

计算机组织结构

1 计算机系统概述

刘博涵

2023年9月5日



南京大學
NANJING UNIVERSITY

教材对应章节

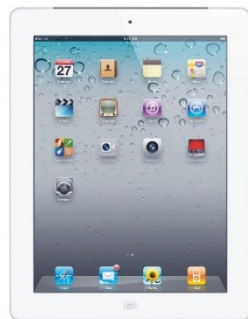


第1章 计算机系统概述



第1章 导论 第2章 计算机的演变和性能

计算机无处不在



什么改变了，什么没有变？

回顾: 什么是计算机?



什么是计算机？

- 计算机是指 “**通用电子数字计算机**（ general-purpose electronic digital computer ）”
 - **通用**：不是一种专用设备
 - 所有计算机在给予足够时间和容量存储器的条件下，都可以完成同样的计算
 - 当希望完成新的计算时，不需要对计算机重新设计
 - **电子**（非机械）：采用电子元器件
 - **数字**（非模拟）：信息采用数字化的形式表示
- 计算机系统
 - **硬件**：处理器，存储器，外部设备，……
 - **软件**：程序，文档，……



什么是“组织”与“结构”？

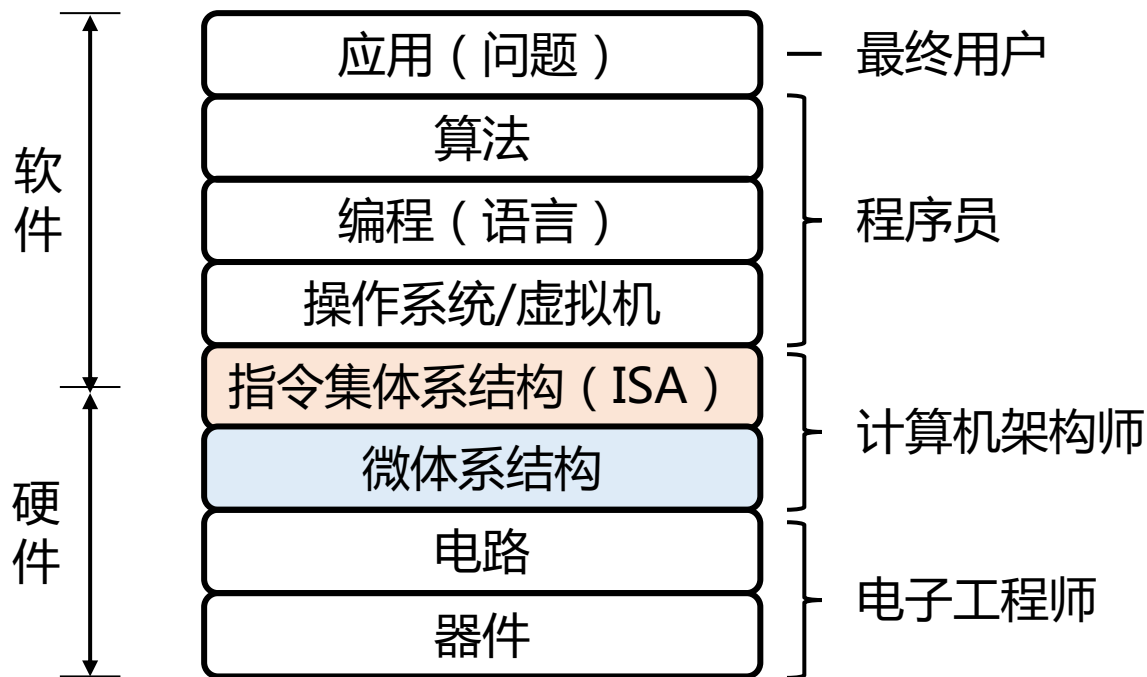


计算机系统抽象层

计算机体系结构即为ISA

ISA是对硬件的抽象
所有软件功能都建立在ISA之上

**微体系结构即为
计算机硬件组成**



组织与结构

组织 (Organization) : 对编程人员不可见

- 操作单元及其相互连接
- 包括：控制信号，存储技术，.....
 - 例如：实现乘法是通过硬件单元还是重复加法？

结构 (Architecture) : 对编程人员可见

- 直接影响程序逻辑执行的属性
- 包括：指令集，表示数据类型的位数，.....
 - 例如：是否有乘法指令？



指令集体系结构 (ISA)

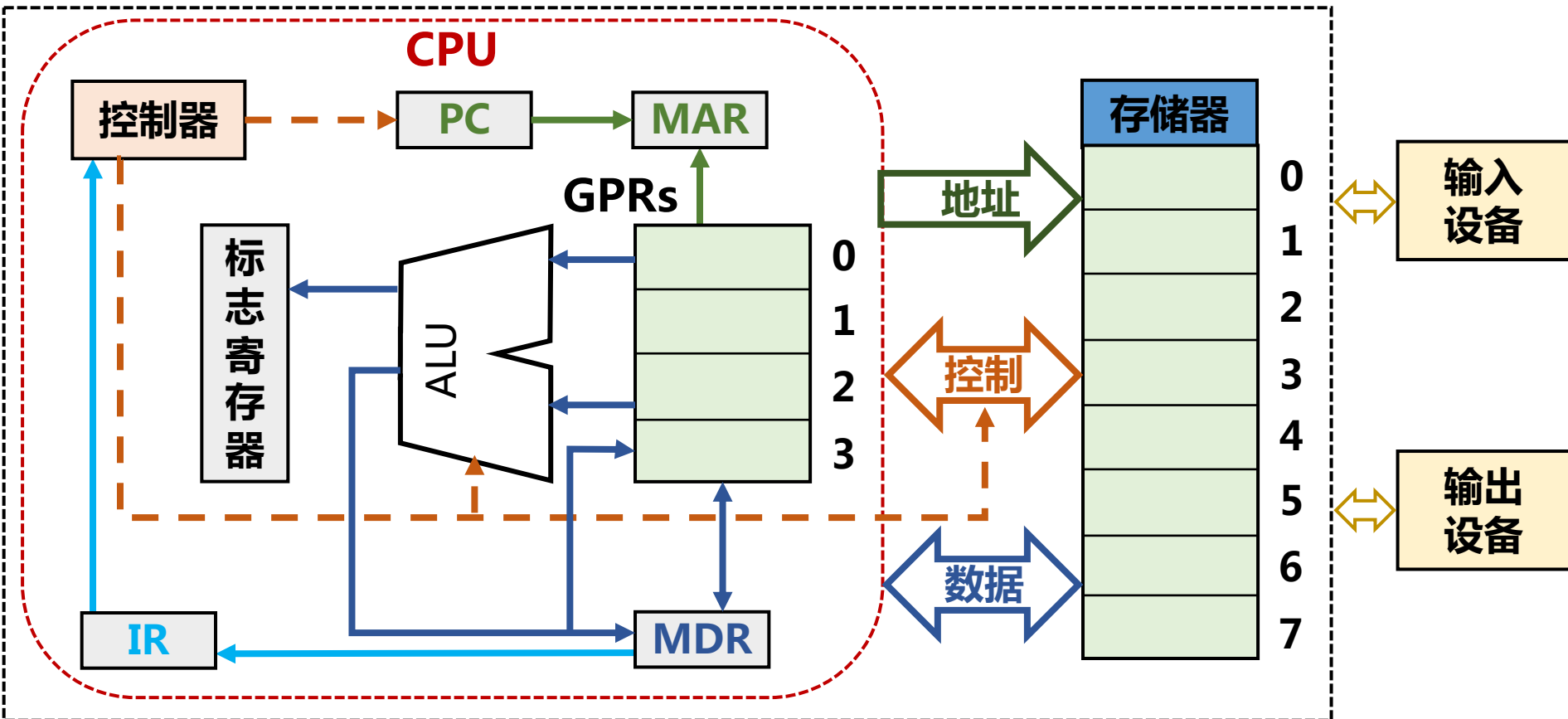
Instruction Set Architecture (ISA) , 有时简称为**指令系统**。常与**计算机体系结构混用**。

- ISA是一种规约 (Specification) , 它规定了**如何使用硬件**
 - 可执行的指令的集合, 包括**指令格式**、**操作种类**以及每种操作对应的操作数的相应规定;
 - 指令可以接受的**操作数类型**;
 - 操作数所能存放的寄存器组的结构, 包括每个**寄存器的名称、编号、长度和用途**;
 - 操作数所能存放的**存储空间的大小和编址方式**;
 - 操作数在存储空间存放时按照**大端还是小端方式存放**;
 - 指令获取操作数的方式, 即**寻址方式**;
 - 指令执行过程的控制方式, 包括**程序计数器 (PC)、条件码定义**等。



ISA在**通用计算机系统是必不可少**的一个**抽象层**

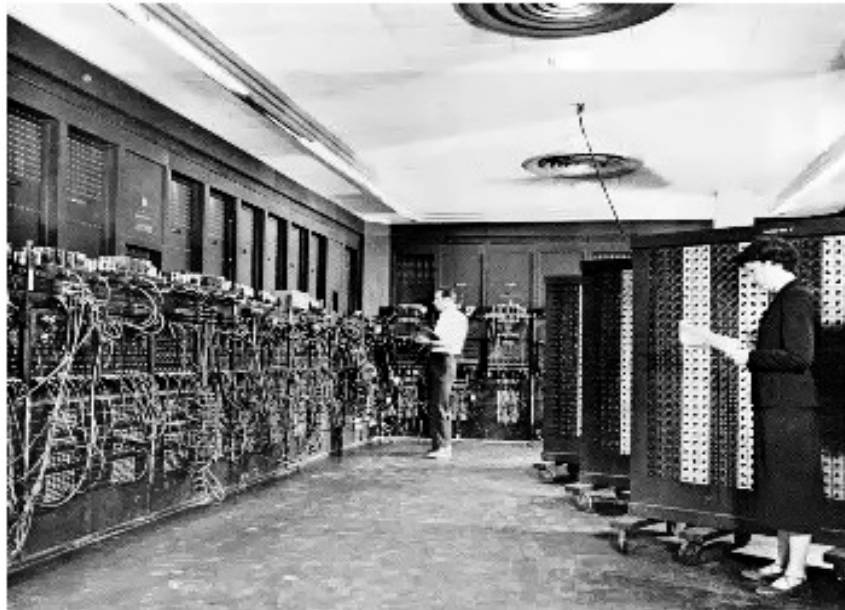
ISA与组成（微结构）之间的关系



不同ISA规定的指令集不同，如IA-32，MIPS，ARM等
计算机组成必须能够实现ISA规定的功能，如提供GPR、标志、运算电路等
同一种ISA可以有不同的计算机组成，如乘法指令可用ALU或乘法器实现

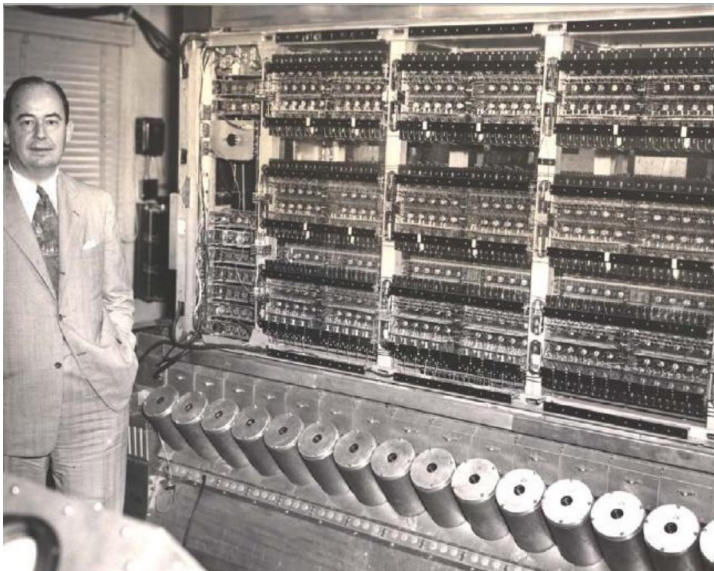
计算机简史

- 第一代：真空管（1946-1957）
 - ENIAC（1946-1955）：第一台**通用**计算机，十进制，手动编程
 - Electronic Numerical Integrator And Computer
 - ABC（1937）：世界上第一台**电子**计算机，不可编程
 - Atanasoff-Berry Computer

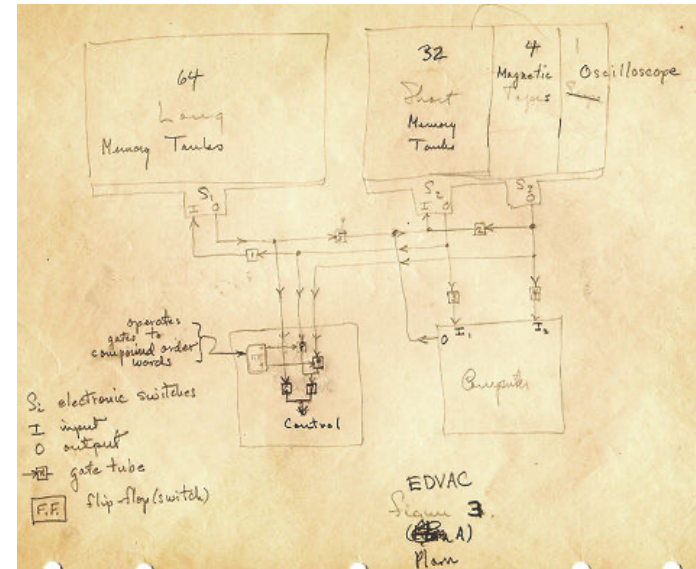


计算机简史

- 第一代：真空管 (1946-1957)
 - EDVAC (1944-1951) : **冯·诺伊曼结构**
 - Electronic Discrete Variable Automatic Computer



冯·诺伊曼 (von Neumann)



The First Draft Report on the EDVAC
von Neumann (1945)

回顾: 什么是“冯·诺伊曼结构”？



冯·诺伊曼结构

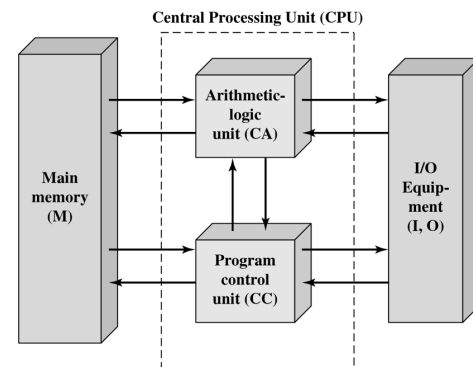
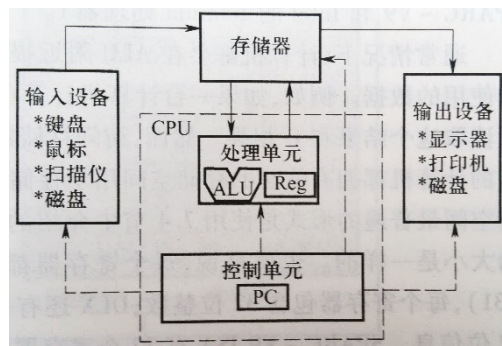
- 又称为“普林斯顿结构”

- 三个基本原则

- 二进制
- 存储程序

- 5个组成部分

- 主存储器：地址和存储的内容
- 算术逻辑单元 / 处理单元：执行信息的实际处理
- 程序控制单元 / 控制单元：指挥信息的处理
- 输入设备：将信息送入计算机中
- 输出设备：将处理结果以某种形式显示在计算机外



计算机简史

- **第二代：晶体管（1958-1964）**

- NCR和RCA, IBM 7000：晶体管体积更小、更便宜、发热更少，而且能以与电子管相同的方式建造计算机
- 采用更复杂的算术逻辑单元和控制器，使用高级编程语言，并为计算机提供了系统软件



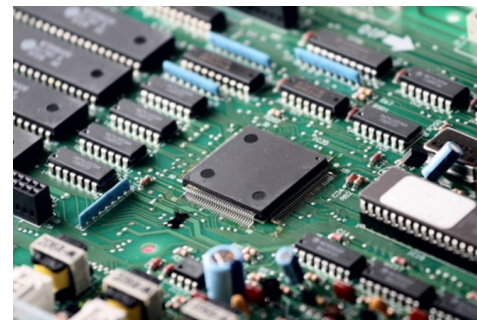
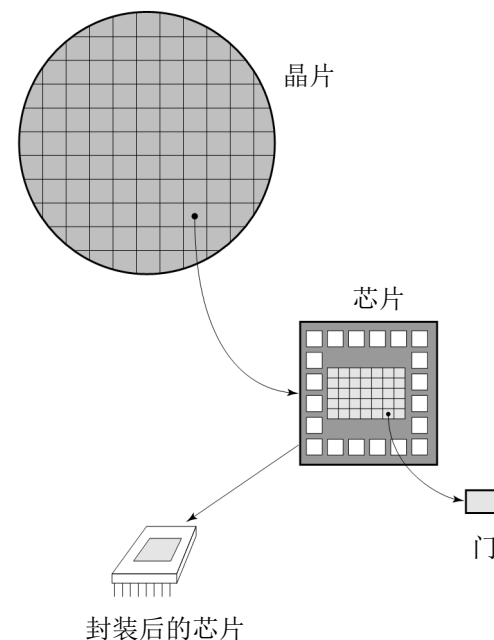
计算机简史

- 第三代：中小规模集成电路（1965-1970）
- 第四代：（超）大规模集成电路（1970-至今）
 - 思想：
 - 将整个电路安装在很小的硅片上，而不是用分立元件搭成的等价电路
 - 晶体管可以通过金属化过程相互连接，以形成电路
 - 规模：
 - 小 → 大 → 超大 → 巨大 ...

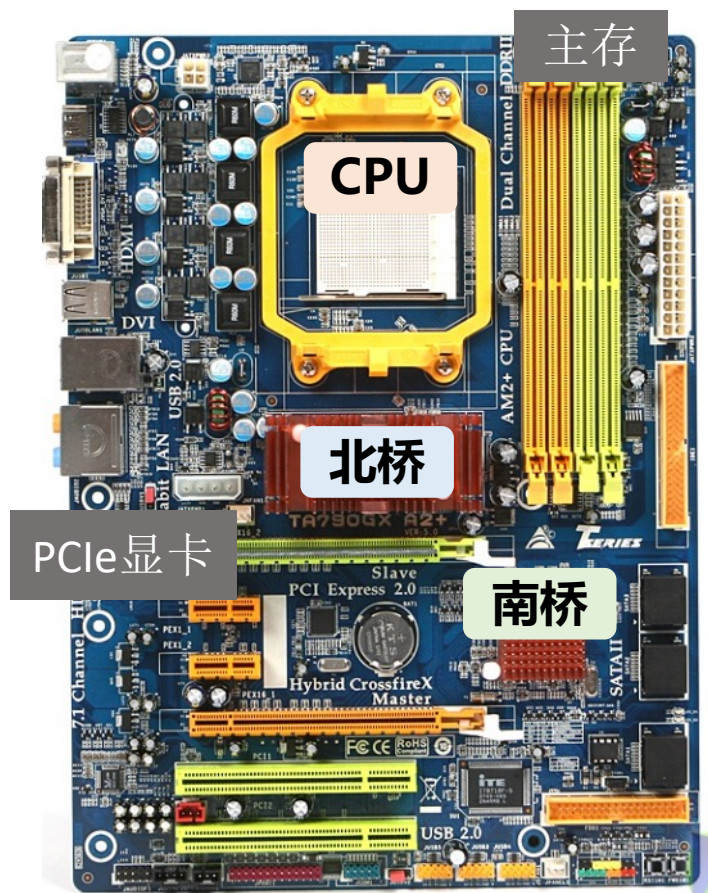
1968年，“软件工程”的概念被提出

1972年，首届软件测试正式会议

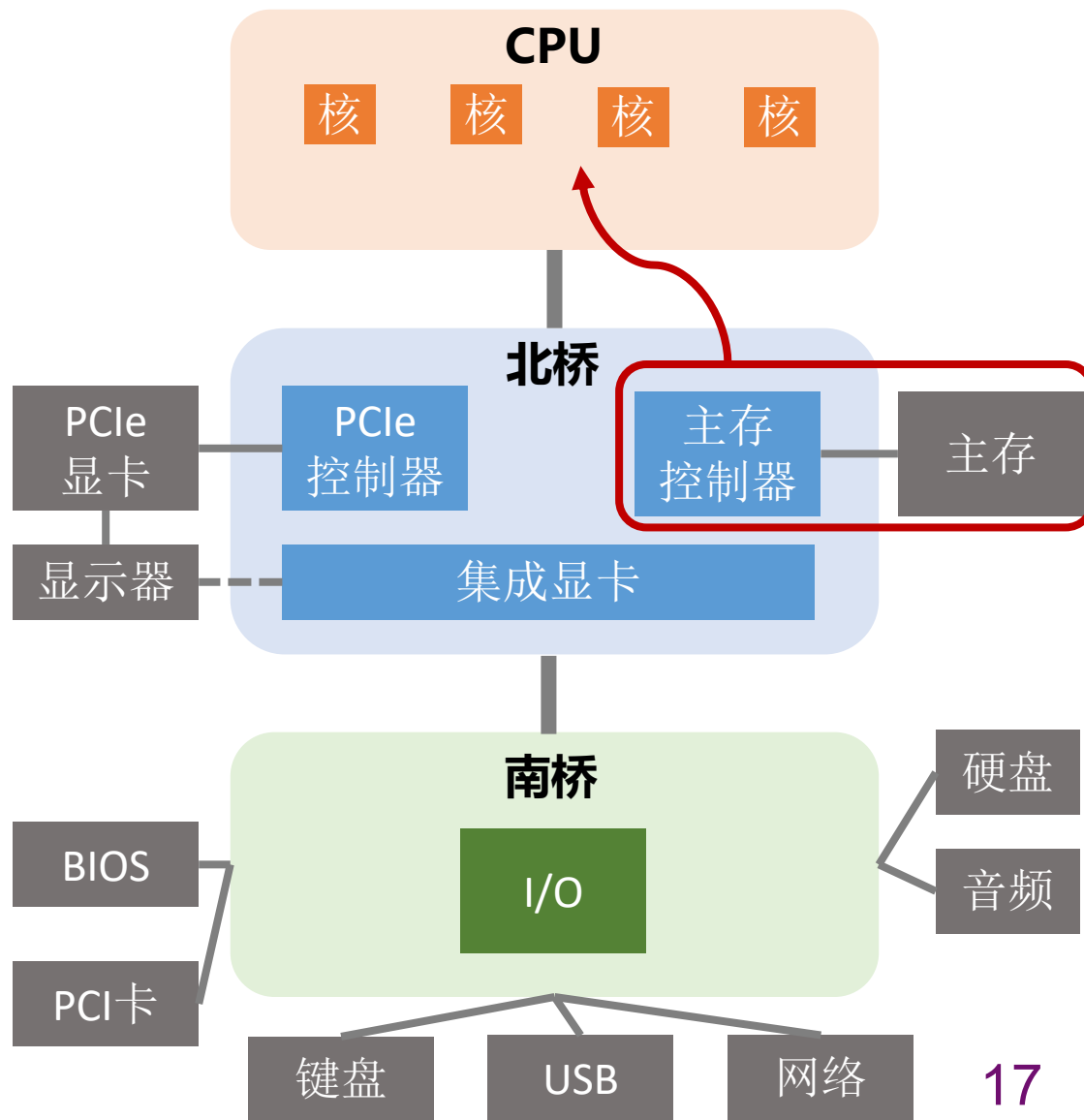
1975年，首届国际软件工程会议ICSE



冯·诺伊曼结构的实现与演变

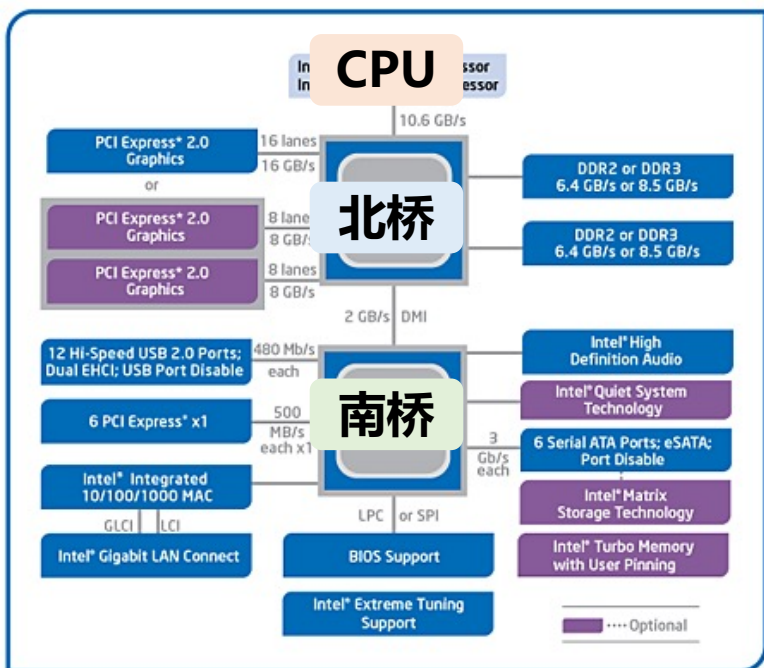


2009年停产



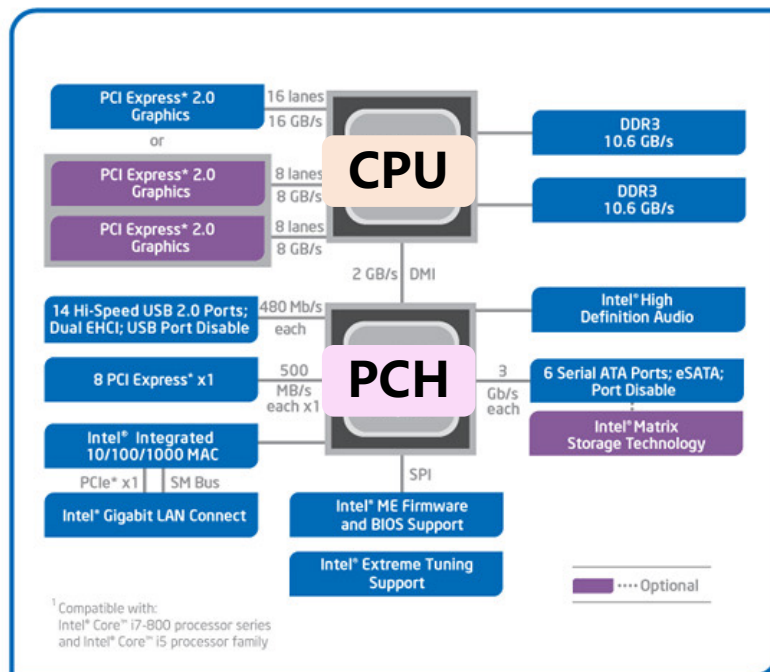
冯·诺伊曼结构的实现与演变

2008年底进入酷睿时代，南北桥变成PCH



Intel® P45 Express Chipset Block Diagram

40系列



¹ Compatible with:
Intel® Core™ i7-800 processor series
and Intel® Core™ i5 processor family

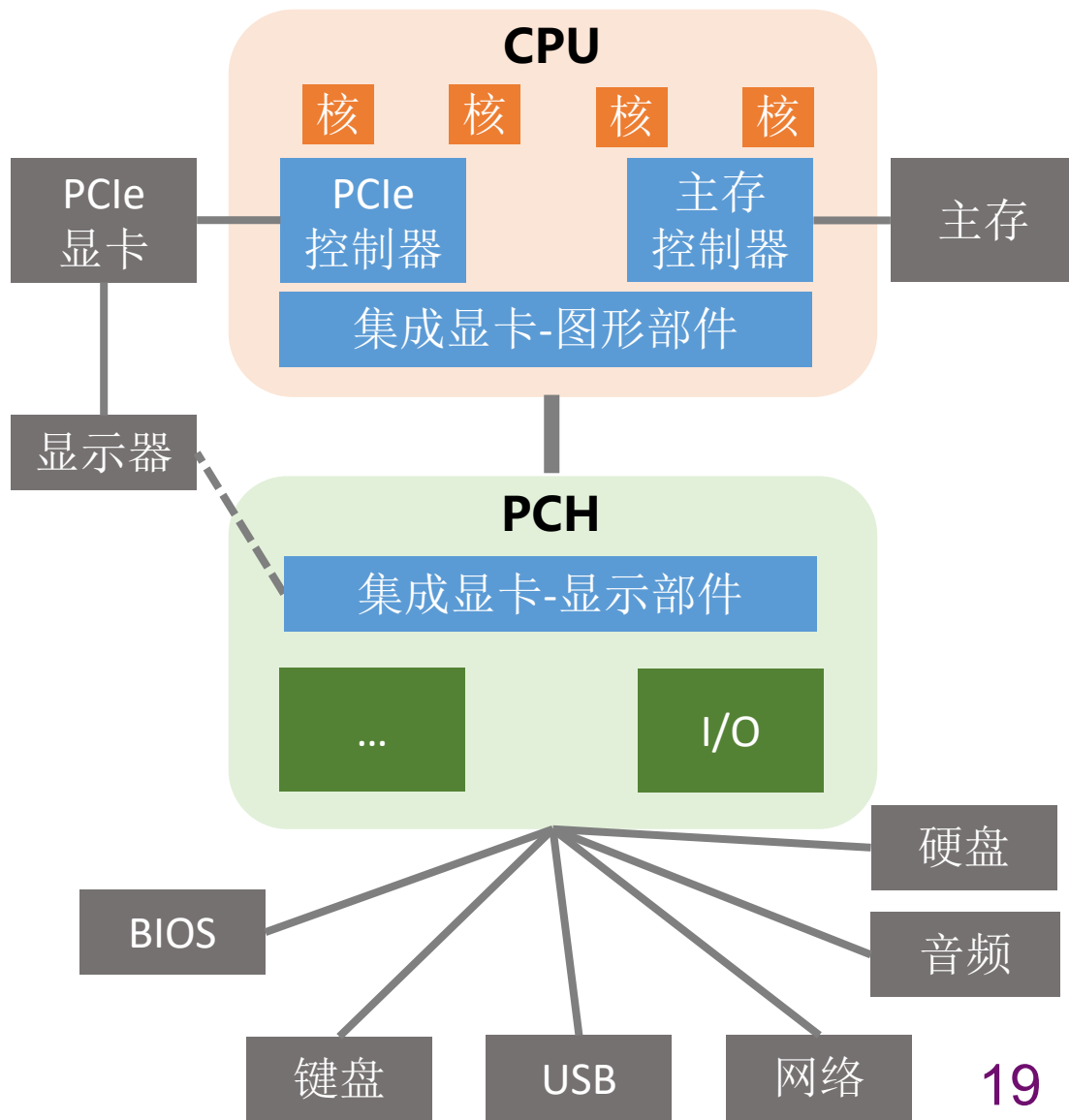
50系列



冯·诺伊曼结构的实现与演变



2022年上市



回顾: 什么是“摩尔定律”？



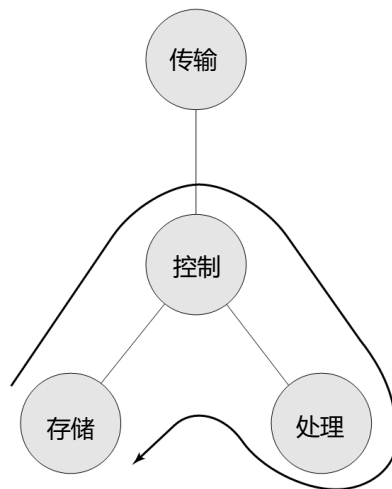
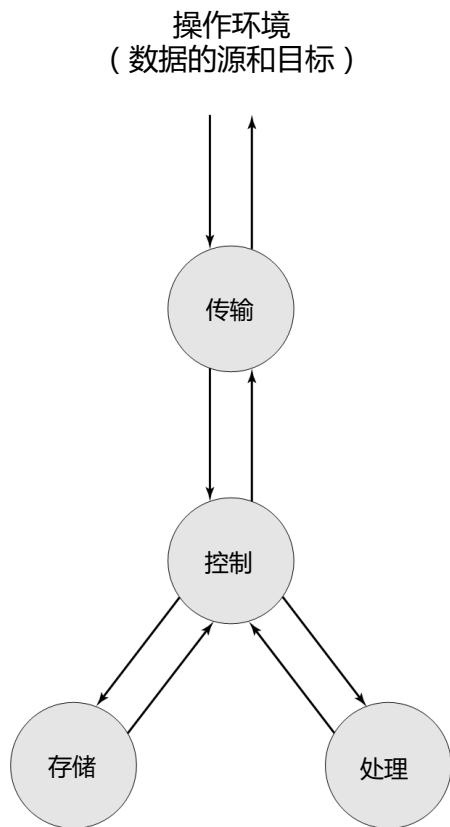
摩尔定律

- **摩尔定律** (Gordon Moore, 1965)
 - 当**价格不变**时，单芯片上所能包含的晶体管数量每年翻一番 (1965-1969) / 1970年起减慢为每18个月翻一番
 - 影响
 - 更小的尺寸带来更多**灵活性**和**可能性**
 - 由于单个芯片的成本几乎不变，计算机逻辑电路和存储电路的**成本显著下降**
 - 减小了对**电能消耗和冷却**的要求
 - 集成电路上的内部连接比焊接更可靠，**芯片间的连接更少**

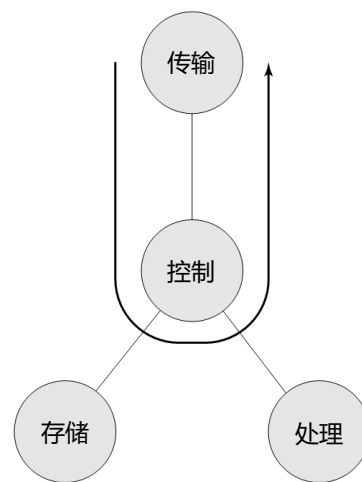


计算机发展：变与不变

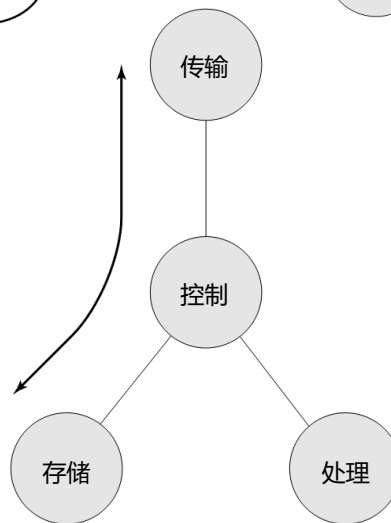
- 基本功能



数据处理



数据传输



数据存储

计算机发展：变与不变

- 运算速度

发展阶段	大致时间	技术	典型速度（每秒的操作次数）
1	1946–1957	真空管（电子管）	40,000
2	1958–1964	晶体管	200,000
3	1965–1971	小规模和中规模 集成电路	1,000,000
4	1972–1977	大规模集成电路	10,000,000
5	1978–1991	超大规模集成电路	100,000,000
6	1991–	巨大规模集成电路	1,000,000,000

2023年，苹果的M2 Ultra芯片做到了**1340亿**晶体管
比2022年发布的M1 Ultra**多200亿**
是52年前的**1亿倍**



计算机性能

- 计算机的关键参数之一
 - 性能，成本，尺寸，安全性，可靠性，能耗，.....
- 性能评价标准
 - CPU：速度
 - 存储器：速度，容量
 - I/O：速度，容量
 -

计算机设计的主要目标是：提高CPU性能



CPU性能

- 系统时钟

- **时钟频率** / 时钟速度（单位：Hz）：计算机在单位时间内（例如1秒钟）执行最基本操作的次数
- **时钟周期** / 周期时间（单位：s）：执行每次最基本操作的时间
 - 时钟滴答（有时也称为“时钟周期”）：CPU 中用于**同步**执行最基本操作的单个电子脉冲
 - 因此，周期时间即为两个电子脉冲之间的时间
 - 时钟周期是时钟频率的倒数



CPU性能

- 指令执行
 - 处理器由时钟驱动，时钟具有固定的频率 f ，或等价于固定的时钟周期 t
 - 如果用 CPI_i 来表示指定类型 i 所需要的周期数，用 I_i 表示在某一给定程序中所执行的 i 类指令的条数
 - 则我们可以计算整个CPI如下：

$$CPI = \frac{\sum_{i=1}^n (CPI_i \times I_i)}{I_c}, \quad I_c = \sum_{i=1}^n I_i$$

执行一个给定程序的处理时间表示为：

$$T = I_c \times CPI \times t$$

$$T = I_c \times [p + [m \times k]] \times t$$

存储器周期时间和处理器周期时间之比

译码和执行指令 存储器访问次数

在处理器和存储器之间传输数据



CPU性能

- 每秒百万条指令 (MIPS) :

$$MIPS = \frac{I_c}{T \times 10^6} = \frac{f}{CPI \times 10^6}$$

- 每秒百万条浮点操作 (MFLOPS) :

$$MFLOPS = \frac{N_{floating-point\ op}}{T \times 10^6}$$



CPU性能

- 基准程序
 - 使用一系列基准程序来测量系统的性能
 - 平均结果：
 - 算数平均值：
$$R_A = \frac{1}{m} \sum_{i=1}^m R_i$$
 - 调和平均值：
$$R_H = \frac{m}{\sum_{i=1}^m \frac{1}{R_i}}$$
 - 几何均值：
$$R_G = (\prod_{i=1}^n R_i)^{\frac{1}{n}}$$



性能设计的基本原则

大概率事件优先原则

- 对于大概率事件（或最常见的事件），赋予它优先的处理权和资源使用权

$$\text{系统加速比} = \frac{\text{总执行时间}_{\text{改进前}}}{\text{总执行时间}_{\text{改进后}}}$$

Amdahl定律

- 加快某部件执行速度所获得的系统性能加速比，受限于该部件在系统中所占的重要性比例
- 性能增加的递减规则**：如果仅仅对计算机中的一部分做性能改进，改进越多，系统获得的效果越小。

$$\text{系统总加速比} = \frac{\text{总执行时间}_{\text{改进前}}}{\text{总执行时间}_{\text{改进后}}} = \frac{1}{(1 - \text{局部占比}) + \frac{\text{局部占比}}{\text{提升的性能}}}$$



性能设计的基本原则

只提高CPU性能够吗？

怎么提高计算机性能？



总结

- 概念
 - 计算机，组织，结构
- 计算机发展历史
 - 真空管 → 晶体管 → 集成电路
 - 冯·诺伊曼结构，摩尔定律，
- 计算机发展
 - 基本功能，运算速度
- 计算机性能
 - CPU性能评价：时钟频率，CPI，MIPS，MFLOPS，基准程序
 - 性能设计的基本原则：大概率事件优先原则 + Amdahl定律



作业

- 作业类型：书面作业
- 发布时间：本周五晚12点前
- 发布平台：教学立方
- 最迟提交时间：下周五晚12点前
- 题目正误不影响评分！！！！
- 不会可以写不知道
- 有空题或未按时提交：0分



教学立方

202302-计算机组织结构-2

如何使用? ^

课程邀请码：QTDCVG2M 刷新 | 关闭加课



谢谢

bohanliu@nju.edu.cn



南京大學
NANJING UNIVERSITY