Вестник Приднестровского университета.
Серия: Физико-математические и технические науки.
№ 1-2 (12), 2000. - С. 99-110.

В С Выхованец

КРАТНЫЕ ВЫЧИСЛЕНИЯ В СОВРЕМЕННЫХ ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ СИСТЕМАХ ОБРАБОТКИ ДАННЫХ

Показано, что известные методы структурной организации вычислительного процесса (повторение, распараллеливание и конвейеризация) могут быть дополнены новым способом - параллельными вычислениями во времени. Этот

метод выявлен в результате анализа их комплексной эффективности, учитыва ющей как временные затраты, так и сложность требуемых аппаратных средств. Возникающие при этом вычисления названы кратными, т. е. выполняемыми сразу на нескольких наборах данных. Приведены примеры реализации парал лельных вычислений во времени на основе параллельной декомпозиции сис темы логических функций, осуществляемой при ее представлении в полино миальной форме в базисе поразрядных логических операций.

Введение

В настоящее время возможности дальнейшего наращивания производительности вычислительной техники в рамках последовательного принципа обработки данных считаются практически исчерпанными, что обусловлено в основном конечной скоростью распространения сигналов [1]. Поиск решений проблемы повышения производительности идет в направлении развития как нетрадиционных принципов обработки данных [2], так и методов структурной организации вычислений: распараллеливания [3] и конвейеризации [4]. Последние предназначены для выполнения одних и тех же вычислений на нескольких наборах исходных данных.

Под кратными вычислениями будем понимать процесс обработки не на одном, а на множестве наборов данных. Кратные вычисления в виде повторного вызова подпрограмм и распараллеливания вычислительного процесса возникают на программном уровне, в виде конвейерного и параллельного выполнения команд- на командном уровне, в виде распараллеливания и конвейеризации операций - на операционном уровне.

Наименее изученными остаются вопросы исследования комплексной эффективности кратных вычислений на разных уровнях конвейерных и параллельных вычислительных средств, учитывающих как временные затраты, так и сложность используемых аппаратных средств.

Критерии эффективности

Для оценки эффективности кратных вычислений введем следующие критерии: эффективность по времени £ эффективность по сложности *J]* и комплексную эффективность *3,* равную произведению эффективностей по времени £ и сложности *г*. Эти эффективности будем вычислять в сравнении с однократным случаем, когда обработка осуществляется на одном наборе данных. Время и сложность однократных вычислений принимаем за единицы измерения времени и сложности.

Эффективность по времени £ определяет во сколько раз время т-кратных вычислений m_m для исследуемого метода меньше, чем время повторения вычислений *m* раз в однократном случае:

$$\xi = \frac{m \cdot 1}{\tau_m} = \frac{m}{\tau_m}.$$

Эффективность по сложности *n* определяет во сколько раз аппаратные затраты v_m (объем памяти при программной реализации) меньше аналогичных затрат при однократных вычислениях:

$$\eta = \frac{1}{v_m}.$$

Комплексная эффективность характеризует эффективность вычислений с учетом эффективностей по времени и сложности:

$$\vartheta = \xi\eta = \frac{m}{\tau_m v_m}.$$

Методы структурной организации кратных вычислений

В основу классификации методов кратных вычислений положим следующие принципы повторности: параллельная или последовательная повторность, повторность в пространстве или во времени. На основе введенных принципов можно выделить параллельные и последовательные вычисления, осуществляемые как в пространстве, так и во времени (рис. 1).

Последовательные вычисления во времени (повторная обработка) представляют собой такую организацию вычислений, при которой различные наборы данных $X^{(r)}$ $(r = 1, m)$ подаются на вход вычислительного средства в различные моменты времени *t,* отстоящие друг от друга на время, достаточное для однократного вычисления, т. е. последовательность входных наборов перемен ных.

$$\{X^{(1)}(t), X^{(2)}(t+1), \ldots, X^{(m)}(t+m-1)\}$$

порождает соответствующую последовательность результатов:

$$\{Y^{(1)}(t+1), Y^{(2)}(t+2), \ldots, Y^{(m)}(t+m)\}.$$

В итоге эффективность повторной обработки определим так:

$$\xi_1 = \frac{m}{m} = 1,\ \eta_1 = \frac{1}{1} = 1,\ \vartheta_1 = \xi_1\eta_1 = 1.$$

При последовательных вычислениях в пространстве (конвейеризация или многостадийная обработка) различные наборы переменных $X^{(r)}$ подаются на вход также в разные моменты времени, но отстоящие друг от друга на время

Δt, которого достаточно для вычислений, осуществляемых самой медленной ступенью конвейера [1, 4]. Первый набор переменных $X^{(1)}$ поступает на первую ступень конвейера в некоторое время t. Через время Δt ступень завершает свою работу, и на ее вход может быть подан второй набор $X^{(2)}$. Результат вычисления на первой ступени $Y_1^{(1)}(t + \Delta t)$ подается на вход второй ступени и т. д. Первый итоговый результат появляется на выходе через время $d \bullet \Delta t$, равное суммарному времени работы всех d ступеней; каждый последующий – через время Δt. Таким образом, последовательность входных данных

$$\{X^{(1)}(t), X^{(2)}(t + \Delta t), \ldots, X^{(m)}(t + (m - 1)\Delta t)\}$$

порождает соответствующую последовательность результатов

$$\{Y^{(1)}(t + d\Delta t), Y^{(2)}(t + (d + 1)\Delta t), \ldots, Y^{(m)}(t + (d + m - 1)\Delta t)\}.$$

В технических системах получили распространение такие реализации, в которых каждая ступень вычисляет результат за время, меньшее в d раз времени однократного вычисления, т. е,, $\Delta t \approx 1/d$, а общая сложность вычислений незначительно превосходит сложность в однократном случае, т. е. $v_m \approx 1$. С учетом этого получаем:

$$\xi_2 = \frac{m}{1+(m-1)\Delta t} \underset{(m>>d)}{\approx} \frac{d \cdot m}{(m-1)},\ \eta_2 \approx 1,\ \vartheta_2 = \xi_2 \eta_2 \approx \frac{d \cdot m}{(m-1)}.$$

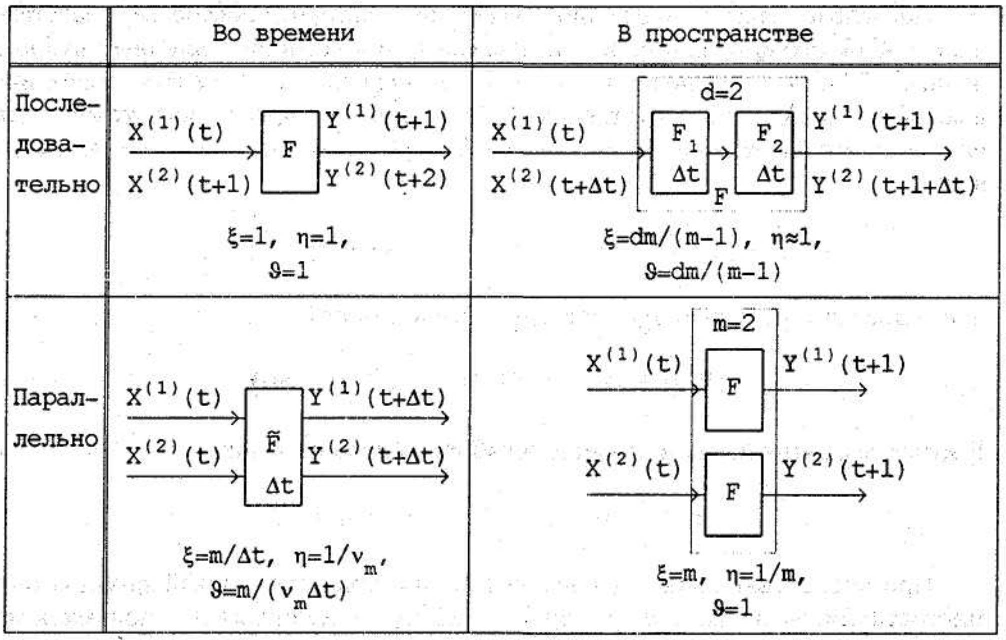

Рис. 1. Методы структурной организации кратных вычислений

Как видно из последнего выражения, при достаточно большом m комплексная эффективность *9* конвейерного метода стремится к d - количеству ступеней конвейера.

Параллельные вычисления в пространстве (распараллеливание или многоэлементная обработка) представляют такую организацию вычислительного процесса, когда имеется несколько вычислительных средств, каждое из которых выполняет вычисление на своем наборе данных [3, 5]. В этом случае результаты появляются на выходе через время однократного вычисления, но при этом требуется дублирование оборудования, равное кратности m. Таким образом, выходные наборы данных

$$\{X^{(1)}(t), X^{(2)}(t), \ldots, X^{(m)}(t)\}$$

подаются одновременно, одновременно появляются на выходе и соответствующие результаты:

$$\{Y^{(1)}(t+1), Y^{(2)}(t+1), \ldots, Y^{(m)}(t+1)\}.$$

Эффективности параллельных в пространстве вычислений равны:

$$\xi_3 = m,\ \eta_3 = \frac{1}{m},\ \vartheta_3 = \xi_3\eta_3 = 1.$$

И, наконец, параллельные вычисления во времени соответствуют такой организации вычислений, при которой на вход вычислительного средства одновременно подаются все m наборов данных

$$\{X^{(1)}(t), X^{(2)}(t), \ldots, X^{(m)}(t)\}$$

и через время At одновременно получаются все m результатов

$$\{Y^{(1)}(t+\Delta t), Y^{(2)}(t+\Delta t), \ldots, Y^{(m)}(t+\Delta t)\}.$$

Очевидно, чтобы отличить параллельные вычисления во времени от параллельных вычислений в пространстве (по аналогии с последовательными вычислениями во времени и в пространстве), комплексный критерий эффективности $\vartheta_4 = m/(\Delta t\ v_m)$ должен быть строго больше единицы, т..е. $\vartheta_4 > 1$.

Таким образом, к параллельным вычислениям во времени могут быть отнесены такие методы, при реализации которых комплексный критерий эффективности будет больше единицы и исходные данные (результаты) подаются (снимаются) одновременно. Заметим, чю комплексная эффективность больше единицы только для последовательных в пространстве и параллельных во времени вычислений, а для остальных методов она равна единице.

Реализация кратных вычислений

Проанализируем основные разновидности современных высокопроизводительных систем обработки данных по эффективности кратных вычислений.

Последовательные вычисления во времени - основной метод структурной организации кратных вычислений в системах обработки с фон-неймановской архитектурой [6]. По такому принципу реализуются основные обрабатывающие устройства (процессорные элементы) на операционном уровне во всех известных системах: матричных, систолических, ассоциативных, транспьютерных, мультипроцессорных, волновых и т. д.

В качестве примера систем с последовательными вычислениями в пространстве можно привести векторно-конвейерные, систолические и волновые системы, нейронные сети (командный уровень), а также процессорные элементы (процессоры) с конвейеризацией команд (командный уровень) и/или конвейеризацией операций (операционный уровень).

Параллельные вычисления в пространстве реализуются в мультипроцессорных системах на программном уровне; в векторно-параллельных (матричных), ассоциативных, потоковых, систолических и волновых системах и нейронных сетях, в процессорных элементах (процессорах) со сверхдлинным командным словом, суперскалярной обработкой данных на командном уровне; в операционных устройствах с поразрядными операциями на операционном уровне.

Анализ современных систем обработки данных показал, что основными методами структурной организации кратных вычислений являются последовательные вычисления во времени, последовательные вычисления в пространстве и параллельные вычисления в пространстве. Исследование возможности реализации параллельных вычислений во времени остается актуальным.

Параллельные вычисления во времени

Основное структурное свойство параллельных вычислений во времени - одновременная подача всех наборов переменных и одновременное получение результата. Комплексная эффективность таких вычислений должна быть больше единицы, что можно обеспечить более быстрым выполнением кратных, чем последовательных вычислений во времени, или с меньшими аппаратными затратами по сравнению с параллельными вычислениями в пространстве.

Последний случай поясним на примере многомодовой оптической обработки (рис. 2), когда по некоторому каналу распространяется совокупность сигналов №\ имеющих разную поляризацию, что обеспечивает отсутствие их существенного влияния друг на друга. В среду обработки сигналы вводятся с помощью пространственного мультиплексора MUX, а после обработки они разделяются пространственным демультиплексором DMX. В среде обработки все сигналы подвергаются моделирующему воздействию со стороны обрабатывающего элемента, являющегося частью среды распространения (показан штриховой

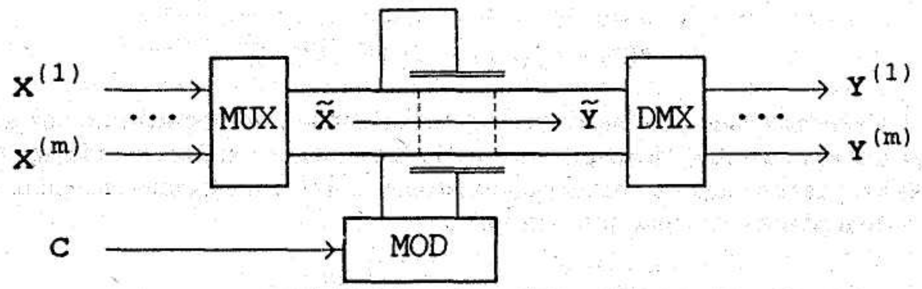

Рис. 2. Многомодовая оптическая обработка

линией). Моделирующий сигнал выдается пространственным модулятором MOD, на вход которого подается обрабатывающий сигнал *C,* задающий закон (алгоритм) преобразования.

В настоящее время наибольшее распространение получили вычислительные устройства дискретного действия. Техника дискретной обработки достигла такого уровня развития, который позволяет решать многие практические задачи, а развитые в последние годы методы дискретного анализа дают возможность получать численный результат с заданной точностью с помощью выполнения некоторой конечной последовательности операций над дискретными данными.

Математическим аппаратом, используемым при проектировании дискретных устройств, является аппарат алгебры логики и теории автоматов, а применение таких устройств в конечном итоге вынуждает сводить решаемую задачу к логической обработке данных. Логическая обработка формализуется в алгебре логики и представляется в виде системы логических функций

$$F(X) = (f_q(X), f_{q-1}(X), \ldots, f_1(X)),$$

заданных на общем множестве переменных

$$X = (X_n, X_{n-1}, \ldots, X_1).$$

При вычислений этой системы реализуется требуемая обработка данных.

Основная задача реализации параллельных во времени вычислений — нахождение таких форм совместного описания кратных вычислений, в которых выделяются некоторые общие выражения, используемые для вычислений на каждом наборе. Такую декомпозицию назовем параллельной в отличие от последовательной декомпозиции, необходимой при конвейерной обработке.

Для конвейерных вычислений система функций разбивается на d более простых подсистем F_t $\{t = 1, d)$ таким образом, чтобы результаты вычисления предыдущей подсистемы F_t являлись аргументами для вычисления следующей подсистемы F_{t+1}:

$$F(X) = F_d\,(F_{d-1}\,(\ldots F_2\,(F_1\,(X))\ldots)).$$

Следовательно, для вычислений, параллельных во времени, необходимо представить систему функций так, чтобы можно было выделить общие выражения, участвующие'в последующих вычислениях для каждого значения системы на разных наборах аргументов.

Параллельная декомпозиция

Пусть требуется вычислить систему функций на m наборах аргументов $X^{(r)}$ $(r = 1,m)$ Представим вычисляемую систему на всем множестве наборов, что эквивалентно системе функций

$$\widetilde{F}(\widetilde{X}) = (\widetilde{f}_q(\widetilde{X}), \widetilde{f}_{q-1}(\widetilde{X}), \ldots, \widetilde{f}_1(\widetilde{X})),$$

от nm переменных:

$$\widetilde{X} = \left(\widetilde{X}_n, \widetilde{X}_{n-1}, \ldots, \widetilde{X}_1\right),$$

$$\widetilde{X}_i = \left(X_i^{(m)}, X_i^{(m-1)}, \ldots, X_i^{(1)}\right)\left(i = \overline{1, n}\right),$$

где $f_i\,(F)$ – форма описания кратных вычислений j-и функции (системы функций),

В работах [7, 8] показано, что кратные вычисления могут быть представлены в полиномиальной форме и реализованы путем вычисления арифметико-логического полинома в базисе поразрядных операций:

$$\widetilde{F}\left(\widetilde{X}\right) = \sum_{i=0}^{k^n-1} A_i \cdot \left(C_i \circ \widetilde{X}_n^{i_n} \circ \ldots \circ \widetilde{X}_n^{i_1}\right) = \sum_{i=0}^{k^n-1} A_i \theta_i\left(\widetilde{X}\right),$$

где $к$ — значность алгебры логики; A_i - i-й коэффициент полинома; C_i- - некоторые константы; o - поразрядные логические операции (возможно, различные); X_i^j - значения переменных $X_i^{(r)}$ в логической степени j (степенные операции также могут различаться); $i = (i_n\ \ldots\ i_2\ i_1)_k$ - k-ичное представление числа i; $\theta_i(X)$ – логическая часть полинома. Обобщенная методика синтеза полиномиальных форм основана на дискретном ортогональном преобразовании и подробно рассмотрена в работе [9].

Для примера представим систему булевых функций

$$\begin{cases} f_1(X) = (X_3 \,\&\, X_2) \vee X_1, \\ f_2(X) = (X_3 \,\&\, X_2) \otimes X_1 \end{cases}$$

в полиномиальной форме для двукратных вычислений

$$\tilde{Y} = \tilde{F}(\tilde{X}) = 5\tilde{X}_1 + 5(\tilde{X}_3 \,\&\, X_2) - 9\cdot(\tilde{X}_3 \,\&\, \tilde{X}_2 \,\&\, \tilde{X}_1)$$

и найдем этот полином на наборах $X^{(1)} = (111)$. $X^{(2)} = (110)$. Вычислив подстановочные булевые векторы: $X_3 =(11)$, $X_2=(11)$, $X_1 = (01)$, получим

$$\tilde{F}(\tilde{X}) = 5\cdot(01) + 5\cdot((11)\&(11)) - 9\cdot((11)\&(11)\&(01)) = (0010),$$

а результат интерпретируем следующим образом:

$$\tilde{Y} = \left(Y_2^{(1)} \middle| Y_2^{(0)} \middle| Y_1^{(1)} \middle| Y_1^{(0)}\right).$$

Итак, общим выражением для всех наборов переменных в полиномиаль ной форме является ее арифметическая часть, зависящая только, от исходной системы функций.

Реализация параллельных вычислений во времени -

Как показано выше, общим оборудованием при параллельных вычислениях вр времени могут быть блоки, хранящие, коэффициенты формы и выполняющие арифметические операции. Структура такого устройства показана на рис. 3, где для двух параллельных каналов выделено общее оборудование: блок хранения коэффициентов *A* и сумматор а, Вычисление логической части полинома ***в*** выполняется, параллельно в пространстве путем использования поразрядных операций. Данные С, как и ранее, определяют алгоритм обработки, а коэффициенты полинома выступают в роли исполняемой программы.

Дадим расширенную интерпретацию полученных результатов. К параллельным вычислениям во времени отнесем все реализации, у которых имеется несколько параллельных каналов с общим оборудованием. Примером в этом случае может служить процессор Pentium фирмы Intel, у которого для нескольких каналов обработки данных имеется общая кэш-память команд и кэш-память данных.

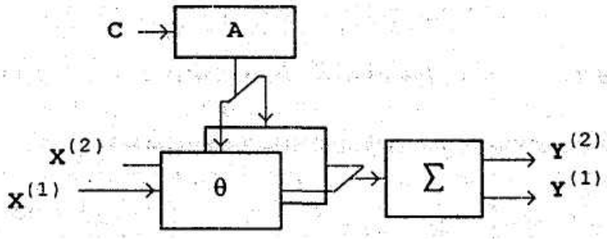

Рис. 3. Параллельные вычисления во времени

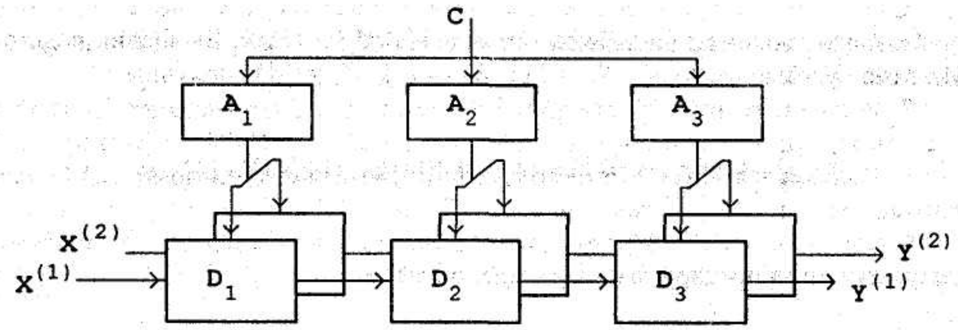

Рис. 4. Структура конвейерно-параллельного процессора

В заключение приведем структуру конвейерно-параллельного процессора, представленную на рис. 4, которая реализует кратные вычисления с наибольшей эффективностью, равной произведению эффективностей конвейерной и параллельной во времени обработки данных.

Выводы

Анализ известных методов структурной организации кратных вычислений показал, что параллельные вычисления могут быть разделены на параллельные вычисления в пространстве и параллельные вычисления во времени, которые отличаются комплексной эффективностью. В первом случае комплексная эффективность равна единице, во втором-строго больше единицы.

Для представления параллельных вычислений во времени необходимо выполнить параллельную декомпозицию вычисляемой функции, которая заключается в выделении общих для всех наборов данных выражений в формуле. При логических вычислениях параллельная декомпозиция выполняется путем представления системы функций в полиномиальной форме в базисе поразрядных операций.

Параллельные вычисления во времени позволяют более интенсивно использовать вычислительные средства и реализовать кратные (повторяющиеся) вычисления с наибольшей эффективностью.

Цитированная литература

1. **Самофалов К.Г., Луцкий Г.М.** Основы теории многоуровневых конвейерных вычислительных систем. М.: Радио и связь, 1989.

2. **Водяхо А.И., Горнец Н.Н., Пузанков** Д. **В.** Высокопроизводительные системы обработки данных. М.: Высшая школа, 1997.

3. Системы параллельной обработки / Под ред. Д. Ивенса. М.: Мир, 1985.

4. **Кухарчук А.Г., Луцкий Г.М.** Конвейерный принцип обработки информации//Кибернетика. 1968. № 6. С. 43-49.

5. **Головкин Б.А.** Параллельные вычислительные системы. М.: Наука, 1980.

6. **Амамия М., Танака Ю.** Архитектура ЭВМ и искусственный интеллект. М.: Мир, 1993.

7. **Выхованец В.С.** Многократные параллельные логические вычисления //Вестник Приднестровского ун-та. 1997.№2 (7). С. 64-74.

8. **Выхованец В.С., Малюгин В.Д.** Кратные логические вычисления // Автоматика и телемеханика. 1998. №6. С. 163-171.

9. **Выхованец В.С.** Параллельные вычисления во времени *II* Автоматика и телемеханика. 1999. № 12. С. 155-165.