

Please put all answers in the examination paper.

"On my honor, I pledge that I have not violated the provisions of the NJIT Academic Integrity Code."

Signature: [Signature]
Printed Name: SAGAR SOOD

1. (5) In a pipelined architecture why should branch instructions be minimized in writing programs?

This is because in a branch instruction, the first instruction decodes the target address, and this address is used to fetch by the following instruction. Hence, the branch instruction stops at step 3 in a pipeline.

2. (5) Given the two instruction sequence below in a pipelined architecture (5 stages as the book example) what must be done to ensure that the result after execution is the same as a non-pipelined architecture?

FDC RL

AND R2, R4, R6
ADD R5, R2, R4

"No-Operation" instructions should be placed between the two instructions so that the pipeline can avoid delays caused due to data-dependence in the two instructions.

3. (5) Direct Memory Access (DMA) is used with high speed I/O devices. Why is this an advantage as opposed to programmed and interrupt I/O?

Direct Memory Access is an advantage with high speed I/O because the DMA controller would handle the processes of getting blocks of data instead of the processor itself. Programmed and interrupt I/O would run on the processor which can cause delays and inefficiency due to interrupt signals. Direct Memory would ensure that the control signals and addresses are managed and recorded correctly. Although it would not run on the processor, the DMA would be managed and tracked by the processor, furthermore making the process efficient and fast.

- 32
4. (5) How is it determined which DMA device has access to the computer bus?

The Direct Memory Access (DMA) controller decides which device has access to the computer bus.

arbitration
(centralized
decentralized)

5. (5) Define hit rate in relation to cache.

A successful hit when the processor tries to access a word/block of data in the cache and finds it. The hit-rate is defined as a fraction of the successful hits over total attempted accesses. For example, a 9 out of 10 hits would be stated as 0.9 or 90% hit-rate.

6. (5) In the ARM assembly language what is the difference during execution of the branch (b) and branch and link (bl) instructions?

The branch instruction compares the value of two registers, ^{with the given conditions} and depending on the result, provides a target to change the sequence of execution in the program. A branch and link instruction does the same but instead of providing a target address, the instruction links the instruction to another instruction in the program.

b
goto
cannot go back

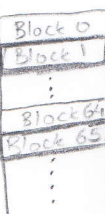
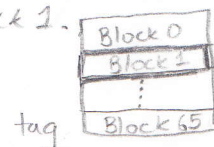
bl
goto
can go back (subroutine/
method/
function)

7. (5) Semiconductor memory was divided into static and dynamic types. Discuss one defining characteristic between them.

Static memory types are faster and consume less power compared to dynamic types. They utilize the power to retain the state of the data. On the contrary, dynamic types consume less power but require periodic refreshing of the memory to update the data and values.

8. (5) In a direct mapped cache how is it determined where the memory block goes in the cache?

In a direct mapped cache, the block of data from the memory is loaded onto the block of cache. The next block of data from the memory is then loaded onto the same block of cache (starting from Block 0). Each block is tagged with the address of the block of memory. For example, Block 0, 64, 128 would be stored in Block 0. Blocks 1, 65, 129 would then be stored in Block 1.



9. (10) You have available 256K x 2 bit memory integrated circuits. You need a 2M x 32 bit memory system. How many address lines do you need for the memory system? How many address lines are connected the decoder?

$$256K \times 2 = 256K \times 2^1 = 1 \text{ address lines}$$

$$2048K \times 32 = 2048K \times 2^5 = 5 \text{ address lines}$$

Therefore we need 4 address lines

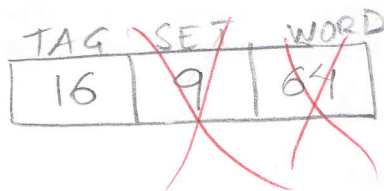
$$256 = 2^8 = 8 \text{ address lines connected to decoder}$$

10. (5) The memory was divided into a hierarchy. Which level of the hierarchy has the slowest (longest access) time?

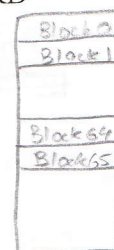
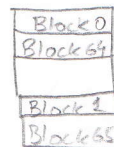
The memory at the bottom in terms of speed would be magnetic disks, example hard-drives. They take the longest access time but have an upperhand in terms of storage capacity.

Cache
↑
Mem
↑
Hard

11. (10) For the memory system which has 8M words of main memory, a cache of 256K words where the cache is organized as a 4 way set associative cache with a block size of 64 words, how many bits are there in the TAG, SET, and WORD fields?



tag
tag



12. (5) In a naïve implementation of virtual memory, the page table is maintained only in memory which necessitates two memory accesses for each read. How is this improved?

Since the page table is in the memory, this slows down the memory accesses. This can be improved by having the page tables be copied in the cache memory so the access can be made faster. This would reduce the clock cycles for access and avoid delays in memory.

13. (5) How are the control signals necessary for the BPU to operate generated?

The control signal help the processes in the BPU to occur in a sequential manner rather than having a chaos of the flow of instruction. They are generated in 2 ways - "hardwired signals" that use the finite-state mechanism (FSM) to control and track step counters, contents of IR, results of ALU and external signals. The other method is microprogrammed signals that use control storage to send signals to words and data in the processes, which is read from a ROM source.

14. (5) If the requirement is to send data over a long distance what type of external input/output is used?

External input/output such as Ethernet would allow a fast and effective method of sending data.

OUTSIDE of comp → SERIAL
INSIDE " " → PARALLEL

15. (5) When accessing memory how does the BPU know when the data is available?

The use of mapping techniques between the cache and main memory allows the BPU to track the presence of data in the memory. As cache memory is faster to access, and holds the copy of instructions and addresses from the main memory, the BPU can access this quickly and know if the data is available in memory.

16. (5) In an associative cache memory how do we determine if a cache line in the cache corresponds to the memory address presented?

In the associative cache memory, a block of data from the main memory in the cache memory is tagged with the respective address of the block in the main memory. This way the blocks are organized efficiently in the cache and can easily determine if the data exists in the memory.

17. (5) Define access time for memory.

The time it takes for the BPU to access the data in the memory is called "access time for memory". This is usually represented in terms of clock cycles and requires more than one clock cycle.

18. (5) What function does the following ARM assembly language perform? The C language driver program is:

```
extern int exam2( char str[] );
```

entry
 ↓
 exam2: stmfd sp!, {v1-v6, lr}
 mov v1, #0 *pointer moves immediate value 0 to variable v1.*
 lop: ldrb v2, [a1], #1 *perform a loop and stores the data into v2.*
 add v1, v1, #1 *increments variable v1 by 1.*
 cmp v2, #0 *compare v2 with 0.*
 bne lop *branch loop*
 mov a1, v1 *moves variable v1 to register a1.*
 ldmfd sp!, {v1-v6, pc} *exit*

load →
is v2=0 →
if not go back →

The function loads and stores the values of an array into variables. The function starts with first variable and stores the value from the array, then the counter is incremented. The loop continues until the array has no more values.

strlen program