ІНФАРМАТЫКА

УДК 519.6

П. И. СОБОЛЕВСКИЙ, С. В. БАХАНОВИЧ

ДВУХУРОВНЕВЫЙ ТАЙЛИНГ И ЕГО ПРИМЕНЕНИЕ ПРИ ПРОСТРАНСТВЕННО-ВРЕМЕННОМ ОТОБРАЖЕНИИ АЛГОРИТМОВ НА ПАРАЛЛЕЛЬНЫЕ АРХИТЕКТУРЫ

Институт математики Национальной академии наук Беларуси, Минск, Беларусь,
e-mail: sobolevsky@im.bas-net.by, bsv@im.bas-net.by

Предложена идея использования двухуровневого тайлинга для решения задачи пространственно-временного отображения алгоритмов на параллельные вычислительные системы заданной размерности и размера. Разработан формализованный метод параметризованного двухуровневого тайлинга. Получены формулы для определения векторов глобальных зависимостей между тайлами первого и второго уровней, а также формальное представление множеств итераций, порождающих эти зависимости, в виде многогранников с явным выражением их границ. На основе разработанного метода двухуровневого тайлинга и локально параллельной глобально последовательной стратегии отображения предложен формализованный метод решения задачи пространственно-временного отображения алгоритмов в интеграции с тайлингом.

Ключевые слова: тайлинг, пространственно-временное отображение, локально параллельная глобально последовательная стратегия, параллельная архитектура, суперкомпьютер.

P. I. SOBOLEVSKY, S. V. BAKHANOVICH

TWO-LEVEL TILING AND ITS APPLICATION IN THE SPACE-TIME MAPPING OF ALGORITHMS ONTO PARALLEL ARCHITECTURES

Institute of Mathematics of the National Academy of Sciences of Belarus, Minsk, Belarus,
e-mail: sobolevsky@im.bas-net.by, bsv@im.bas-net.by

The idea of application of two-level tiling at space-time mapping of algorithms onto parallel computing systems is proposed. A formalized method of parameterized two-level tiling is developed. The formulas for determination of global dependences between different-level tiles are constructed. The formal representation of sets of iterations generating these dependences is obtained. The representation is given in the form of polyhedra with an explicit expression of their boundaries. A formalized method of space-time mapping onto supercomputers is developed. The method is based on the locally parallel globally sequential mapping strategy and the designed method of two-level tiling. The method realizes the proposed idea of space-time mapping in integration with tiling.

Keywords: tiling, space-time mapping, locally sequential globally parallel strategy, parallel architecture, supercomputer.

Введение. Эффективность параллельных программ зависит от ряда факторов. Очевидно, что прежде всего она зависит от выбора алгоритма решения задачи и его распараллеливания. Но не меньшее значение имеет и то, насколько при разработке и программной реализации параллельного алгоритма учитываются особенности архитектуры суперкомпьютера.

На практике при разработке программных продуктов широко используется тайлинг, который является одним из наиболее результативных средств оптимизации программ. Тайлинг позволяет значительно рациональнее использовать многоуровневую память компьютера и, кроме этого, помогает оптимизировать операции обмена данными в параллельных приложениях для вычислительных систем с распределенной памятью.

Высокая эффективность тайлинга привела к его интенсивному исследованию и появлению большого числа работ, посвященных его различным теоретическим и прикладным аспектам [1–6]. В настоящее время существует ряд методов тайлинга, дающих возможность оптимизировать как последовательные, так и параллельные программы. Тем не менее тема тайлинга продолжает активно развиваться, и одним из перспективных направлений является идея многоуровневого тайлинга [7]. Многоуровневый тайлинг в сравнении с одноуровневым тайлингом в большей степени может увеличить эффективность использования многоуровневой памяти компьютера. Кроме того, тайлинг такого типа может быть использован для решения и других задач.

В данной работе предлагается идея использования двухуровневого тайлинга для решения задачи пространственно-временного отображения алгоритмов на параллельные вычислительные системы заданной размерности и фиксированного размера. Суть идеи заключается в интеграции локально параллельной глобально последовательной стратегии отображения алгоритмов [8] и техники двухуровневого тайлинга с целью комплексного решения задач пространственно-временного отображения и тайлинга.

Изначально суть тайлинга состоит в увеличении зернистости алгоритма, множество операций которого разбивается на группы-тайлы, каждый тайл рассматривается как зерно вычислений или макрооперация. Поскольку тайл является атомарным, то при распределении операций алгоритма между процессорами суперкомпьютера операции, принадлежащие одному и тому же тайлу, отображаются на один и тот же процессор. Зависимости между операциями алгоритма порождают глобальные зависимости между тайлами, которые, в свою очередь, определяют зависимости между процессорами при отображении тайлов на параллельные архитектуры.

Для реализации идеи отображения алгоритмов в интеграции с двухуровневым тайлингом в работе предложен необходимый математический аппарат: дано формальное определение тайлов первого и второго уровней, исследована задача построения векторов глобальных зависимостей между тайлами фиксированного уровня, получены формулы, определяющие зависимости между тайлами. Кроме этого построены формулы, которые позволяют определить зависимости между процессорами при отображении тайлов на параллельную архитектуру. Предложенный математический аппарат является развитием формализованной методики тайлинга, представленной в работе [6]. На основе разработанного метода двухуровневого тайлинга и локально параллельной глобально последовательной стратегии отображения (LPGS-стратегии) построен формализованный метод решения задачи пространственно-временного отображения алгоритмов в интеграции с тайлингом.

Тайлинг. Тайлинг (tiling) – это преобразование вычислительного алгоритма с целью укрупнения зернистости путем объединения определенного числа мелких операций алгоритма в крупные зерна вычислений – тайлы. При тайлинге укрупнение зернистости осуществляется покрытием области вычислений однотипными n-мерными параллелепипедами (тайлами) [1–6].

Будем предполагать, что область вычислений алгоритма (индексное множество) – это n-мерный выпуклый многогранник, описание которого в декартовой системе координат в арифметическом пространстве Z^n имеет вид

$$V = \left\{ J \in Z^n \mid LJ \geq I \right\}, \tag{1}$$

где целочисленная матрица $L \in Z^{m \times n}$ и целочисленный вектор $I \in Z^m$ являются параметрами, определяющими область вычислений. Каждой точке области вычислений V соответствует ограниченный набор операций алгоритма. В данной работе все векторные равенства и неравенства понимаются в обычном смысле (выполняются для каждой координаты).

Кроме области вычислений алгоритмы характеризуются множеством векторов зависимостей Φ. Данное множество отражает информационные зависимости между операциями алгоритма. Наличие вектора $\varphi \in \Phi$ означает, что в области вычислений V существуют точки J и $J + \varphi$ такие, что операции, приписанные точке $J + \varphi$, информационно зависят от результатов операций, приписанных точке J.

Разбиение арифметического пространства Z^n и, следовательно, области вычислений алгоритма на тайлы осуществляется нелинейным отображением $Z^n \xrightarrow{f} Z^n$:

$$J^{gl} = f(J) = \left\lfloor R^{-1}H(J - J^{\vec{0}}) \right\rfloor,\ J \in Z^n. \tag{2}$$

Отображение f определяется тремя параметрами $H \in Z^{n\times n}$, $R \in Z^{n\times n}$, $J^{\vec{0}} \in Z^n$ и позволяет идентифицировать тайлы вектором $J^{gl} \in Z^n$. Параметры тайлинга *R*, *H* и $J^{\vec{0}}$ имеют геометрический смысл и определены следующим образом.

1. H – нижняя треугольная матрица с единицами на главной диагонали; составлена построчно из координат нормальных векторов $h_1 = (1,0,\ldots,0)$, $h_k = (h_{k1}, h_{k2},\ldots,h_{kk-1},1,0,\ldots,0) \in Z^n$, $k = 2,3,\ldots,n$, гиперплоскостей, проходящих через точки пространства Z^n. Векторы h_k, $k = 1,2,\ldots,n$, определяют параллелепипеидальную форму тайла.

2. $R = \mathrm{diag}(r_1, r_2,\ldots,r_n)$ – диагональная матрица, определяющая размеры тайла, где r_k – количество различных параллельных гиперплоскостей с нормальными векторами h_k, $k = 1,2,\ldots,n$, которые проходят через все целочисленные точки одного тайла. Обозначим дополнительно $\vec{R} = (r_1, r_2,\ldots,r_n)$, $\vec{1} = (1,1,\ldots,1)$.

3. Точку тайла J^{gl}, в которой достигается наименьшее значение функции HJ, будем называть начальной точкой тайла и обозначать $J^{\vec{0},gl}$. В формуле (2) целочисленный вектор $J^{\vec{0}} \in Z^n$ определяет начальную точку нулевого тайла $J^{gl} = \vec{0}$. Если точка $J^{\vec{0}}$ определена, то начальная точка $J^{\vec{0},gl}$ любого тайла $J^{gl} \in Z^n$ определяется условием $RJ^{gl} = H(J^{\vec{0},gl} - J^{\vec{0}})$ и равна $J^{\vec{0},gl} = J^{\vec{0}} + H^{-1}RJ^{gl}$.

Каждый тайл, идентифицированный вектором $J^{gl} \in Z^n$, состоит из $\prod_{m=1}^{n} r_m$ точек с целочисленными координатами. Обозначим множество этих точек $T(J^{gl})$,

$$T(J^{gl}) = \left\{ J \in Z^n \mid J^{gl} = \left\lfloor R^{-1}H(J - J^{\vec{0}}) \right\rfloor \right\} = \left\{ J \in Z^n \mid \vec{0} \le H\left(J - J^{\vec{0},gl}\right) \le \vec{R} - \vec{1} \right\}, \tag{3}$$

а множество точек из области *V*, принадлежащих тайлу $J^{gl} \in Z^n$, обозначим $V^{loc}(J^{gl})$,

$$V^{loc}(J^{gl}) = T(J^{gl}) \cap V = \left\{ J \in Z^n \mid \vec{0} \le H\left(J - J^{\vec{0},gl}\right) \le \vec{R} - \vec{1},\ J \in V \right\}. \tag{4}$$

О п р е д е л е н и е 1 [6]. Тайл J^{gl} будем называть пустым, если $V^{loc}(J^{gl}) = \varnothing$, неполным, если $V^{loc}(J^{gl}) \ne T(J^{gl})$, и будем называть полным, если $V^{loc}(J^{gl}) = T(J^{gl})$.

Из приведенных определений следует, что множество тайлов, имеющих с областью *V* непустое пересечение (множество тайлов, покрывающих область вычислений), которое будем обозначать V^{gl}, можно представить в виде

$$V^{gl} = \left\{ J^{gl} \in Z^n \mid T(J^{gl}) \cap V \ne \varnothing \right\} = \left\{ J^{gl} \in Z^n \mid J^{gl} = \left\lfloor R^{-1}H(J - J^{\vec{0}}) \right\rfloor,\ J \in V \right\}. \tag{5}$$

Согласно концепции тайлинга, непустой тайл рассматривается как новое укрупненное зерно вычислений или макрооперация, которое обладает свойством атомарности. В результате применения тайлинга алгоритм преобразуется к блочному виду (блок – макрооперация) и выполняется как последовательность макроопераций по двухуровневой схеме: внешний уровень – это уровень макроопераций, внутренний – уровень операций в пределах тайла. При параллельной реализации алгоритма операции тайла, в силу его атомарности, отображаются на один и тот же процессор, при этом все необходимые обмены данными выполняются либо до выполнения макрооперации (прием начальных данных), либо после нее (передача результатов).

Информационные зависимости между операциями алгоритма порождают глобальные зависимости между тайлами. Поскольку тайлы являются атомарными, глобальные зависимости между тайлами не должны быть взаимообратными. Данное требование формализуется в виде условия на матрицу H [1]:

$$H\varphi \geq \vec{0}, \quad \varphi \in \Phi. \tag{6}$$

Условие (6) является условием корректности тайлинга, которое необходимо учитывать при выборе параметра H.

Двухуровневый тайлинг. Метод двухуровневого тайлинга, представленный в данной работе, предназначен для алгоритмов с однородными зависимостями. Решение задачи поиска глобальных зависимостей между тайлами и определения множеств информационно зависимых операций, ассоциированных с глобальными зависимостями, выполнено только для алгоритмов такого типа. Кроме этого, метод предполагает разбиение области вычислений алгоритма только на полные тайлы, что накладывает некоторые ограничения на форму и размеры области вычислений. В случае невырожденного тайлинга область вычислений должна представлять собой n-мерный параллелепипед.

Пусть область вычислений – n-мерный параллелепипед вида

$$V = \left\{ J \in \boldsymbol{Z}^n \mid m \leq HJ \leq M \right\} = \left\{ J \in \boldsymbol{Z}^n \mid m_k \leq h_k \cdot J \leq M_k,\, k = 1,2,\ldots,n \right\}, \tag{7}$$

где матрица H удовлетворяет условию корректности тайлинга (6). Матрицу H будем использовать в качестве параметра, определяющего форму тайла, а параметры тайлинга R и $J^{\vec{0}}$ выбирать исходя из условия целочисленности векторов $R^{-1}\left(m - HJ^{\vec{0}}\right)$ и $R^{-1}(M - m + \vec{1})$, а также условия $\vec{R} \leq \vec{R}^{\max}$, $r_k^{\max} = M_k - m_k + 1$, $k = 1,2,\ldots,n$.

Для областей вида (7) в работе [6] получено явное представление множества тайлов

$$V^{gl} = \left\{ J^{gl} \in Z^n \mid \left\lfloor R^{-1}(m - HJ^{\vec{0}}) \right\rfloor \leq J^{gl} \leq \left\lfloor R^{-1}(M - HJ^{\vec{0}}) \right\rfloor \right\},$$

которое с учетом целочисленности векторов $R^{-1}\left(m - HJ^{\vec{0}}\right)$ и $R^{-1}(M - m + \vec{1})$ принимает вид

$$V^{gl} = \left\{ J^{gl} \in Z^n \mid \vec{0} \leq J^{gl} - R^{-1}(m - HJ^{\vec{0}}) \leq R^{-1}\vec{R}^{\max} - \vec{1} \right\} \tag{8}$$

и состоит только из полных тайлов, т. е. $V^{loc}\left(J^{gl}\right) = T\left(J^{gl}\right)$, $J^{gl} \in V^{gl}$.

Условие целочисленности векторов $R^{-1}\left(m - HJ^{\vec{0}}\right)$ и $R^{-1}(M - m + \vec{1})$ может быть удовлетворено, например, если положить $J^{\vec{0}} = H^{-1}m$ и выбрать в качестве диагональных элементов матрицы R делители соответствующих координат вектора $\vec{R}^{\max} = M - m + \vec{1}$. В этом случае

$$V^{gl} = \left\{ J^{gl} \in Z^n \mid \vec{0} \leq J^{gl} \leq R^{-1}\vec{R}^{\max} - \vec{1} \right\}. \tag{9}$$

Далее мы будем предполагать именно такой выбор параметров тайлинга.

Тайлы, полученные при разбиении области вычислений, будем называть тайлами первого уровня. Каждый тайл первого уровня представляет собой n-мерный параллелепипед, который может быть представлен по аналогии с определением области вычислений (7) в виде

$$\begin{gathered} T\left(J^{gl}\right) = \left\{ J \in Z^n \mid \vec{0} \leq H\left(J - J^{\vec{0},gl}\right) \leq \vec{R} - \vec{1} \right\} = \left\{ J \in Z^n \mid \tilde{m} \leq HJ \leq \tilde{M} \right\}, \\ \tilde{m} = m + RJ^{gl},\ \tilde{M} = m + RJ^{gl} + \vec{R} - \vec{1}. \end{gathered} \tag{10}$$

В соответствии с идеей многоуровневого тайлинга, каждый тайл первого уровня разбивается на множество тайлов следующего, второго, уровня. При определенном выборе параметров тайлинга второго уровня, которые будем обозначать $\tilde{H}$, $\tilde{R}$ и $\tilde{J}^{\vec{0}}$, можно осуществить покрытие каждого тайла первого уровня J^{gl} полными тайлами второго уровня. Для этого достаточно определить параметры тайлинга второго уровня, положив $\tilde{H} = H,\ \tilde{J}^{\vec{0}} = H^{-1}\tilde{m} = H^{-1}\left(m + RJ^{gl}\right) = J^{\vec{0},gl}$, а в качестве диагональных элементов $\tilde{r}_k$, $k = 1,2,\ldots,n,$ матрицы $\tilde{R}$, определяющей размеры тайла второго уровня, взять делители соответствующих координат $\tilde{r}_k^{\max}$ вектора $\vec{\tilde{R}}^{\max} = \tilde{M} - \tilde{m} + \vec{1} = \vec{R}$. При таком выборе параметров тайлинга получим множество тайлов второго уровня, состоящее только из полных тайлов.

Тайл второго уровня будем идентифицировать вектором $\tilde{J}^{gl}$ в привязке к тайлу первого уровня J^{gl} , которому этот тайл принадлежит, или, что то же самое, $2n$-мерным вектором $\left(\tilde{J}^{gl}, J^{gl}\right) \in Z^{2n}.$ Множество точек, принадлежащих тайлу второго уровня, имеет вид

$$T\left(\tilde{J}^{gl}, J^{gl}\right) = \left\{J \in Z^n \mid \vec{0} \le H\left(J - \tilde{J}^{\vec{0},gl}\right) \le \vec{\tilde{R}} - \vec{1}\right\} =$$
$$= \left\{J \in Z^n \mid \tilde{R}\tilde{J}^{gl} + RJ^{gl} + m \le HJ \le \tilde{R}\tilde{J}^{gl} + RJ^{gl} + m + \vec{\tilde{R}} - \vec{1}\right\},\ \tilde{J}^{\vec{0},gl} = H^{-1}\left(\tilde{R}\tilde{J}^{gl} + RJ^{gl} + m\right). \tag{11}$$

С учетом сделанных предположений о параметрах тайлинга, множество тайлов второго уровня, принадлежащих тайлу $J^{gl} \in V^{gl}$, может быть представлено в виде

$$\tilde{V}^{gl}\left(J^{gl}\right) = \left\{\left(\tilde{J}^{gl}, J^{gl}\right) \in Z^{2n} \mid \vec{0} \le \tilde{J}^{gl} \le \tilde{R}^{-1}\vec{R} - \vec{1}\right\}. \tag{12}$$

Значения векторов $\tilde{J}^{gl}$ в определении множества $\tilde{V}^{gl}\left(J^{gl}\right)$ зависят только от размеров тайлов первого и второго уровней. Это является следствием принятого выше выбора параметров тайлинга. Формально тайлинг второго уровня выполняется независимо для каждого тайла в отдельности, поэтому в общем случае значения векторов $\tilde{J}^{gl}$ могут зависеть от идентификатора J^{gl} соответствующего тайла первого уровня и от ассоциированных с ним параметров тайлинга второго уровня, в частности от выбора начальной точки $\tilde{J}^{\vec{0}}$.

Глобальные зависимости между тайлами. Наличие вектора $\varphi \in \Phi$ для алгоритма с однородными зависимостями означает, что для любой пары точек $J, J + \varphi \in V$ операции, приписанные точке $J + \varphi,$ информационно зависят от операций, приписанных точке J. При разбиении области вычислений на тайлы точки с информационно зависимыми операциями могут оказаться в двух разных тайлах – в таком случае между этими тайлами возникает информационная зависимость. Таким образом, существование информационной зависимости между операциями алгоритма обусловливает появление глобальной зависимости между тайлами.

Глобальные зависимости фактически определяют порядок выполнения операций алгоритма на уровне макроопераций. Кроме этого, при отображении макроопераций алгоритма на процессоры параллельной архитектуры глобальные зависимости естественным образом определяют коммуникации между процессорами. С практической точки зрения важно не только установить глобальные зависимости между тайлами, но и определить подмножества точек тайлов, которые обусловливают наличие этих зависимостей. Эти множества необходимы для определения конкретных данных, участвующих в коммуникации между процессорами, и определения их количества (коммуникационного объема).

Зависимости между тайлами можно характеризовать векторами глобальных зависимостей соответствующего уровня. Каждый вектор $\varphi \in \Phi$ может порождать некоторое количество векторов глобальных зависимостей разных уровней.

О п р е д е л е н и е 2. Вектор $\varphi^{gl} \in Z^n$ будем называть вектором глобальной зависимости первого уровня, порожденным вектором $\varphi \in \Phi,$ если существуют точки $J, J + \varphi \in V$ и тайлы $J^{gl}, J^{gl} + \varphi^{gl} \in V^{gl}$ такие, что $J \in T\left(J^{gl}\right),\ J + \varphi \in T\left(J^{gl} + \varphi^{gl}\right).$

Приведенное определение векторов глобальных зависимостей формально включает в себя случай, когда точки с информационно зависимыми операциями принадлежат одному и тому же тайлу и, как следствие, порождается нулевой вектор глобальной зависимости. Нулевой вектор зависимостей не имеет значения при организации взаимодействия между тайлами текущего первого уровня, но имеет существенное значение с точки зрения определения зависимостей между тайлами следующего, второго, уровня.

Множество векторов глобальных зависимостей первого уровня, порождаемых вектором $\varphi \in \Phi$, будем обозначать Φ_{φ}^{gl}. Очевидно, что имеется непосредственная связь между значениями векторов $\varphi \in \Phi$ и значениями векторов φ^{gl}. В следующем ниже утверждении сформулировано точное определение координат векторов глобальных зависимостей, а также соответствующих им множеств точек области вычислений, которые обусловливают наличие этих зависимостей. Утверждение фактически является компиляцией результатов, полученных ранее авторами данной работы. Результаты представлены вместе с доказательствами в работе [6].

Перед формулировкой утверждения необходимо ввести ряд обозначений и определений.

Обозначим x_+ (x_-) вектор, полученный из вектора $x \in Z^n$ заменой всех его отрицательных (положительных) координат нулями. В соответствии с этим множество точек $J \in V$ n-мерного параллелепипеда (7), для которых выполняется условие $J + \alpha \in V$, $\alpha \in Z^n$, можно определить как $V^{\alpha} = \left\{ J \in \boldsymbol{Z}^n \mid m - (H\alpha)_- \le HJ \le M - (H\alpha)_+ \right\}$.

Для каждого тайла $J^{gl} \in V^{gl}$, вектора $\varphi \in \Phi$ и порождаемого им вектора $\varphi^{gl} \in \Phi_{\varphi}^{gl}$ определим следующие множества:

$V^{loc}\left(J^{gl}\right)_{\varphi,\varphi^{gl}}^{in} = \left\{ J \in V^{-\varphi} \,\middle|\, J \in T\left(J^{gl}\right),\ J - \varphi \in T\left(J^{gl} - \varphi^{gl}\right) \right\}$ – множество точек тайла J^{gl}, принадлежащих области вычислений алгоритма и информационно зависимых от точек области вычислений, принадлежащих тайлу $J^{gl} - \varphi^{gl}$;

$V^{loc}\left(J^{gl}\right)_{\varphi,\varphi^{gl}}^{out} = \left\{ J \in V^{\varphi} \,\middle|\, J \in T\left(J^{gl}\right),\ J + \varphi \in T\left(J^{gl} + \varphi^{gl}\right) \right\}$ – множество точек области вычислений, принадлежащих тайлу J^{gl}, операции которых непосредственно влияют на результат выполнения операций, приписанных точкам тайла $J^{gl} + \varphi^{gl}$.

Утверждение 1. *Пусть область вычислений алгоритма с однородными зависимостями представлена в виде* (7)*, а параметры тайлинга H, R и* $J^{\vec{0}}$ *удовлетворяют условиям* $J^{\vec{0}} = H^{-1}m$, $R^{-1}\vec{R}^{\max} \in Z^n$. *Тогда векторы глобальных зависимостей первого уровня* $\varphi^{gl} \in \Phi_{\varphi}^{gl}$, *порожденные вектором* $\varphi \in \Phi$, *принимают значения, удовлетворяющие неравенству*

$$\left\lfloor R^{-1}H\varphi \right\rfloor \le \varphi^{gl} \le \min\left(\left\lceil R^{-1}H\varphi \right\rceil, R^{-1}\vec{R}^{\max} - \vec{1} \right).$$

Глобальная зависимость между тайлами $J^{gl} \in V^{gl}$ и $J^{gl} + \varphi^{gl} \in V^{gl}$, *порожденная вектором* φ, *обусловлена информационной зависимостью операций, которые приписаны точкам множеств*

$$V^{loc}\left(J^{gl}\right)_{\varphi,\varphi^{gl}}^{out} = \left\{ J \in V^{\varphi} \,\middle|\, \left(R\varphi^{gl} - H\varphi\right)_+ \le H\left(J - J^{\vec{0},gl}\right) \le \vec{R} - \vec{1} + \left(R\varphi^{gl} - H\varphi\right)_- \right\},$$

$$V^{loc}\left(J^{gl} + \varphi^{gl}\right)_{\varphi,\varphi^{gl}}^{in} = \left\{ J \in V^{-\varphi} \,\middle|\, \left(R\varphi^{gl} - H\varphi\right)_+ \le H\left(J - \varphi - J^{\vec{0},gl}\right) \le \vec{R} - \vec{1} + \left(R\varphi^{gl} - H\varphi\right)_- \right\},$$

где $J^{\vec{0},gl}$ *– начальная точка тайла* J^{gl}. *Количество точек области вычислений, принадлежащих этим множествам, определяется формулой*

$$\left| V^{loc}\left(J^{gl}\right)_{\varphi,\varphi^{gl}}^{out} \right| = \left| V^{loc}\left(J^{gl} + \varphi^{gl}\right)_{\varphi,\varphi^{gl}}^{in} \right| = \prod_{i=1}^{n} \left(r_i - \left| r_i \varphi_i^{gl} - h_i \varphi \right| \right).$$

Очевидно, что векторы зависимостей $\varphi \in \Phi$ порождают зависимости не только между тайлами первого, но и второго уровня.

О п р е д е л е н и е 3. Пусть вектор $\varphi \in \Phi$ – фиксированный вектор зависимостей, а $\varphi^{gl} \in \Phi_{\varphi}^{gl}$ – порождаемый им вектор глобальной зависимости первого уровня. Вектор $\psi^{gl} = \left(\tilde{\varphi}^{gl}, \varphi^{gl}\right) \in Z^{2n}$ будем называть вектором глобальной зависимости второго уровня, порожденным векторами φ и φ^{gl}, если существуют точки J и $J + \varphi$ области вычислений V, принадлежащие соответственно тайлам $\left(\tilde{J}^{gl}, J^{gl}\right) \in \tilde{V}^{gl}\left(J^{gl}\right)$ и $\left(\tilde{J}^{gl} + \tilde{\varphi}^{gl}, J^{gl} + \varphi^{gl}\right) \in \tilde{V}^{gl}\left(J^{gl} + \varphi^{gl}\right)$.

Множество векторов глобальных зависимостей второго уровня, порождаемых парой векторов $\varphi \in \Phi$ и $\varphi^{gl} \in \Phi_{\varphi}^{gl}$, будем обозначать $\Psi_{\varphi,\varphi^{gl}}^{gl}$.

В следующем ниже утверждении отражена связь между значениями векторов $\varphi \in \Phi$, $\varphi^{gl} \in \Phi_{\varphi}^{gl}$ и $\psi^{gl} \in \Psi_{\varphi,\varphi^{gl}}^{gl}$, а также явно определены множества точек области вычислений, которые обусловливают наличие глобальных зависимостей второго уровня. Предварительно введем обозначения и определения, необходимые для формулировки утверждения.

Для каждого тайла второго уровня $\left(\tilde{J}^{gl}, J^{gl}\right) \in \tilde{V}^{gl}\left(J^{gl}\right)$, $J^{gl} \in V^{gl}$, вектора $\varphi \in \Phi$ и порождаемых им векторов $\varphi^{gl} \in \Phi_{\varphi}^{gl}$ и $\psi^{gl} \in \Psi_{\varphi,\varphi^{gl}}^{gl}$ определим следующие множества:

$V^{loc}\left(\tilde{J}^{gl}, J^{gl}\right)_{\varphi,\psi^{gl}}^{in} = \left\{ J \in V^{-\varphi} \,\middle|\, J \in T\left(\tilde{J}^{gl}, J^{gl}\right),\ J - \varphi \in T\left(\tilde{J}^{gl} - \tilde{\varphi}^{gl}, J^{gl} - \varphi^{gl}\right) \right\}$ – множество точек тайла $\left(\tilde{J}^{gl}, J^{gl}\right)$, принадлежащих области вычислений алгоритма и информационно зависимых от точек области вычислений, принадлежащих тайлу $\left(\tilde{J}^{gl} - \tilde{\varphi}^{gl}, J^{gl} - \varphi^{gl}\right)$;

$V^{loc}\left(\tilde{J}^{gl}, J^{gl}\right)_{\varphi,\psi^{gl}}^{out} = \left\{ J \in V^{\varphi} \,\middle|\, J \in T\left(\tilde{J}^{gl}, J^{gl}\right),\ J + \varphi \in T\left(\tilde{J}^{gl} + \tilde{\varphi}^{gl}, J^{gl} + \varphi^{gl}\right) \right\}$ – множество точек области вычислений, принадлежащих тайлу $\left(\tilde{J}^{gl}, J^{gl}\right)$, операции которых непосредственно влияют на результат выполнения операций, приписанным точкам тайла $\left(\tilde{J}^{gl} + \tilde{\varphi}^{gl}, J^{gl} + \varphi^{gl}\right)$.

У т в е р ж д е н и е 2. *Пусть область вычислений алгоритма с однородными зависимостями представлена в виде* (7)*, а параметры тайлинга первого и второго уровня удовлетворяют условиям* $J^{\vec{0}} = H^{-1}m$, $R^{-1}\vec{R}^{\max} \in Z^n$, $\tilde{J}^{\vec{0}} = H^{-1}\left(m + RJ^{gl}\right)$, $\tilde{R}^{-1}\vec{\tilde{R}} \in Z^n$. *Тогда векторы глобальных зависимостей второго уровня* $\psi^{gl} \in \Psi_{\varphi,\varphi^{gl}}^{gl}$, *порожденные парой векторов* $\varphi \in \Phi$ *и* $\varphi^{gl} \in \Phi_{\varphi}^{gl}$, *принимают значения, удовлетворяющие неравенству*

$$\max\left(\left\lfloor \tilde{R}^{-1}H\varphi \right\rfloor - \tilde{R}^{-1}R\varphi^{gl},\ -\tilde{R}^{-1}\vec{R} + \vec{1}\right) \le \tilde{\varphi}^{gl} \le \min\left(\left\lceil \tilde{R}^{-1}H\varphi \right\rceil - \tilde{R}^{-1}R\varphi^{gl},\ \tilde{R}^{-1}\vec{R} - \vec{1}\right).$$

Глобальная зависимость между тайлами $\left(\tilde{J}^{gl}, J^{gl}\right) \in \tilde{V}^{gl}\left(J^{gl}\right)$ *и* $\left(\tilde{J}^{gl} + \tilde{\varphi}^{gl}, J^{gl} + \varphi^{gl}\right) \in \tilde{V}^{gl}\left(J^{gl} + \varphi^{gl}\right)$, *порожденная вектором* φ, *обусловлена информационной зависимостью операций, соответствующих точкам множеств* $V^{out} = V^{loc}\left(\tilde{J}^{gl}, J^{gl}\right)_{\varphi,\psi^{gl}}^{out}$ *и* $V^{in} = V^{loc}\left(\tilde{J}^{gl} + \tilde{\varphi}^{gl}, J^{gl} + \varphi^{gl}\right)_{\varphi,\psi^{gl}}^{in}$:

$$V^{out} = \left\{ J \in V^{\varphi} \,\middle|\, \left(\tilde{R}\tilde{\varphi}^{gl} + R\varphi^{gl} - H\varphi\right)_+ \le H\left(J - \tilde{J}^{\vec{0},gl}\right) \le \vec{\tilde{R}} - \vec{1} + \left(\tilde{R}\tilde{\varphi}^{gl} + R\varphi^{gl} - H\varphi\right)_- \right\},$$

$$V^{in} = \left\{ J \in V^{-\varphi} \,\middle|\, \left(\tilde{R}\tilde{\varphi}^{gl} + R\varphi^{gl} - H\varphi\right)_+ \le H\left(J - \varphi - \tilde{J}^{\vec{0},gl}\right) \le \vec{\tilde{R}} - \vec{1} + \left(\tilde{R}\tilde{\varphi}^{gl} + R\varphi^{gl} - H\varphi\right)_- \right\},$$

где $\tilde{J}^{\vec{0},gl}$ *– начальная точка тайла* $\left(\tilde{J}^{gl}, J^{gl}\right)$. *Количество точек области вычислений, принадлежащих этим множествам, определяется формулой*

$$\left| V^{loc}\left(\tilde{J}^{gl}, J^{gl}\right)_{\varphi,\psi^{gl}}^{out} \right| = \left| V^{loc}\left(\tilde{J}^{gl} + \tilde{\varphi}^{gl}, J^{gl} + \varphi^{gl}\right)_{\varphi,\psi^{gl}}^{in} \right| = \prod_{i=1}^{n}\left(\tilde{r}_i - \left|\tilde{r}_i\tilde{\varphi}_i^{gl} + r_i\varphi_i^{gl} - h_i\varphi\right|\right).$$

Д о к а з а т е л ь с т в о. Пусть $\psi^{gl} \in \Psi^{gl}_{\varphi,\varphi^{gl}}$ – вектор глобальной зависимости второго уровня, порожденный парой векторов $\varphi \in \Phi$ и $\varphi^{gl} \in \Phi^{gl}_{\varphi}$. В соответствии с определением 3 векторов зависимостей второго уровня, это означает, что в области вычислений алгоритма существуют точки J и $J+\varphi$, принадлежащие соответственно тайлам $\left(\tilde{J}^{gl}, J^{gl}\right) \in \tilde{V}^{gl}\left(J^{gl}\right)$ и $\left(\tilde{J}^{gl}+\tilde{\varphi}^{gl}, J^{gl}+\varphi^{gl}\right) \in \tilde{V}^{gl}\left(J^{gl}+\varphi^{gl}\right)$, $J^{gl}, J^{gl}+\varphi^{gl} \in V^{gl}$, что в силу предположений о параметрах тайлинга и определения множества точек тайла в виде (11) равносильно выполнению двух неравенств:

$$\begin{gathered}\vec{0} \le H\left(J-\tilde{J}^{\vec{0},gl}\right) \le \vec{\tilde{R}}-\vec{1},\\ \tilde{R}\tilde{\varphi}^{gl}+R\varphi^{gl}-H\varphi \le H\left(J-\tilde{J}^{\vec{0},gl}\right) \le \vec{\tilde{R}}-\vec{1}+\tilde{R}\tilde{\varphi}^{gl}+R\varphi^{gl}-H\varphi,\end{gathered} \tag{13}$$

где $\tilde{J}^{\vec{0},gl}$ – начальная точка тайла $\left(\tilde{J}^{gl}, J^{gl}\right)$. Одновременное выполнение этих неравенств эквивалентно неравенству

$$\max\left(\vec{0}, \tilde{R}\tilde{\varphi}^{gl}+R\varphi^{gl}-H\varphi\right) \le H\left(J-\tilde{J}^{\vec{0},gl}\right) \le \min\left(\vec{\tilde{R}}-\vec{1}, \vec{\tilde{R}}-\vec{1}+\tilde{R}\tilde{\varphi}^{gl}+R\varphi^{gl}-H\varphi\right),$$

или, что то же самое, неравенству

$$\left(\tilde{R}\tilde{\varphi}^{gl}+R\varphi^{gl}-H\varphi\right)_{+} \le H\left(J-\tilde{J}^{\vec{0},gl}\right) \le \vec{\tilde{R}}-\vec{1}+\left(\tilde{R}\tilde{\varphi}^{gl}+R\varphi^{gl}-H\varphi\right)_{-}.$$

Из последнего неравенства следует оценка для модулей координат вектора $\tilde{R}\tilde{\varphi}^{gl}+R\varphi^{gl}-H\varphi$:

$$\left|\left(\tilde{R}\tilde{\varphi}^{gl}+R\varphi^{gl}-H\varphi\right)_{k}\right| \le \tilde{r}_k-1, \quad 1 \le k \le n.$$

Раскрывая модуль в виде двойного неравенства и выражая координаты вектора $\tilde{R}\tilde{\varphi}^{gl}$, получаем

$$H\varphi-R\varphi^{gl}-\vec{\tilde{R}}+\vec{1} \le \tilde{R}\tilde{\varphi}^{gl} \le H\varphi-R\varphi^{gl}+\vec{\tilde{R}}-\vec{1}.$$

Умножая полученное неравенство на обратную матрицу $\tilde{R}^{-1}$, с учетом целочисленности вектора $\tilde{R}^{-1}\vec{R}$, получаем соотношение, которому удовлетворяют компоненты вектора ψ^{gl}:

$$\left\lfloor \tilde{R}^{-1}H\varphi \right\rfloor-\tilde{R}^{-1}R\varphi^{gl} \le \tilde{\varphi}^{gl} \le \left\lceil \tilde{R}^{-1}H\varphi \right\rceil-\tilde{R}^{-1}R\varphi^{gl}.$$

С учетом определения множества $\tilde{V}^{gl}\left(J^{gl}\right)$ в виде (12) имеем неравенство $-\tilde{R}^{-1}\vec{R}+\vec{1} \le \tilde{\varphi}^{gl} \le \tilde{R}^{-1}\vec{R}-\vec{1}$, что в сочетании с предыдущим неравенством приводит к получению сформулированного в утверждении соотношения, связывающего компоненты вектора ψ^{gl} с вектором φ.

Множество точек тайла $\left(\tilde{J}^{gl}, J^{gl}\right)$, операции которых непосредственно влияют на результат выполнения операций, приписанных точкам тайла $\left(\tilde{J}^{gl}+\tilde{\varphi}^{gl}, J^{gl}+\varphi^{gl}\right)$, т. е. множество V^{out}, определяется как совокупность целочисленных решений системы, составленной из неравенств (13). Из этих же соотношений, с учетом унимодулярности матрицы H, следует оценка мощности множества V^{out}. Аналогичным образом, с помощью формулы (11), определяется вид и оценка мощности множества V^{in}. Утверждение доказано.

Отображение алгоритма на параллельную архитектуру. Для параллельной реализации алгоритма необходимо решить задачу его пространственно-временного отображения на многопроцессорную вычислительную систему заданной конфигурации. Пространственное отображение заключается в распределении операций алгоритма и данных по процессорам (вычислительным процессам). Временное отображение состоит в определении порядка выполнения операций, отображенных на каждый из процессоров, и в согласовании параллельных вычислительных процессов между собой в случае необходимости взаимодействия между ними.

На практике параллельная реализация алгоритма проходит, как правило, в условиях ограниченности системных ресурсов. В первую очередь это ограничения на размер и размерность топологии параллельной архитектуры. В этом случае для организации параллельных вычислительных процессов применяются различные стратегии отображения алгоритма. Одной из известных и используемых стратегий является локально параллельная глобально последовательная стратегия отображения алгоритмов [8].

LPGS-стратегия определяет общий подход к отображению алгоритма на параллельную архитектуру и заключается в следующем: множество всех точек области вычислений алгоритма и, соответственно, его операций разбивается на непересекающиеся подмножества; каждое из полученных подмножеств точек области вычисления последовательно и в общем случае независимо отображается на всю параллельную архитектуру. С точки зрения временного отображения выполнение алгоритма происходит по двухуровневой схеме, включающей в себя глобальный уровень или уровень подмножеств операций алгоритма, полученных при разбиении области вычислений, и локальный уровень – уровень операций фиксированного подмножества. На глобальном уровне подмножества операций алгоритма последовательно выполняются на процессорах параллельной архитектуры. На локальном уровне в пределах одного подмножества операции алгоритма выполняются параллельно. Очевидно, что порядок выполнения алгоритма на обоих уровнях должен определяться в соответствии с зависимостями между операциями и их распределением по процессорам. Необходимо также отметить, что выбор разбиения области вычислений существенным образом влияет на решение задачи пространственно-временного отображения алгоритма и на потенциал к параллельной обработке в целом. Например, задача временного отображения на глобальном уровне существенно упрощается в случае, когда между подмножествами операций отсутствуют обратные связи.

Двухуровневый тайлинг по своей идее близок к LPGS-стратегии. Тайлинг первого уровня фактически соответствует разбиению области вычислений на глобальном уровне стратегии. Тайлинг второго уровня можно поставить в соответствие локальному уровню стратегии с учетом отображения операций алгоритма на процессоры. Такое соответствие дает возможность решения задачи пространственно-временного отображения в интеграции с тайлингом – общий принцип отображения алгоритма можно сформулировать следующим образом: на глобальном уровне параллельная вычислительная система последовательно выполняет тайлы первого уровня; на локальном уровне стратегии фиксированный тайл первого уровня разбивается на тайлы второго уровня, которые в свою очередь, как атомарные макрооперации, отображаются на процессоры и выполняются параллельно с учетом информационных зависимостей между ними. Такой интегрированный подход позволяет использовать математический аппарат двухуровневого тайлинга для решения задачи пространственно-временного отображения алгоритмов в рамках LPGS-стратегии и получать при этом ряд преимуществ, которые дает тайлинг как средство оптимизации программ.

Далее в разделе представлен формализованный метод решения задачи пространственно-временного отображения алгоритмов на параллельные архитектуры заданного размера и размерности. Метод разработан в рамках интегрированного подхода, предложенного выше, и предназначен для алгоритмов, область вычислений которых имеет форму n-мерного параллелепипеда вида (7).

Пусть целевая параллельная архитектура имеет размерность p и состоит из $l_1 \times l_2 \times \ldots \times l_p$ процессорных элементов $\text{ПЭ}\left(j_1, j_2, \ldots, j_p\right)$, координаты которых, как вектор, принадлежат множеству

$$\tilde{V}_p = \left\{ \left(j_1, j_2, \ldots, j_p\right) \in Z^p \,\middle|\, 0 \le j_k \le l_k - 1,\ k = 1, 2, \ldots, p \right\}.$$

Определим параметры тайлинга первого уровня, положив $J^{\vec{0}} = H^{-1}m$ и выбрав в качестве диагональных элементов матрицы R делители соответствующих координат вектора $\vec{R}^{\max} = M - m + \vec{1}$. В качестве матрицы, задающей форму тайлов, используем матрицу H из определения области вычислений алгоритма (7).

Зафиксируем p-элементную упорядоченную выборку попарно различных элементов из множества целочисленных значений $\{1,2,\ldots,n\}$ и обозначим ее как множество $\{k_1,k_2,\ldots,k_p\}$. В соответствии с этой выборкой определим параметры тайлинга второго уровня, исходя из условий $\tilde{H}=H$, $\tilde{J}^{\vec{0}}=H^{-1}\left(m+RJ^{gl}\right)$, $\tilde{R}^{-1}\vec{R}\in Z^n$ и дополнительного условия на размеры тайлов $\tilde{r}_{k_i}=r_{k_i}/l_i$, $i=1,2,\ldots,p$. Такой выбор параметров тайлинга приводит к получению множества тайлов второго уровня $\tilde{V}^{gl}\left(J^{gl}\right)$ в виде

$$\tilde{V}^{gl}\left(J^{gl}\right)=\left\{\left(\tilde{J}^{gl},J^{gl}\right)\in Z^{2n}\mid 0\le\tilde{J}^{gl}_{k_i}\le l_i-1,\ i=1,2,\ldots,p,\ 0\le\tilde{J}^{gl}_{k}\le r_k/\tilde{r}_k-1,\ k\notin\left\{k_1,k_2,\ldots,k_p\right\}\right\}.$$

Для пространственного отображения алгоритма необходимо определить функцию размещения, которая устанавливает соответствие между операциями алгоритма и процессорами, на которых они должны выполняться. Реализуемый в рамках метода принцип отображения алгоритмов и свойство атомарности тайлов позволяют задавать пространственное отображение не на множестве исходных операций алгоритма, а на множестве макроопераций второго уровня, что является более рациональным с точки зрения организации параллельных вычислительных процессов. Таким образом, пространственное отображение алгоритма будем задавать функцией размещения $g:Z^{2n}\to Z^p$ следующего вида:

$$g\left(\tilde{J}^{gl},J^{gl}\right)=\left(\tilde{J}^{gl}_{k_1},\tilde{J}^{gl}_{k_2},\ldots,\tilde{J}^{gl}_{k_p}\right),\ \left(\tilde{J}^{gl},J^{gl}\right)\in\tilde{V}^{gl}\left(J^{gl}\right),\ J^{gl}\in V^{gl}. \tag{14}$$

Функция g ставит в соответствие каждому тайлу второго уровня процессор, на котором он должен выполняться как макрооперация. Данная функция реализует так называемый «координатный» способ отображения элементов $2n$-мерного пространства в пространство меньшей размерности p. Координаты, по которым осуществляется отображение, непосредственно определяются выборкой $\{k_1,k_2,\ldots,k_p\}$. Отображение вида (14) многовариантно; количество вариантов равно количеству различных выборок $\{k_1,k_2,\ldots,k_p\}$, которое, в свою очередь, равно числу размещений из n по p.

В результате пространственного отображения тайлы второго уровня распределяются между процессорами параллельной архитектуры. Каждый процессор выполняет свой набор макроопераций второго уровня, последовательность которых в рамках одного процессора составляет один из параллельных процессов. Существование глобальных зависимостей между тайлами второго уровня в зависимости от выбранного варианта пространственного отображения может привести к необходимости организации обмена данными между параллельными процессами, протекающими на разных процессорах. Необходимость передачи данных возникает в том случае, когда информационно зависимые тайлы отображаются на разные процессоры. Таким образом, глобальные зависимости второго уровня порождают зависимости между процессорами. Очевидно, что для организации взаимодействия между параллельными процессами необходимо определить все зависимости между процессорами.

Рассматривая пары информационно зависимых тайлов второго уровня и их пространственное отображение в соответствии с функцией размещения вида (14), нетрудно показать, что зависимости между процессорами, по аналогии с зависимостями между операциями, определяются ненулевыми p-мерными векторами вида

$$d_{\psi^{gl}}=\left(\tilde{\varphi}^{gl}_{k_1},\tilde{\varphi}^{gl}_{k_2},\ldots,\tilde{\varphi}^{gl}_{k_p}\right),\ \psi^{gl}=(\tilde{\varphi}^{gl},\varphi^{gl})\in\Psi^{gl}_{\varphi,\varphi^{gl}},\ \varphi^{gl}\in\Phi^{gl}_{\varphi},\ \varphi\in\Phi. \tag{15}$$

Векторы $d_{\psi^{gl}}$ фактически являются образами соответствующих векторов глобальных зависимостей второго уровня ψ^{gl} при их отображении функцией g в p-мерное пространство. Каждый вектор ψ^{gl}, у которого существует хотя бы одна ненулевая координата с номером из набора $\{k_1,k_2,\ldots,k_p\}$, порождает вектор зависимостей на уровне процессоров. Необходимо заметить, что при применении предлагаемого метода отображения для однородных алгоритмов, с учетом разбиения только на полные тайлы, свойство однородности зависимостей сохраняется не только на уровне тайлов, но и на уровне процессоров.

Временное отображение алгоритма будем осуществлять на множестве макроопераций в соответствии с LPGS-стратегией по предложенной выше двухуровневой схеме. Для реализации такого способа отображения на глобальном уровне необходимо установить порядок выполнения на параллельной вычислительной системе тайлов первого уровня. На локальном уровне требуется для каждого из процессоров определить порядок выполнения тайлов второго уровня, отображенных на него функцией размещения g и являющихся в совокупности фрагментом зафиксированного глобально тайла первого уровня. Таким образом, каждый процессор параллельной системы последовательно выполняет фрагменты тайлов первого уровня, которые представляют собой подмножества тайлов второго уровня, полученные при разбиении и отображении тайлов первого уровня на параллельную архитектуру. Кроме этого, для синхронизации параллельных процессов необходимо реализовать на уровне тайлов второго уровня передачу данных между процессорами в соответствии с векторами зависимостей $d_{\psi^{gl}}$.

Зафиксируем на множестве целых чисел $\{1,2,\ldots,n\}$ перестановку и обозначим ее $\{i_1,i_2,\ldots,i_n\}$. На множестве чисел $\{t\in Z \mid 1\le t\le n, t\notin\{k_1,k_2,\ldots,k_p\}\}$ определим перестановку $\{t_1,t_2,\ldots,t_{n-p}\}$.

В силу условия корректности тайлинга (6), которому удовлетворяет параметр H, векторы глобальных зависимостей первого уровня имеют неотрицательные координаты. Это следует непосредственно из формулы, определяющей значения координат векторов, приведенной в утверждении 1. Данный факт позволяет установить на множестве макроопераций первого уровня лексикографический порядок их выполнения в соответствии со значениями координат векторов J^{gl}, упорядоченных в соответствии с перестановкой $\{i_1,i_2,\ldots,i_n\}$: тайл $J^{gl,1}$ выполняется раньше тайла $J^{gl,2}$, если существует $z\in N$, $1\le z\le n$, такое что выполняются условия $J^{gl,1}_{i_k}=J^{gl,2}_{i_k}$, $k=1,2,\ldots,z-1$, и $J^{gl,1}_{i_z}<J^{gl,2}_{i_z}$.

Глобальные зависимости между тайлами второго уровня в пределах фиксированного тайла первого уровня в соответствии с утверждением 2 также имеют неотрицательные координаты. Это позволяет установить лексикографический порядок выполнения макроопераций второго уровня на каждом процессоре по аналогии с глобальным уровнем отображения в соответствии с перестановкой $\{t_1,t_2,\ldots,t_{n-p}\}$. Тогда процесс выполнения алгоритма на произвольном процессоре ПЭ$(j_1,j_2,\ldots,j_p)$, может быть представлен в рамках модели обмена сообщениями в виде псевдокода следующего вида:

for $J^{gl}_{i_1}=0$ to $r^{\max}_{i_1}/r_{i_1}-1$
  for $J^{gl}_{i_2}=0$ to $r^{\max}_{i_2}/r_{i_2}-1$
    …
      for $J^{gl}_{i_n}=0$ to $r^{\max}_{i_n}/r_{i_n}-1$
        for $\tilde{J}^{gl}_{t_1}=0$ to $r_{t_1}/\tilde{r}_{t_1}-1$
          for $\tilde{J}^{gl}_{t_2}=0$ to $r_{t_2}/\tilde{r}_{t_2}-1$
            …
              for $\tilde{J}^{gl}_{t_{n-p}}=0$ to $r_{t_{n-p}}/\tilde{r}_{t_{n-p}}-1$
              begin
                Receive data in directions $d_{\psi^{gl}}$;
                Calculate tile $\left(\tilde{J}^{gl},J^{gl}\right)$, where $\tilde{J}^{gl}_{k_i}=j_i$, $i=1,2,\ldots,p$;
                Send data in directions $d_{\psi^{gl}}$;
              end

Первые n циклов псевдокода реализуют лексикографический порядок выполнения тайлов первого уровня в соответствии с заданной перестановкой $\{i_1, i_2, \ldots, i_n\}$, фиксируя при конкретных значениях счетчиков $J_{i_k}^{gl}, k = 1,2,\ldots,n$, тайл первого уровня J^{gl}. Следующие за ними $n-p$ циклов описывают очередность выполнения тайлов второго уровня $\left(\tilde{J}^{gl}, J^{gl}\right)$, принадлежащих фиксированному тайлу первого уровня J^{gl} и отображенных на процессор ПЭ$\left(j_1, j_2, \ldots, j_p\right)$. В теле полученного гнезда циклов фиксированные значения счетчиков $J_{i_k}^{gl}$, $k = 1,2,\ldots,n$, и $\tilde{J}_{t_k}^{gl}$, $k = 1,2,\ldots,n-p$, а также условие $\tilde{J}_{k_i}^{gl} = j_i$, $i = 1,2,\ldots,p$, однозначным образом определяют для выполнения на процессоре ПЭ$\left(j_1, j_2, \ldots, j_p\right)$ соответствующий тайл $\left(\tilde{J}^{gl}, J^{gl}\right)$.

В силу атомарности тайлов второго уровня все обмены данными, необходимые для выполнения тайла $\left(\tilde{J}^{gl}, J^{gl}\right)$, осуществляются либо до, либо после выполнения макрооперации. Перед выполнением тайла необходимо осуществить ряд операций приема данных от процессоров по всем возможным направлениям, определяемым векторами $d_{\psi^{gl}}$, где $\psi^{gl} = (\tilde{\varphi}^{gl}, \varphi^{gl}) \in \Psi_{\varphi,\varphi^{gl}}^{gl}$, $\varphi^{gl} \in \Phi_{\varphi}^{gl}$, $\varphi \in \Phi$. Аналогичным образом после выполнения макрооперации процессор осуществляет по направлениям $d_{\psi^{gl}}$ передачу ее результатов. В случае приема данных по направлению $d_{\psi^{gl}}$ процессор получает данные от процессора ПЭ$\left(j_1 - \tilde{\varphi}_{k_1}^{gl}, j_2 - \tilde{\varphi}_{k_2}^{gl}, \ldots, j_p - \tilde{\varphi}_{k_p}^{gl}\right)$. В случае передачи результатов по направлению $d_{\psi^{gl}}$ процессор посылает данные процессору ПЭ$\left(j_1 + \tilde{\varphi}_{k_1}^{gl}, j_2 + \tilde{\varphi}_{k_2}^{gl}, \ldots, j_p + \tilde{\varphi}_{k_p}^{gl}\right)$.

Операции приема-передачи данных по своей актуальности и содержанию определяются в соответствии с векторами глобальных зависимостей второго уровня ψ^{gl} и их проекциями $d_{\psi^{gl}}$ на множество процессоров. В общем случае каждому вектору ψ^{gl}, для которого $d_{\psi^{gl}} \neq \vec{0}$, соответствует отдельная операция обмена данными. Конкретные данные, которые участвуют в операциях приема и передачи, определяются соответственно множествами $V^{loc}\left(\tilde{J}^{gl}, J^{gl}\right)_{\varphi,\psi^{gl}}^{in}$ и $V^{loc}\left(\tilde{J}^{gl}, J^{gl}\right)_{\varphi,\psi^{gl}}^{out}$. Мощность этих множеств обусловливает количество таких данных.

В предложенном подходе к решению задачи отображения алгоритм рассматривается на уровне макроопераций. Представленный псевдокод реализует исходный алгоритм также на уровне макроопераций. Для параллельной реализации алгоритма на уровне исходных операций необходимо дополнительно установить и программно реализовать порядок выполнения операций внутри тайла.

Матрица H, по определению тайлинга, должна удовлетворять условию (6). Это означает, что на множестве точек $J \in T\left(\tilde{J}^{gl}, J^{gl}\right)$ можно установить лексикографический порядок выполнения приписанных им операций в соответствии со значением вектора HJ. Реализация порядка выполнения операций в программном коде представляет собой отдельную задачу.

В соответствии с формализованным подходом к тайлингу, представленным в данной работе, множество точек тайла образует параллелепипед вида (11). В случае, когда матрица тайлинга H является нижней треугольной с единицами на главной диагонали, параллелепипед имеет представление с явным определением зависимости граничных значений координат своих точек от параметров тайлинга, параметров, определяющих область вычислений и координат соответствующих точек с меньшими номерами:

$$T\left(\tilde{J}^{gl}, J^{gl}\right) = \left\{ J \in Z^n \mid \tilde{r}_k \tilde{J}_k^{gl} + r_k J_k^{gl} + m_k - \sum_{i=1}^{k-1} h_{ki} J_i \le J_k \le \tilde{r}_k \tilde{J}_k^{gl} + r_k J_k^{gl} + m_k + \tilde{r}_k - \sum_{i=1}^{k-1} h_{ki} J_i - 1 \right\}.$$

Такое представление множества точек тайла позволяет реализовать лексикографический порядок выполнения операций тайла в виде гнезда циклов

$$
\begin{array}{l}
\text{for}\quad J_1=\tilde r_1\tilde J_1^{gl}+r_1J_1^{gl}+m_1\quad\text{to}\quad \tilde r_1\tilde J_1^{gl}+r_1J_1^{gl}+m_1+\tilde r_1-1\\
\quad\text{for}\quad J_2=\tilde r_2\tilde J_2^{gl}+r_2J_2^{gl}+m_2-h_{21}J_1\quad\text{to}\quad \tilde r_2\tilde J_2^{gl}+r_2J_2^{gl}+m_2+\tilde r_2-h_{21}J_1-1\\
\qquad\ldots\\
\qquad\quad\text{for}\quad J_n=\tilde r_n\tilde J_n^{gl}+r_nJ_n^{gl}+m_n-\sum\limits_{i=1}^{n-1}h_{ni}J_i\quad\text{to}\quad \tilde r_n\tilde J_n^{gl}+r_nJ_n^{gl}+m_n+\tilde r_n-\sum\limits_{i=1}^{n-1}h_{ni}J_i-1\\
\qquad\qquad\text{Calculate } S(J);
\end{array}
$$

где $S(J)$ – операция алгоритма, ассоциированная с точкой J области вычислений V.

В общем случае формализацию порядка выполнения операций тайла можно выполнить за счет аффинного преобразования тайла матрицей H к n-мерному многограннику и применения метода исключения Фурье – Моцкина для генерации программного кода.

В заключение необходимо отметить, что вариативность порядка выполнения макроопераций первого и второго уровней не ограничивается вариантами перестановок $\{i_1, i_2, \ldots, i_n\}$ и $\{t_1, t_2, \ldots, t_{n-p}\}$. Корректным будет любой порядок, сохраняющий зависимости между макрооперациями. В качестве альтернативного подхода к упорядочиванию макроопераций, а также операций внутри тайла второго уровня, можно использовать аппарат векторных функций таймирования [9].

Работа выполнена в рамках подпрограммы «Математические методы» Государственной программы научных исследований «Конвергенция».

Список использованной литературы

1. *Xue, J.* Loop Tiling For Parallelism / J. Xue. – Norwell: Kluwer Academic Publishers, 2000.

2. *Irigoin, F.* Supernode partitioning / F. Irigoin, R. Triolet // Proceedings of the ACM SIGPLAN Symposium on Principles of Programming Languages, San Diego, California, Jan. 1988. – San Diego, 1988. – P. 319–329.

3. Parameterized tiled loops for free / L. Renganarayanan [et al.] // SIGPLAN Conf. on Programming Language Design and Implementation. – New York: ASM Press, 2007. – P. 405–414.

4. DynTile: Parametric Tiled Loop Generation for Parallel Execution on Multicore Processors / A. Hartono [et al.] // 24th International Parallel and Distributed Proc. Symposium (2010 IPDPS Conference), Atlanta, April 2010. – Atlanta, 2010.

5. Parametric tiling of affine loop nests / S. Tavarageri [et al.] // Proc. 15th Workshop on Compilers for Parallel Computers, Vienna, Austria, July 2010. – Vienna, 2010.

6. *Баханович, С. В.* Параметризованный тайлинг: точные аппроксимации и анализ глобальных зависимостей / С. В. Баханович, П. И. Соболевский // ЖВМ. – 2014. – Т. 54, № 11. – С. 1817–1828.

7. Primetile: A Parametric Multi-Level Tiler for Imperfect Loop Nests / A. Hartono [et al.] // Proc. of the 23rd Int. Conf. on Supercomputing (ICS'09). – P. 147–157.

8. *Баханович, С. В.* Отображение алгоритмов на параллельные архитектуры заданной размерности и заданного размера / С. В. Баханович, Н. А. Лиходед // Вес. Нац. акад. навук Беларусі. Сер. фіз.-мат. навук. – 2003. – № 1. – С. 101–108.

9. *Лиходед, Н. А.* Получение аффинных преобразований для улучшения локальности гнезд циклов / Н. А. Лиходед, С. В. Баханович, А. В. Жерело // Программирование. – 2005. – № 5. – С. 52–65.

Поступила в редакцию 28.04.2016