

УДК 004.021

doi 10.26089/NumMet.v16r448

ПАКЕТ ПАРАЛЛЕЛЬНОЙ ДЕКОМПОЗИЦИИ БОЛЬШИХ СЕТОК GRIDSPIDERPAR**Е. Н. Головченко¹, М. В. Якововский²**

Задача рациональной декомпозиции расчетных сеток возникает при численном моделировании на высокопроизводительных вычислительных системах проблем механики сплошных сред, импульсной энергетики, электродинамики и др. Число процессоров, на котором будет считаться вычислительная задача, как правило, заранее не известно. В этой связи имеет смысл предварительно однократно разбить сетку на большое число микродоменов, а затем формировать из них домены. Методы разбиения графов параллельных пакетов ParMETIS, Jostle, PT-Scotch и Zoltan основываются на иерархических алгоритмах, недостатком которых является образование несвязных доменов. Другим недостатком указанных пакетов является получение сильно несбалансированных разбиений. Разработан пакет программ GridSpiderPar для параллельной декомпозиции больших сеток. Проведены вычислительные эксперименты по сравнению различных разбиений на микродомены, разбиений графов микродоменов на домены, а также разбиений сразу на домены нескольких сеток (10^8 вершин, 10^9 элементов), полученных методами созданного комплекса программ GridSpiderPar и пакетов ParMETIS, Zoltan и PT-Scotch. Качество разбиений проверялось по дисбалансу числа вершин в доменах, числу несвязных доменов и числу разрезанных ребер, а также по эффективности параллельного счета задач газовой динамики при распределении сеток по ядрам в соответствии с различными разбиениями. Полученные результаты выявили преимущества разработанных алгоритмов.

Ключевые слова: параллельные вычисления, разбиение графов, декомпозиция сеток.

1. Введение. Задача рациональной декомпозиции расчетных сеток возникает при численном моделировании на высокопроизводительных вычислительных системах проблем механики сплошных сред, импульсной энергетики, электродинамики и многих других. При распараллеливании подобных вычислительных приложений используется метод геометрического параллелизма, при котором сетка, аппроксимирующая расчетную область, распределяется между процессорами по геометрическому признаку. В ходе расчета каждый процессор обрабатывает свою часть сетки. Эффективность работы многопроцессорной вычислительной системы определяется тем, насколько равномерно распределена сетка по процессорам и насколько минимизированы затраты на передачу данных между процессорами. Объем передаваемых между процессорами данных зависит от числа связей между распределенными по процессорам доменами (частями сеток).

Задача сбалансированного разбиения сетки на домены сводится к более общей задаче разбиения графа на домены. В этом случае выполняется разбиение графа, аппроксимирующего вычислительные и коммуникационные нагрузки сетки. Сетка аппроксимируется неориентированным графом $G = (V, E)$, где V — множество вершин, E — множество ребер. И вершины, и ребра имеют вес. Оптимальным считается разбиение на домены, при котором выровнен суммарный вес вершин в доменах и минимизирован суммарный вес разрезанных ребер (разрезанное ребро — это ребро, соединяющее вершины из разных доменов). Поставленная задача декомпозиции графа является NP-полной, поэтому для ее решения используются различные эвристические методы.

Число процессоров, на котором будет считаться вычислительная задача, как правило, заранее не известно. Поэтому имеет смысл предварительно однократно разбить сетку на большое число микродоменов, а потом формировать из них домены. Количество микродоменов на несколько порядков меньше числа вершин, поэтому многократное разбиение микродоменов на домены быстрее многократного разбиения всей сетки.

Широко известны методы декомпозиции областей, используемые для решения линейных и нелинейных систем уравнений, получающихся в результате дискретизации дифференциальных уравнений с част-

¹ Институт прикладной математики им. М. В. Келдыша РАН (ИПМ РАН), Миусская пл., 4А, 125047, Москва; науч. сотр., e-mail: ge03@imamod.ru

² Институт прикладной математики им. М. В. Келдыша РАН (ИПМ РАН), Миусская пл., 4А, 125047, Москва; зав. сектором, e-mail: lira@imamod.ru

ными производными, например метод Шварца [1]. В этом методе геометрическая область разбивается на множество микродоменов, что позволяет организовать эффективные параллельные вычисления.

Еще одной областью использования разбиения сеток на микродомены является хранение больших сеток. Разбиение сеток на микродомены позволяет увеличить коэффициент компрессии сеточных данных.

Областью данного исследования являются нерегулярные сетки, содержащие 10^9 и более вершин. В настоящее время такие сетки невозможно разместить в памяти одного процессора, поэтому для декомпозиции нужны параллельные алгоритмы. Методы разбиения графов параллельных пакетов ParMETIS, Jostle, PT-Scotch и Zoltan основываются на иерархических алгоритмах, состоящих из следующих частей: поэтапное огрубление графа, декомпозиция самого маленького из полученных графов и отображение разбиения на предыдущие графы с периодическим локальным уточнением границ доменов. Недостатком таких алгоритмов является образование доменов, границы которых состоят из неоптимальных наборов сегментов. В частности, домены могут оказаться несвязными. Такое ухудшение качества доменов для некоторых задач является критичным. На доменах с длинными границами или сложной конфигурацией алгоритмы решения систем линейных уравнений сходятся за большее число итераций. Связность микродоменов важна при хранении больших сеток, поскольку на связных микродоменах коэффициент сжатия информации о сеточных данных, как правило, будет больше. В алгоритме композиции подобластей у несвязных подобластей длиннее приграничные полосы, в которых требуется повторное вычисление значений, а на узких приграничных полосах возникают проблемы с применимостью метода [2]. Несвязные домены с оторванными ячейками являются неприемлемыми, например, для распараллеливания методики ТИМ-2D решения задач механики сплошной среды на нерегулярных многоугольных сетках произвольной структуры [3]. Другим недостатком указанных пакетов является получение сильно несбалансированных разбиений. В частности, в разбиениях, получаемых пакетом ParMETIS, числа вершин в доменах могут отличаться в два раза. К тому же разбиения больших сеток на большое число микродоменов не всегда удается получить методами существующих пакетов разбиения графов. Вышесказанное обусловило разработку пакета GridSpiderPar для параллельной декомпозиции больших сеток.

2. Алгоритмы параллельной декомпозиции. Разработаны два алгоритма: параллельный алгоритм геометрической декомпозиции сеточных данных и параллельный инкрементный алгоритм декомпозиции графов, на основе которых был создан пакет программ GridSpiderPar декомпозиции больших сеток. Разработанные алгоритмы поддерживают два основных этапа декомпозиции больших сеток: предварительную декомпозицию сетки по процессорам и высокого качества параллельную декомпозицию сетки. Оба алгоритма рассчитаны на нерегулярные сетки, содержащие 10^9 и более вершин.

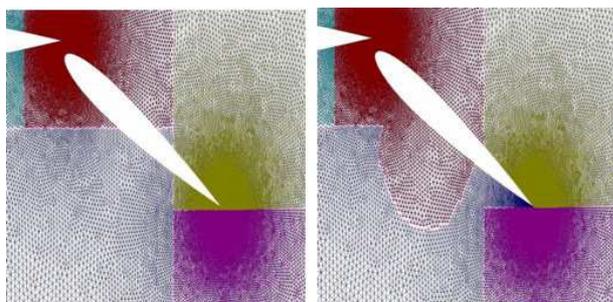


Рис. 1. Фрагменты геометрического распределения вершин по процессорам (слева) и перераспределения малых блоков вершин (справа)

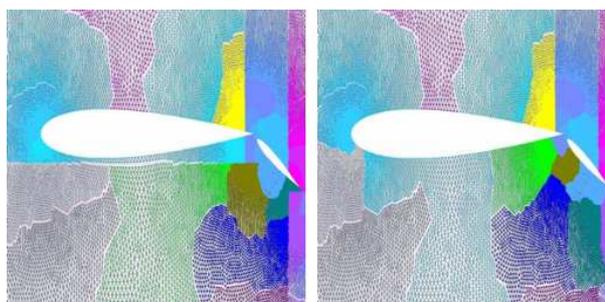


Рис. 2. Локальное разбиение (слева), сбор плохих групп доменов и их повторное разбиение (справа)

2.1. Параллельный инкрементный алгоритм декомпозиции графов. Параллельный инкрементный алгоритм декомпозиции графов пакета GridSpiderPar основан на последовательном инкрементном алгоритме декомпозиции графов [4]. Достоинством инкрементного алгоритма является формирование преимущественно связных доменов.

Параллельный инкрементный алгоритм предполагает выполнение этапов:

- геометрическое распределение вершин по процессорам с помощью разработанного параллельного алгоритма геометрической декомпозиции сеточных данных;
- перераспределение малых блоков вершин (рис. 1);
- локальное разбиение вершин на каждом из процессоров на домены последовательным инкрементным

алгоритмом декомпозиции графов; на рис. 2 слева четко выражены границы между процессорами, домены не пересекают эти границы;

- перераспределение плохих групп доменов; сбор каждой группы плохих доменов на одном процессоре;
- локальное повторное разбиение плохих групп доменов инкрементным алгоритмом декомпозиции графов; на рис. 2 справа видно, что после перераспределения плохих групп доменов и их повторного разбиения домены вышли за границы между процессорами.

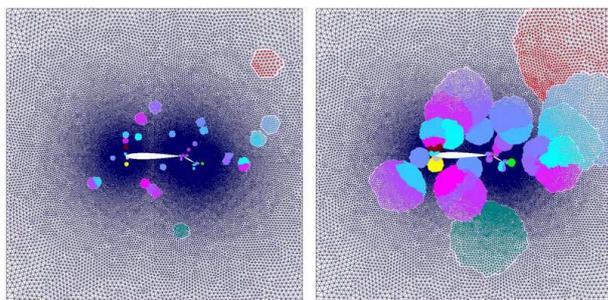


Рис. 3. Инкрементный рост доменов

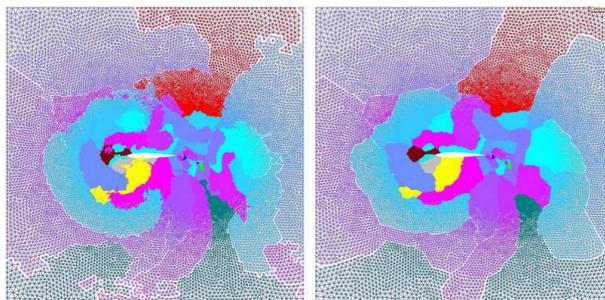


Рис. 4. Диффузное перераспределение вершин между доменами (слева) и локальное уточнение (справа)

При выполнении локального разбиения вначале выбираются инициализирующие вершины для доменов. Далее разбиение проводится посредством итерационного процесса, на каждом шаге которого выполняются следующие действия:

- инкрементный рост доменов и диффузное перераспределение вершин между доменами (рис. 3 и 4);
- локальное уточнение доменов алгоритмом KL/FM (рис. 4); границы доменов стали более гладкими, однако в разбиении присутствуют несвязные домены (желтый домен);
- проверка качества доменов; если качество доменов соответствует заданному, то разбиение считается найденным и происходит выход из цикла, иначе — переход к следующему этапу;

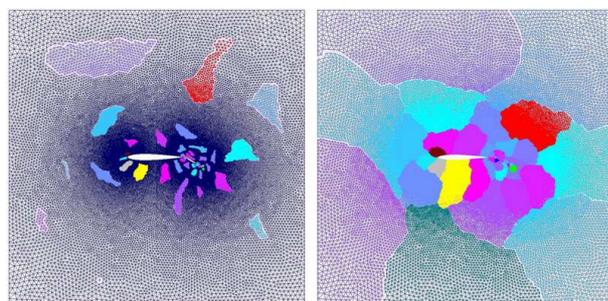


Рис. 5. Освобождение части вершин плохих доменов (слева) и результирующее разбиение (справа)

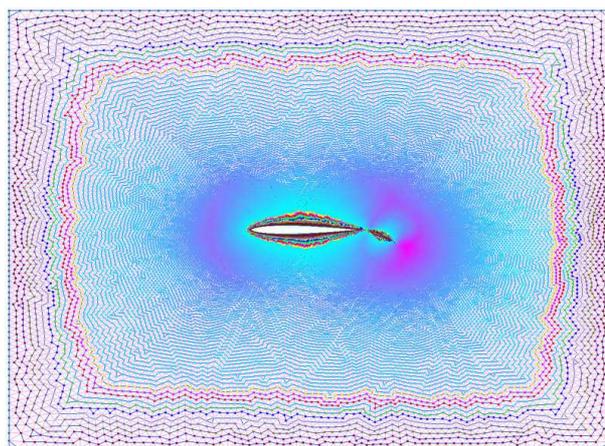


Рис. 6. Оболочки домена

- освобождение части вершин плохих доменов и переход к первому этапу; плохими доменами считаются домены, качество которых не соответствует заданному, а также соседи таких доменов; в плохих доменах часть вершин освобождается, т.е. вновь считается нераспределенной (рис. 5); освобождается часть внешних оболочек, а затем все компоненты связности, кроме той, которая содержит наибольшее число вершин.

В результирующем разбиении на рис. 5 видно, что теперь желтый домен является связным.

Качество доменов проверяется следующим образом. Все вершины, расположенные на геометрической границе графа (вершины, аппроксимирующие границу моделируемой области) и на границах между доменами, считаются принадлежащими первой оболочке своего домена. В каждом домене вершинами второй оболочки считаются соседи вершин из первой оболочки данного домена, которые так же принадлежат этому домену и не попали в первую оболочку. Остальные оболочки вычисляются аналогично. На рис. 6 представлены оболочки домена при условии, что вся сетка принадлежит одному домену. Проверяется связность оболочек каждого домена и вычисляется номер несвязной оболочки с наименьшим номером (номер первой несвязной оболочки) в каждом домене. Хорошими считаются те домены, номер первой несвязной оболочки в которых не меньше заданного порогового значения.

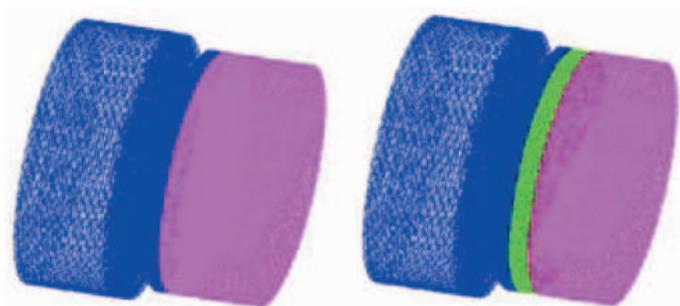


Рис. 7. Геометрическое разбиение сетки на семь доменов на трех процессорах. Первые два этапа разбиения: распределение вершин по процессорам



Рис. 8. Результат геометрического разбиения сетки на семь доменов на трех процессорах

2.2. Параллельный алгоритм геометрической декомпозиции сеточных данных. Параллельный алгоритм геометрической декомпозиции сеточных данных пакета GridSpiderPar основывается на методе рекурсивной координатной бисекции [5]. На каждом этапе рекурсивной бисекции область разбивается на две части. Полученные подобласти разбиваются дальше аналогичным образом до тех пор, пока в подобластях не останется по одному домену.

Основные этапы алгоритма следующие:

- случайное начальное распределение вершин по процессорам; начальное распределение может быть любым, например в соответствии с порядковыми номерами вершин;
- рекурсивная координатная бисекция вершин по процессорам (рис. 7):
 - параллелепипед, заключающий в себе сетку, разбивается на две части; выбирается координатная ось, вдоль которой параллелепипед имеет наибольшую протяженность; параллелепипед разрезается перпендикулярно выбранной оси;
 - группа процессоров делится на две, далее каждая из групп делит свой блок вершин аналогичным образом; для разделения блока вершин используется параллельная сортировка [6] по выбранной оси координат (а также по остальным для разрезания секущей плоскости);
- локальная рекурсивная координатная бисекция вершин по доменам; дальнейшее разбиение на домены проводится локально на каждом процессоре (рис. 8).

Достоинством геометрического алгоритма является то, что при разбиении на равные домены числа вершин в получаемых доменах отличаются не больше, чем на единицу.

Подобный алгоритм реализован в пакете Zoltan [5]. Отличие рекурсивной координатной бисекции созданного алгоритма от аналогичного алгоритма в пакете Zoltan состоит в том, что в нем секущая плоскость (медиана) при необходимости разрезается по нескольким координатам, что позволяет обрабатывать ситуации наличия на одной плоскости множества вершин с одинаковым значением координаты (рис. 9). В пакете Zoltan вершины из медианы распределяются по областям произвольным образом, что увеличивает число разрезанных ребер.

Более подробное описание разработанных алгоритмов можно найти в статье [7].

3. Результаты.

3.1. Разбиение на микродомены и домены. Проведены вычислительные эксперименты по сравнению разбиений на микродомены, разбиений графов микродоменов на домены, а также разбиений сразу на домены нескольких тетраэдральных сеток ($10^8 \div 2.7 \times 10^8$ вершин, $7 \times 10^8 \div 1.6 \times 10^9$ тетраэдров, $8 \times 10^8 \div 1.9 \times 10^9$ ребер) (рис. 10), полученных методами разработанного пакета GridSpiderPar и пакетов ParMETIS, Zoltan и PT-Scotch. Качество разбиений проверялось по дисбалансу числа вершин в доменах, числу несвязных доменов и числу разрезанных ребер.

В сравнении участвовали следующие методы:

- IncrDecomp – параллельный инкрементный алгоритм декомпозиции графов пакета GridSpiderPar;
- PartKway – иерархический алгоритм разбиения графов пакета ParMETIS;
- PartGeomKway – иерархический алгоритм разбиения графов пакета ParMETIS, выполняющий предварительное геометрическое разбиение с использованием кривой Гильберта;
- PT-Scotch – иерархический диффузионный алгоритм пакета PT-Scotch;
- GeomDecomp – параллельный алгоритм геометрической декомпозиции сеточных данных пакета GridSpiderPar;
- RCB – алгоритм рекурсивной координатной бисекции пакета Zoltan.

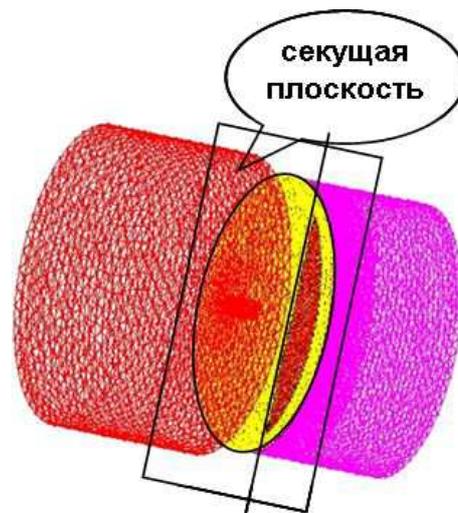


Рис. 9. Разрезание секущей плоскости

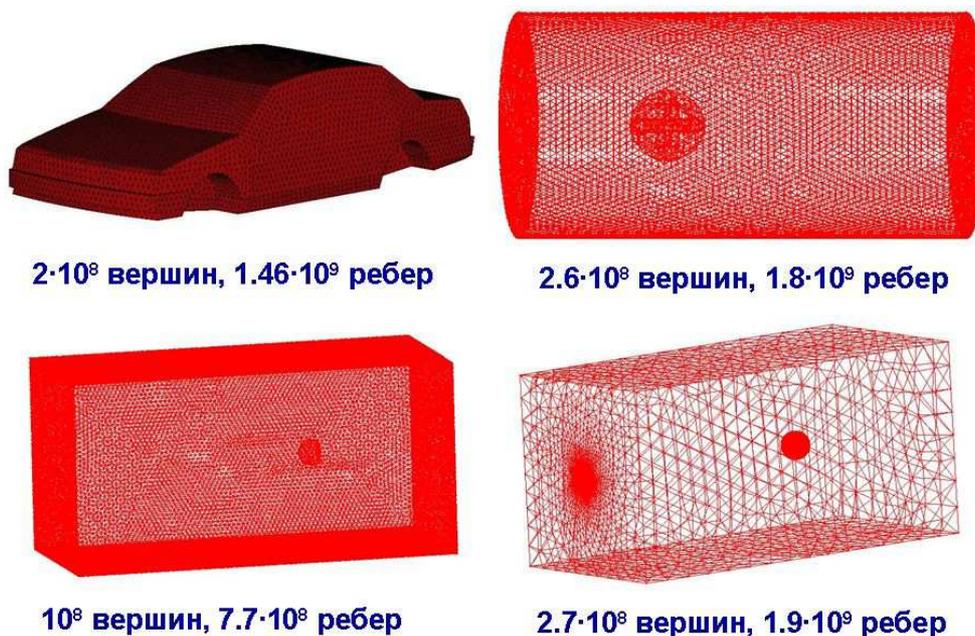


Рис. 10. Тетраэдральные сетки

Вначале сетки были разбиты на 25 600 микродоменов (табл. 1 и 2). Результаты табл. 1 показывают, что дисбаланс числа вершин в микродоменах в разбиениях, полученных методами пакета ParMETIS, достигает 60%, а в разбиениях, сформированных пакетом PT-Scotch, — 8%, в то время как почти во всех разбиениях, полученных методом IncrDecomp, дисбаланс меньше 1%. В разбиениях, образованных геометрическими методами, как и предполагалось, число вершин в микродоменах отличаются не более чем на единицу. Здесь и далее под дисбалансом подразумевается процентное отношение максимального модуля отклонения от среднего арифметического числа вершин в микродомене.

Как видно из табл. 2, число несвязных микродоменов в разбиениях, полученных методами пакета ParMETIS, достигает 69 из 25 600, как и в разбиениях, образованных геометрическими методами. Для пакета PT-Scotch это число достигает 7. Обычно в разбиениях, получаемых пакетом PT-Scotch, почти все микродомены связны, но PT-Scotch не гарантирует связность формируемых микродоменов. Почти во

всех разбиениях, образуемых методом IncrDecomp, все микродомены связны.

Затем по разбиениям тетраэдральных сеток на 25 600 микродоменов, полученных методами пакета ParMETIS и разработанного пакета GridSpiderPar, были составлены графы связей между микродоменами с весами вершин, соответствующими количеству вершин в микродоменах. Графы микродоменов были разбиты на 512 доменов методами PartGraphRecursive и PartGraphKway пакета METIS, методом PartKway пакета ParMETIS и методом IncrDecomp пакета GridSpiderPar, запущенными на одном процессоре. Проведено сравнение различных вариантов разбиений графов микродоменов на домены между собой и с разбиениями сразу на 512 доменов, полученных методами PartKway и PartGeomKway пакета ParMETIS, диффузионным алгоритмом пакета PT-Scotch и методом GeomDecomp пакета GridSpiderPar.

Таблица 1
Дисбаланс числа вершин в
25 600 микродоменах, %

Методы	Сетки			
	1	2	3	4
Разбиение графов				
IncrDecomp	3.5	0.1	0.3	0.2
PartKway	53.4	59.8	58.6	64.3
PartGeomKway	48.7	50.4	62.4	56.5
PT-Scotch	8.3	8.3	8.3	8.3
Геометрические методы				
GeomDecomp	0.01	0.01	0.02	0.01
RCB	0.01	0.01	0.02	0.01

Таблица 2
Число несвязных
микродоменов из 25 600

Методы	Сетки			
	1	2	3	4
Разбиение графов				
IncrDecomp	0	0	0	1
PartKway	69	35	37	29
PartGeomKway	67	34	28	37
PT-Scotch	7	0	2	4
Геометрические методы				
GeomDecomp	62	38	16	33
RCB	64	43	14	44

Результаты табл. 3 показывают, что дисбаланс числа вершин в доменах в разбиениях, полученных напрямую методами пакета ParMETIS, достигает 50%, а в разбиениях, сформированных пакетом PT-Scotch — 5%. Дисбаланс числа вершин в доменах, сформированных из микродоменов, не зависит от дисбаланса числа вершин в микродоменах и составляет около 5%. Видимо, это связано с малым числом микродоменов в домене (50) и недостаточной чувствительностью алгоритмов разбиения графов к весам вершин. Наименьший дисбаланс числа вершин в доменах оказался в разбиениях графов микродоменов, сформированных методами пакета METIS.

Время работы алгоритма IncrDecomp в несколько раз больше времени работы остальных алгоритмов. Время работы алгоритма GeomDecomp не отличается от времени работы других геометрических алгоритмов. Однако алгоритмы разрабатывались в качестве инструментов статической декомпозиции, которая проводится один раз перед расчетом, а время расчета физических задач в разы больше, чем время получения разбиений. Поэтому увеличение времени получения разбиений не является таким существенным.

3.2. Тестирование различных разбиений на задачах газовой динамики. На задачах газовой динамики проведено тестирование разбиений, полученных методами разработанного пакета GridSpiderPar и пакетов ParMETIS, Zoltan и PT-Scotch.

Первая задача — моделирование газоплазменных потоков в диверторе токамака ITER. Токамак — тороидальная установка для магнитного удержания плазмы с целью достижения условий, необходимых для протекания управляемого термоядерного синтеза. Дивертор является одним из ключевых компонентов токамака ITER (рис. 11). Он расположен вдоль нижней части вакуумной камеры и служит для приема

Таблица 3
Дисбаланс числа вершин
в 512 доменах, %

Методы	Сетки			
	1	2	3	4
Разбиение графов				
PartKway	12.9	20.6	17.6	28.4
PartGeomKway	31.1	35.7	44.2	51.4
PT-Scotch	4.9	1.7	2.8	2.9
Геометрические методы				
GeomDecomp	0	0	0	0
Разбиение графов микродоменов				
Среднее арифм.	5.3	5.4	3.7	5.1

потоков примесей и излучений из плазмы.

Расчетная область аппроксимировалась тетраэдральной сеткой, содержащей порядка 2.8×10^6 ячеек (divertor). Считалась полная система уравнений радиационной магнитной газовой динамики с учетом радиационного и кондуктивного теплопереноса и турбулентной вязкости. Использовались явные и неявные схемы. Вычисления проводились на 256 ядрах.

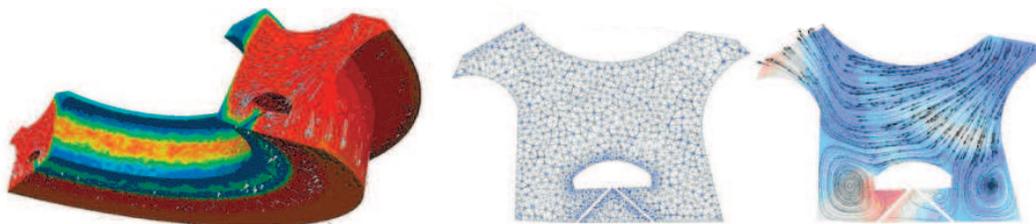


Рис. 11. Моделирования газоплазменных течений в диверторе токамака ITER

Вторая задача — моделирование распространения ударной волны от приземного источника энергии взрывного типа (рис. 12). Для моделирования приземного взрыва была выбрана кубическая область, которая аппроксимировалась гексаэдральными сетками, содержащими порядка 6.1×10^7 (boom) и 1.16×10^8 (boomL) ячеек со сгущением в области взрыва. Вычисления проводились на 4096 и 10 080 ядрах соответственно. Считалась полная система уравнений газовой динамики с учетом кондуктивного теплопереноса. Турбулентные потоки не учитывались. Использовались явные и неявные схемы.

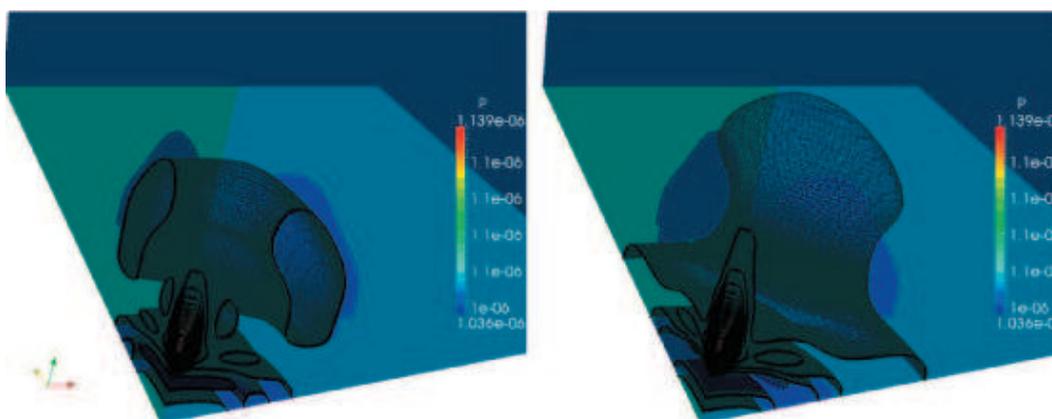


Рис. 12. Аппроксимация изоповерхностей давления на расчетную сетку в момент времени $t=1000$ мс

Для всех расчетных сеток были построены дуальные графы с числом вершин $2.8 \times 10^6 \div 1.2 \times 10^8$ и числом ребер $2.3 \times 10^7 \div 1.0 \times 10^9$. Разбиения дуальных графов были получены методами пакетов GridSpiderPar, ParMETIS, Zoltan и PT-Scotch.

К списку сравниваемых методов были добавлены два метода:

- RIB — алгоритм рекурсивной инерциальной бисекции пакета Zoltan;
- HSFC — алгоритм пакета Zoltan, выполняющий геометрическое разбиение с использованием кривой Гильберта.

На рис. 13–15 метод IncrDecomp обозначен через I, PartKway — PK, PartGeomKway — PGK, PT-Scotch — PTScotch и GeomDecomp — G. Методы IncrDecomp и GeomDecomp окрашены отличающимися цветами, а алгоритмы разбиения графов и геометрические алгоритмы разделены между собой.

На рис. 13, 14 отображен дисбаланс числа вершин в доменах с точки зрения недостатка числа вершин и избытка числа вершин соответственно. Недостаток числа вершин в доменах в разбиениях, полученных методами пакета ParMETIS, достигает 80%, избыток — 5%. Дисбаланс в разбиениях, сформированных пакетом PT-Scotch, в обоих случаях около 5%, а в разбиениях, образованных методом IncrDecomp, дисбаланс меньше 0.1%.

Наименьшее число разрезанных ребер получено методами пакета ParMETIS и алгоритмом IncrDecomp пакета GridSpiderPar.

Проведено сравнение эффективности параллельного счета физических задач пакетом MARPLE3D [8] при распределении сеток по ядрам в соответствии с различными разбиениями. Параллельный программ-

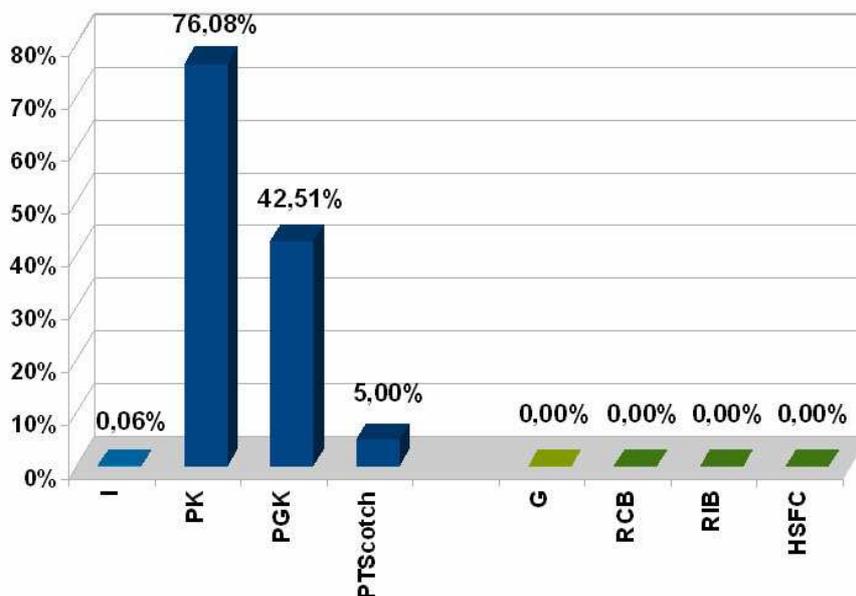


Рис. 13. Дисбаланс числа вершин в доменах: недостаток вершин (boom)

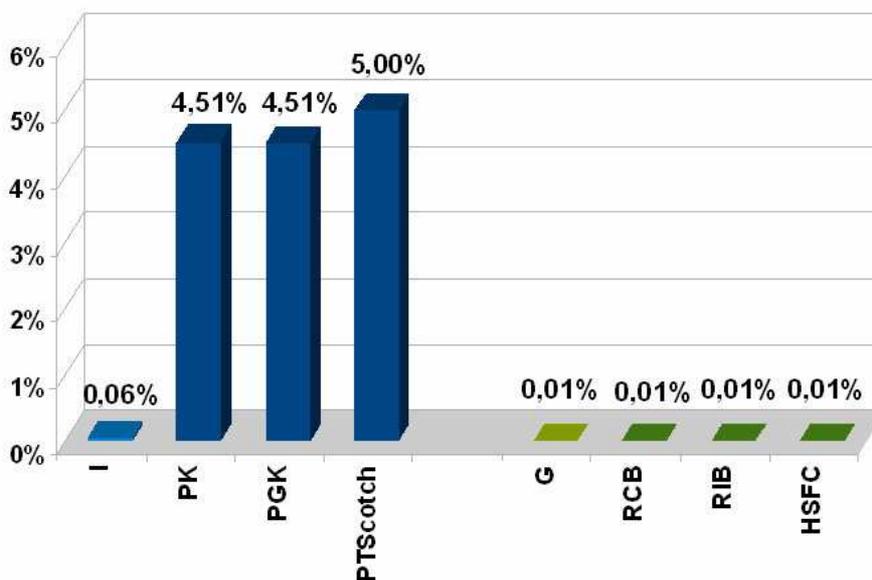


Рис. 14. Дисбаланс числа вершин в доменах: избыток вершин (boom)

ный комплекс MARPLE3D создан в ИПМ им. М. В. Келдыша РАН, и его предметной областью являются задачи двухтемпературной радиационной магнитной гидродинамики. Для расчета каждой физической задачи на всех разбиениях выделялось одинаковое машинное время. Были получены числа шагов по времени, до которых досчитали задачи.

Полученные результаты (рис. 15) демонстрируют, что по числу шагов по времени метод IncDecomp опережает остальные методы декомпозиции графов, а метод GeomDecomp несколько опережает другие геометрические методы. На разбиениях, полученных методами разбиения графов, числа шагов по времени больше, чем на разбиениях, полученных геометрическими методами, не учитывающими связи между вершинами.

3.3. Тестирование разбиений графов микродоменов на задаче приземного взрыва. Дуальный граф boomL, содержащий 1.2×10^8 вершин и 1.0×10^9 ребер, был разбит на различное число микродоменов (от 24576 до 196608) и сразу на 3072 домена алгоритмом IncDecomp разработанного пакета

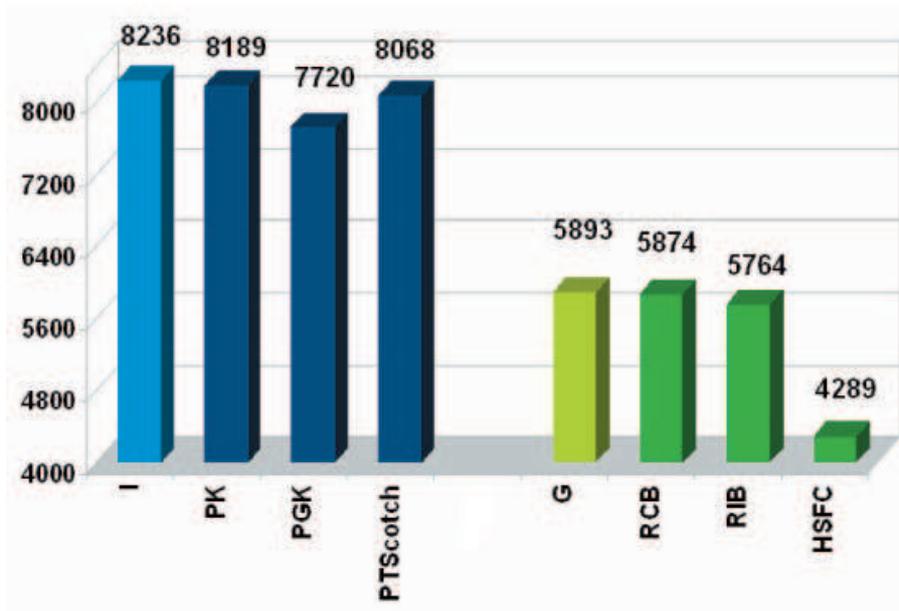


Рис. 15. Число шагов по времени за 1 час (divertor)

GridSpiderPar. Составлены графы связей микродоменов с весами вершин, соответствующими количеству вершин в микродоменах. Графы микродоменов были разбиты алгоритмом IncrDecomp на 3072 домена.

Для расчета задачи моделирования распространения ударной волны от приземного взрыва на всех разбиениях выделялось одинаковое машинное время (5 часов). Были получены числа шагов по времени, до которых досчитала задача.

Таблица 4

Тестирование разбиений графов микродоменов на задаче приземного взрыва

Информация о сетке	Имя: BoomL (116 214 272 гексаэдров)				
Микродомены	3072	24 576	49 152	98 304	196 608
Микродомены в домене	1	8	16	32	64
Дисбаланс, %	9.1	62.5	37.5	18.7	7.9
Разрезанные ребра	5.31×10^7	6.46×10^7	6.65×10^7	6.88×10^7	6.82×10^7
Соседние домены (макс.)	28	25	25	23	21
Шаги по времени	1107	833	880	949	999

Как видно из табл. 4, разбиения содержали от 1 (разбиение сразу на 3072 домена) до 64 микродоменов в домене. Чем больше микродоменов, тем меньше дисбаланс получаемых разбиений и тем меньше максимальное число соседних доменов, но тем больше общее число разрезанных ребер. Увеличение общего числа разрезанных ребер объясняется тем, что при составлении графов микродоменов не учитывались веса ребер между микродоменами. Максимальное число соседних доменов влияет на количество обменов между процессорами, обрабатывающими данные домены. С увеличением числа микродоменов увеличивается также число шагов по времени, полученных на разбиениях. Это говорит о том, что для задачи моделирования приземного взрыва равномерность распределения вычислительной нагрузки по процессорам и количество обменов между процессорами критичнее, чем объем передаваемых данных. При сравнении разбиения сразу на 3072 домена и разбиений графов микродоменов заметно, что в разбиении, составленном из 196 608 микродоменов (64 микродомена в домене), дисбаланс числа вершин в доменах меньше, чем в разбиении сразу на домены, и меньше максимальное число соседних доменов. Этот результат объясняется тем, что при разбиении на определенное количество доменов не всегда удается получить требуемый дисбаланс. Например, при разбиении данного графа на 4096 доменов получаемый дисбаланс составлял 0.03%, что значительно меньше 9.1%, полученных при разбиении на 3072 домена. Число шагов по времени, полученных на разбиении, составленном из 196 608 микродоменов, не намного меньше, чем полученных

на разбиении сразу на домены.

Таким образом, можно сделать вывод о том, что при достаточном количестве микродоменов в доменах разбиения графов микродоменов не уступают по качеству разбиению сразу на домены, что подтверждается малым уменьшением скорости счета рассматриваемой физической задачи. К тому же на декомпозицию графа микродоменов при массовых расчетах требуется меньше процессорочасов.

4. Заключение. Разработан пакет параллельной декомпозиции больших сеток GridSpiderPar, в который вошли два алгоритма: параллельный алгоритм геометрической декомпозиции сеточных данных и параллельный инкрементный алгоритм декомпозиции графов. Разработанные алгоритмы поддерживают два основных этапа декомпозиции больших сеток: предварительную декомпозицию сетки по процессорам и высокого качества параллельную декомпозицию сетки. Оба алгоритма рассчитаны на нерегулярные сетки, содержащие 10^9 и более вершин. Достоинством инкрементного алгоритма является формирование преимущественно связанных доменов.

Проведены вычислительные эксперименты по сравнению различных разбиений на микродомены, разбиений графов микродоменов на домены, а также разбиений сразу на домены нескольких сеток (10^8 вершин, 10^9 элементов), полученных методами созданного комплекса программ GridSpiderPar и пакетов ParMETIS, Zoltan и PT-Scotch. Качество разбиений проверялось по дисбалансу числа вершин в доменах, числу несвязных доменов и числу разрезанных ребер, а также по эффективности параллельного счета задач газовой динамики при распределении сеток по ядрам в соответствии с различными разбиениями. Полученные результаты выявили преимущества разработанных алгоритмов.

Вычисления проводились на кластерах МВС-100К, “Ломоносов” и Helios.

Работа выполнена при поддержке РФФИ (гранты 13-01-12073 офи_м, 14-01-00663 А, 14-07-00712 А и 15-07-04213 А).

Статья рекомендована к публикации Программным комитетом Международной научной конференции “Суперкомпьютерные дни в России 2015” (<http://russianscdays.org>).

СПИСОК ЛИТЕРАТУРЫ

1. *Smith B., Bjorstad P., Gropp W.* Domain decomposition: parallel multilevel methods for elliptic partial differential equations. New York: Cambridge University Press, 1996.
2. *Илюшин А.И., Колмаков А.А., Меньшов И.С.* Построение параллельной вычислительной модели путем композиции вычислительных объектов // Математическое моделирование. 2011. **23**, № 7. 97–113.
3. *Воропинов А.А.* Декомпозиция данных для распараллеливания методики ТИМ-2D и критерии оценки ее качества // Вестник ЮУрГУ. Серия “Математическое моделирование и программирование”. 2009. Вып. 4. 40–50.
4. *Якововский М.В.* Инкрементный алгоритм декомпозиции графов // Вестник Нижегородского университета им. Н.И. Лобачевского. Серия “Математическое моделирование и оптимальное управление”. 2005. Вып. 1. 243–250.
5. *Boman E., Devine K., Catalyurek U., Bozdag D., Hendrickson B., Mitchell W.F., Teresco J.* Zoltan: parallel partitioning, load balancing and Data-Management Services. Albuquerque: Sandia National Laboratories, 2010 (http://www.cs.sandia.gov/Zoltan/dev_html/dev.html).
6. *Якововский М.В.* Параллельные алгоритмы сортировки больших объемов данных // Фундаментальные физико-математические проблемы и моделирование технико-технологических систем. Вып. 7. М.: Янус, 2004. 235–249.
7. *Головченко Е.Н.* Параллельный пакет декомпозиции больших сеток // Математическое моделирование. 2011. **23**, № 10. 3–18.
8. *Гасилов В.А., Болдарев А.С., Дьяченко С.В. и др.* Пакет прикладных программ MARPLE3D для моделирования на высокопроизводительных ЭВМ импульсной магнитоускоренной плазмы // Математическое моделирование. 2012. **24**, № 1. 55–87.

Поступила в редакцию
2.09.2015

Parallel Partitioning Tool GridSpiderPar for Large Mesh Decomposition

E. N. Golovchenko¹ and M. V. Yakobovskiy²

¹ *Keldysh Institute of Applied Mathematics, Russian Academy of Sciences; Miusskaya ploshchad' 4, Moscow, 125047, Russia; Ph.D., Scientist, e-mail: ge03@imamod.ru*

² *Keldysh Institute of Applied Mathematics, Russian Academy of Sciences; Miusskaya ploshchad' 4, Moscow, 125047, Russia; Dr. Sci., Professor, Head of Sector, e-mail: lira@imamod.ru*

Received September 2, 2015

Abstract: The problem of load balancing arises in parallel mesh-based numerical solution of problems of continuum mechanics, energetics, electrodynamics etc. on high-performance computing systems. The number of processors to run a computational problem is often unknown. It makes sense, therefore, to partition a mesh into a great number of microdomains which then are used to create subdomains. Graph partitioning methods implemented in state-of-the-art parallel partitioning tools ParMETIS, Jostle, PT-Scotch and Zoltan are based on multilevel algorithms. That approach has a shortcoming of forming unconnected subdomains. Another shortcoming of present graph partitioning methods is generation of strongly imbalanced partitions. The program package for parallel large mesh decomposition GridSpiderPar was developed. We compared different partitions into microdomains, microdomain graph partitions and partitions into subdomains of several meshes (10^8 vertices, 10^9 elements) obtained by means of the partitioning tool GridSpiderPar and the packages ParMETIS, Zoltan and PT-Scotch. Balance of the partitions, edge-cut and number of unconnected subdomains in different partitions were compared as well as the computational performance of gas-dynamic problem simulations run on different partitions. The obtained results demonstrate advantages of the devised algorithms.

Keywords: parallel programming, graph partitioning, mesh decomposition.

References

1. B. Smith, P. Bjorstad, and W. Gropp, *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations* (Cambridge Univ. Press, New York, 1996).
2. A. I. Ilyushin, A. A. Kolmakov, and I. S. Menshov, "Construction of a Parallel Computational Model by Composition of Computational Facilities," *Mat. Model.* **23** (7), 97–113 (2011) [*Math. Models Comput. Simul.* **4** (1), 118–128 (2012)].
3. A. A. Voropinov, "Data Decomposition for TIM-2D Code Parallelizing Method and Its Quality Evaluation Criteria," *Vestn. South Ural State Univ. Ser. Math. Model. Comput. Softw.*, No. 4, 40–50 (2009).
4. M. V. Yakobovskii, "An Incremental Algorithm for Graph Decomposition," *Vestn. Lobachevskii Univ. Nizhni Novgorod*, No. 1, 243–250 (2005).
5. E. Boman, K. Devine, U. Catalyurek, et al., *Zoltan: Parallel Partitioning, Load Balancing and Data-Management Services*, Sandia Nat. Labs. http://www.cs.sandia.gov/Zoltan/dev_html/dev.html. Cited October 22, 2015.
6. M. V. Yakobovskii, "Parallel Sorting Algorithms for Large Data Volumes," in *Fundamental Physico-Mathematical Problems and Modeling of Technological Systems* (Yanus, Moscow, 2004), Issue 7, pp. 235–249.
7. E. N. Golovchenko, "Parallel Partitioning Tool for Large Mesh Decomposition," *Mat. Model.* **23** (10), 3–18 (2011).
8. V. A. Gasilov, A. S. Boldarev, S. V. D'yachenko, et al., "Program Package MARPLE3D for Simulation of Pulsed Magnetically Driven Plasma Using High Performance Computing," *Mat. Model.* **24** (1), 55–87 (2012).