

AGILER: Adaptacyjna heterogeniczna architektura wielordzeniowa oparta na kafelkach dla procesorów RISC-V

By kama3 kama3
cze 26, 2025

W tej pracy zaproponowano AGILER — nowatorską adaptacyjną architekturę wielordzeniową opartą na kafelkach dla heterogenicznych procesorów RISC-V odpowiednich dla platform FPGA. Proponowana architektura charakteryzuje się wysokim stopniem skalowalności i regularności projektu przy użyciu heterogenicznych PE RISC-V dla wielordzeniowych/jednordzeniowych układów obliczeniowych.

AGILER opiera się na modułowym i konfigurowalnym zestawie heterogenicznych układów obliczeniowych połączonych za pomocą skalowalnej architektury NoC.

Każdy układ obliczeniowy obsługuje systemy pamięci współdzielonej i lokalnej scratchpad. Ponadto proponowana architektura obsługuje rekonfiguracje w celu zmiany liczby i typów układów obliczeniowych dla kilku konfiguracji wielordzeniowych. W celu rekonfiguracji w czasie wykonywania zaimplementowano menedżera rekonfiguracji o dużej prędkości. Proponowana architektura ma na celu ułatwienie rozwoju i realizacji heterogenicznych architektur wielordzeniowych opartych na RISC-V poprzez skrócenie czasu projektowania i jednorazowych kosztów inżynierskich.

Architektura AGILER jest oceniana na podstawie wykorzystania zasobów sprzętowych i skalowalności wydajności obliczeniowej poprzez kilka konfiguracji wielordzeniowych i testów porównawczych. Wyniki pokazują wysoki stopień skalowalności obliczeniowej przy użyciu skalowalnej liczby heterogenicznych kafelków obliczeniowych. Jeśli chodzi o przyszłe prace, planowane jest zapewnienie bardziej heterogenicznych kafelków obliczeniowych z rozszerzeniami RISC-V ISA dla operacji zmiennoprzecinkowych (np. RV32F, RV64D) i instrukcji wektorowych w celu obsługi większej liczby wymagań aplikacji. Ponadto zostanie zaimplementowany pomocniczy kafelek pamięci w celu obsługi dostępu do pamięci poza układem bezpośrednio przez NoC, aby uzyskać do niego bezpośredni dostęp ze wszystkich kafelków obliczeniowych bez przechodzenia przez główny kafelek przetwarzania w celu uzyskania dostępu do pamięci zewnętrznej.

Dlatego Blackparrot [26] został zaproponowany jako pierwsza heterogeniczna i obsługująca Linuksa platforma wielordzeniowa typu open source. Platforma charakteryzuje się spójną architekturą opartą na NoC jako międzykafelkową strukturą komunikacyjną, trzy NoC są używane do wewnętrznych danych, pamięci współdzielonej i zewnętrznych urządzeń peryferyjnych. Trzy heterogeniczne typy kafelków obliczeniowych są dostarczane do obliczeń ogólnego przeznaczenia (GPP) opartych na RV64, koherentnych i strumieniowych niestandardowych akceleratorach sprzętowych. Platforma obsługuje kilka modeli programowania w oparciu o wymagania aplikacji. Ponadto wiele platform wielordzeniowych [27], [28] obsługuje wiele ISA RISC-V w celu uzyskania większych zasobów obliczeniowych i wydajności w celu obsługi wielu wymagań obciążeń roboczych. Manticore [27] został zaproponowany do obciążeń obliczeniowych o wysokiej wydajności, składa się z setek 32-/64-bitowych rdzeni RISC-V zgrupowanych w wiele kafelków. Każdy kafelek zawiera osiem rdzeni RV32 z pamięcią współdzieloną L1 i współdzieloną pamięcią podręczną instrukcji. Wiele kafelków tworzy klaster, każdy klaster zawiera czterordzeniowy RV64 do sterowania i zarządzania. Kaskadowe poprzeczki są używane do komunikacji międzykafelkowej

i między klastrami. Z drugiej strony, struktura ESP [28] została zaproponowana do realizacji różnych konfiguracji wielordzeniowych dla platform wbudowanych. ESP charakteryzuje się skalowalną, spójną architekturą opartą na NoC, składającą się z wielu heterogenicznych kafelków obliczeniowych procesorów opartych na RISC-V i niestandardowych akceleratorów sprzętowych. Każdy kafelek obliczeniowy hostuje pojedynczy element obliczeniowy. ISA RV64 jest obsługiwany wewnątrz kafelka ogólnego przeznaczenia, a także kafelek obliczeniowy dla akceleratora opartego na HLS. Podobnie jak platforma Tapasco [23], ESP automatycznie generuje również wymaganą logikę opakowaniową i interfejsy w celu płynnej integracji niestandardowych akceleratorów sprzętowych z kafelkiem akceleratora. Ponadto kilka modeli programowania jest obsługiwanych przez goły metal lub system operacyjny Linux. W związku z tym nasza proponowana architektura AGILER różni się od wyżej wymienionych prac, jak pokazano w Tabeli 1, zapewniając więcej przestrzeni projektowej do tworzenia kilku heterogenicznych konfiguracji wielordzeniowych opartych na kafelkach. Proponowana architektura zapewnia konfigurowalny zestaw kafelków obliczeniowych, który może obsługiwać dwa ISA RISC-V (RV32, RV64). Ponadto kafelki obliczeniowe można skonfigurować tak, aby obsługiwały architektury jednordzeniowe i wielordzeniowe z elastycznością obsługi kilku hierarchii pamięci. Każdy kafelek obliczeniowy ma prywatną przestrzeń adresową, która umożliwia komunikację między wszystkimi PE i współdzielonymi urządzeniami peryferyjnymi kafelków poprzez współdzieloną pamięć danych za pośrednictwem połączenia AXI. Do komunikacji między kafelkami używana jest synchroniczna skalowalna architektura NoC ARTNoC [29] z modelem komunikacji opartym na wiadomościach w celu zarządzania transferem danych między kafelkami obliczeniowymi. Ponadto proponowana architektura jest adaptowalna w czasie wykonywania, co zapewnia elastyczność zmiany typów i liczby aktywnych kafelków obliczeniowych w czasie wykonywania. Dzięki wewnętrznemu menedżerowi rekonfiguracji w czasie rzeczywistym, który zarządza i kontroluje proces DPR na układach FPGA.

Dlatego w tej pracy proponujemy AGILER jako nowatorską samoadaptacyjną architekturę wielordzeniową opartą na kafelkach dla heterogenicznych konfiguracji wielordzeniowych opartych na RISC-V, ukierunkowanych na układy FPGA bez potrzeby zewnętrznego urządzenia /menedżera do rekonfiguracji i transferu danych. Proponowana architektura spełnia wymagania samoadaptowalności i modułowości dla wysoce skalowalnych heterogenicznych systemów wielordzeniowych opartych na RISC-V na układach FPGA. Nasz wkład w tę pracę polega na następujących kwestiach: Implementacja modułowych heterogenicznych elementów przetwarzania opartych na RISC-V obsługujących RV32/RV64-ISA ze ściśle sprzężoną parametryzowaną pamięcią podręczną i zunifikowanymi interfejsami AXI do wykorzystania jako modułowe PE w ramach płytek obliczeniowych. Dostarczamy heterogeniczny zestaw płytek obliczeniowych dla wielu ISA RISC-V. Płytki obliczeniowe charakteryzują się architekturą wielo-/jednordzeniową wykorzystującą heterogeniczne PE oparte na RISC-V dla ISA 32-/64-bitowych. Ponadto, główny/podstawowy kafelek przetwarzania jest rozwijany przy użyciu wielu procesorów RISC-V 64-bitowych jako głównego i stałego kafelka obliczeniowego dla proponowanego systemu wielordzeniowego. Rozwój wysoce skalowalnej architektury wielordzeniowej opartej na kafelkach przy użyciu ogólnej architektury NoC jako głównego medium połączeń międzykafelkowych. Wymagany model komunikacji i metoda programowania są dostarczane w celu obsługi równoległego wykonywania zadań i interakcji w heterogenicznych kafelkach obliczeniowych. Menedżer rekonfiguracji w czasie wykonywania jest rozwijany dla wielordzeniowych kafelków obliczeniowych RISC-V w celu obsługi funkcji adaptacji dla proponowanej architektury poprzez

samodzielne zarządzanie procesem dynamicznej częściowej rekonfiguracji (DPR) z głównego kafelka przetwarzania w czasie wykonywania.

Streszczenie: Architektury wielordzeniowe oparte na kafelkach są szeroko stosowane w nowoczesnych projektach systemów na chipie w celu osiągnięcia skalowalnej wydajności obliczeniowej przy odpowiedniej efektywności energetycznej. Heterogeniczność jest kluczowym elementem zwiększającym wydajność obliczeniową i utrzymującym zużycie energii poniżej pewnych limitów dla kilku domen aplikacji. Jednak stały wzrost wykorzystania wielu niestandardowych heterogenicznych kafelków prowadzi do wzrostu kosztów projektowania i integracji przy ograniczonej możliwości ponownego wykorzystania kafelków. Ostatnie rozpowszechnienie oprogramowania ISA typu open source RISC-V zapewnia potencjał do opracowywania modułowych jednostek obliczeniowych, które mogą być używane w wielu domenach aplikacji przy dużej redukcji jednorazowych kosztów inżynierskich. Motywacją tej pracy jest wprowadzenie modułowości projektowania i funkcji adaptacji dla heterogenicznych architektur wielordzeniowych opartych na kafelkach poprzez zwiększenie ich elastyczności w celu realizacji różnych konfiguracji wielordzeniowych przy krótszym czasie projektowania i niższych kosztach. W tej pracy zaproponowano AGILER jako adaptacyjną architekturę wielordzeniową opartą na kafelkach dla heterogenicznych procesorów opartych na RISC-V. Proponowana architektura składa się z modułowych i adaptowalnych heterogenicznych wielordzeniowych /jednordzeniowych płytek obliczeniowych, które obsługują 32-/64-bitowe ISA RISC-V z różnymi hierarchiami pamięci. Komunikacja między płytkami jest rozwijana na podstawie skalowalnej architektury sieciowej na chipie, aby osiągnąć wysoki stopień skalowalności systemu. AGILER obsługuje adaptację w czasie wykonywania za pomocą niestandardowego wewnętrznego menedżera rekonfiguracji do dynamicznej i częściowej rekonfiguracji na układach FPGA Xilinx. Wyniki oceny pokazują, że proponowana architektura charakteryzuje się skalowalną wydajnością obliczeniową do 685 MOPS dla płytek 8×328\times 32-bitowych i 316 MOPS dla płytek 8×648\times 64-bitowych ze skalowalną przepustowością pamięci do 7,4 GB/s. AGILER jest oceniany na układzie FPGA Xilinx Virtex Ultrascale+ z maksymalnym czasem rekonfiguracji 38,1 ms dla pojedynczej płytki obliczeniowej.

[Read more](#)