

操作系统

主讲人：张杨

101001010100111101000010010111010010
0041000010100101001001010000101101001010140000111101001010100111101000010010111010010
11010101010111010000110000010100100100100000110100101014000011110100101

总体安排

- **学时：2学分（32学时）**
- **时间：周二第三大节（1-8周），周四第一大节（1-8周）**
- **地点：致理A112**
- **教材：**
 - **汤小丹, 梁红兵, 哲凤屏等. 计算机操作系统[M]（第四版）. 西安：西安电子科技大学出版社**
- **参考资料：**
 - **西尔伯查茨. 操作系统概念[M]. 北京：高等教育出版社**

考试安排

- **最终成绩组成：**
 - **平时成绩 30%：考勤、作业、课堂表现**
 - **期末闭卷考试 70%**
- **期末考试安排：第10周**

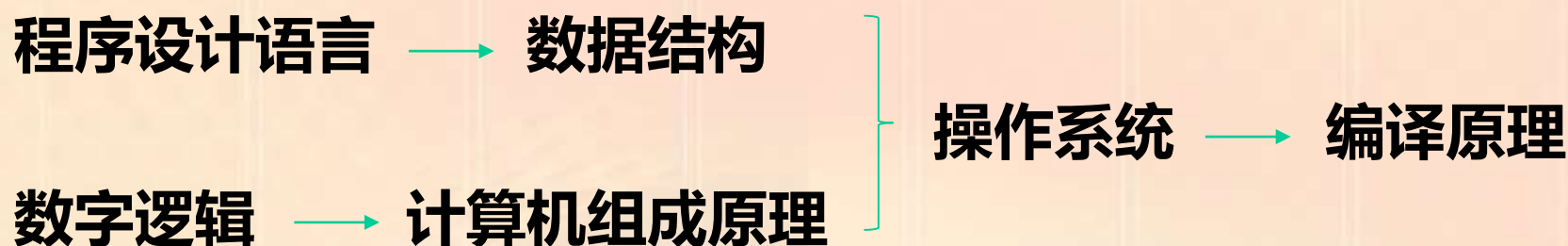
- **课程简介**：操作系统是一门涉及较多硬件的系统软件课程，在计算机软件课程的设置上起到承上启下的作用。既涉及硬件资源管理又涉及软件算法，是计算机课程的专业主干课。
- **课程特点**：概念多、较抽象和涉及面广，整体实现思想和技术又比较难于理解。
- **学习方法**：牢记概念、理解功能、联系实际、前后贯通。
- **主要知识**：操作系统的四大管理功能。

- **OS是什么？做什么？如何做？**

一、《操作系统》课程的重要性

1. 掌握核心系统软件
2. 掌握并发处理的思想方法
3. 考研专业课的重要科目
4. 为后继课程打好基础

二、《操作系统》的课程地位 专业基础课



三、课程内容

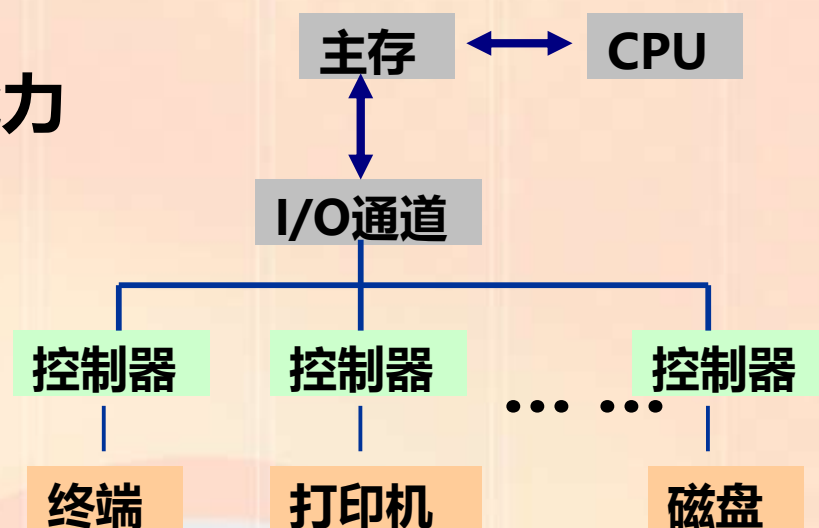
- 操作系统引论（3学时）
 - 进程管理（7学时）
 - 处理机调度与死锁（5学时）
 - 存储器管理（4学时）
 - 虚拟存储器（5学时）
 - 设备管理（3学时）
 - 文件管理（4学时）
 - 总结（1学时）
- } CPU管理

四、OS课程所站的角度

- 单CPU
- 多任务
- CPU具备与外设并行操作的能力

特别指出：

- | 程序均由CPU执行
- | CPU只能执行机器语言程序
- | CPU执行的程序均在内存中



- **1.1 什么是操作系统**
- **1.2 操作系统的发展历史**
- **1.3 操作系统的特征和服务**
- **1.4 操作系统的功能**
- **1.5 操作系统的分类**

1.1 什么是操作系统

一、操作系统（OS）的定义

你用过哪些OS？

Windows ; Unix ; Linux ; Dos

OS能做什么？

- ü 各种命令：dir copy del format
- ü 启动、结束用户程序
- ü 系统调用：例如INT指令
- ü UNIX 等提供多任务、多用户环境

1.1 什么是操作系统

OS不能做什么？

不做文档编辑

不做房屋设计

不是编译程序

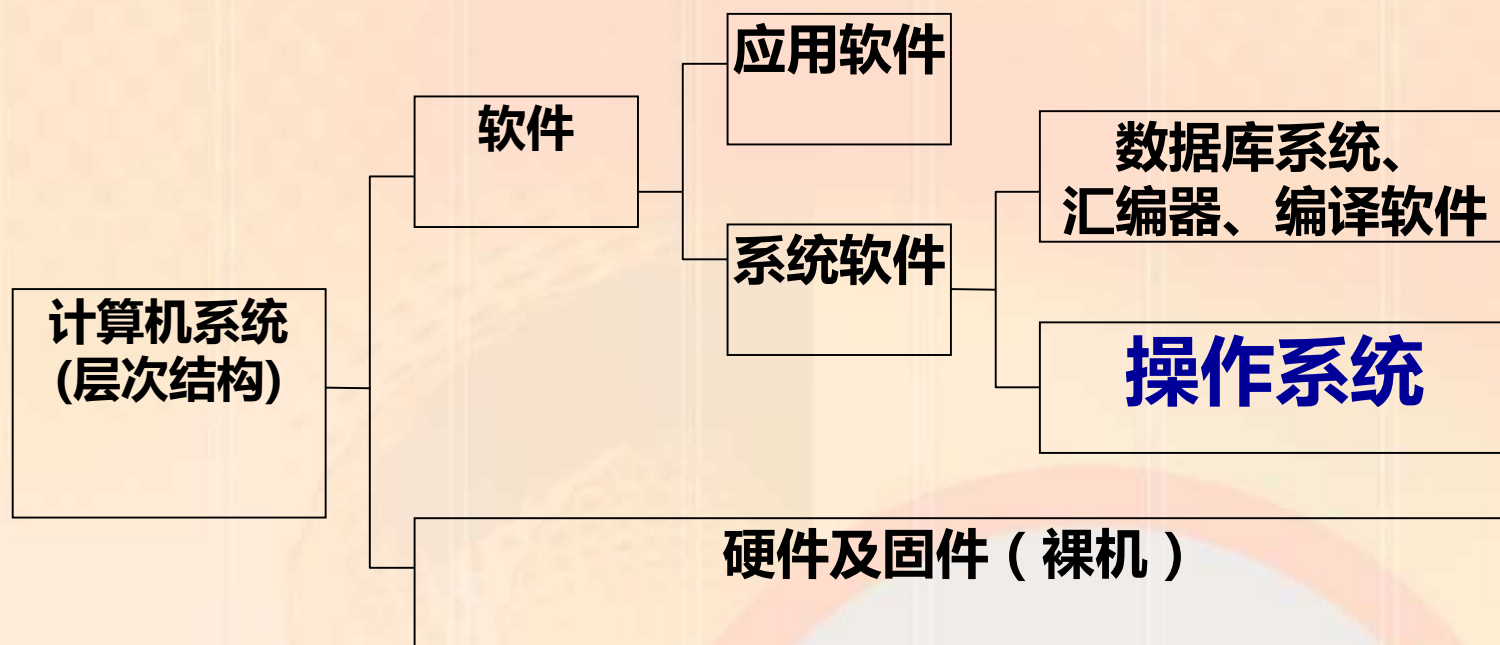
OS是什么？

结论：OS不直接解决最终具体应用问题，为你完成所有“硬件相关、应用无关”的工作，以给你方便、效率、安全。

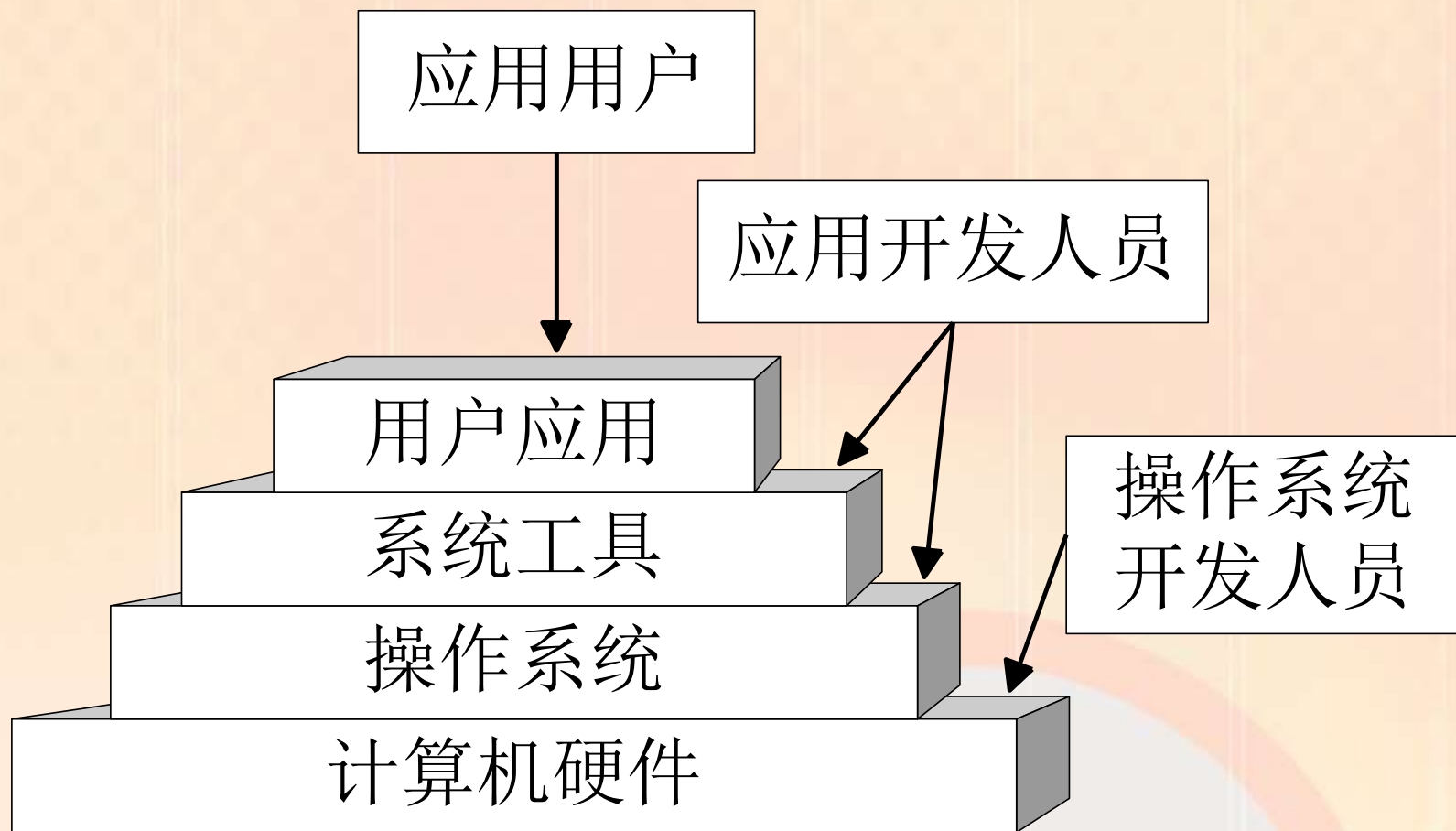
定义：OS是直接控制和管理计算机硬件、软件资源，合理地对各类作业进行调度，以方便用户使用的程序集合

1.1 什么是操作系统

计算机系统的组成：

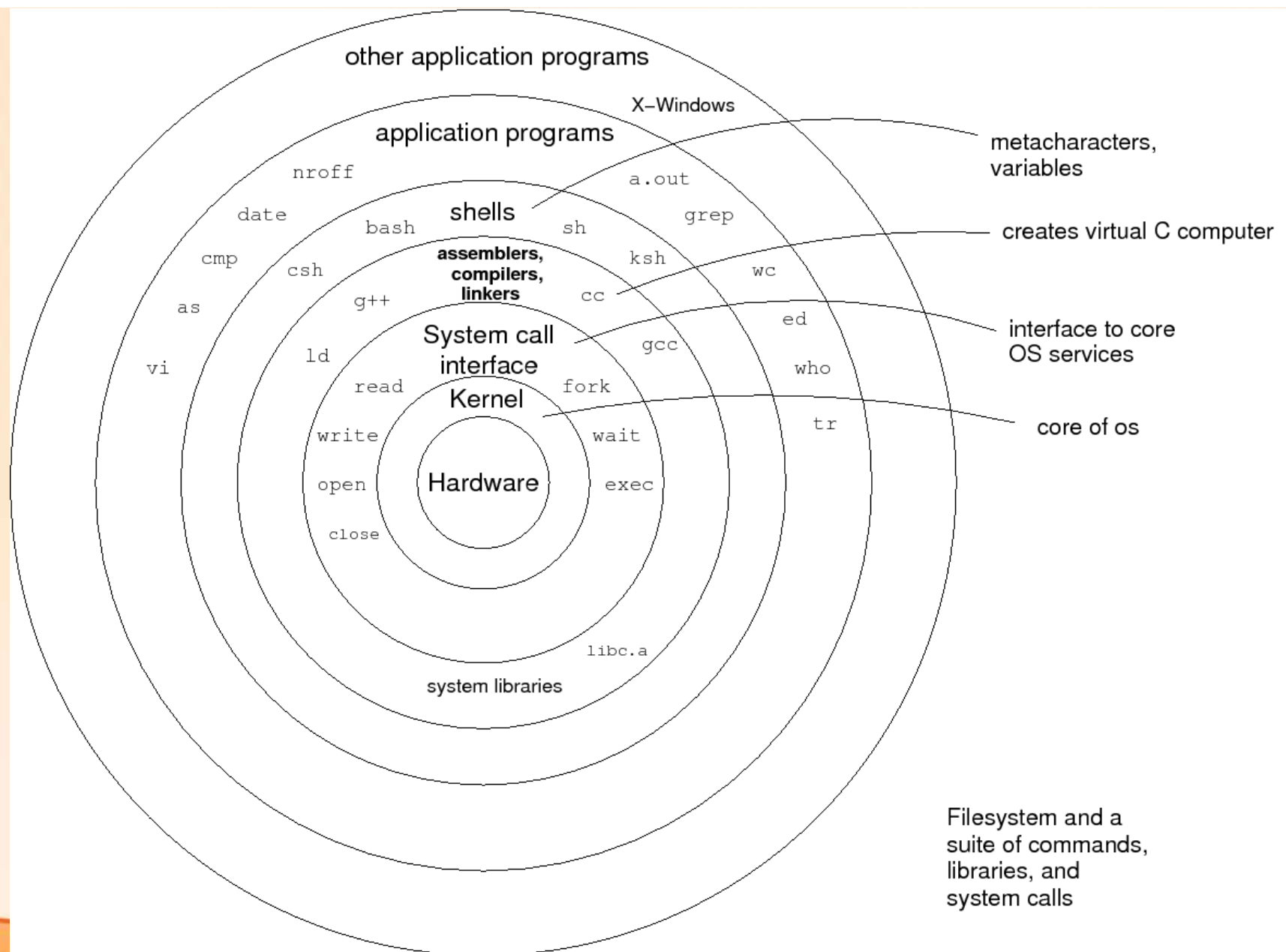


1.1.1 操作系统的地位和目标



- 紧贴系统硬件之上，所有其他软件之下（是其他软件的共同环境）

1.1.1 操作系统的地位和目标



Conceptual Architecture of UNIX SYSTEMS

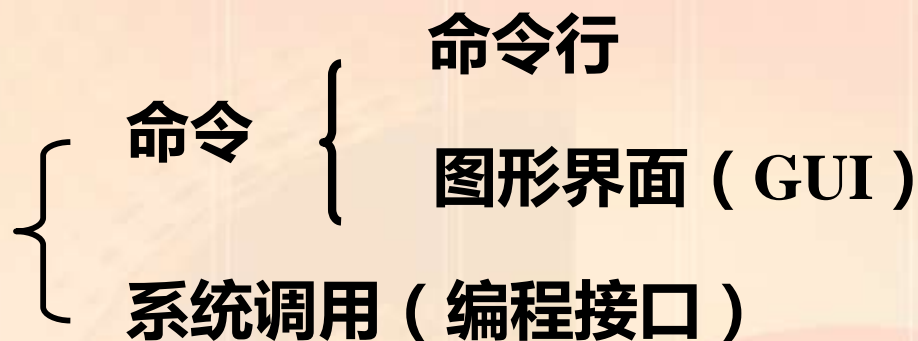
操作系统的定义

- 操作系统 operating system(OS)是配置在计算机硬件上的**第一层软件**，是对硬件系统的首次扩充。
- 操作系统是一组控制和管理计算机硬件和软件资源，方便用户使用的程序的集合。
 - 控制程序执行以防错误和不当使用
 - 执行用户程序并提供服务
 - 管理各种计算机资源：CPU、内存、硬盘、硬件外设等
 - 有效解决冲突请求并确保资源公平使用

几种观点

(1) OS是计算机硬件、软件资源的管理者。

(2) OS是用户使用系统的接口



(3) OS是扩展机(extended machine)/虚拟机(virtual machine)。

- 在裸机上添加：设备管理、文件管理、存储管理（针对内存和外存）、处理机管理

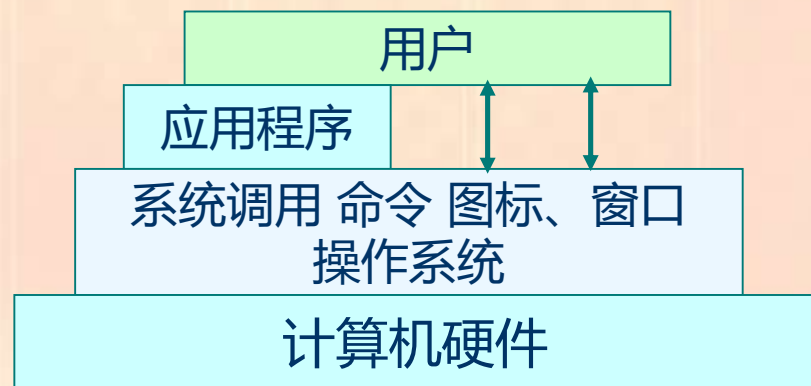
引入操作系统的目标

- **有效性**（系统管理人员的观点）：管理和分配硬件、软件资源，合理地组织计算机的工作流程
- **方便性**（用户的观点）：提供良好的、一致的用户接口，弥补硬件系统的类型和数量差别
- **可扩充性**（开放的观点）：硬件的类型和规模、操作系统本身的功能和管理策略、多个系统之间的资源共享和互操作

操作系统的作用

I 作为用户和计算机间的接口

OS处于用户与计算机硬件系统之间
用户通过OS来使用计算机系统



I 作为计算机系统资源的管理者

处理机管理、存储器管理、
设备管理、文件管理



I 实现了对计算机资源的抽象



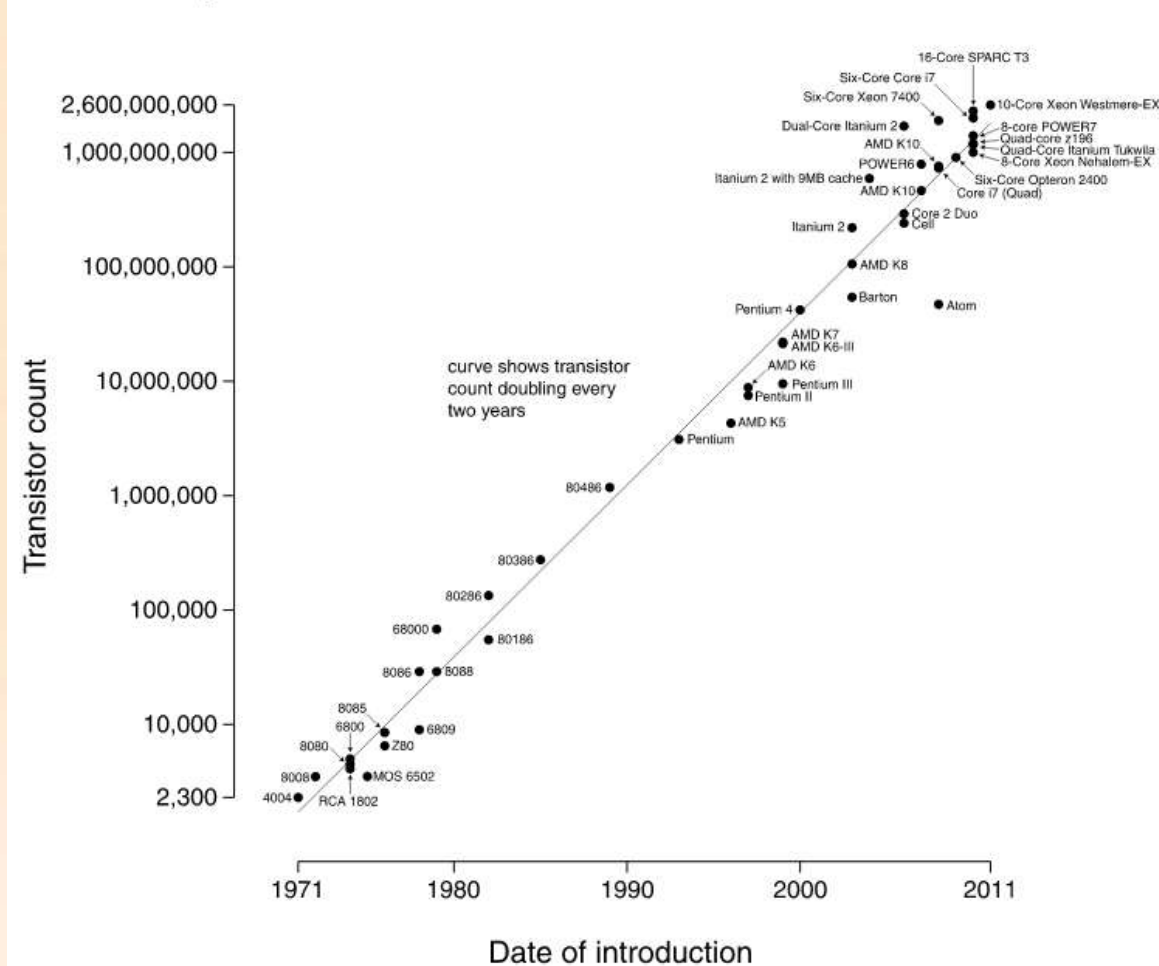
1.2 操作系统的发展历史

“需求推动发展”

1. **提高资源的利用率和系统性能：**计算机发展的初期，计算机系统昂贵，尽力提高各种资源的利用率，有了批处理系统。
2. **方便用户：**用户上机、调试程序，分散计算时的事务处理，有了分时操作系统。
3. **器件的发展：**CPU的位宽度（指令和数据）、快速外存

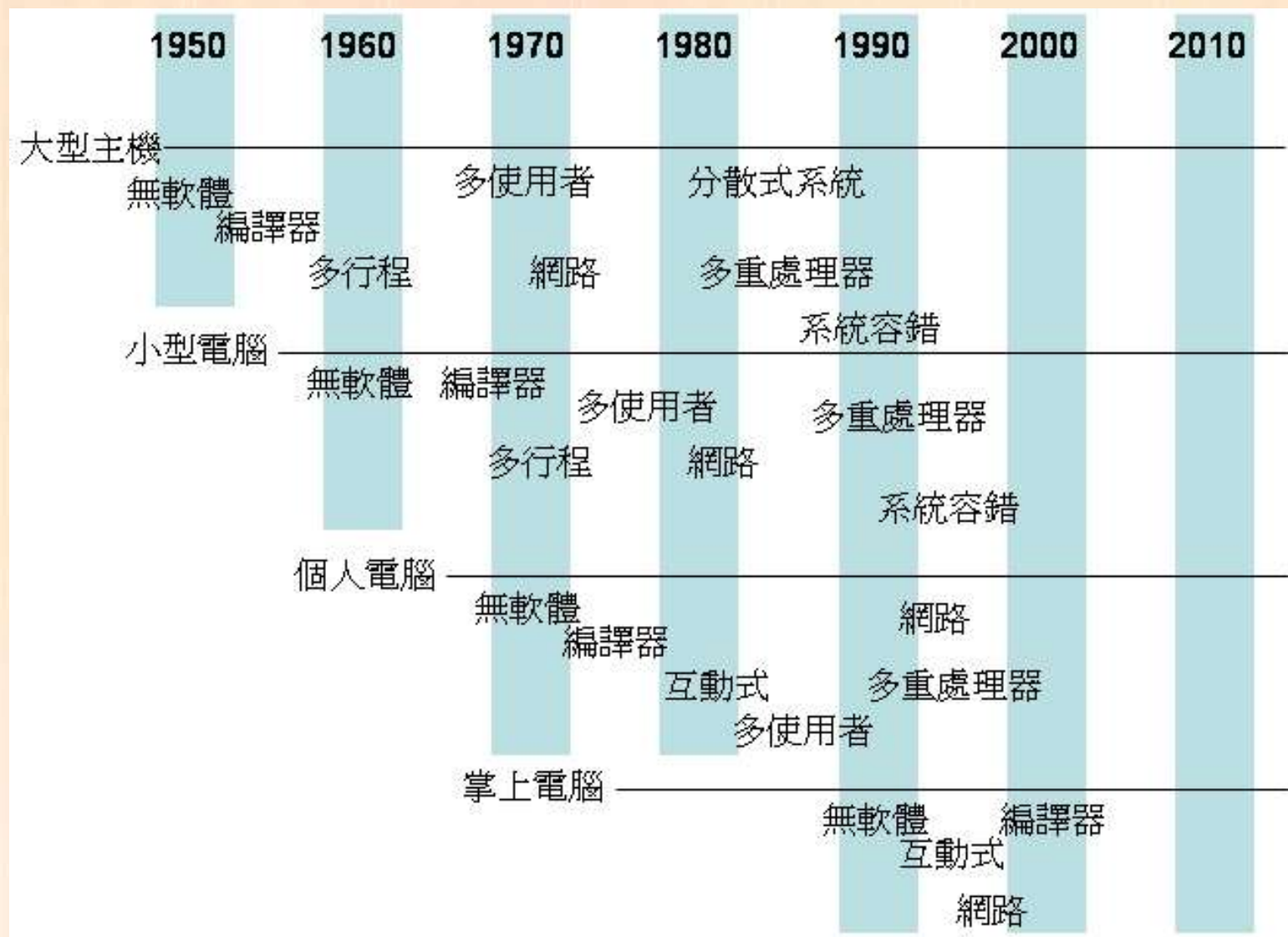
1.2 操作系统的发展历史

Microprocessor Transistor Counts 1971-2011 & Moore's Law



- 摩尔定律：
 - 集成电路芯片上所集成的电路的数目，每隔18个月就翻一倍。
 - 微处理器的性能每隔18个月提高一倍，或价格下降一半。
 - 用一个美元所能买到的计算机性能，每隔18个月翻两倍。

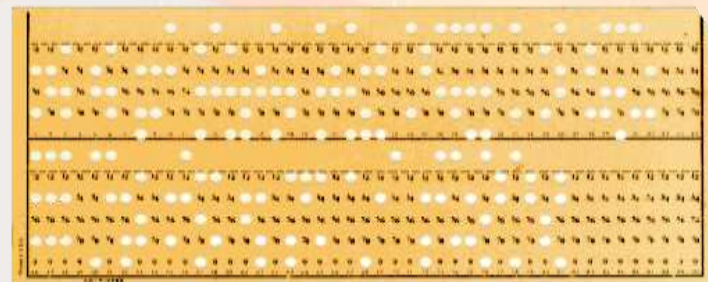
1.2 操作系统的发展历史



1.2.1 无操作系统阶段

人工操作方式

- 1946-50年代中：电子管时代，计算机速度慢，无操作系统，集中计算，计算机资源昂贵；
- 工作方式：
用户：既是程序员又是操作员；用户是专业人员；
输入输出：纸带或卡片；
缺点：用户独占全机，
CPU等待人工操作



1.2.1 无操作系统阶段

时间

1946 ~ 50年代（电子管），集中计算（计算中心），计算机资源昂贵；

需求

二战对武器设计的需要，美国、英国和德国等国家开始了电子数字计算机的研究工作。

哈佛大学的Howard Aiken

普林斯顿高等研究院的John Neumann（冯·诺依曼）

宾夕法尼亚大学的J.Presper Eckert和William Mauchley

德国电话公司的Konraad Zuse以及其他一些人都使用真空管成功地建造了运算机器。

1.2.1 无操作系统阶段

ENIAC计算机(电子数值积分计算机, Electronic Numerical Integrator And Computer, 1946年)

运算速度：5000次/秒, 18000个真空管, 占地182平方米, 130吨

花费：近五十万美元（相当于2011年的六百五十万美元）

没有程序设计语言（甚至没有汇编），更谈不上操作系统。

程序员提前预约一段时间,然后到机房将他的插件板插到计算机里。

期盼着在接下来的时间中几个真空管不会烧断从而可以计算自己的题目。



1.2.1 无操作系统阶段

人工操作方式特点：

用户独占全机

CPU等待人工操作

串行性

人工操作方式缺点：

昂贵组件的低利用率

利用率 = 执行时间 / (执行时间 + 读卡时间)

1.2.2 单道批处理系统(simple batch processing)

- | 用户使用系统提供的作业控制语言（JCL）来描述自己对作业运行的控制意图，并将这些控制信息连同作业一起提交给计算机。
- | 由OS去控制、调度各作业的运行并输出结果。
- | 由于作业进入系统后用户不再干预，从而提高了效率。

设计目标：

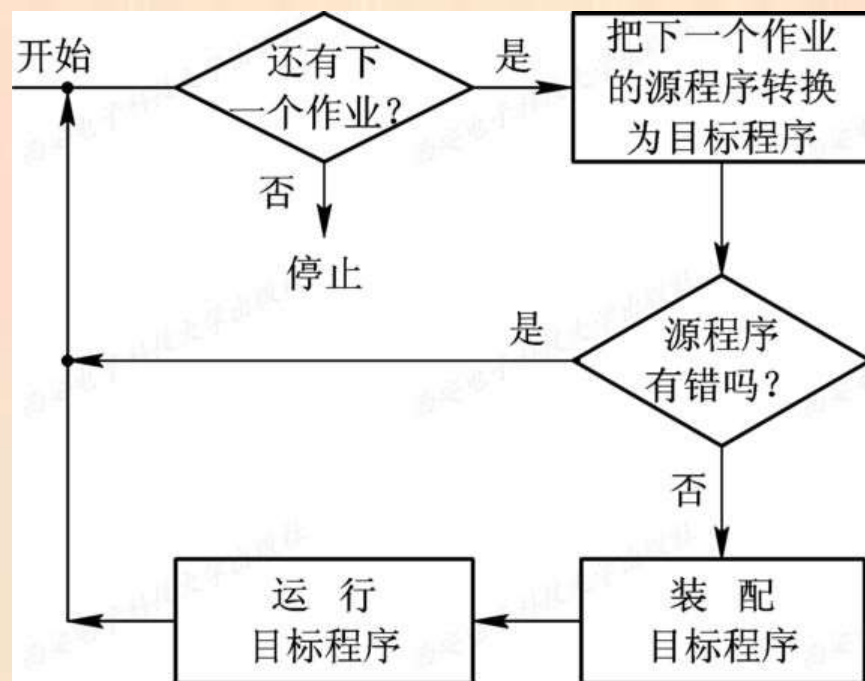
- | 提高系统资源的使用效率；提高作业吞吐量

单道批处理系统；多道批处理系统

- 1、语文P34
- 2、数学P25
- 3、英语P36

1.2.2 单道批处理系统(simple batch processing)

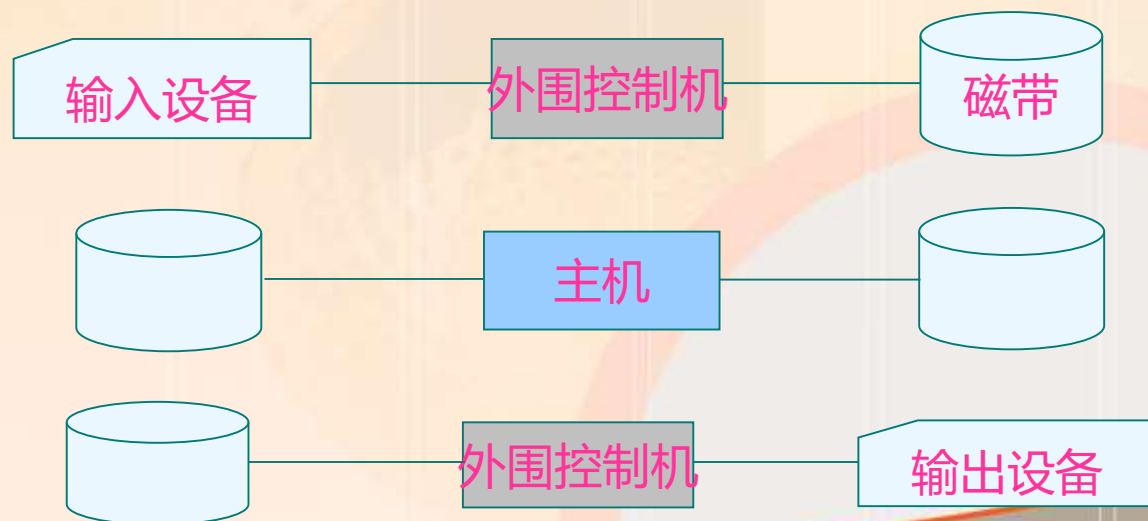
为对作业的连续处理，需要先把一批作业以脱机方式输入到磁带上，并在系统中配上监督程序(Monitor)，在它的控制下，使这批作业能一个接一个地连续处理。



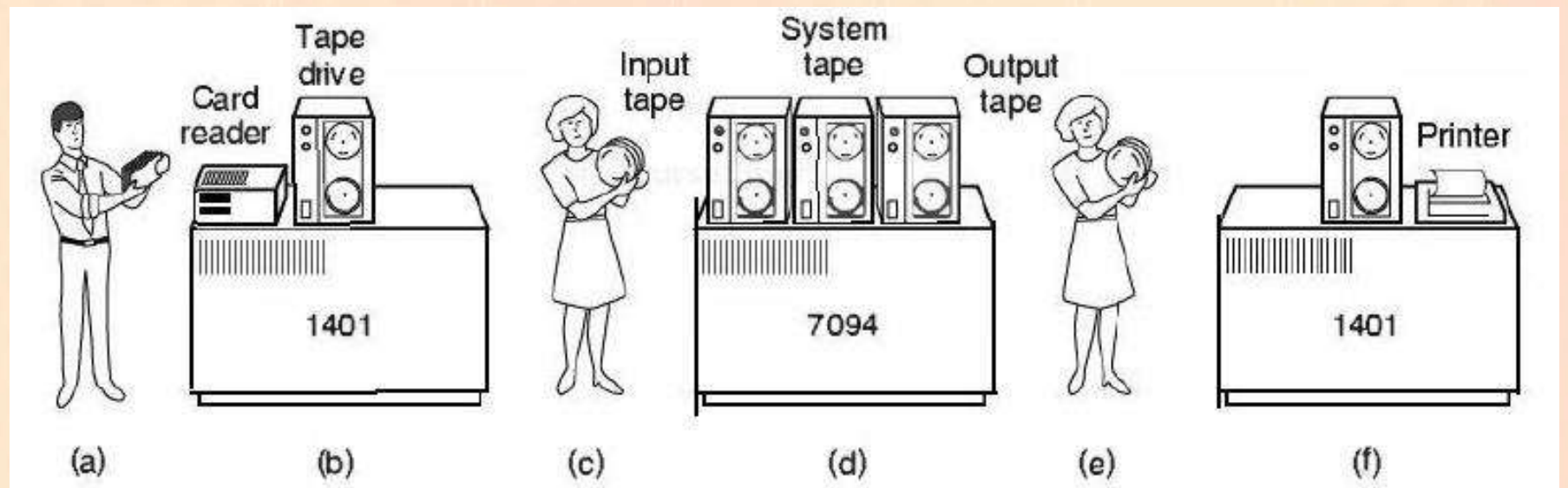
单道批处理系统的处理过程

1.2.2 单道批处理系统(simple batch processing)

- 50年代末 ~ 60年代中（晶体管）：利用磁带把若干个作业分类编成作业执行序列，每个批作业由一个专门的监督程序（Monitor）自动依次处理。可使用汇编语言开发。
- 批处理中的作业的组成：
 - 包括用户**程序**、**数据**和**作业说明书**（作业控制语言）
- **“批”**：
 - 供一次加载的磁带或磁盘，通常由若干个作业组装成，在处理中使用一组相同的系统软件（系统带）



1.2.2 单道批处理系统(simple batch processing)



An early batch system. (a) Programmers bring cards to 1401. (b) 1401 reads batch of jobs onto tape. (c) Operator carries input tape to 7094. (d) 7094 does computing. (e) Operator carries output tape to 1401. (f) 1401 prints output.



1.2.2 单道批处理系统(simple batch processing)

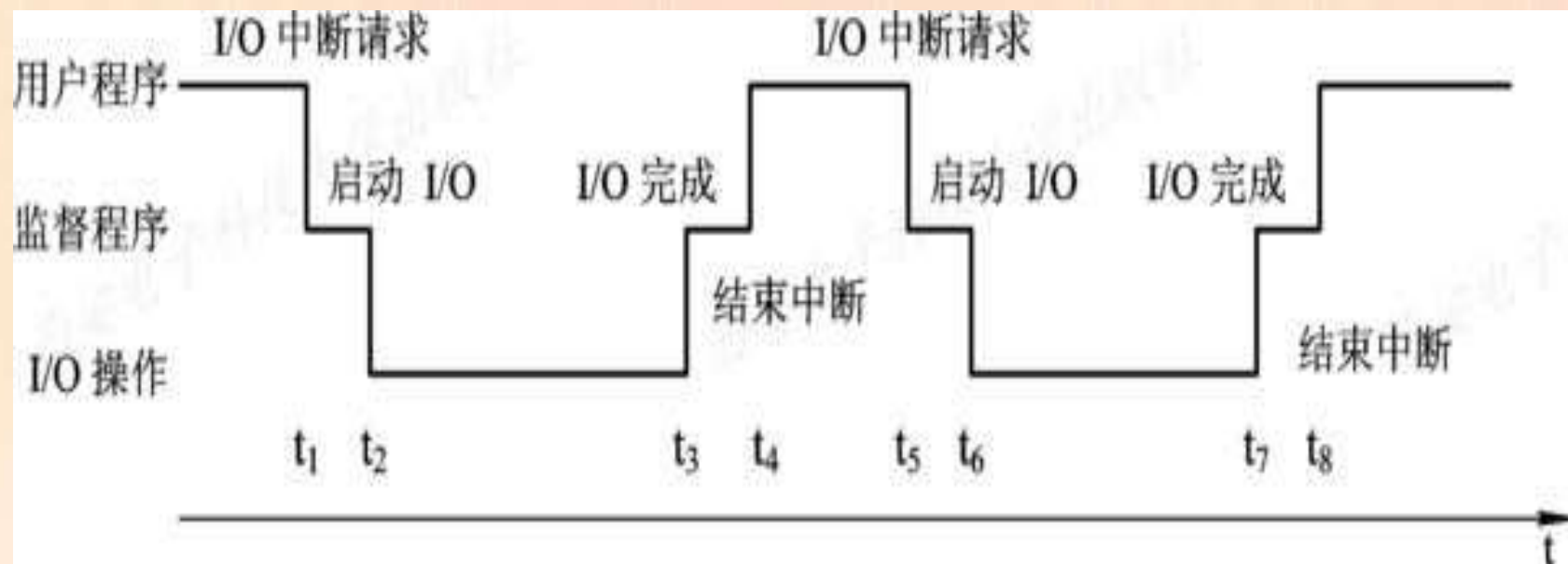
- **优点**

- 同一批内各作业的自动依次更替，改善了主机CPU和I/O设备的使用效率，提高了吞吐量

- **缺点：**

- 内存中仅有一道程序运行，当该程序运行结束或出故障，由监控程序换入另一个程序。不能很好地利用系统资源。
- CPU和I/O设备使用忙闲不均（取决于当前作业的特性）。对计算为主的作业，外设空闲；对I/O为主的作业，CPU空闲。

1.2.2 单道批处理系统(simple batch processing)



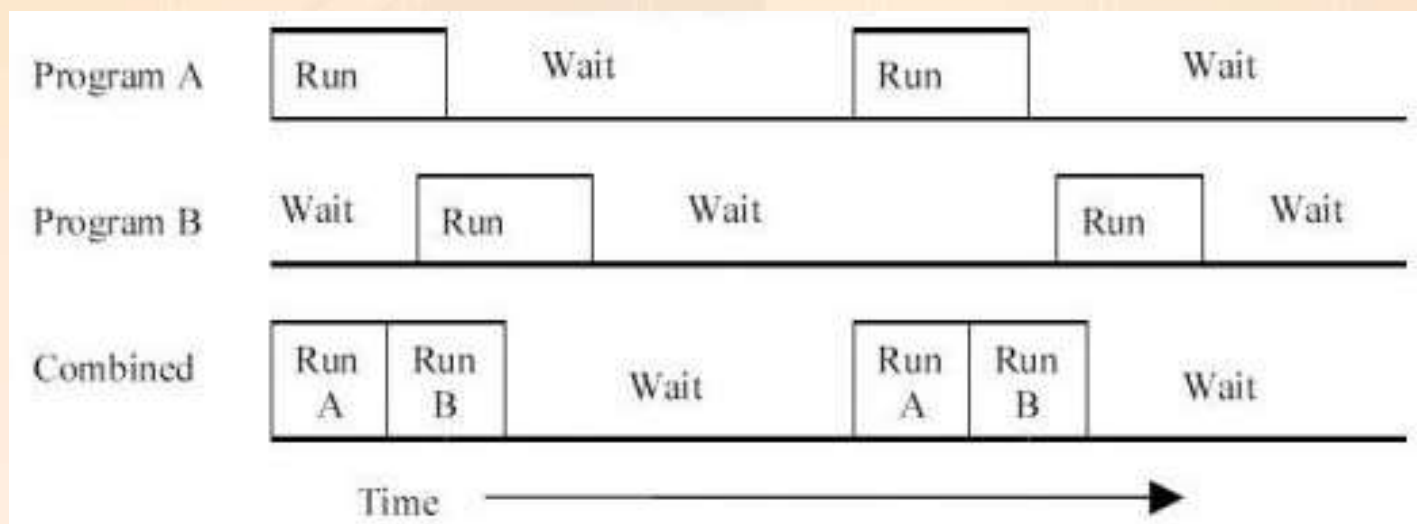
单道程序的运行情况示意图

1.2.2 单道批处理系统(simple batch processing)

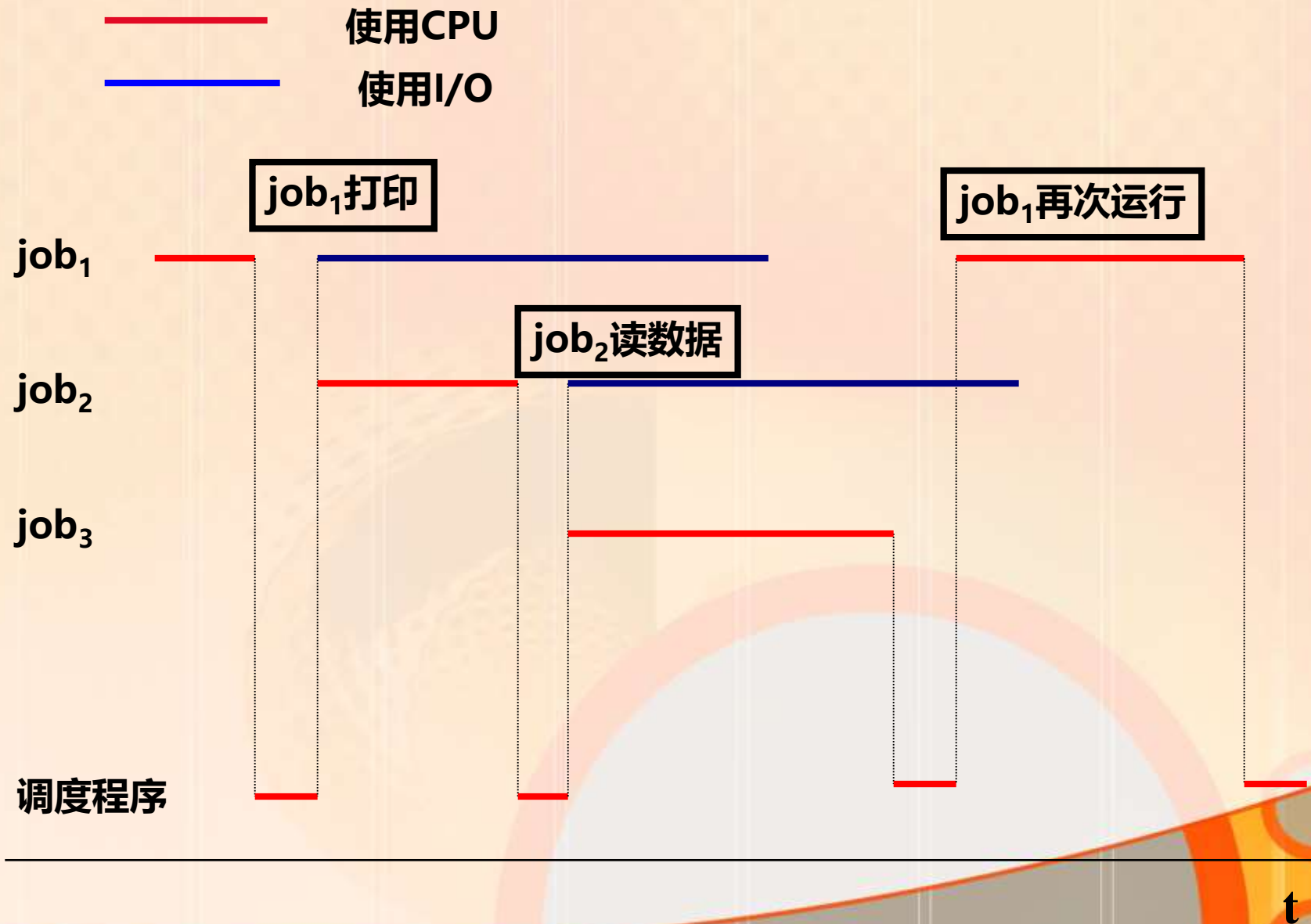
- 单道批处理系统是最早出现的一种OS，严格地说，它只能算作是 OS 的前身而并非是现在人们所理解的OS。尽管如此，该系统比起人工操作方式的系统已有很大进步。
- 系统的主要特征如下：
 - 自动性
 - 顺序性
 - 单道性

1.2.3 多道批处理系统(multiprogrammed batch system)

- 60年代中~70年代中（集成电路），利用多道批处理提高资源的利用率。
- 在该系统中,用户所提交的作业都先存放在外存上并排成一个队列,称为“后备队列”；然后,由作业调度程序按一定的算法从后备队列中选择若干个作业调入内存,使它们共享 CPU 和系统中的各种资源。
- 所以为了进一步提高资源的利用率和系统吞吐量,在20世纪60年代中期引入了多道程序设计技术,由此形成了多道批处理系统。



1.2.3 多道批处理系统(multiprogrammed batch system)



1.2.3 多道批处理系统(multiprogrammed batch system)

特点：

多道：内存中同时存放几个作业；

宏观：并行运行，都处于运行状态；

微观：串行运行，各作业交替使用CPU。

优点：

资源利用率高：CPU和内存利用率较高；

作业吞吐量大：单位时间内完成的工作总量大。

缺点：

用户交互性差：整个作业完成后或中间出错时，才与用户交互，不利于调试和修改；

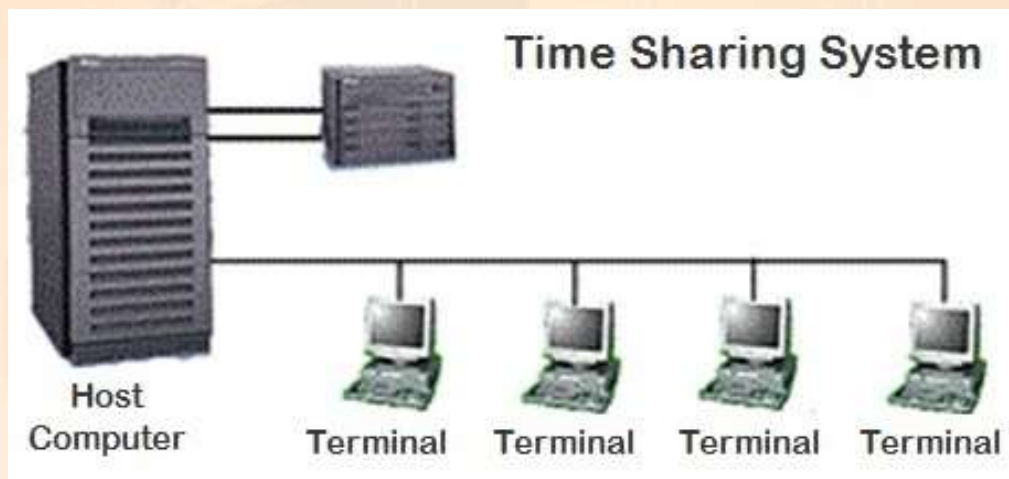
作业平均周转时间长：短作业的周转时间显著增长。

1.2.3 多道批处理系统(multiprogrammed batch system)

- **多道程序系统的技术问题：**
 - **并行程序的运行需要共享软硬件资源，需要同步和互斥机制。**
 - **多道程序需要提高内存的使用效率，需要覆盖技术、交互技术、虚拟存储等技术。**
 - **多道程序在内存中要保证系统存储区和用户存储区安全可靠，需要内存保护。**
- **操作系统经过20世纪60年代、70年代的发展，到20世纪80年代已趋于成熟。批处理操作系统、分时操作系统、实时操作系统是操作系统的三个基本类型。**

1.2.4 分时系统(time-sharing system)

- 70年代中期至今
- “分时”：
 - 多个用户分享使用同一台计算机。每个用户给一定的时间运行，然后切换到另一个用户，一个轮转周期在宏观上相当短，用户感觉不到。



1.2.4 分时系统(time-sharing system)

- 分时系统的基本思想
 - **时间片 (time slice)**：是把计算机的系统资源（尤其是 CPU 时间）进行时间上的分割，每个时间段称为一个时间片，每个用户依次轮流使用时间片。
 - **分时技术**：把处理机的运行时间分为很短的时间片，按时间片轮流把处理机分给各联机作业使用。
 - **分时操作系统**：是一种联机的多用户交互式的操作系统。一般采用时间片轮转的方式使一台计算机为多个终端服务。对每个用户能保证足够快的响应时间，并提供交互会话能力。
 - **设计目标**：对用户的请求及时响应，并在可能条件下尽量提高系统资源的利用率。



1.2.4 分时系统(time-sharing system)

- 分时系统特征：
 - **交互性**：用户与系统进行人机对话。
 - **多路性**：多用户同时在各自终端上使用同一CPU。
 - **独立性**：用户可彼此独立操作，互不干扰，互不混淆。
 - **及时性**：用户在短时间内可得到系统的及时回答。
- 影响响应时间的因素：
 - 终端数目多少
 - 时间片的大小
 - 信息交换量
 - 信息交换速度

1.2.5 实时系统(real-time system)

- **实时操作系统：**
 - 所谓“实时”即“及时”，是指系统能**及时**(或即时)响应外部事件的请求，在**规定的时间内**完成对该事件的处理，并控制所有实时任务协调一致地运行。
 - 必须保证实时性和高可靠性，对系统的效率则放在第二位。
- **分类：**
 - **实时控制系统：**通常是指以计算机为中心的生产过程控制系统，又称计算机控制系统。
 - 例如：铁冶炼和钢板轧制的自动控制、炼油、化工生产过程的自动控制等。
 - **实时信息处理系统：**计算机及时接收从远程终端发来的服务请求，根据用户提出的问题对信息进行检索和处理，并在很短时间内对用户做出正确回答。
 - 例如：机票订购系统、情报检索系统等。

1.2.5 实时系统(real-time system)

- **实时任务的类型：**
 - **按任务执行时是否周期性**
 - **周期性实时任务**：温度控制
 - **非周期性实时任务**：截止时间(deadline)，开始截止时间（最晚开始时间）和完成截止时间（最晚完成时间）
 - **按截止时间的要求**
 - 硬实时任务(hard real-time task)
 - 软实时任务(soft real-time task)

实时系统与分时系统的比较

	实 时 系 统	分 时 系 统
多路性	体现在对多路的现场信息进行采集、对多个对象或多个执行机构进行控制。	按分时原则为多个终端用户服务。
独立性	每个终端用户向系统提出服务请求时，彼此独立操作，互不干扰；对信息的采集和对对象的控制也彼此互不干扰。	每个用户各占一个终端，彼此互不干扰，独立操作，。
及时性	以控制对象所要求的开始截止时间或完成截止时间来确定，一般为秒级。	用户的请求能在短时间（用户能接受的）内获得响应。
交互性	人与系统的交互，仅限于访问系统中某些特定的专用服务程序。	用户与系统进行广泛的人机对话，系统能向终端用户提供数据处理服务，资源共享等服务。
可靠性	高度可靠	可靠程度较低



三种基本操作系统的比较

	多路性	独立性	及时性	交互性	可靠性
批处理系统	无	无	差 (天,时)	差	一般
分时系统	多终端服务	有	好 (分,秒)	好	可靠
实时系统	多路采集、 多路控制	有	最好 (ms,μs)	一般	高度 可靠

- 同时兼有批处理、分时、实时处理的功能，或同时兼有其中两个以上功能的操作系统称为**通用操作系统**。
- 可适用于计算、事务处理等多种领域，能运行在多种硬件平台上，如 UNIX系统、Windows 等。 - - 通用化、小型化。
- 操作系统经过20世纪60年代、70年代的发展，到20世纪80年代已趋于成熟，随着VLSI（超大规模集成电路，very-large-scale integration）和计算机体系结构的发展，操作系统随之发展，形成了**多处理机操作系统、微机操作系统、网络操作系统、分布式操作系统和嵌入式操作系统**等。

1.2.6 多处理机操作系统(Multi-processor Operating System)

- 75年前后出现，为提高计算机系统性能和可靠性。提高性能有两条途径：提高各个组成部分的速度、增大处理的并行程度。
- 多处理机系统的优点：
 - 增加系统的吞吐量。但N个处理器加速比达不到N倍（额外的调度开销，算法的并行化）
 - 提高系统可靠性：故障时系统降级运行
- 多处理机系统的类型：
 - **紧密耦合(tightly-coupled)** - - 并行（parallel）系统：各处理机之间通过快速总线或开关阵列相连，共享内存，整体系统由一个统一的OS管理（一个OS核心）。
 - **松散耦合(loosely-coupled)** - - 分布式系统：各处理机带有各自的存储器、I/O设备和操作系统，通过通道或通信线路相连。每个处理机上独立运行OS。

1.2.6 多处理机操作系统(Multi-processor Operating System)

- 并行系统的类型

- **非对称式多重处理(Asymmetric Multiprocessing, ASMP)**：又称主从模式(Master-slave mode)。

- 主处理器：只有一个，运行OS。管理整个系统的资源，为从处理器分配任务；
 - 从处理器：可有多多个，执行应用程序或I/O处理。
 - 特点：不同性质任务的负载不均，可靠性不够高，不易移植（通常要求硬件也是“非对称”）。

- **对称式多重处理(Symmetric Multiprocessing, SMP)**：OS交替在各个处理器上执行。任务负载较为平均，性能调节容易 - - “傻瓜式”

1.2.7 网络操作系统(NOS, Network Operating System)

- 网络操作系统是在通常操作系统功能的基础上提供**网络通信**和**网络服务**功能的操作系统。
- 网络操作系统为网上计算机进行方便而有效的网络资源共享，提供网络用户所需各种服务的软件和相关程序的集合。
- 计算机网络
 - 一些自主的计算机系统，通过通信设施相互连接，完成信息交换、资源共享、互操作和协同工作等功能。
- 引入计算机网络的目的是：完成新的应用（进行自动的信息交换），提高性能 - 价格比（共享昂贵资源）

1.2.7 分布式操作系统(Distributed Operating System)

- 分布式系统：处理和控制是分散的（相对于集中式系统）
- 分布式系统是以计算机网络为基础的，它的基本特征是处理上的分布，即功能和任务的分布。分布式操作系统的所有系统任务可在系统中任何处理机上运行，自动实现全系统范围内的任务分配并自动调度各处理机的工作负载。



分布式操作系统与网络操作系统的比较

(1)耦合程度：

- 分布式系统是紧密耦合系统：分布式OS是在各机上统一建立的“OS同质”，直接管理CPU、存储器和外设；统一进行全系统的管理；
- 网络通常容许异种OS互连，各机上各种服务程序需按不同网络协议“协议同质”。

(2)并行性：

- 分布式OS可以将一个进程分散在各机上并行执行“进程迁移”；
- 网络则各机上的进程独立。

(3)透明性：用户是否知道或指定资源在哪个机器上（如CPU、内存或外设）。

- 分布式系统的网络资源调度对用户透明，用户不了解所占有资源的位置；
- 网络操作系统中对网络资源的使用要由用户明确指定；

(4)健壮性：分布式系统要求更强的容错能力（工作时系统重构）

1.2.9 个人计算机操作系统(Personal Computer Operating System)

- 针对单用户使用的个人计算机进行优化的操作系统。
- 个人计算机操作系统的特征
 - 应用领域：事务处理、个人娱乐，
 - 系统要求：使用方便、支持多种硬件和外部设备（多媒体设备、网络、远程通信）、效率不必很高，但响应性（交互性）要好。

- **单用户单任务**：只允许一个用户上机，且只允许用户程序作为一个任务运行。
 - 最具代表性有CP/M和MS-DOS：
 - CP/M：在1975年由Digital Research公司率先推出的8位微机OS，具有较好的层次结构、可适应性、可移植性及易学易用性，是8位微机OS的标准。
 - MS-DOS：该操作系统在CP/M的基础上进行了较大的扩充，增加了许多内、外部命令，使该操作系统具有较强的功能及性能优良的文件系统。随着IBM-PC机及其兼容机的畅销，MS-DOS成为16位单用户单任务OS的标准MS-DOS

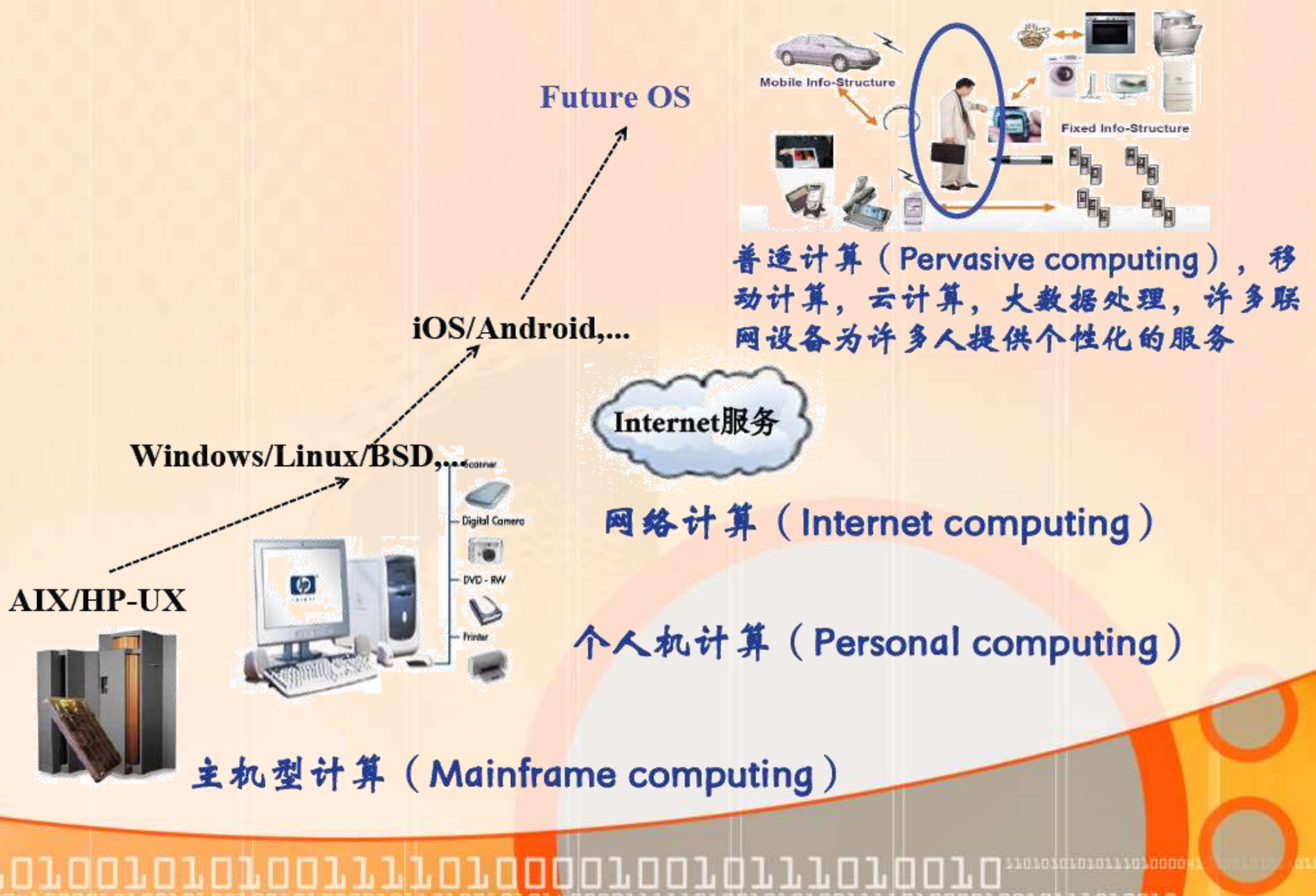


- **单用户多任务**：只允许一个用户上机，但允许将一个用户程序分为若干个任务并发执行，从而有效地改善系统的性能。
 - 最具代表性的是OS/2和WINDOWS：
 - OS/2：1987年4月，IBM同时发表了OS/2和下一代个人系统PS/2。最初版本OS/21.X针对80286开发，属16位微机OS，但已能实现真正的多任务处理。后来的OS/22.X版本，针对80386和80486，属32位微机OS。
 - WINDOWS：Microsoft提供的一种图形用户界面（GUI）方式的操作系统。1990年5月推出了WINDOWS 3.0，1993年推出的WINDOWS NT，之后又推出 WINDOWS 95/ WINDOWS 98/2000/XP。WINDOWS采取了动态内存管理方式，对应用程序以段为单位进行管理。

- **多用户多任务** : UNIX(SCO UNIX, Solaris x86, FreeBSD) , Windows 7



总结：操作系统的发展历史



总结：操作系统分类

批处理系统（单道批处理、多道批处理）

分时系统

实时系统

多处理机操作系统(Multi-processor Operating System)

网络操作系统(NOS, Network Operating System)

分布式操作系统(Distributed Operating System)

个人计算机操作系统

1.3 操作系统的特征和服务

- **操作系统基本特征**
 - **并发**
 - **共享**
 - **虚拟**
 - **异步**

1.3.1 操作系统的特征

1. 并发(concurrency)

- | 多个事件在**同一时间段**内发生。操作系统是一个并发系统，各进程间的并发，系统程序与应用程序间的并发。操作系统要完成这些并发过程的管理。

| 并行(parallel)

- | 是指在**同一时刻**发生。在多道程序处理时，宏观上并发，微观上交替执行（在单处理器情况下）。

1.3.1 操作系统的特征

2.共享(sharing)

- | 多个进程共享有限的计算机系统资源。操作系统要对系统资源进行合理分配和使用。资源在一个时间段内交替被多个进程所用。
- | 互斥共享（如音频设备，打印机）：资源分配后到释放前，不能被其他进程所用。
- | 同时访问（如可重入代码，磁盘文件）

并发和共享是操作系统两个最基本的特征，这两者之间又是互为存在条件的。

- | 资源共享是以程序的并发为条件的，若系统不允许程序并发执行，自然不存在资源共享问题。
- | 若系统不能对资源共享实施有效的管理，也必将影响到程序的并发执行，甚至根本无法并发执行。

1.3.1 操作系统的特征

3.虚拟(virtual)

- 一个物理实体映射为若干个对应的逻辑实体 - - **分时或分空间**。虚拟是操作系统管理系统资源的重要手段，可提高资源利用率。
- CPU - - 每个用户（进程）的“虚处理机”
- 存储器 - - 每个进程都占有的地址空间（指令 + 数据 + 堆栈）
- 显示设备 - - 多窗口或虚拟终端(virtual terminal)

4.异步性(asynchronism)

- 也称不确定性，指进程的**执行顺序和执行时间的不确定性**；
 - 进程的**运行速度不可预知**
 - 判据：无论快慢，应该结果相同 - - 通过**进程互斥和同步手段**来保证
 - 难以重现系统在某个时刻的状态（包括重现运行中的错误）
 - 性能保证：实时系统与分时系统相似，但通过**资源预留**以保证性能

