Q1 What is the different between Princeton architecture and Harvard architecture？

For Harvard architecture, program and data are stored in two different memory structures. For Princeton architecture, program and data are stored in a shared memory structure.

Q2 In 2017, the ACM Turing Award was awarded to David A Patterson and John L. Hennessy for their invention of which type of ISA (CISC or RISC)?

RISC. (<https://www.acm.org/media-center/2018/march/turing-award-2017> )

Q3 If the CPU can do one addition of two 32-bit values for every 1 ns, and a memory transfer takes 8 ns, how would you design the memory interface (data bus width) and why?

8\*2\*32bits=512bits

During one memory transfer, CPU can do 8 additions(8ns/1ns=8).8 additions need 16 32-bit values, which need to be transferred from memory to CPU every 8ns to keep CPU 100% busy.

Q4. Please rank the size and access time of memory structures in the computers: cache, disk, register, main memory.

Size:

disk>main memory>cache>register

Access time:

disk>main memory>cache>register

Q5. List the difference between address bus, data bus, and control bus. Which one determine the memory capacity?

System bus is used to connected different components within computer (CPU, memory, IO), and system bus is composed of address bus, data bus and control bus:

1. Address bus is uni-directional. It transmits address information from MAR (memory address register) in CPU to memory or IO devices.
2. Data bus is bidirectional. It transmits data information between CPU and memory (IO).

For example, CPU reads instructions from memory through data bus and writes the instruction execution results back to memory through data bus.

1. Control bus is bidirectional. It can transmit control signal from CPU to memory (IO) or feedback information from memory(IO) to CPU.

6. Can you derive the truth table for the following CMOS circuit?

|  |  |  |
| --- | --- | --- |
| A | B | F |
| 0 | 0 | 1 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 0 |

This is a NOR gate.

1.只有当A、B都是低电平时，上方的两个串联的PMOS才会同时导通，并且下方并联的NMOS都截止。此时F拉高成高电平输出1。

2.当A、B存在高电平时，上方串联的PMOS不会都导通，并且下方并联的NMOS至少1个导通。此时F拉低成低电平输出0。

7. Suppose a following C program:

int result = b = 0;

int array[1000];

for (int i = 0; i < 1000; i ++)

  result = array[i] + b;

Suppose a CISC CPU can execute the program with 100 instructions per second while a RISC CPU can execute the program with 150 instructions per second. Can you determine which CPU is able to finish the program faster? Why? (You can assume all variables first exist in the main memory. Since we haven't taught the details of CISC/RISC, one general rule that you can use is that the CISC instruction is more complex than the RISC instruction. For example, a CISC add instruction can operate on the main memory address directly while the RISC add instruction can only operate on the register.)

We cannot determine which CPU is faster. A compiler will compile this C program to the assembly program in each CPU. We cannot determine the number of instructions in the compiled assembly programs for the two CPUs. (Although we know the assembly program in the CISC CPU will probably have less instructions than it on the RISC CPU, the specific numbers cannot be derived.)

8. For the same C program:

int result = b = 0;

int array[1000];

for (int i = 0; i < 1000; i ++)

  result = array[i] + b;

You can assume all variables first exist in the main memory. We have taught in the class that the register is faster to access than the main memory. But register capacity is limited. If you want to select several variables to put in to the register, which variable(s) would you choose?

It is best to put variable i, result, and b in the register because they are used in every loop iteration.