

REZUMATUL TEZEI DE DOCTORAT

“Integrarea în medii de calcul distribuit a acceleratoarelor bazate pe circuite FPGA”

Autor: inf. Laurențiu DUMITRU, tel. 0727.737.660, email: dumitru.laurentiu@cmsystems.ro

Conducător de doctorat: Prof. univ. dr. ing. Ciprian RĂCUCIU

Teza de doctorat se încadrează în domeniul “Inginerie electronică, telecomunicații și tehnologii informaționale” prin subiectul abordat care tratează modalitatea de utilizare transparentă a acceleratoarelor FPGA în medii de calcul distribuit de mare putere.

Principalul obiectiv al tezei a fost dezvoltarea unei platforme care să asigure gestiunea automată a acceleratoarelor FPGA în medii de calcul distribuit, stabilirea unei structuri fixe de interconectare a modulelor de accelerare și implementarea unor interfețe API bine definite care să fie utilizate de către dezvoltatorii aplicațiilor finale.

Lucrarea este structurată pe opt capitole. Primul capitolul, cel introductiv, expune contextul tehnologic, contribuțiile autorului și obiectivul final al lucrării.

Capitolul 2 analizează beneficiile aduse de accelerarea hardware în diverse domenii, pe baza literaturii de specialitate. Din lucrările analizate se observă îmbunătățirea timpilor de execuție pentru diverși algoritmi și eficiența energetică a soluțiilor FPGA. Nu există elemente tehnice comune de acces la resursele FPGA și a unor canale de transfer ale datelor specifice între soluții. Prin introducerea unui substrat de control și acces se poate reduce complexitatea soluțiilor bazate pe acceleratoare FPGA, facilitând astfel adoptarea accelerării hardware la scară largă.

În Capitolul 3 sunt evaluate soluții tehnice care tratează abstractizarea utilizării dispozitivelor de accelerare FPGA, atât din mediul open-source cât și din mediul comercial. Sunt prezentate avantajele și dezavantajele fiecărei variante atât în comparație cu celelalte cât și cu platforma care face obiectul lucrării. De asemenea, sunt prezentate și câteva idei de îmbunătățire a soluțiilor analizate.

Capitolul 4 prezintă arhitectura generală a platformei, canalul de comunicație între placa de accelerare și sistemul gazdă, canalele de comunicație din interiorul plăcii de accelerare și elementele software necesare la nivel kernel al sistemului de operare al gazdei. Sunt detaliate modulele IP specifice, metode de comunicație și proceduri de transfer a memoriei.

În Capitolul 5 este descris ce se întâmplă în interiorul plăcii de accelerare atunci când aceasta se află în uz. Este detaliată structura modulelor de accelerare, a elementelor încapsulate și procedurile de intrare/ieșire pentru datele de lucru. Sunt prezentate interfețele standard care facilitează reutilizarea modulelor de accelerare în proiecte diferite. Sunt expuse funcțiile API high-level necesare programatorilor software și seturile de semnale necesare programatorilor HDL. De asemenea, în acest capitol sunt expuse metodele de reconfigurare parțială și totală, elementele necesare gestiunii plăcii FPGA și fluxurile de lucru pentru toate acțiunile plăcii de accelerare.

Capitolul 6 se concentrează pe detalierea implementării platformei de accelerare într-un mediu de calcul distribuit. Sunt descrise modurile de operare ale platformei și procedeele de integrare transparentă într-un cluster ale cărui noduri de procesare sunt echipate cu plăci de accelerare FPGA. Este prezentat fluxul complet de lucru, plecând de la momentul lansării unui job de către un utilizator, trecând prin etapele specifice de încărcare, execuție sau reprogramare și terminând cu transferul datelor de ieșire înapoi spre aplicația care a beneficiat de accelerare.

Capitolul 7 prezintă două studii de caz. Primul este efectuat într-un mediu comercial iar al doilea într-un mediu de cercetare. Ambele demonstrează utilitatea platformei și beneficiile aduse atât din punct de vedere tehnic cât și din punct de vedere administrativ/financiar.

Concluziile și direcțiile de dezvoltare sunt prezentate în Capitolul 8. Este expusă o sumarizare a principalelor beneficii aduse de plăcile de accelerare FPGA și de gestionarea centralizată acestora. Sunt prezentate și eventuale direcții viitoare de cercetare care au în vedere extinderea unor capabilități ale platformei, capabilități care nu sunt specifice mediilor de calcul distribuit.