

Bus, IO, External memory

最后一部分，比较简单，主要是概念

Bus, IO, External memory

Bus(总线)

基本参数

通信的方式

Bus scalability(总线可扩展性)

Bus Timing(总线时序)

Input/output Interfaces (输入输出接口)

I/O modules的使用原因

I/O modules (I/O模块)

Programmed I/O(程序化 I/O) 和 Interrupt Driven I/O(中断驱动)

Addressing memory-mapped I/O Devices (寻址内存映射I/O设备)

DMA – direct memory access (直接内存访问)

External memory

Disk performance

Magnetic Tape

Flash vs. Disks

Question

Bus(总线)

把CPU, RAM, Input/output devices etc连接起来

Multiple senders and multiple receivers, but one sender at a time (多个发送者和多个接收者，但一次只能发送一个发送者)

基本参数

- **Bandwidth (带宽)** : The amount of data that can be transferred at once (e.g., per clock cycle)
- **Clock speed (时钟速度)** :
 - 1 MHz (megahertz) = 1 million cycles per second
 - 1 GHz = 1000 Mhz
 - Faster clock speed = lower transmission latency
- **Theoretical throughput (理论吞吐量) = Bandwidth X Clock Speed**

通信的方式

Memory-like communications(内存类通信)

- **initiator issues read/write transactions to an address (device)**: 这里的“initiator”指的是发起通信的设备或组件，通常是CPU。它会发出读取或写入的请求（即“transactions”），这些请求是针对特定的地址或设备的。例如，CPU可能会发出一个请求，要求从内存的某个地址读取数据，或者将数据写入到某个地址。
- **each target listens to an address range to respond, by returning or accepting data**: “target”指的是目标设备或组件，例如内存模块或I/O设备。每个目标设备都会监听它负责的地址范围。当它检测到发起者发出的请求是针对它所监听的地址范围时，它会做出响应。对于读取请求，目标设备会返回数据给发起者；对于写入请求，它会接受数据并将其存储在相应的地址中

Bus scalability(总线可扩展性)

将多个设备连接到单个总线会导致：

- long data paths and hence slower communication due to propagation delays;
- more complicated communication with the aggregate demand for bus access

大多数系统使用多总线来克服这些问题

- They are generally hierarchical (分层的)

Bus Timing(总线时序)

1. 同步时序 (Synchronous Timing) :

- 在同步时序中，总线上的事件发生是由时钟信号决定的。
- 例如，所有的“事件”都在时钟周期的开始时发生。
- 这种方式的优点是事件的发生是可预测的，因为它们与时钟信号同步。每个事件都发生在固定的时钟周期内，因此系统的时序是固定的和可预测的。

2. 异步时序 (Asynchronous Timing) :

- 在异步时序中，总线上的一个事件的发生依赖于前一个事件的发生。
- 响应时间可以变化，不是一个固定的时钟周期数。
- 这种方式的优点是更加灵活，因为事件的发生不需要等待时钟周期的开始，而是可以立即响应前一个事件的完成。然而，这也使得系统的时序更加复杂和不可预测，因为事件的发生时间可能会因各种因素而变化。

Input/output Interfaces (输入输出接口)

I/O modules的使用原因

- **A wide variety of peripherals exist with various methods of operation.**
存在各种各样的外设，它们的运行方式各不相同。
- **They deliver different amounts of data...**
它们传输的数据量不同...
- **...at different speeds...**
...传输速度也不同...
- **...and in different formats**
...并且数据格式各异。
- **Most are much slower than the CPU and RAM**
大多数外设的速度远低于CPU和RAM。

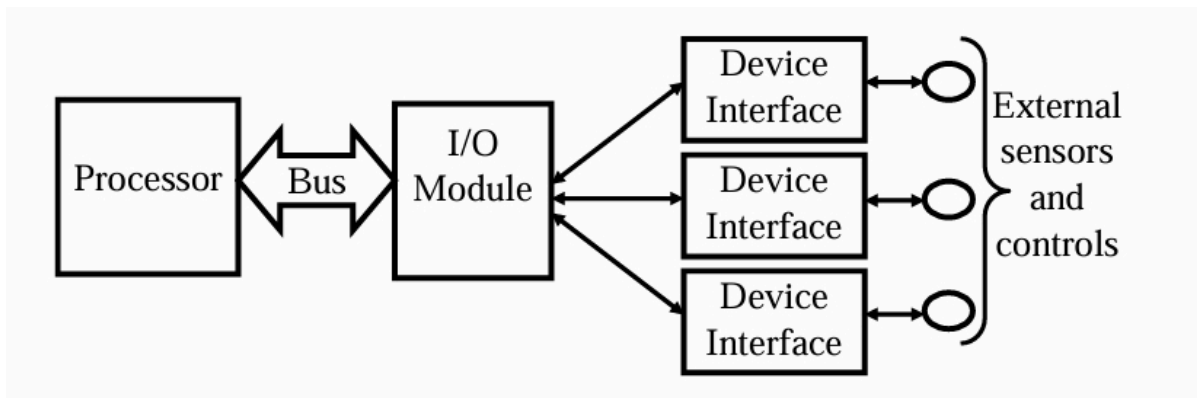
I/O modules (I/O模块)

Need I/O modules to act as bridge between processor/memory bus and the peripherals

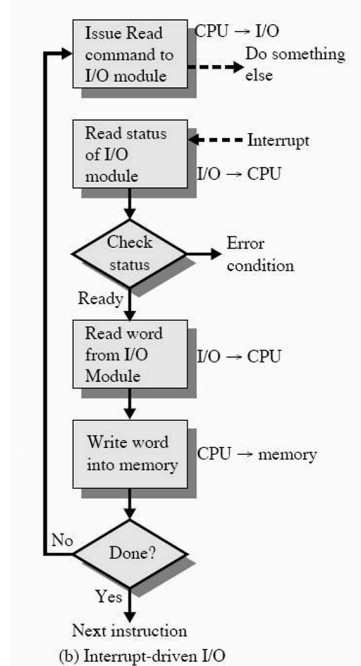
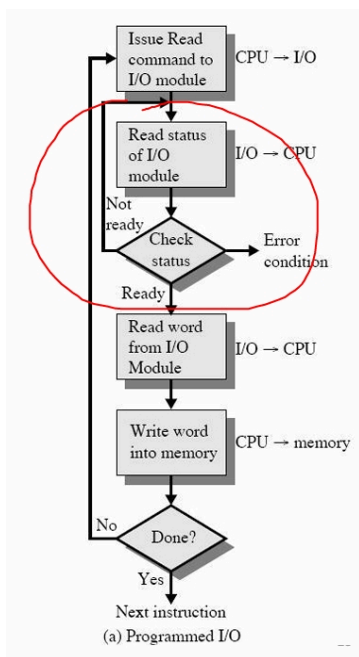
需要I/O模块充当处理器/内存总线和外围设备之间的桥梁

I/O modules provide interfaces to the CPU, memory and external devices

I/O模块为CPU、内存和外部设备提供接口



Programmed I/O(程序化 I/O) 和 Interrupt Driven I/O(中断驱动)



应当注意程序化 I/O 的 CPU 查询的方式是轮询 (Polling)

Interrupt Driven I/O(中断驱动)的 CPU 查询的方式是中断

Programmed I/O (程序化 I/O) 采用轮询而不是中断的原因主要与它的简单性和适用场景有关。以下是几个关键点:

1. 简单性:

- **实现简单:** 轮询机制的实现相对简单, 不需要复杂的中断处理机制。CPU 只需定期检查 I/O 设备的状态, 而不需要处理中断信号和中断服务程序。
- **控制明确:** 在轮询中, CPU 对 I/O 操作的控制更加明确和直接。程序员可以精确地知道何时检查设备状态, 以及何时进行数据传输。

2. 适用场景:

- **低速设备:** 对于一些速度较慢的 I/O 设备, 轮询是合适的。因为这些设备的数据传输速度较慢, CPU 可以轻松地在其他任务之间轮询这些设备, 而不会对整体系统性能造成太大影响。
- **简单任务:** 对于一些简单的 I/O 任务, 轮询可以避免中断带来的额外开销。中断处理需要保存和恢复 CPU 状态, 这在某些情况下可能不划算。

3. 避免中断开销:

- **中断开销:** 中断处理需要保存 CPU 的当前状态, 执行中断服务程序, 然后再恢复到被中断的任务。这个过程会带来一定的开销, 尤其是在频繁中断的情况下。
- **实时性要求不高:** 如果 I/O 任务的实时性要求不高, 轮询可以满足需求, 而不需要中断带来的快速响应能力。

4. 系统设计考虑:

- **资源限制**：在某些资源受限的系统中，可能没有足够的硬件支持来实现复杂的中断机制，或者中断资源已经被其他更重要的任务占用。
- **避免中断冲突**：在多设备环境中，频繁的中断可能导致中断冲突或优先级管理问题，而轮询可以避免这些问题。

Addressing memory-mapped I/O Devices（寻址内存映射I/O设备）

- Each device is given a unique identifier(address).
- The CPU commands contain the identifier.
- Each I/O device interprets the address lines on the bus to determine if a command is for itself

DMA – direct memory access（直接内存访问）

不必再去调用cpu

直接访问主内存

在游戏的加载阶段，DMA（Direct Memory Access，直接内存访问）通常用于将数据从存储设备（如硬盘）快速传输到内存中。DMA允许外设直接与内存进行数据交换，而不需要CPU的直接参与。这样做的好处是可以显著提高数据传输的速度和效率，因为CPU可以继续执行其他任务，而不需要等待数据传输完成。

External memory

这里的设备就是硬盘，磁盘什么的，我建议还是多看一下课本

Disk performance

Disk latency = Seek Time + Rotational Time + Transfer time

Magnetic Tape

- Secondary storage
- Good for backup
- Sequential access time

Flash vs. Disks

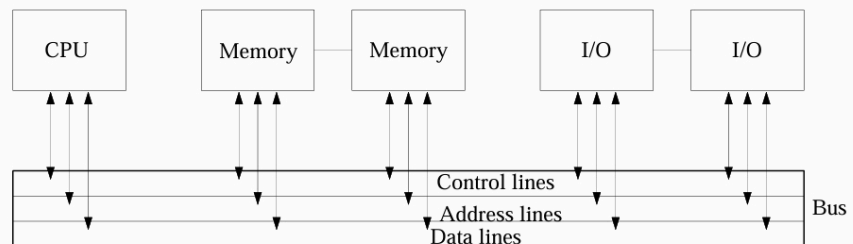
- Nearly instant standby wake-up
- No platter spinning
- Random access to data stored
- Constant access time – no head/platter movement (no seek, rotational latency)
- Tolerant to shock and vibration
- Charge can leak over time without power

以下是翻译：

- **几乎即时的待机唤醒**：从待机状态唤醒的速度非常快，几乎可以瞬间完成。
- **没有磁盘旋转**：与传统的机械硬盘不同，没有磁盘旋转部件。
- **随机访问存储的数据**：可以随机访问存储在设备中的数据，不需要按顺序读取。
- **恒定的访问时间 - 没有磁头/磁盘移动（没有寻道、旋转延迟）**：访问数据的时间是恒定的，因为不需要像机械硬盘那样移动磁头或等待磁盘旋转到特定位置，从而避免了寻道时间和旋转延迟。
- **抗冲击和振动**：对冲击和振动具有较高的耐受性，不易因这些因素而损坏。
- **没有电源时电荷会逐渐泄漏**：如果长时间不供电，存储在设备中的电荷可能会逐渐泄漏，导致数据丢失。

Question

- A system bus connects major components.
- It typically consists of 50–100s of separate lines.
- Each line has a particular meaning or function
 - data lines;
 - address lines;
 - control lines.



这里的独立线路是并行的意思吗

Load/store to mmap locations can have side-effects