A organização e arquitetura do microprocessador IBM Power5

Rodrigo Santos de Souza

Escola de Informática – Universidade Católica de Pelotas (UCPEL)
Rua Félix da Cunha, 412, CEP: 96010-000

rsouza@ucpel.tche.br

Resumo. *Esse trabalho apresenta as características do microprocessador POWER5, explicando brevemente seu esquema geral de funcionamento, paralelismo, previsão de desvio, gerenciamento de memória, bem como as características físicas e as tecnologias utilizadas pelo microprocessador.*

1. Introdução

O microprocessador Power5 é um processador RISC de 64 bits da IBM. Foi desenvolvido introduzindo-se melhorias no seu antecessor, o Power4, com o intuito de aumentar a performance. Um dos objetivos do seu projeto era de manter a compatibilidade em relação aos dois processadores, não somente os programas compilados para o Power 4 deviam executar corretamente, mas as otimizações feitas nas aplicações deveriam continuar válidas. Assim como o seu antecessor, o Power5 apresenta dois cores no mesmo chip, e alterações no seu design, como na localização da cache L3, permitiram o aumento do poder de processamento. Foi introduzido o paralelismo em nível de thread com duas vias para cada core, o que acarreta um aproveitamento maior de cada ciclo de clock e dos recursos do sistema, e com isso um ganho de desempenho.

Este texto trás as principais características do Power5 e discute as principais mudanças em relação ao Power4 que acarretaram em um ganho de desempenho.

2. Visão geral do chip

O Power5 é a segunda geração de processadores da IBM que incorpora dois processadores RISC de 64 bits idênticos no mesmo chip. Possui 276 milhões de transistores por chip e ocupando uma área de 389mm^2. Foi construído utilizando a tecnologia IBM de semicondutores 130 nanômetros e Silicon-On-Insulator (SOI), que diminui as capacitâncias e consequentemente um aumento de performance. Apresenta uma maior capacidade de processamento que seu antecessor e possui uma arquitetura de sistema que permite uma alta largura de banda de comunicação com a memória. Cada core possui múltiplas unidades de execução e cache L1 própria de instruções e dados. Possui cache L2 compartilhada entre ambas CPUs, e também o diretório e o controlador da cache L3, no entanto a cache L3 propriamente, foi deixada fora do chip devido as suas dimensões físicas (ver Figura 2). Além disso, o chip abriga também controladores de memória e unidade de controle distribuída. A unidade de controle proporciona a interconexão entre os diferentes chips para uso em sistemas de memória compartilhada.

Apresenta também caches maiores se comparado ao Power4 (1.9 MB L2 e 36 MB L3), que reduz a necessidade de acesso à memória e aumenta o cache hit e por conseqüência o desempenho como um todo.

A maior eficiência de uso dos recursos devido as características multithread e o elevado clock causam um alto consumo de energia. Para minimizar isso e deixar a dissipação de calor em níveis aceitáveis, o Power5 possui gerencia Dinâmica de potência (Dynamic Power Management) que reduz o consumo de energia sem perda de desempenho significativa.

O Power5 aumentou a escalabilidade SMP para até 64 vias. Para evitar que este aumento do tráfego sature a interconexão da unidade de controle, a cache L3 opera com um barramento próprio, um para leitura e outro para escrita (ver Figura 1). Isso não só reduz a latência no acesso a esta cache, mas também libera a unidade de controle do tráfego correspondente.

O controlador de memória, que no Power4 era um chip próprio, foi integrado ao chip do Power5, fazendo ligação diretamente com a cache L2 e não mais com a L3, reduzindo drasticamente a latência no acesso à memória (ver Figura 1).

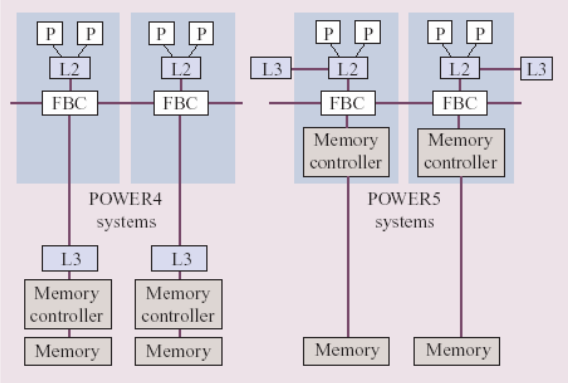

Figura 1. Organização da cache

3. Gerência de memória

O Power5 trabalha com três níveis de memória cache além da memória principal. Fisicamente a cache de nível 1 (L1) fica localizada dentro de cada core. A cache de nível 2 fica fora dos cores e a cache de nível 3 (L3) fica fora do chip devido as suas dimensões (ver Figura 2).

A cache L1 é separada em cache de dados e de instruções. A parte de instruções tem 64Kbytes e associatividade de 2 vias. A parte de dados tem 32Kbytes e associatividade de 4 vias.

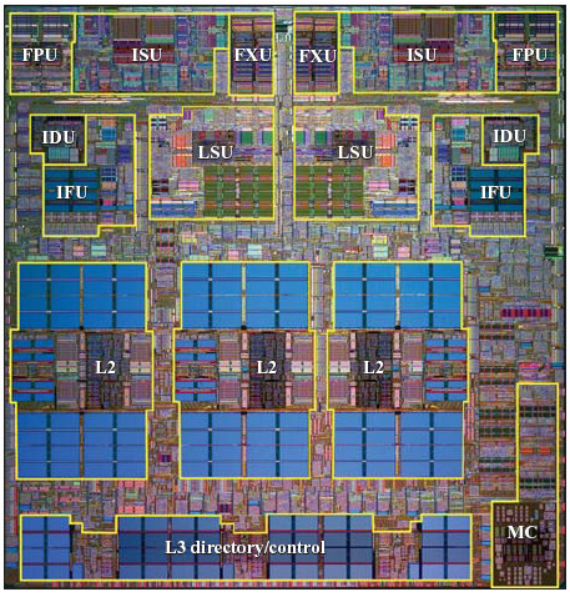

Figura 2. Chip Power5

A cache L2 possui 1,875 MB e é compartilhada pelos dois cores processadores. É dividida em três partes (ou slice), cada uma possui 10 vias set-associative com classe de congruência 512 e 128 bytes por linha. As partes são idênticas e possuem cada uma o seu controlador que pode ser acessado independentemente por cada core. O endereço físico de acesso determina o slice da cache em que a linha desejada está localizada.

O terceiro nível da cache, a cache L3, tem 36MB de memória e fica localizado fora do chip devido ao seu tamanho físico, porém dentro do chip tem um diretório da cache L3 para que a cache somente seja acessada se realmente o dado desejado estiver nela. Dentro do chip também fica o controlador da cache reduzindo os atrasos causados no acesso a controladores externo. A L3 é implementada usando 3 slices, uma para cada slice da cache L2. Cada slice é de 12 vias set-associative, com classe de congruência 4096 com linhas de 256 bytes com dois setores de 128 bytes para equiparar ao tamanho das linhas da cache L2. O acesso a cache L3 se dá através de dois barramentos 16-byte-wide separados, um para leitura e outro pra escrita que operam na metade da freqüência do processador.

A comunicação entre a memória principal e o controlador que fica dentro do chip, se dá através de dois barramentos unidirecionais com os módulos DDR ou DDR2 dependendo do modelo do processador. Os dados lidos do barramento são 16-byte-wide e escritos 8-byte-wide.

4. Pipeline de instruções

O Power5 foi desenvolvido para suportar paralelismo em nível de threads, mais especificamente simultâneos multithread (SMT) que será tratado na sessão 5, no entanto o seu pipeline é idêntico ao do seu antecessor o Power4.

O modelo de execução é fora de ordem com um máximo de oito instruções despachadas a cada ciclo. São oito as unidades de execução existentes: duas unidades para operações com ponto flutuante, duas de loadstore, duas para operações com ponto fixo, uma para branch e uma para operações lógicas.

A figura 3 mostra o pipeline do processador Power 5. Os estágios IF, IC e BP correspondem a predição de fetching e branch. Do estágio D0 até o GD é feita a é decodificação das instruções e são formados os grupos. No MP são determinadas as dependências. Durante o ISS as instruções são despachadas apropriadamente para a unidade de execução, então lê os registradores no estágio RF, executa a instrução no EX e escreve os resultados nos registradores adequados no estágio WB. Para que a instrução possa ser finalizada é necessário que todas as instruções do seu grupo sejam também finalizadas, isto é feito nos estágios Xfer e CP. Na sessão 5 este processo será descrito com mais detalhes.

Dividir os pipelines em vários estágios são uma estratégia utilizada para aumentar a freqüência de operação do processador e diminuir o tempo total de cada instrução. O efeito contrário que isso traz é que uma previsão de desvio errada acarreta numa perca de muitas instruções que já estiverem no pipeline. Para minimizar este problema, o Power5 tem um mecanismo sofisticado de previsão dinâmica de desvio, através da utilização de três tabelas de histórico. Duas delas são usadas para previsão de desvio usando os mecanismos de Previsão Bimodal e Correlacionada e a terceira é usada para informar qual dos dois métodos é o mais provável de estar correto.

As instruções são executadas fora de ordem, por isso é necessário guardar a ordem do programa para todas instruções que estão no pipeline. Com o objetivo de minimizar a lógica necessária para guardar ordem de um grande número de instruções, são formados grupos. Os grupos são seguidos individualmente durante o fluxo de instruções, isto é, o estado da máquina é mantido por grupo e não por instruções individuais.

As instruções são despachadas um grupo por vez. Quando o grupo é despachado as informações de controle do grupo são gravados na tabela GCT (Group Completion Table) que pode armazenar até 20 grupos. Assim que as instruções são finaliza esta informação é gravada na tabela GCT que é mantida até que o grupo seja retirado, isto é, até que todas as instruções do grupo tenham sido executadas.

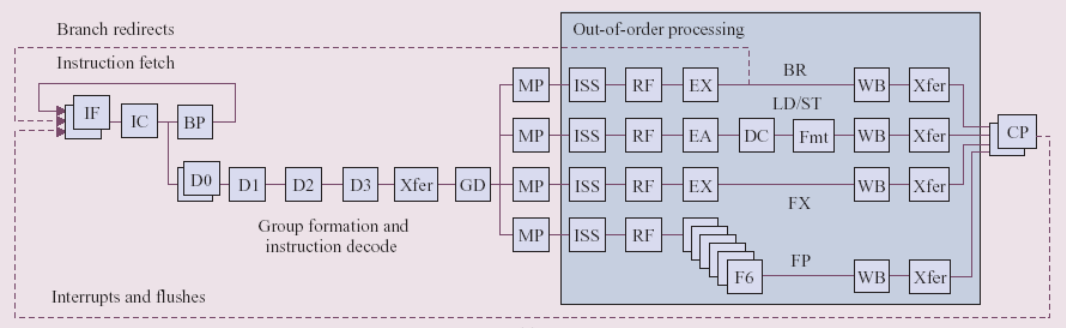

Figura 3. Pipeline do Power5 - (IF: instruction fetch; IC: instruction cache; BP: branch predict; D0: decode stage 0; Xfer: transfer; GD: group dispatch; MP: mapping; ISS: instruction issue; RF: register file read; EX: execute; EA: compute address; DC: data caches; F6: six-cycle floating-point execution pipe; Fmt: data format; WB: write back; CP: group commit)

5. Paralelismo em nível de thread

Em processadores superescalares convencionais, apenas uma thread é executada por vez, e embora o paralelismo exista em nível de instruções, e processador fica ocioso em muitos ciclos devido a eventos de longa latência como os acessos a memória devido a situações de cache miss. A implementação de multithread em nível de hardware diminui estes ciclos perdidos, pois enquanto uma thread está parada devido a um evento de longa latência, o processador pode executar uma outra thread. Para o sistema operacional, um processamento a multithread é visto de forma semelhante a um processamento simétrico.

Existe basicamente três métodos de implementação de paralelismo multithread em nível de processador: coarse-grain multithread, fine-grain multithread e simultâneous multithread (SMT), que o método que o processador Power5 utiliza.

No método coarse-grain multithread somente uma thread executa a cada instante de tempo. Quando esta thread encontra um evento de longa latência, como um cache miss, o hardware troca para a outra thread que passa a usar os recursos do processador. Já no método fine-grain a thread é trocada a cada ciclo independente de alguma delas estar aguardando um estado de longa latência ou não. O método simultâneos multithread, que é implementado no Power5, também faz a troca de thread a cada ciclo, porém quando uma delas encontra um estado de longa latência o hardware troca a execução para a outra thread e desta forma tem-se uma espécie de misto dos dois métodos anteriores, fazendo uso das vantagens de cada um.

O Power5 suporta dois modos de operação: simultaneos multithread (SMT) ou single thread (ST). No modo SMT dois contadores de programa (PC) são usados, um para cada thread, de modo as instruções são buscadas alternadamente para uma thread e outra. Similarmente, as predições de desvio também são feitas de forma alternada entre as threads. No modo ST apenas um contador de programa é usado e as instruções são buscadas para esta thread em todos os ciclos. A execução na forma multithread é descritas abaixo.

Após carregar o contador de programa com o endereço da próxima instrução, até oito instruções podem ser buscadas na cache de instruções a cada ciclo. A cache de instruções é compartilhada pelas duas threads. Em um dado ciclo, instruções de uma mesma thread são buscadas e então é verificada a existência de algum branch entre as instruções. Se algum for encontrado a direção do branch é prevista utilizando-se as três tabelas de previsão de branches, que também são compartilhadas entre as threads. Depois que as instruções são buscadas, são colocadas em dois buffers de instruções separados, um para cada thread (ver Figura 4). Cada um deles pode ter até 24 instruções. Baseado na prioridade das threads, até cinco instruções são buscadas no buffer e os grupos são formados. Cada grupo só possui instruções de uma mesma thread e todas são decodificadas em paralelo. Quando os recursos necessários ao grupo estão disponíveis, o grupo é despachado na ordem do programa. Depois do despacho, é feita a renomeação e o mapeamento dos registradores lógicos para os físicos. Todos os registradores físicos são compartilhados pelas duas threads. Feitas todas as renomeações, as instruções são colocadas nas filas de despacho. As informações de cada grupo de instruções despachado são gravadas numa tabela (GCT) para que possa ser controlada a finalização de cada instrução do grupo e posteriormente reordenar as instruções na ordem real do programa. Assim que os dados de entrada das instruções estão disponíveis, elas são delegadas para a execução. Para a delegação não existe distinção entre instruções de diferentes threads, prioridades ou grupos. Até oito instruções podem ser delegadas ao mesmo tempo, uma em cada unidade de execução. Então é realizada a execução e em seguida são gravados os resultados nos registradores físicos. Quando todas as instruções de um grupo tiverem sido finalizadas, então o grupo todo é liberado.

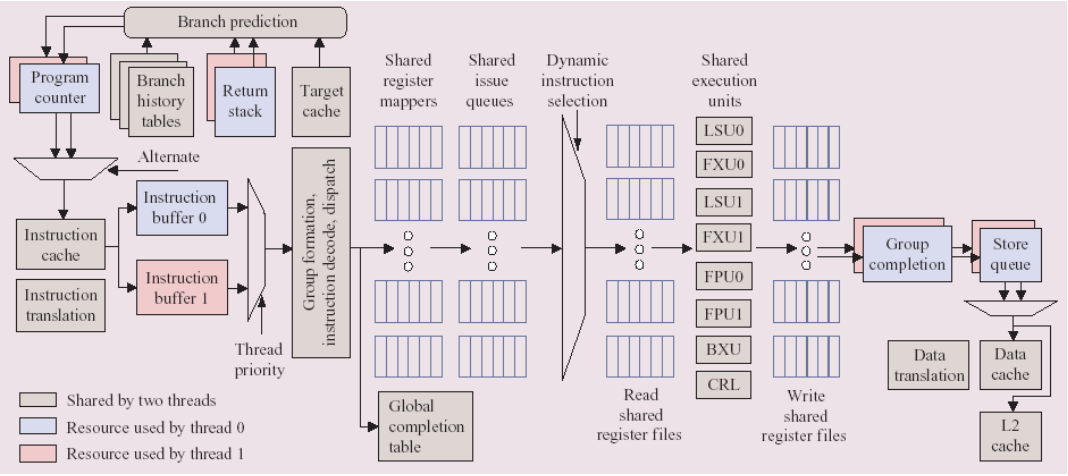

Figura 4. Fluxo de Instruções

6. Considerações finais

O processador Power5 é uma evolução do Power4 que trás como principal característica o uso de simultaneous multithread (SMT) com o intuito de fazer um melhor aproveitamento dos recursos de hardware disponíveis. Outras mudanças significativas foram realizadas que resultaram em um maior desempenho que o Power4, mesmo quando utilizado em modo single thread. A melhora na tecnologia dos semicondutores, a utilização de barramento exclusivo para a cache L3 e o aumento das caches propiciou um ganho significativo de desempenho.

Referências

Sinharoy, B., Kalla, R. N., Tendler, J. M., Eickemeyer, R., J. and Joyner, J.,B. (2005) "POWER5 system microarchitecture", In IBM J. Res. & Dev. vol. 49 num. 4/5, p. 505-521

Kalla, R., Sinharoy, B., Tendler, J. M., (2004) "IBM Power5 Chip: A dual-core Multithreaded Processor", Micro, IEEE, Vol. 24, Issue 2, p. 40 – 47

Tendler, J. M., Dodson, S., Fields, S., Le, H., Sinharoy, B. (2001) "POWER4 System Microarchitecture", Technical White Paper, IBM Server Group, http://www-03.ibm.com/servers/eserver/pseries/hardware/whitepapers/power4.pdf