

Posudek oponenta diplomové práce

Student: Vosyka Pavel, Bc.
Téma: Ukázky hardwarové akcelerace na přípravku Pynq Z2 (id 25156)
Oponent: Kekely Lukáš, Ing., Ph.D., UPSY FIT VUT

- 1. Náročnost zadání** **průměrně obtížné zadání**
Práce byla zaměřena na návrh a implementaci několik ukázkových aplikací využívajících FPGA programovatelnou logiku na platformě PYNQ-Z2 osazené čipem Zynq. Hlavním cílem bylo ukázat výhody paralelního zpracování dat v FPGA pro zrychlení zvolených úloh.
- 2. Splnění požadavků zadání** **zadání splněno s drobnými výhradami**
Student zadání v podstatě splnil. Vytvořil tři přímočaré příklady akcelerace zjednodušených úloh z oblasti zpracování textu, úpravy videa a klasifikace rozhodovacím stromem. Výhrady ke kvalitě a způsobu jejich realizace jsou podrobněji rozebrány v sekci 7 posudku. Z pohledu splnění všech náležitostí zadání je ale problematický jeho bod 5, část se simulacemi. V textu práce je sice uvedeno, že implementace byly otestovány v simulacích a zdrojové kódy obsahují pro to potřebné soubory. Zadání ale vyžaduje také sepsání návodu, jak simulace akceleratorů spouštět a obsluhovat. V odevzdaném návodu je popsáno jen testování na přípravku PYNQ-Z2, simulace nikoliv.
- 3. Rozsah technické zprávy** **splňuje pouze minimální požadavky**
Délka textové části práce je na spodní hranici požadavek pro diplomové práce. V textu jsou navíc podle mně uvedeny některé nepotřebné informace a jiné zajímavé naopak chybí. Například chybí kapitola s popisem implementace a kapitola s výsledky mohla být podrobnější. Nakonec použití místy zcela zbytečných nebo zbytečně velkých obrázků působí jako snaha dohnat minimální rozsah bez přidání obsahu.
- 4. Prezentací úroveň předložené práce** **50 b. (E)**
Text práce působí nevyváženým dojmem. Zhruba první polovinu rozsahu tvoří teoretický rozbor, který je skoro celý věnovaný platformě Zynq resp. Pynq a nástrojům pro jejich programování z obecného a povrchního pohledu. Specifické oblasti a znalosti pro vybrané demonstrační úlohy jsou pak jen velice stručně rozebrány. Pro každou úlohu je to v podstatě jedna strana čistého textu (plus nějaké obrázky) bez referencí na zdroje. Druhou polovinu tvoří popis návrhu a stručné shrnutí výsledků práce. Zde zcela chybí kapitola s popisem detailů implementace, ty jsou místy vloženy přímo do návrhu. Z popsaného návrhu navíc není zřejmé proč byly zvoleny některé klíčové postupy. Uvedeny jsou jen prvoplánové myšlenky autora bez odkazů na (chybějící) teoretický rozbor a běžně zaužívané postupy.
- 5. Formální úprava technické zprávy** **50 b. (E)**
Celkově působí text práce narychlo dodělaným a nevytříštěným dojmem. Obrázky jsou vesměs všechny přímo přebírané z citovaných zdrojů. Často jsou navíc do textu vloženy nevhodným způsobem, kdy jsou kusy stránek (někdy až polovina) kvůli tomu ponechány prázdné. Postřehl jsem též chyby v odkazech na číslování obrázků a kapitol. Samotný text je místy zmatený a jakoby napsán bez zpětné kontroly. Jazykovou stránku a českou gramatiku však neumím plnohodnotně posoudit.
- 6. Práce s literaturou** **50 b. (E)**
Práce odkazuje jen velice malé množství literatury. Všechny prameny jsou v elektronické podobě a většina z nich jsou navíc jen příručky od společnosti Xilinx zaměřené na architekturu Zynq a Pynq platform. Zcela chybí reference na relevantní zdroje z pohledu problematiky řešených úloh a vhodného přepojení akceleratorů s procesorem. V textu jsou navíc prameny nejčastěji odkazovány jen jako zdroj přebíraných obrázků.
- 7. Realizační výstup** **50 b. (E)**
Výstupem práce jsou implementace tří zvolených úloh pro platformu PYNQ-Z2. Pro každou úlohu je implementován akcelerator pro FPGA programovatelnou logiku, jeho softwarové řízení v jazyku Python a referenční implementace v C++ pro procesor.

Všechny tři implementace akcelerace používají podivnou a nestandardní metodu komunikace s procesorem skrz komponentu GPIO kontroléru, která má originálně sloužit jen pro pomalé ovládání externích pinů platformy. Zde je použita pro přenos dat po AXI sběrnici z procesoru na adresování interních registrů akceleratorů. Výstupní piny GPIO kontroléru jsou tedy použity jako interní signály pro adresovou a datovou sběrnici realizující přenos dat do/z akceleratoru. Pro takovýto přenos dat skrz AXI sběrnici existují efektivnější standardní komponenty dostupné přímo v nástroji Vivado. Pro pomalé (konfigurační) přenosy se dá automatizovaně v HLS nebo HDL

vygenerovat šablona IP komponenty kde je zpracování AXI komunikace pro zápis a čtení registrů již správně implementováno. Pro rychlejší přenosy se používá standardní komponenta DMA modulu volitelně doplněná o instanci interní blokové paměti.

Jednotlivé akcelerátory implementují také velice naivní nebo zjednodušené přístupy k řešení zvolených úloh. Vyhledání v textu je implementováno jen jako sada komparátorů s hledanými slovy vůči posuvnému registru vstupních dat, pracuje se bajt po bajtu. Rozhodovací strom pracuje nad jedinou dimenzí (položkou) vstupních dat. Zpracování obrazu nepracuje nad proudem pixelů v běžném pořadí po řádcích, ale nad dopředu nachystanými 9-okolíma. Všechny aplikace též pracují nevhodným sekvenčním způsobem: naplní paměť vstupů, provede výpočty nad dávkou, vyprázdni paměť výstupů. Měřené výsledky výkonu a zrychlení přitom velice vhodně uvažují jen s čas potřebným pro výpočet. Ignorují režii naplnění a vyprázdnění paměti ze strany procesoru, která bude díky nevhodnému napojení na AXI sběrnici celkem významná.

8. Využitelnost výsledků

Práce je implementačního charakteru. Zamýšleným cílem bylo využití vytvořených ukázkových aplikací jako demonstračních příkladů pro studenty ve výuce na fakultě. Nicméně kvůli použití nestandardních a nevhodných postupů není takovéto využití vhodné. Navíc je také úroveň a kvalita uživatelské dokumentace nedostačující.

9. Otázky k obhajobě

1. Co vedlo na výběr GPIO kontroléru jako vhodného způsobu pro komunikaci mezi procesorem a FPGA akcelerátorem? Proč byl tento přístup upřednostněn před standardními cestami skrz AXI Endpoint šablona nebo DMA modul? V použitém nástroji Vivado jsou obě tyto cesty jednoduše dostupné.
2. Proč byly pro všechny úlohy zvolené implementace zjednodušených verzí problémů?
3. Je autor sám spokojen s kvalitou textové a realizační části své diplomové práce nebo by své výsledky raději doplnil/opravil?

10. Souhrnné hodnocení

50 b. dostatečně (E)

Práce svým rozsahem a kvalitou zodpovídá spíše úrovni bakalářské práce. Student sice zadání v podstatě splnil a vytvořil funkční implementace akcelerátorů zvolených úloh. Využil u toho ale nestandardní a neoptimální postupy, které zamezují využití výsledků práce jako vhodných ukázek pro další studenty. Kvalita textové části práce taktéž není nejlepší. Navrhují proto hodnocení stupněm **E**.

Práci by určitě prospělo dodatečné doplnění a oprava. Nechám na studentovi samotném, hodnocení od jeho vedoucího a/nebo komise zdali práci v této podobě uznají, nebo bude přes léto vylepšena.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 2. června 2022

Kekely Lukáš, Ing., Ph.D.
oponent