

Unidade de Ensino: 4

Competência da Unidade: Aprofundar o conhecimento sobre arquiteturas de alto desempenho.

Resumo: Nesta unidade será apresentado as arquiteturas CISC e RISC, paralelas, multithread e multicore

Palavras-chave: CISC, RISC, paralelas, multithread e multicore

Título da Tele aula: Arquiteturas de Alto Desmpenho

Tele aula nº: 04

Contextualização

Você foi contratado como um dos analistas do projeto do desenvolvimento da arquitetura de um novo processador de uma importante fabricante internacional de processadores, conhecida pela qualidade e pela inovação.

Esse novo processador deve ser projetado para oferecer um

Esse novo processador deve ser projetado para oferecer un desempenho superior ao dos concorrentes.

Existem diversas questões que implicam decisões impactantes para o futuro produto que exigem um compromisso (ou trade-off, em inglês) no desempenho desse processador.

Uma das primeiras decisões que a equipe de projeto deve tomar é em relação à quantidade de instruções. Ela deve escolher entre um número maior de instruções com instruções mais complexas _ou_um_número_menor de instruções com sintaxe simplificada.___



Fonte: Shutterstock

Contextualização

Haverá uma reunião de trabalho entre toda a equipe de trabalho e o gerente, e você ficou responsável de trazer diversos elementos para enriquecer a discussão.

Ao final dessa reunião vocês chegarão a um consenso sobre quantidade de instruções que o processador suportará, e complexidade de cada instrução.

O projeto já está atrasado em relação ao cronograma, e caso a decisão seja equivocada trará consequências nas próximas etapas de definição da arquitetura.

Além disso, o principal concorrente da sua empresa também está desenvolvendo um projeto similar.



Contextualização

Você já participou de reuniões anteriores que não chegaram a um consenso sobre as características desejadas.

Foi debatido o tamanho da instrução: se deveria ter a mesma quantidade de elementos na instrução ou se quantidade desses elementos poderia variar.

A conclusão dessa discussão contribuirá para definição da quantidade e complexidade das instruções com que o processador vai trabalhar.



onte: Shutterstock

Conceitos

Introdução a Arquiteturas de Alto Desempenho



Sua missão:

Como material para essa reunião, você deve levantar diversas recomendações para o projeto com base nas arquiteturas CISC e RISC, indicando quais características seriam mais adequadas para o projeto.

Prepare um relatório com as características das arquiteturas CISC, RISC e as desejadas no projeto e verifique qual arquitetura está mais próxima do desejado. O futuro desse importante projeto está em suas mãos.



Contextualizando

- Pesquisadores estudam métodos e técnicas para melhoria do desempenho e da capacidade de sistemas computacionais.
- Software = com desenvolvimento de sistemas operacionais,
- Hardware = arquitetura dos processadores. Evolução do hardware: a velocidade de processamento da frequência do clock (velocidade em ciclos por segundo - hertz - para executar as funções) do computador passou de poucos MHz em 1980 para GHz na década de 2010, além da criação de novas tecnologias de processamento como pipelin processamento paralelo.
- Linguagem de máquina x linguagem de programação
- Inicialmente foram utilizados conjuntos extensos de instruções, dezenas de modos de endereçamento.



Contextualizando

- Este tipo de arquitetura de computadores foi denominado Complex Instruction Set Computers, ou computadores com conjunto de instruções complexas.
- Avanços na área de arquitetura de processadores = com um conjunto menor de instruções, seria possível melhorar o desempenho do processador com uma arquitetura denominada RISC - Reduced Instruction Set Computers



Características da Arquitetura CISC

A arquitetura de processadores CISC (Complex Instruction Set Computers) ou processadores com conjunto de instruções complexas, apresentava grande quantidade de instruções com múltiplos modos de endereç

Os processadores com tecnologia CISC são capazes de processar $\,$ centenas de conjuntos complexos de instruções simples. Isto significa que cada instrução isoladamente é considerada simples, curta e pouco potente, porém várias destas instruções agrupadas formam um conjunto complexo que é executado processador.

A ideia dos fabricantes era produzir processadores cada vez mais potentes baseados na complexidade destes conjuntos de instruções (BROOKSHEAR, 2013).



Características da Arquitetura CISC

Inicialmente, existia uma grande tendência a esta tecnologia de processadores, porém havia algumas desvantagens, como o desempenho reduzido justamente pelo excesso de instruções executadas pelo processador e pela velocidade de processamento ter que ser elevada para que o desempenho fosse melhorado.

Dessa forma, pode-se concluir que os projetistas de arquiteturas CISC consideram três aspectos básicos (MONTEIRO, 2010): Utilização de microcódigo.

Construção de conjuntos com instruções completas e eficientes. Criação de instruções de máquina com complexidade similar aos

comandos de linguagens de alto nível. Na arquitetura CISC, o objetivo era a criação de um conjunto numeroso de funções, na maioria das vezes complexas e com vários operandos, o que aumentava o tempo de execução.



Características da Arquitetura RISC

A ideia original era entregar um conjunto mínimo de instruções que poderiam realizar todas as operações essenciais: movimentação de dados, operações para Unidade Lógico Aritmética e desvios. Somente instruções explícitas LOAD e STORE tinham permissão para acessar a memória.

Alguns elementos constituem a base da arquitetura RISC: o pequeno conjunto de instruções como tamanho fixo e execução rápida da instrução.

A cada ciclo de clock uma instrução é selecionada (buscada) e executada. Essa arquitetura contém uma menor quantidade de modos de endereçamento, maior quantidade de registradores e

O pipeline (ou paralelismo) é a divisão de uma instrução em muitas partes, para que cada parte seja manipulada por uma



Características da Arquitetura RISC

O pipeline (ou paralelismo) é a divisão de uma instrução em muitas partes, para que cada parte seja manipulada por uma parte de hardware dedicada e, com isso, todas as partes possam ser executadas em paralelo (TANEMBAUM, 2013).

Na arquitetura RISC existe uma diminuição da quantidade de instruções disponíveis e o modo de endereçamento em relação à arquitetura CISC

As máquimas RISC utilizam os registradores da Unidade Central de Processamento para armazenamento dos parâmetros e variáveis em chamadas de funções e rotinas.

Processadores ARM usados em celulares são um bom exemplo de uso da arquitetura RISC. Outro exemplo de uso dessa arquitetura é em consoles, como o Nintendo 64 e o Playstation.



13

RISC x CISC

Uma vez que o tempo para buscar uma instrução não é mais uma preocupação e com memória barata para armazenar uma quantidade grande de instruções, não existe uma vantagem real em instruções CISC. (Murdocca, 2000).

Todas as instruções RISC têm tamanho fixo, uma palavra.

Todas as instruções RISC efetuam operações simples que podem ser iniciadas no pipeline à taxa de uma por ciclo de clock.

Arquiteturas RISC x CISC CARACTERÍSTICAS RISC CISC ARQUITETURA Registrador-registrador Registrador-memória TIPOS DE DADOS Pouca variedade Muito variada FORMATO DAS Poucos endereços Muitos endereços INSTRUÇÕES MODO DE Pouca variedade ENDEREÇAMENTO ESTÁGIOS DE PIPELINE Entre 4 e 10 Entre 20 e 30 ACESSOS AOS DADOS Via registradores Via memória



No relatório que você preparará para a reunião, os elementos quantidade e complexidade das instruções são importantes da arquitetura, e você deve levantar diversas características que sejam relevantes, tais como:

Arquiteturas RISC não apresentam microprogramação e microcódigo; todas as instruções são executadas diretamente pelo hardware.

Na arquitetura CISC era comum, em estruturas computacionais, adicionar uma camada de software entre o hardware e os demais programas, providenciando um novo nível de interpretação, o microcódigo. O microcódigo oferece a possibilidade de incluir ou alterar instruções sem alterar o hardware.

Entretanto, ter a maior flexibilidade não é considerada uma boa prática, pois o nível adicional de interpretação acarreta tempo de processamento para interpretação e, com isso, perda de desempenho, que não compensa às novas instruções que possam ser adicionadas. As instruções devem ser simples de decodificar, uma característica da arquitetura RISC. Quanto menos complexas forem e menos formatos alternativos tiverem as opções de execução de uma instrução, será melhor.

Isto ocorre porque quando a instrução tem poucos formatos, o tempo de identificação da instrução será reduzido. Logo, o tempo de processamento será otimizado.

Somente as instruções LOAD e STORE devem referenciar a memória.

O acesso à memória principal é uma operação mais demorada que o acesso a registradores. Por esse motivo, é recomendado que apenas as instruções LOAD (leitura) e STORE (gravação) tenham acesso à memória. Isto ocorre na arquitetura RISC. Maximize a taxa de execução de instruções.

3

Além do aumento da velocidade de clock, uma das formas de maximizar o desempenho do computador é por meio do paralelismo, amplamente utilizado na arquitetura RISC.

Implemente muitos registradores de uso geral.

Quando o processador não tem um registrador disponível para armazenar um valor resultante, ele transfere esse valor para a memória principal.

Já sabemos que a transferência de dados entre o processador e a memória principal é um processo mais lento do que a movimentação de dados dentro do processador. Assim, quanto mais registradores possíveis e de uso geral, melhor.

A arquitetura RISC apresenta maior quantidade de registradore que a CISC.

Conceitos

Arquiteturas
Paralelas

Introdução a Arquiteturas Paralelas

SISD - Single Instruction, Single Data

MEM = memória

"Uma arquitetura paralela fornece uma estrutura explícita e de alto nível para o desenvolvimento de soluções utilizando o processamento paralelo, através da existência de múltiplos processadores que cooperam para resolver problemas através de execução concorrente" (Duncan, 1990)

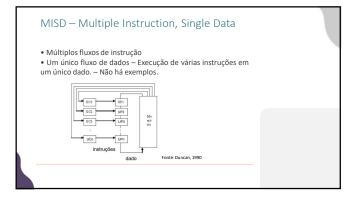
Classificação de Máquinas Paralelas

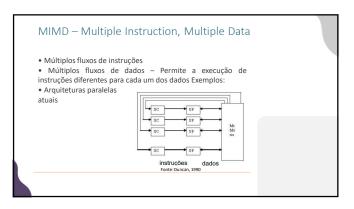
✓ Classificação de Flynn — Classificação de arquiteturas paralelas genérica de acordo com as características do fluxo de instruções e o fluxo de dados (único ou múltiplo).

	Sequência de Instruções	Sequência de Dados	Nome	Exemplos
_	1	1	SISD	Máquina clássica de Von Neumann
	1	Várias	SIMD	Supercomputador vetorial, processador matricial
	Várias	1	MISD	Nenhum exemplo
	Várias	Várias	MIMD	Multiprocessador, multicomputador

Um único fluxo de instruções Um único fluxo de dados Contém as arquiteturas tradicionais não paralelas Máquinas de Von Neumann. Instruction Steram Data Steram Fonte: Duncan, 1990 UC = unidade de controle UP = unidade de processsamento

SIMD - — Single Instruction, Multiple Data • Um único fluxo de instrução • Múltiplos fluxos de dados — Execução síncrona da instrução para todos os dados — Exemplos: • Processadores vetoriais (Cray 1) • GPU • GPU Fonte: Duncan, 1990







Nesta etapa do projeto de novo processador de alto desempenho, foram apresentadas as arquiteturas para processamento paralelo com a implementação de pipeline, abordando suas classificações, características e utilizações. Foram detalhadas as classificações de Flynn

Essa classificação divide os processadores da seguinte maneira: com instrução simples ou múltiplas instruções e com dados únicos ou dados múltiplos. Os quatros tipos são: SISD (única instrução, único dado), SIMD (única instrução, múltiplos dados), MISD (múltiplas instruções, único dado) e MIMD (múltiplas instruções, único dado) e o miltiplos dados). No SISD há apenas um processador, e o exemplo são máquinas uniprocessadas, sendo recomendada para processamento de texto. SIMD executa uma instrução específica para vários grupos de dados, sendo recomendada para operações matriciais.

MISD não é uma arquitetura utilizada comercialmente. Já a arquitetura MIMD consiste em multiprocessadores, e a maior parte de sistemas paralelos utilizam essa arquitetura.

Nessa etapa do projeto será decidido o tipo de arquitetura em relação ao processamento, e você pesquisou sobre as arquiteturas paralelas segundo a classificação de Flynn.

Com base nos requisitos do projeto, para desenvolver um processador de alto desempenho a arquitetura mais adequada é a MIMD, mais avançada tecnologicamente e que apresenta elevado desempenho do sistema de computação.

É um conjunto de processadores que executam sequências de instruções diferentes simultaneamente em diferentes conjuntos de dados. Máquinas MIMD, que adotam vários pontos de controle, apresentam sequências de instruções e dados independentes.

Assim, nesse projeto, a arquitetura mais recomendada para ser utilizada é a MIMD.







Sua missão

Você deve pesquisar as arquiteturas multithread e preparar um relatório com as suas principais características, que vai contribuir para a escolha da arquitetura mais apropriada para o projeto.



Contextualizando

Dentro do projeto de desenvolvimento de um novo processador, iniciou-se a fase de definição do fluxo de um processo e de seus componentes, sendo as threads um desses componentes.

Você deve conhecer a arquitetura do computador, que resulta em elementos visíveis a um programador. Você já participou de reuniões anteriores que definiram por

consenso que o projeto terá troca de threads, que são mais ra

Foi definido que a arquitetura multithread deve buscar as seguintes características: capacidade de envio de instruções supe utilização de múltiplos contextos de threads. Entretanto, não houve a definição de arquitetura multithread a ser utilizada.



Contextualizando

Desde os primeiros computadores, os cientistas se empenham para fazer as máquinas resolverem problemas de forma melhor

A aplicação de técnicas de microeletrônica resultou em circuitos integrados (CIs) de alta complexidade e encapsulados em um único chip, isto é, o CI está dentro de um invólucro protetor. Os clocks ou relógios aumentaram a frequência para a faixa de Gigahertz (GHz). Entretanto, existem barreiras físicas que

delimitam o desempenho de uniprocessadores.



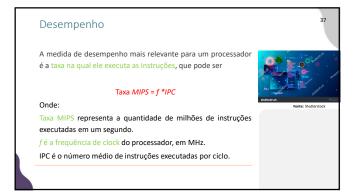
Desde os primeiros computadores, os cientistas se empenham para fazer as máquinas resolverem problemas de forma melhor

Contextualizando

A aplicação de técnicas de microeletrônica resultou em circuitos integrados (CIs) de alta complexidade e encapsulados em um único chip, isto é, o CI está dentro de um invólucro protetor.

Os clocks ou relógios aumentaram a frequência para a faixa de Gigahertz (GHz). Entretanto, existem barreiras físicas que delimitam o desempenho de uniprocessadores.

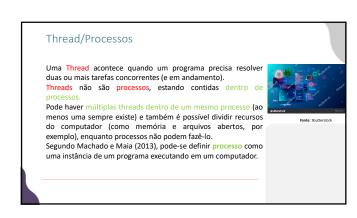


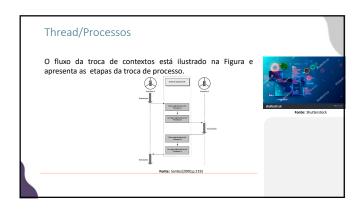


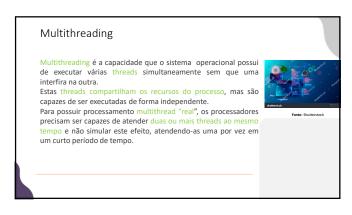
uma ordem diferente da forma que ocorrem no fue de instruções e começar a execução de instruções que podem nunca ser necessárias. Entretanto, essa abordagem pode estar alcançando o limite, por causa da complexidade e dos problemas de consumo de energia. Segundo Monteiro (2010), uma dessas estratégias para melhoria do desempenho é conhecida como superpipelining e ocorre quando um pipeline tem

estágios que requerem menos da metade de um ciclo









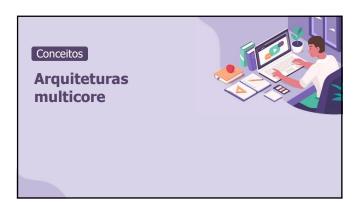


Analisando a situação-problema proposta, é necessário a determinação de qual arquitetura multithread será adequada. Por sua vez, multithreading simultâneo (SMT) possibilita execução paralela e simultânea de várias threads. Esse é um sistema capaz de emitir diversas instruções ao mesmo tempo. Se um thread apresenta um alto grau de paralelismo em nível de instruções, ela pode, em alguns ciclos, ser capaz de preencher todos os slots horizontais. Em outros ciclos, as instruções de duas ou mais threads podem ser enviados. Se threads suficientes estão ativos, normalmente seria possível enviar o número máximo de instruções em cada ciclo, fornecendo um nível alto de eficiência.

Já no caso de chip multiprocessadores ou multicore, cada núcleo tem um processador superescalar de envio de duas instruções simultâneas.

Para cada processador é atribuído um thread, a partir do qual ele pode enviar até duas instruções por ciclo. É possível combinar as potencialidades das arquiteturas

E possivei combinar as potencialidades das arquiteturas multithreading simultâneo (SMT) e chip multiprocessadores ao utilizar multithread dentro de cada processador em um chip multicore, para potencializar o ganho de processamento, fato que é implementado em diversos processadores atuais.



Sua Missão

Chegou o momento de mostrar seu conhecimento como analista de projetos para o desenvolvimento de um novo processador de alto desempenho.

Já está acertado com a equipe que serão utilizados processadores multicore, e isso tornará mais viável a ideia de paralelismo entre os núcleos do processador, ou seja: trabalhar com uma arquitetura que possibilite a execução simultânea, em que cada núcleo admite uma tarefa.

Você deverá apresentar para a equipe um relatório do funcionamento dos núcleos do processador em um ambiente Windows.



Fonte: Shutterstock

Sua Missão

Você poderá utilizar a ferramenta de gerenciamento de tarefas para facilitar suas explicações.

Faça um print (cópia da tela) da execução das tarefas dos respectivos núcleos e, finalizando o seu relatório, contextualize o funcionamento dos threads e dos caches envolvidos em cada core (núcleo) do processador.

Fique à vontade para caracterizar um processador em específico.



Fonte: Shutterstock

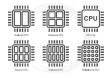
Introdução a Arquiteturas Multicore

Hoje, com a necessidade de otimização de espaço físico, são desenvolvidos chips de processamento multicore (múltiplos núcleos), nos quais é possível otimizar os transistores, gerando economia de energia e a redução de calor.

Um processador multicore pode ser definido como dois ou mais núcleos de processamento em uma única pastilha de silício (chip). As tarefas a serem executadas são distribuídas entre esses núcleos do processador. Em outras palavras, cada núcleo pode executar múltiplos processos simultaneamente. A Figura ilustra a estrutura de processadores multicore.

Processadores com múltiplos núcleos de processamento

São processadores com dois, quatro, um chip, seis, oito e dezesseis núcleos.



Processadores com múltiplos núcleos de processamento

Cada núcleo de processamento é organizado como uma unidade de central de processamento independente, composta por unidade lógica aritmética e unidade de controle e registradores,

Além dessas, há as memórias cache, que podem ser compartilhadas entre os núcleos e podem ser dedicadas. Por exemplo: temos as caches Level 1 (L1), Level 2 (L2) e Level 3 (L3) para dados e de instruções.

Ao visitar sites de fabricantes de processadores, temos um termo frequentemente utilizado, o Smart

Processadores com múltiplos núcleos de processamento

processamento do processador multicore, sendo as caches L2 e L3 compartilhadas entre os núcleos.

Os engenheiros implementaram alterações na organização de processadores permitindo o aumento do nível de paralelismo em nível de instruções, de forma a executar mais de uma instrução a cada ciclo de clock.

Posteriormente, com o surgimento da arquitetura superescalar, foram implementados vários pipelines com recursos de execução replicados para possibilitar

Processadores com múltiplos núcleos de processamento

Como forma de melhorar o desempenho da arquitetura superescalar, foi desenvolvida a arquitetura multiprocessada (Smultaneous Multiprocessing – SMP), que consiste em arquitetura com mais de um processador.

Esse tipo de arquitetura tem recursos de memória, disco compartilhados e roda em um mesmo sistema operacional. Nesse sentido, dois processos podem ser executados ao mesmo tempo em dois processos podem ser executados ao mesmo tempo em dois processos dores.

Ainda como evolução, surgiu a arquitetura multithread simultânea (SMT, do inglês Simultaneous Multithreading), também chamada de hyper-threading.

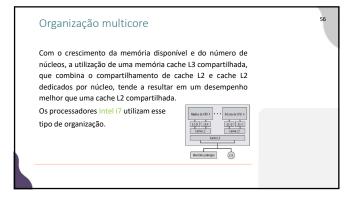
Processadores com múltiplos núcleos de processamento

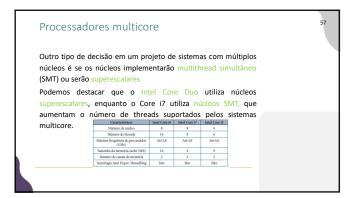
Nessa técnica é explorado o pipelining em nível de threads em um processador. Assim, um processador físico pode simular dois processadores lógicos. Cada um dos processadores lógicos tem o seu próprio controlador de interrupção, bem como um banco de registradores.

Essa arquitetura conta também com alguns recursos compartilhados, tais como, cache, ULA e barramentos.

Na Figura é apresentada uma arquitetura de quatro processadores físicos, sendo dez núcleos por processador que serão compartilhados com a memória cache L3 (30 MB compartilhado entre os dez core).





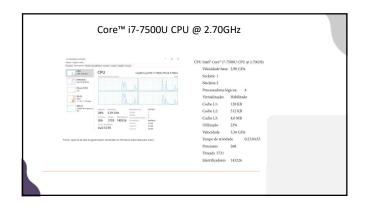




Você ficou encarregado de elaborar um relatório para a equipe de desenvolvimento do novo processador, pelo qual deverá mostrar o funcionamento dos núcleos do processador em um ambiente Windows.

Nesse caso específico foi realizada uma demonstração pelo gerenciador de tarefas do Windows, mas você poderá usar outras ferramentas para suas colocações. Finalize o relatório com uma contextualização sobre o funcionamento dos threads e dos caches envolvidos em cada core (núcleo) do processador.

Funcionamento dos núcleos do processador em um i7 de 7ª geração da Intel. Core™ i7-7500U CPU @ 2.70GHz



Podemos considerar que a execução das threads no processador pode ter o compartilhamento das informações. Cabe relatar que esse armazenamento na cache pode ser feito distintamente em outros core do processador.

Lembre-se de que as memórias caches apresentam três níveis: a L1 é a de menor nível e menor espaço, porém muito mais rápida que a L2 e L3. A L1 é dividida em duas partes, uma para instruções e outro para os dados. A L2 é destinada para gravação de dados e informações, e a L3 é a maior entre os três níveis, e pode dar apoio a qualquer core para execuções de tarefas.

De acordo com Stallings (2013), existe um ganho quando há um compartilhamento do cache L2 e em paralelo há cache L2 dedicada, como ocorre no caso de Intel Core i7, em que cada núcleo tem a sua cache L2 dedicada, e os núcleos compartilham uma memória cache L3.

Uma das vantagens é a que interferência construtiva pode reduzir as taxas de falhas do sistema, ou seja, se uma thread em um core acessa uma posição de memória, esse acesso traz para linha de conteúdo a posição de referenciada para a cache compartilhada. Caso outra thread de outro core tente acessar depois, as posições de memória já estarão disponíveis na cache compartilhada no chip, gerando maior rapidez de acesso.

Utilizando algoritmos adequados de substituição de linha, existe uma alocação dinâmica da quantidade de cache compartilhada para cada core, de modo que threads com maiores conjuntos de trabalho apresentam mais cache. Além disso, a comunicação entre processadores é facilitada quando implementada por meio das posições de memória compartilhadas.

Assim, a estrutura multicore do processor pode ser representada na Figura.

Núdeo de CPU 1

Núdeo de CPU 1

L1-D L1-L

Cache L2

Cache L3

Memória principal

E/S

Para finalizar, a 10ª geração de processadores Intel® Core™ dispõe de gráficos Intel® Iris® Plus, que proporcionam uso de inteligência artificial (IA) no computador em larga escala para acelerar o desempenho.

Esses processadores possibilitam um novo nível de integração para aprimorar as experiências com PCs atuais e as necessárias para o futuro.

Como pudemos verificar, a organização e arquitetura do processador evoluíram muito com o tempo. Como essa arquitetura pode mudar ainda nos próximos anos?



- Introdução à arquitetura de alto desempenho;
- Arquiteturas Paralelas;
- Arquiteturas Multithreaded;
- Arquitetura Multicore.

