

A Mechanistic Performance Model for Superscalar Out-of-Order Processors  
Stijn Eyerman, Lieven Eeckhout, Tejas Karkhanis e James E. Smith

Eyerman, S., Eeckhout, L., Karkhanis, T., and Smith, J. E. **A mechanistic performance model for superscalar out-of-order processors**. ACM Trans. Comput. Syst. 27, 3, Article 3 (May 2009).

Resumo: Hamilton José Brumatto – RA 096389.

Este artigo propõe um modelo mecanicista para estimar a performance de processadores de execução fora de ordem. A figura 1 (apêndice) mostra um processador típico superescalar de execução fora de ordem. Este processador incorpora em sua microarquitetura um conjunto de parâmetros representando a largura de banda, ou largura (Width -  $W$ ) medida em instruções por ciclo (IPC). Na figura estão indicadas diversas larguras: busca ( $F$  - fetch), despacho ( $D$  - dispatch), distribuição ( $I$  - issue), retirada ( $R$  - retire). Na proposta de um processador balanceado, a largura é a mesma em todos estágios e é ditada por  $D$ .

O modelo mecanicista proposto constrói um intervalo de análise que quebra o tempo total de execução em intervalos baseados em eventos de perdas/erros (miss events). A figura 2 (apêndice) mostra um conjunto de eventos associados a estas perdas. No modelo considera-se que em um intervalo de  $N$  instruções, leva-se  $\lceil N/D \rceil$  ciclos para despachar as instruções, se houver uma perda, haverá uma penalidade de tempo  $c$  associada à perda: Perdas de I-cache/I-TLB:  $c_{iL1}$  para I-cache L1,  $c_{L2}$  I-cache L2 e  $c_{ITLB}$  I-TLB; Erro em predição de desvios:  $c_{fe}$  (estágios no pipeline de entrada) +  $c_{dr}$  (descarte de execução do desvio); Perda de cache de dados (L2), o mesmo para D-TLB:  $c_{L2}$ . Estes tempos podem ser sobrepostos na ocorrência de duas perdas (as duas se equivalem), ou serializados (somam-se os tempos). Ineficiência do despacho: como o despacho múltiplo pode encontrar um intervalo de instruções no qual o número não é um múltiplo da largura do despacho, o termo  $\lceil N/D \rceil$  pode ser escrito como:  $N_{total}/D + (D-1/2D).(m_{iL1}+m_{iL2}+m_{br}+m_{dL2}(W))$ , o efeito da borda na função teto pode ser representada como uma perda por despacho ineficiente,  $m$  representa o número de eventos associado a cada tipo de perda.

O tempo total de execução em ciclos é: 
$$N_{total}/D + (D-1/2D).(m_{iL1}+m_{iL2}+m_{br}+m_{dL2}(W)) + m_{iL1}.c_{iL1} + m_{iL2}.c_{L2} + m_{br}.(c_{dr} + c_{fe}) + m_{dL2}(W).c_{L2}$$

A performance geral do processador pode ser analisada e estimada, desta forma, através de intervalos individuais com diferentes tipos e comprimentos. A diferença média entre o resultado obtido no modelo mecanicista e simulações é de 7% em um processador superescalar de execução fora de ordem e largura 4 de despacho.

A partir do modelo foi realizado um estudo do efeito de escala de recursos nos processadores de execução fora de ordem. Para os processadores balanceados foi explorada a configuração do pipeline: profundidade, largura e o relacionamento entre ambos. A profundidade ótima de pipeline  $p^*$  para um dado processador obtida no estudo é aproximadamente proporcional à raiz da largura  $D$ :

$p^*(D).D^{-1/2} \sim cte.$  Em outras palavras, ao aumentar a largura do processador de um fator  $c$ , deve-se diminuir a profundidade de pipeline de um fator  $c^{-1/2}$ .

A parte final estuda o superdimensionamento (ou desbalanceamento) do processador. Normalmente se um estágio do pipeline apresenta uma largura maior, esta não será aproveitada, pois o estágio anterior não consegue entregar em uma performance maior, nem o posterior aproveitar tal performance. Por outro lado, se um estágio é subdimensionado, todos os demais apresentam perda, pois tal estágio passa a ser um gargalo.

Uma exceção ao superdimensionamento ocorre no despacho inicial, o aumento da largura no despacho inicial, em comparação com a largura de distribuição, aumenta a performance pois o processador consegue atingir o próximo evento de perda de forma mais rápida. Os resultados mostram que para um teste de performance houve um aumento de até 9% fazendo uma largura de 6 no despacho em um processador de largura 4 na distribuição e demais estágios.

O trabalho propõe uma continuidade no estudo considerando limitações na largura de banda de memória, efeitos de pré-busca de instrução por hardware e também no estudo de modelos não balanceados de projeto de processadores.

Apêndice

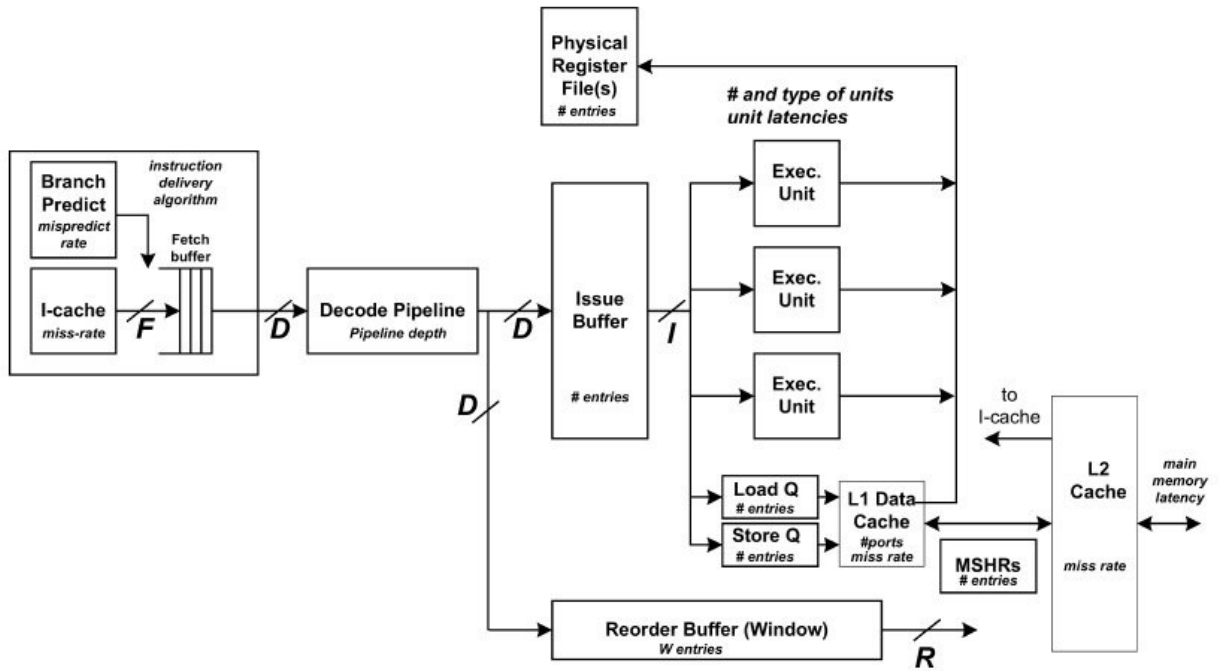


Figura 1 - Processador superescalar de execução fora de ordem parametrizado (extraído do artigo em revisão).

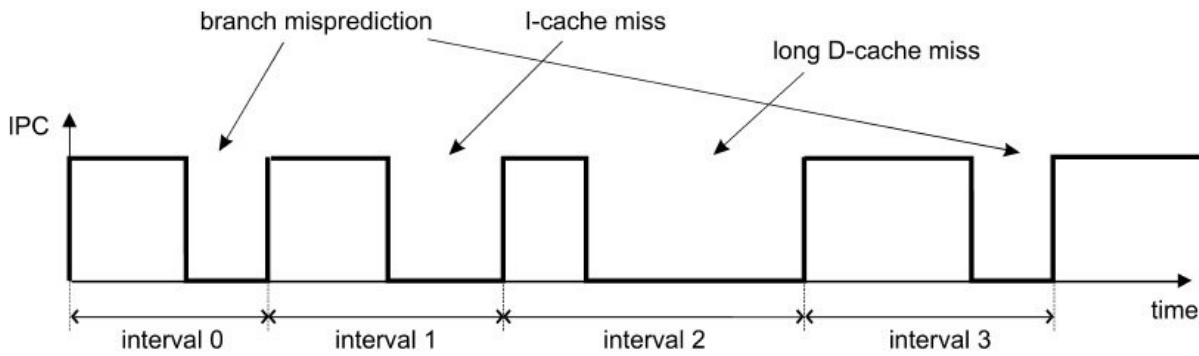


Figura 2 - Espaço de tempo dividido em intervalos de execução com perdas para análise de performance (extraído do artigo em revisão).