

MO401 – Trabalho 1 (Resumo de Artigo)

Título do Artigo: Tradeoffs in Designing Accelerator Architectures for Visual Computing - Aqeel Mahesri, Daniel Johnson, Neal Crago, Sanjay J. Patel - MICRO 2008

Autor do Resumo: Marcos Vinícius Mussel Cirne

RA: 045116

O artigo fala sobre os tradeoffs que existem ao desenvolver uma arquitetura aceleradora (também denominada como xPU) voltada para aplicações na área de Computação Visual. Esse procedimento consiste basicamente em projetar um co-processador (definido de acelerador) que faz com que a CPU acelere a computação de tarefas intensivas. A arquitetura proposta no artigo apresenta um alto grau de paralelismo, com o intuito de melhorar a sua performance, e uma grande demanda de throughput. Para realizar as avaliações de benchmark desta arquitetura, foi utilizada uma ferramenta denominada VISBench, que consiste em uma compilação de diversas aplicações gráficas, tais como renderização em alta qualidade, codificação de vídeos, simulação, visão computacional e processamento de imagens médicas.

A arquitetura básica de uma xPU consiste em um conjunto vasto de pequenos núcleos (cores), agrupados em clusters e conectados a uma grande cache compartilhada. Ela também apresenta um sistema de memória com uma alta largura de banda e um sistema de gerenciamento de threads, além de um núcleo de controle, responsável pela inicialização do programa a ser executado no acelerador. Uma vez que estes recursos são compartilhados por todos os cores presentes nesta arquitetura, ela tem um ganho na sua performance, através de paralelismo. Além disso, outros fatores importantes são a sincronização e a comunicação entre os elementos embutidos nessa arquitetura paralela. No caso das xPU's, objetivou-se uma abordagem próxima do que é normalmente feito em outros tipos de arquitetura, como as multicore, os supercomputadores e as GPU's.

Um outro aspecto fundamental para se considerar nas arquiteturas de xPU's é a avaliação da performance dos seus componentes em relação aos seus custos (definidos pela área do chip utilizado na construção de cada um deles). Para se fazer essa avaliação, é utilizada uma metodologia baseada em simulações, determinando-se o custo total para uma arquitetura com um dado conjunto de características. A ideia é definir uma configuração que tenha a melhor relação performance/área (em relação aos núcleos desenvolvidos) possível, além de maximizar o throughput da arquitetura. Para tal, foram testadas três diferentes configurações de pipeline (1-wide in-order, 2-wide in-order e 2-wide out-of-order), analisando-se as suas respectivas performances em relação às áreas totais ocupadas, bem como os comportamentos quando sujeitas a diversas características da arquitetura, como multithreading, presença de uma unidade que calcula funções trigonométricas, pipelines com alta frequência, etc.

Após essas análises, concluiu-se que uma xPU desenvolvida a partir de núcleos pequenos e simples atinge uma performance bem maior do que um processador de propósito geral. Outros fatores que contribuem para uma melhora no desempenho são o suporte limitado para ILP nos núcleos, um multithreading granulado, uma quantidade suficiente de memória e a largura de banda da cache.