

# ДЕКОМПОЗИЦИЯ СТРУКТУРЫ ГИБРИДНЫХ ПРОЦЕССОРНЫХ ЭЛЕМЕНТОВ И БАЛАНСИРОВКА ВЫЧИСЛЕНИЙ

*С.А. Степаненко*

Российский федеральный ядерный центр Всероссийский научно-исследовательский институт экспериментальной физики

## АННОТАЦИЯ

Ключевые слова: *гибридный процессорных элемент, оценки ускорения вычислений, MIMD, SIMD дисциплины вычислений.*

Предложен метод декомпозиции структуры гибридного процессорного элемента в соответствии с особенностями его конструктивной реализации и определенными первичными параметрами исполняемого вычислительного процесса. Метод позволяет получить минимальную в заданных условиях длительность вычислений.

Получены значения коэффициентов ускорения вычислений, достигаемых гибридным процессорным элементом, в частности, сформулированы аналитические условия достижения максимальных значений ускорений для заданной структуры вычислителя и данного вычислительного процесса.

Предложен основанный на полученных соотношениях метод балансировки (распределения) вычислительного процесса, позволяющий достигать максимального ускорения вычислений.

Приведены иллюстрации применения полученных соотношений, демонстрирующие возможности предложенного метода.

## ВВЕДЕНИЕ

В [1] показана необходимость применения гибридных процессорных элементов для создания эксафлопсных супер-ЭВМ. В составе гибридного процессорного элемента задействованы многоядерные универсальные процессоры, реализующие MIMD дисциплину вычислений, и арифметические ускорители (сопроцессоры), реализующие SIMD дисциплину вычислений.

В [2, 3] получены аналитические оценки длительностей вычислений, достигаемых гибридными процессорными элементами, и коэффициентов ускорения вычислений по сравнению с ядром универсального процессора.

В основе применяемого в [2, 3] метода – два параметра, характеризующих выполняемый процесс:  $\varphi$  - доля MIMD фрагмента в исполняемом процессе, и  $\rho$  - значение ускорения вычислений на SIMD фрагменте, достигаемое применением ускорителя; предполагается, что MIMD и SIMD фрагменты выполняются последовательно.

Полученные в [2, 3] оценки позволяют оценить длительности выполнения заданного процесса в результате задействования того или иного количества ядер и ускорителей. Кроме того, в указанных работах получены предельные значения ускорения, достигаемые при сколь угодно больших количествах ядер и ускорителей.

В этой работе исследуются применения полученных зависимостей и оценок в условиях практических ограничений – количество ядер универсального процессора и количество ускорителей в составе процессорного элемента фиксированы и конструктивно неизменны.

Для достижения максимального ускорения вычислений предлагается метод декомпозиции структуры процессорного элемента, предусматривающий выделение однородной подструктуры и гибридной подструктуры. Однородная подструктура состоит либо из ядер универсального процессора, либо из ускорителей. Гибридная подструктура содержит и универсальные процессоры и ускорители.

Составы однородной и гибридной подструктур определяются:

- первичными параметрами  $\varphi$  и  $\rho$  выполняемого процесса;
- количеством ядер и количеством ускорителей в процессорном элементе;
- производительностью ядра и производительностью ускорителя; значения производительности выражаются через длительности исполняемого процесса.

Предлагаемый метод декомпозиции структуры процессорного элемента и соответствующего этой декомпозиции размещения вычислительного процесса позволяет достигнуть наименьшей в заданных условиях длительности выполнения этого процесса, т.е. достигнуть наибольшей производительности.

В отличие от известных методов динамической балансировки – см., например, [4] – применяющих варьирование потока заданий (процессов), предлагаемый в этой работе метод предназначен для выполнения заданного процесса заданным элементом за минимально возможный временной интервал.

## 1 ОЦЕНКИ ДЛИТЕЛЬНОСТЕЙ ВЫЧИСЛЕНИЙ И КОЭФФИЦИЕНТОВ УСКОРЕНИЯ

Воспроизведем согласно [2, 3] основные обозначения и определения, используемые в этой работе.

Структура гибридного процессорного элемента приведена на рисунке 1, он содержит  $q$  ядер, выполняющих MIMD вычисления, и  $r$  ускорителей; каждый ускоритель выполняет SIMD вычисления. Коммутатор позволяет соединять любые подмножества ядер с любыми подмножествами ускорителей.

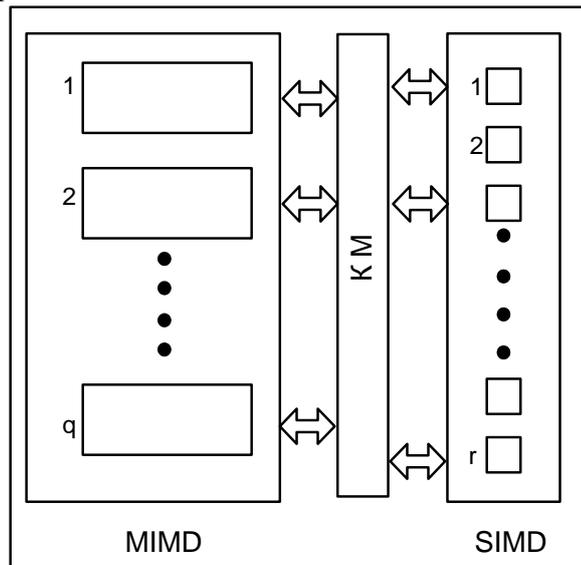


Рисунок 1 – Структура гибридного вычислителя

Длительность выполнения вычислительного процесса ядром универсального процессора обозначаем  $T_1$ . Длительность выполнения этого же процесса гибридным элементом, содержащим  $q$  ядер и  $r$  ускорителей, обозначаем  $T_{q,r}$ .

Различаем два режима распараллеливания:

- режим «умножения вычислений», предусматривающий увеличение в  $\max(q, r)$  раз по сравнению с одним ядром количества вычислений, выполняемых гибридным вычислителем из  $q$  ядер и  $r$  ускорителей;
- режим «деления вычислений», предусматривающий неизменность, по сравнению с одним ядром, общего количества вычислений, выполняемых гибридным вычислителем из  $q$  ядер и  $r$  ускорителей.

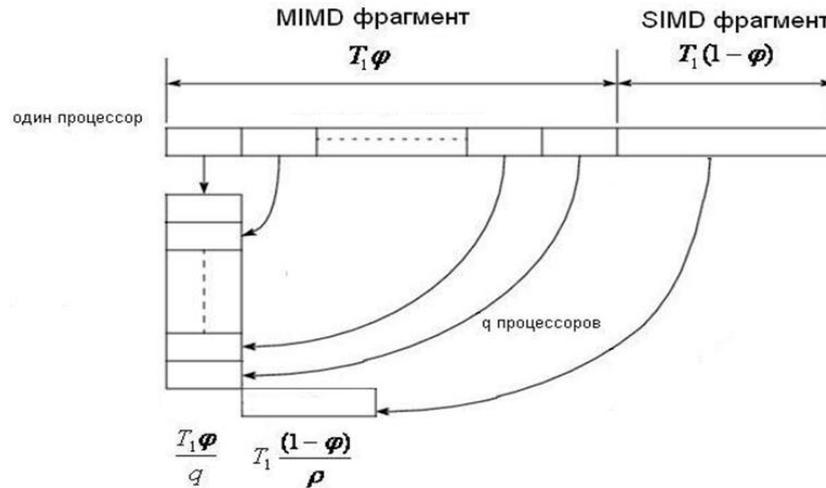
Коэффициенты ускорения вычислений для режима умножения и режима деления вычислений имеют вид:

$$K_{q,r} = \frac{qT_1}{T_{q,r}},$$

и, соответственно,

$$K_{q,r} = \frac{T_1}{T_{q,r}}.$$

При задействовании процессорного элемента вычислительный процесс, выполнявшийся одним ядром универсального процессора, представляется в виде двух последовательных фрагментов (см. рис.2): MIMD фрагмента, по-прежнему исполняемого ядром универсального процесса, и SIMD фрагмента, исполняемого ускорителем.



$$\hat{T}_{q,1} = \frac{T_1\varphi}{q} + T_1 \frac{1-\varphi}{\rho};$$

$$K_{q,1} = \frac{1}{\varphi + (1-\varphi)\frac{q}{\rho}}$$

$$\tilde{T}_{q,1} = T_1\varphi + T_1(1-\varphi)\frac{q}{\rho}.$$

**Рисунок 2** – Параллелизуемые фрагменты

Длительность выполнения одним ядром MIMD фрагмента равна  $T_M = T_1\varphi$ , где  $\varphi < 1$  - доля MIMD фрагмента,  $T_1$  - длительность исполнения процесса одним ядром. Длительность выполнения одним ускорителем SIMD фрагмента составляет  $T_S = \frac{T_1(1-\varphi)}{\rho}$ , т.е. SIMD фрагмент исполняется ускорителем в  $\rho$  раз быстрее, чем ядром универсального процесса.

Общая длительность выполнения процесса элементарным процессорным элементом, содержащим одно ядро и один ускоритель составляет:

$$T_{1,1} = \varphi T_1 + \frac{(1-\varphi)T_1}{\rho}.$$

Коэффициент ускорения вычислений элементарным процессорным элементом по сравнению с одним ядром составляет (для одного подпроцесса):

$$K_{1,1} = \frac{T_1}{T_{1,1}} = \frac{1}{\varphi + \frac{1-\varphi}{\rho}} = \frac{\rho}{\rho\varphi + 1 - \varphi} = \frac{\rho}{1 + \varphi(\rho - 1)}.$$

Коэффициент ускорения вычислений, по сравнению с одним ядром, достигаемый гибридным процессорным элементом, содержащим  $q$  ядер и один ускоритель, обозначим  $K_{q,1}$ , а коэффициент ускорения для элемента из 1 ядра и  $r$  ускорителей обозначим  $K_{1,r}$ .

Оценки длительностей вычислений и коэффициентов ускорения, включая их максимальные предельные значения, получены в [2, 3]. Эти оценки приведены в таблице 1 и таблице 2.

**Таблица 1** – Оценки длительностей вычислений

| Режим умножения  | Режим деления   | Состав вычислителя                                   |
|--|---|--|
| $\tilde{T}_{q,1}^* = T_1\varphi + T_1(1 - \varphi) \frac{q}{\rho}$   | $\hat{T}_{q,1}^* = T_1 \frac{\varphi}{q} + T_1(1 - \varphi) \frac{1}{\rho}$   | $q$ ядер, 1 ускоритель                               |
| $\tilde{T}_{1,r}^* = T_1\varphi r + T_1(1 - \varphi) \frac{1}{\rho}$ | $\hat{T}_{1,r}^* = T_1\varphi + T_1(1 - \varphi) \frac{1}{r\rho}$             | 1 ядро, $r$ ускорителей                              |
| $\tilde{T}_{q,r}^* = T_1\varphi + T_1(1 - \varphi) \frac{\mu}{\rho}$ | $\hat{T}_{q,r}^* = T_1 \frac{\varphi}{\mu} + T_1(1 - \varphi) \frac{1}{\rho}$ | $q$ ядер, $r$ ускорителей<br>$\frac{q}{r} = \mu > 1$ |

**Таблица 2** - Значения коэффициентов ускорения

| Режим умножения   | Режим деления   | Состав вычислителя                                |
|---|---|---|
| $\tilde{K}_{q,1} = \frac{q}{\varphi + (1 - \varphi) \frac{q}{\rho}}$<br>$\tilde{K}_{q,1} = \frac{\rho}{1 - \varphi}$ , при $q \rightarrow \infty$<br>$\tilde{K}_{q,1} \geq q$ , при $q < \rho$                        | $\hat{K}_{q,1} = \frac{q}{\varphi + (1 - \varphi) \frac{q}{\rho}}$<br>$\hat{K}_{q,1} = \frac{\rho}{1 - \varphi}$ , при $q \rightarrow \infty$<br>$\hat{K}_{q,1} \geq q$ , при $q < \rho$                        | $q$ ядер, 1 ускоритель                            |
| $\tilde{K}_{1,r} = \frac{r}{\varphi r + \frac{1 - \varphi}{\rho}}$<br>$\tilde{K}_{1,r} = \frac{1}{\varphi}$ , при $r \rightarrow \infty$<br>$\tilde{K}_{r,1} \geq r$ , при $\rho > \frac{1 - \varphi}{1 - \varphi r}$ | $\hat{K}_{1,r} = \frac{r}{\varphi r + \frac{1 - \varphi}{\rho}}$<br>$\hat{K}_{1,r} = \frac{1}{\varphi}$ , при $r \rightarrow \infty$<br>$\hat{K}_{r,1} \geq r$ , при $\rho > \frac{1 - \varphi}{1 - \varphi r}$ | 1 ядро, $r$ ускорителей                           |
| $\tilde{K}_{q,q} = \frac{q}{\varphi + \frac{1 - \varphi}{\rho}}$  | $\hat{K}_{q,q} = \frac{q}{\varphi + \frac{1 - \varphi}{\rho}}$  | $q$ ядер, $q$ ускорителей                         |
| $\tilde{K}_{q,1} > \tilde{K}_{1,r}$ при $\rho > \frac{1 - \varphi}{\varphi}$  | $\hat{K}_{q,1} > \hat{K}_{1,r}$ при $\rho > \frac{1 - \varphi}{\varphi}$  | $q$ ядер, 1 ускоритель<br>1 ядро, $r$ ускорителей |

Для обоих режимов при одинаковых составах вычислителей коэффициенты совпадают.

Целесообразность увеличения количества ядер или количества ускорителей, задействуемых в гибридном процессорном элементе для выполнения данного вычислительного процесса, определяется значениями  $\rho$  и  $\varphi$  этого процесса. В частности, если  $\rho > \frac{1 - \varphi}{\varphi}$ , то  $K_{q,1} > K_{1,r}$  и, следовательно, целесообразно увеличивать количество ядер.

С ростом  $q$  наиболее «быстро» растущим является  $K_{q,q}$ . В [3] показано, что применение реконфигурации – изменение количества ядер и количества взаимодействующих с ними ускорителей, позволяет для процессорного элемента, содержащего  $q$  ядер и  $q$  ускорителей, получить коэффициент ускорения  $K_R$  больший, чем  $K_{q,q}$ , но не более чем в два раза.

В продолжение полученных в [2, 3] оценок представляется необходимым оценить максимальное значение  $K_{q,r}$ , достигаемое при фиксированных  $q$ ,  $r$  и заданных  $\varphi$  и  $\rho$ , а также распределение вычислений по ядрам и ускорителям процессорного элемента, обеспечивающее этот максимум.

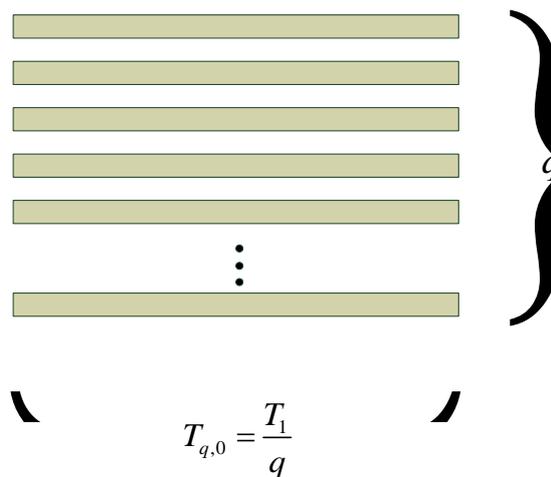
## 2 МЕТОД ДЕКОМПОЗИЦИИ СТРУКТУРЫ ГИБРИДНОГО ПРОЦЕССОРНОГО ЭЛЕМЕНТА

В этом разделе предлагается алгоритм распределения ядер и ускорителей, позволяющий для фиксированного состава оборудования процессорного элемента и заданных значений параметров вычислительного процесса, достигнуть наибольший коэффициент ускорения, т.е. получить наибольшую производительность этого процессорного элемента на данном процессе.

Процессорный элемент содержит  $q$  ядер и  $r$  ускорителей (см. рис. 1). Необходимо выполнить вычислительный процесс с параметрами  $\varphi$  и  $\rho$ , где  $\varphi$  - доля SIMD фрагмента и  $\rho$  - ускорение выполнения SIMD фрагмента, достигаемое применением ускорителя (см. рис. 2).

Пусть длительность выполнения этого процесса одним ядром универсального процессора составляет  $T_1$ .

Очевидно, если бы в процессорном элементе отсутствовали ускорители, то максимальное значение ускорения не превышало бы  $q$ . Одновременно  $q$  ядер выполняли бы  $q$  подпроцессов (см. рис. 3) и в принятых обозначениях  $T_{q,0} = \frac{T_1}{q}$ . В идеальном случае достигалось бы значение  $K_{q,0} = q$ .



**Рисунок 3** – Распределение вычислительного процесса по ядрам

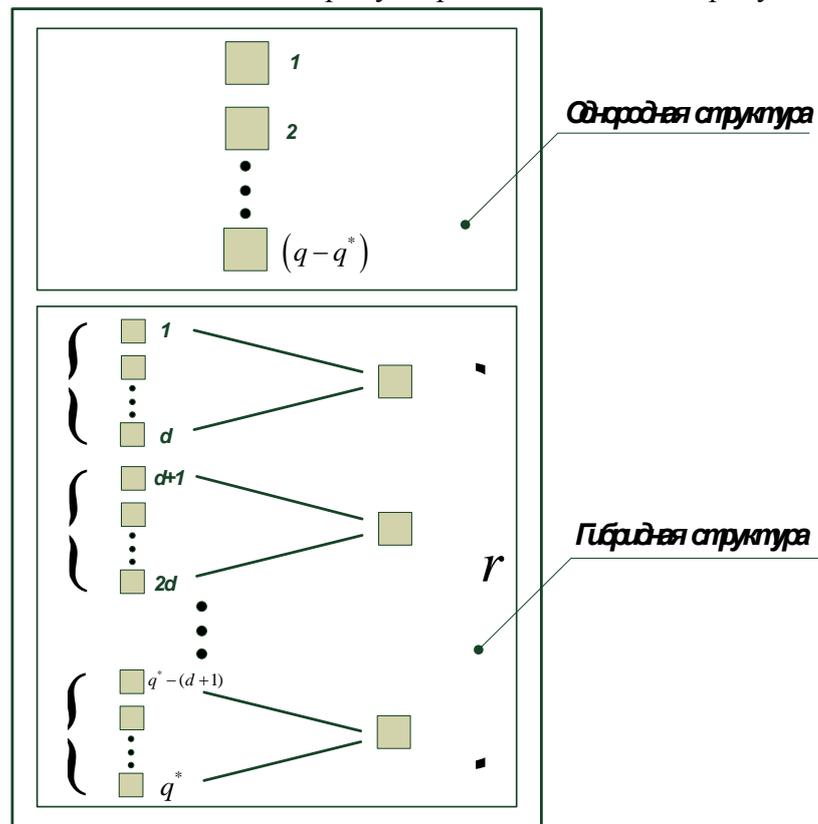
Применение ускорителей должно ускорить вычисления.

Целесообразность увеличения количества ядер или количества ускорителей, задействуемых в гибридном вычислителе для выполнения данного вычислительного

процесса, определяется значениями  $\rho$  и  $\varphi$  этого процесса; в частности, если  $\rho > \frac{1-\varphi}{\varphi}$ , то  $K_{q,1} > K_{1,r}$  и, следовательно, целесообразно увеличивать количество ядер.

На практике количество ядер и ускорителей в составе гибридного процессорного элемента фиксировано. Как правило, вследствие конструктивных особенностей, выполняется  $q > r$ . Начиная с некоторого значения  $q^* < q$ , прирост коэффициента ускорения при заданном  $r = const$  становится незначительным. Поэтому оказывается целесообразным образование в процессорном элементе двух структур. Одна – однородная MIMD структура из  $(q - q^*)$  универсальных ядер, другая – гибридная MIMD/SIMD структура, содержащая  $q^* < q$  ядер и  $r$  ускорителей, причем с каждым ускорителем взаимодействуют  $d = \frac{q^*}{r}$  ядер.

Этот вариант задействования ядер и ускорителей показан на рисунке 4.



**Рисунок 4** – Структурная декомпозиция процессорного элемента,  $r < q$

Распределим вычисления согласно структуре рисунка 4.

Коэффициент ускорения вычислений, достигаемый однородной универсальной компонентой, составит  $(q - q^*)$ .

Коэффициент ускорения  $K_{q^*,r}$ , достигаемый гибридной компонентой из  $q^*$  ядер и  $r$  ускорителей, составит -  $r \cdot K_{d,1}$ , где  $d = \frac{q^*}{r}$  - количество ядер в подструктуре, содержащей 1 ускоритель.

В [2] показано (см. таблицу 2):

$$K_{d,1} = \frac{d}{\varphi + (1 - \varphi) \frac{d}{\rho}} .$$

В результате получаем:

$$K_{q^*,r} = r \frac{d}{\varphi + (1 - \varphi) \frac{d}{\rho}}$$

Общее ускорение вычислений, достигаемое гибридным процессорным элементом, имеющим представленную на рисунке 4 структуру, имеет следующий вид:

$$K_{q,r}^D = q - q^* + r \frac{d}{\varphi + (1 - \varphi) \frac{d}{\rho}} = q - \nu q + \frac{\nu q}{\varphi + (1 - \varphi) \frac{\nu q}{r \rho}},$$

где  $\nu = \frac{q^*}{q}$ .

Аналогично, для общности, рассмотрим вариант, когда  $q < r$  и при этом  $\rho < \frac{1 - \varphi}{\varphi}$ .

В этом варианте, представленном на рис. 5, имеем:

$$K_{q,r}^D = q - q^* + q^* \frac{d}{\varphi d + (1 - \varphi) \frac{1}{\rho}},$$

где  $d = \frac{r}{q^*}$ .

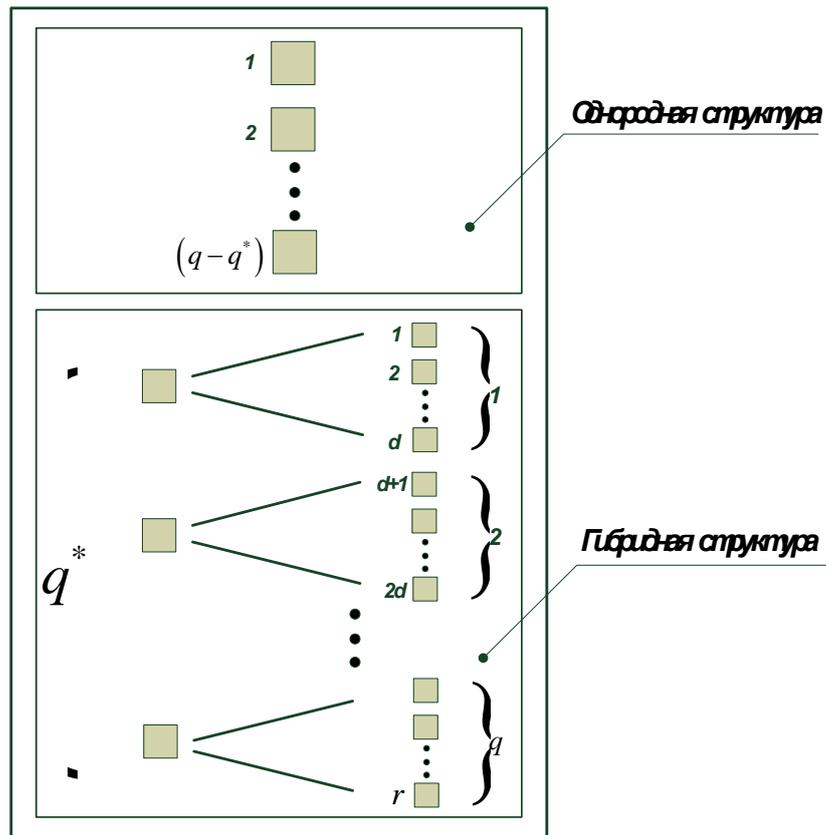


Рисунок 5 - Структурная декомпозиция процессорного элемента,  $r > q$

Далее получаем:

$$K_{q,r}^D = q - q^* + \frac{r}{\varphi \frac{r}{q^*} + (1 - \varphi) \frac{1}{\rho}} = q - \nu q + \frac{\nu q}{\varphi + (1 - \varphi) \frac{\nu q}{r \rho}}.$$

Итак, для обоих вариантов коэффициент ускорения  $K_{q,r}^D$ , имеет вид:

$$K_{q,r}^D = q - \nu q + \frac{\nu q}{\varphi + (1 - \varphi) \frac{\nu q}{r\rho}},$$

где

$$\nu = \frac{q^*}{q}$$

В приложении А показано, что  $K_{q,r}^D$  достигает максимального значения  $K_{q,r}^{\max}$  при

$$\nu = \frac{r\rho}{q} \frac{\sqrt{\varphi}}{(1 + \sqrt{\varphi})}.$$

Значение  $K_{q,r}^{\max}$  оценено в приложении Б; оно составляет:

$$K_{q,r}^{\max} = q + r\rho \frac{1 - \sqrt{\varphi}}{1 + \sqrt{\varphi}}.$$

Последнее выражение позволяет оценить максимально достижимое ускорение вычислений при заданных  $q$ ,  $r$  и  $\rho$ ,  $\varphi$ .

Возвращаясь к выражению  $K_{q,r}^D = q - \nu q + \frac{\nu q}{\varphi + (1 - \varphi) \frac{\nu q}{r\rho}}$ , заметим, что если

$q^* = \nu q > r\rho$ , то  $K^D < q$ .

Идентичный вывод, а именно  $K_{q,r}^D < q$  при  $q > r\rho$ , следующий из оценок таблицы 2, получаем для  $\nu = 1$ , когда  $q = q^*$ ; это означает задействование в структуре вычислителя только гибридной подструктуры.

Таким образом, совместное применение ядер и ускорителей нецелесообразно, если достигаемое применением ускорителя (сопроцессора) значение ускорения  $\rho$  таково, что выполняется неравенство  $q > r\rho$ . Коэффициент ускорения вычислений в данном случае меньше, чем достигаемый без ускорителей.

### 3 БАЛАНСИРОВКА ВЫЧИСЛИТЕЛЬНОГО ПРОЦЕССА

Балансировка вычислений сводится к распределению процесса по ядрам и ускорителям. Распределим вычисления в соответствии с полученным коэффициентом ускорения:

$$K_{q,r}^D = q - \nu q + \frac{\nu q}{\varphi + (1 - \varphi) \frac{\nu q}{r\rho}}.$$

Ускорение, достигаемое  $q$  ядрами и  $r$  ускорителями, исчисляется по сравнению с одним ядром. Очевидно, вычислительные подпроцессы, выполняемые однородной MIMD структурой и гибридной MIMD/SIMD структурой, должны оканчиваться одновременно.

В режиме деления это означает, что на каждое ядро MIMD компоненты загружается  $\frac{1}{K_{q,r}^D}$  часть (доля) вычислительного процесса (если какое-либо ядро будет выполнять вычисления медленнее, то ускорение будет меньше  $K_{q,r}^D$ ).

Всего однородная MIMD компонента выполняет приходящуюся на ее долю в  $(q - q^*) \frac{1}{K_{q,r}^D}$  вычислений в параллельном режиме.

Гибридная MIMD/SIMD компонента содержит  $q^*$  ядер и  $r$  ускорителей; на нее

приходится оставшаяся доля  $\left( \frac{q^*}{\varphi + (1 - \varphi) \frac{\nu q}{r \rho}} \right) \frac{1}{K_{q,r}^D}$  вычислений.

Эти вычисления равными долями распределены по  $r$  гибридным вычислителям, в каждом из которых 1 ускоритель и  $d = \frac{q^*}{r}$  соединенных с ним ядер. На один такой

вычислитель приходится  $\frac{1}{r} \left( \frac{q}{\varphi + (1 - \varphi) \frac{\nu q}{r \rho}} \right) \frac{1}{K_{q,r}^D}$  вычислений.

Этот коэффициент ускорения достигается по сравнению с одним ядром в режиме деления.

Напомним, что в этом режиме

$$K_{d,1} = \frac{T_1}{T_{d,1}},$$

где

$$T_{d,1} = T_1 \frac{\varphi}{d} + T_1 (1 - \varphi) \frac{1}{\rho}.$$

Первое слагаемое – длительность вычислений, исполняемых параллельно ядрами, образующими MIMD компоненту, второе – длительность вычислений, выполняемых ускорителем.

Потребуем, чтобы гибридная компонента выполняла свою часть вычислений одновременно с другими компонентами, т.е. в течение  $\frac{1}{K_{q,r}^D} T_1$ . Тогда  $\frac{1}{K_{q,r}^D} K_{d,1} T_{d,1} = \frac{1}{K_{q,r}^D} T_1$ .

Это означает, что на одно ядро приходится  $\frac{T_1}{K_{q,q}^D} \frac{K_{d,1}}{d} \varphi$  вычислений, а на ускоритель

$\frac{T_1}{K_{q,d}^D} (1 - \varphi) \frac{K_{d,1}}{\rho}$  вычислений.

Аналогично, в режиме умножения имеем  $K_{d,1} = \frac{dT_1}{T_{d,1}}$ , где  $T_{d,1} = T_1 \varphi + T_1 (1 - \varphi) \frac{d}{\rho}$ .

Потребуем, как и ранее, чтобы гибридная компонента выполняла вычисления одновременно с другими компонентами в данном режиме в течение  $T_1$ .

Тогда

$$\frac{K_{d,1} T_{d,1}}{d} = T_1.$$

На одно ядро приходится  $\frac{K_{d,1}}{d} T_1 \varphi$  вычислений, на ускоритель приходится

$$\frac{K_{d,1} T_1 (1 - \varphi) d}{d \rho} = \frac{K_{d,1} T_1 (1 - \varphi)}{\rho}.$$

Изложенные соотношения позволяют распределять вычисления в соответствии со структурой вычислителя. При этом достигается одинаковая длительность процессов,

выполняемых различными компонентами процессорного элемента, соответствующая коэффициенту ускорения  $K_{q,r}^D$ .

Таким образом, в результате применения изложенного алгоритма выполняется распределение вычислений, позволяющее получить коэффициент ускорения равный  $K_{q,r}^D$ .

#### 4 ПРИМЕРЫ ДЕКОМПОЗИЦИИ И БАЛАНСИРОВКИ ВЫЧИСЛЕНИЙ

Пусть гибридный процессорный элемент содержит два универсальных процессора, в каждом процессоре 14 ядер, и три ускорителя. Таким образом,  $q = 28$  и  $r = 3$ .

Оценим значения коэффициентов ускорения, достигаемые различными структурами процессорного элемента. Вычислительный процесс имеет следующие параметры [5]:

- длительность выполнения всего процесса одним ядром универсального процессора  $T_1 = 22,98$  с;
- длительность выполнения MIMD фрагмента одним ядром  $T_M = 7,05$  с;
- длительность выполнения SIMD фрагмента одним ускорителем  $T_S = 2,82$  с.

$$\text{Находим, } \varphi = \frac{T_M}{T_1} = 0,3, \quad \rho = \frac{(1 - \varphi)T_1}{T_S} = 5,7.$$

Полагаем, что длительность обменов между ядрами, а также между ядрами и ускорителем достаточно мала, и ею можно пренебречь.

Это, в частности, означает, что задействование  $q = 28$  ядер для выполнения данного процесса позволяет без применения ускорителей получить  $K_{28,0} = 28$ .

Поскольку  $\rho > \frac{1 - \varphi}{\varphi} = 2,33$  целесообразно наращивать количество ядер.

Рассмотрим «пропорциональный» вариант; в этом варианте каждому ускорителю сопоставлено  $\frac{q}{r}$  ядер.

Представим процессорный элемент совокупностью трех структур, в каждой по 9 ядер и одному ускорителю (одно ядро из 28 задействуем автономно).

$$\text{Получаем } K_{28,3} = 1 + 3K_{9,1} = 1 + 3 \cdot 6,05 = 19,1.$$

Т.е. применение ускорителей в таком сочетании замедляет вычисления.

Оценим возможности ускорения вычислений.

Предельное значение ускорения, которое может быть достигнуто в рассматриваемом примере, определяется выражением:

$$K_{28,3}^{\max} = q + r\rho \frac{1 - \sqrt{\varphi}}{1 + \sqrt{\varphi}} = 28 + 3 \cdot 5,7 \frac{1 - 0,548}{1 + 0,548} = 32,99.$$

Построим структуру на основе полученных выше оптимизационных соотношений.

Во-первых, находим:

$$v = \frac{r\rho\sqrt{\varphi}}{q(1 + \sqrt{\varphi})} = \frac{3 \cdot 5,7 \cdot 0,548}{28 \cdot 1,548} = 0,216.$$

Следовательно,  $q^* = 6$  и  $d = \frac{q^*}{r} = 2$ . На каждый ускоритель целесообразно задействовать два ядра. Получаем:

$$K_{28,3}^D = 28 - 6 + \frac{6}{0,3 + (1 - 0,3) \frac{6}{3 \cdot 5,7}} = 32,98.$$

Таким образом, ускорение, полученное в результате применения предложенного метода, совпало с максимально достижимым в заданных условиях. Оно в 1,83 раза больше получаемого в «пропорциональном» варианте, когда ядра равными частями распределяются между ускорителями.

Далее выполним балансировку для режима «деления» вычислений. Вычислительный процесс делится на  $K_{28,3}^D = 33$  долей (порций).

Универсальная однородная компонента содержит  $q - \nu q = 22$  ядра. Эти ядра выполняют 22 порции вычислений из  $\approx 33$ . На каждое ядро в режиме деления приходится  $\frac{1}{33} T_1$  вычислений.

$$\text{Гибридная компонента выполняет } K_{q,r} = \frac{\nu q}{\varphi + (1 - \varphi) \frac{\nu q}{r \rho}} = \frac{6}{0,3 + 0,7 \frac{6}{5,7 \cdot 3}} \approx 11,01$$

порций.

$$\text{В режиме деления на каждое ядро приходится } \frac{T_1}{K_{q,q}^D} \frac{K_{d,1}}{d} \varphi = T_1 \frac{1}{33} \frac{3,67}{2} \cdot 0,3 = \frac{1}{33} 0,55 T_1$$

вычислений (в однородной компоненте на ядро приходится  $\frac{T_1}{33}$  вычислений). На один

ускоритель приходится  $\frac{T_1}{K_{q,q}^D} \frac{K_{d,1}}{\rho} (1 - \varphi) = \frac{T_1}{33} 0,45$  вычислений.

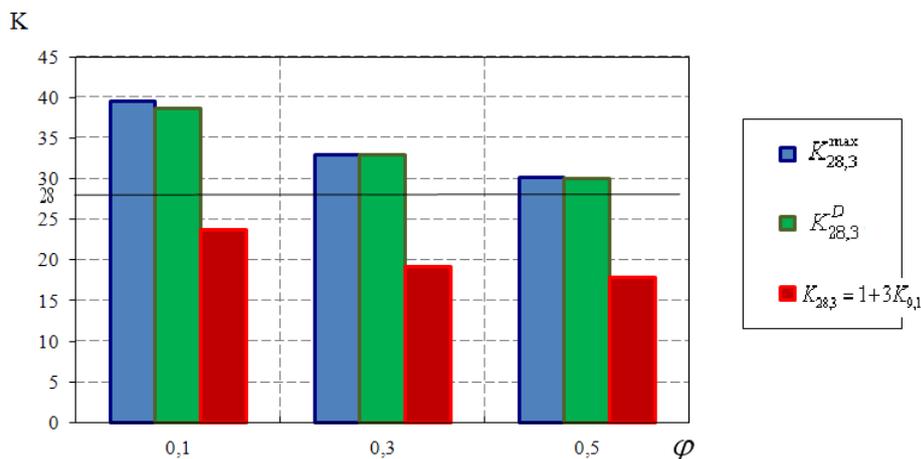
В режиме умножения на каждое ядро в неоднородной компоненте приходится по  $\frac{K_{d,1}}{d} \varphi T_1 = 0,55 T_1$ , на ускоритель -  $\frac{K_{d,1}(1 - \varphi)}{\rho} T_1 = 0,45 T_1$ .

Аналогичные действия выполняются и при других значениях  $q$ ,  $r$ ,  $\varphi$  и  $\rho$ .

Значения ускорения  $K_{28,3}^{\max}$ ,  $K_{28,3}^D$  и «пропорциональный» вариант  $K_{28,3} = 1 + 3 K_{9,1}$ , иллюстрирующие полученные соотношения, приведены для различных  $\varphi$  в таблице 3 и на рисунке 6. Они демонстрируют эффективность предложенного метода.

**Таблица 3** – Значения коэффициентов ускорения для  $q = 28$  и  $r = r$ .

| $\varphi$ | $\rho$ | $\nu$ | $K_{28,3}^{\max}$ | $K_{28,3}^D$ | $K_{28,3} = 1 + 3 K_{9,1}$ |
|-----------|--------|-------|-------------------|--------------|----------------------------|
| 0,1       | 7,4    | 0,216 | 39,5              | 38,6         | 23,7                       |
| 0,3       | 5,7    | 0,19  | 33                | 33           | 19,1                       |
| 0,5       | 4,07   | 0,18  | 30,15             | 30,05        | 17,8                       |



**Рисунок 6** – Значения ускорения

## ЗАКЛЮЧЕНИЕ

Предложен метод декомпозиции структуры гибридного процессорного элемента в соответствии с особенностями конструктивной реализации этого элемента и определенными первичными параметрами исполняемого вычислительного процесса. Метод позволяет получить минимальную в заданных условиях длительность вычислений.

Получены значения коэффициентов ускорения вычислений, достигаемых применением гибридных процессорных элементов, в частности, сформулированы аналитические условия достижения максимальных значений ускорений для заданной структуры вычислителя и данного вычислительного процесса.

Предложен основанный на полученных соотношениях метод балансировки (распределения) вычислительного процесса, позволяющий достигать максимального для заданных условий ускорения вычислений.

Приведены иллюстрации применения полученных соотношений, демонстрирующие возможности предложенного метода.

Применение метода требует средств программной или аппаратной коммутации ядер универсального процессора и ускорителей.

Предложенные средства позволяют повысить производительность создаваемых и перспективных вычислительных систем, использующих гибридную архитектуру.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

[1] Степаненко С.А., Южаков В.В. Эксафлопные суперЭВМ. Контуры архитектуры // Программные системы: теория и приложения: электрон. научн. журн. 2013. Т. 4, № 3(17). Режим доступа: <http://psta.psisras.ru/read/>

[2] Степаненко С.А. Оценки ускорения вычислений гибридными системами. Пленарные доклады Пятой международной конференции <Параллельные вычисления и задачи управления> РАСО 2010 Москва 26-28 октября 2010г. М.: Учреждение Российской академии наук. Институт проблем управления им.В.А.Трапезникова РАН стр.61-71, ISBN 978-5-91450-062-4.

[3] Степаненко С.А. Оценки ускорения вычислений гибридными реконфигурируемыми системами. Известия Южного Федерального Университета. Технические науки № 12, декабрь 2014г., с.74-83 ISSN 1999-9429.

[4] Mohan R., Gopalan N.P., Dynamic Load Balancing using Graphics Processors. L.J. Intelligent Systems and Applications, 2014, 05, 70-75.

[5] Воронин Б.Л., Ерофеев А.М., Копкин С.В., Крючков И.А., Рыбкин А.С., Степаненко С.А., Южаков В.В. Применение арифметических ускорителей для расчета задач молекулярной динамики по программному комплексу МД: «Вопросы атомной науки и техники» Сер. Математическое моделирование физических процессов. 2009г., вып.2.

## ПРИЛОЖЕНИЕ А

$$K = q - \nu q + \frac{\nu q}{\varphi + (1 - \varphi) \frac{\nu q}{r \rho}};$$

$$\begin{aligned} \frac{dK}{d\nu} &= -q + \frac{\left( \varphi + (1 - \varphi) \frac{\nu q}{r \rho} \right) q - \nu q (1 - \varphi) \frac{q}{r \rho}}{\left( \varphi + (1 - \varphi) \frac{\nu q}{r \rho} \right)^2} = \\ &= \frac{-q \left( \varphi + (1 - \varphi) \frac{\nu q}{r \rho} \right)^2 + q \left( \varphi + (1 - \varphi) \frac{\nu q}{r \rho} \right) - \nu q (1 - \varphi) \frac{q}{r \rho}}{\left( \varphi + (1 - \varphi) \frac{\nu q}{r \rho} \right)^2} \end{aligned}$$

Приравняем числитель к 0.

$$-q \left( \varphi + (1 - \varphi) \frac{\nu q}{r \rho} \right)^2 + q \left( \varphi + (1 - \varphi) \frac{\nu q}{r \rho} \right) - \nu q (1 - \varphi) \frac{q}{r \rho} = 0.$$

Раскрывая скобки во втором слагаемом, получаем:

$$\begin{aligned} -q \left( \varphi + (1 - \varphi) \frac{\nu q}{r \rho} \right)^2 + q \varphi &= 0 \\ \varphi + (1 - \varphi) \frac{\nu q}{r \rho} &= \sqrt{\varphi} \\ (1 - \varphi) \frac{\nu q}{r \rho} &= \sqrt{\varphi} - \varphi \\ \nu &= \frac{r \rho (\sqrt{\varphi} - \varphi)}{q (1 - \varphi)} \\ \nu &= \frac{r \rho (\sqrt{\varphi} - \varphi)}{q (1 - \varphi)} = \\ &= \frac{r \rho \frac{\sqrt{\varphi} (1 - \sqrt{\varphi})}{(1 - \sqrt{\varphi})^2 - 2\varphi + 2\sqrt{\varphi}}}{q} = \\ &= \frac{r \rho \frac{\sqrt{\varphi} (1 - \sqrt{\varphi})}{(1 - \sqrt{\varphi})^2 + 2\sqrt{\varphi} (1 - \sqrt{\varphi})}}{q} = \\ &= \frac{r \rho \frac{\sqrt{\varphi}}{1 - \sqrt{\varphi} + 2\sqrt{\varphi}}}{q} = \frac{r \rho \frac{\sqrt{\varphi}}{1 + \sqrt{\varphi}}}{q}. \end{aligned}$$

Окончательно получаем:

$$\nu = \frac{r \rho \frac{\sqrt{\varphi}}{1 + \sqrt{\varphi}}}{q}.$$

## ПРИЛОЖЕНИЕ Б

Подставляя полученное выражение  $v$  в  $K = q - vq + \frac{vq}{\varphi + (1-\varphi)\frac{vq}{r\rho}}$ , находим:

$$\begin{aligned}
 K &= q - \frac{r\rho(\sqrt{\varphi} - \varphi)}{(1-\varphi)} + \frac{\frac{r\rho(\sqrt{\varphi} - \varphi)}{(1-\varphi)}}{\varphi + (1-\varphi)\frac{r\rho(\sqrt{\varphi} - \varphi)}{(1-\varphi)r\rho}} = \\
 &= q - \frac{r\rho(\sqrt{\varphi} - \varphi)}{(1-\varphi)} + \frac{r\rho(\sqrt{\varphi} - \varphi)}{(1-\varphi)\sqrt{\varphi}} = \\
 &= q + \frac{r\rho(\sqrt{\varphi} - \varphi)}{(1-\varphi)} \left( \frac{1}{\sqrt{\varphi}} - 1 \right) = \\
 &= q + r\rho \frac{(\sqrt{\varphi} - \varphi)(1 - \sqrt{\varphi})}{(1-\varphi)\sqrt{\varphi}} = \\
 &= q + r\rho \frac{\sqrt{\varphi}(1 - \sqrt{\varphi})^2}{(1-\varphi)\sqrt{\varphi}} = \\
 &= q + r\rho \frac{(1 - \sqrt{\varphi})^2}{(1 - \sqrt{\varphi})^2 + 2\sqrt{\varphi} - 2\varphi} = \\
 &= q + r\rho \frac{(1 - \sqrt{\varphi})^2}{(1 - \sqrt{\varphi})^2 + 2\sqrt{\varphi}(1 - \sqrt{\varphi})} = \\
 &= q + r\rho \frac{1 - \sqrt{\varphi}}{1 - \sqrt{\varphi} + 2\sqrt{\varphi}} = \\
 &= q + r\rho \frac{1 - \sqrt{\varphi}}{1 + \sqrt{\varphi}}.
 \end{aligned}$$