COMPUTER ARCHITECTURE

Submitted To: Prof. Amin

Submitted By: Awais Khan

ID: 6858

Dated: 26/06/2020

Q.1 Give answers to each of the following:

Discuss the concept of word, addressable units, and unit of transfer for internal memories.

In computing, a word is the normal data unit used in the design of a specific processor. A word is a fixed-sized piece of data which the instruction set or the processor's hardware handles as a unit. An important characteristic of any particular processor design or computer architecture is the number of bits in a word (the word size, word width, or word length).

How least recently used (LRU) and least frequently used (LFU) replacement algorithms are implemented for a cache memory with two-way set associative mapping?

Least Recently User (LRU):

We are given total number of possible pages which can be referenced. We are also given cache (or memory) size (Number of page frames that cache can hold at a time). The LRU caching scheme is programmed to delete the least recently used frame when the cache is full and when a new page is accessed that is not in cache.

We use two data structures to implement an LRU Cache.

1. **Queue** that is implemented using a list of double ties. The maximum queue size would be equal to the total number of available frames (cache size). The most recent pages used will be near the front end and the least recent pages will be close to the rear end.
2. A **Hash** with page number as key and the corresponding queue node address as tag.

The requisite page may be in the memory when a page is referenced. If it's in the brain, then we need to remove the list node and put it to the queue front. If the page you need isn't in memory, we'll put it in memory. We add a new node to the front of the queue in simple words, and update the corresponding node address in the hash. When the queue is full, i.e. all frames are full, we delete a node from the queue 's rear, and connect the current node to the queue's front.

Least Frequently Used (LFU):

Least Frequently Used is a caching algorithm which removes the least frequently used cache block when the cache is overflowing. In LFU we test the old page as well as the page's frequency and if the page 's frequency is higher than the old page we can't delete it and if all the old pages have the same frequency then use the last i.e. FIFO option to delete the page.

Min-heap data structure is a good choice for applying this algorithm, as it manages logarithmic time complexity addition, deletion and update. Deleting the least recently used cache block can resolve a tie. To solve the problem the following two containers were used:

1. The cache was represented by a matrix of integer pairs, where each pair consists of the block number and the number of times it was used. The vector is ordered in the form of a min-heap enabling us to access in constant time the least frequently used block.
2. A HashMap was used to store the cache block indices that enables searching in constant time.

How read and write operations are performed in SRAM cell?

Reading Data from SRAM Cell:

Theoretically, reading only involves claiming the WL word line and reading the SRAM cell state by a single access transistor and bit line, e.g. M6, BL. However, bit lines are fairly long and have a broad parasitic power. In action, a more complicated method is used to speed up reading: the read cycle begins by preloading both BL and BL bit lines to high (logic 1) voltage. Then claiming the term line WL allows both the M5 and M6 control transistors, which causes the BL voltage of one bit to drop slightly. Then the BL and BL lines should have a minor voltage difference. The sensor amplifier can detect which line has a higher voltage and thus decide if 1 or 0 has been deposited. The higher the sensitivity of the sensor amplifier, the faster the read process. As the NMOS is stronger, it's easier to pull down. As a result, the bit lines are traditionally preloaded to high voltage. Several researchers are now attempting to pre-charge at a slightly low voltage to reduce power consumption.

Writing Data to SRAM Cell:

The writing process starts by adding the value to be written to the lines of the piece. If we want to write a 0, we will add a 0 to the bit row, i.e. set BL to 1 and BL to 0. This is equivalent to applying a reset pulse to the SR-latch, which causes the flip-flop to alter the state. A 1 is written by inverting the values of the lines of the row. WL is then asserted, and the value to be stored is latched in. It works because the bit line input-drivers are engineered to be much stronger than the comparatively weak transistors in the cell itself so that they can easily bypass the previous cross-coupled inverters.

For action, exposure to NMOS transistors M5 and M6 must be greater than either the bottom NMOS (M1, M3) or the top PMOS (M2, M4) transistors. It is easily accomplished as PMOS transistors are much weaker than NMOS in the same amount. As a result, when one transistor pair (e.g. M3 and M4) is only slightly overridden by the write process, the opposite transistor pair (M1 and M2) of the gate voltage is also changed. This means that the M1 and M2 transistors can be quickly overridden, and so on. Cross-coupled inverters thus magnify the writing process.

Discuss 16-Mbit DRAM (4M x 4) organization in detail.

DRAM is normally arranged in a rectangular array of load storage cells consisting of one capacitor and one transistor per data bit. The figure on the right provides a basic example of a four-by-four cell matrix. Many DRAM matrixes are thousands of cells in height and width.

Long horizontal lines connecting to each row are known as word-lines. Every cell column consists of two bit-lines, every connected to each other storage cell in the column (this essential detail is not shown in the diagram to the right). These are commonly referred to as "+" and "−" bit lines.

A sense amplifier is basically a pair of cross-connected inverters between the bit-lines. The first inverter is connected to the input from the + bit-line and the output to the − bit-line. The second inverter input is from the − bit-line output to the + bit-line input. It results in positive feedback, which stabilizes when one bit-line is completely at its peak voltage and the other bit-line is at the lowest possible voltage.

What are the reasons for DVD's greater capacity over CD?

DVD uses a red laser instead of an infrared laser. Red light is shorter in wavelength than infrared light. So, consequently Partly because of the aforementioned, the DVD has a smaller numerical aperture than the CD does. This helps you to

Shorter pits, The individual pits and lands that represent 1s and 0s (which use non-return-to-zero encoding) in the bitstream that represents the data on the disk are physically shorter and narrower than on the CD, allowing more of them to be compressed into the same physical space (which is

referred to as the area density). The actual spiral data track on the disk is more tightly wound than the CD, i.e. it has a narrower pitch than the data track(s) on the CD. It also raises the area density-allowing more data to be crammed into the same physical space.

Q.2 Differentiate each of the following in detail:

a) EEPROM and flash memory

EEPROM	**flash memory**
EEPROM is a type of data memory device that uses an electronic device to delete or write digital data. It has the ability to erase-and-write byte, which makes it slow. Flash memory is a distinct form of EEPROM that is programmed and deleted in large blocks. Nonetheless, the trend seems to be to use AND flash for devices that only allow large-block erasing. Flash memory has a lot of features. It is much cheaper than EEPROM and does not need solid-state storage batteries such as static RAM.	Flash memory is a type of programmable read-only memory (EEPROM), but it can also be a stand-alone storage device, such as a USB drive.
EEPROM uses NOR type.	This is a non-volatile memory chip used for storing and transmitting data between a PC and other digital devices. This is also used in USB flash drives, MP3 players, digital cameras and solid-state drives.
EEPROM is byte-wise erasable.	Flash memory includes the use of floating-gate transistors to store data.

Soft Errors	**Hard Failures**
In electronics and computation, a soft error is a type of error where a signal or data is incorrect. Errors can be caused by a fault that is generally known to be either a design or construction flaw or a broken part. A soft error is also a signal or datum which is wrong, but is not assumed to imply such a mistake or breakage. After observing a soft error, there is no implication that the system is any less reliable than before. One cause of soft errors is single event upsets from cosmic rays.	Hard failures can be caused by excess temperature, excess current or voltage, ionizing radiation, mechanical shock, stress or effects, and many other causes. For semiconductor devices, problems in the product package can cause failure due to contamination, mechanical stress of the system, or open or short circuits.

c) Read and write Mechanisms for magnetic disk

Data Reading	Data Writing
The magnetic disk drive consists of a set of disks (disks) covered with magnetic material. They 're spinning at about 7200 rpm. The data is translated into bits and written to the surface as a sequence of shifts in the direction of magnetization. The data is read by detecting the changes in direction.	A hard drive, heads 'float' over the disk surface with a clearance of as little as 3 nanometers. The "flying height" is continuously rising to allow for higher area density. The flight height of the head is controlled by an air-bearing mechanism engraved on the disk-facing surface of the slider. The function of the air bearing is to keep the flying height constant as the head travels over the surface of the disk. When the head hits the surface of the disk, a catastrophic head crash can result.

d) Parallel access and independent access RAID schemes

Parallel access technique	Independent access technique
RAID Levels 2 and 3 use a parallel control strategy. All member disks participate in the execution of each I / O request in a parallel access sequence. Usually, the spindles of the individual drives are coordinated such that each disk head is at the same location on each disk at any given time.	RAID Levels 4 through 6 use an independent access technique. Each member disk operates independently in an independent access array, so that separate I / O requests can be satisfied in parallel. As a result, independent access arrays are more suitable for applications requiring high I / O request rates and are comparatively less suited for applications requiring high data transfer rates.
Like for other RAID systems, data stripping is used. In the case of RAID 2 and 3, strips are very small, sometimes as short as a single byte or a single word. For RAID 2, the error correction code is measured through the corresponding bits on each data disk, and the bits of the code are stored in the corresponding bit positions on multiple parity disks. Usually, a Hamming code is used to correct single-bit errors and to prevent double-bit errors.	Like for other RAID systems, data stripping is used. The strips are fairly large in the case of RAID 4 through 6. With RAID 4, a bit-by - bit parity strip is calculated across the corresponding stripes on each data disk and the parity bits are stored in the corresponding strip on the parity disk.

e) HD DVD and Blu-ray DVD

Blu-Ray has 25 GB capacity (50 GB for dual-layer) and is more expensive.	HD-DVD has 15 GB (30GB for dual layer) and is cheaper than Blu-Ray.
Laser wavelength 405 nm (blue-violet laser)	Laser wavelength 405 nm (blue-violet laser)
Expenisve machine to read data from disk	Cheaper machine to read data from disk

Q.3 Write note on each of the following:

a) Memory access methods

1 Sequential Access:-

In this process, the memory is accessed in a common linear sequential way, such as in a single Linked List. The time of access depends on the location of the data. The applications of this sequential memory access are magnetic tapes, magnetic disk and optical memory.

Random Access: -

In this method, any location of the memory can be accessed at random, just like accessing the Array. Specific locations are independent of this access process.

Applications of this random memory access are RAM and ROM.

4 Direct Access: -

In this process, the unique location of the memory can be reached directly by accessing the array. This method is a combination of the two access methods described above. Access time depends on both memory structure and storage technology characteristics. Access is either semi-random or direct.

The application of direct memory access is a magnetic hard disk, read / write header.

Associate Access:-

A word is accessed in this memory rather than its address. This access method is a special type of random-access method.

Cache memory is the use of direct memory access.

b) Principle of locality

Reference locality refers to a phenomenon in which a computer program tends to have access to the same set of memory locations for a specific time period. In other words, the Reference Locality refers to the tendency of a computer program to access instructions whose addresses are close to each other.

c) Possible approaches to cache coherency

Write Through (WT) Protocol

Once a processor writes a new value to its cache, the new value is also written to the memory module that holds the cache block updated. Some copies of this block may exist in other caches, and these copies must be updated to reflect the changes caused by the write operation.

We update other cache copies by sending up-to - date data to all processor modules on the system. That processor module receives broadcast data, updates the contents of the affected cache block if this block is present in its cache.

Write Back (WB) Protocol

Multiple copies of a cache block can exist in the WB protocol if different processors have loaded (read) the block into their caches.

In this approach, if a processor wants to change this block, it must first become the exclusive owner of the block.

When this processor is granted ownership, the memory module is the home location of the device. Any other copies, including one in the memory module, would be invalidated.

The owner of the block will now modify the memory contents.

If another processor decides to read this file, the data will be sent to this processor by the current owner. Data is also sent to the home memory node, which includes ownership and updates of the block to contain the latest value.

d) Practical Issues peculiar to SSDs

There are two practical issues peculiar to SSDs that are not faced by HDDs

SSD performance has a tendency to slow down as the device is used

- The entire block must be read from the flash memory and placed in a RAM buffer
- Before the block can be written back to flash memory, the entire block of flash memory must be erased
- The entire block from the buffer is now written back to the flash memory

Flash memory becomes unusable after a certain number of writes

- Techniques for prolonging life:
    - Front-ending the flash with a cache to delay and group write operations
    - Using wear-leveling algorithms that evenly distribute writes across block of cells
    - Bad-block management technique
- Most flash devices estimate their own remaining lifetimes so systems can anticipate failure and take preemptive action

a) CD read and write operation

Normal CDs cannot be modified — they are read-only devices. The CD-R disk needs to allow the drive to write data to the disk. In order for the CD-R to operate, there must be a way for the laser to create a non-reflective region on the disk. Therefore, the CD-R disk has an extra layer that the laser can change. This additional coating is a greenish tint. In a normal CD, you have a plastic substrate covered with a reflective layer of aluminum or gold. You have a plastic substrate, a dye layer, and a reflective gold layer in a CD-R. On a new CD-R disk, the entire surface of the disk is reflective — the laser will shine through the dye and reflect the golden layer.

When writing data to a CD-R, the writing laser (which is much more powerful than the reading laser) heats up the dye layer and changes its transparency. The alteration in the hue produces the appearance of a non-reflective hump. This is a permanent change, and both CD and CD-R drives can read the modified dye as a bump later.

It turns out that the dye is quite sensitive to light — it has to be so that it can be quickly modified by the laser. Therefore, you want to stop exposing CD-R disks to sunlight.