

Título: Single-ISA Heterogeneous Multi-Core Architectures for Multithreaded Workload Performance

Referência: Rakesh Kumar, Dean M. Tullsen, Parthasarathy Ranganathan, Norman P. Jouppi, Keith I. Farkas. Single-ISA Heterogeneous Multi-Core Architectures for Multithreaded Workload Performance; pp. 64-75; ISCA '04

Autor do resumo: Daniel Henricus de Knegt Dutra Nicácio **RA:** 057612

Os processadores atuais precisam atender a dois requisitos: bom desempenho com um alto fluxo de *threads* e boa performance com uma única *thread*. A abordagem mais promissora é a construção de chips com multi-processadores. Sendo assim, o artigo tem como objetivo mostrar que o uso de processadores heterogêneos em um mesmo chip é mais eficiente do que o uso de processadores iguais.

A segunda parte do artigo apresenta duas grandes vantagens do uso de arquiteturas heterogêneas. A primeira delas é a eficiente adaptação para a diversidade de aplicações, ou seja, diferentes programas ou diferentes fases de um mesmo programa exigem diferentes tipos de processadores. A segunda vantagem é o uso eficiente do *die* para um dado paralelismo de *threads*, ou seja, para um certo número de *threads*, a área do *die* pode ser melhor aproveitada utilizando processadores de tamanhos e eficiências diferentes. No final desta parte, é apresentado um gráfico comparativo contendo diferentes combinações de processadores, mostrando a área utilizada por cada uma dessas combinações e seus *throughputs*.

A terceira seção do artigo descreve a metodologia utilizada para realizar as comparações entre as diferentes arquiteturas heterogêneas e homogêneas. Os hardwares utilizados foram os processadores da família *Alpha*: EV5(21164), EV6(21264) e um hipotético EV6+ com suporte *multi-thread*. A avaliação foi feita com o número de *threads* variando de um ao máximo número de contextos dos processadores. A métrica de comparação foi o tempo médio de resposta das aplicações, assim como o tamanho das listas de espera do sistema.

A quarta seção apresenta as vantagens de performance das arquiteturas heterogêneas e mecanismos de designação de tarefas para seus processadores, os quais contribuem para o aumento da eficiência das arquiteturas heterogêneas. O primeiro tópico desta parte é o escalonamento estático para diversidade entre *threads*. Utilizando esta técnica padrão, a arquitetura composta por 3 EV6 e 5 EV5 teve uma melhora média de 26% em relação à arquitetura com 4 EV6 e 23% em relação à arquitetura com 20 EV5. Esta comparação utilizando o alcance de 20 *threads* mostra um significativo ganho com o uso de processadores heterogêneos.

A segunda análise explora a eficiência de arquiteturas heterogêneas em sistemas abertos. Para tal, foi utilizado um *framework* de simulação que modela chegadas randômicas de tarefas com tamanhos também randômicos. Dessa forma, foi medido o tempo médio de resposta, levando em consideração o número de tarefas designadas em um intervalo de tempo. O resultado obtido entre as arquiteturas 4EV6 e a 3EV6 & 5EV5 mostra que a segunda se satura com um *throughput* muito maior do que a primeira. Além disso, o tempo de resposta da arquitetura heterogênea se degrada de forma muito mais suave quando submetida a uma alta carga de tarefas.

O terceiro tópico enfatiza o escalonamento dinâmico para *threads* diversas, detalhando políticas de designação de tarefas. As estratégias, chamadas de '*core sampling strategies*', atribuem tarefas aos processadores de forma que o processador mais robusto nunca fique a toa e o sistema como um todo tenha o melhor desempenho possível. A primeira, *sample-one*, testa cada *thread* em cada diferente processador; a segunda, *sample-avg*, testa cada *thread* várias vezes em cada tipo de processador; e a terceira, *sample-sch*, testa cada forma de escalonamento das tarefas. O resultado da comparação mostra que a terceira técnica é a mais eficiente, aumentando a eficiência em relação à atribuição randômica em até 22%.

Estas estratégias requerem a minimização da sobrecarga das amostragens e uma reação rápida a mudanças do sistema. Com este fim, são sugeridas duas técnicas: variação do intervalo de tempo entre as amostragens e o monitoramento do IPC, este pode ser realizado em uma única *thread*, em todas elas, ou especificando um intervalo geral de IPC. Destas técnicas, a última apresentou o melhor resultado, com melhora de 20% em relação à técnica estática.

Para finalizar esta quarta seção, o artigo relata o uso de processadores com suporte *multi-thread* em arquiteturas heterogêneas. O resultado obtido foi um desempenho ainda maior para situações com grande quantidade de tarefas.

Por fim, o artigo conclui que arquiteturas multi-processadores heterogêneas provêm uma significativa vantagem em relação às arquiteturas homogêneas de mesma área. Além disso, o artigo também conclui que o uso de estratégias de escalonamento em arquiteturas heterogêneas aumentam ainda mais o potencial destas arquiteturas.