій.

Радіотехнічні задачі вимагають проведення великого числа операцій швидкого перетворення Фур'є (ШПФ), що зазвичай виконуються з плаваючою точкою, оскільки операції з фіксованою точкою обмежують динамічний діапазон одержуваного рішення, тобто при створенні цих пристроїв слід застосовувати ЦСП. Крім того, операції обертання (або ділення) матриці також краще виконувати за допомогою ЦСП або СП. Таким чином, ПЛІС доцільно застосовувати в системах, що працюють з фіксованою точкою і потребують високого рівня паралелізму, ЦСП і ЦП більше підходять для роботи зі складними алгоритмами з плаваючою точкою. Тому актуальною задачею є застосування такого методу реалізації ЦОС в РТС, який характеризується високою швидкодією у разі багатоканальної обробки (з високошвидкісними паралельними обчисленнями), зручністю програмної та апаратної версійності, простотою реалізацією складних алгоритмами з плаваючою точкою.

Ще один спосіб вирішення задач ЦОС, який виник зовсім нещодавно, є використання спеціалізованих апаратних засобів – графічних процесорів (ГП), які мають значні обчислювальні ресурси [4].

Головне призначення ГП – побудова зображення з урахуванням фізичних властивостей об'єктів, а ці завдання вимагають паралельності в обробці. Через наявність великої кількості ядер (кілька тисяч) ГП стали використовуватися для моделювання та розрахунків, а після появи спеціалізованого інтерфейсу прикладного програмування CUDA реалізація специфічних алгоритмів для ГП дозволила використання загальної для всіх мови С.

В результаті аналізу існуючих пристроїв за архітектурою та на основі балансу між ціною і показниками ефективності, було обрано наступні пристрої:

- Чотирьох'ядерний ЦП Intel Core i5-4460;
- ЦСП TMS320C6747;
- ГП GeForce GTX750.

Основні характеристики цих пристроїв описані відповідно в [5-8].

Для аналізу ефективності відповідних пристроїв було реалізовано три алгоритми з паралельним виконанням: множення матриць, порівняння за значенням в матриці та перетворення Фур'є, які є складовими для реалізації задач в радіолокації. Тестування запропонованих алгоритмів проводилась з числами з плаваючою точкою одинарної точності для матриць різних розмірів.

Основна задача алгоритму множення матриць різної довжини — показати ефективність роботи системи з плаваючою точкою. Алгоритм порівняння за значенням в матриці та запис знайдених елементів у масив наочно показує ефективність роботи системи в багатопотоковому режимі. Алгоритм перетворення Фур'є був обраний як один з основних, який використовується в цифровій обробці сигналів РТС. Для реалізації цього алгоритму використані стандартні бібліотеки, вони рекомендовані для подібних систем і є оптимальними. Оскільки при використанні стандартних бібліотек в ЦП, ЦСП та ГП немає доступу до відповідного коду програми, то результат оцінки ефективності пристроїв пропонується представити у вигляді часу виконання.

На основі аналізу методик вимірювання ефективності вищезазначених ЦП, ЦСП та ГП, для кожного з них було проведено тестування, загальні результати якого зведені в табл. 1.

Робота проводилась з числами з плаваючою точкою одинарної точності. Як показують результати, що наведені в таблиці 1, з трьох обраних пристроїв за швидкодією найефективнішим є ГП. Оптимізація може в значній мірі збільшити ефективність паралельних обчислень, але в даних тестах основна увага приділяється не оптимізації, а визначенню швидкодії платформ виходячи з однакових умов. Варто зазначити, що вже зараз існують ГП, які в кілька разів перевершують тестований по числу ядер, що підвищує продуктивність ГП.

Особлива потреба в імітаційному моделюванні ЦОС радіолокаційного сигналу для вирішення радіолокаційних задач виникає у разі оцінки роботи радіолокаційної станції (РЛС) в заводській обстановці. В [9] представлені етапи моделювання цифрової обробки радіолокаційного сигналу, який відбитий від цілей

на фоні адитивних флуктуаційних та пасивних завад. За основу була обрана імпульсно-доплерівська РЛС. Модель була реалізована в середовищі MatLab, перша частина якої дозволяла отримати радіолокаційний сигнал, а друга — побудувати та дослідити тракт сигнальної обробки, результатом якої є виявлення цілей та їх параметрів.

Таблиця 1

Порівняння універсального, сигнального та графічного процесорів за швидкістю

Алгоритми		Час виконання		
		ЦП	ЦСП	ГП
множення матриць				
512 x 512	мс	34.9958	5817.176	4.148832
1024 x 1024	мс	689.531	46286.120	33.07335
порівняння за значенням в матрицях				
512 x 512	мс	0.0799926	10.765	0.024096
1024 x 1024	мс	0.288284	43.026	0.093888
швидкого перетворення Фур'є				
512	мс	2.65478	5.378	0.350656
1024	мс	19.6767	20.460	0.86608

Моделювання дозволило проаналізувати та наочно оцінити всі етапи цифрової обробки радіолокаційного сигналу. Це стало основою для реалізації паралельної цифрової обробки в середовищі Visual Studio з використанням ГП та програмно-апаратної платформи CUDA.

З відкриттям доступу до компілятора нової мови CUDA С спростилося програмування ГП для обчислень загального призначення. В [10] зазначено, що основними перевагами технології CUDA є: простота — усі програми пишуться "розширеною" мовою С; наявність документації; набір готових інструментів, що включають профайлер (налагоджувач, який показує структуру виконання програми); набір готових бібліотек; багатоплатформність (підтримуються Microsoft Windows, Linux і Mac OS X).

Базові етапи програмування з використанням CUDA складаються з таких дій: виділення пам'яті на ГП; копіювання даних з пам'яті ЦП у виділену пам'ять ГП; запуск ядра (чи послідовний запуск декількох ядер); копіювання результатів обчислень назад в пам'ять ЦП; звільнення виділеної пам'яті ЦП.

Ядра ГП створені для швидкого виконання великої кількості потоків інструкцій, що паралельно виконуються, підтримується більше тисячі потоків на кожен мультипроцесор, яких в чіпі кілька штук. І, як зазначено в [10], якщо перемикання з одного потоку на інший для центрального процесора коштує сотні тактів, то ГП переключає кілька потоків за один такт.

Ієрархія потоків, що запущені на виконання в ГП, представлена на рис. 1. Верхній рівень ієрархії — сітка (англ., grid) відповідає усім потокам (англ., thread), що виконує відповідне ядро. Вона представляє собою одновимірний або двовірний масив блоків (англ., block), а кожен блок — це одновимірний, двовірний або тривірний масив потоків. При цьому усі блоки, що утворюють сітку, мають однакову розмірність і розмір.

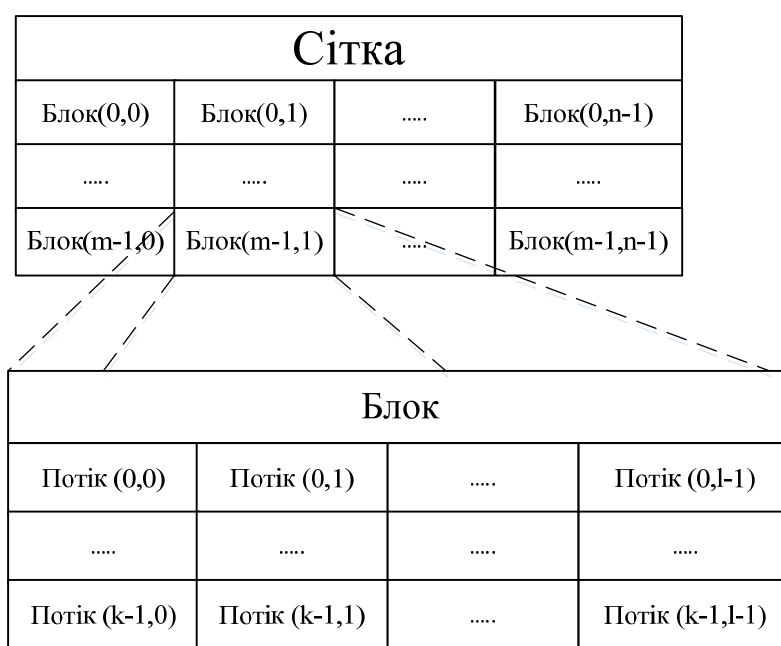


Рис. 1. Ієрархія потоків в CUDA

Кожен блок в сітці має свою адресу, що складається з одного або двох додатних цілих чисел (індекс блоку в сітці). Кожний потік усередині блоку також має свою адресу — одно, два або три додатних цілих числа, які задають індекс потоку усередині блоку. Однозначне визначення ядром номеру потоку здійснюється завдяки змінним, що є тривимірним цілочисловим вектором. Відповідне розділення усіх потоків сформував загальний прийом використання CUDA, де початкова задача розбивається на набір окремих підзадач, що вирішуються незалежно одна від одної, і кожній такій підзадачі відповідає свій блок потоків, який її вирішує. Для реалізації роботи з платформою CUDA пропонується використання бібліотеки Cudafy.NET, яка в є набором бібліотек, які дозволяють написання додатків NVIDIA CUDA з Microsoft NET Framework.

Радіолокаційний сигнал, який було отримано з середовища Matlab, представляє собою матрицю. Над цією матрицею радіолокаційних даних виконується ЦОС, яка складається з чотирьох етапів:

1. Узгоджена фільтрація (стиснення радіолокаційного сигналу), для чого застосовується фільтр на основі операції згортки або фільтрація на основі швидкого перетворення Фур'є;
2. Фільтрація пасивних завад на основі швидкого перетворення Фур'є;
3. Багатоканальна доплерівська фільтрація на основі швидкого перетворення Фур'є;
4. Детектування сигналу для розпізнавання цілей та знаходження параметрів цілі.

Для паралельних обчислень в обробці радіолокаційного сигналу були виділені основні алгоритми, які необхідні для реалізації даної радіотехнічної задачі на ГП:

- Перетворення матриці у вектор;
- Повернення масиву комплексно-сполучених значень в векторі;
- Ділення вектору комплексних значень на комплексне значення;
- Ділення матриці комплексних значень на комплексне значення;
- Множення матриці комплексних значень на комплексне значення;
- Множення вектору комплексних значень на матрицю комплексних значень;
- Трансформування матриці комплексних значень;
- Добуток матриці комплексних значень на матрицю комплексно-сполучених значень;
- Додавання рядків в матриці комплексних значень;
- Множення вектору комплексних значень на вектор комплексних значень;

Висновки

Обґрунтовано використання графічного процесора та апаратно-програмної платформи CUDA для реалізації паралельних обчислень. Розроблено алгоритми та відповідні додатки для кожного пристрою, представлено результати тестування пристроїв для порівняння універсального, сигнального та графічного процесорів за швидкодією. Результати показали, що з трьох обраних пристроїв за швидкодією найефективнішим є графічний процесор. Даний метод підвищення ефективності паралельних обчислень є придатним для будь-яких задач, де є високоємні обчислення.

Запропоновано використання графічного процесора для підвищення ефективності паралельних обчислень в цифровій обробці радіолокаційного сигналу, який відбитий від цілей на фоні адитивних флукуаційних та пасивних завад в імпульсно-доплерівській радіолокаційній станції для виявлення цілей та знаходження параметрів цілей. На основі моделювання цифрової обробки радіолокаційного сигналу в середовищі MatLab визначені основні алгоритми для реалізації даної радіотехнічної задачі на графічному процесорі.

Література

1. Радиотехнические системы: учебник для студ. высш. учеб. заведений / [под ред. Ю. М. Казаринова]. — М.: Издательский центр «Академия», 2008. — 592 с.
2. Беляев А. Современные устройства цифровой обработки сигналов. Вместе или врозь? / А. Беляев, Т. Солохина, В. Юдинцев // ЭЛЕКТРОНИКА: Наука, Технология, Бизнес. — 2009. — №1. — С. 28 – 35.
3. Shuai Che. Accelerating Compute-Intensive Applications with GPUs and FPGAs / Shuai Che, Jie Li, W Sheaffer, K. Skadron, J. Lach. In Proceeding // SASP '08 Proceedings of the 2008 Symposium on Application Specific Processors // IEEE Computer Society Washington, DC, USA. —2008, pages 101-107. [Электронный ресурс]. — Режим доступа: http://www.cs.virginia.edu/~skadron/Papers/che_sasp08.pdf
4. Шаршавин П. В. Применение графического процессора ПК для цифровой обработки сигналов. — Красноярск: ООО НПП «Автономные аэрокосмические системы – ГеоСервис». — 2013. [Электронный ресурс]. Режим доступа: <http://uav-siberia.com/content/primenenie-graficheskogo-processora-pk-dlya-cifrovoy-obrabotki-signalov>.
5. Шаповалов А. Обзор и тестирование процессора Intel Core i5-4460 [Электронный ресурс]. — Режим доступа: http://ru.gecid.com/cpu/intel_core_i5-4460/?s=all/.
6. Рудникович А.С. Архитектура процессоров TMS320C6000 фирмы TEXAS INSTRUMENTS и отладочная программная среда CODE COMPOSER STUDIO: Учебное пособие. – Томск: кафедра ТУ, ТУСУР, 2012. — 81 с.
7. Офіційний сайт компанії Texas Instruments [Електронний ресурс]. — Режим доступу:

<http://www.ti.com/>.

8. Офіційний сайт компанії NVIDIA [Електронний ресурс]. — Режим доступу: <http://www.nvidia.com/>.

9. Реутська Ю. Ю. Моделювання цифрової обробки радіолокаційного сигналу імпульсно-доплерівського радару, який працює в заводській обстановці. Міжнародна науково-технічна конференція «Радіотехнічні поля, сигнали, апарати та системи», Київ, 16 – 22 березня 2015 р.: матеріали конференції — Київ, 2015. — 286 с. — с. 184 – 186.

References

1. Radiotekhnicheskie sistemy: uchebnyk dlya stud. vyssh. ucheb. zavedenii / [pod red. Yu. M. Kazarinova]. — M.: Izdatel'skii tsentr «Akademiya», 2008. — 592 s.
2. Belyaev A. Sovremennyye ustroystva tsifrovoi obrabotki signalov. Vmeste ili vroz'? / A. Belyaev, T. Solokhina, V. Yudinsev // ELEKTRONIKA: Nauka, Tekhnologiya, Biznes. — 2009. — #1. — S. 28 – 35.
3. Shuai Che. Accelerating Compute-Intensive Applications with GPUs and FPGAs / Shuai Che, Jie Li, W Sheaffer, K. Skadron, J. Lach. In Proceeding // SASP '08 Proceedings of the 2008 Symposium on Application Specific Processors // IEEE Computer Society Washington, DC, USA. — 2008, pages 101-107. [Электронный ресурс]. — Режим доступа: http://www.cs.virginia.edu/~skadron/Papers/che_sasp08.pdf
4. Sharshavin P. V. Primenenie graficheskogo protsessora PK dlya tsifrovoi obrabotki signalov. — Krasnoyarsk: OOO NPP «Avtonomnye aerokosmicheskie sistemy – GeoServis». — 2013. [Elektronnyi resurs]. Rezhim dostupa: <http://uav-siberia.com/content/primenenie-graficheskogo-processora-pk-dlya-cifrovoy-obrabotki-signalov>.
5. Shapovalov A. Obzor i testirovanie protsessora Intel Core i5-4460 [Elektronnyi resurs]. — Rezhim dostupa: http://ru.gecid.com/cpu/intel_core_i5-4460/?s=all/.
6. Rudnikovich A.S. Arkhitektura protsessorov TMS320C6000 firmy TEXAS INSTRUMENTS i otladochnaya programmaya sreda CODE COMPOSER STUDIO: Uchebnoe posobie. – Tomsk: kafedra TU, TUSUR, 2012. — 81 s.
7. Ofitsiyniy sait kompanii Texas Instruments [Elektronnyi resurs]. — Rezhim dostupa: <http://www.ti.com/>.
8. Ofitsiyniy sait kompanii NVIDIA [Elektronnyi resurs]. — Rezhim dostupa: <http://www.nvidia.com/>.
9. Reutska Yu. Yu. Modeliuvannya tsyfrovoy obrobky radiolokatsiinoho syhnalu impulsno-doplerivskoho radaru, iakyi pratsiuie v zavadvoyi obstanovtsi. Mizhnarodna naukovo-tekhnichna konferentsiia «Radiotekhnichni polia, syhnaly, aparaty ta systemy», Kyiv, 16 – 22 bereznia 2015 r.: materialy konferentsii — Kyiv, 2015. — 286 s. — s. 184 – 186.
10. Borekov A. V. Osnovy raboty s tekhnologiei CUDA. / A. V Borekov, A. A Kharlamov. — M.: DMK Press, 2010. — 232 s.

Рецензія/Peer review : 9.1.2017 р. Надрукована/Printed :22.3.2017 р.
Рецензент: к.т.н., доцент Савенко Я.В.