**Homework 4**

**Requirement：**

Please type your answers in this .docx file. Do **not** submit handwriting documents.

1. (20 points) Problems in this exercise refer to the following sequence of instructions, and assume that it is executed on a five-stage pipelined datapath:

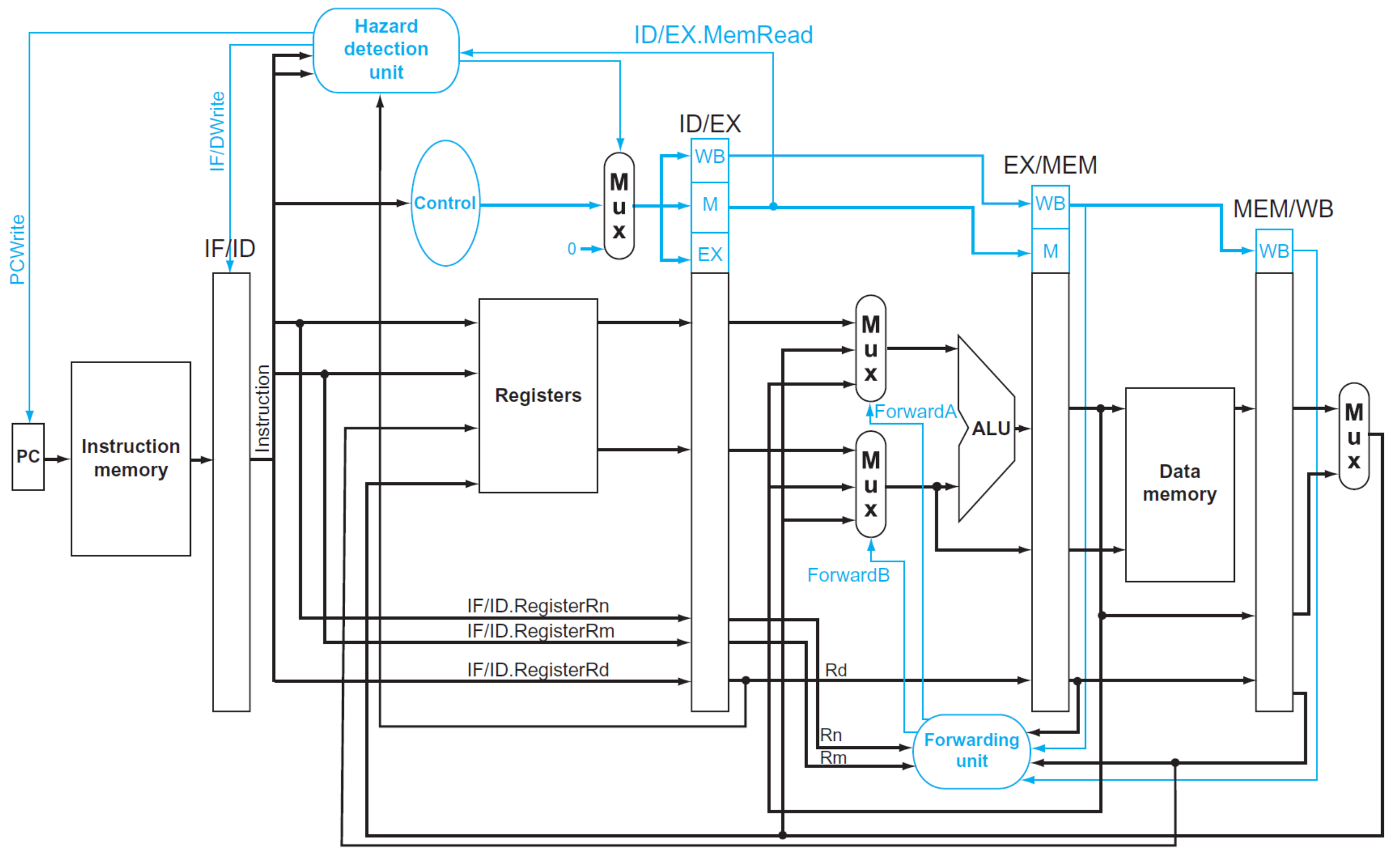
ADD X5, X2, X1

LDUR X3, [X5, #4]

LDUR X2, [X2, #0]

ORR X3, X5, X3

STUR X3, [X5, #0]



1. If there is no forwarding or hazard detection, insert NOPs (stalls) to ensure correct execution.
2. Is it possible to reduce the number of NOPs by rearranging the sequence of the code?
3. If there are forwarding and hazard detection units, for the first seven cycles during the execution of this code, specify which signals (including ForwardA, ForwardB, PCWrite, and IF/IDWrite) are asserted in each cycle by hazard detection and forwarding units in the above figure.
4. If the processor has forwarding, but we forgot to implement the hazard detection unit, what happens when the original code executes?
5. (26 points) For a direct-mapped cache design with a 64-bit address, the following bits of the address are used to access the cache.

|  |  |  |
| --- | --- | --- |
| Tag | Index | Offset |
| 63-10 | 9-5 | 4-0 |

1. What is the cache block size (in words)?
2. How many blocks does the cache have?
3. What is the ratio between total bits required for such a cache implementation over the data storage bits?
4. (24 points) Cache block size (B) can affect both miss rate and miss latency. Assuming a

machine with a base CPI of 1, find the block size that minimizes the total miss latency

given the following miss rates for various block sizes.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Block Sizes | 8 | 16 | 32 | 64 |
| Miss Rate | 4% | 3% | 2% | 1.5% |

(1) What is the optimal block size for a miss latency of 20 × B cycles?

(2) What is the optimal block size for a miss latency of 24 + B cycles?

(3) For constant miss latency, what is the optimal block size?

1. (30 points) Multilevel caching is an important technique to overcome the limited amount of space that a first-level cache can provide while still maintaining its speed. Consider a processor with the following parameters:

Basic CPI (with no memory stalls) is 1.5, processor speed is 2GHz, main memory access time is 100ns, first-level cache miss rate per instruction is 7%, second-level cache (direct mapped) access time is 12 cycles, miss rate of the second-level cache (direct mapped) is 3.5%, second-level cache (eight-way set associative) access time is 28 cycles, miss rate of the second-level cache (eight-way set associative) is 1.5%

(1) Calculate the CPI for the processor using:

1) only a first-level cache,

2) a first-level cache and a second-level direct-mapped cache, and

3) a first-level cache and a second-level eight-way set associative cache.

(2) How do the results in question “(1)” change if main memory access time doubles? (Give each change as both an absolute CPI and a percent change.)