| | |
|---|---|
| In a PC, there tend to be two chip-sets: the northbridge; and the southbridge. What is the essential difference between the two? [2] | What are the two main components of an integrated circuit? (Both of these components usually appear in their millions.) [1] |
| What is the current trend in microprocessor design so as to overcome difficulties with power dissipation as we build faster and faster single processors? [2] | What is Moore's Law? [1] |
| What is a Boolean function? In relation to Boolean functions, what is so special about NOT-, AND- and OR-gates? [2] | Explain the basic principle behind using half-adders and full-adders to build a circuit that computes the sum of two n-bit strings of 0s and 1s. (You need not describe the resulting circuit in full detail; just give an overview of its construction.) [6] |
| What is the ultimate aim of research in formal methods? [1] | A hardware description language is used so as to better enable the design of integrated circuits. Name two distinctive attributes of computer hardware that are normally expressible in a hardware description language. [2] |
| What is the von Neumann bottleneck? Which component of a modern CPU did the von Neumann bottleneck give rise to? [3] | How does the Harvard architecture differ from the von Neumann architecture? [2] |

| | |
|---|---|
| Transistors [$\frac{1}{2}$] interconnected by microscopic wires [$\frac{1}{2}$]. | The northbridge handles fast communications amongst the CPU, the memory and the graphics processing unit [1], whereas the southbridge handles (slower) communications involving external hard-disks, the mouse, the keyboard, the Internet, the printer and other such devices [1]. |
| A rough description that long-term transistor capacity doubles every 18–24 months (coined by Gordon Moore in 1965) [1]. | Multi-core processors [1] where one CPU with a high clock-speed is replaced with a number of CPUs with lower clock-speeds but which, when working together, can give better computational power [1]. |
| Let the two n-bit strings be $X_1X_2...X_n$ and $Y_1Y_2...Y_n$. A half-adder takes $X_1$ and $Y_1$ as input and outputs the sum, with the carry fed into a full-adder [2]. This full-adder also has inputs $X_2$ and $Y_2$ and its sum is output, with its carry fed into a full-adder [2]. This full-adder also has $X_3$ and $Y_3$ as inputs and its sum is output, with its carry fed into a full-adder, and so on [2]. Both the sum and the carry of the final full-adder are output. | A Boolean function is a function $f : \{0,1\}^n \to \{0,1\}$ [1]. We can build a circuit to compute any Boolean function using only NOT-, AND- and OR-gates [1]. |
| Unlike a normal programming language, an HDL includes explicit notations for expressing **time** [1] and **concurrency** [1], which are primary attributes of computer hardware. | The ultimate aim of formal methods is to enable us to mathematically prove properties of designs, programs and so on, so as not just to rely on empirical testing [1]. |
| The Harvard architecture has memory that is partitioned into data memory and instruction memory with dedicated buses for each of them [2]. | The von Neumann bottleneck is a limitation of the rate of data transfer between the CPU and memory (data and instructions have to be fetched in sequential order and idle time is wasted whilst waiting for data items or instructions to be fetched from memory) [2]. The von Neumann bottleneck gave rise to the use of caches [1]. |

| | |
|---|---|
| Explain the difference between a byte and a word in relation to computer architecture. How is the size of a word related to the width of a bus? [2] | What is a basic principle as regards different types of memory and its physical distance from the CPU? [1] |
| What is the difference between static RAM and dynamic RAM? [3] | What is the purpose of cache memory in a CPU? [1] |
| What are registers in the CPU? [1] | Modern parallel computing can come in many shapes and forms. Briefly explain the fundamental difference between multi-core computing and GPGPU computing. [2] |
| Give two illustrations of principles of Computational Thinking in the context of hardware. [2] | |
| | |

| | |
|---|---|
| The cost and performance of memory is generally proportional to its physical distance from the CPU [1]. | 1 byte of storage consists of 8 bits [1]. A word, being a number of bytes, is the amount of data that can be handled by a processor as one unit [1]. |
| Caches are expensive memory that are used to store rapidly accessed items [1]. | Dynamic RAM (DRAM) is where a bit of data is stored using a transistor/capacitor combination [1]. Static RAM (SRAM) is where a bit of data is stored by a flip-flop, which incorporates 4-6 transistors [1]. Static RAM is stable and fast but takes up more memory [$\frac{1}{2}$] whereas dynamic RAM is cheap, slow and needs to be refreshed because of 'leaky' capacitors [$\frac{1}{2}$]. |
| A multi-core computer is a computer where more than one processor is integrated on one IC [1]. GPGPU computing uses the processors within the graphics processing unit to compute with rather than deal with the screen pixels [1]. | Registers are on-chip memory locations that are limited in number. They provide the fastest way to access data [1]. |
| | 'Processing in parallel' in multi-core or GPGPU computing. 'Using abstraction and decomposition in tackling a large complex task' in IC design and computer system design. 'Prefetching and caching in anticipation of future use' in caching within CPUs. 'Interpreting code as data and data as code' in processor architectures. [1] for each illustration. |
| | |