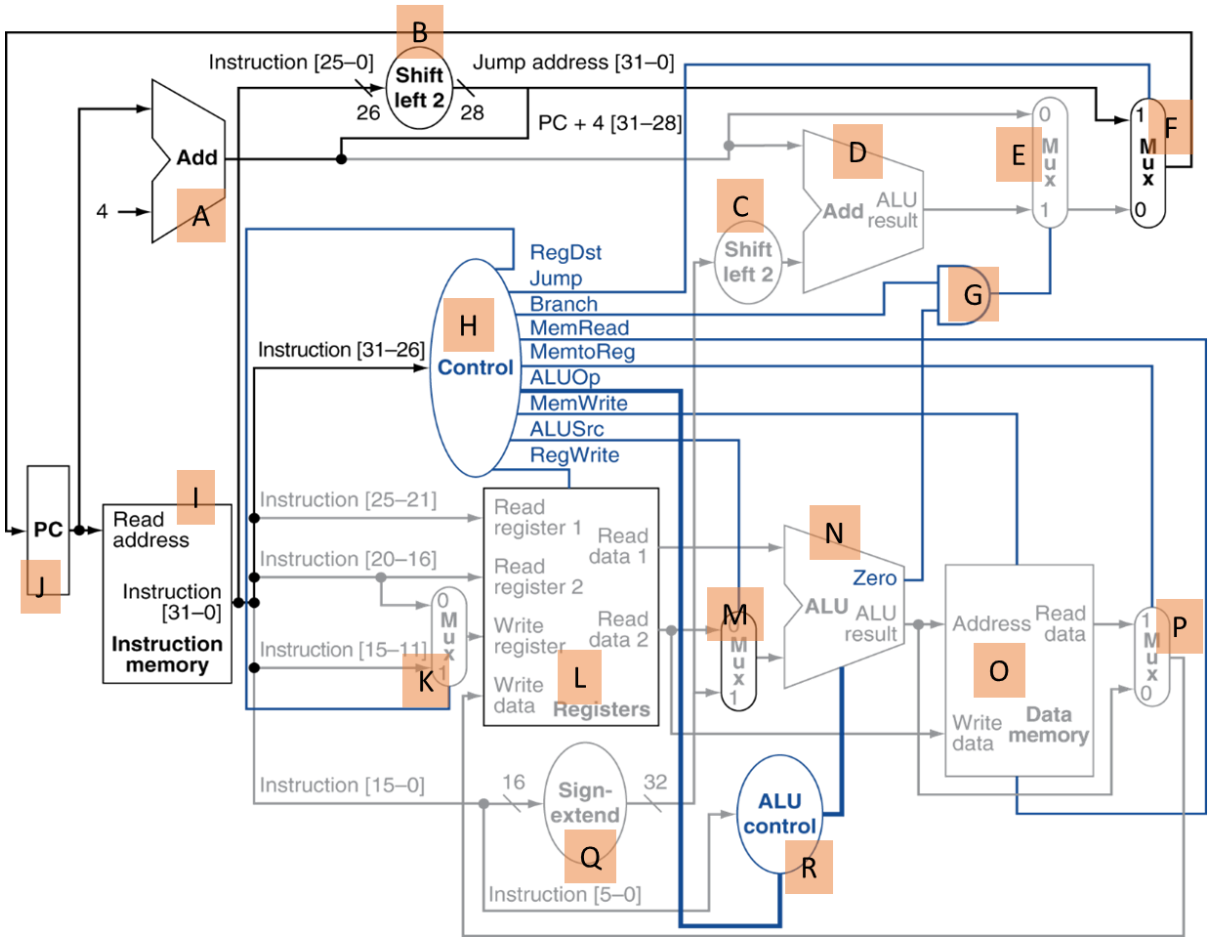
Department of Computer Science and Engineering, IIT Delhi
Computer Architecture (COL 216)
II Semester 2012-21
Major Exam
Maximum Marks: 75
10 May 2021, 9:30 to 11:00 AM

1. **[10 Marks]** We studied that in MIPS assembly language, when a function F calls function G, the function parameters are copied into registers a0, a1, a2, etc., before calling G. Why is this step needed? Why can't we just use in function G the appropriate register from the function F that contains the parameter data?
2. **[6 Marks]** The **bne** instruction has a 16-bit immediate operand. How is this instruction implemented in hardware, as instruction addresses won't fit into 16 bits?
3. **[9 Marks]** If a direct-mapped cache has 32 blocks, and we need to access data from Memory Block A, comment on the suitability of the following mapping functions for determining the cache location for the block.
    a. Can we use cache block (A mod 16) to store memory block A? Why?
    b. Can we use cache block (A mod 32) to store memory block A? Why?
    c. Can we use cache block (A mod 64) to store memory block A? Why?
4. **[10 Marks]** Which data values are stored in pipeline register ID/EX (which separates the Instruction Decode/Register Fetch stage from the ALU/Execute stage)? Why?
5. **[20 Marks]** Consider a loop with only 5 instructions within it, that runs for 10,000 iterations on a pipelined processor. It seems that repeatedly fetching the same 5 instructions from the instruction memory is wasteful of energy. This is true even if we have a separate instruction cache; a typical instruction cache may be 16 KB or 32 KB, so accessing this memory wastes energy. Let us try out an energy-efficient architectural variation where we detect small loops at run time and keep them in a memory structure closer to the pipeline, so that as long as we are in the loop, we can skip accesses from the instruction cache most of the time. Show how you would implement such a feature and draw the relevant parts of the modified MIPS pipeline to accommodate this feature. What is the impact of such a feature on area and performance?

[Continued on the next page…]

6. **[10 + 10 = 20 Marks]** Consider the architecture diagram of the single-cycle processor below, with the component delays labelled A, B,…, etc., as shown.
   a. For each of the following relationships between the delays, indicate whether it is TRUE, FALSE, or NOT ENOUGH INFORMATION. Justify briefly. (1) A < N, (2) B = zero, (3) C < Q, (4) B > R, (5) K < M, (6) M = P (7) L = O [note: "O", not zero], (8) H = R, (9) J < L, (10) G < E. Ignore wire delays and focus only on gate/logic delays. Note that higher fan-out leads to longer delays.
   b. We would like to compute the fastest clock frequency at which we can execute the LOAD (lw) instruction on this architecture. How do we compute this, assuming the correct delay values of all components are known? Assume the following:
   - The delays for the Register File and Memories are calculated from the time the input signals are ready to the time the operation completes (i.e., they are not edge-triggered).
   - For a Register File, WRITE operation just sends the data to the write port. It is transferred to the register outputs at the next clock edge.
   - For combinational components, "Delay" refers to the maximum delay from any input to any output. For individual registers (such as PC), "Delay" refers to the time required from the rising clock edge to the output Q stabilising.



[Picture credit: David A. Patterson and John L. Hennessy, Computer Organization and Design: The Hardware/Software Interface, Morgan Kaufmann Publisher]