

УДК 681.3:004.272

А.А. ЯРОВИЙ¹, Н.І. КОКРЯЦЬКА², С.В. НАКОНЕЧНА², М.С. МАТЕЙЧУК¹, Т.Д. ПОЛЬГУЛЬ¹

АНАЛІЗ ОБЧИСЛЮВАЛЬНОЇ СКЛАДНОСТІ GPU-ОРІЄНТОВАНИХ ПАРАЛЕЛЬНО-ІЄРАРХІЧНИХ ОБЧИСЛЮВАЛЬНИХ СИСТЕМ ТА ОЦІНЮВАННЯ ПРОДУКТИВНОСТІ ЇХ АПАРАТНОГО ЗАБЕЗПЕЧЕННЯ

¹ *Вінницький національний технічний університет,
Хмельницьке шосе, 95, Вінниця, Україна, 21021
Тел.: +380 (432) 598243, E-mail: axa@vinnitsa.com*

² *Державний економіко-технологічний університет
транспортного
вул. Лукашевича, 19, Київ, Україна, 03049,
Тел.: +380 (44) 4654280, E-mail: timchen@list.ru*

Анотація. У проведених дослідженнях здійснено аналіз процесів паралельно-ієрархічного перетворення, які базуються на новій багаторівневій концепції організації обчислень, визначено їх обчислювальну складність. Здійснено оцінку продуктивності апаратного забезпечення паралельно-ієрархічних обчислювальних систем на основі GPU-орієнтованої платформи, як перспективного засобу організації високопродуктивних обчислень.

Аннотация. В проведенных исследованиях осуществлен анализ процессов параллельно-иерархического преобразования, которые базируются на новой многоуровневой концепции организации вычислений, определена их вычислительная сложность. Осуществлена оценка производительности аппаратного обеспечения параллельно-иерархических вычислительных систем на базе GPU-ориентированной платформы, как перспективного средства организации высокопроизводительных вычислений.

Abstract. The analysis of processes of parallel-hierarchical transformation based on modern multilevel conception of computing organization is carry out in researches. The computational complexity of parallel-hierarchical transformation is estimated. The performance evaluation of hardware of parallel-hierarchical computer system based on GPU-oriented platform as perspective facilities of organization of high-performance computing is carry out.

Ключові слова: паралельні обчислення, паралельно-ієрархічне перетворення, GPGPU, обчислювальна складність, продуктивність апаратного забезпечення.

ВСТУП

На сучасному етапі попит на підвищення гетерогенності в обчислювальних системах пов'язаний із постійним зростанням потреби у високопродуктивних інтерактивних системах, які динамічно взаємодіють із зовнішнім середовищем (відео-системи, системи керування, мережеві системи, тощо). Оскільки в сучасних умовах зростаючі обсяги даних та обчислень, необхідних для обробки великих масивів інформації потребують все більшої продуктивності використовуваних для цього обчислювальних систем. Натепер, із врахуванням вказаних додаткових обмежень, основним методом одержання додаткової продуктивності обчислювальних систем є введення додаткових спеціалізованих ресурсів, в результаті чого обчислювальна система стає гетерогенною. Це дозволяє розробникам використовувати декілька типів обчислювальних елементів, кожен з яких здатен найбільш ефективно виконувати завдання. Введення додаткових, незалежних обчислювальних ресурсів неминуче приводить до того, що більшість гетерогенних систем розглядаються як паралельні багатоядерні обчислювальні системи [1-4].

Також відзначимо, що останнім часом сильний поштовх у розвитку одержала досить специфічна галузь високопродуктивних обчислень – застосування в обчислювальних системах додаткових акселераторів обчислень, що ефективно виконують частину обчислювального навантаження. До таких акселераторів, зокрема, належать GPU-пристрої. У зв'язку з тим, що продуктивність таких пристроїв

останнім часом суттєво зросла, постає питання про використання їх потужностей (до 6 Тфлопс для сучасних відеоадаптерів) для розв'язання складних і ресурсомістких обчислювальних задач. Актуальність цього напряму досліджень підтверджується також тим, що провідні виробники відеоадаптерів почали випускати спеціалізовані GPU-акселератори, призначені для високопродуктивних обчислень загального характеру: nVidia Tesla і AMD Firestream із спеціалізованими програмними засобами розробки: nVidia CUDA і AMD Firestream SDK. Дані засоби дозволяють розробнику абстрагуватися від специфіки комп'ютерної графіки при використанні GPU-обчислювачів у задачах організації високопродуктивної обробки інформації. Це є істотною перевагою зазначених систем, оскільки існуючі засоби для багатопроцесорних систем досить дороговартісні і, як правило, орієнтовані на гомогенну архітектуру обчислювальних систем, що обмежує їх поширення. Тому, розробка більш дешевих і більш варіативних (щодо структури устаткування) засобів управління багатопроцесорними обчислювальними системами є перспективною [5-8].

В даному напрямі варто виділити наукові дослідження з розробки суперкомп'ютерів Національної лабораторії "Oak Ridge", США (проект Jaguar), НДОО МДУ ім. М.В. Ломоносова, Росія (проект Lomonosov), тощо. Вагомих наукових результатів в організації паралельних обчислень досягнуто в: Лабораторії паралельних інформаційних технологій (проф. В.В.Воєводін, МДУ), Центрі суперкомп'ютерних і масивно-паралельних застосувань (Обчисл. центр РАН), Відділі космічних інформаційних технологій та систем (Інститут космічних досліджень НАНУ-НКАУ), Інституті кібернетики ім. В.М. Глушкова НАНУ.

У проведених наукових дослідженнях завдяки сполученню ідей паралельно-ієрархічного перетворення (ППП) інформації і технологій GPGPU з масовим паралелізмом стало можливим створення ефективних обчислювальних структур, у яких розширення функціональних та інтелектуальних можливостей технічних і алгоритмічних засобів на різноманітних рівнях ієрархії випереджає ріст їхньої складності. Це особливо виявляється при реалізації багатоетапного оброблення зображень, де перспективним є використання PPP на основі графічних прискорювачів, рекурсивного паралельно-ієрархічного оброблення з компактною кластеризацією і сортуванням одержуваних проміжних результатів [4, 9-11].

Тематика даних досліджень пов'язана з обласною Програмою "Розвитку інформаційних, телекомунікаційних та інноваційних технологій в закладах освіти Вінницького регіону до 2015 року".

У попередніх роботах було розглянуто методи та математичні моделі організації високопродуктивних паралельно-ієрархічних обчислювальних процесів та особливості організації комп'ютерних засобів для паралельно-ієрархічного перетворення цифрових сигналів, а також деякі питання їх прикладного застосування [12-14]. Тому, детальніше зупинимося на особливостях процесу перетворення інформації в багаторівневих паралельно-ієрархічних обчислювальних системах (ПООС) та аналізі їх обчислювальної складності.

МЕТА ДОСЛІДЖЕННЯ

Метою роботи є розвиток теоретичних основ паралельно-ієрархічного перетворення інформації шляхом аналізу процесів PPP та визначення їх обчислювальної складності, які базуються на новій багаторівневій концепції організації обчислень, що дозволяє формалізувати багатоетапні процедури їх паралельної взаємодії у часі на різних рівнях ієрархії ПООС при реалізації на високопродуктивній GPU-орієнтованій апаратній платформі.

АНАЛІЗ ОБЧИСЛЮВАЛЬНОЇ СКЛАДНОСТІ ПАРАЛЕЛЬНО-ІЄРАРХІЧНОГО ПЕРЕТВОРЕННЯ

Паралельно-ієрархічна обчислювальна система складається з рівнів, а рівні – з паралельних гілок, що містять процесорні елементи (рис. 1). Причому, починаючи із другого рівня, кожна наступна гілка, що формується, зсунута у часі щодо попередньої (позначено зафарбованим колом на рис. 1). Початковий масив даних подається на елементи паралельних гілок першого рівня, а елементи гілок другого й наступних рівнів формуються послідовно відповідно з елементів гілок першого й попередніх рівнів. Необхідно зазначити, що довжина обчислювального алгоритму ПООС й відповідно кількість гілок та елементів у них, а також рівнів не є постійними величинами, а залежать від характеру розподілу вхідних даних [12-13].

Основне призначення ПООС – розв'язання задач швидкого перетворення, класифікації й прогнозування, орієнтованих на оброблення великих масивів даних (наприклад, відеоданих) у реальному часі. Здійснимо аналіз обчислювальної складності PPP за критерієм кількості використовуваних операцій

[12]. Нехай є k масивів. Позначимо через n_i – кількість елементів у масиві під номером i (очевидно, що $n_i \geq 1$), $i = \overline{1, k}$, причому $\sum_{i=1}^k n_i = N$ – загальна кількість оброблюваних елементів; через m_i – кількість різноманітних елементів у i -му масиві (очевидно, що $m_i \geq 1$), причому $\sum_{i=1}^k m_i = M$ – загальна кількість різноманітних оброблюваних елементів; через p_i^j – ймовірність появи j -го елемента в i -му масиві, де $j = \overline{1, m_i}$, а $m = \max_{i=1, k} m_i$.

Для одержання першої множини елементів гілки ППП необхідно здійснити N раз операцію порівняння та $\sum_{i=1}^k n_i p_i^1$ –операцію вибору спільної частини. Тому для оброблення залишається:

$$N_1 = \sum_{i=1}^k n_i (1 - p_i^1) \text{ елементів.}$$

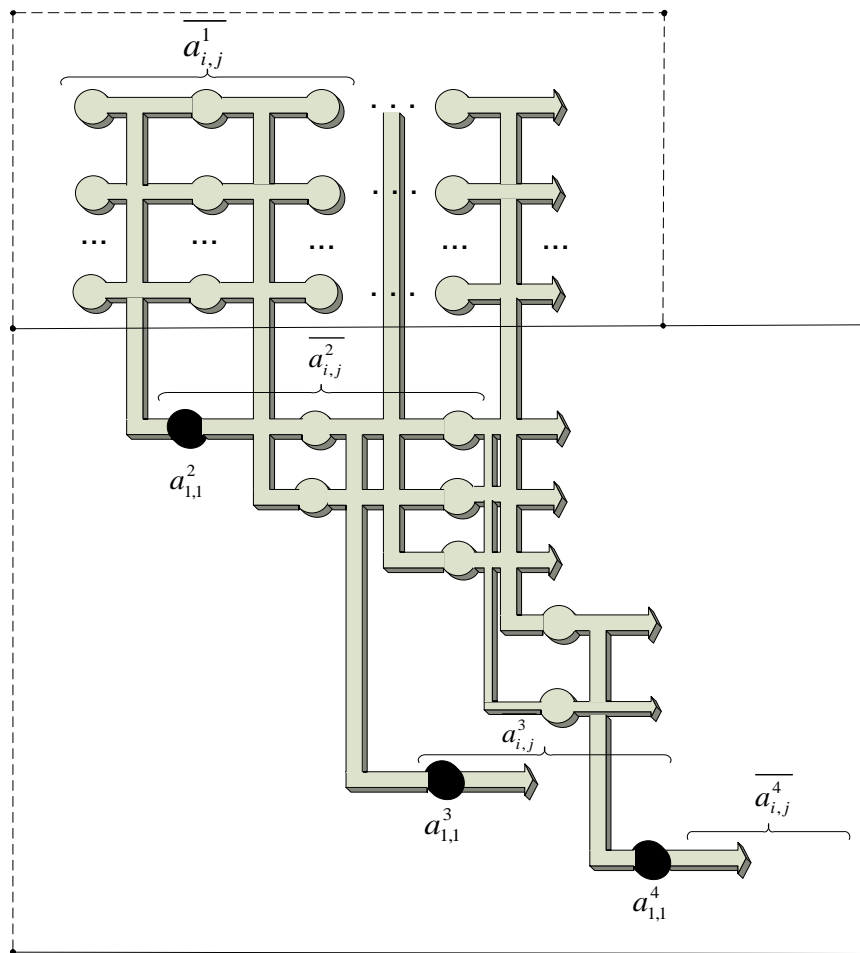


Рис. 1. Узагальнена структура ППОС

Для одержання другої множини елементів гілки ППОС необхідно здійснити N_1 раз операцію порівняння та $\sum_{i=1}^k n_i (1 - p_i^1) p_i^2$ раз операцію вибору. Тому для оброблення залишається:

$$N_2 = \sum_{i=1}^k n_i (1 - p_i^1) (1 - p_i^2) \text{ елементів.}$$

Для одержання m -ої множини елементів гілки необхідно здійснити $N_{m-1} = \sum_{i=1}^k \left[n_i \prod_{j=1}^{m-1} (1 - p_i^j) \right]$ раз

операцію порівняння та $\sum_{i=1}^k \left[n_i \prod_{j=1}^{m-1} (1-p_i^j) p_i^m \right]$ раз операцію вибору.

Таким чином, кількість операцій буде:

$$\begin{aligned}
 & N + \sum_{i=1}^k n_i p_i^1 + \sum_{i=1}^k n_i (1-p_i^1) + \sum_{i=1}^k n_i (1-p_i^1) p_i^2 + \sum_{i=1}^k n_i (1-p_i^1)(1-p_i^2) + \\
 & \quad + \dots + \sum_{i=1}^k \left[n_i \prod_{j=1}^{m-1} (1-p_i^j) p_i^m \right] + \sum_{i=1}^k \left[n_i \prod_{j=1}^{m-1} (1-p_i^j) \right] = \\
 & = N + \sum_{i=1}^k n_i (p_i^1 + 1-p_i^1 + (1-p_i^1) p_i^2 + (1-p_i^1)(1-p_i^2) + \dots + \\
 & \quad + \prod_{j=1}^{m-1} (1-p_i^j) p_i^m + \prod_{j=1}^{m-1} (1-p_i^j)) = N + \sum_{i=1}^k [n_i (1 + (1-p_i^1) + (1-p_i^1)(1-p_i^2) + \dots + \\
 & \quad + \prod_{j=1}^{m-1} (1-p_i^j))] = N + \sum_{i=1}^k \left[n_i \left(1 + \sum_{z=1}^{m-1} \prod_{j=1}^z (1-p_i^j) \right) \right]. \tag{1}
 \end{aligned}$$

Доведемо вираз (1) методом математичної індукції. Нехай кількість множин $k = 1$, тоді кількість елементів $N = n$, а також

$$n + \sum_{i=1}^1 \left[n_i \left(1 + \sum_{z=1}^{m-1} \prod_{j=1}^z (1-p_i^j) \right) \right] = 2n + n \sum_{z=1}^{m-1} \prod_{j=1}^z (1-p_i^j) = n \left(2 + \sum_{z=1}^{m-1} \prod_{j=1}^z (1-p_i^j) \right)$$

Нехай $k = 2$, тоді:

$$\begin{aligned}
 & n_1 + n_2 + \sum_{i=1}^2 \left[n_i \left(1 + \sum_{z=1}^{m-1} \prod_{j=1}^z (1-p_i^j) \right) \right] = 2n_1 + 2n_2 + n_1 \sum_{z=1}^{m-1} \prod_{j=1}^z (1-p_i^j) + \\
 & \quad + n_2 \sum_{z=1}^{m-1} \prod_{j=1}^z (1-p_i^j) = 2(n_1 + n_2) + (n_1 + n_2) \sum_{z=1}^{m-1} \prod_{j=1}^z (1-p_i^j) = (n_1 + n_2) \left(2 + \sum_{z=1}^{m-1} \prod_{j=1}^z (1-p_i^j) \right)
 \end{aligned}$$

З допущення, що вираз (1) правильний для k множин, доведемо, що він правильний для $k+1$ множини.

$$\begin{aligned}
 & N + \sum_{i=1}^k \left[n_i \left(1 + \sum_{z=1}^{m-1} \prod_{j=1}^z (1-p_i^j) \right) \right] + n_{k+1} \left(2 + \sum_{z=1}^{m-1} \prod_{j=1}^z (1-p_i^j) \right) = n_1 + n_2 + \dots + n_k + \\
 & \quad + \sum_{i=1}^k \left[n_i \left(1 + \sum_{z=1}^{m-1} \prod_{j=1}^z (1-p_i^j) \right) \right] + n_{k+1} + n_{k+1} \left(1 + \sum_{z=1}^{m-1} \prod_{j=1}^z (1-p_i^j) \right) = n_1 + n_2 + \dots + n_{k+1} + \\
 & \quad + \sum_{i=1}^{k+1} \left[n_i \left(1 + \sum_{z=1}^{m-1} \prod_{j=1}^z (1-p_i^j) \right) \right] = N + \sum_{i=1}^{k+1} \left[n_i \left(1 + \sum_{z=1}^{m-1} \prod_{j=1}^z (1-p_i^j) \right) \right],
 \end{aligned}$$

де $N = \sum_{i=1}^{k+1} n_i$. Отже, отримано вираз вигляду (1), що і було потрібно довести.

Нехай поява елементів у масиві буде подією рівновірогідною із ймовірністю p , тоді для N оброблюваних елементів із k n -елементних множин справедливо:

$$\begin{aligned}
 & N + \sum_{i=1}^k \left[n_i \left(1 + \sum_{z=1}^{m-1} \prod_{j=1}^z (1-p_i^j) \right) \right] = N + \sum_{i=1}^k \left[n_i \left(1 + \sum_{z=1}^{m-1} (1-p)^z \right) \right] = \\
 & = N + \left(1 + \sum_{z=1}^{m-1} (1-p)^z \right) \sum_{i=1}^k n_i = N + N \left(1 + \sum_{z=1}^{m-1} (1-p)^z \right) = \\
 & = N \left(2 + \sum_{z=1}^{m-1} (1-p)^z \right) = N \left(2 + \frac{1-p}{p} \right) = \left(1 + \frac{1}{p} \right) N, \tag{2}
 \end{aligned}$$

якщо $n = m$, тоді $p = \frac{1}{N}$.

Отже, кількість операцій для ППП – $N(N+1)$.

ОЦІНЮВАННЯ ПРОДУКТИВНОСТІ АПАРАТНОГО ЗАБЕЗПЕЧЕННЯ ПАРАЛЕЛЬНО-ІЄРАРХІЧНИХ ОБЧИСЛЮВАЛЬНИХ СИСТЕМ НА ОСНОВІ GPU-ОРІЄНТОВАНОЇ ПЛАТФОРМИ

Усі сучасні платформи API для програмування відеоадаптерів підтримують можливість використання декількох відеоадаптерів у межах однієї системи (наприклад, SLI у NVIDIA або Crossfire у AMD). Для кожного відеоадаптера можна призначити декілька контекстів виконання, і з кожним контекстом асоціюється один потік операційної системи. В рамках контексту створюються шейдери та виділяється пам'ять під дані користувача. Одночасно на одному відеоадаптері може виконуватись лише один шейдер із одного контексту. Процеси виклику і передавання даних є асинхронними [15].

Для реалізації ППП на основі GPU необхідно модифікувати його класичну математичну модель [16]. Архітектура графічних адаптерів дає змогу запускати величезну кількість потоків одночасно, також ядро відеоадаптера здатне виконувати арифметичні операції швидше за процесорне ядро (CPU), але на відміну від центрального процесора звернення до пам'яті на відеоадаптері відбувається повільніше, а час на копіювання даних пам'яті з відеоадаптера на центральний процесор та навпаки, може значно вплинути на час роботи програми, та позбавити виграшу у швидкодії. Також обмеженням є те, що неможливо контролювати усі потоки без взаємодії з центральним процесором. Саме тому оптимізація алгоритму повинна бути направлена на зменшення кількості звернень до глобальної пам'яті усіма потоками одночасно, зменшення обсягів пам'яті, які пересилаються з центрального процесора на відеоадаптер та навпаки [16-18].

Апаратне забезпечення для організації масивно-паралельних обчислень на основі високопродуктивної SLI-системи. В якості апаратного забезпечення, при організації масивно-паралельних обчислень, було обрано технологію NVIDIA SLI, яка також підтримує програмну платформу NVIDIA CUDA.

Необхідно зазначити, що технологія NVIDIA SLI розроблена для "PCI Express". Ця нова шина даних має збільшену пропускну здатність (в два або чотири рази більше, ніж AGP 8X), підтримку ізохронної передачі даних і можливість роботи з кількома високошвидкісними графічними пристроями. Стосовно ж AGP 8X, то вони обмежені одним високошвидкісним графічним пристроєм і не досягають такої високої ефективності як у SLI [19].

Розроблено ПІОС з рекурсивною архітектурою на базі високопродуктивного обчислювального мережевого комплексу на основі сучасних GPGPU технологій, який містить 2 відеоадаптера GeForce GTX590 (кожен з яких працює у 1024 потоки), що є пристроєм останнього покоління NVIDIA [16]. Кожен відеоадаптер GeForce GTX590 містить два 512-ядерних GPU GTX 500 (з архітектурою GF110), що в сукупності становить $512 \times 4 = 2048$ -ядерне апаратне забезпечення. Таким чином, розроблена на базі технології Quad SLI високопродуктивна ПІОС, що містить два таких відеоадаптери надає можливість обробляти інформацію у 2048 потоків.

Оцінювання продуктивності апаратного забезпечення. Оцінювання теоретичної продуктивності апаратного забезпечення є дуже важливим етапом, так як це може допомогти розробнику виявити проблемні аспекти при впровадженні та організації обчислень на GPU [8, 15].

Розрахунок теоретичної продуктивності. Кожен потоковий мультипроцесор (SM) в GPU архітектури GF110 містить 32 шейдерних процесори (SP) і 4 блока спеціалізованих функцій (SFU). Кожен SP може виконувати до двох операцій FMA (Fused ADD+MUL) одинарної точності за такт, а кожен SFU – до чотирьох операцій за такт. Пікове співвідношення операцій, які виконує SP, до операцій, які виконує SFU, становить для GF100 – SFU=4:1 [15].

Теоретична продуктивність $FLOPS_{sp+sfu}$ (GFLOPS) блоків SP розраховується за виразом [15]:

$$FLOPS_{sp} \approx f \times n \times 2,$$

де n – кількість SP; f – їх частота (ГГц).

Відповідно до вищенаведеного, вираз для GF110 буде таким [15]:

$$FLOPS_{sp} \approx f \times m \times (32 \text{ SPs} \times 2 \text{ (FMA)}),$$

де m – кількість SM.

Повну продуктивність, що характеризує кількість операцій із плаваючою комою в секунду, яку

виконує графічний процесор GF110 можна розрахувати за виразом [8, 15]:

$$FLOPS_{sp} \approx f \times m \times (32 \text{ SPs} \times 2(\text{FMA}) + 4 \times 4 \text{ SFUs}) = 1,215 \times 16 \times (32 \times 2 + 4 \times 4) = \\ = 1,215 \times 16 \times (64 + 16) = 1555,2 \text{ GFLOPS}$$

або

$$FLOPS_{sp} \approx f \times n \times 2,5 = 1,215 \times 512 \times 2,5 = 1555,2 \text{ GFLOPS.}$$

Відповідно до даних NVIDIA для відеоадаптера NVIDIA GeForce GTX 590, який містить два графічних процесори GF110, повна теоретична продуктивність буде [8, 15]:

$$FLOPS_{GF110} \approx FLOPS_{sp} \times 1,6 = 1555,2 \times 1,6 = 2488,32 \text{ GFLOPS.}$$

Таким чином, розроблена SLI система з двох відеоадаптерів NVIDIA GeForce GTX 590 надаватиме можливість обробляти інформацію у 2048 потоків з граничною теоретичною продуктивністю 4976,6 GFLOPS.

Оцінювання продуктивності роботи пам'яті. Розглянемо процедуру розділення шейдера на команди. Адже в подальшому ця інформація буде потрібна для теоретичних оцінок. Додатковою інформацією, необхідною для оцінювання теоретичної продуктивності, є:

- 1) # блоки ALU;
- 2) # текстурні блоки;
- 3) розмір системної шини;
- 4) частота ядра;
- 5) частота пам'яті.

Для GF110, кількість ALU-блоків є кількістю SPU (512), які виконують VLIW-команди. Розмір шини пам'яті – 384 біти. Частота пам'яті і ядра буде залежати від використаної плати, і може бути знайдена в технічних специфікаціях плати. Відеоадаптер NVIDIA GeForce GTX 590 базуються на двох GF110 з частотою ядра 607 МГц і частотою пам'яті 3414 МГц [8, 15].

Починаємо з шейдеру тільки з ALU-командами. Теоретична продуктивність роботи пам'яті (теоретичний час роботи з одним ALU, одним текстурним блоком, одним вхідним і вихідним бітом) буде обчислюватися за виразом [15]:

$$T_1 = \frac{(\# \text{ потоки}) \times (\# \text{ VLIW ALU команди})}{(\text{ALU} / \text{clk}) \times (3D_engine_speed)}$$

Кількість потоків будемо вважати за обсяг домену виконання. На прикладі GF110 одна ALU-команда шейдера з доменом 2M потоків (орієнтовний розмір висоти definition frame) будуть обчислюватись так [8, 15]:

$$T_1 = \frac{(2M \text{ потоків}) \times (1 \text{ команда})}{(512 \text{ ALU} / \text{clk}) \times (607 \text{ МГц})} = 0,006 \text{ мс.}$$

Шейдер має тільки одну текстурну команду, тоді теоретична робота буде дорівнювати [8, 15]:

$$T_2 = \frac{(\# \text{ потоки}) \times (\# \text{ текстурні_команди})}{(\text{текстура} / \text{clk}) \times (3D_engine_speed)} = \frac{(2M) \times (1)}{(64) \times (607 \text{ МГц})} = 0,051 \text{ мс.}$$

Текстурні блоки використовують частоту ядра і застосовують тривимірну швидкість в оцінюванні. Оцінка роботи пам'яті основана на загальній кількості записаних і прочитаних даних в пам'ять потоку обчислюється так [15]:

$$T_3 = \frac{(\# \text{ потоки}) \times (\text{вхідні} + \text{вихідні_біти_поток})}{(\text{частота_шини}) \times (\text{memory_speed})}$$

В простій програмі (один байт на виході іншого) теоретична робота пам'яті була б такою [8, 15]:

$$T_3 = \frac{(2M \text{ потоків}) \times (64 \text{ біт})}{(384 \text{ біт}) \times (3414 \text{ МГц} \times 2 \text{ DDR})} = 0,048 \text{ мс}.$$

На прикладі шейдеру з одним ALU, одним текстурним блоком, одним вхідним і вихідним бітом – теоретичний час роботи буде 0,051 мс. Проте цей шейдер буде обмежений по швидкодії, тому що текстурні блоки є „вузьким місцем” [8, 15].

Звичайно теоретична робота може лише служити вказівками при збільшенні складності шейдера, адже практична їх реалізація має свою специфіку для різних класів задач і є достатньо складним процесом. Необхідно відзначити, що представлена модель роботи пам'яті основана на „ідеальних” патернах доступу до пам'яті (послідовних або високої когерентності). Важливим аспектом є те, що якщо швидкодія в певній задачі далека від теоретичної, то слід внести алгоритмічні зміни.

ВИСНОВКИ

У ході досліджень набула подальшого розвитку теорія організації високопродуктивних ПІОС, що містить нові моделі та методи, які враховують просторово-багаторівневе подання даних і часово-мережевий принцип їх аналізу, що дозволяє реалізацію високопродуктивних паралельно-ієрархічних обчислювальних процесів в багатоядерних системах з GPU-прискоренням. Проаналізовано процеси паралельно-ієрархічного перетворення та їх обчислювальну складність, в основу яких покладена нова концепція багаторівневої організації обчислень, що дозволяє здійснювати паралельне перетворення не лише просторово ізольованих локальних областей, а досліджувати і формалізувати багатетапні процедури їх паралельної взаємодії у часі на різних рівнях ієрархії ПІОС. Разом з тим, відсутність трудомістких обчислювальних операцій свідчить про достатню простоту алгоритму обчислень, які реалізують паралельно-ієрархічне перетворення, що забезпечує його ефективність при застосуванні у різноманітних прикладних областях, де потрібно сполучення високого ступеня паралелізму і компактної форми представлення даних.

У зв'язку із тим, що перед розробниками гетерогенних обчислювальних систем постають задачі, пов'язані з неоднозначністю вибору можливих варіантів побудови комбінацій функціональних блоків, які можуть бути реалізовані різноманітними апаратно-програмними засобами, здійснено оцінку продуктивності апаратного забезпечення ПІОС на основі GPU-орієнтованої платформи, що переконливо свідчить про її перспективність.

СПИСОК ЛІТЕРАТУРИ

1. Гергель В.П. Высокопроизводительные вычисления для многоядерных многопроцессорных систем. / Гергель В.П. – Н.: ННГУ им. Н.И.Лобачевского, 2010. – 421 с.
2. М.А. Аляутдинов Перспективные средства построения высокопроизводительных масштабируемых вычислительных систем / М.А. Аляутдинов, А. И. Галушкин, Г. В. Тропольская // Нейрокомпьютеры: разработка, применение : Научно-технический журнал. – 2008. – № 8. – С. 70-80.
3. Суперкомпьютерные технологии в науке, образовании и промышленности / Под редакцией: акад. В.А. Садовниченко, акад. Г.И. Савина, чл.-корр. РАН Вл.В. Воеводина. – М.: Издательство Московского университета, 2009. – 232 с.
4. Високопродуктивні гетерогенні обчислювальні комплекси паралельно-ієрархічного оброблення зображень / Тимченко Л.І., Яровий А.А., Мудрик В.В., Кокряцька Н.І. : Proceedings of 2 International Conference [High Performance Computing (HPC-UA'2012)], (Київ, 8-10 жовтня 2012 р.) – Київ, ТОВ "Три К", 2012. – С. 322-327.
5. Скрибцов П.В. Сравнение производительности графических ускорителей и центрального процессора при вычислениях для больших объемов обрабатываемых данных / Скрибцов П.В., Долгополов А.В. // Нейрокомпьютеры: разработка, применение – Москва, Издательство "Радиотехника", 2007. – № 9. – С. 421-425.
6. Owens J. A survey of General-Purpose Computation on Graphics Hardware. / J. Owens, D. Luebke, N. Govindaraju, M. Harris, J. Kruger, A. Lefohn, T. Purcell. –Eurographics. – State of the Art Reports. – 2005. – P. 21-51.
7. Кривов М.А. Тестируем GPU-заменители / Кривов М.А. // Суперкомпьютеры, вып. "Осень-2012", 2012. – С. 38-41.
8. Яровий А.А. Аналіз функціонування програмної моделі GPGPU в контексті організації паралельних обчислень в нейроподібних паралельно-ієрархічних системах / А.А. Яровий // Оптико-електронні інформаційно-енергетичні технології. – 2009. – №1 (17). – С. 42-49.

9. Applied Realization of Neural Network and Neurolike Parallel-Hierarchical System Based on GPGPU / A.A. Yarovyuy : Development and application systems : Proceedings of the 10th International Conference on DAS-2010, May 27-29, 2010, Suceava, Romania – Suceava, Universitatea Stefan cel Mare Suceava, 2010 – P. 351-356. – Режим доступу: <http://www.dasconference.ro/cd2010/data/papers/D50.pdf>
10. Яровий А.А. Прикладні аспекти і перспективи побудови кластерів на основі GPU для реалізації паралельної та паралельно-ієрархічної обробки інформації / А.А. Яровий // Оптико-електронні інформаційно-енергетичні технології. – 2009. – №2 (18). – С. 119-126.
11. Яровий А.А. Прикладна реалізація масштабних нейронних та нейроподібних паралельно-ієрархічних мереж на основі технологій GPGPU [Електронний ресурс] : [Електронне наукове фахове видання] / А. А. Яровий, Ю. С. Богомолов, К. Ю. Вознесенський // Наукові праці Вінницького національного технічного університету. – 2009. – №2. – С. 1-8. – Режим доступу до журн.: http://www.nbu.gov.ua/e-journals/VNTU/2009_2_ua/2009-2.files/uk/09aayogt_ua.pdf. – Назва з екрану.
12. Кожем'яко В.П. Паралельно-ієрархічні мережі як структурно-функціональний базис для побудови спеціалізованих моделей образного комп'ютера : [Монографія] / В.П. Кожем'яко, Л.І. Тимченко, А.А. Яровий. – Вінниця: Універсум-Вінниця, 2005. – 161 с.
13. Тимченко Л.І. Теоретические и прикладные аспекты параллельно-иерархического многоуровневого преобразования цифровых сигналов / Л.И. Тимченко, А.А. Яровой, Н.И. Кокряцкая // Электронное моделирование. – 2013. – Т.35. – №2. – С. 35-54.
14. Яровий А.А. Аналіз нових обчислювальних властивостей паралельно-ієрархічного перетворення інформаційних середовищ та їх комп'ютерне моделювання / А.А. Яровий // Оптико-електронні інформаційно-енергетичні технології. – 2012. – №1 (23). – С. 16-24.
15. GPGPU: General Purpose computations on Graphic Processing Unit [Електронний ресурс]. – Режим доступу: <http://www.gpgpu.org>.
16. Яровий А. А. Методи та засоби організації високопродуктивних паралельно-ієрархічних обчислювальних систем із рекурсивною архітектурою : автореф. дис. на здобуття наук. ступеня докт. техн. наук : спец. 05.13.05 "Комп'ютерні системи та компоненти" / Яровий Андрій Анатолійович ; Вінницький національний технічний університет. – Вінниця, 2013. – 40 с.
17. Свідоцтво про реєстрацію авторського права на твір № 45426. Комп'ютерна програма "Комп'ютерна програма кодування / декодування інформації модифікованими для GPGPU методами прямого та зворотного паралельно-ієрархічного перетворення з оптимізацією масок" / Яровий А.А., Сугак І.М., Трошина А.В. Дата реєстрації Державною службою інтелектуальної власності України 03.09.2012.
18. Свідоцтво про реєстрацію авторського права на твір № 45427. Комп'ютерна програма "Комп'ютерна програма для визначення оптимальних параметрів груп паралельних потоків при організації обчислювального GPGPU-процесу в паралельно-ієрархічному перетворенні" / Яровий А.А., Сугак І.М., Трошина А.В. Дата реєстрації Державною службою інтелектуальної власності України 03.09.2012.
19. Жогов Н. С удвоенной силой. Обзор технологий NVIDIA SLI и ATI CrossFire / Жогов Н. // ЛКИ. – № 4. – 2008. – С. 3-5.

Надійшла до редакції 30.01.2014 р.

ЯРОВИЙ А.А. – д.т.н., доцент, професор кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця, Україна.

КОКРЯЦЬКА Н.І. – к.т.н., доцент, доцент кафедри телекомунікаційних технологій та автоматики, Державний економіко-технологічний університет транспорту, м.Київ, Україна.

НАКОНЕЧНА С.В. – аспірант кафедри телекомунікаційних технологій та автоматики, Державний економіко-технологічний університет транспорту, м. Київ, Україна.

МАТЕЙЧУК М.С. – магістрант кафедри комп'ютерних наук, Вінницький національний технічний університет, м.Вінниця, Україна.

ПОЛЬГУЛЬ Т.Д. – студент кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця, Україна.