Name the three basic components of every computer.

1. 用来解释和执行程序的处理器② 用来存储数据和程序的存储器③ 与外界进行数据传输的机制

Name and explain the main components of a von Neumann computer.

Three hardware systems:

central processing unit (CPU)：由control unit控制单元、program counter程序计数器、arithmetic logic unit算数逻辑单元、register寄存器组成

main memory system：保存控制计算机操作的各种程序

I/O system

具有执行顺序指令的处理能力

在主存与CPU控制单元之间，包含一条物理或逻辑上的单一通道，可强制改变指令和执行周期

CPU

Central processing unit 中央处理器，任务是负责提取程序指令，并对指令进行译码，然后按照程序规定的顺序对正确的数据执行各种操作。由数据通道datapath和控制单元control unit组成。

ALU

arithmetic logic unit 算术逻辑单元：在程序执行过程中用于进行逻辑运算和算术运算。

Accumulator architectures

Accumulator 累加器：用来保持CPU需要处理的数据值，是一个通用寄存器。现在大部分计算机都有多个通用寄存器

MAR

Memory address register 存储器地址寄存器：用来保持被引用数据的存储器地址

MBR

Memory buffer register 存储器缓存寄存器：用来保持刚从存储器中读取或者将要写入存储器的数据

PC

Program counter 程序计数器：用来保持程序将要执行的下一条指令的地址

IR

Instruction register 指令寄存器：用来保存将要执行的下一条指令

InREG

Input register 输入寄存器：用来保持自输入设备的数据

OutREG

Output register 输出寄存器：用来保持输出到输出设备的书

How does the fetch-decode-execute cycle work?

CPU首先提取一条指令，即将指令从主存移动到指令存储器；接着对指令进行译码，即确定指令的操作码和提取执行该指令所需的数据；然后执行这条指令，即执行指令规定的各种操作。

****

步骤1、2完成取址，3完成译码，4完成执行：

1. 将PC中的内容复制到MAR:MAR←PC；

2. CPU转向主存储器，提取由MAR给出的地址单元中的指令，并将指令放入指令寄存器IR中。同时，PC自动加1，现在PC指向程序的下一条指令：IR←M[MAR]和PC←PC+1;

3. 将IR中最右边的12位复制到MAR，并对IR最左边的4位进行译码，以确定操作码：MAR←IR[11-0]和译码IR[15-12];

4. 如果需要，CPU将使用MAR中的地址转向存储器提取数据，并将数据放入MBR（可能是AC）中，然后执行该命令：MBR←M[MAR]，并且执行实际的指令。

What is the difference between a byte and a word?

一个字节8位，一个字不一定是8位，可为16位、32位等。

Explain the difference between memory-mapped I/O and instruction-based I/O.

Memory-mapped I/O 内存交换（映射）输入输出：接口中的寄存器地址就在内存地址的分配表中。CPU对I/O设备的访问和CPU对内存的访问完全相同。

Instruction-based I/O 指令实现的输入输出：CPU需要有专门的指令来实现输入输出操作，虽然不占内存空间，但需要特殊的I/O指令。

Describe how an interrupt works and name four different types interrupt at least.

中断：改变或中断系统正常执行流程的各种事件，大部分的计算机采用的中断处理方法是：在机器的每一个取指-译码-执行周期开始出，首先检查是否有中断请求存在。如果有，则CPU先处理中断任务。处理完成后，CPU会继续进行机器的执行过程。若没有，CPU则进行正常的程序执行过程；I/O请求，算数错误，算数下溢或上溢，硬件故障，用户定义的中断点，页面错误，非法指令

What is the difference between synchronous buses and nonsynchronous buses?

同步：事件发生的顺序由时钟脉冲来控制。异步：采用较为复杂的握手协议来强制实现与计算机其他操作同步

What is a stack?

是一种后进先出的数据结构，堆栈中所有数据的插入删除必须从堆栈顶部开始。

What is an address mode? List five types of address mode

Addressing mode：指定指令中操作数位置的方法

立即寻址（immediate addressing），直接寻址（direct addressing），间接寻址（indirect addressing），变址寻址（indexed addressing），基址寻址（based addressing），堆栈寻址（stack addressing）

What are immediate, direct, register, indirect, register indirect, and indexed addressing

Immediate addressing立即寻址：指在指令中操作代码后面的数值会被立即引用。即要操作的数据本身是指令的一部分

Direct addressing 直接寻址：指在指令中直接指定要引用的数值的存储器地址

Indirect addressing 间接寻址：地址域中的二进制数用来指定一个存储器地址，该地址中的内容将被用作一个指针，操作数的有效地址是通过访问这个存储器地址来获得的。

Register addressing 寄存器寻址: 是采用一个寄存器，而不是存储器来指定操作数

Indexed addressing 变址寻址：一个变址寄存器用来存储一个偏移量或称为位移量。将这个偏移量与操作数相加，就产生了指令所要求数据的有效地址。

What is an opcode?

操作码：指定要执行的操作类型

Operand

操作数：指出数据所处的寄存器和内存单元。

What are the advantages and disadvantages of fixed-length and variable-length instructions? Which is currently more popular?

Fixed length：指令执行速度快，会浪费内存空间；

variable length：可以节省内存空间，译码会变得复杂

通常会考虑一些折中的方案，采用两到三种不同的指令长度。

Explain the concept of pipelining.

计算机使用时钟脉冲来精确控制各个操作步骤的顺利执行，但是有时还可以使用一些额外的脉冲来控制某个操作步骤中的一些小细节，有些CPU会将取指-译码-执行周期分成一些较小的步骤，其中的某些较小的步骤可以并行执行，这种时间上的交叠可以加快CPU的执行速度，这种方法称为流水线（pipeline）。

What are the pipeline conflicts that can cause a slowdown in the pipeline?

Pipeline conflict 流水线冲突：

Resource conflict 资源冲突：指令级并行执行过程中要考虑的主要问题，比如两个操作同时需要访问内存。

Data dependency 数据关联：当一条指令的执行尚未结束时，后面某条指令却要求该指令的执行结果作为操作数。

（Conditional branch） 条件分支语句

What is an ISA?

ISA，instruction set architecture是在机器上运行的所有软件和执行这些软件的硬件之间的协定接口，规定了计算机可执行的每条指令及其格式

What is SSI, MSI, LSI, and VLSI

小规模集成电路（SSI）：每块芯片上只有10-100个元件；中规模集成电路（MSI）：每块芯片上集成100-1000个元件；大规模集成电路（LSI）：每块芯片上集成的元件多达1000-10000个；最后是超大规模集成电路（VLSI），每块芯片上集成的元件超过10000.超大规模集成电路标志着第四代计算机的开始。

What is a bus in a computer.

总线：一组导电线路的组合，它作为一个共享和公用的数据通道将系统内的各个子系统连接在一起

What is a bus master?

总线主设备：最初启动的设备

Bus slave

总线从设备：响应主设备请求的设备

What is the difference between a point-to-point bus and a multipoint bus?

点对点总线连接的是两个特定的设备；多点总线则要求多个设备共享同一条总线

Explain the differences between data buses, address buses, and control buses

Data buses：用于数据传输的总线；

address buses：指出数据读写的位置；

control buses：指示哪个设备允许使用总线，以及使用目的

What is a bus cycle

是完成总线信息传送所需的时钟脉冲间的时间间隔

What are the four types of bus arbitration?

1. Daisy chain（菊花链）
2. Centralized parallel（集中式平行）
3. Distributed using self-detection（自选择的分布式）
4. Distributed using collision-detection（冲突检测的分布式）

What is an expanding opcode?

一个指令总长度固定的指令系统体系结构，但是却可以允许操作数域的位数根据需要而改变

What is a dirty block?

已经被修改过的数据块。

State Amdahl’s Law in words

第一章，基本表述：对某种特定的系统改进，系统性能增强的可能性收到被改进特征部位的使用次数的限制

第七章，概念：计算机系统整体的性能的速度提升（称为加速度，speed up）取决于某个特定部件本身的加速率和该部件在系统中的使用率

What is speedup?

计算机系统整体的性能的速度提升，称加速度

Name and explain the seven commonly accepted layers of the Computer Level Hierarchy.

6、user level用户层，由各种应用任务组成。

5、high-level language level高级语言层，由各种高级编程语言组成

4、assembly language level汇编语言层，包含某种类型的汇编语言。

3、system software level系统软件层，主要处理操作系统的命令。

2、指令集体系结构（Instruction Set Architecture），或者称为machine level机器层，由特殊的计算机系统结构所能识别的机器语言组成。

1、control level控制层。在这一层次上，控制单元将确保正确地译码并执行指令，并且适时地将数据传送到正确的位置。

0、digital logic level数字逻辑层，在这里我们面对的是计算机系统的物理构成：各种逻辑门和引线。

What is assembler?

即编译程序，使用助记标识符将汇编语言转换成机器语言

What is effective access time (EAT)?

EAT是使用命中率与相连存储器层次的相对访问时间产生的加权平均值

EAT=H×AccessC+(1-H)×AccessMM

其中H为高速缓存命中率，AccessC是高速缓存的访问时间，AccessMM是主存储器的访问时间

Define the terms seek time, rotational delay, and transfer time.

Seek time 寻道时间：指磁盘驱动臂定位到指定的磁道上所需的时间。并不包括磁头读取磁盘目录。

Rotational delay 旋转延迟：是读写头定位到指定扇区上方所需的时间。

Transfer time 传输时间：寻道时间+旋转延迟+从磁盘上实际读取数据所需时间

What is a protocol?

协议：在发送设备和接受设备之间交换的各种信号的具体形式和信号所代表意义。包括：命令信号；状态信号；数据传递信号

Name the four types of I/O architectures

Programmed I/O 程序控制的I/O，用于某些专用系统；interrupt-driven I/O 中断控制的I/O，用于当今许多流行的操作系统；direct memory access 直接存储器存取，多用于小型单用户计算机系统，管理开销很低；channel I/O 通道控制的I/O，多用于“慢速”设备

What is interrupt-driven I/O?

中断控制的I/O：当CPU要执行输入或输出指令时，首先通知相应的I/O设备；然后，继续处理其他的一些工作任务，直到该I/O设备准备就绪；这时，I/O设备会向CPU发送一个中断请求信号；随后，CPU会响应和处理这个中断请求；完成输入或输出操作后，CPU会继续其正常的取指-译码-执行周期。

What is polling?

程序控制的I/O>轮询：CPU会持续不断地监视每个寄存器，等待数据的到达。

Explain how programmed I/O is different from interrupt-driven I/O.

程序控制的I/O：计算机系统需要为每个I/O设备至少准备一个专用的寄存器，CPU会不断地监视每个寄存器，等待数据的到达。每次传输一个字节；中断控制的I/O：CPU不再需要持续的查询其附属设备是否有任何输入请求，而是在有数据发送需求时由外部设备来通知CPU.每次可按一个字节，或者小数据块的形式处理数据。

**通道控制的I/O与中断控制的I/O有什么不同？**

中断控制的I/O：通过传输字或者小数据块，这要取决与传输的设备。

而通道控制I/O，一个I/O处理器可以控制几条I/O通道；对于小规模数据的输入I/O如键盘等，可以将它们合并起来一起使用，称为多道复用通道，而对于大量数据的传输处理器。则使用选择通道。

How does direct memory access (DMA) work？

一旦在存储器中装入了所要求的数值后，CPU就会发送信号到DMA子系统，由DMA来负责I/O的细节过程，而CPU会继续执行其下一个任务，在完成I/O处理后，DMA子系统会发送另一次中断请求通知CPU

What is the Principle of Equivalence of Hardware and Software?

任何可以利用软件实现的事情可以利用硬件来实现。反之，任何可以利用硬件来实现的事件也同样可以利用软件来实现。

Locality of reference三个引用局部性

Temporal locality时间局部性：最近访问过的内容很可能不久的将来再次被访问

Spatial locality空间局部性：对存储器地址空间的访问形式形成团簇的集中倾向

Sequential locality顺序局部性：访问存储器的指令倾向于按顺序执行

Give the two factors and four possible combinations for categorizing computer architecture in Flynn’s taxonomy

主要考虑的两大因素（分类依据）：指令的数目和流入处理器的数据流的数目。

SISD（单指令流，单数据流）

SIMD（单指令流，多数据流）

MISD（多指令流，单数据流）

MIMD（多指令流，多数据流）

Assume you have a machine that uses 32-bit integers and you are storing the hex value

1234 at address 0.

a. Show how this is stored on a big endian machine.

b. Show how this is stored on a little endian machine.

c. If you wanted to increase the hex value to 123456, which byte assignment would be

more efficient, big or little endian? Explain your answer.

Ans.

a and b.

Address 00 01 10 11

Big Endian 00 00 12 34

Little Endian 34 12 00 00

c. Little endian is more efficient because the additional information simply needs to be

appended. With big endian, the "12" and "34" would need to shift to maintain the

correct byte ordering.

State Moore’s Law.

硅芯片的密度每18个月翻一番

Rock’s Law

制造半导体集成电路所需要的主要设备成本每4年翻一番

若moore’s law 成立，rock’s law失效

RAM and ROM

Random access memory 随机存储器：执行程序时，用来存放数据或者程序，是一种易失性存储器

Read only memory 只读存储器：非易失性存储器

高速缓存的映射模式：

Direct mapped cache 直接映射的高速缓存

概念：采用模块方式来指定高速缓存和主存储器之间的映射关系。

将主存储器中的块X映射到高速缓存的块Y，对应的模量为N。其中，N是高速缓存中所划分的存储空间块的总数。

求模运算： Y = X mod N

优点： 简单 缺点： 浪费资源

域的划分：标记、块、字

Full associative cache 全关联高速缓存

域的划分：标记域、字域

缺点：查找麻烦（需要遍历）

Set associative cache组关联高速缓存

组内是全关联，组组之间是直接映射。

域的划分：标记、组、字

What is Virtual Memory?

虚拟存储器：使用硬盘作为RAM存储器的扩充，增加了进程可以使用的有效地址空间。

Virtual address

虚拟地址：进程所使用的逻辑地址或者程序地址，

Physical address

物理地址：物理存储器的实际地址

Mapping

映射：一种地址变换机制，通过映射可将虚拟地址转化为物理地址

Page frame

页帧：由主存（物理存储器）分成的相等大小的数据块

Pages

页：由虚拟存储器（逻辑地址空间）划分成的数据块

Paging

分页：将一个虚拟页从硬盘复制到贮存的某个页帧的过程

Fragmentation

存储碎片：变得不能用的存储单元

What is a page fault?

缺页：一个请求页在主存储器中没有找到时所发生的的事件。必须将请求页从硬盘复制到存储器

What is a TLB and how does it improve EAT?

Translation look-aside buffer 转换旁视缓冲器，可以加速页表的查询时间。每个TLB的入口目录由一个虚拟页码和对应的物理页帧的帧数组成。TLB通常由关联高速缓存来实现，虚拟页码和物理页帧对可以映射到TLB高速缓存的任何位置，加速页表查询时间。

Explain the concept of a memory hierarchy

计算机采用不同类型的存储器的组合配置。基本类型包括：寄存器、高速缓存、主储存器和辅助储存器。计算机存储器由层次结构组成，最顶端是规模最小速度最快的存储器，最底层的是规模最大速度最慢的存储器。