

«УТВЕРЖДАЮ»

Директор Федерального
государственного учреждения
«Федеральный научный центр Научно-
исследовательский институт системных
исследований Российской академии
наук», доктор технических наук



Власов С.Е.

«13» ноября 2023 г.

Отзыв

ведущей организации на диссертационную работу Хилько Дмитрия
Владимировича

«Исследование и разработка потоковой рекуррентной архитектуры для
эффективной реализации параллелизма в области цифровой обработки сигналов»,
представленную на соискание ученой степени кандидата технических наук по
специальности 2.3.2 – «Вычислительные системы и их элементы».

Актуальность темы исследования

Исследование, результаты которого представлены в диссертации Д.В. Хилько, посвящено разработке прототипа нетрадиционной потоковой рекуррентной архитектуры, который предназначен для решения задач цифровой обработки сигналов (ЦОС) реального времени, а также способов и средств его программирования и отладки. На сегодняшний день задачи ЦОС решаются во множестве областей человеческой деятельности, в том числе, в таких критических областях, как автоматизированные системы управления, связь и телекоммуникации. Как следствие, в мире постоянно ведутся исследования и разработка высокопроизводительных цифровых сигнальных процессоров (ЦСП), большинство из которых основаны на принципах традиционной фон-Неймановской или Гарвардской архитектуры. Высокая производительность современных ЦСП достигается путем реализации широкого набора средств поддержки параллельных вычислений за счет значительной аппаратной избыточности, обусловленной изначально последовательным характером вычислительного процесса в традиционных архитектурах.

В тоже время, потоковая архитектура, в которой продвижение вычислительного процесса осуществляется по готовности данных, поддерживает параллелизм естественным образом, что, теоретически, должно устранить

необходимость значительной аппаратной избыточности. Исследователи потоковой архитектуры также отмечают, что ее принципы хорошо сочетаются с требованиями задач ЦОС. Тем не менее, зарубежные разработки потоковых архитектур не продвинулись дальше прототипов ввиду целого ряда проблем и были практически остановлены. Наиболее активно в данной области исследования продвигаются в Российской Федерации. В частности, была разработана концепция потоковой рекуррентной архитектуры, ключевые принципы которой позволяют частично или полностью решить упомянутые проблемы потоковых архитектур. В диссертационной работе Д.В. Хилько указанным вопросам уделяется большое внимание. Поэтому тема диссертационного исследования, безусловно, актуальна.

Структура и содержание диссертации

Во Введении представлено краткое описание истории разработки потоковых архитектур как зарубежными, так и российскими исследователями. Дана краткая характеристика проблем реализации потоковых архитектур, а также основных направлений, в которых ведутся исследования в данной области. Автором обозначены ключевые особенности новой рекуррентной архитектуры, которые, теоретически, должны обеспечить ей конкурентное преимущество. В общем виде формулируются: цель диссертации; задачи, которые решены в ходе ее подготовки.

В главе 1 автор представляет результаты анализа предметной области параллельных вычислительных систем вообще и потоковых архитектур в частности, в том числе и объект исследования диссертационной работы – многоядерная потоковая рекуррентная архитектура (МПРА).

В разделе 1.1 приводится классификация параллельных вычислительных систем и способов реализации параллелизма.

В разделе 1.2 со ссылками на многочисленные источники рассматриваются ключевые принципы потоковых архитектур и модели вычислений, на которой они основаны. Приводятся результаты анализа основных проблем реализации потоковых архитектур, к которым относятся: высокая избыточность памяти данных; построение и обработка неизменяемых структур данных; балансировка вычислительной нагрузки; низкая эффективность конвейеризации последовательных вычислений; построение высокопроизводительной микроархитектуры; обработка константных данных; организация циклов. Также приводится классификация потоковых архитектур и исследуемой МПРА.

В разделе 1.3 автор приводит ключевые принципы новой рекуррентно-потоковой модели вычислений и МПРА (со ссылками на труды ее создателей), которая является одной из возможных реализаций данной модели и предназначена для организации параллельных вычислений ограниченной размерности. Автор исследует прототип МПРА, специфицированный в виде двухуровневой гибридной архитектуры рекуррентного обработчика сигналов (ГАРОС) и осуществляет анализ

его функциональных возможностей, применительно к обозначенным в разделе 1.2 проблемам. В разделе также показано, что ГАРОС в той или иной степени реализует решения для всех рассмотренных проблем, но не обладает высокопроизводительной микроархитектурой.

В разделах 1.4-1.5 рассматриваются разработки российских исследователей в области потоковых архитектур, а именно, параллельная потоковая вычислительная система «Буран» и «Мультиклет». Осуществляется их сравнение с ГАРОС, по результатам которого делается вывод, что наиболее близким аналогом является «Мультиклет». Однако данные вычислительные системы предназначены для решения различных классов задач и различных условий применения.

В разделе 1.6 автор формулирует основные выводы по результатам анализа прототипа рекуррентной архитектуры и конкретизирует постановку задач диссертационного исследования.

В главе 2 приводятся основные результаты разработки прототипа потоковой рекуррентной архитектуры.

В разделе 2.1 представлены результаты анализа проблем реализации задач ЦОС в уже существовавшем прототипе МПРА. Автором выделены 9 проблем, касающихся как модели программирования и отдельных функциональных блоков, так и ГАРОС в целом.

Раздел 2.2 посвящен наиболее значимым результатам разработки микроархитектуры вычислительных блоков МПРА. Автор существенно усовершенствовал структуру вычислительного блока, по сравнению с исходной реализацией, представленной в разделе 1.3, что позволило ГАРОС функционировать в двухзадачном режиме. Ключевым результатом является предложенный метод поддержки многозадачности, суть которого заключается в исполнении за 1 такт до двух совместимых инструкций. При этом допускается как одновременное исполнение независимых по данным инструкций, так и последовательное исполнение зависимых по данным инструкций (два режима суперскалярных вычислений).

Для этого Д.В. Хилько разбил множество всех кодов операций системы команд ГАРОС на несколько подмножеств и разработал алгоритмы, которые позволяют: установить совместимость полученных инструкций; разрешить возможные конфликты доступа к ресурсам микроархитектуры; в зависимости от установленного режима суперскалярных вычислений настроить схему вычислительного процесса на текущем такте. Упомянутые алгоритмы размещены в приложении к основному тексту диссертации.

В разделах 2.3-2.6 приводятся различные дополнительные архитектурные и алгоритмические решения, которые позволили решить выявленные в разделе 2.2 проблемы.

Раздел 2.7 посвящен еще одному значимому результату разработки ГАРОС – средствам аппаратной поддержки алгоритма Быстрого преобразования Фурье (БПФ). Автором справедливо указано, что БПФ, описание которого представлено в **разделе 2.7.1**, является ключевым алгоритмом ЦОС, который используется для решения широкого круга задач. Поэтому любой современный ЦСП должен обеспечивать высокую производительность и эффективность использования ресурсов при вычислении БПФ.

В **разделе 2.7.2** представлены результаты анализа возможных путей реализации средств аппаратной поддержки БПФ в современных ЦСП, а также имевшегося в ГАРОС решения.

В **разделе 2.7.3** приводятся результаты разработки средств аппаратной поддержки БПФ, реализация которых затронула большинство функциональных блоков ГАРОС. В основе предложенной реализации лежит специальная четырех цикловая инструкция, которая на каждом шаге исполнения нагружает вычислительное ядро в четырех задачном режиме. Также приводятся сравнительные результаты синтеза в ПЛИС данных средств, из которых видно, что разработанная автором версия требует существенно меньше аппаратных ресурсов и вычисляет типовой 256-точечный БПФ на 12% быстрее имевшегося в ГАРОС решения.

В **главе 3** рассматриваются результаты разработки отдельных элементов методологии программирования и отладки ГАРОС в совокупности методик, алгоритмов, моделей и программных средств.

В **разделе 3.1** представлены результаты доработки модели программирования в соответствии с решениями, предложенными в главе 2.

В **разделе 3.2** приводится концептуальное описание рекуррентно-поточковой методологии программирования, отдельные элементы которой разработаны автором в совокупности: методик, алгоритмов (их блок-схемы или вербальное описание представлены в приложении к основному тексту диссертации) и обобщенной технологии программирования. Данные теоретические результаты являются важными, т.к. позволяют систематизировать процесс разработки специализированного программного обеспечения для ГАРОС (называемого капсулами).

В **разделе 3.3** описываются разработанные инструменты моделирования ГАРОС в составе программной имитационной и аппаратной VHDL-модели.

В **разделе 3.4** Д.В. Хилько приводит описание программного комплекса ПК ПОТОК, который включает в себя следующие подсистемы: имитационного моделирования; аппаратного моделирования; автоматизированного построения граф-капсул (специальная нотация описания вычислительного процесса); автоматизированной верификации и валидации ГАРОС. Разработанные автором программные средства являются приложением к рекуррентно-поточковой

методологии программирования и позволяют частично или полностью автоматизировать ее ключевые этапы.

В главе 4 представлены сведения о тестовых испытаниях ГАРОС как на уровне моделей, так и на уровне синтезированного в ПЛИС макетного образца.

В разделе 4.1 дается описание демонстрационной задачи распознавания изолированных слов (РИС). Данная задача имеет эталонное решение на языке С для микроконтроллера dsPIC30F компании Microchip, а ее тестовые данные используются в качестве основы для верификации и валидации ГАРОС.

В разделе 4.2 автор демонстрирует разработанные методики программирования для реализации ключевых алгоритмов задачи РИС в виде капсул.

В разделе 4.3 приводятся результаты испытаний ГАРОС на уровне моделей. На рисунке 1.6 автор указал, что операционный уровень ГАРОС включает 4 вычислительных ядра. Результаты модельных испытаний сравниваются с эталонной одноядерной реализацией по количеству требуемых циклов вычислений. Средний коэффициент ускорения составил $\sim 4,5$, что свидетельствует об эффективности предложенных автором в главе 2 решений.

В разделе 4.4 представлены результаты синтеза на основе верифицированной средствами ПК ПОТОК VHDL-модели ПЛИС прототипа архитектуры, а также результаты его натурных испытаний на реальных произнесениях 100 слов из библиотеки распознавания. Автором представлены данные испытаний для двух версий программы управляющего уровня, реализующих как синхронную, так и асинхронную дисциплину взаимодействия с операционным уровнем. Результаты для синхронной дисциплины оказались ожидаемо низкими, а для асинхронной — коэффициент ускорения составил $\sim 3,6$. Автор сделал предположение о причинах расхождения с модельными результатами и сформулировал дальнейшие направления исследований для их устранения.

В разделе 4.5 Д.В. Хилько привел описание набора эталонных алгоритмических ядер, используемых для измерения производительности современных ЦСП, а также предоставил модельные оценки производительности ГАРОС на данном наборе в сравнении с ЦСП серии TMS C55x компании Texas Instruments. Подробности реализации капсул каждого из алгоритмических ядер описаны в приложении к основному тексту диссертации. Согласно полученным оценкам, ГАРОС имеет сопоставимый с C55x уровень производительности, что свидетельствует об эффективности предложенных автором диссертации решений.

В Заключение сформулированы основные научный и практические результаты, представленные в диссертации, а также предложены направления дальнейших исследований для развития МПРА и ее прототипа ГАРОС.

Кроме описания методов, алгоритмов, методик программирования и результатов исследований, в приложении к основному тексту диссертационной

работы содержится большой объем справочной информации в виде: блок-схем алгоритмов, таблиц, вербального описания алгоритмов и деталей реализации алгоритмических ядер.

С учетом изложенного выше можно считать, что **научно-техническая новизна диссертационной работы** состоит в следующем:

1. Предложена микроархитектура вычислительного ядра ГАРОС, максимальная степень многозадачности которой равна четырем.
2. Разработаны методы и алгоритмы организации двух режимов суперскалярных вычислений, которые позволяют нагружать вычислительное ядро ГАРОС в двух- (при определенных конфигурациях трех-) задачном режиме.
3. Предложены архитектура средств аппаратной поддержки БПФ, а также методы и алгоритмы реализации данного режима функционирования, при котором задействуется специальная память отсчетов и поворотных коэффициентов, а вычислительное ядро исполняет специальную четырех цикловую инструкцию в четырех задачном режиме.
4. Разработаны концептуальные основы рекуррентно-поточковой методологии программирования в виде совокупности: методик, алгоритмов, моделей и программных средств, которые позволили автору успешно решить демонстрационную задачу, а также осуществить верификацию и валидацию ГАРОС.
5. Синтезирован ПЛИС прототип ГАРОС.

Достоверность результатов и выводов, представленных в диссертации, подтверждается следующими факторами:

- Разработанные автором программная имитационная и аппаратная модели прошли процесс верификации на предмет соответствия спецификации средствами ПК ПОТОК.
- Демонстрационная задача РИС имеет эталонное решение, поэтому ее тестовые данные корректны и могут быть использованы для валидации.
- Разработанные автором программная имитационная и аппаратная модели прошли процесс валидации на задаче РИС средствами ПК ПОТОК.
- Результаты запуска ПЛИС прототипа согласуются бит-экзектным образом с результатами запуска моделей и эталонными данными задачи РИС.
- Эти результаты неоднократно докладывались на российских и международных научных форумах. Они опубликованы в российских и зарубежных рецензируемых изданиях, в числе которых 24 из списка ВАК РФ, и получили положительный отклик авторитетных специалистов в области исследования диссертационной работы.

Значимость для науки и практики полученных результатов

Теоретическая значимость результатов диссертационной работы для развития предметной области заключается в разработке: набора методов и алгоритмов организации рекуррентного вычислительного процесса, которые позволили создать прототип ГАРОС для решения задач ЦОС реального времени; и разработке элементов рекуррентно-поточковой методологии программирования, которые позволили организовать процесс программирования и отладки этого прототипа.

Практическая значимость работы заключается в создании работоспособного макетного образца на ПЛИС на основе принципиально новой рекуррентной архитектуры, не имеющей аналогов в мире, который способен решать демонстрационную задачу распознавания слов и другие типовые задачи ЦОС на уровне современных высокопроизводительных ЦСП в случае его реализации в виде полноценного процессора.

Рекомендации по использованию результатов и выводов диссертации

Целесообразно продолжить теоретическую и практическую работы по развитию средств программирования и рекуррентной архитектуры в целом в соответствии с направлениями, предложенными Д.В. Хилько в Заключении к диссертационной работе.

Результаты исследований, представленные в диссертации, могут заинтересовать ученых, занимающихся вопросами разработки архитектур вычислительных систем в качестве инновационного подхода к организации вычислительного процесса, управляемого потоком данных. Также результаты могут использоваться на практике компаниями, которые занимаются разработкой ЦСП на основе отечественных технологических решений.

Соответствие диссертации и автореферата требованиям «Положения о порядке присуждения ученых степеней»

Диссертация Д.В. Хилько обладает внутренним единством. Автореферат диссертации адекватно отражает ее содержание. Диссертация и автореферат удовлетворяют требованиям «Положения о присуждении ученых степеней».

По тексту диссертации есть следующие замечания

1. В работе рассматривается ряд российских исследований и разработок в области потоковых архитектур и приводится их сравнительный анализ с рекуррентной архитектурой. Однако, полнота сравнительного анализа является неполной. В частности, не затрагивается опыт ФГУ ФНЦ НИИСИ РАН в области разработок микропроцессоров для цифровой обработки сигналов, таких как 1890BM7Я.
2. В таблице 4.1 раздела 4.3 приводятся размеры капсул алгоритмов распознавания и сравнение с эталоном Microchip по скорости выполнения.

Однако, отсутствует сопоставление размера программных ядер эталона и предлагаемой архитектуры РИС. Возможно ли, в принципе, сравнение размера исполнительного кода (ядер преобразований) для традиционных вычислительных средств и исследуемых архитектур?

Отмеченные замечания не меняют общего положительного впечатления о диссертационной работе.

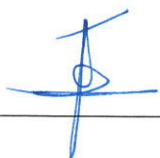
Заключение

Диссертация Д.В. Хилько обладает внутренним единством, представляет завершённые результаты на актуальную тему и является научно-квалификационной работой, в которой на основании выполненных автором исследований разработаны методы, алгоритмы, методики программирования, вносящие значительный вклад в разработку вычислительных систем и их элементов. Автореферат диссертации адекватно отражает ее содержание.

Диссертационная работа соответствует паспорту научной специальности 2.3.2 – «Вычислительные системы и их элементы», а именно: п. 5. Разработка научных методов и алгоритмов организации арифметической, логической, символьной и специальной обработки данных, хранения и ввода-вывода информации; п. 7. Разработка научных методов и алгоритмов организации параллельной и распределенной обработки информации, многопроцессорных, многомашинных и специальных вычислительных систем. Автор диссертации заслуживает присуждения ему ученой степени кандидата технических наук.

Отзыв заслушан и одобрен на заседании Отделения разработки вычислительных систем ФГУ ФНЦ НИИСИ РАН (протокол №1 от 09.11.2023). Присутствовало – 11 чел. Результаты голосования: «за» – 11 чел., «против» – 0 чел., «воздержалось» – 0 чел.

Отзыв составил
заместитель заведующего Отделением
разработки вычислительных систем,
старший научный сотрудник,
кандидат технических наук


Петров К.А.
«09» ноября 2023 г.

*Подпись руки Петрова К.А. за верю
инспектор отдела кадров Телмисова И.А.*



Петров Константин Александрович - кандидат технических наук (05.13.05 – Элементы и устройства вычислительной техники и систем управления); старший научный сотрудник, заместитель заведующего Отделением разработки вычислительных систем Научно-исследовательского института системных исследований РАН.

Сведения об организации

Федеральное государственное учреждение "Федеральный научный центр научно-исследовательский институт системных исследований российской академии наук"
117218, Российская Федерация, Москва, Нахимовский просп., 36, к.1; +7 495 718 21 10; электронная почта: niisi@niisi.msk.ru; адрес в сети Интернет: <https://www.niisi.ru/>