

УДК 681.3

Оценка быстродействия устройства с систолической структурой для умножения бинарных матриц

Гвоздева С.Н., аспирант

Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Юго-Западный государственный университет», г. Курск

e-mail: svetka-gvozdeva@yandex.ru

Ватутин Э.И., канд. техн. наук

Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Юго-Западный государственный университет», г. Курск

e-mail: evatutin@rambler.ru

Титов В.С., докт. техн. наук, проф.

Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Юго-Западный государственный университет», г. Курск

e-mail: titov-swsu@rambler.ru

В данной статье представлена оценка быстродействия для разработанного устройства для умножения бинарных матриц размером $n \times n$ с систолической структурой и выполнено ее сравнение с быстродействием прототипа, приведены временные диаграммы для устройства умножения бинарных матриц и для прототипа, на базе которых сделаны выводы по быстродействию разработанного устройства по отношению к прототипу данного устройства. В лучшем случае ($n < 2^{|\log_2 m|+2}$, где m – разрядность обрабатываемых данных) разработанное устройство обеспечивает выигрыш во времени до 57,7 раза;

в худшем случае быстроедействие предложенного устройства и прототипа совпадает и по прежнему лимитируются доступом в специализированную многопортовую память.

Ключевые слова: *умножение матриц, систолические структуры, многопортовая память, специализированные вычислительные средства.*

В настоящее время для некоторых областей научно-технического направления большое значение имеет задача умножения матриц. Время ее решения во многих случаях является важным компонентом, поэтому существует большое количество различных подходов, связанных с оптимизацией и распараллеливанием выполняемых действий [1]. В ряде научных работ описаны алгоритмические оптимизации, оказывающие значительное влияние на время выполнения умножения и направленные на повышение эффективности работы кэш-памяти CPU или разделяемой памяти GPU[2–3]. Кроме того, известен ряд аппаратно-ориентированных решений рассматриваемой задачи [4]

На данный момент известны эффективные способы реализации векторных и матричных операций на аппаратном уровне, способные существенно снизить время обработки. Они подразделяются на 3 основные группы.

К первой из них можно отнести схемы на оптических элементах [7–11], которые по ряду причин не получили широкого распространения и в настоящее время практически не применяются.

Вторая группа включает в себя устройства умножения матриц, имеющие вероятностные свойства и присущую им статистическую погрешность [12–14]. В настоящее время они не используются на практике, причиной этому является несоответствие общепринятому стандарту IEEE 754 на операции с плавающей точкой и практически полное вытеснение с рынка устройствами третьего направления.

Третья и наиболее обширная группа образована устройствами, в основу работы которых положен принцип параллельной [15–19] иногда в сочетании

с конвейерной, матричной и/или систолической обработкой данных [15–21]. Известными схемами умножения являются также умножители Брауна, Бута, Бо-Вули и пр. [4], что в различных вариациях используется в приведенных выше патентах. Данная группа аппаратных реализаций характеризуется существенным выигрышем во времени выполнения операции, в некоторых случаях позволяя выполнение умножения матриц за линейное время, однако данным устройствам необходима специализированная многопортовая память [6], которая должна обеспечить достаточный темп поступления исходных данных, иначе быстродействие устройства будет лимитировано именно им, а не скоростью работы операционной части.

В данной статье приведен расчет быстродействия устройства для умножения бинарных матриц размером $n \times n$. Технической задачей предложенного устройства является снижение аппаратной сложности и увеличения выигрыша во времени за счет реализации возможности умножения бинарных квадратных матриц $C = A \times B$ размером $n \times n$ элементов.

Учитывая специфику выполняемой операции, общая стратегия умножения матриц сводится к выполнению следующих шагов:

- 1) загрузка исходных данных (коэффициентов) матриц A и B из оперативной памяти в память устройства;
- 2) выполнение умножения;
- 3) передача результирующей матрицы C из памяти устройства в оперативную память.

В общем случае кроме непосредственно времени умножения необходимо учитывать время загрузки исходных данных и время возврата результата, для чего определяющим фактором является пропускная способность шины, обеспечивающей связь специализированного устройства и оперативной памяти (контроллера прямого доступа в память).

Время работы устройства будет зависеть от времени работы каждого его этапа отдельно. Его можно определить как

$$t_{общ} = t_{зан} + t_{ин} + t_{раб} + t_{выгр}, \quad (1)$$

где $t_{зан}$ – время передачи данных из оперативной памяти в память устройства; $t_{ин}$ – время инициализации триггеров в составе ячеек операционной части устройства; $t_{раб}$ – время обработки данных; $t_{выгр}$ – время передачи данных из памяти устройства в оперативную память.

На рисунке 1 представлена схема блока хранения одной ячейки памяти. При реализации операции записи сигналу с входа данных 23 необходимо переключить D -триггер 16 (необходимое время – $2t_0$, где t_0 – время работы одного эквивалентного вентиля) по приходу синхросигнала, которому требуется пройти через трехвходовой элемент И 17 (необходимое время – $2t_0$). Таким образом, общее время записи одной ячейки составляет

$$t_{зан} = \underbrace{2t_0}_{D16} + \underbrace{2t_0}_{И17} = 4t_0. \quad (1)$$

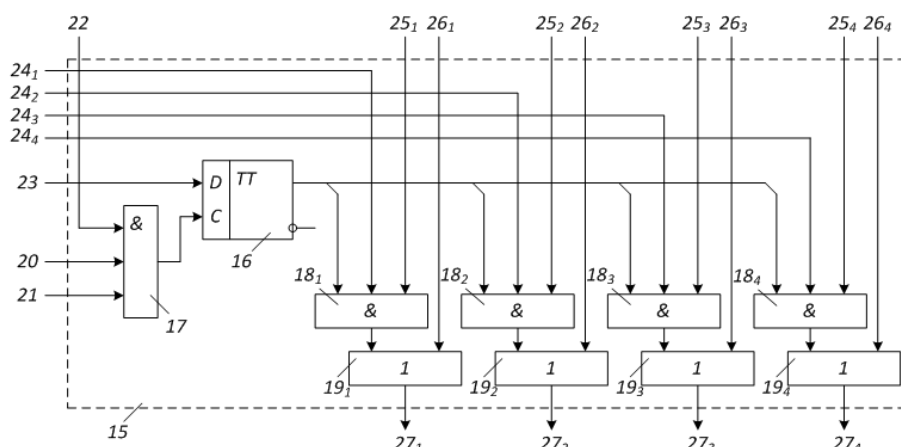


Рис.1 – Схема блока хранения предложенного устройства

На этапе загрузки исходных данных значения a_{ij} , $i, j = \overline{1, n}$ элементов первой матрицы подаются поочередно. С учетом того, что время записи одной ячейки составляет $4t_0$ (1), общее время последовательной загрузки данных одной матрицы, включающей в своем составе $n \times n$ ячеек, составляет $4t_0 n^2$. Аналогичное время требуется для загрузки данных второй матрицы, а общее время загрузки составляет $t_{зан} = 8t_0 n^2$.

На этапе инициализации устройства производится сброс значений триггеров 6, 7, 8, на что требуется время $t_{ин} = 2t_0$.

Этап работы устройства включает в своем составе $2n - 1$ итерацию. На каждой итерации операционная часть работает как группа линейных синхронных конвейеров под управлением общего синхросигнала. Конвейерный такт определяется максимумом из времени чтения данных из специализированной памяти и времени работы ячейки операционного блока.

При чтении данных сигнал с выхода D -триггера 16 проходит через один из трехходовых элементов И 18, на что затрачивается время $2t_0$, через один из двухходовых элементов ИЛИ 19 и через n -входовой элемент ИЛИ 34 (рис. 2), на что с использованием пирамидальной схемы из двухходовых элементов необходимо время $\lceil \log_2 n \rceil t_0$, где $\lceil x \rceil$ – операция округления вверх. Таким образом, общее время чтения данных составляет

$$t_{ум} = (2t_0 + t_0)n + \lceil \log_2 n \rceil t_0 = 3t_0n + \lceil \log_2 n \rceil t_0.$$

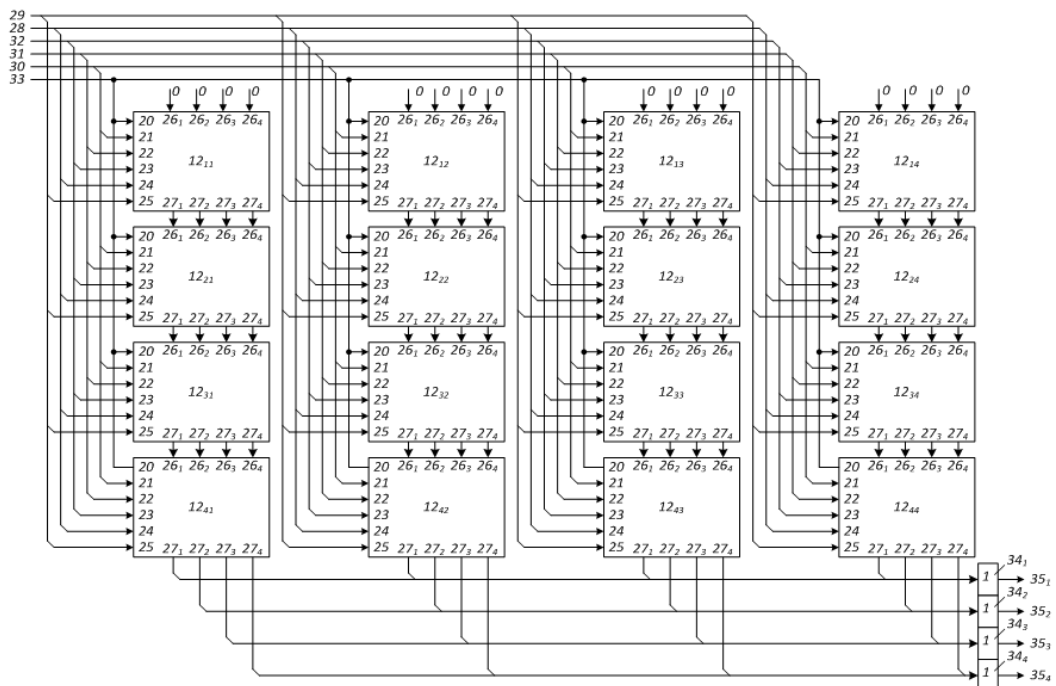


Рис.2 – Схема блока коэффициентов матрицы

На рисунке 3 представлена схема операционного блока предложенного устройства.

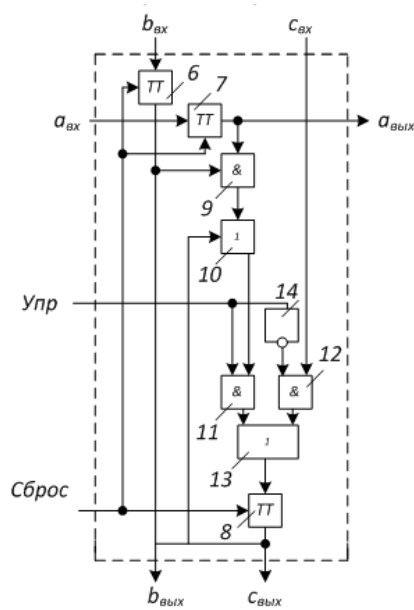


Рис.3 – Схема операционного блока предложенного устройства

Во время работы ячейки операционного блока предложенного устройства сигнал $a_{вх}$ проходит через двухступенчатый триггер 7 за время $4t_0$, параллельно с этим, сигнал $b_{вх}$ аналогично проходит через двухступенчатый триггер 6. Изначально запись сигналов $a_{вх}$ и $b_{вх}$ происходит в первую ступень триггеров 6 и 7, соответственно, за время $2t_0$. Далее, сигналы переходят на вторую ступень данных триггеров, за время равное $2t_0$. Наиболее длинный путь, возникающий при прохождении сигнала через логические элементы по схеме операционного блока, представлен

$$\text{элементами } \max \left(\underbrace{TT\ 6}_{2t_0}, \underbrace{TT\ 7}_{2t_0} \right) \rightarrow \underbrace{И\ 9}_{t_0} \rightarrow \underbrace{ИЛИ\ 10}_{t_0} \rightarrow \underbrace{И\ 11}_{t_0} \rightarrow \underbrace{ИЛИ\ 13}_{t_0} \rightarrow \underbrace{TT\ 8}_{2t_0},$$

на что необходимо время $8t_0$. Таким образом, общее время работы ячейки операционного блока составляет $10t_0$. Поясняющая временная диаграмма приведена на рис.4.

Конвейерный такт может быть определен как:

$$t_k = \max(10t_0, 3t_0n + \lceil \log_2 n \rceil t_0).$$

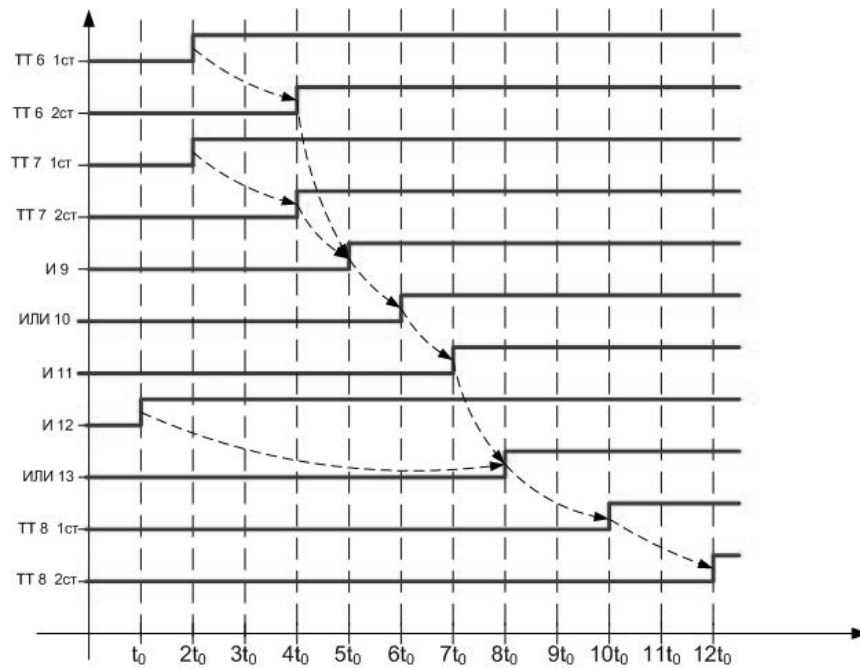


Рис.4 – Временная диаграмма, поясняющая работу предложенного устройства умножения бинарных матриц

Для практически важных случаев ($n > 4$)

$$t_k = \max(10t_0, 3t_0n + \lceil \log_2 n \rceil t_0) = 3t_0n + \lceil \log_2 n \rceil t_0,$$

то есть время работы схемы лимитируется темпом поступления входных данных из специализированной многопортовой памяти. Время обработки данных $t_{раб} = t_k \times (2n - 1)$ ввиду того, что при выбранном алгоритме умножения матриц подача коэффициентов производится за $2n - 1$ итерацию [2].

На этапе получения результата происходит выгрузка данных, хранящихся в двухступенчатых триггерах 8 ячеек среды. Выгрузка происходит в конвейерном режиме: перемещение данных между ступенями триггеров 8 занимает время, равное $4t_0$, для перемещения данных ячеек верхнего ряда требуется n итераций, следовательно $t_{выгр} = 4t_0n$.

$$\begin{aligned} t_{общ} &= 8t_0n^2 + 2t_0 + (3t_0n + \lceil \log_2 n \rceil t_0) \times (2n - 1) + 4t_0n = \\ &= 8t_0n^2 + 2t_0 + (6t_0n^2 + 2\lceil \log_2 n \rceil t_0 - 3t_0n - \lceil \log_2 n \rceil t_0) + 4t_0n = \\ &= 14t_0n^2 + 2t_0 + \lceil \log_2 n \rceil t_0 + t_0n = (14n^2 + 2 + \lceil \log_2 n \rceil + n)t_0 \end{aligned}$$

Ниже приведено описание устройства прототипа [2], производящего умножение матриц общего вида. Для определения оценки выигрыша устройства необходимо сравнить его с прототипом, для этого ниже приведен расчет времени его работы.

Так же, как и для предложенного устройства, время работы прототипа будет зависеть от времени работы каждого этапа отдельно. Его можно определить по формуле (1) аналогично рассмотренному выше.

На рисунке 5 представлена схема блока хранения одной ячейки памяти прототипа. При реализации операции записи сигналу с входа данных 20 необходимо переключить регистр 13 (необходимое время – $2t_0$) по приходу синхросигнала, которому требуется пройти через трехходовой элемент И 14 (необходимое время – $2t_0$). Таким образом, общее время записи одной ячейки составляет

$$t_{зан}^{(n)} = \underbrace{2t_0}_{PI\ 13} + \underbrace{2t_0}_{II\ 14} = 4t_0. \quad (3)$$

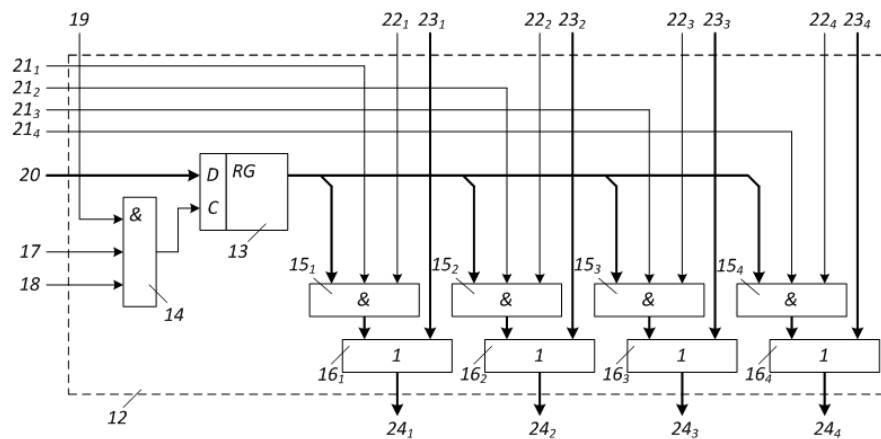


Рис.5 – Схема блока хранения прототипа

Загрузка исходных данных и инициализация для прототипа производятся аналогично рассмотренным выше операциям для предложенного устройства и требуют $t_{зан}^{(n)} = 8t_0n^2$ и $t_{ин}^{(n)} = 2t_0$. соответственно.

Этап работы прототипа концептуально не отличается от рассмотренного выше и выполняется за $2n-1$ итерацию. При чтении данных сигнал с выхода регистра 13 проходит через один из трехходовых элементов

И 15, на что затрачивается время $2t_0$, через один из двухвходовых элементов ИЛИ 16 и через n -входовой элемент ИЛИ 31 (рис. 6), на что с использованием пирамидальной схемы из двухвходовых элементов необходимо время $\lceil \log_2 n \rceil t_0$. Таким образом, общее время чтения данных, аналогично рассмотренному выше, составляет

$$t_{ум}^{(n)} = (2t_0 + t_0)n + \lceil \log_2 n \rceil t_0 = 3t_0n + \lceil \log_2 n \rceil t_0.$$

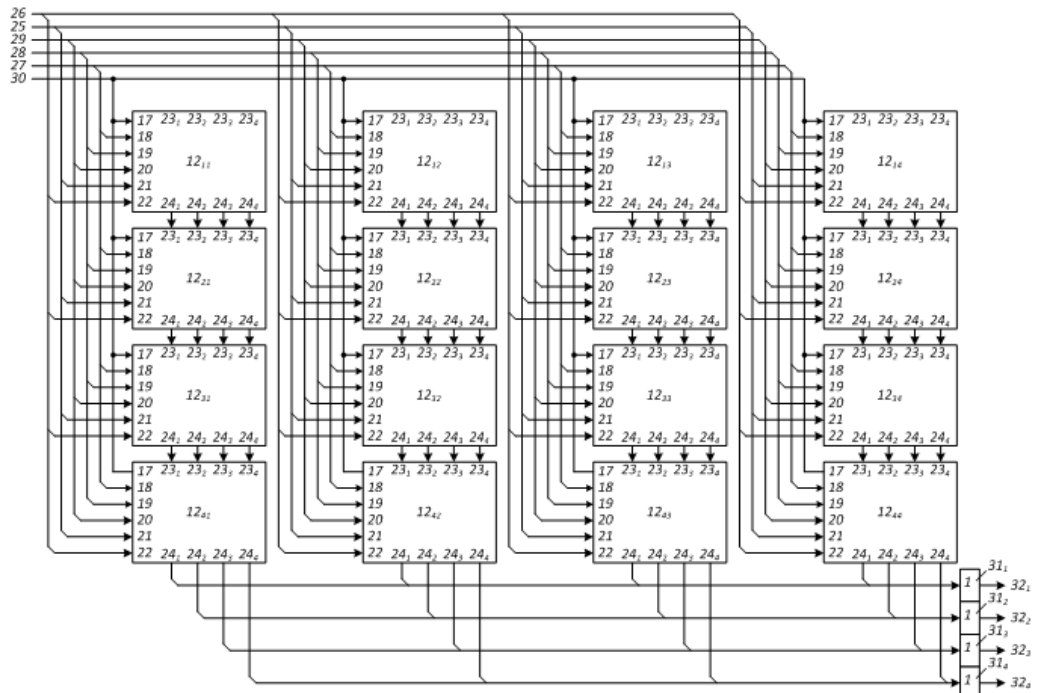


Рис.6 – Схема блока коэффициентов матрицы прототипа

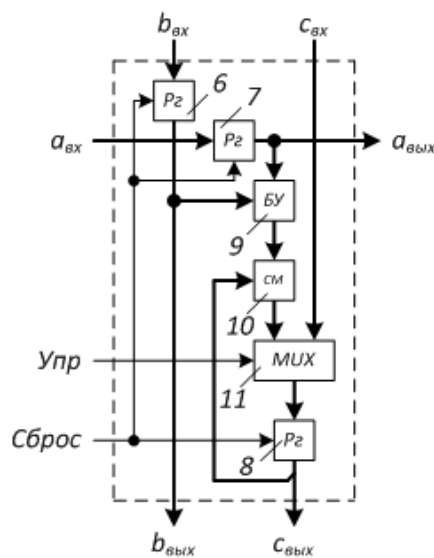


Рис.7 – Схема операционного блока прототипа

Во время работы ячейки операционного блока сигнал a_{ex} проходит через регистр 7 за время $2t_0$, параллельно с этим, сигнал b_{ex} аналогично проходит через регистр 6. Наиболее длинный путь, возникающий при прохождении сигнала через логические элементы по схеме операционного блока, представлен элементами

$$\max \left(\underbrace{PG\ 6, PG\ 7}_{2t_0} \right) \rightarrow \underbrace{БУ\ 9}_{3t_0(2m-2)} \rightarrow \underbrace{CM\ 10}_{3t_0m} \rightarrow \underbrace{MUX\ 11}_{3t_0} \rightarrow \underbrace{PG\ 8}_{2t_0}.$$

В данной схеме на рисунке 7 применяется простейший вид сумматора с последовательным переносом, в качестве блока умножения применяется каскад из матрично-суммирующих блоков [22], выполненных в виде комбинационной схемы. Общее время работы ячейки операционного блока составляет $(t_0 + 9t_0m)$. Поясняющая временная диаграмма приведена на рис. 8.

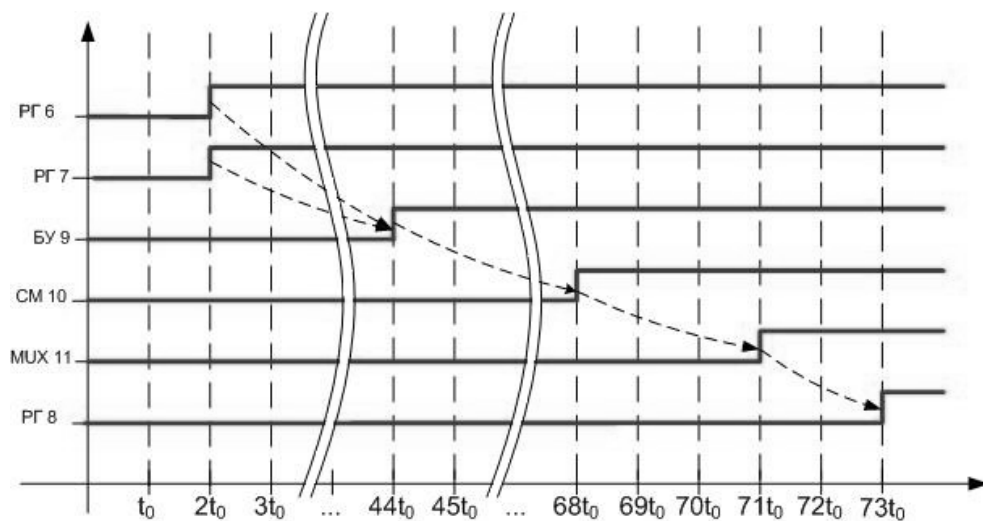


Рис.8 – Временная диаграмма, поясняющая принцип работы операционной части прототипа

Конвейерный такт в данном случае может быть определен как:

$$t_k^{(n)} = \max(t_0 + 9t_0m, 3t_0n + \lceil \log_2 n \rceil t_0).$$

Для случаев, когда быстродействие лимитируется операционной частью,

$$t_k^{(n)} = \max(t_0 + 9t_0m, 3t_0n + \lceil \log_2 n \rceil t_0) = t_0 + 9t_0m.$$

Для случаев, когда быстродействие лимитируется темпом поступления данных из памяти

$$t_k^{(n)} = \max(t_0 + 9t_0m, 3t_0n + \lceil \log_2 n \rceil t_0) = 3t_0n + \lceil \log_2 n \rceil t_0.$$

На этапе получения результата выгрузка данных происходит аналогично предложенному устройству за время $t_{\text{выгр}}^{(n)} = 4t_0n$.

Для случая ($m = 8, n \geq 32; m = 16, n \geq 64; m = 32, n \geq 128...$) время работы схемы будет лимитироваться временем поступления данных из памяти, так же как и в предложенном устройстве. Поэтому, время работы двух устройств при данном значении будет одинаковым.

$$\begin{aligned} t_{\text{общ}}^{(n)} &= 8t_0n^2 + 2t_0 + (3t_0n + \lceil \log_2 n \rceil t_0) \times (2n - 1) + 4t_0n = \\ &= 8t_0n^2 + 2t_0 + (6t_0n^2 + 2\lceil \log_2 n \rceil t_0 - 3t_0n - \lceil \log_2 n \rceil t_0) + 4t_0n = \\ &= 14t_0n^2 + 2t_0 + \lceil \log_2 n \rceil t_0 + t_0n = (14n^2 + 2 + \lceil \log_2 n \rceil + n)t_0. \end{aligned}$$

Рассмотрим более подробно случай, когда время работы лимитируется операционной частью ($m = 8, n < 32; m = 16, n < 64; m = 32, n < 128...$). Тогда

$$\begin{aligned} t_{\text{общ}}^{(n)} &= 8t_0n^2 + 2t_0 + (t_0 + 9t_0m) \times (2n - 1) + 4t_0n = \\ &= 8t_0n^2 + 2t_0 + 2t_0n + 18t_0mn - t_0 - 9t_0m + 4t_0n = \\ &= 8t_0n^2 + t_0 + 6t_0n + 9t_0m(2n - 1). \end{aligned}$$

Проведем сравнение быстродействия двух устройств при одинаковых n

$$\eta = \frac{t_{\text{общ}}^{(n)}}{t_{\text{общ}}}$$

В таблицах 1–4 приведены результаты расчеты быстродействия двух устройств, и оценка выигрыша устройств при различных значениях m .

Таблица 1

Оценка быстродействия устройств при $m=8, t_0 = 1нс$

n	Оценка быстродействия прототипа устройства, мс	Оценка быстродействия нового устройства, мс	Выигрыш, раз
2	0,073	0,010	7,30

4	0,073	0,014	5,21
8	0,073	0,027	2,70
16	0,073	0,052	1,40
32	0,101	0,101	1,00
64	0,198	0,198	1,00
128	0,391	0,391	1,00
256	0,776	0,776	1,00
512	1,545	1,545	1,00
1024	3,082	3,082	1,00
2048	6,155	6,155	1,00

Таблица 2

Оценка быстродействия устройств при $m=16$, $t_0 = 1нс$

n	Оценка быстродействия прототипа устройства, мс	Оценка быстродействия нового устройства, мс	Выигрыш, раз
2	0,145	0,010	14,50
4	0,145	0,014	10,36
8	0,145	0,027	5,37
16	0,145	0,052	2,79
32	0,145	0,101	1,44
64	0,198	0,198	1,00
128	0,391	0,391	1,00
256	0,776	0,776	1,00
512	1,545	1,545	1,00
1024	3,082	3,082	1,00
2048	6,155	6,155	1,00

Таблица 3

Оценка быстродействия устройств при $m=32$, $t_0 = 1нс$

n	Оценка быстродействия прототипа устройства, мс	Оценка быстродействия нового устройства, мс	Выигрыш, раз
2	0,289	0,010	28,90
4	0,289	0,014	20,64
8	0,289	0,027	10,70
16	0,289	0,052	5,56
32	0,289	0,101	2,86
64	0,289	0,198	1,46
128	0,391	0,391	1,00
256	0,776	0,776	1,00
512	1,545	1,545	1,00
1024	3,082	3,082	1,00
2048	6,155	6,155	1,00

Таблица 4

Оценка быстродействия устройств при $m=64$, $t_0 = 1нс$

n	Оценка быстродействия прототипа устройства, мс	Оценка быстродействия нового устройства, мс	Выигрыш, раз
2	0,577	0,010	57,7
4	0,577	0,014	41,21
8	0,577	0,027	21,37
16	0,577	0,052	11,10
32	0,577	0,101	5,71
64	0,577	0,198	2,91
128	0,577	0,391	1,48
256	0,776	0,776	1,00
512	1,545	1,545	1,00
1024	3,082	3,082	1,00
2048	6,155	6,155	1,00

Из приведенных расчетов следует, что существует закономерность: при $m=8$ выигрыш в 1 раз получен при $n=32$, при $m=16$ выигрыш в 1 раз – при $n=64$, при $m=32$ – при $n=128$, при $m=64$ – при $n=256$. На основании этого можно заметить, что пороговое значение, определяющее минимальное значение выигрыша, определяется как

$$2^{\lceil \log_2 m \rceil + 2} \equiv 4m,$$

где символ \equiv обозначает точное равенство в случае, если разрядность обрабатываемых данных m является степенью двойки.

Таким образом, разработанное устройство обеспечивает выигрыш в быстродействии у прототипа в 57,7 – 1,0 раз при определенных соотношениях размерности обрабатываемых матриц n и разрядности обрабатываемых данных m :

$$n \geq 2^{\lceil \log_2 m \rceil + 2} - \eta = 1,$$

$$n < 2^{\lceil \log_2 m \rceil + 2} - \eta > 1.$$

Несмотря на то, что предложенное устройство характеризуется как меньшей аппаратной сложностью [23], так и более высоким быстродействием в ряде случаев, существуют такие соотношения размерности обрабатываемых матриц n и разрядности обрабатываемых

данных m , при которых быстродействие прототипа и предложенного устройства совпадает.

Список литературы

1. Ватутин Э.И., Мартынов И.А., Титов В.С. Оценка реальной производительности современных процессоров в задаче умножения матриц для однопоточной программной реализации // Известия Юго-Западного государственного университета. Серия «Управление, вычислительная техника, информатика. Медицинское приборостроение». 2013. № 4. С. 11–20.
2. Ватутин Э.И., Мартынов И.А., Титов В.С. Оценка реальной производительности современных видеокарт с поддержкой технологии CUDA в задаче умножения матриц // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. 2014. № 2. С. 8–17.
3. Затолокин Ю.А., Ватутин Э.И., Титов В.С. Алгоритмическая оптимизация программной реализации алгоритмов умножения плотных вещественных матриц на графических процессорах с поддержкой технологии OpenCL // Известия Юго-Западного государственного университета. 2017. Т. 21. № 5 (74). С. 6–15. DOI: 10.21869/2223-1560-2017-21-5-06-15.
4. Кун С. Матричные процессоры на СБИС: Пер. с англ. М.: Мир, 1991. 672 с.
5. Штейнберг, Б. Я. Блочнo-рекурсивное параллельное перемножение матриц / Б.Я. Штейнберг // Известия ВУЗов. Приборостроение. — Т. 52, №10. — 2009. — С. 33-41.
6. Патент РФ на полезную модель № 157948. Устройство для умножения матриц / Ватутин Э.И., Мартынов И.А., Титов В.С. Заявл. 08.07.2015, опубл. 20.12.2015. Бюл. № 35.
7. А. с. СССР № 588548, кл. G 06 G 9/00, 1976. Оптическое устройство для умножения матриц / Михляев С.В., Твердохлеб П.Е., Чугуй Ю.В. Заявл. 08.07.1976. Опубл. 15.01.1978.

8. А. с. СССР 1427394 МПК G06G9/00. Оптическое устройство для вычисления произведения трех матриц / Михляев С.В., Твердохлеб П.Е. Заявл. 20.03.1987. Оpubл. 30.09.1988.
9. Патент РФ 2022334 МПК G06F 1/04, G06F 15/347. Устройство для перемножения числовых матриц / Красиленко В.Г., Заболотная Н.И. Заявл. 21.10.1991. Оpubл. 30.10.1994.
10. А.с. СССР 1781679 МПК G06F1/04, G06F15/347. Устройство для умножения квадратных матриц картин-изображений / Красиленко В.Г., Заболотная Н.И. Заявл. 07.07.1989. Оpubл. 15.12.1992.
11. Патент РФ 2018916, МПК G06E 1/04, G06F 15/347. Устройство для умножения матриц картин-изображений / Красиленко В.Г., Заболотная Н.И., Евтихийев Н.Н. Заявл. 04.06.1991. Оpubл. 30.08.1994.
12. А.с. СССР 264797 МПК G06G7/16. Устройство для перемножения алгебраических матриц / Н.И. Денисенко, А.А. Лор. Оpubл. 01.01.1970.
13. А.с. СССР 647687 МПК G06F17/16. Устройство для операций над матрицами / Гладкий В.С., Гук Л.Б. Заявл. 21.12.1976. Оpubл. 15.02.1979.
14. А.с. СССР 1056192 МПК G06F7/70. Вероятностное устройство для умножения матриц / Яковлев В.В., Мальченкова О.С., Яковлев А.В. Заявл. 28.07.1982. Оpubл. 23.11.1983.
15. А.с. СССР 1345211 МПК G06F17/16. Устройство для операций над матрицами / Якуш В.П., Мищенко В.А., Соболевский П.И., Авгуль Л.Б., Лазаревич Э.Г. Заявл. 15.04.1986. Оpubл. 15.10.1987.
16. А.с. СССР 1418749 МПК G06F17/16. Устройство для умножения матриц / Обод И.И. Заявл. 23.01.1987. Оpubл. 23.08.1988.
17. А.с. СССР 1444819 МПК G06F17/16. Устройство поклеточного умножения матриц / Вышинский В.А., Рабинович З.Л., Тихонов Б.М., Фесенко Н.Б. Заявл. 22.12.1986. Оpubл. 15.12.1988.

18. А.с. СССР 1471201 МПК G06F17/16. Устройство для умножения матриц / Грищенко В.А., Калалб А.Д., Царев А.П. Заявл. 19.08.1987. Оpubл. 07.04.1989.

19. А.с. СССР 1807499 G06F15/347. Устройство для умножения матриц / Аванесян Г.Р., Антоненков В.Б., Айдаров Г.А. Заявл. 13.03.1991. Оpubл. 07.04.1993.

20. Патент РФ 2006937 G06F 15/347. Устройство для перемножения матриц / Выжиковски Р., Каневский Ю.С., Клименко М.К., Овраменко С.Г. Заявл. 19.10.1990. Оpubл. 30.01.1994.

21. А. с. СССР 1226484 МПК G06F17/16. Устройство умножения матрицы на вектор / Выжиковска А.В., Выжиковски Р., Каневский Ю.С., Лозинский В.И.. Заявл. 23.10.1984. Оpubл. 23.04.1986.

22. Угрюмов Е.П. Цифровая схемотехника. / Угрюмов Е.П. // Учебное пособие для вузов. 3-е издание, переработанное и дополненное. – Санкт-Петербург: Издательство «БХВ-Петербург», – 2010 – 800 с.

23. Гвоздева С.Н. Оценка аппаратной сложности устройства умножения квадратных бинарных матриц размером $n \times n$ / Гвоздева С.Н., Ватутин ЭИ. // Оптико-электронные приборы и устройства в системах распознавания образов и обработки изображений. Распознавание – 2019: сб. материалов XV Междунар. науч.-техн. конф. / ред. кол.: С.Г. Емельянов [и др.]; Юго-Зап. гос. ун-т. Курск, 2019. Стр. 66-68.