

УДК 681.323

Задержка обработки информации в многопроцессорных вычислителях различной конфигурации

О.Ю. Аксенов

Создание многопроцессорных вычислительных устройств, предназначенных для обработки непрерывных потоков информации, связано с определением информационных связей между процессорами и банками памяти, а также с выбором типа используемой памяти. Особенно это актуально для встраиваемых устройств, когда коммуникационные возможности памяти различных типов влияют не только на вычислительные возможности всего устройства, но и на его сложность, стоимость и, главное, габариты. Ниже анализируется влияние конфигурации многопроцессорного вычислительного устройства на задержку обработки информации.

Рассмотрение ведется для следующих условий.

1. Задача, решаемая при обработке потока информации, допускает распараллеливание. Применительно к задачам обработки потока изображений это означает, что площадь каждого изображения при обработке может быть разбита на N равных независимых участков, либо процедура обработки может быть разделена на N параллельных ветвей.
2. Применяемую память по характеру взаимодействия с источником и приемником информации (например, с процессором), можно отнести к одному из следующих типов:
 - память с независимым доступом (память типа D) – обеспечивающая одновременную работу с двумя процессорами (двухпортовая);
 - разделяемая память (память типа S) – работающая с двумя процессорами в режиме разделения времени;
 - локальная память (память типа L) – работающая только с одним процессором.
3. На входе и на выходе рассматриваемого устройства установлена память типа D (D_{in} и D_{out} соответственно).
4. Входная информация (например, отдельный кадр видеопотока) приходит в многопроцессорное вычислительное устройство равномерно на некотором временном интервале.

Рассматриваются встраиваемые многопроцессорные вычислители, предназначенные для осуществления в реальном масштабе времени обработки потоков информации, формируемых, например, видеокамерой. Анализируется связь конфигурации вычислителей с задержкой обработки информации. Определяются подходы к выбору конфигурации вычислителя для различных условий обработки.

5. Многопроцессорное вычислительное устройство должно обеспечивать:

- обработку информации с темпом поступления;
- минимальную задержку обработки.

Предполагается, что требуемый темп обработки информации обеспечивается за счет одновременного использования нескольких однотипных процессоров.

Для оперативной обработки потока информации, например потока кадров, формируемого видеокамерой, можно использовать различные конфигурации устройства обработки. Ниже рассматриваются некоторые из них, обозначенные как схемы обработки информации “А”, “Б”, “В”, “Г”, “Д”. Символом B на схемах обозначены процессоры, символами D, S, L – банки памяти соответствующих типов.

Схема “А” (рис. 1) соответствует однопроцессорному вычислителю. Эта схема проигрывает по производительности всем остальным, рассмотренным ниже, имеющим в своем составе N процессоров. Схема “Б” (рис. 2) представляет собой конвейер с использованием между процессорами памяти типа D . Схема “В” (рис. 3) – конвейер с использованием памяти типа S .



Рис. 1. Схема “А” обработки информации

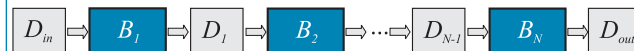


Рис. 2. Схема “Б” обработки информации

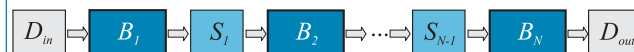


Рис. 3. Схема “В” обработки информации

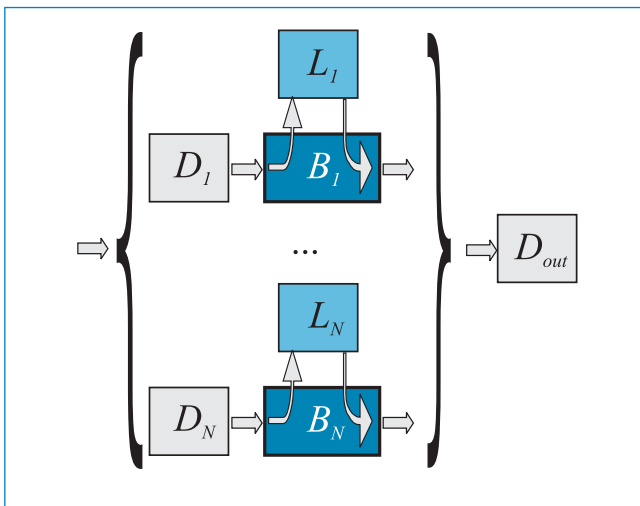


Рис. 4. Схема "Г" обработки информации

Рис. 4 представляет схему "Г" – параллельное соединение процессоров при обеспечении одновременной зачки входной информации во входные банки памяти каждого процессора ($D_1...D_N$). Такая схема целесообразна, если одну и ту же информацию (например, изображение) требуется обработать по различным алгоритмам. Для каждого (n -ого) процессора этой схемы характерны два этапа работы. На первом производятся собственно вычисления. Исходная информация берется из банка памяти D_n , а результат записывается в банк L_n . Второй этап – копирование из банка L_n в выходной банк D_{out} . Этот этап осуществляется в режиме программного разделения между процессорами времени доступа к памяти.

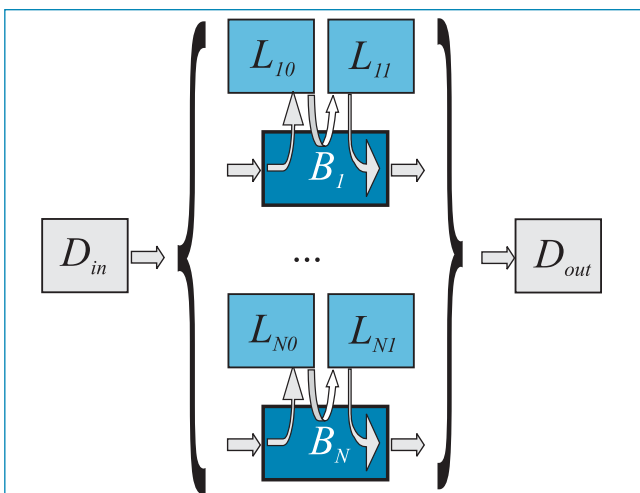


Рис. 5. Схема "Д" обработки информации

Схему "Д" отображает рис. 5. Эта схема предполагает, что каждый (n -ый) процессор обрабатывает только $1/N$ часть общей входной информации (кадра). При этом на первом этапе осуществляется копирование этой части информации в локальный банк памяти L_{n0} . На втором – собственно вычисления, при которых исходные данные берутся из банка памяти L_{n0} , а результат записывается в банк L_{n1} . На третьем этапе производится копирование из локального банка памяти L_{n1} в выходной банк D_{out} . Первый и третий этап тут осуществляются с программным разделением между процессорами времени доступа к памяти D_{in} и D_{out} соответственно.

Временные диаграммы работы рассмотренных схем обработки информации представлены на рис. 6–10. Используются следующие обозначения:

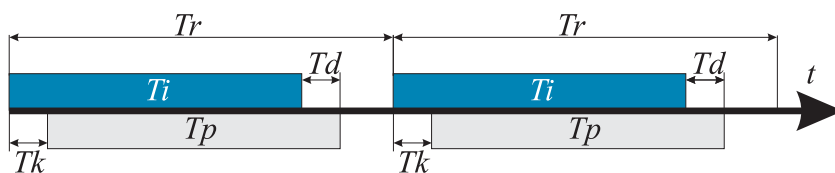


Рис. 6. Временная диаграмма обработки информации по схеме "А"

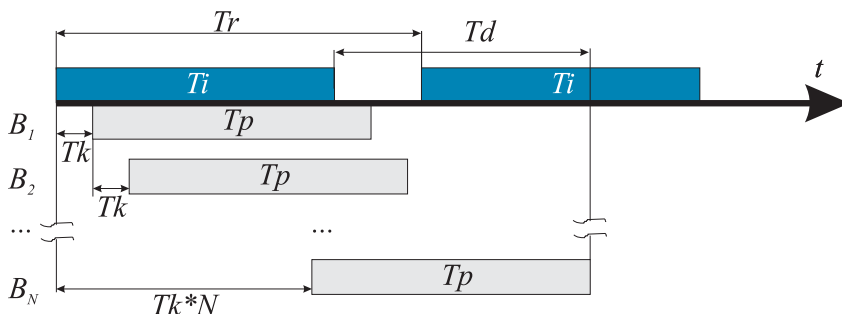


Рис. 7. Временная диаграмма обработки информации по схеме "Б"

T_r – период прихода входной информации;
 T_i – время прихода входной информации;
 T_p – время на обработку кадра одним процессором
 $T_s = N \cdot T_p$ – время на обработку кадра процессорами;
 T_c – время копирования кадра (переноса из одного банка памяти в другой);
 T_k – время прихода минимальной достаточной для начала обработки порции информации ();
 T_d – общая задержка обработки.

Анализ временных диаграмм работы позволяет оценить для рассматриваемых схем допустимый период поступления информации и задержку ее обработки. При этом число процессоров – N определяется исходя из того, чтобы обеспечить требуемый темп обработки: $N = T_s / T_p \Big|_{T_p = T_r} = T_s / T_r$. Соответствующие выражения содержит таб. 1.

Удобно использовать относительные значения времени, приняв за единицу измерения время обработки кадра одним процессором – T_p . Кроме того, исходя из условий обеспечения обработки непрерывного потока информации, целесообразно принять $T_i \approx T_p$. Тогда можно ввести коэффициенты, характеризующие:

$s = T_s / T_p$ – сложность задачи ($s \geq 1$);

$k = T_k / T_p$ – скорость прихода минимальной порции данных ($k \ll 1$);

$c = T_c / T_p$ – скорость копирования ($c < 1$);

$i = T_i / T_p$ – скорость прихода данных ($i \approx 1$).

Использование этих коэффициентов позволяет осуществить переход к относительным значениям допустимого периода поступления информации ($\tau = T_r / T_p$) и задержки обработки информации ($t = T_d / T_p$). В этом случае соотношения, входящие в таб. 1, примут вид, который отображает таб. 2.

Зависимость относительной задержки, обеспечиваемой различными схемами, для $s=4$; $c=0,1$; $i=0,9$ от k при N иллюстрирует рис. 11, а при $N=2$ и $N=4$ – рис. 12. Из представленных рисунков следует, например, что при $N=4$ и $k > 0,3$, наименьшую задержку обеспечивает схема “Д”, тогда как при меньших значениях k более выгодной становится схема “Б”. Ана-

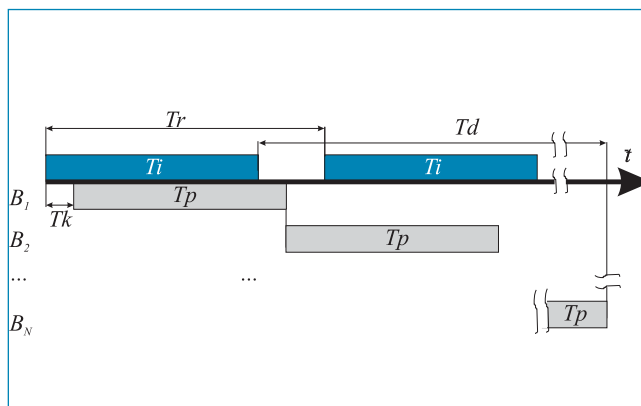


Рис. 8. Временная диаграмма обработки информации по схеме “В”

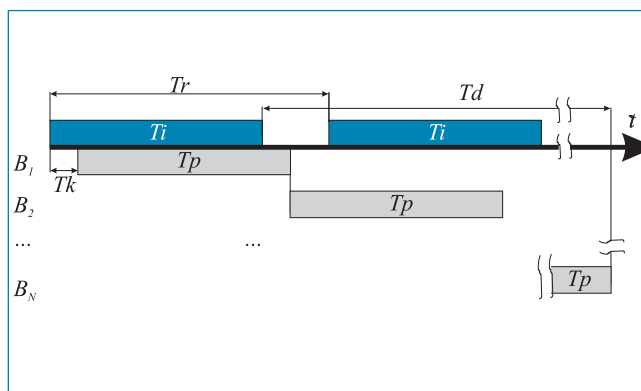


Рис. 9. Временная диаграмма обработки информации по схеме “Г”

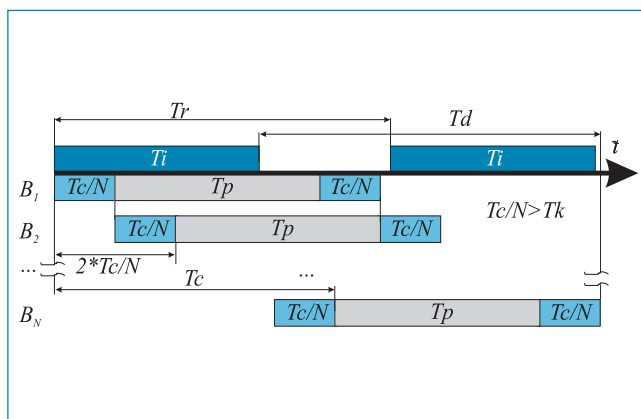


Рис. 10. Временная диаграмма обработки информации по схеме “Д”

Таблица 1

Схема	Период поступления	Задержка обработки	Сложность задачи
“А”	$T_r \geq T_k + T_p$	$T_d = T_k + T_p - T_i$	$T_s = T_p$
“Б”	$T_r \geq T_k + T_p$	$T_d = T_k \cdot N + T_p - T_i$	$T_s = N \cdot T_p$
“В”	$T_r \geq T_k + T_p$	$T_d = T_k + N \cdot T_p - T_i$	$T_s = N \cdot T_p$
“Г”	$T_r \geq T_k + T_p$	$T_d = T_k + T_p + T_c - T_i$	$T_s = N \cdot T_p$
“Д”	$T_r \geq T_p + 2 \cdot T_c / N$	$T_d = T_p + T_c \cdot (1 + 1/N) - T_i$	$T_s = N \cdot T_p$

Таблица 2

Схема	Относительный период поступления ($\tau = T_r / T_p$)	Относительная задержка обработки ($t = T_d / T_p$)
“А”	$\tau_A \geq k + s$	$t_A = k + s - i$
“Б”	$\tau_B \geq k + s / N$	$t_B = k \cdot N + s / N - i$
“В”	$\tau_B \geq k + s / N$	$t_B = k + s - i$
“Г”	$\tau_G \geq k + s / N$	$t_G = k + s / N + c - i$
“Д”	$\tau_D \geq (s + 2 \cdot c) / N$	$t_D = s / N + c \cdot (1 + 1/N) - i$

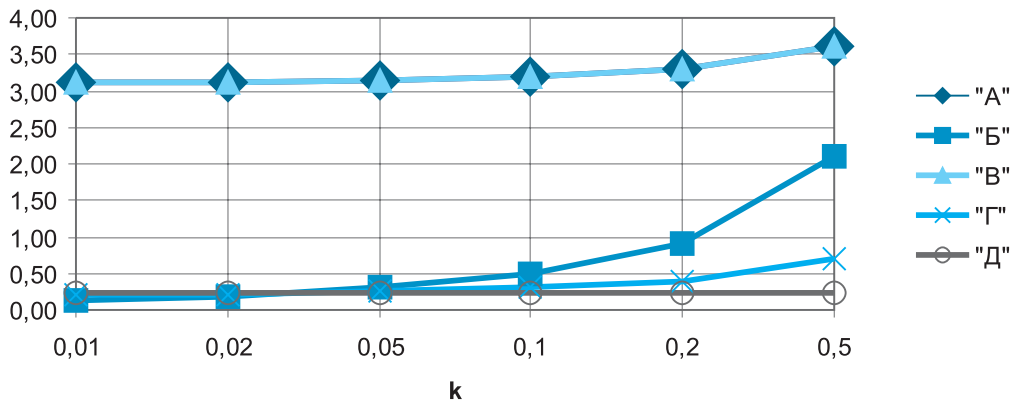


Рис. 11 Относительная задержка обработки для $s=4$; $c=0,1$; $i=0,9$; $N=4$.

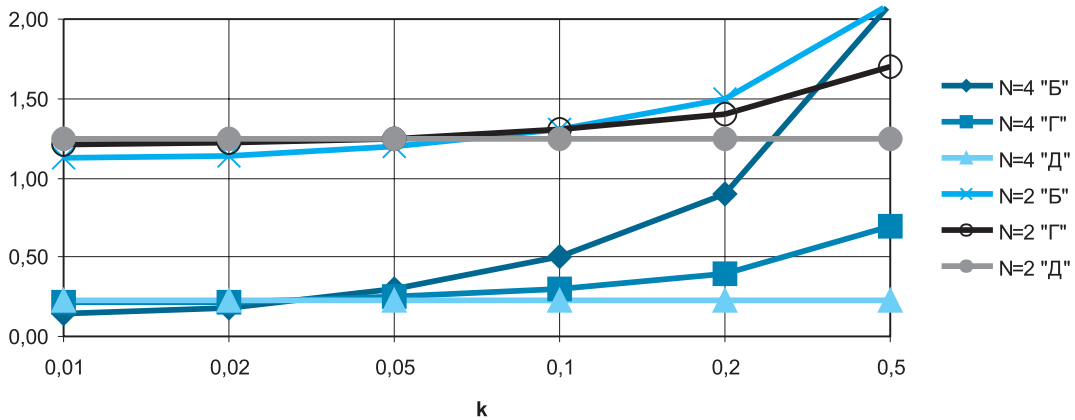


Рис. 12. Относительная задержка обработки для $s=4$; $c=0,1$; $i=0,9$.

логично для $N=2$ пограничным значением оказывается $k \approx 0,7$.

Сравнение выражений, представленных в таб. 2, позволяет определить значения параметров, при которых та или иная схема обеспечивает меньшую задержку или способна функционировать при меньшем значении периода поступления информации. Эти значения соответствуют точкам пересечения графиков. Так $\tau_B = \tau_G = \tau_D$ имеет место при $k\tau > 2 \cdot c/N$. Аналогично:

- $t_B > t_G$ при $k_{BG} > c/(N-1)$;
- $t_B > t_D$ при $k_{BD} > c \cdot (N+1)/N^2$;
- $t_A > t_B$ при $k_{AB} > s/N$;
- $t_G > t_D$ при $k_{GD} > c/N$.

В приведенные соотношения входит величина k , характеризующая минимальную порцию исходной информации, получив которую, можно начинать обработку. Эта величина определяется источником информации и способом обработки. Так, например, пусть требуется обработать изображение размером 1024×768 пикселей фильтром с апертурой 9×9 пиксе-

лов. Исходная информация поступает построчно. Значит, для начала обработки надо получить 8 строк плюс 9 первых элементов 9-ой строки. Отсюда $Tk/Ti = (1024 \cdot 8 + 9)/(1024 \cdot 768) \approx 0,01$. При $i \approx 1$ это означает, что $k \approx 0,01$. Тогда для схем "Б", "В", "Г", "Д" и случая $s=4$; $c=0,1$; $N=4$ наименьшую задержку обеспечит схема "Б". Однако, если исходная информация поступает полукадрами (сначала только четные строки, потом только нечетные), то $Tk/Ti = (1024 \cdot 4 + 9 + 1024 \cdot 768/2)/(1024 \cdot 768) \approx 0,5$, и при $k \approx 0,05$ выгоднее использовать схему "Д".

Следует отметить, также, что существенным моментом, определяющим на практике выбор схемы обработки является удобство («стоимость») аппаратной реализации памяти рассмотренных типов.

Таким образом, предложенный подход и полученные соотношения позволяют в конкретных случаях выбрать наилучшую (по критериям «наибольший темп поступления данных» или «наименьшая задержка обработки») схему обработки информации и соответственно конфигурацию многопроцессорного вычислителя.