

РАЗРАБОТКА ПАРАЛЛЕЛЬНЫХ РЕАЛИЗАЦИЙ АЛГОРИТМОВ ПАРАМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ ТРЕХМЕРНЫХ КРИСТАЛЛИЧЕСКИХ РЕШЁТОК

Д.В. Кириш, А.В. Куприянов

Институт систем обработки изображений РАН, Самара, Россия,
Самарский государственный аэрокосмический университет имени академика С.П. Королёва (национальный исследовательский университет) (СГАУ), Самара, Россия

В статье описываются параллельные реализации двух алгоритмов параметрической идентификации: на основе оценивания параметров элементарной ячейки Браве и на основе оценивания объёмов ячеек Вигнера-Зейтца. Разработанные параллельные реализации базируются на двухуровневой модели параллелизма. Внешний уровень параллелизма использует технологию MPI для распределения задач параметрической идентификации между вычислительными узлами. Внутренний уровень параллелизма использует технологию OpenMP для разбиения отдельных шагов алгоритмов параметрической идентификации на независимые подзадачи. Результаты исследований на многопроцессорных/многоядерных системах показали достижение почти линейного ускорения и подтвердили эффективность применённой двухуровневой модели параллелизма.

Ключевые слова: кристаллическая решётка, элементарная ячейка, параметрическая идентификация, параллельный алгоритм, MPI, OpenMP, масштабируемость.

Введение

Структурная идентификация кристаллических решёток в трёхмерном пространстве является одной из главных задач рентген-дифракционного анализа. Основной проблемой на пути её решения оказывается неоднозначность выбора элементарной ячейки, а следовательно, и неоднозначность определения типа кристаллической системы [1].

Для решения данной проблемы было предложено два подхода: на основе методов параметрической идентификации [2] и на основе применения нечётких нейронных сетей [3]. Наиболее универсальным оказался первый из предложенных подходов. Согласно ему, в начале, у исследуемой кристаллической решётки оценивается ряд параметров: стороны и углы элементарной ячейки Браве [3] и объём ячейки Вигнера-Зейтца [4]. Затем вычисляются меры схожести оценённых параметров с набором параметров эталонных решёток. Однако проведённые вычислительные эксперименты показали, что для достижения высокой точности структурной идентификации (более 95%) требуется огромная база эталонных решёток [5].

Таким образом, обратной стороной эффективности предложенного подхода стала его высокая вычислительная сложность. Для преодоления указанного недостатка было решено разработать параллельные реализации алгоритмов параметрической идентификации, позволяющие задействовать вычислительные мощности современных многопроцессорных/многоядерных систем.

Главным требованием при создании необходимых параллельных реализаций стала эффективность масштабирования алгоритма. Таким образом, нужно было обеспечить большое количество независимых вычислительных задач, которые могли бы выполняться параллельно. Для решения проблемы масштабирования была предложена двухуровневая модель параллелизма, описываемая в настоящей статье.

1. Внешний уровень параллельной реализации алгоритмов параметрической идентификации

Разработанные методы параметрической идентификации оценивают и сравнивают параметры элементарных ячеек, с помощью которых была образована вся кристаллическая решётка. Очевидно, что рассматриваемые методы обладают высокой степенью локальностью: для их работы необходимо знание координат лишь нескольких, наиболее близко расположенных, узлов (рис. 1).

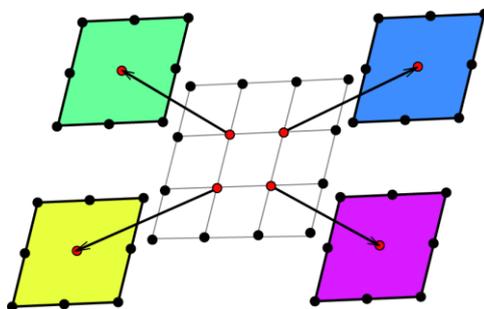


Рис.1. Пример разделения двумерной решётки на области элементарных ячеек (как Браве, так и Вигнера-Зейтца)

Кроме того, вычисление параметров для каждой элементарной ячейки может происходить независимо друг от друга. Необходимость в обмене данными возникает только на заключительном этапе, когда происходит анализ параметров всех элементарных ячеек и вычисление итоговых параметров для всей решётки.

Таким образом, на уровне выделения задач оценивания параметров элементарных ячеек для всех методов параметрической идентификации был предложен единый алгоритм параллельного выполнения (внешний уровень параллелизма):

1. Проверка принадлежности узла грани решётки – выполняется одним из вычислительных процессов.
2. Распределение элементарных ячеек – координаты узлов, попавших внутрь шара заданного радиуса с центром в негранеобразующем узле, равномерно распределяются по вычислительным процессам.
3. Вычисление параметров – для каждой элементарной ячейки оцениваются её параметры, вычисления выполняется всеми процессами параллельно.
4. Сбор результатов – оценивание итоговых параметров для всей решётки, выполняется одним из вычислительных процессов.

Очевидно, что данный алгоритм параллельного выполнения в наибольшей степени соответствует модели MPI [6]: локальные вычисления, коммуникации между задачами минимальны и присутствуют только на начальном и заключительном этапах.

Выбранный подход позволяет избежать несбалансированность загрузки вычислительных процессов (число задач у разных вычислительных процессов не будет отличаться больше, чем на одну задачу).

2. Внутренний уровень параллельной реализации алгоритмов параметрической идентификации

Однако при использовании большого количества многоядерных процессоров число независимых задач вычисления параметров элементарных ячеек может оказаться недостаточным для загрузки всех доступных вычислительных ядер. В связи с невозможностью выделения дополнительных задач на внешнем уровне параллелизма решено было для каждого метода идентификации организовать дополнительный внутренний уровень параллелизма за счёт параллельной реализации самих задач вычисления параметров.

2.1 Параллельная реализация алгоритма оценивания параметров элементарной ячейки Браве

Алгоритм оценивания параметров элементарной ячейки Браве представляет собой последовательность шагов, которые должны выполняться строго по порядку [3]. Следовательно, организация параллелизма возможна только внутри самих шагов. На рис. 2 представлена схема предлагаемой параллельной реализации для данного алгоритма.



Рис. 2. Схема параллельной реализации алгоритма оценивания параметров элементарной ячейки Браве

Вычислительная сложность каждого этапа алгоритма для решётки, содержащей N узлов, составляет: 1) $O(N)$; 2) $O(N)$; 3) $O(N)$; 4) $O(N)$; 5) $O(N)$; 6) $O(1)$.

В связи с тем, что наибольшей вычислительной сложностью обладают первые 5 этапов, именно внутри них и следует организовывать параллельное выполнение.

2.2 Параллельная реализация алгоритма оценивания объёма ячейки Вигнера-Зейтца

Алгоритм оценивания объёма ячейки Вигнера-Зейтца также представляет собой последовательность шагов, которые должны выполняться строго по порядку [4]. Как и в предыдущем случае, организация параллелизма возможна только внутри самих шагов. На рис. 3 представлена схема предлагаемой параллельной реализации.



Рис. 3. Схема параллельной реализации алгоритма оценивания объёма ячейки Вигнера-Зейтца

Вычислительная сложность каждого этапа алгоритма для решётки, содержащей N узлов, при вбрасывании L точек составляет: 1) $O(N)$; 2) $O(L)$; 3) $O(L \cdot N)$; 4) $O(1)$.

В связи с тем, что вычислительная сложность второго этапа алгоритма существенно выше вычислительной сложности первого этапа (минимальное число вбрасываемых точек, обеспечивающих достаточный уровень точности вычисления объёма, составляет 10 000), именно внутри него и следует организовывать параллельное выполнение, оставив первый этап последовательным.

Общей чертой выполненных параллельных реализаций алгоритмов оценивания параметров элементарных ячеек является их зависимость по данным и большое число коммуникаций на каждом этапе. Таким образом, наиболее подходящей моделью для реализации параллелизма является модель OpenMP [6].

3. Исследование ускорения разработанных параллельных реализаций

Целью данного исследования являлось экспериментальное определение зависимости ускорения разработанных параллельных реализаций от количества вычислительных узлов/ядер и размера исследуемой решётки.

Ускорение, получаемое при использовании параллельной реализации для p вычислительных ядер, по сравнению с последовательным вариантом выполнения определялось следующей величиной [7]:

$$S_p = \frac{T_1}{T_p}, \quad (1)$$

где T_1 – время выполнения на 1 ядре;

T_p – время выполнения на p ядрах.

Согласно формуле (1), наилучшее значение ускорения $S_p = p$ (линейное ускорение), то есть время выполнения параллельной реализации линейно уменьшается с ростом числа вычислительных ядер.

В качестве исследуемой использовались две триклинные решётки: малая (м) и большая (б). Размеры решёток подбирались таким образом, чтобы обеспечить время вычисления параметров для малой решётки – более 1 сек., а для большой – в районе 5 мин. на одном вычислительном ядре. Данный выбор позволяет как оценить ускорение при малом количестве параллельных задач, когда главную роль играют накладные расходы при организации параллельных процессов, так и спрогнозировать изменение ускорения с ростом размера решётки и, соответственно, количества и сложности задач.

3.1. Исследование ускорения на многоядерной персональной ЭВМ

Данное исследование проводилось на персональной ЭВМ на базе 4-ядерного процессора Intel Core i5 3470 (частота ядер 3.8 GHz).

Размеры решёток для каждого метода параметрической идентификации составили:

а) Метод оценивания параметров элементарных ячеек Браве – 24 и 51 трансляций по каждому направлению;

б) Метод оценивания объёмов ячеек Вигнера-Зейтца – 8 и 17 трансляций по каждому направлению.

По полученным данным была сформирована диаграмма, отображающую величины ускорения параллельных реализаций для разного числа вычислительных ядер (рис. 4.).

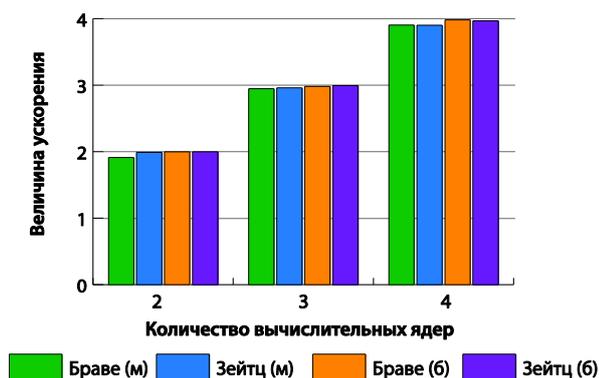


Рис. 4. Ускорение на многоядерной персональной ЭВМ

Полученные результаты подтверждают, что параллельные реализации обоих методов параметрической идентификации обеспечивают почти линейное ускорение на всём диапазоне используемых размеров решёток. Различие в полученных значениях ускорений для малых и больших решёток составляет менее 5 %.

3.2. Исследование ускорения на многопроцессорной/многоядерной кластерной системе

Данное исследование проводилось на 16 узлах кластера «Сергей Королёв», содержащих по два 4-ядерных процессора Intel Xeon X5560 (частота ядер 2.80 GHz).

Размеры решёток для каждого метода параметрической идентификации составили:

- а) Метод оценивания параметров элементарных ячеек Браве – 19 и 37 трансляций по каждому направлению;
- б) Метод оценивания объёмов ячеек Вигнера-Зейтца – 4 и 13 трансляций по каждому направлению.

Следует заметить, что решётка, состоящая из 4 трансляций, не сможет обеспечить достаточного числа задач вычисления параметров, необходимого для равномерной загрузки всех ядер каждого вычислительного узла кластера (27 параллельных задач). В данном случае важную роль будет играть внутренний уровень параллелизма, который должен сбалансировать нагрузку на вычислительные ядра.

По полученным данным была сформирована диаграмма, отображающую величины ускорения параллельных реализаций для разного числа вычислительных ядер (рис. 5.).

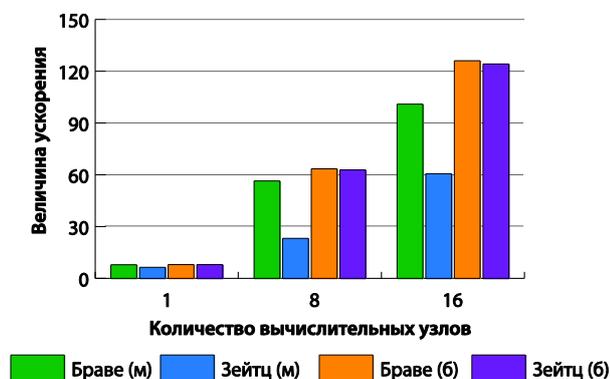


Рис. 5. Ускорение на многопроцессорной/многоядерной кластерной системе

Из представленной диаграммы видно, что параллельные реализации обеспечивают почти линейное ускорение на одном 8-ядерном узле кластера, тем самым косвенно подтверждая результаты исследования из пункта 3.1. Наихудший результат для решётки малого размера ожидаемо продемонстрировала параллельная реализация алгоритма оценивания объёма ячейки Вигнера-Зейтца. Однако даже в случае решётки малого размера параллельная реализация обеспечивает ускорение, существенно большее, чем число используемых узлов кластера, что подтверждает эффективность применения двухуровневой модели параллелизма.

Заключение

Разработанные параллельные реализации алгоритмов параметрической идентификации кристаллических решёток в трёхмерном пространстве эффективно объединяют в себе два базовых подхода к параллельному программированию: MPI и OpenMP.

Проведённые исследования показали, что использование двухуровневой модели параллелизма позволяет максимально полно использовать вычислительные мощности многопроцессорных/многоядерных систем.

Благодарности

Работа выполнена при частичной поддержке Министерства образования и науки РФ в рамках реализации мероприятий Программы повышения конкурентоспособности СГАУ среди ведущих мировых научно-образовательных центров на 2013–2020 годы; грантов РФФИ 14-07-97040-р_поволжье_a, 15-29-03823, 15-29-07077; программы № 6 фундаментальных исследований ОНИТ РАН «Биоинформатика, современные информационные технологии и математические методы в медицине» 2016 г.

Литература

1. Tilley, R. Crystals and crystal structure / R. Tilley. – West Sussex: John Wiley & Sons Ltd, 2006. – P. 17-32.
2. Куприянов, А. В. Оценка меры схожести кристаллических решёток по координатам их узлов в трёхмерном пространстве / А. В. Куприянов, Д. В. Кириш // Компьютерная оптика. – 2012. – Т. 36, № 4. – С. 590-595.
3. Солдатова, О. П. Применение нечётких нейронных сетей для определения типа кристаллических решёток, наблюдаемых на наномасштабных изображениях / О. П. Солдатова, И. А. Лёзин, И. В. Лёзина, А. В. Куприянов, Д. В. Кириш // Компьютерная оптика. – 2015. – Т. 39, № 5. – С. 787-795.
4. Кириш, Д. В. Идентификация кристаллических решёток на основе оценивания параметров элементарных ячеек / Д. В. Кириш // Вестник СГАУ. – 2014. – Т. 45, № 3. – С. 130-137
5. Kirsh, D. V. Crystal lattice identification by coordinates of their nodes in three dimensional space / D. V. Kirsh, A. V. Kupriyanov // Pattern recognition and image analysis. – 2015. – Vol. 25, № 3. – P. 456-460.
6. Антонов, А. С. Технологии параллельного программирования MPI и OpenMP // Издательство МГУ – М., 2012, 344 с.
7. Гергель, В. П. Теория и практика параллельных вычислений: учеб. пособие / В. П. Гергель. – М.: Интернет-университет информационных технологий; БИНОМ. Лаборатория знаний, 2007. – 423 с.