

# 第 1 章 绪论

在本章里,主要是描述计算机的构成及性能,使读者在开始了解全书内容之前,首先对计算机有一概念上的认识。

## 1.1 计算机的发展史

电子计算机的发明,无疑是人类科学技术发展史上最伟大的发明之一。由于计算机的出现,它深刻地影响着人类的精神文明和物质文明的发展。

尽管在上世纪七十年代前,电子计算机包括电子模拟计算机和电子数字计算机。前者是由模拟电子器件(如模拟加法器、减法器、乘法器、除法器等等)构成,它早已完成了它的历史使命,再无人提及。因此,本书中所描述的就是电子数字计算机并称计算机。电子数字计算机,是指能对离散逻辑符号表示的数据或信息进行自动处理的电子装置。

### 1.1.1 发展经历

#### 1. 第 0 代: 机械计算器

早在二千多年前,我们的祖先就发明了算盘,这是最简单实用的机械计算器。在这漫长的历史中,算盘传遍了全世界,为人类社会做出了巨大贡献。

近代,人们又发明了计算尺及手摇机械计算器。在电子计算器(机)普及之后,它们也逐步退出历史舞台。

#### 2. 第 1 代: 电子管计算机

第 1 代计算机从 1946 年到 1954 年。这一代计算机是由电子管、电磁继电器等器件构成。软件主要使用机器语言编程。典型机器有 ENIAC、IBM 701 等。

#### 3. 第 2 代: 晶体管计算机

第 2 代计算机所在年代为 1955 年到 1964 年。这一代计算机采用晶体管、磁芯存储器等构成。软件上有了监控程序对计算机进行管理,并且开始使用高级语言。这个时期的计算机有很多种,如 IBM 7030、Univac LARC 等。

#### 4. 第 3 代: 集成电路计算机

这一代计算机从 1965 年到 1974 年。这代计算机利用小规模及中规模集成电路芯片,多层印刷电路板及磁芯存储器等构成。在软件上,高级语言迅速发展并出现了分时操作系统。在这个时期里,计算机的应用领域不断扩展,开始向国民经济各部门及军事领域渗透。典型机器如 IBM 360、370, DEC PDP-8 等等。

#### 5. 第 4 代: 超大规模集成电路计算机

从 1975 年到 1990 年出现的是第 4 代计算机,这些计算机是利用大规模、超大规模集成电路构成,其主存也是由超大规模集成电路构成的半导体存储器来实现。这一代计算机在结构上有了很大的发展,在性能上有了很大的提高。

同时,在这一时期由于微细加工技术的发展,超净环境的实现,超纯材料的研制成功,推动着超大规模集成技术的发展。于是,出现了依赖于这种技术的微型计算机、单片微型计算机等。

与硬件发展的同时,这一代计算机的软件飞速发展。出现了许多著名的操作系统,

如 DOS、Windows、Unix 等。

这一时期出现了一些典型的计算机，如 IBM 3090，VAX 9000 等。而这一时期应用最多，也是最广的还是个人微型计算机，诸如 PC 机、苹果机等等。

#### 6. 第 5 代：高性能智能计算机

从 1991 年开始，进入了计算机发展的第 5 代，采用超大规模、超高速集成电路构成计算机。在结构上，构成计算机已从单处理器向多处理器发展。即使构成微型机也采用多核处理器，目前常见的是双核处理器，4 核处理器及更多核的处理器。用这样的多核处理器构成计算机可获得很高的性能。此前，Intel 公司已可以做出一块芯片，内含 80 个核的处理器，用这样的一块 80 核处理器芯片构成的计算机，其速度已超过一万亿次。

可以想象，若用几百、几千、甚至上万块双核（或更多核）处理器芯片构成一台计算机，如集群系统。那么，该计算机系统的性能将是非常高的。例如，目前用这种思路做出来的计算机系统，其速度可达到 1600 万亿次。

第 5 代计算机不仅在速度等性能上不断提高，而且希望计算机更加人性化、智能化，包括能听、会看、会说、有感情等。

第 5 代计算机的发展必定对软件提出更高的要求，因此也必然会促使包括操作系统、应用软件等各种软件的快速发展。

### 1.1.2 摩尔定律

#### 1.1.2.1. 摩尔定律的由来

1965 年 4 月，《电子学》杂志刊登了戈登·摩尔(Gordon Moore)撰写的一篇文章。戈登·摩尔当时是飞兆半导体公司研发部门的主管。文章中讲述了他将 50 个晶体管集成在一块芯片中。并且预言，到 1975 年，就可能将 6.5 万只这样的元件密植在一块芯片上，制成高度复杂的集成电路。

当时，集成电路问世才 6 年，摩尔的预测听起来不可思议。但那篇文章的核心思想——即预测集成电路芯片内可集成的元件差不多每年可增加一倍。在后来的技术发展过程中被证明这一预测是正确的。尽管，现在人们根据几十年走过的技术历程将“摩尔定律”描述为：集成电路芯片的集成度每 18 个月翻一番。经过了四十多年，到今天摩尔定律依然有效，而且许多人确信该定律在未来很多年仍将成立。

摩尔的预言不仅对他本人，而且对整个社会而言都是意义深远的。后来摩尔与他人共同成立了英特尔公司，并通过他所开创的技术创造了无数的财富。

摩尔定律并不是一个物理定律(定律是放之四海而皆准)，而是一种预言。但它鞭策工业界不断地改进，并努力去实现它。从根本上讲摩尔定律是一种自我激励的机制，它让人们无法抗拒，并努力追赶。从人们认识摩尔定律开始，无论是英特尔公司、摩托罗拉公司还是其他的半导体器件公司，无一不是在不断地努力去实现摩尔定律，促使各厂家不断地推出集成度更高的产品。在上世纪九十年代中期，英特尔利用 350nm 技术制造出集成度达 120 万的 80486。但很快线宽就实现了 250nm、180nm、130nm、90nm、65nm。今天，已经可以用 45nm 的生产线制造处理器、DRAM 等器件。今天的处理器的集成度已超过 10 亿。根据摩尔定律，芯片的集成度还会迅速提高。有人曾经说过，集成度提高 100 倍，则相对价格可以已降低 100 倍、性能可提高 100 倍、可靠性也可以提高 100 倍。当然，也许不一定是 100 倍，但是，随

着集成度的提高，性能及可靠性都会大大提高、价格会大大降低将是毋庸置疑的。正是摩尔定律使得计算机日新月异地发展。

归根结底，在 40 多年的实践证明摩尔定律有利于工业的发展及人类的需求。直至今日，半导体工业还是按照 DRAM 每 18 个月、微处理器每 24 个月集成度翻倍的规律发展着。

如果按照旧有方式制造电路，即将晶体管、电阻器和电容器安装在电路板上构成电子设备，那么个人电脑、移动电话、计算机辅助设计等都是不可能问世。

#### 1.1.2.2. 摩尔定律的未来

2004 年以后，半导体工业无论从技术上或者是成本上的挑战都越来越激烈，各种关键问题综合在一起，而且几乎要同时得到解决。如硅片尺寸的继续扩大；新材料、新工艺和新电路结构的采用都使得工业制造难度显著提高。同时为了实现以上的要求，固定资产的投入必然加大，这就导致电路成本急剧增加且工业的风险程度明显上升。

晶体管是处理数字信号的微电子开关。其中栅极用来打开或闭合晶体管，而栅极介质是用来将栅极从电流通道隔离出来的绝缘体底层。过去数十年来，芯片厂商一直采用二氧化硅来做栅极的介质，采用多晶硅做栅电极。但是，使用现有的材料，晶体管的尺寸缩小几乎已达到极限，如 Intel 在其 65nm 制程技术中，已经成功地将二氧化硅栅介质的厚度缩小至 1.2nm（相当于五个原子的厚度），但是栅介质的漏电流也随之迅速增加，这就导致功耗和发热等问题日益严重，让芯片厂商头痛不已。

在最近的处理器中，随着刻蚀线宽从 45nm 过渡到 32nm，晶体管的尺也愈来愈小，其漏电及功耗问题越来越严重，性能和功耗之间的平衡也越来越困难。这是过去近十年里摩尔定律遇到的最大的技术难题，这就引起了人们的疑虑，芯片的集成度还能继续增加吗？摩尔定律是否已走到了尽头？

为了顺利地实现处理器制程由现在的 65nm 向 45nm 再到 32nm 的转变，Intel 采用了新的基于铅的高-K 绝缘体材料，用以取代二氧化硅和多晶硅两种化合物，从而可以大大地减少电流的泄漏。Intel 认为，新材料和新技术的使用可以提升处理器芯片的性能，并能达到降低功耗、减小噪声和降低成本的要求。由于 32nm 晶体管远小于上一代晶体管，因此，晶体管开关所需电量也大为减少，使晶体管功耗大约降低了 30%，栅漏电流减少 10 倍以上。随着晶体管新材料的改进，既可以通过大幅提升 CPU 主频来实现更高性能，也可以通过降低主频来提升能源效率。

同时，微细加工技术的线宽仍在不断的减小，目前生产线上所用的是 45nm 和 30nm。据报道当前实验室所能做到的最小线宽为 9nm。

最新发展的 nm 技术使得制造性能更佳的计算机成为可能，而这已经突破了硅技术所能达到的极限。这项新技术通过以碳为成份的 nm 管来制造元件，而 nm 管的直径只有 10 个原子那么大，是现在的基于硅成份的晶体管体积的 1/500。该技术的出现向制造分子级电子设备的目标迈进了一大步。研究显示以碳 nm 管在性能上不会逊色于硅晶体管，还因为它们的体积要小得多，所以有很大希望成为将来 nm 电子技术的基础。nm 技术的前景是非常广阔，这样的技术会使摩尔定律得以延续。

可以看到，新技术、新材料的出现，使芯片的集成度进一步提高成为可能。尽管集成度不可能无限地增加下去，但人们预计今后十年、二十年甚至三十年，芯片的制造依旧会继续遵循摩尔定律。摩尔定律将会继续激励人向更高的高度攀登。

## 1.2 计算机的基本组成

### 1.2.1 硬件系统

#### 1.2.1.1 早期的冯·诺依曼计算机

计算机是由两大部分：硬件系统和软件系统组成。硬件系统是指计算机中那些看得见摸得着的物理实体。

##### 1. 硬件组成

早期的计算机硬件结构如图 1.1 所示。图 1.1 所示的计算机结构最早是在 1946

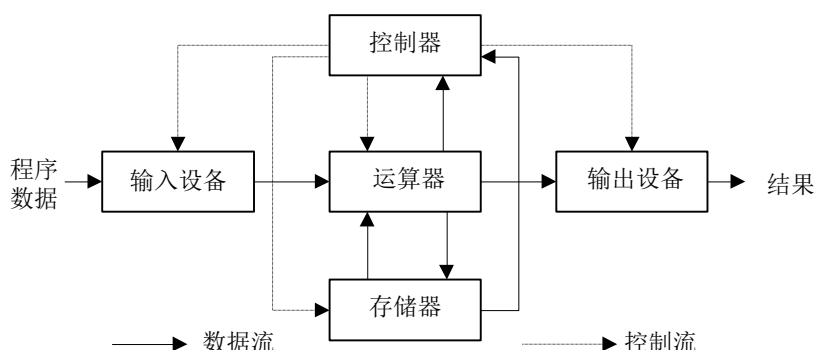


图1.1 早期计算机（硬件）的组成

由冯·诺依曼提出。同时，在此硬件结构的基础上，还提出了计算机是依据存储程序、程序控制的方式工作。这就是冯·诺依曼的计算机设计思想。

当时的计算机由运算器、控制器、存储器、输入设备和输出设备五大部分构成。运算器用以实现算术运算和逻辑运算；控制器根据指令的功能产生相应的控制信号，控制其他部分的工作以便实现指令的功能；存储器用来存放数据和程序；输入设备可将外部的信息输入到计算机中；输出设备可将结果显示或记录下来。

##### 2. 冯·诺依曼计算机的特点

冯·诺依曼计算机工作的基本思想，就是将计算机要处理的问题用指令编成程序，并将程序存放在存储器中。而后，在控制器的控制下，逐条指令从存储器中取出来执行指令。通过执行程序最终解决计算机所要处理的问题。尽管经历了几十年的发展，又出现了新的设计思想，但冯·诺依曼的这种思路直到今天仍然还在广泛地应用。

在冯·诺依曼计算机工作过程中，总是一条指令接一条指令地执行，执行指令会产生控制流，在控制流的驱动下完成指令的功能。在此过程中，数据（流）则是被动地调用。

冯·诺依曼计算机的另一特点就是进入计算机的指令、数据及其他信息均是用二进制编码来表示的。用二进制不仅电路简单、使用方便而且抗干扰性强。因此，二进制一直延用至今。

#### 1.2.1.2. PC 机的结构

计算机的发展已走过了六十多年，尤其是最近三十年，其发展更是日新月异。

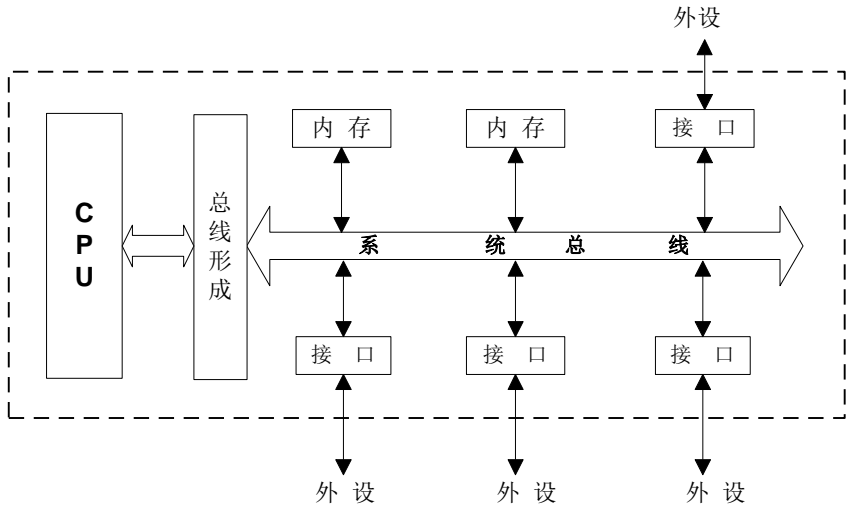


图1.2 微型计算机结构框图

本书的后面将会介绍一系列的计算机结构形式。从中可以看到，在过去的年代里计算机从最简单的冯·诺依曼计算机到现在都有哪些发展与改变，这些发展和改变是由什么原因引起的，是如何实现这些发展它改变的，实现的依据又是怎样的。这些问题正是本书要呈现给读者的内容。

另外，应当特别提及的就是个人计算机—PC机。从1981年PC机诞生以来，由于其规模小、结构简单，人们称其为微型计算机。在二十多年的时间里，PC机一代接一代的发展。现在，PC机已遍布全世界。尽管今天的PC机的功能已十分强大，但人们仍然称之为微型计算机。目前读者所使用的、所看到的绝大多数是PC机。

早期的PC机结构非常简单，其框图可以简化如图1.2所示。该框图也是那个年代微型计算机的基本结构形式。

由图1.2可以看到，中央处理单元CPU是将图1.1中的运算器（算术逻辑单元ALU）和控制器（还有一些寄存器）集成在一块集成电路芯片中。在执行指令的过程中，它可以产生地址信号、数据信号和一系列控制信号，从而形成系统总线信号。图1.2中的存储器也采用大

表 1.1 英特尔微处理器发展历程

年 份	产 品
1971 年	4004 微处理器
1972 年	8008 微处理器
1974 年	8080 微处理器
1978 年	8086—8088 微处理器
1982 年	80286 微处理器
1985 年	80386 微处理器
1989 年	80486 DX 微处理器
1993 年	Pentium 处理器
1995 年	Pentium Pro 处理器
1997 年	Pentium II 处理器
1998 年	Pentium Pro Xeon 处理器
1999 年	Celeron 处理器
1999 年	Pentium III 处理器
1999 年	Pentium III Xeon 处理器
2000 年	Pentium 4 处理器
2001 年	Xeon 处理器
2001 年	Itanium 处理器
2002 年	Itanium 2 处理器
2003 年	Pentium M 处理器
2005 年	Pentium D 处理器
2006 年	Core 2 Duo 处理器（酷睿）

规模集成的半导体存储器芯片构成。外设，包括输入设备、输出设备及外部存储器均经过接口与系统总线相联接。因此，图 1.2 与图 1.1 已有很大的不同，它们是不同时代的计算机。同时，处理器厂家将图 1.2 中用虚线框起来的部分集成在一块芯片中，这就是单片微型计算机。

在摩尔定律的激励下，半导体集成电路，包括 CPU、半导体存储器、接口芯片以及外设所用到的芯片的集成度均以每 18~24 个月翻 1 倍速度飞速发展。尤其值得说明的是 CPU，它也被简称为处理器。各处理器厂家不遗余力地、一代接一代、一种型号接一种型号地被开发研制出来。在这里我们仅将英特尔（Intel）公司的 80X86 系列处理器的发展历程列于表 1.1 中。由于本书篇幅的限制，表中并未列出各处理器的集成度及性能指标。表 1.1 只列到 2006 年，已可以看到处理器的更新速度，每年都会开发出多种芯片。今天的 Core i7 处理器的集成度已达到六核、8.2 亿晶体管、3.2GHz。

在处理器的开发过程中，过去用于大型计算机的许多技术，在摩尔定律的支持下不断地运用到处理器的开发过程中。例如，超标量、超流水、多级 Cache、虚拟技术、RISC、多机系统概念下的多核处理器等等技术均已应用在近几年的处理器中。这些技术正是本书中所要讨论的重要内容。

随着处理器性能的不断提高，从上世纪九十年代开始，PC 机的结构发生了很大的变化。基本上是以主板芯片组为基础构成主板，而后插上相关的部件构成 PC 机。主板芯片组有 1 到 3 块芯片，经常采用 2 片，分别称为北桥和南桥。随着处理器的不断更新换代，芯片组也在不断地更新换代。一般是每开发出一种新的处理器，就会开发出与之相对应的芯片组。目前常见的 PC 机主板框图如图 1.3 所示。

图 1.3 中，北桥芯片是主板芯片组中起主导作用的最重要的组成部分，也称为主桥。北桥芯片主要决定主板的规格、对硬件的支持以及系统的性能，它连接着 CPU、内存、AGP 总线。主板支持哪种处理器，支持何种显卡，支持什么频率的内存条，都是北桥芯片决定的。北桥芯片往往有较高的工作频率，所以发热量较大，故需要为北桥加一个散热器。

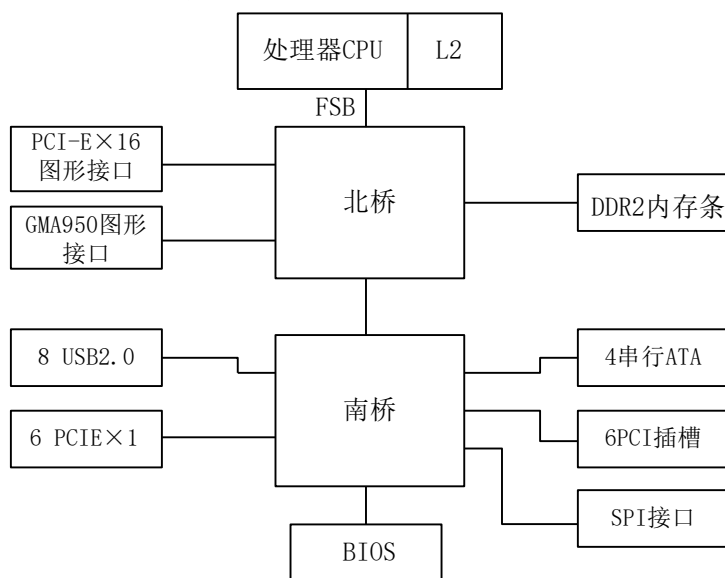


图1.3 PC机主板结构

南桥芯片主要决定主板的接口，该芯片所提供的各种接口（如串口、USB）、PCI 总线、串行 ATA（接硬盘、光驱）、PCIE 总线（接声卡、RAID 卡、网卡等），都归南桥芯片控制。

南北桥间可随时进行数据传递，需要一条通道，这条通道就是南北桥总线。南北桥总线越宽，数据传输越便捷。

目前的北桥都支持双核、四核甚至更多核性能更高的处理器。

随着技术的发展和摩尔定律的继续有效，PC 机这种双桥结构形式可望在不久的将来就会消失。到那时，处理器、北桥、南桥以至于内存条会集成在一块集成芯表中，构成人们所谓的片上系统（SOC），也有人称之为系统芯片。届时，PC 机可能只需要一块 SOC，在此芯片上接上外设就可以构成。为此，在本书的最后一章将向读者介绍如何设计处理器及 SOC。

### 1.2.2 软件系统

对计算机而言，只有上面提到的硬件系统，计算机是不能工作的。必须配上软件计算机才能工作。

计算机软件通常是指计算机所配置的各类程序和文件，由于它们是存放在内存或外存中的二进制编码信息，是不能直接触摸而且修改相对比较容易，故称之为软件。在计算机系统中，各种软件相互配合，很好地支持计算机有条不紊地工作，这一系列软件就构成了计算机的软件系统。软件系统一般包括两大部分：系统软件和应用软件。

#### 1.2.2.1 系统软件

系统软件是一系列保障计算机很好地运行的程序的集合。它们的功能是对系统的各种资源（硬件和软件）进行管理和调度，使计算机能有条不紊地工作，为用户提供有效地服务，充分发挥其效能。系统软件包括：

##### 1. 操作系统

操作系统是最重要的系统软件，它是管理计算机硬、软件资源、控制程序运行、改善人机交互并为应用软件提供支持的一种软件。通常，操作系统包括五大功能：处理器管理、存储管理、文件管理、设备管理及作业管理。

##### 2. 语言处理程序

每一台计算机都会配置多种语言以利于用户使用。从各种高级语言到汇编语言均会涉及到。当用户使用某种语言编写程序后，在该语言编译程序的支持下，可将用户的源程序转换为目的程序。

##### 3. 通用程序

在操作系统下提供给用户使用的一些通用程序。最常见的如在 Windows 下的 Word、PSW、Excel 等等。

##### 4. 各种服务支持软件

这是一些帮助用户使用和维护计算机的软件。例如，各种调试程序、诊断程序、提示警告程序等。

#### 1.2.2.2 应用软件

应用软件是指用户在各自的应用中，为解决自己的有关问题而编写的程序。这是一类直接由用户的需求为目标的程序。由于用户的多样性（各行各业、各种部门）和用户需求的多样性，使得这类软件也具有非常多样性。例如，用于科学计算、信息管理、过程控制、武器装备等等方面的应用软件。

### 1.3 计算机的分类及性能描述

#### 1.3.1 计算机的分类

了解计算机的分类，有助于理解计算机的结构及工作原理。不同的时期，不同的角度，产生不同的分类方法。

##### 1.3.1.1 早期的计算机分类方法

在上世纪八十年代前，人们根据计算机的字长、规模、价格等指标，将计算机分为：微型机、小型机、中型机、大型机、巨型机。随着计算机的发展，现在，它们的界限已十分模糊。那一时期的小型机、中型机、大型机已经见不到了，而且很少再有人提到这些机型。这种分类方法基本上已不再使用。今天，人们经常提到的是个人计算机 PC、服务器或高性能计算机。尽管它们在性能上有差别，但并没有人做出严格的界定。而且，今天性能好的 PC 机也许比前些年的某些服务器性能还要好。

##### 1.3.1.2 以用途分类

###### 1. 按照计算机的用途可分为通用计算机和嵌入式计算机（专用计算机）

通用计算机的硬件系统及系统软件均由有关的计算机公司设计制造，其用途不是针对某一个或某一类用户的，而是可以满足许许多多用户的。例如，目前国内外广泛使用的台式 PC 机或笔记本计算机，用户可直接由市场购买，在厂家提供的软件支持下可工作。也许用户需要有配上少量的软件或硬件，即可满足用户的需求。

对于各种服务器或高性能计算机，它们具有更高的性能，可以适用于许多领域或部门的需求。它们也可以看成是通用计算机。

嵌入式计算机系统的定义可表述如下：以应用为目标，以计算机技术为基础，软硬件可裁剪，适应对功能、实时性、可靠性、安全、体积、重量、成本、功耗、环境、安装方式等方面有严格要求的专用计算机系统。

可见，嵌入式计算机是一种专用计算机，它广泛应用于工业企业、军事装备的测量与控制。这类计算机通常采用系统集成，即利用一些工业控制机各部件的提供商所提供的部件，进行硬件系统和软件系统集成来实现嵌入式计算机。对某些特殊要求的计算机，如要求体积特别小、工作温度特别高、震动将特别剧烈等等无法进行系统集成时，则需要由设计者从元器件开始设计嵌入式系统或者采用 SOC 进行嵌入式系统的设计。

###### 2. 按照计算机的用途通用计算机还可分为个人计算机 PC 和服务器

现在，许多计算机许多都在联网工作，甚至有人提出网络才是计算机。在联网工作中经常有两种结构：客户机/服务器（C / S）结构和浏览器/服务器（B / S）结构。



网络中，用做客户机或浏览器的多是各种型号的个人计算机，通常它们性能要低一些，满足用户要求即可。

服务器是用于实现某种服务的。例如，目前常见的有 WEB 服务器，FTP 服务器，MAIL 服务器，文件共享服务器，数据库应用服务器，域名服务器，网站的网关服务器，DNS 服务器，流媒体服务器等等。一般而言，对服务器的性能要求是会很高的。但也要根据服务器服务对象的要求来设计，目前高档的服务器是由多达千个处理器构成集群系统来实现，其速度超过万亿次。也有用性能好一些的 PC 机来充当最简单的服务器的。

### 1.3.1.3 Flynn 分类法

Flynn 分类法是按照计算机在执行程序的过程中信息流的特征进行分类的。在程序执行过程中存在三种信息流：

指令流 (IS) —— 机器执行的指令序列，它由存储器流入控制单元 (CU)；

数据流 (DS) —— 由指令流所使用的数据，包括输入数据、中间数据和结果。数据在处理单元 (PU) 中进行处理。

控制流 (CS) —— 指令流进入 CU，由 CU 产生一系列的控制流 (信号)，在控制流的控制下完成指令的功能。

Flynn 将计算机分为四类，如图 1.4 所示。

#### 1. 单指流单数据流 (SISD) 计算机

图 1.4 (a) 所示的就是单指令流单数据流计算机的示意图。该计算机是由单一一个控制单元，单一一个处理单元和单一主存储器组成。每次从存储器取一条指令到控制器，由控制器对指令译码产生控制信号，控制处理单元完成指令规定的功能。这是最简单的一类计算机，本书后面的章节将予以说明。

#### 2. 单指流多数据流 (SIMD) 计算机

单指流多数据流 (SIMD) 计算机的结构示意图如图 1.4 (b) 所示。它是由一个控制单元，多个处理单元和多个存储器模块组成。每次从存储器取一条指令到控制器，由控制器对指令译码产生多种相同控制信号，并将相同的控制命令分别下到多个处理单元上，控制多个处理单元执行相同的操作，完成这条指令对多个数据的处理。最终实现一条指令所规定的功能。由于该计算机是能够通过执行一条指令实现对多个数据的处理，这就是单指令流多数据流的由来。这类计算机本书后面的章节将予描述。

#### 3. 多指令流单数据流 (MISD) 计算机

这种计算机的含义是由多个控制单元同时执行多条指令对同一数据进行处理，其结构示意图如图 1.4 (c) 所示。但这种方式尚无实际的实例。

#### 4. 多指令流多数据流 (MIMD) 计算机

多指令流多数据流计算机的结构示意图如图 1.4 (d) 所示。它是由多个控制单元、多个处理单元和多个存储模块构成，实际上是由多处理机由某种方式连在一起构成的计算机系统，通常称为多处理机系统。这类计算机是各个处理机分别执行多条指令、处理多个数据，并行工作而实现某种功能。目前性能好的多核处理器、多计算机的集群系统都属于这一类计算机，本书后面的章节将予描述。

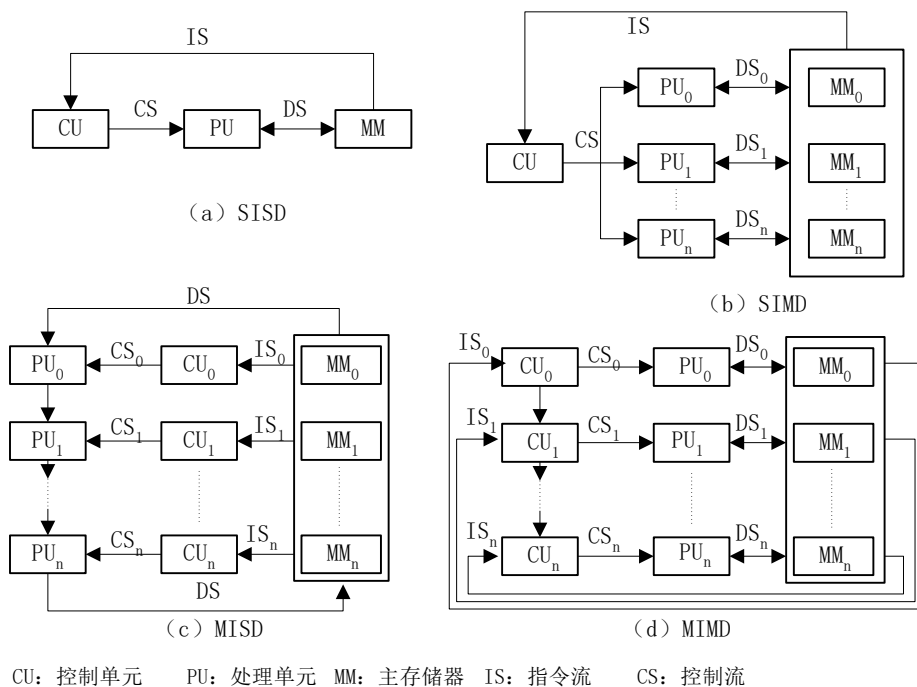


图1.4 Flynn分类法各类计算机示意图

计算机的分类方法还有诸如以并行度为准则的冯氏分类法，以并行度和流水线为准则的 **Handler** 分类法等，本书不再说明。

### 1.3.2 计算机的性能描述

每一处理器都有自己的性能指标，例如，某处理器是 32 位或 64 位处理器，时钟频率为多少 GHz，一级 Cache 和二级 Cache 各为多少，是否采用超流水、超标量技术，引出脚为多少，封装形式，集成度为多少亿个晶体管，工作电压及功耗，制造工艺多少 nm 等等。利用这些性能参数，人们可以大致了解这种处理器的性能。

由处理器（CPU）构成的计算机系统的性能，与所采用的处理器有很大关系。但是，并不完全由处理器来决定。计算机系统的性能还与构成计算机的系统总线、主存容量、外存容量、接口总线、外设以及系统软件的配置均有密切关系。因此，计算机性能应取决于该系统的硬、软件的配置。

人们所关心的计算机的性能包括速度、存储容量（内存及外存）、外设的配置、软件的配置、接口的配置、可靠性、可操作性、功耗、体积、重量、价格等。其中最重要的当属速度。

#### 1.3.2.1 计算机系统配置描述

通过计算机系统的配置，可以知道计算机系统的基本性能。下面从个人计算机 PC 和服务服务器分别做简单说明。

##### 1. 个人计算机配置

个人计算机最常见到的是台式（桌面）计算机和笔记本（便携式）计算机，它们包括多个厂家和多种型号。在这里我们随便列出一种台式机的配置，从配置中可以了解该机的大致性能。

- 处理器：英特尔<sup>®</sup> 酷睿<sup>™</sup>2 双核处理器 E6300  
（处理器速度 1.86GHz，2MB 二级高速缓存，1.066GHz 前端总线，硬件防病毒，65nm，节能技术）；
- 正版 Windows Vista Home Premium 简体中文版；
- 支持英特尔<sup>®</sup> 64 位内存扩展技术；
- 采用英特尔<sup>®</sup> LGA775 封装技术；
- 内存：2GB DDR II 533；
- 硬盘：SATA 250GB（7200 转）；
- 显卡：NV7950 256MB PCI-E 高性能显示卡
- 19" 宽屏液晶显示器；
- SATA DVD RW 光驱；
- 集成 100M 网卡；
- 集成 5.1 声卡；
- 防水抗菌键盘；
- 光电鼠标；
- MTBF 6 万小时。

对于台式 PC 机而言，尽管在配置上还有一些配置没有说明，但就上述配置已经可以为我们提供该计算机的性能。该计算机采用的是高性能 32 位英特尔公司的双核 Core 处理器。内存比较大，达到 2GB。显存 256MB 也较大，显示控制器性能比较好。具有串行 ATA 接口、容量达 500GB 的硬盘，平均故障间隔时间（MTBF）高达 60000 小时等等。可以想象，该计算机一定会有多个 SUB、1394 接口，还可以在这些接口上接上其他用户所需要的外设。我们认为上面所列出的个人计算机的配置，就是该计算机的主要性能参数。作为个人计算机，上述配置使整个计算机的性能相对比较好。

2. 服务器的配置

前面已经提到，不同用途服务器的性能相差非常多，最简单的服务器可用一台一般的 PC 机配上相应的应用软件就可以实现，功能强的服务器可能需要成百上千个处理器构成集群系统来实现。下面表 1.2 同样是随便列出一种服务器的配置，从配置中可以了解该服务器的大致性能。

表 1.2 某服务器的基本配置

基本类别	
类型	企业级
类别	机架式
结构	4U
处理器	
CPU 类型	Xeon MP E7430
CPU 频率	2130MHz
处理器描述	标配 2 个 Xeon MP E7430 处理器

最大处理器数量	4
制程工艺	45 纳米
CPU 核心	四核（Dunnington）
<b>主板</b>	
FSB（总线）	1066MHz
扩展槽	7 个 PCI-Express 插槽(4 个 x8,3 个 x4)
<b>内存</b>	
内存类型	FB-DIMM
内存大小	2GB*16
最大内存容量	256GB
<b>存储</b>	
硬盘大小	3*300GB
硬盘类型	SAS
硬盘最大容量	1.5TB*5
内部硬盘架数	内置硬盘托架可支持 5 块 3.5 英寸 SAS 热插拔硬盘、内置硬盘托架可支持 8 块 2.5 英寸 SAS 热插拔硬盘
磁盘阵列卡	PERC 6/iR 支持 RAID0、1、5、6、10
光驱	DVD-ROM
<b>网络</b>	
网络控制器	4 个嵌入式 Broadcom NetXtreme IITM5708 Gigabit5 以太网卡,具有故障恢复和负载平衡功能
<b>显示性能</b>	
显示芯片	ATI-Radeon ES1000 集成显卡,含 32MB SDRAM
<b>接口类型</b>	
标准接口	后部:2 个通用串行总线(USB) 2.0 端口,9 针串行端口,视频接口,适合 DRAC 的千兆以太网接口,带有蓝/橙色 LED 指示灯的 ID 按钮、前部:2 个通用串行总线(USB) 2.0 端口,带有系统警报滚动显示屏的 LCD 面板, 15 针视频接口,系统电源开/关按钮、内部:1 个通用串行总线(USB) 2.0 端口
<b>其他参数</b>	
散热系统	热插拔冗余冷却风扇
<b>电源性能</b>	
电源	热插拔冗余电源
电源数量	2
电压	180-240 VAC/90V-180 VAC
功率（W）	1570/1030
<b>外观特征</b>	

尺寸	172.7*447*701mm
重量	41.73Kg
软件系统	
系统支持	Microsoft Windows Server 2003 R2 SP2 企业版 Microsoft Windows Server 2003 R2 SP2 标准版 Red Hat Linux Enterprise v5,高级平台 SUSE Linux Enterprise Server 10 EM64T SP2
适用环境	
工作温度	10℃-35℃
工作湿度	8%-85%(无冷凝)
储存温度	-40℃-65℃
储存湿度	5%-95%(无冷凝)

根据上述服务器的配置信息，可以想象，其处理速度和处理能力要比上面的个人计算机高出很多。

### 3. 高性能计算机

不同时期，对高性能计算机有不同的解释。目前，一般认为性能过到或起过万亿次的计算机为高性能计算机。例如，天河一号的一些配置如下：

这套系统采用了采用混合异构的环境：主计算系统为 Intel 至强 5500CPU 和 ATI 的 GPU。CPU 和 GPU 数目分别为 6144 个和 5120 个，峰值性能可以达到 1.206PFlops。内存总容量 98TB，点对点通信带宽每秒 40Gb，共享磁盘容量为 1PB。该机由 103 个机柜组成、占地面积近千平方米、总重量达 155 吨。

#### 1.3.2.2 每秒钟执行指令的百万条数 MIPS

用计算机平均每秒钟能执行多少百万条指令来衡量计算机的性能，对于一个给定的程序，其定义如式（1-1）所示：

$$\text{MIPS} = \frac{\text{指令条数}}{\text{执行时间} \times 10^6} \quad (1-1)$$

可以看到，计算机执行指令的速度愈快，其 MIPS 就愈大。同时可以想象，同一计算机执行不同的程序（指令），可能会得到不同的 MIPS。尽管如此，该指标仍经常用来描述计算机的性能。如某计算机的速度达到 2000MIPS。

#### 1.3.2.3 每秒钟执行浮点数的百万次操作的数量 MFLOPS

另一种描述计算机性能的指标就是每秒钟执行浮点数的百万次操作的数量 MFLOPS，定义如（1-2）式：

$$\text{MFLOPS} = \frac{\text{浮点操作次数}}{\text{执行时间} \times 10^6} \quad (1-2)$$

显然，式（1-2）所示的标准是用来描述计算机执行浮点操作的能力，它也是

从这一侧面来描述计算机的性能，必然也带有一定的局限性。但在现实中也会看到有人用 MFLOPS 来说明计算机的性能。

#### 1.3.2.4 用测试程序来测评计算机的性能

##### 1. 关于基准测试程序

以往对计算机的测试采用过如下几种程序：

①实际应用程序，也就是计算机工作的真实程序。

②修正的实际应用程序，即对上面真实程序进行某些修改构成测试程序。

③核心程序，提取真实程序中的核心部分构成测试程序。

④小测试程序，具有特定目的的、100 行以内的测试程序。

⑤合成测试程序，就是选择具有各种代表性的一系列测试程序，将它们组合在一起。这种测试程序的集合称为测试程序组件，或称合成测试程序，或称基准测试程序。

利用基准测试程序的优点可以避免单个测试程序的片面性，可以更加全面地测试计算机的性能。因此，目前利用合成测试程序进行计算机性能评估已广泛被采用。

目前常见的基准测试程序有：

**TPC-C:** 对系统在线处理事务的能力进行评价。以每分钟处理新订单个数 (tpmC 为单位)。

**TPC-H:** 对系统在线数据库资料的查询能力进行评价。以每小时完成查询的数量。

**SPEC web2005:** 用于评价系统同时响应 http 连接的最大数量。

**SPEC jApp Server2004:** 用于评价系统基于 java 平台每秒钟所完成的 java 操作的最大数量。

**SPEC CPU2000:** 用于对特定程序包执行时的评估，

**Linpack:** 在每秒钟内，利用高斯消元法求解一元 N 次线性方程组的次数来评价系统的性能。

**HPCC:** 利用双精度矩阵乘法、傅立叶变换、并行矩阵转置等七个子项全面评价系统的性能。

**SAP SD:** 测试系统的响应时间及每小时完成的定单数，用以衡量系统同时执行应用程序及数据库的能力。

限于篇幅，上述基准测试程序不再详细说明。下面就 SPEC CPU200 基准测试程序予以说明。

##### 2. SPEC CPU200 基准测试程序

SPEC 成立于 1988 年，全称最初是 System Performance Evaluation Cooperative (即系统性能评估合作社)，现在已经更名为 Standard Performance Evaluation Corporation (即系统性能评估公司)。先后开发出一系列的测试程序，其主要版本有：SPEC CPU 89、SPEC CPU 92、SPEC CPU95、SPEC CPU2000 等。

CPU2000 是 SPEC 提供的 CPU 性能评估测试包，测试的源程序包含了整数和浮点两大部分，即 CINT2000 和 CFP2000。

##### 3. PC 机性能测试

上面提到的基准测试程序，在过去多用于测试大型的计算机的性能。今天，也可以用这些基准测试程序测试 PC 机。但是，在测试 PC 机时人们常用另外一些基准

程序。

对 PC 机的测试分为三部分，第一部分是 CPU 基本性能测试，通过测试软件对 CPU 的性能作一个基本了解。第二部分是基准测试软件测试。使用的测试工具大多是模拟了真实的软件工作环境，所以测试得分客观地反映 CPU 在实际应用中的性能差异。第三部分是实际应用软件性能测试，通过使用现实生活中常用的软件，记录在实际应用中的测试结果，以最直接的方式体现 CPU 的处理速度。

#### (1) CPU 基本性能测试

对 CPU 基本性能测试采用如下一些程序：

- PC Mark2002：整机综合性能测试软件，其中包含对 CPU、内存、硬盘等子系统的性能测试。

- Super pi（104 万位）：将圆周率计算到 104 万位。这一科学计算软件不但对 CPU 处理能力要求很高，而且测试结果受内存带宽和处理速度影响很大，是检验 CPU、内存和主板北桥芯片内存控制器的常用基准测试工具。

#### (2) 基准测试软件测试

##### ① 办公应用

- Business Winstone2001。

##### ② 网络/多媒体创作

- CC Winstone2002。

##### ③ 3D 游戏处理性能

- 3D Mark2001 SE(330 版)。

- Quake3 1.17Ver。

- Serious Sam-second Encounter。

##### ④ 3D 图形性能

- Cinema 4D XL V6.103：。

#### (3) 实际应用软件测试

##### ① MPEG4 编码

- FlacK MPEG0.594+DivX5.0.2。

##### ② MP3 编码

- Lame3.89 alpha。

##### ③ 文件压缩

- Winzip 8.1。

### 1.3.2.5 与 CPU 性能有关的几个问题

在这里还要讨论涉及到 CPU 性能的几个问题。

#### 1. 响应时间

响应时间是指计算机完成某一任务所耗费的全部时间。显然，响应时间里包括为完成任务对磁盘的访问、内存的访问、I/O 操作、操作系统的开销等。两台计算机 A 和 B，如果说 A 比 B 快，就是说对于给定的任务，A 的响应时间比 B 的响应时间短。通常，计算机 A 比计算机 B 快 n 倍可用（1—3）式表示：

$$n = \frac{\text{计算机B的响应时间}}{\text{计算机A的响应时间}} \quad (1-3)$$

## 2. CPU 时间

CPU 时间表示 CPU 工作的时间，不包括 I/O 等待时间。因为，在多任务系统中当某一任务进入等待状态时，CPU 可以运行其他的任务。

通常，计算机的时钟频率是固定的，CPU 执行某一程序共用多少时钟周期是可以测出来的（不包括 I/O 等待时间）。因此，CPU 时间可以用如下两种形式表示：

$$\begin{aligned} & \text{CPU 时间} = \text{总的时钟周期数} \times \text{时钟周期长度} \\ \text{或} & \quad \text{CPU 时间} = \text{总的时钟周期数} / \text{时钟频率} \end{aligned} \quad (1-4)$$

另外，我们一定能够知道上面某一程序有多少条指令，假定上面的程序有 IC 条指令，执行用的总时钟周期数也是知道的，则就可以计算出一条指令从头到尾处理完的平均时钟周期数 CPI：

$$\text{CPI} = \text{总的时钟周期数} / \text{IC} \quad (1-5)$$

将 (1-5) 代入 (1-4) 还可以得到 CPU 时间的另外的表达式：

$$\begin{aligned} & \text{CPU 时间} = \text{IC} \times \text{CPI} \times \text{时钟周期长度} \\ \text{或} & \quad \text{CPU 时间} = \text{IC} \times \text{CPI} / \text{时钟频率} \end{aligned} \quad (1-6)$$

从上 (1-6) 式可以看到，影响 CPU 的性能的因素是：①CPU 的时钟频率，它受 CPU 硬件工艺及结构的影响。②完成某任务所需要的指令条数，这与 CPU 的指令集和编译技术有关。③每条指令执行所需时钟周期数，它主要由 CPU 的组织结构来决定。因此，为提高 CPU 的性能，可以从改进和提高上述三种因素入手，提高它们的性能指标，达到改进 CPU 性能的目的。

假定在计算机系统中有 n 种指令，其中第 i 种指令的处理时间为  $\text{CPI}_i$ ，在程序中第 i 种指令的出现次数为  $\text{IC}_i$ ，则程序执行的 CPU 时间可用下 (1-7) 式计算：

$$\text{CPU 时间} = \frac{\sum_{i=1}^n (\text{CPI}_i \times \text{IC}_i)}{\text{时钟频率}} = \sum_{i=1}^n (\text{CPI}_i \times \text{IC}_i) \times \text{时钟周期长度} \quad (1-7)$$

将 (1-6) 和 (1-7) 合并，可以得到：

式 (1-8) 中， $\text{IC}_i / \text{IC}$  表示第 i 种指令在程序中所占的比例。

利用上面所描述的概念及算式，可以描述 CPU 的性能。

$$\text{CPI} = \frac{\sum_{i=1}^n (\text{CPI}_i \times \text{IC}_i)}{\text{IC}} = \sum_{i=1}^n (\text{CPI}_i \times \frac{\text{IC}_i}{\text{IC}}) \quad (1-8)$$

通常认为系统性能指的是响应时间，而 CPU 性能对应于 CPU 时间。

## 3. Amdahl 定律

Amdahl 定律是上世纪六十年代由 IBM 360 系列计算机的主要设计者 Amdahl 提出来的。其内容为：计算机系统中某一部件由于采用某种更快的执行方式后，整个系统性能的提高与这种执行方式的使用频率或占总执行时间的比例有关。Amdahl 定律给出了加速比的定义如下 (1-9) 所示：

$$\begin{aligned} \text{加速比} &= \frac{\text{改进后的系统性能}}{\text{改进前的系统性能}} \\ &= \frac{\text{改进前的系统总执行时间}}{\text{改进后的系统总执行时间}} \end{aligned} \quad (1-9)$$



从 Amdahl 定律所描述的内容可以看到，通过对计算机某一部分的改进，计算机比改进前快了多少倍，即加速比。计算机系统的加速比取决于下面两个因素：

①可改进部分在原系统总执行时间中所占的比，称为可改进比例，用  $f_e$  表示。例如，程序的总执行时间为 100s，可改进的部分是其中的 20s，则  $f_e=0.2$ 。可见， $f_e$  总是小于或等 1 的。

②可改进部分改进后性能提高的程度，通常用部件加速比  $r_e$  来表示某部件改进后性能提高的比例。例如，某部件改进后，执行时间由原来的 20s 减少到 5s，则部件加速比  $r_e=20/5=4$ 。可见， $r_e$  一般是大于 1 的。

根据上述分析，若假设改进前的系统总执行时间为  $T_0$ ，可以得出改进后的系统总执行时间  $T_n$  为：

$$T_n = T_0 (1 - f_e + \frac{f_e}{r_e}) \quad (1-10)$$

若加速比用  $S_p$  表示，根据 (1-9) 和 (1-10)，则加速比  $S_p$  可表示为：

$$S_p = \frac{1}{(1 - f_e) + \frac{f_e}{r_e}} \quad (1-11)$$

式 (1-11) 中， $(1 - f_e)$  为不可改进的部分，当可改进部分为 0 时，那没有可改进的了，则系统的加速比  $S_p$  就是 1。随着可改进部分的增加 ( $f_e$  加大) 和改进效果的提高 ( $r_e$  增加)，系统的加速比  $S_p$  就会增加。当系统可改进的部分  $f_e$  确定后，即使这一部分改进后不再需要时间，即  $r_e \rightarrow \infty$ ，则  $S_p = 1 / (1 - f_e)$ 。从而可见，系统性能的改善受可改进部分  $f_e$  的限制。

例 1.1 某计算机系统的某一部件的处理时间为总处理时间的 40%，该部件改进后部件加速比为 10，试计算改进后系统的加速比  $S_p$  为多少。

解：由上述题意可知， $f_e=0.4$ ， $r_e=10$ ，则

$$S_p = 1 / [(1 - 0.4) + 0.4/10] \approx 1.56$$

由计算可见，即使某一部件的部件加速比已达 10 倍，但该部件仅影响到总执行时间的小部分，对整个计算机系统的贡献是有限的。所以，改进后系统的加速比只有 1.5 倍左右。

例 1.2 若计算机系统有三个部件 a、b、c 是可改进的，它们的部件加速比分别为 30、30、20。它们在总执行时间中所占的比例分别是 30%、30%、20%。试计算这三部件同时改进后系统的加速比。

解：多个部件可同时改进的情况下，Amdahl 定律可表示为：

$$S_p = \frac{1}{(1 - \sum f_e) + (\sum f_e / r_e)} \quad (1-12)$$

将已知条件代入式 (1-12)，计算出  $S_p \approx 4.35$

## 习 题

1.1 第一代计算机所采用器件是\_\_\_\_\_；第二代计算机所采用器件是\_\_\_\_\_；第三代计算机所采用器件是\_\_\_\_\_；第四代计算机所采用器件是\_\_\_\_\_。

1.2 参见图 1.1，在计算机中有两种信息在流动。一种是\_\_\_\_\_，即操作命令，它是由\_\_\_\_\_产生流向各个部件。另一种是\_\_\_\_\_，它在计算机中被加工处理。

1.3 计算机软件通常分为两大类，一类是\_\_\_\_\_；另一类是\_\_\_\_\_。操作系统属于\_\_\_\_\_。

1.4 描述摩尔定律并说明在最近的一、二十年里摩尔定律得以延续的理由。

1.5 早期的冯·诺依曼计算机硬件系统主要由哪几部分组成？各部分的功能是什么？

1.6 叙述冯·诺依曼计算机的基本工作过程。

1.7 说明 Flynn 对计算机分类的依据，并描述不同类型计算机的特点。

1.8 为大致了解某种处理器（CPU）的性能，人们应注意哪些性能参数？

1.9 描述一台计算机的性能主要从哪些方面来说明？

1.10 某 500MHz 计算机，执行标准测试程序，程序中的指令类型、数量及指令执行的平均周期数如下表 1.3 所示：

表 1.3 题 1.10 所用表

指令类型	指令数量	指令执行周期数
整数	50000	1
数据传送	80000	2
浮点	10000	4
控制传送	5000	2

求该计算机的有效 CPI，MIPS 及程序的执行时间。

1.11 目前常见的基准测试程序有哪些？

1.12 用于对 PC 机的性能测试分为哪三部分？

1.13 若计算机系统有三个部件 a、b、c 是可改进的，它们的部件加速比分别为 30、30、20。若部件 a 和 b 在总执行时间中所占的比例分别是 30%、30%。若要使整个系统的加速比达到 10，部件 c 在总执行时间中所占的比例应为多少？