

І.Г. Цмоць, д.т.н., проф., В.Я. Антонів,
Національний університет “Львівська політехніка”

ВЕРТИКАЛЬНО-ПАРАЛЕЛЬНИЙ МЕТОД СОРТУВАННЯ МАСИВІВ ЧИСЕЛ

Розроблено метод і потокові графи алгоритмів вертикально-паралельного сортування одновимірних і двовимірних масивів чисел, оцінено швидкодію сортування масивів чисел за даним методом.

Ключові слова: сортування, вертикально-паралельний метод, алгоритми, масиви, граф алгоритму.

Постановка проблеми

Сучасний етап розвитку інформаційних технологій характеризується накопиченням великих масивів даних. При опрацюванні таких масивів найчастіше доводиться використовувати операції сортування та пошуку даних. Метою сортування масивів даних є прискорення пошуку необхідної інформації. Основними шляхами підвищення швидкодії операцій сортування є розробка нових алгоритмів сортування, адаптація їх до архітектури сучасних масово-паралельних комп'ютерних засобів (програмна реалізація) та їх апаратна реалізація у вигляді надвеликої інтегральної схеми (НВІС).

До масово-паралельних комп'ютерних засобів відноситься графічні процесори (GPU – Graphics Processing Unit), які є процесорами класу SIMD (Single Instruction Multiple Data). Особливістю процесорів класу SIMD є те, що в них одна операція використовується одночасно для опрацювання множини незалежних даних. Для розробки програмного забезпечення сортування великих масивів даних, частина якого працює на CPU (центральний процесор), а частина на GPU доцільно використати кросплатформову систему компіляції та виконання програм CUDA.

Програмно-апаратна реалізація операції сортування даних на сучасній елементній базі вимагає розроблення нових паралельних алгоритмів і структур. Паралельні алгоритми сортування, які програмно-апаратно орієнтовані на сучасну елементну базу повинні бути:

- добре структурованими з детермінованим переміщенням даних;
- ґрунтуватися на однотипних операціях з регулярними та локальними зв'язками;
- використовувати конвеєризацію та просторовий паралелізм;
- мати мінімальну кількість виводів інтерфейсу.

У зв'язку з цим особливої актуальності набуває проблема розроблення нових паралельних алгоритмів, орієнтованих на графічні процесори та на НВІС-реалізацію.

Аналіз останніх досліджень і публікацій

Аналіз відомих методів сортування показав, що вони ґрунтуються на базовій операції - порівняння та перестановки, яка зводиться до попарного порівняння значень з масиву даних, який необхідно відсортувати, та їх перестановки у випадку коли порядок цих значень не відповідає умовам сортування [1-3]. Методи сортування на основі такої базової операції відрізняються один від одного вибором пар значень для порівняння. Дані методи сортування можна об'єднати в такі групи: сортування вставкою, обмінне сортування, сортування злиттям і сортування вибором. Кожна з цих груп сортування орієнтовна на послідовну реалізацію.

З аналізу публікацій [1-11] випливає, що основним шляхом зменшення часу сортування масиву чисел є розпаралелення процесу сортування.

Формування цілі статті

Метою роботи є розроблення методу та потокових графів алгоритмів вертикально-паралельного сортування масивів чисел, орієнтованих на їх програмно-апаратну реалізацію.

Основний матеріал

Вертикально-паралельний метод сортування одновимірних масивів чисел. Даний метод сортування передбачає паралельне надходження N чисел розрядними зрізами старшими розрядами уперед і паралельне формування розрядних зрізів відсортованих чисел. Сортування одновимірного масиву чисел $\{D_k\}_{k=1}^N$ здійснюється за n макротактів. При виконанні кожного i -о ($i=1, \dots, n$) макротакту сортування масиву чисел одночасно виконується N базових операцій, які забезпечують формування i -х розрядів N відсортованих чисел. На першому макротакті сортування здійснюється формування старших розрядів всіх чисел. Даний макротакт зводиться до таких операцій:

- 1) підрахунку кількості одиниць у розрядному зрізі

$$S_1 = \sum_{k=1}^N D_{1k} \wedge y_{1k}, \quad (1)$$

де y_{1k} – значення k -о розряду 1-о слова управління, яке дорівнює $y_{1k}=1$;

- 2) формування для $D_1^*, \dots, D_{S_1}^*$ виходів значення 1-о розрядного зрізу P_1 за формулою:

$$P_1 = \bigvee_{k=1}^N D_{k1} \wedge y_{1k}, \quad (2)$$

де D_{k1} – значення 1-о розряду k -о числа масиву;

- 3) формування для $D_{S_1+1}^*, \dots, D_N^*$ виходів значення 1-о розрядного зрізу P_1 за формулою:

$$P_1 = \bigvee_{k=1}^N D_{k1} \wedge \bar{y}_{1k}, \quad (3)$$

де \bar{y}_{1k} – інверсне значення k -о розряду 1-о слова управління;

4) визначення 1-х (старших) розрядів для виходів $D_{11}^*, \dots, D_{N1}^*$ за виразом:

$$D_{k1}^* = \begin{cases} 0, & \text{коли } P_1 = 0 \\ 1, & \text{коли } P_1 = 1 \end{cases} \quad (4)$$

5) формування 2-о слова управління для виходів $D_1^*, \dots, D_{S_1}^*$ за формулою:

$$y_{21}, \dots, y_{2S_1} = \begin{cases} 0, & \text{коли } P_1 = 1, D_{k1} = 0 \\ 1, & \text{коли } P_1 = 1, D_{k1} = 1 \end{cases} \quad (5)$$

6) формування 2-о слова управління для виходів $D_{S_1+1}^*, \dots, D_N^*$ за формулою:

$$y_{2(S_1+1)}, \dots, y_{2N} = \begin{cases} 0, & \text{коли } P_1 = 0, D_{k1} = 1 \\ 1, & \text{коли } P_1 = 0, D_{k1} = 0 \end{cases} \quad (6)$$

Наступні макротакти сортування для кожної групи виходів (чисел) ($D_1^*, \dots, D_{S_1}^*$) і ($D_{S_1+1}^*, \dots, D_N^*$) виконується незалежно та аналогічно до першого макротакту.

Сортування двовимірних масивів чисел. Сортування двовимірного масиву чисел $\{D_{kj}\}_{k=1; j=1}^{N;M}$ з використанням засобів вертикально-паралельного сортування одновимірних масивів чисел передбачає, що вхідні дані надходять паралельно N каналами. Таке сортування доцільно виконувати методом витіснення, базовою макрооперацією якого є вертикально-паралельне сортування одновимірного масиву із $2N$ чисел. В кожному макротакті реалізації даного метода виконується сортування $2N$ чисел, з яких N є вхідними числами, а N – більшими числами з попереднього макротакту роботи. Відсортовані N більших чисел залишаються для виконання наступного макротакту роботи, а N менших записуються у пам'ять. Кількість макротактів, необхідних для сортування двовимірного масиву чисел $\{D_{kj}\}_{k=1; j=1}^{N;M}$, визначається за формулою:

$$p = \sum_{j=1}^M (M - j). \quad (7)$$

Одним із шляхів підвищення швидкодії сортування двовимірного масиву чисел є збільшення кількості базових макрооперацій, які виконуються

паралельно.

Графи алгоритмів вертикально-паралельного сортування масивів чисел. Поточковий граф вертикально-паралельного сортування одновимірного масиву чисел. Такий граф алгоритму повинен забезпечити просторово-часове відображення процесу вертикально-паралельного сортування одновимірного масиву $\{D_k\}_{k=1}^N$ чисел. Особливістю вертикально-паралельного сортування одновимірного масиву чисел є паралельне порозрядне надходження N вхідних чисел (розрядного зрізу) старшими розрядами вперед і паралельне порозрядне формування N відсортованих чисел. Поточковий граф алгоритму вертикально-паралельного сортування одновимірного масиву чисел наведений на рис.1, де Φ_1 і Φ_y - відповідно функціональний та управляючий оператори, ПЕ – процесорний елемент.

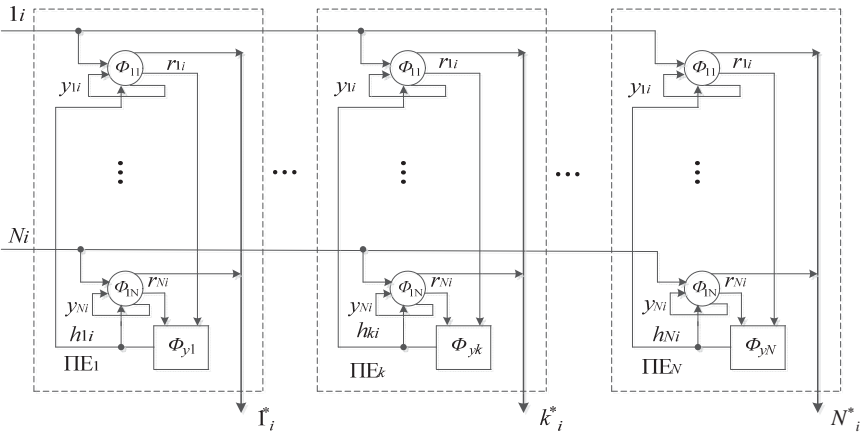


Рис.1. Поточковий граф алгоритму вертикально-паралельного сортування одновимірного масиву чисел

Особливістю поточкового графу алгоритму вертикально-паралельного сортування одновимірного масиву чисел є те, що з кожного k входу одночасно надходить на всі ПЕ. Кількість ПЕ визначається розмірністю масиву, який сортується. В кожному i -у такті роботи в ПЕ $_k$ формується i -й розряд відсортованого k -о числа. Найбільше відсортоване число формується на виході ПЕ $_1$, а найменше – на виході ПЕ $_N$. Кожний ПЕ реалізується на основі N функціональних операторів Φ_1 і одного функціонального оператора Φ_y . У ПЕ $_k$ функціональний оператор Φ_{1k} у кожному i -у такті роботи забезпечує виконання таких операцій:

- 1) обчислення значення i -о розряду для k -о входу за формулою:

$$P_{ki} = \overline{D_{ki}} \wedge y_{ki}, \quad (8)$$

- 2) формування сигналу для підрахунку кількості одиниць за формулою:

$$r_{ki} = (\bar{D}_{ki}^* \vee D_{ki}) \wedge y_i \quad (9)$$

3) визначення k розряду $(i+1)$ -о слова управління за формулою:

$$y_{k(i+1)} = r_{ki} \oplus h_{ki}, \quad (10)$$

де h_{ki} - сигнал управління передачею формування $(i+1)$ -о слова управління, \bar{D}_{ki}^* - інверсне значення i -о розряду k -о відсортованого числа, яке обчислюється в ПЕ $_k$ за формулою:

$$\bar{D}_{ki}^* = \bigvee_{k=1}^N P_{ki}. \quad (11)$$

У ПЕ $_k$ функціональний оператор управління Φ_{yk} забезпечує виконання таких операцій:

1) підрахунку кількості одиниць за формулою:

$$R_{ki} = \sum_{k=1}^N r_{ki} \quad (12)$$

2) визначення h_{ki} за формулою:

$$h_{ki} = \begin{cases} 0, & \text{коли } \sum_k \geq P_{2ki} \\ 1, & \text{коли } \sum_k < P_{2ki} \end{cases} \quad (13)$$

де значення регістра P_{2ki} в i -у такті.

На початку роботи у всіх ПЕ всі розряди слів управління встановлюються в одиницю, а в регістр P_{2k} кожного k -о ПЕ $_k$ записується значення k .

Потоковий граф вертикально-паралельного сортування двовимірного масиву чисел. Потоковий граф модифікованого алгоритму сортування двовимірного масиву чисел $\{D_{kj}\}_{k=1; j=1}^{N;M}$ методом витіснення, наведений на рис.2, де Φ_{mo} – функціональний макрооператор, Φ_y – оператор управління, Φ_k – оператор комутації.

У даному потоковому графі кожний функціональний макрооператор Φ_{mo} виконує сортування одновимірного масиву розміром $2N$ чисел, оператор управління Φ_y – управляє процесом сортування двовимірного масиву чисел, оператор комутації Φ_k – забезпечує з'єднання з функціональними макрооператорами Φ_{mo} у відповідності до алгоритму сортування. Сортування одновимірного масиву розміром $2N$ чисел за вертикально-паралельним методом виконується за n тактів, де n – розрядність чисел, які утворюють макротакт. За макротакт виконується сортування $2N$ чисел, з яких перші N є вхідними, а другі N - більшими числами з попереднього такту роботи. У кожному макрооператорі Φ_{mo} з $2N$ відсортованих чисел N більші чисел затримуються, а N менших поступають на входи наступних макрооператорів.

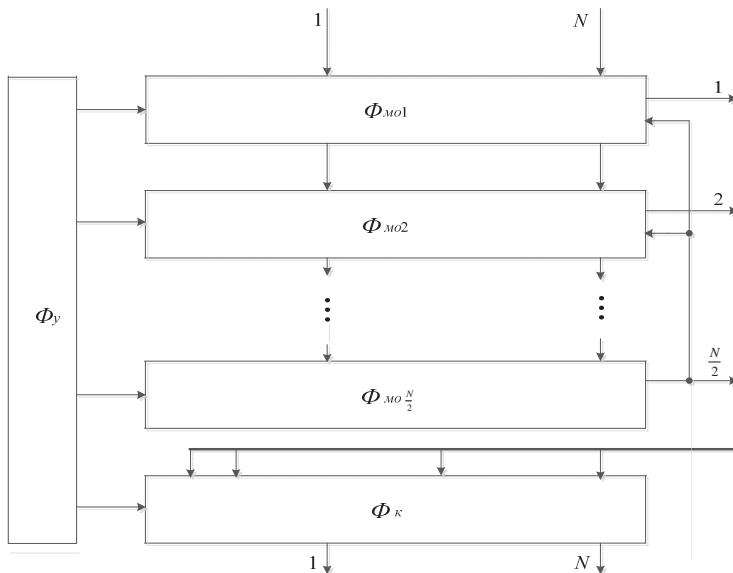


Рис.2. Поточковий граф модифікованого алгоритму сортування двовимірного масиву чисел методом витіснення

Особливістю поточкового графу модифікованого алгоритму сортування двовимірного масиву чисел методом витіснення є використання нелінійних зв'язків між макрооператорами Φ_{mo} . Кількість макротактів, необхідних для сортування двовимірного масиву чисел $\{D_{kj}\}_{k=1; j=1}^{N; M}$ за даним поточковим графом виконується за $2N$ макротакти.

Висновки

1. Метод вертикально-паралельного сортування масивів чисел за рахунок паралельного опрацювання i -о розрядного зрізу масиву чисел забезпечує паралельне формування i -о розрядного зрізу відсортованого масиву чисел, що забезпечує збільшення швидкодії, яка в основному залежить від розрядності чисел, а не їх кількості.

2. Просторово-часове відображення алгоритмів вертикально-паралельного сортування масивів чисел вигляді поточкових графів, в яких за рахунок виявлення паралелізм і керування ним забезпечується ефективна програмно-апаратна їх реалізація.

3. Одним із шляхів підвищення швидкодії сортування двовимірного масиву чисел є збільшення кількості базових макрооперацій, які виконуються паралельно.

1. Д. Кнут. Искусство программирования, том 3: Сортировка и поиск, 2-е изд. - М., 2000.-832с.

2. Цмоць І.Г. Інформаційні технології та спеціалізовані засоби обробки сигналів і зображень у реальному часі. – Львів: УАД, 2005.- 227с.
3. В.П. Гергель. Высокопроизводительные вычисления для многопроцессорных многоядерных систем // Издательство Московского университета , 2010. – 544с.
4. А.В. Борескови др. Параллельные вычисления на GPU, Архитектура и программная модель CUDA// Издательство Московского университета , 2012. – 336с.
5. Грушицкий Р.И., Мурсаев А.Х., Узрюмов Е.П. Проектирование систем на микросхемах программируемой логики. – СПб.: БХВ-Петербург, 2002. – 608с.
6. С.Кун. Матричные процессоры на СБИС:-М.:Мир,1991.-672 с.
7. Кормен Т., Лейзерсон Ч., Ривест Р. Алгоритмы: построение и анализ: Пер. с англ. / Под ред. А. Шеня. – М.: МЦНМО: БИНОМ. Лаборатория знаний, 2004. – 960 с.
8. Левитин Ананий. Алгоритмы: введение в разработку и анализ. :Пер. с англ. - М.: Издательский дом «Вильямс», 2006. - 576с.
9. Лорин Г. Сортировка и системы сортировки. – М. : Мир. 1983. – 384 с.
10. Мельничук А.С., Луценко С.П., Громовий Д.С., Трофимова К. В. Аналіз методів сортування масиву чисел. Технологический аудит и резервы производства - №4/1(12), 2013. – С. 37-40.
11. Воеводин В. В., Воеводин Вл. В. Параллельные вычисления. — СПб: БХВ-Петербург, 2002. — 608 с.

Поступила 6.10.2016р.

УДК 004.75

І.Г. Цмоць, д.т.н., О.В. Скорохода, к.т.н., В.І. Роман
кафедра автоматизованих систем управління
Національний університет «Львівська політехніка», м. Львів

СХОВИЩА ДАНИХ БАГАТОРІВНЕВИХ СИСТЕМ УПРАВЛІННЯ ЕНЕРГОЕФЕКТИВНІСТЮ

Анотація. Досліджено та проаналізовано існуючі методи і засоби організації та розробки високонавантажених сховищ даних для організації доступу до великих даних (big data). Розглянуто лямбда-архітектуру, запропоновано технологічне рішення для реалізації сховища даних на її основі.

Abstract. Existing methods and means of organization and development of high-loaded warehouses for providing access to big data have been investigated and analysed. The lambda architecture has been considered, a technological solution for implementing a data warehouse based on it has been proposed.

Ключові слова: сховища даних, великі дані, багаторівневі системи управління, енергоефективність.

Keywords: data warehouses, big data, multi-level management system, energy efficiency.