Name   **Mehran Ali Shah**

Id No   **13943**

Subject   **Computer Architecture**

Date   **21-08-2020**

Question No 1:

Part a:

Answer:

Pipelining is when computers receives multiple instructions and carry them out as they're received. Branch prediction is the process of being able to predict the next set of instructions so that they can be carried out. Superscalar execution is when you're able to give more than one set of instructions at time. Data flow analysis analyzes instructions that need each other Speculative execution carry out instructions before they are actually executed.

Part b:

Answer:

The speedup using a parallel processor with N processors that fully exploits the parallel portion of the program is as follows:
Speedup =Time to execute program on a single processor/Time to execute program on N parallel   processors
$= T(1 - f) + Tf/T(1 - f) + Tf/N = 1/(1 - f) + f/N$

Part c:

Answer:

Physical:

Consists of the actual wires carrying the signals, as well as circuitry and logic to support ancillary features required in the transmission and reception of the 1s and 0s.

Link:

Responsible for reliable transmission and flow control.

Routing:

Provides the framework for directing packets through the fabric.

Protocol:

The high-level set of rules for exchanging packets of data between devices. A packet is comprised of an integral number of Flits.
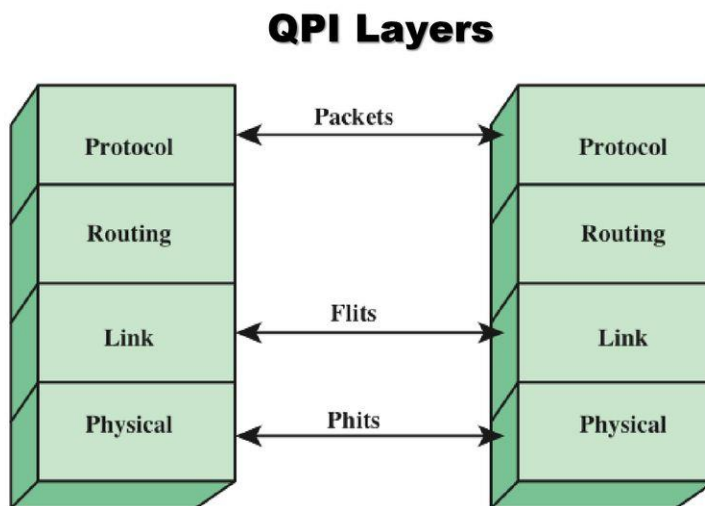
QPI Layers

Protocol	Packets	Protocol
Routing		Routing
Link	Flits	Link
Physical	Phits	Physical

Figure 3.21 QPI Layers

Part d:

Answer:

A root complex device, also referred to as a chipset or a host bridge, connects the processor and memory subsystem to the PCI Express switch fabric comprising one or more PCIe and PCIe switch devices. The root complex acts as a buffering device, to deal with difference in data rates between I/O controllers and memory and processor components. The root complex also translates between PCIe transaction formats and the processor and memory signal and control requirements. The chipset will typically support multiple PCIe ports, some of which attach directly to a PCIe device, and one or more that attach to a switch that manages multiple PCIe streams. PCIe links from the chipset may attach to the following kinds of devices that implement PCIe:

Switch:

The switch manages multiple PCIe streams.

PCIe Endpoint:

An I/O device or controller that implements PCIe, such as a Gigabit ethernet switch, a graphics or video controller, disk interface, or a communications controller.

Legacy endpoint:

Legacy endpoint category is intended for existing designs that have been migrated to PCI Express, and it allows legacy behaviors such as use of I/O space and locked transactions. PCI Express endpoints are not permitted to require the use of I/O space at runtime and must not use locked transactions. By distinguishing these categories, it is possible for a system designer to restrict or eliminate legacy behaviors that have negative impacts on system performance and robustness.

PCIe/PCI bridge:

Allows older PCI devices to be connected to PCIe-based systems. As with QPI, PCIe interactions are defined using a protocol architecture. The PCIe protocol architecture encompasses the following layers.

Physical:

Consists of the actual wires carrying the signals, as well as circuitry and logic to support ancillary features required in the transmission and receipt of the 1s and 0s.

Data link:

Is responsible for reliable transmission and flow control. Data packets generated and consumed by the DLL are called Data Link Layer Packets (DLLPs).

Transaction:

Generates and consumes data packets used to implement load/store data transfer mechanisms and also manages the flow control of those packets between the two components on a link. Data packets generated and consumed by the TL are called Transaction Layer Packets (TLPs).
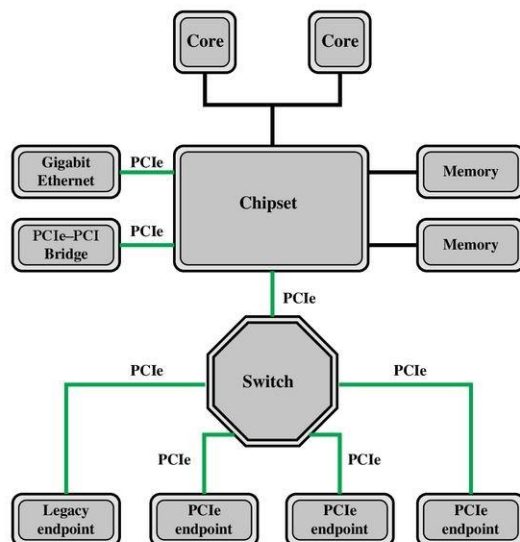
Question No 2:

Part a:

Answer:

Moore's Law:

Moore observed that the number of transistors that could be put on a single chip was doubling every year, and correctly predicted that this pace would continue into the near future. To the surprise of many, including Moore, the pace continued year after year and decade after decade

The consequences of Moore's law are profound:

1. The cost of a chip has remained virtually unchanged during this period of rapid growth in density. This means that the cost of computer logic and memory circuitry has fallen at a dramatic rate.
2. Because logic and memory elements are placed closer together on more densely packed chips, the electrical path length is shortened, increasing operating speed.
3. The computer becomes smaller, making it more convenient to place in a variety of environments.
4. There is a reduction in power requirements.
5. The interconnections on the integrated circuit are much more reliable than solder connections. With more circuitry on each chip, there are fewer inter-chip connections.


Part c:

Answer:

A program residing in the memory unit of a computer consists of a sequence of instructions. These instructions are executed by the processor by going through a cycle for each instruction.

In a basic computer, each instruction cycle consists of the following phases:

1. Fetch instruction from memory.
2. Decode the instruction.
3. Read the effective address from memory.
4. Execute the instruction.
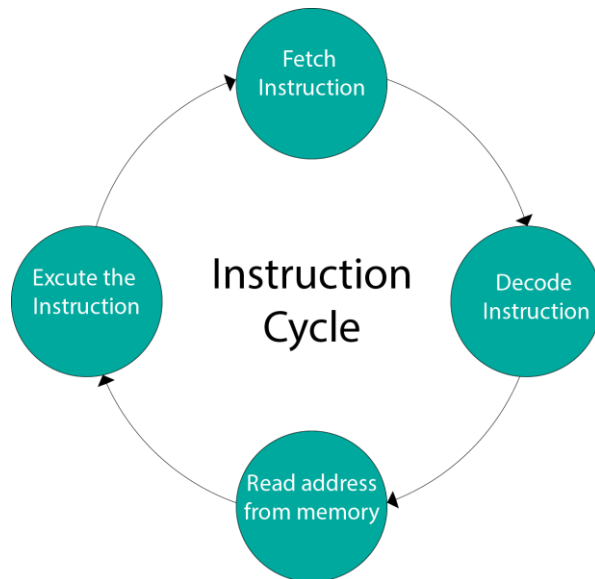
Part d:

Answer:

An interrupt is a signal to the processor that is emitted by one of the three classes of interrupts indicating an event needs to be handled. There are three interrupt classes when referring to computer architecture, interrupts caused by: hardware failure, external events, or executed instructions.

Hardware Failures

This class of interrupts are caused by power outages or memory parity errors. A lot of the time these are out of your hands, but you need to know how to properly handle a hardware interrupt.

External Events

A "soft" or "hard" reset are generally known to fall in this class. In addition, I/O devices can cause these types of interrupts.

Executed Instructions

Exceptions and Systems call fall into this interrupt class. Think address errors, reserved instructions, integer overflow, floating point errors.

Question No 3:

Part b:

Answer:

Cortex-A:

The cortex A are application processor intended for mobile devices such as smartphone, digital TV etc.

Cortex-R:

The cortex R is designed to support real-time application in which the time of event needs to be constable to with rapid response to events. They run as a high frequency.

CarMax-M:

Cortex M sense processor has been developed primarily for the microcontroller domain where the need for fast high deterministic interrupt. Management is coupled with the desire for extremely possible power consumption.

Part b:

Answer:

Multicore:

The use of multiple processors on the same chip provides the potential to increase performance without increasing the clock rate. Strategy is to use two simpler processors on the chip rather than one more complex processor. With two processors larger caches are justified. As caches became larger it made performance sense to create two and then three levels of cache on a chip.

MIC:

Leap in performance as well as the challenges in developing software to exploit such a large number of cores. The multicore and MIC strategy involves a homogeneous collection of general purpose processors on a single chip.

GPUs:

Core designed to perform parallel operations on graphics data. Traditionally found on a plug-in graphics card, it is used to encode and render 2D and 3D graphics as well as process video. Used as vector processors for a variety of applications that require repetitive computations.

Part c:

Answer:

Disable Interrupt:

Simply means that the processor can end will ignore that interrupt request signal.

Nested Interrupt:

It allows an interrupt of higher priority to cases to lower priority interrupt handler to be itself interrupt.

Question No 4:

Part a:

Answer:

ISU (instruction sequence unit):

Determines the sequence in which instructions are executed in what is referred to as a superscalar architecture.

IFU (instruction fetch unit):

Logic for fetching instructions. IDU (instruction decode unit): The IDU is fed from the IFU buffers, and is responsible for the parsing and decoding of all z/Architecture operation codes.

LSU (load-store unit):

The LSU contains the 96-kB L1 data cache, 1 and manages data traffic between the L2 data cache and the functional execution units. It is responsible for handling all types of operand accesses of all lengths, modes, and formats as defined in the z/Architecture.

XU (translation unit):

This unit translates logical addresses from instructions into physical addresses in main memory. The XU also contains a translation lookaside buffer (TLB) used to speed up memory access.

FXU (fixed-point unit):

The FXU executes fixed-point arithmetic operations.

BFU (binary floating- point unit):

The BFU handles all binary and hexadecimal floating-point operations, as well as fixed- point multiplication operations.

DFU (decimal floating-point unit):

The DFU handles both fixed-point and floating-point operations on numbers that are stored as decimal digits.

RU (recovery unit):

The RU keeps a copy of the complete state of the system that includes all registers, collects hardware fault signals, and manages the hardware recovery actions.

COP (dedicated co-processor):

The COP is responsible for data compression and encryption functions for each core.

I- cache:

This is a 64-kB L1 instruction cache, allowing the IFU to pre fetch instructions before they are needed.

L2 control:

This is the control logic that manages the traffic through the two L2 caches.

Data- L2:

A 1-MB L2 data cache for all memory traffic other than instructions.

Instr-L2:

A 1-MB L2 instruction cache.

Question No 5:

Answer:

Calculating CPI:

The formula of calculating CPI is:

CPI= _____ Instruction count * cycles per second _____

Number of instruction the executed program consists

$$= (45000 * 1) + (23000 * 2) + (15000 * 2) + (8000 * 2) / 100000$$

$$=155000 / 100000$$

$$CPI = 1.55$$

Calculating MIPS:

The million instruction per second (MIPS) rate can be calculated with following constraints.

Processor time, $T = I_c * CPI * \tau$

"τ" represents constant cycle time can be calculated as $1/f$.

$MIPS = I_c / T * 10^6$

$$= I_c / I_c * CPI * \tau * 10^6$$

$$= f / CPI * 10^6$$

$MIPS = 40 * 10^6 / 1.55 * 10^6$

$MIPS = 25.8$

Calculation of T:

$T = 100000 * 1.55 / 40 * 10^6$

$$= 155000 / 40000000$$

$$= 0.003875$$

$$= 3.875 \text{ ms}$$