XII Международная конференция

РАЗВИТИЕ ИНФОРМАТИЗАЦИИ И ГОСУДАРСТВЕННОЙ СИСТЕМЫ НАУЧНО-ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ

РИНТИ-2013

20 ноября 2013 года, Минск

ДОКЛАДЫ

Объединенный институт проблем информатики
Национальной академии наук Беларуси

XII Международная конференция

РАЗВИТИЕ ИНФОРМАТИЗАЦИИ И ГОСУДАРСТВЕННОЙ СИСТЕМЫ НАУЧНО-ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ

РИНТИ-2013

20 ноября 2013 года, Минск

Доклады

Минск
ОИПИ НАН Беларуси
2013

УДК 002; 004

Развитие информатизации и государственной системы научно-технической информации (РИНТИ-2013) : доклады XII Международной конференции, Минск, 20 ноября 2013 г. – Минск : ОИПИ НАН Беларуси, 2013. – 412 с. – ISBN 978-985-6744-78-8.

Представлены доклады XII Международной конференции «Развитие информатизации и государственной системы научно-технической информации» (РИНТИ-2013), Минск, 20 ноября 2013 г., в которых рассматриваются вопросы научно-методического, информационного и технологического обеспечения развития информатизации, создания и развития автоматизированных систем научно-технической информации, корпоративных библиотечно-информационных систем и технологий, а также направления использования информационно-коммуникационных технологий и результаты научных исследований в данных областях.

Материалы конференции будут полезны специалистам в области информационно-коммуникационных технологий, занимающимся разработкой и внедрением автоматизированных информационных систем управления, развитием информационной инфраструктуры Беларуси, реализацией проектов государственных программ в сфере информатизации и систем научно-технической информации.

Одобрены программным комитетом и печатаются по решению редакционной коллегии Объединенного института проблем информатики Национальной академии наук Беларуси в виде, представленном авторами.

Научные редакторы:

доктор физико-математических наук, профессор А.В. Тузиков;
кандидат технических наук, доцент Р.Б. Григянец;
кандидат технических наук В.Н. Венгеров

ISBN 978-985-6744-78-8

УДК 681.3.06:519

ОРГАНИЗАЦИЯ ОБРАБОТКИ ДАННЫХ В МНОГОПРОЦЕССОРНЫХ СИСТЕМАХ МАКРОКОНВЕЙЕРНОГО ТИПА

Н.С. Коваленко[1], В.Н. Венгеров[2] , П.А. Павлов[3]
[1]Белорусский государственный экономический университет, Минск;
[2]Объединенный институт проблем информатики НАН Беларуси, Минск;
[3] Полесский государственный университет, Пинск, Беларусь

Предложены математическая модель организации вычислений неоднородных конкурирующих процессов в многопроцессорных системах макроконвейерного типа и решение задач определения характеристик данной организации вычислений по времени реализации процессов при ограниченном числе каналов обмена. Для данной модели получены аналитические методы и формулы расчета минимального общего времени реализации заданных объемов вычислений при ограниченном числе каналов обмена в условиях асинхронного режима взаимодействия процессов, процессоров и каналов.

Введение

В настоящее время одной из наиболее перспективных концепций параллельной обработки является макроконвейерная организация вычислений над структурами данных. Этот подход впервые предложен в начале 1980-х гг. академиком В.М. Глушковым и его учениками. Основная идея концепции *макроконвейерной* организации вычислений заключается в том, что при распараллеливании и распределении вычислений между процессорами (процессорными узлами) «каждому отдельному процессору на очередном шаге вычислений дается такое задание, которое позволяет ему длительное время работать автономно без взаимодействия с другими процессорами» [1]. Уменьшение числа и объемов обмена сообщениями, которыми обмениваются параллельно работающие узлы, как правило, приводит к уменьшению общего времени выполнения заданных обьемов вычислений, что является одним из главных критериев качества распараллеливания вычислений. Интерес к данной концепции постоянно растет в связи с развитием и широким применением локальных и глобальных сетей, созданием вычислительных многопроцессорных систем (МС) и комплексов, сетевого аппаратного и прикладного программного обеспечения.

1. Метод структурирования программных ресурсов и макроконвейерная обработка

Структурирование (декомпозиция) – это основной способ уменьшения сложности больших задач, программ, систем и т. д., основная идея которого состоит в обеспечении специального способа структурирования программного ресурса на блоки $Q_1, Q_2, ..., Q_s$ и организации параллельного использования этих блоков множеством конкурирующих процессов [2].

Макроконвейерная технология вычислений предполагает декомпозицию структуры данных на большие информационно слабозависимые подструктуры, способные занимать процессор длительное время. Работа процессоров при этом организуется таким образом, чтобы обмен данными между ними занимал небольшое время по сравнению с временем вычислений.

Пусть PR – программный ресурс, который могут использовать два и более конкурирующих процессов, причем их число $n \geq 2$; $p \geq 2$ – число процессоров макроконвейерной системы, имеющих как локальную, так и общую для всех процессоров память. Применительно к программным ресурсам, одновременно используемым множеством процессов, при макроконвейерной обработке возможны следующие два способа организации вычислений:

– каждому i-му процессу, $i = \overline{1,n}$, предоставляется отдельная копия программного ресурса PR. При такой стратегии в случае $p \geq n$ все n процессов могут выполняться одновременно при условии, что в МС достаточно памяти для размещения n копий программного ресурса (в случае с общей памятью) или память каждого процессора МС вмещает отдельную копию программного ресурса (в случае с распределенной памятью). Если же $p < n$, то возможна организация циклического выполнения n процессов группами по p;

– программный ресурс PR может быть структурирован на блоки $Q_1, Q_2, \ldots, Q_s$, а вычисления в этом случае организуются в соответствии с методом структурирования. Эта стратегия может применяться при организации вычислений в МС всякий раз, если имеются ограничения на оперативную память, как общую, так и память каждого процессора.

Заметим, что второй способ наиболее предпочтителен при организации вычислений в МС конвейерного и макроконвейерного типов с целью эффективного использования основных вычислительных ресурсов, а также при организации процессов в операционных системах, распараллеливании и конвейеризации циклов.

Пусть МС характеризуется следующими параметрами: p – число процессоров, каждый из которых имеет собственную локальную память, $p \geq 2$; k – число каналов, через которые каждый из процессоров имеет доступ к внешней памяти, общей для всех процессоров, $k \geq 1$.

Предполагается, что в МС выполняется n процессов, $n \geq 2$, каждый из которых состоит из s блоков обмена и s блоков счета, $s \geq 1$. Времена обмена и счета для каждого из процессов представлены в виде матриц $t = [t_{ij}]_{n \times s}$ и $T = [T_{ij}]_{n \times s}$ размерности $n \times s$, в которых i-е строки соответствуют i-му процессу.

Взаимодействие процессов с каналами и процессорами характеризуется следующими пятью условиями:

– к выполнению одновременно готовы p процессов из n;

– в каждый момент времени k процессов из n, одновременно протекающих в МС, выполняются синхронно, остальные в очереди ждут освобождения каналов;

– во время обмена каждый процесс монополизирует один и тот же канал, во время счета – процессор;

– очередной j-й блок счета на каждом процессоре выполняется только после завершения соответствующего j-го блока обмена, а каждый $(j+1)$-й блок обмена выполняется после завершения j-го блока счета;

– процессы считаются равноприоритетными, а режим работы каналов является циклическим.

Данные условия определяют *асинхронный* режим взаимодействия процессов, каналов и процессоров, который допускает как простои каналов из-за занятости процессоров, так и простои процессоров из-за занятости каналов обмена.

Рассмотрим решение задач получения математических соотношений для вычисления минимального общего времени реализации множества конкурирующих процессов при условии выполнения сформулированных выше предположений.

2. Время реализации асинхронных процессов в макроконвейерных системах с одним каналом обмена

Обозначим через $T_n(k)$ общее время выполнения всех n процессов, которые используют k каналов. Заметим, что при $p \geq k \geq n$ в рамках принятой модели макроконвейерных вычислений $T_n(k)$ составит величину $T_n(k) = T_n(n) = \max\limits_{1 \leq i \leq n} \sum\limits_{j=1}^{s} (t_{ij} + T_{ij})$. Если окажется, что $p > k > n$, то $k - n$ каналов будут не задействованы, а $p - n$ процессоров будут простаивать.

Пусть имеется один канал, т. е. $k = 1$. Предположим, что $n \leq p$. На рис. 1 приведена несовмещенная диаграмма Ганта, отображающая взаимодействие n процессов (номер процесса изображен справа в прямоугольнике) с одним каналом и p процессорами. Причем каждый процесс состоит из $2s$ блоков, $s \geq 1$, которые периодически повторяются в порядке обмена, счета. При этом осуществляется конвейеризация каждого из блоков счета по всем n процессорам, причем одновременно могут выполняться n блоков.

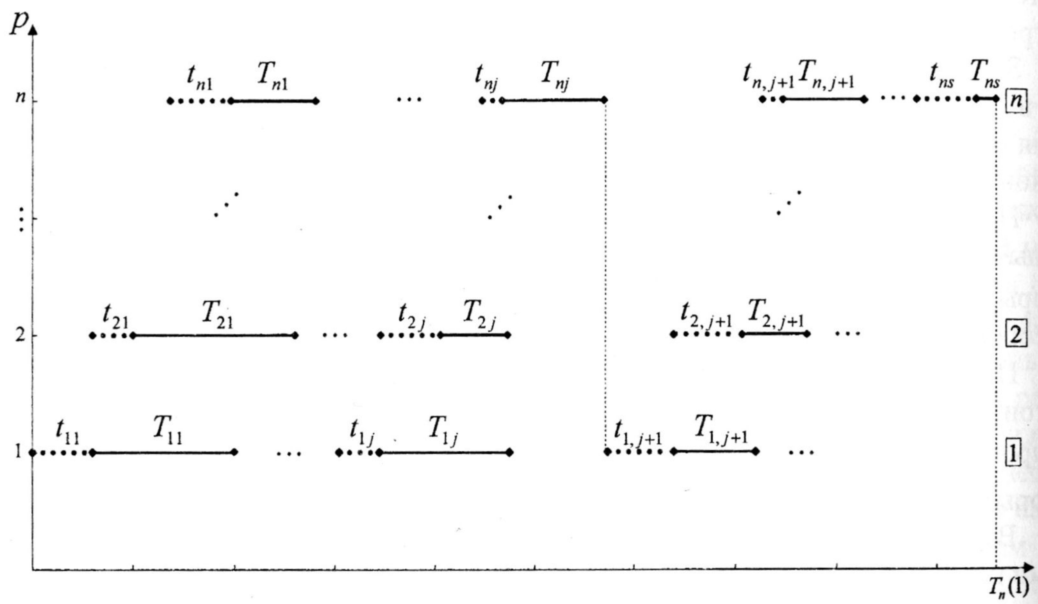

Рис. 1. Несовмещенная диаграмма Ганта с одним каналом обмена

Из анализа данной диаграммы Ганта следует, что $T_n(1)$ можно существенно сократить, если воспользоваться совмещением соседних диаграмм Ганта, начиная со второй, справа налево на максимально возможную величину, не нарушающую условий взаимодействия процессов с каналами и процессорами. Для этого необходимо составить расписание моментов начала выполнения *j*-го блока обмена, $j = \overline{1, s}$, для *i*-го процесса, $i = \overline{1, n}$ [3].

Анализируя две соседние диаграммы Ганта (рис. 1), соответствующие *j*-му и (*j*+1)-му блокам обмена и счета, с временами t_{ij}, T_{ij} и $t_{i,j+1}$, $T_{i,j+1}$ $i = \overline{1, n}$, $j = \overline{1, s-1}$, видно, что моменты начала выполнения обмена для каждого процесса определяются из соотношений:

– для первого блока

$$sb_{11} = 0,\ sb_{21} = sb_{11} + t_{11},\ \ldots,\ sb_{i1} = sb_{i-1,1} + t_{i-1,1},\ \ldots,\ sb_{n1} = sb_{n-1,1} + t_{n-1,1};$$

170

– *для второго блока*

$$sb_{12} = \max(sb_{11} + t_{11} + T_{11}, sb_{31} + t_{31}),\ sb_{22} = \max(sb_{21} + t_{21} + T_{21}, sb_{12} + t_{12}),\ \ldots,$$

$$sb_{i2} = \max(sb_{i1} + t_{i1} + T_{i1}, sb_{i-1,2} + t_{i-1,2}),\ \ldots,\ sb_{n2} = \max(sb_{n1} + t_{n1} + T_{n1}, sb_{n-1,2} + t_{n-1,2})\ \ldots;$$

– *для s-го блока*

$$sb_{1s} = \max(sb_{1,s-1} + t_{1,s-1} + T_{1,s-1}, sb_{3,s-1} + t_{3,s-1}),$$

$$sb_{2s} = \max(sb_{2,s-1} + t_{2,s-1} + T_{2,s-1}, sb_{1s} + t_{1s}),\ \ldots,$$

$$sb_{i3} = \max(sb_{i,s-1} + t_{i,s-1} + T_{i,s-1}, sb_{i-1,s} + t_{i-1,s}),\ \ldots,$$

$$sb_{ns} = \max(sb_{n,s-1} + t_{n,s-1} + T_{n,s-1}, sb_{n-1,s} + t_{n-1,s}).$$

Таким образом, имеет место следующая теорема.

Теорема 1. *Общее время выполнения n $(n \geq 2)$ процессов p $(p \geq 2)$ процессорами, конкурирующими за использование одного канала, в случае $n \leq p$ определяется по формуле*

$$T_n(1) = \max_{1 \leq i \leq n} (sb_{is} + t_{is} + T_{is}), \tag{1}$$

где sb_{ij} – моменты начала выполнения j-го блока обмена для i-го процесса, определяемые из соотношений

$$sb_{11} = 0,\ sb_{i1} = sb_{i-1,1} + t_{i-1,1},\ sb_{1j} = \max(sb_{1,j-1} + t_{1,j-1} + T_{1,j-1}, sb_{n,j-1} + t_{n.j-1}), \tag{2}$$

$$sb_{ij} = \max(sb_{i,j-1} + t_{i,j-1} + T_{i,j-1}, sb_{i-1,j} + t_{i-1,j}),\ i = \overline{2,n},\ j = \overline{2,s}.$$

В результате совмещения диаграмма Ганта будет иметь вид, показанный на рис. 2.

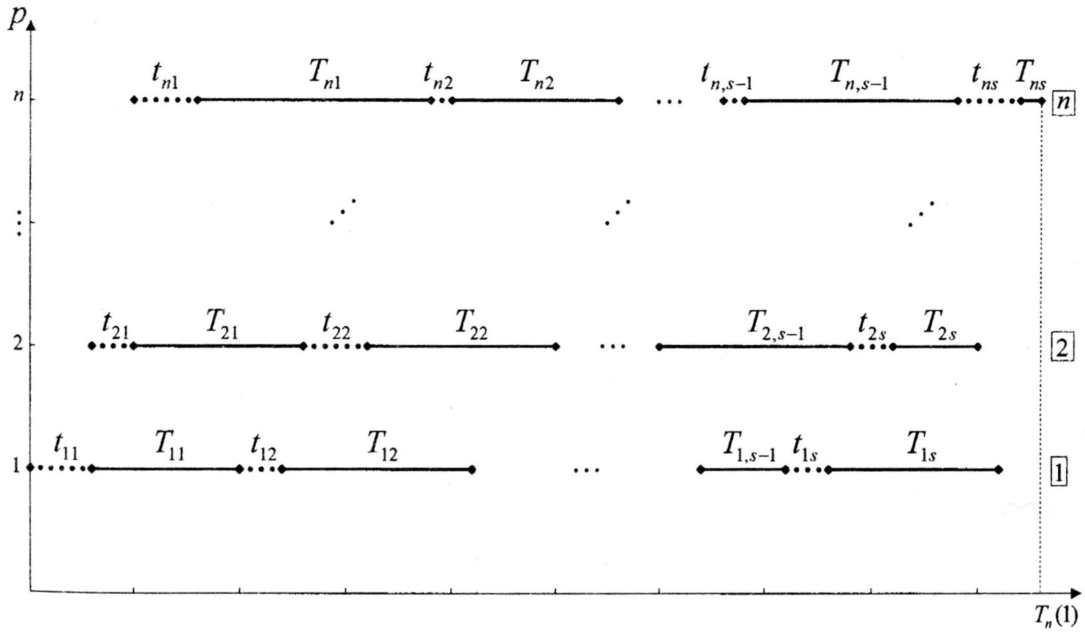

Рис. 2. Совмещенная диаграмма Ганта с одним каналом обмена

3. Макроконвейерные системы с ограниченным числом каналов обмена

В рамках концепции макроконвейерных вычислений из физических соображений наибольший интерес представляет случай ограниченного числа каналов, т. е. когда $k << n$, $n = mk$, $m > 1$, что означает, что процессы конкурируют за использование каналов. Будем считать, что $n \le p$. Рассмотрим следующие способы взаимодействия процессов с каналами и процессорами.

При *первом способе* каждый g-й канал, $g=\overline{1,k}$, обслуживает очередные m процессов, которые выполняются на m процессорах, т. е. *1-й канал* обслуживает процессы с номерами 1, 2, …, m; *2-й* – с номерами $m+1$, $m+2$, …, $2m$; *k-й* – с номерами – $(k-1)m+1$, $(k-1)m+2$, …, n. На рис. 3 показано взаимодействие процессов с каналами и процессорами после совмещения.

Рис. 3. Первый способ взаимодействия процессов, каналов и процессоров

Теорема 2. *Общее время выполнения p процессорами* $(p \ge 2)$ $n = mk$ $(m>1)$ *процессов, которые конкурируют за использование k каналов* $(k \ge 1)$, *в случае* $n \le p$ *определяется из соотношения*

$$T_n(k) = \max_{1\le g\le k} T_m^g(1) = \max_{1\le g\le k}\left(\max_{(g-1)m+1\,\le\, i\,\le\, gm}(sb_{is}+t_{is}+T_{is})\right), \tag{3}$$

где sb_{ij} *– моменты начала выполнения j-го блока обмена для i-го процесса, определяемые из соотношений*

$$\begin{gathered} sb_{gm+1,1}=0,\ \ g=\overline{0,k-1},\ sb_{i1}=sb_{i-1,1}+t_{i-1,1}, \\ sb_{(g-1)m+1,j}=\max(sb_{(g-1)m+1,j-1}+t_{(g-1)m+1,j-1}+T_{(g-1)m+1,j-1}, sb_{mg,j-1}+t_{mg,j-1}), \\ sb_{ij}=\max(sb_{i,j-1}+t_{i,j-1}+T_{i,j-1}, sb_{i-1,j}+t_{i-1,j}),\ i=\overline{(g-1)m+2,gm},\ j=\overline{2,s},\ g=\overline{1,k}. \end{gathered} \tag{4}$$

Доказательство теоремы следует из того факта, что каждый g-й канал, $g=\overline{1,k}$, согласно формулам (1)–(2), обслуживает группу из m процессов m процессорами за время $T_m^g(1)$, определяемое по следующим формулам:

$$T_m^1(1)=\max_{1\le i\le m}(sb_{is}+t_{is}+T_{is}),$$

где $sb_{11}=0$, $sb_{i1}=sb_{i-1,1}+t_{i-1,1}$, $sb_{1j}=\max(sb_{1,j-1}+t_{1,j-1}+T_{1,j-1},sb_{mg,j-1}+t_{mg,j-1})$,

$$sb_{ij}=\max(sb_{i,j-1}+t_{i,j-1}+T_{i,j-1},sb_{i-1,j}+t_{i-1,j}),\ i=\overline{2,m},\ \ j=\overline{2,s};$$

$$T_m^2(1)=\max_{m+1\le i\le 2m}(sb_{is}+t_{is}+T_{is}),$$

где $sb_{m+1,1}=0$, $sb_{i1}=sb_{i-1,1}+t_{i-1,1}$, $sb_{m+1,j}=\max(sb_{m+1,j-1}+t_{m+1,j-1}+T_{m+1,j-1},sb_{2m,j-1}+t_{2m,j-1})$,

$$sb_{ij}=\max(sb_{i,j-1}+t_{i,j-1}+T_{i,j-1},sb_{i-1,j}+t_{i-1,j}),\ i=\overline{m+2,2m},\ \ j=\overline{2,s};\ \ldots;$$

$$T_m^k(1)=\max_{(k-1)m+1\le i\le mk}(sb_{is}+t_{is}+T_{is}),$$

где $sb_{(k-1)m+1,1}=0$, $sb_{i1}=sb_{i-1,1}+t_{i-1,1}$,

$$sb_{(k-1)m+1,j}=\max(sb_{(k-1)m+1,j-1}+t_{(k-1)m+1,j-1}+T_{(k-1)m+1,j-1},sb_{mk,j-1}+t_{mk,j-1}),$$

$$sb_{ij}=\max(sb_{i,j-1}+t_{i,j-1}+T_{i,j-1},sb_{i-1,j}+t_{i-1,j}),\ i=\overline{(k-1)m+2,mk},\ \ j=\overline{2,s}.$$

Следовательно, общее время $T_n(k)$ будет равно $\max\limits_{1\le g\le k} T_m^g(1)$.

При *втором способе* взаимодействия процессов, каналов и процессоров все множество из n процессов разбивается на k групп по m процессов в каждой. Причем каждый g-й канал, $g=\overline{1,k}$, обслуживает группу из m процессов с номерами $(l-1)k+g$, где $l=\overline{1,m}$.

В этом случае согласно формулам (1)–(2) время, затраченное на выполнение каждой группы из m процессов m процессорами для каждого g-го канала, $g=\overline{1,k}$, составит

$$T_m^1(1)=\max_{1\le l\le m}(sb_{[(l-1)k+1],s}+t_{[(l-1)k+1],s}+T_{[(l-1)k+1],s}),$$

где $sb_{11}=0$, $sb_{lk+1,1}=sb_{[(l-1)k+1],1}+t_{[(l-1)k+1],1}$,

$$sb_{[(l-1)k+1],j}=\max(sb_{[(l-1)k+1],j-1}+t_{[(l-1)k+1],j-1}+T_{[(l-1)k+1],j-1},sb_{lk+1,j-1}+t_{lk+1,j-1}),$$

$$sb_{lk+1,j}=\max(sb_{lk+1,j-1}+t_{lk+1,j-1}+T_{lk+1,j-1},sb_{[(l-1)k+1],j}+t_{[(l-1)k+1],j}),\ l=\overline{1,m-1},\ \ j=\overline{2,s};$$

$$T_m^2(1)=\max_{1\le l\le m}(sb_{[(l-1)k+2],s}+t_{[(l-1)k+2],s}+T_{[(l-1)k+2],s}),$$

где $sb_{21}=0$, $sb_{lk+2,1}=sb_{[(l-1)k+2],1}+t_{[(l-1)k+2],1}$,

$$sb_{[(l-1)k+2],j}=\max(sb_{[(l-1)k+2],j-1}+t_{[(l-1)k+2],j-1}+T_{[(l-1)k+2],j-1},sb_{lk+2,j-1}+t_{lk+2,j-1}),$$

$$sb_{lk+2,j}=\max(sb_{lk+2,j-1}+t_{lk+2,j-1}+T_{lk+2,j-1},sb_{[(l-1)k+2],j}+t_{[(l-1)k+2],j}),\ l=\overline{1,m-1},\ \ j=\overline{2,s};\ \ldots;$$

$$T_m^k(1)=\max_{1\le l\le m}(sb_{lk,s}+t_{lk,s}+T_{lk,s}),$$

где $sb_{k1}=0$, $sb_{(l+1)k,1}=sb_{lk,1}+t_{lk,1}$, $sb_{lk,j}=\max(sb_{lk,j-1}+t_{lk,j-1}+T_{lk,j-1},sb_{(l+1)k,j-1}+t_{(l+1)k,j-1})$,

$$sb_{(l+1)k,j}=\max(sb_{(l+1)k,j-1}+t_{(l+1)k,j-1}+T_{(l+1)k,j-1},sb_{lk,j}+t_{lk,j}),\ l=\overline{1,m-1},\ \ j=\overline{2,s}.$$

Таким образом, справедлива следующая теорема.

Теорема 3. *Общее время выполнения p процессорами* $(p \geq 2)$ $n = mk$ $(m > 1)$ *процессов, которые конкурируют за использование k каналов* $(k \geq 1)$, *в случае* $n \leq p$ *определяется из соотношения*

$$T_m^g(1) = \max_{1 \leq l \leq m} (sb_{[(l-1)k+g],s} + t_{[(l-1)k+g],s} + T_{[(l-1)k+g],s}),$$

где sb_{ij} *– моменты начала выполнения j-го блока обмена для i-го процесса, определяемые из соотношений*

$$sb_{g1} = 0,\ sb_{lk+g,1} = sb_{[(l-1)k+g],1} + t_{[(l-1)k+g],1},$$

$$sb_{[(l-1)k+g],j} = \max(sb_{[(l-1)k+g],j-1} + t_{[(l-1)k+g],j-1} + T_{[(l-1)k+g],j-1}, sb_{lk+g,j-1} + t_{lk+g,j-1}),$$

$$sb_{lk+g,j} = \max(sb_{lk+g,j-1} + t_{lk+g,j-1} + T_{lk+g,j-1}, sb_{[(l-1)k+g],j} + t_{[(l-1)k+g],j}),\ l = \overline{1,m-1},\ j = \overline{2,s},\ g = \overline{1,k}.$$

Заключение

Построенная модель организации макроконвейерных вычислений над структурами данных при ограниченном числе каналов обмена и разработанные аналитические методы расчета общего времени выполнения множества неоднородных конкурирующих процессов являются основой для постановки и решения ряда важных практических оптимизационных задач по расчету оптимальной балансировки числа процессоров и каналов, оптимизации числа блоков счета и обмена, минимизации общего времени выполнения процессов и др. Данная модель включает следующие основные параметры: число процессоров, процессов, блоков и каналов обмена; матрицы времен счета и обмена; условия взаимодействия процессов, процессоров, блоков и каналов обмена; синхронные и асинхронный режимы взаимодействия процессов, процессоров и каналов.

Список литературы

1. Капитонова, Ю.В. Математическая теория проектирования вычислительных систем / Ю.В. Капитонова, А.А. Летичевский. – М. : Наука, 1988. – 296 с.
2. Коваленко, Н.С. Математическое моделирование параллельных процессов / Н.С. Коваленко, П.А. Павлов. – Saarbrucken, Germany : LAP Lambert Academic Publishing GmbH, 2011. – 246 с.
3. Танаев, В.С. Теория расписаний. Групповые технологии / В.С. Танаев, М.Я. Ковалев, Я.М. Шафранский. – Минск : Ин-т техн. кибернетики НАН Беларуси, 1998. – 290 с.

174