УДК 004.2

Массово-параллельная вычислительная архитектура на основе конфигурируемых процессорных кластеров

Ходченков М.С. Федеральное государственное автономное образовательное учреждение высшего образования “Национальный исследовательский университет ИТМО”
Научный руководитель – доцент, к.т.н. Антонов А.А. Федеральное государственное автономное образовательное учреждение высшего образования “Национальный исследовательский университет ИТМО”

Введение. Современные вычислительные нагрузки такие как нейронные сети, обработка сигналов и распознавание образов, предъявляют повышенные требования к процессорам в части производительности и вычислительной эффективности. Массово-параллельные архитектуры на основе процессорных кластеров удовлетворяют данным задачам, однако, остаются открытыми вопросы планирования выполнения нагрузки, структуры процессорных элементов и организации коммуникационной инфраструктуры. Основные исследования направлены на разработку SIMD и MIMD архитектур, содержащих легковесные RISC-V ядра, реализующие стандартные ISA. Между тем, во многих прикладных задачах существует потребность в гибкой адаптации SIMD/MIMD конфигураций. В описанных в литературе проектах (например, ParaNut [1]) представлены архитектуры систем, в которых управляющий тракт процессоров является отключаемым, а тракт данных подключается к управляющему тракту другого процессора. Однако, управляющий тракт в отключенном режиме занимает аппаратные ресурсы. Данный проект нацелен на создание массово-параллельных архитектур, которые, в зависимости от задачи, могут работать в различных SIMD/MIMD режимах и обеспечивают повторное использование ресурсов для управляющих и обрабатывающих трактов за счет различных механизмов конфигурирования.

Основная часть.

Основная идея заключается в исследовании конфигурируемого кластера интегрированного в массово-параллельную кластерную архитектуру. В кластере содержится несколько трактов управления и данных, общая память, интерфейсы настройки режима работы и сопряжения с сетью-на-кристалле (Network-on-Chip, NoC). В кластере предусмотрена возможность изменять конфигурацию и соединять соответствующие тракты управления и данных для более эффективного использования вычислительных ресурсов в зависимости от прикладной задачи. Особенностью данного подхода является то, что имеется возможность динамически изменять структуру кластера без использования технологий динамической частичной реконфигурации (DPR) Xilinx которая применяется в представленных в литературе проектах [2], тем самым такой подход является более универсальным.

Для обеспечения энергоэффективности предусматривается отключение от тактового генератора и/или от питания тех трактов данных и управления, которые не используются в данный момент. В случае, когда в кластере собран один векторный процессор, то есть, используется только один тракт управления, часть высвобожденной памяти, которая использовалась для хранения команд незадействованных трактов управления, может использоваться для размещения данных. Данный подход актуален для задач, сильно зависимых от данных, так как обращение к участкам памяти вне кластера через NoC имеет задержку, существенно большую чем для локальной памяти. Недостатком данного подхода является неэффективность с точки зрения площади. Но в сложных задачах, задействующих в разный момент времени разные вычислительные ресурсы, данных подход имеет потенциал. В работе представлена оценка актуальности данного подхода.

Для организации связи между кластерами в представленном проекте задействуется NoC Noplite[3], так как данная сеть имеет хорошую пропускную способность, требует мало аппаратных ресурсов, а также хорошо адаптирована для FPGA [4]. В отличие от представленных в литературе аналогов, в качестве основы процессорного кластера используется открытое промышленное ядро SCR1 (Syntacore) [5]. Также, в отличие от проекта ParaNut, написанного на SystemC, в качестве языка проектирования используется Verilog HDL, хорошо поддержанный как открытыми, так и коммерческими инструментами синтеза, что, как ожидается, позволит улучшить площадь, частоту и энергопотребление разрабатываемой системы.

Выводы. Проведен анализ существующих концепций гетерогенных платформ, на основе массово-параллельных вычислительных архитектур. Предложена архитектура на основе вычислительных кластеров с динамически настраиваемыми SIMD/MIMD режимами. Данная архитектура эффективна для решения сложных параллельных задач, таких как распознавание образов, извлечение признаков в сигналах и нейронные сети.

Список использованных источников:

1. Bahle A., Kiefer G., The ParaNut/RISC-V Processor - An Open, Parallel, and Highly Scalable Processor Architecture for FPGA-based Systems // Embedded Word 2020 – 2020.

2. Kamaleldin A., Göhringer D., AGILER: An Adaptive Heterogeneous Tile-Based Many-Core Architecture for RISC-V Processors // IEEE Access – 2022. – Vol. 10, P.43895-43913.

3. Kapre N., Gray J., Hoplite: building austere overlay NoCs for FPGAs // 25th Int'l Conf. on Field-Programmable Logic and Applications – 2015.

4. Gray J., GRVI phalanx: A massively parallel RISC-V FPGA accelerator accelerator // FCCM – 2016. – P.17–20.

5. Scr1 URL: https://github.com/syntacore/scr1 (дата обращения: 27.02.2023).

Ходченков М.С. (автор) Подпись

Антонов А.А. (научный руководитель) Подпись