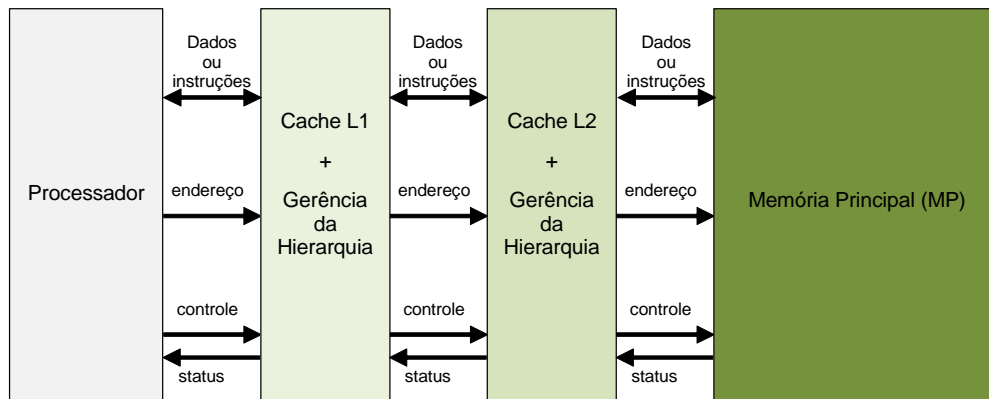


## Hierarquia de Memória Cache – Versão 14

Este documento descreve o terceiro trabalho da disciplina de Arquitetura de Computadores II, que consiste na implementação de um sistema de gerenciamento da hierarquia de memórias cache. Este gerenciamento deve ser descrito em VHDL, e simulado junto com o processador utilizado no Trabalho 2. A hierarquia de memória de dados, que deve ser implementada com três níveis: (i) cache de nível 1 (L1), (ii) cache de nível 2 (L2) e (iii) Memória Principal (MP), tal como ilustrado na figura abaixo.



O processador se comunica com a hierarquia de memória através da porta de endereços, da porta de controle que define o acesso à memória, da porta de status para indicar se a informação lida da memória já encerrou, além da porta bidirecional de dados.

Independente da implementação, pois o aluno tem algumas liberdades de escolha, os sinais provenientes do processador devem ir diretamente para o nível logo abaixo da hierarquia de memória, e cada nível será responsável por gerar os endereços para o nível subsequente.

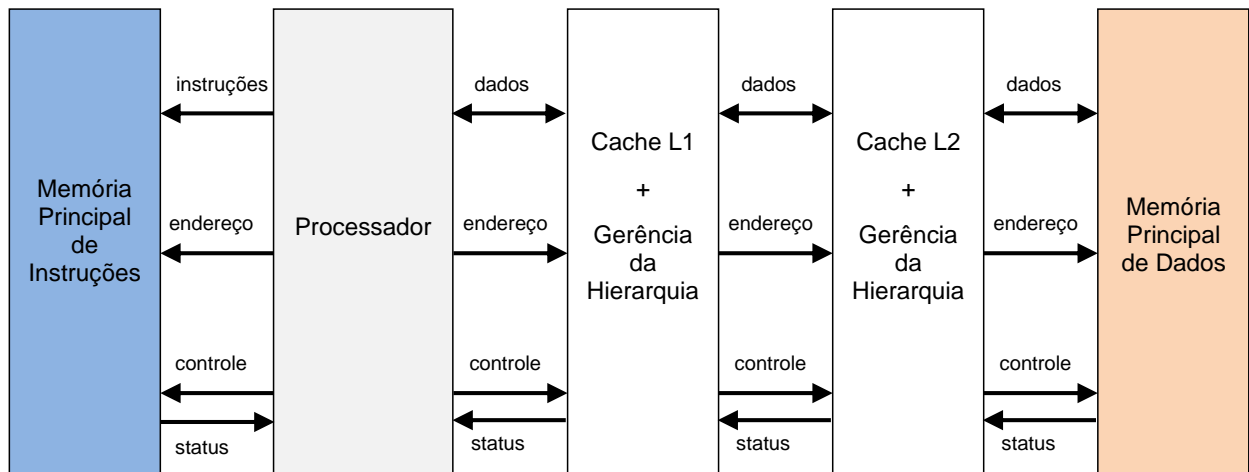
Cada nível de cache tem um conjunto de sinais que reportam o resultado da operação de leitura, tais como miss ou hit. Para cada nível, também, devem ser direcionados sinais de controle, bem como endereços.

A memória de instruções deve ser implementada através de um arquivo contendo um código executável do processador. **Este código deve ser descrito de forma a verificar a funcionalidade da hierarquia da memória que está sendo implementada.** Assim, por exemplo, um programa que teste a hierarquia de instruções deve fazer diversos acessos à memória, forçando que ocorram alguns casos de cache miss e outros de cache hit de forma a explorar a localidade espacial e temporal do programa. O arquivo que contém o código executável do processador pode ser obtido com um programa *assembly* do MIPS sendo entrada para o montador do processador (e.g., MARS).

Considere que para o nível 1 a cache terá 4 linhas, e cada bloco terá 4 palavras. A cache de nível 2 terá 8 linhas e cada linha com 8 palavras. Considere, também, uma MP com 8 K palavras.

Entre cada nível da hierarquia de memória deve ser considerado um atraso parametrizável. A cache L1 opera na mesma frequência do processador (acesso com um ciclo de relógio). A cache L2 tem tempo de acesso dois ciclos de relógio, enquanto que a MP tem tempo de acesso de 16 ciclos de relógio. Demais características de implementação, não definidas aqui, estão livres para serem escolhidas. Para implementação das caches de dados devem ser consideradas as seguintes características: (i) Cache L1: mapeamento direto; (ii) Cache L2: mapeamento associativo. Para ambas as técnicas, o aluno deve implementar um mecanismo para manter a integridade de dados, tal como *write-through* ou *write-back*.

Ao final do trabalho o aluno terá uma arquitetura semelhante a que segue na figura abaixo.



### Realização do trabalho

O trabalho deverá ser realizado em *no máximo* grupos de até 2 (dois) alunos, e ser entregue até o dia descrito na agenda da disciplina. ***Deve ser entregue um relatório descrevendo sucintamente as atividades feitas, juntamente com os arquivos fonte VHDL da implementação do trabalho e os arquivos de teste utilizados.*** É importante descrever a validação da gerência da hierarquia de memória. Note que validar este tipo de sistema implica diversos testes: tais como sucessivos miss e hits. ***Espera-se que os grupos, apresentem gráficos de como ficam as latências de quatro programas (feitos pelos alunos, eventualmente copiados de livros ou Internet) para uma arquitetura sem cache, com cache apenas de nível 1 e com cache de nível 1 e 2, tal como exemplificado a seguir.***

