

THEME A: CONCEPTS · A1.1

Computer hardware and operation — Mark scheme

35 marks · covers A1.1.1–A1.1.9

How to mark. Each creditworthy point is on its own line and ends with a semicolon (;). A line beginning with ... is an *expansion* of the line above it, so the pair is worth the marks shown. A slash (/) separates acceptable alternative wordings. Words in (brackets) are not needed for the mark. Mark positively: credit what the student has got right rather than penalising what is missing, and accept any valid alternative not listed here. Do not award more than the maximum for a part. Where a later part uses a wrong answer from an earlier part correctly, award follow-through marks (FT). HL marks parts that assess higher level only content.

Q	ANSWER	MARKS
1(a) A1.1.1	Award [1] max (The) program counter (PC);	1
1(b) A1.1.1	Award [2] max Award [1] for the definition and [1] for the expansion. (It) holds/stores the address of the memory location to be read from or written to; ...(so that) this address is placed on the address bus (for memory to locate the data); (It) holds the address the CPU wants to access next; ...(so that) the correct instruction or data can be fetched from / written to primary memory;	2
1(c) A1.1.5	Award [3] max Award [1] per point up to [3] max. The control unit (CU) reads/interprets the bit pattern held in the instruction register; It separates the instruction into the opcode (which operation to perform) and the operand (the data, or the address of the data); It looks up/identifies which operation the opcode represents; It works out whether any further data must be fetched from memory before the instruction can run; It generates the control signals that set up the circuits (for example the ALU) needed to carry the instruction out; Do not credit points that belong to the fetch stage (using the PC, placing an address on the address bus) or to the execute stage.	3
2 A1.1.1 A1.1.5	Award [4] max Award [1] per point up to [4] max. Accept the points in any order. The address of the instruction is copied from the program counter (PC) into the memory address register (MAR); The address bus carries this address from the CPU to primary memory; The address bus is unidirectional / carries addresses in one direction only; The control bus carries the read signal / timing signals telling memory that a read is required; The data bus carries the instruction back from memory to the memory data register (MDR); The data bus is bidirectional / carries data both to and from the CPU; The instruction is then copied from the MDR into the instruction register (IR);	4
3(a)		

Q	ANSWER	MARKS
A1.1.4	Award [2] max Award up to [1] per difference, [2] max. The answer must make the contrast clear. RAM is volatile (loses its contents when power is removed) whereas ROM is non-volatile (keeps its contents); RAM can be read from and written to whereas ROM is read-only (in normal use); RAM holds programs and data currently in use whereas ROM holds the start-up/boot firmware;	2
3(b) A1.1.4	Award [3] max Award [1] per point up to [3] max. Cache is much faster to access than main memory (RAM); It stores copies of frequently or recently used instructions and data; A cache hit means the data is found in cache, so the CPU does not have to wait for the slower RAM; A cache miss means the data must be fetched from RAM, which takes longer; Fewer/shorter waits for memory mean the CPU spends less time idle, so more instructions are executed per second; Levels (L1, L2, L3) trade off speed against capacity, with L1 the fastest and smallest;	3
3(c) A1.1.7	Award [1] max Any one, [1] max. Faster read/write access (no moving parts to wait for); More resistant to physical shock/damage; Silent in operation; Lower power consumption;	1
4(a) A1.1.8	Award [2] max Award up to [1] per difference, [2] max. Lossy compression permanently discards some data whereas lossless compression keeps all of the original data; The original file cannot be exactly restored after lossy compression, whereas it can be restored exactly after lossless compression; Lossy usually achieves a much greater reduction in file size than lossless;	2
4(b) A1.1.8	Award [2] max Award [1] for the point and [1] for the expansion. Subtitles are text, so every character must be recovered exactly; ...(if) characters were discarded the subtitles would be misspelled or unreadable / the meaning would change; Lossless compression allows the file to be restored exactly to the original; ...(so) the subtitle text displayed to the viewer is identical to what was written;	2
4(c) A1.1.8	Award [2] max Award [1] per point up to [2] max. A run of identical consecutive values is replaced by a single code giving the length of the run and the value; For example the sequence AAAAA is stored as 5A, so five items become two; Fewer items need to be stored, so the file takes up less space; It is lossless, because the original run can be rebuilt exactly from the value and the count;	2

Q	ANSWER	MARKS
	It is most effective when the data contains many long runs (for example large areas of one colour);	
5(a) A1.1.9	Award [1] max Infrastructure as a service (IaaS);	1
5(b) A1.1.9	Award [5] max This is a discuss question (AO3), so a balanced response is required. Award [1] per point up to [5] max, but award a maximum of [3] if only advantages or only disadvantages are given. Credit any valid conclusion. Arguments for PaaS No need to buy, house or maintain physical servers, so the up-front cost is much lower; The provider manages the operating system, updates and security patching, so a small team can focus on writing the application; Development tools, databases and deployment pipelines are supplied ready to use, so the application can be built more quickly; Resources can be scaled up or down as demand changes, which suits a start-up whose usage is unpredictable; The team can work on the same hosted environment from anywhere; Arguments against PaaS There is an ongoing subscription cost, which may exceed the cost of owning servers once the company is large and its demand is steady; The company has less control over the environment and cannot configure the underlying hardware or operating system; It depends on the provider's availability, so an outage at the provider takes the booking application offline; Data is stored on a third party's servers, raising privacy, security and jurisdiction concerns; Moving to another provider later can be difficult (vendor lock-in); Conclusion (credit a supported judgement, for example) PaaS is the better choice at this stage because low up-front cost and speed of development matter more to a start-up than long-run unit cost or full control;	5
6(a) A1.1.2	Award [2] max Award [1] per point up to [2] max. A GPU carries out the calculations needed to render graphics/images for display; It contains a very large number of simple cores that work in parallel; This makes it suited to tasks where the same operation is applied to very many data items at once; It is used for graphics rendering in games, and also for machine learning, simulations and other highly parallel computation;	2
HL 6(b) A1.1.3	Award [3] max Award up to [2] per difference (point plus expansion), [3] max. The difference must be linked to why the GPU suits this task. A GPU has thousands of simple cores whereas a CPU has a small number of powerful cores; ... (so) the many independent multiplications in training can be carried out at the same time rather than one after another;	3

Q	ANSWER	MARKS
	A CPU is designed for low-latency general-purpose work with complex control logic and branching, whereas a GPU is designed for high throughput on the same operation applied to many items; ...(and) training a model is exactly this kind of repetitive arithmetic over large matrices of image data; A GPU has much higher memory bandwidth than a CPU; ...(so) the large volumes of image data can be moved to the cores fast enough to keep them busy; A GPU delivers more calculations per watt for this kind of parallel work; ...(so) training costs less energy/time than doing the same work on the CPU;	
HL 6(c) A1.1.6	Award [2] max Award [1] per point up to [2] max. The stages (fetch, decode, execute, write-back) are overlapped, so a new instruction can start before the previous one has finished; Each stage of the processor is kept busy instead of standing idle waiting for the instruction ahead of it; A completed instruction therefore leaves the pipeline every clock cycle once it is full, rather than one every four cycles; The clock speed is unchanged; it is the number of instructions finished per second (throughput) that rises; Award [1] max if the answer only states that instructions are processed 'at the same time' with no reason. Accept an analogy (for example a laundry pipeline) if the overlap is made clear.	2

Total for this paper: 35 marks