Question 1: Latency and Bandwidth **[6 Total]**

#1

$$\text{Total Latency} = \frac{1\,024 \text{ B} \cdot 8 \text{ b/B}}{50 \cdot 10^6 \text{ b/s}} + \frac{200 \text{ m}}{200 \cdot 10^6 \text{ m/s}} + 2 \cdot 100 \cdot 10^{-6} \text{ s} = 364.84 \; \mu\text{s}$$

$$\text{Effective Bandwidth} = \frac{1\,024 \text{ B} \cdot 8 \text{ b/B}}{\text{Total Latency}} \approx 22.45 \text{ Mb/s}$$

[6]

Question 2: Communication **[6 Total]**

#1 A barrier is a point in the code where multiple threads in a parallel program must stop execution and wait for each other. Once all these threads have reached the barrier, the threads are synchronised and allowed to continue.
[3]

#2 Broadcast communication is where one message is sent to multiple receivers simultaneously. Scatter communication is where different messages are sent to each receiver, from the same source.
[3]

Question 3: Cloud Computing [7 Total]

#1 Platform as a service is a cloud computing service that provides a development platform (some user-friendly API) on which to build applications. These applications are run on the cloud service. The Google App Engine, for example, provides a platform on which to develop web-based applications, which are then hosted by Google's cloud servers.

[3]

#2 Virtualisation is the process of creating a virtual (as opposed to actual) version of something. In computing terms, the thing being virtualised is most often a hardware platform, but it could be other things as well. With regards to cloud computing, it is the creation of multiple instances of virtual hardware, operating systems and software platforms, even though the underlying real hardware and software platform might be something completely different.

[4]

Question 4: Seminar Related Multiple Choice [8 Total]

#1 (b)
THT $\Rightarrow$ Through-hole technology
DIL $\Rightarrow$ Dual in-line
DIP $\Rightarrow$ Dual in-line package (implies through-hole by convention)

[2]

#2 (b)

[2]

#3 1. This technology is reprogrammable (FPGA)

2. This technology is the faster of the two (ASIC)

3. This technology wastes very little space (ASIC)

4. This technology is more common for low-volume production (FPGA – referring to end-product production)

[4]

Question 5: GPGPU [7 Total]

#1
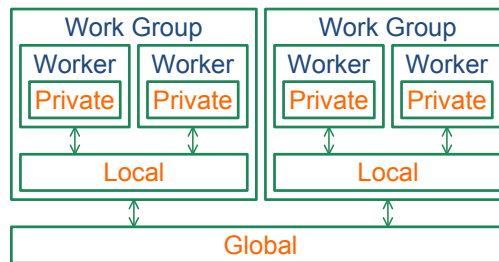
Memory is organised into 3 levels: private, local and global. Each thread, or worker, has a small section of private memory. Workers are grouped into sets of work groups. The workers of each work group share that work group's local memory. All workers have access to global memory. [5]

#2 Consider the kernel below:

```
__kernel void Add(
  __global float* A, // Input
  __global float* B, // Input
  __global float* Y  // Output
){
  const int i = get_global_id(0); // Get loop index
  Y[i] = A[i] + B[i];
}
```

Each kernel instance has a unique global ID set (which can be up to 3 dimensions). The get_global_id(n) function obtains the ID for the current kernel instance, for the nth dimension. It is used to know which portion of the work the current kernel instance is supposed to be doing. [2]

Question 6: FPGA and ASIC [16 Total]

#1 An ASIC standard cell is an ASIC manufacturer-provided building block that ASIC designers can use in their designs. The ASIC manufacturer guarantees certain properties of the cell, including timing delays and correct functionality.
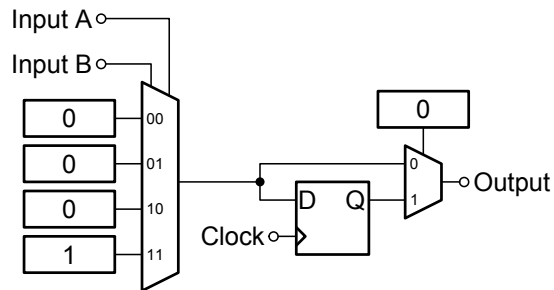
Typical standard cells include gates (AND, OR, NAND, NOR, XOR, NOT, etc.), registers, phase-locked loops, adders, multipliers, etc. [2]
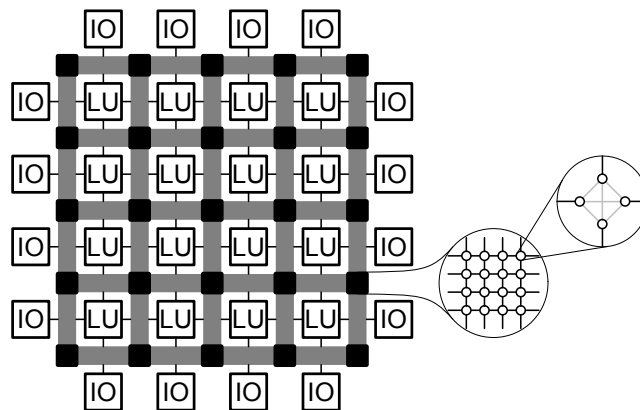
#2
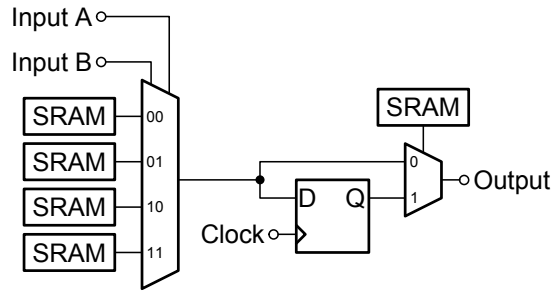
[4]

#3 Program the SRAM cells as follows:

[2]

#4

In the figure, IO indicates an input-output block and LU indicates a logic unit. The logic units are arranged in a grid pattern and surrounded by interconnecting wires. Wherever these wires cross is a junction box. Each junction box (of 4 incoming wires) has 6 programmable switches. By programming these switches, the interconnect can be configured in any way desired. **[6]**

#5 Extra features include embedded memory, multipliers, phase locked loops, embedded processors, Ethernet MACs, DDR memory controllers, floating-point units, etc. **[2]**