

ГОСУДАРСТВЕННОЕ НАУЧНОЕ УЧРЕЖДЕНИЕ  
«ИНСТИТУТ МАТЕМАТИКИ  
НАЦИОНАЛЬНОЙ АКАДЕМИИ НАУК БЕЛАРУСИ»

Объект авторского права  
УДК 519.6

**ТОЛСТИКОВ**  
Алексей Александрович

**МЕТОДЫ ПОСТРОЕНИЯ И ИССЛЕДОВАНИЯ  
ПАРАЛЛЕЛЬНЫХ АЛГОРИТМОВ  
ДЛЯ РЕАЛИЗАЦИИ НА СУПЕРКОМПЬЮТЕРАХ С  
РАСПРЕДЕЛЕННОЙ ПАМЯТЬЮ**

**АВТОРЕФЕРАТ**  
диссертации на соискание ученой степени  
кандидата физико-математических наук  
по специальности 01.01.07 — вычислительная математика

Минск, 2023

Работа выполнена в Белорусском государственном университете.

Научный руководитель: **Лиходед Николай Александрович**,  
доктор физико-математических наук, профессор  
кафедры вычислительной математики  
Белорусского государственного университета.

Официальные оппоненты: **Волков Василий Михайлович**,  
доктор физико-математических наук, доцент,  
заведующего кафедрой веб-технологий и  
компьютерного моделирования механико-  
математического факультета Белорусского  
государственного университета;

**Чиж Олег Петрович**,  
кандидат физико-математических наук,  
заведующий лабораторией высокопроизводи-  
тельных систем Объединенного института про-  
блем информатики НАН Беларуси.

Оппонирующая организация: **Учреждение образования «Брестский го-  
сударственный университет имени А.С.  
Пушкина».**

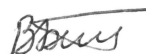
Защита состоится «01» июня 2023 года в 14:30 на заседании совета по защите диссертаций Д 01.02.02 при Государственном научном учреждении «Институт математики Национальной академии наук Беларуси» по адресу: 220072, Республика Беларусь, г. Минск, ул. Сурганова, д. 11.

Тел. ученого секретаря совета: (+375-17) 378-17-62, e-mail:  
vbened@im.bas-net.by.

С диссертацией можно ознакомиться в библиотеке Государственного научного учреждения «Институт математики Национальной академии наук Беларуси».

Автореферат разослан «28» апреля 2023 года.

Ученый секретарь совета  
по защите диссертаций, кандидат  
физико-математических наук, доцент



В. И. Бенедиктович

# Введение

Диссертационная работа относится к области вычислительной математики — распараллеливание вычислительных алгоритмов — изучающей методы преобразования и разработки алгоритмов для выполнения на многопроцессорных системах. Основные результаты работы посвящены особенностям параллельных алгоритмов для выполнения на вычислительных системах с распределенной памятью, но некоторые результаты могут быть использованы и в системах с общей памятью.

Бурное развитие вычислительных систем в последние два-три десятилетия является следствием развития общества и его информационных потребностей с одной стороны и стало катализатором увеличения интереса к более масштабным и точным моделям физического и цифрового мира вокруг нас с другой. В наши дни мы не удивляемся точному расписанию общественного транспорта, адаптивному регулированию светофоров, практически поминутному прогнозу погоды в мобильных устройствах, приближающейся эре автономного транспорта и метавселенных.

Большинство масштабных цифровых проектов моделируются и эксплуатируются на распределенных многопроцессорных системах, которые в свою очередь могут дублироваться в нескольких датацентрах, расположенных на различных континентах. Фактически мы могли бы представить многоуровневую систему обмена данными для работы приложений от коммуникации между датацентрами до обмена обновленными данными в регистрах ядер процессора. При этом каждый из уровней этой иерархии коммуникаций имеет свои характеристики, сильные и слабые стороны. Также стоит отметить, что последнее десятилетие ознаменовано огромным интересом к децентрализованным системам, с одной стороны это технология блокчейн, а с другой системы децентрализованного обучения моделей машинного обучения и хранения данных.

Мировое научное и инженерное сообщество уделяет большое внимание развитию суперкомпьютеров. Большинство ведущих университетов имеют доступ либо к собственной многопроцессорной системе, либо к арендованным или предоставленным для исследований облачным ресурсам. При этом специализированные кластерные системы тоже продолжают создаваться и развиваться. Об этом можно смело судить по регулярно публикуемым рейтингам Top-500 (производительность на бенчмарке LINPACK), HPCG (производительность в тесте на вычисление сопряженных градиентов), Green-500 (производительность, которую удается показывать суперкомпьютеру, в расчете на один ватт) и Graph-500 (производительность в задачах обхода графа в ширину и поиска кратчайших путей).

Стоит отметить, что использование многопроцессорных систем в наши дни позволяет углубляться в очень разнообразные аспекты исследований. С одной стороны, для увеличения производительности систем необходимо решать сложные инженерные задачи по увеличению тактовой частоты ядер процессоров и

объединению узлов в одну систему, изменению архитектуры доступа к памяти и увеличению скорости подачи данных к процессору, усовершенствованию надежности и пропускной способности сетевых протоколов. С другой стороны, создавать высокопроизводительное программное обеспечение и инструменты для него.

Если задачи обработки данных обычно решаются потоковыми алгоритмами, то для задач математического моделирования методы решения, как правило, являются итерационными. Объектом исследования данной работы являются преобразования алгоритмов, основная вычислительная трудоемкость которых описывается многомерным вложенным циклом, к параллельной реализации. Предложенный в работе метод позволяет получить целое семейство параллельных алгоритмов, среди которых могут быть и не самые эффективные. Поэтому необходимо, проверив корректность преобразований, сравнить полученные модификации. Изучается возникающая сложность распределения исходных данных и результатов, направление и объем коммуникаций между узлами вычислительной системы.

В качестве целевого компьютера для реализации алгоритмов будем рассматривать многопроцессорную систему с распределенной памятью. К компьютерам этого класса, в частности, относятся белорусско-российские мультипроцессорные вычислительные системы «СКИФ».

Для высокой степени полезной эксплуатации компьютеров с распределенной памятью параллельные алгоритмы должны быть зернистыми: множество операций алгоритма должно быть разбито на множества, называемые зернами вычислений. Мы будем использовать тайлинг (от англ. tiling) — преобразование алгоритма для получения макроопераций-тайлов. Операции одного тайла выполняются атомарно, как одна единица вычислений, а обмен данными происходит массивами. В диссертационной работе предлагается метод получения параллельных последовательностей зернистых вычислений, задание которых позволяет организовать параллельные вычислительные процессы. Разработан математический аппарат оценки степени параллелизма, получены условия параллельности последовательностей.

Для распараллеливания алгоритма требуется получить его информационную структуру, исследовать параллельную структуру и возможности параллелизма, т. е. наличие независимых в какие-то моменты времени множеств операций. Мощным математическим аппаратом для описания информационных связей на уровне отдельных операций алгоритма являются функции зависимостей. Доказано (теорема В. В. Воеводина об информационном покрытии), что все зависимости алгоритмов, представленных довольно широким классом программ, могут быть заданы функциями, линейно зависящими от параметров циклов и внешних переменных.

При реализации параллельного алгоритма требуется добавить коммуникационные операции между вычислительными процессами, для чего требуется полу-

чить зависимости между тайлами. Ранее<sup>1</sup> разработан метод получения всех векторов зависимостей глобального уровня, т. е. уровня макроопераций-тайлов. В диссертационной работе сформулированы и доказаны утверждения, позволяющие получить все глобальные зависимости зернистых версий алгоритмов, представленных многомерными циклами с аффинными зависимостями: функции глобальных зависимостей определяются из функций на уровне операций. Сведения об информационной структуре глобального уровня позволяют формально получать и автоматизировать коммуникационные операции зернистых вычислений.

Локальность параллельного алгоритма, предназначенного для реализации на компьютерах с распределенной памятью, характеризует коммуникационные затраты: чем меньше при заданном числе вычислительных ядер суммарный объем операций обмена данными, тем лучше локальность. Заключительная часть диссертационной работы посвящена разработке метода оценки коммуникационных затрат альтернативных вариантов организации одномерных вычислительных процессов; используется анализ информационных зависимостей, порождающих коммуникационные операции; сформулированы и доказаны утверждения, позволяющие получить асимптотические оценки объема коммуникационных операций.

## Общая характеристика работы

**Связь работы с крупными научными программами (проектами) и темами.** Тема диссертации соответствует направлению научных исследований «Развитие теоретико-методологических основ информатики и информационных технологий», определенному Перечнем приоритетных направлений научных исследований Республики Беларусь на 2006–2010 (постановления Совета Министров Республики Беларусь от 17 мая 2005 г. № 512) и направлению «Цифровые информационно-коммуникационные и междисциплинарные технологии и основанные на них производства: математика и моделирование сложных функциональных систем» Перечня приоритетных направлений научной, научно-технической и инновационной деятельности на 2021–2025 годы (Указ Президента Республики Беларусь от 07 мая 2020 г. № 156).

Диссертационная работа выполнялась на кафедре вычислительной математики факультета прикладной математики и информатики Белорусского государственного университета. Основные результаты получены в рамках выполнения заданий государственной комплексной программы научных исследований «Инфотех» на 2006–2010 гг., договоров с БРФФИ, а также государственных программ научных исследований Республики Беларусь «Конвергенция-2020», подпрограмма «Методы математического моделирования сложных систем», и «Конвергенция-2025», подпрограмма «Математические модели и методы»:

---

<sup>1</sup> Лиходед, Н. А. Информационная структура зернистых алгоритмов с однородными зависимостями / Н. А. Лиходед, П. И. Соболевский // Доклады НАН Беларуси. – 2011. – Т. 55, № 2. – С. 22–26.

1. Государственная комплексная программа научных исследований «Инфотех» на 2006–2010 гг., подраздел «прикладные исследования». Тема: «Разработка, исследование и программная реализация методов глубокого статического распараллеливания программ» (Инфотех-24).

2. Государственная комплексная программа научных исследований «Инфотех» на 2006–2010 гг., подраздел «прикладные исследования». Тема: «Автоматизированная адаптация последовательных программ для параллельных компьютеров с распределенной памятью» (Инфотех-55).

3. Договор с БРФФИ № Ф08-016. Тема: «Разработка и программная реализация методов отображения вычислительных и коммуникационных операций алгоритмов на процессоры параллельного компьютера».

4. Договор с БРФФИ № Ф10-085. Тема: «Обобщенный тайлинг для распараллеливания и улучшения локальности гнезд циклов».

5. Договор с БРФФИ № Ф12ОБ-005. Тема: «Методы и алгоритмы для организации параллельных зернистых вычислений».

6. Государственная программа научных исследований Республики Беларусь «Конвергенция-2020», подпрограмма «Методы математического моделирования сложных систем». Тема: «Разработка методов получения параллельных зернистых алгоритмов, построение эффективных алгоритмов решения многомерных задач математической физики».

7. Государственная программа научных исследований Республики Беларусь «Конвергенция-2025», подпрограмма «Математические модели и методы». Тема НИР: «Разработка методов численного решения дифференциальных задач большой размерности» в рамках задания «Методы вычислительной математики и математического моделирования физических, технических и технологических процессов».

**Цель и задачи исследования.** *Целью исследования* является комплексное изучение возможности построения зернистых версий вычислительных алгоритмов для параллельного исполнения, формализация и анализ порождаемых коммуникационных операций, а также построение метода оценки локальности данных при реализации алгоритмов на вычислительных системах с распределенной памятью.

Цель диссертации достигается путем исследования информационной структуры алгоритмов, взаимосвязи между распределением операций алгоритма и данных между вычислительными узлами, построением математической модели пространства итераций и преобразованием линейных отображений, задаваемых функциями зависимостей. Исследуется возможность описания зернистых алгоритмов инструментами, используемыми для программ с аффинными зависимостями: вектора и функции зависимостей, области определения функций. Сформулированы условия допустимости тайлинга в методе, порождающем целые семейства зернистых алгоритмов, имеющих различные параметры производительности. Кроме теоретического исследования проведены вычислительные

эксперименты на кластерных системах с распределенной памятью.

*Объектами исследования* являются зернистые версии алгоритмов и их свойства. *Предмет исследования* включает в себя оценку возможности параллельного исполнения, коммуникационные операции, порождаемые при реализации алгоритмов под вычислительные системы с распределенной памятью, а также асимптотическая оценка коммуникационных затрат.

**Научная новизна.** В работе рассмотрен подход создания зернистых версий вычислительных алгоритмов: сформулирован метод построения параллельных вычислительных процессов, доказаны условия применимости тайлинга определенного вида, показана оценка возможного параллелизма; разработан метод получения информационной структуры зернистых алгоритмов с аффинными зависимостями, обобщающий метод нахождения векторов зависимостей уровня макроопераций. Формализовано получение коммуникационных операций — результата адаптации алгоритма для параллельного исполнения на системах с распределенной памятью, исследованы объем и направление коммуникационных операций зернистых алгоритмов. Разработан метод получения асимптотических оценок объема коммуникационных операций, что позволяет в достаточной мере сравнить различные версии параллельных алгоритмов с точки зрения локальности данных.

#### **Положения, выносимые на защиту:**

1. Метод построения параллельных вычислительных процессов на компьютерах с распределенной памятью; условия параллельности выполнения макроопераций-тайлов, условия корректности тайлинга.

2. Метод получения информационной структуры зернистых алгоритмов с аффинными зависимостями; утверждения, позволяющие получить все глобальные зависимости зернистых версий алгоритмов, представленных многомерными циклами с аффинными зависимостями.

3. Подход к формализации и включению в схему параллельного зернистого алгоритма с однородными или аффинными зависимостями коммуникационных операций получения и отправки массивов данных; утверждения, позволяющие в явном виде записать новые коммуникационные операции.

4. Оценки коммуникационных затрат при реализации алгоритмов на параллельных компьютерах с распределенной памятью; выражения, характеризующие локальность параллельного зернистого алгоритма: число данных, для которых требуются коммуникации, и число процессов, вовлеченных в пересылки этих данных.

**Личный вклад соискателя ученой степени.** Все новые результаты, представленные в диссертационной работе, получены автором работы лично. Научным руководителем, доктором физико-математических наук, профессором Н. А. Лиходедом была предложена общая тематика исследований, было про-

ведено ознакомление с широким перечнем существующих результатов, были сформулированы рабочие гипотезы и подходы, производились обсуждение и проверка полученных результатов. В совместной статье с Соболевским П. И. соавтору принадлежит вид векторов глобальных зависимостей и соответствующих областей входа-выхода. В совместной статье с Бахановичем С. В. соавтору принадлежит 2D-алгоритм с улучшенной локальностью.

**Апробация диссертации и информация об использовании ее результатов.** Результаты, вошедшие в диссертационную работу, были представлены на следующих конференциях:

1. X Белорусская математическая конференция (Минск, 3–7 ноября 2008 года).

2. Международный конгресс по информатике: информационные системы и технологии, CSIST (Минск, 31 октября – 3 ноября 2011 года).

3. XI Белорусская математическая конференция (Минск, 5–9 ноября 2012 года).

4. XII Белорусская математическая конференция (Минск, 5–10 сентября 2016 года).

5. Международная научная конференция «Параллельные вычислительные технологии», ПАВТ (Пермь, Россия, 31 марта – 2 апреля 2020 года).

6. Международная научная конференция «Параллельные вычислительные технологии», ПАВТ (Волгоград, Россия, 30 марта – 1 апреля 2021 года).

7. XIII Белорусская математическая конференция (Минск, 22–25 ноября 2021 года).

Система «ParProc» — экспериментальный образец программного обеспечения для реализации многомерных циклов на параллельных компьютерах с распределенной памятью. Свидетельство о регистрации компьютерной программы от 20 января 2011 г. № 283).

Результаты диссертации внедрены в учебный процесс на факультете прикладной математики и информатики БГУ: акты от 20 декабря 2010 г., 26 декабря 2018 г. и 20 октября 2022 г.

**Опубликование результатов диссертации.** Список работ по теме диссертации состоит из 15 наименований: 7 публикаций, соответствующих пункту 19 Положения о присуждении ученых степеней и присвоении ученых званий в Республике Беларусь (общим объемом 3,28 авторских листа), одна публикация в других научных изданиях, 3 публикации в сборниках материалов научных конференций и 4 тезиса докладов. Общий объем публикаций по теме диссертации составляет 5,40 авторских листа.

**Структура и объем диссертации.** Диссертация состоит из титульного листа, оглавления, введения, общей характеристики работы, четырех глав, заключения, списка литературы и приложений.



Полный объем диссертации составляет 122 страницы, включая 11 рисунков и 17 листингов алгоритмов, занимающих в совокупности 14,74 страницы. Список литературы содержит 93 наименования, включая публикации автора диссертации, и занимает 9 страниц.

## Основная часть

**Во введении** обоснована актуальность темы диссертационной работы, изложена структура работы и полученные в диссертации результаты.

**Первая глава** содержит необходимые обозначения и понятия, определения, а также известные результаты по теме диссертации, аналитический обзор литературы.

**В разделе 1.1** вводятся необходимые обозначения и понятия, даются определения.

Будем считать, что алгоритм задан последовательной программой линейного класса<sup>2</sup>. Основную вычислительную часть такой программы составляют циклы произвольной структуры вложенности; границы изменения параметров циклов задаются неоднородными формами, линейными по совокупности параметров циклов и внешних переменных; шаги изменения параметров циклов равны 1.

Пусть в гнезде циклов имеется  $K$  выполняемых операторов  $S_\beta$  и используется  $L$  массивов  $a_l$ . Область изменения параметров циклов для оператора  $S_\beta$  и размерность этой области обозначим соответственно  $V_\beta$  и  $n_\beta$ ; через  $\nu_l$  обозначим размерности массивов  $a_l$ .

Пусть в гнезде циклов имеется  $\Theta$  наборов выполняемых операторов. Под набором операторов будем понимать один или несколько выполняемых операторов, окруженных одним и тем же множеством циклов. Выполняемые операторы и наборы операторов линейно упорядочены расположением их в записи алгоритма. Обозначим  $V_\theta$ ,  $1 \leq \theta \leq \Theta$ , — область изменения параметров циклов, окружающих  $\theta$ -й набор операторов,  $n^\theta$  — размерность этой области (число циклов, окружающих  $\theta$ -й набор операторов).

Зависимости между операциями можно задать функциями вида

$$\begin{aligned} \overline{\Phi}_{\alpha,\beta}(J) &= \Phi_{\alpha,\beta}J + \Psi_{\alpha,\beta}N - \varphi^{\alpha,\beta}, \\ J \in V_{\alpha,\beta}, \Phi_{\alpha,\beta} &\in \mathbb{Z}^{n_\alpha \times n_\beta}, \Psi_{\alpha,\beta} \in \mathbb{Z}^{n_\alpha \times e}, \varphi^{\alpha,\beta} \in \mathbb{Z}^{n_\alpha}, \end{aligned} \quad (1)$$

где  $N \in \mathbb{Z}^e$  — вектор внешних переменных алгоритма,  $e$  — число внешних переменных. Функция зависимостей  $\overline{\Phi}_{\alpha,\beta}(J)$  позволяет для операции  $S_\beta(J)$  найти операцию  $S_\alpha(I)$ , от которой  $S_\beta(J)$  зависит.

В случае однородных зависимостей матрица  $\Phi_{\alpha,\beta}$  — единичная, а матрица  $\Psi_{\alpha,\beta}$  — нулевая, поэтому запись функции зависимости может быть упрощена  $\overline{\Phi}_{\alpha,\beta}(J) = J - \varphi^{\alpha,\beta}$ , т. е. сама зависимость описывается вектором  $\varphi^{\alpha,\beta}$ . Векторы  $\varphi^{\alpha,\beta}$  называются векторами зависимостей.

<sup>2</sup> Воеводин, В. В. Параллельные вычисления / В. В. Воеводин, Вл. В. Воеводин. — Санкт-Петербург, БХВ-Петербург, 2002. — 608 с.

Пусть  $m_\zeta^\theta = \min \{j_\zeta \mid J(j_1, \dots, j_{n^\theta}) \in V^\theta\}$  и  $M_\zeta^\theta = \max \{j_\zeta \mid J(j_1, \dots, j_{n^\theta}) \in V^\theta\}$ ,  $1 \leq \zeta \leq n^\theta$ , — предельные значения изменения параметров циклов. Будем считать, что  $m_\zeta^\theta$  не зависит от внешних переменных, а  $M_\zeta^\theta$  может зависеть от внешних переменных:  $M_\zeta^\theta = M^{\theta, N}(\zeta)N + M_\zeta^{\theta, 1}$  ( $M^{\theta, N}(\zeta)$  — вектор-строка). Если два набора операторов  $\theta_1$  и  $\theta_2$  имеют общий цикл с параметром  $j_\zeta$ , то  $m_\zeta^{\theta_1} = m_\zeta^{\theta_2}$ ,  $M_\zeta^{\theta_1} = M_\zeta^{\theta_2}$ .

Размеры тайлов задаются натуральными числами  $r_1^\theta, \dots, r_{n^\theta}^\theta$ . Параметр  $r_\zeta^\theta$  обозначает число значений параметра цикла  $j_\zeta$ , приходящихся на один тайл  $\theta$ -го набора операторов. Если два набора операторов имеют общий цикл с параметром  $j_\zeta$ , то  $r_\zeta^{\theta_1} = r_\zeta^{\theta_2}$ .

Число частей  $Q_\zeta^\theta$ , на которые при формировании тайлов разбивается область значений параметра  $j_\zeta$  цикла, окружающего  $\theta$ -й набор операторов, находится согласно  $Q_\zeta^\theta = \lceil (M_\zeta^\theta - m_\zeta^\theta + 1)/r_\zeta^\theta \rceil$ . Тайлы нумеруются по каждой координате в пределах от 0 до  $Q_\zeta^\theta - 1$ ,  $1 \leq \zeta \leq n^\theta$ . Стоит отметить, что при необходимости выбора зерна вычислений под фиксированное количество тайлов вдоль некоторой координаты, можно воспользоваться симметричной формулой  $r_\zeta^\theta = \lceil (M_\zeta^\theta - m_\zeta^\theta + 1)/Q_\zeta^\theta \rceil$ .

Обозначим  $V^{\theta, gl} = \left\{ J^{gl}(j_1^{gl}, \dots, j_{n^\theta}^{gl}) \mid 0 \leq j_\zeta^{gl} \leq Q_\zeta^\theta - 1, 1 \leq \zeta \leq n^\theta \right\}$  — область изменения параметров глобальных, т. е. уровня тайлов, циклов. Если цикл является глобальным не разбиваемым, то условимся, что границы изменения этого цикла тайлинг оставляет прежними. Каждый тайл можно обозначить некоторым вектором  $J^{gl}$  или, подробнее, вектором  $J^{gl}(j_1^{gl}, \dots, j_{n^\theta}^{gl})$ . Область изменения параметров локальных (уровня операций тайлов) циклов при фиксированных значениях параметров глобальных циклов  $J^{gl} \in V^{\theta, gl}$ , обозначим

$$V_{J^{gl}}^\theta = \left\{ J(j_1, \dots, j_{n^\theta}) \in V^\theta \mid m_\zeta^\theta + j_\zeta^{gl} r_\zeta^\theta \leq j_\zeta \leq m_\zeta^\theta - 1 + (j_\zeta^{gl} + 1)r_\zeta^\theta, 1 \leq \zeta \leq n^\theta \right\}.$$

Отметим, если  $J(j_1, \dots, j_{n^\theta}) \in V_{J^{gl}}^\theta$ , то имеют место равенства

$$j_\zeta^{gl} = \left\lfloor \frac{j_\zeta - m_\zeta^\theta}{r_\zeta^\theta} \right\rfloor, \quad j_\zeta^{gl} = \left\lfloor \frac{j_\zeta - m_\zeta^\theta + 1}{r_\zeta^\theta} \right\rfloor - 1.$$

В разделе 1.2 приводится обзор известных результатов, касающихся преимущественно методов анализа природы зависимостей по данным в алгоритмах и методам организации вычислений на многопроцессорных системах с распределенной памятью.

**Вторая глава** посвящена результатам автора диссертации, связанным с созданием параллельных вычислительных процессов, и анализу эффективности получаемых версий алгоритмов с точки зрения параллельной загрузки.

Рассмотрим гнездо вложенных циклов, которое в общем виде можно представить псевдокодом алгоритма 1. Циклы необязательно вложены тесно: между

операторами **for** и между операторами **endfor** допускается наличие выполняемых операторов. Нижние и верхние границы каждого из циклов являются аффинными выражения от переменных внешних циклов.

---

**Алгоритм 1** Общий вид вложенного цикла

---

```

1: for  $j_1 = l_1, u_1$  do
2:   for  $j_2 = l_2(j_1), u_2(j_1)$  do
3:     .....
4:     for  $j_n = l_n(j_1, \dots, j_{n-1}), u_n(j_1, \dots, j_{n-1})$  do
5:       операторы
6:     end for
7:     .....
8:   end for
9: end for

```

---

Основным результатом **раздела 2.1** являются следующие лемма и теорема.

Пусть  $\Xi_z = \{\zeta_1, \dots, \zeta_z\} \subseteq \{\xi + 1, \dots, n\}$  — множество разбиваемых циклов алгоритма 1, и  $Q_1, \dots, Q_z, z \geq 1$ , — заданные натуральные числа, превосходящие 1. Обозначим  $\bar{B}_k = \left\lceil \frac{M_{\zeta_k} - m_{\zeta_k} + 1}{Q_k} \right\rceil$  и разобьем каждое  $V_{\bar{J}, p}$  на  $Q_1 \times \dots \times Q_z$  частей посредством разбиения циклов  $j_{\zeta_1}, \dots, j_{\zeta_z}$ :

$$V_{\bar{J}, p, \bar{q}} = \left\{ J \in V_{\bar{J}, p} \mid m_{\zeta_k} + (q_k - 1)\bar{B}_k \leq j_{\zeta_k}(J) \leq m_{\zeta_k} - 1 + q_k\bar{B}_k, 1 \leq k \leq z \right\},$$

где  $\bar{q} = (q_1, \dots, q_z)$ .

**Лемма 2.3.** *Множества операций  $V_{\bar{J}, p, \bar{q}}$  могут являться зерном вычислений, если*

$$j_{\zeta} - i_{\zeta} \geq 0, \zeta \in \Xi_z, \quad (2)$$

для всех зависимостей

$$S_{\alpha}(I(i_1, \dots, i_n)) \rightarrow S_{\beta}(J(j_1, \dots, j_n))$$

таких, что  $i_1 = j_1, \dots, i_{\xi-1} = j_{\xi-1}$ .

Для анализа возможности параллельного исполнения полученных зерен вычислений введем обозначения:  $\delta_{\zeta_i}^{(\alpha, \beta)} = \min_{I, J} (j_{\zeta_i} - i_{\zeta_i})$ , где  $\min_{I, J}$  вычисляется по всем  $I(i_1, \dots, i_n)$  и  $J(j_1, \dots, j_n)$  таким, что операция  $S_{\beta}(J)$  зависит от операции  $S_{\alpha}(I)$ , и

$$\Delta'' = \max \left\{ 0, \max_{\alpha, \beta} \left[ \frac{\left\lceil \frac{-\delta_{\xi}^{(\alpha, \beta)}}{B_{\xi}} \right\rceil + \sum_{k=1}^z \left( \prod_{l=k+1}^z Q_l \right) \left\lceil \frac{-\delta_{\zeta_k}^{(\alpha, \beta)}}{\bar{B}_k} \right\rceil}{\delta^{(\alpha, \beta)} \prod_{l=1}^z Q_l} \right] \right\}.$$

**Теорема 2.3.** *Если  $\Delta'' = 0$  и удовлетворяются условия (2), то зернистые вычисления, задаваемые множествами операций  $V_{\bar{J}, p, \bar{q}}$  можно выполнять параллельно.*

В разделе 2.2 изучаются подходы организации зернистых вычислений для циклов произвольной структуры вложенности.

Для каждого из  $\Theta$  наборов операторов многомерных циклов произвольной структуры вложенности поставим в соответствие набор тесно вложенных циклов, удовлетворяющих условию: любой оператор многомерного цикла окружен ровно одним набором этих тесно вложенных циклов.

Многомерный цикл после применения к нему преобразования тайлинга имеет следующую структуру: параметры циклов  $j_{\zeta}^{gl}$  изменяются в соответствии с неравенствами  $0 \leq j_{\zeta}^{gl} \leq Q_{\zeta}^{\theta} - 1$ , для каждого набора операторов имеется столько локальных циклов, сколько  $r_1^{\theta}, \dots, r_{n^{\theta}}^{\theta}$  превосходят единицу.

Для распределения вычислений по вычислительным процессам будем использовать следующий вид отображения тайлов  $V_{J^{gl}}^{\theta}$  на процессы:

$$\text{Pr}^{\theta}(J^{gl}(j_1^{gl}, \dots, j_{n^{\theta}}^{gl})) = \chi^{\theta} j_{\xi^{\theta}}^{gl} + \delta^{\theta} Q_{\xi^{\theta}}^{\theta} + \gamma^{\theta}, \quad (3)$$

где  $\chi^{\theta} \in \{-1, 1\}$ ,  $\delta^{\theta} \in \{0, 1\}$ ,  $\gamma^{\theta} \in \mathbb{Z}$ . Здесь для каждого  $\theta$  зафиксировано число  $\xi^{\theta}$ ,  $1 \leq \xi^{\theta} \leq n^{\theta}$ .

На практике чаще других используются функции прямого отображения

$$\text{Pr}^{\theta}(J^{gl}(j_1^{gl}, \dots, j_{n^{\theta}}^{gl})) = j_{\xi^{\theta}}^{gl}, (\chi^{\theta} = 1, \delta^{\theta} = 0, \gamma^{\theta} = 0) \quad (4)$$

и функции реверсивного отображения

$$\text{Pr}^{\theta}(J^{gl}(j_1^{gl}, \dots, j_{n^{\theta}}^{gl})) = Q_{\xi^{\theta}}^{\theta} - j_{\xi^{\theta}}^{gl} - 1, (\chi^{\theta} = -1, \delta^{\theta} = 1, \gamma^{\theta} = -1). \quad (5)$$

**Лемма 2.5.**<sup>3</sup> Пусть существует хотя бы одна зависимость  $S_{\alpha}(I) \rightarrow S_{\beta}(J)$ , что в окружении операторов  $S_{\alpha}(I)$  и  $S_{\beta}(J)$  имеется хотя бы один общий разбиваемый цикл и равны параметры всех внешних не разбиваемых циклов:  $(i_1, \dots, i_{k_1-1}) = (j_1, \dots, j_{k_1-1})$ ,  $k_1 \neq 1$ . Тайлинг является допустимым, если для любой такой зависимости выполняются следующие условия:

$$j_{k_{\zeta}} \geq i_{k_{\zeta}}, \quad 1 \leq \zeta \leq c_{\theta\alpha\beta}, \quad (6)$$

$$\beta \geq \alpha. \quad (7)$$

При отсутствии указанных зависимостей тайлинг допустим.

Основным результатом раздела 2.2 является следующая теорема.

**Теорема 2.4.** Если для всех наборов операторов выполнено хотя бы одно из условий:

1) по крайней мере для одного цикла, допускающего разбиение, имеет место  $Q_{\zeta}^{\theta} \gg Q_{\xi^{\theta}}^{\theta}$ ,  $\zeta \neq \xi^{\theta}$ ;

<sup>3</sup>Лиходед, Н. А. Обобщенный тайлинг / Н. А. Лиходед // Доклады НАН Беларуси. – 2011. – Т. 55, № 1. – С. 16–21.

2) в наборе циклов, окружающем набор операторов  $\Theta$ , цикл  $\xi^\theta$  параллельный;

и выполнены условия леммы 2.5, то последовательности зернистых вычислений, задаваемые формулой (3), являются параллельными.

В **третьей главе** представлен метод получения информационных зависимостей уровня макроопераций для алгоритмов с аффинными зависимостями. Сформулированы и доказаны утверждения, позволяющие получить все глобальные зависимости зернистого алгоритма, представленного многомерным циклом с аффинными зависимостями. По заданным функциям, определяющим зависимости на уровне операций, построены функции глобальных зависимостей. Предложен подход к формализации коммуникационных операций алгоритмов, представленных последовательными программами и решена задача генерации (включения) коммуникационных операций в параллельный зернистый алгоритм с однородными информационными зависимостями.

В **разделе 3.1** сформулированы математические основы возникновения коммуникационных операций при реализации параллельных программ на компьютерах с распределенной памятью.

В **разделе 3.2** исследуется задача выявления и записи коммуникационных операций зернистого алгоритма, порождаемые однородными информационными зависимостями. Разработан метод формализации такого вида коммуникационных операций получения и отправки массивов данных в параллельном зернистом алгоритме.

Коммуникационные операции зернистого алгоритма порождаются информационными зависимостями между тайлами. Глобальные (уровня тайлов) зависимости задаются векторами  $\Phi^{sl}$ , которые можно найти согласно результатам работы<sup>4</sup>.

Фиксированный вектор зависимостей  $\Phi$  порождает  $2^{P\Phi}$  глобальных векторов зависимостей; множество этих векторов содержит все различные векторы, у которых  $\zeta$ -я координата равна:

$$\left\lfloor \frac{\Phi_\zeta}{r_\zeta^\theta} \right\rfloor \text{ или } \left\lceil \frac{\Phi_\zeta}{r_\zeta^\theta} \right\rceil, \text{ если } r_\zeta^\theta \neq r_\zeta^{\theta, \max} \text{ и } \frac{\Phi_\zeta}{r_\zeta^\theta} \text{ не является целым числом;}$$

$$\frac{\Phi_\zeta}{r_\zeta^\theta}, \text{ если } r_\zeta^\theta \neq r_\zeta^{\theta, \max} \text{ и } \frac{\Phi_\zeta}{r_\zeta^\theta} \text{ — целое число;}$$

$$0, \text{ если } r_\zeta^\theta = r_\zeta^{\theta, \max}.$$

Множество итераций  $J$  таких, что аргументы выполняемых на них операций являются результатами операций, выполняемых на итерации  $J - \Phi$ , имеет следующий вид:

$$V_{\alpha, \beta} = \left\{ J \in Z^{n^0} \mid \Lambda J + \lambda \leq J \leq P J + \rho \right\},$$

<sup>4</sup>Лиходед, Н. А. Информационная структура зернистых алгоритмов с однородными зависимостями / Н. А. Лиходед, П. И. Соболевский // Доклады НАН Беларуси. – 2011. – Т. 55, № 2. – С. 22–26.

где  $\Lambda$  и  $P$  — нижние треугольные матрицы с нулевыми элементами на главной диагонали,  $\lambda$  и  $\rho$  —  $n^\theta$ -мерные векторы.

Введем в рассмотрение множества

$$V_{J^{gl}, \varphi, \varphi^{gl}}^{in} = \left\{ J \in \mathbb{Z}^{n^\theta} \left| \begin{array}{l} \Lambda J + \lambda \leq J, \\ m^\theta + R^\theta J^{gl} - \min \{0, R^\theta \varphi^{gl} - \varphi\} \leq J, \\ J \leq P J + \rho, \\ J \leq m^\theta + R^\theta (J^{gl} + 1_{n^\theta}) - 1_{n^\theta} - \max \{0, R^\theta \varphi^{gl} - \varphi\} \end{array} \right. \right\}, \quad (8)$$

$$V_{J^{gl}, \varphi, \varphi^{gl}}^{out} = \left\{ J \in \mathbb{Z}^{n^\theta} \left| \begin{array}{l} \Lambda (J + \varphi) + \lambda - \varphi \leq J, \\ m^\theta + R^\theta J^{gl} + \max \{0, R^\theta \varphi^{gl} - \varphi\} \leq J, \\ J \leq P (J + \varphi) + \rho - \varphi, \\ J \leq m^\theta + R^\theta (J^{gl} + 1_{n^\theta}) - 1_{n^\theta} + \min \{0, R^\theta \varphi^{gl} - \varphi\} \end{array} \right. \right\}, \quad (9)$$

где  $1_{n^\theta} = (1, 1, \dots, 1)$  —  $n^\theta$ -мерный вектор с единичными координатами, а матрица  $R^\theta$  диагональная  $\text{diag}(r_1^\theta, \dots, r_{n^\theta}^\theta)$ .

**Теорема 3.1.** Пусть вычислительные процессы заданы функциями  $\text{Pr}^\theta$  вида (4) или (5) и зафиксированы произвольные вектор  $\varphi^{gl}$  и тайл  $J^{gl}$ , относящиеся к  $\theta$ -му набору операторов. Если  $\varphi_\xi^{gl} \neq 0$ , то вектор  $\varphi^{gl}$  порождает коммуникационную операцию получения процессом  $\text{Pr}^\theta(J^{gl})$  массива  $a(\overline{F}_{a, S_{\beta, q}}(J))$ , где  $J \in V_{J^{gl}, \varphi, \varphi^{gl}}^{in}$ , от процесса  $\text{Pr}^\theta(J^{gl} - \varphi^{gl})$ . Если  $\varphi_\xi^{gl} \neq 0$ , то вектор  $\varphi^{gl}$  порождает также коммуникационную операцию отправки процессом  $\text{Pr}^\theta(J^{gl})$  массива  $a(\overline{F}_{a, S_{\alpha, 1}}(J))$ , где  $J \in V_{J^{gl}, \varphi, \varphi^{gl}}^{out}$ , процессу  $\text{Pr}^\theta(J^{gl} + \varphi^{gl})$ .

Если  $\varphi_\xi^{gl} = 0$ , то вектор  $\varphi^{gl}$  коммуникационных операций между процессами не порождает.

В разделе 3.3 для получения функций глобальных зависимостей по аналогии с функциями зависимостей  $\overline{\Phi}_{\alpha, \beta}(J)$  уровня операций, рассматриваются функции глобальных зависимостей в виде

$$I^{gl} = \overline{\Phi}_{\alpha, \beta}^{gl}(J^{gl}) = \Phi_{\alpha, \beta}^{gl} J^{gl} + \Psi_{\alpha, \beta}^{gl} N - \varphi^{\alpha, \beta, gl}, \quad (10)$$

где  $\Phi_{\alpha, \beta}^{gl}$  и  $\Psi_{\alpha, \beta}^{gl}$  — матрицы,  $\varphi^{\alpha, \beta, gl}$  — вектор.

Обозначим:  $\Phi_{\alpha, \beta, +}$  — матрица, получаемая из матрицы  $\Phi_{\alpha, \beta}$  обнулением всех ее отрицательных элементов,  $\Phi_{\alpha, \beta, -}$  — матрица, получаемая из матрицы  $\Phi_{\alpha, \beta}$  обнулением всех ее положительных элементов. Неравенства векторов означают неравенства соответствующих координат векторов.

**Теорема 3.2.** Если вектор  $(R^{\theta^\alpha})^{-1} \Psi_{\alpha, \beta} N$  и матрица  $(R^{\theta^\alpha})^{-1} \Phi_{\alpha, \beta} R^{\theta^\beta}$  являются целочисленными, то аффинная функция зависимостей  $\overline{\Phi}_{\alpha, \beta}(J)$  вида (1) порождает функции глобальных зависимостей вида (10), параметры которых определяются следующим образом:

$$\Phi_{\alpha, \beta}^{gl} = (R^{\theta^\alpha})^{-1} \Phi_{\alpha, \beta} R^{\theta^\beta}, \quad (11)$$

$$\Psi_{\alpha,\beta}^{gl} = \left(R^{\theta^\alpha}\right)^{-1} \Psi_{\alpha,\beta}, \quad (12)$$

$\varphi^{\alpha,\beta,gl}$  — векторы с целыми координатами, удовлетворяющие следующим неравенствам

$$\begin{aligned} & \left(R^{\theta^\alpha}\right)^{-1} \left(m^{\theta^\alpha} + \varphi^{\alpha,\beta} - \Phi_{\alpha,\beta} m^{\theta^\beta} + \Phi_{\alpha,\beta,+} 1_{n_\beta} - \Phi_{\alpha,\beta,+} R^{\theta^\beta} 1_{n_\beta}\right) \leq \varphi^{\alpha,\beta,gl} \leq \\ & \leq \left(R^{\theta^\alpha}\right)^{-1} \left(m^{\theta^\alpha} + \varphi^{\alpha,\beta} - 1_{n_\alpha} + R^{\theta^\alpha} 1_{n_\alpha} - \Phi_{\alpha,\beta} m^{\theta^\beta} - \Phi_{\alpha,\beta,-} R^{\theta^\beta} 1_{n_\beta} + \Phi_{\alpha,\beta,-} 1_{n_\beta}\right) \end{aligned} \quad (13)$$

Одна функция зависимостей может породить столько функций глобальных зависимостей, сколько векторов  $\varphi^{\alpha,\beta,gl}$  удовлетворяет условию (13).

**Раздел 3.4** посвящен анализу коммуникационных операций параллельного алгоритма, порождаемых аффинными зависимостями.

**Четвертая глава** посвящена результатам автора диссертации, связанным с построением метода оценки локальности параллельных алгоритмов.

В **разделе 4.1** сформулированы теоремы позволяющие получить оценку объема пересылаемых данных и количества коммуникационных операций между параллельными вычислительными процессами.

Обозначим через  $e_\zeta^{(n_\beta)}$  вектор-строку размера  $n_\beta$ , у которой координата с номером  $\zeta$  равна 1, а остальные координаты нулевые,  $\rho_{a_l, S_\beta, q} = \text{rank} F_{a_l, S_\beta, q}$ ,

$$\rho_{a_l, S_\beta, q}^\zeta = \text{rank} \begin{pmatrix} F_{a_l, S_\beta, q} \\ e_\zeta^{(n_\beta)} \end{pmatrix}, \quad \rho_{a_l, S_\beta, q}^\Phi = \text{rank} \begin{pmatrix} F_{a_l, S_\beta, q} \\ \Phi_{\alpha,\beta} \end{pmatrix}, \quad \rho_{a_l, S_\beta, q}^{\Phi, \zeta} = \text{rank} \begin{pmatrix} F_{a_l, S_\beta, q} \\ \Phi_{\alpha,\beta} \\ e_\zeta^{(n_\beta)} \end{pmatrix}.$$

**Теорема 4.1.** Пусть определение элемента некоторого массива  $a_l$  происходит на вхождении  $(a_l, S_\alpha, 1)$  в левой части оператора  $S_\alpha$ , а использование — на вхождении  $(a_l, S_\beta, q)$  в правой части оператора  $S_\beta$ ,

$$\begin{aligned} \text{Pr}^\alpha \left( J^{gl}(j_1^{gl}, \dots, j_{n_\alpha}^{gl}) \right) &= j_{\zeta^\alpha}^{gl}, \\ \text{Pr}^\beta \left( J^{gl}(j_1^{gl}, \dots, j_{n_\beta}^{gl}) \right) &= j_{\zeta^\beta}^{gl}. \end{aligned}$$

Тогда, при реализации алгоритма на компьютере с распределенной памятью, объем коммуникационных операций, связанных с вхождением  $(a_l, S_\beta, q)$ , имеет следующие оценки:

1) если выполняются условия

$$(\Phi_{\alpha,\beta})_{\zeta^\alpha} = e_{\zeta^\beta}^{(n_\beta)}, \quad (14)$$

$$(\Psi_{\alpha,\beta})_{\zeta^\alpha} = 0, \quad (15)$$

$$r_{\zeta^\beta}^\beta = r_{\zeta^\alpha}^\alpha \quad (16)$$

и  $d_{\alpha,\beta} = 0$ , то коммуникационных операций не требуется;

2) если выполняются условия (14)–(16) и  $0 < |d_{\alpha,\beta}| < r_{\zeta^\beta}^\beta$ , то процессам требуется получить  $O(Q_{\zeta^\beta}^\beta M^{\rho_{a_l, S_\beta, q}^\Phi}^{-1})$  данных; коммуникационные операции происходят между процессами, номера которых отличаются на 1;

3) если выполняются условия (14)–(16) и  $|d_{\alpha,\beta}| \geq r_{\zeta^\beta}^\beta$ , или выполняются условия (14), (16), но условие (15) не выполняется, то процессам требуется получить  $O(M^{\rho_{a_l, S_\beta, q}^\Phi})$  данных; коммуникационные операции происходят между процессами, номера которых отличаются не менее чем на 1.

В случае, когда вхождение  $(a_l, S_\beta, q)$  не порождает истинной зависимости между итерациями алгоритма (происходит обращение к входным данным) или равенство (14) не выполняется (тогда вычисления требуют групповых коммуникационных операций), оценки следующие:

4) если выполняется условие  $\rho_{a_l, S_\beta, q}^{\Phi, \zeta^\beta} = \rho_{a_l, S_\beta, q}^\Phi$ , то требуется получить данные в количестве  $O(M^{\rho_{a_l, S_\beta, q}^\Phi})$ ;

5) если выполняется условие  $\rho_{a_l, S_\beta, q}^{\Phi, \zeta^\beta} - 1 = \rho_{a_l, S_\beta, q}^\Phi$ , то требуется получить данные в количестве  $O(Q_{\zeta^\beta}^\beta M^{\rho_{a_l, S_\beta, q}^\Phi})$ .

**Теорема 4.2.** Пусть определение элемента некоторого массива  $a_l$  происходит на вхождении  $(a_l, S_\alpha, 1)$  в левой части оператора  $S_\alpha$ , а использование — на вхождении  $(a_l, S_\beta, q)$  в правой части оператора  $S_\beta$  и

$$\text{Pr}^\alpha \left( J^{gl} (j_1^{gl}, \dots, j_{n^\alpha}^{gl}) \right) = j_{\zeta^\alpha}^{gl}, \text{Pr}^\beta \left( J^{gl} (j_1^{gl}, \dots, j_{n^\beta}^{gl}) \right) = Q_{\zeta^\beta}^\beta - j_{\zeta^\beta}^{gl} - 1$$

или

$$\text{Pr}^\alpha \left( J^{gl} (j_1^{gl}, \dots, j_{n^\beta}^{gl}) \right) = Q_{\zeta^\alpha}^\alpha - j_{\zeta^\alpha}^{gl} - 1, \text{Pr}^\beta \left( J^{gl} (j_1^{gl}, \dots, j_{n^\beta}^{gl}) \right) = j_{\zeta^\beta}^{gl}.$$

Тогда, при реализации алгоритма на компьютере с распределенной памятью, объем коммуникационных операций, связанных с вхождением  $(a_l, S_\beta, q)$ , имеет следующие оценки:

1) если выполняются условия

$$(\Phi_{\alpha,\beta})_{\zeta^\alpha} = -e_{\zeta^\beta}^{(n^\beta)}, \quad (17)$$

$$(\Psi_{\alpha,\beta})_{\zeta^\alpha} = M^{\beta, N}(\zeta^\beta), \quad (18)$$

$$r_{\zeta^\beta}^\beta = r_{\zeta^\alpha}^\alpha, Q_{\zeta^\beta}^\beta = Q_{\zeta^\alpha}^\alpha, \quad (19)$$

и  $\bar{d}_{\alpha,\beta} = 0$ , то коммуникационных операций не требуется;

2) если выполняются условия (17)–(19) и  $0 < |\bar{d}_{\alpha,\beta}| < r_{\zeta^\beta}^\beta$ , то процессам требуется получить  $O(Q_{\zeta^\beta}^\beta M^{\rho_{a_l, S_\beta, q}^\Phi}^{-1})$  данных; коммуникационные операции происходят между процессами, номера которых отличаются на 1;



3) если выполняются условия (17)–(19) и  $|\bar{d}_{\alpha,\beta}| \geq r_{\zeta^\beta}^\beta$ , или выполняются условия (17), (19), но условие (18) не выполняется, то процессам требуется получить  $O(M^{\rho_{a_l, S_\beta, q}^\Phi})$  данных; коммуникационные операции происходят между процессами, номера которых отличаются не менее чем на 1.

В случае, когда вхождение  $(a_l, S_\beta, q)$  не порождает истинной зависимости между итерациями алгоритма (происходит обращение к входным данным) или равенство (17) не выполняется (тогда вычисления требуют групповых коммуникационных), оценки следующие:

4) если выполняется условие  $\rho_{a_l, S_\beta, q}^{\Phi, \zeta^\beta} = \rho_{a_l, S_\beta, q}^\Phi$ , то требуется получить данные в количестве  $O(M^{\rho_{a_l, S_\beta, q}^\Phi})$ ;

5) если выполняется условие  $\rho_{a_l, S_\beta, q}^{\Phi, \zeta^\beta} - 1 = \rho_{a_l, S_\beta, q}^\Phi$ , то требуется получить данные в количестве  $O(Q_{\zeta^\beta}^\beta M^{\rho_{a_l, S_\beta, q}^\Phi})$ .

В разделах 4.2 и 4.3 рассматривается параллельный алгоритм метода расщепления для реализации на суперкомпьютерах с распределенной памятью.

Рассматривается трехмерное уравнение теплопроводности в области  $Q_T = G \times I$ ,  $G = [0 < x_1 < l_1] \times [0 < x_2 < l_2] \times [0 < x_3 < l_3]$ ,  $I = [0 < t \leq T]$ :

$$\frac{\partial u}{\partial t} = Lu + f(x_1, x_2, x_3, t), L = \sum_{m=1}^3 L_m, L_m = k \frac{\partial^2 u}{\partial x_m^2}, k = \text{const} > 0, \quad (20)$$

с начальными и краевыми условиями

$$u(x_1, x_2, x_3, 0) = u_0(x_1, x_2, x_3), (x_1, x_2, x_3) \in G, \quad (21)$$

$$u(x_1, x_2, x_3, t) \Big|_{S_t} = \mu(x_1, x_2, x_3, t), (x_1, x_2, x_3, t) \in S_t \equiv \partial Q_T. \quad (22)$$

Введем на отрезке  $[0 < t \leq T]$  сетку узлов

$$\omega_\tau = \{t_j = j\tau, j = 0, 1, \dots, j_0, j_0\tau = T\},$$

в области  $G$  и на ее границе введем сетку узлов

$$\bar{\omega}_h = \left\{ \begin{array}{l} x_1^{(i_1)} = i_1 h_1, \quad i_1 = 0, 1, \dots, N_1, \quad N_1 h_1 = l_1, \\ x_2^{(i_2)} = i_2 h_2, \quad i_2 = 0, 1, \dots, N_2, \quad N_2 h_2 = l_2, \\ x_3^{(i_3)} = i_3 h_3, \quad i_3 = 0, 1, \dots, N_3, \quad N_3 h_3 = l_3 \end{array} \right\}.$$

Вычислительные эксперименты с реализациями 1D и 2D параллельных алгоритмов с улучшенной локальностью согласуются с утверждениями следующих теорем.

**Теорема 4.3.** Пусть  $N_1 - 1 = N_2 - 1 = N_3 - 1 = M$ . Вычисления алгоритма с заменой части естественного параллелизма на конвейерный параллелизм требуют коммуникационных операций только между процессами, номера которых отличаются на 1. Суммарный объем коммуникационных операций, требуемых на одном слое, равен  $5(P - 1)M^2$ .

**Теорема 4.4.** Пусть  $N_1 - 1 = N_2 - 1 = N_3 - 1 = M$ . Существует распределение операций между 2D-процессами с коммуникационными операциями только между процессами, номера которых отличаются по первой или по второй координате на 1. Суммарный объем коммуникационных операций, требуемых на одном слое, равен  $5(P_\eta + P_\xi - 2)M^2$ .

## Заключение

### Основные результаты диссертации

Диссертационная работа посвящена методам построения и оценки эффективности параллельных версий вычислительных алгоритмов для выполнения на многопроцессорных системах с распределенной памятью.

В ходе выполнения диссертационной работы получены следующие результаты:

1. Метод построения параллельных вычислительных процессов на компьютерах с распределенной памятью [1, 2, 9, 12].
2. Метод получения информационной структуры зернистых алгоритмов [5, 13].
3. Формализация коммуникационных операций параллельного алгоритма [3, 4, 6, 14].
4. Метод оценки локальности параллельных алгоритмов, ориентированных на компьютеры с распределенной памятью [7, 8, 10, 11, 15].

### Дополнительные результаты диссертации

Программная реализация метода автоматизации построения параллельных вычислительных процессов.

### Рекомендации по практическому использованию результатов

Диссертация относится к теоретическим исследованиям в области параллельных вычислений и построения вычислительных процессов. Методы построения зернистых алгоритмов и анализа эффективности построенного параллельного алгоритма могут использоваться при решении вычислительно объемных задач и проведении дальнейших исследований.

Результаты диссертации, обобщающие ранее известные методы построения зернистых вычислений, позволяют строить семейства алгоритмов для последующего выбора оптимального размера зерна вычислений. При построении зернистых функций зависимостей обобщаются методы исследования информационной структуры алгоритмов и глобальных векторов зависимостей. Дополнительно введен метод оценки объема коммуникационных операций, что позволяет сравнивать локальность параллельных версий алгоритмов.

Результаты диссертации могут быть использованы при автоматизированной разработке параллельных версий классических алгоритмов вычислитель-

ной математики с высокой степенью утилизации вычислительных ресурсов суперкомпьютеров с распределенной памятью.

Результаты диссертации используются в учебном процессе на факультете прикладной математики и информатики БГУ в курсе «Параллельные и распределённые вычисления». Их также можно использовать при чтении специальных курсов по параллельным вычислениям.

## Список работ, опубликованных автором по теме диссертации

### Статьи:

1. Лиходед, Н. А. Параллельные версии реализации многомерных циклов / Н. А. Лиходед, А. А. Толстикова // Информатика. – 2008. – Т. 15, № 2. – С. 129–136.

2. Лиходед, Н. А. Параллельные последовательности зернистых вычислений / Н. А. Лиходед, А. А. Толстикова // Доклады НАН Беларуси. – 2010. – Т. 54, № 4. – С. 36–41.

3. Лиходед, Н. А. Формализация коммуникационных операций многомерных циклов / Н. А. Лиходед, А. А. Толстикова // Весці НАН Беларусі. Сер. фіз.-мат. навук. – 2010. – № 3. – С. 109–114.

4. Лиходед, Н. А. Коммуникационные операции параллельного алгоритма, порождаемые однородными зависимостями / Н. А. Лиходед, П. И. Соболевский, А. А. Толстикова // Доклады НАН Беларуси – 2011. – Т. 55, № 3. – С. 21–26.

5. Лиходед, Н. А. Функции, задающие зависимости зернистых алгоритмов / Н. А. Лиходед, А. А. Толстикова // Доклады НАН Беларуси. – 2014. – Т. 58, № 4. – С. 35–41.

6. Лиходед, Н. А. Формализованное получение коммуникационных операций параллельных зернистых алгоритмов / Н. А. Лиходед, А. А. Толстикова // Весці НАН Беларусі. Сер. фіз.-мат. навук. – 2018. – Т. 54, № 1. – С. 50–61.

7. Лиходед, Н. А. Метод оценки локальности параллельных алгоритмов, ориентированных на компьютеры с распределенной памятью / Н. А. Лиходед, А. А. Толстикова // Доклады НАН Беларуси. – 2020. – Т. 64, № 6. – С. 647–656.

### Статьи в других научных изданиях:

8. Толстикова, А. А. Подход к оценке локальности зернистых вычислительных процессов, логически организованных в двумерную структуру / А. А. Толстикова, С. В. Баханович, Н. А. Лиходед // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. – 2021. – Т. 10, № 3. – С. 37–55.

## Материалы конференций:

9. Толстикова, А. А. Корректность разбиений алгоритмов при организации зернистых параллельных вычислительных процессов / А. А. Толстикова, Н. А. Лиходед // Международный конгресс по информатике: информационные системы и технологии CSIST'2011, 31 октября – 3 ноября 2011 г., / Минск, Беларусь. Бел. гос. ун-т. – Минск, 2011. – Т. 2. – С. 122–126.

10. Толстикова, А. А. Параллельный алгоритм метода расщепления для реализации на суперкомпьютерах с распределенной памятью / А. А. Толстикова, Н. А. Лиходед // Международная научная конференция «Параллельные вычислительные технологии» (ПаВТ'2020): г. Пермь, 31 марта – 2 апреля 2020 г., / Короткие статьи и описания плакатов. Челябинск: Издательский центр ЮУрГУ, 2020. – С. 287–297.

11. Толстикова, А. А. Построение и исследование локальности параллельного алгоритма, реализующего схему расщепления на 2D структуре процессов / А. А. Толстикова., С. В. Баханович, Н. А. Лиходед // Международная научная конференция «Параллельные вычислительные технологии» (ПаВТ'2021): 30 марта – 1 апреля 2021 г., / Короткие статьи и описания плакатов. Челябинск: Издательский центр ЮУрГУ, 2021. – С. 263–274.

## Тезисы докладов:

12. Толстикова, А. А. Организация вычислительных процессов при параллельной реализации вложенных циклов / А. А. Толстикова, Н. А. Лиходед // Тезисы докладов Международной научной конференции «X Белорусская математическая конференция». Часть 4. Минск, Республика Беларусь, 3–7 ноября 2008 г. – Минск, 2008. – С. 27–28.

13. Толстикова, А. А. Функции, определяющие информационную структуру зернистых алгоритмов / А. А. Толстикова, Н. А. Лиходед // Тезисы докладов Международной научной конференции «XI Белорусская математическая конференция». Часть 3. Минск, Республика Беларусь, 5–8 ноября 2012 г. – Минск, 2012. – С. 23–24.

14. Толстикова, А. А. Получение коммуникационных операций параллельных зернистых алгоритмов / А. А. Толстикова, Н. А. Лиходед // Тезисы докладов Международной научной конференции «XII Белорусская математическая конференция». Часть 3. Минск, Республика Беларусь, 5–10 сентября 2016 г. – Минск, 2016. – С. 11–12.

15. Толстикова, А. А. Метод оценки локальности параллельных зернистых вычислительных процессов / А. А. Толстикова, Н. А. Лиходед // XIII Белорусская математическая конференция: материалы Международной научной конференции, Минск, 22–25 ноября 2021 г.: в 2 ч. / сост. В. В. Лепин; Национальная академия наук Беларуси, Институт математики, Белорусский государственный университет. Минск: Беларуская навука, 2021. Ч. 2. – С. 49–50.

## Рэзюмэ

### Толсцікаў Аляксей Аляксандравіч МЕТАДЫ ПАБУДОВЫ І ДАСЛЕДВАННЯ ПАРАЛЛЕЛЬНЫХ АЛГАРЫТМАЎ ДЛЯ РЭАЛІЗАЦЫІ НА СУПЕРКАМПУТАРАХ З РАЗМЕРКАВАНАЙ ПАМЯЦЦЮ

**Ключавыя словы:** паралельныя вылічэнні, тайлінг, паралельныя паслядоўнасці вылічэнняў, камунікацыйныя аперацыі, глабальныя функцыі залежнасцяў, лакальнасць дадзеных, ацэнка аб'ёму камунікацый.

**Мэта работы** — комплекснае вывучэнне магчымасці пабудовы крупчастых версій вылічальных алгарытмаў для паралельнага выканання, фармалізацыя і аналіз спараджальных камунікацыйных аперацый, а таксама пабудова метаду ацэнкі лакальнасці дадзеных пры рэалізацыі алгарытмаў на вылічальных сістэмах з размеркаванай памяццю.

**Метады даследавання:** аналіз інфармацыйнай структуры алгарытмаў і ўзаемасувязі размеркаванняў аперацый алгарытму і дадзеных паміж вылічальнымі вузламі, пабудова матэматычнай мадэлі прасторы ітэрацый і вывучэнне лінейных адлюстраванняў, задаваных функцыямі залежнасцяў.

У дысертацыі атрыманы наступныя **навуковыя вынікі:**

1. Метад пабудовы паралельных вылічальных працэсаў на кампутарах з размеркаванай памяццю; умовы паралельнасці выканання тайлаў, умовы карэктнасці тайлінгу.

2. Метад атрымання інфармацыйнай структуры крупчастых алгарытмаў з афіннымі залежнасцямі; сцвярджэнні, якія дазваляюць атрымаць усе глабальныя залежнасці крупчастых версій алгарытмаў, прадстаўленых шматмернымі цыкламі з афіннымі залежнасцямі.

3. Падыход да фармалізацыі і ўключэння ў схему паралельнага алгарытму з аднароднымі або афіннымі залежнасцямі камунікацыйных аперацый атрымання і адпраўкі масіваў дадзеных; сцвярджэнні, якія дазваляюць у відавочным выглядзе запісаць новыя камунікацыйныя аперацыі.

4. Ацэнкі камунікацыйных выдаткаў пры рэалізацыі алгарытмаў на паралельных кампутарах з размеркаванай памяццю; выразы, якія характарызуюць лакальнасць паралельнага крупчастага алгарытму: колькасць дадзеных, для якіх патрабуюцца камунікацыі, і лік працэсаў, уцягнутых у перасылкі гэтых дадзеных.

**Рэкамендацыі па практычным выкарыстанні.** Дысертацыя адносіцца да тэарэтычных даследаванняў у вобласці паралельных вылічэнняў і пабудовы вылічальных працэсаў. Метады пабудовы крупчастых алгарытмаў і аналізу эфектыўнасці пабудаванага паралельнага алгарытму могуць выкарыстоўвацца пры вырашэнні вылічальна аб'ёмных задач і правядзенні далейшых даследаванняў.

**Галіна прымянення:** пабудова паралельных алгарытмаў.

# Резюме

## Толстиков Алексей Александрович МЕТОДЫ ПОСТРОЕНИЯ И ИССЛЕДОВАНИЯ ПАРАЛЛЕЛЬНЫХ АЛГОРИТМОВ ДЛЯ РЕАЛИЗАЦИИ НА СУПЕРКОМПЬЮТЕРАХ С РАСПРЕДЕЛЕННОЙ ПАМЯТЬЮ

**Ключевые слова:** параллельные вычисления, тайлинг, параллельные последовательности вычислений, коммуникационные операции, глобальные функции зависимостей, локальность данных, оценка объема коммуникаций.

**Цель работы** — комплексное изучение возможности построения зернистых версий вычислительных алгоритмов для параллельного исполнения, формализация и анализ порождаемых коммуникационных операций, а также построение метода оценки локальности данных при реализации алгоритмов на вычислительных системах с распределенной памятью.

**Методы исследования:** анализ информационной структуры алгоритмов и взаимосвязи распределений операций алгоритма и данных между вычислительными узлами, построение математической модели пространства итераций и изучение линейных отображений, задаваемых функциями зависимостей.

В диссертации получены следующие **научные результаты:**

1. Метод построения параллельных вычислительных процессов на компьютерах с распределенной памятью; условия параллельности выполнения тайлов, условия корректности тайлинга.

2. Метод получения информационной структуры зернистых алгоритмов с аффинными зависимостями; утверждения, позволяющие получить все глобальные зависимости зернистых версий алгоритмов, представленных многомерными циклами с аффинными зависимостями.

3. Подход к формализации и включению в схему параллельного зернистого алгоритма с однородными или аффинными зависимостями коммуникационных операций получения и отправки массивов данных; утверждения, позволяющие в явном виде записать новые коммуникационные операции.

4. Оценки коммуникационных затрат при реализации алгоритмов на параллельных компьютерах с распределенной памятью; выражения, характеризующие локальность параллельного зернистого алгоритма: число данных, для которых требуются коммуникации, и число процессов, вовлеченных в пересылки этих данных.

**Рекомендации по практическому использованию.** Диссертация относится к теоретическим исследованиям в области параллельных вычислений и построения вычислительных процессов. Методы построения зернистых алгоритмов и анализа эффективности построенного параллельного алгоритма могут использоваться при решении вычислительно объемных задач и проведении дальнейших исследований.

**Область применения:** построение параллельных алгоритмов.

# Summary

Tolstikau, Aliaksei A.

## METHODS TO BUILD AND TO RESEARCH PARALLEL ALGORITHMS FOR SUPERCOMPUTERS WITH DISTRIBUTED MEMORY

**Key words:** parallel computing, tiling, parallel sequences of calculations, communication operations, global dependency functions, data locality, estimation of the volume of communications.

The **goal of the work** are a comprehensive study of computational algorithms versions construction for parallel execution, formalization and analysis of generated communication operations, as well as the construction of a method to assess data locality when implementing algorithms for distributed memory computing systems.

The **methods of the investigation** are the analysis of the information structure of algorithms and the relationship of the distributions of algorithm operations and data between computing nodes, construction of a mathematical model of the iteration space and the study of linear maps defined by dependency functions

The following **scientific results** were obtained in the thesis:

1. A method for constructing parallel computing processes on computers with distributed memory; the conditions for parallel execution of tiles, the conditions for tiling to be correct.

2. A method for obtaining the information structure of granular algorithms with affine dependencies; the statements allowing to get all the global dependencies of granular versions of algorithms with affine dependencies.

3. An approach to formalize and to include the communication operations of receiving and sending data arrays; the statements that allow to explicitly record new communication operations.

4. The estimations of communication costs when implementing algorithms on parallel computers with distributed memory; the expressions characterizing the locality of a parallel granular algorithm: the number of data for which communications are required and the number of processes, involved in the transmission of this data.

**Recommendations on practical application of the results.** The dissertation relates to theoretical research in the field of parallel computing and the construction of computational processes. The methods of constructing granular algorithms and analyzing the effectiveness of the constructed parallel algorithm can be used in solving numerous problems and conducting further research.

**Application field:** construction of parallel algorithms.