ПРОГРАММНО-ТЕХНИЧЕСКИЕ КОМПЛЕКСЫ

П.И. СОБОЛЕВСКИЙ, С.В. БАХАНОВИЧ, А.Н. ГОРБАЧ

УДК 681.3.012+519.6

ОПТИМИЗАЦИЯ ТАЙЛИНГА ПРИ ЧИСЛЕННОМ РЕШЕНИИ МНОГОМЕРНОГО УРАВНЕНИЯ ТЕПЛОПРОВОДНОСТИ НА КОЛЬЦЕ ПРОЦЕССОРОВ [1]

Ключевые слова: *тайлинг, тайл, вычислительная система с распределенной памятью, оптимизация, уравнение теплопроводности.*

Один из важнейших аспектов при разработке и оптимизации параллельных приложений для вычислительных систем с распределенной памятью — эффективная организация коммуникации. На практике коммуникации часто являются узким местом параллельных приложений вследствие высокой латентности каналов связи по сравнению с производительностью узлов. Одно из эффективных средств оптимизации коммуникаций — техника тайлинга, позволяющая уменьшить долю коммуникаций по сравнению с долей вычислений параллельного алгоритма и дополнительно повысить эффективность использования памяти вычислительных узлов.

В данной работе исследуется задача оптимизации тайлинга при решении первой краевой задачи для многомерного уравнения теплопроводности

$$\frac{\partial u}{\partial x_1}=\sum_{m=2}^{n}\frac{\partial^2 u}{\partial x_m^2},\ 0<x_m<1,\ 2\le m\le n,\ 0<x_1<T, \tag{1}$$

на вычислительных системах с распределенной памятью, построены оценки объема коммуникаций и вычислений при заданных вариантах тайлинга. На их основе для случаев $n=2$ и $n=3$ получены оценки полного времени реализации алгоритма, представляющие собой функцию от параметров тайлинга и параметров целевого суперкомпьютера: размерности и размеров вычислительной среды, производительности процессоров, времени инициализации и пропускной способности каналов связи. Минимизация этой функции позволяет оптимизировать тайлинг.

1. ТАЙЛИНГ

Предположим, что параллельный алгоритм, как это часто бывает на практике, строится на основе соответствующего последовательного алгоритма. Рассматрим n-мерные алгоритмы с однородными зависимостями, записанные в виде гнезда тесновложенных циклов:

$$\begin{array}{l}
\text{for } i_1=1 \text{ to } I_1 \text{ do}\\
\quad \ldots\ldots\ldots\ldots\ldots\ldots\\
\quad \text{for } i_n=1 \text{ to } I_n \text{ do}\\
\qquad \text{begin}\\
\qquad\quad S_1(i_1,i_2,\ldots,i_n);\\
\qquad\quad \ldots\ldots\ldots\ldots\ldots\\
\qquad\quad S_k(i_1,i_2,\ldots,i_n)\\
\qquad \text{end}
\end{array} \tag{2}$$

[1] Работа выполнена при финансовой поддержке Белорусского республиканского фонда фундаментальных исследований (проект № Ф08-016).

Алгоритмы такого вида характеризуются областью вычислений V и множеством векторов зависимостей Ф. Область вычислений есть подмножество точек целочисленного пространства $\mathbf{Z}^n$, которое определяется как $V=\{v(i_1,i_2,\ldots,i_n)\in \in \mathbf{Z}^n \mid 1\le i_m \le I_m, 1\le m\le n\}$. Каждой точке $v(i_1,i_2,\ldots,i_n)\in V$ в соответствии с алгоритмом приписано k операций $S_i(v)=S_i(i_1,i_2,\ldots,i_n)$, $1\le i\le k$. Множество векторов зависимостей отражает информационные зависимости между операциями алгоритма: наличие вектора $\varphi^{(i,j)}$ множества Ф означает, что существуют такие $v\in V$ и $i,j\in N$, $1\le i,j\le k$, что операция $S_j(v+\varphi^{(i,j)})$ информационно зависит от операции $S_i(v)$.

Техника тайлинга состоит в разбиении области вычислений алгоритма (и соответственно множества его операций) n семействами параллельных гиперплоскостей на множество n-мерных параллелепипедов-тайлов. Множество операций, принадлежащих одному тайлу, составляют новое (укрупненное) зерно вычислений. Семейства гиперплоскостей задаются множеством из n линейно независимых векторов $\{h_1,h_2,\ldots,h_n\}$. Каждому семейству соответствует свой вектор h_m, нормальный для гиперплоскостей этого семейства. Получаемое разбиение области V на тайлы должно удовлетворять следующим условиям:

1) тайлы должны быть однотипными, т.е. одинаковыми по форме и размеру (за исключением возможно граничных);

2) между тайлами не должно существовать обратных связей.

Введем ряд формальных определений и обозначений.

• $H=(h_1,h_2,\ldots,h_n)^{\mathrm{T}}$ — квадратная матрица порядка n, строками которой выступают координаты линейно независимых целочисленных векторов h_m, $1\le m\le n$, — нормальных векторов гиперплоскостей, осуществляющих разбиение области вычислений на тайлы. Будем считать, что наибольший общий делитель модулей координат каждого из этих векторов равен единице.

• $R=\mathrm{diag}\,(r_1,r_2,\ldots,r_n)$ — диагональная матрица. Натуральное число r_m обозначает количество гиперплоскостей с нормальным вектором h_m, $1\le m\le n$, проходящих через каждый тайл.

• $v^{\min}=(v_1^{\min},v_2^{\min},\ldots,v_n^{\min})^{\mathrm{T}}$ — вектор, определяемый условиями $h_m\cdot v_m^{\min} = =\min\limits_{v\in V} h_m\cdot v$, $1\le m\le n$.

• $1_n=(1,1,\ldots,1)^{\mathrm{T}}$ — вектор размерности n, составленный из единиц.

• $I=(I_1,I_2,\ldots,I_n)^{\mathrm{T}}$ — вектор, построенный из значений I_m, $1\le m\le n$.

• $\lfloor a\rfloor$ — вектор, составленный из целых частей координат вектора a. $\{a\}$ — вектор, построенный из дробных частей координат вектора a.

Формально тайлинг можно рассматривать как нелинейное отображение множества V на множество $V'\subset \mathbf{Z}^{2n}$: $v\to(j,\alpha)$, $v\in V\subset \mathbf{Z}^n$, $(j,\alpha)\in V'\subset \mathbf{Z}^{2n}$, где

$$j=(j_1,j_2,\ldots,j_n)^{\mathrm{T}}=1_n+\lfloor R^{-1}H(v-v^{\min})\rfloor,$$

$$\alpha=(\alpha_1,\alpha_2,\ldots,\alpha_n)^{\mathrm{T}}=R(1_n-\{R^{-1}H(v-v^{\min})\}). \tag{3}$$

Вектор $j=(j_1,j_2,\ldots,j_n)^{\mathrm{T}}$, $j_m=1,2,\ldots,J_m$, $J_m=\left\lceil \frac{1}{r_m}\sum\limits_{k=1}^{n}|h_{m,k}|(I_k-1)+1\right\rceil$, $1\le m\le n$, определяет глобальные координаты тайла, вектор $\alpha=(\alpha_1,\alpha_2,\ldots,\alpha_n)^{\mathrm{T}}$, $\alpha_m=1,2,\ldots,r_m$, $1\le m\le n$, — локальные (в каждом тайле) координаты точки $v=(i_1,i_2,\ldots,i_n)\in V$. Координата α_m обозначает порядковый номер гиперплоскости, проходящей через тайл с нормальным вектором h_m.

Векторы зависимостей для алгоритма вида (2) обладают таким свойством: первая ненулевая координата вектора больше нуля. Отсюда, а также из определения тайлинга и обозначений, приведенных выше, следует, что точка $v^{\min}$ принадлежит тайлу с единичными координатами и является точкой пересечения гиперплоскостей с номерами, определяемыми вектором $\alpha = R1_n = (r_1, r_2, \ldots, r_n)^{\mathrm{T}}$.

Пусть H_+ и H_- — матрицы, полученные из матрицы H заменой соответственно отрицательных и положительных элементов нулями. Тогда имеет место равенство $Hv^{\min} = H_+ 1_n + H_- I$.

Очевидно соотношение

$$H(v - v^{\min}) = Rj - \alpha, \tag{4}$$

из которого при условии целочисленности вектора $v^{\min} + H^{-1}(Rj - \alpha)$ следует $v = v^{\min} + H^{-1}(Rj - \alpha) \in V$ или с учетом приведенных определений

$$v = v^{\min} + \frac{1}{|H|} \begin{Bmatrix} \sum_{m=1}^{n} (r_m j_m - \alpha_m) \mathrm{A}_{m1} \\ \sum_{m=1}^{n} (r_m j_m - \alpha_m) \mathrm{A}_{m2} \\ \vdots \\ \sum_{m=1}^{n} (r_m j_m - \alpha_m) \mathrm{A}_{mn} \end{Bmatrix}, \tag{5}$$

где A_{mk} — алгебраическое дополнение элемента h_{mk} матрицы H, $|H|$ — определитель матрицы H.

Определение 1. Тайлы j', j'' называются однотипными, если существует такой вектор $b \in \mathbf{Z}^n$, что тайл j'' может быть получен путем параллельного переноса тайла j' на вектор b.

Утверждение 1. Пусть $I_m \to \infty$, $1 \le m \le n$. Для того чтобы все тайлы были однотипными, необходимо и достаточно, чтобы матрица $H^{-1}R$ была целочисленной, т.е. чтобы все диагональные элементы r_m матрицы r делились соответственно на $\Delta_m = \dfrac{|H|}{\text{НОД}(|H|, A_{m1}, A_{m2}, \ldots, A_{mn})}$, $1 \le m \le n$.

Действительно, пусть j', j'' — два тайла, $v' = v^{\min} + H^{-1}(Rj' - \alpha)$, $v'' = v^{\min} + H^{-1}(Rj'' - \alpha)$ — две точки, являющиеся точками пересечения гиперплоскостей с одинаковыми номерами, определяемыми одним и тем же вектором α. Из равенства $v'' - v' = H^{-1}R(j'' - j')$ следует, что тайл j'' можно получить из тайла j' параллельным переносом тайла j' на вектор $H^{-1}R(j'' - j')$. Следовательно, координаты вектора $H^{-1}R(j'' - j')$ должны принимать целочисленные значения. В свою очередь целочисленность координат вектора $H^{-1}R(j'' - j')$ при условии неограниченности области вычислений эквивалентна целочисленности элементов матрицы $H^{-1}R$.

Утверждение 2[1]. Для того чтобы между тайлами отсутствовали обратные связи, необходимо и достаточно выполнения условия

$$H\varphi \ge 0, \ \varphi \in \Phi. \tag{6}$$

Условие (6) эквивалентно тому, что проекции векторов зависимостей на прямые с направляющими векторами h_m, $1 \le m \le n$, либо нулевые, либо одинаково направлены.

Способ разбиения на супервершины (их форма, размеры) и способ отображения множества супервершин на параллельную архитектуру определяют эффективность реализации алгоритма. Кроме того, на эффективность реализации влияет и то, в каком объеме при разбиении учитываются характеристики суперкомпьютера, такие как производительность процессоров, время инициализации каналов связи и их пропускная способность.

Перспективность тайлинга повлекла за собой ряд исследований в целях разработки методов построения оптимальных разбиений. Результатом этих исследований стал ряд методов оптимизации тайлинга [1–9]. Отличия этих методов состоят в определении критериев оптимальности: в какой степени при разбиении учитываются особенности алгоритма, характеристики вычислительной системы, а также, как решается задача отображения тайлов на множество вычислительных узлов параллельного компьютера. Недостаток методов заключается в том, что они не в полной мере, а некоторые из них вообще не учитывают характеристики целевой вычислительной системы. Исключение составляют методы, представленные в [8, 9]. Однако они пригодны только тогда, когда множество векторов зависимостей алгоритма и множество нормальных векторов гиперплоскостей, осуществляющих разбиение на тайлы, состоит только из координатных векторов e_m, $1 \le m \le n$ (n-мерные векторы, у которых m-я координата равна единице, а остальные нулевые).

2. АЛГОРИТМ РЕШЕНИЯ ПЕРВОЙ КРАЕВОЙ ЗАДАЧИ ДЛЯ МНОГОМЕРНОГО УРАВНЕНИЯ ТЕПЛОПРОВОДНОСТИ

Последовательный алгоритм вида (2) решения задачи (1), основанный на использовании явной разностной схемы, можно записать в следующем виде:

$$\begin{aligned} &\text{for } i_1 = 1 \text{ to } I_1 \text{ do} \\ &\quad \text{for } i_2 = 1 \text{ to } I_2 \text{ do} \\ &\qquad \cdots\cdots\cdots\cdots \\ &\qquad\quad \text{for } i_n = 1 \text{ to } I_n \text{ do} \\ y(i_1, i_2, \dots, i_n) &= \left(1 - 2\sum_{m=2}^{n} \frac{\tau_1}{\tau_m^2}\right) y(i_1 - 1, i_2, \dots, i_n) + \\ &+ \sum_{k=2}^{n} \frac{\tau_1}{\tau_m^2} (y(i_1 - 1, i_2, \dots i_{k-1}, i_k - 1, i_{k+1}, \dots, i_n) + \\ &+ y(i_1 - 1, i_2, \dots i_{k-1}, i_k + 1, i_{k+1}, \dots, i_n)). \end{aligned} \tag{7}$$

Областью вычислений V для данного алгоритма есть подмножество точек пространства $\mathbf{Z}^n$, определяемое как $V = \{v(i_1, i_2, \dots, i_n) \in \mathbf{Z}^n \,|\, 1 \le i_m \le I_m, 1 \le m \le n\}$. Множество $\Phi = \{\varphi_1, \varphi_k^+, \varphi_k^- \mid \varphi_1 = e_1, \ \varphi_k^+ = e_1 + e_k, \ \varphi_k^- = e_1 - e_k, 2 \le k \le n\}$ отражает информационные зависимости между операциями алгоритма.

3. ОЦЕНКА ВЫЧИСЛИТЕЛЬНОГО И КОММУНИКАЦИОННОГО ОБЪЕМОВ ПРИ $n = 2$

Рассмотрим вариант разбиения на тайлы, определяемый матрицей $H_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$.

Очевидно, что для этого варианта тайлинга условие (6) выполняется и имеют место следующие равенства:

$$v^{\min} = (1,1)^{\mathrm{T}}, \ J = \left(\left\lceil \frac{I_1}{r_1} \right\rceil, \left\lceil \frac{I_1 + I_2 - 1}{r_2} \right\rceil \right)^{\mathrm{T}},$$

$$j = \left(\left\lfloor 1 + \frac{i_1 - 1}{r_1} \right\rfloor, 1 + \left\lfloor \frac{i_1 + i_2 - 2}{r_2} \right\rfloor \right)^{\mathrm{T}}, \ \alpha = \left(r_1 - r_1 \left\{ \frac{i_1 - 1}{r_1} \right\}, r_2 - r_2 \left\{ \frac{i_1 + i_2 - 2}{r_2} \right\} \right)^{\mathrm{T}}.$$

Множество векторов $\alpha = (\alpha_1, \alpha_2)^{\mathrm{T}}$, $\alpha_m = 1, 2, \ldots, r_m$, $m = 1, 2$, обозначим $\Lambda(R)$, а количество различных векторов во множестве $\Lambda(R)$ запишем $|\Lambda(R)|$.

Требование однотипности не накладывает никаких ограничений на параметры r_1, r_2, кроме их целочисленности, так как $\Delta_1 = 1$ и $\Delta_2 = 1$.

Тайлы можно рассматривать как новые укрупненные зерна вычислений (макрооперации), а множество всех тайлов, обозначим его U, — как область вычислений алгоритма (7) на уровне макроопераций.

Информационная зависимость между тайлами $j, j+\beta \in U$ определяется наличием точек $v \in V$ и $v + \varphi \in V$, вектора зависимости $\varphi \in \Phi$, таких что $v \in j$, $v + \varphi \in j + \beta$. Множество всех векторов $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$ — это множество векторов зависимостей алгоритма на уровне тайлов, обозначим его B.

Множество всех точек $v \in j$, для каждой из которых существуют точки $v + \varphi \in j + \beta$, обозначим $V_{j,\beta}$. Множество векторов $\varphi \in \Phi$, определяющих информационную зависимость между тайлами j и $j+\beta$, запишем Φ_β.

Пусть $v = v^{\min} + H^{-1}(Rj - \alpha^1)$, $v + \varphi = v^{\min} + H^{-1}(R(j+\beta) - \alpha^2)$, $\alpha^1, \alpha^2 \in \Lambda(R)$. Тогда $\Phi_\beta = \{\varphi \in \Phi \,|\, H\varphi = R\beta + \alpha^1 - \alpha^2,\ \alpha^1, \alpha^2 \in \Lambda(R)\}$. Другими словами, вектор $\varphi \in \Phi$ принадлежит множеству Φ_β, если существуют векторы $\alpha^1, \alpha^2 \in \Lambda(R)$, удовлетворяющие равенству $H\varphi = R\beta + \alpha^1 - \alpha^2$. Если для всех $\varphi \in \Phi$ при фиксированном β таких векторов $\alpha^1, \alpha^2 \in \Lambda(R)$ не существует, то множество Φ_β пусто и тайлы j и $j+\beta$ информационно независимы. Заметим, что множество Φ_β будет пустым, если вектор β имеет хотя бы одну отрицательную координату. Действительно, пусть m-я координата β_m вектора β отрицательна. Тогда из определения множества Φ_β следует $h_m \cdot \varphi = \alpha_m^1 - \alpha_m^2 + \beta_m r_m \le r_m - 1 + \beta_m r_m < 0$, что противоречит условию (6).

Определение 2. Два тайла: $j, j+\beta$ будем называть соседними по направлению ненулевого вектора $\beta \in \mathbf{Z}^n$, если координаты вектора β равны 0 либо 1, а $\max\limits_{\varphi \in \Phi_\beta} h_m \cdot \varphi \le r_m - 1$, $1 \le m \le n$.

Определим коммуникационный объем данных, передаваемых от тайла j к соседнему тайлу $j+\beta$, как $T_{\text{comm}}^{\beta} = a + b|V_{j,\beta}|$. Здесь a — время инициализации канала связи, b — время, необходимое на непосредственную передачу единицы данных, $|V_{j,\beta}|$ — количество точек $v \in j$ в множестве $V_{j,\beta}$, или, что то же самое, количество различных векторов $\alpha^1 \in \Lambda(R)$, для которых существуют соответствующие векторы $\alpha^2 \in \Lambda(R)$, удовлетворяющие всем равенствам

$$H\varphi = R\beta + \alpha^1 - \alpha^2,\ \varphi \in \Phi_\beta. \tag{8}$$

Утверждение 3. Пусть $n = 2$, $H = H_1$ и $r_1 \ge 2$, $r_2 \ge 3$. Тогда вычислительный объем тайла равен $T_{\text{comp}} = t_0 |\Lambda(R)| = t_0 r_1 r_2$, множество B имеет вид $B = \{\beta^1 = (0,1),\ \beta^2 = (1,0), \beta^3 = (1,1)\}$, а коммуникационные объемы тайла равны соответственно $T_{\text{comm}}^{\beta^1} = a + 2b(r_1 - 1)$, $T_{\text{comm}}^{\beta^2} = a + br_2$, $T_{\text{comm}}^{\beta^3} = a + 2b$, где множества векторов, определяющих информационную зависимость между соседними тайлами j и $j + \beta^k$, $k = 1, 2, 3$, имеют вид $\Phi_{\beta^1} = \{\varphi_1, \varphi_2^+\}$, $\Phi_{\beta^2} = \{\varphi_1, \varphi_2^-, \varphi_2^+\}$, $\Phi_{\beta^3} = \{\varphi_1, \varphi_2^+\}$.

Покоординатно равенства (8) запишем $h_m \cdot \varphi = \alpha_m^1 - \alpha_m^2 + \beta_m^k r_m$, $\varphi \in \Phi_{\beta^k}$, $k = 1, 2, 3$, причем $h_m \cdot \varphi \in \{0, 1, 2\}$, $m = 1, 2$. Но равенство $r_m + \alpha_m^1 - \alpha_m^2 = 0$ невозможно ни при каком выборе векторов $\alpha^1, \alpha^2 \in \Lambda(R)$, равенство $r_m + \alpha_m^1 - \alpha_m^2 = 1$ имеет место, только если $\alpha_m^1 = 1, \alpha_m^2 = r_m$, а равенство $r_m + \alpha_m^1 - \alpha_m^2 = 2$ возможно, только если $\alpha_m^1 = 1, \alpha_m^2 = r_m - 1$, либо $\alpha_m^1 = 2, \alpha_m^2 = r_m$, $m = 1, 2$. Это позволяет определить множество различных пар векторов $\alpha^1, \alpha^2 \in \Lambda(R)$, удовлетворяющих (покоординатно) всем равенствам (8), и, следовательно, определить значение $|V_{j,\beta}|$.

Поскольку множество векторов $\Lambda(R)$ имеет вид $\Lambda(R) = \{\alpha \in \mathbf{Z}^2 \mid 1 \le \alpha_m \le r_m, m = 1, 2\}$, то $|\Lambda(R)| = r_1 r_2$. Вектор φ_1 принадлежит множеству Φ_{β^1}, так как имеется $r_1 - 1$ пар векторов $\alpha^1, \alpha^2 \in \Lambda(R)$, $\alpha^1 = (k, 1)$, $\alpha^2 = (k-1, r_2)$, $k = 2, 3, \ldots, r_1$, удовлетворяющих равенствам $\alpha_1^1 - \alpha_1^2 = 1$, $r_2 + \alpha_2^1 - \alpha_2^2 = 1$. Аналогично $\varphi_1 \in \Phi_{\beta^2}$, так как имеется $r_2 - 1$ пар векторов $\alpha^1, \alpha^2 \in \Lambda(R)$, $\alpha^1 = (1, k)$, $\alpha^2 = (r_1, k-1)$, $k = 2, 3, \ldots, r_2$, удовлетворяющих равенствам $\alpha_2^1 - \alpha_2^2 = 1$, $r_1 + \alpha_1^1 - \alpha_1^2 = 1$. Вектор φ_1 принадлежит также множеству Φ_{β^3}, поскольку имеется одна пара векторов $\alpha^1, \alpha^2 \in \Lambda(R)$, $\alpha^1 = (1, 1)$, $\alpha^2 = (r_1, r_2)$, удовлетворяющих равенствам $r_1 + \alpha_1^1 - \alpha_1^2 = 1$, $r_2 + \alpha_2^1 - \alpha_2^2 = 1$.

Аналогично определяется, что вектор φ_2^- принадлежит множеству Φ_{β^2}, а вектор φ_2^+ — всем множествам Φ_{β^k}, $k = 1, 2, 3$.

В приведенном утверждении не учтена особенность граничных тайлов. Это оправдано, если размеры области V достаточно велики и влиянием граничных тайлов можно пренебречь.

4. ОЦЕНКА ВЫЧИСЛИТЕЛЬНОГО И КОММУНИКАЦИОННОГО ОБЪЕМОВ ПРИ $n \ge 3$

Пусть $n = 3$. В качестве матрицы H рассмотрим матрицу $H_2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix}$. Условие однотипности тайлов эквивалентно требованию четности всех диагональных элементов матрицы R, так как $|H_3| = 4$, $\Delta_m = 2$, $1 \le m \le 3$.

Утверждение 4. Пусть I_2 — четное число, а I_3 — нечетное. Тогда вычислительный объем тайла равен $T_{\text{comp}} = t_0 \dfrac{r_1 r_2 r_3}{4}$; множество B состоит из векторов $\beta^1 = (0, 0, 1)$, $\beta^2 = (0, 1, 0)$, $\beta^3 = (0, 1, 1)$, $\beta^4 = (1, 0, 0)$, $\beta^5 = (1, 1, 0)$; коммуникационный объем тайла равен $T_{\text{comm}}^{\beta} = a + b|V_{j,\beta}|$, где $\beta = \beta^k$, $1 \le k \le 5$, $|V_{j,\beta^1}| = r_1 r_2 / 2$, $|V_{j,\beta^2}| = r_1 r_3 / 2 - 2$, $|V_{j,\beta^3}| = r_1$, $|V_{j,\beta^4}| = r_2 r_3 / 2$, $|V_{j,\beta^5}| = r_3$.

Доказательство проводится аналогично предыдущему с учетом того, что если I_2 — четное число, а I_3 — нечетное, то структура множества векторов $\Lambda(R)$ имеет вид $\Lambda(R) = \{(2k_1, 2k_2 - 1, 2k_3 - 1), (2k_1 - 1, 2k_2, 2k_3), 1 \le k_m \le r_m / 2, m = 1, 2, 3\}$, а информационная зависимость соседних по направлениям β^k, $1 \le k \le 5$, тайлов определяется соответствующими множествами векторов $\Phi_{\beta^1} = \{\varphi_1, \varphi_2^-, \varphi_3^-\}$, $\Phi_{\beta^2} = \{\varphi_1, \varphi_2^-, \varphi_3^+\}$, $\Phi_{\beta^3} = \{\varphi_1, \varphi_2^-\}$, $\Phi_{\beta^4} = \{\varphi_1, \varphi_2^+, \varphi_3^+\}$, $\Phi_{\beta^5} = \{\varphi_3^+\}$.

Пусть $n \ge 3$. В качестве матрицы H рассмотрим матрицу

$$H_3 = \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & \cdots & 1 & -1 \\ 1 & 1 & \cdots & -1 & 1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 1 & -1 & \cdots & 1 & 1 \end{pmatrix}.$$

Тогда $|H_3| = (-1)^{\frac{n(n-1)}{2}} 2^{n-1}$. Условие однотипности тайлов эквивалентно требованию четности всех диагональных элементов матрицы R. Вычислительный объем тайла равен $T_{\text{comp}} = t_0 \dfrac{1}{2^{n-1}} \prod_{m=1}^{n} r_m$. При определении коммуникационных объемов тайла учитывается структура множества $\Lambda(R)$, которая зависит также от четности параметров I_m, $2 \le m \le n$. В частности, если в алгоритме (2) все параметры I_m, $2 \le m \le n$, — нечетные числа, то

$$\Lambda(R) = \{(2k_1, 2k_2, \ldots, 2k_n), (2k_1 - 1, 2k_2 - 1, \ldots, 2k_n - 1),\ 1 \le k_m \le r_m / 2,\ 1 \le m \le n\},$$

если все параметры I_m, $2 \le m \le n$, — четные числа, то

$$\Lambda(R) = \{(2k_1 - 1, 2k_2, \ldots, 2k_n), (2k_1, 2k_2 - 1, \ldots, 2k_n - 1),\ 1 \le k_m \le r_m / 2,\ 1 \le m \le n\}.$$

Аналогично определяются все допустимые векторы, удовлетворяющие равенствам (8), так как все равенства (8) имеют вид $h_m \cdot \varphi = \alpha_m^1 - \alpha_m^2 + \beta_m^k r_m$, $\varphi \in \Phi_{\beta^k}$, $\beta^k \in B$, причем $h_m \cdot \varphi \in \{0, 1, 2\}$, $1 \le m \le n$.

5. АППРОКСИМАЦИЯ МНОЖЕСТВА ТАЙЛОВ

Для построения оценки полного времени реализации алгоритма необходимо явно определить множество тайлов.

Пусть $D = \dfrac{1}{|H|} \operatorname{diag}(d_1, d_2, \ldots, d_n)$, $d_k = \text{НОД}(A_{1k} r_1, A_{2k} r_2, \ldots, A_{nk} r_n)$, $1 \le k \le n$.

Введем в рассмотрение векторы

$$C^+ = D^{-1} H_+^{-1} R 1_n + \lfloor D^{-1} (H^{-1} H_+ (I - 1_n) + H_-^{-1} 1_n) \rfloor$$

и

$$C^- = D^{-1} H_-^{-1} R 1_n - \lfloor D^{-1} (H^{-1} H_- (I - 1_n) - H_+^{-1} 1_n) \rfloor .$$

Утверждение 5. Пусть разбиение на тайлы определяется матрицей H. Тогда множество тайлов ограничено n-мерным многогранником

$$\overline{U} = \{ j \in Z^n \,|\, C^- \le D^{-1} H^{-1} R j \le C^+,\ 1_n \le j \le J \}. \tag{9}$$

Доказательство следует из оценок

$$H^{-1} R j = H_+^{-1} R j + H_-^{-1} R j = H_+^{-1} R (1_n + \lfloor R^{-1} (H_+ (v - 1_n) + H_- (v - I)) \rfloor) +$$

$$+ H_-^{-1} R (\lceil R^{-1} (H_+ (v - 1_n) + H_- (v - I) + 1_n) \rceil) \le$$

$$\le H_+^{-1} R (1_n + R^{-1} (H_+ (v - 1_n) + H_- (v - I))) +$$

$$+ H_-^{-1} R (R^{-1} (H_+ (v - 1_n) + H_- (v - I) + 1_n)) =$$

$$= H_+^{-1} R 1_n + H_-^{-1} 1_n + v - H^{-1} H_+ 1_n - H^{-1} H_- I \le$$

$$\le H_+^{-1} R 1_n + H_-^{-1} 1_n + H^{-1} H_+ (I - 1_n),$$

$$H^{-1} R j = H_+^{-1} R j + H_-^{-1} R j = H_+^{-1} R \lceil R^{-1} (H_+ (v - 1_n) + H_- (v - I) + 1_n) \rceil +$$

$$+ H_-^{-1} R (1_n + \lfloor R^{-1} (H_+ (v - 1_n) + H_- (v - I)) \rfloor) \ge$$

$$\ge H_+^{-1} (H_+ (v - 1_n) + H_- (v - I) + 1_n) + H_-^{-1} R 1_n + H_-^{-1} (H_+ (v - 1_n) + H_- (v - I)) =$$

$$= H_-^{-1} R 1_n + H_+^{-1} 1_n + v - H^{-1} H_+ 1_n - H^{-1} H_- I \ge$$

$$\ge H_1^{-1} R 1_n + H_+^{-1} 1_n - H^{-1} H_- (I - 1_n).$$

6. ОТОБРАЖЕНИЕ НА КОЛЬЦО ПРОЦЕССОРОВ

Применим LPGS-стратегию пространственно-временного отображения полученного алгоритма на кольцо процессоров, состоящее из заданного числа P вычислительных узлов (процессорных элементов) [8]. Оценим время решения задачи сверху функцией, зависящей явным образом от параметров компьютера, размеров сетки узлов в алгоритме и от чисел r_m, $1 \le m \le n$, характеризующих размеры тайла.

Пусть $n = 2$, J_2 кратно P и r_2 кратно r_1. Применяя LPGS-стратегию отображения в соответствии с методикой, описанной в [8], получаем, что полное время решения задачи не превосходит значения T:

$$T(P, t_0, a, b, I_1, I_2, r_1, r_2) =$$

$$= t(j^{\max}) = \tau_2 (J_2 - 1) + \tau_1 \left(\frac{r_2}{r_1} (J_2 - 1) + l - 1 \right) + (\Gamma - 1) \max \left\{ 0, \tau_1 \left(l - \frac{r_2}{r_1} P \right) - \tau_2 P \right\},$$

где $\overline{\tau} = (\max(T_{\text{comp}}, T_{\text{comm}}), T_{\text{comp}} + T_{\text{comm}})$, $l = 1 + \frac{r_2}{r_1} + \left\lfloor \frac{I_2 - 2}{r_1} \right\rfloor$, $\Gamma = J_2 / P$, $T_{\text{comm}} =$ $= T_{\text{comm}}^{\beta^1} + T_{\text{comm}}^{\beta^3}$.

При построении оценки $T(P, t_0, a, b, I_1, I_2, r_1, r_2)$ учтена возможность совмещения процесса передачи с процессом выполнения операций тайлов, отображенных на один и тот же вычислительный узел. Время передачи данных между тайлами, отображенными на один узел, принято равным 0, так как физически данные остаются в памяти процессорного элемента.

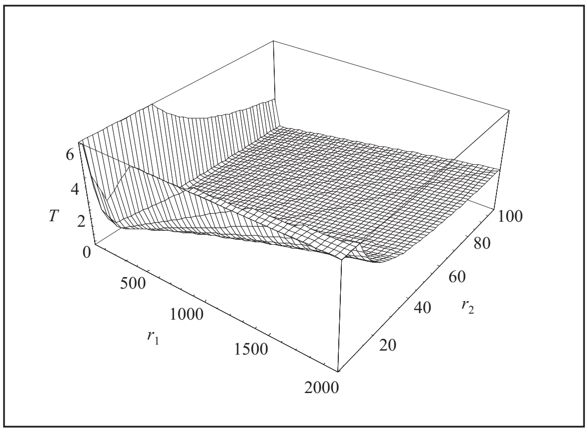

Рис. 1

Минимизация функции $T(P, t_0, a, b, I_1, I_2, r_1, r_2)$ позволяет получить алгоритм с наилучшим временем реализации на заданном кольце процессоров. В частности если принять значения параметров $P = 10$, $t_0 = 10^{-9}$ с, $a = 10^{-6}$ с, $b = 10^{-8}$ с, $I_1 = 2 \cdot 10^6$ и $I_2 = 10^3$, то оптимальное значения функции T достигаются при $r_1 = r_2 = 56$ и равно 0,207 с. На рис. 1 изображена поверхность значений функции T в зависимости от параметров тайла r_1 и r_2.

В случае $n = 3$ применение LPGS-стратегии отображения по методике [8] позволяет оценить время решения задачи функцией

$$T(P, t_0, a, b, I_1, I_2, I_3, r_1, r_2, r_3) =$$

$$= \tau_2 (J_2 - 1) + \tau_3 (J_3 - 1) + (\Gamma - 1) \max \left\{ 0, \tau_1 \left(l J_3 + \frac{r_2}{r_1} (J_2 - 1) \right) - \tau_2 P \right\},$$

где

$$\tau_1 = \max \left(t_0 \frac{r_1 r_2 r_3}{4}, a + b \left(\frac{r_1 r_2}{2} - 2 \right) \right), \ \tau_2 = t_0 \frac{r_1 r_2 r_3}{4} + a + b \left(\frac{r_1 r_2}{2} - 2 \right), \ \tau_3 = l \tau_1,$$

$$J_2 = \lceil (I_1 + I_2 + I_3 - 2) / r_2 \rceil, \ J_3 = \lceil (I_1 + I_2 + I_3 - 2) / r_3 \rceil, \ \Gamma = J_2 / P,$$

$$l = 2 + \frac{r_2}{r_1} + 2 \left\lfloor \frac{I_2 - 2}{r_1} \right\rfloor.$$

СПИСОК ЛИТЕРАТУРЫ

1. Ramanujam J., Sadayappaan P. Tiling of iteration spaces for multicomputers // Proc. of Int. Conf. on Parallel Proces. — 1990. — **2**. — P. 179–186.
2. Boulet P., Darte A., Risset T., Robert Y. (Pen)-ultimate tiling? // Integration, The VLSI J. — 1994. — **17**. — P. 33–51.
3. Xue J. Communication-minimal tiling of uniform dependence loops // J. of Parallel and Distributed Comput. — 1997. — **1**, N 42. — P. 42–59.
4. Schreiber R., Dongarra J. Automatic blocking of nested loops // Tech. rep. 90.38, RIACS. 1997.
5. Ohta H., Saito Y., Kainaga M., Ono H. Optimal tile size adjustment in compiling general DOACROSS loop nests // Proc. of Int. Conf. on Supercomput. ACM Press, 1995. — P. 270–279.
6. Hodzic E., Shang W. On supernode transformation with minimized total running time // IEEE Trans. Parallel and Distributed Systems. — 1998. — **9**, N 5. — P. 417–428.
7. Hodzic E., Shang W. On-time optimal supernode shape // IEEE Trans. Parallel and Distributed Systems. — 2002. — **13**, N 10. — P. 1220–1223.
8. Баханович С.В., Соболевский П.И. Отображение алгоритмов на вычислительные системы с распределенной памятью: оптимизация тайлинга для одно- и двумерных топологий // Весці НАН Беларусі. Сер. фіз.-мат. навук. — 2006. — № 2. — С. 106–112.
9. Баханович С.В., Соболевский П.И. Оптимизация тайлинга при LSGP стратегии отображения алгоритмов на суперкомпьютеры с распределенной памятью // Там же — 2007. — № 3. — С. 113–118.

Поступила 24.07.2008