КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ

На правах рукопису

МАСЛЕННІКОВ Олег Володимирович

УДК 681.322

МЕТОД ПРОЕКТУВАННЯ ТА СТРУКТУРИ СИСТОЛІЧНИХ ПРОЦЕСОРІВ ДЛЯ ВИРІШЕННЯ ЗАДАЧ ЛІНІЙНОЇ АЛГЕБРИ

Спеціальність 05.13.13 — обчислювальні машини, комплекси, системи та мережі

АВТОРЕФЕРАТ

дисертації на здобуття вченого ступеня кандидата технічних наук

КИЇВ — 1993

КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ

На правах рукопису

МАСЛЕННІКОВ ОЛЕГ ВОЛОДИМИРОВИЧ

УДК 681.322

МЕТОД ПРОЕКТУВАННЯ ТА СТРУКТУРИ СИСТОЛІЧНИХ ПРОЦЕСОРІВ ДЛЯ ВИРІШЕННЯ ЗАДАЧ ЛІНІЙНОЇ АЛГЕБРИ

Спеціальність 05.13.13 - обчислювальні машини, комплекси, системи та мережі

Автореферат
дисертації на здобуття вченого ступеню
кандидата технічних наук

Київ - 1993

Робота виконана у Київському політехнічному інституті.

Науковий керівник: доктор технічних наук, професор КАНЄВСЬКИЙ Ю.С.

Офіційні опоненти: доктор технічних наук, професор ДОДОНОВ О.Г., кандидат технічних наук НЕКРАСОВ Б.А.

Провідна установа: Інститут проблем моделювання в енергетиці АН України

Захист відбудеться "15" 12 1993 р. о 14.30 на засіданні спеціалізованої Ради Д 068.14.09 в Київському політехнічному інституті (м.Київ, пр.Перемоги,37, корпус 18, ауд.306).

Відгуки на автореферат у двох примірниках, засвідчені печаткою установи, просимо надсилати за адресою: 252056, м.Київ, проспект Перемоги,37, КПІ, Ученому секретареві КПІ.

З дисертацією можна ознайомитись в бібліотеці Київського політехнічного інституту.

Автореферат розісланий "14" 10 1993 р.

Учений секретар спеціалізованої Ради Д 068.14.09 д.т.н., професор Бузовський О.В.

АНОТАЦІЯ

Метою роботи являється підвищення ефективності методів проектування та синтез з їх допомогою структурних схем спеціалізованих паралельних обчислювальних пристроїв (СПОП), орієнтованих на реалізацію методами НВІС-технології (СПОП на НВІС), а також розробка паралельних обчислювальних схем алгоритмів (ОСА) вирішення задач лінійної алгебри (ЛА).

В відповідності з поставленою метою вирішуються наступні задачі:

1. аналіз основних задач ЛА та методів їх вирішення з метою визначення ступеню їх придатності для паралельної реалізації на СБІС;
2. аналіз відомих методів структурного проектування СПОП на НВІС та деякі направлення їх розвитку з точки зору практичної реалізації синтезованих структур;
3. розробка метода синтезу синхронізуючої функції, описуючої роботу структури СПОП в часі;
4. розробка підходів до синтезу структур СПОП фіксованого розміру для вирішення задач ЛА;
5. розробка способів перетворення решітчатих функціональних графів (РФГ) алгоритмів ЛА для оптимізації апаратно-часових витрат при їх реалізації на паралельних обчислювальних системах (ОС);
6. розробка та дослідження паралельних ОСА та структурних схем СПОП на НВІС для вирішення деяких задач ЛА.

Автор захищає такі основні положення та результати:

1) метод синтезу часової компоненти відображення РФГ алгоритму в структурні схеми СПОП на НВІС;
2) способи ізоморфних та гомоморфних перетворень РФГ ряду алгоритмів ЛА для їх ефективної паралельної реалізації;
3) спосіб проектування структурних схем СПОП на НВІС, незалежних від розмірів вирішуємих задач;
4) паралельні ОСА та структурні схеми СПОП на НВІС для вирішення різних задач ЛА.

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність досліджень. До задач ЛА зводиться цілий ряд проблем з самих різних областей науки і техніки. Однак вирішення більшості задач ЛА потребує виконання $O(n^3)$ (де n - порядок оброблюваних матриць) операцій типу множення з додаванням. Тому в ряді випадків тільки СПОП, орієнтовані на вирішення визначеного класу задач і максимально враховуючі їх специфіку, придатні для виконання останніх в реальному масштабі часу. Поява високошвидкісних НВІС, які мають на одному кристалі декілька процесорних елементів (ПЕ) середньої складності, прискорило дослідження в структурному проектувані СПОП на НВІС, основними представниками яких являються систолічні процесори (СП) та процесорні матриці (ПМ). При цому постійне підвищення складності вирішуваних задач та скорочення строків на розробку таких СПОП, а також постійний ріст вартості помилок евристичного проектування привели до появи ряду формалізованих та частково автоматизованих методів структурного проектування СПОП на НВІС.

Значними недоліками відомих методів проектування СПОП на НВІС є беспосередня залежність одержаних архітектур ОС від розмірів структур початкових даних вирішуваної задачі, а також складність одержання синхронізуючої функції, описуючої роботу СПОП в часі, в випадках зниження розмірності графа структури СПОП відносно розмірності РФГ початкового алгоритму (ПА) більше, ніж на одиницю. Тому актуальною являється задача їх удосконалення шляхом усунення названих недоліків. Крім того, множина одержаних з допомогою указаних методів структурних рішень однозначно визначається базисним РФГ ПА. Разом з тим, із-за відносно високої вартості реалізації з'єднань в НВІС, ріст ефективності застосування СПОП на НВІС можна чекати тільки тоді, коли ПА підготовлений (модифікований) з урахуванням не тільки збалансованого розподілу робіт, але і з дотриманням вимог локальності (тобто коротких ліній зв'язку). Отже, актуальною також є розробка ОСА вирішення задач ЛА (НВІС-алгоритмів), дозволяючих оптимізувати апаратні витрати на побудову СПОП на НВІС та мінімізувати час вирішення вказаних задач на паралельних ОС.

Методи дослідження. В роботі використанні методи ЛА, теорії графів, теорії лінійних операторів, теорії множин, теорії проектування ЕОМ та систем.

Наукова новітність роботи полягає в слідуючому:

1. розроблено метод синтезу часової компоненти відображення графу алгоритму розмірності n в структурні схеми СПОП розмірності $m = n-2$, який на відміну від відомих дозволяє зменшити трудомісткість цього процесу та повністю формалізувати його;

2. запропоновані способи ізоморфних та гомоморфних перетворень РФГ ряду алгоритмів ЛА, які дозволяють оптимізувати параметри відповідних структур СПОП на НВІС чи реалізацію вказаних задач на паралельних ОС.

3. розроблено спосіб проектування структурних схем СПОП на НВІС, розміри яких не залежать від розмірів реалізуємих іми задач.

Практична цінність роботи. Запропоновані в роботі метод та способи дозволяють спростити, систематизувати, а також в деякій мірі автоматизувати розробку структурних схем СПОП на НВІС. При іх допомозі автором розроблено ряд структур СПОП на НВІС та НВІС-алгоритмів для вирішення основних задач ЛА, практична користь та новітність яких підтвержуються авторськими свідоцтвами СРСР .

Реалізація результатів роботи. Теоретичні та прикладні результати дисертаційної роботи покладені в основу лабораторних робіт з курсу "Спеціалізовані ЕОМ" для студентів КПІ спеціальності 2201 - обчислювальні машини, комплекси, системи та мережі.

Апробація работи. Основні положення дисертаційної роботи обговорювались на:

- 2-й Всесоюзній нараді "Конвейерні обчислювальні системи", Київ, 1988 р.;
- Всесоюзній конференції "Методи та мікроелектронні засоби цифрового перетворення та обробки сигналів", Рига, 1989 р.;
- 1-й Всесоюзній конференції "Однорідні обчислювальні середовища та систолічні структури", Львів, 1990 р.;
- Всесоюзному семінарі "Теорія та практика створення систем технічного зору" Москва, 1990 р.
- Всесоюзних семінарах "Однорідні обчислювальні середовища та систолічні структури", Львів, 1989 - 1992 рр.

Публікаціі. По темі дисертаціі опубліковано 26 друкованих робіт, включаючи 14 авторських свідоцтв СРСР та рішень про видачу патентів.

Структура та обсяг работи. Дисертація складається з вступу, чотирьох глав, висновку та списку використаноі літератури з 92 найменувань. Робота містить 168 сторінок машинописного тексту та 27 сторінок графічного матеріалу.

В першій главі проведено аналіз основних задач ЛА та методів іх вирішення з метою визначення придатності останніх до паралельноі реалізаціі на НВІС. Досліджені відомі методи структурного проектування СПОП на НВІС та визначені основні напрямки іх удосконалення з точки зору безпосередньоі практичноі придатності.

В другій главі для випадку $n-m=2$ розроблено більш простий та суворо формалізований метод синтезу часовоі компоненти відображення РФГ алгоритму в структурні схеми СПОП, яка в останніх описує часову розгортку обчислень. Розроблені способи ізоморфних та гомоморфних перетворень РФГ ряду алгоритмів ЛА для їх ефективної паралельної реалізації, в тому числі й на НВІС. Розроблено спосіб структурного синтезу СПОП на НВІС, дозволяючий проектувати структурні схеми СПОП, незалежно від розмірів вирішувальних задач.

В третій главі показані можливості та переваги розроблених способів ізоморфних та гомоморфних перетворень РФГ алгоритмів ЛА та методу визначення часової розгортки обчислень при синтезі структурних схем СПОП та НВІС-алгоритмів для вирішення систем лінійних алгебраїчних рівнянь (СЛАР), LU - та LL^t - розкладення матриць.

В четвертій главі на основі розробленого способу структурного проектування СПОП на НВІС синтезовані структурні схеми СПОП фіксованого розміру, кожна з яких призначена для вирішення цілого ряду задач ЛА незалежно від їх розмірів.

ОСНОВНИЙ ЗМІСТ РОБОТИ

На основі аналізу відомих методів вирішення основних задач ЛА встановлено, що більшості з них характерні: потреба в виконанні великого обсягу періодично повторюючихся обчислень, а також висока ступінь внутрішнього паралелізму та регулярність зв'язків між функціональними операторами (ФО) алгоритмів. Виділені ті з них, які найбільше придатні для паралельної реалізації на НВІС.

На основі відомої процедури побудови РФГ безумовних регулярних алгоритмів, заданих у вигляді програми на мові високого рівня, побудовано набір базових РФГ виділених алгоритмів ЛА та виконана їх класифікація по размірності n простору представлення РФГ, типу та довжині дуг між вершинами, функціональному змісту і внутрішній будові вершин, довжині критичного шляху.

На основі аналізу існуючих методів структурного проектування СПОП на НВІС для вирішення задач ЛА та вимог до них встановлено, що а) синтез по будь-якому методу являється лише автоматизованим, а не автоматичним, тобто вимагає участі проектувальника в процесі проектування; б) лише небагато методів допускають одержання структур СПОП довільної розмірності m незалежно від розмірності n РФГ алгоритму, але навіть в них в випадку $n-m>1$ пошук часової компоненти відображення РФГ в структуру СПОП є досить важким.

Існуючі методи, як правило, базуються на представленні початкового безумовного регулярного алгоритму А у вигляді РФГ G такого, що множина його вершин відповідає множині ФО алгоритму, причому кожна вершина описується одним і тільки одним елементом (точкою або вектором з початку координат в цю точку) декартового добутку (цілочисленої решітки) $K^n = \langle \bar{K} = [k_1\ k_2\ \ldots\ k_n]^t \mid k_i \in Z^n,\ u_i \le k_i \le h_i \rangle$, де Z^n - лінійний простір, в якому знаходиться K^n. Дуги графу G відповідають безпосереднім інформаційним залежностям між його ФО. В деяких методах РФГ неявно задається в компактній формі у вигляді характеристичної матриці $D = [\bar{d}_1\ \bar{d}_2\ \ldots\ \bar{d}_l]^t$ векторів -стовпців безпосередньої інформаційної залежності (тобто дуг графа G) по змінних алгоритма та інтервалів змін індексів цих змінних. Далі граф G за допомогою лінійного, монотоного та ін'єктивного оператора F відображається в структурну схему $C = \langle S, T, \Phi \rangle$ СПОП, де S - орграф структури, заданий в просторі Z^m $(m<n)$, T - синхронізуюча функція, Φ - набір операцій, виконуюмих ПЕ. Оператор F називають ST-відображенням, оскільки він складається з просторової F_S та часової F_T компонент, перша з яких визначає структуру S, а друга - функцію T, а обидві - множину Φ:

$$F_S:\ K^n \to K^m_F, \qquad F_S(\bar{K}) = [k^F_1\ k^F_2\ \ldots\ k^F_m]^t,$$
$$F_T:\ K^n \to K^1_F, \qquad F_T(\bar{K}) = [\hat{t}],$$

де $\hat{t}$ - дискретний період часу (машинний такт), в якому ФО, відповідаючий вершині з координатами $\bar{K}$, виконується ПЕ з координатами $F_S(\bar{K})$ (вважається одиничний час виконання будь-якого ФО). Вказаний ПЕ в цей час не може виконувати ніяких інших ФО, що приводить до умови ін'єктивності

$$\forall \bar{K}_1,\ \bar{K}_2 \in K^n\ (F_S(\bar{K}_1) = F_S(\bar{K}_2) \to F_T(\bar{K}_1) \ne F_T(\bar{K}_2)), \qquad (I)$$

в той час як монотонність функції F (точніше, її компоненти F_T) визначається слідуючим чином:

$$F_T \times \bar{d}_i \ge 1, \qquad i = 1,\ 2, \ldots, l. \qquad (2)$$

На сьогодні методика синтезу структурних схем C для часткового випадку $r=n-m=1$, відповідного як раз СП, досить добре розроблена. Зауважимо тільки, що умова (I) в цьому випадку буде виконуватись, якщо

$$\det F_S \ne 0. \qquad (3)$$

Однак, якщо врахувати, по-перше, що для подавляючої більшості алгоритмів ЛА з ФО на рівні слів $n = 3$, а з ФО на рівні розрядів $n = 4$, та, по-друге, що сучасні можливості технології та ширина каналу вводу-виводу НВІС дозволяють практично реалізувати тільки одномірні структури (m = I) з обробкою на уровні слів або двумірні ($m = 2$) з обробкою на рівні розрядів, з'являється необхідність в розгляді випадку, коли $r=2$.

Синтез структурних схем С в цьому випадку, відповідному ПМ, також здійснюється на основі ST-відображення F, причому пошук просторової компоненти F_s може не відрізнятися від попереднього випадку, коли $r = 1$. Однак, значно складнішою та трудомістською задачею стає при цьому одержання синхронізуючої функції Т структури. Відомі методи, працюючі при $r>1$, використовують або багатоступінчату стратегію пошуку ST-відображення, яка потребує евристичних дій для одержання результуючої функції F_T, або багатомірну часову розгортку обчислень, яка ускладнює формулювання умови (3).

Разом з тим з практики проектування відомо, що найкращі структури зазначених СПОП одержують при проекції РФГ G відповідних їм алгоритмів або на осі координат простору Z^r, або на координатні площини. Це пов'язано з тим, що граничні вершини РФГ G алгоритмів ЛА, як правило, належать гіперплощинам, паралельним координатним площинам простору Z^r. Тому будь-яка проекція цих графів на прямі або площини, відмінні від координатних, приводить до збільшення кількості ПЕ та/або зв'зків між ними в результуючій структурі, а також нерідко погіршує часові характеристики одержаного СПОП. В роботі показано, що прийняття до уваги зазначеного обмеження на вибір проекції графа G значно зпрощує пошук часової компоненти F_T при $r=2$, та пропонується новий метод синтезу часової розгортки обчислень в СПОП на НВІС, придатний для пошуку F_T в цьому випадку.

У відповідності з ним для пошуку F_T виконується перехід від представлення РФГ алгоритма в стандартному (що складається з стовпців одиничної матриці) ортонормованому базисі лінійного простору Z^r до його представлення в ортогональному базисі структури $B = [\,\bar{f}^t_1 \ldots \bar{f}^t_m\ \bar{v}_1\ \bar{v}_2\,]$, де вектори-стовпці $\bar{f}^t_1, \ldots, \bar{f}^t_m$ є векторами-строками матриці F_s, а вектори $\bar{v}_1$, $\bar{v}_2$ створюють ортогональний базис ядра $\mathrm{Ker}\ F_s$ оператора F_s. Тоді, якщо вектор $K_B = [k^B_1, k^B_2 \ldots k^B_n]$ є представлення в базисі структури довільного вектора $K \in K^n$, то співвідношення $K_B = B^t \cdot K$ буде задавати перетворення координат вектора при переході від одного базису до іншого.

З визначення базису В витікає, що перші m координат будь-якого вектора K_B - це представлені в новому базисі координати деякого ПЕ структури S. Інакше кажучи, співвідношення $K_{B1} = [k^B_1\ k^B_2 \ldots k^B_m]^t = B^t_1 \cdot K$, де $K \in K^n$, а B_1 - підматриця матриці В, створена її першими r вектор-стовпцями, означає що ФО, відповідаючий вершині К виконується ПЕ, координати якого в новому базисі дорівнюють K_{B1}. Тоді множина $\Pi(K_{B1})$ така, що $\Pi(K_{B1}) = \langle K_B = B^t \cdot K = [k^B_1\ k^B_2\ \dots\ k^B_n]^t \mid k^B_1, k^B_2, \dots, k^B_m = const;\ \bar{K} \in K^n \rangle$, буде задавати в цьому ж базисі координати всіх тих вершин графу G, які проецюються в зазначений ПЕ і будуть в ньому виконуватись.

Перехід від стандартного до ортогонального базису B приводить також до трансформації лінійного оператора F_T в оператор F^B_T слідуючим чином: $F^B_T = [f^B_{m+1,1}\ f^B_{m+1,2}\ \dots\ f^B_{m+1,n}] = F_T \cdot B$. Тоді, якщо для будь-якої множини $\Pi(K_{B1})$ гіперплощина з нормаллю F^B_T при будь-якому своєму положенні містить не більше одного елементу зазначеної множини, то для будь-яких двох вершин графа G з координатами $K_1, K_2 \in K^n$ такими, що $F_S(K_1) = F_S(K_2)$, буде виконуватись $F_T(K_1) \neq F_T(K_2)$. Отже, для будь-яких двох вершин $K_{B1} = [k^B_{1,1}\ k^B_{1,2} \dots\ k^B_{1,n}]^t$ та $K_{B,2} = [k^B_{2,1}\ k^B_{2,2} \dots k^B_{2,n}]^t$ графа G, з однієї множини $\Pi(K_{B1})$, різниця між моментами виконання ФО, які їм відповідають, складає $\Delta t = F^B_T(\bar{K}_{B,2}) - F^B_T(\bar{K}_{B,1}) = \sum_{i=n-1}^{n} f^B_{m+1,i}(k^B_{2,i} - k^B_{1,i})$ тактів, тобто визначається тільки двома останніми координатами як оператора F^B_T, так і вершин $K_{B,1}$, $K_{B,2}$.

На основі приведених викладок в роботі доводяться слідуючі прості умови ін'єктивності ST-відображення для випадку $r=2$:

$$|f^B_{m+1,n}| \geq M^B_{n-1},\quad f^B_{m+1,n-1} = \gamma_n,\ \text{або}$$
$$|f^B_{m+1,n-1}| \geq M^B_n,\ -f^B_{m+1,n} = \gamma_{n-1},$$

та обгрунтовується слідуюча процедура пошуку $F_T = [f_{m+1,1}\ f_{m+1,2} \dots f_{m+1,n}]$, яка проводиться на основі вибраного F_S, а також відомих діапазонів змін M^B_i та мінімальних приристів γ_i $(i=1,\dots,n)$ n різних індексів змінних по всім n координатам простору Z^n:

1. Серед n стовпців вибраної у відповідності з прийнятим обмеженням просторової компоненти F_S, що визначає m-мірну структуру S СПОП, находять $r=2$ нульових стовпців, номера яких присвоюються змінним, наприклад, x та y.
2. Із заданних значень M^B_i та γ_i вибираються значення M^B_x та γ_y або M^B_y та γ_x, які присвоюються відповідно $f_{m+1,y}$ та $f_{m+1,x}$.
3. Решті елементів $f_{m+1,i}$ $(i=1,\dots,n;\ i \neq x,\ i \neq y)$ $(m+1)$-й строки матриці F присвоюються значення з множини $\{-1,0,1\}$.
4. З одержаної множини значень F_T вибираються ті, які задовольняють умові монотонності (2) та залишають попарно різними всі стовпці та строки матриці F.

На основі аналізу існуючих методів структурного проектування

СПОП на НВІС установлено, що всі вони дозволяють одержати деякі множини структур, адекватні саме заданому запису алгоритму вирішення задачі, що значно звужує клас одержаних архітектур та не гарантує одержання рішень, відрізняючихся високою ефективністю. Тому при структурному синтезі СПОП на НВІС етапу просторово-часового відображення РФГ алгоритму в адекватні апаратно-програмні засоби пропонується ввести етап трансформації РФГ з допомогою цілеспрямованих ізоморфних та/або гомоморфних перетворень. Їх можна виконати, по-перше, шляхом ізоморфного перетворення (занурення) РФГ в граф другої структури. По-друге, дякуючи великій свободі в порядку передачі непереобчислюваних змінних алгоритму між вершинами графу появляється можливість цілеспрямовано генерувати на основі початкового запису алгоритму вирішення задачі множини його РФГ. В-третіх, навіть для переобчислюваних змінних можна іноді допустити зміну порядку їх формування, використовуючи властивості комутативності та асоціативності математичних операцій.

Відомо, що трьохмірні базисні РФГ ряду алгоритмів вирішення СЛАР і трикутного розкладення матриць можна звести до двухмірного систолічного трикутного графу (СТГ), оскільки всі вони породжуються єдиною обчислювальною схемою, на i-му кроці якої ($i = 1,2,\dots,N$; де N - кількість стовпців початкової матриці A^*) обробці підлягає зпочатку стовпець $\bar{a}_i^i$ матриці (де $\bar{a}_i^1 = \bar{a}_i^*$), який потім взаємодіє з усіма іншими стовпцями, за виключенням перших $(i-1)$ стовпців.

В зв'язку з цим, в роботі пропонується спосіб ізоморфного перетворення (занурення) СТГ в систолічний тороідальний граф (СТОГ). Припустимо, що з любої внутрішньої вершини зазначеного СТОГ виходять (а також входять в неї) дві дуги з координатами $\bar{d}_1=(1,1)$ та $\bar{d}_2=(1,-1)$. В крайніх вершинах СТОГ, розміщених на його лівій або правій границях, замість однієї з вказанних дуг з'являється дуга $\bar{d}_3=(1,0)$. Тоді суть пропонованого занурення, яке справедливо тільки в випадку непарних значень N може бути описана таким чином:

1. Вершина (i,i) початкового СТГ (де $i = 1,2,\dots,N$) розташовується в вершині $\bar{K}_i = (i^*,j^*) = (2i-1,\ 1)$ СТОГ.

2. Для данного i інші вершини (i,j) СТГ (де $j = i+1,\ i+2,\ \dots,N$) розміщаються послідовним образом вздовж напрямку вектора $\bar{d}_1$ (починаючи з вузла $\bar{K}_i + \bar{d}_1$) до зустрічі з правою границею СТОГ. Тоді проходить поступова, з участю вектора $\bar{d}_3=(1,0)$, зміна напрямку розміщення вершин на $\bar{d}_2= (1,-1)$.

Результатом описаної трансформації є зменьшення ширини початкового СТГ G^* зі значення N вершин вздовж осі j до значення $N^* = (N+1)/2$ вершин вздовж осі j^* в одержаному графі G^*, тобто майже вдвоє. Вказане зменшення, в свою чергу, дозволяє скоротити число ПЕ та повисити їх завантаженість в лінійних СПОП, реалізуючих відповідні алгоритми ЛА. При цьому в роботі доведено, що оптимальна по швидкості виконання алгоритму часова разгортка відповідає значенню $F_T = (1,0)$, яке у випадку алгоритмів, описуваних трьохмірним РФГ, задає разгортку уже не по тактах, а по макротактах. Кожний з останніх містить, як правило, M тактів (де M - кількість строк матриці A^*), необхідних для поелементної обробки стовпця матриці.

На основі запропонованого способу перетворення СТГ в роботі синтезовані паралельні ОСА та структурні схеми лінійних СПОП на НВІС, які релізують численно стійкі методи Жордана-Гауса для вирішення СЛАР та Холецького для трикутного розкладу симетричних матриць. Синтезовані структурні схеми мають $N/2$ ПЕ кожна (де N - порядок початкової матриці) та характеризуються асимптотичною завантаженістю при вирішені вказаних задач відповідно $\eta_4^{\wedge} \approx I$ і $\eta_9^{\wedge} \approx 0,66$, тобто вдвоє більшою, ніж кращі з відомих структур СПОП. Такою ж завантаженістю буде характеризоватись будь - яка паралельна ОС (яка містить таку ж кількість ПЕ та міжпроцесорних зв'язків) при реалізації на ній запропонованих ОС. Важливою особливістю одержаних СПОП є також те, що всі складні операції типу ділення та знаходження зворотньої величини (а також порівняння) реалізуються в першому ПЕ, в той час як інші ПЕ виконують тільки множення з додаванням та зберігання інформації. Крім того, з допомогою ряду цілеспрямованих ізоморфних та гомоморфних перетворень базисного РФГ матрично-векторного множення спроектована структурна схема СПОП на НВІС, реалізуюча методи простої ітерації та Гауса-Зейделя для ітераціонного вирішення СЛАР, яка має в чотири рази менший час виконання однієї ітерації ціною менш ніж двукратного збільшення апаратурних затрат в порівнянні з кращими з відомих схем СПВУ, реалізуючих зазначені методи.

На основі аналізу існуючих методів структурного проектування СПОП на НВІС також встановлено, що ні один з них не дозволяє проектувати СПОП фіксованого розміру для реалізації задач довільних розмірів. В то й же час архітектури реальних паралельних ОС характеризуються рядом обмежень на свої ресурси, найбільш істотними з яких являються кількість ПЕ та об·єм локальної пам·яті (ЛП), а також пропускна можливість каналів обміну як між ПЕ, так і з зовнішньою памяттю.

Для синтезу структурних схем СПОП на НВІС, незалежних від розмірів вирішуємих задач запропоновано використовувати декомпозицію (розбивку) віртуального обчислювального процесу великих розмірів на множину взаємозв'язаних підпроцесів менших розмірів, та виконувати потім відображення останніх на структуру СПОП фіксованого розміру таким чином, щоб виконання підпроцесів в них здійснювалось частково паралельно або послідовно. На основі аналізу рівнів представлення обчислювального процесу (рівня початкових алгоритмів, РФГ та ОСА), встановлено, що кращим для використання процедур декомпозиції є рівень РФГ.

Для здійснення занурення алгоритмів довільних розмірів в архітектуру СПОП з фіксованою кількістю ПЕ РФГ алгоритму будем регулярно розбивати на множину підграфів. При цьому відображення декомпонованого графу в структурну схему з фіксованою кількістю ПЕ може виконуватися згідно з однією із слідуючих двох запропонованих стратегій (або їх комбінацій).

Нехай початковий РФГ, вершини якого розміщені в n-мірному просторі, за допомогою максимум n-1 сімейств паралельних гіперплощин розбивається на множину підграфів. Тоді згідно з локально паралельною глобально послідовною (ЛПГПС) стратегією будь який граф відображається в одну і ту ж результуючу схему. При цьому одержані підграфи повинні виконуватись один за другим в відповідності з деяким порядком. Це означає, що вихідні данні одного підграфу являються вхідними для послідуючих і з·являється необхідність буферізації у зовнішній памяті проміжних результатів, зумовлених інформаційними залежностями між підграфами. Таким чином, за рахунок додаткових обмінів з зовнішньою памяттю, виконання яких не завжди можно організувати на фоні обчислень, вдається, як правило, уникнути додаткової (порівняно з випадком необмеженого паралелізму) локальної памяті ПЕ. Це дозволяє одержати СПОП з ПЕ, оптимізованими по об·єму ЛП. Однак, при наявності зустрічних обмінів проміжними результатами між підграфами вимагаємий послідовний порядок виконання підграфів може не існувати, що обмежує область використання данної стратегії. Накінець, відмінності в будові підграфів, які відображаються в одну і ту ж структурну схему, можуть привести до збільшення накладних витрат на її реконфігурацію та управління.

Протилежні, в основному, властивості має локально послідовна глобально паралельна (ЛПСГП) стратегія, що припускає виділення свого ПЕ для кожного одержаного підграфу, в зв'язку з чим кількість ПЕ схеми рівняється числу підграфів, а кожний ПЕ повинен послідовно виконувати оператори (вершини) відповідного йому підграфу. В результаті структура міжпроцесорних зв'язків СПОП буде визначатися інформаційними залежностями між підграфами і виникає потреба вводу в кожний ПЕ додаткової ЛП, яка призначена для зберігання проміжних результатів, народжених виконанням підграфу. Таким чином, при додержанні обмежень на обсяг ЛП, а також з урахуванням додаткових витрат на його управління, ця стратегія, по-перше, дозволяє уникнути обмінів проміжними результатами з будь-якою зовнішньою памяттю. По-друге, з'являється можливість зменшити вимоги до пропускної можливості каналів зв'язку ПЕ, оскільки час реалізації підграфу, який визначається загальною кількістю його вершин, при збільшенні розмірів підграфу росте швидше, ніж об'єм данних, передаваємих між сусідніми підграфами, який визначається кількістю граничних вершин підграфу. При цьому межа використання зазначеної можливості визначається обмеженнями на обсяг ЛП, який залежить від розмірів підграфу. В-третіх, на відміну від ЛПГПС, ця стратегія дозволяє оперувати з графами теоретично любого вигляду, оскільки будь-якому способу вибору підграфів відповідає тепер непорожня множина синхронізуючих функцій, задающих в часі порядок виконання вершин початкового РФГ процесорними елементами СПОП.

Враховуючи викладене, в роботі обгрунтовується доцільність використання узагальненої (або ієрархічної) стратегії декомпозиції, включаючої обидві розглянуті вище стратегії, та на цій основі пропонується слідуючий спосіб синтезу структурних схем СПОП на НВІС фіксованого розміру:

I. Цілочислена решітка K^n, в вузлах якої розміщені вершини решітчатого графу G алгоритму, n способами разбивається на однакові прямокутні підрешітки, розміри яких для j-го способу розбивки задовольняють слідуючим відношенням:

$$q_j = M_j, \qquad j = 1, 2, \ldots, n,$$
$$1 \le q_i \le M_i, \qquad i \in \{1, 2, \ldots, n\},\ i \ne j,$$

де M_i, M_j - розміри решітки K^n вздовж одноіменних координатних вісей простору Z^n.

2. Кожній підрешітці ставиться у взаємно однозначне співвідношення В-вершина. Співкупність всіх В-вершин для фіксованого j-го способу розбивки утворюють В-граф G^j_B. Вершини останнього розташовані в вузлах решітки $K^{n-1}_{B,j}$, а дуги одержуються шляхом проекції дуг графу G вздовж j-ї координатної осі простору Z^n.

3. Для кожного з одержаних таким чином графів G^j_B знаходиться множина допустимих часових розгорток, задаючих суворо послідовний порядок виконання функціональних операторів всередині В-вершин. На цій основі визначаються вимоги до організації та обсягу ЛП ПЕ синтезованої структурної схеми, вираженому як функція розмірів q^j_i підрешіток. Виходячи потім з обмежень на організацію та обсяг ЛП ПЕ цільової ОС, знаходяться максимально допустимі розміри q^j_i підрешіток, що дозволяє мінімізувати вимоги до пропускної можливості каналів зв'язку ПЕ.

4. Кожна з одержаних решіток $K^{n-1}_{B,j}$, в вузлах якої розміщені вершини графу G^j_B, $N^m_v = C^m_{n-1}$ способами (де C^m_{n-1} - число комбінацій з $n-1$ по m) розбивається на однакові прямокутні підрешітки, розміри p^j_i яких при $m=1$ задаються співвідношеннями

$$\left.\begin{aligned} I \le p^j_i \le L^j_i =]M^j_i/q^j_i[, &\quad \text{якщо } i = r, \\ p^j_i = L^j_i, &\quad \text{якщо } i \ne r, \end{aligned}\right\} \qquad (4)$$

де r - номер тієї координатної осі простору Z^{n-1}_j, перпендикулярними до якої являються всі розсікаючі гіперплощини єдиного їх сімейства, причому $r \in \{I, 2, \ldots, n-1\}$. У випадку ж $m = 2$ маємо

$$\left.\begin{aligned} I \le p^j_i \le L^j_i, &\quad \text{якщо } i \in \{r_1, r_2\}, \\ p^j_i = L^j_i, &\quad \text{якщо } i \notin \{r_1, r_2\}, \end{aligned}\right\} \qquad (5)$$

де r_1, r_2 - номера тих координатних осей простору Z^{n-1}_j, перпендикулярними до яких являються гіперплощини відповідно першого та другого їх сімейства, причому $r_1, r_2 \in \{I, 2, \ldots, n-1\}$, $r_1 \ne r_2$. При цьому довільний підграф графу G^j_B, народжений В-вершинами, які відповідають будь-якій з вказаних підрешіток, утворюють деякий В-кластер.

5. З метою одержання альтернативної множини структур S, описуючих топологію міжпроцесорних зв'язків синтезованих ОС, для кожного фіксованого способу розбивки, задаваємого або єдиним індексом r при $m = I$, або парою індексів $\{r_1, r_2\}$ при $m = 2$ (де r, r_1, $r_2 \in \{I, 2, \ldots, n-I\}$ і $r_1 \ne r_2$), всі В-кластери графу G^j_B проецюються при $m = I$ на r-ю координатну вісь простору Z^{n-1}_j, в якому розміщена решітка $K^{n-1}_{B,j}$, або при $m = 2$ - на r_1-ю та r_2-ю координатні

осі цього ж простору.

6. Підставляючи задану кількість ПЕ цільової ОС замість p^j_r (при $m = 1$) або p^j_{r1}, p^j_{r2} (при $m = 2$) в співвідношення (4, 5), які виражають кількість ПЕ в кожній з результуючих структур S як функцію розмірів M_l області K^n визначення графу G та знайдених раніше на кроці 3 розмірів q^j_l підрешіток $K^{n-1}_{B,j}$, визначаємо розміри В-кластерів.

7. Для кожного фіксованого способу розбивки графу G^j_B знаходиться множина допустимих макророзгорток, які задають суворо послідовний порядок виконання В-кластерів в часі, і тому існують тільки при умові відсутності зустрічних інформаційних обмінів між кластерами. Для кожної допустимої макророзгортки визначається відповідна їй мікророзгортка, задаюча в часі паралельний порядок виконання В-вершин кожного кластеру на цільовій ОС, а також величини зміщення початку виконання слідуючого кластеру по відношенню до попереднього. При цьому критерієм при знаходженні мікророзгортки та зміщень використовується час реалізації алгоритму з урахуванням обмежень, накладених організацією обмінів з зовнішньою памятю цільової ОС, а також накладними витратами на реконфігурацію та управління.

8. Якщо множина всіх схем, народжених n графами G^j_B та N^m_v способами розбивки кожного з них на В-кластери, а також допустимими часовими розгортками, являється непорожньою, серед схем зазначеноімножини вибираються найбільш придатні з урахуванням заданих вимог до результуючої архітектури СПОП.

У випадку, коли допустимі макророзгортки відсутні взагалі, як альтернативний підхід до синтезу можна рекомендувати застосування винятково ЛПСГП стратегії декомпозиції. Проте при обмеженному обсязі ЛП це пов'язано зі значно зростаючим навантаженням на менш швидкісну зовнішню пам'ять, яка повинна тепер використовуватися як більш швидкісна ЛП ПЕ , що в результаті може привести до простою ПЕ та до падіння продуктивності ОС.

На основі розробленого способу проектування структурних схем СПОП на НВІС фіксованого розміру синтезовано ряд архітектур СПОП на НВІС, реалізуючих набір задач ЛА з використанням алгоритму Фаддєєва. Всі СПОП поряд з однотипністю модулів характеризуються однорідними та локальними міжмодульними зв'язками. При цьому на основі цілеспрямованих перетворень G-графу початкового алгоритму одержано одномірні СПОП фіксованого розміру з мінімальною кількістю каналів вводу/виводу (1/1) та близькою до одиниці завантаженістю.

Запропонована модифікація відомого підходу до побудови багатофункціональних (проблемно-орієнтованих) СПОП на НВІС, яка заключається в цілеспрямованих ізоморфних та/або гомоморфних перетвореннях РФГ вимагаємого набору алгоритмів до виду, близького по топології зв'язків з цільовим (вибраним) РФГ. Лише потім застосовується відомий підхід, який заключається в накладенні РФГ всіх алгоритмів один на одного, одержанні в результаті об'єднаного графу, а потім у синтезі за допомогою відомих методів множини структурних схем СПОП для реалізації вказаних алгоритмів.

На основі модифікованого підходу проведено узагальнення запропонованого способу розбивки РФГ алгоритму Фаддєєва при його реалізації на СПОП фіксованого розміру на інші методи ЛА, зокрема, методи Гаусса, Жордана-Гаусса, Холецького, Хаусхолдера та Гівенса. Показано, що синтезовані в роботі структури СПОП можуть виконувати крім алгоритму Фаддєєва всі перераховані алгоритми. Для цього необхідна лише попередня настройка внутрішньої структури їх ПЕ на виконання заданого алгоритму, що у випадку використання в якості ПЕ універсальних (програмованих) обчислювальних пристроїв, не представляє особливої складності.

Показано, що алгоритми, які мають різні по топології зв'язків базисні РФГ, можуть мати ідентичні або більш близькі В-графи чи В-кластери. Це свідчить про те, що при синтезі СПОП на НВІС фіксованого розміру, реалізуючого набір алгоритмів, описуємих різними базисними РФГ, перед використанням різних перетворень останніх доцільно розглянути всі можливі множини відповідних ІМ В-графів та В-кластерів, і тільки у випадку негативного результату використовувати перетворення їх графів.

ОСНОВНІ РЕЗУЛЬТАТИ РОБОТИ

1. Розроблено метод синтезу часової компоненти відображення РФГ алгоритму розмірності n в структурні схеми СПОП на НВІС розмірності $m=n-2$, який на відміну від відомих дозволяє зменшити трудомісткість цього процесу та повністю формалізувати його.

2. Розроблено спосіб ізоморфного перетворення РФГ трикутної форми, відповідаючого алгоритмам вирішення ряду основних задач ЛА, в систолічний тороїдальний РФГ, що дозволило розробити більш ефективні НВІС-алгоритми вирішення цих задач та за рахунок цього вдвоє скоротити кількість необхідних ПЕ та підвищити завантаженість паралельних ОС при реалізації вказаних задач.

3. Розроблено спосіб проектування структурних схем СПОП на НВІС, який на відміну від відомих дозволяє враховувати фіксовані розміри виділених під задачу ресурсів.

4. На основі цього способу синтезовано ряд одномірних СПОП фіксованого розміру, реалізуючих набір задач ЛА, які відрізняються від відомих мінімальною кількістю каналів вводу/виводу та близькою до повної завантаженістю обладнання.

5. Проведено узагальнення запропонованого способу розбивки РФГ алгоритму Фаддєєва при його реалізації на СПОП фіксованого розміру на інші методи ЛА, в тому числі, методи Гауса, Жордана-Гауса, Холецького, Хаусхолдера та Гівенса. Показано, що синтезовані структури СПОП можуть виконувати, крім алгоритму Фаддєєва, всі зазначені алгоритми, для чого необхідна лише попередня підстройка внутрішньої структури їх ПЕ.

8. Розроблено ряд інших структурних схем СПОП на НВІС, які відрізняються від відомих більш високими технічними параметрами.

РОБОТИ, ОПУБЛІКОВАНІ ПО ТЕМІ ДИСЕРТАЦІЇ

1. А.с. СССР №1520542 МКИ G 06 F 15/347. Устройство для LL^t-разложения симметричных матриц/ Выжиковски Р., Каневский Ю.С., Масленников О.В.// - Б.И.41, 1989г.

2. А.с. СССР №1509933 МКИ G 06 F 15/347. Устройство для LU - разложения матриц/ Каневский Ю.С., Котов С.Э., Масленников О.В.// - Б.И.35, 1989г.

3. Выжиковски Р., Каневский Ю.С., Клименко М.К., Масленников О.В. Устройство для операций над матрицами. - решение о выдаче авт. свид. СССР по заявке №4847777/24 от 27.09.91г.

4. А.с. СССР №1777154 МКИ G 06 F 15/347. Устройство для решения систем линейных алгебраических уравнений. / Выжиковски Р., Каневский Ю.С., Масленников О.В.//- Б.И.43, 1992 г.

5. Выжиковски Р., Каневский Ю.С., Масленников О.В. Устройство для треугольного разложения матриц. - решение о выдаче авт. свид.СССР по заявке № 4774437/24 от 18.06.91г.

6. Выжиковски Р., Каневский Ю.С., Масленников О.В. Устройство для решения систем линейных алгебраических уравнений. - решение о выдаче авт. свид.СССР по заявке № 4878784/24 от 16.01.92г.

7. Выжиковски Р., Каневский Ю.С., Клименко М.К., Масленников О.В. Устройство для решения систем линейных алгебраических уравнений. - решение о выдаче авт. свид.СССР по заявке №

4916907/24 от 15.01.92г.

8. А.с. СССР №1741153 МКИ G 06 F 15/347. Устройство для операций над матрицами./ Выжиковски Р., Каневский Ю.С., Клименко М.К., Масленников О.В.// - Б.И.22, 1992г.

9. А.с. СССР №1735868 МКИ G 06 F 15/347. Устройство для операций над матрицами./ Выжиковски Р., Каневский Ю.С., Масленников О.В.// - Б.И.19, 1992г.

10. Выжиковски Р., Масленников О.В. Синтез систолических процессоров для треугольного разложения матриц различного вида// Конвейерные вычислительные системы: Тез. докл. и сообщ., 2 Всесоюзное совещание - Киев: КПИ. 1988.- С. 79.

11. Масленников О.В., Овраменко С.Г., Синичук И.И. Методика и программные средства автоматизированного проектирования систолических структур // Конвейерные вычислительные системы: Тез. докл. и сообщ., 2 Всесоюзное совещание - Киев: КПИ. 1988.- С. 71.

12. Выжиковски Р., Каневский Ю.С., Масленников О.В. Систолический процессор для треугольного разложения ленточных матриц// Методы и микроэлектронные средства цифрового преобразования и обработки сигналов (SIAP-89): Тез. докл. конф., - Рига: ИЭиВТ, 1989.- т.1., - С.302.

13. Выжиковски Р., Каневский Ю.С., Масленников О.В. Реализация алгоритма исключения Гаусса с частичным выбором на систолических процессорах// Тез. докл. I Всесоюзной конф. "Однородные вычислительные среды и систолические структуры" - Львов. 1990.- т.1.- С.41.

14. Выжиковски Р., Каневский Ю.С., Масленников О.В. Систолический процессор с активным управлением для решения систем линейных алгебраических уравнений// Теория и практика создания систем технического зрения : Матер. семинара МДНТП им.Дзержинского - Москва. 1990. -С. 88.

15. Выжиковски Р., Каневский Ю.С., Масленников О.В. Отображение алгоритма исключения Гаусса на архитектуру систолических массивов. //Электронное моделирование.-1991.-№2.-с.14-18.

16. Выжиковски Р., Каневский Ю.С., Масленников О.В. Решение одного класса задач линейной алгебры на линейных систолических вычислителях.//Электронное моделирование.-1993.-№4, с. 26-33.

Автор

НВЛ ПФ КПИ 1993. Зак.

ИВЛ ПФ КПІ 1993. Зам. 650 — 100