На правах рукописи

ГВОЗДЕВА СВЕТЛАНА НИКОЛАЕВНА

МОДЕЛЬ И АППАРАТНО-ОРИЕНТИРОВАННЫЙ АЛГОРИТМ ВЫЧИСЛИТЕЛЬНОГО УСТРОЙСТВА ДЛЯ ОБРАБОТКИ БИНАРНЫХ МАТРИЦ

Специальность 05.13.05 – Элементы и устройства
вычислительной техники и систем управления

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Курск 2021

Работа выполнена на кафедре «Вычислительная техника» Федерального государственного бюджетного образовательного учреждения высшего образования «Юго-Западный государственный университет»

Научный руководитель: Доктор технических наук, профессор,
Заслуженный деятель науки РФ
Юго-Западный государственный университет
Титов Виталий Семенович

Официальные оппоненты: **Ларкин Евгений Васильевич**
Доктор технических наук, профессор,
Тульский государственный университет, кафедра «Робототехника и автоматизация производства», заведующий кафедрой

Курочкин Илья Ильич
кандидат технических наук, с.н.с.,
Институт проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН), лаборатория Ц-1 «Моделирование и анализ телекоммуникационных систем», руководитель лаборатории

Ведущая организация: Федеральное государственное бюджетное образовательное учреждение высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»

Защита диссертации состоится «5» октября 2021 г. в 14:00 на заседании диссертационного совета Д 212.105.02, созданного на базе Юго-Западного государственного университета по адресу: 305040, г. Курск, ул. 50 лет Октября, 94, конференц-зал.

С диссертацией можно ознакомиться в библиотеке Юго-Западного государственного университета и на сайте https://swsu.ru/upload/iblock/6d0/8ccc5jq2qmprj0nuk6qzmz9855wg0aeo/Kandidatskaya-dissertatsiya-Gvozdeva-S.N.-_Sayt_.pdf

Автореферат разослан «___»_____2021 г.

Ученый секретарь
диссертационного совета Титенко Евгений Анатольевич

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность. Одними из распространенных классов цифровых управляющих систем являются системы логического управления (СЛУ). Они представляют собой параллельные многомодульные однородные системы, которые связывают тысячи параллельно работающих логических контроллеров, в совокупности решающих возложенную на них задачу логического управления некоторым объектом управления в соответствии с заданным алгоритмом логического управления. На протяжении многих лет в центре внимания многих ученых остаются проблемы анализа и синтеза систем логического управления (В.А. Горбатов, В.З. Магергут, И.В. Зотов, А.А. Баркалов, В.Г. Лазарев, Е.Н. Турута, А.Д. Закревский, С.И. Баранов, Е.И. Пийль, В.С. Харченко, А.А. Шалыто, С.А. Юдицкий, S. Husson, В.И. Варшавский, R. Puri, T. Agerwala и др.). Современные СЛУ, именуемые также логическими мультиконтроллерами (ЛМК), представляют собой параллельные многомодульные однородные мультисистемы, объединяющие тысячи параллельно работающих контроллеров. При проектировании ЛМК возникает задача разбиения комплексного параллельного алгоритма управления на блоки разбиения ограниченной сложности в соответствии со структурными и функциональными ограничениями базиса СЛУ. Одним из ограничений при построении разбиений, упрощающим внутреннюю структуру контроллеров в составе ЛМК, является отсутствие параллельных вершин в одном блоке разбиения.

Построение разбиений при ограниченных временных затратах на их получение приводит к необходимости переноса вычислительно сложных процедур, одной из которых является определение состава бинарных отношений вершин (в том числе – отношения параллельности) заданной граф-схемы алгоритма управления, с программного уровня на аппаратный путем разработки специализированных устройств-акселераторов, адаптированных к особенностям решаемой задачи. Одной из наиболее трудоемких операций при определении состава бинарных отношений вершин является задача транзитивного замыкания бинарного отношения следования, которая допускает сведение к задаче умножения бинарных матриц. Существующие устройства для умножения матриц и решения схожих задач на графах (С. Кун, В.М. Курейчик, Н.А. Лиходед, P. Dighe и др.) характеризуются либо высокой аппаратной сложностью, либо недостаточными функциональными возможностями, что ограничивает сферу их практического применения.

Таким образом, существует **противоречие**, выражающееся в необходимости снижения аппаратной сложности специализированных вычислительных устройств и недостаточным уровнем ее обеспечения в существующих технических средствах. В связи с этим **актуальной научно-технической задачей** является разработка математической модели, аппаратно-ориентированного алгоритма функционирования устройств обработки бинарных матриц, использующихся для определения состава бинарных отношений вершин параллельных граф-схем алгоритмов управления,

позволяющего снизить до практически реализуемых значений их аппаратную сложность.

Работа выполнена при поддержке гранта Президента РФ для поддержки молодых ученых – кандидатов наук «Разработка эвристических методов, алгоритмов и аппаратно-программных средств с параллельной архитектурой для решения задач дискретной комбинаторной оптимизации при проектировании однородных многомодульных мультисистем» (МК-9445.2016.8, 2016-2017 гг.) и гранта РФФИ «Разработка и анализ эффективности эвристических итерационных методов при проектировании логических мультиконтроллеров с использованием грид-систем на добровольной основе» (РФФИ 17-07-00317-а, 2017-2019 гг.).

Цель диссертационной работы – снижение аппаратной сложности устройства обработки бинарных матриц, применяемых для ускорения определения состава бинарных отношений при построении параллельных алгоритмов.

В соответствии с целью работы сформулированы следующие **основные задачи**:

1. Анализ существующих алгоритмов и практических реализаций устройства обработки бинарных матриц с целью определения состава бинарных отношений и определения направления исследования.
2. Разработка модифицированной математической модели и аппаратно-ориентированного алгоритма для умножения бинарных матриц.
3. Разработка структурно-функциональной организации устройства обработки бинарных матриц для реализации вычислительно сложных операций при определении состава бинарных отношений.
4. Оценка аппаратной сложности разработанного устройства обработки бинарных матриц.

Объект исследования: многомодульные однородные системы логического управления, направленные на реализацию параллельных алгоритмов логического управления.

Предмет исследования: алгоритмы и устройства обработки бинарных матриц.

Методы исследования: теории множеств и графов, методы математической логики, дискретных систем и устройств ЭВМ, теории проектирования конечных автоматов.

Научная новизна и основные положения, выносимые на защиту:

1. Модифицированная математическая модель бинарных отношений следования и связи вершин граф-схем параллельных алгоритмов, основанная на бинарных отношениях достижимости и контрдостижимости графов общего вида, отличающаяся введением бинарных отношений связи, альтернативы и параллельности вершин граф-схем параллельных алгоритмов, позволяющая при определении состава бинарных отношений обеспечить организацию параллельной обработки данных.
2. Алгоритм обработки бинарных матриц, позволяющий выполнить транзитивное замыкание бинарного отношения следования параллельных

алгоритмов, основанный на алгоритме умножения матриц, отличающийся возможностью досрочного прерывания вычислительного процесса при нахождении произведения бинарных матриц.

3. Структурно-функциональная организация устройства обработки бинарных матриц, основанная на систолических вычислительных структурах, отличающаяся использованием многопортового матричного запоминающего устройства с двухкоординатной адресацией и операционной части с возможностью досрочного прерывания вычислительного процесса при нахождении произведений бинарных матриц, позволяющая снижение аппаратной сложности по сравнению с аналогами.

Практическая ценность результатов исследования заключается в том, что реализация разработанных теоретических положений позволила:

– снизить аппаратную сложность разработанного устройства обработки бинарных матриц в зависимости от размера матрицы и разрядности обрабатываемых данных не менее чем в 5 раз;

– разработать устройство обработки бинарных матриц предназначенное, в том числе, для решения ряда задач на графах, которые сводятся к операциям матричного умножения (например, достижимость и контрдостижимость).

Реализация результатов работы. Результаты диссертационного исследования используются в учебном процессе Юго-Западного государственного университета в рамках следующих дисциплин по направлению 09.03.01 «Информатика и вычислительная техника»: «Параллельное программирование», «Теоретические основы организации многопроцессорных комплексов и систем», а также внедрены в ООО «РедСофтЦентр» (г.Муром) и ООО «Инвитро вижн» (г.Белгород), что подтверждается соответствующими актами.

Соответствие паспорту специальности. Согласно паспорту специальности 05.13.05 – Элементы и устройства вычислительной техники и систем управления, научная проблема, рассмотренная в диссертационной работе, соответствует пунктам 1 и 2 паспорта специальности:

1. Разработка научных основ создания и исследования общих свойств и принципов функционирования элементов, схем и устройств вычислительной техники и систем управления, в части разработки принципов функционирования устройств обработки бинарных матриц.

2. Теоретический анализ и экспериментальное исследование функционирования элементов и устройств вычислительной техники и систем управления в нормальных и специальных условиях с целью улучшения технико-экономических и эксплуатационных характеристик, в части разработки аппаратно-ориентированного алгоритма, устройства обработки бинарных матриц, обеспечивающего снижение аппаратной сложности.

Апробация результатов исследования. Основные положения диссертационной работы докладывались и получили положительную оценку на всероссийских и международных конференциях: XV Международной научно-технической конференции «Оптико-электронные приборы и устройства в системах распознавания образов и обработки изображений» (г. Курск, 2019 г.);

Всероссийской научно - технической конференции «Интеллектуальные и информационные системы» (г. Тула, 2019 г.); XXIII Международной научно-технической конференции «Медико-экологические информационные технологии – 2020» (г. Курск, 2020); на научно-технических семинарах, проводимых кафедрой «Вычислительная техника» Юго-Западного государственного университета в течение 2016-2020 гг.

Личный вклад автора. Все выносимые на защиту научные результаты получены соискателем лично. В опубликованных работах предложены: в [1,2] приведено описание особенностей использования метода взвешенного случайного перебора в задаче поиска субоптимальных разбиений граф-схем параллельных алгоритмов, программной реализации жадного метода в задаче поиска хроматического числа графа; в [3,10] представлены результаты изучения и анализа эффективности эвристических итерационных методов на базе модификации существующих решений в тестовой задаче поиска кратчайшего пути в графе с целью оценки потенциала их использования при построении разбиений; в [4, 8, 11] разработано устройство для умножения бинарных матриц, приведены результаты оценки быстродействия и аппаратная сложности; в [5] представлено описание математической модели определения состава бинарных отношений и алгоритма умножения бинарных матриц; в [6] реализована программа, предназначенная для построения разбиений граф-схем параллельных алгоритмов логического управления методом взвешенного случайного перебора, в [7] – программа для умножения плотных вещественных матриц на GPU с поддержкой технологии OpenCL; в [9, 14] создано устройство для возведения бинарной матрицы в квадрат и приведена оценка его аппаратной сложности; в [12] представлена схемотехническая реализация операции булева умножения двух квадратных бинарных матриц на базе схемы умножения i-й строки на j-й столбец матрицы; в [13] представлены результаты вычислительного эксперимента по выбору начального цвета первой вершины эвристических методов случайного перебора в задаче поиска квазиоптимальной раскраски неориентированного графа при построении разбиений.

Публикации. Результаты проведенных диссертационных исследований опубликованы в 14 научных трудах, из них четыре статьи (3 опубликованы в центральных рецензируемых научных журналах и изданиях по перечню ВАК при Минобрнауки России, 1 опубликована в журнале, индексируемом в Scopus). Получено 2 патента РФ (на полезную модель № 193927, от 21 ноября 2019 г.; на изобретение №2744239, от 04.03.2021) и 2 свидетельства о государственной регистрации программы для ЭВМ (№ 2019613452 от 18 марта 2019 г.; № 2019611362 от 01 февраля 2018 г.).

Структура и объем диссертации. Диссертационная работа состоит из введения, четырех глав, заключения, списка литературы, включающего 181 наименование, изложена на 131 странице.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИОННОЙ РАБОТЫ

Во введении обоснована актуальность работы, сформулированы цель и задачи исследований, научная новизна и положения, выносимые на защиту, практическая ценность результатов диссертационного исследования.

В первом разделе проведен анализ задач, методов, алгоритмов и их практических реализаций устройств обработки бинарных матриц, позволяющих выполнять определение состава бинарных отношений.

В связи с тем, что задача разбиения алгоритмов управления относится к классу *NP*–сложных и не допускает получение оптимального решения за приемлемое время для управляющих алгоритмов реальной сложности, на практике для ее решения применяются эвристические методы (методы ограниченного перебора, жадные методы, методы случайного перебора, метод взвешенного случайного перебора, метод муравьиной колонии, генетические алгоритмы, метод пчелиной колонии, метод капель воды, роевые методы, метод параллельно-последовательной декомпозиции). Они требуют в процессе работы информации о бинарных отношениях вершин параллельной граф-схемы алгоритма управления, для которых производится поиск разбиений, с целью организации проверки на отсутствие параллельных вершин в составе формируемых блоков разбиения. При разработке программных реализаций алгоритмов поиска разбиений существенную часть вычислительного времени при определении состава отношений занимают матричные операции, что делает актуальной задачу снижения затрат вычислительного времени путем разработки специализированных устройств обработки бинарных матриц.

В настоящее время существует большое количество важных практических задач, включающих в своем составе выполнение операции умножения матриц. Задача умножения бинарных матриц имеет ряд особенностей, к которым относятся однородность расположения исходных данных и идентичность алгоритмов их обработки, позволяющие разработку большого количества параллельных программных или аппаратных реализаций. При программной реализации одним из главных недостатков классических методов умножения является низкая эффективность использования подсистемы памяти используемой вычислительной системы (например, связки CPU – RAM или GPU – графическая память), что приводит к увеличению времени обработки и снижению реальной производительности вычислительной системы. С целью снижения влияния указанного недостатка возможен ряд оптимизаций, выполняющихся программно на алгоритмическом уровне и позволяющих повысить эффективность использования кэш-памяти CPU (снизить число промахов кэша) или разделяемой памяти GPU и, как следствие, реальную производительность используемой вычислительной системы. Другим направлением для снижения времени выполнения матричного умножения является использование специализированных вычислительных структур в виде устройств в заказном исполнении или на базе ПЛИС. Большинство известных устройств основаны на систолических вычислительных структурах, могут выполнять умножение за линейное время, однако они характеризуются большой аппаратной сложностью, что не позволяет их практическую

реализацию на современном уровне развития полупроводниковых цифровых схем для матриц большой размерности.

Существующие способы реализации матричных операций на аппаратном уровне, способные снизить время обработки, могут быть разделены на три основных направления.

1. Схемы на оптических элементах, которые по ряду причин не получили широкого распространения и в настоящее время практически не применяются.

2. Устройства умножения матриц, имеющие вероятностные свойства и присущую им статистическую погрешность, в настоящее время на практике не используются.

3. Устройства, в основу работы которых положен принцип параллельной, иногда в сочетании с конвейерной, матричной и/или систолической обработкой данных (в том числе на базе систолических структур). Они характеризуется существенным выигрышем во времени выполнения операции, в некоторых случаях позволяя выполнение умножения матриц за линейное время, однако данным устройствам необходима специализированная многопортовая память, которая должна обеспечить достаточный темп поступления исходных данных, иначе быстродействие устройства будет лимитировано именно им, а не скоростью работы операционной части.

Кроме рассмотренных выше аппаратных реализаций известны алгоритмы умножения матриц Штрассена, Штрассена-Винограда, Копперсмита-Винограда, отличительной особенность которых является более низкая временная асимптотика. Однако практический выигрыш от их использования наблюдается на матрицах значительно большего размера (миллионы элементов).

На основании вышеизложенного сделан вывод о необходимости разработки устройств обработки бинарных матриц, в основу работы которых положен принцип параллельной, в некоторых случаях в сочетании с конвейерной, матричной и/или систолической обработкой данных, что позволяет снизить их аппаратную сложность и расширить сферу практического применения.

Во втором разделе приведена модифицированная математическая модель бинарных отношений и аппаратно-ориентированный алгоритм умножения бинарных матриц. Для этой цели введены основные понятия и свойства бинарных отношений вершин граф-схем параллельных алгоритмов.

Бинарные отношения обладают такими свойствами как

- транзитивность ($\forall x, y, z \in S : xRy, yRz \Rightarrow xRz$);
- рефлексивность ($\forall x \in S : xRx$);
- симметричность ($\forall x, y \in S : xRy \Rightarrow yRx$).

С учетом рассмотренных выше свойств бинарных отношений на множестве вершин граф-схемы алгоритма используются следующие бинарные отношения:

– следования (ν) – обладает свойствами антирефлексивности, несимметричности и транзитивности;

– связи (φ) – обладает свойствами рефлексивности, симметричности и не обладает свойством транзитивности;

– параллельности (ω) – обладает свойствами рефлексивности, симметричности и нетранзитивности;

– альтернативы (ψ) – обладает свойствами антирефлексивности, симметричности и нетранзитивности.

Аналогичными отношениям следования и связи, которые определены для граф-схем параллельных алгоритмов, являются бинарные отношения контрдостижимости, достижимости и связности вершин графов общего вида.

В первую очередь выполняется определение отношения следования. Для этого производится его транзитивное замыкание, отталкиваясь от начальных значений, получаемых исходя из рассмотрения всех дуг передачи управления между вершинами в составе граф-схемы:

$$\left(a_i \nu a_j\right) \wedge \left(a_j \nu a_k\right) \Rightarrow \left(a_i \nu a_k\right), \tag{1}$$

где a_i, a_j, a_k – вершины граф-схемы параллельных алгоритмов.

Для хранения бинарного отношения следования применяется матрица отношений $M_R^{\nu} = \left(m_{ij}^{\nu}\right)$, $i, j = \overline{1, N}$, N – число вершин в граф-схеме алгоритма. В исходной матрице необходимо найти такие i, j и k, что истинно выражение

$$\left(a_i \nu a_j\right) \wedge \left(a_j \nu a_k\right) \wedge \neg \left(a_i \nu a_k\right), \tag{2}$$

или, что то же самое:

$$\left(m_{ij}^{\nu} = 1\right) \wedge \left(m_{jk}^{\nu} = 1\right) \wedge \left(m_{ik}^{\nu} = 0\right) \tag{3}$$

и положить $m_{ik}^{\nu} := 1$, где m_{ik}^{ν} – значение бинарного отношения в составе соответствующей матрицы отношений M_R^{ν}, повторяя указанное действие до тех пор, пока возможно нахождение элементов, удовлетворяющих (2).

При практической реализации указанного действия существует два подхода. Первый базируется на алгоритме Флойда-Уоршала, позволяющим выполнить транзитивное замыкание отношения за один проход ввиду использования особого порядка рассмотрения элементов матриц, что ограничивает возможность его распараллеливания. Второй основан на возведении матрицы в степень.

Данная операция реализуется двумя способами. Первый из них выполняется путем умножения исходной матрицы бинарного отношения самой на себя:

$$M_R^{\nu\prime} = M_R^{\nu} \times M_R^{\nu} \times \ldots \times M_R^{\nu}. \tag{4}$$

Число матричных умножений в формуле (4) определяется исходными данными и ограничено сверху значением N, т.е. в худшем случае искомое транзитивное замыкание будет найдено путем возведения матрицы в N-ю

степень. На практике, умножения прерывается начиная с момента, когда после выполнения очередного умножения матрица не изменится.

Второй способ основан на возведении матрицы в квадрат по следующей схеме:

$$\begin{aligned} M_2^{\nu} &= M_R^{\nu} \times M_R^{\nu}, \\ M_4^{\nu} &= M_2^{\nu} \times M_2^{\nu}, \\ M_8^{\nu} &= M_4^{\nu} \times M_4^{\nu}, \\ &\ldots \end{aligned} \tag{5}$$

Результирующее значение матрицы, обладающей свойством транзитивного замыкания, будет получено в худшем случае за $\lceil \log_2 N \rceil$ шагов. При практической реализации действия, как и в рассмотренной выше ситуации, изменение матрицы может прекратиться значительно раньше, однако в худшем случае потребуется число шагов, пропорциональное логарифму от N.

В процессе нахождения произведения бинарных матриц базовой операцией является операция нахождения бинарного скалярного произведения i-го столбца и j-й строки матриц. Оно выполняется по формуле:

$$m_{ij}^{\nu\prime} = m_{ij}^{\nu} \vee \left(\bigvee_{k=\overline{1,N}} m_{ik}^{\nu} m_{kj}^{\nu} \right). \tag{6}$$

В результате выполнения данной операции N^2 раз для всех $i, j = \overline{1, N}$ получается результирующая матрица. Новизной алгоритма является сокращение числа итераций внутреннего цикла (по переменной k) при получении единичного значения на одной из итераций нахождения скалярного произведения двоичных векторов.

При умножении матриц с большим числом единиц число выполняемых итераций по k будет малым, что уменьшает как число выполняемых операций (конъюнкций и дизъюнкций), так и число обращений к памяти. Программная реализация с использованием универсальных конвейерных суперскалярных вычислителей данного способа экономии числа выполняемых операций имеет ряд недостатков, основным из которых является нерегулярно срабатывающее условие прерывания внутреннего цикла, приводящее к сбросам конвейера при спекулятивном выполнении соответствующего программного кода. Они снижают реальную производительность вычислительной системы при выполнении умножения неплотных бинарных матриц, выполняемого для реализации транзитивного замыкания.

После выяснения отношения следования производится определение отношения связи путем выполнения покомпонентной конъюнкции:

$$M_R^{\varphi} = M_R^{\nu} \vee \left(M_R^{\nu} \right)^T. \tag{7}$$

Выяснение отношения альтернативы производится путем анализа путей между условными вершинами и вершинами объединения альтернативных дуг, в результате чего оказывается сформирована матрица M_R^{ψ}. Указанная операция

не зависит по данным от рассмотренных выше операций выяснения отношения следования и связи и может быть выполнена параллельно с ними. На завершающем этапе определения состава бинарных отношений выполняется выяснение отношения параллельности:

$$M_R^{\omega} = \overline{M_R^{\varphi}} \wedge \overline{M_R^{\psi}}. \tag{8}$$

Граф-схема параллельного алгоритма определения состава бинарных отношений представлена на рис. 1.

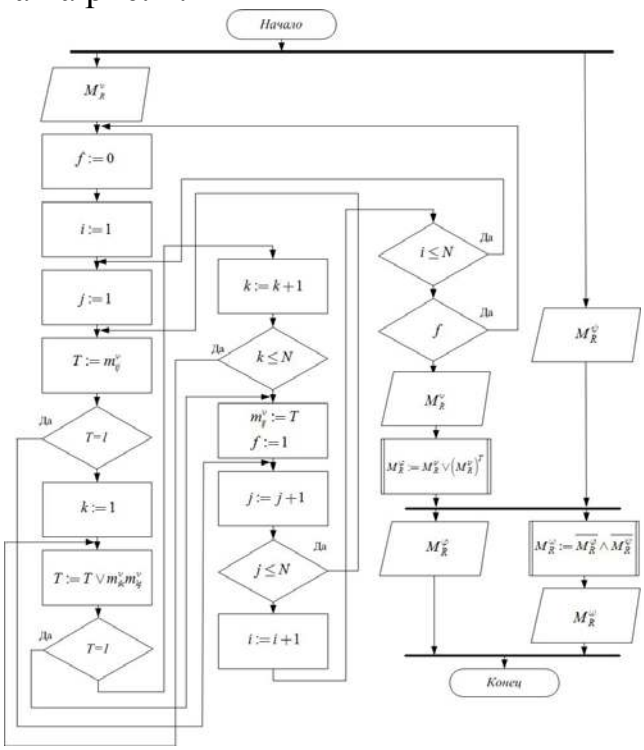


Рис. 1. Параллельный алгоритм определения состава бинарных отношений

Формулы (1) – (3) определяют свойства используемых бинарных отношений и в совокупности с (6) – (8) образуют модифицированную математическую модель работы устройства обработки бинарных матриц. Новизной математической модели является введение бинарных отношений связи, альтернативы и параллельности вершин граф-схем параллельных алгоритмов в дополнение к отношению следования, известному в теории графов.

Таким образом, эффективным способом практической реализации является разработка специализированной аппаратно-ориентированной структурно-функциональной организации устройства обработки бинарных матриц. Разработанная математическая модель и алгоритм позволяют осуществить практическую реализацию устройства для умножения бинарных матриц с прерыванием внутреннего цикла.

В третьем разделе разработана структурно-функциональная организация устройства обработки бинарных матриц, приведены соответствующие структурные схемы и описание их работы.

На рисунке 2 приведена структурная схема подключения устройства обработки бинарных матриц (УОБМ) к современным вычислительным

системам с использованием шины PCI Express. Она включает в своем составе запоминающее устройство (ЗУ) 1, устройство обработки бинарных матриц, матрицы логических элементов ИЛИ 3 (М ИЛИ) и И 4 (М И), использующиеся при выяснении бинарных отношений связи и параллельности. Запоминающее устройство подключено к шине PCI Express с целью загрузки исходных данных для построения матрицы отношений и выгрузки результата. УОБМ 2 включает в своем составе одно из предложенных в диссертации устройств.

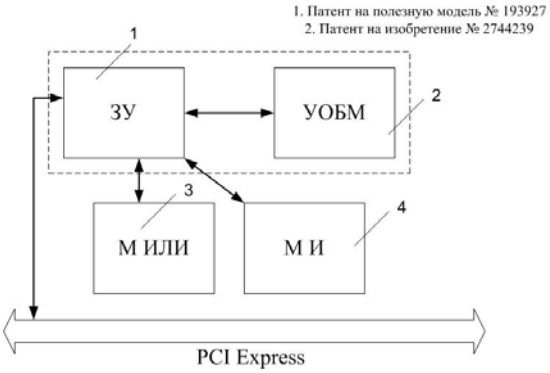


Рис. 2. Структурная схема устройства обработки бинарных матриц

При программной обработке матрица хранится в оперативной памяти в виде двумерного массива бинарных значений, элементы матрицы располагаются в памяти подряд, а для обращения к элементу требуется вычисление адреса. При схемотехнической реализации это приводит к появлению в схеме дополнительных сумматоров и блоков умножения, увеличивая аппаратную сложность, тепловыделение и время задержки распространения сигнала. Выходом из данной ситуации является разработка специализированного запоминающего устройства с двухкоординатной адресацией (блока коэффициентов матрицы), ориентированного на хранение именно матричной информации. На рисунке 3 приведена схема блока коэффициентов матрицы, на рисунке 4 – функциональная схема ячейки блока хранения битовых признаков бинарного отношения.

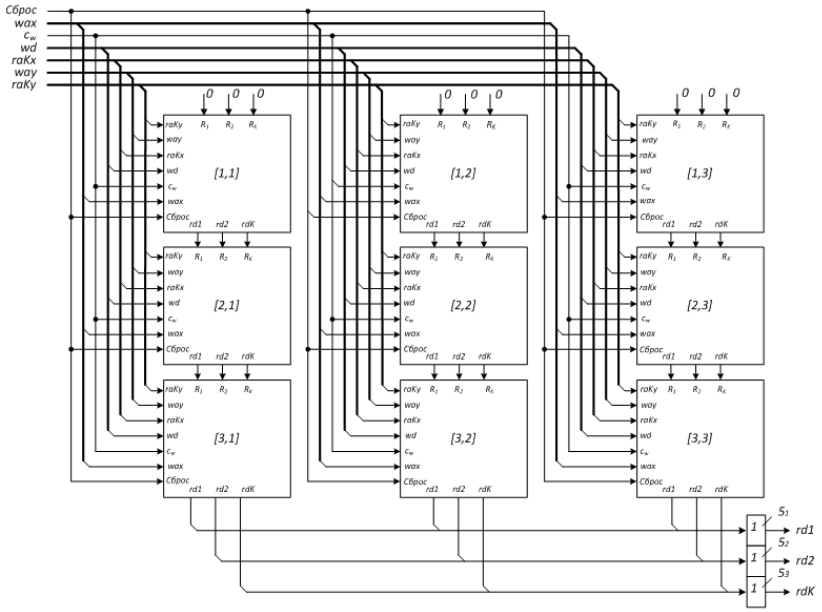


Рис. 3. Схема блока коэффициентов матрицы
(пример запоминающего устройства со структурой $3\times3\times1$)

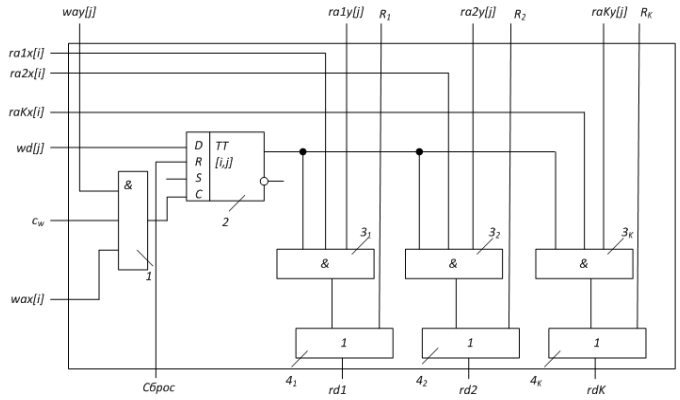


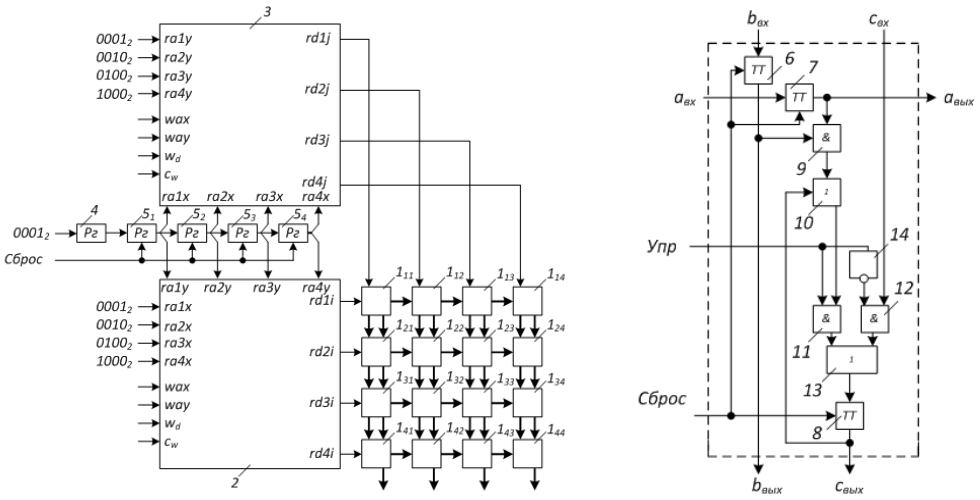
Рис. 4. Функциональная схема ячейки блоков хранения битовых признаков бинарного отношения

Существуют два основных направления разработки устройств обработки бинарных матриц: на базе систолических структур и на базе аппаратной реализации алгоритмов классического умножения матриц с возможностью параллельной обработки информации.

Устройства первого направления характеризуется высоким быстродействием, однако они обладают чрезмерно большой аппаратной сложностью, являющейся препятствием для их практической реализации при умножении матриц большого размера. Устройства обработки бинарных матриц, ориентированные на аппаратную реализацию алгоритмов классического умножения, характеризуются умеренным быстродействием и низкой аппаратной сложностью и позволяют осуществить прерывание вычислительного процесса при умножении битовых векторов в соответствии с алгоритмом, приведенным на рис. 1. В данном диссертационном исследовании выполнена разработка двух способов реализации устройства обработки бинарных матриц (по одному для каждого направления), с целью сравнения их аппаратной сложности и быстродействия.

Устройство обработки бинарных матриц на базе систолических структур, его функциональная схема и схема операционного блока приведены на рисунке 5.

Значения элементов умножаемых матриц загружаются в блоки коэффициентов матриц 2 и 3, загрузка элементов первой и второй матриц может быть совмещена во времени.



а) б)

Рис. 5. Функциональная схема устройства обработки бинарных матриц на базе систолических структур (а); схема операционного блока устройства обработки бинарных матриц на базе систолических структур (б)

Значения с выходов блоков коэффициентов матрицы 2 и 3 поступают на вход триггеров 6 и 7, соответственно, матрицы операционных блоков. С выходов триггеров 6 и 7 значения подаются на входы элемента И 9, на выходе которого формируется их конъюнкция. Данные с выхода элемента И 9 поступают на второй вход элемента ИЛИ 10, а на первый вход подается значение из триггера 8, с выхода которого сформированное значение поступает на второй вход элемента И 11. Значение со второго входа элемента И 11 проходит на его выход. Значение с выхода элемента И 11 проходит через элемент ИЛИ 13 на вход первой ступени триггера 8, где фиксируется по пришествии соответствующего синхросигнала, что обеспечивает формирование в триггерах 8 операционных блоков 1 искомых значений.

Функциональная схема устройства обработки бинарных матриц на базе аппаратной реализации алгоритмов классического умножения матриц приведена на рисунке 7.

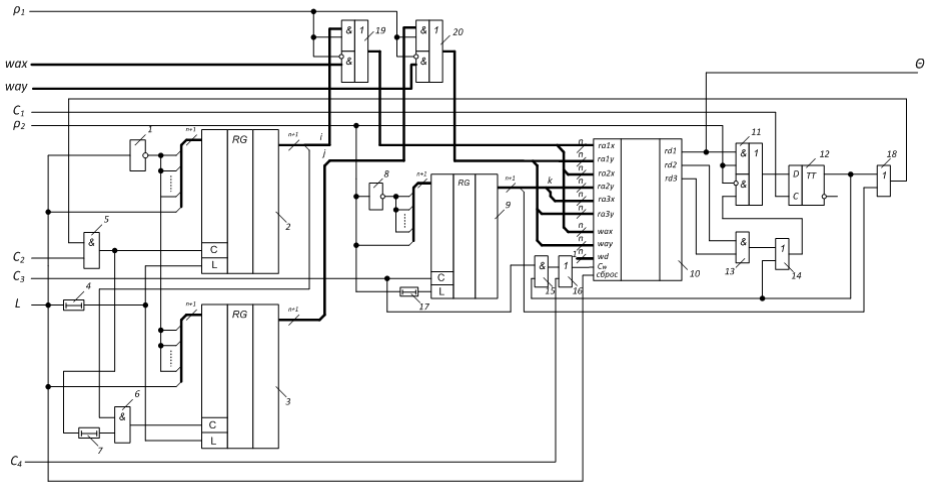


Рис. 7. Схема устройства обработки бинарных матриц на базе аппаратной реализации алгоритмов классического умножения матриц

Перед началом работы устройства на информационных входах сдвиговых регистров 2 и 3 формируются значения «00…01». На входы адреса записи строки и адреса записи столбца устройства внешним устройством поочередно подаются адреса строк и столбцов элементов загружаемой матрицы, имеющих единичное значение. При умножении матрицы значение с выхода *rd1* блока коэффициентов матрицы 10 проходит через коммутатор 11 и попадает на информационный вход двухступенчатого триггера 12. Сигнал со входа ρ_2 проходит через инвертор 8 и формирует на информационном входе сдвигового регистра 9 значение «00…01». На выходе элемента ИЛИ 14 формируется значение $t \vee m_{ik} m_{kj} = m_{ij} \vee m_{i1} m_{1j} \vee m_{i2} m_{2j} \vee ... \vee m_{ik} m_{kj}$, которое подается на второй вход коммутатора 11.

Синхросигнал со входа C_3 устанавливается в единичное значение, которое производит сдвиг влево содержимого сдвигового регистра 9. Если на данной итерации в двухступенчатом триггере 12 хранится единичное значение, оно открывает элемент И 15 для прохождения синхросигнала со входа C_3 через элемент ИЛИ 16 на синхровход записи c_w блока коэффициентов матрицы 10, что приводит к записи единичного значения в элемент m_{ij} матрицы взамен предыдущего нулевого. В противном случае новое значение k обеспечивает чтение новых значений m_{ik} и m_{kj} матрицы из блока коэффициентов матрицы 10 – производится переход к новой итерации.

После завершения умножения *i*-й строки на *j*-ый столбец единичное значение с выхода элемента ИЛИ 18 открывает элемент И 5 для прохождения синхросигнала со входа C_2 на синхровход регистра 2, обеспечивая сдвиг его содержимого в сторону старших разрядов. Далее процесс умножения повторяется для *(i+1)*-й строки и *j*-го столбца аналогично рассмотренному выше. После выполнения *n* итераций умножения единичное значение в составе сдвигового регистра 2 попадает в *(n+1)*-й разряд. С выхода сдвигового регистра 2 указанное значение открывает элемент И 6 для прохождения синхросигнала со входа C_2 через элементы И 5 и элемент задержки 7 на синхровход сдвигового регистра 3. Поступление n^2 синхроимпульсов на вход C_2 обеспечивает получение n^2 результатов умножения m_{ij}.

Таким образом, предложена структурно-функциональная организация устройства обработки бинарных матриц: на базе систолических структур, отличающаяся от прототипа узкой ориентацией на умножение бинарных матриц, и на базе аппаратной реализации алгоритмов классического умножения матриц, позволяющего аппаратную реализацию прерывания внутреннего цикла в соответствии с алгоритмом, приведенным на рисунке 1, позволяющие сокращение аппаратной сложности по сравнению с известными устройствами, что подтверждается расчетами, приведенными в разделе 4.

В четвертом разделе приведены оценки аппаратной сложности и быстродействия разработанного устройства обработки бинарных матриц: на базе систолических структур и на базе аппаратной реализации алгоритмов

классического умножения матриц и их сравнение с устройством для умножения матриц.

При умножении бинарных матриц по формуле (6) работу соответствующего алгоритма и его практической реализации (программной или аппаратной) досрочно прерывается в случае, если текущее значение частичной конъюнкции при умножении бинарных векторов равно 1 (дальнейшие действия не выполняются ввиду того, что $1 \vee x = 1$). Это позволяет экономить время на умножение матриц за счет сокращения числа итераций внутреннего цикла.

С целью оценки соответствующей экономии используется понятие плотности *d* умножаемых матриц размера $N \times N$:

$$d = \frac{M}{N^2}, \tag{7}$$

где *M* – число единиц в умножаемых матрицах, $0 \le d \le 1$, и вероятности α досрочного прекращения операции умножения бинарных векторов, $0 \le \alpha \le 1$.

Зависимость вероятности $\alpha = 1 - \beta$ (β – вероятность выполнения операций умножения) досрочного прекращения умножения бинарных векторов от размера *N* умножаемых матриц и их плотности *d* является нетривиальной, не выражается аналитическими формулами и была установлена эмпирически в ходе вычислительного эксперимента. Для этого формируется случайная бинарная матрица размера $N \times N$, включающая в своем составе $M = \lfloor dN^2 \rfloor$ единиц, для нее производится возведение в квадрат, в ходе выполнения которого подсчитывается необходимое для этого число конъюнкций (логических умножений) *K*, по которому вычисляется вероятность досрочного прекращения умножения $\alpha = 1 - \beta = 1 - \frac{K}{N^3}$ (N^3 – число умножений элементов матриц без досрочного прерывания). Соответствующая зависимость $\beta(d)$ для *N=64* представлена на рисунке 8.

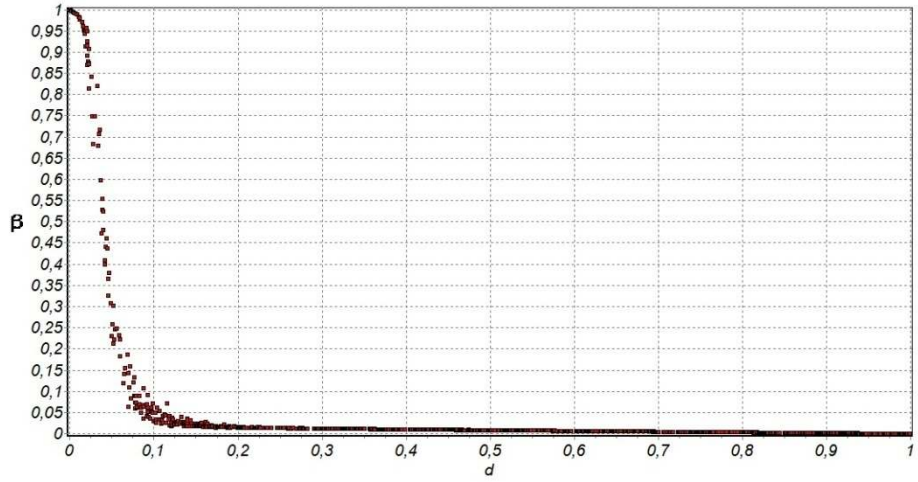


Рис. 8 – Зависимость вероятности выполнения умножения (β) от плотности умножаемых матриц (d) (пример для $N = 64$)

Анализ полученных результатов позволяет сделать вывод о том, что с ростом размера *N* умножаемых матриц число выполняемых умножений

$K \ll N^3$, значение величины $\beta < 0{,}1$ для $N = 10$, $d > 0{,}5$; $\beta < 0{,}01$ для $N = 100$, $d > 0{,}1$, и раннее прекращение операции умножения битовых векторов (в приведенном примере – на 2–3 порядка) сокращает необходимое число умножений элементов матриц при вычислении произведения бинарных матриц.

Оценка аппаратной сложности произведена в эквивалентных вентилях (ЭВ) (одно- или двухвходовых логических элементах, выполняющих элементарную логическую операцию).

Аппаратная сложность *устройства для умножения матриц* (R_3) (прототип, патент РФ на полезную модель № 157948) в целом складывается из сложности блоков блока хранения (R_1), набора регистров и матрицы $n \times n$ операционных блоков (R_2) и вычисляется по формулам:

$$R_1 = 4mn^2 + 3mn^3 + 3n^2 - n,$$
$$R_2 = 35mn^2 + 3m^2n^2,$$
$$R_3 = 43mn^2 + 3m^2n^2 + 6mn^3 + 14n^2 + 2n, \quad (8)$$

где n – размер обрабатываемых матриц, m – разрядность обрабатываемых данных.

Аппаратная сложность устройства обработки бинарных матриц на базе систолических структур (R_6) (патент на полезную модель № 193927) в целом складывается из сложности блоков блока хранения (R_4), набора регистров и матрицы $n \times n$ операционных блоков (R_5) и вычисляется по формулам:

$$R_4 = 7n^2 + 3n^3 - n,$$
$$R_5 = 30n^2,$$
$$R_6 = 6n^3 + 52n^2 + 2n. \quad (9)$$

Аппаратная сложность устройства обработки бинарных матриц на базе аппаратной реализации алгоритмов классического умножения матриц (R_8) (патент на изобретение № 2744239) складывается из аппаратной сложности блока коэффициентов матрицы (R_7) и аппаратной сложности элементов устройства, образующих его операционную часть и вычисляется по формулам:

$$R_7 = 15n^2 + 3n - 3.$$
$$R_8 = 15n^2 + 45n + 59. \quad (10)$$

Значения величины аппаратной сложности устройства для умножения матриц (R_3) и устройства обработки бинарных матриц: на базе систолических структур (R_6) и на базе аппаратной реализации алгоритмов классического умножения матриц (R_8) приведены в таблице 1.

Таблица 1.

Результаты оценки аппаратной сложности разработанных устройств

m	n	Устройство для	Устройство обработки бинарных матриц

		умножения матриц (прототип, патент РФ на полезную модель № 157948), ЭВ (R_3)	на базе систолических структур (патент на полезную модель № 193927), ЭВ (R_6)	на базе аппаратной реализации алгоритмов классического умножения матриц (патент на изобретение № 2744239), ЭВ (R_8)
8	10	$1{,}0\times10^5$	$1{,}1\times10^4$	$2{,}0\times10^3$
8	100	$5{,}4\times10^7$	$6{,}5\times10^6$	$1{,}5\times10^5$
8	1000	$4{,}9\times10^{10}$	$6{,}1\times10^9$	$1{,}5\times10^7$
16	10	$2{,}4\times10^5$	$1{,}1\times10^4$	$2{,}0\times10^3$
16	100	$1{,}1\times10^8$	$6{,}5\times10^6$	$1{,}5\times10^5$
16	1000	$9{,}7\times10^{10}$	$6{,}1\times10^9$	$1{,}5\times10^7$
32	10	$6{,}4\times10^5$	$1{,}1\times10^4$	$2{,}0\times10^3$
32	100	$2{,}4\times10^8$	$6{,}5\times10^6$	$1{,}5\times10^5$
32	1000	$2{,}0\times10^{11}$	$6{,}1\times10^9$	$1{,}5\times10^7$
64	10	$1{,}9\times10^6$	$1{,}1\times10^4$	$2{,}0\times10^3$
64	100	$5{,}3\times10^8$	$6{,}5\times10^6$	$1{,}5\times10^5$
64	1000	$4{,}0\times10^{11}$	$6{,}1\times10^9$	$1{,}5\times10^7$
128	10	$6{,}2\times10^6$	$1{,}1\times10^4$	$2{,}0\times10^3$
128	100	$1{,}3\times10^9$	$6{,}5\times10^6$	$1{,}5\times10^5$
128	1000	$8{,}2\times10^{11}$	$6{,}1\times10^9$	$1{,}5\times10^7$

Снижение аппаратной сложности устройства обработки бинарных матриц на базе систолических структур достигается за счет хранения и обработки 1 бита информации для каждого коэффициента обрабатываемых бинарных матриц вместо m бит в составе устройства для умножения матриц, что позволяет снизить аппаратную сложность как схем хранения (блоков коэффициентов матриц), так и схем обработки информации (матрица операционных блоков) не менее чем в 8 раз в зависимости от размера матрицы n и разрядности обрабатываемых данных m.

Из результатов расчета видно, что устройство обработки бинарных матриц на базе аппаратной реализации алгоритмов классического умножения матриц обладает не менее чем в 5 раз меньшей аппаратной сложностью по сравнению с устройством обработки бинарных матриц на базе систолических структур в зависимости от размера матрицы n.

Устройство обработки бинарных матриц на базе аппаратной реализации алгоритмов классического умножения матриц обладает не менее чем в 5 раз меньшей аппаратной сложностью по сравнению с устройством обработки бинарных матриц на базе систолических структур, и не менее чем в 50 раз меньшей аппаратной сложностью по сравнению с устройством для умножения матриц в зависимости от размера матрицы n и разрядности обрабатываемых данных m за счет более простой операционной части.

Выполнена сравнительная оценка быстродействия устройства для умножения матриц и устройства обработки бинарных матриц: на базе систолических структур и на базе аппаратной реализации алгоритмов классического умножения матриц.

Оценка быстродействия устройств обработки бинарных матриц производится по формулам, приведенным в таблице 2:

Таблица 2.

Формулы для оценки быстродействия разработанных устройств, t_0 – время работы одного эквивалентного вентиля

Название	Формула для оценки быстродействия
Устройство для умножения матриц (прототип, патент РФ на полезную модель № 157948)	$t_{общ1} = 8t_0n^2 + t_0 + 6t_0n + 9t_0m(2n-1),$
Устройство обработки бинарных матриц	
на базе систолических структур (патент на полезную модель №193927)	$t_{общ2}^{(n)} = \left(14n^2 + 2 + \lceil \log_2 n \rceil + n\right)t_0,$
на базе аппаратной реализации алгоритмов классического умножения матриц (патент на изобретение № 2744239)	$t_{общ3} = 3t_0 + 4t_0 \times d \times n + \\ + \left(3t_0n^2 + t_0n \times \left(12 + \lceil \log_2 n \rceil\right)\right) \times \beta \times n^2 + \\ + 2t_0n^2 + 3t_0n^3 + \lceil \log_2 n \rceil t_0n^2.$

В таблице 3 приведены результаты оценки быстродействия устройств обработки бинарных матриц (для сравнения взяты усредненные значения $m = 16,\ \beta = 0{,}5,\ t_0 = 1$ нс).

Таблица 3.

Результаты оценки быстродействия разработанных устройств

n	Устройство для умножения матриц (прототип, патент РФ на полезную модель № 157948), мс	Устройство обработки бинарных матриц	
		на базе систолических структур (патент на полезную модель № 193927), мс	на базе аппаратной реализации алгоритмов классического умножения матриц (патент на изобретение № 2744239), мс
4	0,145	0,014	0,0008
8	0,145	0,027	0,0039
16	0,145	0,052	0,0243
32	0,145	0,101	0,1759
64	0,198	0,198	1,3148
128	0,391	0,391	10,0733

Полученные оценки быстродействия позволяют сделать вывод о том, что быстродействие предложенных устройств сопоставимо с быстродействием прототипа, в то время как их аппаратная сложность на 1 – 2 порядка ниже. Таким образом, поставленная в диссертационном исследовании цель, заключающаяся в снижении аппаратной сложности разработанных устройств, достигнута.

В заключении сформулированы основные результаты диссертационного исследования.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

В диссертационной работе в рамках решения поставленной научно-технической задачи, заключающейся в сокращении аппаратной сложности

устройства обработки бинарных матриц, используемого для определения состава бинарных отношений при построении разбиений граф-схем параллельных алгоритмов логического управления получены следующие результаты:

1. Проведен анализ существующих алгоритмов, программных и аппаратных реализаций устройств для обработки матриц, в ходе которого выявлены три способа реализации матричных операций на аппаратном уровне, способных снизить время обработки. По итогу обзора сделан вывод о необходимости разработки устройств обработки бинарных матриц, в основу работы которых положен принцип параллельной, в некоторых случаях в сочетании с конвейерной, матричной и/или систолической обработкой данных, что позволяет снизить их аппаратную сложность и расширить сферу практического применения
2. Разработана модифицированная математическая модель бинарных отношений граф-схем параллельных алгоритмов, позволяющая свести задачу определения состава указанной пары отношений к задаче транзитивного замыкания бинарного отношения, решаемую путем нахождения произведения бинарной матрицы самой на себя.
3. На базе модифицированной математической модели бинарных отношений разработан аппаратно-ориентированный алгоритм для умножения бинарных матриц, ориентированный на параллельную аппаратную реализацию, позволяющий перенести вычислительно сложные процедуры умножения бинарных матриц на аппаратный уровень. Особенностью данного алгоритма является сокращение числа итераций внутреннего цикла за счет возможности досрочного прерывания умножения матриц, что реализовано в рамках разработанного устройства обработки бинарных матриц на базе аппаратной реализации алгоритмов классического умножения матриц.
4. Разработана структурно-функциональная организация устройства обработки бинарных матриц: на базе систолических структур и ориентированное на аппаратную реализацию алгоритмов классического умножения матриц, отличающихся низкой аппаратной сложностью по сравнению с известными устройствами за счет их узкой ориентации на обработку бинарных матриц.
5. В ходе проведенных вычислительных экспериментов были получены зависимости вероятности досрочного прекращения умножения бинарных векторов от размера умножаемых матриц и их плотности, анализ которых позволил сделать вывод о том, что досрочное прерывание процесса умножения сокращает временные затраты на 2 – 3 порядка по сравнению с умножением матриц общего вида. На основе полученных зависимостей была произведена оценка быстродействия разработанного устройства, которая показала сопоставимые значения в сравнении с устройством прототипом. При этом аппаратную сложность разработанного устройства удалось снизить не менее чем в 50 раз для устройства обработки бинарных матриц на базе аппаратной реализации алгоритмов классического умножения матриц по сравнению с прототипом в зависимости от размера матрицы и разрядности

обрабатываемых данных, и в не менее чем в 5 раз для устройства обработки бинарных матриц на базе систолических структур по сравнению с прототипом.

Перспективы дальнейшей разработки темы. Созданные алгоритм и устройства могут найти применение в составе аппаратно-программных комплексов по разработке СЛУ в базисе ЛМК, позволяя снизить время, затрачиваемое на проектирование. Дальнейшее повышение быстродействия предложенных устройств возможно за счет введения параллельной и конвейерной обработки матриц.

СПИСОК НАУЧНЫХ РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в рецензируемых научных журналах

1. Гвоздева С.Н. Метод взвешенного случайного перебора для построения разбиений граф-схем параллельных алгоритмов при проектировании логических мультиконтроллеров / Ватутин Э.И., Панищев В.С., Гвоздева С.Н., Титов В.С. // Известия ЮЗГУ. 2017. Т. 21. № 6 (75). С. 6–21. DOI: 10.21869/2223-1560-2017-21-6-6-21.
2. Гвоздева С.Н. О влиянии порядка рассмотрения вершин при поиске раскрасок графов общего вида с использованием жадного алгоритма / Пшеничных А.О., Ватутин Э.И., Гвоздева С.Н. // Высокопроизводительные вычислительные системы и технологии. Т. 3., № 1, 2019. С. 101-106.
3. Vatutin E., Panishchev V., Gvozdeva S., Titov V. Comparison of Decisions Quality of Heuristic Methods Based on Modifying Operations in the Graph Shortest Path Problem // Problems of Information Technology. No. 1. 2020. pp. 3–15. DOI: 10.25045/jpit.v11.i1.01.(Scopus)
4. Гвоздева С.Н. Оценка быстродействия устройства с систолической структурой для умножения бинарных матриц / Гвоздева С.Н., Ватутин Э.И., Титов В.С. // Телекоммуникации. № 3. 2020. С. 2-10.
5. Гвоздева С.Н. Математическая модель определения состава бинарных отношений и алгоритм умножения бинарных матриц / Гвоздева С.Н. // Известия ЮЗГУ. 2021. № 2 (75). С. 81–98.

Патенты, программы

6. Гвоздева С.Н. Программа для построения разбиений граф-схем параллельных алгоритмов логического управления методом взвешенного случайного перебора / Ватутин Э.И., Панищев В.С., Гвоздева С.Н. // Свидетельство об официальной регистрации программы для ЭВМ №2018611362 Российская Федерация, заявл. 04.12.2017; зарегистрировано 01.02.2018.
7. Гвоздева С.Н. Программа для умножения плотных вещественных матриц на GPU с поддержкой технологии OpenCL / Ватутин Э.И., Затолокин Ю.А., Гвоздева С.Н., Титов В.С. // Свидетельство об официальной регистрации программы для ЭВМ №2019613452 Российская Федерация, заявл. 28.02.2019; зарегистрировано 18.03.2019.

8. Гвоздева С.Н. Устройство для умножения бинарных матриц / Гвоздева С.Н., Ватутин Э.И., Пшеничных А.О., Титов В.С.// Патент на полезную модель №193927. Заявка № 2019119879 от 21.11.2019.Опубликовано 21.11.2019г. Бюл.№33

9. Гвоздева С.Н. Устройство для возведения бинарной матрицы в квадрат / Гвоздева С.Н., Ватутин Э.И., Титов В.С. // Патент на изобретение № 2744239. Заявка № 2020122205 от 05.07.2020. .Опубликовано 04.03.2021г. Был.№7.

Материалы конференций

10. Gvozdeva S.N. Comparison of Decisions Quality of Heuristic Methods Based on Modifying Operations in the Graph Shortest Path Problem / Vatutin E.I., Panishchev V.S., Titov V.S., Gvozdeva S.N. // IX International Conference on Optimization Methods and Applications "Optimization and Applications (OPTIMA-2018)", Book of Abstracts. Moscow, Petrovac, 2018. P. 171.

11. Гвоздева С.Н. Оценка аппаратной сложности устройства умножения квадратных бинарных матриц размером $n \times n$ / Гвоздева С.Н., Ватутин Э.И. // Оптико-электронные приборы и устройства в системах распознавания образов и обработки изображений (Распознавание – 2019): сборник материалов XV международной научно-технической конференции / ред. кол.: С.Г. Емельянов [и др.]; Юго-Зап. гос. ун-т. Курск, 2019. С. 66-68.

12. Гвоздева С.Н. Последовательное устройство для умножения бинарных матриц / Гвоздева С.Н., Мартынов И.А., Ватутин Э.И. // Интеллектуальные и информационные системы. Труды Всероссийской научно-технической конференции. Тула: Изд-во ТулГУ, 2019, 422 с. – С. 37-43.

13. Гвоздева С.Н. О влиянии вероятности выбора минимально допустимого или случайного цвета для метода случайного перебора эвристической оценки хроматического числа графа / Пшеничных А.О., Гвоздева С.Н., Панищев В.С., Ватутин Э.И. // Интеллектуальные и информационные системы. Труды Всероссийской научно-технической конференции. Тула: Изд-во ТулГУ, 2019. С. 59-63

14. Гвоздева С.Н. Оценка аппаратной сложности устройства для возведения бинарной матрицы в квадрат / Гвоздева С.Н., Ватутин Э.И. // Медико-экологические информационные технологии -2020: сборник научных статей по материалам XXIII Международной научно-технической конференции: в 2ч. Ч.2 / редкол.: Кореневский Н.А. (отв.ред.) и [и др.]; Юго-Зап.гос.ун-т. Курск, 2020. С. 62-65.

Подписано в печать _____2021. Формат 60х84 1/16.
Печатных листов 1,0. Тираж 120 экз. Заказ _____.
Юго-Западный государственный университет,
305040, Курск, ул. 50 лет Октября, 94.